Study of Cryptocurrency Market and Bitcoin Price Prediction

Group 8

Aaditi Parab
Komal Ginwal

SPRING 2021

Course: FE 582 A

Last updated on: 05/16/2021

ABSTRACT

In this project, we conduct time series analysis of the returns and volatilities of cryptocurrencies and interdependency over various factors. We initially aim to analyze different cryptocurrencies, their volatility, and their performance over time. Also, we aim to understand what external factors have caused the values of certain cryptocurrencies to unexpectedly dwindle or soar during a particular period. We also aim to create forecast models for certain cryptocurrencies which have seen a large amount of volatility in recent times.

INTRODUCTION

There is a significant interest in the growth and development of cryptocurrencies, the most notable one being Bitcoin. Global trading in these cryptocurrencies has led to highly speculative and "bubble-like" price movements. Cryptocurrencies are a digital currency, and they act as a medium for exchanges and transactions. They are decentralized, which means they are not processed by any banking system and go straight to the consumers. Users' identities are protected through an encryption key, which is a feature that Bitcoin has.

Bitcoin is one of the more popular choices of cryptocurrency. Since its introduction into the market in 2009, it has drastically increased and decreased in value. The analysis below will offer insights on the characteristics of the cryptocurrency and its and trend. Because of an increasing interest in cryptocurrency investments, there is a need to quantify their variation over time.

RESEARCH QUESTIONS

- What happened to the historical values and market capitalizations of different currencies over time?
- What currencies are more volatile than others?
- What are the factors affecting the Bitcoin price?
- Bitcoin Price Prediction

DATASET

Cryptocurrency Historical Price Data – This dataset includes historical pricing records for some of the most valuable crypto currencies according to market capitalization. The data for forecasting of Bitcoin and analysis of multiple cryptocurrencies together is taken from CoinMarketCap.

Historical Data for Bitcoin

Date Range

Date	Open*	High	Low	Close**	Volume	Market Cap
Mar 15, 2021	$59,267.43	$60,540.99	$55,393.17	$55,907.20	$66,419,369,890	$1,042,946,024,860
Mar 14, 2021	$61,221.13	$61,597.92	$59,302.32	$59,302.32	$43,901,225,564	$1,106,226,132,525

The historical price dataset consists of 23 csv files - one file for each currency.
The price data is available on a daily basis from April 29, 2013 to February 27, 2021 and have 34,115 records in total.

Variables	**Description**
Date	Observation Date
Open	Opening price in USD on the given day
High	Highest price in USD on the given day
Low	Lowest price in USD on the given day
Close	Closing price in USD on the given day
Volume	Volume represents the monetary value of the currency traded in a 24 hour period, denoted in USD
Market Cap	Toal market value of the coin on the given day

Custom Columns for analysis purpose:

Variables	**Description**
Name	Name of coin
Symbol	Symbol of coin

DATA ANALYSIS

- **Here is the glimpse of the data we plan to use:**

Global Environment

Data	
Aave	146 obs. of 9 variables
BinanceCoin	1313 obs. of 9 variables
Bitcoin	2862 obs. of 9 variables
Cardano	1245 obs. of 9 variables
ChainLink	1256 obs. of 9 variables
Cosmos	716 obs. of 9 variables
Cryptocom	806 obs. of 9 variables
Dogecoin	2631 obs. of 9 variables
EOS	1337 obs. of 9 variables
Ethereum	2031 obs. of 9 variables
Iota	1355 obs. of 9 variables
Litecoin	2862 obs. of 9 variables
Monero	2473 obs. of 9 variables
NEM	2159 obs. of 9 variables
Polkadot	191 obs. of 9 variables
price_data	34115 obs. of 9 variables
Solana	323 obs. of 9 variables
Stellar	2398 obs. of 9 variables
Tether	2189 obs. of 9 variables
Tron	1263 obs. of 9 variables
Uniswap	163 obs. of 9 variables
USDCoin	873 obs. of 9 variables
WrappedBitcoin	759 obs. of 9 variables
XRP	2764 obs. of 9 variables

*Price_data is consolidated data of 23 currencies with a total of 34115 records

- **Data Cleaning**

<u>Checking for NAs</u>

```
> count(is.na(price_data))
  x.Name x.Symbol x.Date x.High x.Low x.Open x.Close x.Volume x.Marketcap  freq
1  FALSE    FALSE  FALSE  FALSE FALSE  FALSE   FALSE    FALSE       FALSE 34115
>
```

There are no NA values in the dataset for any currency.

Checking for variable class type:

```
> sapply(price_data, class)
     Name    Symbol      Date      High       Low      Open     Close    Volume Marketcap
 "factor"  "factor"  "factor" "numeric" "numeric" "numeric" "numeric" "numeric" "numeric"
```

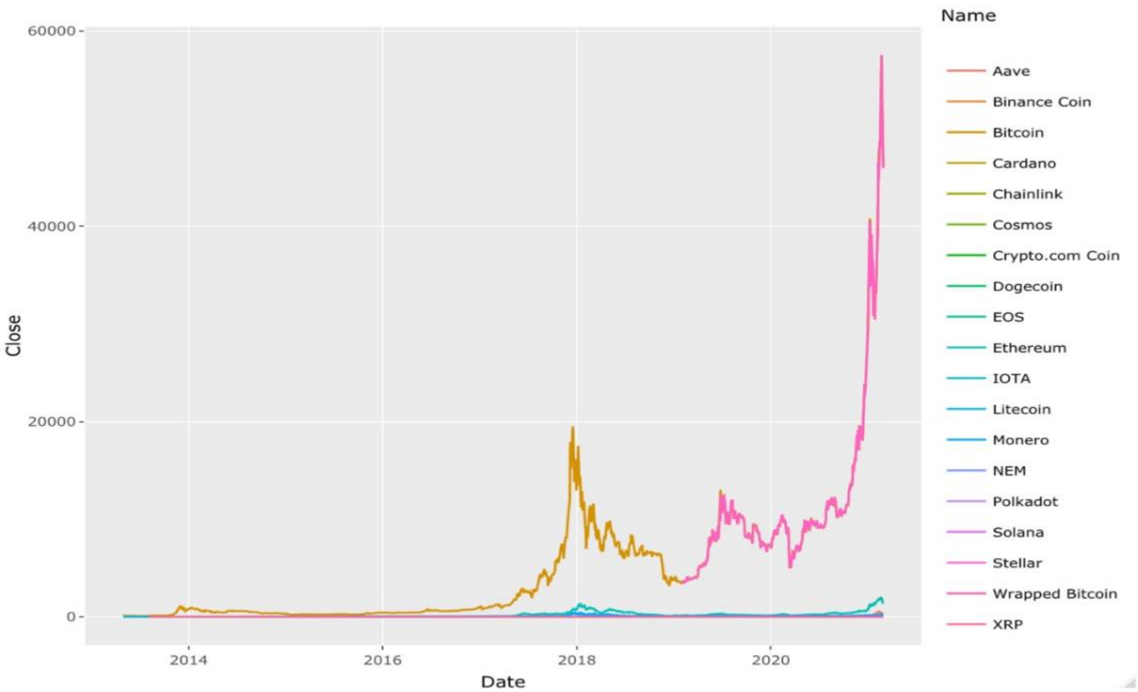
We need to correct the format of the Date Variable into Date Format.

Converting the Date variable into Date Format for all currencies and price_data

```
Aave$Date<- anydate(Aave$Date)
BinanceCoin$Date<- anydate(BinanceCoin$Date)
Bitcoin$Date <- anydate(Bitcoin$Date)
Cardano$Date <- anydate(Cardano$Date)
ChainLink$Date <- anydate(ChainLink$Date)
Cosmos$Date <- anydate(Cosmos$Date)
Cryptocom$Date <- anydate(Cryptocom$Date)
Dogecoin$Date <- anydate(Dogecoin$Date)
Ethereum$Date <- anydate(Ethereum$Date)
EOS$Date <- anydate(EOS$Date)
Iota$Date <- anydate(Iota$Date)
Litecoin$Date <- anydate(Litecoin$Date)
Monero$Date <- anydate(Monero$Date)
NEM$Date <- anydate(NEM$Date)
Polkadot$Date <- anydate(Polkadot$Date)
Solana$Date <- anydate(Solana$Date)
Stellar$Date <- anydate(Stellar$Date)
Tether$Date <- anydate(Tether$Date)
Tron$Date <- anydate(Tron$Date)
Uniswap$Date <- anydate(Uniswap$Date)
USDCoin$Date <- anydate(USDCoin$Date)
WrappedBitcoin$Date <- anydate(WrappedBitcoin$Date)
XRP$Date <- anydate(XRP$Date)
price_data$Date <- anydate(price_data$Date)
```

```
> sapply(price_data,class)
     Name    Symbol      Date      High       Low      Open     Close    Volume Marketcap
 "factor"  "factor"    "Date" "numeric" "numeric" "numeric" "numeric" "numeric" "numeric"
> 
```

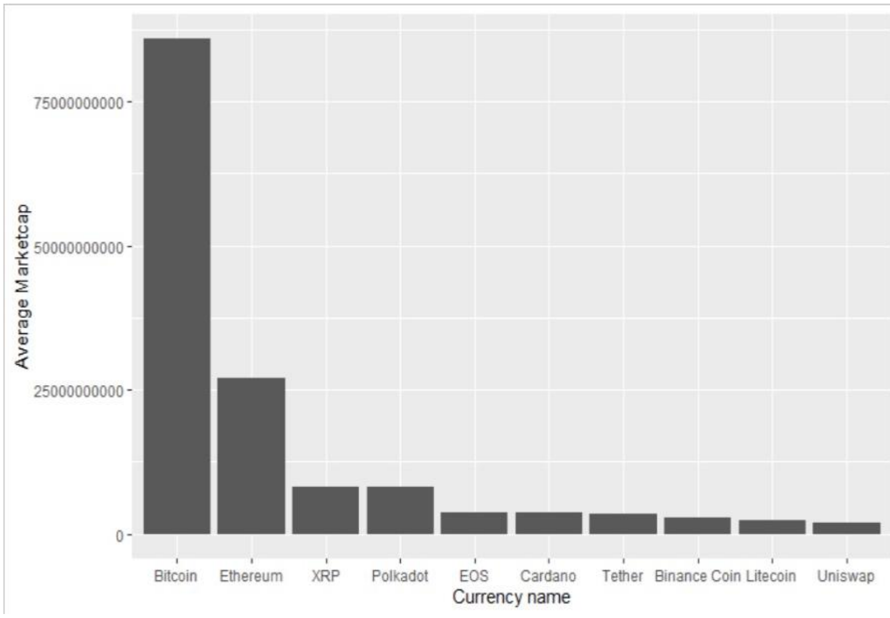
- **Closing Price of 23 Cryptocurrencies over time**



Inference:
Bitcoin has seen the highest increase in closing price over the years followed by Ethereum.

- **Market capitalization of 23 cryptocurrencies over time**

We calculated the average market capitalization of top 10 cryptocurrencies

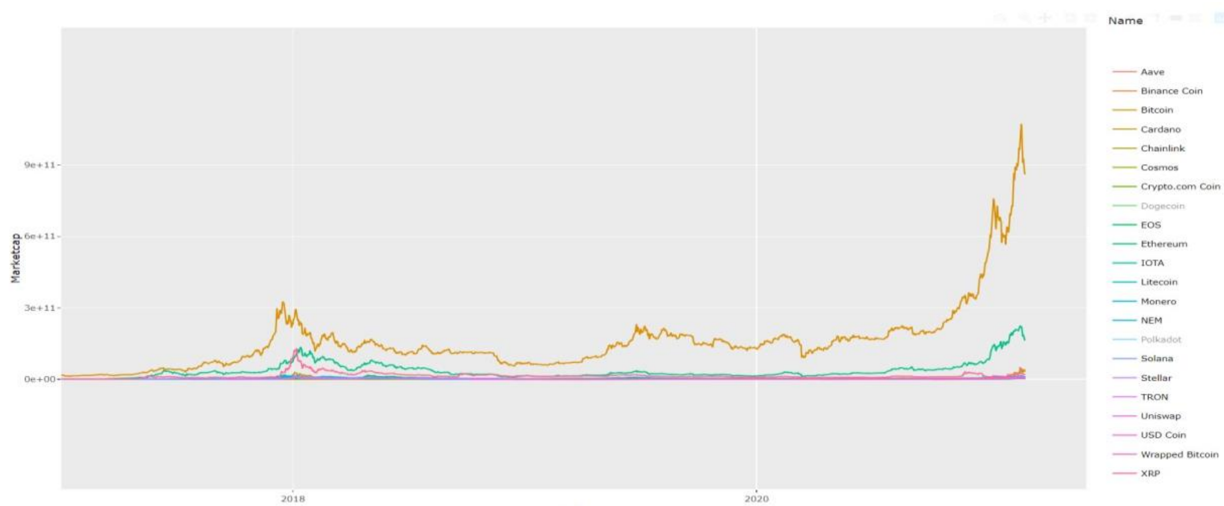


	name	mean.marketcap
1	Bitcoin	85916224373
2	Ethereum	26986559906
3	XRP	8208209621
4	Polkadot	8201284054
5	EOS	3795451404
6	Cardano	3788101188
7	Tether	3453176960
8	Binance Coin	2714482635
9	Litecoin	2429446537
10	Uniswap	2007867885

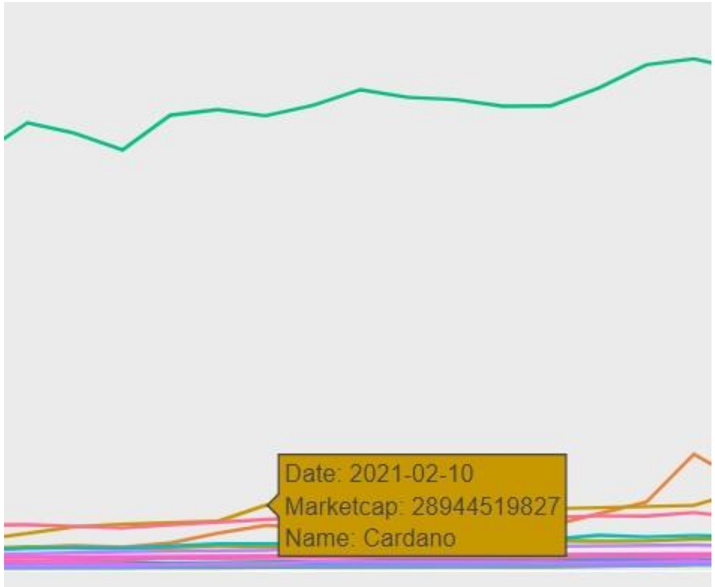
Top 10 currencies according to average market cap

Inference:

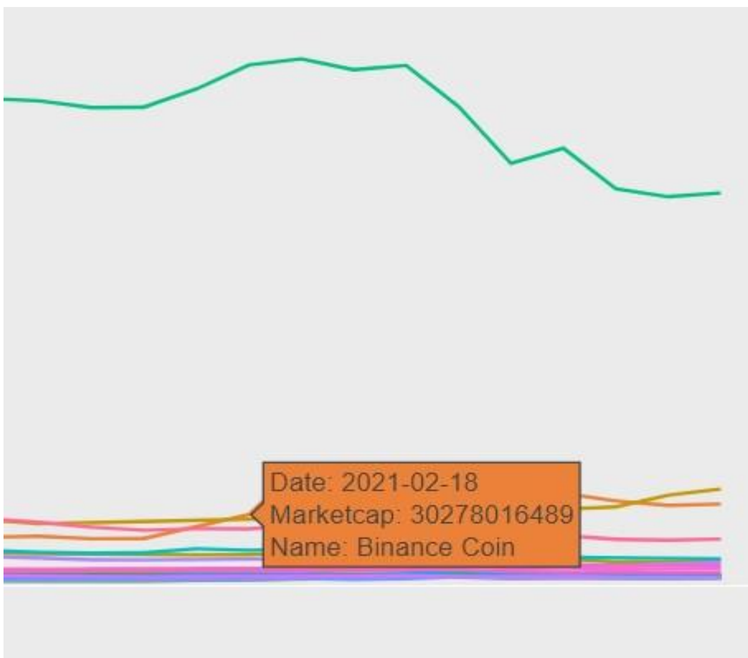
Bitcoin represents the highest average marketcap over the years followed by Ethereum. Average market capitalization of Ripple(XRP) and Polkadot is very close.



Date v/s Market Capitalization



Date v/s Market Capitalization (Cardano)



Date v/s Market Capitalization (Binance Coin)

Inference:
We can see from the following charts that the Market Cap started to increase for Bitcoin followed by Ethereum and then Ripple in, 2017. However, as of February 2021, the Market capitalization for Ripple went down and was exceeded by Cardano and Binance Coin; putting them 3rd and 4th respectively, on the list of top cryptocurrencies with respect to Market Cap.

• **Volatility of Bitcoin, Ethereum, Ripple, Cardano and Binance Coin during crisis period. (Mar'20-Feb'21 – Covid period)**

Volatility of cryptocurrency could be correlated to the real economy, and uncertainty could thus be linked to the stock market. Since the data covers the COVID-19 pandemic and therefore the financial crisis, we checked for volatility of some cryptocurrencies during the pandemic.

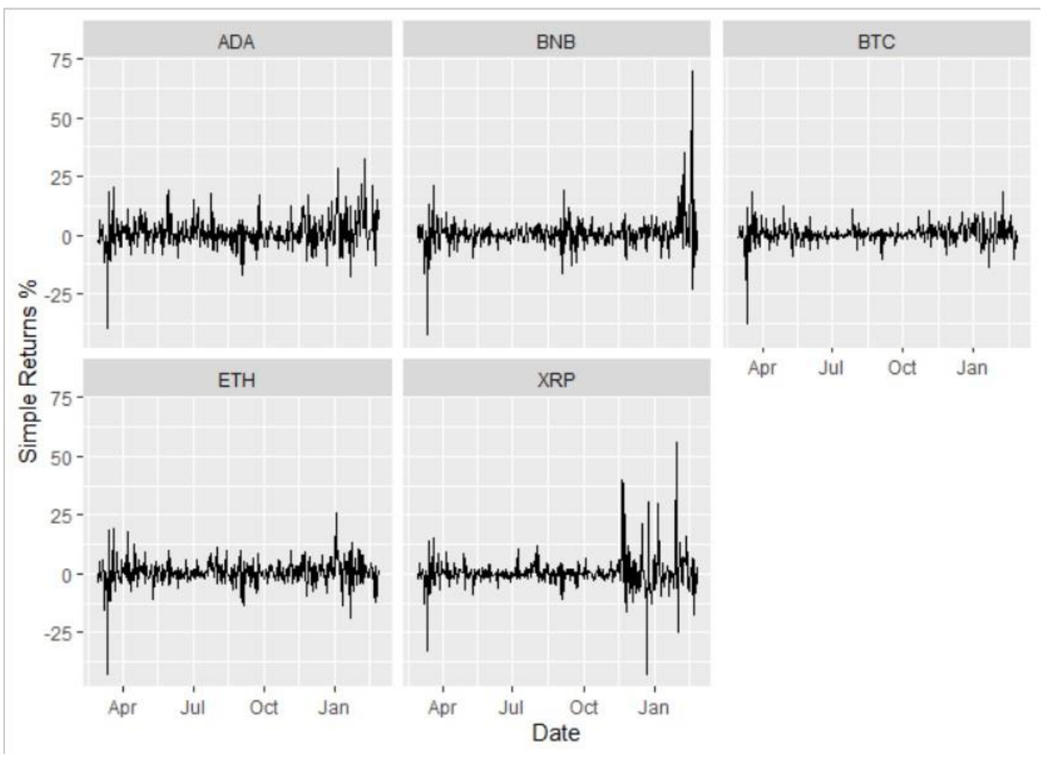
Calculating daily volatility using close price over a period of year from 27th February 2020 to 27th February 2021.

Formula used:

$R_t = \ln(P_t / P_{t-1})$

where R = simple log return

P = close price t = time instant



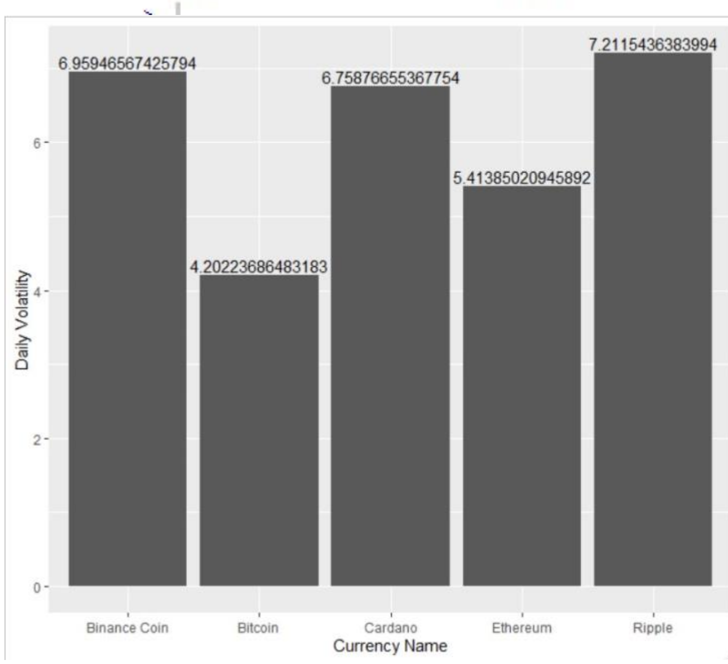
Inference:

According to the above plot, all 5 currencies showed drop in the daily returns in during the month of March which is when the Covid lockdown began. Bitcoin seems to be the least volatile currency in 2020. XRP seems to be show more changes in daily returns after October. Binance Coin seems to show an increase in daily returns after January.

$$\text{Daily Volatility} = \text{Standard Deviation } (R_t) = \sqrt{\frac{\sum |r-r'|^2}{n}}$$

where n = The number of days

```
# A tibble: 5 x 2
  Symbol `Daily Volatility`
  <chr>                <dbl>
1 XRP                   7.21
2 BNB                   6.96
3 ADA                   6.76
4 ETH                   5.41
5 BTC                   4.20
```



Inference:

We can see that Covid-19 pandemic does not have statistically significant effect on the future volatility of the cryptocurrencies. However, all of them showed some increased volatility during the month when lockdown began. Ripple seems to be the most volatile currency with the highest standard deviation (7.21) of returns followed by Binance Coin and Cardano.

- **Cumulative Return over a period of year from 27th February 2020 to 27th February 2021**

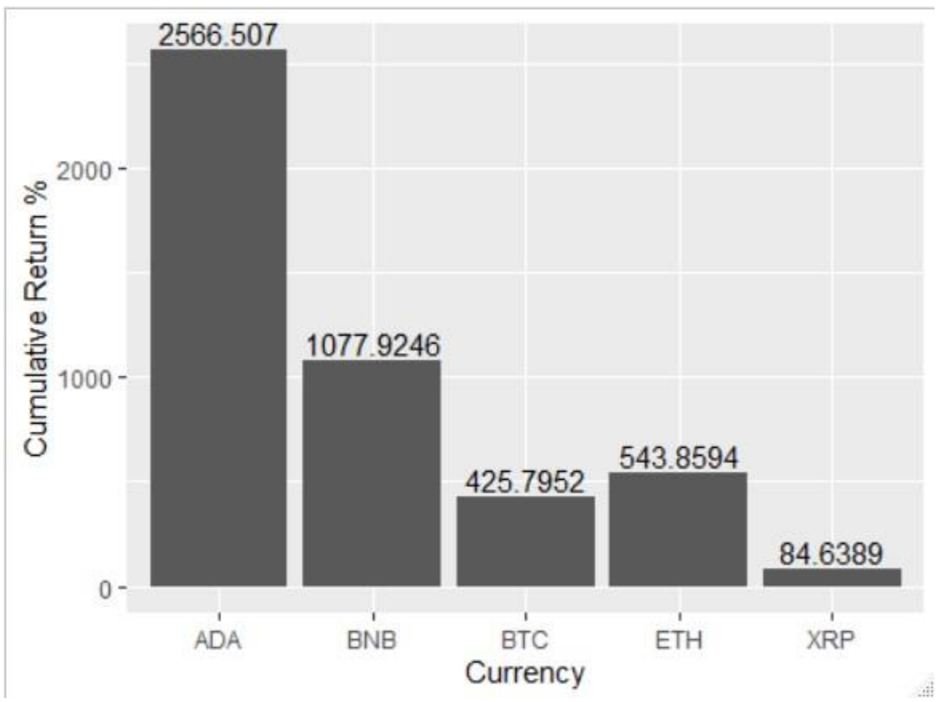
  We calculated the cumulative return on the entire period of 27th Feb 2020 to 27th Feb 2021 with the following formula:

  Cumulative Return Percentage= $[(P_1 - P_0)/ P_0]*100$

  Where P_0 = Initial/Original stock price,

  P_1 = Ending stock price (period end),

  $(P_1 - P_0)/ P_0$ = Rates of change

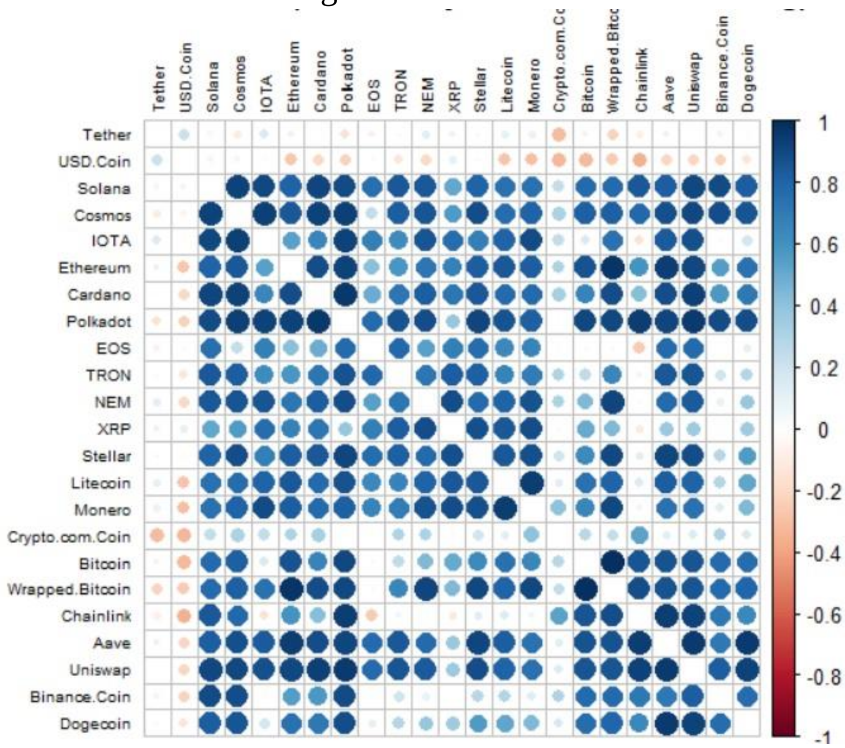


Inference:

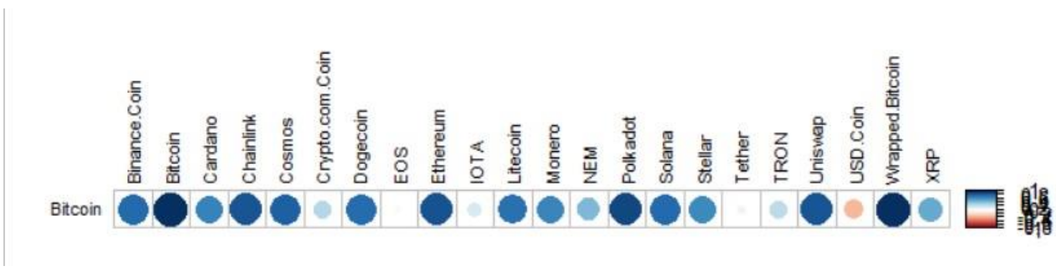
Cardano shows the highest growth in its closing prices with 2566.5% increase followed by Binance Coin with 1078% growth, Bitcoin with 426% growth, Ethereum with 544% growth and Ripple with 85% growth.

- **Hierarchical Clustering**

  Here, correlation between all the 23 currencies is shown by correlation matrix ordered by Hierarchical Clustering.



Correlation between all the currencies



Correlation of Bitcoin with other currencies

Inference:

Top 5 High correlation with Bitcoin

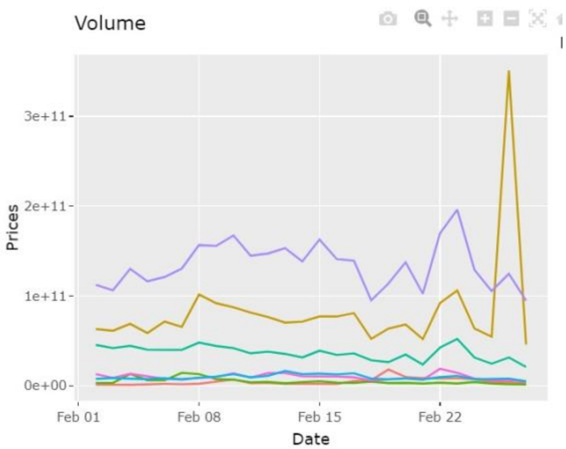
	cor
Wrapped.Bitcoin	0.9999
Polkadot	0.9050
Aave	0.8687
Ethereum	0.8642
Chainlink	0.8540

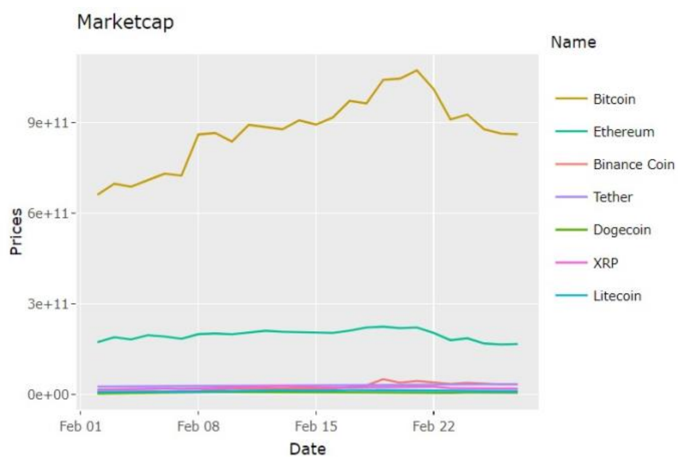
Top 5 Low Correlation with Bitcoin

	cor
USD.Coin	-0.32875
EOS	0.03197
Tether	0.06128
IOTA	0.16835
TRON	0.25048

- **Volume to Marketcap Ratio**

It is natural to believe that the greater a coin's market capitalization, the more value exchanged in any given amount of time. However, it is not always true and following graphs may help in understanding the same.
We compared Volume and Market Cap values from Feb 01, 2021 to Feb 27, 2021





Ranking of currencies as Feb 27

#	w.r.t Volume	w.r.t Market Cap
1	**Tether**	Bitcoin
2	Bitcoin	Ethereum
3	Ethereum	**Binance Coin**
4	**Litecoin**	**Tether**
5	XRP	XRP
6	Dogecoin	**Litecoin**
7	**Binance Coin**	Dogecoin

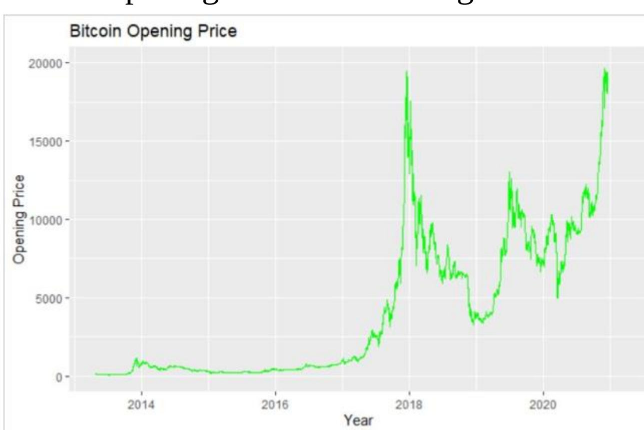
Inference:
We can observe that even though Tether has the highest volume, it ranks 4th with respect to market capitalization at around $3.5 billion. We can say that the coin is undervalued. As opposed to Binance Coin with low volume and still ranking 3rd with respect to Market cap indicating the high value of each coin.

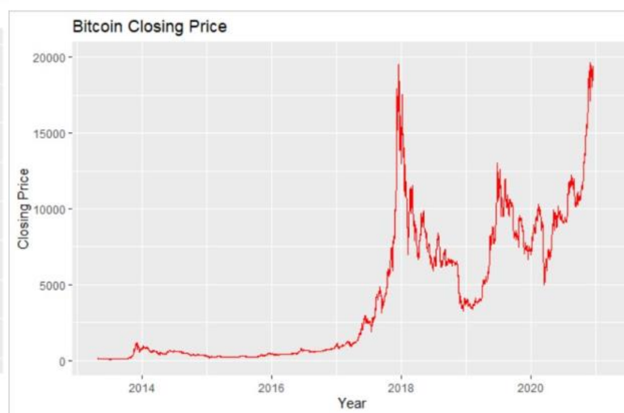
- **Bitcoin Analysis**

    - Summary of Bitcoin

```
> summary(Bitcoin)
      SNO              Name              Symbol               Date                High
 Min.   :   1.0   Length:2862        Length:2862        Min.   :2013-04-29   Min.   :   74.56
 1st Qu.: 716.2   Class :character   Class :character   1st Qu.:2015-04-14   1st Qu.:  426.05
 Median :1431.5   Mode  :character   Mode  :character   Median :2017-03-29   Median : 1197.33
 Mean   :1431.5                                         Mean   :2017-03-29   Mean   : 4974.04
 3rd Qu.:2146.8                                         3rd Qu.:2019-03-14   3rd Qu.: 8138.05
 Max.   :2862.0                                         Max.   :2021-02-27   Max.   :58330.57
      Low                Open              Close              Volume
 Min.   :   65.53   Min.   :   68.5   Min.   :   68.43   Min.   :0.000e+00
 1st Qu.:  415.68   1st Qu.:  421.2   1st Qu.:  420.99   1st Qu.:2.786e+07
 Median : 1164.17   Median : 1180.1   Median : 1182.81   Median :3.302e+08
 Mean   : 4695.10   Mean   : 4836.3   Mean   : 4852.09   Mean   :8.978e+09
 3rd Qu.: 7703.36   3rd Qu.: 7924.6   3rd Qu.: 7926.70   3rd Qu.:1.297e+10
 Max.   :55672.61   Max.   :57532.7   Max.   :57539.94   Max.   :3.510e+11
   Marketcap
 Min.   :7.784e+08
 1st Qu.:5.989e+09
 Median :1.924e+10
 Mean   :8.592e+10
 3rd Qu.:1.388e+11
 Max.   :1.072e+12
> head(Bitcoin,5)
     Name Symbol                Date    High      Low    Open  Close Volume  Marketcap
1 Bitcoin    BTC 2013-04-29 23:59:59 147.488 134.0000 134.444 144.54      0 1603768865
2 Bitcoin    BTC 2013-04-30 23:59:59 146.930 134.0500 144.000 139.00      0 1542813125
3 Bitcoin    BTC 2013-05-01 23:59:59 139.890 107.7200 139.000 116.99      0 1298954594
4 Bitcoin    BTC 2013-05-02 23:59:59 125.600  92.2819 116.380 105.21      0 1168517495
5 Bitcoin    BTC 2013-05-03 23:59:59 108.128  79.1000 106.250  97.75      0 1085995169
```

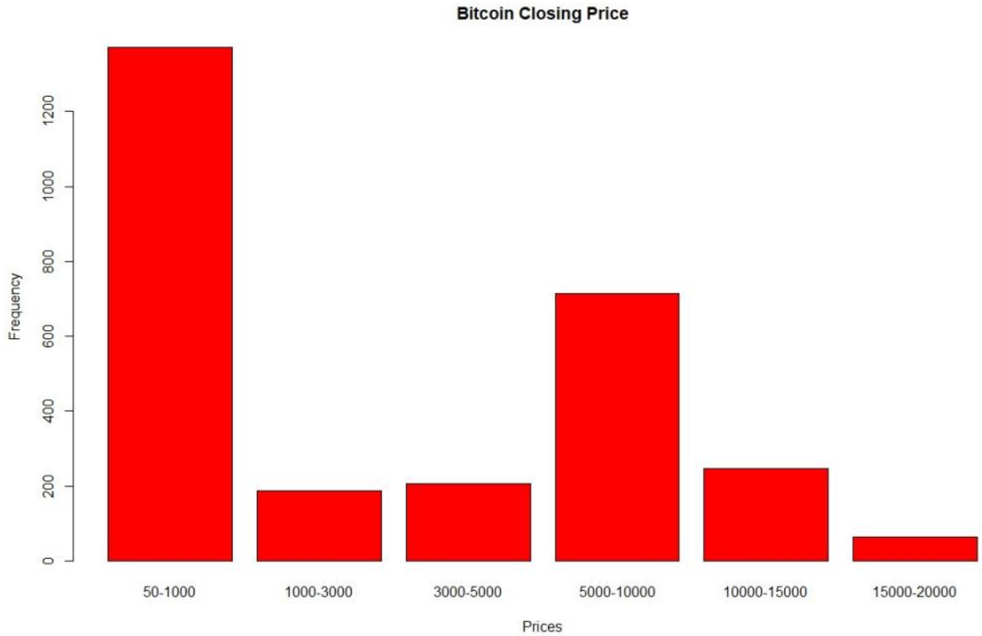
    - Opening Price and Closing Price

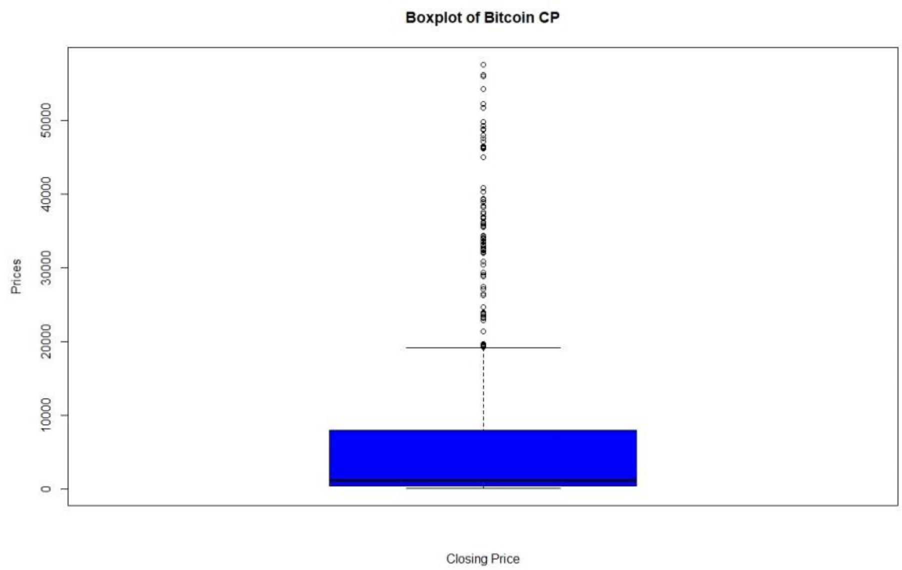




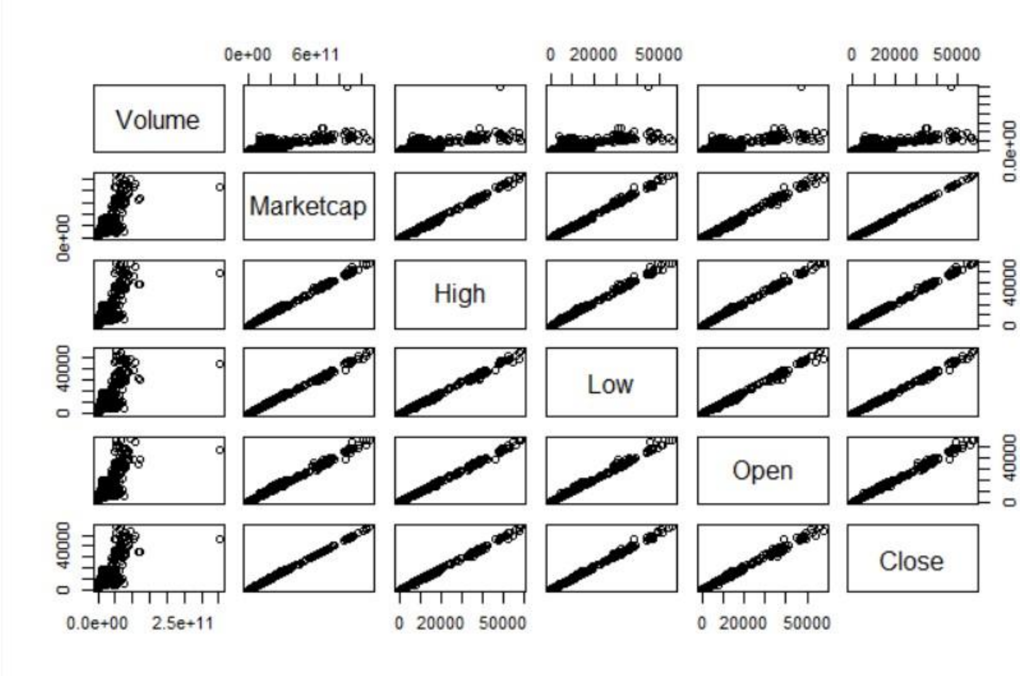
Inference:
By comparing the Opening and Closing Price, we can see that there is not much difference between the two as both the prices open and close nearly at the same range of amount.

- Histogram and Box plot of Bitcoin Closing Price





- **Multiple Linear Regression to study the relationships within the datasets.**



Inference:
We can see that Close price have a strong positive linear correlation with High, Low, Open and Market Cap and can be considered as the most relevant values in predicting the closing price.

The relationship between Market Cap and Volume is not linear and can be studied further to understand the trend.

- **Regression Model**

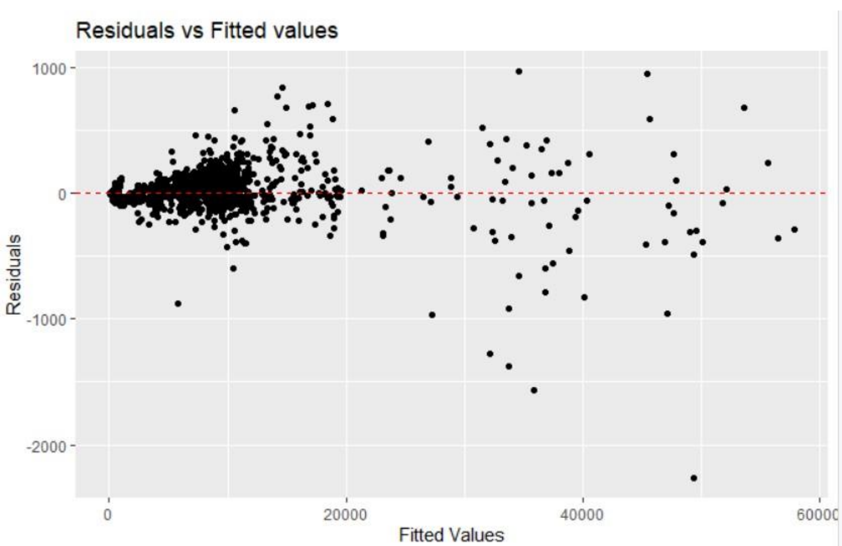
- Forming a multiple regression model for bitcoin since most relationships look linear and assessing the quality of the fit of the model.

```
Call:
lm(formula = Close ~ Open + High + Low + Volume + Marketcap,
    data = Bitcoin)

Residuals:
     Min       1Q   Median       3Q      Max
-2344.76   -43.43   -22.11    36.97  1031.49

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.356e+01  3.447e+00   18.44   <2e-16 ***
Open        -2.982e-01  1.022e-02  -29.18   <2e-16 ***
High         6.321e-01  1.100e-02   57.45   <2e-16 ***
Low          3.418e-01  9.055e-03   37.75   <2e-16 ***
Volume      -3.368e-09  2.675e-10  -12.59   <2e-16 ***
Marketcap    1.760e-08  4.498e-10   39.12   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 130.4 on 2856 degrees of freedom
Multiple R-squared:  0.9997,    Adjusted R-squared:  0.9997
F-statistic: 1.637e+06 on 5 and 2856 DF,  p-value: < 2.2e-16
```

Inference:

We can consider this as a good model because of the following reasons: ○ The p-values of all the predictors are smaller than 0.05 and are considered to be significant.

- The R-squared value is 0.9997 indicating that the model can explain 99.9% of the total variability.
- The sum of the residuals is as low as possible.

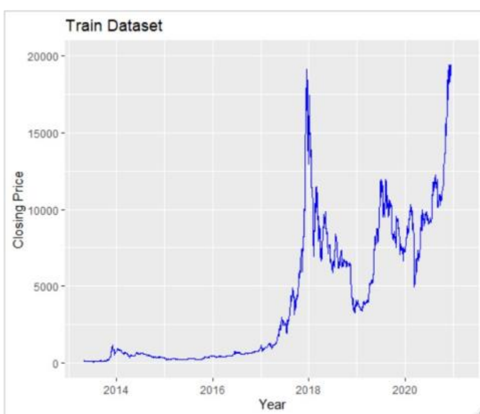


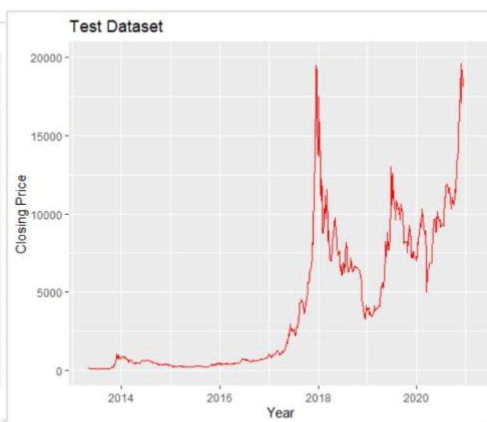
We can observe some heteroskedasticity in the above plot.

- **Prediction model**

Main aim is to predict the closing price of Bitcoin.
There needs to be a training data and validation data if we want to create classification models. Here, we divided the data in such a way that 70% of our data is reserved for testing and 30% data for validation.

	Rows	Columns
Train	2003	9
Test	859	9

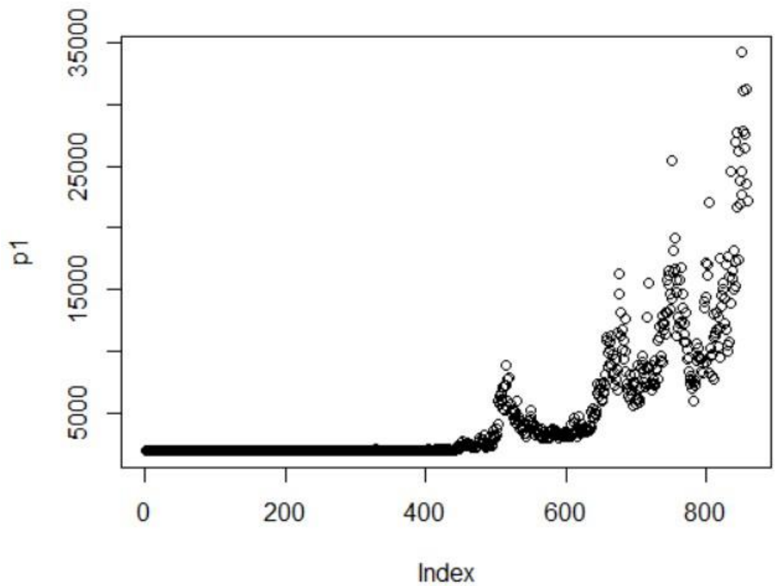




RESULTS

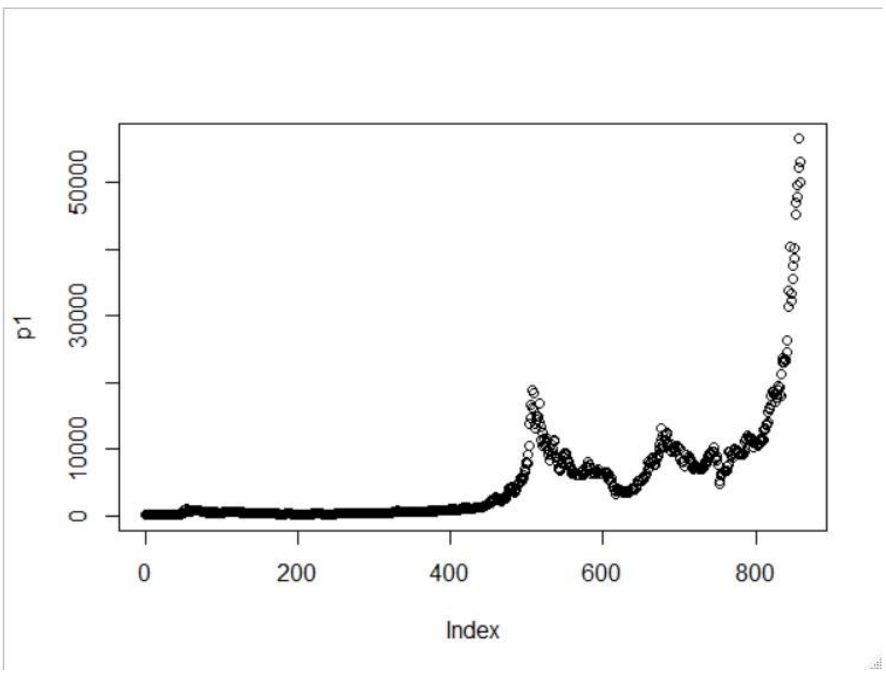
- **Only Volume as the predictor:**

```
Call:
lm(formula = Close ~ Volume, data = train)

Residuals:
   Min     1Q Median     3Q    Max
-67102  -1667  -1311   1470  39111

Coefficients:
                     Estimate       Std. Error t value              Pr(>|t|)
(Intercept) 1941.564775069526  104.918483357306   18.50 <0.0000000000000002
Volume         0.000000317694    0.000000005493   57.83 <0.0000000000000002

Residual standard error: 4157 on 2001 degrees of freedom
Multiple R-squared:  0.6257,    Adjusted R-squared:  0.6255
F-statistic:  3345 on 1 and 2001 DF,  p-value: < 0.00000000000000022

> p1 <- predict(model1,test)
> head(p1)
       2        5        6       18       20       21
1941.565 1941.565 1941.565 1941.565 1941.565 1941.565
> error1 <- p1 - test[["Close"]]
> sqrt(mean(error1^2))
[1] 3982.537
> plot(p1)
```



- **Volume, Open, Low, High, Marketcap as the predictor:**

```
Call:
lm(formula = Close ~ Volume + High + Low + Open + Marketcap,
    data = train)

Residuals:
     Min       1Q   Median       3Q      Max
-2231.28   -44.72   -21.91    36.18  1221.07

Coefficients:
                      Estimate        Std. Error t value            Pr(>|t|)
(Intercept) 65.2215897387429  4.2155708356701  15.472 <0.0000000000000002
Volume      -0.0000000029859  0.0000000003068  -9.731 <0.0000000000000002
High         0.5955213664098  0.0138843315174  42.892 <0.0000000000000002
Low          0.3742436601638  0.0113888538185  32.861 <0.0000000000000002
Open        -0.3039379295200  0.0132963549025 -22.859 <0.0000000000000002
Marketcap    0.0000000181937  0.0000000005600  32.486 <0.0000000000000002

Residual standard error: 132.9 on 1997 degrees of freedom
Multiple R-squared:  0.9996,    Adjusted R-squared:  0.9996
F-statistic: 1.045e+06 on 5 and 1997 DF,  p-value: < 0.00000000000000022

> p2 <- predict(model2,test)
> head(p2)
       2        5        6       18       20       21
187.1914 146.6817 161.2557 167.3400 173.1439 171.4660
> error2 <- p2 - test[["Close"]]
> sqrt(mean(error2^2))
[1] 127.4255
> plot(p2)
```



The plot is more defined with the increase in significant predictors.

CONCLUSION

In conclusion for our assessment, we would like to outline the techniques used for understanding different performance metrics for cryptocurrencies as well as go over the inferences of each study. The closing price and average market capitalization studies were both performed using the ggplot package and from these studies we can infer that Bitcoin leads the chart in both cases.

Further we calculated volatility of each cryptocurrency by computing the standard deviations for daily returns for each cryptocurrency. According to this study, Ripple showed the greatest volatility over a one-year period.

We also calculated cumulative returns for the "Top 5" cryptocurrencies and Cardano had the greatest cumulative returns over a one-year period. We also tried to understand Bitcoin correlation with all other cryptos, and it showed positive correlation with Polkadot, Aave and Ethereum.

We also performed a study to understand factors that affect Bitcoin prices. High, Low, Open and Martketcap have strong positive correlation with closing prices of Bitcoin. We also built a prediction model to understand bitcoin price behavior, and using Volume, Open, High, Low and Marketcap as the predictors gives us a price trend that closely follows the current price trend of Bitcoin. However, the plot becomes strongly defined with the increase in significant predictors. Cryptocurrency prices can be affected by a large number of factors, and we believe it is crucial to understand and leverage these factors as well before cryptocurrencies can be deemed as a reliable investment opportunity.

REFERENCE

- Automated Bitcoin Trading via Machine Learning Algorithms – • https://www.investopedia.com/terms/c/cryptocurrency.asp
- https://towardsdatascience.com/a-foray-into-time-series-forecasting-usingethereum-closing-prices-5ca69dbbd76a
- https://www.toptal.com/finance/market-research-analysts/cryptocurrency-market
- https://www.kaggle.com/archit9406/analyzing-bitcoin-trends/comments
- https://www.quora.com/How-important-is-volume-and-market-cap-when-tradingcrypto-altcoins-I-have-read-if-they-are-low-you-shouldnt-trade-it-but-how-low-istoo-low
- https://www.datacamp.com/community/tutorials/linear-regression-R
- https://datascienceplus.com/time-series-analysis-using-arima-model-in-r/