Data Analytics
Assignment -6

Name : R Siva Girish SRN : PES1201700159
Dataset : Sensex Dataset 2014 - 2019

The dataset is based on Sensex data collected over the years 2014 to 2019.Until current date.This dataset consists of attribute date,open,high,low,close and volume. Volume is the number of trading instances taking place. But we are more interested in the closing value of the share.Hence we shall evaluate a time series model over the entire dataset to forecast the closing values of share in the next few years.

	Date	Open	High	Low	Close	Volume
1	2014-10-16	26260.349609	26462.080078	25933.980469	25999.339844	10700
2	2014-10-17	25950.000000	26248.539063	25910.769531	26108.529297	9400
3	2014-10-20	26434.160156	26517.900391	26368.939453	26429.849609	9300
4	2014-10-21	26552.449219	26615.410156	26407.000000	26575.650391	8400
5	2014-10-22	26782.570313	26818.330078	26712.210938	26787.230469	6000
6	2014-10-27	26959.570313	26994.960938	26726.839844	26752.900391	6500
7	2014-10-28	26788.730469	26907.140625	26764.150391	26880.820313	6800
8	2014-10-29	27017.439453	27126.300781	26971.160156	27098.169922	8100
9	2014-10-30	27098.939453	27390.599609	27088.650391	27346.330078	7000
10	2014-10-31	27439.060547	27894.320313	27438.279297	27865.830078	11600

❖ Time Series Models

- ➢ A time series is a set of observations,each one being recorded at a specific time t
- ➢ Time Series modelling is a method by which we can predict the future values of a variable and gain useful insights regarding the trends followed by the data.
- ➢ Our Goal is to test different Time Series models and pick the most suitable model.
- ➢ Models to be performed
  - AR
  - MA
  - Holt Winter
  - ARIMA

❖ Naive Model :

Naive<-naive(Bst)

accuracy(Naive)

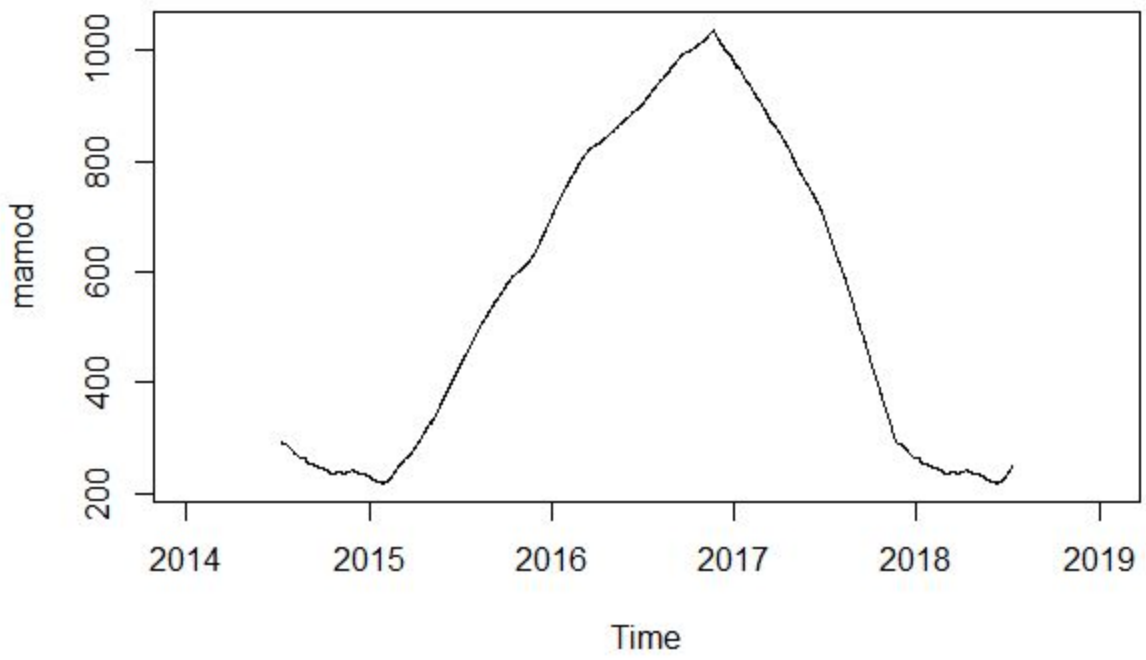
```
> accuracy(Naive)
                    ME     RMSE      MAE       MPE     MAPE      MASE       ACF1
Training set 0.2345205 92.05889 28.49753 -4.034641 12.50867 0.06425446 -0.3007802
>
```

Mape values is 12.50867

Therefore Accuracy of model is around 87%.

❖ Moving Average Model

```
mamod<-ma(Bst,order=365)
plot.ts(mamod)
mfore<-forecast(mamod)
accuracy(mfore)
```



Accuracy :

```
> accuracy(mfore)
                     ME      RMSE       MAE          MPE       MAPE         MASE         ACF1
Training set 0.005805189 0.649125 0.2129152 0.005412662 0.05837852 0.0006078316 0.002248723
> |
```

Mape Value is around 0.05 and RMSE is very low as well.

Accuracy of the model is around 99.95%

This is a very good model

❖ Autoregression Model

armod<-ar(Bst)
frmod<-forecast(armod)
accuracy(frmod)

```
> accuracy(frmod)
                     ME     RMSE      MAE      MPE     MAPE      MASE         ACF1
Training set 0.3537945 81.04302 33.83996 -10.805 18.92937 0.07630024 -0.008830074
>
```

Mape ~ 19
Accuracy ~ 81%
RMSE is also pretty High ~81
Therefore this model doesn't perform very well when compared to Moving Average model.

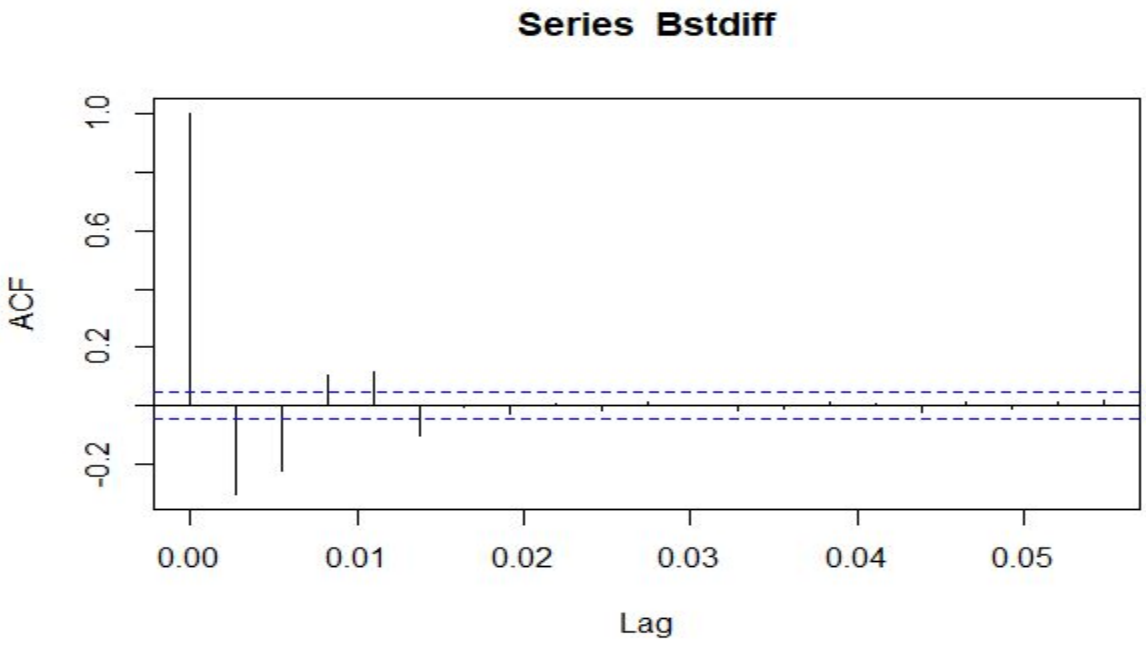
- Performing Augmented Sickey Fuller Test on data to check whether it is stationary or not.
  adfe_test <- adf.test(Bst,alternative = 'stationary')
  print(adfe_test)

```
> adfe_test <- adf.test(Bst,alternative = 'stationary')
> print(adfe_test)

        Augmented Dickey-Fuller Test

data:  Bst
Dickey-Fuller = -2.1008, Lag order = 12, p-value = 0.5357
alternative hypothesis: stationary
```

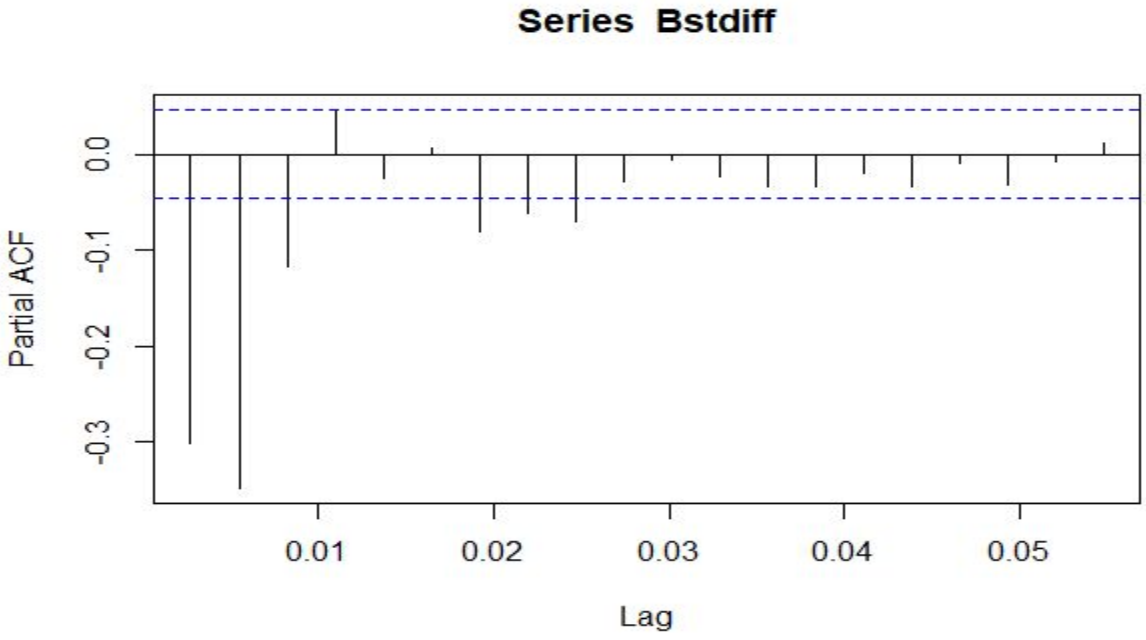
Clearly the dataset is not stationary.
So we difference the data 1 time.

```
> print(adf_test)

        Augmented Dickey-Fuller Test

data:  Bstdiff
Dickey-Fuller = -15.136, Lag order = 12, p-value = 0.01
alternative hypothesis: stationary
```

❖Auto Regression Integrated Moving Average Model



Autocorrelation Function (ACF) on differenced data



Passive Autocorrelation Function (PACF) on differenced data.

```
> acf(Bstdiff,lag.max=20,plot=FALSE)

Autocorrelations of series 'Bstdiff', by lag

0.00000 0.00274 0.00548 0.00822 0.01096 0.01370 0.01644 0.01918 0.02192 0.02466 0.02740 0.03014 0.03288
  1.000  -0.301  -0.225   0.102   0.115  -0.101  -0.006  -0.028   0.006  -0.018   0.014   0.002  -0.018
0.03562 0.03836 0.04110 0.04384 0.04658 0.04932 0.05205 0.05479
 -0.011   0.010   0.007  -0.021   0.011  -0.013   0.013   0.017
> pacf(Bstdiff,lag.max=20,plot=FALSE)

Partial autocorrelations of series 'Bstdiff', by lag

0.00274 0.00548 0.00822 0.01096 0.01370 0.01644 0.01918 0.02192 0.02466 0.02740 0.03014 0.03288 0.03562
 -0.301  -0.347  -0.117   0.046  -0.025   0.006  -0.081  -0.062  -0.069  -0.028  -0.005  -0.022  -0.033
0.03836 0.04110 0.04384 0.04658 0.04932 0.05205 0.05479
 -0.034  -0.019  -0.033  -0.008  -0.032  -0.007   0.012
>
```

In the ACF and the PACF plot the values of the lag are not integral.Therefore it is difficult to determine exact integral values for the factors p and q in arima. So we use the auto.arima function.

```
> auto.arima(Bst)
Series: Bst
ARIMA(4,1,2)

Coefficients:
          ar1     ar2     ar3     ar4      ma1      ma2
      -0.1623  0.3245  0.2359  0.2582  -0.2873  -0.6025
s.e.   0.1256  0.0680  0.0346  0.0338   0.1270   0.1100

sigma^2 estimated as 6622:  log likelihood=-10615.08
AIC=21244.17   AICc=21244.23   BIC=21282.73
>
```

The auto.arima function tells us that the optimal values of p and q are 4 and 2 respectively.The values of d as calculated before is 1 based on the augmented dickey fuller test.We use the optimal values for arima model and estimate our time series data using this model.

```
> accuracy(farimamod)
                     ME      RMSE       MAE        MPE      MAPE        MASE         ACF1
Training set 0.668951 81.21772 33.79523 -8.091526 16.93045 0.07619938 0.002192935
>
```

Arima model RMSE is pretty high 81.2722
Mape is around 17%
Accuracy of the model is around 83%
Therefore the arima model is definitely not the most suitable model.

❖ Different models and their Accuracy's

Model Name	RMSE	Accuracy% (100-MAPE)
Naive	92.05	87
AR	81.04	81
MA(Best Model)	0.649	99.5
Holt Winter (Previous Assignment)	91.36	81
ARIMA	81.272	83

Conclusion :

Therefore Based on the aforementioned statistics we can conclude that the moving Average model is the best time series model for the Sensex dataset.