**LAB MANUAL**

**Day – 2 | ARTIFICIAL INTELLEGENCE**

**Lab – 2.6**



A close up of a sign

Description automatically generated

**Understanding and Demonstration of Regression problems using ML Studio Classic**

**Understanding and Demonstration of Regression problems using ML Studio Classic**

**Objective:**

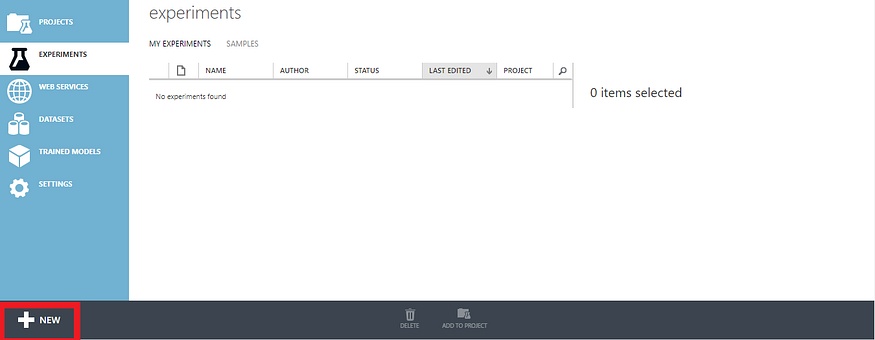
To keep it simple we need to predict a straight line which best fits the data. So without going into math behind linear regression, we will train a model in Azure ML Studio without writing any piece of code.

**Steps:**

**Azure ML Studio (Classic )**

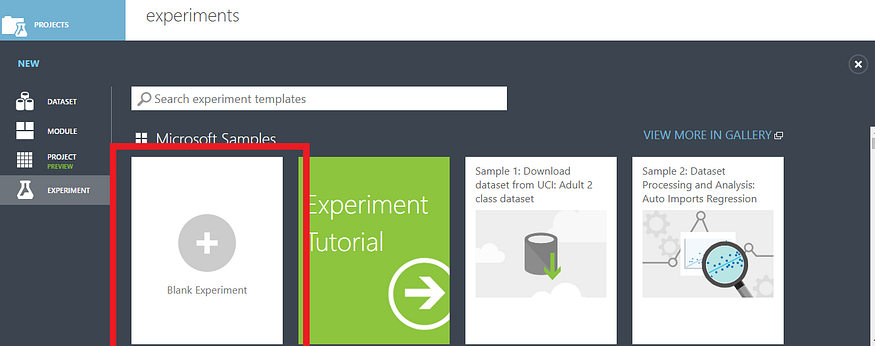
We can freely sign up [here](https://studio.azureml.net/) without providing any credit card details

**Azure ML studio** is a collaborative, drag and drop tool where we can build, test and deploy machine learning models. Azure ML studio looks like below once we sign in.

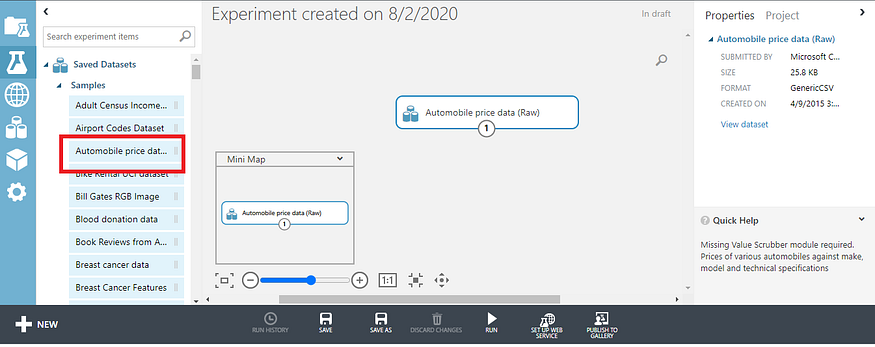


To know more about Azure Machine Learning Studio , Please refer official documentation from microsoft :  
[https://docs.microsoft.com/en-us/azure/machine-learning/studio/what-is-ml-studio#](https://docs.microsoft.com/en-us/azure/machine-learning/studio/what-is-ml-studio)

Without wasting anytime let us create an experiment. To create an experiment, Select Experiments on the left menu and select Blank Experiment.Azure provides Many sample Experiments to work with, for now we will create our own experiment by selecting blank Experiment.

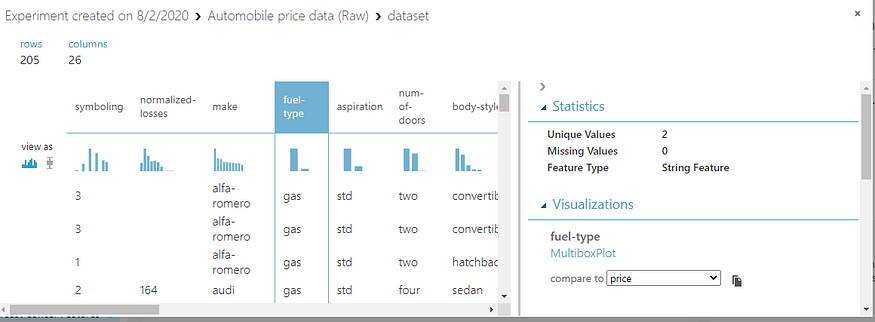


Azure provides sample datasets under saved datasets. Any of these datasets can be dragged into experiment. You can upload your own datasets from local file system .Here we will work with builtin data set .Drag and drop Automobile price data set into experiment canvas .



**Exploring Data**

We can have a quick visualization of our data on right clicking our dataset and selecting visualization.We can select any column and have a glance at some stats like unique values, missing values, feature types.

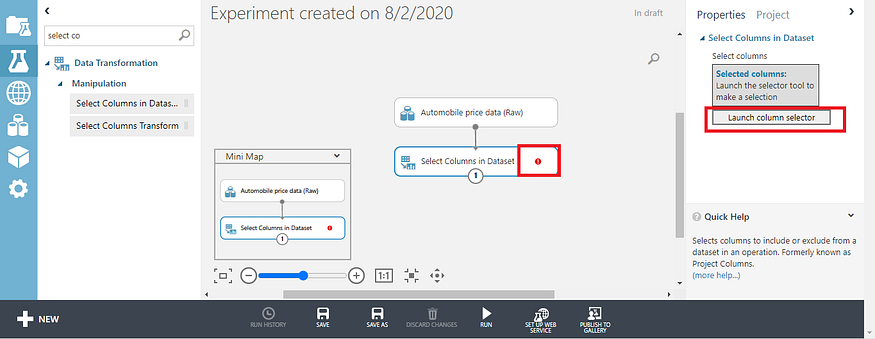


Under visualizations we can compare two columns and have a quick view at box-plot analysis or histograms.So with one click on data we are able to analyse each feature ( column ), compare two features using many visualization techniques without any code.

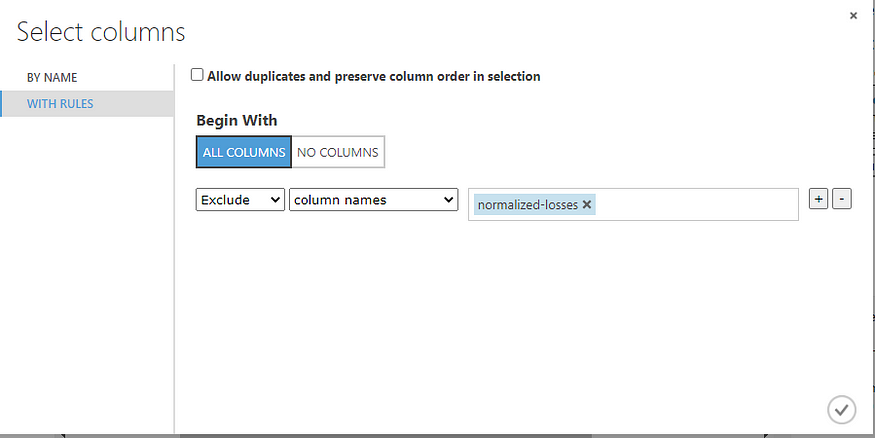


**Selecting Columns**

Next we need to select columns on which our model is going to be trained . Here we can exclude some columns which are unnecessary or don’t provide much information . We can search the Select columns in Dataset from modules and connect to the dataset output .Select Launch Column Editor to include or exclude columns we needed.



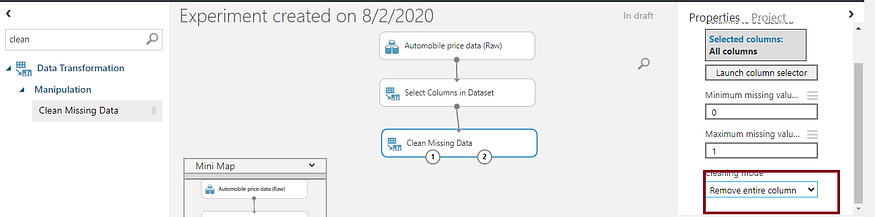
We are excluding column normalized-losses column here as it has more missing values .



**Data Cleaning**

Data cleaning is most important step in Machine Learning process.Most common methods of data cleaning include  
1.Handling Missing values  
2.Dealing with outliers  
3.Data standardization/normalization  
and many more data transformation techniques

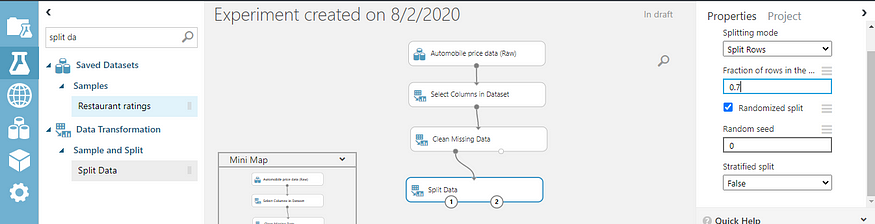
Next, we need to clean Missing Data. Drag and drop Clean Missing Data and connect from Select Column in Dataset. On selecting Clean missing data module, from the properties on right hand side we can select specific columns or select All columns.



From the cleaning mode we can select an option how to clean missing data. we can replace missing values with mean, median, mode specific to problem we are solving. Here we are just removing entire row .  
To know more about handling missing values in Azure ML Studio refer below  
[https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/clean-missing-data#](https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/clean-missing-data)

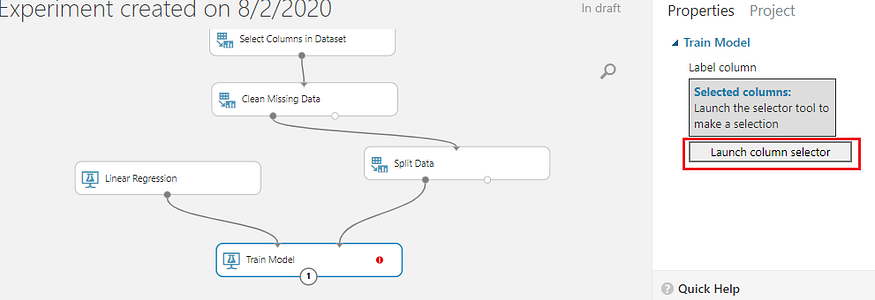
**Split Data:**

Next, we need to split data for training and evaluating (test data).Drag and drop split Data module and connect the input to output of clean missing data.From the properties menu fraction of rows to be split (0.7) .So here our data will be divided into Train data(70 %) and test data (30%).



**Train a model:**

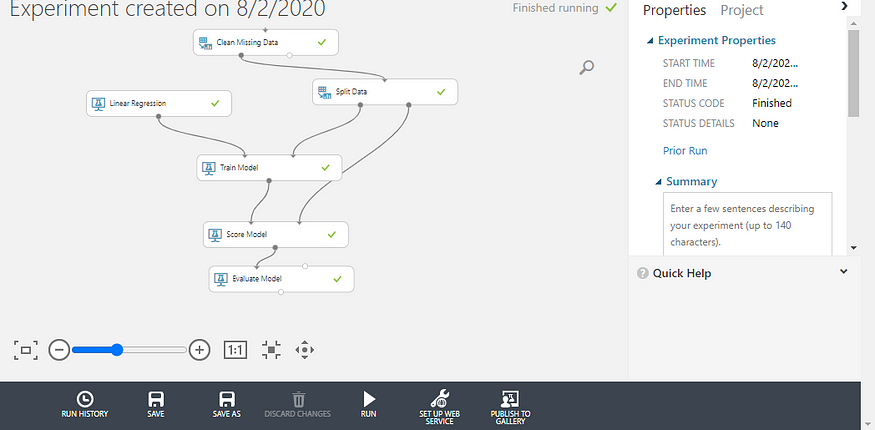
Here we need to select the machine learning algorithm, to train a model which is linear regression in this case.



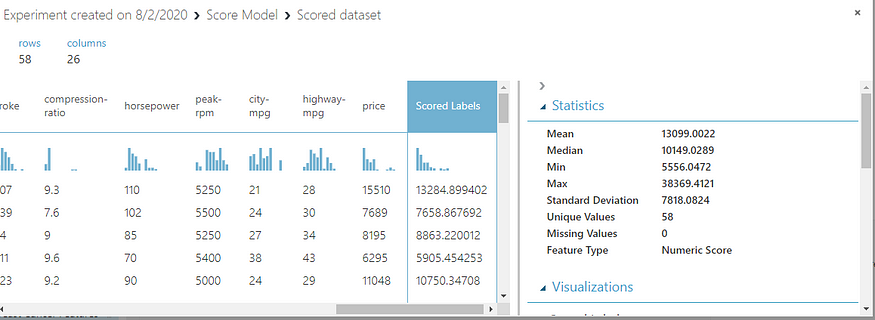
Drag and drop linear regression and Train modules (under machine learning modules ). Connect the linear regression module to the first input of Train model and first output of split data to second input of Train model.So now Train model (module ) trains linear regression with 70% data ( Train data ).Launch the column selector for train model and select the column which we need to predict, in this case it is price column we need to predict.

**Score & Evaluate Model:**

So now we finished our model training and we need to predict for test data.Drag and drop Score model ( under score )and connect output of train model to first input of score model and second output of split data to second input of score model (Test Data).Now place evaluate model (under evaluate )and connect to score model.

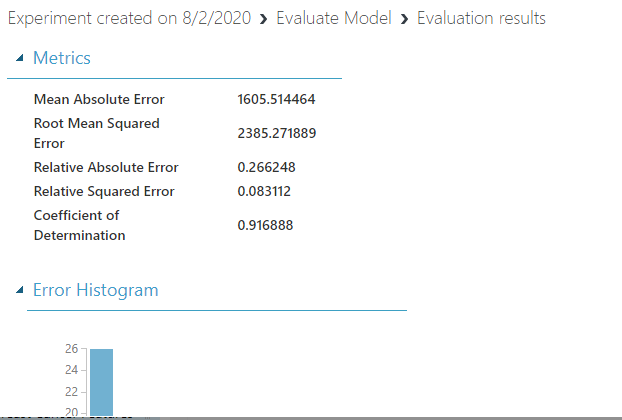


Click on Run to complete the experiment. Once finished we can see scored labels(Predicted Price) by right clicking on score model -> Dataset-> Visualize



We can compare statistics of original prices and predicted prices ( Scored labels ), like mean and standard deviation .

We can check evaluation metrics by right clicking on Evaluation model->Evaluation results -> Visualization



**Coefficient of Determination:**

This coefficient is commonly known as R-squared is represented as a value between 0 and 1, where a value nearer to 1 indicates a good fit.