# Developing ML Model In Cloud Environment Using MS Azure

Here is the step-by-step process of building the prediction model.

Step 1:- Search for Azure Machine Learning Studio on Google and click on the first link. Login with the credentials **(Create Free account – Please don't enter any credit or debit Card details)** and we can see the studio.
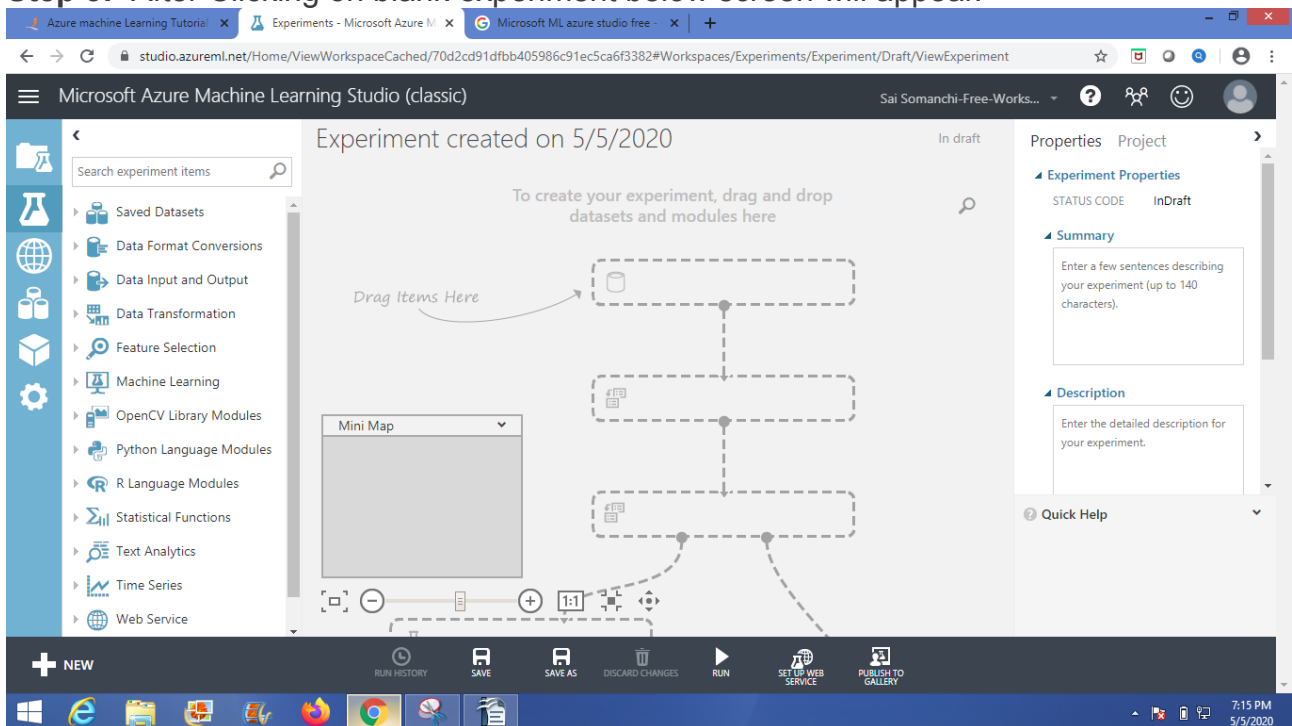To create a new experiment, click on **NEW(+NEW at bottom left side)** which is on the bar at the bottom of the studio.
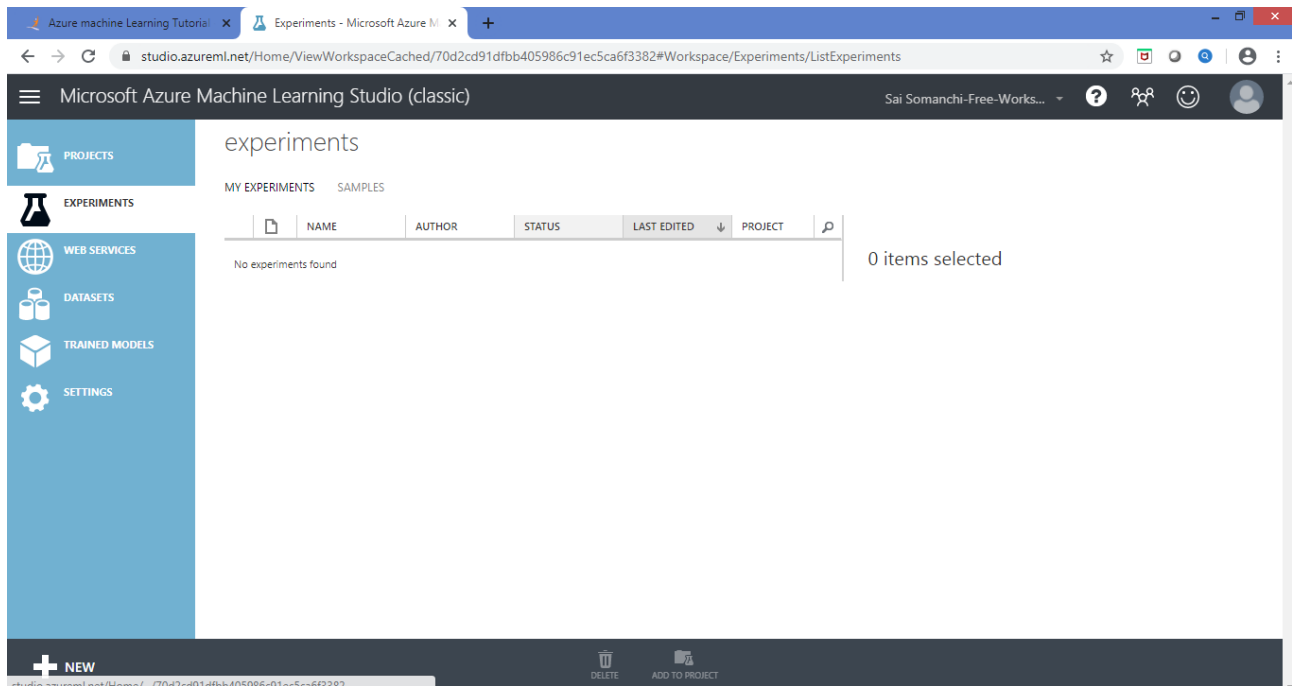
Step 2: When we click on **NEW**, here are the options that will pop up. Click on **Blank Experiment** and we will be redirected to our workspace, where we can start with our experiment.
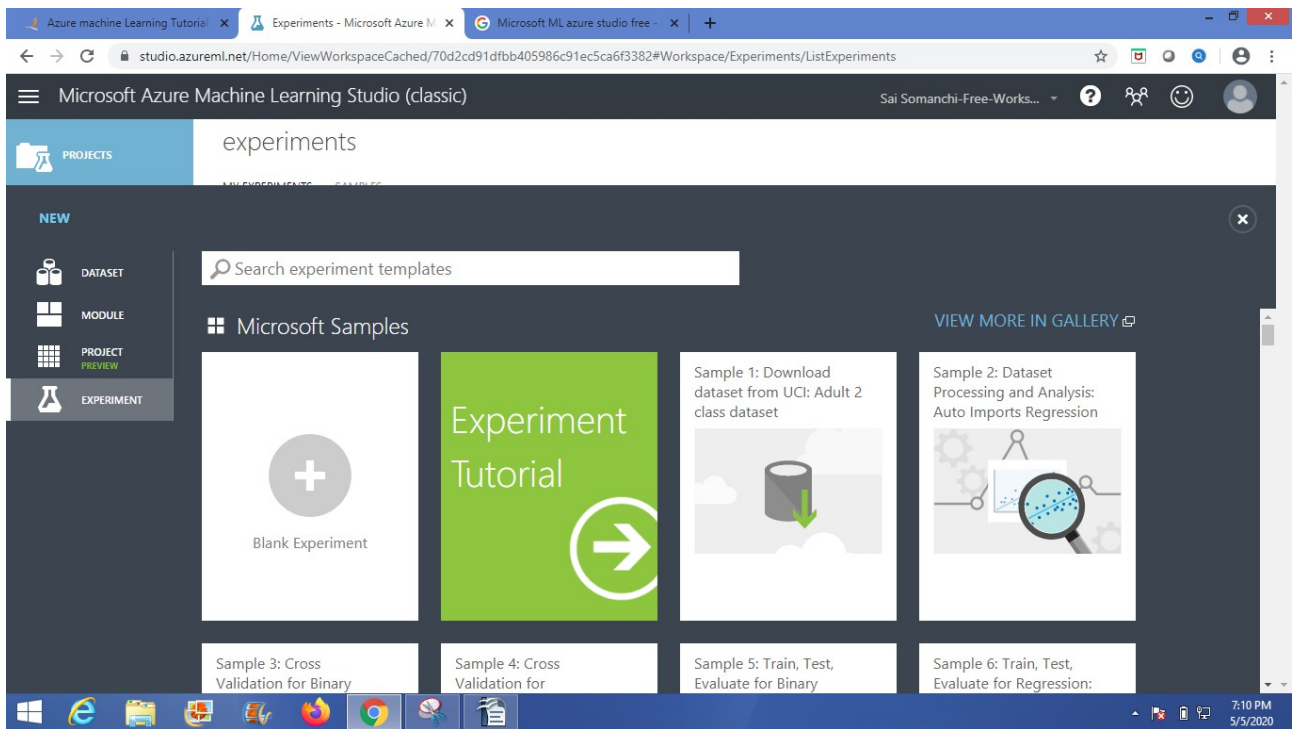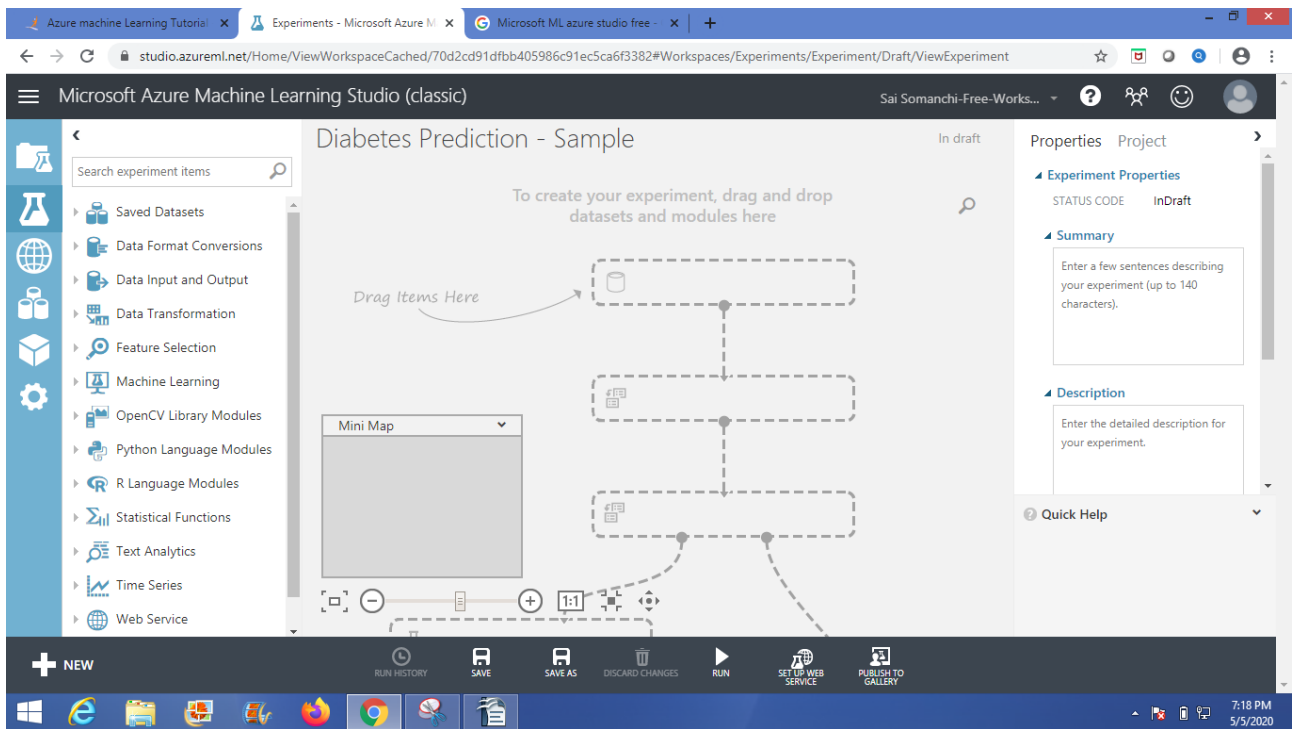
Step 3:- After Clicking on blank experiment below screen will appear.

In place of "**Experiment created on 5/5/2020**" give your own experiment name.

New name Diabetes Prediciton – Sample.

Step 4: Now, select a dataset. There are many sample datasets available for experiment. Take the sample dataset of Diabetes binary classification.

1. From the menu on the left, select **Saved Datasets**

2. For more insights on sample datasets, click on **Samples** and go through the list of sample datasets available.

Select **Pima Indian Diabetes Binary Classification Dataset**, drag it to the center of the screen and drop it.
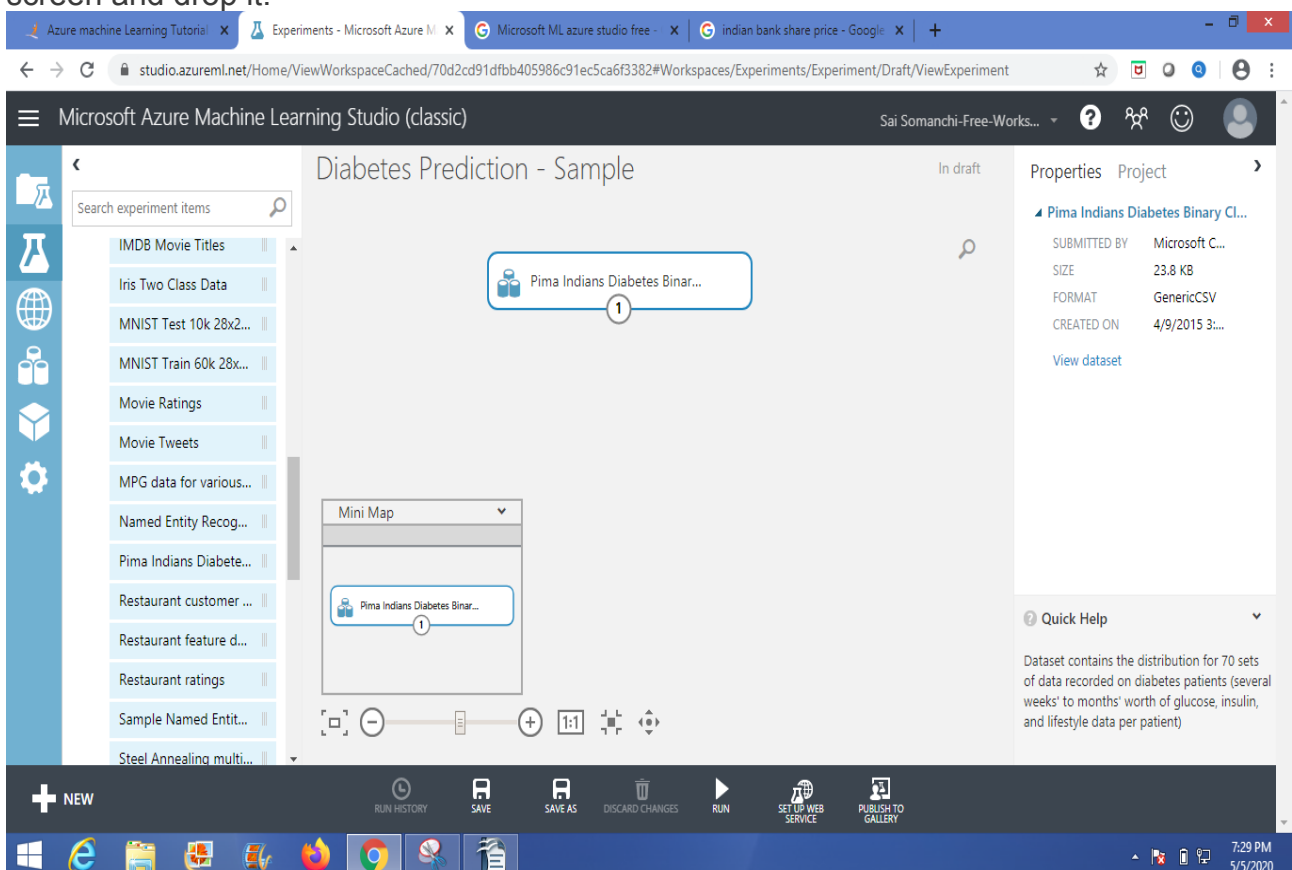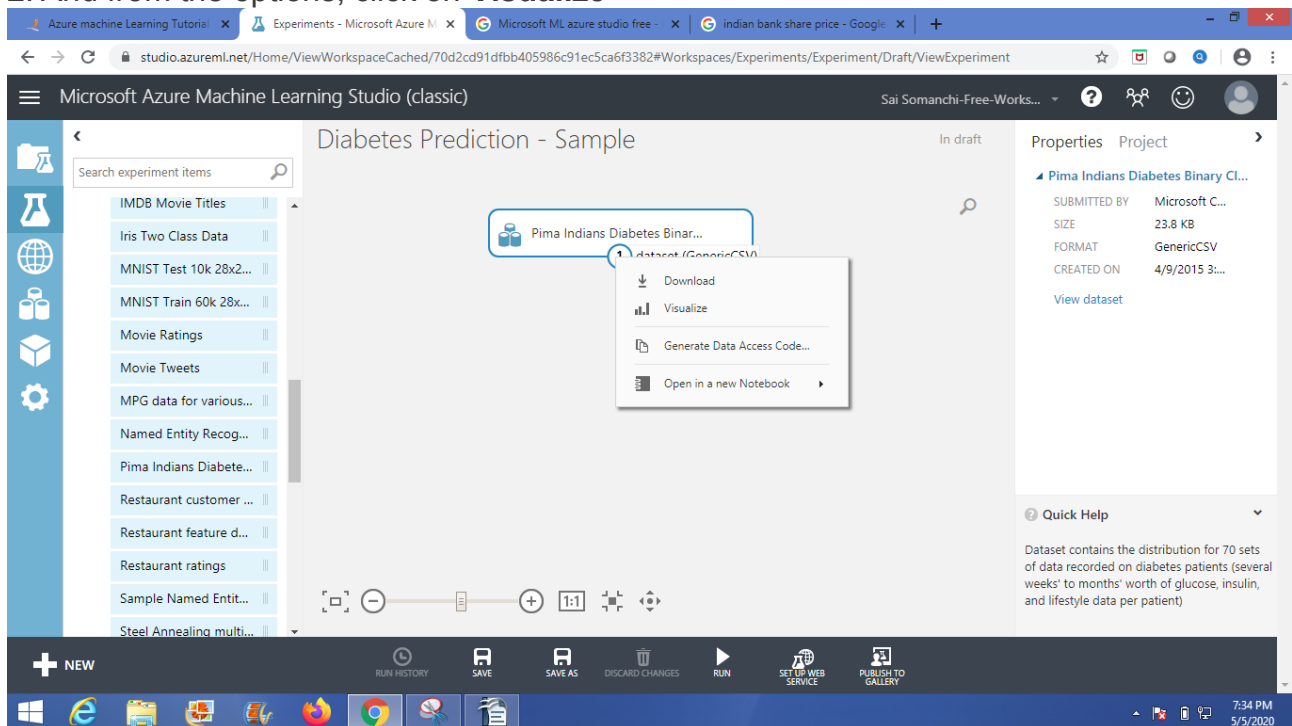
Then Below screen appear.
In above screen – small table named as MINI MAP is there – you can minimize that by clicking
Down arrow – Then below screen appear.
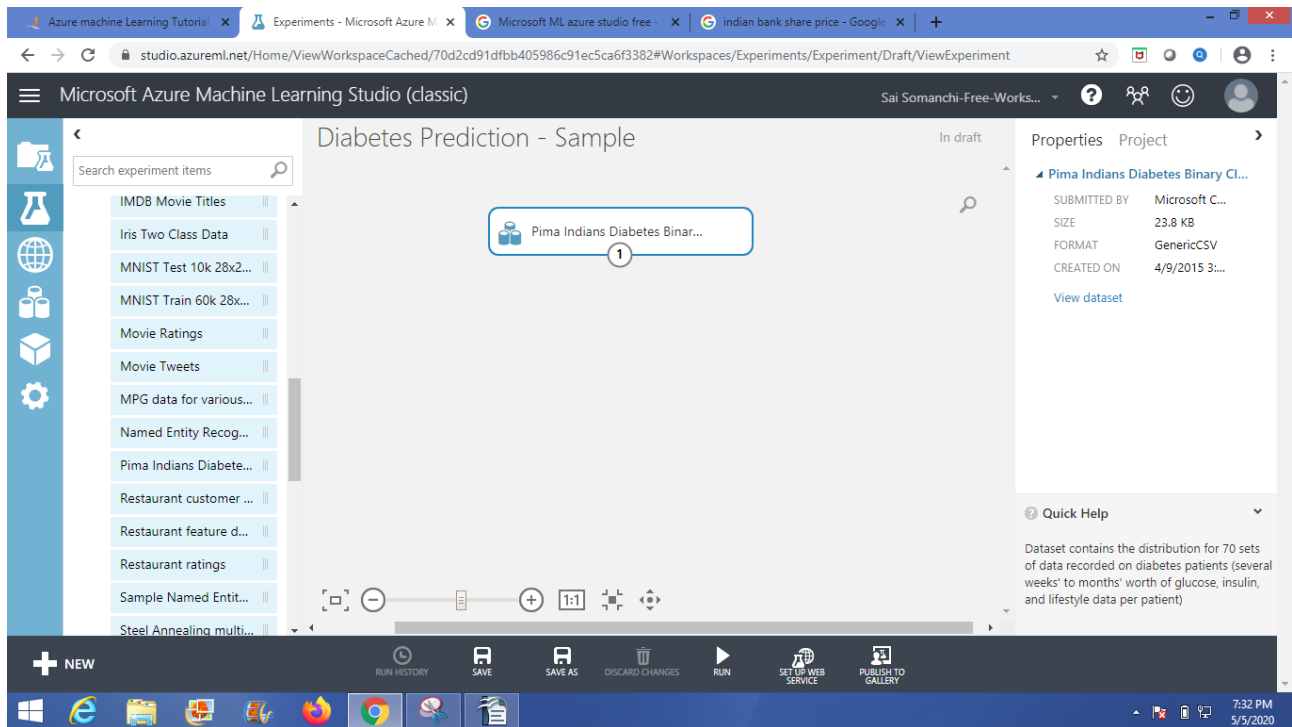
Step 5: Now that we have our dataset, let us see what it has.

1. Click on 1 in circle.

2. And from the options, click on **Visualize**

Once we click on Visualize, we will be able to see what our data looks like. Here is our data that we will be working with: **Below screen appear.**
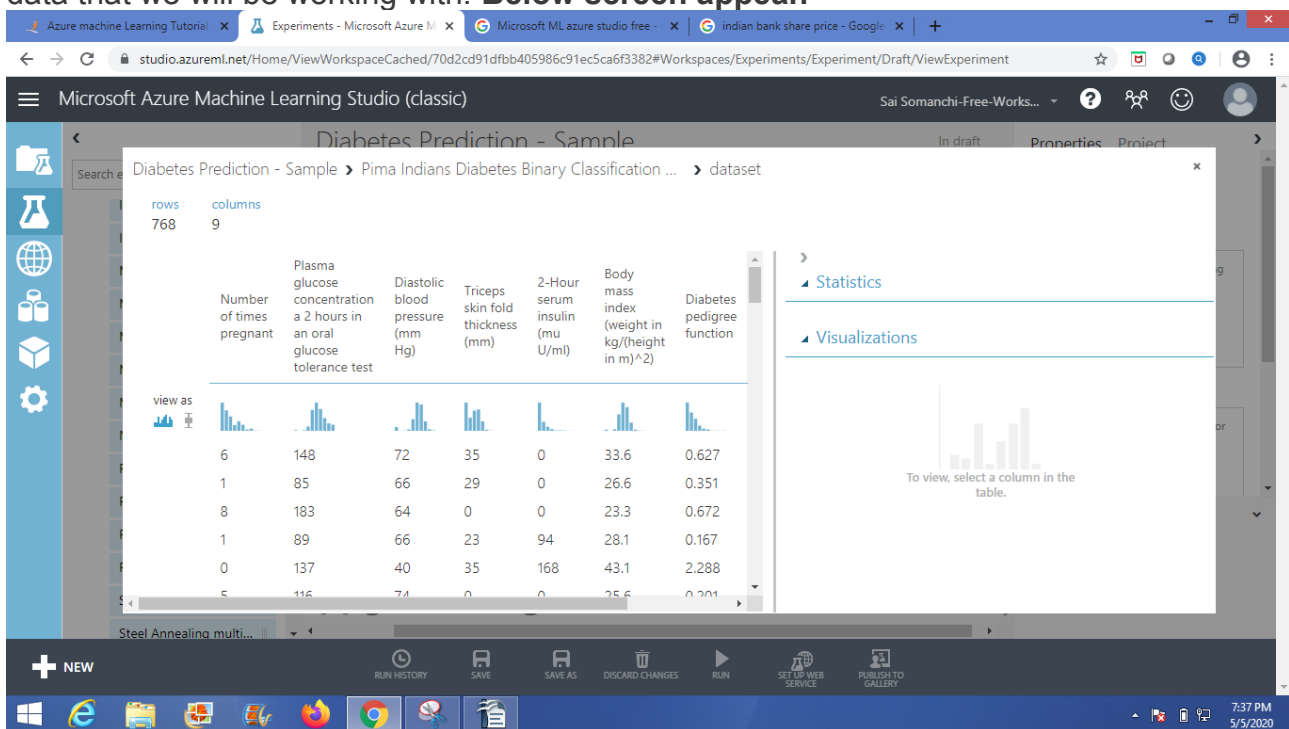
You Can generate statistics and Visualization by selecting a column.

Here I selected Column *"Number of times pregnant"* – you can see in below screen right side Statistics and Visualizations are generated.

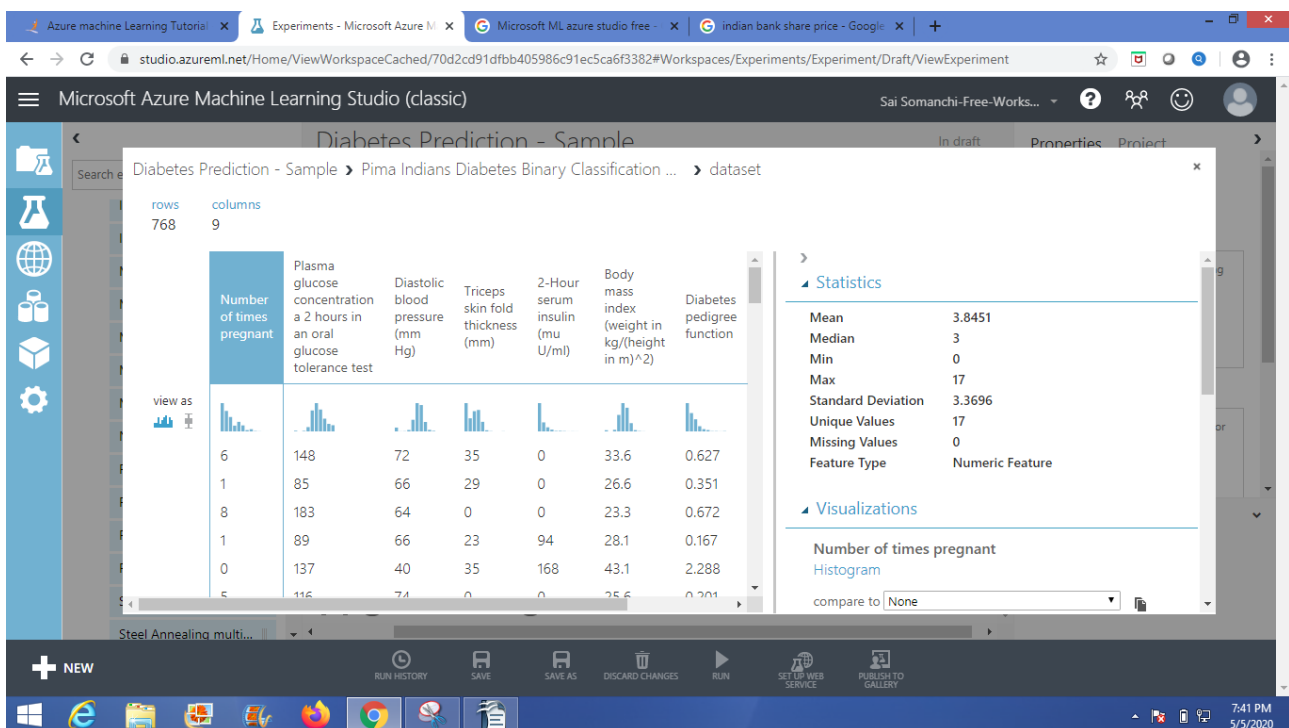**Step 6:** Now, we have to select columns that are relevant to train our model. Here is how we do it:

1. On the left-hand side, there is a search bar, where we will search(**Under Manipulation**) for **Select Columns in Dataset**

2.Drag and drop the item below the dataset and connect the two.
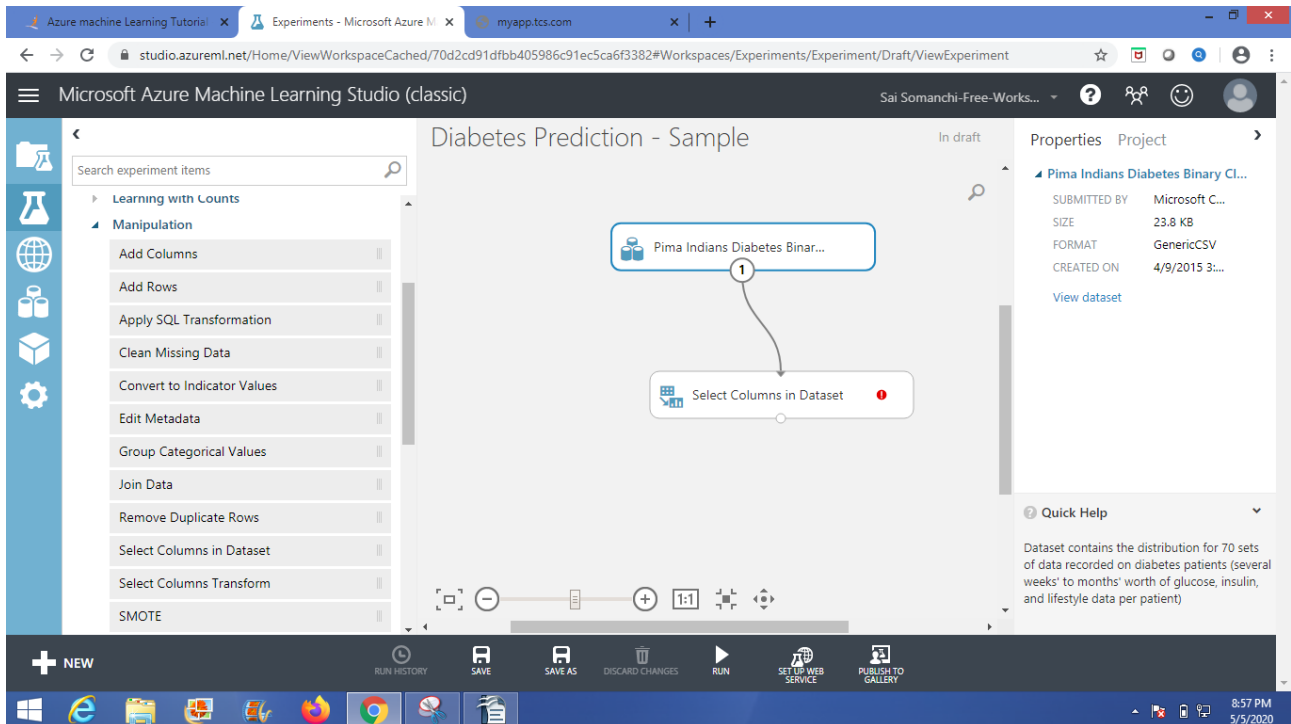
Under Manipulation there are other things which we can explor.

Step 7: After clicking on that entity(Select Column in data set)

Right side two pop-up windows will appear.
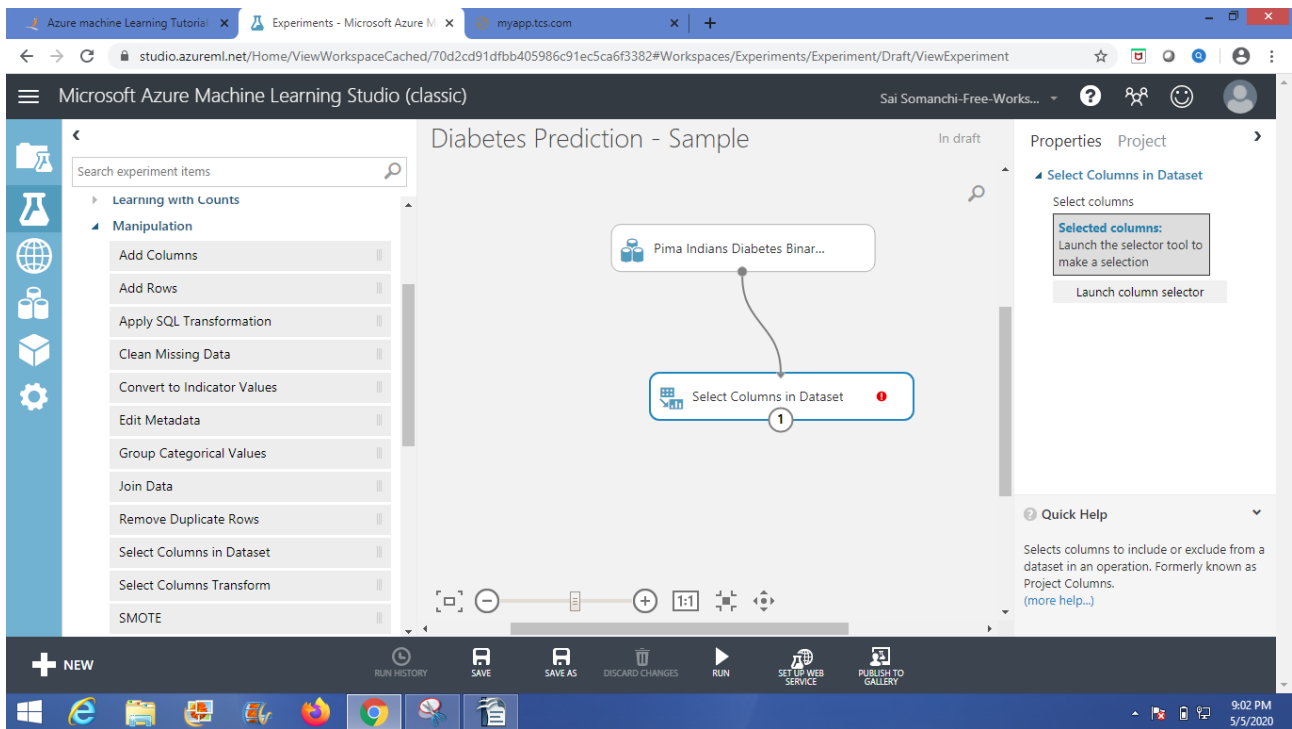
In above screenshot right side "Launch Column Selector" – pop window is there. Please click on that to get below screen.
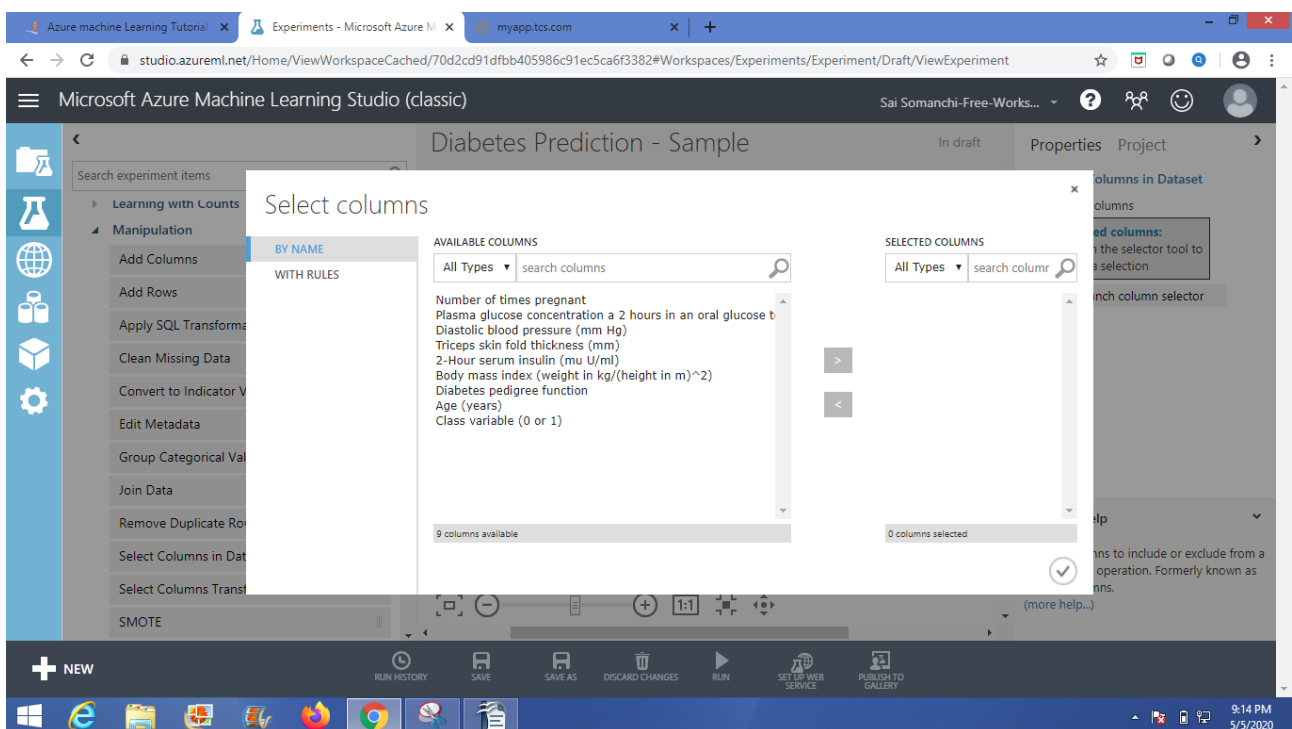
Select column and click on right arrow to select particular column for model.

Here I seleted all columns.

After selecting all columns screen will become like below.

Then click on correct tick mark in below of screen and close the window.

Step 8: Now that we have selected the columns we want to train our model on, we need to split the data into training and test datasets

1. For that, search(Under Sample and split) for **Split Data** and drop it on the workspace

2. Join **Column selector** and **split data**

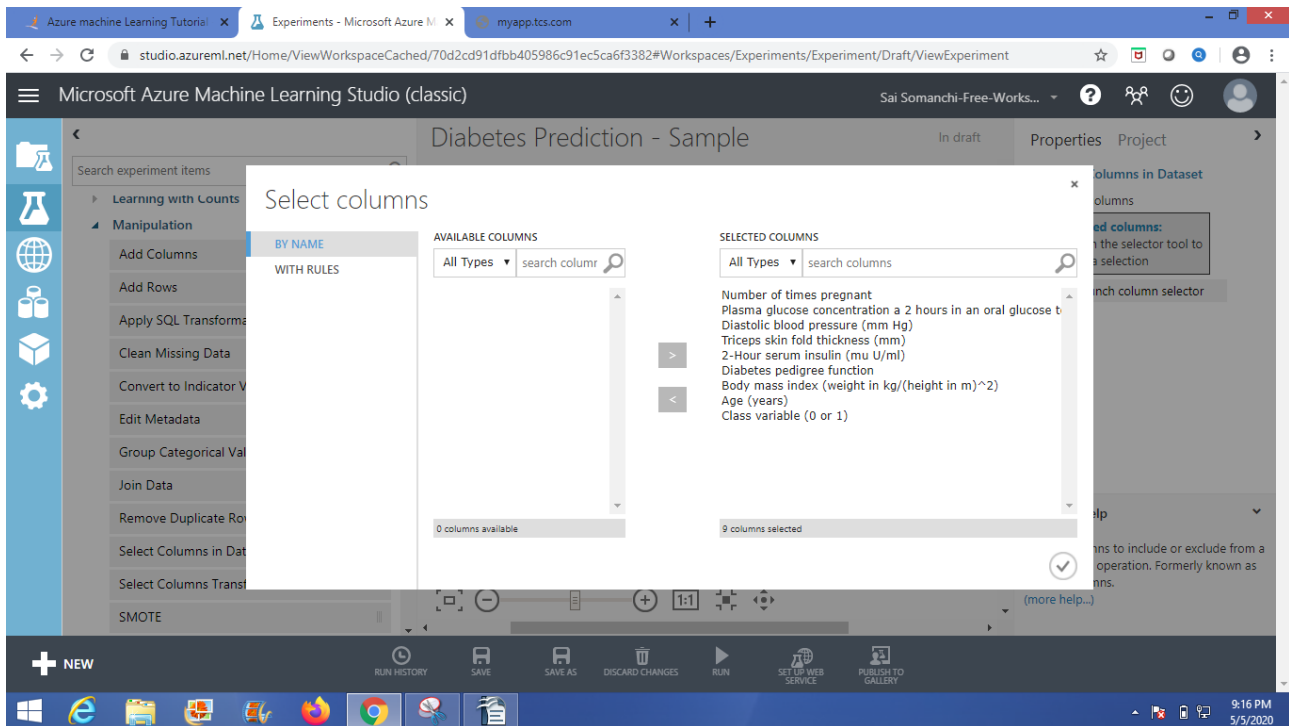3. Toward the right, we can change the percentage of train and test datasets. It is initially 0.5, but we want 70 percent training data and 30 percent test data. Hence, we will change it to 0.7. We can make the ratio 80-20 as well or as per your requirement

In this, is our training dataset and is our test dataset.
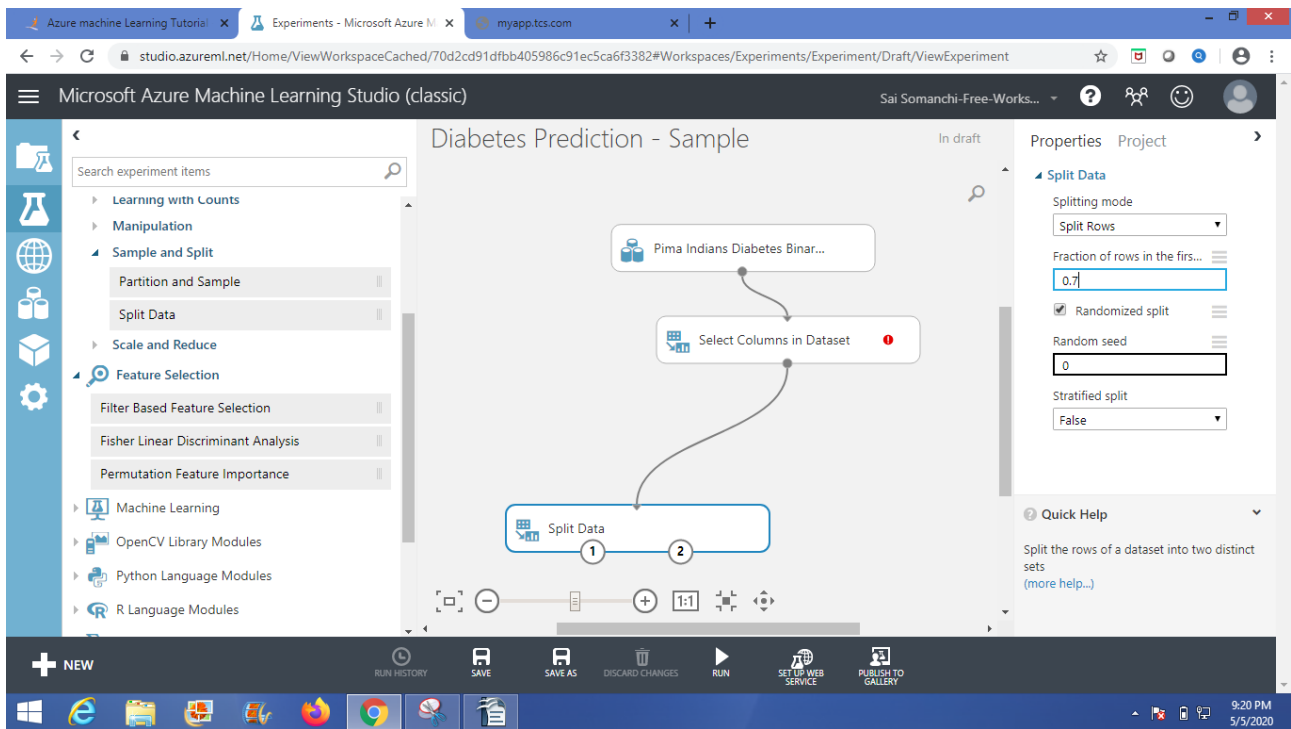
Step 9: We now have our training and testing datasets. Next, we need an algorithm to train our model. The algorithm we are choosing is **two-class logistic regression**.

Logistic regression is used to predict the probability of an outcome.

It predicts the probability of the appearance of an event by providing data to a logistic function.
Since there are two outcomes, it is two-class logistic regression.

For a single value, we can use linear regression. Also, our aim of this prediction model is to find if a person is diabetic or not. Hence, this falls under classification. Search for **Classification**, and under the category we will be able to find this algorithm
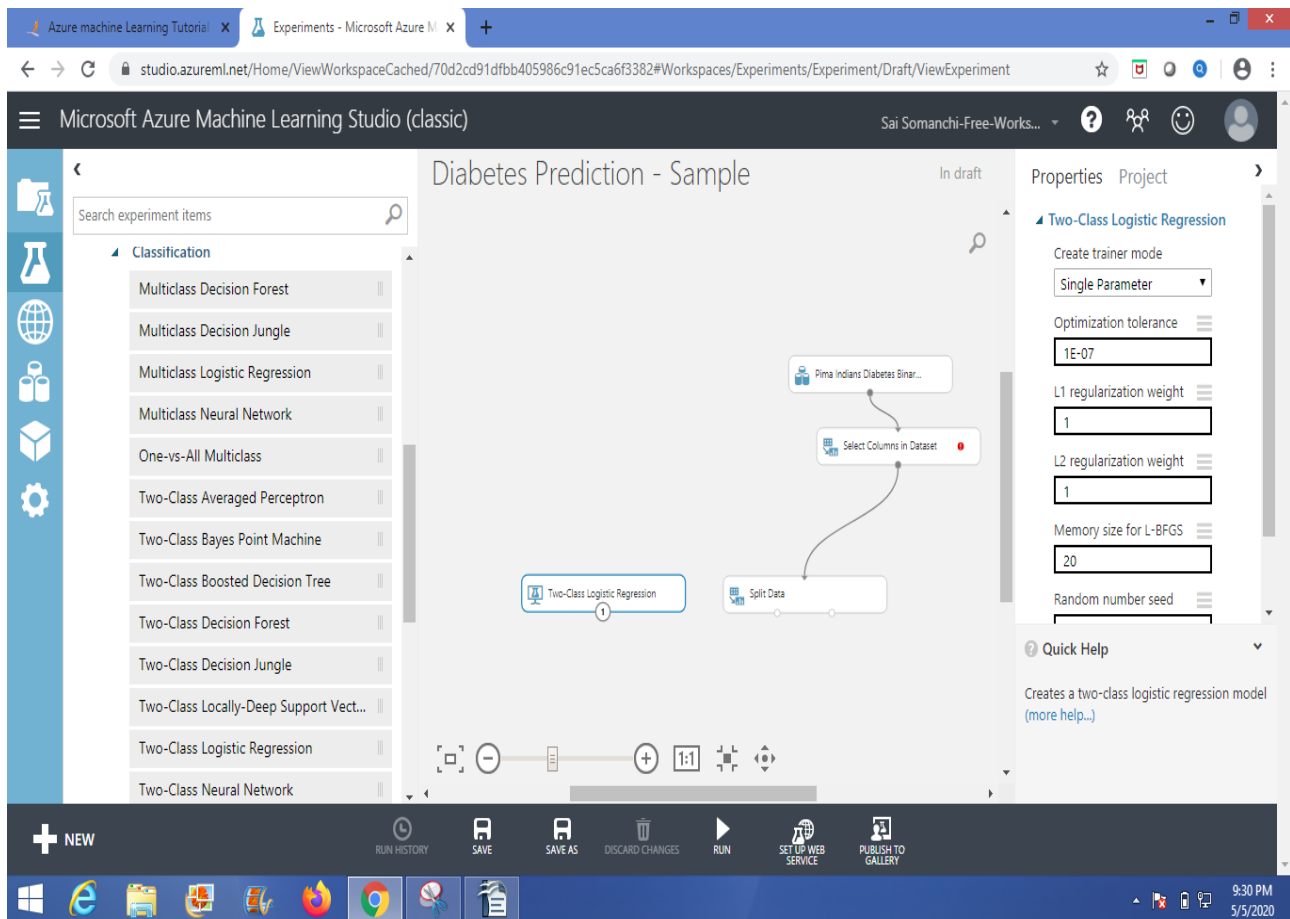
Step 10: It is now time for us to train our model

1. Search for **Train Model** and drop it on the workspace

2. Connect the **Algorithm** to the **train model**.

3. Connect the **training dataset** from **split data to train model**

4.Then, we need to select the column that we need to test (In our case class variable)
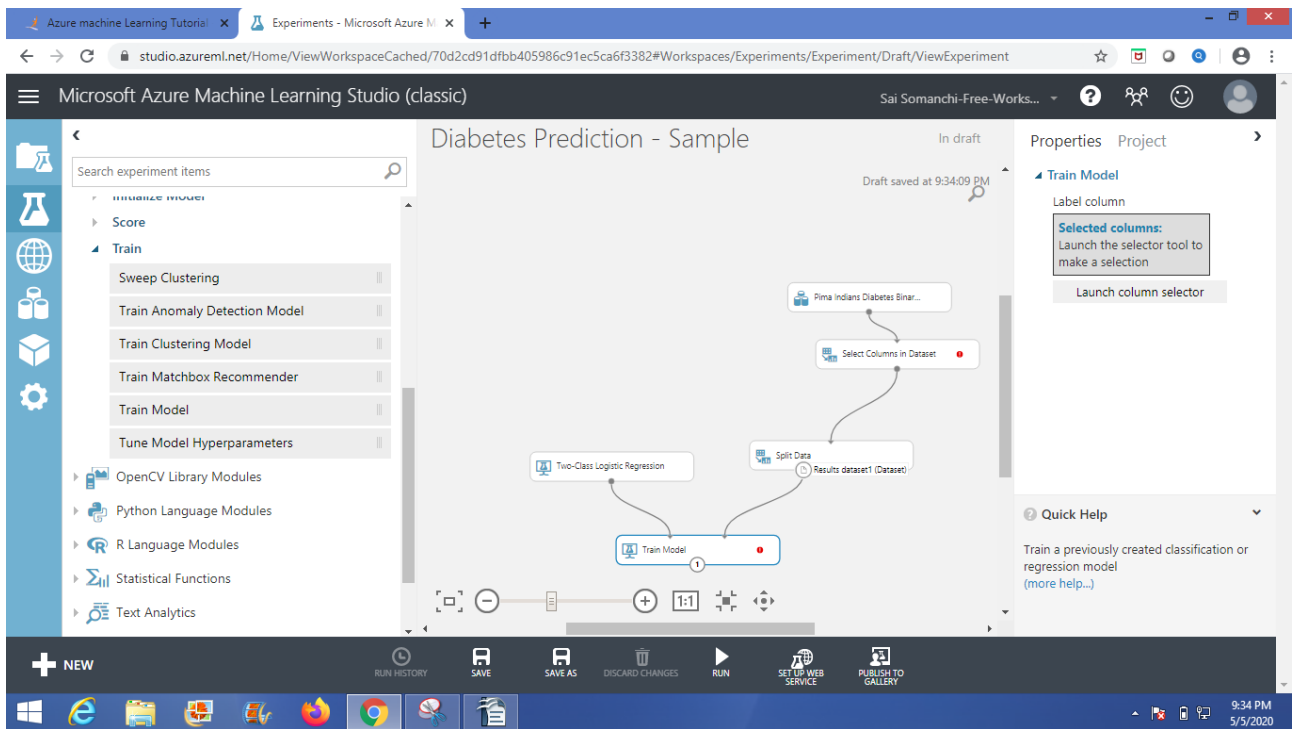
Step 11: Now, we need to score our trained model and then evaluate it

1. Drag and drop the **Score Model** on to the center

2. Connect the **Train Model** and

3. Connect the **Training Dataset** from **Split Data** to **train model**

4. Drag and drop **Evaluate Model**

5. Connect **Score Model** with it

6. **Save** the experiment

7. **Run** the experiment

If you observed All Green Tick marks indicate it run succefully.

Step 12: After running the model, we need to visualize the result and find precision, accuracy, etc.

1. Click on **Evaluate** and then on

2. Select **Visualize** and there will be a pop-up window like below

Here are the results:
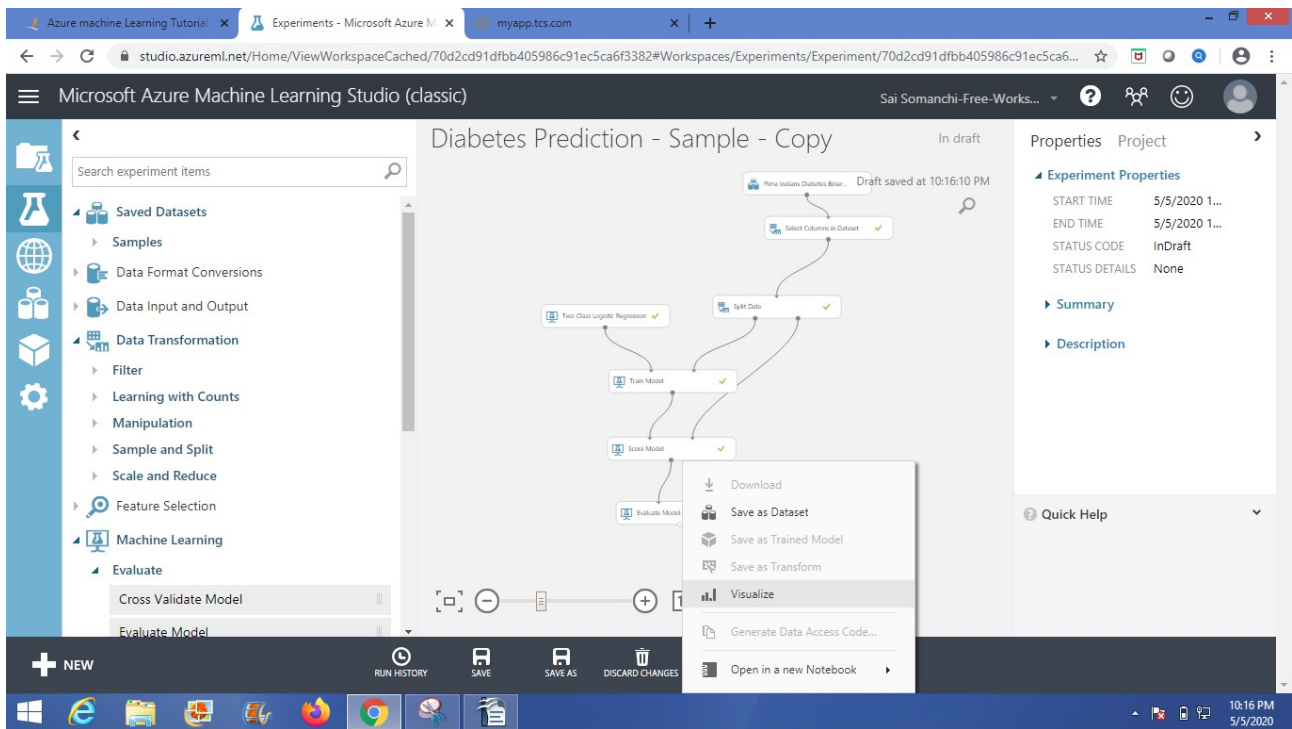
Select Visualize

you will get below results.

Diabetes Prediction - Sample - Copy ➤ Evaluate Model ➤ Evaluation results

True Positive	False Negative	Accuracy	Precision	Threshold		AUC
35	39	0.778	0.745	0.5		0.795

False Positive	True Negative	Recall	F1 Score
12	144	0.473	0.579

Positive Label	Negative Label
1	0

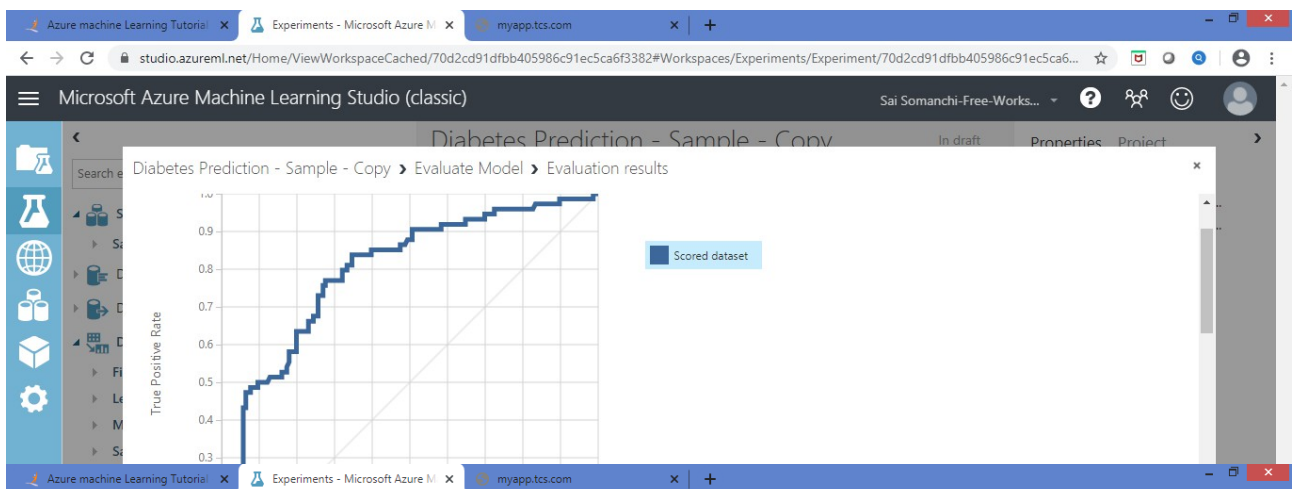| Score Bin	Positive Examples	Negative Examples	Fraction Above Threshold	Accuracy	F1 Score	Precision	Recall	Negative Precision	Negative Recall	Cumulative AUC
(0.900,1.000]	0	0	0.000	0.678	0.000	1.000	0.000	0.678	1.000	0.000
(0.800,0.900]	3	2	0.022	0.683	0.076	0.600	0.041	0.684	0.987	0.000
(0.700,0.800]	6	0	0.048	0.709	0.212	0.818	0.122	0.703	0.987	0.000
(0.600,0.700]	7	3	0.091	0.726	0.337	0.762	0.216	0.722	0.968	0.004
(0.500,0.600]	19	7	0.204	0.778	0.579	0.745	0.473	0.787	0.923	0.019
(0.400,0.500]	12	20	0.343	0.743	0.614	0.595	0.635	0.821	0.795	0.087

With this, model development has been completed.

********************************Happy Learning**

Please watch below video's for more clarity.

1.https://www.youtube.com/watch?v=F_o9GYBDTF4

2.https://www.youtube.com/watch?v=nNHSlpJMOJs

Please find links for detailed process.

Azure ML Studio:
https://studio.azureml.net/
How to use Azure ML Studio
https://docs.microsoft.com/en-us/azure/machine-learning/studio/create-experiment
https://towardsdatascience.com/how-microsoft-azure-machine-learning-studio-clarifies-data-science-8e8d3e6ed64e

**********************************Assignment**

1.Please take any dataset from avalaible datasets in sample and do the similar operations and prepare your own document using ML Azure free framework.

Cover all the below mentioned topics in model development using Azure.

1.Select and load dataset.

2.Generate all basic statistics and draw graphs using Visualization tool with various column combinations.

3.Clean data – Remove or impute null values, remove duplicates and other operations.

4.Split dataset in 80:20 ratio.

5.Select Proper ML model and apply against dataset.

6.Evaluate model and Cross validate.

7.Generate all results using Visualization and other parameters.

8.Fine tune the model further by changing the various inputs.

9.Try to workout the same model in normal anaconda environment (with out using cloud)and compare the process.

10.Try to understand entire life cycle.

11.Please try to know other cloud platforms for ML like Amazon and Google.

****