**CRYPTO CURRENCY FUTURE PRICE PREDICTION**

-DATA SCIENCE WITH PYTHON PROJECT

***INTRODUCTION:***

A cryptocurrency is a type digital currency which can be used a mode of online payment or make an individual to sell/purchase cryptocurrency to utilizing various monetary forms. The Cryptocurrency transaction are stored and maintained in an electronic ledger and even supports blockchain transactional databases. Bitcoin is the earliest invented cryptocurrency and has been in trading since 2009. There are around 7577 different Cryptocurrencies available in the market. We can use straightforward utilization methods like Machine learning or Data science to predict its values. The model can be used to predict the market price of each cryptocurrency. This model can be useful for Several traders who are scared about investing in cryptocurrency because of the fear of frequent drop in price. But with data science and constant monitoring, it will be easier to identify the trend in stock price. It also becomes easier to find the right time to invest and to draw the money.

***OVERVIEW:***

This project aims to predict the future price of crypto currency.

**Input:** The previous prices of crypto currency and volume.

**Output:** Future prices of crypto currency

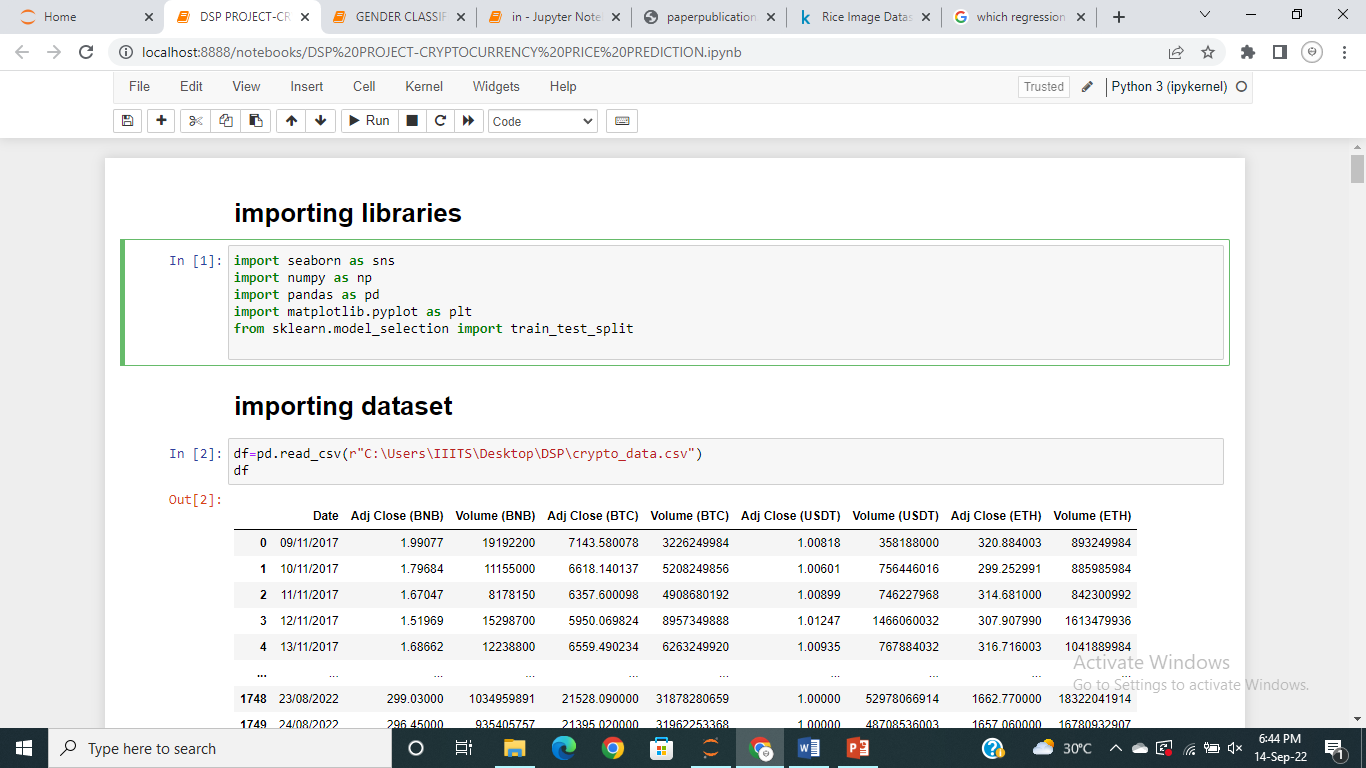
-By considering the past prices as input we can predict the future prices.

***About Dataset:***

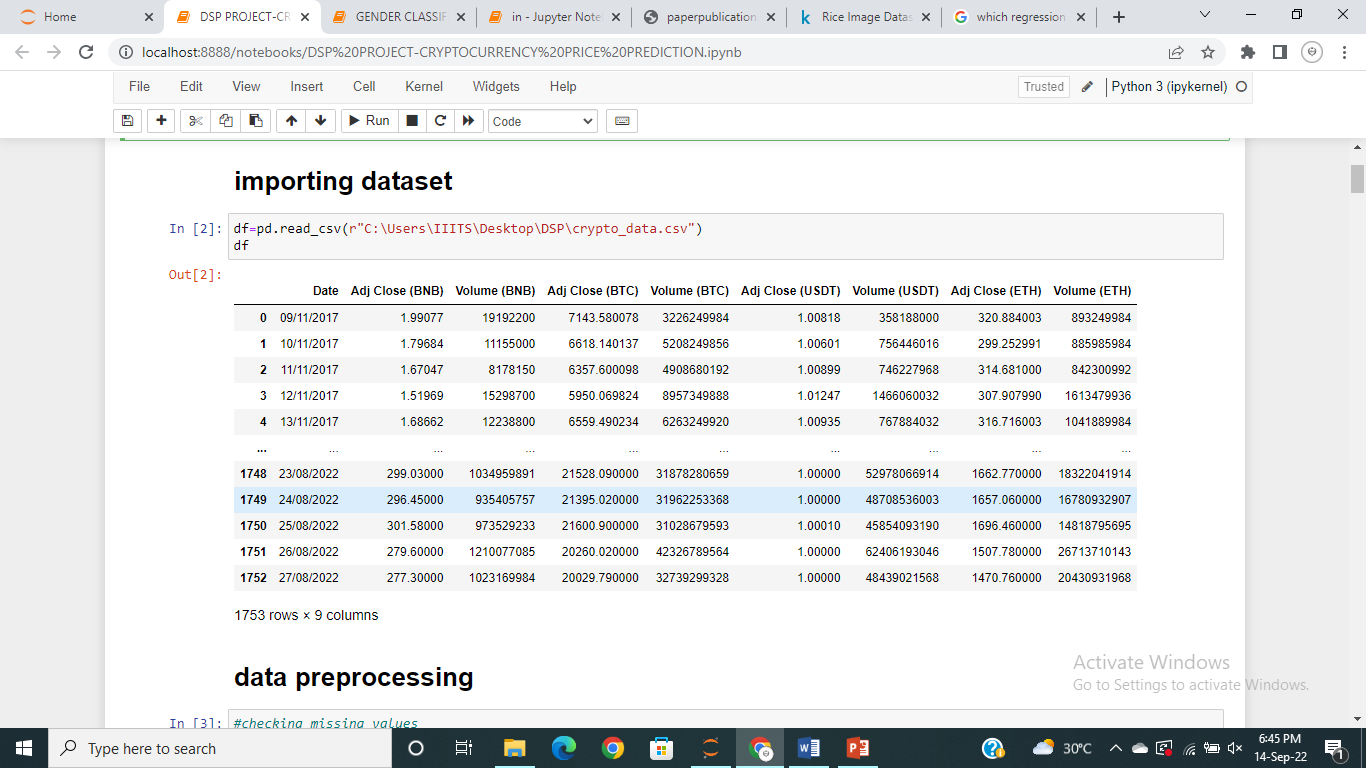
* It is having 9 columns and 1753 rows.
* Each row is having date and bitcoin(BTC) ,binance(BNB),ethereum(ETH),Tether(USDT) adjacent close price and volumes of that coins on that day
* I took the dataset from kaggle

***Steps Involved in Project:***

***1.Importing Libraries:***



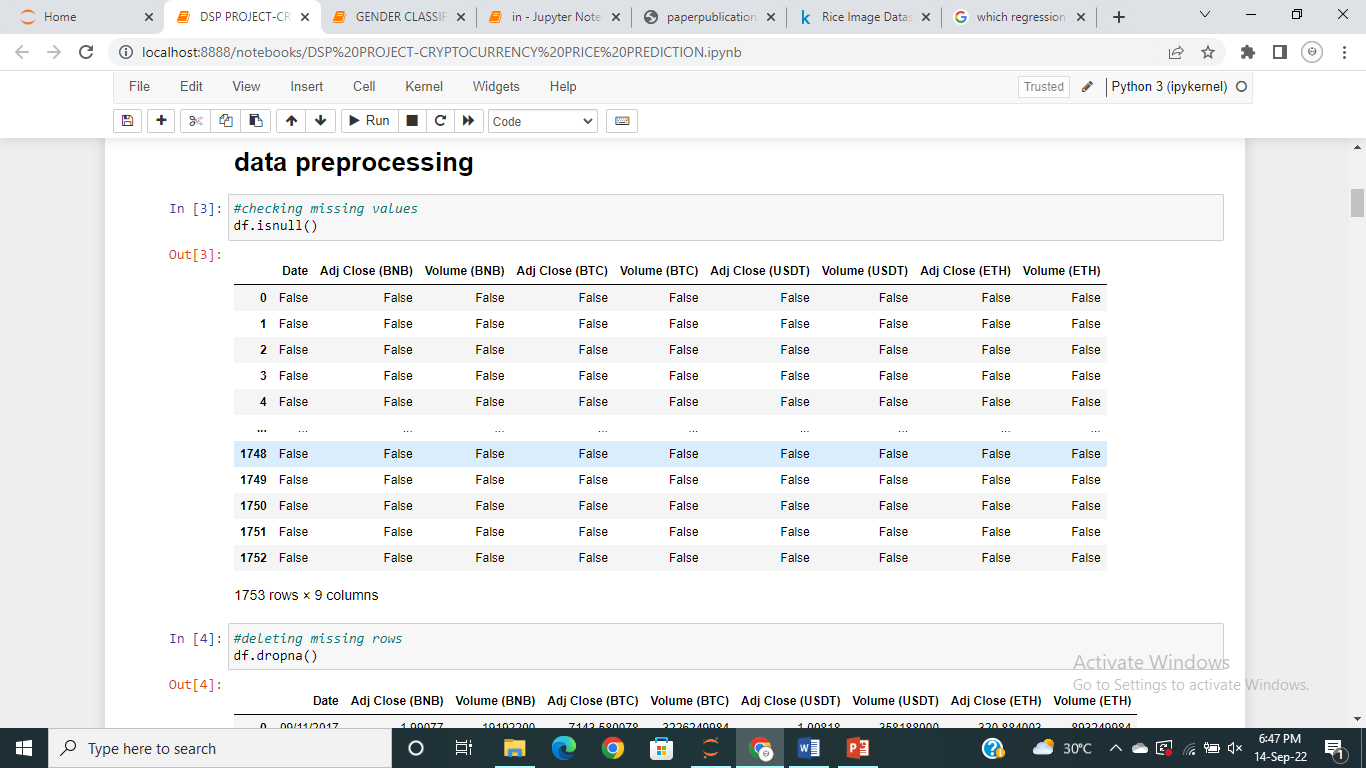
* Firstly we have imported the libraries used in this model
* ***Seaborn:*** Seaborn is a Python data visualization library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics.Here I have used this to Visualise the data.
* ***Numpy:*** NumPy is a Python library used for working with arrays.It also has functions for working in domain of linear algebra, fourier transform, and matrices.here I have used this to work with arrays in my data
* ***Pandas:*** **pandas** is a fast, powerful, flexible and easy to use open source data analysis and manipulation tool  
  built on top of the [Python](https://www.python.org/) programming language.here I have used to work with data.
* ***Matplotlib****:* Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python. Most of the Matplotlib utilities lies under the pyplot submodule, and are usually imported under the plt alias.Here I have used this in visualizing analysis.
* ***Sklearn:***Scikit learn has various [classification](https://en.wikipedia.org/wiki/Statistical_classification) , [regression](https://en.wikipedia.org/wiki/Regression_analysis) and [clustering](https://en.wikipedia.org/wiki/Cluster_analysis) algorithms including [support-vector machines](https://en.wikipedia.org/wiki/Support_vector_machine), [random forests](https://en.wikipedia.org/wiki/Random_forests), [gradient boosting](https://en.wikipedia.org/wiki/Gradient_boosting), [*k*-means](https://en.wikipedia.org/wiki/K-means_clustering) and [DBSCAN](https://en.wikipedia.org/wiki/DBSCAN), and is designed to interoperate with the Python numerical and scientific libraries [NumPy](https://en.wikipedia.org/wiki/NumPy) and [SciPy](https://en.wikipedia.org/wiki/SciPy). Here I have used this for training and testing the data and model selection.

***2.Importing Dataset:*** 

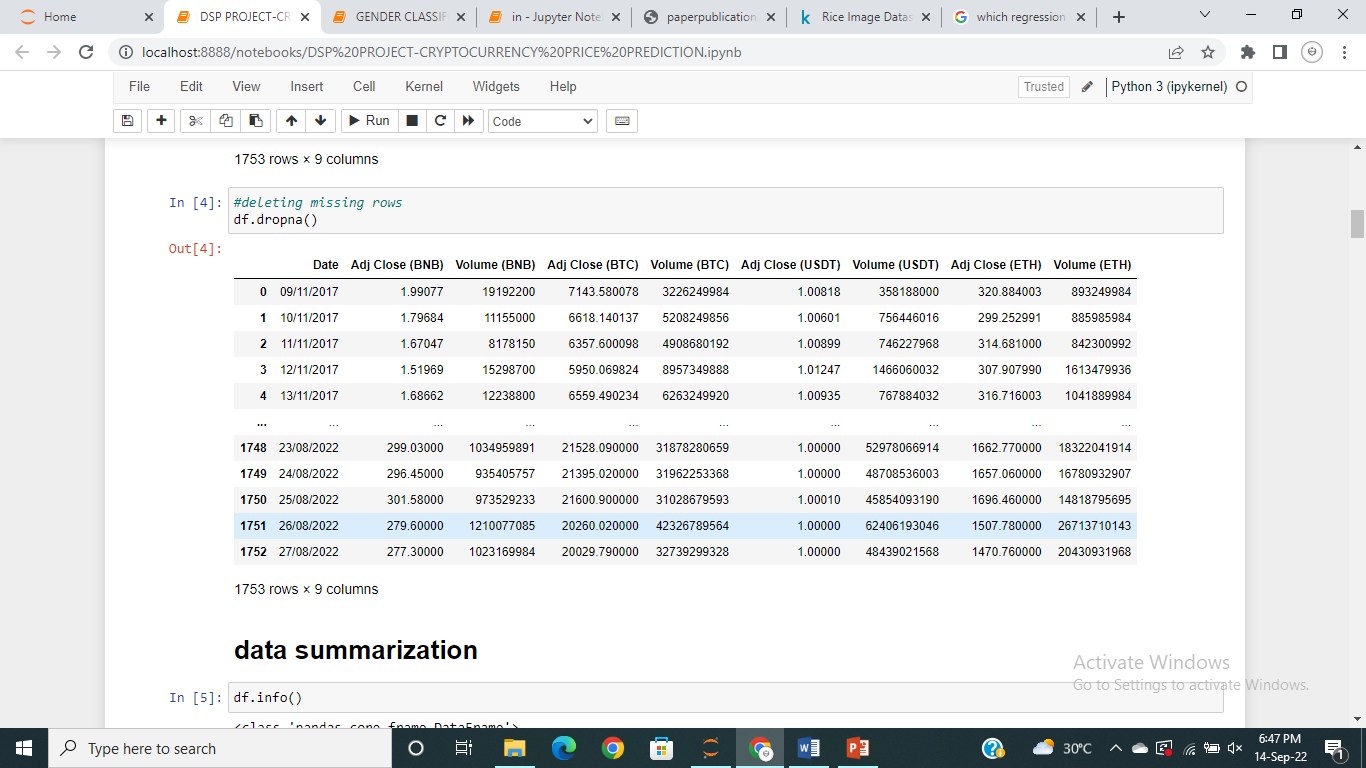
Here I have imported the dataset by specifing its path.

***3.Data Cleaning:***

Data Preprocessing is a technique that is used to convert the raw data into a clean data set.

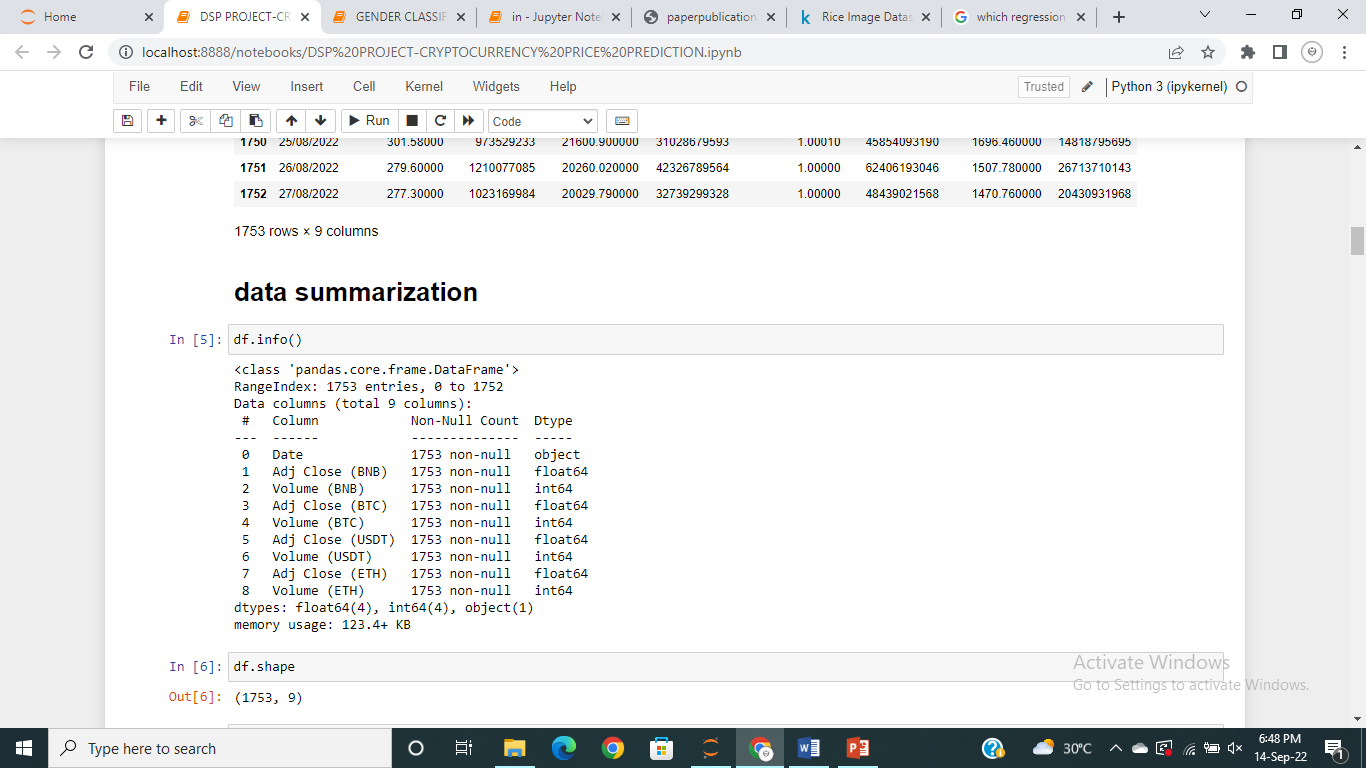


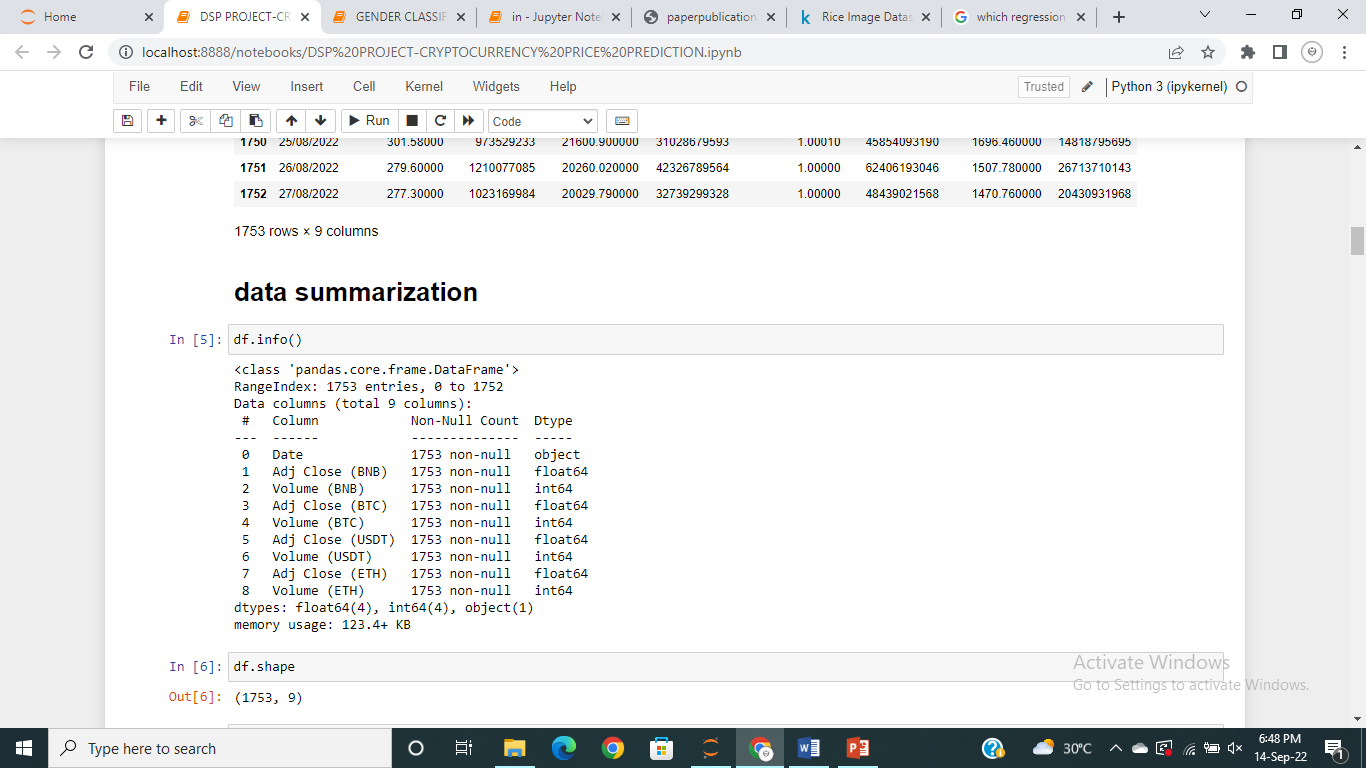
In the step of data preprocessing I have checked for the missing values and then I dropped the missing values.

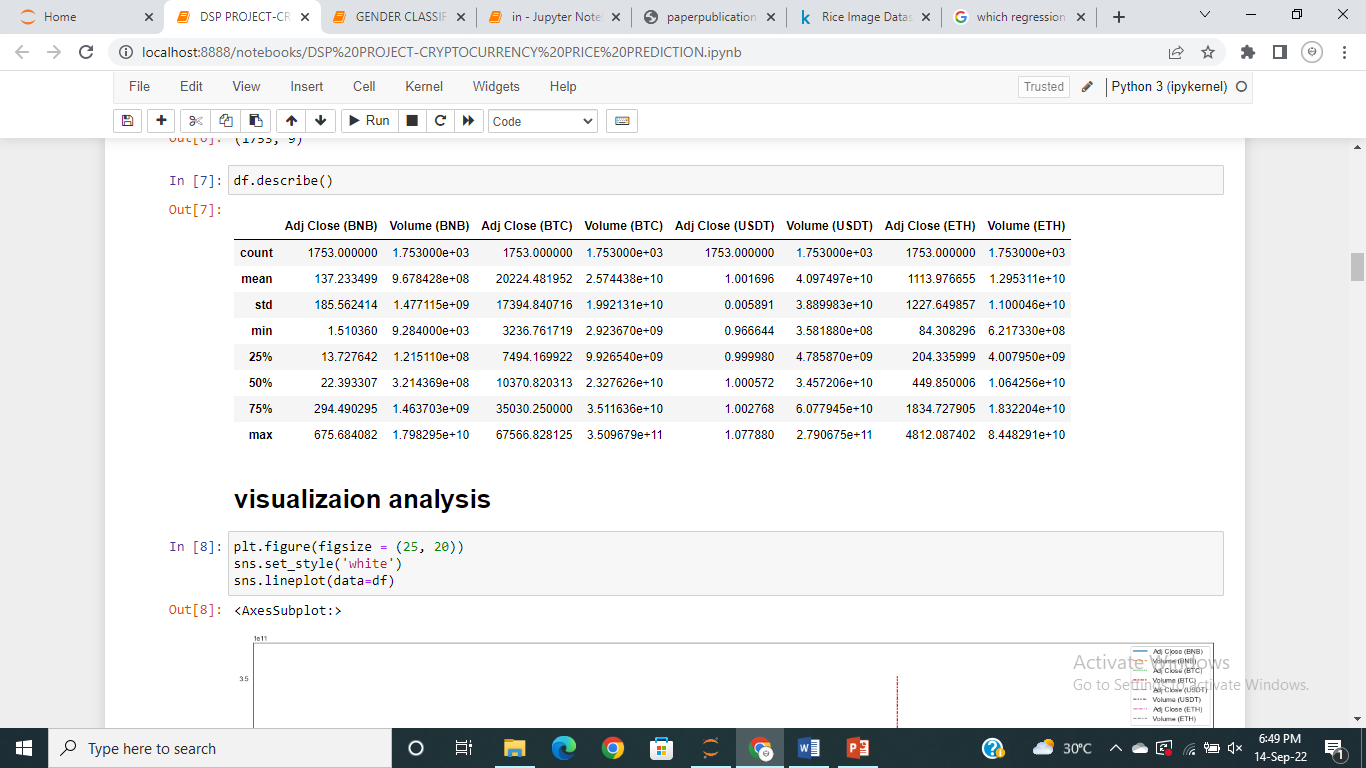


***4.Summarization of Data:***

The term Data Summarization refers to presenting the summary of generated data in an easily comprehensible and informative manner.





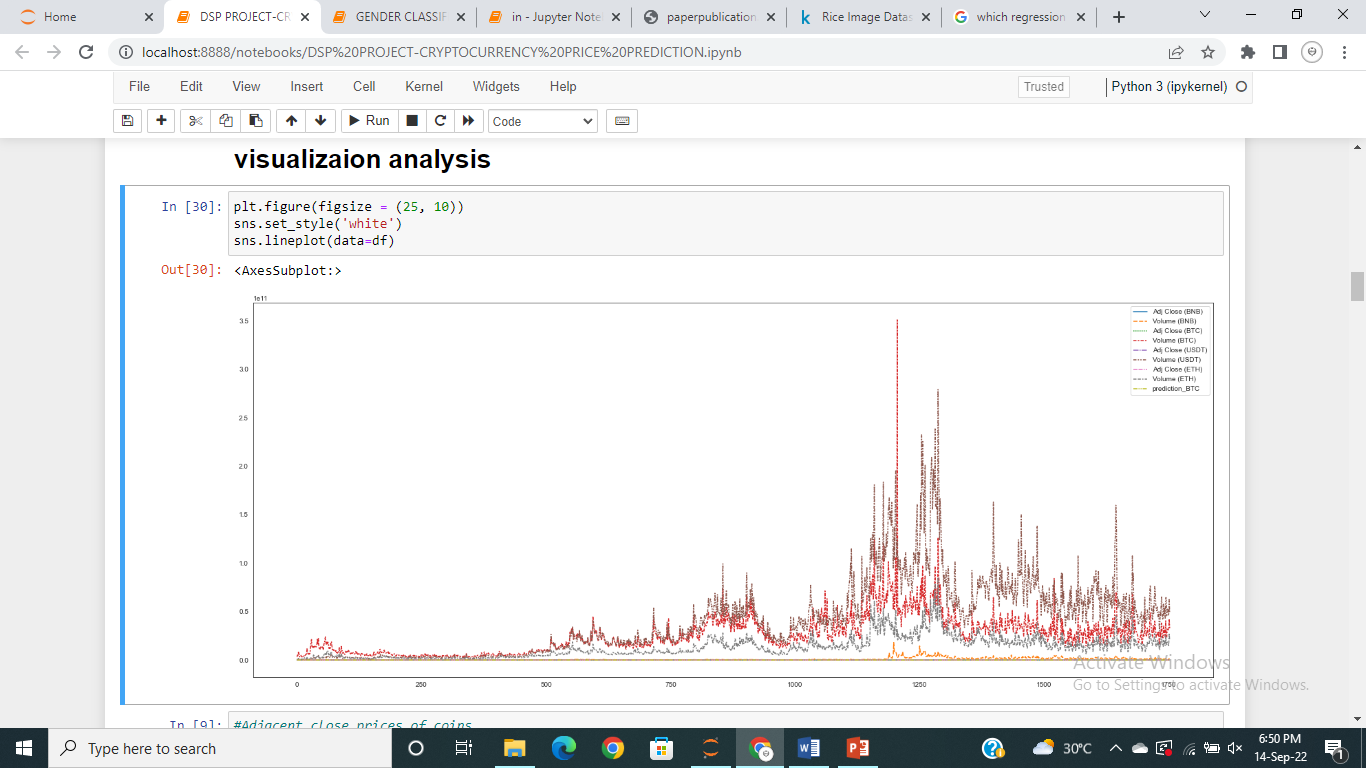


***5.Visualization Analysis:***

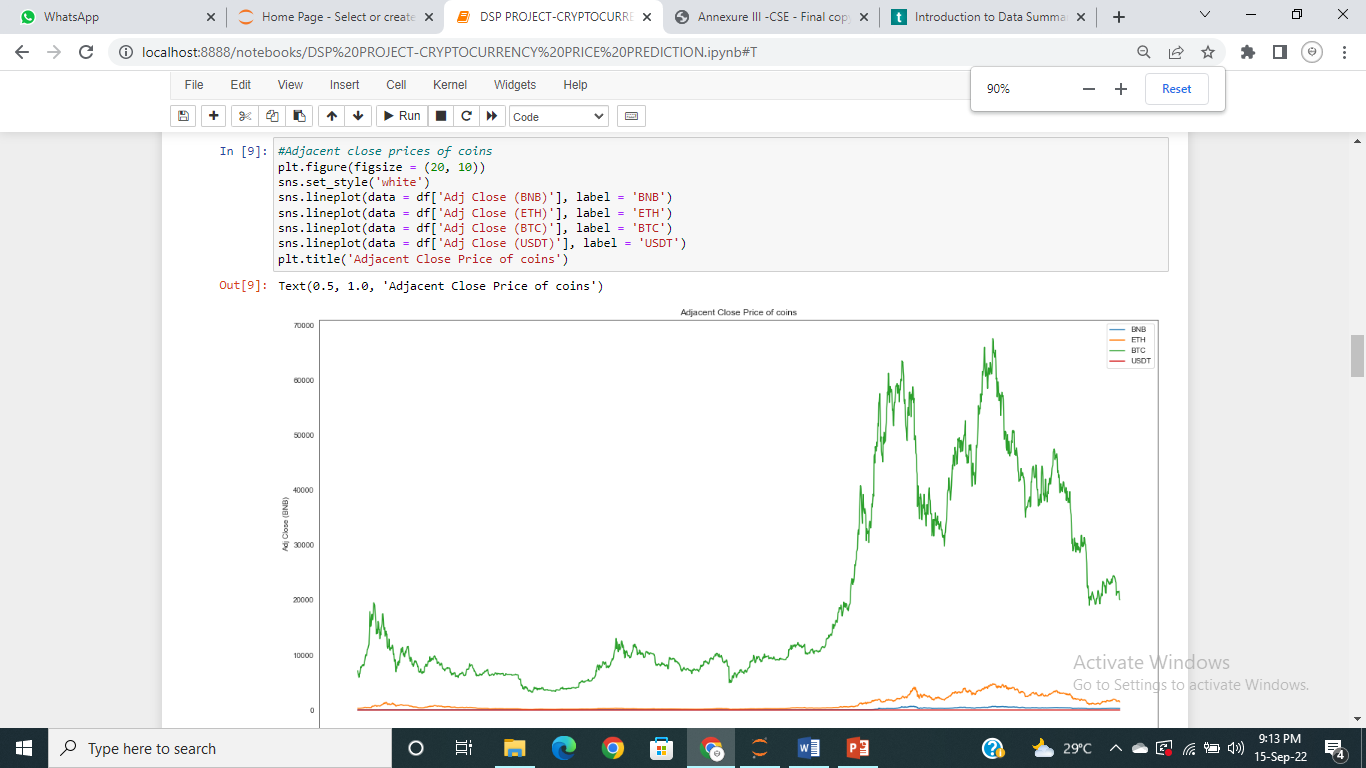
Here in this step we visualize the data of the model.

We visualize the relationship between the columns.

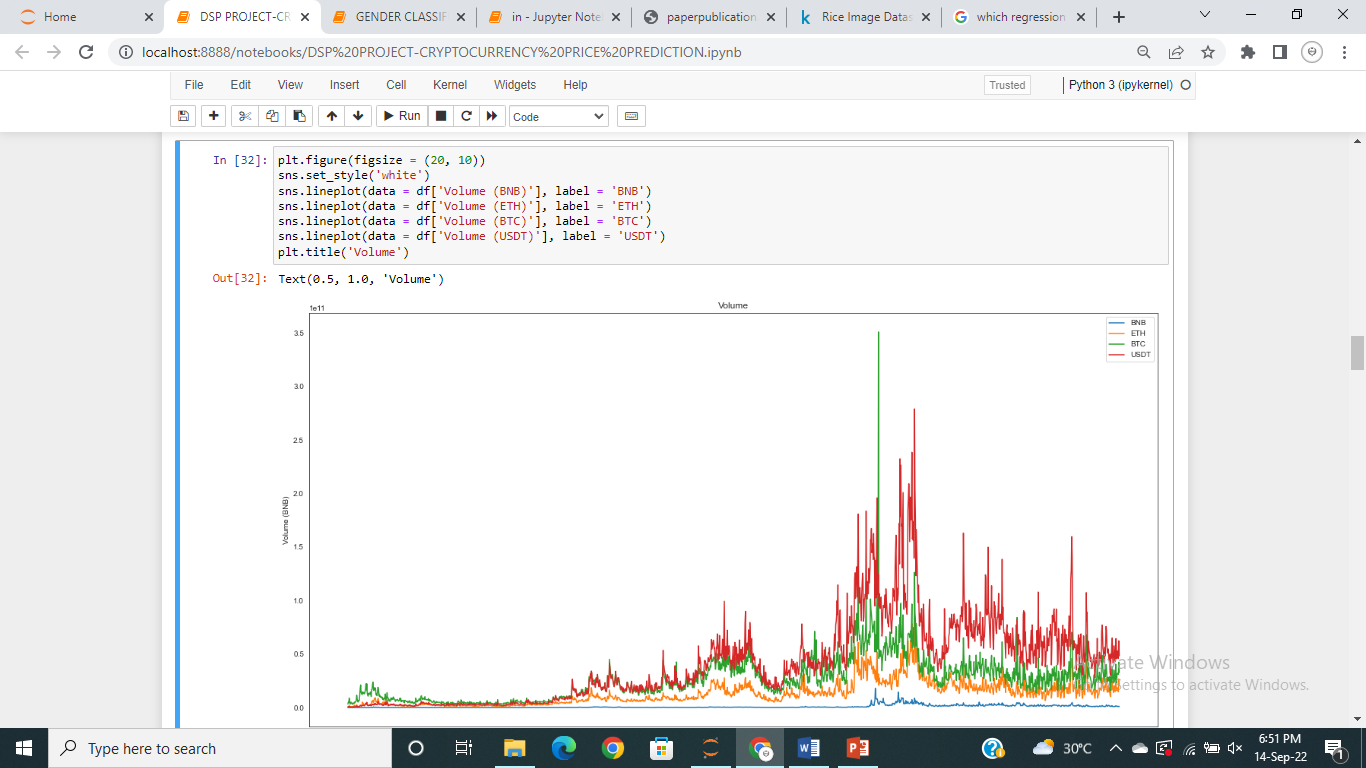
Visualization of all close prices and volumes:



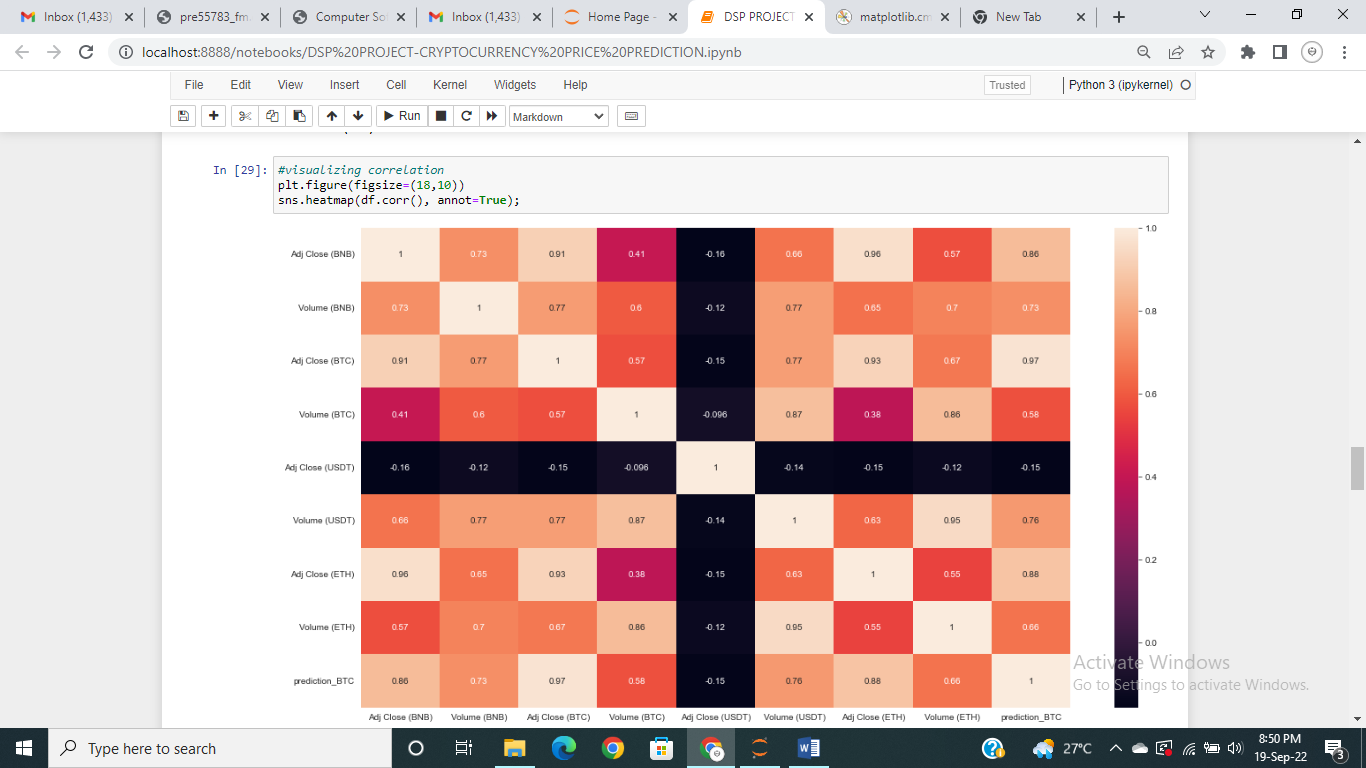
Visualization of adjacent close prices:



Visualization of volumes:



Visualization of correlation:



# *6. Choosing Independent and dependent values:*

# Here in this step split the data to predict the values.

# Here I am forecasting the Bitcoin prices for next 14 days(2 weeks).

# So, we shift the values of Adj close(BTC) values into new column named as prediction\_BTC from the index 14 because we are calculating the price of 14th day from the corresponding day .

# By shifting the values we will get nan values at last 14 rows

# 

# Now we will remove the last 14 rows to train and test the data because we can’t train the data with nan values*.*

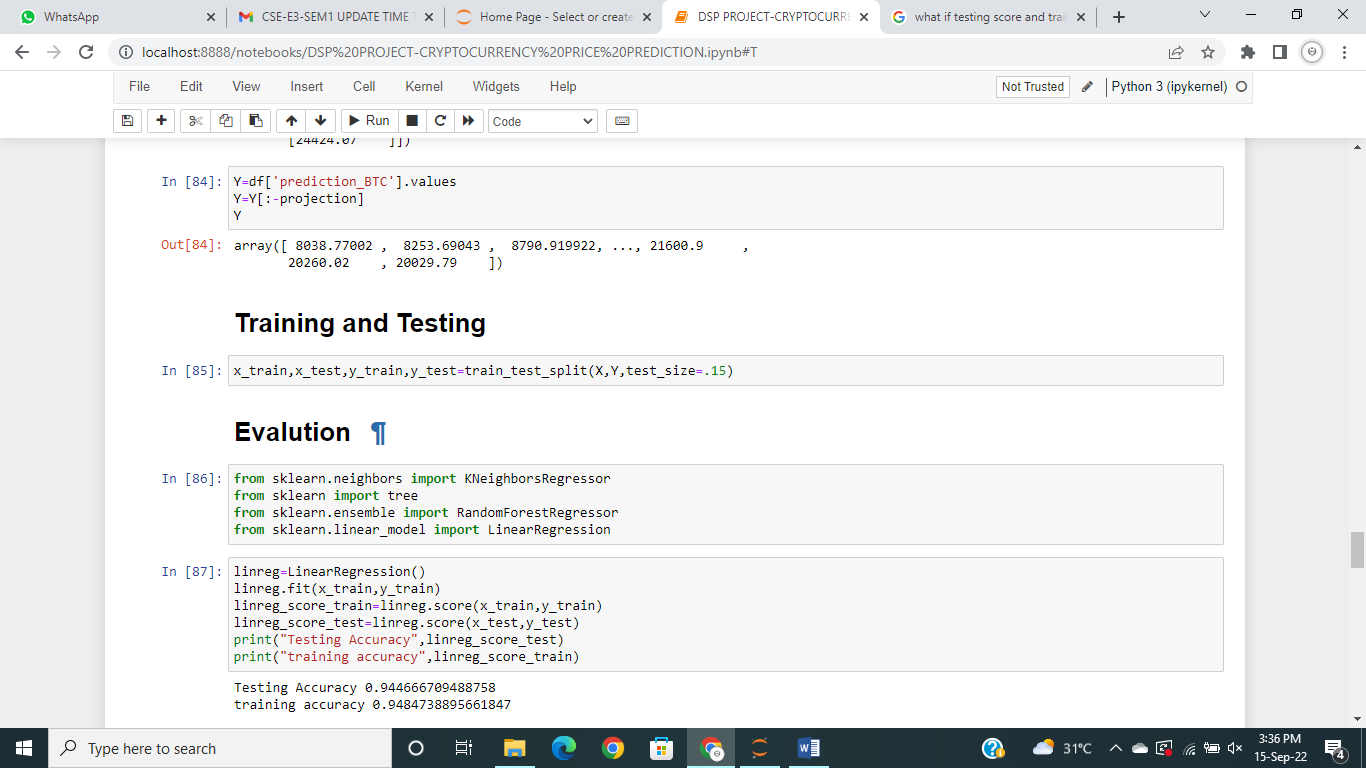
# Then we will take Adj close(BTC) and volume(BTC) into X variable as independent variable because future price depending on the past prices.

# And take Prediction\_BTC into Y variable as dependent variable.

# 

***7.Training and testing***

Now we will split the training data(85%) and testing data(15%)

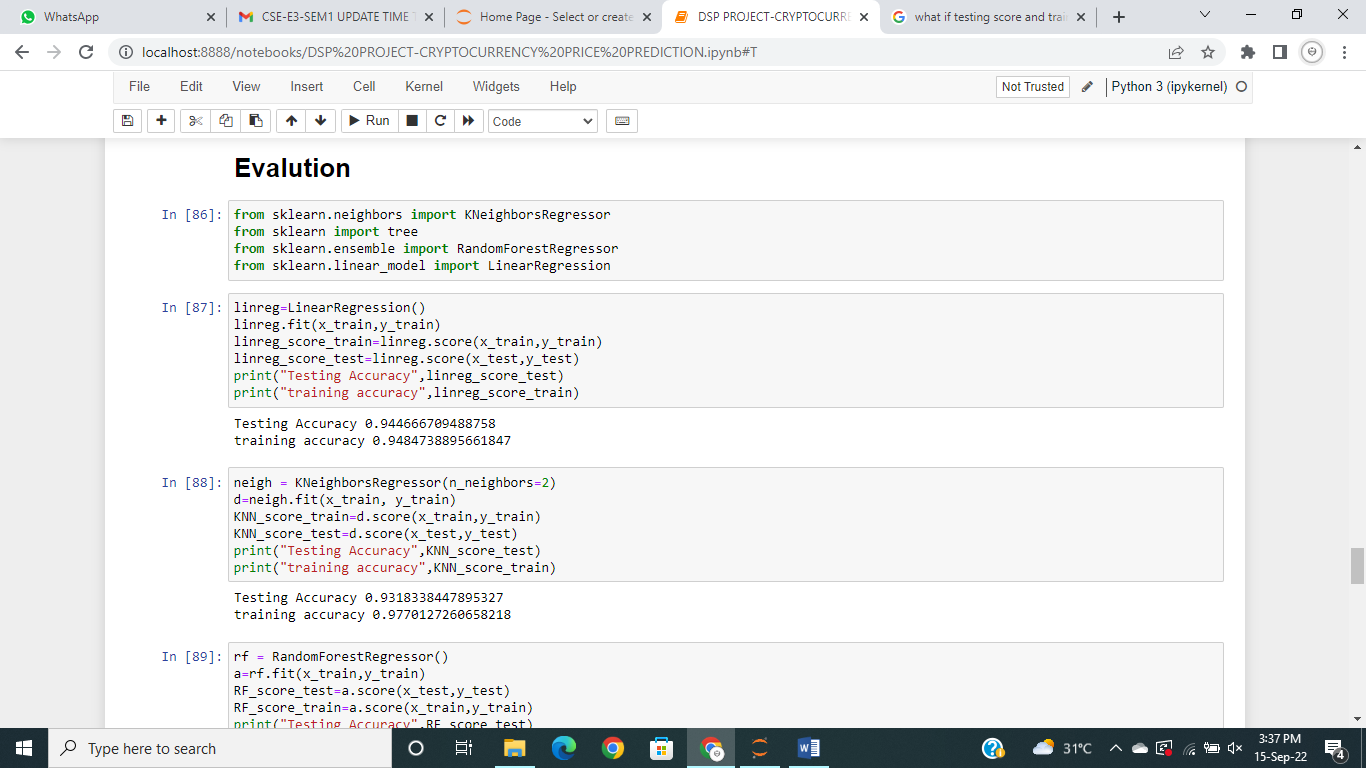


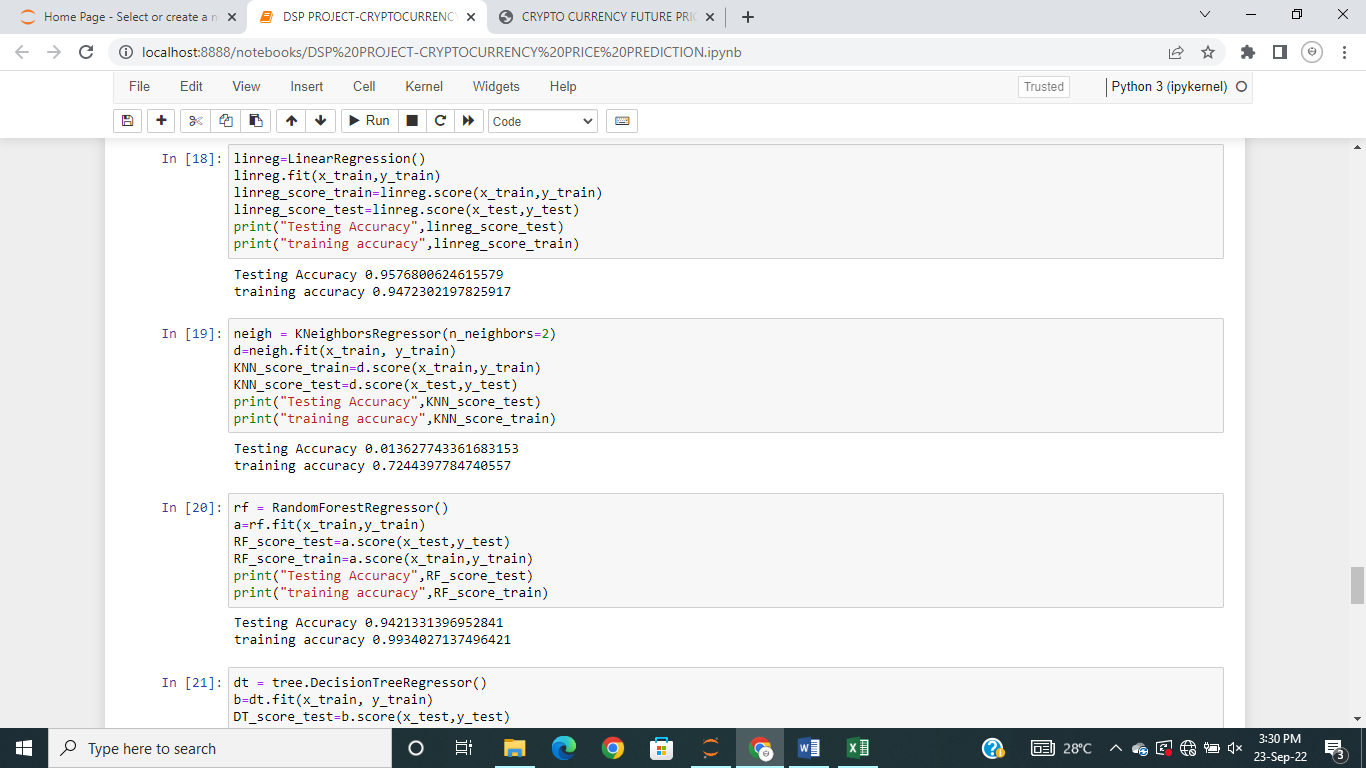
***8.Model Building***

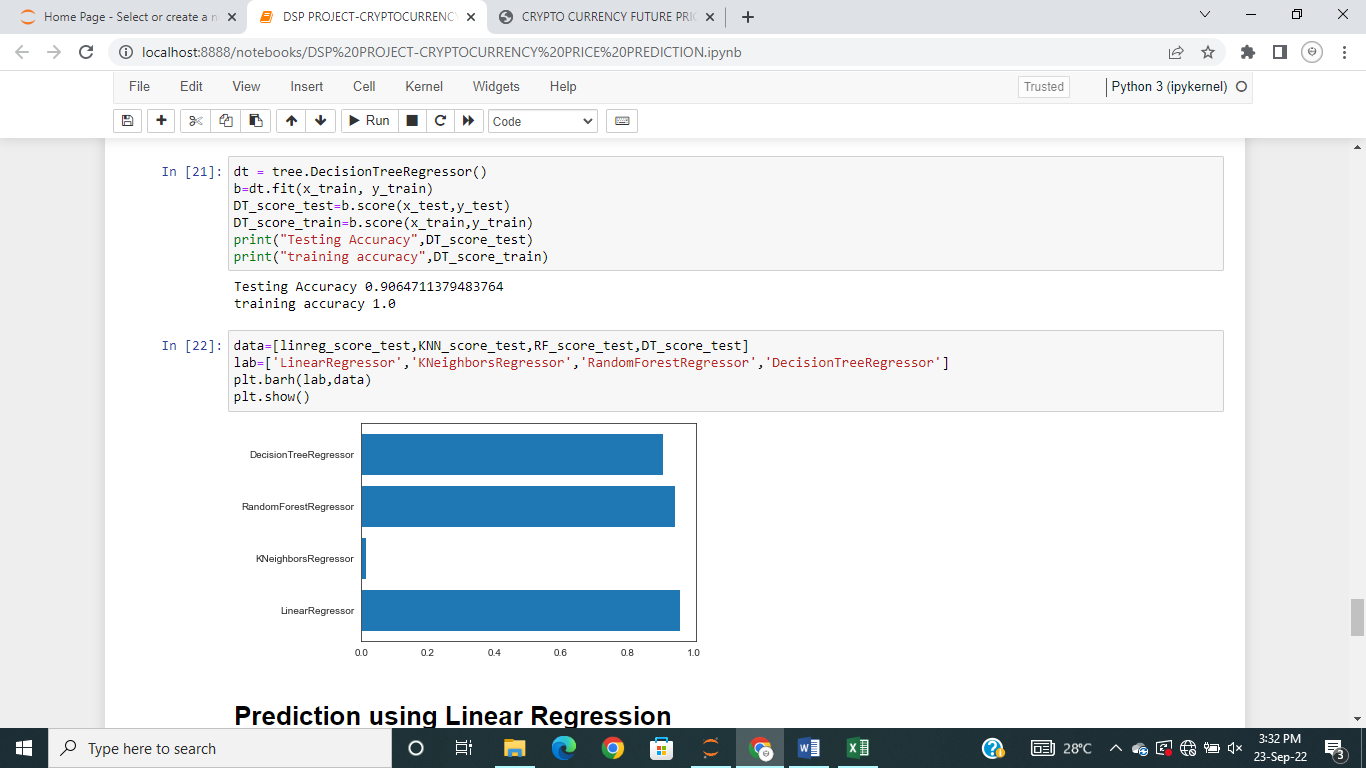
Here we will select the model based on highest testing accuracy.

I have took four algorithms for this model and we will test for accuracy of testing data***.***

* ***KNN:*** K-Nearest Neighbour is one of the simplest Machine Learning algorithms based on Supervised Learning technique. K-NN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories.
* ***DECISION TREE***: A decision tree is a [decision support](https://en.wikipedia.org/wiki/Decision_support_system) tool that uses a [tree-like](https://en.wikipedia.org/wiki/Tree_(graph_theory)) [model](https://en.wikipedia.org/wiki/Causal_model) of decisions and their possible consequences, including [chance](https://en.wikipedia.org/wiki/Probability) event outcomes, resource costs, and [utility](https://en.wikipedia.org/wiki/Utility). It is one way to display an [algorithm](https://en.wikipedia.org/wiki/Algorithm) that only contains conditional control statement.
* ***Random forest Regression:*** Random Forest has multiple decision trees as base learning models. We randomly perform row sampling and feature sampling from the dataset forming sample datasets for every model.
* ***Linear Regression:*** Linear regression analysis is used to predict the value of a variable based on the value of another variable. The variable you want to predict is called the dependent variable. The variable you are using to predict the other variable's value is called the independent variable.







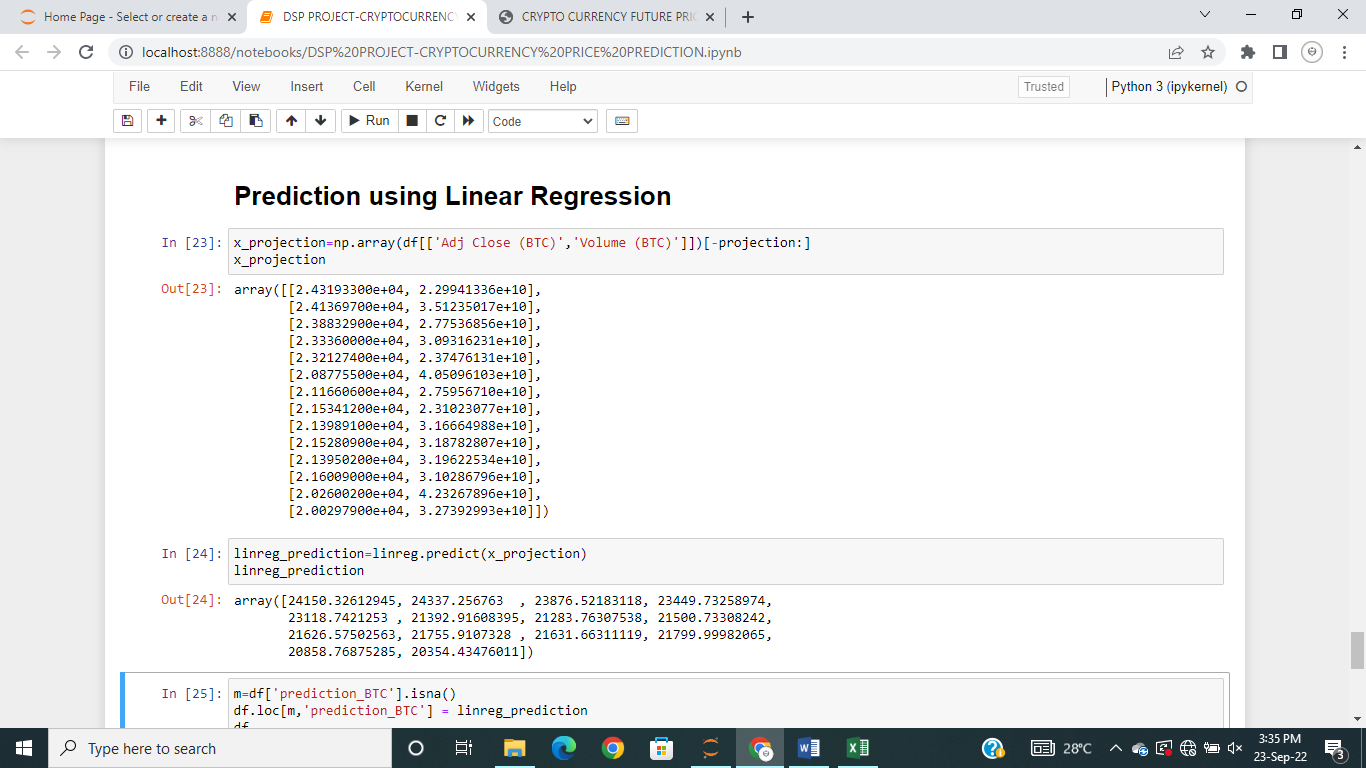
Here all have nearly same accuracy but linear regression is having high accuracy than others so we will select the linear regression model.

***9.Prediction Using Linear Regression:***

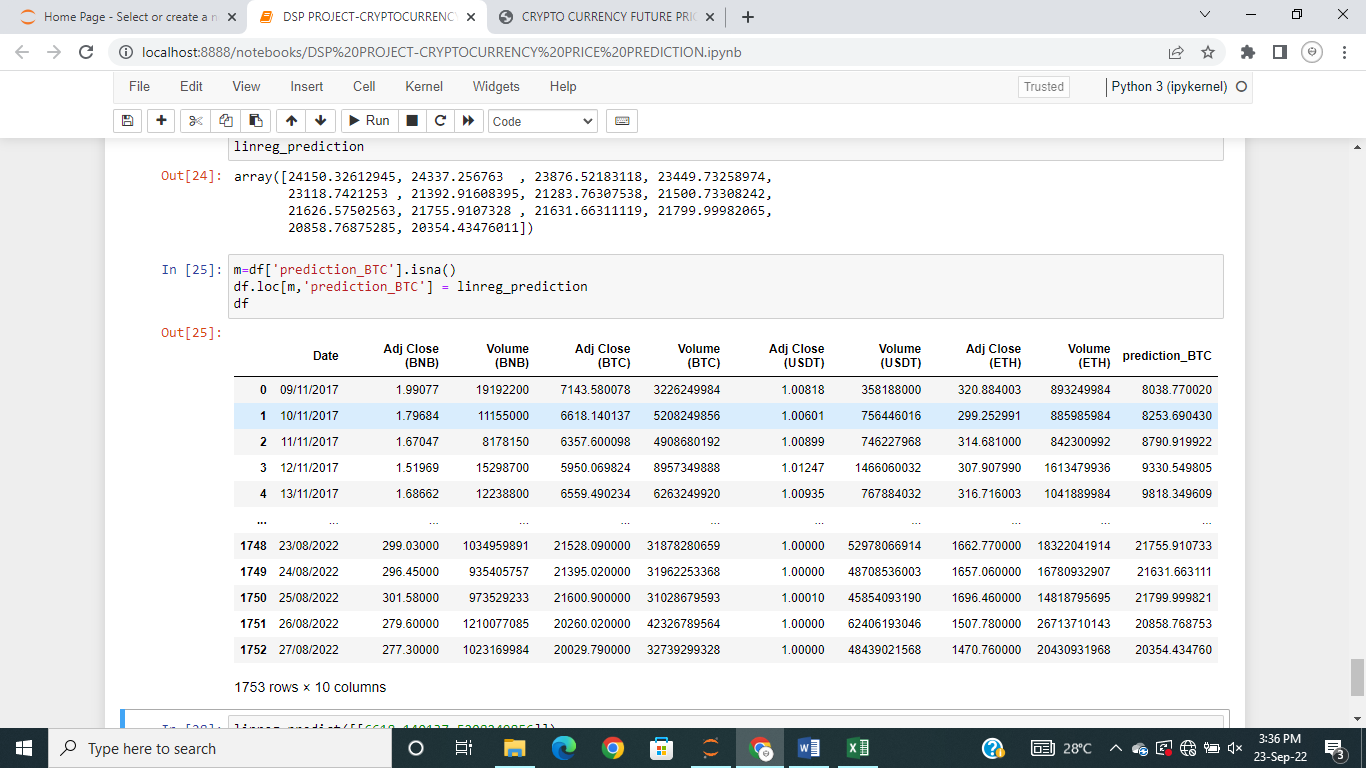
Now the prediction of future prices takes place here.

We are taking the last 14 rows of Adj close(BTC) and Volume (BTC) for which we have to predict the values in the column Prediction\_BTC.

Then we will predict the values using predict() function.



Now we will fill the predicted values into our dataset by using following code



***Conclusion:***

This is the systematic process used to build the model starting from the importing dataset to predicting values.This model can be used to forecast the future prices of other cryptocurrencies also.

Prepared by

THANNIRU YASHASVINI

S180493

CSE-2C

***GITHUB:*** [***https://github.com/YASHASVINI-THANNIRU/DSP-PROJECT.git***](https://github.com/YASHASVINI-THANNIRU/DSP-PROJECT.git)