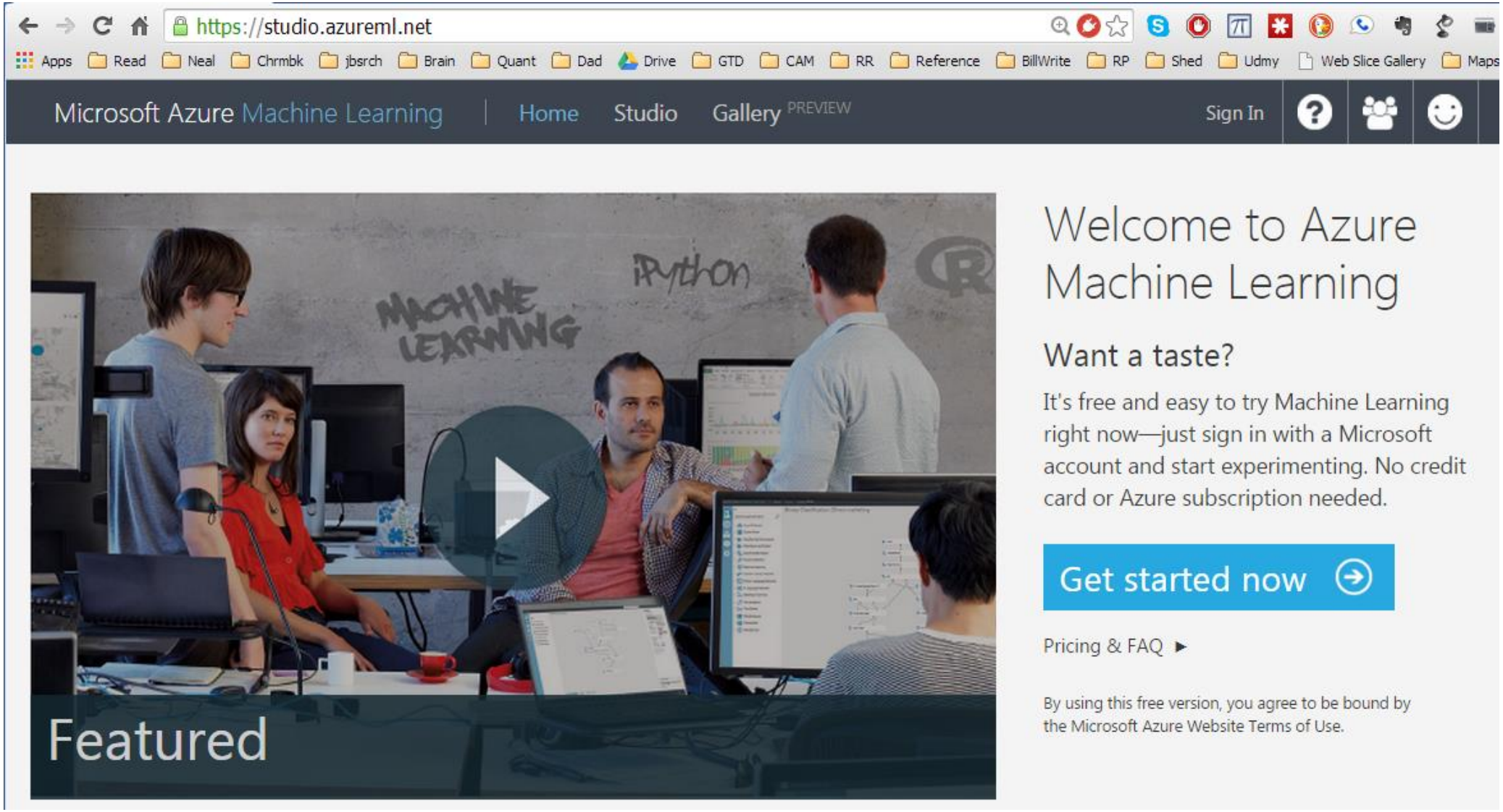
Import Data in Azure ML

Exploratory Data Analysis – Lab Session

- Introduction to Notebooks
- Setting up and loading data into Notebook
- Data descriptive
- Data visualizations
- Univariate and Bi Variate plots
- Correlation
- Principal Component Analysis – dimensionality reduction
- Feature Engineering

Microsoft NEAL ANALYTICS

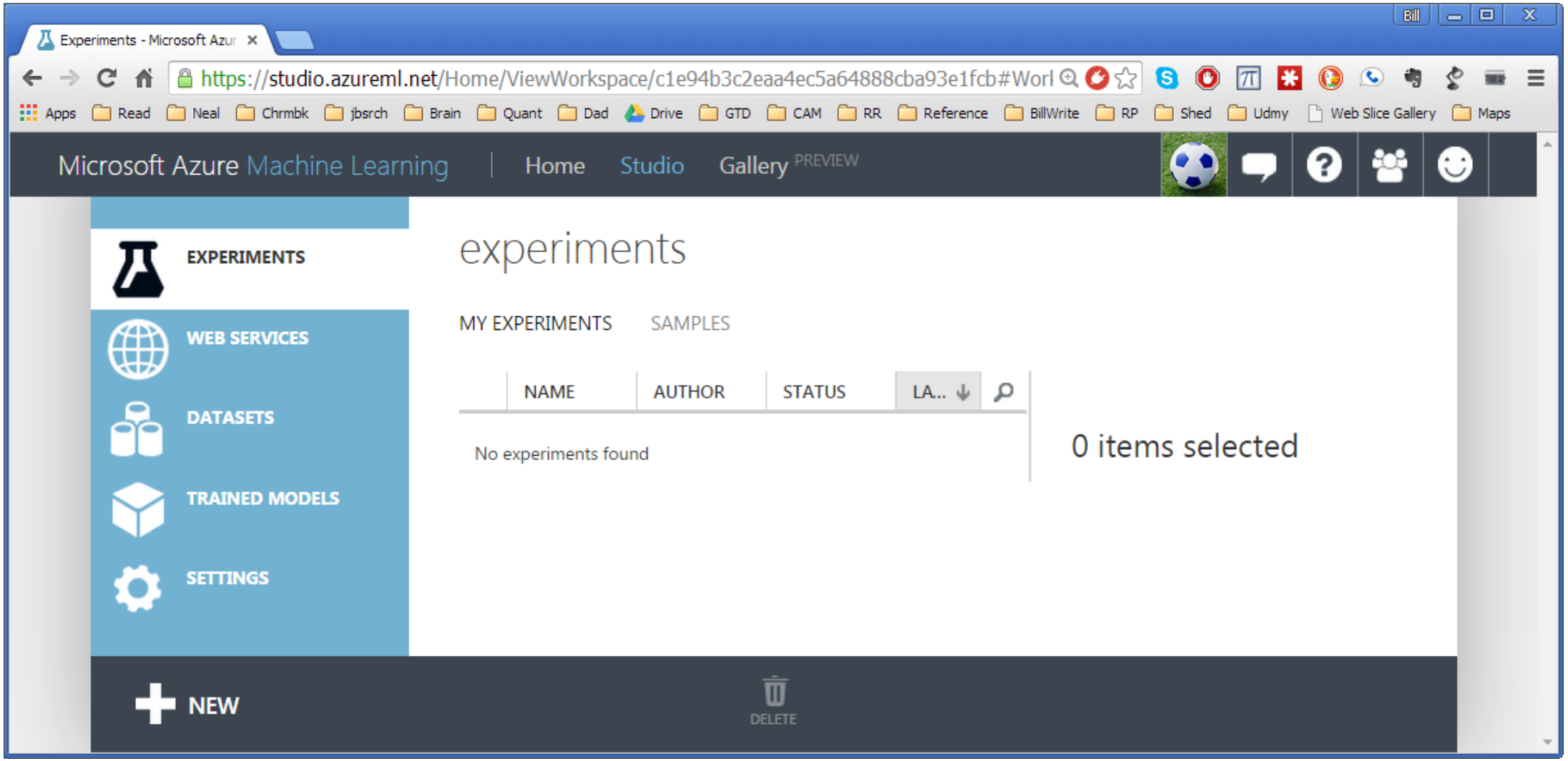
Step 1 : Go to https://studio.azureml.net/

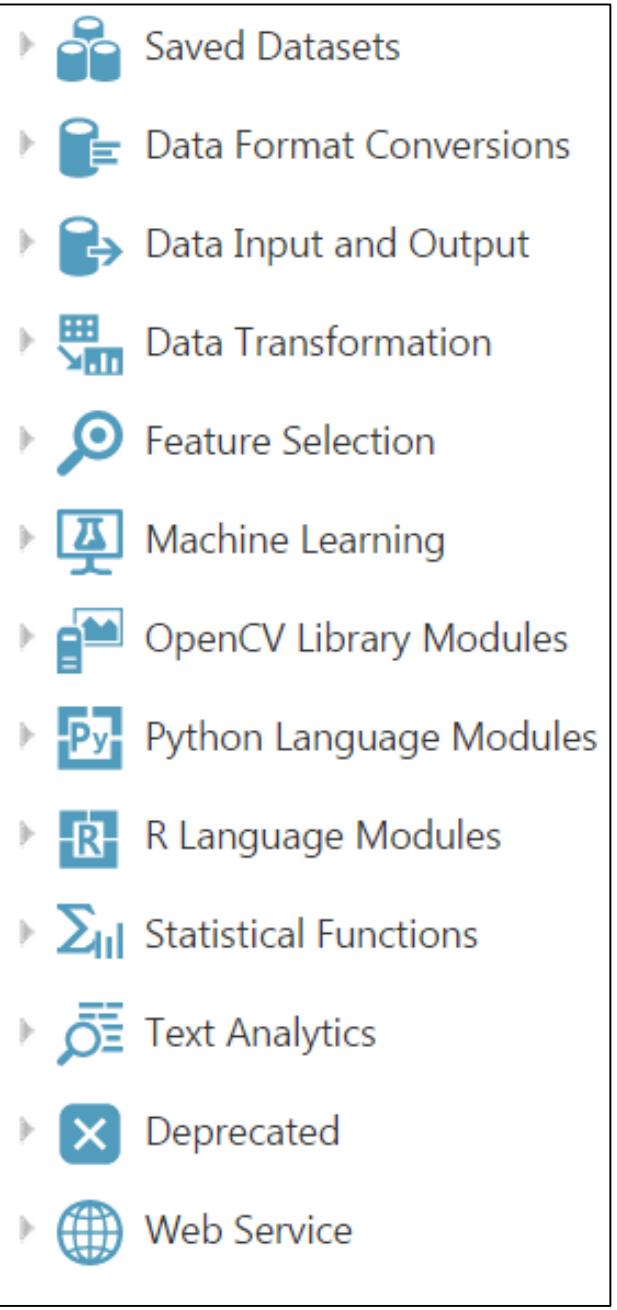


Microsoft

NEAL ANALYTICS

Step 2 : Log in to your account

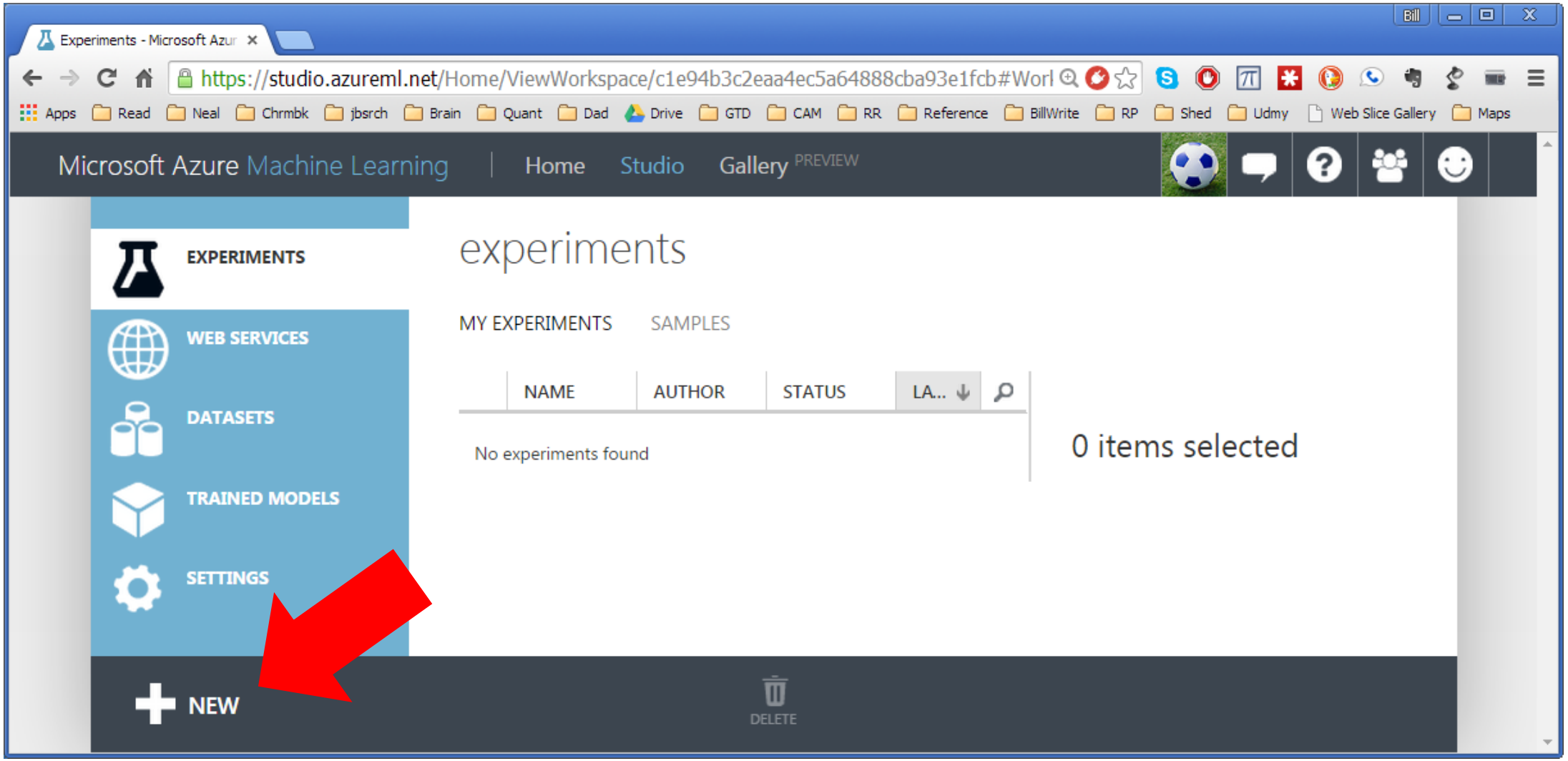




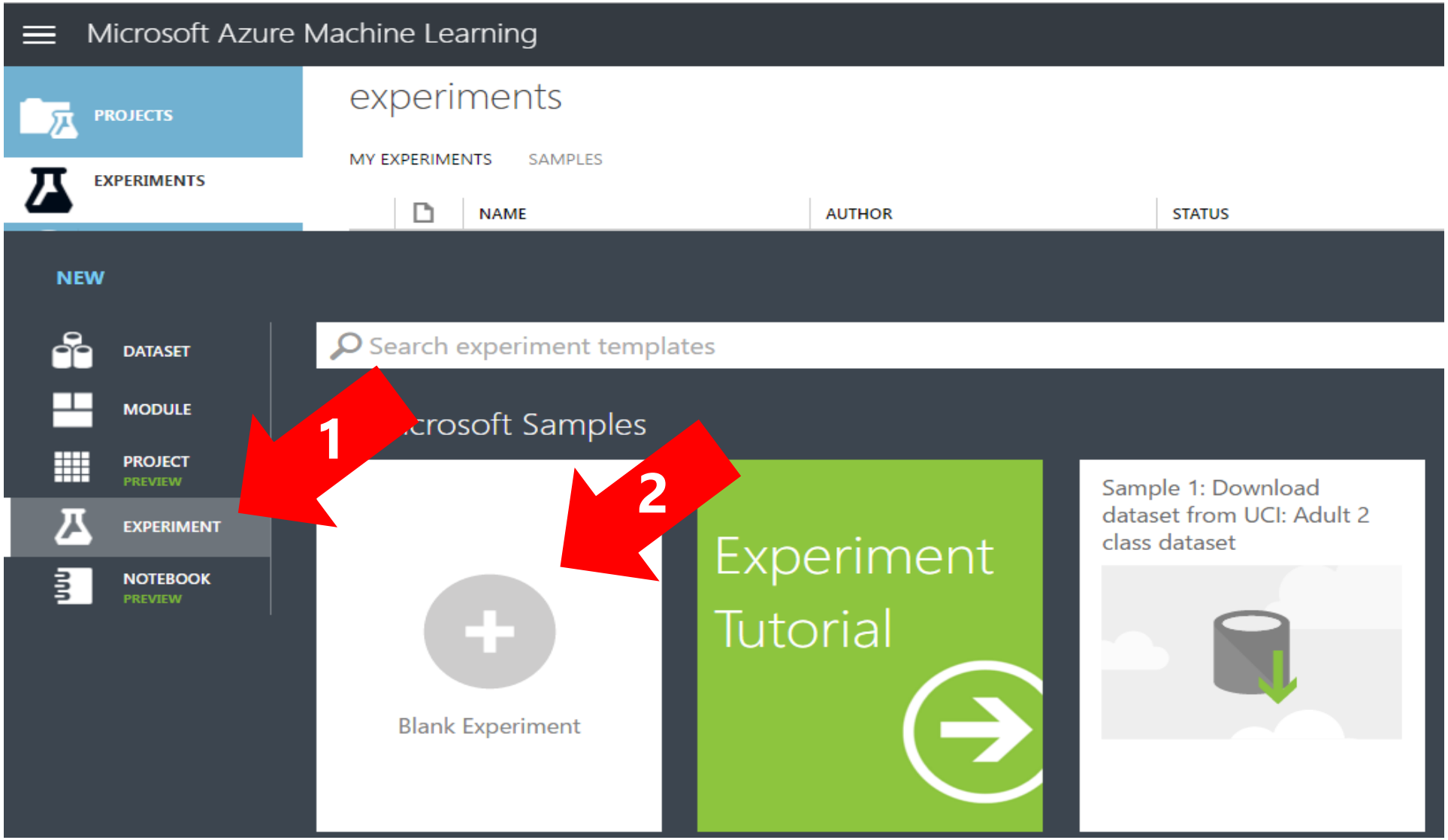
Step 3.1 : Create New Experiment

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

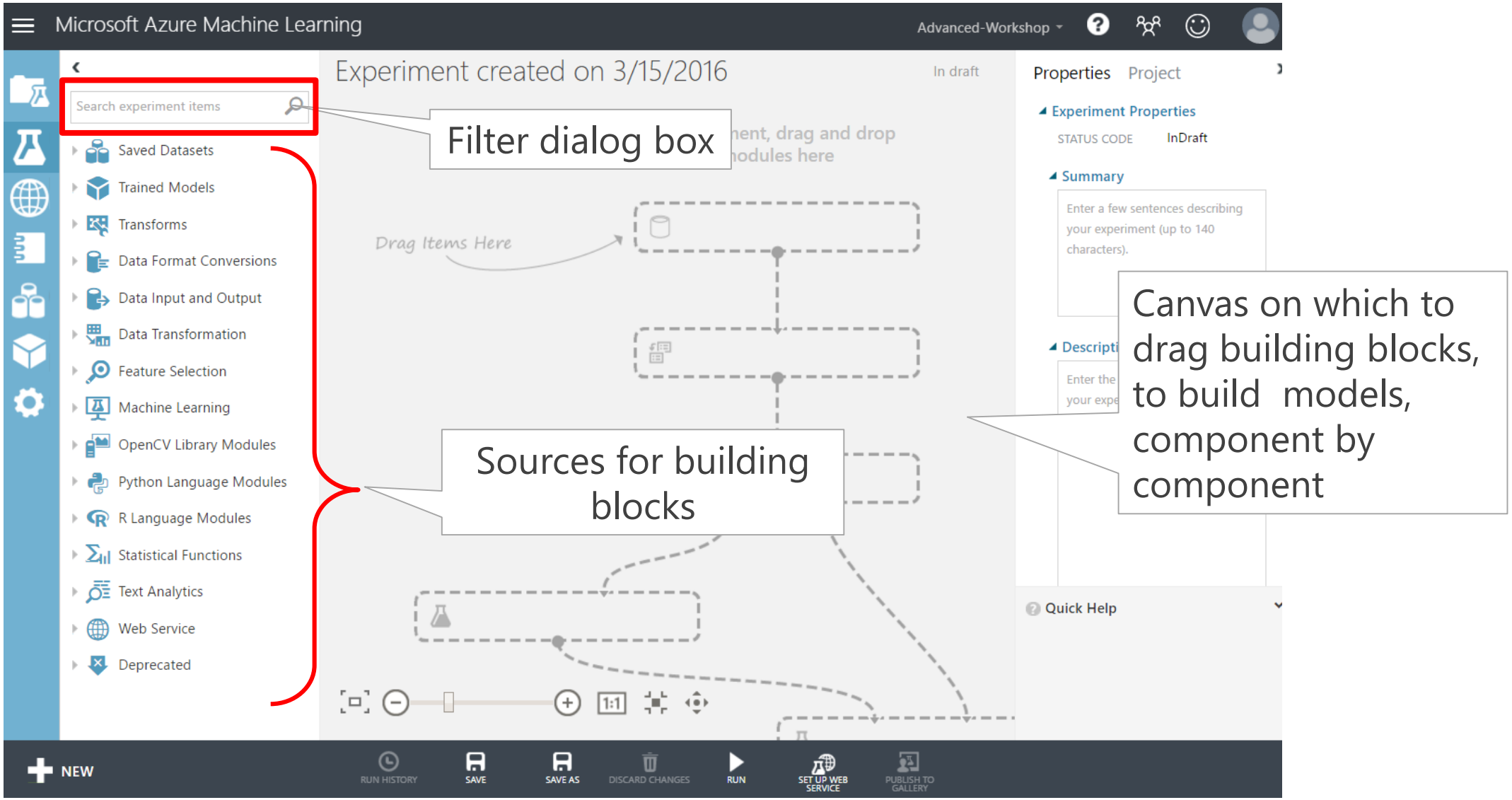
Step 3.2 : Create a new experiment



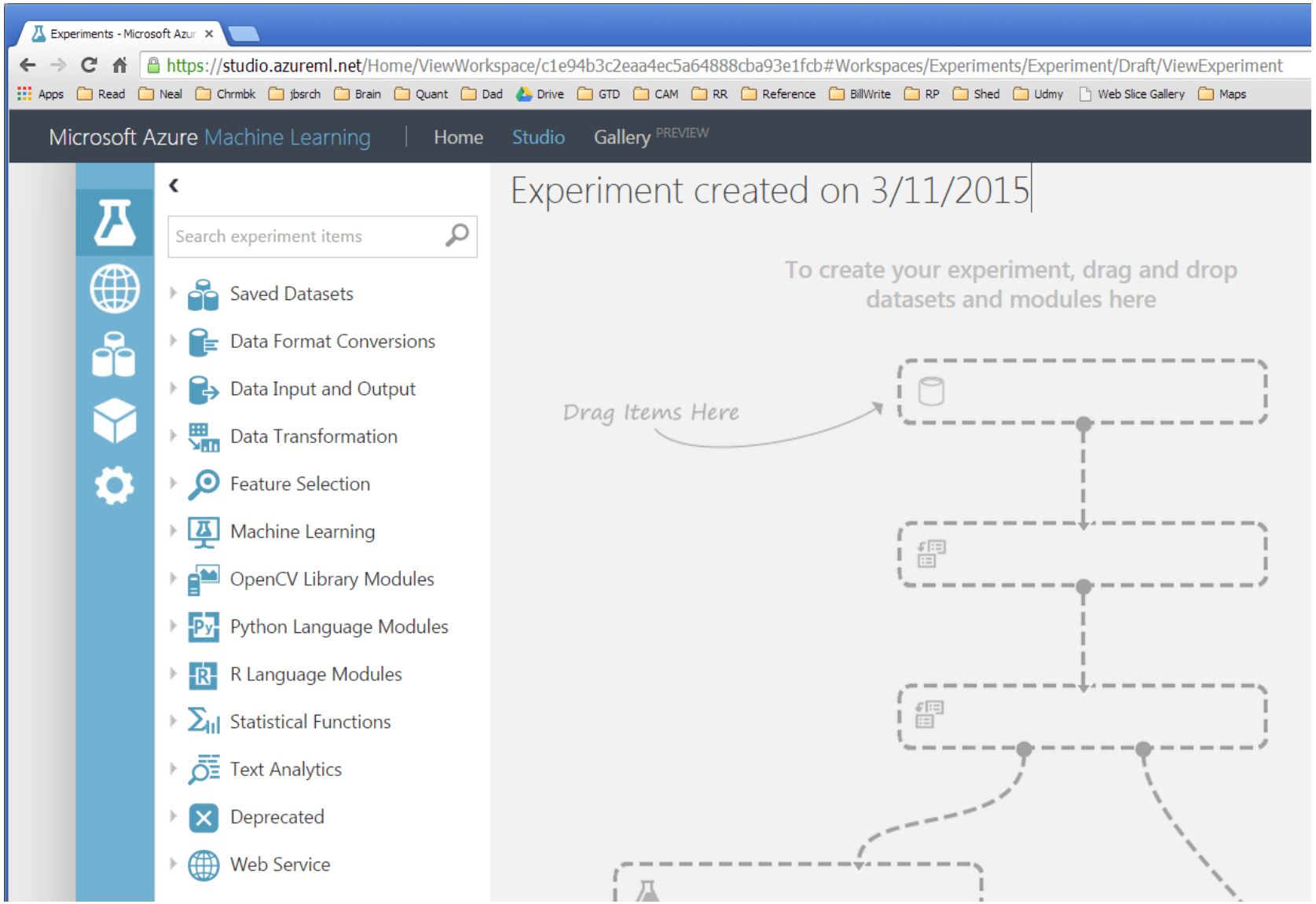
Step 3.3 : Create a new experiment



Step 3.4 : Slow down and look at the experiment



Step 3.5 : This is a “drag and drop” environment

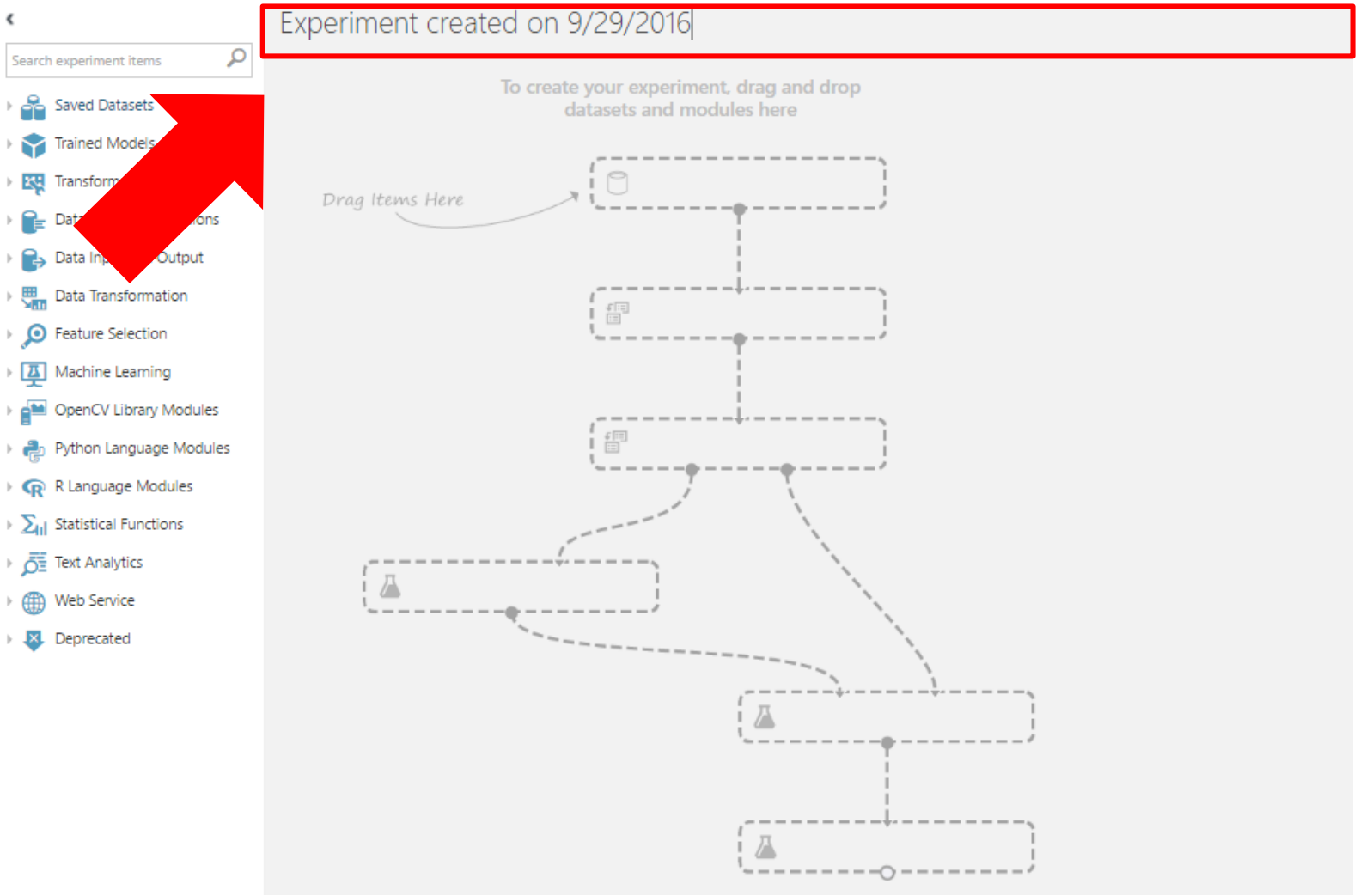


Azure Machine Learning is to modelling as …

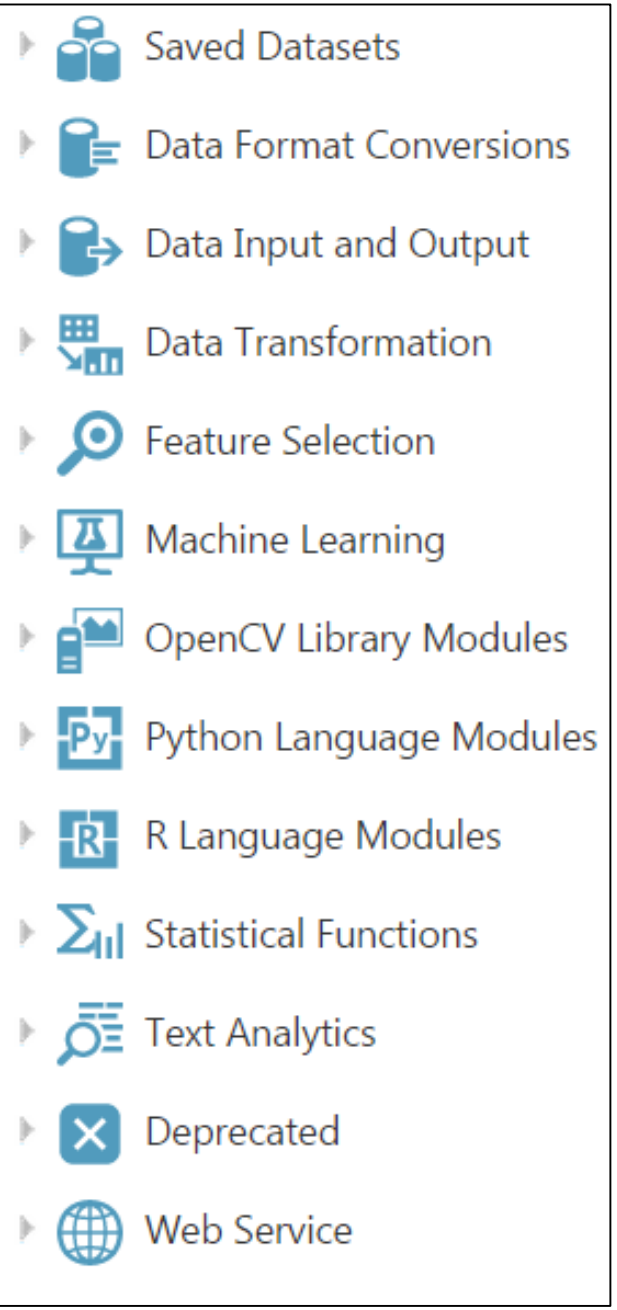
Inserting shapes into PowerPoint is to drawing

= Modelling by dragging and dropping

Step 3.6 : Name your experiment



- Click on the title box at the top that says "Experiment Created on ...."
- Give it a name of your choice



Step 4.1 : Import the dataset

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

Import Data

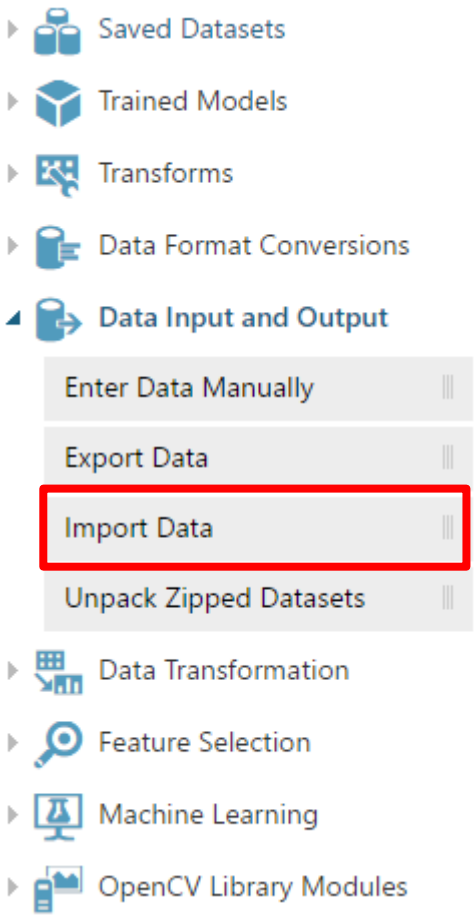
- We have three options for importing data -
  - Get data from Hive (seamless and most optimal)
  - Get data from Azure Blob Storage
  - Get data from csv

- Final datasets for EDA --

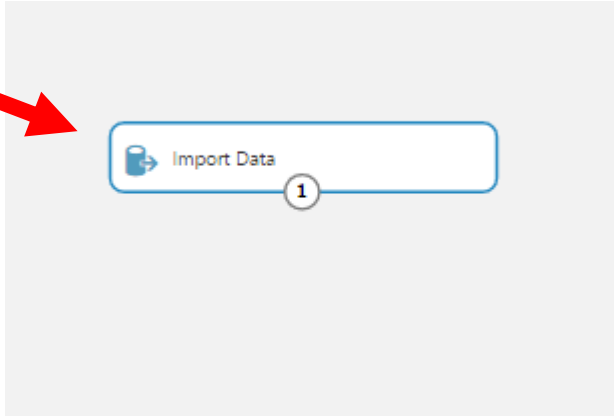
Scenario	Dataset Name
Dynamometer	Dynamometer Card Reading Data Raw Dynamometer Clusters
Predictive Maintenance of Compressors	Predictive Maintenance Raw Data
Tank Level Forecasting	Tank Level Forecasting Raw Data

Option 1 - Import the training dataset using a Hive Query. If option 1 did not work, skip to option 2.

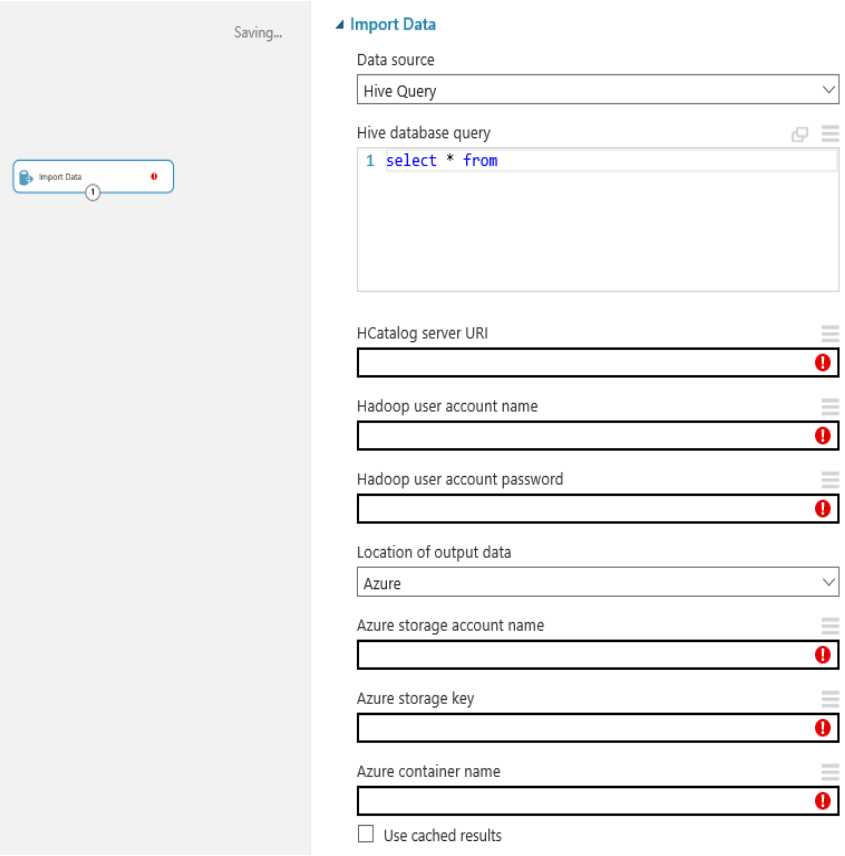
Step 4.2 : Q? How to import data using Hive Query?



- 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.



Step 4.3 : Q? How to import data using Hive Query?



1. Machine Learning
   - Click on Import Data
   - Chose “HiveQuery” for Data source
   - Add in Hive query to pull in data
   - Update information on the cluster and storage account details
2. Pull data in as per scenario
3. Need to add select * from <tablename>
4. Add in Data Import module specific for each data set.

Scenario	Dataset Name
Dynamometer	Dynamometer Card Reading Data Raw Dynamometer Clusters
Predictive Maintenance of Compressors	Predictive Maintenance Raw Data
Tank Level Forecasting	Tank Level Forecasting Raw Data

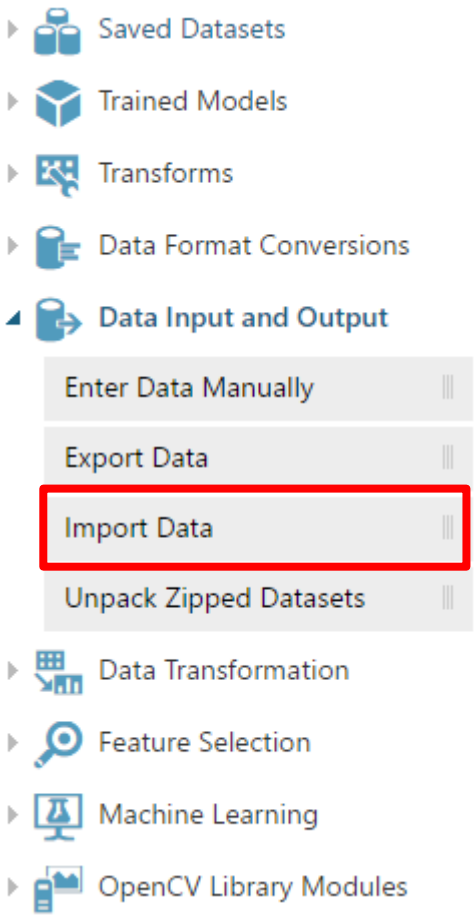
After updating value, click “Cached Results” and hit RUN.

If Option 1 worked, move on to Step 5.1

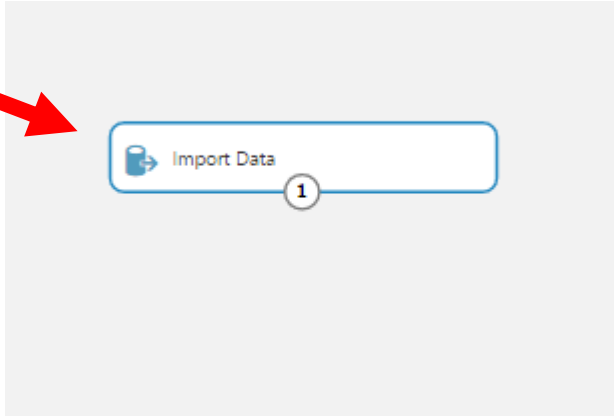
Option 2 - Import the training dataset from the Azure Blob Storage.

Microsoft NEAL ANALYTICS

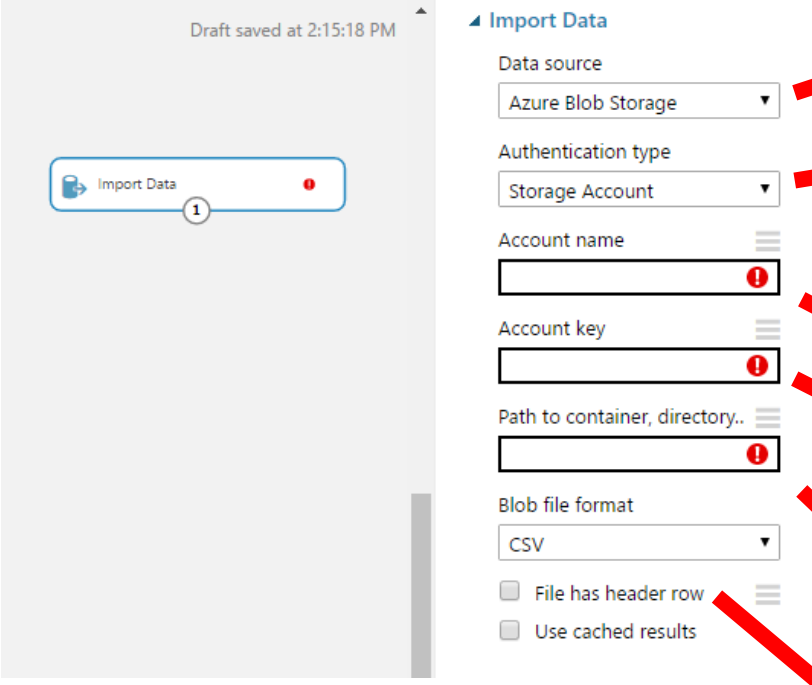
Step 4.4 : Q? How to import data from Azure Blob?



- 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



Step 4.5 : 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+wIalzfWKxyIjA==**"
    - Path to container – "**/oilandgas/_____.csv**"
      (Look next slide for the dataset name)
    - Check the File has header row check box

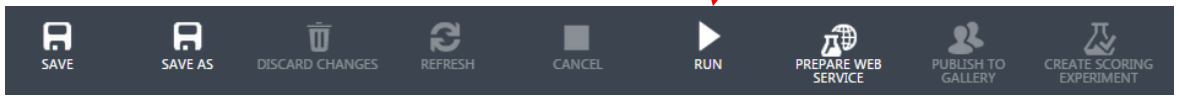
Step 4.6 : Q? How to import data from Azure Blob?

Fill the _____.CSV in the previous slide depending on which scenario you picked

Scenario	Dataset Name
Dynamometer	Dynamometer Card Reading Data Raw Dynamometer Clusters
Predictive Maintenance of Compressors	Predictive Maintenance Raw Data
Tank Level Forecasting	Tank Level Forecasting Raw Data

Step 1: Click “Save”

Step 2: Click “Run”

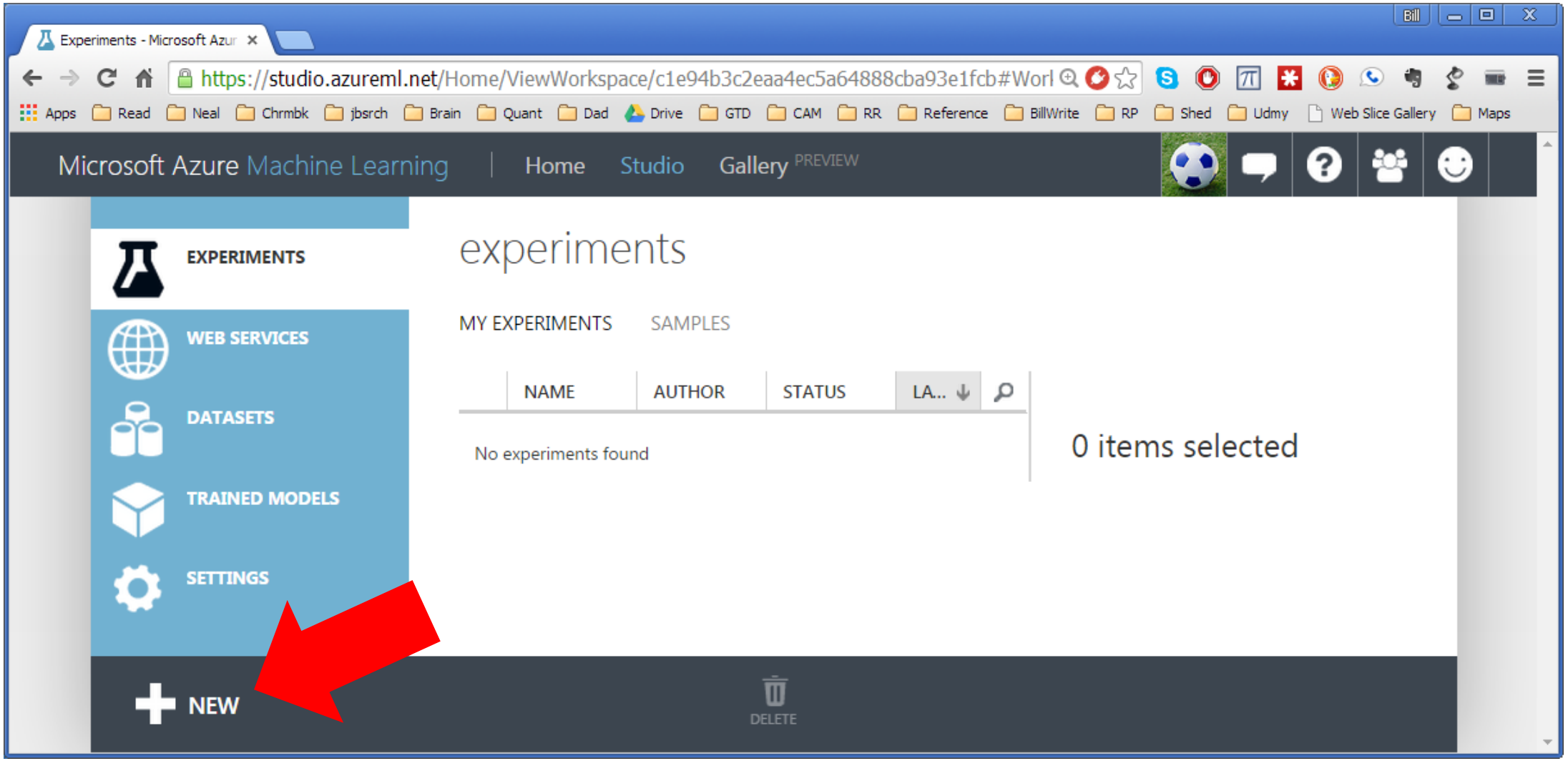


If Option 2 worked, move on to Step 5.1

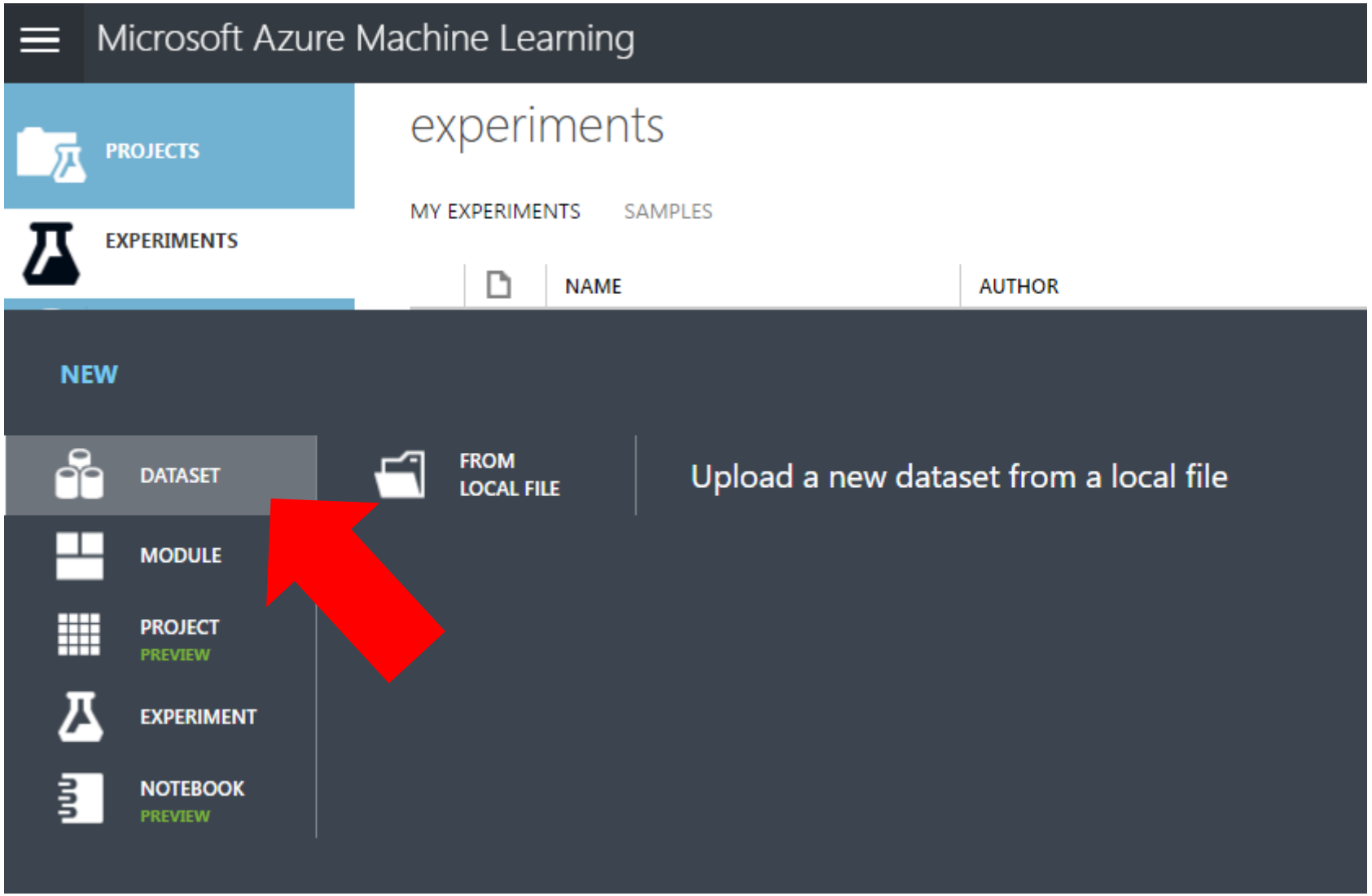
Option 3 - Import the training dataset from the saved datasets.

Microsoft NEAL ANALYTICS

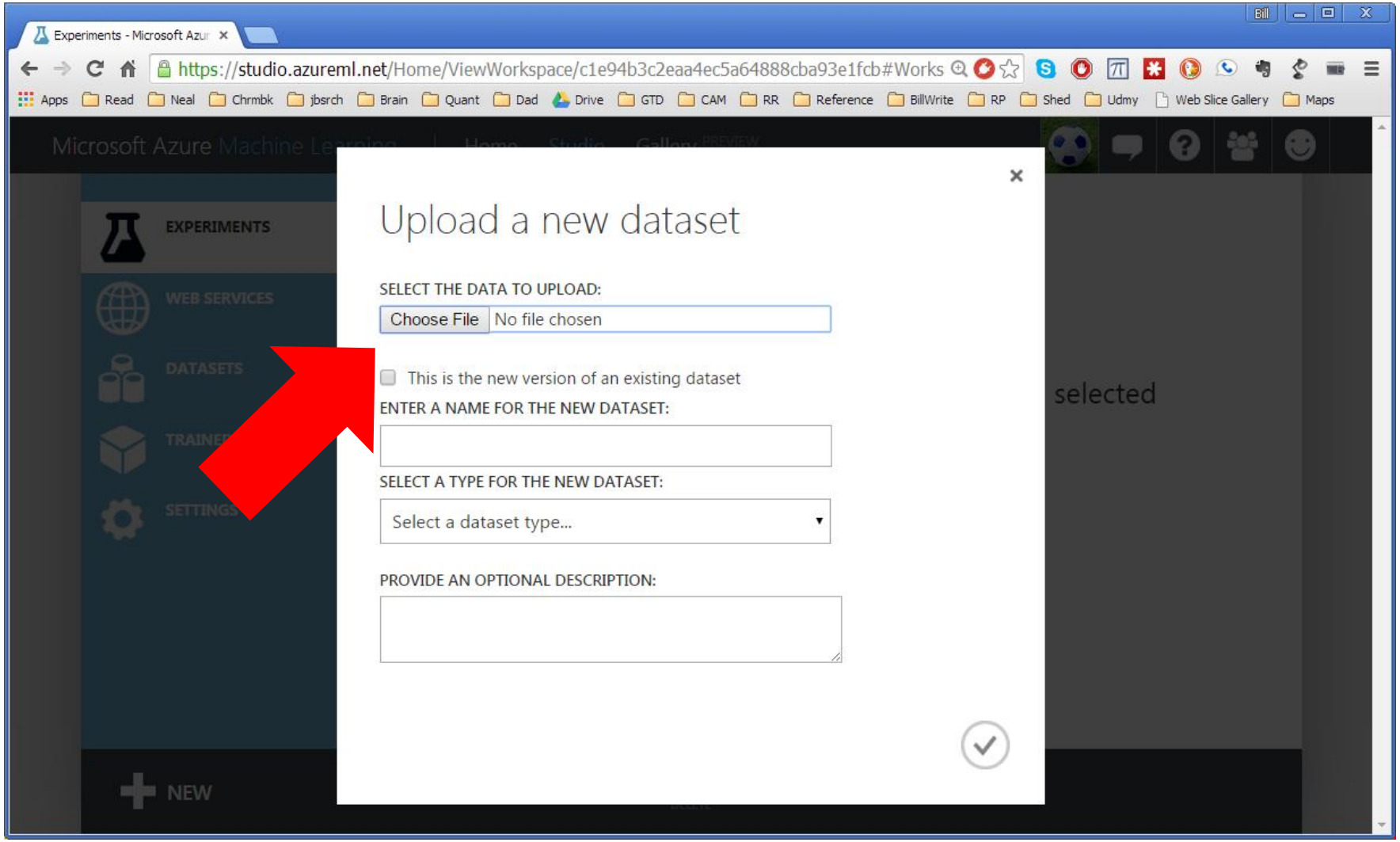
Step 4.7 : Import this tutorial's training dataset



Step 4.8 : Import this tutorial's training dataset



Step 4.9 : Import this tutorial's training dataset



Step 4.10 : Import this tutorial's training dataset

Microsoft Azure Machine Learning Studio

Advanced-Workshop

PROJECTS
EXPERIMENTS
WEB SERVICES
NOTEBOOKS
DATASETS
TRAINED MODELS
SETTINGS

experiments

MY EXPERIMENTS SAMPLES

NAME	AUTHOR	STATUS	LAST EDITED	PROJECT
Oil & Gas - Tank Level For...	achal_mallaya	Draft	9/28/2016 5:39:58 PM	None
Oil & Gas - Tank Level For...	achal_mallaya	Finished	9/28/2016 5:25:19 PM	None
OK Training - Tank Level F...	sailaja.karthik	Draft	9/28/2016 5:19:53 PM	None
Oil & Gas - Brine Analysis ...	achal_mallaya	Draft	9/28/2016 3:53:47 PM	Oklahoma Training Worksh...
Oil & Gas - Dynamometer ...	achal_mallaya	Draft	9/28/2016 1:58:14 PM	Oklahoma Training Worksh...
Oil & Gas - Dynamometer ...	achal_mallaya	Draft	9/28/2016 1:57:51 PM	Oklahoma Training Worksh...
Oil & Gas - Dynamometer ...	achal_mallaya	Draft	9/28/2016 1:57:30 PM	Oklahoma Training Worksh...
Oil & Gas – Binary Classifi...	achal_mallaya	Draft	9/28/2016 5:33:52 AM	Oklahoma Training Worksh...
Oil & Gas - Linear Regressi...	achal_mallaya	Finished	9/28/2016 5:26:28 AM	Oklahoma Training Worksh...
Oil & Gas - Linear Regressi...	achal_mallaya	Draft	9/28/2016 5:25:59 AM	Oklahoma Training Worksh...
Oil & Gas - Dynamometer ...	achal_mallaya	Finished	9/27/2016 6:33:37 PM	Oklahoma Training Worksh...
Oil & Gas - Linear Regressi...	Ryan	Draft	9/27/2016 4:33:20 PM	None
OK Cluster, Classify, Regre...	eric.hullander	Finished	9/26/2016 12:32:05 PM	None
Brine Analysis K Means Cl...	sailaja.karthik	Draft	9/26/2016 4:00:55 AM	None
OK - K Means Clustering -...	sailaja.karthik	Finished	9/26/2016 3:54:17 AM	None
OK Predictive Maintenanc...	eric.hullander	Draft	9/22/2016 2:02:02 PM	OK Oil and Gas Workshop
Oil & Gas – Binary Classifi...	sailaja.karthik	Failed	9/20/2016 12:30:00 PM	None
Binary Classification: Brea...	Microsoft	Draft	9/14/2016 4:14:28 PM	None
OK Pre ntenanc...	Microso	Draft	9/14/2016 1	OK Oil and G
Aerosp r Regres...	zperkel	Finished	9/14/2016 1	Aerospace W

1 2

Select Columns in Dataset
Normalize Data
Edit Metadata
Edit Metadata
Split Data
Select Columns in Dataset
Filter Based Feature Selection
Two-Class Decision Forest
SMOTE
Two-Class Boosted Decision T...
Train Model
Train Model
Score Model
Score Model
Evaluate Model

Upload of the dataset 'Tank Level Forecasting Training Data.csv' has completed.

OK

NEW

DELETE

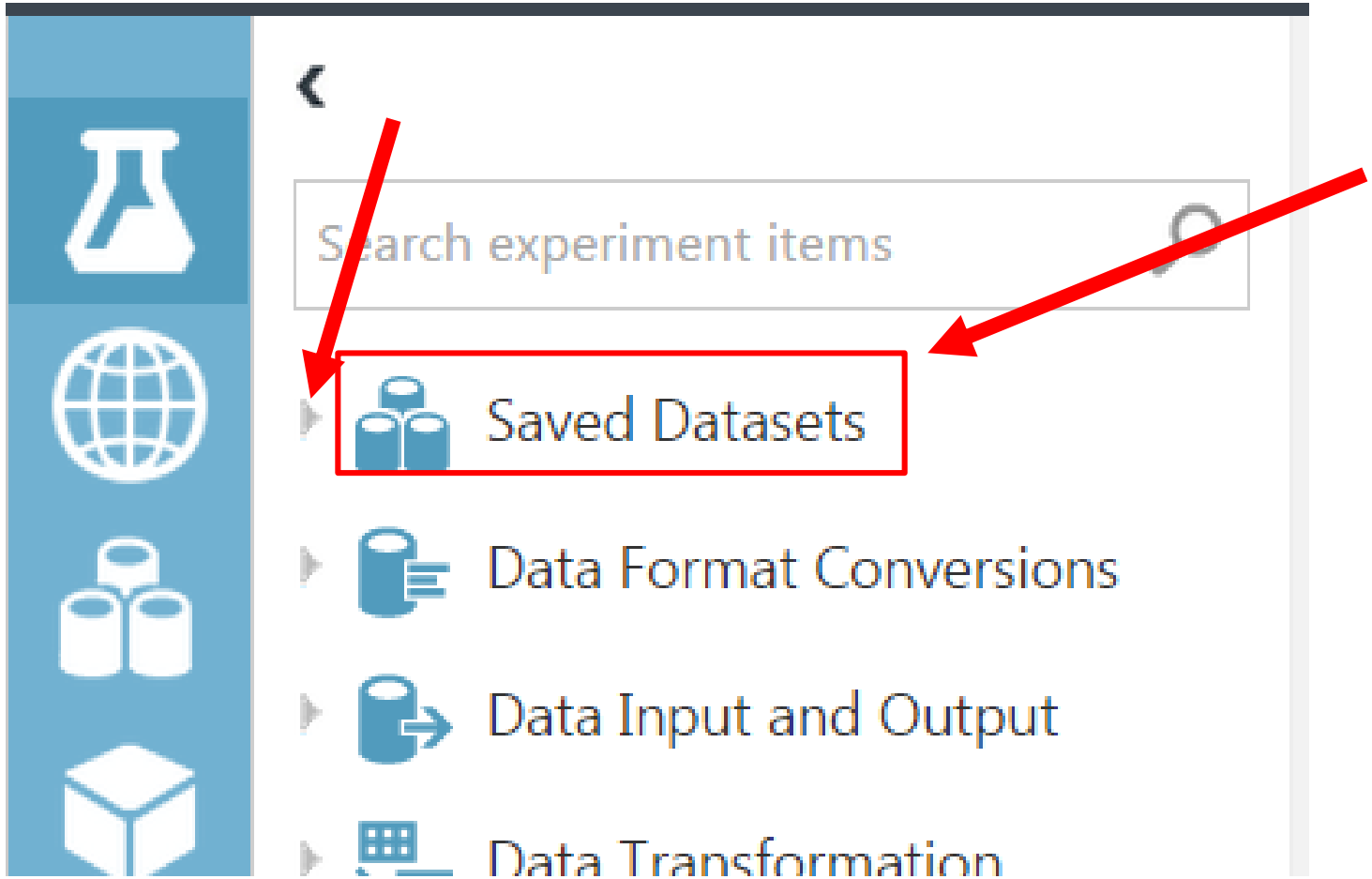
ADD TO PROJECT

COPY TO WORKSPACE

Microsoft

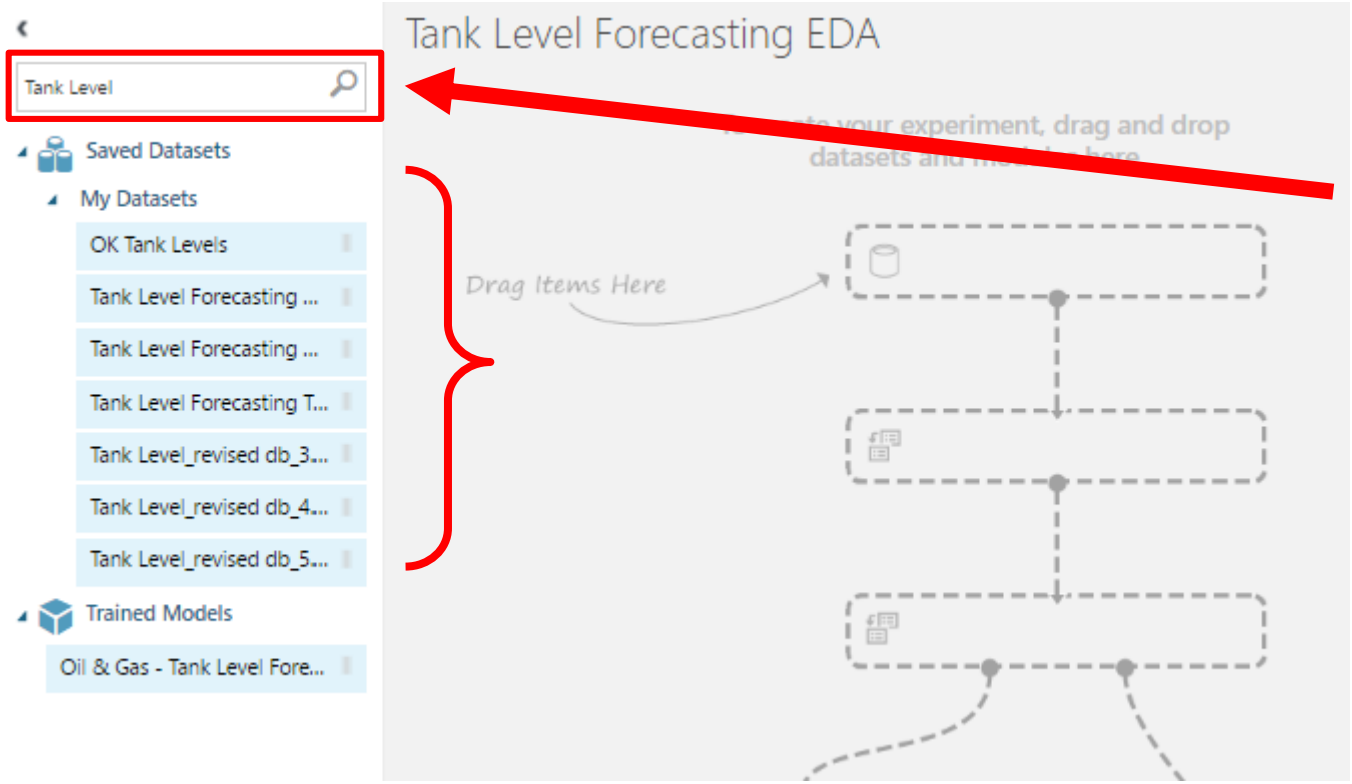
NEAL ANALYTICS

Step 4.11 : Open “Saved Datasets”



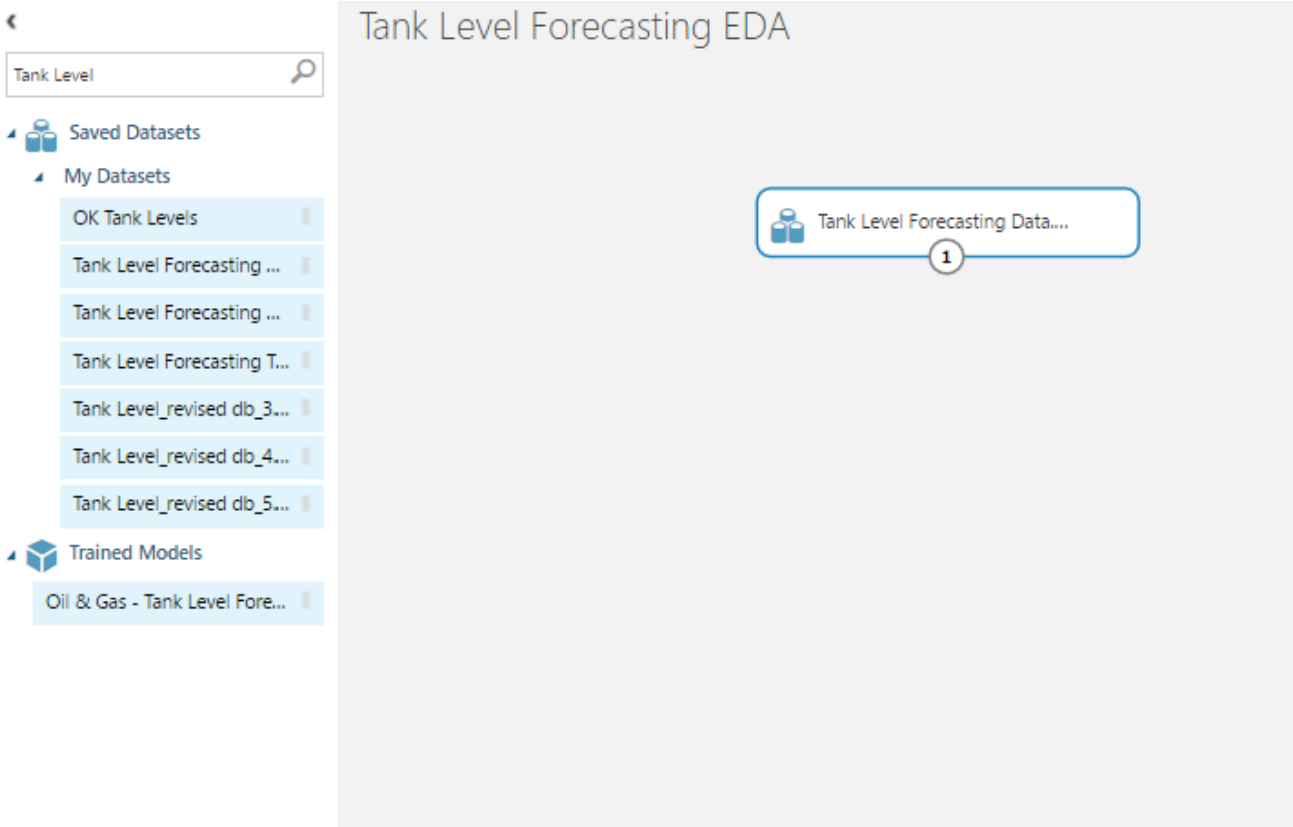
- By clicking on the triangle at the left
  of “Saved Datasets”

Step 4.12 : Take a second to notice the MANY datasets



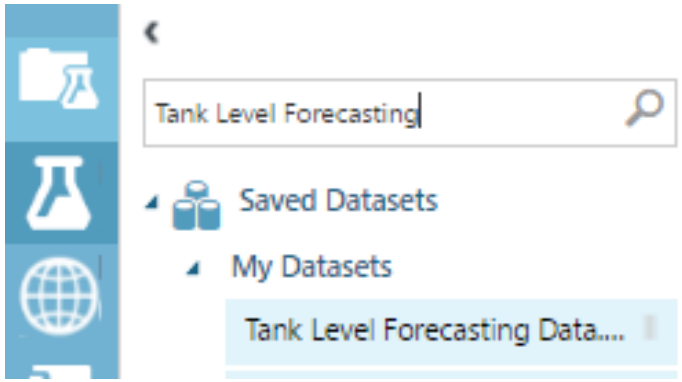
- To filter to the data set for this tutorial …
- Type the name of your dataset in "Search experiment items" dialog box
- The data set list will reduce to our data set for this tutorial

Step 4.13 : Drag the data set to the experiment

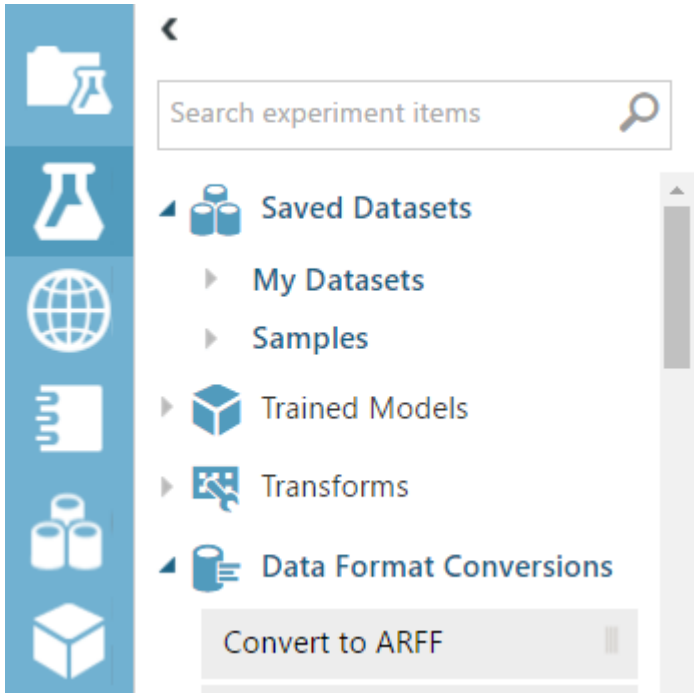


- *Note* when you drag the first element of your model to the canvas … all the guides disappear
- Now, where are all the tools that were at the left Azure Machine Learning?
- They are still there, … but we need to un-filter to see them

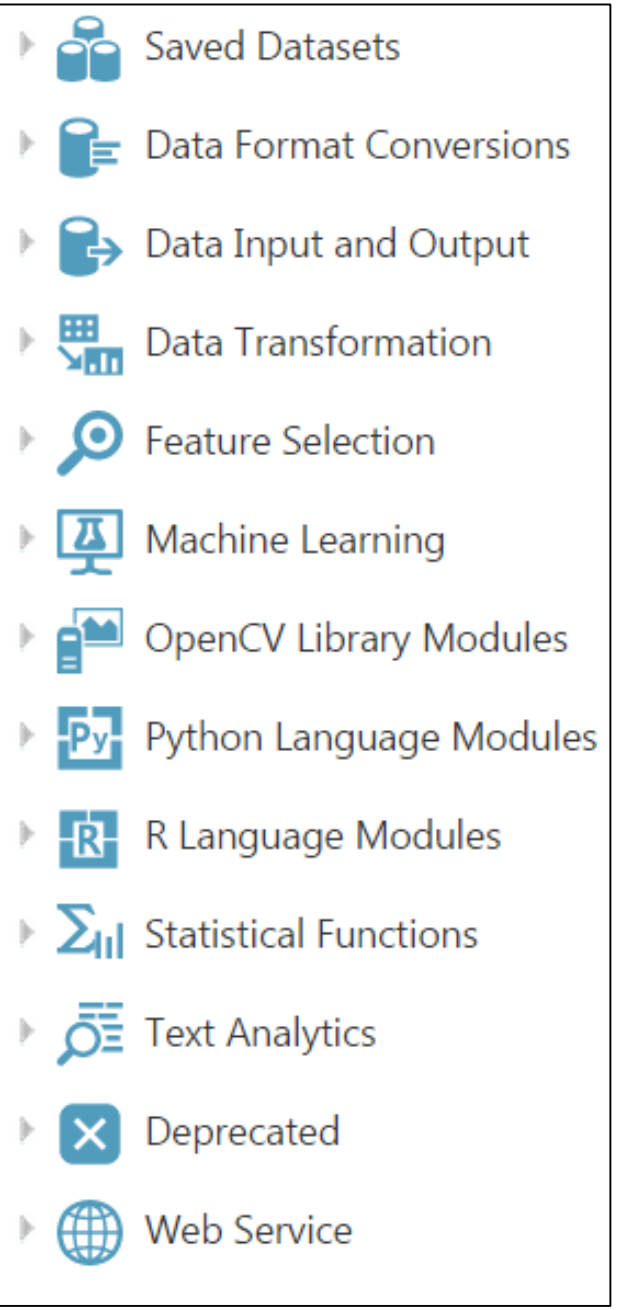
Step 4.14 : Backspace over “data”



- Backspace to remove the dataset name from the “Search experiment items” dialog



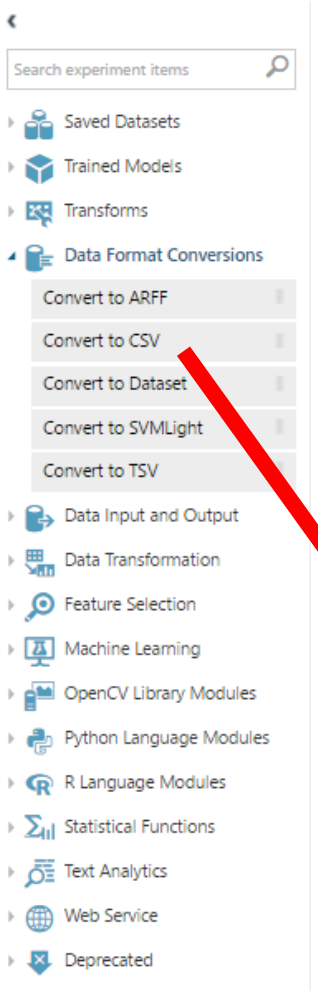
- Then click the triangle at left of “Saved Datasets” to close the dataset list



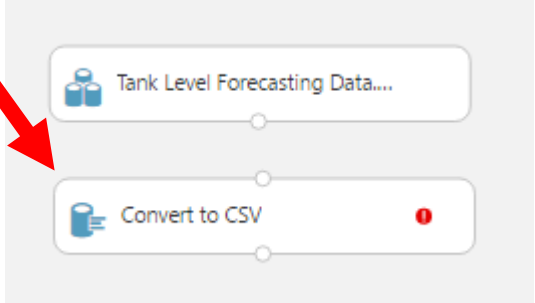
Step 5.1 : Convert to CSV

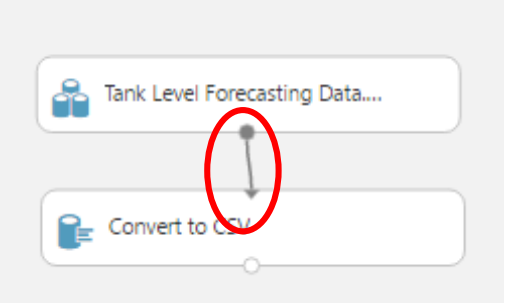
- AML modelling … a checklist approach
  - ~~Create new experiment~~
  - ~~Import data set~~
  - Convert to CSV

Step 5.2 : Add convert to CSV module to your experiment



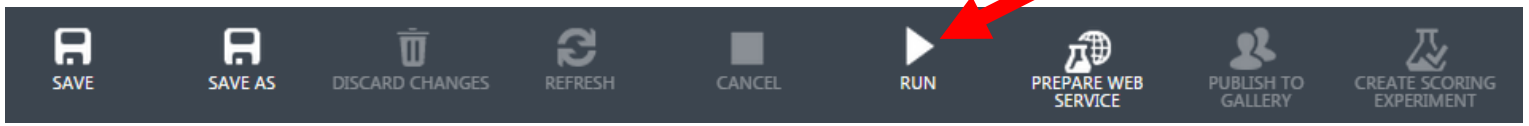
- Open "Data Format Conversion" from the navigation pane at the left
- Drag "Convert to CSV" to the canvas
  - "Convert to CSV" Converts data input to a comma-separated values format
- Next, click and hold on the bottom middle circle of your "Dataset" module
- While holding down the mouse button, drag the line to the top middle circle of "Convert to CSV" module

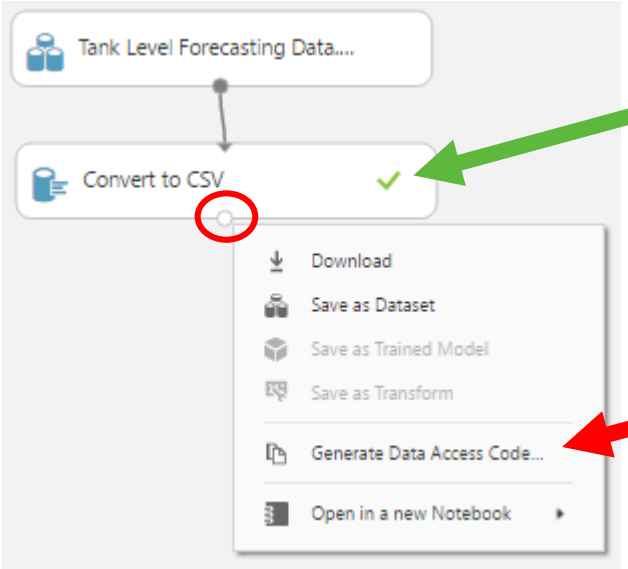




Step 5.3 : Run the experiment

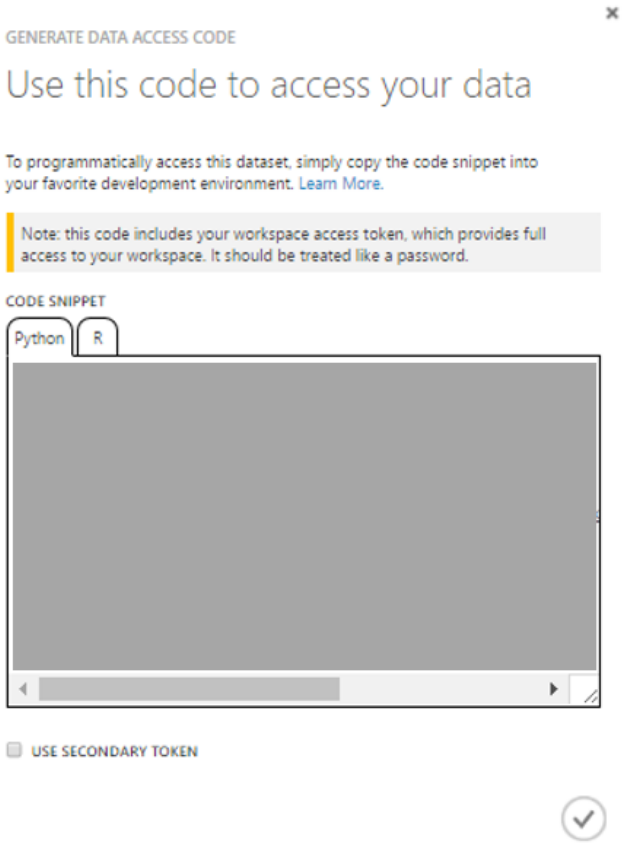
Click “Run”





- After Azure Machine Learning “runs” all the steps in your model, you will see a green check mark in the “Convert to CSV” box
- Next right click on the bottom-middle circle in “Convert to CSV” and select “Generate Data Access Code”

Step 5.4 : Generate access code

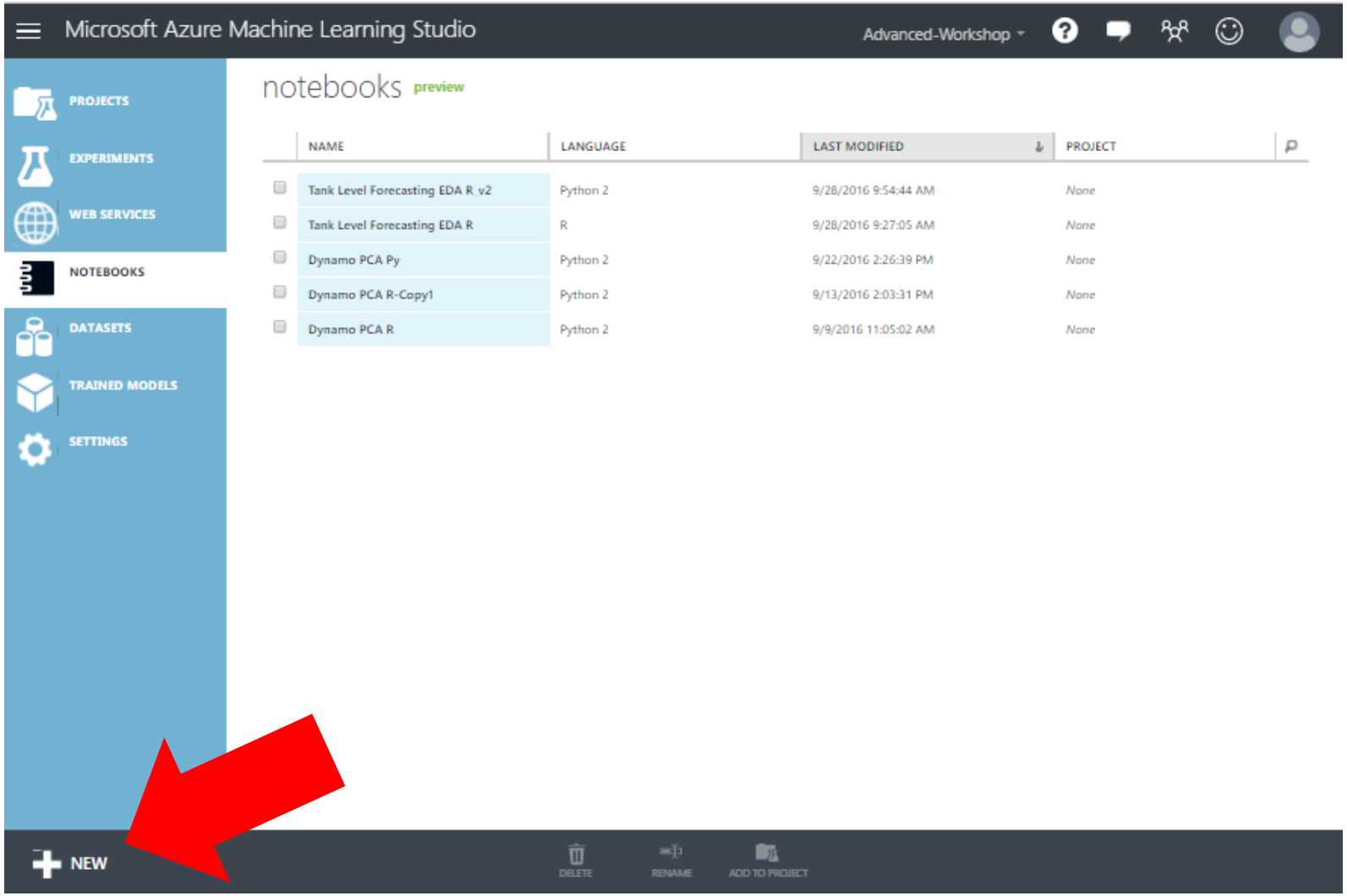


- You will find data access code one each for Python and R which you will need to import data into your Python Notebooks
- Copy it to your clipboard (Ctrl-C)

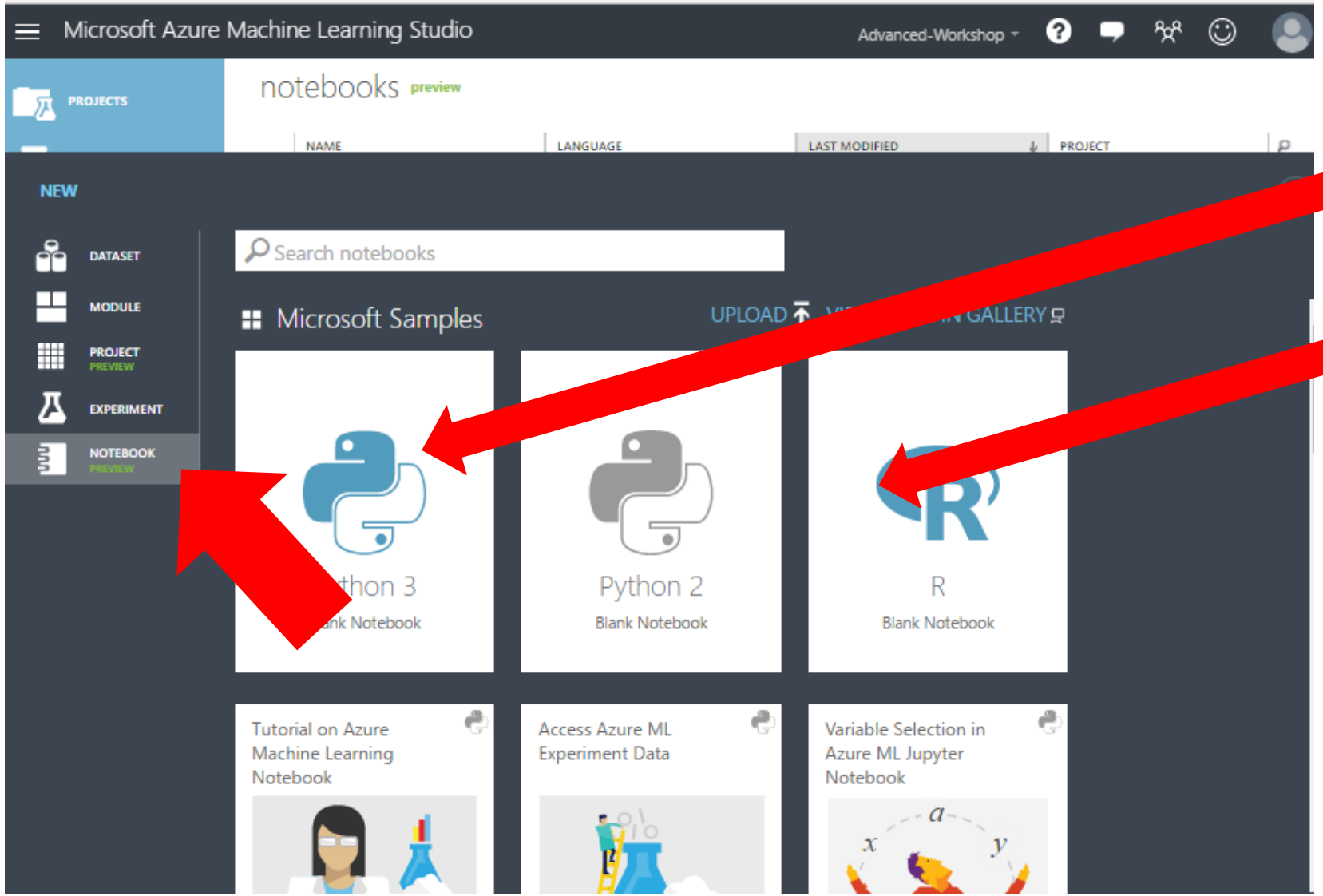
Step 5.5 : Open Python or R Notebook in Azure ML

- AML modelling … a checklist approach
  - ☑ ~~Create new experiment~~
  - ☑ ~~Import data set~~
  - ☑ ~~Convert to CSV~~
  - ❑ Open notebook in AML

Step 5.6 : Create a new Python or R Notebook

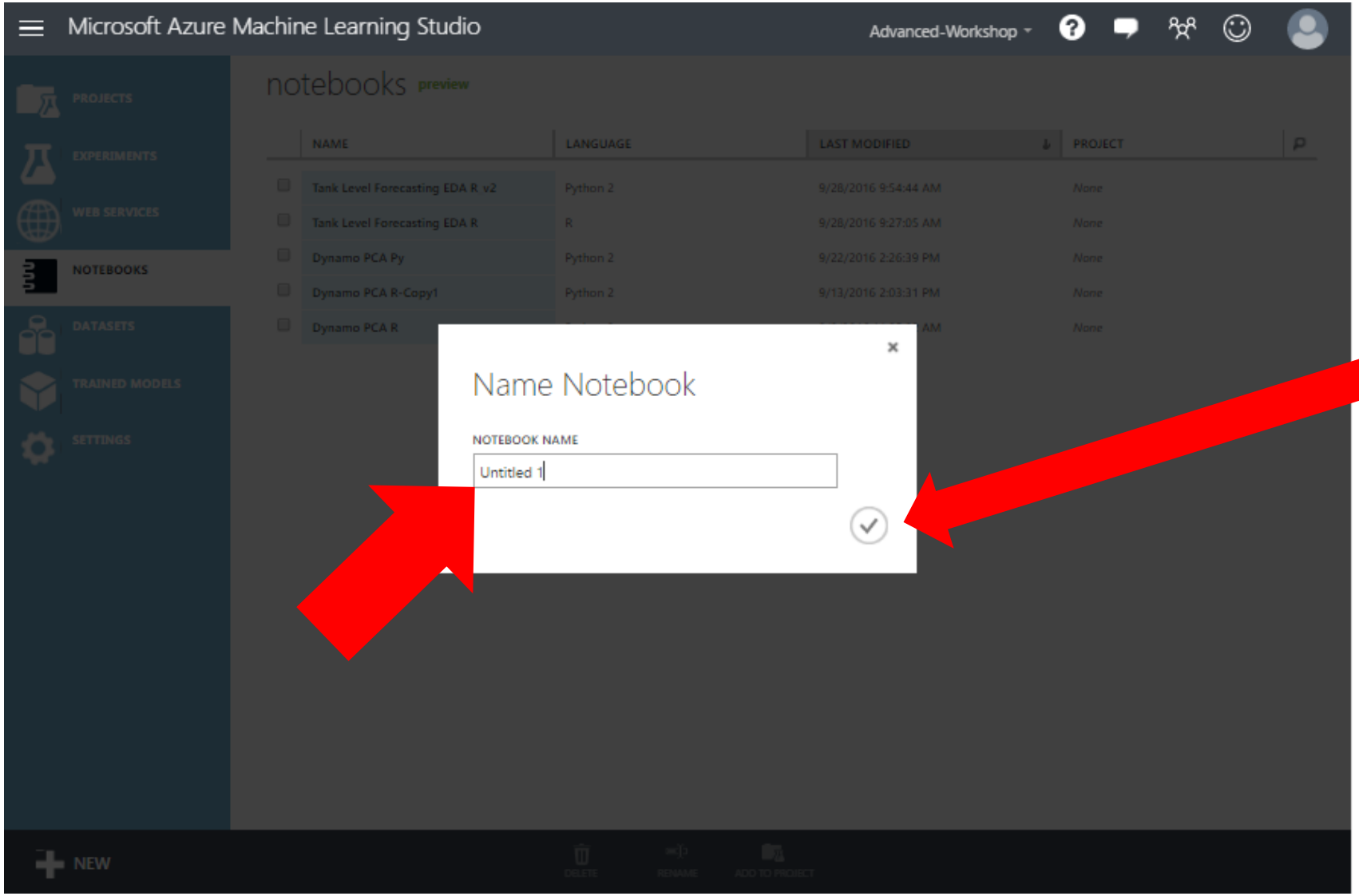


Step 5.6 : Create a new Python or R Notebook



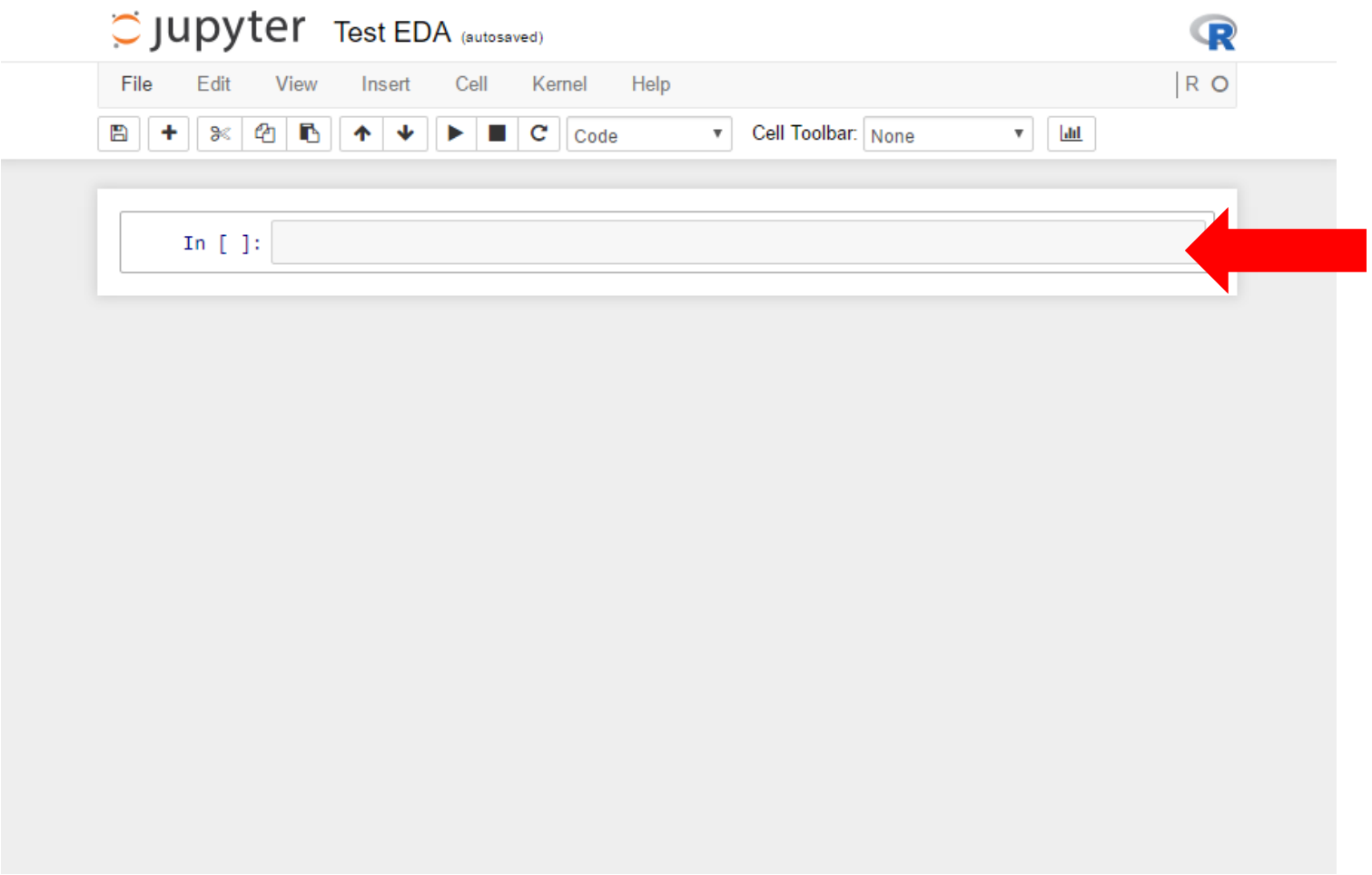
- Chose an R or Python Notebook based on your choice

Step 5.7 : Import this tutorial's training dataset



- Give a name to your notebook
- Click on check mark

Step 5.8 : Import this tutorial's training dataset



- Paste in the data access code you had copied earlier onto the clipboard depending on whether it is a Python or R Notebook