

Milestone 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 cryptocurrency 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 cryptocurrency.

These findings and model would be useful to the aforementioned users who trade and invest in cryptocurrencies 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 cryptocurrencies. Each entry has 13 variables, the slug, symbol and name of the cryptocurrency 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 cryptocurrency peaked at during that day, the lowest price in dollars the cryptocurrency 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 columns record the close ratio and the spread.

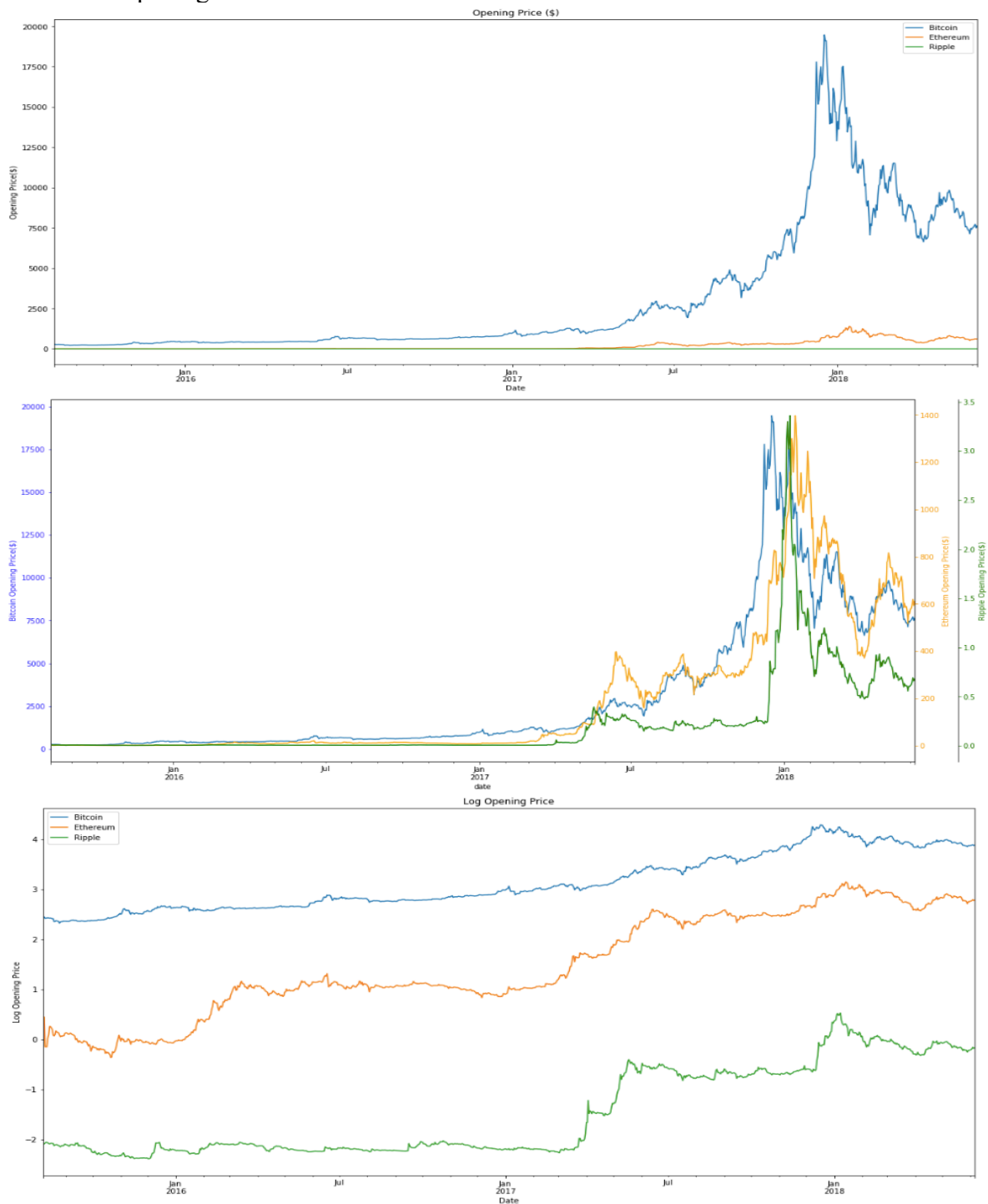
For this analysis only the top three ranked cryptocurrencies will be utilized, Bitcoin (BTC), Ethereum (ETH) and Ripple (RFX). The observations for these cryptocurrencies have been sliced out of the original data frame and put into their own data frames respectively. The date column is then used as the 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.

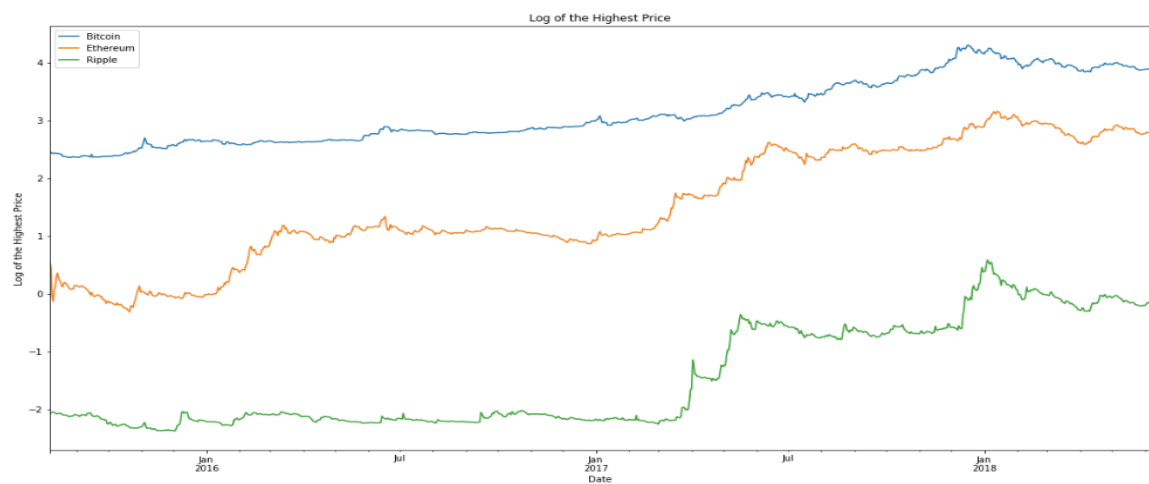
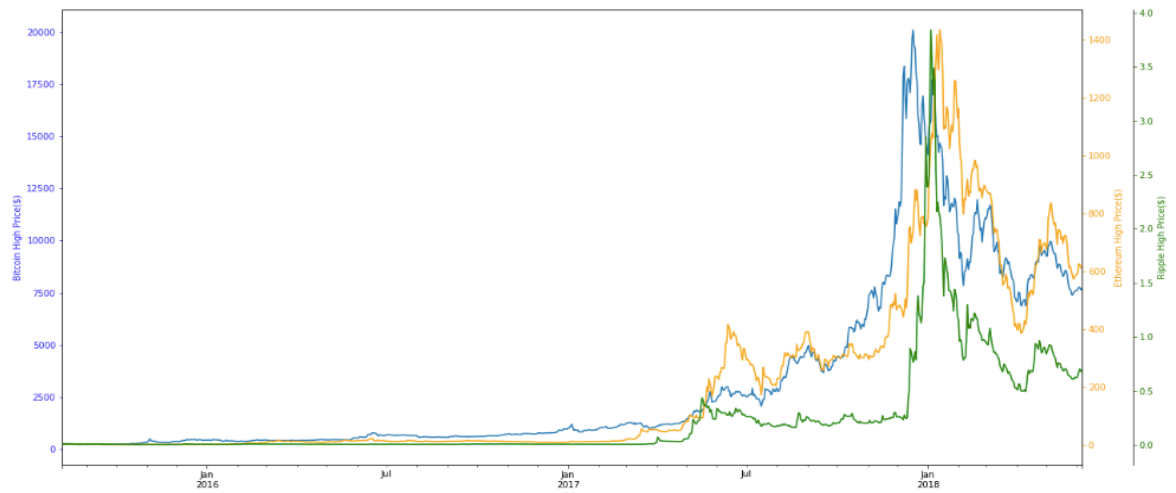
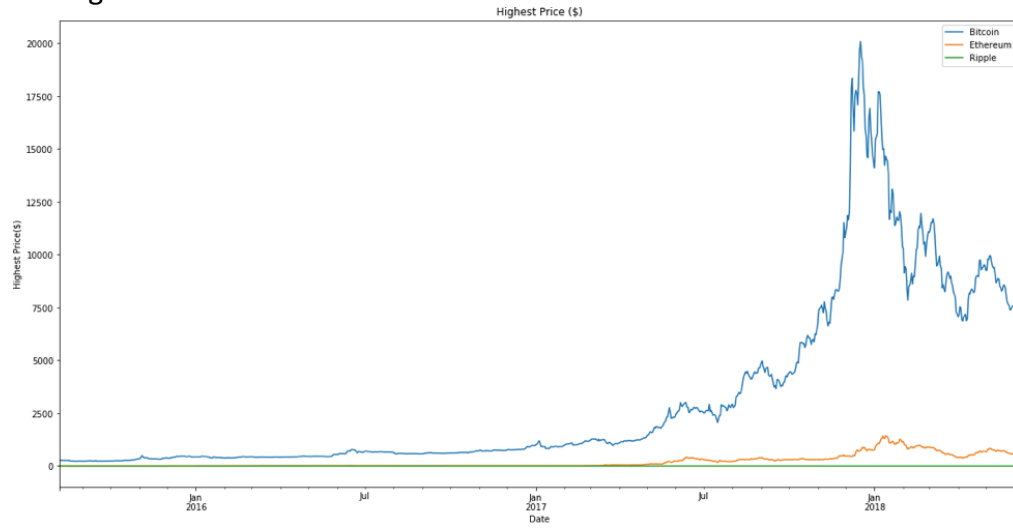
Exploratory Data Analysis

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.

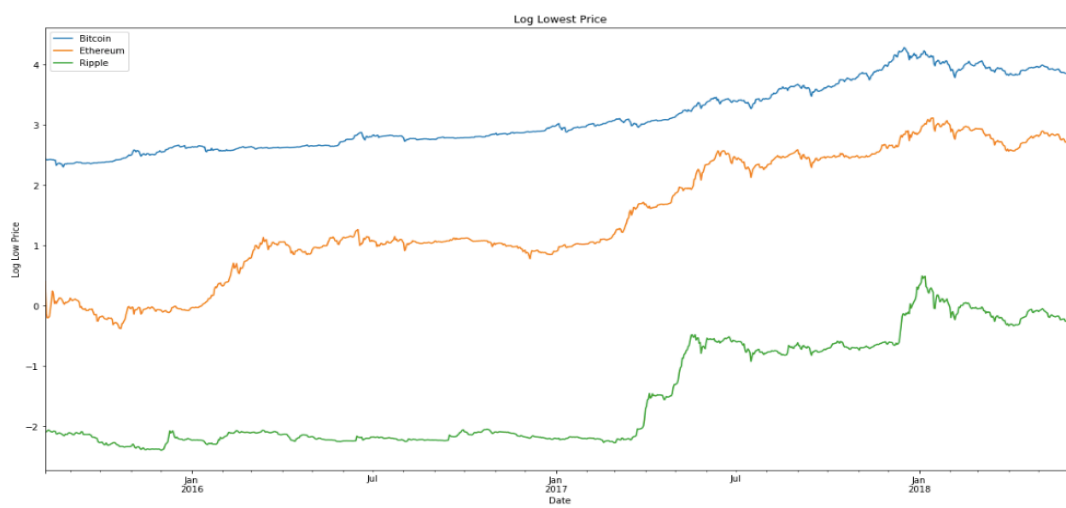
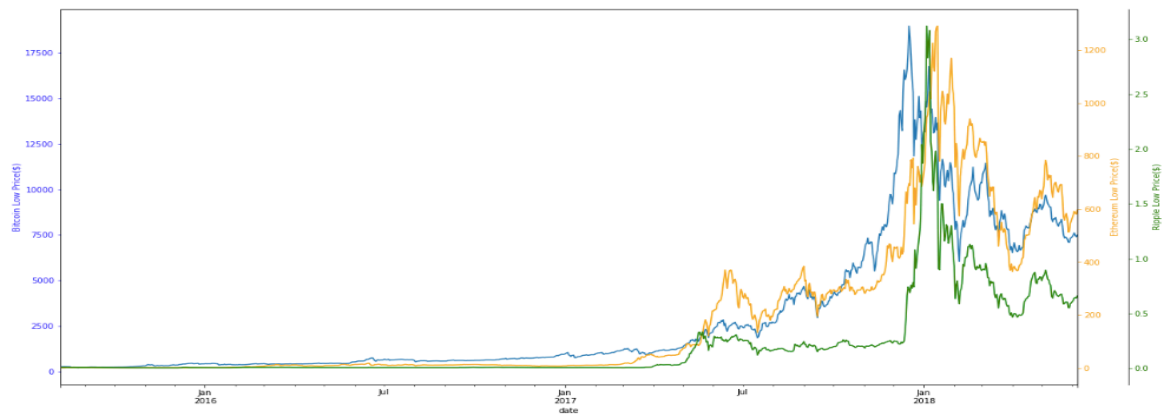
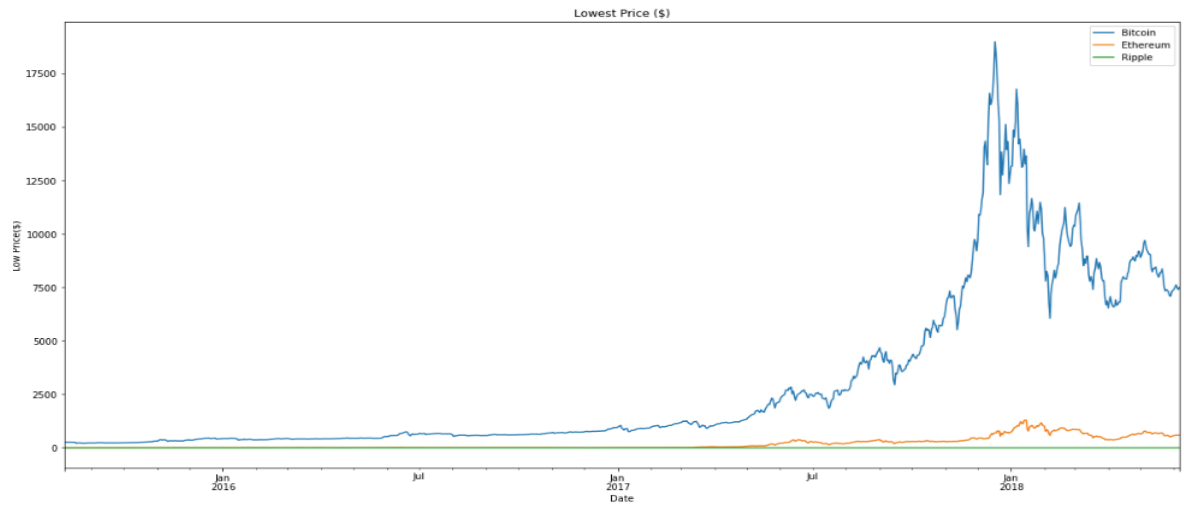
i. Opening Price



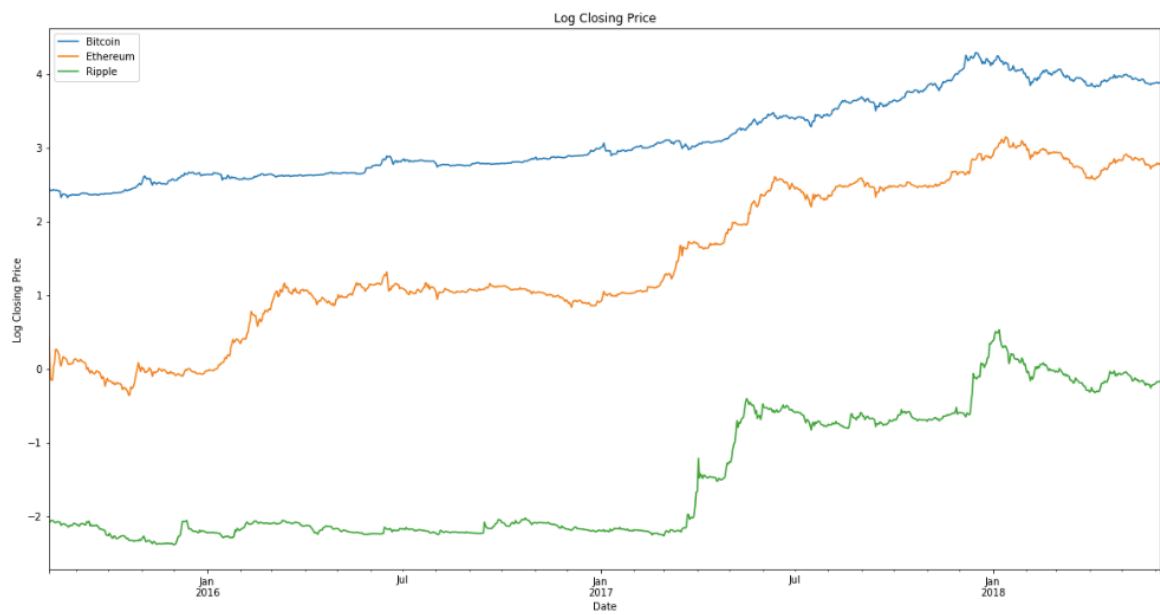
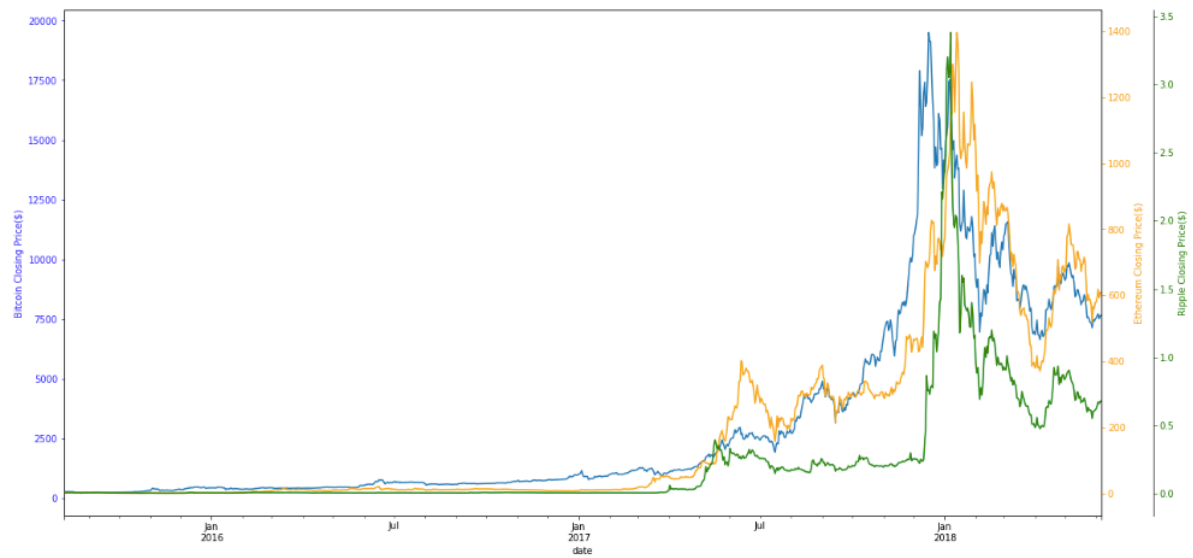
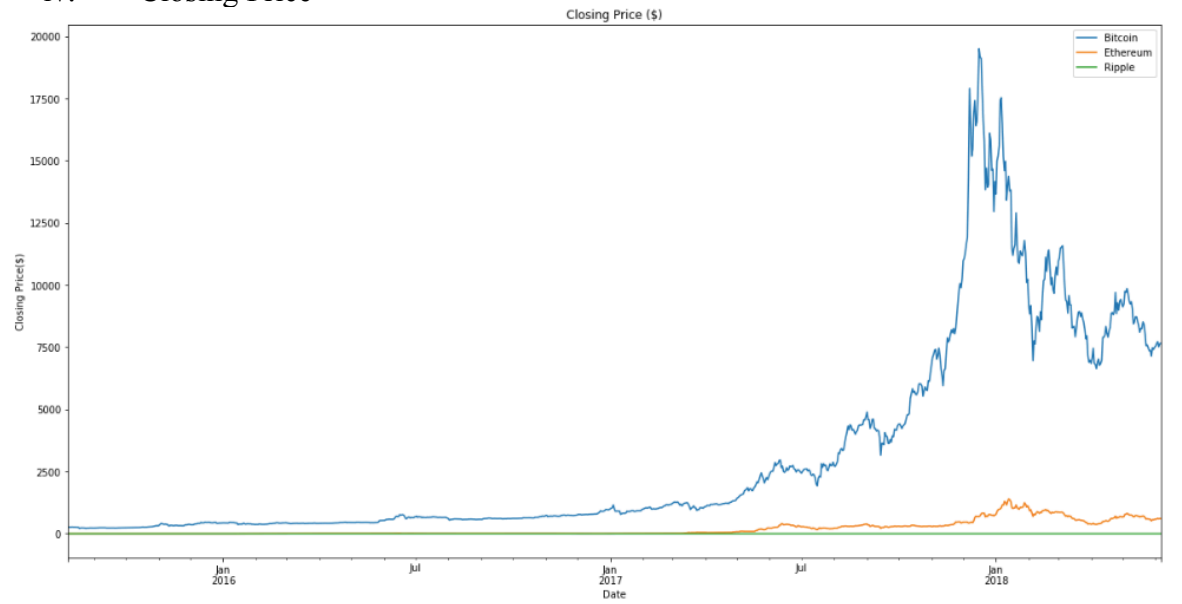
ii. Highest Price



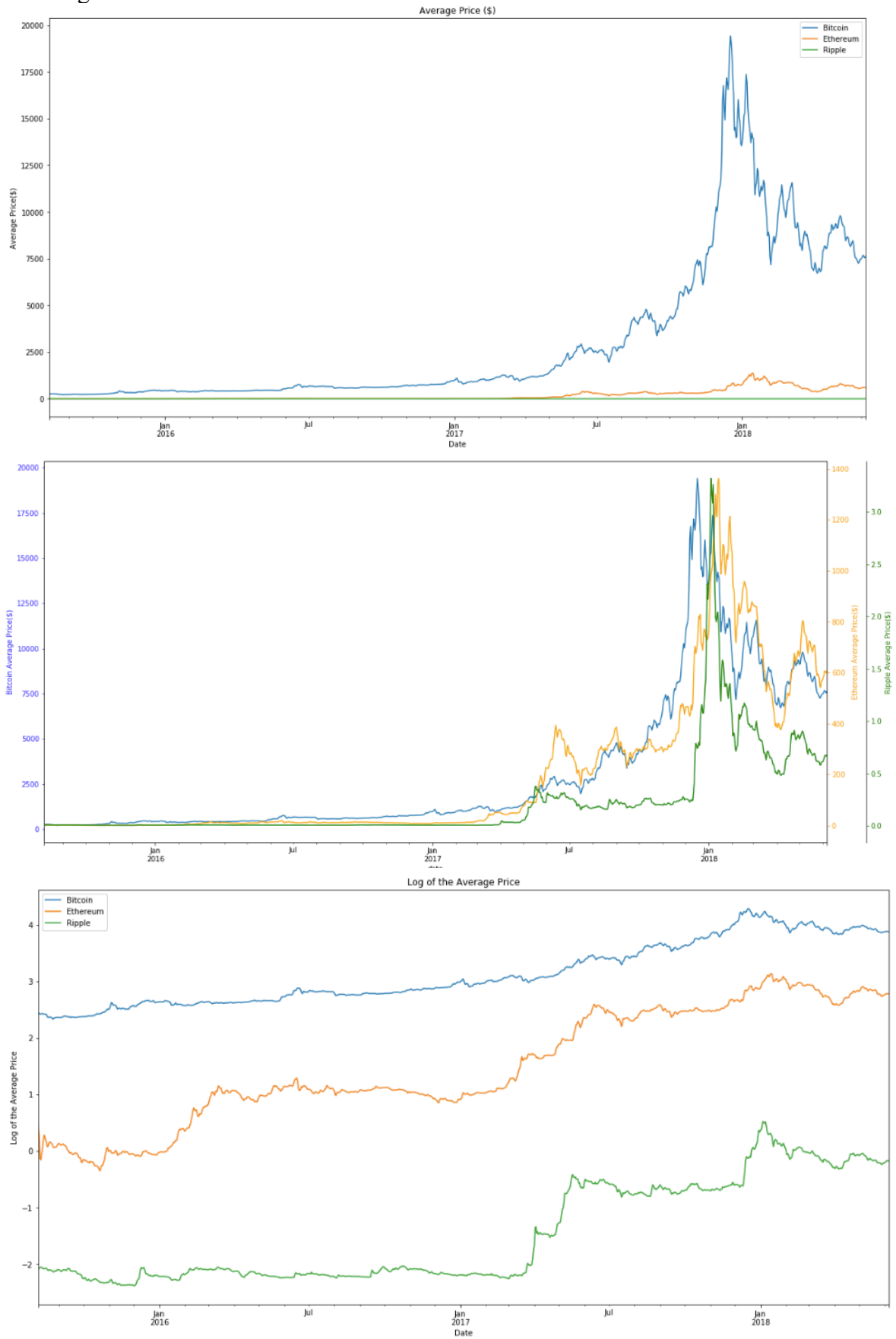
iii. Lowest Price



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.

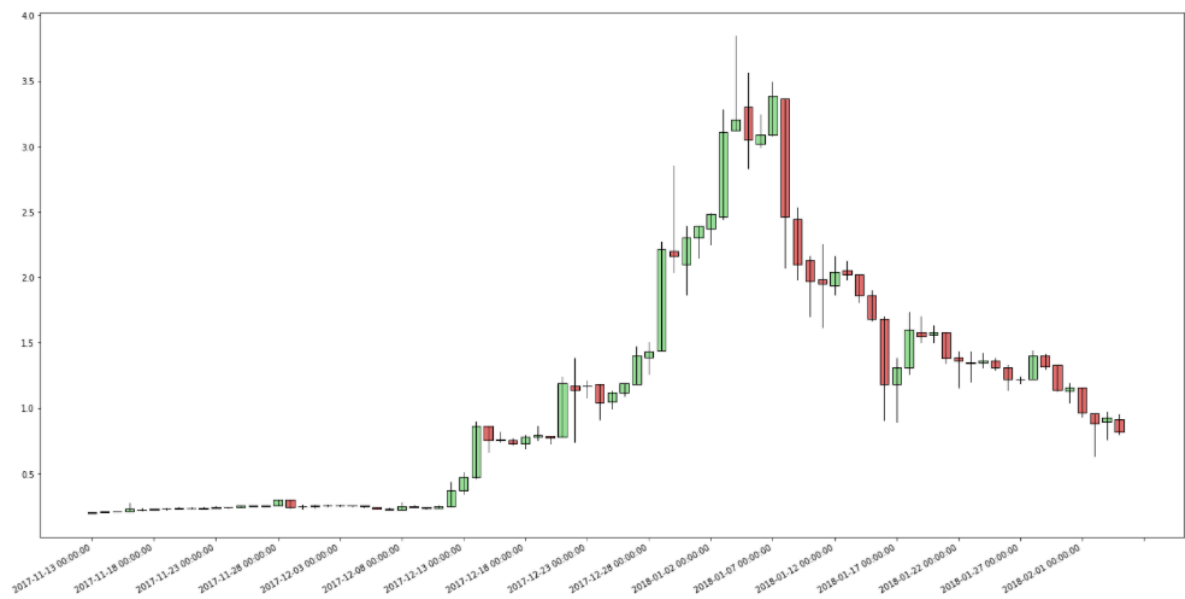
i. Bitcoin Price (\$)



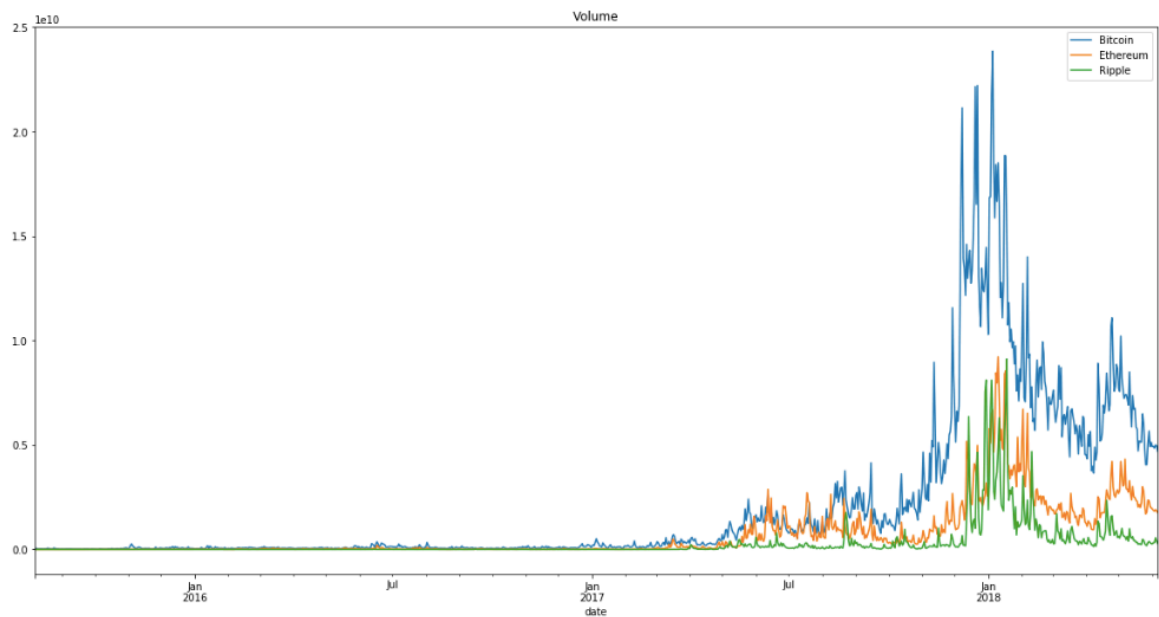
ii. Ethereum Price (\$)

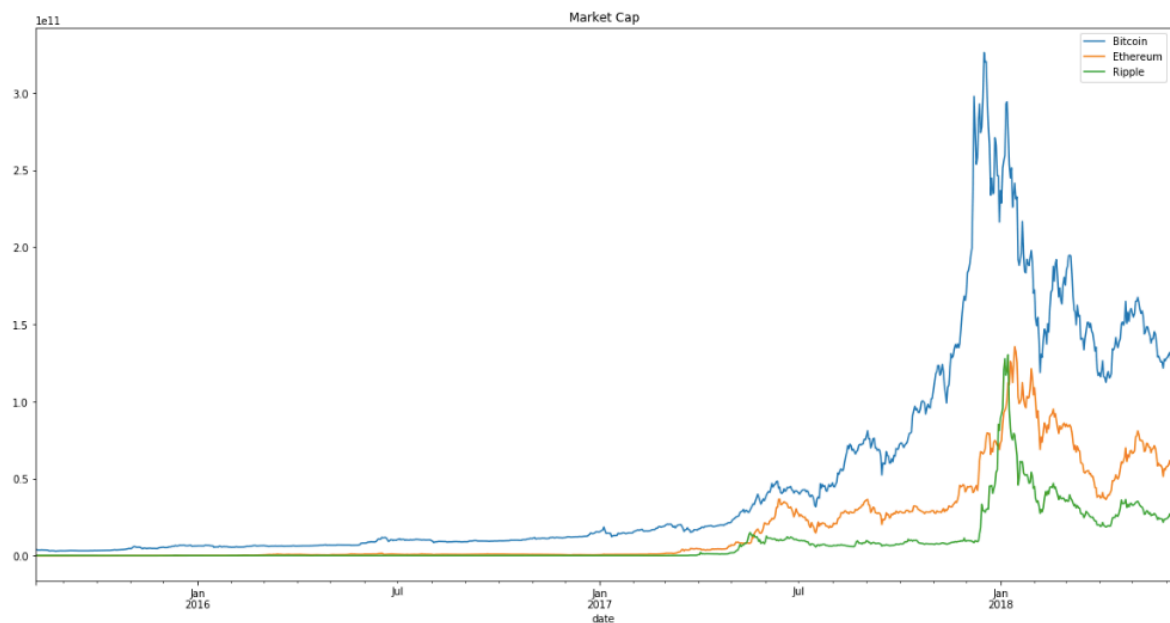


iii. Ripple Price (\$)



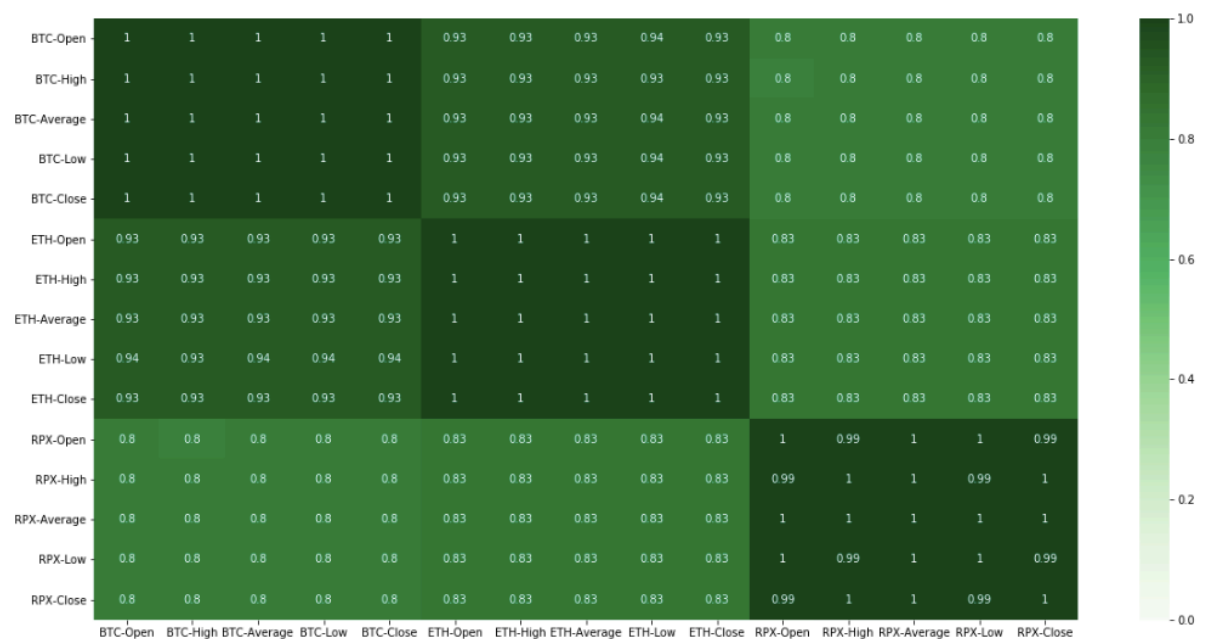
Further plotted below were the Market Cap and Volume of coins being traded and it is inferred that as the prices of the coins rose or fell so did the market cap and volume of coins being traded.





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 values 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.