MSCI 718 Final Project

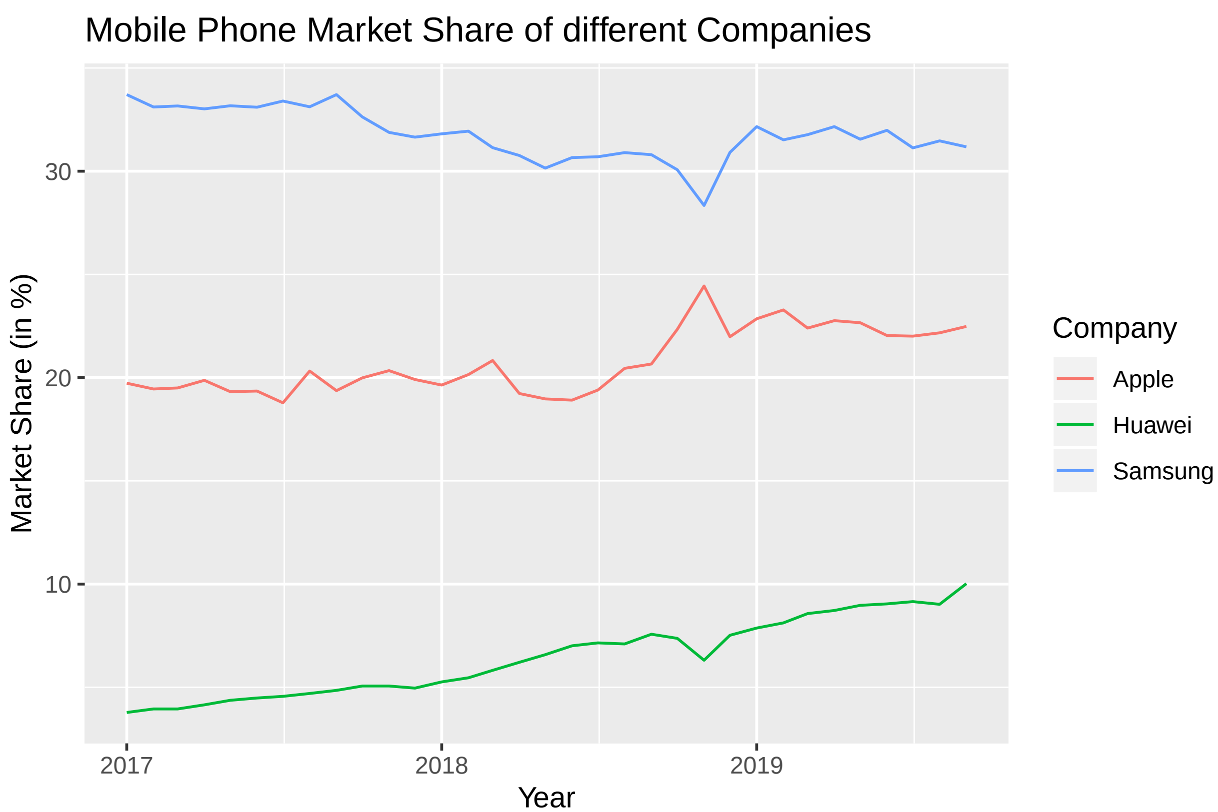
Karan Kohli and Rishabh Karwayun

April 2020

Whole World In Your Hand

According to the Oxford English Dictionary, one of the earliest uses of the word *"mobile"* was in association with a Latin phrase *"mobile vulgus"* meaning *"excitable crowd"*. Today, mobile phones live up to these origins. In the contemporary world, mobile phones are not just a rich man's accessory. It is transforming the way the world leads their day to day lives. There are estimates that by 2021 around 45% of the world's population is going to own a smartphone[1]. This means that mobile business is one of the most lucrative and competitive one with numerous vendors trying to grab a bigger share in the market. We became interested in analyzing the market share of companies and how they are projected to vary in the coming few months.

The data considered[2] includes the yearly worldwide market share of different manufacturers. Samsung is the leading vendor worldwide having the highest mobile phone market share in recent times with Apple in second place and Huwaei in the third position.



Fig(1)

Last year, Samsung was the leader in the mobile market with a market share of 31.37% at the end of the year. Apple being at second number with 24.79% and Huwaei coming in at 3rd place with 9.95%. It would be interesting to know how these numbers of market share for these 3 big companies might change in the coming few months.

Exploring The Dataset

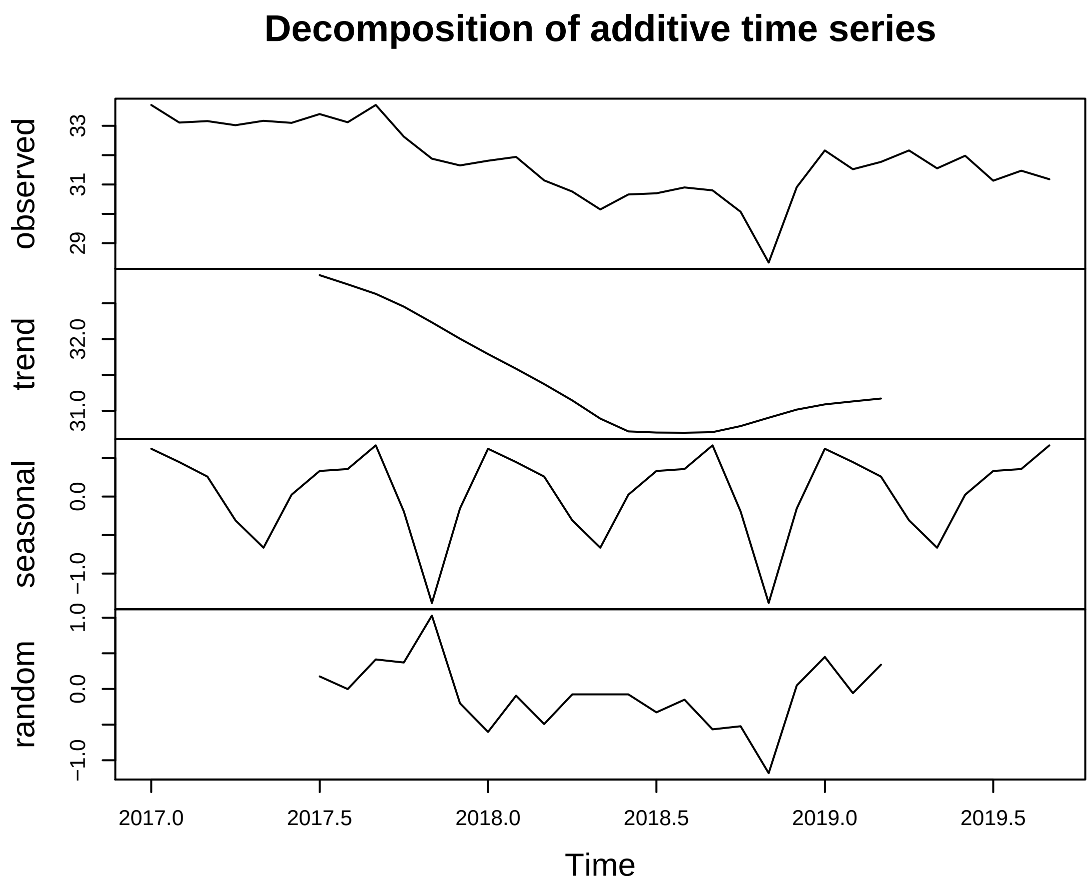
In our final data, we have 33 observations which are the market shares of 3 companies namely Samsung, Apple and Huawei. Samsung has been the global leader in the mobile phone market share with a mean of 31.8% since 2017. Samsung's market share for mobile phones saw a big dip in November of 2011 coming down to 28.34% which is the minimum share Samsung has gone since 2017. Apple with an average of 20.78% has been as a strong 2nd place in the race for market share for mobile phones with a maximum share being 24.44% at the same time when Samsung's share dipped. This could have been due to Apple's launch of iPhone models XS, XS Max and XR, which were extremely successful. Huawei has consistently stayed 3rd in the global leaders for mobile phone market share, with an average of 6.4% share since 2017 with a standard deviation of 1.9%. There are no missing or NA values in the dataset and we found no outliers worth discarding in the data (Appendix 1). Summary of the dataset and descriptive statistics can be found in Appendix 2.

```
##          min  max range median mean std.dev
## Samsung 28.3 33.7   5.4   31.8 31.8     1.2
## Apple   18.8 24.4   5.7   20.3 20.8     1.5
## Huawei   3.8 10.0   6.2    6.3  6.4     1.9
```

We will be predicting the values of market share for the top 3 global leaders in the mobile phone market namely Samsung, Apple, Huwaei with the values from January 2016 to September 2019 as our training data and values from October 2019 to March 2020 as our test data. We will be coming out with a model with the best accuracy for each of the companies and then give final predictions for the next 6 months. We will be looking in detail on the process for Samsung.

Predicting Samsung's Market Share For Next Six Months

We will start with the analysis of Samsung's mobile phone market share. Decomposing the time series we can that there is a seasonal trend in the series from Fig(2). We will keep this in mind while making our model. This seasonal trend is possible as the market share is expected to change due to the launch of new models both by Samsung and by their competitors. These releases are mostly periodic. There is also a downward trend in the series with some random or unexplainable parts of the series.



Fig(2)

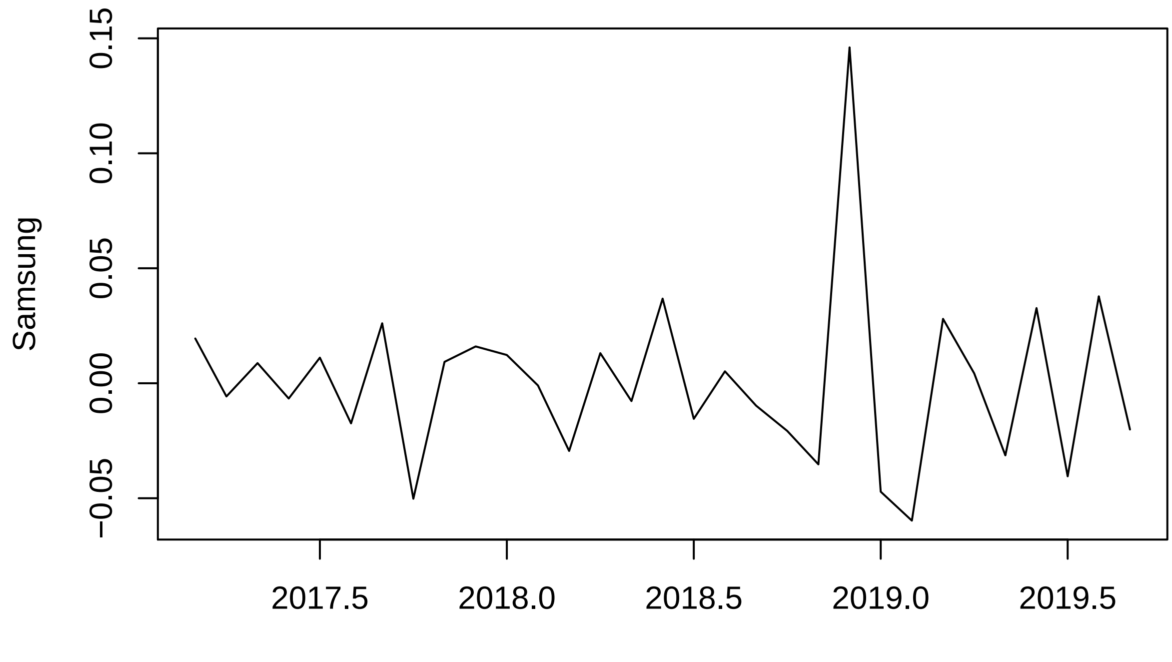
Stationarity

Primary assumption of ARIMA model is Stationarity. A time series is stationary when the mean and variance are constant over time[4]. There are tests which can determine whether a time series is stationary or not. We will be using the Augmented Dicky-Fuller Test to test the stationarity of the series which has the null hypothesis that a unit root is present in a time series sample. The alternative hypothesis is different depending on which version of the test is used, but is usually stationarity[5].

Looking at the test results (p-value = 0.8224), we can say that data does not pass the stationarity test. We will apply log transformation to the data and test the stationarity again. But, we see that the series is still not stationary (p-value = 0.8135). The results of the two Dickey-Fuller tests can be seen in Appendix 3.

The next step to make the data stationary is by differencing the series. When we run the test with a difference of d = 2 on the log-transformed data, we get a p-value less than 0.01 which means that we can reject the null hypothesis and proceed with the alternate hypothesis (that the data is stationary). Now, we have a stationary time series at our hands which can be seen in Fig(3).

```
##
##  Augmented Dickey-Fuller Test
##
## data:  ts_data_diff_samsung
## Dickey-Fuller = -4.4034, Lag order = 3, p-value = 0.01
## alternative hypothesis: stationary
```



Fig(3)

Model

We move on to make our ARIMA model. We use auto.arima() function to get an initial model with default seasonal argument set to TRUE, so that seasonality of the series is also considered in the model and then tune the model with graphs of Autocorrelation and Lag (acf and pacf).

```
## Series: log_ts_data_samsung
## ARIMA(1,2,4)
##
## Coefficients:
##           ar1      ma1      ma2      ma3     ma4
##       -0.8978  -0.0349  -1.2189  -0.1339  0.4592
## s.e.   0.1147   0.2162   0.2164   0.2174  0.2041
##
## sigma^2 estimated as 0.0006307:  log likelihood=70.95
## AIC=-129.9   AICc=-126.4   BIC=-121.3
```

The initial model given by auto.arima() is an ARIMA(1,2,4). We see that neither of the p and q values in the ARIMA(p,d,q) model is 0. Hence, the acf and pacf graphs won't give a conclusive value of p and q in this case[3]. Hence we move on with the model that auto.arima() gives us. From the ACF and PACF graphs of the residuals of the model (Appendix 5), we see that there are no spikes in any lags, hence we can move on to predicting the values saying the model seems decent enough.

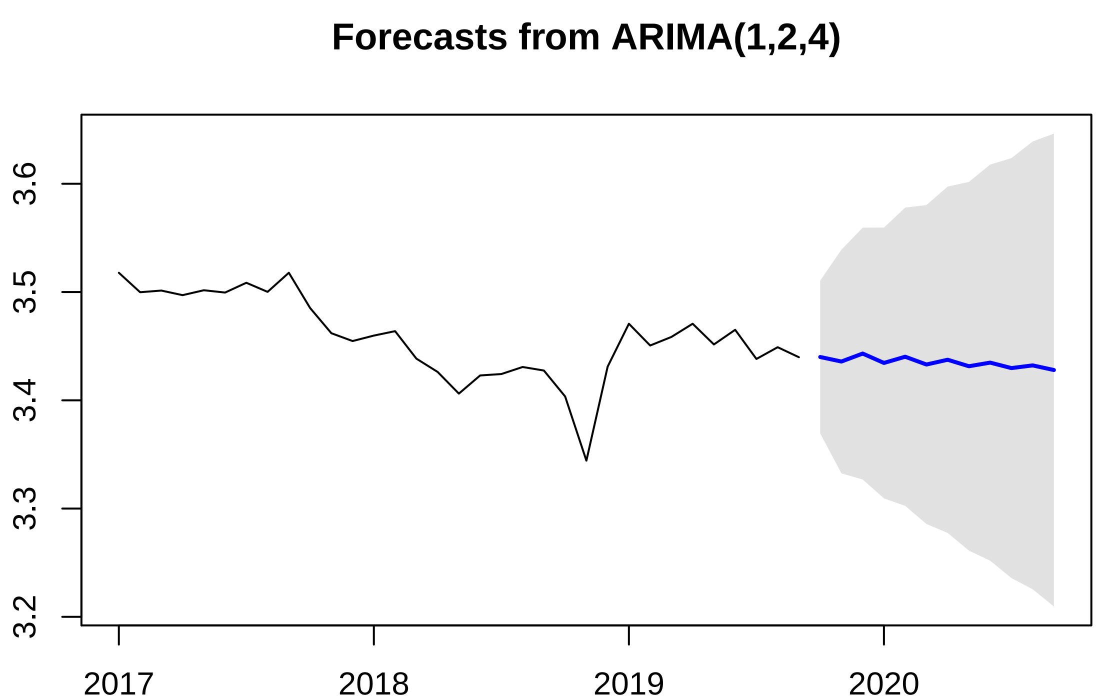
Prediction

The test data consists of values from 2019-10 to 2020-03. We predict the values of same duration from our model. Looking at the side by side view of our forecast values and original values, we can say that they are quite close and the original values are well within the 99.5% CI of forecasted values, hence our model is justifiable. The accuracy of the model can be seen through the errors in the model. We can see that RMSE and MAPE are quite low for the model which can be seen in Appendix 4.

```
##          Original Point_Forecast Lower_CI Upper_CI
## Oct 2019    31.49       31.18679 29.06172 33.46725
## Nov 2019    31.36       31.05627 28.00787 34.43645
## Dec 2019    31.37       31.28610 27.85145 35.14431
## Jan 2020    31.29       31.01614 27.37059 35.14726
## Feb 2020    30.91       31.19449 27.18090 35.80074
## Mar 2020    30.80       30.97088 26.72873 35.88630
```

Now, we will use this model to predict the values for coming six months.

```
##           Jan      Feb      Mar      Apr      May      Jun      Jul      Aug
## 2019
## 2020 31.01614 31.19449 30.97088 31.10784 30.92151 31.02518 30.86885 30.94577
##           Sep      Oct      Nov      Dec
## 2019          31.18679 31.05627 31.28610
## 2020 30.81358
```



Looking at the forecasted values and graph, we see an overall declining trend in this year for Samsung's market share in the mobile phone market.

Dashboard

The Dashboard is accessible online at : https://rproject718.shinyapps.io/try_shiny/

We have also created a DashBoard using RShiny. In the dashboard, it is possible to select the company (Samsung, Apple and Huawei) and an option to select analysis (Market Share History, Predictions, Residuals and Decomposition).
Our dashboard comes with a prediction analysis of the chosen company. It gives the detailed decomposition graphs which show trends and seasoanilty of the series through which multiple things can been explored. It also shows the graphs of residuals through which one can see how good of a fit the model is to the data. The prediction plots show the curve of market share of company chosen till September 2020. It also shows the exisiting trends of each company's market share of mobile phones. This DashBoard can be useful during presentations where the presenter needs pictorial representation of the data rather than a full report. The existing dashboards don't provide a full fledged analysis while they just act as a tool to visualize the existing or prediction data of these companies. There is generally no provision to see the seasonilty or trend componnents differently to visualize the time series. Also, they dont provide a visual to how good the fit of the model is. We aim to compensate all these in our dashboard.

Some visuals of the Dashboard are included in Appendix 8.

Conclusion and Future Work

The project analyzed the market share of the 3 largest mobile manufacturers in the world and their projected share for the coming six months using ARIMA model. The report outlines the process in detail for Samsung. We did a similar analysis for Apple. After differencing the data and achieving a stationary series, we run auto.arima() which gives a Seasonal Arima model with orders as (0,2,1)(0,1,0). We tune our p and q looking at the ACF and PACF graphs and obtain a model with orders (0,2,2).By looking at the errors and AIC values, we conclude that the model given by auto.arima() has less error and AIC value and hence we procede to predcit further values with that data. We see an increasing overall trend in Apple's Mobile Phone market share in second half of 2020. The full analysis can be found in Appendix 6.

Analysis of Huawei was also done. After differencing the data and achieving a stationary series, we run auto.arima() which gives an Arima model with orders as (2,2,0). We tune our p and q looking at the ACF and PACF graphs and obtain a model with orders (0,2,2). By looking at the errors and AIC values, we conclude that the model given by auto.arima() has less error and AIC value and hence we procede to predcit further values with that data. We see a strong increasing overall trend in Huawei's Mobile Phone market share in second half of 2020. The full analysis can be found in Appendix 7.

Time Series Analysis does not incorporate any external factors. For example, sales for some vendors may be affected more adversely by the ongoing coronavirus epidemic than others. But this model will not be able to incorporate such conditions. A possible extension to this project could be making a different prediction model that is robust to such external circumstances.

References

[1] https://www.statista.com/topics/840/smartphones/

[2] https://gs.statcounter.com/vendor-market-share/mobile/worldwide

[3] https://otexts.com/fpp2/non-seasonal-arima.html

[4] https://towardsdatascience.com/time-series-forecasting-arima-models-7f221e9eee06

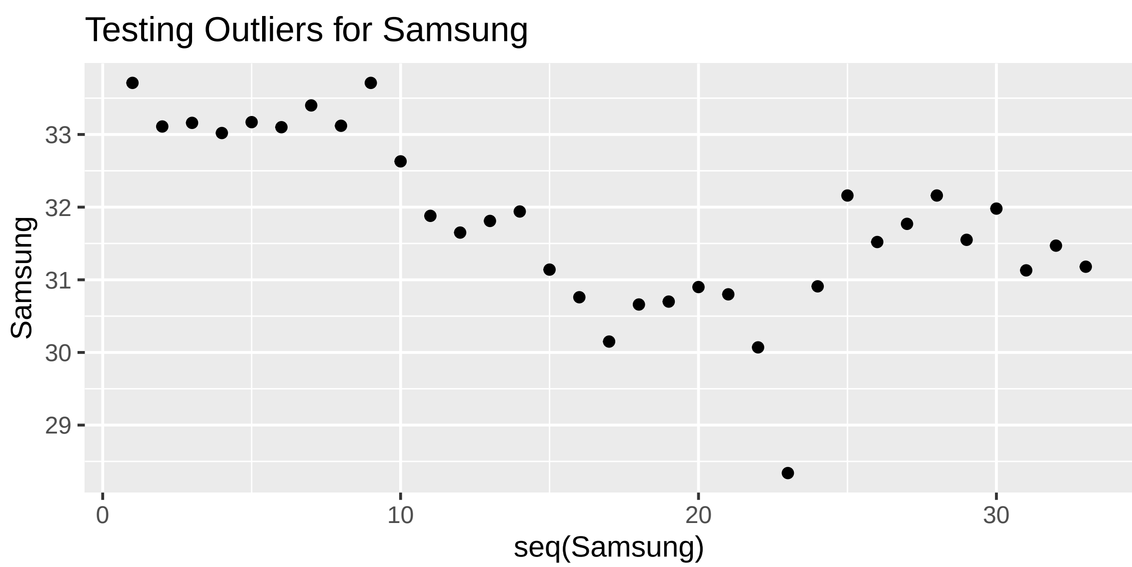
[5] https://en.wikipedia.org/wiki/Augmented_Dickey%E2%80%93Fuller_test

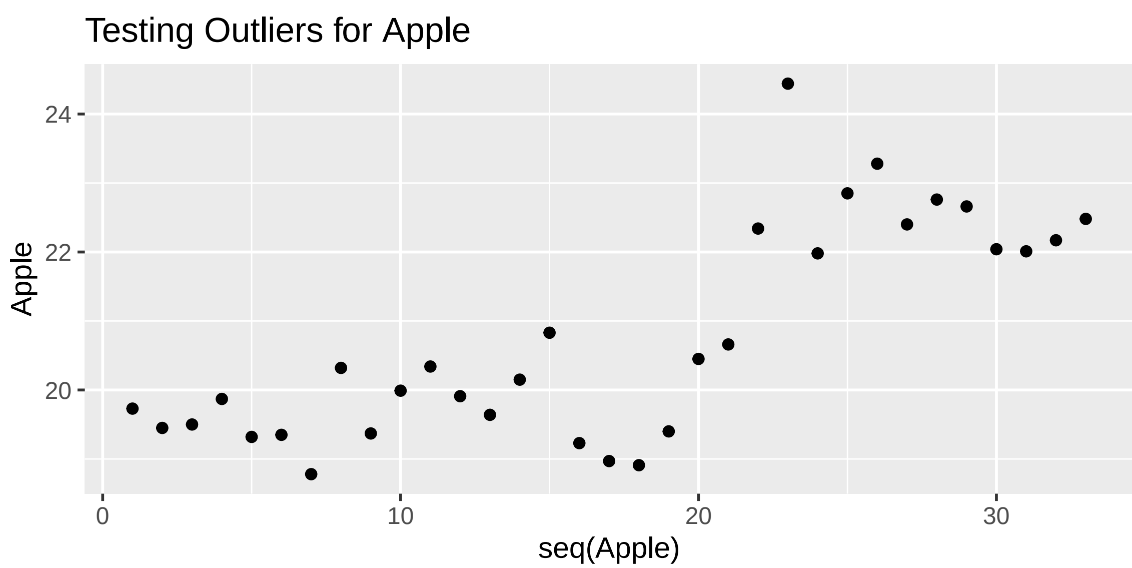
Contributions

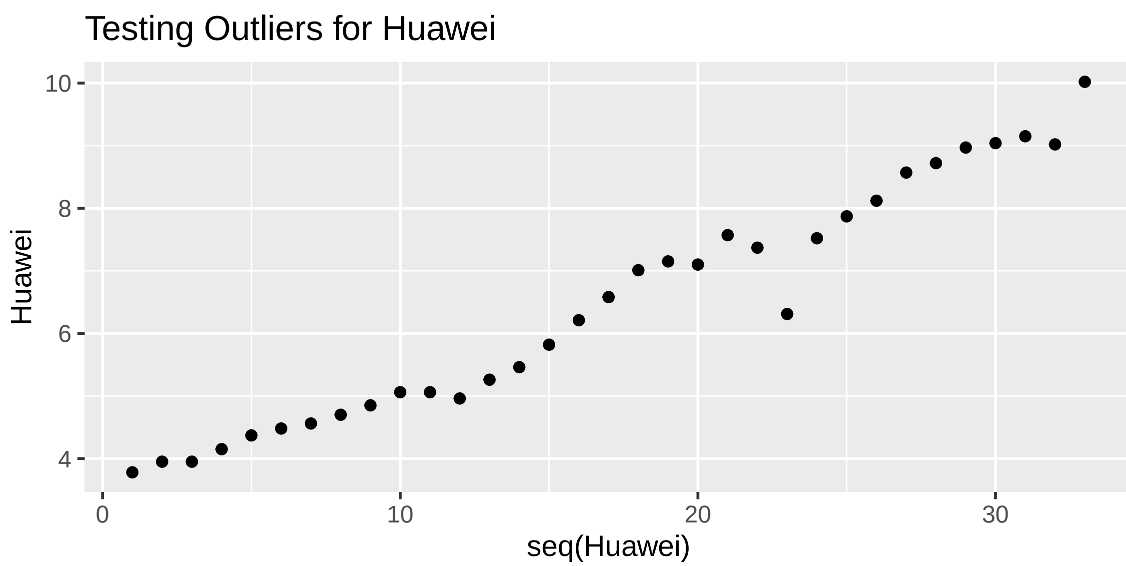
Both of us surveyed the data and made the models independently. After this, we split the companies for analysis. Post that, Rishabh focused more on compiling the report whereas Karanjot worked on making the dashboard.

Appendix

Appendix 1








```
##   ï..Date Samsung Apple Huawei Date
## 1       0       0     0      0    0
```

Appendix 2

```
##     ï..Date      Samsung          Apple           Huawei
##  2017-01: 1   Min.   :28.34   Min.   :18.78   Min.   : 3.780
##  2017-02: 1   1st Qu.:30.91   1st Qu.:19.45   1st Qu.: 4.850
##  2017-03: 1   Median :31.77   Median :20.32   Median : 6.310
##  2017-04: 1   Mean   :31.78   Mean   :20.78   Mean   : 6.446
##  2017-05: 1   3rd Qu.:33.02   3rd Qu.:22.17   3rd Qu.: 7.870
##  2017-06: 1   Max.   :33.71   Max.   :24.44   Max.   :10.020
##  (Other):27
##       Date
##  Min.   :2017-01-01
##  1st Qu.:2017-09-01
##  Median :2018-05-01
##  Mean   :2018-05-01
##  3rd Qu.:2019-01-01
##  Max.   :2019-09-01
##
```

```
## 'data.frame':    33 obs. of  5 variables:
##  $ ï..Date: Factor w/ 33 levels "2017-01","2017-02",..: 1 2 3 4 5 6 7 8 9 10 ...
##  $ Samsung: num  33.7 33.1 33.2 33 33.2 ...
##  $ Apple  : num  19.7 19.4 19.5 19.9 19.3 ...
##  $ Huawei : num  3.78 3.95 3.95 4.15 4.37 4.48 4.56 4.7 4.85 5.06 ...
##  $ Date   : Date, format: "2017-01-01" "2017-02-01" ...
```

Appendix 3

```
##
##  Augmented Dickey-Fuller Test
```

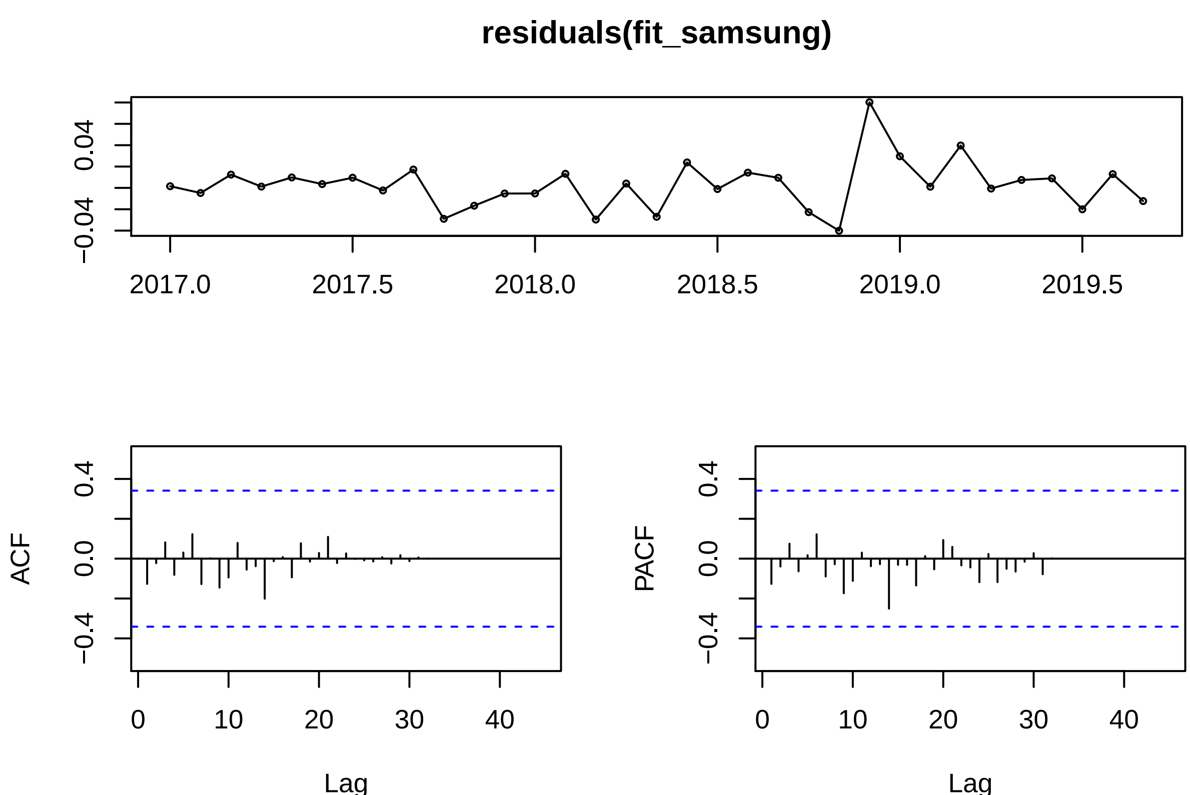
```
##
## data:  ts_data_samsung
## Dickey-Fuller = -1.3546, Lag order = 3, p-value = 0.8224
## alternative hypothesis: stationary
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  log(ts_data_samsung)
## Dickey-Fuller = -1.3777, Lag order = 3, p-value = 0.8135
## alternative hypothesis: stationary
```

Appendix 4

```
##                  ME      RMSE       MAE       MPE      MAPE
## Test set 0.08488703 0.2504532 0.2366786 0.2664794 0.7582135
```

Appendix 5



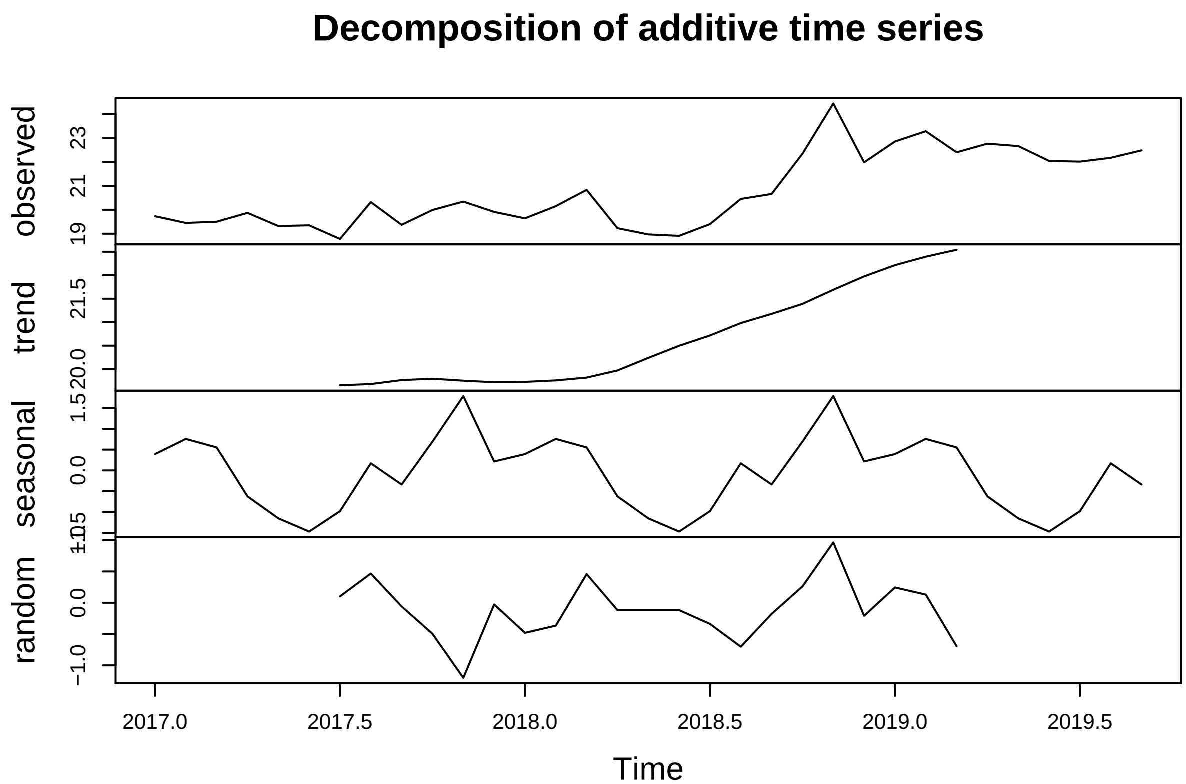
Fig(4)

Appendix 6

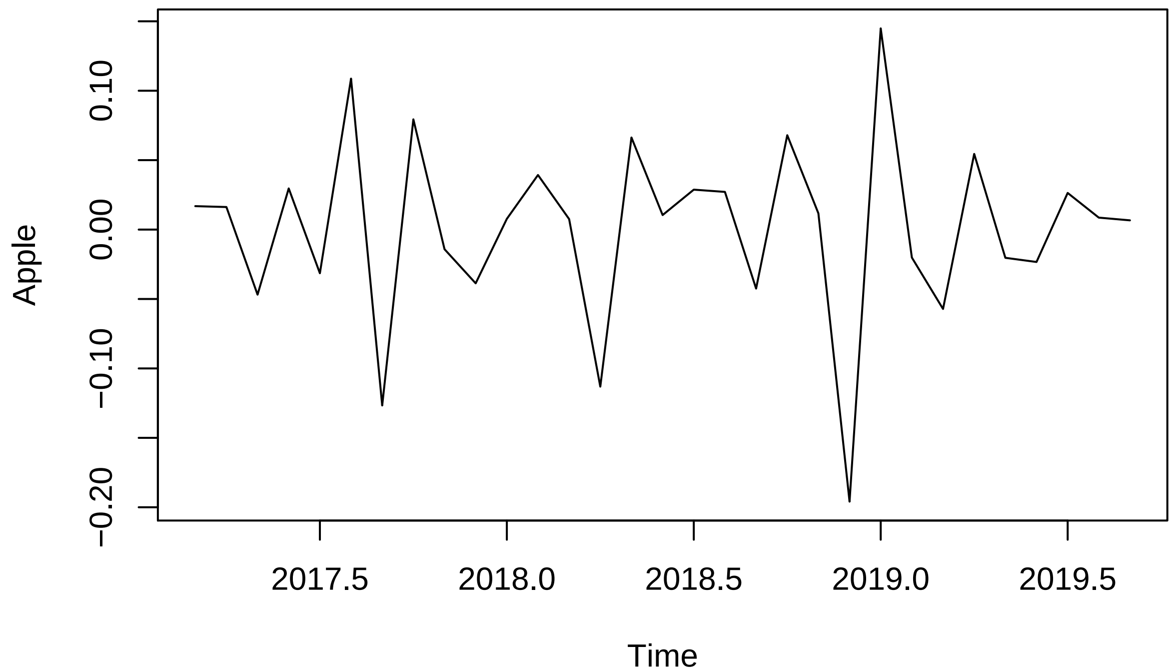
```
##
##  Augmented Dickey-Fuller Test
```

```
##
## data:  ts_data_apple
## Dickey-Fuller = -2.6372, Lag order = 3, p-value = 0.3265
## alternative hypothesis: stationary
```

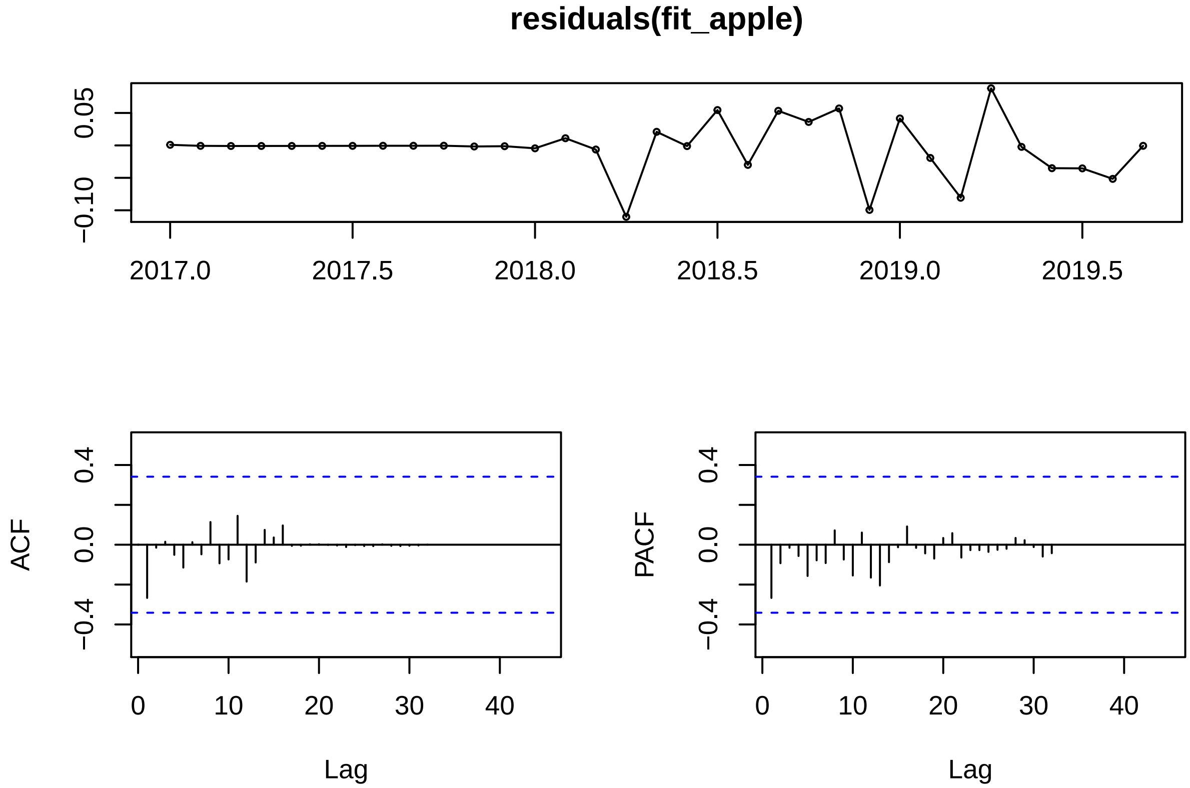
```
##
##  Augmented Dickey-Fuller Test
##
## data:  log_ts_data_apple
## Dickey-Fuller = -2.6567, Lag order = 3, p-value = 0.319
## alternative hypothesis: stationary
```

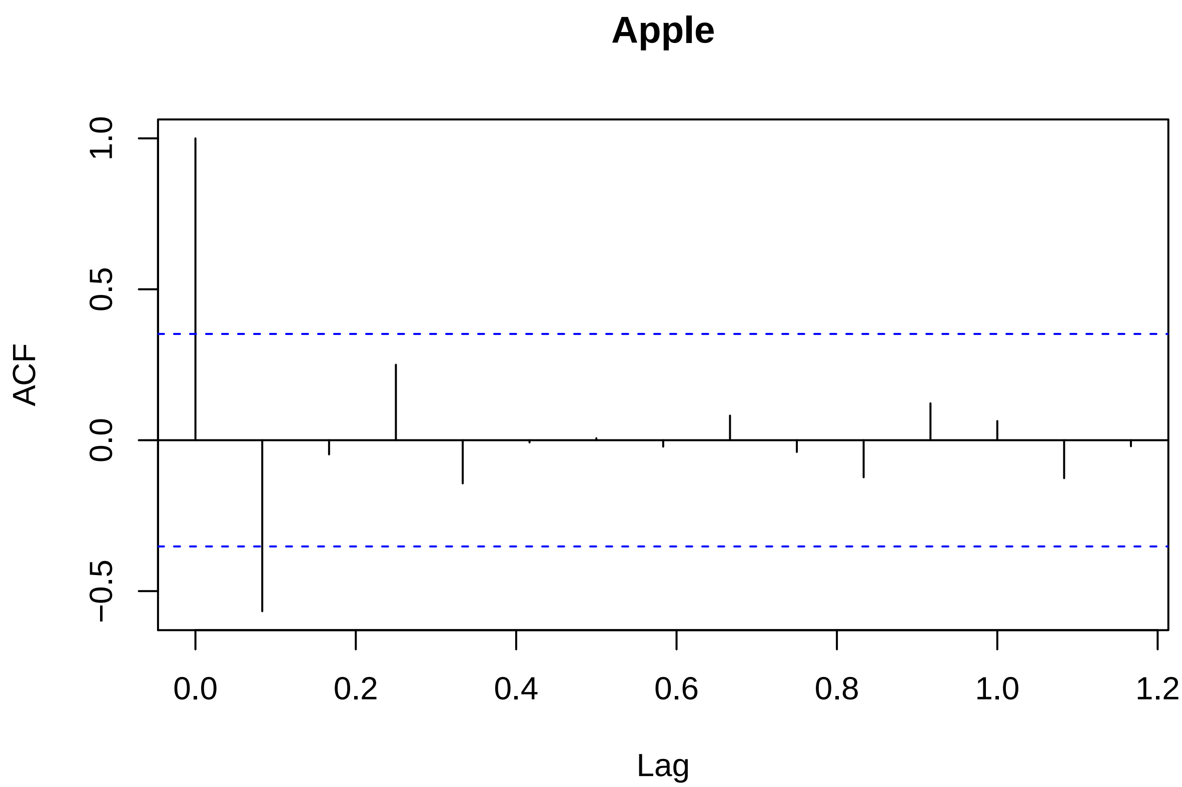



```
##
##  Augmented Dickey-Fuller Test
##
## data:  ts_data_diff_apple
## Dickey-Fuller = -3.6887, Lag order = 3, p-value = 0.04201
## alternative hypothesis: stationary
```

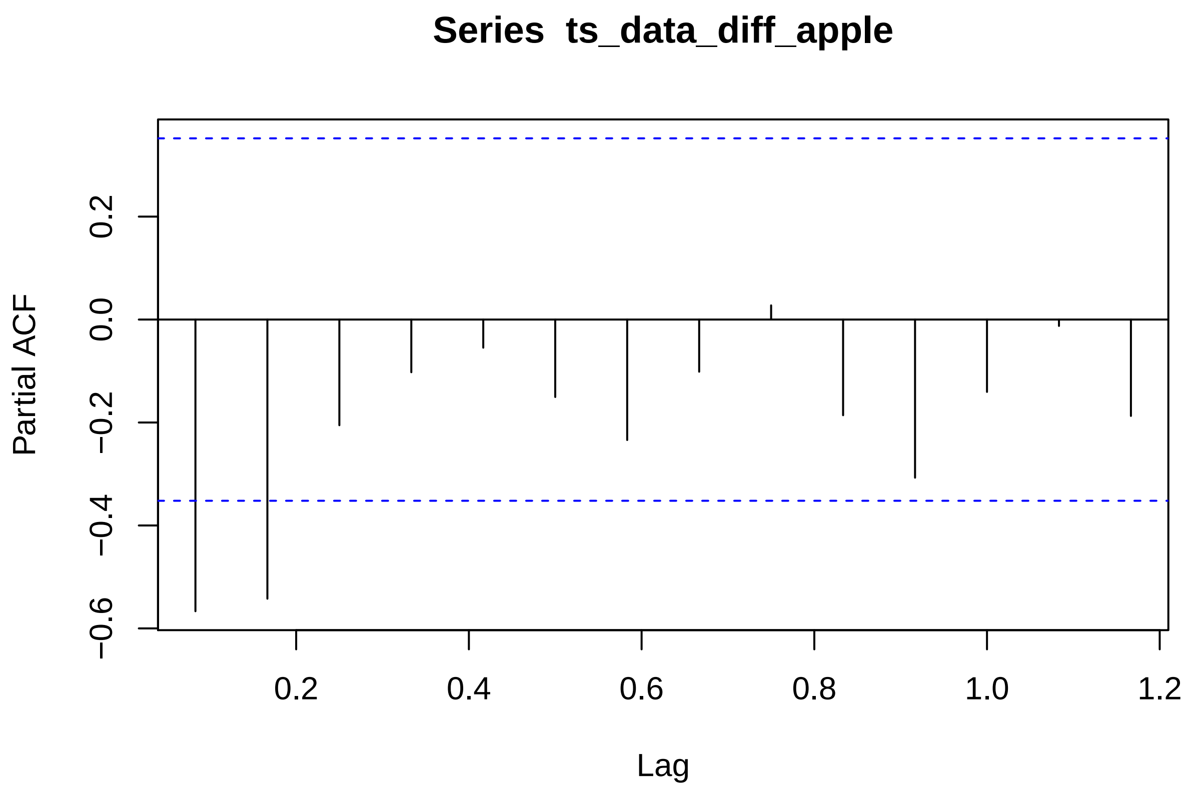



```
## Warning: The chosen seasonal unit root test encountered an error when testing for the second differer
## From stl(): series is not periodic or has less than two periods
## 1 seasonal differences will be used. Consider using a different unit root test.

## Warning in auto.arima(log_ts_data_apple, stepwise = FALSE, approximation =
## FALSE, : Having 3 or more differencing operations is not recommended. Please
## consider reducing the total number of differences.

## Series: log_ts_data_apple
## ARIMA(0,2,1)(0,1,0)[12]
##
## Coefficients:
##           ma1
##       -0.9853
## s.e.   0.1749
##
## sigma^2 estimated as 0.003078:  log likelihood=24.46
## AIC=-44.91   AICc=-44.16   BIC=-43.02
```


residuals(fit_apple)
0.05
−0.10
2017.0
2017.5
2018.0
2018.5
2019.0
2019.5
ACF
0.4
0.0
−0.4
0
10
20
30
40
Lag
PACF
0.4
0.0
−0.4
0
10
20
30
40
Lag



Apple
ACF
1.0
0.5
0.0
−0.5
0.0
0.2
0.4
0.6
0.8
1.0
1.2
Lag



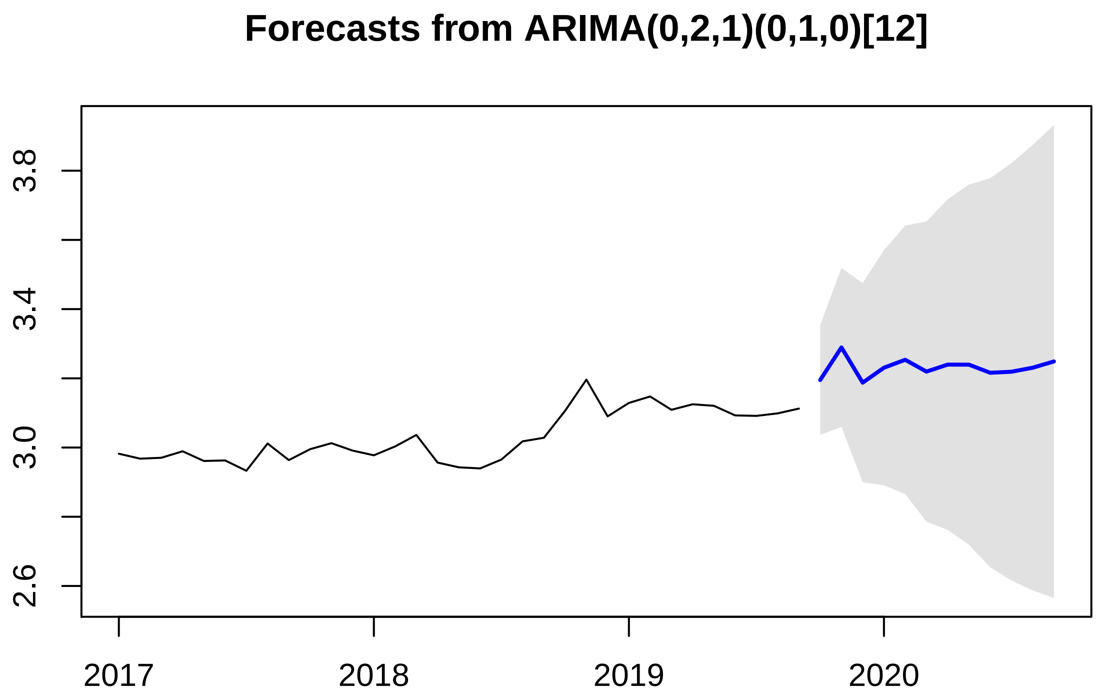

```
## Series: log_ts_data_apple
## ARIMA(0,2,2)
##
## Coefficients:
##           ma1     ma2
##       -1.2474  0.2474
## s.e.   0.2027  0.1744
##
## sigma^2 estimated as 0.001725:  log likelihood=53.63
## AIC=-101.26   AICc=-100.37   BIC=-96.96
```

```
##                  ME     RMSE      MAE       MPE     MAPE
## Test set -0.6994016 2.060086 1.561429 -3.342941 6.595084
```

```
##                ME     RMSE      MAE      MPE     MAPE
## Test set 1.821953 2.370556 1.965127 7.021224 7.669366
```

```
##          Original_Apple Point_Forecast_Apple Lower_CI_Apple Upper_CI_Apple
## Oct 2019          22.09             24.41313       20.83413       28.60694
## Nov 2019          22.90             26.82352       21.31998       33.74774
## Dec 2019          24.79             24.22794       18.16933       32.30681
## Jan 2020          24.76             25.29585       18.01058       35.52801
## Feb 2020          25.89             25.88334       17.56160       38.14842
## Mar 2020          27.03             25.01264       16.21195       38.59081
```

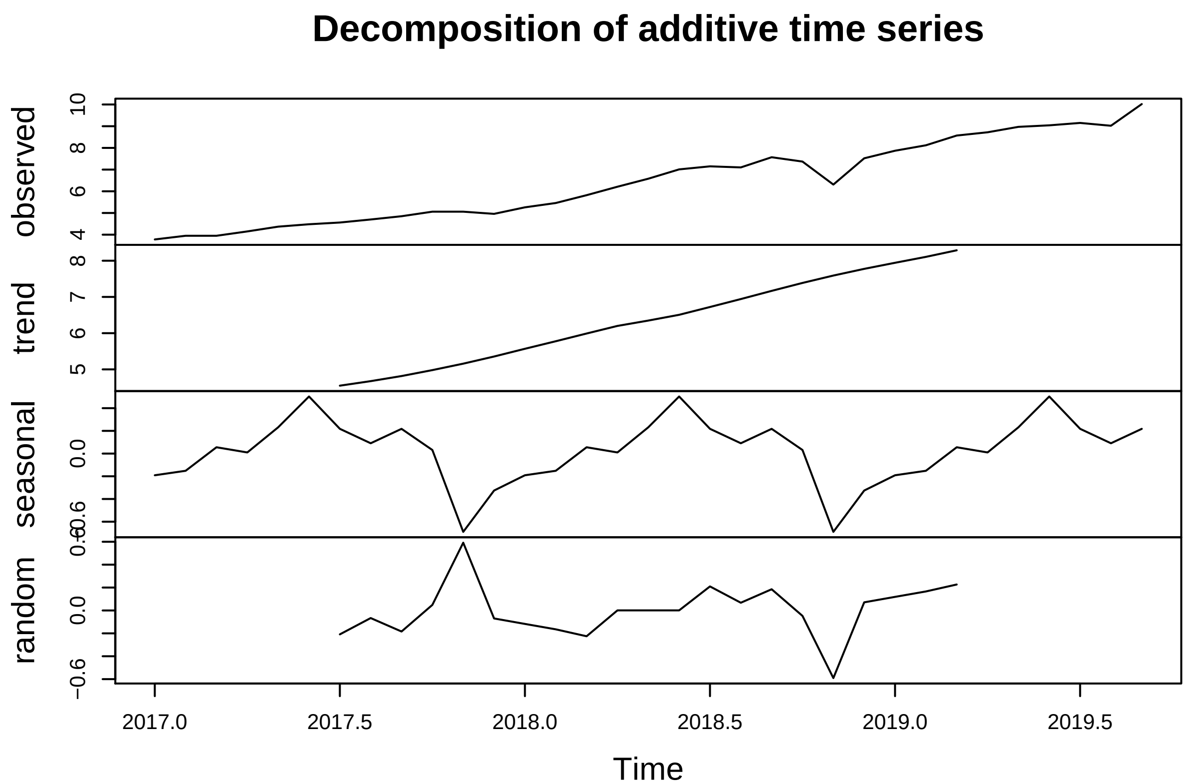
```
##           Jan      Feb      Mar      Apr      May      Jun      Jul      Aug
## 2019
## 2020 25.29585 25.88334 25.01264 25.52455 25.52231 24.93136 25.00510 25.29581
##           Sep      Oct      Nov      Dec
## 2019          24.41313 26.82352 24.22794
## 2020 25.76045
```



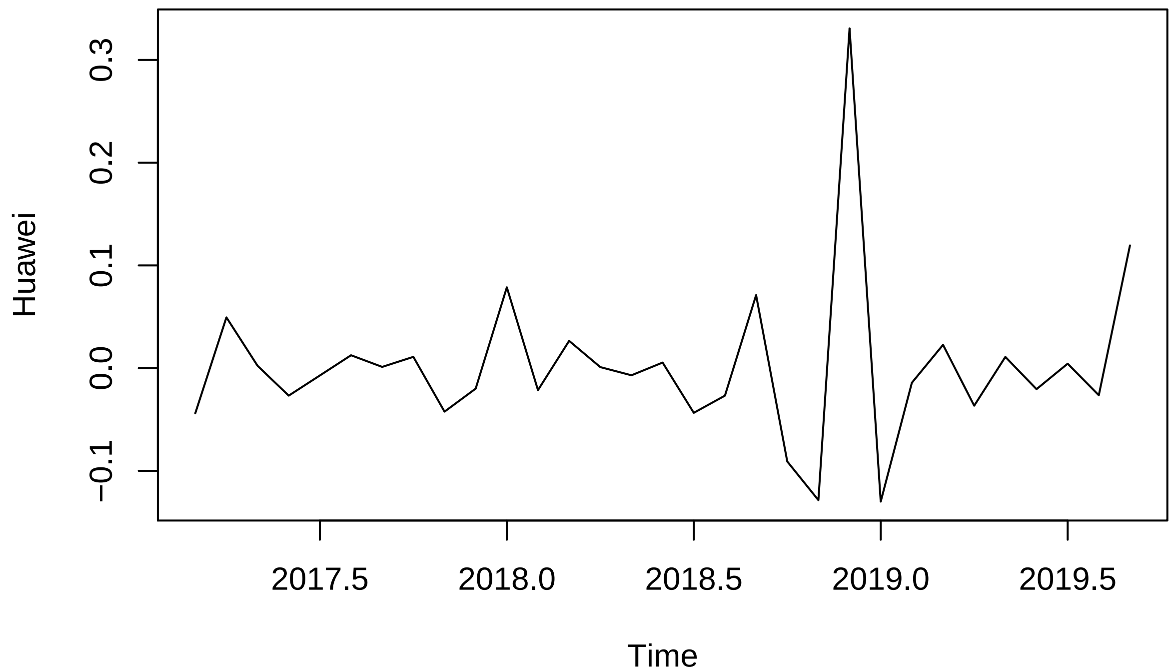
Appendix 7

```
##
##  Augmented Dickey-Fuller Test
##
## data:  ts_data_huawei
## Dickey-Fuller = -2.835, Lag order = 3, p-value = 0.2501
## alternative hypothesis: stationary
```

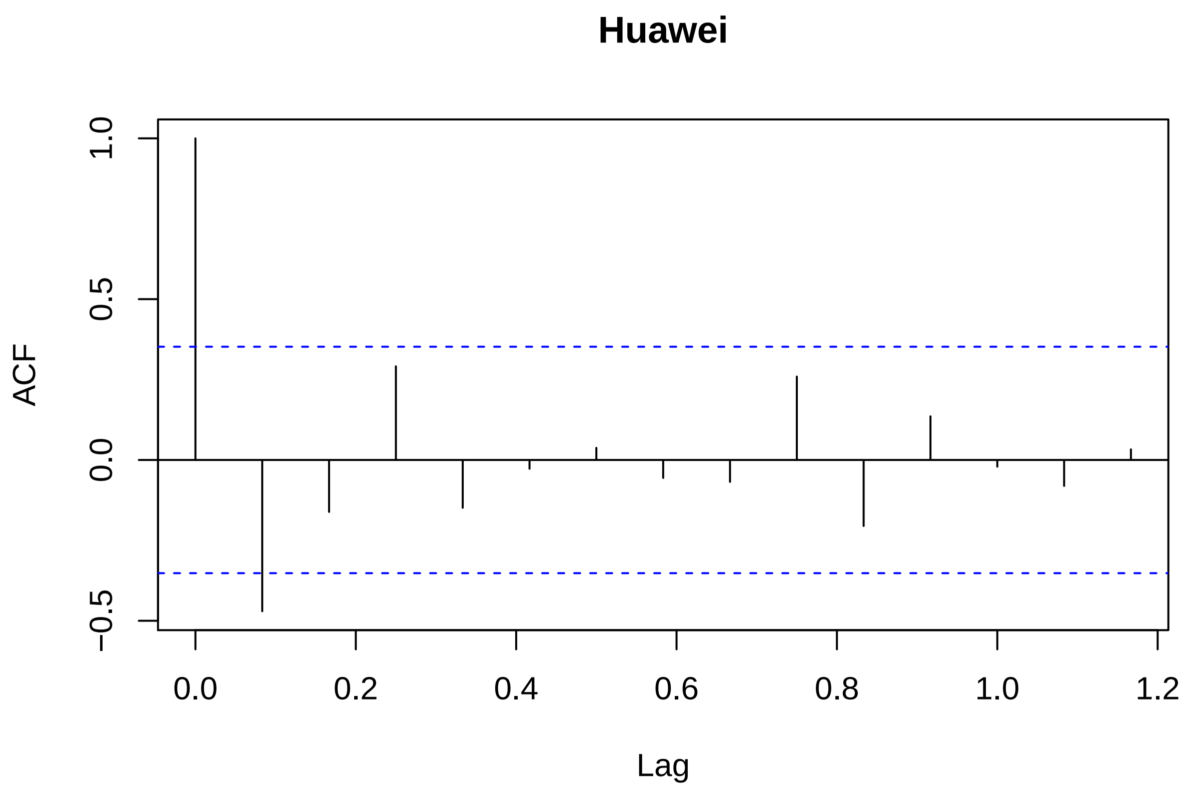
```
##
##  Augmented Dickey-Fuller Test
##
## data:  log_ts_data_huawei
## Dickey-Fuller = -3.0662, Lag order = 3, p-value = 0.1607
## alternative hypothesis: stationary
```

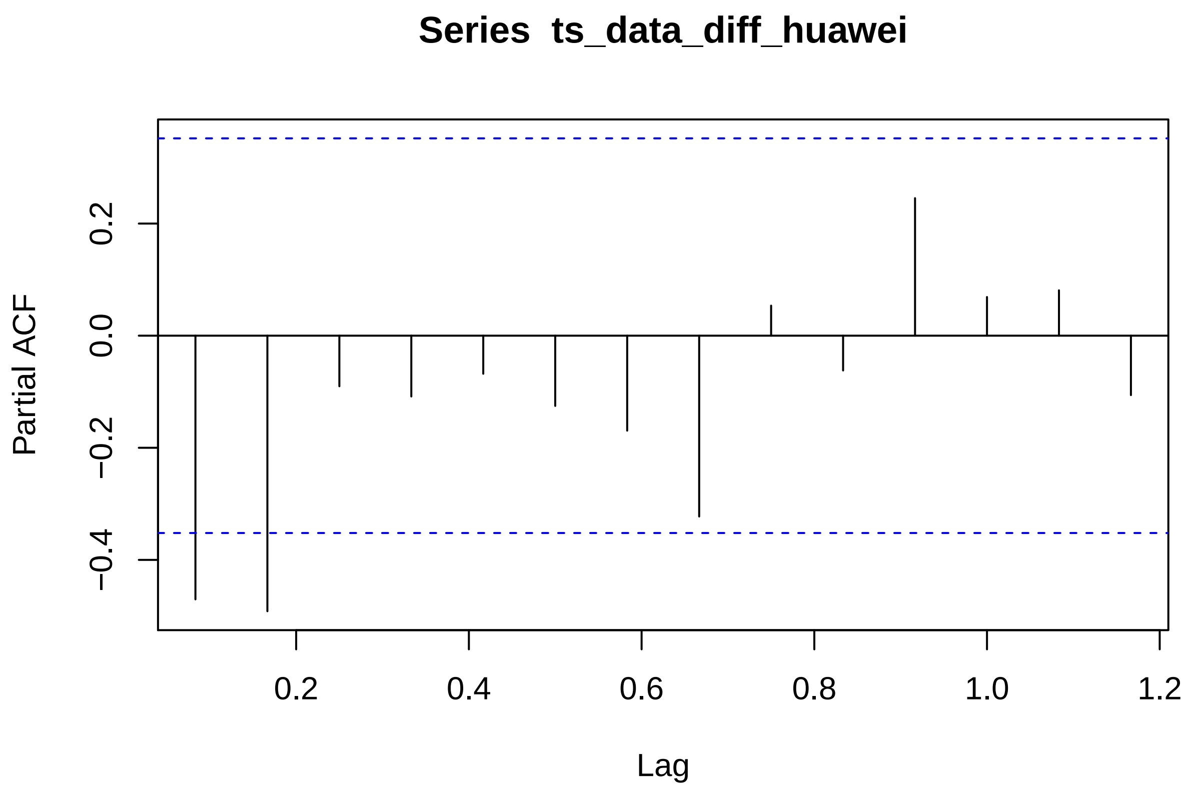



```
##
##  Augmented Dickey-Fuller Test
##
## data:  ts_data_diff_huawei
## Dickey-Fuller = -3.748, Lag order = 3, p-value = 0.03765
## alternative hypothesis: stationary
```



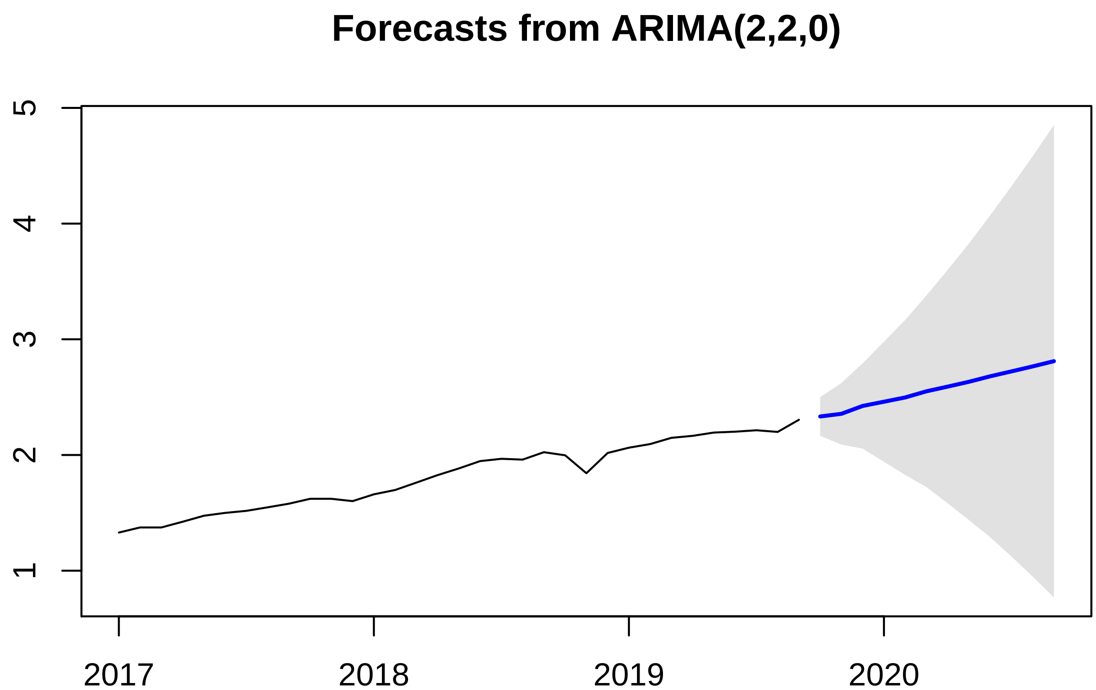

```
## Series: log_ts_data_huawei
## ARIMA(2,2,0)
##
## Coefficients:
##           ar1      ar2
##       -0.7620  -0.5306
## s.e.   0.1552   0.1504
##
## sigma^2 estimated as 0.003553:  log likelihood=43.99
## AIC=-81.99   AICc=-81.1   BIC=-77.69
```


Huawei
ACF
1.0
0.5
0.0
−0.5
0.0
0.2
0.4
0.6
0.8
1.0
1.2
Lag

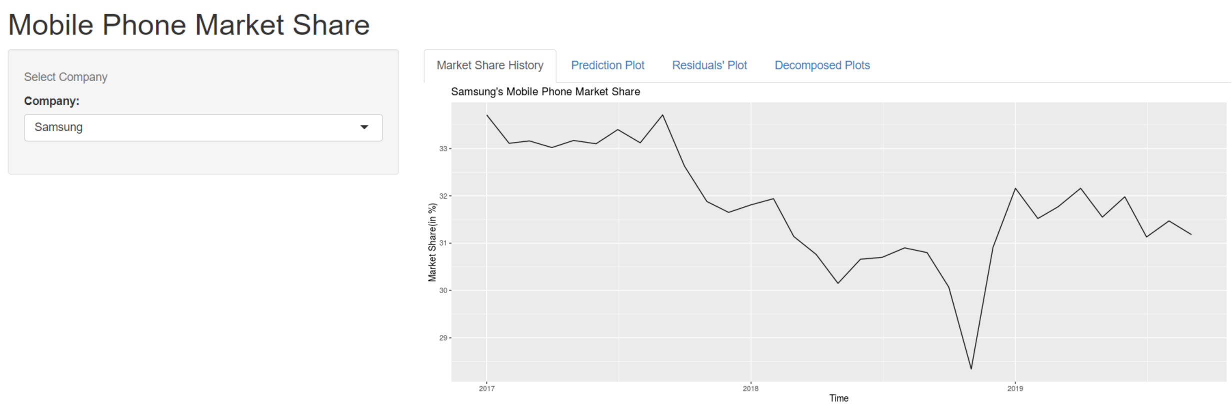


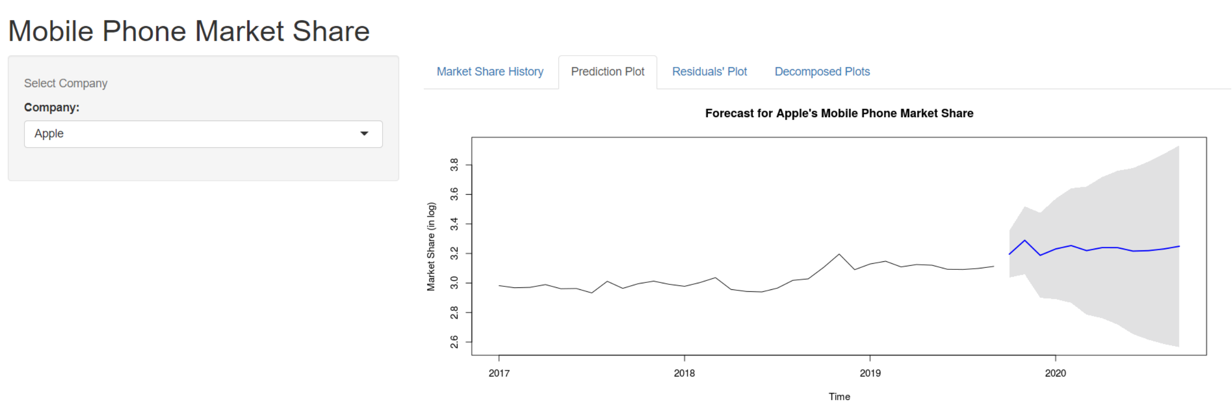

```
## Series: log_ts_data_huawei
## ARIMA(0,2,2)
##
## Coefficients:
##           ma1     ma2
##       -1.2899  0.2899
## s.e.   0.2335  0.2159
##
## sigma^2 estimated as 0.002599:  log likelihood=47.2
## AIC=-88.41   AICc=-87.52   BIC=-84.11

##                ME     RMSE      MAE      MPE     MAPE
## Test set 13.10617 13.13228 13.10617 53.34854 53.34854

##                ME     RMSE      MAE      MPE     MAPE
## Test set 13.63229 13.67933 13.63229 55.41172 55.41172

##          Original_Huawei Point_Forecast_Huawei Lower_CI_Huawei Upper_CI_Huawei
## Oct 2019           10.02              10.30591        8.718079        12.18293
## Nov 2019           10.18              10.55023        8.083832        13.76913
## Dec 2019            9.95              11.29116        7.807591        16.32901
## Jan 2020           10.61              11.71096        6.970539        19.67518
## Feb 2020           10.98              12.15025        6.220454        23.73278
## Mar 2020           10.70              12.81447        5.596452        29.34190
```

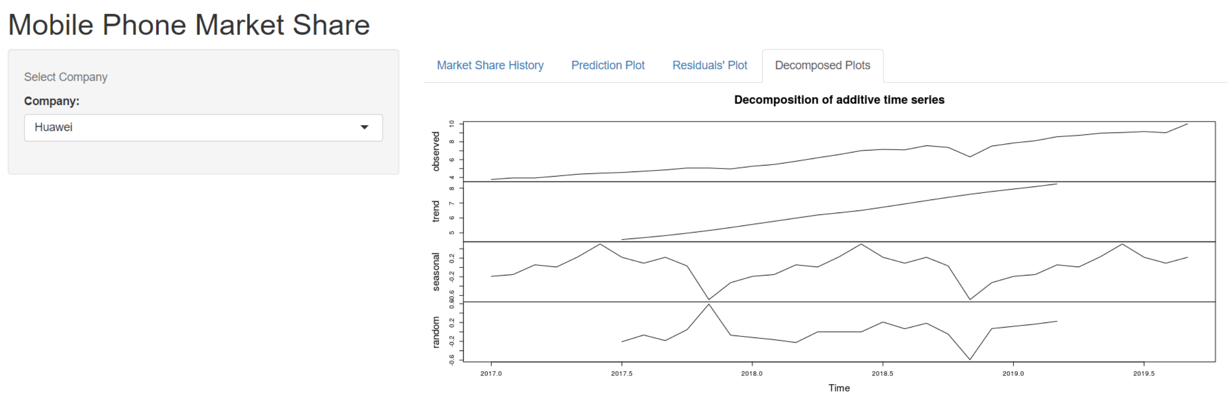
```
##           Jan      Feb      Mar      Apr      May      Jun      Jul      Aug
## 2019
## 2020 11.71096 12.15025 12.81447 13.34489 13.91047 14.58724 15.21953 15.89008
##           Sep      Oct      Nov      Dec
## 2019          10.30591 10.55023 11.29116
## 2020 16.62623
```

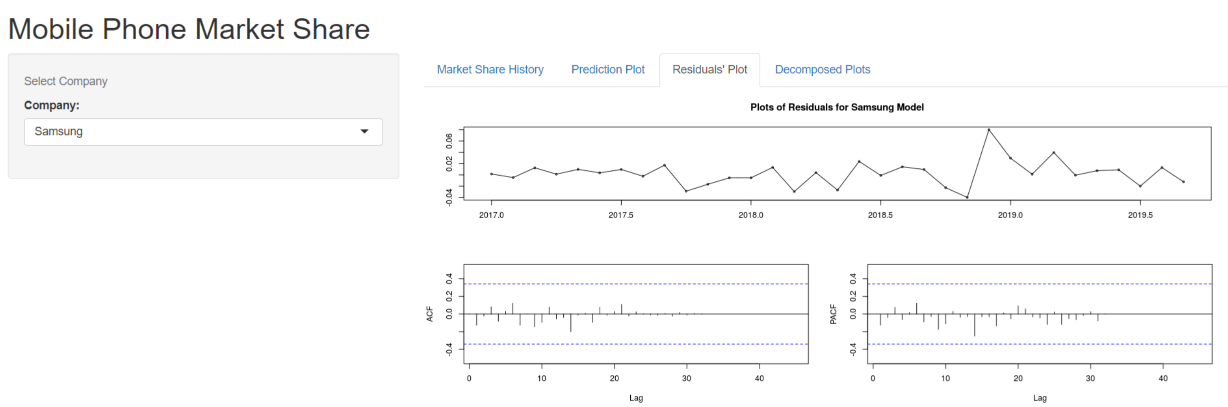


Appendix 8


Mobile Phone Market Share
Select Company
Company:
Samsung
Market Share History
Prediction Plot
Residuals' Plot
Decomposed Plots
Samsung's Mobile Phone Market Share
Market Share(in %)
33
32
31
30
29
2017
2018
2019
Time


Mobile Phone Market Share
Select Company
Company:
Apple
Market Share History
Prediction Plot
Residuals' Plot
Decomposed Plots
Forecast for Apple's Mobile Phone Market Share
Market Share (in log)
2.6
2.8
3.0
3.2
3.4
3.6
3.8
2017
2018
2019
2020
Time


Mobile Phone Market Share
Select Company
Company:
Huawei
Market Share History
Prediction Plot
Residuals' Plot
Decomposed Plots
Decomposition of additive time series
observed
trend
seasonal
random
Time


Mobile Phone Market Share
Select Company
Company:
Samsung
Market Share History
Prediction Plot
Residuals' Plot
Decomposed Plots
Plots of Residuals for Samsung Model
ACF
PACF
Lag
Lag