TIME SERIES FORECASTING PROJECT

For this particular assignment, the data of different types of wine sales in the 20th century is to be analyzed. Both of these data are from the same company but of different wines. As an analyst in the ABC Estate Wines, you are tasked to analyze and forecast Wine Sales in the 20th century.

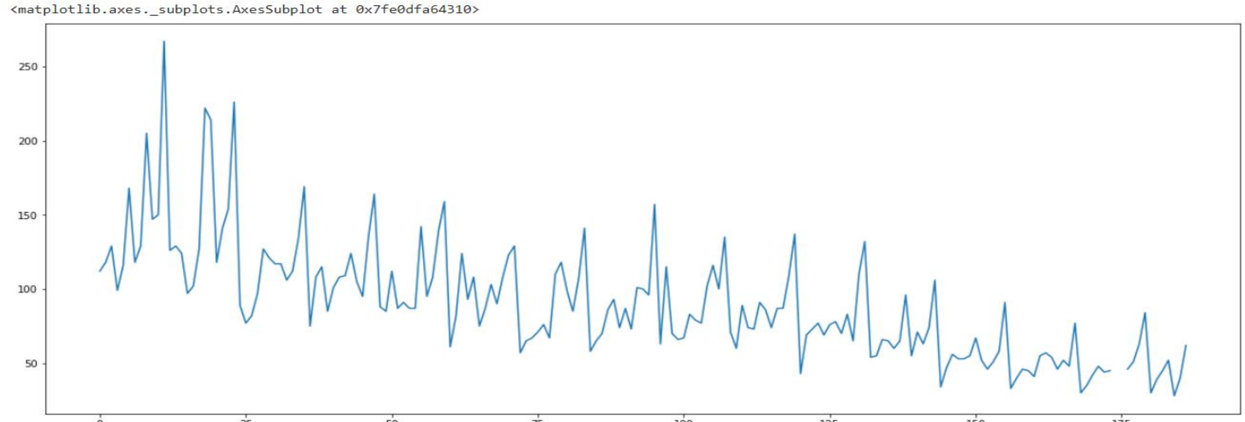
ROSE.CSV

1. Read the data as an appropriate Time Series data and plot the data.
   Time Series is a sequence of observations recorded at regular time intervals.

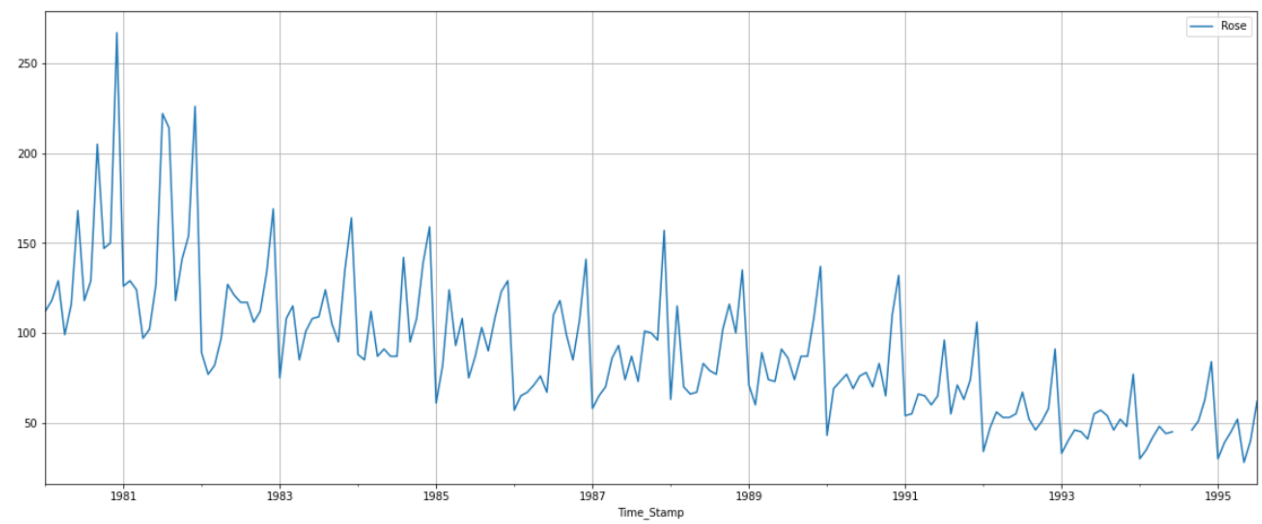
Time_Stamp	YearMonth	Rose
1980-01-31	1980-01	112.0
1980-02-29	1980-02	118.0
1980-03-31	1980-03	129.0
1980-04-30	1980-04	99.0
1980-05-31	1980-05	116.0

Time_Stamp	YearMonth	Rose
1995-03-31	1995-03	45.0
1995-04-30	1995-04	52.0
1995-05-31	1995-05	28.0
1995-06-30	1995-06	40.0
1995-07-31	1995-07	62.0

<matplotlib.axes._subplots.AxesSubplot at 0x7fe0dfa64310>



Given data is not time. So we parse the date range and create a timestamp. We also notice the fluctuations in the trend in the initial years and slowly decreasing the following years.

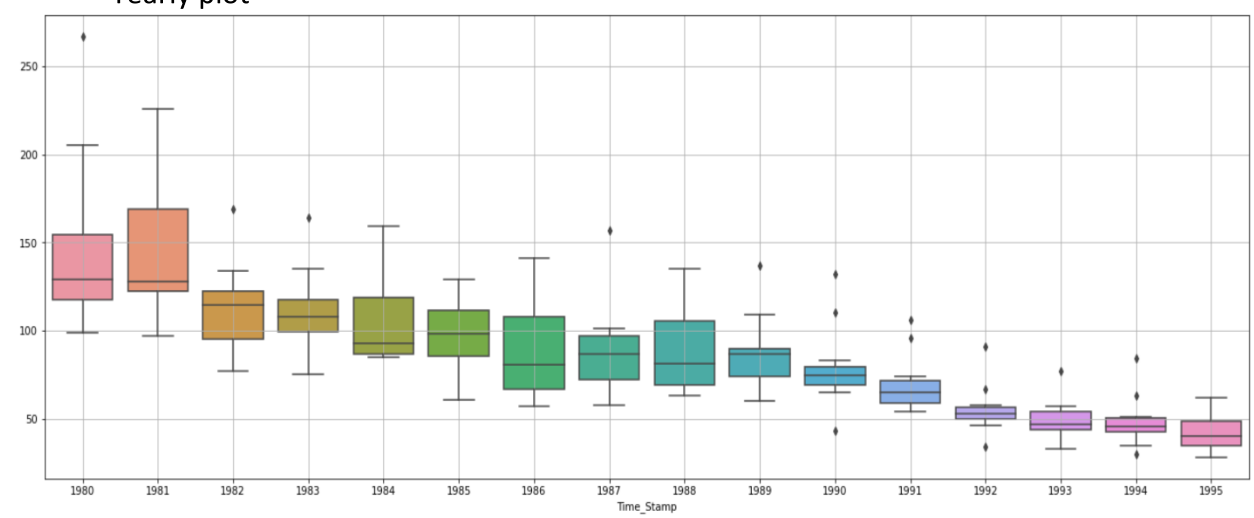


- Data consist of 187 data points
- It seems to be contain seasonality

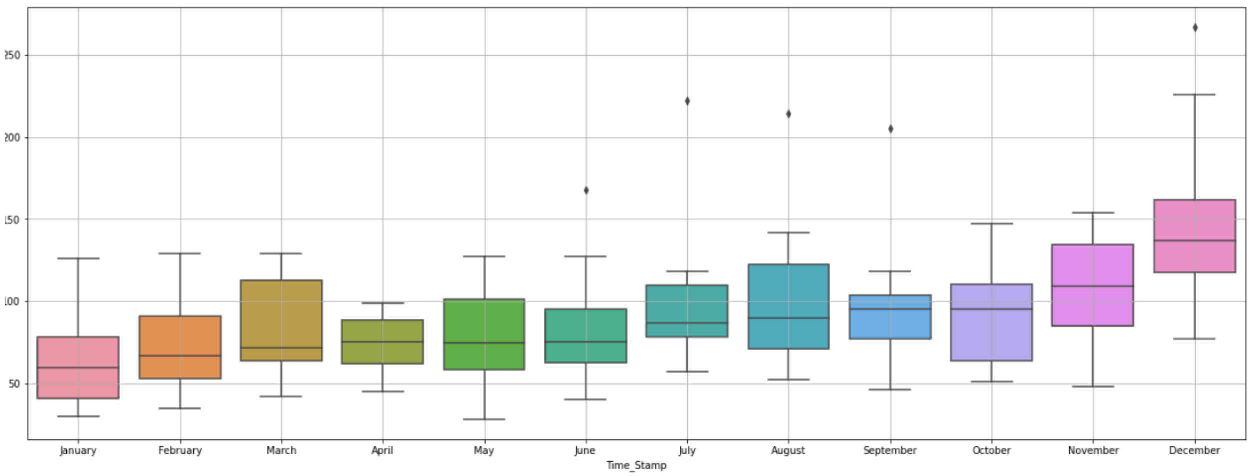
2. Perform appropriate Exploratory Data Analysis to understand the data and also perform decomposition.

There is a huge increase in 1981 and slowly decrease in the rose data. After that, there is a rise and fall equally. Seasonality is seen from the stable fluctuations repeating over the data.

- To understand the spread of the data, we use plotting.
- Boxplot helps to check the outliers in each year and month.
- Yearly plot

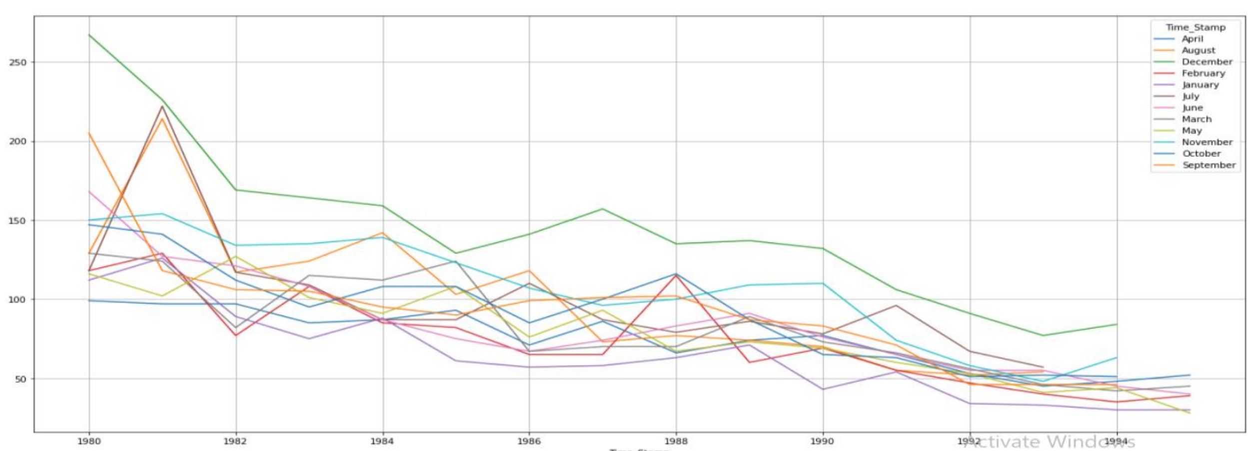


- We can clearly see some of the outliers in the plot.
- Monthly plot
- The box plot for various months is plotted
- Monthly plot contains outliers in the month of June, July, August, September and December.



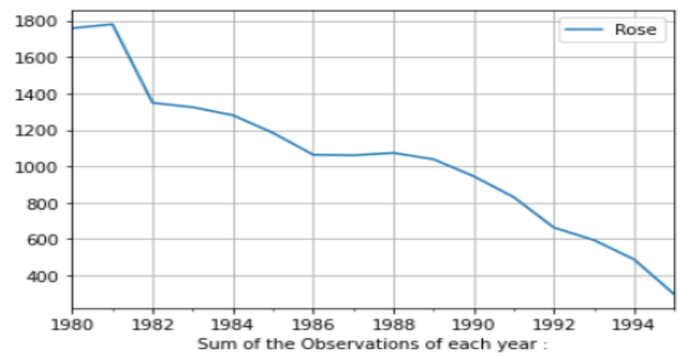
- Plot for different months and different years

Time_Stamp	April	August	December	February	January	July	June	March	May	November	October	September
Time_Stamp												
1980	99.0	129.0	267.0	118.0	112.0	118.0	168.0	129.0	116.0	150.0	147.0	205.0
1981	97.0	214.0	226.0	129.0	126.0	222.0	127.0	124.0	102.0	154.0	141.0	118.0
1982	97.0	117.0	169.0	77.0	89.0	117.0	121.0	82.0	127.0	134.0	112.0	106.0
1983	85.0	124.0	164.0	108.0	75.0	109.0	108.0	115.0	101.0	135.0	95.0	105.0
1984	87.0	142.0	159.0	85.0	88.0	87.0	87.0	112.0	91.0	139.0	108.0	95.0
1985	93.0	103.0	129.0	82.0	61.0	87.0	75.0	124.0	108.0	123.0	108.0	90.0
1986	71.0	118.0	141.0	65.0	57.0	110.0	67.0	67.0	76.0	107.0	85.0	99.0
1987	86.0	73.0	157.0	65.0	58.0	87.0	74.0	70.0	93.0	96.0	100.0	101.0
1988	66.0	77.0	135.0	115.0	63.0	79.0	83.0	70.0	67.0	100.0	116.0	102.0
1989	74.0	74.0	137.0	60.0	71.0	86.0	91.0	89.0	73.0	109.0	87.0	87.0
1990	77.0	70.0	132.0	69.0	43.0	78.0	76.0	73.0	69.0	110.0	65.0	83.0
1991	65.0	55.0	106.0	55.0	54.0	96.0	65.0	66.0	60.0	74.0	63.0	71.0
1992	53.0	52.0	91.0	47.0	34.0	67.0	55.0	56.0	53.0	58.0	51.0	46.0
1993	45.0	54.0	77.0	40.0	33.0	57.0	55.0	46.0	41.0	48.0	52.0	46.0
1994	48.0	NaN	84.0	35.0	30.0	NaN	45.0	42.0	44.0	63.0	51.0	46.0
1995	52.0	NaN	NaN	39.0	30.0	62.0	40.0	45.0	28.0	NaN	NaN	NaN

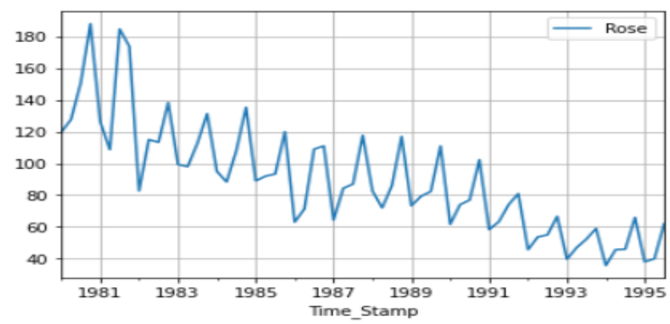


- December records have the high number of rose wine sales

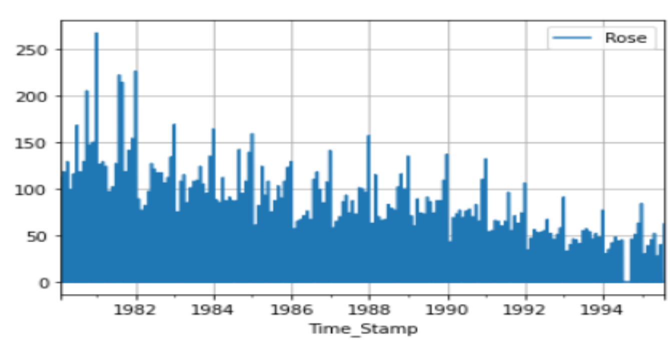
- May, January have low number of wine sales.
- Yearly Plot – aggregate the time series from an annual perspective and summing up the observations



- The plot shows that in 1982 there is a fall in the wine sales and there is a steep downfall is observed.
- Quarterly plot – aggregate the time series from a quarterly perspective and sum the observations of each quarter.

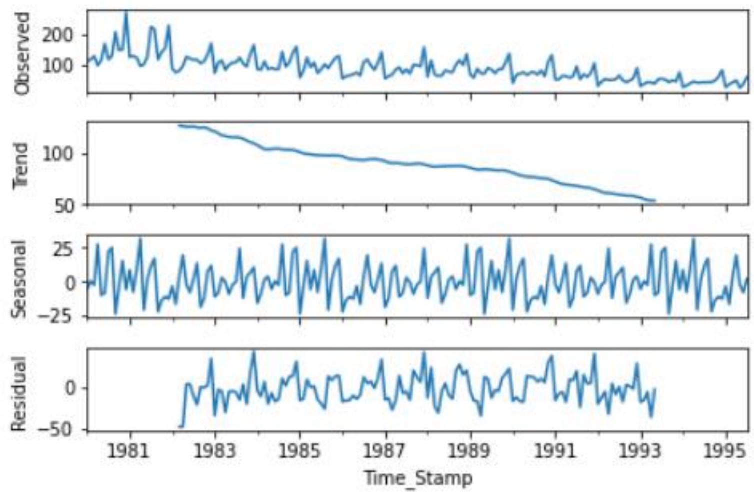


- High rise is found in 1988
- Daily plot –aggregate the data from a daily perspective
- Resampling can also be used for better overview



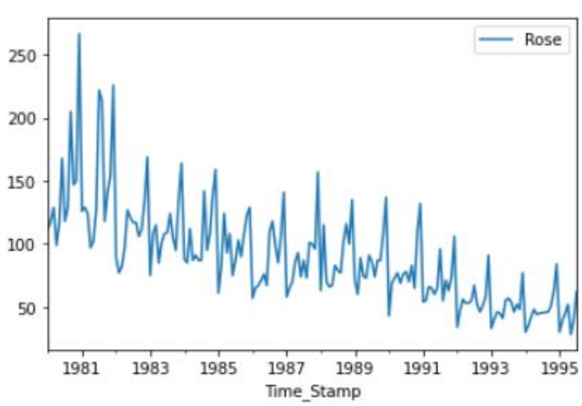
```
<class 'pandas.core.frame.DataFrame'>
DatetimeIndex: 187 entries, 1980-01-31 to 1995-07-31
Data columns (total 2 columns):
 #   Column     Non-Null Count  Dtype  
---  ------     --------------  -----  
 0   YearMonth  187 non-null    object 
 1   Rose       185 non-null    float64
dtypes: float64(1), object(1)
memory usage: 4.4+ KB
```

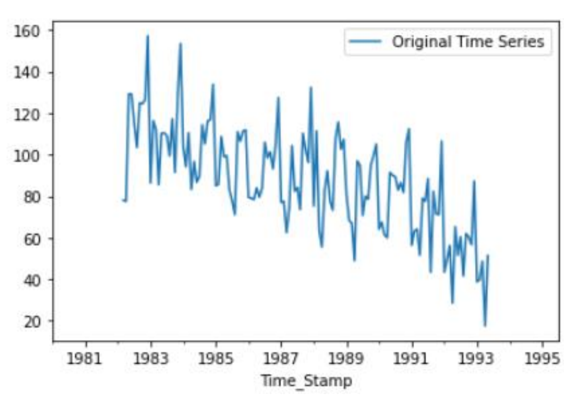
	Rose
count	185.000000
mean	90.394595
std	39.175344
min	28.000000
25%	63.000000
50%	86.000000
75%	112.000000
max	267.000000



- From the decomposition, there is seasonality in the data.

```
Seasonality
                  Rose
Time_Stamp
1980-01-31   -5.661140
1980-02-29    0.268347
1980-03-31   -2.603448
1980-04-30   27.640142
1980-05-31  -10.257294
1980-06-30   -8.308576
1980-07-31   22.162578
1980-08-31   25.109693
1980-09-30  -23.981653
1980-10-31   -3.750883
1980-11-30   15.316424
1980-12-31   -5.947999
```

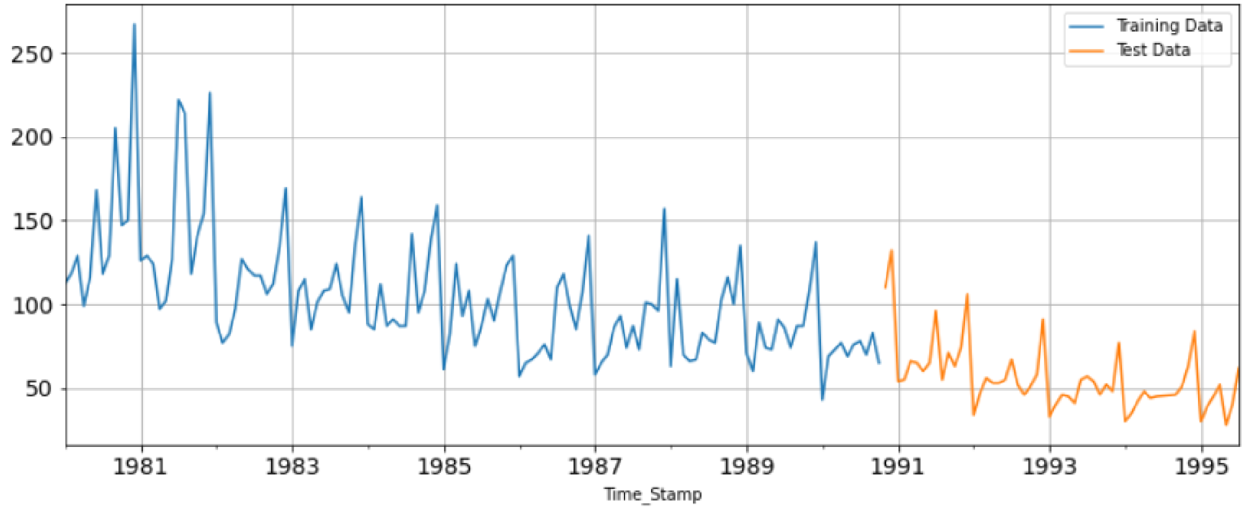




3. Split the data into training and test. The test data should start in 1991.

Train and test shapes

```
(130, 1)
(57, 1)
```



- The test data starts from 1991
- It is difficult to predict the future if the past is not happened. From the above split, we are predicting similar to the past data.

4. Build various exponential smoothing models on the training data and evaluate the model using RMSE on the test data. Other models such as regression,naïve forecast models, simple average models etc. should also be built on the training data and check the performance on the test data using RMSE.

Model1: Linear Regression

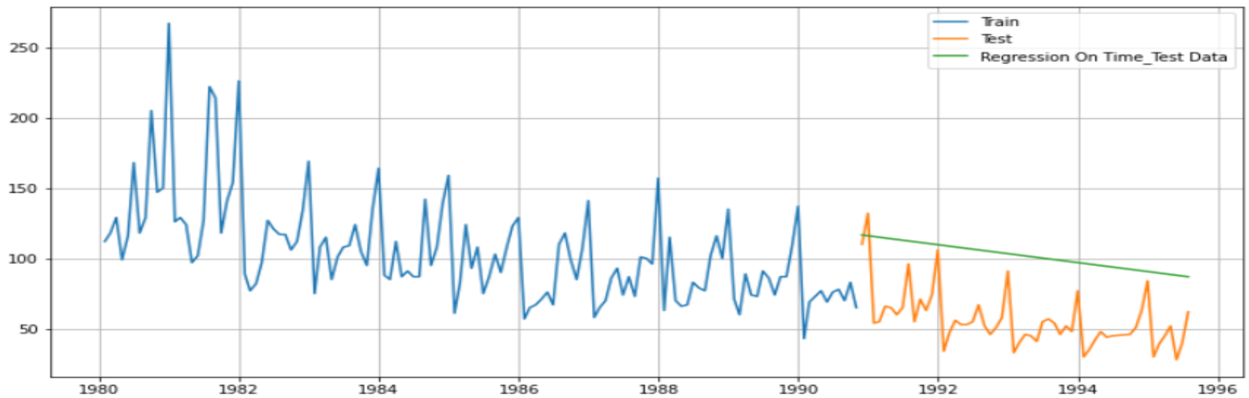
- Regress the "Rose" variable against the order of occurrence.
- Modifying the training set
- Generate the numerical instance order for both training and test set
- Printing the head and tail of test and train data

```
First few rows of Training Data
             Rose  time
Time_Stamp
1980-01-31  112.0     1
1980-02-29  118.0     2
1980-03-31  129.0     3
1980-04-30   99.0     4
1980-05-31  116.0     5

Last few rows of Training Data
            Rose  time
Time_Stamp
1990-06-30  76.0   126
1990-07-31  78.0   127
1990-08-31  70.0   128
1990-09-30  83.0   129
1990-10-31  65.0   130
```

```
First few rows of Test Data
             Rose  time
Time_Stamp
1990-11-30  110.0    43
1990-12-31  132.0    44
1991-01-31   54.0    45
1991-02-28   55.0    46
1991-03-31   66.0    47

Last few rows of Test Data
            Rose  time
Time_Stamp
1995-03-31  45.0    95
1995-04-30  52.0    96
1995-05-31  28.0    97
1995-06-30  40.0    98
1995-07-31  62.0    99
```

- Linear Regression is built on the training and test dataset

1992-12-31	91.000000	68	103.369335
1993-01-31	33.000000	69	102.840145
1993-02-28	40.000000	70	102.310956
1993-03-31	46.000000	71	101.781767
1993-04-30	45.000000	72	101.252578
1993-05-31	41.000000	73	100.723388
1993-06-30	55.000000	74	100.194199



- Defining the accuracy metrics
- Evaluating the model

```
For RegressionOnTime forecast on the Test Data,  RMSE is 48.754
```

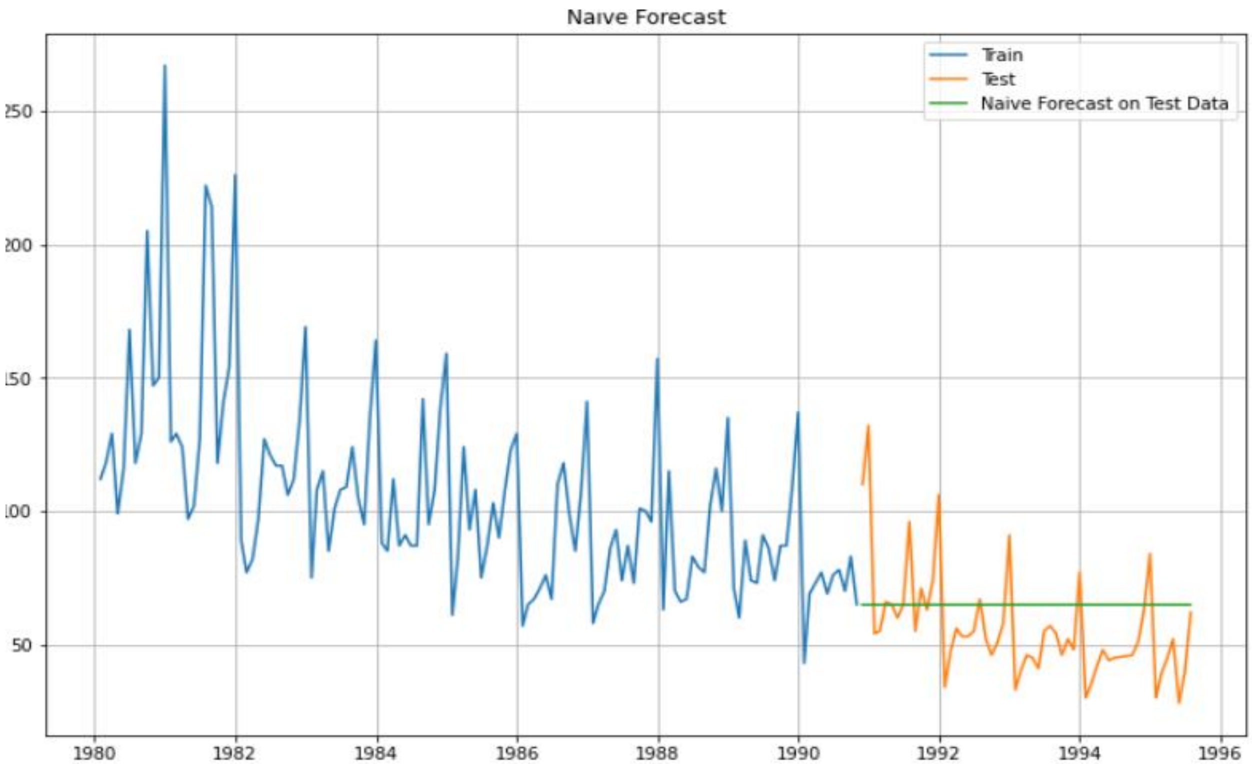
	Test RMSE
RegressionOnTime	48.754146

Model2 – Naïve model

We say that the prediction for tomorrow is the same as today and the prediction for day after tomorrow is tomorrow

```
129
```

```
Time_Stamp
1990-11-30    65.0
1990-12-31    65.0
1991-01-31    65.0
1991-02-28    65.0
1991-03-31    65.0
Name: naive, dtype: float64
```



For RegressionOnTime forecast on the Test Data, RMSE is 21.767

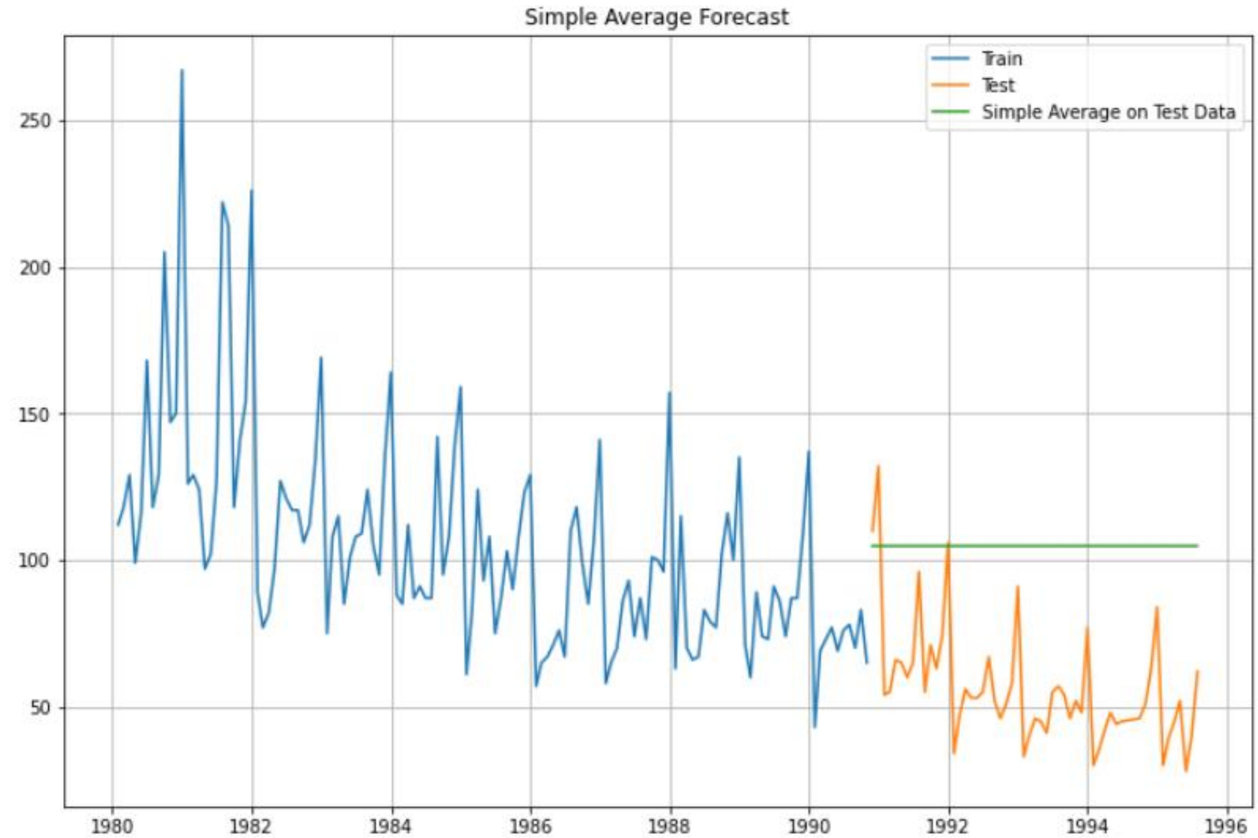
	Test RMSE
RegressionOnTime	48.754146
NaiveModel	21.766930

Model3 – Simple Average – Forecast using the average of training values

Time_Stamp	Rose	mean_forecast
1990-11-30	110.0	104.692308
1990-12-31	132.0	104.692308
1991-01-31	54.0	104.692308
1991-02-28	55.0	104.692308
1991-03-31	66.0	104.692308

For Simple Average forecast on the Test Data, RMSE is 52.412

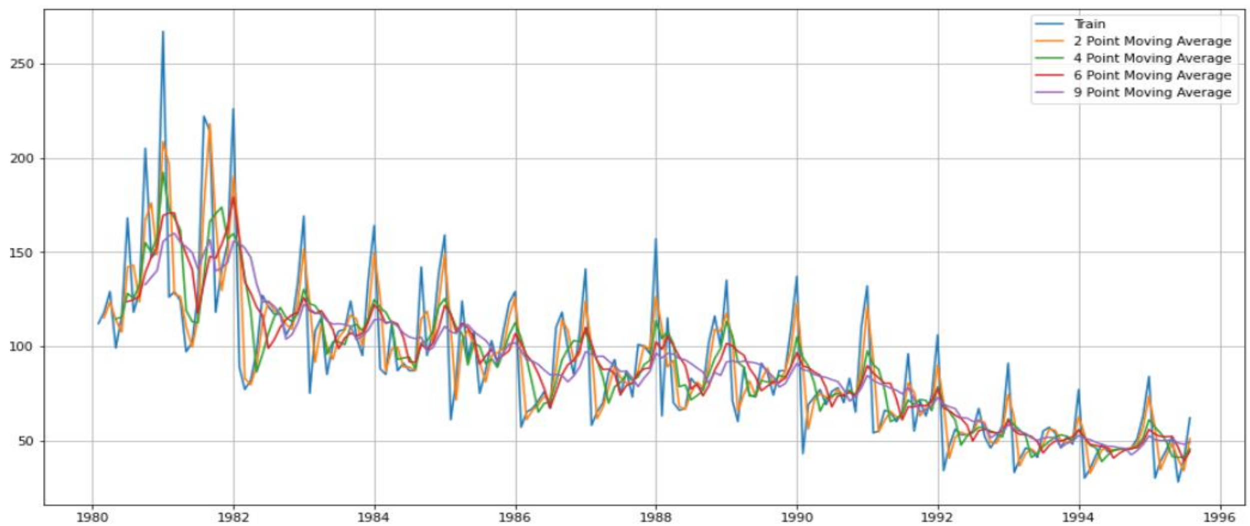
	Test RMSE
RegressionOnTime	48.754146
NaiveModel	21.766930
SimpleAverageModel	52.412093



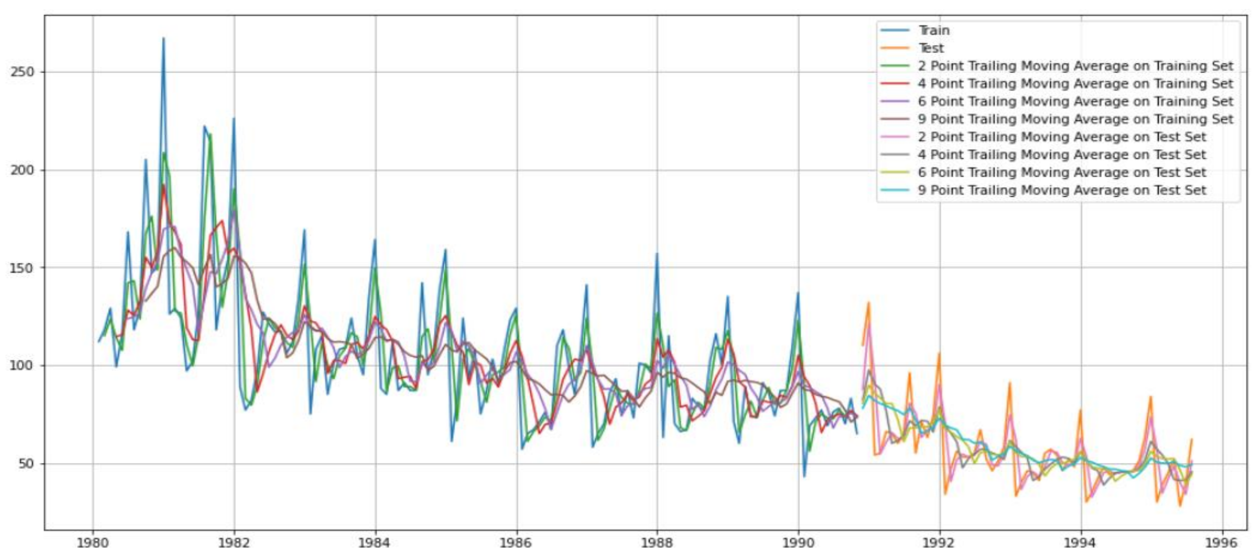
Model4- Moving Average – Calculating the rolling means (or moving averages) for different intervals. The best interval can be determined by the maximum accuracy (or the minimum error) over here.

Time_Stamp	Rose	Trailing_2	Trailing_4	Trailing_6	Trailing_9
1980-01-31	112.0	NaN	NaN	NaN	NaN
1980-02-29	118.0	115.0	NaN	NaN	NaN
1980-03-31	129.0	123.5	NaN	NaN	NaN
1980-04-30	99.0	114.0	114.5	NaN	NaN
1980-05-31	116.0	107.5	115.5	NaN	NaN

```
For 2 point Moving Average Model forecast on the Training Data,  RMSE is 11.801
For 4 point Moving Average Model forecast on the Training Data,  RMSE is 15.367
For 6 point Moving Average Model forecast on the Training Data,  RMSE is 15.862
For 9 point Moving Average Model forecast on the Training Data,  RMSE is 16.342
```

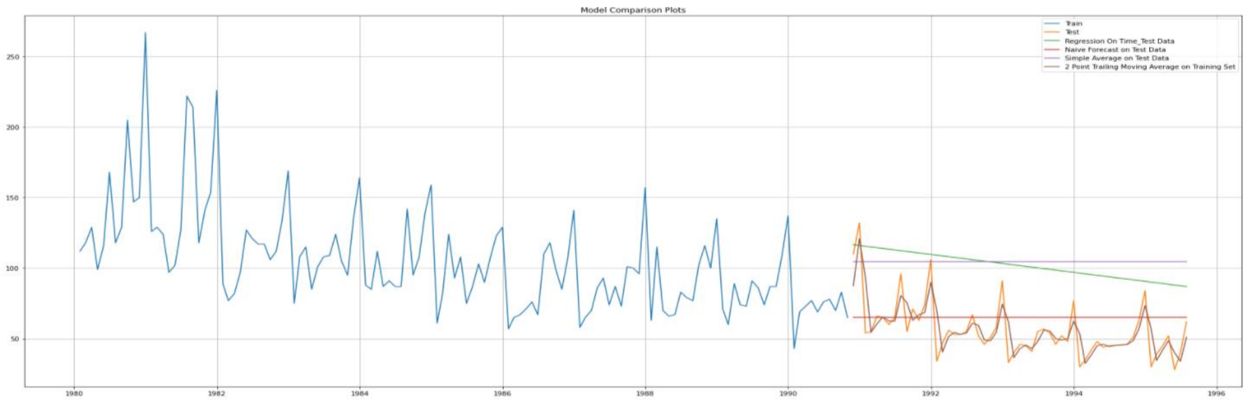


Let us split the data into train and test and plot this Time Series. The window of the moving average is need to be carefully selected as too big a window will result in not having any test set as the whole series might get averaged over.



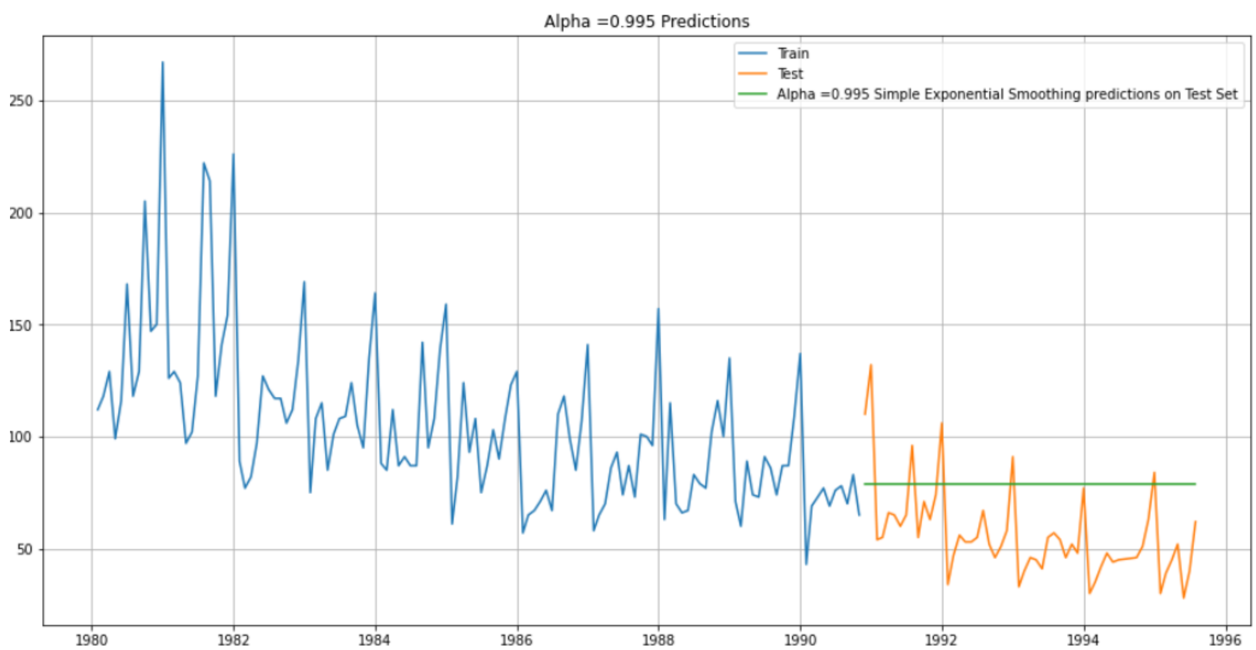
	Test RMSE
RegressionOnTime	48.754146
NaiveModel	21.766930
SimpleAverageModel	52.412093
2pointTrailingMovingAverage	11.801043
4pointTrailingMovingAverage	15.367212
6pointTrailingMovingAverage	15.862350
9pointTrailingMovingAverage	16.341919

Before we go on to build the various Exponential Smoothing models, let us plot all the models and compare the Time Series plots



Model -5- Exponential Smoothing

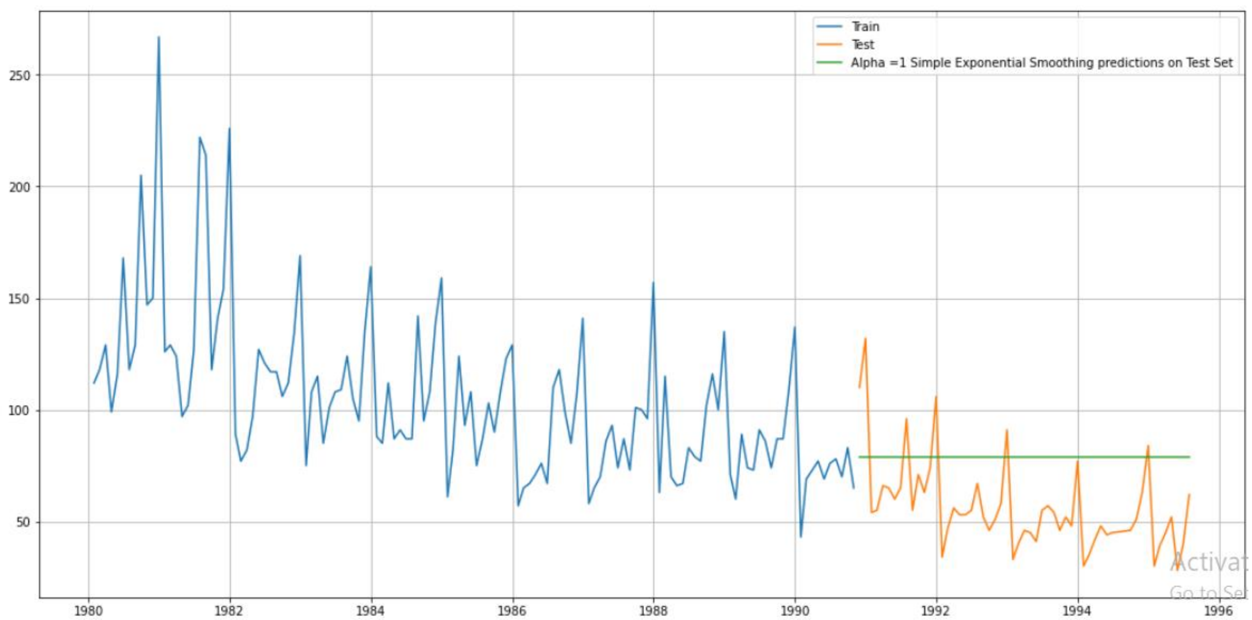
Time_Stamp	Rose	predict
1990-11-30	110.0	78.89952
1990-12-31	132.0	78.89952
1991-01-31	54.0	78.89952
1991-02-28	55.0	78.89952
1991-03-31	66.0	78.89952



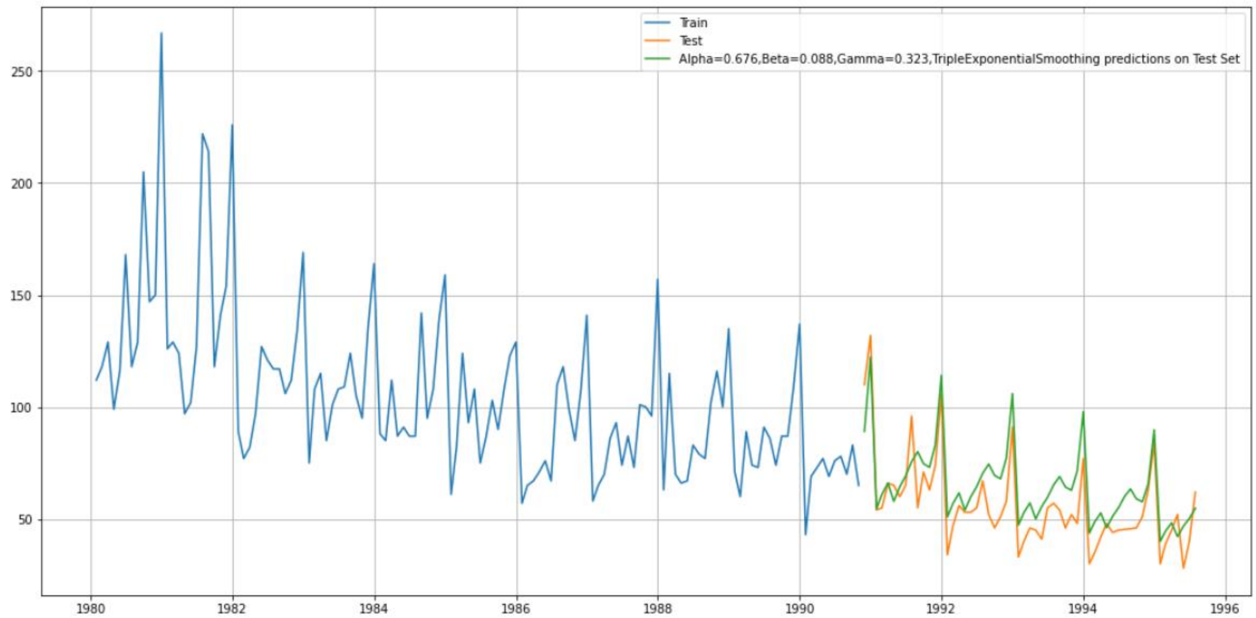
```
For Alpha =0.995 Simple Exponential Smoothing Model forecast on the Test Data, RMSE is 30.188
```

	Test RMSE
RegressionOnTime	48.754146
NaiveModel	21.766930
SimpleAverageModel	52.412093
2pointTrailingMovingAverage	11.801043
4pointTrailingMovingAverage	15.367212
6pointTrailingMovingAverage	15.862350
9pointTrailingMovingAverage	16.341919
Alpha=0.995,SimpleExponentialSmoothing	30.188322

Setting different alpha values. Higher the alpha, the more weightage is given to more recent observation.



Time_Stamp	Rose	auto_predict
1990-11-30	110.0	89.115658
1990-12-31	132.0	122.254704
1991-01-31	54.0	54.549530
1991-02-28	55.0	61.040149
1991-03-31	66.0	66.100106

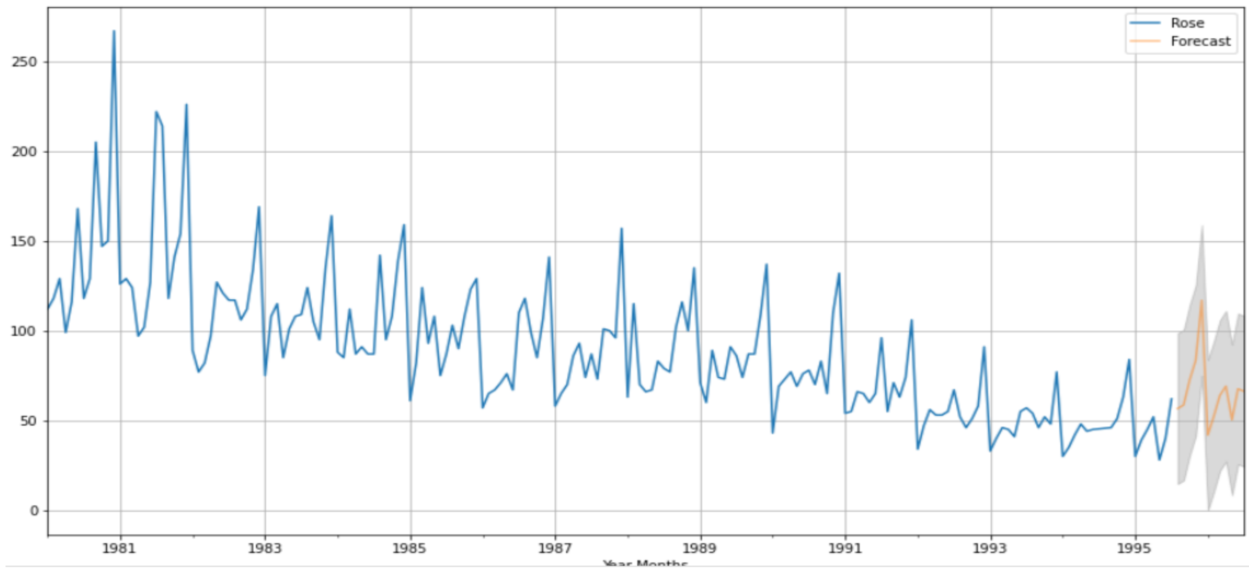


For Alpha=0.676,Beta=0.088,Gamma=0.323, Triple Exponential Smoothing Model forecast on the Test Data, RMSE is 12.666

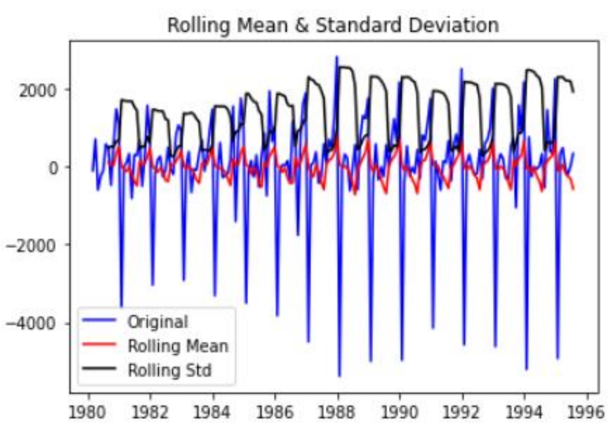
	Test RMSE
RegressionOnTime	48.754146
NaiveModel	21.766930
SimpleAverageModel	52.412093
2pointTrailingMovingAverage	11.801043
4pointTrailingMovingAverage	15.367212
6pointTrailingMovingAverage	15.862350
9pointTrailingMovingAverage	16.341919
Alpha=0.995,SimpleExponentialSmoothing	30.188322
Alpha=0.676,Beta=0.088,Gamma=0.323,TripleExponentialSmoothing	12.666350

	Test RMSE
2pointTrailingMovingAverage	11.801043
Alpha=0.676,Beta=0.088,Gamma=0.323,TripleExponentialSmoothing	12.666350
4pointTrailingMovingAverage	15.367212
6pointTrailingMovingAverage	15.862350
9pointTrailingMovingAverage	16.341919
NaiveModel	21.766930
Alpha=0.995,SimpleExponentialSmoothing	30.188322
RegressionOnTime	48.754146
SimpleAverageModel	52.412093

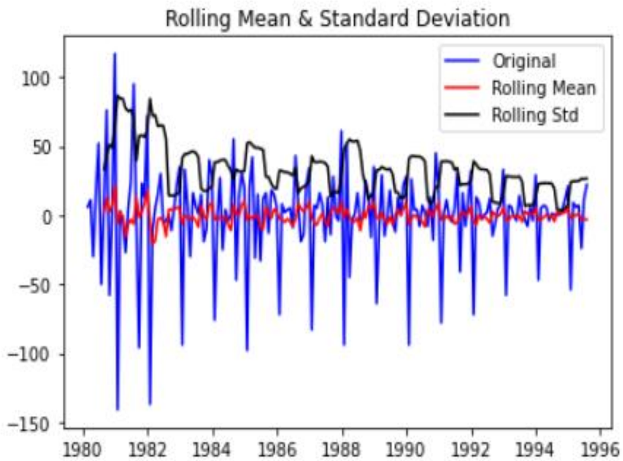
	lower_CI	prediction	upper_ci
1995-08-31	8.740506	50.719439	92.698373
1995-09-30	8.148766	50.127699	92.106633
1995-10-31	14.391781	56.370714	98.349648
1995-11-30	24.477306	66.456239	108.435173
1995-12-31	53.865719	95.844653	137.823586



5. Check for the stationarity of the data on which the model is being built on using appropriate statistical tests and also mention the hypothesis for the statistical test. If the data is found to be non-stationary, take appropriate steps to make it stationary. Check the new data for stationarity and comment. Note: Stationarity should be checked at alpha = 0.05.




```
Results of Dickey-Fuller Test:
Test Statistic                  -1.876699
p-value                          0.343101
#Lags Used                      13.000000
Number of Observations Used    173.000000
Critical Value (1%)             -3.468726
Critical Value (5%)             -2.878396
Critical Value (10%)            -2.575756
dtype: float64
```

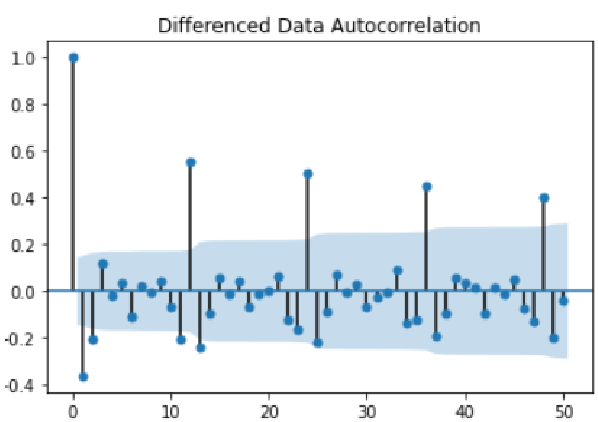


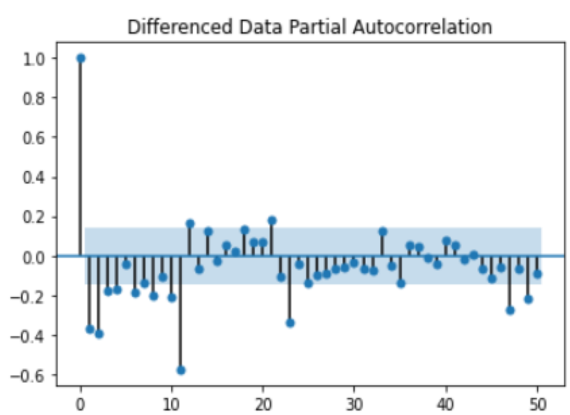

```
Results of Dickey-Fuller Test:
Test Statistic                 -8.044392e+00
p-value                         1.810895e-12
#Lags Used                      1.200000e+01
Number of Observations Used     1.730000e+02
Critical Value (1%)            -3.468726e+00
Critical Value (5%)            -2.878396e+00
Critical Value (10%)           -2.575756e+00
dtype: float64
```

- When the time series data is not stationary we need to convert it into stationary before applying models.
- We use Augmented Dickey fuller test.
- It determines how strongly a time series is defined by the trend.
- From the null and alternate hypothesis, we define time series data is stationary or not.
- We see that 5% significant level the time series is non-stationarity
- P value >0.05 – Failed to reject null hypothesis - Stationary
- Let us take a difference of order 1 and check whether the Time Series is stationary or not
- At α = 0.05 the Time Series is indeed stationary.

6. Build an automated version of the ARIMA/SARIMA model in which the parameters are selected using the lowest Akaike Information Criteria (AIC) on the training data and evaluate this model on the test data using RMSE.

7. Build ARIMA/SARIMA models based on the cut-off points of ACF and PACF on the training data and evaluate this model on the test data using RMSE.





- From the above plot, we see seasonality in the data.

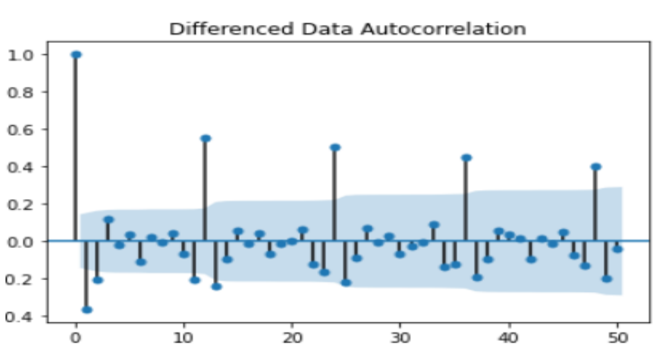
```
ARIMA(0, 1, 0) - AIC:1315.1645929217334
ARIMA(0, 1, 1) - AIC:1258.039048575531
ARIMA(0, 1, 2) - AIC:1254.786677530965
ARIMA(1, 1, 0) - AIC:1299.0560243200116
ARIMA(1, 1, 1) - AIC:1255.6046151581302
ARIMA(1, 1, 2) - AIC:1255.2709084666585
ARIMA(2, 1, 0) - AIC:1280.0806971753402
ARIMA(2, 1, 1) - AIC:1256.862594546858
ARIMA(2, 1, 2) - AIC:1257.2401826353614
```

	param	AIC
2	(0, 1, 2)	1254.786678
5	(1, 1, 2)	1255.270908
4	(1, 1, 1)	1255.604615
7	(2, 1, 1)	1256.862595
8	(2, 1, 2)	1257.240183
1	(0, 1, 1)	1258.039049
6	(2, 1, 0)	1280.080697
3	(1, 1, 0)	1299.056024
0	(0, 1, 0)	1315.164593

- If we have seasonality, then we should go for SARIMA model.
- We are building ARIMA model by looking at minimum AIC values and ACF and PACF plots.
- Sorting the AIC values to see the lower AIC value.

```
                             ARIMA Model Results
==============================================================================
Dep. Variable:                 D.Rose   No. Observations:                  129
Model:                 ARIMA(2, 1, 1)   Log Likelihood                -623.431
Method:                       css-mle   S.D. of innovations             29.839
Date:                Fri, 02 Jul 2021   AIC                           1256.863
Time:                        06:31:10   BIC                           1271.162
Sample:                    02-29-1980   HQIC                          1262.673
                         - 10-31-1990
================================================================================
                   coef    std err          z      P>|z|      [0.025      0.975]
--------------------------------------------------------------------------------
const           -0.5291      0.079     -6.673      0.000      -0.685      -0.374
ar.L1.D.Rose     0.1971      0.088      2.240      0.027       0.025       0.369
ar.L2.D.Rose    -0.0758      0.088     -0.863      0.390      -0.248       0.096
ma.L1.D.Rose    -1.0000      0.032    -31.129      0.000      -1.063      -0.937
                                    Roots
=============================================================================
                  Real          Imaginary           Modulus         Frequency
-----------------------------------------------------------------------------
AR.1            1.2999           -3.3914j            3.6319           -0.1917
AR.2            1.2999           +3.3914j            3.6319            0.1917
MA.1            1.0000           +0.0000j            1.0000            0.0000
```

	RMSE
ARIMA(2,1,1)	17.436875



- Again we plot ACF to see and understand the seasonal parameter of SARIMA model.
- First iteration by setting 6 as the seasonality
- We sort the AIC values to see the lowest of all vales.
- Next predicting the data using the SARIMA model and evaluating the model.
- We get the summary of the data

```
Examples of some parameter combinations for Model.
Model: (0, 1, 1)(0, 0, 1, 6)
Model: (0, 1, 2)(0, 0, 2, 6)
Model: (1, 1, 0)(1, 0, 0, 6)
Model: (1, 1, 1)(1, 0, 1, 6)
Model: (1, 1, 2)(1, 0, 2, 6)
Model: (2, 1, 0)(2, 0, 0, 6)
Model: (2, 1, 1)(2, 0, 1, 6)
Model: (2, 1, 2)(2, 0, 2, 6)
```

```
SARIMA(2, 1, 1)x(2, 0, 2, 6) - AIC:1034.1310578958723
SARIMA(2, 1, 2)x(0, 0, 0, 6) - AIC:1233.5045954992431
SARIMA(2, 1, 2)x(0, 0, 1, 6) - AIC:1165.708439811226
SARIMA(2, 1, 2)x(0, 0, 2, 6) - AIC:1065.2709643814117
SARIMA(2, 1, 2)x(1, 0, 0, 6) - AIC:1179.9266583946824
SARIMA(2, 1, 2)x(1, 0, 1, 6) - AIC:1128.9638001128374
SARIMA(2, 1, 2)x(1, 0, 2, 6) - AIC:1043.9776059292521
SARIMA(2, 1, 2)x(2, 0, 0, 6) - AIC:1056.3458634986323
SARIMA(2, 1, 2)x(2, 0, 1, 6) - AIC:1051.9522744457843
SARIMA(2, 1, 2)x(2, 0, 2, 6) - AIC:1029.198822670194
```

	param	seasonal	AIC
26	(0, 1, 2)	(2, 0, 2, 6)	1025.823325
80	(2, 1, 2)	(2, 0, 2, 6)	1029.198823
53	(1, 1, 2)	(2, 0, 2, 6)	1031.313972
71	(2, 1, 1)	(2, 0, 2, 6)	1034.131058
44	(1, 1, 1)	(2, 0, 2, 6)	1034.920837

```
                                     Statespace Model Results
==========================================================================================
Dep. Variable:                                  y   No. Observations:                  130
Model:             SARIMAX(0, 1, 2)x(2, 0, 2, 6)   Log Likelihood                -505.912
Date:                            Fri, 02 Jul 2021   AIC                           1025.823
Time:                                    06:32:32   BIC                           1044.977
Sample:                                         0   HQIC                          1033.597
                                            - 130
Covariance Type:                              opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ma.L1         -0.7887    652.856     -0.001      0.999   -1280.364    1278.786
ma.L2         -0.2113    137.890     -0.002      0.999    -270.470     270.047
ar.S.L6       -0.0636      0.038     -1.692      0.091      -0.137       0.010
ar.S.L12       0.8373      0.042     20.141      0.000       0.756       0.919
ma.S.L6        0.2034    652.867      0.000      1.000   -1279.393    1279.799
ma.S.L12      -0.7966    520.014     -0.002      0.999   -1020.005    1018.412
sigma2       345.1849      3.084    111.926      0.000     339.140     351.230
===================================================================================
Ljung-Box (Q):                       24.55   Jarque-Bera (JB):                94.11
Prob(Q):                              0.97   Prob(JB):                         0.00
Heteroskedasticity (H):               0.41   Skew:                             0.42
Prob(H) (two-sided):                  0.01   Kurtosis:                         7.37
```

y	mean	mean_se	mean_ci_lower	mean_ci_upper
0	85.176012	19.135821	47.670491	122.681532
1	109.402253	19.605757	70.975675	147.828831
2	66.211885	19.605759	27.785304	104.638466
3	67.805730	19.607598	29.375545	106.235915
4	75.481847	19.599286	37.067951	113.895742

	RMSE
ARIMA(2,1,1)	17.436875
SARIMA(0,1,2)(2,0,2,6)	25.766527

- There is a huge gain in the RMSE value by including seasonal parameters
- Keeping 12 as seasonal parameter for second iteration

	param	seasonal	AIC
26	(0, 1, 2)	(2, 0, 2, 12)	871.075238
53	(1, 1, 2)	(2, 0, 2, 12)	873.003875
80	(2, 1, 2)	(2, 0, 2, 12)	874.213962
69	(2, 1, 1)	(2, 0, 0, 12)	879.792363
78	(2, 1, 2)	(2, 0, 0, 12)	880.763857

```
                                     Statespace Model Results
==========================================================================================
Dep. Variable:                                  y   No. Observations:                  130
Model:             SARIMAX(1, 1, 2)x(2, 0, 2, 12)   Log Likelihood                -428.502
Date:                            Fri, 02 Jul 2021   AIC                            873.004
Time:                                    06:35:13   BIC                            894.004
Sample:                                         0   HQIC                           881.507
                                            - 130
Covariance Type:                              opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.1002      0.350      0.286      0.775      -0.587       0.787
ma.L1         -0.9391    484.578     -0.002      0.998    -950.694     948.815
ma.L2         -0.0609     29.580     -0.002      0.998     -58.036      57.915
ar.S.L12       0.3490      0.077      4.534      0.000       0.198       0.500
ar.S.L24       0.3067      0.073      4.193      0.000       0.163       0.450
ma.S.L12       0.0506      0.133      0.379      0.705      -0.211       0.312
ma.S.L24      -0.0896      0.146     -0.615      0.539      -0.375       0.196
sigma2       250.4588   1.21e+05      0.002      0.998   -2.38e+05    2.38e+05
===================================================================================
Ljung-Box (Q):                       23.26   Jarque-Bera (JB):                 2.89
Prob(Q):                              0.98   Prob(JB):                         0.24
Heteroskedasticity (H):               0.88   Skew:                             0.41
Prob(H) (two-sided):                  0.70   Kurtosis:                         3.02
```

y	mean	mean_se	mean_ci_lower	mean_ci_upper
0	91.010006	15.904143	59.838457	122.181554
1	114.697036	16.134059	83.074861	146.319211
2	60.835502	16.138989	29.203665	92.467339
3	70.598439	16.139302	38.965988	102.230891
4	76.889513	16.139329	45.257010	108.522017

8. Build a table (create a data frame) with all the models built along with their corresponding parameters and the respective RMSE values on the test data.

	RMSE
ARIMA(2,1,1)	17.436875
SARIMA(0,1,2)(2,0,2,6)	25.766527
SARIMA(1,1,2)(2,0,2,12)	25.420876

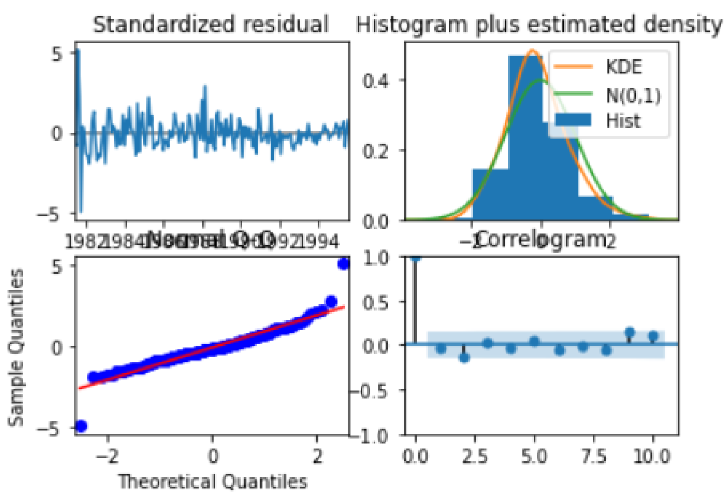
- It is clear that SARIMA(0,1,2)(2,0,2,6) has the higher RMSE and ARIMA(2,1,1) has the lowest value.

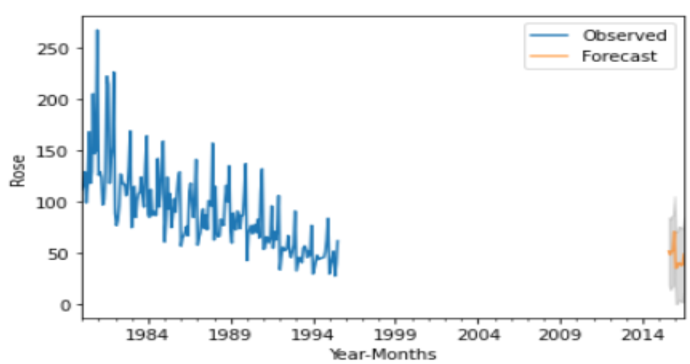
9. Based on the model-building exercise, build the most optimum model(s) on the complete data and predict 12 months into the future with appropriate confidence intervals/bands.

```
                                  Statespace Model Results
==========================================================================================
Dep. Variable:                                 Rose   No. Observations:                187
Model:             SARIMAX(0, 1, 2)x(2, 0, 2, 6)   Log Likelihood                -734.147
Date:                              Fri, 02 Jul 2021   AIC                           1482.294
Time:                                      06:35:49   BIC                           1504.286
Sample:                                  01-31-1980   HQIC                          1491.218
                                       - 07-31-1995
Covariance Type:                                opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ma.L1         -0.7297      0.070    -10.350      0.000      -0.868      -0.592
ma.L2         -0.1899      0.066     -2.883      0.004      -0.319      -0.061
ar.S.L6       -0.0496      0.029     -1.687      0.092      -0.107       0.008
ar.S.L12       0.8766      0.030     29.411      0.000       0.818       0.935
ma.S.L6        0.1746      0.235      0.744      0.457      -0.286       0.635
ma.S.L12      -0.7825      0.194     -4.033      0.000      -1.163      -0.402
sigma2       283.5320     58.864      4.817      0.000     168.161     398.903
===================================================================================
Ljung-Box (Q):                       26.88   Jarque-Bera (JB):               297.19
Prob(Q):                              0.94   Prob(JB):                         0.00
Heteroskedasticity (H):               0.17   Skew:                             0.45
Prob(H) (two-sided):                  0.00   Kurtosis:                         9.40
===================================================================================
```

Rose	mean	mean_se	mean_ci_lower	mean_ci_upper
1995-08-31	51.906635	16.960539	18.664588	85.148681
1995-09-30	48.560882	17.562344	14.139321	82.982443
1995-10-31	51.876503	17.615197	17.351352	86.401654
1995-11-30	52.523982	17.667892	17.895550	87.152413
1995-12-31	70.922129	17.720430	36.190724	105.653533
1996-01-31	35.608491	17.772813	0.774418	70.442564
1996-02-29	36.103892	18.066583	0.694041	71.513744
1996-03-31	40.520377	18.159129	4.929139	76.111616
1996-04-30	40.337886	18.222257	4.622919	76.052853
1996-05-31	38.152511	18.285167	2.314243	73.990780
1996-06-30	38.640570	18.347861	2.679423	74.601718
1996-07-31	48.980236	18.410343	12.896628	85.063845

```
RMSE of the Full Model 28.050812588685464
```





10. Comment on the model thus built and report your findings and suggest the measures that the company should be taking for future sales.

- To find the most optimum model, we run the model on the full data
- Correlogram, histogram, residual and quartiles are shown.
- We predict for the next 12 months for next years.
- We get forecast.
- RMSE of the full complete data is 28.05
- Plotting the forecast with the confidence band
- It is clear that SARIMA(0,1,2)(2,0,2,6) has the higher RMSE and ARIMA(2,1,1) has the lower value.