

Source: C# Corner (www.c-sharpcorner.com)

PRINT

Article



Creating An Experiment In Azure Machine Learning Studio

By [Abdul Rasheed Feroz Khan](#) on Jan 13 2017

Introduction

This article is a continuation of my previous series of articles and here we will be discussing about Azure Machine Learning Studio. We will create an experiment, and we will be discussing more with the sample datasets and other Workflows of the experiment.

Note

Surf the articles given below before you work on this experiment creation with Machine Learning Studio.

- [Click here](#) for why we should choose Microsoft Azure Machine Learning.
- [Click here](#) for starting with Machine Learning on Microsoft Azure – Part One.
- [Click here](#) for starting with Machine Learning on Microsoft Azure – Part Two.

About Azure Machine Learning Studio

Azure Machine Learning Studio is the IDE, which we will be using here for Machine Learning. We will be creating and deploying our Azure Machine Learning Solutions with help of this Studio. It supports us with all the phases of development, where the interface is extremely easy to work on, we can just drag and drop and we can see the graphical view of how the data flows. We need not write any program codes here but still the programming languages like Python are supported in it.

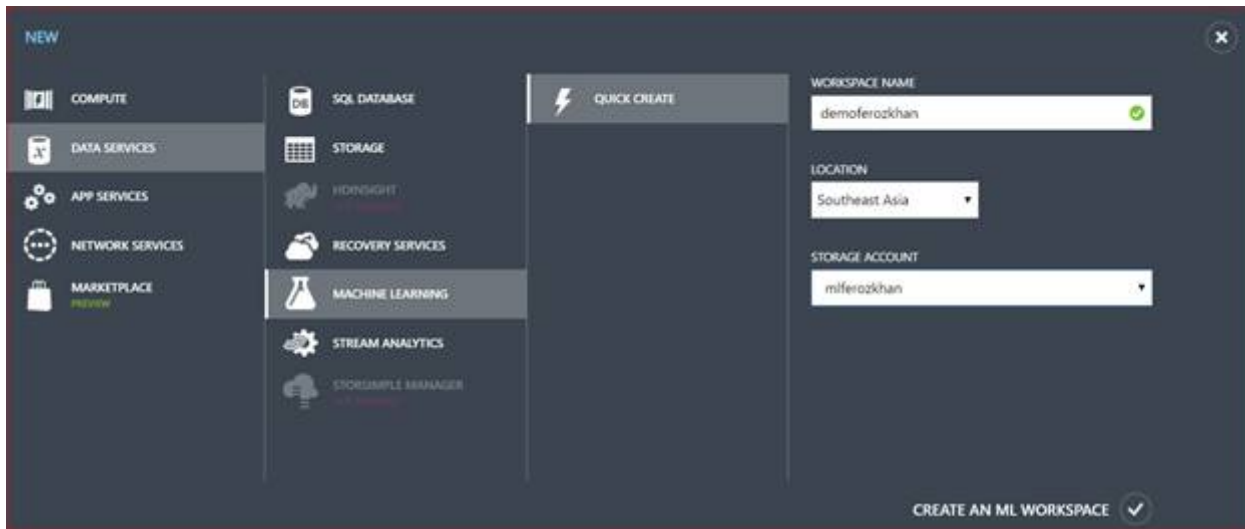
As mentioned in the previous articles, we can get access for Machine Learning with the free trial or an event with the help of Microsoft Azure Account.

- [Go here](#) for free trial.
- [Go here](#) for Azure Machine Learning.

Here, we will be going with Machine Learning from Microsoft Azure.

Login for Azure portal – www.manage.windowsazure.com

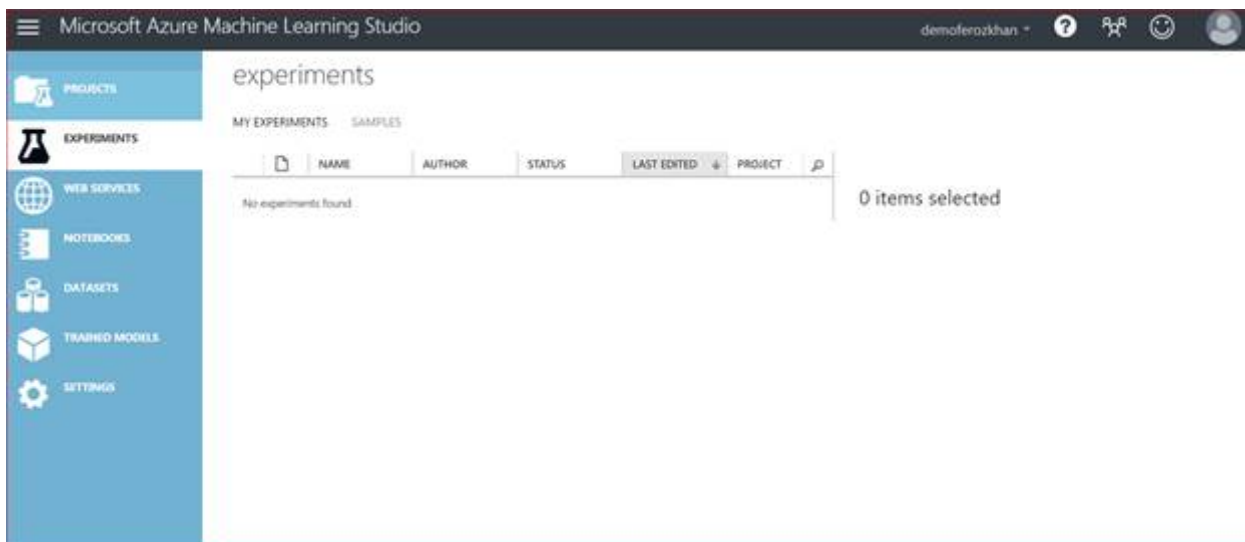
Click New - Data Services - Machine Learning - Quick Create - Workspace name, Location and Storage Account. Click Create an ML Workspace.



Click Sign in to ML Studio, once it is created.



Here, we can find the menu options in Azure Machine Learning Studio, as shown below.



- Projects
- Experiments
- Web Services
- Notebooks
- Datasets
- Trained Models
- Settings

Projects

Projects helps us with listing whatever projects are available. This organizes the collection of experiments, Web Services, data sets and the trained models. The same experiments can also be used for other projects.

Experiments

Experiments lists out the graphical view for the steps such as getting the data, pre-processing it, transforming it, training the model etc.

My experiments will list out the experiments, which we have on our Machine Learning Studio.



Samples in Machine Learning Studio will help us to go with sample Machine Learning experiments.

NAME	AUTHOR	STATUS	LAST EDITED	START TIME	END TIME
Sample 1: Download dataset fr...	Microsoft Corporation	Finished	6/16/2016 12:21:43 AM	6/16/2016 12:21:45 AM	6/16/2016 12:22:33 AM
Sample 2: Dataset Processing ...	Microsoft Corporation	Finished	6/16/2016 12:21:51 AM	6/16/2016 12:21:57 AM	6/16/2016 12:22:52 AM
Sample 3: Cross Validation for ...	Microsoft Corporation	Finished	6/16/2016 12:22:00 AM	6/16/2016 12:22:04 AM	6/16/2016 12:22:27 AM
Sample 4: Cross Validation for ...	Microsoft Corporation	Finished	6/16/2016 12:22:09 AM	6/16/2016 12:22:14 AM	6/16/2016 12:22:34 AM
Sample 5: Train, Test, Evaluate ...	Microsoft Corporation	Finished	6/16/2016 12:22:17 AM	6/16/2016 12:22:19 AM	6/16/2016 12:22:54 AM
Sample 6: Train, Test, Evaluate ...	Microsoft Corporation	Finished	6/16/2016 12:22:27 AM	6/16/2016 12:22:29 AM	6/16/2016 12:23:12 AM
Sample 7: Train, Test, Evaluate ...	Microsoft Corporation	Finished	6/16/2016 12:22:34 AM	6/16/2016 12:22:36 AM	6/16/2016 12:23:30 AM
Sample 8: Apply SQL transfor...	Microsoft Corporation	Finished	6/16/2016 12:22:48 AM	6/16/2016 12:22:52 AM	6/16/2016 12:23:02 AM
Sample 9: Split, partition and s...	Microsoft Corporation	Finished	6/16/2016 12:22:56 AM	6/16/2016 12:22:57 AM	6/16/2016 12:23:05 AM
Anomaly Detection: Credit Risk	Microsoft Corporation	Finished	6/16/2016 12:26:38 AM	6/16/2016 12:26:44 AM	6/16/2016 12:28:10 AM
Binary Classification: Credit risk...	Microsoft Corporation	Finished	6/16/2016 12:28:11 AM	6/16/2016 12:28:15 AM	6/16/2016 12:30:30 AM
Binary Classification: Breast ca...	Microsoft Corporation	Finished	6/16/2016 12:28:31 AM	6/16/2016 12:28:39 AM	6/16/2016 12:39:45 AM

Web Services

Web Services are used to go with our experiments to support the clients with the application on how we help them to work with the predictive models.



Datasets

Datasets are created with the help of files, which we have uploaded from our machine or with the help of Azure Data stores. This also holds the samples.

Trained Models

This is created as we run our experiments. We can store them and we can even use the same on other projects as we need.

trained models						
NAME	SUBMITTED BY	DESCRIPTION	DATA TYPE	CREATED		PROJECT
No trained models found						

Creating an Experiment in Azure Machine Learning Studio

We can create the experiments with the help of a blank or a predefined experiment, which is available. A blank experiment will help us to define the specific experiment, a pre-defined experiment is a best way to go, if we are working on a solution for a specific problem.

About the Experiment

We are going to create a Machine Learning project here, which can predict the price of a car with the help of its features.

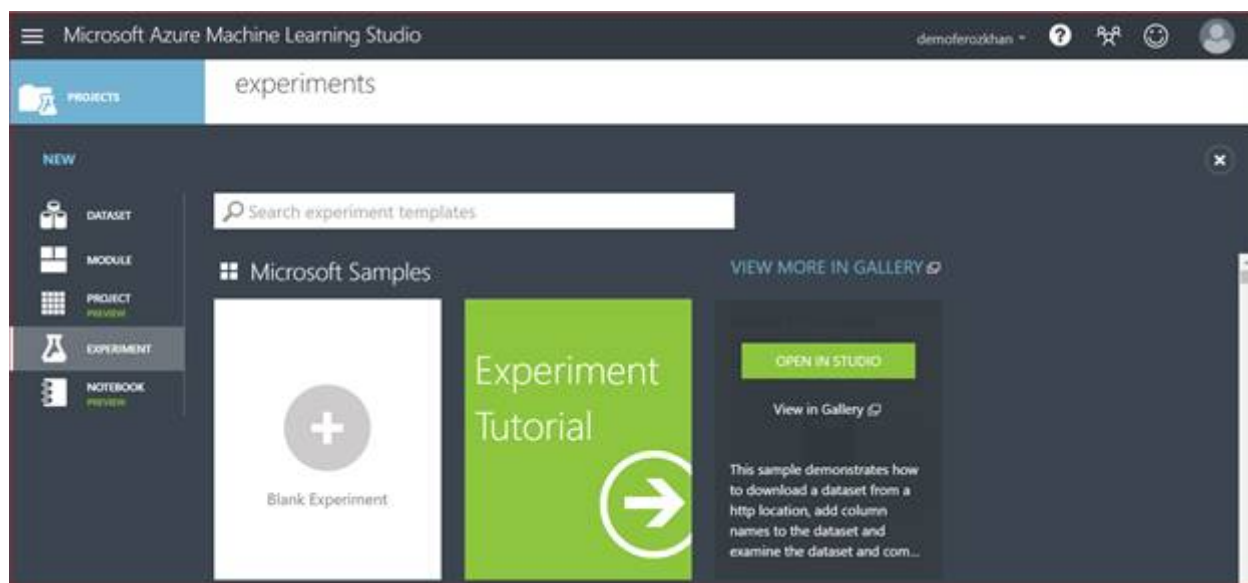
Major steps on Experiments of Microsoft Azure Machine Learning Studio

- Getting data.
- Selecting the algorithm.
- Training the model.
- Evaluating the model.

Keep the steps given above in mind, whenever you work with Machine Learning.

Let's work on a blank experiment here.

New - Experiment - Blank Experiment.

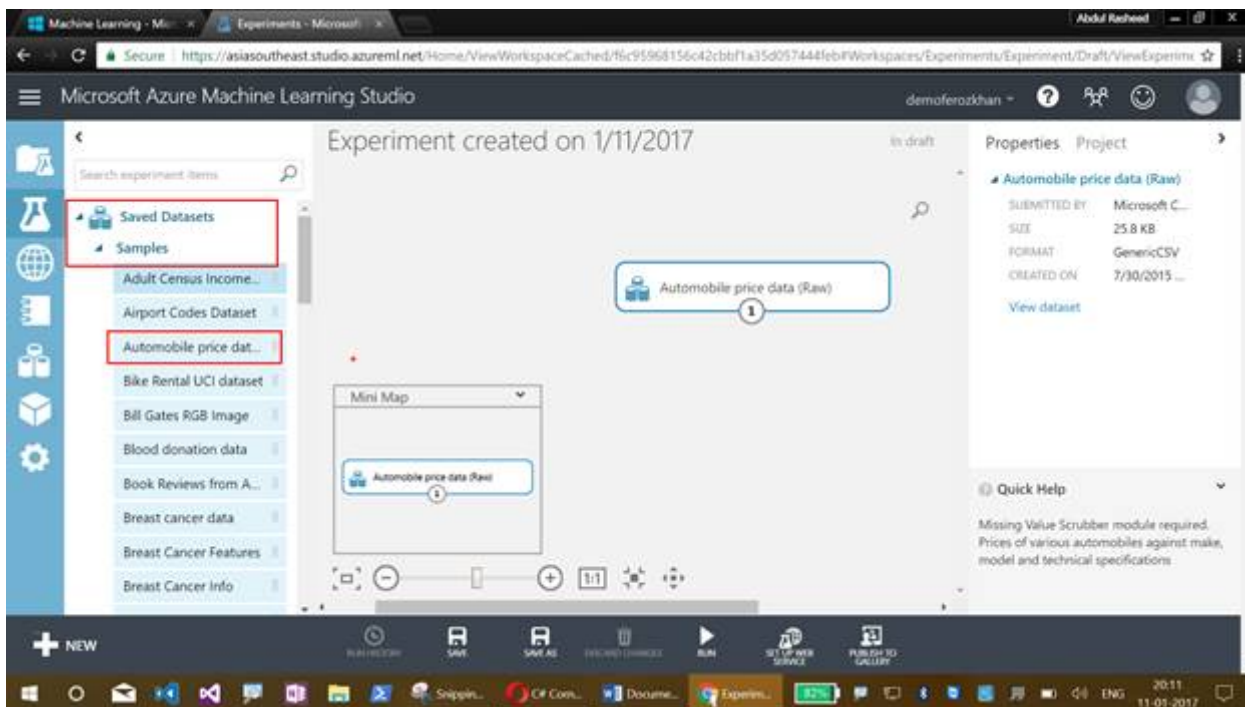


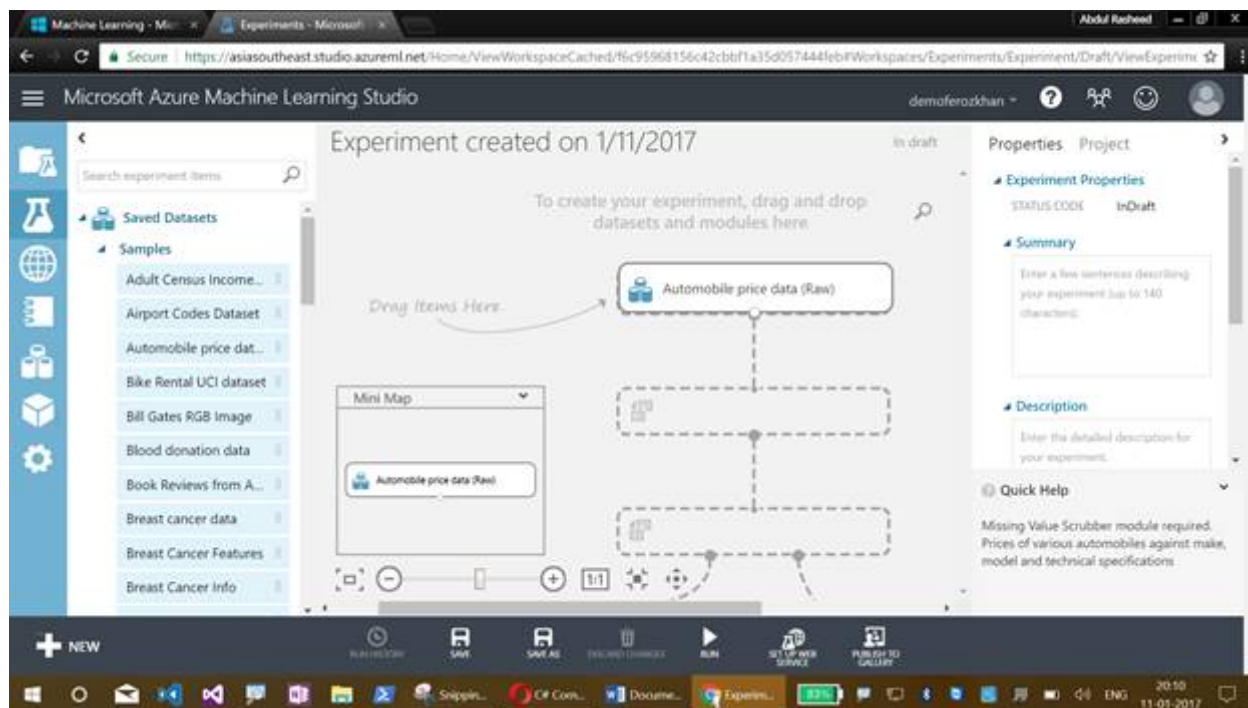
Now comes the graphical display of our Azure Machine Learning Studio experiment, which will help us to drag the items. This also illustrates the flow of an experiment.



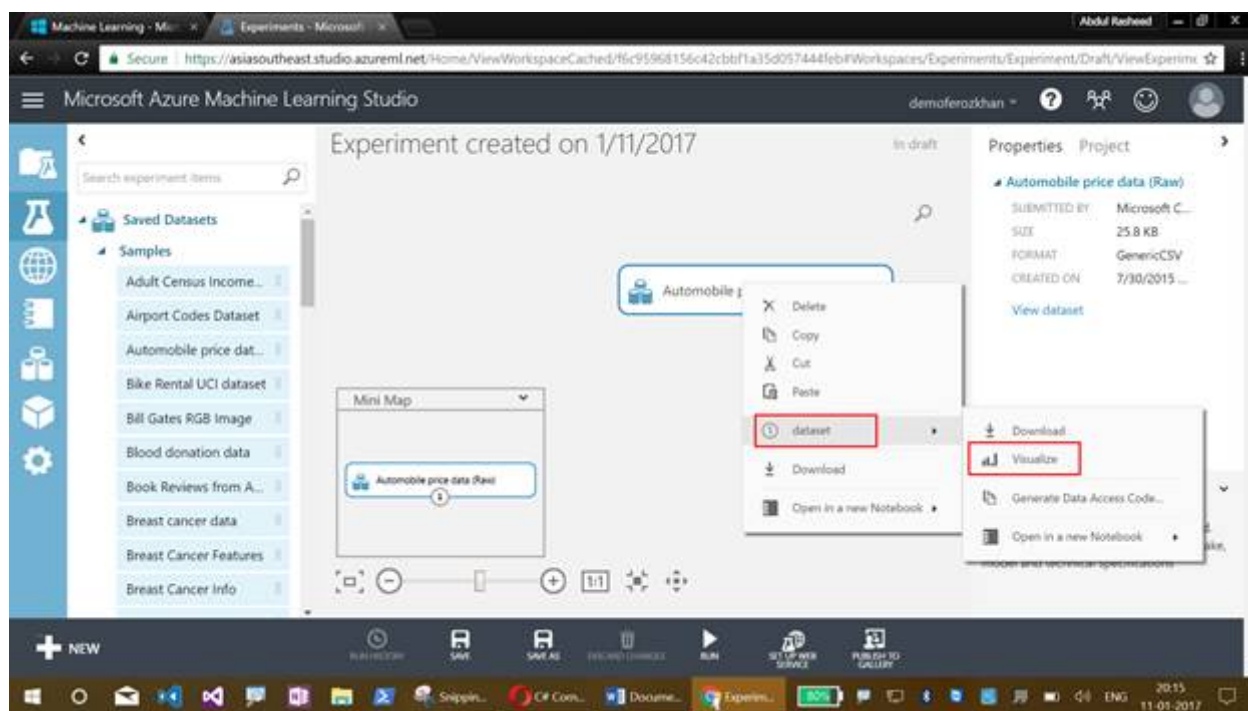
Now, the features of the cars holds the data like Manufacturer, number of doors, number of seats and other automobile features. This data is already available with Microsoft Azure; and I am going to just drag it off.

Click on Saved Datasets -> Samples -> Automobile price data (Raw) -> Drag it on for the dataset Workflow.

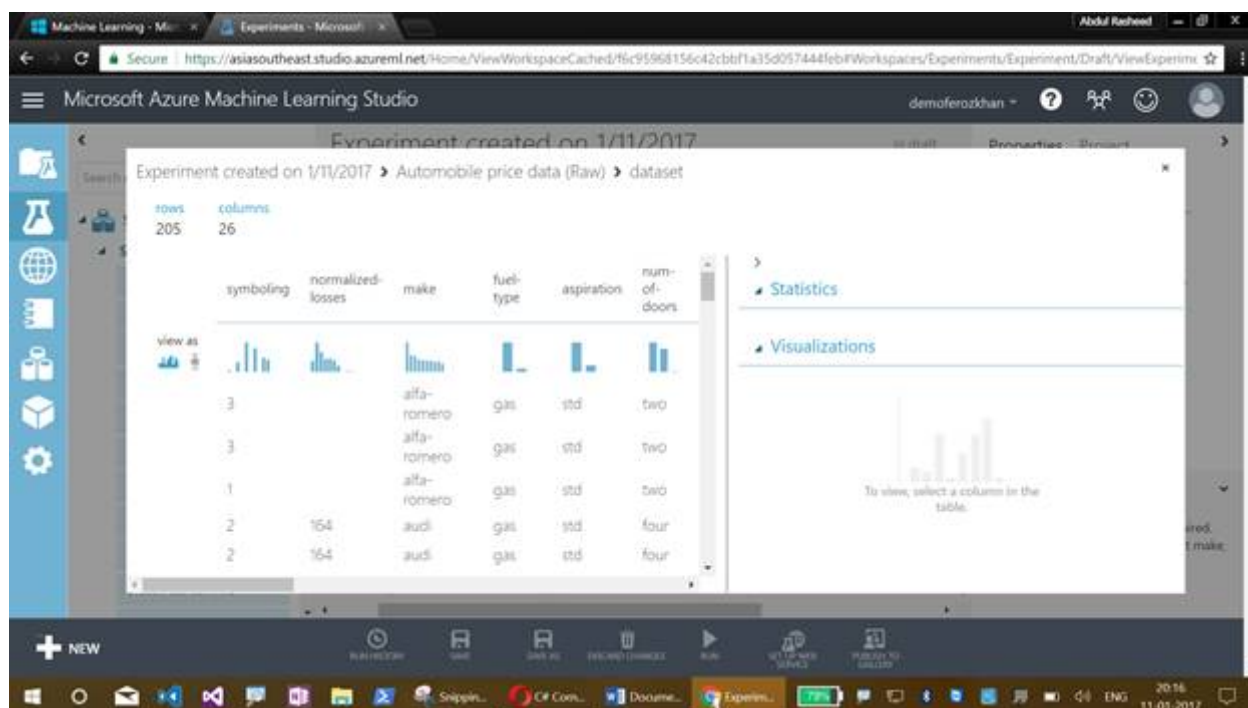




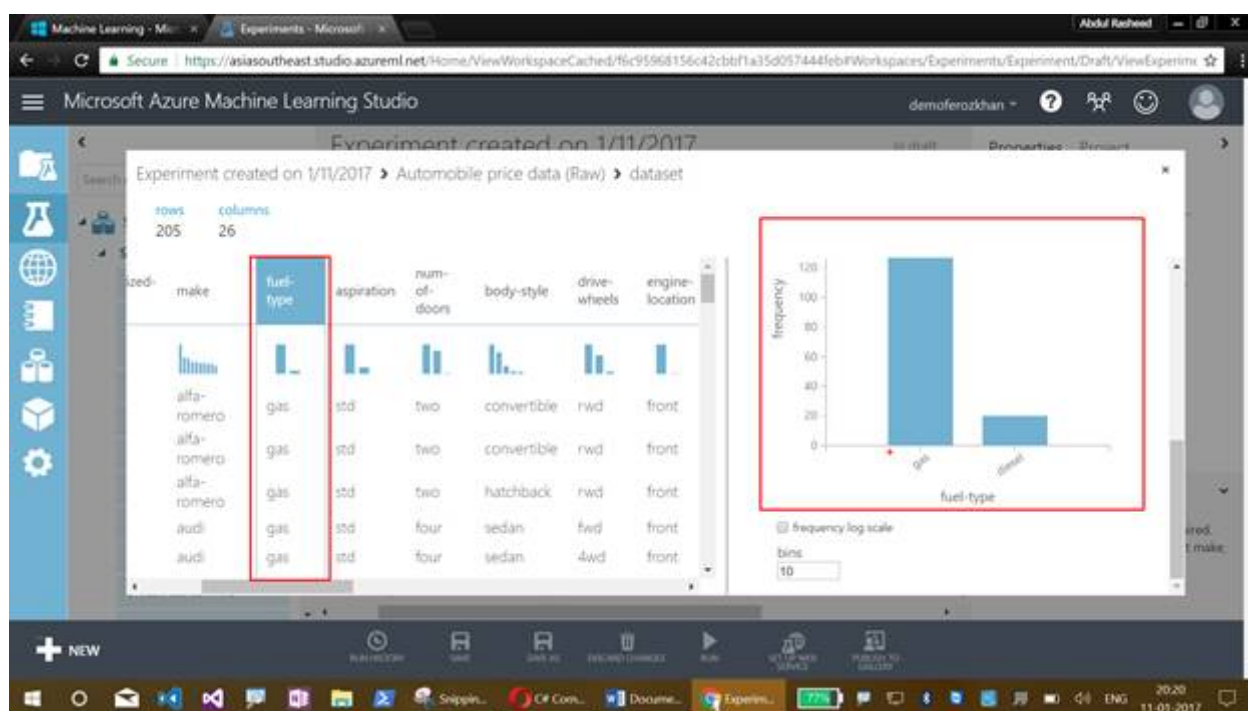
Right click on the Dataset-> Dataset-> Visualize.



Here, we can find more data available in the dataset with different features.

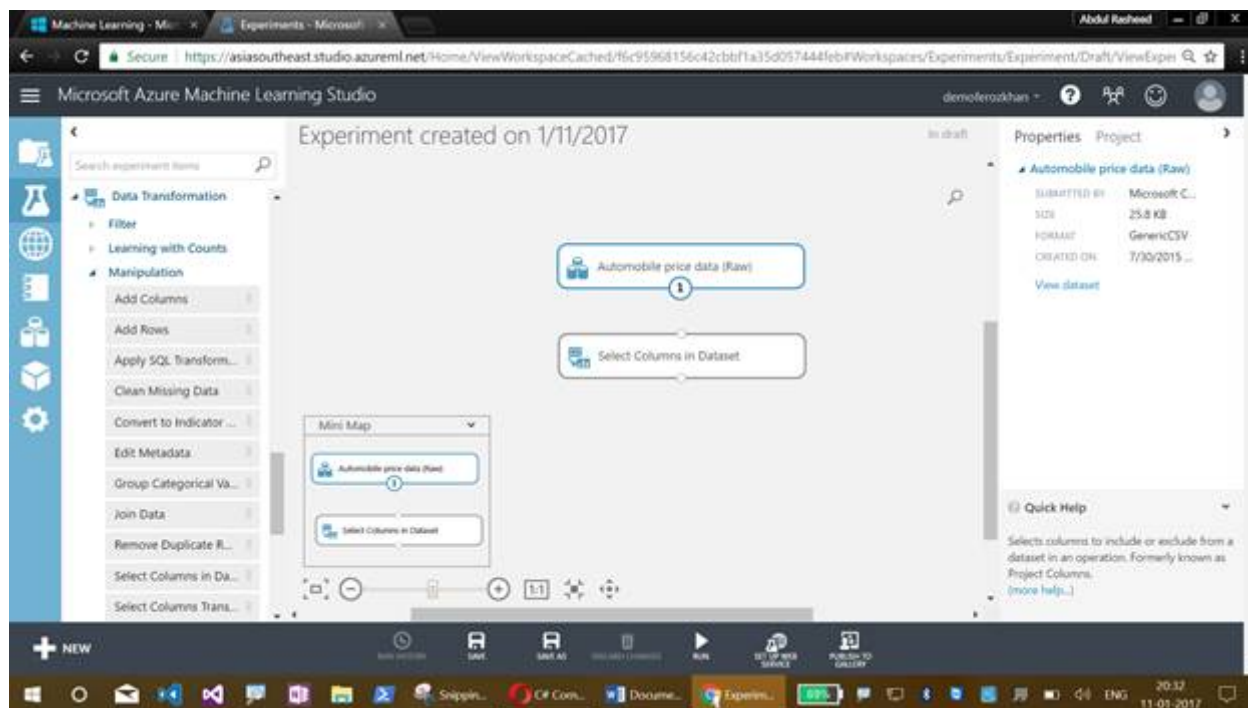
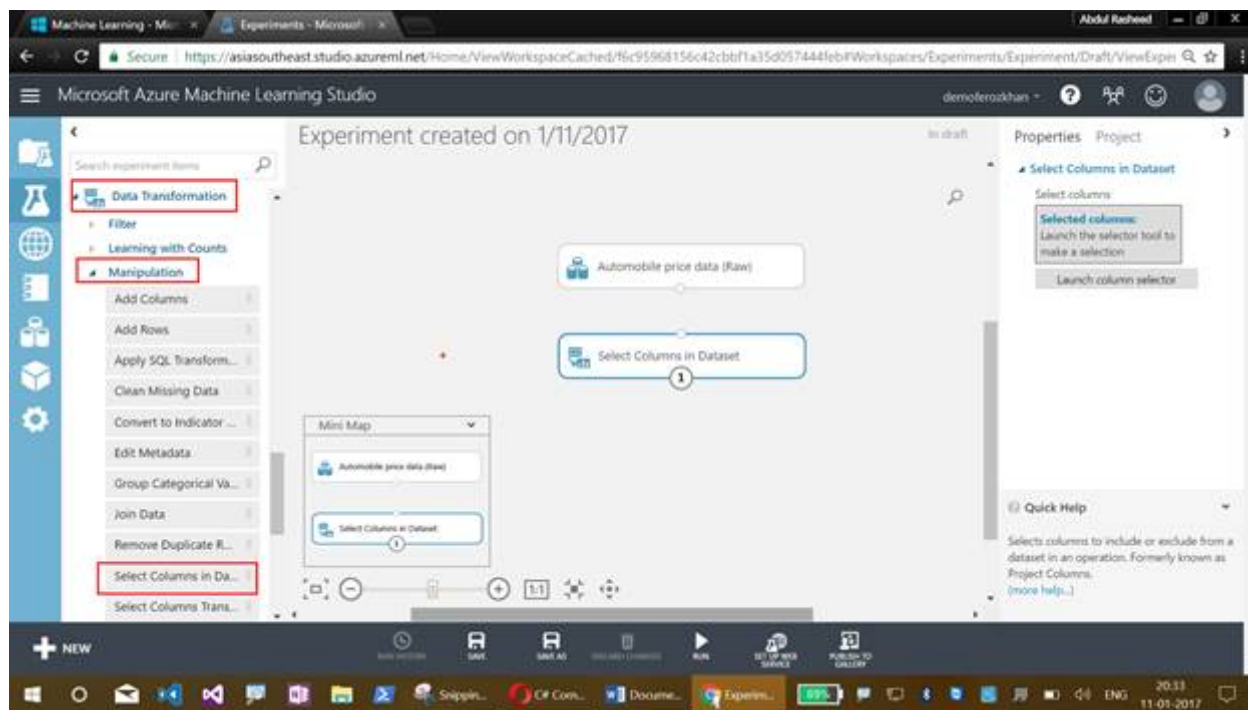


We can also get more details about a feature, when we click on it from the sample dataset.

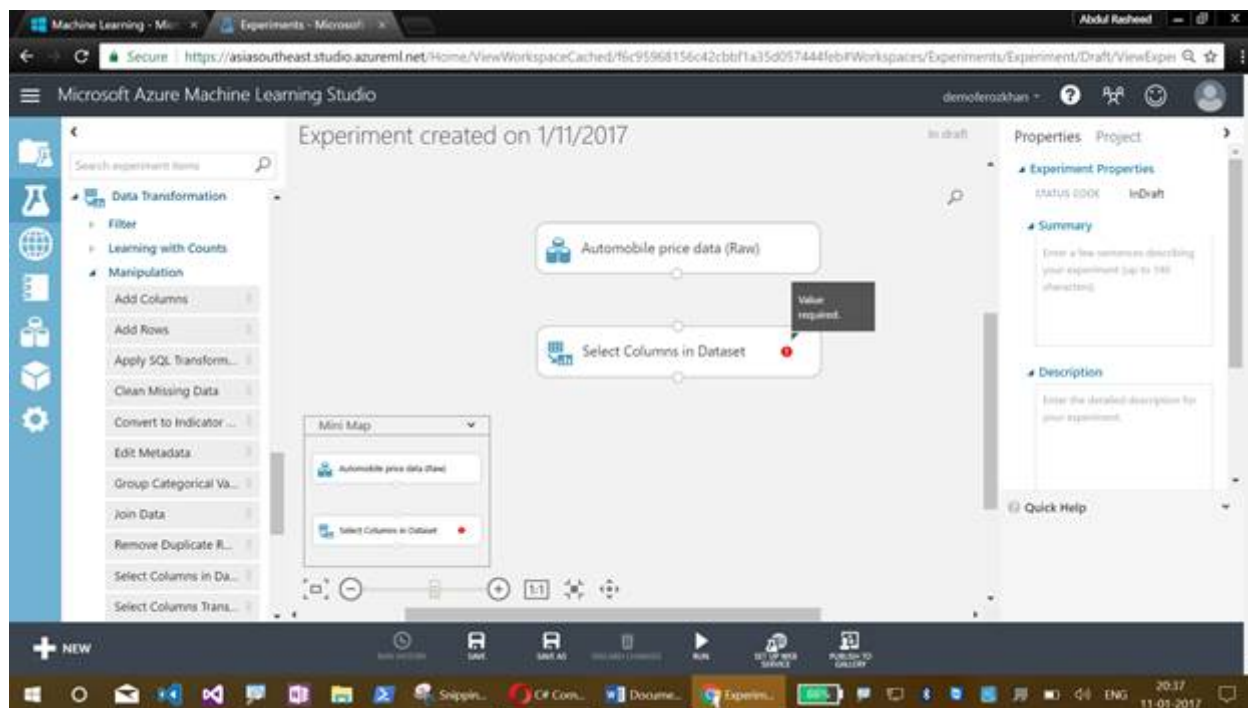


Let's work with a manipulation for the dataset now.

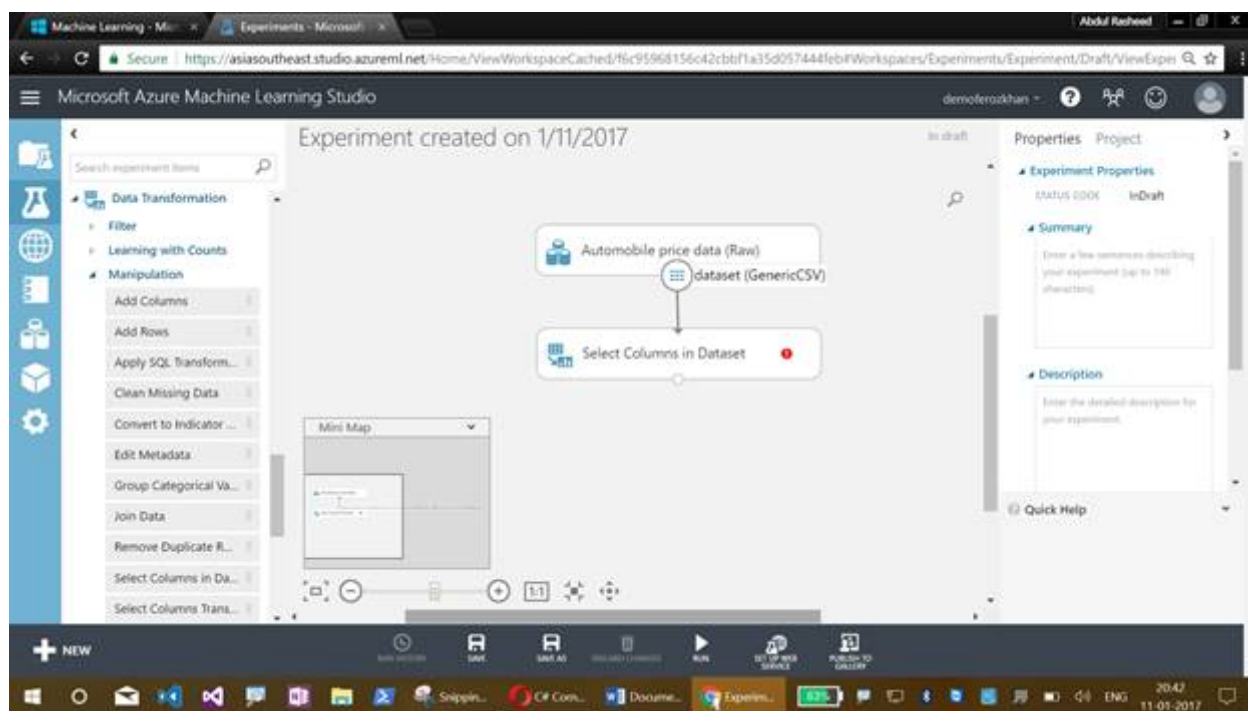
Click on Manipulation - Select Columns in Dataset - Drag it to the Workflow.

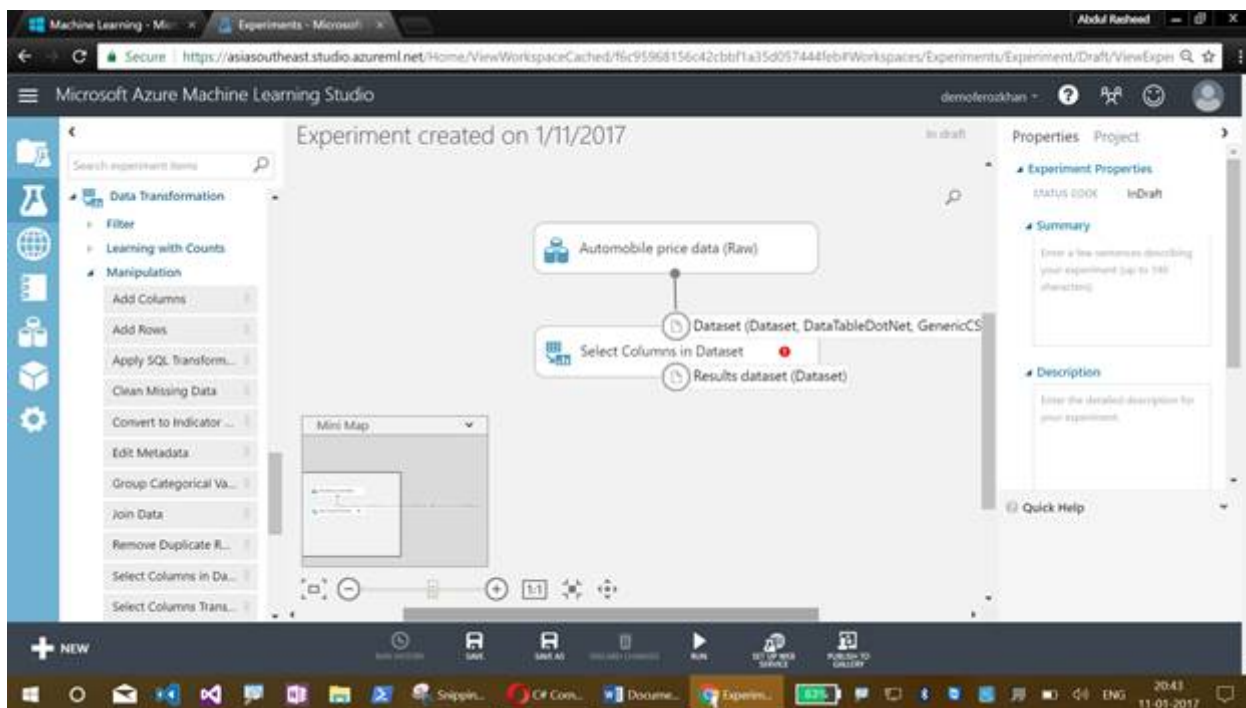


Now, we can find an exclamation symbol over here, as shown below.



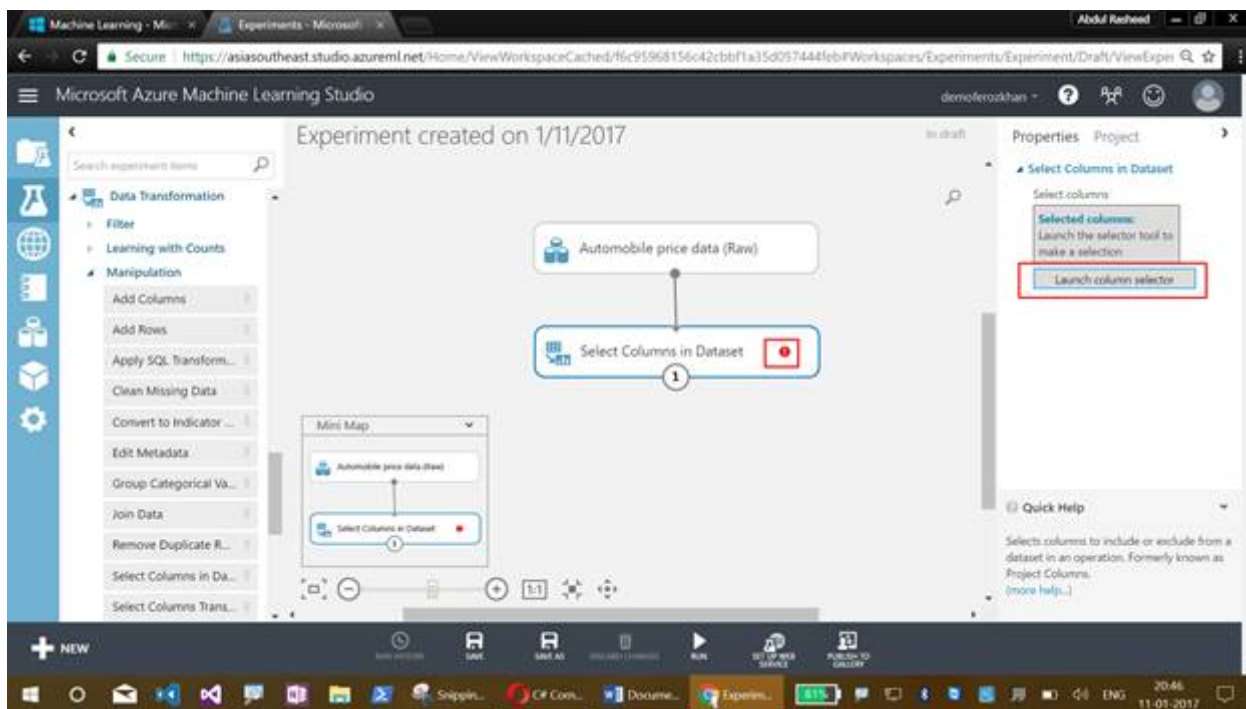
This exclamation symbol defines that it is not connected with any datasets. We can fix a solution for this by just drawing a line, which will flow the output of the price dataset as an input for the column dataset. We can also find the format for the dataset of each and this line will not connect, if the datasets format doesn't match.



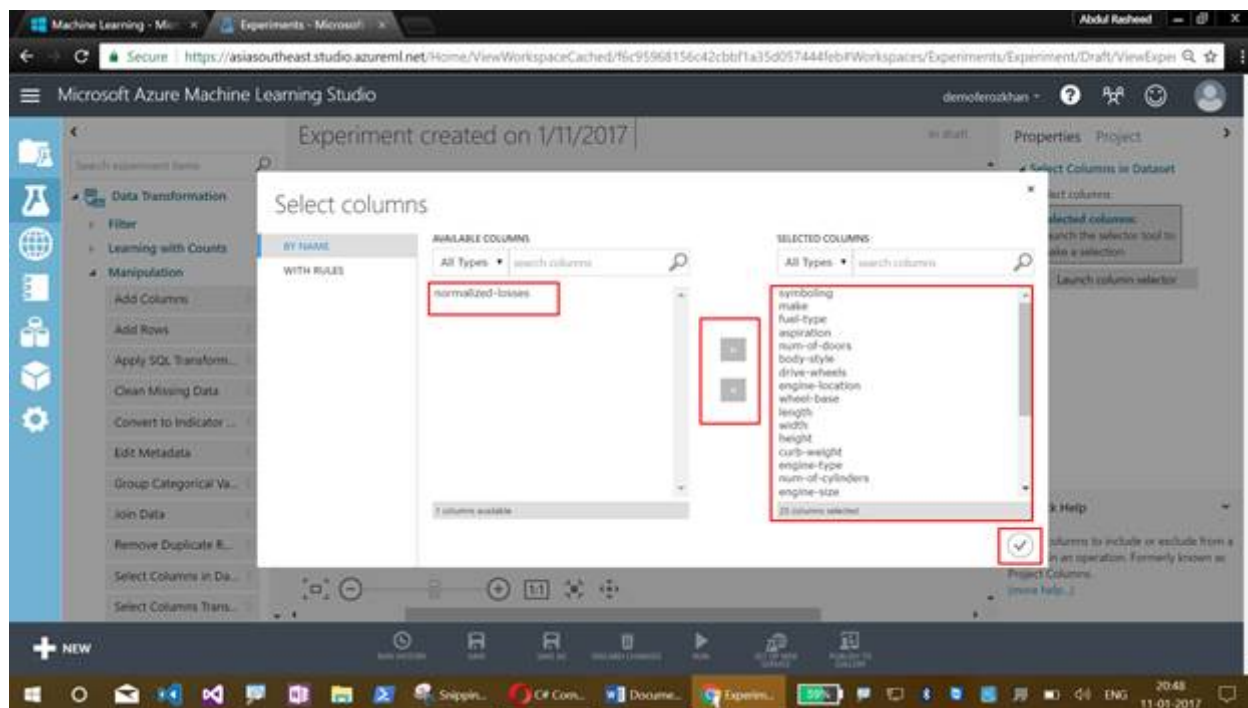


We still have an exclamation mark, which means that we have to specify a value for the column in the datasets.

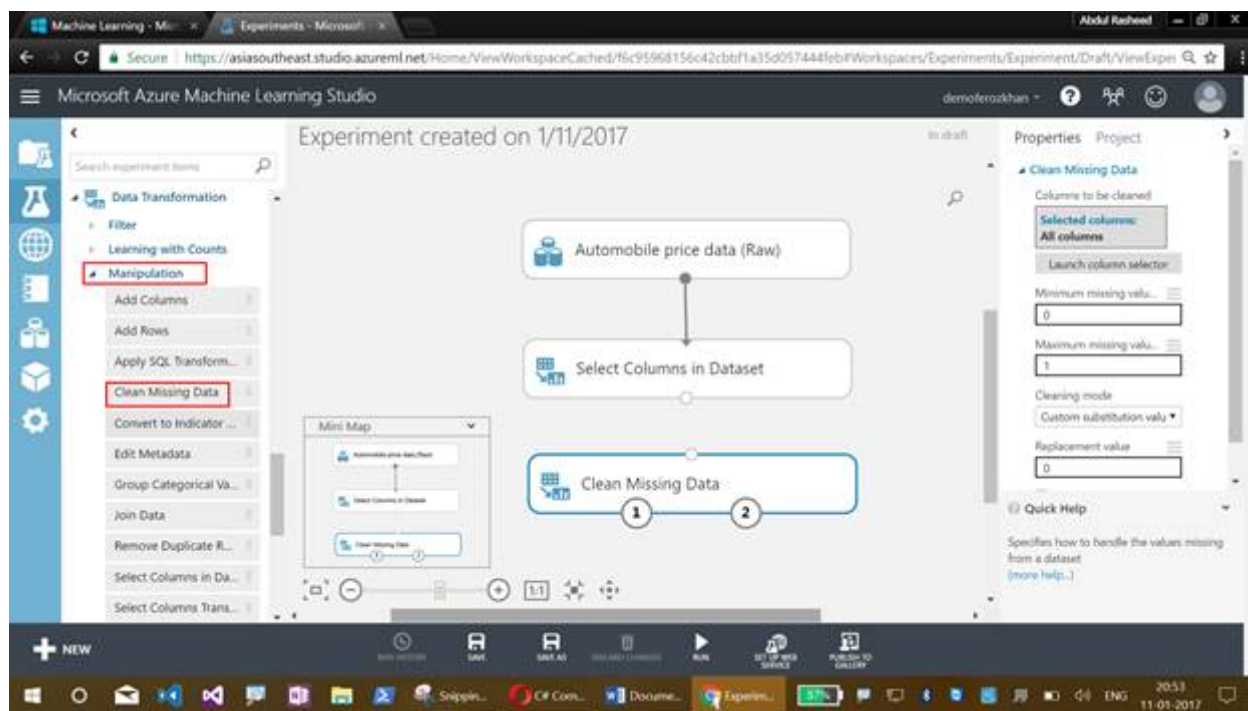
Click on the ! - Launch Column Selector.

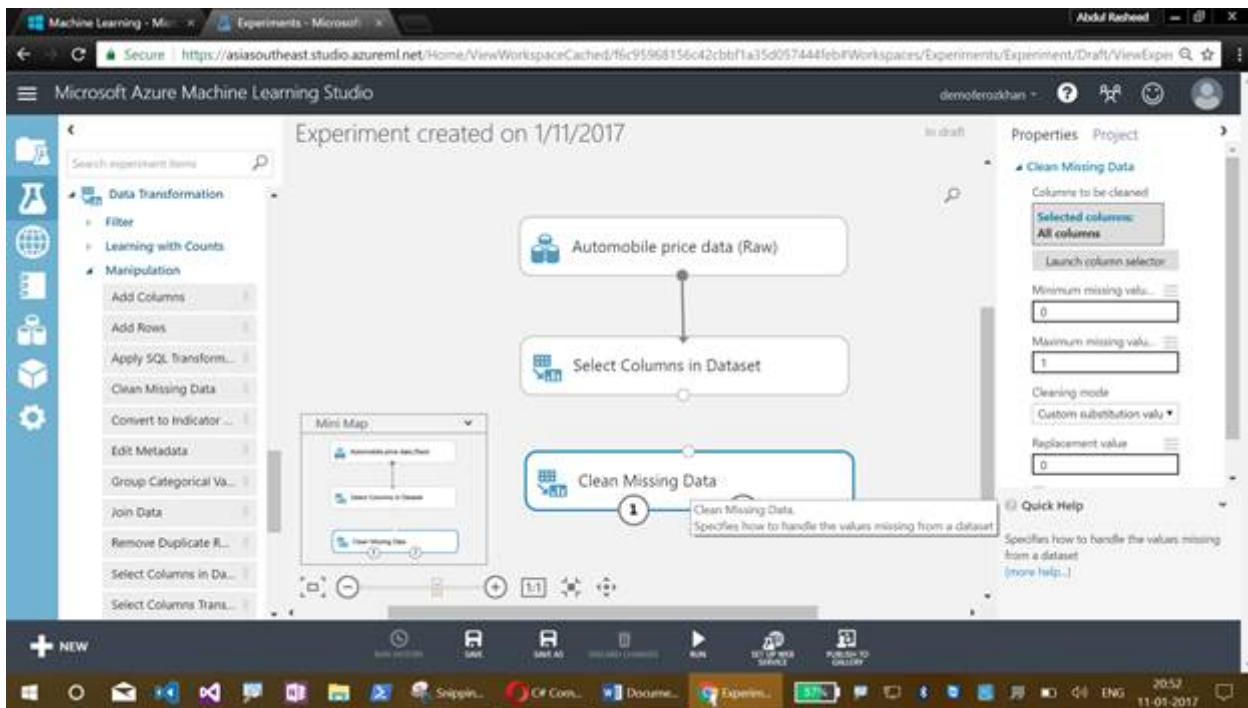


The list of columns which has to be selected is not mentioned in the Datasets and due to this reason, we get the exclamation mark here. To overcome this, click on the Launch Column Selector, except normalized losses move, all the other columns will load the columns for the next step. Now, the exclamation mark will not be available on the Workspace.



We do this with the Clean Missing Data, so let's drag it off to the Workspace and connect it to the datasets in the column.

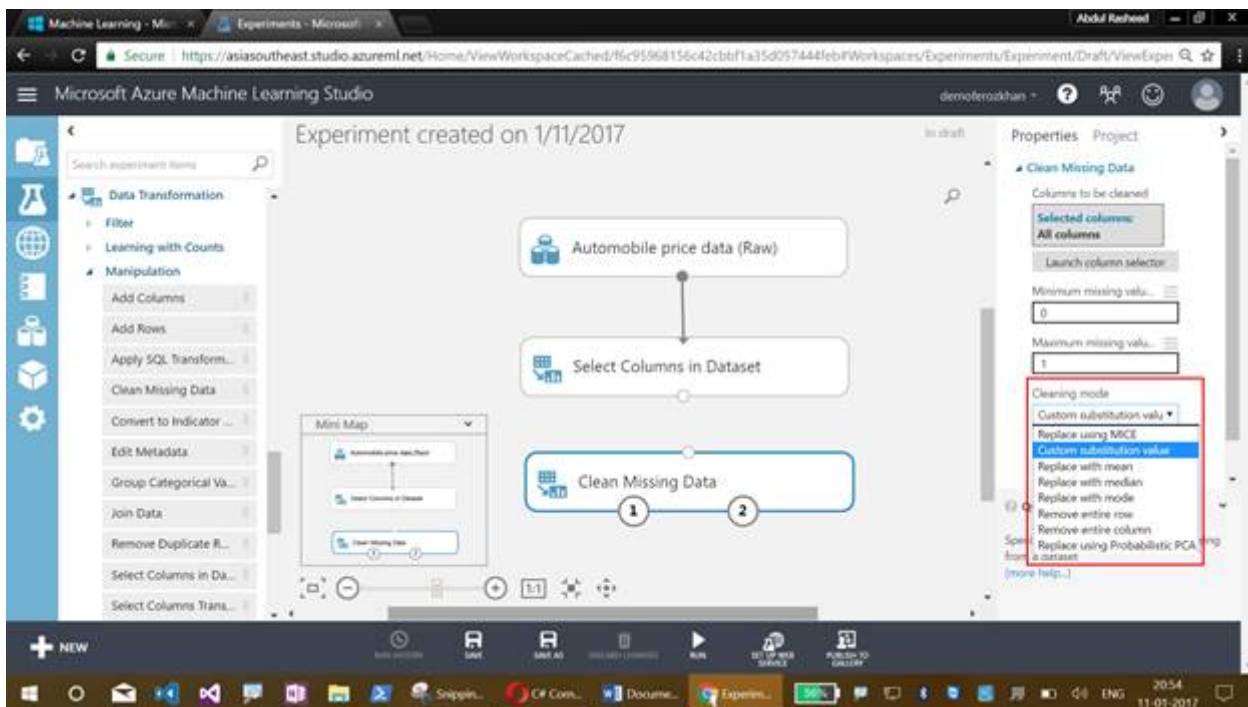


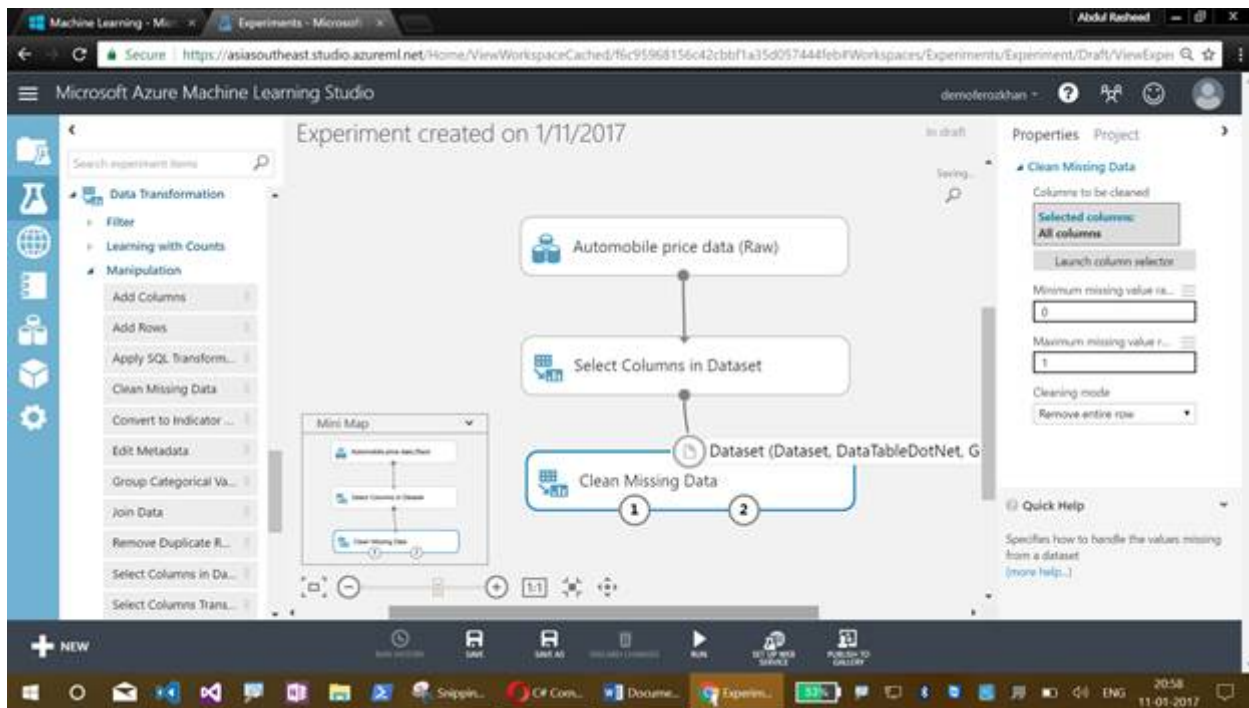
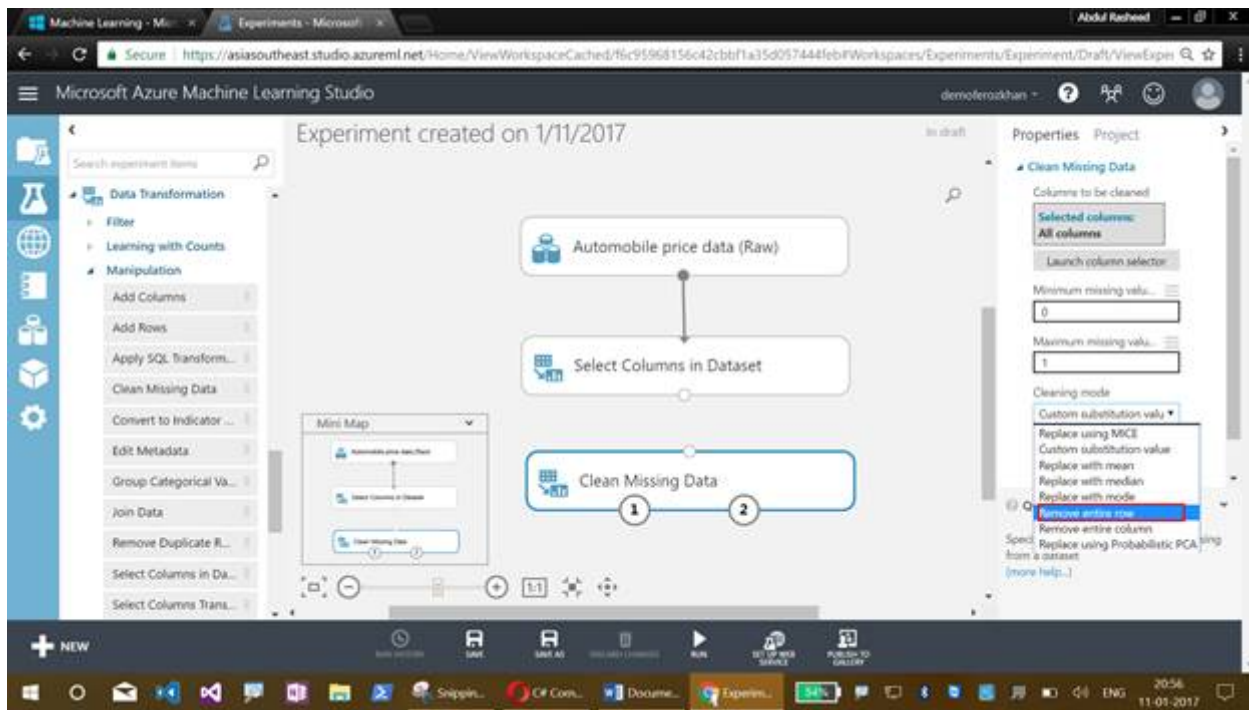


Specify the mode for the Cleaning module.

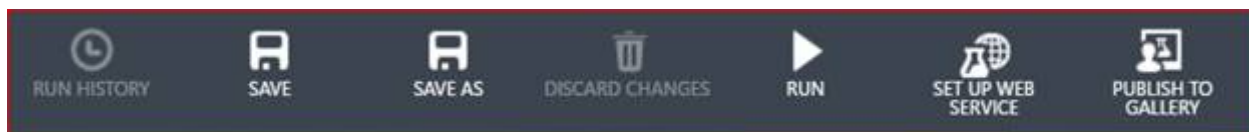
- *Replace using Probabilistic PCA*- This will help you to replace with the complex value.
- *Custom Substitution Value*- This helps us in inner replacement value.
- *Replace with mean/median/mode* helps us in replacing the specified statistic for the column.
- *Remove entire row/column* helps us in moving the missing values.

Let's select Remove entire row here.





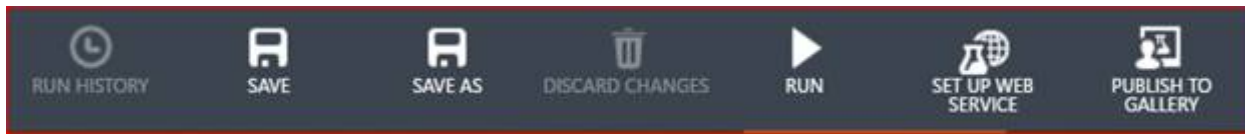
Now, click on the Save button at the bottom pane to save the work, which we have done so far.



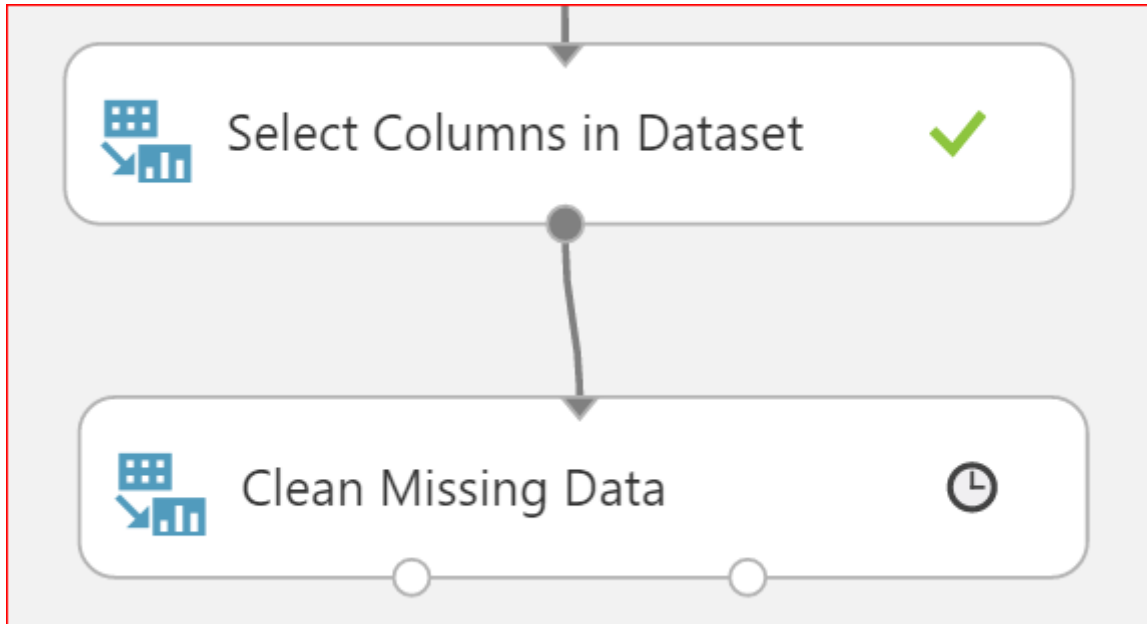
Once the saving process is complete, you can find a notification that the draft has been saved with the time, as shown below.

Draft saved at 8:59:02 PM

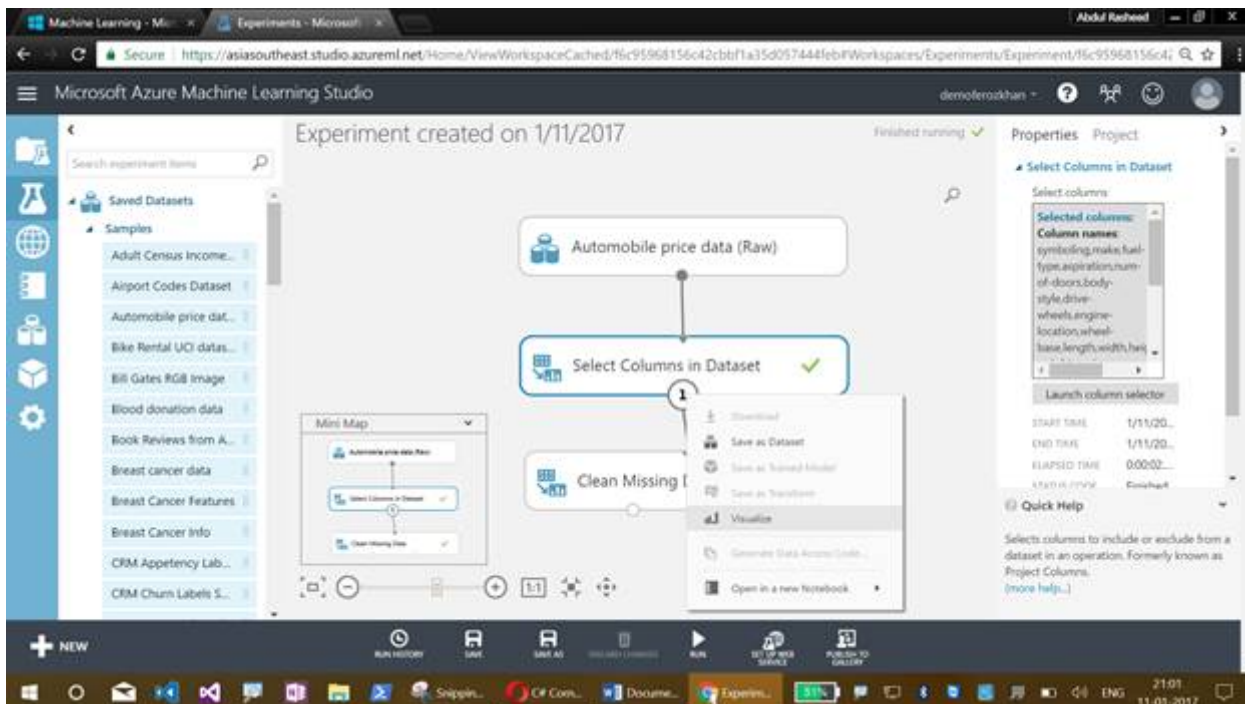
Again, click on Run at the bottom pane of your panel to run the experiment.



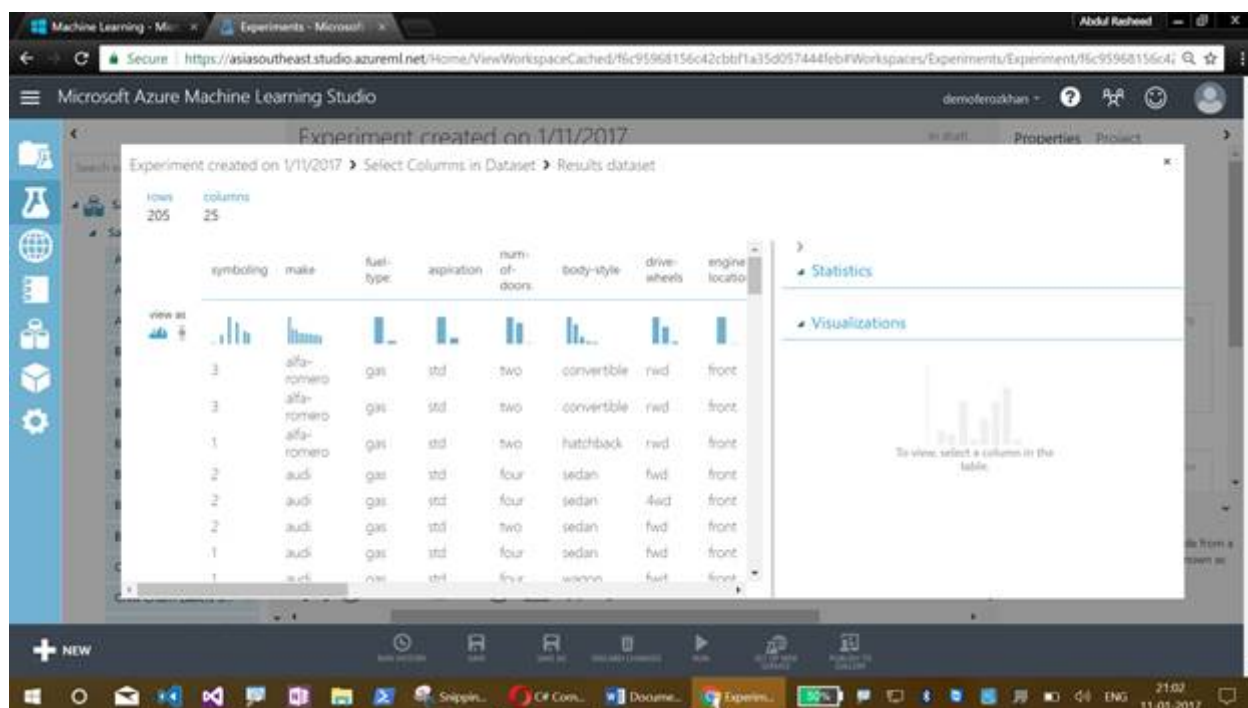
As each of the steps run, we can find the green color tick mark, which specifies the status of completion.



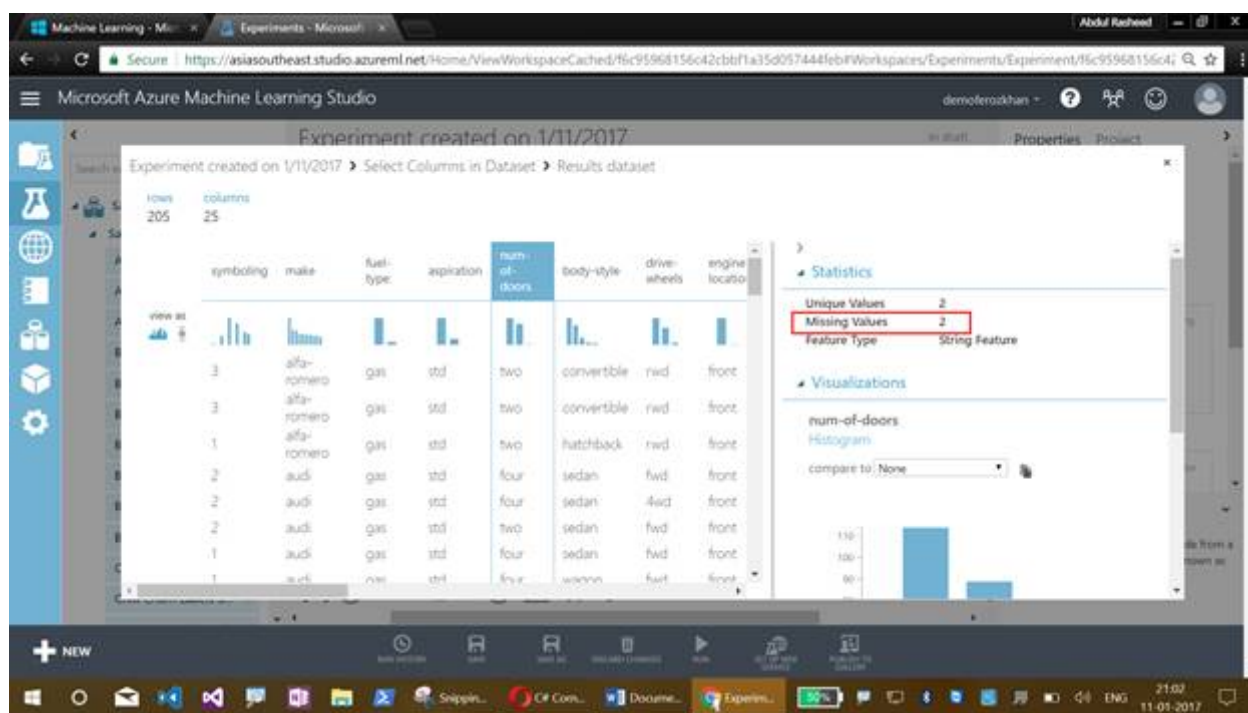
Now, check the data and see how it has been changed. Right click on the output from select column and datasets and click on Visualize.



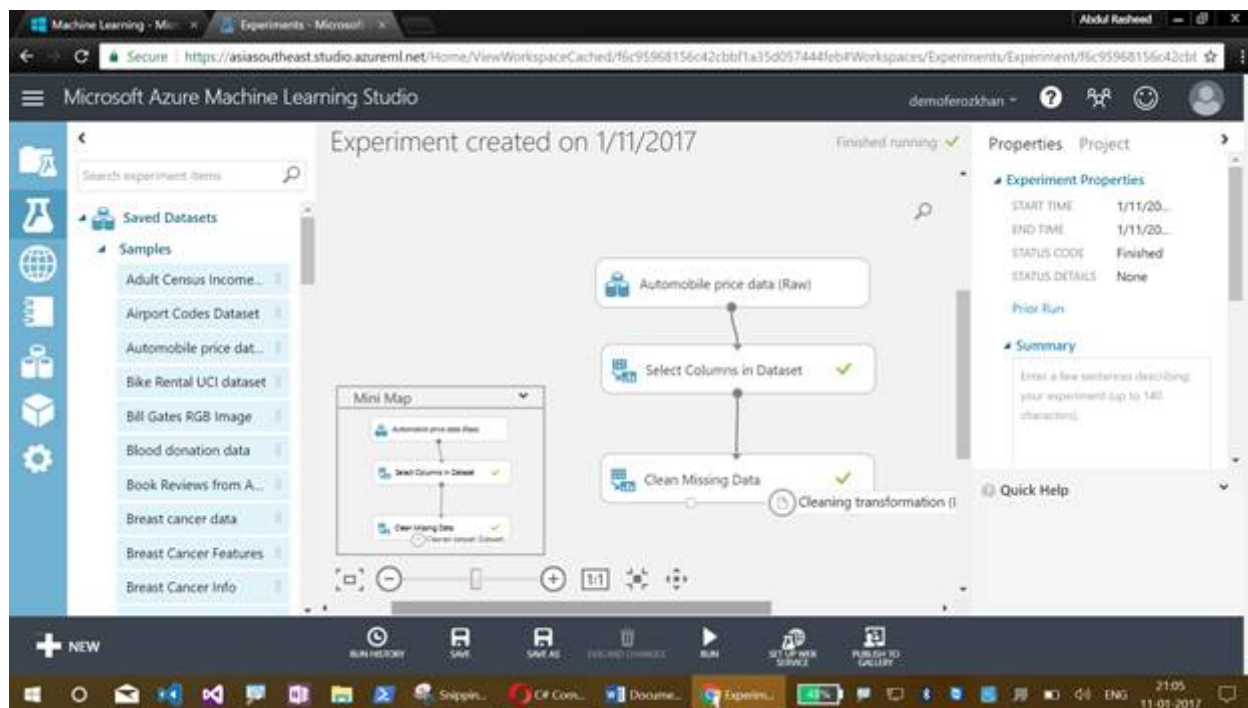
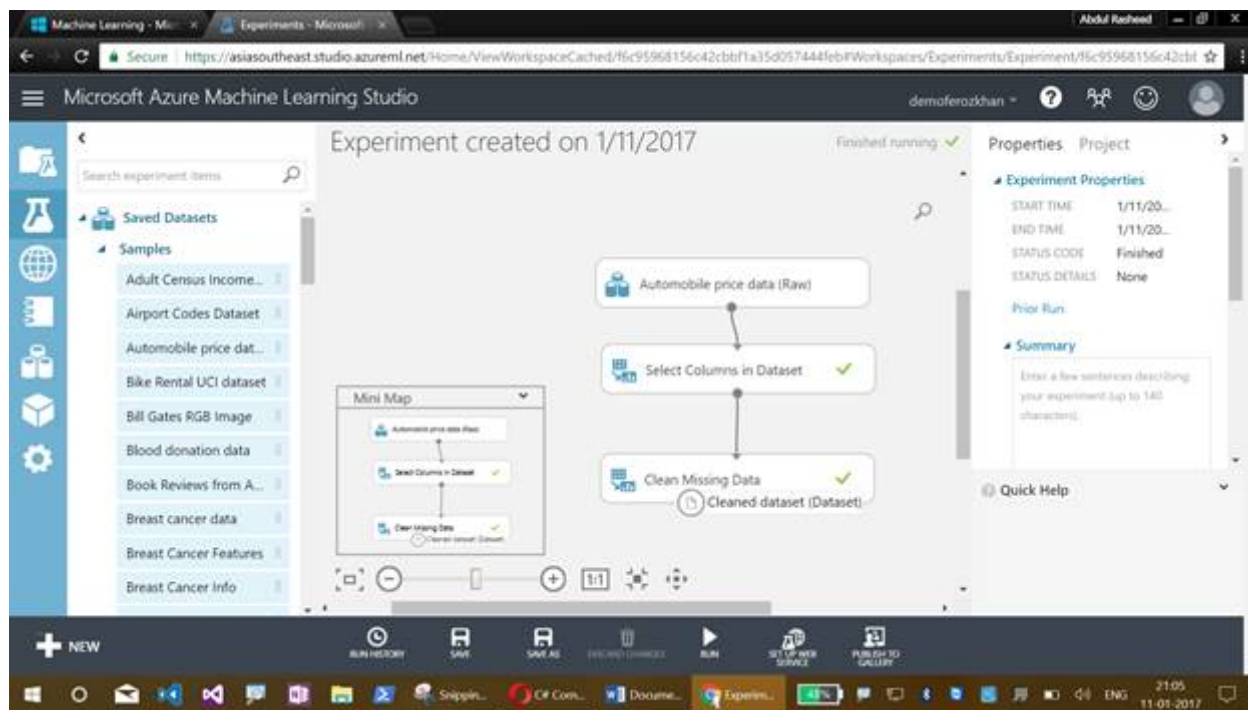
Here, you can find that the normal loss column, which will not be available.



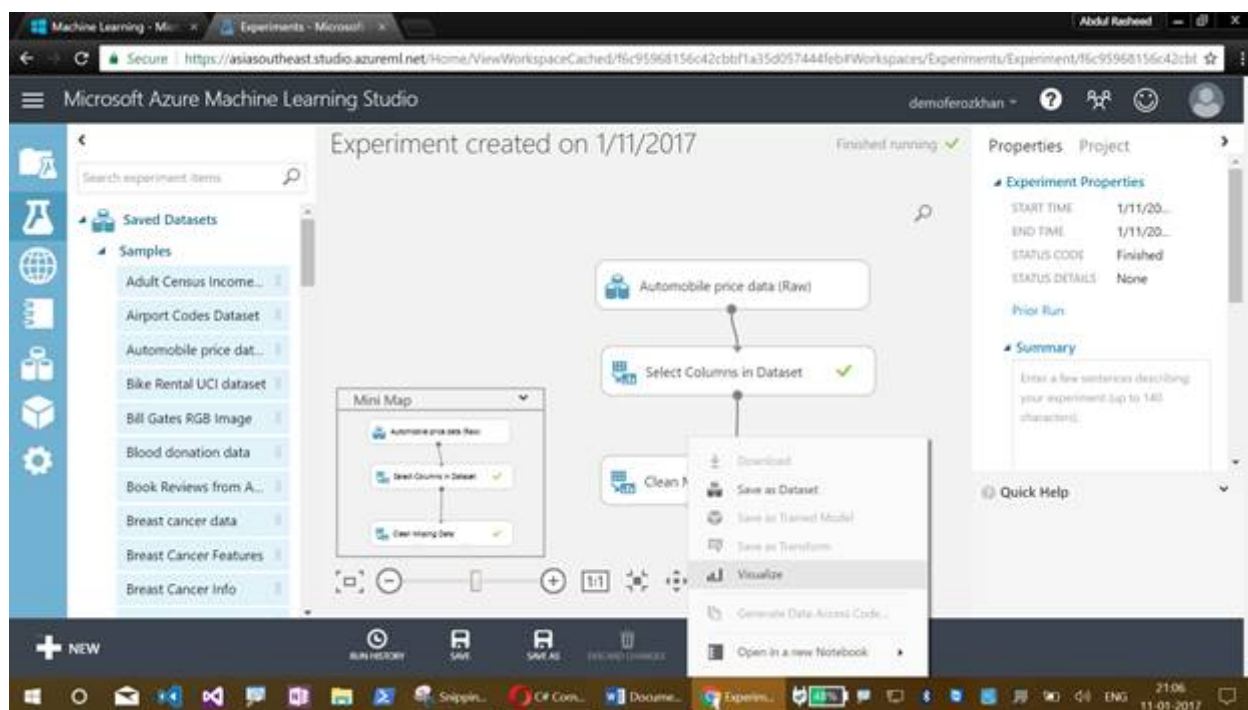
We still have missing values available for the number of doors as 2. To remove this, we will be using Cleaning Missing Data Model.



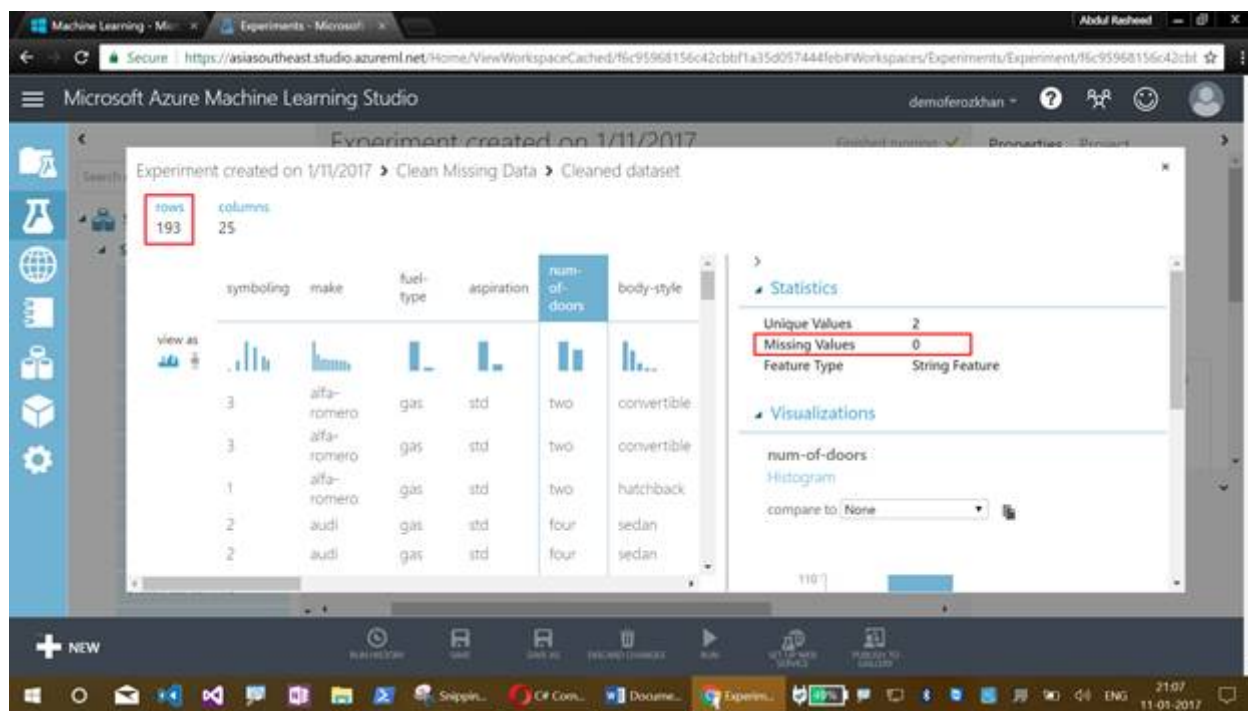
Clean missing data has two outputs- Cleaned Dataset and Cleaning Transformation.



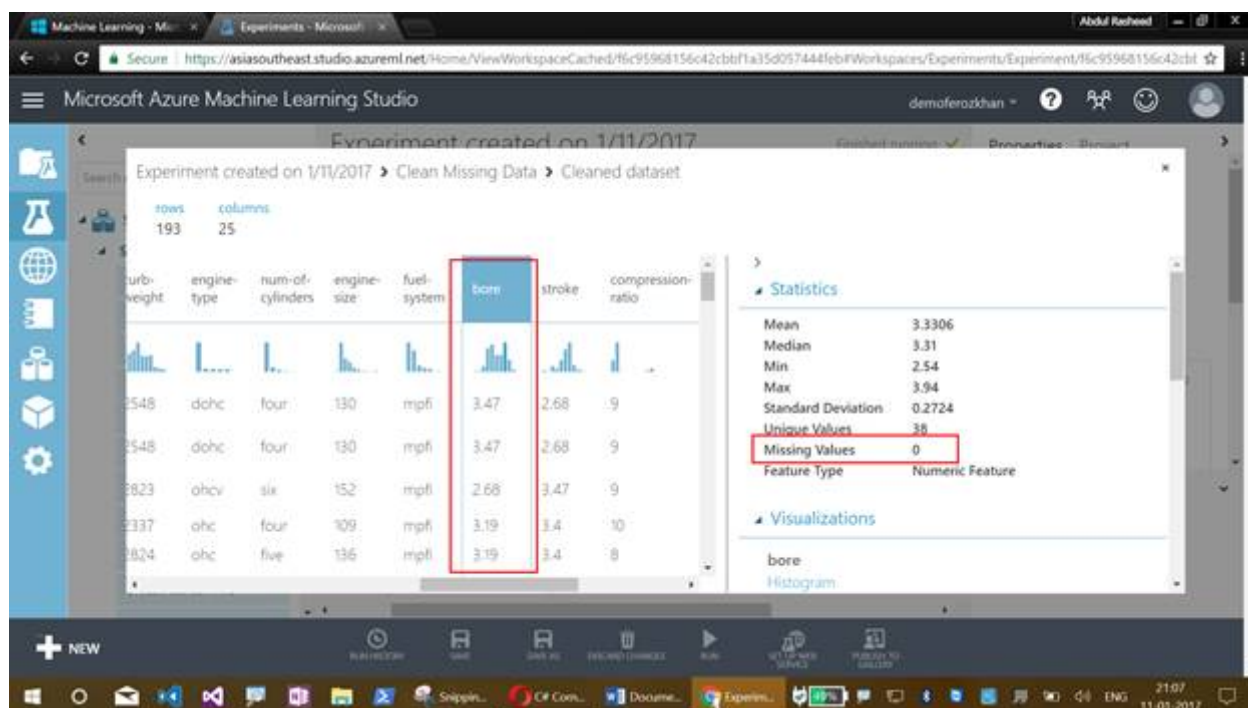
Right click on the Cleaned Dataset Model and go for Visualize.



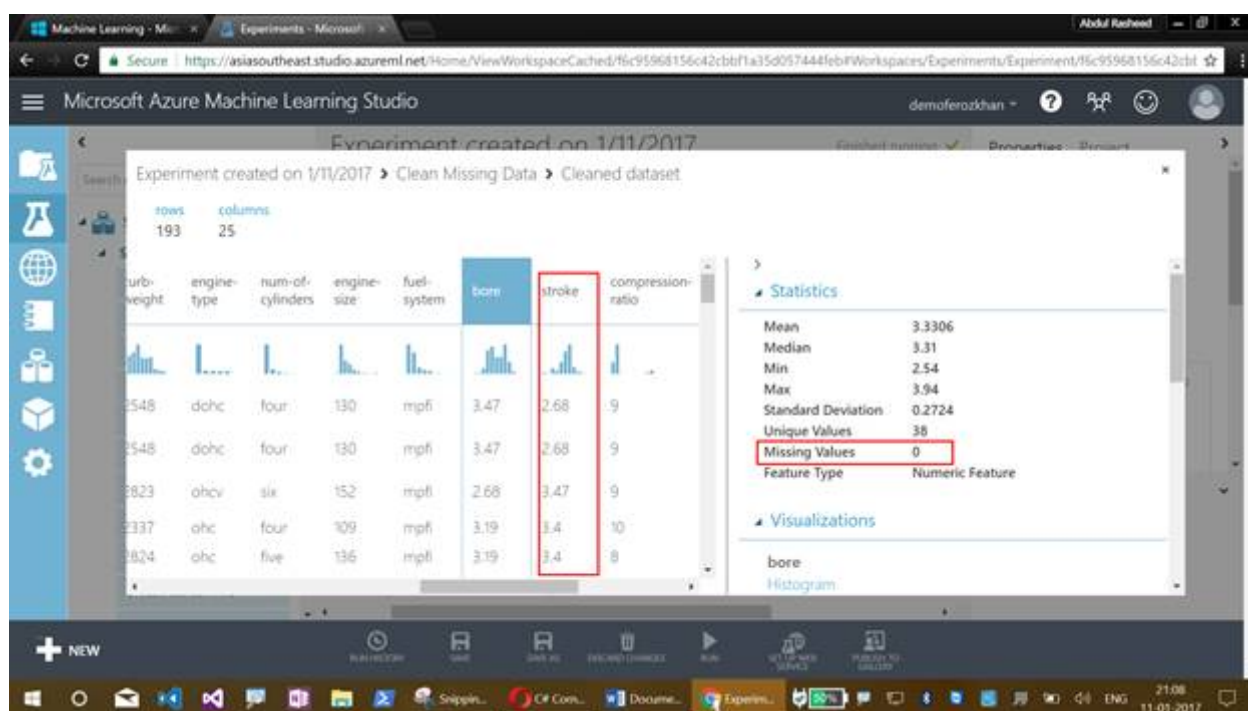
Search for the number of doors and you can find the missing value as 0. The rows will also be reduced for 193.



Kindly find the screenshot given below.



For Stroke, kindly find the screenshot given below.



Now, we can use this cleaned data to predict the price of the car, where we don't have any missing values.

Summary

I have discussed about the Machine Learning Studio, their menus, creating an experiment, working with sample datasets, refining with the missing values and getting clean data for training the model.

Follow my next article for Training the model with this data.

Thank you for using C# Corner