BITCOIN TRADING STRATEGY PREDICTOR

PRAKHER GUPTA (UIN: 677991441)

Background

Bitcoin is a digital cryptocurrency and payment system that is entirely decentralized, meaning it is based on peer-to-peer transactions with no bureaucratic oversight. Transactions and liquidity within the network are instead based on cryptography. The system first emerged formally in 2009 and is currently a thriving open-source community and payment network. Based on the uniqueness of Bitcoin's payment protocol and its growing adoption, the Bitcoin ecosystem is gaining lots of attention from businesses, consumers, and investors alike. Namely, for the ecosystem to thrive, we need to replicate financial services and products that currently exist in our traditional, fiat currency world and make them available and custom-tailored to Bitcoin, as well as other emerging cryptocurrencies.

In this Project, we will try to predict the value of Daily values of Open, High, Low, Close to help the trader plan the trades accordingly for the day. As part of the future scope the same thing will be replicated for Hours and Minutes to increase the accuracy of the trades done. We will try to build a statistical model around the current data and fuse it to forecast the values for 2018.

Prior Work:

Bitcoin has been recently very popular, drawing a lot of attention towards the use of machine learning to predict the values. Hence, we can find bits and pieces of work done to fit a model with some accuracy. Please find the details in references.

Data Collection and Exploration

Every minute data is available on Kaggle for all the crypto currencies out of which we will be concentrating on btceUSD_1-min_data_2012-01-01_to_2017-05-31 which is the data in USD from 2012-01-01 to 2017-05-31. Below are the data columns:

Timestamp – Unix time

Open - Open Value

High - Highest bid value

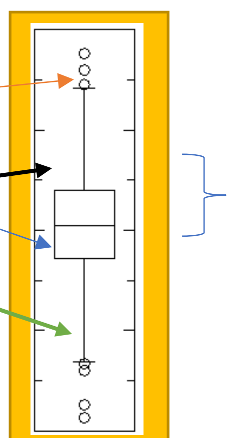
Low - Lowest Bid value

Close - Closed value

Volume_.BTC. - Volume of Bitcoin traded

Volume_.Currency. – Volumn of currency traded

Weighted_Price - Weighted price of Bitcoin



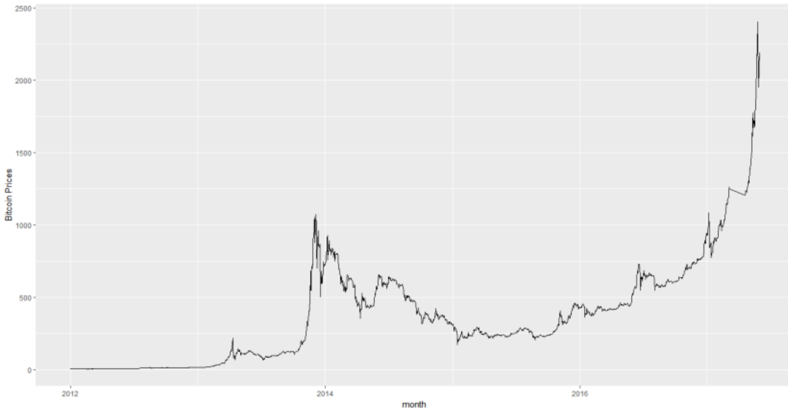
Note: The Open and Close value can vary as per the 'Red' and the 'Green' Candle/Box Plot

Following steps were taken as part of data cleansing:

1. Convert UNIX Time into Date Time Format
2. Omit Null values
3. Convert data into Daily Frequency Data

Note:

We want a "Trading Strategy predictor", which means our model should identify the 'Open', 'High', 'Low', 'Close' values for the day which could help a trader realize any anticipated spikes in the market. In this report we will be concentrating to fit a model which could forecast the 'Open' value which can later be replicated for High, Low and Close and a box plot can be obtained.



Observation

- We observe an exponentially increasing trend for bitcoin with few outliers which we will get rid of

What we will try

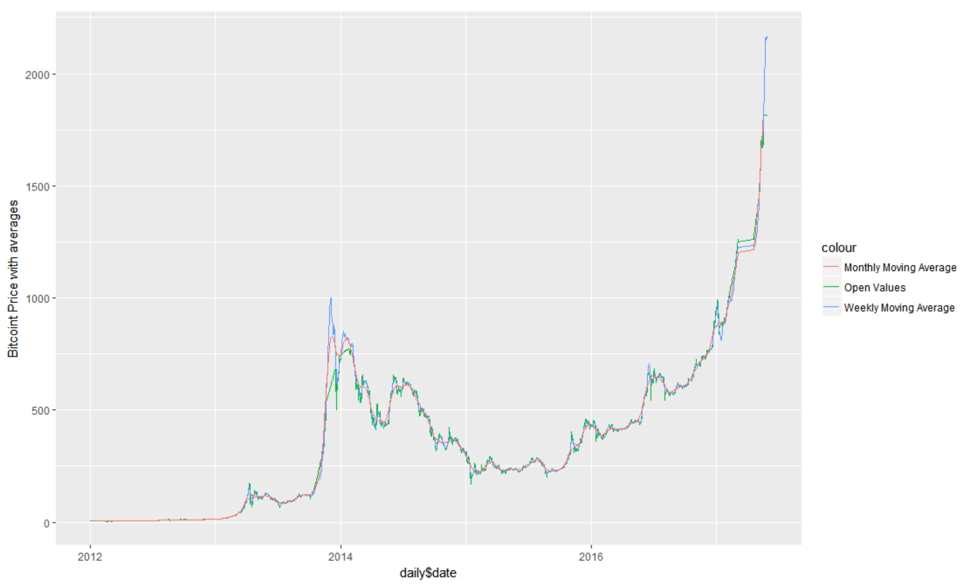
- We will try to predict the value for Bitcoin beyond this data

Model Selection

Since, we need a model which can analyze the past values and understand the seasonal variations, we will be going ahead with ARIMA model. For this we first need the moving averages for the Open values for every day.

1. Adding Moving Averages

ARIMA is combination of Auto Regression and Moving averages. Here, we calculate the weekly and monthly moving averages. The plot of these moving averages along with the daily open value is plotted as shown in the below chart.

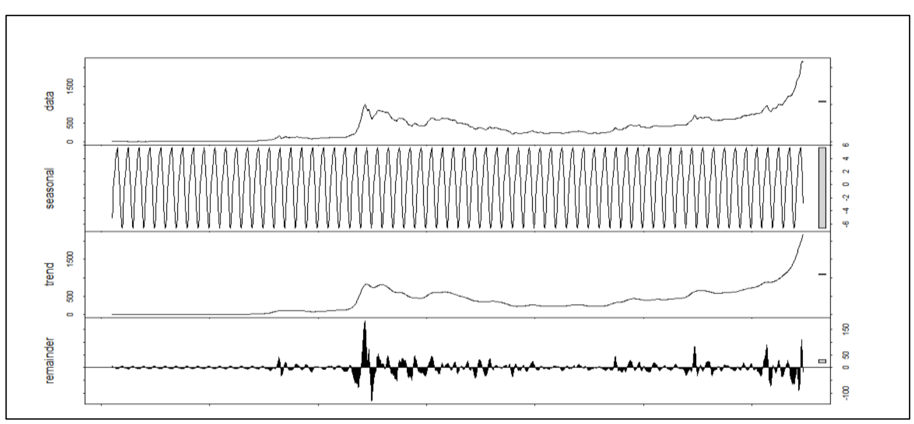


Plot of Daily open values against weekly and monthly moving averages

2. Decomposing Data

Now, we try to decompose our data. Decomposing data includes following components as shown in the plot:

- Seasonal Component
- Trend Component
- Cycle Component
- Residual or error



3. Stationary Check

ARIMA model needs the series to be stationary i.e. the mean, Variance and Autocovariance should be independent on time. Hence, we will be using Augmented Dickey-Fuller (ADF) test to formally test the stationarity.

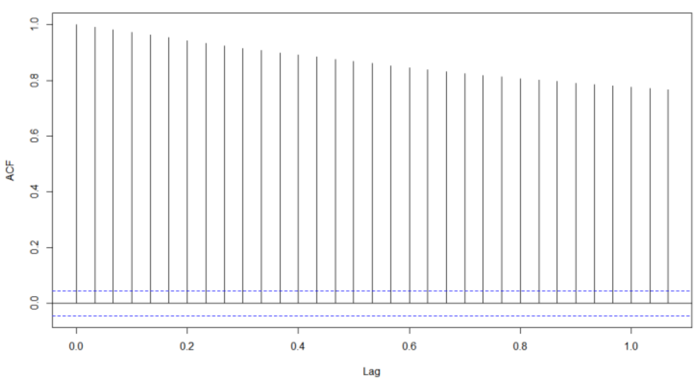
<u>Test Statistics:</u>

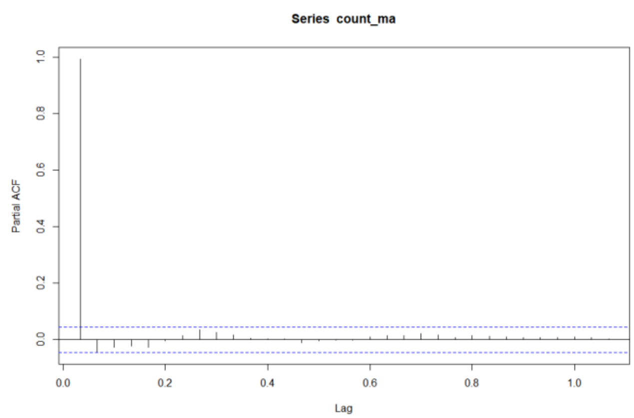
```
> adf.test(count_ma, alternative = "stationary")

        Augmented Dickey-Fuller Test

data:  count_ma
Dickey-Fuller = 0.13567, Lag order = 12, p-value = 0.99
alternative hypothesis: stationary
```

Since, the P-value is huge,we cannot reject the null hypothesis, i.e. weekly moving average is not stationary. Hence, we will check for the stationarity for the difference of weekly moving average. Below are the plots for ACF and PACF for weekly moving averages which strengthens the belief.





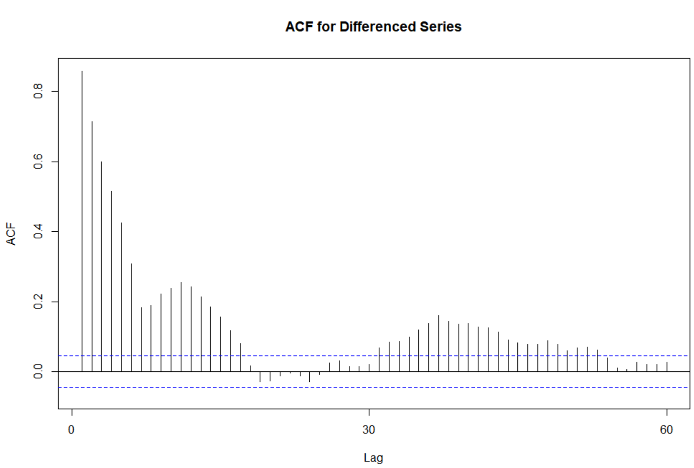
<u>Test Statistics for Difference of weekly moving average:</u>

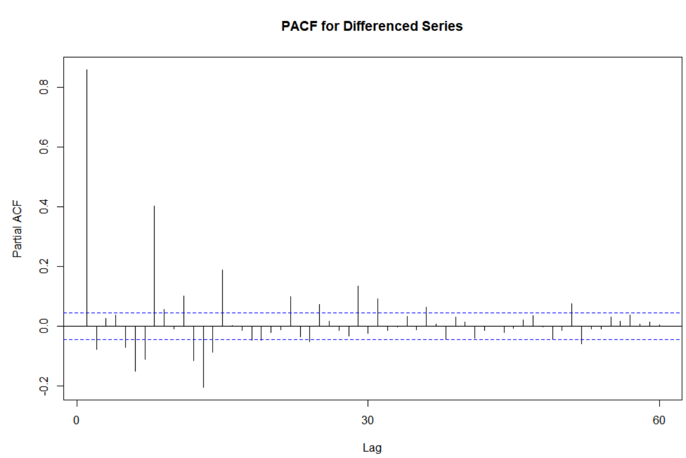
```
> adf.test(count_d1, alternative = "stationary")

        Augmented Dickey-Fuller Test

data:  count_d1
Dickey-Fuller = -9.0799, Lag order = 12, p-value = 0.01
alternative hypothesis: stationary
```

Since, the P-value is less than the alpha values,we reject the null hypothesis, i.e. the difference of weekly moving average is stationary. Hence, we will go ahead with the difference for building the ARIMA model. Below are the plots for ACF and PACF for difference of weekly moving averages which strengthens the belief.



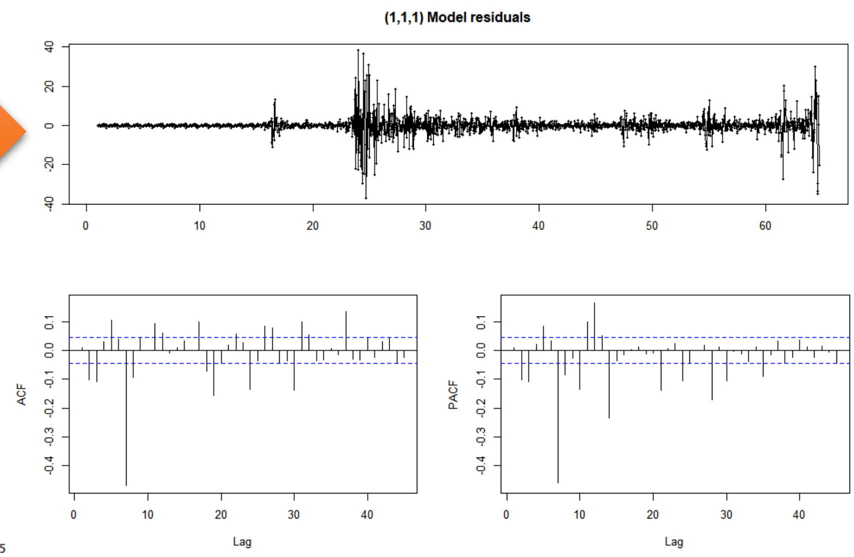


4. ARIMA Model

We try to fit an ARIMA model with c (1,1,1) for which below are the statistics

T-Statistics for the ARIMA model

Residual plot for ARIMA with c(1,1,1)




```
> summary(try)

Call:
arima(x = deseasonal_cnt, order = c(1, 1, 1))

Coefficients:
         ar1     ma1
      0.8346  0.1006
s.e.  0.0146  0.0266

sigma^2 estimated as 17.73:  log likelihood = -5462.62,  aic = 10931.25
```

Since, we observe that we have an outlying ACF value for lag 7, we can try an ARIMA with c(1,1,7) which gives following statistics.

```
> fit2

Call:
arima(x = count_d1, order = c(1, 1, 7))

Coefficients:
          ar1      ma1     ma2     ma3     ma4     ma5     ma6      ma7
      -0.0019  -0.0103  0.0082  0.0117  0.0140  0.0146  0.0347  -0.9864
s.e.   0.0236   0.0064  0.0080  0.0066  0.0072  0.0076  0.0065   0.0083

sigma^2 estimated as 10.12:  log likelihood = -4940.24,  aic = 9898.47
```

Hence, we will go ahead with the second model and will try to predict the value of Open.

5. Forecasting

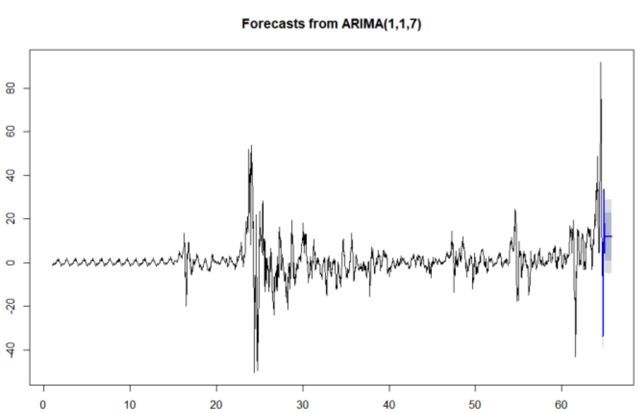


Figure (v). a

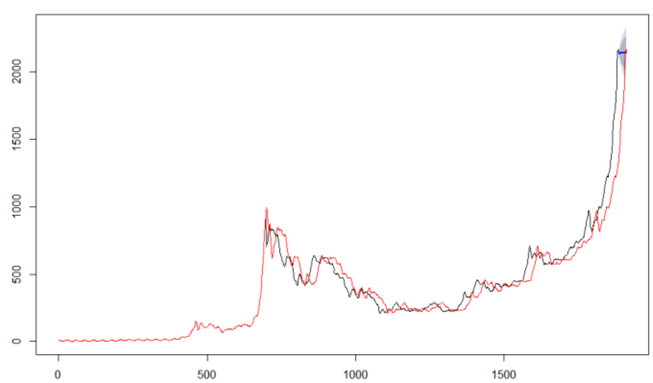
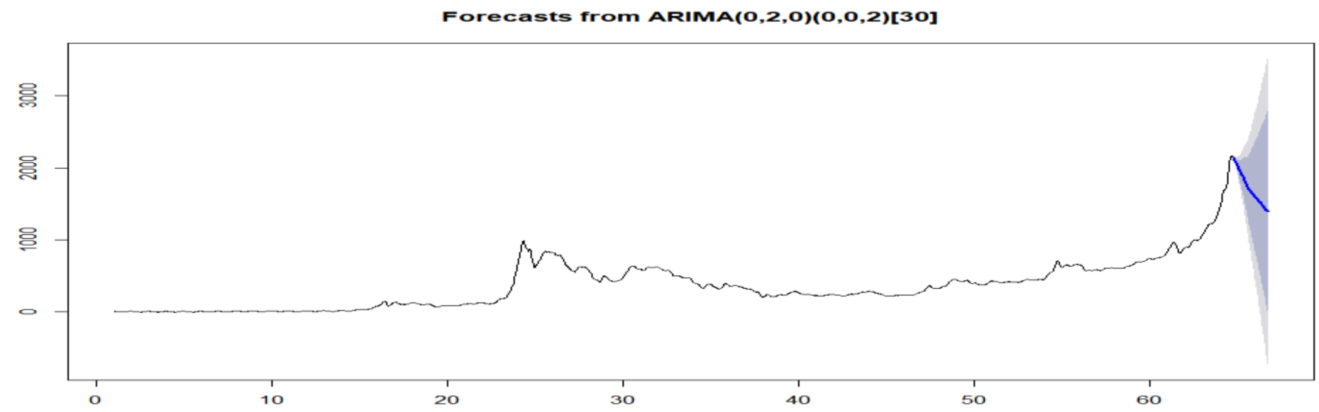


Figure (v). b

Figure (v). a depicts the forecast in ‘Blue’ line which is in line with the past values. Comparing it with the given values in Figure (v). b shows the trend which is again in line with the current values. The figures also depict the confidence bound of 80% and 95% confidence in both the plots. But these plots do not consider the seasonal variations because of which they become flat quickly.

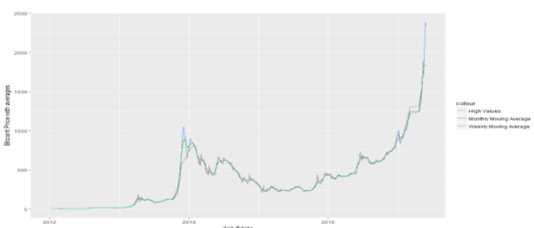
Hence, after considering the seasonal variation we observed the below plot where the Blue line depicts the Forecasted line which was what we wanted.

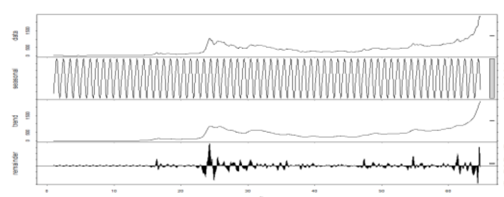


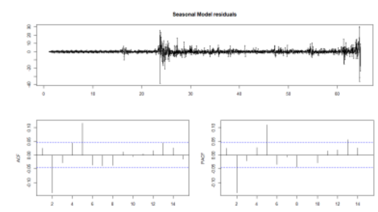
Replicating model for High, Low, Close values

1) Low Values

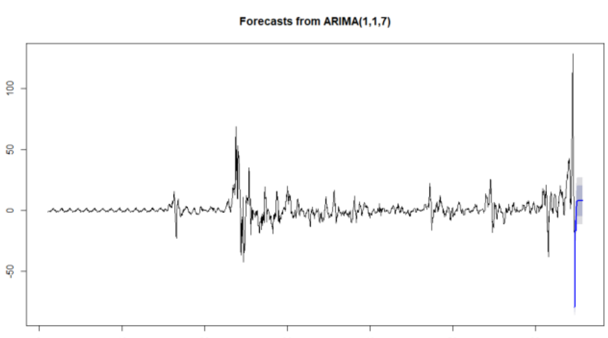
Applying the similar approach to obtain the prediction for High Values, we observe below test results and prediction charts.

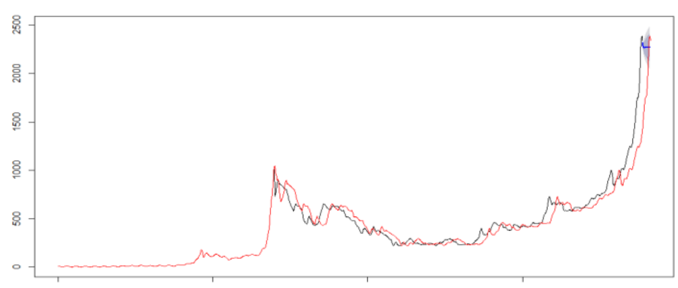


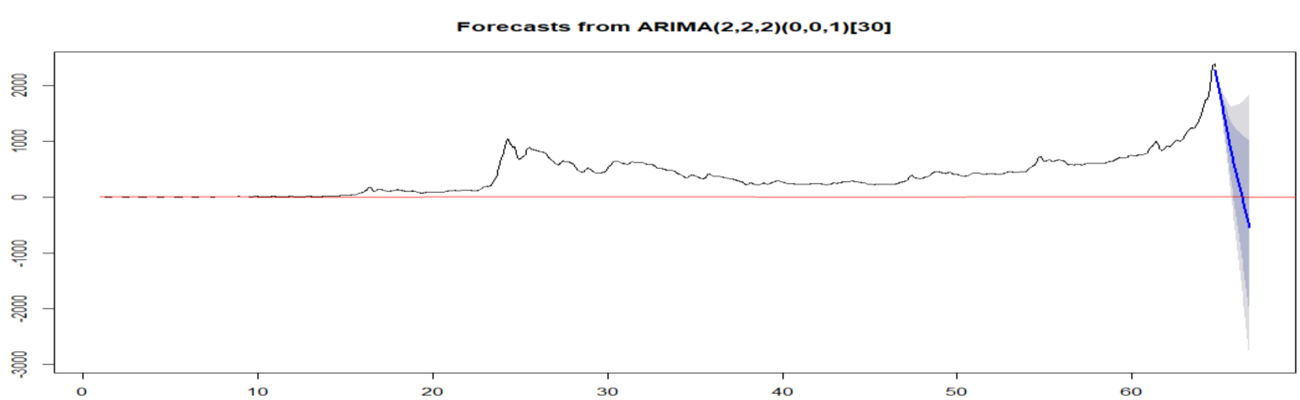




Prediction Charts for ***<u>LOW</u>*** *Values*

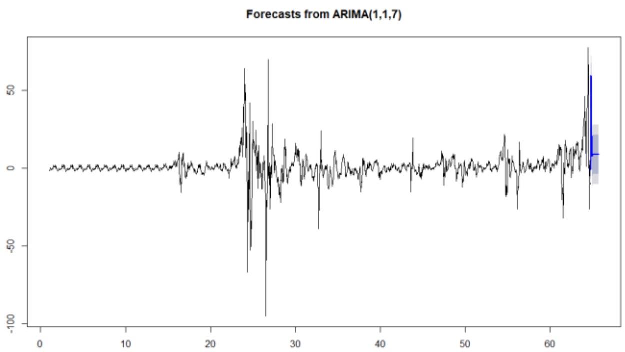


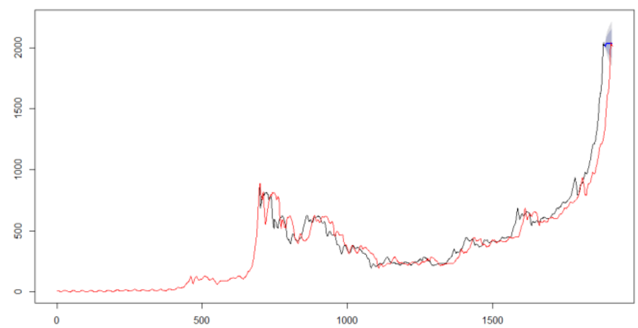


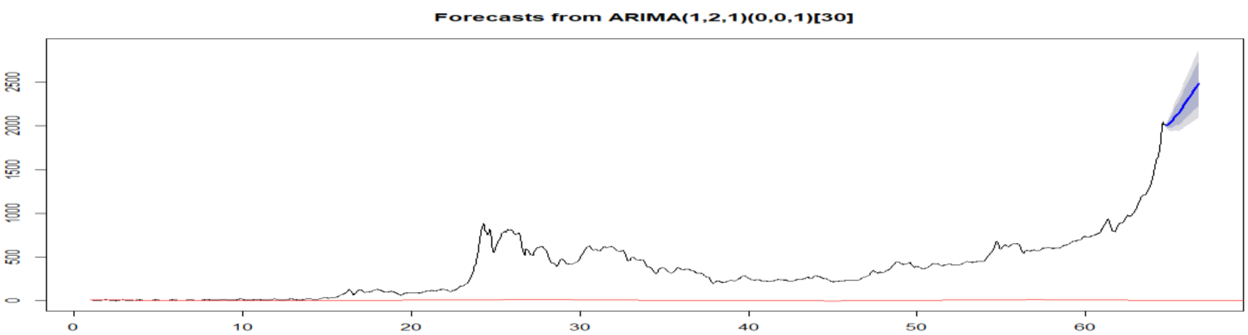


2) High values

While similar test results were obtained below are the Prediction charts for Low Values

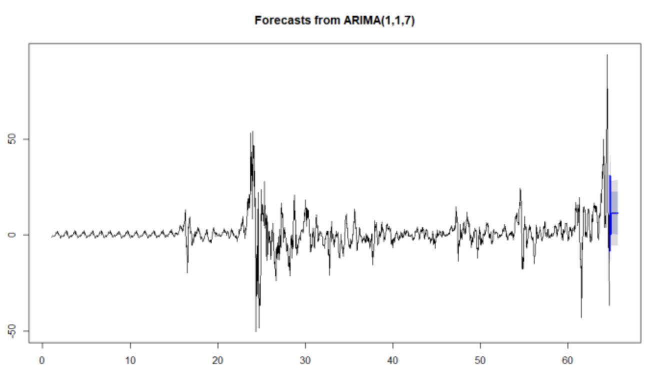


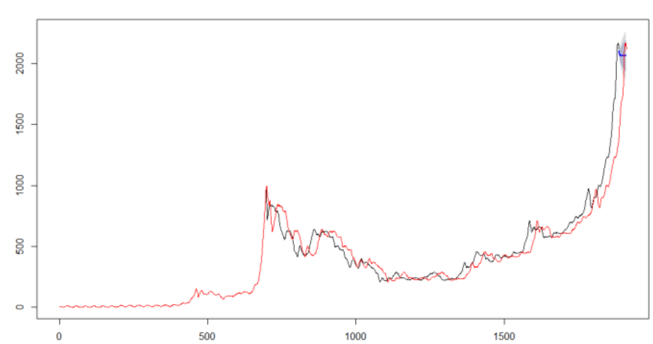


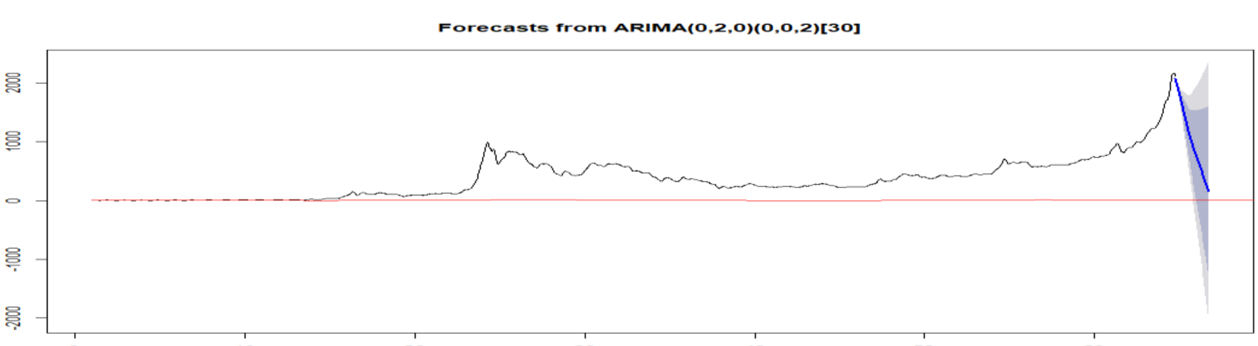


3) Close Values

Similarly, for close values, below are the prediction plots



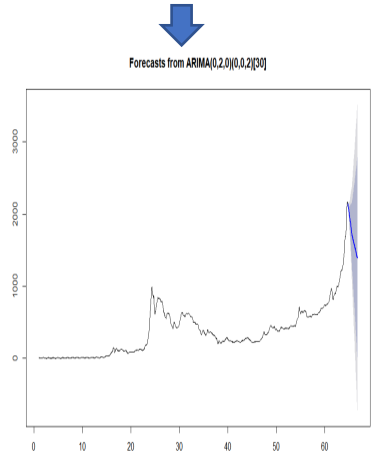




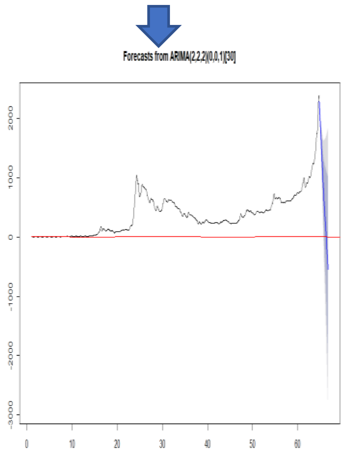
Trading Strategy

We predicted the Open, High, Low, Close values for bitcoin using ARIMA model here and now all a trader needs to do is look at these charts for the day and make his trades accordingly. Thank us later!

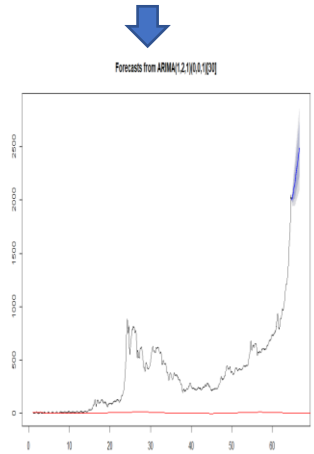
OPEN



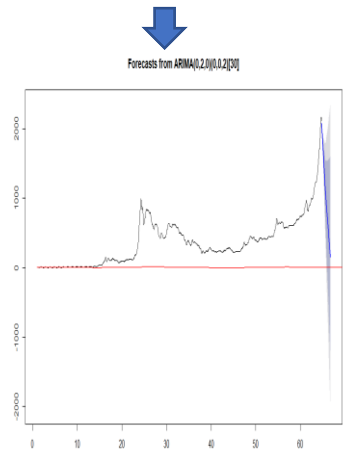
LOW



HIGH



CLOSE



Future Scope

Now that we have predicted values for Open, High, Low, Close daily, we would try to break it down into hours and minutes of the day and will try to predict the best trades along with the profit values which a trader can make to maximize his profit.

References

1) www.kaggle.com (Data Source)
2) MIT Report on Machine learning prediction
   http://ai2-s2-pdfs.s3.amazonaws.com/e065/3631b4a476abf5276a264f6bbff40b132061.pdf
3) https://www.kaggle.com/myonin/bitcoin-price-prediction-by-arima/notebook