Final Report: Cryptocurrency Price Prediction using Time Series Forecasting

Mining of crypto currencies began in 2009, with Bitcoin being the first to be mined using a software through which new Bitcoins are created and transactions are recorded and verified on the blockchain. Eventually trading of bitcoins in exchange for money begins and with the added popularity of the idea of a decentralized and encrypted currency led to the emergence of alternative cryptocurrencies. These new currencies try improving on the initial design of Bitcoin offering added benefits in speed or anonymity or something else.

By 2017, the value of a Bitcoin had peaked at $19,475 and it was evident that more and more users are engaging and more money is being invested into the cryptocoin ecosystem. Even banks like Barclays and CitiBank have said that they are looking into ways to work with cryptocurrencies.

The following report presents my findings in investigating a time series analysis in the prices of most popularly and highest valued cryptocurrencies and finally in developing a Autoregressive Integrated Moving Average(ARIMA) model to predict the future prices of these coins based both on its historical data and also how the value of different coins impact the prices of a certain cryptocoin.

These finding and model would be useful to the aforementioned users who trade and invest in crypto currencies and also for banks looking for methods to be involved in cryptocurrency mining and trading

Describing the dataset and how it was wrangled

The data set was obtained from kaggle – ‘ Every Cryptocurrency daily market price’(https://www.kaggle.com/jessevent/all-crypto-currencies).

After downloading and reading the CSV file on python, and running a preliminary exploratory analysis it is inferred that the initial data set consisted 785020 entries of 1643 unique crypto currencies. Each entry has 13 variables, the slug, symbol and name of the crypto currency are the first three variables. The next two columns recorded the date that each observation was recorded, the rank the coin held during that date respectively. The next few columns are as follows, the opening price in dollars, the highest price in dollars the crypto currency peaked at during that day, the lowest price in dollars the crypto currency dropped to during that day and the closing price in dollars. Then the total volume of coins traded that day and the market cap are recorded and finally the last two column record the close ratio and the spread.

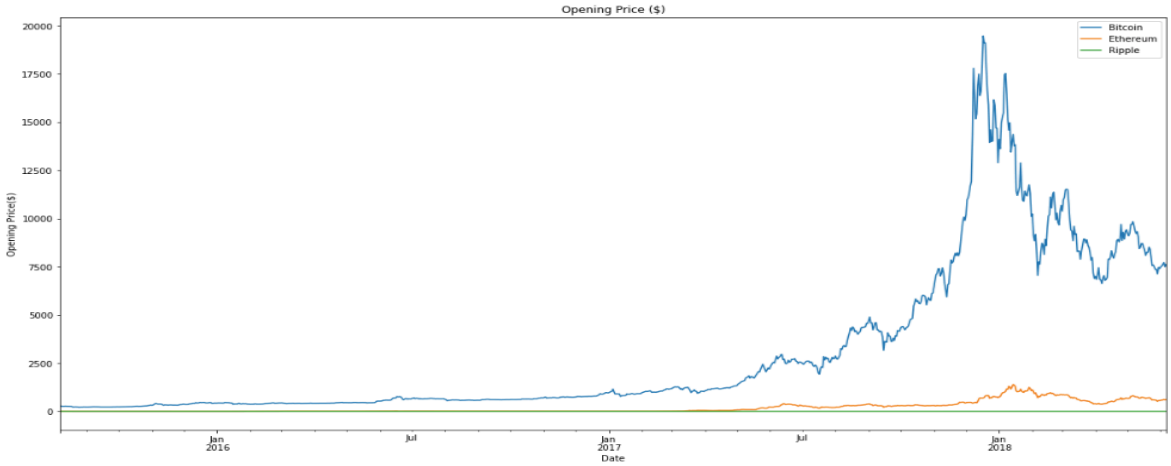
For this analysis only the top three ranked crypto currencies will be utilized, Bitcoin(BTC), Ethereum(ETH) and Ripple(RPX). The observations for these crypto currencies have been sliced out of the original data frame and put into their own data frames respectively. The date column is then used as the index for the data frames and it is sorted in an ascending order.

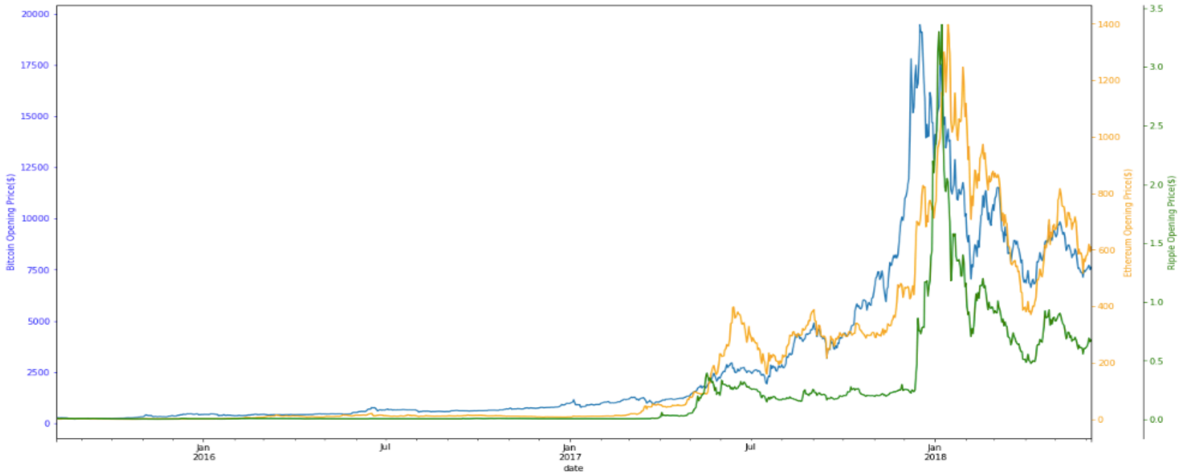
An additional column called ‘average’ is created as that records the average of the opening, high, low and closing prices per day for all three coins as like stock market prices the average price may be important and the most representative of the value of the coins for some users.

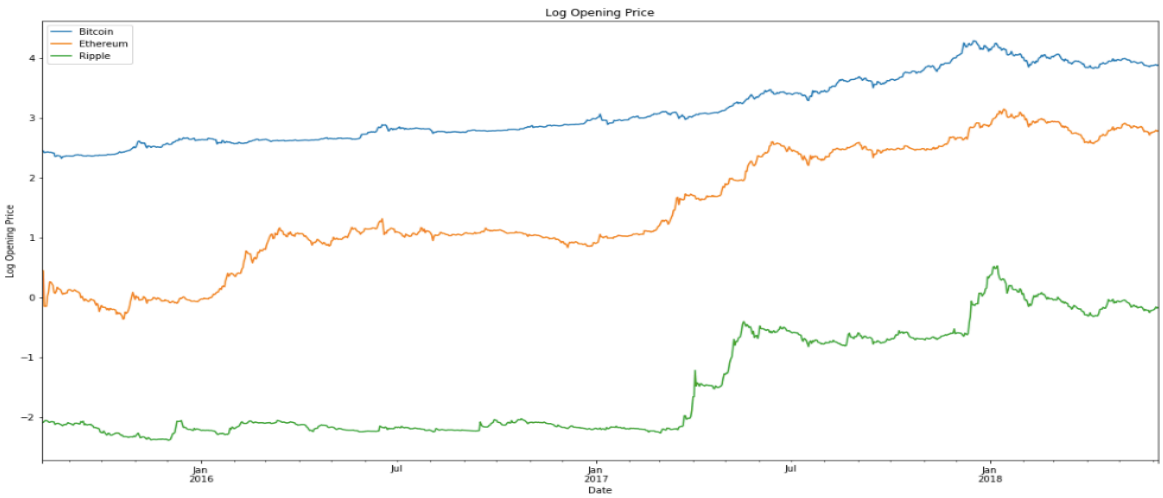
Exploratory Data Analysis

A series of plots are made as shown below, for the opening, high, low, closing and average prices of the cryptocurrencies recorded from August 8th 2015 to June 2nd 2018. For each price three kinds of plots are drawn, the first is a time series plot of the respective price and the currency in that time period. Second a time series plot with different y axes scaled to account for the substantial difference in prices between the three coins. Thirdly a time series plot of the log of the prices of the three currencies in the same time frame.

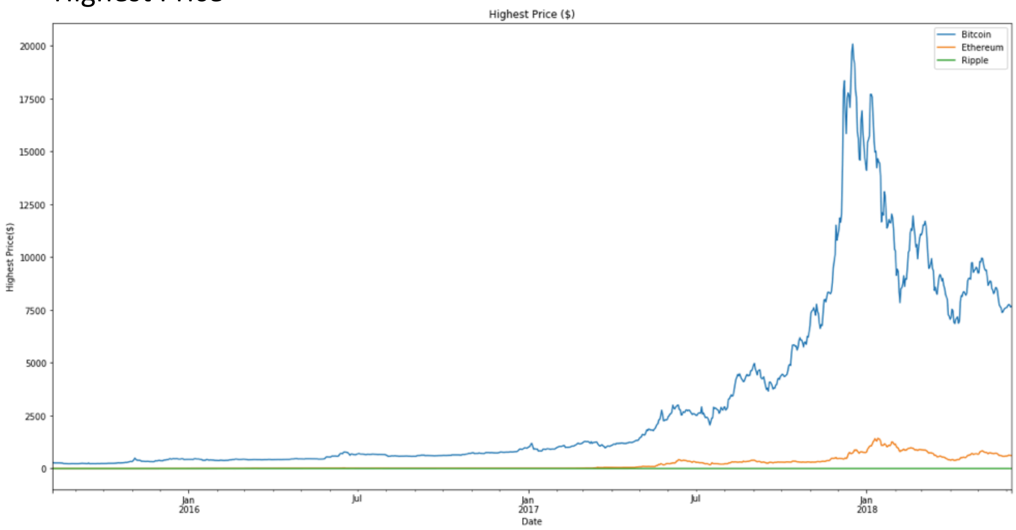
i. Opening Price

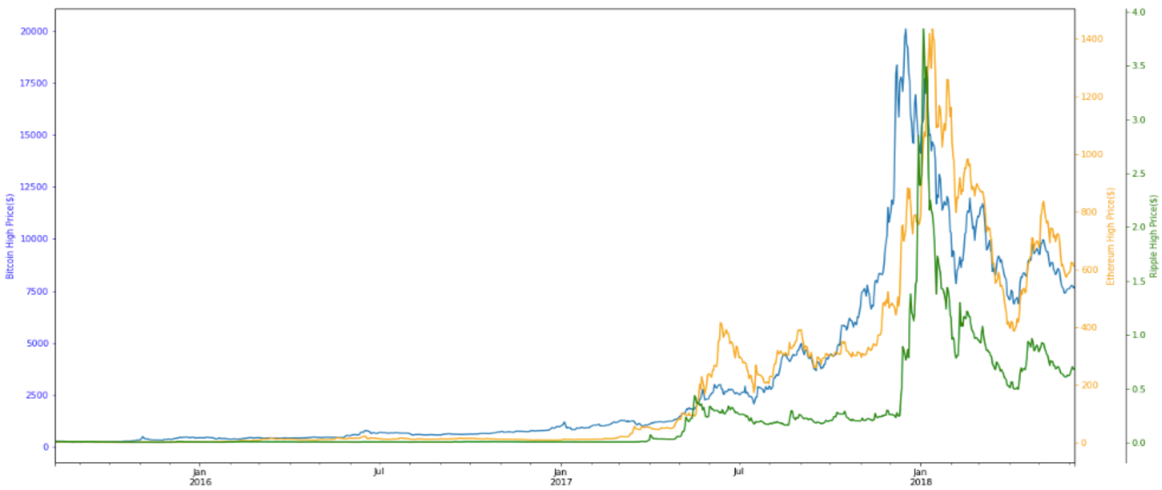


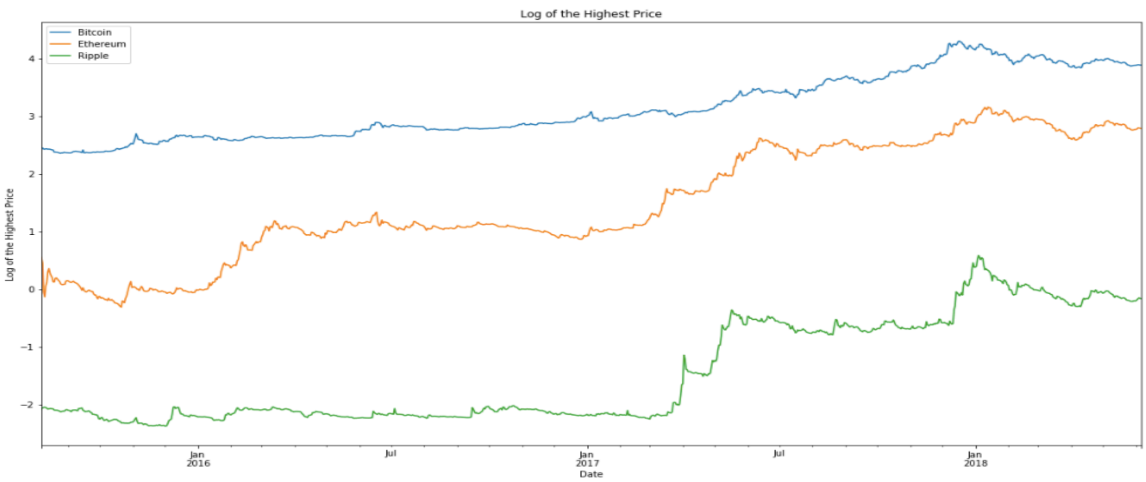




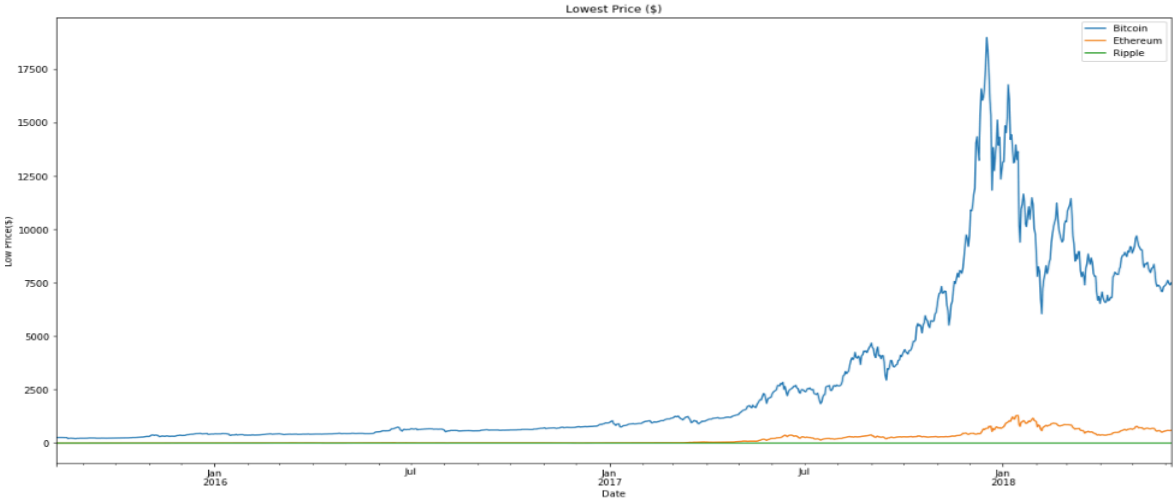
ii. Highest Price

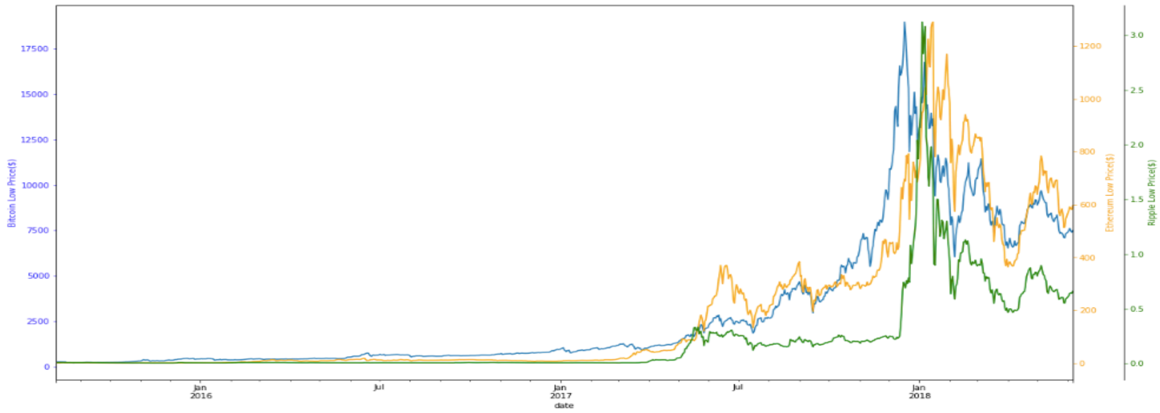


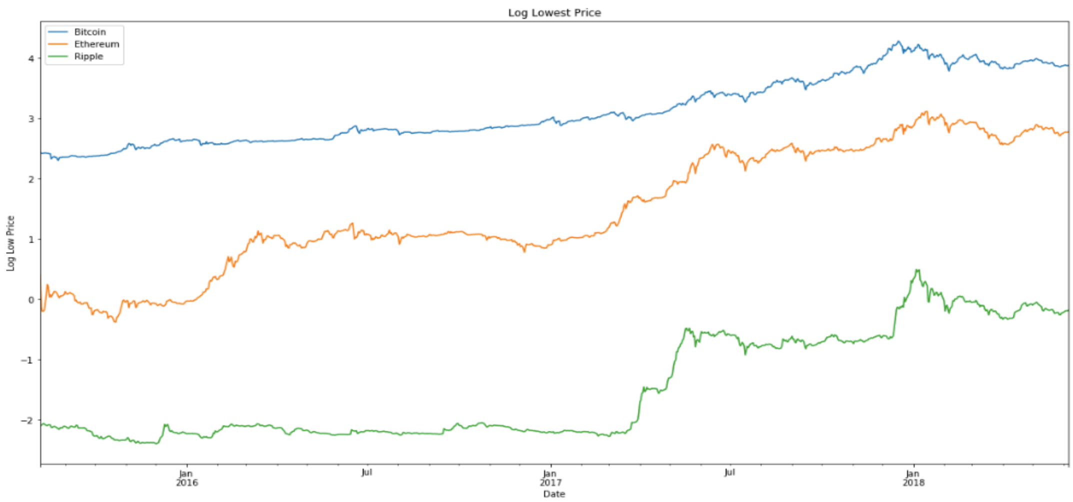




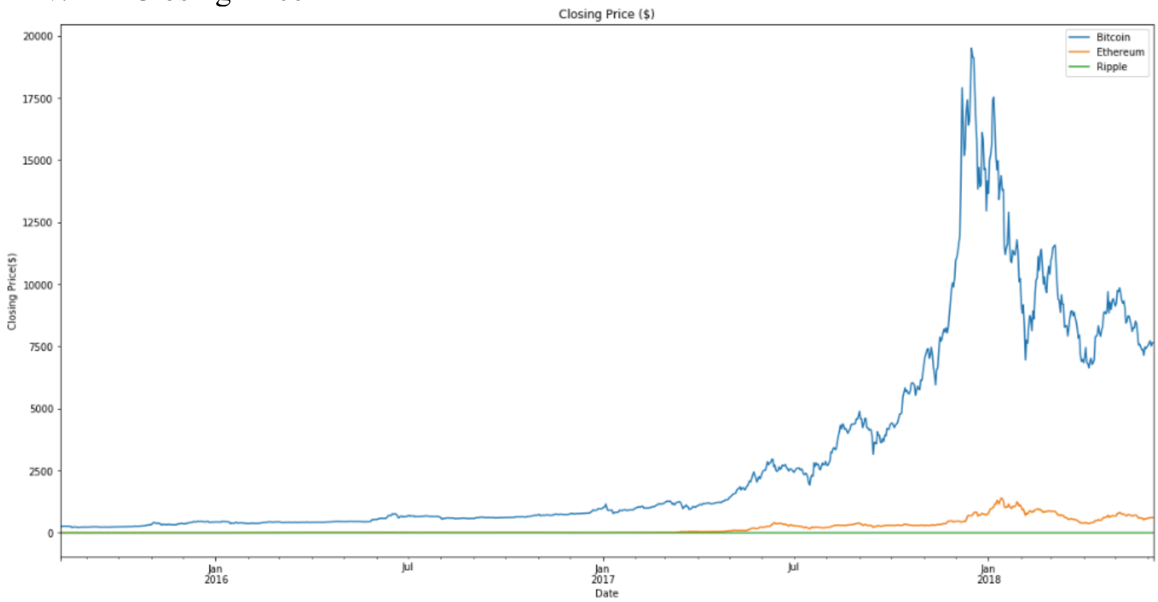
iii. Lowest Price

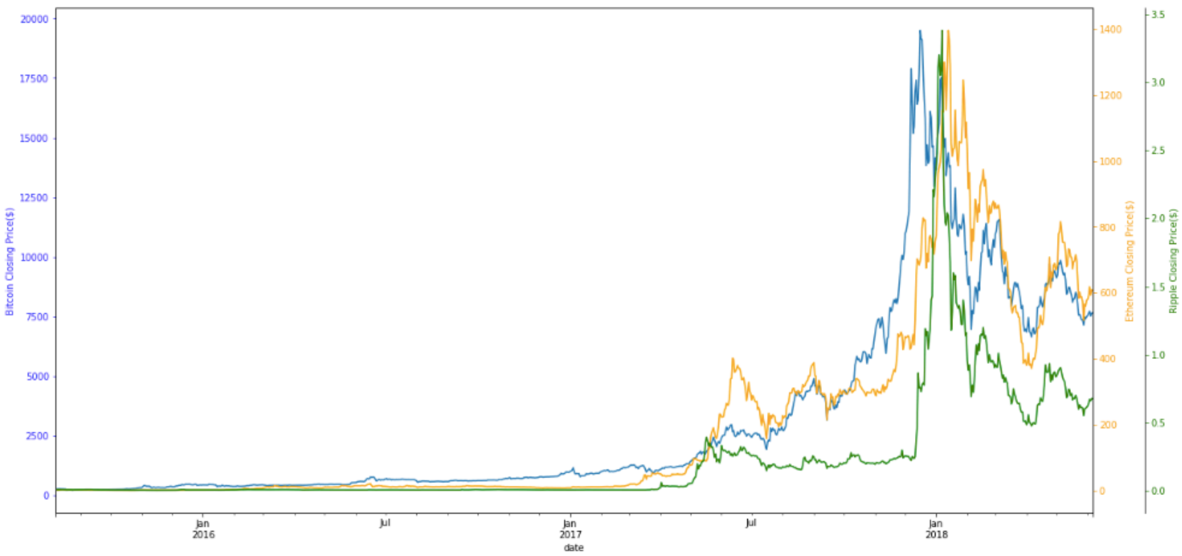


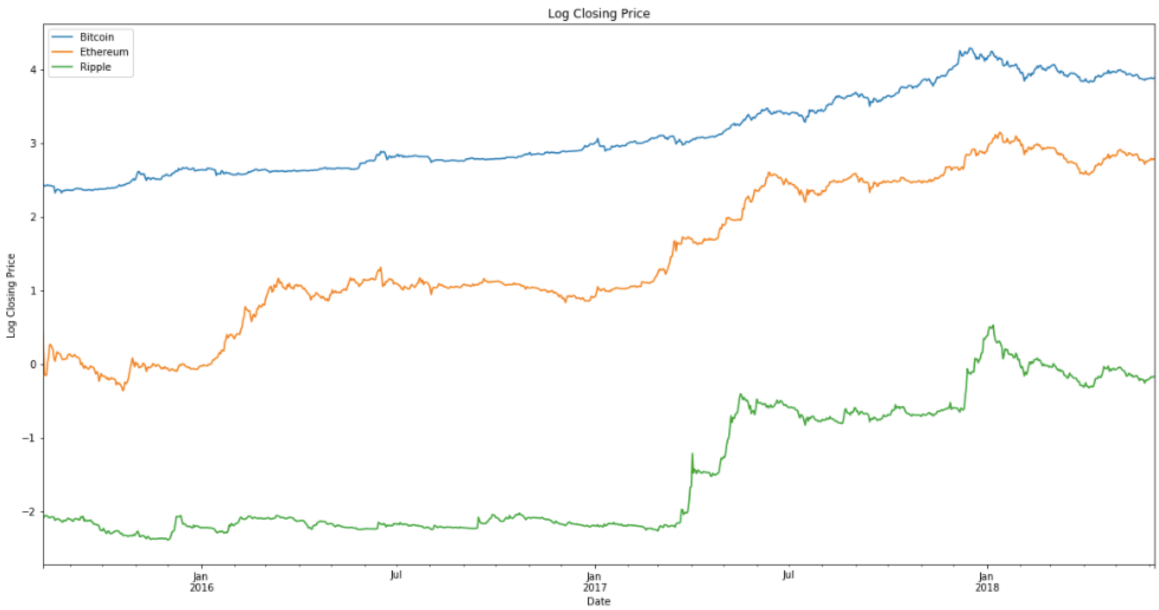




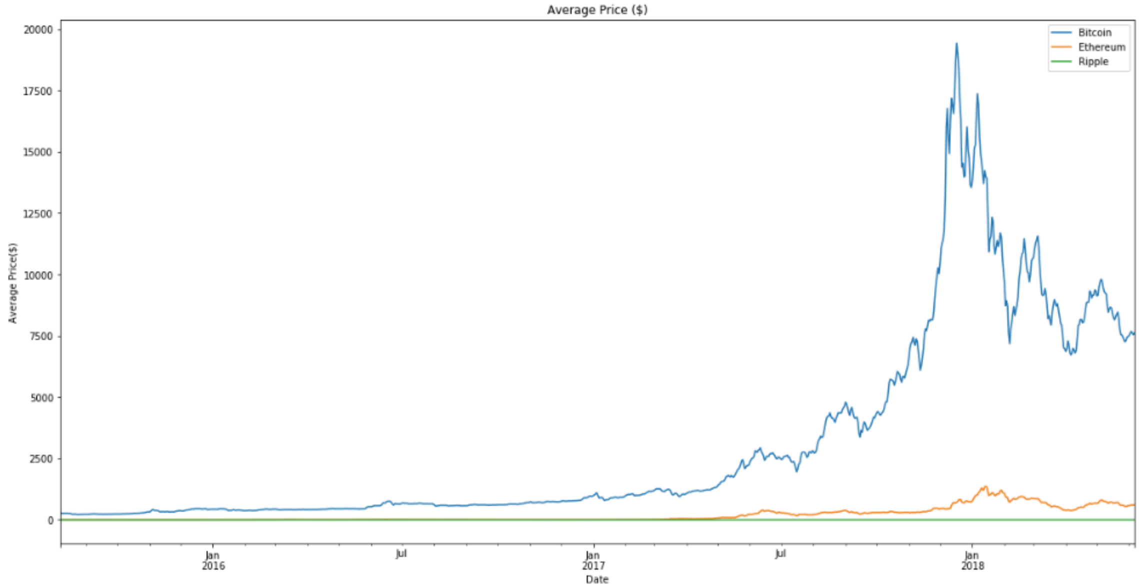
iv.	Closing Price

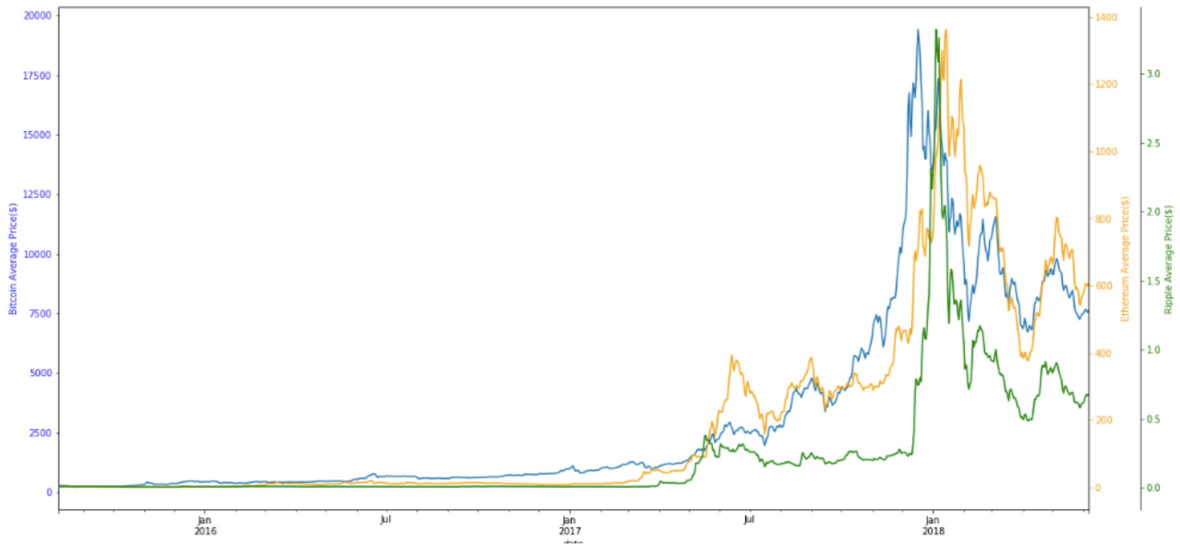


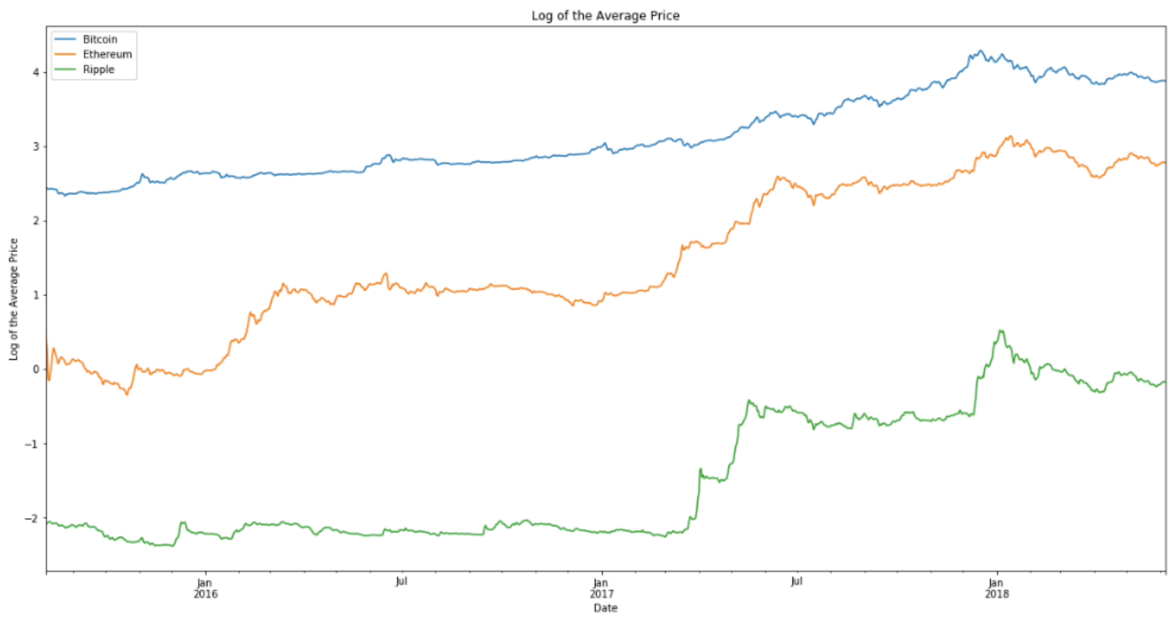




v. Average Price







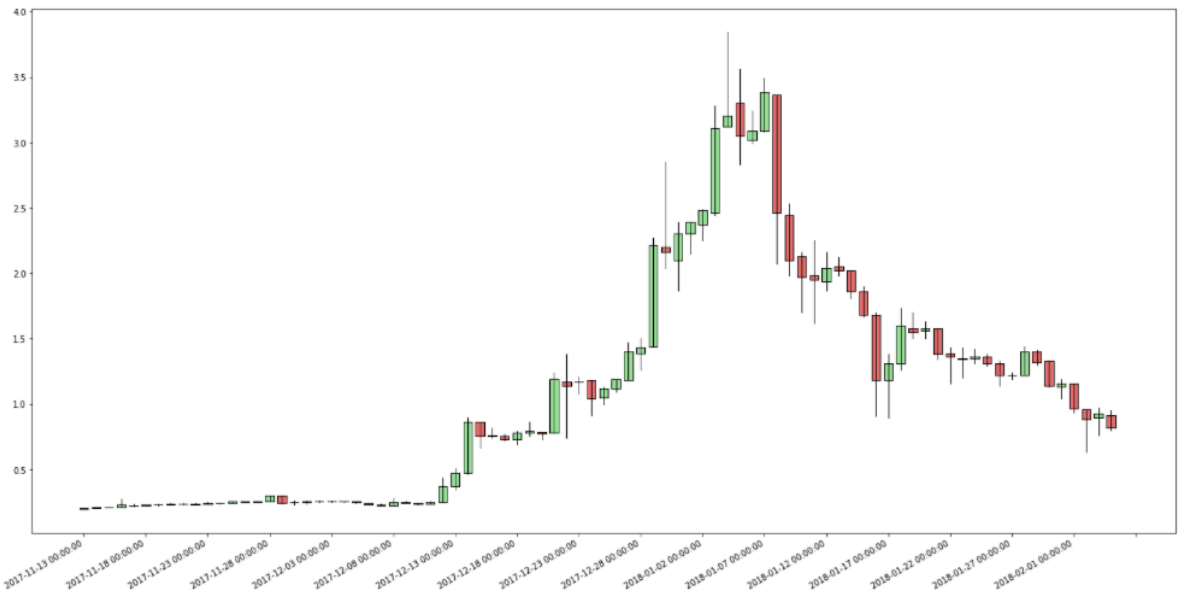
From the graphs above it is noted that bitcoin has always a higher value than ethereum and ripple. It is also noted that the rise and drop in prices of the currencies all happen at the same time.

Plotted below are candle stick plots of the prices of the three coins.

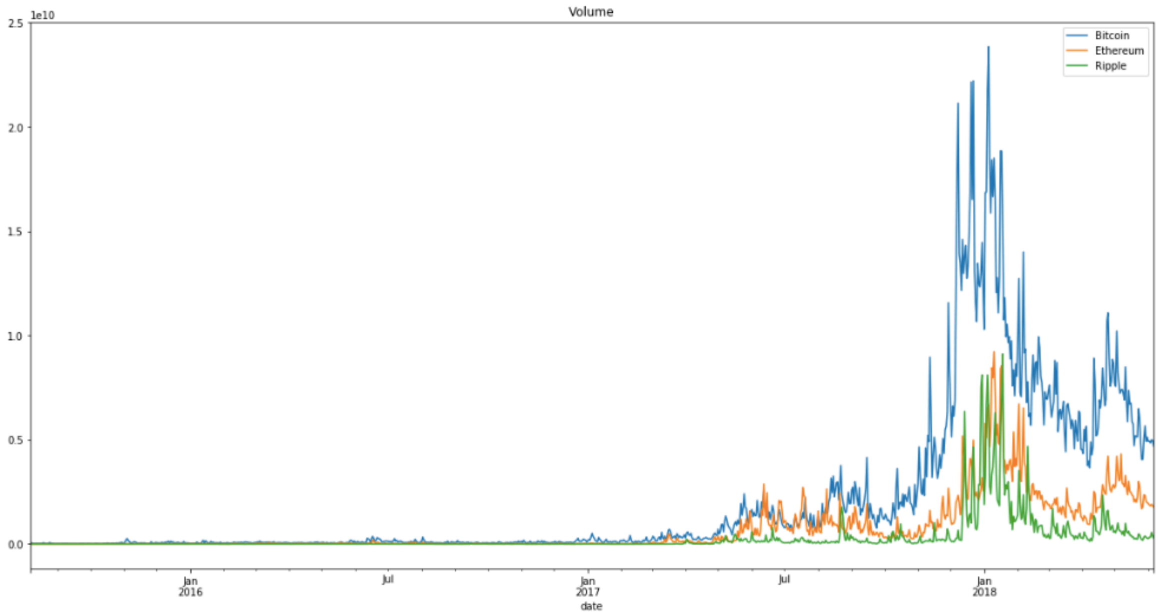
i. Bitcoin Price ($)

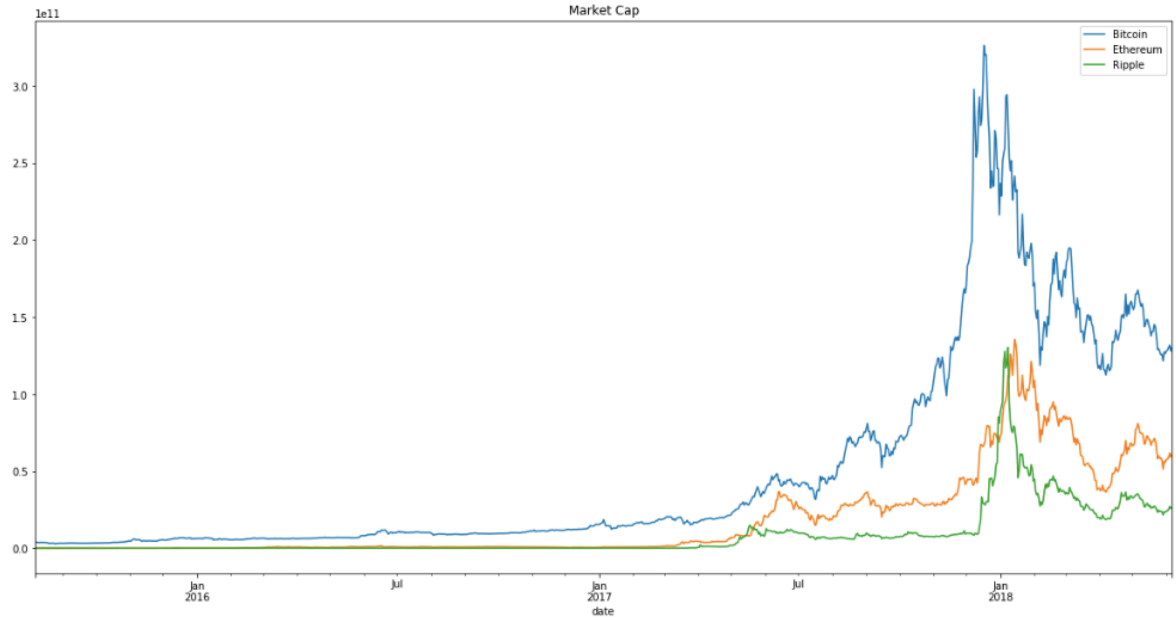
ii. Ethereum Price ($)

iii. Ripple Price ($)



Further plotted below were the Market Cap and Volume of coins being traded and it is inferred that as the prices of the coins rose or fell so did the market cap and volume of coins being traded.

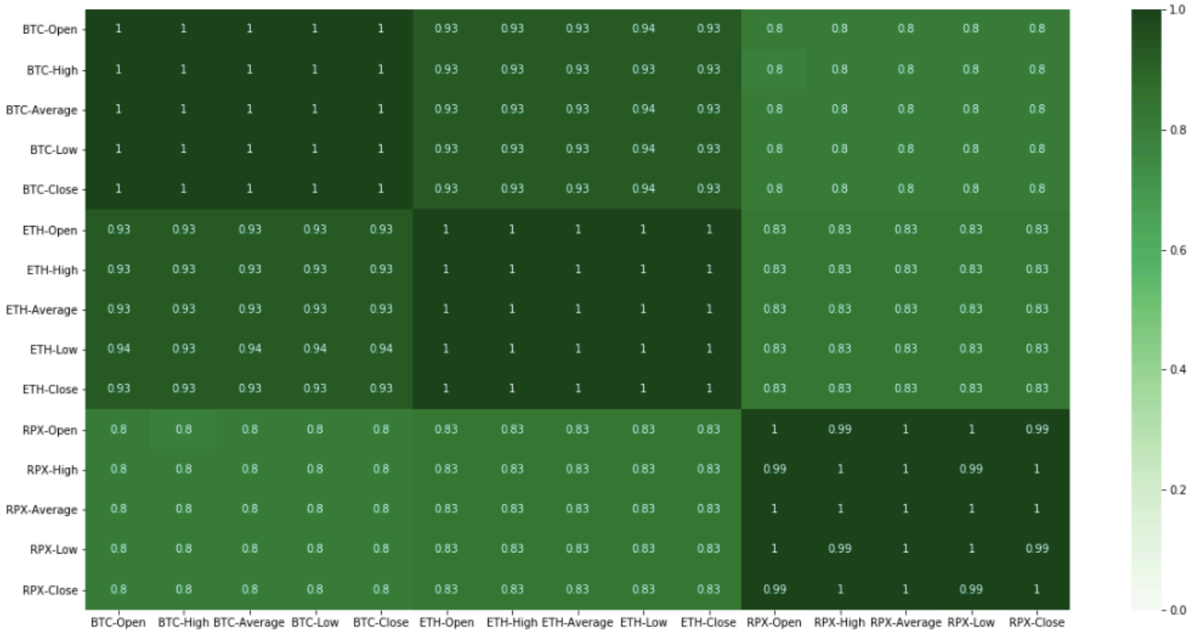




From all the plots above very similar conclusions can been drawn about the trends of the prices of crypto coins. Firstly, the value of the coins all increase and decrease at around similar times. This shows that there may be relation between how the change in the trends of one type of coin affects the other. Secondly it looks that even the rate at which the value of the coins drop are similar. The changing trends in the prices also directly affect the volume of coins being traded and market cap.

Statistical Inference

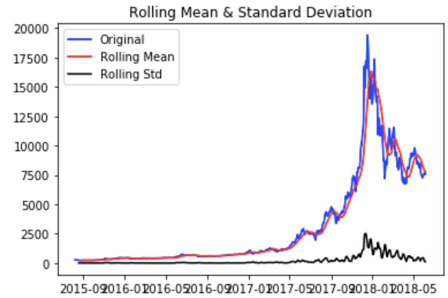
First a spearman correlation heat map is plotted between different values of all the three coins.



Then the relation between their average prices were checked using a Granger Causality test to check if the price of one coin affected the other. The results obtained all had p vales less than 0.05 except for the test checking weather change in the ethereum average price affects the average price of ripple (p = 0.29). This result was used to conclude that change in the value of Bitcoin affected the prices of Ethereum and Ripple.

Machine Learning: ARIMA

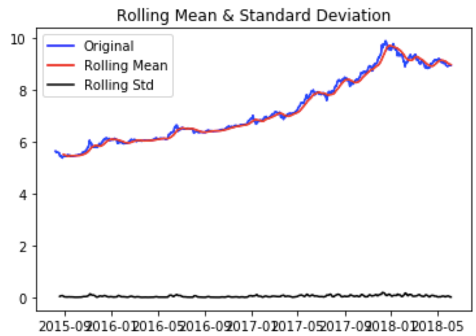
First the data is checked if it is stationary using the Dicky Fuller Test. The Null Hypothesis of the test is not stationary.




```
ADF Stastistic: -1.541109
p-value: 0.513086
The graph is non stationery
Critical values:
        1%: -3.437
        5%: -2.864
        10%: -2.568
```

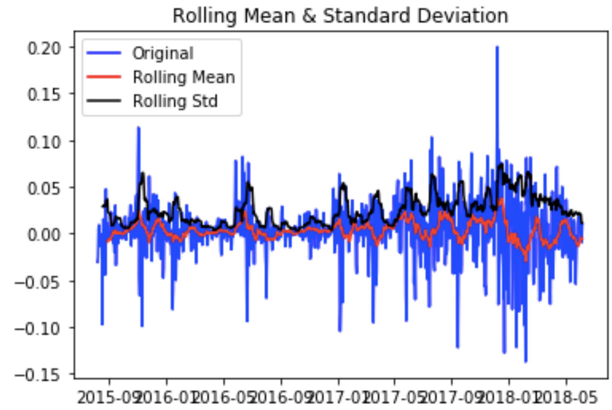
Since the p value is greater than 0.05 the null hypothesis is accepted and the data is non stationary hence it must be transformed.

The data is log transformed to unskew the data and the same test is conducted.




```
ADF Stastistic: -0.464491
p-value: 0.898762
The graph is non stationery
Critical values:
        1%: -3.437
        5%: -2.864
        10%: -2.568
```

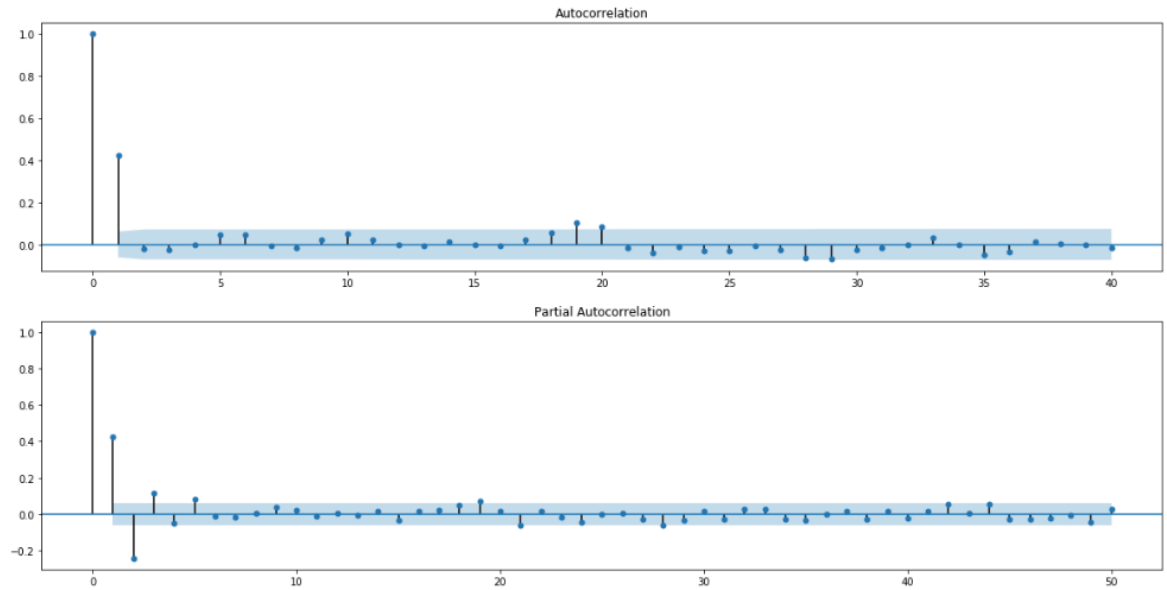
Since the p value is still greater than 0.05 the data is still non stationary and must be further transformed to remove trend and seasonality. This was done by differencing the current value with the previous value so that the mean is stabilized and the chances of the data being stationary is increased.




```
ADF Stastistic: -12.649901
p-value: 0.000000
The graph is stationery
Critical values:
        1%: -3.437
        5%: -2.864
        10%: -2.568
```

Since the p value is less than 0 the data is stationary the time forecast model can be used.

The ARIMA model has 3 parameters that can be tuned in order to achieve the right model. This is done by plotting a ACF and PACF graph.



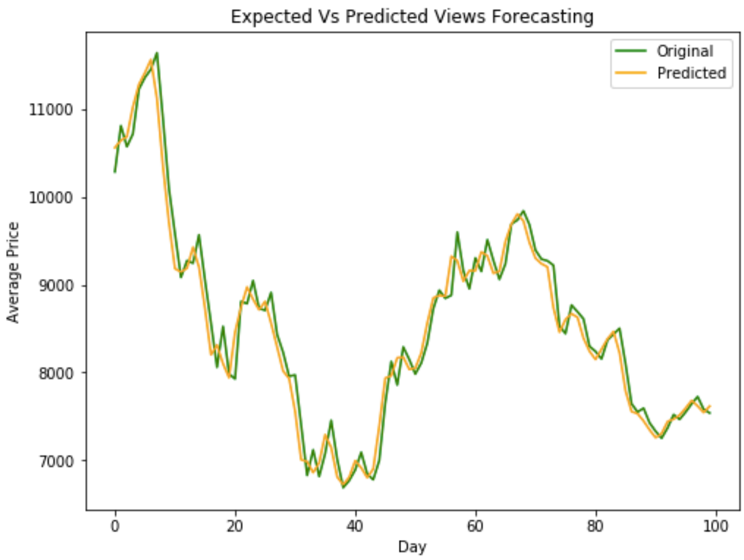
The ACF can be used to estimate the MA-part, i.e q-value, the PACF can be used to estimate the AR-part, i.e. p-value

Thus an ARIMA(2,1,0) model was chosen for analysis

Results

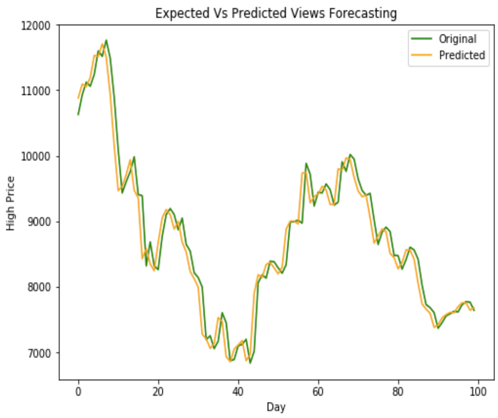
i. Bitcoin Prices

Mean Error in Predicting Test Case Articles : 2.019849 %

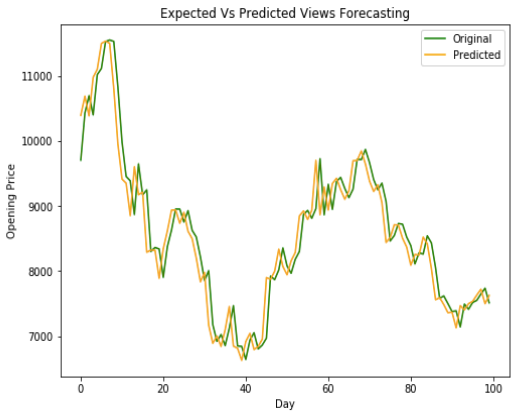


From the plot it is inferred that the ARIMA model predicts the average price of bitcoins accurately. The model can similarly be used to predict the opening, high, low and closing prices of the coins

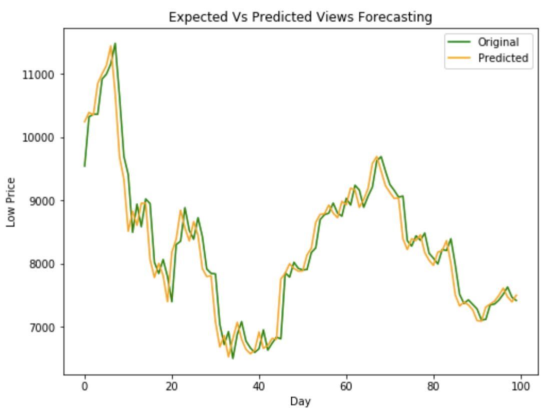
Mean Error in Predicting Test Case Articles : 2.289227 %



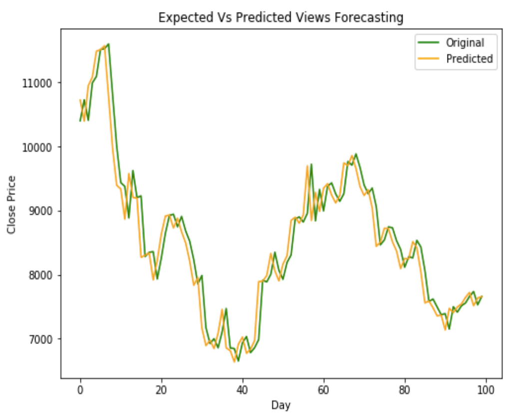
Mean Error in Predicting Test Case Articles : 3.122299 %



Mean Error in Predicting Test Case Articles : 2.816136 %



Mean Error in Predicting Test Case Articles : 3.008508 %



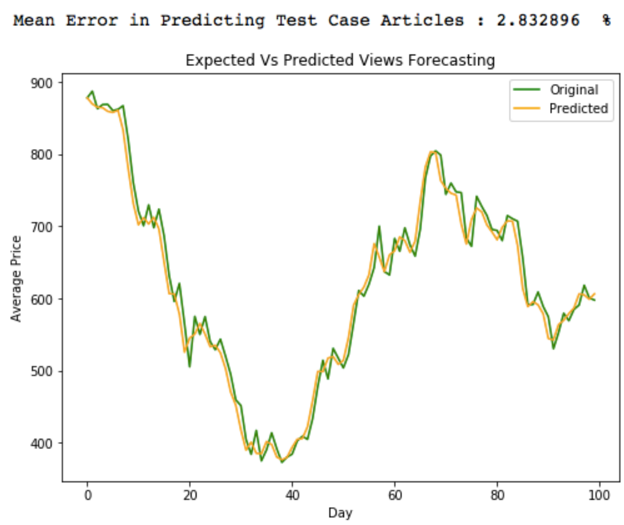
The mean error in predicting the open, high, low and closing prices is higher than the error in predicting the average price howerver the values still do no differ by much and fall within an acceptable range in prediction.

The candlestick plot super imposes the actual prices (light green: increase in value and pink: decrease in value) of Bitcoins over the predicted prices (dark green: increase in value and red: decrease in value).
It is inferred that the predictions made from the model in checking if prices increase or decrease tend to lag behind a little. For example, if the price of bit coin reduces its being predicted as it is the same as the previous day an then predicts the actual drop in prices the next day accurately.

The model was then used to predict Etherum and Ripple prices in a similar manner

ii. Ethereum Prices

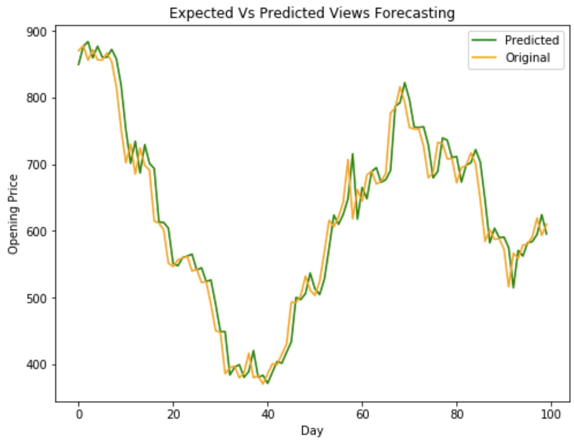
Plotted below is the average price of Ethereum both actual and the price predicted by the ARIMA model



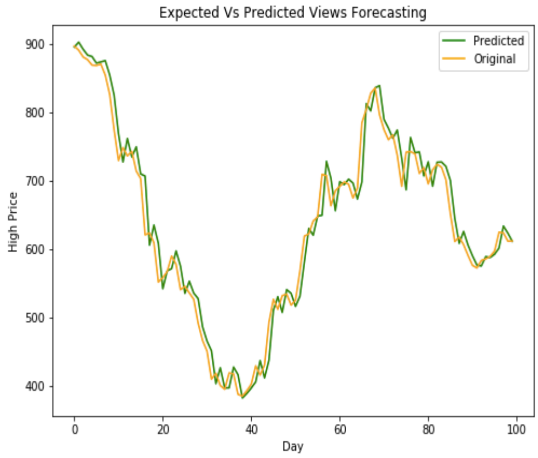
The average Etherum price is also predicted at a similar accuracy as the model predicted the bitcoin prices.

Plotted below are the opening, high, low and closing prices (left to right) of Ethereum both actual and predicted prices.

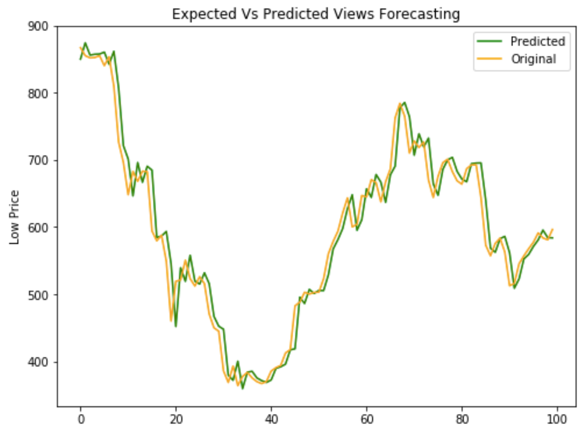
Mean Error in Predicting Test Case Articles : 3.972729 %



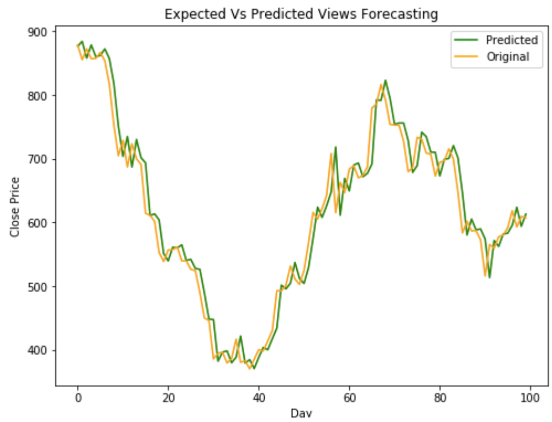
Mean Error in Predicting Test Case Articles : 3.338154 %



Mean Error in Predicting Test Case Articles : 3.795721 %

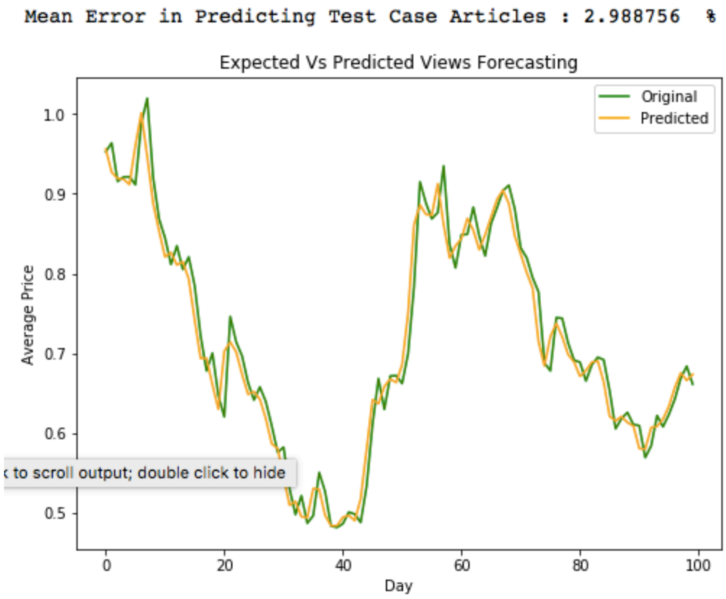


Mean Error in Predicting Test Case Articles : 4.011982 %

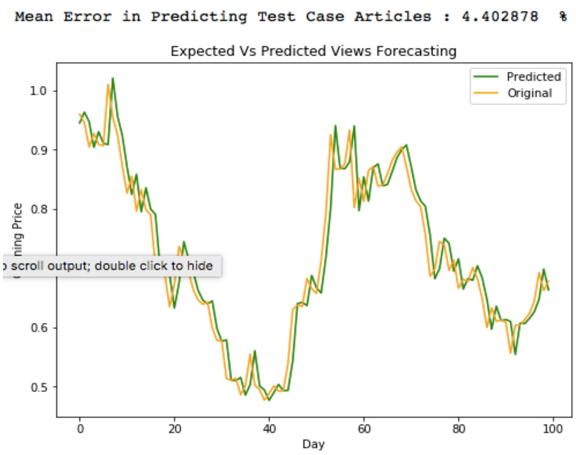


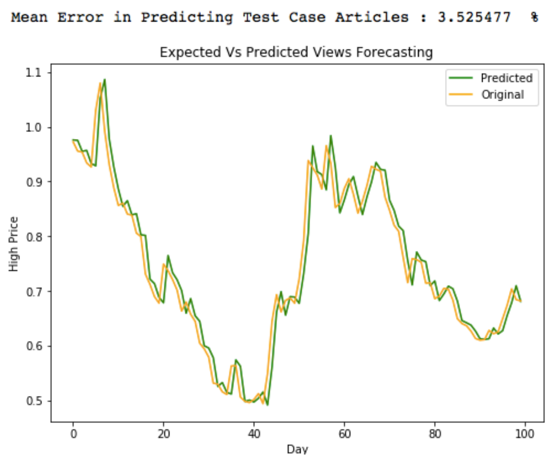
iii. Ripple Prices

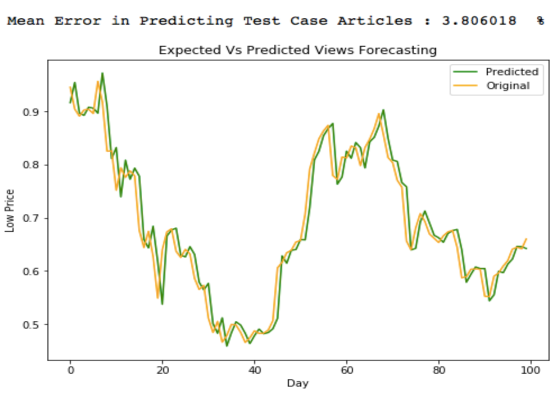
Plotted below is the average price of Ethereum both actual and the price predicted by the ARIMA model

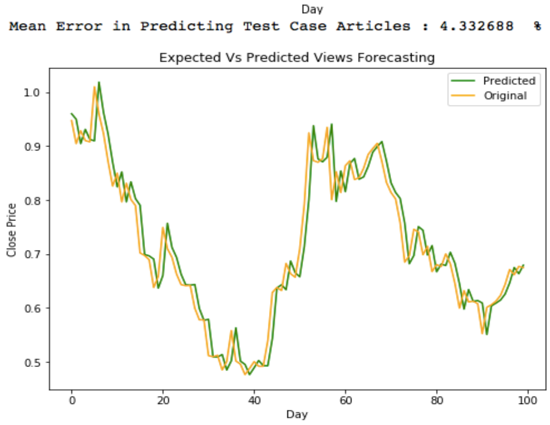


The average Ripple price is also predicted at a similar accuracy as the model predicted the bitcoin prices.
Plotted below are the opening, high, low and closing prices (left to right) of Ripple both actual and predicted prices.









From the plots above the following is inferred. Firstly for each of the coins the model tends to predict the average prices of the coins more accurately that it does in predicting the opening, high,low and closing prices.

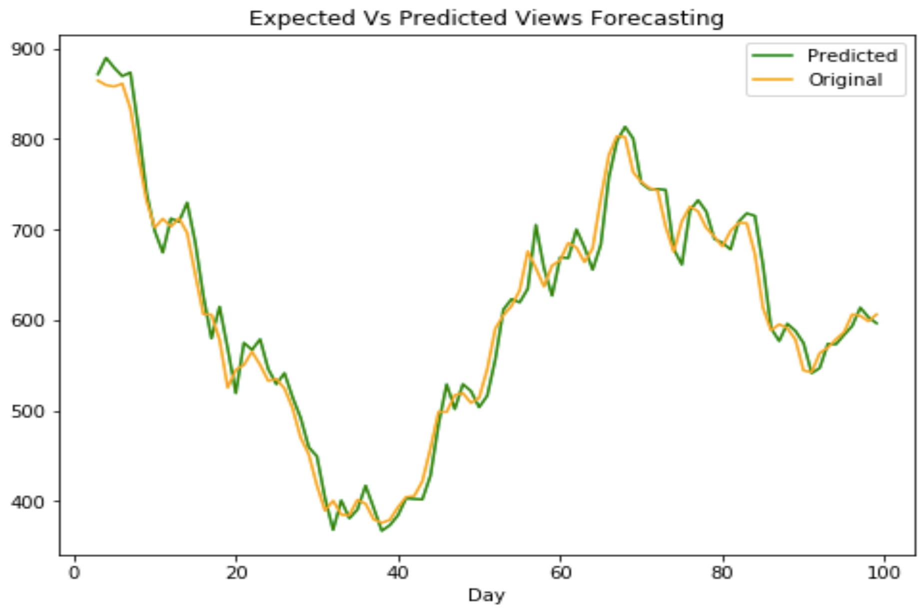
Secondly from the candle stick plots it is again evident when there are large drops or increases in the values of the coins the model tends to lag behind in the prediction a bit.

<u>ARIMAX Model</u>

The ARIMAX model is an extended version of the ARIMA model. It includes also other independent (predictor) variables.
Here it is required to specify the endogenous (target value) and exogenous(independent predictor) in the model.
From the above statistical inference using Granger Causality and spearman coefficient it is evident that there seems to be an effect in the prices of Ethereum and Ripple with changes in Bitcoin value. Here the goal us to predict the average Ethereum and Ripple prices using both the historical data of the coin's values as the endogenous variable and the history of the average bitcoin value as the independent exogenous variable.

i. Average Ethereum value using Average Bitcoin value

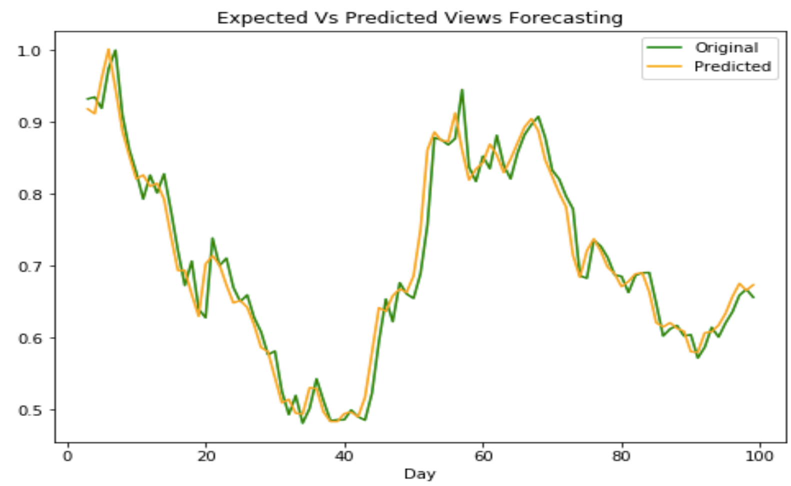
Mean Error in Predicting Test Case Articles : 2.957909 %



ii. Average Ripple value using Average Bitcoin value

Mean Error in Predicting Test Case Articles : 2.940953 %

<matplotlib.figure.Figure at 0x1137ee978>



It is seen that the model has a slight improvement in its accuracy when predicting the prices for Ripple.

It is noted that a similar approach can be used in predicting different values of the coins for example the opening price can be used to predict the closing price etc.
But it should be noted that it can only be used if there is a relation in the trend between the values. For example because it is seen from grangers casuality test that Ethereums prices does not cause a change in the price for Ripple the model was not run using the same.