**MS Azure Machine Learning Tutorial**

Clustering: K-means clustering

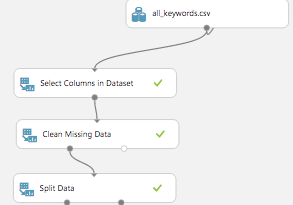
1. If you have not already done so, open a browser and browse to https://studio.azureml.net. Then sign in using the Microsoft account associated with your Azure ML account.

2. Create a new blank experiment, and give it the title **bikeshare**. Add your downloaded **all\_keywords.csv** to saved datasets, and add it to the experiment canvas.

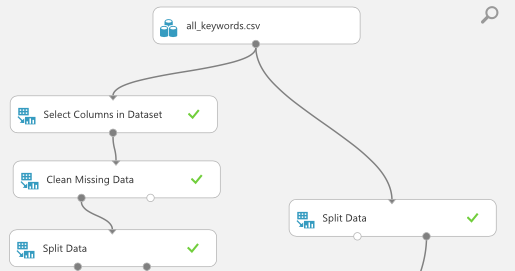
3. Find the **Select Columns in Dataset** module, and drag it to the experiment canvas under all\_keywords.csv. Select columns by names, select Start Station Latitude and Start Station Longitude.

4. Find the **Clean Missing Data** module, and drag it to the experiment canvas under the Select Columns in Dataset module. Select all columns. Set the minimum missing value ratio to 0 and set maximum missing value ratio to 1. Select custom substitution value for the cleaning mode and set replacement value to 0. Uncheck Generate missing value.

5. Find the **Split Data** module, and drag it to the experiment canvas under the clean missing data module. Connect to the cleaned dataset from the clean missing data module (Left most). Select Split mode to Split Rows, set fraction of rows value to 0.5. Checked Randomized split, set random seed to 0. Select Stratified split to False. Your experiment canvas should look like this for now.



6. Duplicate the Split Data model and put it on the right of the first one.



7. Find the **Train Clustering Model**, and drag it to the experiment canvas. Select columns: Column type: Numeric, All. Checked the box of check for append

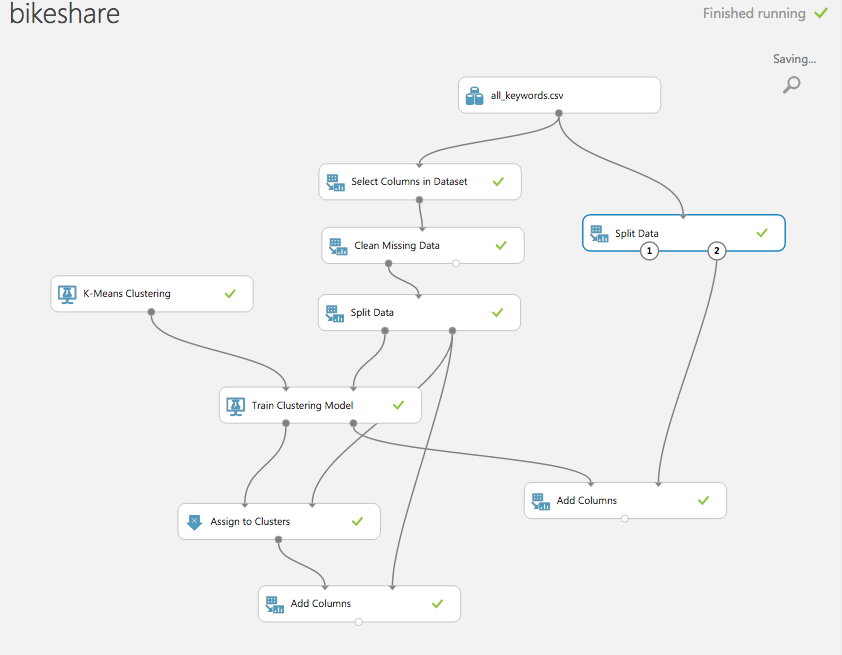
8. Find the **K-Means Clustering**, and drag it to the experiment canvas, connect the output of this model to the left most input of the Train Clustering Model, and connect the left most output from Split Data to the right most input of the Train Clustering Model.

9. Find the **Assign to Clusters**, and drag it to the experiment canvas under the Train Clustering Model. Column select, all columns, column type: Numeric, All.Connect the top left dot to the train clustering model and connect the top right most input to the first split data model right most output.

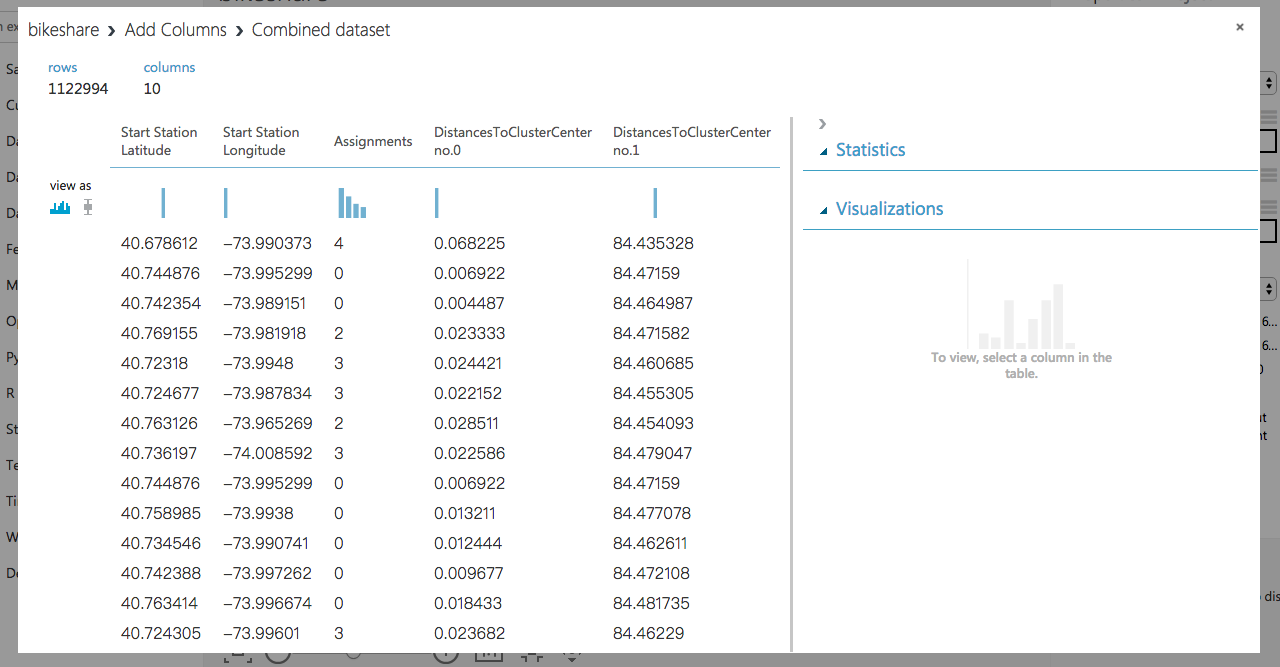
10. Find the **Add Columns** module, and drag it to the experiment canvas under the Assign to Clusters. Duplicate another one and put in under the second split data model

11. Find the **Evaluate model** and drag it to the experiment canvas under the score model.

12. Your experiment canvas should look like the following screenshot. Save and run the experiment.



13. Visualize the left one Add Columns model and the K-means clustering algorithm cluster the 5 clusters information just show below.

**

*[Azure ML tutorial end]*