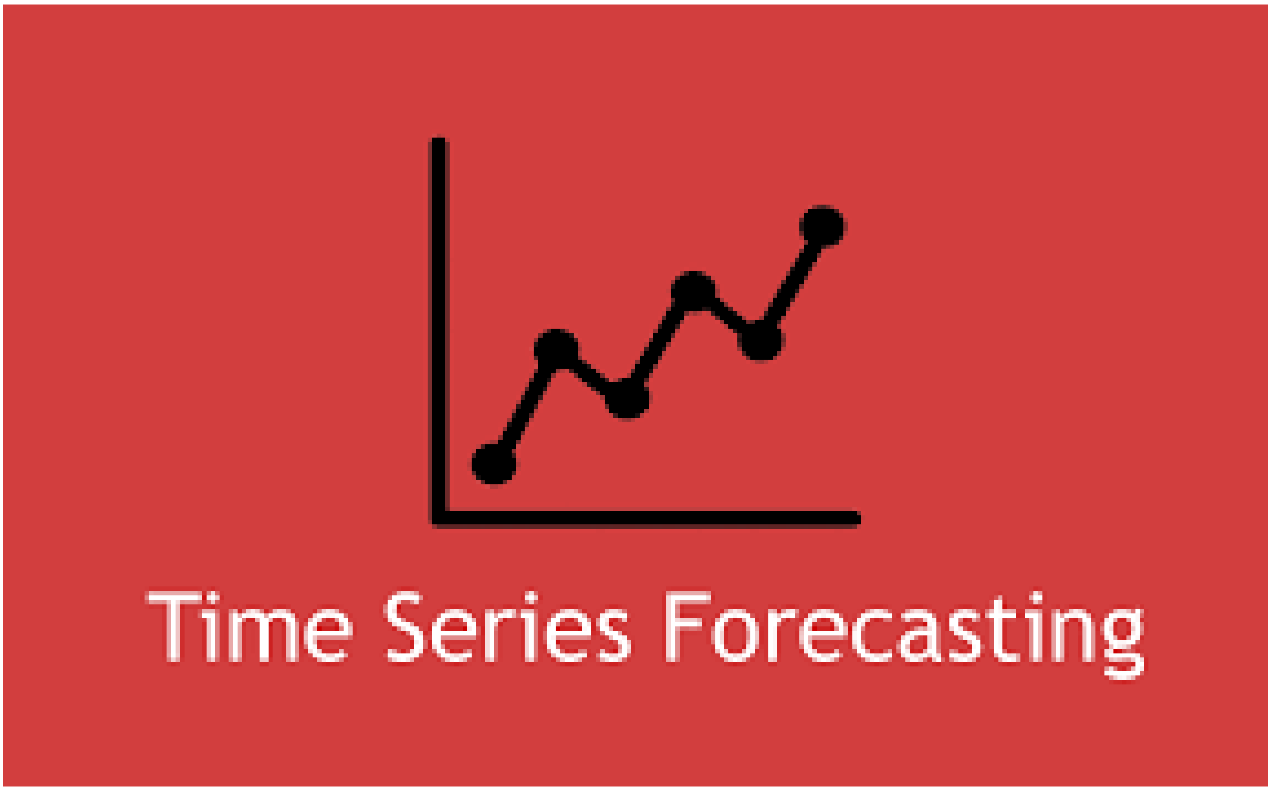
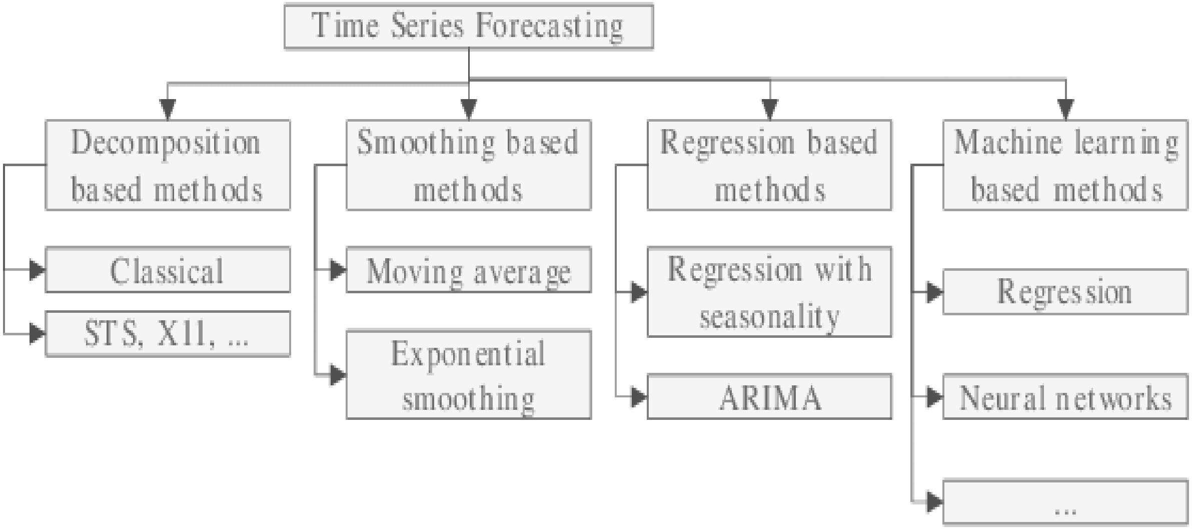
PARAKANDLA. MANISHA – PGP DSBA ONLINE

PROJECT: TIME SERIES FORECASTING (20-02-2022)





This Business report will provide the detailed explanation of how we performed analysis according to the problem statement given in the assignment. It will also provide the relative resolution and explanation with regards to the problem statement.

Contents

List of Figures:

TIME SERIES ANALYSIS ON WINESALES WITH ROSE DATASET

PROBLEM STATEMENT:

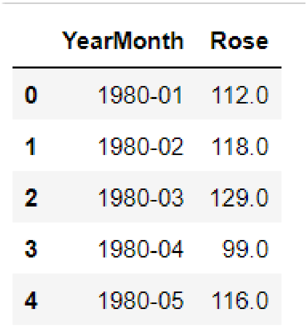
For this particular assignment, the data of different types of wine sales in the 20th century is to be analysed. Both of these data are from the same company but of different wines. As an analyst in the ABC Estate Wines, you are tasked to analyse and forecast Wine Sales in the 20th century.

Data set for the Problem: **Rose.csv**

First, we import all the necessary libraries such as seaborn, pandas, sklearn etc to perform our analysis.
Next, we import the dataset **Rose.csv**

1.Read the data as an appropriate Time Series data and plot the data.

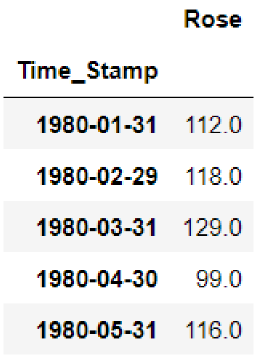
When we read the data using pandas read_csv function, the sample of the data looks like:



	YearMonth	Rose
0	1980-01	112.0
1	1980-02	118.0
2	1980-03	129.0
3	1980-04	99.0
4	1980-05	116.0

We need to transform this data into appropriate time series data using pandas date_range function with (start='1/1/1980', end='8/1/1995', freq='M').

After that, I have created timestamp for this data and made timestamp as index for the data and dropped the YearMonth column. Now, the data looks like:



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

ROSE WINE YEAR WISE SALE:

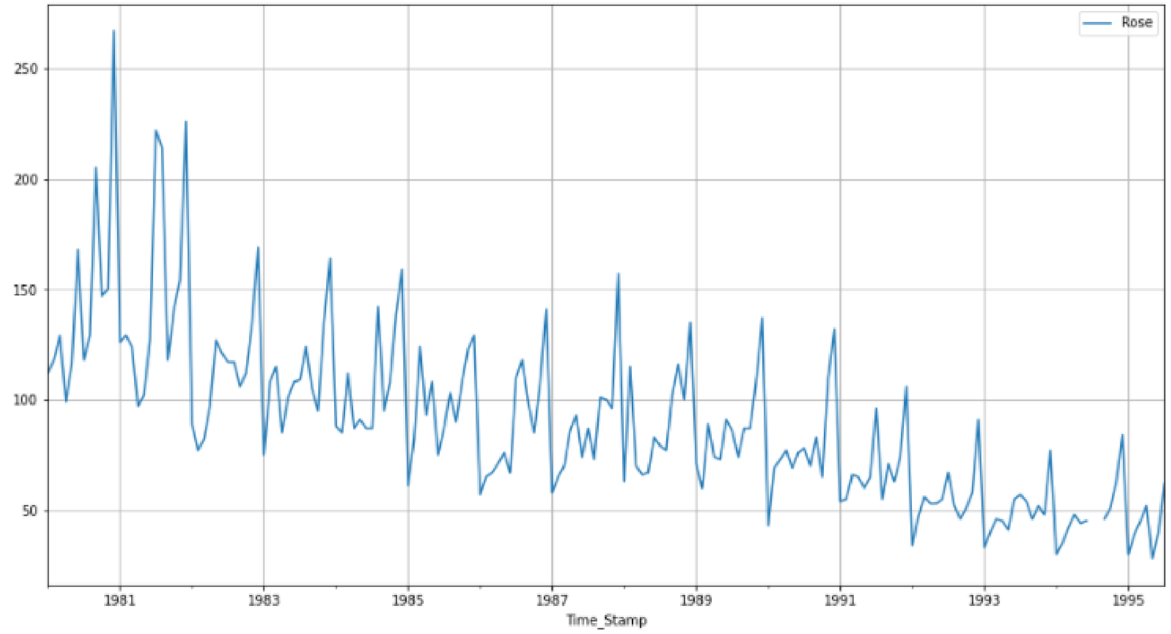


Fig1. Rose year wise sales

From the above plot, we observe that there is a decreasing trend in the initial years and stabilizes over the years.

We can also see that the seasonality in the data trend and pattern seems to repeat.

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

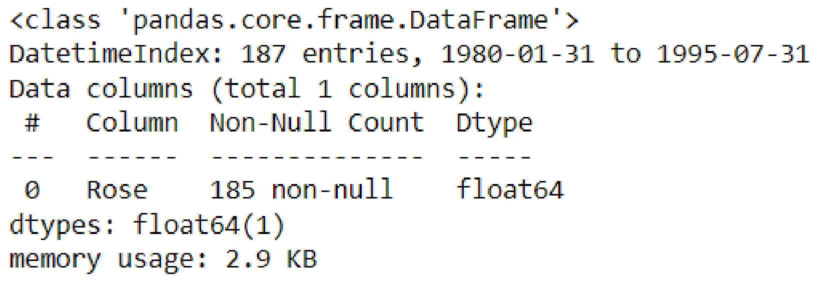
The Shape of the dataset is (187,1).

1.There are 187 observations which represent the monthly sales of respective wines from the year 1980 to July 1995.

2.The data has two variables the YearMonth of sales and the sales for the respective month of the year.

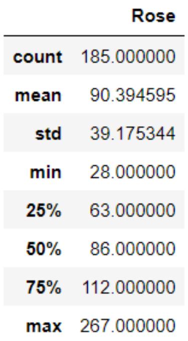
3.There are 2 null values present in the data, which were interpolated using linear method.

4.Checking the info of the data:



```
<class 'pandas.core.frame.DataFrame'>
DatetimeIndex: 187 entries, 1980-01-31 to 1995-07-31
Data columns (total 1 columns):
 #   Column  Non-Null Count  Dtype  
---  ------  --------------  -----  
 0   Rose    185 non-null    float64
dtypes: float64(1)
memory usage: 2.9 KB
```

5.Description of the data:



	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

Boxplots for yearly/Monthly sales for rose and Rose wine:

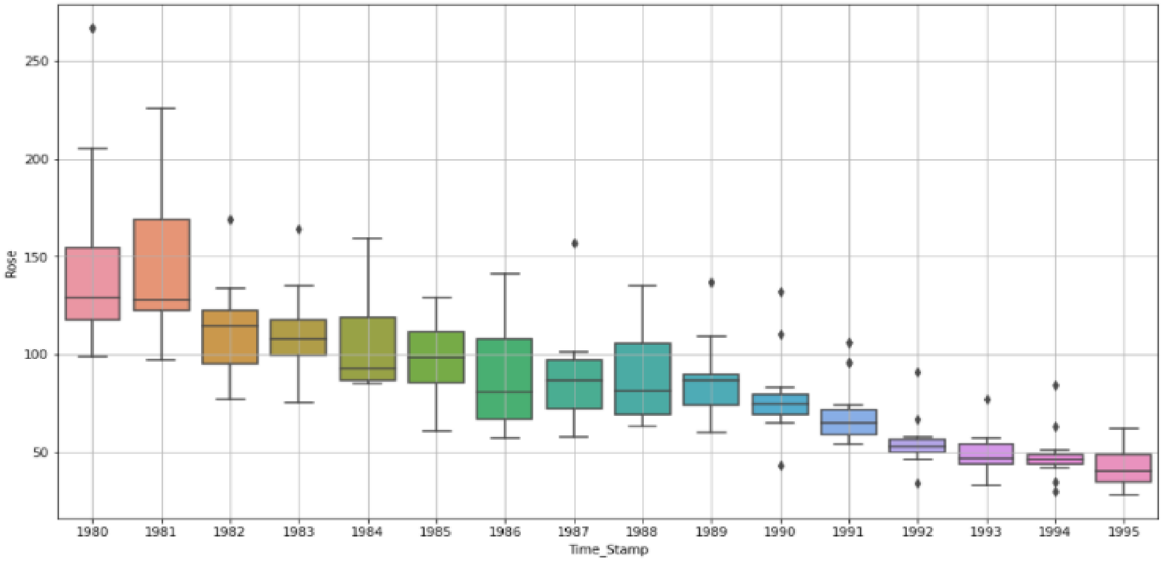


Fig2. Boxplot for yearly sales

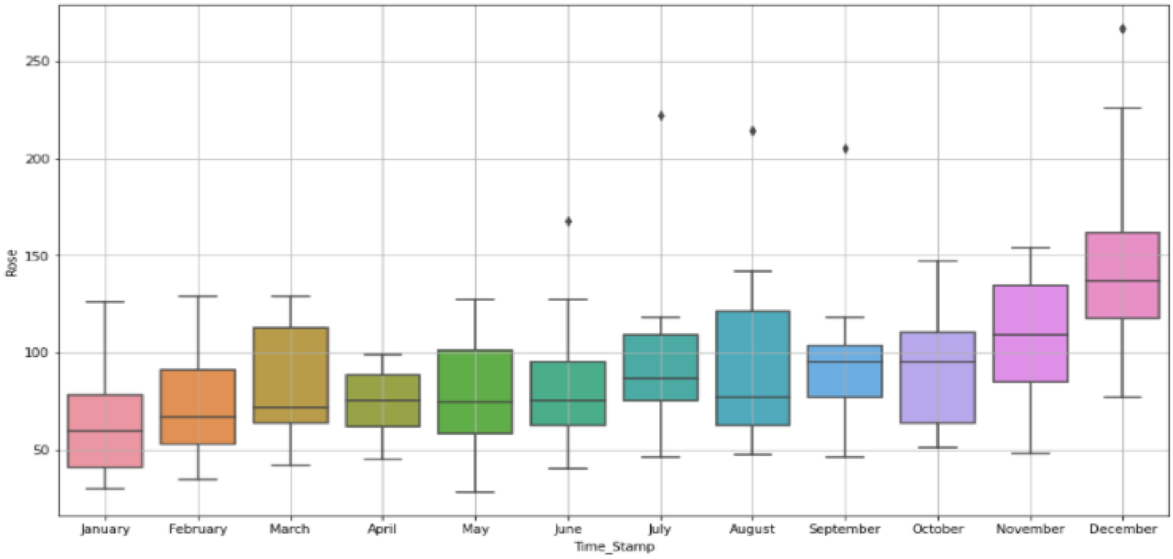


Fig3. Boxplot for Monthly sales

Inference:

In agreement with the timeseries plot the year wise plot indicate a measure of downward trend.

The sales of ROSE wine having some outliers for certain years.

December seems to have the highest sales od Rose wine and there are also outlier in June, July, August and September months.

Cumulative % and Month on Month % sales plots of Rose wine:

Time series month plot to understand the spread of Sales across different years and within different months across years.

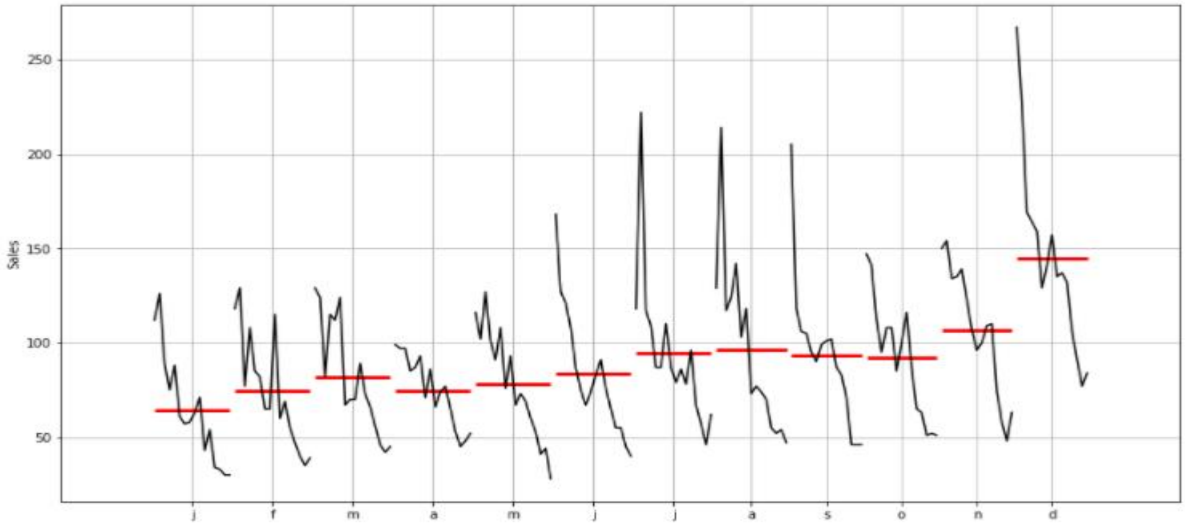


Fig4.Timeseries month plot for spread of sales

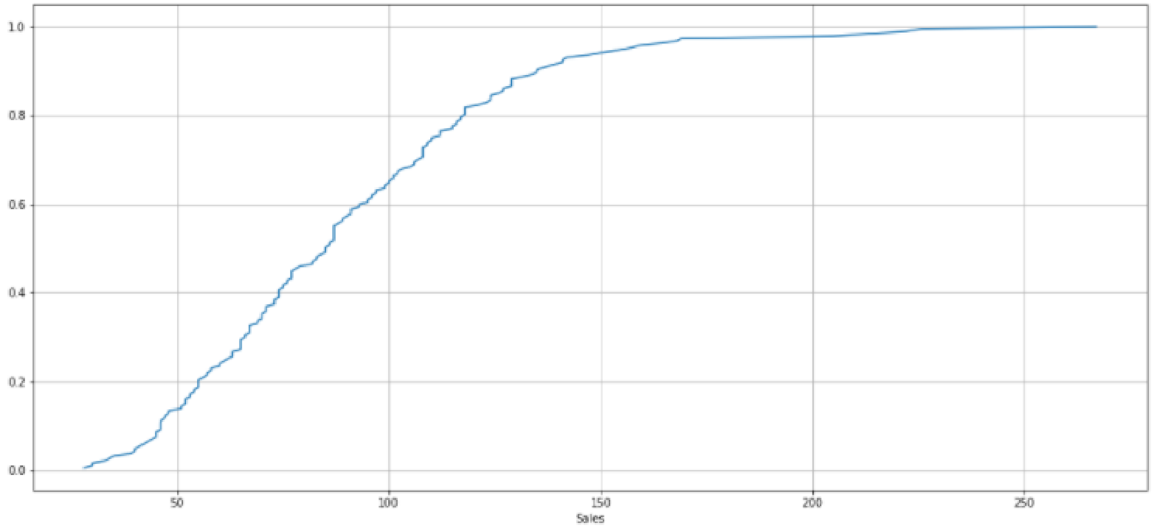


Fig5. Empirical Cumulative Distribution Curve

The ECD curve tells us what percentage of data points refer to what number of sales.

Line plot for monthly sales for Rose wine:

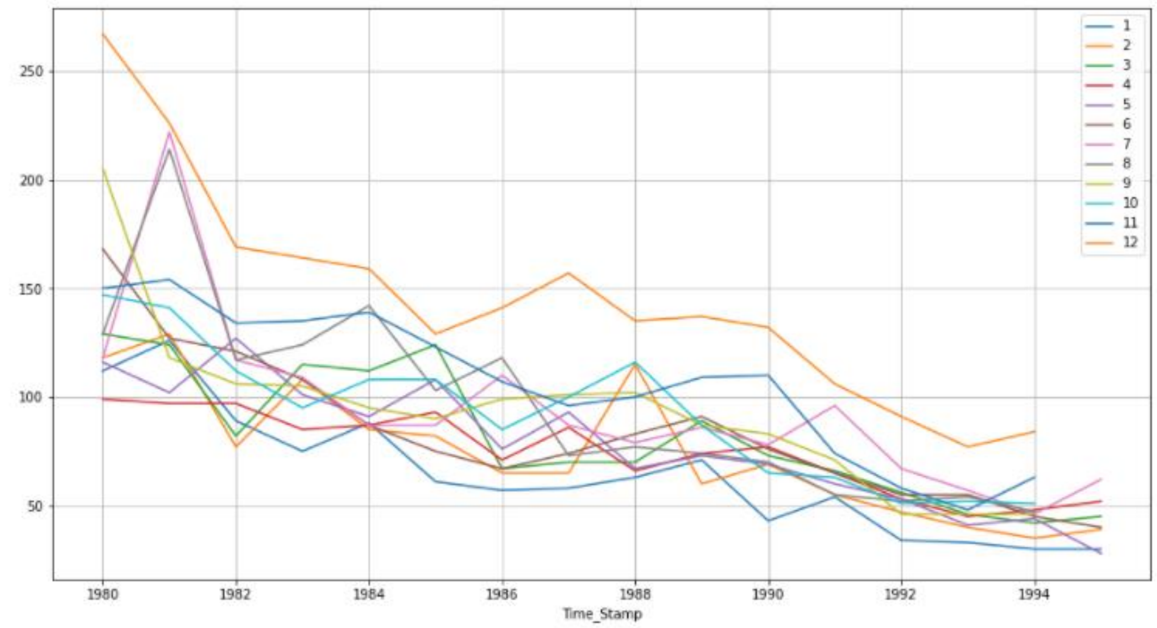
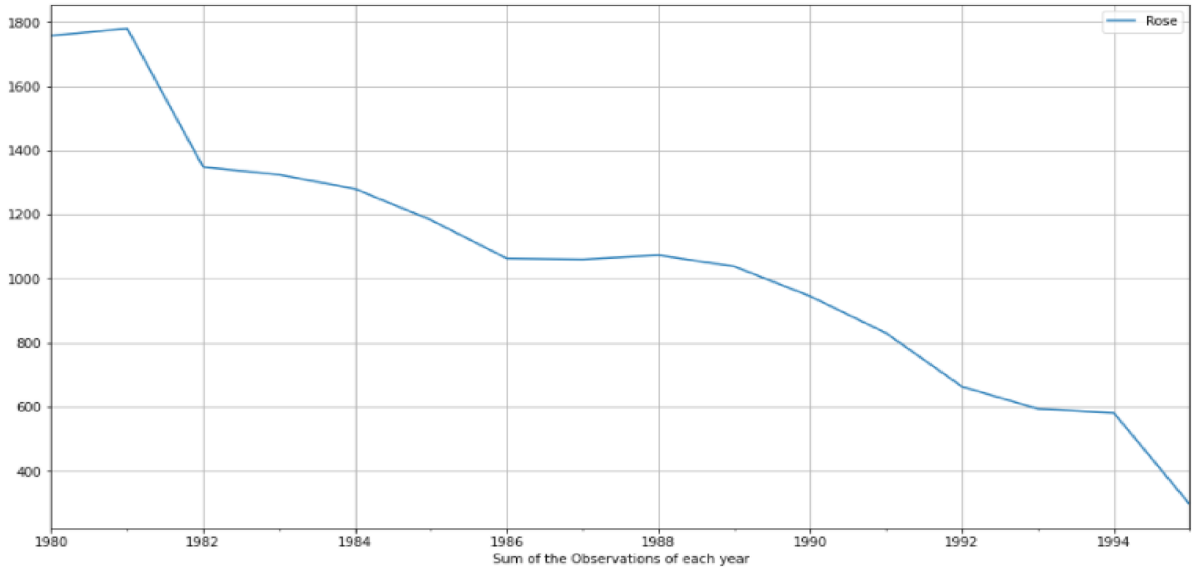


Fig6. Line plot for monthly sales

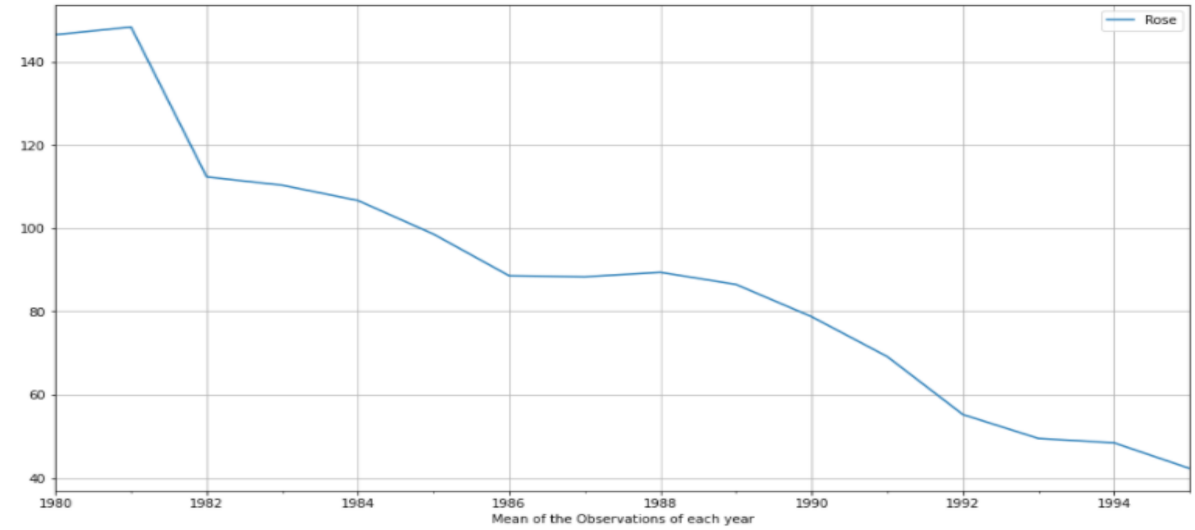
We can observe that, the line plot for monthly sales of Rose wine shows that the December month the highest sale and May, January and February show lower sale values.

Read the monthly data into a quarterly and yearly format. Compare the time series plot and draw inferences.

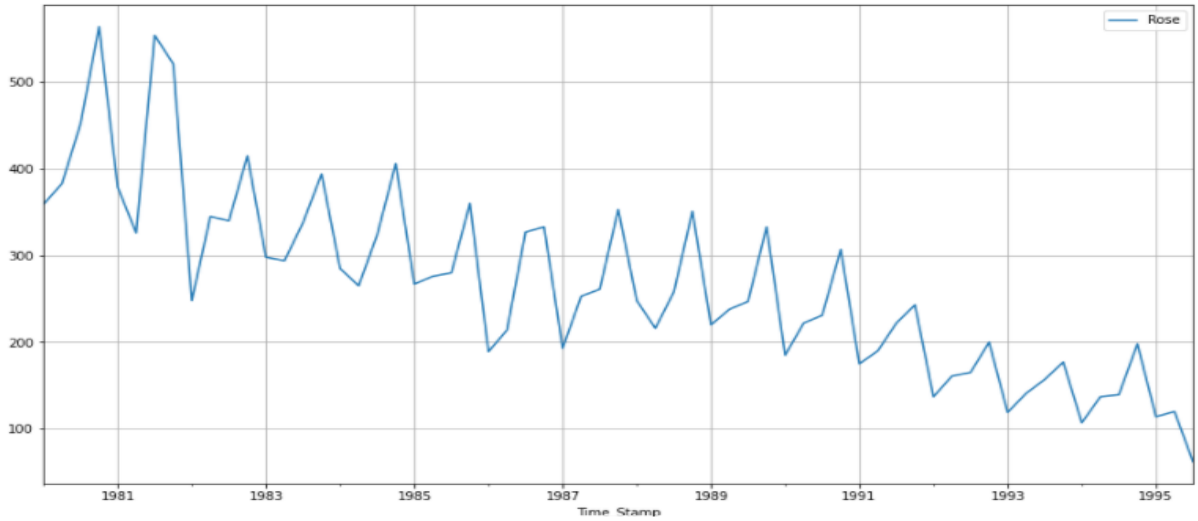
Sum of observations of each year:

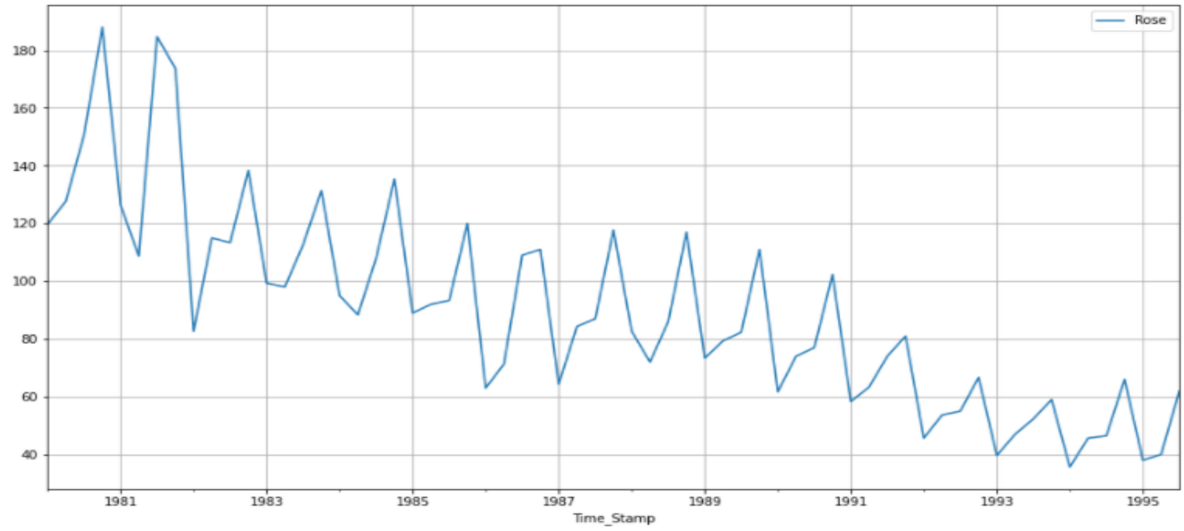


Mean observations of each year:

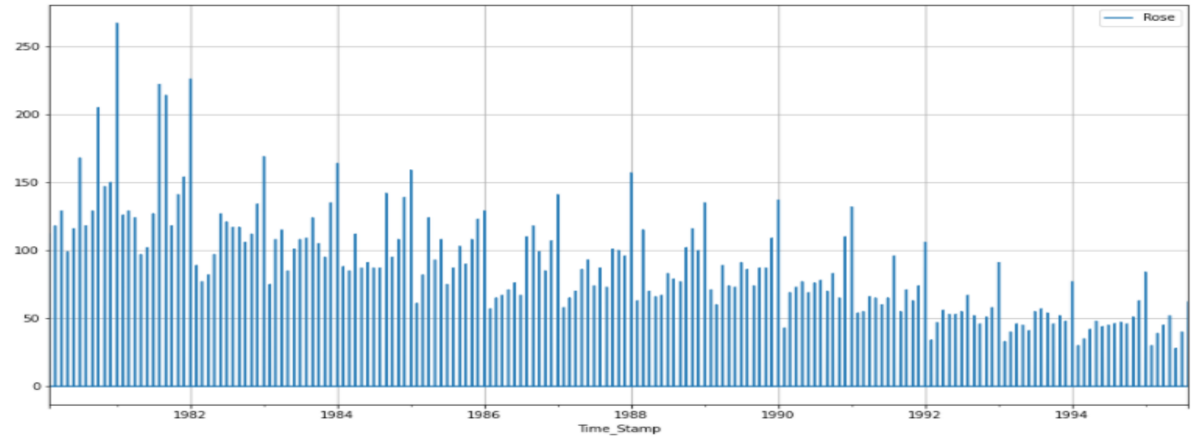


Quarterly plots:

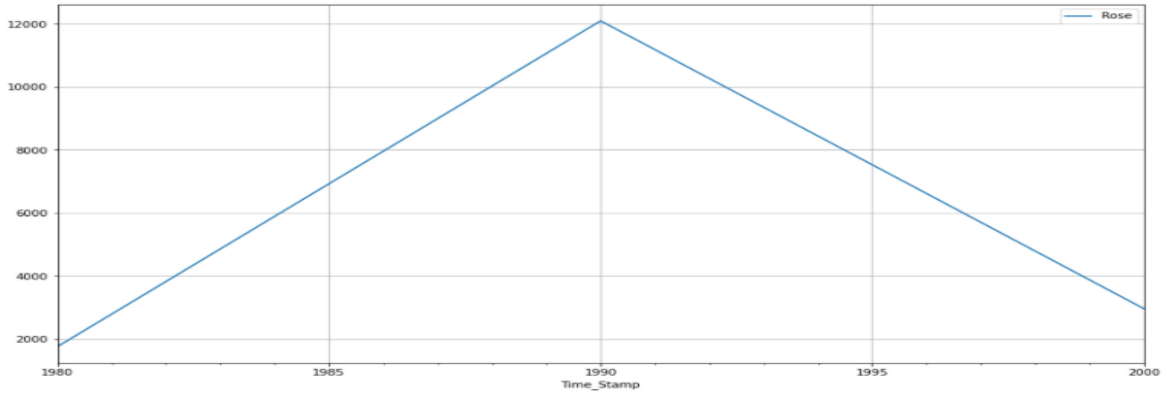




Daily plot:



Decade Plot:



Average sales & Percentage change of Sales with respect to time:

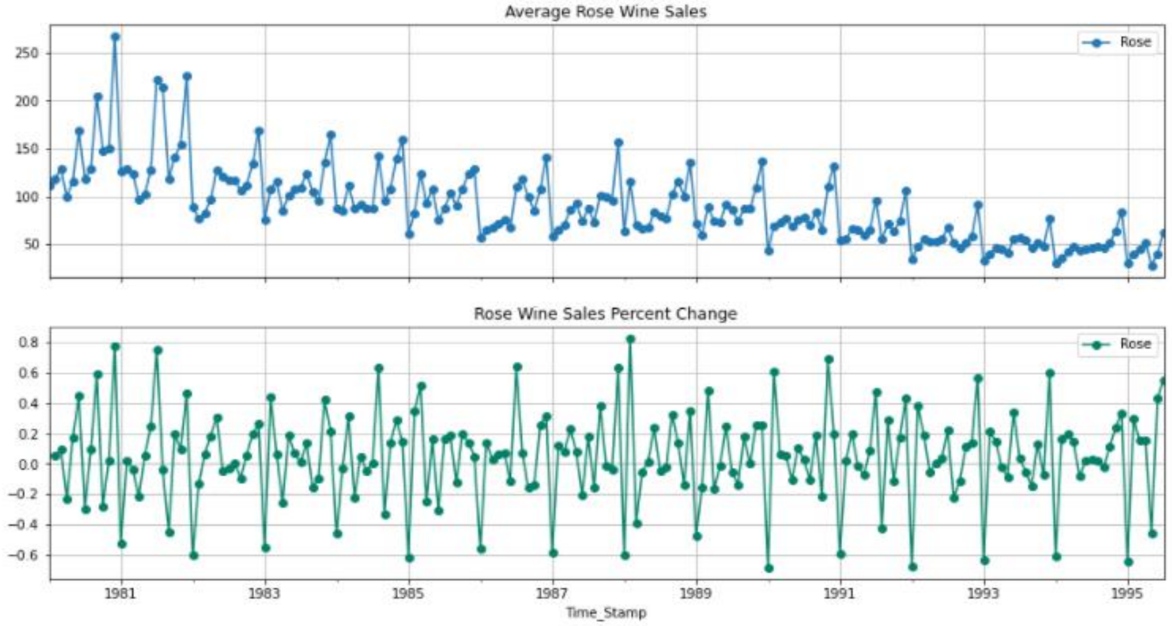


Fig7. Average sales & Percentage change of Sales

The median values keep increasing from January to December months. The Average sales value also shows a decreasing trend.

Additive Decomposition of Sparkling wine sales:

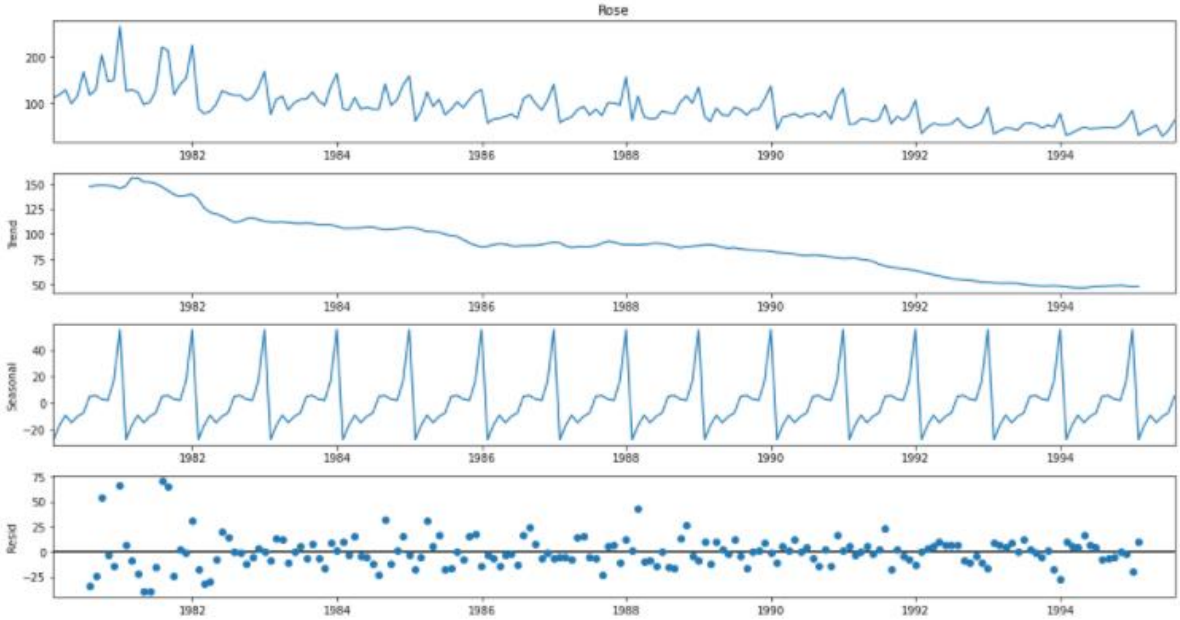


Fig8. Additive Decomposition

We can see that the residuals are located around 0 from the plot of the residuals in the decomposition.

Multiplicative Decomposition of Sparkling wine sales:

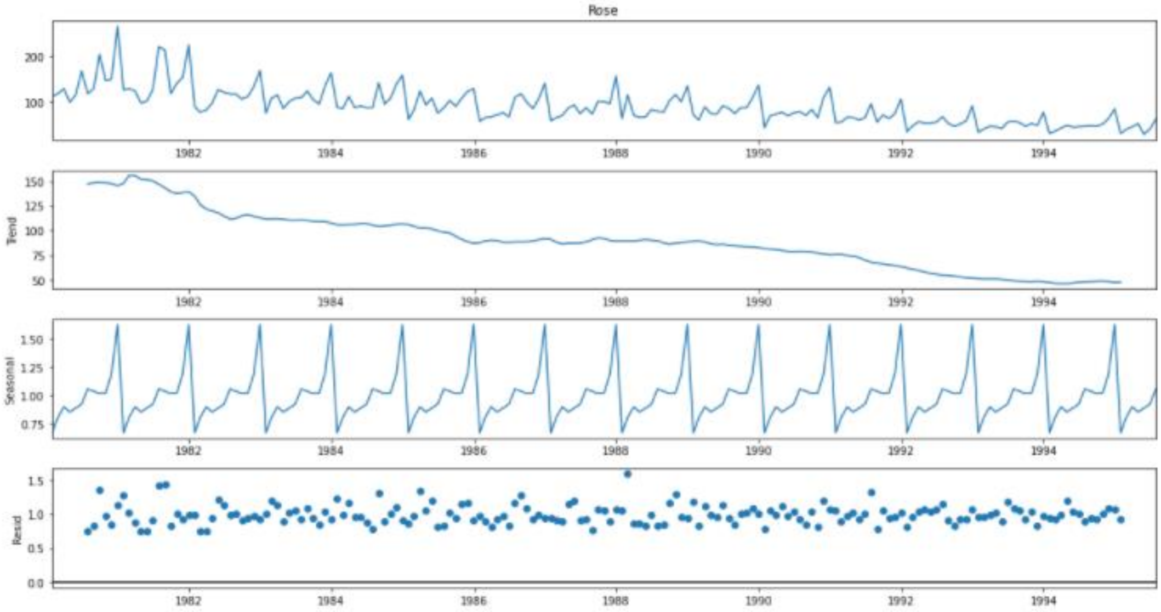


Fig9. Multiplicative Decomposition

For Multiplicative decomposition, we can see that a lot of residuals are located around 1.

Trend

```
 Time_Stamp
1980-01-31           NaN
1980-02-29           NaN
1980-03-31           NaN
1980-04-30           NaN
1980-05-31           NaN
1980-06-30           NaN
1980-07-31    147.083333
1980-08-31    148.125000
1980-09-30    148.375000
1980-10-31    148.083333
1980-11-30    147.416667
1980-12-31    145.125000
Name: trend, dtype: float64
```

Seasonality

```
 Time_Stamp
1980-01-31    0.669945
1980-02-29    0.806018
1980-03-31    0.900897
1980-04-30    0.853717
1980-05-31    0.889141
1980-06-30    0.923716
1980-07-31    1.058922
1980-08-31    1.037766
1980-09-30    1.017401
1980-10-31    1.022301
1980-11-30    1.192005
1980-12-31    1.628171
Name: seasonal, dtype: float64
```

Residual

```
 Time_Stamp
1980-01-31         NaN
1980-02-29         NaN
1980-03-31         NaN
1980-04-30         NaN
1980-05-31         NaN
1980-06-30         NaN
1980-07-31    0.757626
1980-08-31    0.839193
1980-09-30    1.358004
1980-10-31    0.971029
```

```
1980-11-30    0.853624
1980-12-31    1.129976
Name: resid, dtype: float64
```

Summary of Rose:

→Rose dataset shows a clear decreasing trend as well as seasonality, multiplicative decomposition.

→Dictates the series the noise is reduced considerably in it also the seasonal patterns increase and decrease in the size across different years.

→The sales tend to go up during the July-August and also during end of the year.

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

The train data of Sparkling has been split up to the year 1990 and has 132 data points.

The test data has been split from the year 1991 a 3. Split the data into training and test. The test data should start in 1991 and has 55 data points.

From train-test split we will be predicting the future sales in comparison with past years' sale.

Shape of train data: (132,1)

Shape of test data: (55,1)

First/last few rows of training and testing data:

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

Last few rows of Training Data
              Rose
Time_Stamp
1990-08-31   70.0
1990-09-30   83.0
1990-10-31   65.0
1990-11-30  110.0
1990-12-31  132.0

First few rows of Test Data
             Rose
Time_Stamp
```

```
1991-01-31  54.0
1991-02-28  55.0
1991-03-31  66.0
1991-04-30  65.0
1991-05-31  60.0
```

Last few rows of Test Data

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

Plot for train-test data:

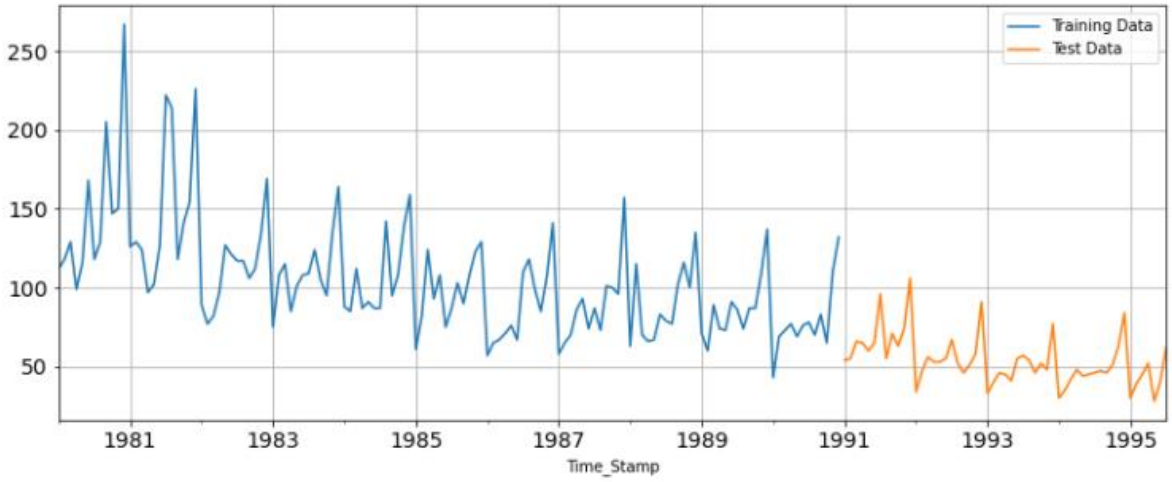


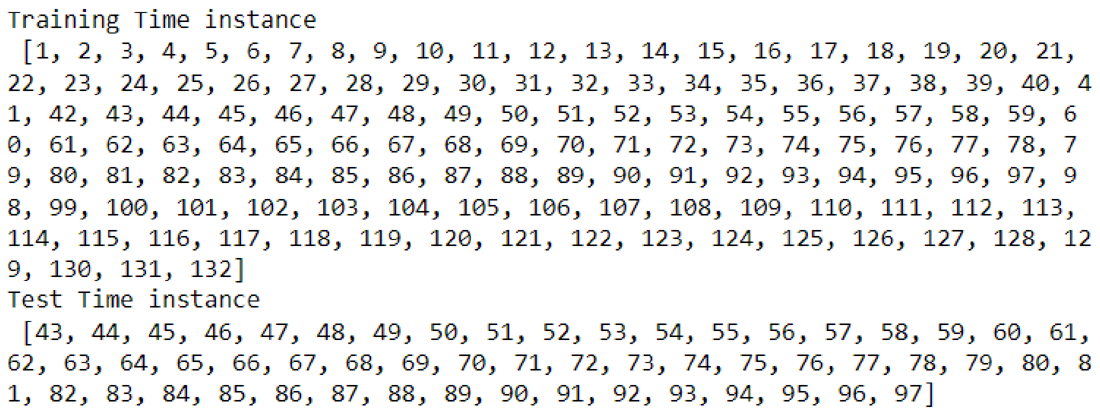
Fig10. Train-test split

4. Build all the 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 and simple average models. should also be built on the training data and check the performance on the test data using RMSE.

MODEL1: LINEAR REGRESSION

For Linear Regression, we regress the sales variable against the order of the occurrence.

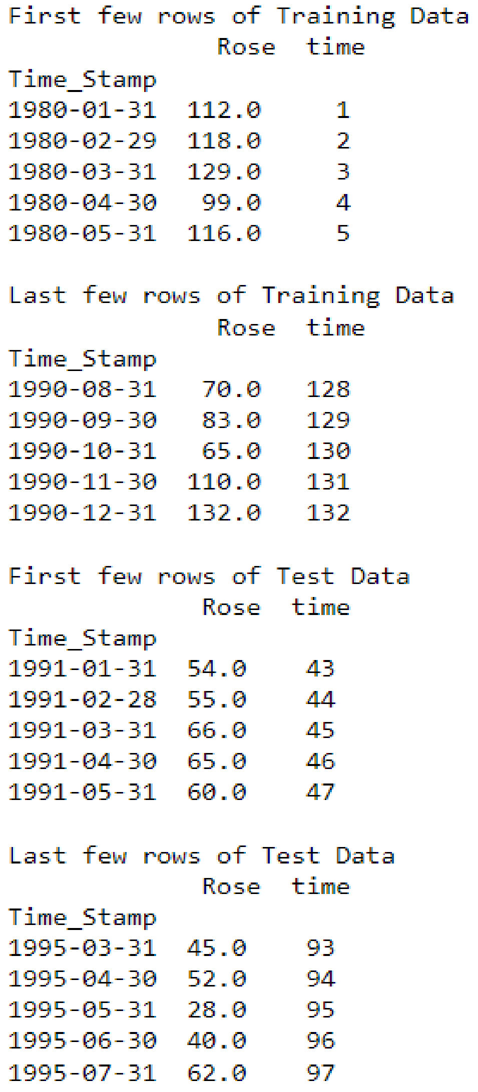
Then we generate the numerical time instance order for both train and test set.



```
Training Time instance
 [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21,
22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 4
1, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 6
0, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 7
9, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 9
8, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113,
114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 12
9, 130, 131, 132]
Test Time instance
 [43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61,
62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 8
1, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97]
```

We will add these values in the training and test set.

Hence, the train and test are thus modified to perform Linear Regression.



```
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
1990-08-31    70.0   128
1990-09-30    83.0   129
1990-10-31    65.0   130
1990-11-30   110.0   131
1990-12-31   132.0   132

First few rows of Test Data
             Rose  time
Time_Stamp
1991-01-31   54.0    43
1991-02-28   55.0    44
1991-03-31   66.0    45
1991-04-30   65.0    46
1991-05-31   60.0    47

Last few rows of Test Data
             Rose  time
Time_Stamp
1995-03-31   45.0    93
1995-04-30   52.0    94
1995-05-31   28.0    95
1995-06-30   40.0    96
1995-07-31   62.0    97
```

Linear Regression train test plot:

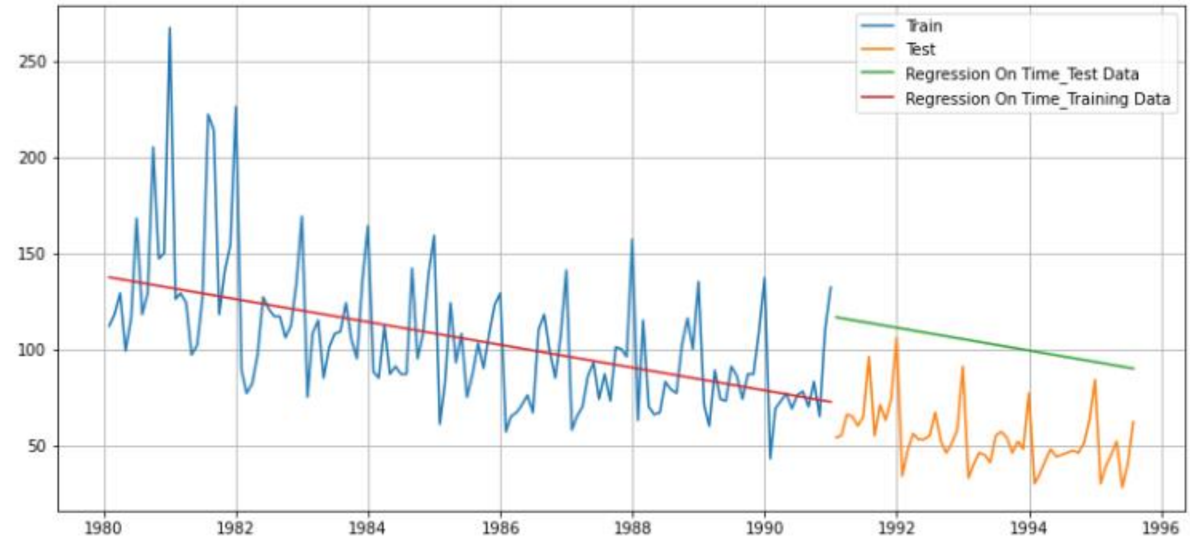


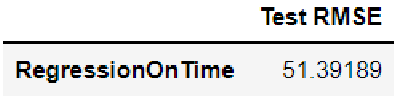
Fig11. Linear Regression train test

The predicted values for the test data using linear regression model is shown as a straight line with slope.

Regression On Time forecast on the Training Data: RMSE is 30.718

Regression On Time forecast on the Test Data: RMSE is 51.392

The RMSE for the linear regression model generated for test data



	Test RMSE
RegressionOnTime	51.39189

<u>**MODEL2: NAÏVE FORECAST MODEL**</u>

Naive Model forecast on the Training Data: RMSE is 45.064

Naive Model forecast on the Test Data: RMSE is 79.672

For the Naïve model, we observe that the red line in the plot shows a straight line which predicts sale of tomorrow is same as today.

And the prediction for day after tomorrow is same as tomorrow.

Hence, it applies to all the future years.

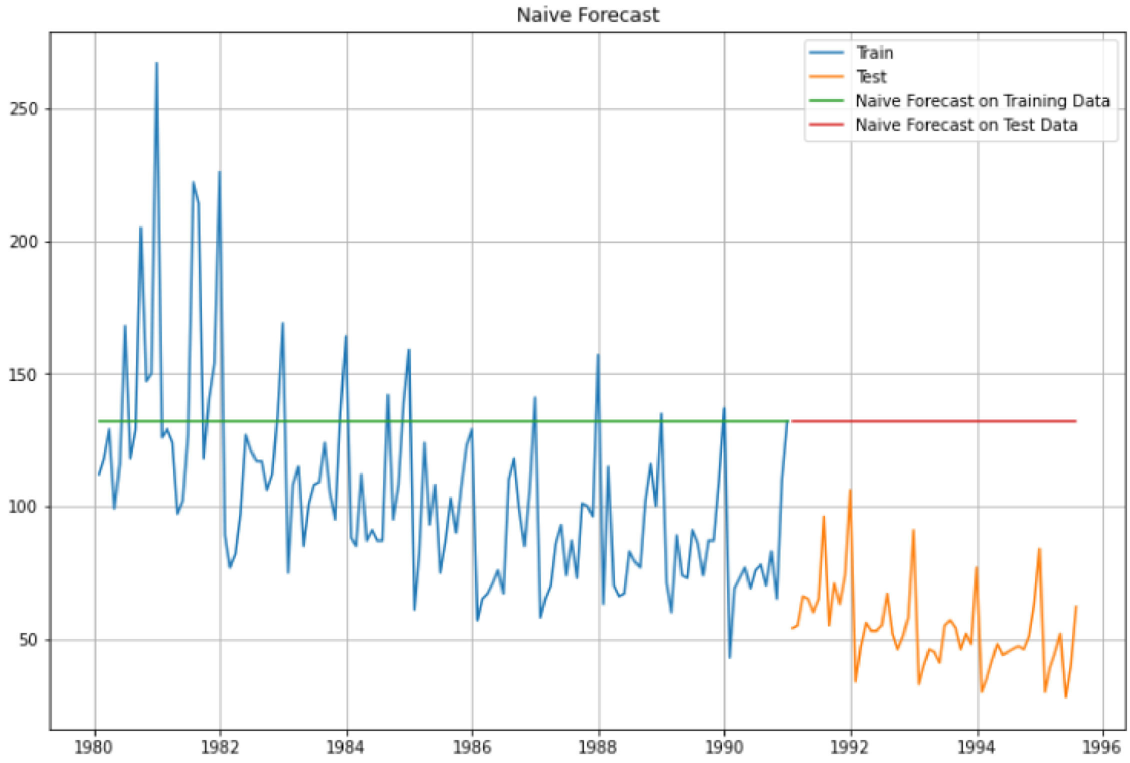


Fig12. Train-Test Naïve

The RMSE for the Naive model generated for test data: 79.672238

<u>MODEL3: SIMPLE AVERAGE MODEL</u>

In Simple Average method, we will forecast the data using the average of the training values.

From the plot below, we observe that the red line is straight and shows the simple average forecasting.

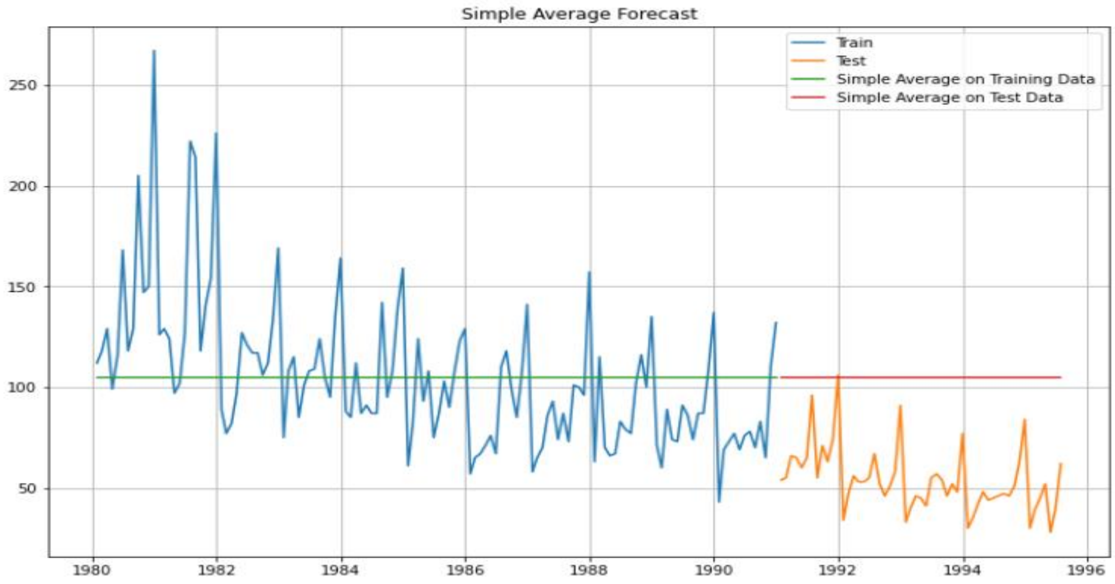


Fig13. SA-train-test

Simple Average Model forecast on the Training Data: RMSE is 36.034

Simple Average forecast on the Test Data: RMSE is 53.413

The RMSE for the Simple Average model generated for test data: 53.413057

<u>MODEL4: MOVING AVERAGE MODEL</u>

In Moving Average Model, we compute moving averages for 2,4,6,9 point intervals.

Then the best interval is determined by the maximum accuracy.

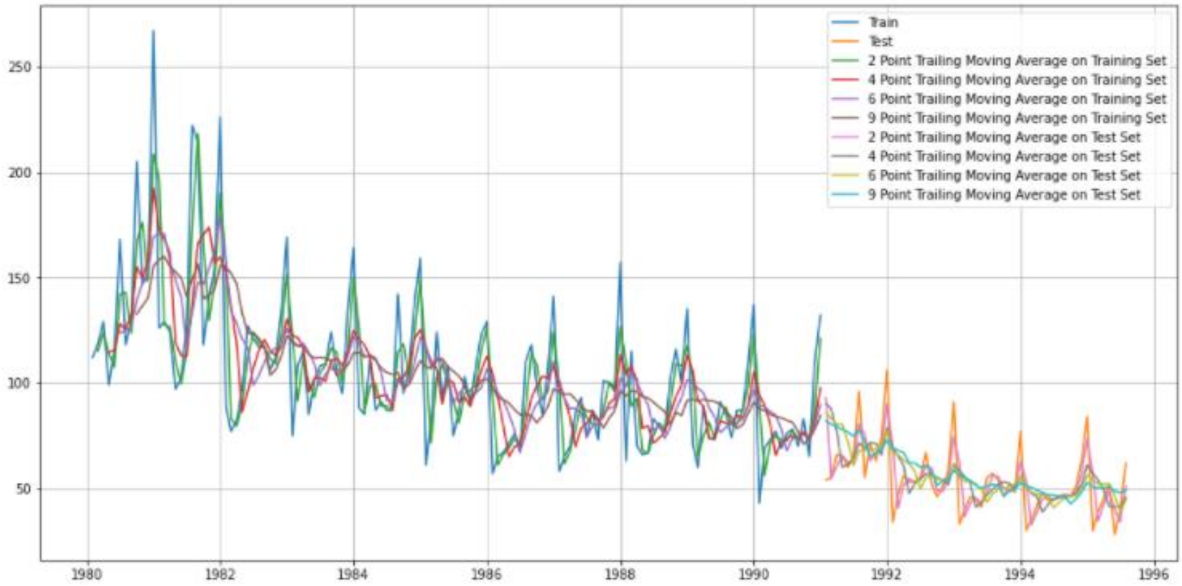
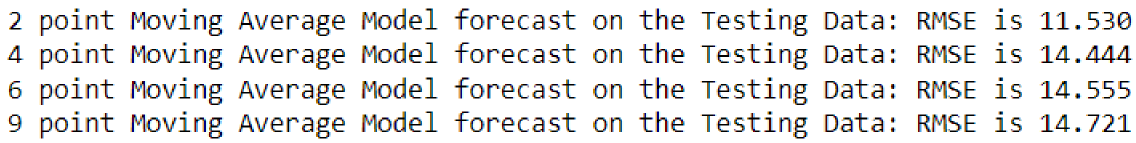
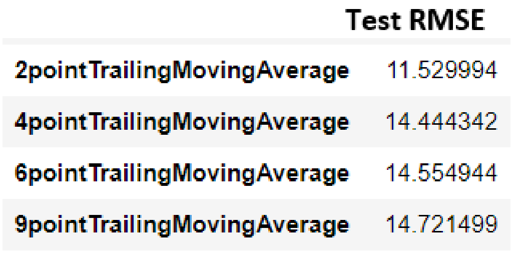


Fig14.MA Train-Test



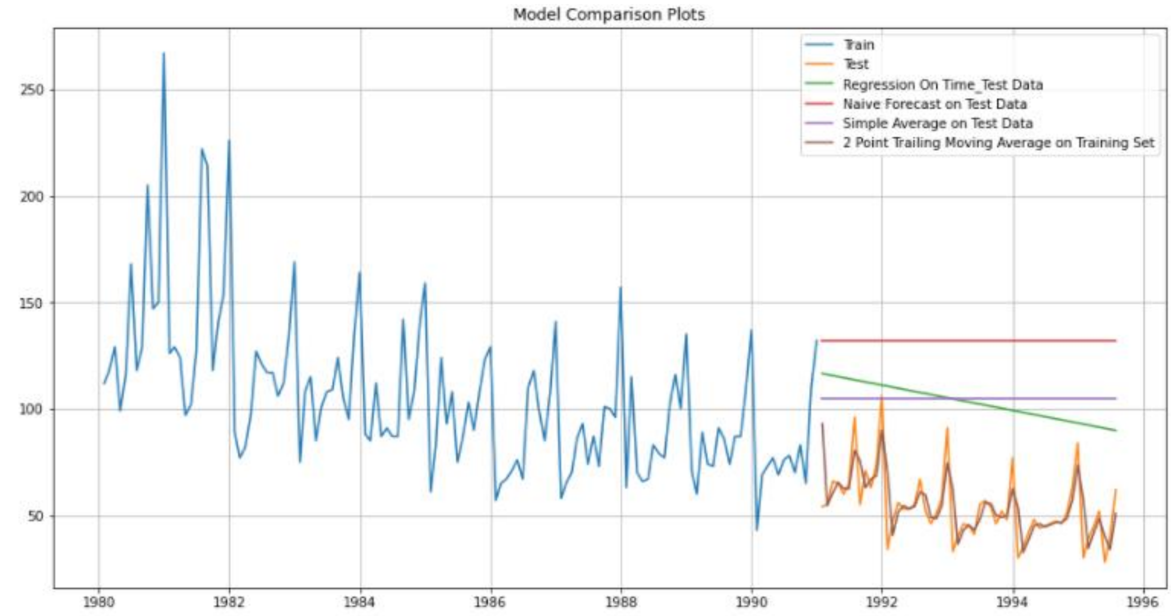
```
2 point Moving Average Model forecast on the Testing Data: RMSE is 11.530
4 point Moving Average Model forecast on the Testing Data: RMSE is 14.444
6 point Moving Average Model forecast on the Testing Data: RMSE is 14.555
9 point Moving Average Model forecast on the Testing Data: RMSE is 14.721
```



	Test RMSE
2pointTrailingMovingAverage	11.529994
4pointTrailingMovingAverage	14.444342
6pointTrailingMovingAverage	14.554944
9pointTrailingMovingAverage	14.721499

From the above table, we see that 2point trailing moving average has the least score.

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



MODEL5: SIMPLE EXPONENTIAL SMOOTHING

Simple Exponential Smoothing, is **a time series forecasting method for univariate data without a trend or seasonality**. It requires a single parameter, called alpha (a), also called the smoothing factor or smoothing coefficient._This method is suitable for forecasting data with no clear trend or seasonal pattern.

It requires a single parameter, called *alpha* (*a*), also called the smoothing factor or smoothing coefficient.
This parameter controls the rate at which the influence of the observations at prior time steps decay exponentially. Alpha is often set to a value between 0 and 1.

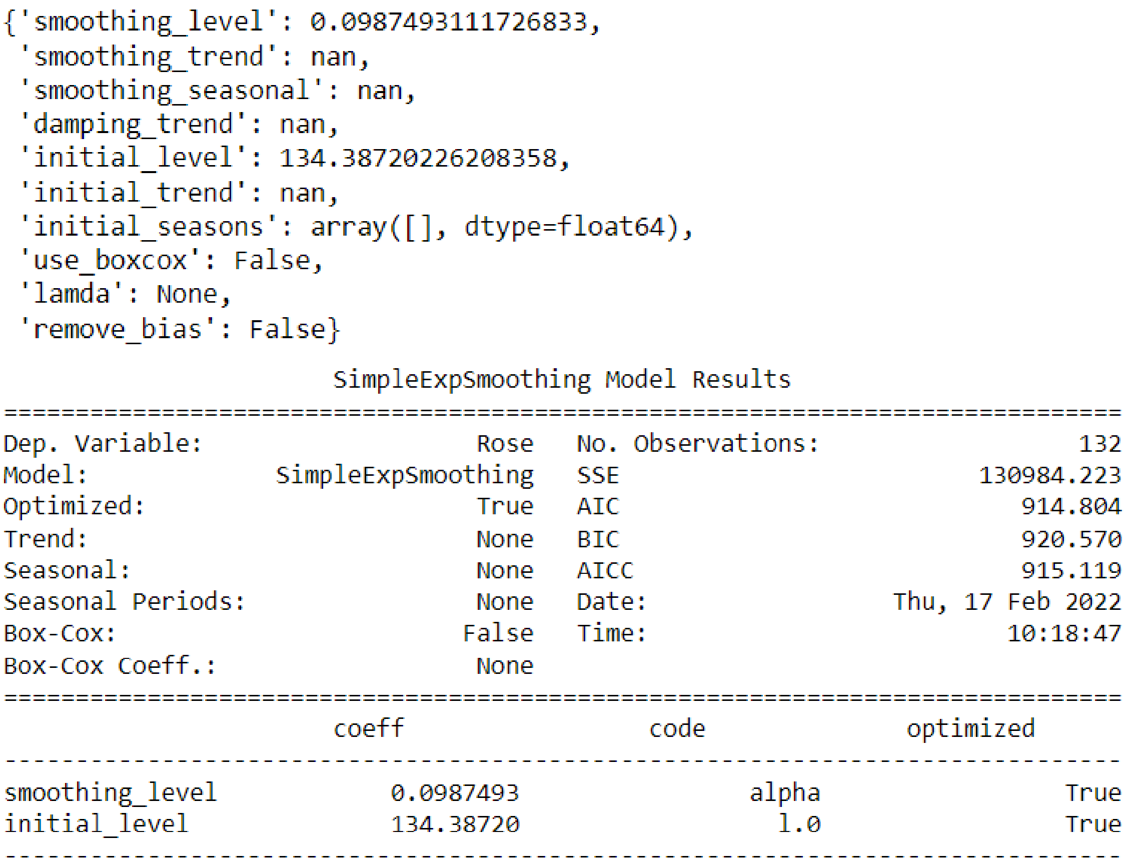
Large values mean that the model pays attention mainly to the most recent past observations, whereas smaller values mean more of the history is taken into account when making a prediction.

A value close to 1 indicates fast learning (that is, only the most recent values influence the forecasts), whereas a value close to 0 indicates slow learning (past observations have a large influence on forecasts).

Hyperparameters:

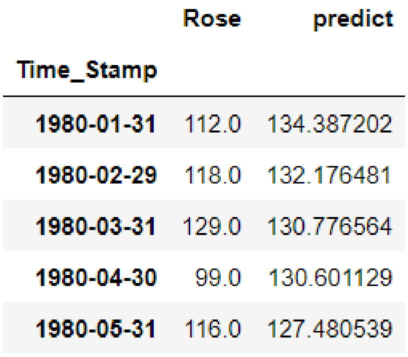
Alpha: Smoothing factor for the level.

Parameters:



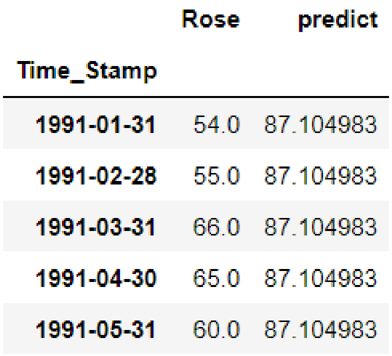
```
{'smoothing_level': 0.0987493111726833,
 'smoothing_trend': nan,
 'smoothing_seasonal': nan,
 'damping_trend': nan,
 'initial_level': 134.38720226208358,
 'initial_trend': nan,
 'initial_seasons': array([], dtype=float64),
 'use_boxcox': False,
 'lamda': None,
 'remove_bias': False}
                       SimpleExpSmoothing Model Results
==============================================================================
Dep. Variable:                     Rose   No. Observations:                  132
Model:               SimpleExpSmoothing   SSE                         130984.223
Optimized:                         True   AIC                            914.804
Trend:                             None   BIC                            920.570
Seasonal:                          None   AICC                           915.119
Seasonal Periods:                  None   Date:                 Thu, 17 Feb 2022
Box-Cox:                          False   Time:                         10:18:47
Box-Cox Coeff.:                    None
==============================================================================
                       coeff                 code              optimized
------------------------------------------------------------------------------
smoothing_level            0.0987493                alpha                 True
initial_level              134.38720                  1.0                 True
------------------------------------------------------------------------------
```

Predict on train:



	Rose	predict
Time_Stamp		
1980-01-31	112.0	134.387202
1980-02-29	118.0	132.176481
1980-03-31	129.0	130.776564
1980-04-30	99.0	130.601129
1980-05-31	116.0	127.480539

Predict on Test:



	Rose	predict
Time_Stamp		
1991-01-31	54.0	87.104983
1991-02-28	55.0	87.104983
1991-03-31	66.0	87.104983
1991-04-30	65.0	87.104983
1991-05-31	60.0	87.104983

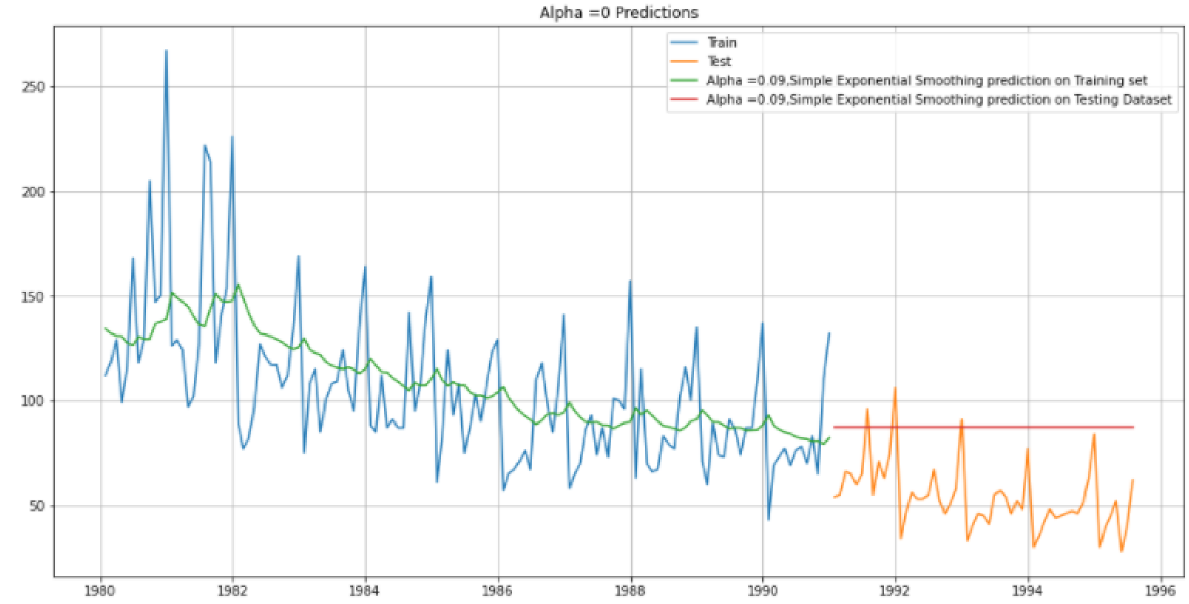


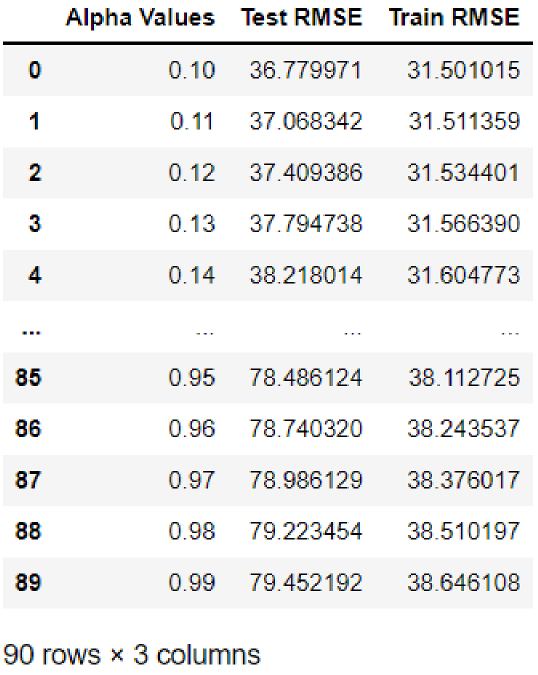
Fig15. SES Train-Test

Alpha =0 Simple Exponential Smoothing Model forecast on the Training Data: RMSE is 31.50 1
Alpha =0 Simple Exponential Smoothing Model forecast on the Training Data: **RMSE is 36.74 8**

MODEL6: SIMPLE EXPONENTIAL SMOOTHING WITH ALPHA IN RANGE OF 0.01 TO 0.1

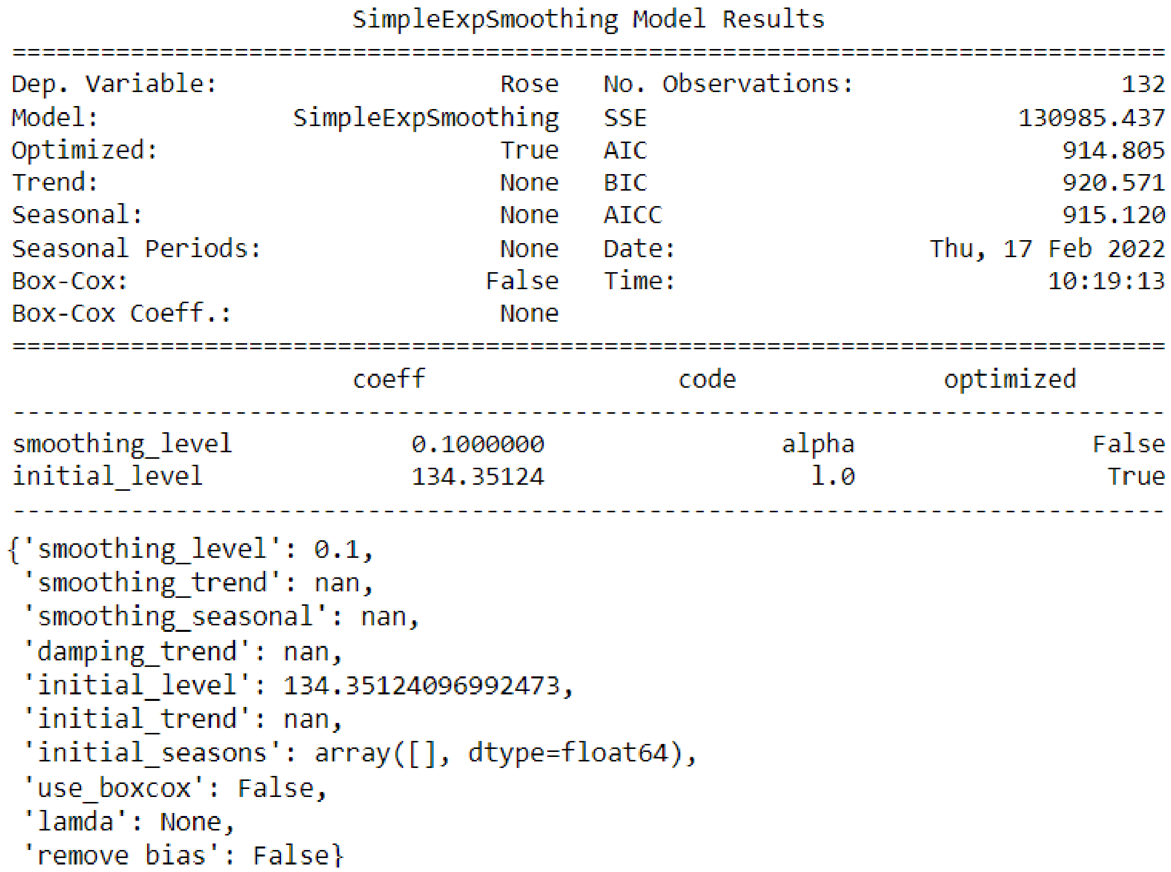
In Simple Exponential Smoothing Model, we will run a loop with different alpha values to understand which particular value is best.

Alpha value ranges from 0.01 to 0.1



	Alpha Values	Test RMSE	Train RMSE
0	0.10	36.779971	31.501015
1	0.11	37.068342	31.511359
2	0.12	37.409386	31.534401
3	0.13	37.794738	31.566390
4	0.14	38.218014	31.604773
...	...	...	...
85	0.95	78.486124	38.112725
86	0.96	78.740320	38.243537
87	0.97	78.986129	38.376017
88	0.98	79.223454	38.510197
89	0.99	79.452192	38.646108

90 rows × 3 columns



```
                        SimpleExpSmoothing Model Results
==============================================================================
Dep. Variable:                     Rose   No. Observations:                  132
Model:               SimpleExpSmoothing   SSE                         130985.437
Optimized:                         True   AIC                            914.805
Trend:                             None   BIC                            920.571
Seasonal:                          None   AICC                           915.120
Seasonal Periods:                  None   Date:                 Thu, 17 Feb 2022
Box-Cox:                          False   Time:                         10:19:13
Box-Cox Coeff.:                    None
==============================================================================
                       coeff                 code              optimized
------------------------------------------------------------------------------
smoothing_level            0.1000000                alpha                False
initial_level              134.35124                  1.0                 True
------------------------------------------------------------------------------
{'smoothing_level': 0.1,
 'smoothing_trend': nan,
 'smoothing_seasonal': nan,
 'damping_trend': nan,
 'initial_level': 134.35124096992473,
 'initial_trend': nan,
 'initial_seasons': array([], dtype=float64),
 'use_boxcox': False,
 'lamda': None,
 'remove_bias': False}
```

The RMSE for the Simple Exponential smoothing model (Alpha=0.01) generated for test data: 36.779971

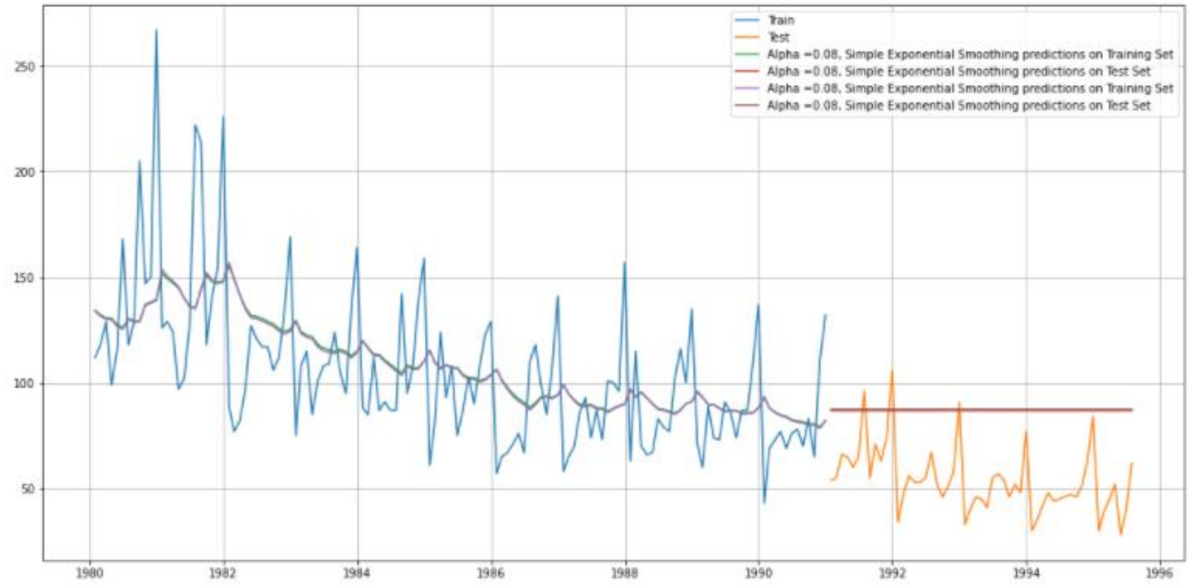


Fig16. SES (Alpha) Train-Test

MODEL7: DOUBLE EXPONENTIAL SMOOTHING

Double Exponential Smoothing is an extension to Exponential Smoothing that explicitly adds support for trends in the univariate time series.

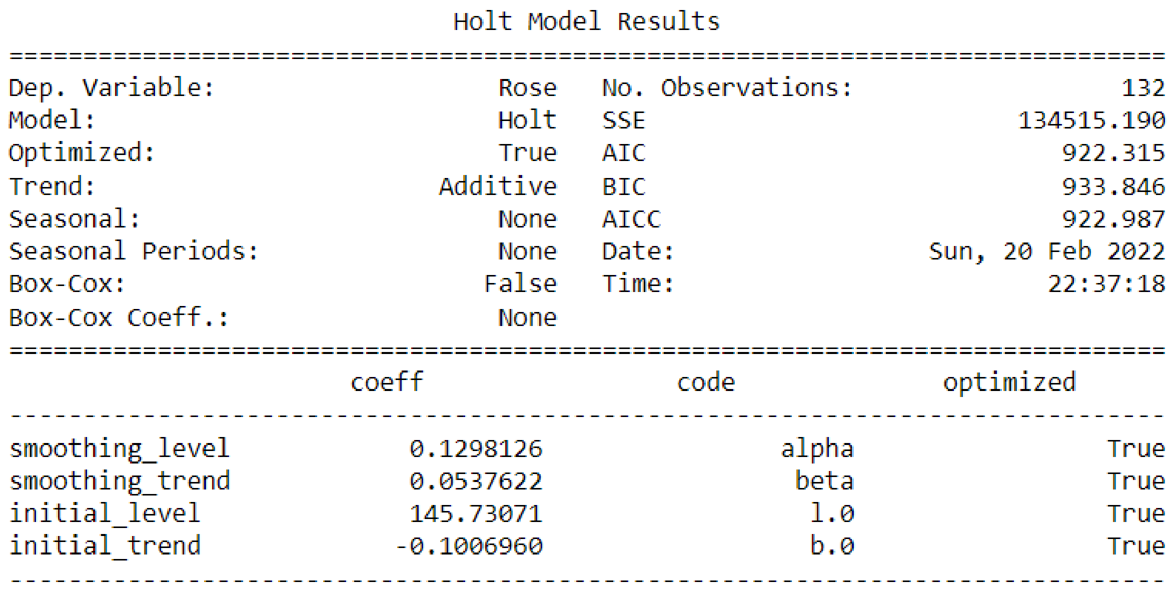
In addition to the alpha parameter for controlling smoothing factor for the level, an additional smoothing factor is added to control the decay of the influence of the change in trend called beta(*b*).

The method supports trends that change in different ways: an additive and a multiplicative, depending on whether the trend is linear or exponential respectively.

Double Exponential Smoothing with an additive trend is classically referred to as Holt's linear trend model, named for the developer of the method Charles Holt.

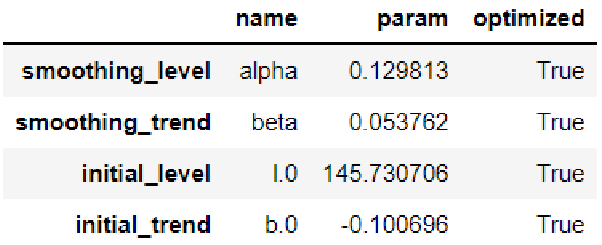
- **Additive Trend**: Double Exponential Smoothing with a linear trend.
- **Multiplicative Trend**: Double Exponential Smoothing with an exponential trend.

For longer range (multi-step) forecasts, the trend may continue on unrealistically. As such, it can be useful to dampen the trend over time.



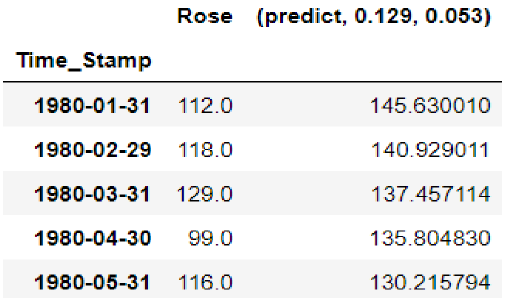
```
                              Holt Model Results
==============================================================================
Dep. Variable:                   Rose   No. Observations:                  132
Model:                           Holt   SSE                         134515.190
Optimized:                       True   AIC                            922.315
Trend:                       Additive   BIC                            933.846
Seasonal:                        None   AICC                           922.987
Seasonal Periods:                None   Date:                 Sun, 20 Feb 2022
Box-Cox:                        False   Time:                         22:37:18
Box-Cox Coeff.:                  None
==============================================================================
                       coeff                 code              optimized
------------------------------------------------------------------------------
smoothing_level            0.1298126                alpha                 True
smoothing_trend            0.0537622                 beta                 True
initial_level              145.73071                  l.0                 True
initial_trend             -0.1006960                  b.0                 True
------------------------------------------------------------------------------
```

Parameters formatted:



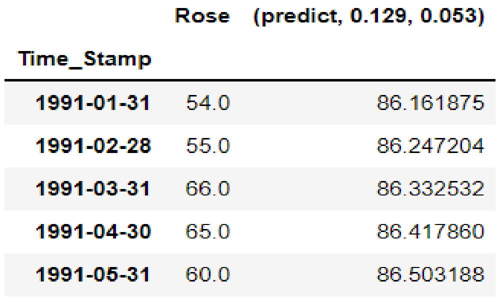
	name	param	optimized
smoothing_level	alpha	0.129813	True
smoothing_trend	beta	0.053762	True
initial_level	l.0	145.730706	True
initial_trend	b.0	-0.100696	True

Predictions on Training data:



Time_Stamp	Rose	(predict, 0.129, 0.053)
1980-01-31	112.0	145.630010
1980-02-29	118.0	140.929011
1980-03-31	129.0	137.457114
1980-04-30	99.0	135.804830
1980-05-31	116.0	130.215794

Predictions on Testing data:



Time_Stamp	Rose	(predict, 0.129, 0.053)
1991-01-31	54.0	86.161875
1991-02-28	55.0	86.247204
1991-03-31	66.0	86.332532
1991-04-30	65.0	86.417860
1991-05-31	60.0	86.503188

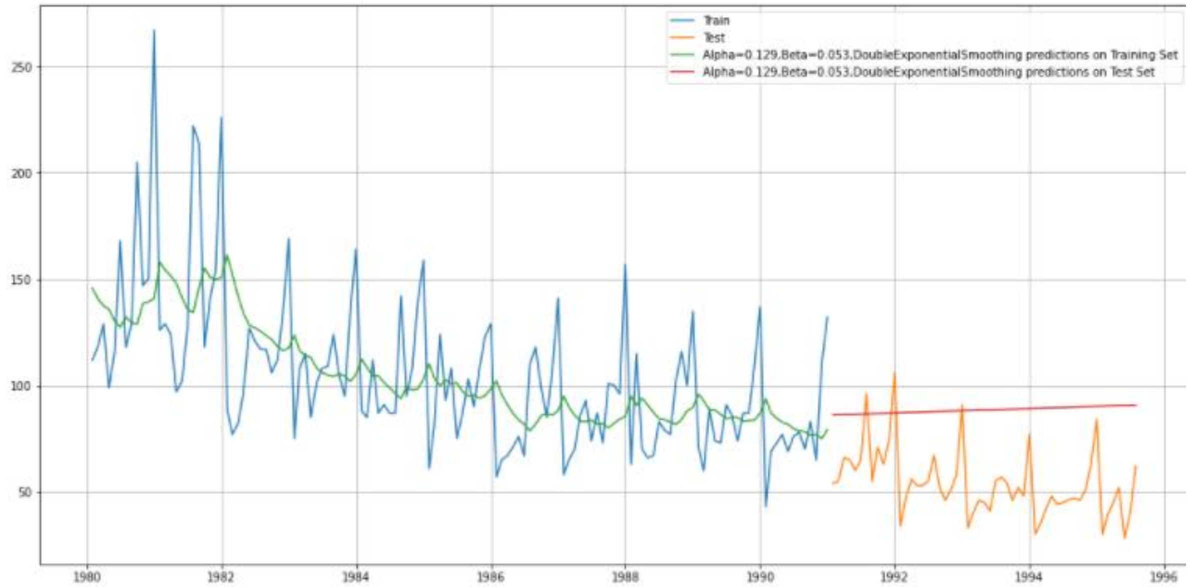
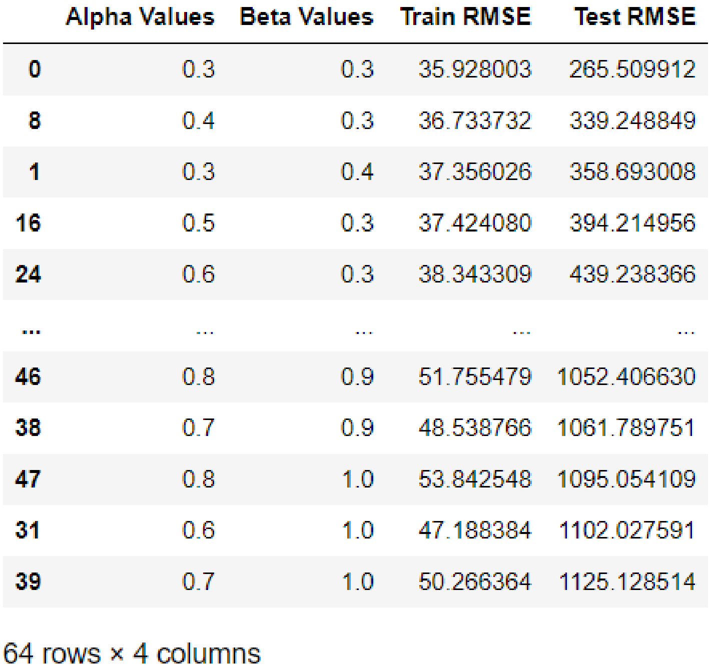


Fig17. DES Train-Test

Alpha=0.129 and Beta=0.053 Double Exponential Smoothing Model forecast on the Training Data: RMSE is 31.923
Alpha=0.129 and Beta=0.053 Double Exponential Smoothing Model forecast on the Testing Data: RMSE is 38.232

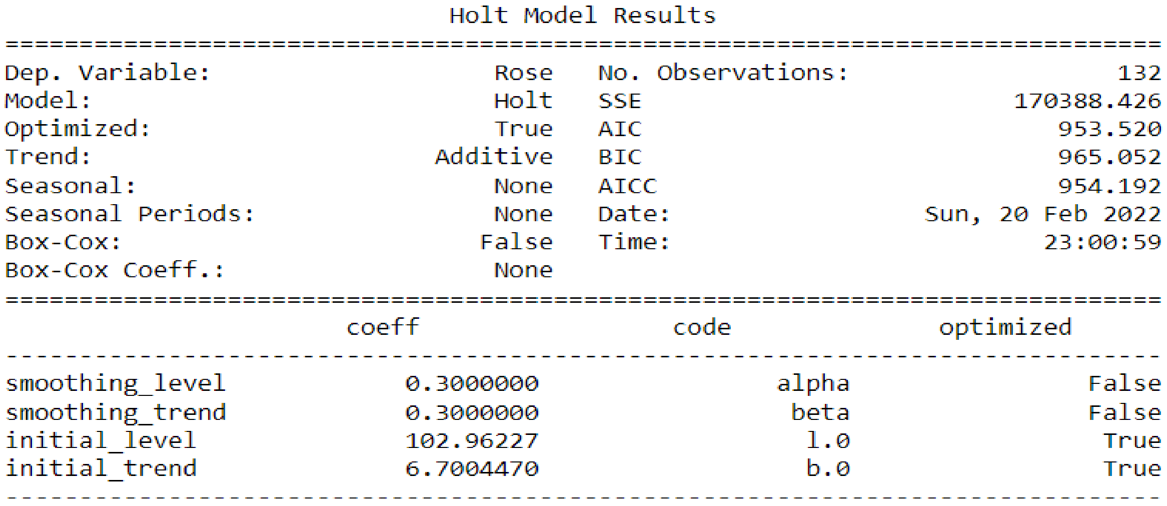
MODEL8: DOUBLE EXPONENTIAL SMOOTHING IN RANGE 0.01 TO 1

In Double Exponential Smoothing Model, we will run a loop with different alpha, beta values to understand which particular value is best.



	Alpha Values	Beta Values	Train RMSE	Test RMSE
0	0.3	0.3	35.928003	265.509912
8	0.4	0.3	36.733732	339.248849
1	0.3	0.4	37.356026	358.693008
16	0.5	0.3	37.424080	394.214956
24	0.6	0.3	38.343309	439.238366
...	...	...	...	...
46	0.8	0.9	51.755479	1052.406630
38	0.7	0.9	48.538766	1061.789751
47	0.8	1.0	53.842548	1095.054109
31	0.6	1.0	47.188384	1102.027591
39	0.7	1.0	50.266364	1125.128514

64 rows × 4 columns



```
                              Holt Model Results
==============================================================================
Dep. Variable:                     Rose   No. Observations:                  132
Model:                             Holt   SSE                         170388.426
Optimized:                         True   AIC                            953.520
Trend:                         Additive   BIC                            965.052
Seasonal:                          None   AICC                           954.192
Seasonal Periods:                  None   Date:                 Sun, 20 Feb 2022
Box-Cox:                          False   Time:                         23:00:59
Box-Cox Coeff.:                    None
==============================================================================
                       coeff                 code               optimized
------------------------------------------------------------------------------
smoothing_level            0.3000000                alpha                False
smoothing_trend            0.3000000                 beta                False
initial_level              102.96227                  l.0                 True
initial_trend              6.7004470                  b.0                 True
------------------------------------------------------------------------------
```

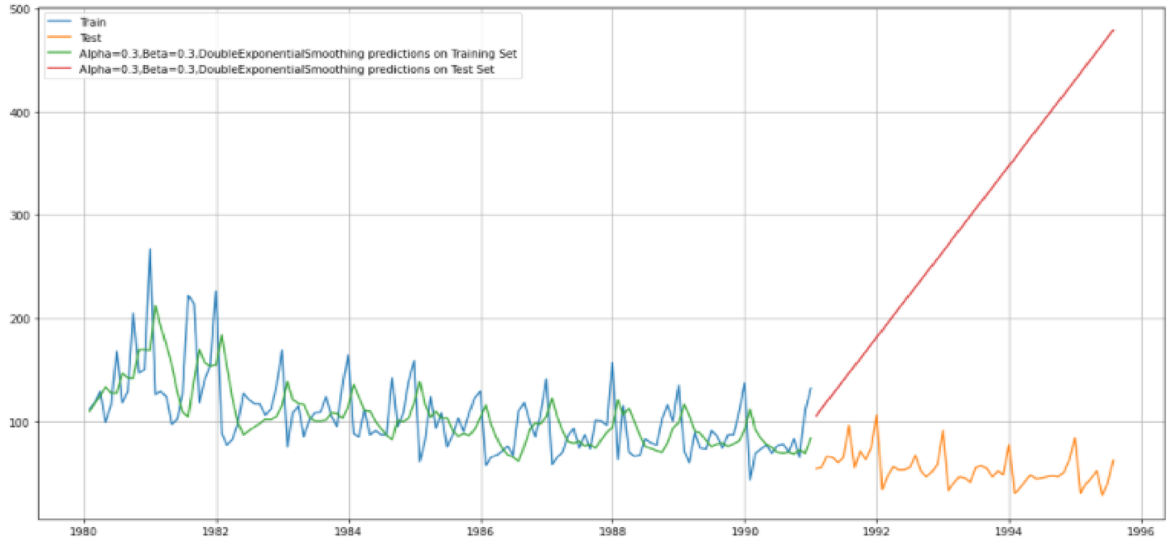


Fig18. DES(Alpha,beta) Train-Test

MODEL9: TRIPLE EXPONENTIAL SMOOTHING

Triple Exponential Smoothing is an extension of Exponential Smoothing that explicitly adds support for seasonality to the univariate time series.

This method is sometimes called Holt-Winters Exponential Smoothing, named for two contributors to the method: Charles Holt and Peter Winters.

In addition to the alpha and beta smoothing factors, a new parameter is added called *gamma* (*g*) that controls the influence on the seasonal component.

As with the trend, the seasonality may be modeled as either an additive or multiplicative process for a linear or exponential change in the seasonality.

- **Additive Seasonality**: Triple Exponential Smoothing with a linear seasonality.
- **Multiplicative Seasonality**: Triple Exponential Smoothing with an exponential seasonality.

Triple exponential smoothing is the most advanced variation of exponential smoothing and through configuration, it can also develop double and single exponential smoothing models.

Being an adaptive method, Holt-Winter's exponential smoothing allows the level, trend and seasonality patterns to change over time.

In Triple Exponential smoothing we have three parameters:

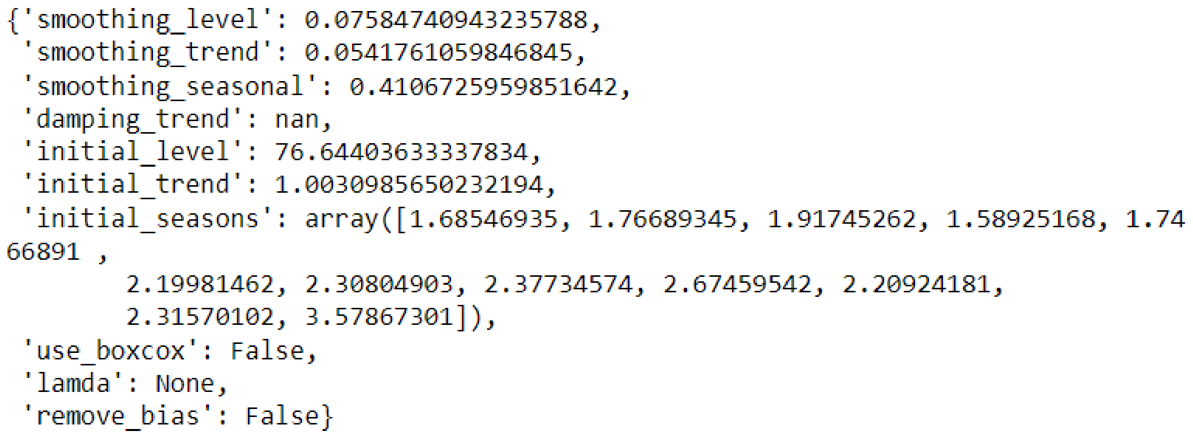
Alpha, Beta, Gamma

Smoothing level value represents Alpha

Smoothing trend value represents Beta

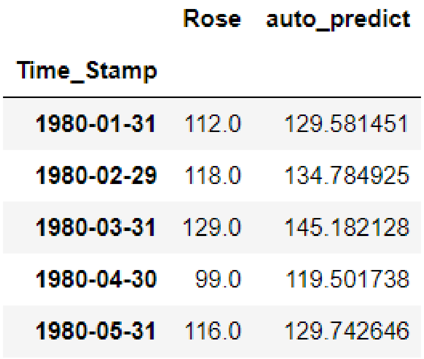
Smoothing Seasonality value represents Gamma

Parameters:



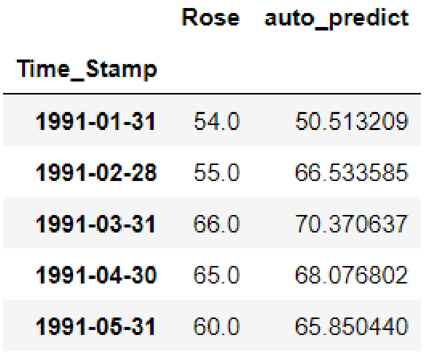
```
{'smoothing_level': 0.07584740943235788,
 'smoothing_trend': 0.0541761059846845,
 'smoothing_seasonal': 0.4106725959851642,
 'damping_trend': nan,
 'initial_level': 76.64403633337834,
 'initial_trend': 1.0030985650232194,
 'initial_seasons': array([1.68546935, 1.76689345, 1.91745262, 1.58925168, 1.74
66891 ,
        2.19981462, 2.30804903, 2.37734574, 2.67459542, 2.20924181,
        2.31570102, 3.57867301]),
 'use_boxcox': False,
 'lamda': None,
 'remove_bias': False}
```

Prediction on train data

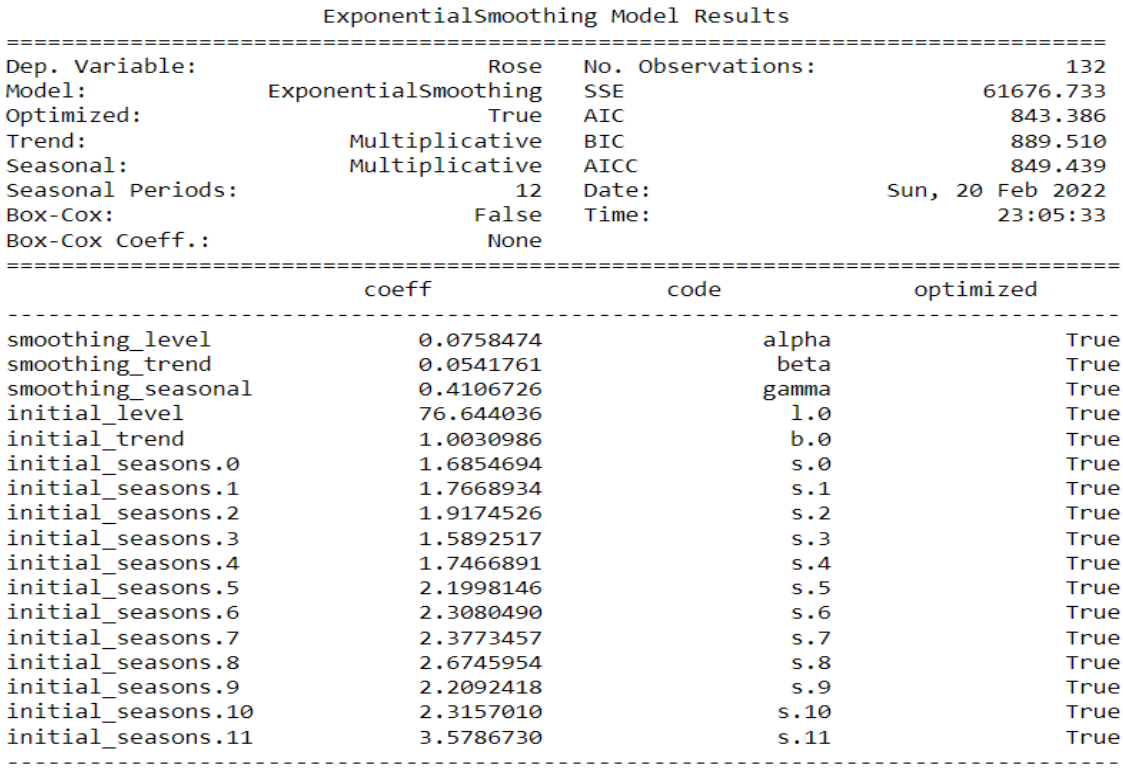


	Rose	auto_predict
Time_Stamp		
1980-01-31	112.0	129.581451
1980-02-29	118.0	134.784925
1980-03-31	129.0	145.182128
1980-04-30	99.0	119.501738
1980-05-31	116.0	129.742646

Prediction on test data:



	Rose	auto_predict
Time_Stamp		
1991-01-31	54.0	50.513209
1991-02-28	55.0	66.533585
1991-03-31	66.0	70.370637
1991-04-30	65.0	68.076802
1991-05-31	60.0	65.850440



```
                       ExponentialSmoothing Model Results
================================================================================
Dep. Variable:                          Rose   No. Observations:                  132
Model:                  ExponentialSmoothing   SSE                          61676.733
Optimized:                              True   AIC                            843.386
Trend:                        Multiplicative   BIC                            889.510
Seasonal:                     Multiplicative   AICC                           849.439
Seasonal Periods:                         12   Date:                 Sun, 20 Feb 2022
Box-Cox:                               False   Time:                         23:05:33
Box-Cox Coeff.:                         None
=================================================================================
                          coeff                 code              optimized
---------------------------------------------------------------------------------
smoothing_level               0.0758474                alpha                 True
smoothing_trend               0.0541761                 beta                 True
smoothing_seasonal            0.4106726                gamma                 True
initial_level                 76.644036                  l.0                 True
initial_trend                 1.0030986                  b.0                 True
initial_seasons.0             1.6854694                  s.0                 True
initial_seasons.1             1.7668934                  s.1                 True
initial_seasons.2             1.9174526                  s.2                 True
initial_seasons.3             1.5892517                  s.3                 True
initial_seasons.4             1.7466891                  s.4                 True
initial_seasons.5             2.1998146                  s.5                 True
initial_seasons.6             2.3080490                  s.6                 True
initial_seasons.7             2.3773457                  s.7                 True
initial_seasons.8             2.6745954                  s.8                 True
initial_seasons.9             2.2092418                  s.9                 True
initial_seasons.10            2.3157010                 s.10                 True
initial_seasons.11            3.5786730                 s.11                 True
---------------------------------------------------------------------------------
```

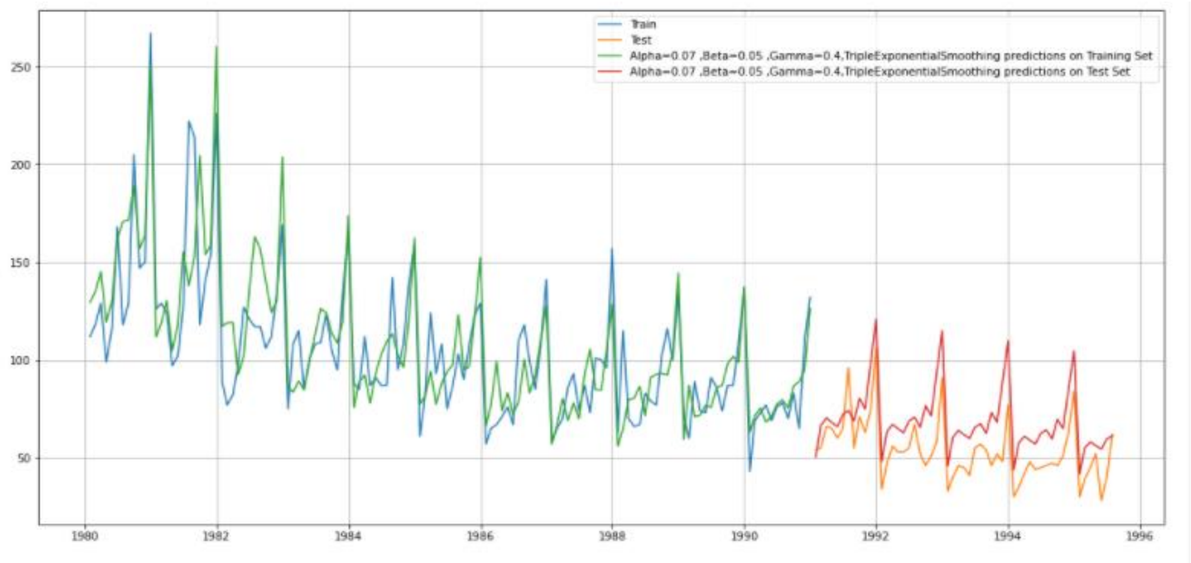


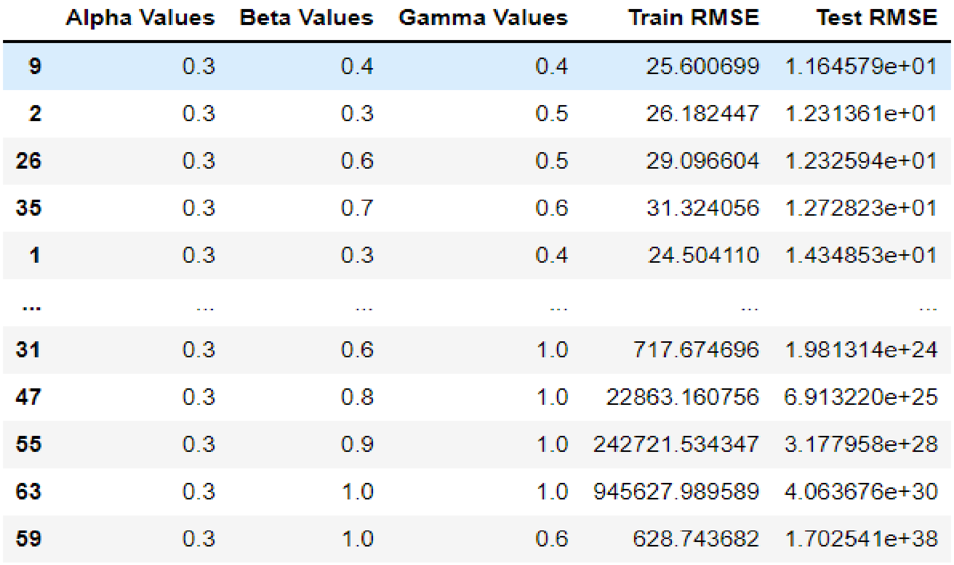
Fig19. TES Train-Test

Alpha: 0.07, Beta: 0.05 and Gamma: 0.4, Triple Exponential Smoothing Model forecast on th e Training Data: RMSE is 21.616
Alpha: 0.07, Beta: 0.05 and Gamma: 0.4, Triple Exponential Smoothing Model forecast on th e Test Data: RMSE is 17.760

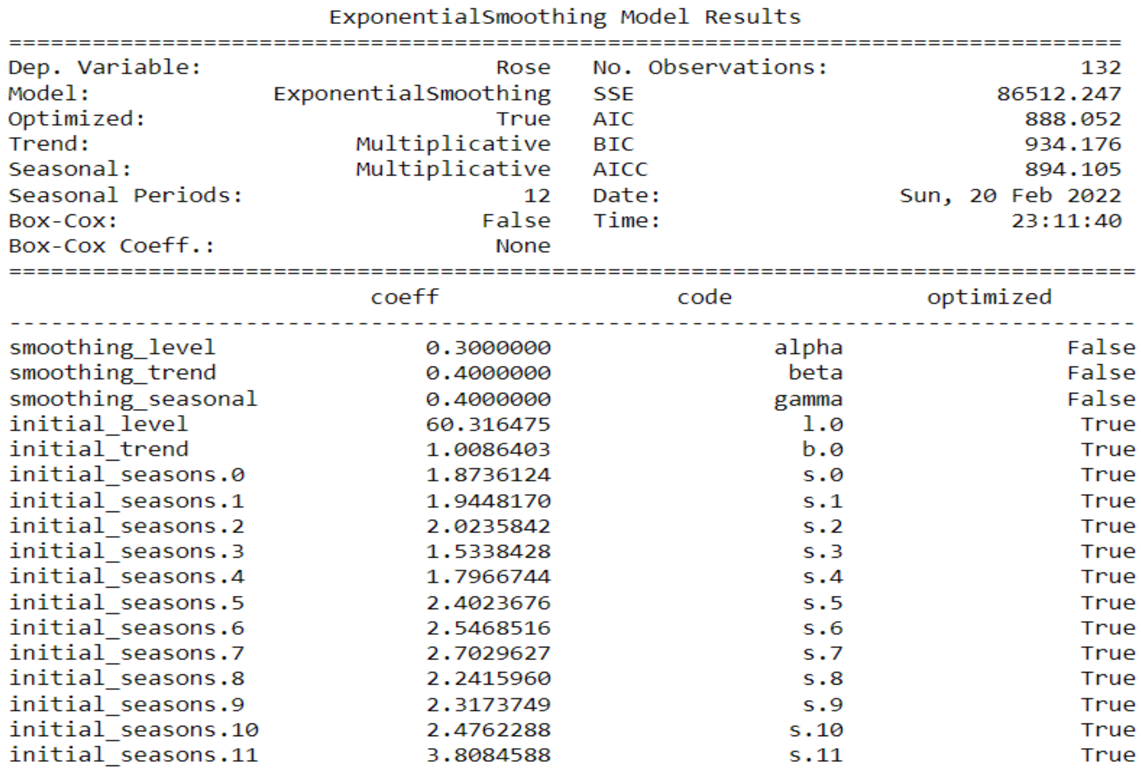
MODEL10: TRIPLE EXPONENTIAL SMOOTHING IN RANGE 0.3 TO 1.1

In Triple Exponential Smoothing Model, we will run a loop with different alpha, beta and gamma values to understand which particular set of value is best.

The results are stored in a data frame as shown below,



	Alpha Values	Beta Values	Gamma Values	Train RMSE	Test RMSE
9	0.3	0.4	0.4	25.600699	1.164579e+01
2	0.3	0.3	0.5	26.182447	1.231361e+01
26	0.3	0.6	0.5	29.096604	1.232594e+01
35	0.3	0.7	0.6	31.324056	1.272823e+01
1	0.3	0.3	0.4	24.504110	1.434853e+01
...	...	...	...	...	...
31	0.3	0.6	1.0	717.674696	1.981314e+24
47	0.3	0.8	1.0	22863.160756	6.913220e+25
55	0.3	0.9	1.0	242721.534347	3.177958e+28
63	0.3	1.0	1.0	945627.989589	4.063676e+30
59	0.3	1.0	0.6	628.743682	1.702541e+38



```
                       ExponentialSmoothing Model Results
================================================================================
Dep. Variable:                     Rose   No. Observations:                  132
Model:             ExponentialSmoothing   SSE                          86512.247
Optimized:                         True   AIC                            888.052
Trend:                   Multiplicative   BIC                            934.176
Seasonal:                Multiplicative   AICC                           894.105
Seasonal Periods:                    12   Date:                 Sun, 20 Feb 2022
Box-Cox:                          False   Time:                         23:11:40
Box-Cox Coeff.:                    None
=================================================================================
                          coeff                 code              optimized
---------------------------------------------------------------------------------
smoothing_level               0.3000000                alpha                False
smoothing_trend               0.4000000                 beta                False
smoothing_seasonal            0.4000000                gamma                False
initial_level                 60.316475                  l.0                 True
initial_trend                 1.0086403                  b.0                 True
initial_seasons.0             1.8736124                  s.0                 True
initial_seasons.1             1.9448170                  s.1                 True
initial_seasons.2             2.0235842                  s.2                 True
initial_seasons.3             1.5338428                  s.3                 True
initial_seasons.4             1.7966744                  s.4                 True
initial_seasons.5             2.4023676                  s.5                 True
initial_seasons.6             2.5468516                  s.6                 True
initial_seasons.7             2.7029627                  s.7                 True
initial_seasons.8             2.2415960                  s.8                 True
initial_seasons.9             2.3173749                  s.9                 True
initial_seasons.10            2.4762288                 s.10                 True
initial_seasons.11            3.8084588                 s.11                 True
```

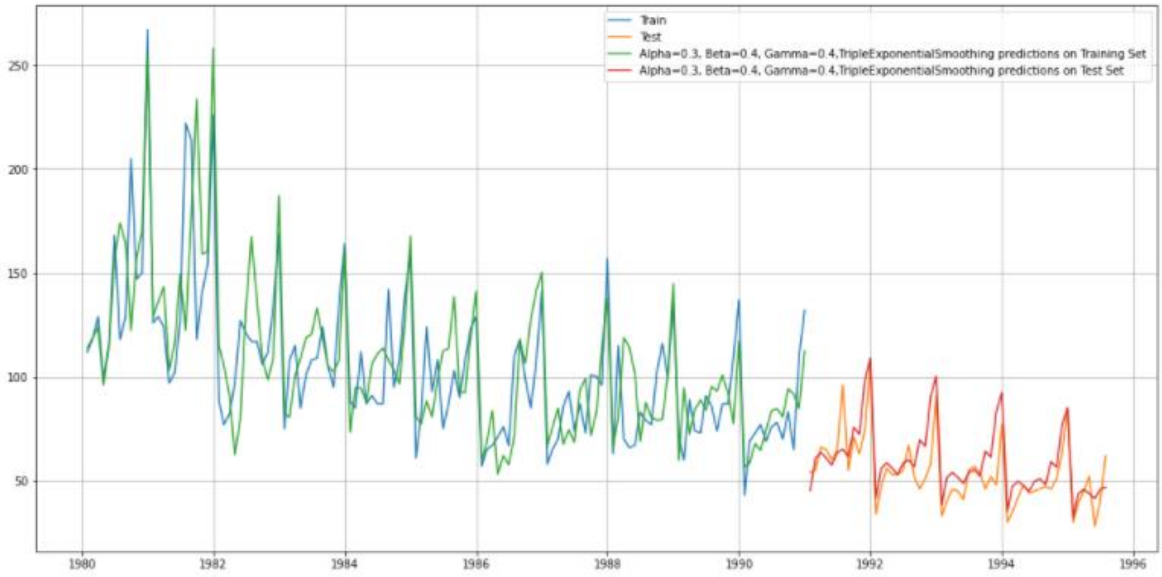
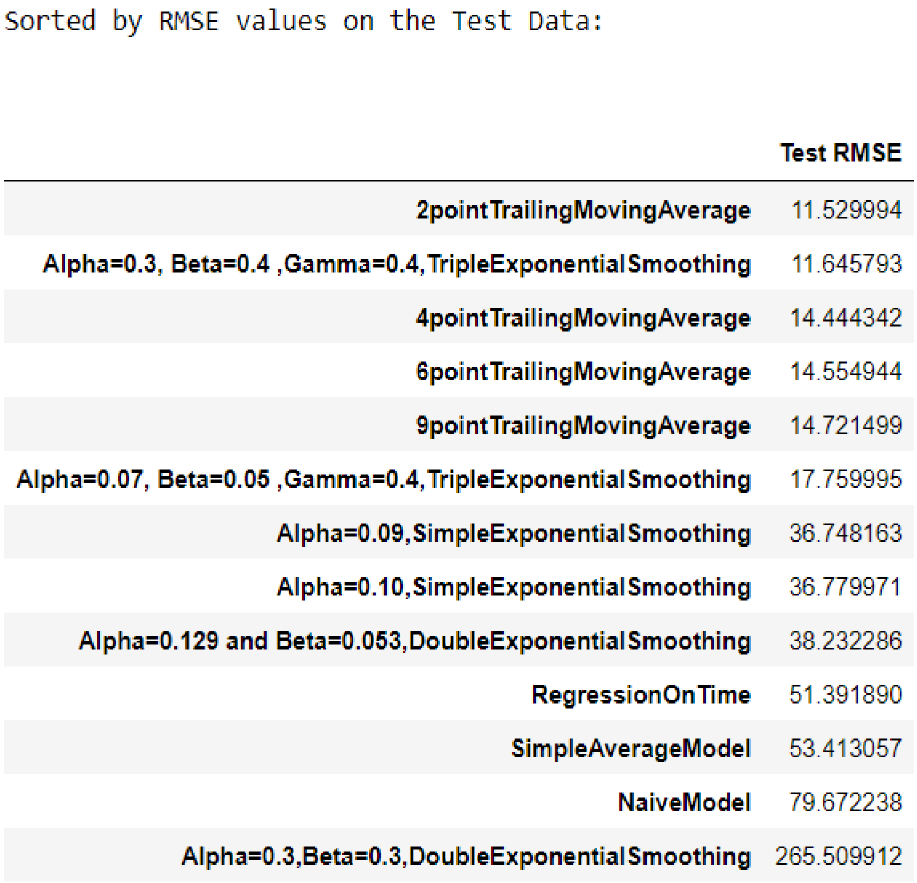


Fig20. TES (Alpha, beta, gamma) Train-Test

The RMSE for the Triple Exponential smoothing model (Alpha=0.4, beta=0.4, gamma=0.3) generated for test data: 11.645793

Results of RMSE values of the models on Test data:



```
Sorted by RMSE values on the Test Data:
```

	Test RMSE
2pointTrailingMovingAverage	11.529994
Alpha=0.3, Beta=0.4 ,Gamma=0.4,TripleExponentialSmoothing	11.645793
4pointTrailingMovingAverage	14.444342
6pointTrailingMovingAverage	14.554944
9pointTrailingMovingAverage	14.721499
Alpha=0.07, Beta=0.05 ,Gamma=0.4,TripleExponentialSmoothing	17.759995
Alpha=0.09,SimpleExponentialSmoothing	36.748163
Alpha=0.10,SimpleExponentialSmoothing	36.779971
Alpha=0.129 and Beta=0.053,DoubleExponentialSmoothing	38.232286
RegressionOnTime	51.391890
SimpleAverageModel	53.413057
NaiveModel	79.672238
Alpha=0.3,Beta=0.3,DoubleExponentialSmoothing	265.509912

Model Comparison Plot:

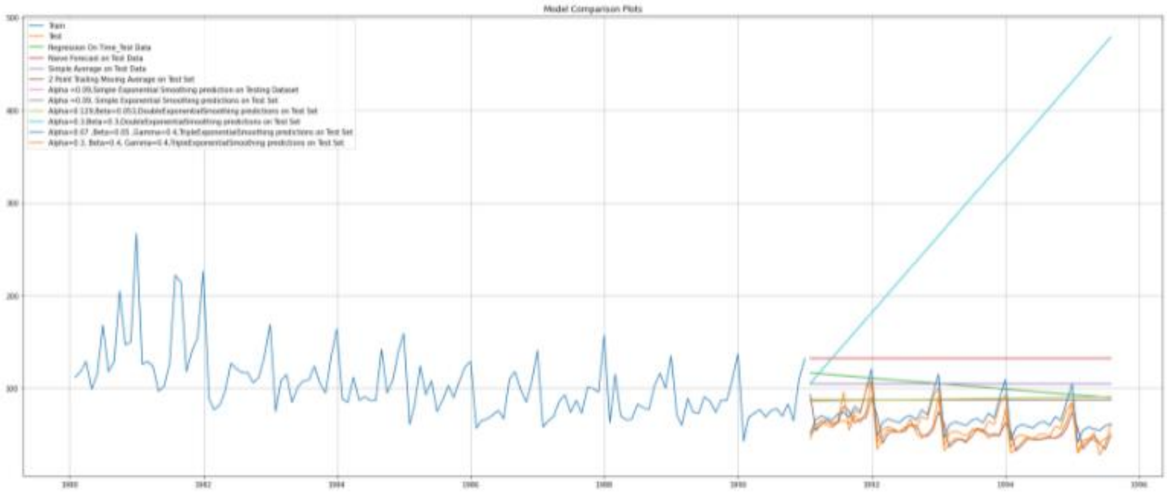
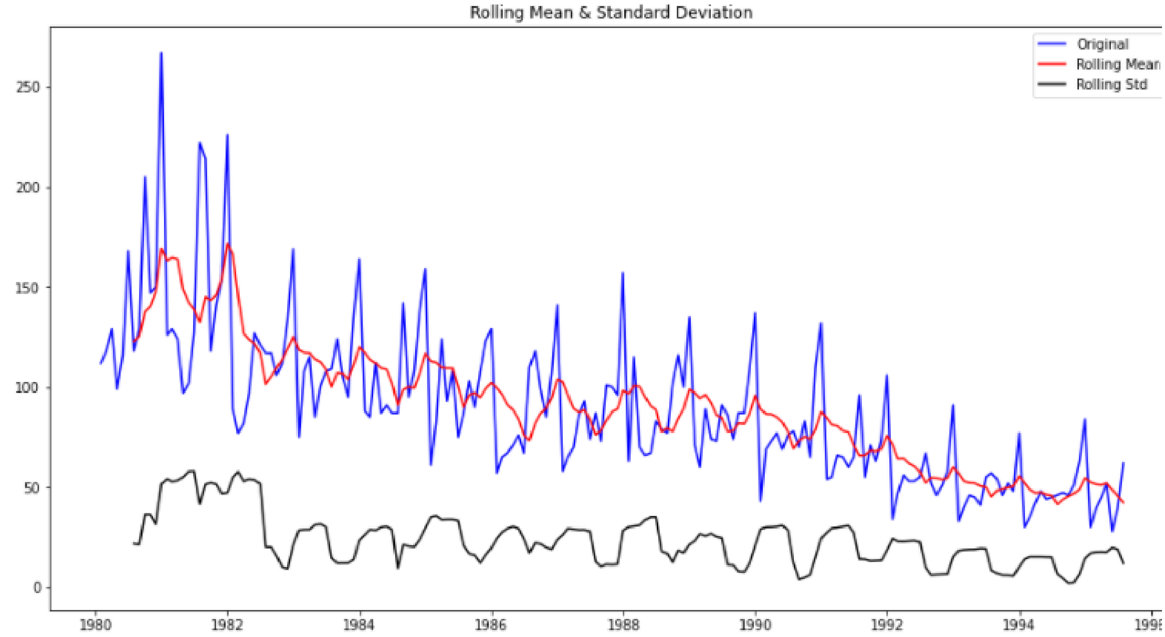
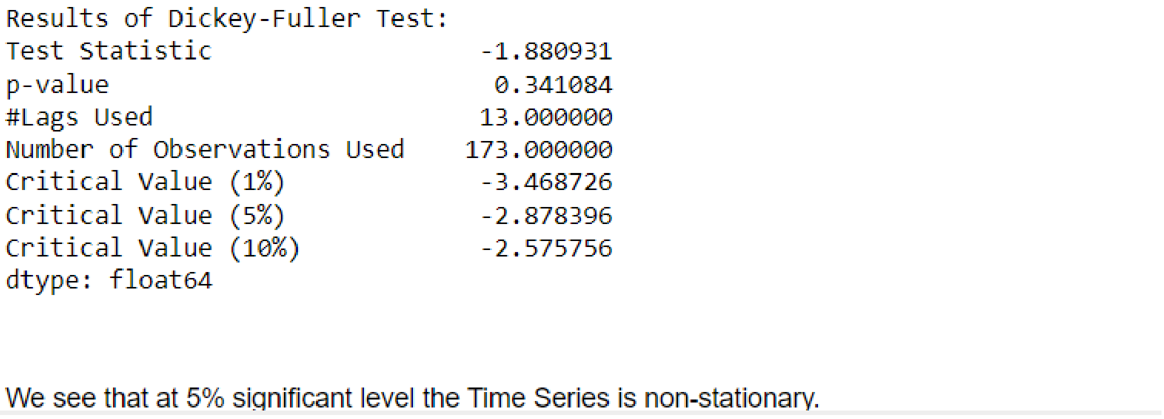


Fig21. Model Comparison Plot

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                  -1.880931
p-value                          0.341084
#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
```

We see that at 5% significant level the Time Series is non-stationary.

The null hypothesis for ADF test (H0) is that the time series is non-stationary.

The alternate hypothesis for ADF test (H1) is that time series is stationary.

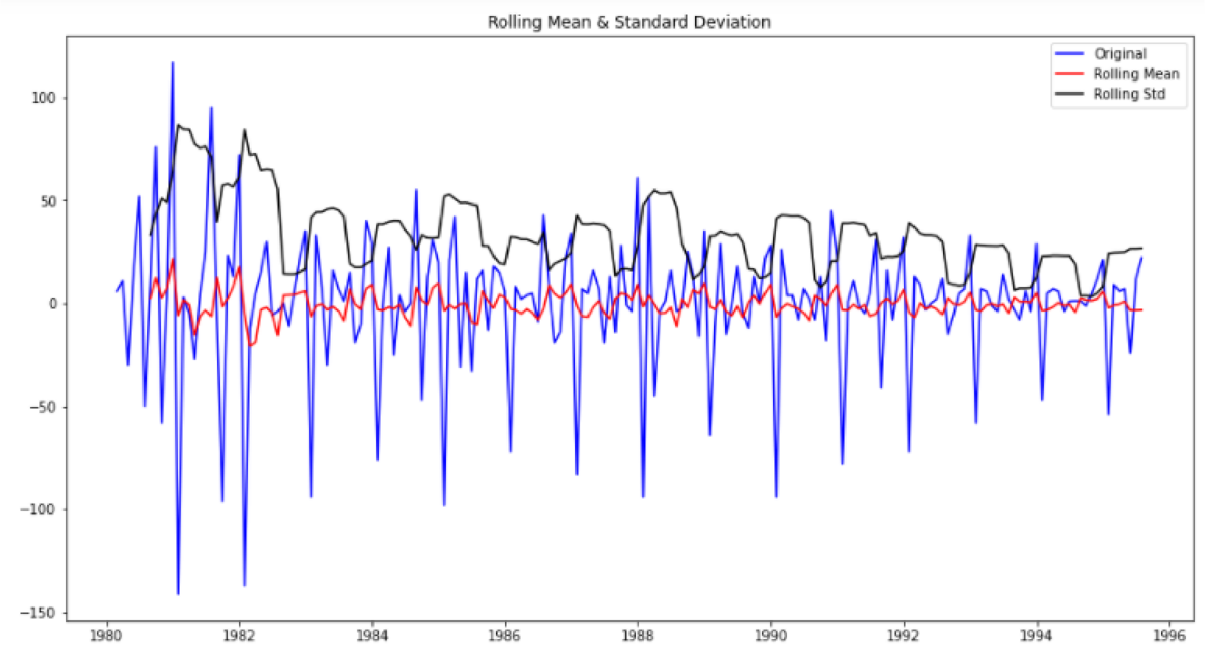
since the p-value of the ADF test is greater than the critical value at 5%, we cannot reject the null hypothesis

Thus, the given time given series is non stationary

To check the stationarity of Sparkling data, we need to check if the alpha value is less than 0.05.

From the above result, we can see that alpha = 0.34 which is higher than 0.05.

Hence, we fail to reject the null hypothesis.



```
Results of Dickey-Fuller Test:
Test Statistic                -8.044820e+00
p-value                        1.806363e-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
```

After taking a difference of order 1, we see that at α = 0.05 the Time Series is indeed stationary.

Therefore, we apply a difference of 1 and check for stationarity.

Now, the result for alpha value is less than 0.05.

Hence the null hypothesis is rejected and the data is stationary.

If the series is non-stationary, stationarize the Time Series by taking a difference of the Time Series. Then we can use this particular differenced series to train the ARIMA models. We do not need to worry about stationarity for the Test Data because we are not building any models on the Test Data, we are evaluating our models over there. You can look at other kinds of transformations as part of making the time series stationary like taking logarithms.

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.

MODEL11: AUTOMATED ARIMA BASED ON AIC CRITERIA

ARIMA stands for Auto Regressive Integrated Moving Average. It is a class of model that captures a suite of different standard temporal structures in time series data.

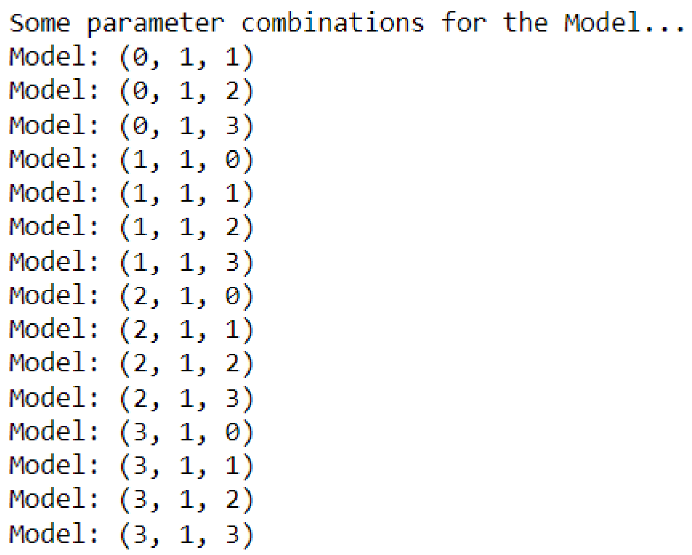
ARIMA description, capturing the key aspects of the model itself. Briefly, they are:

- **AR**: *Autoregression.* A model that uses the dependent relationship between an observation and some number of lagged observations.
- **I**: *Integrated*. The use of differencing of raw observations (e. g. subtracting an observation from an observation at the previous time step) in order to make the time series stationary.
- **MA**: *Moving Average*. A model that uses the dependency between an observation and a residual error from a moving average model applied to lagged observations.

Each of these components are explicitly specified in the model as a parameter. A standard notation is used of ARIMA (p, d, q) where the parameters are substituted with integer values to quickly indicate the specific ARIMA model being used.

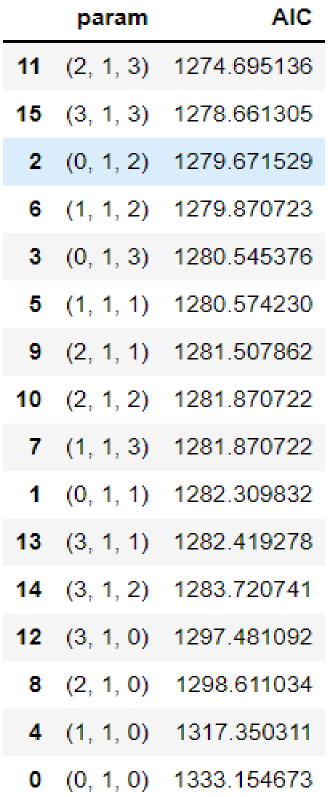
The parameters of the ARIMA model are defined as follows:

- **p**: The number of lag observations included in the model, also called the lag order.
- **d**: The number of times that the raw observations are differenced, also called the degree of differencing.
- **q**: The size of the moving average window, also called the order of moving average.

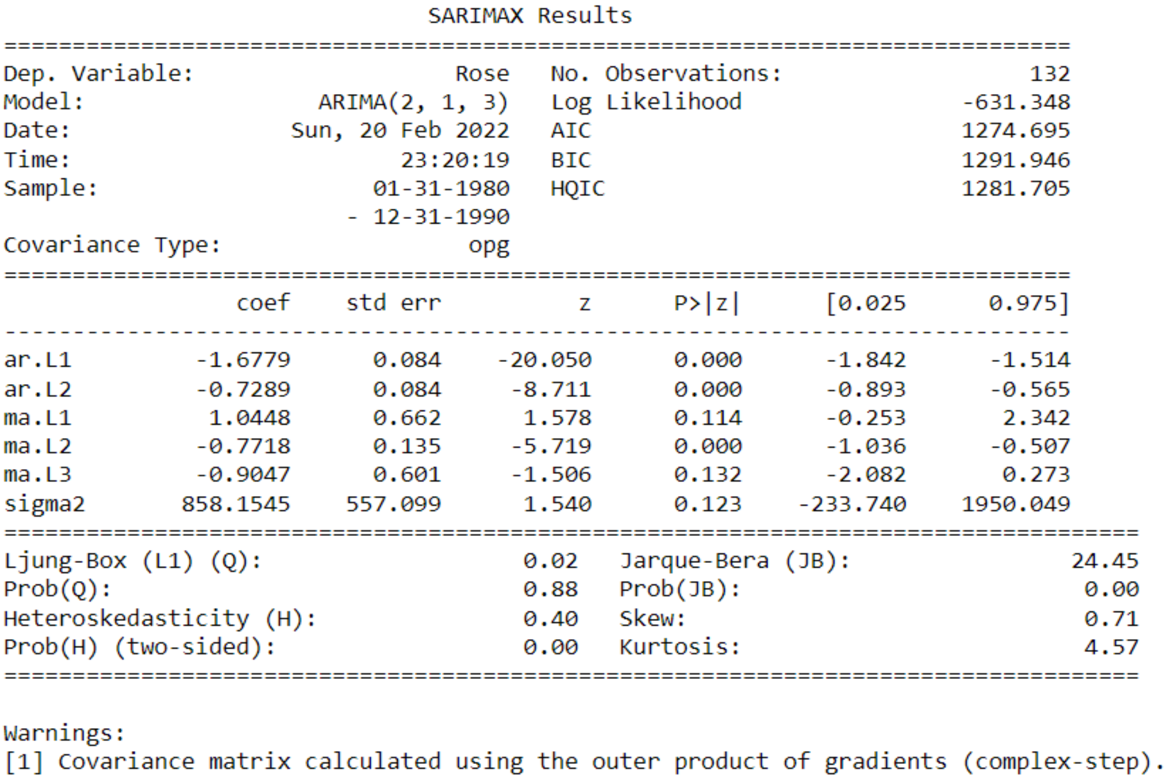


```
Some parameter combinations for the Model...
Model: (0, 1, 1)
Model: (0, 1, 2)
Model: (0, 1, 3)
Model: (1, 1, 0)
Model: (1, 1, 1)
Model: (1, 1, 2)
Model: (1, 1, 3)
Model: (2, 1, 0)
Model: (2, 1, 1)
Model: (2, 1, 2)
Model: (2, 1, 3)
Model: (3, 1, 0)
Model: (3, 1, 1)
Model: (3, 1, 2)
Model: (3, 1, 3)
```

The table showing AIC values arranged in descending order with continuous combinations of p, d and q:

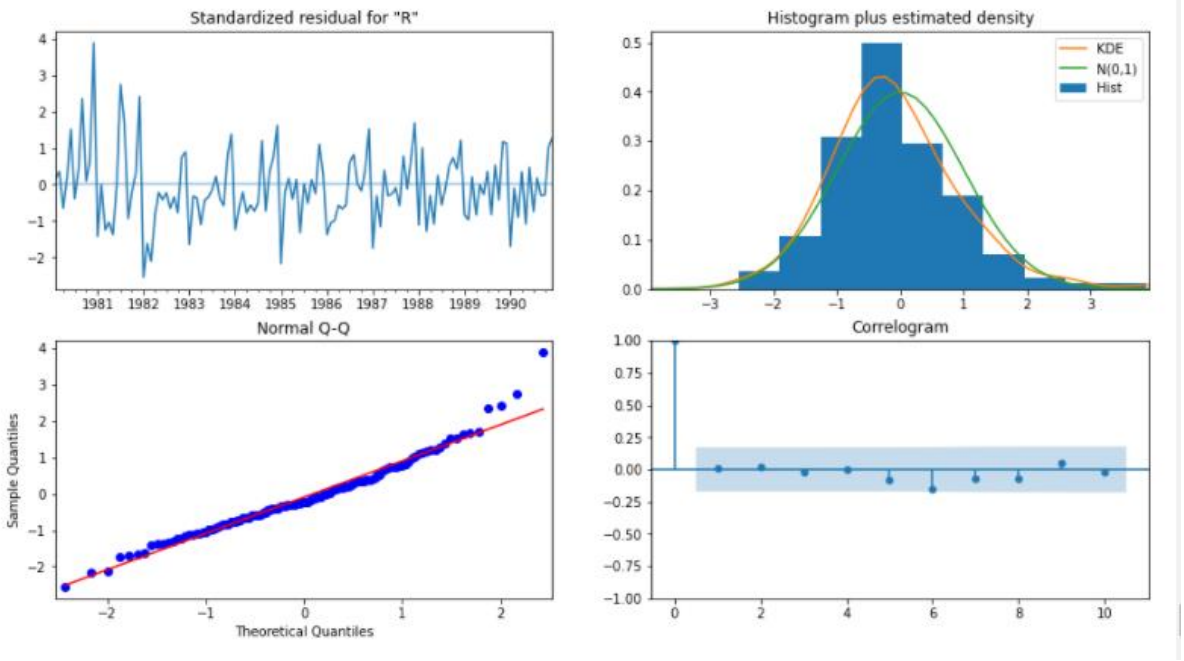


	param	AIC
11	(2, 1, 3)	1274.695136
15	(3, 1, 3)	1278.661305
2	(0, 1, 2)	1279.671529
6	(1, 1, 2)	1279.870723
3	(0, 1, 3)	1280.545376
5	(1, 1, 1)	1280.574230
9	(2, 1, 1)	1281.507862
10	(2, 1, 2)	1281.870722
7	(1, 1, 3)	1281.870722
1	(0, 1, 1)	1282.309832
13	(3, 1, 1)	1282.419278
14	(3, 1, 2)	1283.720741
12	(3, 1, 0)	1297.481092
8	(2, 1, 0)	1298.611034
4	(1, 1, 0)	1317.350311
0	(0, 1, 0)	1333.154673



```
                               SARIMAX Results
==============================================================================
Dep. Variable:                   Rose   No. Observations:                  132
Model:                 ARIMA(2, 1, 3)   Log Likelihood                -631.348
Date:                Sun, 20 Feb 2022   AIC                           1274.695
Time:                        23:20:19   BIC                           1291.946
Sample:                    01-31-1980   HQIC                          1281.705
                         - 12-31-1990
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -1.6779      0.084    -20.050      0.000      -1.842      -1.514
ar.L2         -0.7289      0.084     -8.711      0.000      -0.893      -0.565
ma.L1          1.0448      0.662      1.578      0.114      -0.253       2.342
ma.L2         -0.7718      0.135     -5.719      0.000      -1.036      -0.507
ma.L3         -0.9047      0.601     -1.506      0.132      -2.082       0.273
sigma2       858.1545    557.099      1.540      0.123    -233.740    1950.049
===================================================================================
Ljung-Box (L1) (Q):                   0.02   Jarque-Bera (JB):                24.45
Prob(Q):                              0.88   Prob(JB):                         0.00
Heteroskedasticity (H):               0.40   Skew:                             0.71
Prob(H) (two-sided):                  0.00   Kurtosis:                         4.57
===================================================================================

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

Plot diagnostics:



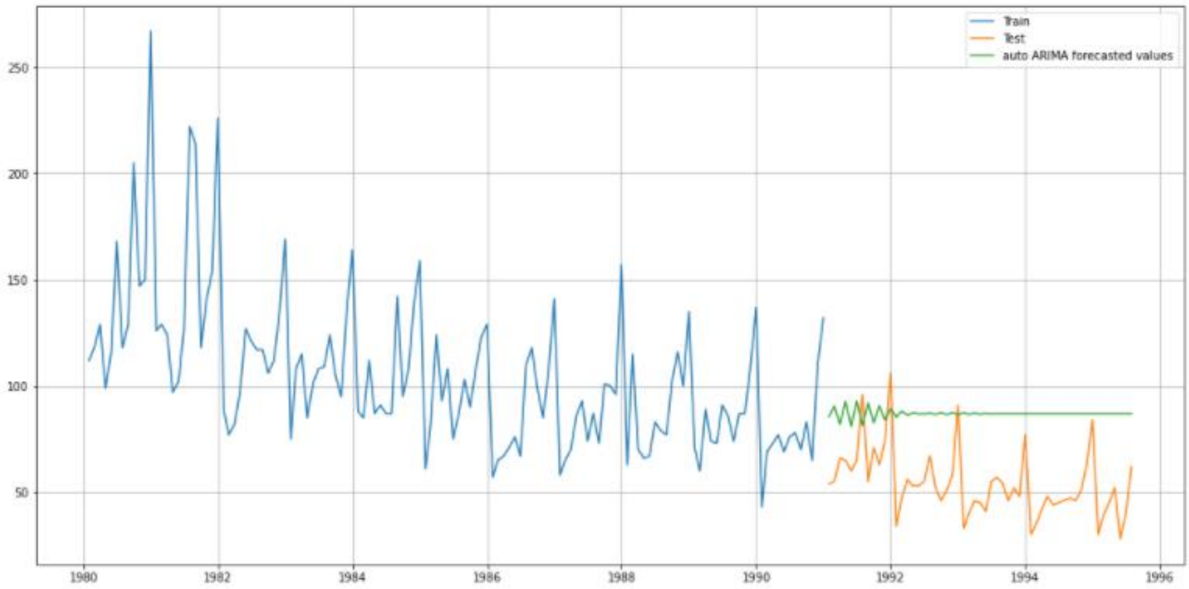


Fig23. Automated Arima

We ran the automated ARIMA model for Rose Sales and sorted AIC values output from lowest to highest.

We then proceeded to build the ARIMA model with the lowest Akaike Information Criteria.

The ARIMA model is built with the best parameters based on the least AIC value in the above table.

RMSE for the autofit ARIMA model: **36.76789623595707**

MODEL12: AUTOMATED SARIMA MODEL WITH SEASONALITY 6 & 12:

A seasonal autoregressive integrated moving average (SARIMA) model is one step different from an ARIMA model based on the concept of seasonal trends. In many time series data, frequent seasonal effects come into play. Take for example the average temperature measured in a location with four seasons. There will be a seasonal effect on a yearly basis, and the temperature in this particular season will definitely have a strong correlation with the temperature measured last year in the same season.

It adds three new hyperparameters to specify the autoregression (AR), differencing (I) and moving average (MA) for the seasonal component of the series, as well as an additional parameter for the period of the seasonality.

Configuring a SARIMA requires selecting hyperparameters for both the trend and seasonal elements of the series.

Trend Elements

There are three trend elements that require configuration.

They are the same as the ARIMA model; specifically:

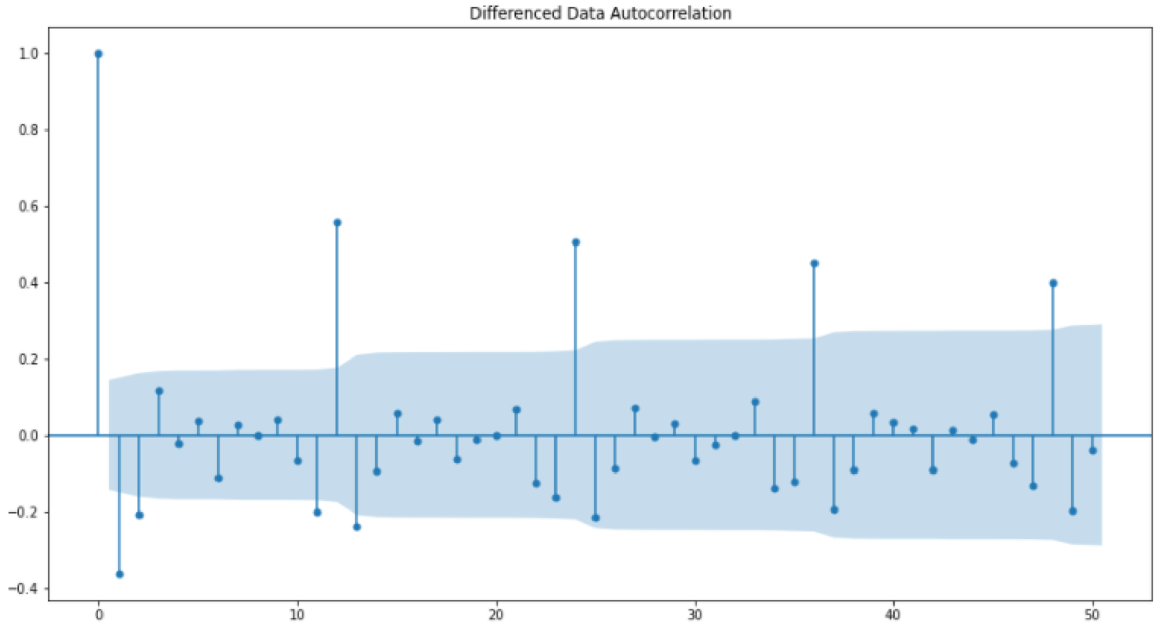
- **p**: Trend autoregression order.
- **d**: Trend difference order.
- **q**: Trend moving average order.

Seasonal Elements

There are four seasonal elements that are not part of ARIMA that must be configured; they are:

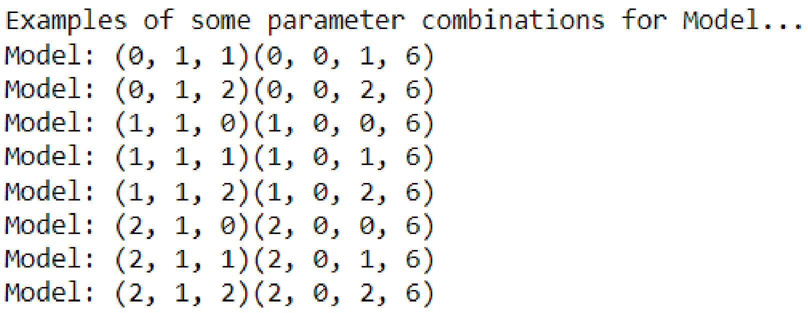
- **P**: Seasonal autoregressive order.
- **D**: Seasonal difference order.
- **Q**: Seasonal moving average order.
- **m**: The number of time steps for a single seasonal period.

Let us look at the ACF plot once more to understand the seasonal parameter for the SARIMA model.

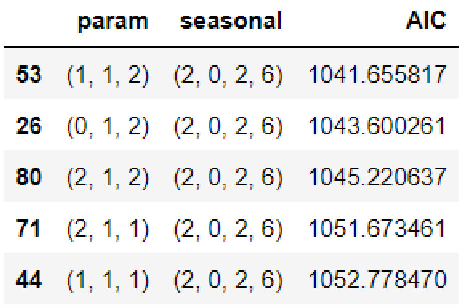


We see that there can be a seasonality of 6 as well as 12. We will run our auto SARIMA models by setting seasonality both as 6 and 12.

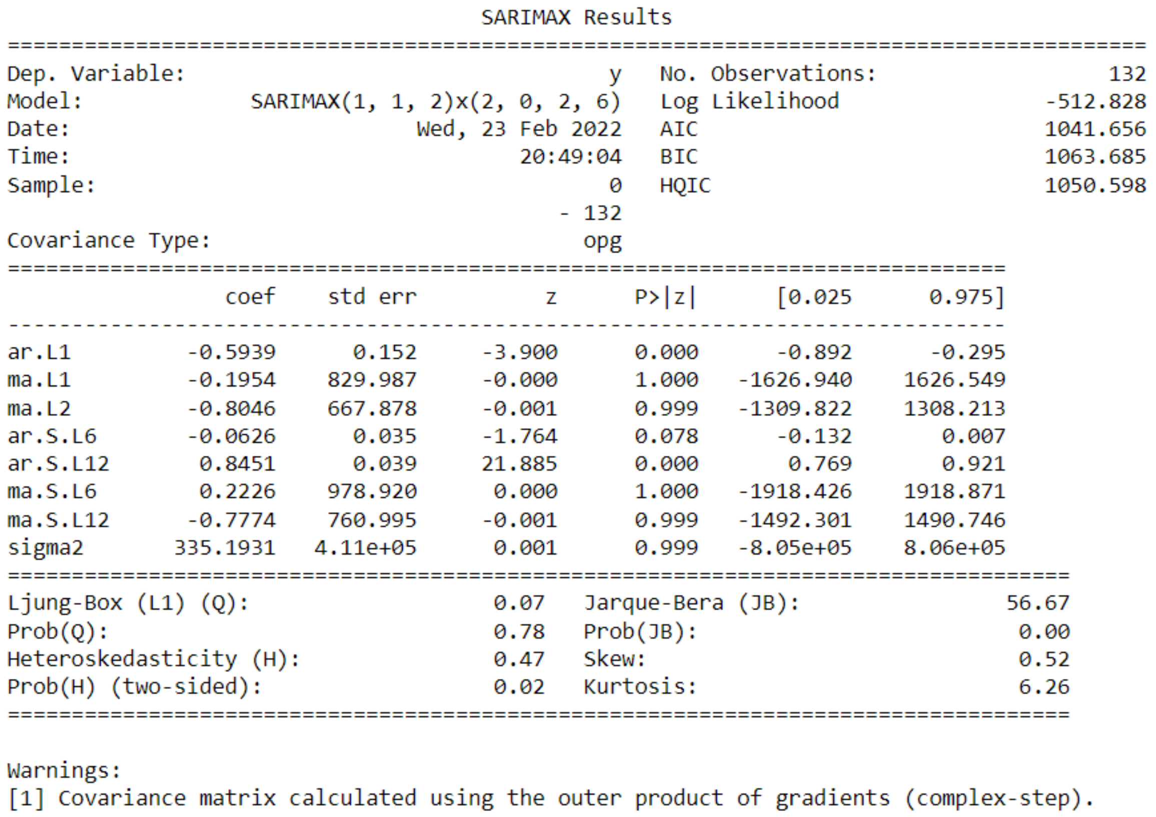
Setting the seasonality as 6 for the first iteration of the auto SARIMA model



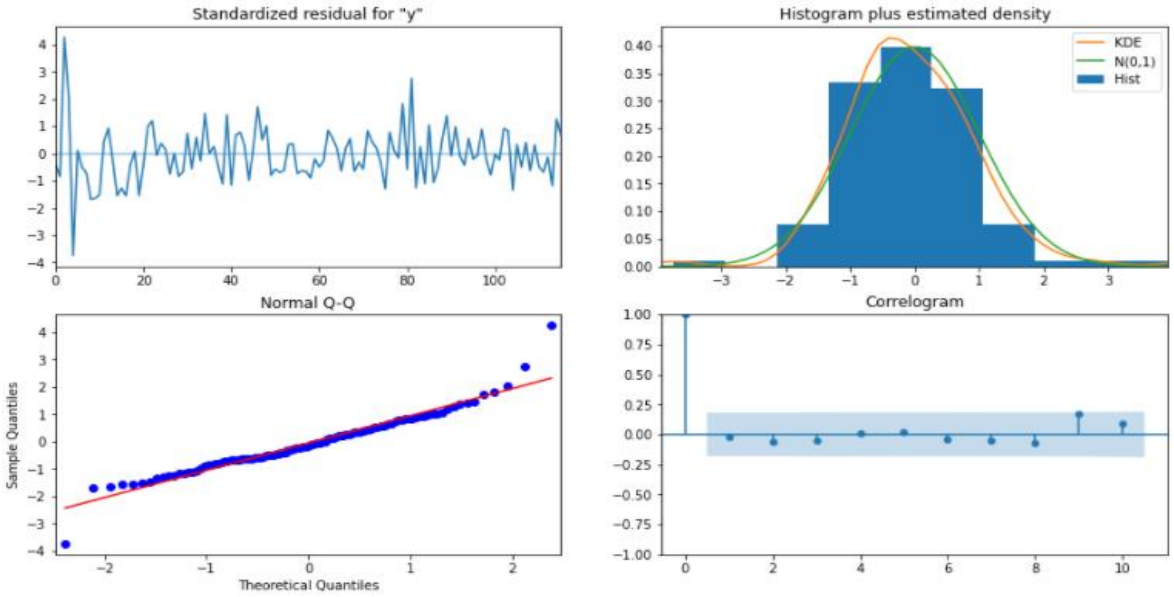
```
Examples of some parameter combinations for Model...
Model: (0, 1, 1)(0, 0, 1, 6)
Model: (0, 1, 2)(0, 0, 2, 6)
Model: (1, 1, 0)(1, 0, 0, 6)
Model: (1, 1, 1)(1, 0, 1, 6)
Model: (1, 1, 2)(1, 0, 2, 6)
Model: (2, 1, 0)(2, 0, 0, 6)
Model: (2, 1, 1)(2, 0, 1, 6)
Model: (2, 1, 2)(2, 0, 2, 6)
```

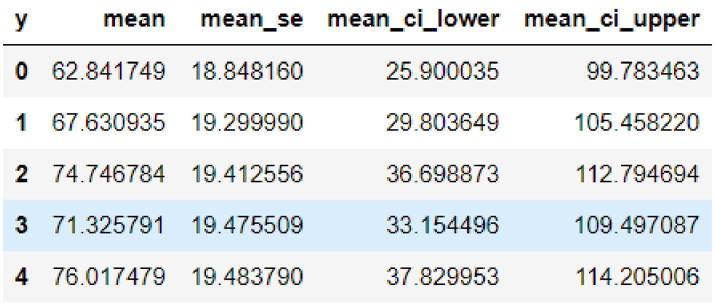


	param	seasonal	AIC
53	(1, 1, 2)	(2, 0, 2, 6)	1041.655817
26	(0, 1, 2)	(2, 0, 2, 6)	1043.600261
80	(2, 1, 2)	(2, 0, 2, 6)	1045.220637
71	(2, 1, 1)	(2, 0, 2, 6)	1051.673461
44	(1, 1, 1)	(2, 0, 2, 6)	1052.778470



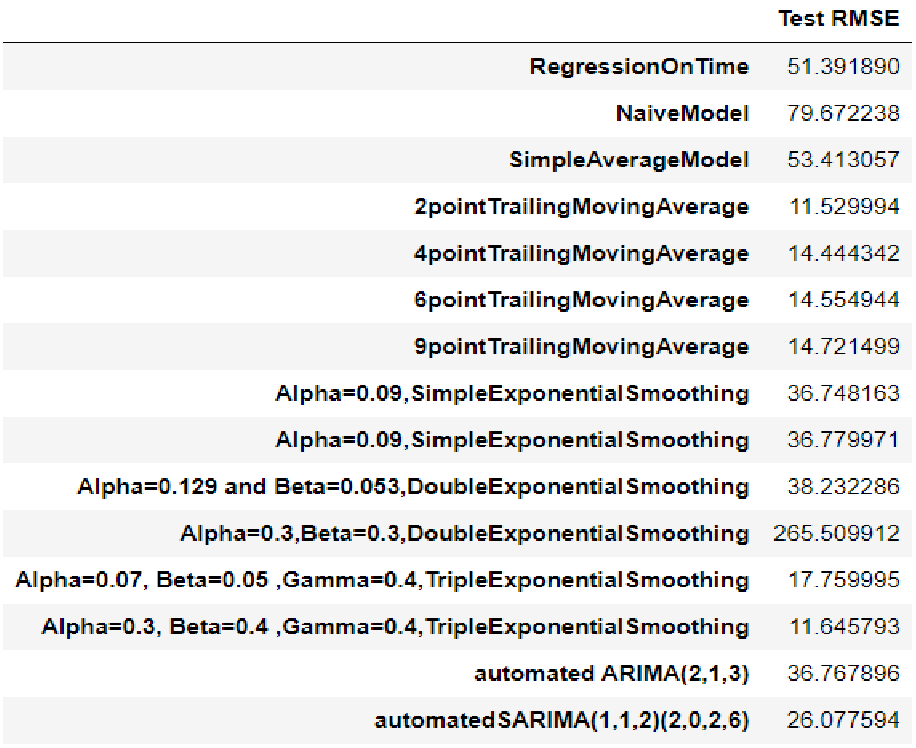
```
                                     SARIMAX Results                                     
=========================================================================================
Dep. Variable:                                  y   No. Observations:                 132
Model:             SARIMAX(1, 1, 2)x(2, 0, 2, 6)   Log Likelihood                -512.828
Date:                            Wed, 23 Feb 2022   AIC                           1041.656
Time:                                    20:49:04   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      0.152     -3.900      0.000      -0.892      -0.295
ma.L1         -0.1954    829.987     -0.000      1.000   -1626.940    1626.549
ma.L2         -0.8046    667.878     -0.001      0.999   -1309.822    1308.213
ar.S.L6       -0.0626      0.035     -1.764      0.078      -0.132       0.007
ar.S.L12       0.8451      0.039     21.885      0.000       0.769       0.921
ma.S.L6        0.2226    978.920      0.000      1.000   -1918.426    1918.871
ma.S.L12      -0.7774    760.995     -0.001      0.999   -1492.301    1490.746
sigma2       335.1931   4.11e+05      0.001      0.999   -8.05e+05    8.06e+05
===================================================================================
Ljung-Box (L1) (Q):                   0.07   Jarque-Bera (JB):                56.67
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).
```



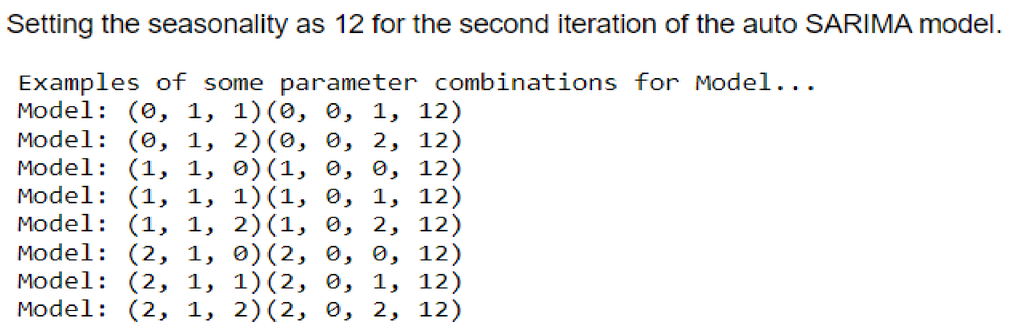


y	mean	mean_se	mean_ci_lower	mean_ci_upper
0	62.841749	18.848160	25.900035	99.783463
1	67.630935	19.299990	29.803649	105.458220
2	74.746784	19.412556	36.698873	112.794694
3	71.325791	19.475509	33.154496	109.497087
4	76.017479	19.483790	37.829953	114.205006

RMSE for the autofit ARIMA model: 26.077593566351386



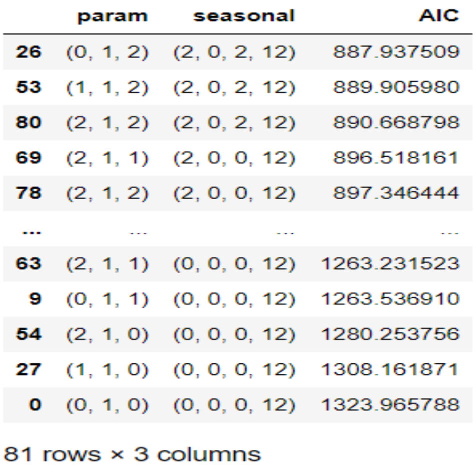
	Test RMSE
RegressionOnTime	51.391890
NaiveModel	79.672238
SimpleAverageModel	53.413057
2pointTrailingMovingAverage	11.529994
4pointTrailingMovingAverage	14.444342
6pointTrailingMovingAverage	14.554944
9pointTrailingMovingAverage	14.721499
Alpha=0.09,SimpleExponentialSmoothing	36.748163
Alpha=0.09,SimpleExponentialSmoothing	36.779971
Alpha=0.129 and Beta=0.053,DoubleExponentialSmoothing	38.232286
Alpha=0.3,Beta=0.3,DoubleExponentialSmoothing	265.509912
Alpha=0.07, Beta=0.05 ,Gamma=0.4,TripleExponentialSmoothing	17.759995
Alpha=0.3, Beta=0.4 ,Gamma=0.4,TripleExponentialSmoothing	11.645793
automated ARIMA(2,1,3)	36.767896
automatedSARIMA(1,1,2)(2,0,2,6)	26.077594



Setting the seasonality as 12 for the second iteration of the auto SARIMA model.

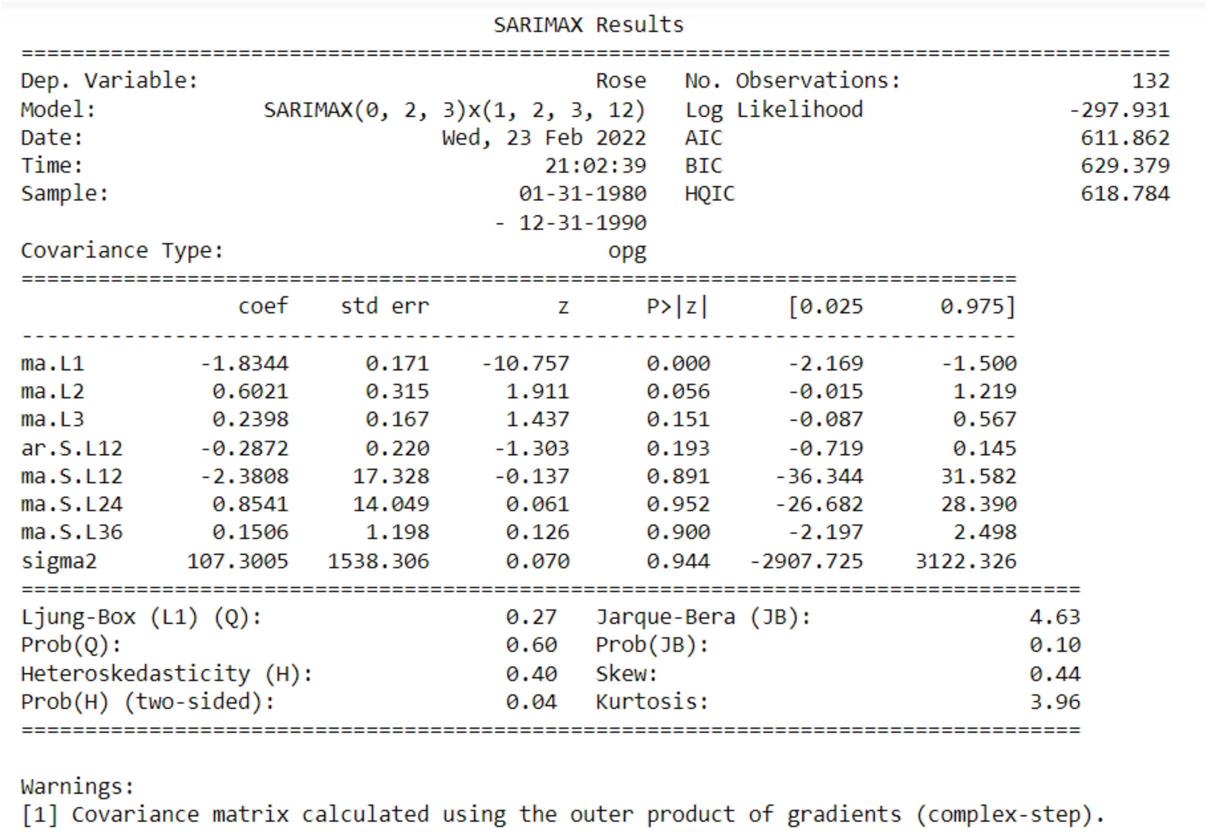
```
Examples of some parameter combinations for Model...
Model: (0, 1, 1)(0, 0, 1, 12)
Model: (0, 1, 2)(0, 0, 2, 12)
Model: (1, 1, 0)(1, 0, 0, 12)
Model: (1, 1, 1)(1, 0, 1, 12)
Model: (1, 1, 2)(1, 0, 2, 12)
Model: (2, 1, 0)(2, 0, 0, 12)
Model: (2, 1, 1)(2, 0, 1, 12)
Model: (2, 1, 2)(2, 0, 2, 12)
```

Sorted the AIC values output from lowest to highest, we then proceed to build the SARIMA MODEL with the lowest Akaike Information Criteria.

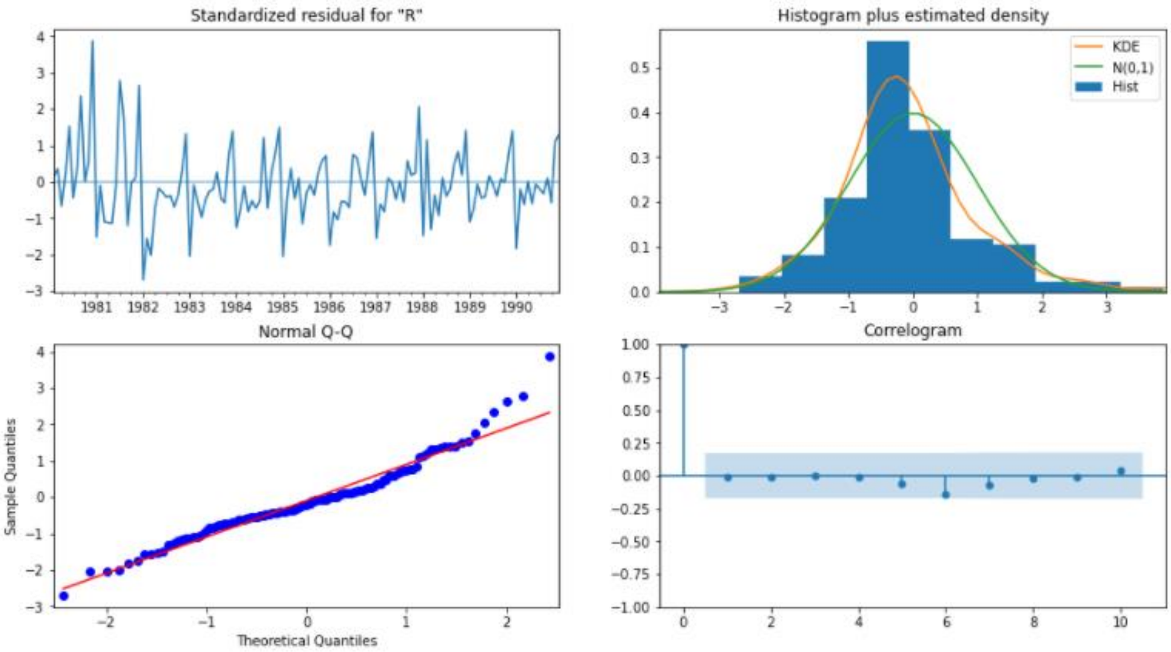


	param	seasonal	AIC
26	(0, 1, 2)	(2, 0, 2, 12)	887.937509
53	(1, 1, 2)	(2, 0, 2, 12)	889.905980
80	(2, 1, 2)	(2, 0, 2, 12)	890.668798
69	(2, 1, 1)	(2, 0, 0, 12)	896.518161
78	(2, 1, 2)	(2, 0, 0, 12)	897.346444
...	...	...	...
63	(2, 1, 1)	(0, 0, 0, 12)	1263.231523
9	(0, 1, 1)	(0, 0, 0, 12)	1263.536910
54	(2, 1, 0)	(0, 0, 0, 12)	1280.253756
27	(1, 1, 0)	(0, 0, 0, 12)	1308.161871
0	(0, 1, 0)	(0, 0, 0, 12)	1323.965788

81 rows × 3 columns



```
                                     SARIMAX Results
==========================================================================================
Dep. Variable:                                Rose   No. Observations:                 132
Model:             SARIMAX(0, 2, 3)x(1, 2, 3, 12)   Log Likelihood                -297.931
Date:                             Wed, 23 Feb 2022   AIC                            611.862
Time:                                     21:02:39   BIC                            629.379
Sample:                                 01-31-1980   HQIC                           618.784
                                      - 12-31-1990
Covariance Type:                               opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ma.L1         -1.8344      0.171    -10.757      0.000      -2.169      -1.500
ma.L2          0.6021      0.315      1.911      0.056      -0.015       1.219
ma.L3          0.2398      0.167      1.437      0.151      -0.087       0.567
ar.S.L12      -0.2872      0.220     -1.303      0.193      -0.719       0.145
ma.S.L12      -2.3808     17.328     -0.137      0.891     -36.344      31.582
ma.S.L24       0.8541     14.049      0.061      0.952     -26.682      28.390
ma.S.L36       0.1506      1.198      0.126      0.900      -2.197       2.498
sigma2       107.3005   1538.306      0.070      0.944   -2907.725    3122.326
===================================================================================
Ljung-Box (L1) (Q):                   0.27   Jarque-Bera (JB):                 4.63
Prob(Q):                              0.60   Prob(JB):                         0.10
Heteroskedasticity (H):               0.40   Skew:                             0.44
Prob(H) (two-sided):                  0.04   Kurtosis:                         3.96
===================================================================================

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



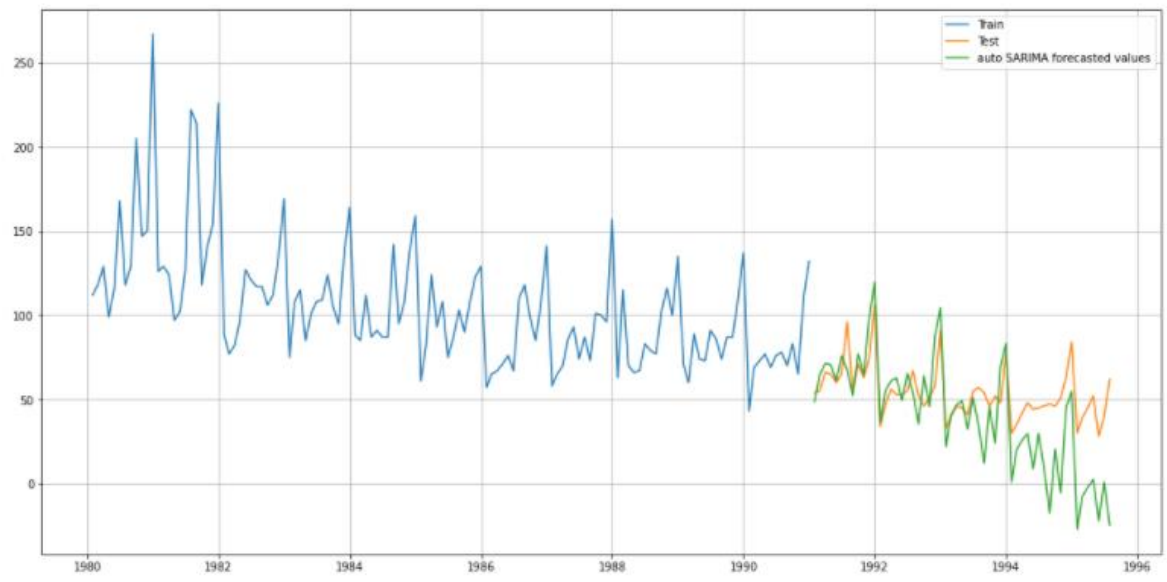


Fig24. Automated Sarima

RMSE for the autofit SARIMA model: 28.051354

Inference on Model diagnostics confirms that the model residuals are normally distributed.

Standardized residual: Do not display any obvious seasonality

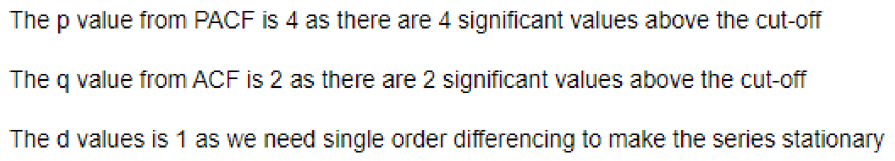
Histogram plus estimated density: The KDE plot of the residuals is similar with the normal distribution. Hence the model residuals are normally distributed based.

Normal Q-Q plot: There is an ordered distribution of residuals (blue dots) following the linear trend of samples taken from a standard normal distribution.

Correlogram: The time series residuals have low correlation with lagged versions itself.

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.

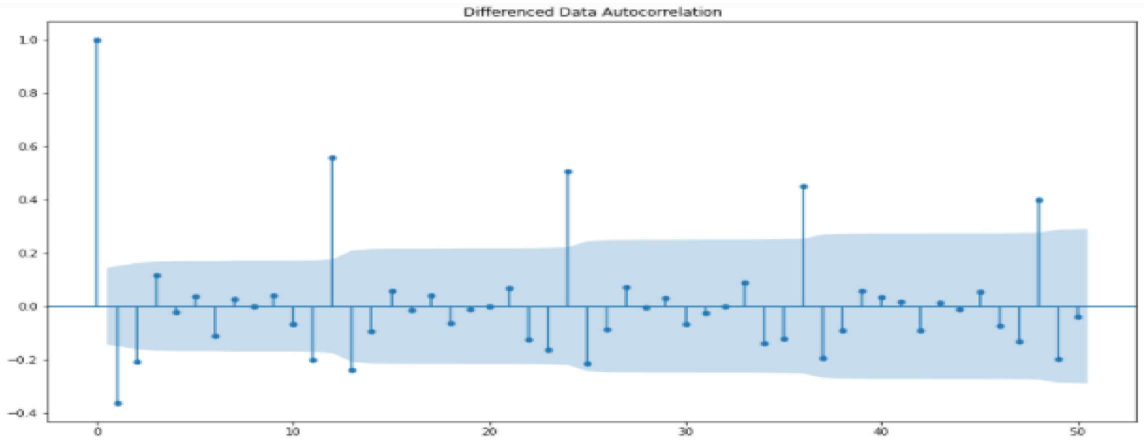
<u>MODEL13: MANUAL ARIMA WITH CUT-OFF VALUES FROM ACF AND PACF</u>

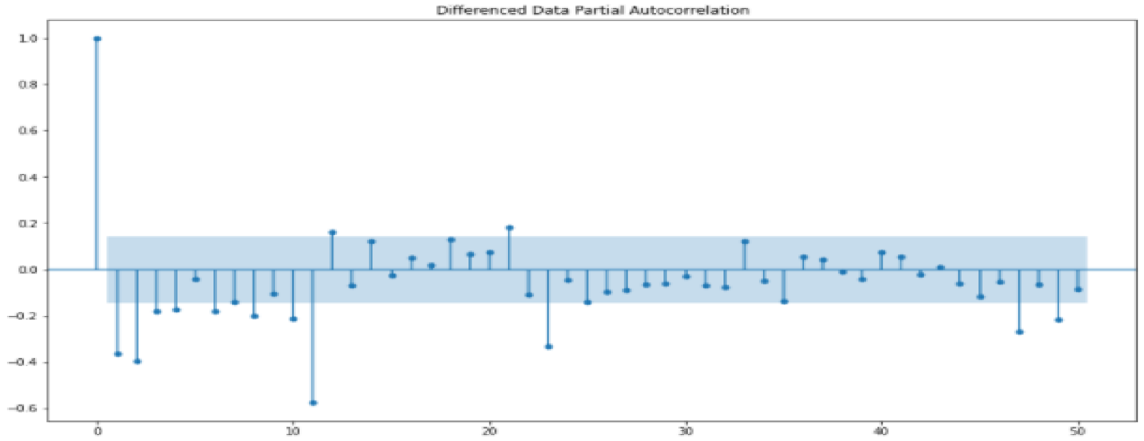


The p value from PACF is 4 as there are 4 significant values above the cut-off

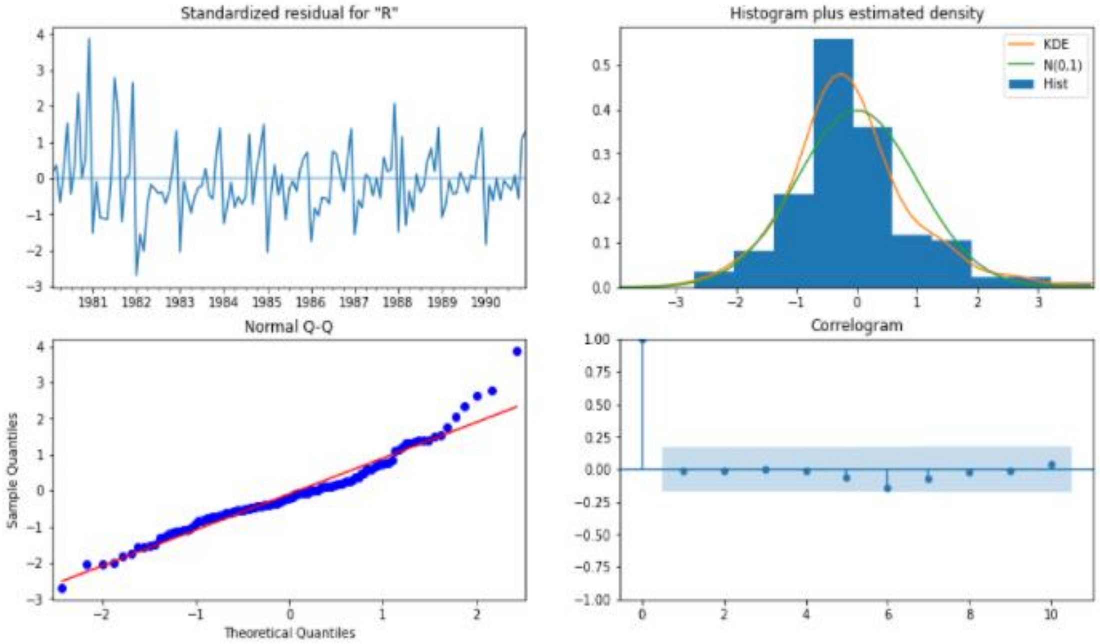
The q value from ACF is 2 as there are 2 significant values above the cut-off

The d values is 1 as we need single order differencing to make the series stationary





```
                               SARIMAX Results
==============================================================================
Dep. Variable:                   Rose   No. Observations:                  132
Model:                 ARIMA(4, 1, 2)   Log Likelihood                -635.859
Date:                Mon, 21 Feb 2022   AIC                           1285.718
Time:                        11:35:45   BIC                           1305.845
Sample:                    01-31-1980   HQIC                          1293.896
                         - 12-31-1990
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.3838      0.923     -0.416      0.677      -2.192       1.425
ar.L2          0.0046      0.258      0.018      0.986      -0.502       0.511
ar.L3          0.0414      0.113      0.366      0.714      -0.180       0.263
ar.L4         -0.0054      0.177     -0.031      0.976      -0.353       0.342
ma.L1         -0.3239      0.933     -0.347      0.729      -2.153       1.505
ma.L2         -0.5407      0.874     -0.619      0.536      -2.254       1.172
sigma2       951.1524     93.870     10.133      0.000     767.170    1135.135
===============================================================================
====
Ljung-Box (L1) (Q):                   0.02   Jarque-Bera (JB):                 3
2.85
Prob(Q):                              0.88   Prob(JB):
0.00
Heteroskedasticity (H):               0.37   Skew:
0.77
Prob(H) (two-sided):                  0.00   Kurtosis:
4.91
===============================================================================
====

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



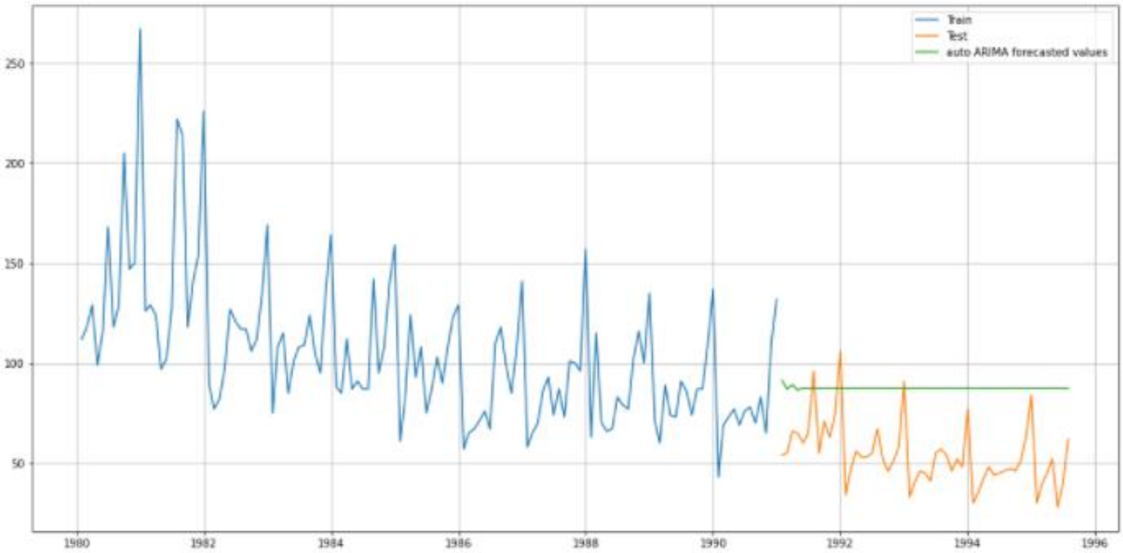


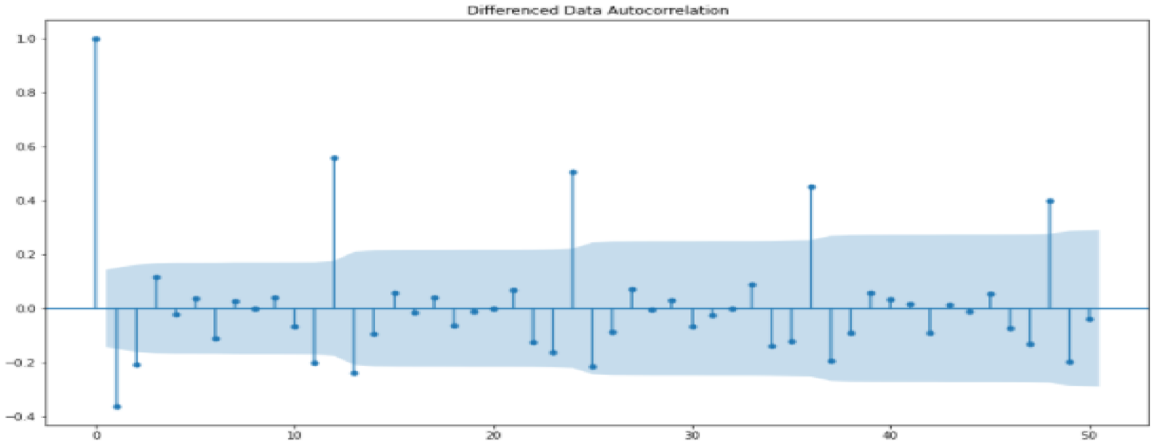
Fig25. Manual Arima

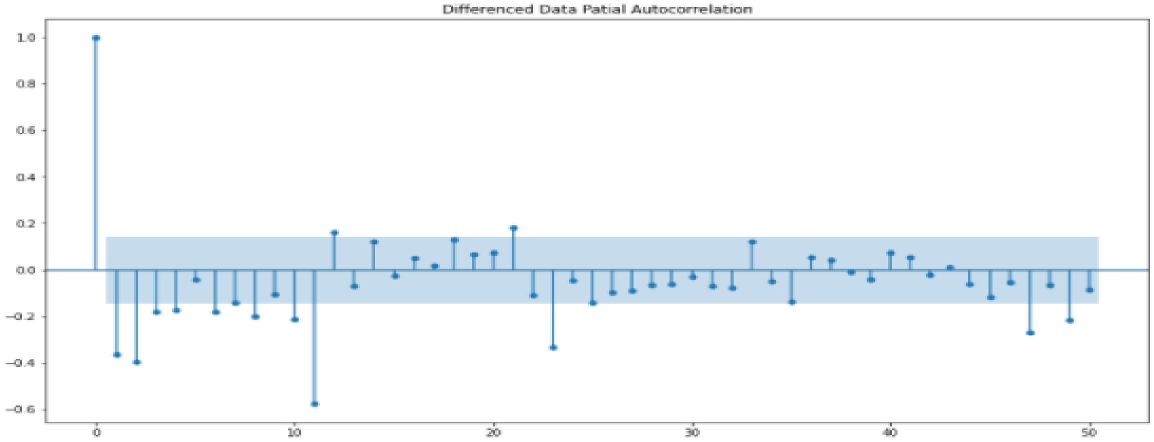
RMSE for the manual ARIMA model: 36.98966342788272

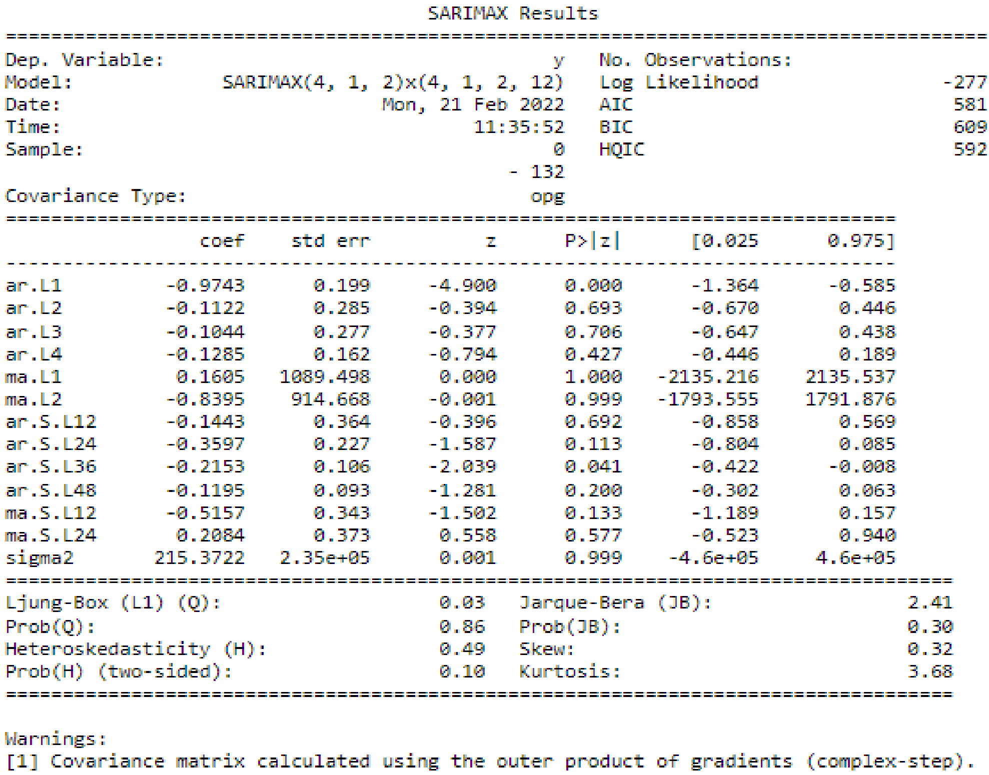
Manual ARIMA is built based on ACF plot and PACF plot.

Hence, we choose AR parameter value as p and moving average parameter value to be q.

MODEL14: MANUAL SARIMA


Differenced Data Autocorrelation


Differenced Data Patial Autocorrelation



```
                                     SARIMAX Results
==========================================================================================
Dep. Variable:                                  y   No. Observations:                  132
Model:             SARIMAX(4, 1, 2)x(4, 1, 2, 12)   Log Likelihood                -277.661
Date:                            Mon, 21 Feb 2022   AIC                            581.322
Time:                                    11:35:52   BIC                            609.983
Sample:                                         0   HQIC                           592.663
                                            - 132
Covariance Type:                              opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.9743      0.199     -4.900      0.000      -1.364      -0.585
ar.L2         -0.1122      0.285     -0.394      0.693      -0.670       0.446
ar.L3         -0.1044      0.277     -0.377      0.706      -0.647       0.438
ar.L4         -0.1285      0.162     -0.794      0.427      -0.446       0.189
ma.L1          0.1605   1089.498      0.000      1.000   -2135.216    2135.537
ma.L2         -0.8395    914.668     -0.001      0.999   -1793.555    1791.876
ar.S.L12      -0.1443      0.364     -0.396      0.692      -0.858       0.569
ar.S.L24      -0.3597      0.227     -1.587      0.113      -0.804       0.085
ar.S.L36      -0.2153      0.106     -2.039      0.041      -0.422      -0.008
ar.S.L48      -0.1195      0.093     -1.281      0.200      -0.302       0.063
ma.S.L12      -0.5157      0.343     -1.502      0.133      -1.189       0.157
ma.S.L24       0.2084      0.373      0.558      0.577      -0.523       0.940
sigma2       215.3722   2.35e+05      0.001      0.999    -4.6e+05     4.6e+05
===================================================================================
Ljung-Box (L1) (Q):                   0.03   Jarque-Bera (JB):                 2.41
Prob(Q):                              0.86   Prob(JB):                         0.30
Heteroskedasticity (H):               0.49   Skew:                             0.32
Prob(H) (two-sided):                  0.10   Kurtosis:                         3.68
===================================================================================

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

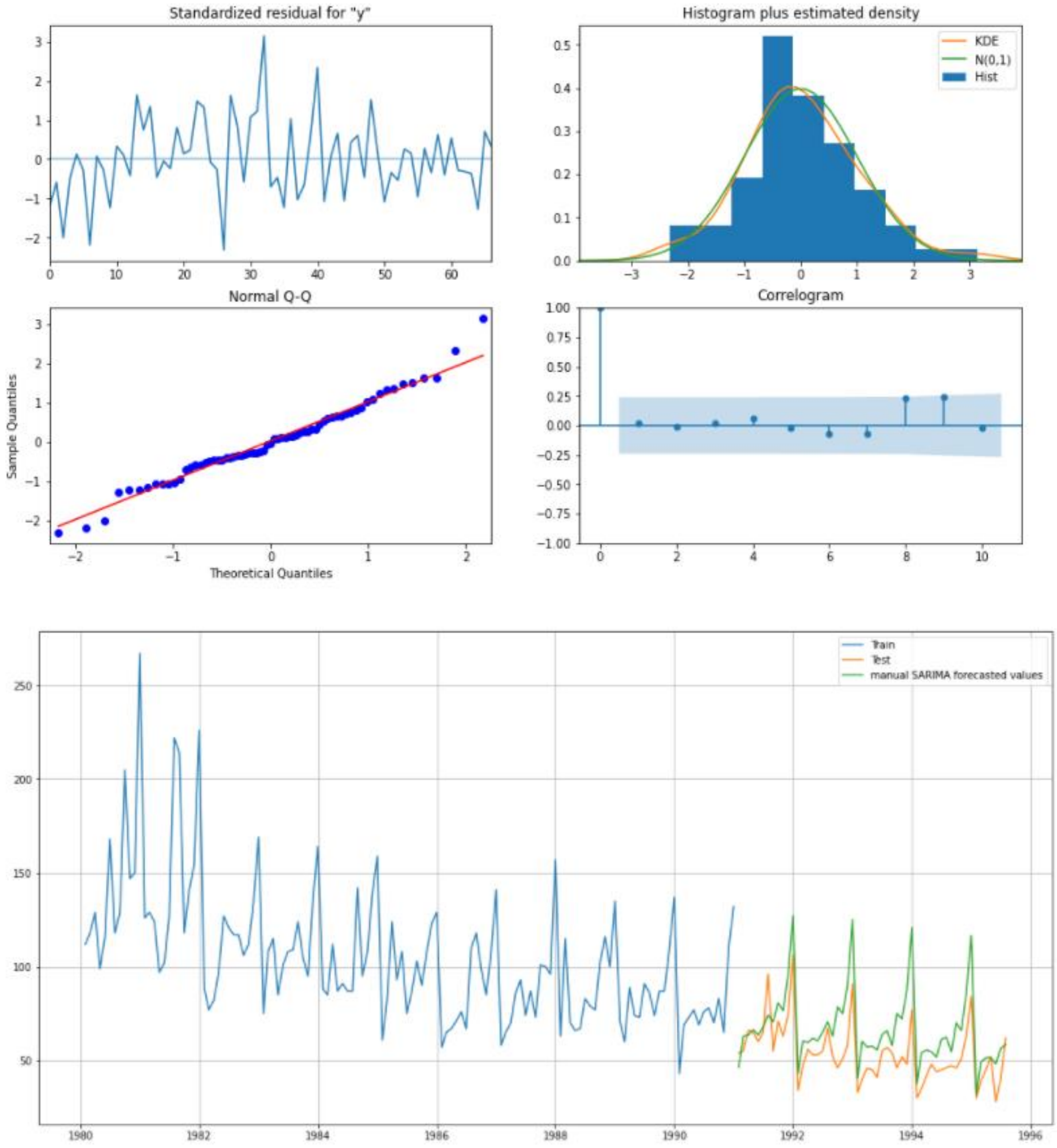


Fig26. Manual Sarima

RMSE for the manual SARIMA model: 17.501825369836837

Manual SARIMA is built based on ACF plot and PACF plot.

Hence, we choose AR parameter value as p. moving average parameter value to be q and d(difference) value to be 1.

We then derive the seasonal parameters based on the seasonal cut-off.

Inference on Model diagnostics confirms that the model residuals are normally distributed.

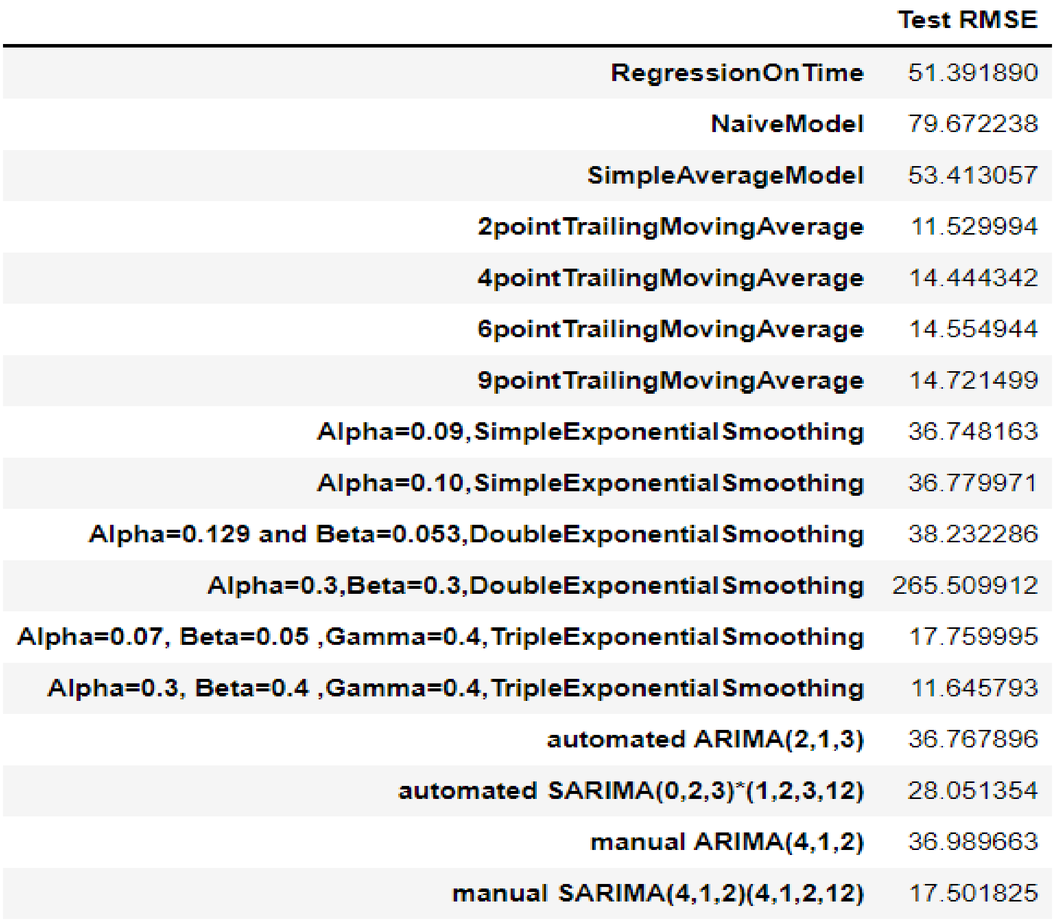
Standardized residual: Do not display any obvious seasonality

Histogram plus estimated density: The KDE plot of the residuals is similar with the normal distribution. Hence the model residuals are normally distributed based.

Normal Q-Q plot: There is an ordered distribution of residuals (blue dots) following the linear trend of samples taken from a standard normal distribution.

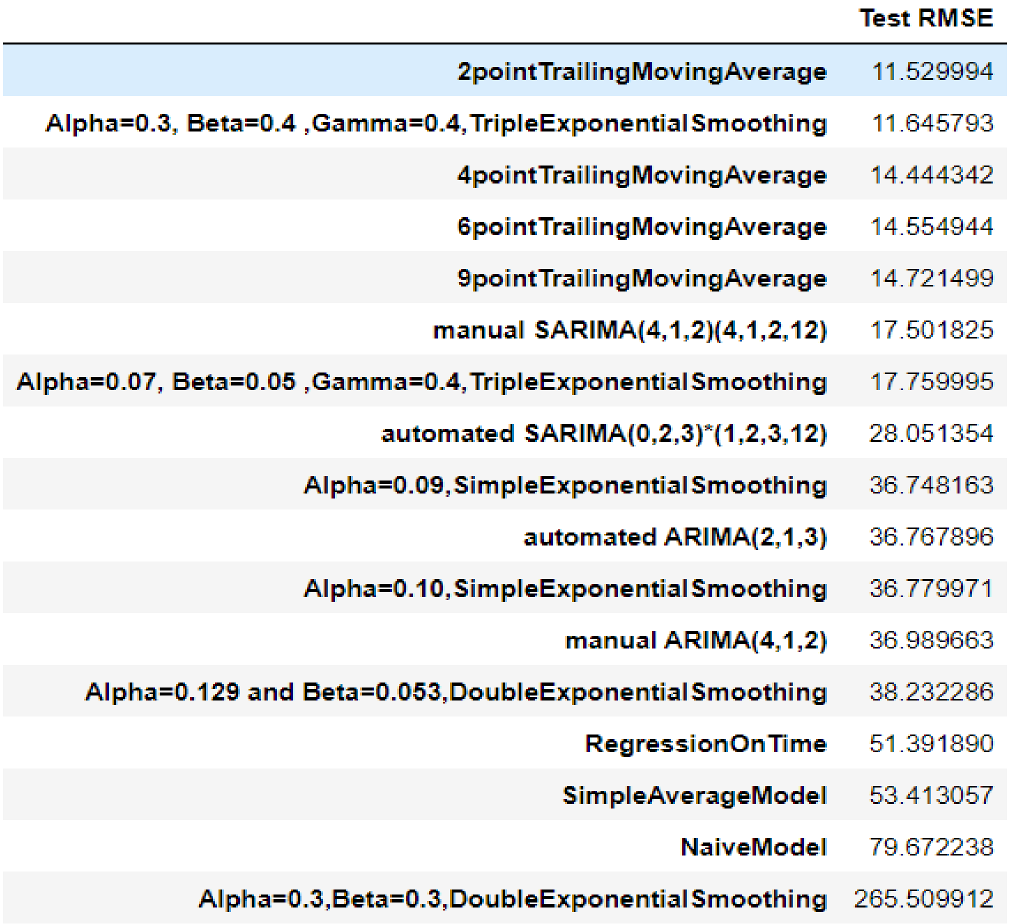
Correlogram: The time series residuals have low correlation with lagged versions itself.

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.



	Test RMSE
RegressionOnTime	51.391890
NaiveModel	79.672238
SimpleAverageModel	53.413057
2pointTrailingMovingAverage	11.529994
4pointTrailingMovingAverage	14.444342
6pointTrailingMovingAverage	14.554944
9pointTrailingMovingAverage	14.721499
Alpha=0.09,SimpleExponentialSmoothing	36.748163
Alpha=0.10,SimpleExponentialSmoothing	36.779971
Alpha=0.129 and Beta=0.053,DoubleExponentialSmoothing	38.232286
Alpha=0.3,Beta=0.3,DoubleExponentialSmoothing	265.509912
Alpha=0.07, Beta=0.05 ,Gamma=0.4,TripleExponentialSmoothing	17.759995
Alpha=0.3, Beta=0.4 ,Gamma=0.4,TripleExponentialSmoothing	11.645793
automated ARIMA(2,1,3)	36.767896
automated SARIMA(0,2,3)*(1,2,3,12)	28.051354
manual ARIMA(4,1,2)	36.989663
manual SARIMA(4,1,2)(4,1,2,12)	17.501825

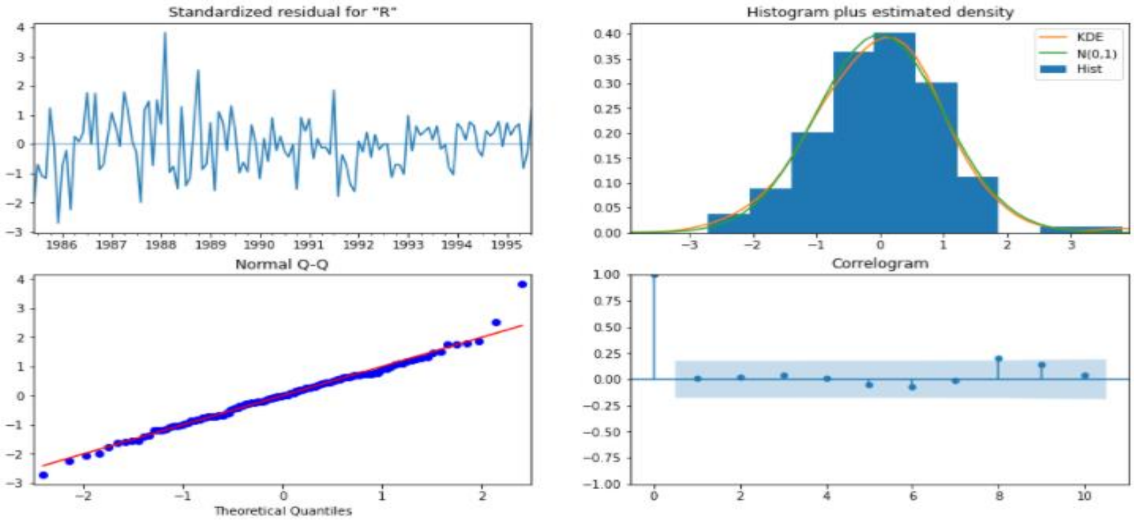
Results Sorted by RMSE:

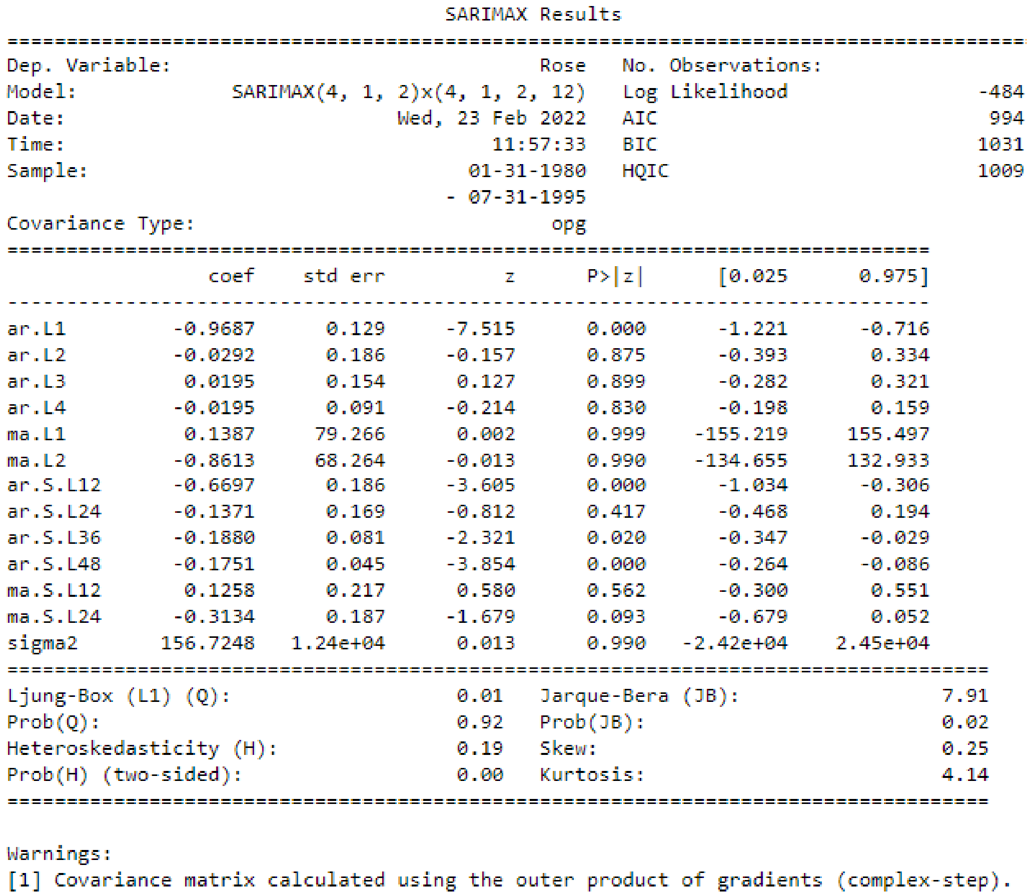


	Test RMSE
2pointTrailingMovingAverage	11.529994
Alpha=0.3, Beta=0.4 ,Gamma=0.4,TripleExponentialSmoothing	11.645793
4pointTrailingMovingAverage	14.444342
6pointTrailingMovingAverage	14.554944
9pointTrailingMovingAverage	14.721499
manual SARIMA(4,1,2)(4,1,2,12)	17.501825
Alpha=0.07, Beta=0.05 ,Gamma=0.4,TripleExponentialSmoothing	17.759995
automated SARIMA(0,2,3)*(1,2,3,12)	28.051354
Alpha=0.09,SimpleExponentialSmoothing	36.748163
automated ARIMA(2,1,3)	36.767896
Alpha=0.10,SimpleExponentialSmoothing	36.779971
manual ARIMA(4,1,2)	36.989663
Alpha=0.129 and Beta=0.053,DoubleExponentialSmoothing	38.232286
RegressionOnTime	51.391890
SimpleAverageModel	53.413057
NaiveModel	79.672238
Alpha=0.3,Beta=0.3,DoubleExponentialSmoothing	265.509912

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.

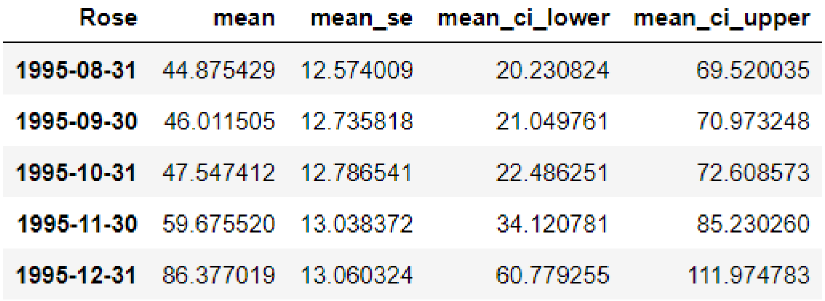
Optimum Model on Complete Dataset:



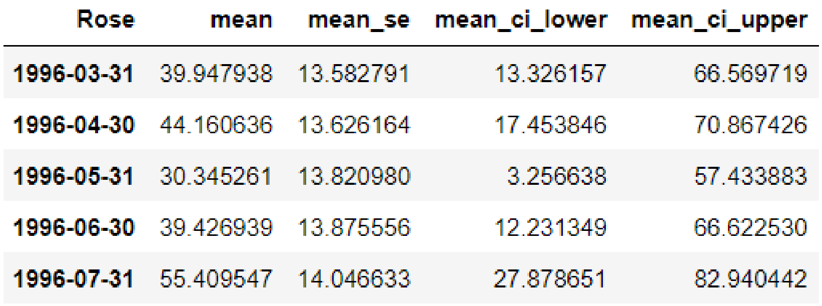


```
                                      SARIMAX Results                                       
==========================================================================================
Dep. Variable:                                  Rose   No. Observations:                  187
Model:             SARIMAX(4, 1, 2)x(4, 1, 2, 12)   Log Likelihood                -484.421
Date:                            Wed, 23 Feb 2022   AIC                            994.841
Time:                                    11:57:33   BIC                           1031.294
Sample:                                01-31-1980   HQIC                          1009.647
                                     - 07-31-1995                                         
Covariance Type:                              opg                                         
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.9687      0.129     -7.515      0.000      -1.221      -0.716
ar.L2         -0.0292      0.186     -0.157      0.875      -0.393       0.334
ar.L3          0.0195      0.154      0.127      0.899      -0.282       0.321
ar.L4         -0.0195      0.091     -0.214      0.830      -0.198       0.159
ma.L1          0.1387     79.266      0.002      0.999    -155.219     155.497
ma.L2         -0.8613     68.264     -0.013      0.990    -134.655     132.933
ar.S.L12      -0.6697      0.186     -3.605      0.000      -1.034      -0.306
ar.S.L24      -0.1371      0.169     -0.812      0.417      -0.468       0.194
ar.S.L36      -0.1880      0.081     -2.321      0.020      -0.347      -0.029
ar.S.L48      -0.1751      0.045     -3.854      0.000      -0.264      -0.086
ma.S.L12       0.1258      0.217      0.580      0.562      -0.300       0.551
ma.S.L24      -0.3134      0.187     -1.679      0.093      -0.679       0.052
sigma2       156.7248   1.24e+04      0.013      0.990   -2.42e+04    2.45e+04
===================================================================================
Ljung-Box (L1) (Q):                   0.01   Jarque-Bera (JB):                 7.91
Prob(Q):                              0.92   Prob(JB):                         0.02
Heteroskedasticity (H):               0.19   Skew:                             0.25
Prob(H) (two-sided):                  0.00   Kurtosis:                         4.14
===================================================================================

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

Forecasting 12 months into the future with the complete model

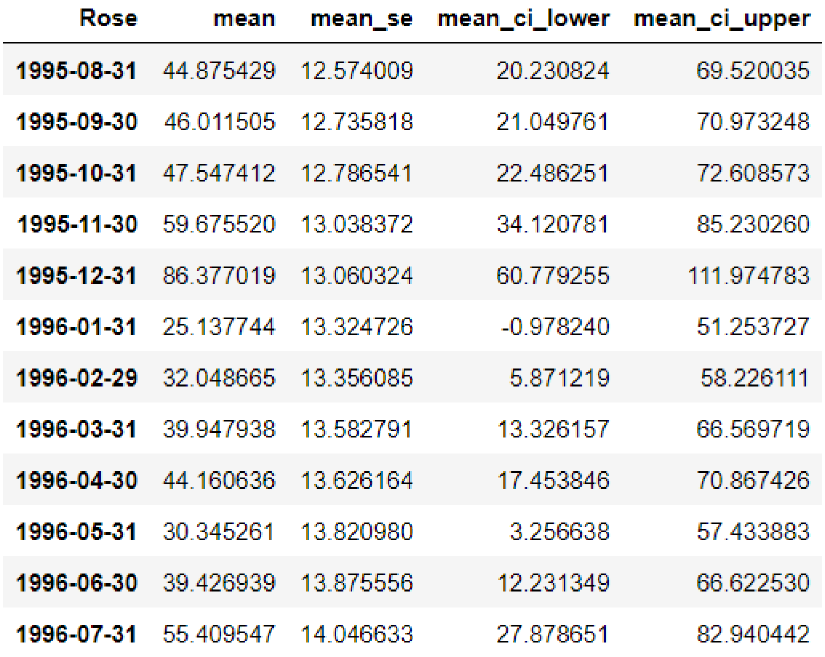


Rose	mean	mean_se	mean_ci_lower	mean_ci_upper
1995-08-31	44.875429	12.574009	20.230824	69.520035
1995-09-30	46.011505	12.735818	21.049761	70.973248
1995-10-31	47.547412	12.786541	22.486251	72.608573
1995-11-30	59.675520	13.038372	34.120781	85.230260
1995-12-31	86.377019	13.060324	60.779255	111.974783



Rose	mean	mean_se	mean_ci_lower	mean_ci_upper
1996-03-31	39.947938	13.582791	13.326157	66.569719
1996-04-30	44.160636	13.626164	17.453846	70.867426
1996-05-31	30.345261	13.820980	3.256638	57.433883
1996-06-30	39.426939	13.875556	12.231349	66.622530
1996-07-31	55.409547	14.046633	27.878651	82.940442

RMSE of the Final Model1 34.868942267557216



Rose	mean	mean_se	mean_ci_lower	mean_ci_upper
1995-08-31	44.875429	12.574009	20.230824	69.520035
1995-09-30	46.011505	12.735818	21.049761	70.973248
1995-10-31	47.547412	12.786541	22.486251	72.608573
1995-11-30	59.675520	13.038372	34.120781	85.230260
1995-12-31	86.377019	13.060324	60.779255	111.974783
1996-01-31	25.137744	13.324726	-0.978240	51.253727
1996-02-29	32.048665	13.356085	5.871219	58.226111
1996-03-31	39.947938	13.582791	13.326157	66.569719
1996-04-30	44.160636	13.626164	17.453846	70.867426
1996-05-31	30.345261	13.820980	3.256638	57.433883
1996-06-30	39.426939	13.875556	12.231349	66.622530
1996-07-31	55.409547	14.046633	27.878651	82.940442

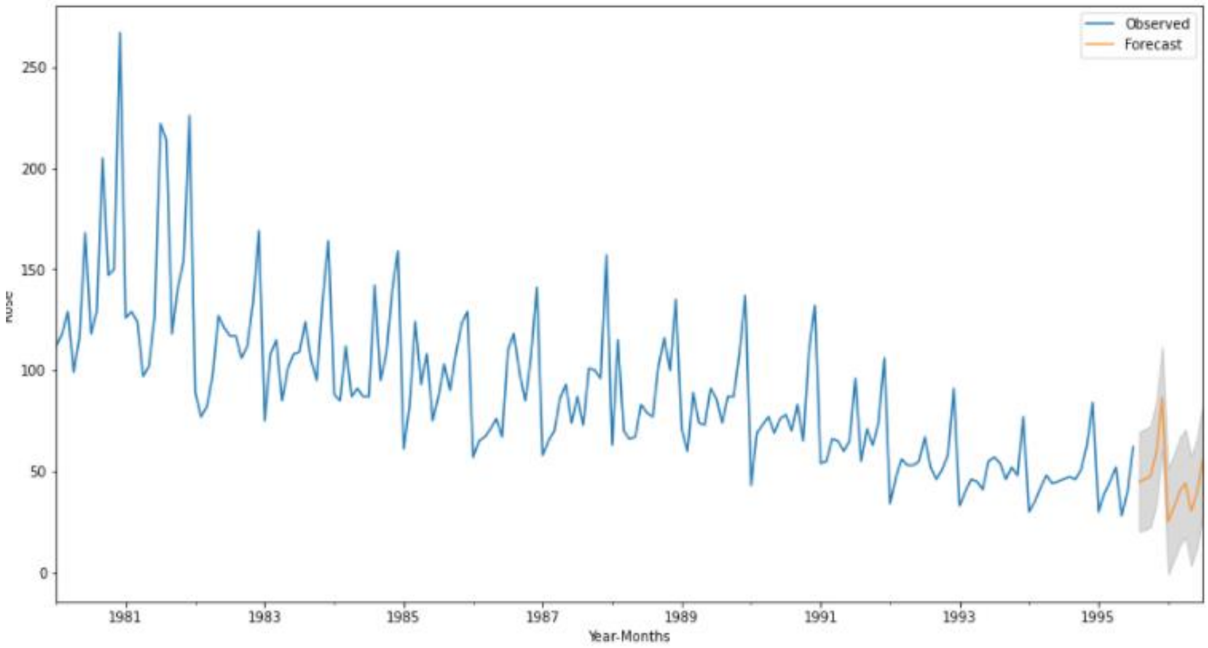
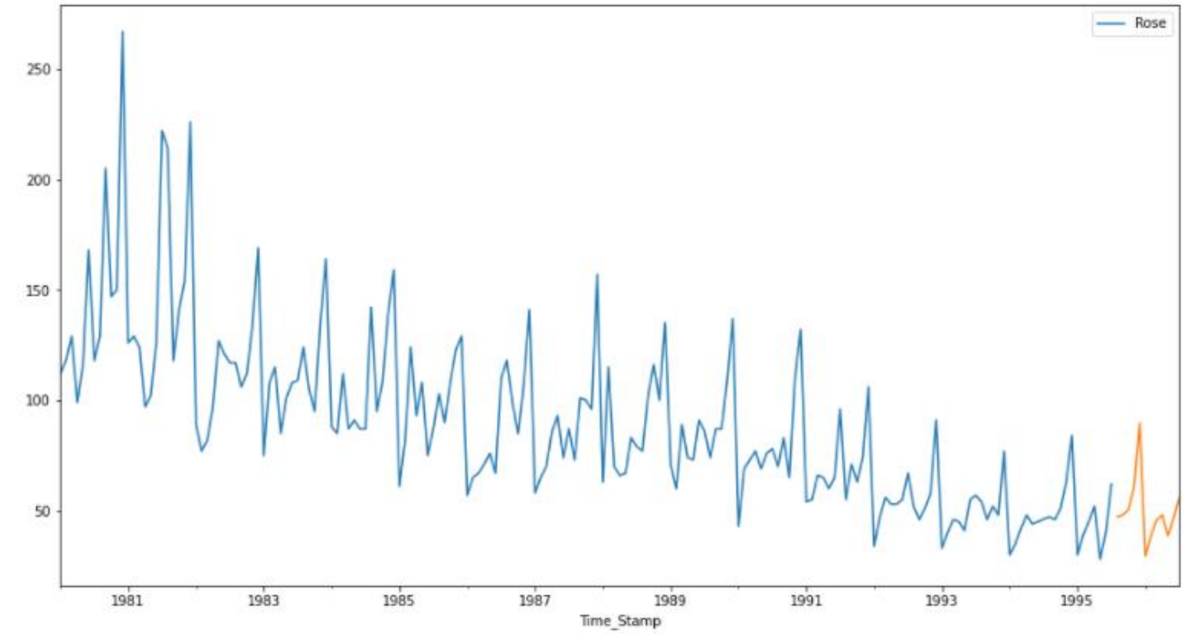


Fig27. Final Model1

Evaluate the model on the whole and predict 12 months into the future with Exponential smoothing.

RMSE: 19.447341066473314

Prediction Plot:



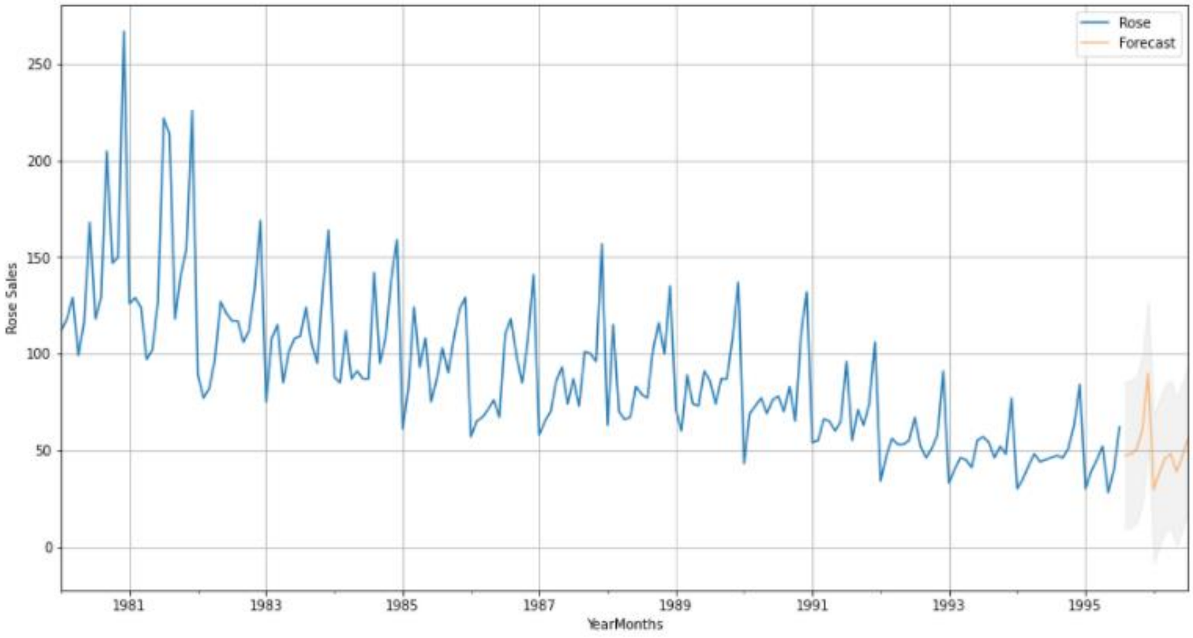


Fig28. Final Model2

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

Time series analysis involves understanding various aspects about the inherent nature of the series so that you are better informed to create meaningful and accurate forecasts**.**

Any time series may be split into the following components:

Base Level + Trend + Seasonality + Error

Observations:

Rose sales shows decrease in trend compared to the previous years.

December month shows the highest sales.

The models are chosen based on least RMSE score.

The sales of Rose wine is seasonal and also had trend. Therefore, the company cannot have the same stock throughout the year.

The Company should use prediction results to plan about future stock.

Insights:

The models are built considering the Trend and Seasonality in to account and we see from the output plot that the future prediction is in line with the trend and seasonality in the previous years.

The company should use the prediction results and capitalize on the high demand seasons and ensure to source and supply the high demand.

The company should use the prediction results to plan the low demand seasons to stock as per the demand.

Products that are discounted should be highlighted so consumers can see the savings prominently Discounts can compel consumers to buy.

As we know how the seasonality is in the prediction company cannot have the same stock through the year.

You should create a dynamic consumer experience with fresh point -of-sale materials and well stocked displays.

Displays need to look fresh and interesting and tell a compelling story about why the consumer should purchase the product.

Seasonal memberships and discounts can be introduced. Consumers get very excited about savings and appreciate discounts being passed on.

Many prominent retailers also have loyalty programs or club member cards that create excitement. A club -member price brings consumers back and improve sales

Events and tastings help draw consumers to your store and generate sales. Retailers with economies of scale successfully sample consumers on more profitable wines.

Some even comparison -taste customers on national brands that are more expensive to demonstrate they are offering a less expensive but superior product.

And bringing in celebrities, sommeliers or trade reps for tastings can help create excitement and drive traffic.