Hands on Tutorial
k-means Clustering Model

Tutorial #3 Questions:

- What are we trying to accomplish?
  - Build a clustering algorithm to classify wells based on dynamometer card data.
  - Which are similar?
  - Which are dissimilar?
- How can we use Azure Machine Learning to address this question?

Microsoft NEAL ANALYTICS

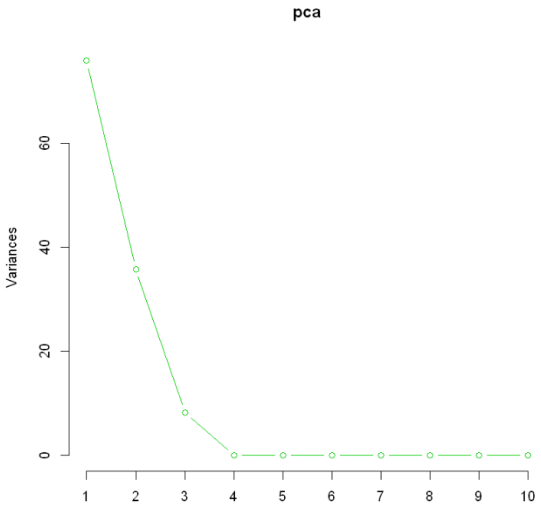
Scenario Data:

well	date	idx	x	y
Alabama	8/16/2013 1:55	1	0.00074002	-5327.23698
Alabama	8/16/2013 1:55	2	0.15595834	-4476.758361
Alabama	8/16/2013 1:55	3	1.536036006	-3475.778052
Alabama	8/16/2013 1:55	4	3.779102308	-2636.548613
Alabama	8/16/2013 1:55	5	6.817062356	-1999.545532
Alabama	8/16/2013 1:55	6	10.7311654	-1622.668567
Alabama	8/16/2013 1:55	7	15.5308829	-1432.091513
Alabama	8/16/2013 1:55	8	21.09257393	-1346.745211
Alabama	8/16/2013 1:55	9	27.24284435	-1325.808833

- ~3.4 million rows of data
- 263 unique wells
- The data comprises of the following:
  - Date-Time at which the readings were taken
  - Idx - Reading number (1-60): During every up and down motion of the Pump Jack, 60 sets of reading will be taken at various positions
  - X - Position of the Pump Jack
  - Y – Load on the Pump Jack
- We then perform the Principal Component Analysis over this data which was covered during your Exploratory Data Analysis lecture

Scenario Data:

- After the PCA was performed we ended up with~57K rows of data with 3 Principal Components (PC1, PC2, PC3)
- 263 unique wells
- Training data comprises of the following:
  - Date-Time at which the readings were taken
  - PC1
  - PC2
  - PC3

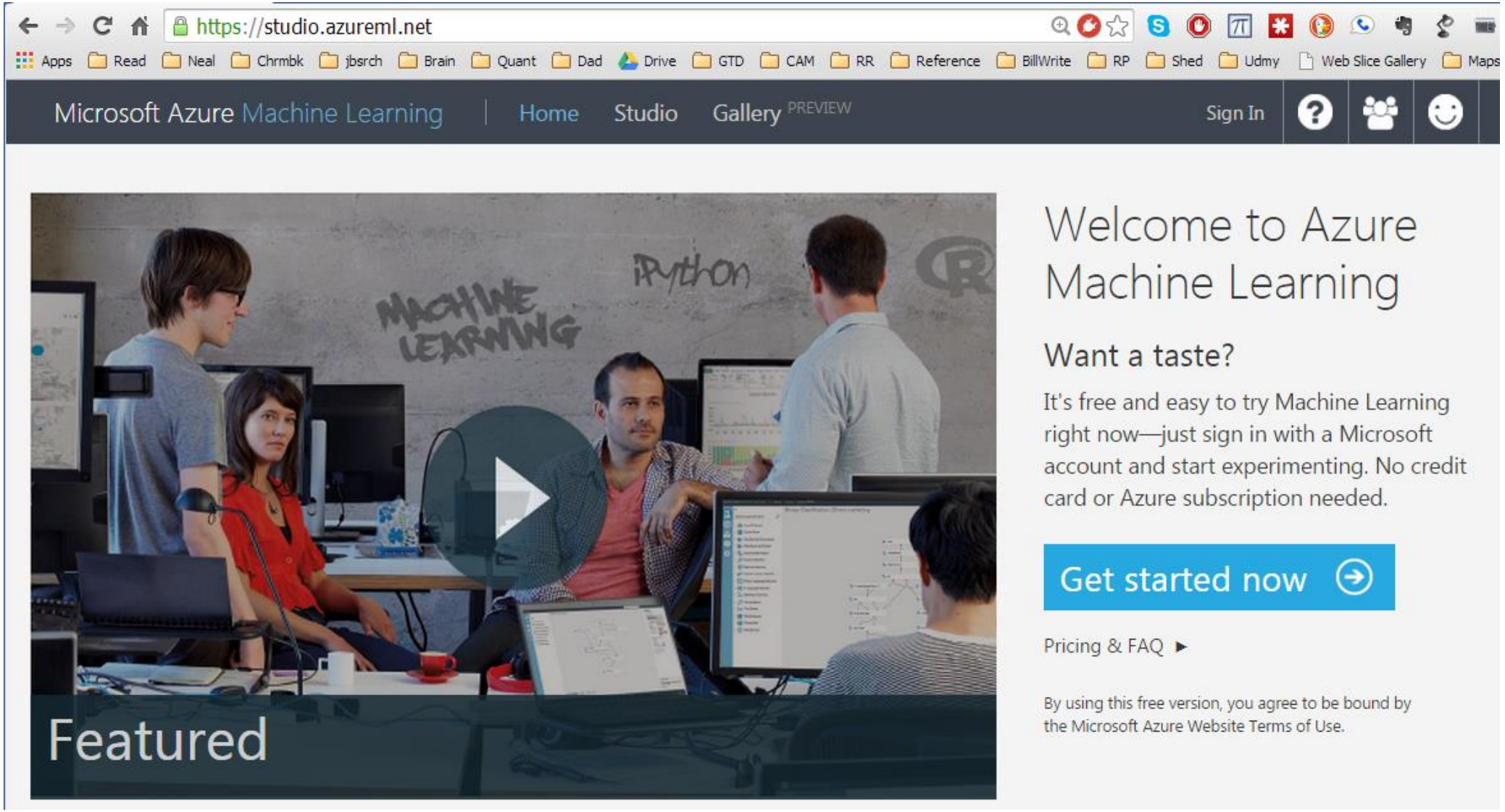


well	date	PC1	PC2	PC3
Alabama	8/16/2013 1:55	-8.243340971	2.918070697	2.023029116
Alabama	8/16/2013 23:49	-8.243340971	2.918070697	2.023029116
Alabama	9/3/2013 0:17	-8.495327766	2.311736767	2.154066737
Alabama	9/6/2013 9:38	-8.148383208	2.731807776	2.341956526
Alabama	9/6/2013 16:13	-8.148383208	2.731807776	2.341956526
Alabama	9/13/2013 3:48	-8.206877815	2.58376763	2.282912974
Alabama	9/13/2013 20:28	-8.206877815	2.58376763	2.282912974
Alabama	11/5/2013 13:17	-8.640142409	2.146508336	2.971266786
Alabama	11/5/2013 17:26	-8.640142409	2.146508336	2.971266786
Alabama	12/12/2013 9:46	-3.139291754	7.206502009	-2.257469677
Alabama	12/12/2013 14:13	-3.139291754	7.206502009	-2.257469677

Microsoft NEAL ANALYTICS

Step by Step Guide for Building a Clustering Model

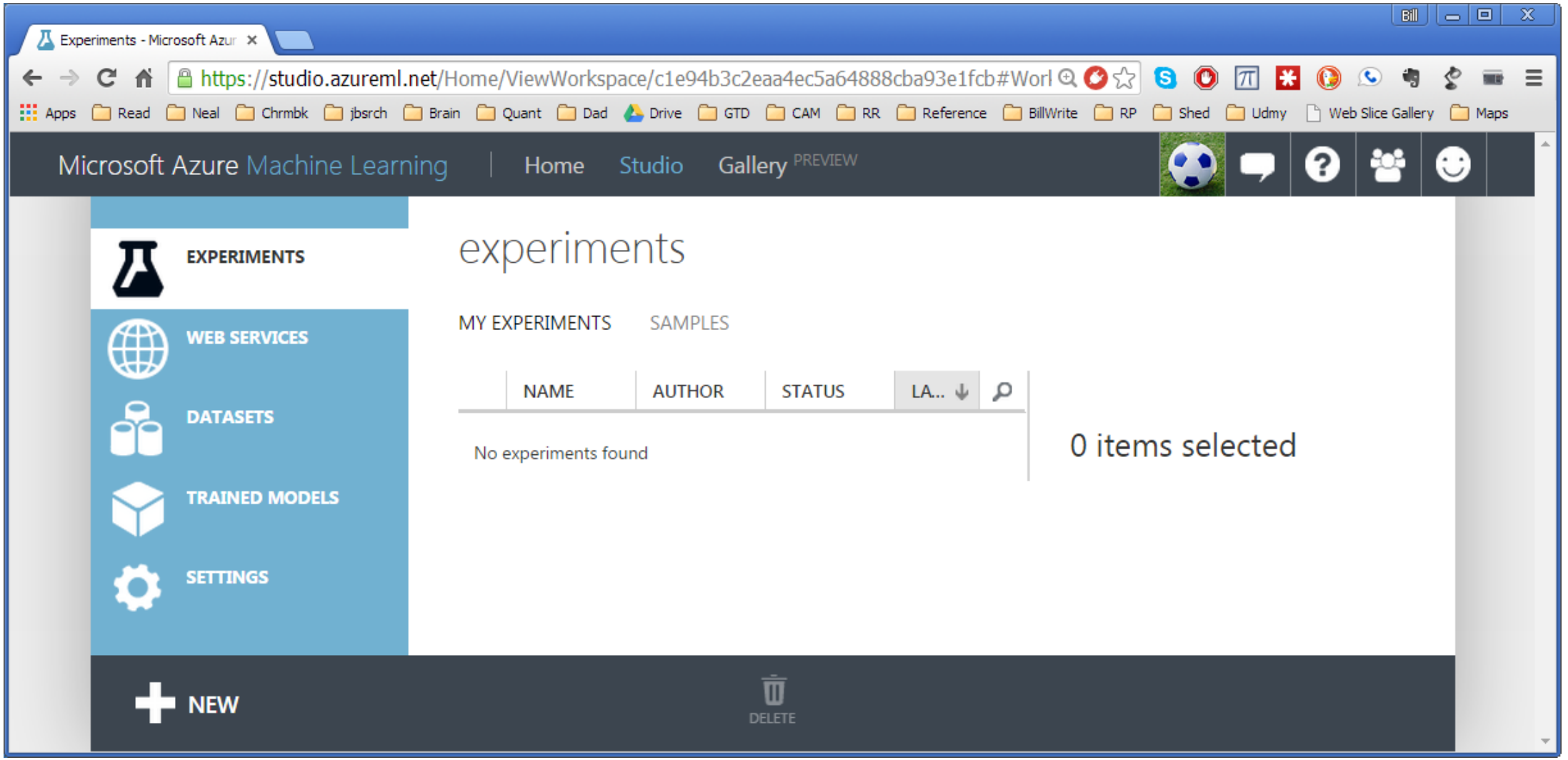
Step 1 : Go to https://studio.azureml.net/



Microsoft

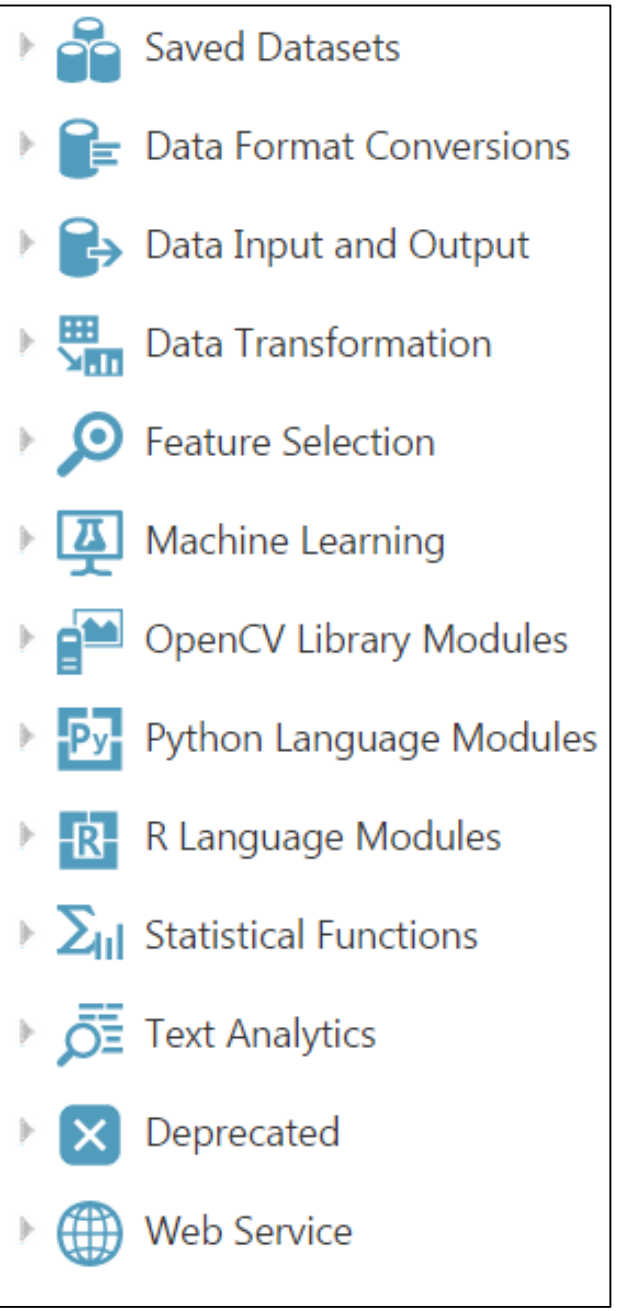
NEAL ANALYTICS

Step 2 : Log in to your account



Microsoft

NEAL ANALYTICS

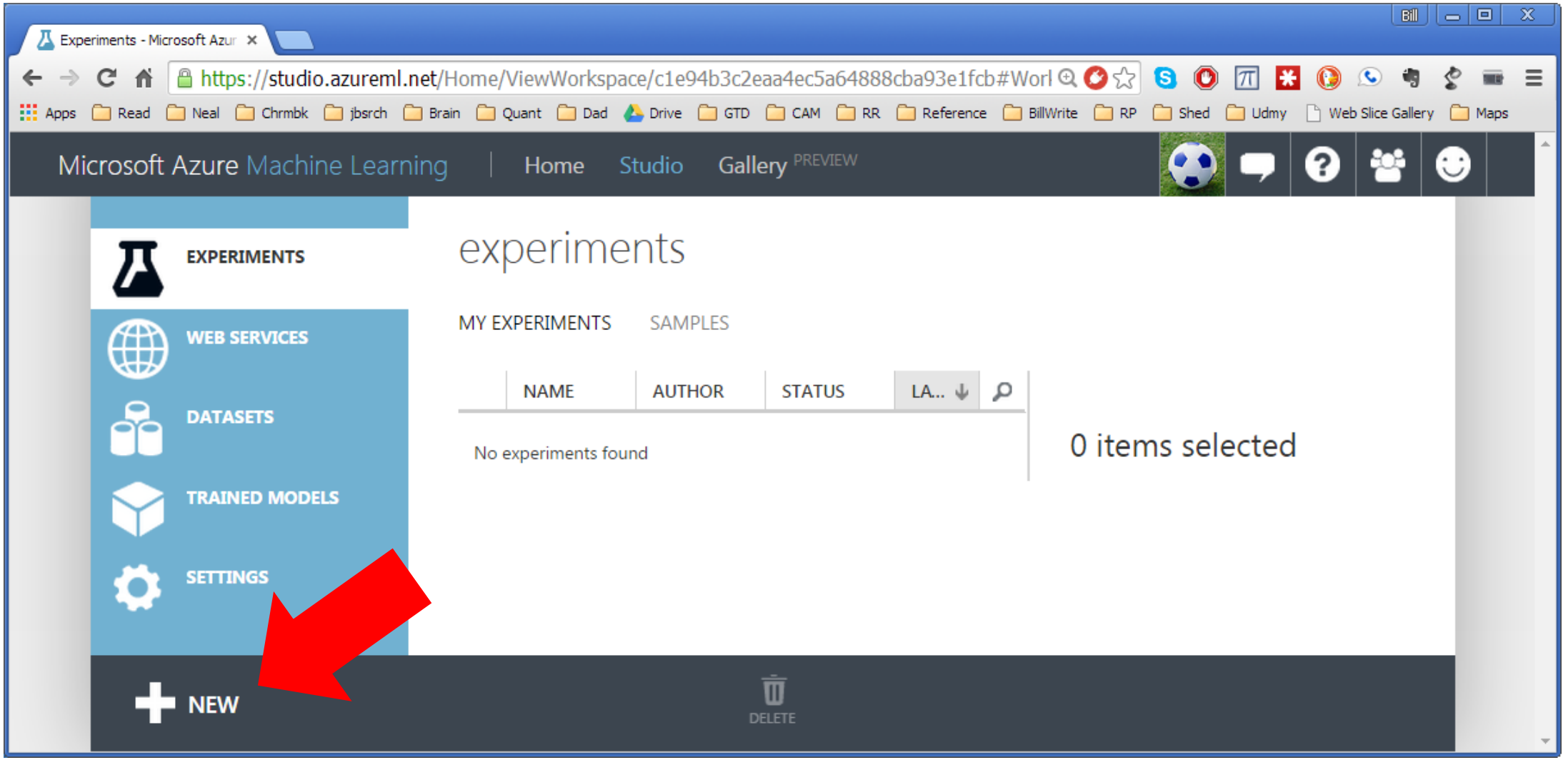


Step 3.1 : Create New Experiment

- AML modelling … a checklist approach
  - ❑ Create new experiment

Microsoft NEAL ANALYTICS

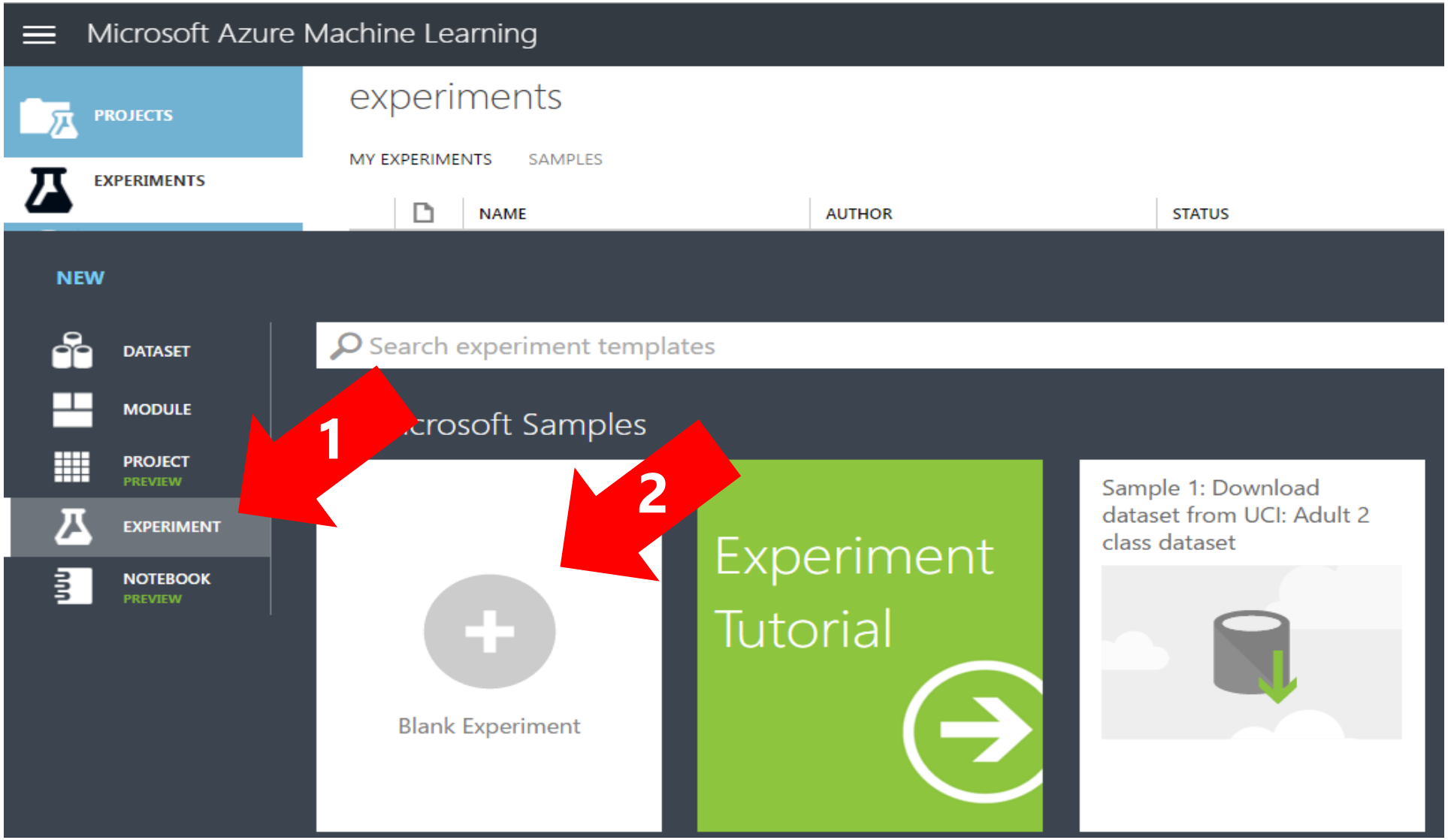
Step 3.2 : Create a new experiment



Microsoft

NEAL ANALYTICS

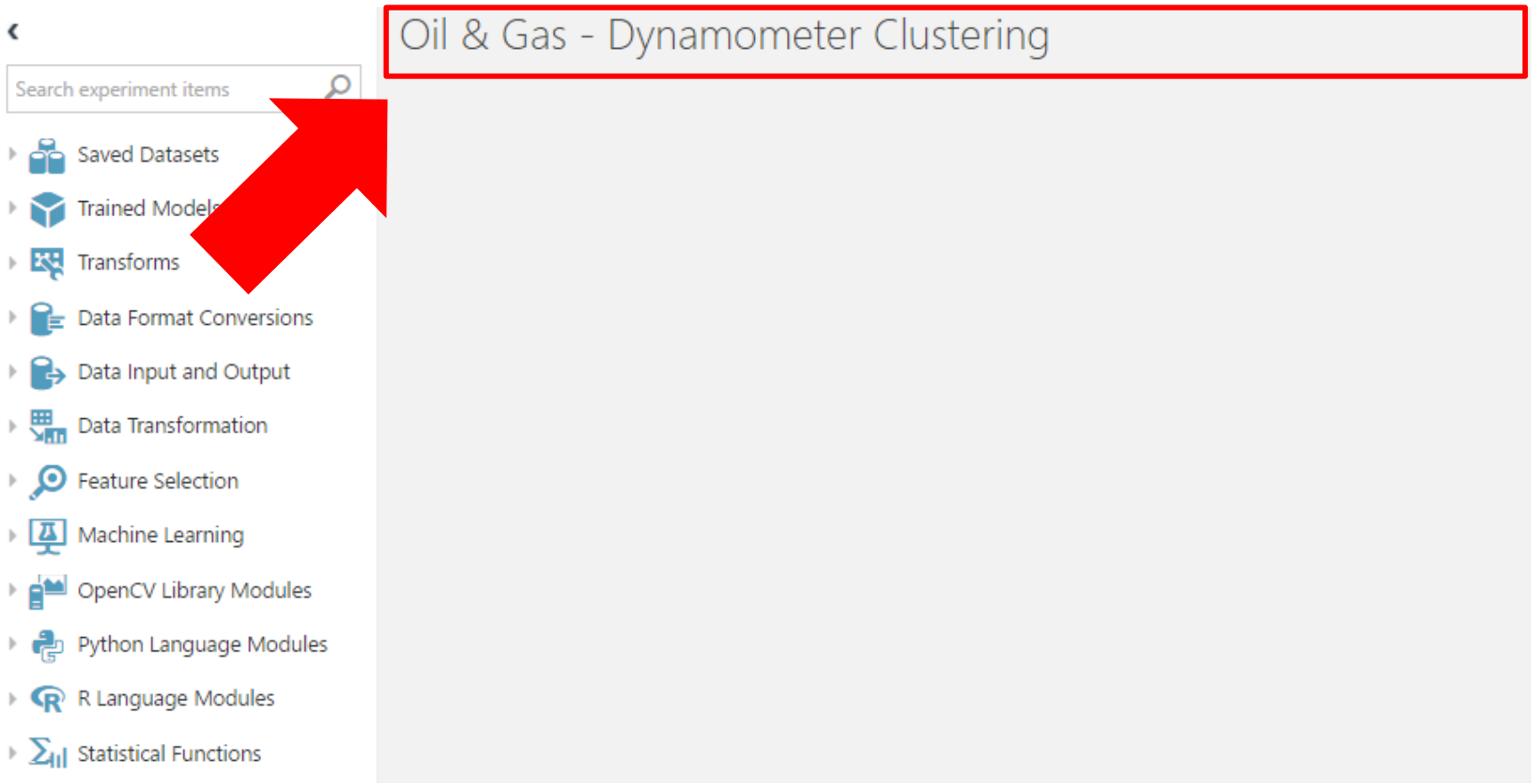
Step 3.3 : Create a new experiment



Microsoft

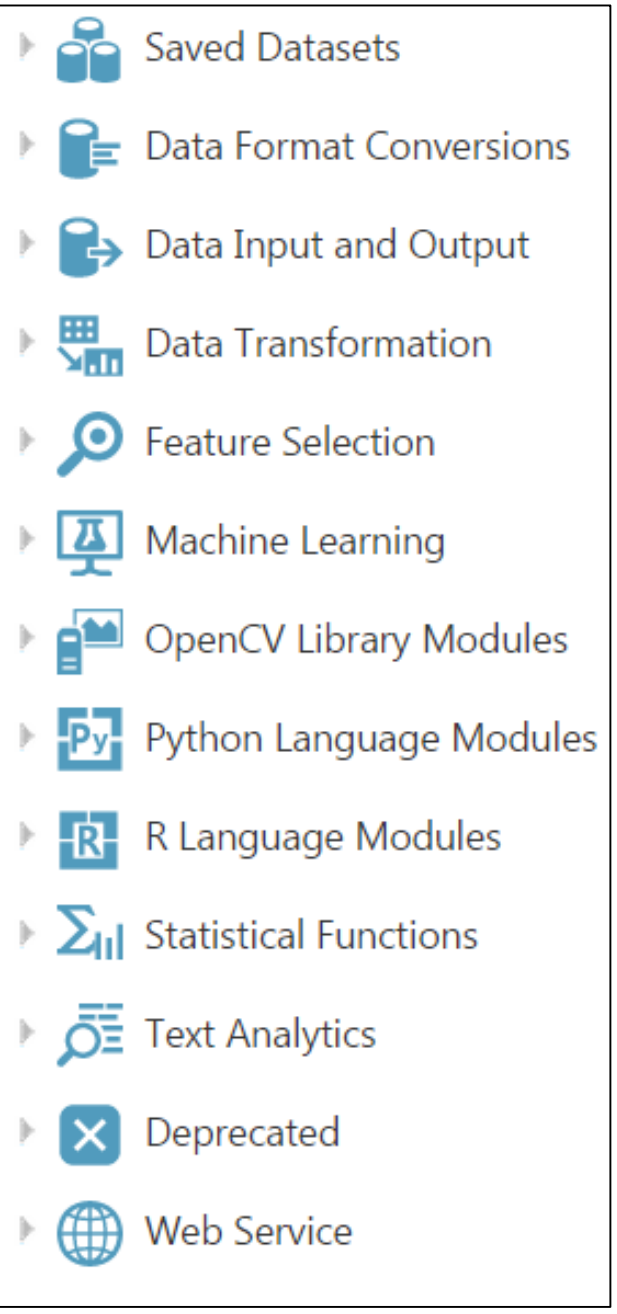
NEAL ANALYTICS

Step 3.4 : Name your experiment



- Click on the title box at the top that says "Experiment Created on ...."
- Give the experiment the following title: "Oil & Gas – Dynamometer Clustering"

Microsoft NEAL ANALYTICS



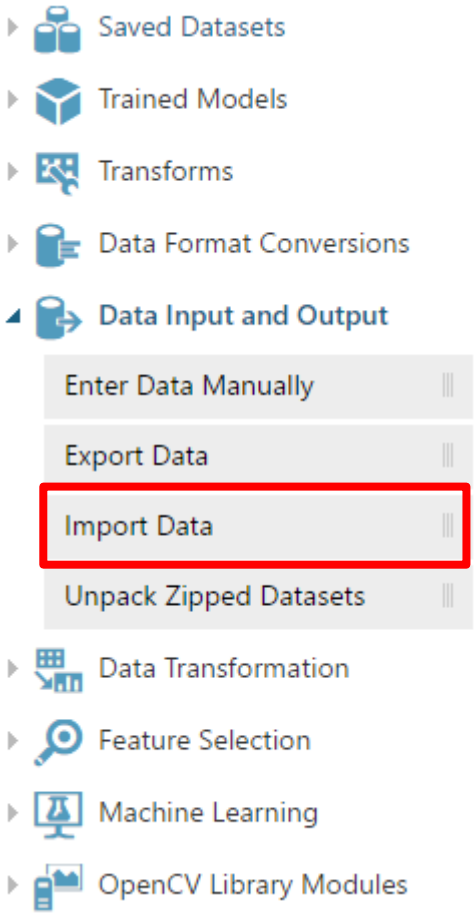
Step 4.1 : Import the dataset

- AML modelling … a checklist approach
  - ❑ ~~Create new experiment~~
  - ❑ Import data set

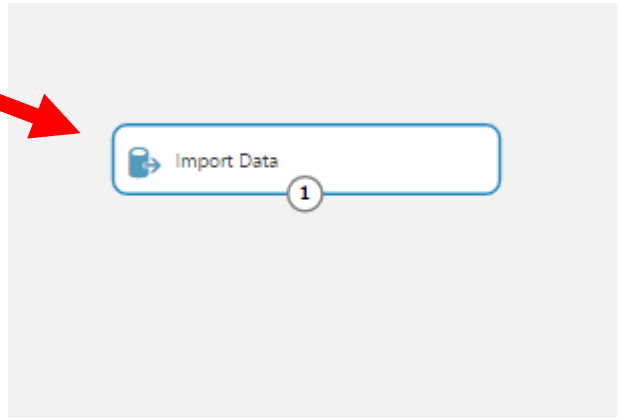
Microsoft NEAL ANALYTICS

Option 1 - Import the training dataset from the Azure Blob Storage. If option 1 did not work, skip to option 2.

Step 4.2 : Q? How to import data from Azure Blob?

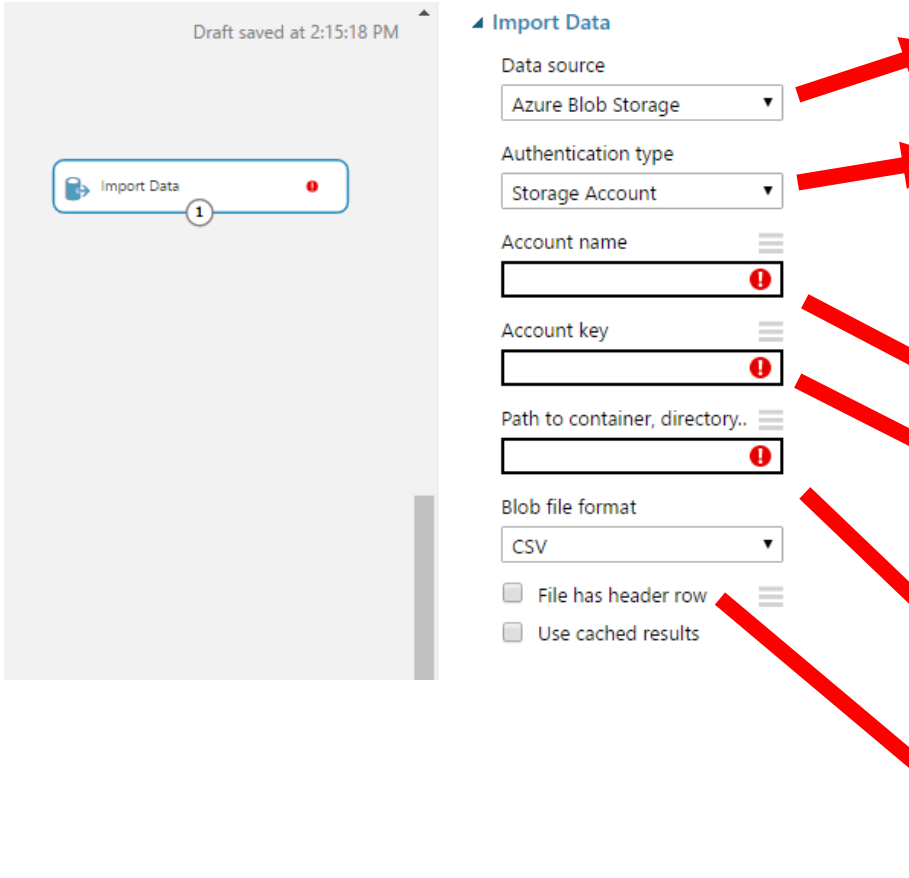


- Open “Data Input and Output” from the navigation pane at the left
- Drag “Import Data” to the canvas
  - “Import Data” loads data from sources such as the Web, Azure SQL, Windows Azure Blob storage, etc



Microsoft NEAL ANALYTICS

Step 4.3 : Q? How to import data from Azure Blob?



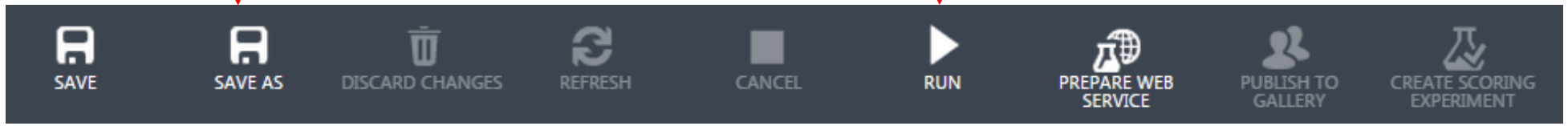
- Machine Learning
  - Click on Import Data
  - Chose "Azure Blob Storage" for Data source
  - Chose "Storage Account" for Authentication type
  - Copy and Paste the following information without the quotes in the Import Data module
    - Account Name – "**nealworkshop**"
    - Account key – "**RER9c7kfM1e67p7p7gI+TbkE5Y6aIzURg4PQc9lew+l8O+ZfU58gFjNgBW/WQm0u8N0YZQUG+wIaIzfWKxyIjA==**"
    - Path to container – **"/oilandgas/Dynamometer Card Reading Data PCA.csv**"
    - Check the File has header row check box

Microsoft NEAL ANALYTICS

Step 4.4 : Q? How to import data from Azure Blob?

Step 1: Click "Save"

Step 2: Click "Run"



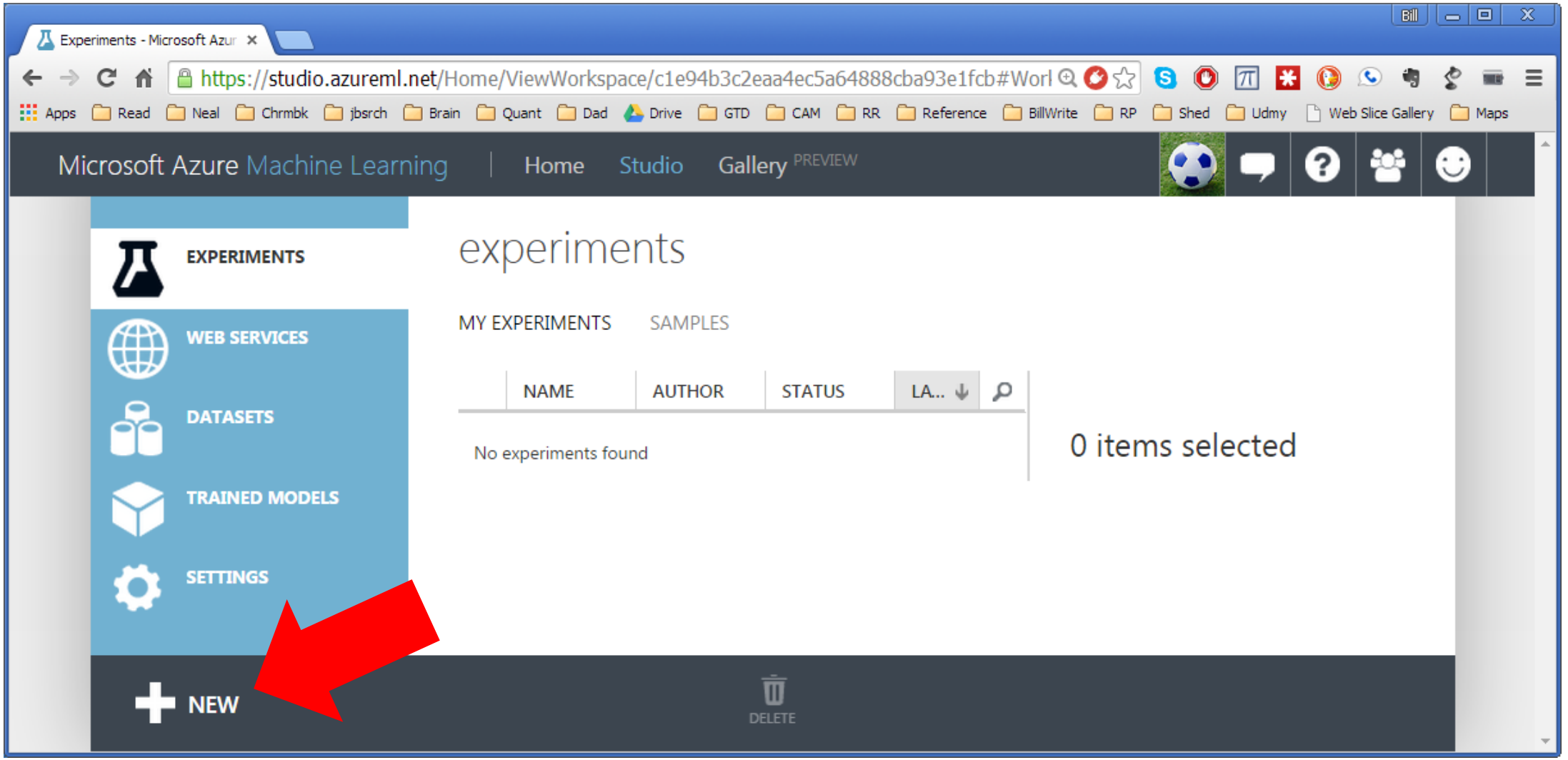
Microsoft

NEAL ANALYTICS

If Option 1 worked, move on to Step 4.12

Option 2 - Import the training dataset from the saved datasets.

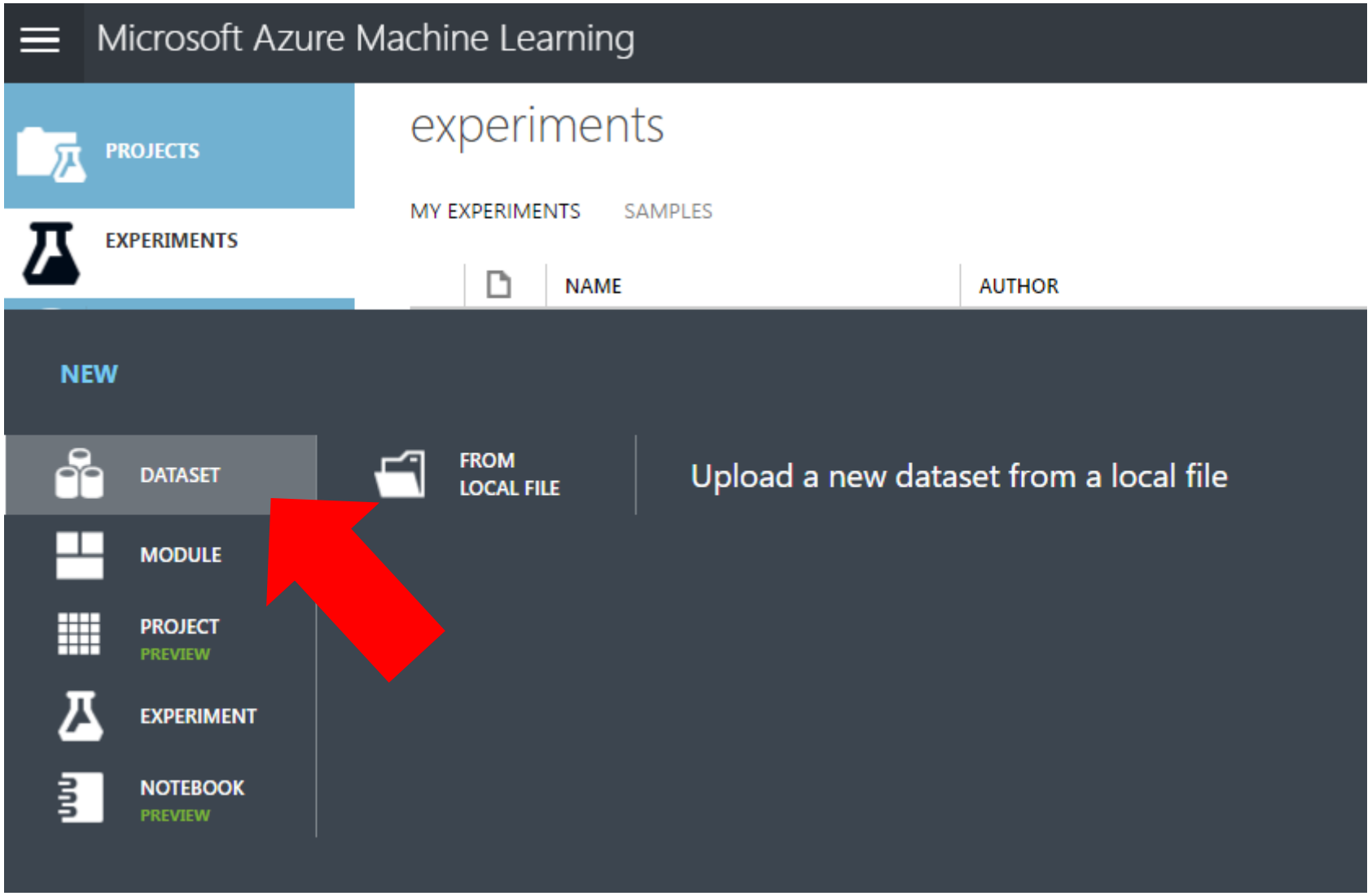
Step 4.5 : Import this tutorial's training dataset



Microsoft

NEAL ANALYTICS

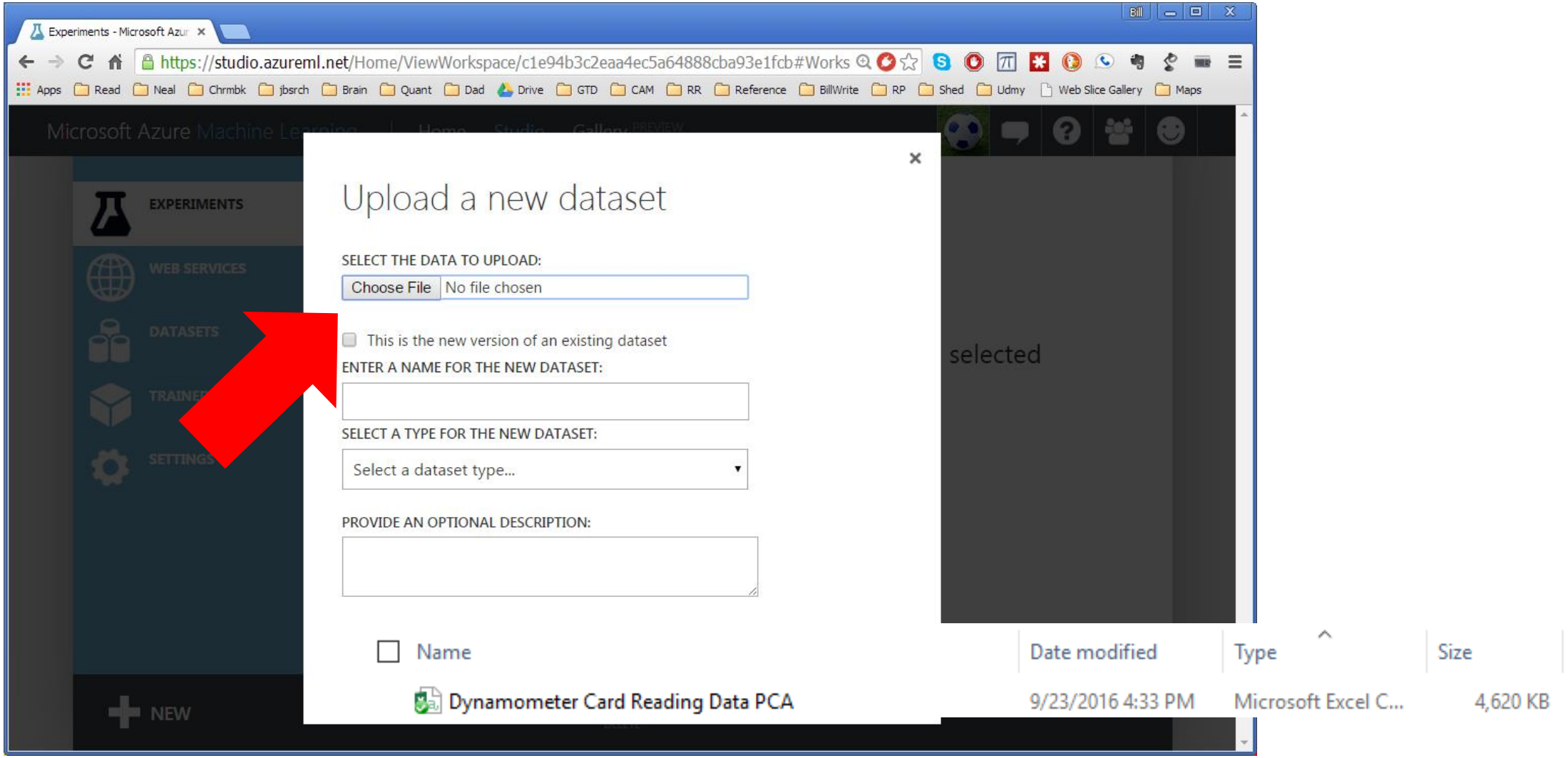
Step 4.6 : Import this tutorial's training dataset



Microsoft

NEAL ANALYTICS

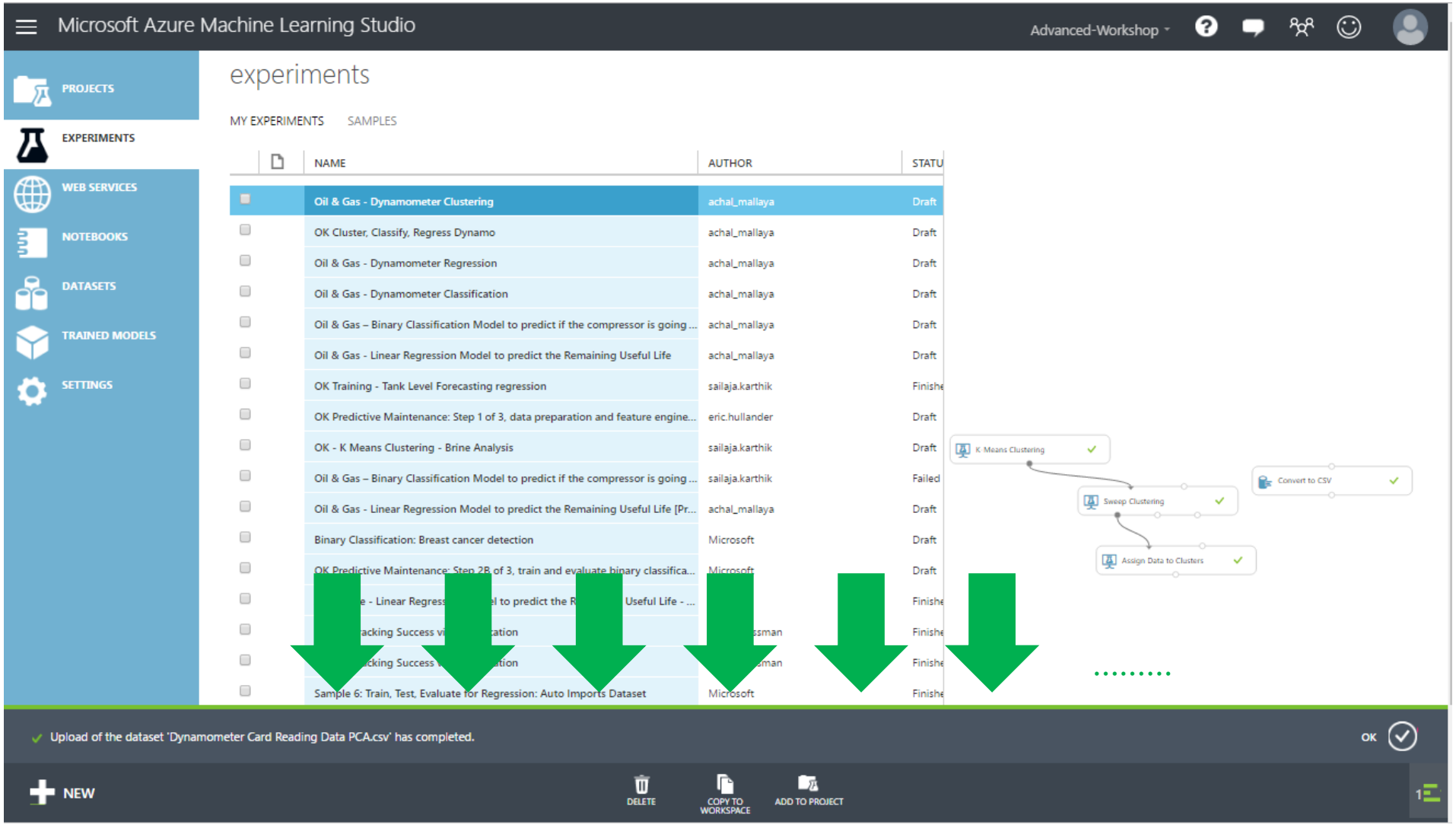
Step 4.7 : Import this tutorial's training dataset



Microsoft

NEAL ANALYTICS

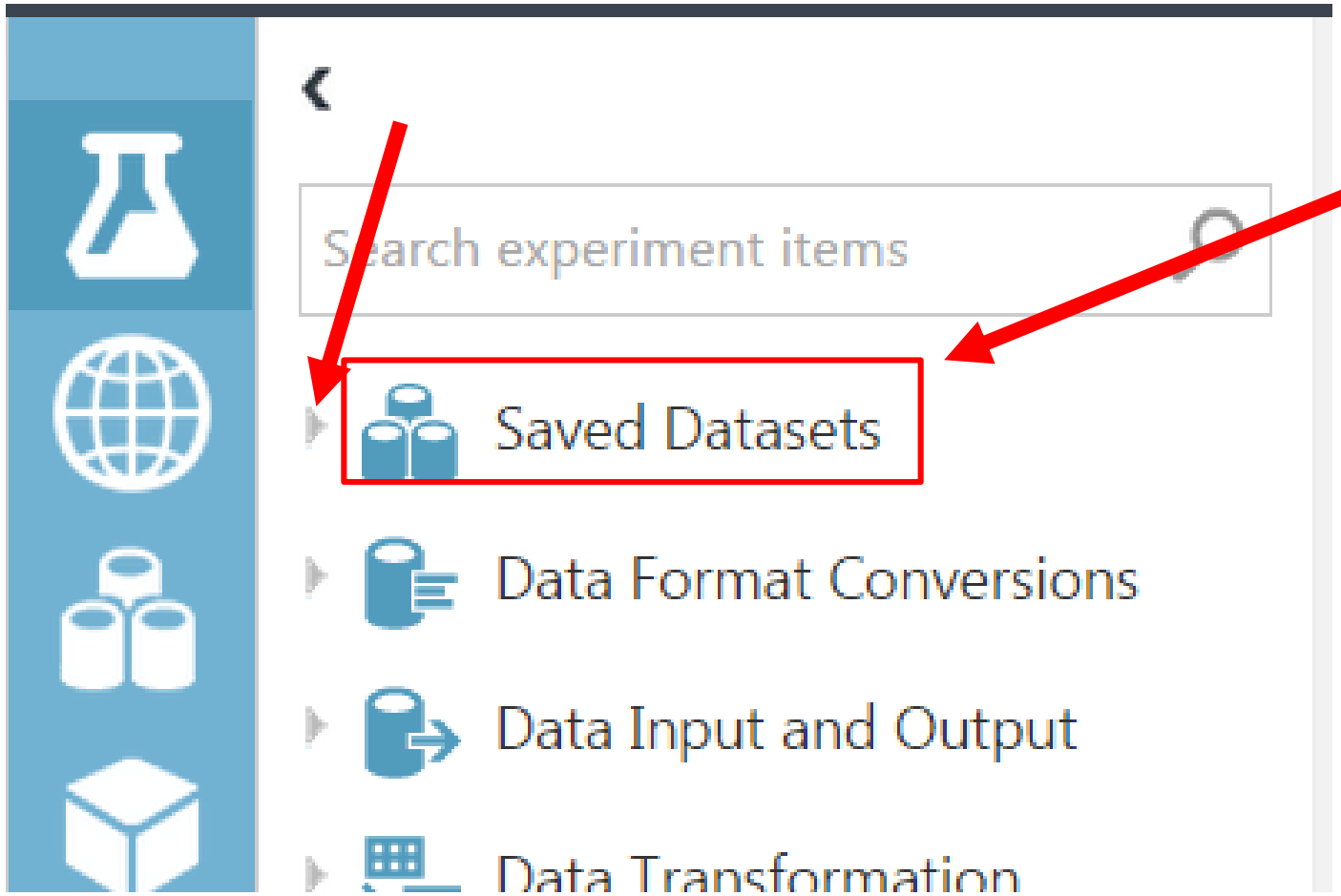
Step 4.8 : Import this tutorial's training dataset



Microsoft

NEAL ANALYTICS

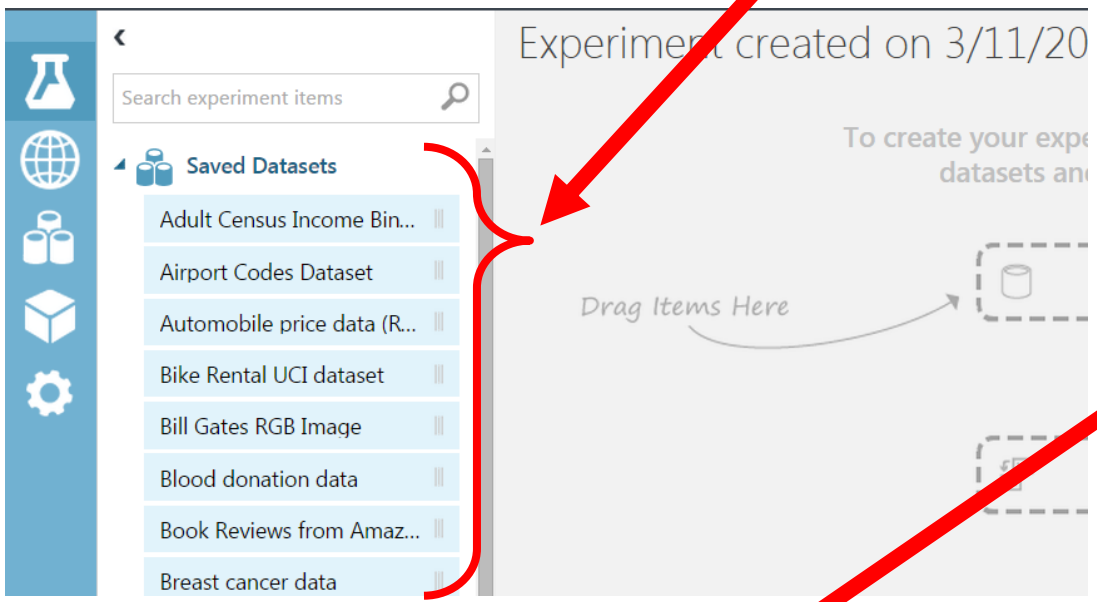
Step 4.9 : Open “Saved Datasets”



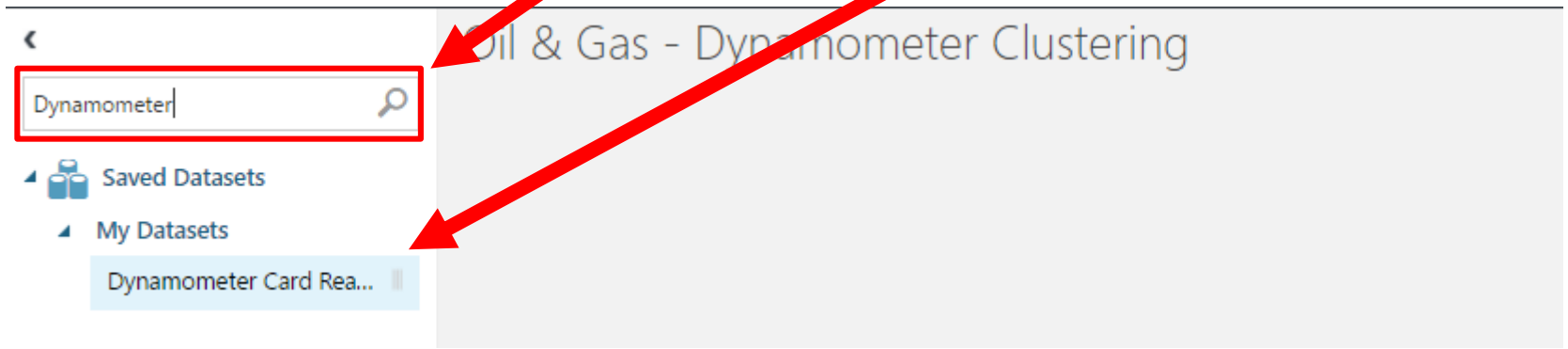
- By clicking on the triangle at the left
of “Saved Datasets”

Microsoft NEAL ANALYTICS

Step 4.10 : Take a second to notice the MANY datasets



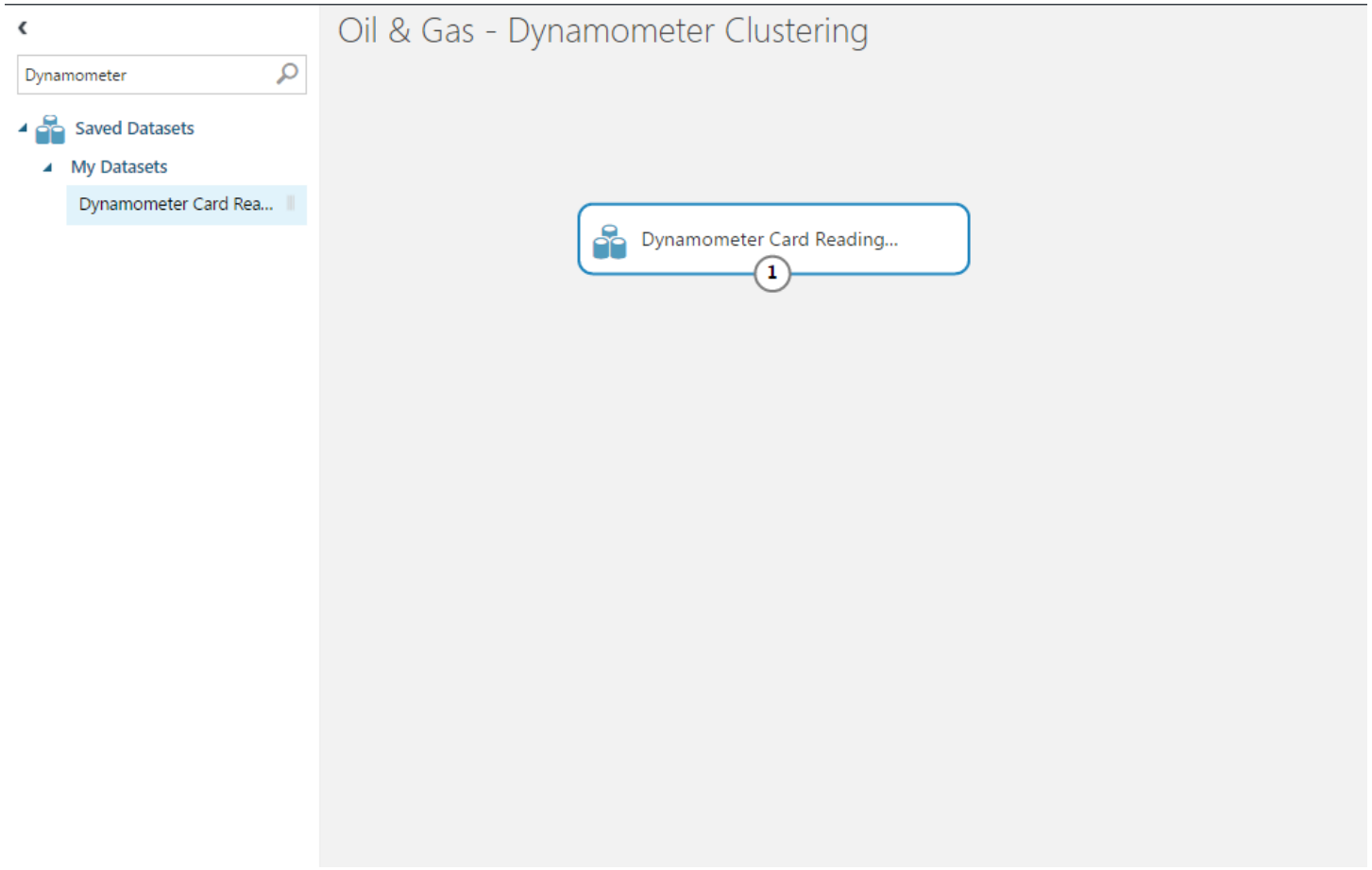
- To filter to the data set for this tutorial …
- Type “Dynamometer Card Reading” in the “Search experiment items” dialog box
- The data set list will reduce to our data set for this tutorial



Microsoft

NEAL ANALYTICS

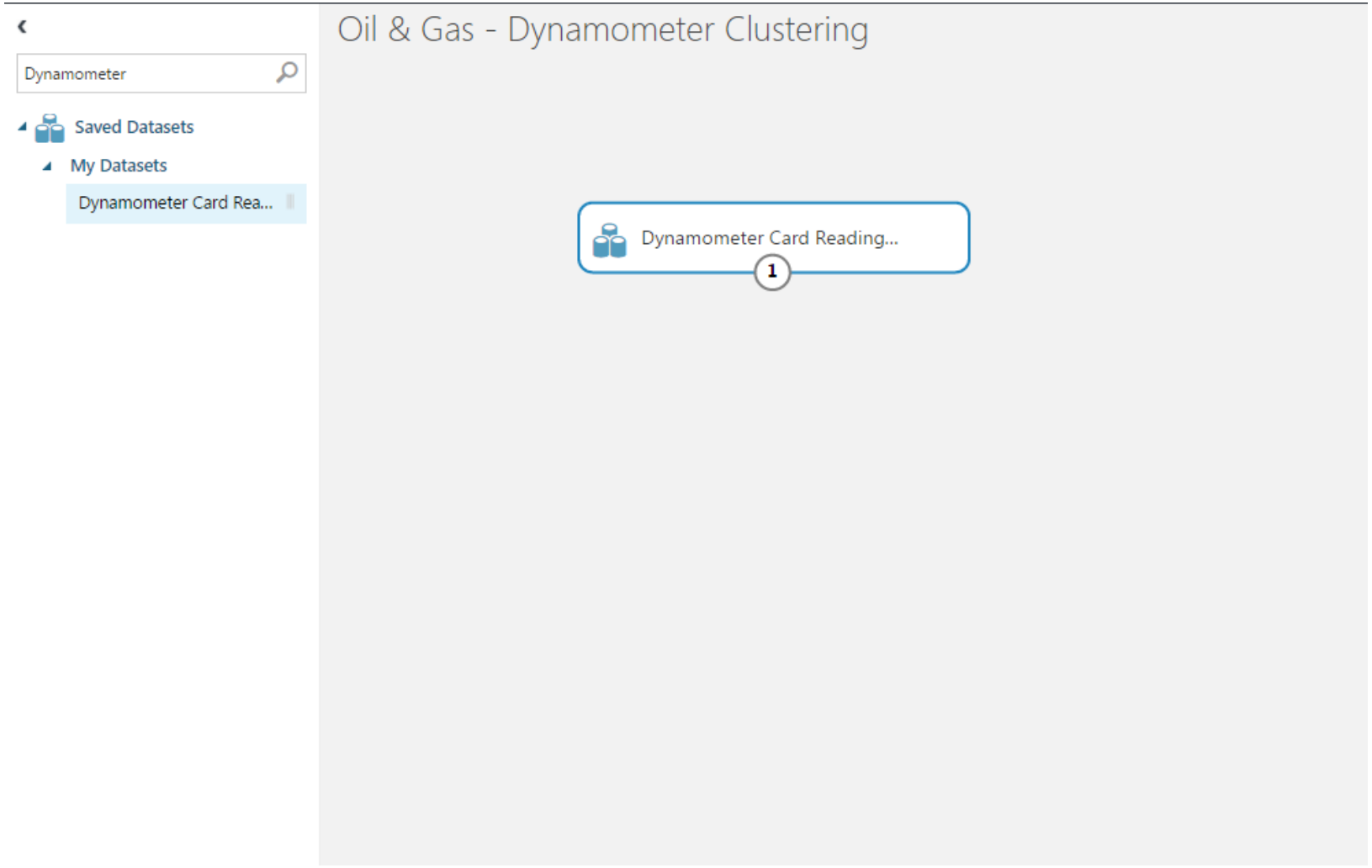
Step 4.11 : Drag the data set to the experiment



- *Note* when you drag the first element of your model to the canvas … all the guides disappear
- Now, where are all the tools that were at the left Azure Machine Learning?
- They are still there, … but we need to un-filter to see them

Microsoft NEAL ANALYTICS

Step 4.12 : Admire your data set living in your experiment

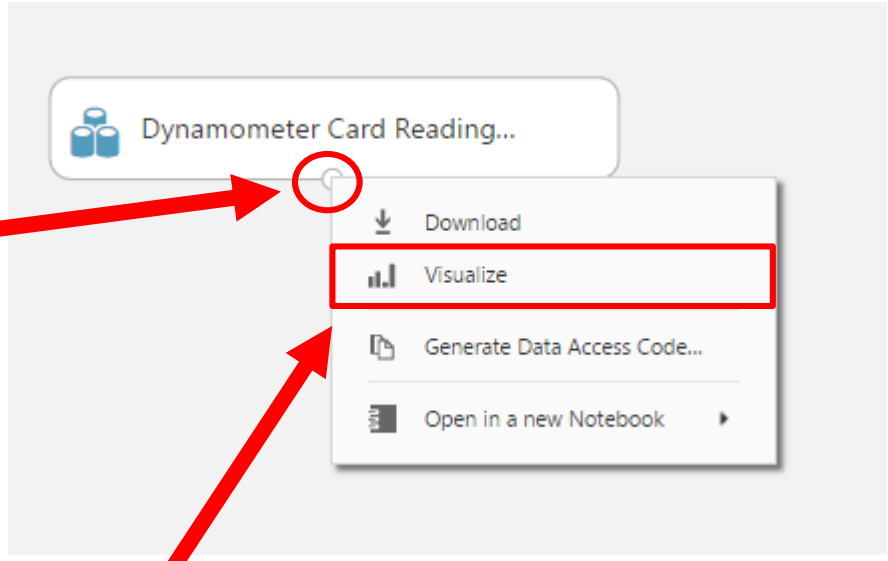


Microsoft

NEAL ANALYTICS

Step 4.13 : Visualize the data set

Right click on the bottom-middle circle of the module

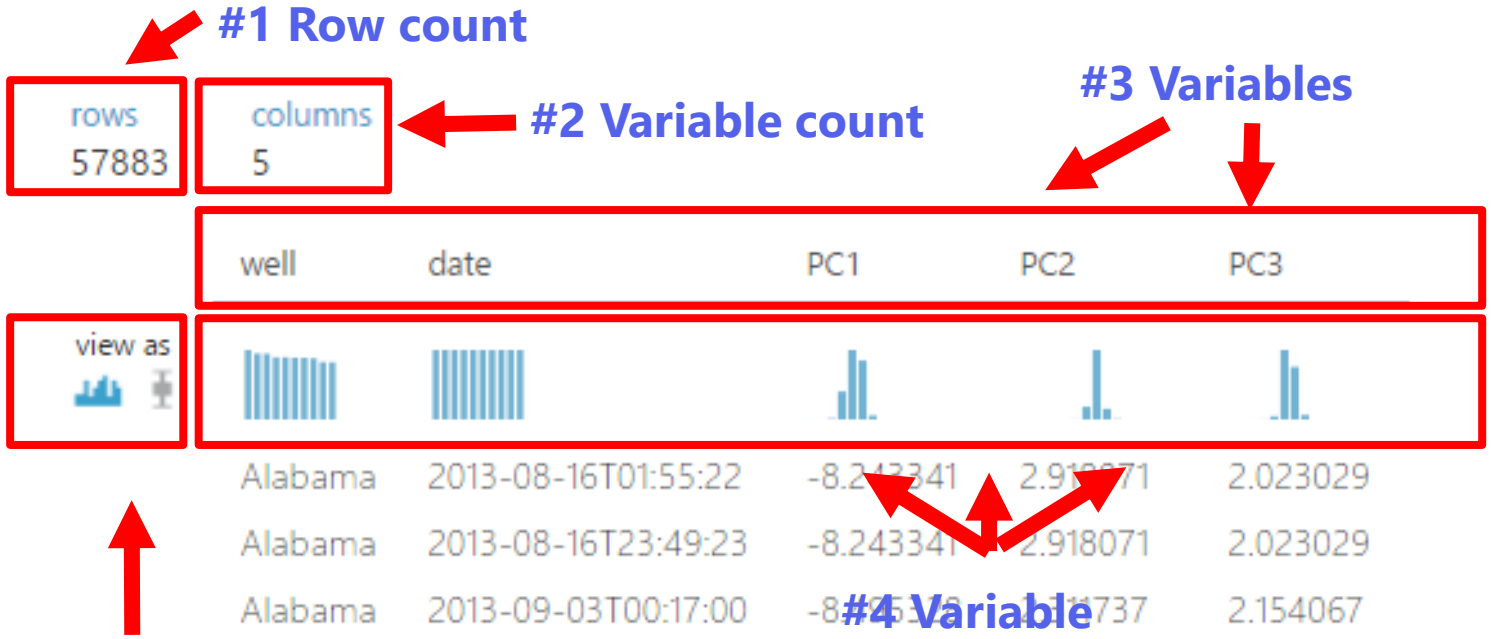


Then click “Visualize”

Microsoft

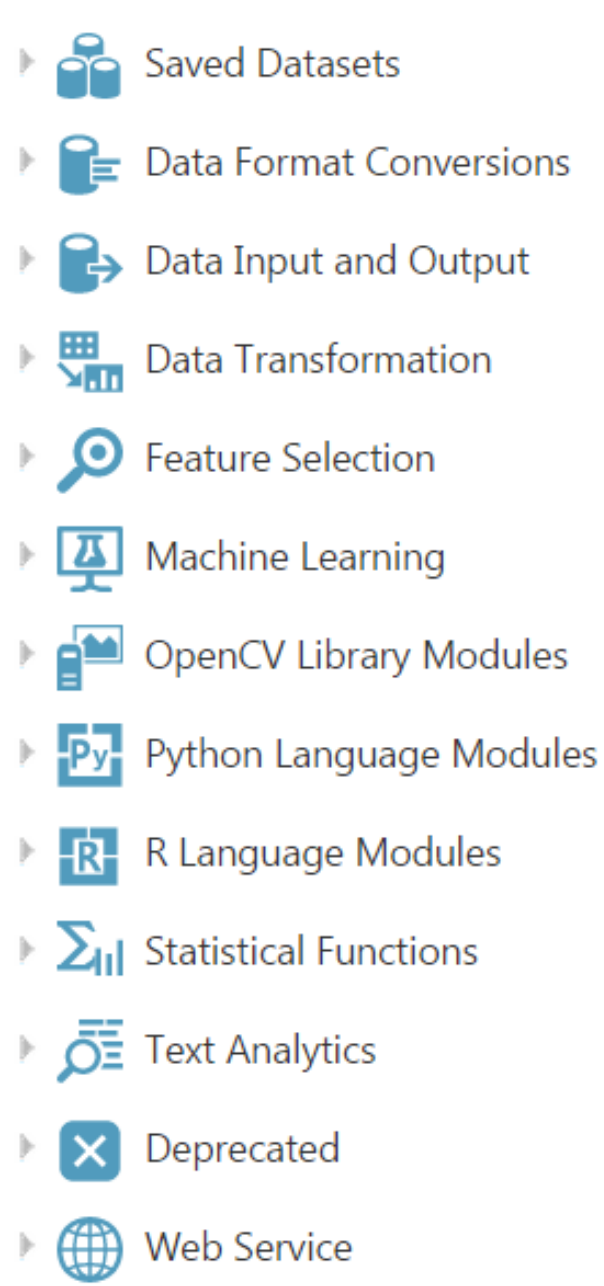
NEAL ANALYTICS

Step 4.14 : Now look at the data for 4 attributes



Microsoft NEAL ANALYTICS

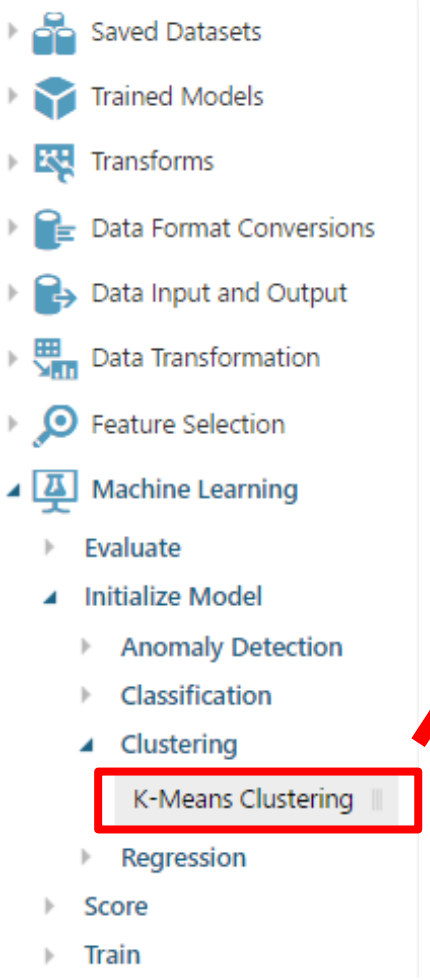
Step 5.1 : AML Modeling Components

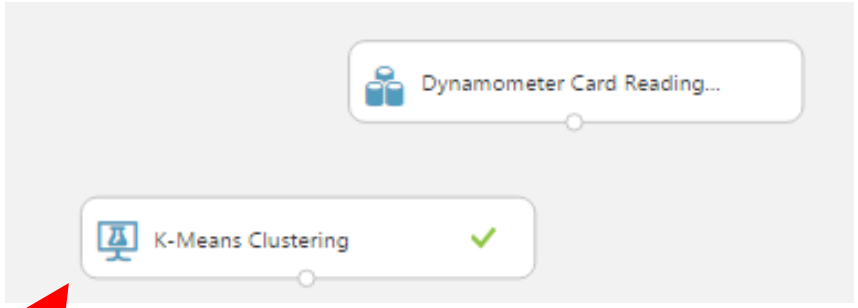


- AML modelling … a checklist approach
  - ~~Initial data set~~
  - ~~Create new experiment~~
  - ~~Import data set~~
  - ❑ Machine Learning

Microsoft NEAL ANALYTICS

Step 5.2 : Q? How to develop the Machine Learning Algorithm?

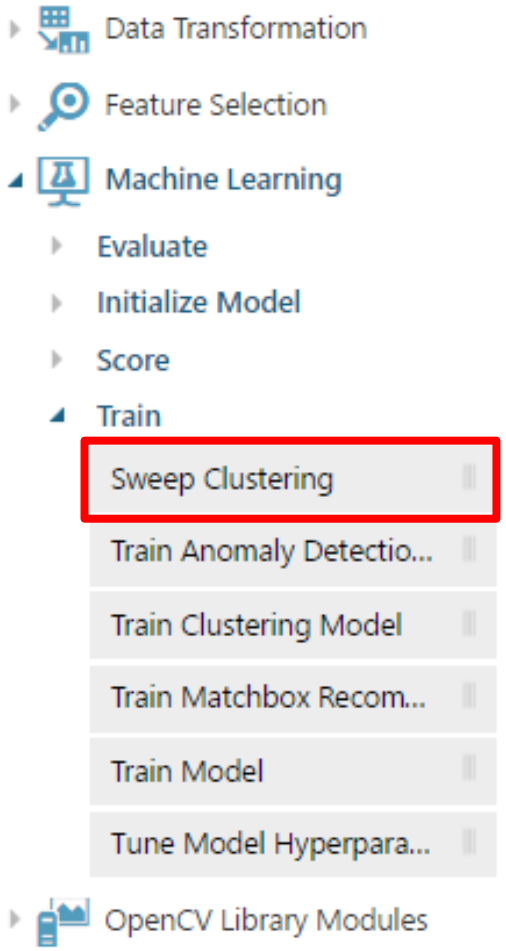


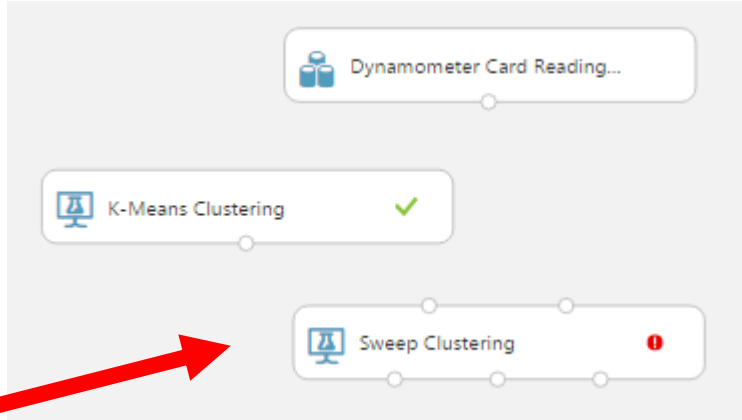


- Open “Machine Learning” -> “Initialize Model” -> “Clustering” from the navigation pane at the left
- Drag “K-Means Clustering” to the canvas

Microsoft NEAL ANALYTICS

Step 5.3 : Q? How to develop the Machine Learning Algorithm?

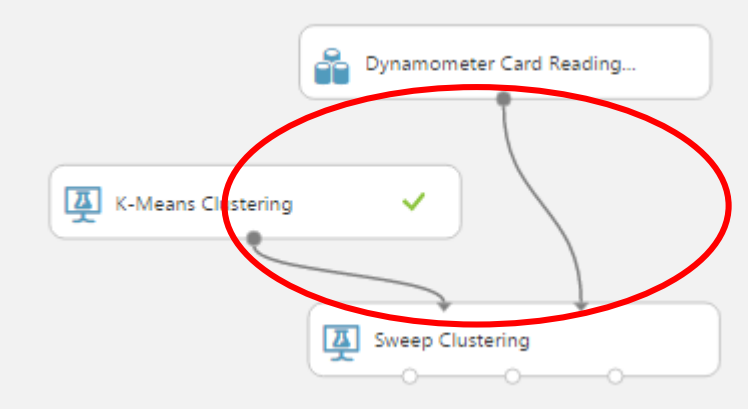




- Open “Machine Learning” and then “Train” from the navigation pane at the left
- Drag “Sweep Clustering” module to the canvas
  - “Train Clustering Model” performs a sweep on the clustering model to determine the optimum parameter settings

Microsoft NEAL ANALYTICS

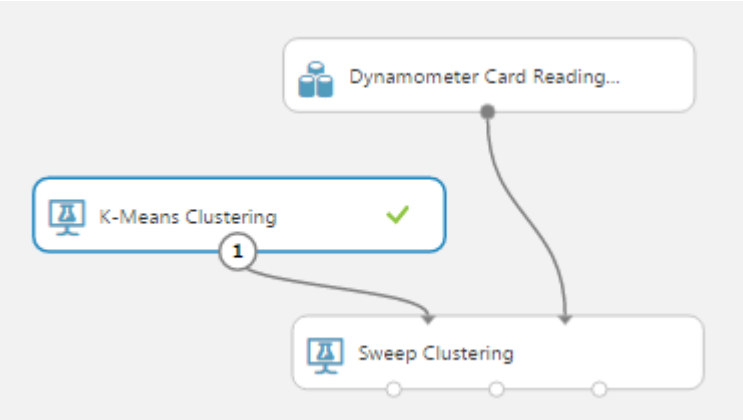
Step 5.4 : Q? How to develop the Machine Learning Algorithm?



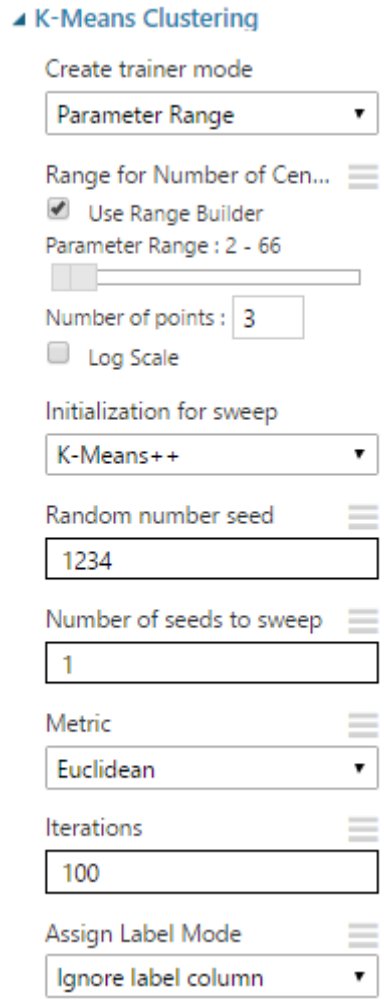
- Next, click and hold on the bottom middle circle of your “K-Means Clustering” module and drag the line to the top left circle of the “Sweep Clustering” module.
- Similarly, click and hold on the bottom left circle of your “Dataset” module and drag the line to the top right circle of the “Sweep Clustering” module.

Microsoft NEAL ANALYTICS

Step 5.5: Q? How to develop the Machine Learning Algorithm?



- Machine Learning
  - Initialize parameters for K-Means Clustering algorithm
  - Click on “K-Means Clustering” module on your canvas, and make sure it has a blue outline
  - A properties window will open on extreme right

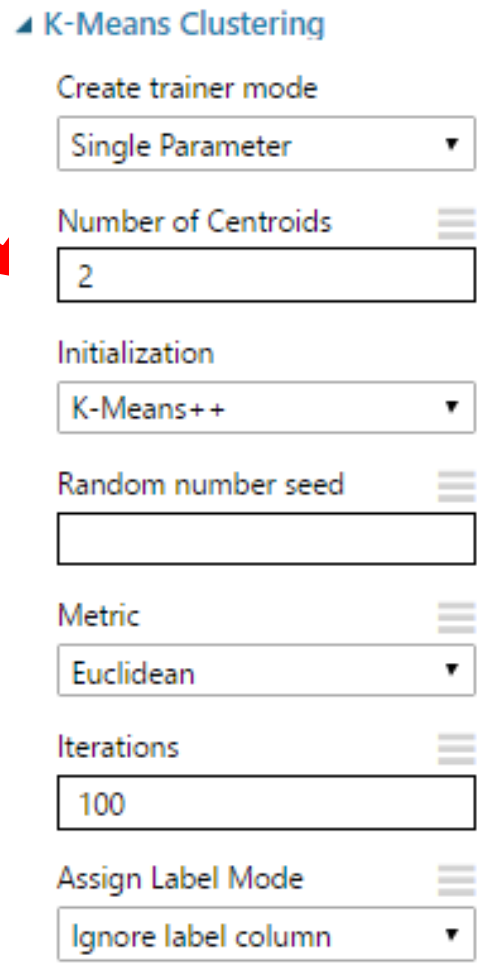


Microsoft NEAL ANALYTICS

Step 5.6: Q? How to develop the Machine Learning Algorithm?

- Machine Learning
  - ~~Click on K-Means Module and a properties window will open on extreme right~~
  - Initialize parameters

Default Values



Number of Clusters needed

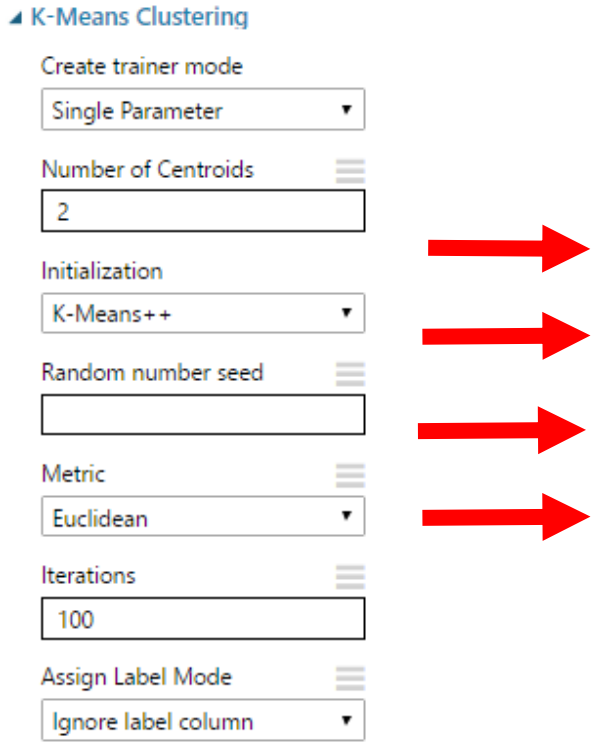
Initialization – Algorithm used for initial cluster configuration

Metric – Function used to measure distance between cluster vectors

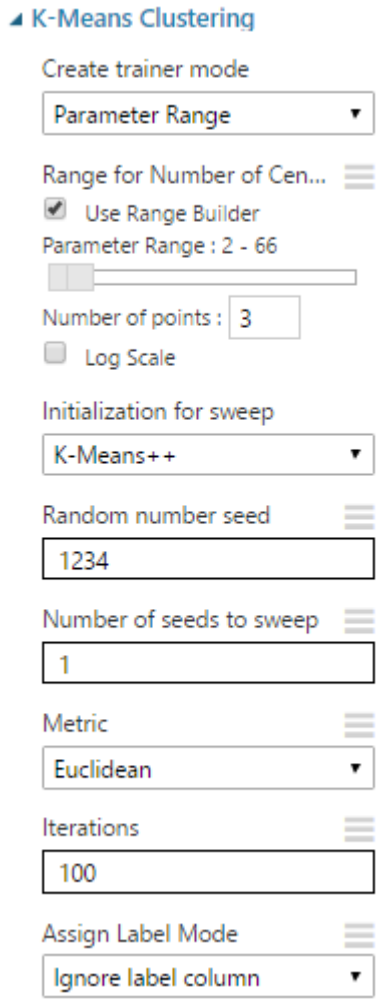
Iterations – Number of times the allocation is retested

Microsoft NEAL ANALYTICS

Step 5.7 : Q? How to develop the Machine Learning Algorithm



Assign Values

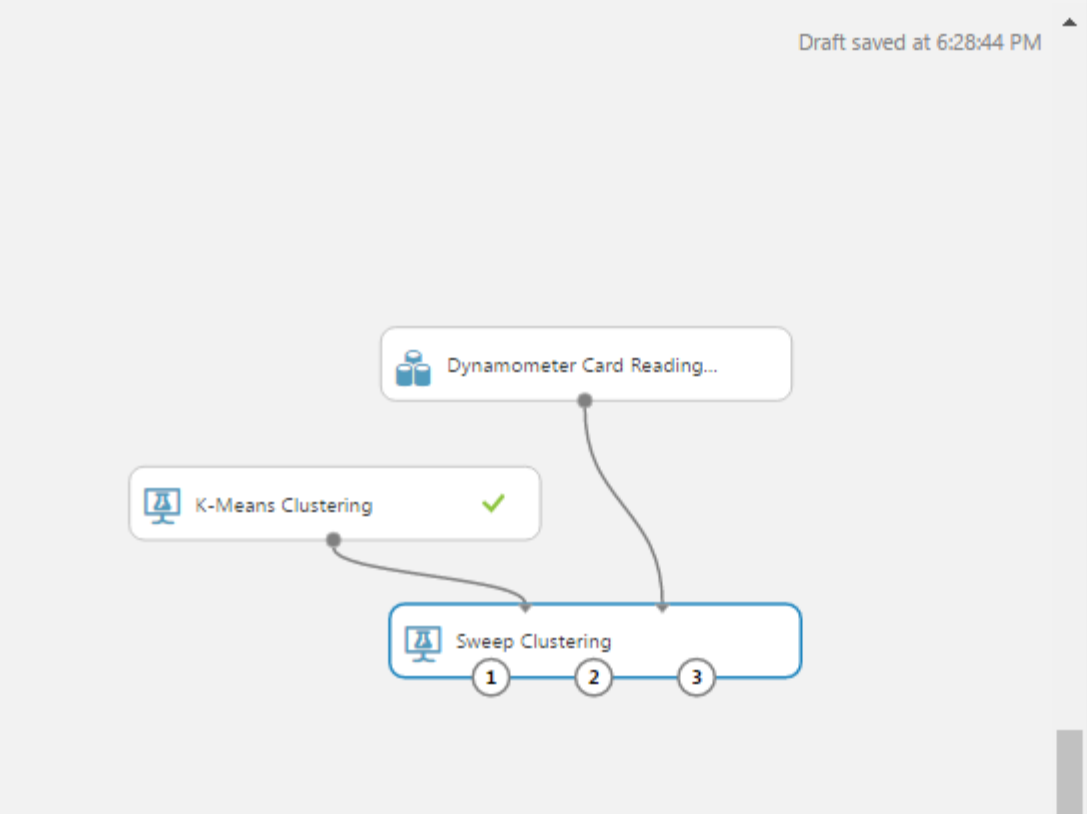


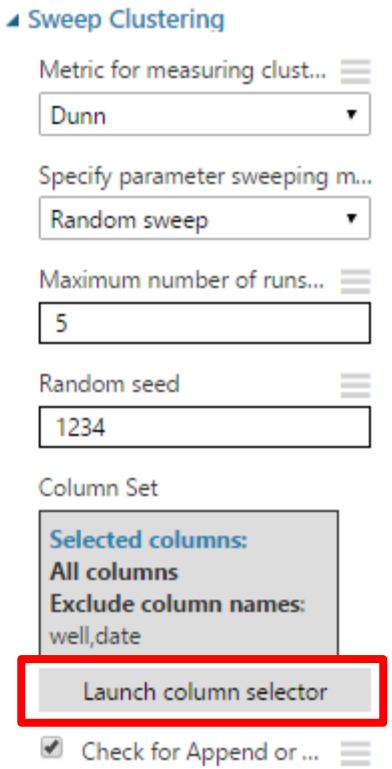
- Machine Learning
  - ~~Click on K-Means Module and a properties window will open on extreme right~~
  - Initialize parameters

Starting with 4 Clusters and rest as default options.

Microsoft NEAL ANALYTICS

Step 5.8 : Q? How to develop the Machine Learning Algorithm

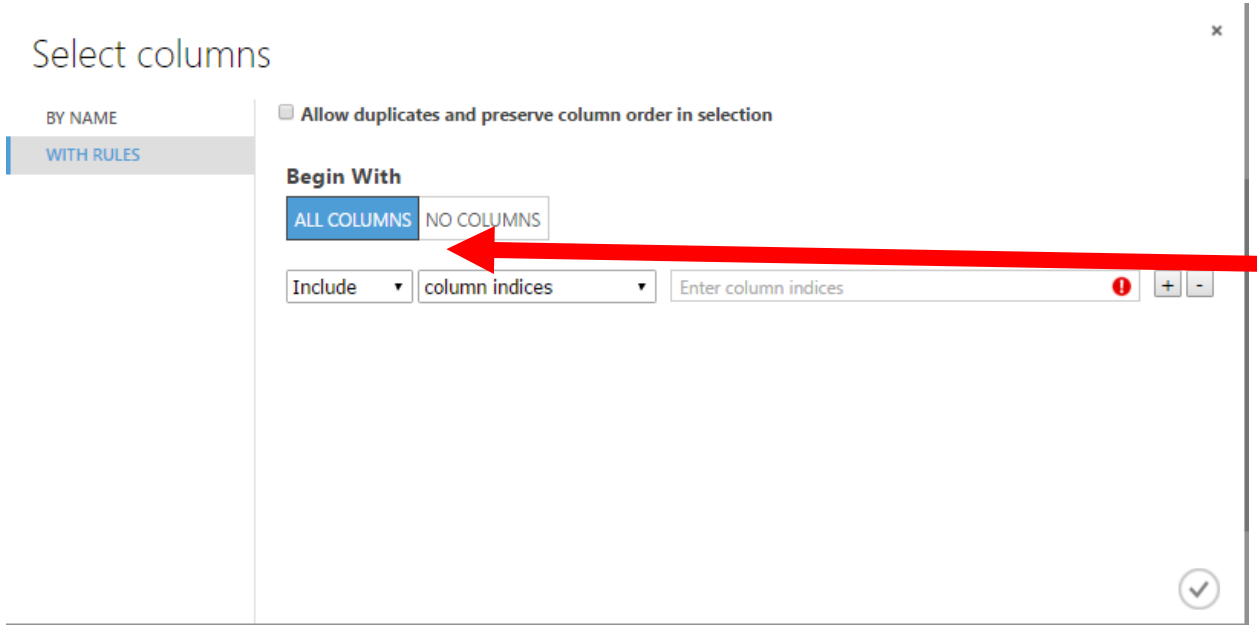




- Machine Learning
  - Click on Sweep Cluster
  - Set Metric for measuring clustering to “Dunn”
  - Set the Random seed to “1234”
  - Then click on “Launch column selector”

Microsoft NEAL ANALYTICS

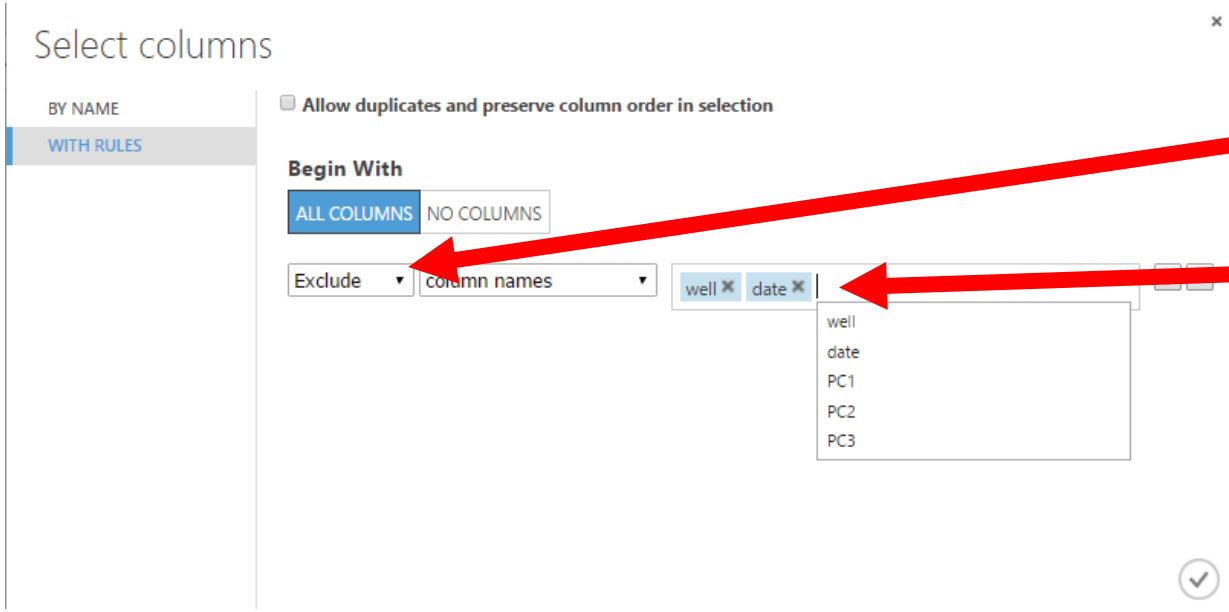
Step 5.9 : Q? How to develop the Machine Learning Algorithm



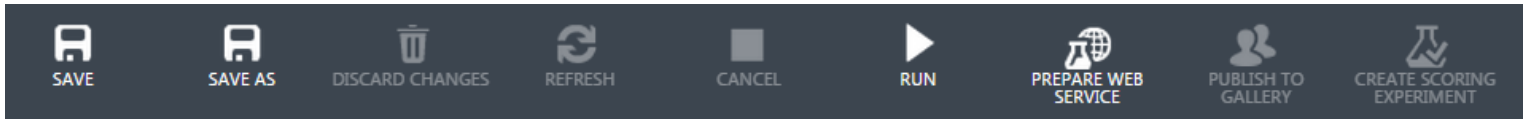
- Machine Learning
  - Select “With Rules”
  - Select “All Columns” from list ***Begin with***

Microsoft NEAL ANALYTICS

Step 5.10 : Q? How to develop the Machine Learning Algorithm

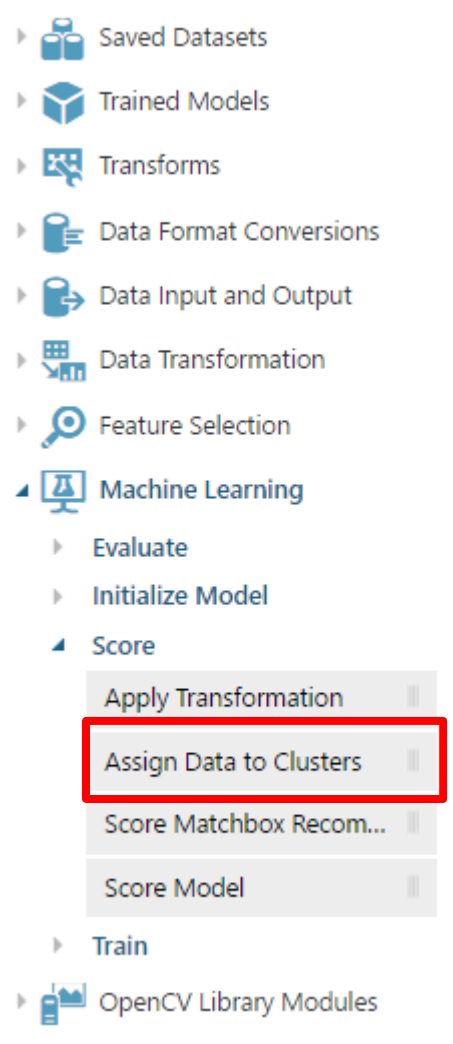


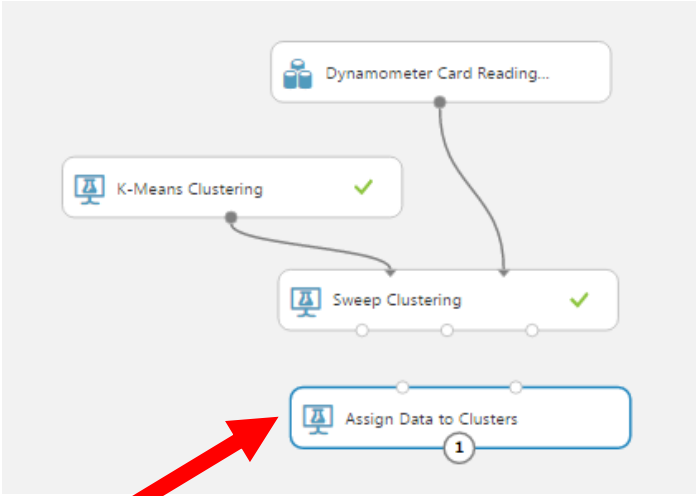
- Machine Learning
  - Select Exclude from the dropdown
  - Chose the following columns:
    - well
    - date
  - Then click on Check sign to close window
  - Click “Run”



Microsoft NEAL ANALYTICS

Step 5.11 : Q? How to develop the Machine Learning Algorithm?

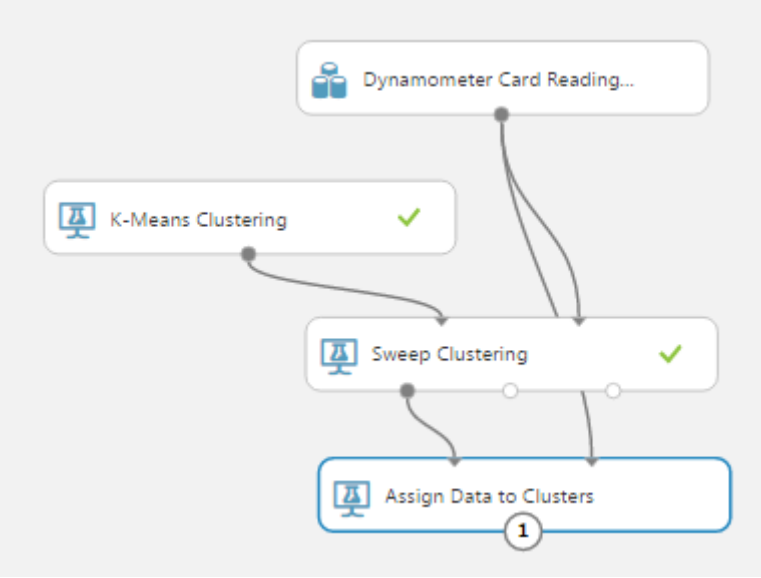




- Open “Machine Learning” and then “Score” from the navigation pane at the left
- Drag “Assign Data to Clusters” module to the canvas
  - “Assign Data to Clusters” assigns data to clusters using an existing trained clustering model

Microsoft NEAL ANALYTICS

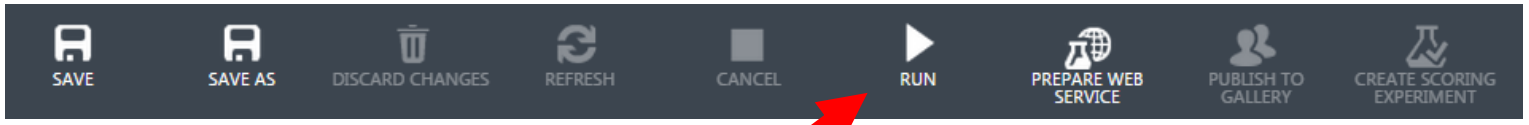
Step 5.12 : Q? How to develop the Machine Learning Algorithm?



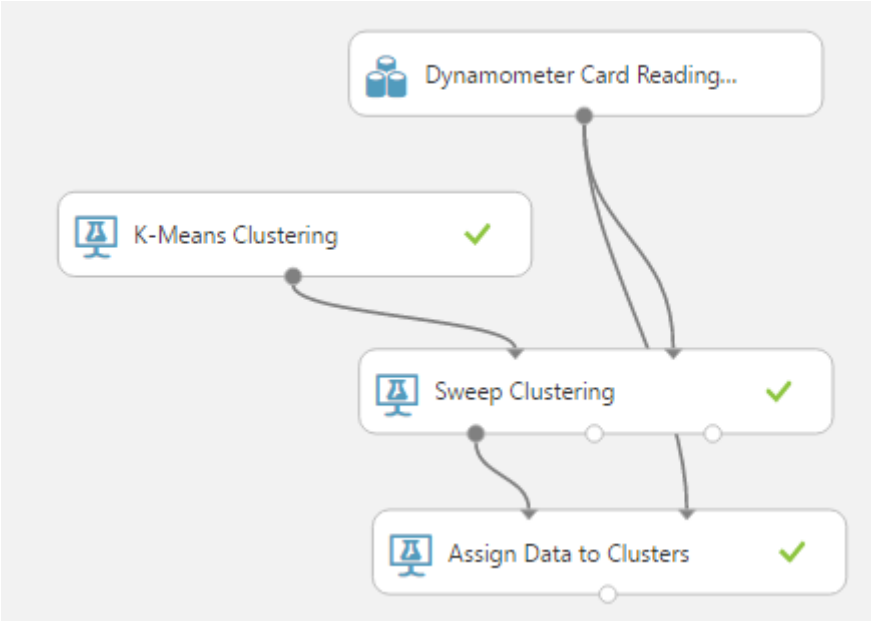
- Next, click and hold on the bottom left circle of your “Sweep Clustering” module and drag the line to the top left circle of the “Assign Data to Clusters” module.
- Similarly, click and hold on the bottom middle circle of your “Dataset” module and drag the line to the top right circle of the “Assign Data to Clusters” module.

Microsoft NEAL ANALYTICS

Step 5.13 : Run the Experiment



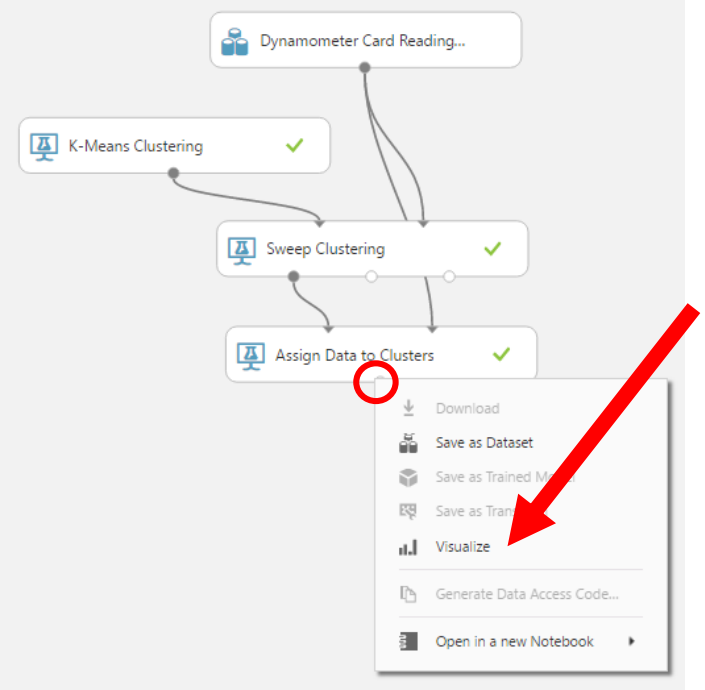
Click “Run”



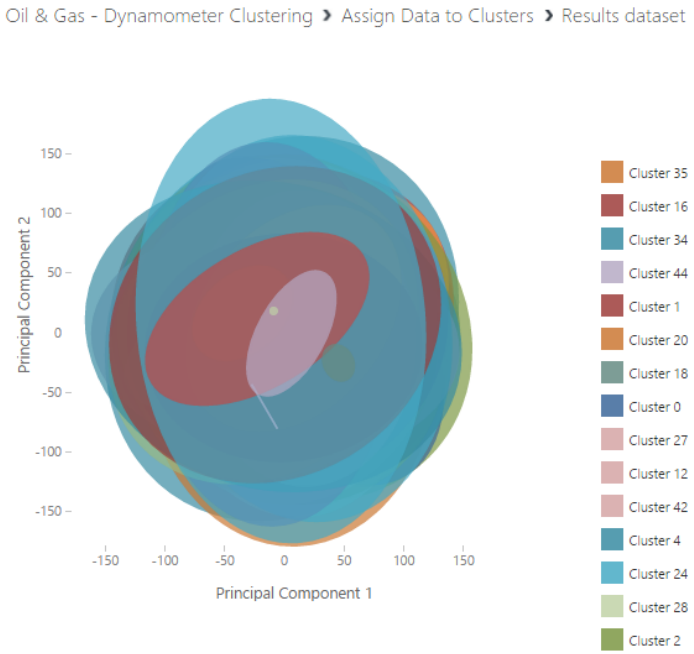
Microsoft

NEAL ANALYTICS

Step 5.14 : Visualize the Results



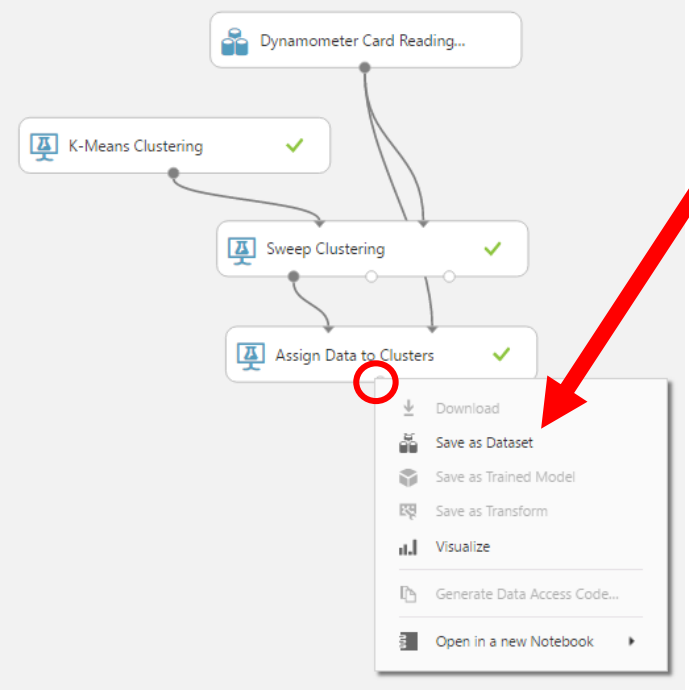
- The Clustering algorithm has assigned a cluster across each item given by variable “Assignments”
- Right click on the bottom-middle circle of “Assign Data to Clusters” module and select “Visualize”



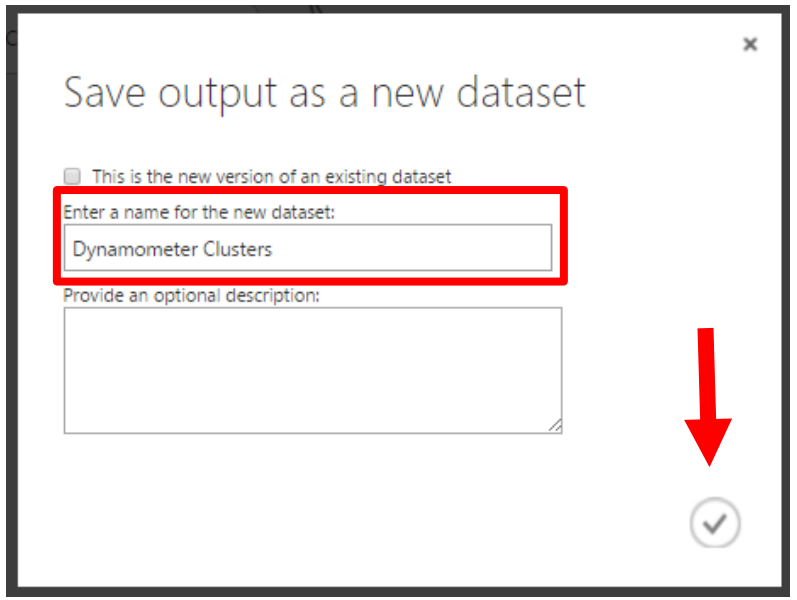
Microsoft

NEAL ANALYTICS

Step 5.15 : Save the Result as a Dataset

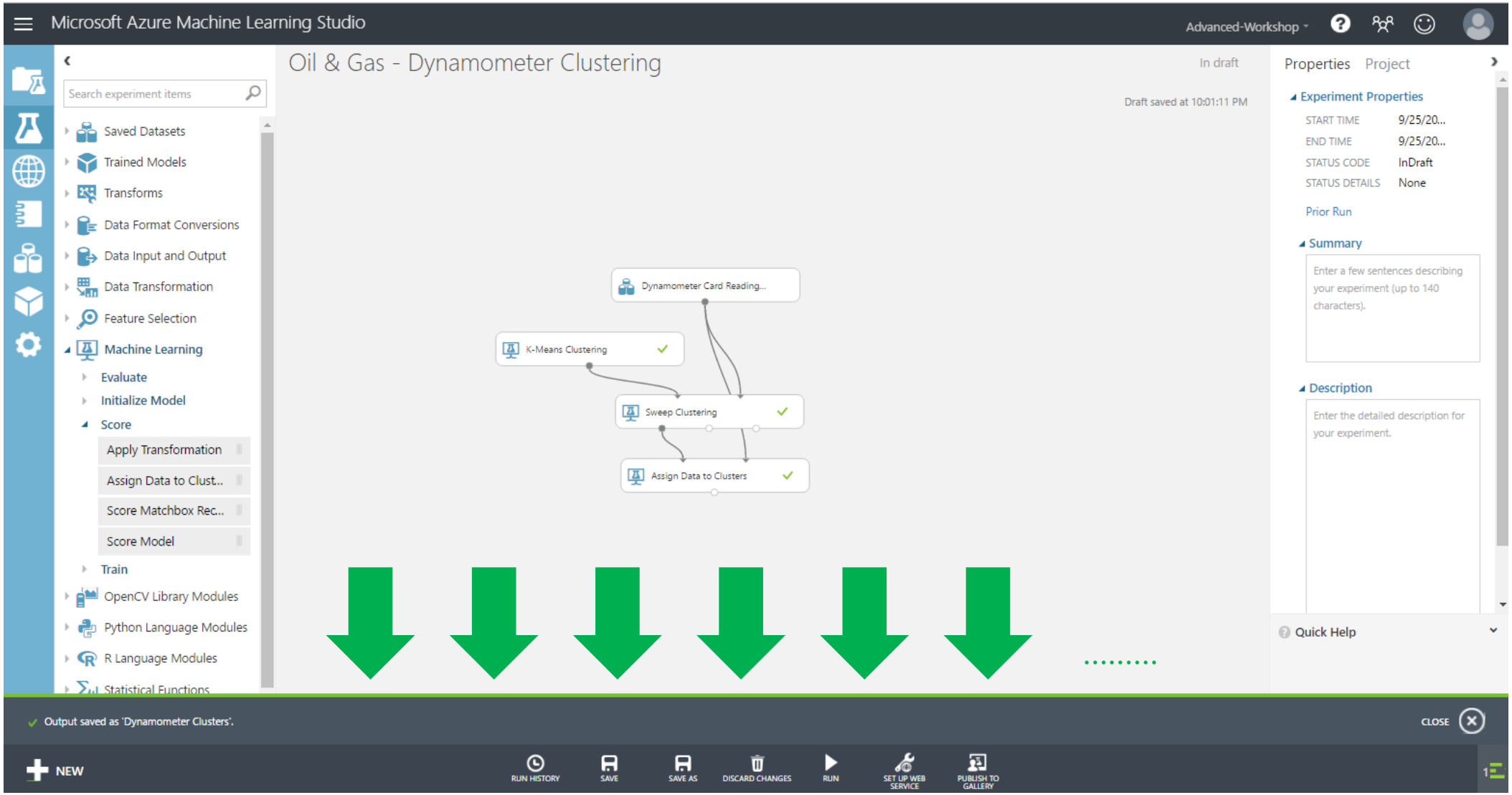


- Right click on the bottom-middle circle of “Assign Data to Clusters” module and select “Save as Dataset”
- A pop-up will appear. Call the new dataset “Dynamometer Clusters”
- Click on the check mark



Microsoft NEAL ANALYTICS

Step 5.16 : Save the Result as a Dataset



Microsoft

NEAL ANALYTICS