Microsoft

Cluster Analysis

Contents

Goals and Requirements

Estimated time to complete lab is 30-35 minutes

Goals

1. Implement and design a Classification model for Call Center Data.

2. Approach of using K-Means

Requirements:

1. Access to an Azure Machine Learning Studio

Cluster Analysis

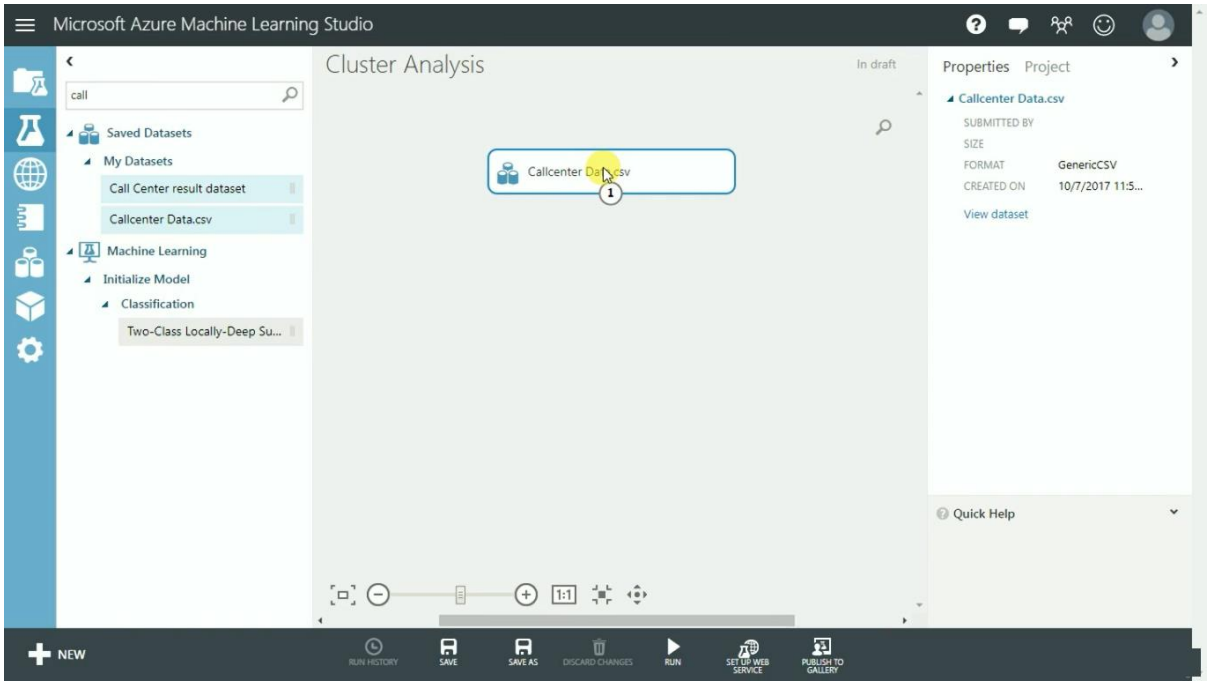
Clustering Experiment

Dataset for Callcenter Data

For cluster analysis experiment, Download the csv from course material and upload

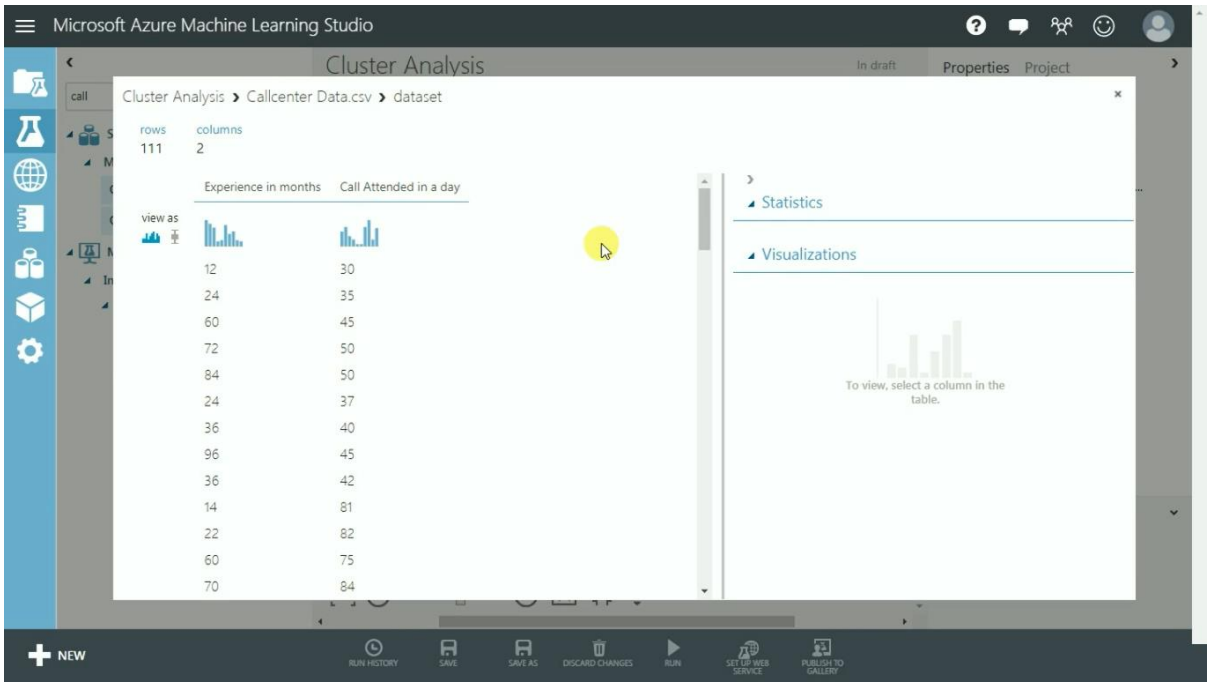
In workspace

The objective is to identify the strategy to improve the performance by clusters of employees.

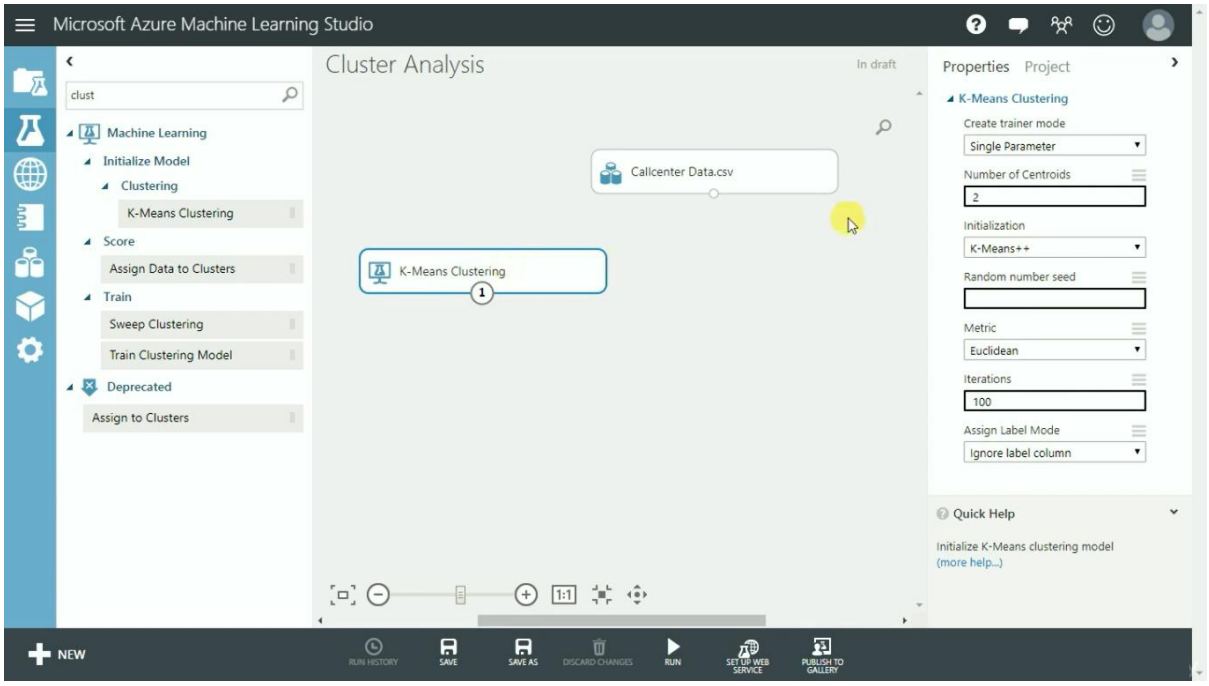


Dataset for K-means Clustering

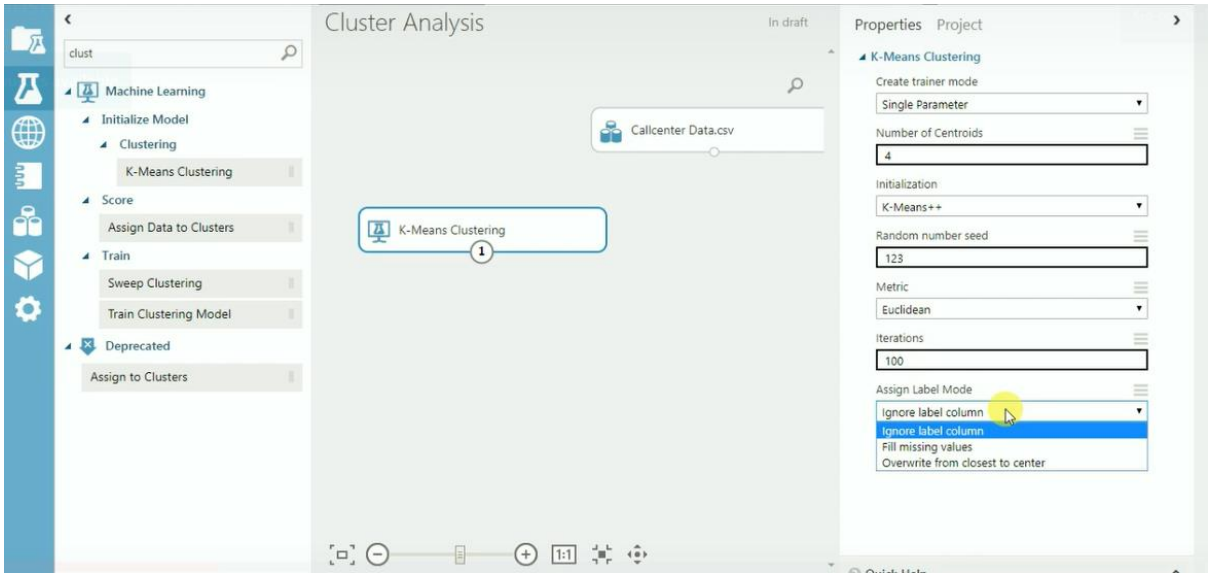
Visualize the dataset



Insert K-means clustering in canvas and check parameters

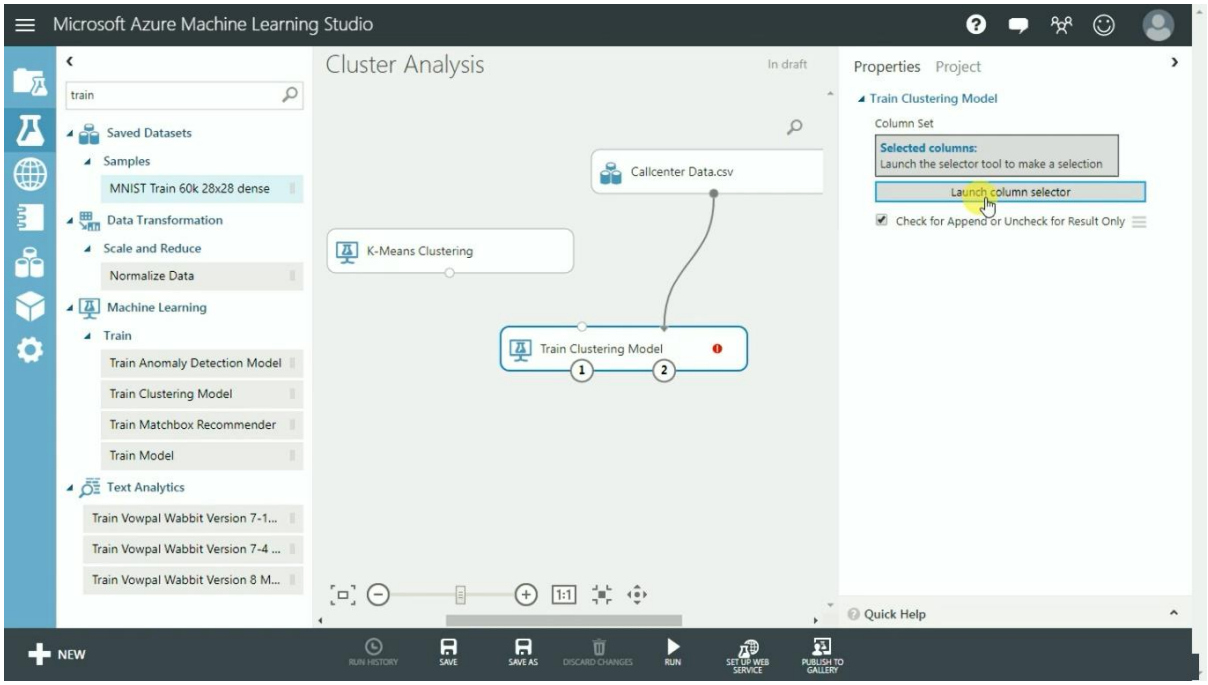


Input the parameters as shown

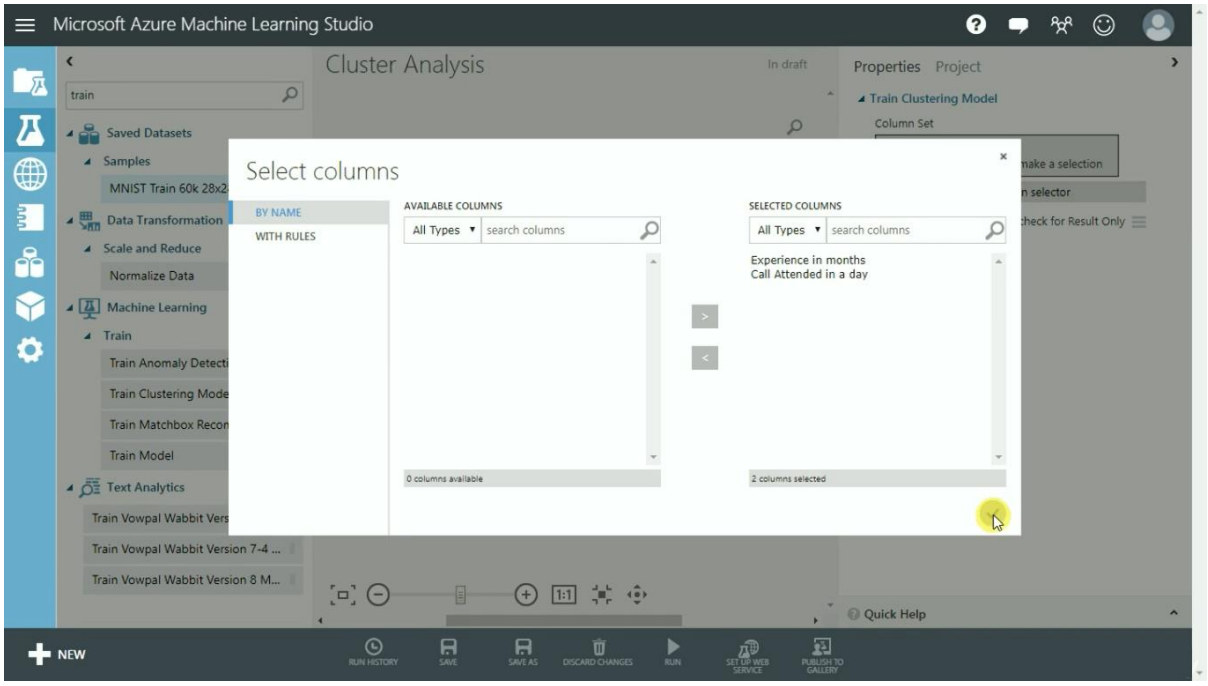


Train Clustering Model Dataset

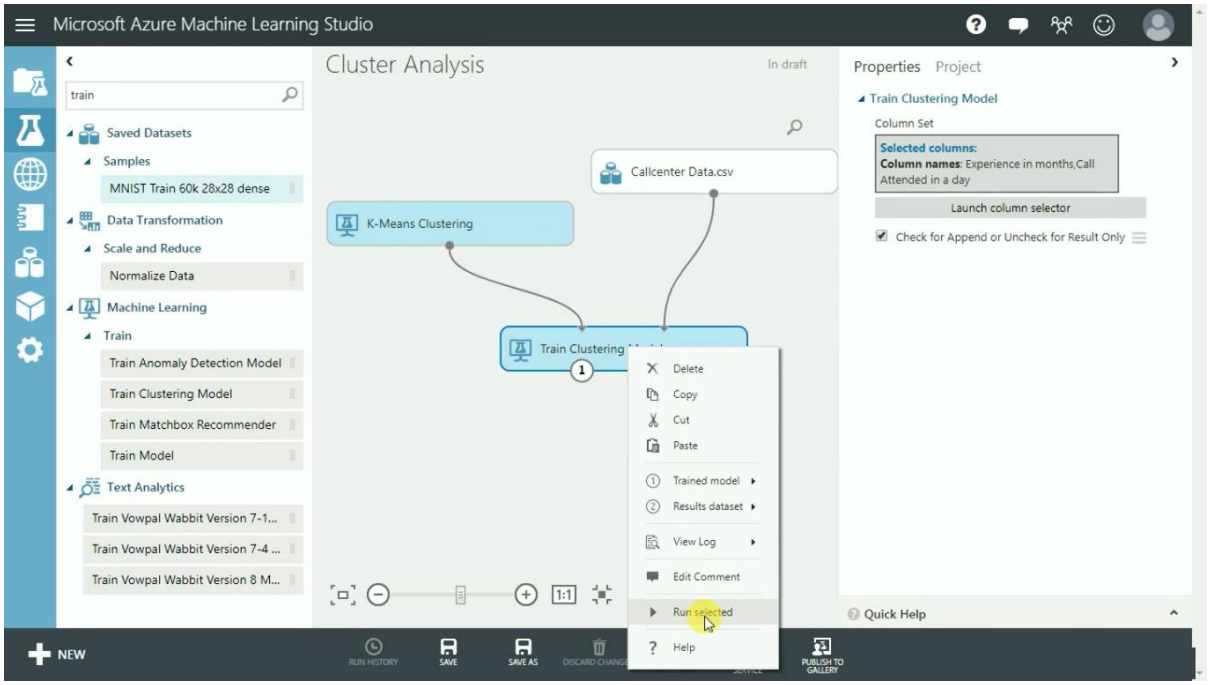
Add train clustering model in canvas and connect with call center data.csv



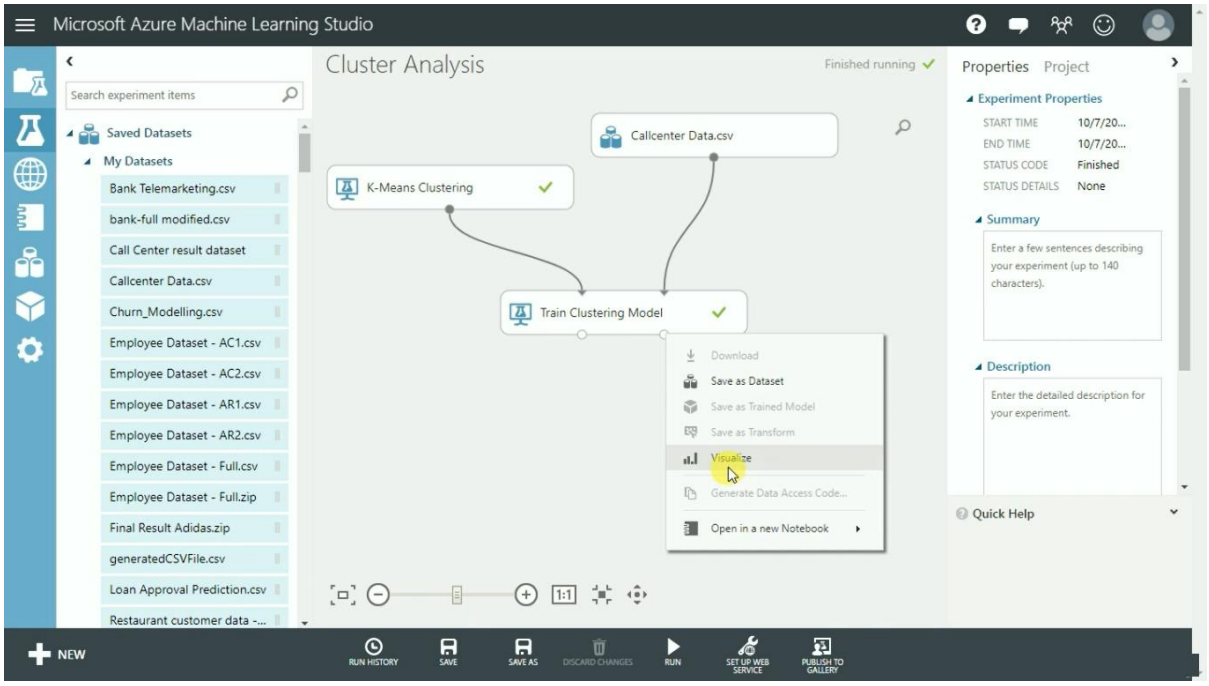
Launch column selector and select both the columns and click ok



Connect k means clustering with train clustering and run the module

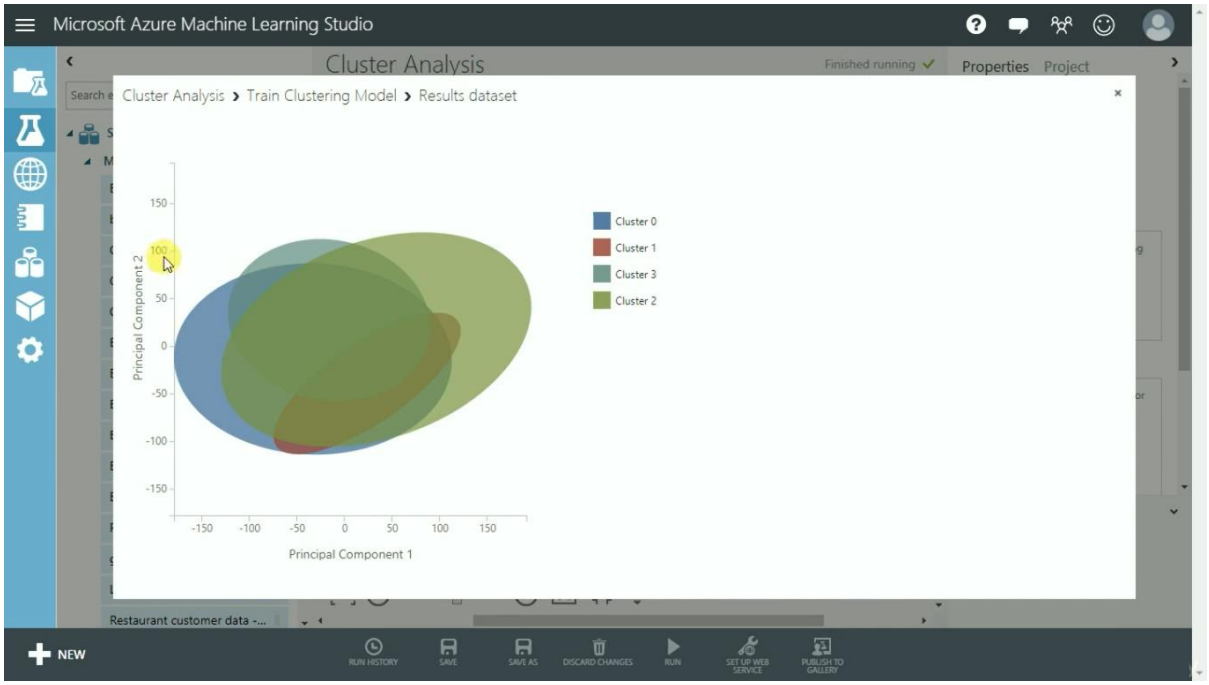


Visualize the module



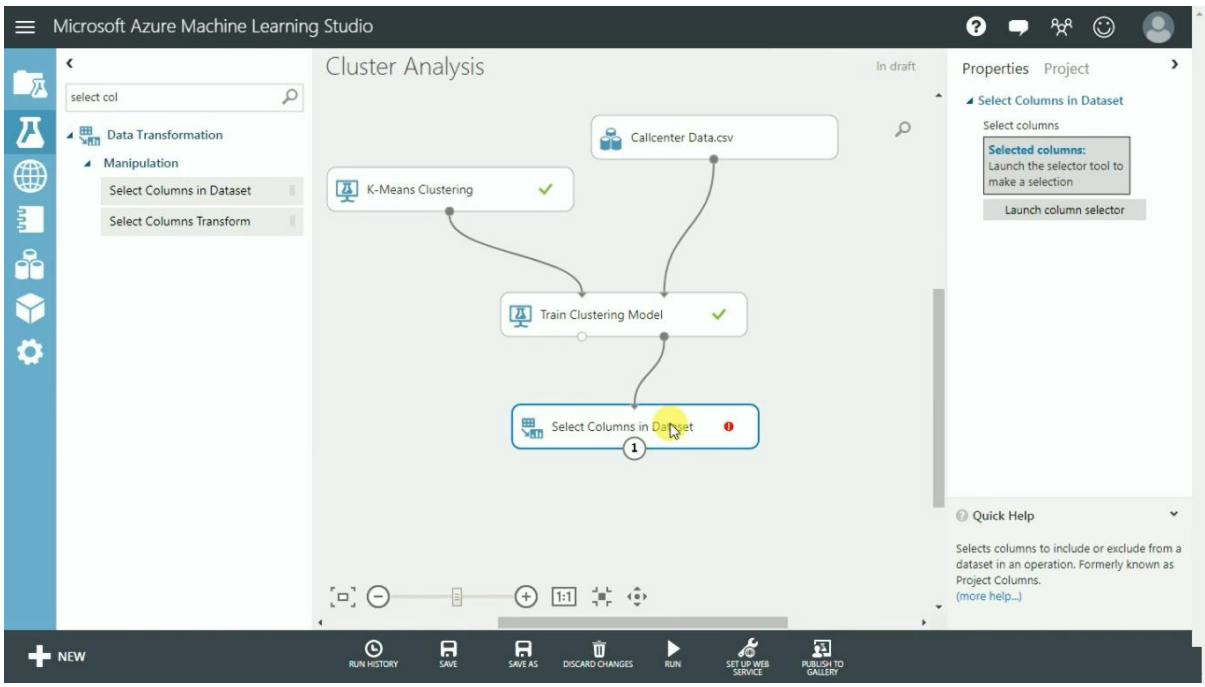
Result with two-dimension component, which seems not to be a good output

Let us implement the result in another method

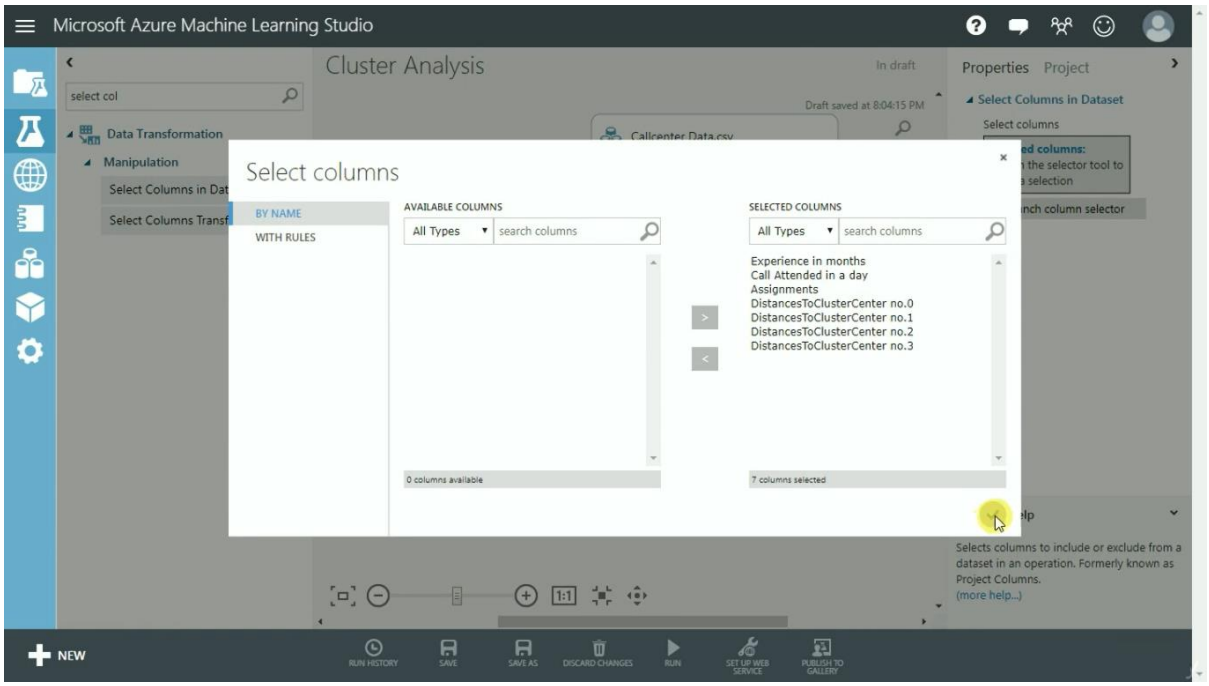


Select Columns in Dataset

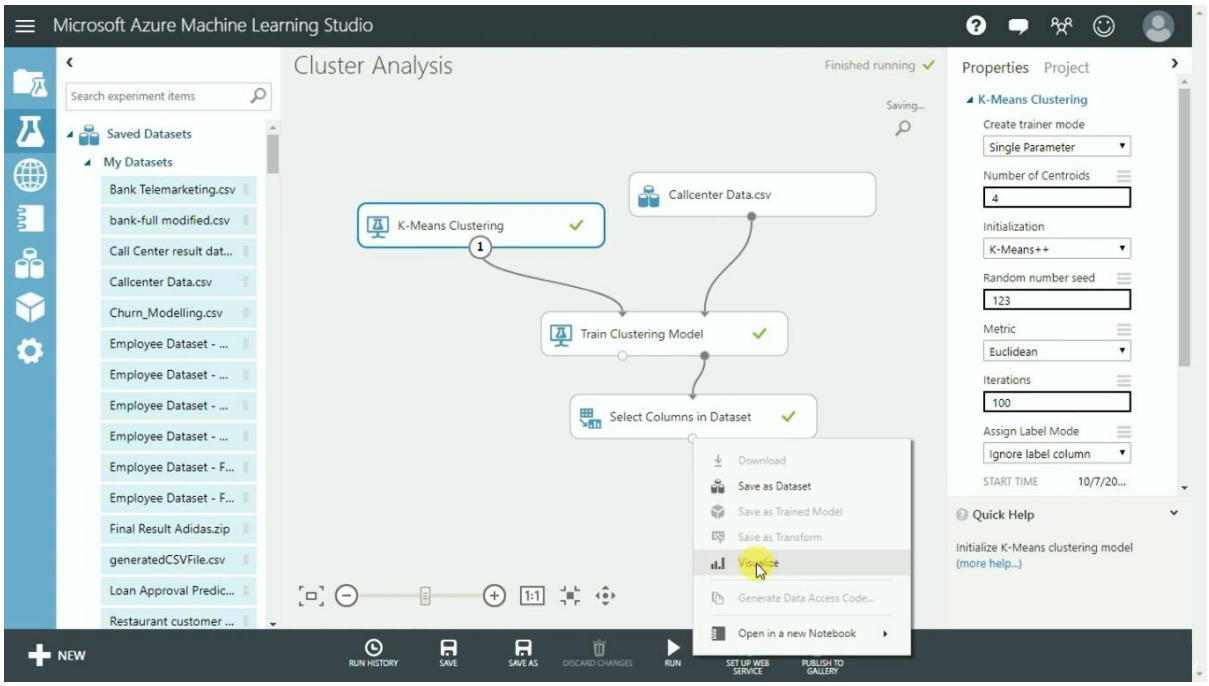
Insert select columns in dataset and connect with output node of train cluster model



Now select all the columns from launch column selector and click ok



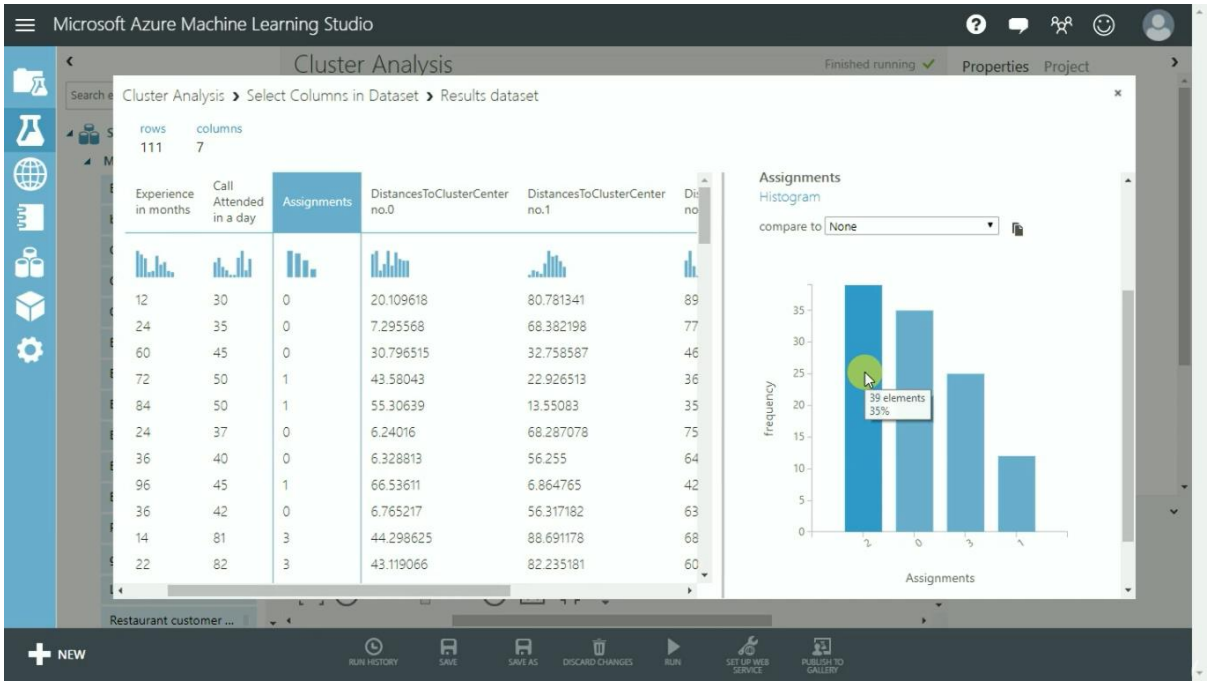
Run and visualize for result



View the result and observe the same

Now we assigned to different clusters in result. However, the objective is to identify

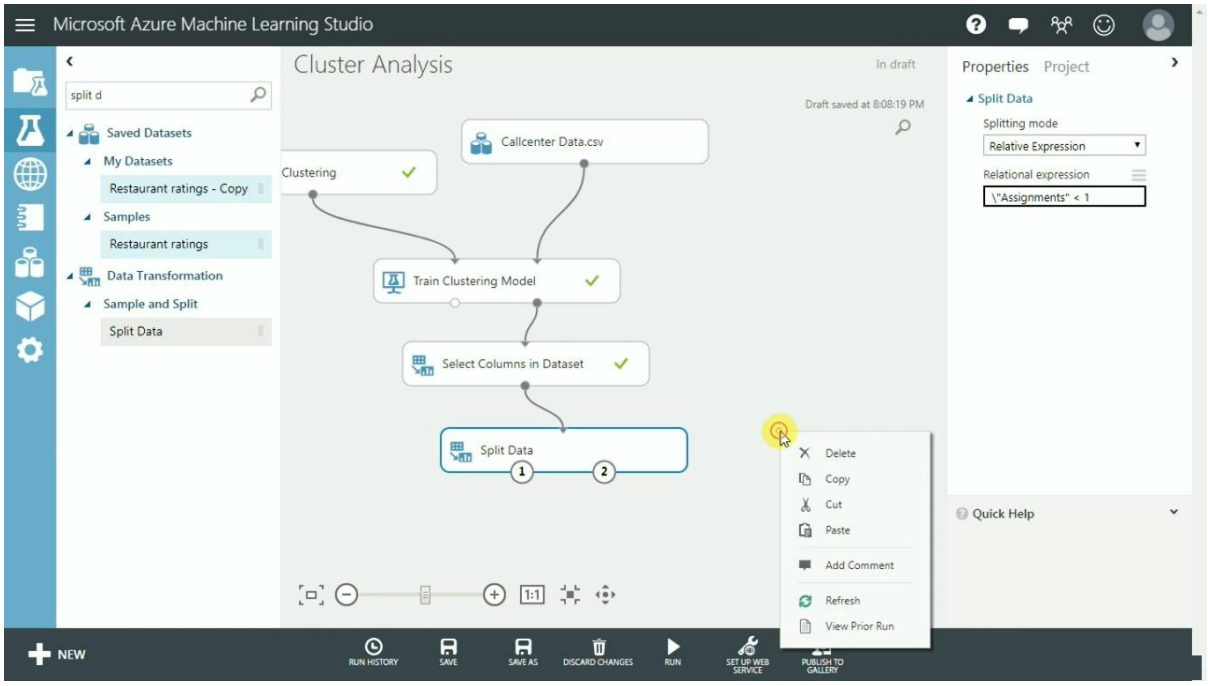
The strategy to improve the performance by clusters of employees



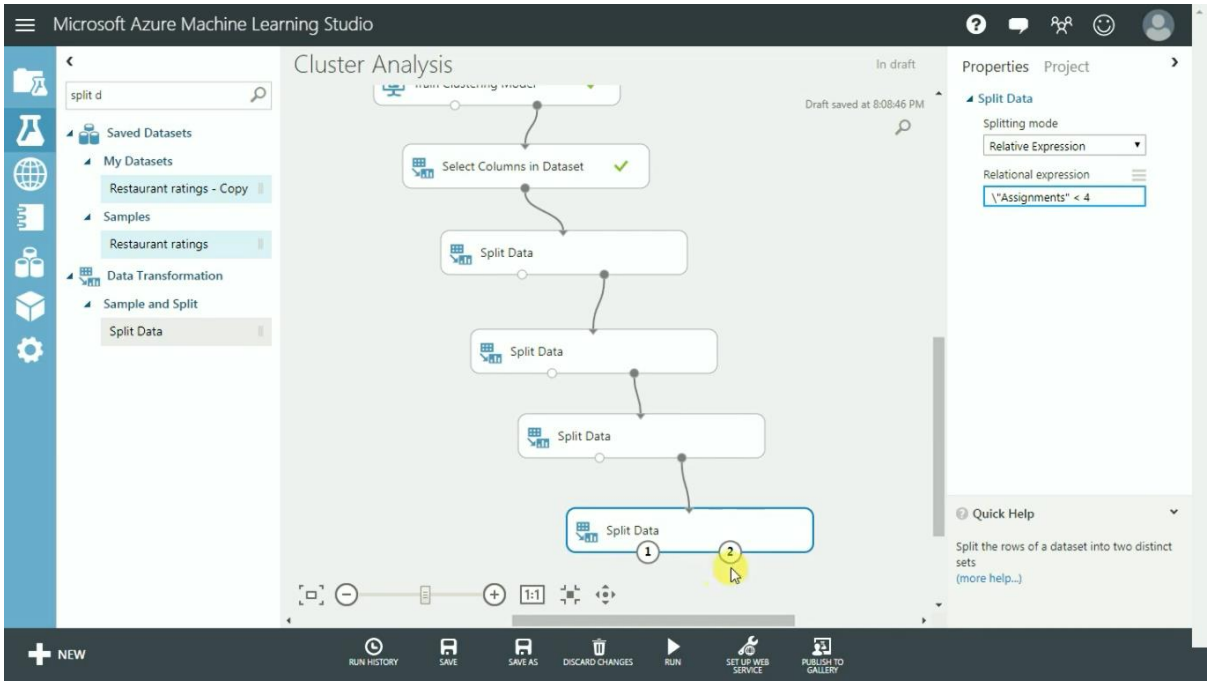
Dataset for Split data

Insert split data and change parameter as shown

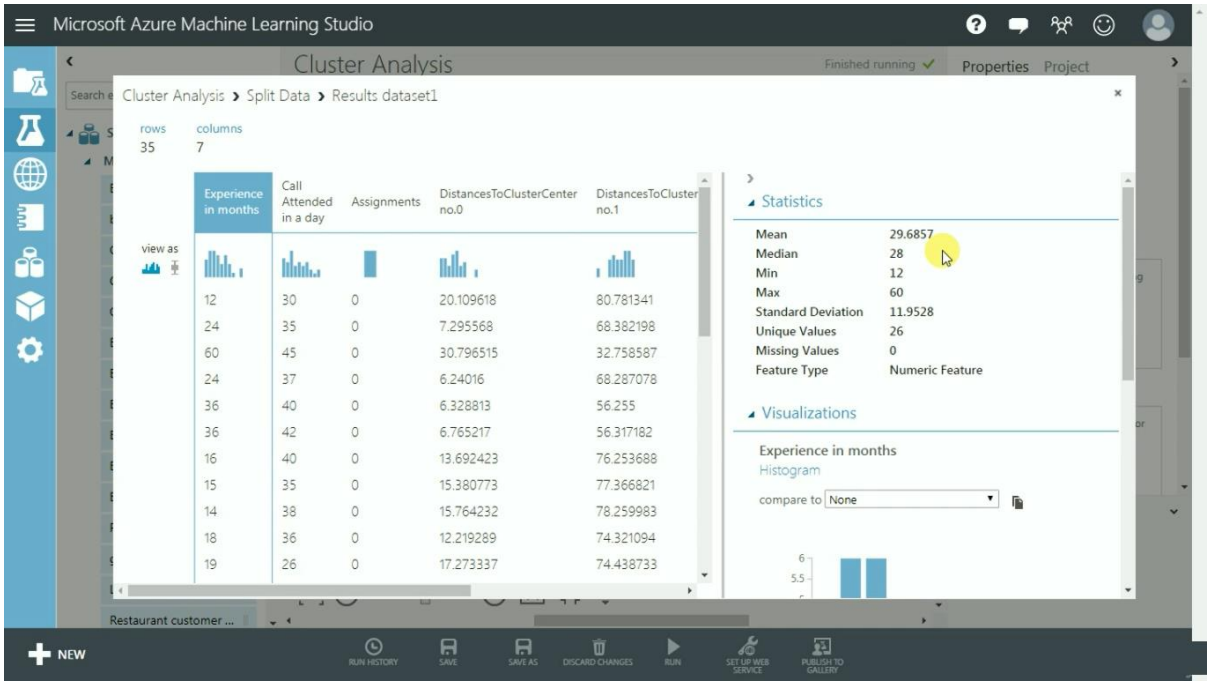
And add 3 more split data with Assignments 2 to 4 in Relational expression parameter



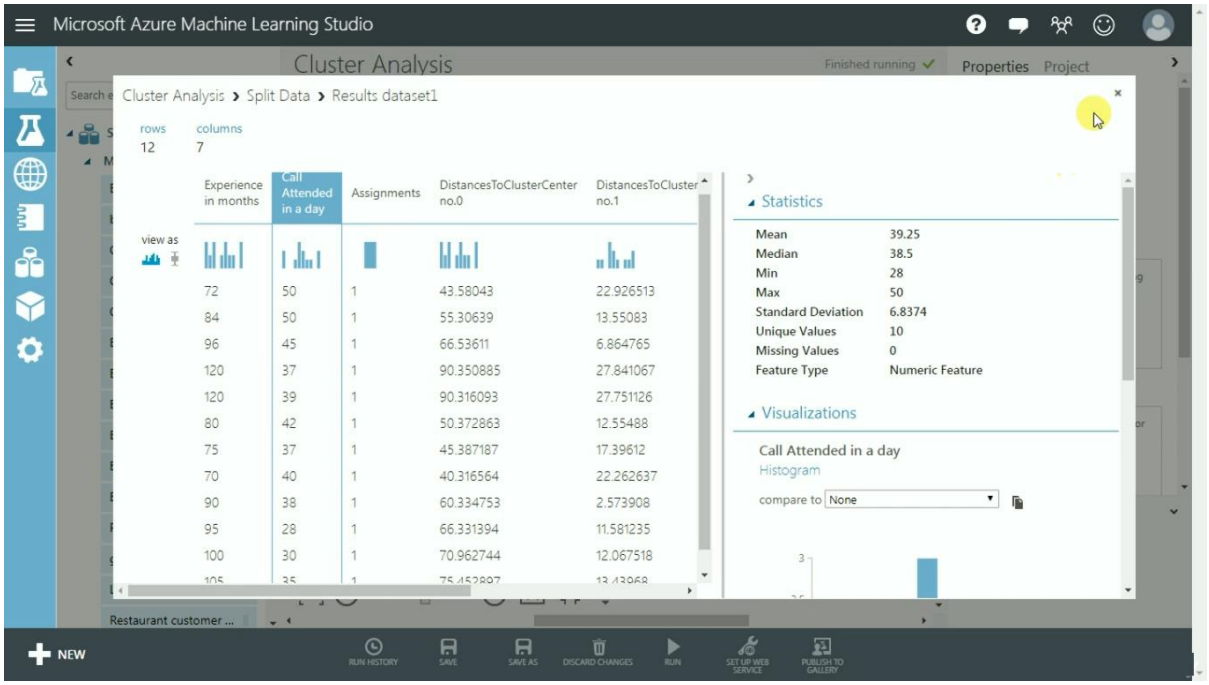
After inserting four split data in canvas, its ready to run and visualize the module one by one



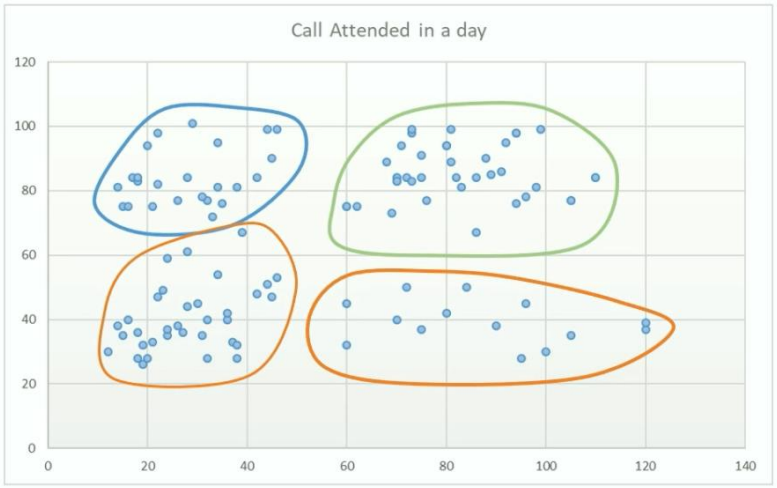
Visualize split data 1 and view the performance of the 1st group



Similarly view the other split results

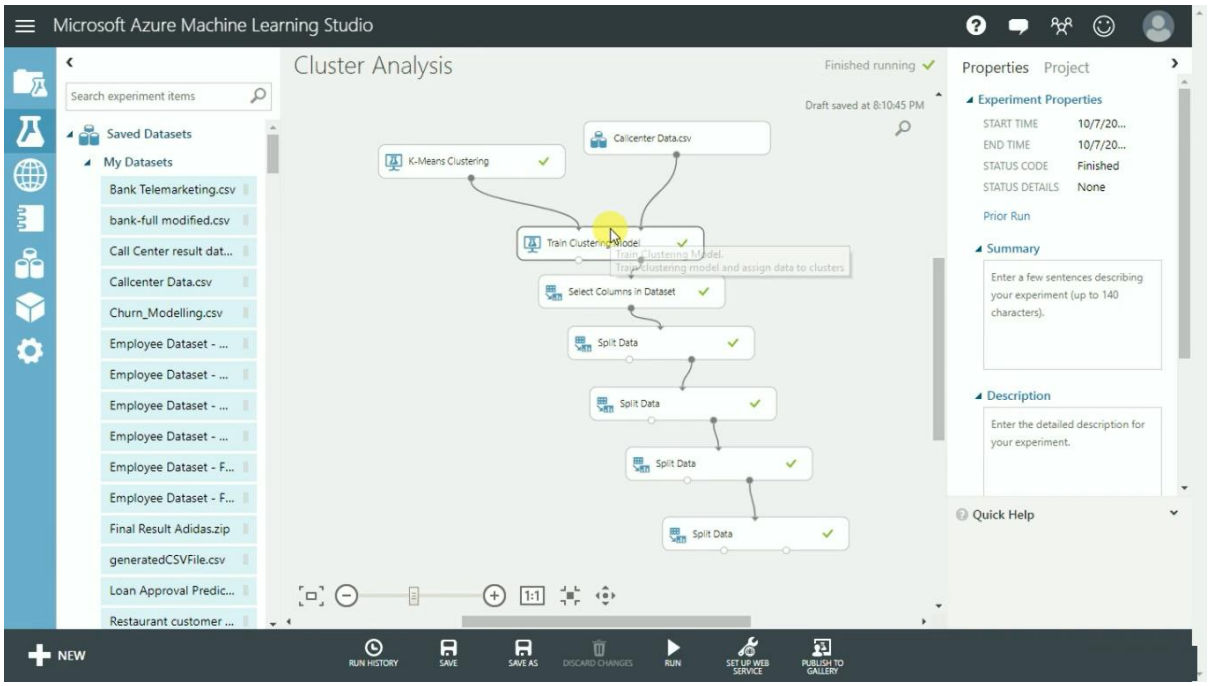


Clusters formed

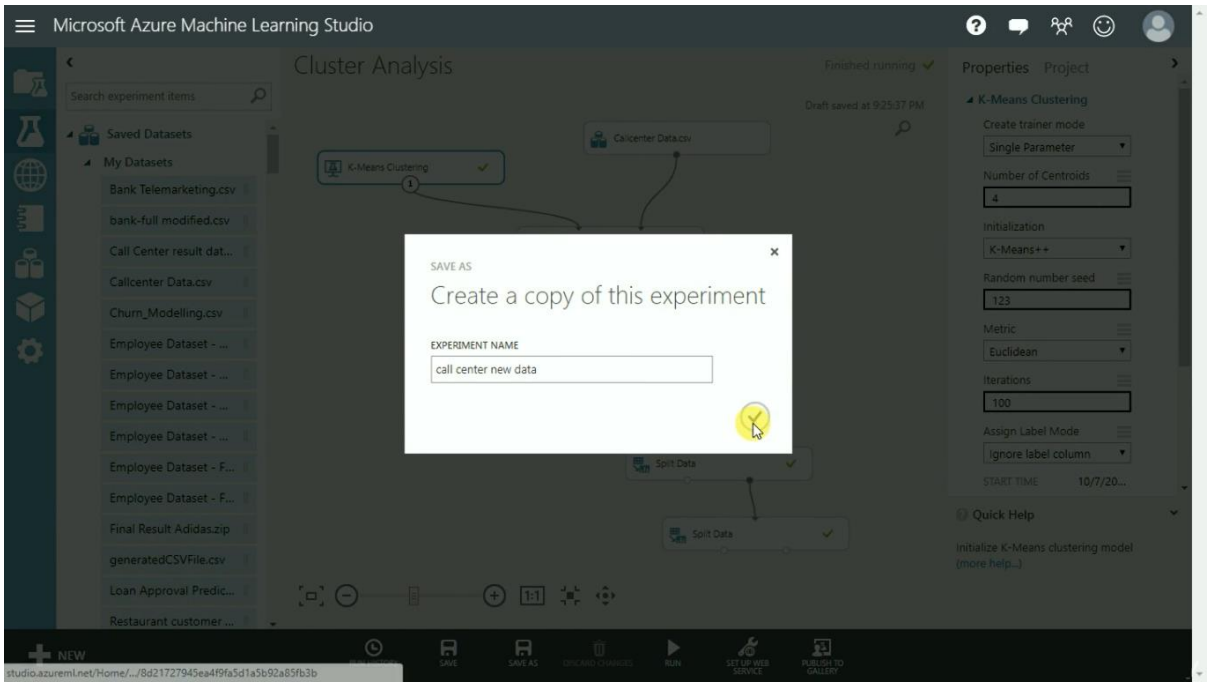


SCORE AND EVALUATE MODEL

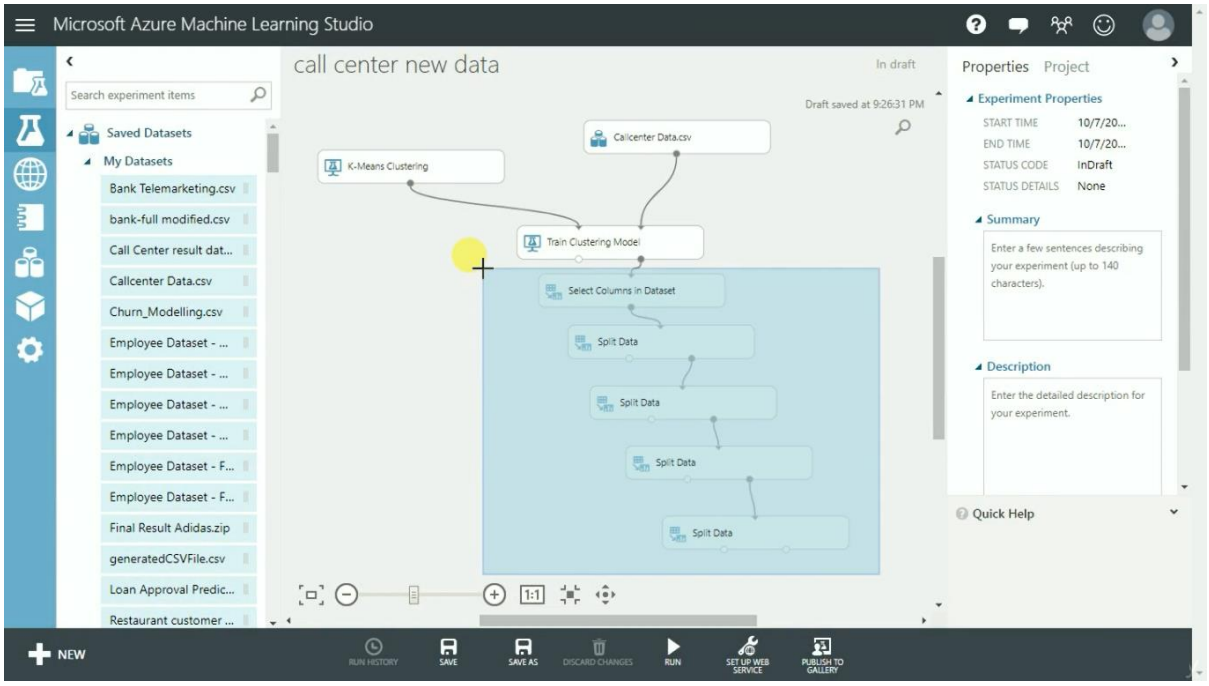
Previous experiment for identifying the strategy of cluster analysis in performance of employees



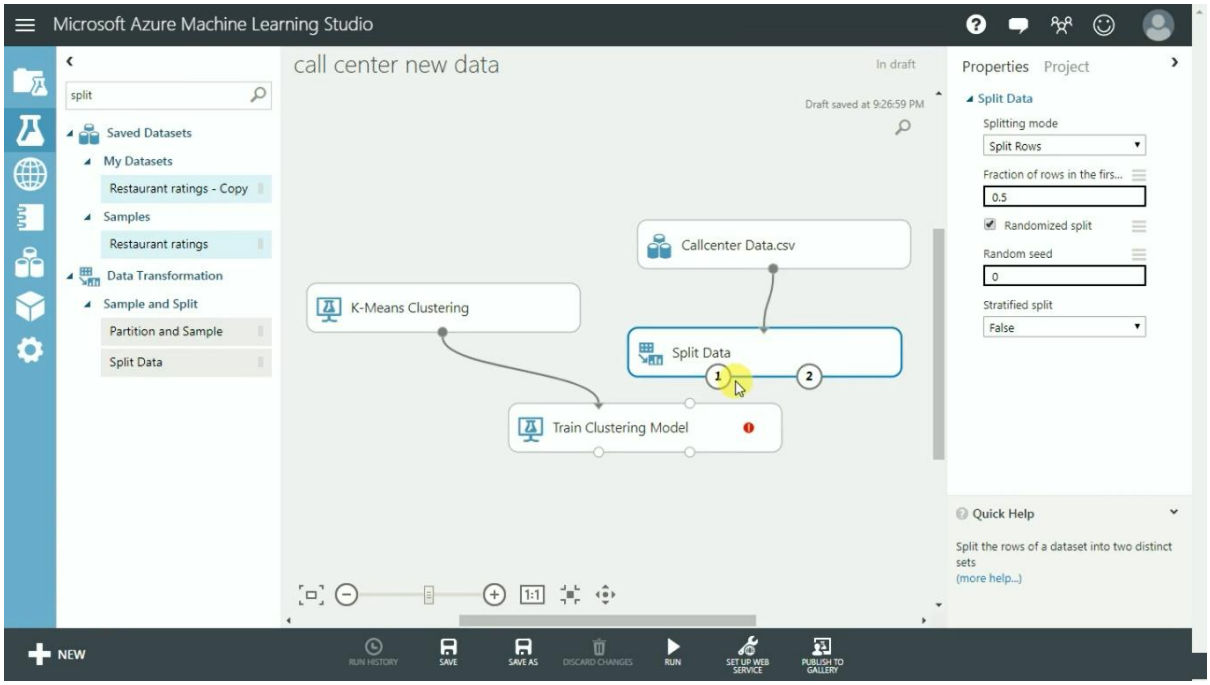
Save as call center new data



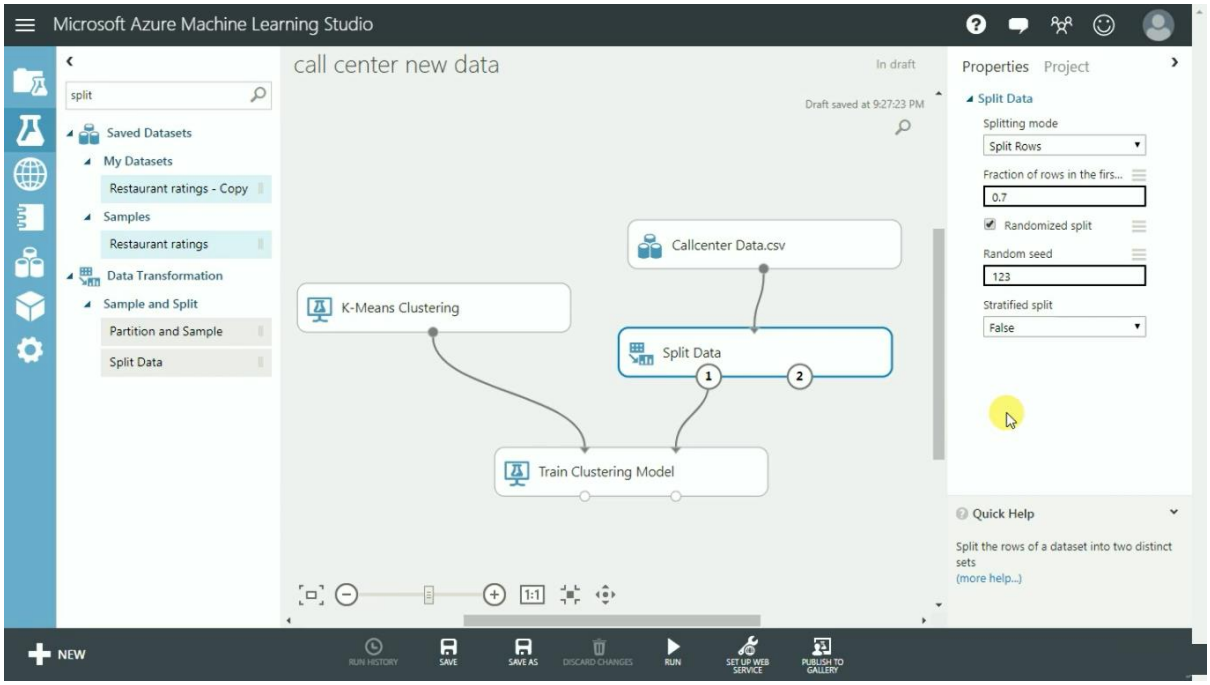
Delete split data and select columns in dataset and keep the required datasets



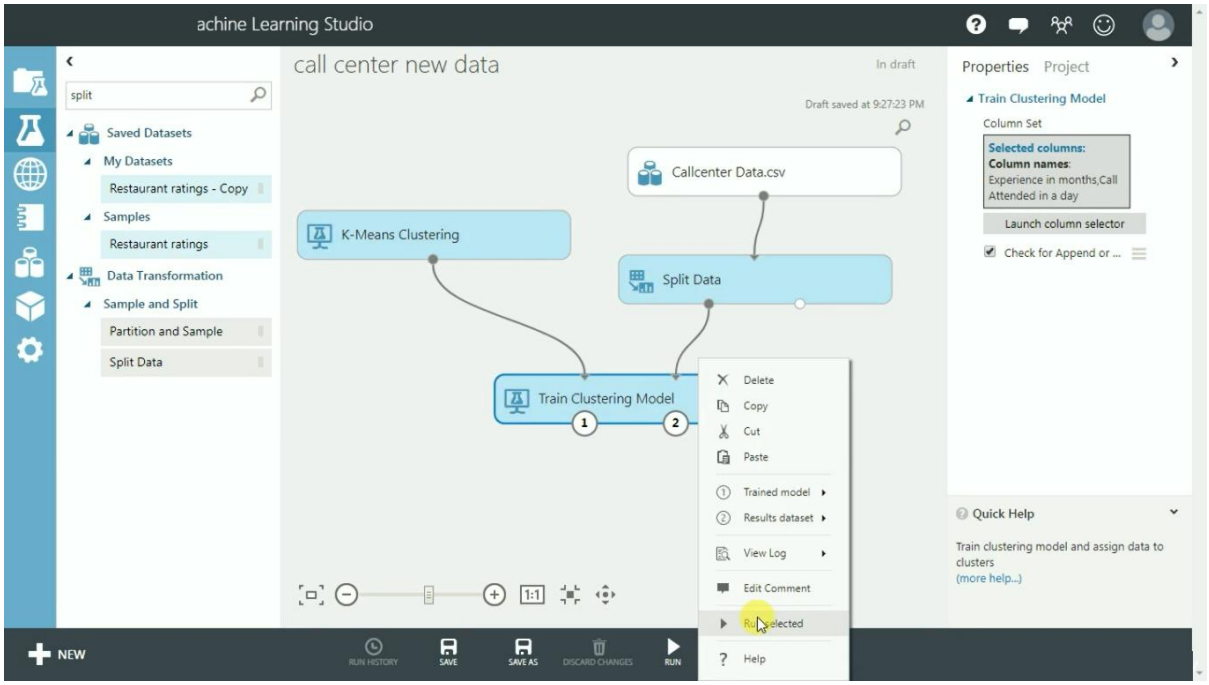
Add split data and connect the nodes from Call center dataset



Connect split data to train clustering model

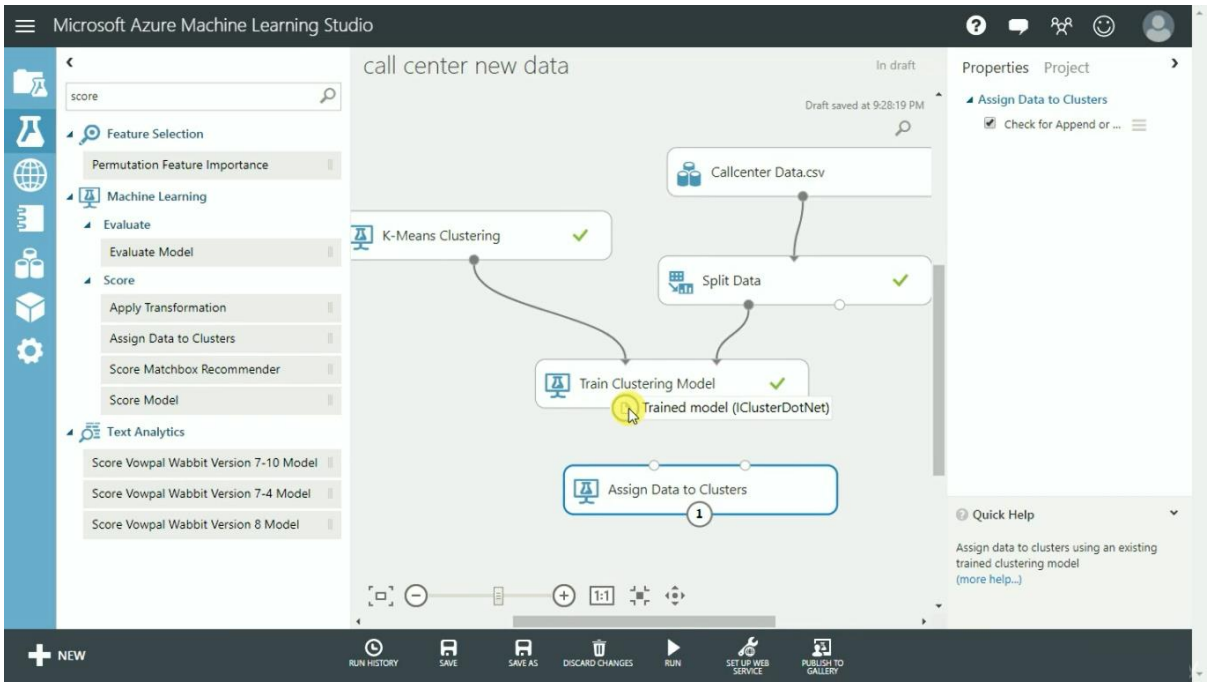


Now run the train clustering module

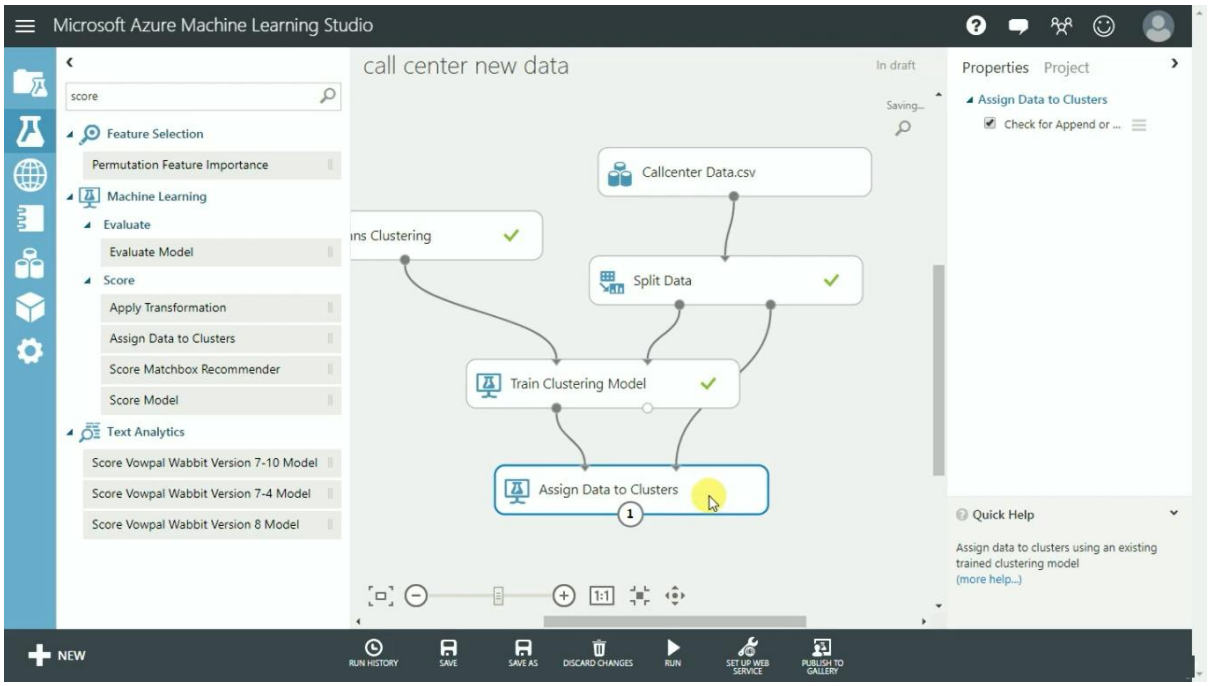


Assign data to Clusters Dataset

For cluster analysis model, apply Assign data to clusters instead of score model



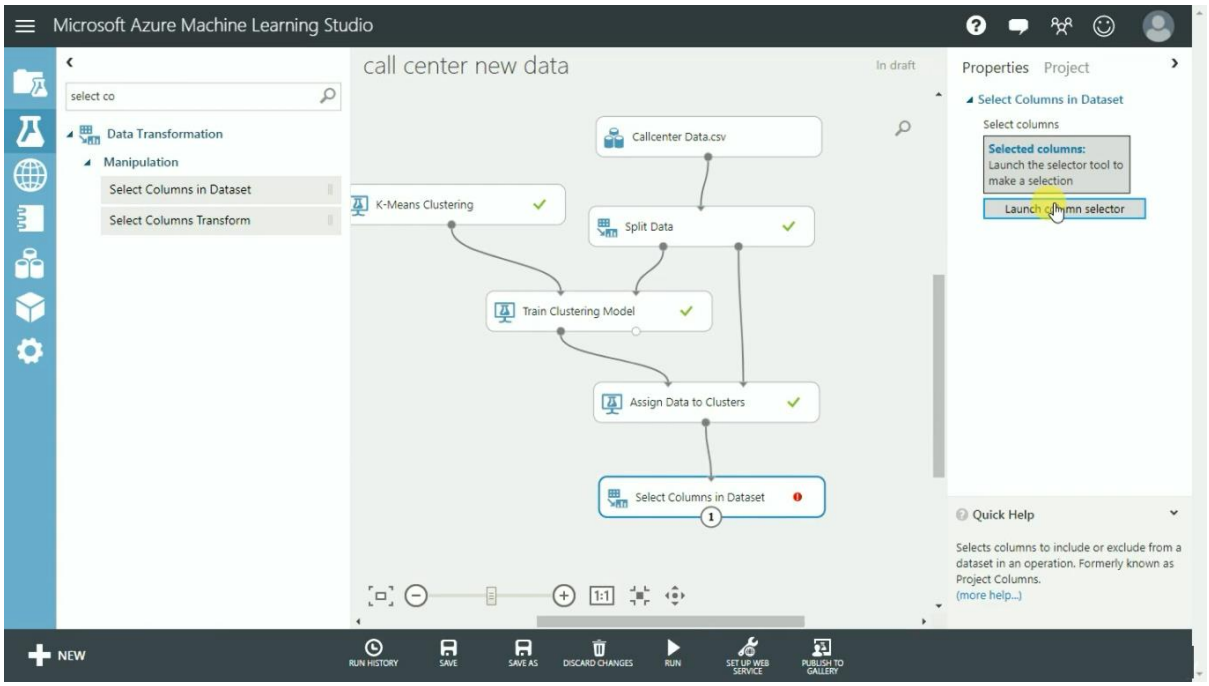
Connect the nodes from split data node and train clustering model



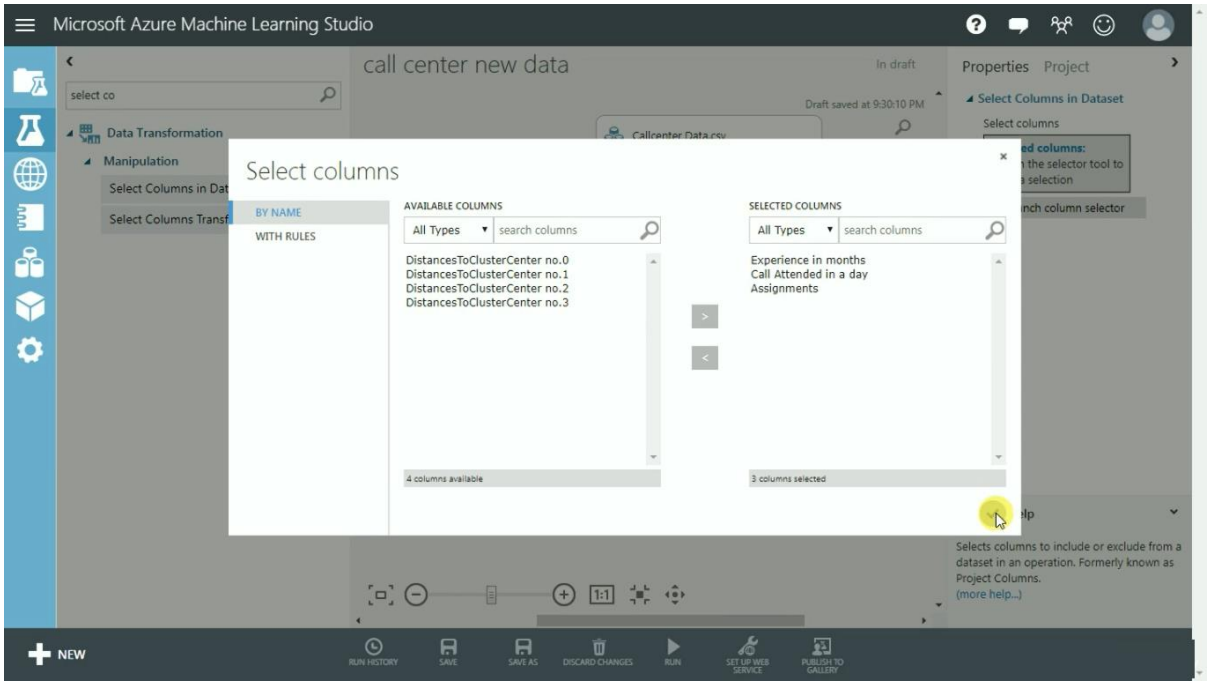
Run and visualize the result

Select columns in Dataset

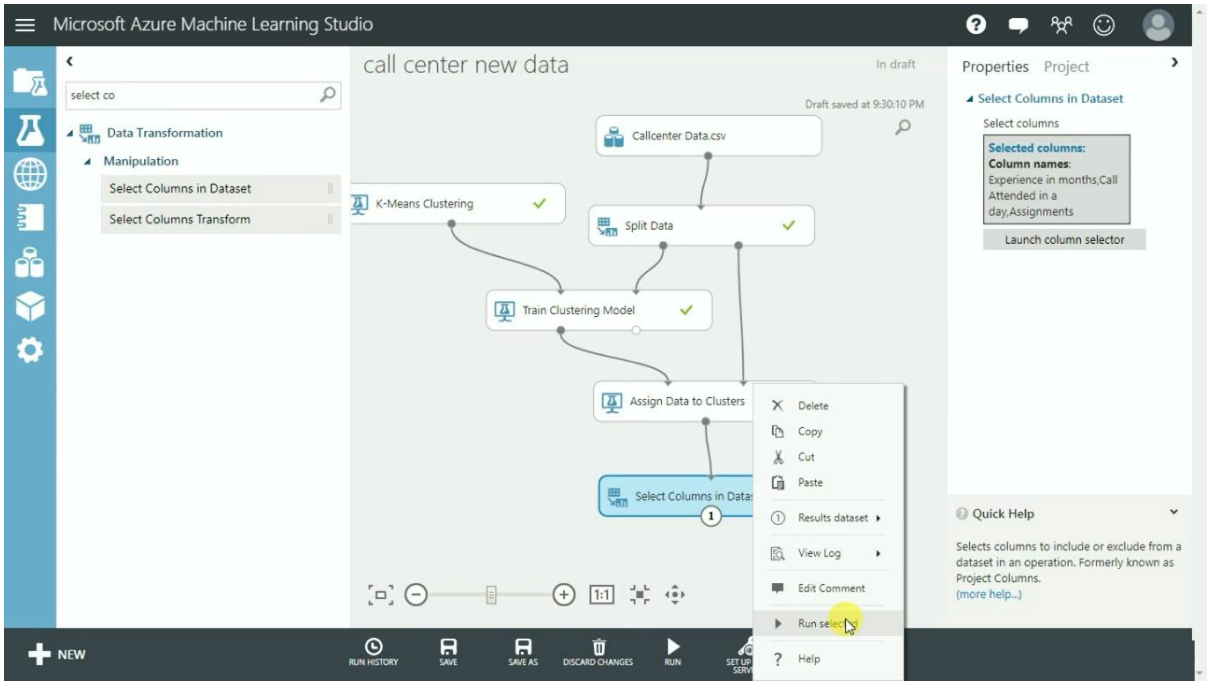
Now introduce select columns in dataset in picture



Launch column selector and select columns required

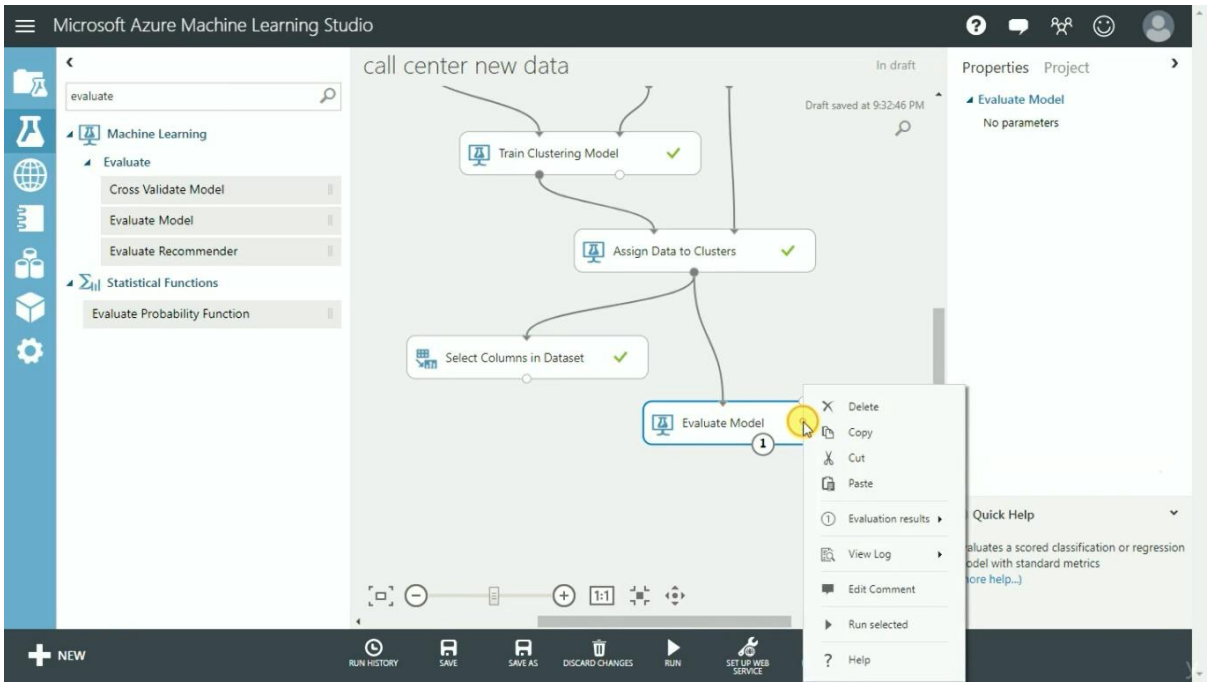


Run select column in dataset and get ready for evaluation

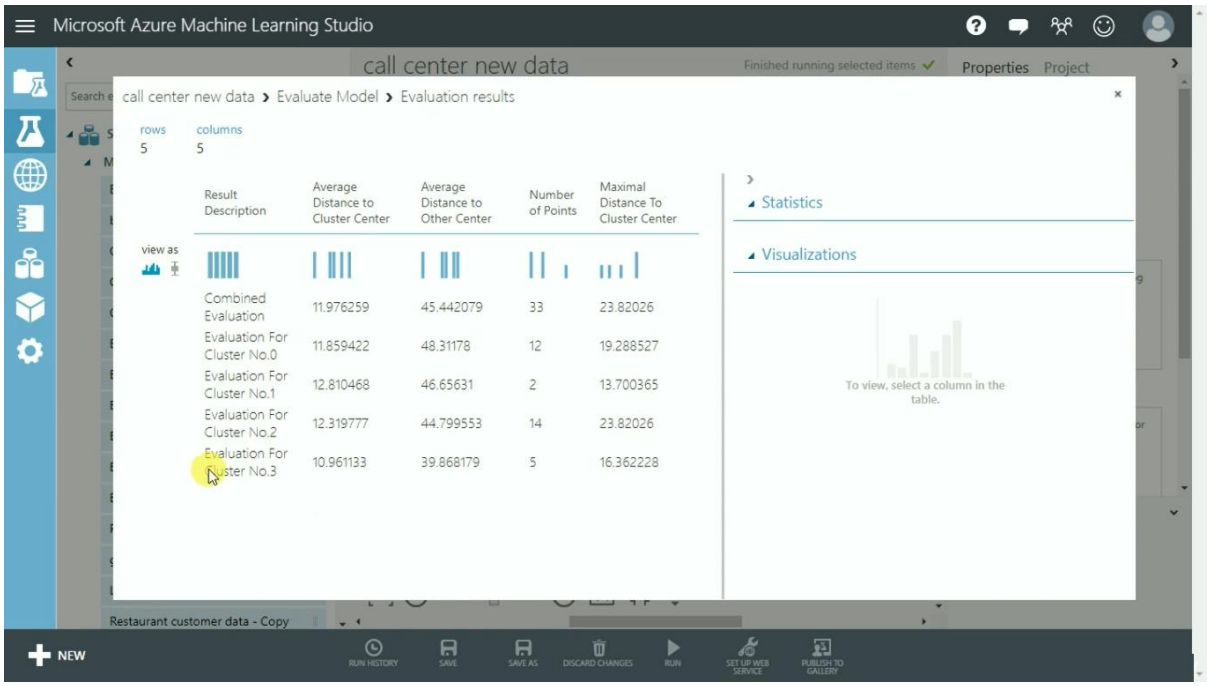


Evaluate Model Dataset

Now add the evaluate model and connect from assign data to clusters and run the module



Result obtained



Thus experimented score and evaluate model