# Preparing Python Packages

## Loading Data

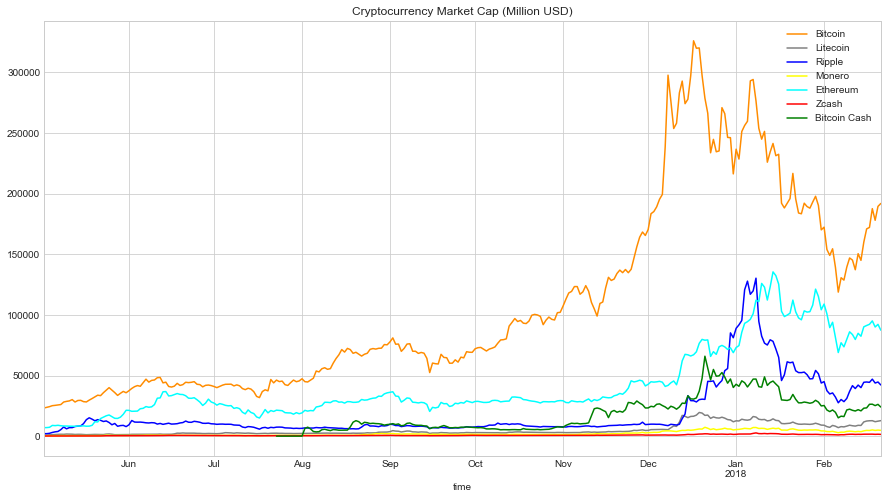
For the project we have imported pandas package which is used for providing fast and expressive data structures designed to make working with relational data very easy to understand. As this project had timeseries element in it as well we have also imported the datetime package to use. Matplotlib is imported for plotting purposes while plotly is also imported to plot different forms of graphs in this project. Sklearn is imported for data analysis purposes in the project.

The dataset is loaded in csv format, with columns such as date, name, symbol, open (opening price of that date), high (highest rate of that date), low (lowest rate of that date), close (closing rate of that date), volume of the transaction etc. We have chosen the top 7 cryptocurrencies according to Business Insider.

We have checked for missing values and duplicate values as well and found out that there were no missing values in the data set, neither were there any duplicate records in the data set. The data was clean.

# Analysis

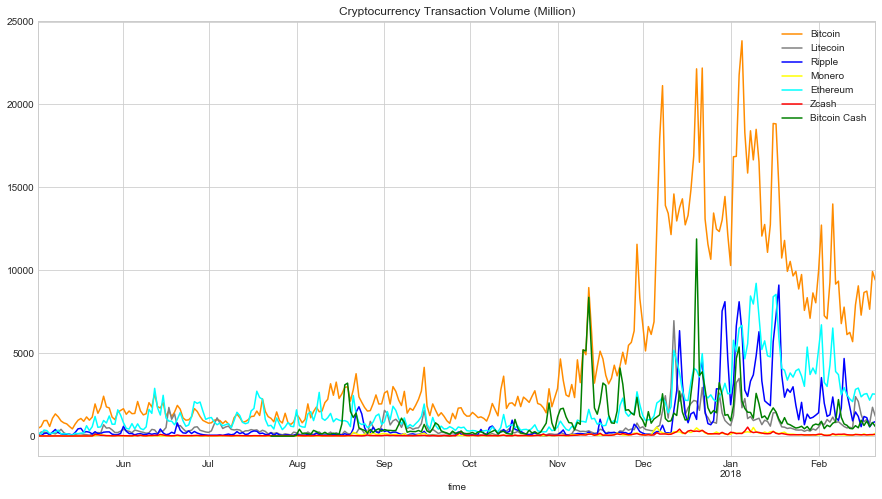
## Market Capitalization



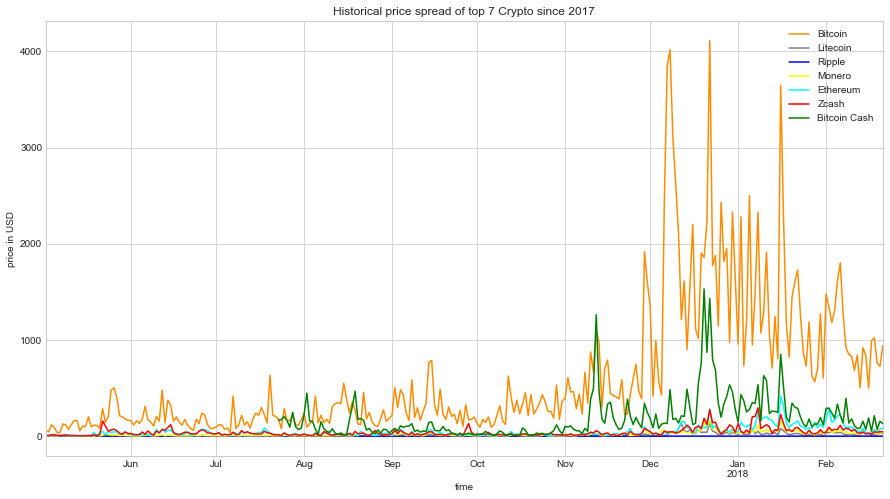
According to Investopedia, Market capitalization refers to the total dollar market value of a company’s outstanding shares. Commonly referred to as market cap, it is calculated by multiplying company’s shares outstanding by the current market price of one share.

It is important to find out the market cap as it tells us the value of a company or in this case the cryptocurrency. As we can see from the graph that the market cap for Ethereum came second only to Bitcoin since June 2017, going to third number for a short period of time in January 2018, but bouncing back to second position in January 2018. We can see the trend from the graph. The market cap for Bitcoin started to increase manifolds since December 2017, and other cryptocurrencies followed the suit and with a lagged effect the market cap for other cryptocurrencies started to increase around end of December-start of January 2018.

## Transaction volume

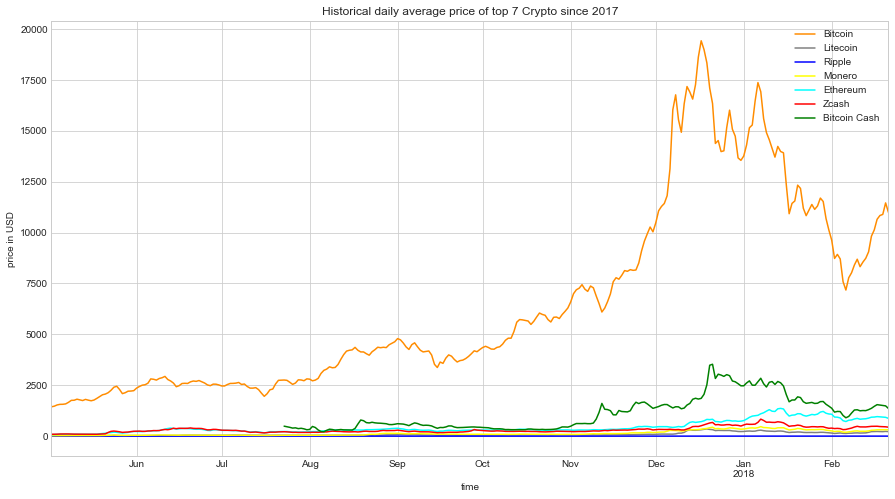


The transaction volume graph is denoted in million. The graph shows that Bitcoin, by far, is the leading cryptocurrency in terms of volume traded. Interestingly, there are some points, where Ethereum overthrew Bitcoin to be the leading cryptocurrency in terms of volume during the months of June till August 2017. We can see the spikes in the graph (denoting a rise in volume) for Bitcoin are way ahead of every other cryptocurrency. After December 2017, the gap between Bitcoin and other cryptocurrencies widened and no one could catch up in terms of volume.

Price spread since 2017

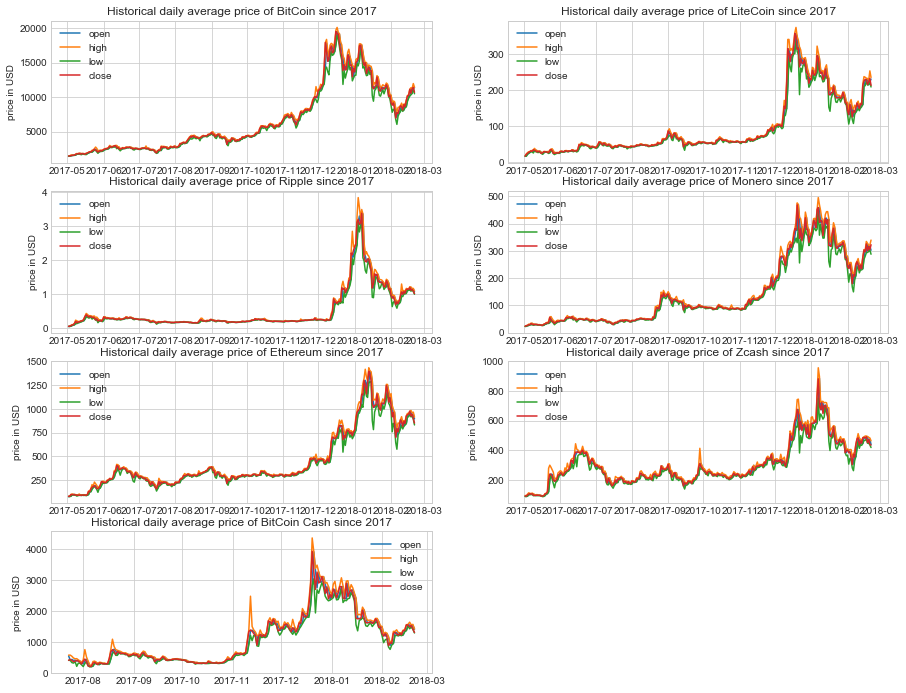
The above graph shows the historical price spread of top 7 Cryptocurrencies since 2017. We can analyse that even here Bitcoin is way ahead of every other cryptocurrency. Remarkably, Bitcoin had a price spread of 4000 USD, which could have given investors a chance to earn quick money. Just two out of the seven cryptocurrencies managed to touch 1000 USD price spread while others merely crossed 500 USD. Ethereum saw an increase in price spread after December 2017.

## Historical daily average prices of top 7 Crypto since 2017



Graph A

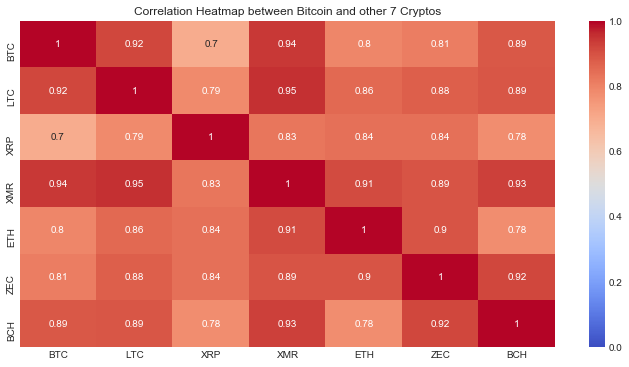
The above graph shows historic daily average price of top 7 crypto since 2017. The daily average price was calculated by adding up the highest price, lowest price, opening price and the closing price. Then dividing the sum by 4. We can see that average price of Bitcoin is higher than any other crypto throughout our time frame. The average price of Ethereum peaked in later part of January 2018.



The set of graphs above this text complements the graph A. We can see that the average prices of each cryptocurrencies started to increase after December 2017. Each graph has 4 different lines representing open, high, low and closing rate.

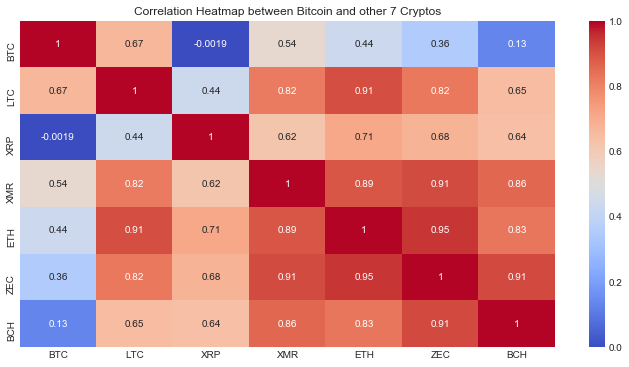
## Correlation between different cryptocurrencies

|  |
| --- |
| plt.figure(figsize=(12,6))  sns.heatmap(close.corr(),vmin=0, vmax=1, cmap='coolwarm', annot=True)  plt.title('Correlation Heatmap between Bitcoin and other 7 Cryptos')  plt.show() |



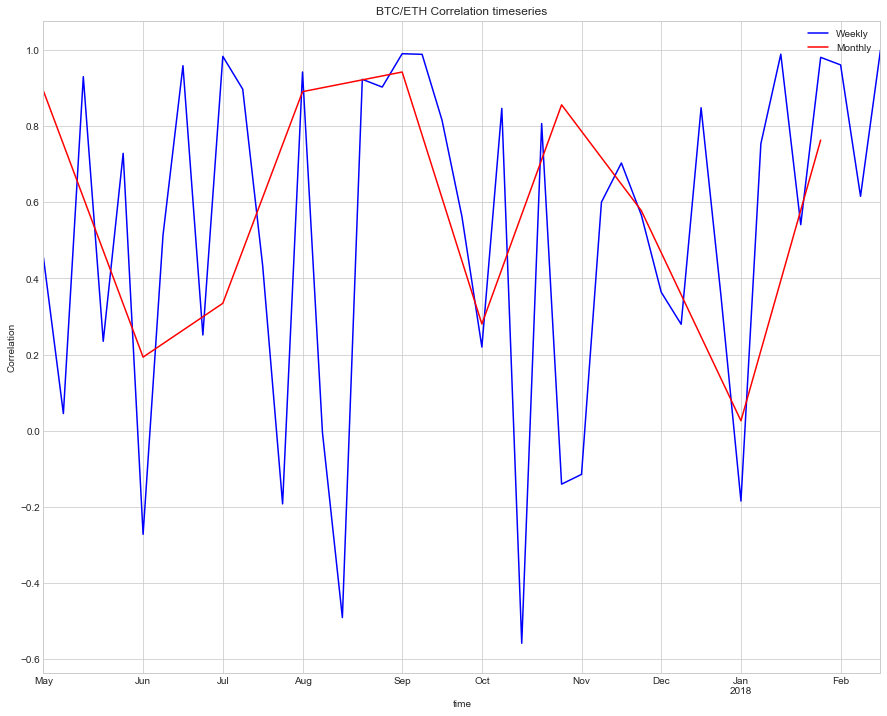
The above figure shows correlation heatmap between Ethereum and 7 other crypto-currencies. The higher the correlation, the darker it gets. Red indicates higher correlation (closer to 1), while blue indicates no correlation (close to 0). The figure is a heatmap from May 2017 till February 2018. As we can see from the graph above, the correlation between Ethereum and Bitcoin is 0.8 which is quite high. The highest correlation of ETH occurs with XMR, which is 0.91, while BCH has the lowest correlation with ETH (0.78).

|  |
| --- |
| #Correlation for a specific time period  import datetime as dt  plt.figure(figsize=(12,6))  sns.heatmap(close['30-11-2017':'01-01-2018'].corr(),vmin=0, vmax=1, cmap='coolwarm', annot=True)  plt.title('Correlation Heatmap between Bitcoin and other 7 Cryptos')  plt.show() |



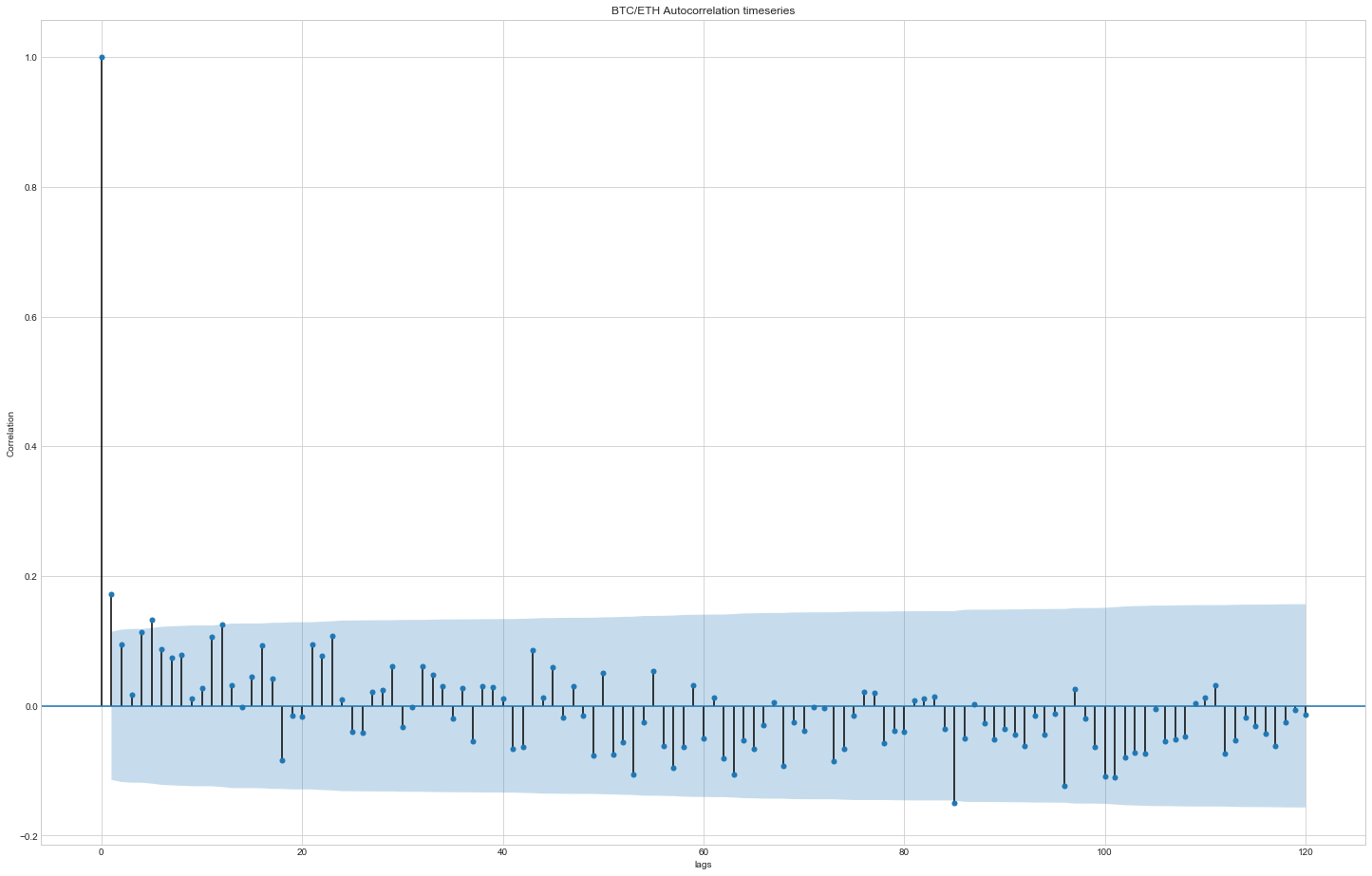
The figure above is also a heatmap for correlation between Ethereum and 7 other cryptocurrencies, but the timeline for this heatmap is December 2017. WE can see that this figure can see quiet a few blue boxes as compared to the first figure. Correlation between different crypto-currencies have dropped in the month of December. The correlation between ETH and BTC have dropped to 0.44 in this figure as compared to 0.8 from the first figure.

## Correlation Time series



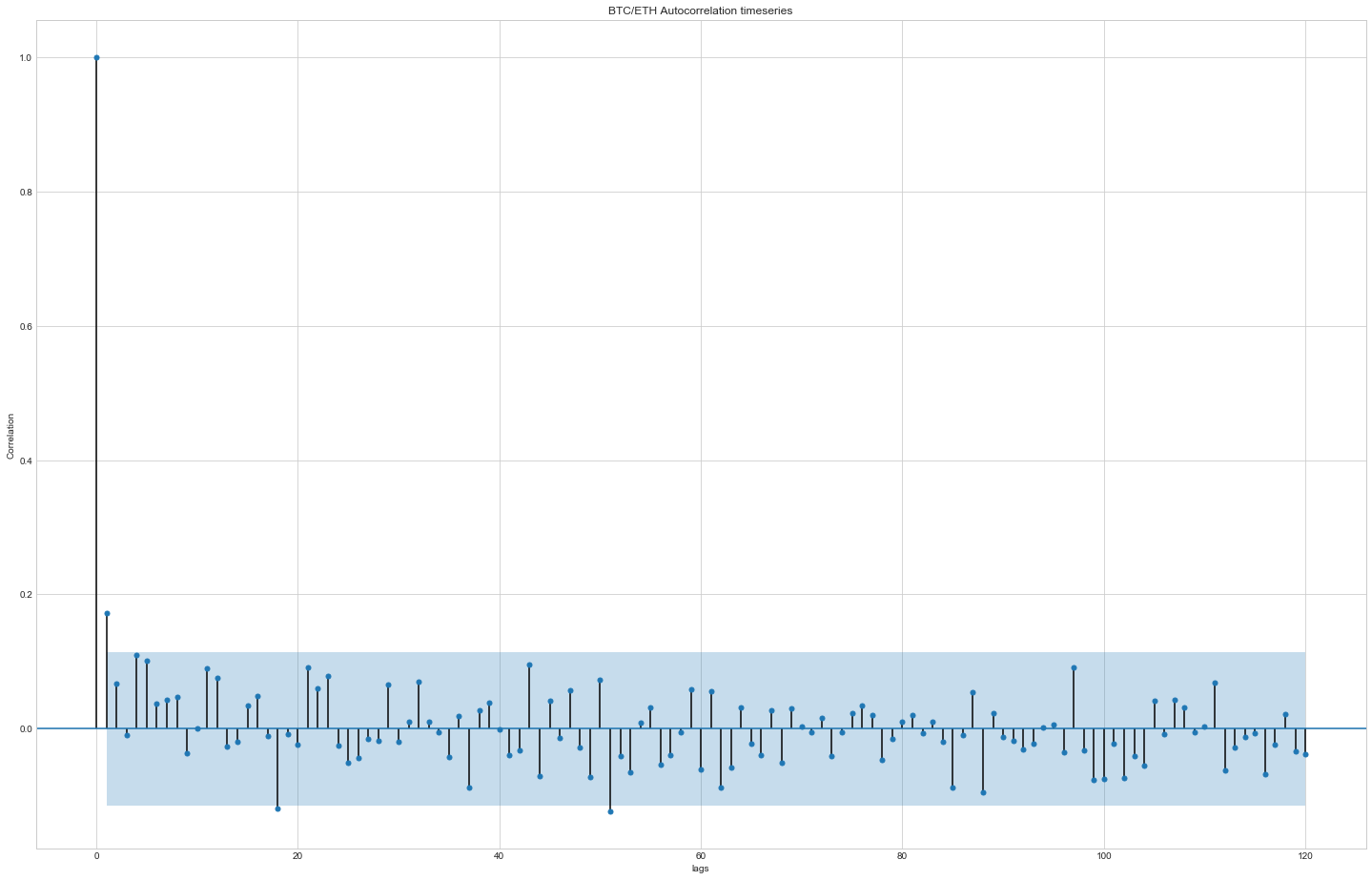
The graph shows correlation between BTC and ETH with timeseries prospective. WE can see seasonality in the graph with a distant form of sin wave curve. The red line represents the monthly correlation between ETH and BTC while the blue line represents the weekly correlation between the two cryptocurrencies. WE can see during the month of May till June the weekly correlation between the two varied from 0.9 to -0.3(start of June). WE can see that the correlations changed weekly in the month of May, as it decreased in the first week, and increased in the next week and again decreasing in the third week. During the first week of September the correlation between the two cryptocurrencies stayed the same as previous week (close to 1). This is the only instant where the correlation did not change between two weeks. This is also, in fact the time-period when weekly correlation between the two was at its highest point. The monthly correlation was also at its highest point during the month of September. The lowest monthly correlation between the wo occurred at the start of January (close to 0.05). However, the lowest weekly correlation was -0.5 during second week of October. Interestingly, during the month of October, the monthly correlation went up from 0.3 in the end of September to approximately 0.9 during the end of October.

## Autocorrelation



The above autocorrelation plot, also known as correlogram, shows Autocorrelation between Bitcoin and Ethereum in timeseries. This type of plot is used to check randomness in the dataset. We can see that the plot nears zero value and hence the data set is random. If observed closely, we can also look at a distant form of a sine function wave.

## Partial autocorrelation



The above plot shows partial autocorrelation function. We can also notice the sine wave function here in this plot. Simply it is a regression of the series against its past lags.

## Regression functions

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Bitcoin | Ethereum | Litecoin | Bitcoin Cash | Ripple | Monero | Zcash |
| Random Forest Regressor | R2: 0.85  MAE: 1010.13  MSE: 2540545 | R2: 0.84  MAE:66.67  MSE:8603.4 | R2: 0.75  MAE:17.7  MSE:1068.1 | R2: 0.81  MAE: 258  MSE: 121403 | R2: 0.67  MAE: 0.14  MSE: 0.11 | R2: 0.8  MAE: 30.45  MSE: 2641.2 | R2: 0.93  MAE: 29.23  MSE: 1654.13 |
| Gradient Boosting | R2: 0.86  MAE: 1016.9  MSE: 2464682.2 | R2: 0.82  MAE: 70.71  MSE: 9799.4 | R2: 0.7  MAE:20.7  MSE:1272.7 | R2: 0.79  MAE: 265.55  MSE: 134112.3 | R2: 0.68  MAE: 0.16  MSE: 0.11 | R2: 0.76  MAE: 31.08  MSE: 3114.3 | R2: 0.93  MAE: 27.7  MSE: 1567.1 |
| Extra Trees | R2: 0.87  MAE: 900.37  MSE: 2222611.46 | R2: 0.81  MAE: 70.1  MSE: 10288.24 | R2: 0.71  MAE: 18.7  MSE: 1233.9 | R2: 0.85  MAE: 239.23  MSE: 97964.84 | R2: 0.56  MAE: 0.17  MSE: 0.15 | R2: 0.82  MAE: 30.03  MSE: 2306.7 | R2: 0.94  MAE: 25.1  MSE: 1319.42 |
| Bayesian Ridge | R2: 0.7  MAE: 1714.85  MSE: 5178364.12 | R2: 0.47  MAE: 106.9  MSE: 29131.3 | R2: 0.29  MAE: 41.35  MSE: 2975.29 | R2: 0.32  MAE: 541.09  MSE: 435625.1 | R2: 0.17  MAE: 0.27  MSE: 0.28 | R2: 0.59  MAE: 51.88  MSE: 5314.11 | R2: 0.46  MAE: 88.31  MSE: 12648.8 |
| Elastic Net CV | R2: 0.66  MAE: 1831.29  MSE: 5841023.25 | R2: 0.35  MAE: 143.7  MSE: 36031.04 | R2: 0.31  MAE: 41.19  MSE: 2889.68 | R2: 0.31  MAE: 552.45  MSE: 442785.5 | R2: 0.17  MAE: 0.27  MSE: 0.28 | R2: 0.58  MAE: 52.46  MSE: 5393.12 | R2: 0.44  MAE: 89.3  MSE: 13065.29 |

The above values show different regression functions for Bitcoin. Five different regression functions are used: Random Forest, Gradient Boosting Regressor, Extra Trees Regressor, Bayesian Ridge and Elastic Net CV. We can see the value of R2, Mean Average Error and Mean Square Error.

Random Forest is one of the methods used for classification and regression. This model is a type of additive model that makes predictions by combining decisions from a sequence of models. Extra Trees regressor (extremely randomized trees) is used to further randomizing tree building in the context of numerical input features. It is closely associated with Random Forest, however in this method, instead of finding an optimal point for each of K randomly chosen features at each node, it selects a cut point at random. It often leads to higher accuracy as is the case with Bitcoin here.

Bayesian Ridge regression was used as an approach to linear regression in context to Bayesian inference. On the other hand, gradient boosting is a machine learning method for classification and regression problems. It builds the model is stage-wise way,

For Bitcoin, Extra Trees Regressor is the most suitable as the R2 value is the highest amongst all and the MAE plus MSE values are the lowest. The same goes for Bitcoin Cash, as the values for R2 are the highest in Extra Trees.

For Ethereum, unlike Bitcoin, the most suitable regression function is the Random Forest Regressor as the R2 value is the highest (0.84). MAE and the MSE value are also the lowest amongst all the functions.

For Litecoin, like Ethereum, the most suitable algorithm is Random Forest Regressor taking account for all the values of R2, MSE and MAE of all five algorithms. The highest R2 value was of Random Forest Regressor in Litecoin.

For Monero and Z Cash, we would go with Extra Trees Regressor due to the values of R2, MSE and MAE. However, for Ripple, the most suitable algorithm is Random Forest Regressor.