Stock Market Prediction with Time Series Analysis

Project Update Part- 2

Hypothesis:

1. For forecasting the stocks of various companies ARIMA model performs better than the other more basic models like Naïve model or the Moving Average model.
2. The Stocks of the four big companies – Google, Amazon, Microsoft, Apple, IBM will have increasing trend over the period of 10 years.
3. The highest open price will be highest for Google among all the companies.
4. The Arima Model which I will use by choosing with the appropriate (p, d, q) values after analyzing the data will perform better in forecasting than the model using the function auto.arima().
5. The Google stocks over the 10 years are more stable as compared to Amazon, Microsoft, IBM and Apple. I will be analyzing the various components of Time series decomposition for
6. Time Series Forecasting with Neural Networks will provide better results compared to the forecasting with other traditional models like ARIMA model.

EDA:

The dataset which I have used for my analysis is the "DJIA 30 Stock Time Series" dataset from Kaggle (https://www.kaggle.com/szrlee/stock-time-series-20050101-to-20171231)
This data contains past year's stock data from 30 companies that posted on Dow Jones Industrial Average.
Data has 93612 observations in total with 7 variables which are –

- **Date - in format: yy-mm-dd**
- **Open - price of the stock at market open (this is NYSE data so all in USD)**
- **High - Highest price reached in the day**
- **Low Close - Lowest price reached in the day**
- **Volume - Number of shares traded**
- **Name - the stock's ticker name**

Structure of the dataset:

```
## 'data.frame':    93612 obs. of  7 variables:
##  $ Date  : Factor w/ 3020 levels "2006-01-03","2006-01-04",..: 1 2 3 4 5 6 7 8 9 10 ...
##  $ Open  : num  77.8 79.5 78.4 78.6 78.5 ...
##  $ High  : num  79.3 79.5 78.7 78.9 79.8 ...
##  $ Low   : num  77.2 78.2 77.6 77.6 78.5 ...
##  $ Close : num  79.1 78.7 78 78.6 79 ...
##  $ Volume: int  3117200 2558000 2529500 2479500 1845600 1919900 1911900 2121100 1925300 2073400 ...
##  $ Name  : Factor w/ 31 levels "AABA","AAPL",..: 20 20 20 20 20 20 20 20 20 20 ...
```

Summary of the dataset:

```
##       Date           Open            High            Low
##  2006-01-03:   31   Min.   :   6.75   Min.   :   7.17   Min.   :   0.00
##  2006-01-04:   31   1st Qu.:  33.95   1st Qu.:  34.29   1st Qu.:  33.60
##  2006-01-05:   31   Median :  60.04   Median :  60.63   Median :  59.49
##  2006-01-06:   31   Mean   :  85.62   Mean   :  86.39   Mean   :  84.84
##  2006-01-09:   31   3rd Qu.:  94.00   3rd Qu.:  94.74   3rd Qu.:  93.25
##  2006-01-10:   31   Max.   :1204.88   Max.   :1213.41   Max.   :1191.15
##  (Other)   :93426   NA's   :25        NA's   :10        NA's   :20
##      Close           Volume            Name
##  Min.   :   6.66   Min.   :        0   AXP    :  3020
##  1st Qu.:  33.96   1st Qu.:  5040180   BA     :  3020
##  Median :  60.05   Median :  9701142   CAT    :  3020
##  Mean   :  85.64   Mean   : 20156670   CVX    :  3020
##  3rd Qu.:  94.01   3rd Qu.: 20752222   DIS    :  3020
##  Max.   :1195.83   Max.   :843264044   GE     :  3020
##                                        (Other):75492
```

Observations:

1. The plot below represents the probability density of each column – Open, Close, High, Low. We can see that for Open and High, the values lie mostly between 1-100 and then decreases gradually till 400.We can see that for Close and Low, the values lie mostly between 1-75 and then decreases gradually till 300-350.
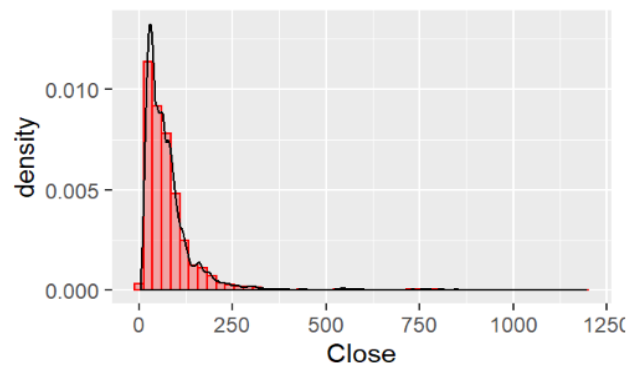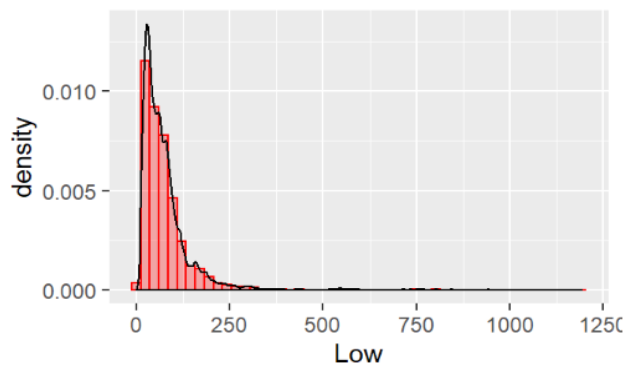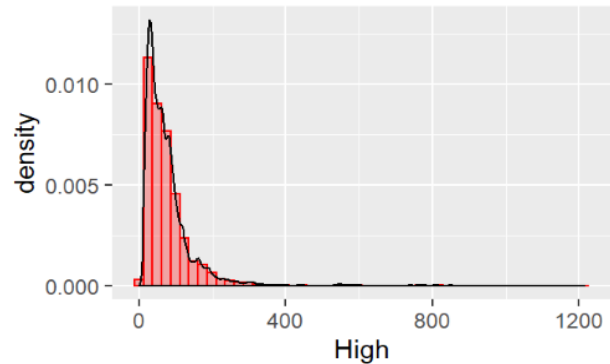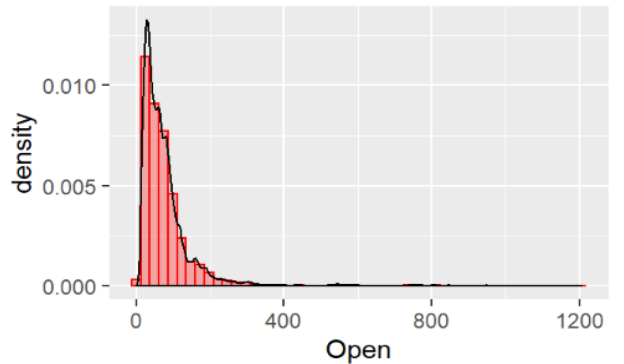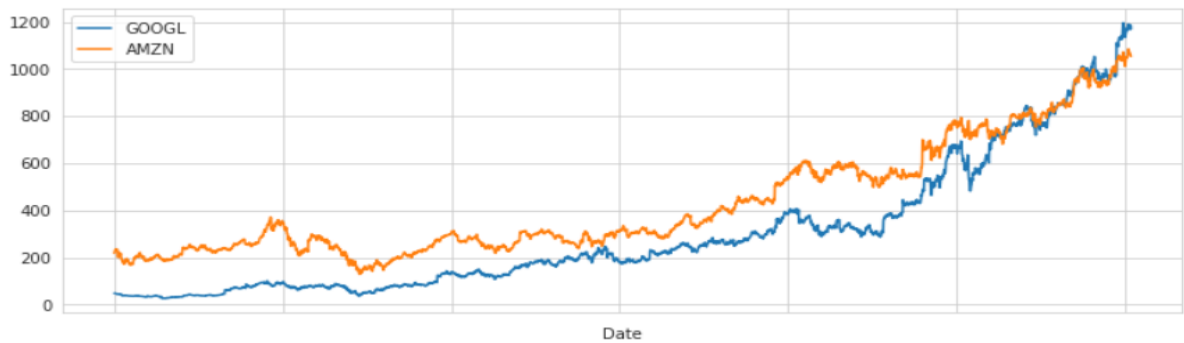

2. The highest open price of Amazon gradually went up over years and finally exceeded Google's open price.
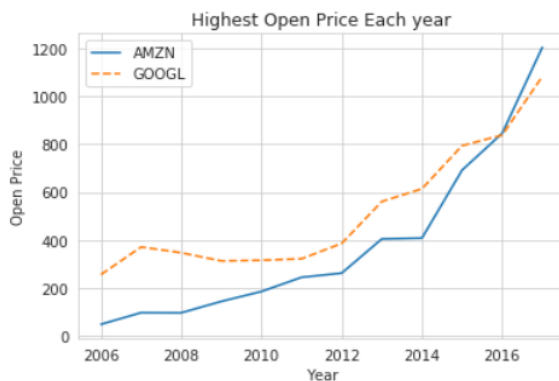
3. The lowest volume of Google over year is always lower than Amazon especially in 2009.
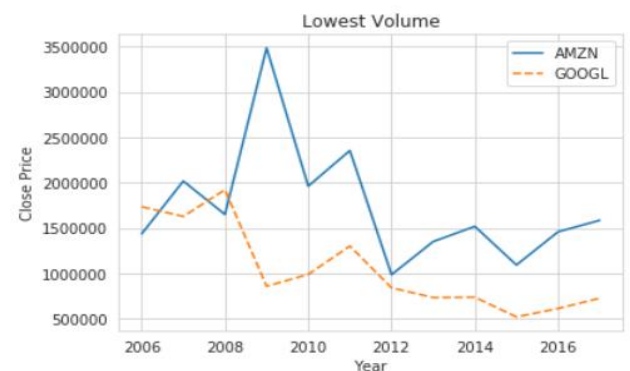
4. The close price of Amazon is more stable around the 30 days average of the stock close price than Googles from 2010 to 2011.


5. I have plotted four charts- Closing Stock with time, Opening stock with time, Decomposition of additive time series and Decomposition of multiplicative time series of Amazon, Google and Microsoft.
   a. We can see that the closing stocks of Amazon increases more steeply than that of Google and Microsoft.
   b. The closing and opening stocks of Microsoft are fluctuating in almost same range till 2012and then is increasing steeply till 2016.
   c. For Amazon, both the opening and closing values are increasing very steeply after 2015
   d. The trend is upward increasing for all the four companies and there is seasonality in all four as well.
   e. The residual for Amazon and Google has almost no pattern but for Microsoft there seems to be some pattern, so we need to analyze it in more depth.
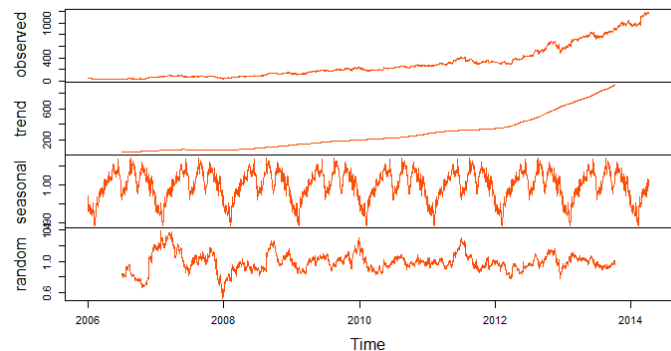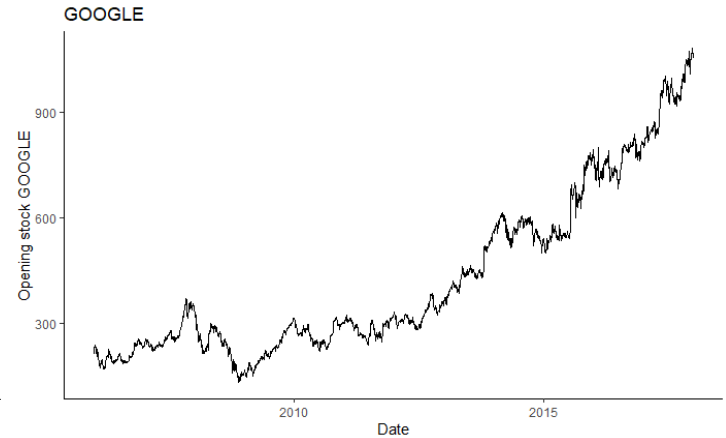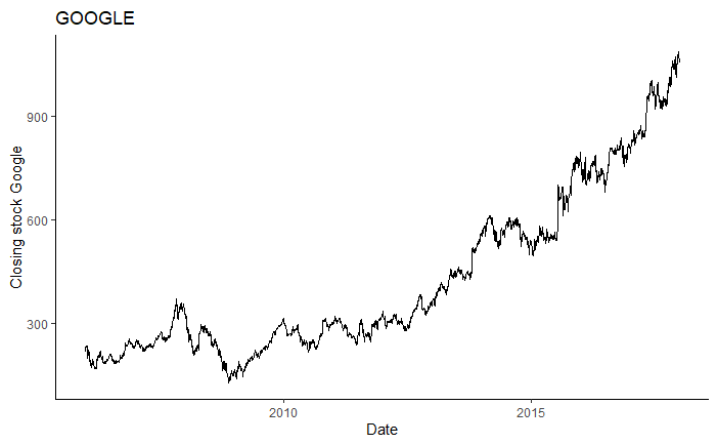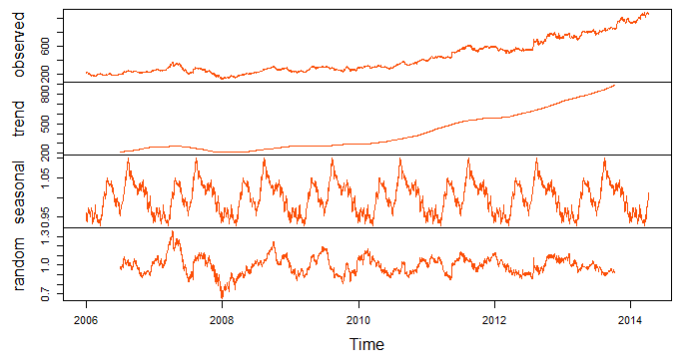
AMAZON
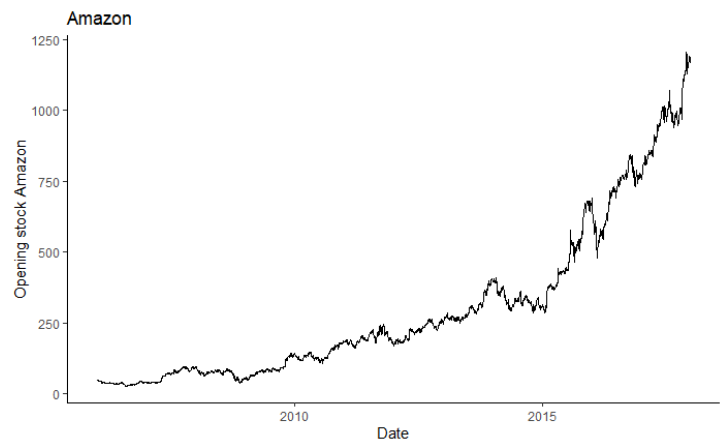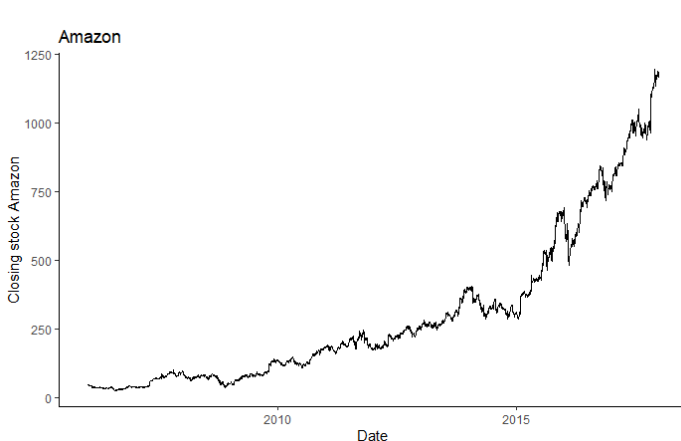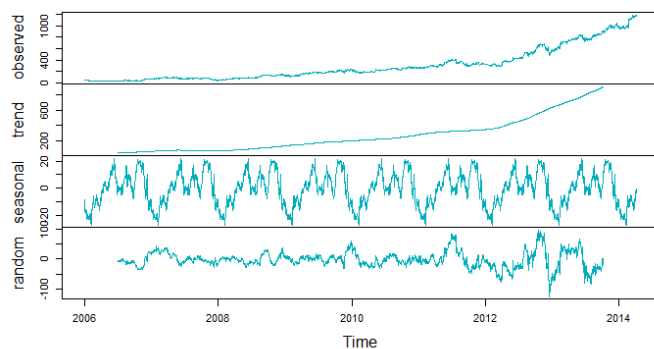
GOOGLE

GOOGLE

GOOGLE

Decomposition of additive time series
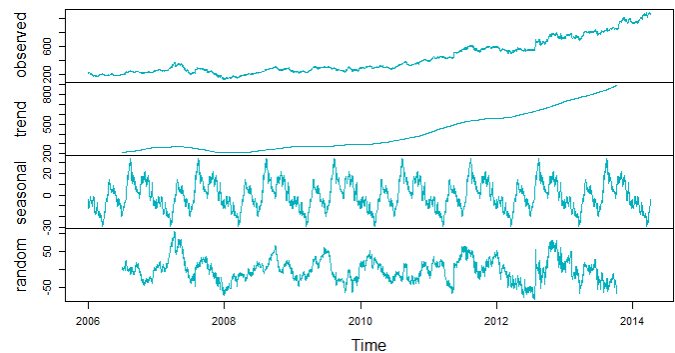
Decomposition of multiplicative time series

MICROSOFT

Microsoft

Microsoft

Decomposition of additive time series

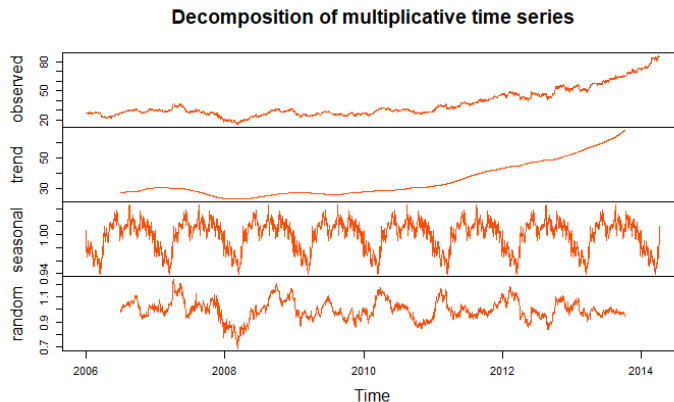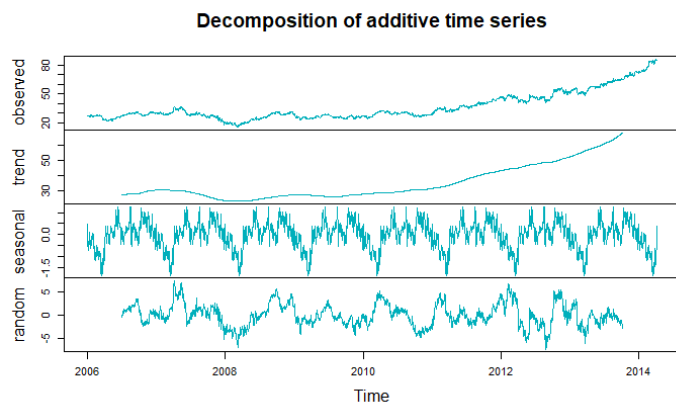**Decomposition of multiplicative time series**

6. Following is the time series plot of Google with IBM, Apple, Goldman Sachs and Boeing. We can see that Google High values are significantly greater than the rest of the company's stock and also it increases very steeply as compared to others.

   Note: The graph is more dynamic, you can move it and see the values for all the years and months in R but here in pdf I couldn't add it in the same way.
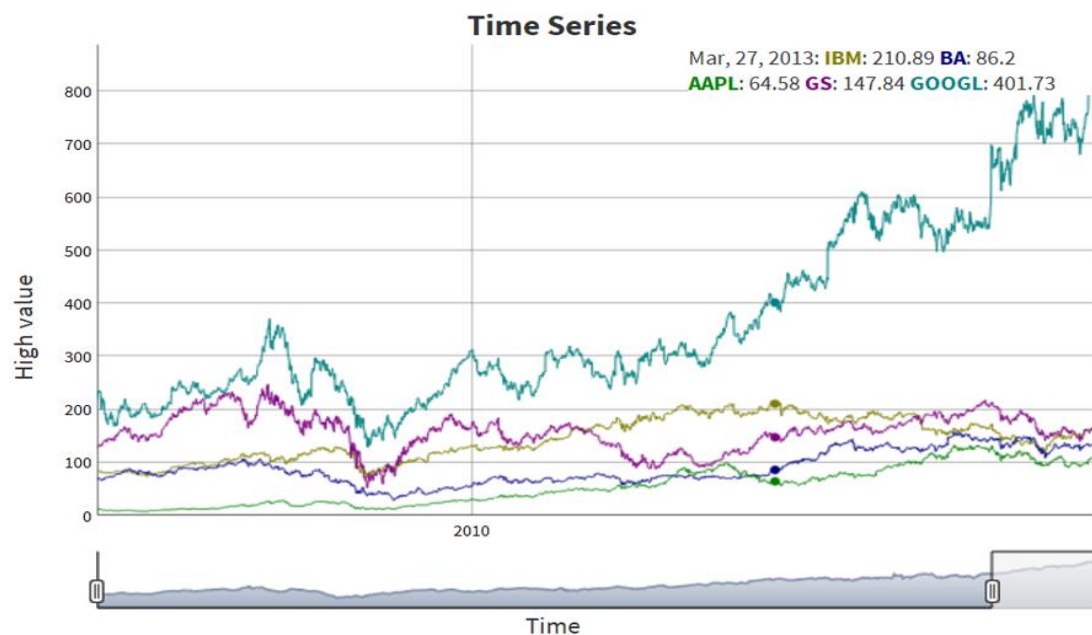
Summary of preliminary analysis

I performed the various models for forecast of Google, Amazon, Microsoft and Apple using various time series models including Holt Winter's, ETS, ETS with Box-Cox transformation, Naïve, Auto- Arima and Arima model and till now I got best results with the ARIMA model including manual chosen p, d and q values. (I am not attaching the results here as the report will extend a lot and also, I am experimenting with a few things and so the results might change).

What additional analysis you are planning to conduct?

Once I will choose the best model, I will forecast and compare different values to answer the hypothesis.

Also, I am planning to conduct same analysis and forecasting with the Neural Networks (RNN and LSTM) and will be comparing the results obtained by the above models.