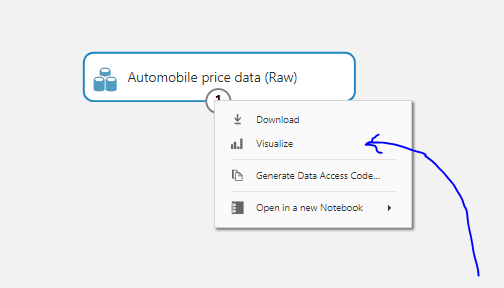
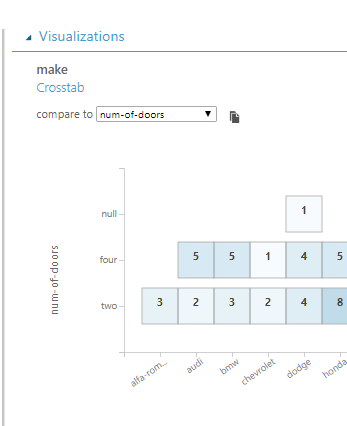
Toptim Kelly Duffy

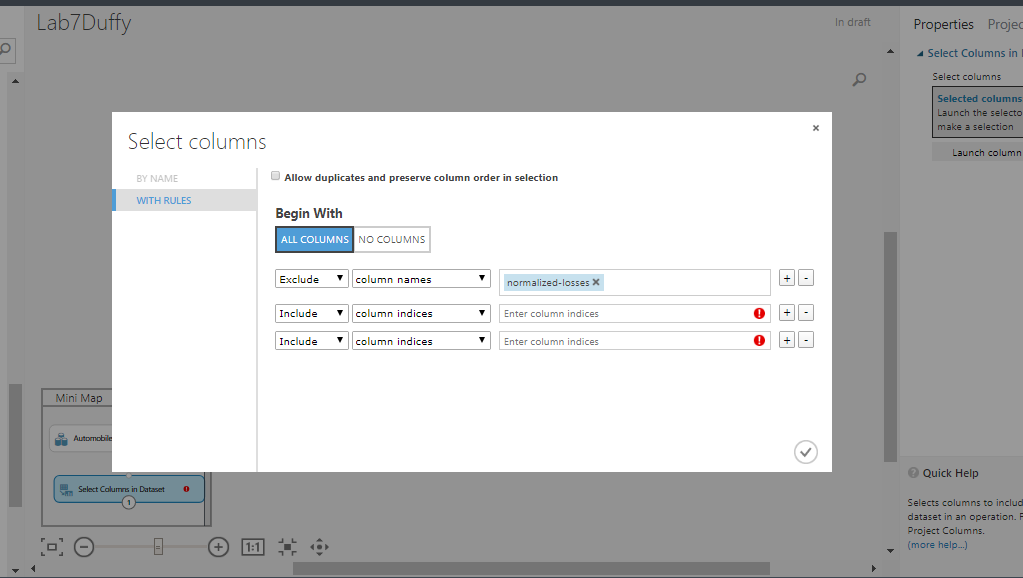
Lab 7 - Microsoft Azure Machine Learning

Starting off with Automobile Price Data(raw) that can be found on the left side column under the samples tab. By clicking ‘1’ in the circle then the visualize option, you can see all the columns and rows of the data set. In addition to viewing the table, you can visualize the data as scatterplots, box plots, histograms, multibox plots, and crosstabs for example. You can also pick a single data set/column and compare it to another column and ML will visualize the comparison accordingly.

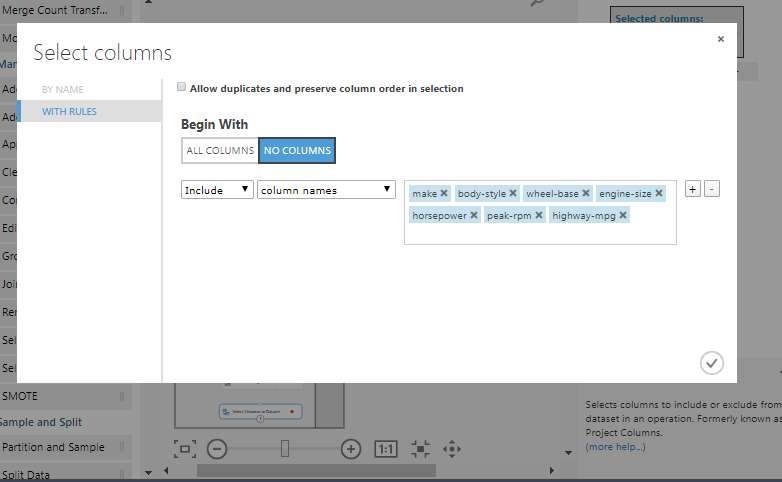


* Make by Number of Doors for example as displayed on a crosstab visualization.

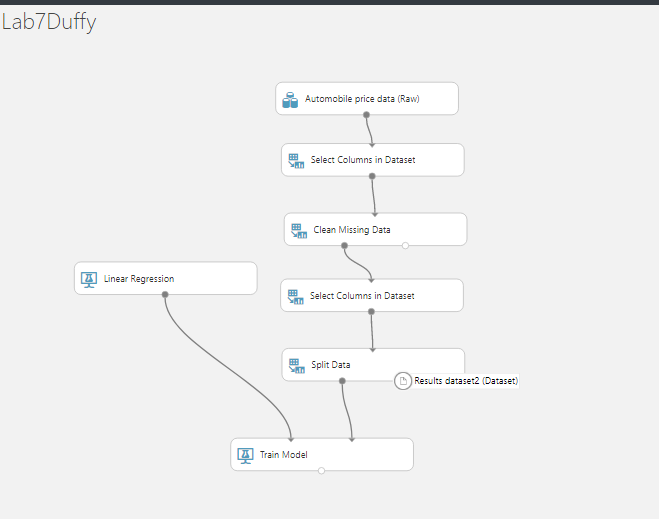


One of the first steps is to exclude normalized losses. This can be done in the search bar by searching ‘select columns in the dataset’, then ‘launch column director.’ You can then use the all columns option and exclude normalized-losses.

Normalized losses is the relative average lost payment per insured vehicle by year. Since we are predicting prices, we remove normalized losses as it can make the outcome muddy since we are omitting the effects of non recurring charges or gains from the insured vehicles. Afterwards, we would clean the missing data using the data manipulation tab to ensure our dataset reflects this exclusion of data which can be found using the search option to the top left.

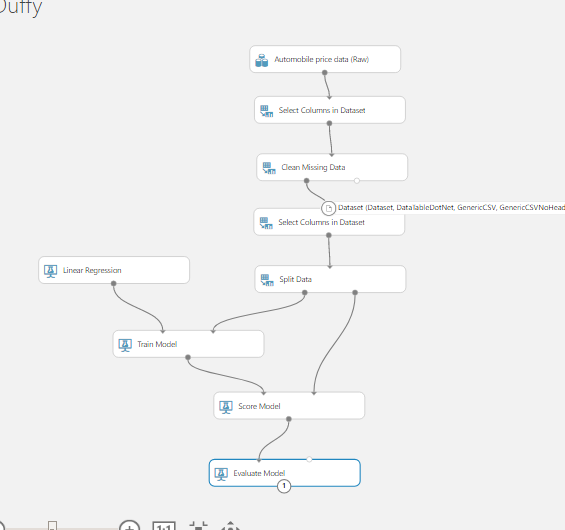


After cleaning the missing data, the columns that were included were make, body style, wheel base, engine size, horsepower, peak rpm, and highway mpg. Price was left out of the data set as we are trying to predict it using a simple linear regression model.Confirm with the checkmark. Connect all the models together dragging the arrow to other ones in order they were created.

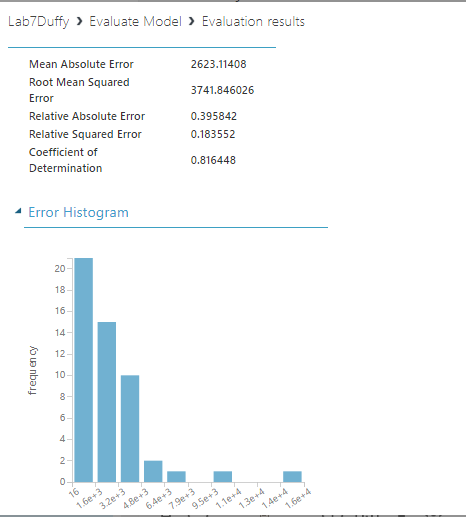


The train model is set to ‘price.’ 

Linear regression is set to the train model that is specified to the ‘price’ column as we are trying to create a linear regression model to predict price, thus the connection.



Connect split data to score model. A score model is like a scorecard/table in which all the elements that influence the outcome are separated into individual characteristics, as shown in the model. The Score model trains the classification or regression model. It is pretty important to include as it brings everything together. Lastly, evaluation of the model will allow us to run the model and visualize the results.



The new visualized results.