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.

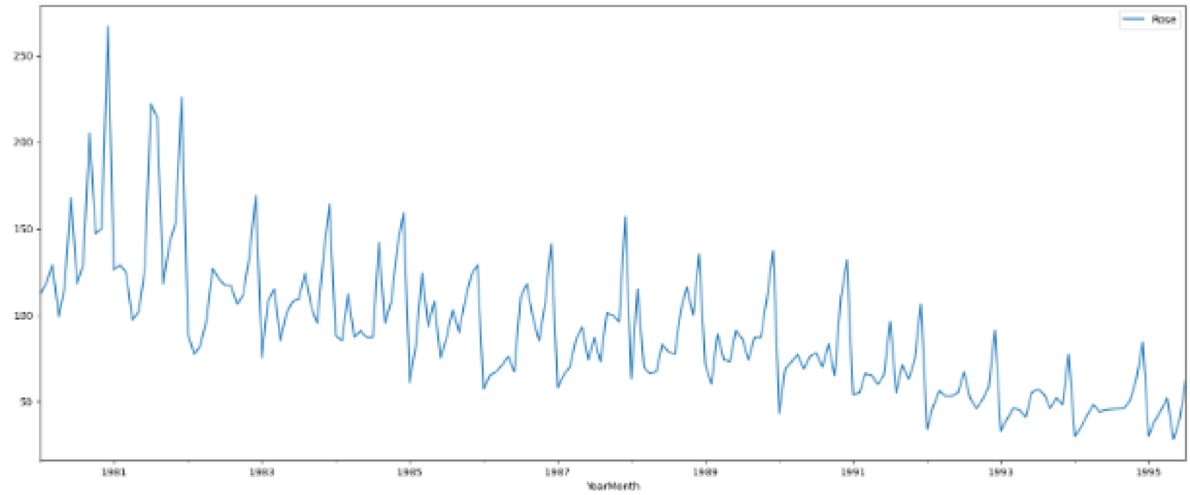
Rose.csv

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

YearMonth	Rose
1980-01-01	112.0
1980-02-01	118.0
1980-03-01	129.0
1980-04-01	99.0
1980-05-01	116.0

- Shape – (187,1)
- Null Values present, were replaced using ingerpolate.

```
<class 'pandas.core.frame.DataFrame'>
DatetimeIndex: 187 entries, 1980-01-01 to 1995-07-01
Data columns (total 1 columns):
 #   Column  Non-Null Count  Dtype  
---  ------  --------------  -----  
 0   Rose    187 non-null    float64
dtypes: float64(1)
memory usage: 2.9 KB
```



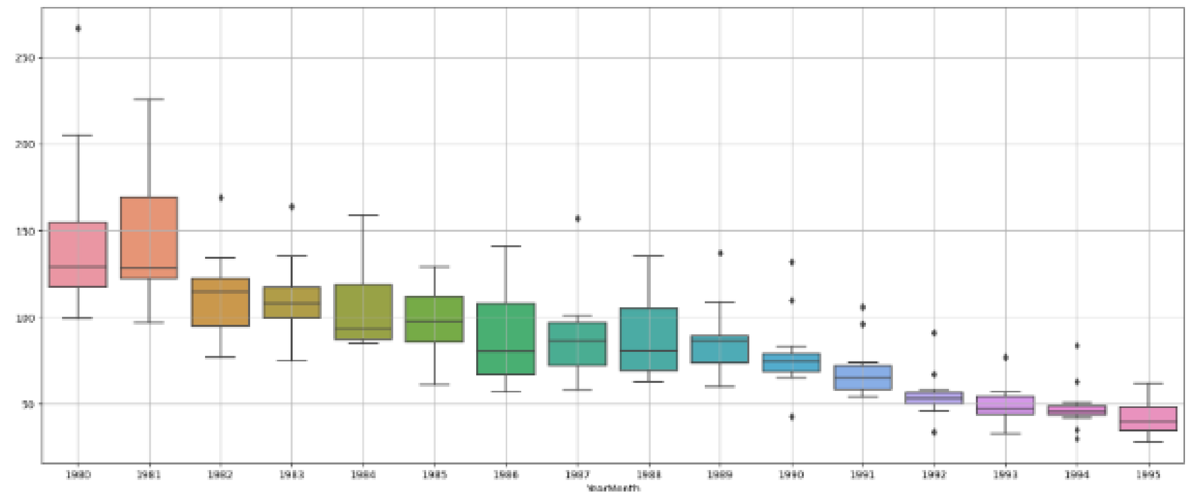
Given data is not time. So we parse the date range and create a timestamp. We also notice the fluctuations in the trend in the initial years and slowly decreasing the following years.

- 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.

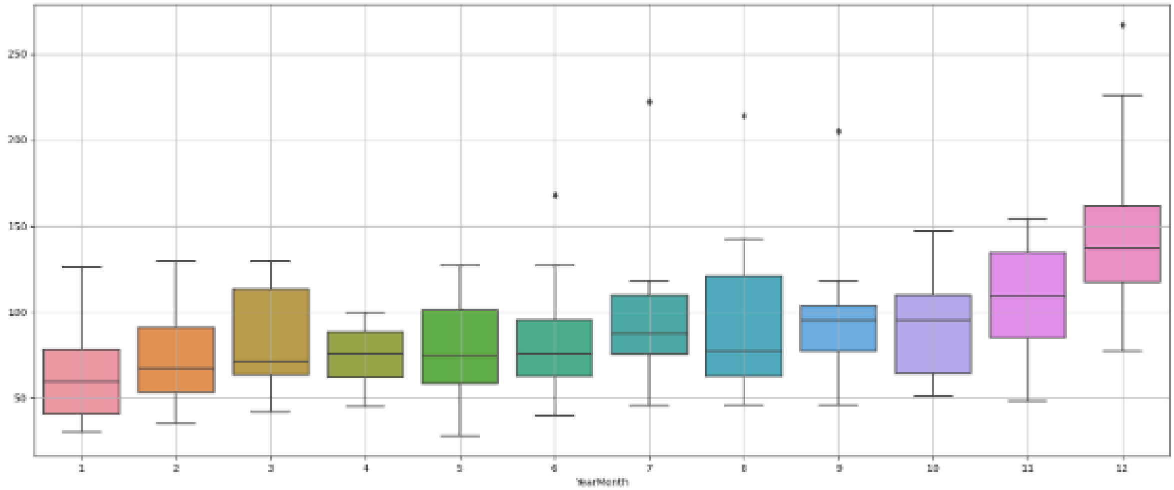
There is a huge increase in 1981 and slowly decrease in the rose data. After that, there is a rise and fall equally. Seasonality is seen from the stable fluctuations repeating over the data.

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



- We van clearly see some of the outliers in the plot.

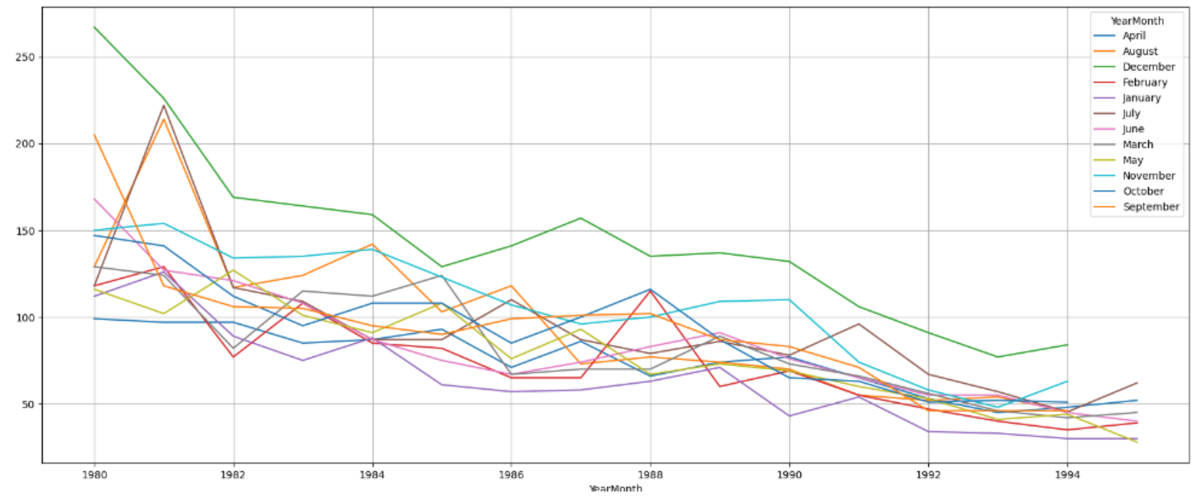
- Monthly plot



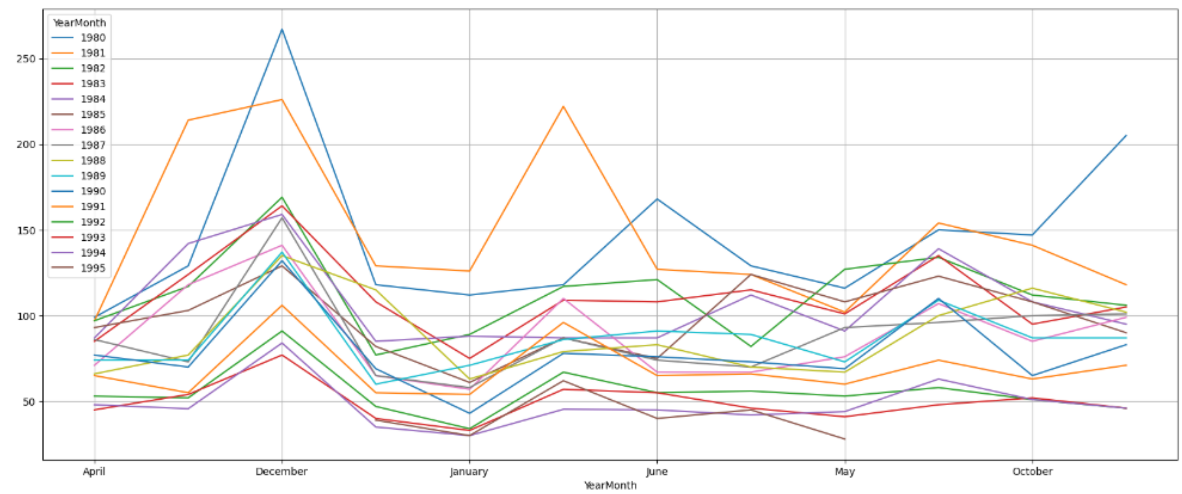
- The boxplot for various months is plotted
- Monthly plot contains outliers in the month of June, July, August, September and December.

- Plot for different months and different years

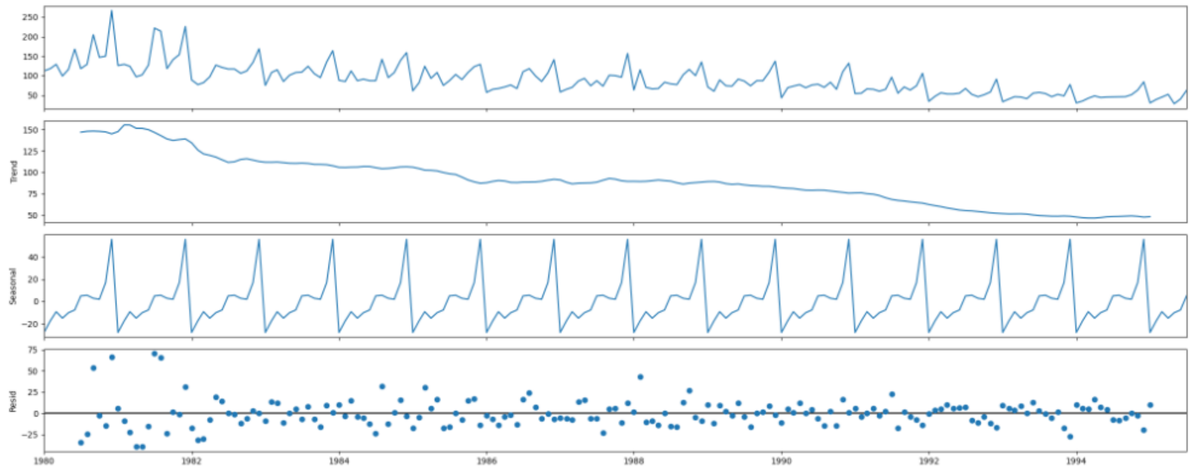
YearMonth	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995
YearMonth																
April	99.0	97.0	97.0	85.0	87.0	93.0	71.0	86.0	66.0	74.0	77.0	65.0	53.0	45.0	48.000000	52.0
August	129.0	214.0	117.0	124.0	142.0	103.0	118.0	73.0	77.0	74.0	70.0	55.0	52.0	54.0	45.666667	NaN
December	267.0	226.0	169.0	164.0	159.0	129.0	141.0	157.0	135.0	137.0	132.0	106.0	91.0	77.0	84.000000	NaN
February	118.0	129.0	77.0	108.0	85.0	82.0	65.0	65.0	115.0	60.0	69.0	55.0	47.0	40.0	35.000000	39.0
January	112.0	126.0	89.0	75.0	88.0	61.0	57.0	58.0	63.0	71.0	43.0	54.0	34.0	33.0	30.000000	30.0
July	118.0	222.0	117.0	109.0	87.0	87.0	110.0	87.0	79.0	86.0	78.0	96.0	67.0	57.0	45.333333	62.0
June	168.0	127.0	121.0	108.0	87.0	75.0	67.0	74.0	83.0	91.0	76.0	65.0	55.0	55.0	45.000000	40.0
March	129.0	124.0	82.0	115.0	112.0	124.0	67.0	70.0	70.0	89.0	73.0	66.0	56.0	46.0	42.000000	45.0
May	116.0	102.0	127.0	101.0	91.0	108.0	76.0	93.0	67.0	73.0	69.0	60.0	53.0	41.0	44.000000	28.0
November	150.0	154.0	134.0	135.0	139.0	123.0	107.0	96.0	100.0	109.0	110.0	74.0	58.0	48.0	63.000000	NaN
October	147.0	141.0	112.0	95.0	108.0	108.0	85.0	100.0	116.0	87.0	65.0	63.0	51.0	52.0	51.000000	NaN
September	205.0	118.0	106.0	105.0	95.0	90.0	99.0	101.0	102.0	87.0	83.0	71.0	46.0	46.0	46.000000	NaN

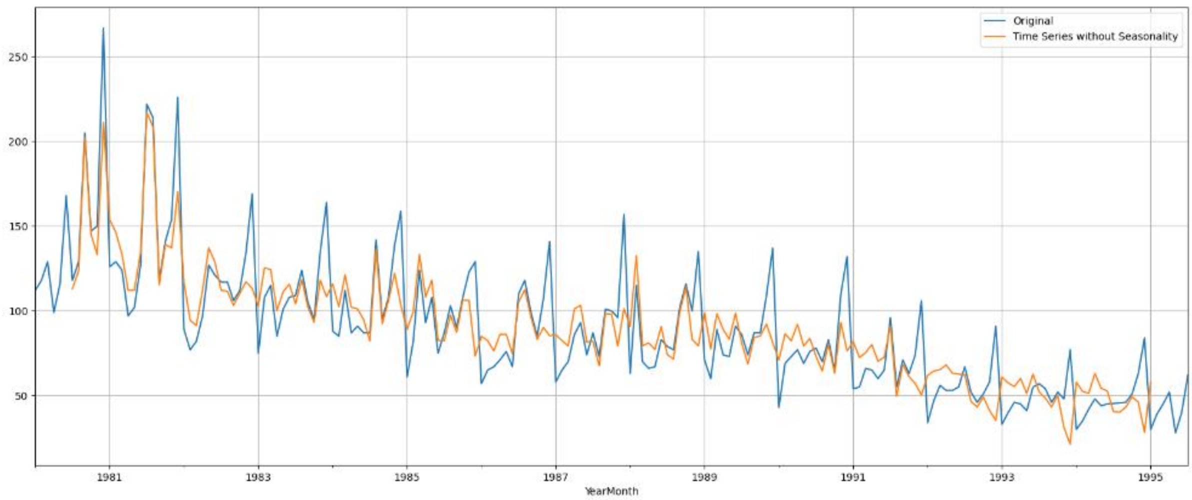


YearMonth	April	August	December	February	January	July	June	March	May	November	October	September
YearMonth												
1980	99.0	129.000000	267.0	118.0	112.0	118.000000	168.0	129.0	116.0	150.0	147.0	205.0
1981	97.0	214.000000	226.0	129.0	126.0	222.000000	127.0	124.0	102.0	154.0	141.0	118.0
1982	97.0	117.000000	169.0	77.0	89.0	117.000000	121.0	82.0	127.0	134.0	112.0	106.0
1983	85.0	124.000000	164.0	108.0	75.0	109.000000	108.0	115.0	101.0	135.0	95.0	105.0
1984	87.0	142.000000	159.0	85.0	88.0	87.000000	87.0	112.0	91.0	139.0	108.0	95.0
1985	93.0	103.000000	129.0	82.0	61.0	87.000000	75.0	124.0	108.0	123.0	108.0	90.0
1986	71.0	118.000000	141.0	65.0	57.0	110.000000	67.0	67.0	76.0	107.0	85.0	99.0
1987	86.0	73.000000	157.0	65.0	58.0	87.000000	74.0	70.0	93.0	96.0	100.0	101.0
1988	66.0	77.000000	135.0	115.0	63.0	79.000000	83.0	70.0	67.0	100.0	116.0	102.0
1989	74.0	74.000000	137.0	60.0	71.0	86.000000	91.0	89.0	73.0	109.0	87.0	87.0
1990	77.0	70.000000	132.0	69.0	43.0	78.000000	76.0	73.0	69.0	110.0	65.0	83.0
1991	65.0	55.000000	106.0	55.0	54.0	96.000000	65.0	66.0	60.0	74.0	63.0	71.0
1992	53.0	52.000000	91.0	47.0	34.0	67.000000	55.0	56.0	53.0	58.0	51.0	46.0
1993	45.0	54.000000	77.0	40.0	33.0	57.000000	55.0	46.0	41.0	48.0	52.0	46.0
1994	48.0	45.666667	84.0	35.0	30.0	45.333333	45.0	42.0	44.0	63.0	51.0	46.0
1995	52.0	NaN	NaN	39.0	30.0	62.000000	40.0	45.0	28.0	NaN	NaN	NaN

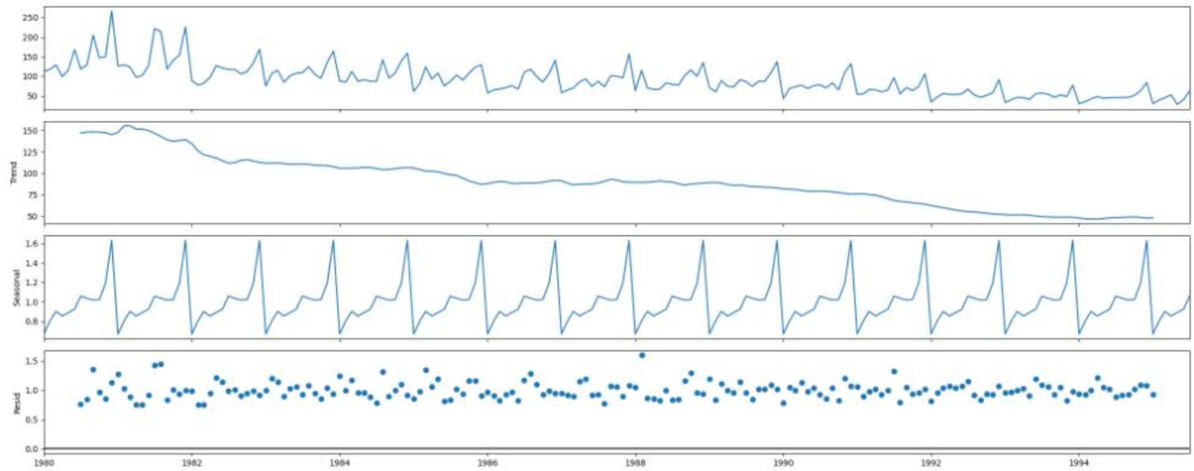


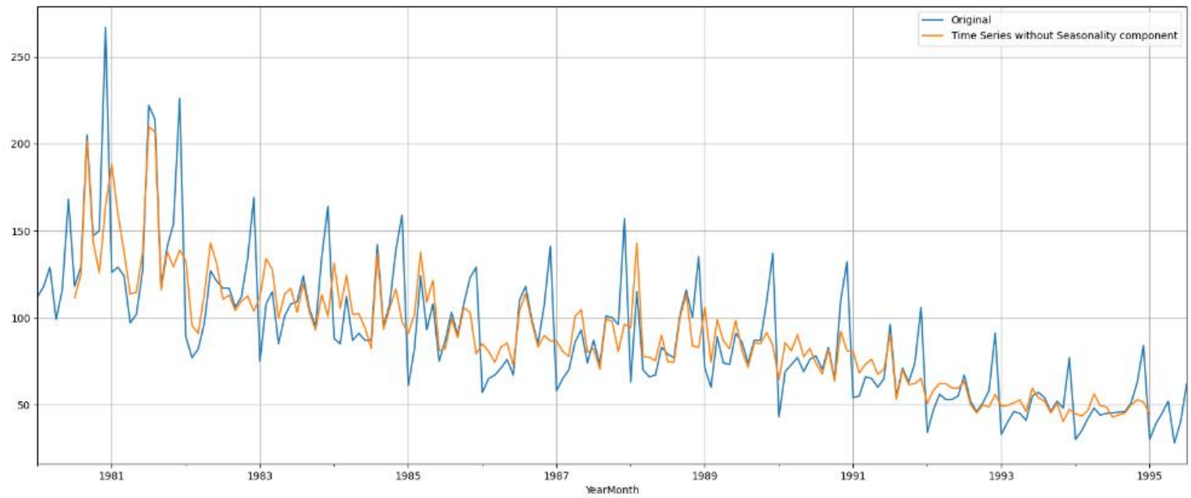
- December records have the high number of rose wine sales
- May, January have low number of wine sales.
- Yearly Plot – aggregate the time series from an annual perspective and summing up the observations

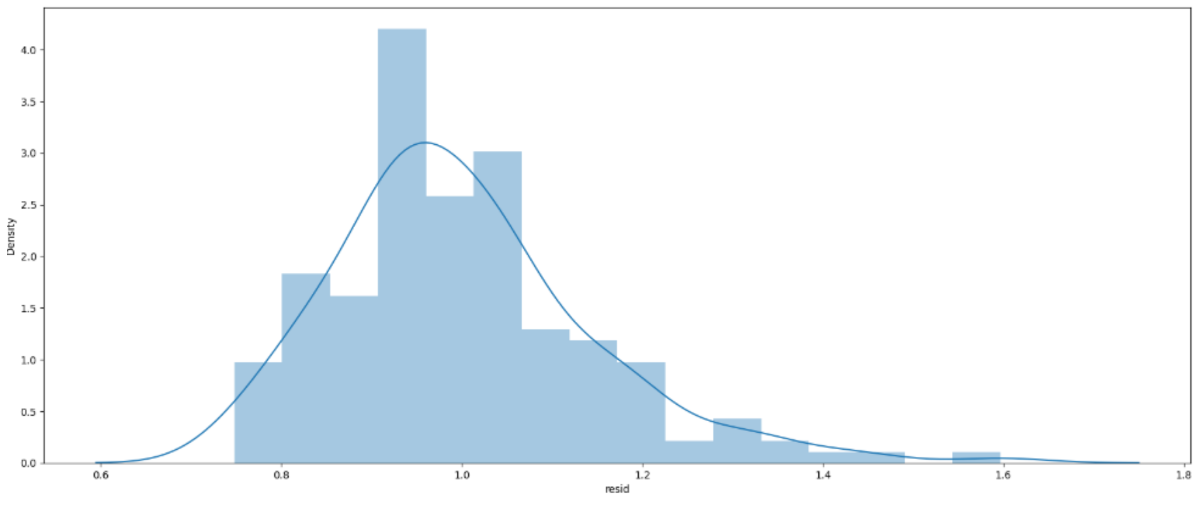




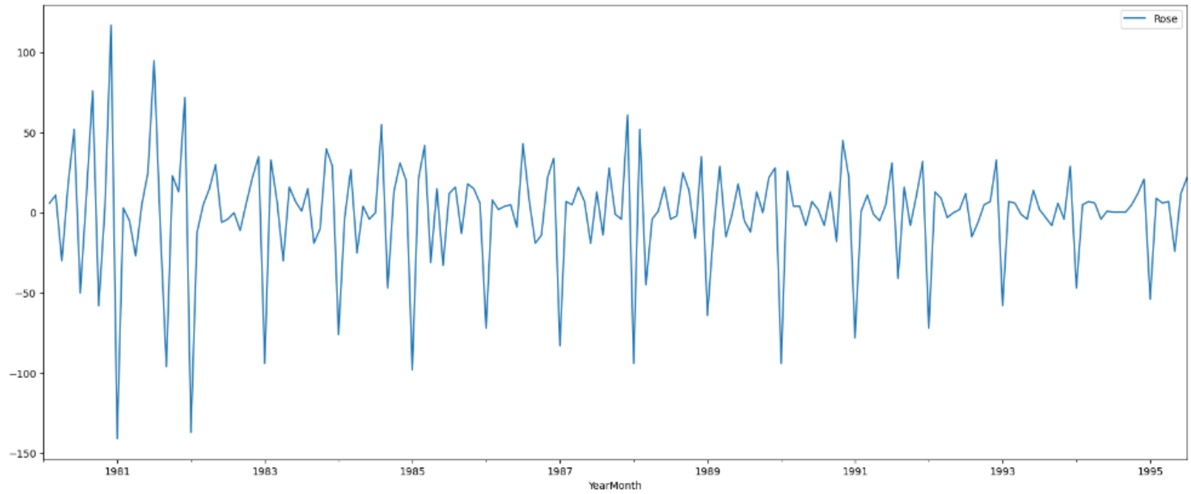
- From the decomposition, there is seasonality in the data.







ShapiroResult(statistic=0.9489824771881104, pvalue=6.11626865065773e-06)

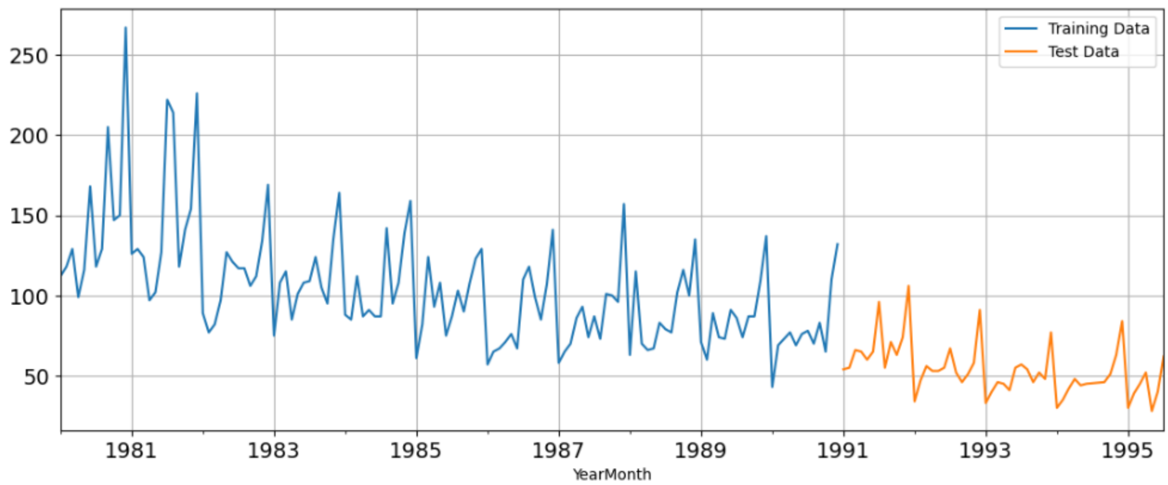


- Actual data without seasonality

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

Train and test shapes.

```
(132, 1)
(55, 1)
```



```
             Rose
YearMonth
1980-01-01  112.0
1980-02-01  118.0
1980-03-01  129.0
1980-04-01   99.0
1980-05-01  116.0
            Rose
YearMonth
1991-01-01  54.0
1991-02-01  55.0
1991-03-01  66.0
1991-04-01  65.0
1991-05-01  60.0
```

- 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

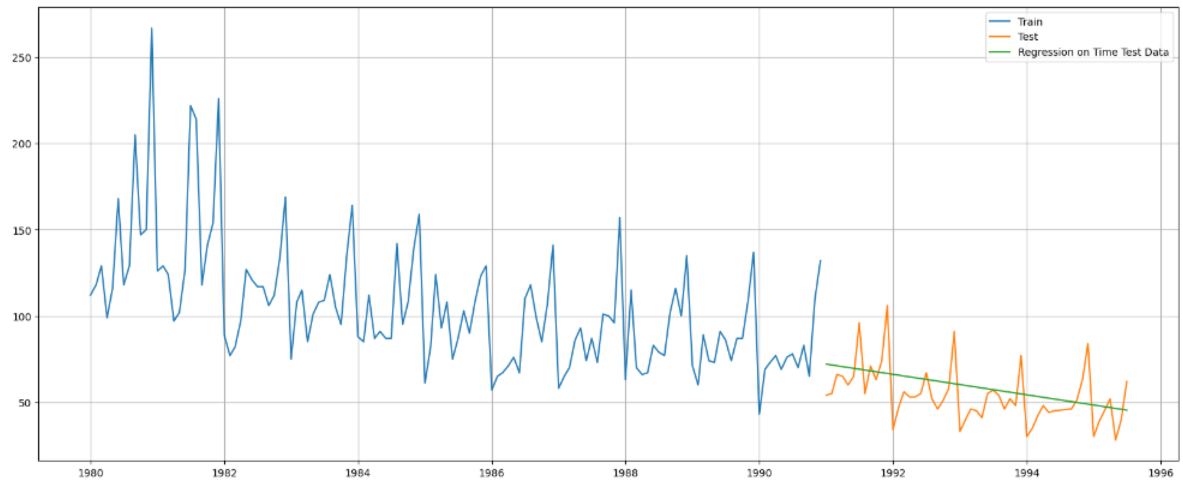
- Regress the "Rose" variable against the order of occurrence.
- Modifying the training set
- Generate the numerical instance order for both training and test set
- Printing the head and tail of train and test data

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

Last few rows of Training Data
              Rose  time
YearMonth
1990-08-01   70.0   128
1990-09-01   83.0   129
1990-10-01   65.0   130
1990-11-01  110.0   131
1990-12-01  132.0   132
```

```
First few rows of Test Data
             Rose  time
YearMonth
1991-01-01   54.0   133
1991-02-01   55.0   134
1991-03-01   66.0   135
1991-04-01   65.0   136
1991-05-01   60.0   137

Last few rows of Test Data
             Rose  time
YearMonth
1995-03-01   45.0   183
1995-04-01   52.0   184
1995-05-01   28.0   185
1995-06-01   40.0   186
1995-07-01   62.0   187
```

- Linear Regression is built on the training and test dataset

YearMonth	Rose	time	RegOnTime
1991-01-01	54.000000	133	72.063266
1991-02-01	55.000000	134	71.568888
1991-03-01	66.000000	135	71.074511
1991-04-01	65.000000	136	70.580133
1991-05-01	60.000000	137	70.085755
1991-06-01	65.000000	138	69.591377
1991-07-01	96.000000	139	69.096999
1991-08-01	55.000000	140	68.602621
1991-09-01	71.000000	141	68.108243
1991-10-01	63.000000	142	67.613866



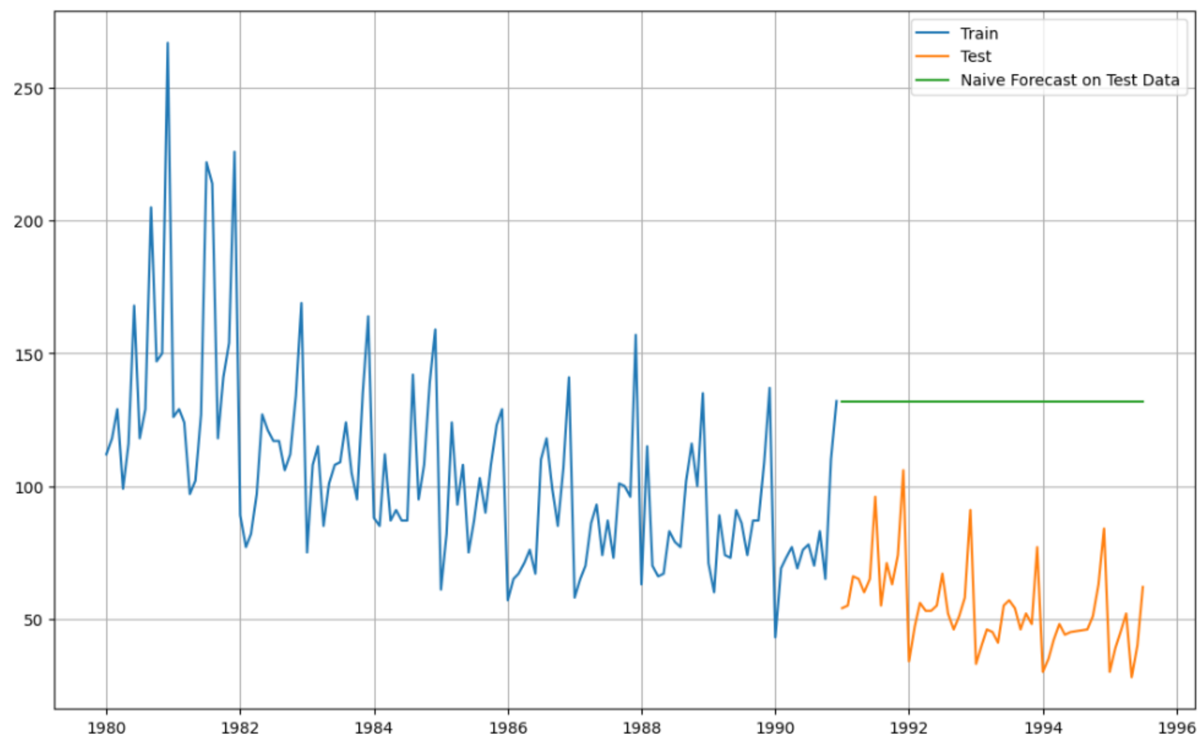
- Defining the accuracy metrics
- Evaluating the model

	Test RMSE
RegressionOnTime	15.268955

Model 2 – Naïve Model

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

```
YearMonth
1991-01-01    132.0
1991-02-01    132.0
1991-03-01    132.0
1991-04-01    132.0
1991-05-01    132.0
Name: naive, dtype: float64
```

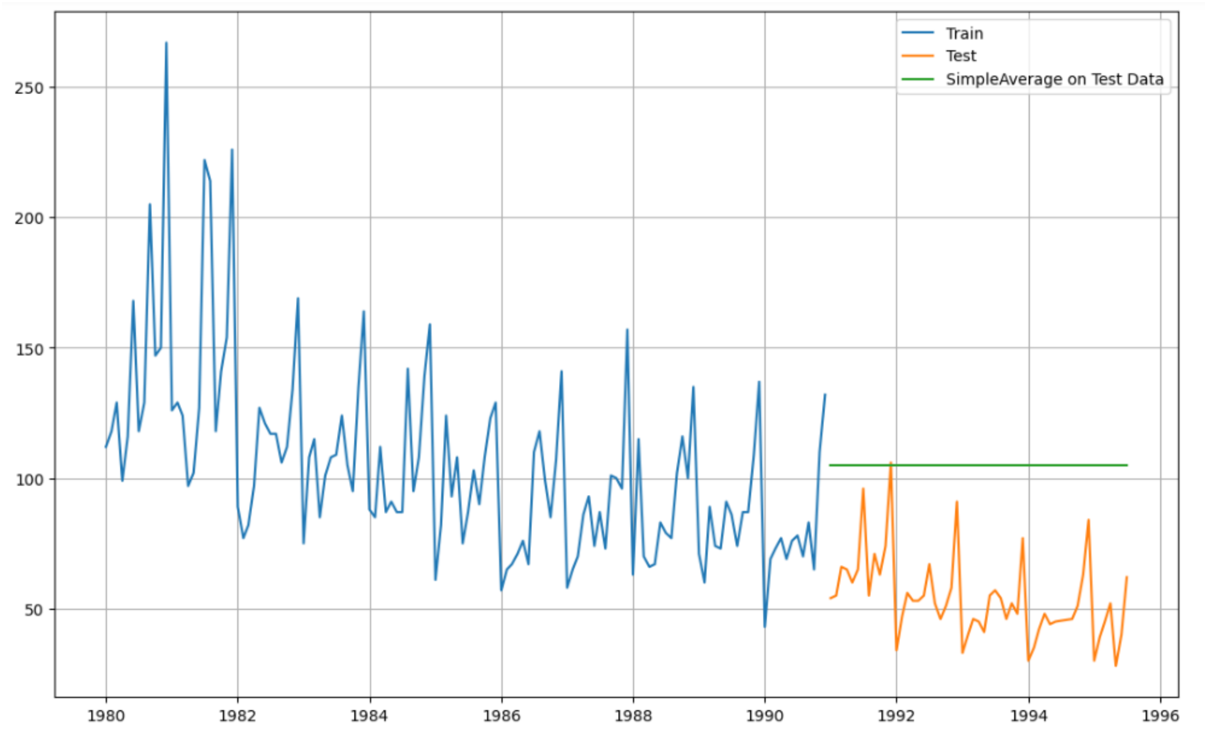



```
For Naive Forecast on Test Data, RMSE is 79.719
```

	Test RMSE
RegressionOnTime	15.268955
NaiveModel	79.718773

Model 3 – Simple Average – Forecast using the average of training values

```
YearMonth
1991-01-01    104.939394
1991-02-01    104.939394
1991-03-01    104.939394
1991-04-01    104.939394
1991-05-01    104.939394
Name: mean_forecast, dtype: float64
```



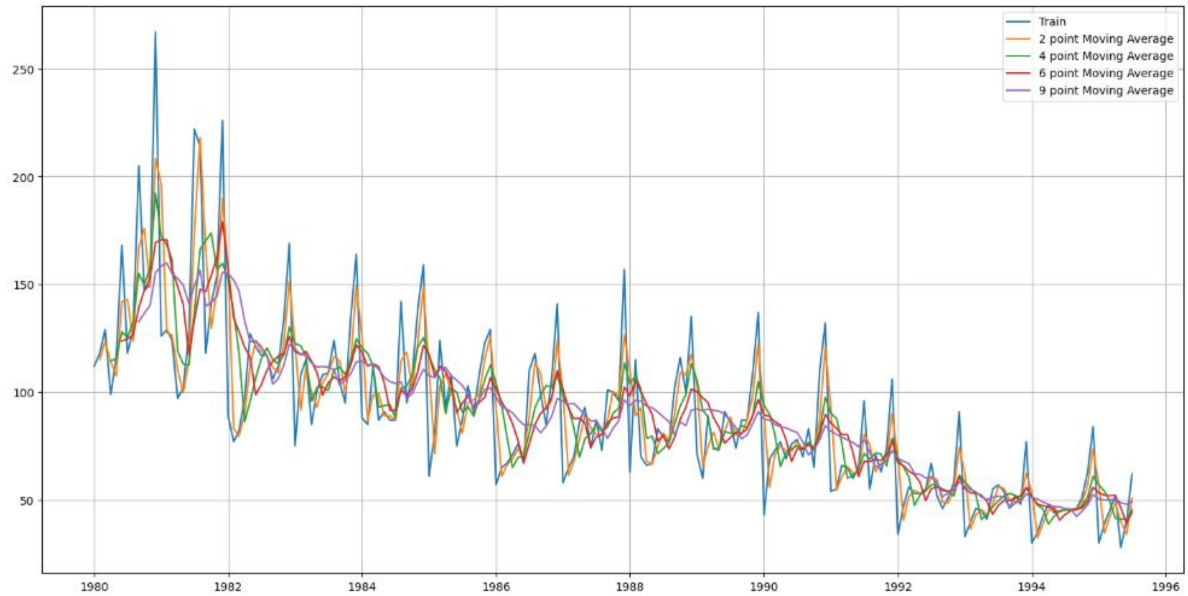

```
For SimpleAverage forecast on Test Data, RMSE is 53.461
```

	Test RMSE
RegressionOnTime	15.268955
NaiveModel	79.718773
SimpleAverage	53.460570

Model 4 – Moving Average – Calculating the rolling means (or moving average) for different intervals. The best interval can be determined by the maximum accuracy (or the minimum error) over here.

	Rose
YearMonth	
1980-01-01	112.0
1980-02-01	118.0
1980-03-01	129.0
1980-04-01	99.0
1980-05-01	116.0

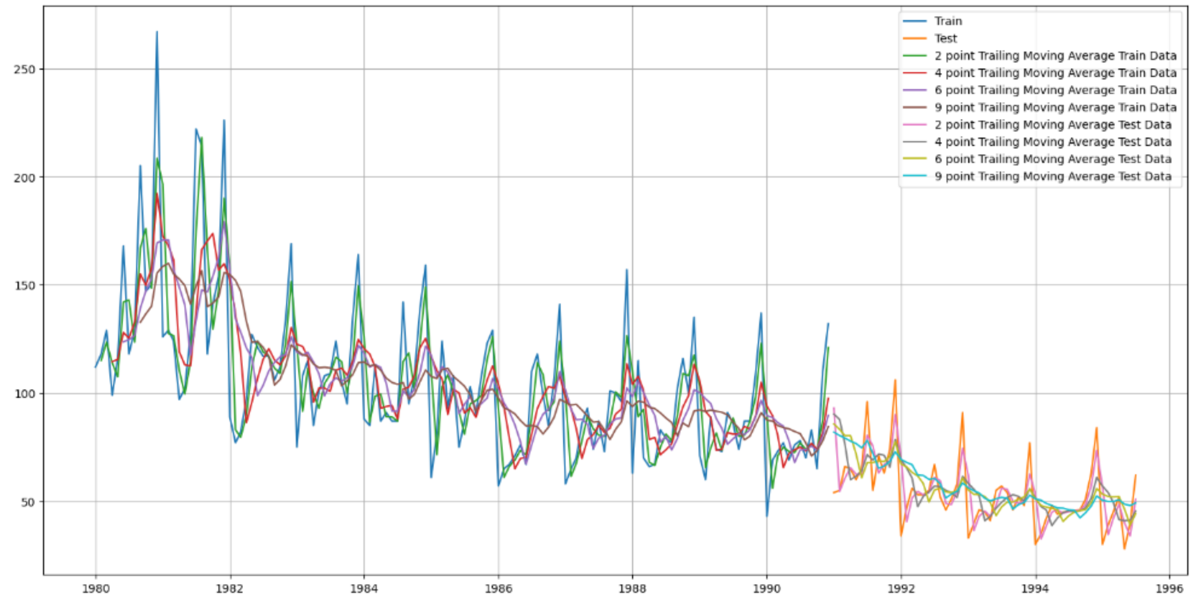
	Rose	Trailing_2	Trailing_4	Trailing_6	Trailing_9
YearMonth					
1980-01-01	112.0	NaN	NaN	NaN	NaN
1980-02-01	118.0	115.0	NaN	NaN	NaN
1980-03-01	129.0	123.5	NaN	NaN	NaN
1980-04-01	99.0	114.0	114.5	NaN	NaN
1980-05-01	116.0	107.5	115.5	NaN	NaN



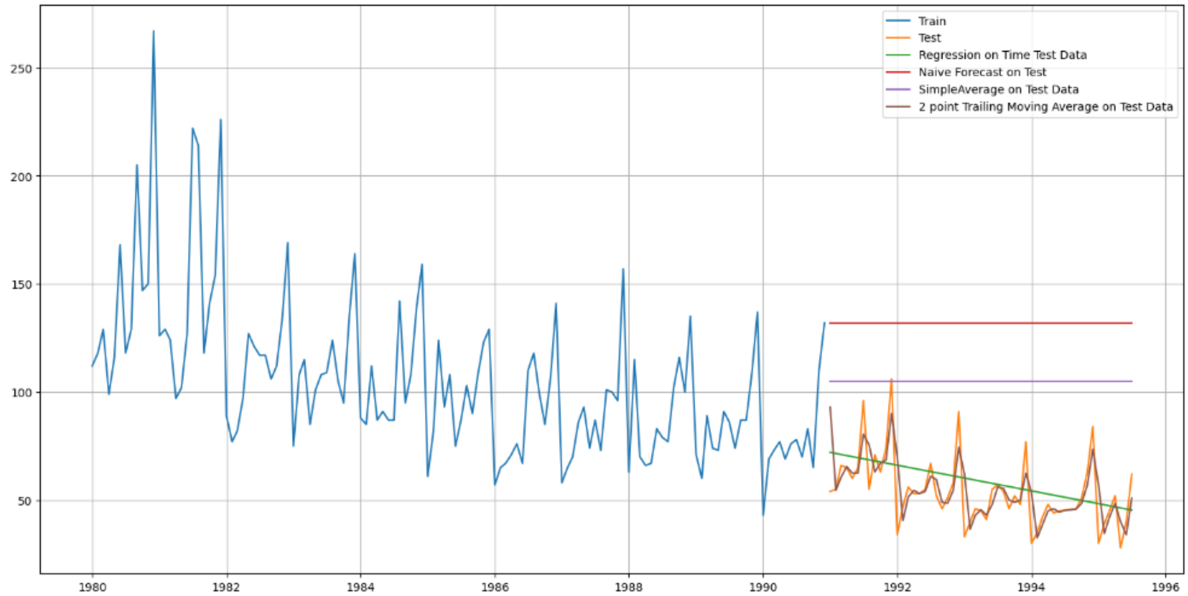
	Test RMSE
RegressionOnTime	15.268955
NaiveModel	79.718773
SimpleAverage	53.460570
2pointTrailingMovingAverage	11.529278
4pointTrailingMovingAverage	14.451403
6pointTrailingMovingAverage	14.566327
9pointTrailingMovingAverage	14.727630

```
For 2 point Moving Average Model forecast on Test Data, RMSE is 11.529
For 4 Point Moving Average Model forecast on Test Data, RMSE is 14.451
For 6 point Moving Average Model forecast on Test Data, RMSE is 14.566
For 9 point Moving Average Model forecast on Test Data, RMSE is 14.728
```

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.



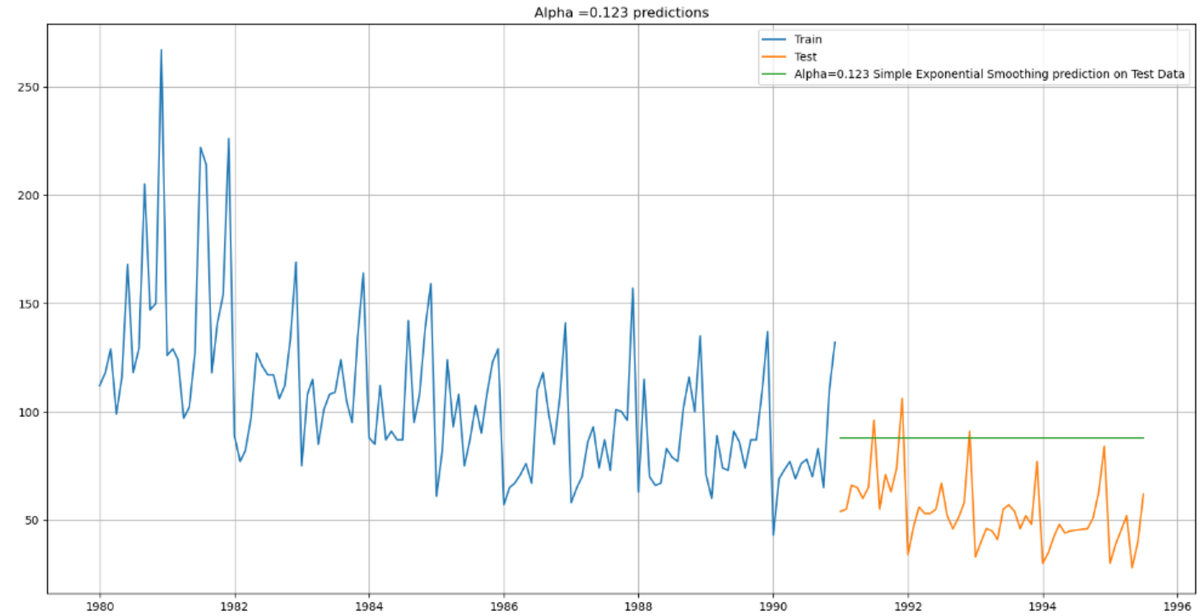
Before we go on to build the various Exponential Smoothing models, let us plot all the models and compare the Time Series plots



Model – 5 – Exponential Smoothing

```
YearMonth
1991-01-01    87.983765
1991-02-01    87.983765
1991-03-01    87.983765
1991-04-01    87.983765
1991-05-01    87.983765
Name: predict, dtype: float64
```

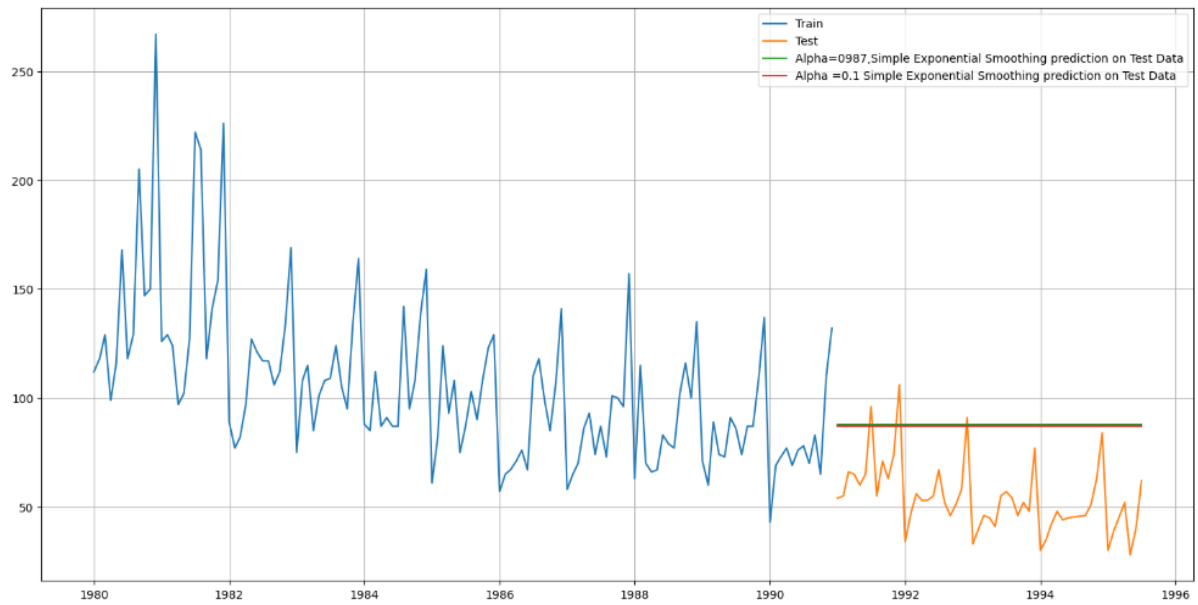
```
                  name      param  optimized
smoothing_level  alpha    0.12362       True
initial_level      1.0  112.00000      False
```




```
For Alpha =0.123 Simple Exponential Smoothing Model forecast on Test Data, RMSE is 37.592
```

	Test RMSE
RegressionOnTime	15.268955
NaiveModel	79.718773
SimpleAverage	53.460570
2pointTrailingMovingAverage	11.529278
4pointTrailingMovingAverage	14.451403
6pointTrailingMovingAverage	14.566327
9pointTrailingMovingAverage	14.727630
Alpha=0.123,SimpleExponential	37.592212

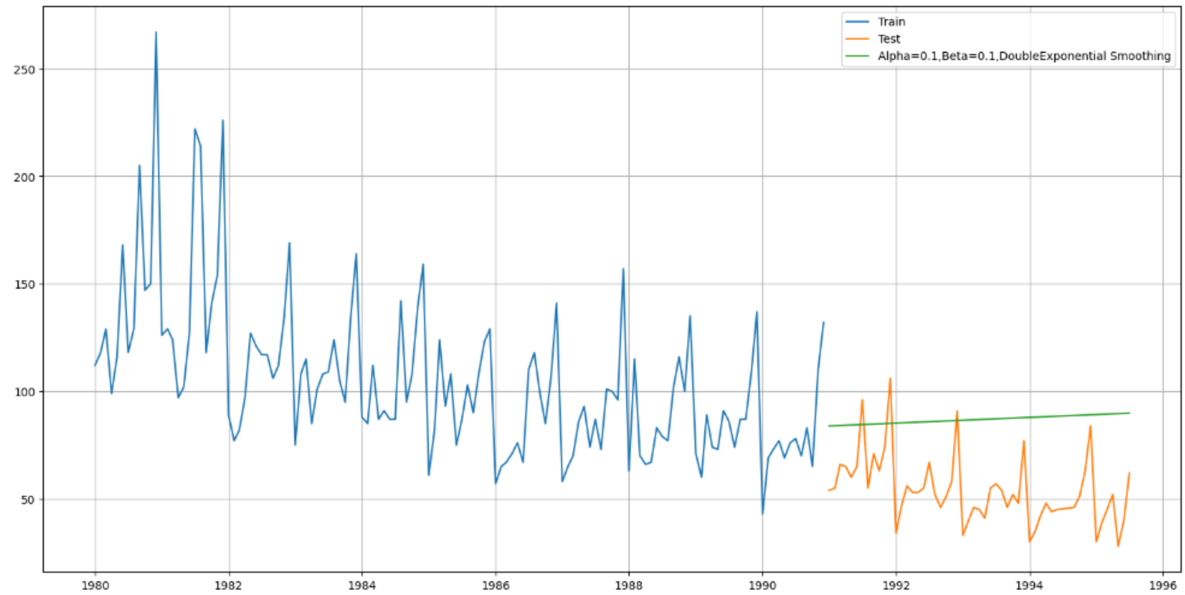
Setting different alpha values. Higher the alpha, the more weightage is given to more recent observation.



	Test RMSE
RegressionOnTime	15.268955
NaiveModel	79.718773
SimpleAverage	53.460570
2pointTrailingMovingAverage	11.529278
4pointTrailingMovingAverage	14.451403
6pointTrailingMovingAverage	14.566327
9pointTrailingMovingAverage	14.727630
Alpha=0.123,SimpleExponential	37.592212
Alpha=0.1,SimpleExponentialSmooting	36.828033

```
                 name        param  optimized
smoothing_level  alpha    0.162133       True
smoothing_trend   beta    0.131522       True
initial_level      l.0  112.000000      False
initial_trend      b.0    6.000000      False
```

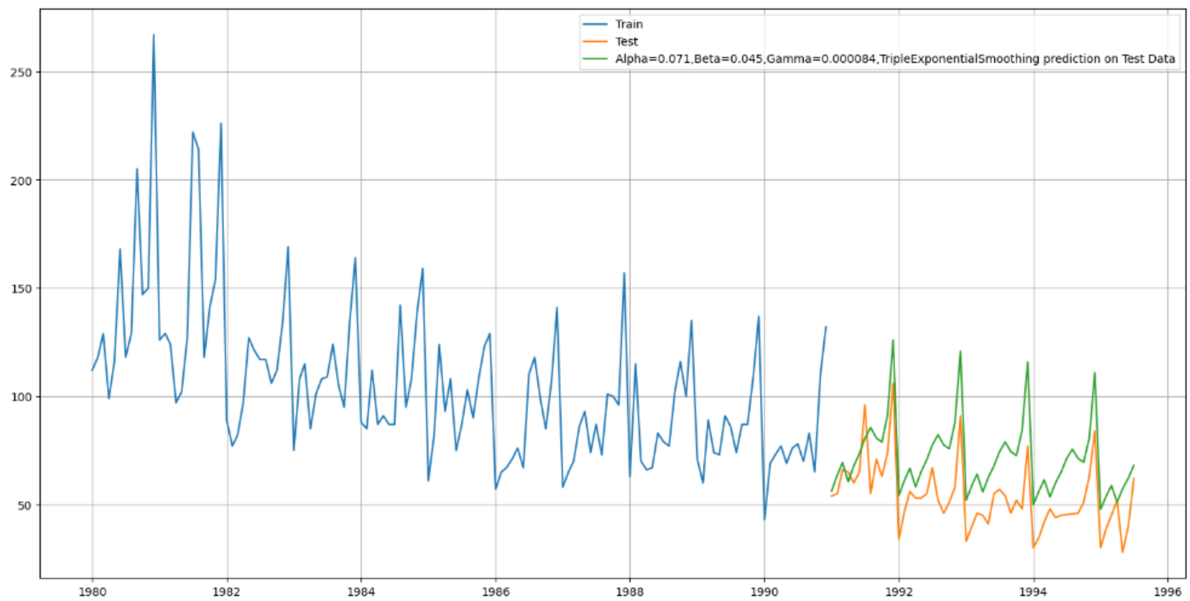
	Alpha Values	Beta Values	Train RMSE	Test RMSE
0	0.1	0.1	34.439111	36.923416
1	0.1	0.2	33.450729	48.688648
10	0.2	0.1	33.097427	65.731702
2	0.1	0.3	33.145789	78.156641
20	0.3	0.1	33.611269	98.653317



	Test RMSE
RegressionOnTime	15.268955
NaiveModel	79.718773
SimpleAverage	53.460570
2pointTrailingMovingAverage	11.529278
4pointTrailingMovingAverage	14.451403
6pointTrailingMovingAverage	14.566327
9pointTrailingMovingAverage	14.727630
Alpha=0.123,SimpleExponential	37.592212
Alpha=0.1,SimpleExponentialSmooting	36.828033
Alpha=0.1,Beta=0.1,DoubleExponentialSmoothing	36.923416

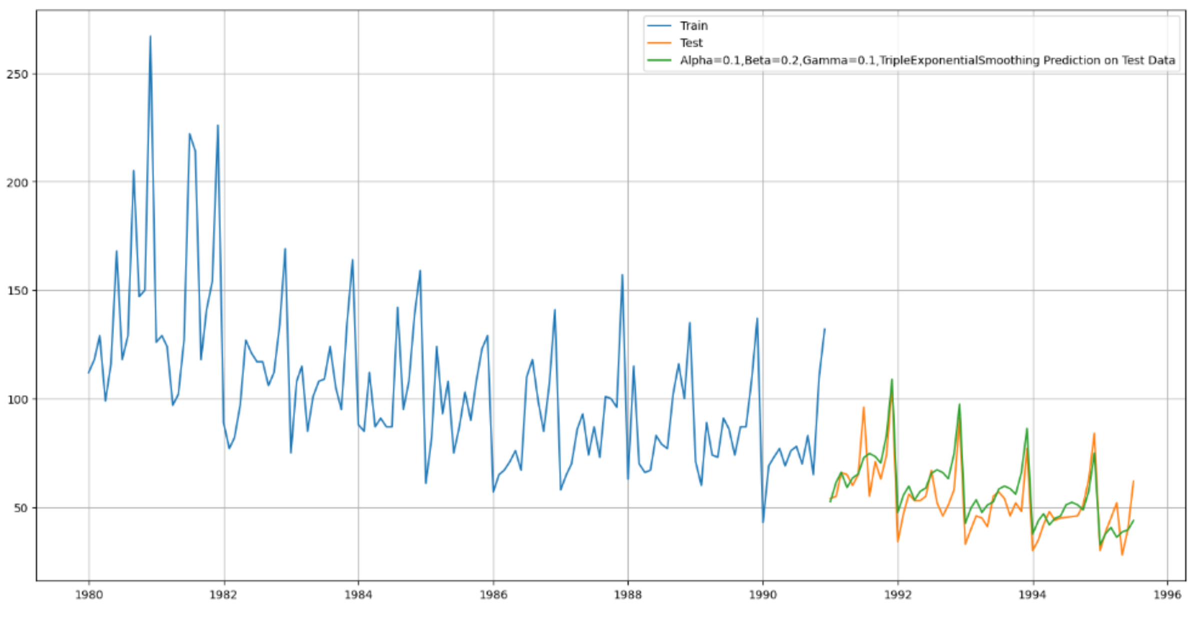
	name	param	optimized
smoothing_level	alpha	0.071303	True
smoothing_trend	beta	0.045508	True
smoothing_seasonal	gamma	0.000084	True
initial_level	l.0	163.600927	True
initial_trend	b.0	-0.980484	True
initial_seasons.0	s.0	0.687142	True
initial_seasons.1	s.1	0.779361	True
initial_seasons.2	s.2	0.851847	True
initial_seasons.3	s.3	0.744464	True
initial_seasons.4	s.4	0.837295	True
initial_seasons.5	s.5	0.911822	True
initial_seasons.6	s.6	1.002823	True
initial_seasons.7	s.7	1.067453	True
initial_seasons.8	s.8	1.010252	True
initial_seasons.9	s.9	0.989574	True
initial_seasons.10	s.10	1.153515	True
initial_seasons.11	s.11	1.590371	True

	Rose	auto_predict
YearMonth		
1991-01-01	54.0	56.332862
1991-02-01	55.0	63.693063
1991-03-01	66.0	69.394575
1991-04-01	65.0	60.454513
1991-05-01	60.0	67.772390



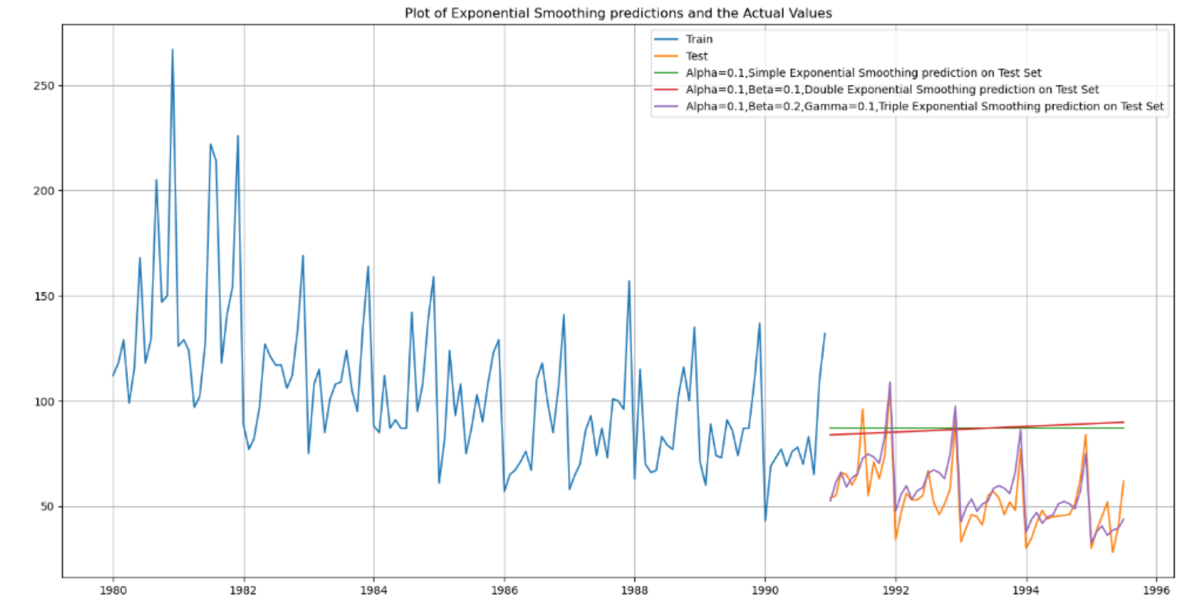
	Test RMSE
RegressionOnTime	15.268955
NaiveModel	79.718773
SimpleAverage	53.460570
2pointTrailingMovingAverage	11.529278
4pointTrailingMovingAverage	14.451403
6pointTrailingMovingAverage	14.566327
9pointTrailingMovingAverage	14.727630
Alpha=0.123,SimpleExponential	37.592212
Alpha=0.1,SimpleExponentialSmooting	36.828033
Alpha=0.1,Beta=0.1,DoubleExponentialSmoothing	36.923416
Alpha=0.071,Beta=0.045,Gamma=0.000084,TripleExponentialSmoothing	20.189764

	Alpha Values	Beta Values	Gamma Values	Train RMSE	Test RMSE
10	0.1	0.2	0.1	19.770392	9.223504
11	0.1	0.2	0.2	20.253487	9.496152
151	0.2	0.6	0.2	23.129850	9.565988
12	0.1	0.2	0.3	20.871304	9.888106
142	0.2	0.5	0.3	23.656276	9.891550

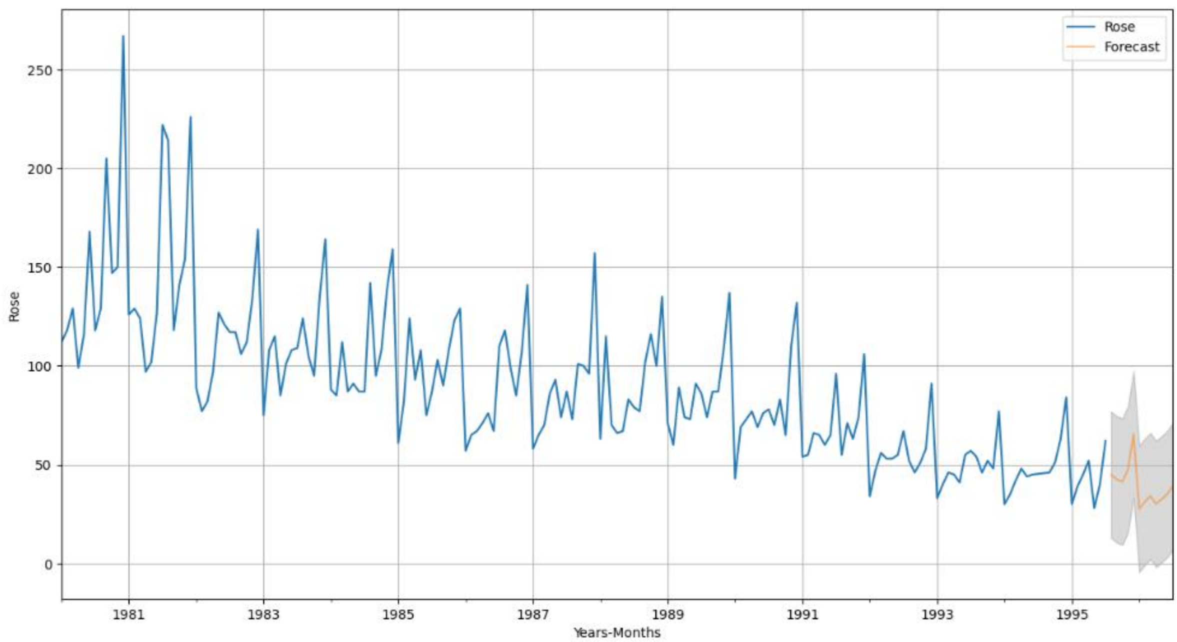


	Test RMSE
RegressionOnTime	15.268955
NaiveModel	79.718773
SimpleAverage	53.460570
2pointTrailingMovingAverage	11.529278
4pointTrailingMovingAverage	14.451403
6pointTrailingMovingAverage	14.566327
9pointTrailingMovingAverage	14.727630
Alpha=0.123,SimpleExponential	37.592212
Alpha=0.1,SimpleExponentialSmooting	36.828033
Alpha=0.1,Beta=0.1,DoubleExponentialSmoothing	36.923416
Alpha=0.071,Beta=0.045,Gamma=0.000084,TripleExponentialSmoothing	20.189764
Alpha=0.1,Beta=0.2,Gamma=0.1,TripleExponentialSmoothing	9.223504

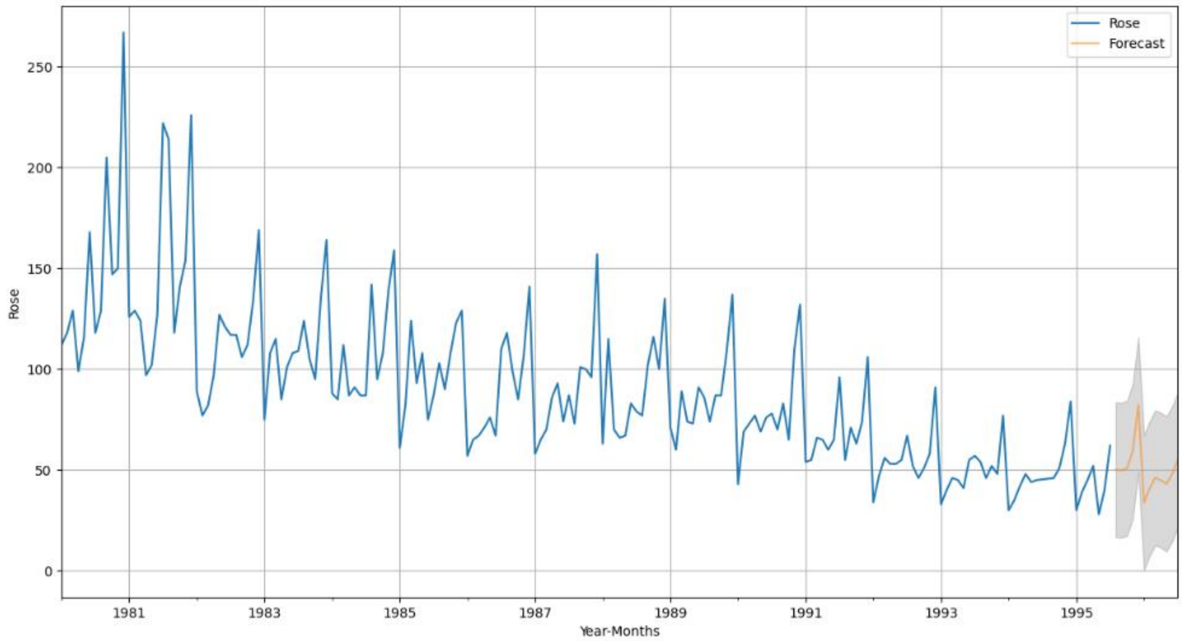
	Test RMSE
Alpha=0.1,Beta=0.2,Gamma=0.1,TripleExponentialSmoothing	9.223504
2pointTrailingMovingAverage	11.529278
4pointTrailingMovingAverage	14.451403
6pointTrailingMovingAverage	14.566327
9pointTrailingMovingAverage	14.727630
RegressionOnTime	15.268955
Alpha=0.071,Beta=0.045,Gamma=0.000084,TripleExponentialSmoothing	20.189764
Alpha=0.1,SimpleExponentialSmooting	36.828033
Alpha=0.1,Beta=0.1,DoubleExponentialSmoothing	36.923416
Alpha=0.123,SimpleExponential	37.592212
SimpleAverage	53.460570
NaiveModel	79.718773



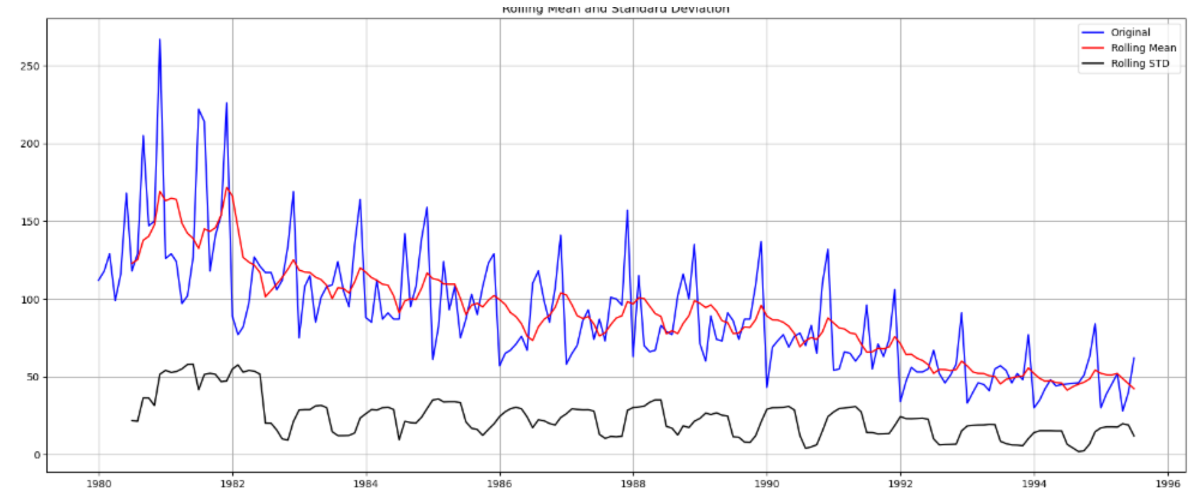
	lower_CI	prediction	upper_ci
1995-08-01	12.891228	44.899800	76.908372
1995-09-01	10.531140	42.539712	74.548285
1995-10-01	9.375796	41.384368	73.392940
1995-11-01	15.680354	47.688927	79.697499
1995-12-01	33.411405	65.419977	97.428549



	lower_CI	prediction	upper_ci
1995-08-01	16.636530	50.084245	83.531960
1995-09-01	16.427461	49.875176	83.322891
1995-10-01	17.384872	50.832587	84.280302
1995-11-01	25.743396	59.191111	92.638826
1995-12-01	48.902697	82.350412	115.798127

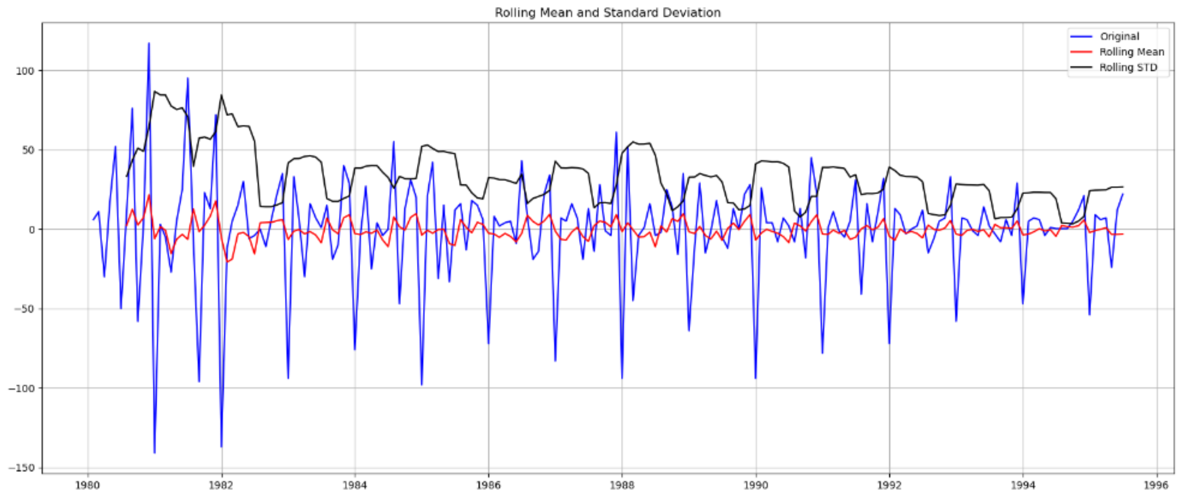


5. **Check for the stationarity of the data on which the model is being built on using appropriate statistical tests and also mention the hypothesis for the statistical test. If the data is found to be non-stationary, take appropriate steps to make it stationary. Check the new data for stationarity and 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 Observation Used     173.000000
Critical Value (1%)             -3.468726
Critical Value (5%)             -2.878396
Critical Value (10%)            -2.575756
dtype: float64
```

- **Applying difference**



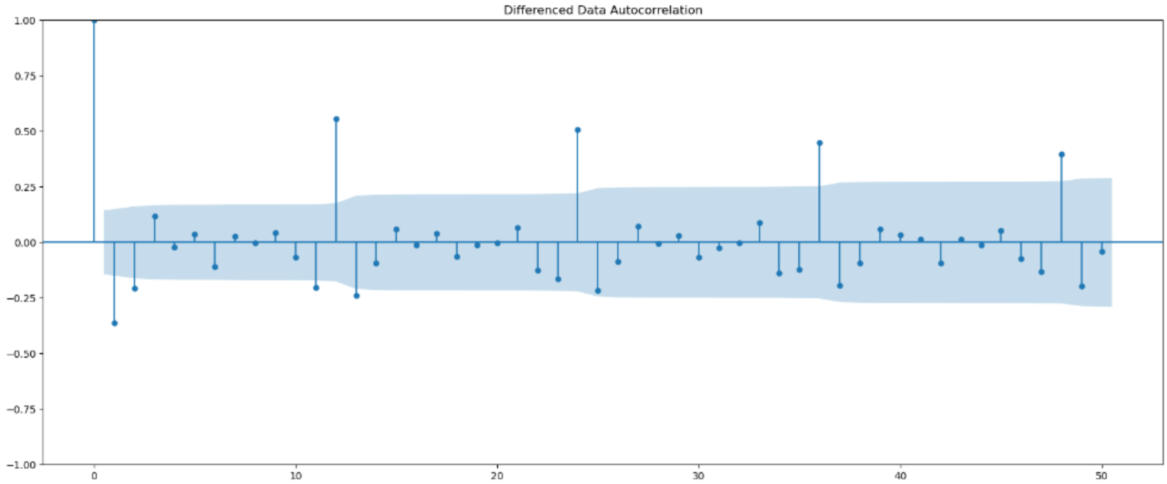
```
Results of Dickey-Fuller Test:
Test Statistic                -8.044392e+00
p-value                        1.810895e-12
#Lags Used:                    1.200000e+01
Number of Observation Used     1.730000e+02
Critical Value (1%)           -3.468726e+00
Critical Value (5%)           -2.878396e+00
Critical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

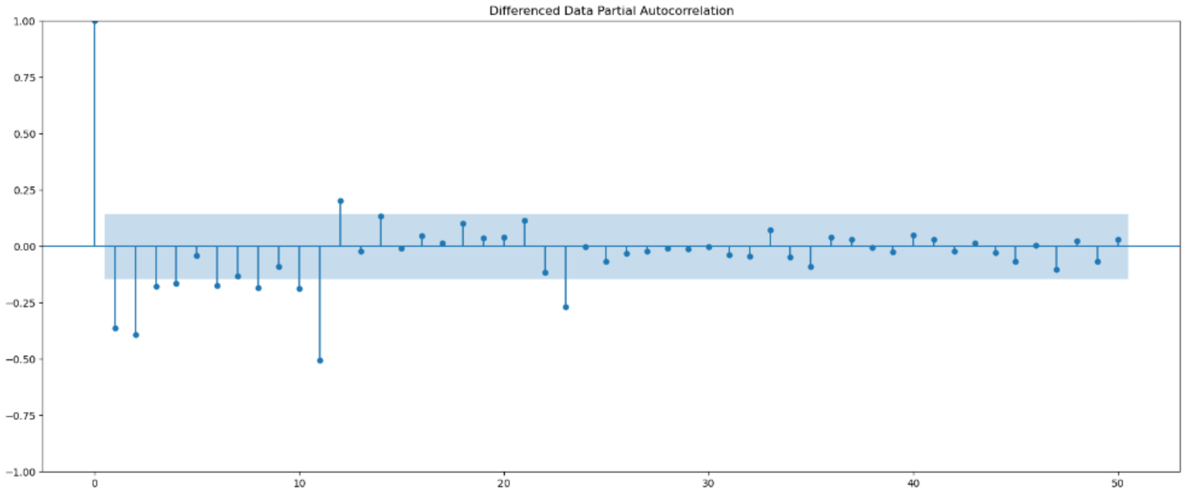
- When the time series data is not stationary we need to convert it into stationary before applying models.
- We use Augmented Dickey Fuller test.
- It determines how strongly a time series is defined by the trend.
- From the null and alternate hypothesis, we define time series data is stationary or 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 in stationary or not
- At alpha = 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.**

- Autoregression means regression of a variable on itself which means Autoregressive model use previous time period values to predict the current time period values.
- One of the fundamental assumptions of an AR model is that the time series is assumed to be a stationary process.
- We look at the Partial Autocorrelations of a stationary Time Series to understand the order of Auto-Regressive model.

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.

```
Some parameter combinations for the Model...
Model:(0, 1, 1)
Model:(0, 1, 2)
Model:(1, 1, 0)
Model:(1, 1, 1)
Model:(1, 1, 2)
Model:(2, 1, 0)
Model:(2, 1, 1)
Model:(2, 1, 2)
```

```
ARIMA(0, 1, 0) - AIC1333.1546729124348
ARIMA(0, 1, 1) - AIC1282.3098319748333
ARIMA(0, 1, 2) - AIC1279.671528853576
ARIMA(1, 1, 0) - AIC1317.3503105381546
ARIMA(1, 1, 1) - AIC1280.5742295380055
ARIMA(1, 1, 2) - AIC1279.8707234231922
ARIMA(2, 1, 0) - AIC1298.6110341604983
ARIMA(2, 1, 1) - AIC1281.5078621868515
ARIMA(2, 1, 2) - AIC1281.8707222264436
```

- 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.

	param	AIC
2	(0, 1, 2)	1279.671529
5	(1, 1, 2)	1279.870723
4	(1, 1, 1)	1280.574230
7	(2, 1, 1)	1281.507862
8	(2, 1, 2)	1281.870722
1	(0, 1, 1)	1282.309832
6	(2, 1, 0)	1298.611034
3	(1, 1, 0)	1317.350311
0	(0, 1, 0)	1333.154673

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                   Rose   No. Observations:                  132
Model:                 ARIMA(0, 1, 2)   Log Likelihood                -636.836
Date:                Sun, 22 Oct 2023   AIC                           1279.672
Time:                        16:04:29   BIC                           1288.297
Sample:                    01-01-1980   HQIC                          1283.176
                         - 12-01-1990
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ma.L1         -0.6970      0.072     -9.689      0.000      -0.838      -0.556
ma.L2         -0.2042      0.073     -2.794      0.005      -0.347      -0.061
sigma2       965.8407     88.305     10.938      0.000     792.766    1138.915
===================================================================================
Ljung-Box (L1) (Q):                   0.14   Jarque-Bera (JB):                39.24
Prob(Q):                              0.71   Prob(JB):                         0.00
Heteroskedasticity (H):               0.36   Skew:                             0.82
Prob(H) (two-sided):                  0.00   Kurtosis:                         5.13
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```

	RMSE
ARMIMA(0,1,2)	37.30648

```
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(0, 1, 0)x(0, 0, 0, 6) - AIC1323.9657875279158
SARIMA(0, 1, 0)x(0, 0, 1, 6) - AIC1264.4996261113868
SARIMA(0, 1, 0)x(0, 0, 2, 6) - AIC1144.7077471827024
SARIMA(0, 1, 0)x(1, 0, 0, 6) - AIC1274.7897737087983
SARIMA(0, 1, 0)x(1, 0, 1, 6) - AIC1241.7870945149134
SARIMA(0, 1, 0)x(1, 0, 2, 6) - AIC1146.3093266722567
SARIMA(0, 1, 0)x(2, 0, 0, 6) - AIC1137.9167236212038
SARIMA(0, 1, 0)x(2, 0, 1, 6) - AIC1137.4533629515188
SARIMA(0, 1, 0)x(2, 0, 2, 6) - AIC1117.022442614149
SARIMA(0, 1, 1)x(0, 0, 0, 6) - AIC1263.5369097383968
SARIMA(0, 1, 1)x(0, 0, 1, 6) - AIC1201.383254802955
SARIMA(0, 1, 1)x(0, 0, 2, 6) - AIC1097.190821775278
SARIMA(0, 1, 1)x(1, 0, 0, 6) - AIC1222.4354735745042
SARIMA(0, 1, 1)x(1, 0, 1, 6) - AIC1160.4386253746786
SARIMA(0, 1, 1)x(1, 0, 2, 6) - AIC1084.8564123296583
SARIMA(0, 1, 1)x(2, 0, 0, 6) - AIC1095.7490379981891
SARIMA(0, 1, 1)x(2, 0, 1, 6) - AIC1097.6455189518385
SARIMA(0, 1, 1)x(2, 0, 2, 6) - AIC1053.0044082633433
SARIMA(0, 1, 2)x(0, 0, 0, 6) - AIC1251.667543054105
SARIMA(0, 1, 2)x(0, 0, 1, 6) - AIC1192.001719456311
```

	param	seasonal	AIC
53	(1, 1, 2)	(2, 0, 2, 6)	1041.655818
26	(0, 1, 2)	(2, 0, 2, 6)	1043.600261
80	(2, 1, 2)	(2, 0, 2, 6)	1045.220373
71	(2, 1, 1)	(2, 0, 2, 6)	1051.673461
44	(1, 1, 1)	(2, 0, 2, 6)	1052.778470
...	...	...	...
1	(0, 1, 0)	(0, 0, 1, 6)	1264.499626
3	(0, 1, 0)	(1, 0, 0, 6)	1274.789774
54	(2, 1, 0)	(0, 0, 0, 6)	1280.253756
27	(1, 1, 0)	(0, 0, 0, 6)	1308.161871
0	(0, 1, 0)	(0, 0, 0, 6)	1323.965788

```
                                     SARIMAX Results
==========================================================================================
Dep. Variable:                                    y   No. Observations:                  132
Model:             SARIMAX(1, 1, 2)x(2, 0, 2, 6)   Log Likelihood                -512.828
Date:                              Sun, 22 Oct 2023   AIC                           1041.656
Time:                                      16:09:18   BIC                           1063.685
Sample:                                           0   HQIC                          1050.598
                                              - 132
Covariance Type:                                opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.5939      0.149     -3.988      0.000      -0.886      -0.302
ma.L1         -0.1954   1944.856     -0.000      1.000   -3812.042    3811.651
ma.L2         -0.8046   1564.872     -0.001      1.000   -3067.897    3066.287
ar.S.L6       -0.0626      0.034     -1.825      0.068      -0.130       0.005
ar.S.L12       0.8451      0.035     24.094      0.000       0.776       0.914
ma.S.L6        0.2226   1944.846      0.000      1.000   -3811.605    3812.050
ma.S.L12      -0.7774   1511.984     -0.001      1.000   -2964.212    2962.658
sigma2       335.2051      9.255     36.218      0.000     317.065     353.345
===================================================================================
Ljung-Box (L1) (Q):                   0.07   Jarque-Bera (JB):                56.68
Prob(Q):                              0.78   Prob(JB):                         0.00
Heteroskedasticity (H):               0.47   Skew:                             0.52
Prob(H) (two-sided):                  0.02   Kurtosis:                         6.26
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
[2] Covariance matrix is singular or near-singular, with condition number 3.25e+21. Standard errors may be unstable.
```

y	mean	mean_se	mean_ci_lower	mean_ci_upper
0	62.842300	18.848311	25.900288	99.784311
1	67.631232	19.300117	29.803699	105.458766
2	74.747320	19.412679	36.699167	112.795472
3	71.326316	19.475631	33.154780	109.497852
4	76.018426	19.483913	37.830658	114.206194

	RMSE
ARMIMA(0,1,2)	37.306480
SARIMA(1,1,2)(2,0,2,6)	26.135733

```
Examples of some paramete combinations for model...
Model: (0, 1, 1)(0, 0, 1, 12)
Model: (0, 1, 2)(0, 0, 2, 12)
Model: (1, 1, 0)(1, 0, 0, 12)
Model: (1, 1, 1)(1, 0, 1, 12)
Model: (1, 1, 2)(1, 0, 2, 12)
Model: (2, 1, 0)(2, 0, 0, 12)
Model: (2, 1, 1)(2, 0, 1, 12)
Model: (2, 1, 2)(2, 0, 2, 12)
```

```
SARIMA(0, 1, 0)x(2, 0, 2, 12) - AIC955.5735408945677
SARIMA(0, 1, 1)x(2, 0, 2, 12) - AIC901.1988265993291
SARIMA(0, 1, 2)x(2, 0, 2, 12) - AIC887.9375085680122
SARIMA(1, 1, 0)x(2, 0, 2, 12) - AIC942.297310307097
SARIMA(1, 1, 1)x(2, 0, 2, 12) - AIC900.6725795936877
SARIMA(1, 1, 2)x(2, 0, 2, 12) - AIC889.871767321596
SARIMA(2, 1, 0)x(2, 0, 2, 12) - AIC927.8380693280794
SARIMA(2, 1, 1)x(2, 0, 2, 12) - AIC899.4835866291729
SARIMA(2, 1, 2)x(2, 0, 2, 12) - AIC890.6687980993004
```

	param	seasonal	AIC
83	(0, 1, 2)	(2, 0, 2, 12)	887.937509
86	(1, 1, 2)	(2, 0, 2, 12)	889.871767
89	(2, 1, 2)	(2, 0, 2, 12)	890.668798
88	(2, 1, 1)	(2, 0, 2, 12)	899.483587
85	(1, 1, 1)	(2, 0, 2, 12)	900.672580

```
                                      SARIMAX Results
==========================================================================================
Dep. Variable:                                  y   No. Observations:                  132
Model:             SARIMAX(0, 1, 2)x(2, 0, 2, 12)   Log Likelihood                -436.969
Date:                            Sun, 22 Oct 2023   AIC                            887.938
Time:                                    16:12:37   BIC                            906.448
Sample:                                         0   HQIC                           895.437
                                            - 132
Covariance Type:                              opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ma.L1         -0.8427    189.835     -0.004      0.996    -372.912     371.227
ma.L2         -0.1573     29.824     -0.005      0.996     -58.611      58.297
ar.S.L12       0.3467      0.079      4.375      0.000       0.191       0.502
ar.S.L24       0.3023      0.076      3.996      0.000       0.154       0.451
ma.S.L12       0.0767      0.133      0.577      0.564      -0.184       0.337
ma.S.L24      -0.0726      0.146     -0.498      0.618      -0.358       0.213
sigma2       251.3137   4.77e+04      0.005      0.996   -9.33e+04    9.38e+04
===================================================================================
Ljung-Box (L1) (Q):                   0.10   Jarque-Bera (JB):                 2.33
Prob(Q):                              0.75   Prob(JB):                         0.31
Heteroskedasticity (H):               0.88   Skew:                             0.37
Prob(H) (two-sided):                  0.70   Kurtosis:                         3.03
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```

y	mean	mean_se	mean_ci_lower	mean_ci_upper
0	62.867264	15.928501	31.647976	94.086552
1	70.541190	16.147659	38.892361	102.190020
2	77.356411	16.147656	45.707586	109.005235
3	76.208814	16.147656	44.559989	107.857638
4	72.747398	16.147656	41.098573	104.396223

	RMSE
ARMIMA(0,1,2)	37.306480
SARIMA(1,1,2)(2,0,2,6)	26.135733
SARIMA(0,1,2)(2,0,2,12)	26.928362

- 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 values.
- Next predicting the data using the SARIMA model and evaluating the model.
- We get the summary of the data

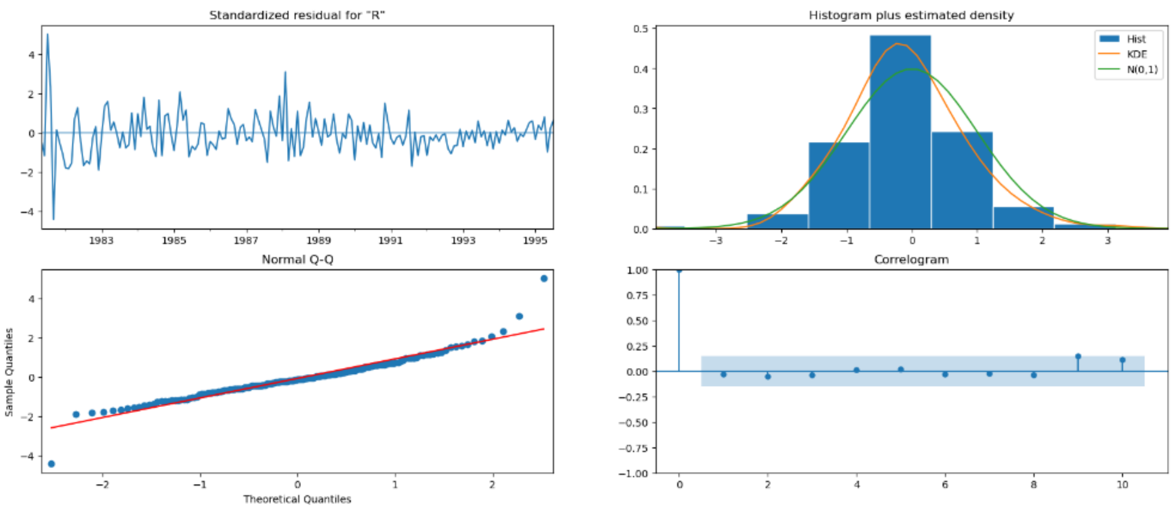
- There is a huge gain in the RMSE value by including seasonal parameters
- Keeping as seasonal parameter for second iteration

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.

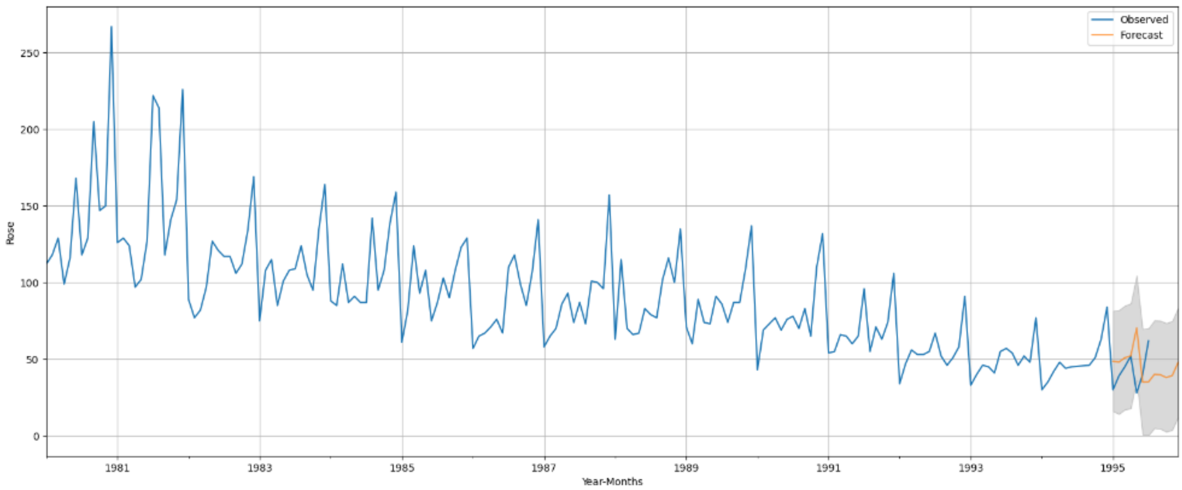
	RMSE
ARMIMA(0,1,2)	37.306480
SARIMA(1,1,2)(2,0,2,6)	26.135733
SARIMA(0,1,2)(2,0,2,12)	26.928362

9. Build 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/brands

```
                                     SARIMAX Results
==========================================================================================
Dep. Variable:                               Rose   No. Observations:                  187
Model:             SARIMAX(1, 1, 2)x(2, 0, 2, 6)   Log Likelihood                -731.652
Date:                            Sun, 22 Oct 2023   AIC                           1479.303
Time:                                    16:13:20   BIC                           1504.437
Sample:                                01-01-1980   HQIC                          1489.501
                                     - 07-01-1995
Covariance Type:                              opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.5906      0.129     -4.585      0.000      -0.843      -0.338
ma.L1         -0.1368      0.136     -1.009      0.313      -0.403       0.129
ma.L2         -0.7238      0.115     -6.310      0.000      -0.949      -0.499
ar.S.L6       -0.0367      0.026     -1.401      0.161      -0.088       0.015
ar.S.L12       0.8860      0.026     34.140      0.000       0.835       0.937
ma.S.L6        0.1504      0.199      0.754      0.451      -0.241       0.541
ma.S.L12      -0.7973      0.170     -4.692      0.000      -1.130      -0.464
sigma2       275.7336     50.003      5.514      0.000     177.730     373.737
===================================================================================
Ljung-Box (L1) (Q):                   0.15   Jarque-Bera (JB):               190.97
Prob(Q):                              0.70   Prob(JB):                         0.00
Heteroskedasticity (H):               0.19   Skew:                             0.59
Prob(H) (two-sided):                  0.00   Kurtosis:                         8.04
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```



Rose	mean	mean_se	mean_ci_lower	mean_ci_upper
1995-08-01	48.664589	16.699666	15.933845	81.395333
1995-09-01	48.031579	17.308145	14.108238	81.954919
1995-10-01	50.938839	17.311503	17.008916	84.868763
1995-11-01	52.039684	17.496849	17.746490	86.332878
1995-12-01	70.318415	17.516313	35.987073	104.649757
1996-01-01	35.081967	17.612574	0.561956	69.601978
1996-02-01	35.161390	17.854785	0.166655	70.156125
1996-03-01	40.104021	17.972333	4.878896	75.329145
1996-04-01	39.783532	18.022224	4.460622	75.106442
1996-05-01	37.989311	18.108818	2.496680	73.481942
1996-06-01	39.279412	18.172457	3.662050	74.896774
1996-07-01	47.831576	18.248832	12.064522	83.598629



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 26.2598
- Plotting the forecast with the confidence band
- It is clear that ARIMA(0,1,2) has the higher RMSE and SARIMA(1,1,2)(2,0,2,6) has the lower value.