Working with datasets on Azure Machine Learning Studio

Introduction:

This article is again a continuation from my previous one which leads up for the next step of working with datasets on Azure ML. Here we will be working with uploading the datasets and understanding them.

What are datasets?

Datasets – Datasets holds a collection of information or data that are related towards a unique identification or for a manipulation towards an experiment that we build here in Machine Learning.

Microsoft Azure here comes up with few datasets which are already available with Azure Machine Learning Studio, we call it as sample datasets in Azure ML which holds collection of data's for various manipulation tasks.

Azure Machine Learning Studio has several sample datasets, additional datasets are available from other sources also which we can google for or we can get from other sources. We do have other Machine Learning Repositories for which we can build a Machine Learning models.

Other sources for datasets –

http://archive.ics.uci.edu/ml/index.php

https://www.kaggle.com/

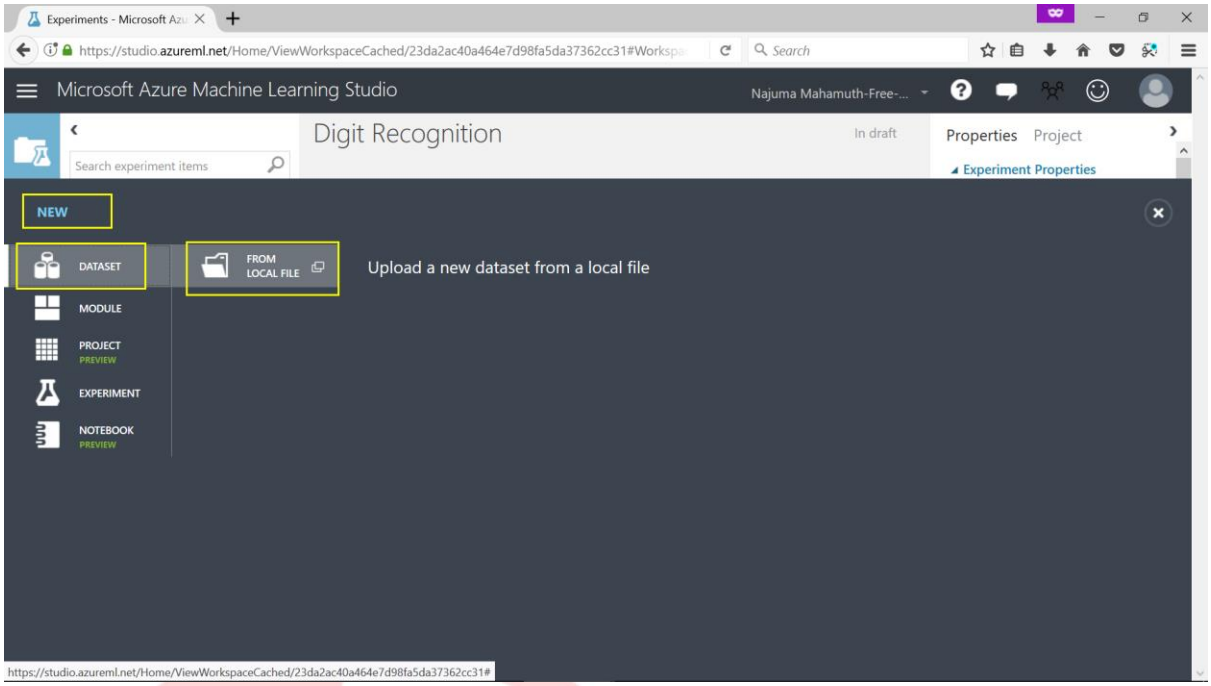
http://data.gov

Upload the datasets and visualising them:

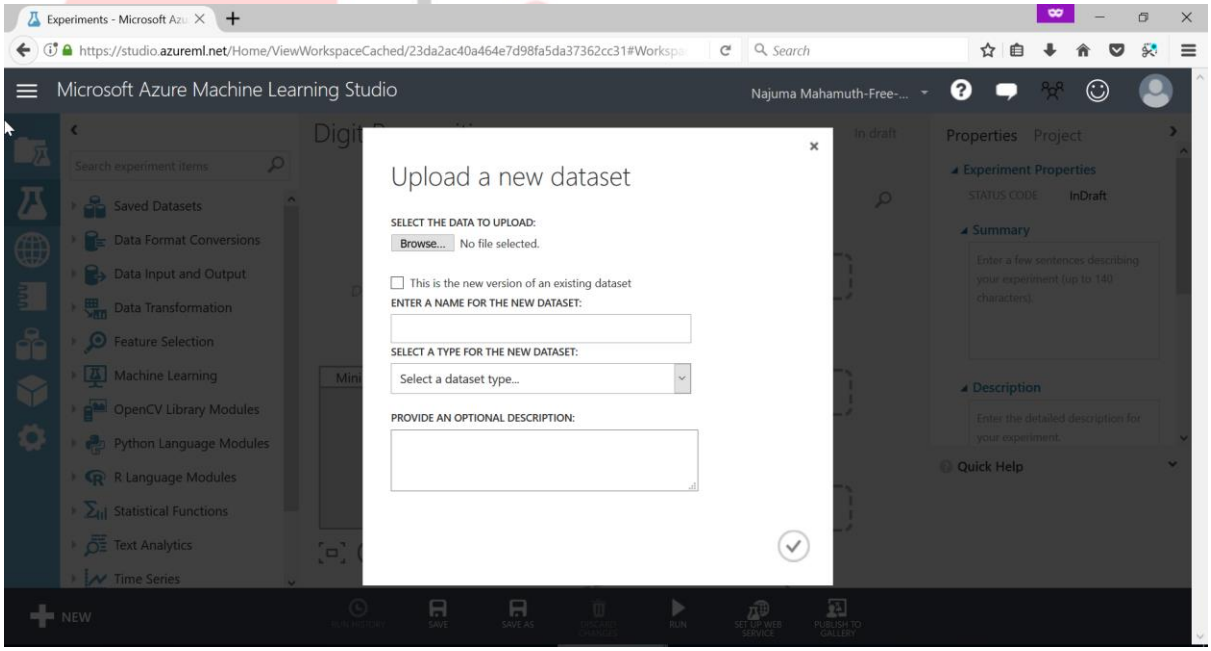
By this demo we will be uploading a dataset which will be having a collection of data generated with help of hand written and scanned digits within 0 to 9 from a few set of people. Followed up by that we will be using the same dataset to train the ML model which can predict the handwritten digits.

- Go for New → Dataset → From Local File and select the dataset which has to be uploaded.

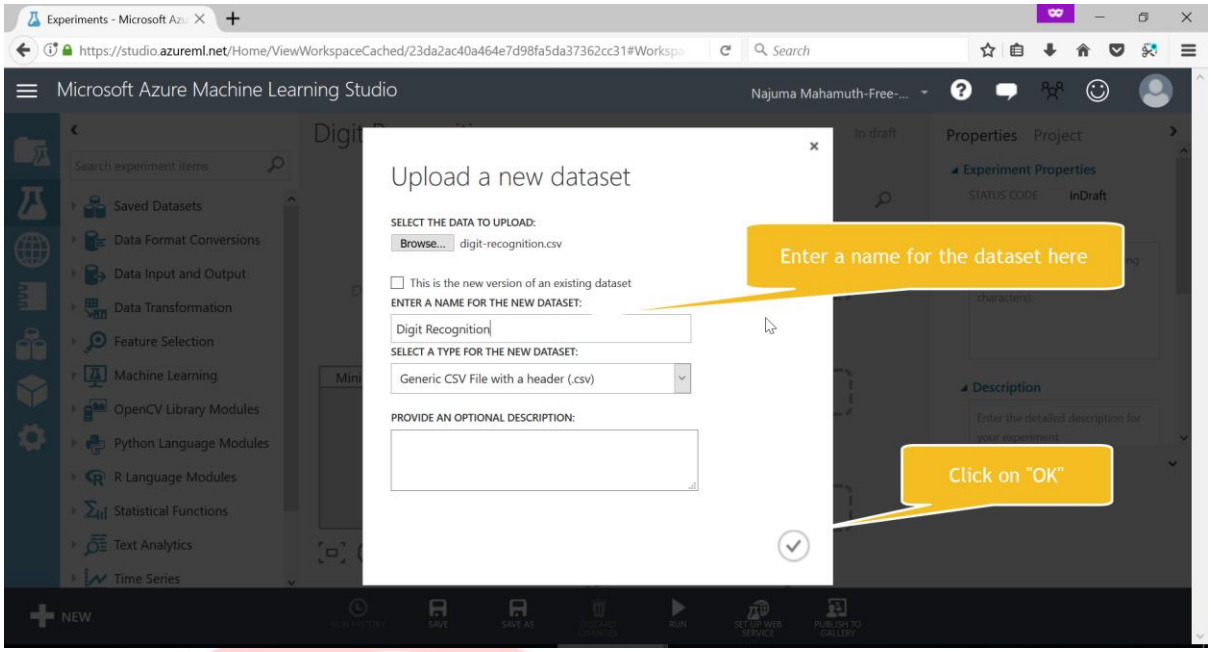
Note – You can find the dataset "digit-recognition.csv" from our GitHub repository.



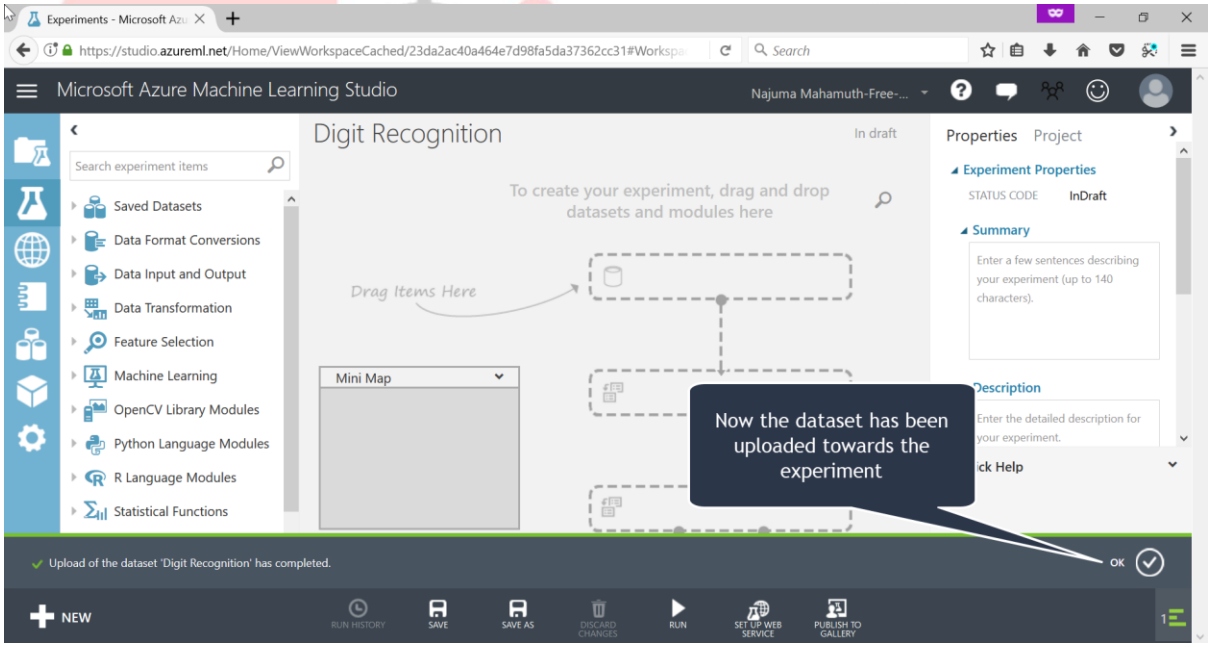
We will be landing up on a page as shown below in which we can upload the dataset and we should enter a name for the new dataset which has been uploaded and type of the dataset.



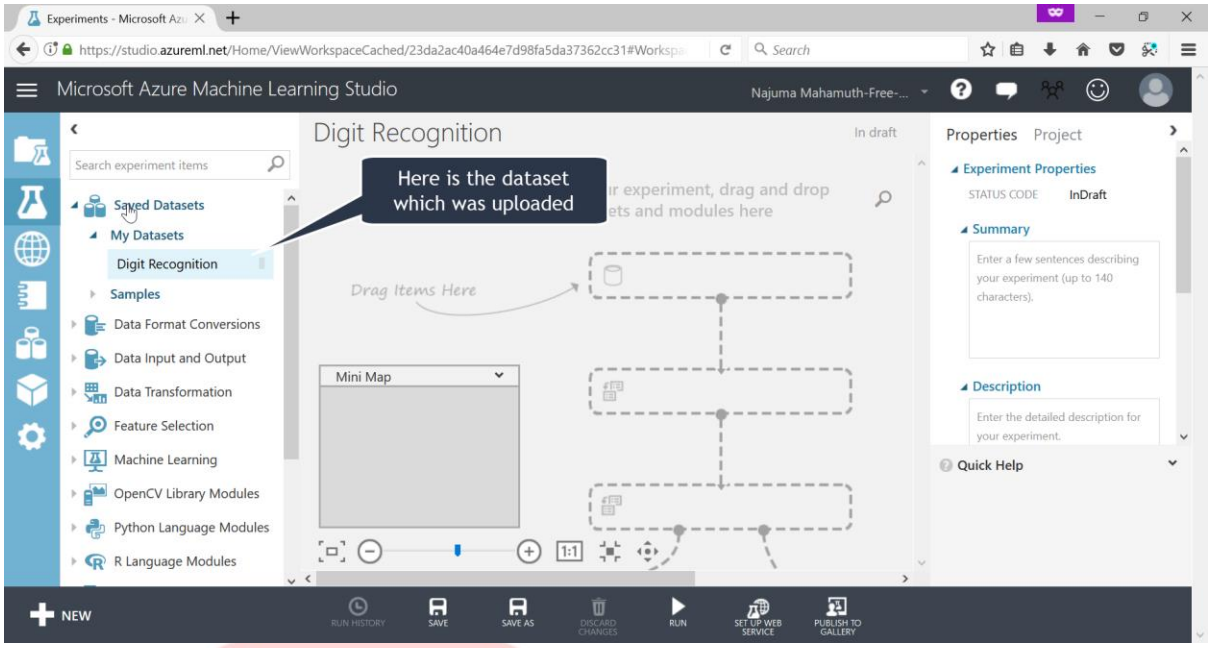
- Select the dataset which has been downloaded from the GitHub repository and upload the same, enter the name as shown below and click on OK.



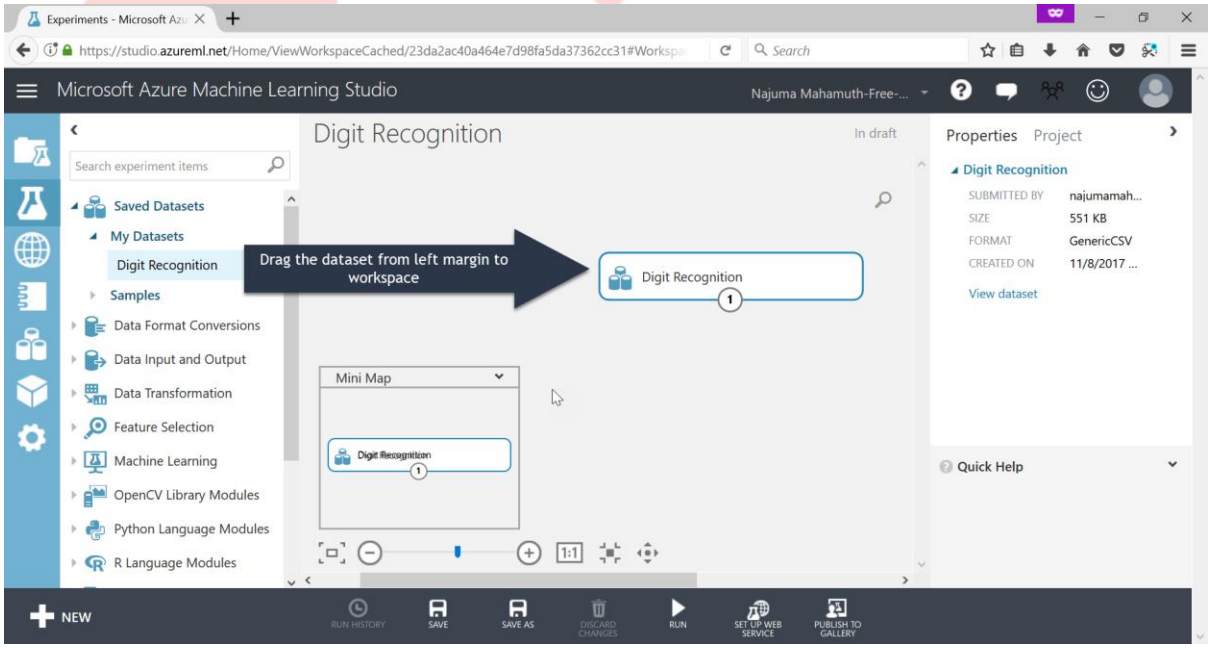
We can find a notification at the bottom part of Azure ML Studio page stating that the dataset has been uploaded successfully towards the experiment as shown in the image below: Click on “OK” to close it.



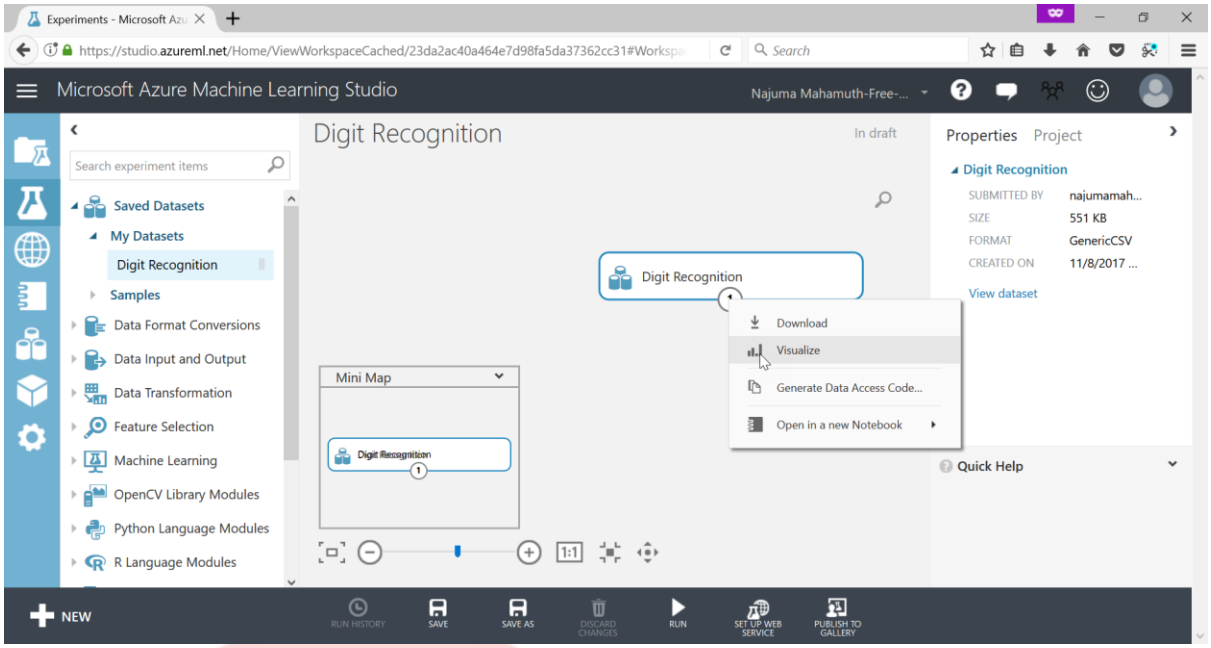
Clicking on “Saved Datasets → My Datasets →Digit Recognition” will help us to find the dataset which we have uploaded.



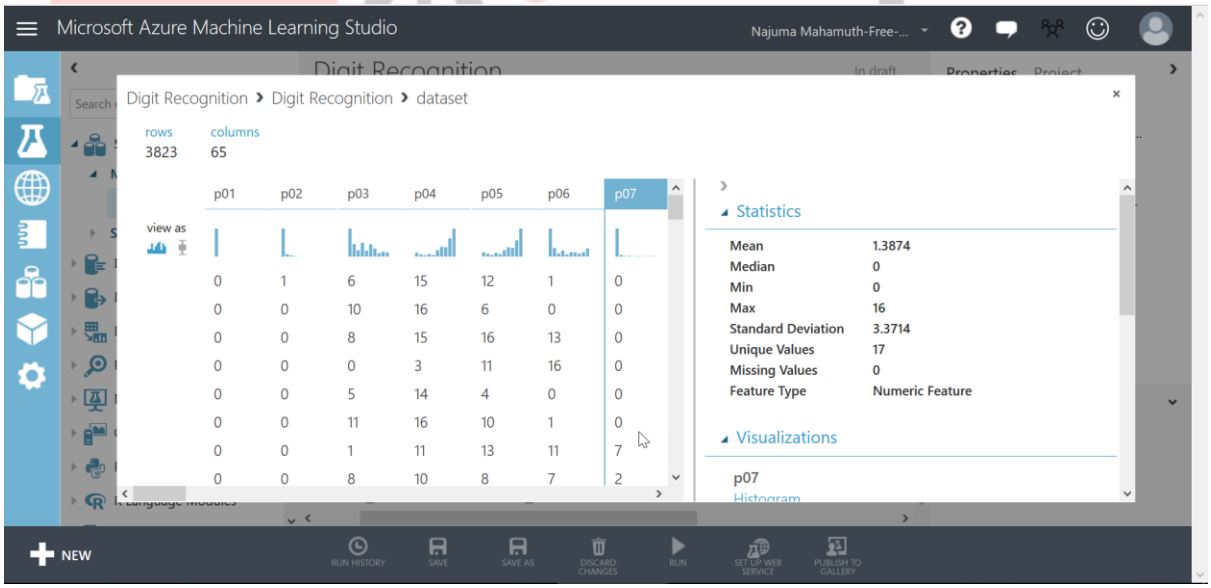
- Now drag the uploaded dataset towards Azure ML Studio from the left pane towards the workspace to work on it for the Digit Recognition experiment.



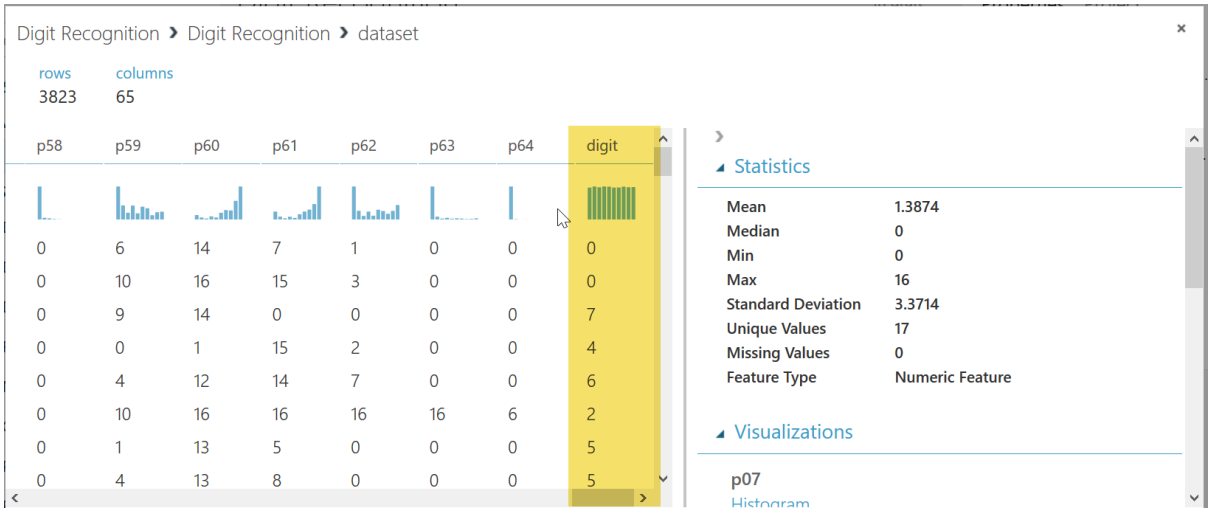
- Right click on the “Dataset” and go for “Visualise” to visualise the dataset which has been uploaded.



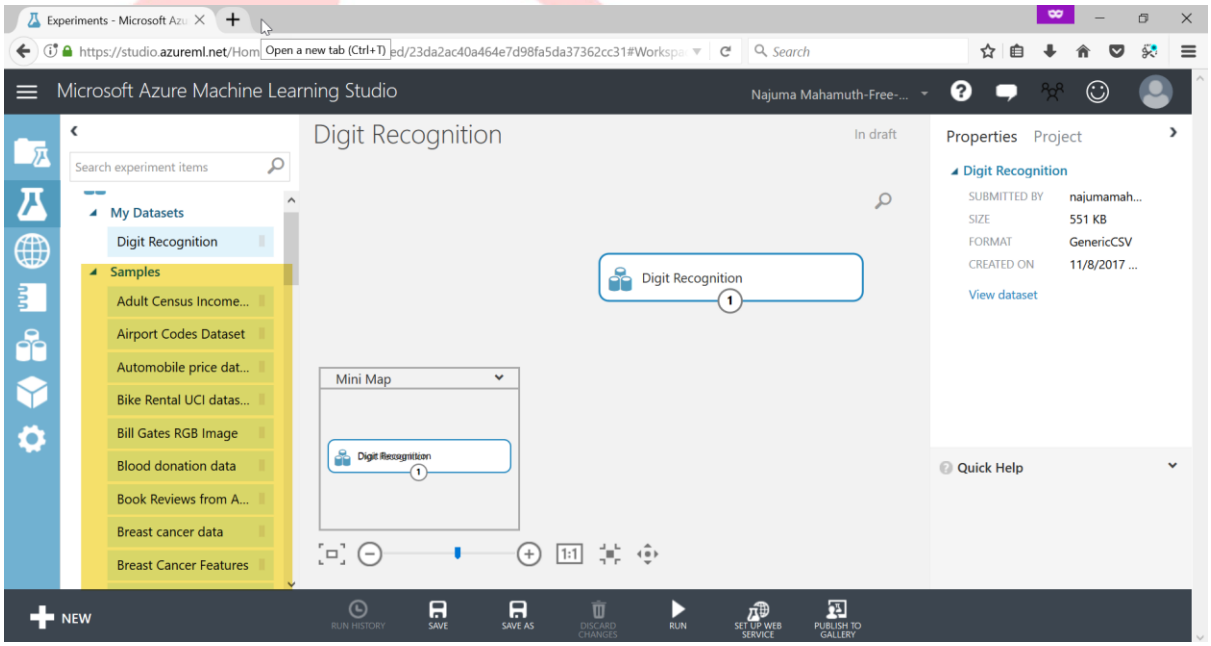
This will open up a page as shown below which holds the data with 3823 rows and 65 columns, normally all datasets will be having a particular set of rows and columns, here the datas available represents the digits from 0 to 9. The first 64 columns contains values from 0 to 16 representing 4*4 blocks of pixels.



The last column named “digit” represents the value of corresponding data in that particular row. It’s a target value which the model will be predicting by the way we train them. Clicking on the “x” at the top right corner will help us in closing this dataset visualising page.



We can go for the Azure ML Studio page and click on “My Datasets” in which we can find “Samples” and inside samples we will be having the sample datasets which azure provides for us. This datasets can help us in checking the airport codes, automobile price prediction, prediction for breast cancer, prediction for rent of bikes, credit card fraudulent prediction, etc.,



Follow my next writing to go with modelling.