

HOL Guide for Enterprise Risk Analysis

USING DATAMEER, HDINSIGHT, TRENDMICRO DEEP SECURITY
AND CHEF

The purpose of this section is to capture all changes made to the content of the document.

Contact for Enquiries and Proposed Changes

If you have any questions regarding this document, please contact:

Email Address	azuremarketplace@avyanconsulting.com
---------------	--

1 Table of Contents

1	Overview	3
2	How to deploy this solution	3
3	How to configure the components.....	7
3.1	Datameer.....	7
3.2	TrendMicro.....	7
4	Signing into Datameer UI	8
5	Configure Datameer to Fetch Data from Azure Storage	10
6	Link, Clean and Prepare the Data	14
7	Perform Analysis to Identify Outliers.....	25
8	Logging in to the TrendMicro DSM	43
8.1	Server name	43
8.2	Server login.....	43
9	Perform policy configuration on the TrendMicro DSM	44
10	Exercises.....	46
10.1	Datameer – Visualize the Data	46
10.2	TrendMicro – Malware test.....	46
11	Visualize the Data	46
12	Malware Test	49
12.1	Generating Malware alert in the computer	49
12.2	Dashboard – Malware Alert	50
12.3	Malware Alert verification	50
13	References, Attachments & Definitions	51
13.1	References.....	51

1 Overview

The purpose of this document is to provide the step-by-step instructions of deploying and configuring the Enterprise Risk Analysis using Datameer Business Intelligence and TrendMicro DeepSecurity solution and lab exercises.

The exercises includes creation of credit fraud risk awareness using sample (representative) data, building powerful Infographics of the Datameer and the security intelligence in the malware detection of the TrendMicro DeepSecurity

2 How to deploy this solution

This section will provide you the details of how to deploy this solution in the Microsoft Azure

- 1) Go to the below link available in the Github

<https://github.com/AvyanConsultingCorp/azure-quickstart-templates/tree/master/datameer-trend-chef-businessintelligence>

- 2) Click on the “Deploy to Azure” in the page, this will take you to the page where you need to provide the parameters



- 3) Provide the custom parameters for the solution accordingly and click “Next”

Microsoft Azure

New > Custom deployment > Parameters

Custom deployment
Deploy from a custom template

Parameters
Customize your template parameters

* Template
Edit template

* Parameters
Edit parameters

* Subscription
datameerhdi

* Resource group
demo

* Resource group location
West US

* Legal terms
Review legal terms

☐ Pin to dashboard

Create

LOCATION (string) ①
westus

HDICLUSTERTYPE (string) ①
hadoop

HDICLUSTERNAME (string) ①
datameerhdi

HDICLUSTERLOGINUSERNAME (string) ①
demo

* HDICLUSTERLOGINPASSWORD (securestring) ①
[Red arrow pointing to the password field]

HDISSHUSERNAME (string) ①
demo

* HDISHPASSWORD (securestring) ①
[Red arrow pointing to the password field]

HDISTORAGEACCOUNT (string) ①
datameerhdi

OK

- 4) You need to select the subscription you want to deploy this solution

Microsoft Azure

New > Custom deployment

Custom deployment
Deploy from a custom template

* Template
Edit template

* Parameters
Edit parameters

* Subscription
[Red arrow pointing to the subscription dropdown]

* Resource group ①
demo

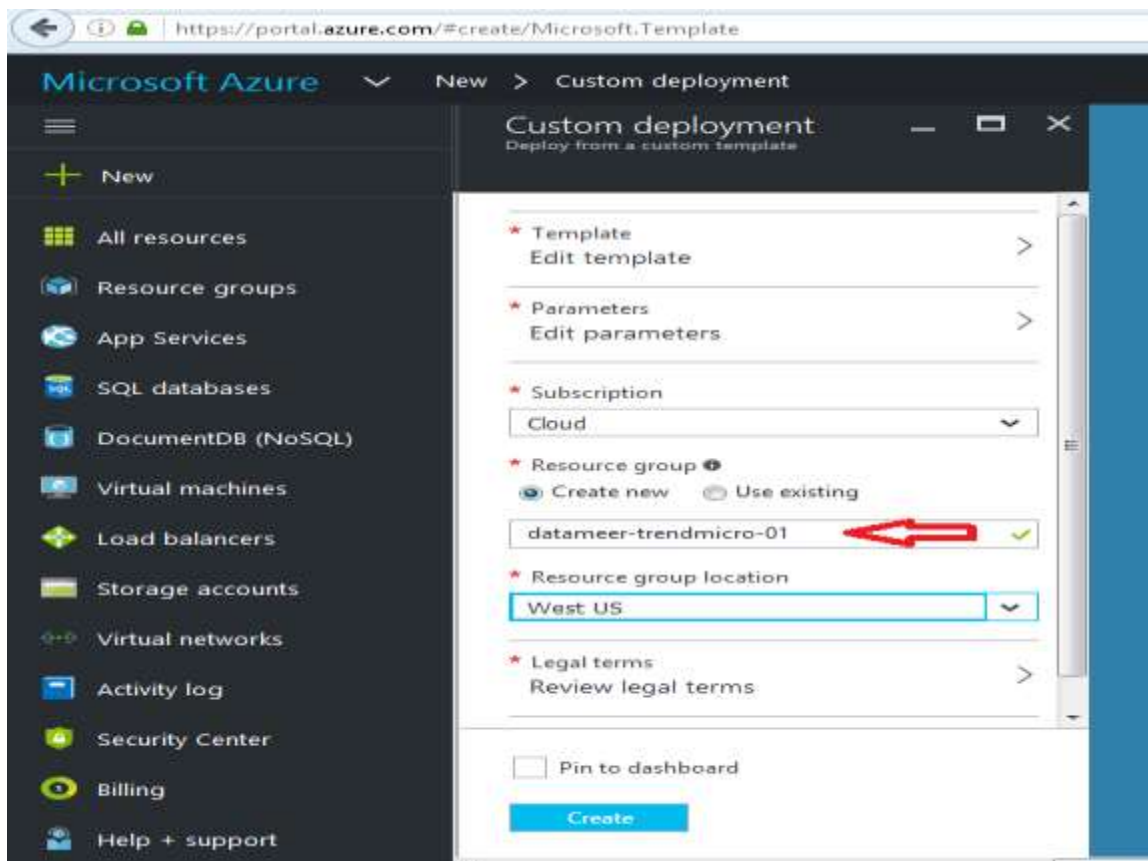
* Resource group location
West US

* Legal terms
Review legal terms

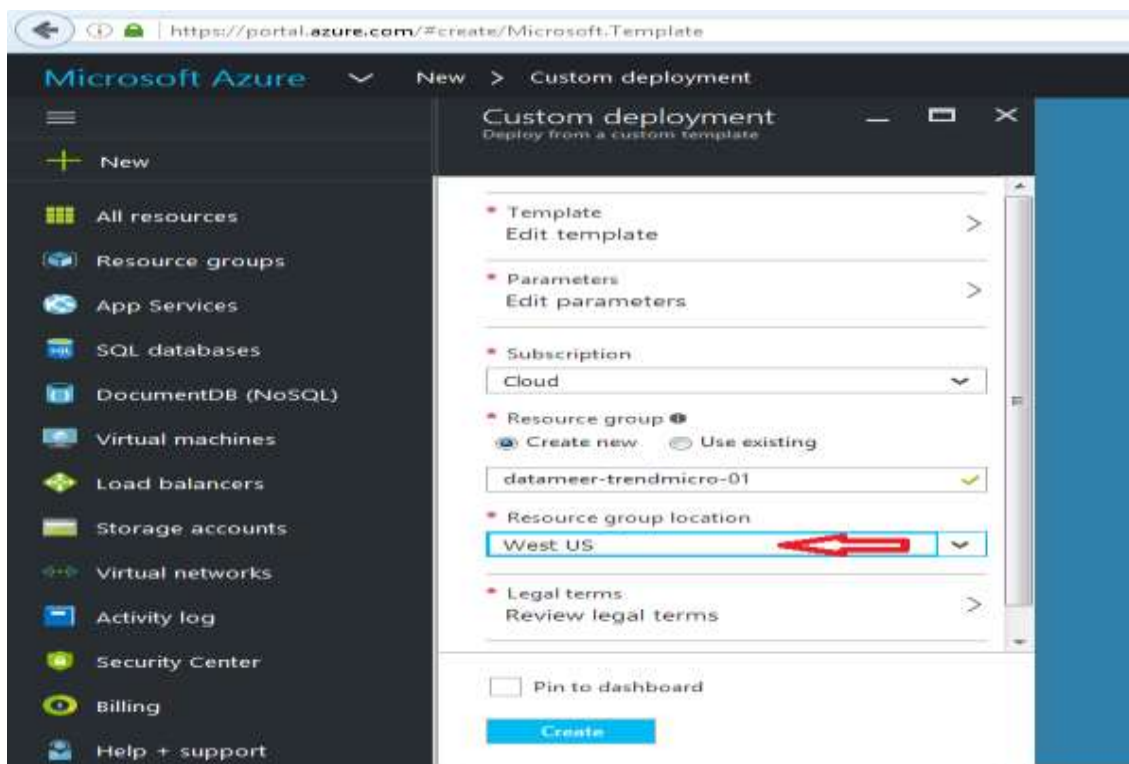
☐ Pin to dashboard

Create

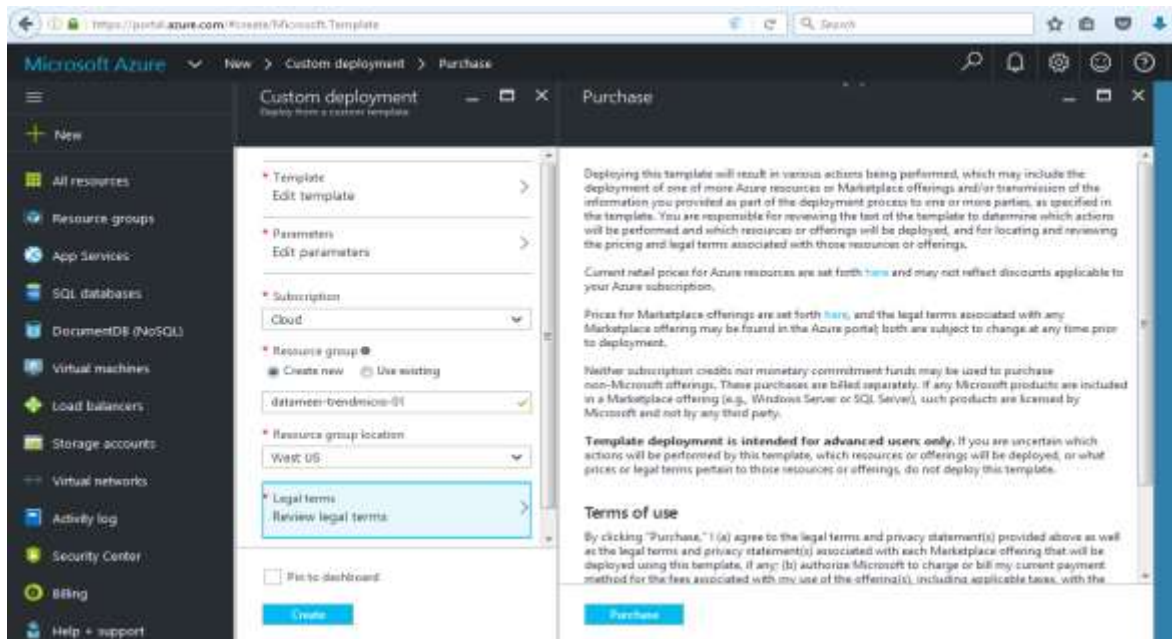
- 5) Either you can create a new “Resource Group” or use the existing resource group to deploy this solution



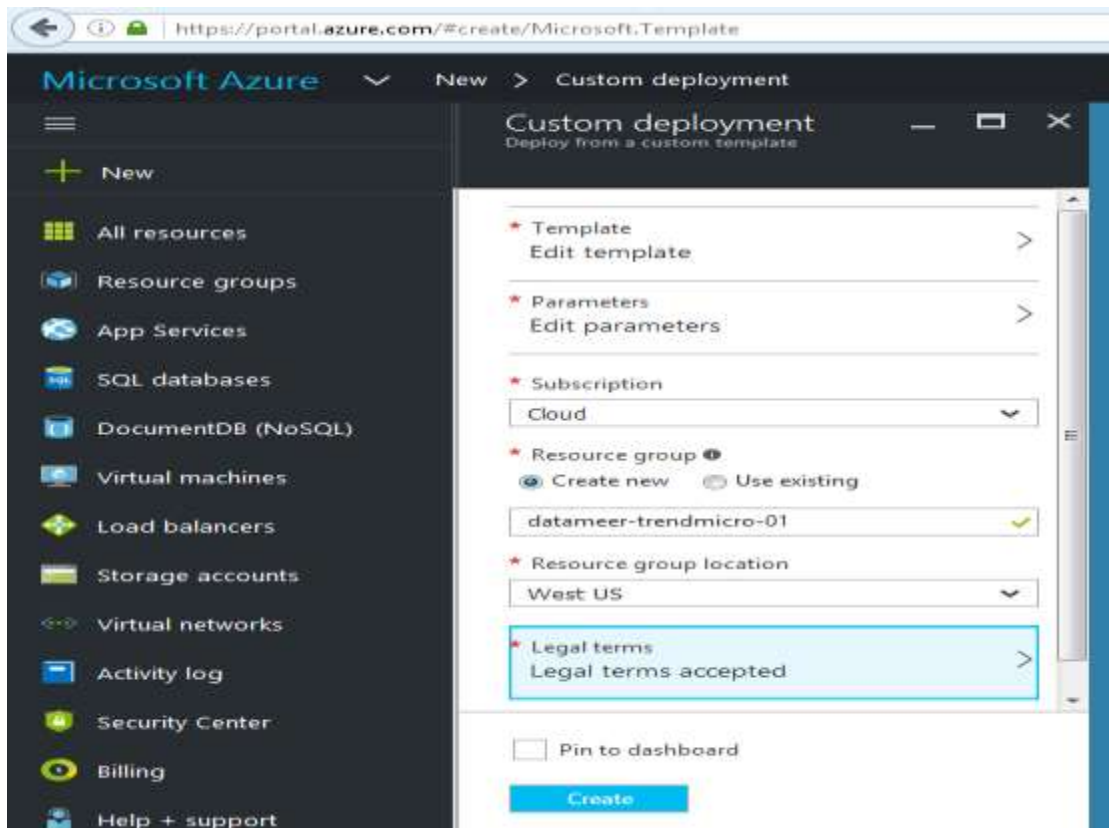
- 6) Select your choice of Region to deploy this solution,



- 7) Accept the legal terms to deply the products from the Azure marketplace which includes, Datameer, TrendMicro and Chef. Click on the “Purchase” button for the same.



- 8) Click on the “Create” button to start deply the solution now



3 How to configure the components

3.1 Datameer

Datameer is the product used for the Big Data Analysis. It can be used many types of data and can connect to different data sources like storage, database etc. In this solution, the data (.csv) from the azure blob storage will be used too identify the Fraud detection using the credit card. The below sections will provide the details of the configuration of the data in the Datameer for the Big Data Analysis

3.2 TrendMicro

TrendMicro is the industry leading security product, which has the capabilities of

- Anti-Virus/Anti-malware detection and prevention.
- Web reputation
- Host based firewall
- Host based Intrusion detection and prevention
- File Integrity monitoring
- Log Inspection

TrendMicro DeepSecurity is an agent based security solution which will help the organisations to comply with all their security requirements.

This is solution, showcases the Anti-Malware capabilities of the TrendMicro deepSecurity and below sections will provide the details of the configuration on the same.

4 Signing into Datameer UI

Copy samples to your storage account

Typically an enterprise will send the payment and other transactions to a data lake. For the purposes of a Hand-on-lab, our team has created a samples file with approx. 1.5 Million records and is made available to you during the ignite 2016 time period here

https://msignite2016stg.blob.core.windows.net/samples/online-transactions-cc_masked.csv

If for some reason this location is not accessible to you, please do not hesitate to reach out to us @ azuremarketplace@avyanconsulting.com

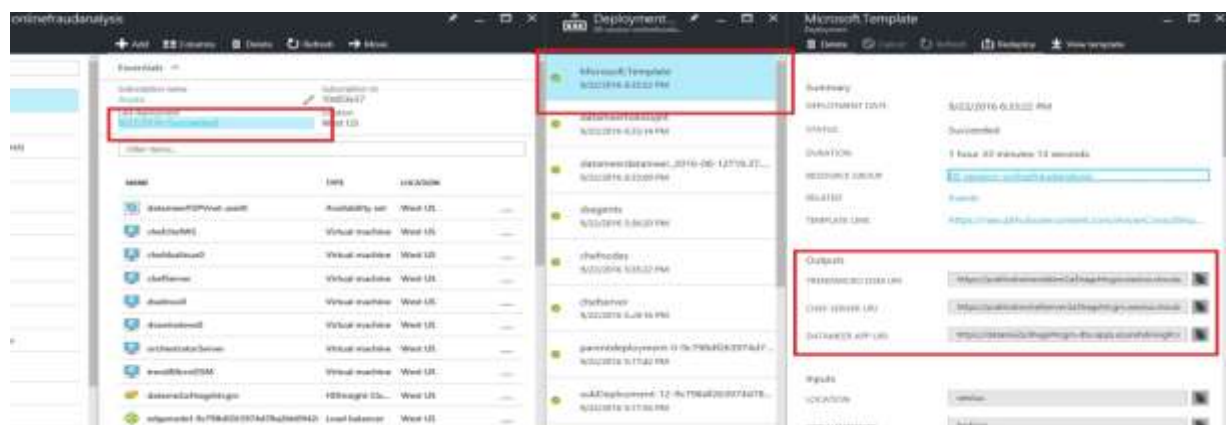
1. Download the file to your desktop
2. Navigate to the deployment resourcegroup and open the datameer storage account
3. Create a new blob container called "samples"
4. Upload the file online-transactions-cc_masked.csv to this container.

The fraud analysis is performed with the Business Analytics components of the solution, and namely Datameer and Azure HDInsight. All steps are executed in the Datameer UI. There are two ways that you can use to access the Datameer UI:

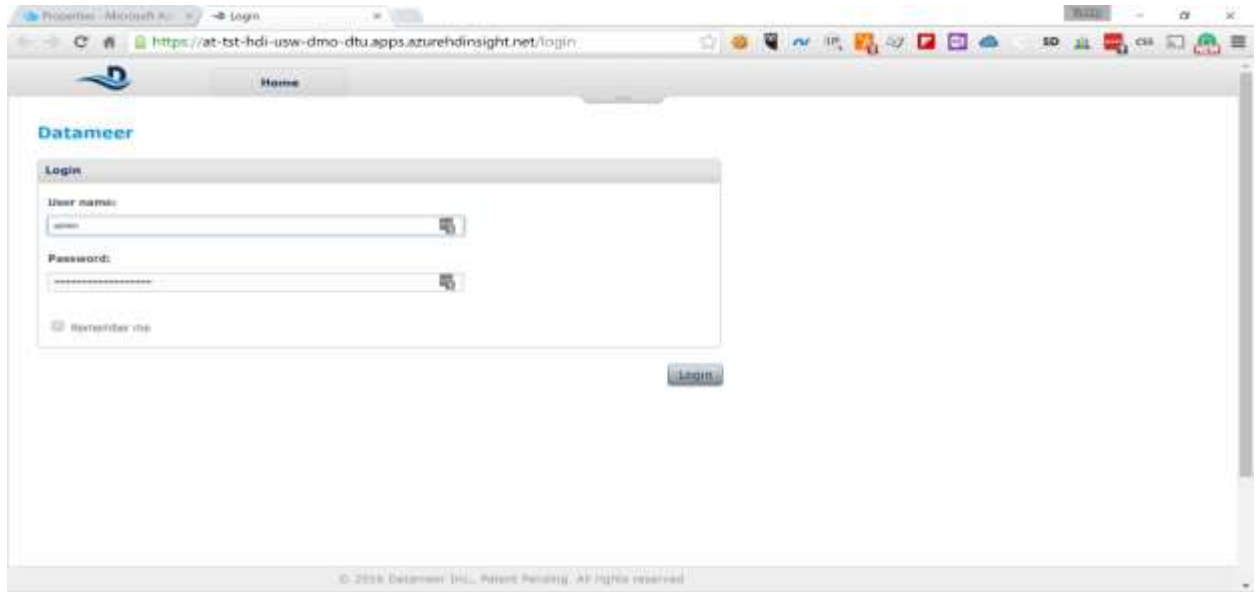
- Accessing it directly via the UI URL
- Accessing it via the Azure Management Portal

For the purposes of this HOL we will access the Datameer UI from the Azure Management Portal. Follow these steps:

1. In Azure Management Portal (<http://portal.azure.com>)
 - a. Click on specific resource group
 - b. Click on Last deployment date
 - c. Click on the Microsoft Template
 - d. Copy the Datameer URI from the Outputs



2. Paste the URI in your browser to load the Datameer UI



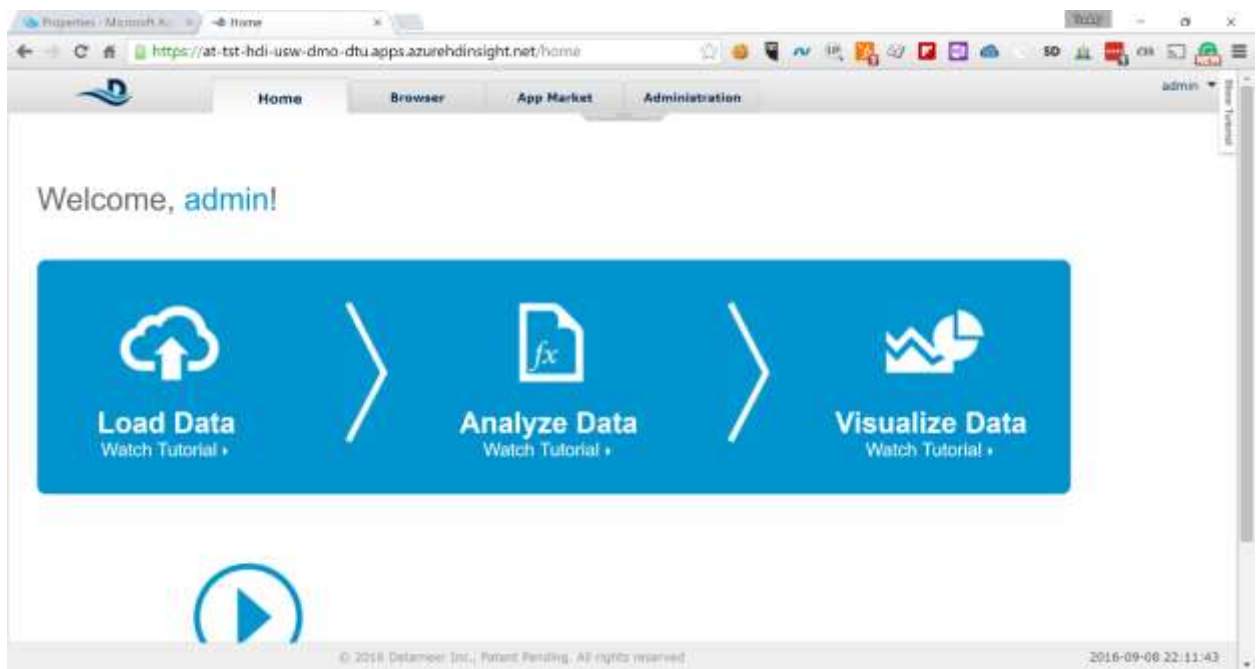
You will be prompted to sign in using your Datameer username and password

3. Sign into Datameer UI using the following default credentials:

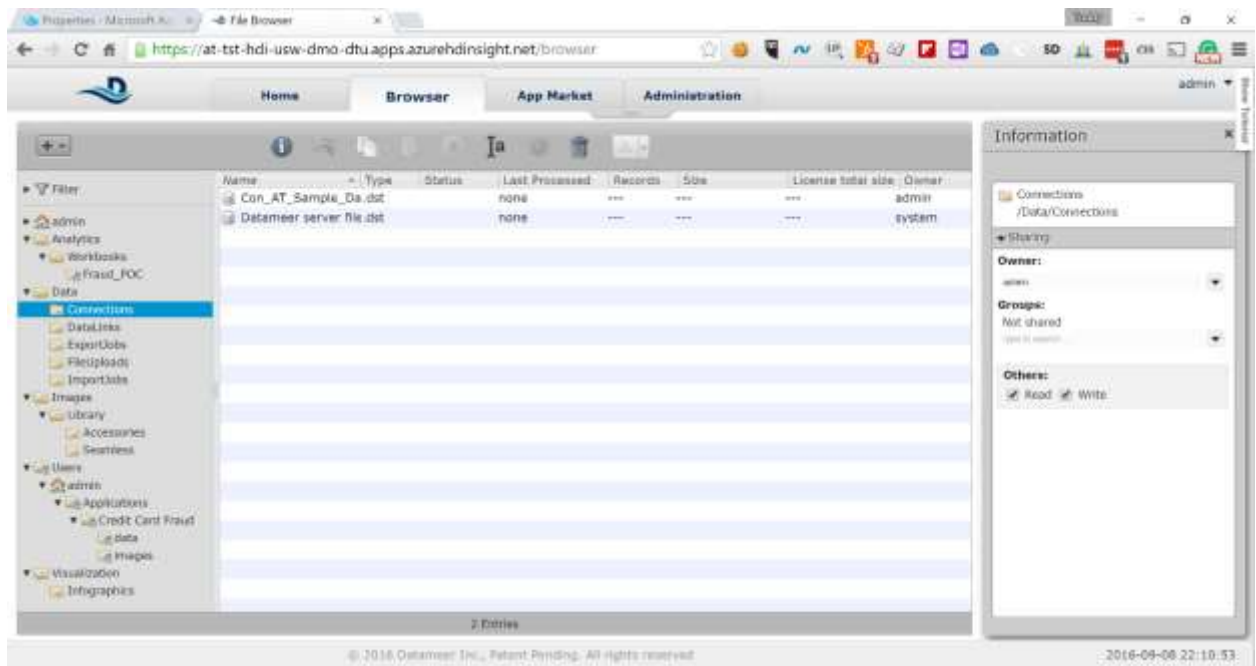
username: *admin*

password: *admin*

You will see the Welcome Screen for Datameer and an introduction video will pop up



4. Close the introduction video pop-up and click on the Browse tab

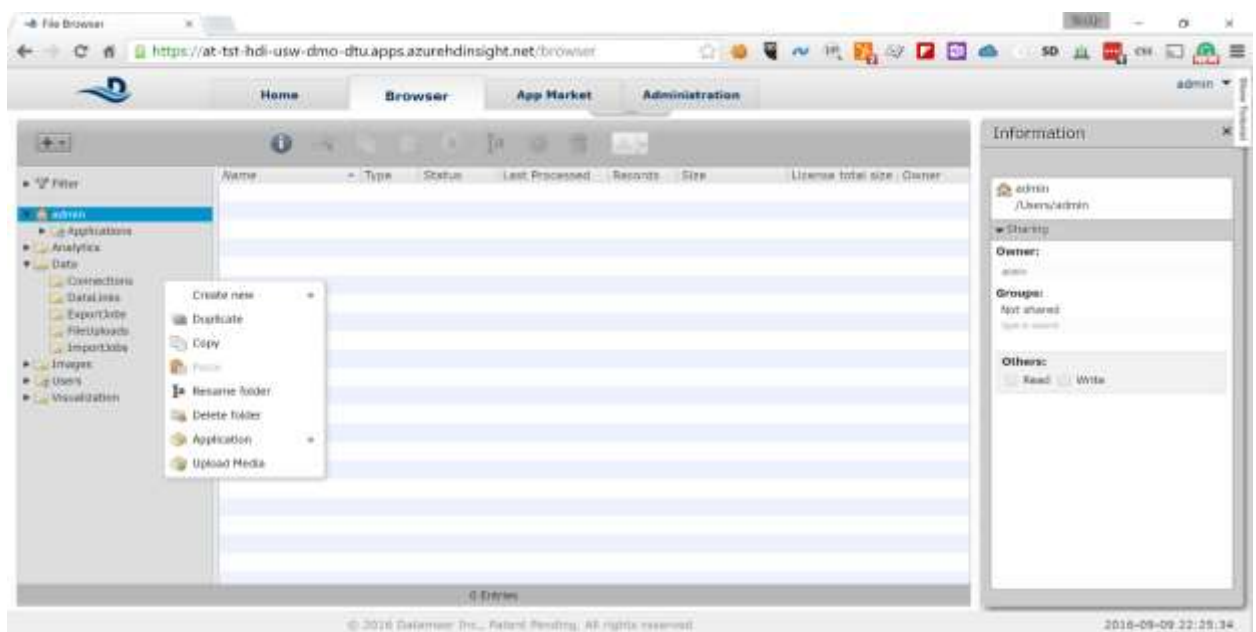


5 Configure Datameer to Fetch Data from Azure Storage

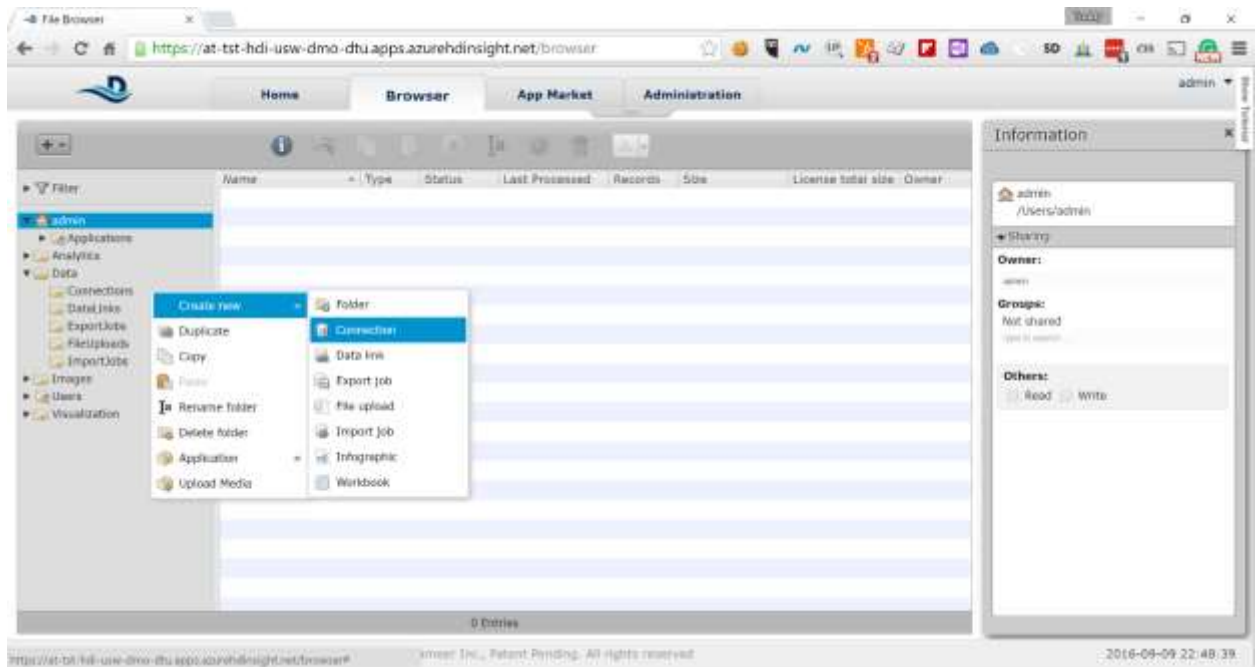
Datameer has more than 65 connectors built in, that allow various systems as data sources. For the purpose of this HOL we will use the Azure Storage connector and fetch the data from there. The assumption is that you have storage account data that contains the transaction data.

In order to configure Datameer to fetch the data from Azure Storage account you need to go through the following steps:

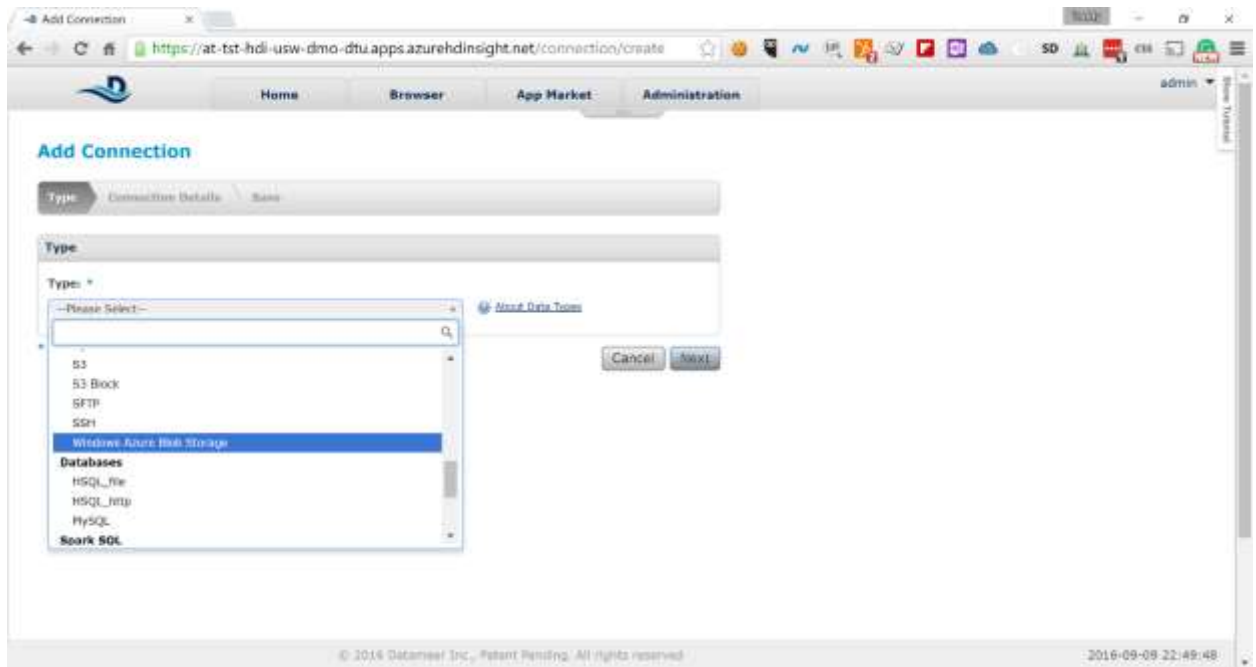
1. Expand the *Data* node in the left-side navigation and right-click on *Connections*



2. Select *Create new* -> *Connection*



3. In the *Type* drop-down, scroll down to *File* section and select *Windows Azure Blob Storage*



4. Click *Next* and fill in the following information on the next screen

The screenshot shows a web browser window with the URL <https://at-tst-hdi-usw-dmo-dtu.apps.azurehdinsight.net/connection/create/con>. The page title is "Add Connection - Windows Azure Blob Storage". The navigation bar includes "Home", "Browser", "App Market", and "Administration". The user is logged in as "admin". The form has two tabs: "Type" and "Connection Details". The "Connection Details" tab is active, showing the following fields:

- Storage name:** * (required) [text input] Name of the storage account. <http://mystorage.blob.core.windows.net/>
- Container name:** * (required) [text input] Name of the storage container
- Access key:** * (required) [text input] Storage access key
- Protocol:** * (required) [dropdown menu] Choose to use a secure or non-secure data transfer channel.
- Connection usage:** [dropdown menu] Import/Export

At the bottom of the form are "Cancel", "Back", and "Next" buttons. A footer note states: "© 2016 Delameter Inc., Patent Pending. All rights reserved." The timestamp "2016-09-09 22:52:18" is visible in the bottom right corner.

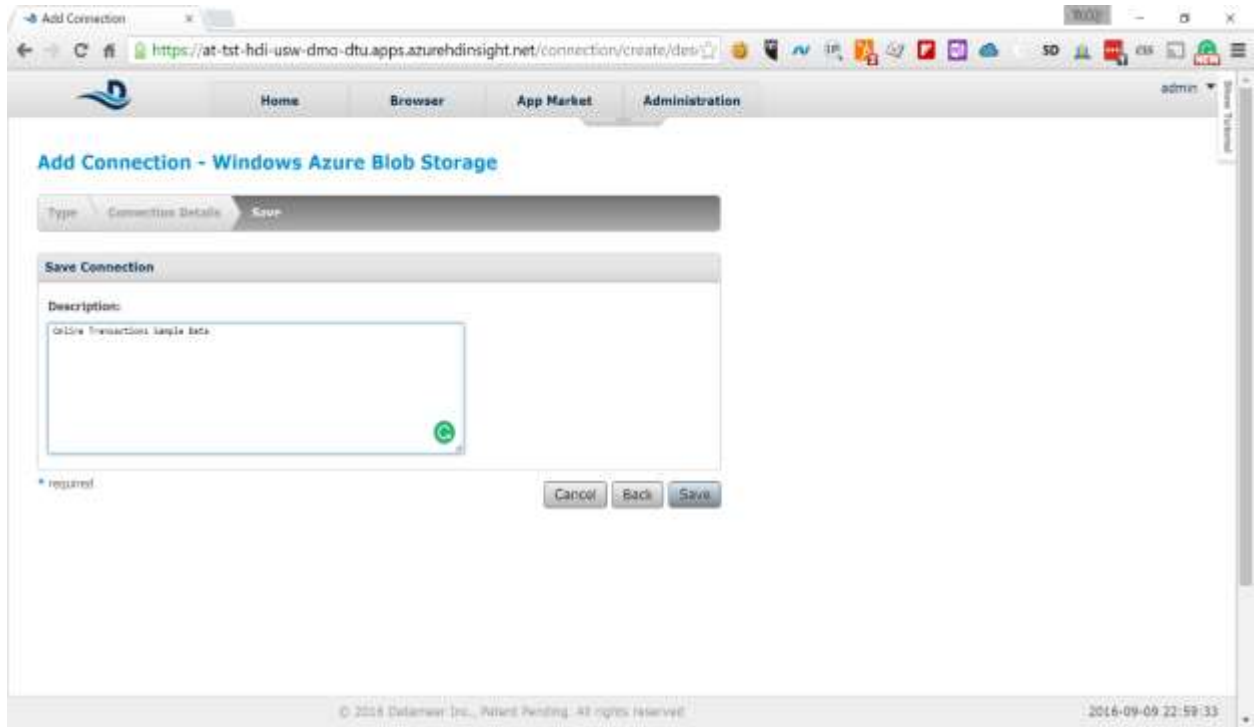
Storage Name: The name of the storage account where your data is

Container Name: The name of the container where your data is

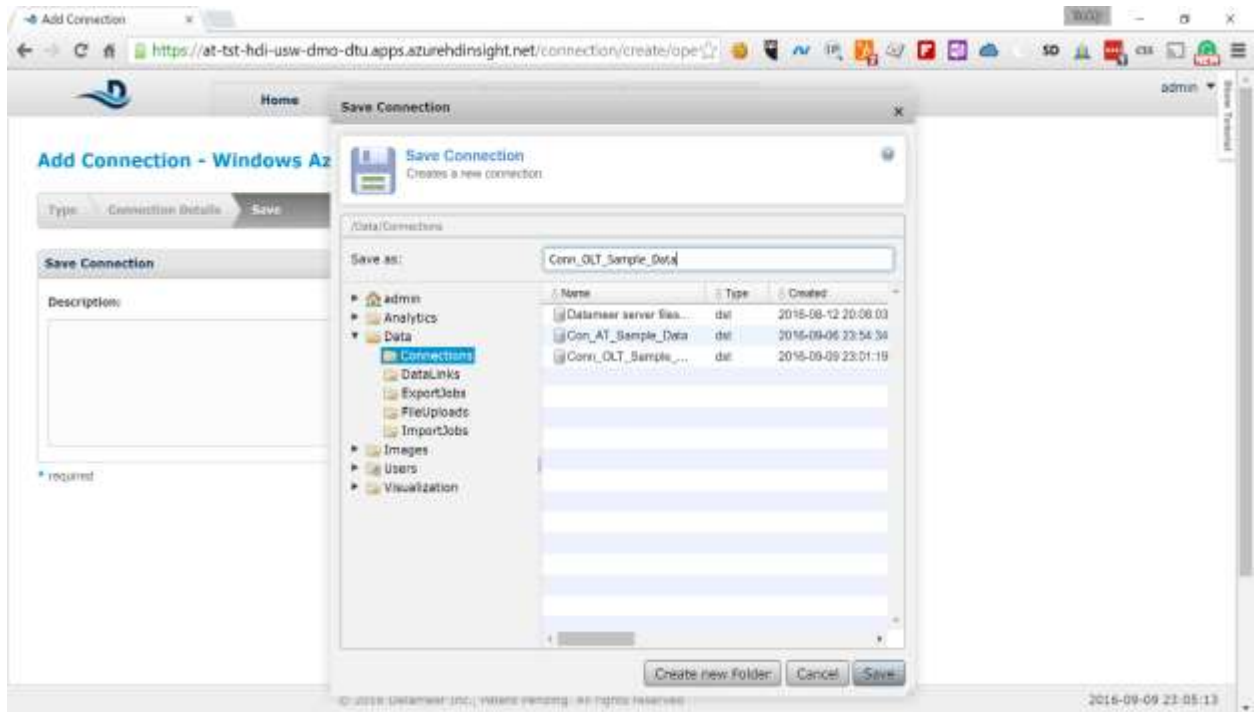
Access Key: The key used to access the above storage account

Leave the default values for *Protocol* and *Connection usage*.

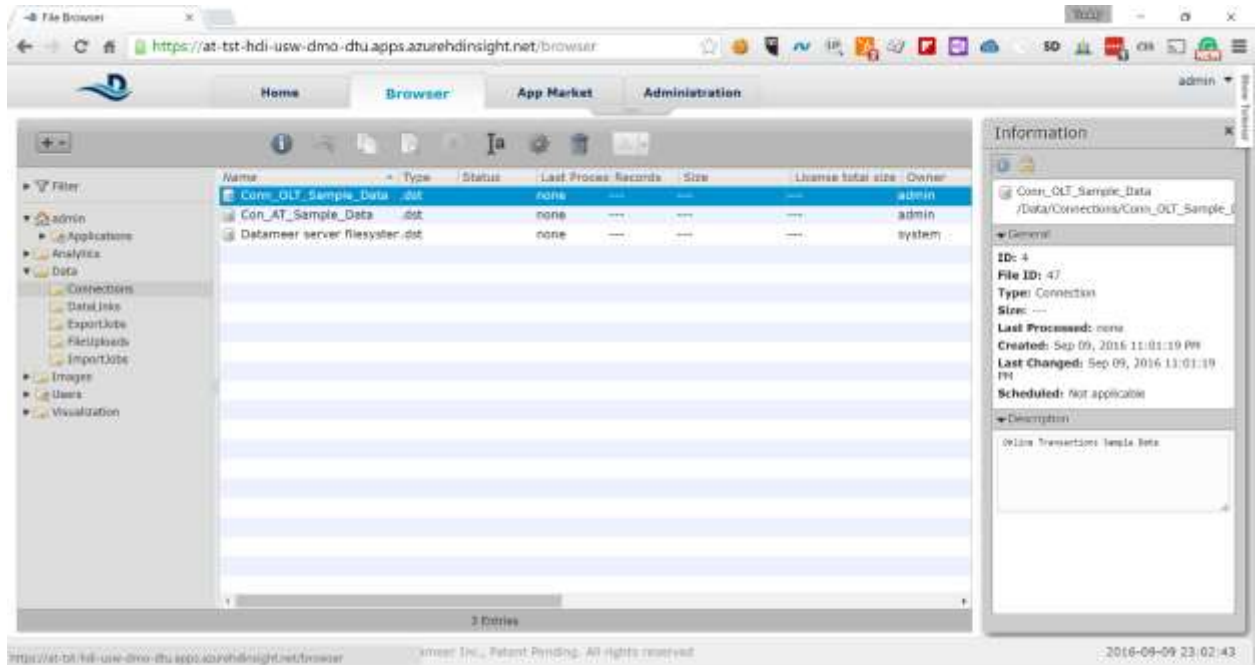
5. Click *Next* and on the next screen type the following description for the connection:
"Online Transactions Sample Data"



- Click **Save** to save the connection and type the following name in the **Save as** field:
Conn_OLT_Sample_Data



- Click **Save** again and you will see the new connection in the list



Now you have Datameer configured to look for data in the specified Azure Storage account and you can start creating your analysis.

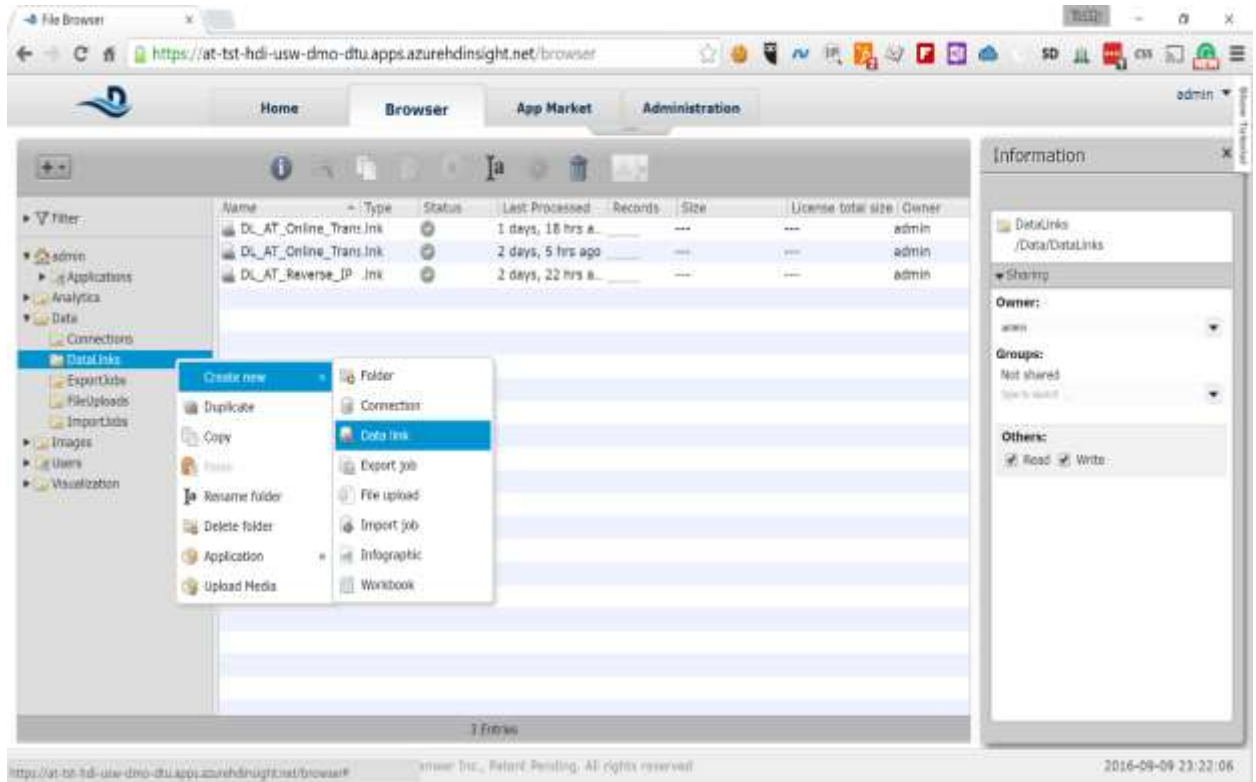
6 Link, Clean and Prepare the Data

Before we start our analysis we need to tell Datameer which data exactly we want to analyze and make sure that it is in the correct format. The sample data we provided has the following two fields that need to be fixed before it is usable for analysis:

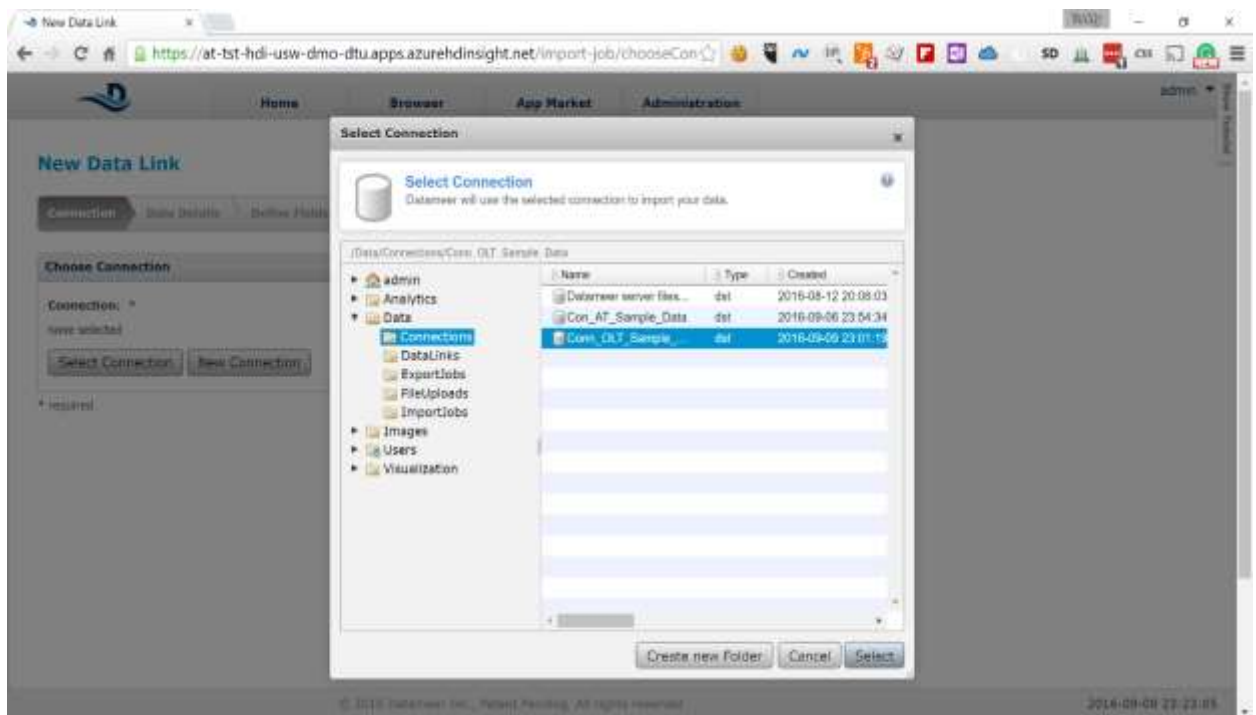
- The *timestamp* field is in ISO-8601 format, which needs to be converted into date/time field that Datameer can understand. We can do this conversion while we are linking the data.
- The *purchase_amount* field is a money field that is interpreted as a *STRING* by Datameer. We need to convert this to *FLOAT* in order to be able to do calculations. We will do that using Datameer formulas once we start our analysis.

Here are the steps to link the data for analysis.

1. Right-click on the *DataLinks* node in the left-side navigation and select *Create new -> Data link*



2. On the next screen click on the *Select Connection* button and select the *Conn_OLT_Sample_Data* connection you created previously



- Click on the *Select* button in the pop-up. Keep the default value *CSV/TSV* in the *File Type* drop down and click *Next*

The screenshot shows the 'New Data Link' web interface. The 'Connection' tab is active, showing a 'Choose Connection' section with a dropdown menu set to 'Conn_OJT_Sample_Data' and buttons for 'Select Connection' and 'New Connection'. Below this is the 'File Type' section with a dropdown menu set to 'CSV / TSV'. At the bottom right are 'Cancel' and 'Next' buttons. The footer shows '© 2018 Datameer Inc., Patent Pending. All rights reserved.' and the date '2018-09-09 23:27:03'.

- On the next screen type the following in the *File Or Folder* field:
/online-transactions-cc_masked.csv
(screenshot below shows ***"/samples/online-transactions-cc_masked.csv"***. Please use ***/online-transactions-cc_masked.csv*** instead)

The screenshot shows the 'New Data Link' web interface with the 'Data Details' tab active. The 'Basic' section is visible, showing a 'Path Prefix' field with a '/' character. Below it is the 'File Or Folder' field, which contains the text '/samples/online-transactions-cc_masked.csv'. To the right of this field is a text box explaining the path format: 'Enter a relative path to a file or folder. Wildcards are accepted. For example: /dev/data/nyfls.txt or /dev/data/nyfls.txt*. Use patterns %year%, %month%, %day%, %hour%, %minute% to be replaced with current date values. Use %pattern% in the file path when entering the data expression for the time based partitions below. [Learn more](#)'. Below the 'File Or Folder' field is a 'Delimiter' field with a dropdown menu set to 'CSV'. At the bottom is a 'Schema' field. The footer shows '© 2018 Datameer Inc., Patent Pending. All rights reserved.' and the date '2018-09-09 23:29:30'.

Scroll down to the bottom, keeping the default values for the rest of the fields, and click on Next

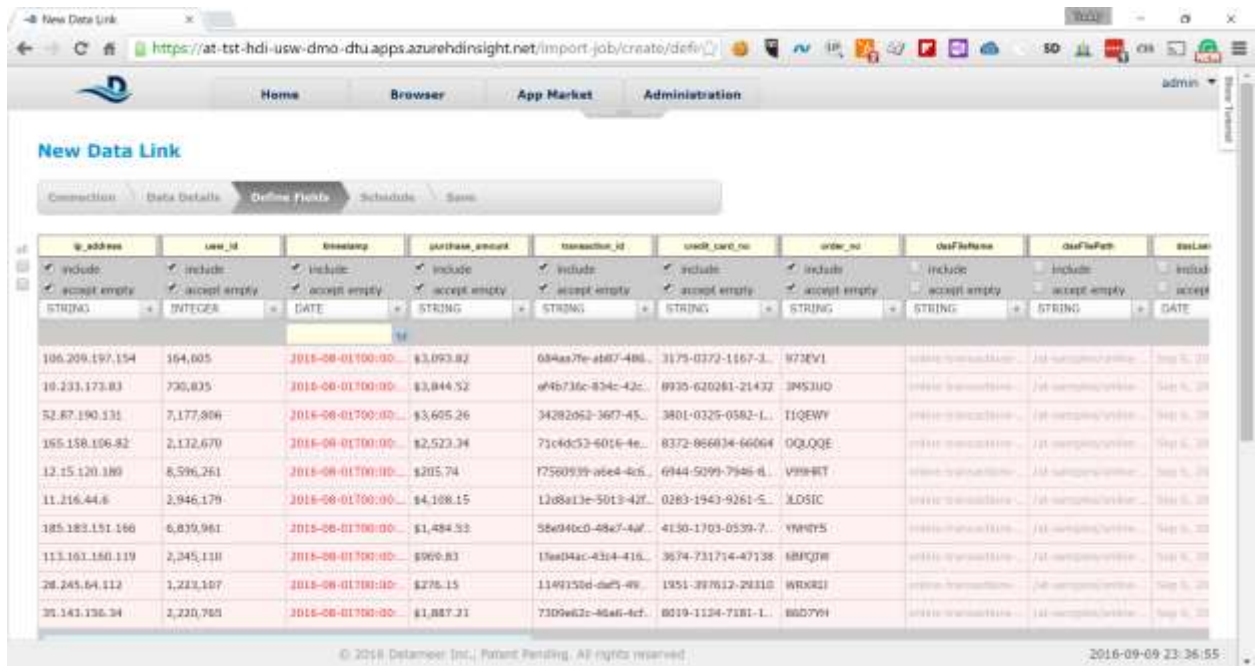
5. Datameer pre-fetches a representative sample of the data and shows it on the next screen

The screenshot shows the 'New Data Link' interface in the Datameer application. The 'Define Fields' tab is active, displaying a table of data. The columns are: ip_address, user_id, timestamp, purchase_amount, transaction_id, credit_card_no, order_no, dsf/fileName, dsf/FilePath, and dsf/Len. The 'timestamp' column is currently set to 'STRING'. Below the table, there is a 'Rescan Schema' button and a copyright notice for 2016 Datameer Inc.

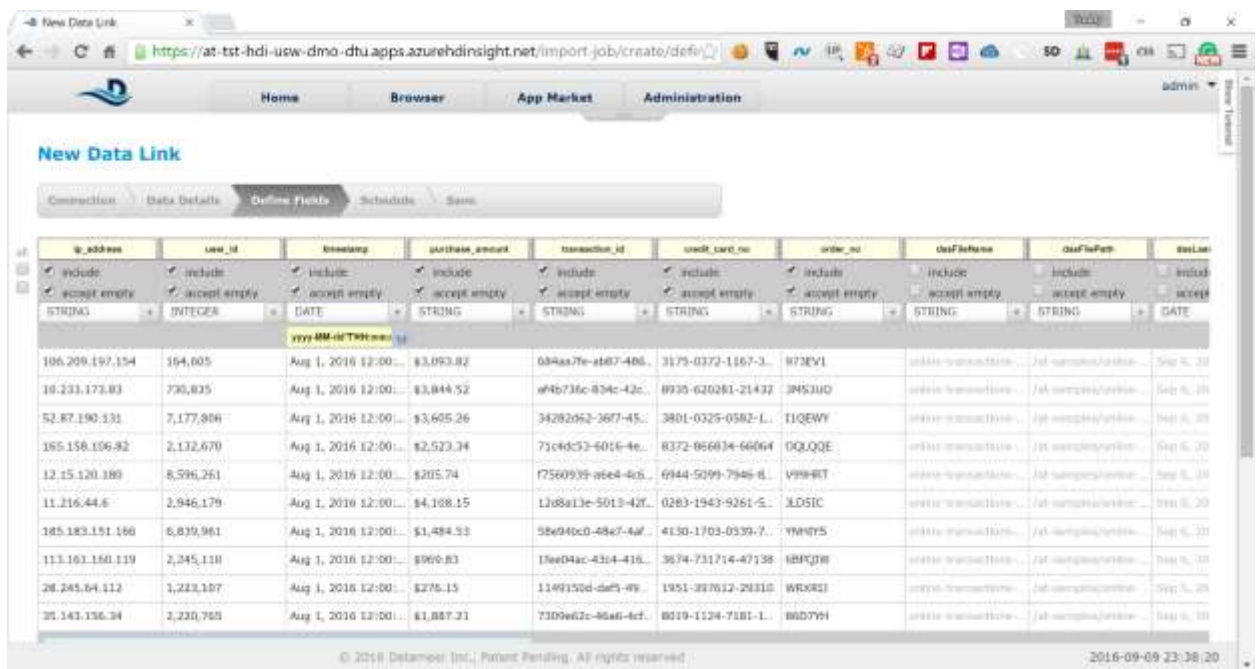
6. Click on the down-arrow for the field type under *timestamp* and change the type from *STRING* to *DATE*

This screenshot shows the same 'New Data Link' interface, but the 'timestamp' field type dropdown menu is open. The options visible are: STRING, INTEGER, FLOAT, DATE, BOOLEAN, BIG_DECIMAL, and BIG_INTEGER. The 'DATE' option is highlighted. The rest of the interface remains the same as in the previous screenshot.

- The dates in the timestamp are automatically marked in red because Datameer cannot parse the ISO-8601 date by default and an input field appears under the field type drop-down



- Type the following pattern in the field `yyyy-MM-dd'T'HH:mm:ss'Z'`



Datameer immediately parses the data in the field and shows it in the correct format. Scroll to the bottom of the screen and click on *Next*

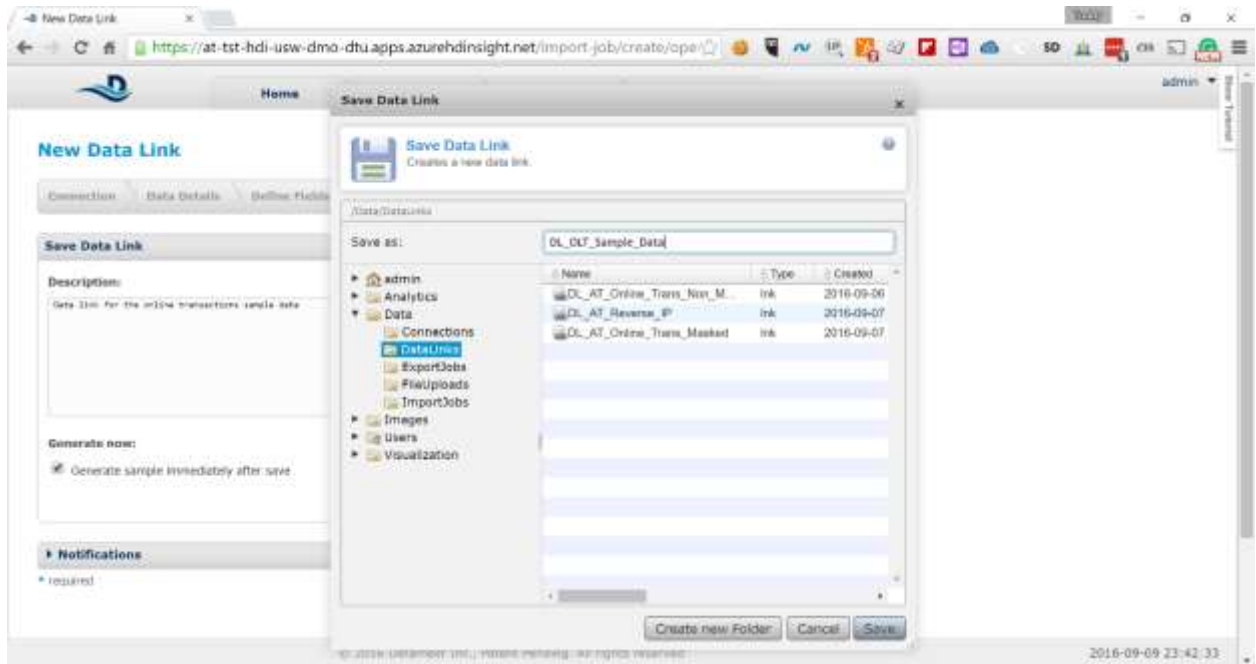
9. On the next screen keep the default value for *Trigger* and click on *Next*

The screenshot shows the 'New Data Link' configuration page in the Azure HDInsight portal, specifically the 'Schedule' tab. The page has a navigation bar with 'Home', 'Browse', 'App Market', and 'Administration'. The 'Schedule' tab is active, and the 'Save' button is highlighted. The 'Refresh Sample Data' section shows the 'Trigger' set to 'Manually' (selected) and 'On a schedule'. A tooltip explains that the trigger determines how and when to refresh the sample data. Below this, the 'Advanced' section is expanded, showing a 'Required' checkbox. At the bottom, there are 'Cancel', 'Back', and 'Next' buttons. The footer indicates '© 2018 Delameter Inc., Patent Pending. All rights reserved.' and the timestamp '2016-09-09 23:40:21'.

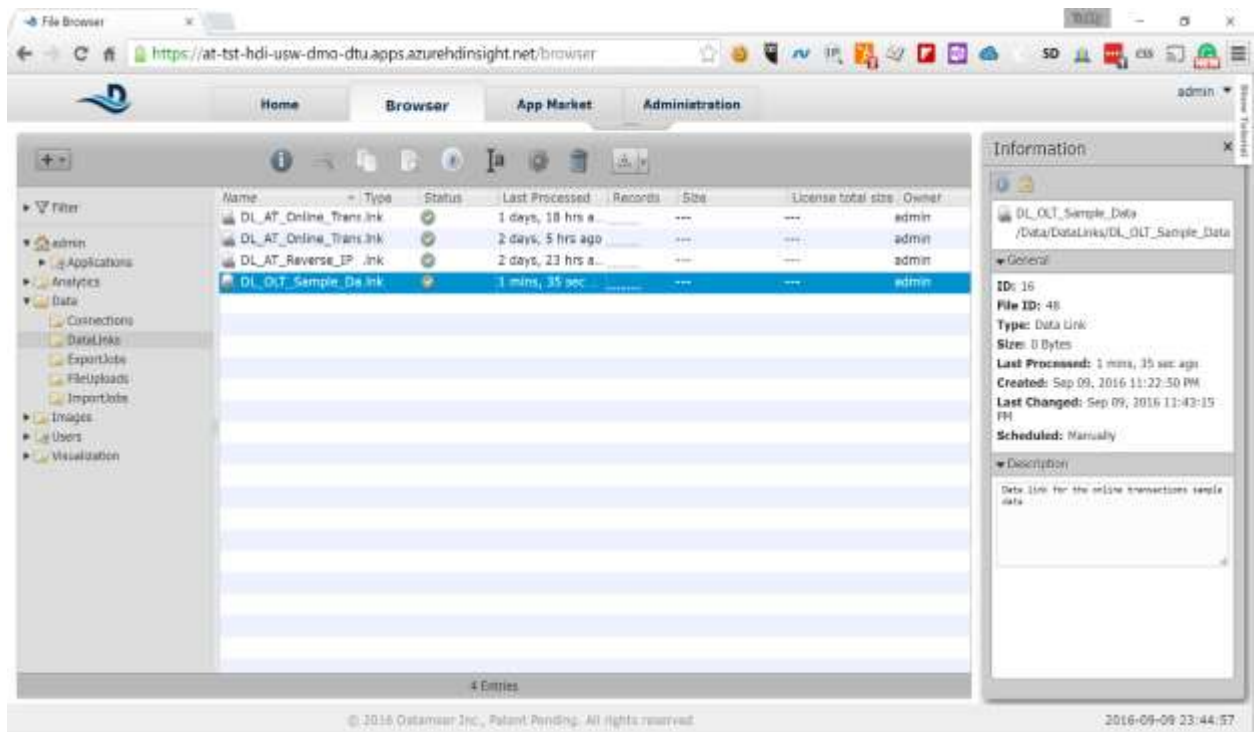
10. On the next screen type a meaningful description for the DataLink and click on *Save*

The screenshot shows the 'New Data Link' configuration page in the Azure HDInsight portal, specifically the 'Save' tab. The page has a navigation bar with 'Home', 'Browse', 'App Market', and 'Administration'. The 'Save' tab is active, and the 'Save' button is highlighted. The 'Save Data Link' section shows a 'Description' field with the text 'Data link for the online transactions sample data'. Below this, the 'Generate now' section is expanded, showing a 'Generate sample immediately after save' checkbox. At the bottom, there are 'Cancel', 'Back', and 'Save' buttons. The footer indicates '© 2018 Delameter Inc., Patent Pending. All rights reserved.' and the timestamp '2016-09-09 23:41:43'.

11. Type the following name in the *Save as* field for the DataLink and click on the *Save* button

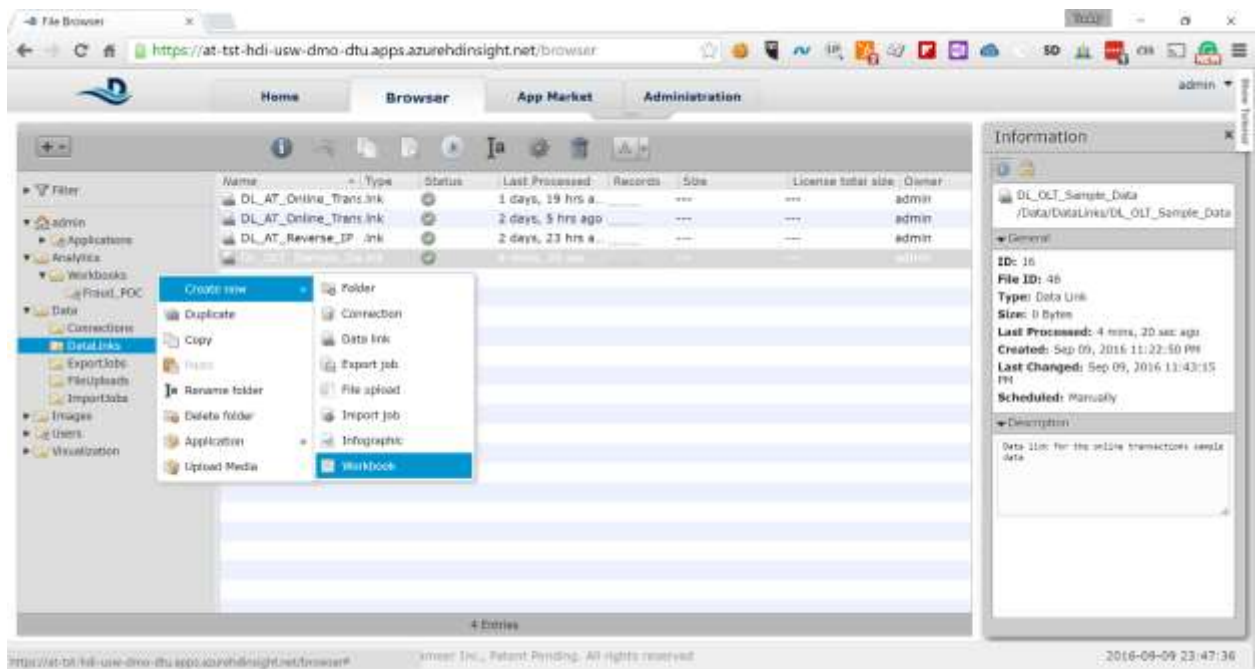


12. The new data link will appear in the list of data links on the next screen

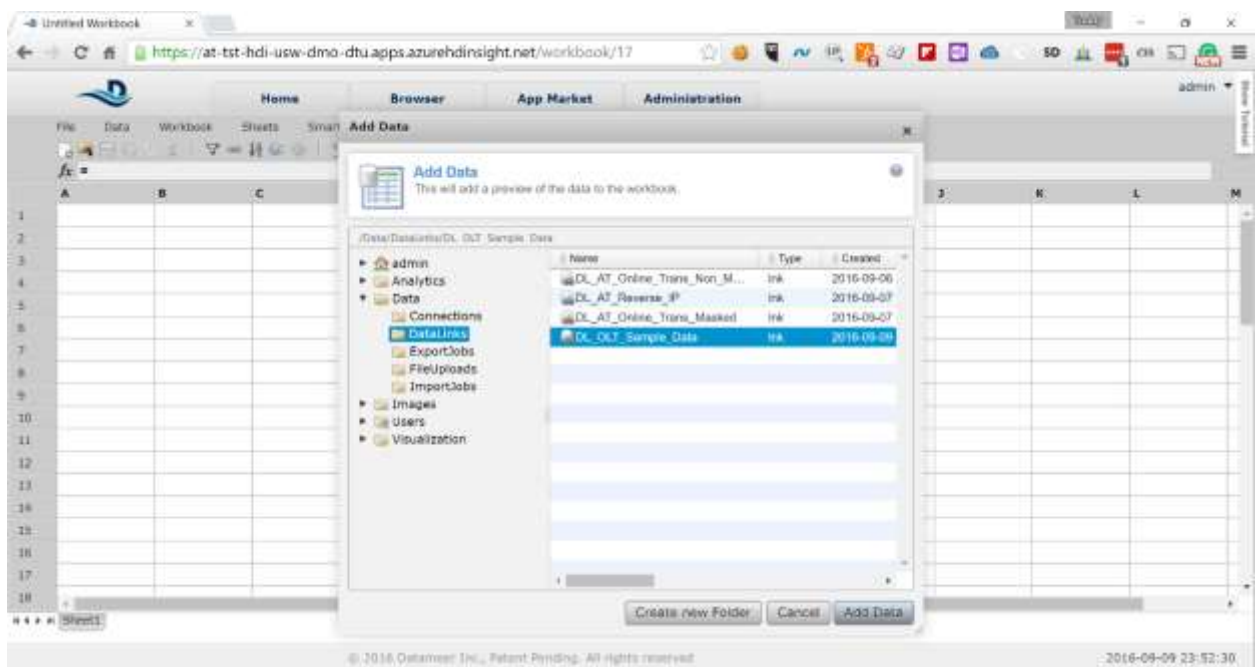


Now you have your dataset linked and have done some preliminary clean-up of the data. Next we will create a workbook where we will finish cleaning up our data and do our analysis. Here are the steps.

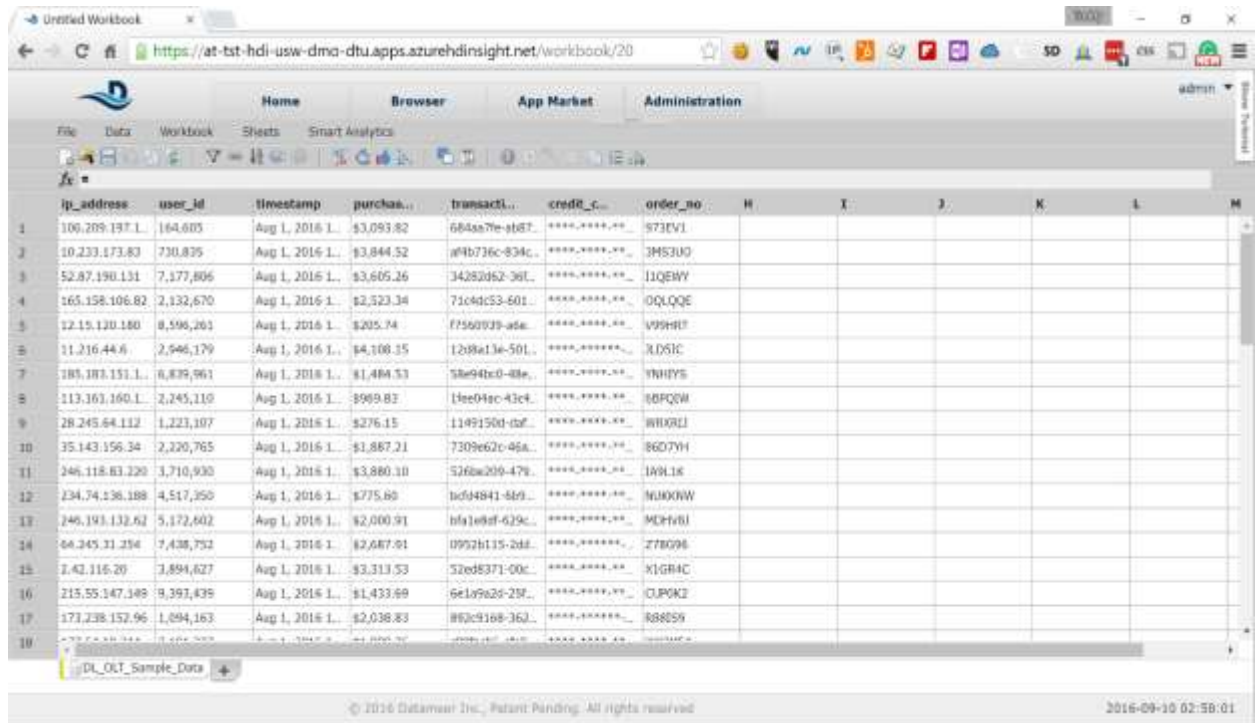
13. Expand the *Analytics* node in the left-side navigation, right-click on the *Workbook* node and select *Create new -> Workbook*



14. A new workbook is created and a pop-up window is shown asking you to select the dataset you want to use for analysis. Expand the *Data* node in the pop-up navigation and click on the *DataLinks* node. In the right-side window select the data link that you just created.



15. Click on *Add Data* to load a sample of the dataset in the workbook sheet

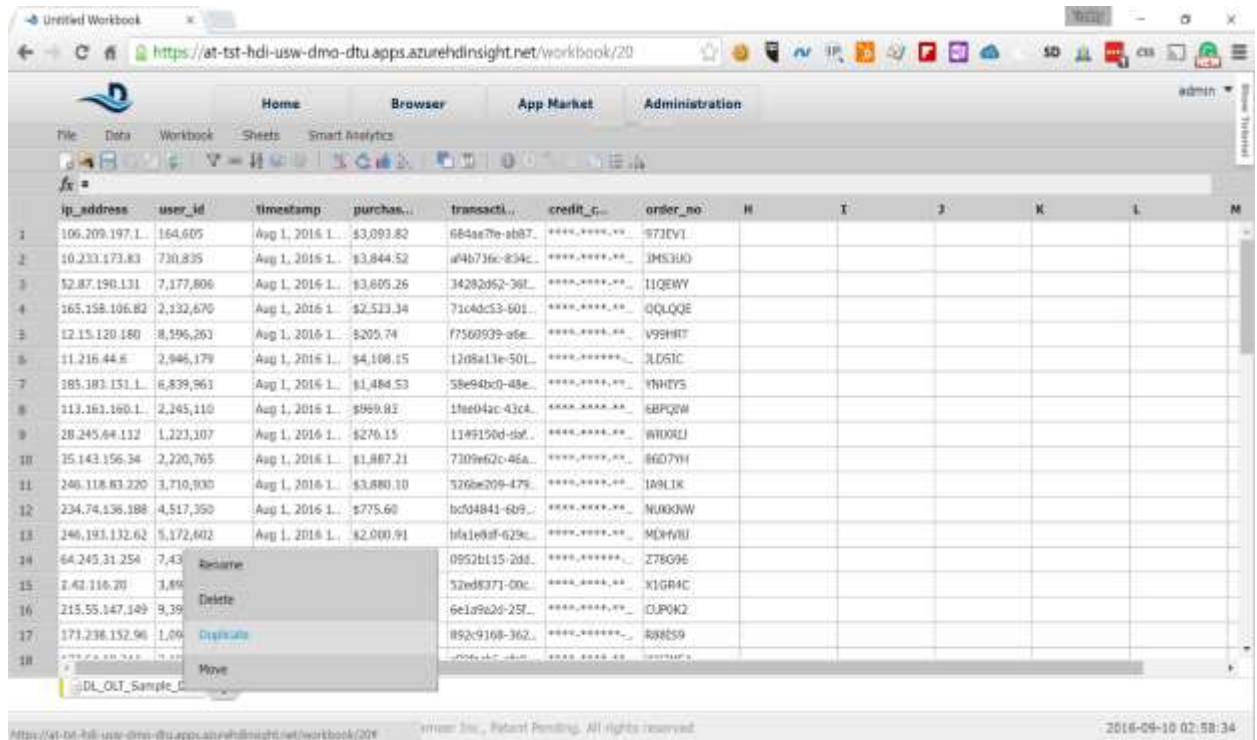


The screenshot shows the Data Manager interface with a workbook titled 'Untitled Workbook'. The interface includes a navigation bar with 'Home', 'Browser', 'App Market', and 'Administration' tabs. Below the navigation bar is a menu with 'File', 'Data', 'Workbook', 'Sheets', and 'Smart Analytics'. The main area displays a table with columns: 'ip_address', 'user_id', 'timestamp', 'purchase...', 'transacti...', 'credit_c...', 'order_no', and several empty columns labeled 'H' through 'M'. The table contains 18 rows of data. At the bottom of the table, there is a tab labeled 'DL_OLT_Sample_Data'.

	ip_address	user_id	timestamp	purchase...	transacti...	credit_c...	order_no	H	I	J	K	L	M
1	100.209.197.1	164,605	Aug 1, 2016 1..	\$3,093.82	684a7fe-9b87..	****,****,xx	973EV1						
2	10.233.173.83	730,835	Aug 1, 2016 1..	\$3,844.52	a94b736c-834c..	****,****,xx	3MS3U0						
3	52.87.190.131	7,177,806	Aug 1, 2016 1..	\$3,605.26	34282052-361..	****,****,xx	11QEWY						
4	165.158.106.82	2,132,670	Aug 1, 2016 1..	\$2,523.34	71c4dc53-601..	****,****,xx	0QLQQE						
5	12.15.120.180	8,596,263	Aug 1, 2016 1..	\$205.74	f7560939-ad6..	****,****,xx	V99HRT						
6	11.216.44.6	2,946,179	Aug 1, 2016 1..	\$4,108.35	1208a13e-501..	****,****,xx	3LDS1C						
7	185.183.151.1	6,839,961	Aug 1, 2016 1..	\$1,484.53	58e94bc0-48e..	****,****,xx	VNHYS						
8	113.161.160.1	2,245,110	Aug 1, 2016 1..	\$969.83	1f6e04ac-43c4..	****,****,xx	6BPQW						
9	28.245.64.132	1,223,307	Aug 1, 2016 1..	\$276.15	1149150d-daf..	****,****,xx	WDXRLJ						
10	35.143.156.34	2,220,765	Aug 1, 2016 1..	\$1,887.21	7309e62c-46a..	****,****,xx	8607YH						
11	246.118.83.220	3,710,930	Aug 1, 2016 1..	\$3,880.10	526be209-479..	****,****,xx	1A9L1K						
12	234.74.136.188	4,517,350	Aug 1, 2016 1..	\$775.60	bc94841-6b9..	****,****,xx	NXODW						
13	246.193.132.62	5,172,602	Aug 1, 2016 1..	\$2,000.91	bf91e8df-629c..	****,****,xx	MDHVB						
14	64.245.31.254	7,438,752	Aug 1, 2016 1..	\$2,687.01	0952b115-2d8..	****,****,xx	Z78G96						
15	2.42.116.20	3,894,627	Aug 1, 2016 1..	\$3,313.53	52ed8371-00c..	****,****,xx	X1GRAC						
16	215.55.147.149	9,393,439	Aug 1, 2016 1..	\$1,433.69	6e1a9a2d-25f..	****,****,xx	CLPKK2						
17	173.238.152.96	1,694,163	Aug 1, 2016 1..	\$2,038.83	852c9168-362..	****,****,xx	888ES9						
18													

The UI you are presented with is very similar to Excel and uses the same concepts.

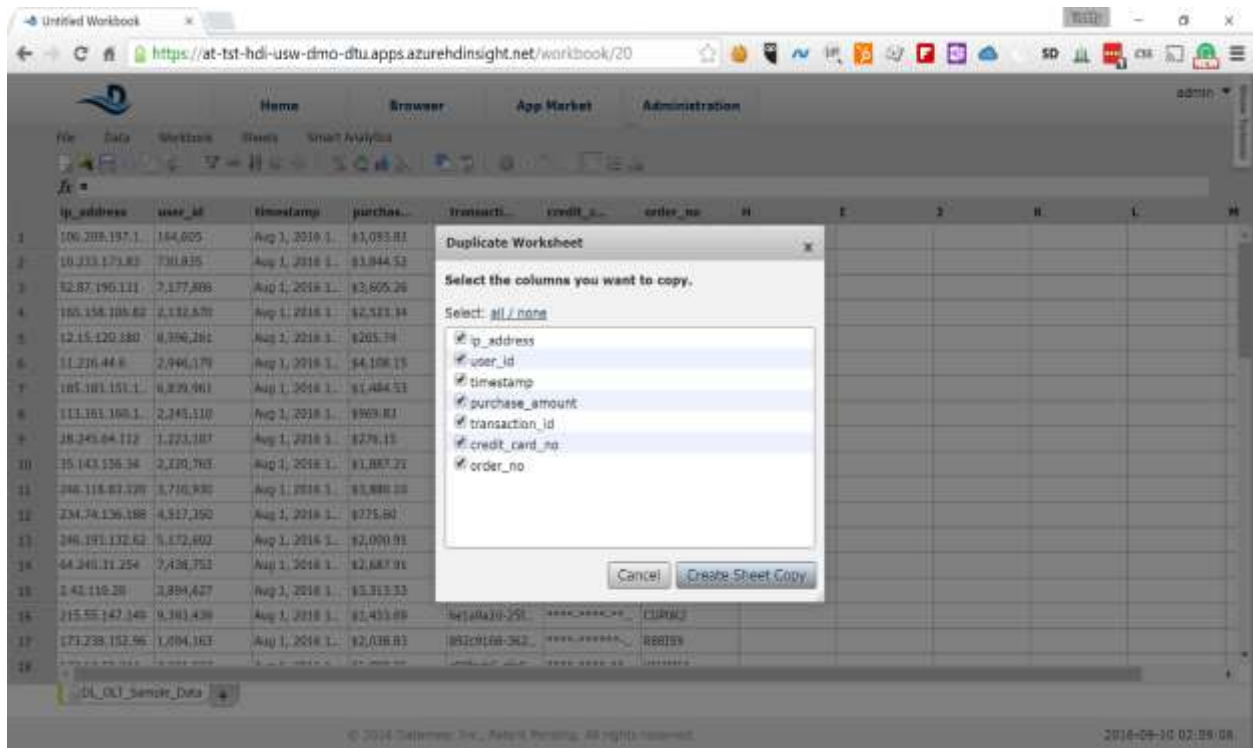
16. Right-click on the *DL_OLT_Sample_Data* sheet at the bottom of the screen next and select *Duplicate*



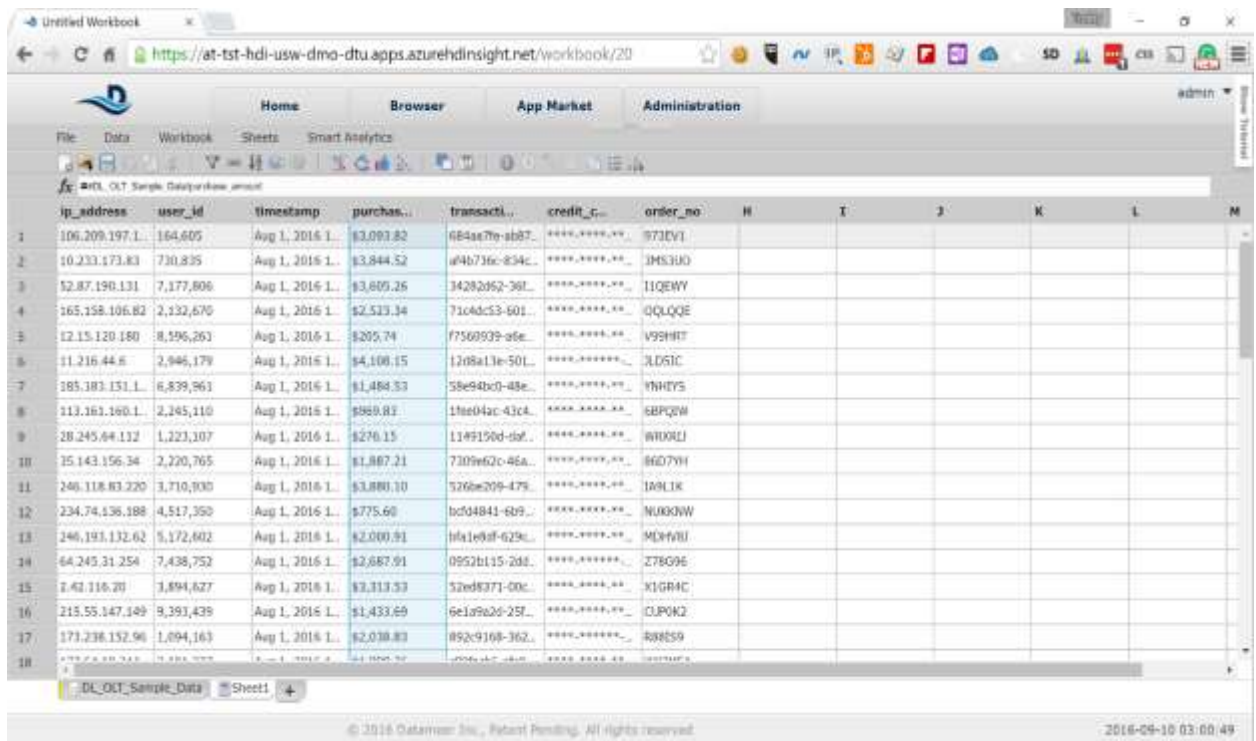
The screenshot shows the same Data Manager interface as before, but with a right-click context menu open over the 'DL_OLT_Sample_Data' tab at the bottom. The menu options are: 'Rename', 'Delete', 'Duplicate', and 'Move'. The 'Duplicate' option is highlighted in blue.

	ip_address	user_id	timestamp	purchase...	transacti...	credit_c...	order_no	H	I	J	K	L	M
1	100.209.197.1	164,605	Aug 1, 2016 1..	\$3,093.82	684a7fe-9b87..	****,****,xx	973EV1						
2	10.233.173.83	730,835	Aug 1, 2016 1..	\$3,844.52	a94b736c-834c..	****,****,xx	3MS3U0						
3	52.87.190.131	7,177,806	Aug 1, 2016 1..	\$3,605.26	34282052-361..	****,****,xx	11QEWY						
4	165.158.106.82	2,132,670	Aug 1, 2016 1..	\$2,523.34	71c4dc53-601..	****,****,xx	0QLQQE						
5	12.15.120.180	8,596,263	Aug 1, 2016 1..	\$205.74	f7560939-ad6..	****,****,xx	V99HRT						
6	11.216.44.6	2,946,179	Aug 1, 2016 1..	\$4,108.35	1208a13e-501..	****,****,xx	3LDS1C						
7	185.183.151.1	6,839,961	Aug 1, 2016 1..	\$1,484.53	58e94bc0-48e..	****,****,xx	VNHYS						
8	113.161.160.1	2,245,110	Aug 1, 2016 1..	\$969.83	1f6e04ac-43c4..	****,****,xx	6BPQW						
9	28.245.64.132	1,223,307	Aug 1, 2016 1..	\$276.15	1149150d-daf..	****,****,xx	WDXRLJ						
10	35.143.156.34	2,220,765	Aug 1, 2016 1..	\$1,887.21	7309e62c-46a..	****,****,xx	8607YH						
11	246.118.83.220	3,710,930	Aug 1, 2016 1..	\$3,880.10	526be209-479..	****,****,xx	1A9L1K						
12	234.74.136.188	4,517,350	Aug 1, 2016 1..	\$775.60	bc94841-6b9..	****,****,xx	NXODW						
13	246.193.132.62	5,172,602	Aug 1, 2016 1..	\$2,000.91	bf91e8df-629c..	****,****,xx	MDHVB						
14	64.245.31.254	7,438,752	Aug 1, 2016 1..	\$2,687.01	0952b115-2d8..	****,****,xx	Z78G96						
15	2.42.116.20	3,894,627	Aug 1, 2016 1..	\$3,313.53	52ed8371-00c..	****,****,xx	X1GRAC						
16	215.55.147.149	9,393,439	Aug 1, 2016 1..	\$1,433.69	6e1a9a2d-25f..	****,****,xx	CLPKK2						
17	173.238.152.96	1,694,163	Aug 1, 2016 1..	\$2,038.83	852c9168-362..	****,****,xx	888ES9						
18													

- Keep all of the fields selected in the pop-up and click on the *Create Sheet Copy* button



- A new copy of the sheet is created that contains all of the data from the original sheet. Click on the *purchase_amount* column to enable the f_x field for that column available on top of the sheet



19. Type the following in the f_x field and press *Enter*

```

FLOAT(SUBSTITUTEALL(SUBSTR(#DL_OLT_Sample_Data!purchase_amount;1);",";""))
    
```

The formula strips the \$ (dollar) sign in front of the amount, removes all commas and converts the string to FLOAT. Now you can use numeric functions to perform calculations on the field.

	ip_address	user_id	timestamp	purchase...	transacti...	credit_c...	order_no...	H	I	J	K	L	M
1	106.209.197.1...	164,605	Aug 1, 2016 1...	1,093.82	684aa76-ad87...	****,****,...	973EV1						
2	10.233.173.83	730,835	Aug 1, 2016 1...	1,044.53	a4b736c-e34c...	****,****,...	IM53U0						
3	52.87.190.131	7,177,606	Aug 1, 2016 1...	1,005.26	34292452-36f...	****,****,...	11QEWY						
4	165.158.106.82	2,132,670	Aug 1, 2016 1...	1,523.34	73e1dc53-601...	****,****,...	QQLQQE						
5	12.15.120.180	8,596,261	Aug 1, 2016 1...	105.74	77560929-a6e...	****,****,...	V99H8T						
6	11.216.44.8	2,946,179	Aug 1, 2016 1...	4,108.15	1209e13e-50L...	****,****,...	3LDSIC						
7	185.183.151.1...	6,839,961	Aug 1, 2016 1...	1,494.53	58e94fc0-48e...	****,****,...	V8H7YS						
8	113.181.160.1...	3,245,110	Aug 1, 2016 1...	969.83	1fee04ec-43c4...	****,****,...	68PQ2W						
9	28.249.44.112	1,223,107	Aug 1, 2016 1...	276.15	1149150d-daf...	****,****,...	WR0REJ						
10	35.143.156.34	2,226,765	Aug 1, 2016 1...	1,887.21	7309e62c-86a...	****,****,...	66D7YH						
11	246.116.83.220	3,710,930	Aug 1, 2016 1...	1,880.1	5260e206-479...	****,****,...	1A9L1K						
12	234.74.136.188	4,517,350	Aug 1, 2016 1...	775.6	bcd94841-6b9...	****,****,...	9L0K9W						
13	246.193.132.62	5,172,602	Aug 1, 2016 1...	2,000.91	bfa1e8df-629c...	****,****,...	MDHVB1						
14	64.245.11.254	7,438,792	Aug 1, 2016 1...	2,087.91	0952b115-2dd...	****,****,...	27BGB6						
15	2.42.116.20	3,894,627	Aug 1, 2016 1...	3,313.53	52ed8371-00c...	****,****,...	KLGR4C						
16	215.55.147.149	9,393,439	Aug 1, 2016 1...	1,433.89	ee1af62d-25f...	****,****,...	OLPKK2						
17	171.238.152.96	1,094,163	Aug 1, 2016 1...	2,038.83	890c9168-362...	****,****,...	8882S9						
18						

20. Right-click on the sheet name at the bottom of the screen to show the context menu for *Sheet1* and select *Rename*

	ip_address	user_id	timestamp	purchase...	transacti...	credit_c...	order_no...	H	I	J	K	L	M
1	106.209.197.1...	164,605	Aug 1, 2016 1...	1,093.82	684aa76-ad87...	****,****,...	973EV1						
2	10.233.173.83	730,835	Aug 1, 2016 1...	1,044.53	a4b736c-e34c...	****,****,...	IM53U0						
3	52.87.190.131	7,177,606	Aug 1, 2016 1...	1,005.26	34292452-36f...	****,****,...	11QEWY						
4	165.158.106.82	2,132,670	Aug 1, 2016 1...	1,523.34	73e1dc53-601...	****,****,...	QQLQQE						
5	12.15.120.180	8,596,261	Aug 1, 2016 1...	105.74	77560929-a6e...	****,****,...	V99H8T						
6	11.216.44.8	2,946,179	Aug 1, 2016 1...	4,108.15	1209e13e-50L...	****,****,...	3LDSIC						
7	185.183.151.1...	6,839,961	Aug 1, 2016 1...	1,494.53	58e94fc0-48e...	****,****,...	V8H7YS						
8	113.181.160.1...	3,245,110	Aug 1, 2016 1...	969.83	1fee04ec-43c4...	****,****,...	68PQ2W						
9	28.249.44.112	1,223,107	Aug 1, 2016 1...	276.15	1149150d-daf...	****,****,...	WR0REJ						
10	35.143.156.34	2,226,765	Aug 1, 2016 1...	1,887.21	7309e62c-86a...	****,****,...	66D7YH						
11	246.116.83.220	3,710,930	Aug 1, 2016 1...	1,880.1	5260e206-479...	****,****,...	1A9L1K						
12	234.74.136.188	4,517,350	Aug 1, 2016 1...	775.6	bcd94841-6b9...	****,****,...	9L0K9W						
13	246.193.132.62	5,172,602	Aug 1, 2016 1...	2,000.91	bfa1e8df-629c...	****,****,...	MDHVB1						
14	64.245.11.254	7,438,792	Aug 1, 2016 1...	2,087.91	0952b115-2dd...	****,****,...	27BGB6						
15	2.42.116.20	3,894,627	Aug 1, 2016 1...	3,313.53	52ed8371-00c...	****,****,...	KLGR4C						
16	215.55.147.149	9,393,439	Aug 1, 2016 1...	1,433.89	ee1af62d-25f...	****,****,...	OLPKK2						
17	171.238.152.96	1,094,163	Aug 1, 2016 1...	2,038.83	890c9168-362...	****,****,...	8882S9						
18						

21. Rename *Sheet1* to *Transaction_Data*

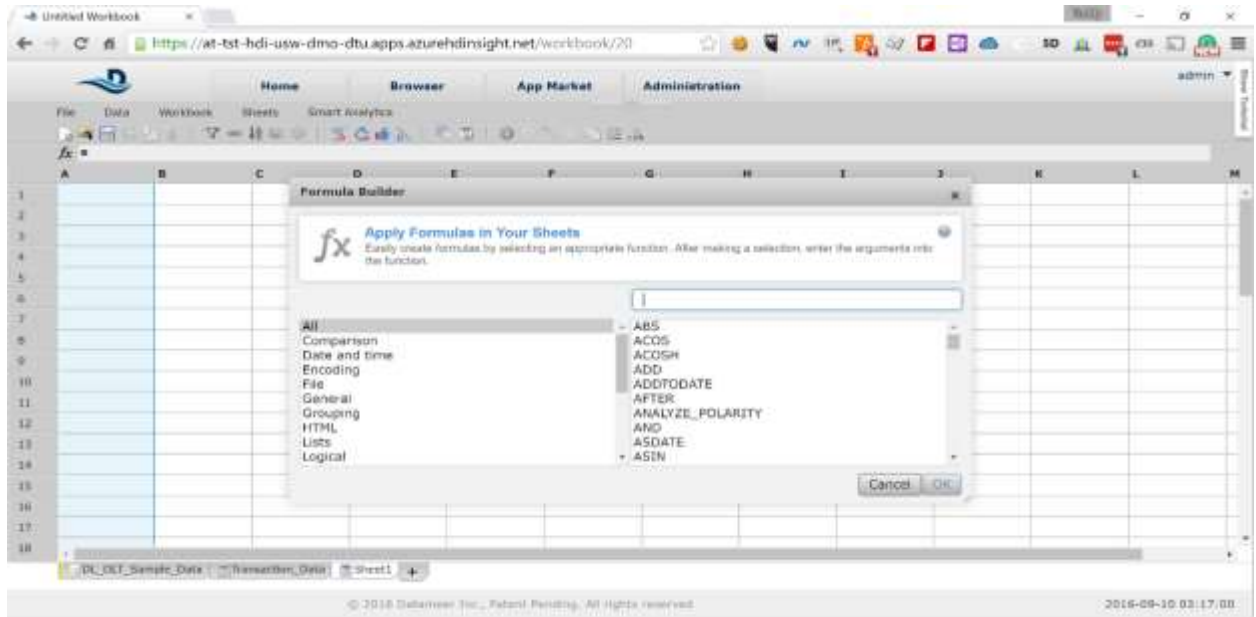
	ip_address	user_id	timestamp	purchase...	transacti...	credit_c...	order_no	H	I	J	K	L	M
1	106.109.197.1...	154,605	Aug 1, 2016 1...	1,092.82	684aa70e-ad87...	****,****,****	873EV1						
2	10.233.173.83	730,835	Aug 1, 2016 1...	1,844.53	a64b73bc-834c...	****,****,****	9M5TUC						
3	52.87.190.131	7,177,806	Aug 1, 2016 1...	1,605.26	34262d52-36f...	****,****,****	11QETW						
4	165.158.106.82	2,132,670	Aug 1, 2016 1...	1,523.34	71c4dc53-601...	****,****,****	0QJQQE						
5	12.15.120.180	8,596,261	Aug 1, 2016 1...	105.74	77560939-a6e...	****,****,****	V99HRT						
6	11.216.44.8	2,946,179	Aug 1, 2016 1...	4,108.15	1208e13e-301...	****,****,****	3LDGIC						
7	185.183.151.1...	6,839,961	Aug 1, 2016 1...	1,484.53	36e94bc0-48a...	****,****,****	VNHYYS						
8	113.161.160.1...	3,245,110	Aug 1, 2016 1...	869.83	1aee04ae-43c4...	****,****,****	8BPCQW						
9	28.245.84.112	1,223,187	Aug 1, 2016 1...	176.15	1149150d-09e...	****,****,****	WRXRJ3						
10	35.143.156.34	2,220,765	Aug 1, 2016 1...	1,887.21	7309e62c-46e...	****,****,****	8607YH						
11	246.118.83.220	3,710,930	Aug 1, 2016 1...	1,880.1	5268e209-479...	****,****,****	1A9L1K						
12	234.74.136.188	4,517,350	Aug 1, 2016 1...	775.8	bd94841-6b9...	****,****,****	9UDCKW						
13	246.103.132.62	5,172,602	Aug 1, 2016 1...	2,000.01	bfa1e8df-629c...	****,****,****	MDP4BI						
14	64.245.71.254	7,438,752	Aug 1, 2016 1...	1,687.91	0952b135-2dd...	****,****,****	27BG86						
15	2.42.116.20	3,894,627	Aug 1, 2016 1...	3,313.53	52e18371-00c...	****,****,****	K1GR4C						
16	215.55.147.149	9,393,439	Aug 1, 2016 1...	1,432.65	6e1a9a3d-25f...	****,****,****	CLPQK2						
17	173.238.152.96	1,094,163	Aug 1, 2016 1...	2,038.81	892c9168-362...	****,****,****	888259						
18													

With this we are done with the clean-up of our data and are ready to perform our analysis.

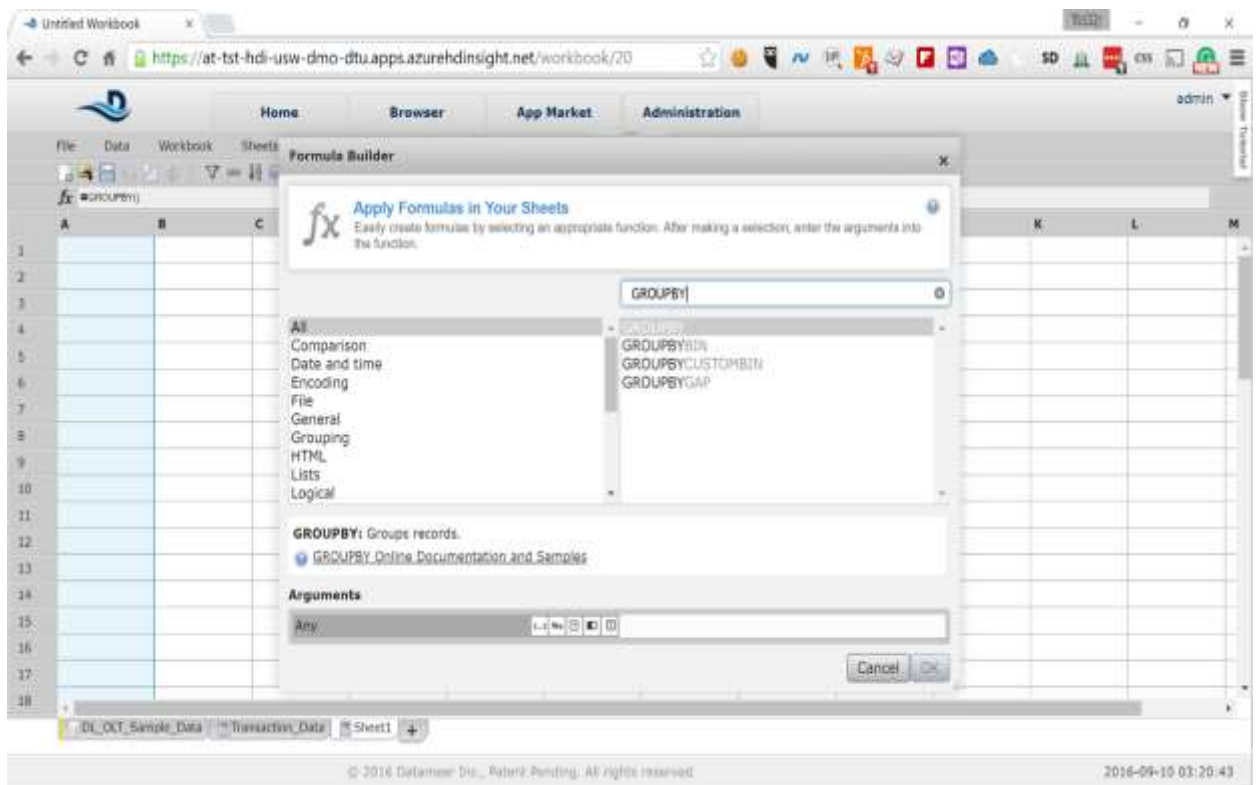
7 Perform Analysis to Identify Outliers

The goal of our analysis is to identify unusual purchasing patterns that deviate from a well-established norm. If we notice something unusual this may be sign that fraud may be committed. For the purpose of this HOL we will be looking for period during the month, in which the transactions significantly deviate from the normal patters during the rest of the month. Here the steps:

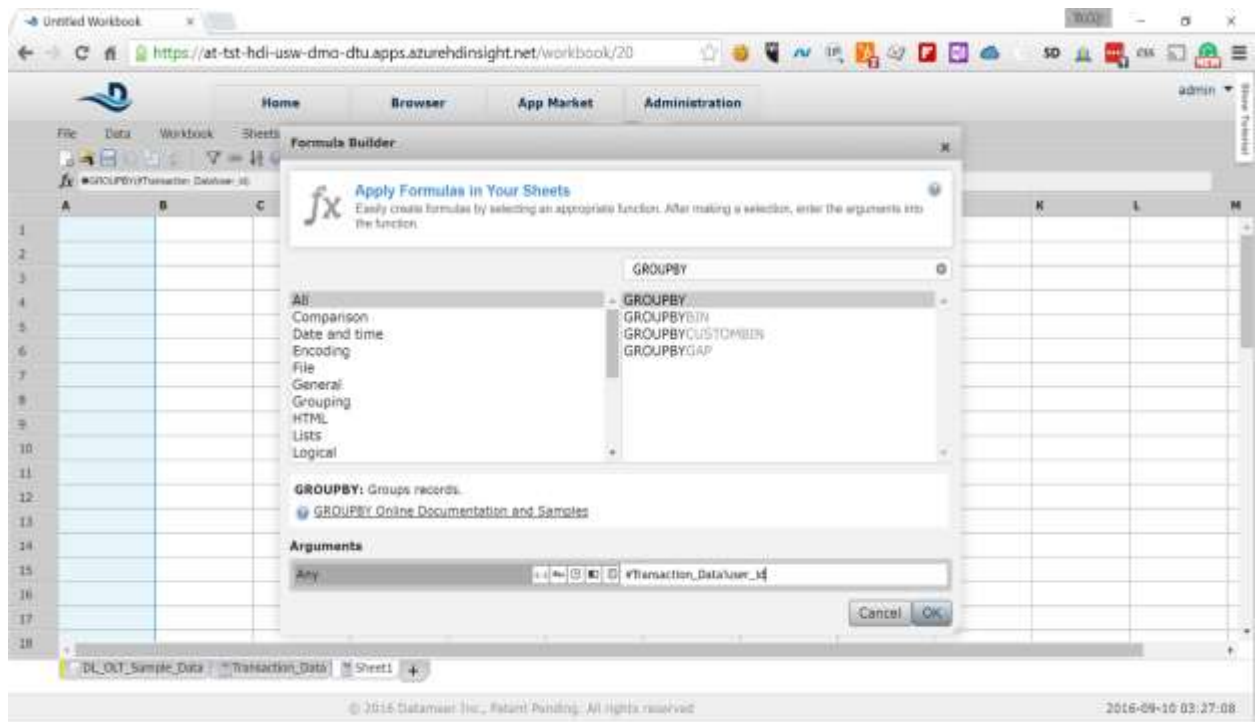
1. Click on the + sign next to the *Transaction_Data* sheet to create an empty sheet for analysis



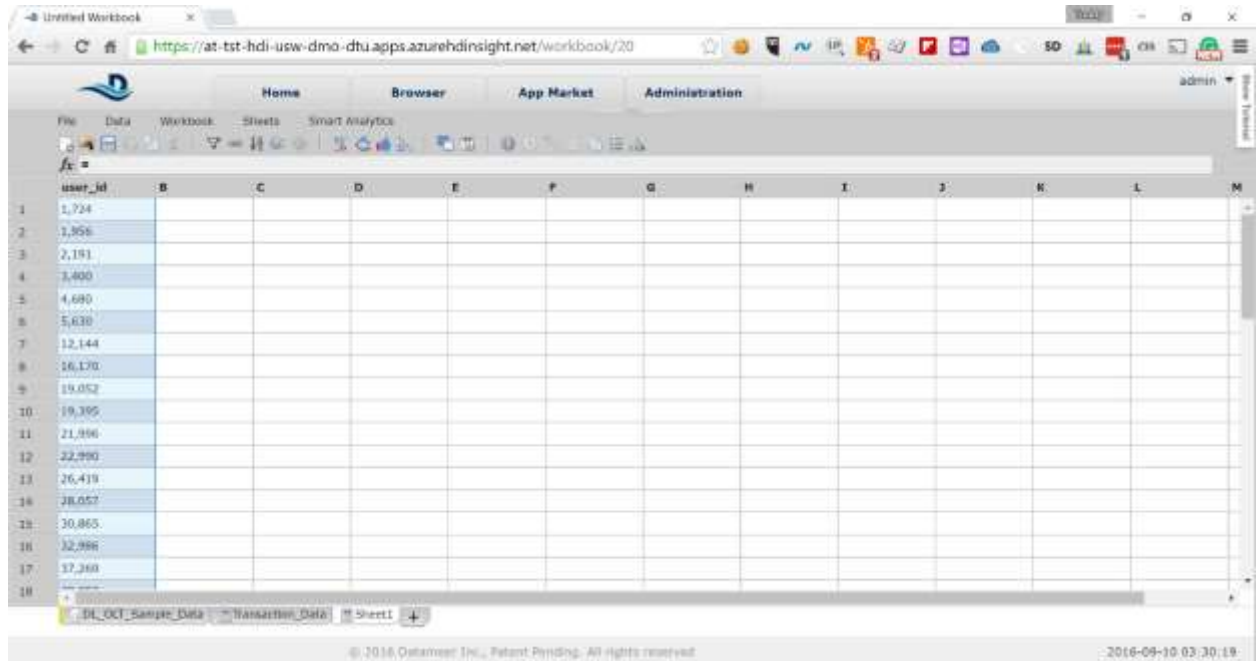
2. In the input field in the pop-up type *GROUPBY* to filter the functions and select *GROUPBY* from the list



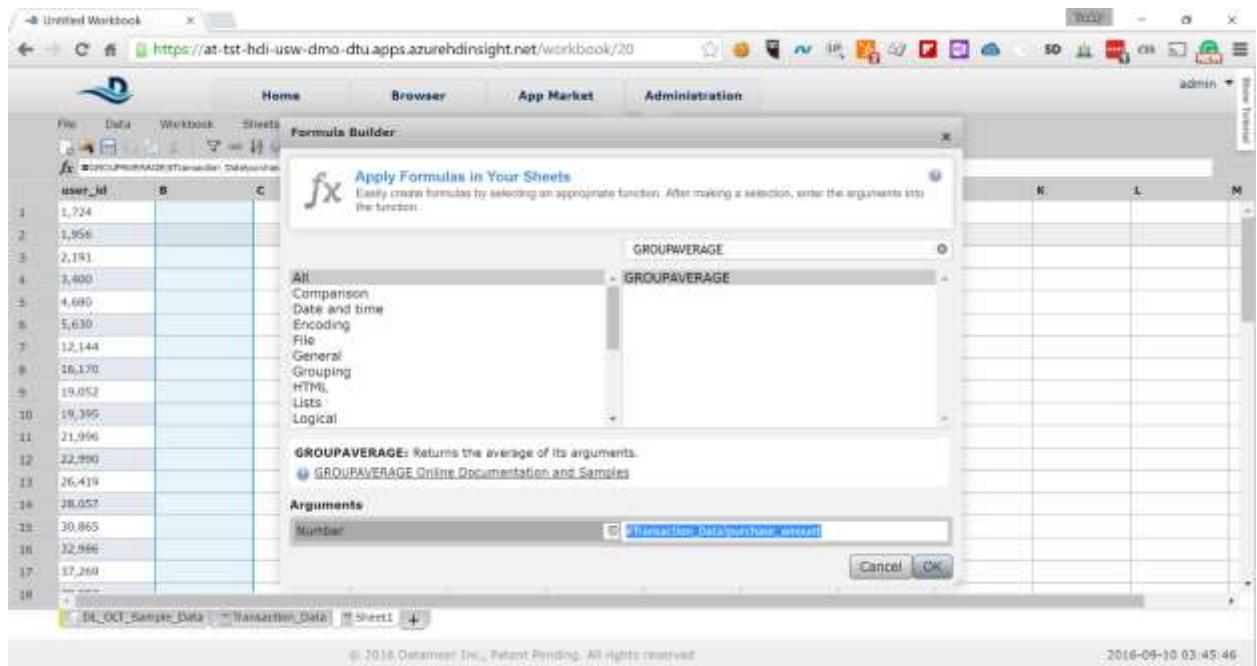
3. In the *Arguments* input field at the bottom of the pop-up type *#Transaction_Data!user_id* to group the data by user identifier and click on the OK button



4. The first column of the sheet will be populated with list of unique user identifiers

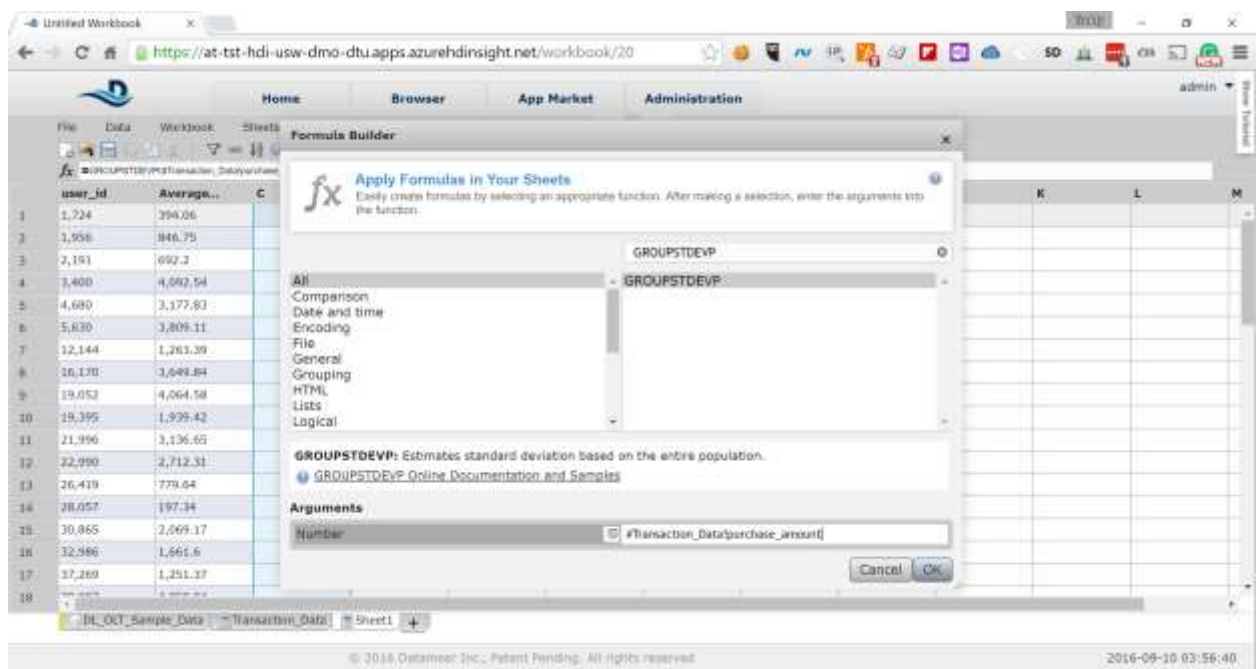


5. Click on the second column to show the functions pop-up again and type GROUPAVERAGE in the filter box and select the GROUPAVERAGE function. In the Arguments field type `#Transaction_Data!purchase_amount`



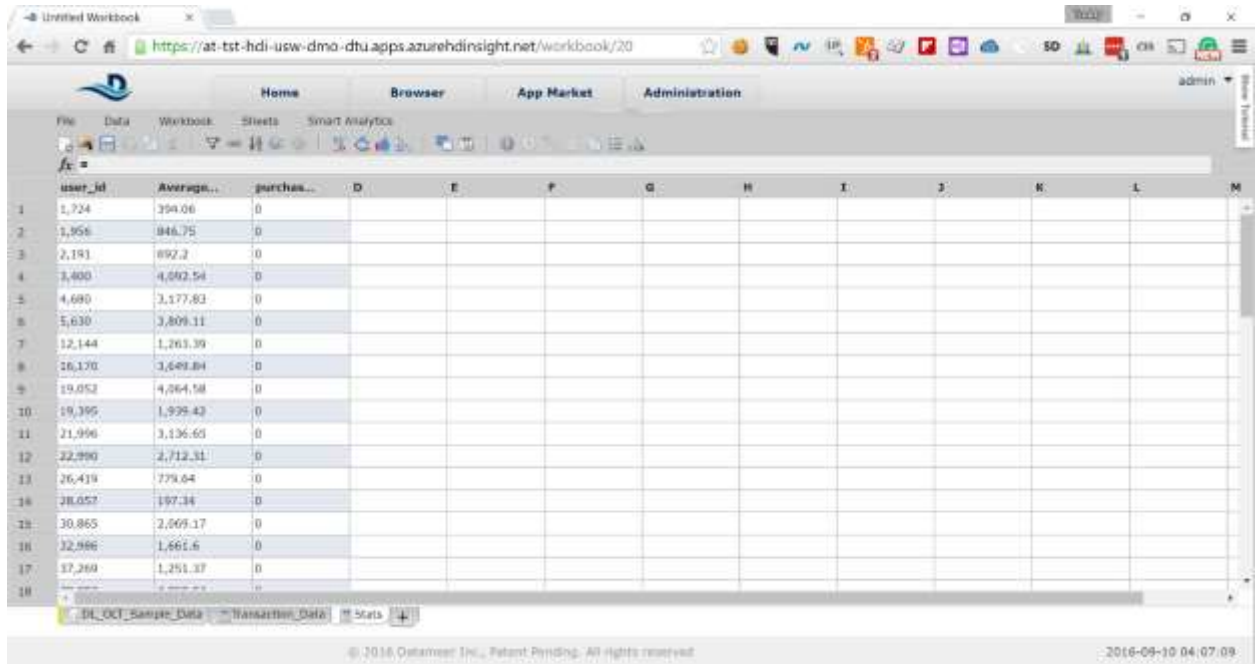
This will calculate the average purchase amount for each of the users.

- Click on the third column to show the function pop-up again and type *GROUPSTDEVP* and select the *GROUPSTDEVP* function

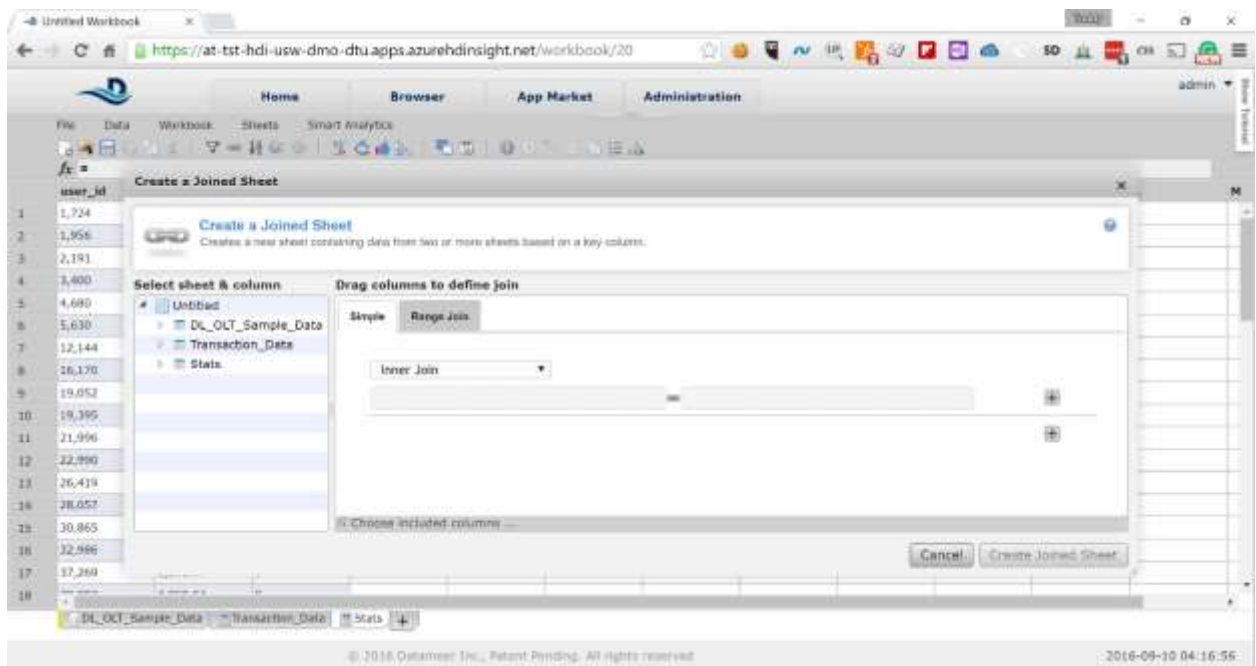


In the Arguments field type `#Transaction_Data!purchase_amount` to calculate the standard deviation for the *purchase_amount* field

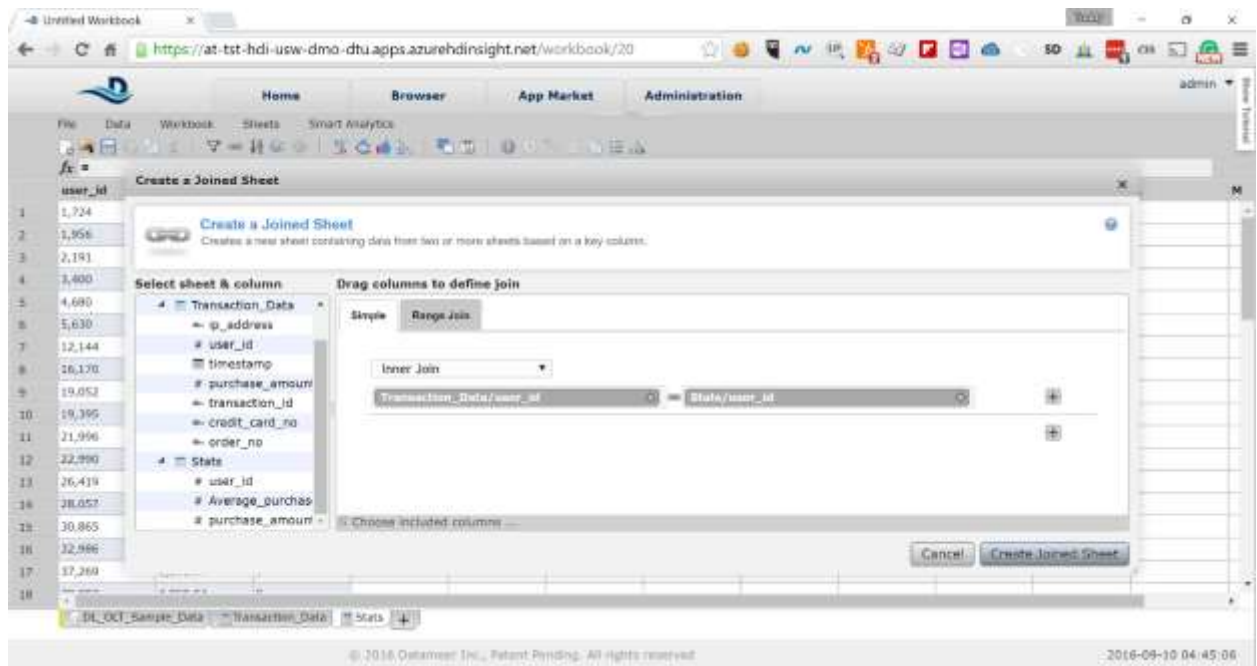
7. Right-click on the sheet name and select *Rename* from the context menu to rename the sheet. Choose the following name for the sheet:
Stats



8. Next we need to join the transaction data for each user with the statistical data for each user to determine how much particular transaction differentiates from the common norm. From the menu bar select *Data* -> *Join* to create a joined sheet

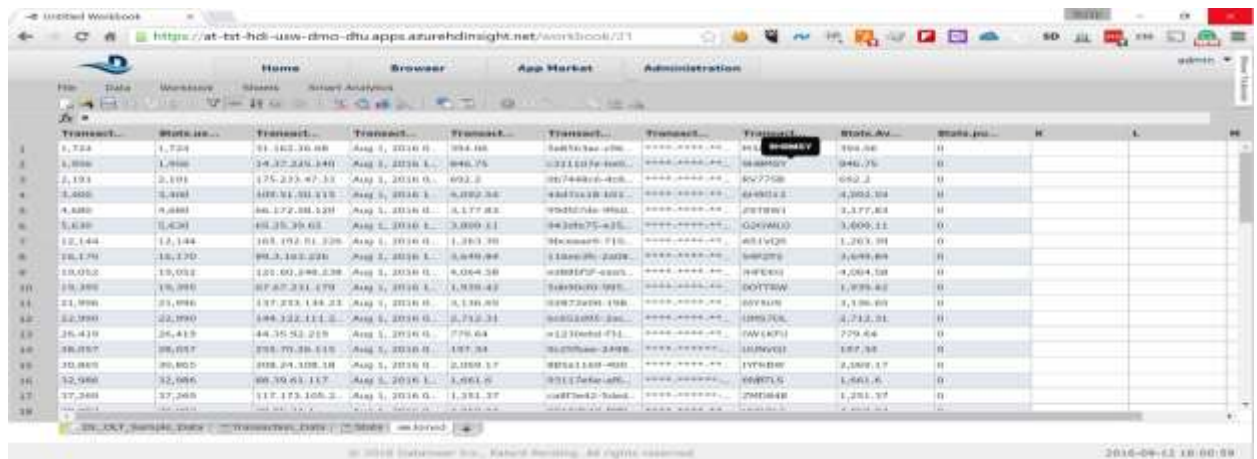


9. Expand the *Transaction_Data* node in the pop-up navigation tree and drag the *user_id* field to the right. Do the same with the *user_id* field from the *Stats* node.



Click on the *Create Joined Sheet* button to create the joined sheet.

- The resulting sheet will show the joined data from both *Transaction_Data* and *Stats* sheets. For convenience let's rename few of the columns. Right-click and rename the columns as below:



For convenience let's rename few of the columns. Right-click and rename the columns as below:

Transaction_Data.user_id -> *user_id*
Transaction_Data.purchase_amount -> *purchase_amount*
Stats.Average_purchase_amount -> *average_purchase_amount*
Stats.purchase_amount Stddevp -> *purchase_amount_deviation*

Also, right-click on the sheet name and rename it to *Joined_Data_and_Stats*

- Next, we will identify the outliers by creating a copy of the joined data and filtering it. Right-click on the *Joined_Data_and_Stats* sheet and select *Duplicate*. We will select only the data we need and ignore the

rest. In the pop-up select only the following fields:

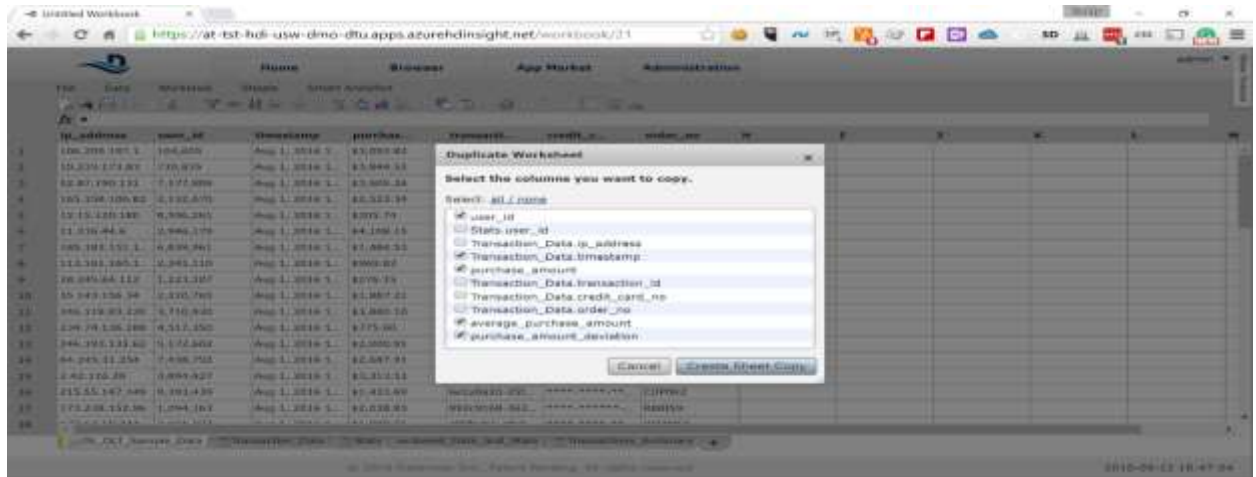
user_id

purchase_amount

Transaction_data.timestamp

average_purchase_amount

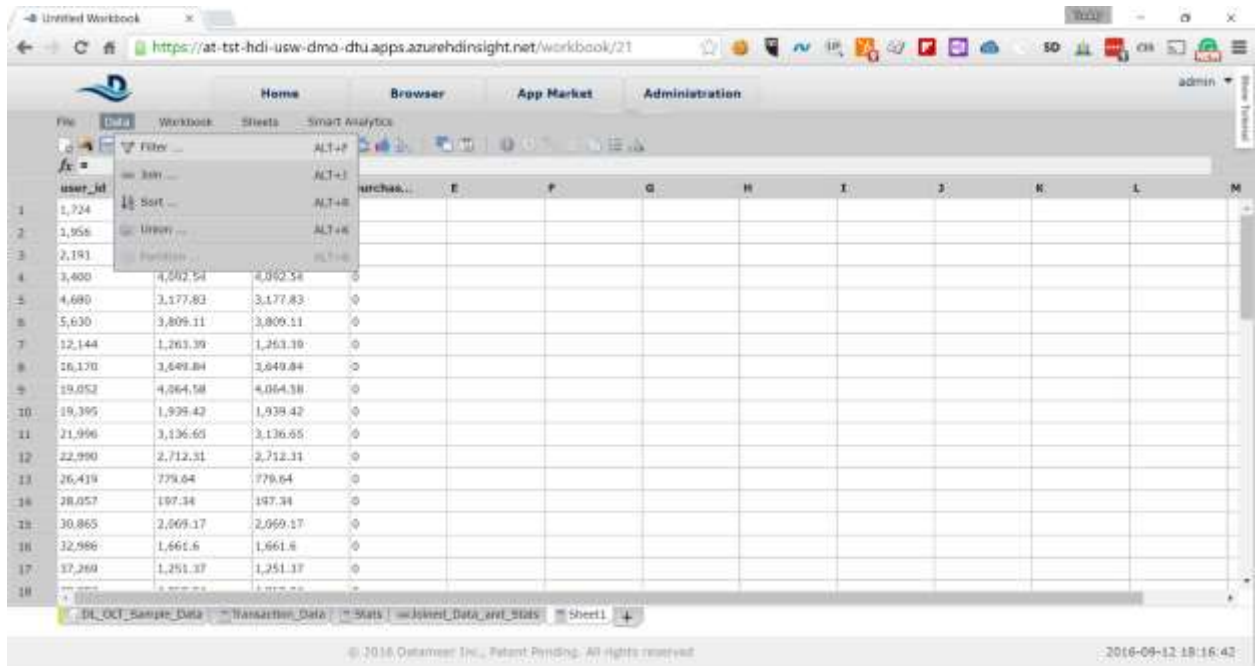
purchase_amount_deviation



Click on *Create Sheet Copy* button

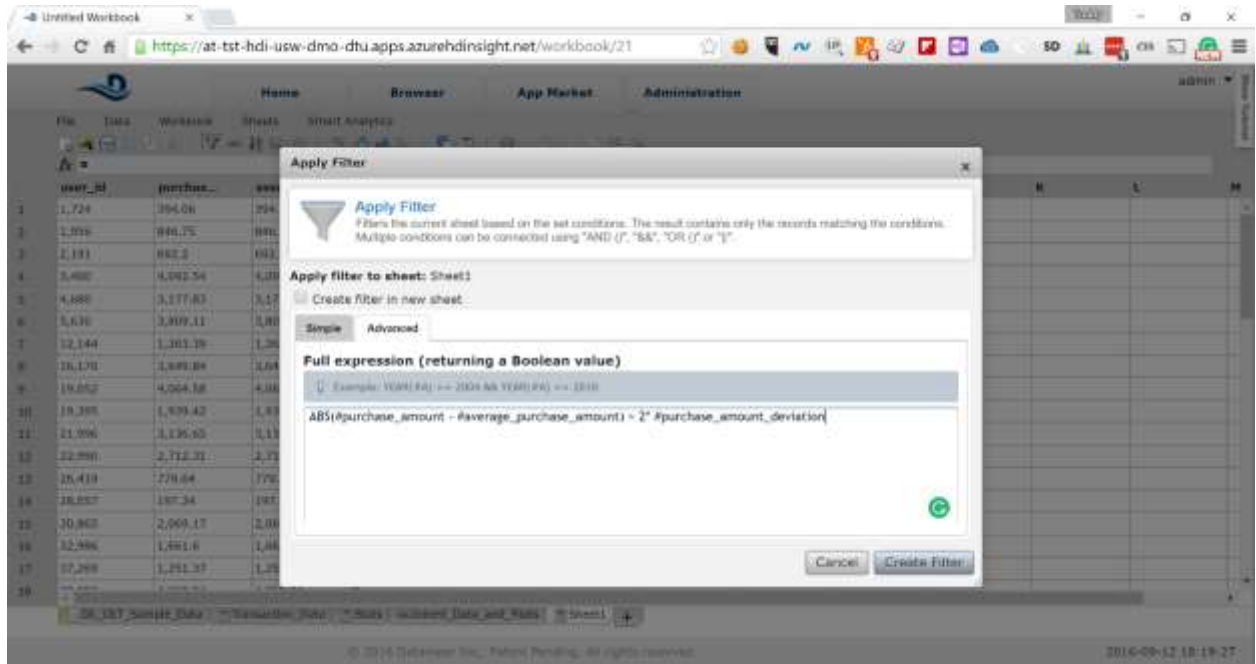
12. Right-click on the *Transaction_Data.timestamp* field and rename it to *timestamp* only.

13. For the purpose of our analysis we will consider transactions with deviation two times more than standard deviation as outliers. In the new sheet select *Data* -> *Filter* from the menu.

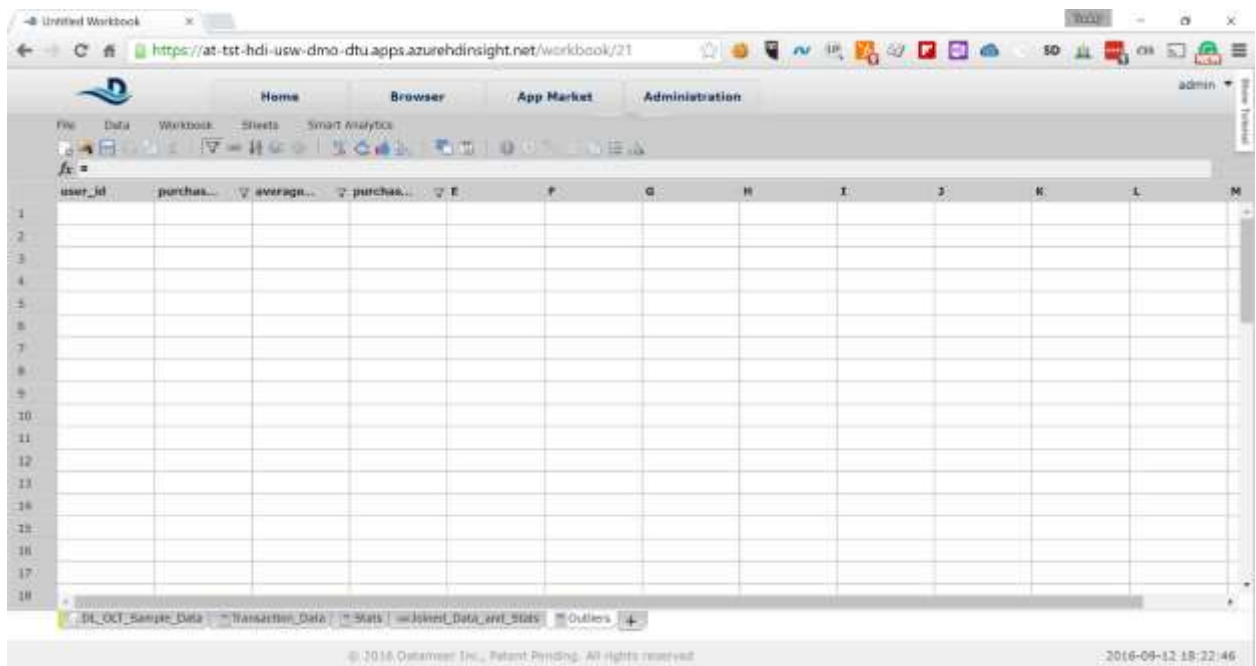


14. Select the *Advanced* tab in the pop-up and type the following formula:

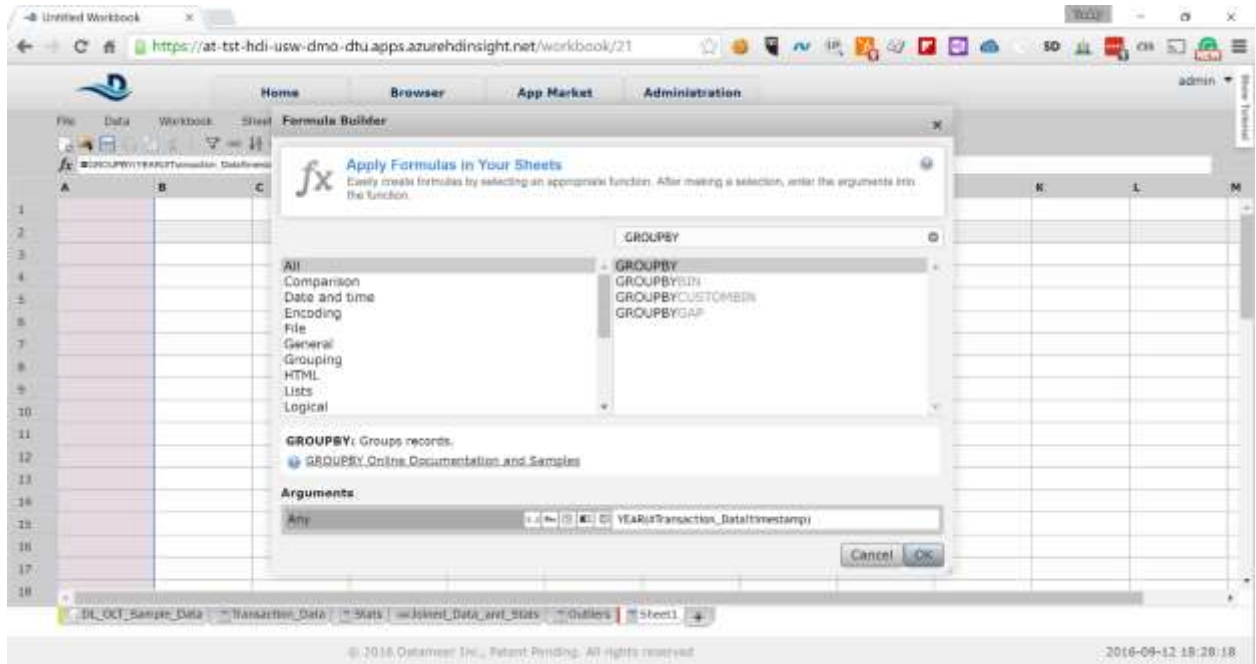
$ABS(\#purchase_amount - \#average_purchase_amount) > 2 * \#purchase_amount_deviation$



15. The resulting sheet may be empty because the representative sample that Datameer has selected may not have transactions that are considered outliers. Right-click on the sheet name and rename it to *Outliers*



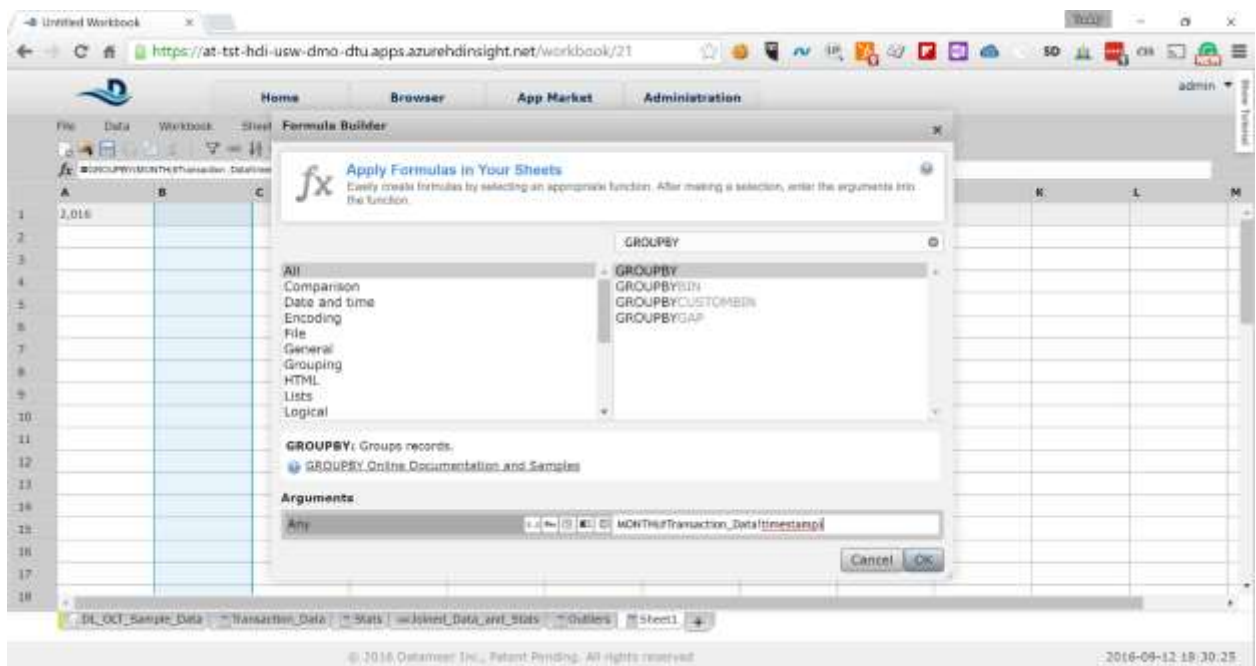
16. Finally, we would like to create a summary of the data that we would like to visualize. Let's start with summary of the *Transaction_Data*. Create new sheet and in the formula pop-up select the *GROUPBY* function



In the *Arguments* field type the following formula:
YEAR(#Transaction_Data.timestamp)

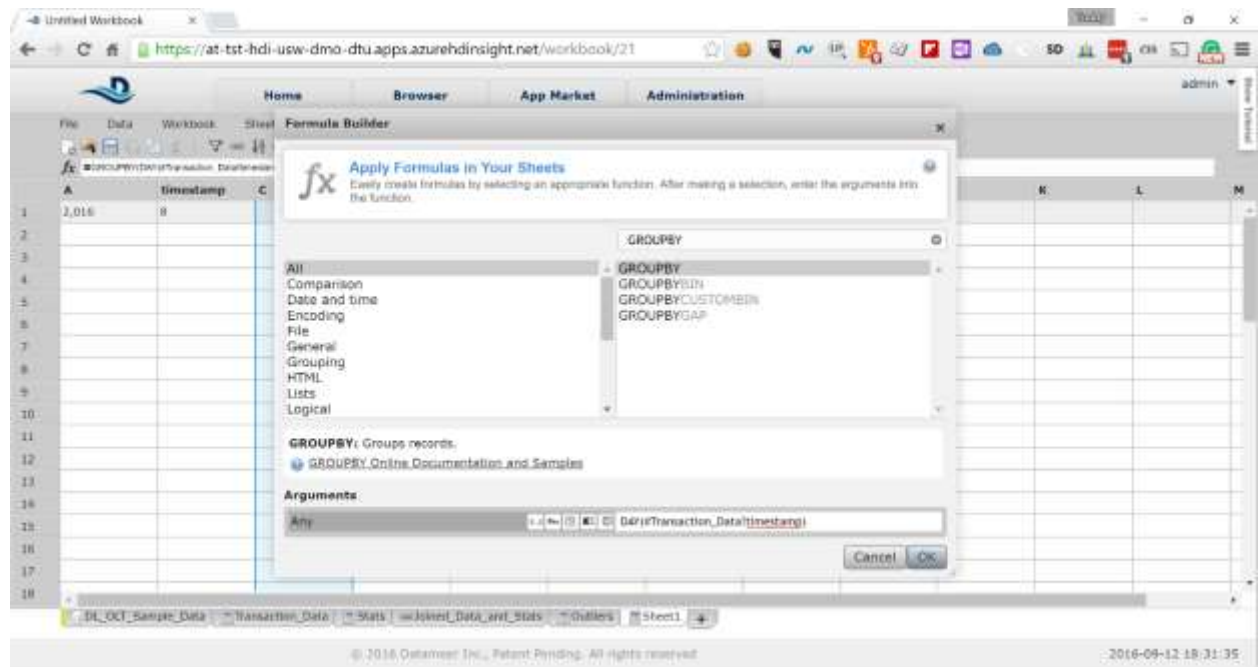
17. Click on the next column and in the formula pop-up select again the *GROUPBY* function and paste the following formula in the *Arguments* field:

MONTH(#Transaction_Data.timestamp)

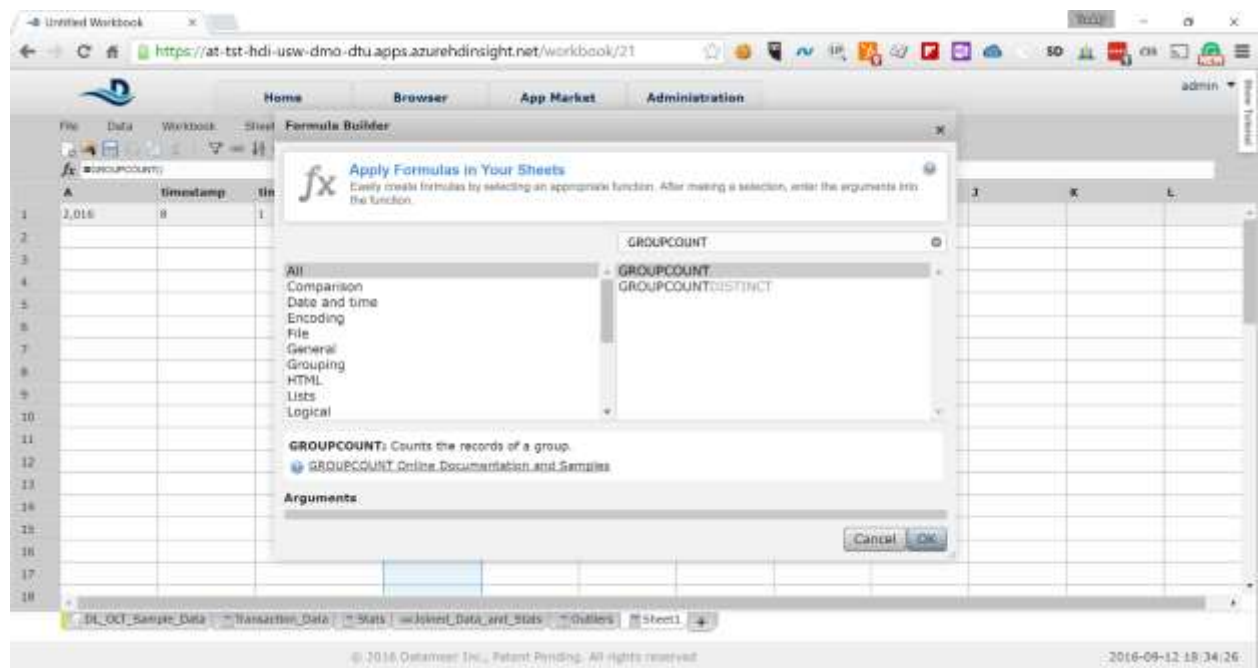


- Click on the third column and in the formula pop-up select again the *GROUPBY* function and paste the following formula in the *Arguments* field:

DAY(#Transaction_Data!timestamp)



- Click on the fourth column and in the formula pop-up select the *GROUPCOUNT* function and click the *OK* button



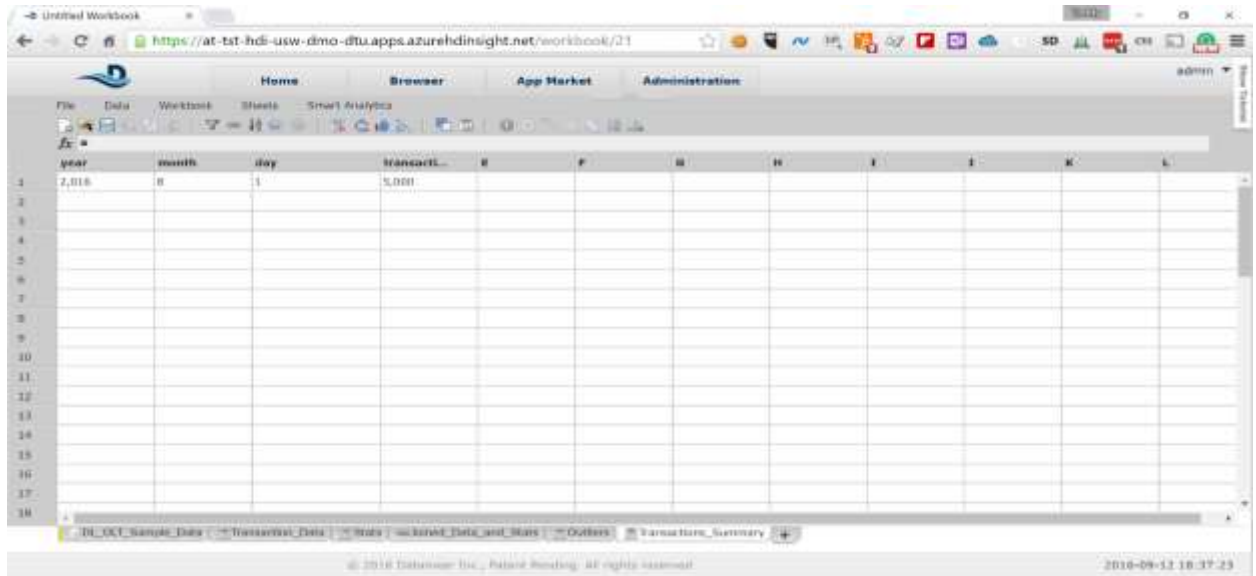
20. We have created summary sheet for our transaction data. Rename the field names as follows:

year

month

day

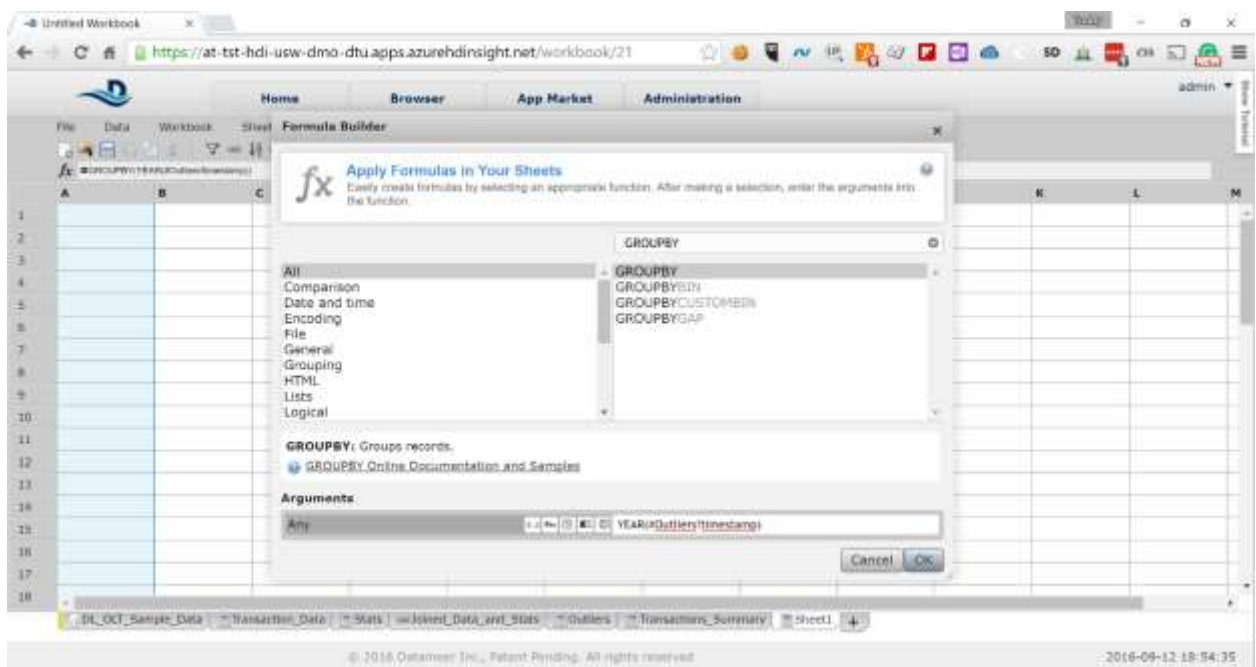
transactions_count



Also, rename the sheet to *Transactions_Summary*

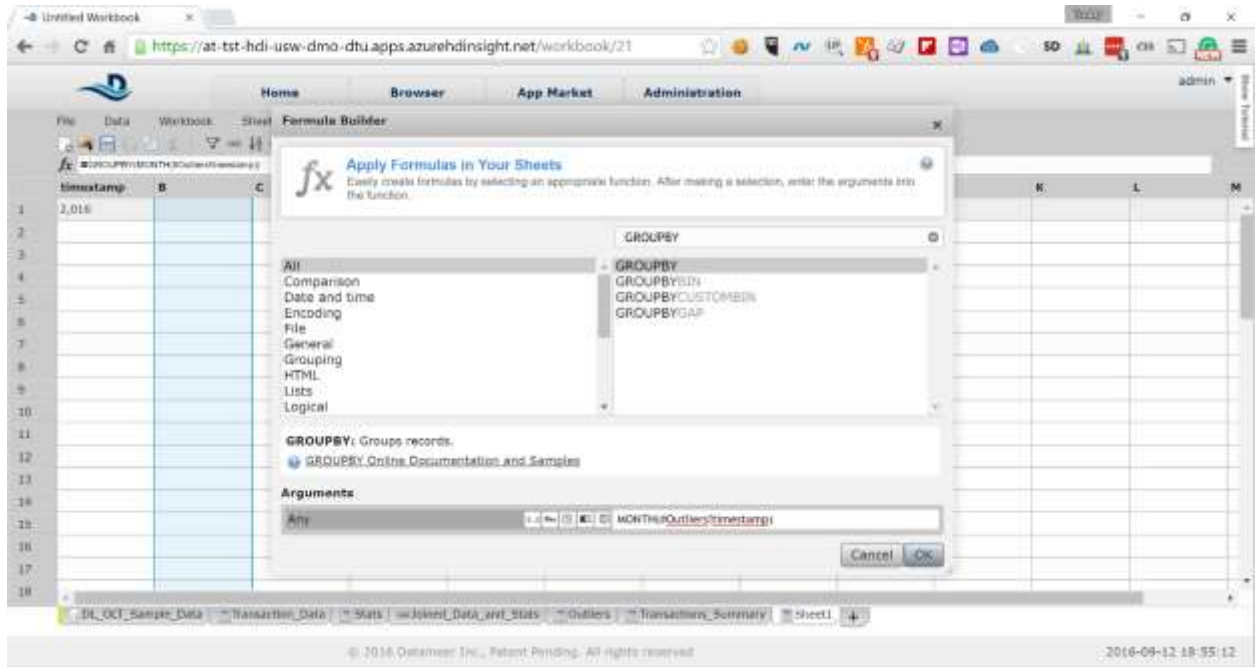
21. Let's create similar summary for the outliers. Create new sheet and in the formula pop-up select the *GROUPBY* function. In the *Arguments* field type the following formula:

YEAR(#Outliers!timestamp)



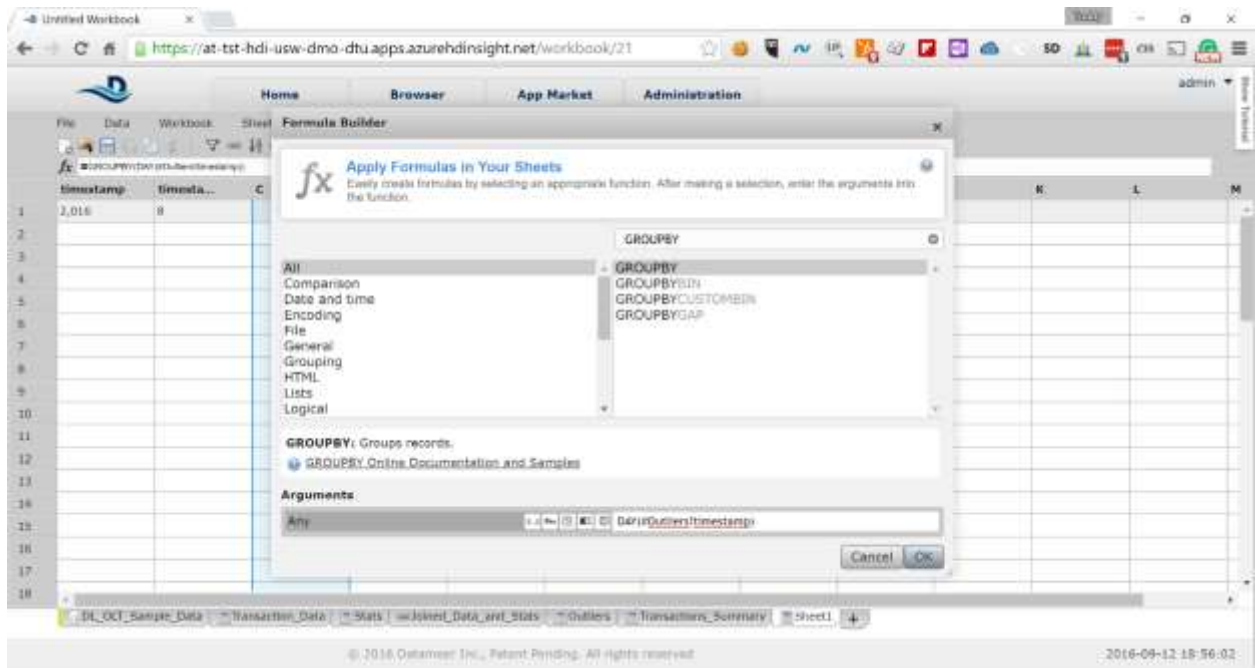
22. Click on the next column and in the formula pop-up select again the *GROUPBY* function and paste the following formula in the *Arguments* field:

MONTH(#Outliers!timestamp)

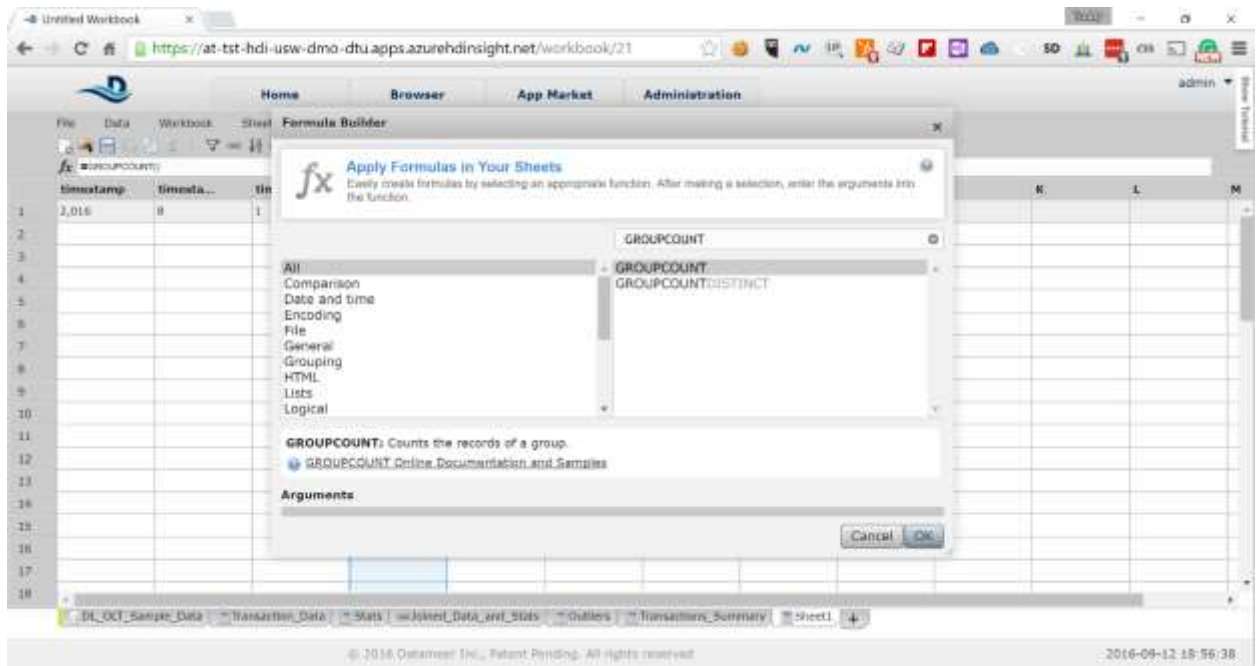


23. Click on the third column and in the formula pop-up select again the *GROUPBY* function and paste the following formula in the *Arguments* field:

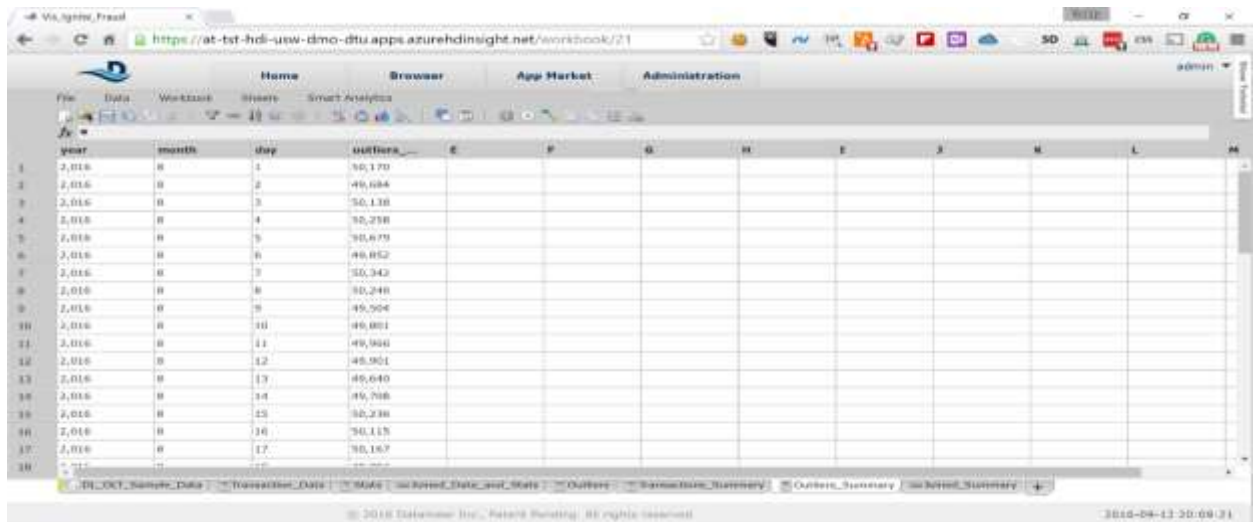
DAY(#Outliers!timestamp)



24. Click on the fourth column and in the formula pop-up select the *GROUPCOUNT* function and click the *OK* button



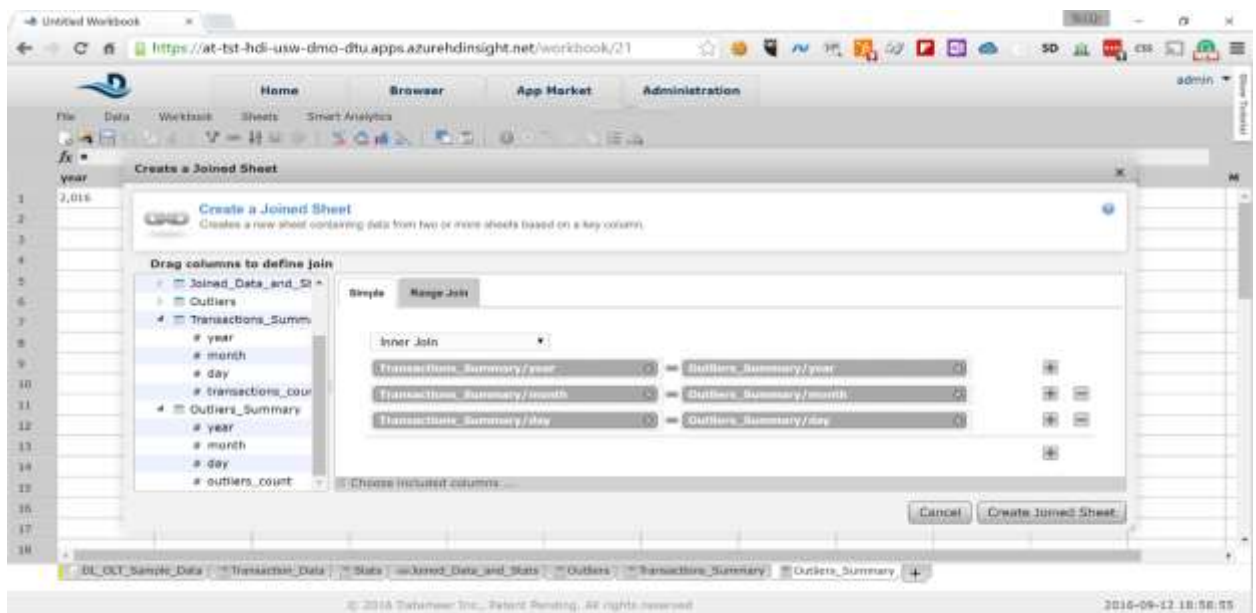
25. We have created summary sheet for our transaction data. Rename the field names as follows:
- year*
 - month*
 - day*
 - outliers_count*



year	month	day	outliers_count	transactions_count
2016	0	1	50,170	49,504
2016	0	2	49,684	49,504
2016	0	3	50,138	49,504
2016	0	4	50,238	49,504
2016	0	5	50,679	49,504
2016	0	6	49,852	49,504
2016	0	7	50,342	49,504
2016	0	8	50,248	49,504
2016	0	9	49,504	49,504
2016	0	10	49,881	49,504
2016	0	11	49,901	49,504
2016	0	12	49,901	49,504
2016	0	13	49,640	49,504
2016	0	14	49,708	49,504
2016	0	15	50,238	49,504
2016	0	16	50,115	49,504
2016	0	17	50,187	49,504

Also, rename the sheet to *Outliers_Summary*

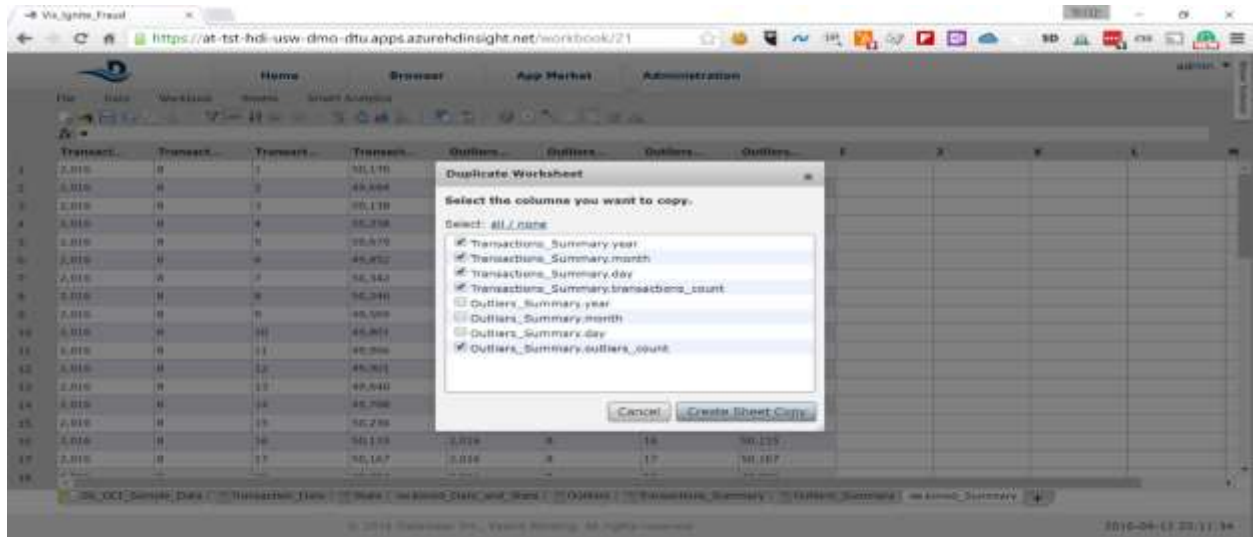
26. We need to join the two summary sheets to have the results available in a single sheet for visualization. Select *Data* -> *Join* and join the *Transactions_Summary* and *Outliers_Summary* sheets by year, month and date as on the picture below by clicking on the *Create Joined Sheet* button



Rename the joined sheet to *Joined_Summary*

27. Let's copy the joined sheet and remove the duplicate data from it. Right-click on the *Joined_Summary* sheet and select *Duplicate*. Select the following fields in the pop-up:

Transactions_Summary.year
Transactions_Summary.month
Transactions_Summary.day
Transactions_Summary.transactions_count
Outliers_Summary.outliers_count



28. Rename the sheet to *Summary* and the columns as follows:

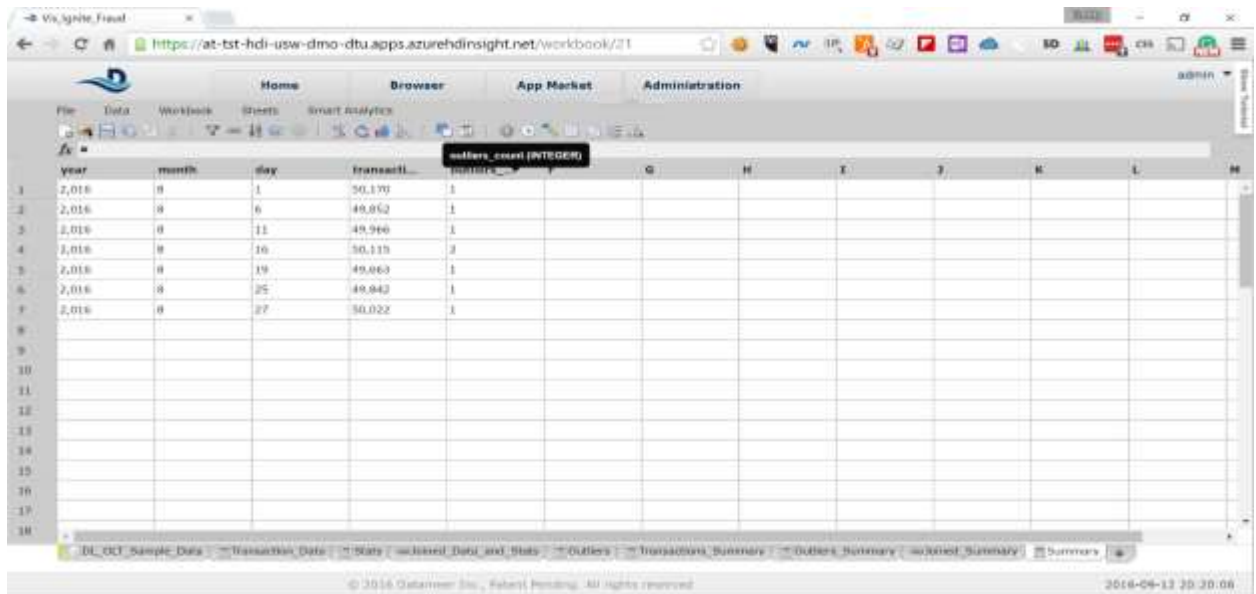
Transactions_Summary.year -> *year*

Transactions_Summary.month -> *month*

Transactions_Summary.day -> *day*

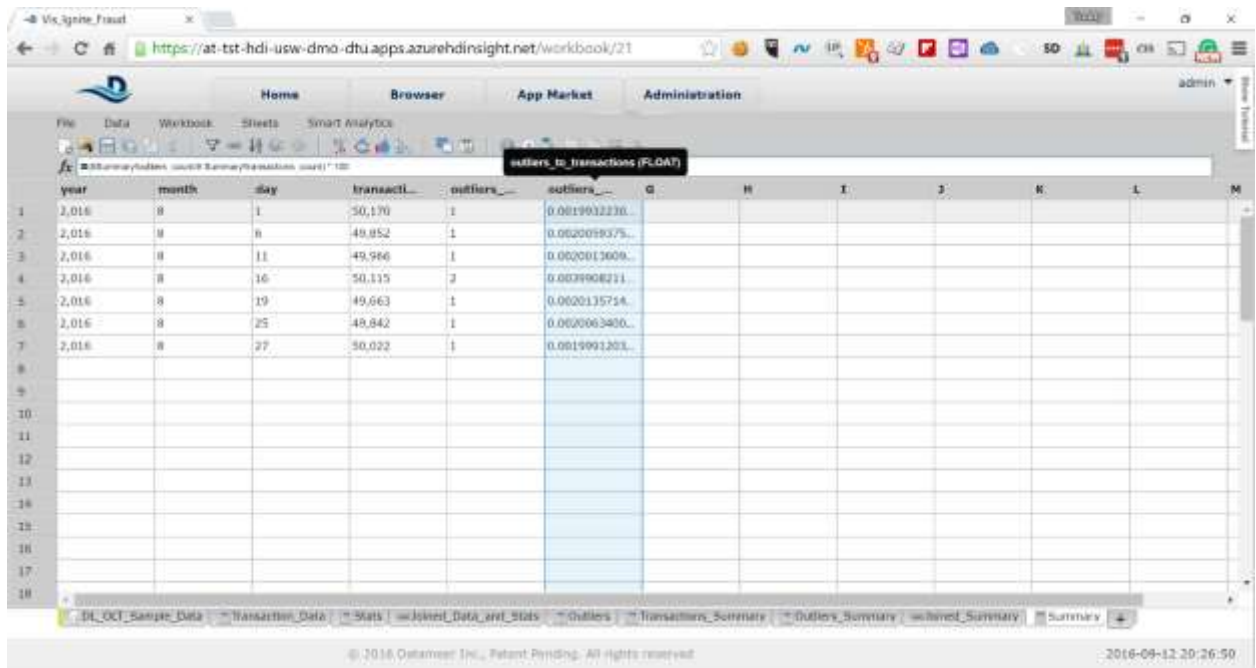
Transactions_Summary.transactions_count -> *transactions_count*

Outliers_Summary.outliers_count -> *outliers_count*



29. Click on the sixth column and cancel the formula pop-up. In the f_x field on top of the sheet type the following:

$(\#Summary!outliers_count/\#Summary!transactions_count) * 100$

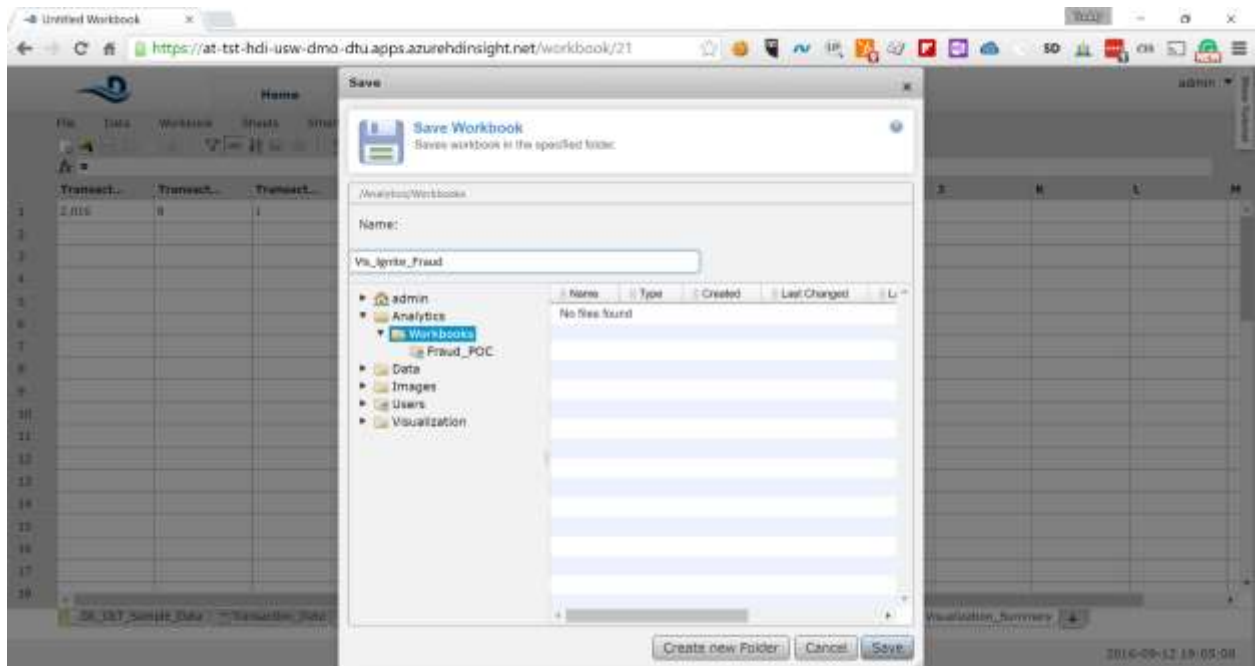


The screenshot shows the 'Vis_Ignite_Fraud' workbook in the Azure Data Explorer interface. The 'outliers_to_transactions (FLOAT)' column is highlighted in blue. The data is as follows:

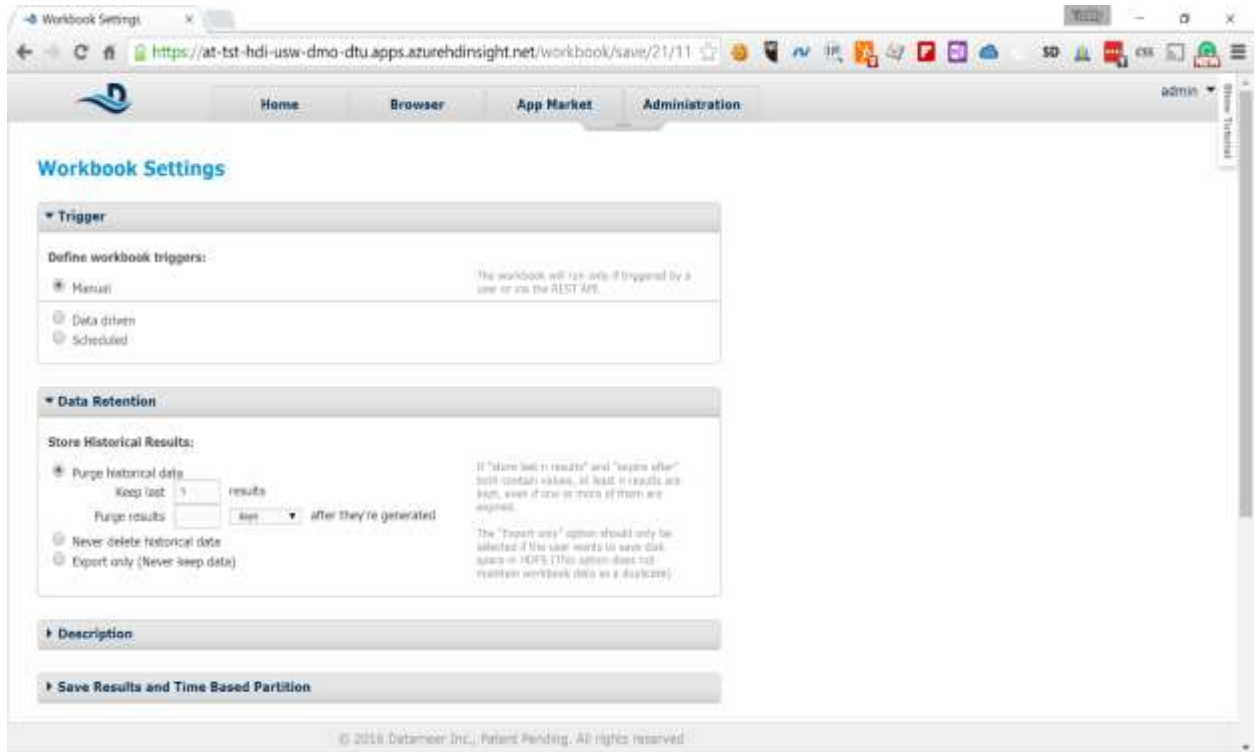
	year	month	day	transacti...	outliers_...	outliers_...	G	H	I	J	K	L	M
1	2,016	8	1	50,170	1	0.0019012230...							
2	2,016	8	8	49,952	1	0.0020019375...							
3	2,016	8	11	49,986	1	0.0020013609...							
4	2,016	8	16	50,115	2	0.0039908211...							
5	2,016	8	19	49,663	1	0.0020135714...							
6	2,016	8	25	48,842	1	0.0020063400...							
7	2,016	8	27	50,022	1	0.0019991203...							
8													
9													
10													
11													
12													
13													
14													
15													
16													
17													
18													

Also, rename the field to *outliers_to_transactions*

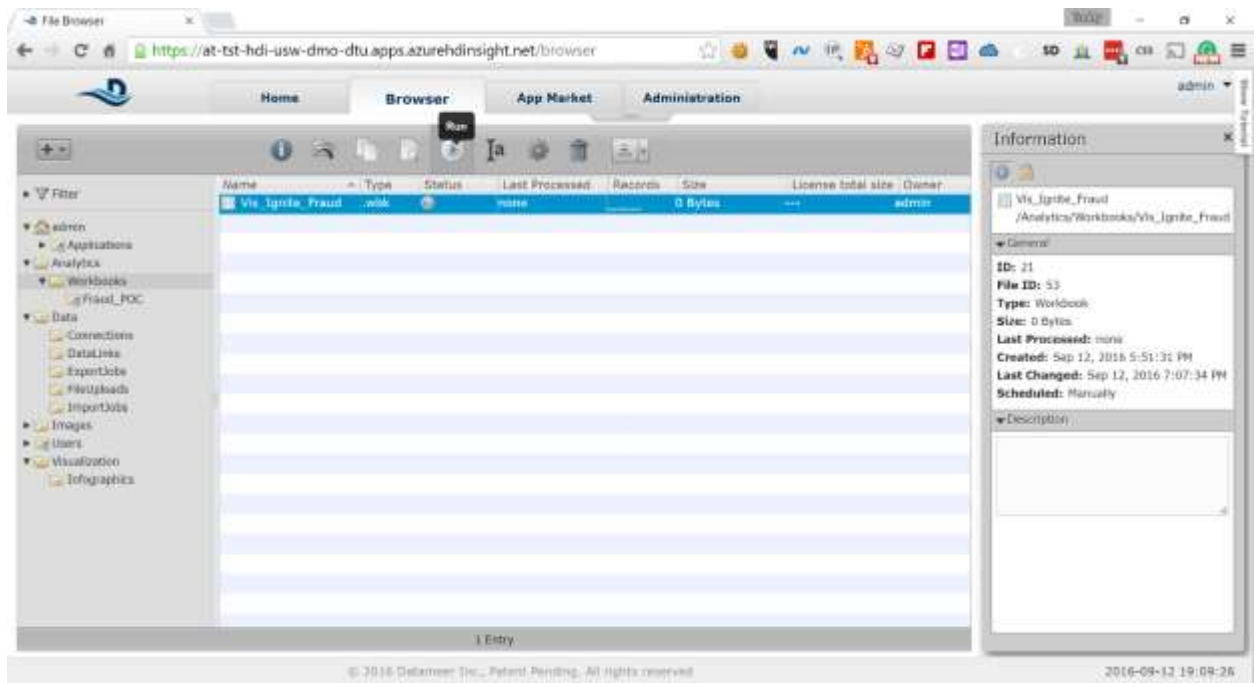
30. Select *File* -> *Save* from the menu and type the *Vis_Ignite_Fraud* in the *Name* field



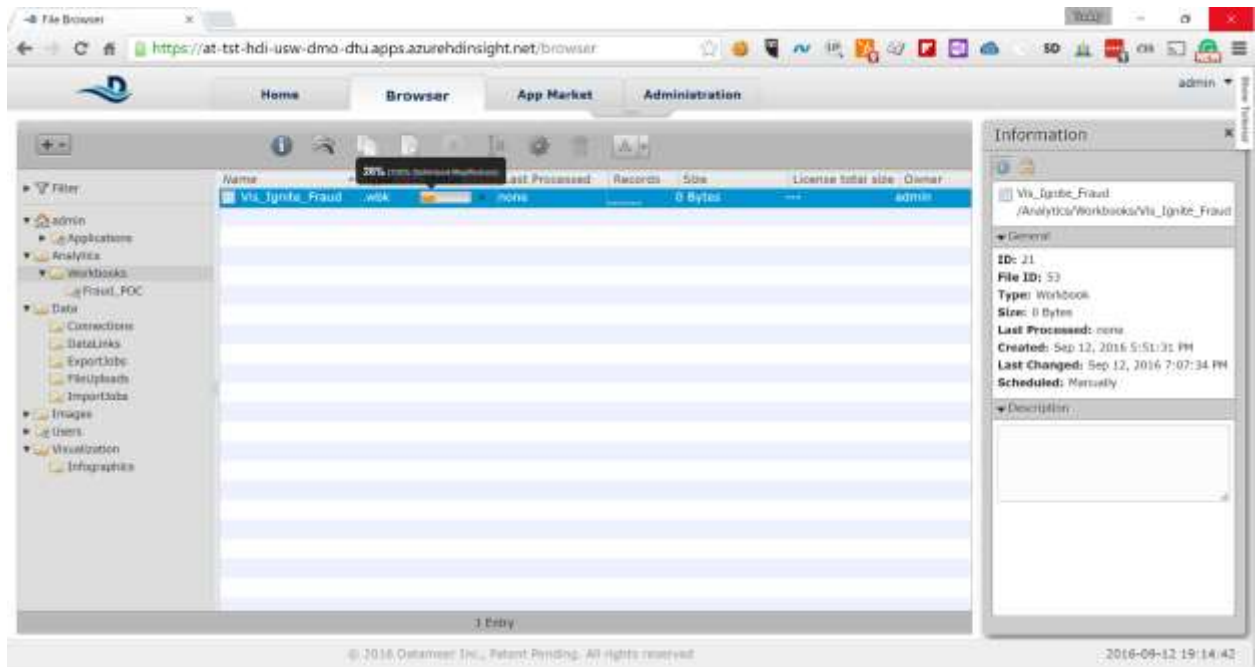
31. On the next screen keep the default values for all the fields. Scroll down and click on the *Save* button again



32. In the list of workbooks select the newly created workbook and click on the run button from the toolbar. This will trigger the calculation on the full data set



You will see updates in the *Status* column, showing you how the Hadoop job is progressing.



8 Logging in to the TrendMicro DSM

8.1 Server name

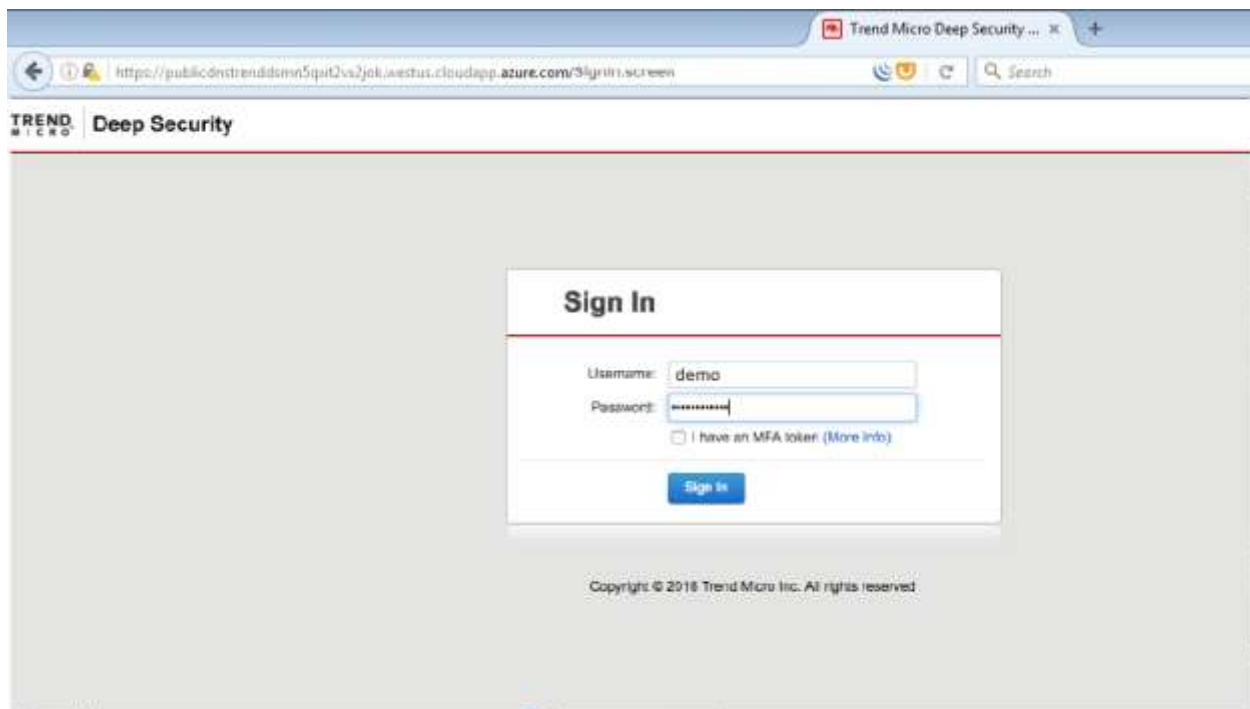
From the output section of the deployment you can get the URL for TrendMicroDSM, Splunk and Chef Server (Microsoft.Template)



8.2 Server login

To login to TrendMicro DSM

- Paste the TrendMicro DSM URL in the browser
- Enter the **Username** and **Password** provided in the parameter section during the deployment



9 Perform policy configuration on the TrendMicro DSM

1. Changing the base policy

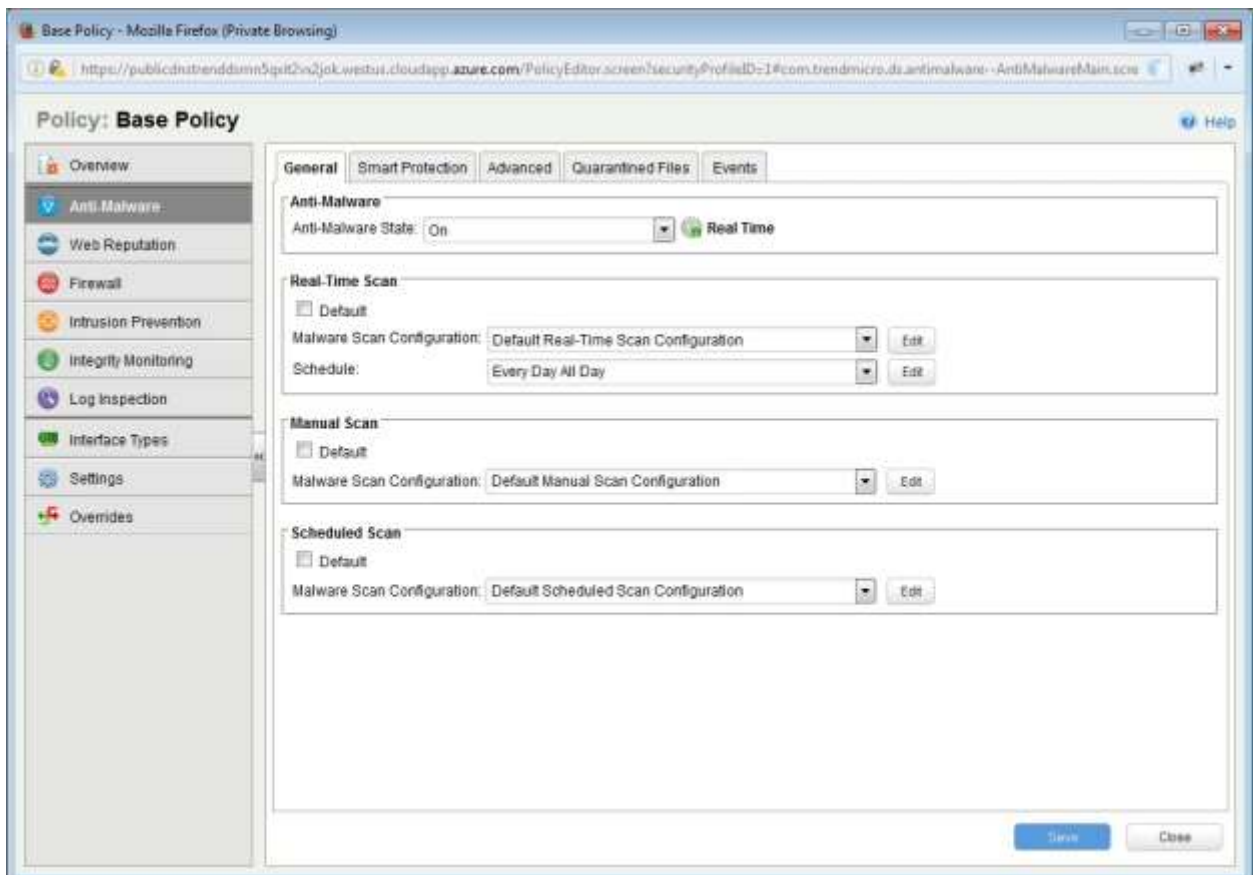
Go to policies->Base Policy



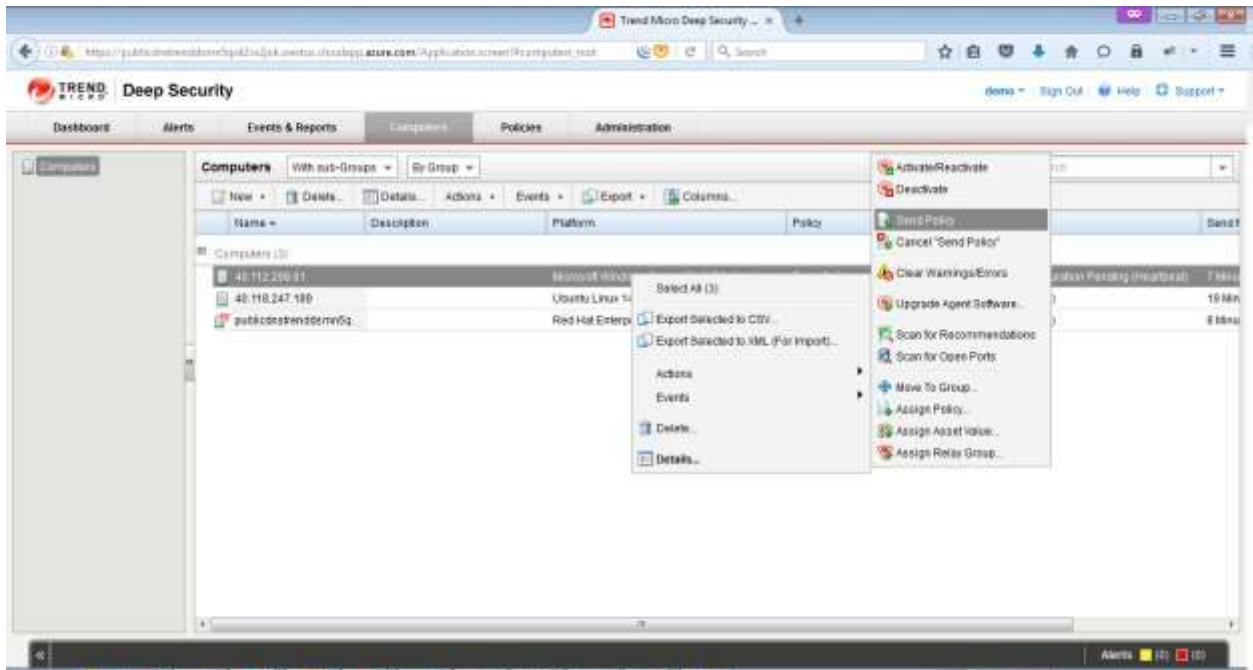
2. Enable Anti-Malware

Go to Anti-malware->Anti-Malware State->On

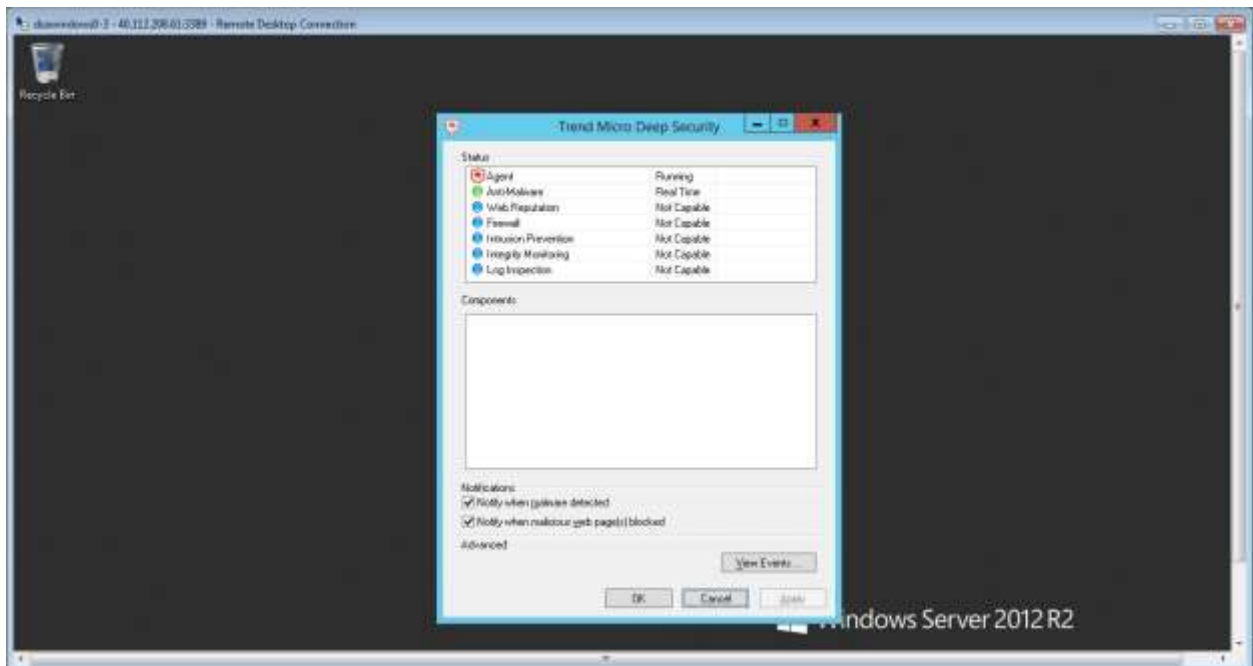
Click "Save"



3. Applying policies to computer



4. Verifying policy in the computer



10 Exercises

10.1 Datameer – Visualize the Data

Datameer has powerful Infographics to Visualise the data. In this exercise, the data analysed in the above configuration will be displayed graphically.

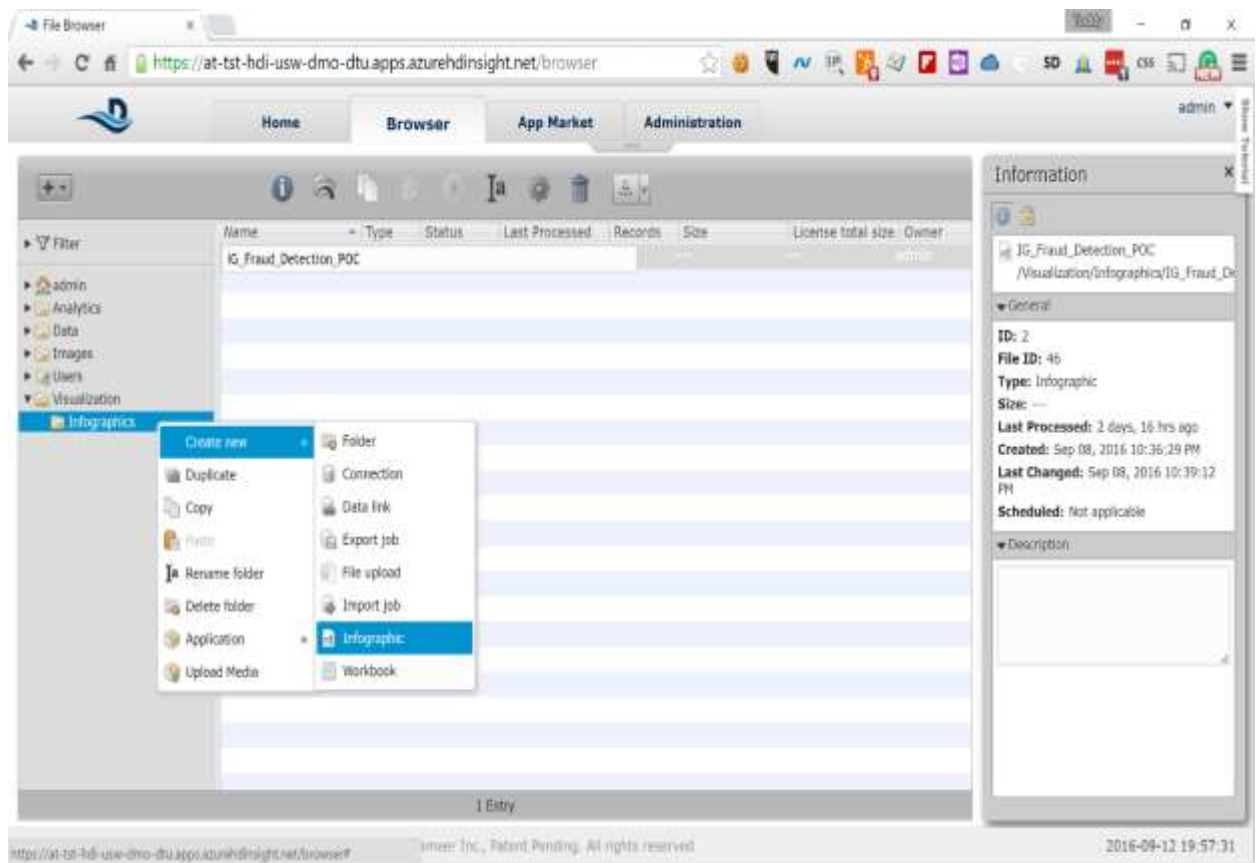
10.2 TrendMicro – Malware test

TrendMicro has security intelligence built-in to protect the systems against the malwares. In this exercise, showcases the TrendMicro DSM malware detection capability

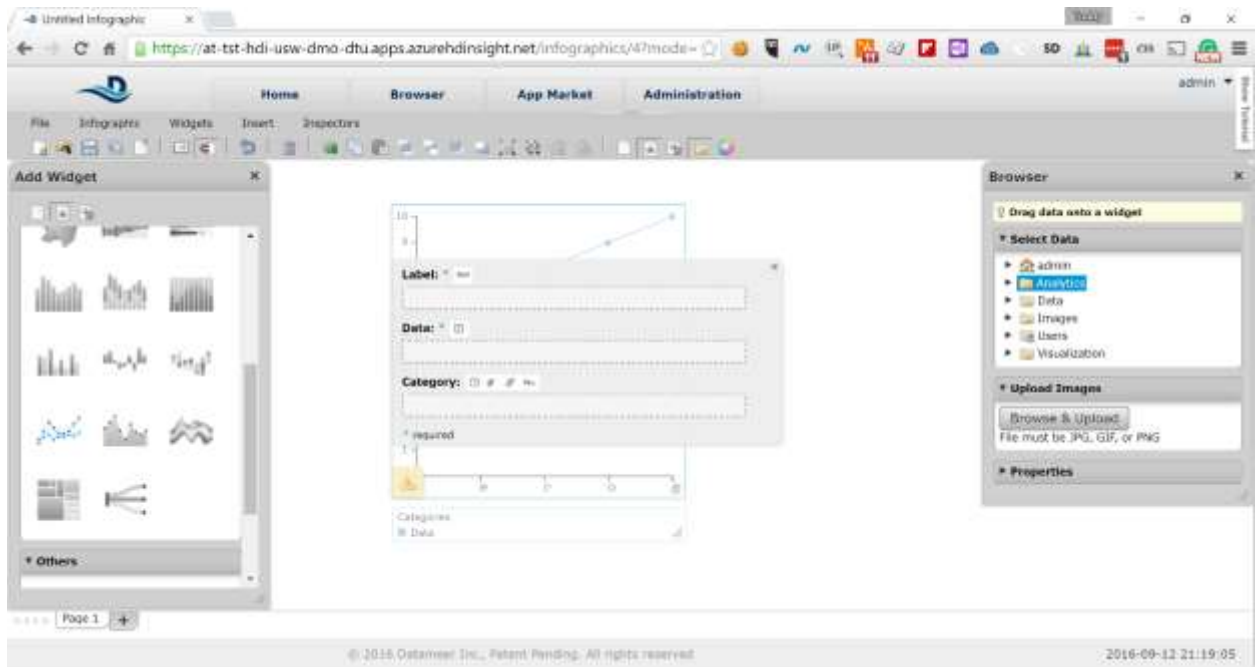
11 Visualize the Data

The last exercise in this HOL is to visualize the data and identify certain days when the irregular transactions have spiked. To do that we will use the following steps:

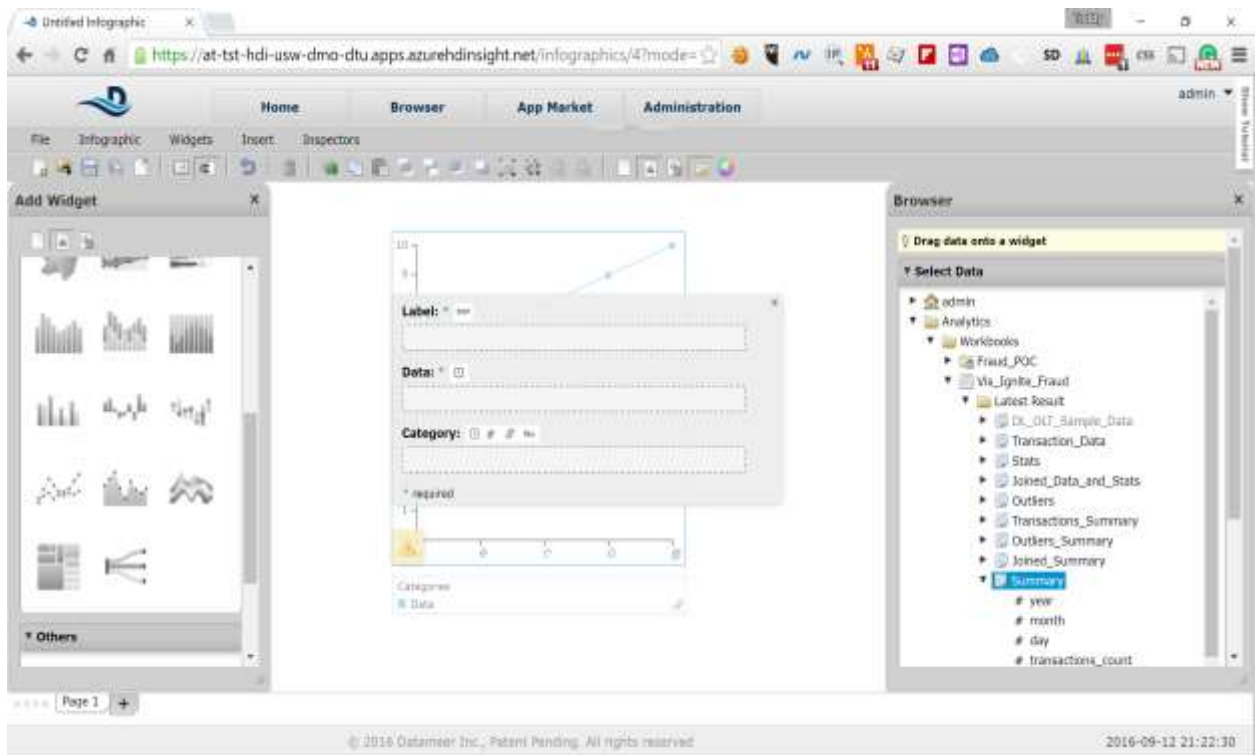
1. In Datameer's Browser view expand the Visualization node and right click on Infographics -> Create New -> Infographic



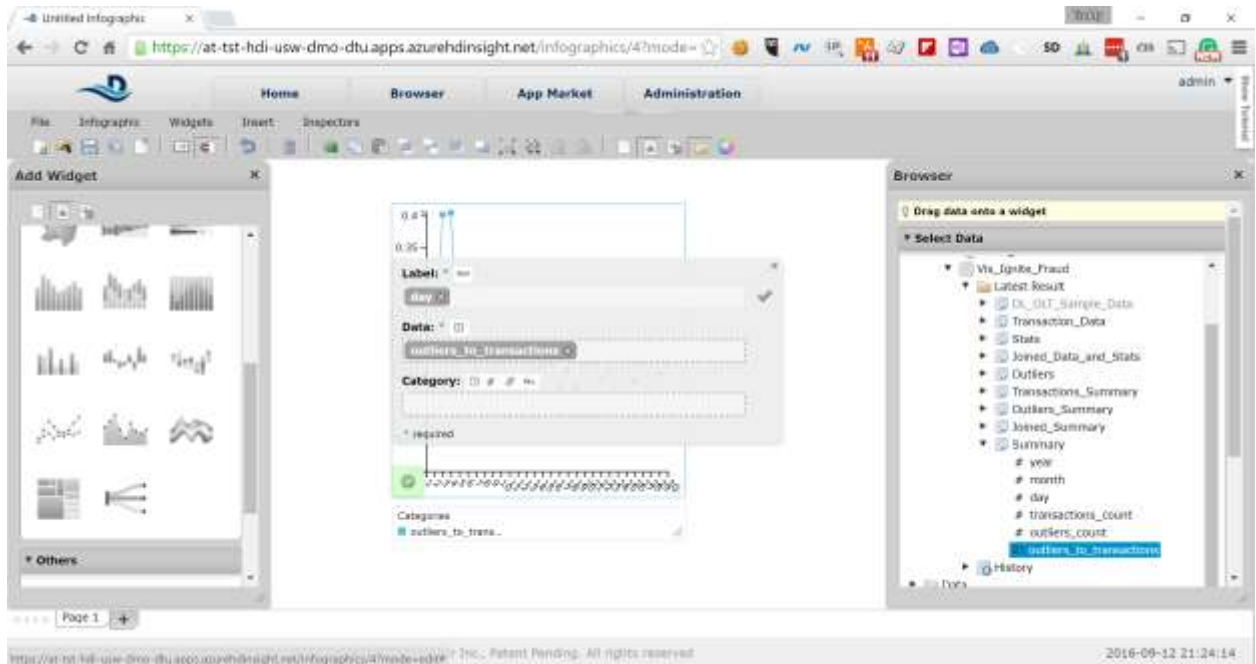
2. Drag the *Line and Area Chart* from the *Add Widget* pane on the left to the work pane in the middle



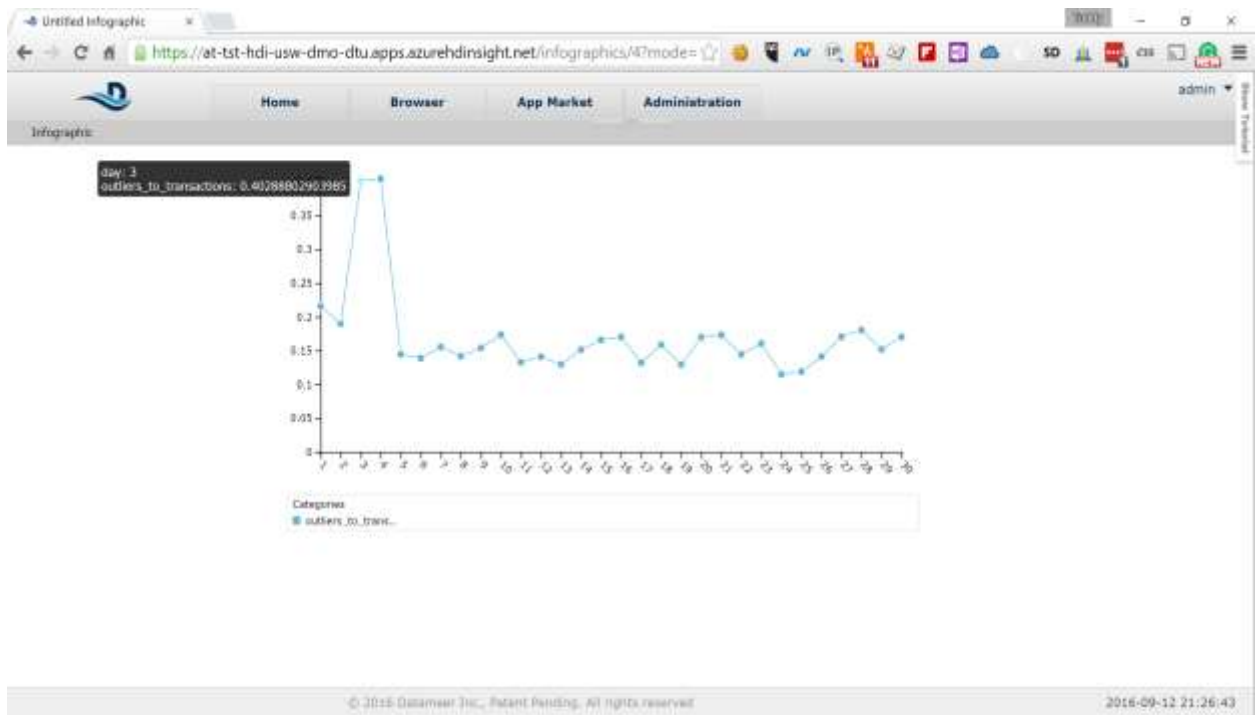
3. In the Browser pane expand *Analytics* node and then *Workbooks* -> *Vis_Ignite_Fraud* -> *Latest Results* -> *Summary*



4. Drag the *day* field to the *Label* input field and the *outliers_to_transactions* field to the *Data* input field in the Work pane

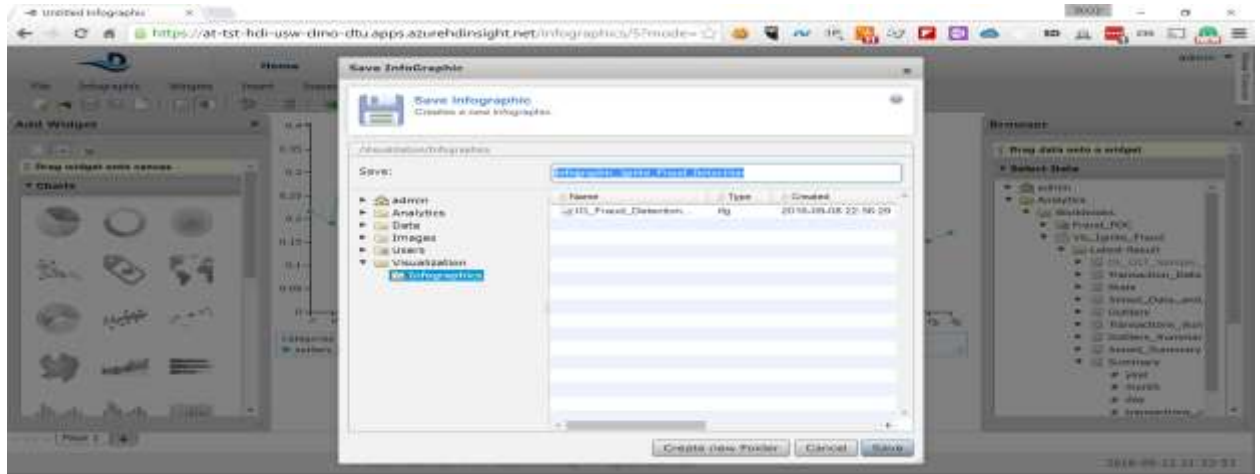


5. Select Infographic -> View from the menu to present the infographic. You can easily see that on the 3rd and 4th day of the month the outliers significantly spiked, which is a sign of something unusual going on those two days



6. Select Infographic -> Edit from the Menu and then File -> Save. Type the following in the Name field:

Infographic_Ignite_Fraud_Detection
and click on the Save button



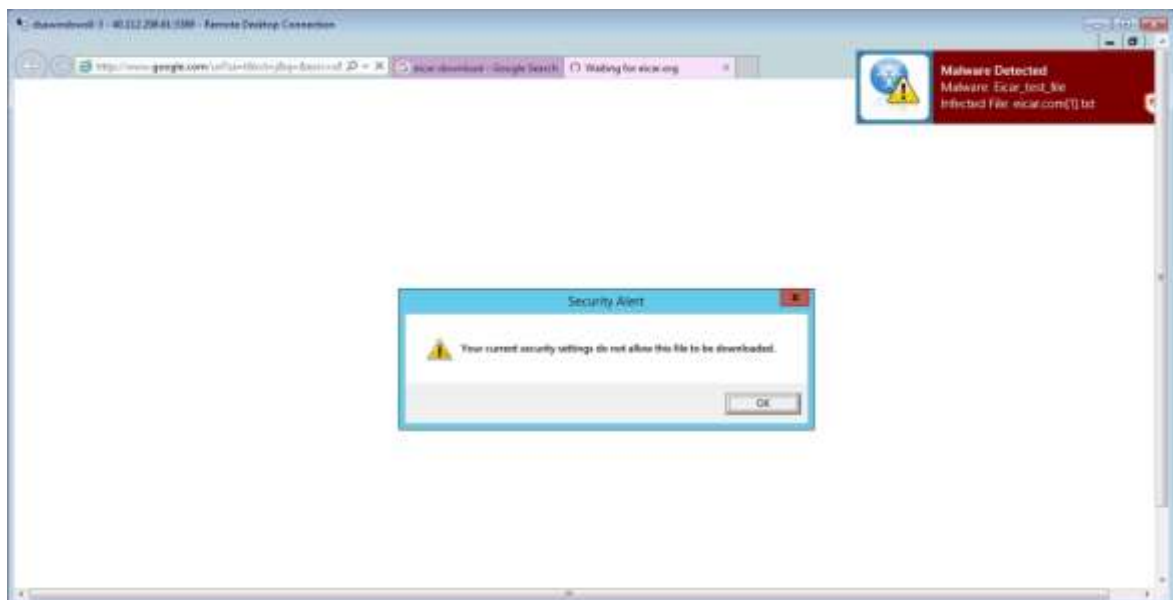
12 Malware Test

12.1 Generating Malware alert in the computer

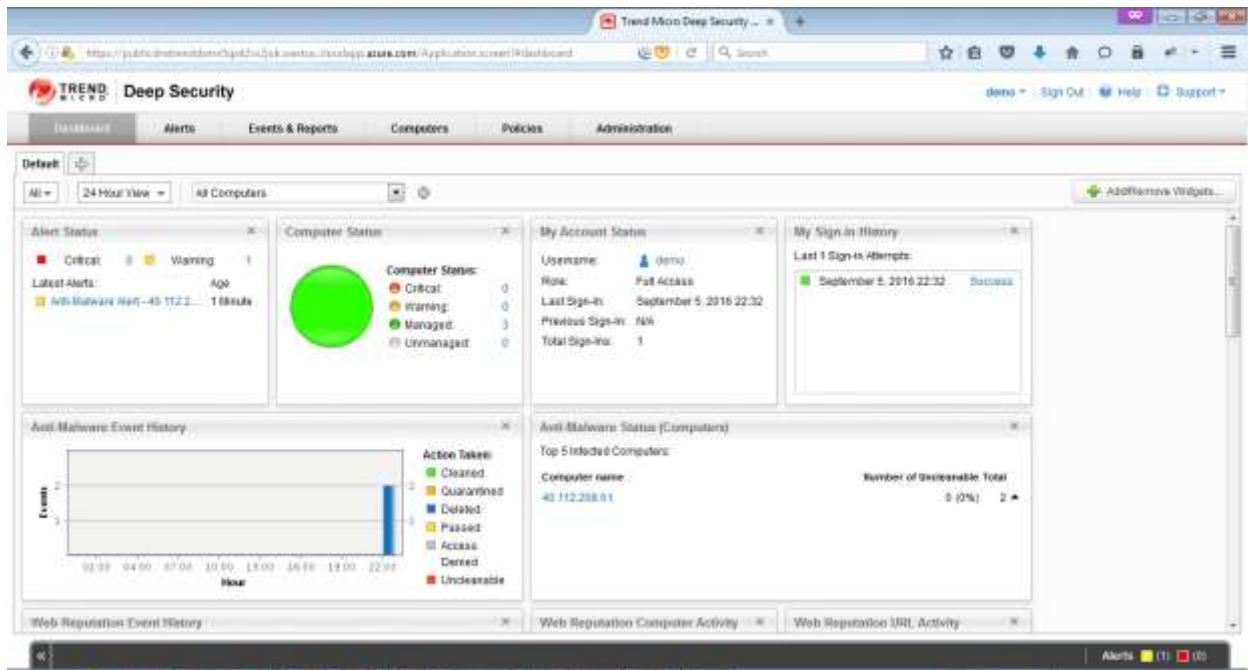
The Malware test can be performed by going to the url

<https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=2&cad=rja&uact=8&sqi=2&ved=0ahUKewj6n6H1-IXPAhUSzGMKHZMGC5AQFggmMAE&url=http%3A%2F%2Fwww.eicar.org%2Fdownload%2Feicar.com.txt&usg=AFQjCNE8DvVI7BE5Nd2hq1zNDTP6hNjclA&bvm=bv.132479545,d.cGc>

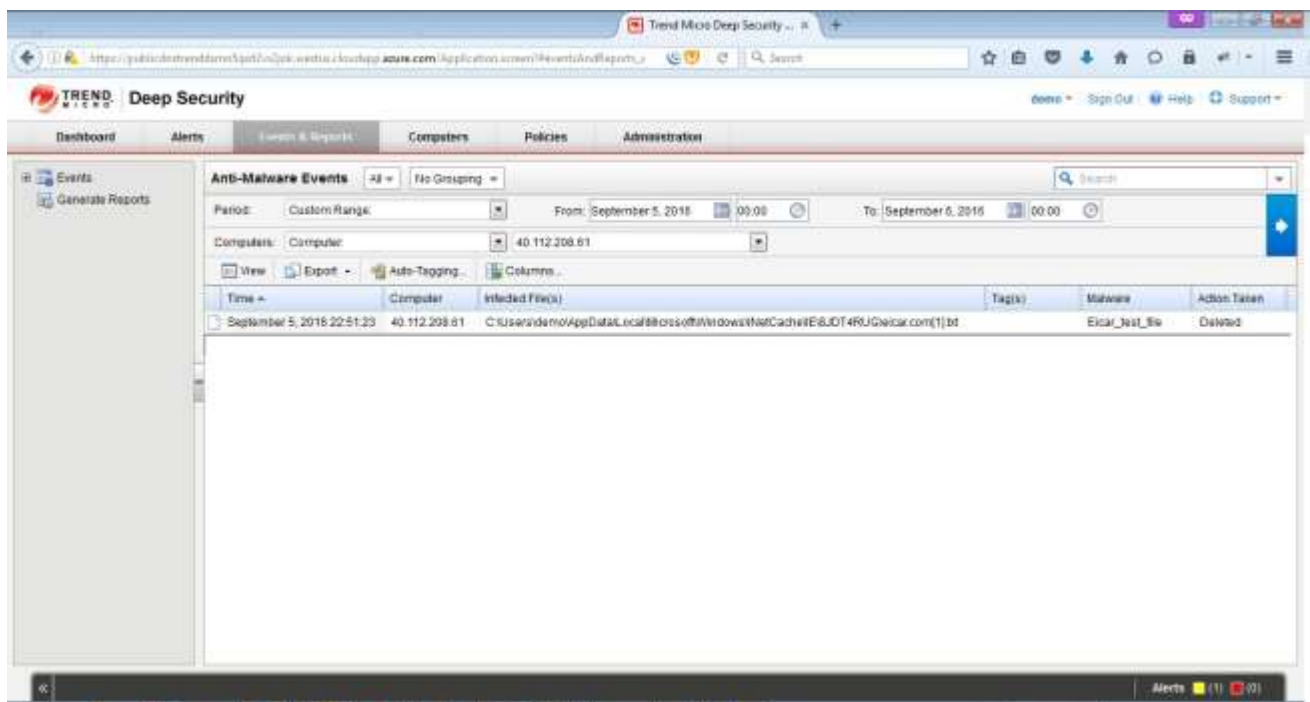
this is eicar malware test



12.2 Dashboard – Malware Alert



12.3 Malware Alert verification



13References, Attachments & Definitions

13.1 References

No.	Document Title	Link/ Attachment	Comments
1			