

Table of Contents

Emcien Patterns Quickstart	4
Installing Emcien	13
Installing to a VM	13
Installing Emcien Software with Oracle VirtualBox	13
Installing Emcien with VMware vCenter	20
Enabling Virtualization	28
Emcien Cloud POC	30
Configuring SFTP for Open VM	37
Emcien Quick Requirements Guide	40
Troubleshooting your VM installation	43
Using the Emcien Software	46
How to Analyze Data with Emcien	46
Features	55
Connections Map	55
Perspectives	56
Substitutions	61
Clusters	63
Predictive Rules and Analysis	68
Creating Prediction Rules	68
Prediction Rules	71
Predictive Analysis	74
Prediction Dashboard	77
Search	80
Anomalies	91
Resources	103
Preparing Data	103
Preparing Your Data	103
Common Pitfalls	109
Adding Derivative Data	111
Using Bandit to Prepare Data	113
Bandit Commands Reference	121
Implementing Band and Predict	123
Using the Software	133
Loading Your Data	133
Analyze a Data Set	137
Opening a Report	141
Video Tutorials	143
Categories Video	143
Category Details Video	144
Clusters Video	145
Dashboard Video	146
Data Overview Video	147
Loading Data with FTP Video	148
Navigation Video	149
Predictors Video	150
Long Data Format Video	151
Substitutions Video	152

Wide Data Format Video	153
Sample Data Sets	154
Reports in Patterns via cURL	155
File Formatting	159
The Wide Format	159
The Long Format	164
The Tagged Format	177
Use Cases	179
Retail and Inventory	179
Product Recommendations	179
Inventory Management	200
Predictions	213
Sales/CRM: Predicting Wins & Losses	213
Making Predictions from a Range of Outcomes	227
Real-Time Predictions	238
Predicting Truck Brake System Failure	248
Predicting Titanic Survivors	253
Predicting Return on Assets	267
Telecom	278
Correlating Network Attack Data	278
Manufacturing	289
Improving Parts Management	289
Advanced	305
Administration	305
Administration Hierarchy	305
Administration	308
Relicensing Emcien	308
Setting Your Item URL	312
Company Management	315
Emcien APIs	329
Thank You	339
Contact Us	340
Release Notes	341
Emcien Release Notes	341
Bandit Release Notes	360
Previous Versions	362

Emcien Patterns Quickstart

Getting Started

If you are just starting out with Emcien, or would like to see the basics of how we analyze data, follow the links below for a quick tutorial.

Select

Use Case Data

Prep Automated

Analysis Review

Results

Select Use Case

To get started, choose what use case you would like to do with Emcien Patterns.

For the most part, use cases fall into one of two categories:

1. Find Relationships within a set of data

In this use case, we will use Emcien Patterns to examine a set of bank marketing data to determine what attributes lead to someone making a sale. We will look at different relationships and clusters between the data and illustrate how using Emcien Patterns can enable us to easily and quickly find connections within our data.

2. Make Predictions for a set of data

In this use case, we will use Emcien Patterns to find the unique patterns and connections in a dataset of bank marketing data with customer attribute data. We will then use that data set to find rules and predictors that would allow us to predict whether or not a sale will be made in a given situation when only having

the attribute data. This exercise will enable us to see how easy it is to use the connections that Emcien Patterns finds within data to predict future outcomes.

In this Walkthrough, we will demonstrate both use cases with the same data set. When you know the use case you'd like to continue with, scroll to the [top](#) of the page and click on 'Data Prep'.

Data Prep

Click the link below to download the sales data set.

[sales-data.csv.zip](#)

Before we move on, let's take a moment to point out some key attributes about the prepared data:

1. Notice that every row is a separate person. We call this our 'wide' format, where each row is a different "transaction" of data
2. Also, see how every column across the top is its own distinct category, and none of the categories are blank or repeat
3. Because the file contained numerical data, those columns were banded using our Banding tool, Bandit. For more details on how we band out data, follow along in the [Advanced Walkthrough](#)

After the data has been prepared and we understand the traits of good data, scroll to the [top](#) of the page and click on 'Automated Analysis'.

Automated Analysis

To begin, let's go to our home page and click on 'Analyze Data'

Home

[Analyze Data](#)

Default Project

Analysis Created

Sort: Created

[Sales Analysis V1](#)May 6
2 Days Ago**10,005**
Transactions**235**
Items[ACME truck parts](#)May 6
2 Days Ago**10,005**
Transactions**235**
Items

Next, select the file we want to analyze and scroll down

New Analysis Nick Admin

Data

Filter by Name

Name	File Size	Modified Time
banded_sales_data.wide.csv	9.00 MB	May 5 3:20PM
<input checked="" type="checkbox"/> sales-data.wide.csv	4.9 MB	Apr 6 5:16PM
titanic-upload.wide.csv	140 KB	May 7 9:42AM

Now, if you are doing Use Case 2 and running a prediction for Sale here, type 'Sale' into the predictions field. If you are just trying to find relationships within the data without making predictions, leave this field blank and hit 'Analyze'.

- Predictions

Prediction Category Name

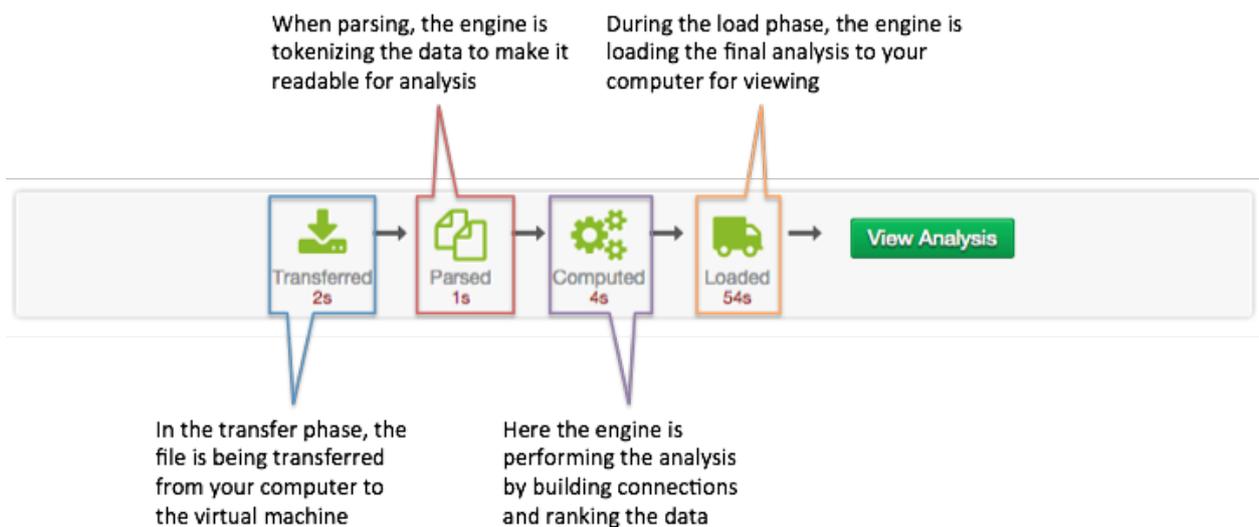
Prediction Category may have 20 items or less

+ Override Defaults

Report Name

Start Analysis

The engine will have animated steps displaying what it's doing as it tokenizes, connects, and ranks the data. The picture below describes what is happening in each step.



Now that the data has been analyzed, it's time to see the analyzed product. Scroll to the [top](#) of the page and click on 'Review Results'.

Review Results

You can easily navigate different views of the analysis using tabs

The engine found 22 categories (columns) and 119 items (values) within the data

The Connections map is a visual representation of the connections found within the data

The engine found 119 unique items (values) within the data set, and determined that 97% of them contained data in a meaningful connection

41,188 distinct transactions (rows) were analyzed, and there was a meaningful data connection in 100% of the rows

This is the home page of Emcien Patterns. The colorful chart in the middle is a visual representation of how the data is connected, and gives us an idea of what the engine found in the analysis. To explore the connections within the data a little, let's go ahead and click on the 'Clusters' tab at the top of the page to learn more.

Each of the 22 categories are shown here, with the percent of its items that are connected

This arrow chart illustrates how our selected category is connected with other categories, and gives the number of clusters we find in each

The connections page begins to tell us the story of how our data is connected. The blue boxes at the top represent categories, and are there to show you how items in the selected category are connected with items in other categories. In the picture above, we have selected the category "Age" and we can see how many clusters items in that category make with other categories (a cluster is a

connection between 2 or more items). Clicking on a 'Tell Me' Button will present each row in a more readable format. Let's scroll down to find out more about this category.

Each blue box represents an item in the data set, and the arrow shows the connection between it and another item or set of items

Column labels tell us more about the connection. Hover over the column name in your browser to learn more

Clicking on a 'tell me' box will give you an easily readable summary of the row

Age is connected with items from **Any Category** and has **72,540** clusters.

#	Items	Strength	Freq	Cnd Prob	Trend	
1	[17.0_to_32.0] age → single marital status	S	5,642	60.5%	N/A	🗨️
2	[17.0_to_32.0] age → no in default	S	8,330	89.3%	N/A	🗨️
3	[32.0_to_37.0] age → no in default	S	7,703	85.9%	N/A	🗨️
4	[17.0_to_32.0] age → cellular contact	S	6,357	68.1%	N/A	🗨️
5	[17.0_to_32.0] age → no personal loan	S	7,693	82.5%	N/A	🗨️

Further down the page we see the individual connections for items in this category. Each line represents a connection, where the item on the left suggests the presence of the item on the right. So for the first row in our example, when the item "[17.0-32.0] Age" is present in a row, it suggests that the item "single marital status" will be present in that row as well, with a conditional probability of 100%. After reading a few more lines to get an understanding of this category, let's scroll back to the top and select the 'Sale' category to learn more.

22 Categories By Strength | A - Z

- CCI 100%
- CPI 100%
- Sale 100%**
- age 100%
- contact 100%
- contacts this campaign 100%
- day_of_week 100%
- days since last contact 100%
- education 100%
- employment variation rate 100%
- euribor3m 100%
- housing loan 100%
- job 100%
- last contact duration 100%

Sale
2 of the 2 items are in clusters across 21 categories.
The number of items with connections make this category's connect

- Sale** ←
- job 22,298 Clusters
- month 25,842 Clusters
- nr.employed 21,482 Clusters
- employment variation rate 25,397 Clusters
- education 23,292 Clusters
- day_of_week 26,309 Clusters
- age 21,655 Clusters

Sale is connected with items from **Any Category** and has **145,964** clusters.

#	Items	Strength	Freq
1	no Sale → [7.0_to_999.0] days since last contact	VS	36,16
2	no ID → nonexistent previous campaign outcome no Sale	VS	32,42
3	Sale Category → no Sale	VS	32,42
4	103 Ranked Cnx 139,776 Clusters → no Sale	VS	32,42

Clicking on 'Sale' will bring up the category detail page

After clicking the Sale category, we can see the connections that lead to a 'Yes' or 'No' for the sale. Clicking the arrow inside of a 'no' box will allow us to go to the category detail page to learn more about the different outcomes of 'Sale'.

Category Details Sale

Sale contains:

Download CSV

#	Item	Freq	Int Cx	Rnk Cx	Clstrs	Trend
1	no Sale	36,548	117	103	139,776	N/A
2	yes Sale	4,640	115	109	6,188	N/A

The category detail page shows us the different values for 'Sale' that were found in the data set. To learn more about the patterns associated with a successful sale, click the 'yes' item.

Here is a graphical representation of the previous screen, showing our selected item ('yes' for sale) and all of the items it is most strongly connected with. By highlighting individual items we see the connection patterns and can begin to understand the data within our data set.

Now that we know the patterns and relationships within our sales data, we have finished our first use case. Now, if we'd like to use those connections to find the unique predictors for what causes a successful sale, click on 'View Predictors' on the right hand side of your screen.

Downloading the CSV file will give you a complete list of all predictive rules with their corresponding strengths and information

Predictors for yes Sale

Displaying **1,778** Predictive Clusters

#	Predictor Cluster	Probability	Item	Size	Freq
1	success previous campaign outcome + [94.027_to_94.767] CPI + [371.0_to_4918.0] last contact duration	90.1%	yes Sale	4	127
2	success previous campaign outcome + [-41.8_to_-36.4] CCI + [371.0_to_4918.0] last contact duration	89.4%	yes Sale	4	126
3	mar month + [0.0_to_7.0] days since last contact	86.0%	yes Sale	3	49
4	[5000.15_to_5013.1] nr_employed + [93.749_to_93.994] CPI + [371.0_to_4918.0] last contact duration	83.9%	yes Sale	4	99
5	success previous campaign outcome + [-41.8_to_-36.4] CCI + [224.0_to_371.0] last contact duration	83.9%	yes Sale	4	151

[Download CSV](#)

The Predictions screen shows us the individual combinations of items that, when found in a transaction together, yield a high probability of our predicted item being present in that transaction. Think of these as rules- when the items on the left are found in a row, there is the displayed probability that the item on the right will be present as well. To download a list of these rules, click the 'Download CSV' button in the top right.

Installing Emcien Software with Oracle VirtualBox

Emcien supports two different deployment methods, via the Cloud or Via an On-premise installation.

Cloud Access

The fastest way to get started with Emcien is to leverage [Emcien's Cloud solution](#) where you can log in immediately and begin analyzing your data. Contact our sales team if you don't have a login account yet and get started today!

On-Premise Installation

For those users who need to analyze their data on-premise this article will walk you through the process of installing Emcien on a local machine. With virtualization software, you will be able to run Emcien on almost any hardware, including laptops, desktops, and enterprise scale hardware.

Minimum Desktop or Laptop Requirements

Virtualization Software	The simplest option is to use the Emcien software on your laptop or desktop. We recommend Oracle's VirtualBox for easy local installation on laptops and smaller desktops.
-------------------------	--

Disk Space	500GB of Hard Drive Space
------------	---------------------------

Memory	8GB of RAM
Processors	4 CPU
Motherboard	Virtualization Technology is enabled. For more information on how to enable this on your system, click here

Step 1: Download the Emcien Software

[Click here to download the Emcien software.](#)

Step 2: Install your preferred virtualization software

If you do not already have a preferred virtualization software, Emcien recommends [VirtualBox](#). Downloading and installing VirtualBox will only take a few minutes.

Step 3: Install Emcien using VirtualBox

1. Unzip the emcien-x.x.zip file
2. Double click emcien-x.x.ova

1. Configure the number of CPUs and the amount of RAM you will allow Emcien to use.

CPU: The more CPUs Emcien can use, the faster it will process data. You can allow up to the number of CPUs on your laptop or desktop. This will slow down other applications (like Microsoft Word). We suggest that you use half of your laptop or desktop CPUs to ensure Emcien and other applications can function properly. For example, if you have 4 CPUs, give Emcien 2 CPUs.

RAM: The more RAM Emcien can access, the larger the data sets it can analyze. However, the memory you give Emcien will not be available to your computer while the EmcienVM is running, so do not specify more RAM than necessary. We suggest you give Emcien 50% of your laptop or desktop's RAM. For example, if you have 16GB of RAM, give Emcien 8GB.

2. Click Import.

Appliance settings

These are the virtual machines contained in the appliance and the suggested settings of the imported VirtualBox machines. You can change many of the properties shown by double-clicking on the items and disable others using the check boxes below.

Description	Configuration
Virtual System 1	
 Name	emcien-2.13
 Description	Emcien Virtual Appli...
 Guest OS Type	 Ubuntu (64 bit)
 CPU	2
 RAM	4096 MB

Reinitialize the MAC address of all network cards

Restore Defaults

Go Back

Import

Cancel

Step 4: Setting up your first user

Now that you have Emcien installed, it is time to start the software and create your first user.

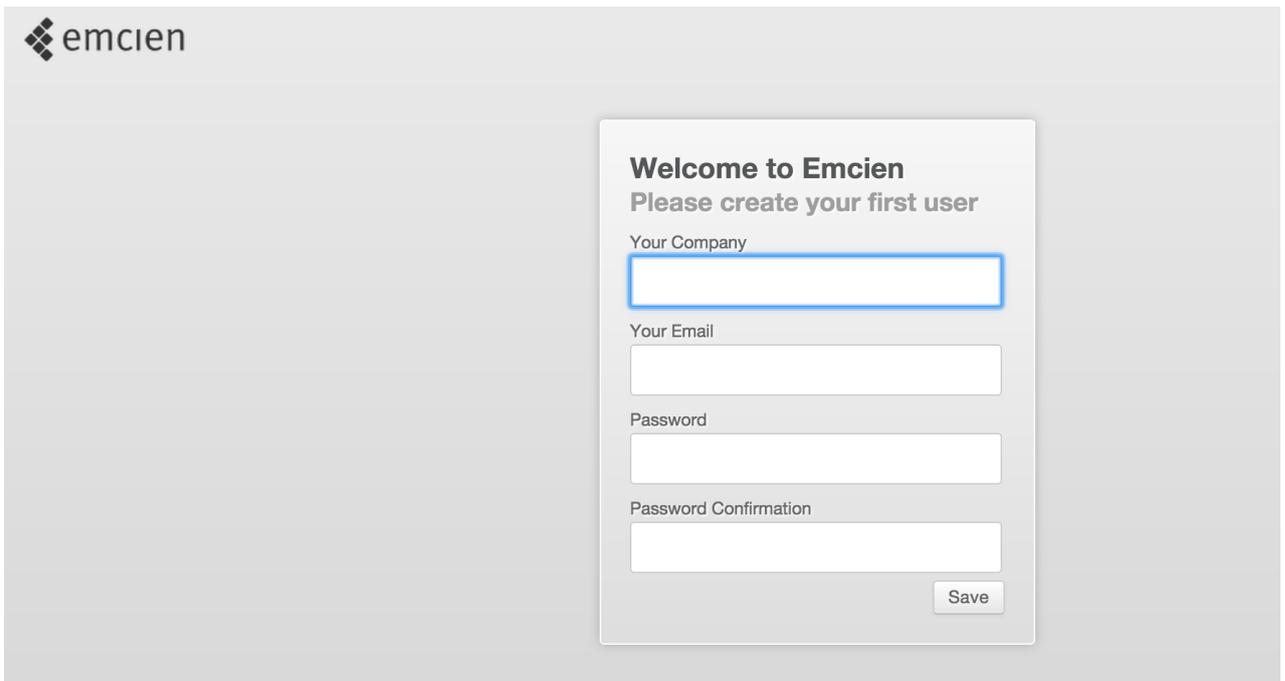
1. Click the Start Arrow



2. Clicking start will open a window with instructions to navigate to the software. In your web browser (ie. Google Chrome, Firefox, etc.) go to <http://localhost:5115>.

```
Welcome to Emcien
Open a browser to http://localhost:5115
IP Address: 10.0.2.15
MAC Address: 08:00:27:40:b3:3c
For support visit http://support.emcien.com
Copyright Emcien Corporation
localhost login: _
```

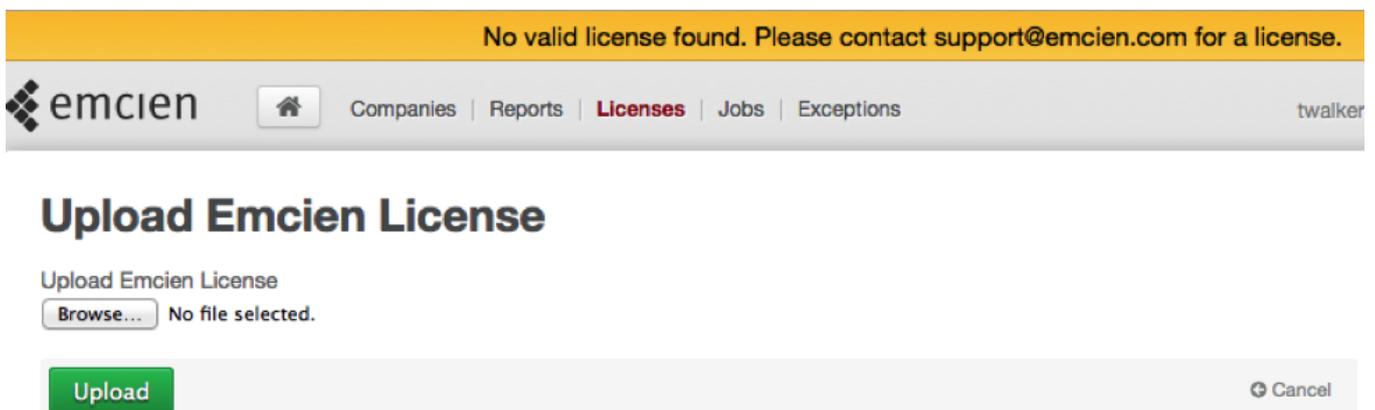
3. You can also click this link <http://localhost:5115> to open the EmcienVM software. Follow the instructions within the software to create your first Company and User.



Step 5: Licensing

You should already have the Emcien software installed and your first user created. The next step is to upload an Emcien License. If you do not have an Emcien License, please contact support@emcien.com.

Select the license file and click "Upload". This process may take as long as two minutes.



Congratulations!

Your Emcien Automated Analytics Platform is ready to use! For walkthroughs, help documentation, and instructional videos please visit <http://support.emcien.com>.

The EmcienVM is an Ubuntu Linux operating system. You will log into the VM software to manage some settings including hostnames, SSL certifications, etc. or manage backups. If you need a username and password please contact support@emcien.com.

Installing Emcien with VMware vCenter

Emcien supports two different deployment methods, via the Cloud or Via an On-premise installation.

Cloud Access

The fastest way to get started with Emcien is to leverage [Emcien's Cloud solution](#) where you can log in immediately and begin analyzing your data. Contact our sales team if you don't have a login account yet and get started today!

On-Premise Installation

For those users who need to analyze their data on-premise this article will walk you through the process of installing Emcien on a local machine.

This article walks you through the process of installing Emcien on VMware vCenter.

(NOTE: If you are using Oracle Virtualbox, click [here for VirtualBox installation instructions](#))

Generally, installation is only performed by IT team members and Emcien support.

Not an IT person? Emcien can help you get things set up for free. Contact us at support@emcien.com.

The VM ships as Version 7. Follow the directions [here](#) to upgrade to Version 8 or 9 for more CPU support.

This process includes the following steps:

[Minimum System Requirements](#)

[Installing the Emcien VM](#)

[Configuring Hardware Settings](#)

[Launching the Emcien VM](#)

[Licensing Emcien](#)

[Validating the Emcien VM](#)

Minimum System Requirements

To get started, you'll need a server that meets the following requirements:

RAM	16GB or more
Disk Space	500GB or more
Processor	Intel or AMD x86 with at least 8 cores, each with a speed of 2GHz or more
VM Software	VMware vCenter 5.1 or later

Virtualization Technology is enabled. For more information on how to enable this on your system, click [here](#)

Installing the Emcien VM

After ensuring your server meets the minimum requirements, follow the steps below to install the Emcien VM.

1. [Download the Emcien Open Virtualization Archive \(OVA\) file.](#)
2. Deploy the Emcien OVA file as an Open Virtualization Format (OVF) template in VMWare vCenter.

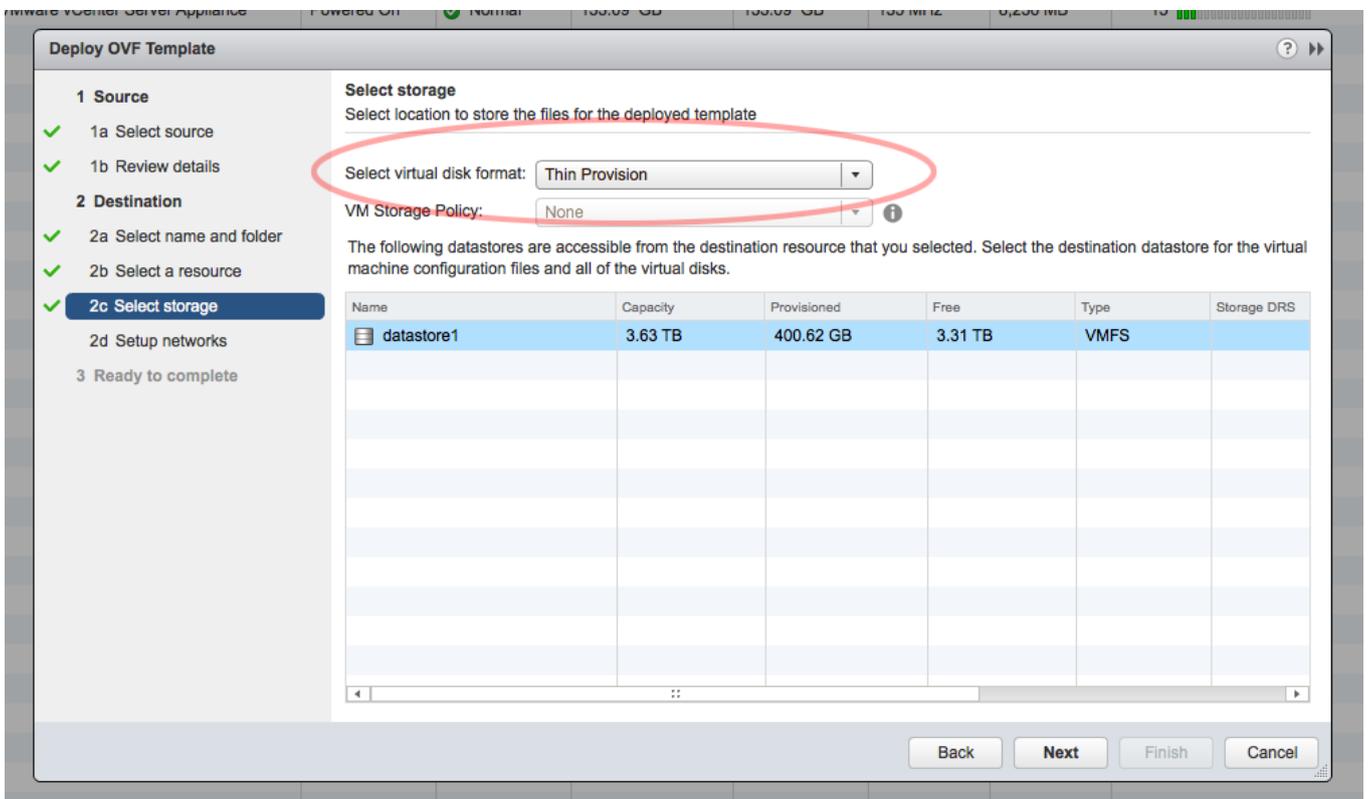
[Refer to VCenter's documentation for more information](#) about deploying OVF templates.

Configuring Hardware Settings

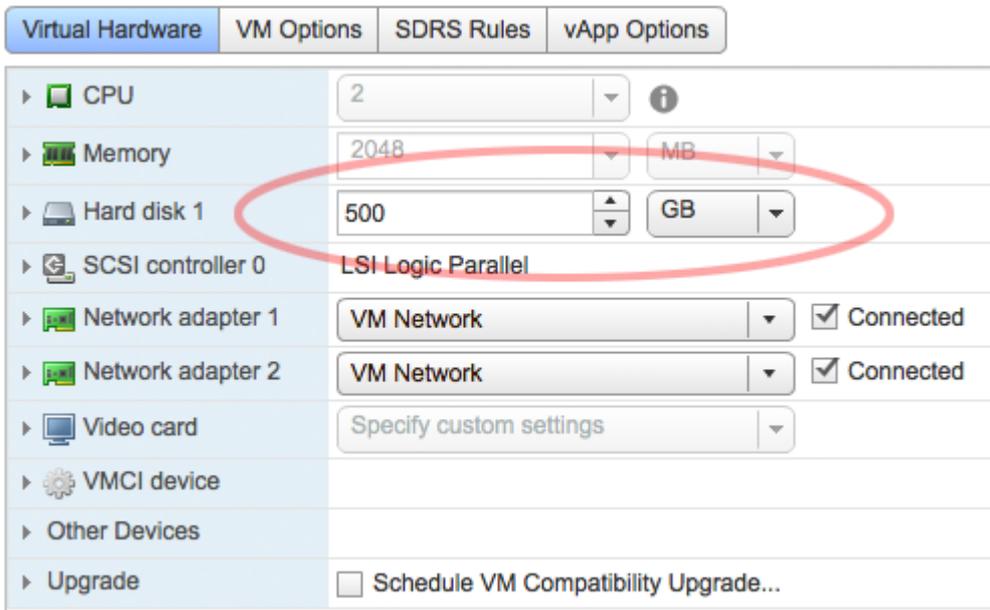
By default, the hard disk has an initial size of 100GB, we strongly recommend you increase the size of the data disk to at least 500GB or more. The disk size can be increased when the Emcien VM is powered off.

CPU	Defaults to 2. Supports 1 - 64. Increasing this will improve analysis speed.
Memory	Defaults to 2GB, suggested is 16GB for larger data sets. Increasing memory allows for larger data sets to be analyzed.
Data Disk	Ships with 100GB, increase to at least 500GB. Will need to be increased based on your data file sizes.

Installation Step: Please use "Thin Provision" storage. This means the Emcien VM will use hard drive space only when needed.



Configuration Step: When the VM is powered off, you can increase the disk space in the VM Hardware Edit Settings window. You can only increase disk space (you cannot decrease disk space).



Note: If you need to access the VM, you can email Emcien for a username and password for access.

Launching the Emcien VM

After configuring your hardware settings, deploy your VM.

Once started you will need to use your browser and go to the IP Address for the new VM provided by vCenter.

The screenshot shows the vCenter interface for a virtual machine named 'emcien-2.13-8641253'. The 'Summary' tab is active, displaying the VM's status as 'Powered Off'. To the right, system details are listed: Guest OS is Ubuntu Linux (64-bit), Compatibility is ESX/ESXi 4.0 and later (VM version 7), VMware Tools are not running, and the DNS Name is localhost. The IP Address is 10.1.1.140, which is circled in red. A link to 'View all 4 IP addresses' is provided below. The Host is 10.1.1.178. A 'Launch Console' button is visible at the bottom left of the VM details area.

emcien-2.13-8641253	
Guest OS:	Ubuntu Linux (64-bit)
Compatibility:	ESX/ESXi 4.0 and later (VM version 7)
VMware Tools:	Not running, version:2147483647 (Guest Managed)
DNS Name:	localhost
IP Addresses:	10.1.1.140 View all 4 IP addresses
Host:	10.1.1.178

NOTE: Use the IP address shown on YOUR vCenter screen, it will be different from the example shown above.

Once your browser refreshes, Emcien will show the first user creation screen to allow you to get started.

1. Enter a "Company" name where you analyzed results will be organized under.
2. Enter your email address that you will use to login with.
3. Enter your password (and confirmation password).



Welcome to Emcien
Please create your first user

Your Company

Your Email

Password

Password Confirmation

Save

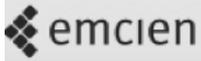
Licensing Emcien

After launching the Emcien VM, you can license your installation of Emcien. You must license your installation before you can begin using Emcien.

You should already have the Emcien software installed and your first user created. The next step is to upload an Emcien License. If you do not have an Emcien License, please contact support@emcien.com.

Select the license file and click “Upload”. This process may take as long as two minutes.

No valid license found. Please contact support@emcien.com for a license.



[Companies](#) | [Reports](#) | **[Licenses](#)** | [Jobs](#) | [Exceptions](#)

twalker

Upload Emcien License

Upload Emcien License

No file selected.

Congratulations!

Your Emcien Automated Analytics Platform is ready to use! For walkthroughs, help documentation, and instructional videos please visit <http://support.emcien.com>.

The Emcien VM is an Ubuntu Linux operating system. You will log into the VM software to manage some settings including hostnames, SSL certifications, etc. or manage backups. If you need a username and password please contact support@emcien.com.

Enabling Virtualization

Confirm that virtualization technology is enabled on your computer

Prior to running a virtual machine (VM) on your computer, the computer's virtualization technology (vt-x / amd-v) must be enabled in the motherboard BIOS. Virtualization technology is by default enabled on Intel-based Macs. If you are using a Windows computer, free utilities are available to help you determine if your computer processor supports hardware virtualization, as well as to check the virtualization setting to ensure it is enabled. Below we have provided links to two such utilities for Windows OS machines.

Windows OS computers with either Intel or AMD processor:

<http://www.microsoft.com/en-us/download/details.aspx?id=592>

Windows OS computer with an Intel processor: <https://downloadcenter.intel.com/download/7838>

If the virtualization technology is NOT enabled on your Windows machine, please refer to the manufacturer's support site for instructions on how to enable the vt-x or amd-v setting in the BIOS. Below are general instructions on steps to follow to enter the BIOS:

1. Boot the computer. As the manufacturer's logo appears, press the designated setup button to enter the BIOS. The setup key varies by manufacturer and often by machine model within a manufacturer. Typical setup keys are F2, F5, F10, F12 and Del. The specific key for your computer will normally display on the screen with the manufacturer's logo while the machine is booting up. The physical key on the key board will often have the manufacturer's logo printed on it.

2. Rapidly press the setup key as the computer boots up. If your computer boots very quickly, you may have to make a few attempts prior to getting the correct timing. After successfully hitting the setup key the BIOS will load.
3. The layout of the BIOS settings menu varies by manufacturer, but most share common settings. Note that BIOS menus Do Not support mouse input. The arrow keys can be used to navigate the menu.

Emcien Cloud POC

Loading Data to the Emcien Cloud

You can upload data files to Emcien using SFTP and your preferred FTP client, such as [FileZilla](#) or [Cyberduck](#).

To connect to Emcien using your FTP client, use the following credentials:

Host: feeds.emcien.com

Username: {Your Emcien Feeds server Username}

Password: {Your Emcien Feeds server Password}

Accessing your Emcien Cloud account for analysis.

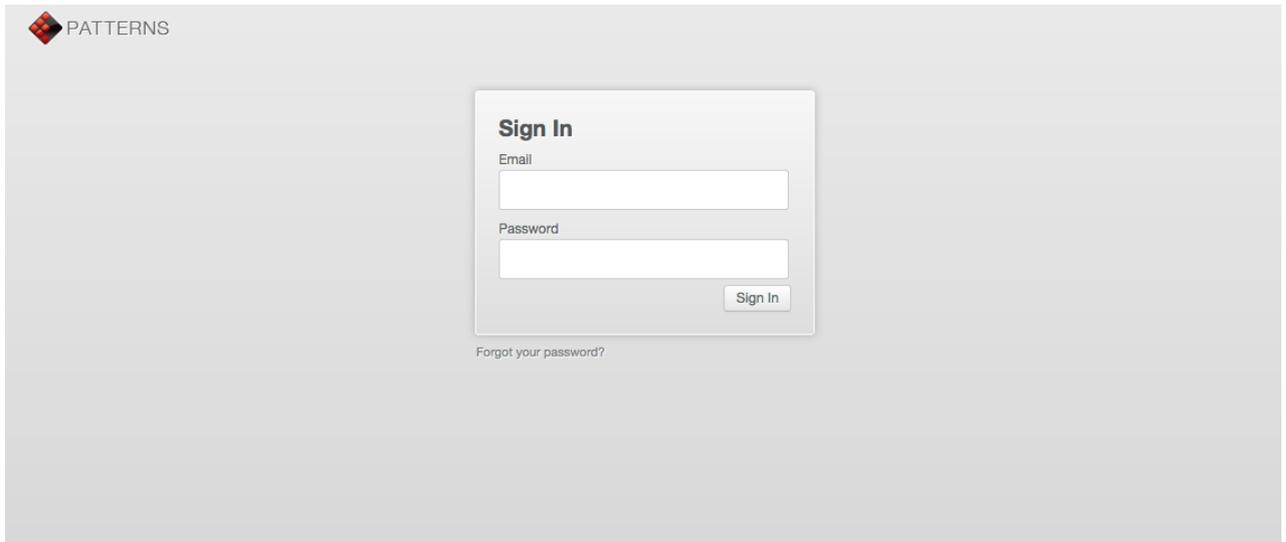
Emcien is a web-based application that works right in your favorite browser. To get started, just sign in. There's no need to download or install any software to your computer to use Emcien.

To sign in to Emcien follow the steps below:

1. Using your internet browser, navigate to the Sign In page.

For Emcien sandbox users, the URL for the Sign In page is <https://patterns-sandbox.emcien.com>

For Emcien Cloud account users the URL for the Sign In page is <http://patterns.emcien.com/>



2. Type your email address in the Email field.
3. Type your password in the Password field.
4. Click Sign In.

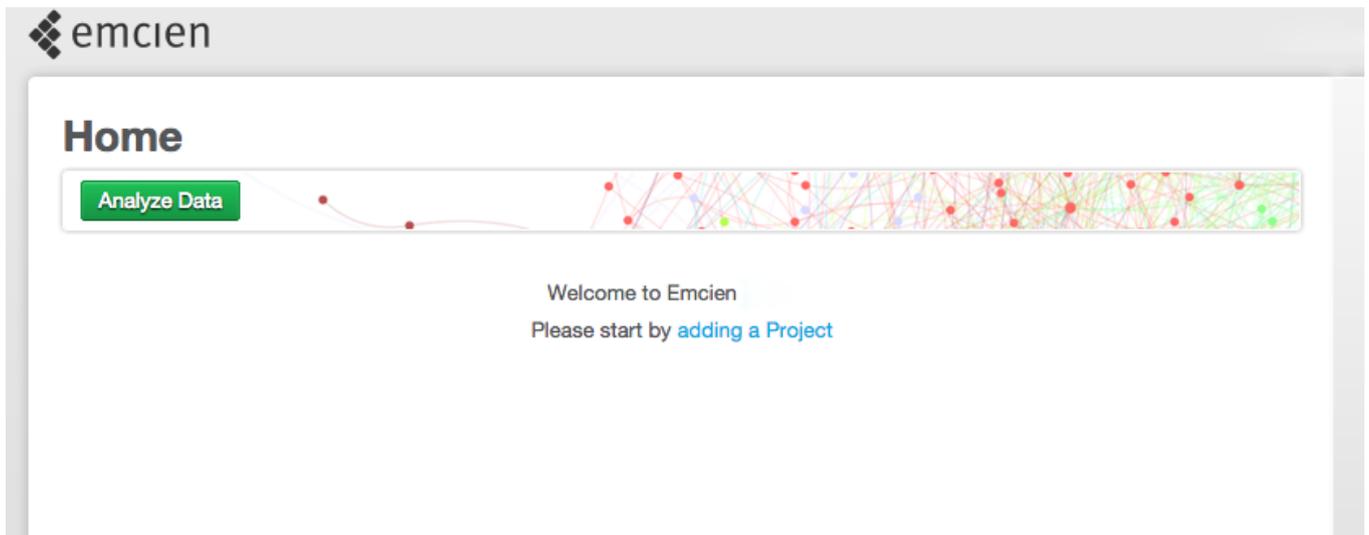
The homepage is displayed.

If your account is locked, contact your Emcien administrator or Emcien support at support@emcien.com.

Make sure you've prepared your data before uploading to Emcien. Check out the [Preparing Your Data](#) article for more information. You can also email us at support@emcien.com for help getting started.

Analyzing Data

Signing in to your Emcien Cloud account will bring you to the home screen.



To create a new report, click on the Analyze Data icon and the data files uploaded to Emcien will be displayed.

New Analysis ACME Admin

Data

Please select a Data File

Filter by Name

Name	File Size	Modified Time
ACME-Office-1M-Load2.csv	370 MB	Aug 27 3:04PM
Telecom-Churn-2.wide.csv	510 KB	Aug 27 3:47PM

Projects

Select a Project to display Report on Home Page

Office Supply

Override Defaults

Report Name

Start Analysis

Cancel

Click on the file that you would like to analyze and it will be updated with a check mark. The Report Name will auto-populate with the file name, but you can change the file name to another name if you prefer.

Override Defaults

The Override Defaults allows for changing the default settings used in the analysis of the data. If you would like to change the default report settings click on "Override Defaults".

Override Defaults

Allow Limited Invalid Lines
Report will load if less than 10% of the file lines are invalid

Cluster Threshold
Number of times a Cluster must occur

Maximum Cluster Size
Set the maximum Clusters size

File Encoding
Specify an encoding

Time Period
Auto option chooses the best period for the data

Partial File Import

Specify the Number of Lines to Import

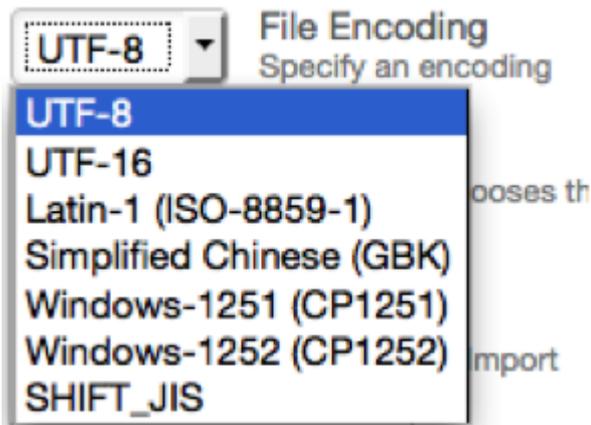
Messy Data

Enabling the option “Import with Limited Invalid Lines” will allow the application to process a data set that contains some invalid lines, as long as less than 10% of the entire file is invalid. This can be helpful if you are still in the process of modifying your input data.

Character Encoding

Data Files are interpreted as UTF-8 character encoding. However, non-UTF-8 data files with one of the listed character encodings can be processed if Patterns is informed of the specific encoding prior to processing. UTF-8 files with a BOM are

accepted using the default UTF-8 selection.



Naming Reports

The last step is to name the report. By default the system uses the filename of the data file. Make any necessary changes and then click on Start Analysis.

The Load Page

The Load Page allows the user to watch the analysis progress through the various phases.

#	Message	Time
49	--> Load Clusters starting	48s
48	--> Maps dumped	48s
47	==> Building Shard at patterns_services_rep27161984_v00048	46s
46	... Preconditions checked out	46s
45	Testing Preconditions...	46s

If one or more data files are already in process a message will display indicating how many Reports are in the queue in front of your Report:

ACME Office 1M Load Log Demo Admin



Please Note: Your report is 1st in line to be analyzed and will continue when resources are available.

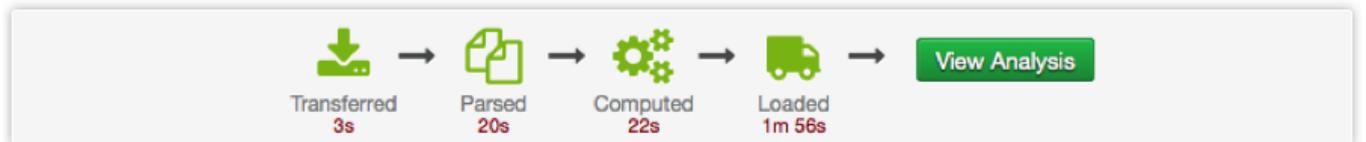
Log: [Summary](#) | **Expanded**

Run Time: **4s**

#	Message	Time
4	Found file ACME-Office-1M-Load.csv	1s
3	SFTP Connecting to Demo Data Server at HOST: localhost and PATH: /~/services/Test	1s
2	Report State: importing	1s
1	Creating New Report with file ACME-Office-1M-Load.csv for Emcien (ID: 1) by twalker@emcien.com	1s

Once the analysis has completed click on the View Analysis icon to view the Report.

ACME Office 1M Load Log Demo Admin

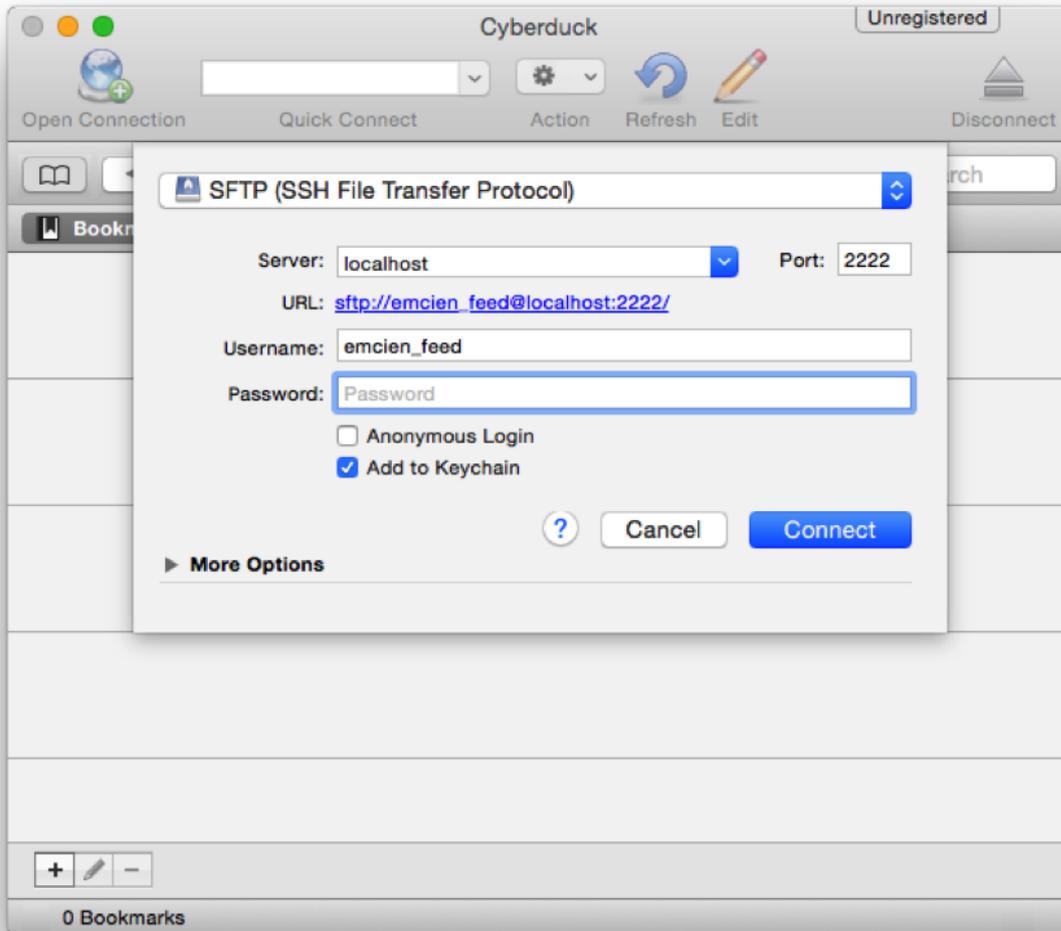


Configuring SFTP for Open VM

Below are the settings for configuring your FTP connection using a client such as Filezilla or Cyberduck.

Emcien OpenVM Information

Protocol	SFTP
Server	localhost
Port	2222
Username	emcien_feed
Password	



The password is already available in the Company settings.

Click on "System" on the upper right of the EmcienPatterns screen



Click on the Company name and then click on "Edit Company"



You will then see the path in the Data server which contains the password.

Data Server

Data Server must accept SFTP

```
sftp://emcien_feed:SAMPLE_PASSWORD@localhost/data
```

Format `sftp://USERNAME:PASSWORD@HOST/PATH`

Simply copy the password from your Data server path and paste it into the appropriate place in the FTP client.

Emcien Quick Requirements Guide

Step 1: Minimum Requirements

Performed by IT Resource

1. What do I need to get started?

VMware vCenter 5.1 or higher ([See here](#) for more details)

Compatible hardware:

At least 16GB RAM

At least 500 GB hard drive

Intel or AMD x86 processor with at least 4 logical cores, each with a speed of at least 2GHz

2. What will Emcien provide?

Emcien analytics software

Pre-loaded sample data sets

Bandit: Emcien's automated banding module (optional)

Step 2: Installing Emcien

Performed by IT Resources

1. How do I install the Emcien?

[Download file here](#)

Refer to the steps outlined here: [Installing Emcien](#)

Send your UUID to support@emcien.com to receive your Emcien license

2. What if I'm interested in the Bandit?

Download [Mac version](#) or [Linux version](#)

Step 3: Uploading Data to Emcien

Performed by IT Resource

How do I upload data into Emcien?

1. Follow [these instructions](#) to connect to Emcien using FTP on a Mac/Linux
2. Follow [these instructions](#) to connect to Emcien using FTP on a Windows

If you do not have a FTP application, we recommend [Cyberduck](#).

Step 4: Scheduling a Working Session

Performed by User

After I have set up my working session, what do I need to bring?

1,000 rows of sample data

Use cases of interest

Any and all questions you have for us

Troubleshooting your VM installation

Having problems with your VM?

This page will guide you through some of the common issues.

Issue: Virtualization not enabled in the PC BIOS (VT-x error)

This is a common setting issue for VirtualBox and other virtualization applications. Performing a Google search on the message will provide you with the steps to make the change needed on the BIOS of the PC.

Click [here](#) or [here](#) to go to a couple of the links.

Issue: License fails to apply

1. Make sure that the VM can access the internet by connecting to licensing.emcien.com in a browser on the VM.

You should see a login screen when you open the link:

Sign In

Email

Password

You may have to change your network settings in the VM console in order to connect. To login to the VM you will need to open the SYSTEM_PASSWORD.txt file that was included with the OVA in the zip file.

Still not able to resolve the license file upload issue?

When contacting Emcien [support](#) please make sure to include the Diagnostic logs from your Patterns VM. To run the logs in the application simply click on "System"

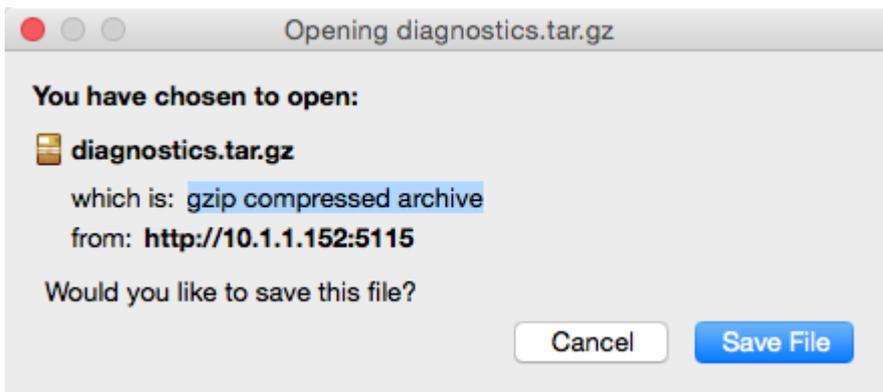
| [System](#) | [Admin](#) | [Settings](#) | [Help](#) | [Sign Out](#)

Then you will click on "Diagnostics"

[Companies](#) | [Reports](#) | [Licenses](#) | [Jobs](#) | [Exceptions](#) | **[Diagnostics](#)**

Click on the "Run all scripts" button

Once it has completed you will be prompted to save the file:



Click on "Save File" and copy it to your local PC. Email the compressed file to Emcien support.

How to Analyze Data with Emcien

Analysis with Emcien

If you're familiar with Emcien's software and would like a more detailed tutorial of how we work, click the links below for a walkthrough.

Select

Use Case Data

Prep Automated

Analysis Review

Results

Select Use Case

To get started, you should figure out what you're trying to accomplish using Emcien.

Emcien works great with a number of use cases, including:

[Sales/CRM: Predicting Wins & Losses](#)

[IT/Networking: Correlating Network Attack Data](#)

[Demographic Patterns: Predicting Titanic Survivors](#)

[FDIC Banking Data: Predicting Return on Assets \(ROA\)](#)

[Predicting Airline Delays \(more than two outcome variables\)](#)

[Using Emcien's Tier 2 Prediction Engine for Real-Time Predictions](#)

[Product Recommendations and Affinities](#)

Don't see your use case? We can help create a custom walkthrough for you. Just contact us at support@emcien.com.

Data Prep

While Emcien offers multiple file formats, the one named "wide format" is the more universal and most commonly used Emcien format. You can use the wide format for multi-dimensional data, such as demographics or configurable products. In the wide format, each transaction is identified by a single row of data.

An example of the wide format is displayed below.

Wide Format Example

User Defined Data

date	volume	ID	Age Group	Race	Birth State	Residence State	Gender	Education	Visits per month	...
2/3/13	1	C-1002010	25-35yrs	Asian	Birth-CA	Residence-FL	Male	High School	3-5 visits	
2/3/13	1	C-1002011	25-35yrs	Caucasian	Birth-OR	Residence-OR	Female	Graduate school	Less than 3	...
2/3/13	1	C-1005401	55-65yrs	Latin	Birth-GA	Residence-FL	Female	University	Over 10 visits	
2/3/13	1	C-1005660	55-65yrs	Caucasian	Birth-NY	Residence-CT	Female	University	Less than 3	

Volume is not a required column but must be in the first column if it exists and there is no date. If the date exists, volume must be in the second column.

Date is not required but must be in first column

Each row represents a Transaction

This example contains four transactions, each represented by a single row of data. In this data, each transaction represents a client and their associated demographics data.

Unlike the receipt format, the wide format has no required columns and supports user-defined columns.

Wide Format Details

Headers are required for each column in the wide format. Your file must contain at least two user-defined columns with a maximum of 1,000 total columns.

While the data is allowed to contain any UTF-8 characters the header must be in lower ASCII.

For all the details on the "Wide" and other data formats visit article for

[Preparing your data in an Emcien file format](#)

Depending on your data, Data Prep can include the following steps:

[Common Pitfalls](#)

[Numeric Data? Automatic Binning of data using Emcien Bandit](#)

Automated Analysis

To begin, let's go to our home page and click on 'Analyze Data'



The screenshot shows the Emcien web interface. At the top left is the 'emcien' logo and at the top right is the user 'nziech@'. The main heading is 'Home'. A green 'Analyze Data' button is highlighted with a red box. Below this is a table of 'Default Project' entries. The table has columns for the project name, 'Analysis Created' (date and time), 'Transactions', and 'Items'. There are two entries: 'Sales Analysis V1' and 'ACME truck parts', both created on May 6, 2 Days Ago, with 10,005 Transactions and 235 Items each. A 'Sort: Created' dropdown menu is visible in the top right of the table area.

Default Project	Analysis Created	Transactions	Items
Sales Analysis V1	May 6 2 Days Ago	10,005	235
ACME truck parts	May 6 2 Days Ago	10,005	235

Next, select the file we want to analyze and scroll down

New Analysis Nick Admin

Data			Filter by Name
Name	File Size	Modified Time	
banded_sales_data.wide.csv	9.00 MB	May 5 3:20PM	
✓ sales-data.wide.csv	4.9 MB	Apr 6 5:16PM	
titanic-upload.wide.csv	140 KB	May 7 9:42AM	

Now, if you are running a prediction analysis, enter the category for which you want to extract predictive rules for (ex. 'Sale' for the sample data file for CRM Sales data). If you are just trying to find relationships within the data without making predictions, leave this field blank and hit 'Analyze'.

Predictions

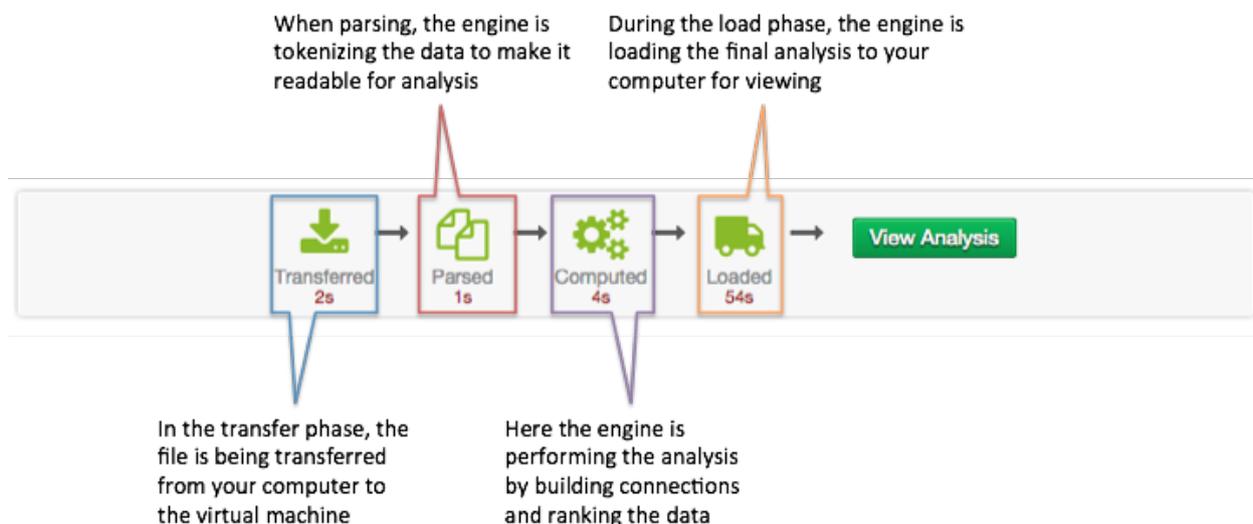
Prediction Category Name

Prediction Category may have 20 items or less

Override Defaults

Report Name

The engine will have animated steps displaying what it's doing as it tokenizes, connects, and ranks the data. The picture below describes what is happening in each step.



Now that the data has been analyzed, it's time to see the analyzed product. Scroll to the [top](#) of the page and click on 'Review Results'.

Review Results

Because Emcien analyzes all of the data, the results reflect a holistic view of your data and the patterns within it. Here is a quick breakdown of some of the main results you will be provided at the end of an analysis.

Dashboard

You can easily navigate different views of the analysis using tabs

The Connections map is a visual representation of the connections found within the data

The engine found 119 unique items (values) within the data set, and determined that 97% of them contained data in a meaningful connection

The engine found 22 categories (columns) and 119 items (values) within the data

41,188 distinct transactions (rows) were analyzed, and there was a meaningful data connection in 100% of the rows

Dashboard Sales Analysis V1

Connections Map

Overview

- 97% of the 119 Items with connections
- 100% of the 41,188 Transactions with clusters

Predicting: Sale

- 13K Predictions
- 2 Outcome Items
- 1 Predictions

Clusters Strength

Your items have ext strong clusters. 97% of the 119 items have clusters.

This is the home page of Emcien Patterns. The colorful chart in the middle is a visual representation of how the data is connected, and gives us an idea of what the engine found

in the analysis. To explore the connections within the data a little, let's go ahead and click on the 'Clusters' tab at the top of the page to learn more.

Clusters

Each of the 22 categories are shown here, with the percent of its items that are connected

This arrow chart illustrates how our selected category is connected with other categories, and gives the number of clusters we find in each

Overview

97% of the 119
Items with connections [View](#) →

100% of the 41,188
Transactions with clusters

Overall Cluster Strength

Your items have **ext strong** clusters.
97% of the 119 items have clusters.

The Clusters page begins to tell us the story of how our data is connected. The blue boxes at the top represent categories, and are there to show you how items in the selected category are connected with items in other categories. In the picture above, we have selected the category “Age” and we can see how many clusters items in that category make with other categories (a cluster is a connection between 2 or more items). Clicking on a ‘Tell Me’ Button will present each row in a more readable format. Let’s scroll down to find out more about this category.

Each blue box represents an item in the data set, and the arrow shows the connection between it and another item or set of items

Column labels tell us more about the connection. Hover over the column name in your browser to learn more

Clicking on a 'tell me' box will give you an easily readable summary of the row

#	Items	Strength	Freq	Cnd Prob	Trend
1	[17.0_to_32.0] age → single marital status	S	5,642	60.5%	N/A
2	[17.0_to_32.0] age → no in default	S	8,330	89.3%	N/A
3	[32.0_to_37.0] age → no in default	S	7,703	85.9%	N/A
4	[17.0_to_32.0] age → cellular contact	S	6,357	68.1%	N/A
5	[17.0_to_32.0] age → no personal loan	S	7,693	82.5%	N/A

Further down the page we see the individual connections for items in this category. Each line represents a connection, where the item on the left suggests the presence of the item on the right. So for the first row in our example, when the item “[17.0-32.0] Age” is present in a row, it suggests that the item “single marital status” will be present in that row as well, with a conditional probability of 100%. After reading a few more lines to get an understanding of this category, let’s scroll back to the top and select the ‘Sale’ category to learn more.

22 Categories By Strength | A - Z

- CCI 100%
- CPI 100%
- **Sale 100%**
- age 100%
- contact 100%
- contacts this campaign 100%
- day_of_week 100%
- days since last contact 100%
- education 100%
- employment variation rate 100%
- euribor3m 100%
- housing loan 100%
- job 100%
- last contact duration 100%

Sale
2 of the 2 items are in clusters across 21 categories.
The number of items with connections make this category's connect

- ← **Sale** → **job** 22,298 Clusters
- **month** 25,842 Clusters
- **nr.employed** 21,482 Clusters
- **employment variation rate** 25,397 Clusters
- **education** 23,292 Clusters
- **day_of_week** 26,309 Clusters
- **age** 21,655 Clusters

Sale is connected with items from Any Category and has **145,964** clusters.

#	Items	Strength	Freq
1	no Sale → [7.0_to_999.0] days since last contact	VS	36,16
2	no ID → nonexistent previous campaign outcome no Sale	VS	32,42
3	Sale Category → no Sale	VS	32,42
4	103 Ranked Cnx 139,776 Clusters → no Sale	VS	32,42

Clicking on 'Sale' will bring up the category detail page

Category Details

After clicking the Sale category, we can see the connections that lead to a 'Yes' or 'No' for the sale. Clicking the arrow inside of a 'no' box will allow us to go to the category detail page to learn more about the different outcomes of 'Sale'.

Category Details Sale

Sale contains:

Download CSV

#	Item	Freq	Int Cx	Rnk Cx	Clstrs	Trend
1	no Sale	36,548	117	103	139,776	N/A
2	yes Sale	4,640	115	109	6,188	N/A

The category detail page shows us the different values for 'Sale' that were found in the data set. To learn more about the patterns associated with a successful sale, click the 'yes' item.

Item Details

Item Detail yes Sale

[Ranked Connections](#) | [Clusters](#) | [Initial Connections](#)

Viewing 2-item clusters after ranking Download Connections CSV

#	Items	Strength	Freq	Cnd Prob	Trend
1	[371.0, to 4918.0] last contact duration	8	2,702	58.2%	N/A
2	[0.634, to 1.313] euribor3m	8	2,561	55.2%	N/A
3	success previous campaign outcome	8	894	19.3%	N/A
4	cellular contact	8	3,653	83.0%	N/A
5	no in default	8	4,197	90.5%	N/A
6	[0.0, to 7.0] days since last contact	8	734	15.8%	N/A
7	no personal loan	8	3,850	83.0%	N/A
8	[7.0, to 999.0] days since last contact	8	3,906	84.2%	N/A
9	[92.201, to 92.963] CPI	8	1,523	32.8%	N/A
10	[-36.1, to -26.9] CCI	8	1,390	30.0%	N/A

Item Overview

yes
ID

Sale
Category [View](#) →

6,188 Clusters [View](#) → **4,640** Transactions

115 Initial Connections **0** Substitutions

Transactions containing this Item

Avg Qty per Trans **1.0**

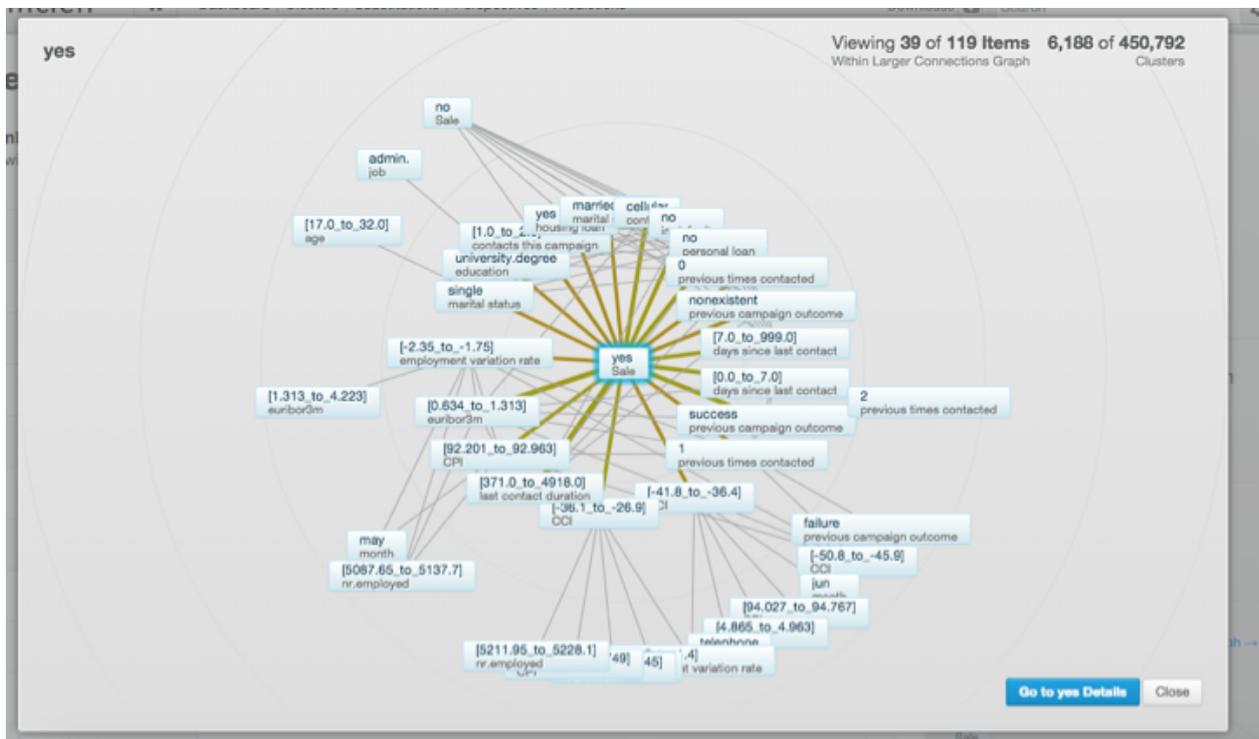
Avg Trans Size **22.0**

This Item has 1,778 Predictors
Containing probabilities as high as **90%**

[View Predictors](#) →

Connections Graph [Explore Graph](#) →

The item detail screen shows us the other items that are most connected to the item we've chosen. What this means is: For the selected item (in this case, 'yes' for sale), which other data values occur in transactions most often with our selected item. By seeing this, we can start to see the relation of our item to other values, and get a sense for what is held in the data. To understand this view a bit more, click on 'Explore Graph', as seen in the bottom right of the picture.



Here is a graphical representation of the previous screen, showing our selected item ('yes' for sale) and all of the items it is most strongly connected with. by highlighting individual

items we see the connection patterns and can begin to understand the data within our data set.

Predictive Rules

Now that we know the patterns and relationships within our sales data, we have finished our first use case. Now, if we'd like to use those connections to find the unique predictors for what causes a successful sale, click on 'View Predictors' on the right hand side of your screen.

Downloading the CSV file will give you a complete list of all predictive rules with their corresponding strengths and information

Predictors for yes Sale

Displaying **1,778** Predictive Clusters

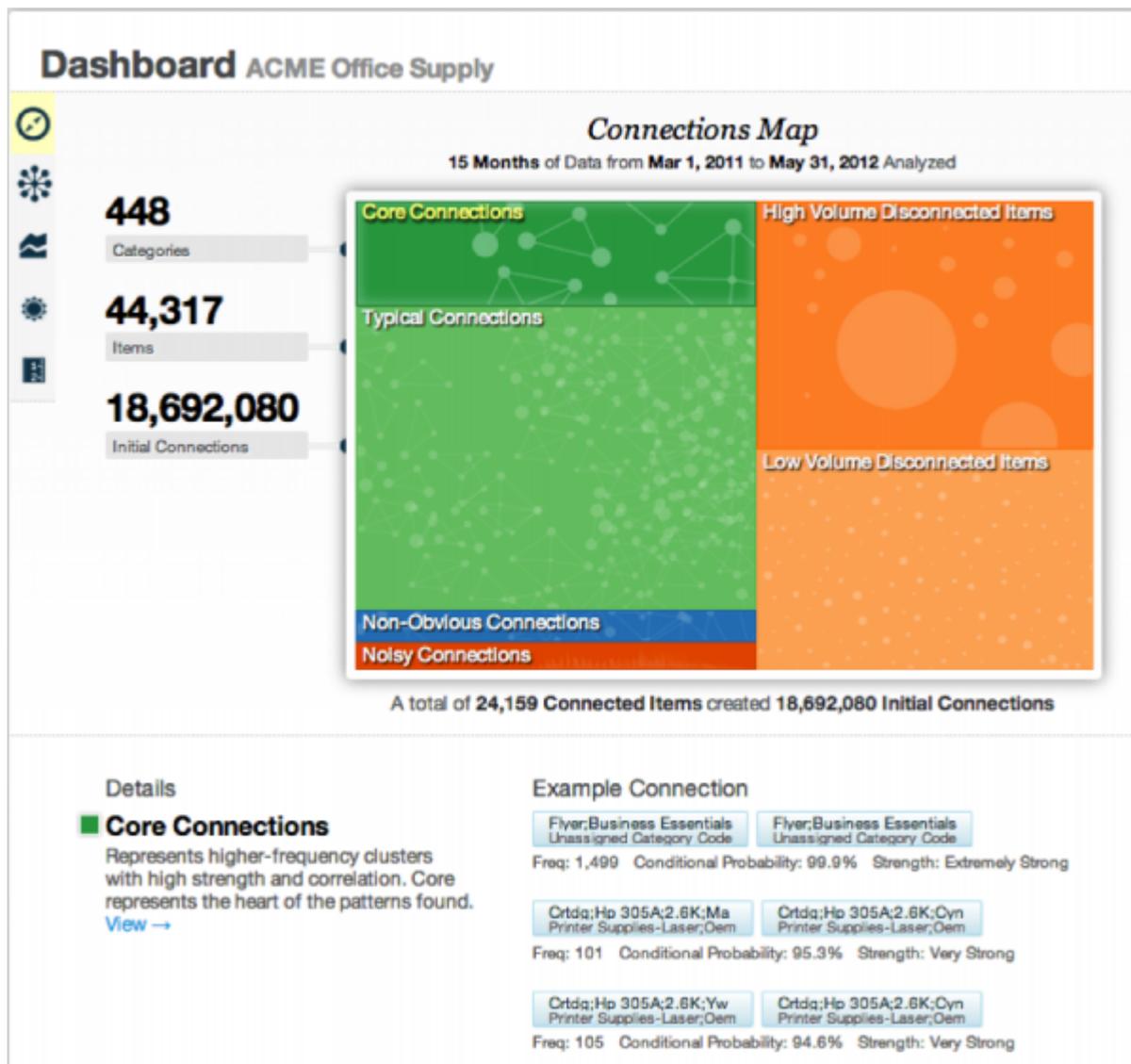
#	Predictor Cluster	Probability ¹	Item	Size ²	Freq
1	success previous campaign outcome + [94.027_to_94.767] CPI + [371.0_to_4918.0] last contact duration	90.1% →	yes Sale	4	127
2	success previous campaign outcome + [-41.8_to_-36.4] CCI + [371.0_to_4918.0] last contact duration	89.4% →	yes Sale	4	126
3	mar month + [0.0_to_7.0] days since last contact	86.0% →	yes Sale	3	49
4	[5000.15_to_5013.1] nr_employed + [93.749_to_93.994] CPI + [371.0_to_4918.0] last contact duration	83.9% →	yes Sale	4	99
5	success previous campaign outcome + [-41.8_to_-36.4] CCI + [224.0_to_371.0] last contact duration	83.9% →	yes Sale	4	151

[Download CSV](#)

The Predictions screen shows us the individual combinations of items that, when found in a transaction together, yield a high probability of our predicted item being present in that transaction. Think of these as rules- when the items on the left are found in a row, there is the displayed probability that the item on the right will be present as well. To download a list of these rules, click the 'Download CSV' button in the top right.

Connections Map

The Connections Map shows a graphical representation of the connections in the processed data. Hovering over a section of the map will update the examples and description in the lower part of the screen. Clicking on that section of the map will take you to the corresponding page in the Report.



The Connections Map can be thought of in two different sections. On the left are the connections that the software has identified and ranked, while the columns on the right represent the individual items that do not have significant connections.

Perspectives

Core Connections

Items in the Core Connections list are both high frequency and very dependent. These items are good candidates for cross selling, bundling together as a special offer, or even stocking closer together. Core Clusters identifies the Clusters that can be bundled, and details the number of Items and Clusters.

Perspectives ACME Office Supply

Core Clusters

Represents higher-frequency Clusters with high strength & correlation. Core represents the heart of the patterns found.



Non-Obvious Clusters

Lower-frequency Clusters that have surprisingly high strength & correlations. These patterns would typically fly "under the radar."



Disconnected Items

Low-frequency Items that do not have a repeated connection with any other Item. These are often opportunities for simplification.



Core Clusters 3,155 Items Created 7,906 Clusters

#	Items	Strength ²	Freq	Size	Cnd Prob	Trend	
1	CARD;OOPS;VILLAGE Unassigned Category Code	CANDY;ALL TYME MIX;ASS... Breakroom-Refreshments	S	335	2	97.4%	
2	TONER;CLP325W;CLX3185... Printer Supplies-Oem;Lsr;Clr	TONER;CLP325W;CLX3185... Printer Supplies-Oem;Lsr;Clr	VS	152	2	92.7%	
3	CRTDG;TONER;MFC9460C... Printer Supplies-Oem;Lsr;Clr	CRTDG;TONER;MFC9460C... Printer Supplies-Oem;Lsr;Clr	VS	261	2	92.6%	
4	CRTDG;INKJET;HP951XL;M... Printer Supplies-Inkjet;Oem	CRTDG;INKJET;HP 951XL;YW Printer Supplies-Inkjet;Oem	VS	474	2	90.1%	
5	CRTDG;INK;MFCJ6510DW;... Printer Supplies-Inkjet;Oem	CRTDG;INK;MFCJ6510DW;... Printer Supplies-Inkjet;Oem	VS	170	2	89.9%	
6	CRTDG;INKJT;PGMNT;MAG... Printer Supplies-Inkjet;Oem	CRTDG;INKJT;PGMNT;CYN;... Printer Supplies-Inkjet;Oem	VS	592	2	89.8%	
7	CRTDG;TONER;MFC9460C... Printer Supplies-Oem;Lsr;Clr	CRTDG;TONER;MFC9460C... Printer Supplies-Oem;Lsr;Clr	VS	254	2	89.8%	
8	CRTDG;HP128A;CM1415;M... Printer Supplies-Oem;Lsr;Clr	CRTDG;HP128A;CM1415;Y... Printer Supplies-Oem;Lsr;Clr	VS	630	2	89.7%	
9	TONER;CLP325W;CLX3185... Printer Supplies-Oem;Lsr;Clr	TONER;CLP325W;CLX3185... Printer Supplies-Oem;Lsr;Clr	VS	147	2	89.6%	
10	CRTDG;INKJET;HP 951;MA... Printer Supplies-Inkjet;Oem	CRTDG;INNKJET;HP 951;YW Printer Supplies-Inkjet;Oem	VS	242	2	89.0%	

If there are multiple pages to view simply click on the Page Number, or "Next" to go to the next page. Clicking "Previous" will take you back a page. Clicking "First" will take you to the beginning of the report and "Last" will take you to the final page.

The Core List is sorted by Strength, but it can also be sorted by Frequency of the connection, dependency, size of the connection, and the Conditional Probability. Each Core Cluster also displays a Trend Graph if time is included in the data.

Non-Obvious Connections

The items in the Non-Obvious list are very dependent but have a low frequency. Removing one item from inventory could impact the sales of any dependent items. The list of Non-Obvious Clusters can also be sorted by Frequency of the connection, dependency, size of the connection, and the Conditional Probability. Each Non-Obvious Cluster can will also display a Trend Graph if time is included in the data.

[Download CSV](#)

Perspectives ACME Office Supply

<p>Core Clusters</p> <p>Represents higher-frequency Clusters with high strength & correlation. Core represents the heart of the patterns found.</p>  <p>8.45% Clusters are Core</p>	<p>Non-Obvious Clusters</p> <p>Lower-frequency Clusters that have surprisingly high strength & correlations. These patterns would typically fly "under the radar."</p>  <p>2.44% Clusters are Non-Obvious</p>	<p>Disconnected Items</p> <p>Low-frequency Items that do not have a repeated connection with any other Item. These are often opportunities for simplification.</p>  <p>28.37% Items are Disconnected</p>
---	---	--

Non-Obvious Clusters 1,622 Items Created 2,286 Clusters

#	Items	Strength	Freq	Size	Cnd Prob	Trend
1	<div style="display: flex; justify-content: space-between;"> <div style="width: 45%;"> <p>INDEX;LTR;1/25;SDTB;301-... Indexes-Avery Legal</p> <p>INDEX;LTR;1/25;SDTB;351-... Indexes-Avery Legal</p> <p>INDEX;LTR;1/25;SDTB;401-... Indexes-Avery Legal</p> <p>INDEX;LTR;1/25;SDTB;451-... Indexes-Avery Legal</p> </div> <div style="width: 45%;"> <p>INDEX;LTR;1/25;SDTB;326-... Indexes-Avery Legal</p> <p>INDEX;LTR;1/25;SDTB;376-... Indexes-Avery Legal</p> <p>INDEX;LTR;1/25;SDTB;426-... Indexes-Avery Legal</p> </div> </div>	ES	3	7	100%	
2	<div style="display: flex; justify-content: space-between;"> <div style="width: 45%;"> <p>INDEX;LTR;1/25;SDTB;301-... Indexes-Avery Legal</p> <p>INDEX;LTR;1/25;SDTB;351-... Indexes-Avery Legal</p> <p>INDEX;LTR;1/25;SDTB;401-... Indexes-Avery Legal</p> </div> <div style="width: 45%;"> <p>INDEX;LTR;1/25;SDTB;326-... Indexes-Avery Legal</p> <p>INDEX;LTR;1/25;SDTB;376-... Indexes-Avery Legal</p> <p>INDEX;LTR;1/25;SDTB;426-... Indexes-Avery Legal</p> </div> </div>	ES	3	6	100%	
3	<div style="display: flex; justify-content: space-between;"> <div style="width: 45%;"> <p>INDEX;LTR;1/25;SDTB;301-... Indexes-Avery Legal</p> <p>INDEX;LTR;1/25;SDTB;351-... Indexes-Avery Legal</p> <p>INDEX;LTR;1/25;SDTB;401-... Indexes-Avery Legal</p> </div> <div style="width: 45%;"> <p>INDEX;LTR;1/25;SDTB;326-... Indexes-Avery Legal</p> <p>INDEX;LTR;1/25;SDTB;376-... Indexes-Avery Legal</p> </div> </div>	ES	3	5	100%	
4	<div style="display: flex; justify-content: space-between;"> <div style="width: 45%;"> <p>CRTDG;XEROX 6360;BK Printer Supplies-Remans;Laser</p> <p>CRTDG;XEROX 6360;MA Printer Supplies-Remans;Laser</p> </div> <div style="width: 45%;"> <p>CRTDG;XEROX 6360;CYN Printer Supplies-Remans;Laser</p> <p>CRTDG;XEROX 6360;YW Printer Supplies-Remans;Laser</p> </div> </div>	ES	3	4	100%	

The Non-Obvious list is sorted by Strength, but it also contains the Frequency of the items' dependency, size of the connection, the Conditional Probability and the Trend graph. In some cases the Conditional Probability may have a 100% value with a lock icon .

When the lock icon is present it means that these items always appear together in a cluster and the order of appearance within the cluster does not matter.

Disconnected Items

These items have a low frequency and do not have any dependencies since they are not found with any other items. Note that this does not display any correlations. These could be considered for inventory reduction since taking them out would not impact other items. The analysis includes a percentage of items that are Disconnected, and on the details it lists the number of items as well.

Perspectives ACME Office Supply

Core Clusters

Represents higher-frequency Clusters with high strength & correlation. Core represents the heart of the patterns found.



Non-Obvious Clusters

Lower-frequency Clusters that have surprisingly high strength & correlations. These patterns would typically fly "under the radar."



Disconnected Items

Low-frequency Items that do not have a repeated connection with any other Item. These are often opportunities for simplification.



Disconnected Items 9,935 Items

#	Item	Category	Freq ¹	Int Cx	Rnk Cx	Clstrs	Trend
1	CATALOG;QUALITY OFFI... Unassigned Category Code	Unassigne...	1	5	0	0	
2	INDEX;RNGBK;LTR;1-31 Index Dividers-Tabbing	Index Divi...	1	12	0	0	
3	FLDR CLASS 4FAST 2" L... Filing;Top Tab;Tt-C/Fldr	Filing;Top ...	1	14	0	0	
4	BAG;DEPOSIT;NIGHT;NY... Bank Supplies	Bank Sup...	1	2	0	0	
5	PCKT;ET;EXP;1-3/4;LTR;... Filing-End Tab;Et-Pkt	Filing-End ...	1	2	0	0	
6	CLIPS;COATED;VINYL;M... Clips;Clamps & Fasteners	Clips;Cla...	1	0	0	0	
7	BOARD;PRCLN;36X38;W... Organizers-Wall Planners	Organizers...	1	1	0	0	
8	PEN; PAPERMATE GEL 2... Pens-Gel Ink	Pens-Gel Ink	1	0	0	0	
9	FLDR;RECY;1/3 CUT;LTR;... Filing;Top Tab;Tt-Fldr	Filing;Top ...	1	9	0	0	
10	BOARD;STANDING LETT... Board-Message & Display	Board-Me...	1	1	0	0	

The Disconnected List contains the Item and it's Category, the Frequency, number of Initial Connections, number of Ranked Connections, number of Clusters, and the Trend graph.

Substitutions

Substitutions show the common connections that items share. They both have the same “friends” but are not found together in a transaction. Items that are substitutes could possibly replace each other in the same basket of transactions.

The upper left hand side of the screen contains the Categories. These are ordered by default by Strength. It can also be alphabetized by Category name. Simply click on the A-Z and the display will update. To change it back click on “By Strength”.

Emcien’s Strength ranking is a proprietary measure the Pattern Detection Engine assigns to connections within an analyzed data set. Think of Strength as a metric that combines Cluster Frequency, the number of times that a specific group of items has occurred in a transaction together, and Correlation, the rate of co-occurrence shared by a selected item and the cluster being displayed.

When a category is selected it will update the upper right hand side to show the items in that category along with the items that have substitutions.

Substitutions ACME Office Supply

Download CSV

112 Categories By Percent | A - Z

- **Breakroom-Refreshments** 48%
- Motivational Marterials 44%
- Filing-End Tab;Et-Access 44%
- Forms-Payroll & Tax 43%
- Paper-Colored 41%
- Art Board & Paper 39%
- Breakroom-Coffee & Tea 39%
- Calendars-Desk Pads 36%
- Labelmaker-Electrnc Lblr Tape 29%
- Labelmakers-Electronic 28%
- Printer Supply-Oem;l;Lg Frmt 27%
- Calendars-Wall 26%
- Labelmakers-Label Printer Lbl 26%

Breakroom-Refreshments

151 items in this category are in 1,004 substitutions. [View All →](#)

- RITZ BITS CHEESE BIG BAG Breakroom-Refreshments
 - SNACK;TRADL;CHEX MIX;3.75OZ Breakroom-Refreshments
 - MIX;SWEET 'N SPICY;KAR'S Breakroom-Refreshments
 - POPTART;STRAWBERRY Breakroom-Refreshments
- PEANUTS;SALTED;KAR'S;2OZ Breakroom-Refreshments
 - SNACK;BAR;STRWBRY YOGRT;NV Breakroom-Refreshments
 - BAR; GRANOLA; PEANUT BUTTER Breakroom-Refreshments
- CANDY;RIESEN;30OZ BAG Breakroom-Refreshments
 - CANDY;TOP OF THE LINE POPS Breakroom-Refreshments
- CEREAL;CUP;CHEERIOS;ORIG Breakroom-Refreshments
 - CEREAL;FRSTD FLAKES;CUPS Breakroom-Refreshments

Substitute and friends

This Substitution is **S**



Showing All 3 Friends

← **Friends**
 The gray items are "Friends". Each substitute item is found in transactions with these friends.

 There are no connections, after ranking, between these substitute items.

The key of Substitutes is that the 2 Light Blue items are not found together in a transaction, and the "friends" that they share are their only "friends". This means that one item could be substituted for the other.

If one of the categories on the right is selected it will become highlighted and the lower detail screen will be updated. The outside Light Blue icons are the substitutes. The gray icons in the middle are "Friend" items found with either of the Light Blue items.

The detail also shows the total number of transactions for each item and the number of friends being displayed. Also being shown in the detail is the Strength of the substitution as mathematically calculated by Patterns.

Clusters

The Clusters page allows you to see how items and categories are connected.

The upper left hand side contains the Categories. These are ordered by default by Strength. It can also be alphabetized by Category Name. Simply click on A-Z and the display will update. To change it back, click on By Strength.

Emcien's Strength ranking is a proprietary measure the Pattern Detection Engine assigns to connections within an analyzed data set. Think of Strength as a metric that combines Cluster Frequency, the number of times that a specific group of items has occurred in a transaction together; Conditional Probability, how likely the connected Items occur in a transaction together when the Anchor Item is present; and mutual information from the graph.

Clusters ACME Office Supply

400 Categories By Strength | A - Z

- **Audio Visual-Tripods** 100%
- Computer-Input;Mice;Lsr;Corde 100%
- Copiers-Analog 100%
- Fax Machines-Thermal Transfer 100%
- Labelmakers-Manual 100%
- Printers-Dot Matrix 100%
- Video-Cassette;8Mm 100%
- Filing;Hanging;Hf-Interior 93%
- Filing-End Tab;Et-Pkt 89%
- Printer Supplies-Ij;Elite Ima 88%
- Printer Supplies-Inkjet;Oem 88%
- Filing-End Tab;Et-Access 87%
- Paper-Colored 87%
- Filing;Top Tab;Tt-Pkt 85%

Audio Visual-Tripods

1 of the 1 Items are in Clusters across 2 categories.
The number of Items with Connections make this Category's connections **Extremely Strong**.



Audio Visual-Tripods is connected with Items from and has 3 clusters.

#	Items	Strength	Freq	Cnd Prob	Trend
1	Camcorder;Hard Drive;120Gb Audio Visual-Video Cameras;Dq	M	5	35.7%	
2	Camcorder;16Gb Cameras-Digital	M	4	36.4%	
3	Camcorder;16Gb Intrnl Me... Audio Visual-Video Cameras;Dq	SW	4	22.2%	

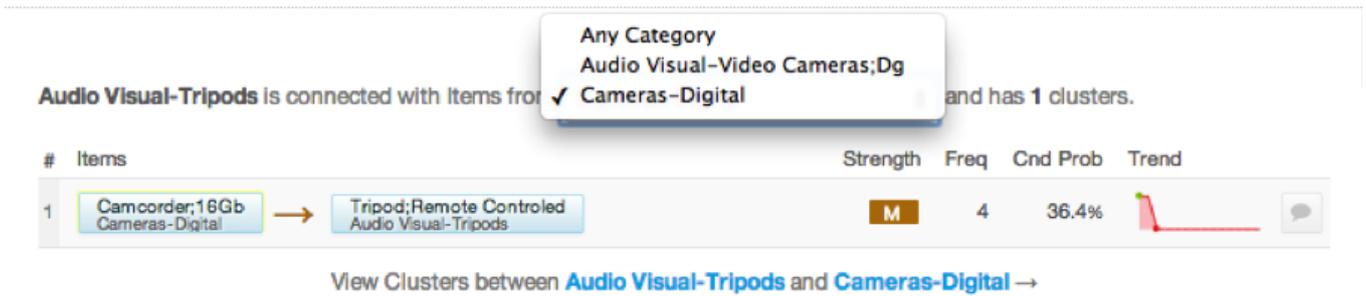
[View Clusters in Audio Visual-Tripods](#) →

When a category is selected it will update the upper right hand side to show the selected category along with the related categories that have dependencies.

Categories are displayed with a Dark Blue icon.

Items are displayed throughout Patterns with a Light Blue icon.

If one of the categories on the right is selected it will become highlighted and the lower detail screen will be updated. Or you can select one of the Categories from the list in the pull down menu:



Each line in the list contains a Cluster. For each Cluster we display the Cluster Items, Strength, Frequency, Conditional Probability, and Trend.

The Item on the left is the Anchor Item. The arrow indicates the dependency relationship between the anchor Item and the remaining Items in the Cluster.

Emcien’s Strength ranking is a proprietary measure the Pattern Detection Engine assigns to connections within an analyzed data set.

Frequency is the number of transactions that contain the Cluster.

Conditional Probability is the likelihood that the connected Items will be present when the anchor Item is present.

Trend is a spark line graph that shows the periodic trends of the data. A green line means a positive or upward trend while a red means a negative or downward trend.

An example of a dependency relationship would be a printer and a USB cable. If you purchase a printer you are likely to purchase a USB cable with it. If you are purchasing a USB cable it is unlikely that you are going to add a printer to your purchase. For this 2-Item cluster the printer is the anchor Item. Purchase of the USB cable was driven by the printer purchase.

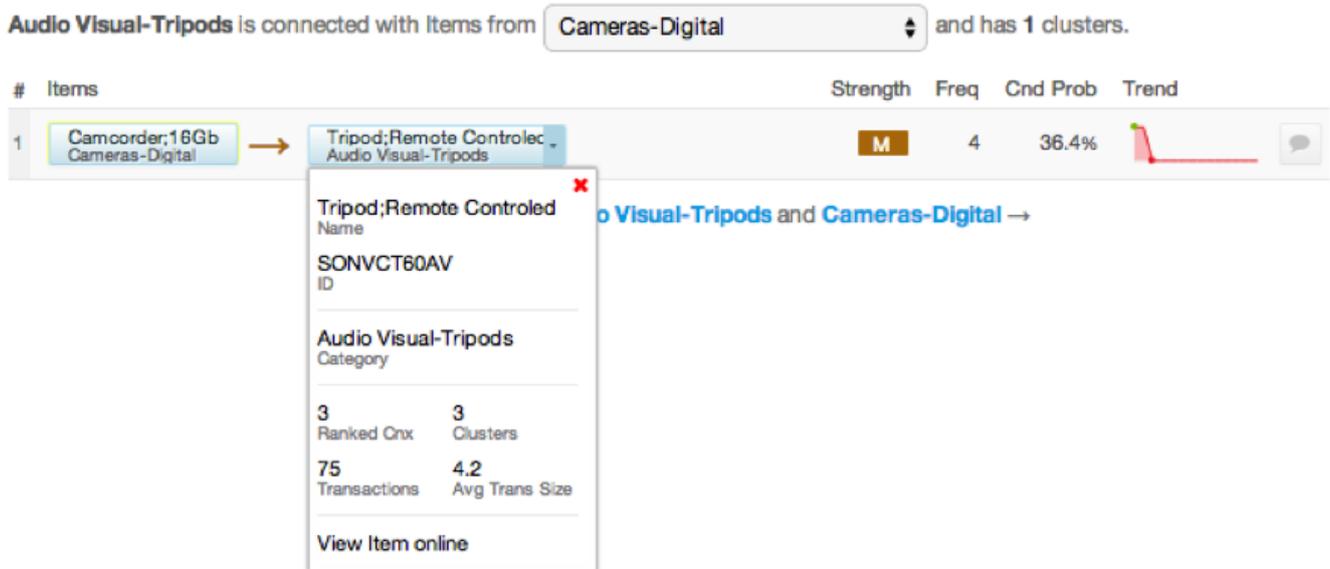
At the end of each line there is a “Tell Me” icon  that will bring up more

detailed information about the connection from the perspective of the anchor Item.

Example of “Tell Me” pop up details:



Information on each item can be viewed by clicking on the right hand side of the item name where a pull down arrow will appear. Selecting it will display the item name, Item ID, Category, Ranked Connections, Clusters, Transactions, and Average Transaction Size. Clicking on the red X in the menu will close it.



If a Company URL has been provided for on-line Item look-up, it will be displayed at the bottom of the Item information.

To view the complete list of Clusters for one of the menu pull down selections you would click on

View Clusters between **Audio Visual-Tripods** and **Cameras-Digital** →

when it is set to “Any Category” or click on

View Clusters between Audio Visual-Tripods and Cameras-Digital →

when you have selected a specific category from the list.

Creating Prediction Rules

EmcienPatterns can generate Prediction Rules based on a specific outcome category in the data. In this section of Predictive Rules and Analysis we will be using the Network Attack sample data. To download the files simply click [Network_Attack.zip](#).

To process an analysis report to generate Prediction Rules simply follow the steps in this example:

1. From the Home screen click on the Analyze Data button:



2. Select the data file from the list

New Analysis Demo Admin

Source Data

Filter by Name

Name	File Size	Modified
banded_network_attack_test.wide.csv	750 KB	May 27 4:27PM
✓ banded_NetworkAttack.wide.csv	2.8 MB	May 27 4:27PM

[Upload New Source Data](#)

3. Click on the plus sign in front of Predictions and type in the Category name "Classification". This is the category that will be used to generate the Predictive Rules for this data.

 Predictions

Outcome Category Name

CLASSIFICATION|

Must have 20 items or less

4. Update the Report Name and click on Start Analysis.

Report Name

banded NetworkAttack

Start Analysis

5. Once the Analysis is complete click on View Analysis.

banded NetworkAttack Demo Admin



Log: **Summary** | Expanded

Run Time: 1m 8s

#	Message	Time
109	Queueing Backup Job	1m 8s
108	-----> Import successful for report 65798386	1m 8s
107	Report State: Report-ready	1m 7s
106	Report State: report-verify	1m 4s
105	Report State: downloads-built	1m 3s

6. At the top of the Analysis screen click on Rules.

Click here to go to the next section in the knowledgebase "[Prediction Rules](#)"

Prediction Rules

Prediction Rules

So now that we have a Prediction Analysis processed, what do the Prediction Rules tell us and how do we use them?

The Prediction Rules screen allows a user to view the rules in the User Interface to verify to what level they are contributing and the coverage of the data. These rules could then be exported to an external system for real-time dashboard reporting or EmcienPatterns' real-time Prediction Engine.

The Rules screen displays all of the predictive rules for the outcomes in the category that we want to predict. To easily understand the rules, the engine divides them into three different sections based on how often the rules occur within the original data.

Let's start with the area on the right side of the screen that shows the Targeted Rules.

Prediction Rules banded NetworkAttack Download CSV

139 Obvious Rules
Rules that are obvious because they exist at high frequencies.

61 Common Rules
Rules which are less obvious because they are decreasing in frequency.

108 Targeted Rules
Rules which contribute highly to prediction results yet are challenging to detect because of their low frequency.

Rules for: **CLASSIFICATION**

14 Outcome Items
Isolate one or more items below. Select **All** or **None**

Displaying **108** of **308** total prediction rules.

#	Freq	Prediction Rule	Probability	Outcome Item	Size	Cover	Contrib	Trend
1	17	ICMP Echo Reply SIGNATURE_NAME	100% →	Misc activity CLASSIFICATION	1	100%	0.71	
2	17	408 SIGNATURE_ID	100% →	Misc activity CLASSIFICATION	1	100%	0.71	
3	4	2001669 SIGNATURE_ID	100% →	Potentially Ba... CLASSIFICATION	1	50.0%	0.50	
4	3	3815 SIGNATURE_ID	100% →	Misc Attack CLASSIFICATION	1	50.0%	0.50	
5	12	222.100.5.233 DST_IP	92.3% →	Misc activity CLASSIFICATION	1	70.6%	0.46	
6	27	1112 SIGNATURE_ID	100% →	Attempted Inf... CLASSIFICATION	1	1.34%	0.45	
7	24	1142 SIGNATURE_ID	100% →	Attempted Inf... CLASSIFICATION	1	1.20%	0.45	
8	25	10.2.23.141 SRC_IP	100% →	Attempted Ad... CLASSIFICATION	1	4.26%	0.44	

- A Network Trojan was detected CLASSIFICATION
- access to a potentially vulnera... CLASSIFICATION
- Attempt to login by a default u... CLASSIFICATION
- Attempted Administrator Privil... CLASSIFICATION
- Attempted Denial of Service CLASSIFICATION
- Attempted Information Leak CLASSIFICATION
- Attempted User Privilege Gain CLASSIFICATION
- Detection of a non-standard p... CLASSIFICATION
- Executable code was detected CLASSIFICATION
- Misc activity CLASSIFICATION
- Misc Attack CLASSIFICATION
- Potentially Bad Traffic CLASSIFICATION
- UNKNOWN CLASSIFICATION
- Web Application Attack CLASSIFICATION

The Targeted Rules are our rules that have very high probabilities, but have very low frequencies. This means that these items don't come together very often- but when they do, they are highly predictive.

Clicking in the middle of the screen will bring up a list of Common Rules.

Prediction Rules banded NetworkAttack Download CSV

139 Obvious Rules
Rules that are obvious because they exist at high frequencies.

61 Common Rules
Rules which are less obvious because they are decreasing in frequency.

108 Targeted Rules
Rules which contribute highly to prediction results yet are challenging to detect because of their low frequency.

Rules for: **CLASSIFICATION**

14 Outcome Items
Isolate one or more items below. Select **All** or **None**

Displaying **61** of **308** total prediction rules.

#	Freq	Prediction Rule	Probability	Outcome Item	Size	Cover	Contrib	Trend
1	107	1156 SIGNATURE_ID	100% →	Attempted De... CLASSIFICATION	1	93.9%	1.93	
2	64	1147 SIGNATURE_ID	100% →	Attempted Inf... CLASSIFICATION	1	3.19%	0.46	
3	47	10.2.23.122 SRC_IP	95.9% →	Attempted Ad... CLASSIFICATION	1	8.01%	0.44	
4	127	2152 SIGNATURE_ID	100% →	access to a p... CLASSIFICATION	1	5.07%	0.37	
5	158	1188 SRC_PORT	98.8% →	access to a p... CLASSIFICATION	1	6.31%	0.37	
6	51	2565 SIGNATURE_ID	100% →	access to a p... CLASSIFICATION	1	2.04%	0.36	
7	116	10.2.199.236 SRC_IP [13.0_to_14.0] Hour 2 PRIORITY	97.5% →	access to a p... CLASSIFICATION	3	4.63%	0.36	
8	148	1250 SIGNATURE_ID	100% →	Web Applicati... CLASSIFICATION	1	2.18%	0.35	

- A Network Trojan was detected CLASSIFICATION
- access to a potentially vulnera... CLASSIFICATION
- Attempt to login by a default u... CLASSIFICATION
- Attempted Administrator Privil... CLASSIFICATION
- Attempted Denial of Service CLASSIFICATION
- Attempted Information Leak CLASSIFICATION
- Attempted User Privilege Gain CLASSIFICATION
- Detection of a non-standard p... CLASSIFICATION
- Executable code was detected CLASSIFICATION
- Misc activity CLASSIFICATION
- Misc Attack CLASSIFICATION
- Potentially Bad Traffic CLASSIFICATION
- UNKNOWN CLASSIFICATION
- Web Application Attack CLASSIFICATION

The Common Rules occur somewhat frequently, and when they occur they are fairly predictive. These items come together more often than the Targeted Rules, but when they do they may not lead to a targeted outcome as often.

Clicking on the left side of the screen will display the Obvious Rules.

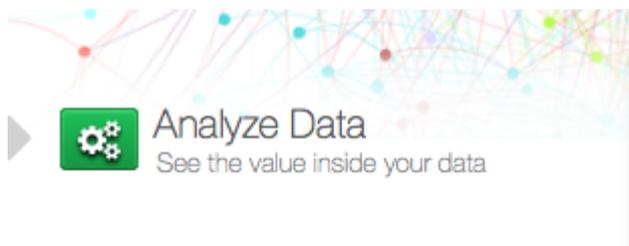


Predictive Analysis

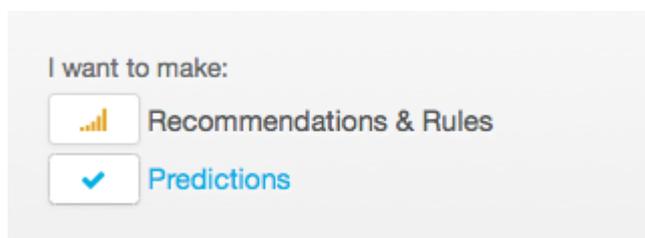
Predictive Analysis

Now that we have created a Predictive Analysis and Rules that correspond to it we want to run a separate data set against these rules to see the outcome.

To run a Prediction, click on the Analyze Data button on the home screen.



On the right side of the screen you will see two options: "Recommendations & Rules" and "Predictions"



Click on "Predictions"

Next select the Training Rules from the pull down menu:

New Prediction Demo

Select Training Rules

Select Test Data

Filter by Name

Name	File Size	Modified
banded_network_attack_test.wide.csv	750 KB	May 27 4:27PM
banded_NetworkAttack.wide.csv	2.8 MB	May 27 4:27PM

[Upload New Source Data](#)

When you select the Rule set from the list, the ID number will match the Predictive Analysis that created the rule set.

New Prediction Demo

Select Training Rules

Select Test Data

Filter by Name

Name	File Size	Modified
✓ banded_network_attack_test.wide.csv	750 KB	May 27 4:27PM
banded_NetworkAttack.wide.csv	2.8 MB	May 27 4:27PM

[Upload New Source Data](#)

Select the file that you want to Predict from the list of available files.

Click on "Make Prediction" after making any changes to the name of the analysis.

Name

Once the Prediction has completed processing, simply click on "View Predictions"

Prediction Log banded network attack test

Successfully Predicted

[View Predictions](#)

#	Log Message	Time
96	▶ Complete	22s
95	Building characterization of the extraction	22s

Clicking on View Predictions will take you to the Predictions Dashboard. For the detailed knowledgebase article on the Predictions Dashboard click [here](#).

Prediction Dashboard

Prediction Dashboard

The predictions page is helpful to see how accurate the engine is at making predictions, and displays statistics on the engine's performance. The columns are defined below:

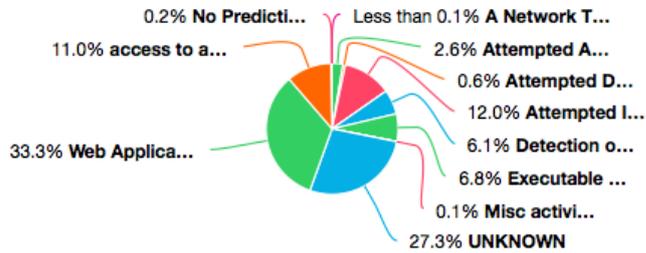
Prediction Dashboard banded network attack test

Starting with your training data... your test data predicted the following:

banded NetworkAttack
Training Report Name

Nov 6 2:29PM
Training File Created

308
Rules from Training File



Explore the **Outcome(s)** you most care about:

#	Predicted	Outcome	Captured	TP	FP	P%	TN	FN	N%	Actual	Skipped
1	0 Predictions	A Network Trojan was detected	--	0	0	--	5,268	0	100.0%	0	0
2	0 Predictions	Attempt to login by a default username and password	--	0	0	--	5,268	0	100.0%	0	0
3	139 Predictions	Attempted Administrator Privilege Gain	96.45%	136	3	97.8%	5,124	5	99.9%	141	0
4	33 Predictions	Attempted Denial of Service	94.29%	33	0	100.0%	5,235	0	100.0%	35	2
5	633 Predictions	Attempted Information Leak	90.0%	522	111	82.5%	4,580	55	98.7%	580	3
6	0 Predictions	Attempted User Privilege Gain	0.0%	0	0	--	5,262	6	99.9%	6	0
7	320 Predictions	Detection of a non-standard protocol or event	100.0%	320	0	100.0%	4,948	0	100.0%	320	0
8	360 Predictions	Executable code was detected	99.71%	348	12	96.7%	4,907	1	100.0%	349	0
9	0 Predictions	Misc Attack	--	0	0	--	5,268	0	100.0%	0	0

Captured- This is the percentage of predictions that the engine made for the item that the engine was able to correctly predict.

TP- True Positives, this is the number of times the engine predicted an item would occur, and it did occur (e.g. engine predicted "sick" and the patient was sick).

FP- False Positives, this is the number of times the engine predicted an item would

occur and the item did not occur (e.g. engine predicted "sick" and the patient was not sick).

P%- The number of True Positives divided by the total number of time the engine predicted a positive outcome, displayed as a percentage.

TN- True Negatives, the number of times the engine confidently predicted the item would not occur in a transaction and was right (e.g. engine predicted healthy people were not sick).

FN- False Negatives, the number of times the engine confidently predicted the item would not occur in a transaction and the item actually was in the transaction. (e.g. engine predicted healthy people were sick).

N%- similar to P%, this is the number of True Negatives divided by Total Negatives.

Actual- The actual number of times the item occurred in the data set.

Skip- The number of times the engine did not have enough information to confidently list an outcome.

Search

Overview

Searches can be done in a number of different ways. Search options include: Items Search, Categories Search, Substitutions Search and Prediction Rule Search. Each search will have different filtering criteria.

To select the search that you wish to use click on the gear icon and select it from the list:

- Searches
- Item Search
- Category Search
- Cluster Search
- Substitution Search
- Prediction Rule Search

Item Searches

This search will display the item along with the category. It will also have the Frequency, Initial Connections, Ranked Connections, Clusters, and the Spark Trend line.

Item Search ACME Office Supply

Your search matched **35,020** of the **35,020** total Items.

[Download CSV](#)

#	Item	Category	Freq ↓	Int Cx	Rnk Cx	Clstrs	Trend
1	FLDR;FILE;LTR;1/3;MLA;1... <small>Filing;Top Tab;T1-Fldr</small>	Filing;Top ...	21,984	12,703	252	282	
2	MARKER;PERM;SHARPI... <small>Pens & Markers</small>	Pens & M...	15,280	12,077	273	295	
3	PAPER;8.5X11;20#;WE <small>Paper-Multipurpose</small>	Paper-Mul...	13,375	8,117	36	36	
4	STAPLES;CHISEL PT;STD... <small>Staples</small>	Staples	12,826	10,863	191	326	
5	CLIP;BINDER;SM;12EA <small>Clips;Clamps & Fasteners</small>	Clips;Cla...	12,340	11,368	133	143	
6	CLIP;BINDER;MED;12EA <small>Clips;Clamps & Fasteners</small>	Clips;Cla...	11,731	11,235	97	108	
7	PAPER;COPY;8.5X11;WE... <small>Paper-Copy</small>	Paper-Copy	11,174	8,845	95	110	

Advanced Search

Add search criteria for more specific searches

Name or ID

Contains

[+ Add Criteria](#) Search

Other Searches

[Category Search](#)
Perform complex queries for Category

[Cluster Search](#)
Perform complex queries for Cluster

[Substitution Search](#)
Perform complex queries for Substitution

The frequency is the primary sort in the report and clicking on that column heading will reverse the sort order. The Item, Frequency, Initial Connections, Ranked Connections, Clusters, and Trend headings can also be sorted in the report. Simply click on any of those headings to change the order of the sort.

Category Searches

This will display the Category name along with the number of items in the category, the number of Clusters, the number of items that have Substitutes, the number of transactions, the average number of transactions, and the Spark Trend line.

Category Search ACME Office Supply

Download CSV

Your search matched 506 of the 506 total Categories.

#	Category	Items	Clrts	Sub	Trans ¹	Trns #	Trend
1	Printer Supplies-Inkjet;Oem	560	4,160	79	80,903	4.7	
2	Printer Supplies-Oem;Lsr;Mono	578	955	13	80,308	4.0	
3	Pads-Adhesive	365	1,584	30	79,265	7.1	
4	Clips;Clamps & Fasteners	249	1,667	4	78,617	7.0	
5	Pens-Retractable & Ballpoint	544	2,849	42	76,639	6.9	
6	Paper-Copy	54	751	0	75,014	4.6	
7	Tape & Dispensers	331	1,755	6	70,287	6.5	
8	Filing;Top Tab;T1-Fldr	363	1,624	45	67,190	6.1	

Advanced Search

Add search criteria for more specific searches

Name
Contains

+ Add Criteria

Search

Other Searches

Item Search

Perform complex queries for Item

Cluster Search

Perform complex queries for Cluster

Substitution Search

Perform complex queries for Substitution

The number of Transactions is the primary sort on this report and clicking on that column heading will reverse the sort order. The Category, Items, dependencies, Substitutions, and Average Transactions can also be sorted in the report. Simply click on any of those headings to change the order of the sort.

Cluster Searches

This will show the item along with it's dependent items, the mathematically calculated "Strength" of the connection, the frequency, the number of items that are connected, the Conditional Probability, and the Spark Trend line.

Cluster Search ACME Office Supply

Download CSV

Your search matched **93,588** of the **93,588** total Clusters. Search criteria: Size at least "1".

#	Items	Strength	Freq	Size	Cnd Prob	Trend
1	INDEX:LTR;1/25;SDTB;301-... Indexes-Avery Legal INDEX:LTR;1/25;SDTB;326-... Indexes-Avery Legal INDEX:LTR;1/25;SDTB;351-... Indexes-Avery Legal INDEX:LTR;1/25;SDTB;376-... Indexes-Avery Legal INDEX:LTR;1/25;SDTB;401-... Indexes-Avery Legal INDEX:LTR;1/25;SDTB;426-... Indexes-Avery Legal INDEX:LTR;1/25;SDTB;451-... Indexes-Avery Legal	ES	3	7	100%	
2	INDEX:LTR;1/25;SDTB;301-... Indexes-Avery Legal INDEX:LTR;1/25;SDTB;326-... Indexes-Avery Legal INDEX:LTR;1/25;SDTB;351-... Indexes-Avery Legal INDEX:LTR;1/25;SDTB;376-... Indexes-Avery Legal INDEX:LTR;1/25;SDTB;401-... Indexes-Avery Legal INDEX:LTR;1/25;SDTB;426-... Indexes-Avery Legal	ES	3	6	100%	
3	INDEX:LTR;1/25;SDTB;301-... Indexes-Avery Legal INDEX:LTR;1/25;SDTB;326-... Indexes-Avery Legal INDEX:LTR;1/25;SDTB;351-... Indexes-Avery Legal INDEX:LTR;1/25;SDTB;376-... Indexes-Avery Legal INDEX:LTR;1/25;SDTB;401-... Indexes-Avery Legal	ES	3	5	100%	
4	CRTDG:XEROX 6360;BK Printer Supplies-Remans;Laser CRTDG:XEROX 6360;CYN Printer Supplies-Remans;Laser CRTDG:XEROX 6360;MA Printer Supplies-Remans;Laser CRTDG:XEROX 6360;YW Printer Supplies-Remans;Laser	ES	3	4	100%	
5	INDEX:LTR;1/25;SDTB;301-... Indexes-Avery Legal INDEX:LTR;1/25;SDTB;326-... Indexes-Avery Legal INDEX:LTR;1/25;SDTB;351-... Indexes-Avery Legal INDEX:LTR;1/25;SDTB;376-... Indexes-Avery Legal	ES	3	4	100%	

Advanced Search

Add search criteria for more specific searches

Size
At least

+ Add Criteria

Search

Display Options

Display or Hide Child Clusters

A B C Parent Clusters
 B Child Clusters

Display Lockstep Items Only

Locksteps

Other Searches

[Item Search](#)
Perform complex queries for Item

[Category Search](#)
Perform complex queries for Category

[Substitution Search](#)
Perform complex queries for Substitution

Strength is the primary sort on the report and clicking on that column heading will reverse the sort order.

The frequency, and size headings can also be sorted in the report. Simply click on any of those headings to change the order of the sort.

The Cluster search has an option that allows the user to display the "Child Clusters". This is enabled by default so that the user can view all of the clusters.



Simply click on the check box next to "Child Clusters" to remove this option. You

will notice that the number of “Matched” versus “Total Clusters” changes.

Cluster Search ACME Office Supply
Your search matched **66,559** of the **93,588** total Clusters. Search criteria: Size at least "1".

#	Items	Strength	Freq	Size	Cnd Prob	Trend
1	INDEX;LTR;1/25;SDB;301-... Indexes-Avery Legal INDEX;LTR;1/25;SDB;351-... Indexes-Avery Legal INDEX;LTR;1/25;SDB;401-... Indexes-Avery Legal INDEX;LTR;1/25;SDB;451-... Indexes-Avery Legal	ES	3	7	100%	
2	CRTDG;XEROX 6360;BK Printer Supplies-Remans;Laser CRTDG;XEROX 6360;MA Printer Supplies-Remans;Laser	ES	3	4	100%	
3	TONER;CLP365W;CLX3305... Printer Supplies-Oem;Lsr;Clr TONER;CLP-365W;CLX330... Printer Supplies-Oem;Lsr;Clr	ES	4	3	100%	
4	TONER;X792;RTN P;CYN;20K Printer Supplies-Oem;Lsr;Clr TONER;X792;RTN P;MA;20K Printer Supplies-Oem;Lsr;Clr	ES	4	3	100%	
5	CRTDG;TONER;HC;C530;YW Printer Supplies-Oem;Lsr;Clr CRTDG;TONER;HC;C530;C... Printer Supplies-Oem;Lsr;Clr	ES	4	3	100%	
6	INKS;SOLID;CQ8700;4.2K;... Printer Supply-Oem;Inksticka INKS;SOLID;CQ8700;4.2K;... Printer Supply-Oem;Inksticka	ES	3	3	100%	

Advanced Search
Add search criteria for more specific searches
Size: [dropdown] 1
At least: [dropdown]
+ Add Criteria Search

Display Options
Display or Hide Child Clusters
A B C Parent Clusters
B C Child Clusters
Display Lockstep Items Only
 Locksteps

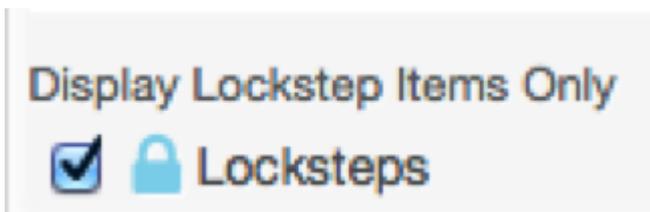
Other Searches
Item Search
Perform complex queries for Item
Category Search
Perform complex queries for Category
Substitution Search
Perform complex queries for Substitution

In some cases the Conditional Probability may have a 100% value with a lock icon



When the lock icon is present it means that these items always appear together in a cluster and the order of appearance within the cluster does not matter.

There is an option to “Display Lockstep Items Only” which when enabled will change the display to only show those Clusters that have lockstep items. Click on the checkbox to enable or disable it.



The count at the top of the screen will then be updated:

Cluster Search ACME Office Supply

Download CSV

Your search matched 21 of the 319,459 total Clusters. Search criteria: Size at least "1".

#	Items	Strength	Freq	Size	Cnd Prob	Trend
1	Index;Ltr;1/26;Sdtab;Q Indexes-Avery Legal Index;Ltr;1/26;Sdtab;R Indexes-Avery Legal Index;Ltr;1/26;Sdtab;T Indexes-Avery Legal	ES	4	3	100%	
2	Desk;L 72";10f2;Mo Office Furnishings-Contemprar Desk;L 72";20f2;Mo Office Furnishings-Contemprar	ES	12	2	100%	
3	Desk;L 60";10f2;Mo Office Furnishings-Contemprar Desk;L 60";20f2;Mo Office Furnishings-Contemprar	ES	8	2	100%	
4	Crtgd;Toner;7000;Yw Ribbons-Cash Register&Finance Ink;Cart;Ltrmg;Stylpro 7500 Printer Supplies-Inkjet;Oem	ES	8	2	100%	
5	Toner;Konica2400;2500;Cyan Printer Supplies-Laser;Oem Toner;Konica2400;2500;Yel Printer Supplies-Laser;Oem	ES	7	2	100%	
6	Tower;Rght;Modchy;Bx1/2 Office Furnishings-Contemprar Tower;Rgt;Chy;Bx2/2 Office Furnishings-Contemprar	ES	7	2	100%	
7	Top;Table;Conf;Round;Bbc Office Furnishings-Conferene Base;Table;Conf;X;42";Cy Office Furnishings-Conferene	ES	7	2	100%	
8	Index;Legal;851-875 Indexes-Avery Legal Index;Legal;876-900 Indexes-Avery Legal	ES	6	2	100%	
9	Hutch;Stackon;72";W;Door;Cy Office Furnishings-Contemprar Hutch;Stack-On;72";Cy;Bx2/2 Office Furnishings-Contemprar	ES	5	2	100%	
10	Index;Legal;Sdtab;V Indexes-Avery Legal Index;Legal;Sdtab;X Indexes-Avery Legal	ES	4	2	100%	
11	Index;Ltr;1/25;Sdtab;58 Indexes-Avery Legal Index;Ltr;1/25;Sdtab;59 Indexes-Avery Legal	ES	4	2	100%	

Advanced Search

Add search criteria for more specific searches

Size
At least

+ Add Criteria

Search

Display Options

Display or Hide Child Clusters

A B C Parent Clusters
 B C Child Clusters

Display Lockstep Items Only

Locksteps

Other Searches

Item Search

Perform complex queries for Item

Category Search

Perform complex queries for Category

Substitution Search

Perform complex queries for Substitution

Substitution Searches

This displays the 2 items (A and B) that are connected by mutual items but are not found together in a transaction. It lists the number of transactions for A and B, and also the mathematically calculated "Strength" of the connection.

Strength is the primary sort on the report and if you click on it, the sort order will be reversed.

Substitution Search ACME Office Supply

Download CSV

Your search matched 5,360 of the 5,360 total substitutions .

#	Item A	Item B	Trans A	Trans B	Strength
1	INDEX;LEGAL;SDTAB;8 Indexes-Avery Legal	INDEX;LEGAL;SDTAB;2 Indexes-Avery Legal	4	9	Extremely Strong
2	CRTDG;MAINT;DSNJT;HP... Printer Supply-Oem;Lg Frmt	CRTDG;HP 771;775ML;M... Printer Supply-Oem;Lg Frmt	6	7	Extremely Strong
3	CRTDG;TONER;LJ;HP 13... Printer Supplies-Oem;Lar,Cir	CRTDG;TONER;LJ;HP 13... Printer Supplies-Oem;Lar,Cir	7	20	Extremely Strong
4	COLORSTX;8400;6 STX;... Printer Supply-Oem;Inksticks	INK;SOLID;PHASER8400;... Printer Suply-Oem;Inksticks	8	14	Extremely Strong
5	CRTDG;INK;SP 2200;MAT... Printer Supplies-Inkjet,Oem	CRTDG;INK;SP 2200;YW Printer Supplies-Inkjet,Oem	8	9	Extremely Strong
6	CRTDG;LSR;ICMF8170C;... Printer Supplies-Oem;Lar,Cir	DRUM;LSR;ICMF8170C Printer Supplies-Oem;Lar,Cir	8	27	Extremely Strong
7	PRINTER;MFC;COLOR;DI... Multifunction-Laser,Color	PRINTER;MFC;COLOR;DI... Multifunction-Laser,Color	10	19	Extremely Strong

Advanced Search

Add search criteria for more specific searches

Category
Exactly

+ Add Criteria

Search

Other Searches

Item Search

Perform complex queries for Item

Category Search

Perform complex queries for Category

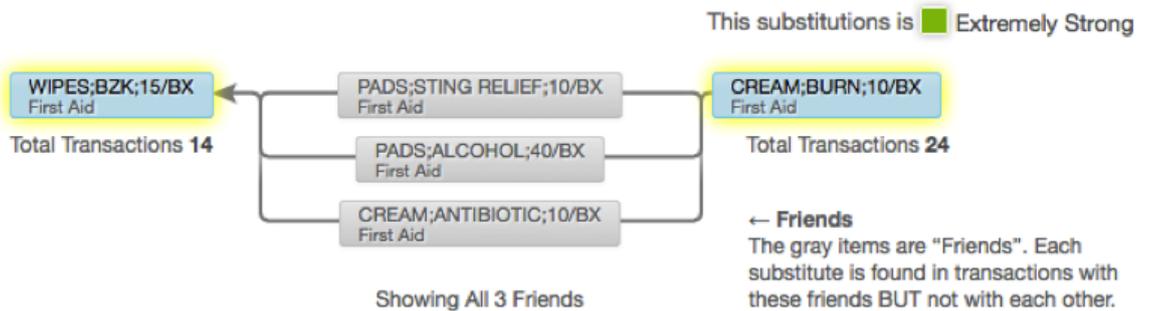
Cluster Search

Perform complex queries for Cluster

To view the details of a specific item in the list click on the icon:



Friends Details



It will show the items, transaction counts and the gray items that are "Friends". Each substitute is found in transactions with these friends BUT not with each other.

Click on the X in the upper right hand corner to close the pop up window.

Prediction Rule Search

The Prediction Rule Search allows the user to search through predictive rules for the selected outcome category, and displays the conditional probability for each rule next to the outcome item.

Prediction Rule Search Diabetes Download CSV

Your search matched **182** of the **182** total prediction rules. Search criteria: Size at least "1".

#	Freq	Prediction Rule	Probability	Outcome Item	Size	Cover	Contro
1	83	[161.0_to_199.0] Plasma glucose concentration	82.2% →	Yes Diabetes	1	31.0%	0.15
2	58	[161.0_to_199.0] Plasma glucose concentration [29.6_to_41.8] Body mass index (weight in kg/(h...	87.9% →	Yes Diabetes	2	21.6%	0.11
3	202	[0.0_to_102.0] Plasma glucose concentration	90.6% →	No Diabetes	1	40.4%	0.10
4	271	[21.0_to_27.364599] Age (years)	81.6% →	No Diabetes	1	54.2%	0.090
5	126	[0.0_to_102.0] Plasma glucose concentration [21.0_to_27.364599] Age (years)	95.5% →	No Diabetes	2	25.2%	0.076
6	39	[161.0_to_199.0] Plasma glucose concentration [122.0_to_846.0] 2-Hour serum insulin (mu U/ml)	84.8% →	Yes Diabetes	2	14.6%	0.073

Advanced Search
Add search criteria for more specific searches

Size
At least Numeric Value

+ Add Criteria Search

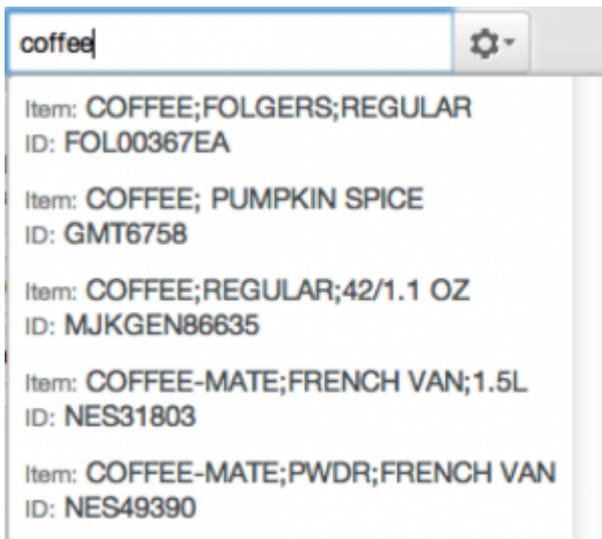
Other Searches

- [Item Search](#)
Perform complex queries for item
- [Category Search](#)
Perform complex queries for category
- [Cluster Search](#)
Perform complex queries for cluster
- [Substitution Search](#)
Perform complex queries for substitution

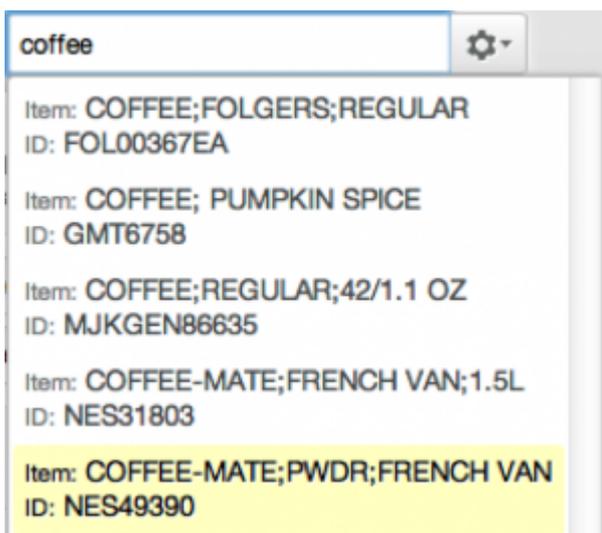
Similar to other searches, filters can be applied by typing in the search box or by dragging items over to the search area on the right.

Keyword Searches

To search for a specific item or SKU in Patterns simply type the keyword in the search field near the gear icon.



It will auto populate any matches that it may find. Hover the mouse over the specific name and click on it to bring up the item details.



Advanced Searches Using AND

While using any of the various searches the user can add criteria to the existing search to make it more granular. This item search is configured to look for the word “paper” to be contained in the Name and Item Name fields.

Item Search ACME Office Supply

Download CSV

Your search matched 1,393 of the 44,317 total Items. Search criteria: Name or ID contains "paper".

#	Item	Category	Freq	Int Cx	Rnk Cx	Clstrs	Trend
1	Paper;Copy;8.5X11;We;9... Paper-Copy	Paper-Copy	123,038	18,643	174	184	
2	Paper;Copy;4200Dp;8.5X... Paper-Copy	Paper-Copy	79,166	18,098	9	9	
3	Paper;Copy;8.5X11;We;9... Paper-Copy	Paper-Copy	65,641	16,109	4	4	
4	Paper;8.5X11;20#;We Paper-Multipurpose	Paper-Mul...	59,635	15,358	6	6	
5	Paper;Tdp;20#;8.5X11;92... Paper-Copy	Paper-Copy	58,183	16,607	6	6	
6	Clip;Paper;Jumbo Clips;Clamps & Fasteners	Clips;Cla...	56,146	17,834	40	42	
7	Paper;Discovery;20Lb;8.5... Paper-Multipurpose	Paper-Mul...	46,769	14,319	12	13	
8	Clip;Paper;#1;Regular Clips;Clamps & Fasteners	Clips;Cla...	44,896	16,920	25	25	

Advanced Search

Add search criteria for more specific searches

Name or ID

Contains

+ Add Criteria

Other Searches

Category Search

Perform complex queries for Category

Cluster Search

Perform complex queries for Cluster

Substitution Search

Perform complex queries for Substitution

We want to add another keyword to this search.

Click on the link + Add Criteria

Advanced Search

Add search criteria for more specific searches

Name or ID

Contains

AND

Name or ID

Contains

+ Add Criteria

Enter the additional search parameter "staple" and click on Search.

Item Search ACME Office Supply

Download CSV

Your search matched 2 of the 44,317 total Items. Search criteria: Name or ID contains "paper" and Name or ID contains "staples".

#	Item	Category	Freq	Int Cx	Rnk Cx	Clstrs	Trend
1	Staples; Paperpro Standard Staples	Staples	4,472	7,824	3	3	
2	Staples; Paperpro Heavyd... Staples	Staples	959	2,764	2	2	

Advanced Search

Add search criteria for more specific searches

Name or ID

Contains

AND

Name or ID

Contains

+ Add Criteria

You will notice that the resulting search has items that have both "paper" and "staple" in the item names.

Advanced Searches Using OR

A search can also be performed using an "Or" between 2 items. For example if you wanted to search for Item Names that start with either "staple" or "remover" you could edit your search to look like this:

Item Search ACME Office Supply

Download CSV

Your search matched 462 of the 44,317 total Items. Search criteria: Name or ID contains "staple|remover".

#	Item	Category	Freq	Int Cx	Rnk Cx	Clstrs	Trend
1	Staples;Chisel Pt;Std;5M/Bx Staples	Staples	68,210	19,377	85	85	
2	Staples;Std;210/Strip;5M/... Staples	Staples	43,990	17,239	18	18	
3	Staples;Chisel Pt;Std;5M/Bx Staples	Staples	40,197	16,395	31	31	
4	Staples;Std;High Quality;5M Staples	Staples	30,008	15,328	9	10	
5	Remover;Staple Staple-Removers	Staple-Re...	22,548	14,892	30	33	
6	Stapler;Desk;Econ;Full;Bk Staplers-Desk	Staplers-D...	18,824	13,347	6	6	
7	Stapler;Desk;Std;Full;Bk Staplers-Desk	Staplers-D...	17,647	12,982	5	5	

Advanced Search

Add search criteria for more specific searches

Name or ID

Contains

+ Add Criteria

Other Searches

[Category Search](#)

Perform complex queries for Category

[Cluster Search](#)

Perform complex queries for Cluster

[Substitution Search](#)

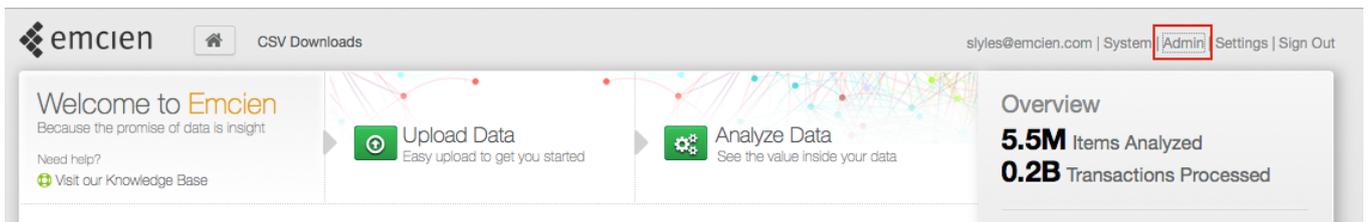
Perform complex queries for Substitution

The key to using the "Or" search is to use the pipe symbol " | " as a separator between the names like this: staple|remover

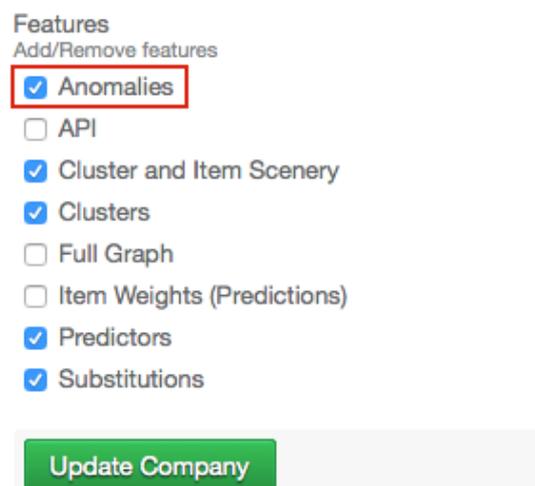
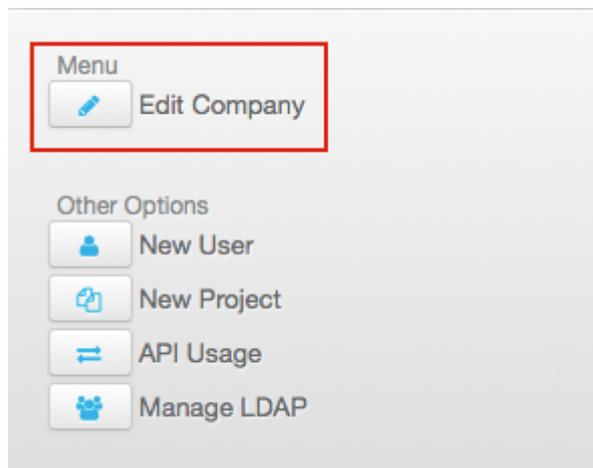
Anomalies

Enabling Anomalies

If *Anomalies* is included in the Patterns license, it can be enabled by editing the company settings. It is disabled by default. You must be logged into Patterns as an Admin. Then click on the “Admin” menu in the upper right.



Now click on “Edit Company” in the upper right.



In the lower left, make sure “Anomalies” is checked and click “Update Company” to save the changes.

Setting Up the Analysis

Data for Anomalies detection must be over a time period. You must choose a category to be the Actor and then another for the Action.

The process is to analyze a moving window of data over time looking at only the new time period compared to the included, previous time periods. For example, you may be analyzing 1 year of data looking for anomalies in the new data of the last month. Each month of new data would be added to the previous 11 months of historical data to produce a new report. Then over time you can see any new anomalies appear compared to previous reports.

NOTE: All the examples in these instructions are based on the analysis of three rolling sets of insurance claims data. The actor in these is the claimant and the action is claim_type.

[ACME_Insurance_11.07_to_10.08.long.csv](#)

[ACME_Insurance_12.07_to_11.08.long.csv](#)

[ACME_Insurance_01.08_to_12.08.long.csv](#)

Start a new analysis as you would any other. Upload you data file and select Analyze Data. In the “New Analysis” window you will have an “Anomaly Detection” selection.

New Analysis

New File to Analyze

Source Data Filter by Name

Name	File Size	Modified
<input checked="" type="checkbox"/> ACME_insurance.long.csv	32 MB	Sep 21 9:07AM
<input type="checkbox"/> Insurance_1K.long.csv	590 KB	Sep 18 2:53PM
<input type="checkbox"/> Insurance_1K.receipt.csv	500 KB	Sep 18 2:48PM
<input type="checkbox"/> Insurance_1K.csv	470 KB	Sep 15 4:01PM
<input type="checkbox"/> Trucks.wide.csv.gz	440 KB	Sep 14 4:50PM
<input type="checkbox"/> FDIC-ROA.wide.csv.gz	160 KB	Sep 14 4:50PM
<input type="checkbox"/> ACME-Retail-Sales.json.gz	70.00 MB	Sep 14 4:50PM
<input type="checkbox"/> Airlines-Delay.wide.csv.gz	2.7 MB	Sep 14 4:50PM

[Upload New Source Data](#)

Projects

Help organize Analysis on the home page

Default Project

Predictions

Anomaly Detection

Anomaly Detection Settings

Actor Category
For Anomaly Detection

claimant

Actor Category

Action Category
For Anomaly Detection

claim_type

Action Category

Tailed Analysis
Analyze Items present in last Time Period

Detect Anomalies in last time period

Override Defaults

Report Name

ACME insurance

[Start Analysis](#)

[Cancel](#)

To detect anomalies just over the last time period in the data check “Tailed Analysis” in the Anomalies settings. The time period for this tail can be set by clicking the “+” beside the

“Override Defaults” settings to view the choices.

Actor Category
For Anomaly Detection

claimant

Action Category
For Anomaly Detection

claim_type

Tailed Analysis
Analyze Items present in last Time Period

Override Defaults

Allow Limited Invalid Lines
Report will load if less than 10% of the file lines are invalid

3 Pattern Threshold
Number of times a group of items must repeat to be considered a Pattern

7 Maximum Cluster Size
Set the maximum Clusters size

UTF-8 File Encoding
Specify an encoding

Comma Field Delimiter
Specify a field delimiter

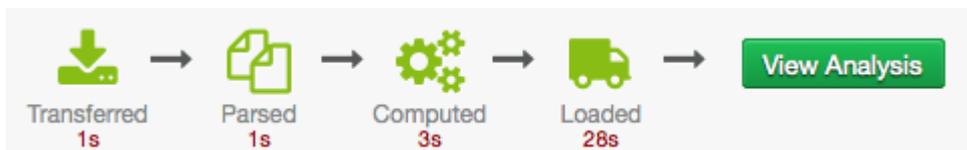
1 Month Time Period
Supports time ranges up to 100 Months.

Select time period for Tailed Analysis

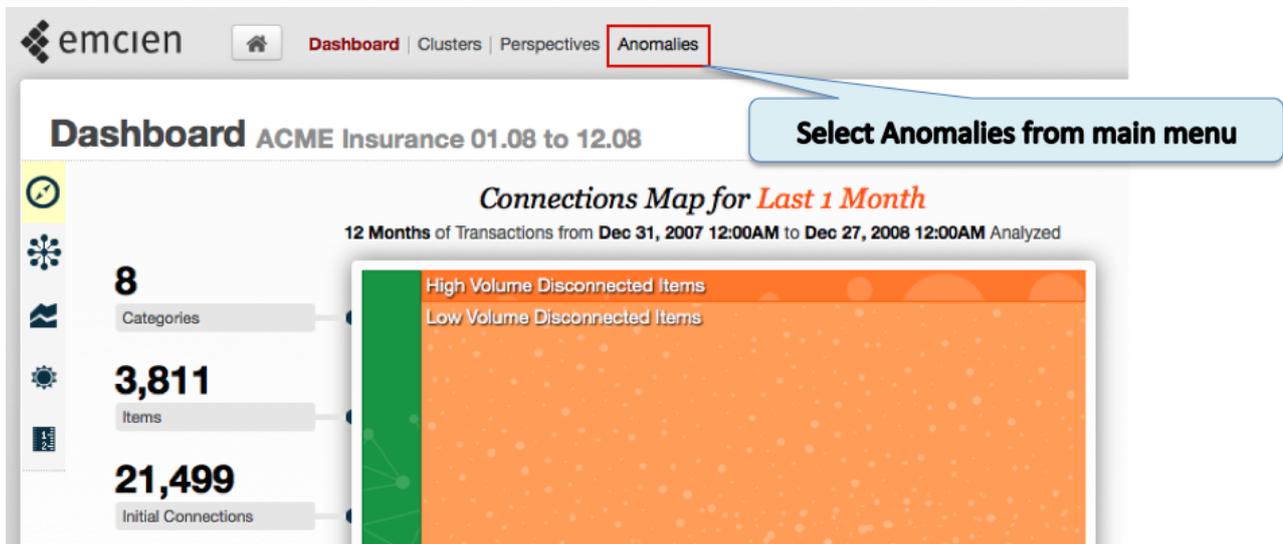
Here you can set the “Time Period” for the tailed analysis. If this parameter is not set the default is “Auto”. In Auto Patterns will select a time period based on the data.

Interpreting the Results

After an analysis completes click on the green “View Analysis” button to go to the Report Dashboard.



On the report Dashboard click on “Anomalies” from the main menu to view the Anomalies report page.



The Anomalies report will give you five selections as shown below.

Anomalies for [Download CSV](#)

<p>Surprising Behaviors</p> <p>When several Connections between an Actor & Action are surprising.</p> 	<p>Brand New</p> <p>When an Actor or Action has never been observed in the historical data.</p> 	<p>Connections Spike</p> <p>When the number of Connections between an Actor & Action dramatically increase.</p> 	<p>Timeline Surprise</p> <p>When an Actor or Action exhibits an unpredictable increase or decrease in frequency.</p> 
--	--	---	---

Actor / Action Dropdown

The dropdown allows you to view anomalies by the Actor or the Action. In the example shown above with insurance data it is showing our Actor of claimant by default.

Surprising Behaviors

Selecting this category will show when there is an unusual change in connections between an Actor and Action compared to the historical data. An example of this could be a claimant that has historically made claims of type A and then suddenly makes several claims of type B.

Brand New

This section of the report will show when an Actor or Action appears in the tailed time period but was not observed in the historical data. For example, in the sample insurance data we are analyzing a 12-month period but only looking for anomalies (Tailed Analysis) in the last month compared to the previous 11 months (historical data).

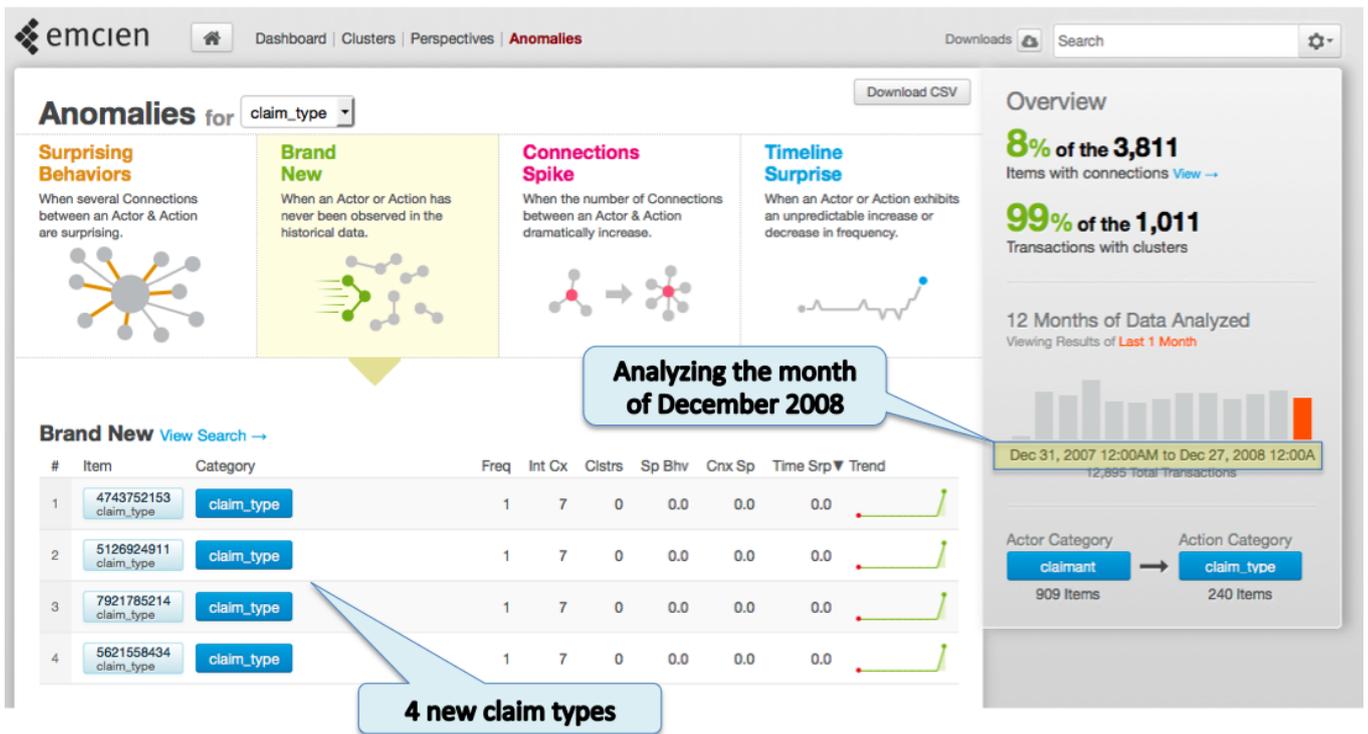
This is not interesting for an Actor of claimant in our sample data. There can be many new claimants each month. A more interesting use case is looking at brand new claim types.

In the insurance data, for example, let's look at the data for Nov 30, 2007 to Nov 27, 2008 looking for anomalies in November. If we click on "Brand New" and choose the action of claim_type from the drop down, we will see that there are no

results – so no new claim types that month.

The screenshot displays the Emcien Anomalies dashboard for the month of November 2008. The dashboard is titled "Anomalies for claim_type" and includes a "Download CSV" button. It features four main anomaly categories: "Surprising Behaviors", "Brand New", "Connections Spike", and "Timeline Surprise". The "Brand New" category is highlighted in yellow. Below these categories is a table titled "Brand New" with columns for "#", "Item", "Category", "Freq", "Int Cx", "Clstrs", "Sp Bhv", "Cnx Sp", "Time Sp", and "Trend". The table shows "No Results". To the right, the "Overview" section displays "9% of the 4,119 Items with connections" and "100% of the 1,130 Transactions with clusters". It also shows "12 Months of Data Analyzed" and "Viewing Results of Last 1 Month". A bar chart shows the distribution of transactions over time, with a highlighted period from "Nov 30, 2007 12:00AM to Nov 27, 2008 12:00A" and "12,638 Total Transactions". At the bottom right, the "Actor Category" is "claimant" (1,008 Items) and the "Action Category" is "claim_type" (247 Items). Two callout boxes are present: one pointing to the "Brand New" table with the text "No new claim types", and another pointing to the "Overview" section with the text "Analyzing the month of November 2008".

Now if we analyze the data for Dec 30, 2007 to Dec 27, 2008 looking at for anomalies in December – the next month - we suddenly see 4 new claim types appear. These are new claim types that have not appeared in the previous 11 months.



Connections Spike

Choosing this section of the report will show when there are an increased number of connections between an actor and action. In our insurance claims example this could be a claimant suddenly making a large number of claims of type A.

If we look at the analysis for Dec 30, 2007 to Dec 27, 2008 again and select “Connections Spike” we will see there is only a single result for claimant.

Anomalies for claimant

Download CSV

Surprising Behaviors

When several Connections between an Actor & Action are surprising.



Brand New

When an Actor or Action has never been observed in the historical data.



Connections Spike

When the number of Connections between an Actor & Action dramatically increase.



Timeline Surprise

When an Actor or Action exhibits an unpredictable increase or decrease in frequency.



Connections Spike [View Search](#) →

#	Item	Category	Freq	Int Cx	Clstrs	Sp Bhv	Cnx Sp▼	Time Srp	Trend
1	ALYCIA CASPER claimant	claimant	4	11	9	0.053	0.001	0.044	

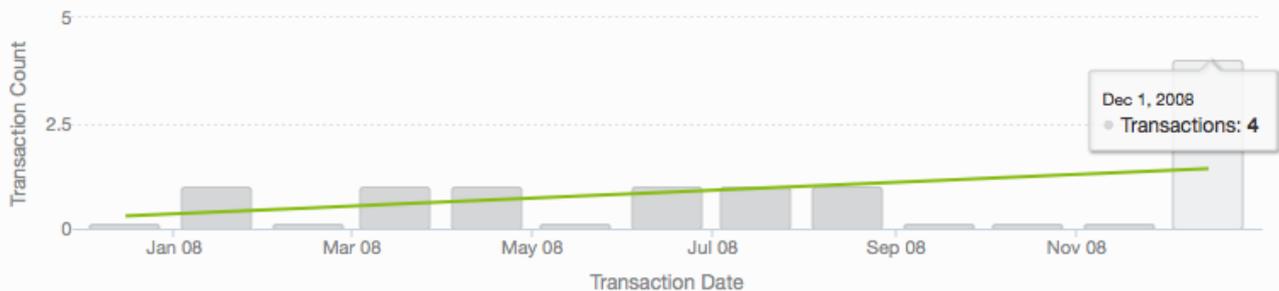
Our claimant Alycia Casper had only a few claims during the year but suddenly make a jump in claims for Dec. 2008. Notice the orange halo around the bubble with her name. This indicates that this actor has anomalies. We can click on the ALYCIA CASPER bubble to get more information. When we do we see the detailed screen below.

Item Detail

ALYCIA CASPER
claimant

Transactions for this item are trending up at less than 1 per month

Transactions from Dec 31, 2007 12:00AM to Dec 27, 2008 12:00AM Analyzed



Ranked Connections | Clusters | Initial Connections

Viewing 2-item clusters after ranking

Download Connections CSV

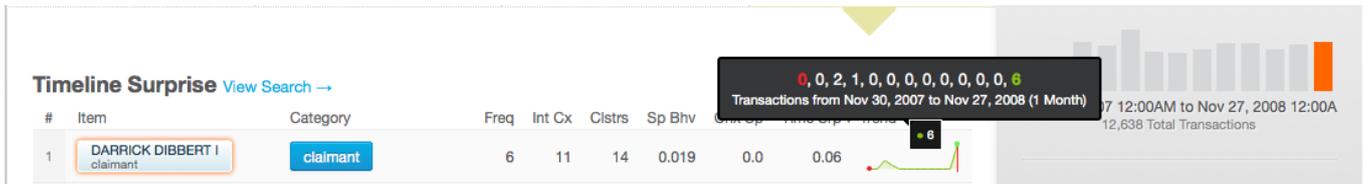
#	Items	Strength ¹	Freq	Cnd Prob	Time Srp	Trend
1	PC1BN9YT58 policy	ES	4	100.0%	0.021	
2	13786-8808 zip	S	4	100.0%	0.017	
3	OKLAHOMA accident_state	S	4	100.0%	0.018	
4	OSINSKI, RYAN AND PAUCEK agent	S	4	100.0%	0.021	
5	2432397705 claim_type	S	4	100.0%	0.002	

Notice that she only made a few single claims during the year but that jumped to 4 in December.

Timeline Surprise

Clicking on the Timeline Surprise section shows when an actor or action as an unpredictable increase or decrease in activity. This is different from connections spike in that it looks at changes in activity compared to the typical or average over the historical data - not just increases. So, it's not necessarily a spike in activity as it is a change compared to the historical norm. If we look at the analysis for the data ranging from Nov 30, 2007 to Nov 27, 2008 and choose

“Timeline Surprise” for claimant we see that the first anomaly is our claimant Darrick Dibbert I.



He made 1 and 2 claims earlier in the year then nothing until Nov when 6 claims were made. This can be a connection spike as well for the claimant, but as a timeline surprise it is a change in activity compared to his normal activity of 0.

Preparing Your Data

This article covers how to format and save your data before uploading to Emcien.

Choosing a File Format

To use Emcien, you'll need to save your data as a CSV file using one of the following Emcien formats:

Wide

Long

Tagged

JSON

The Wide Format

The wide format is the most commonly used and is easy to start with. The wide format consists mostly of user-defined columns with no required column headings.

Wide Format Example

User Defined Data

date	volume	ID	Age Group	Race	Birth State	Residence State	Gender	Education	Visits per month
2/3/13	1	C-1002010	25-35yrs	Asian	Birth-CA	Residence-FL	Male	High School	3-5 visits
2/3/13	1	C-1002011	25-35yrs	Caucasian	Birth-OR	Residence-OR	Female	Graduate school	Less than 3
2/3/13	1	C-1005401	55-65yrs	Latin	Birth-GA	Residence-FL	Female	University	Over 10 visits
2/3/13	1	C-1005660	55-65yrs	Caucasian	Birth-NY	Residence-CT	Female	University	Less than 3

Volume is not a required column but must be in the first column if it exists and there is no date. If the date exists, volume must be in the second column.

Date is not required but must be in first column

Each row represents a Transaction

The Long Format

The long format is typically used for transaction data containing many variables and combinations. In the long format, categories and their corresponding items are represented vertically and grouped based on a unique transaction ID.

Long Format Sample

Optional Data

transaction_id	item_id	transaction_date	item_name	item_category	transaction_price	item_volume	item_price	transaction_volume
101001	395847	12/15/06	4 in SQR COVER	CONDUIT	\$27.39	1	\$16.92	1
101001	457634	12/15/06	8IN NATURAL CABLE TIE	TOOLS & ACCESSORIES	\$27.39	1	\$10.47	1
101002	272840	12/31/06	20A IVORY GFCl	WIRING DEVICES	\$127.67	1	\$15.69	1
101002	668138	12/31/06	4" in SQUARE COVER GFC	CONDUIT	\$127.67	3	\$5.23	1
101002	923352	12/31/06	DRAIN TAPE-ENCASED SINK	PLUMBING	\$127.67	1	\$96.29	1

1st Transaction

2nd Transaction

The Tagged Format

The tagged format consists of two required columns (which may be empty) and no headers. There may be a maximum of 1000 categories per data file, with a maximum of 1000 items per line. An example of the tagged format is below:

Tagged Format Example		User Defined Data			
2012-07-15	transaction1	Model::GT	Color::White Pearl	Automatic	Dark Grey
2012-07-16		Model::Base	"Color::Pearl, Black"	Manual-5 Speed	Brushed Aluminum
2012-07-16	transaction5	Model::Base		Manual-6 Speed	Brushed Aluminum
2012-07-17	transaction6	Model::Base	Color::Gray		Dark Grey

↑ The first column must be the date, but it may be empty

↑ The second column must be the transaction ID, but it may be empty

↑ Items may be represented as Category::Item, or just as the item

The JSON Format

The JSON format contains no headers and is an industry standard format for data storage. JSON files are great for transactional data containing many variables and combinations, and supports large item "baskets" that have up to 1000 items per transaction.

```

{
  "tid": "5698132.2415963",
  "entries": [
    {
      "id": "548E23",
      "desc": "1\ LARGE BRUSH WITH STEEL HANDLE",
      "cat": "BRUSHES",
      "volume": 1,
      "value": 1.58
    },
    {
      "id": "543F52",
      "desc": "1\SMALL ROLLER W/ WOOD HANDLE",
      "cat": "BRUSHES",
      "volume": 1,
      "value": 0.82
    },
    {
      "id": "981A28",
      "desc": "2\PAINT CLOTH",
      "cat": "SCRUBBERS",
      "volume": 1,
      "value": 0.67
    }
  ],
  "date": "2011-01-01",
  "volume": 1,
  "revenue": 21.32
}

```

Each transaction must start with the transaction ID

All entries must have an "id" or "sku"

Double quotes must be escaped as JSON strings (with a backslash)

Not sure which format is best for you? The Emcien team can help prepare your data. Contact us at support@emcien.com.

Compression

Emcien supports CSV files compressed using GZip (.gz) compression. Compressed CSV files must still use either the wide or receipt format.

Character encoding

Emcien supports the following character encodings:

UTF-8

UTF-16

Latin-1

[Simplified Chinese](#)

[Windows-1251](#)

[Windows-1252](#)

[Shift_JIS](#)

Line endings

Patterns uses either Unix-style line endings `\n` or DOS line endings `\r\n`.

Naming Your Data file

Emcien uses a file naming convention to identify the Emcien format used. The filename is limited to 200 characters or less. Files should be named using the following structure:

`<filename>.<file type>.<extension>`

filename	Filenames should use only the following characters: <ul style="list-style-type: none">• ASCII alphanumeric characters in upper (A-Z), lower (a-z) case, 0-9• Whitespace• Periods .• Underscores _• Hyphens -• Parentheses () The following characters <i>are not</i> supported: <ul style="list-style-type: none">• Backslash (\)• Single quote (')• Double quotes (")
file type	receipt or wide Important: If no file type is included, the default receipt format is used. This will result in loading errors for wide format files.
extension	.csv or .csv.gz

File Name Examples

area2-sales.receipt.csv	Uncompressed sales data from area 2 in the receipt format
members.us.wide.csv	Uncompressed U.S. membership data in the wide format
area3-sales.csv.gz	Compressed sales data from area 3 in the receipt format
clinical.all.wide.csv.gz	Compressed clinical test data in the wide format

Common Pitfalls

Having trouble analyzing your data? The following pitfalls could cause an error message or insufficient results:

Replacing blank or missing values with placeholders, such as NULL, false, or n/a. Emcien provides faster, more accurate results when missing values are left blank.

Loading unbanded (or unbinned) numerical data. Unbanded numerical data often results in few connections.

Need help binning your data? You can try our automated data prep module, [Bandit](#) (BETA). If you need further help please contact the Emcien team at support@emcien.com

Loading files that are not in a CSV format, including files using the .csv file extension but that are not comma delimited.

Transforming data into [RDF](#) or Boolean values (1s and 0s) reduces the effectiveness of Emcien's analysis.

Shifted or improperly-entered data in the long format

Using the incorrect date and time format. Check out the [Date and Time](#) section of our [Preparing Your Data](#) article for more information.

Using negative prices in the item_price or transaction_price columns in the long format. Emcien *does not* support negative prices.

Many of these errors can be overcome without any data cleanup or transformation. If your analysis results in a "parse" error and there are only a limited number of lines with errors, the analysis can continue. Simply navigate

back to the "New Analysis" page, open the "Override Defaults" button, and select "Allow Limited Invalid Lines".

Override Defaults

Allow Limited Invalid Lines
Report will load if less than 10% of the file lines are invalid

Cluster Threshold
Number of times a Cluster must occur

Maximum Cluster Size
Set the maximum Clusters size

File Encoding
Specify an encoding

Time Period
Auto select the best period for the data

This will remove those invalid lines and allow the analysis to continue.

Still having issues? We're here to help. Contact the Emcien team at support@emcien.com.

Adding Derivative Data

A similar concept to data banding, derivative data allows analysts to create additional dimensions from existing data. This additive information is perfectly suited for the kinds of analysis in Emcien.

Example 1: If each transaction includes a specific location, either address, GPS coordinates, or even a state, an analyst could derive additional information by creating additional columns. This data could be used to add regions or time zones to the data set, giving the software additional opportunities to find patterns.

Lat.	Long.
33.65	84.25
41.90	87.39
32.73	96.58
33.93	118.24



Location
Atlanta
Chicago
Dallas
Los Angeles

Example 2: In cases where dates are included in the data, it can be beneficial to derive a day of the week attribute from the date value. Patterns may be detected on days of the week more than days of the year.

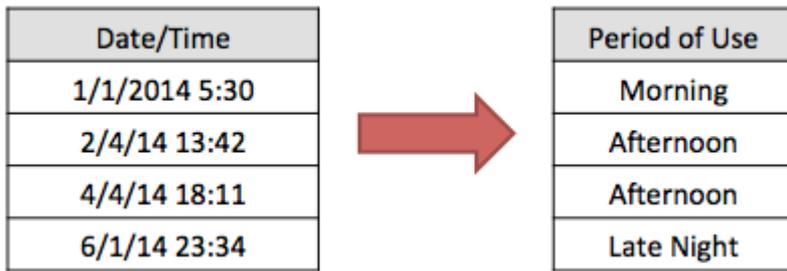
Date
1/1/2014
12/25/2013
11/31/2013
7/4/2014



Day of Week
Wednesday
Wednesday
Thursday
Friday

Example 3: When data contains date and time at a very granular level it can be effective to derive an attribute to create larger groupings of events or usage. For example analysis of web traffic, understand the patterns for different periods of

the day.

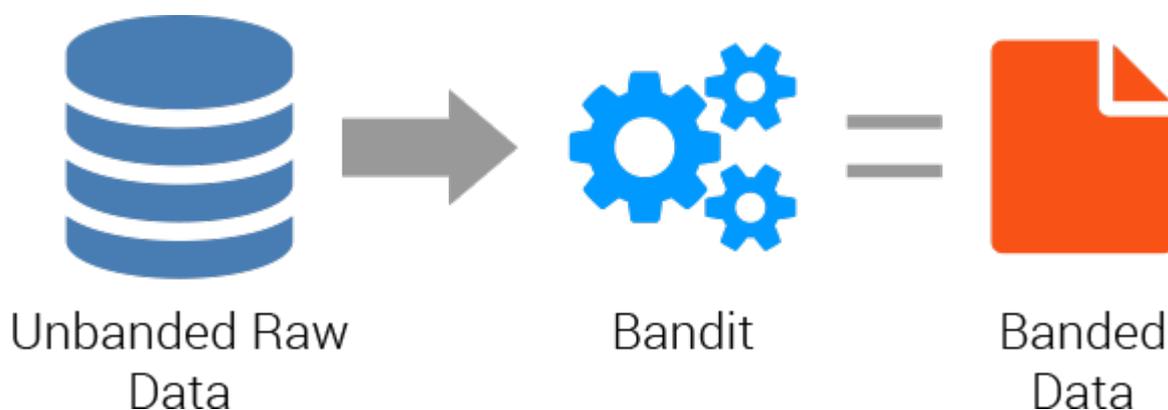


Using Bandit to Prepare Data

What is Bandit?

Emcien Bandit is a command line tool that provides automated, high-speed, smart numeric data banding (or binning). Bandit automatically transforms raw numeric data into categorical data for analysis projects and pipelines. Emcien customers analyzing numeric data can easily pre-process data using Bandit's built-in formulas or create their own bands. You can reference the full list of Bandit Commands

here: <http://support.emcien.com/help/article/link/bandit-commands-reference>



Bandit can work across multiple operating systems, including:

Redhat Linux

OSX

Windows

To download the latest version of Bandit, contact [Emcien support](#).

What input file do I need?

A delimited text file in the Emcien [wide format](#) containing the input data you want to band. One example is this [diabetes research data](#).

How do I use Bandit?

Unzip the package and save the Bandit executable file to a location on your computer. We recommend that the file be renamed to "Bandit" to simplify use of the processing commands. Renaming is not required, but you must use the correct file name in the processing commands.

```
-h          help:      Emcien Bandit Help (www.emcien.com)

-d CHAR    dependent:  name of the dependent category

-f CHAR    form:       form of output file (W for wide or R for receipt)

-s CHAR    separator:  Separator character

-t CHAR    table:      name of table file

-b CHAR    bands:      name of bands file

-r INT     randomize:  1 for 10pct test; 2 for 20pct test

-v INT     verbosity:  Print out lots of extra information on stderr
```

How do I band a file?

Processing data using Bandit involves creating the input files, then passing them along into Bandit using a command line parameter.

Here is an example using the diabetes data set:

```
bandit -t diabetes_input_data_file.csv
```

The above example will:

1. Ingests the input file (-t "diabetes_input_data_file.csv").
2. Returns a banded file (banded_diabetes_input_data_file.csv) optimized for analysis and a breaks file (breaks_diabetes_input_data_file.csv) containing bands and methods used.

What if I have a defined outcome category?

When using Emcien Bandit to process data that contains a defined outcome (dependent) category, you will use the -d parameter to pass in the dependent category name.

Here is an example:

```
bandit -t diabetes_input_data_file.csv -d "Diabetes"
```

The above example will:

1. Ingests the input file (-t "diabetes_input_data_file.csv")
2. Use the category "Diabetes" as the dependent (-d "Diabetes") variable.

3. Returns the following output files:

Banded File (banded_diabetes_input_data_file.csv)

Auto-Banded file optimized for analysis.

Breaks File (breaks_diabetes_input_data_file.csv)

Contains the bands and methods used during the auto banding process.

Mutual Information file (mi_diabetes_input_data_file.csv)

Provides a measurement of the relationship between each category and the dependent category.

What if I want to control how the banding works?

If the banded file has:

columns you would like to exclude

columns that were banded that should not have been

columns that you want to override with another system-defined band, your own bands, or strings,

follow these simple instructions to create a user bands file.

Open up the breaks_diabetes_input_data_file.csv file that was just created.

Sample:

	A	B	C	D	E	F	G
1	Diabetes	cat					
2	Number of times pregnant	wide	raw	0	0.5	111	
3	Number of times pregnant	wide	raw	0.5	1.5	135	
4	Number of times pregnant	wide	raw	1.5	2.5	103	
5	Number of times pregnant	wide	raw	2.5	3.5	75	
6	Number of times pregnant	wide	raw	3.5	4.5	68	
7	Number of times pregnant	wide	raw	4.5	5.5	57	
8	Number of times pregnant	wide	raw	5.5	6.5	50	
9	Number of times pregnant	wide	raw	6.5	7.5	45	
10	Number of times pregnant	wide	raw	7.5	8.5	38	
11	Number of times pregnant	wide	raw	8.5	9.5	28	
12	Number of times pregnant	wide	raw	9.5	10.5	24	
13	Number of times pregnant	wide	raw	10.5	11.5	11	
14	Number of times pregnant	wide	raw	11.5	12.5	9	
15	Number of times pregnant	wide	raw	12.5	13.5	10	
16	Number of times pregnant	wide	raw	13.5	14.5	2	
17	Number of times pregnant	wide	raw	14.5	16	1	
18	Number of times pregnant	wide	raw	16	17	1	
19	Plasma glucose concentration	freq	scale	0	96	160	
20	Plasma glucose concentration	freq	scale	96	110	154	
21	Plasma glucose concentration	freq	scale	110	126	157	
22	Plasma glucose concentration	freq	scale	126	150	154	
23	Plasma glucose concentration	freq	scale	150	199	143	
24	Diastolic blood pressure (mm Hg)	freq	raw	0	61	158	
25	Diastolic blood pressure (mm Hg)	freq	raw	61	70	160	
26	Diastolic blood pressure (mm Hg)	freq	raw	70	76	161	
27	Diastolic blood pressure (mm Hg)	freq	raw	76	84	154	
28	Diastolic blood pressure (mm Hg)	freq	raw	84	122	135	
29	Triceps skin fold thickness (mm)	freq	log	0	7	227	
30	Triceps skin fold thickness (mm)	freq	log	7	24	172	
31	Triceps skin fold thickness (mm)	freq	log	24	33	181	
32	Triceps skin fold thickness (mm)	freq	log	33	46	157	
33	Triceps skin fold thickness (mm)	freq	log	46	99	31	

Column A: Categories within your data

Column B: Method of Banding used

Column C: Ignore

Column D-E: Band range

Column F: Number of rows in the data that contain that band

Column G: Renaming the Bands

Options to create user bands file:

1. To apply a system defined band:

delete all but one line for each category you want to apply a system defined band

	A	B	C	D	E	F	G	H
1	Diabetes	cat						
2	Number of times pregnant	freq3						
3	Number of times pregnant	wide	raw	0.5	1.5	135		
4	Number of times pregnant	wide	raw	1.5	2.5	103		
5	Number of times pregnant	wide	raw	2.5	3.5	75		
6	Number of times pregnant	wide	raw	3.5	4.5	68		
7	Number of times pregnant	wide	raw	4.5	5.5	57		
8	Number of times pregnant	wide	raw	5.5	6.5	50		
9	Number of times pregnant	wide	raw	6.5	7.5	45		
10	Number of times pregnant	wide	raw	7.5	8.5	38		
11	Number of times pregnant	wide	raw	8.5	9.5	28		
12	Number of times pregnant	wide	raw	9.5	10.5	24		
13	Number of times pregnant	wide	raw	10.5	11.5	11		
14	Number of times pregnant	wide	raw	11.5	12.5	9		
15	Number of times pregnant	wide	raw	12.5	13.5	10		
16	Number of times pregnant	wide	raw	13.5	14.5	2		
17	Number of times pregnant	wide	raw	14.5	16	1		
18	Number of times pregnant	wide	raw	16	17	1		
19	Plasma glucose concentration	info	raw	0	102	223		
20	Plasma glucose concentration	info	raw	102	123	214		
21	Plasma glucose concentration	info	raw	123	143	149		
22	Plasma glucose concentration	info	raw	143	161	81		

in column B of the category you want to change, type one of the system defined banding options** and delete the remaining contents of that row

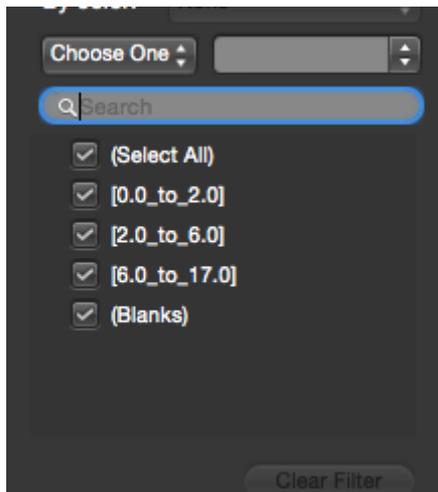
	A	B	C	D	E	F	G	H	I
1	Diabetes	cat							
2	Number of times pregnant	freq3							
3	Plasma glucose concentration	info	raw	0	102	223			
4	Plasma glucose concentration	info	raw	102	123	214			
5	Plasma glucose concentration	info	raw	123	143	149			
6	Plasma glucose concentration	info	raw	143	161	81			
7	Plasma glucose concentration	info	raw	161	199	101			
8	Diastolic blood pressure (mm Hg)	half	raw	0	40.090666	40			
9	Diastolic blood pressure (mm Hg)	half	raw	40.090666	49.762267	11			
10	Diastolic blood pressure (mm Hg)	half	raw	49.762267	59.433868	70			
11	Diastolic blood pressure (mm Hg)	half	raw	59.433868	69.105469	197			
12	Diastolic blood pressure (mm Hg)	half	raw	69.105469	78.77707	245			
13	Diastolic blood pressure (mm Hg)	half	raw	78.77707	88.44867	145			
14	Diastolic blood pressure (mm Hg)	half	raw	88.44867	98.120271	44			
15	Diastolic blood pressure (mm Hg)	half	raw	98.120271	107.791872	9			
16	Diastolic blood pressure (mm Hg)	half	raw	107.791872	117.463473	6			
17	Diastolic blood pressure (mm Hg)	half	raw	117.463473	122	1			

Save this file as something other than breaks_diabetes_input_data_file.csv

Run bandit again with your new user-defined bands file as the -b bands file

Example: `bandit -t diabetes_input_data_file.csv -d "Diabetes" -b user_input_file.csv`

bandit will apply the specified banding method to your data. In this case, the freq banding method, artificially limited to three bands



2. To create a user defined band: open the original `breaks_diabetes_input_data_file.csv` and adjust the max and min of each band range
 - replace the banding method in column B with *user*
 - type in the new min and max in columns D and E respectively
3. To apply a label to your band:
 - replace the banding method in column B with *user*
 - in column G type in the numerical or text string you want to represent your band

Save the modified breaks file (user bands file) under a different name. It is important to use a name other than bandit's format: `breaks_input_file.wide.csv`.

Sample user bands file:

	A	B	C	D	E	F	G	H
1	Diabetes	cat						
2	Number of times pregnant	freq3						
3	Plasma glucose concentration	user	raw	0	102	223	low	
4	Plasma glucose concentration	user	raw	102	123	214	medium	
5	Plasma glucose concentration	user	raw	123	143	149	high	
6	Plasma glucose concentration	user	raw	143	199	1	very high	
7	Diastolic blood pressure (mm Hg)	half	raw	40.090666	49.762267	11		
8	Diastolic blood pressure (mm Hg)	half	raw	49.762267	59.433868	70		
9	Diastolic blood pressure (mm Hg)	half	raw	59.433868	69.105469	197		
10	Diastolic blood pressure (mm Hg)	half	raw	69.105469	78.77707	245		

Here is an example:

```
bandit -t diabetes_input_data_file.csv -d "Diabetes" -  
b user_input_file.csv
```

The above example will:

1. Ingest the input file (-t "diabetes_input_data_file.csv")
2. Use the category "Diabetes" (-d "Diabetes") as the dependent variable.
3. Uses the user bands file (-b "user_input_file.csv") to specify how the data should be transformed.
4. Returns the following output files:
 1. Banded file (banded_diabetes_input_data_file.csv)
 1. User banded file optimized for analysis
 2. Breaks file (breaks_diabetes_input_data_file.csv)
 1. Contains the bands and methods used during user banding process.
 3. Mutual Information file (mi_diabetes_input_data_file.csv)
 1. Provides a measurement of the relationship between each category and the dependent category.

Bandit Commands Reference

This article is a reference

for: <http://support.emcien.com/help/article/link/using-bandit>

An overview of the command line parameters can be seen by simply calling Bandit with the “-h” parameter (bandit -h), in which you will get something like this:

```
-h          help:      Emcien Bandit Help (www.emcien.com)
-d CHAR    dependent:  name of the dependent category
-f CHAR    form:       form of output file (W for wide or R for receipt)
-s CHAR    separator:  Separator character
-t CHAR    table:      name of table file
-b CHAR    bands:      name of bands file
-r INT     randomize:  1 for 10pct test; 2 for 20pct test
-v INT     verbosity:  Print out lots of extra information on stderr
```

There are 9 commands you can pass (7 numerical banding and 2 transformational):

freq - Similar frequency bands - five bands are chosen to divide frequency of transactions into similar sizes. This command can include the number of bands you wish to have. (i.e. “freq8” would create 8 bands based on

frequency)

wide - Equal width bands - range of numerical values is divided evenly into five bands. This command can include the number of bands you wish to have. (i.e. "wide7" would create 7 equal width bands)

fuzz - Fuzzy bands organizes the data into five bands with sizes of approximately 10%, 20%, 40%, 20%, 10% of the transactions. The algorithm begins by placing the highest occurring value into the middle 40% and builds left and right of that to create the other bands.

stat - builds bands based on the standard deviation

half - builds bands based on half of the standard deviation

info - Information Banding - based on "information theory", recursively breaks down the data into chunks with the goal to reduce the entropy/disorder (basically trying to cleanly organize bands such that predictability is maximized). This banding type **REQUIRES** the dependent parameter ("-d") to be passed in the command line.

user - User defined bands allow for custom banding of the numeric ranges and assigning them names. "User" type bands **REQUIRE** the use of a "bands.csv" input file. See section on User Defined Banding for more information.

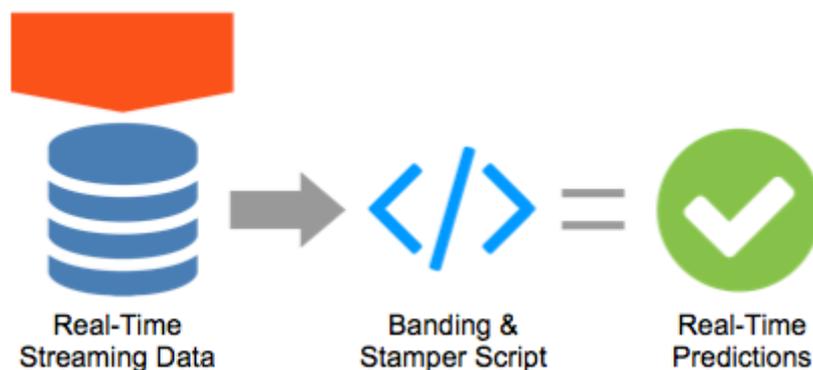
cat - Categorical Data command tells bandit to leave the data alone and pass it onto the output file as is.

skip - Skips the column of data and does not include it in the output file at all.

Implementing Band and Predict

What is band_and_predict?

band_and_predict is an example of a command line script that creates a repeatable process of banding and making predictions on incoming data. This is done by automatically banding a test file, making predictions on that file, and then returns a results file containing those predictions.



Note: band_and_predict requires [Bandit v50](#) or later. When installing Bandit it must be copied to a directory in the user's \$PATH variable (such as /usr/bin/, \$HOME/bin, etc).

What needs to be completed before using band_and_predict?

Before running band_and_predict make sure the training data file has been

banded and analyzed in Emcien. The Bandit breaks file output will be used to band the incoming data for predictions.

What inputs are needed when calling band_and_predict?

Patterns Authentication Token:

Allows the user to view and make predictions

Users

#	Email	Role	Last Session	Status	Menu
1	example@user.com ID: 96	User	A Few Seconds Ago	Active	

1. From the Emcien home page, click on 'Admin'.
2. Click on your email under users.
3. Generate API token and copy that token to use in the band_and_predict

API Token
Used to access the API (Optional)

command.

4. Your API key will be a 24 character string. Example:
ur390s4ju5foh2s0n989mnp7
5. Click Update User to save information.

[Update User](#)

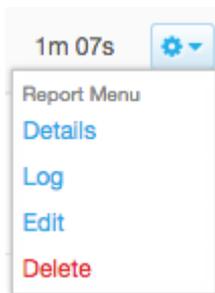
Patterns Report ID:

Identifies which training data has the predictive rules to apply to new data

1. From the the Emcien home page, click on 'Admin'.



2. Hover over settings next to your report and click details.



1. Copy ID information, in our example the ID is 68356633.

Report Attribute	Report Value
ID	68356633

Test File (CSV):

Incoming data that is banded and predictions will be made on.

Save the unbanded test file to the same path as the band_and_predict script. For example our test file could be something like 'sample_test_file.wide.csv'

The test file must have the .wide.csv extension.

Bandit Breaks file:

Contains the numerical bands to use on the incoming test data to match the same ranges from the training data.

The breaks file was created while banding the training data. This file will be used to band the test file. Save this breaks file in the same path as the band_and_predict script. An example of this file would be something like 'training_file_breaks.csv'.

SFTP server information:

Allows band_and_predict to send data to the server

1. From the Emcien home page, click on 'Admin'.



2. Then click 'Edit Company'.



1. Find the SFTP username, host, and path in Data Server.

Data Server

Data Server must accept SFTP

```
sftp://acme:patterns@disk.internal.acme.com/make-predictions
```

Format `sftp://USERNAME:PASSWORD@HOST/PATH`

1. Open the `band_and_predict` script in a text editor.
2. Insert the information about your company's SFTP client into the quotations.

```
sftp_user=${EMCIEN_SFTP_USER:-"acme"}  
sftp_host=${EMCIEN_SFTP_HOST:-"disk.internal.acme.com"}  
sftp_path=${EMCIEN_SFTP_PATH:-"make-predictions"}
```

The SFTP Password:

The password for a secure connection with the data server.

1. From the Emcien home page, click on 'Admin'.



example@user.com | Admin | Settings | Help | Sign Out

2. Then click 'Edit Company'.

Copy the password, it is the string after the ":" but before the "@" in the Data Server field.

Data Server

Data Server must accept SFTP

```
sftp://acme:patterns@disk.internal.acme.com/make-predictions
```

Format sftp://**USERNAME:PASSWORD@HOST/PATH**

Note: If the password has non-alphanumeric characters, be sure to escape them with a backslash.

The Emcien Server URL:

This is the URL for the prediction engine instance that allows `band_and_predict` to connect to the server

1. Use the base URL for your Emcien prediction engine instance (This may be the same URL as your Emcien instance). Be sure to specify http or https.

How do I use `band_and_predict`?

Begin by opening your Command Line tool.

To see an overview of the command line parameters simply call `band_and_predict` with the “-h” parameter (`band_and_predict -h`). It will return:

Emcien Band and Predict (BETA) Usage:

```
band_and_predict [-h] -a AUTH_TOKEN -r REPORT_ID -p PRED_CSV -b  
BREAKS -f PASSWD -u URL
```

Upload and make predictions on a test data set using a specified stamper.

-h Display help and exit

-a A Patterns authentication token for a user with permissions to view and make predictions

-r The Patterns Report ID containing the training data with which to make predictions

-p The CSV file to band and make predictions on

-b The Bandit breaks file that was created when banding the training data file (and will be used to band the prediction CSV)

-f Password for the SFTP user associated with the Patterns client that will be used to make predictions

-u The URL of the Emcien server. Include the protocol("http" or "https")

How Do I band_and_predict a file?

Every band_and_predict call/command requires a number of inputs to connect the results of the Train analysis with the incoming Test file. These calls/commands can be run manually through the terminal each time or can be automated as part of a larger implementation of recurring analysis and predictions.

Here is an example:

```
band_and_predict -a 'ur390s4ju5foh2s0n989mnp7' -r '68356633' -p  
'sample_test_file.wide.csv' -b 'breaks_training_file.wide.csv' -f  
'my_sftp_password' -u 'http://ip.address.of.server'
```

The above example will:

1. Allows access to Emcien API without logging in (-a 'ur390s4ju5foh2s0n989mnp7').
2. Identify in Emcien which training analysis has the rules to make predictions (-r '68356633')
3. Ingest the test file (-p 'sample_test_file.wide.csv')
4. Use the training data breaks file (-b 'breaks_training_file.wide.csv') to specify how the test data should be transformed.
5. Allows band_and_predict to send data to the SFTP server associated with the desired Emcien client (-f 'my_sftp_password').
6. Directs band_and_predict to the Emcien server URL to use (-u 'http://ip.address.of.server').
7. Returns a results file (sample_test_file.results.csv) containing the predictions on your test data.

Note: You can edit the band_and_predict script to be automated on your system or create a similar script that automatically bands and predicts new incoming data.

Advanced Environmental Variable Options:

If you would like to use environment variables instead of parameters, set the flags using the following variables:

"EMCIEN_AUTH_TOKEN"

"EMCIEN_REPORT_ID"

"EMCIEN_PRED_CSV"

"EMCIEN_BREAKS_CSV"

"EMCIEN_SFTP_PASS"

"EMCIEN_URL"

Using a flag for an option will always take precedence over environmental variables.

What if I need to adjust polling intervals?

Adjust polling interval (in seconds) according to the duration of a typical prediction made with this script by editing the "polling_duration" variable near the top of the band_and_predict script's text. For example, to poll every 5 seconds you would use "polling_duration=5".

What additional tools will I need?

While most of the utilities in band_and_predict are typically available on most unix-like systems (OSX and Linux variants), some non-standard tools must be present. To install missing utilities described below, you must be an admin on your system to provide sudo access. If this is not available to you, consult your IT staff.

lftp

OSX: <http://lftp.yar.ru/get.html>

brew install lftp (if homebrew is installed on system)

Ubuntu: sudo apt-get install lftp

RedHat: sudo rpm install lftp

curl

Emcien [Bandit v50](#) (or greater)

copy bandit to directory in the user's \$PATH variable (such as /usr/bin, \$HOME/bin, etc)

Any missing tools can be installed using APT, RPM, Homebrew (if available), or the package manager that is available on your system. Contact your system admin for assistance in doing this or reach out to Emcien's [support site](#) and team for help.

Loading Your Data

This article covers how to upload data files to Emcien using [File Transfer Protocol \(FTP\)](#).

Make sure you've prepared your data before uploading to Emcien. Check out the [Preparing Your Data](#) article for more information. You can also email us at support@emcien.com for help getting started.

Upload using Drag and Drop

Simply open the Emcien application and select "Upload Data"  Select the file that you wish to process, then drag and drop it onto the Upload Data screen  Once it has finished uploading you can then click on "Analyze This Data" to process the

file.  Upload Complete!

 or [Upload More Data](#)

Uploading Data to the Emcien Cloud

You can upload data files to Emcien using [SSH File Transfer Protocol \(SFTP\)](#) and your preferred FTP client, such as [FileZilla](#) or [Cyberduck](#).

To connect to Emcien using your FTP client, use the following credentials:

Host: feeds.emcien.com

Username: {Your Emcien Feeds server Username}

Password: {Your Emcien Feeds server Password}

Uploading Data on Mac or Linux

1. Launch your preferred FTP client. If you do not have a preferred FTP client, we recommend [Cyberduck](#).

For this article, we use screenshots of Cyberduck.

2. Click Open Connection.



3. Enter the below credentials and click Connect.

Server: Your Emcien URL. This URL was set by your IT resource during installation.

Username

Password



Your FTP client will then display the files on your server.

4. Drag and drop your data file to the server files in your FTP application.
5. On the Emcien home page click Analyze Data. You'll see the data files on this screen.

Uploading Data on Windows

1. Click the Start button . The Start menu is displayed.
2. On the Start menu, click Computer. The Computer folder is displayed.
3. Right-click anywhere in the folder and click Add a network location.
4. The Add Network Connection wizard is displayed. Click Next.
5. On the next tab, select Choose a custom network location. Click Next.
6. On the next tab, enter your Emcien URL.

This URL was set by your IT resource during installation. Ensure you include *FTP://* at the beginning of the URL.

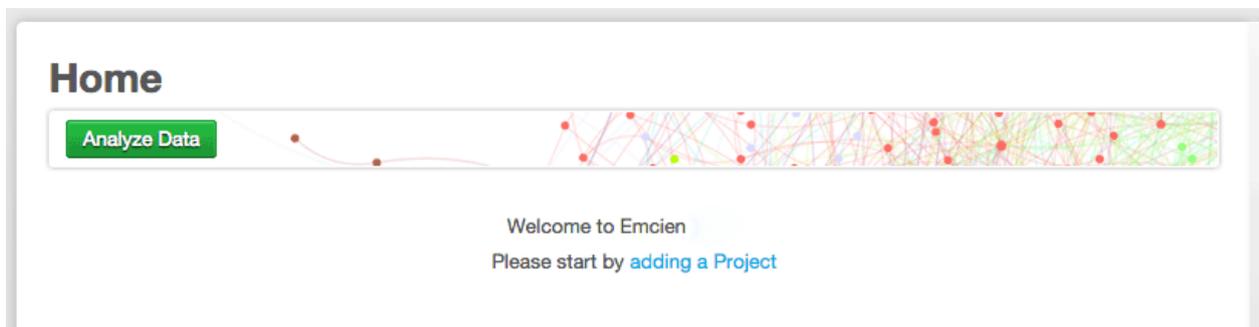
Then click Next.

7. Clear the Log on anonymously checkbox.
8. Type *data* in the username field. Click Next.
9. On the next tab, name your shortcut *Emcien* by typing it in the Type a name for this network location field. Click Next.
10. On the next tab, click Finish.
11. Type *feeds1* in the password field. Check the Save password checkbox so you can connect directly in the future.
12. Click Log On. You can now drag data files into this folder.
13. On the Emcien home page, click Analyze Data. You'll see the data files on this screen.

Analyze a Data Set

1. Using your preferred Internet Browser, navigate to the Emcien Sign In page:

- local VM users, go to: <http://localhost:5115>
- For cloud users, go to: <http://patterns.emcien.com/>



The Analyze Data button will bring up all of your uploaded data. Select your data set.

New Analysis Nick Admin

Data

Filter by Name

Name	File Size	Modified Time
sales-data.wide.csv	4.9 MB	Apr 6 5:16PM
<input checked="" type="checkbox"/> titanic-upload.wide.csv	140 KB	May 7 9:42AM

Projects

Reports in a project will display on home page

[+ Create a New Project](#)

Click on Predictions, and type in the predictive category. For this data set we will be predicting the Category "Survival".

Predictions

Prediction Category Name

Prediction Category may have 20 items or less

Item Weighted Predictions
Allow item_price to weight Prediction Rules

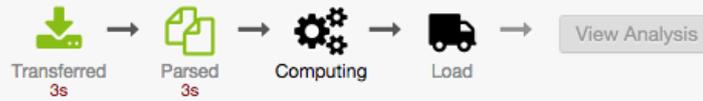
Override Defaults

Report Name

Once your project is organized, click Start Analysis.

The load screen will take you through each stage of the analysis.

Sales Data V1 Log Emcien Marketing Admin



Log: [Summary](#) | [Expanded](#)

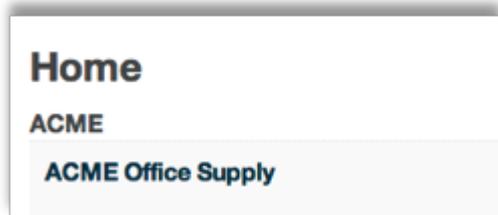
Run Time: **3m 49s**

#	Message	Time
37	--> Building Engine::Clusters::Results as clusters.dat	14s
36	--> Building Engine::Trends::TrendData as item_trends.dat	14s
35	--> Building Engine::RetailSubs::Results as retail_subs.dat	14s
34	--> Building Engine::Graph::DegreesByPeriod as degrees_by_period.dat	14s
33	--> Building Engine::Graph::Lol as graph.dat	12s
32	--> Building Engine::Index::Duplicates as duplicates.dat	12s
31	--> Building Engine::Index::Transaction as index.dat	12s
30	--> Building Engine::Trends::CatTrend as cat_trends.dat	12s
29	--> Building Engine::Index::CatTransaction as catindex.dat	12s
28	--> Building Engine::Input::CatData as catdata.dat	12s

When the analysis is complete you will hear a chime and the View Analysis button will be highlighted in green. Just click the button to see your results.

Opening a Report

From the Emcien Home Page simply click on the report name.



This will take you to the Emcien Dashboard.

Dashboard ACME Office Supply

Connections Map

15 Months of Data from Mar 1, 2011 to May 31, 2012 Analyzed

448

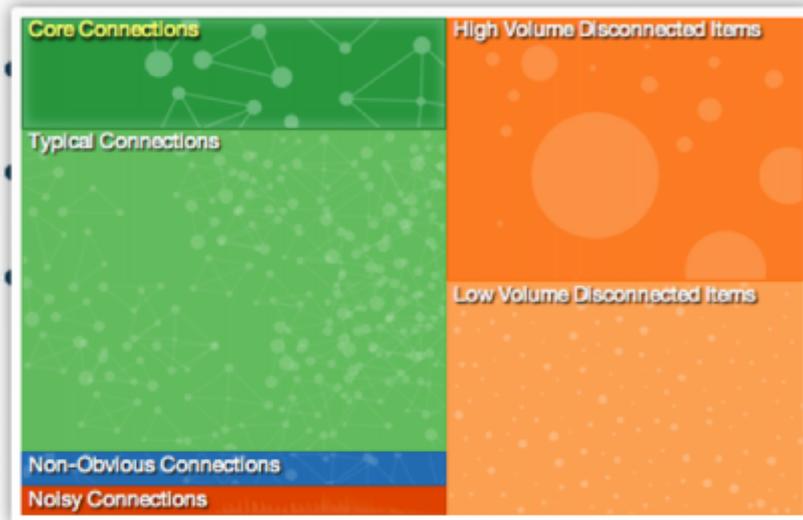
Categories

44,317

Items

18,692,080

Initial Connections



A total of 24,159 Connected Items created 18,692,080 Initial Connections

Details

■ Core Connections

Represents higher-frequency clusters with high strength and correlation. Core represents the heart of the patterns found. [View →](#)

Example Connection

- | | |
|---|---|
| Flyer;Business Essentials
Unassigned Category Code | Flyer;Business Essentials
Unassigned Category Code |
| Freq: 1,499 Conditional Probability: 99.9% Strength: Extremely Strong | |
| Crtdq;Hp 305A;2.6K;Ma
Printer Supplies-Laser;Oem | Crtdq;Hp 305A;2.6K;Cyn
Printer Supplies-Laser;Oem |
| Freq: 101 Conditional Probability: 95.3% Strength: Very Strong | |
| Crtdq;Hp 305A;2.6K;Yw
Printer Supplies-Laser;Oem | Crtdq;Hp 305A;2.6K;Cyn
Printer Supplies-Laser;Oem |
| Freq: 105 Conditional Probability: 94.6% Strength: Very Strong | |

Categories Video

Category Details Video

Clusters Video

Dashboard Video

Data Overview Video

Loading Data with FTP Video

Navigation Video

Predictors Video

Long Data Format Video

Substitutions Video

Wide Data Format Video

Sample Data Sets

[TX Truck Data - Improving Parts Management](#)

[FDIC Bank Data - Predicting Return on Assets](#)

[Retail Sales Data - Inventory Management](#)

[Airline Data - Predicting Flight Delays](#)

[Titanic Data - Demographics Predictions](#)

[Network Data - Correlating Attacks](#)

Reports in Patterns via cURL

Basic Information

Collect Details

You will need the following information from you system:

base url for patterns (like "<https://patterns.example.com>"), referred to as \$BASE_URL from here on out
auth_token for the user who wants to run the report, referred to as \$AUTH_TOKEN from here on out
the company ID for the company to use in patterns, referred to as \$COMPANY_ID from here on out
dependent / outcome category if using for predictions, referred to as \$OUTCOME from here on out
project_id (not name) to put report in (unless the default project is okay), referred to as \$PROJECT_ID from here on out

cURL POST Base Command

In this command, \$DATA and \$URL will be replaced with the post data and actual url.

```
curl --insecure -qSfs -w "\n%{http_code}" -d "$DATA" -D - "$URL"
```

cURL GET Base Command

In this command, \$URL is intended to be replaced with the actual url.

```
curl --insecure -qSfsL -w "\n%{http_code}" "$URL?auth_token=$AUTH_TOKEN"
```

Submit Report

```
curl --insecure -qSfs -w "\n%{http_code}" -d
"auth_token=$AUTH_TOKEN&report[original_file_name]=$INPUT_FILE&report[name]=$REPORT_NAME&outc
ome=$OUTCOME" "http://$BASE_URL/admin/companies/$COMPANY_ID/reports?auth_token=$AUTH_TOKEN" -
H 'Accept: application/json'
```

This command will return a JSON string in the following format:

```
{
  "id":64723871,
  "original_file_name":"banded_titanic_passengers.wide.csv",
  ...
  other stuff
}
```

This JSON string needs to be parsed for the report id which is the first element. This will be needed when getting the report status below and will be referred to from now on as \$REPORT_ID.

Report Status

```
curl --insecure -qSfsL -w "\n%{http_code}"
"http://$BASE_URL/admin/companies/$COMPANY_ID/reports/$REPORT_ID/log.json?auth_token=$AUTH_TO
KEN"
```

This will return a JSON string in the following format:

```
{
  "report":{
    "id": 64723871,
    "state":"report-ready",
    ...
    other stuff
}
```

This string can be parsed for the report state which is the second item under the first element report.

Submit Prediction

```
curl --insecure -qSfs -w "\n%{http_code}" -d
"auth_token=$AUTH_TOKEN&shard[params][filename]=$INPUT_FILE&shard[params][rules_report_id]=$R
EPORT_ID&shard[format]=predict&shard[name]=$PREDICTION_REPORT_NAME"
"http://$BASE_URL/admin/companies/$COMPANY_ID/shards.json?auth_token=$AUTH_TOKEN" -H
'Accept: application/json'
```

This will return a JSON string in the following format:

```
{
  "id":1,
  "client_id":1,
  "creator_id":1,
  "name":"titancic_survival_prediction",
  ...
  other stuff
}
```

The shard id is the first item. This will be needed to get the prediction status below and will be referred to from now on as \$SHARD_ID.

Prediction Status

```
curl --insecure -qSfsL -w "\n%{http_code}"
"http://$BASE_URL/admin/companies/$COMPANY_ID/shards/$SHARD_ID.json?auth_token=$AUTH_TOKEN"
```

This will return a JSON string in the following format:

```
{
  "shard":
  {
    "id":1,
    "client_id":1,
    "state":"ready",
    "waiting_after":0,
    "name":"titancic_survival_prediction",
    ...
    more stuff
  }
}
```

The prediction state is the 3rd sub element of the the first item shard.

Download Prediction Reports

```
curl --insecure -qSfsL -w "\n%{http_code}"  
"http://$BASE_URL/predict/shards/$SHARD_ID/download/$PREDICITON_REPORT?auth_token=$AUTH_TOKEN"  
" >> $PREDICITON_REPORT_NAME
```

where the URLs for the various reports are:

```
http://$BASE_URL/predict/shards/$SHARD_ID/download/prediction_overview.csv.gz  
http://$BASE_URL/predict/shards/$SHARD_ID/download/predictions.csv.gz  
http://$BASE_URL/predict/shards/$SHARD_ID/download/reasons.csv.gz  
http://$BASE_URL/predict/shards/$SHARD_ID/download/rules.csv.gz
```

The Wide Format

This article covers how to format and save your data in the wide format before uploading to Emcien.

The Wide Format

The wide format is the most commonly used and is easy to start with. The wide format consists mostly of user-defined columns with no required column headings.

Formatting

The wide format is the more universal and most commonly used Emcien format. You can use the wide format for multi-dimensional data, such as demographics or configurable products. In the wide format, each transaction is identified by a single row of data.

An example of the wide format is displayed below.

Wide Format Example

User Defined Data										
date	volume	ID	Age Group	Race	Birth State	Residence State	Gender	Education	Visits per month	...
2/3/13	1	C-1002010	25-35yrs	Asian	Birth-CA	Residence-FL	Male	High School	3-5 visits	
2/3/13	1	C-1002011	25-35yrs	Caucasian	Birth-OR	Residence-OR	Female	Graduate school	Less than 3	...
2/3/13	1	C-1005401	55-65yrs	Latin	Birth-GA	Residence-FL	Female	University	Over 10 visits	
2/3/13	1	C-1005660	55-65yrs	Caucasian	Birth-NY	Residence-CT	Female	University	Less than 3	

Volume is not a required column but must be in the first column if it exists and there is no date. If the date exists, volume must be in the second column.

Date is not required but must be in first column

Each row represents a Transaction

This example contains four transactions, each represented by a single row of data. In this data, each transaction represents a client and their associated demographics data.

Unlike the receipt format, the wide format has no required columns and supports user-defined columns.

Wide Format Details

Headers are required for each column in the wide format. Your file must contain at least two user-defined columns with a maximum of 1,000 total columns.

While the data is allowed to contain any UTF-8 characters the header must be in lower ASCII.

Most columns for wide format data are user-defined. Emcien also offers optional columns, which are listed in the table below. You can use these optional columns to use certain Emcien features, such as date and time trends. If used, your optional column header must *exactly* match the column header listed in the table below.

Column Header	Required/Optional	Format	Description
---------------	-------------------	--------	-------------

date	Optional	Date/ Time	<p>The date and time when the transaction occurred. Field values must be formatted using ISO or Unix standards.</p> <p>Field values can contain date only, or date and time. The column heading indicates the date and time format used.</p> <p>Important: If used, this column must be first.</p>
<i>OR</i>			
transaction_date			
<i>OR</i>			
time			
<i>OR</i>			
transaction_time			
<i>OR</i>			
transaction_time_unix			

volume	Optional	Integer	<p>The volume applied to all attributes of the transaction. This column is used to calculate the strength of connections in your data.</p> <p>Field values must be non-negative values consisting only of digits.</p> <p>Important: If you use this column <i>and not</i> the transaction_date, transaction_time, or transaction_time_unix columns, this column must be first.</p> <p>If you use this column <i>and</i> the transaction_date, transaction_time, or transaction_time_unix column, this column must be second.</p>
--------	----------	---------	--

<user-defined>	Optional	String	<p>This column is defined by you. Each user-defined column must use a unique column header.</p> <p>Important: Your data file must contain at least two user-defined columns.</p> <p>Field values must be wrapped in double quotes (") and cannot be longer than 255 characters.</p> <p>All double quotes inside any string should be escaped with another double quote.</p> <p>For example:</p> <p>"String 3" Roll" should be changed to "String 3"" Roll"</p>
----------------	----------	--------	--

Not sure which format is best for you? The Emcien team can help prepare your data. Contact us at support@emcien.com.

The Long Format

The long format is typically used for transaction data containing many variables and combinations. In the long format, categories and their corresponding items are represented vertically and grouped based on a unique transaction ID.

An example of the long format is displayed below.

Long Format Sample

transaction_id	item_id	transaction_date	item_name	item_category	transaction_price	item_volume	item_price	transaction_volume
101001	395847	12/15/06	4 in SQR COVER	CONDUIT	\$27.39	1	\$16.92	1
101001	457634	12/15/06	8IN NATURAL CABLE TIE	TOOLS & ACCESSORIES	\$27.39	1	\$10.47	1
101002	272840	12/31/06	20A IVORY GFCI	WIRING DEVICES	\$127.67	1	\$15.69	1
101002	668138	12/31/06	4" in SQUARE COVER GFC	CONDUIT	\$127.67	3	\$5.23	1
101002	923352	12/31/06	DRAIN TAPE-ENCASED SINK	PLUMBING	\$127.67	1	\$96.29	1

In the above example, two transactions are shown. Each transaction is identified using a unique `transaction_id`: 101001 and 101002. Transaction 101001 is associated with the purchase of 2 items. Transaction 101002 is associated with the purchase of 3 items.

The long format *does not* support user-defined columns.

Using the Long Format for Non-Transaction Data

In addition to transaction data, you can use the long format for any data containing unique identifiers and repeating vertical rows.

An example of network data is displayed below.

transaction_id	item_id	transaction_date	item_category
1000001	Mon	8/1/11	DOW
1000001	Start_Hour:9	8/1/11	Hour
1000001	SIG_ID:4	8/1/11	SIG_id
1000001	TCP_Timestamp_is_outside_of_PA	8/1/11	SIG_name
1000001	UNKNOWN	8/1/11	Classification
1000001	src-10.1.60.203	8/1/11	src_ip
1000001	src-57764	8/1/11	src_port
1000001	dst-7.204.241.161	8/1/11	dest_ip
1000001	dst-25	8/1/11	dest_port
1000002	Tue	11/1/11	DOW
1000002	Start_Hour:9	11/1/11	Hour
1000002	SIG_ID:4	11/1/11	SIG_id
1000002	TCP_Timestamp_is_outside_of_PA	11/1/11	SIG_name
1000002	UNKNOWN	11/1/11	Classification
1000002	src-7.204.241.161	11/1/11	src_ip
1000002	src-25	11/1/11	src_port
1000002	dst-10.1.60.203	11/1/11	dest_ip
1000002	dst-50176	11/1/11	dest_port

In the above example, each network event is tracked by a unique identifier. This identifier is used to associate the event with different information in the item_id column.

Long Format Details

Headers are required for each column in the long format. The minimum columns required are transaction_id and item_id. The acceptable columns are transaction_id, transaction_time, item_id, item_name, item_category, transaction_volume, transaction_price, item_volume, item_price.

The first column must be transaction_id.

While the data is allowed to contain any UTF-8 characters the header must be in lower ASCII.

The long format *does not* support user-defined columns. The required and optional columns are described in the table below.

Column Header	Required/Optional	Format	Description
---------------	-------------------	--------	-------------

transaction_id	Required	String	<p>A unique identifier for the transaction. This identifier can appear in multiple rows but should be different for each transaction. NOTE: Transactions must be contiguous. If you use the same transaction_id in two places in the file, then it will be interpreted as two transactions with the same ID.</p> <p>Rows sharing the same transaction_id must be adjacent.</p> <p>Field values must be wrapped in double quotes (") and cannot be longer than 255 characters.</p> <p>All double quotes inside any string should be escaped with another double quote.</p> <p>For example: "String 3" Roll" should be changed to "String 3"" Roll"</p>
----------------	----------	--------	---

item_id	Required	String	<p>A unique identifier for the item. Typically, this column is used for SKU numbers and product numbers.</p> <p>Field values must be wrapped in double quotes (") and cannot be longer than 255 characters.</p> <p>All double quotes inside any string should be escaped with another double quote.</p> <p>For example: "String 3" Roll" should be changed to "String 3"" Roll"</p>
transaction_date OR transaction_time OR transaction_time_unix	Optional	Date/ Time	<p>The date and time when the transaction occurred. Field values can also be formatted using ISO or AUS standards. (See "Date and Time" section below)</p> <p>Field values can contain date only, or date and time. The column heading indicates the date and time format used.</p>

item_category	Optional	String	<p>A text identifier used to categorize the item. Typically, this column is used for product category.</p> <p>Field values must be wrapped in double quotes (") and cannot be longer than 255 characters.</p> <p>All double quotes inside any string should be escaped with another double quote.</p> <p>For example:</p> <p>"String 3" Roll" should be changed to "String 3"" Roll"</p>
---------------	----------	--------	--

item_name	Optional	String	<p>A descriptive label of the item used in the item_id column.</p> <p>For example, an item_id of RTJ00345 has the associated item_name of Green Striped Lawn Chair.</p> <p>Field values must be wrapped in double quotes (") and cannot be longer than 255 characters.</p> <p>All double quotes inside any string should be escaped with another double quote.</p> <p>For example: "String 3" Roll" should be changed to "String 3"" Roll"</p>
-----------	----------	--------	--

item_volume	Optional	Float	<p>The volume of item associated with the transaction. Typically, this volume is the quantity of a specific item purchased. Field values must be in the standard float format: a string of digits with a decimal point optionally among them. Precision beyond 4 decimal places is not used.</p> <p>If the field value is NULL or blank, a default value of 1 is used. Negative values and 0's are not allowed.</p>
-------------	----------	-------	---

item_price	Optional	Price	<p>The individual price of the item associated with the transaction. Typically, this price is the manufacturer's suggested retail price (MSRP) or other undiscounted price.</p> <p>Field values must be in dollars, and may begin with an optional \$ sign.</p> <p>NULL or blank values are supported. Negative prices <i>are not</i> supported.</p>
------------	----------	-------	--

transaction_volume	Optional	Integer	<p>The total volume of the entire transaction.</p> <p>Note: This column should not be confused with the item_volume column, which is used for the volume associated with specific items.</p> <p>Field values must be non-negative values consisting only of digits.</p> <p>If the field value is NULL or blank, a default value of 1 is used.</p>
--------------------	----------	---------	---

transaction_price	Optional	Price	<p>The total price of the entire transaction.</p> <p>Note: This column should not be confused with the item_price column, which is used for the price associated with specific items.</p> <p>Field values must be in dollars, and may begin with an optional \$ sign.</p> <p>NULL or blank values are supported. Negative prices <i>are not</i> supported.</p>
-------------------	----------	-------	--

Date and Time

When using the transaction_date column, dates should be recorded using in the MM-DD-YYYY or MM/DD/YYYY format. You can zero pad month and day values.

To use the ISO data format (YYYY-MM-DD or YYYY/MM/DD), you can add _ISO or -ISO to the end of the transaction_date column header.

To use the AUS data format (D-M-YY or D/M/YY), you can add _AUS or -AUS to the end of the transaction_date column header.

When using the transaction_time, or transaction_time_unix column, dates should be recorded using the [ISO-8601](#) or [UNIX time](#) formats.

The [ISO-8601](#) format is YYYY-MM-DDTHH:MM:SSZ
T and the Z are literal characters and can be upper or lower case.

The [UNIX time](#) format is the number of seconds since 00:00:00 Coordinated Universal Time (UTC), Thursday, 1 January 1970.

Naming Your Data File

Emcien uses a file naming convention to identify the Emcien format used. The filename is limited to 200 characters or less. Files should be named using the following structure:

<filename>.<file type>.<extension>

filename	Filenames should use only the following characters: <ul style="list-style-type: none">• ASCII alphanumeric characters in upper (A-Z), lower (a-z) case, 0-9• White space• Periods .• Underscores _• Hyphens -• Parentheses () The following characters <i>are not</i> supported: <ul style="list-style-type: none">• Backslash (\)• Single quote (')• Double quotes (")
file type	long

extension

.csv or .csv.gz

File Name Examples

area2-sales.long.csv

Uncompressed sales data from area 2 in the long format

clinical.all.long.csv.gz

Compressed clinical test data in the long format

Not sure which format is best for you? The Emcien team can help prepare your data. Contact us at support@emcien.com.

The Tagged Format

This article covers how to format and save your data in the tagged format before uploading to Emcien.

The Tagged Format

The tagged format contains no headers and is great for unstructured data. The tagged format is most commonly used for analyses containing free-form text.

Formatting

The tagged format consists of two required columns (which may be empty) and no headers. There may be a maximum of 1000 categories per data file, with a maximum of 1000 items per line. An example of the tagged format is below:

Tagged Format Example		User Defined Data			
2012-07-15	transaction1	Model::GT	Color::White Pearl	Automatic	Dark Grey
2012-07-16		Model::Base	"Color::Pearl, Black"	Manual-5 Speed	Brushed Aluminum
2012-07-16	transaction5	Model::Base		Manual-6 Speed	Brushed Aluminum
2012-07-17	transaction6	Model::Base	Color::Gray		Dark Grey

↑ The first column must be the date, but it may be empty

↑ The second column must be the transaction ID, but it may be empty

↑ Items may be represented as Category::Item, or just as the item

Some important things to note about the Tagged Format:

The first column is the date, or date and time. Although the column is required, it is allowed to be empty. Details on the date format are below.

The second column is the transaction ID. As with the date, this column is

required, but may be blank as well. The transaction id can be any unique identifier to a transaction.

An item may be represented as "Category::item" or as "item".

Any cells containing commas must be surrounded by double quotes

Date Format Details

The date column can have 4 different formats, but the format must be consistent throughout the data file. Below are possible examples of what they may be.

Format	Example
YYYY-MM-DD	2012-07-15
YYYY-MM-DDTHH:MM:SS	2012-07-15T02:23:44 (T, t or space must be present between date and time; Z, z, or nothing after time)
unixtime	1431209618 (assumed time zone is UTC+0; 1-10 digit number)
none	No value is required

Not sure which format is best for you? The Emcien team can help prepare your data. Contact us at support@emcien.com.

Product Recommendations

This article outlines how Emcien can drive a product recommendation engine automatically with historical sales data.

Align your data with a use case

Note: This procedure uses the Retail Sales example data set. Your data should be similar to the example data, particularly the structure of a `transaction_id` or something similar to connect the items in each purchase.

[retail_sales_2014_12_02.receipt.csv.zip](#)

If your data doesn't follow this same structure, please consult the [Receipt Format](#) page before contacting Emcien Support.

Data Prep

This data set contains retail transactions in the receipt format. While more difficult to manipulate, the receipt format is particularly helpful for sales data because it allows you to attach multiple items from the same category to each transaction through the `transaction_id`. This feature makes the receipt format ideal for basket analysis.

Receipt Format Example

		1 st Transaction			Optional Data				
transaction_id	item_id	transaction_date	item_category	item_name	item_volume	item_price	transaction_price	transaction_volume	
10001	ABC123	7/4/12	Hardware	"3in, Bolt Steel"	1	\$3.99	\$25.92		1
10001	XYZ456	7/4/12	Cleaner	ACME Bleach	4	\$0.99	\$25.92		1
10001	MNO780	7/4/12	Fasteners	"1 in, Gate Latch"	3	\$5.99	\$25.92		1
10002	QRS123	7/5/12	Live Goods	Lg Potted Fern	2	\$9.99	\$31.97		1
10002	HJK234	7/5/12	Soils	ACME Soil	1	\$11.99	\$31.97		1

Note:
This example shows two transactions, the first with three items and the second with two items.

The sample data set contains five columns: the `transaction_id`, `item_id`, `transaction_date`, `item_name`, and `item_category`. Only the `transaction_id` and `item_id` are required fields, but the added `transaction_date`, `item_name`, and `item_category` give added information to be included in the analysis.

Note that the receipt format does not permit user-defined column headings. Some retail data sets include additional columns that relate more to the product than to the purchase, like product margin. These columns can be moved or rolled into the receipt format using the `transaction_id`.

If your data looks similar to the Receipt Data format, you may only have to rename the column headers or change the column order. Other data will require more work. If you need assistance with the format, first refer to <http://emcien.com/help/article/link/preparing-your-data>

Uploading Data

Upload your data for analysis.

This article covers how to upload data files to Emcien using [File Transfer Protocol \(FTP\)](#).

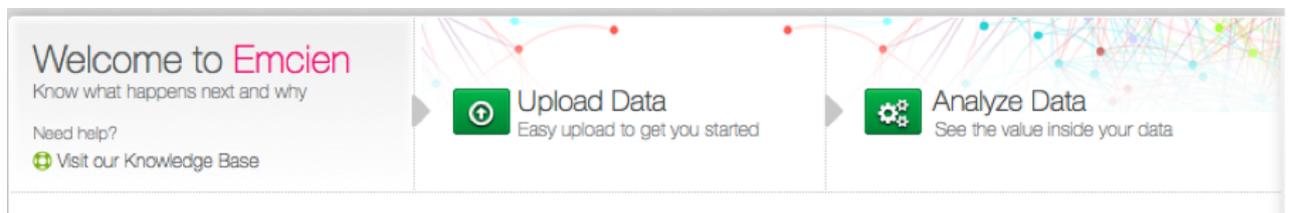
Make sure you've prepared your data before uploading to Emcien. Check out the [Preparing Your Data](#) article for more information. You

can also email us at support@emcien.com for help getting started.

Upload using Drag and Drop

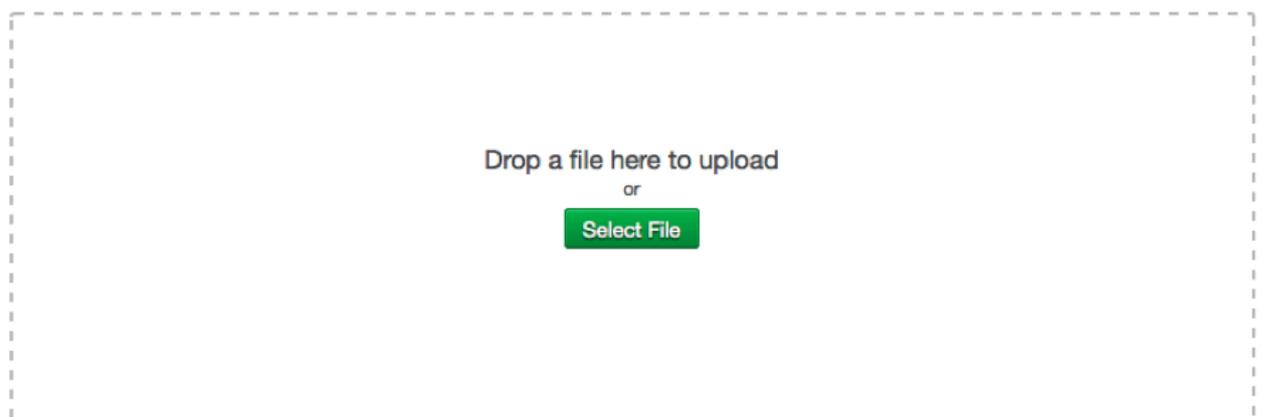
In Emcien version 2.14 you can now drag and drop files into the application for processing. If you are using a previous version you can use one of the other methods listed below.

Simply open the Emcien application and select "Upload Data"



Select the file that you wish to process, then drag and drop it onto the Upload Data screen

Upload Data Services



Once it has finished uploading you can then click on "Analyze This Data" to

process the file.

✔ Upload Complete!

[Analyze This Data](#) or [Upload More Data](#)

Uploading Data to the Emcien Cloud

You can upload data files to Emcien using [SSH File Transfer Protocol \(SFTP\)](#) and your preferred FTP client, such as [FileZilla](#) or [Cyberduck](#).

To connect to Emcien using your FTP client, use the following credentials:

Host: feeds.emcien.com

Username: {Your Emcien Feeds server Username}

Password: {Your Emcien Feeds server Password}

Uploading Data on Mac or Linux

Launch your preferred FTP client. If you do not have a preferred FTP client, we recommend [Cyberduck](#).

For this article, we use screenshots of Cyberduck.

Click Open Connection.



Enter the below credentials and click Connect.

The screenshot shows an FTP client window with the following elements:

- Top toolbar: Open Connection, Quick Connect, Action, Refresh, Edit, Disconnect.
- Address bar: /srv/clients/Emcien_Marketing
- Search bar: Search
- File list table:

Filename	Size	Modified
banded_companies.wide.csv	10.5 MB	10/24/14, 5:29 PM
banded_offenders_table.wide.csv	2.6 MB	10/29/14, 2:03 PM
banded_titanic-table.wide.csv	147.3 KB	12/9/14, 9:21 AM
banded_titanic.wide.csv	122.3 KB	11/7/14, 3:56 PM
banded_WorldPoverty.csv	1.2 MB	10/22/14, 2:10 PM

Your FTP client will then display the files on your server.

Server: Your Emcien URL. This URL was set by your IT resource during installation.

Username

Password

Drag and drop your data file to the server files in your FTP application.

On the Emcien home page click Analyze Data. You'll see the data files on this screen.

Uploading Data on Windows

Click the Start button . The Start menu is displayed.

On the Start menu, click Computer. The Computer folder is displayed.

Right-click anywhere in the folder and click Add a network location.

The Add Network Connection wizard is displayed. Click Next.

On the next tab, select Choose a custom network location. Click Next.

On the next tab, enter your Emcien URL.

This URL was set by your IT resource during installation.
Ensure you include *FTP://* at the beginning of the URL.

Then click Next.

Clear the Log on anonymously checkbox.

Type *data* in the username field. Click Next.

On the next tab, name your shortcut *Emcien* by typing it in the Type a name for this network location field. Click Next.

On the next tab, click Finish.

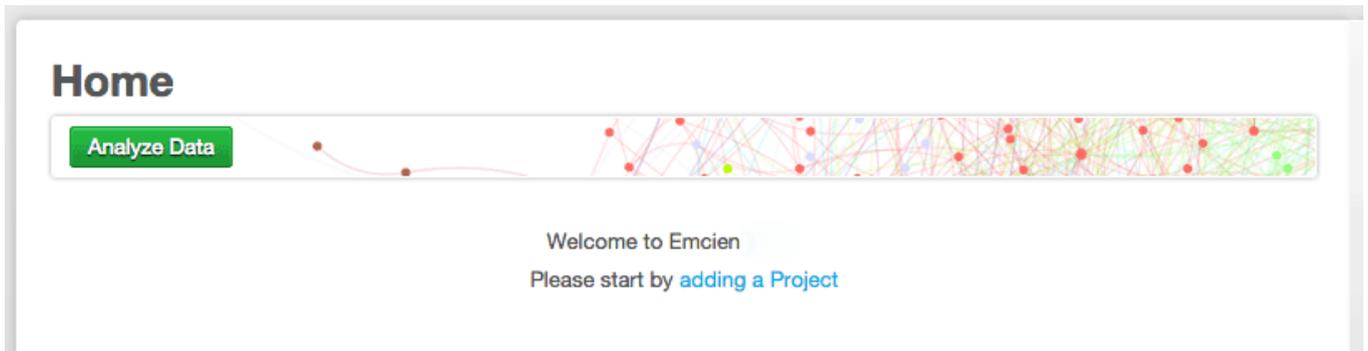
Type *feeds1* in the password field. Check the Save password checkbox so you can connect directly in the future.

Click Log On. You can now drag data files into this folder.

On the Emcien home page, click Analyze Data. You'll see the data files on this screen.

Analysis

On the Home page click Analyze Data.



The Analyze Data button will bring up all of your uploaded data. Select your data set.

New Analysis Demo Admin

Source Data Filter by Name

Name	File Size	Modified Time
Airlines-Delay-Data.wide.csv	33 MB	Jun 4 10:17AM
✓ retail_sales_2014_12_02_receipt.csv	86 MB	Dec 2 4:29PM
ACME-Auto.csv	170 MB	Jun 2 4:38PM
ACMEOfficeSupplyInputFile.csv.gz	390 MB	Jun 2 2:27PM
BankData3Q2013-noemp.wide.csv	1.4 MB	Jun 2 2:00PM
ACMEAutoParts_initial_NoSubCategory.csv	190 MB	Jun 2 1:59PM
ACMEHardwareMediumCleanedupPricing.csv.gz	220 MB	Jun 2 1:58PM
ACME-Office-1M-Load.csv	370 MB	May 28 3:55PM
Consumer_Complaints.wide.csv	27 MB	Mar 31 9:30AM
ACME-Hardware.csv	2.6 GB	Oct 23 1:00PM

Select a Project Folder, and if necessary rename the Report.

Projects

Reports in a project will display on home page

Demo

Predictions

Override Defaults

Report Name

retail sales 2014 12 02.receipt

Once your project is organized, click Start Analysis.

retail sales 2014 12 02.receipt Demo Admin



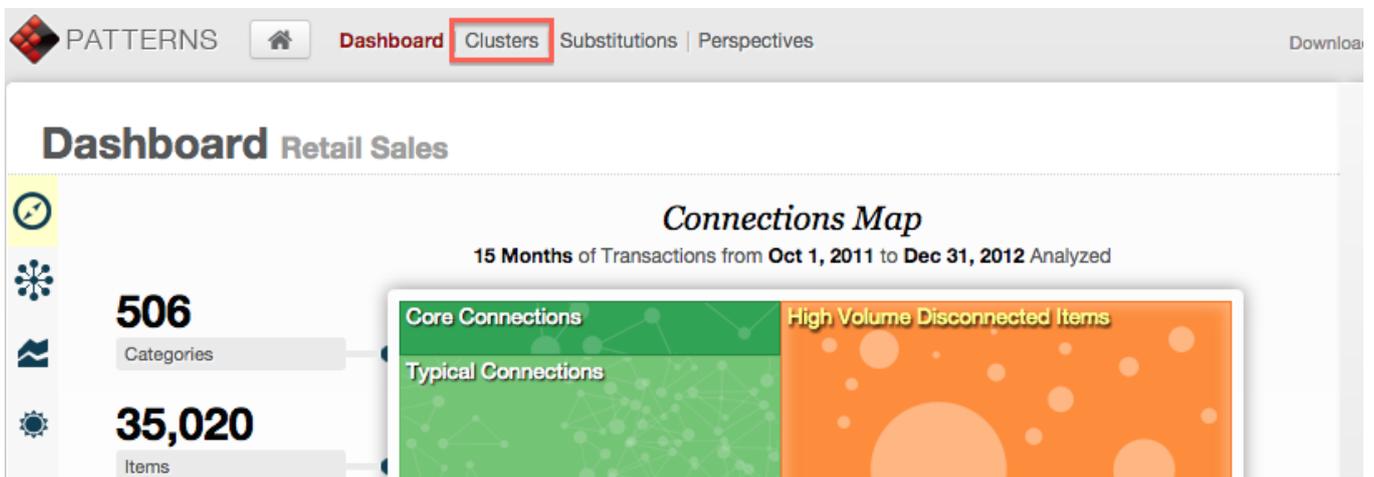
The load screen will take you through each stage of the analysis.

When the analysis is complete you will hear a chime and the View Analysis button will be highlighted in green. Just click the button to see your results.

Results

The Dashboard page for the Retail Sales data set is displayed. On the Dashboard page you can view the number of connected data values, the number of transactions analyzed, and more.

1. On the Dashboard page click Clusters.



The Clusters page is displayed showing an overview of the connections in the data set.

2. On the Clusters page click A-Z to sort the clusters alphabetically.

PATTERNS Dashboard | Clusters | Substitutions | Perspectives Download

Clusters Retail Sales

Download CSV

397 Categories **By Strength** **A-Z**

- Board-Chalk 100%
- Board-Cubicle 100%
- Boards-Accessories 100%
- Labelmaker Tapes (Manual) 100%
- Typewriters 100%

Board-Chalk

1 of the 1 items are in clusters across 1 categories.
The number of items with connections make this category's connections **Extremely Strong**.

Board-Chalk ↔ Board-Cleaners And Acces... 1 Clusters

3. Click the Adhesives category.

PATTERNS Dashboard | Clusters | Substitutions | Perspectives Download

Clusters Retail Sales

Download CSV

397 Categories **Strength** | **By A - Z**

- Address Books 29%
- Adhesive Bandages & Wraps 50%
- Adhesives 49%
- Advanced Ink System 64%
- Air Filters & Humidifiers 15%
- Art Board & Paper 56%

Board-Chalk

1 of the 1 items are in clusters across 1 categories.
The number of items with connections make this category's connections **Extremely Strong**.

Board-Chalk ↔ Board-Cleaners And Acces... 1 Clusters

The categories connected to the Adhesives category is displayed. In this data set, adhesives are connected to other adhesives, tape, and pens.

PATTERNS Dashboard | Clusters | Substitutions | Perspectives Down

Clusters Retail Sales Strength | By A - Z Download CSV

397 Categories

- Address Books 29%
- Adhesive Bandages & Wraps 50%
- Adhesives 49%**
- Advanced Ink System 64%
- Air Filters & Humidifiers 15%
- Art Board & Paper 56%
- Art&Drafting-Layout&Lettering 39%
- Art&Drafting-Tools & Temp 38%
- Arts & Crafts 34%
- Audio Visual-Headsets;Stereo 30%
- Backrests/Footrests 31%
- Badges & Holders 56%
- Bank Supplies 45%
- Batteries 65%

Adhesives
 134 of the 275 items are in clusters across 94 categories.
 The number of items with connections make this category's connections **Strong**.

Adhesives ←

- Adhesives 81 Clusters
- Art Board & Paper 45 Clusters
- Colored Pencils & Markers 34 Clusters
- Tape & Dispensers 28 Clusters
- Pens-Retractable & Ballpoint 26 Clusters
- Janitorial-Paper Products 43 Clusters
- Clips;Clamps & Fasteners 13 Clusters

You can also see that in the Adhesives category glue guns and the sticks for the glue gun are connected.

- On the Clusters page, click on Colored Pencils & Markers.

Clusters Retail Sales Download CSV

397 Categories Strength | By A - Z

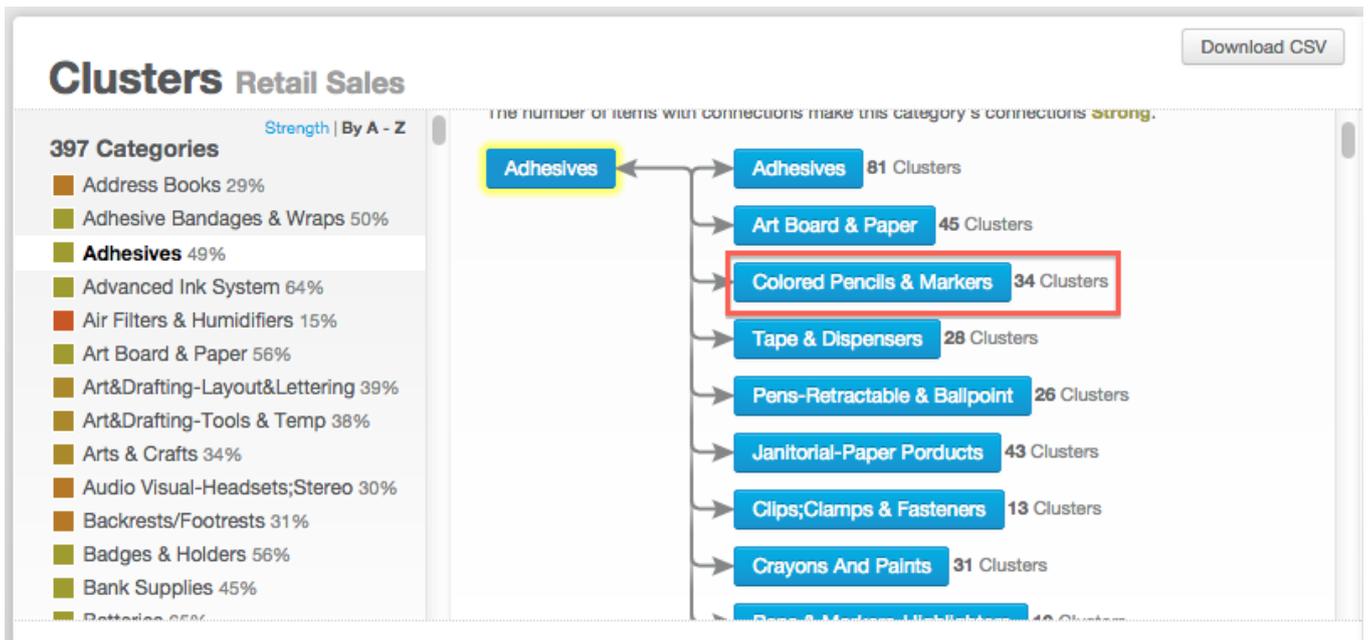
- Address Books 29%
- Adhesive Bandages & Wraps 50%
- **Adhesives 49%**
- Advanced Ink System 64%
- Air Filters & Humidifiers 15%
- Art Board & Paper 56%
- Art&Drafting-Layout&Lettering 39%
- Art&Drafting-Tools & Temp 38%
- Arts & Crafts 34%
- Audio Visual-Headsets;Stereo 30%
- Backrests/Footrests 31%
- Badges & Holders 56%
- Bank Supplies 45%
- Batteries 55%

Adhesives
 134 of the 275 items are in clusters across 94 categories.
 The number of items with connections make this category's connections **Strong**.

Adhesives is connected with items from Any Category and has **543** clusters.

#	Items	Strength	Freq	Cnd Prob	Trend
1	<div style="border: 1px solid #ccc; padding: 2px; display: inline-block;">GUN;GLUE;SFS DUAL MELT Adhesives</div> → <div style="border: 1px solid #ccc; padding: 2px; display: inline-block;">GLUE;STICK;DUAL MELT Adhesives</div>	S	35	74.5%	
2	<div style="border: 1px solid #ccc; padding: 2px; display: inline-block;">REFILL;ADHESIVE;REMOVBLE Adhesives</div> → <div style="border: 1px solid #ccc; padding: 2px; display: inline-block;">ADHESIVE;REMOVBLE;REFI... Adhesives</div>	S	12	42.9%	

Items in the Colored Pencils & Markers category connected to items in the Adhesives category are displayed. You can use the connections on this page to find point-to-point connections.



For example, transactions containing themebook purchases also frequently contain crayon, marker, filler paper, and glue purchases. You can use this information for cross-selling or to put together product bundles.

Adhesives is connected with items from **Colored Pencils & Markers** and has **34** clusters.

#	Items	Strength	Freq	Cnd Prob	Trend
1	ADHESIVE;GLUE-ALL; 4OZ Adhesives → MARKER;BROADLINE;8CT Colored Pencils & Markers	M	6	5.5%	
2	MARKER;BROADLINE;WS... Colored Pencils & Markers → GLUE;ELMERS;4 OZ;WE Adhesives	M	9	9.6%	
3	GLUE;ELMERS;4 OZ;WE Adhesives → PENCIL;COLORED;12/SET Colored Pencils & Markers	M	11	7.0%	
4	GLUE;SCHOOL;7-5/8 OZ;WE Adhesives → MARKER;BROADLINE;WS... Colored Pencils & Markers	SW	4	8.7%	
5	GLUE;SCHOOL;7-5/8 OZ;WE Adhesives → MARKER;BROADLINE;WS... Colored Pencils & Markers CRAYON;CRAYOLA;3-5/8"... Crayons And Paints	SW	3	6.5%	
6	MARKER;BROADLINE;WS... Colored Pencils & Markers → GLUE;ELMERS;4 OZ;WE Adhesives ERASER;PINK PEARL;MED Erasers	SW	5	5.3%	
7	MARKER;BROADLINE;WS... Colored Pencils & Markers → ADHESIVE;STK;ALLPURP;... Adhesives	SW	9	9.6%	
8	PENCIL;COLORED;12/SET Colored Pencils & Markers → GLUE STICK;PERM;0.26OZ Adhesives	SW	15	6.2%	
9	THEME BK WR 1-SUB 92 ... Notebooks → CRAYON;CRAYOLA;3-5/8"... Crayons And Paints MARKER;BROADLINE;WS... Colored Pencils & Markers PAPER;FILLER;10.5X8;200CT Paper-Notebook Filler GLUE;ELMERS;4 OZ;WE Adhesives	SW	4	21.1%	
10	MARKER;BROADLINE;WS... Colored Pencils & Markers → GLUE;ELMERS;4 OZ;WE Adhesives PENCIL;COLORED;12/SET Colored Pencils & Markers	SW	5	5.3%	

5. Click the GLUE;ELMERS;4 OZ;WE item.

Adhesives is connected with items from **Colored Pencils & Markers** and has **34** clusters.

#	Items	Strength	Freq	Cnd Prob	Trend
1	ADHESIVE;GLUE-ALL; 4OZ Adhesives → MARKER;BROADLINE;8CT Colored Pencils & Markers	M	6	5.5%	
2	MARKER;BROADLINE;WS... Colored Pencils & Markers → GLUE;ELMERS;4 OZ;WE Adhesives	M	9	9.6%	
3	GLUE;ELMERS;4 OZ;WE Adhesives → PENCIL;COLORED;12/SET Colored Pencils & Markers	M	11	7.0%	
4	GLUE;SCHOOL;7-5/8 OZ;WE Adhesives → MARKER;BROADLINE;WS... Colored Pencils & Markers	SW	4	8.7%	
5	GLUE;SCHOOL;7-5/8 OZ;WE Adhesives → MARKER;BROADLINE;WS... Colored Pencils & Markers CRAYON;CRAYOLA;3-5/8"... Crayons And Paints	SW	3	6.5%	
6	MARKER;BROADLINE;WS... Colored Pencils & Markers → GLUE;ELMERS;4 OZ;WE Adhesives ERASER;PINK PEARL;MED Erasers	SW	5	5.3%	
7	MARKER;BROADLINE;WS... Colored Pencils & Markers → ADHESIVE;STK;ALLPURP;... Adhesives	SW	9	9.6%	
8	PENCIL;COLORED;12/SET Colored Pencils & Markers → GLUE STICK;PERM;0.26OZ Adhesives	SW	15	6.2%	
9	THEME BK WR 1-SUB 92 ... Notebooks → CRAYON;CRAYOLA;3-5/8"... Crayons And Paints MARKER;BROADLINE;WS... Colored Pencils & Markers PAPER;FILLER;10.5X8;200CT Paper,Notebook Filler GLUE;ELMERS;4 OZ;WE Adhesives	SW	4	21.1%	
10	MARKER;BROADLINE;WS... Colored Pencils & Markers → GLUE;ELMERS;4 OZ;WE Adhesives PENCIL;COLORED;12/SET Colored Pencils & Markers	SW	5	5.3%	

The Item Detail page for the GLUE;ELMERS;4 OZ;WE item is displayed. On the Item Detail page you can see transaction trends for the GLUE;ELMERS;4 OZ;WE item over time.



Item Detail

GLUE;ELMERS;4 OZ;WE
Adhesives



GLUE;ELMERS;4 OZ;WE Connections:

Download CSV

#	Items	Strength	Freq	Cnd Prob	Trend
1	MARKER;BROADLINE;WSHBLE;8CT Colored Pencils & Markers	M	9	5.7%	
2	ERASER;PINK PEARL;MED Erasers	M	8	5.1%	
3	PENCIL;COLORED;12/SET Colored Pencils & Markers	M	11	7.0%	
4	PAPER;FILLER;10.5X8;200CT Paper-Notebook Filler	M	6	3.8%	
5	CRAYON;CRAYOLA;3-5/8"-24CT Crayons And Paints	M	6	3.8%	

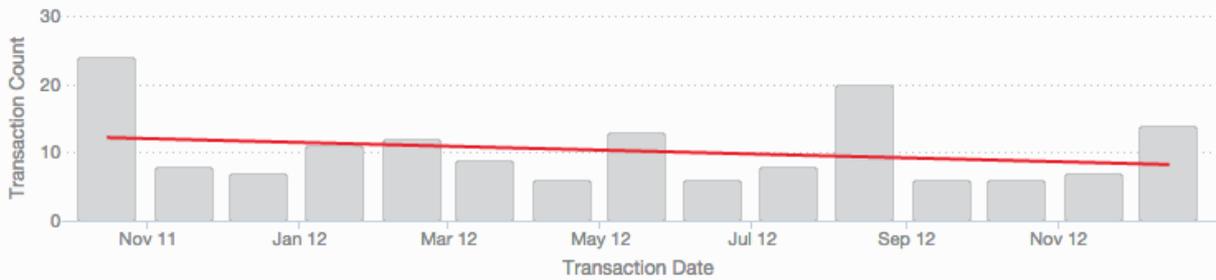
You can also view items frequently connected with the GLUE;ELMERS;4 OZ;WE item.

Item Detail

GLUE;ELMERS;4 OZ;WE
Adhesives

Transactions for this item are trending down at less than 1 per month

Transactions from **Oct 1, 2011** to **Dec 31, 2012** Analyzed



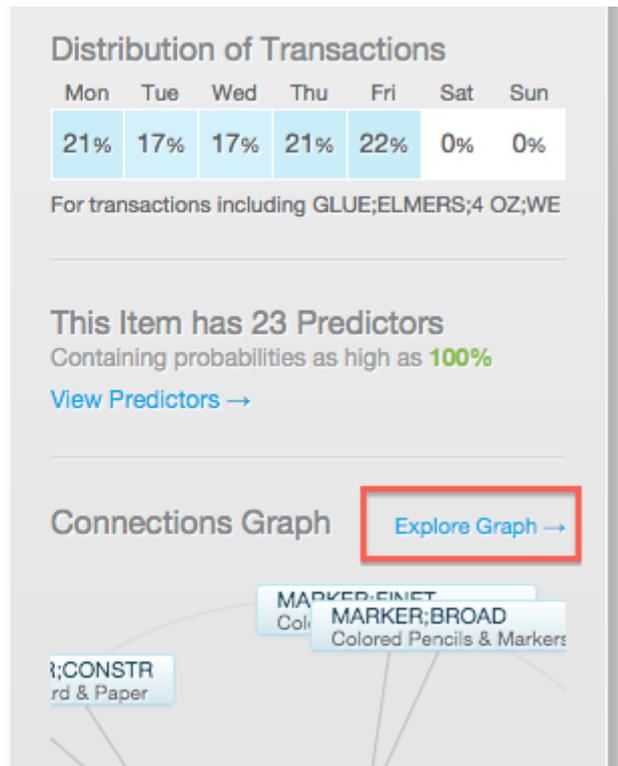
GLUE;ELMERS;4 OZ;WE Connections:

Download CSV

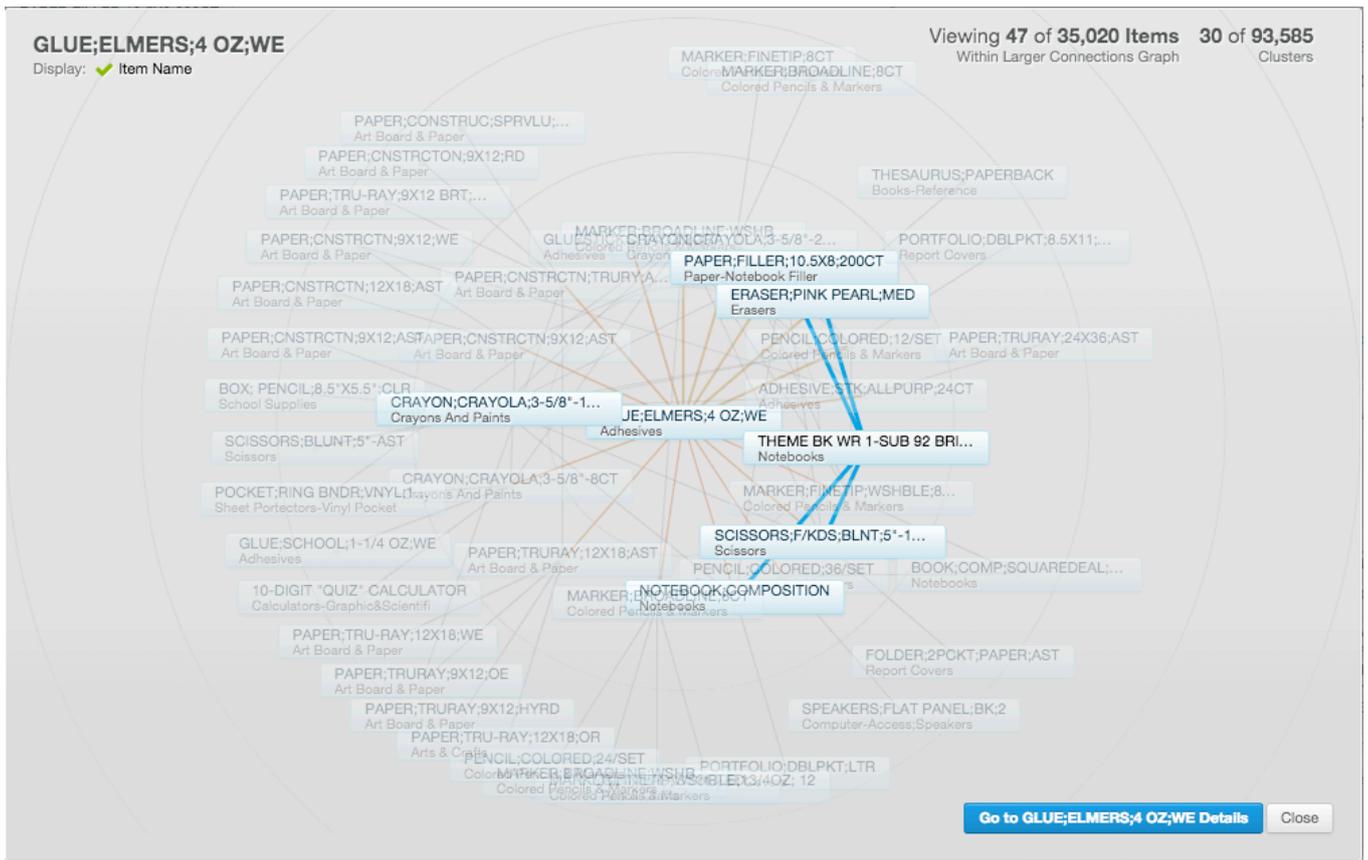
#	Items	Strength	Freq	Cnd Prob	Trend
1	MARKER;BROADLINE;WSHBLE;8CT Colored Pencils & Markers	M	9	5.7%	
2	ERASER;PINK PEARL;MED Erasers	M	8	5.1%	
3	PENCIL;COLORED;12/SET Colored Pencils & Markers	M	11	7.0%	
4	PAPER;FILLER;10.5X8;200CT Paper-Notebook Filler	M	6	3.8%	
5	CRAYON;CRAYOLA;3-5/8"-24CT Crayons And Paints	M	6	3.8%	
6	ADHESIVE;STK;ALLPURP;24CT Adhesives	SW	8	5.1%	

6. Click Explore Graph in the right sidebar.

M	6	3.8%		
M	6	3.8%		
SW	8	5.1%		
SW	7	4.5%		
SW	5	3.2%		
SW	5	3.2%		
SW	4	2.5%		
SW	4	2.5%		
SW	3	1.9%		



A graph containing connections to the GLUE;ELMERS;4 OZ;WE item is displayed. Note that all of the purchased products are represented in this analysis, allowing these outputs to drive recommendations in real time.



Each of these product relationships can be analyzed automatically to deliver intelligent recommendations to customers in real time through an Emcien API. For more information on Emcien APIs click [here](#).

Inventory Management

This article outlines how Emcien can drive a product recommendation engine automatically with historical sales data.

Align your data with a use case

Note: This procedure uses the Retail Sales example data set. Your data should be similar to the example data, particularly the structure of a `transaction_id` or something similar to connect the items in each purchase.

[retail_sales_2014_12_02.receipt.csv.zip](#)

If your data doesn't follow this same structure consult the [Receipt Format](#) page before contacting Emcien Support.

Data Prep

This data set contains retail transactions in the receipt format. While more difficult to manipulate, the receipt format is particularly helpful for sales data because it allows you to attach multiple items from the same category to each transaction through the `transaction_id`. This feature makes the receipt format ideal for basket analysis.

Receipt Format Example

transaction_id	item_id	transaction_date	item_category	item_name	item_volume	item_price	transaction_price	transaction_volume
10001	ABC123	7/4/12	Hardware	"3in, Bolt Steel"	1	\$3.99	\$25.92	1
10001	XYZ456	7/4/12	Cleaner	ACME Bleach	4	\$0.99	\$25.92	1
10001	MNO780	7/4/12	Fasteners	"1 in, Gate Latch"	3	\$5.99	\$25.92	1
10002	QRS123	7/5/12	Live Goods	Lg Potted Fern	2	\$9.99	\$31.97	1
10002	HJK234	7/5/12	Soils	ACME Soil	1	\$11.99	\$31.97	1

1st Transaction (rows 1-3)
Optional Data (columns 5-9)
2nd Transaction (rows 4-5)

Note: This example shows two transactions, the first with three items and the second with two items.

The sample data set contains five columns: the `transaction_id`, `item_id`, `transaction_date`, `item_name`, and `item_category`. Only the `transaction_id` and `item_id` are required fields, but the added `transaction_date`, `item_name`, and `item_category` give added information to be included in the analysis.

Note that the receipt format does not permit user-defined column headings. Some retail data sets include additional columns that relate more to the product than to the purchase, like product margin. These columns can be moved or rolled into the receipt format using the `transaction_id`.

If your data looks similar to the Receipt Data format, you may only have to rename the column headers or change the column order. Other data will require more work. If you need assistance with the format, first refer to <http://emcien.com/help/article/link/preparing-your-data>

Uploading Data

Upload your data for analysis.

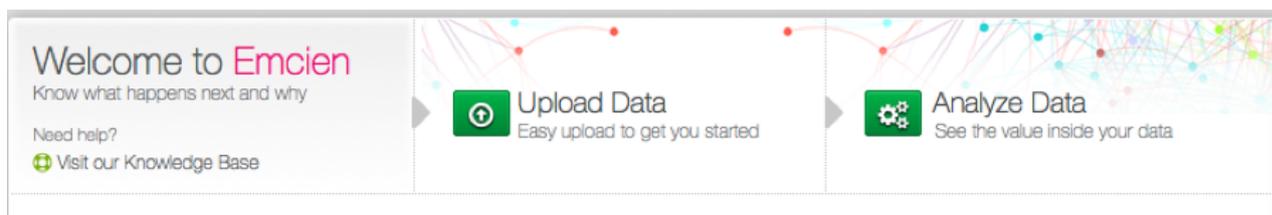
This article covers how to upload data files to Emcien using [File Transfer Protocol \(FTP\)](#).

Make sure you've prepared your data before uploading to Emcien. Check out the [Preparing Your Data](#) article for more information. You can also email us at support@emcien.com for help getting started.

Upload using Drag and Drop

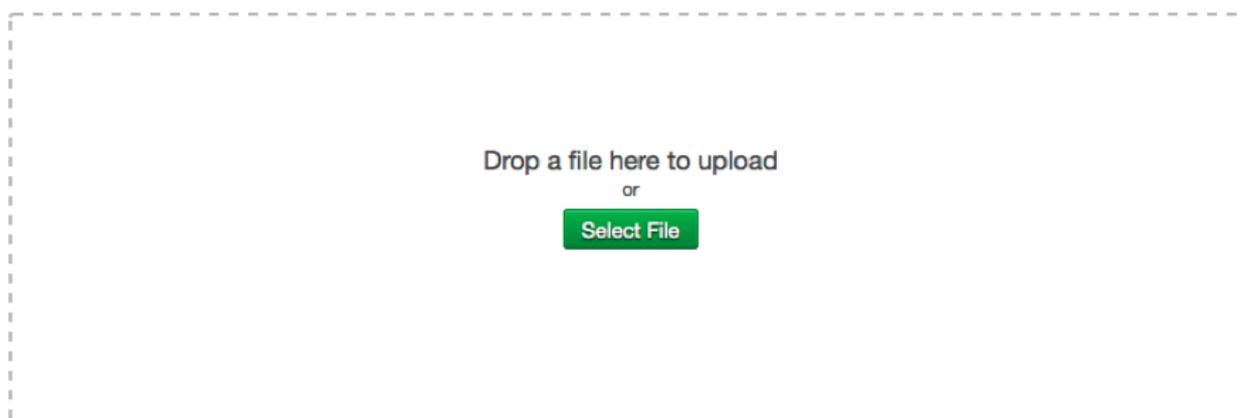
In Emcien version 2.14 you can now drag and drop files into the application for processing. If you are using a previous version you can use one of the other methods listed below.

Simply open the Emcien application and select "Upload Data"

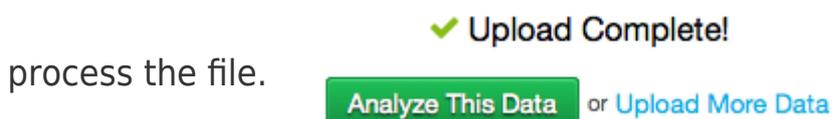


Select the file that you wish to process, then drag and drop it onto the Upload Data screen

Upload Data Services



Once it has finished uploading you can then click on "Analyze This Data" to



Uploading Data to the Emcien Cloud

You can upload data files to Emcien using [SSH File Transfer Protocol \(SFTP\)](#) and your preferred FTP client, such as [FileZilla](#) or [Cyberduck](#).

To connect to Emcien using your FTP client, use the following credentials:

Host: feeds.emcien.com

Username: {Your Emcien Feeds server Username}

Password: {Your Emcien Feeds server Password}

Uploading Data on Mac or Linux

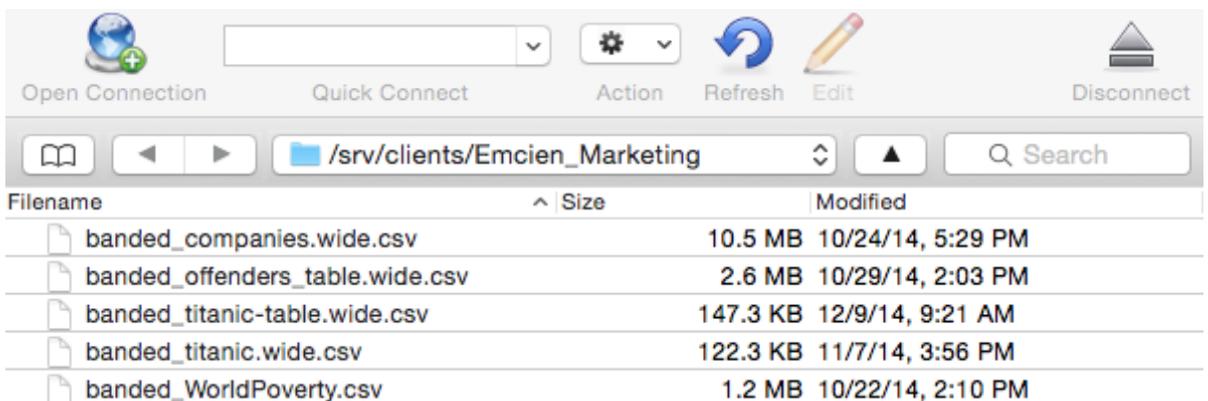
Launch your preferred FTP client. If you do not have a preferred FTP client, we recommend [Cyberduck](#).

For this article, we use screenshots of Cyberduck.

Click Open Connection.



Enter the below credentials and click Connect.



Your FTP client will then display the files on your server.

Server: Your Emcien URL. This URL was set by your IT resource during installation.

Username

Password

Drag and drop your data file to the server files in your FTP application.

On the Emcien home page click Analyze Data. You'll see the data files on this screen.

Uploading Data on Windows

Click the Start button . The Start menu is displayed.

On the Start menu, click Computer. The Computer folder is displayed.

Right-click anywhere in the folder and click Add a network location.

The Add Network Connection wizard is displayed. Click Next.

On the next tab, select Choose a custom network location. Click Next.

On the next tab, enter your Emcien URL.

This URL was set by your IT resource during installation.
Ensure you include *FTP://* at the beginning of the URL.

Then click Next.

Clear the Log on anonymously checkbox.

Type *data* in the username field. Click Next.

On the next tab, name your shortcut *Emcien* by typing it in the Type a name for this network location field. Click Next.

On the next tab, click Finish.

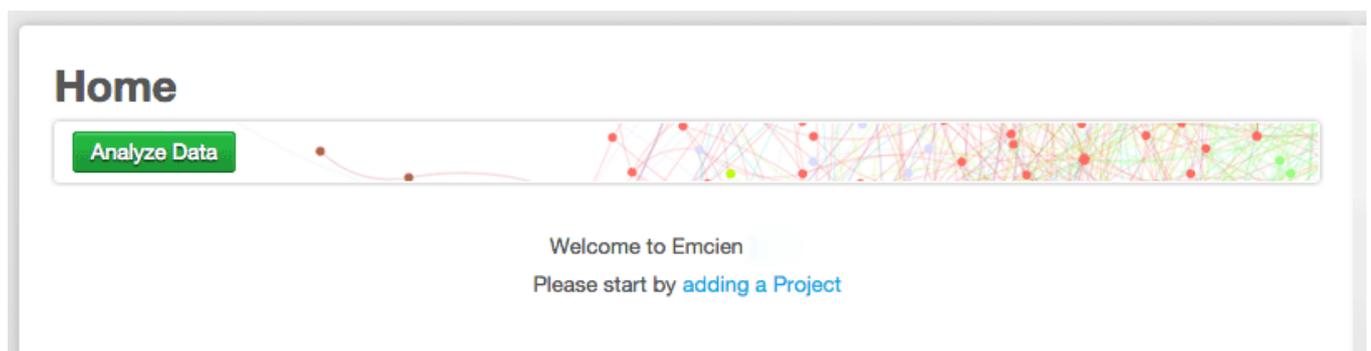
Type *feeds1* in the password field. Check the Save password checkbox so you can connect directly in the future.

Click Log On. You can now drag data files into this folder.

On the Emcien home page, click Analyze Data. You'll see the data files on this screen.

Analysis

On the Home page click Analyze Data.



The Analyze Data button will bring up all of your uploaded data. Select your data

set.

New Analysis Demo Admin

Source Data Filter by Name

Name	File Size	Modified Time
Airlines-Delay-Data.wide.csv	33 MB	Jun 4 10:17AM
✓ retail_sales_2014_12_02.receipt.csv	86 MB	Dec 2 4:29PM
ACME-Auto.csv	170 MB	Jun 2 4:38PM
ACMEOfficeSupplyInputFile.csv.gz	390 MB	Jun 2 2:27PM
BankData3Q2013-noemp.wide.csv	1.4 MB	Jun 2 2:00PM
ACMEAutoParts_initial_NoSubCategory.csv	190 MB	Jun 2 1:59PM
ACMEHardwareMediumCleanedupPricing.csv.gz	220 MB	Jun 2 1:58PM
ACME-Office-1M-Load.csv	370 MB	May 28 3:55PM
Consumer_Complaints.wide.csv	27 MB	Mar 31 9:30AM
ACME-Hardware.csv	2.6 GB	Oct 23 1:00PM

Select a Project Folder, and if necessary rename the Report.

Projects

Reports in a project will display on home page

Demo

Predictions

Override Defaults

Report Name

retail sales 2014 12 02.receipt

Start Analysis

Once your project is organized, click Start Analysis.

retail sales 2014 12 02.receipt Demo Admin



The load screen will take you through each stage of the analysis.

When the analysis is complete you will hear a chime and the View Analysis button will be highlighted in green. Just click the button to see your results.

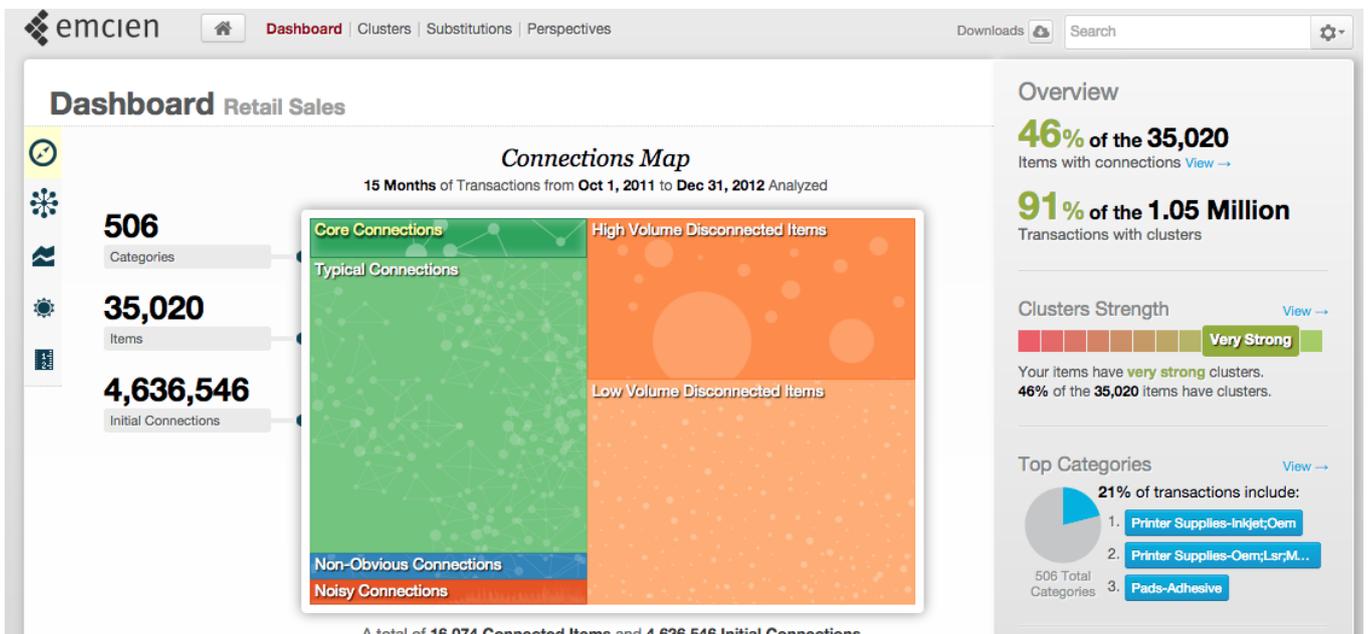
Results

Home

Data Analysis Sort: Created

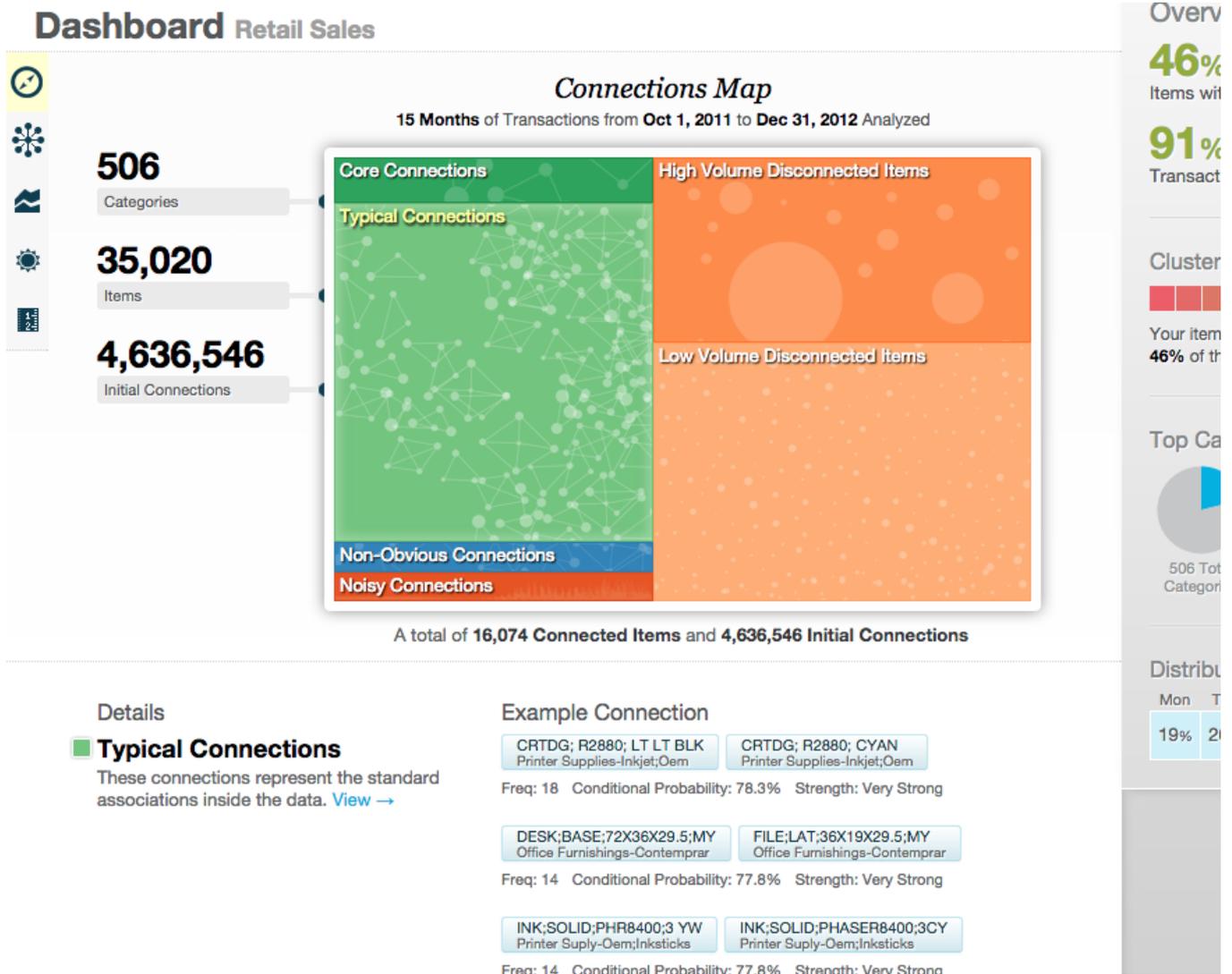
Data Analysis	Analysis Created	Data Ends	Transactions	Items
Retail Sales	4 Hours Ago	December 30 2 Years Ago	1,052,527	35,020
TX Truck	A Day Ago	No Dates	19,829	214
FDIC ROA	October 7 3 Days Ago	No Dates	6,887	6,846
Airline Delay	October 6 4 Days Ago	No Dates	509,519	5,841

The Dashboard page for the Retail Sales data set is displayed. On the Dashboard page you can view the number of connected data values, the number of transactions analyzed, and more.

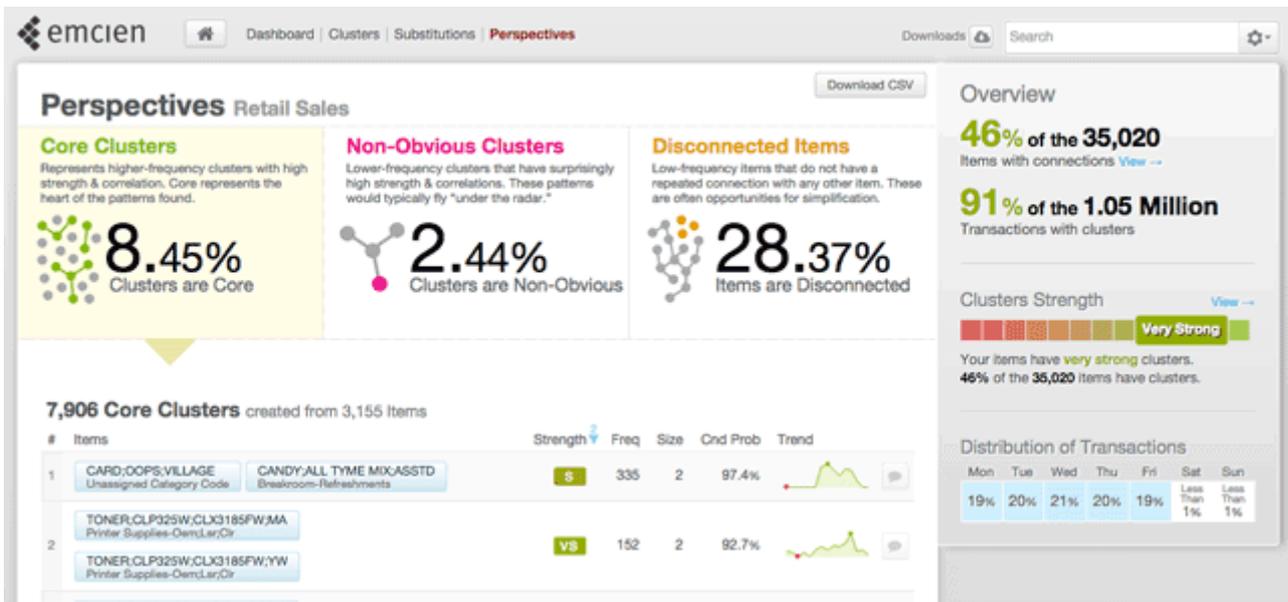


The interactive graphic on the home screen is the Connections Map, a representation of the analyzed data displayed by connectivity.

The green and blue column on the left represents the connections identified as relevant patterns in the data set. Mousing over each section will display examples of the corresponding clusters. These include Core Connections, Typical Connections, and Noisy Connections, and each set is isolated for different purposes.

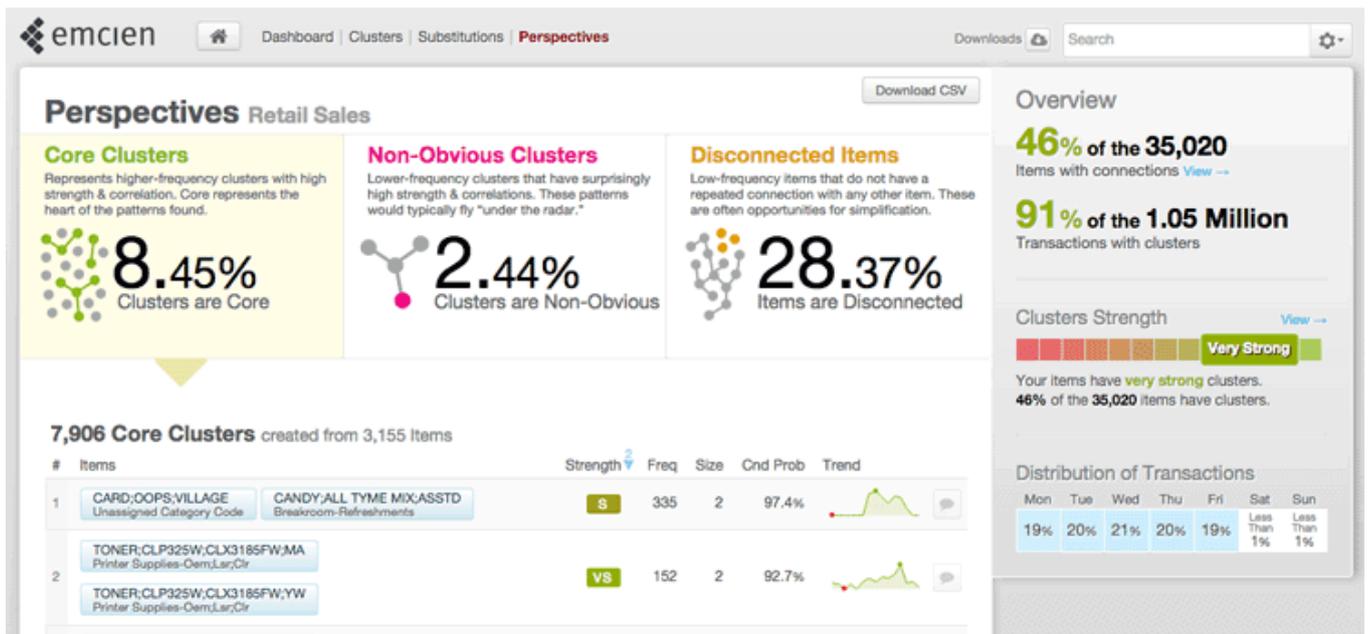


Clicking each of these sections will bring up a list of those connections.



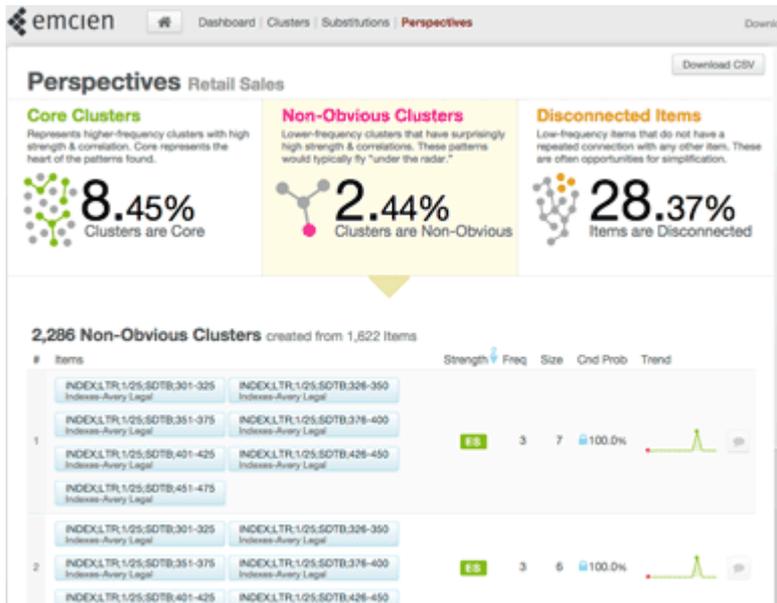
From the Dashboard Home Page you can also click on Perspectives to see these connections listed in greater detail.

Core Clusters are the connections that are central to understanding the data. In retail data these will be the items that often sell together and are sold together frequently.

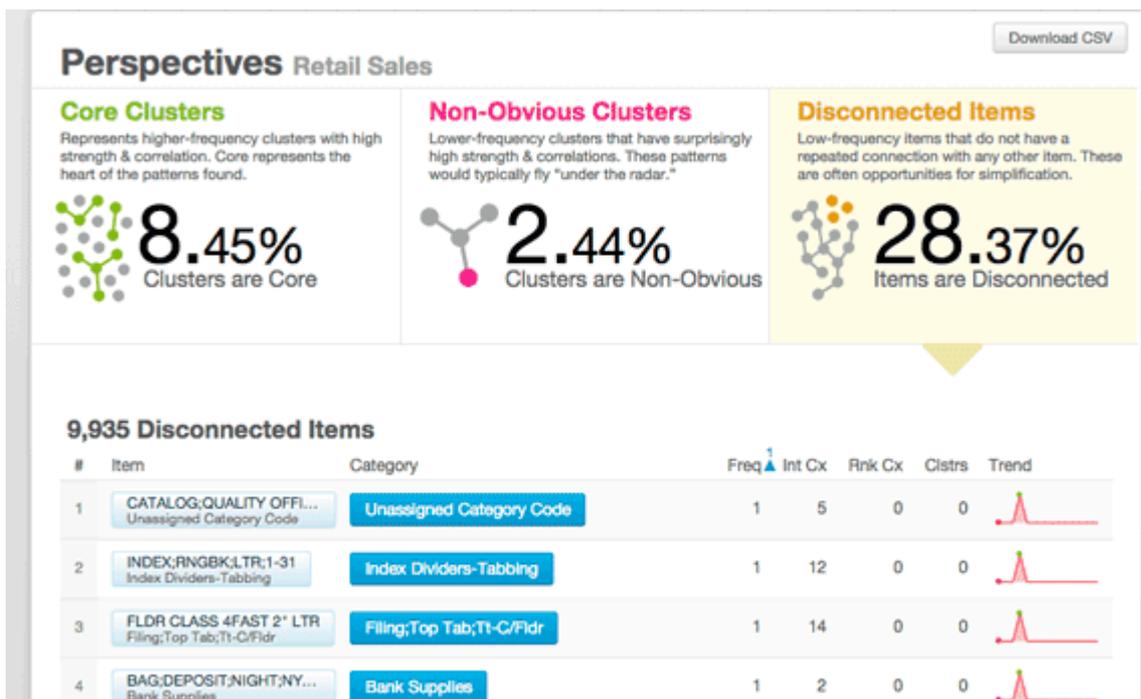


Non-Obvious Clusters represent connections that don't occur frequently but their individual items rarely occur independently of the others. These are important for

inventory management because these pairs can be helpful in avoiding shortages or excess inventory.

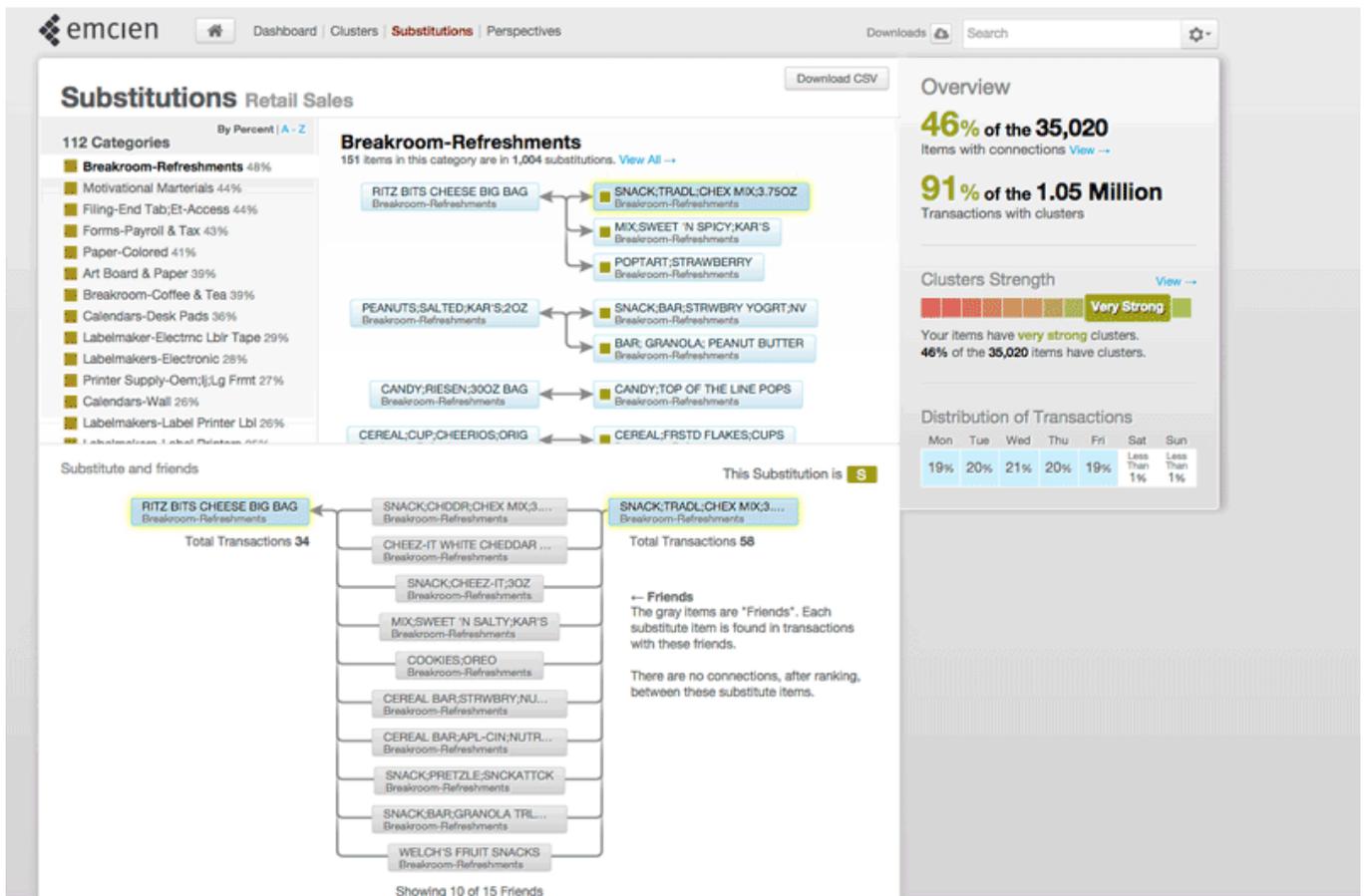


Disconnected Items are particularly helpful for inventory planning because they represent the items that don't have any meaningful connections with other items. These are excellent candidates for inventory reduction. If an item on this list is low margin or has high storage costs it can be considered for reduction.



Another feature that can be helpful for inventory management is the Substitutions feature.

Emcien software automatically identifies the individual items that "act" like other items in the data set. In retail and sales data these are the items that are purchased with the same group of items but are rarely, if ever, seen together.



These are often similar items, competing products, or the same item but in a different size. This is an opportunity to remove the less profitable item.

Each of these product relationships can be analyzed automatically to identify hidden carrying costs and opportunities for bundling and more efficient product planning. For help using Emcien to optimize your inventory you can contact your Services team member directly or email us at hello@emcien.com and we will connect you with a services professional.

Sales/CRM: Predicting Wins & Losses

This article outlines how Emcien can be used to analyze CRM data, like the data that is maintained in salesforce.com, to understand past successes and failures, identify the factors that lead to those outcomes, and apply those results to maximize future sales.

Align your data with a use case

This walkthrough uses the sales-data.csv data set. To apply this same analysis to your own sales data you will need sufficient CRM data.

[sales-data.csv.zip](#)

This demonstration uses a curated and open data set. If your sales data is not as complete or clean as the example set, you may still be able to implement sales analysis. Contact Emcien for help with your specific data challenges or complete automation solutions.

Data Prep

Optimize your data.

In order for Emcien to identify patterns of successful and unsuccessful sales, it is important that the data being analyzed includes dependent variables, or the outcomes you want to predict. For most CRM systems the desired outcome is a concrete outcome of a successful sale. Other CRM events might be more complex,

like identifying the factors that lead to a shorter sales cycle.

In predicting successful sales, it can be helpful to determine what additional data can improve the analysis. Adding data in Emcien is as simple as joining additional data, such as weather data, as the process of analysis is very forgiving.

To further enhance the analysis, or to predict more complex outcomes like shorter sales cycles, it is necessary to derive data from existing sources. For example, if a timestamp is an existing element in the data, it will add depth to the discovered patterns to derive features like seasonality, time of day (e.g. early morning or night shift), geographic region, to the analysis. With multiple timed events you can also derive a sales cycle time. Reference the [Derived Data page](#) and the [Data Prep page](#) for more guidance on data preparation, or you can contact the Emcien Services team for assistance.

Uploading Data

Upload your data for analysis.

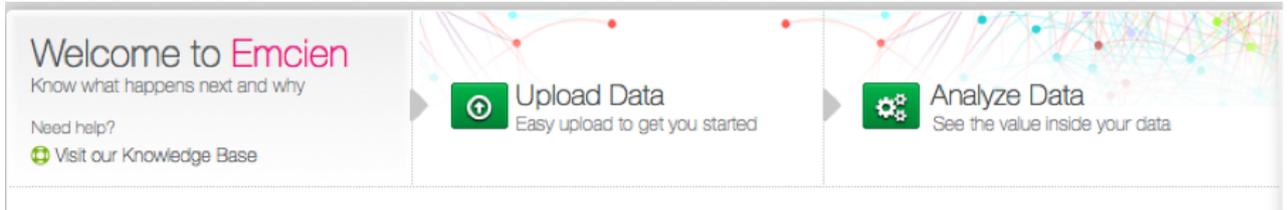
This article covers how to upload data files to Emcien using [File Transfer Protocol \(FTP\)](#).

Make sure you've prepared your data before uploading to Emcien. Check out the [Preparing Your Data](#) article for more information. You can also email us at support@emcien.com for help getting started.

Upload using Drag and Drop

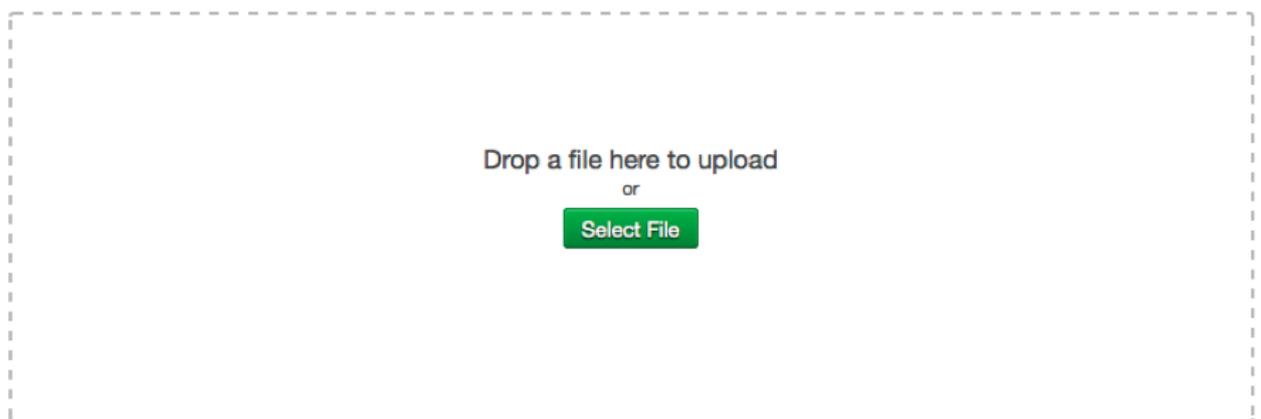
In Emcien version 2.14 you can now drag and drop files into the application for processing. If you are using a previous version you can use one of the other methods listed below.

Simply open the Emcien application and select "Upload Data"



Select the file that you wish to process, then drag and drop it onto the Upload Data screen

Upload Data Services



Once it has finished uploading you can then click on "Analyze This Data" to

process the file.

✔ Upload Complete!

[Analyze This Data](#) or [Upload More Data](#)

Uploading Data to the Emcien Cloud

You can upload data files to Emcien using [SSH File Transfer Protocol \(SFTP\)](#) and your preferred FTP client, such as [FileZilla](#) or [Cyberduck](#).

To connect to Emcien using your FTP client, use the following credentials:

Host: feeds.emcien.com

Username: {Your Emcien Feeds server Username}

Password: {Your Emcien Feeds server Password}

Uploading Data on Mac or Linux

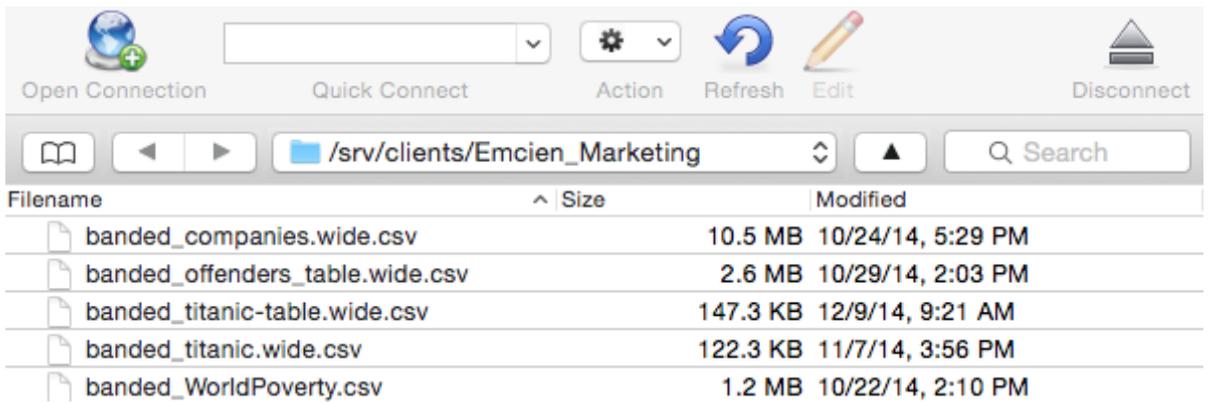
Launch your preferred FTP client. If you do not have a preferred FTP client, we recommend [Cyberduck](#).

For this article, we use screenshots of Cyberduck.

Click Open Connection.



Enter the below credentials and click Connect.



Your FTP client will then display the files on your server.

Server: Your Emcien URL. This URL was set by your IT resource during installation.

Username

Password

Drag and drop your data file to the server files in your FTP application.

On the Emcien home page click Analyze Data. You'll see the data files on this screen.

Uploading Data on Windows

Click the Start button . The Start menu is displayed.

On the Start menu, click Computer. The Computer folder is displayed.

Right-click anywhere in the folder and click Add a network location.

The Add Network Connection wizard is displayed. Click Next.

On the next tab, select Choose a custom network location. Click Next.

On the next tab, enter your Emcien URL.

This URL was set by your IT resource during installation.
Ensure you include *FTP://* at the beginning of the URL.

Then click Next.

Clear the Log on anonymously checkbox.

Type *data* in the username field. Click Next.

On the next tab, name your shortcut *Emcien* by typing it in the Type a name for this network location field. Click Next.

On the next tab, click Finish.

Type *feeds1* in the password field. Check the Save password checkbox so you can connect directly in the future.

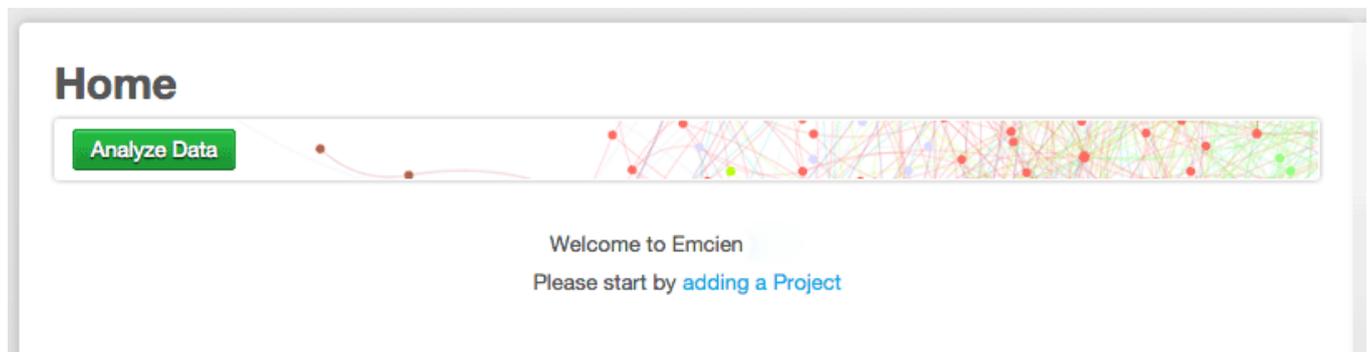
Click Log On. You can now drag data files into this folder.

On the Emcien home page, click Analyze Data. You'll see the data files on this screen.

Analysis

Begin your analysis.

On the Home page click Analyze Data.



The Analyze Data button will bring up all of your uploaded data. Select your data set.

New Analysis Emcien Marketing Admin

Data

Please select a Data File

Filter by Name

Name	File Size	Modified Time
✓ banded_sales-full2.wide.csv	7.9 MB	Apr 3 9:55AM
banded_bank-additional-full.wide.csv	4.4 MB	Apr 2 4:40PM
banded_FortuneGlobal500.wide.csv	980 KB	Apr 1 2:19PM
banded_blind_test.wide.csv	1.1 MB	Apr 1 1:31PM
banded_LANLfailure.wide.csv	5.4 MB	Mar 23 4:56PM
banded_churn-orange.wide.csv	53 MB	Mar 20 3:05PM
ITOCLogInCSV_time.wide.csv	3.3 MB	Mar 18 2:01PM
ITOCLogInCSV_time_TEST_4K.wide.csv	510 KB	Mar 18 11:20AM
banded_Machine.wide.csv	3.9 MB	Mar 16 11:04AM
banded_Machine.csv	3.9 MB	Mar 16 11:02AM

Projects

Select a Project to display Report on Home Page

- Sales Data
- Network Data

Prediction Category

Up to 20 Items allowed in Prediction Category

Select a Project Folder, and if necessary rename the Report. For this data set we will be predicting the Category "Sale".

banded_Machine.csv	3.9 MB	Mar 16 11:02AM
--------------------	--------	----------------

Projects

Select a Project to display Report on Home Page

- Sales Data
- Network Data

Prediction Category

Up to 20 Items allowed in Prediction Category

Override Defaults

Once your project is organized, click Start Analysis.

Prediction Category
Up to 20 Items allowed in Prediction Category

Sale

+ Override Defaults

Report Name

Sales Data V1

Start Analysis Cancel

The load screen will take you through each stage of the analysis.

Sales Data V1 Log Emcien Marketing Admin

Transferred 3s → Parsed 3s → Computing → Load → View Analysis

Log: [Summary](#) | **Expanded** Run Time: 3m 49s

#	Message	Time
37	--> Building Engine::Clusters::Results as clusters.dat	14s
36	--> Building Engine::Trends::TrendData as item_trends.dat	14s
35	--> Building Engine::RetailSubs::Results as retail_subs.dat	14s
34	--> Building Engine::Graph::DegreesByPeriod as degrees_by_period.dat	14s
33	--> Building Engine::Graph::Lol as graph.dat	12s
32	--> Building Engine::Index::Duplicates as duplicates.dat	12s
31	--> Building Engine::Index::Transaction as index.dat	12s
30	--> Building Engine::Trends::CatTrend as cat_trends.dat	12s
29	--> Building Engine::Index::CatTransaction as catindex.dat	12s
28	--> Building Engine::Input::CatData as catdata.dat	12s

When the analysis is complete you will hear a chime and the View Analysis button will be highlighted in green. Just click the button to see your results.

Sales Data V1 Log Emcien Marketing Admin



Results

See your results.

By clicking the “View Analysis” button you are accessing the precomputed results of your analysis. From the Emcien Dashboard Page you can explore the connections in your analyzed data, but the current objective is to determine the factors that signal the outcome of a positive sale.

Dashboard Sales Data V1

- 22
Categories
- 144
Items
- 8,068
Initial Connections

Connections Map

Overview

99% of the **144**
Items with connections [View →](#)

100% of the **32,821**
Transactions with clusters

Predicting: sale

8.1K
Predictive Rules

2
Outcome Items
[View →](#)

0
Predictions
[View →](#)

In the right hand column simply click on the blue button marked “sale” to see the Category Details Page.

Category Details sale

sale contains: Download CSV

#	Item	Freq ¹ ↓	Int Cx	Rnk Cx	Clstrs	Trend
1	no sale	29,140	142	117	141,714	N/A
2	yes sale	3,681	141	127	6,504	N/A

Within the Sale Category are two items, Sale:Yes and Sale:No. Click on the light blue “yes” button to see what items connect to a successful sales outcome. By default these items are ordered by the strength of their connection with the Category Sale and the Item yes. There are many options for exploring the connections with this outcome, but the most helpful will be View Predictors link.

Ranked Connections | [Clusters](#) | [Initial Connections](#)

Viewing 2-item clusters after ranking

[Download Connections CSV](#)

#	Items	Strength	Freq	Cnd Prob	Trend
1	[0.634_to_1.085] euribor 3 month rate	S	1,552	42.2%	N/A
2	[0.0_to_989.01] days since last contact	S	777	21.1%	N/A
3	success previous campaign outcome	S	719	19.5%	N/A
4	[836.0_to_4918.0] last contact duration	S	730	19.8%	N/A
5	cellular contact type	S	3,056	83.0%	N/A
6	no credit in default	S	3,337	90.7%	N/A
7	[1.0_to_2.567593] contacts this campaign	S	2,787	75.7%	N/A
8	no personal loan	S	3,056	83.0%	N/A
9	[-40.8_to_-37.5] consumer confidence index	S	650	17.7%	N/A

yes
ID

sale
Category [View](#) →

6,504 Clusters [View](#) → **3,681** Transactions

141 Initial Connections **0** Substitutions

Transactions containing this Item

Avg Qty per Trans **1.0**

Avg Trans Size **22.0**

This Item has 1,699 Predictive Rule S
Containing probabilities as high as **90%**
[View Predictive Rules](#) →

Connections Graph [Explore Graph](#) →

Clicking on View Predictive Rules will display the predictors of positive sales, combining the frequency and probability of the connection between each item, or even between complex groups of items and your selected outcome.

Prediction Rule Search Sales Data V1

Your search matched **1,699** of the **8,124** total prediction rules. Search criteria: Outcome Item: Item ID is exactly “yes”.

[Download CSV](#)

#	Freq	Prediction Rule	Probability	Outcome Item	Size	Cover	Contro
1	1,552	[0.634_to_1.085] euribor 3 month rate	45.7% →	yes sale	1	42.2%	0.043
2	1,401	[0.634_to_1.085] euribor 3 month rate cellular contact type	48.8% →	yes sale	2	38.1%	0.042
3	1,262	[0.634_to_1.085] euribor 3 month rate [1.0_to_2.567593] contacts this campaign	46.2% →	yes sale	2	34.3%	0.035
4	1,171	[0.634_to_1.085] euribor 3 month rate cellular contact type no personal loan	48.8% →	yes sale	3	31.8%	0.035
5	1,144	[0.634_to_1.085] euribor 3 month rate cellular contact type [1.0_to_2.567593] contacts this campaign	49.3% →	yes sale	3	31.1%	0.035
6	777	[0.0_to_989.01] days since last contact	63.8% →	yes sale	1	21.1%	0.032
7	894	[0.634_to_1.085] euribor 3 month rate [159.0_to_393.0] last contact duration	56.6% →	yes sale	2	24.3%	0.032

Advanced Search
Add search criteria for more specific searches

Outcome Item:

Is Exactly:

[+ Add Criteria](#) [Search](#)

Other Searches

[Item Search](#)
Perform complex queries for item

[Category Search](#)
Perform complex queries for category

[Cluster Search](#)
Perform complex queries for cluster

[Substitution Search](#)
Perform complex queries for substitution

The software displays every one of the 1,699 patterns that have meaningful connections to the successful sales outcome. Now that the historical data has been captured and analyzed you can plan future sales ventures that will maximize

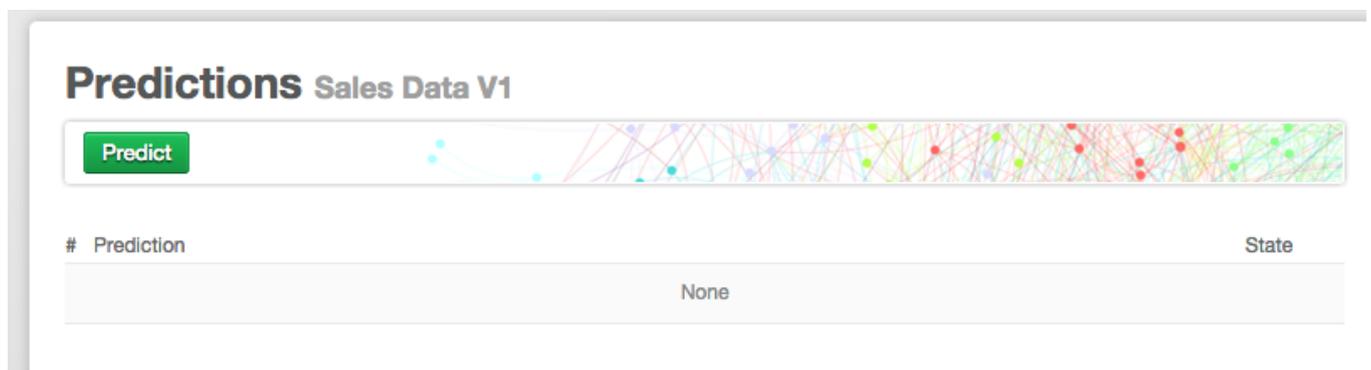
the effectiveness of your sales force and their campaigns.

To take this analysis even further, the Emcien Services team can implement end to end integrations to apply these predictions to new leads and feed directly into your CRM system, even making predictions in real time.

Make Predictions

Apply predictive patterns to new data.

Now that the patterns that identify successful sales opportunities have been identified, Emcien's real-time prediction engine can be used to evaluate and qualify leads as they progress through the sales pipeline. This application of our two-tiered architecture can be demonstrated by clicking on Predictions in the menu at the top of the screen. Simply upload the file of new or incoming data to the same FTP address.



Once your new sales prospects data has been uploaded, click on the green Predict button and select the file. You can let the software append (Prediction) to the existing file name, or you can create a new name for your prediction.

New Prediction Sales Data V1

Source Data Filter by Name

banded_sales-full2.wide.csv	7.9 MB	May 20 1:34PM
✓ banded_sales-test.wide.csv	2.00 MB	May 20 1:34PM
banded_sales-train.wide.csv	7.9 MB	May 20 1:34PM
banded_sales_data.wide.csv	9.00 MB	May 18 10:25AM
banded_Sample-Receipt-Format-Data-Workbook.wide.csv	810 KB	Oct 20 9:53AM + ×
banded_SampleWide.csv	840 Bytes	May 13 1:22PM
banded_worldbank.wide.csv	1.2 MB	Mar 2 8:25AM
BankData3Q2013-noemp.wide.csv	1.4 MB	May 14 10:14AM
BankData3Q2013.wide.csv	1.4 MB	Nov 6 9:55AM
Concatenated_with_quantities.csv	5.8 MB	Apr 3 11:15AM
Concatenated_with_quantities.wide.csv	5.8 MB	Apr 3 11:20AM

Then just click Predict and the Predictions module will apply your analyzed rules to the incoming data.

Prediction Name

banded sales test (Prediction)

Predict

The new data will be scored with all of the rules that justify each prediction and output to a machine and human-readable .CSV file.

Prediction Details Diabetes Test (Prediction)

Working:  Lookups Built

Once the Prediction has completed processing the data you will see the following:

Prediction Overview banded sales test (Prediction)

2 Outcome Items

#	Predicted	Item ¹	Captured	TP	FP	P%	TN	FN	N%	Actual	Skip
1	6,120 Predictions	no sale	81.5%	5,890	230	96.2%	695	1,334	31.4%	7,408	184
2	2,029 Predictions	yes sale	75.1%	695	1,334	34.3%	5,890	230	95.7%	959	34

This screen shows the Outcome Items that were being predicted along with some metrics that will assist you in your analysis. One of the important ones is "Captured". If you hover your mouse over any of the headings a pop up will be displayed that describes each item in more detail.

Captured:
Percentage of test transactions that were correctly predicted.
TP / (Actual Count - Skipped Count)

Prediction Overview banded sales test (Prediction)

2 Outcome Items

#	Predicted	Item ¹	Captured	TP	FP	P%	TN	FN	N%	Actual	Skip
1	6,120 Predictions	no sale	81.5%	5,890	230	96.2%	695	1,334	31.4%	7,408	184
2	2,029 Predictions	yes sale	75.1%	695	1,334	34.3%	5,890	230	95.7%	959	34

Used inside of the Emcien application these predictions can be very useful, but with the help of our Services team you will be able to automate this workflow. Through our automation tools, Emcien can create a custom data flow that will detect new prospects, use the latest analysis to generate a predictive score, and insert that predictive score directly into your CRM system. Contact the [Emcien Services team](mailto:services@emcien.com) at services@emcien.com to find out how.

Making Predictions from a Range of Outcomes

This article outlines how you can find the predictors of significant airline delay using Emcien.

Note: This article uses the Airline Delay example data set.

Align your data with a use case

This walkthrough uses the Airlines-Delay data set. Click on the file name below to download the data.

[Airlines-Delay.wide.zip](#)

Data Prep

Optimize your data.

In order for Emcien to identify patterns of delayed flights, it is important that the data being analyzed includes dependent variables, or the outcomes you want to predict.

Uploading Data

Upload your data for analysis.

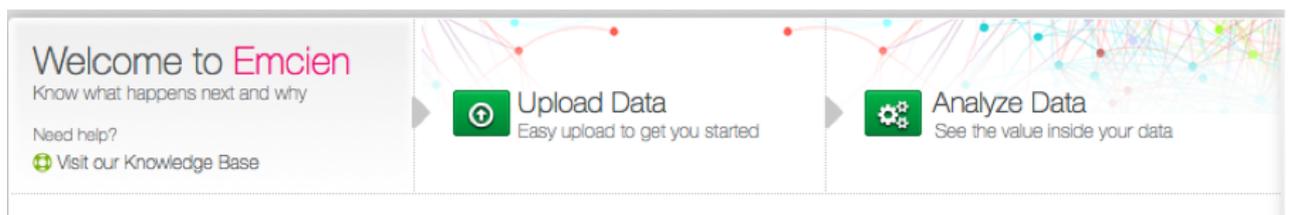
This article covers how to upload data files to Emcien using [File Transfer Protocol \(FTP\)](#).

Make sure you've prepared your data before uploading to Emcien. Check out the [Preparing Your Data](#) article for more information. You can also email us at support@emcien.com for help getting started.

Upload using Drag and Drop

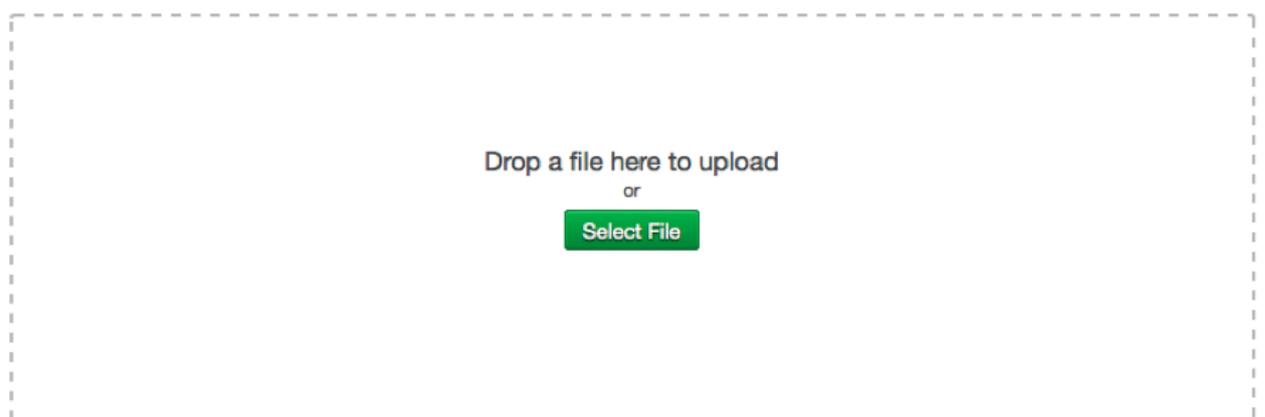
In Emcien version 2.14 you can now drag and drop files into the application for processing. If you are using a previous version you can use one of the other methods listed below.

Simply open the Emcien application and select "Upload Data"



Select the file that you wish to process, then drag and drop it onto the Upload Data screen

Upload Data Services



Once it has finished uploading you can then click on "Analyze This Data" to

✔ Upload Complete!

process the file.

[Analyze This Data](#) or [Upload More Data](#)

Uploading Data to the Emcien Cloud

You can upload data files to Emcien using [SSH File Transfer Protocol \(SFTP\)](#) and your preferred FTP client, such as [FileZilla](#) or [Cyberduck](#).

To connect to Emcien using your FTP client, use the following credentials:

Host: feeds.emcien.com

Username: {Your Emcien Feeds server Username}

Password: {Your Emcien Feeds server Password}

Uploading Data on Mac or Linux

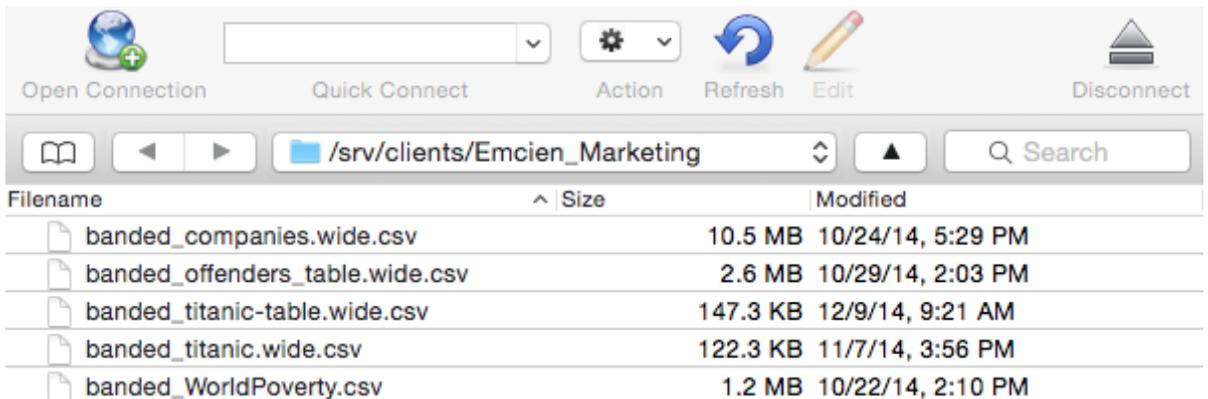
Launch your preferred FTP client. If you do not have a preferred FTP client, we recommend [Cyberduck](#).

For this article, we use screenshots of Cyberduck.

Click Open Connection.



Enter the below credentials and click Connect.



Your FTP client will then display the files on your server.

Server: Your Emcien URL. This URL was set by your IT resource during installation.

Username

Password

Drag and drop your data file to the server files in your FTP application.

On the Emcien home page click Analyze Data. You'll see the data files on this screen.

Uploading Data on Windows

Click the Start button . The Start menu is displayed.

On the Start menu, click Computer. The Computer folder is displayed.

Right-click anywhere in the folder and click Add a network location.

The Add Network Connection wizard is displayed. Click Next.

On the next tab, select Choose a custom network location. Click Next.

On the next tab, enter your Emcien URL.

This URL was set by your IT resource during installation.
Ensure you include *FTP://* at the beginning of the URL.

Then click Next.

Clear the Log on anonymously checkbox.

Type *data* in the username field. Click Next.

On the next tab, name your shortcut *Emcien* by typing it in the Type a name for this network location field. Click Next.

On the next tab, click Finish.

Type *feeds1* in the password field. Check the Save password checkbox so you can connect directly in the future.

Click Log On. You can now drag data files into this folder.

On the Emcien home page, click Analyze Data. You'll see the data files on this screen.

Analysis

Begin your analysis.

Click on the Analyze Data button

Overview Demo Admin

Analyze Data

#	Name	State	Project	▼ Created	Import	Menu
None						

Users

#	Email	Last Sign On
1	demo@emcien.com ID: 11	A Few Seconds Ago

Projects

#	Name	Reports
1	Demo ID: 3	0

You will then need to select the data file that you uploaded or "Drag and Dropped" for analysis.

Source Data			Filter by Name
Name	File Size	Modified Time	
ACME-Auto.csv	170 MB	Jun 2 4:38PM	
ACMEOfficeSupplyInputFile.csv.gz	390 MB	Jun 2 2:27PM	
BankData3Q2013-noemp.wide.csv	1.4 MB	Jun 2 2:00PM	
ACMEAutoParts_initial_NoSubCategory.csv	190 MB	Jun 2 1:59PM	
ACMEHardwareMediumCleanedupPricing.csv.gz	220 MB	Jun 2 1:58PM	
ACME-Office-1M-Load.csv	370 MB	May 28 3:55PM	
Consumer_Complaints.wide.csv	27 MB	Mar 31 9:30AM	
✓ Airlines-Delay-Data.wide.csv	33 MB	Mar 31 9:30AM	
ACME-Hardware.csv	2.6 GB	Oct 23 1:00PM	
ACME-Web-Analvtics-CA-Cleaned.wide.csv.az	310 MB	Oct 22 1:56PM	

[Upload New Source Data](#)

Next you will enter the category that we want to have Emcien Predict: Departure

 Predictions

Delay

Outcome Category Name

Departure Delay

Must have 20 items or less

Enter a report name or except the default created and click "Start Analysis"

Report Name

Airlines Delay Data

Start Analysis

The application will transfer the data into Emcien, parse it for processing, compute all of the connections and predictions, and then load all of the data for the analysis. Go to the next step in the process to "View Analysis"

Results

See your results.

Click on "View Analysis"

Airlines Delay Data Demo Admin



Log: **Summary** | Expanded

Run Time: 1m 56s

#	Message	Time
102	Queueing Backup Job	1m 56s
101	-----> Import successful for report 81968969	1m 56s
100	Report State: Report-ready	1m 55s
99	Report State: report-verify	1m 55s
98	Report State: downloads-built	1m 54s

The Dashboard page for the Airline Delay data set is displayed. On the Dashboard page, you can view the number of connected data values, the number of transactions analyzed, and more information.

Dashboard Airlines Delay Data

- 9 Categories
- 5,841 Items
- 571,088 Initial Connections

Connections Map

A total of 4,258 Connected Items and 571,088 Initial Connections

Overview

73% of the 5,841 Items with connections [View →](#)

100% of the 509,519 Transactions with clusters

Predicting: **DEPARTURE DELAY**

585 Predictive Rules | 5 Outcome Items [View →](#) | 0 Predictions [View →](#)

Clusters Strength [View →](#)

Ext Strong

Your Items have **ext strong** clusters. **73%** of the 5,841 Items have clusters.

Top Categories [View →](#)

100% of transactions include:

- CARRIER
- DEPARTURE DELAY
- DEST_CITY_NAME

9 Total Categories

Details

Core Connections

Represents higher-frequency clusters with high strength and correlation. Core represents the heart of the patterns found. [View →](#)

Example Connection

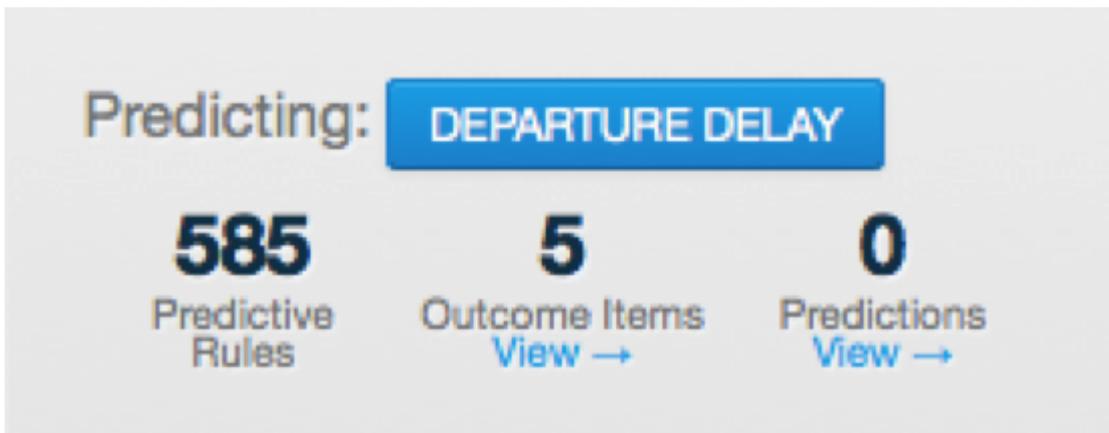
RST ORIGIN Rochester, MN ORIGIN_CITY_NAME

Freq: 164 Conditional Probability: 100.0% Strength: Extremely Strong

RST DEST Rochester, MN DEST_CITY_NAME

Freq: 166 Conditional Probability: 100.0% Strength: Extremely Strong

On the Dashboard page, you will see Predicting: "Departure Delay".



Click on View under the Outcome Items.

Category Details **DEPARTURE DELAY**

DEPARTURE DELAY contains:

[Download CSV](#)

#	Item	Freq ¹	Int Cx	Rnk Cx	Clstrs	Trend
1	ON TIME DEPARTURE DELAY	321,372	5,796	181	1,337	N/A
2	Significant Delay (30) DEPARTURE DELAY	86,292	5,731	362	979	N/A
3	Severe Delay (>30) DEPARTURE DELAY	49,336	5,597	440	882	N/A
4	Minor Delay (5) DEPARTURE DELAY	44,958	5,628	334	719	N/A
5	Cancelled DEPARTURE DELAY	7,561	2,972	121	229	N/A

The Category Details page for the Departure Delay category is displayed. This page displays all the items associated with departure delay.

In our example we are interested in what Predicts a Significant Delay.

Click on "Significant Delay (30)" to display the Item detail.

You will see that there are 109 Predictive Rules for this item.

Rank	Item	Category	Value	Percentage	Other
1	UA CARRIER				N/A
2	OO CARRIER	M	7,708	8.9%	N/A
3	DL CARRIER	M	6,828	7.9%	N/A
4	Chicago, IL	M	7,180	8.3%	N/A

Click on "View Predictive Rules" to display the list of those items that predict "Significant Delay"

A Prediction Rule Search is displayed for the Significant Delay item. This page displays all the predictive rules for significant airline delays. To learn more about the columns in the data you can hover your mouse over each for a more detailed description.

If you click on the "Tell Me" icon ,

a pop up will display more information about that specific item in the list.

Prediction Rule Search Airlines Delay Data

Download CSV

Your search matched **109** of the **585** total prediction rules. Search criteria:
Outcome Item: Item ID is exactly "Significant Delay (30)".

Freq Prediction Rule Probability Outcome Item Size Cover Contro¹

Tell Me:

if **DEN ORIGIN** and **WN CARRIER** are in a transaction there is a
42.1% probability of **Significant Delay (30) DEPARTURE DELAY** being present.

2.19% of the outcome item's transactions (1,894 of 86,292) are predicted

5	1,894	DEN ORIGIN	WN CARRIER	42.1% →	Significant De... DEPARTURE D...	2	2.19%	0.0055		
6	1,891	DEN ORIGIN	WN CARRIER	0 DIVERTED	42.1% →	Significant De... DEPARTURE D...	3	2.19%	0.0055	
7	3,652	DFW ORIGIN	AA CARRIER	27.8% →	Significant De... DEPARTURE D...	2	4.23%	0.0046		
8	3,647	DFW ORIGIN	AA CARRIER	0 DIVERTED	27.8% →	Significant De... DEPARTURE D...	3	4.23%	0.0046	
9	2,058	MDW ORIGIN	WN CARRIER	35.5% →	Significant De... DEPARTURE D...	2	2.38%	0.0044		

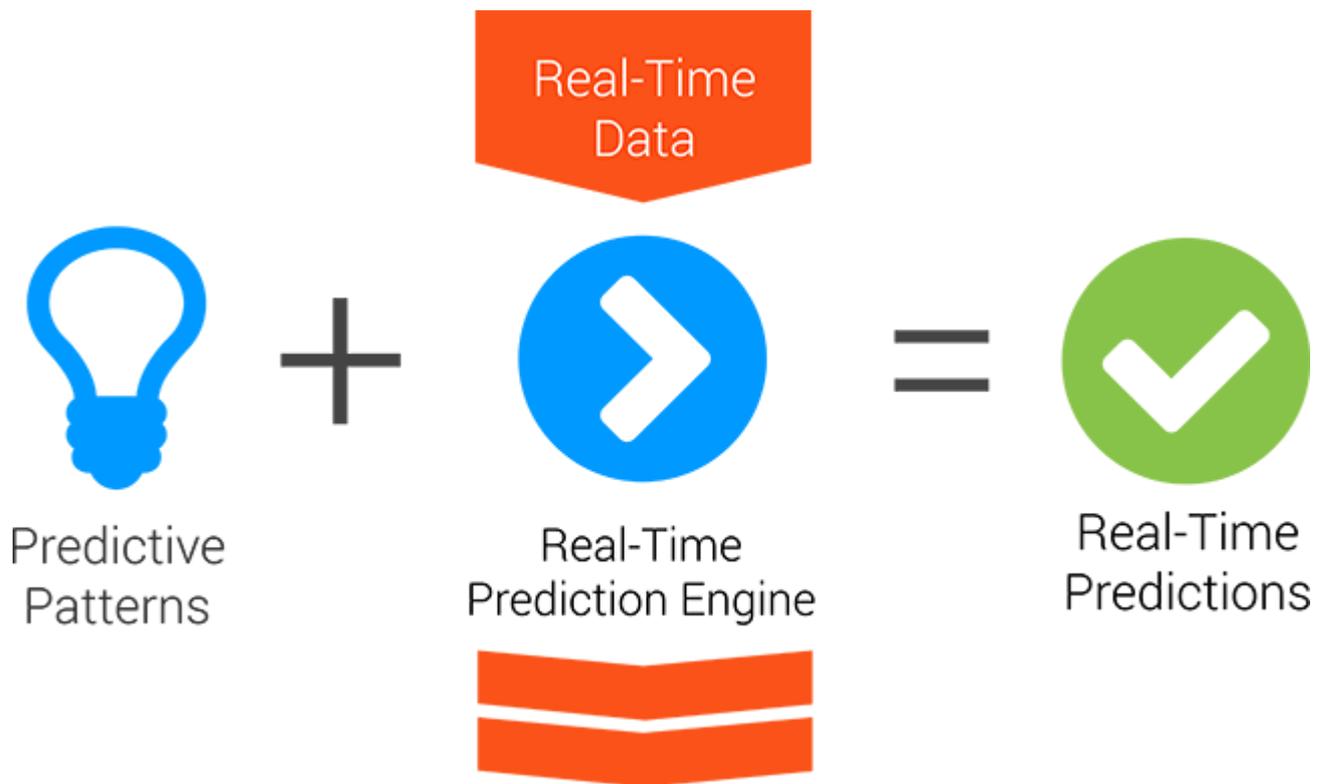
Based on the data that was processed the system can predict that if the Origin is DEN and the Carrier is WN there is a 42% probability of a Significant Delay.

Predictions can be automated through Emcien's APIs to direct other systems or software. For more on this, see:

<http://support.emcien.com/help/article/link/emcien-apis>

Real-Time Predictions

The Emcien real-time prediction engine makes it easy to generate fast and accurate predictions. This article covers everything from the process of automatically analyzing data to applying detected signatures to new data in real time.



How the real-time prediction engine works

Historical data is analyzed automatically by Emcien, extracting predictive signatures from the data. The real-time prediction engine then applies those signatures to incoming data streams.

Exercise: Predicting Diabetes for Patients

We will demonstrate the capabilities of real-time prediction engine using a public dataset that contains patient metrics and a diagnosis of Diabetic or Not Diabetic. This example assumes you already know how to upload a dataset for analysis and will walk you through the steps to apply Emcien's automatically generated signatures to make predictions for diabetes.

Getting Started

1. [Download the diabetes data files](#).
2. Load the banded train and test data sets into Emcien. Check out the [Loading Your Data](#) article for more information.
3. Begin a new analysis by clicking Analyze Data on the Emcien home page.
4. On the Analyze Data page, select the sample diabetes data set you just uploaded.
5. To predict the outcome of the *Diabetes* column in the data set, type Diabetes in the Prediction Category field.

Prediction Category Name

Prediction Category may have 20 items or less

Override Defaults

6. Click Start Analysis.

Report Name

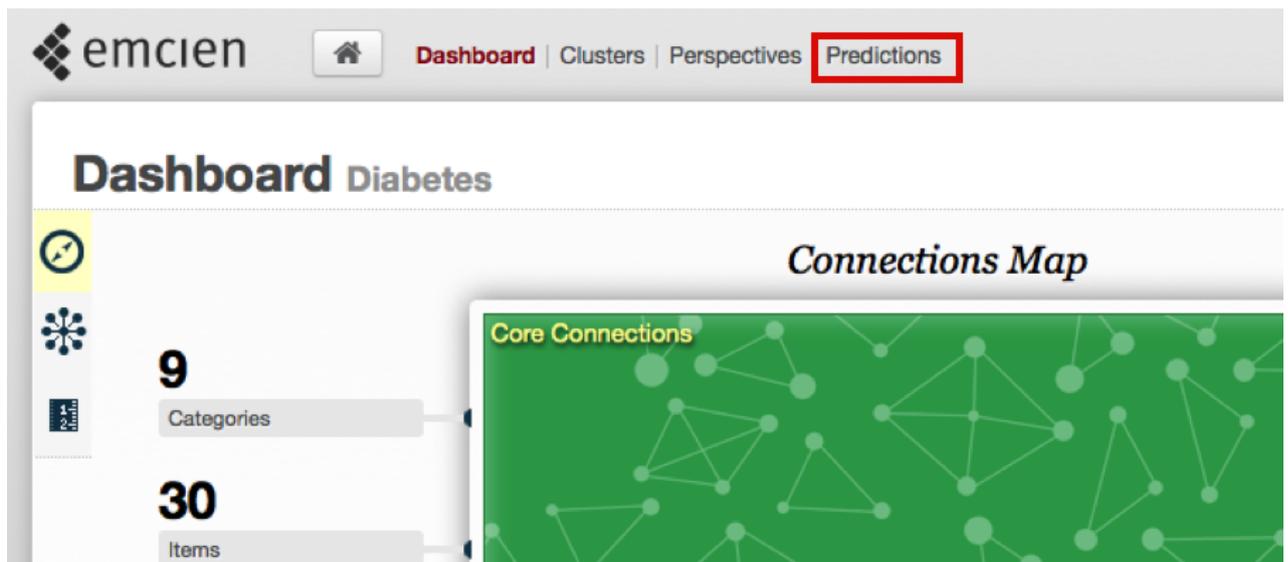
7. When the analysis is complete, View Results is displayed.

Click View Results and the Emcien dashboard is displayed for the sample data set.

Making Predictions

After analyzing the sample data set, you can begin making predictions using new data.

1. Click Predictions heading in the upper navigation.



The Predictions page is displayed.

2. On the Prediction page, click Predict at the top of the page.

Predictions Diabetes

Predict

# Prediction	State
None	

The New Prediction page is displayed.

3. On the New Prediction page, select the sample diabetes data set and click Predict.

New Prediction diabetes train

Data Filter by Name

Please select a Data File

banded_test.csv	1.8 MB	Mar 25 2:11PM
banded_test.wide.csv	1.8 MB	Mar 25 2:13PM
banded_testRevenue.wide.csv	47 MB	Mar 25 5:01PM
banded_trainRevenue.wide.csv	69 KB	Mar 25 4:48PM
diabetes_input_data_file.csv	24 KB	Mar 26 1:03PM
✓ diabetes_test.wide.csv	9.1 KB	Mar 26 1:04PM
diabetes_train.wide.csv	85 KB	Mar 26 1:04PM
FDIC-ROA.wide.csv.gz	160 KB	Feb 23 10:30AM
LANLfailure.wide.csv	2.7 MB	Mar 24 10:25AM
PimaIndiansDiabetes.wide.csv	24 KB	Mar 25 12:16PM
Trucks.wide.csv.gz	440 KB	Feb 23 10:30AM

2 Data Files not displayed due to invalid file names

Quick Predictions: Only Display Summaries
View the summaries quickly by not building all result files

+ Override Defaults

Prediction Name

diabetes test (Prediction)

Predict⌂ Cancel

4. The Emcien real-time prediction engine will then predict the outcome of the Diabetes column for each transaction. The predictions are made using the predictive patterns from Emcien's initial analysis.

Prediction Details Diabetes Test (Prediction)

Working:  Lookups Built

- When finished, the Predictions Details page is displayed. The Predictions Details page contains a report summarizing the prediction results.

Prediction Overview diabetes test (Prediction)

2 Outcome Items and 73 Predictions

#	Item ¹	Captured	TP	FP	P%	TN	FN	N%	Predicted	Actual	
1	No Diabetes	75.0%	33	6	84.6%	23	11	67.7%	39	44	
2	Yes Diabetes	79.3%	23	11	67.7%	33	6	84.6%	34	29	

Hover over the headers and the software will explain each metric. The table below contains summary of those explanations.

Accuracy

Percent of the time a transaction is correctly classified as either having or not having that specific Outcome Item.

$(\text{True Pos} + \text{True Neg}) / \text{Total Predictions}$

True Positive (True Pos)

Outcome Item correctly predicted.

For example, Outcome Item is 'Sick' and sick people were correctly identified as sick.

False Positive (False Pos)	<p>Outcome Item is not correctly predicted.</p> <p>For example, Outcome Item is 'Sick' and healthy people were incorrectly identified as sick.</p>
Positive Accuracy (Pos Acc)	<p>Percentage of transactions that were correctly classified as having this Outcome Item.</p> <p>$\text{True Pos} / (\text{True Pos} + \text{False Pos})$</p>
True Negative (True Neg)	<p>Outcome Item was not predicted, and should not have been.</p> <p>For example, Outcome Item is 'Sick' and healthy people were not identified as sick.</p>
False Negative (False Neg)	<p>Outcome Item was not predicted, but should have been.</p> <p>For example, Outcome Item is 'Sick' and sick people were identified as healthy.</p>
Negative Accuracy (Neg Acc)	<p>Percentage of transactions that were not classified as having this Outcome Item and did not have the Outcome Item.</p> <p>$\text{True Neg} / (\text{True Neg} + \text{False Neg})$</p>
Transaction numbers (Trans #)	Total number of transactions in the Prediction file that contained this Outcome Item.
Predicted numbers (Prdct #)	Total number of predictions made for this Outcome Item.

When the test data does not include the known outcome values, a much simpler table will be shown that describes how many of each prediction were made. Without the outcomes, however there will be no information about the accuracy.

Prediction Overview **diabetes test (Prediction)**

2 Outcome Items and 73 Predictions

#	Item ¹	Predicted	
1	No Diabetes	39	
2	Yes Diabetes	34	

The summary below is a series of links to the real-time prediction engine's output files. When using the real-time prediction engine on streaming data, these files drive a downstream system or save to a database for further action.

Prediction Overview

diabetes test (Prediction)

Prediction Name

Predicting: **Diabetes**

[Predictions with Top Triggers \(CSV\)](#)
Describes top reasons for each prediction

[Prediction with Full Details \(TXT\)](#)
Top reasons for all considered outcomes

[Summary Information](#)
Metrics including true positives/negatives, etc.

The available files include:

Results.csv The row by row predictions made along with the top 5 reasons why. Reasons are the rules (or clusters) from Emcien that matched on the new test row being predicted.

Results.txt Intended for those that want to understand the details of how the real-time prediction engine each prediction it made.

The Results.csv file provides the predictions and the reasons for each prediction. The following is a breakdown of each column within the Results.csv file:

basket_num	predicted_out	actual_out	confidence	Winning Sco	Winner Matc	Reason 1	Reason 2	Reason 3
0	Class::1	Class::1	0.856	0.454	34	(0.034 = 0.47 x 0.77) Age (years)::[28.0_to_81.0] ==> Class::1	(0.032 = 0.53 (0.026 = 0.	
1	Class::0	Class::0	0.916	0.889	53	(0.045 = 0.91 x 0.4) Plasma glucose concentration::[0.0_to_102.0] ==> Class::0	(0.044 = 0.93 (0.043 = 0.	
2	Class::1	Class::1	0.892	0.536	32	(0.051 = 0.82 x 0.31) Plasma glucose concentration::[161.0_to_199.0] ==> Class::1	(0.04 = 0.83 ; (0.035 = 0.	
3	Class::0	Class::0	1	1.268	73	(0.045 = 0.91 x 0.4) Plasma glucose concentration::[0.0_to_102.0] ==> Class::0	(0.044 = 0.93 (0.043 = 0.	
4	Class::1	Class::1	0.848	0.411	33	(0.034 = 0.67 x 0.31) 2-Hour serum insulin (mu U/ml)::[122.0_to_846.0] Age (years)::[28.0_to_81.0] ==> Class::1	(0.034 = 0.47 (0.021 = 0.	
5	Class::0	Class::0	0.605	0.261	20	(0.029 = 0.74 x 0.74) Number of times pregnant::[0.0_to_7.0] Diabetes pedigree function::[0.078_to_0.711] ==> Class::0	(0.022 = 0.71 (0.019 = 0.	
6	Class::0	Class::1	0.939	0.53	24	(0.045 = 0.91 x 0.4) Plasma glucose concentration::[0.0_to_102.0] ==> Class::0	(0.044 = 0.93 (0.043 = 0.	

Note that the rules are presented in order of predictive strength and that all of the rules collectively are needed to maximize the software's predictive capabilities.

Basket number The identification which matches that prediction to a specific transaction in the data set being tested.

Predicted outcome The outcome predicted by the real-time prediction engine for that transaction.

Actual outcome The actual outcome as it appears in the test data set.

Confidence	A metric used by the real-time prediction engine to determine how well the matching rules described the tested data. Primarily an internal metric for the engine, it is not the equivalent of statistical confidence and not intended to be used as a statistical metric.
Winning Score	Highest score, based on the probability and frequency for all the rules that matched the row. For details on the other scores for an outcome see the Results.txt file. This score is primarily an internal mechanism for the real-time prediction engine to identify the best prediction.
Winning matches	The number of rules that matched that transaction.

Reasons

The top five rules that provided the best score to make the prediction. Each reason has one or more items that together had a high conditional probability for the predicted outcome value.

(Reason Score: 0.08 = Cond Prob: 0.92 x Coverage:

0.3) Plasma glucose concentration::[0.0_to_96.0]

The format of a reason includes:

- Scoring metrics within parentheses
- An item or cluster of items of the form

Category::Item values

The example above has two items together:

- Number of times pregnant::[0.0_to_7.0]
- Diabetes pedigree function::[0.078_to_0.711]
- The predicted value after the symbol ==>

Predicting Truck Brake System Failure

This article outlines how Emcien can be used to analyze vehicle warranty data and even identify trends that can be used to preempt costly fleet repairs and recalls.

Note: This article uses the TX Truck Warranty data set.

[tx_truck_2014_11_06.wide.csv](#)

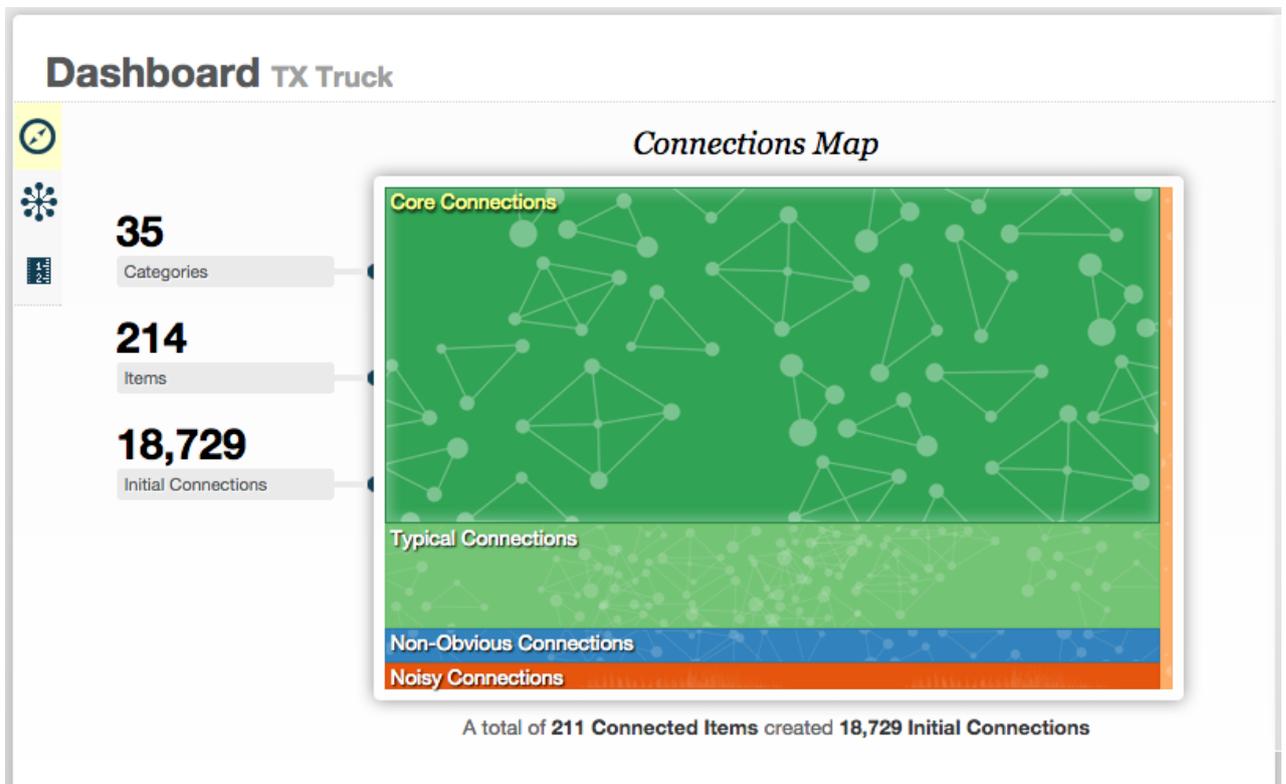
1. Using your preferred Internet Browser, navigate to the Emcien sign in page:
For Sandbox users, navigate to <https://patterns-sandbox.emcien.com/>
For production users, navigate to <http://patterns.emcien.com/>
2. On the Home page click TX Truck.

The screenshot shows the 'Home' dashboard with a table of data analysis items. The 'TX Truck' item is highlighted with a red box. The table has columns for 'Data Analysis', 'Analysis Created', 'Data Ends', 'Sort', 'Transactions', and 'Items'.

Data Analysis	Analysis Created	Data Ends	Sort	Transactions	Items
FDIC ROA	October 27 22 Days Ago	No Dates	Created	6,887	6,032
Retail Sales	October 10 A Month Ago	December 30 2 Years Ago		1,052,527	35,020
TX Truck	October 9 A Month Ago	No Dates		19,829	214
Airline Delay	October 6 A Month Ago	No Dates		509,519	5,841

The Dashboard page for the TX Truck data set is shown. This page gives you a quick overview of the general shape of the data set, including connected and disconnected values, total categories, and the total number of connections.

3. Click on the Categories number to see the complete list of categories in the data.



4. Scroll down the categories page until you see the category you are looking for. In this case we are investigating brake failures and will have to navigate to the second page of categories to find the Brake System Failure category.

Category Search TX Truck

[Download CSV](#)

Your search matched **35** of the **35** total Categories.

#	Category	Items	Clusters	Sub	Trans ¹	Trms #	Trend
21	Interior	4	129,181	0	19,733	29.5	N/A
22	Seat - Driver	5	90,073	0	18,097	29.7	N/A
23	Electrical System Failure	9	15,164	0	17,996	29.6	N/A
24	Brake System Failure	9	14,924	0	17,898	29.6	N/A
25	Fuel System Failure	9	15,027	0	17,897	29.6	N/A
26	Driveline Failure	9	15,035	0	17,866	29.6	N/A
27	Air Conditioner	2	89,500	0	15,917	29.7	N/A

5. At the Category Details page we see the different bands for brake system failure. The lowest range contains 2,580 brake failures before a warranty might have run out. Because this analysis contains 35 different categories for each entry, we can click on the item 51-70K Brake System Failure and see the Item Detail page for that data point.

Brake System Failure contains:

[Download CSV](#)

#	Item	Freq ¹	Int Cx	Rnk Cx	Clstrs	Trend
1	51-70K	2,580	200	196	2,072	N/A
2	71-90K	1,980	200	194	1,675	N/A
3	1-30K	1,958	202	191	1,667	N/A
4	111-130K	1,956	202	192	1,607	N/A
5	31-50K	1,928	200	194	1,644	N/A
6	91-110K	1,908	199	193	1,599	N/A
7	151-170K	1,897	201	197	1,548	N/A
8	171-190K	1,853	200	192	1,614	N/A
9	131-150K	1,838	199	196	1,498	N/A

6. The Item Detail is listed in the order of connection strength, but this view may require some interpretation. For example, the first item listed is the North region. Location may have some affect on how the brakes wear, but features like standard air conditioning or paint may not.

Item Detail 51-70K

Ranked Connections | [Clusters](#) | [Initial Connections](#)
Viewing 2-item clusters after ranking

[Download Connections CSV](#)

#	Items	Strength ↓	Freq	Grd Prob	Trend
1	North	S	1,064	41.2%	N/A
2	F22.5X8.5 (Paint)	S	2,419	93.8%	N/A
3	R22.5X8.5 (Paint)	S	2,390	92.6%	N/A
4	EX-Ceramic	S	2,219	86.0%	N/A
5	170A-12v	S	2,063	80.0%	N/A
6	STD-Slate	S	1,939	75.2%	N/A
7	65gal RH	S	1,744	67.6%	N/A
8	STD-Hydraulic	S	1,015	39.3%	N/A
9	Elec - grill mount	S	1,526	59.1%	N/A
10	TX-2300	S	1,638	63.5%	N/A

Item Overview

51-70K
ID

Brake System Failure
Category [View](#) →

2,072 Clusters [View](#) → **2,580** Transactions

200 Initial Connections **196** Ranked Connections

0 Substitutions

Transactions containing this Item

Avg Qty per Trans **1.0**

Avg Trans Size **29.1**

This Item has 694 Predictors
Containing probabilities as high as **58%**
[View Predictors](#) →

[Connections Graph](#) [Explore Graph](#) →

7. Click View Predictors on the right hand column to see the predictors of 51-70K Brake System Failure. The Predictors page shows items ranked by a combination of strength and frequency, but by sorting by probability we can see the combination of items that are most likely to occur with the 51-70K Brake System Failure item.

Predictors for 51-70K

[Download CSV](#)

Displaying 694 Predictive Clusters

#	Predictor Cluster	Probability	Item	Freq
1	North + STD-Hydraulic + R15000-STD	58%	51-70K	476
2	MD-330 + North + STD-Hydraulic	55%	51-70K	226
3	North + STD-Hydraulic + F8000-STD	54%	51-70K	434
4	North + Company + STD-Hydraulic	52%	51-70K	178
5	North + STD-Hydraulic + Elec - grill mount	50%	51-70K	527
6	North + STD-Hydraulic	50%	51-70K	633
7	Air_Ride XT + North	42%	51-70K	170
8	MD-300 + North	40%	51-70K	111
9	North + F8000-STD + R15000-STD	39%	51-70K	429

Emcien’s analysis engine automatically identifies these predictive rules and allows you to navigate to the item that you want to predict. Through Emcien APIs the analysis can be automated from ingestion, to transformation, to automating service programming to prevent larger costs down the road.

Predicting Titanic Survivors

Use Case: Predicting Titanic Survivors

This article outlines how Emcien can predict the outcome of a situation given historical data. The objective of this use case will be to discover patterns in data from the sinking of the Titanic that would allow us to accurately predict whether a passenger survived or not, given the passenger's demographic data.

Align your data with a use case

Note: This procedure uses the Titanic example data set. Your data should be similar to the example data, with columns for attribute or demographic data for a population. To download, right click the link and click "Save Link As"

titanic-upload.wide.csv

Data Prep

Optimize your data.

The wide format is the more universal and most commonly used Emcien format. You can use the wide format for multi-dimensional data, such as demographics or configurable products. In the wide format, each transaction is identified by a single row of data.

Numerical Data is Banded

Each column header is a unique feature

Parents Aboard	Deck Level	Embarked	Home Destination	Class	Survival	Age	Fare	Cabin
0	B	Southampton	St Louis MO	First	Survived	22.67799 to 29.881214	75.25 to 512.33	0.0 to 21.0
2	C	Southampton	Montreal PQ - Chesterville ON	First	Survived	0.1667 to 8.271542	75.25 to 512.33	21.0 to 35.0
2	C	Southampton	Montreal PQ - Chesterville ON	First	Died	0.1667 to 8.271542	75.25 to 512.33	21.0 to 35.0
2	C	Southampton	Montreal PQ - Chesterville ON	First	Died	29.881214 to 37.084438	75.25 to 512.33	21.0 to 35.0

Each row represents an order

In our example, the columns are attributes (e.g. age, home town) for passengers on the RMS Titanic. Emcien works well with any kinds of demographics data, and is able to provide patterns that lead to discrete outcomes. If your data does not look similar to the wide format data here, don't worry. Emcien Patterns also works with 'receipt' type data. For more information about data preparation and different data types supported, see our [Data Prep Guide](#).

Note: Emcien's engine can also automatically band numeric data given a numeric data source. For more information, see the link above.

Uploading Data

Upload your data for analysis.

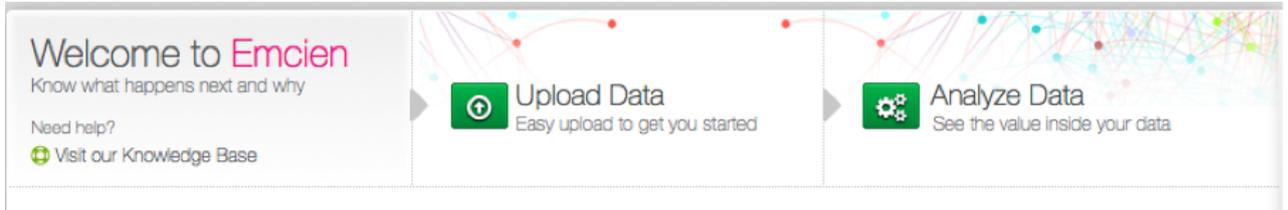
This article covers how to upload data files to Emcien using [File Transfer Protocol \(FTP\)](#).

Make sure you've prepared your data before uploading to Emcien. Check out the [Preparing Your Data](#) article for more information. You can also email us at support@emcien.com for help getting started.

Upload using Drag and Drop

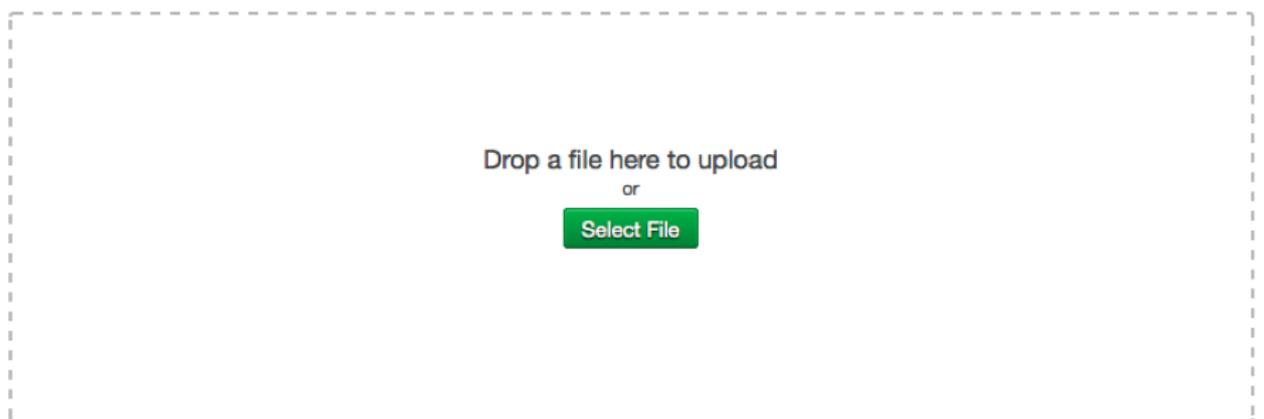
In Emcien version 2.14 you can now drag and drop files into the application for processing. If you are using a previous version you can use one of the other methods listed below.

Simply open the Emcien application and select "Upload Data"



Select the file that you wish to process, then drag and drop it onto the Upload Data screen

Upload Data Services



Once it has finished uploading you can then click on "Analyze This Data" to

process the file.

✔ Upload Complete!

[Analyze This Data](#) or [Upload More Data](#)

Uploading Data to the Emcien Cloud

You can upload data files to Emcien using [SSH File Transfer Protocol \(SFTP\)](#) and your preferred FTP client, such as [FileZilla](#) or [Cyberduck](#).

To connect to Emcien using your FTP client, use the following credentials:

Host: feeds.emcien.com

Username: {Your Emcien Feeds server Username}

Password: {Your Emcien Feeds server Password}

Uploading Data on Mac or Linux

Launch your preferred FTP client. If you do not have a preferred FTP client, we recommend [Cyberduck](#).

For this article, we use screenshots of Cyberduck.

Click Open Connection.



Enter the below credentials and click Connect.

The screenshot shows an FTP client window with a toolbar at the top containing icons for Open Connection, Quick Connect, Action, Refresh, Edit, and Disconnect. Below the toolbar is a navigation bar with a search box and a directory path: /srv/clients/Emcien_Marketing. The main area displays a table of files with columns for Filename, Size, and Modified.

Filename	Size	Modified
banded_companies.wide.csv	10.5 MB	10/24/14, 5:29 PM
banded_offenders_table.wide.csv	2.6 MB	10/29/14, 2:03 PM
banded_titanic-table.wide.csv	147.3 KB	12/9/14, 9:21 AM
banded_titanic.wide.csv	122.3 KB	11/7/14, 3:56 PM
banded_WorldPoverty.csv	1.2 MB	10/22/14, 2:10 PM

Your FTP client will then display the files on your server.

Server: Your Emcien URL. This URL was set by your IT resource during installation.

Username

Password

Drag and drop your data file to the server files in your FTP application.

On the Emcien home page click Analyze Data. You'll see the data files on this screen.

Uploading Data on Windows

Click the Start button . The Start menu is displayed.

On the Start menu, click Computer. The Computer folder is displayed.

Right-click anywhere in the folder and click Add a network location.

The Add Network Connection wizard is displayed. Click Next.

On the next tab, select Choose a custom network location. Click Next.

On the next tab, enter your Emcien URL.

This URL was set by your IT resource during installation.
Ensure you include *FTP://* at the beginning of the URL.

Then click Next.

Clear the Log on anonymously checkbox.

Type *data* in the username field. Click Next.

On the next tab, name your shortcut *Emcien* by typing it in the Type a name for this network location field. Click Next.

On the next tab, click Finish.

Type *feeds1* in the password field. Check the Save password checkbox so you can connect directly in the future.

Click Log On. You can now drag data files into this folder.

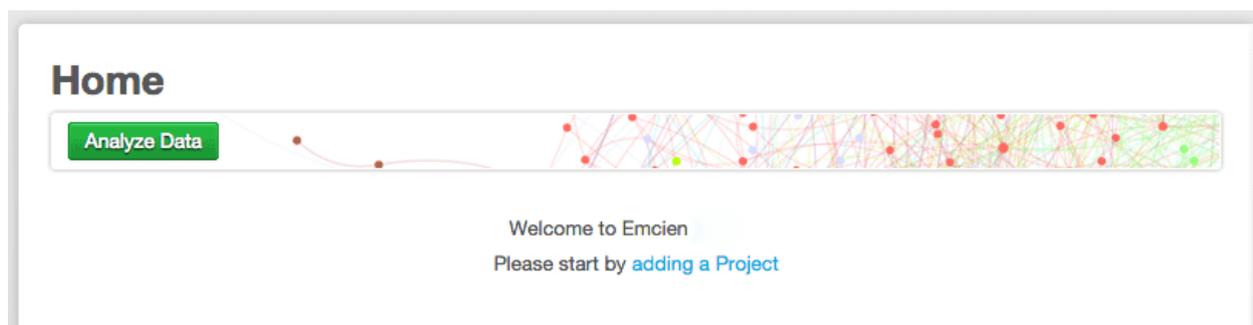
On the Emcien home page, click Analyze Data. You'll see the data files on this screen.

Analysis

Begin your analysis.

1. Using your preferred Internet Browser, navigate to the Emcien Sign In page:

- local VM users, go to: <http://localhost:5115>
- For cloud users, go to: <http://patterns.emcien.com/>



The Analyze Data button will bring up all of your uploaded data. Select your data set.

New Analysis Nick Admin

Data

Filter by Name

Name	File Size	Modified Time
sales-data.wide.csv	4.9 MB	Apr 6 5:16PM
<input checked="" type="checkbox"/> titanic-upload.wide.csv	140 KB	May 7 9:42AM

Projects

Reports in a project will display on home page

[+ Create a New Project](#)

Click on Predictions, and type in the predictive category. For this data set we will be predicting the Category "Survival".

- Predictions

Prediction Category Name

Survival

Prediction Category may have 20 items or less

Item Weighted Predictions
Allow item_price to weight Prediction Rules

+ Override Defaults

Report Name

Titanic Analysis

Start Analysis

Once your project is organized, click Start Analysis.

The load screen will take you through each stage of the analysis.

Sales Data V1 Log Emcien Marketing Admin

Log: [Summary](#) | [Expanded](#) Run Time: 3m 49s

#	Message	Time
37	--> Building Engine::Clusters::Results as clusters.dat	14s
36	--> Building Engine::Trends::TrendData as item_trends.dat	14s
35	--> Building Engine::RetailSubs::Results as retail_subs.dat	14s
34	--> Building Engine::Graph::DegreesByPeriod as degrees_by_period.dat	14s
33	--> Building Engine::Graph::Lol as graph.dat	12s
32	--> Building Engine::Index::Duplicates as duplicates.dat	12s
31	--> Building Engine::Index::Transaction as index.dat	12s
30	--> Building Engine::Trends::CatTrend as cat_trends.dat	12s
29	--> Building Engine::Index::CatTransaction as catindex.dat	12s
28	--> Building Engine::Input::CatData as catdata.dat	12s

When the analysis is complete you will hear a chime and the View Analysis button will be highlighted in green. Just click the button to see your results.

Results

See your results.

Begin by clicking “View Analysis”

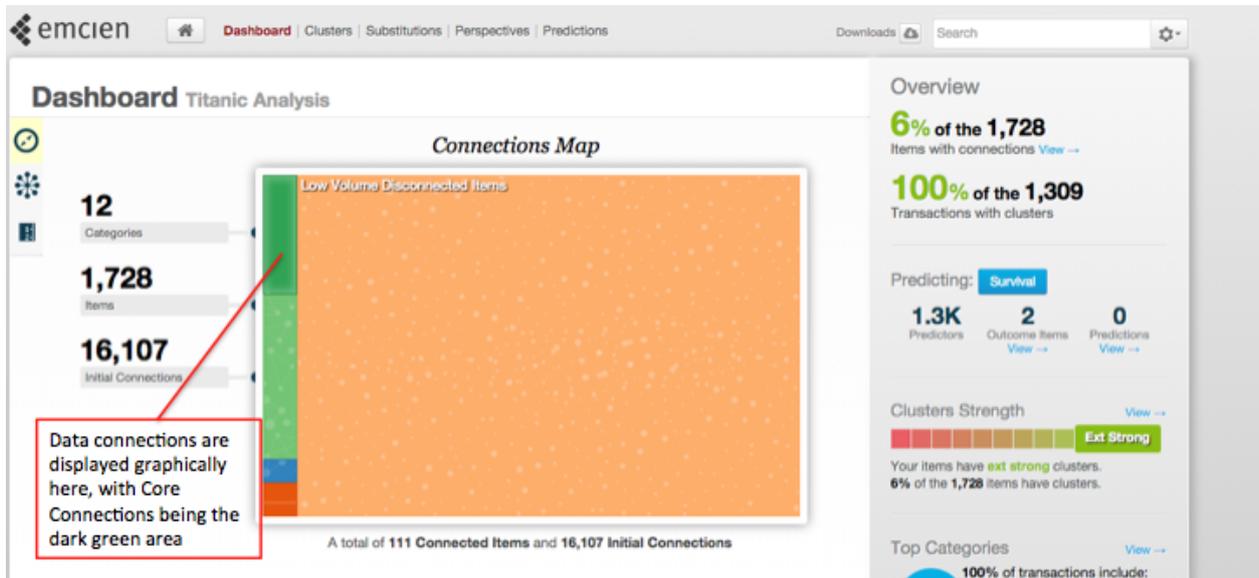
Titanic Analysis Nick Admin

Log: [Summary](#) | [Expanded](#) Run Time: 1m 9s

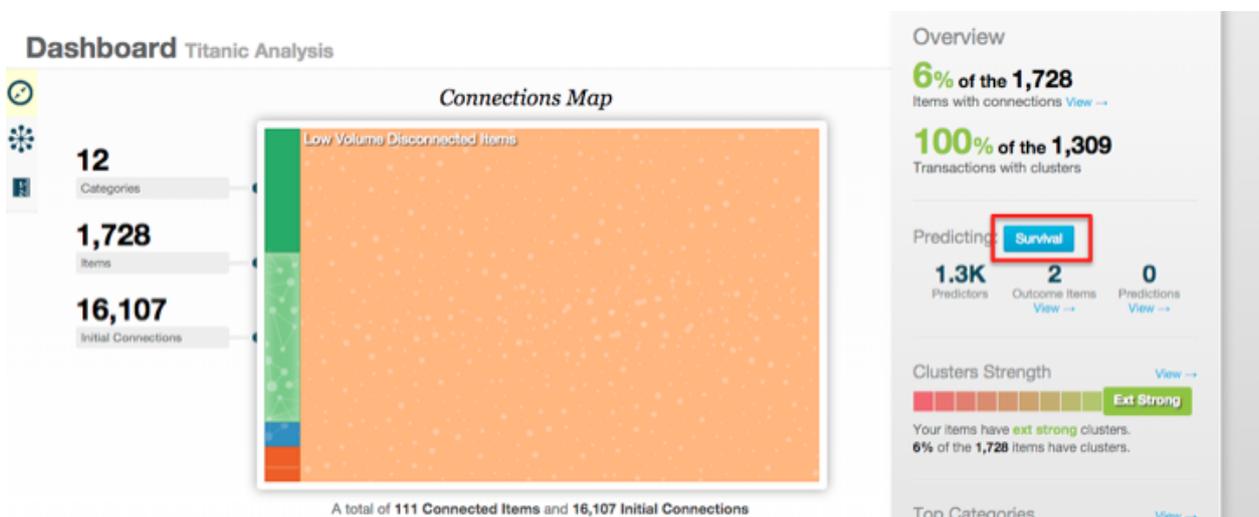
#	Message	Time
106	Queueing Backup Job	1m 9s
105	-----> Import successful for report 55895943	1m 9s

The Dashboard page for the Titanic analysis data set is displayed below. The interactive graphic on the home screen is the Connections Map, a representation

of the analyzed data displayed by how demographics were correlated (or connected) together.



Each color represents the connections identified as relevant patterns in the data set. Mousing over each section will display examples of the corresponding clusters (or groupings of items occurring together). These include Core and Typical Connections (attribute data frequently found together), and Low Volume Disconnected Items (passenger data with no patterns associated with them). Notice at the right of the screen the link to our predictions.



Clicking on the “Survival” button will bring us to the outcomes of the prediction

category “Survival”.

The screenshot shows the 'Category Details' page for the 'Survival' category. On the left, a table lists items: 'Died Survival' (rank 1) and 'Survived Survival' (rank 2). The table columns are '#', 'Item', 'Freq', 'Int Cx', 'Rnk Cx', 'Clstrs', and 'Trend'. A 'Download CSV' button is visible. On the right, a 'Category Overview' sidebar shows: 'Survival' Name, 1,309 Transactions, 2 Items, and 7,976 Clusters. Below this, it states 'Transactions in this Category' with 'Avg Trans Size 9.8' and 'Trans with Cnx 1,309'.

The category details page shows a list of the different items in the category predicted. In this case there are two outcomes: Died or Survived. To see the patterns associated with Survival, click “Survived”

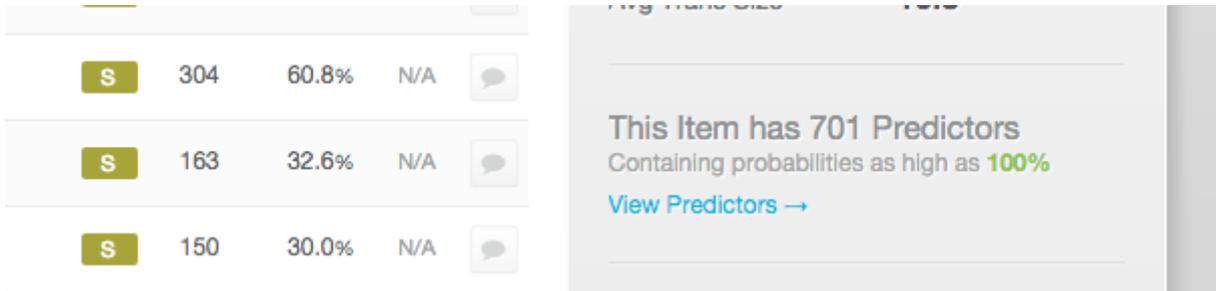
The screenshot shows the 'Item Detail' page for the 'Survived' item. The main table, 'Ranked Connections', lists items connected to 'Survived'. The first item is 'female sex' with a strength of 8, frequency of 339, and a conditional probability of 67.8%. A red box highlights a 'Tell Me' button next to this row. Other items include 'First Class', '0 parents aboard', '0 siblings aboard', '10.5 to 52.55 fare', and 'Southampton embarked'. On the right, an 'Item Overview' sidebar shows: 'Survived' ID, 'Survival' Category, 4,854 Clusters, 500 Transactions, 733 Initial Connections, and 1 Substitution. It also shows 'Transactions containing this Item' with 'Avg Qty per Trans 1.0' and 'Avg Trans Size 10.3'.

The Item Details page shows the strongest patterns related to the item “Survived”. Shown here are the other items (and groups of items) in the data set that connected most strongly to the item we are viewing. To get a better sense of what this page is telling us, click on the “Tell Me” Button to the right of any line.

The screenshot shows a 'Tell Me' tooltip. The text reads: 'For all transactions containing Survived Survival, 67.8% also contain female sex.' Below this, it says: 'This cluster is in 339 of the anchor item's 500 transactions. View Details →'. The tooltip is overlaid on the 'Item Detail' page.

To find the combination of attributes that would have the strongest ability to

predict our item (Survival), click on the “View Predictors” link on the right side of the screen.



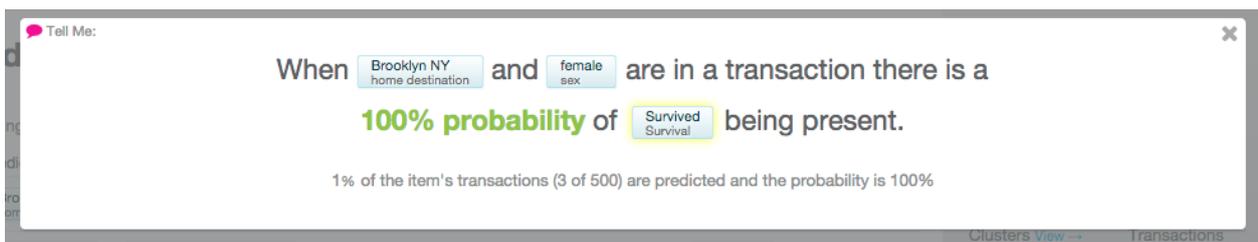
The predictions screen is the combinations of attributes (or set of rules) that leads to the outcome item.

Predictors for Survived Survival Download CSV

Displaying **701** Predictive Clusters

#	Predictor Cluster	Probability ¹	Item	Size ²	Freq
1	Brooklyn NY home destination + female sex	100% →	Survived Survival	3	3
2	3 siblings aboard + 22.67799 to 29.881214 age	100% →	Survived Survival	3	4
3	F Deck Level + 1 parents aboard	100% →	Survived Survival	3	6
4	F Deck Level + female sex	100% →	Survived Survival	3	8
5	E Deck Level + 90.0 to 148.0 Cabin	100% →	Survived Survival	3	5

The way to interpret these rules is “if the items on the left occur together, then the item on the right will have the given probability of occurring”. Click the Tell Me for a better understanding of a row.

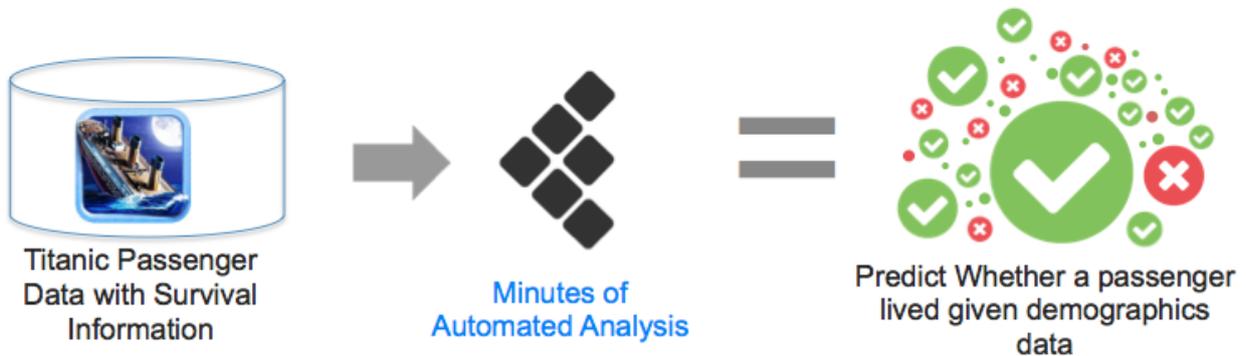


Emcien’s rules and patterns can be easily exported through the “Download CSV” button at the top of each page, and can be used for any predictive scenario.

The screenshot shows the 'Predictors for Survived Survival' interface. At the top right, a 'Download CSV' button is highlighted with a red box. Below the title, it says 'Displaying 701 Predictive Clusters'. A table lists four predictor clusters with their respective probabilities, items, sizes, and frequencies. To the right, an 'Item Overview' sidebar shows statistics for the 'Survived' category, including 4,854 Clusters, 500 Transactions, 733 Initial Connections, and 1 Substitution.

#	Predictor Cluster	Probability	Item	Size	Freq
1	Brooklyn NY home destination + female sex	100%	Survived Survival	3	3
2	3 siblings aboard + 22.67799 to 29.661214 age	100%	Survived Survival	3	4
3	F Deck Level + 1 parents aboard	100%	Survived Survival	3	6
4	F Deck Level + female sex	100%	Survived Survival	3	8

While Emcien is useful for identifying historical patterns in predictions, it can also make real-time predictions given new information. For more information about our real-time prediction engine, see [Our Guide to Making Predictions](#).



For more information or questions about Emcien’s predictive engine, you can contact your Services team member directly or email us at hello@emcien.com and we will connect you with a services professional.

Predicting Return on Assets

This article outlines how you can find the predictors of Return on Assets (ROA) using Emcien.

Note: This article uses the FDIC ROA example data set.

Align your data with a use case

This walkthrough uses the FDIC-ROA data set. Click on the file name below to download the data.

[FDIC-ROA.wide.csv](#)

Data Prep

Optimize your data.

In order for Emcien to identify patterns of what leads to either good or bad ROA, it is important that the data being analyzed includes dependent variables, or the outcomes you want to predict.

Uploading Data

Upload your data for analysis.

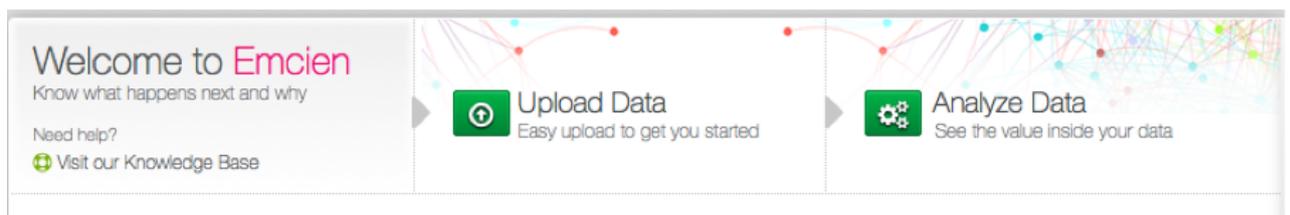
This article covers how to upload data files to Emcien using [File Transfer Protocol \(FTP\)](#).

Make sure you've prepared your data before uploading to Emcien. Check out the [Preparing Your Data](#) article for more information. You can also email us at support@emcien.com for help getting started.

Upload using Drag and Drop

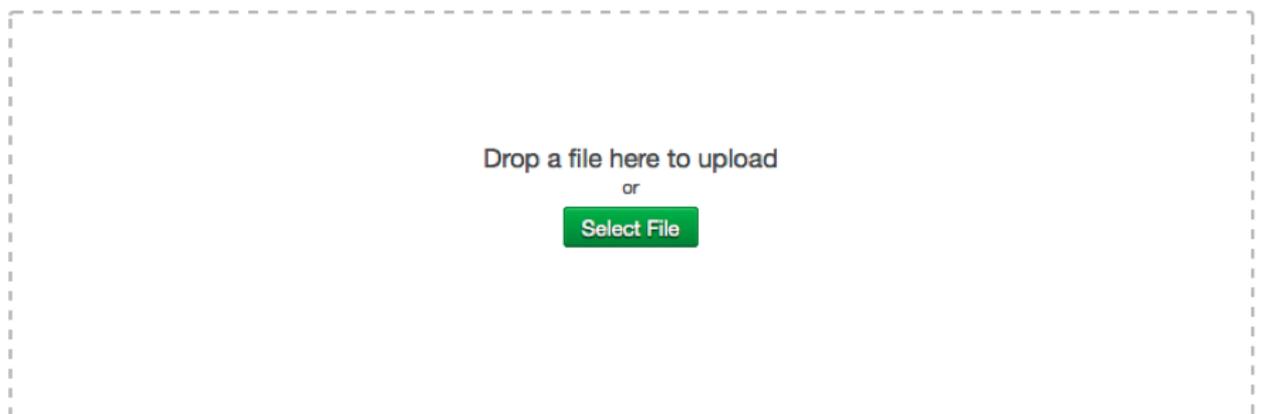
In Emcien version 2.14 you can now drag and drop files into the application for processing. If you are using a previous version you can use one of the other methods listed below.

Simply open the Emcien application and select "Upload Data"



Select the file that you wish to process, then drag and drop it onto the Upload Data screen

Upload Data Services



Once it has finished uploading you can then click on "Analyze This Data" to

✔ Upload Complete!

process the file.

[Analyze This Data](#) or [Upload More Data](#)

Uploading Data to the Emcien Cloud

You can upload data files to Emcien using [SSH File Transfer Protocol \(SFTP\)](#) and your preferred FTP client, such as [FileZilla](#) or [Cyberduck](#).

To connect to Emcien using your FTP client, use the following credentials:

Host: feeds.emcien.com

Username: {Your Emcien Feeds server Username}

Password: {Your Emcien Feeds server Password}

Uploading Data on Mac or Linux

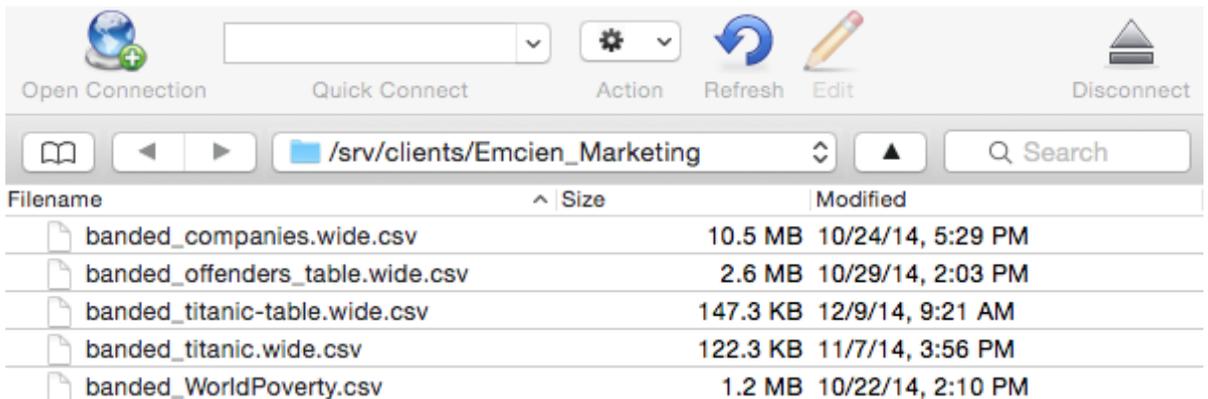
Launch your preferred FTP client. If you do not have a preferred FTP client, we recommend [Cyberduck](#).

For this article, we use screenshots of Cyberduck.

Click Open Connection.



Enter the below credentials and click Connect.



Your FTP client will then display the files on your server.

Server: Your Emcien URL. This URL was set by your IT resource during installation.

Username

Password

Drag and drop your data file to the server files in your FTP application.

On the Emcien home page click Analyze Data. You'll see the data files on this screen.

Uploading Data on Windows

Click the Start button . The Start menu is displayed.

On the Start menu, click Computer. The Computer folder is displayed.

Right-click anywhere in the folder and click Add a network location.

The Add Network Connection wizard is displayed. Click Next.

On the next tab, select Choose a custom network location. Click Next.

On the next tab, enter your Emcien URL.

This URL was set by your IT resource during installation.
Ensure you include *FTP://* at the beginning of the URL.

Then click Next.

Clear the Log on anonymously checkbox.

Type *data* in the username field. Click Next.

On the next tab, name your shortcut *Emcien* by typing it in the Type a name for this network location field. Click Next.

On the next tab, click Finish.

Type *feeds1* in the password field. Check the Save password checkbox so you can connect directly in the future.

Click Log On. You can now drag data files into this folder.

On the Emcien home page, click Analyze Data. You'll see the data files on this screen.

Analysis

Begin your analysis.

Click on the Analyze Data button

Overview Demo Admin

Analyze Data

#	Name	State	Project	▼ Created	Import	Menu
None						

Users

#	Email	Last Sign On
1	demo@emcien.com ID: 11	A Few Seconds Ago

Projects

#	Name	Reports
1	Demo ID: 3	0

You will then need to select the data file that you uploaded or "Drag and Dropped" for analysis.

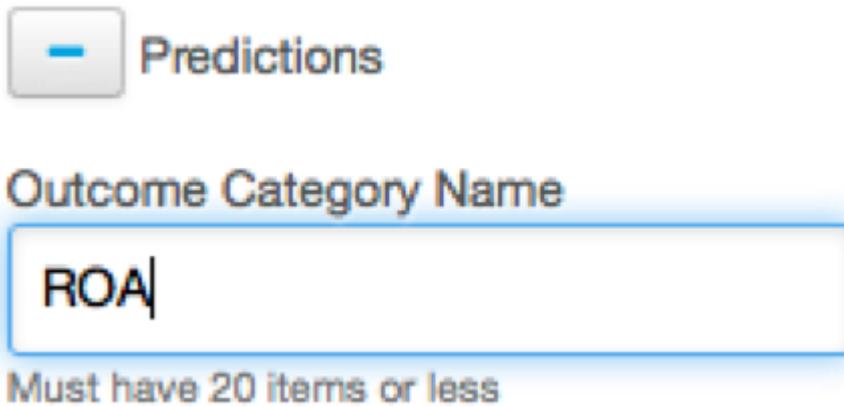
New Analysis Demo Admin

Source Data

Filter by Name

Name	File Size	Modified Time
✓ fdic_roa.wide.csv	1.4 MB	Jun 10 3:09PM
ACME_truck_parts.wide.csv	2.8 MB	Jun 10 2:30PM
retail_sales_2014_12_02.receipt.csv	86 MB	Jun 10 1:44PM
Airlines-Delay-Data.wide.csv	33 MB	Jun 4 10:17AM
ACME-Auto.csv	170 MB	Jun 2 4:38PM
ACMEOfficeSupplyInputFile.csv.gz	390 MB	Jun 2 2:27PM
BankData3Q2013-noemp.wide.csv	1.4 MB	Jun 2 2:00PM
ACMEAutoParts_initial_NoSubCategory.csv	190 MB	Jun 2 1:59PM
ACMEHardwareMediumCleanedupPricing.csv.gz	220 MB	Jun 2 1:58PM
ACME-Office-1M-Load.csv	370 MB	May 28 3:55PM

Next you will enter the category that we want to have Emcien Predict: ROA



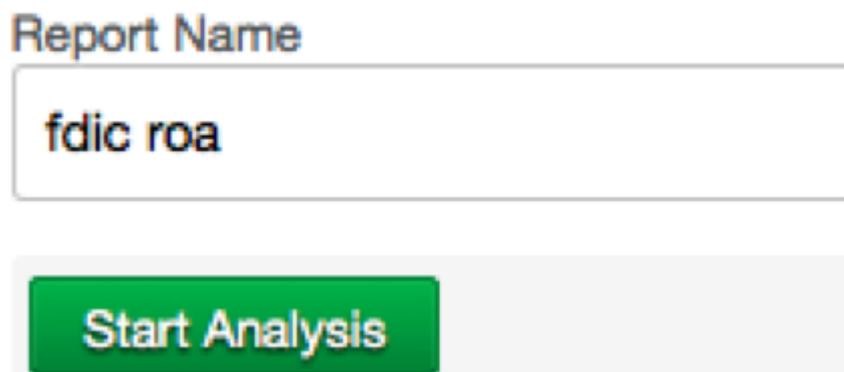
— Predictions

Outcome Category Name

ROA

Must have 20 items or less

Enter a report name or except the default created and click "Start Analysis"



Report Name

fdic roa

Start Analysis

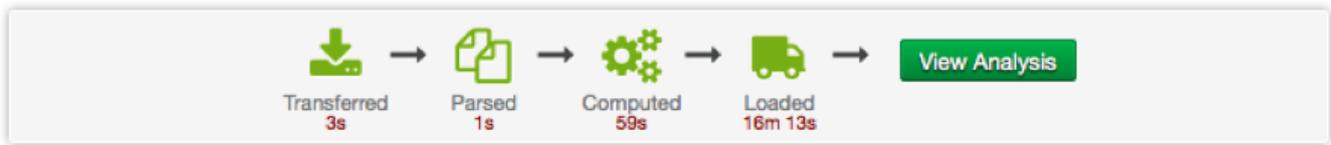
The application will transfer the data into Emcien, parse it for processing, compute all of the connections and predictions, and then load all of the data for the analysis. Go to the next step in the process to "View Analysis"

Results

See your results.

Click on "View Analysis"

fdic roa Demo Admin



Log: [Summary](#) | [Expanded](#)

Run Time: **17m 24s**

#	Message	Time
105	Queueing Backup Job	17m 24s
104	-----> Import successful for report 36479427	17m 24s
103	Report State: Report-ready	17m 23s

The Dashboard page for the FDIC ROA data set is displayed. On the Dashboard page, you can view the number of connected data values, the number of transactions analyzed, and more information.

Dashboard fdic roa

- 27
Categories
- 6,032
Items
- 170,503
Initial Connections

Connections Map

A total of **226 Connected Items** and **170,503 Initial Connections**

Overview

4% of the 6,032
Items with connections [View →](#)

100% of the 6,887
Transactions with clusters

Predicting: **ROA**

6.6K Predictive Rules
6 Outcome Items [View →](#)
0 Predictions [View →](#)

Clusters Strength [View →](#)

Ext Strong

Your items have **ext strong** clusters.
4% of the 6,032 items have clusters.

Top Categories [View →](#)

100% of transactions include:

- Bank**
- State**
- Charter**

27 Total Categories

Details

Core Connections

Represents higher-frequency clusters with high strength and correlation. Core represents the heart of the patterns found. [View →](#)

Example Connection

MA State

Very Low Ag/Loans

Freq: 150 Conditional Probability: 94.9% Strength: Fairly Strong

NJ State

Very Low Ag/Loans

Freq: 101 Conditional Probability: 95.3% Strength: Fairly Strong

On the Dashboard page, you will see Predicting: "ROA".

Predicting: **ROA**

6.6K
Predictive Rules

6
Outcome Items
[View →](#)

0
Predictions
[View →](#)

Click on View under the Outcome Items.

Category Details **ROA**

ROA contains:

[Download CSV](#)

#	Item	Freq ¹	Int Cx	Rnk Cx	Clstrs	Trend
1	Average ROA	3,114	2,943	154	69,478	N/A
2	High ROA	1,731	1,666	158	15,569	N/A
3	Low ROA	893	1,033	162	2,359	N/A
4	Poor ROA	641	783	145	1,877	N/A
5	Superior ROA	431	581	139	1,287	N/A
6	Very Poor ROA	59	198	91	425	N/A

The Category Details page for the ROA category is displayed. This page displays all the items associated with ROA.

In our example we are interested in what Predicts a "Low" ROA.

Click on "Low" to display the Item detail.

You will see that there are 639 Predictive Rules for this item.

Item Detail Poor ROA

Ranked Connections | Clusters | Initial Connections
Viewing 2-Item clusters after

#	Items				
1	Very Inefficient Efficiency				
2	High Loan Problems				
3	Negative Asset Growth				
4	Negative Loan Growth				
5	Negative Deposit Growth				
6	Inefficient Efficiency	S	376	58.7%	N/A
7	High Capital Leverage	S	100	15.6%	N/A

Item Overview
Poor ID

641 Transactions
783 Initial Connections

containing this Item
ns **1.0**
s **26.3**

This Item has 639 Predictive Rules
Containing probabilities as high as **100%**
[View Predictive Rules ->](#)

Click on "View Predictive Rules" to display the list of those items that predict "Poor".

A Prediction Rule Search is displayed for the "Poor" item. This page displays all the predictive rules for "Poor" ROA. To learn more about the columns in the data you can hover your mouse over each for a more detailed description.

If you click on the "Tell Me" icon ,

a pop up will display more information about that specific item in the list.

Prediction Rule Search fdic roa Download CSV

Your search matched **639** of the **6,596** total prediction rules. Search criteria: Outcome Item: Item ID is exactly "Poor".

#	Freq	Prediction Rule	Probability	Outcome Item	Size	Cover	Contro
1	11	Very High Capital Leverage, Very High Loan Problems, Negative Deposit Growth	100% →	Poor ROA	3	1.72%	0.32
2	10	Very High Capital Leverage, Negative Deposit Growth, Negative Loan Growth, High CDs	100% →	Poor ROA	4	1.56%	0.32
7	155	Very Inefficient Efficiency, Average Yield On Loans	82.0% →	Poor ROA	2	24.2%	0.061
8	129	Very Inefficient Efficiency, Average Capital Leverage	86.6% →	Poor ROA	2	20.1%	0.054
9	124	Very Inefficient Efficiency, Average Non-Int Deposits	83.2% →	Poor ROA	2	19.3%	0.049
10	122	Very Inefficient Efficiency, Average Deposit Leverage	80.8% →	Poor ROA	2	19.0%	0.047

Tell Me:

if **Very High Capital Leverage**, **Negative Deposit Growth**, **Negative Loan Growth** and **High CDs** are in a transaction there is a **100% probability of Poor ROA** being present.

1.56% of the outcome item's transactions (10 of 641) are predicted

Based on the data that was processed the system can predict that if the Capital Leverage is "Very High", Deposit Growth is "Negative", Loan Growth is "Negative" and CDs are "High" there is a 100% probability of a "Poor" ROA.

Predictions can be automated through Emcien's APIs to direct other systems or software. For more on this, see:

<http://support.emcien.com/help/article/link/emcien-apis>

Correlating Network Attack Data

This article outlines how Emcien can be used to analyze network traffic to identify patterns of attack and apply those patterns to incoming network data.

Align your data with a use case

Note: This use case uses the network attack example data set. Your data should look similar to the example data, including the necessary dependent variable. In this data set there are fourteen different identified classifications of an attack. Additionally, each transaction in the data has been tagged with a classification score to prioritize attack status.

[NetworkAttack.wide.zip](#)

	A	B	C	D	E	F	G	H	I	J	K	L
1	Time	Hour	SOURCE_NAI	SIGNATURE	SIGNATURE_NAME	PRIORITY	CLASSIFICATION	SRC_IP	SRC_PORT	DST_IP	DST_PORT	
2	2011-11-08T	9	ITOC	4	TCP Timestamp is outside of PAWS window	3	UNKNOWN					25
3	2011-11-08T	9	ITOC	4	TCP Timestamp is outside of PAWS window	3	UNKNOWN					50176
4	2011-11-08T	9	ITOC	4	TCP Timestamp is outside of PAWS window	3	UNKNOWN					50176
5	2011-11-08T	9	ITOC	4	TCP Timestamp is outside of PAWS window	3	UNKNOWN					60708
6	2011-11-08T	9	ITOC	4	TCP Timestamp is outside of PAWS window	3	UNKNOWN					60708
7	2011-11-08T	9	ITOC	4	TCP Timestamp is outside of PAWS window	3	UNKNOWN					443
8	2011-11-08T	9	ITOC	4	TCP Timestamp is outside of PAWS window	3	UNKNOWN					58113
9	2011-11-08T	9	ITOC	4	TCP Timestamp is outside of PAWS window	3	UNKNOWN					443
10	2011-11-08T	9	ITOC	4	TCP Timestamp is outside of PAWS window	3	UNKNOWN					443
11	2011-11-08T	9	ITOC	4	TCP Timestamp is outside of PAWS window	3	UNKNOWN					58113
12	2011-11-08T	9	ITOC	4	TCP Timestamp is outside of PAWS window	3	UNKNOWN					58113
13	2011-11-08T	9	ITOC	4	TCP Timestamp is outside of PAWS window	3	UNKNOWN					443
14	2011-11-08T	9	ITOC	4	TCP Timestamp is outside of PAWS window	3	UNKNOWN					443
15	2011-11-08T	9	ITOC	4	TCP Timestamp is outside of PAWS window	3	UNKNOWN					58113
16	2011-11-08T	9	ITOC	4	TCP Timestamp is outside of PAWS window	3	UNKNOWN					443
17	2011-11-08T	9	ITOC	4	TCP Timestamp is outside of PAWS window	3	UNKNOWN					58113
18	2011-11-08T	9	ITOC	4	TCP Timestamp is outside of PAWS window	3	UNKNOWN					58113
19	2011-11-08T	9	ITOC	4	TCP Timestamp is outside of PAWS window	3	UNKNOWN					58113
20	2011-11-08T	9	ITOC	4	TCP Timestamp is outside of PAWS window	3	UNKNOWN					443
21	2011-11-08T	9	ITOC	4	TCP Timestamp is outside of PAWS window	3	UNKNOWN					443
22	2011-11-08T	9	ITOC	4	TCP Timestamp is outside of PAWS window	3	UNKNOWN					443
23	2011-11-08T	9	ITOC	4	TCP Timestamp is outside of PAWS window	3	UNKNOWN					443
24	2011-11-08T	9	ITOC	4	TCP Timestamp is outside of PAWS window	3	UNKNOWN					58113
25	2011-11-08T	9	ITOC	3	Data sent on stream not accepting data	3	UNKNOWN					1296
26	2011-11-08T	10	ITOC	4	TCP Timestamp is outside of PAWS window	3	UNKNOWN					25
27	2011-11-08T	10	ITOC	3	Data sent on stream not accepting data	3	UNKNOWN					51451
28	2011-11-08T	10	ITOC	3	Data sent on stream not accepting data	3	UNKNOWN					51451
29	2011-11-08T	10	ITOC	3	Data sent on stream not accepting data	3	UNKNOWN					51451
30	2011-11-08T	10	ITOC	3	Data sent on stream not accepting data	3	UNKNOWN					51451
31	2011-11-08T	10	ITOC	3	Data sent on stream not accepting data	3	UNKNOWN					51451
32	2011-11-08T	10	ITOC	3	Data sent on stream not accepting data	3	UNKNOWN					51451
33	2011-11-08T	10	ITOC	3	Data sent on stream not accepting data	3	UNKNOWN					1340
34	2011-11-08T	10	ITOC	4	TCP Timestamp is outside of PAWS window	3	UNKNOWN					443
35	2011-11-08T	10	ITOC	4	TCP Timestamp is outside of PAWS window	3	UNKNOWN					64941
36	2011-11-08T	10	ITOC	4	TCP Timestamp is outside of PAWS window	3	UNKNOWN					443
37	2011-11-08T	10	ITOC	4	TCP Timestamp is outside of PAWS window	3	UNKNOWN	10.1.60.203	443	10.1.60.203		443
38	2011-11-08T	10	ITOC	4	TCP Timestamp is outside of PAWS window	3	UNKNOWN	154.241.88.201				64941

The screenshot shows a 'CLASSIFICATION' panel with a search bar and a list of attack types. The list includes:

- (Select All)
- A Network Trojan was detected
- access to a potentially vulnerable web application
- Attempt to login by a default username and password
- Attempted Administrator Privilege Gain
- Attempted Denial of Service
- Attempted Information Leak
- Attempted User Privilege Gain
- Detection of a non-standard protocol or event
- Executable code was detected
- Misc activity
- Misc Attack
- Potentially Bad Traffic
- UNKNOWN
- Web Application Attack

If your data doesn't follow this same structure, please consult the [Preparing Your Data](#) article. For further information you can contact Emcien's support team.

Data Prep

Optimize your data for analysis.

In order for Emcien to identify the patterns that identify an attack on your network, it is important to first analyze data that includes identified attacks.

If you do not have network data with identified network intrusion attempts, it may be possible to identify those transactions and mark them as such. Adding this "derivative data" is not limited to identifying attacks or other dependent outcomes, but can include day of the week, part of day, seasonality, or any additional information that can be attached to existing data. For more on the concepts behind adding additional data, see the [Derived Data article](#).

This additional data complexity is relatively easy to add and can be very beneficial to Emcien's analysis.

Uploading Data

Upload your data for analysis.

This article covers how to upload data files to Emcien using [File Transfer Protocol \(FTP\)](#).

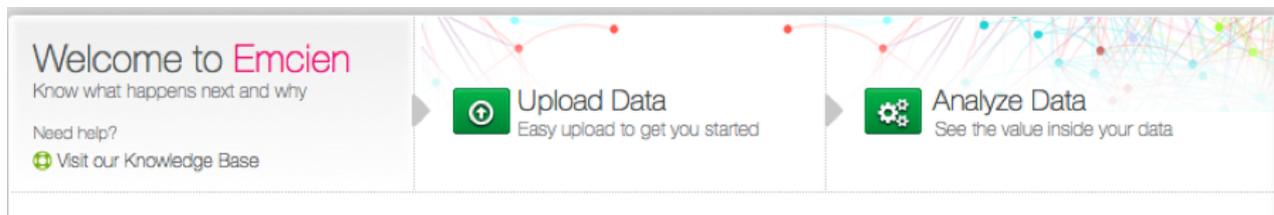
Make sure you've prepared your data before uploading to Emcien. Check out the [Preparing Your Data](#) article for more information. You can also email us at support@emcien.com for help getting started.

Upload using Drag and Drop

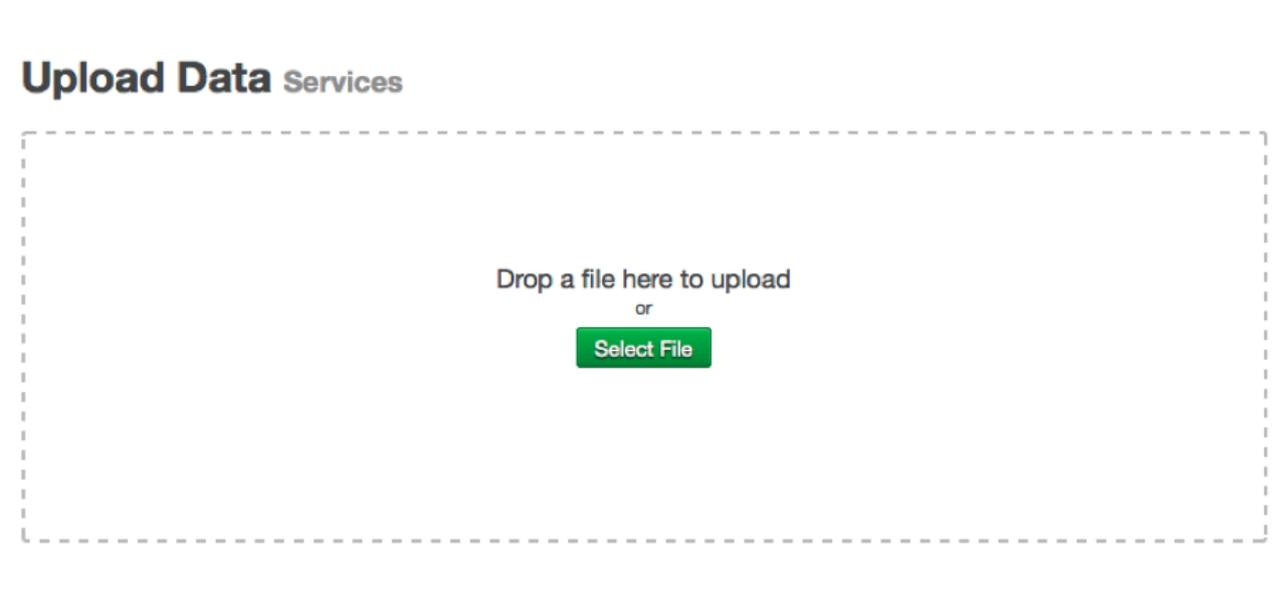
In Emcien version 2.14 you can now drag and drop files into the application

for processing. If you are using a previous version you can use one of the other methods listed below.

Simply open the Emcien application and select "Upload Data"



Select the file that you wish to process, then drag and drop it onto the Upload Data screen



Once it has finished uploading you can then click on "Analyze This Data" to

process the file.  **Upload Complete!**
 or [Upload More Data](#)

Uploading Data to the Emcien Cloud

You can upload data files to Emcien using [SSH File Transfer Protocol \(SFTP\)](#) and your preferred FTP client, such as [FileZilla](#) or [Cyberduck](#).

To connect to Emcien using your FTP client, use the following credentials:

Host: feeds.emcien.com

Username: {Your Emcien Feeds server Username}

Password: {Your Emcien Feeds server Password}

Uploading Data on Mac or Linux

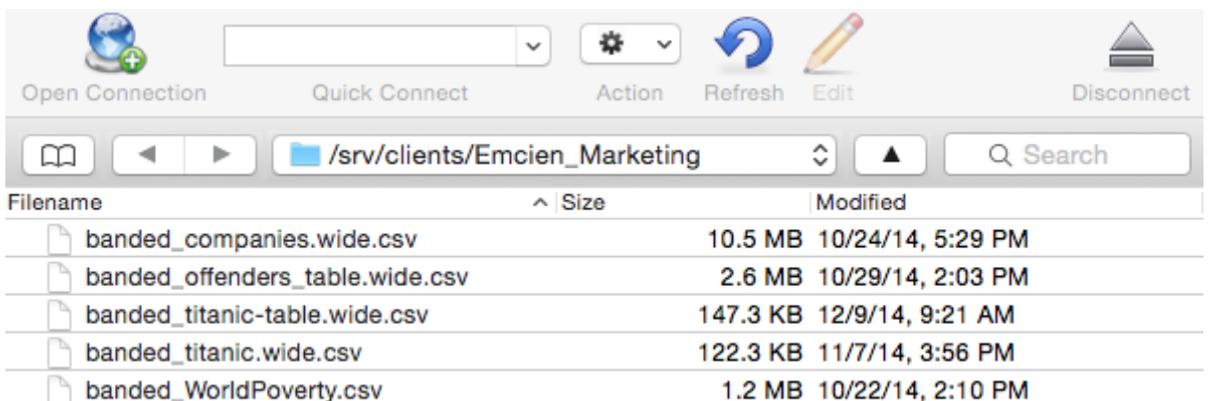
Launch your preferred FTP client. If you do not have a preferred FTP client, we recommend [Cyberduck](#).

For this article, we use screenshots of Cyberduck.

Click Open Connection.



Enter the below credentials and click Connect.



Your FTP client will then display the files on your server.

Server: Your Emcien URL. This URL was set by your IT resource during installation.

Username

Password

Drag and drop your data file to the server files in your FTP application.

On the Emcien home page click Analyze Data. You'll see the data files on this screen.

Uploading Data on Windows

Click the Start button . The Start menu is displayed.

On the Start menu, click Computer. The Computer folder is displayed.

Right-click anywhere in the folder and click Add a network location.

The Add Network Connection wizard is displayed. Click Next.

On the next tab, select Choose a custom network location. Click Next.

On the next tab, enter your Emcien URL.

This URL was set by your IT resource during installation.
Ensure you include *FTP://* at the beginning of the URL.

Then click Next.

Clear the Log on anonymously checkbox.

Type *data* in the username field. Click Next.

On the next tab, name your shortcut *Emcien* by typing it in the Type a name for this network location field. Click Next.

On the next tab, click Finish.

Type *feeds1* in the password field. Check the Save password checkbox so you can connect directly in the future.

Click Log On. You can now drag data files into this folder.

On the Emcien home page, click Analyze Data. You'll see the data files on this screen.

Analysis

Analyze the data you've uploaded.

Using your preferred Internet Browser, navigate to the Emcien Sign In page:

For cloud users, navigate to <http://patterns.emcien.com/> On the Home page click Analyze Data.

Home

Analyze Data

Welcome to Emcien
Please start by [adding a Project](#)

The Analyze Data button will bring up all of your uploaded data. Select your data set. You have the option of selecting a Project Folder or creating a new one.

Please select a Data File Filter by name

GeneralPayments.csv	450 MB	Sep 30 2:25PM
<input checked="" type="checkbox"/> ITOCLogInCSV_time.wide.csv	3.3 MB	Mar 18 2:01PM
ITOCLogInCSV_time_TEST_4K.wide.csv	510 KB	Mar 18 11:20AM
Patient-Clinical.csv	470 KB	Nov 6 1:21PM
retail-sales.csv	410 MB	Nov 20 9:51AM
retail-sales.receipt.csv	3.2 KB	Nov 19 12:17PM
Retail_Sales_2014_12_02.receipt.csv	410 MB	Dec 2 3:34PM
smart_banded_BankData3Q2013.wide.csv	1.6 MB	Oct 21 4:18PM
titanic-upload.wide.csv	140 KB	Dec 29 1:10PM
tx_truck_2014_11_06.wide.csv	5.9 MB	Feb 19 1:30PM
UsMilitaryDeaths.wide.csv	910 KB	Sep 30 9:07AM

Projects
Select a Project to display Report on Home Page
 Network Data

Prediction Category
Up to 20 Items allowed in Prediction Category

Override Defaults

Report Name

Start Analysis

Cancel

Select a Project Folder, enter a Prediction Category if you have one (in this case, Priority), and if necessary rename the Report. For this data set we will be predicting the Category "Priority".

Once your project is organized, click Start Analysis.

Projects
Select a Project to display Report on Home Page

Network Data

Prediction Category
Up to 20 Items allowed in Prediction Category

Priority

Override Defaults

Report Name

Network Attack

The load screen will take you through each stage of the analysis.

Network Attack Log Emcien

Transferred 2s → Parsed 2s → Computed 5s → Loading →

Log: [Summary](#) | **Expanded** Run Time 18s

#	Message	Time
46	==> Building Shard at patt_prod_33fd2c__rep65322108__v00048	14s
45	... Preconditions checked out	14s
44	Testing Preconditions...	14s
43	Report State: loading	14s
42	Report State: Computed	14s
41	Postconditions checked out	14s
40	Testing Postconditions...	14s

When the analysis is complete you will hear a chime and the View Analysis button will be highlighted in green. Just click the button to see your results.

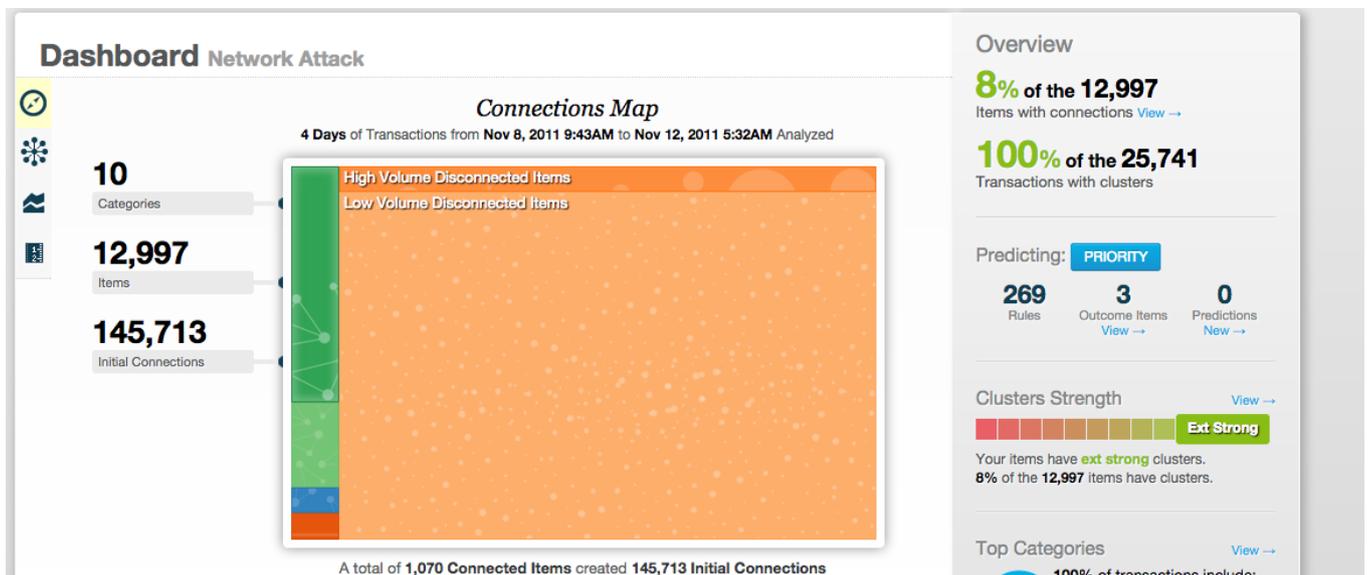
Network Attack Log Emcien



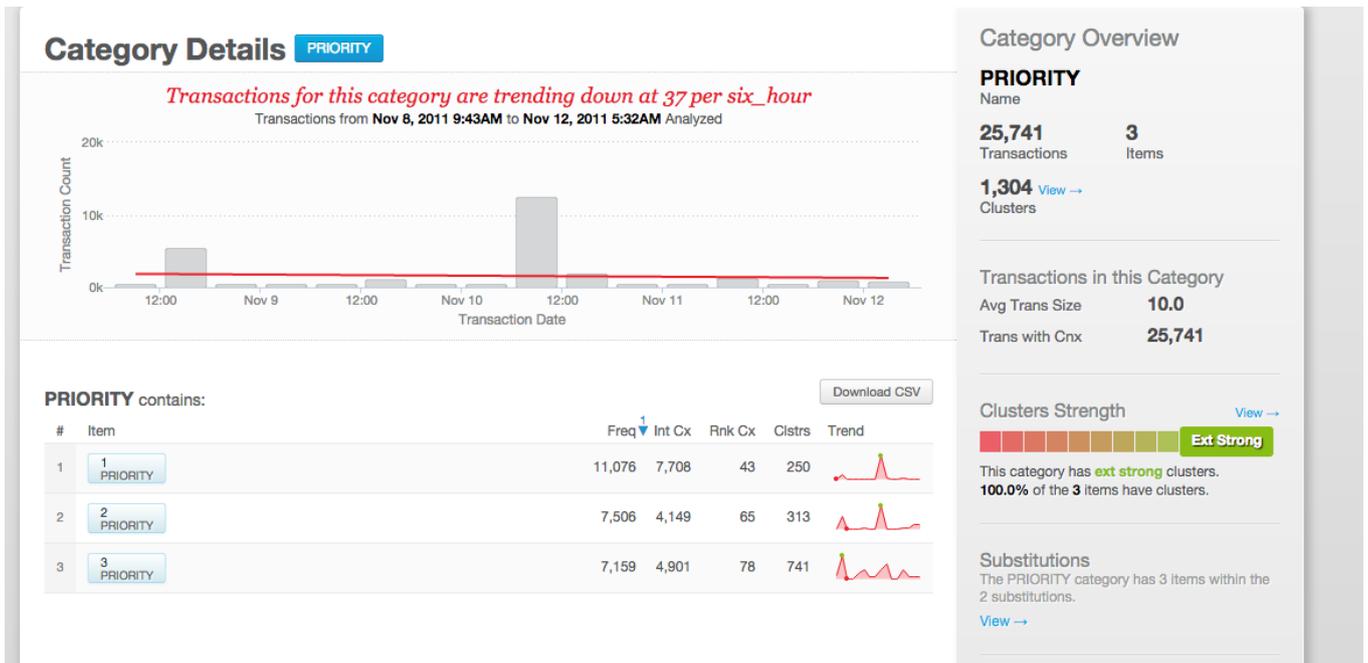
Results

See the results of your analysis.

In our analysis we have directed Emcien to predict the Category "Priority". This will bring up the Dashboard and the Predictions metrics for the targeted outcome. The analysis has identified 269 rules to predict the three outcomes, Priorities 1, 2, and 3.

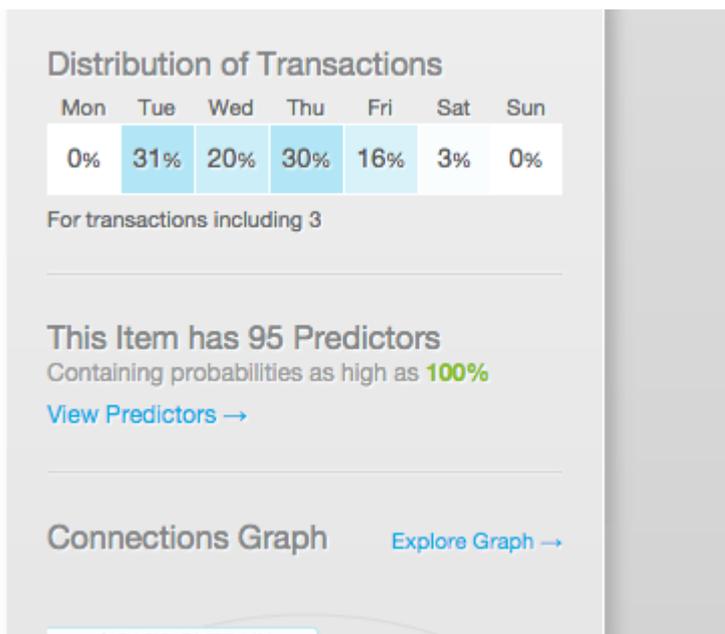


To see the predictions in detail, click the dark blue Priority button. The Category Detail is displayed, showing all three items, their connectivity statistics, and because this data includes a timestamp, a trend line graph.



Because we are interested in only the highest priority attacks we will select the item "Priority:3" at the bottom of the list.

The analysis has found 95 Predictors for Priority:3 attacks, with individual predictions with probabilities as high as 100% in this data. Clicking View Predictors will take us to the list of these rules.



All 95 Predictive Rules are shown for the Item Priority:3, listed in order of strength.

Predictors for 3 PRIORITY Download CSV

Displaying **95** Predictive Clusters

#	Predictor Cluster	Probability	Item	Freq
1	UNKNOWN CLASSIFICATION	100.0% →	3 PRIORITY	7,139
2	7 SIGNATURE_ID	100.0% →	3 PRIORITY	2,340
3	(http_inspect) IIS UNICODE C... SIGNATURE_NAME	100.0% →	3 PRIORITY	2,334
4	4 SIGNATURE_ID	100.0% →	3 PRIORITY	2,120
5	ITOC SOURCE_NAME	27.8% →	3 PRIORITY	7,159
6	TCP Timestamp is outside of ... SIGNATURE_NAME	100.0% →	3 PRIORITY	1,764
7	12 Hour	72.5% →	3 PRIORITY	1,457
8	154.241.88.201 DST_IP	22.1% →	3 PRIORITY	4,544

Item Overview

3 ID

PRIORITY Category [View →](#)

741 Clusters [View →](#) **7,159** Transactions

4,901 Initial Connections **78** Ranked Connections

1 [View →](#) Substitutions

Distribution of Transactions

Mon	Tue	Wed	Thu	Fri	Sat	Sun
0%	31%	20%	30%	16%	3%	0%

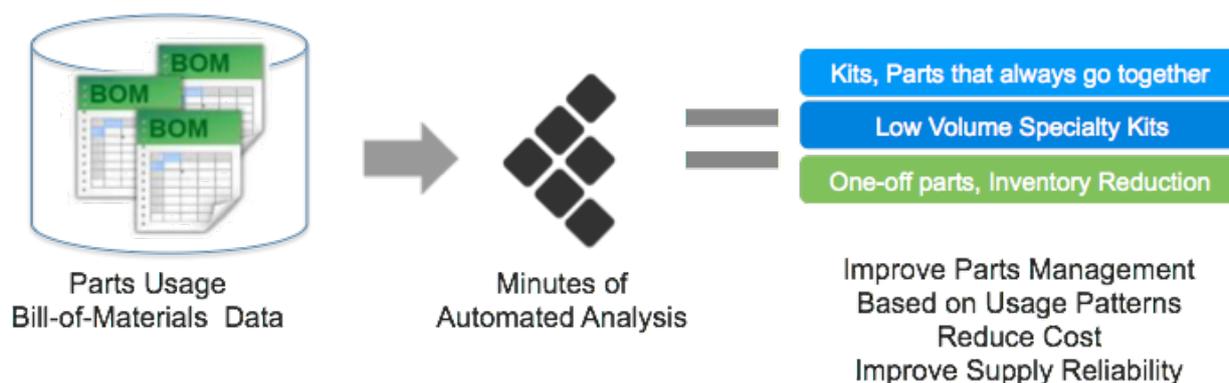
For transactions including 3

You can download these Predictive Clusters by clicking the Download CSV button in the top right corner of the page, but Emcien recommends that these predictions be automated through the real-time [Prediction Server](#).

Improving Parts Management

Use Case: Improving Parts Management

This article outlines how Emcien can drive inventory improvement and cost reduction automatically with parts data to aid in parts management. The objective of this use case will be to discover patterns in parts usage to improve kitting, inventory planning, availability, and minimizing obsolescence costs.



Align your data with a use case

Note: This procedure uses the Truck Manufacturing example data set. Your data should be similar to the example data, with columns for material names for a manufacturing environment.

[ACME_truck_parts.wide.csv](#)

Data Prep

Optimize your data.

This data set contains truck manufacturing data in the wide format. The wide format is the more universal and most commonly used Emcien format. You can use the wide format for multi-dimensional data, such as demographics or configurable products. In the wide format, each transaction is identified by a single row of data.

Each column header is a unique feature

Date	Model	Front Axle	Rear Axle	Transmission	Engine	Weight Rating (Gross)	Gear Speed - 1	Wheelbase
1/1/14	TX-2300	F8000-STD	R15000-STD	7SPD-MAN	MD-330	25200	4.17	469
1/1/14	TX-2400	F8000-STD	R12000-STD	7SPD-MAN	MD-330-NG	36500	6.9	379
1/1/14	TX-2300	F10500-HD	R15000-STD	6SPD-MAN	MD-285-NG	26500	4.17	649
1/1/14	TX-3500	F10000-STD	R15000-HD	6SPD-MAN	MD-300	18750	5.45	424
1/1/14	TX-2300	F10500-HD	R17000-HD	7SPD-MAN	XTD-370	26500	5.64	649
1/1/14	TX-2400	F12000-STD	R17000-HD	6SPD-MAN	MD-330-NG	31000	5.64	369

Each row represents an order

In our example, the column names are different kinds of parts used in the manufacturing of trucks. This parts data is ideal for any analysis that seeks to optimize **bundling**, **kitting**, or to **minimize materials cost** in manufacturing. If your data does not look similar to the wide format data here, don't worry. Emcien Patterns also works with 'receipt' type data. For more information about data preparation and different data types supported, see our [Data Prep Guide](#).

Uploading Data

Upload your data for analysis.

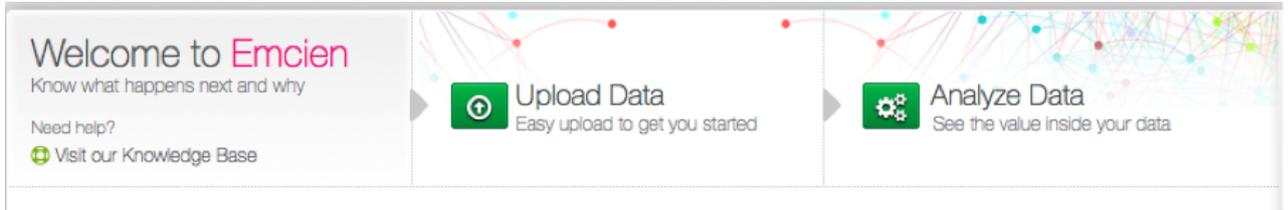
This article covers how to upload data files to Emcien using [File Transfer Protocol \(FTP\)](#).

Make sure you've prepared your data before uploading to Emcien. Check out the [Preparing Your Data](#) article for more information. You can also email us at support@emcien.com for help getting started.

Upload using Drag and Drop

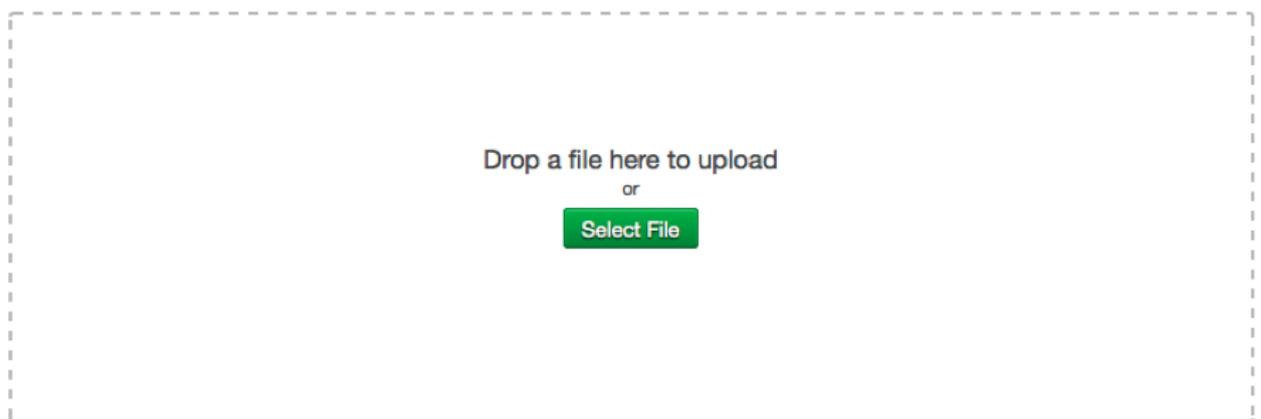
In Emcien version 2.14 you can now drag and drop files into the application for processing. If you are using a previous version you can use one of the other methods listed below.

Simply open the Emcien application and select "Upload Data"



Select the file that you wish to process, then drag and drop it onto the Upload Data screen

Upload Data Services



Once it has finished uploading you can then click on "Analyze This Data" to

process the file.

✔ Upload Complete!

[Analyze This Data](#) or [Upload More Data](#)

Uploading Data to the Emcien Cloud

You can upload data files to Emcien using [SSH File Transfer Protocol \(SFTP\)](#) and your preferred FTP client, such as [FileZilla](#) or [Cyberduck](#).

To connect to Emcien using your FTP client, use the following credentials:

Host: feeds.emcien.com

Username: {Your Emcien Feeds server Username}

Password: {Your Emcien Feeds server Password}

Uploading Data on Mac or Linux

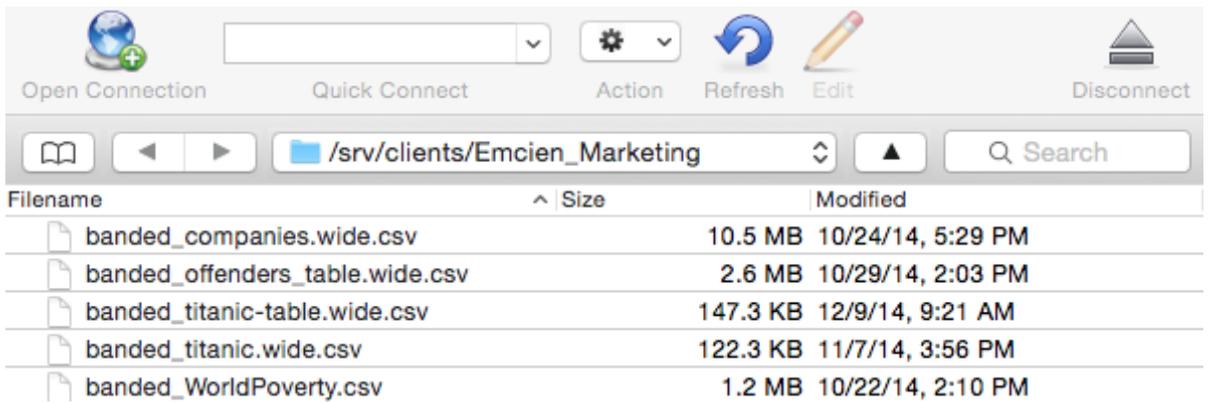
Launch your preferred FTP client. If you do not have a preferred FTP client, we recommend [Cyberduck](#).

For this article, we use screenshots of Cyberduck.

Click Open Connection.



Enter the below credentials and click Connect.



Your FTP client will then display the files on your server.

Server: Your Emcien URL. This URL was set by your IT resource during installation.

Username

Password

Drag and drop your data file to the server files in your FTP application.

On the Emcien home page click Analyze Data. You'll see the data files on this screen.

Uploading Data on Windows

Click the Start button . The Start menu is displayed.

On the Start menu, click Computer. The Computer folder is displayed.

Right-click anywhere in the folder and click Add a network location.

The Add Network Connection wizard is displayed. Click Next.

On the next tab, select Choose a custom network location. Click Next.

On the next tab, enter your Emcien URL.

This URL was set by your IT resource during installation.
Ensure you include *FTP://* at the beginning of the URL.

Then click Next.

Clear the Log on anonymously checkbox.

Type *data* in the username field. Click Next.

On the next tab, name your shortcut *Emcien* by typing it in the Type a name for this network location field. Click Next.

On the next tab, click Finish.

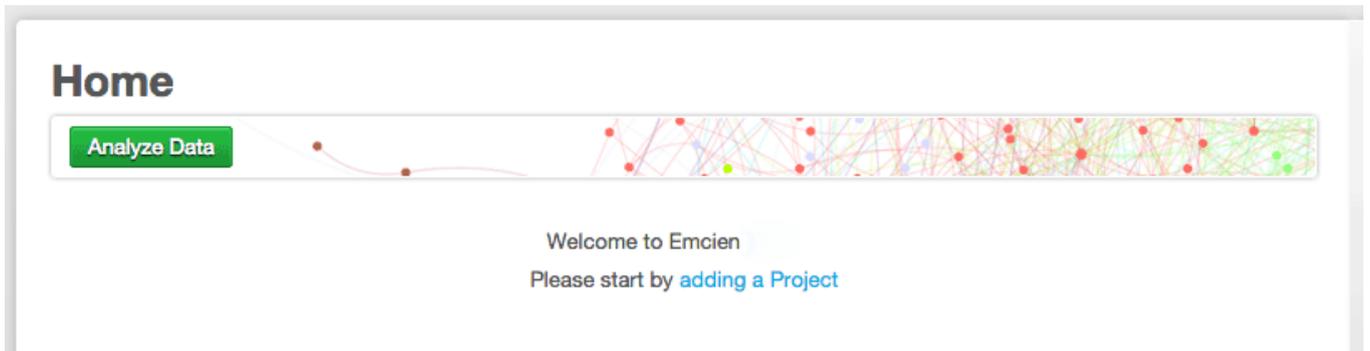
Type *feeds1* in the password field. Check the Save password checkbox so you can connect directly in the future.

Click Log On. You can now drag data files into this folder.

On the Emcien home page, click Analyze Data. You'll see the data files on this screen.

Analysis

On the Home page click Analyze Data.



The Analyze Data button will bring up all of your uploaded data. Select your data set.

New Analysis Demo Admin

Source Data Filter by Name

Name	File Size	Modified Time
✓ ACME_truck_parts.wide.csv	2.8 MB	Jun 10 2:30PM
retail_sales_2014_12_02.receipt.csv	86 MB	Jun 10 1:44PM
Airlines-Delay-Data.wide.csv	33 MB	Jun 4 10:17AM
ACME-Auto.csv	170 MB	Jun 2 4:38PM
ACMEOfficeSupplyInputFile.csv.gz	390 MB	Jun 2 2:27PM
BankData3Q2013-noemp.wide.csv	1.4 MB	Jun 2 2:00PM
ACMEAutoParts_initial_NoSubCategory.csv	190 MB	Jun 2 1:59PM
ACMEHardwareMediumCleanedupPricing.csv.gz	220 MB	Jun 2 1:58PM
ACME-Office-1M-Load.csv	370 MB	May 28 3:55PM
Consumer_Complaints.wide.csv	27 MB	Mar 31 9:30AM

Select a Project Folder, and if necessary rename the Report.

Projects

Reports in a project will display on home page

Demo

Predictions

Override Defaults

Report Name

ACME truck parts

Start Analysis

Once your project is organized, click Start Analysis.

ACME truck parts Demo Admin



The load screen will take you through each stage of the analysis.

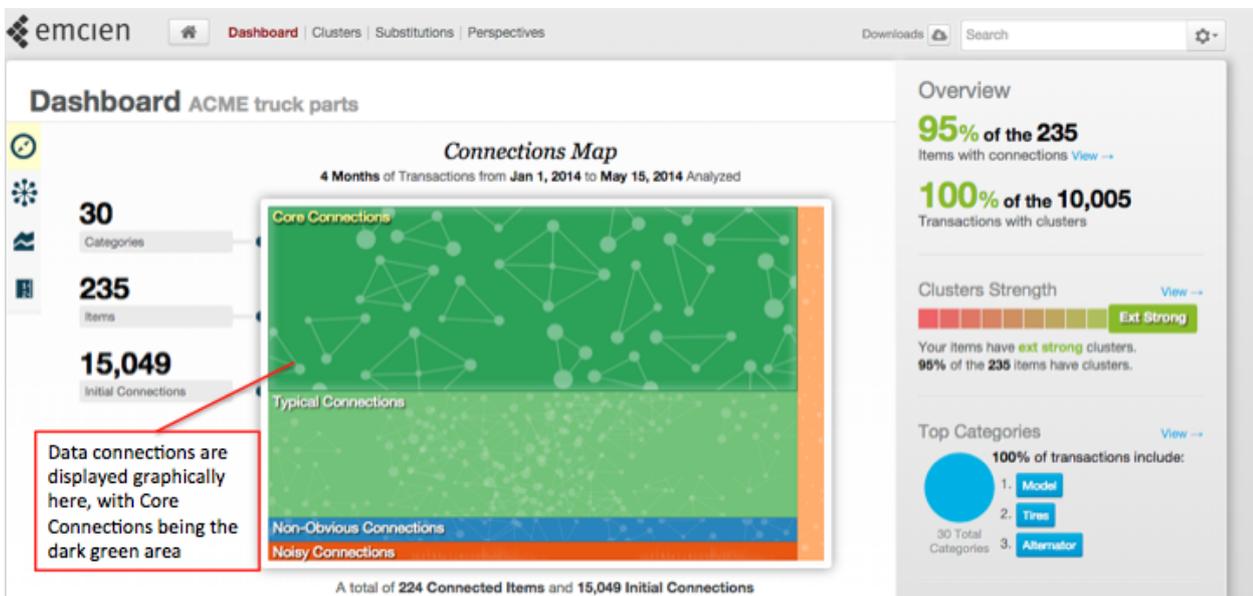
When the analysis is complete you will hear a chime and the View Analysis button will be highlighted in green. Just click the button to see your results.

Results

See your results.

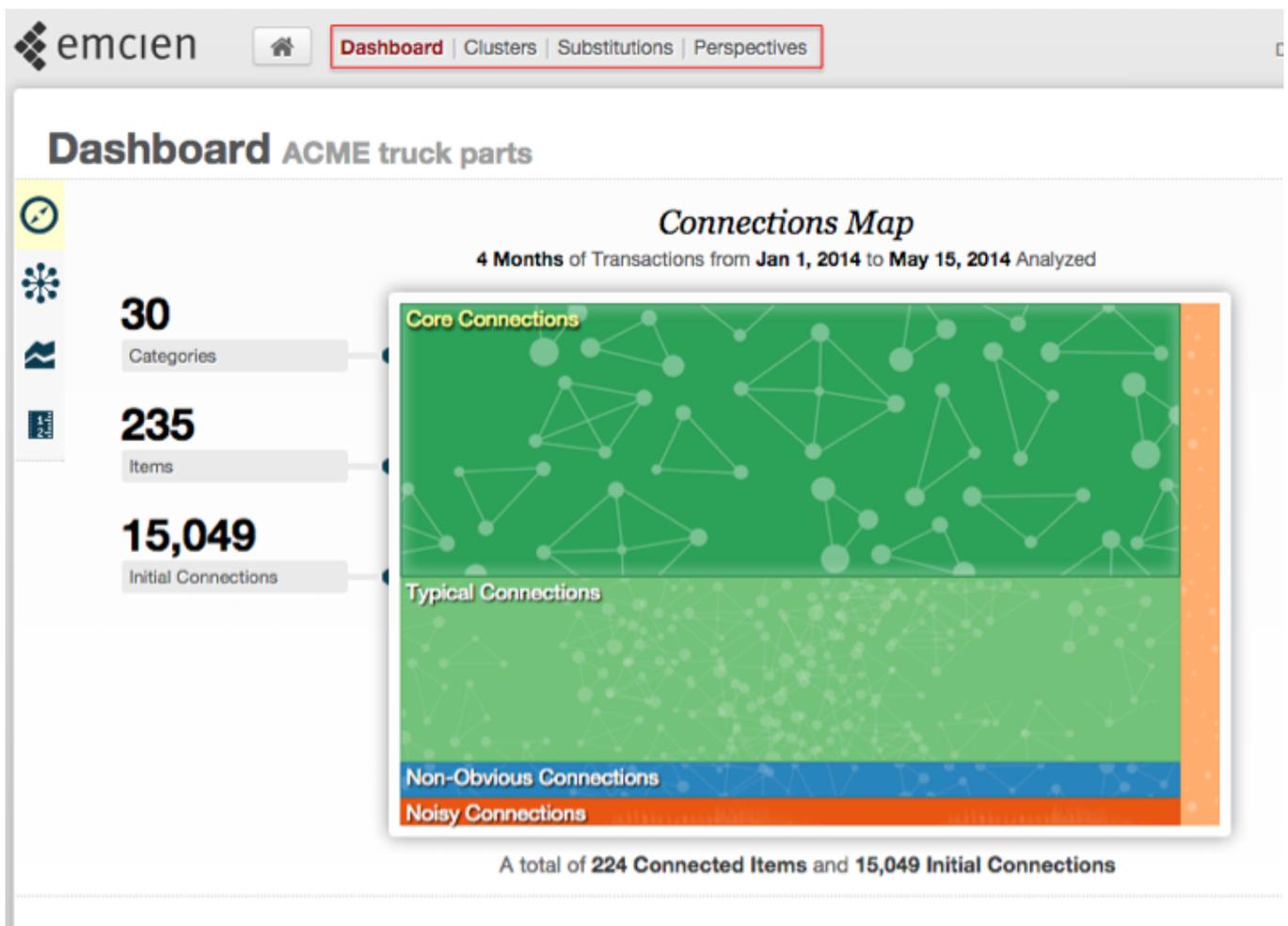


The Dashboard page for the truck analysis data set is displayed below. The interactive graphic on the home screen is the Connections Map, a representation of the analyzed data displayed by how materials were ordered (or connected) together.



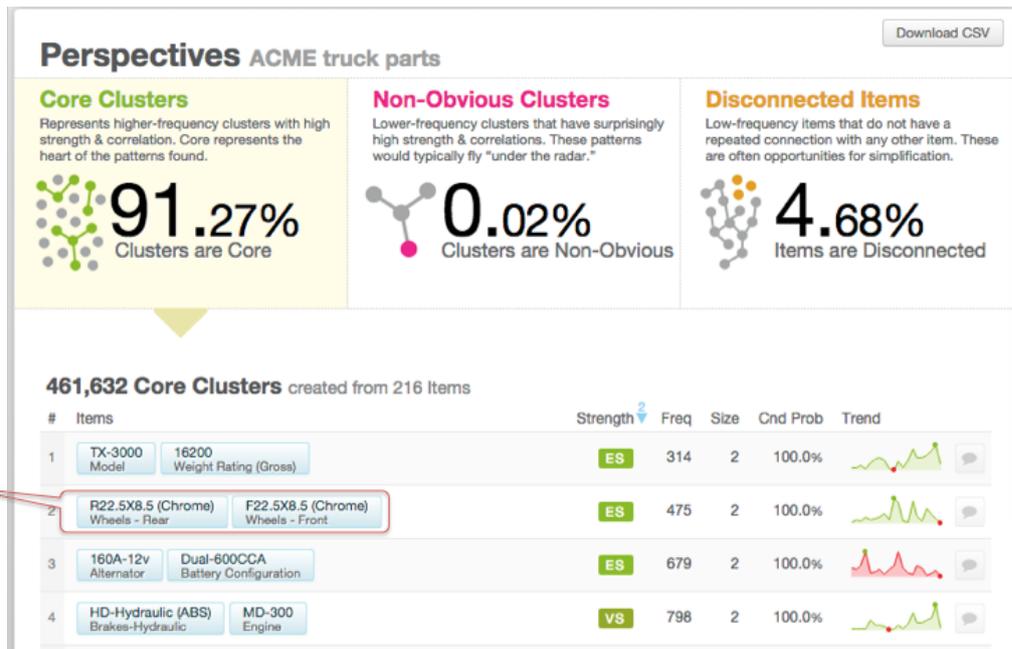
Each color represents the connections identified as relevant patterns in the data set. Mousing over each section will display examples of the corresponding clusters. These include Core and Typical Connections (materials frequently purchased together), and Low Volume Disconnected Items (materials rarely purchased that are prime for elimination). Notice at the top of the screen the links

to separate tabs in the engine.

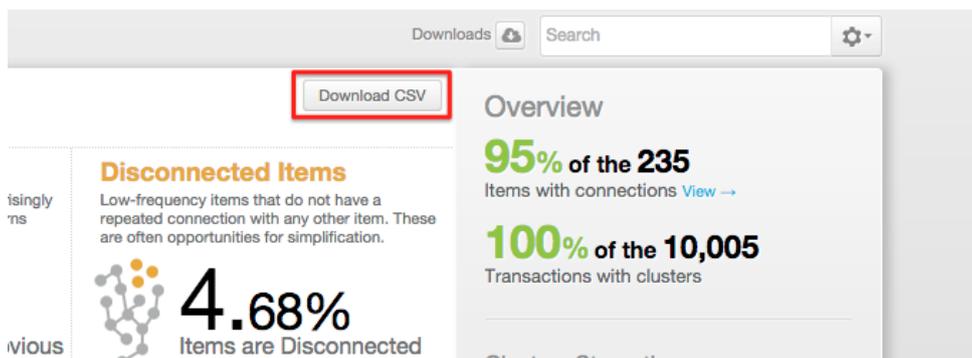


Clicking on the perspectives tab will bring up a list of each kind of data connection found in the analysis.

These two items were sold together 100% of the time; kitting them may reduce materials cost while still answering customer needs



The perspectives screen organizes the analyzed results into three groups of patterns that match many use cases. The default view shows Core Clusters, which are the *high frequency and highly correlated* items in the data. Opportunities for **Bundling and kitting** are in items that often sell together and are sold together frequently, which are seen easily in Core Clusters. Core Clusters are the connections that are central to understanding the data.



To leverage these results, click the “Download CSV” button in the upper right hand corner to save the analysis as a CSV file, to be imported into Excel or other manufacturing solutions (also available as an API).

Download CSV

Perspectives ACME truck parts

Core Clusters

Represents higher-frequency clusters with high strength & correlation. Core represents the heart of the patterns found.



Non-Obvious Clusters

Lower-frequency clusters that have surprisingly high strength & correlations. These patterns would typically fly "under the radar."



Disconnected Items

Low-frequency items that do not have a repeated connection with any other item. These are often opportunities for simplification.



109 Non-Obvious Clusters created from 63 Items

#	Items	Strength ²	Freq	Size	Cnd Prob	Trend
1	Dual-Elect Battery Switch, 451 Wheelbase	FS	3	2	100.0%	
2	Dual-Elect Battery Switch, 451 Wheelbase, 6.5 Gear Speed - 1	FS	3	3	100.0%	
3	48000 Weight Rating (Gross), ECON-275 Engine	FS	3	2	100.0%	

Clicking on the Non-Obvious Clusters shows us opportunities in inventory management for avoiding **shortages or excess inventory**. In the example above, if we were to run out of the Dual-Elect Battery switch, we would potentially effect the sales of Wheelbase 451. These are connections that don't occur frequently, but when they occur, they mostly occur together.

Download CSV

Perspectives ACME truck parts

Core Clusters

Represents higher-frequency clusters with high strength & correlation. Core represents the heart of the patterns found.



Non-Obvious Clusters

Lower-frequency clusters that have surprisingly high strength & correlations. These patterns would typically fly "under the radar."



Disconnected Items

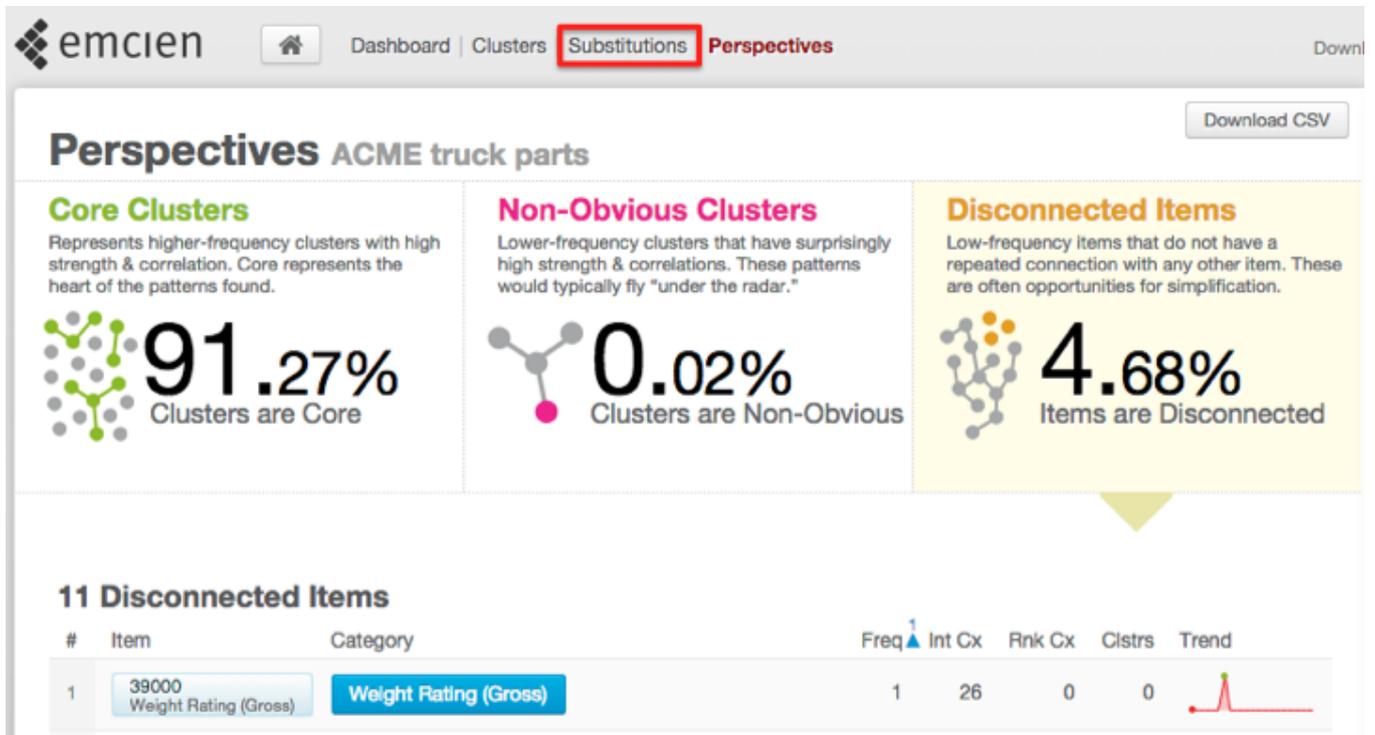
Low-frequency items that do not have a repeated connection with any other item. These are often opportunities for simplification.



11 Disconnected Items

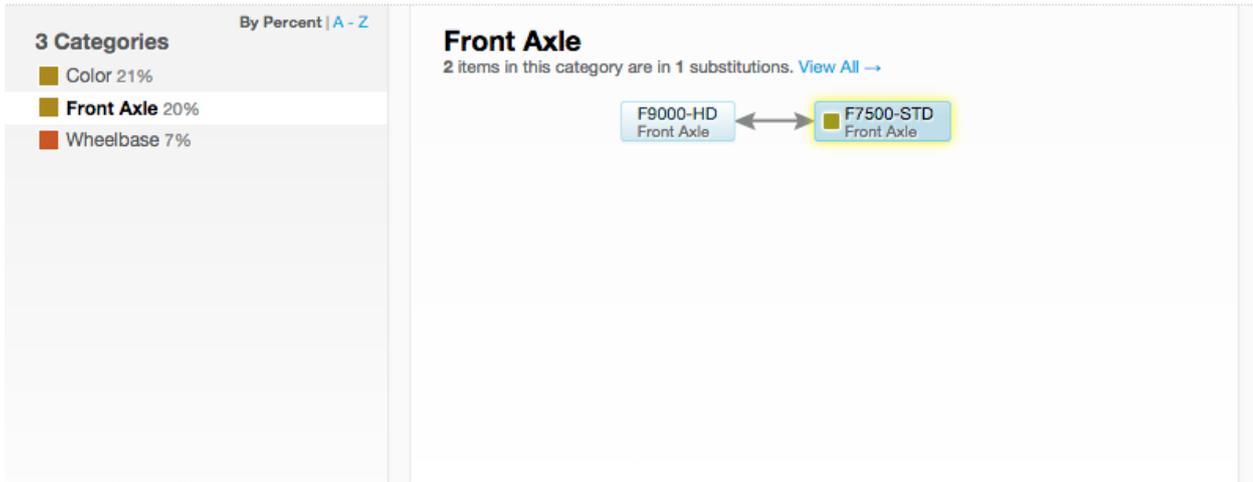
#	Item	Category	Freq ¹	Int Cx	Rnk Cx	Clstrs	Trend
1	39000 Weight Rating (Gross)	Weight Rating (Gross)	1	26	0	0	
2	F22.5X10 (Paint) Wheels - Front	Wheels - Front	1	27	0	0	
3	45000 Weight Rating (Gross)	Weight Rating (Gross)	1	24	0	0	

Clicking the Disconnected Items shows us excellent candidates for **inventory reduction**. Seen here are items that don't have any meaningful connections with other items. If an item on this list is low margin or has high storage cost it can be considered for reduction.



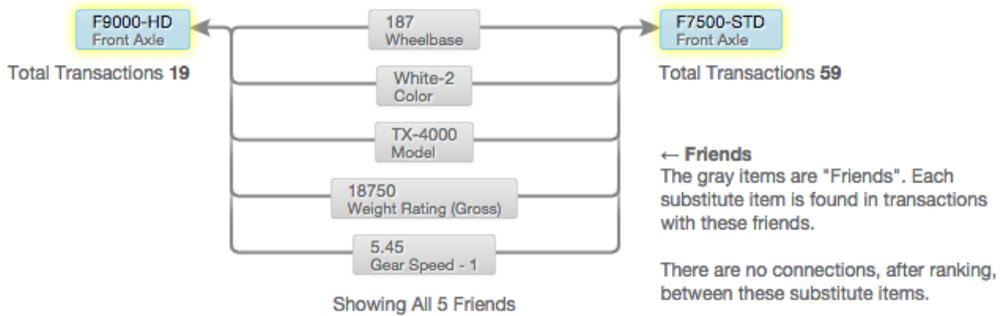
Another feature that can be helpful for inventory management is the Substitutions feature, located at the top of the screen in a tab next to perspectives.

Substitutions ACME truck parts



Substitute and friends

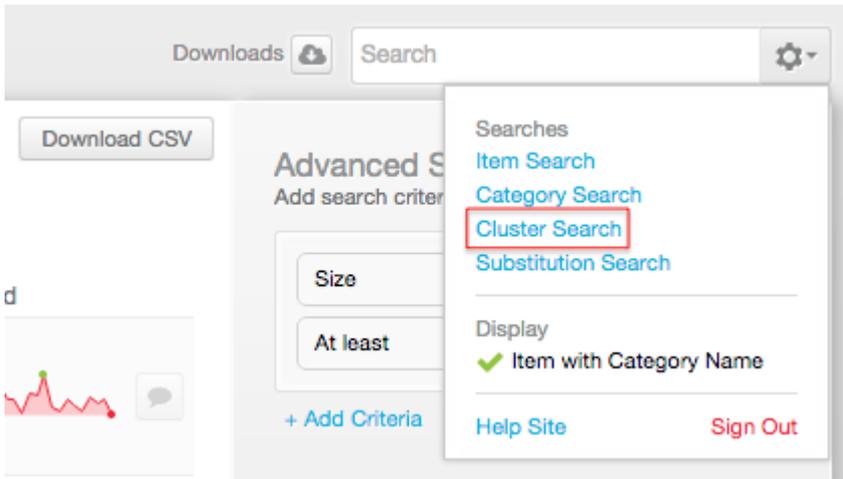
This Substitution is **S**



Emcien software automatically identifies similar items or parts that can be replacements for each other, showing an opportunity to either **remove one of the items from production** or **reduce the inventory of both items**. These are the individual items that "act" like other items in the data set- meaning that these items are ordered with the same group of items but are rarely, if ever, seen together. In this example, we have two similar shades of green which are ordered fairly often, but never ordered together.

Parts Commonality

Finding parts commonality and standardization is a valuable capability for manufacturing and supply chain offered by Emcien. To achieve this, go to the Cluster Search screen using the search drop-down menu:

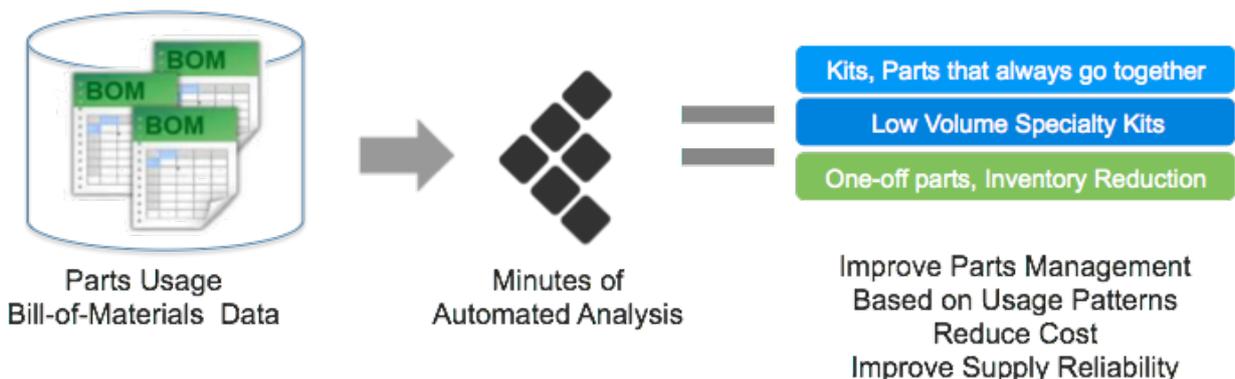


And search for clusters with a large size (pictured below are clusters of at least 7 items):

The screenshot shows a search results page titled 'Cluster Search ACME truck parts'. It indicates that the search matched 99,165 of 505,804 total clusters, with search criteria of 'Size at least 7'. The results are displayed in a table with columns for '# Items', 'Strength', 'Freq', 'Size', 'Cnd Prob', and 'Trend'. Three clusters are shown, each with a strength of 8 and a size of 7. The first cluster has a condition probability of 48.2%, the second 46.4%, and the third 47.0%. To the right, there is an 'Advanced Search' sidebar with a 'Size' dropdown set to 7 and a 'Search' button. Below that, there are 'Display Options' for showing parent or child clusters.

These clusters are groups of items ordered together frequently. These results are similar to our Core Clusters, but provide standard configurations for large groups of items.

For more detail on performing searches, see our [Guide to Searching](#).



Each of these product relationships can be analyzed automatically to identify **hidden carrying costs** and opportunities for **bundling and more efficient product planning**. For help using Emcien to optimize your inventory you can contact your Services team member directly or email us at hello@emcien.com and we will connect you with a services professional.

Administration Hierarchy

There are two levels of Administrator available within Emcien. Whether these are available will depend on the installation type being used.

The Cloud installation is configured as a Multi-Tenant environment, and Emcien is the System Administrator. Customers are able to log in as an Administrator.

If Emcien has been installed in a Virtual Machine environment the customer is the System Administrator. The Customer also continues to manage the Administrator role. Administrators can create users and projects within a company, and process reports. System Administrators have the same privileges as Administrators, but they have an additional privilege that allows them to create Companies.

This list shows the different levels and the corresponding areas that can be accessed.

	System Administrator	Administrator	Data Loader	User
System Level				
Create Company	✓			
View All Reports	✓			
View Jobs	✓			
View Logs	✓			
Edit Company	✓			
Features				

Download Report ✓
Logs

Admin Level

Edit Company ✓ ✓

Create Project ✓ ✓

Edit Project ✓ ✓

View/Edit Users ✓ ✓

View/Edit API ✓ ✓
Token

View API Usage ✓ ✓

View Reports ✓ ✓ ✓

Process Analysis ✓ ✓ ✓
via GUI Loader

User Level

View Analysis ✓ ✓ ✓ ✓
Reports Created
via Autoloader

View Analysis ✓ ✓ ✓ ✓
Reports Created
via GUI Loader

Download CSV	✓	✓	✓	✓
Data from Web UI				
Call Emcien API	✓	✓	✓	✓

Relicensing Emcien

Each installation of Emcien comes with a expiration date. To continue using Emcien, you'll need to relicense your installation.

This article includes everything you'll need to relicense your Emcien installation, including:

[Finding When Your License Expires](#)

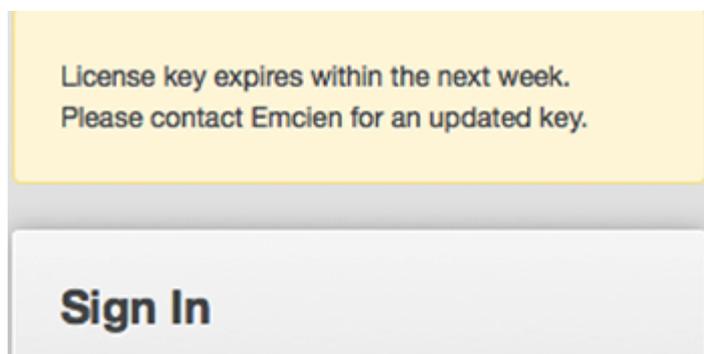
[Requesting a New License](#)

[Updating Your License](#)

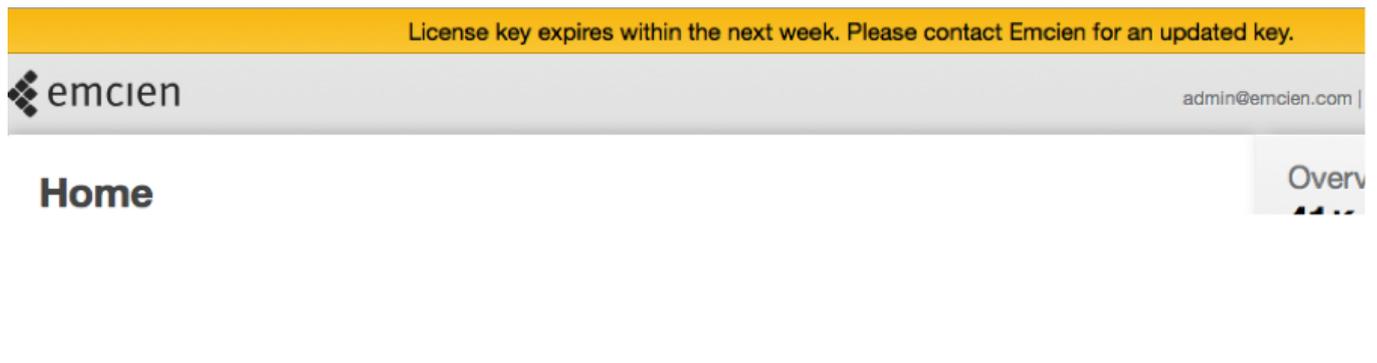
[Confirming Your New License](#)

Finding When Your License Expires

Emcien will display the following message on the Emcien Sign In page one week before your license expires.



After signing in, the following message is displayed at the top of all Emcien pages.



Requesting a New License

When nearing your expiration date, follow the steps below to request a new license.

1. Email us at support@emcien.com. Be sure to include the UUID for your Emcien installation.

If you don't know your UUID, you can retrieve it by:

Contacting your IT team that installed Emcien

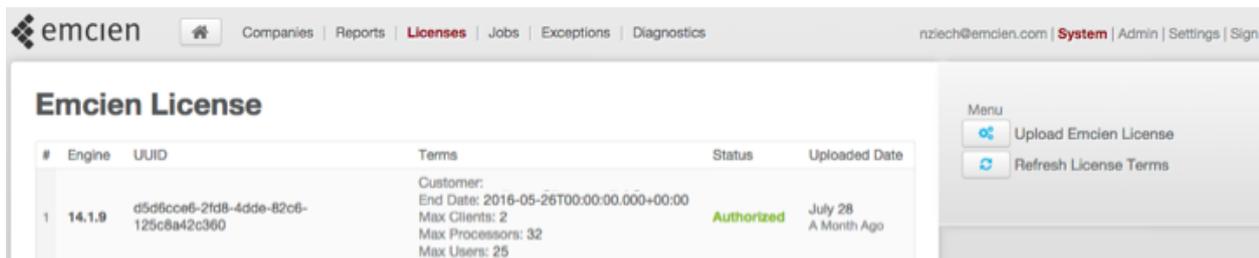
Visit your Emcien URL followed by `/system/licensed_engines`

For example, if you typically use `https://app.emcien.com` to access Emcien, you can find your UUID at `https://app.emcien.com/system/licensed_engines`

2. After we receive your email, the Emcien team will email you a new license key. Save your license key to a location that can be accessed from the Emcien VM.

Updating Your License

After you receive your new license key, you're ready to update your Emcien license.



1. Save your new license key to a location you can access in a web browser while using the Emcien VM.
2. Visit your Emcien URL followed by /system/licensed_engines

For example, if you typically use <https://app.emcien.com> to access Emcien, you can find your UUID at https://app.emcien.com/system/licensed_engines

The License page is displayed.

3. On the License page, click Upload Emcien License and select your license key.
4. Click Upload License.

Emcien will process the license in approximately two minutes.

5. When the license is processed, click Go to Emcien.

The Emcien Sign In page is displayed.

Your expiration messages will no longer display. Your team can continue using Emcien as normal.

Confirming Your License

After updating your license, follow the steps below to verify your new expiration date.

1. On the Emcien home page, click System in the upper right.
2. On the System page, verify that a new expiration date is displayed in the right panel

Setting Your Item URL

This article covers how to set item URLs. Item URLs are useful for retailers that want to view links to their items through Emcien.

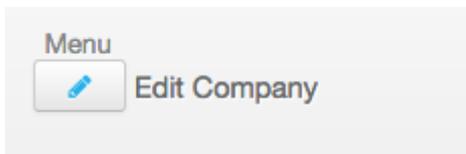
1. Click Admin in the upper right of any Emcien page.



System | Admin | Settings | Help | Sign Out

The Admin Overview page is displayed.

2. On the Admin Overview page, click Edit Company.



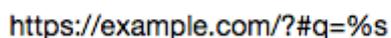
3. On the Edit Company page, change the address in the URL Template field.

For example, you can replace *example.com* with the URL of your search tool or database.

%s will call for each item in your results.

URL Template

View Items on your web site (Optional)



https://example.com/?#q=%s

In this example, the link this will use *google.com* to search for the item.

URL Template

View Items on your web site (Optional)

<https://www.google.com/?#q=%s>

4. Click Update Company.

Update Company

5. Navigate to the report for which you'd like to download APIs.

6. Find an item of interest and click the dropdown arrow.

Item

29.6 to 41.8 kg/m²
Body mass index (weight in kg/(h...

7. In the dropdown menu, click View Item Online.

Item

29.6 to 41.8 kg/m²
Body mass index (weight in kg/(h...

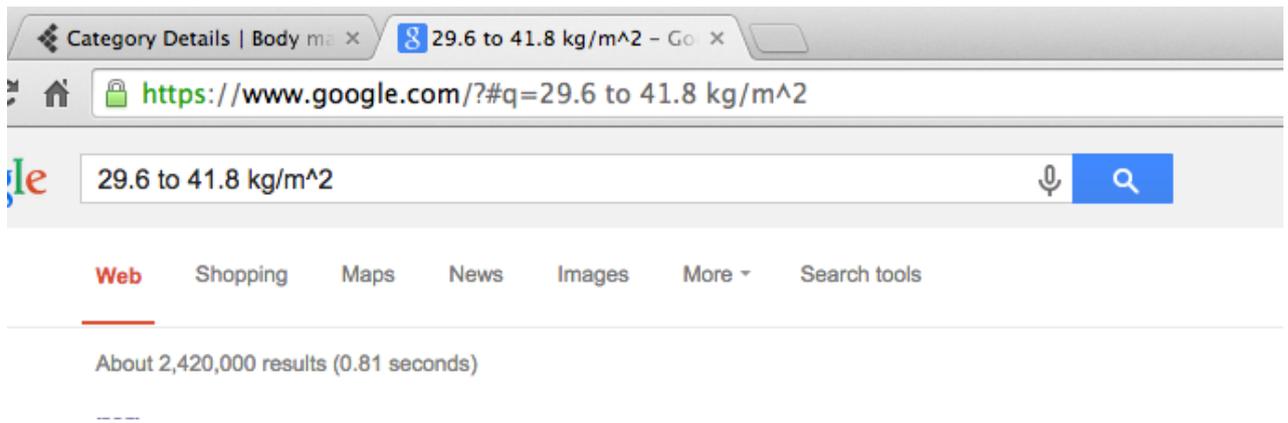
29.6 to 41.8 kg/m²
ID
Body mass index (weight in kg
/(height in m)²)
Category

56	3,413
Ranked Cnx	Clusters
416	9.0
Transactions	Avg Trans Size

[View Item online](#)

The URL set for the item earlier is exchanged.

For example, the View Item online link for the 29.6 to 41.8kg/m² item directs to a Google search for 29.6 to 41.8kg/m².



Company Management

Overview

The Company Overview page is located within the Admin section. This page lists the first five Projects and Users belonging to the Company, along with information specific to each one.

Also listed are up to five Reports. Since the Reports list can be sorted by a number of attributes, the most recent Reports are not always among the five listed on the Overview page. If the Company has more than five Reports, Users, or Projects a link will display to view the entire list.

The screenshot shows the 'Overview Demo Admin' page in the Emcien Patterns Admin interface. The page has a navigation bar with 'Overview' selected, and a user profile 'adm'. A green 'Analyze Data' button is in the top right. The main content is a table of reports with columns for #, Name, State, Project, Created, Import, and Menu. Below the table is a 'View All Reports' link. At the bottom, there are two summary tables: 'Users' and 'Projects'.

#	Name	State	Project	Created	Import	Menu
1	Warehouse Data ID: 2051706	Computing View Log	None	15 Minutes Ago	-	
2	Rail Data ID: 24225321	Report Ready	None	17 Minutes Ago	9m 51s	
3	ACME Transportation ID: 60128301	Report Ready	Demo POC	3 Hours Ago	5m 28s	
4	ACME Office 1M Load ID: 67489798	Report Ready	Demo	An Hour Ago	5m 17s	
5	ACME Office 1M Load ID: 27161984	Report Ready	Demo	20 Hours Ago	2m 38s	

[View All Reports →](#)

#	Email	Last Sign On
1	adminuser@emcien.com ID: 47	3 Hours Ago
2	testuser@emcien.com ID: 46	Never Signed In
3	demo@emcien.com ID: 39	September 24 A Month Ago

#	Name	Reports
1	POC ID: 43	1
2	Demo ID: 42	3

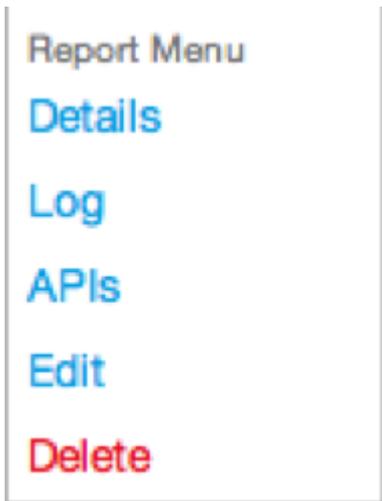
The report list contains the following information for each Report:

Column Name	Description
ID	The report number
Name	The name of the report
State	Status of the report. A report that is Ready can be viewed. Any other status is an indication that the data file failed to process.
Project	The Project(s) where the report is located
Created	The date that the Report was created
Import	The total amount of time it took to process the data file
Menu	A gear allowing access to the Report menu: Details, Log, APIs, Edit, Delete

[Viewing Report APIs](#)

Report data is available for analysis and use outside of Emcien through a set of APIs.

If the API option was enabled for the Company the user can click on the APIs link to display all of the APIs that were created for the report.



Each API output is available as both JSON and CSV. Next to each URL is the name of the file, which has been compressed in GZ format, and the size of the uncompressed (original) file. To see a detailed description of the fields and data types within the API output simply click “Display Details”.

ACME Convenience Store APIs Demo Admin

Items API

+ Display Details

URL	Filename	Original Size
https://patterns-services.internal.emcien.com/reports/17513757/api/items.csv	items.csv.gz	5.74 MB
https://patterns-services.internal.emcien.com/reports/17513757/api/items.json	items.json.gz	13.1 MB

Categories API

+ Display Details

URL	Filename	Original Size
https://patterns-services.internal.emcien.com/reports/17513757/api/categories.csv	categories.csv.gz	79.2 KB
https://patterns-services.internal.emcien.com/reports/17513757/api/categories.json	categories.json.gz	208 KB

Clusters API

+ Display Details

URL	Filename	Original Size
https://patterns-services.internal.emcien.com/reports/17513757/api/clusters.csv	clusters.csv.gz	25.5 MB
https://patterns-services.internal.emcien.com/reports/17513757/api/clusters.json	clusters.json.gz	63.7 MB

Substitutions API

+ Display Details

URL	Filename	Original Size
https://patterns-services.internal.emcien.com/reports/17513757/api/substitutions.csv	substitutions.csv.gz	1.04 MB
https://patterns-services.internal.emcien.com/reports/17513757/api/substitutions.json	substitutions.json.gz	2.73 MB

Item Predictors API

+ Display Details

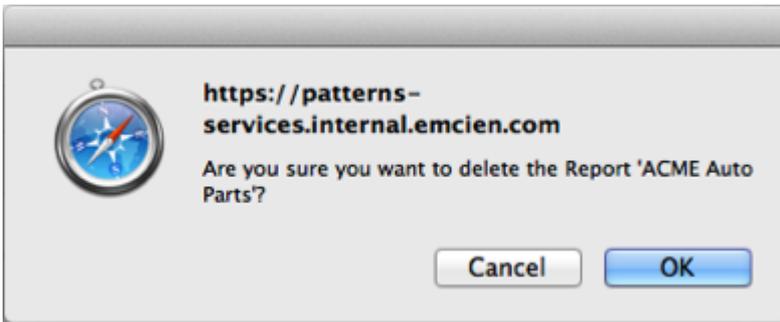
URL	Filename	Original Size
https://patterns-services.internal.emcien.com/reports/17513757/api/predictors.csv	predictors.csv.gz	48.5 MB
https://patterns-services.internal.emcien.com/reports/17513757/api/predictors.json	predictors.json.gz	128 MB

Click on an api link and the data file will be automatically downloaded to your local computer.

Deleting Reports

Reports can be deleted from the Report Details screen as well as from the Report menu on the Company Overview page.

A confirmation message will appear prior to deletion of the Report.



Deleted Reports will disappear from the Admin’s view but will remain visible with a status of ‘Deleted’ for the System Admin view. System Admins can undelete Reports that are marked “Deleted”. System Admins also have the ability to force a permanent deletion of all marked Reports.

Reports Demo Admin Analyze Data

#	Name	State	Project	Created	Import	Menu
1	Warehouse Data ID: 2051706	Deleted	None	23 Minutes Ago	-	⚙️
2	Rail Data ID: 24225321	Report Ready	None	24 Minutes Ago	9m 51s	⚙️
3	ACME Covenience Store ID: 17513757	Report Ready	None	An Hour Ago	5m 13s	⚙️
4	ACME Office 1M Load ID: 67489798	Report Ready	Demo	2 Hours Ago	5m 17s	⚙️
5	ACME Transportation ID: 60128301	Report Ready	Demo POC	3 Hours Ago	5m 28s	⚙️
6	ACME Office 1M Load - as UTF-16 ID: 78614240	Deleted	None	19 Hours Ago	1.00s	⚙️

Company License
Unlimited
Maximum Data File Size

Permanently Remove Deleted Reports

🗑️ **Remove 2 Deleted Reports**

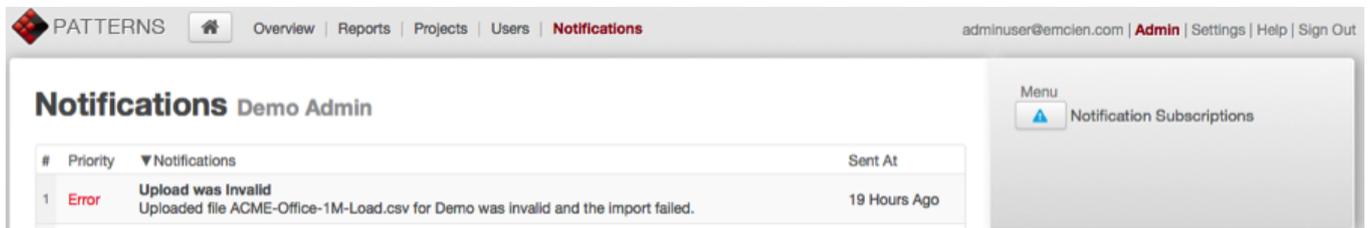
Viewing Failed Reports

When a data file fails to process to completion the Report will show a state other than “Ready”.

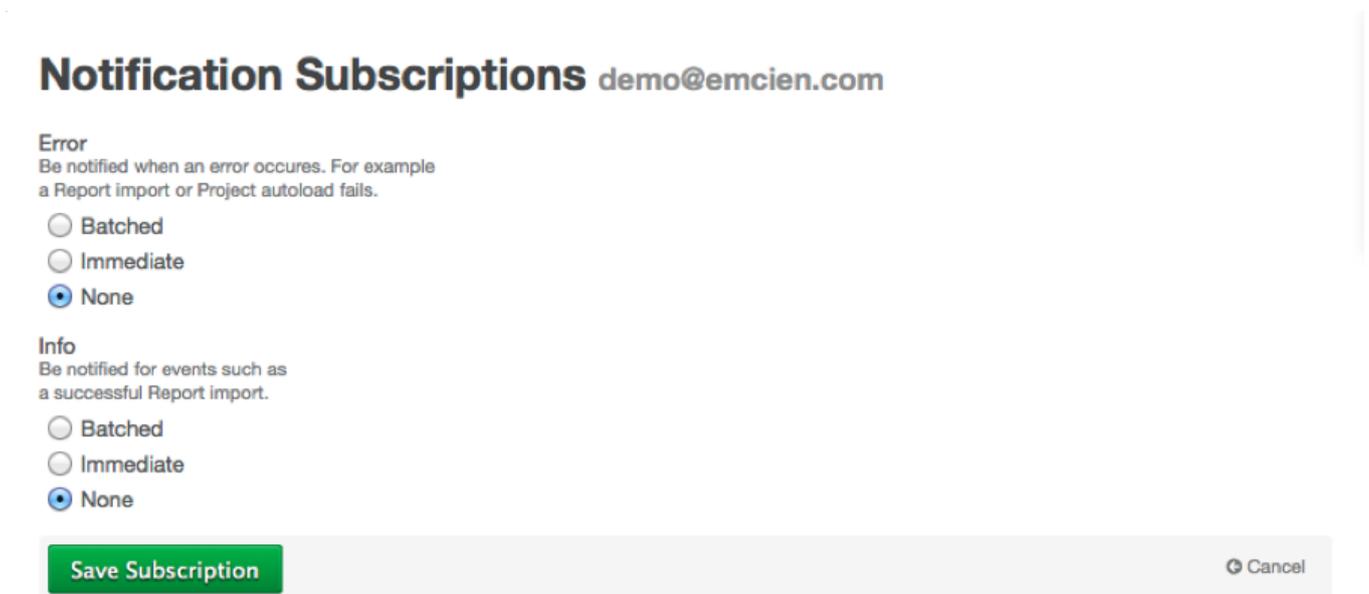
The Report Log will contain an entry that reveals the issue encountered. A Support ticket should be created if the User requires Emcien assistance in such cases. The User may be directed by Emcien Support to submit a copy of the Log for that Report. Simply click on the Log icon and it will take you to the Log page. Copy and paste all of the messages into a text file or email and send it to support@emcien.com.

Viewing Notifications

Administrators in Emcien can configure email notifications to inform them of certain data and processing occurrences.



Click on “Notification Subscriptions” and the following options will be displayed:



If the User selects Immediate notifications an email will be sent as soon as the event occurs. Batched notifications are sent once a day. If the User selects Batched notifications they will only receive information about the event in the daily batched notification email. Make the selections and click on Save Subscription.

The Admin and System Admin can also view all notifications that have been generated for the Company by clicking on the ‘Notifications’ tab in the Admin area. This list populates as the event occurs.

Creating a New Company

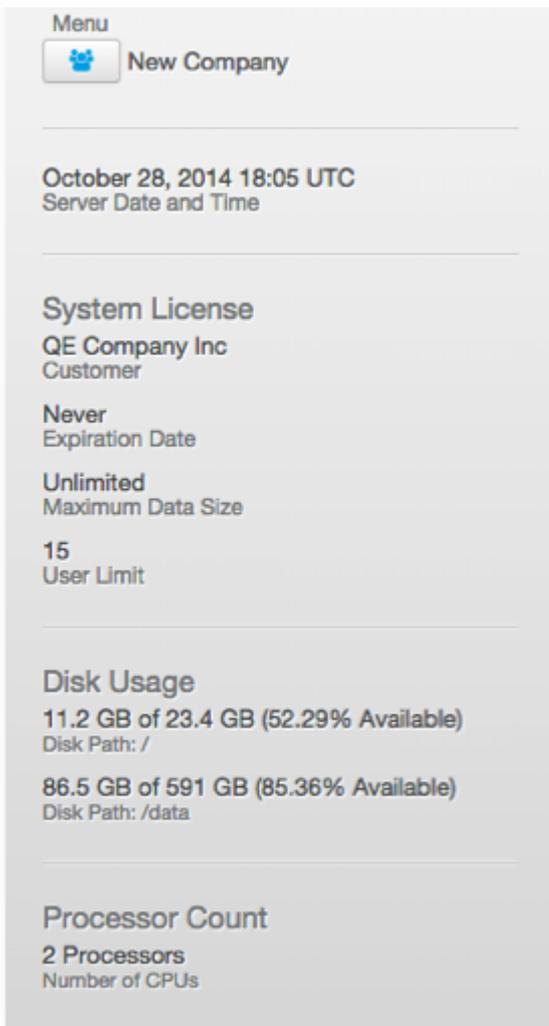
The System Admin can access the System Administrator area by clicking the “System” tab on the top right side of their Homepage.



System | Admin | Settings | Help | Sign Out

In a cloud environment Emcien personnel are the System Admins. In a VM environment someone from the Company that is responsible for setting up and maintaining the VM will serve as System Admin.

System information is available on the right side of the System screen. It displays server date and time, system license settings, disk usage and processor count.



Menu
New Company

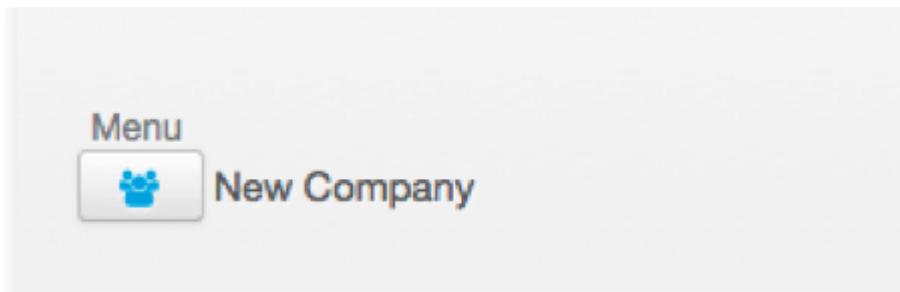
October 28, 2014 18:05 UTC
Server Date and Time

System License
QE Company Inc
Customer
Never
Expiration Date
Unlimited
Maximum Data Size
15
User Limit

Disk Usage
11.2 GB of 23.4 GB (52.29% Available)
Disk Path: /
86.5 GB of 591 GB (85.36% Available)
Disk Path: /data

Processor Count
2 Processors
Number of CPUs

To create a new company, click on the icon “New Company”



The following window will be displayed. Enter the name of the Company, the optional Data Server path, and the optional URL template.

New Company

Company Name

Please name this Company

Data Server

Data Server must accept SFTP

Format `sftp://USERNAME:PASSWORD@HOST/PATH`

URL Template

View Items on your web site (Optional)

[Viewing Disk Usage](#)

The Disk Usage section on the System Administrator Companies page specifies available physical disk space.

Imports should not be attempted after the “Working Area” (/data) portion of the disk reaches 5% Available. If the disk hits 100% capacity while a data file is being processed, not only will the analysis fail, but it may cause other issues with the Emcien application. It is recommended that the System Administrator monitor this space and delete any unnecessary Reports to clear up space on the disk.

VM Space

The Disk Usage section on the System Administrator Companies page specifies available physical disk space.

Imports should not be attempted after the “Working Area” (/data) portion of the disk reaches 5% Available. If the disk hits 100% capacity while a data file is being processed, not only will the analysis fail, but it may cause other issues with the Emcien application. It is recommended that the System Administrator monitor this space and delete any unnecessary Reports to clear up space on the disk.

Changing Available Features

To enable a specific Feature for the company simply click the box.

Any analysis reports created from this point forward will have all of the enabled options. To remove a Feature simply un-check the box next to it.

Features

Add/Remove features

- Anomalies
- API
- Clusters
- Cluster Scenery
- Item Predictors
- Substitutions

Maximum Users and File Size

This option allows the System Administrator to set limits on the maximum number of Users the Company can have and the maximum data file size that can be processed.

Limiting the number of users can be a useful way to keep Company Administrators from creating too many logins. Limiting the size of the data file is useful if there is limited disk space available on the system.

Maximum Users

0 for Unlimited Users

Maximum File Size in Bytes

0 for Unlimited Size | Disk Unit is "binary"

Update Company

To adjust the number of users or file size simply type in the number or click the up or down arrow to adjust it manually. Note that "0" is used for unlimited Users/data size. Also notice that data file size must be entered in bytes.

Maximum Users
0 for Unlimited Users

Maximum File Size in Bytes
0 for Unlimited Size | Disk Unit is "binary"

Update Company

When all of the settings are complete, click the Update Company button and you will be taken back to the Companies list.

[Viewing System Reports](#)

Clicking on the Reports at the top of the screen will allow the System Administrator to display all of the reports for every company that exists in the system.



Reports

Reports

#	Name	State	Company	▼ Created	Import	Database	
1	ACME Office 1M Load ID: 50661186	Report Ready	Company 27	2 Hours Ago	3m 51s	112 MB	
2	ACME Convienence V3 1M ID: 28519394	Report Ready	Company 27	October 27 A Day Ago	54.0s	7.78 MB	
3	Airlines Delay ID: 96041921	Report Ready	Emcien Churn	October 27 A Day Ago	3m 23s	80.3 MB	

The list can be sorted by clicking on one of the column headers.

Viewing System Jobs

Clicking on Jobs at the top of the screen will allow the System Administrator to display the jobs that are currently being processed by the system. This feature is useful if there is a job that has been started but needs to be deleted because it is blocking other more important jobs.

Viewing System Exceptions

All system exceptions are logged in the Exceptions list.

Exceptions

#	Type	Message	Date
1	NoMethodError	"undefined method '[]' for nil:NilClass"	October 8 20 Days Ago
2	NoMethodError	"undefined method '[]' for nil:NilClass"	October 8 20 Days Ago
3	NoMethodError	"undefined method '[]' for nil:NilClass"	October 8 20 Days Ago

If an exception occurs, by clicking on the exception in the list the System Admin can view as well as download full details related to the exception. This information

should be set to Emcien Support as part of a Support ticket.

Emcien APIs

In this article, we'll cover what an API is, how to use authorization tokens to access Emcien accounts, and the information available through Emcien APIs.

Generally, APIs are used by programmers to set up an integration.

Not a programmer? Emcien can help you get things set up. Contact us at support@emcien.com.

What is an API?

An API, or "application programming interface," is a way for web programs to connect using code.

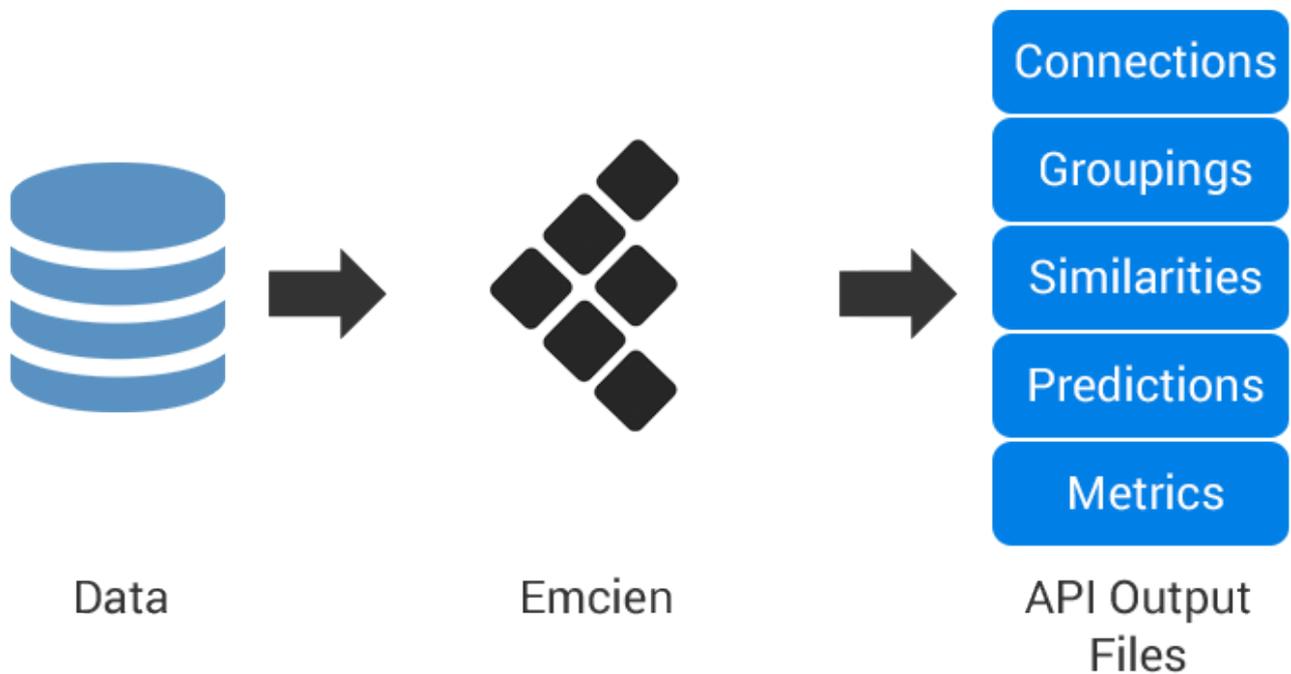
You can use Emcien APIs to perform some of the same tasks you can perform inside the Emcien app. You can also use Emcien APIs to integrate Emcien into your existing systems for end-to-end automation.

Using APIs in Your Workflow

You can use Emcien APIs to quickly integrate Emcien into your existing systems.

For example, you can use Emcien APIs in the following automated workflow:

1. Using an Emcien API call, your existing system sends a data set to Emcien for automated analysis.
2. Emcien automatically analyzes your data.
3. After analysis, your existing system uses an Emcien API call to request analysis results from Emcien. These results are packaged as API output files.



Generating Authorization Tokens

If you or your developer are using Emcien APIs, you will need to generate an authorization token.

Your authorization token works similarly to username and password credentials, granting you access to analysis results in Emcien. To access information through

any API endpoint or call, your authorization token must be included as a parameter in the url.

```
auth_token={your_auth_token}
```

Only Emcien administrators can generate API authorization tokens.

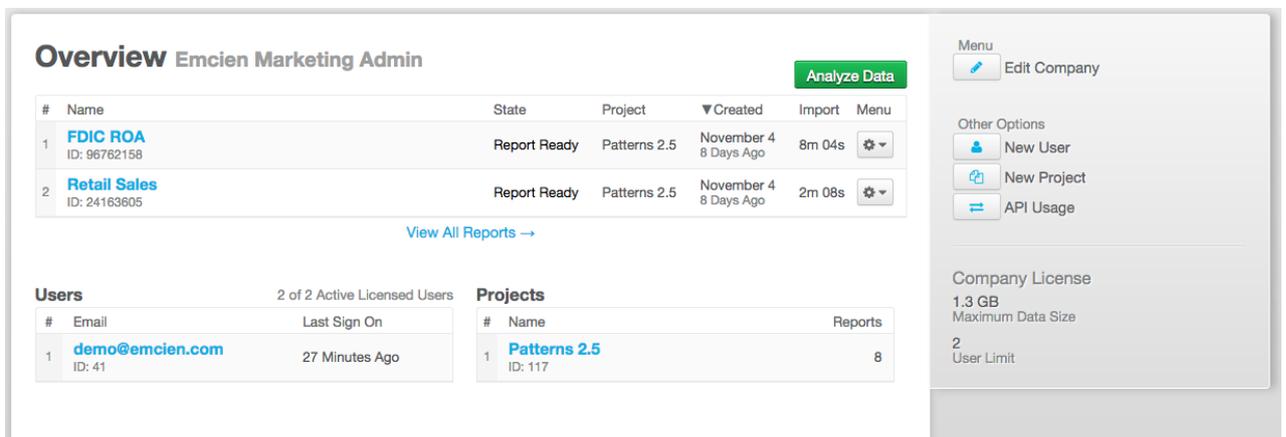
To generate an authorization token, follow the steps below.

1. On any Emcien page, click Admin in the upper left navigation.



Admin Settings | Help | Sign Out

The Admin Overview page is displayed.



The screenshot shows the 'Overview' page for 'Emcien Marketing Admin'. At the top right, there is a green 'Analyze Data' button. Below the header is a table with columns: #, Name, State, Project, Created, Import, and Menu. The table contains two rows: 'FDIC ROA' (ID: 96762158) and 'Retail Sales' (ID: 24163605). Below the table is a link 'View All Reports ->'. There are two summary cards: 'Users' (2 of 2 Active Licensed Users) and 'Projects' (1 project named 'Patterns 2.5' with 8 reports). On the right side, there is a 'Menu' section with 'Edit Company', 'Other Options' (New User, New Project, API Usage), and 'Company License' (1.3 GB Maximum Data Size, 2 User Limit).

2. On the Admin page, click Users in the upper right.



Overview | Reports | Projects | Users | Notifications

The Users page is displayed.

#	Email	Role	Last Session	Last Sign In IP	Status	Menu
1	demo@emcien.com ID: 41	Admin	29 Minutes Ago	10.1.1.155	Active	

3. On the Users page, hover over  beside the user and click Edit for whom you'd like to generate an authorization token.

The Edit User page is displayed.

demo@emcien.com

Email
Email for the user
demo@emcien.com

Password
Leave blank to keep existing password
Note: 8 characters minimum password length

API Token
Used to access the API (Optional)
9d25vn1xadtba928pzgczppu

User Role
Admin

+ Admin Notifications

4. If the user has a current authorization token, the authorization token is displayed in the API Token field.

API Token

Used to access the API (Optional)

Generate

To generate a new authorization token, click Generate. A new authorization token will be displayed.

Project List API Endpoint (JSON Only)

The Project List API endpoint returns a list of current Emcien projects. The Project List API endpoint is only available in the JSON format.

You can use the Project List API endpoint using the url below.

```
http://patterns.emcien.com/projects.json?auth_token=abc
```

Example JSON result:

```
{"id":1,"name":"Office  
Products","client":"ACME","created_at":"2013-10-22T14:15:16.000Z","rep  
ort_count":1,"read_access":true,"last_updated":"2013-10-22T17:47:16.00  
0Z"}
```

Report List API Endpoint (JSON Only)

The Report List API endpoint returns a list of reports and details for a project. Report List is only available in JSON format.

You can use the Report List API endpoint using the url below.

```
http://patterns.emcien.com/projects/{project_id}.json?auth_token=abc
```

Example JSON result:

```
{"id":2,"name":"Automotive","created_at":"2013-10-22T14:15:27.000Z","import_options":{"cluster_threshold":"3","max_cluster_size":"7","clean":false,"long_names":false,"item_communities":false},"reports":[{"id":7,"name":"Auto Parts","file_name":"ACME-Auto.csv","uploaded_at":"2013-10-22T17:46:15.000Z","uploaded_by":2,"item_count":91833,"basket_count":1269093}]
```

Available API Calls

The following calls are available using the Emcien API.

Call

Ranked Connections

Categories

Clusters

Items

Item Predictors

Substitutions

Example API Call

This example uses the Ranked Connections API call. The Affinities API endpoint returns a list of all connected items and the weight of their connection. The greater the weight value, the stronger the connection.

JSON Endpoint

```
http://patterns.emcien.com/reports/{report_id}/affinities.json?auth_token=abc
```

Example JSON Results

```
[ {"item1_id":"224641","item2_id":"436326","strength":0.620422}, {"item1_id":"436326","item2_id":"224641","strength":0.620422}, {"item1_id":"908295","item2_id":"147990","strength":0.975}, {"item1_id":"147990","item2_id":"908295","strength":0.975} ]
```

CSV Endpoint

```
http://patterns.emcien.com/reports/{report_id}/affinities.csv?auth_token=abc
```

Example CSV Result

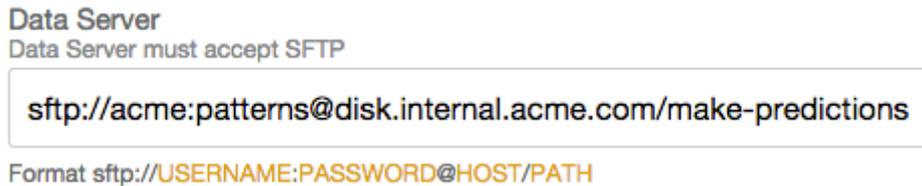
```
224641,436326,0.620422 436326,224641,0.620422 908295,147990,0.975  
147990,908295,0.975 908295,863564,0.711 863564,908295,0.711c
```

Response Redirection

The Ranked Connections API endpoint does not serve data directly; instead it redirects to a temporary link to a file. To retrieve this data with curl, pass it the -L flag to follow the redirect.

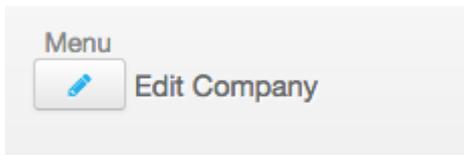
Downloading API Files

1. Click Admin in the upper right of any Emcien page.

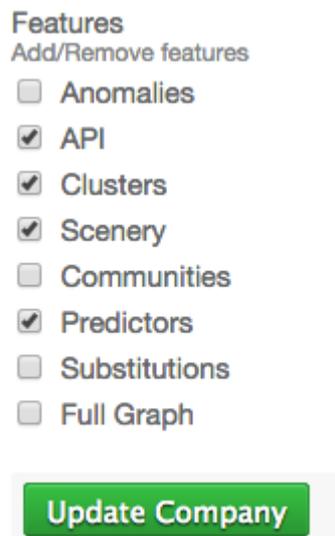


The Admin Overview page is displayed.

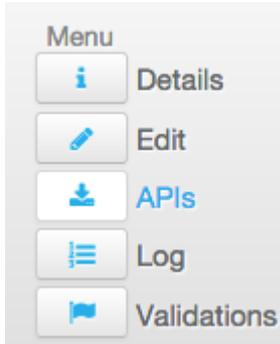
2. On the Admin Overview page, click Edit Company.



3. Ensure the API checkbox is checked and click Update Company.



4. Navigate to the report for which you'd like to download APIs.
5. On the Emcien Dashboard, click APIs.



6. Click URLs for any APIs you want to download.

Item Predictors API + Display Details

URL	Filename	Original Size
https://example.com/reports/60009970/api/predictors.csv	predictors.csv.gz	7 MB
https://example.com/reports/60009970/api/predictors.json	predictors.json.gz	16.4 MB

You can click Display Details to view what is included in the API.

[+ Display Details](#)

An example of the details for the Item Predictors API is included below.

Item Predictors API Description

Key	Data Type	Length/Prec	Description
predicted_item_descriptor	string	512	EmcienPatterns Unique ID (Formatted as: CategoryName::ItemID)
predicted_item_id	string	255	Item ID. This is the Item that is predicted.
predicted_item_name	string	255	Item Name
predicted_item_category_name	string	255	Name of the Category for the Item
predictor_probability	decimal	4	How likely the Item will be present when Predictor Items are present in a Transaction
predictor_ranking	decimal	4	Weighted value using predictor_probability and predictor_frequency
predictor_cluster_size	integer	--	Number of Items in Cluster
predictor_frequency	integer	--	Number of Transactions that contain this Cluster
predictor_frequency_trend	string	255	Time series for Transactions that contain this cluster
item_<N>_item_descriptor	string	512	EmcienPatterns Unique ID (Formatted as: CategoryName::ItemID)
item_<N>_id	string	255	Item ID
item_<N>_name	string	255	Item Name
item_<N>_category_name	string	255	Name of the Category for the Item

Thank You

A member of our team will be contacting you within the next 48 hours.

Contact Us

hello@emcien.com

404.961.6360

Fax: 404.793.7799

2859 Paces Ferry Road SE

Suite 300

Atlanta, Georgia 30339

Contact Us

The Emcien team is always here to help. Fill out the form below and we'll get your message to the right people.

Contact Us

support@emcien.com

404.961.6360

Fax: 404.793.7799

Office Hours

Monday through Friday

9am-6pm EST

Emcien Release Notes

Emcien 2.16

New Functionality

Prediction Reports: predictions have been separated from the Analysis, making the two-tiered prediction process more obvious to users. Reports include:

- Ability to perform searches on prediction details and rules

- Ability to download prediction details, transaction details and rules

Ability to perform a quick-look preview on data files from the New Analysis page.

Ability to auto-populate the prediction category name for wide format files

Ability to declare the file format of prediction test files from the New Prediction page

Enhancements to Existing Functionality

The “long” data format has replaced the “receipt” data format.

Users can now view all of the rules that were triggered during a particular prediction. In previous releases we only listed the top 5 rules that applied to the winning outcome.

Home page redesign:

- Easier access to in-process Reports and the most recent failed Reports

- Reports are displayed by type, rather than project

Enhancements to the usability and supportability of the Emcien VM

VM Upgrades: data from 2.14 and more recent VMs can be exported, and then imported into a 2.16 or later VM. The export and import processes are managed by Patterns bundles.

VM security has been greatly improved. Each VM installation comes with its own unique set of system user and perform user passwords.

Defects Resolved

Reports deleted while in the queue waiting to start processing now appear as deleted. Previously they appeared to process in spite of the fact that they had been deleted, but the 'completed' Report was not viewable. [PAT-826]

Situation where a Data Analysis failed on some data sets when a prediction category was declared and the API feature was enabled. Failure error was "Failed to Load Rules download" or "Load Rules failed" [PAT-973, PAT-1070]

The tool-tip for Negative Accuracy was incorrect and misleading.

Compatible Browsers

Chrome - version 42.0.2311.135

FireFox - version 37.0.2

Safari - version 8.0.8

IE 11

Emcien 2.15

New Functionality

Two new file formats have been added, Tagged and JSON

The JSON file format is a slight modification to the industry standard, great for large multi-attribute transactions

The Tagged format easily handles text data and is our most loosely structured format

Enhancements to Existing Functionality

A parse progress indicator has been added to the Import log in the format of an entry for every 1MM lines parsed, as well as a final entry indicating the total number of lines parsed.

A download id is now associated with all downloads. This id is part of the download name, making it easier for Users with several downloads to keep track of them on their computer.

Defects Resolved

Drag and drop capability to filter by Category on the Item search screen has been restored.

Clear, user friendly error messages now display when a UTF-16 file is improperly declared as a different file encoding

Compatible Browsers

Chrome - version 43.0.2357.124

FireFox - version 38.0.5

Safari - version 8.0.5

Emcien 2.14

New Functionality

The File Upload feature allows the upload of files directly into the company data server, eliminating the need to use an FTP client.

Prediction Rules can now be viewed and downloaded

A Prediction Rules page has been added to the Report. Rules are categorized to distinguish those that most influence the prediction results. Similar to the Predictions page, the Rules page is only visible when the Report was created with an Outcome Category defined.

A Prediction Rules API has been added to facilitate integration with external applications

A Prediction Rules search has been added to the Report

Enhancements to Existing Functionality

Prediction Feature

A prediction is only made against a transaction if there are rules that support the prediction. The prediction results indicate how many predictions were made against the test file, and how many transactions were skipped due to a lack of supporting rules.

The User now has the ability to specify the field delimiter and file encoding on prediction test files.

Data file management

Files in the company data server can now be renamed and deleted from

the New Analysis page.

The data file naming convention has been loosened to permit spaces.

Invalid files display in the file selector box on the New Analysis page, but cannot be selected.

The User Homepage has been redesigned to provide additional information about available Reports

Enhancements to the usability and supportability of the Emcien VM

The New Analysis page now displays available disk space. By providing this information we hope to prevent the User from running out of disk space while processing a data file.

The System license page displays a more specific status of the VM license

Defects Resolved

Failure messages have been improved to better communicate the root cause of the failure, enabling the User to take appropriate action

Compatible Browsers

Chrome - version 42.0.2311.135

FireFox - version 37.0.2

Safari - version 6.2.5 (8537.85.14.17)

IE 11

Emcien 2.13

Enhancements to Existing Functionality

The CSV formatted data file can now be delimited with either a comma, semi-colon, colon, tab, pipe or hash. Comma is the default delimiter. If a different delimiter has been used, it must be specified at time of import.

Navigating within the Report has been made easier by allowing the escape button and background clicking to close “Tell Me” and Scenery information windows.

During GUI data loading the browser tab changes color to reflect the processing state. This functionality is only supported by Chrome, Safari, and Firefox.

Enhancements to the usability and supportability of the EmcienVM

The process for installing a VM and creating the initial company, System Admin user and data URI has been simplified

Auto login for the initial System Admin user

Updates to the VM license terms are now made over the internet, without the need for Emcien Support to send the customer a new license file. The System Admin menu contains a license page that displays license status, license terms, and a history of previous licenses.

The complete hardware profile of the host machine is now displayed on the Company Admin page

A Diagnostics page has been added to the System Admin menu to facilitate trouble shooting when a customer encounters a problem

Defects Resolved

Single quotes (an apostrophe) in the company name are displayed as browser

escape characters.

Wide formatted data files that contain an empty order will fail during the import. An empty order is one that does not contain items. This is characteristic of an order deletion in some Retail environments.

Compatible Browsers

Chrome version 41.0.2272.104

Firefox version 36.0.4

Safari version 6.2.4 (8537.85.13.7)

IE 11

Emcien 2.12

New Functionality

Non-user facing architectural improvements

Enhancements to Existing Functionality

Correction to error message regarding price on receipt format

Compatible Browsers

Chrome version 40.0.2214.91

Firefox version 35.0

Safari version 6.1.3 (8537.75.14)

IE 10

Emcien 2.11

New Functionality

The Partial Import Option has been returned to the New Analysis page. The ability to import a subset of your file was removed from the previous release.

Enhancements to Existing Functionality

Terminology was revised in several UI screens and CSV downloads to more accurately describe the Analysis data.

Several enhancements were made to the Predictions feature to make it easier to make a prediction and understand the results of the prediction.

- Predictors API documentation and downloads updated to specify the format of the unique Item identifier

- Replaced the Dashboard link to make a new Prediction with a link to view existing Predictions.

- Added the ability for a User to perform “quick predictions”. In this situation only the Summary file is generated.

- A “Tell Me” was added to the Predictions Overview page to describe the prediction results for each Outcome Item.

- The Item pill is now active on the Prediction Details page, enabling the User to navigate to the Item Detail page while viewing Prediction results.

The Browser Tab now identifies the specific Report screen the User is viewing, along with the Report name. Prior to this release the Tab only indicated the Report screen followed by “EmcienPatterns”. Stating the Report name facilitates Users who are viewing multiple Reports at a time.

Resolved Issues

Manually aborted prediction jobs now indicate that the “Prediction was terminated by the System Admin”.

Corrections made to error messages to more accurately describe the issue preventing import of the data file

Maximum transaction size for wide files is no longer limited to 100 items. It is now 1000 items.

Compatible Browsers

Chrome version 40.0.2214.91

Firefox version 35.0

Safari version 6.1.3 (8537.75.14)

IE 10

Emcien 2.10

New Functionality

The Predict feature is now fully integrated into Emcien. To make use of this feature a data file is uploaded for analysis with instructions for Emcien to learn which rules (i.e. actions, events, circumstances) produce specific outcomes in a particular category (the prediction category). Predictions can then be made against future data files to determine whether or not the outcomes in the prediction category will occur.

Enhancements to Existing Functionality

The cluster-trend calculation process now makes use of multiple processors, when available, to reduce processing time.

More informative error text when data file format issues prevent the file from successfully loading.

Resolved Issues

Issue preventing the System Admin from downloading the Exception log has been resolved.

Compatible Browsers

Chrome version 39.0.2171.95

Firefox version 33.1.1

Safari version 6.1.3 (8537.75.14)

IE 9 and 10

Emcien 2.9

New Functionality

Emcien now supports LDAP. Many companies utilize an LDAP server to maintain the employee directory along with other data and records related to their organization. LDAP enables 'single signon' to company systems. By integrating with a company's LDAP server, Emcien enables users to sign in using the normal username and password the individual uses to access other company services.

We have implemented functionality that allows Sandbox users to interact with

Emcien Support through the Emcien application.

Enhancements to Existing Functionality

The Emcien import log displays the size of the file that is loading. Compressed files are expressed by their compressed size. Although such files are normally named with a “gzip” extension, there are times when the extension is absent. To ensure that the Admin and Data Loader realize when a compressed file is loading, the import log now indicates “gzipped” next to the local file size.

The import queue message has been improved to take into account the file that is currently processing as well as the situation where a file may complete processing earlier than one that entered the queue before it.

Resolved Issues

When a User’s feed server account password expires, Emcien is unable to connect to the company data server to process a GUI load. The application previously displayed a 500 error in this situation. Emcien now displays a message indicating that there is an error connecting to the data server and suggests that the password may be expired.

The text on the Item Detail page that describes the slope of the Item transaction trend has been corrected.

Predictor probabilities less than 1% now accurately display; they previously rounded down to 0%.

Compatible Browsers

Chrome version 38.0.2125.122

Firefox version 33.0.3

Safari version 6.1.3 (8537.75.14)

IE9 and IE10

Please note that Emcien no longer supports IE8.

Emcien 2.8

New Functionality

Emcien is now able to generate a list of predictive rules from a data set. The rules can be exported and used by EmcienPredict to provide predictions for similar data sets. This feature is available to Companies that have the Predictors feature enabled. The rule set to be generated is pre-defined for auto-load projects, and defined at data loading time when GUI loading.

Enhancements to Existing Functionality

All of the Item's connections (ranked connections, clusters and initial connections) are now accessible from the Item Detail page. By adding multiple views to this page the user is able to obtain a full understanding of the Item's relationships without navigating away from the Item Detail.

Admin and System Admin email notification settings have been moved under the new/edit User form. Email notifications are restricted to Admins only.

Compatible Browsers

Chrome version 38.0.2125.111

Firefox version 33.0.2

Safari version 6.1.3 (8537.75.14)

Emcien 2.7

New Functionality

Item Scenery provides a list, by Category, of every Item the focus Item has been involved in a transaction with. The list of “Scenery” Items includes both pruned and unpruned connections.

Enhancements to Existing Functionality

The Company Projects page now displays the count of Ready reports in each project.

Transaction statistics have been moved from the Trends dashboard to the Transaction Size dashboard.

We now display a more informative data parsing error message when the data file has an invalid line ending.

Added placeholder text below the Search criteria text box to assist Users with entering valid search criteria for advanced searches.

Compatible Browsers

Chrome version 38.0.2125.104

Firefox version 32.0.3

Safari version 6.1.3 (8537.75.14)

IE9 and IE10

Emcien 2.6

New Functionality

Data Loader role: A new role has been added to allow non-Admins to load data and create, edit and delete Reports.

Emcien now accepts UTF-8 files with a BOM. The BOM is detected by the application and stripped off.

An import queue message now displays when multiple files are queued for analysis. If a User loads data to create a Report and the Report is not actively processing due to other Reports in line ahead of it, the message will state how many files are ahead in the queue.

A notification email is sent when an invalid file is uploaded to the auto loader directory. The email alerts the User that the file will not load.

Enhancements to Existing Functionality

The Downloads page has been removed from the Admin view. Each User has the ability to access their downloads from their Homepage.

An index has been added to the Reports page to facilitate keeping count of, and sorting through, long lists of Reports.

Admin pages have been redesigned to improve readability and menu access.

Resolved Issues

The Cluster Scenery “Download CSV” button no longer appears grayed out.

Compatible Browsers

Chrome version 36.0.1985.125

Firefox version 31.0

Safari version 6.1.3 (8537.75.14)

IE9 and IE10

Emcien 2.5

New Functionality

Anomalies Feature: enables Companies to monitor a specific category or attribute of their data to gain insight on when the data deviates from normal, or expected, behavioral patterns.

Tail Report: enables the analysis of data over time, with a deep dive into the active time period. Companies are able to analyze and view trends in Item and Category transaction history, but the detailed analysis is only performed on Items and categories in the final period of data.

Information on available processor count is now displayed on the System overview page for Companies operating Emcien as a virtual appliance

An API Usage section was added to the API page in the Admin area to assist Users with programmatically making API calls.

It is now possible for the Admin to auto-generate the User auth token

A control has been added to close the Item pill dropdown menu.

Improved definitions and display of key metrics

A “Tell Me” feature was added to explain each Cluster in the analysis results.

Allow the import of data files other than UTF-8 character encoding. Please see Admin Guide and Data Requirements document for encoding formats accepted

Enhancements to Existing Functionality

The date field requirement has been extended to accept date and time.

Emcien now accepts both ISO and Unix timestamps.

The defect that prevented the application from recognizing when a change was made to existing Search filters has been corrected

Ability for System Admins to set the number of users a Company can have and the maximum file size that can be loaded

The “last updated” date now displays for Companies and Projects instead of the “Created at” date.

A visual alert has been added to the Admin Project page to inform them when there is a problem that will prevent auto loaded files from processing

A manual start button has been added to allow Admins to override the scheduled autoloader run time

Homepage now displays the end date for Reports when the data contains dates. This is particularly beneficial for Tail Reports

Transaction distribution is now provided graphically and numerically for the Day of Week Eccentric Categories

Changes to the Dashboard Trend chart to improve readability

Stricter adherence to the use of 0 (nothing is affected) and 100 (everything is affected)

The User can now select the time interval the data should be analyzed against.

Compatible Browsers

Chrome version 36.0.1985.125

Firefox version 31.0

Safari version 6.1.3 (8537.75.14)

IE9 and IE10

Emcien 2.4

New Functionality

User Interface: Changed appearance and added functionality to enhance the user experience. Some of these changes include:

- Warn user when they are about to lock out their account

- Cluster Scenery has been added to the application as a new feature.

 - This is the ability to peer into the details of a Cluster and see the Items that appeared in transactions with the Cluster when the feature has been enabled. For more details please see the latest Training Guide or Help files.

Admin Interface: Added functionality to facilitate administration of Company environment.

- License now displays the maximum number of users allowed as well as the maximum file size

- 30 minute lockout of accounts that fail to sign in properly after 3 failed attempts

- Administrators have the ability to unlock user accounts

- Track and display server exceptions in System GUI for VM Clients

Technology updates: Made changes to improve processing and general throughput.

Increased available API calls (see below), and added functionality to deliver a compressed data file for each call: Affinities, Categories, Clusters, Items, Item Predictors, Substitutes

Compatible Browsers

- Chrome version 35.0.1916.153

- Firefox version 29.0.1

Safari version 6.1.3 (8537.75.14)

IE9 and IE10

Emcien 2.3

New Functionality

User Interface: Changed appearance and added functionality to enhance the user experience. Some of these changes include:

- Connections map has been added to the dashboard to provide an overview of patterns within the report

- Retail Substitutions feature has been improved and returned to the solution.

- Item Predictors feature added to predict the probability of an Item appearing in a cluster.

- Substitutions API is provided to obtain Substitutions data

- Improved search capabilities and clarified results being returned

- Advanced Search criteria added to use "OR" as well as "AND"

- Added a Glossary section to the online Help

- Lock Step items can now be filtered in the Cluster Search

- Improved text formatting on dashboard Chord Chart

Technology updates: Made changes to improve processing and general throughput.

- Improved Engine processing to reduce report creation by almost 50%

- Increased the maximum length of item names

Report ID no longer auto-increments by 1. Reports are now assigned a random number.

Compatible Browsers

Chrome version 33.0.1750.152

Firefox version 28.0

Safari version 6.1.2 (8537.74.9)

IE9 and IE10

Emcien 2.2

New Functionality

User Interface: Changed appearance and added functionality to enhance the user experience. Some of these changes include:

- redesigned chord chart

- the ability to drag and drop categories or items onto the search filter

- additional views of daily, weekly, monthly or yearly of the Trends screen if the data supports it

- pop up on clusters screen to display long names or ids

- new search criteria of anchor, not anchor, and without

Technology updates: Made changes to improve processing and general throughput. Added more detailed logging.

Data Formatting: Added new Emcien Wide input format to make data loading easier. Changed file naming format to allow processing of either Receipt (legacy) or Wide formatted files. Also added new date formatting to include: ISO (Y-M-D) and AUS (D-M-Y).

Bandit Release Notes

Bandit has been certified in the following operating systems:

OSX

Linux RedHat

Windows 7 & 8

Bandit 52

Defects Resolved

Fixed the twist to receipt command (-f R).

Failed when trying to band a data set with a dependent variable (numeric or categorical).

Z_get_token_real: (Zero) is not a number (row 7, col 0).

Bandit 55

New Functionality

Enabled banding of 'Long' files (-i R)

Automatically detects numbers that should be categoral data (e.g. phone numbers, social security numbers

The command (-g 2) will enable this feature

The command (-g 1) will skip this logic, but will designate single integers as categoral

Bandit can now prune categories that share little mutual information with the dependent category (-p 1)

These categories will NOT appear in the banded file, but will appear in the breaks and mi files

Enhanced Functionality

Bandit will now round breakpoints of bands to the nearest 2 decimal places

Now sampling up to 10,000,000 lines when banding large files

Improvements made with respect to file size and estimations of the number of rows in a data file

Improved speed on Windows machines

Better 'bad' data detection by the flagging of "silly" numbers as missing data

Defects Resolved

Bandit for Red Hat and Windows will now place the correct bracket labels around bins

Escaped quotes can now be handled, and the missing field issue has been resolved

File headers will no longer cause errors due to skipping the line ending

Corrected instances of auto bands being created with size of zero

Columns containing percent sign are now banding correctly

Previous Versions

Please click on the link below to download a copy of the previous version of the knowledgebase:

[2.14](#)

