HOL Guide for Enterprise Risk Analysis

USING DATAMEER, HDINSIGHT, TRENDMICRO DEEP SECURITY AND CHEF

AVYAN CONSULTING CORP | AZUREMARKETPLACE@AVYANCONSULTING.COM

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

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 TendMicro 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

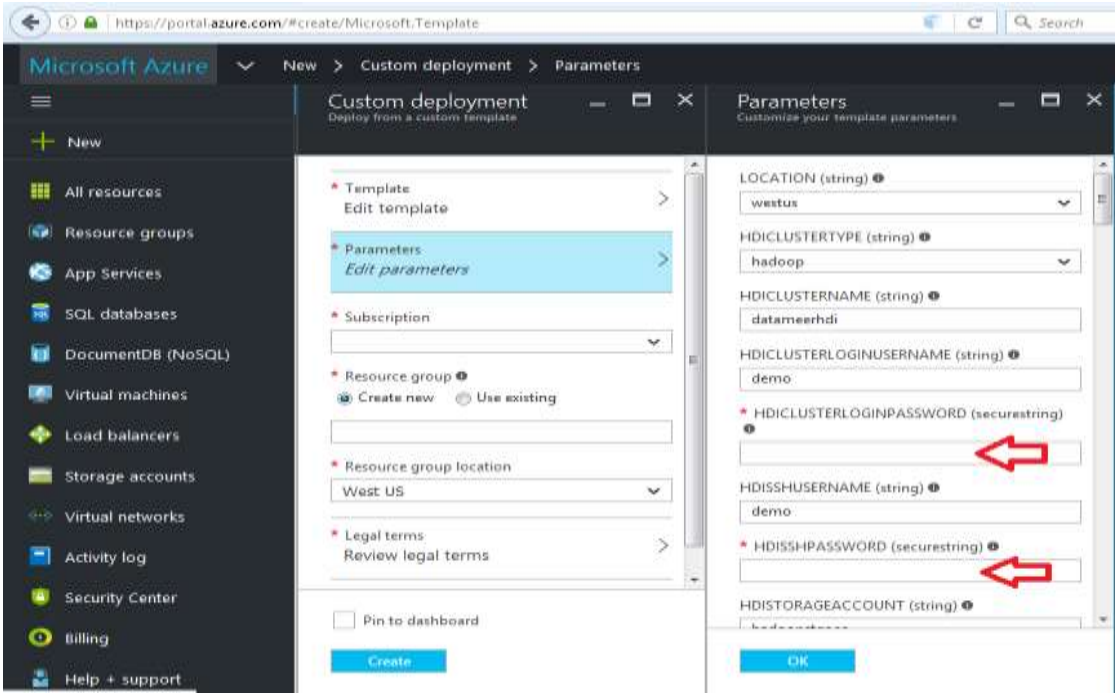
This section will provide you the details of how to deploy this solution in the Microsoft Azure

1) Go to the below link available in the Github

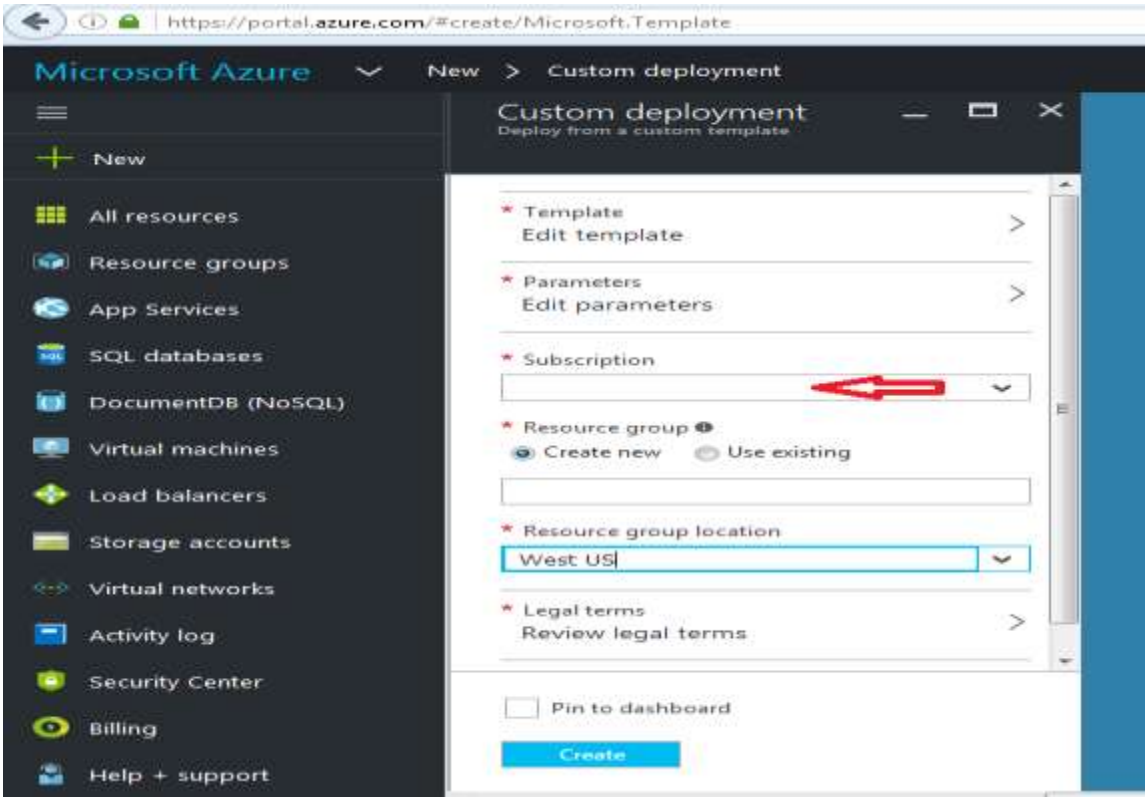
   https://github.com/AvyanConsultingCorp/azure-quickstart-templates/tree/master/datameer-trend-chef-businessintelligence

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

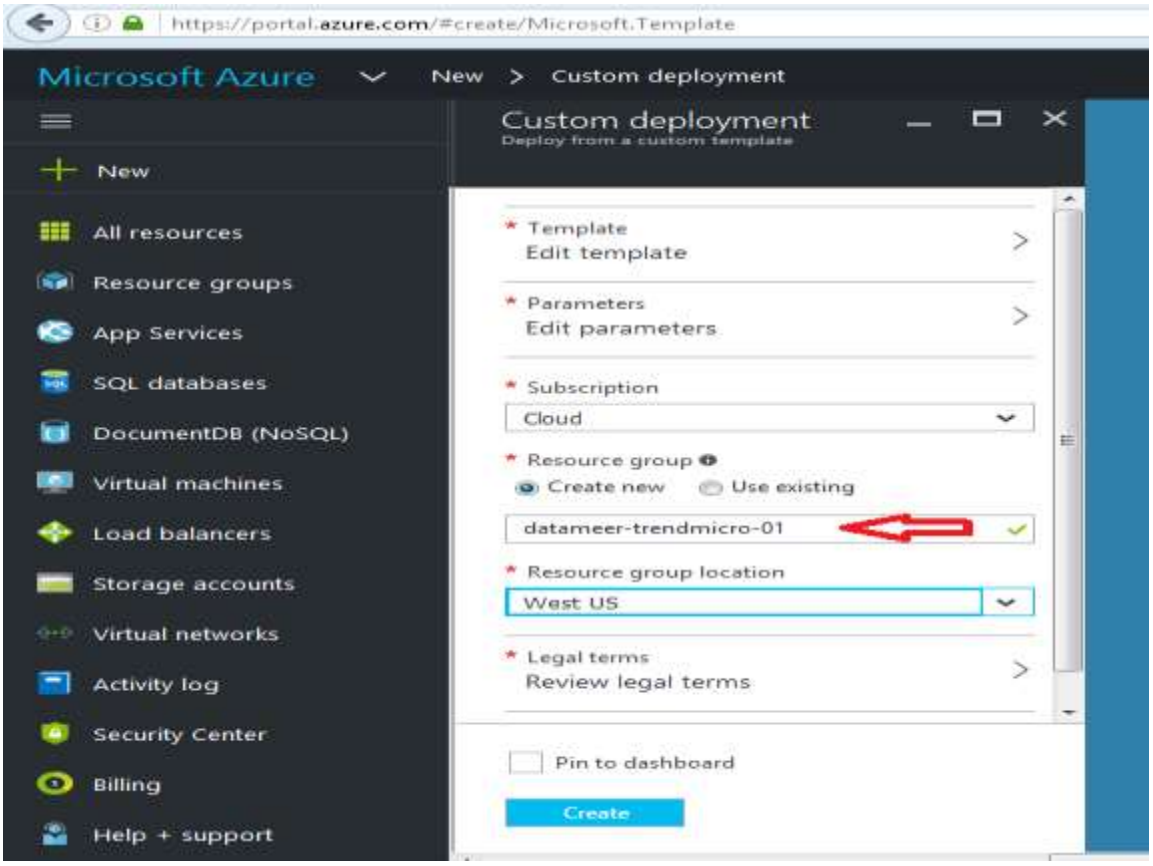
3) Provide the custom paramaters for the solution accordingly and click “Next”



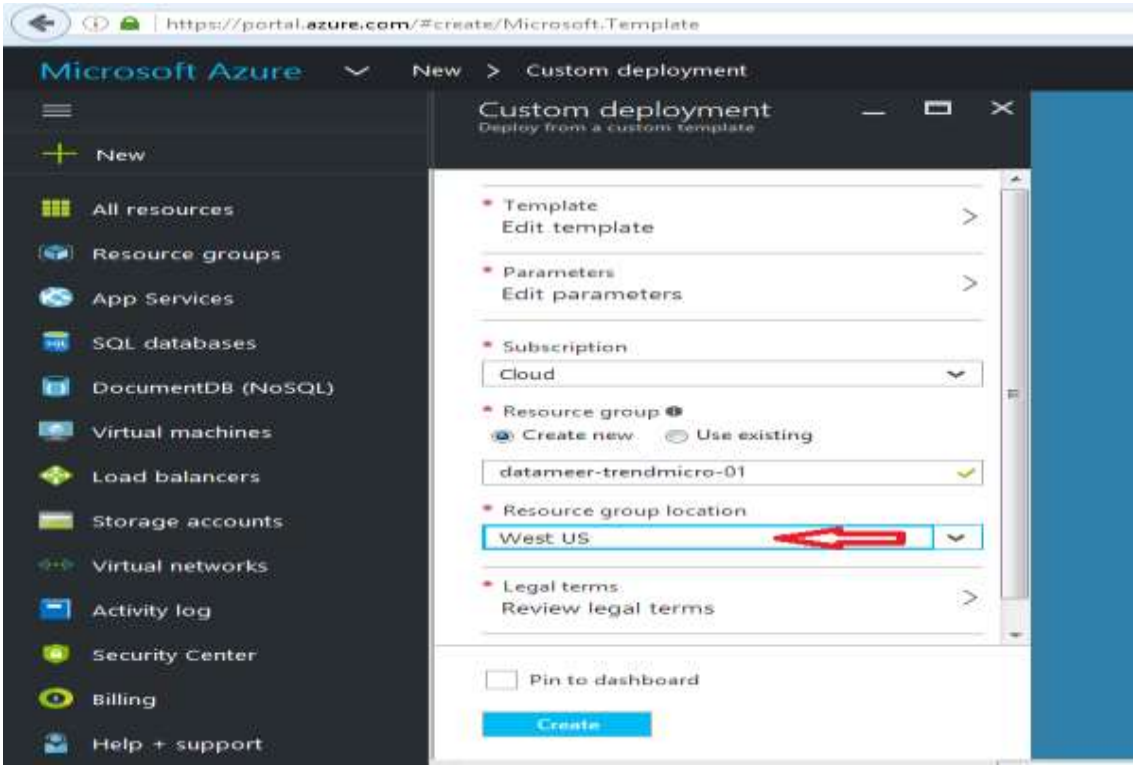
4) You need to select the subscription you want to deploy this solution



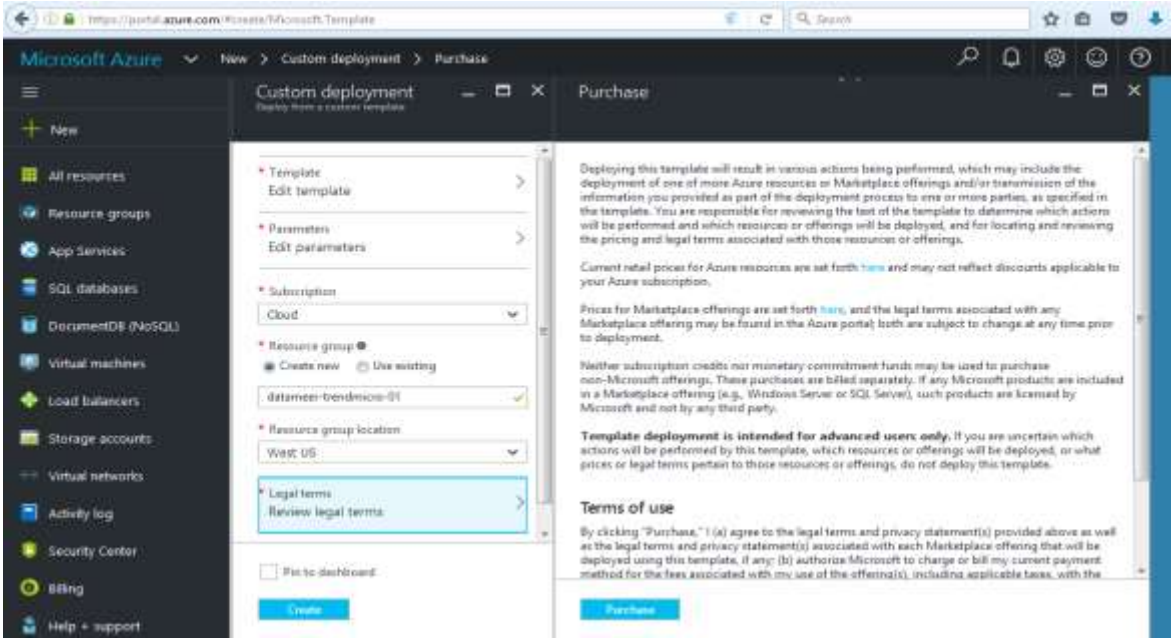
5) Either you can create a new "Resource Group" or use the existing resource group to deploy this solution



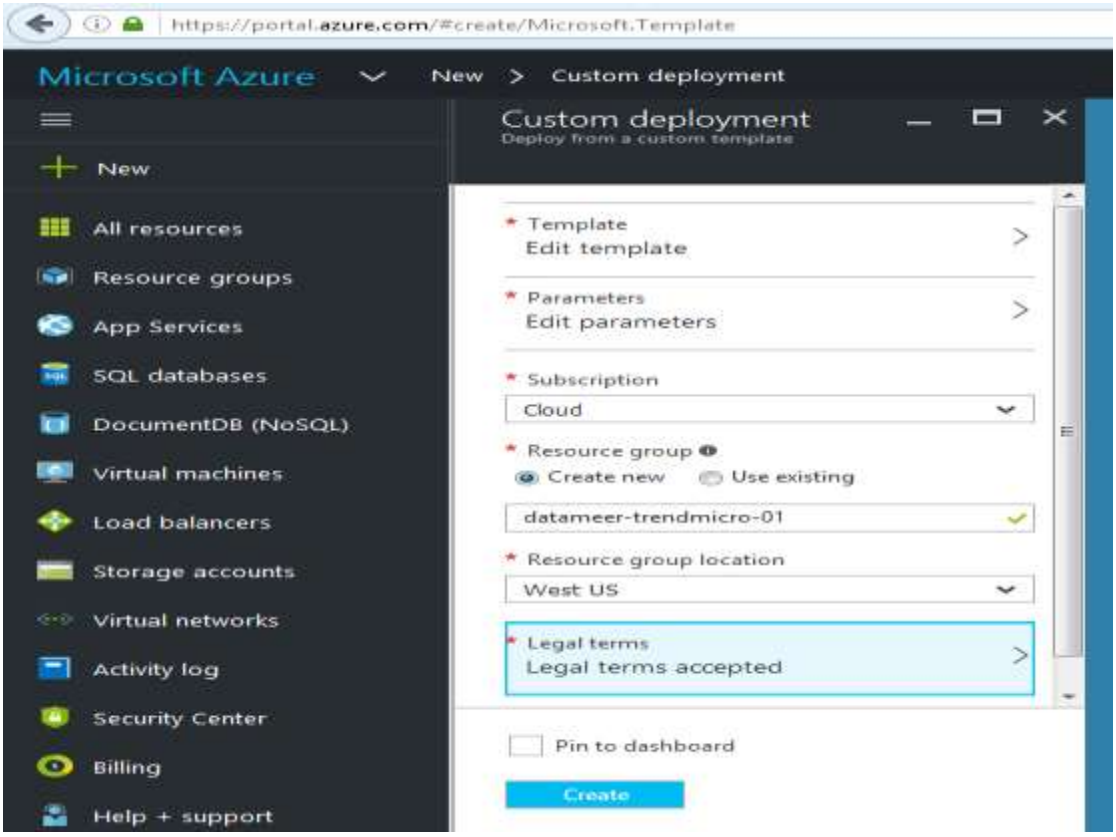
6) Select your choice of Region to deploy this solution,



7) Accept the legal terms to deply the products from the Azure marketplace which includes, Datameer, TredndMicro 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 blog 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 Intrution detection and prevention
- File Integrity monitoring
- Log Inspection

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

This is solution, showcases the Anit-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. 1.5Million 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

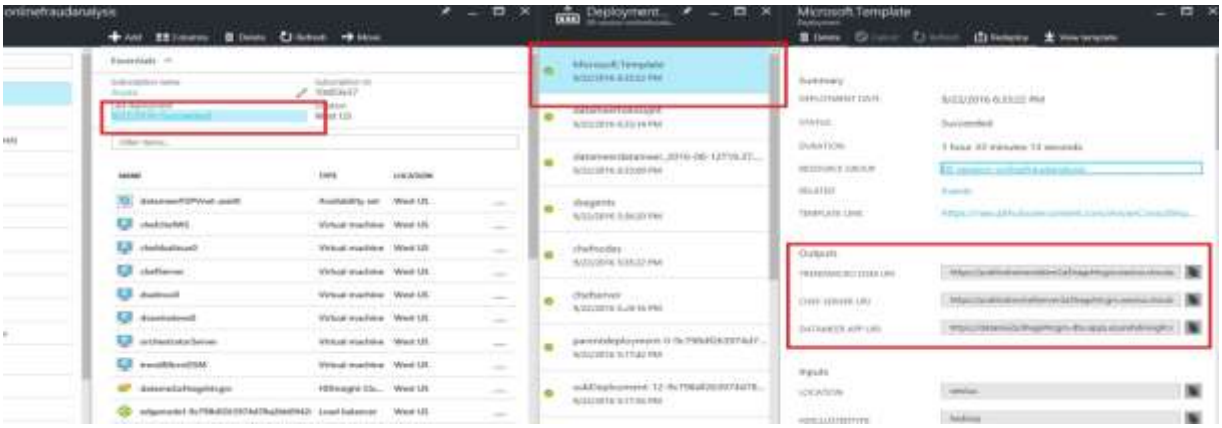
1. Download the file to your desktop
2. Navigate to the deployment resourcegroup and open the datameer storage account
3. Create a new blob container called "samples"
4. Upload the file online-transactions-cc_masked.csv to this container.

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:

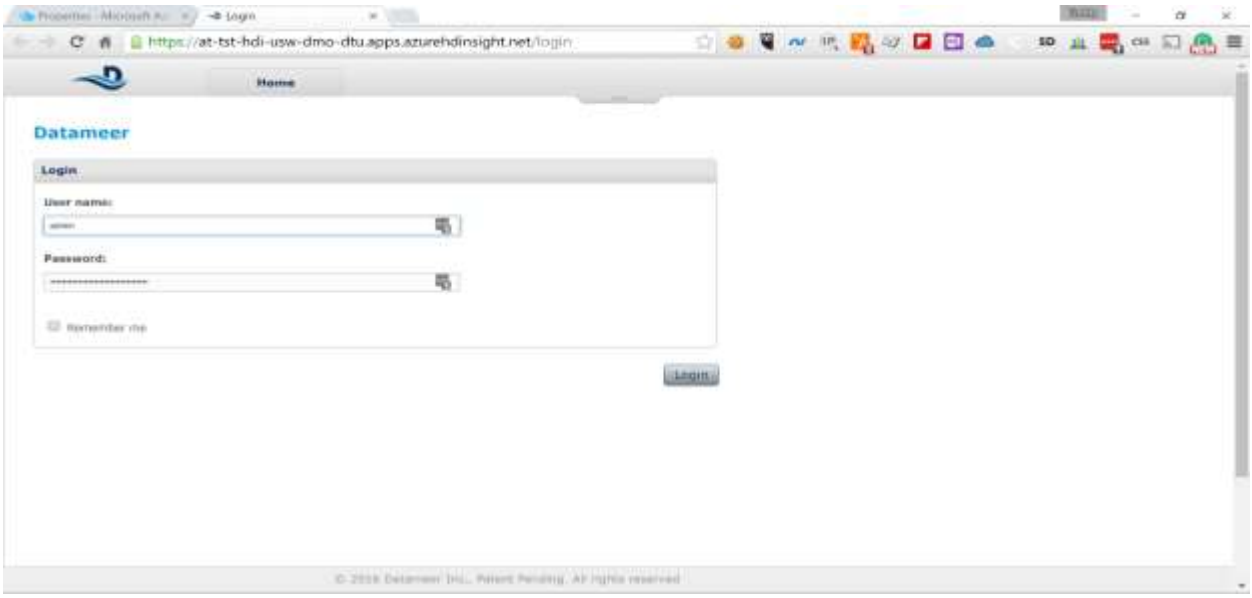
- Accessing it directly via the UI URL
- Accessing it via the Azure Management Portal

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

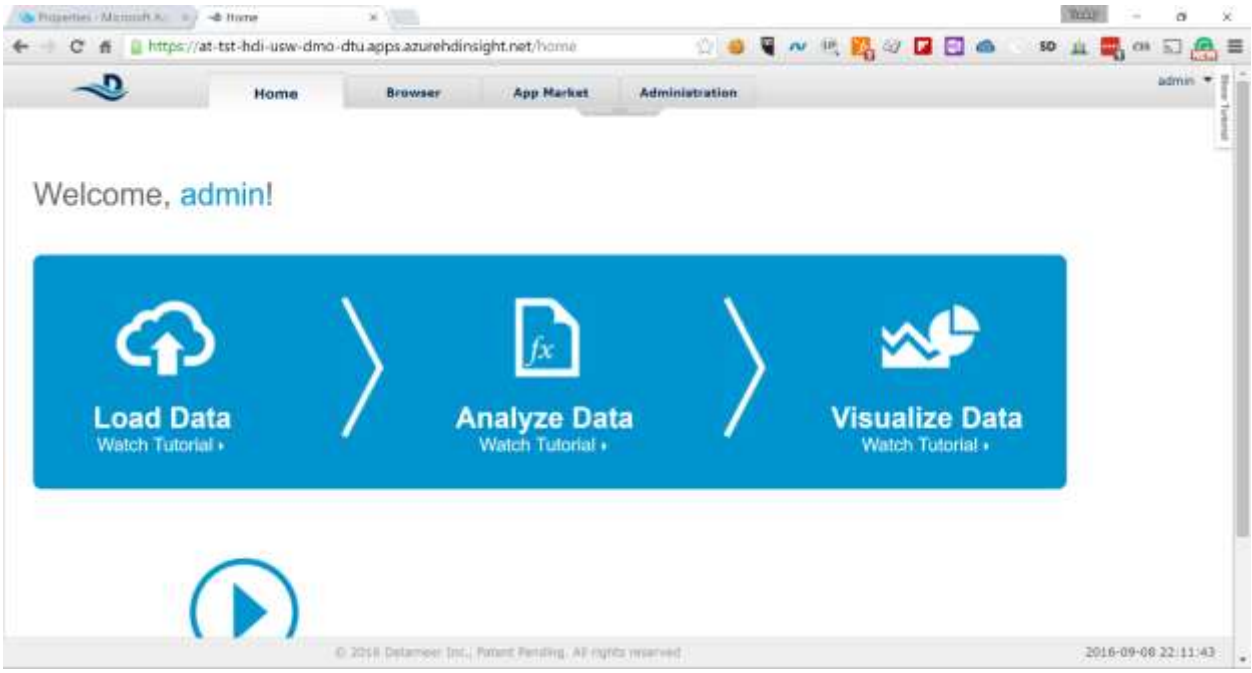


2. Paste the URI in your browser to load the Datameer UI

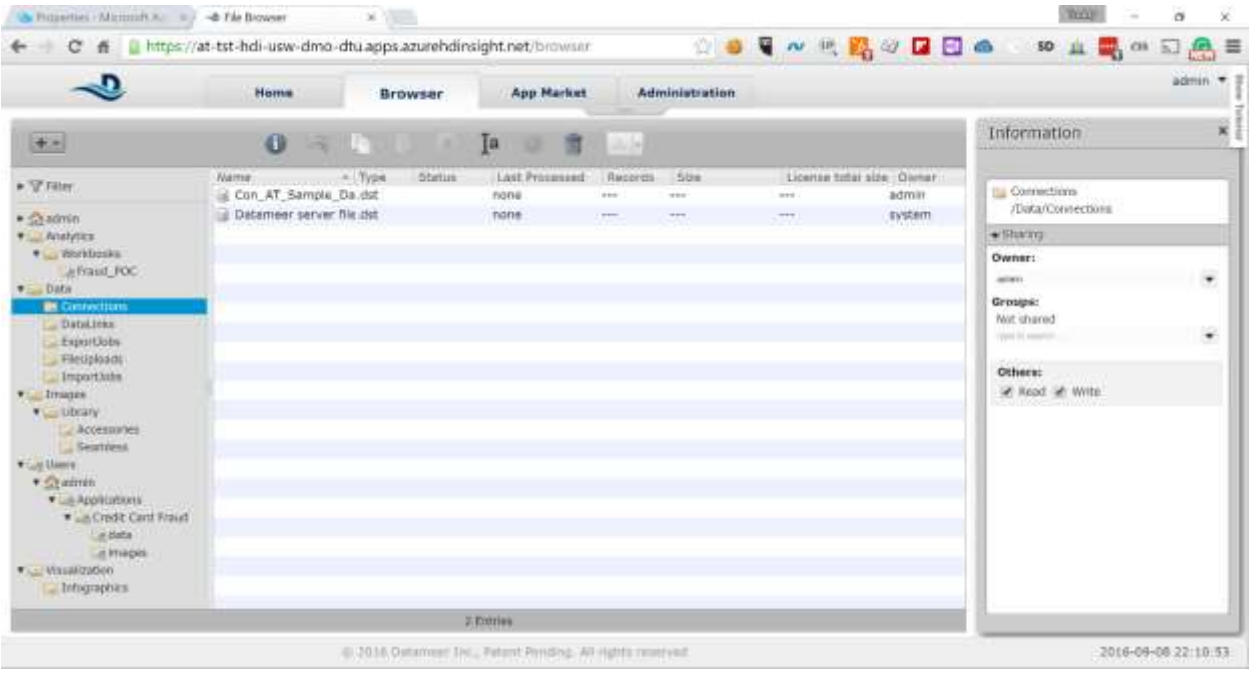


You will be prompted to sign in using your Datameer username and password

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

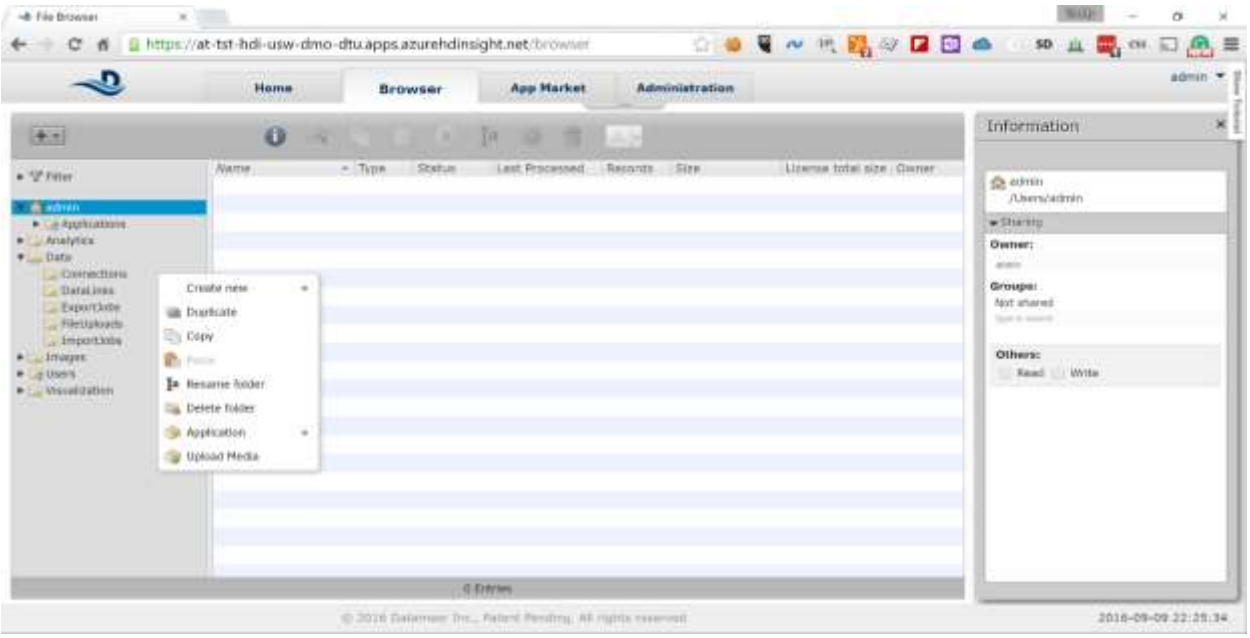


5 Configure Datameer to Fetch Data from Azure Storage

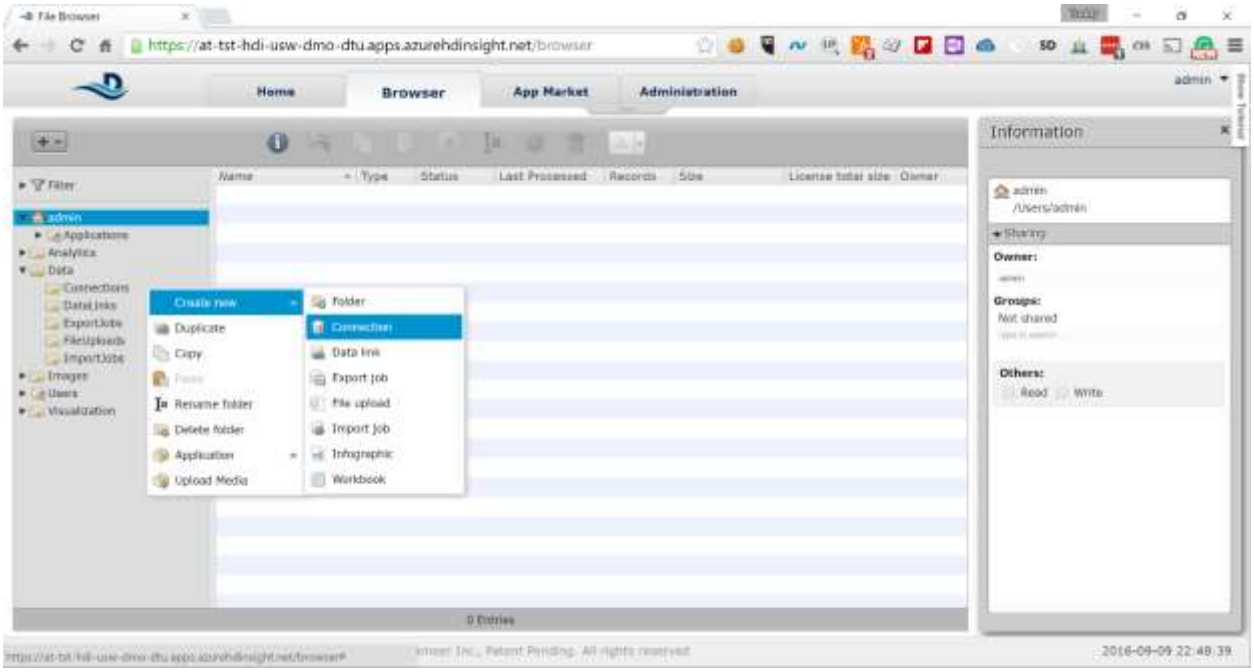
Datameer has more than 65 connectors built in, that allow various systems as data sources. For the purpose of this HOL we will use the Azure Storage connector and fetch the data from there. The assumption is that you have storage account data that contains the transaction data.

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

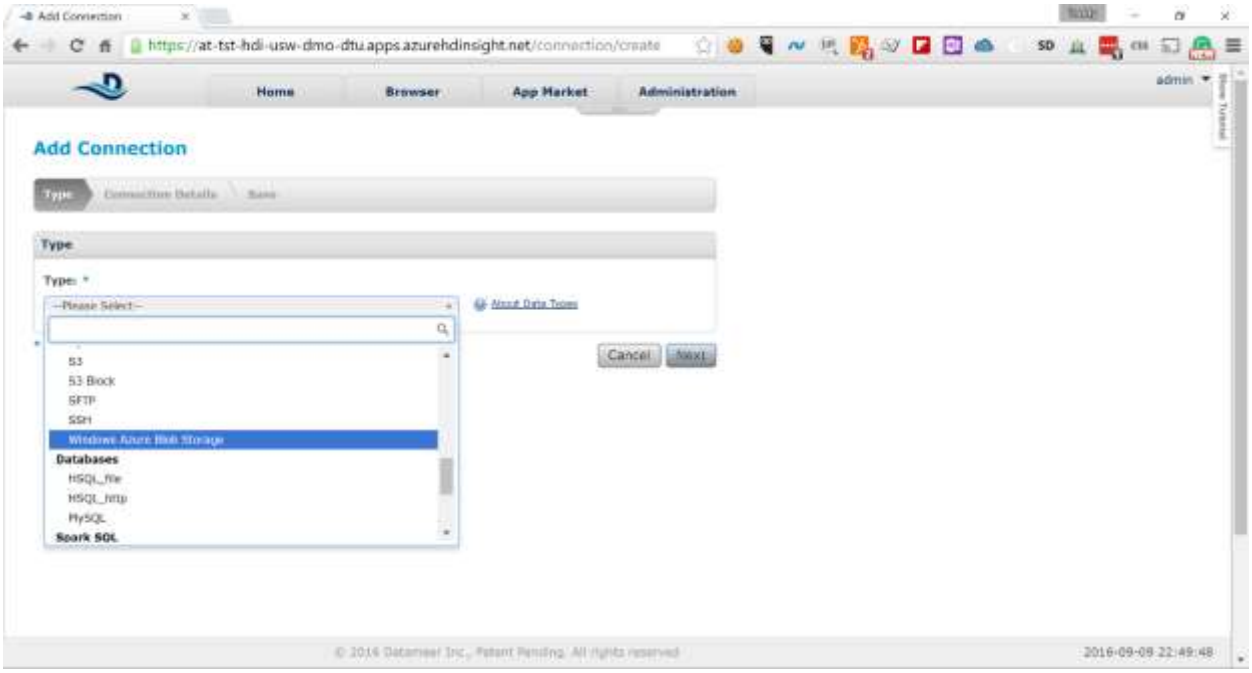
1. Expand the *Data* node in the left-side navigation and right-click on *Connections*



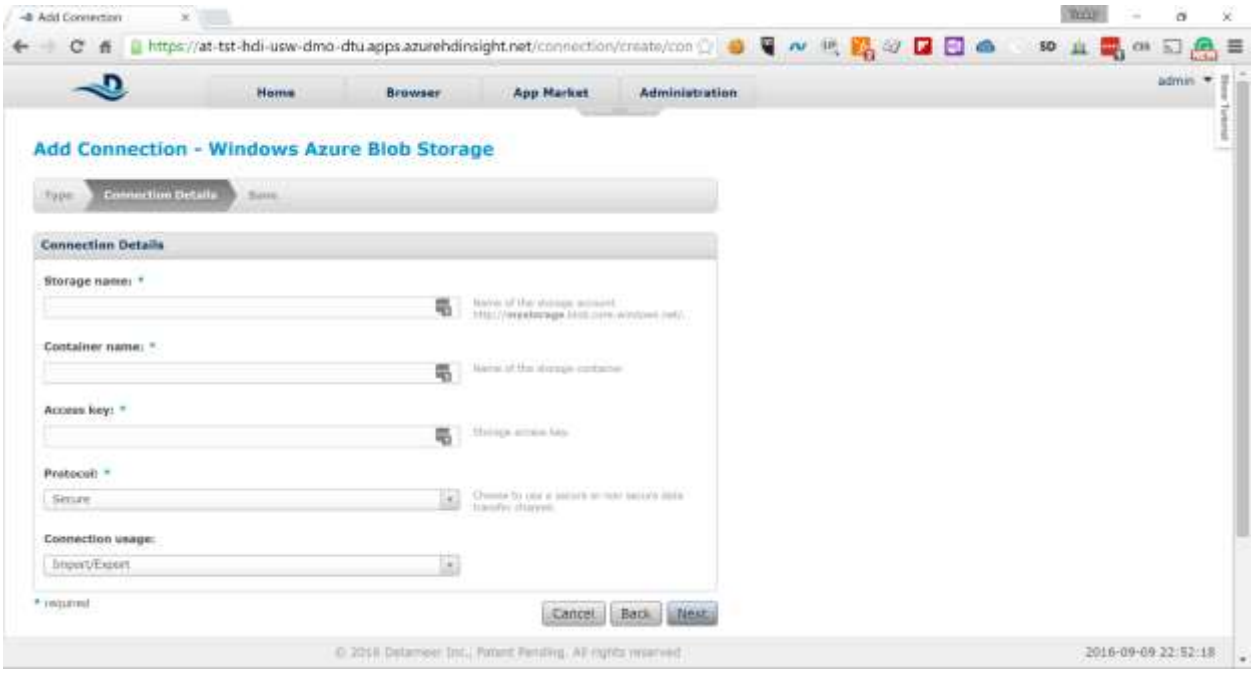
2. Select *Create new -> Connection*



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



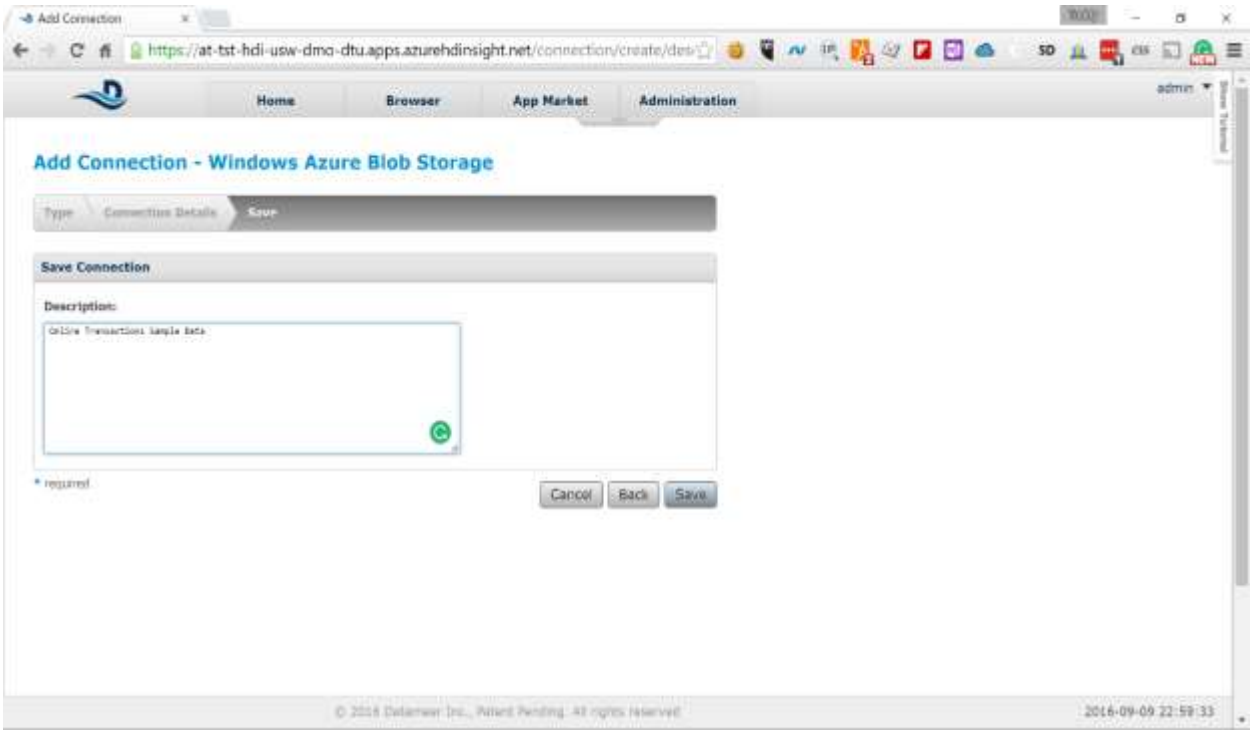
4. Click *Next* and fill in the following information on the next screen



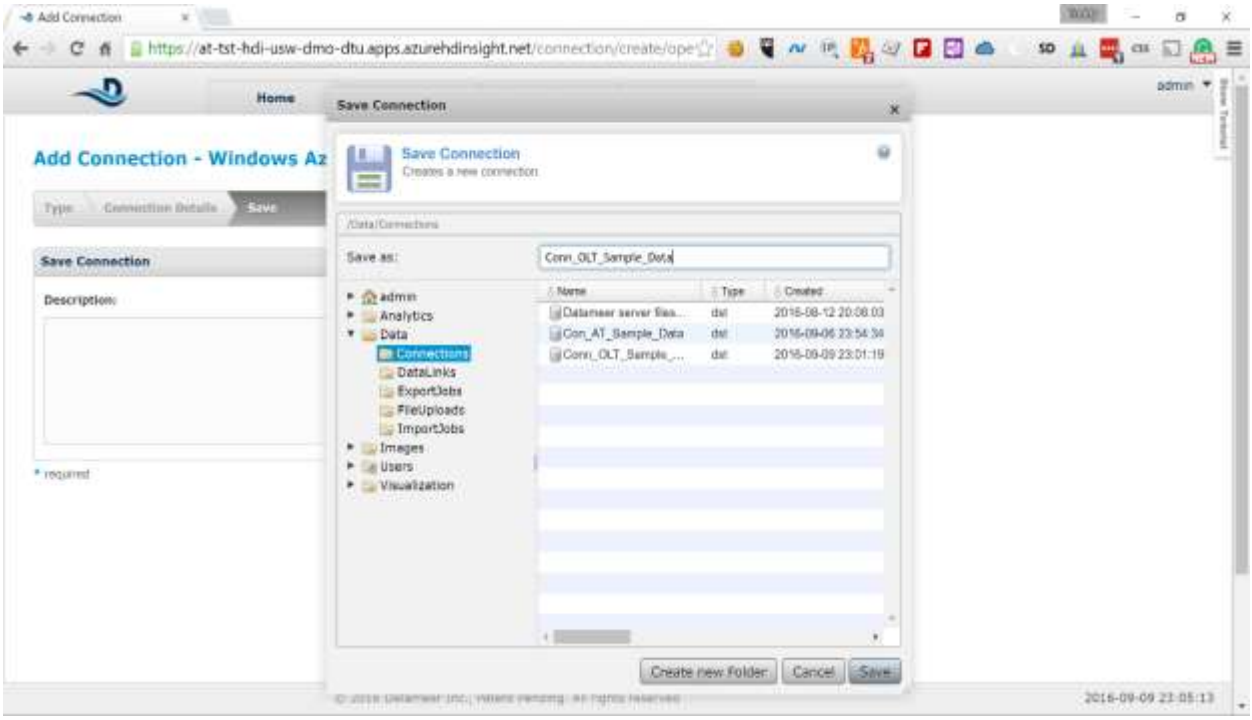
Storage Name: The name of the storage account where your data is
Container Name: The name of the container where your data is
Access Key: The key used to access the above storage account

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

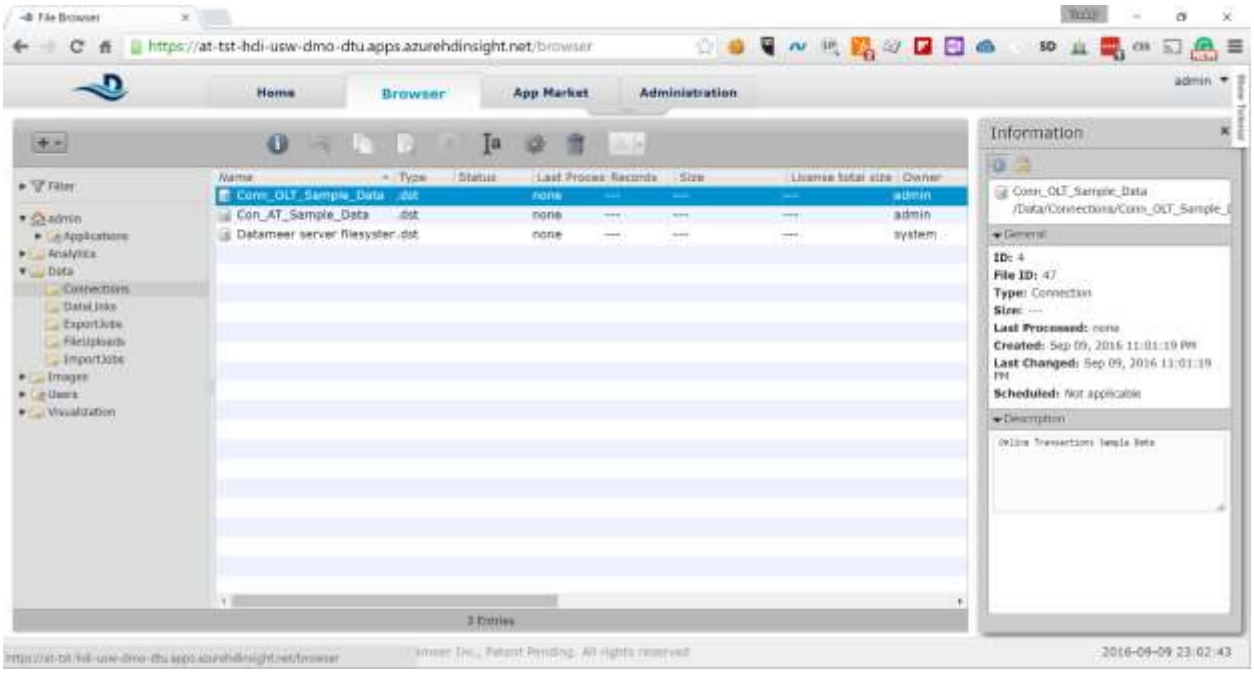
5. Click *Next* and on the next screen type the following description for the connection: "Online Transactions Sample Data"



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



7. 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.

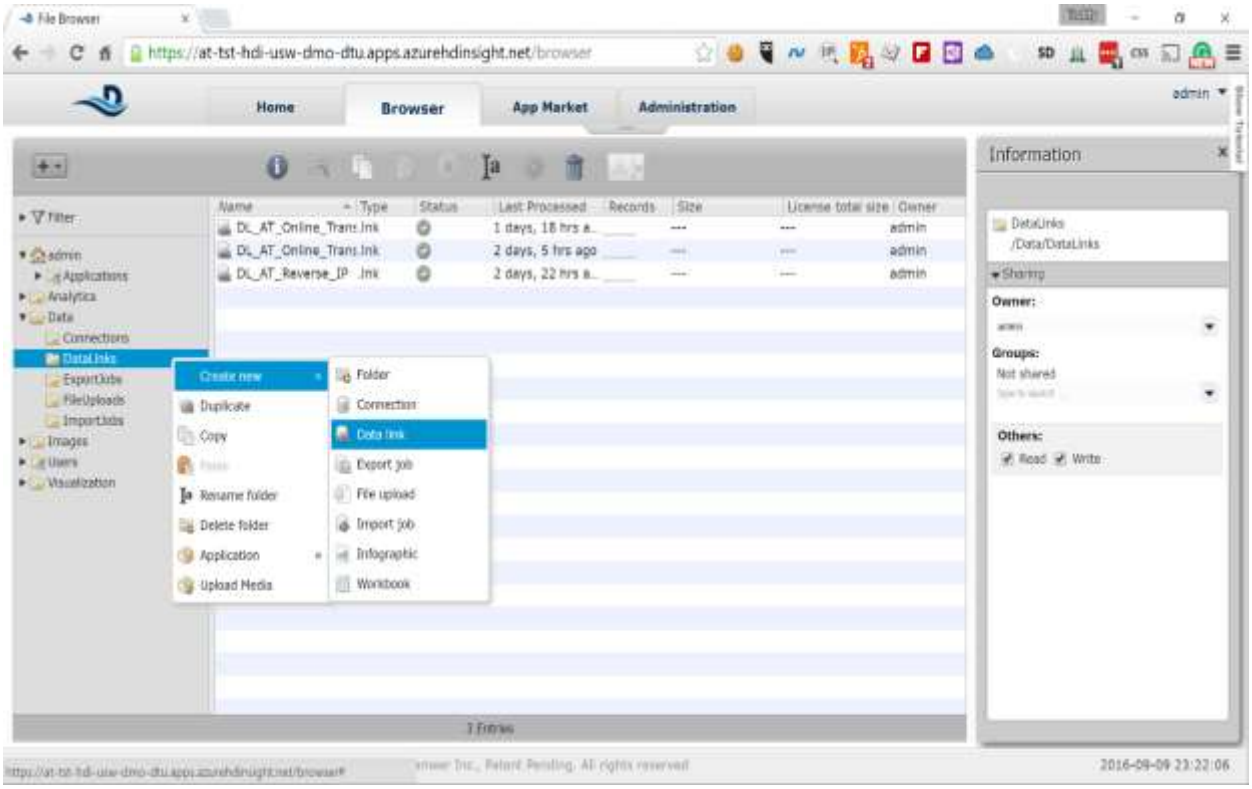
6 Link, Clean and Prepare the Data

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:

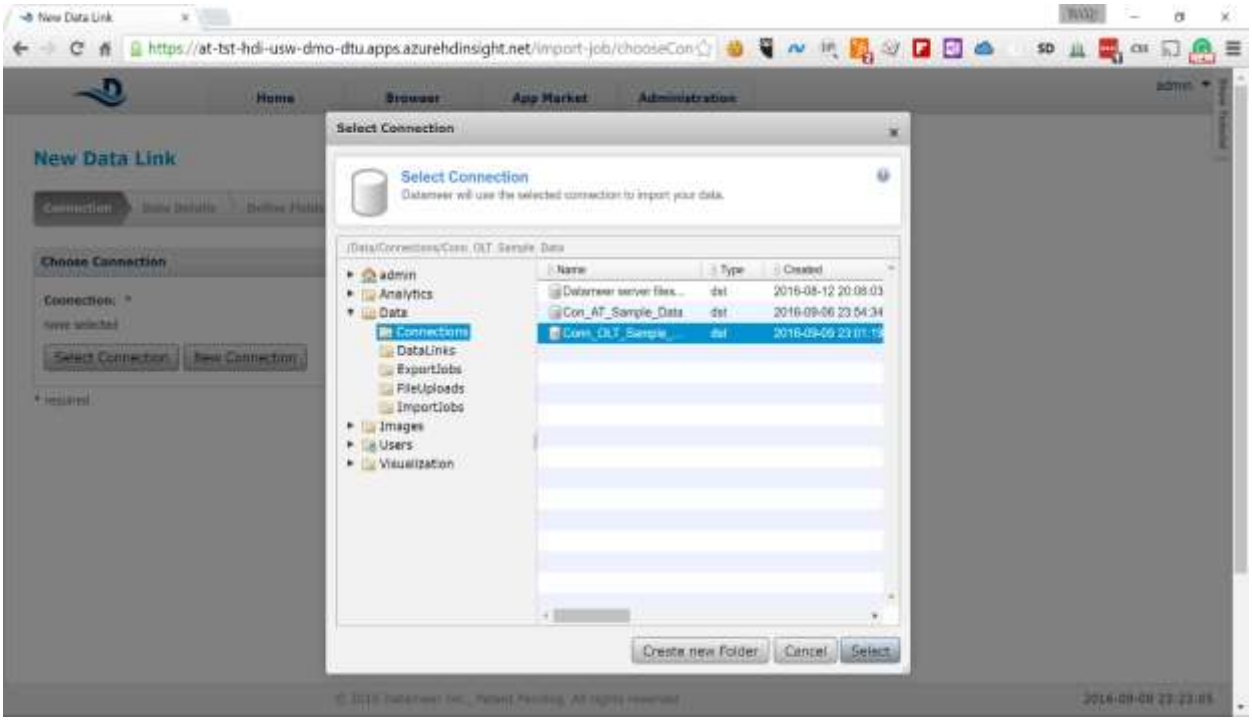
- The *timestamp* field is in ISO-8601 format, which needs to be converted into date/time field that Datameer can understand. We can do this conversion while we are linking the data.
- The *purchase_amount* field is a money field that is interpreted as a *STRING* by Datameer. We need to convert this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

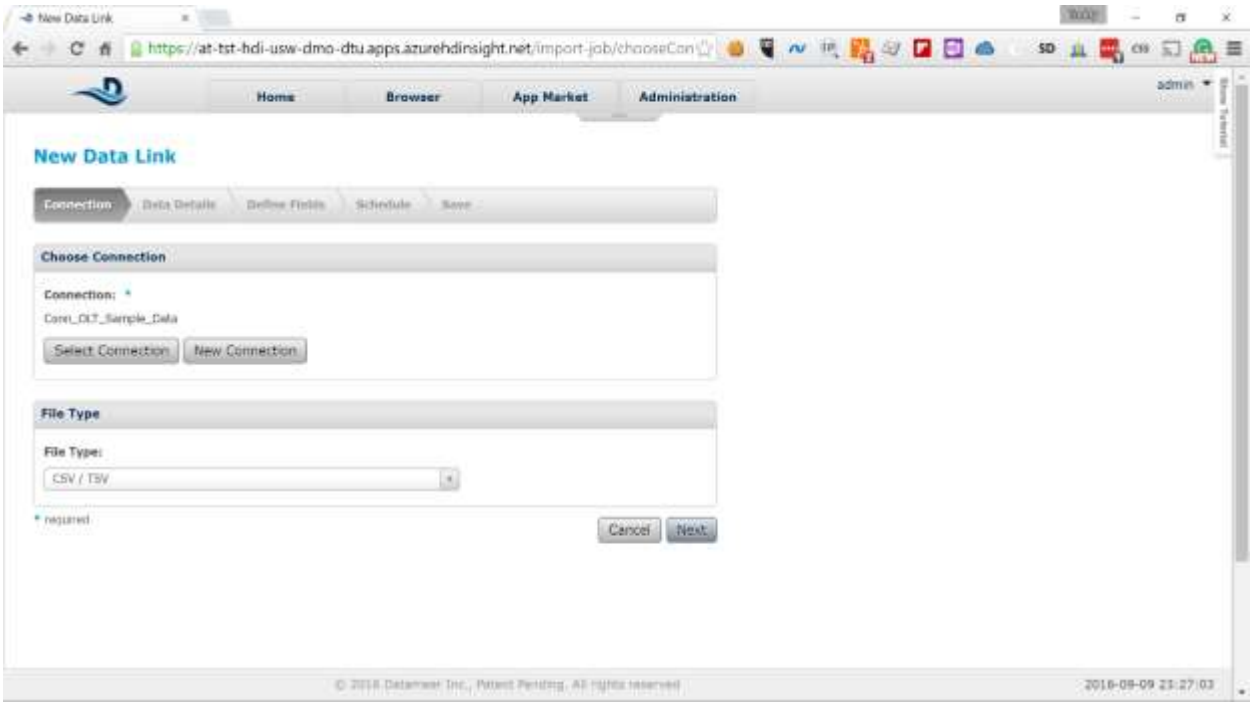
1. Right-click on the *DataLinks* node in the left-side navigation and select *Create new -> Data link*



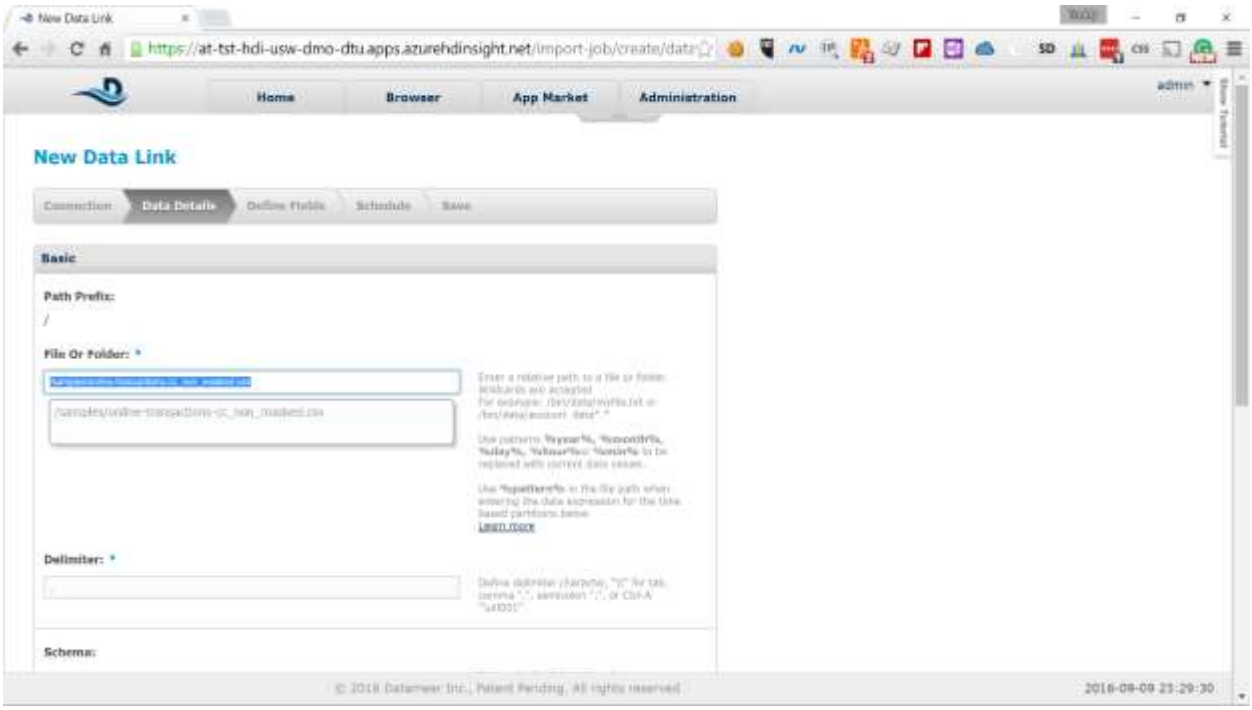
2. On the next screen click on the *Select Connection* button and select the *Conn_OLT_Sample_Data* connection you created previously



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

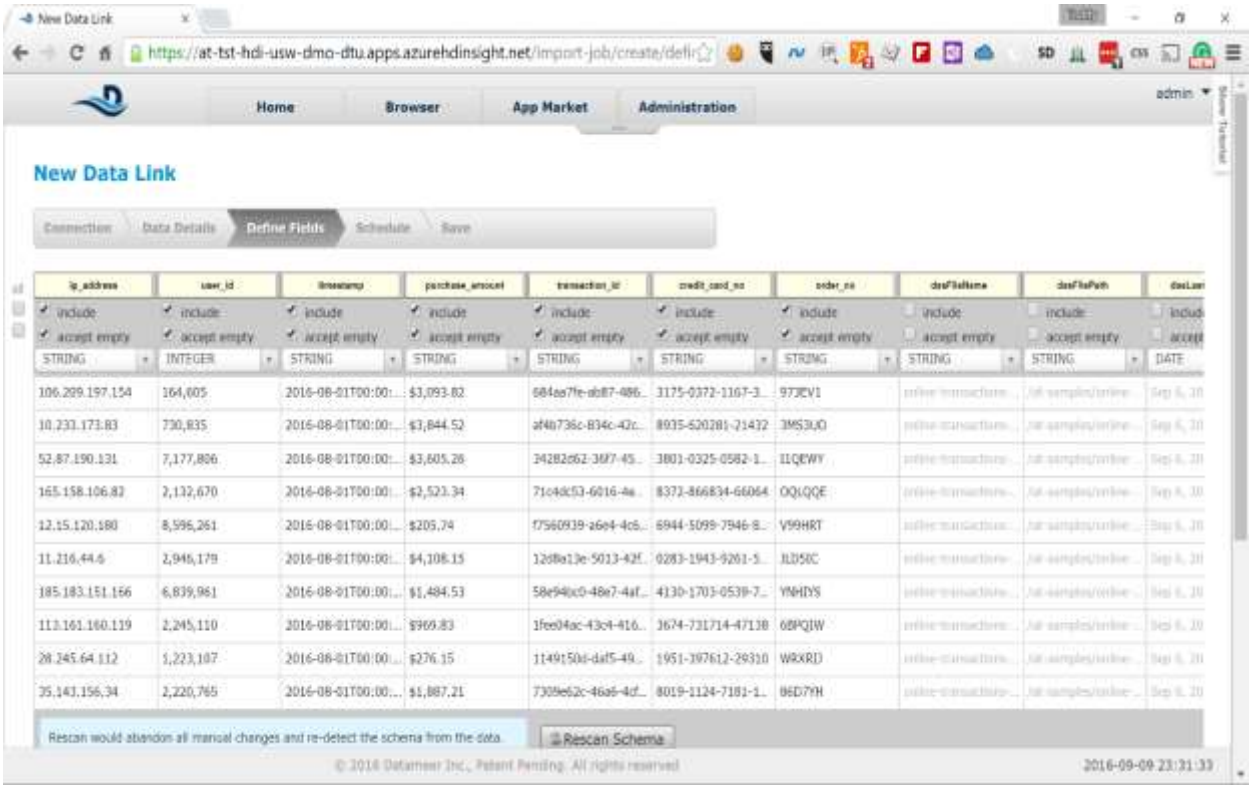


4. On the next screen type the following in the *File Or Folder* field:
/online-transactions-cc_masked.csv
(screenshot below shows ***"/samples/online-transactions-cc_masked.csv***".
Please use ***/online-transactions-cc_masked.csv*** instead)

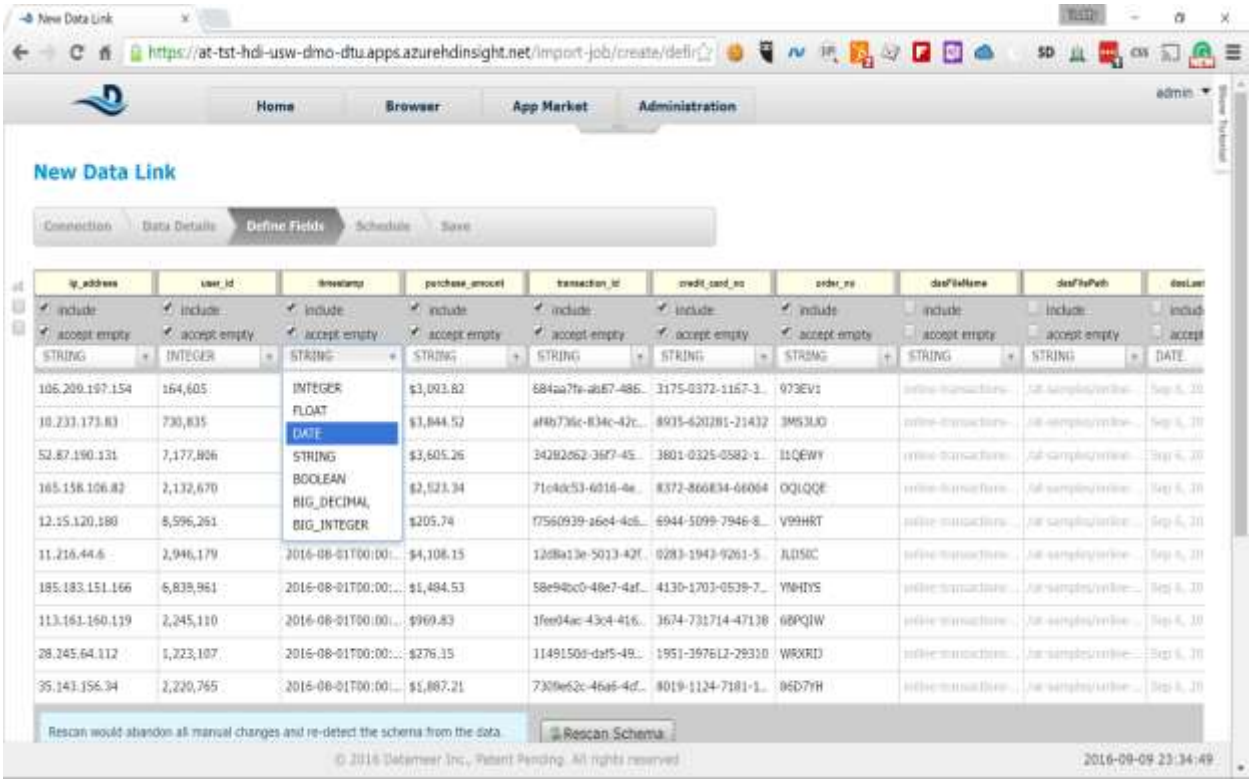


Scroll down to the bottom, keeping the default values for the rest of the fields, and click on *Next*

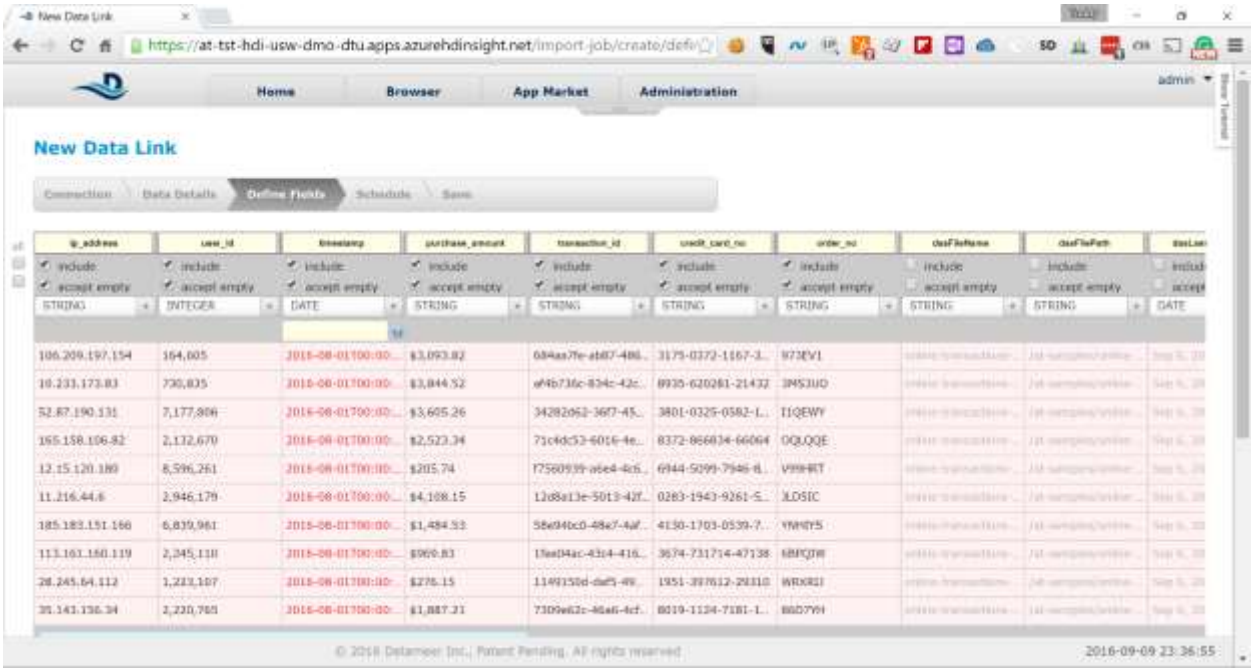
5. Datameer pre-fetches a representative sample of the data and shows it on the next screen



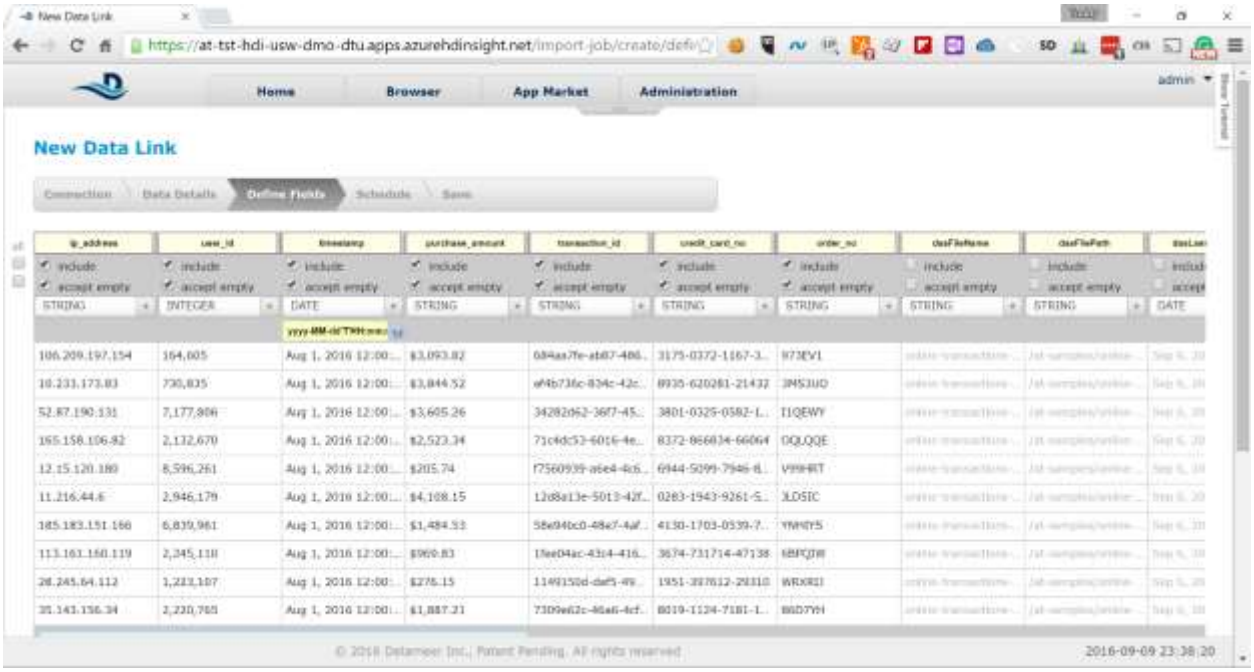
6. Click on the down-arrow for the field type under *timestamp* and change the type from *STRING* to *DATE*



7. 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

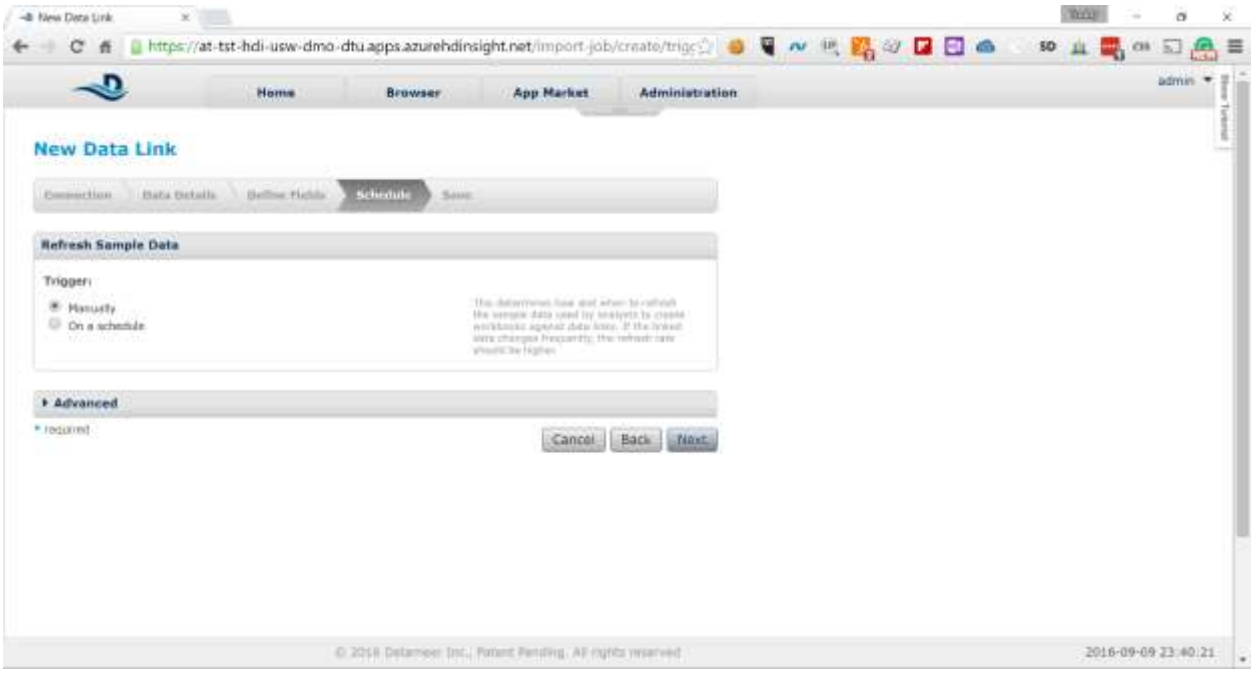


8. Type the following pattern in the field *yyyy-MM-dd'T'HH:mm:ss'Z'*

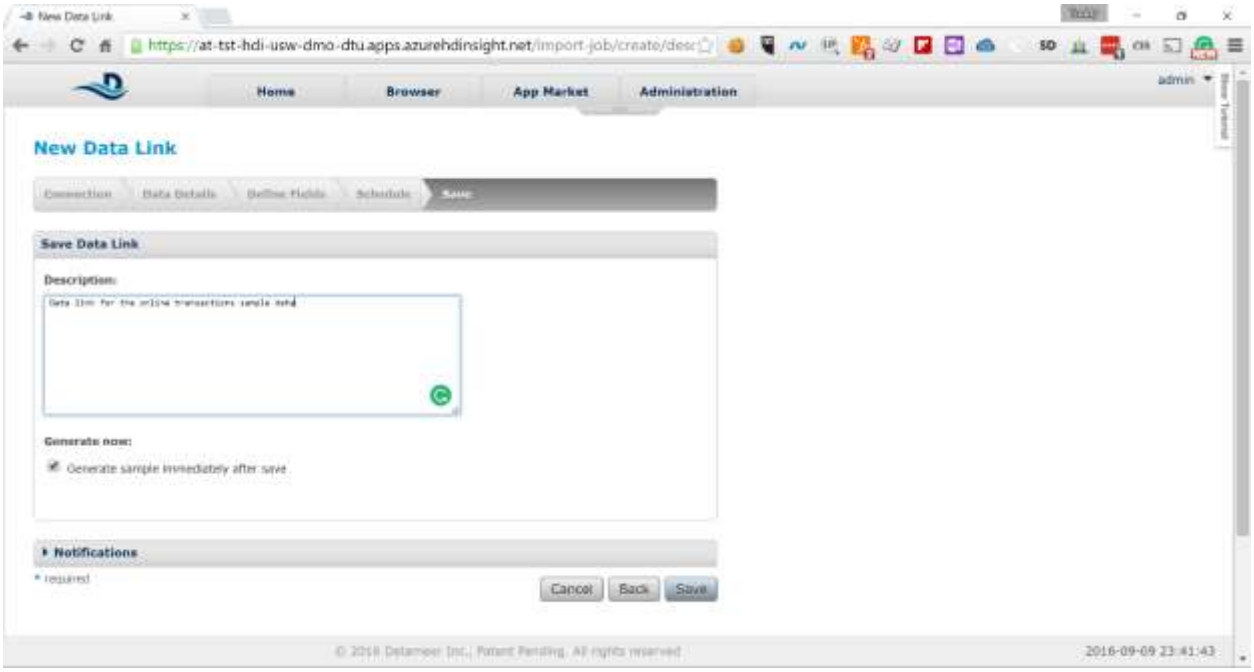


Datameer immediately parses the data in the field and shows it in the correct format. Scroll to the bottom of the screen and click on *Next*

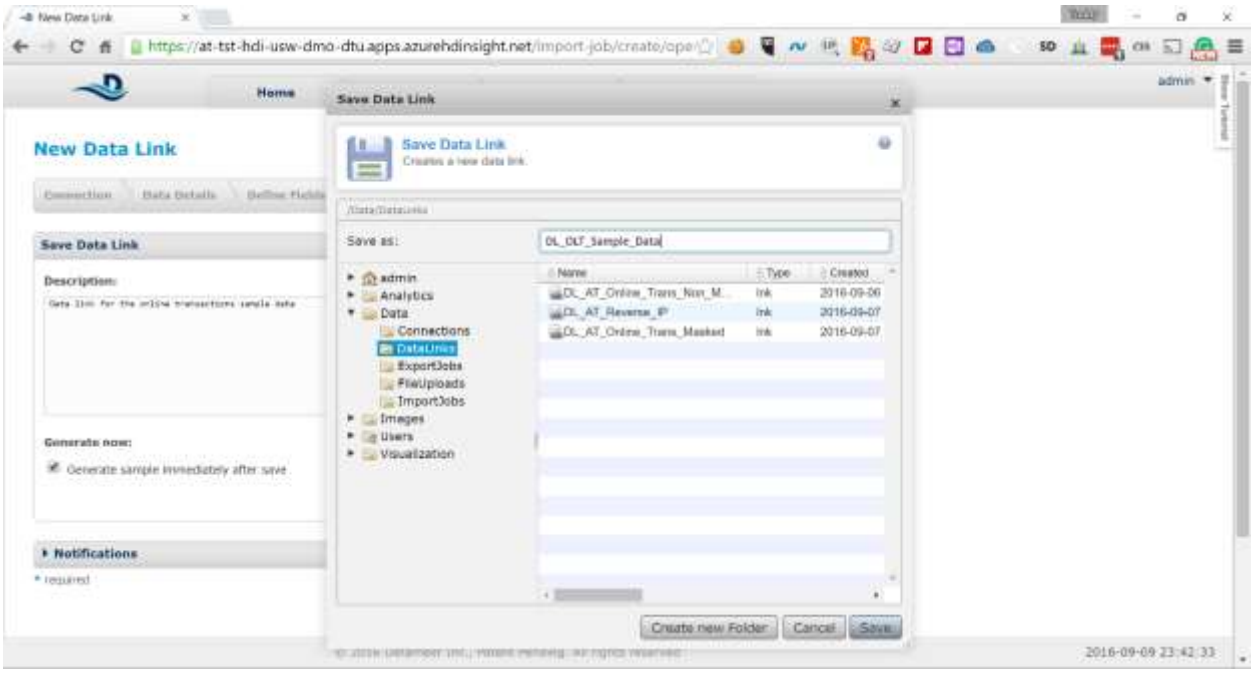
9. On the next screen keep the default value for *Trigger* and click on *Next*



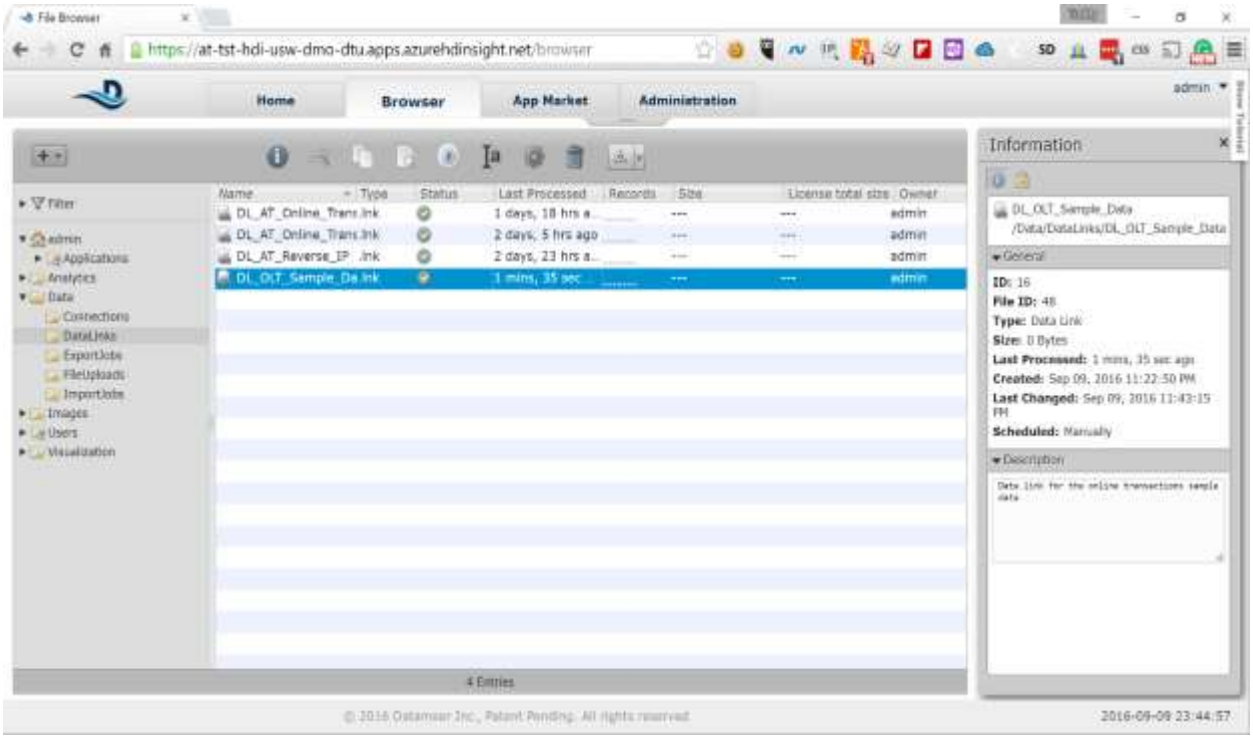
10. On the next screen type a meaningful description for the DataLink and click on *Save*



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

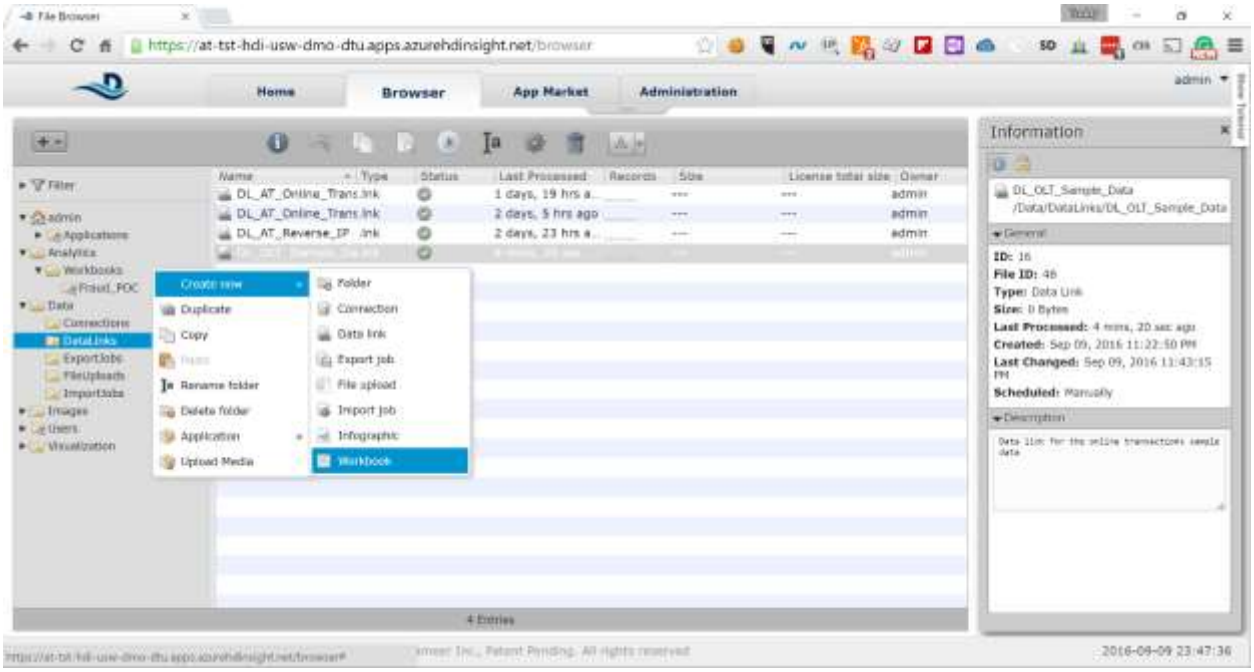


12. The new data link will appear in the list of data links on the next screen

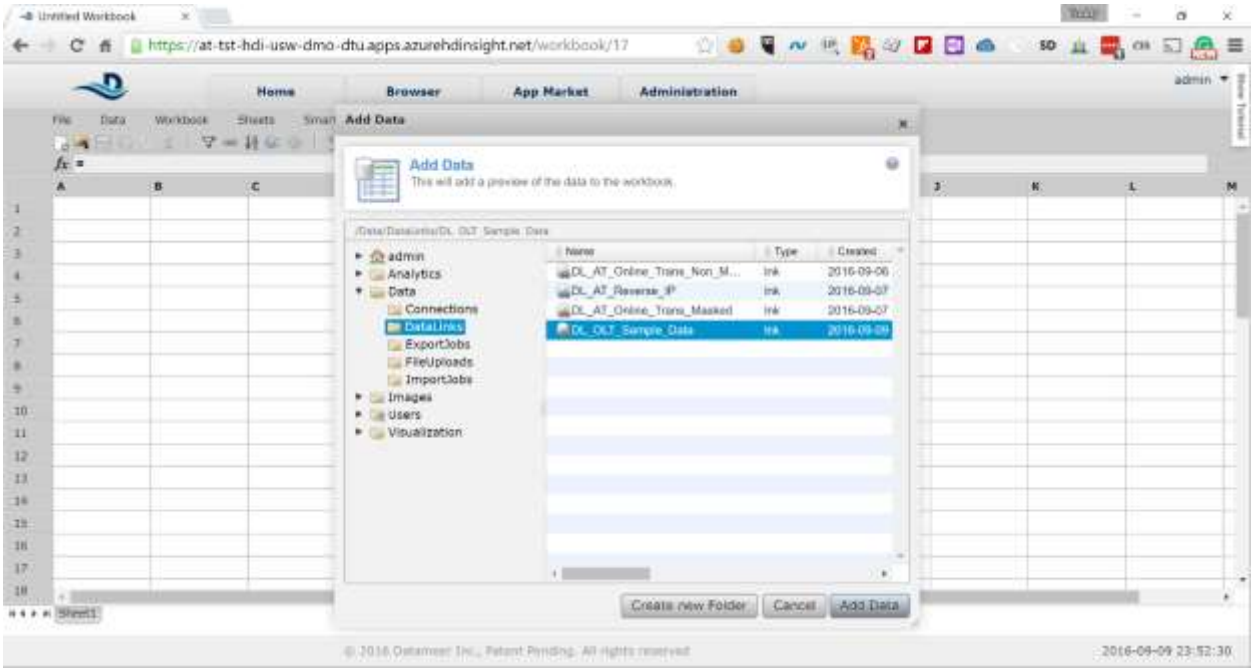


Now you have your dataset linked and have done some preliminary clean-up of the data. Next we will create a workbook where we will finish cleaning up our data and do our analysis. Here are the steps.

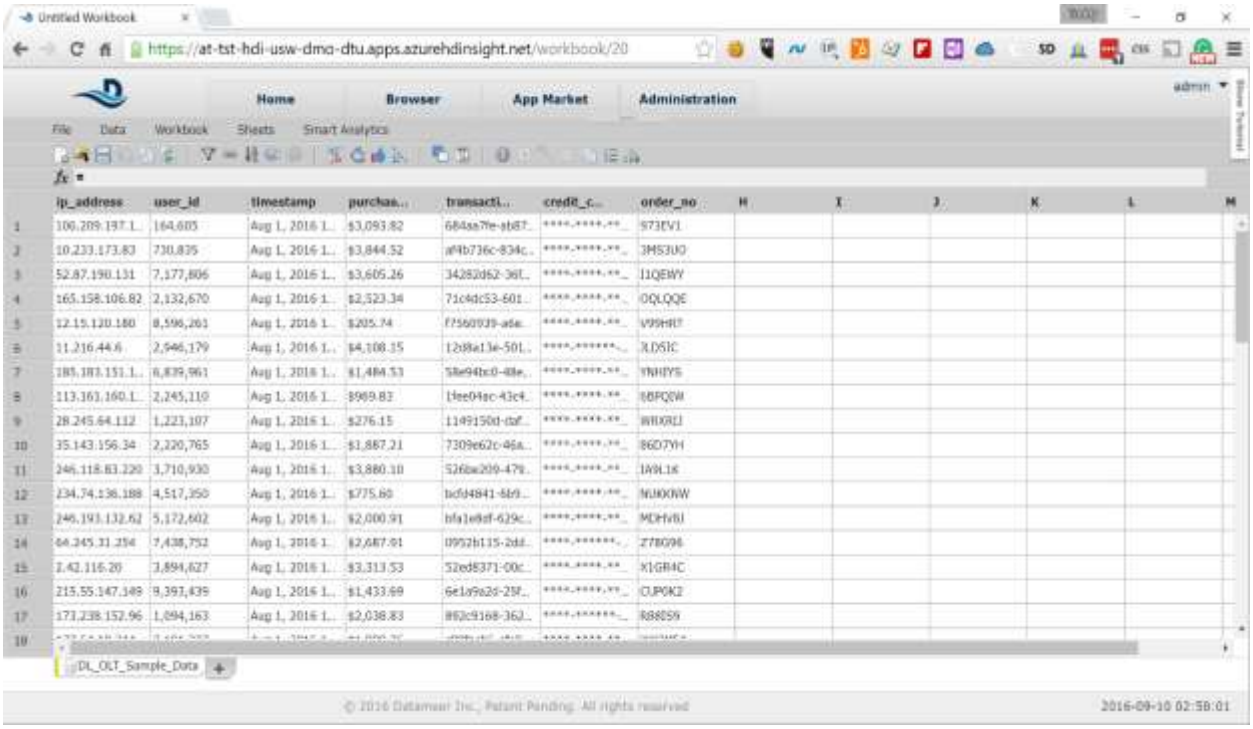
13. Expand the *Analytics* node in the left-side navigation, right-click on the *Workbook* node and select *Create new -> Workbook*



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.

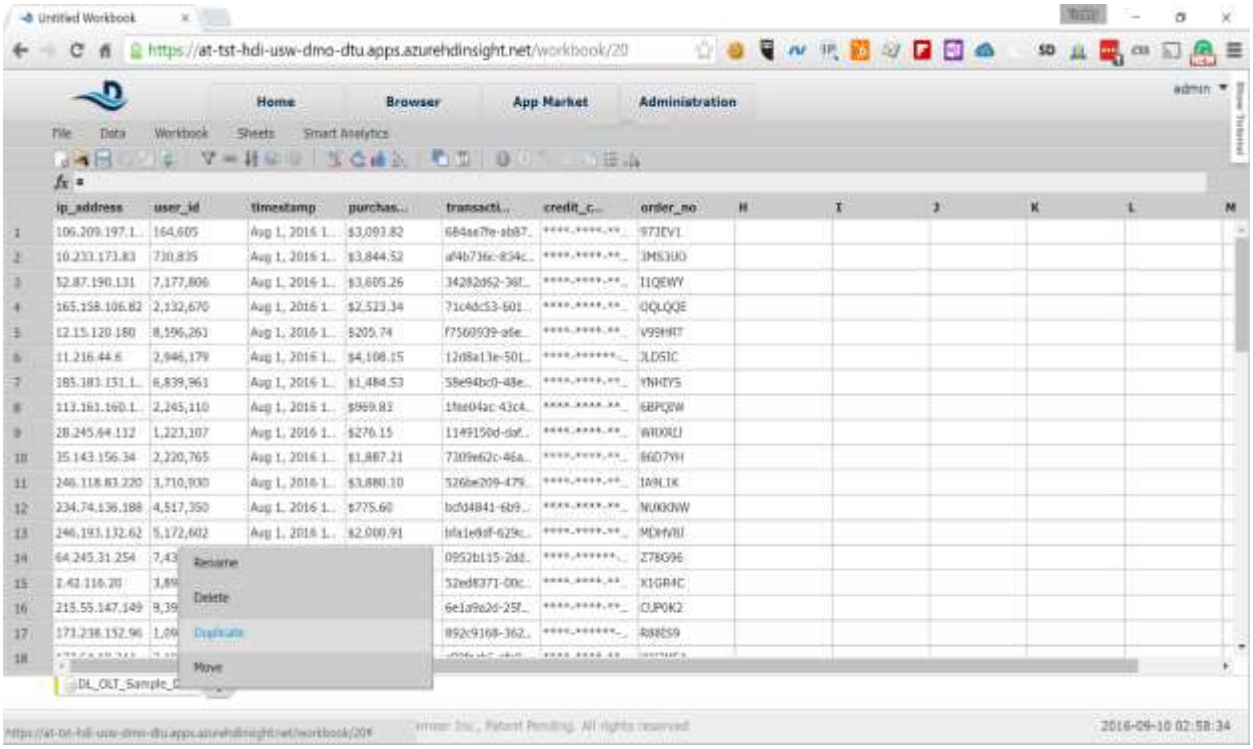


15. Click on *Add Data* to load a sample of the dataset in the workbook sheet

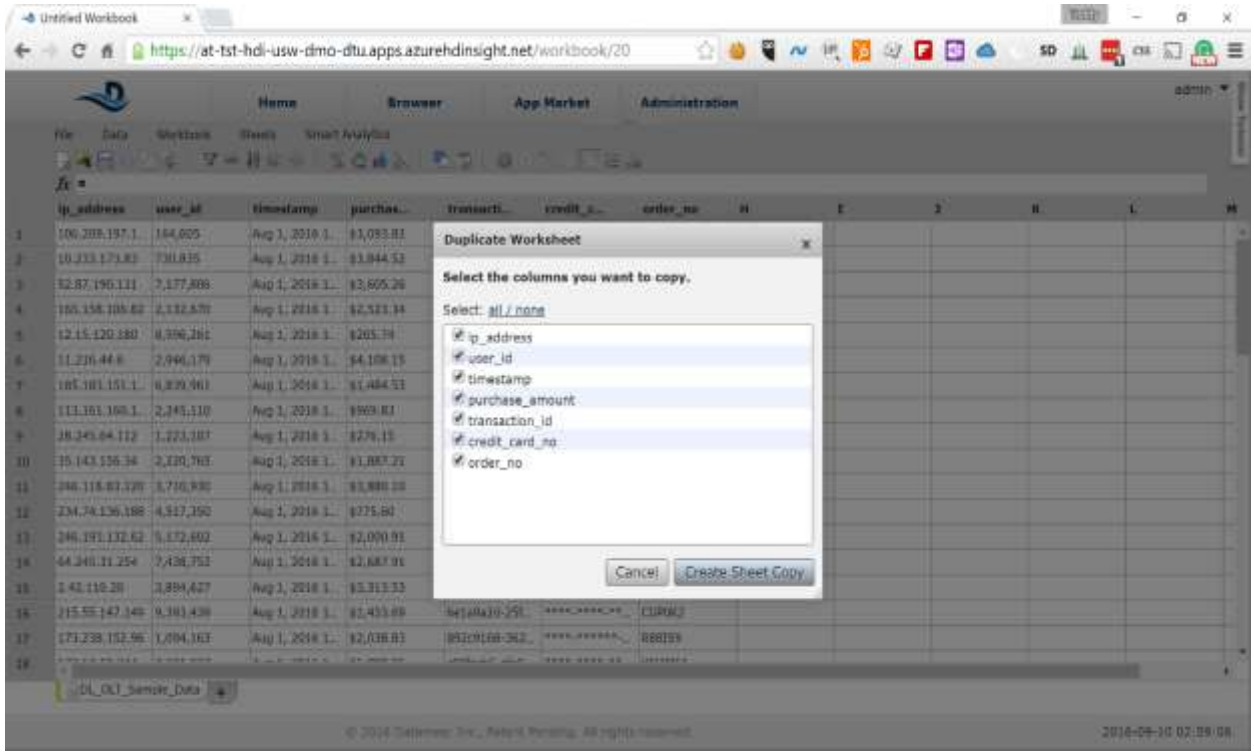


The UI you are presented with is very similar to Excel and uses the same concepts.

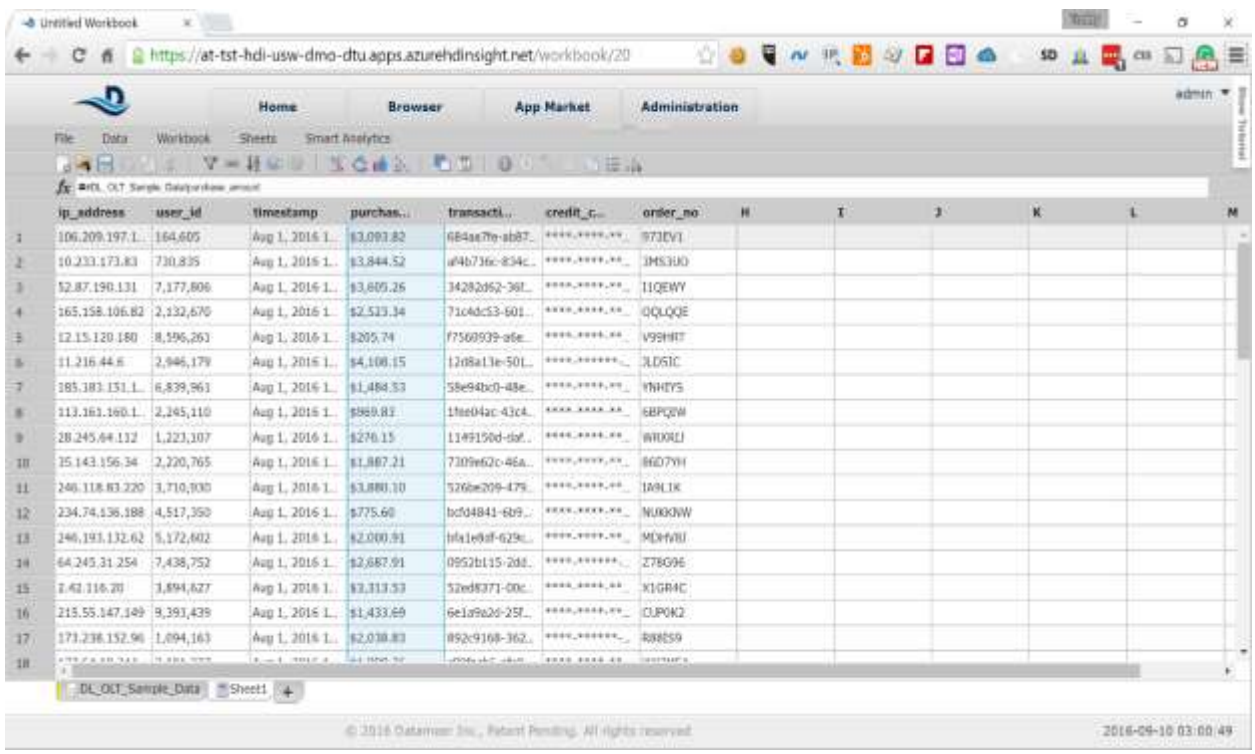
16. Right-click on the *DL_OLT_Sample_Data* sheet at the bottom of the screen next and select *Duplicate*



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



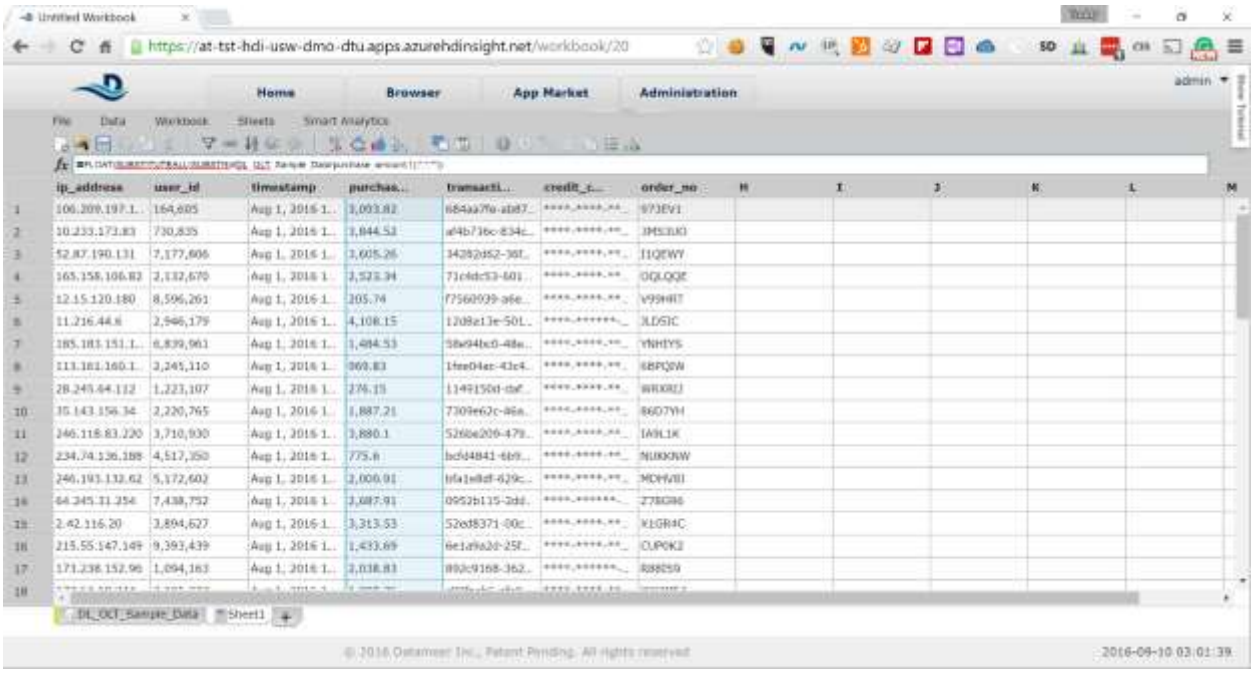
18. A new copy of the sheet is created that contains all of the data from the original sheet. Click on the *purchase_amount* column to enable the f_x field for that column available on top of the sheet



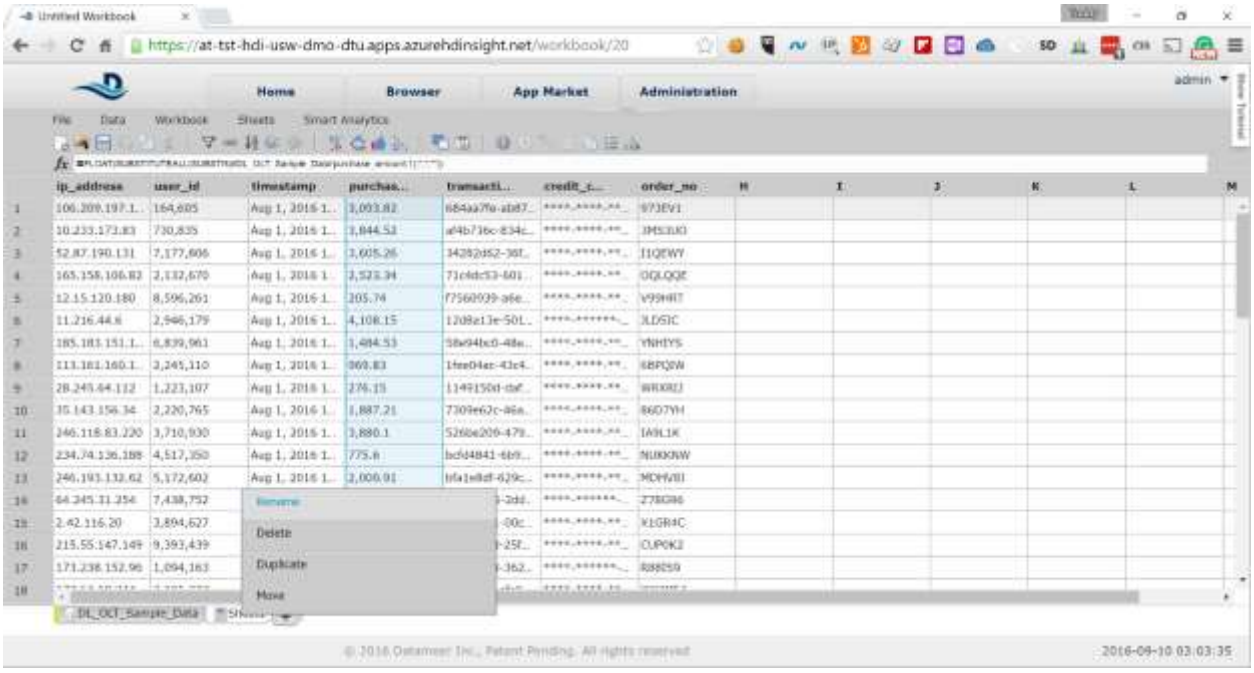
19. Type the following in the f_x field and press *Enter*

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

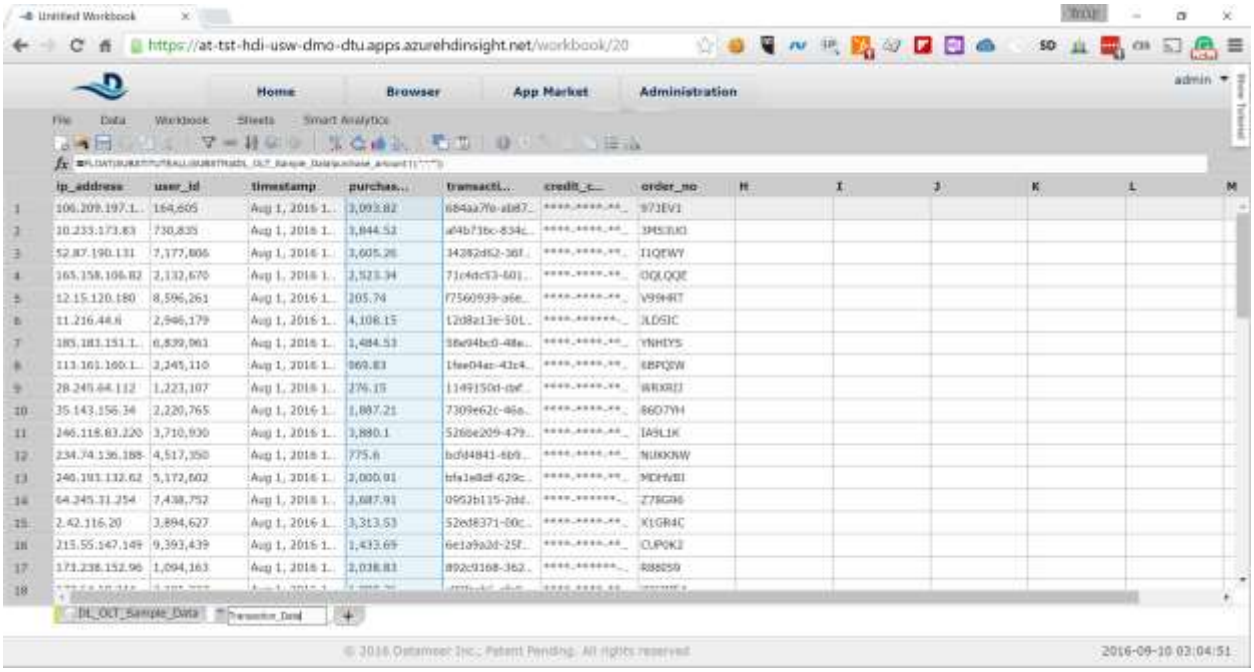
    The formula strips the $ (dollar) sign in front of the amount, removes all commas and converts the string to FLOAT. Now you can use numeric functions to perform calculations on the field.



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



21. Rename *Sheet1* to *Transaction_Data*

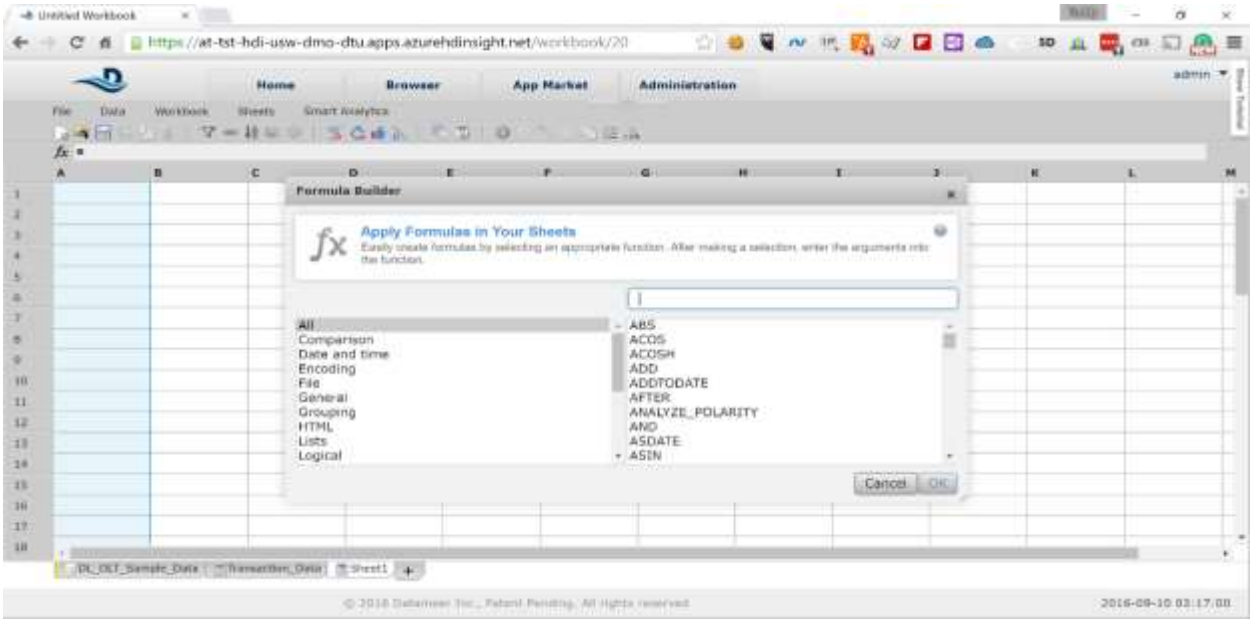


With this we are done with the clean-up of our data and are ready to perform our analysis.

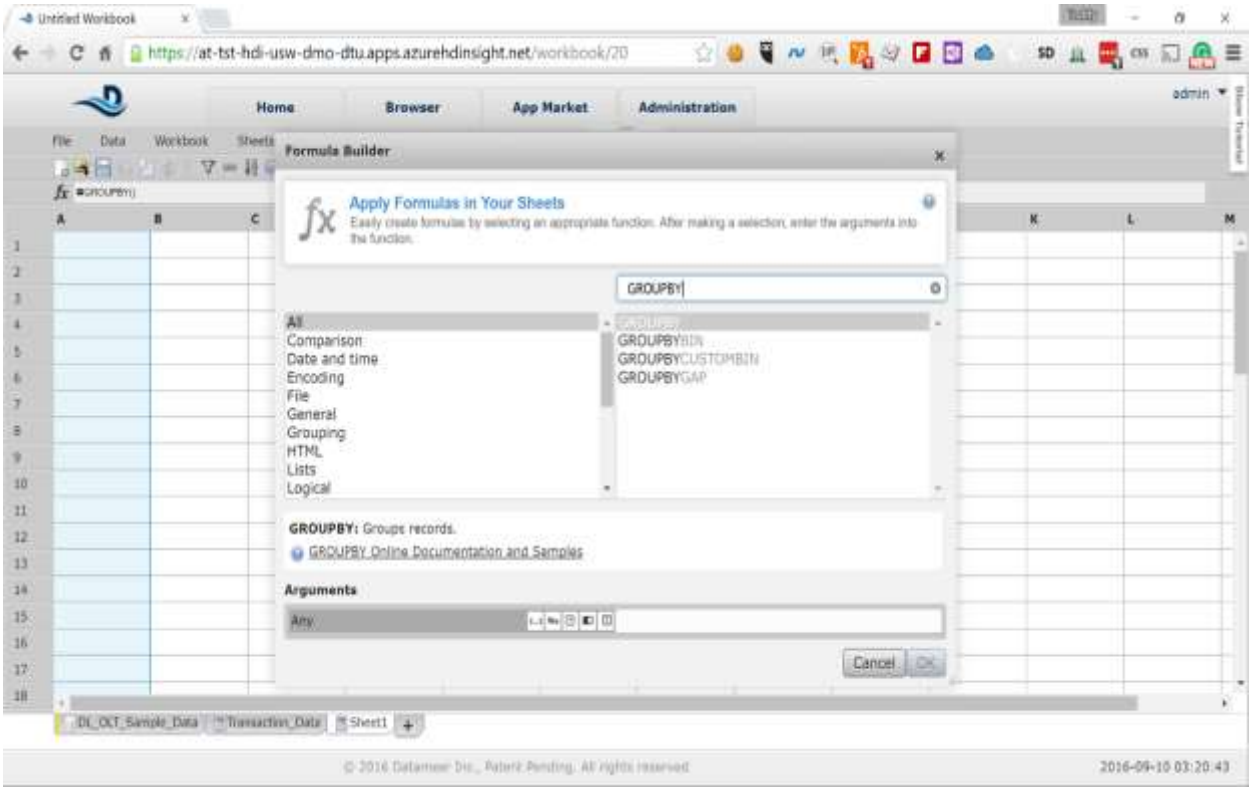
7 Perform Analysis to Identify Outliers

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:

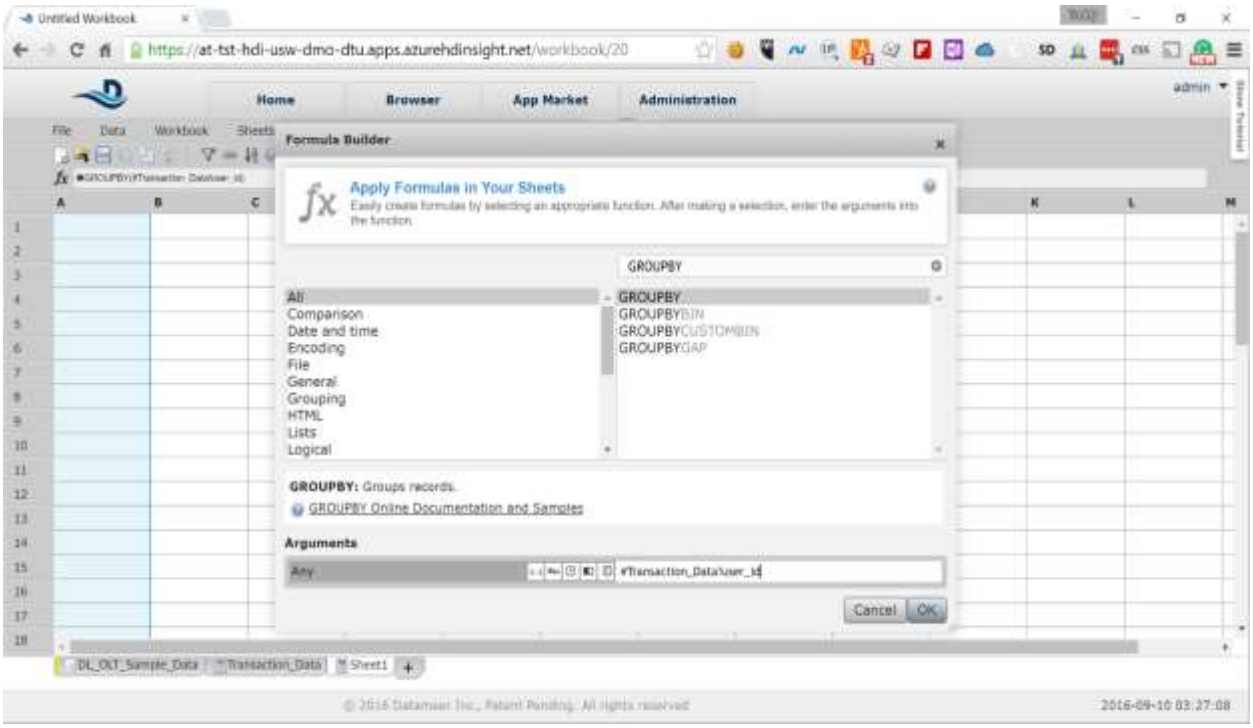
1. Click on the + sign next to the *Transaction_Data* sheet to create an empty sheet for analysis



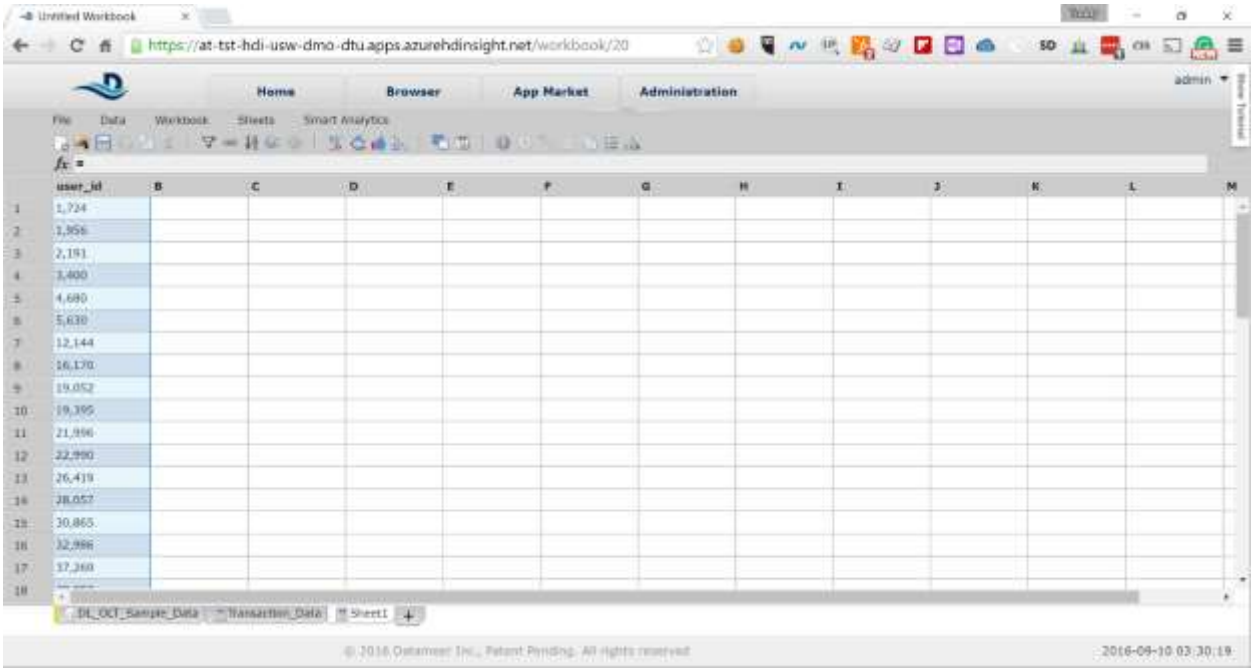
2. In the input field in the pop-up type *GROUPBY* to filter the functions and select *GROUPBY* from the list



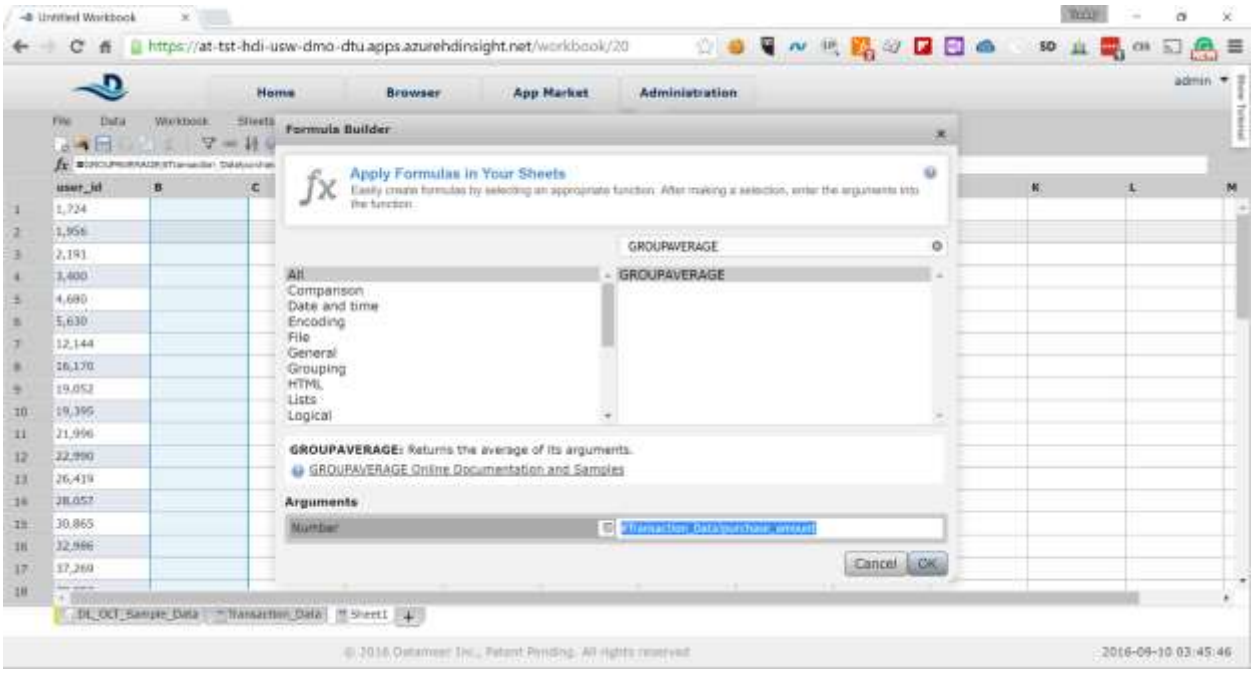
3. In the *Arguments* input field at the bottom of the pop-up type *#Transaction_Data!user_id* to group the data by user identifier and click on the *OK* button



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

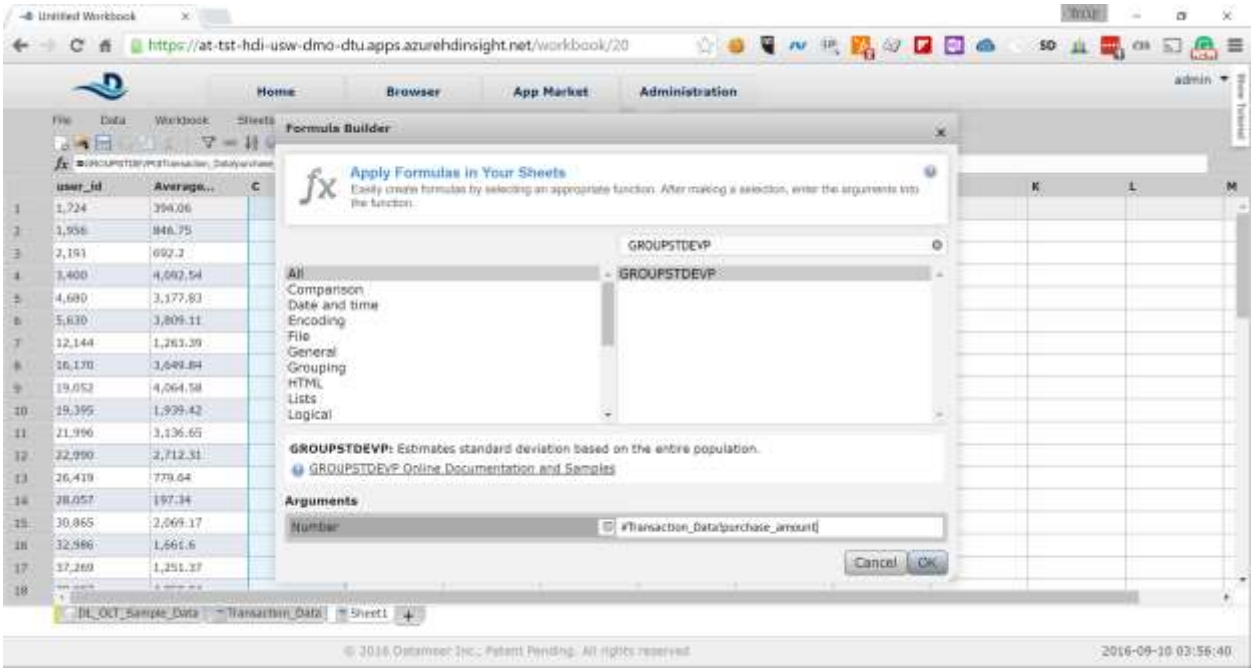


5. Click on the second column to show the functions pop-up again and type GROUPAVERAGE in the filter box and select the GROUPAVERAGE function. In the *Arguments* field type *#Transaction_Data!purchase_amount*



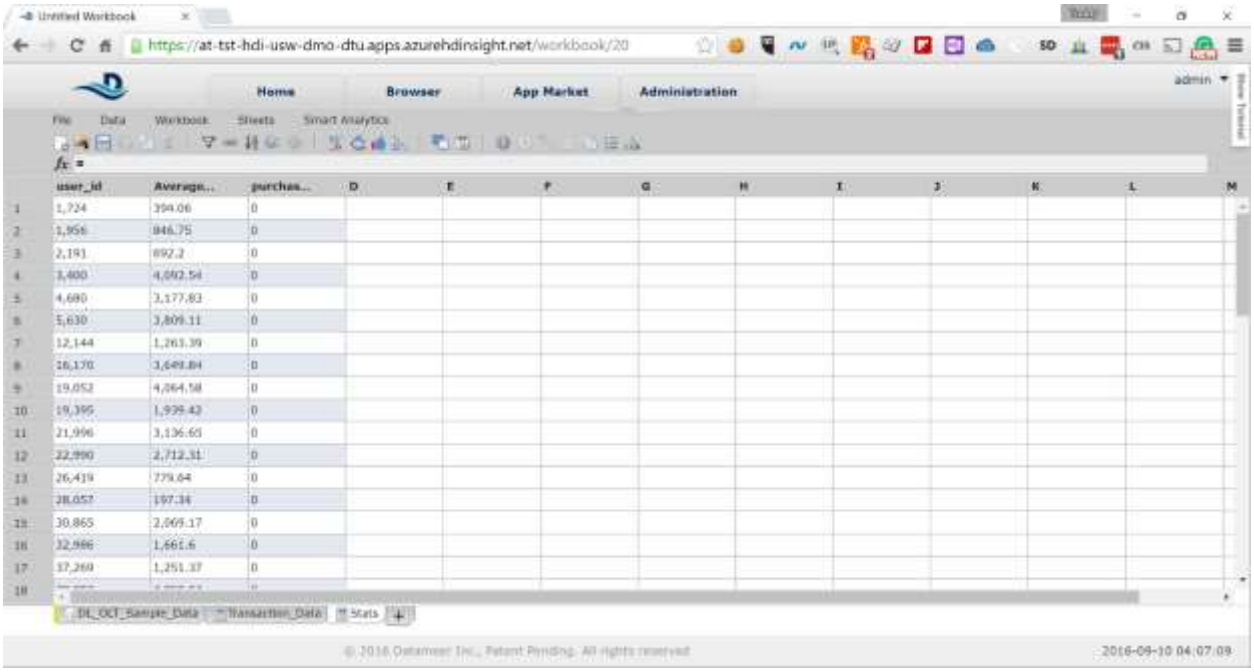
This will calculate the average purchase amount for each of the users.

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

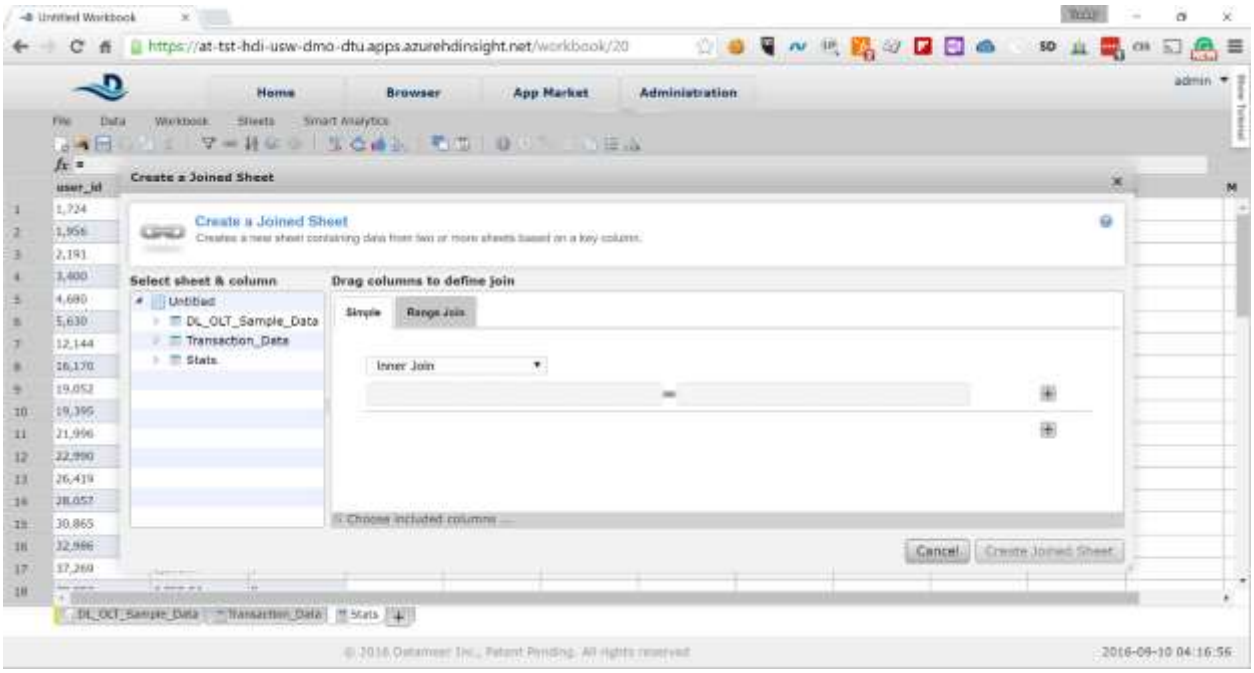


In the Arguments field type #Transaction_Data!purchase_amount to calculate the standard deviation for the *purchase_amount* field

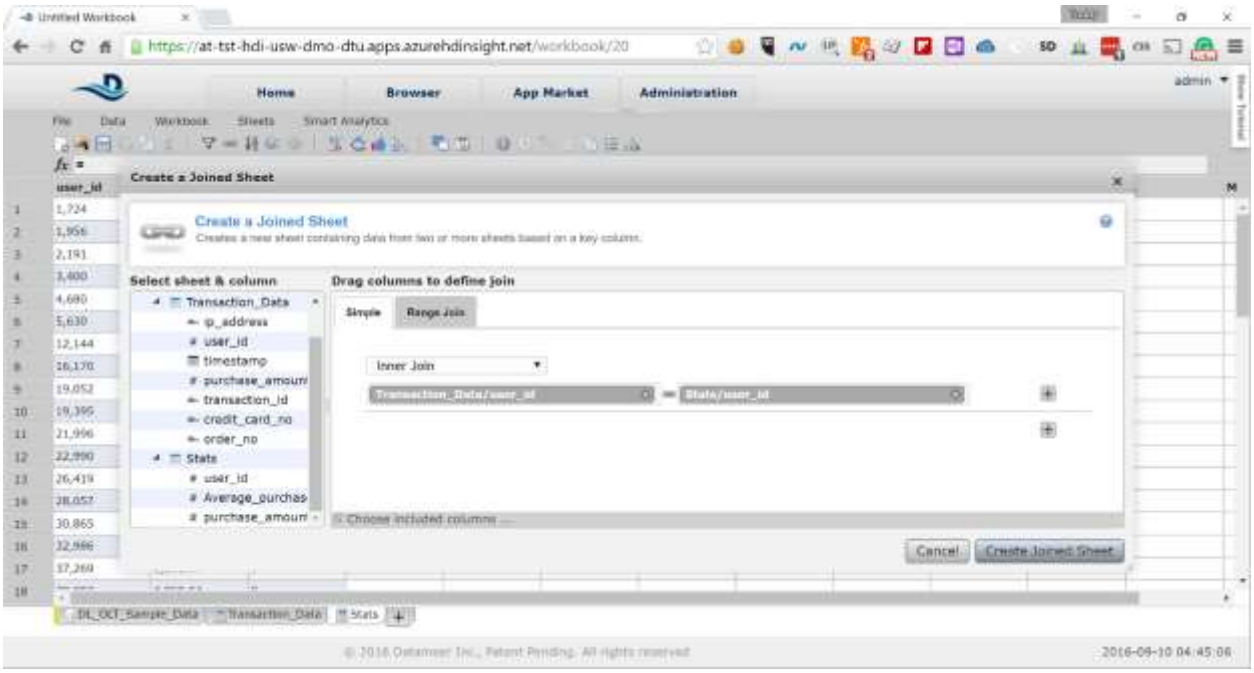
7. Right-click on the sheet name and select *Rename* from the context menu to rename the sheet. Choose the following name for the sheet:
   Stats



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

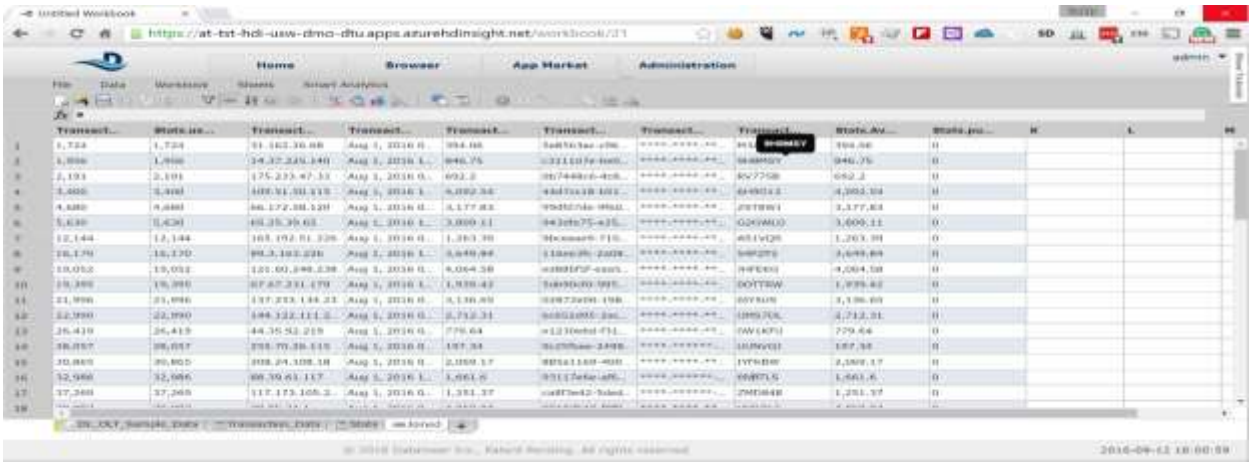


9. Expand the *Transaction_Data* node in the pop-up navigation tree and drag the *user_id* field to the right. Do the same with the *user_id* field from the *Stats* node.



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

10. 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:

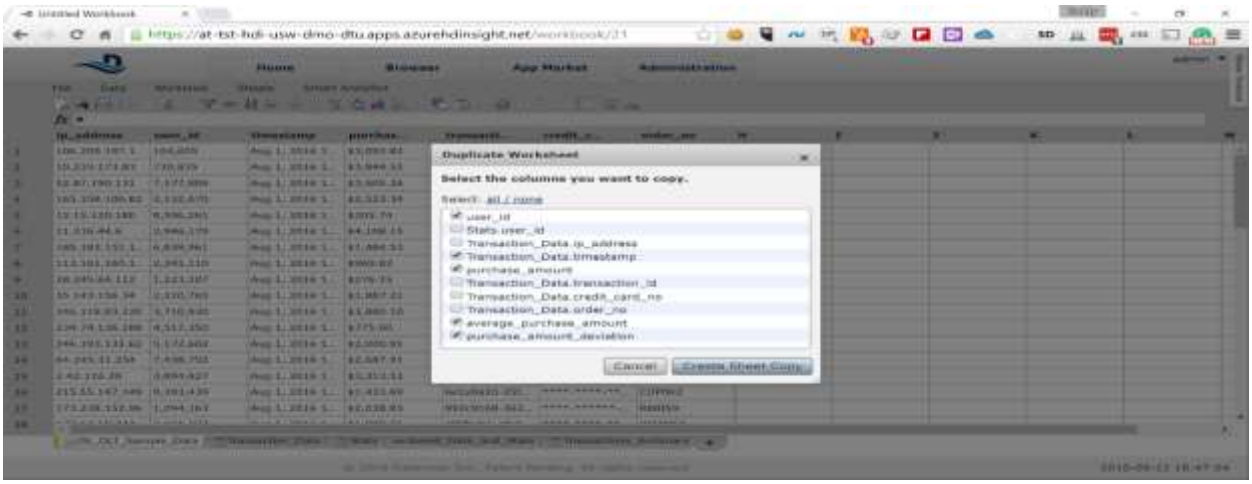


For convenience let's rename few of the columns. Right-click and rename the columns as below:
Transaction_Data.user_id -> user_id
Transaction_Data.purchase_amount -> purchase_amount
Stats.Average_purchase_amount -> average_purchase_amount
Stats.purchase_amount_Stddevp -> purchase_amount_deviation

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

11. 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

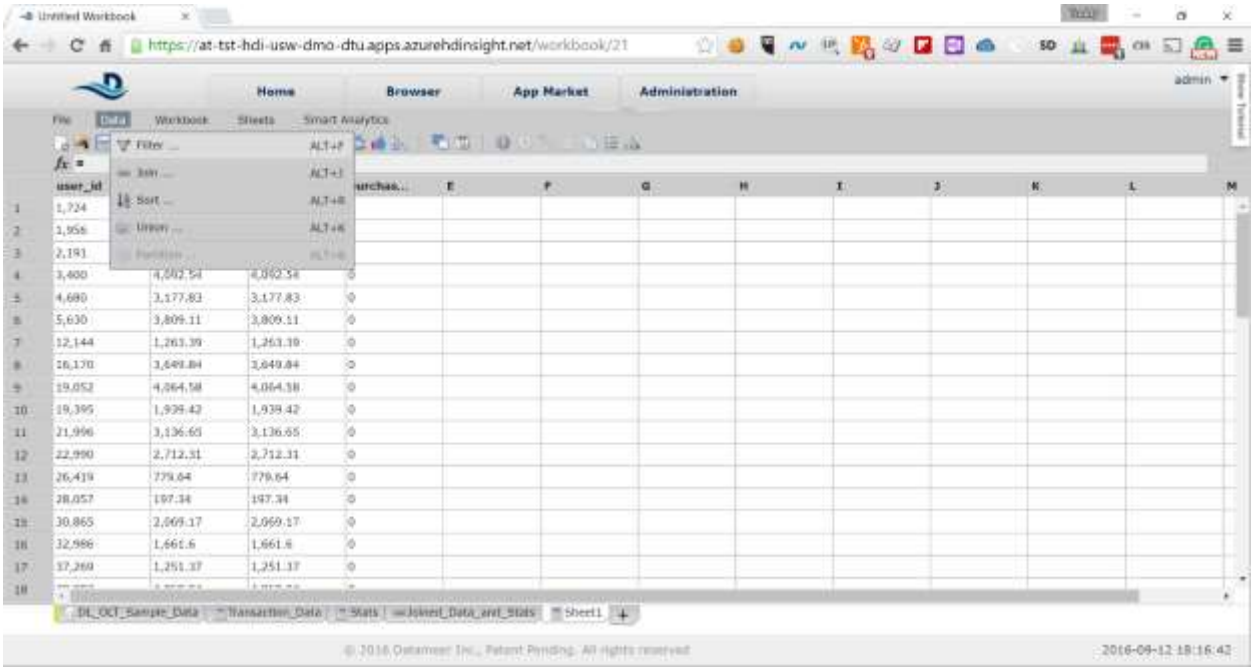
rest. In the pop-up select only the following fields:
user_id
purchase_amount
Transaction_data.timestamp
average_purchase_amount
purchase_amount_deviation



Click on *Create Sheet Copy* button

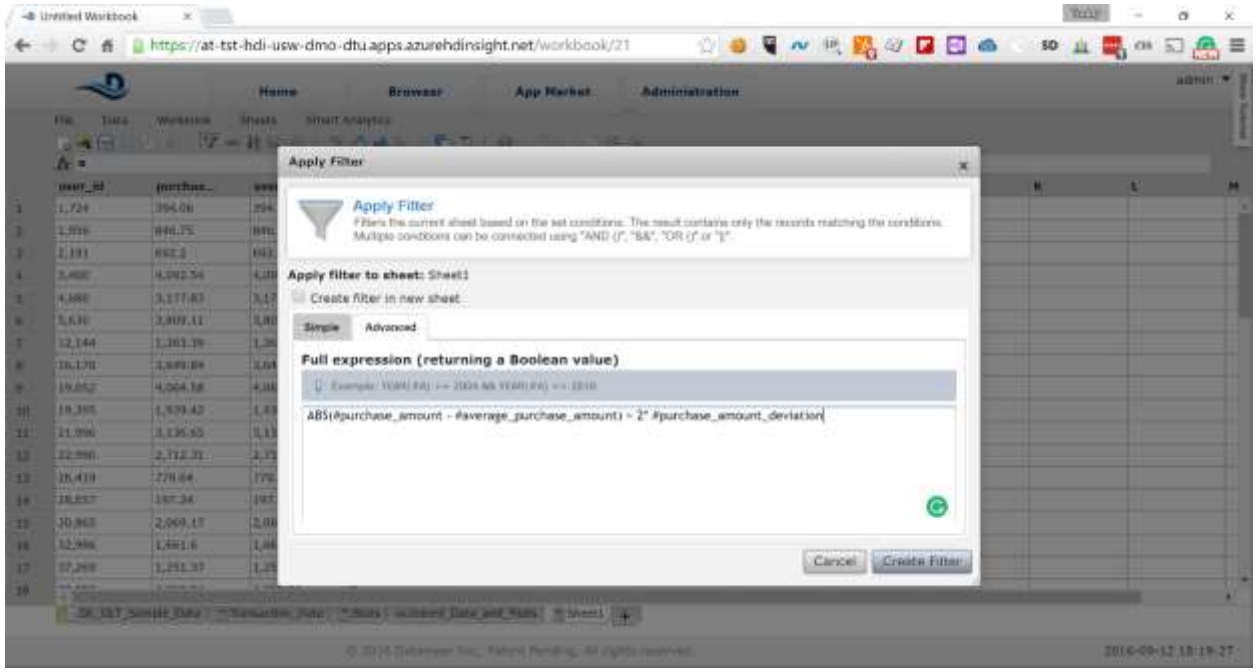
12. Right-click on the *Transaction_Data.timestamp* field and rename it to *timestamp* only.

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.

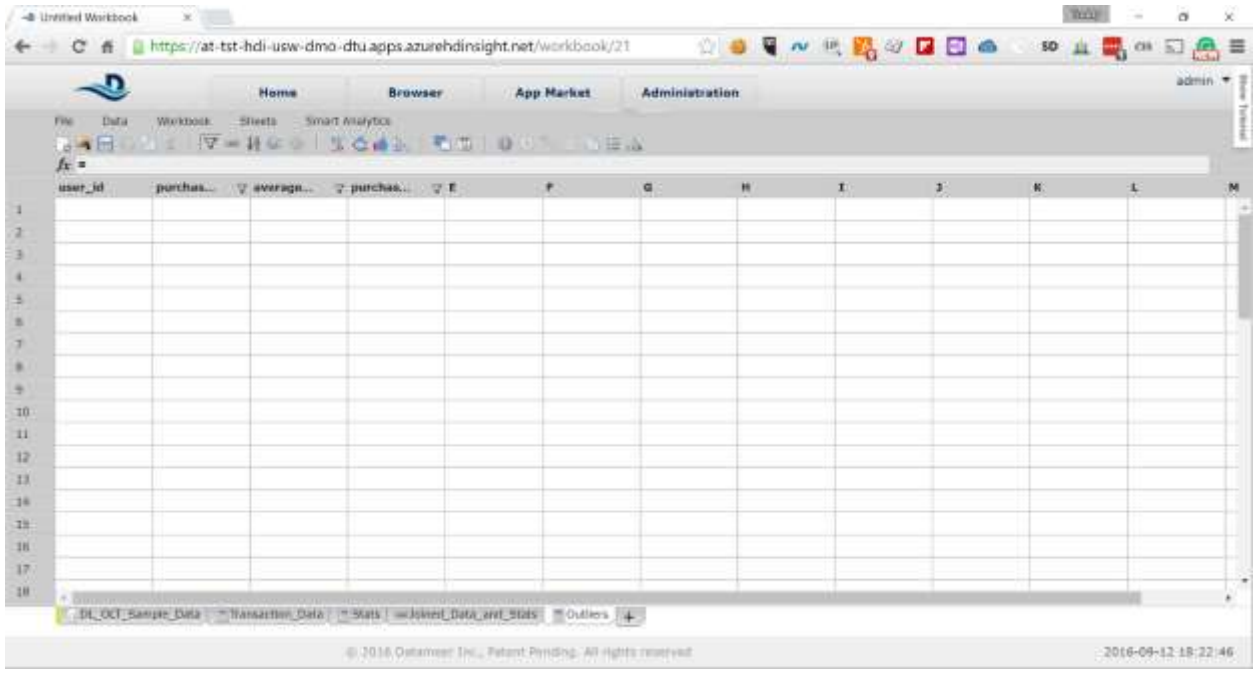


14. Select the *Advanced* tab in the pop-up and type the following formula:

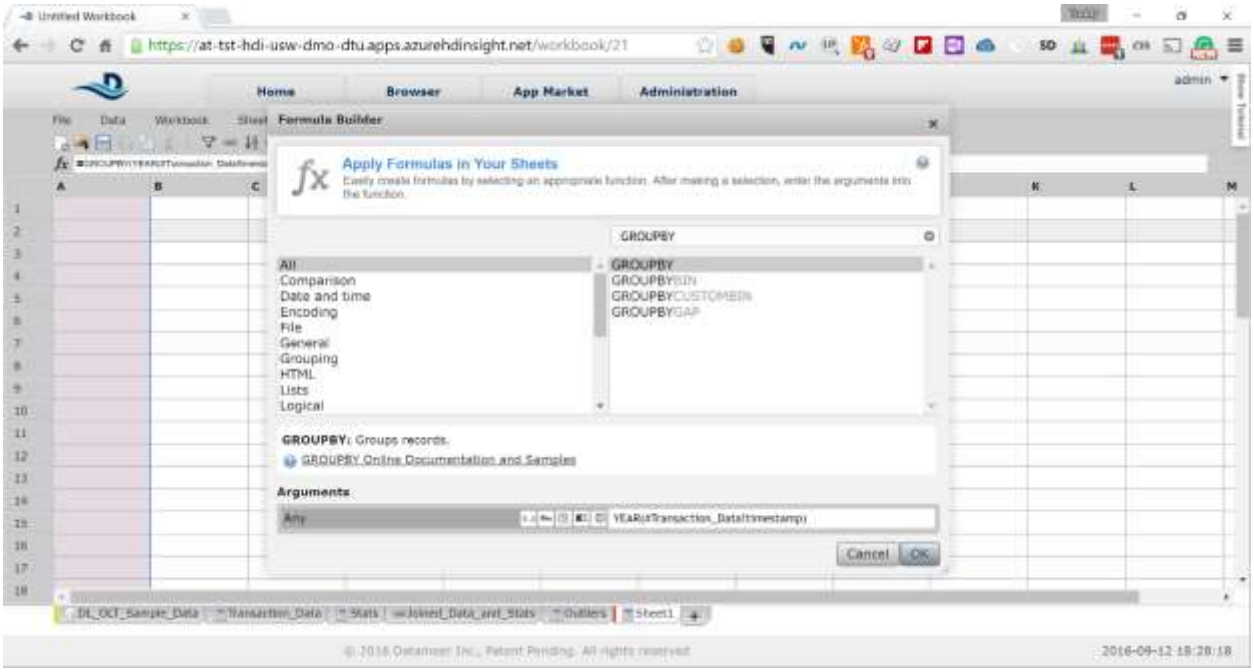
    ABS(#purchase_amount - #average_purchase_amount) > 2 #purchase_amount_deviation*



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*



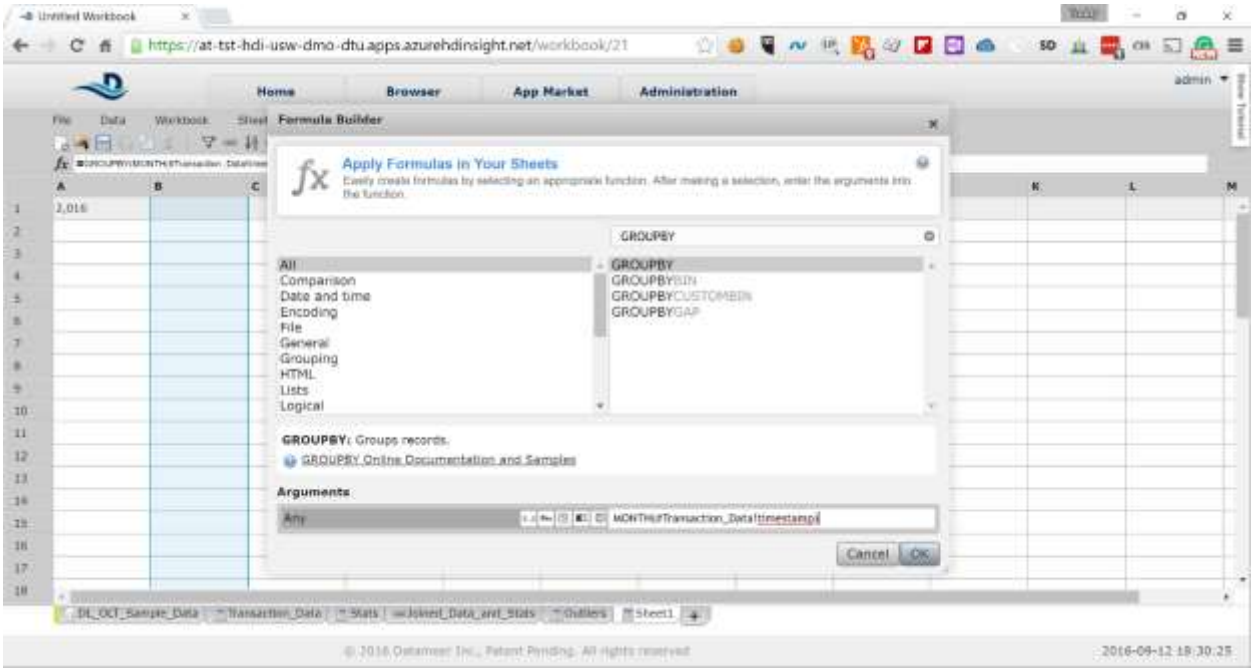
16. Finally, we would like to create a summary of the data that we would like to visualize. Let's start with summary of the *Transaction_Data*. Create new sheet and in the formula pop-up select the *GROUPBY* function



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

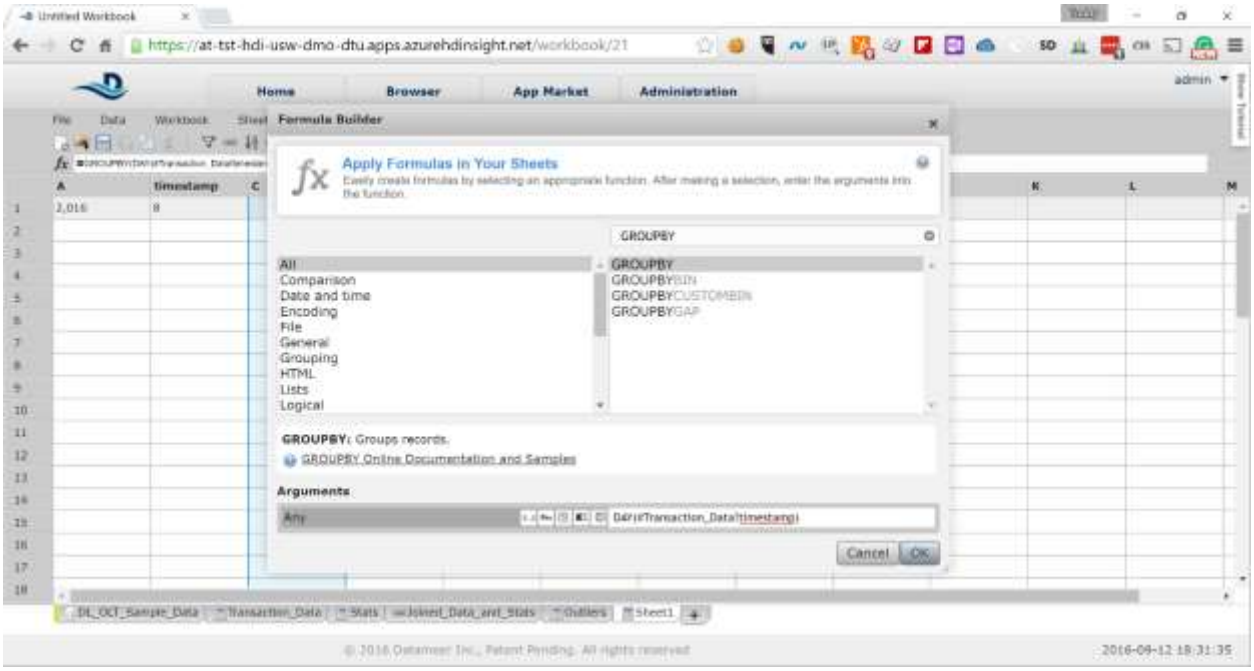
17. Click on the next column and in the formula pop-up select again the *GROUPBY* function and paste the following formula in the *Arguments* field:

MONTH(#Transaction_Data!timestamp)

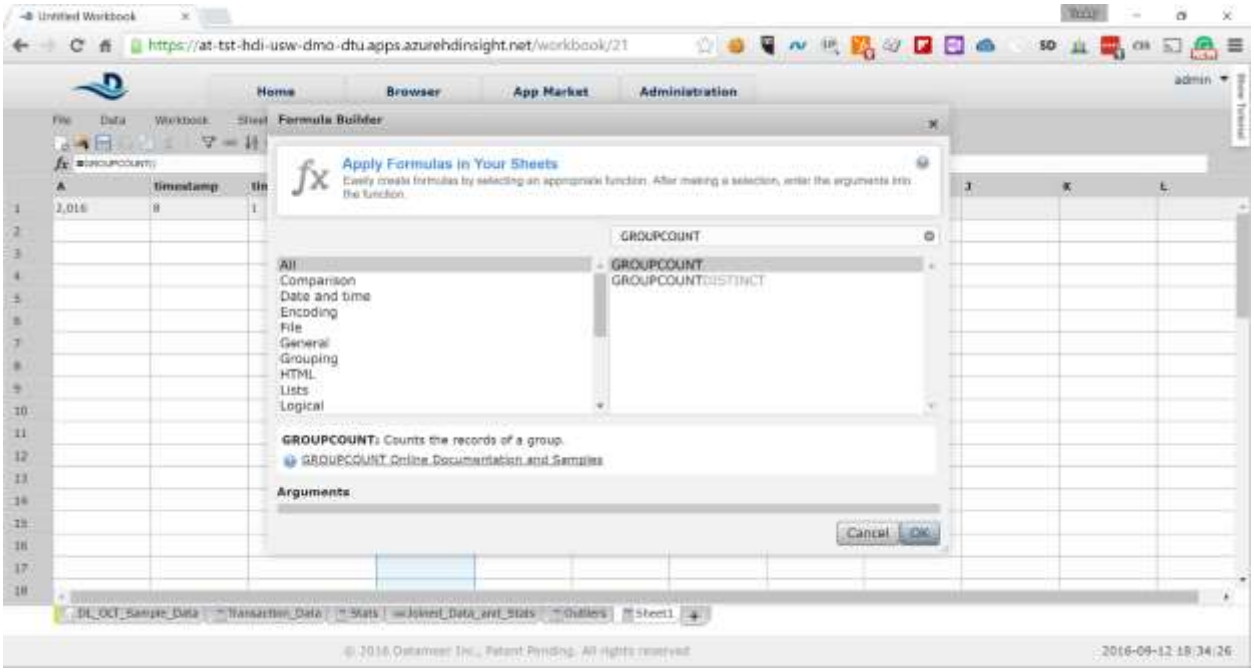


18. Click on the third column and in the formula pop-up select again the *GROUPBY* function and paste the following formula in the *Arguments* field:

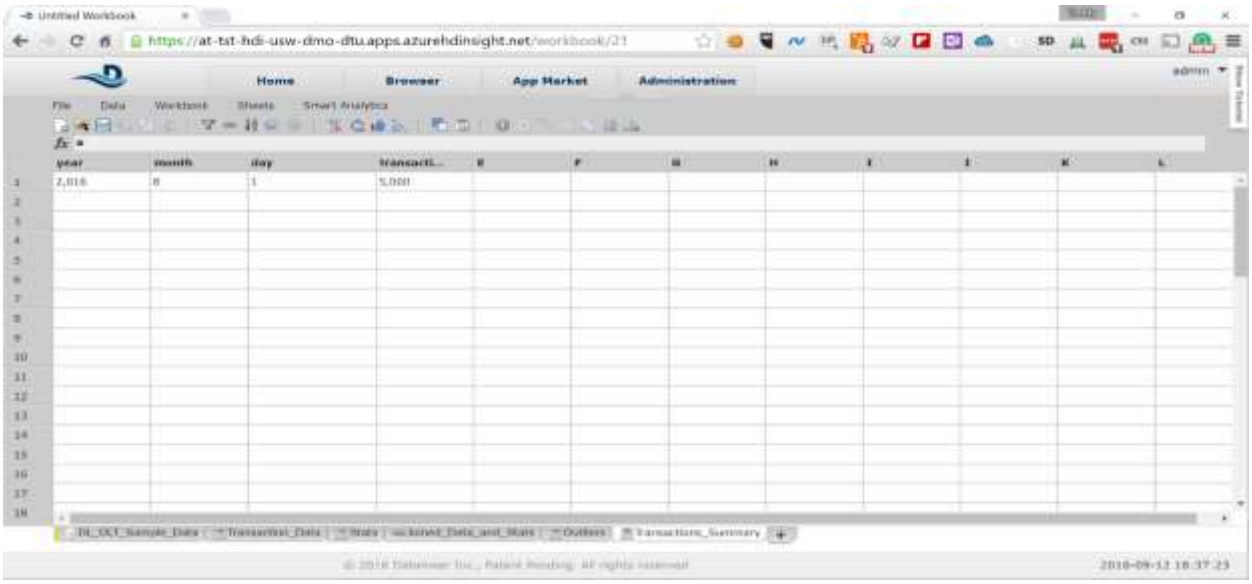
    DAY(#Transaction_Data!timestamp)



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



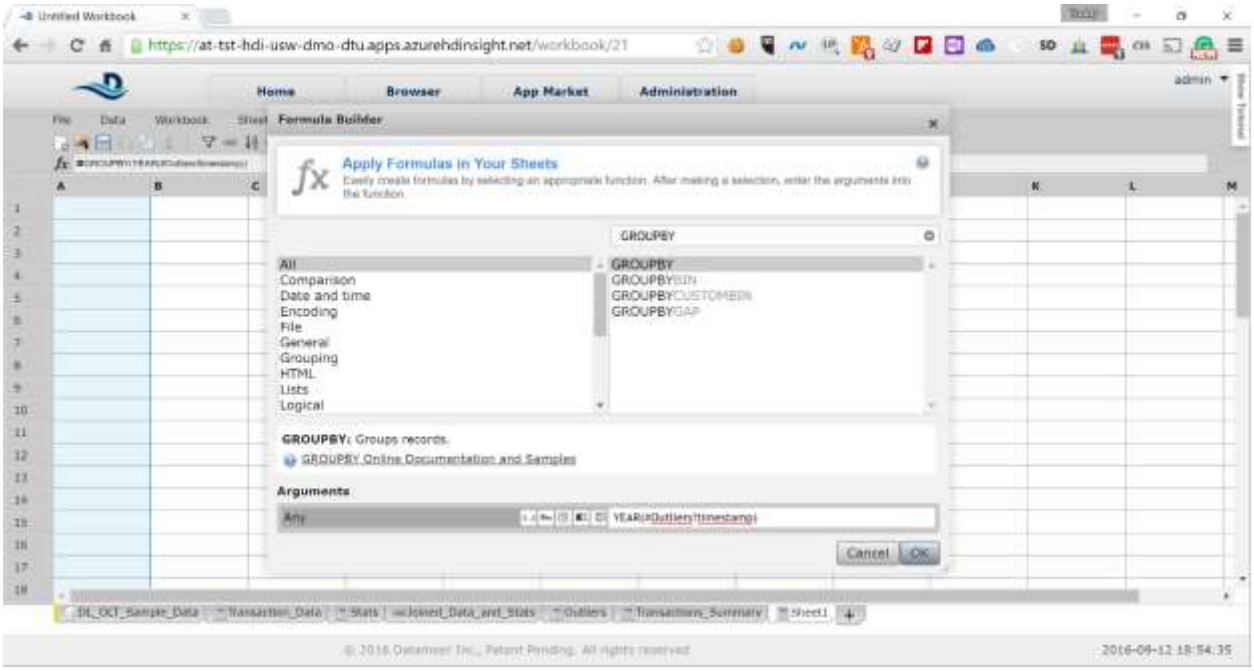
20. We have created summary sheet for our transaction data. Rename the field names as follows:
    year
    month
    day
    transactions_count



Also, rename the sheet to *Transactions_Summary*

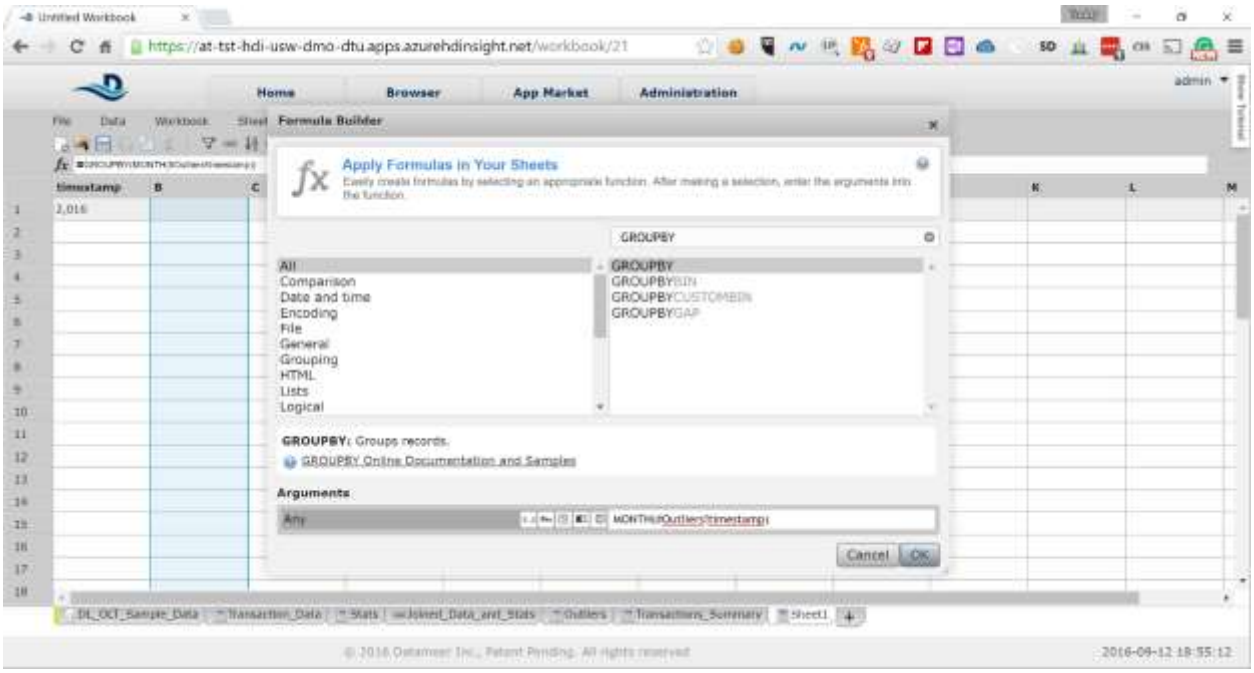
21. Let's create similar summary for the outliers. Create new sheet and in the formula pop-up select the *GROUPBY* function. In the *Arguments* field type the following formula:

    YEAR(#Outliers!timestamp)



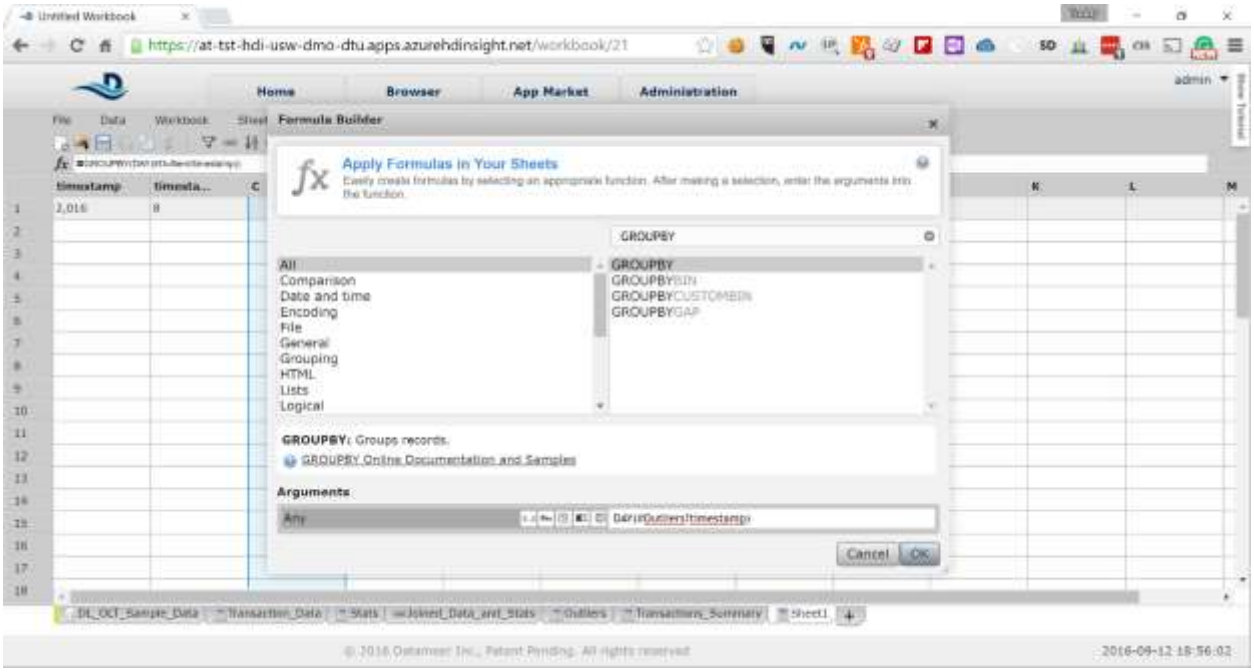
22. Click on the next column and in the formula pop-up select again the *GROUPBY* function and paste the following formula in the *Arguments* field:

    MONTH(#Outliers!timestamp)

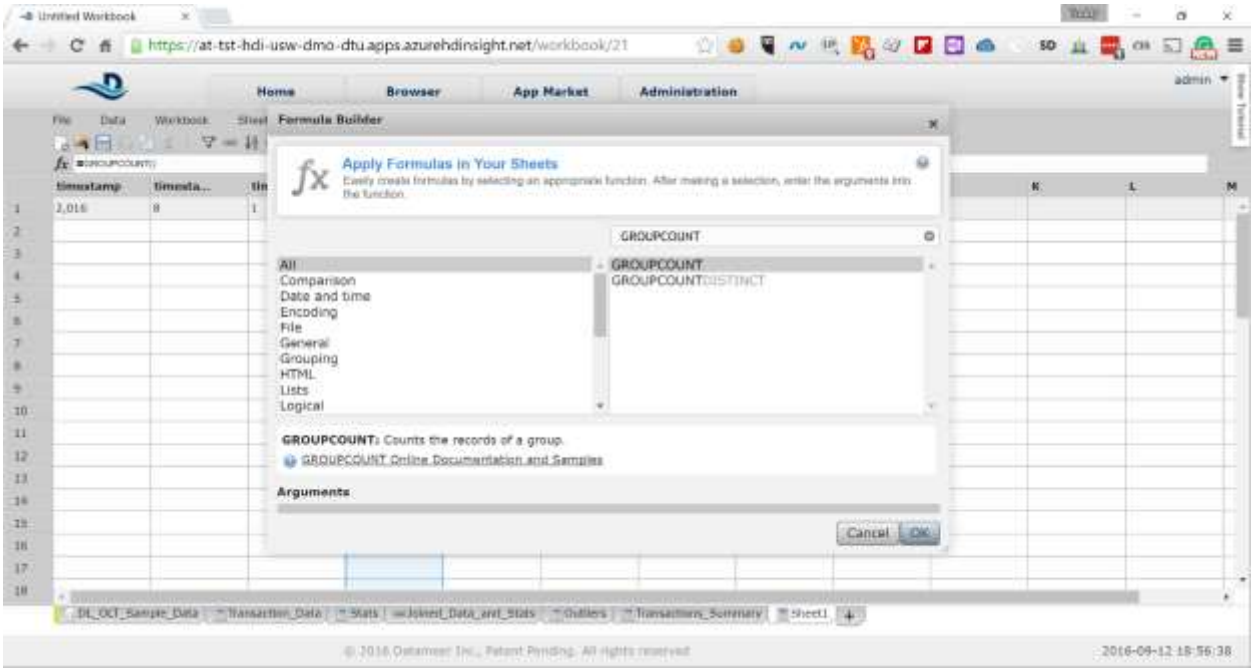


23. Click on the third column and in the formula pop-up select again the *GROUPBY* function and paste the following formula in the *Arguments* field:

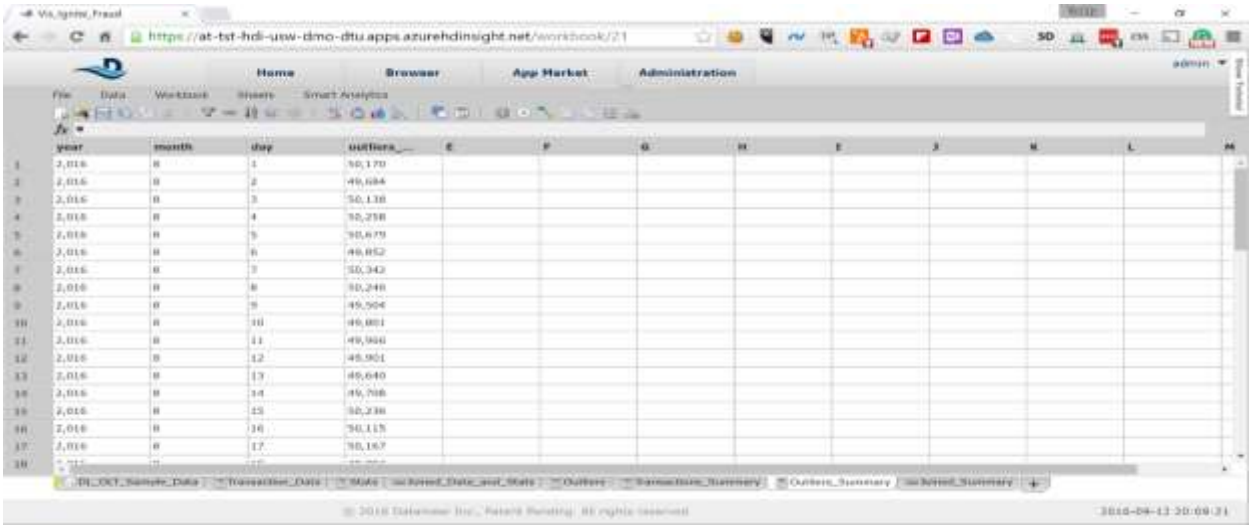
    DAY(#Outliers!timestamp)



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

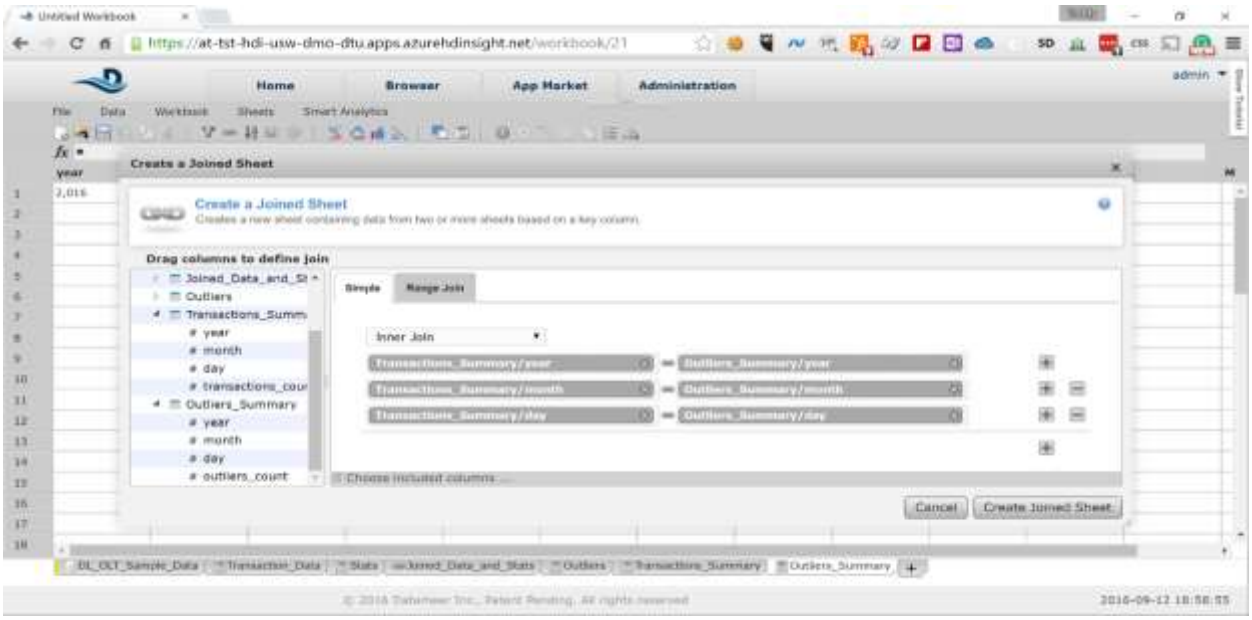


25. We have created summary sheet for our transaction data. Rename the field names as follows:
    year
    month
    day
    outliers_count



Also, rename the sheet to *Outliers_Summary*

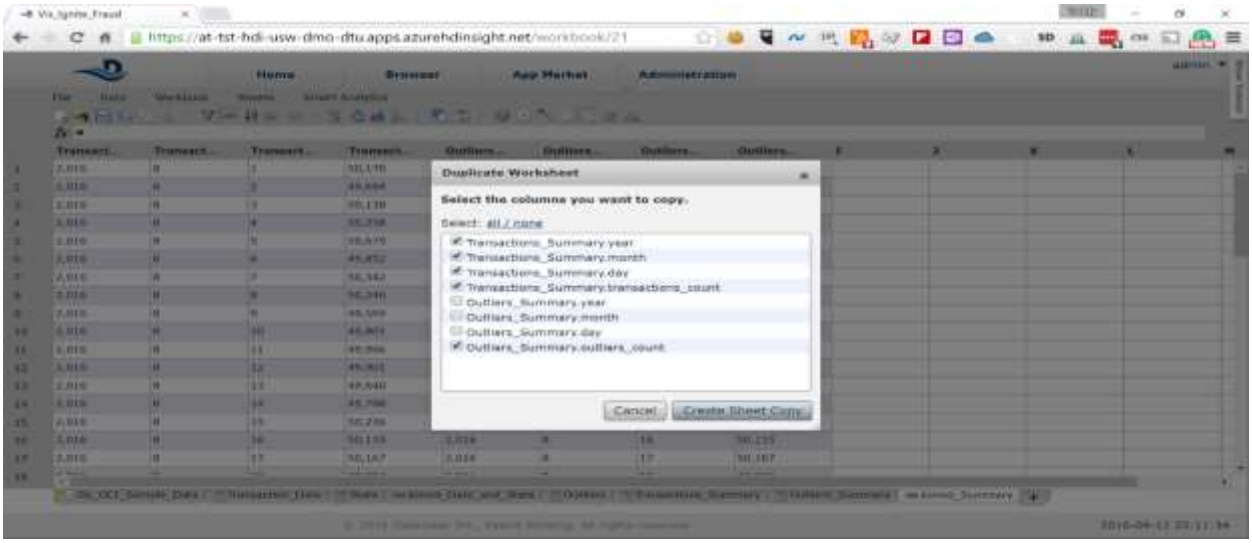
26. We need to join the two summary sheets to have the results available in a single sheet for visualization. Select *Data -> Join* and join the *Transactions_Summary* and *Outliers_Summary* sheets by year, month and date as on the picture below by clicking on the *Create Joined Sheet* button



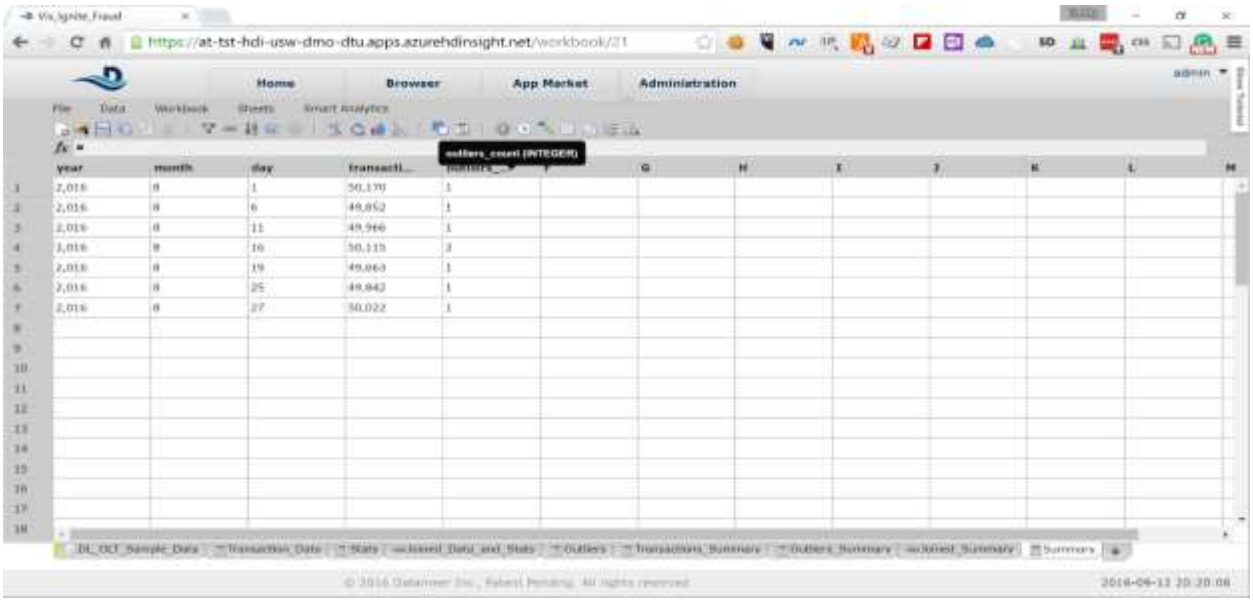
Rename the joined sheet to *Joined_Summary*

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

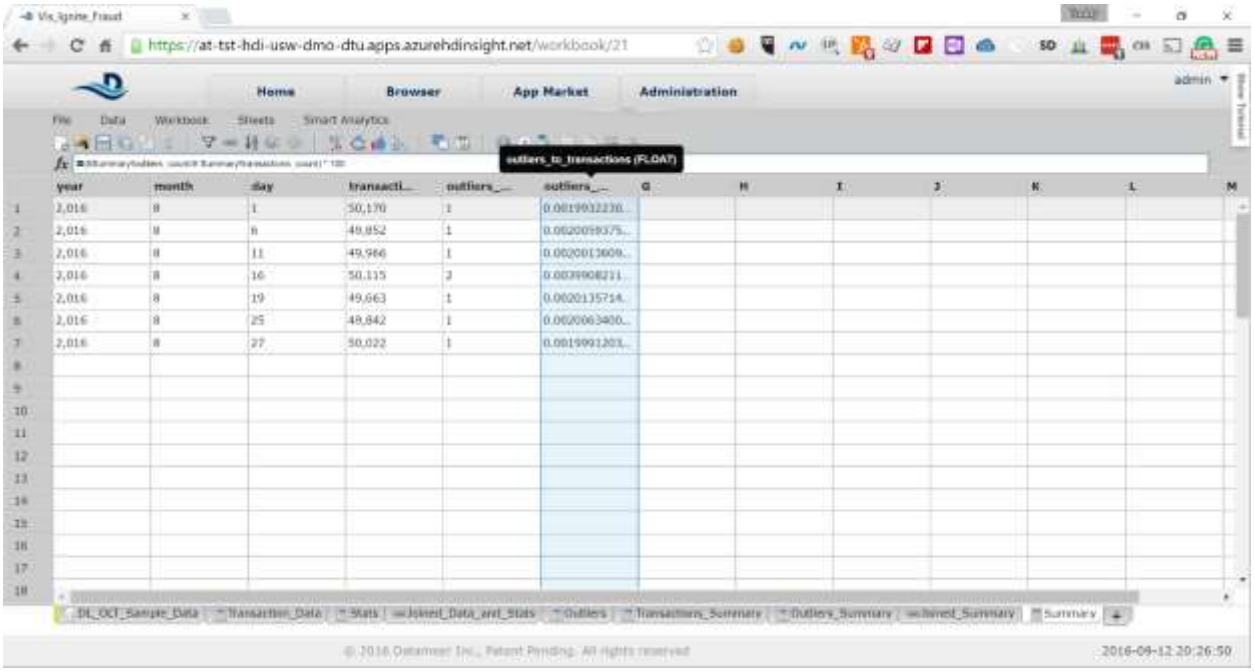


28. Rename the sheet to *Summary* and the columns as follows:
    Transactions_Summary.year -> year
    Transactions_Summary.month -> month
    Transactions_Summary.day -> day
    Transactions_Summary.transactions_count - > transactions_count
    Outliers_Summary.outliers_count -> outliers_count



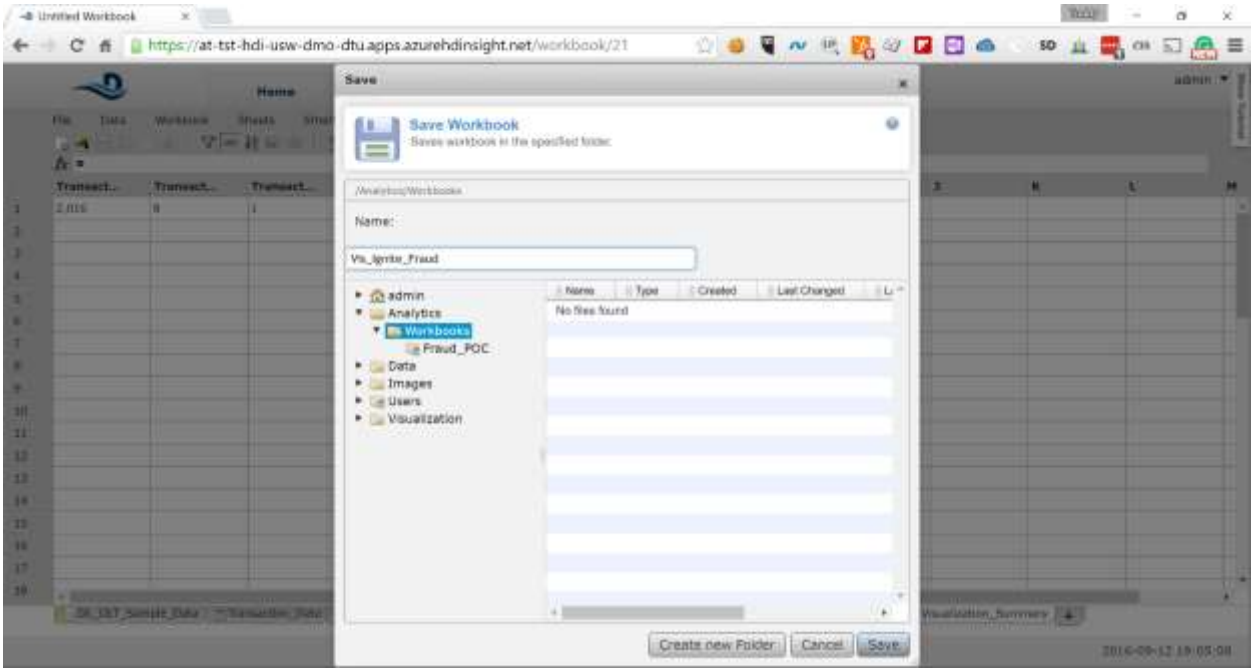
29. Click on the sixth column and cancel the formula pop-up. In the f_x field on top of the sheet type the following:

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

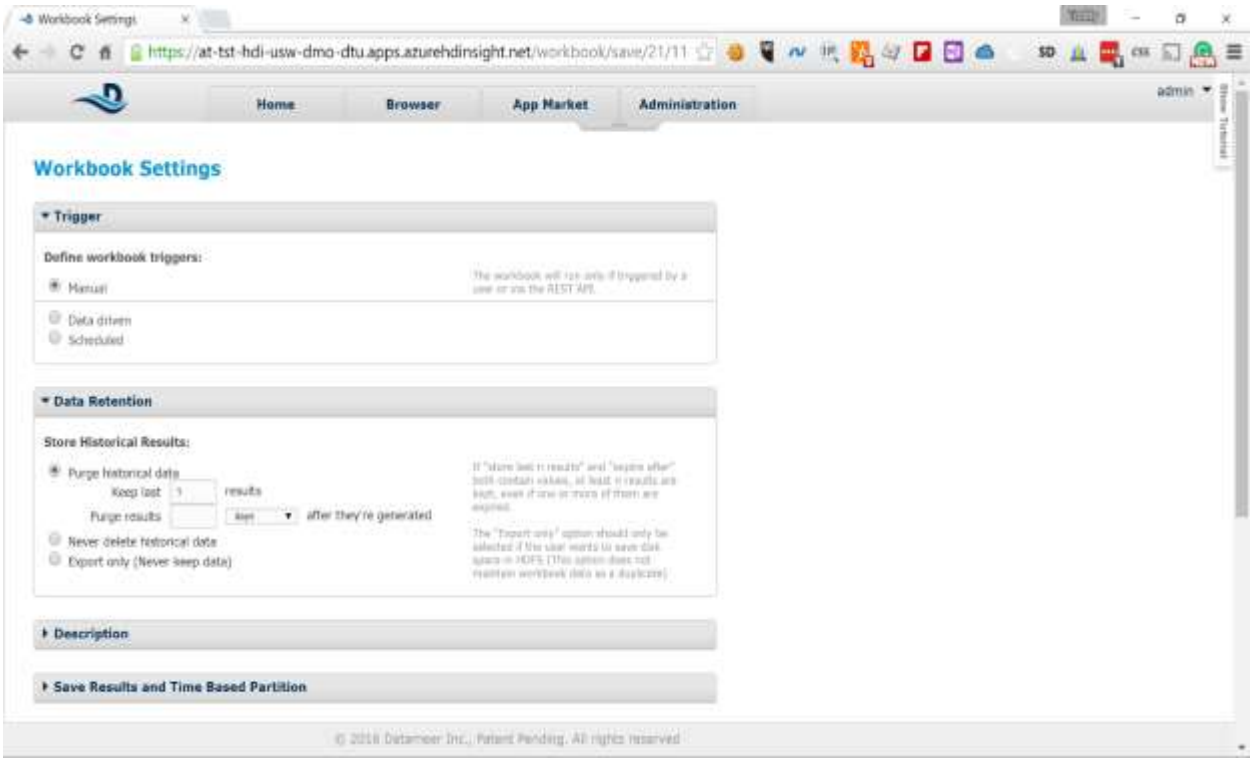


Also, rename the field to *outliers_to_transactions*

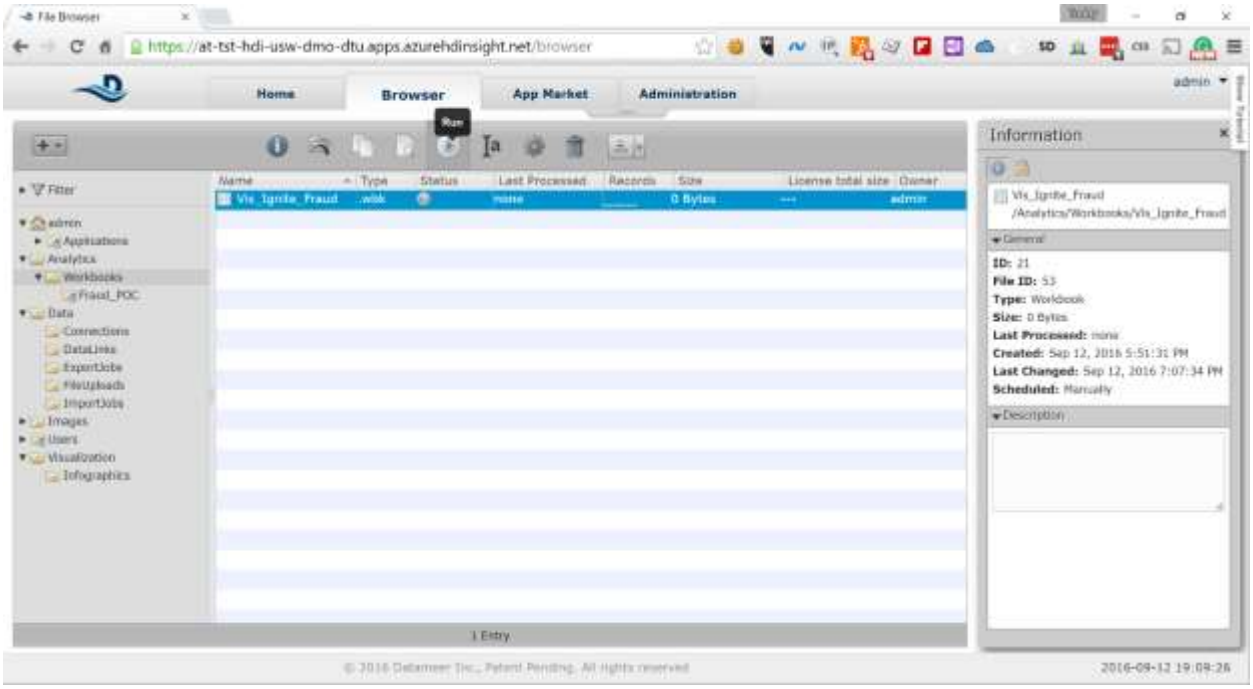
30. Selct *File* -> *Save* from the menu and type the *Vis_Ignite_Fraud* in the *Name* field



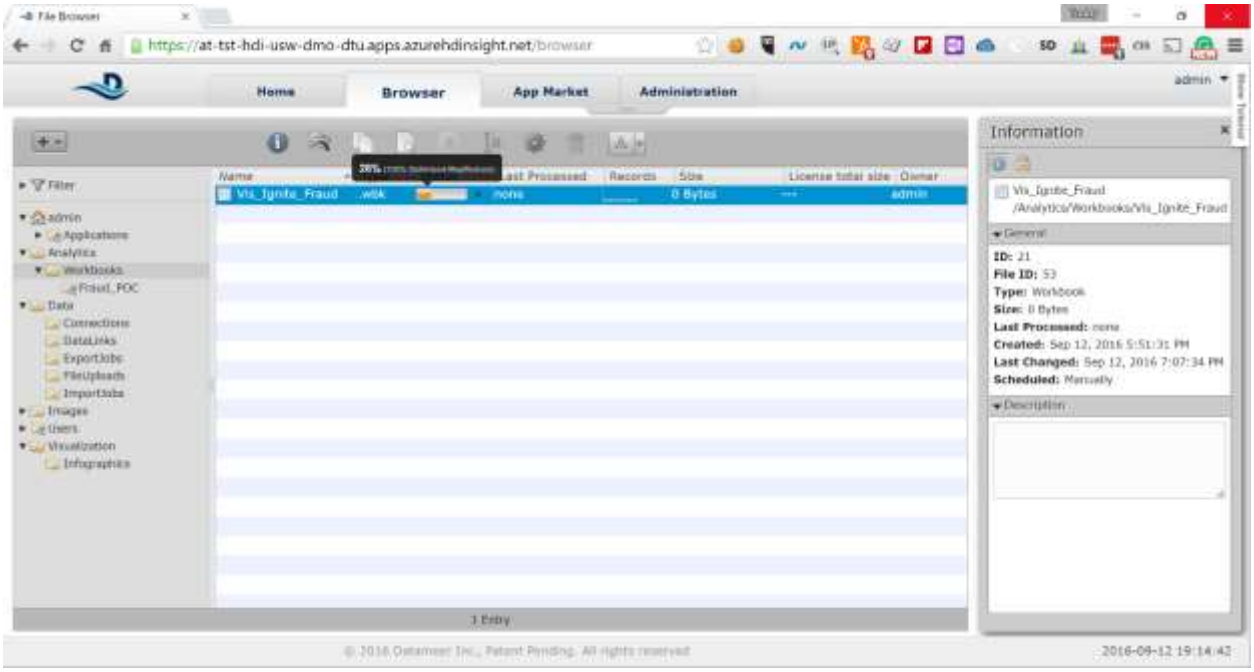
31. On the next screen keep the default values for all the fields. Scroll down and click on the *Save* button again



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

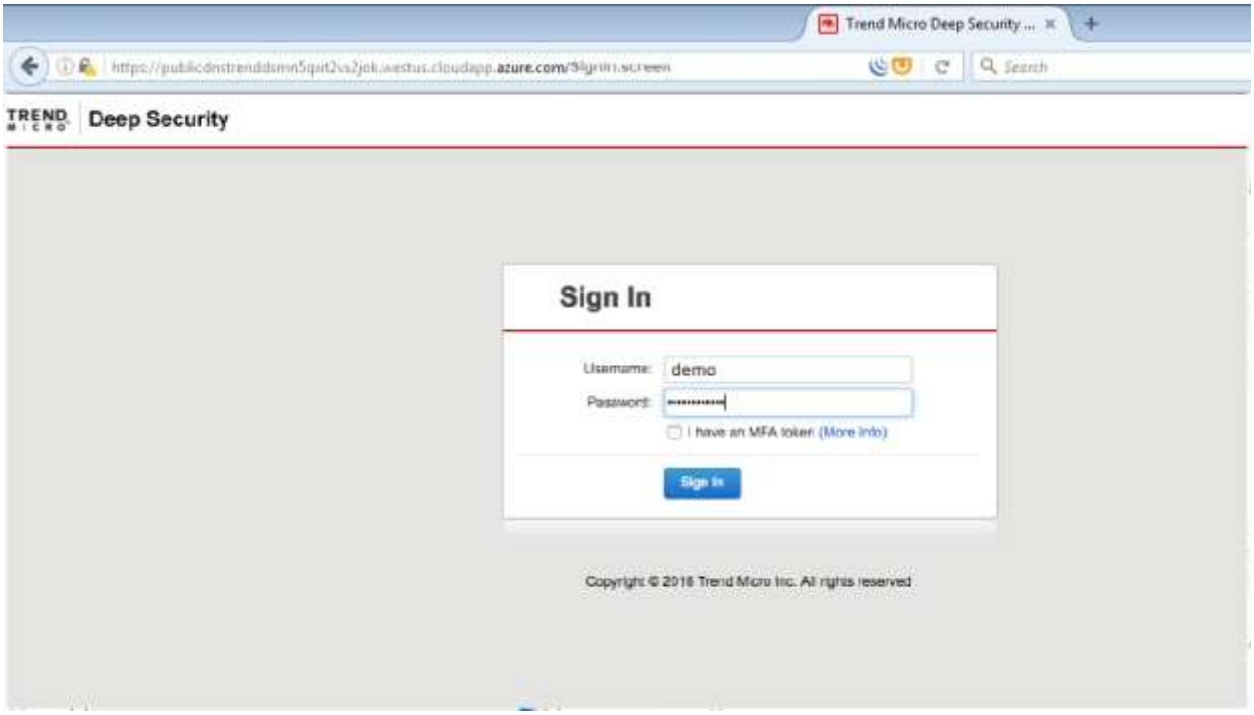
8.1 Server name

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

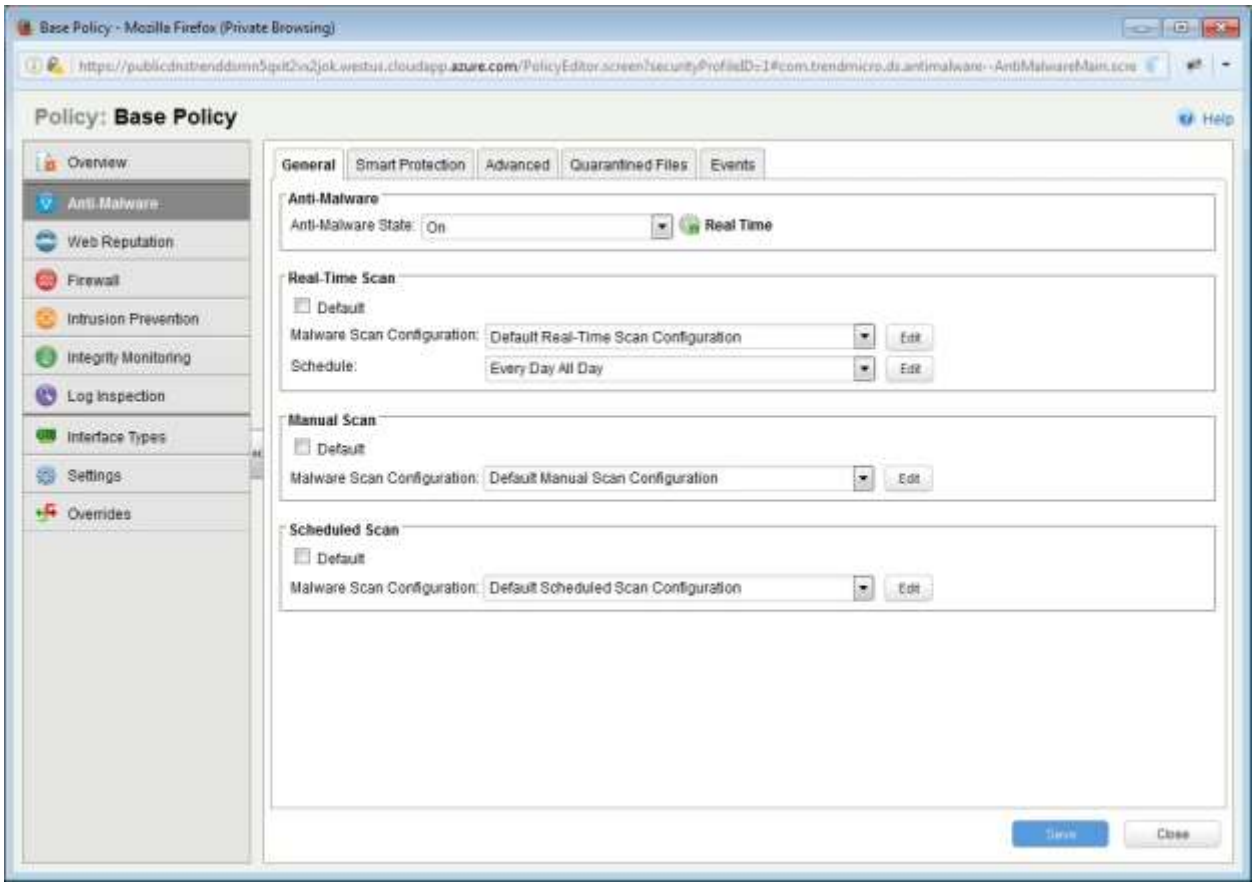
1. Changing the base policy

   Go to policies->Base Policy

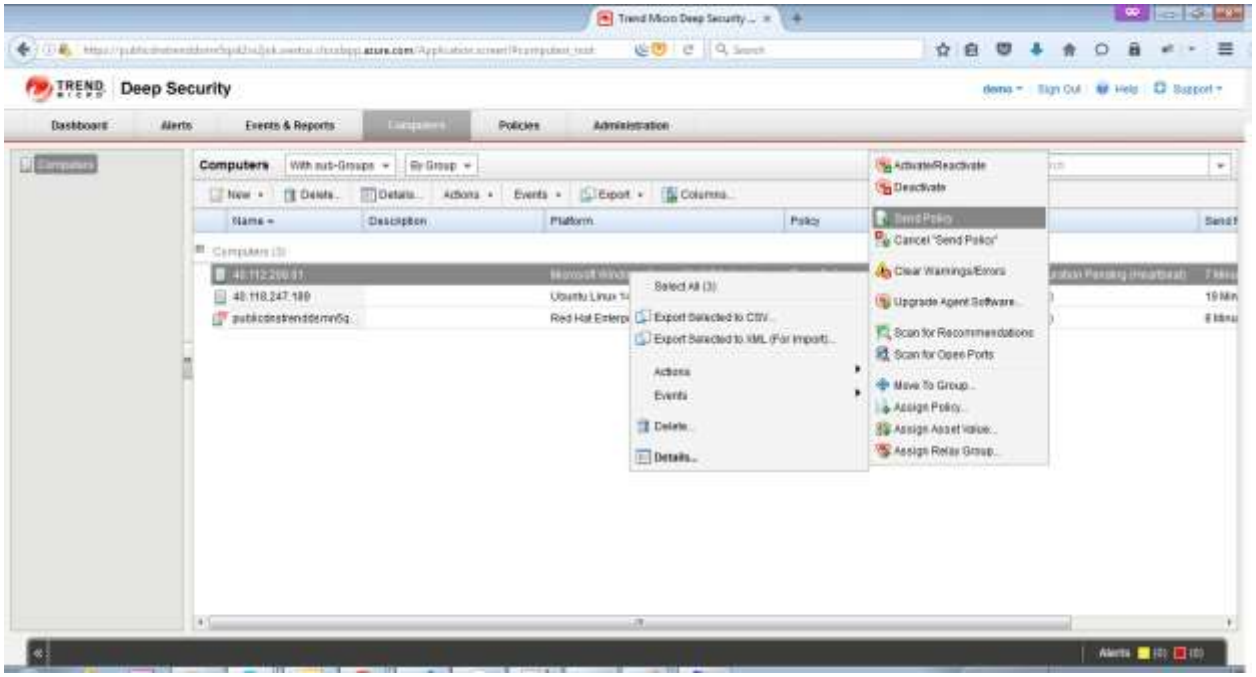
2. Enable Anti-Malware

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

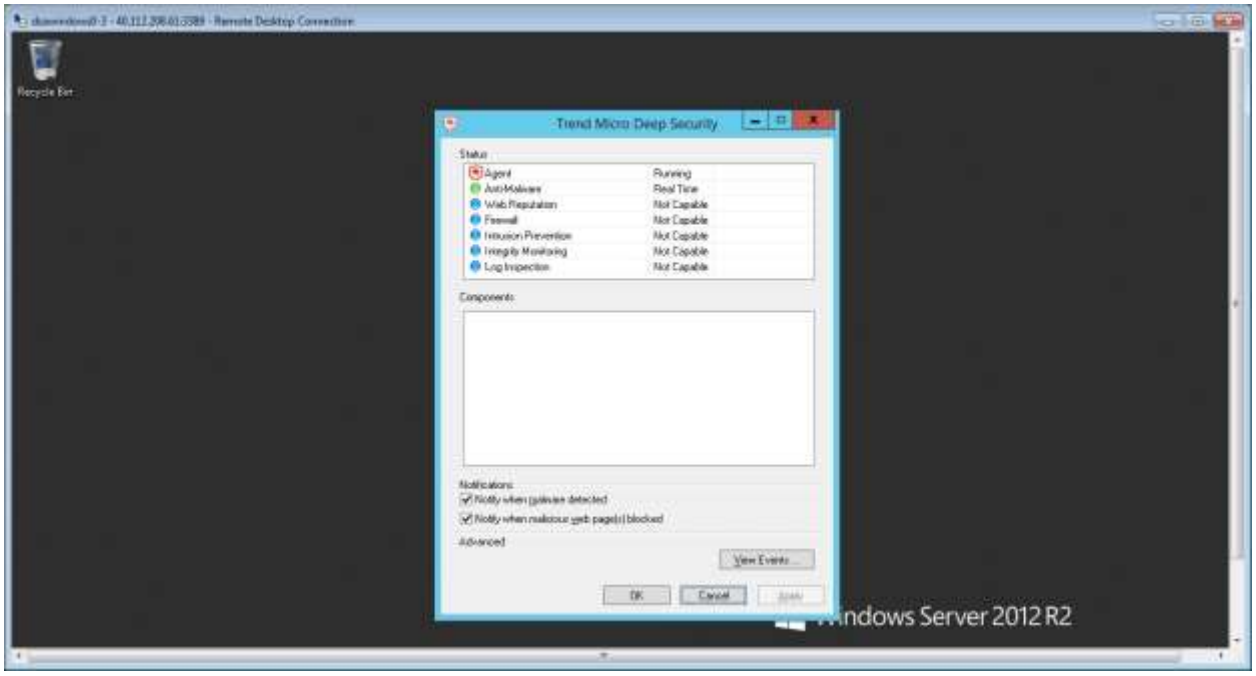
   Click “Save”



3. Applying policies to computer



4. Verifying policy in the computer



10 Exercises

10.1 Datameer – Visualize the Data

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

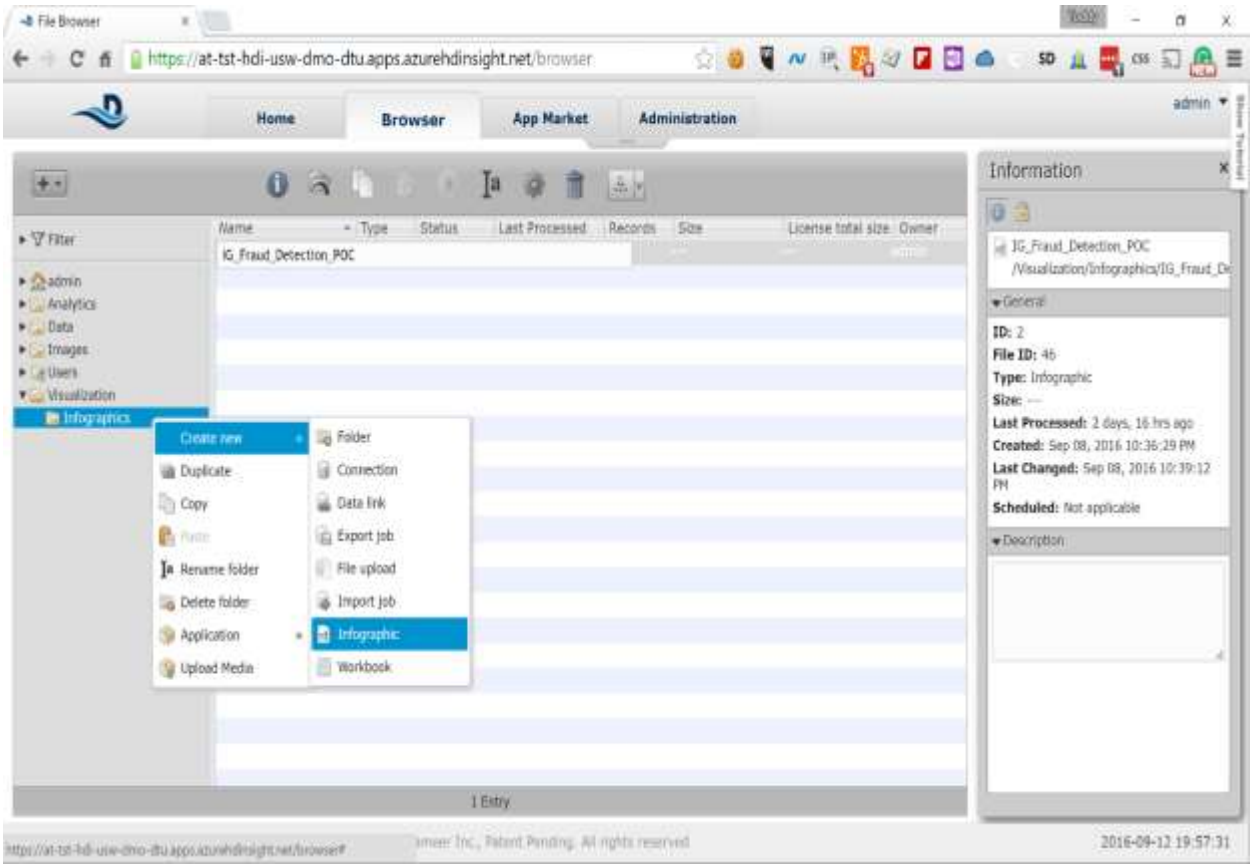
10.2 TrendMicro – Malware test

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

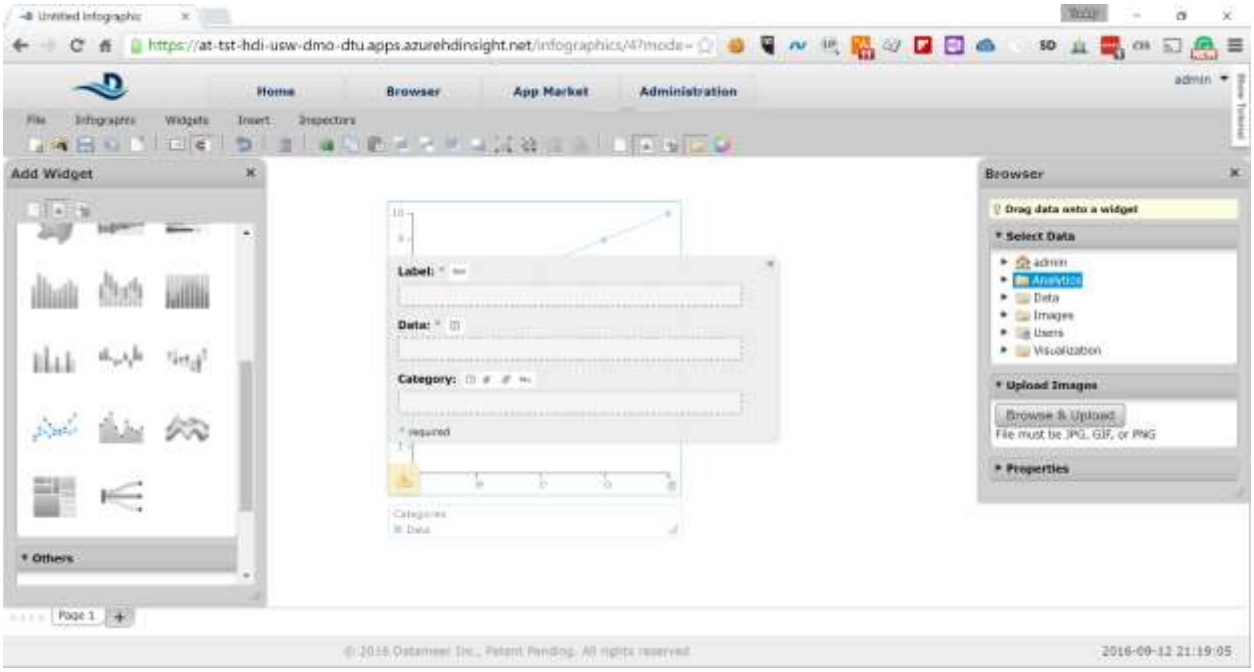
11 Visualize the Data

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:

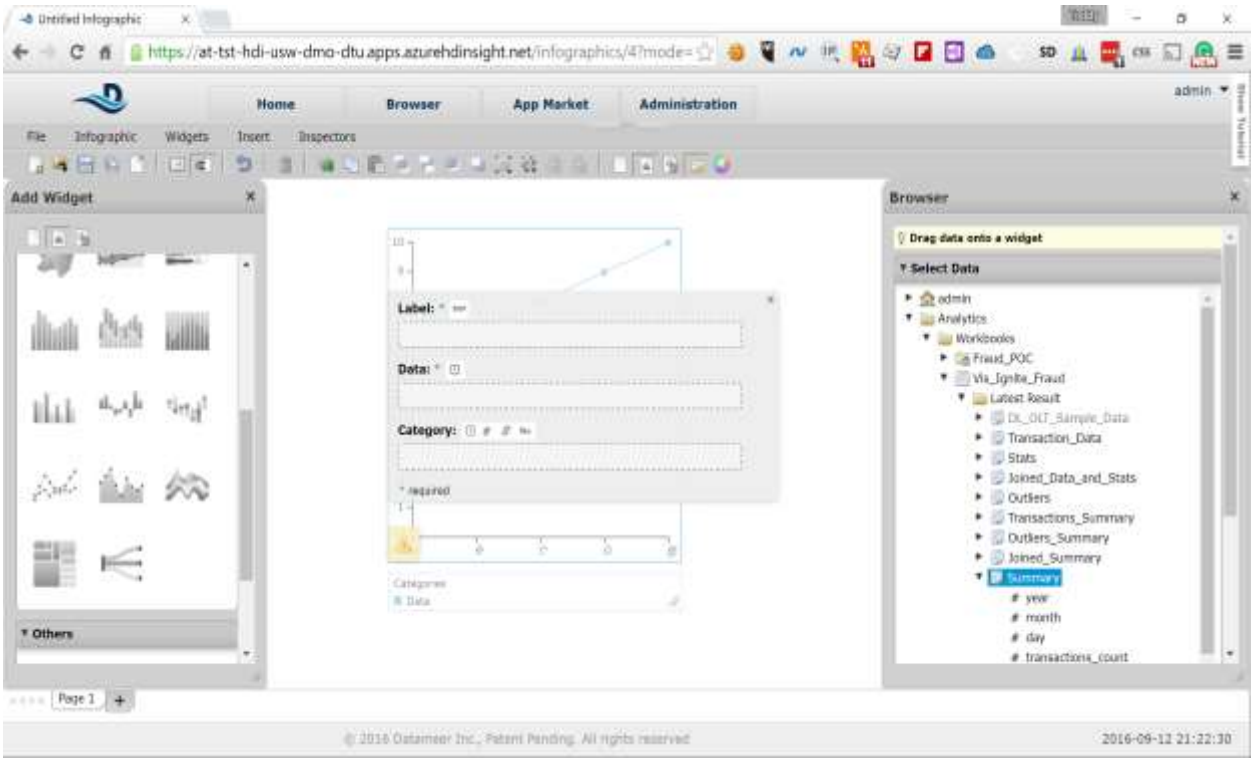
1. In Datameer's Browser view expand the Visualization node and right click on Infographics -> Create New -> Infographic



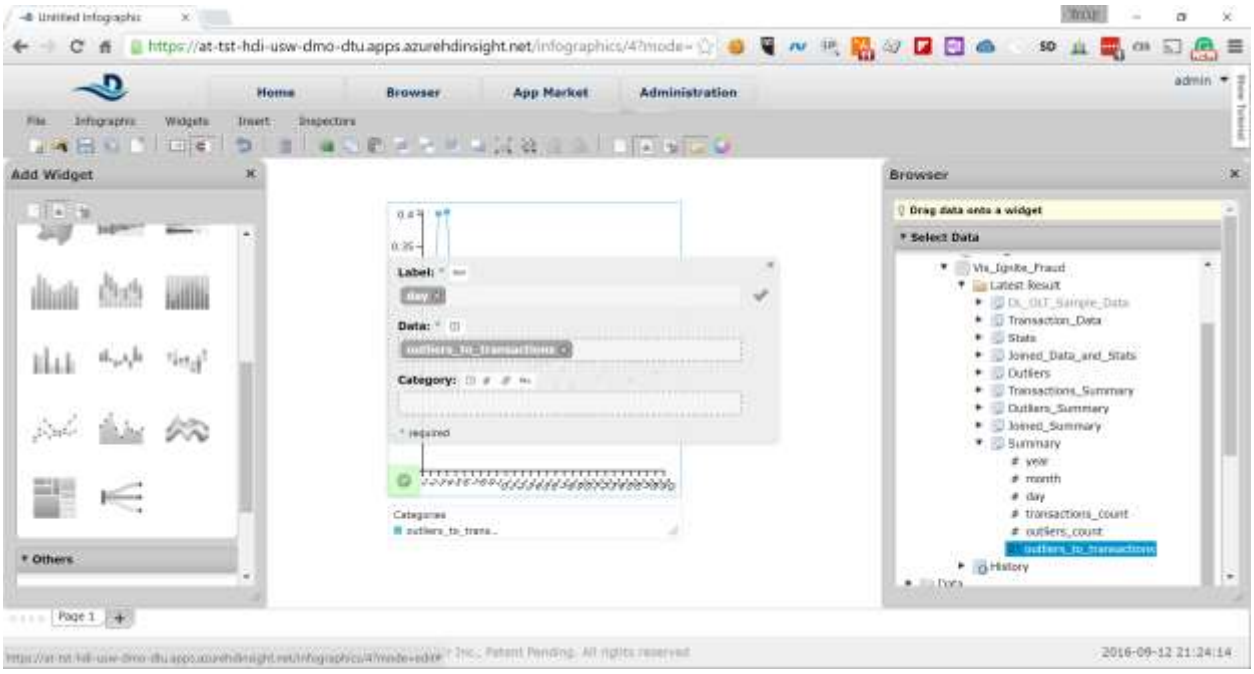
2. Drag the *Line and Area Chart* from the *Add Widget* pane on the left to the work pane in the middle



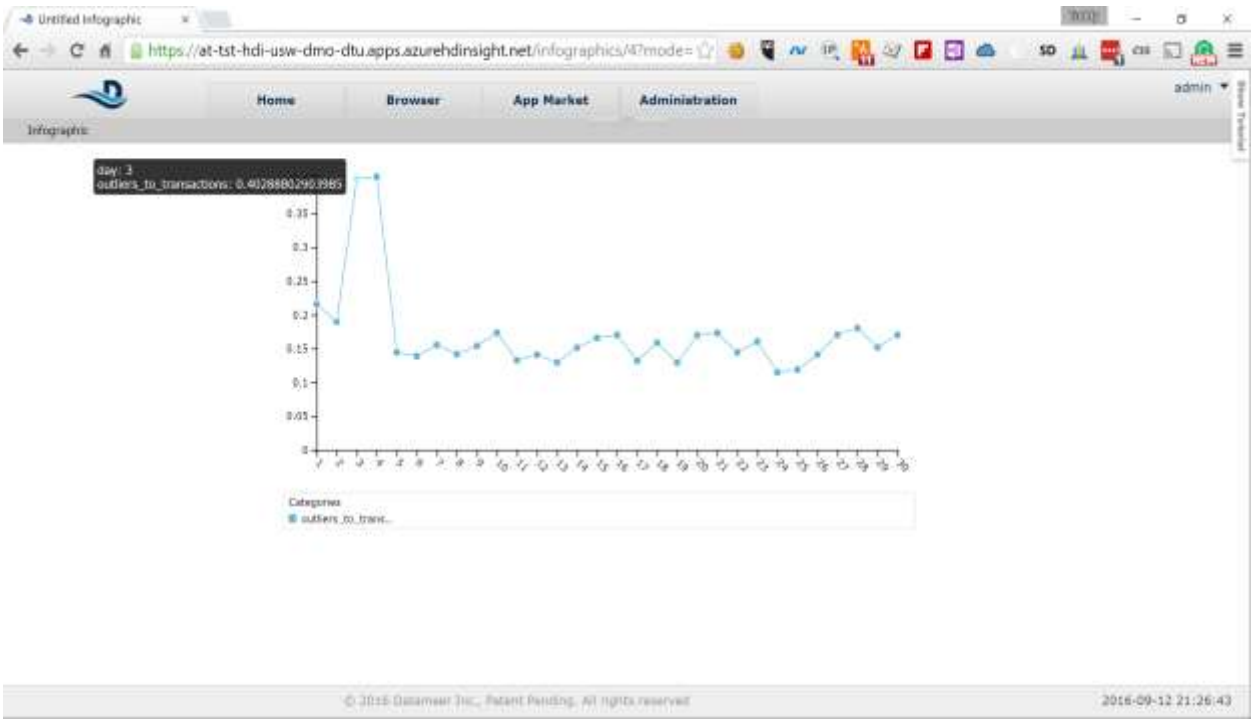
3. In the Browser pane expand *Analytics* node and then *Workbooks -> Vis_Ignite_Fraud -> Latest Results -> Summary*



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

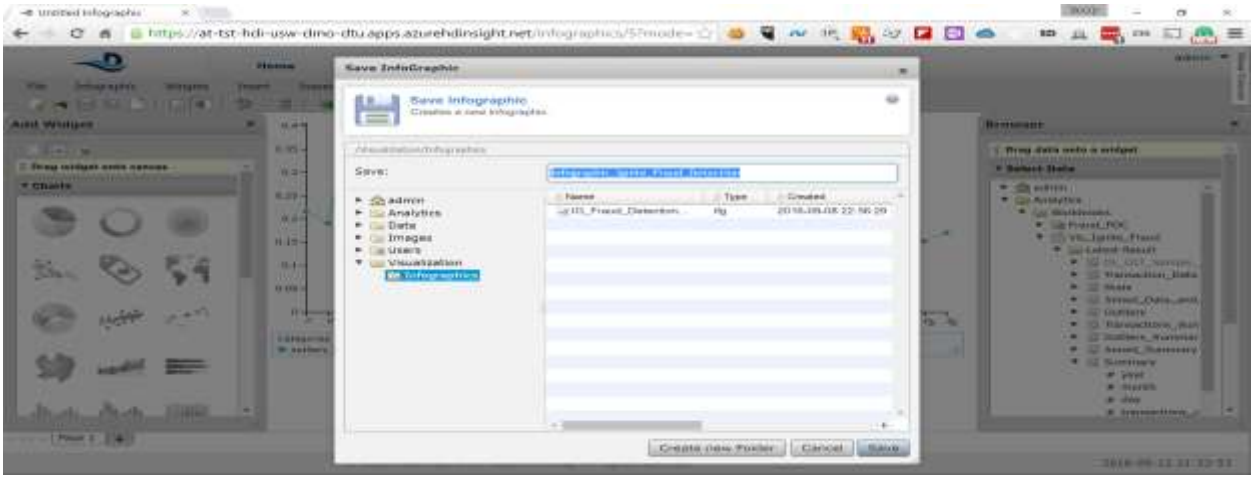


5. Select Infographic -> View from the menu to present the infographic. You can easily see that on the 3rd and 4th day of the month the outliers significantly spiked, which is a sign of something unusual going on those two days



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

   Infographic_Ignite_Fraud_Detection
   and click on the *Save* button



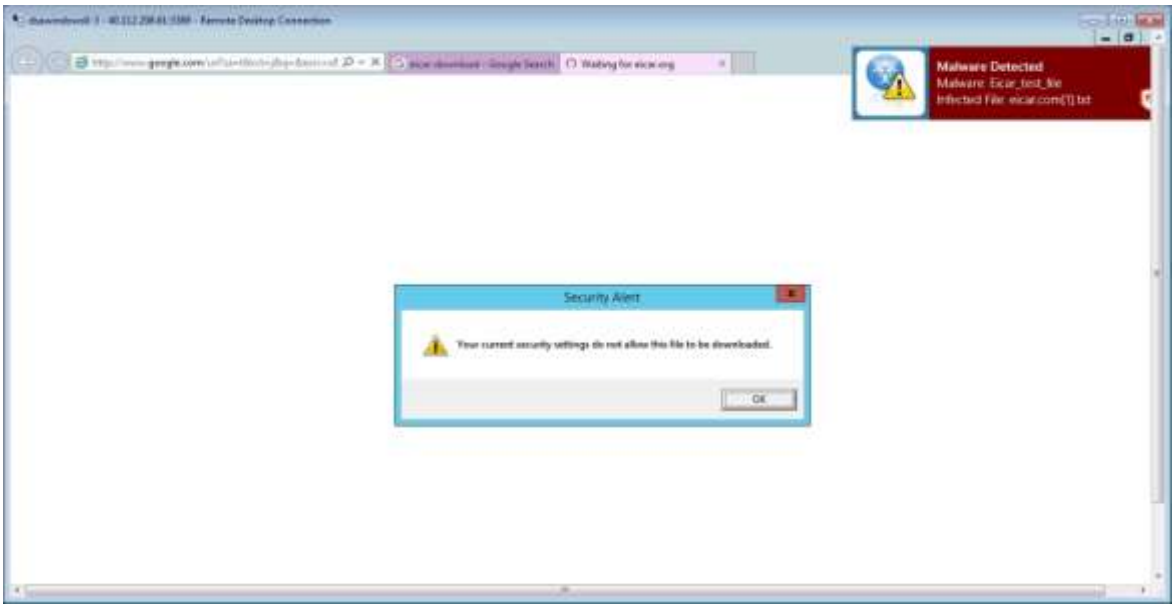
12 Malware Test

12.1 Generating Malware alert in the computer

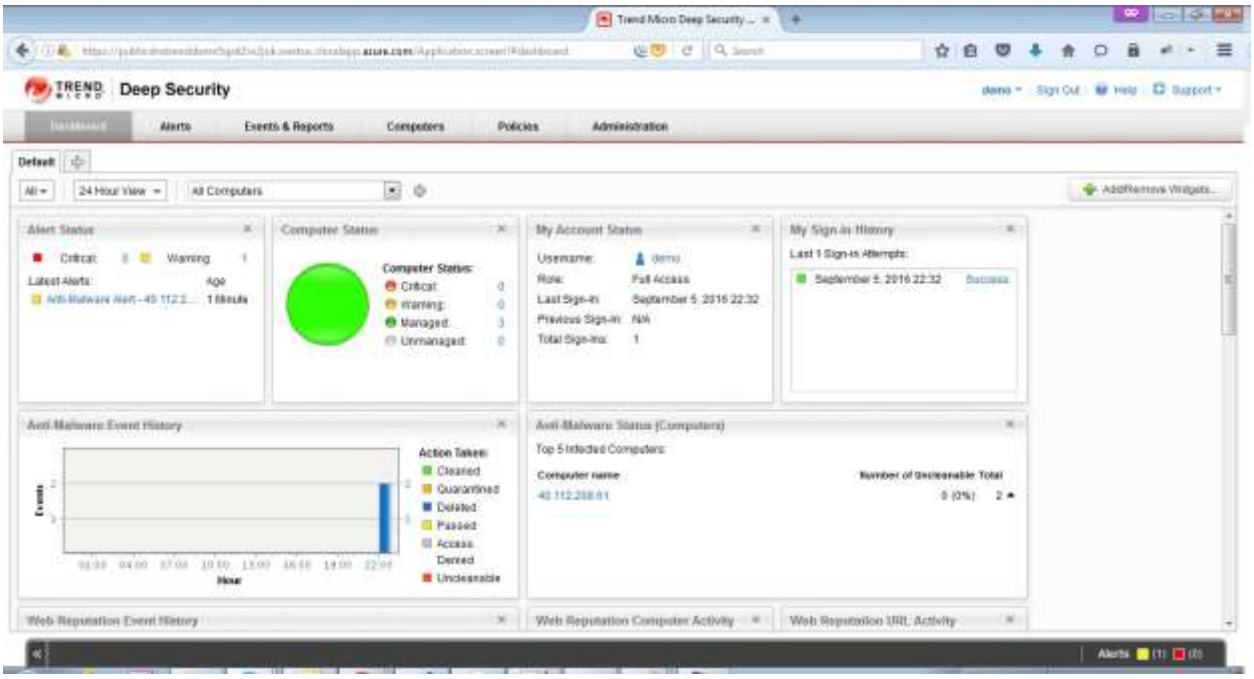
The Malware test can be performed by going to the url

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=AFQjCNE8DvVl7BE5Nd2hq1zNDTP6hNjcIA&bvm=bv.132479545,d.cGc

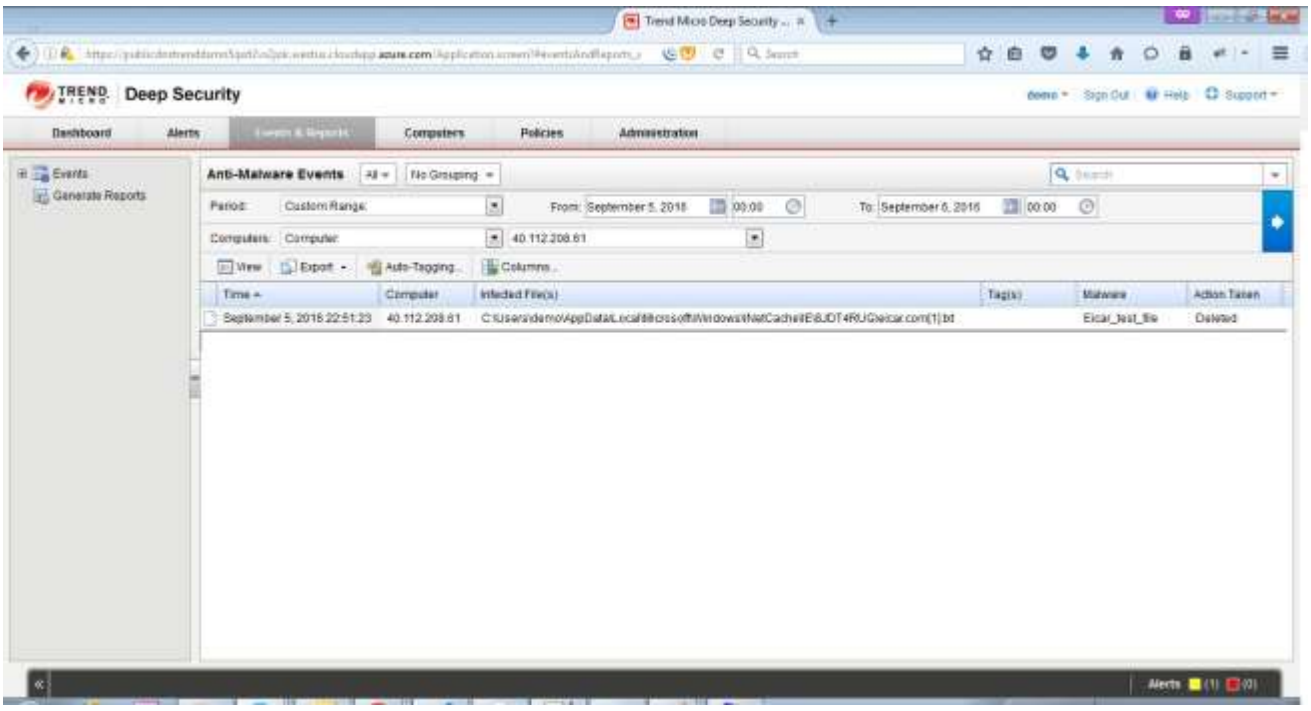
this is eicar malware test



12.2 Dashboard – Malware Alert



12.3 Malware Alert vertification



13 References, Attachments & Definitions

13.1 References

No.	Document Title	Link/ Attachment	Comments
1			