Data Analytics
Assignment -5

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 run Holt Winter model on the dataset and measure its accuracy.

- Holt Winter Model
  - Converting data to Time Series data

Code:

```
Close_Values<-ts(Sensex$Close,frequency=365,start=c(2014,10,16),end = c(2019,10,15))
#Here we convert the closing values of the share to time series data so that it can be modeled.
```

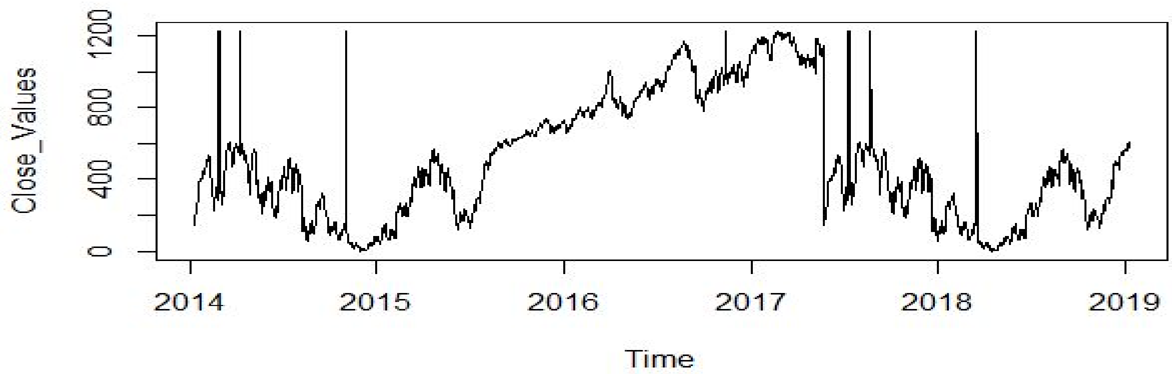
- Applying Holt Winter model

Code:

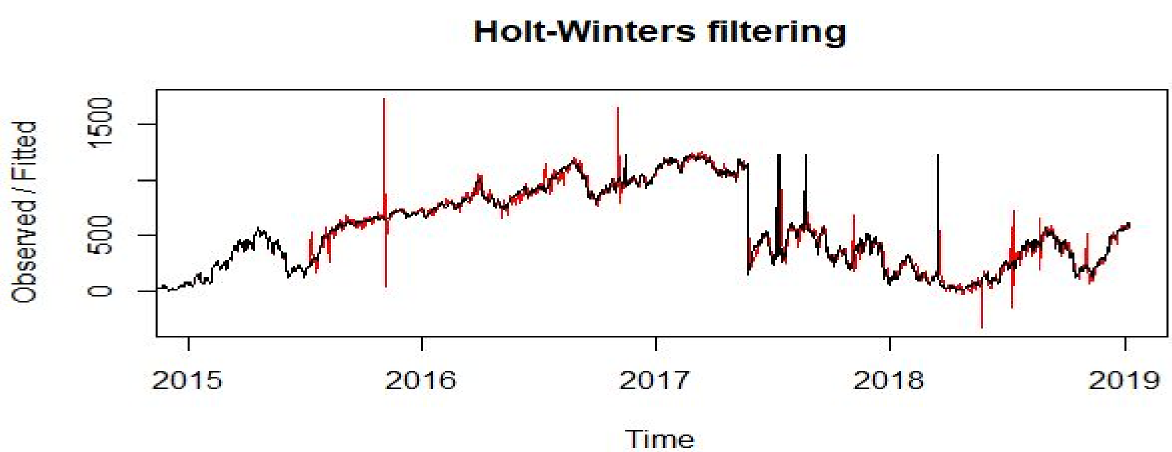
```
CHWF1<-HoltWinters(Close_Values,beta=NULL,gamma=NULL)
#Modeling the time series data as a holt winter model.
```

- Forecasting The data

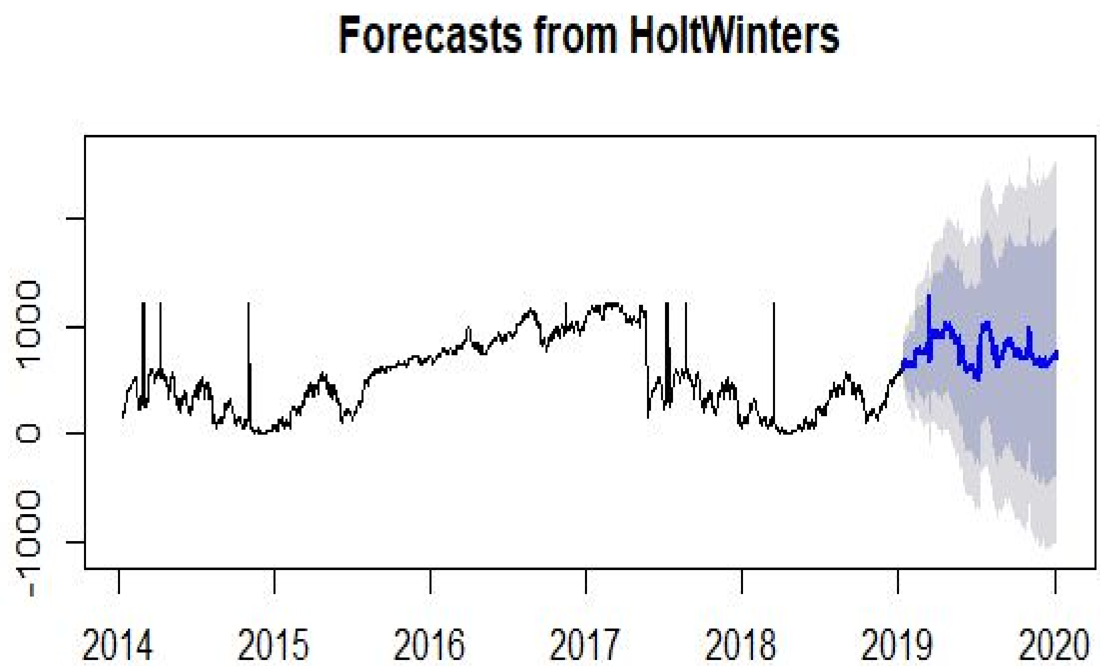
```
fc<-forecast(CHWF1,h=365) [Forecasting for an entire year]
```



In the time series dataset the frequency has been set to 365 days and the time series data is indexed starting from the first record of data to the most recent to get data year wise.



By Setting the Beta parameter of the HoltWinters function to NULL we will obtain a model with exponential smoothing practically suppressing all the outliers.



The model plotted above is the forecast of the HoltWinter model till the year 2020.

```
> accuracy(fc)
                      ME     RMSE     MAE       MPE     MAPE       MASE
Training set -0.03696946 91.36239 40.3543 -1.173932 18.70256 0.09098835
                   ACF1
Training set 0.09533314
>
```

The Mean absolute percentage Error(Mape) is 18.7025% therefore the accuracy of the model is (100% - Mape).

Accuracy of the model = 100 -18.7025 = 81.2975

~81%

❖Linear Model

- ➢Linear model for this dataset is not possible as it doesn’t have any dependent variables.
- ➢All variables are independent therefore we merged the Gold prices for 5 years in our dataset.
- ➢Against USDollars,GBP and Euro prices of gold respectively.

	Date	USD..AM.	USD..PM.	GBP..AM.	GBP..PM.	EURO..AM.	EURO..PM.
1	2019-10-18	1487.50	1490.00	1154.15	1155.64	1336.67	1337.28
2	2019-10-17	1484.45	1492.65	1151.64	1162.63	1336.60	1341.59
3	2019-10-16	1482.55	1485.10	1166.32	1155.85	1344.52	1343.27
4	2019-10-15	1494.75	1487.80	1183.69	1178.34	1357.08	1353.30
5	2019-10-14	1494.20	1490.60	1188.79	1182.94	1354.04	1352.12
6	2019-10-11	1498.35	1479.15	1197.93	1166.01	1359.90	1338.33
7	2019-10-10	1508.20	1494.80	1232.35	1222.75	1368.69	1356.38
8	2019-10-09	1503.40	1507.25	1228.43	1232.93	1369.00	1372.65
9	2019-10-08	1500.00	1505.85	1225.50	1233.14	1365.30	1372.28
10	2019-10-07	1502.15	1501.25	1221.40	1218.11	1369.36	1365.54
11	2019-10-04	1509.50	1499.15	1223.75	1220.01	1374.70	1366.78
12	2019-10-03	1504.00	1517.10	1221.70	1223.84	1372.25	1380.26
13	2019-10-02	1484.05	1492.60	1213.21	1215.13	1359.84	1364.68

Showing 1 to 15 of 13,093 entries, 7 total columns

LBMA Gold Dataset (Dataset used for merging gold prices)

	Date	Open	High	Low	Close	Volume	USD..PM.	GBP..PM.	EURO..PM.
1	2014-10-16	26260.349609	26462.080078	25933.980469	25999.34	10700	1237.75	773.545	971.851
2	2014-10-17	25950.000000	26248.539063	25910.769531	26108.53	9400	1234.25	768.237	967.053
3	2014-10-20	26434.160156	26517.900391	26368.939453	26429.85	9300	1244.50	771.161	974.321
4	2014-10-21	26552.449219	26615.410156	26407.000000	26575.65	8400	1250.25	774.340	982.283
5	2014-10-22	26782.570313	26818.330078	26712.210938	26787.23	6000	1243.75	775.405	982.037
6	2014-10-27	26959.570313	26994.960938	26726.839844	26752.90	6500	1228.75	761.543	966.835
7	2014-10-28	26788.730469	26907.140625	26764.150391	26880.82	6800	1229.25	760.580	964.572
8	2014-10-29	27017.439453	27126.300781	26971.160156	27098.17	8100	1223.50	757.632	959.006
9	2014-10-30	27098.939453	27390.599609	27088.650391	27346.33	7000	1202.00	750.406	952.456
10	2014-10-31	27439.060547	27894.320313	27438.279297	27865.83	11600	1164.25	729.206	931.325
11	2014-11-03	27943.039063	27969.820313	27785.400391	27860.38	9500	1167.75	730.986	935.397
12	2014-11-05	27907.189453	28010.390625	27857.650391	27915.88	8700	1142.00	715.539	914.991
13	2014-11-07	27902.710938	27980.929688	27739.560547	27868.63	10500	1154.50	728.759	928.801

Showing 1 to 15 of 1,186 entries, 9 total columns

- On this merged dataset we perform a linear regression on Closing values of Sensex to USD,GBP and EURO gold prices.

```
> #Applied Linear regression on closing values of Sensex dataset
> linmod<-lm(Nmer$Close~Nmer$USD..PM.+Nmer$GBP..PM.+Nmer$EURO..PM.)
> summary(linmod)

Call:
lm(formula = Nmer$Close ~ Nmer$USD..PM. + Nmer$GBP..PM. + Nmer$EURO..PM.)

Residuals:
    Min      1Q  Median      3Q     Max 
-6584.1 -2291.0  -168.9  2336.3  6979.3 

Coefficients:
                Estimate Std. Error t value Pr(>|t|)    
(Intercept)    19243.883   1708.779  11.262   <2e-16 ***
Nmer$USD..PM.     20.463      2.186   9.359   <2e-16 ***
Nmer$GBP..PM.     35.597      1.720  20.693   <2e-16 ***
Nmer$EURO..PM.   -42.216      2.258 -18.692   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3047 on 1182 degrees of freedom
Multiple R-squared:   0.55,	Adjusted R-squared:  0.5488 
F-statistic: 481.5 on 3 and 1182 DF,  p-value: < 2.2e-16
```

Conclusion :

The accuracy of the model is around 82% which is good but it could possibly be made better.Log functions cannot be applied to this data set as it is the case that some undefined values enter our time series data set and the Mape becomes infinite.Hence this model can neither be log transformed to get a better fit.Linear Regression on this dataset upon merging with gold values gives us an Adjusted R Squared Value of around 55%.Which is very poor.Therefore we must test other models such as Autoregressive Moving Average models(ARMA) to find a better fit for the sensex data.Which may lead to a better fit.