Business Report on Time Series Forecasting Project

-RAHUL SHARMA

Problem (Rose & Sparkling Wines):-

The primary objective of this project is to analyze and forecast wine sales trends for the 20th century based on historical data provided by ABC Estate Wines. We aim to equip ABC Estate Wines with the necessary insights and foresight to enhance sales performance, capitalize on emerging market opportunities, and maintain a competitive edge in the wine industry.

A. Define the problem and perform Exploratory Data Analysis:

Read the data as an appropriate time series data - Plot the data - Perform EDA - Perform Decomposition.

A.1 Read the data as an appropriate time series data:-

ROSE WINE:-

	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

```
Shape of the DataFrame: (187, 2)

Data Types:
YearMonth     object
Rose         float64
dtype: object

Statistical Summary:
             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
```

SPARKLING WINE:-

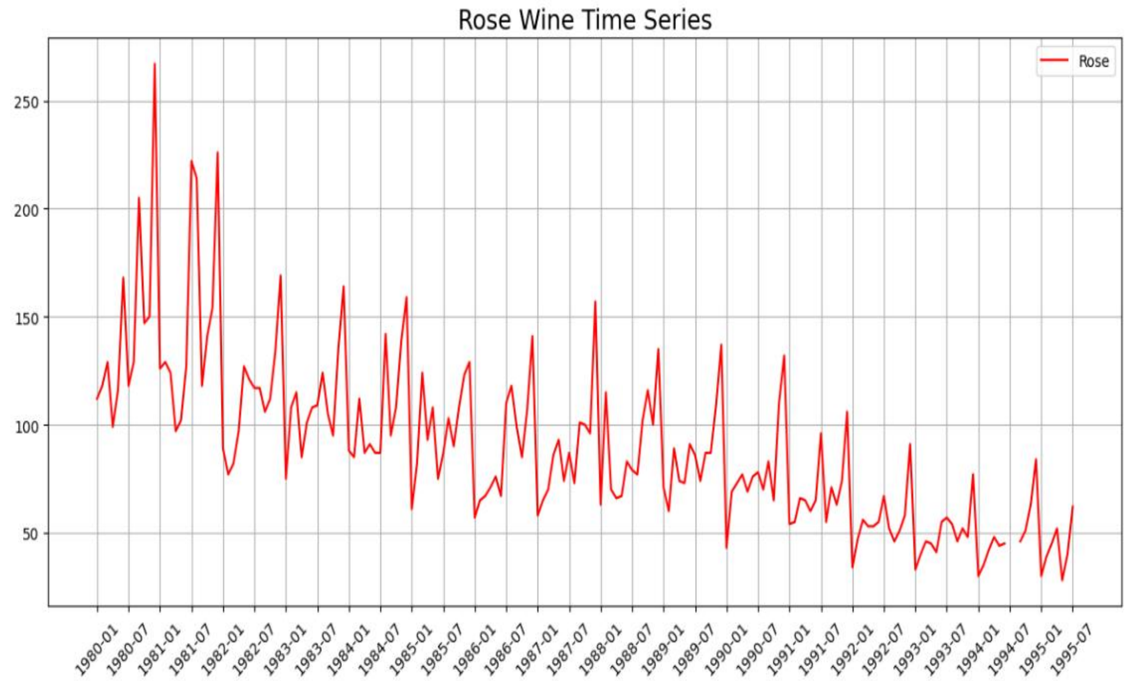
	YearMonth	Sparkling
0	1980-01	1686
1	1980-02	1591
2	1980-03	2304
3	1980-04	1712
4	1980-05	1471

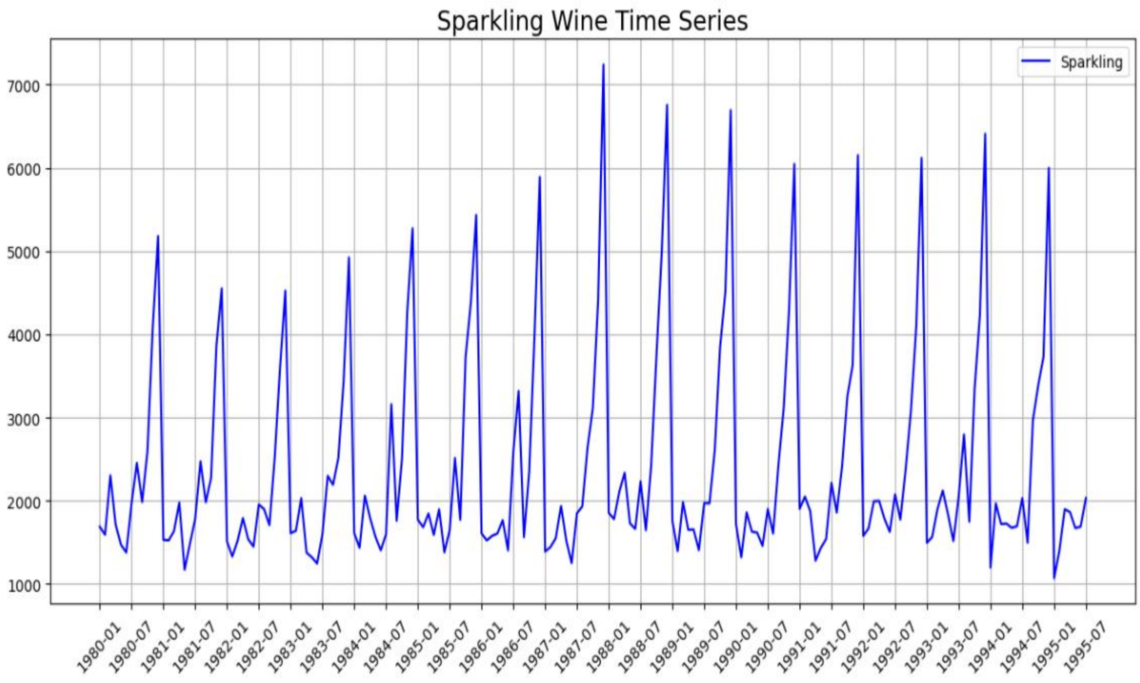
```
Shape of the DataFrame: (187, 2)

Data Types:
YearMonth    object
Sparkling     int64
dtype: object

Statistical Summary:
         Sparkling
count   187.000000
mean   2402.417112
std    1295.111540
min    1070.000000
25%    1605.000000
50%    1874.000000
75%    2549.000000
max    7242.000000
```

A.2 Plot the data:-



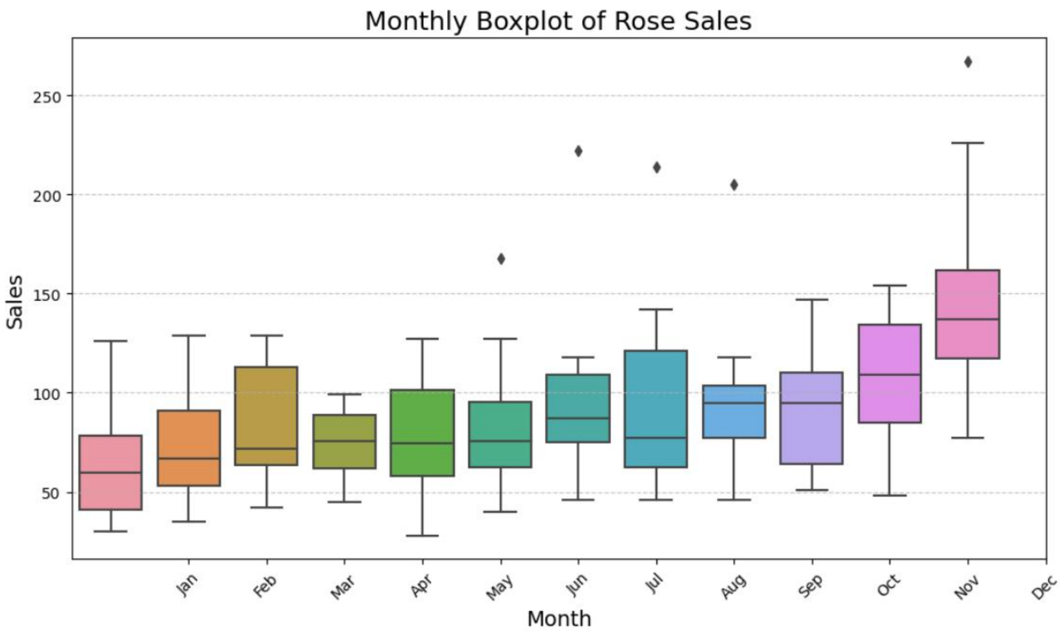


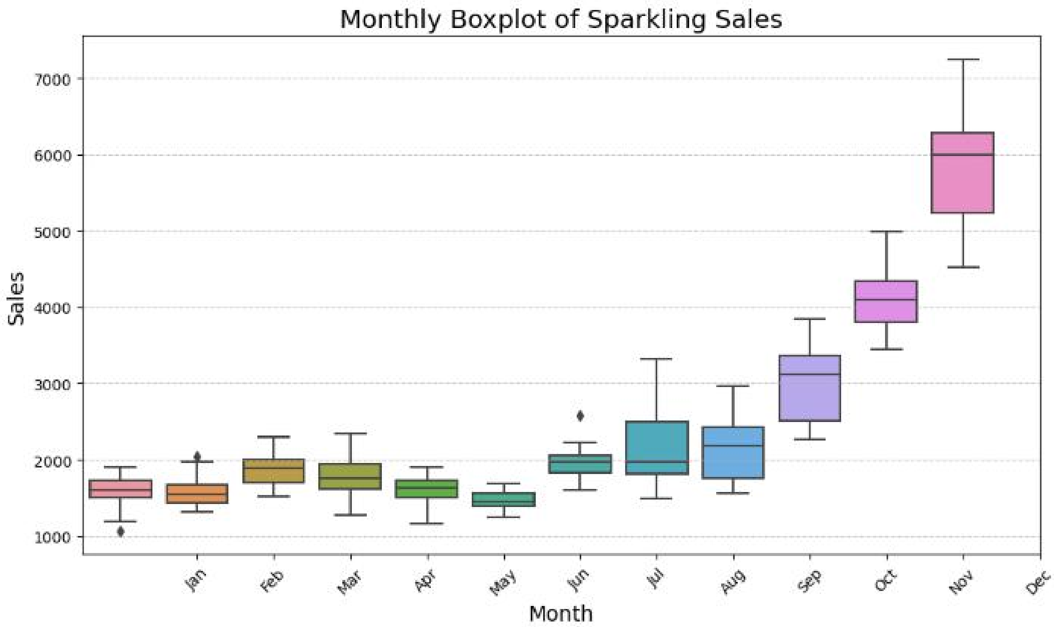
A.3 Perform EDA:-

- Correlation between Rose and Sparkling sales: 0.40457904770543324

```
Summary statistics for Rose dataset:
              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

Summary statistics for Sparkling dataset:
          Sparkling
count    187.000000
mean    2402.417112
std     1295.111540
min     1070.000000
25%     1605.000000
50%     1874.000000
75%     2549.000000
max     7242.000000
```



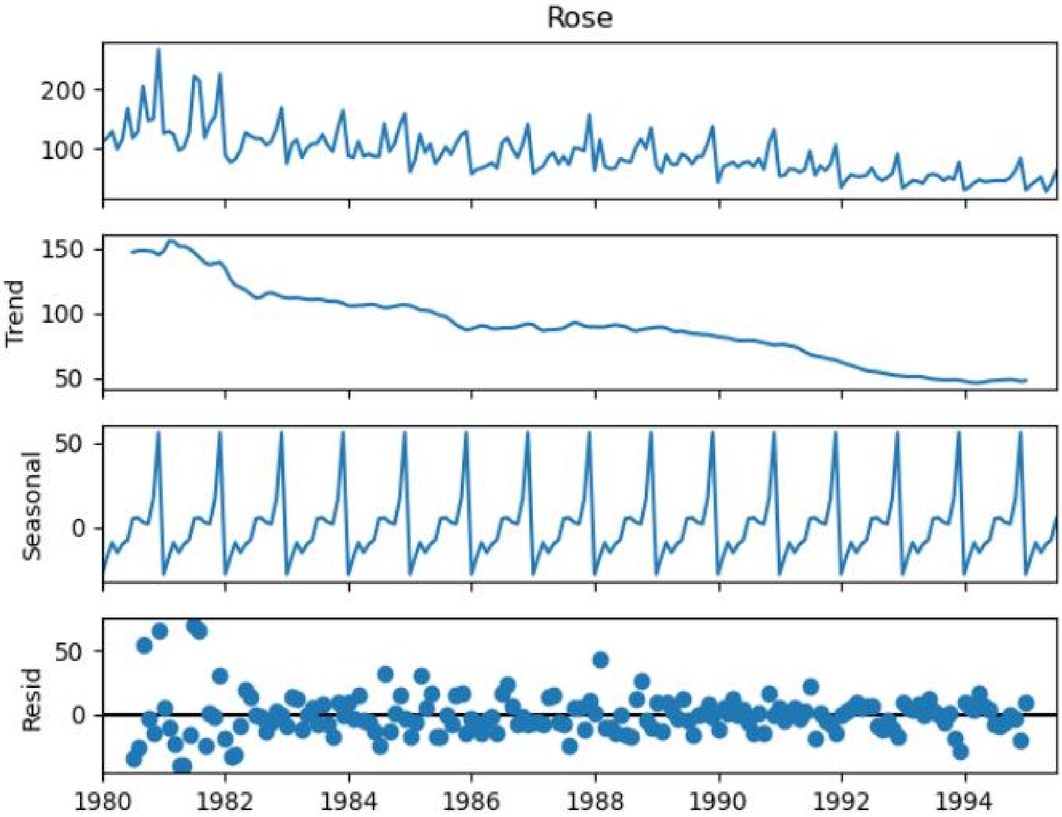


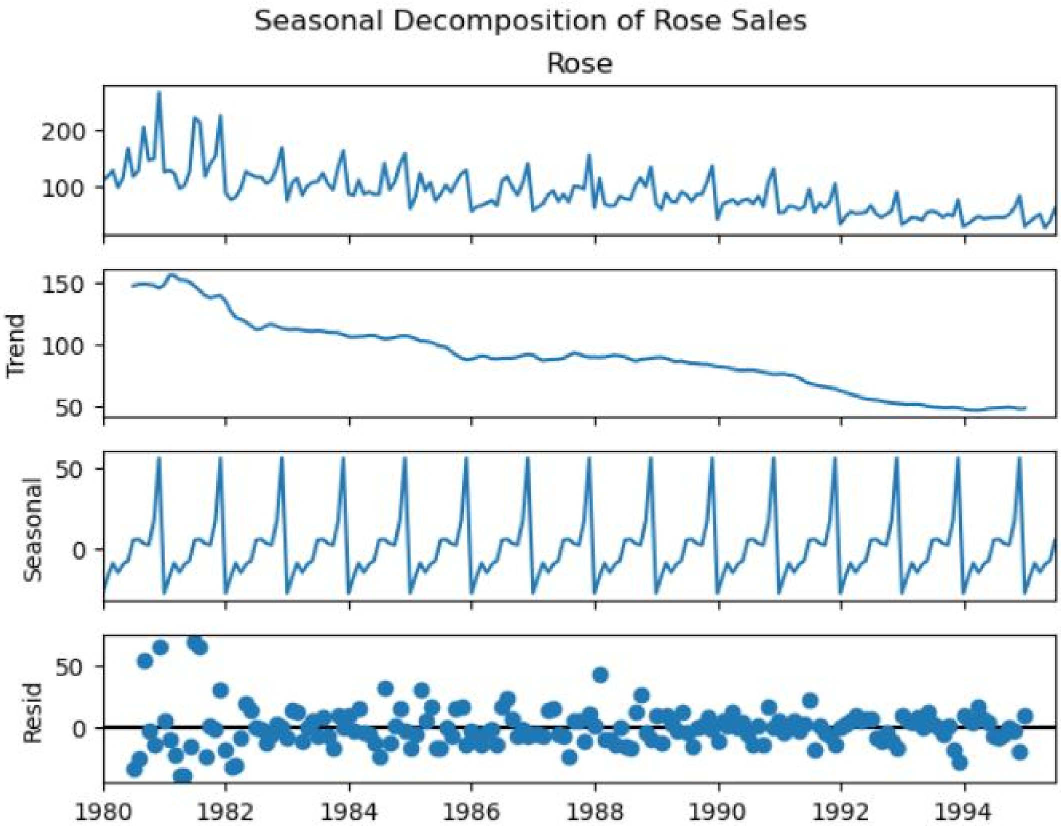
- The center line in each box-plot represents the median sales price.
- The median is the price point where half of the sales were for a higher price and the other half were for a lower price.
- The box in each plot contains the middle 50% of the sales data. The upper edge of the box is the third quartile (Q3) and the lower edge is the first quartile (Q1).
- The whiskers extend from the top and bottom of the box to the highest and lowest sales prices within 1.5 times the interquartile range (IQR). The IQR is the difference between Q3 and Q1. Sales prices outside this range are considered outliers and are shown as individual points in the plot.

ACCORDING TO THE DATA OF ROSE AND SPARKLING WINE:

- The median sales price for roses is lower than the median sales price for sparkling wine.
- The sales price for roses is more spread out than the sales price for sparkling wine. This means that there is a wider range of prices for roses than there is for sparkling wine.
- There are more outliers in the sales price for sparkling wine than there are for roses.

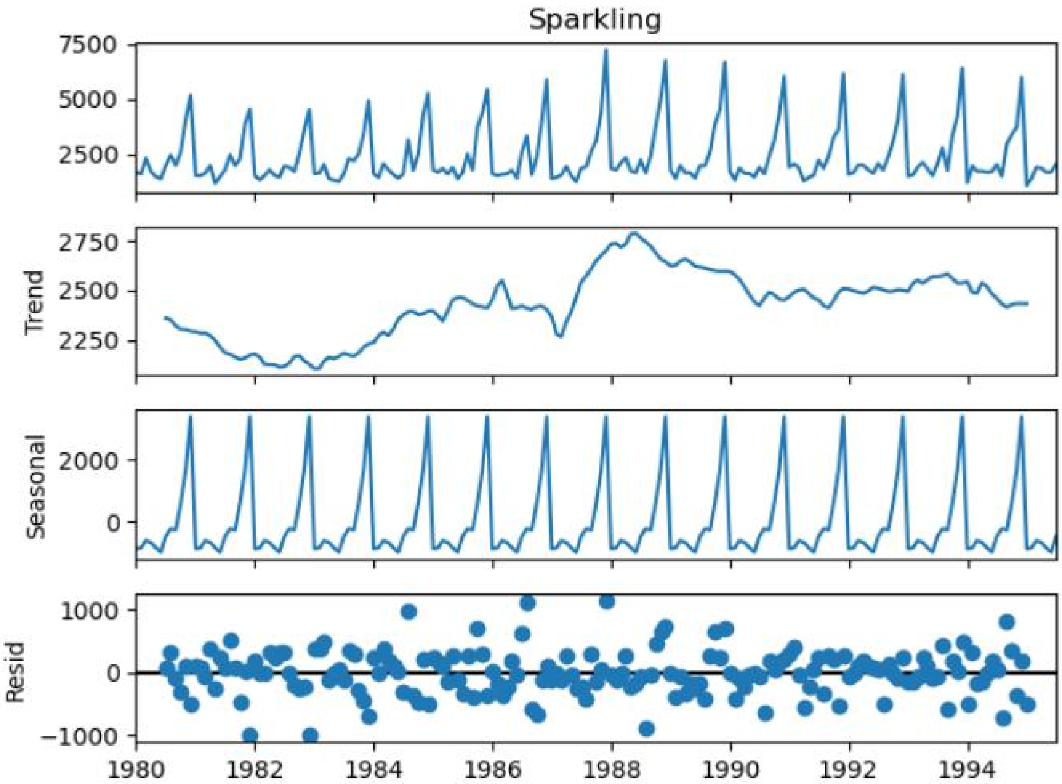
A.4 Perform Decomposition:-

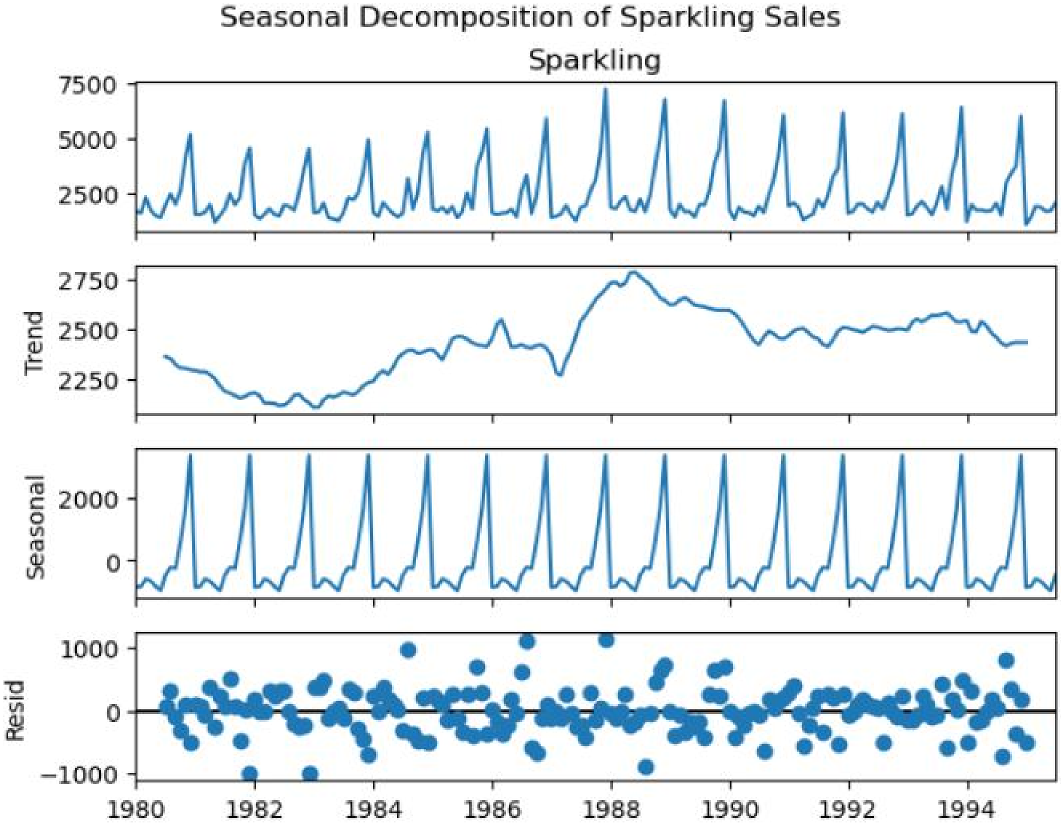




- The top graph shows the original time series data for rose sales.

- The middle graph shows the seasonal component of the sales data. This graph shows how rose sales vary throughout a typical year. For example, rose sales tend to be higher in February around Valentine's Day and in May around Mother's Day.
- The bottom graph shows the trend component of the sales data. This graph shows the overall increase or decrease in rose sales over time.





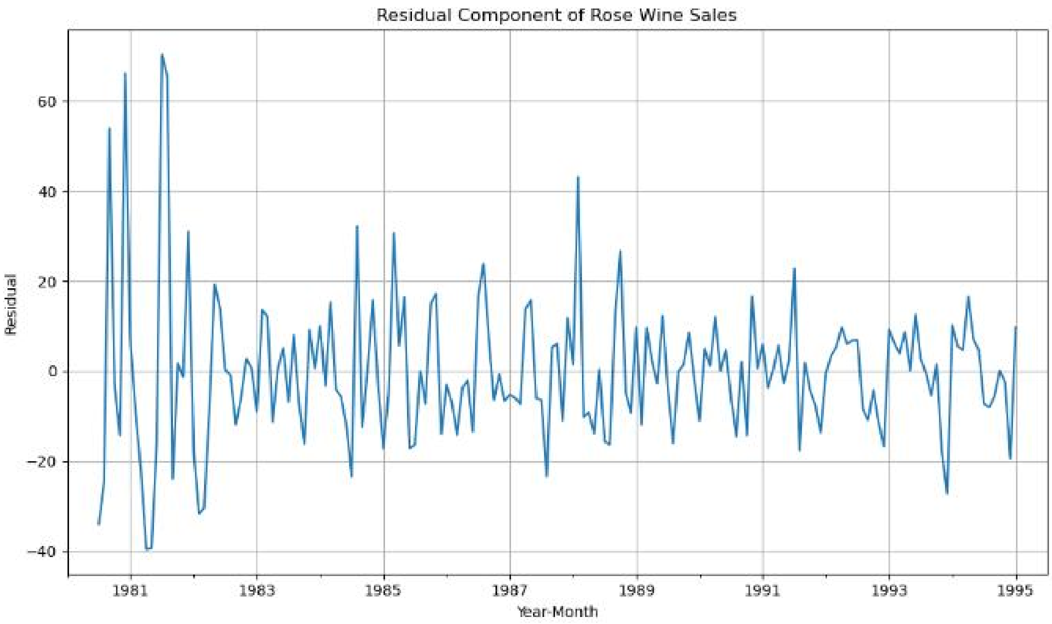
- The top graph shows the original time series data for sparkling sales. It appears to show a declining trend in sparkling wine sales over several years.
- The middle graph shows the seasonal component of the sales data. This graph shows how sparkling sales vary throughout a typical year. There is a peak in December, which could be due to holiday sales.
- The bottom graph shows the trend component of the sales data. This graph confirms the declining trend in sparkling wine sales over time.

TREND, SEASONALITY, RESIDUAL OF ROSE WINE:-

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

Seasonality
 YearMonth
1980-01-01   -27.914120
1980-02-01   -17.440509
1980-03-01    -9.291700
1980-04-01   -15.104200
1980-05-01   -10.202414
1980-06-01    -7.684557
1980-07-01     4.935880
1980-08-01     5.516435
1980-09-01     2.769213
1980-10-01     1.866435
1980-11-01    16.841435
1980-12-01    55.708102
Name: seasonal, dtype: float64

Residual
 YearMonth
1980-01-01          NaN
1980-02-01          NaN
1980-03-01          NaN
1980-04-01          NaN
1980-05-01          NaN
1980-06-01          NaN
1980-07-01   -34.019213
1980-08-01   -24.641435
1980-09-01    53.855787
1980-10-01    -2.949769
1980-11-01   -14.258102
1980-12-01    66.166898
Name: resid, dtype: float64
```



- The title says "Residual Component of Rose Wine Sales" which means this graph shows the difference between the actual sales figures and a predicted baseline sales figure
- The x-axis shows years, ranging from 1983 to 1995.
- The y-axis shows the residual sales. Positive values on the y-axis indicate that sales were higher than predicted in that year. Negative values on the y-axis indicate that sales were lower than predicted in that year.
- Residual sales fluctuate from year to year, with some years having higher than predicted sales and other years having lower than predicted sales.

TREND, SEASONALITY, RESIDUAL OF SPARKLING WINE:-

```
Trend
 YearMonth
1980-01-01            NaN
1980-02-01            NaN
1980-03-01            NaN
1980-04-01            NaN
1980-05-01            NaN
1980-06-01            NaN
1980-07-01    2360.666667
1980-08-01    2351.333333
1980-09-01    2320.541667
1980-10-01    2303.583333
1980-11-01    2302.041667
1980-12-01    2293.791667
Name: trend, dtype: float64

Seasonality
 YearMonth
1980-01-01    -854.260599
1980-02-01    -830.350678
1980-03-01    -592.356630
1980-04-01    -658.490559
1980-05-01    -824.416154
1980-06-01    -967.434011
1980-07-01    -465.502265
1980-08-01    -214.332821
1980-09-01    -254.677265
1980-10-01     599.769957
1980-11-01    1675.067179
1980-12-01    3386.983846
Name: seasonal, dtype: float64

Residual
 YearMonth
1980-01-01           NaN
1980-02-01           NaN
1980-03-01           NaN
1980-04-01           NaN
1980-05-01           NaN
1980-06-01           NaN
1980-07-01     70.835599
1980-08-01    315.999487
1980-09-01    -81.864401
1980-10-01   -307.353290
1980-11-01    109.891154
1980-12-01   -501.775513
Name: resid, dtype: float64
```

Coefficient of Variation for the Residual Component: -296.08335294144075

B. Data Pre-Processing:

Missing value treatment - Visualize the processed data - Train-test split

B.1 Missing values and treatment:-

```
Missing values in Rose dataset after filling:
 Rose    2
dtype: int64

Missing values in Sparkling dataset:
 Sparkling    0
dtype: int64
```

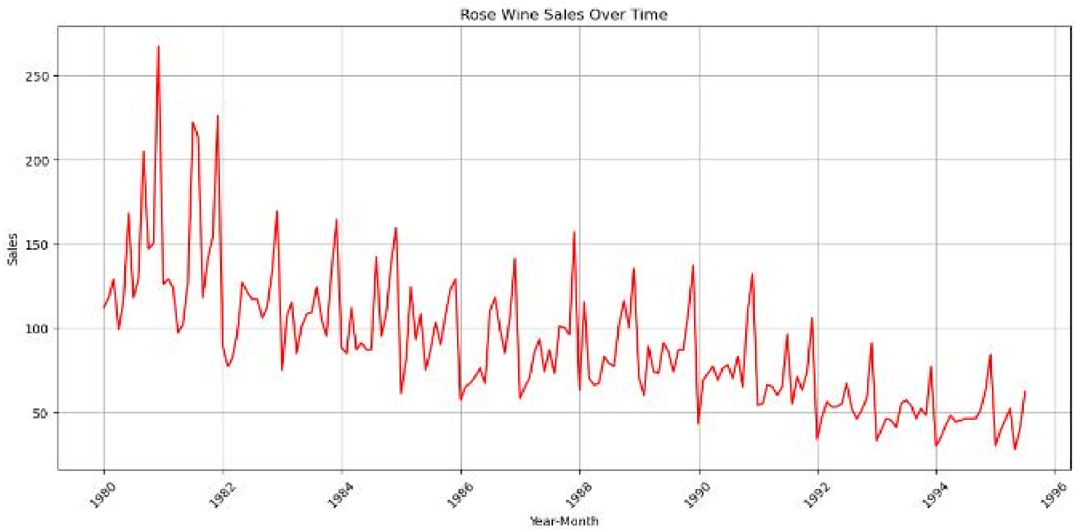
TREATMENT:-

```
Missing values in Rose dataset:
 Rose    0
dtype: int64

Missing values in Sparkling dataset:
 Sparkling    0
dtype: int64
```

B.2 Visualize the processed data:-

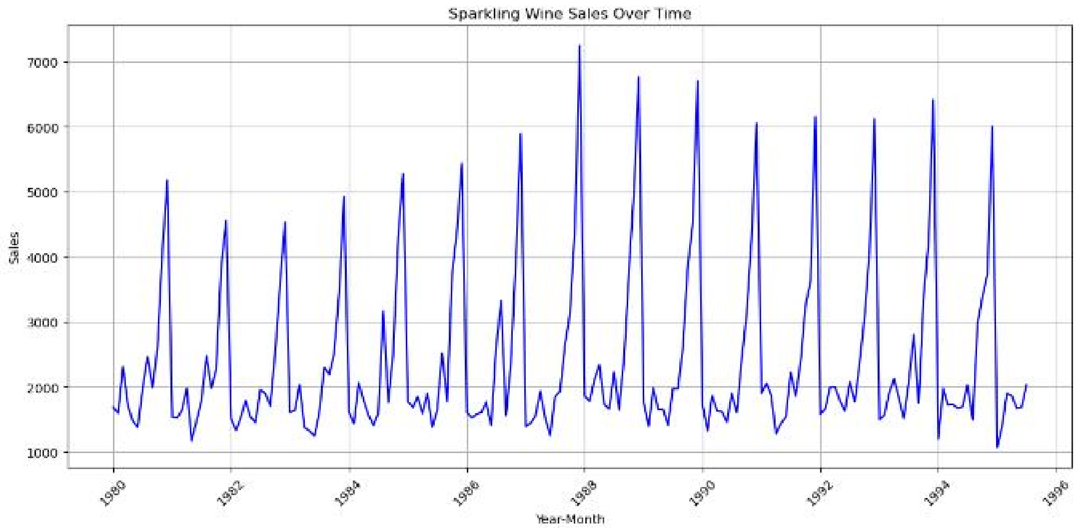
ROSE:-



- The y-axis shows the sales of rose wine.
- The x-axis shows the year and month. The time period ranges from January 1980 to December 1996.

- There appears to be an upward trend in rose wine sales over this time period. Sales appear to be higher in later years (1990s) than in earlier years (1980s).

SPARKLING:-



- The y-axis shows the number of sparkling wine sales.
- The x-axis shows the year, ranging from 1980 to 1996.
- There appears to be some fluctuation in sales throughout the years, but there is no clear upward or downward trend. But in 1988 sales is at highest point as compare to others and lowest in around in 1995.

B.3 Train-Test and Split :-

```
Rose Train shape: (149, 2)
Rose Test shape: (38, 2)
Sparkling Train shape: (149, 2)
Sparkling Test shape: (38, 2)
```

ROSE:-

```
First few rows of Rose Training Data
```

	YearMonth	Rose
0	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

Last few rows of Rose Training Data

	YearMonth	Rose
144	1992-01-01	34.0
145	1992-02-01	47.0
146	1992-03-01	56.0
147	1992-04-01	53.0
148	1992-05-01	53.0

First few rows of Rose Test Data

	YearMonth	Rose
149	1992-06-01	55.0
150	1992-07-01	67.0
151	1992-08-01	52.0
152	1992-09-01	46.0
153	1992-10-01	51.0

Last few rows of Rose Test Data

	YearMonth	Rose
182	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

SPARKILING:-

First few rows of Sparkling Training Data

	YearMonth	Sparkling
0	1980-01-01	1686
1	1980-02-01	1591
2	1980-03-01	2304
3	1980-04-01	1712
4	1980-05-01	1471

Last few rows of Sparkling Training Data

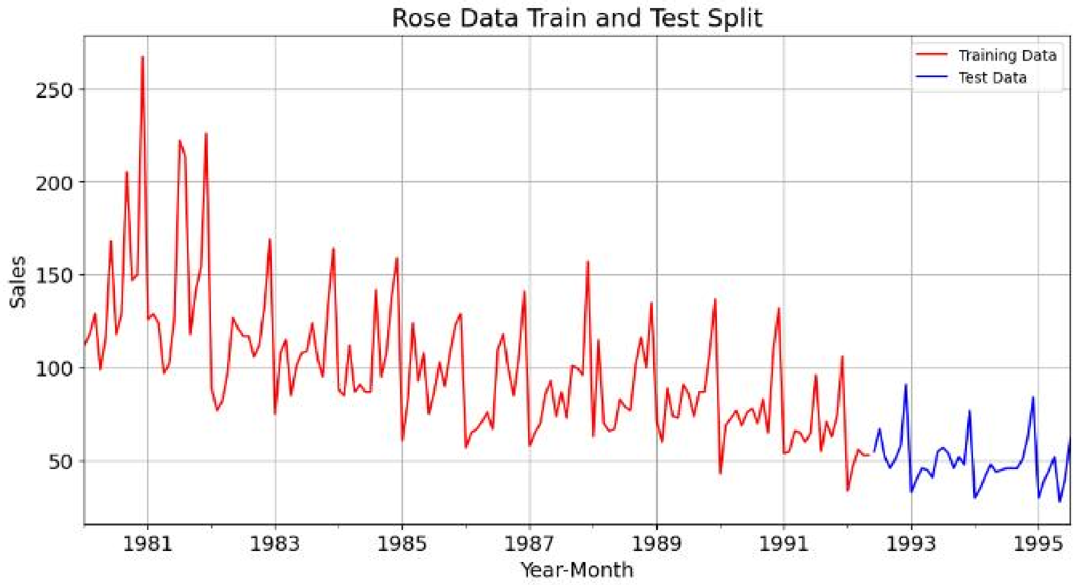
	YearMonth	Sparkling
144	1992-01-01	1577
145	1992-02-01	1667
146	1992-03-01	1993
147	1992-04-01	1997
148	1992-05-01	1783

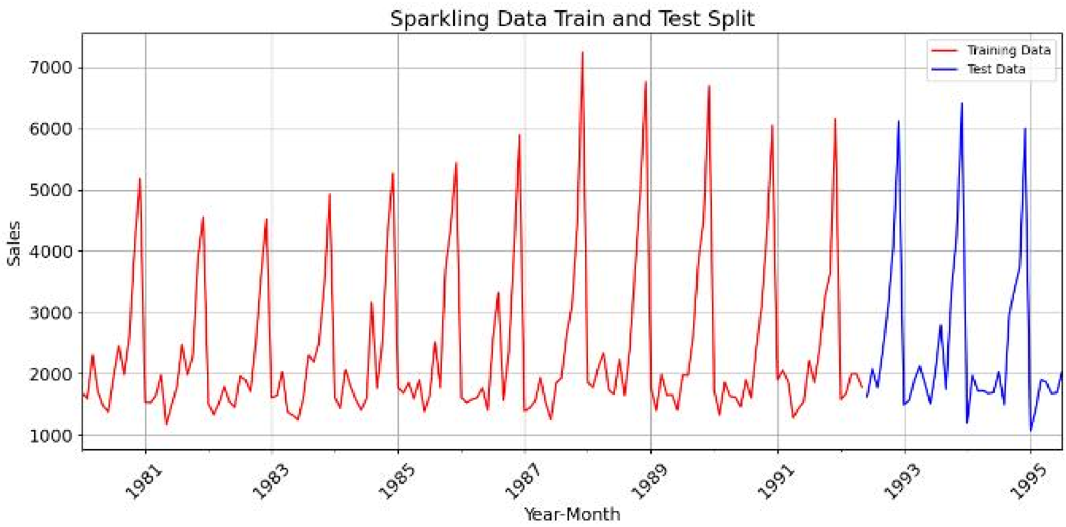
First few rows of Sparkling Test Data

	YearMonth	Sparkling
149	1992-06-01	1625
150	1992-07-01	2076
151	1992-08-01	1773
152	1992-09-01	2377
153	1992-10-01	3088

Last few rows of Sparkling Test Data

	YearMonth	Sparkling
182	1995-03-01	1897
183	1995-04-01	1862
184	1995-05-01	1670
185	1995-06-01	1688
186	1995-07-01	2031

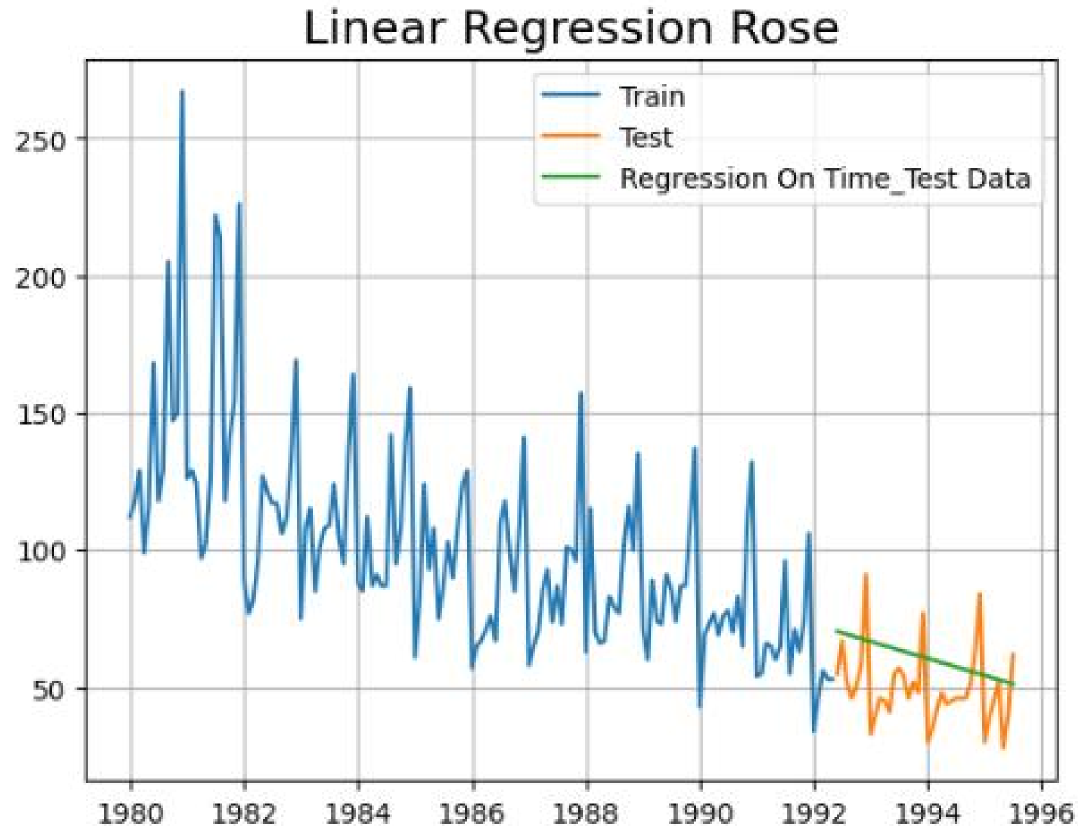


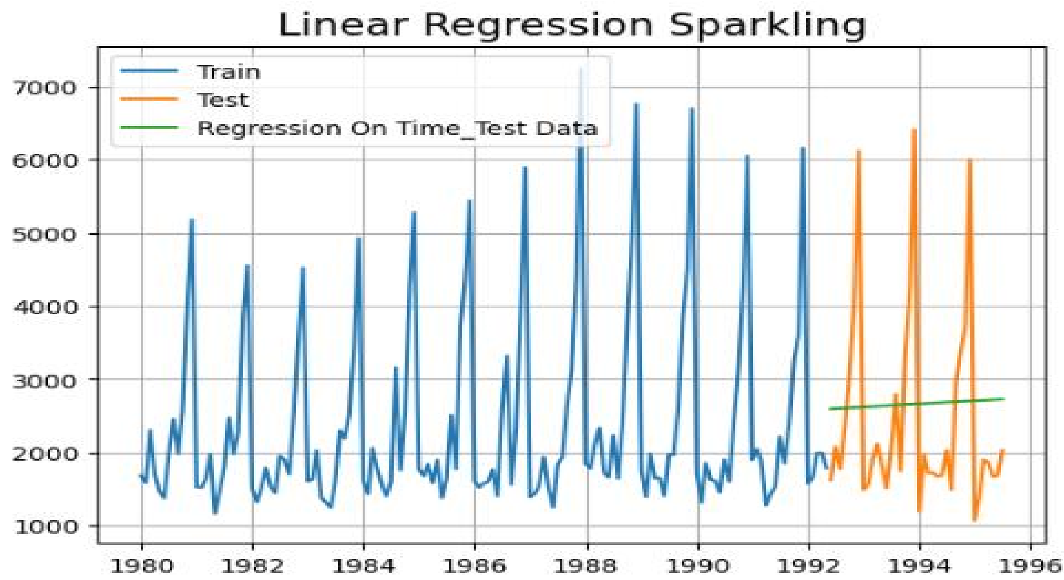


C. <u>Model Building - Original Data:-</u>

Build forecasting models - Linear regression - Simple Average - Moving Average - Exponential Models (Single, Double, Triple) - Check the performance of the models built

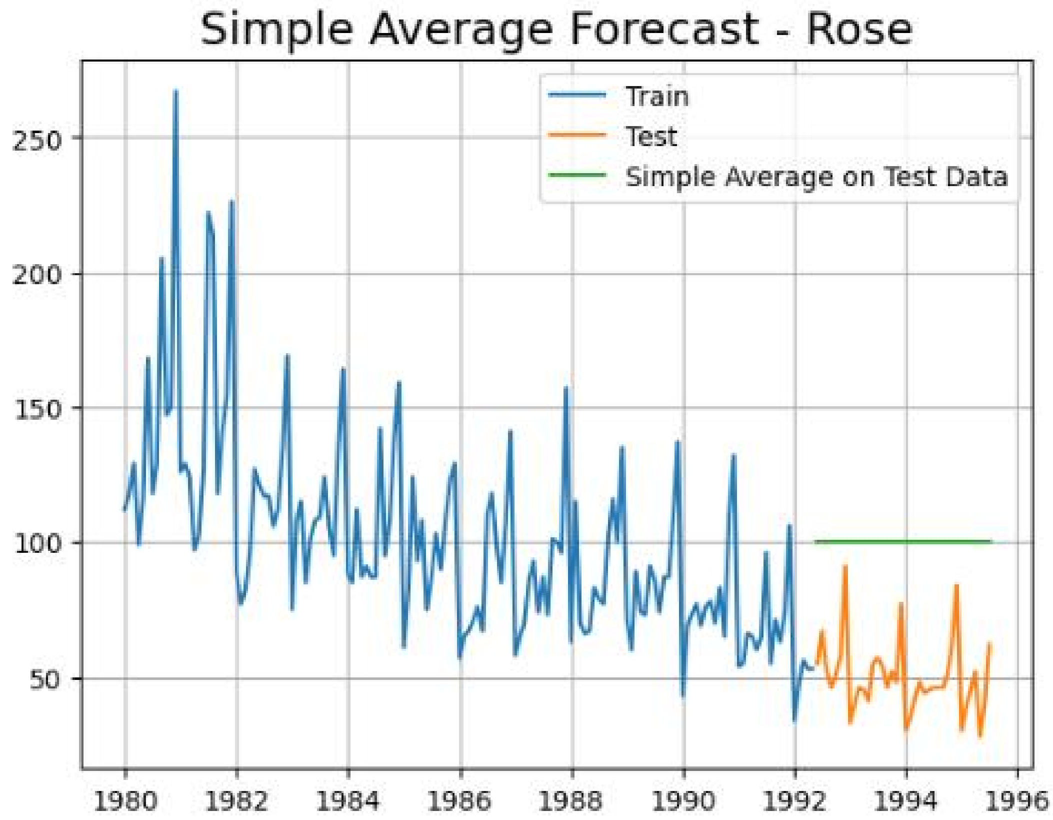
<u>C.1 Model 1-Linear Regression:-</u>

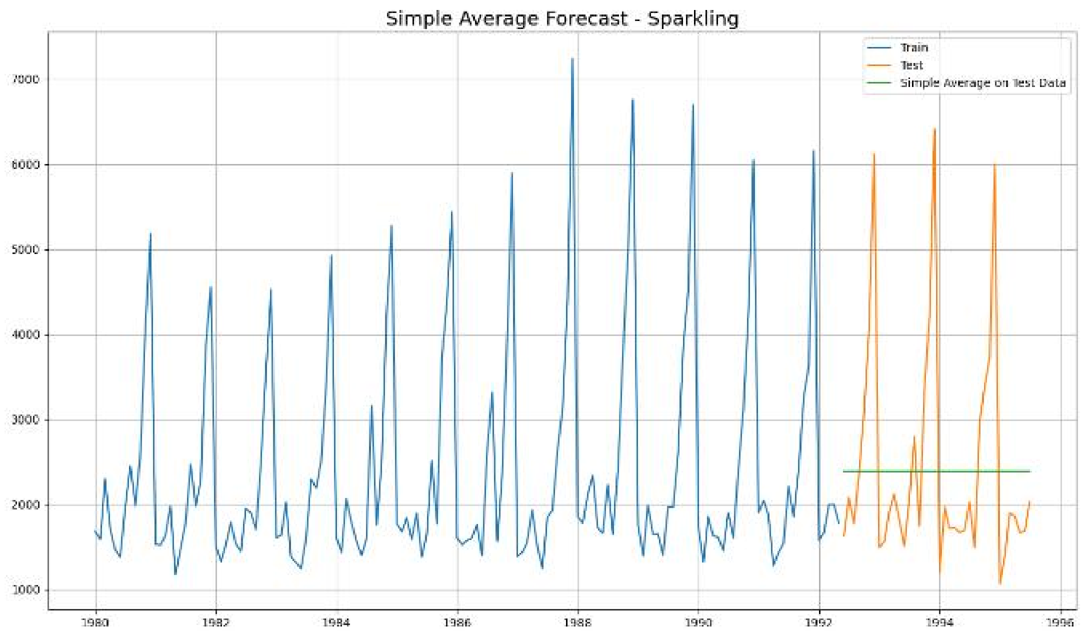




	Test RMSE Rose	Test RMSE Sparkling
RegressionOnTime	17.475054	1349.042457

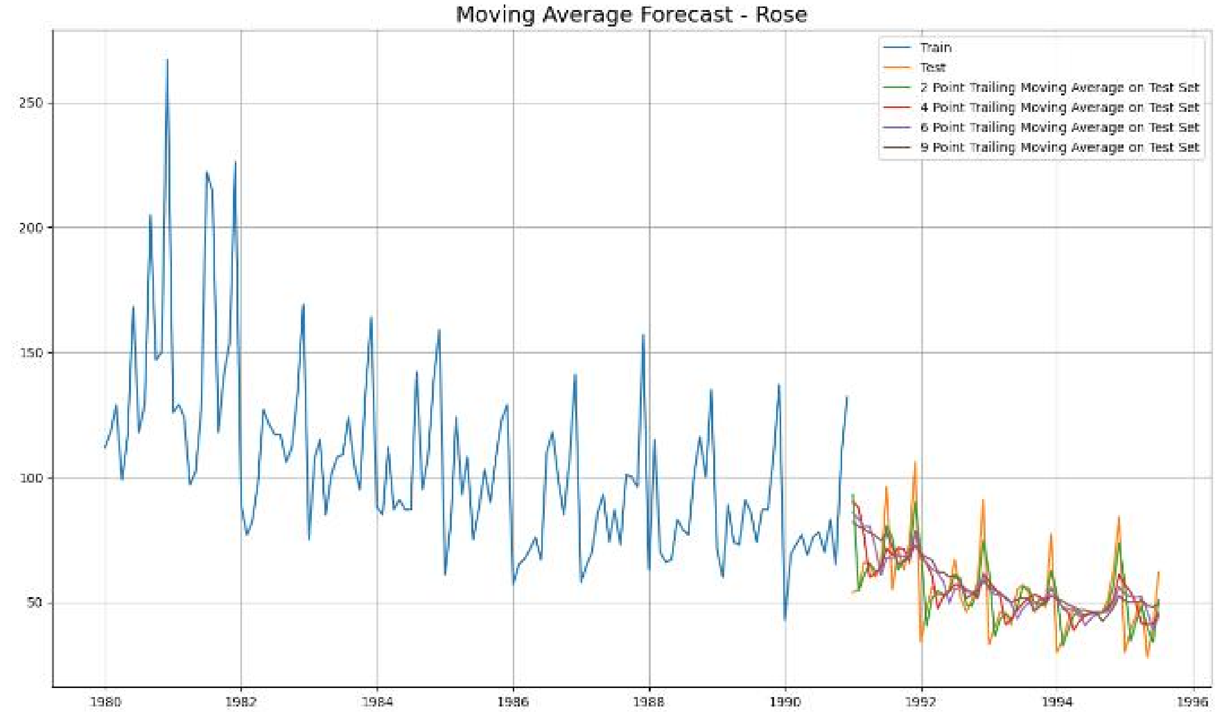
C2. Model 2- Simple Average :-

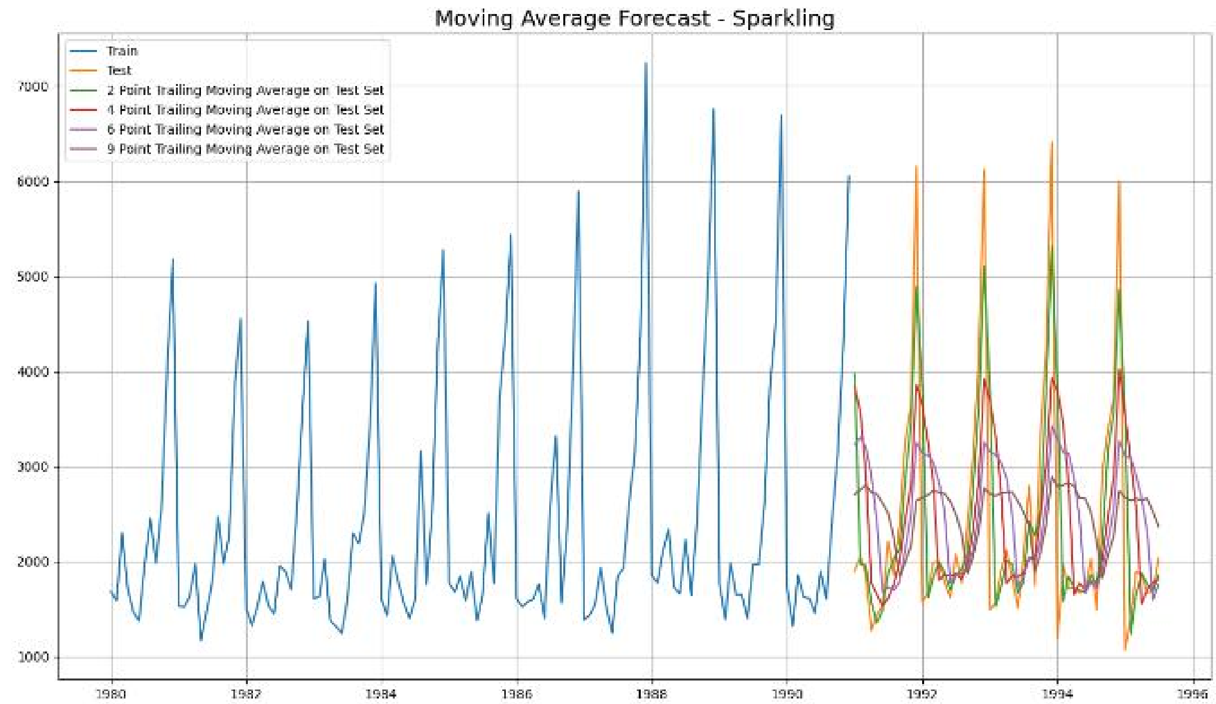




	Test RMSE Rose	Test RMSE Sparkling
SimpleAverageModel	52.184392	2705.474937

C3. Model 3- Moving Average:-





	Test RMSE Rose	Test RMSE Sparkling
2pointTrailingMovingAverage	11.529409	813.400684
4pointTrailingMovingAverage	14.448930	1156.589694
6pointTrailingMovingAverage	14.560046	1283.927428
9pointTrailingMovingAverage	14.724503	1346.278315

JUST COMPARING THE TILL ALL 3 MODELS:-

	Test RMSE Rose	Test RMSE Sparkling
SimpleAverageModel	52.184392	2705.474937
2pointTrailingMovingAverage	11.529409	813.400684
4pointTrailingMovingAverage	14.448930	1156.589694
6pointTrailingMovingAverage	14.560046	1283.927428
9pointTrailingMovingAverage	14.724503	1346.278315
RegressionOnTime	17.475054	1349.042457

Based on the above information , a **2-point moving average model seems to be a good starting point for forecasting both rose and sparkling wine sales due to its low Root Mean Squared Error (RMSE) score.** However, it's important to consider some additional factors before settling on this model:

Simplicity vs. Accuracy: A 2-point moving average is a very simple model that only considers the most recent two data points. This can be beneficial because it's easy to understand and implement. However, it may not capture more complex trends in the data.
Data Availability: A 2-point moving average only needs two data points to make a prediction. This can be an advantage if data is limited.

For further analysis:

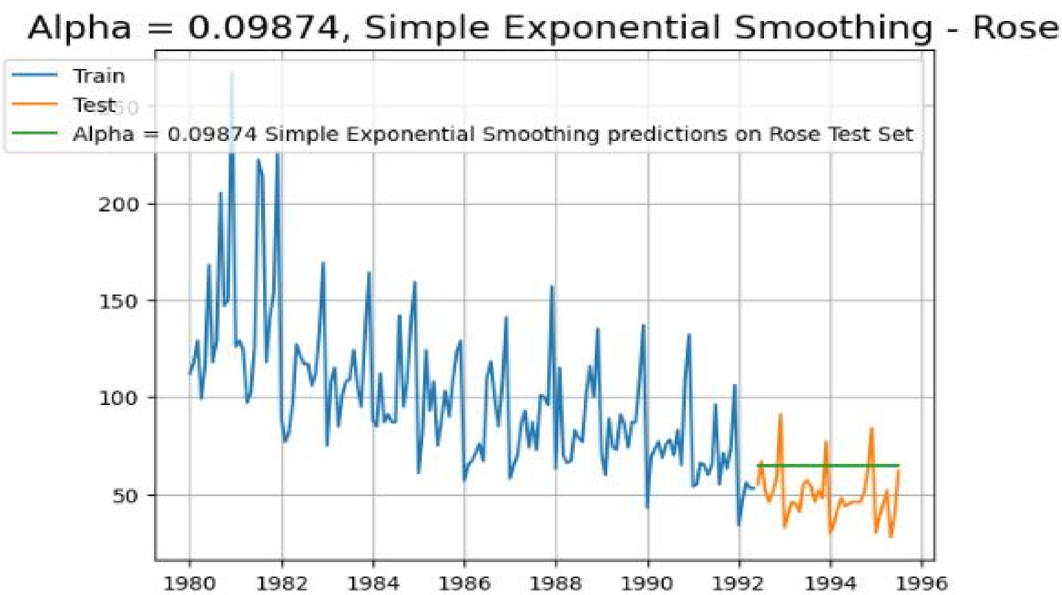
Compare the 2-point moving average model to other models: Try using more sophisticated models like ARIMA or SARIMA which can capture trends and seasonality. Compare the RMSE scores of these models to the 2-point moving average model.
Visualize the forecasts: Plot the actual sales data along with the forecasts from the 2-point moving average model and any other models you consider. This will helps us to see how well the models are capturing the trends in the data.

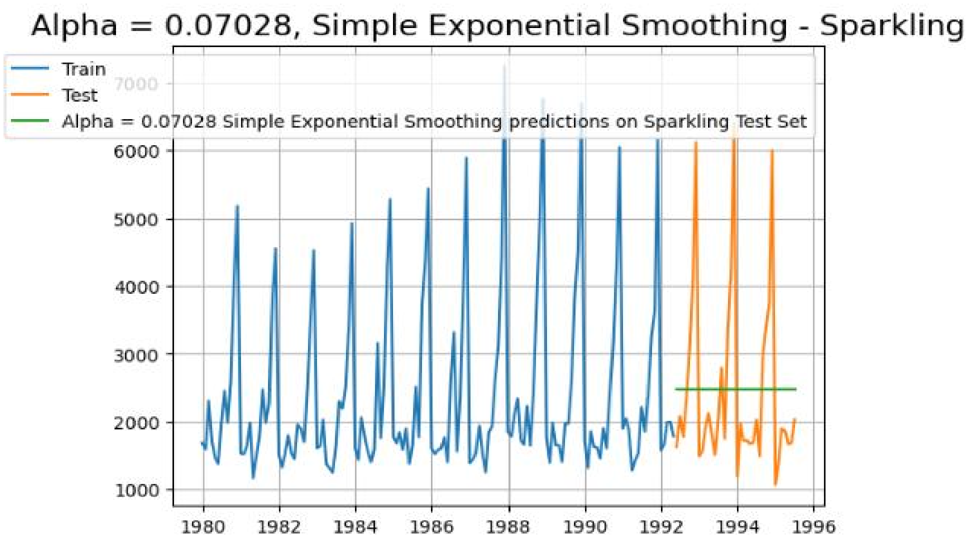
C4. Model 4- Exponential Models (Single, Double, Triple):-

Exponential Smoothing Models -

• Single/Simple Exponential Smoothing with Additive Errors - ETS(A, N, N)
• Double Exponential Smoothing with Additive Errors, Additive Trends - ETS(A, A, N)
• Triple Exponential Smoothing with Additive Errors, Additive Trends, Additive Seasonality - ETS(A, A, A)
• Triple Exponential Smoothing with Additive Errors, Additive Trends, Multiplicative Seasonality - ETS(A, A, M)
• Triple Exponential Smoothing with Additive Errors, Additive DAMPED Trends, Additive Seasonality - ETS(A, Ad, A)
• Triple Exponential Smoothing with Additive Errors, Additive DAMPED Trends, Multiplicative Seasonality - ETS(A, Ad, M).

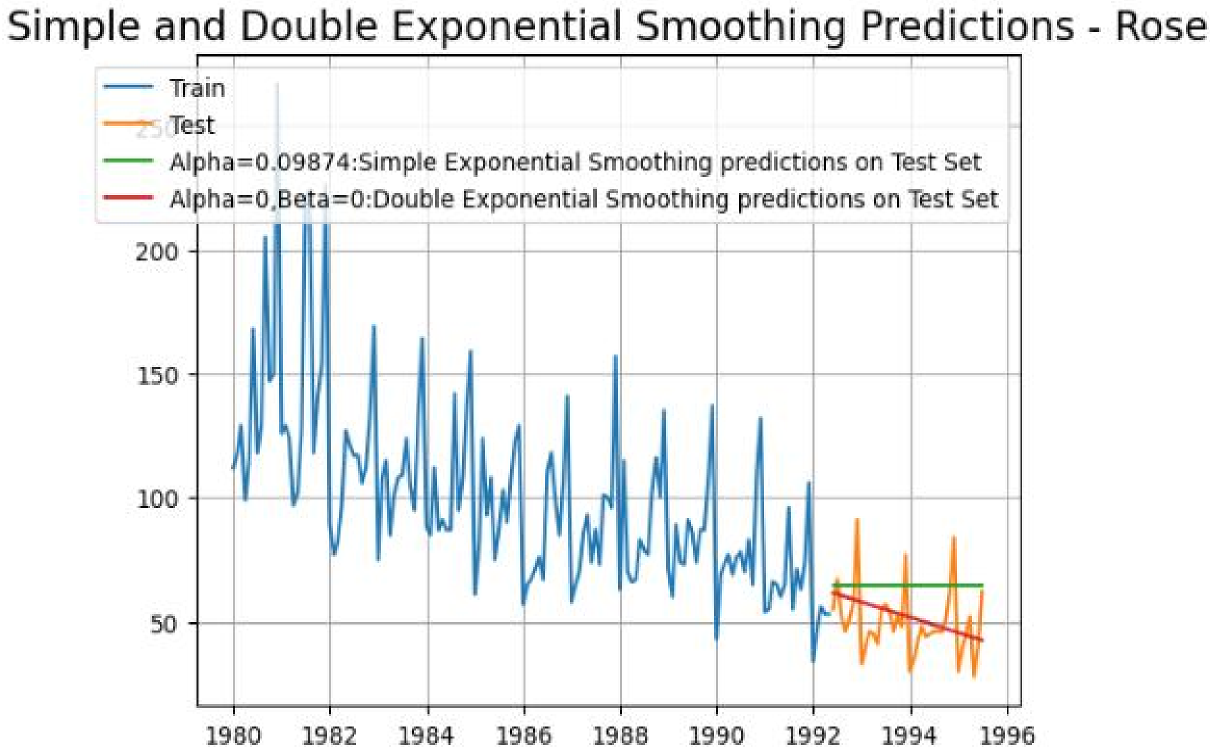
a) SINGLE Exponential Smoothing with additive errors:-

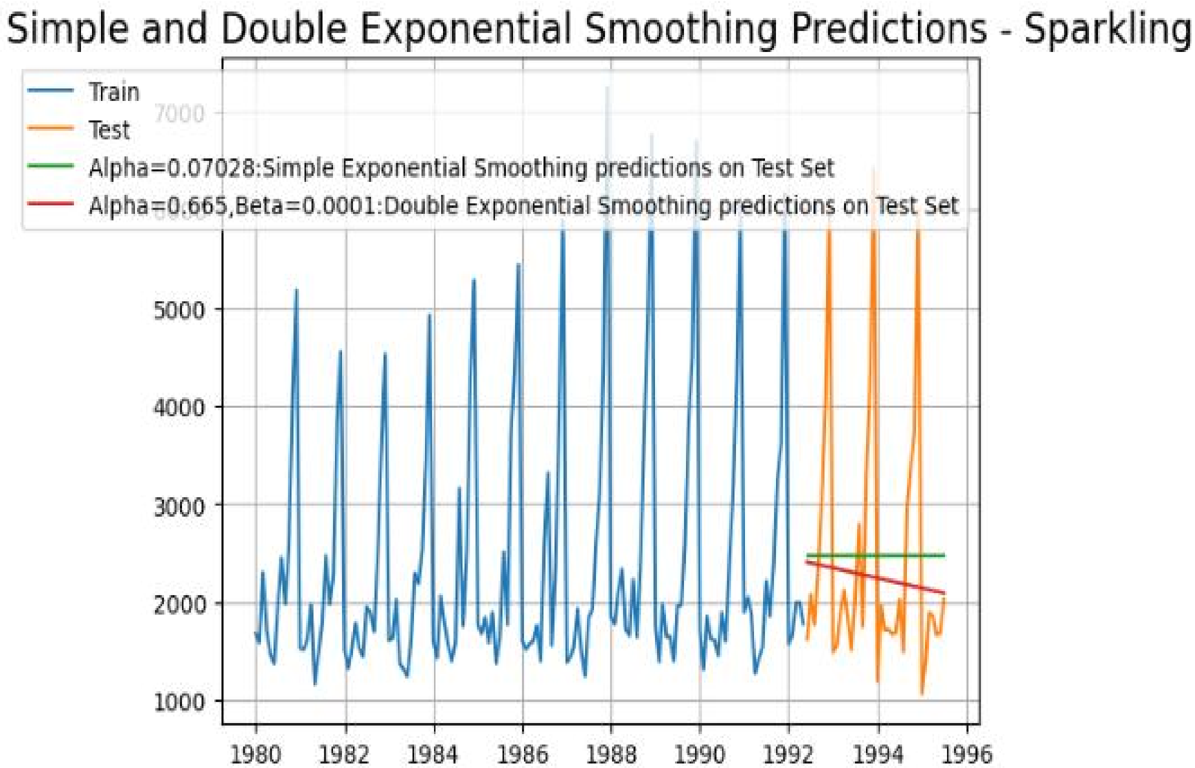




	Test RMSE Rose	Test RMSE Sparkling
Simple Exponential Smoothing	20.263217	1329.402416

b) DOUBLE Exponential Smoothing with additive errors:-



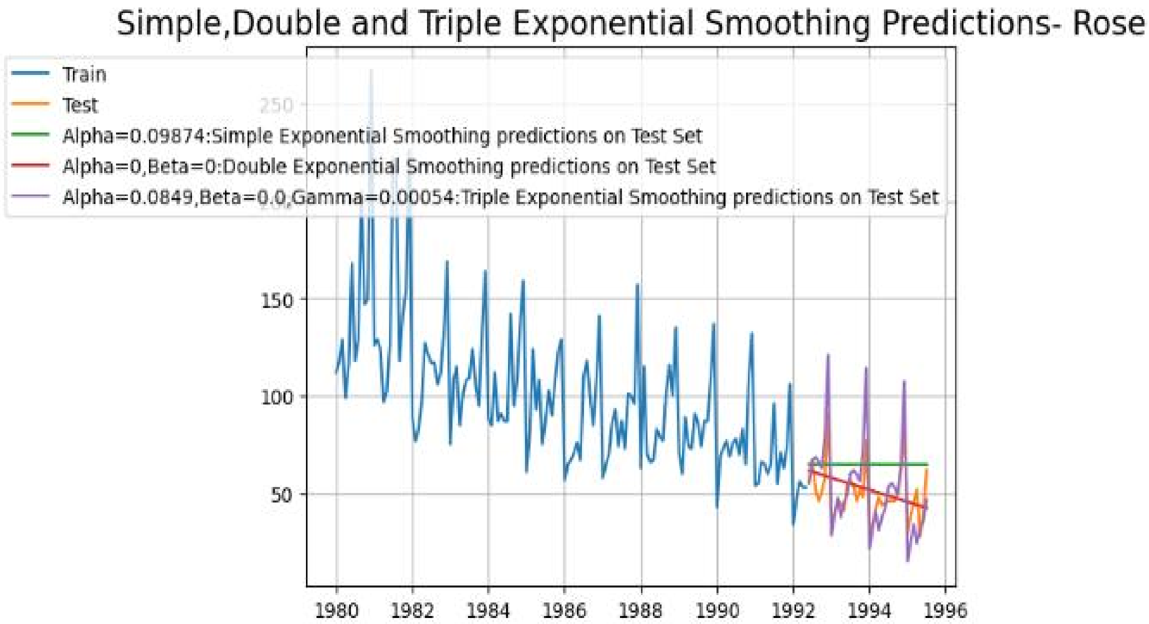


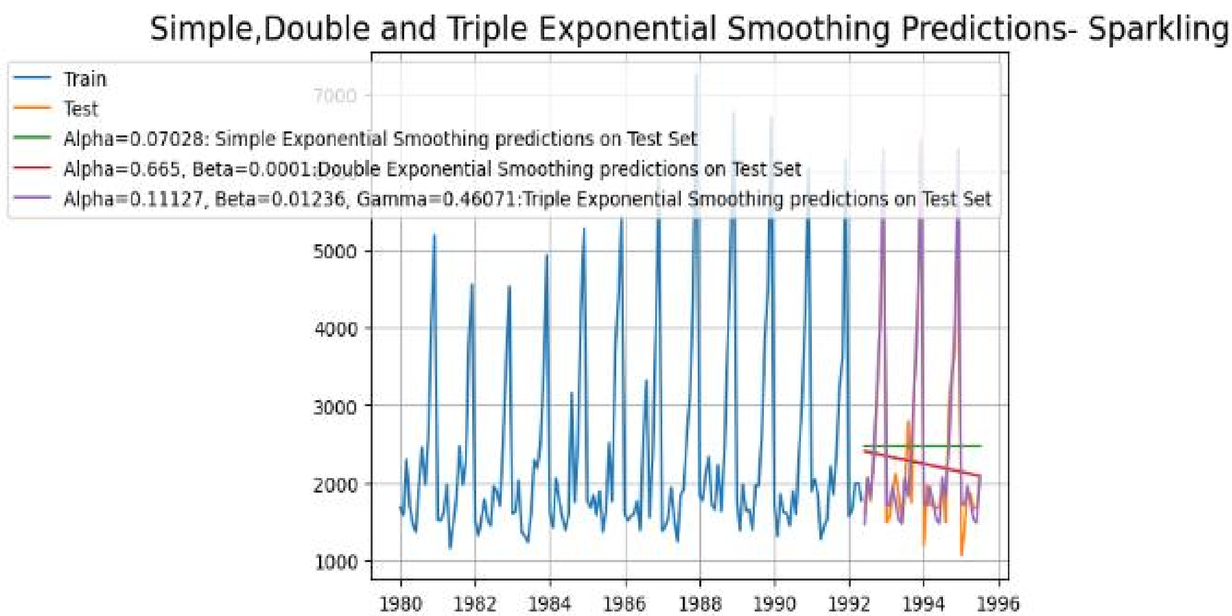
	Test RMSE Rose	Test RMSE Sparkling
Double Exponential Smoothing	13.723188	1340.452773

Rose - Alpha = 0 ; Beta = 0
Sparkling - Alpha = 0.665 ; Beta = 0.0001

Double Exponential Smoothing (DES) is a clear improvement over Single Exponential Smoothing (SES) in this case because it can capture trends in the data.

c) Triple Exponential Smoothing with additive errors:-





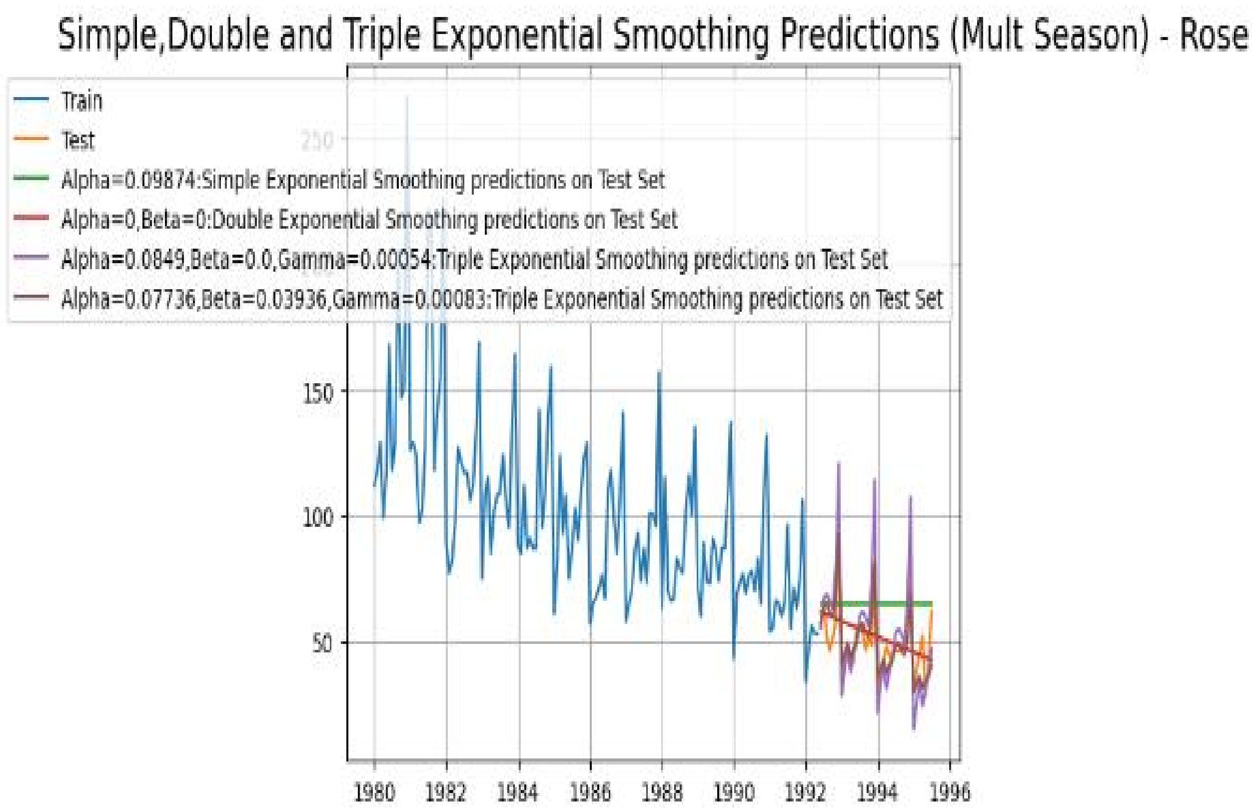
	Test RMSE Rose	Test RMSE Sparkling
Triple Exponential Smoothing (Additive Season)	13.814672	304.269498

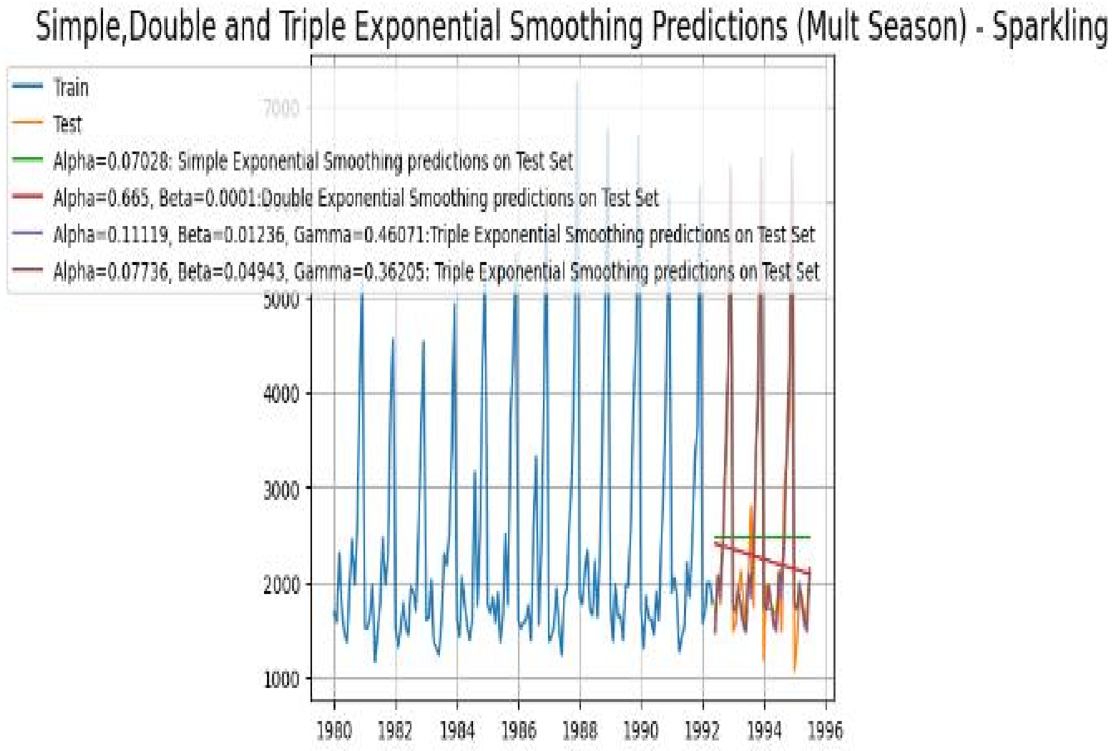
JUST COMPARING:

	Test RMSE Rose	Test RMSE Sparkling
SimpleAverageModel	52.184392	2705.474937
2pointTrailingMovingAverage	11.529409	813.400684
4pointTrailingMovingAverage	14.448930	1156.589694
6pointTrailingMovingAverage	14.560046	1283.927428
9pointTrailingMovingAverage	14.724503	1346.278315
RegressionOnTime	17.475054	1349.042457
Simple Exponential Smoothing	20.263217	1329.402416
Double Exponential Smoothing	13.723188	1340.452773
Triple Exponential Smoothing (Additive Season)	13.814672	304.269498

- **Rose** - Alpha = 0.0849 ; Beta = 0.0 ; Gamma = 0.00054
- **Sparkling** - Alpha = 0.11127 ; Beta = 0.01236 ; Gamma = 0.46071

d) Triple Exponential Smoothing with Additive errors, Additive Trends, Multiplicative Seasonality - ETS(A, A, M):-

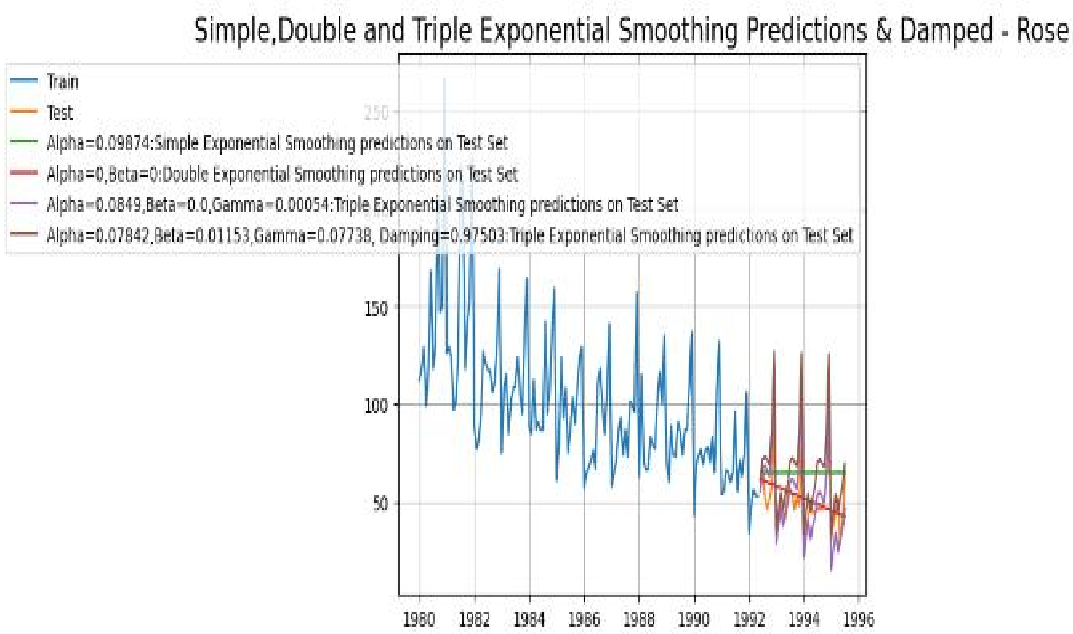


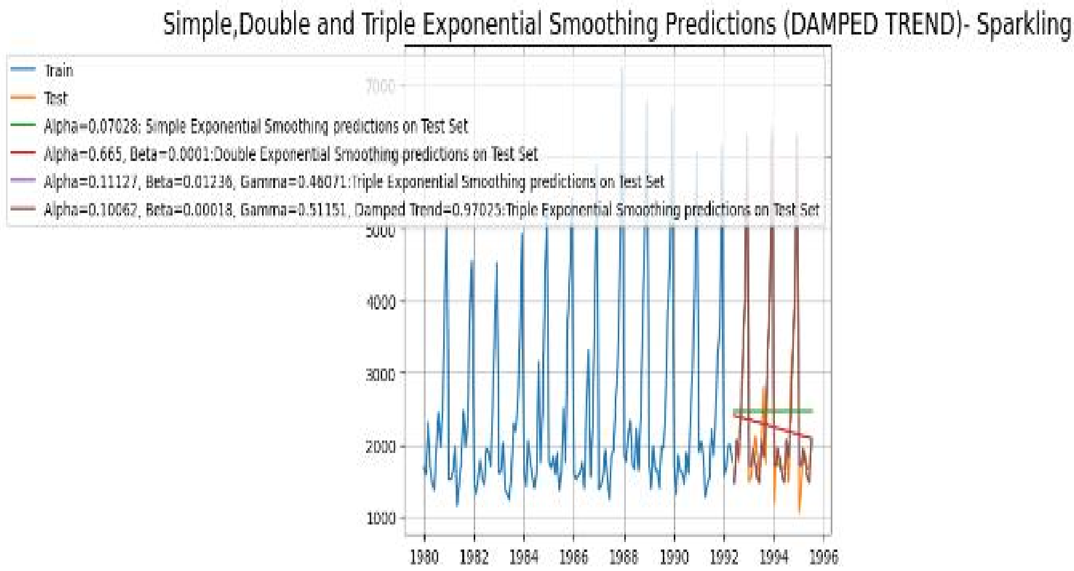


	Test RMSE Rose	Test RMSE Sparkling
Triple Exponential Smoothing (Multiplicative Season)	8.392115	318.448069

- **Rose** - Alpha = 0.07736, Beta = 0.03936, Gamma = 0.00083
- **Sparkling** - Alpha = 0.07736, Beta = 0.04943, Gamma = 0.36205

e) Triple Exponential Smoothing with Additive Errors, Additive DAMPED Trends, Additive Seasonality - ETS(A, Ad, A):-

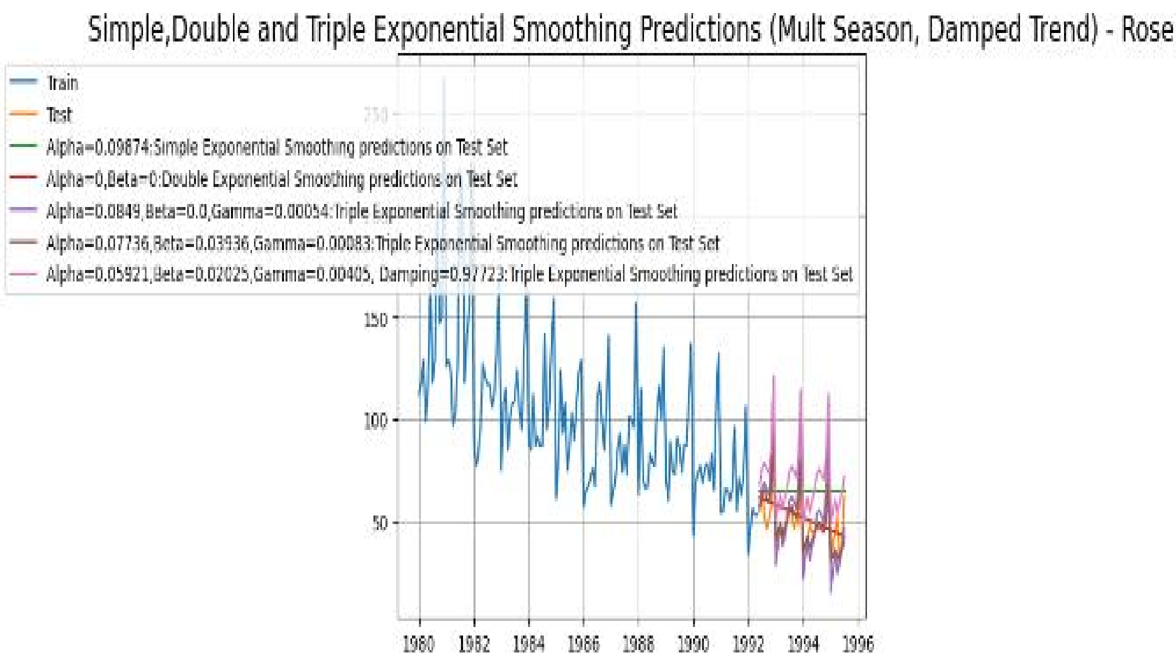


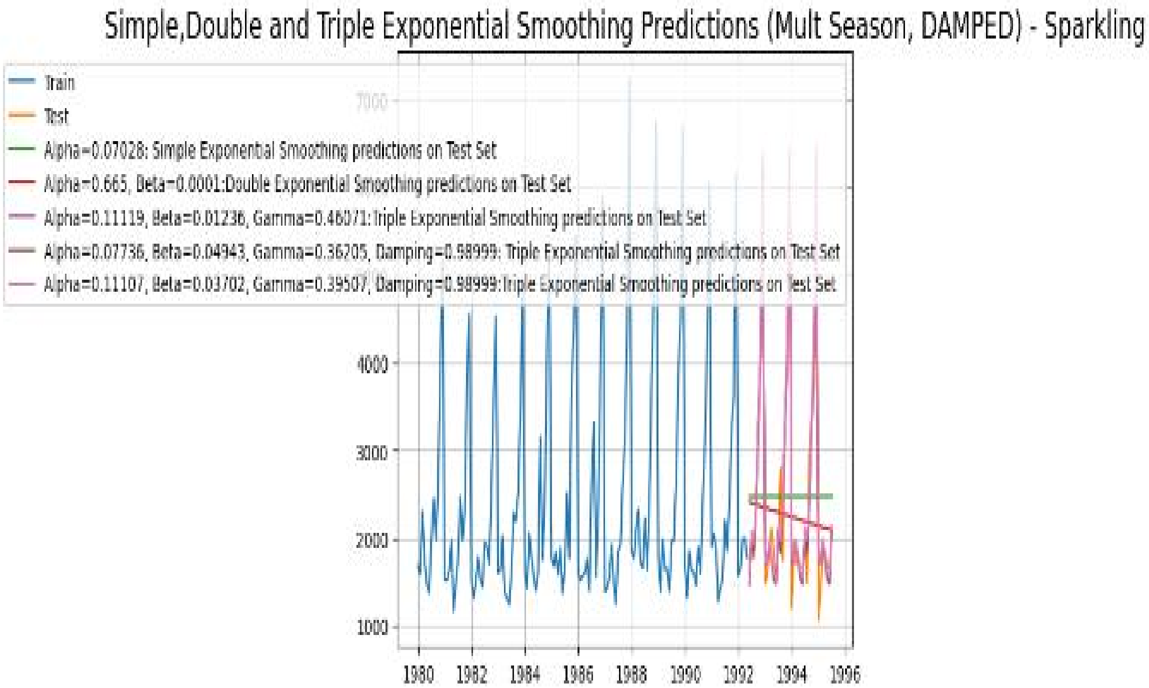


	Test RMSE Rose	Test RMSE Sparkling
Triple Exponential Smoothing (Additive Season, Damped Trend)	19.869175	304.269498

- **Rose** - Alpha= 0.07842, Beta = 0.01153, Gamma = 0.07738, Damping factor = 0.97503
- **Sparkling** - Alpha= 0.10062, Beta = 0.00018, Gamma = 0.51151, Damping factor= 0.97025

f) Triple Exponential Smoothing with Additive Errors, Additive DAMPED Trends, Multiplicative Seasonality - ETS(A, Ad, M):-





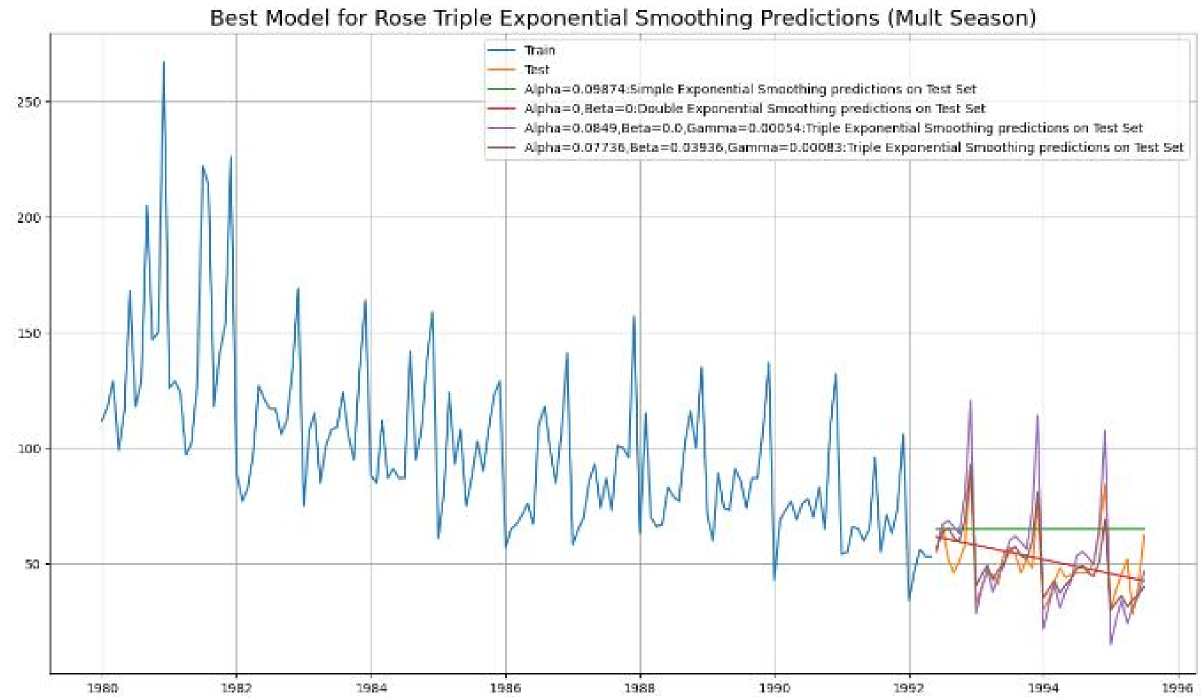
	Test RMSE Rose	Test RMSE Sparkling
Triple Exponential Smoothing (Multiplicative Season, Damped Trend)	21.959062	318.39316

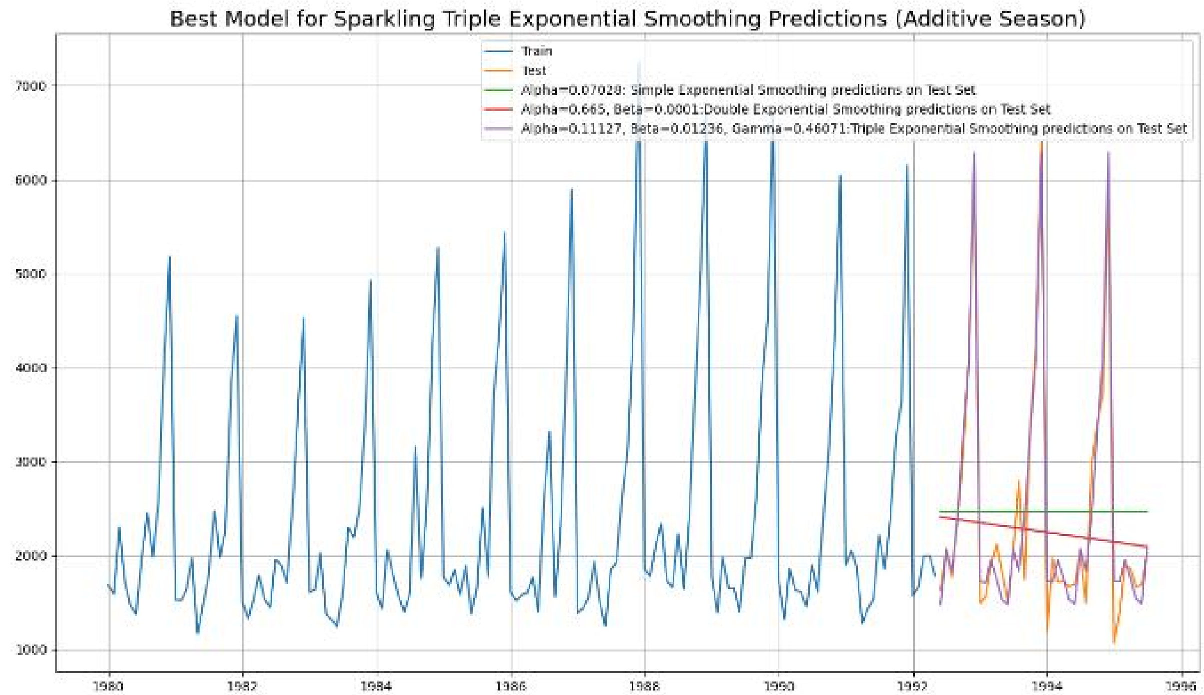
Comparing all the Models:-

	Test RMSE Rose	Test RMSE Sparkling
SimpleAverageModel	52.184392	2705.474937
2pointTrailingMovingAverage	11.529409	813.400684
4pointTrailingMovingAverage	14.448930	1156.589694
6pointTrailingMovingAverage	14.560046	1283.927428
9pointTrailingMovingAverage	14.724503	1346.278315
RegressionOnTime	17.475054	1349.042457
Simple Exponential Smoothing	20.263217	1329.402416
Double Exponential Smoothing	13.723188	1340.452773
Triple Exponential Smoothing (Additive Season)	13.814672	304.269498
Triple Exponential Smoothing (Multiplicative Season)	8.392115	318.448069
Triple Exponential Smoothing (Additive Season, Damped Trend)	19.869175	304.269498
Triple Exponential Smoothing (Multiplicative Season, Damped Trend)	21.959062	318.393160

Best model:-

- **Rose** — Triple Exponential Smoothing (Multiplicative Season)
- **Sparkling** — Triple Exponential Smoothing (Additive Season)





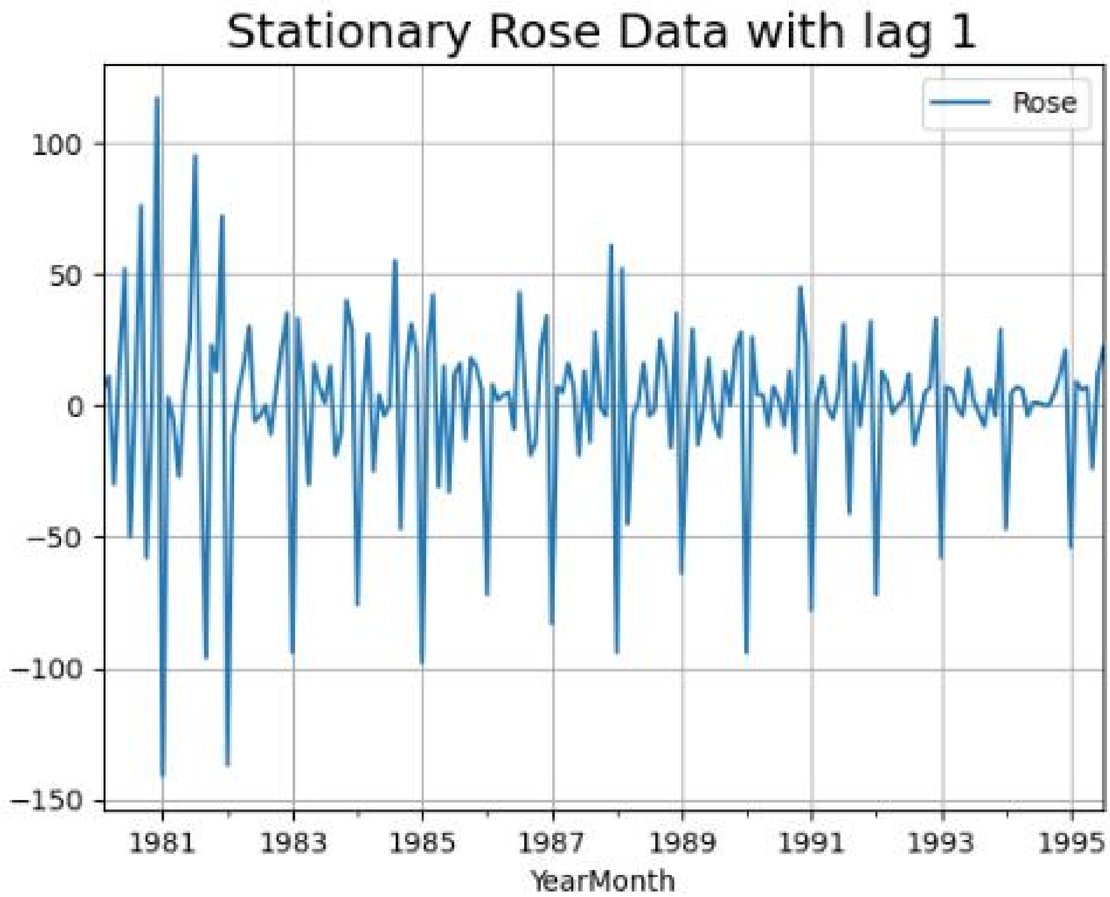
	Test RMSE Rose
Triple Exponential Smoothing (Multiplicative Season)	8.392115

	Test RMSE Sparkling
Triple Exponential Smoothing (Additive Season)	304.269498

D. Check for Stationarity:-

The hypothesis in a simple form for the ADF test is:

H0 : The Time Series has a unit root and is thus non-stationary.;
H1 : The Time Series does not have a unit root and is thus stationary.

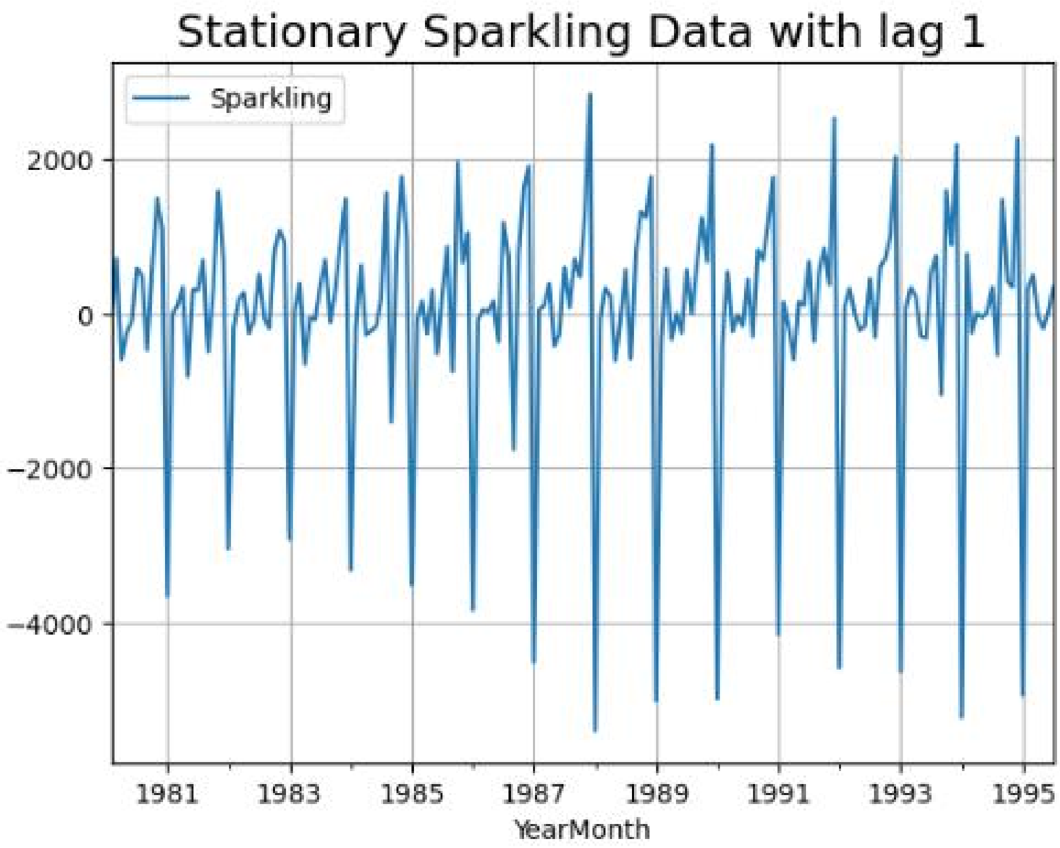



```
DF test statistic is -2.241
DF test p-value is 0.46692256831407475
Number of lags used 13
```

- 5% of significance level reveals its non-stationary.
- To solve this issue, we'll apply a single level of difference to determine, if the series becomes stationary.

```
DF test statistic is -44.912
DF test p-value is 0.0
Number of lags used 10
```

- p-value less than the significance level of 0.0, we reject the Null Hypothesis.
- It conclude that after applying a lag of 10, the Rose data becomes stationary.




```
DF test statistic is -1.798
DF test p-value is 0.7055958459932516
Number of lags used 12
```

- Significance level of 7%, the Time Series appears to be non-stationary.
- Let's apply one level of difference, if the series achieves stationary.

```
DF test statistic is -44.912
DF test p-value is 0.0
Number of lags used 10
```

- p-value less than the significance level 0.05, so we reject the null hypothesis.
- As a result, after applying a lag of 10, the Sparkling data becomes stationary.

• **Acc. to the Industry standard** , the Confidence Interval is 95%
• Alpha = 0.05; IF p-value < alpha :- Reject the Null Hypothesis and hence conclude that given Time Series is Stationarity
• ADF Test, IF p-value > alpha ===> We fail to reject the Null Hypothesis and hence conclude that given Time Series is Not Stationarity
• If Time Series is not Stationarity, then we can apply one level of difference and check
for Stationary again.
• Again, if the Time Series is still not Stationarity, then we again apply one more level of difference and check for Stationarity again
• Generally, Max. 2 levels of difference, Time Series becomes Stationarity
• If Time Series is Stationarity then we are ready to apply ARIMA / SARIMA Models.

E. Model Building - Stationary Data:-

Generate ACF & PACF Plot and find the AR, MA values. - Build different ARIMA models (Auto ARIMA - Manual ARIMA) - Build different SARIMA models (Auto SARIMA - Manual SARIMA) - Check the performance of the models built.

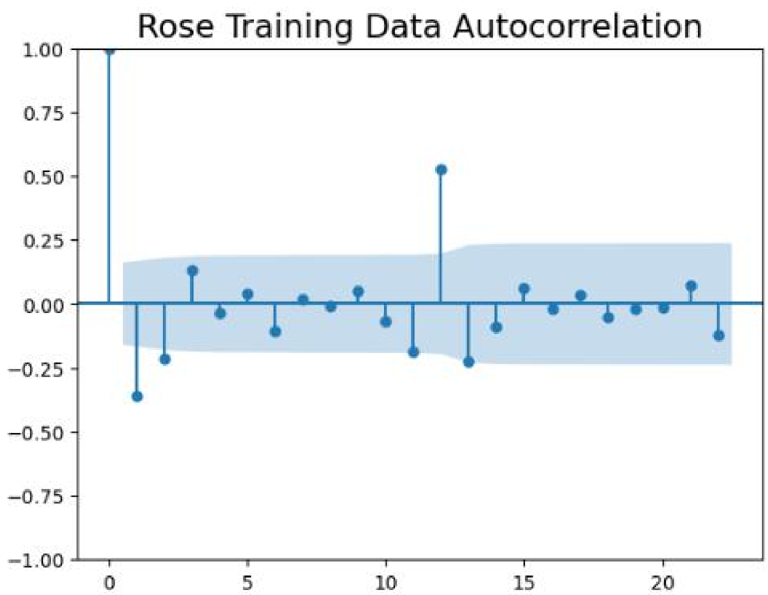
Auto-Correlation Function (ACF) is a statistical tool used to measure the correlation between a time series and its past values. It examines how each point in a time series relates to its previous points. The "auto" aspect of auto-correlation implies that it measures the correlation between a specific time instance and its preceding instances within the same time series. ACF is often visualized through a plot that displays correlations up to a certain lag unit, providing insights into the relationship between consecutive observations.

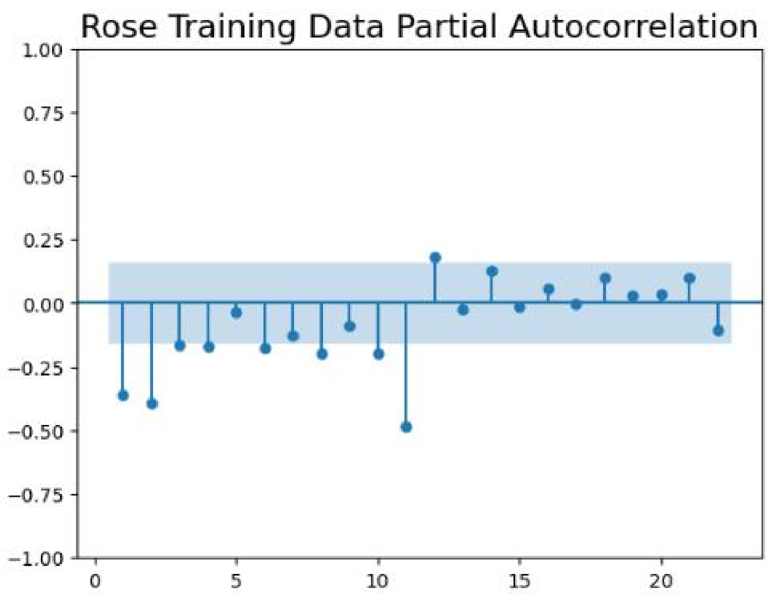
SARIMA (3, 1, 1) (3, 0, 0, 12) Diagnostic Plot - SPARKLING Test RMSE Sparkling Test MAPE Sparkling ARIMA(2,1,2) 1299.98 47.10 SARIMA (3, 1, 1) (3, 0, 0, 12) 601.24 25.87 Time Series Project ACF analysis helps determine the moving average parameter 'q' in ARIMA or SARIMA models.

Partial Auto-Correlation Function (PACF) measures the correlation between a time series and its lagged values, excluding the intermediate instances. For instance, if the lag is denoted as 'k,' PACF computes the correlation between the current value and the value 'k' time units ago, disregarding the impact of observations between them. PACF is represented through a plot that illustrates correlations among lag points. It assists in determining the auto-regressive parameter 'p' in ARIMA or SARIMA models.

E.1 Generate ACF & PACF plot and find the AR, MA Values:-

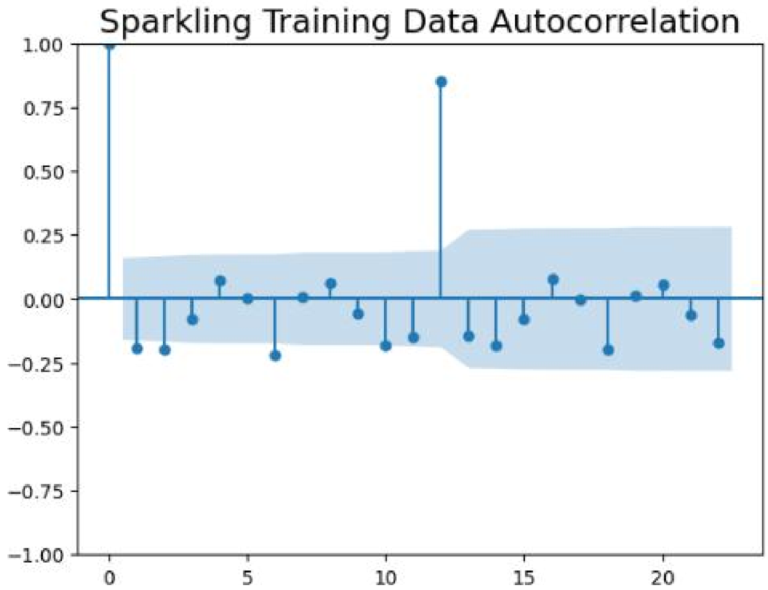
ACF & PACF of Rose:-

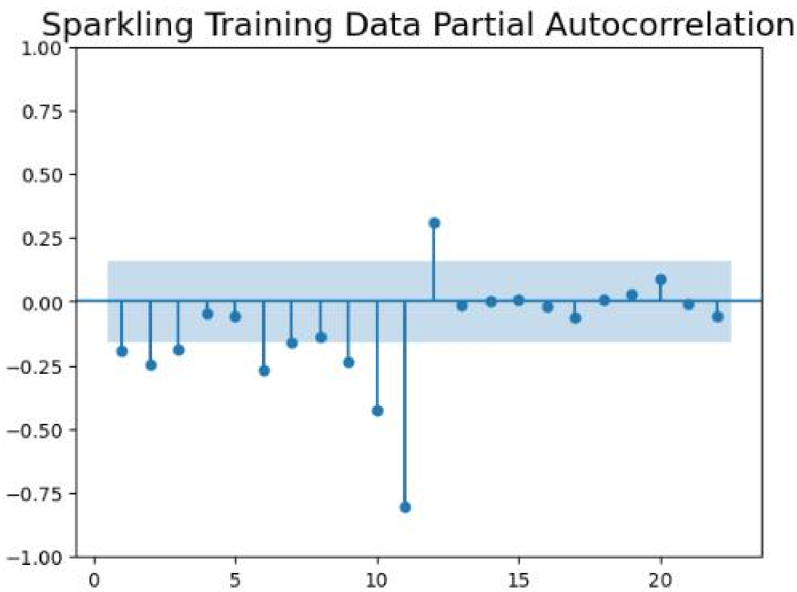




The significance level of 0.05 and analyzing the characteristics of the PACF and ACF plots, we select the Auto-Regressive (AR) parameter 'p' as 2 and the Moving-Average (MA) parameter 'q' also as 2. This decision is guided by identifying significant lags in both plots before they cut off. The significant lag in the PACF plot before it terminates informs the choice of 'p', while the significant lag in the ACF plot before it cuts off guides the selection of 'q'. These parameter values are crucial for constructing ARIMA or SARIMA models, providing insights into the temporal dependencies within the time series data.

ACF & PACF of Sparkling:-





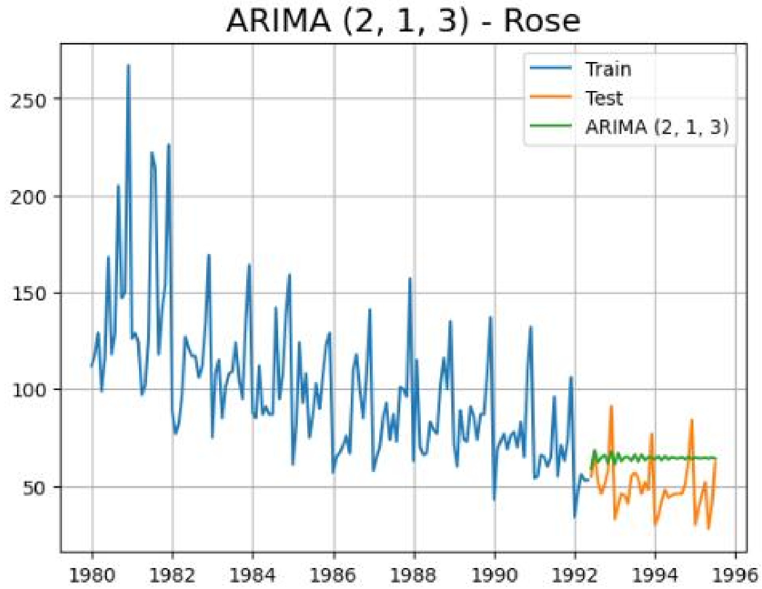
For ARIMA:
Auto-Regressive (AR) parameter (p) = 0
Moving-Average (MA) parameter (q) = 0
Differencing parameter (d) = 1

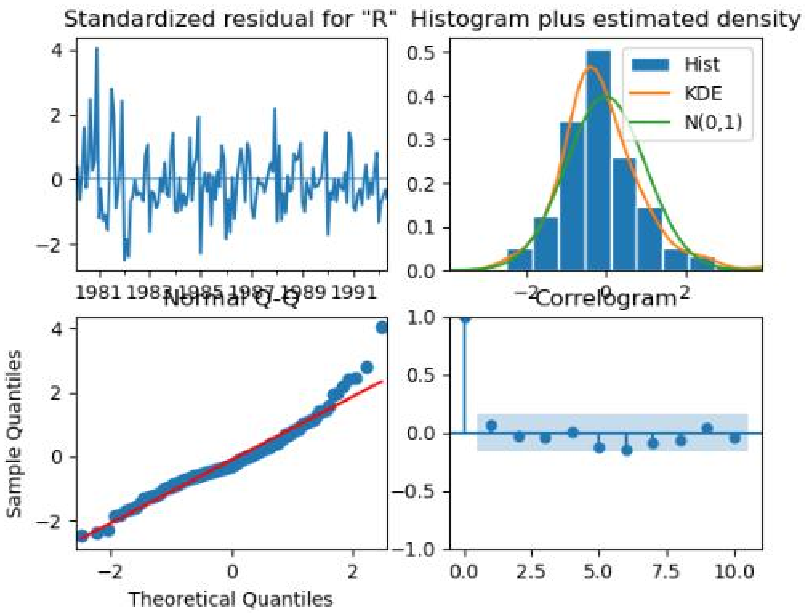
For SARIMA:
Auto-Regressive (AR) parameter (p) = 0
Moving-Average (MA) parameter (q) = 0
Differencing parameter (d) = 1
Seasonal Auto-Regressive (SAR) parameters (P) = 0, 1, 2, 3
Seasonal Moving-Average (SMA) parameters (Q) = 1, 2, 3
Seasonal differencing parameter (D) = 0

E.2 Build different ARIMA models (Auto ARIMA , Manual ARIMA) + E.3 Build different SARIMA models (Auto SARIMA , Manual SARIMA):-

ROSE (MANUAL+AUTO):

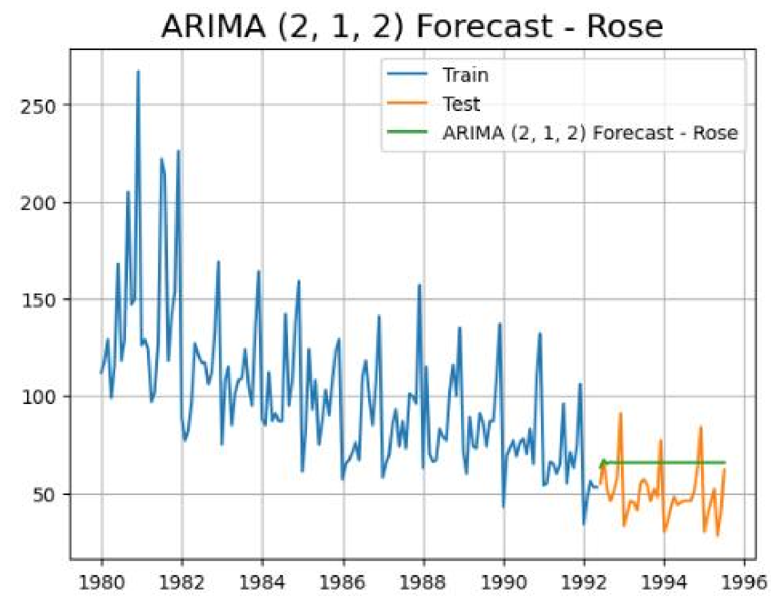
1. ARIMA Auto- Rose (2,1,3):-

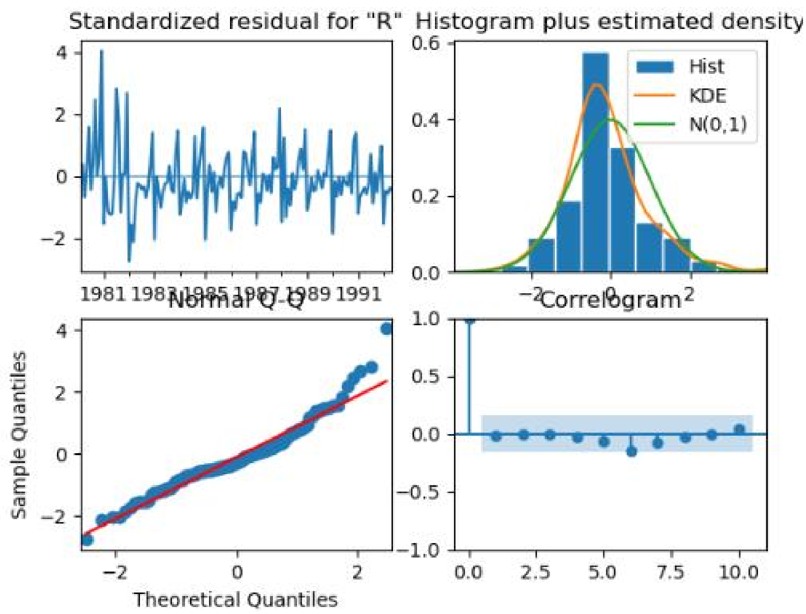




	Test RMSE Rose	Test MAPE Rose
ARIMA(2,1,3)	19.678446	41.475888

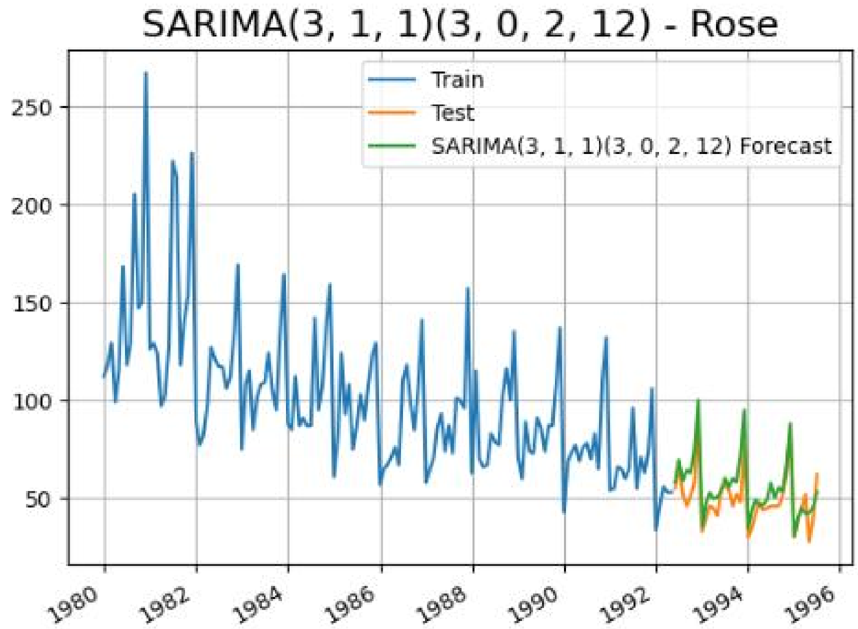
2. ARIMA Manual- Rose (2,1,2):-

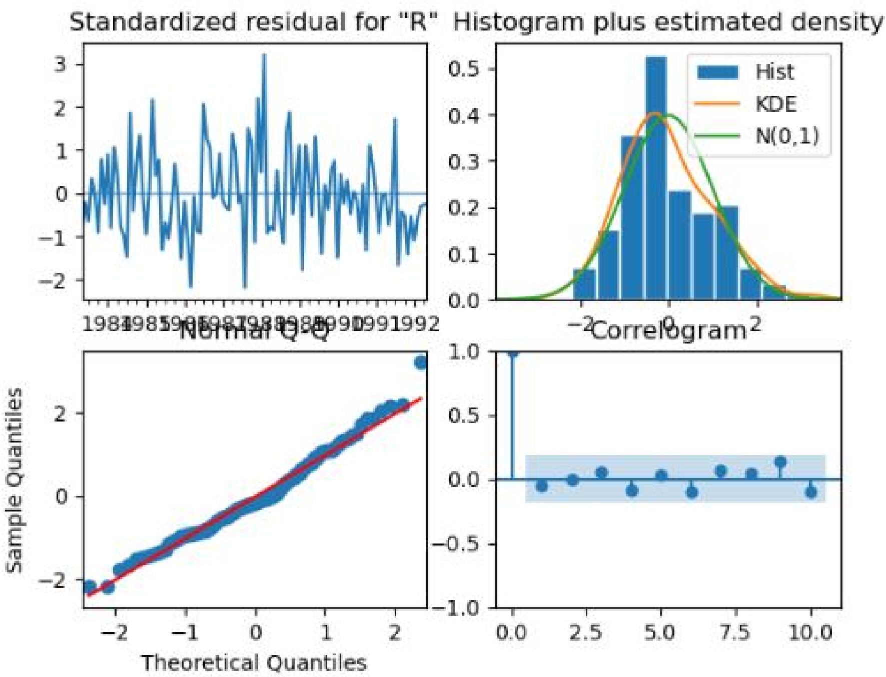




	Test RMSE Rose	Test MAPE Rose
ARIMA(2,1,3)	19.678446	41.475888
ARIMA(2,1,2)	20.756085	44.227060

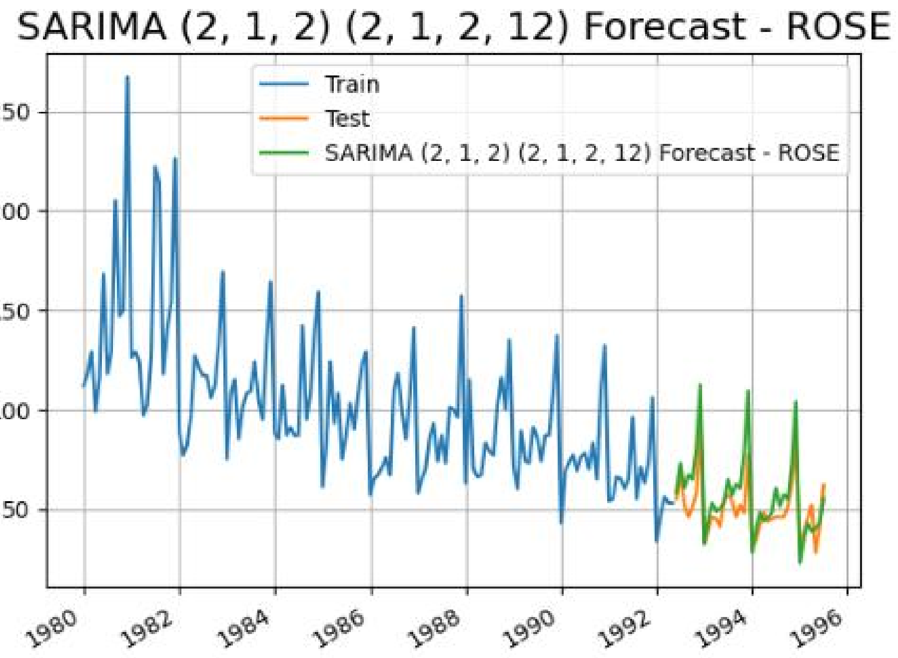
3. SARIMA Auto- Rose (3,1,1):-

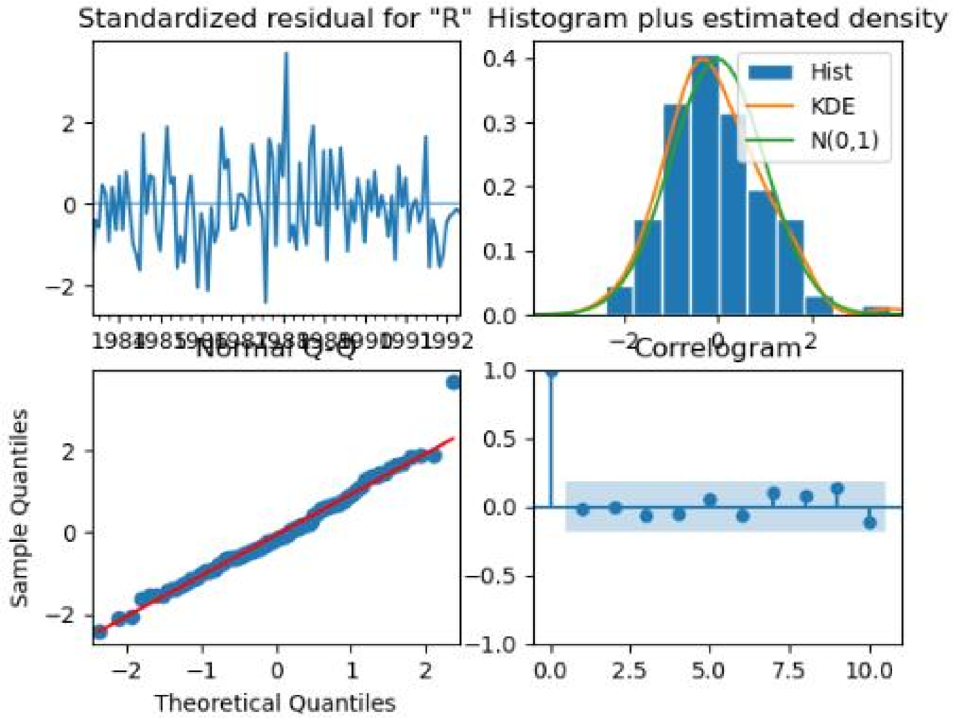




	Test RMSE Rose	Test MAPE Rose
ARIMA(2,1,3)	19.678446	41.475888
ARIMA(2,1,2)	20.756085	44.227060
SARIMA(3, 1, 1)(3, 0, 2, 12)	9.097917	15.202670

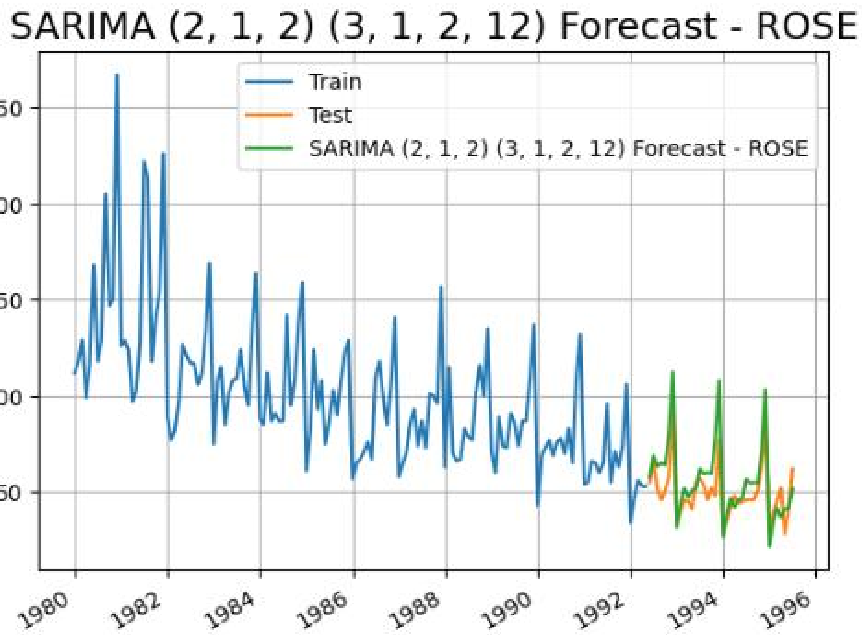
4. SARIMA Manual- Rose (2,1,2)(2,1,2,12):-

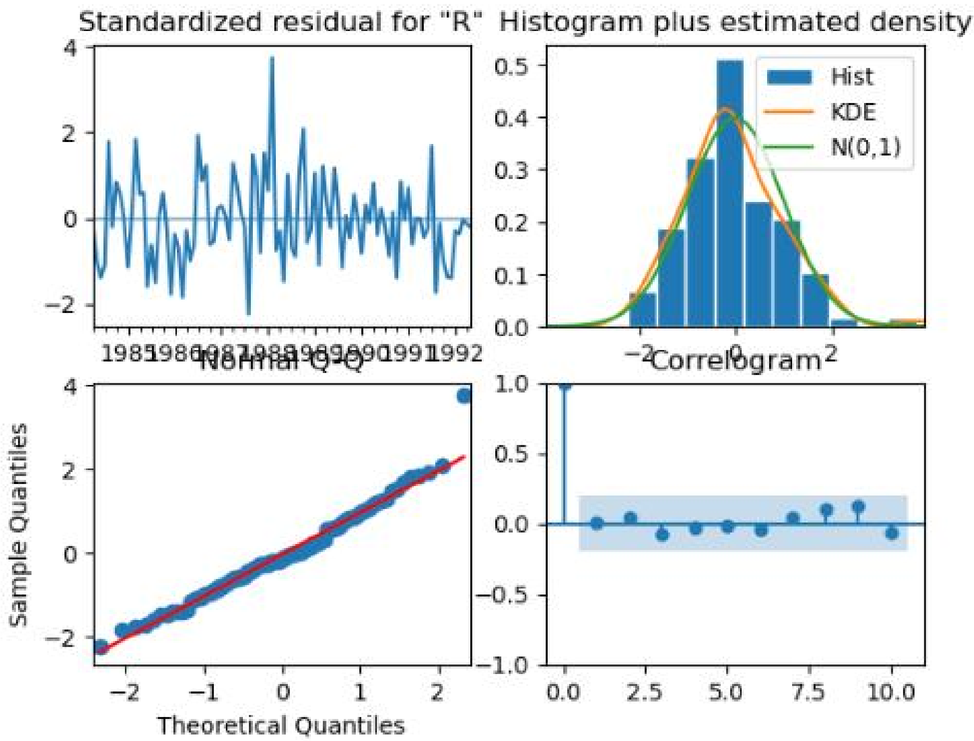




RMSE: 12.166172146045588
MAPE: 18.34290421559828

5. SARIMA Manual- Rose (2,1,2)(3,1,2,12):-



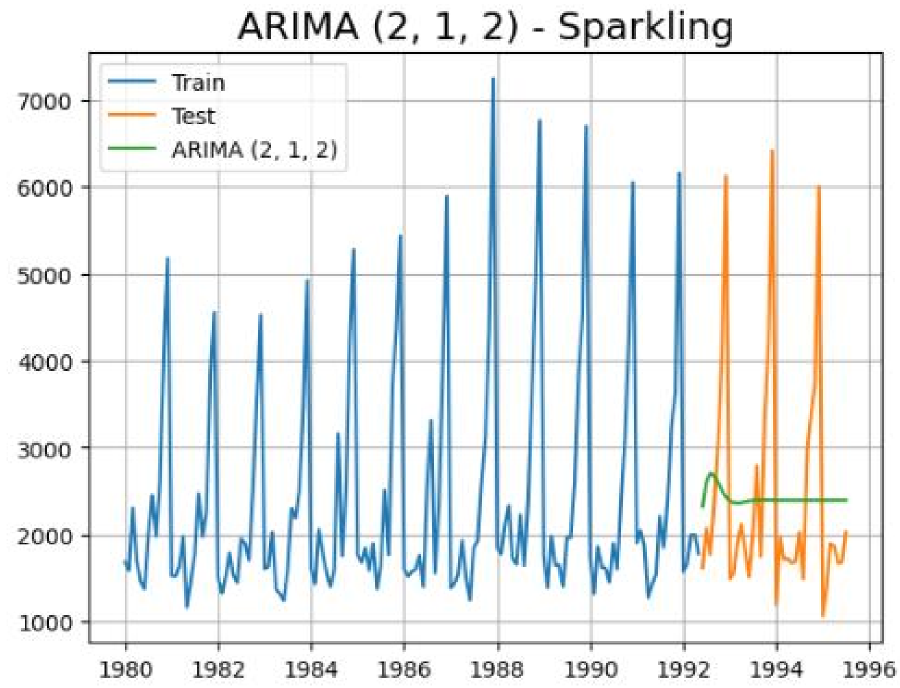


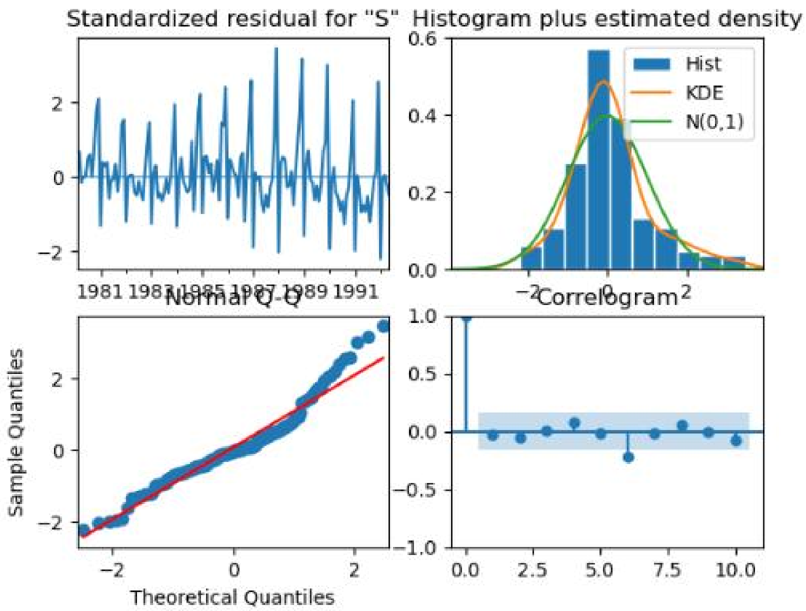
RMSE: 11.958335391994343

MAPE: 18.25040246245028

SPARKLING(MANUAL):

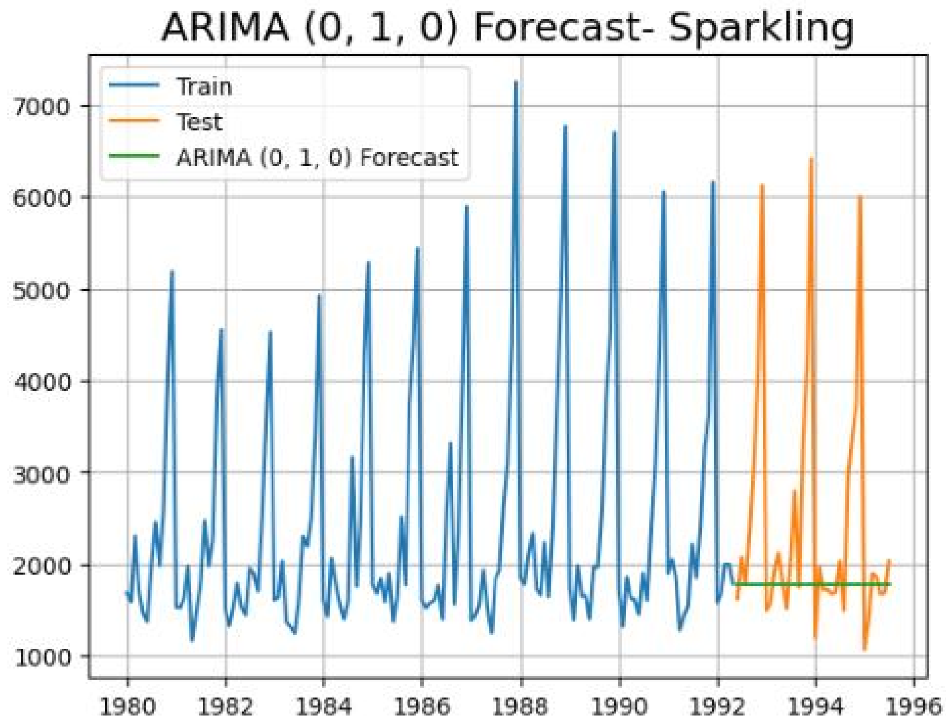
1. ARIMA Auto- Sparkling (2,1,2):-

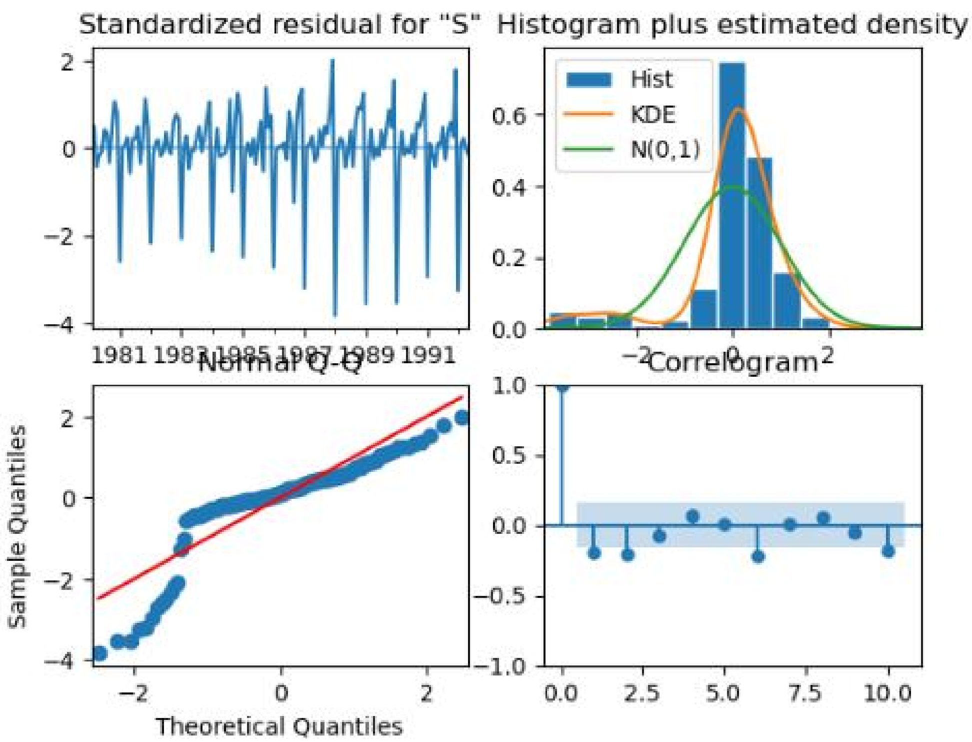




	RMSE	MAPE
ARIMA(2,1,2)	1326.623799	40.73652

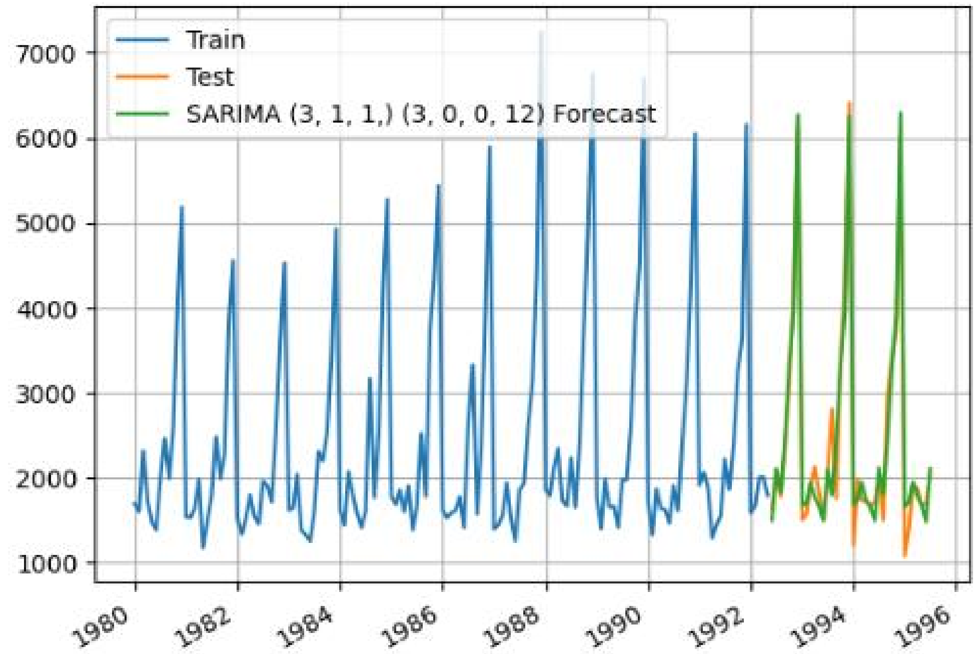
2. ARIMA Manual- Sparkling (0,1,0):-

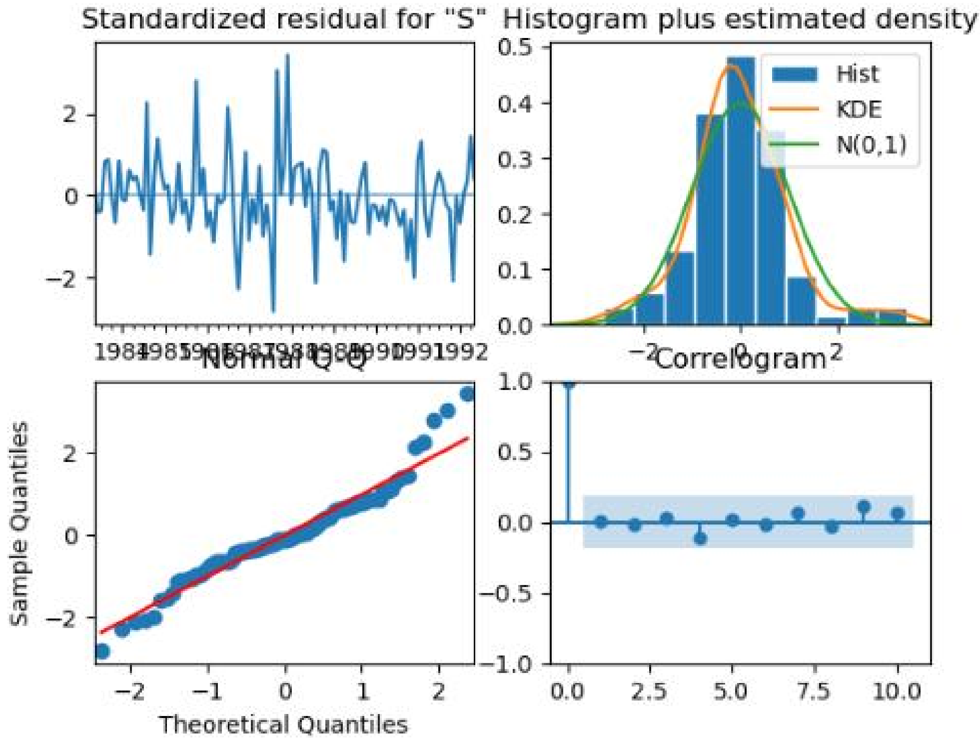




	RMSE	MAPE
ARIMA(2,1,2)	1326.623799	40.736520
ARIMA(0,1,0)	1490.207448	25.510451

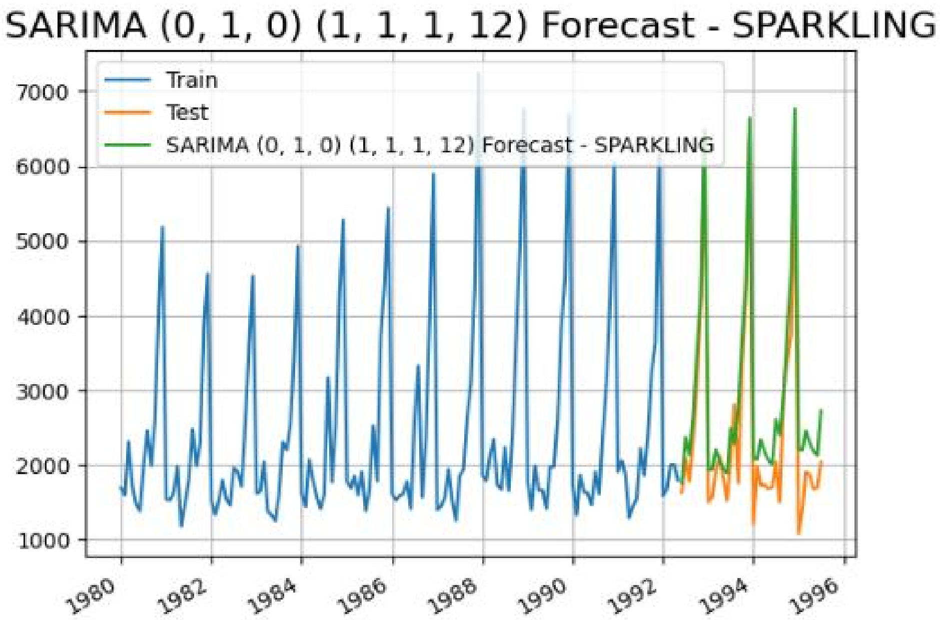
3. SARIMA Auto- Sparkling (3,1,1)(3,0,0,12):-

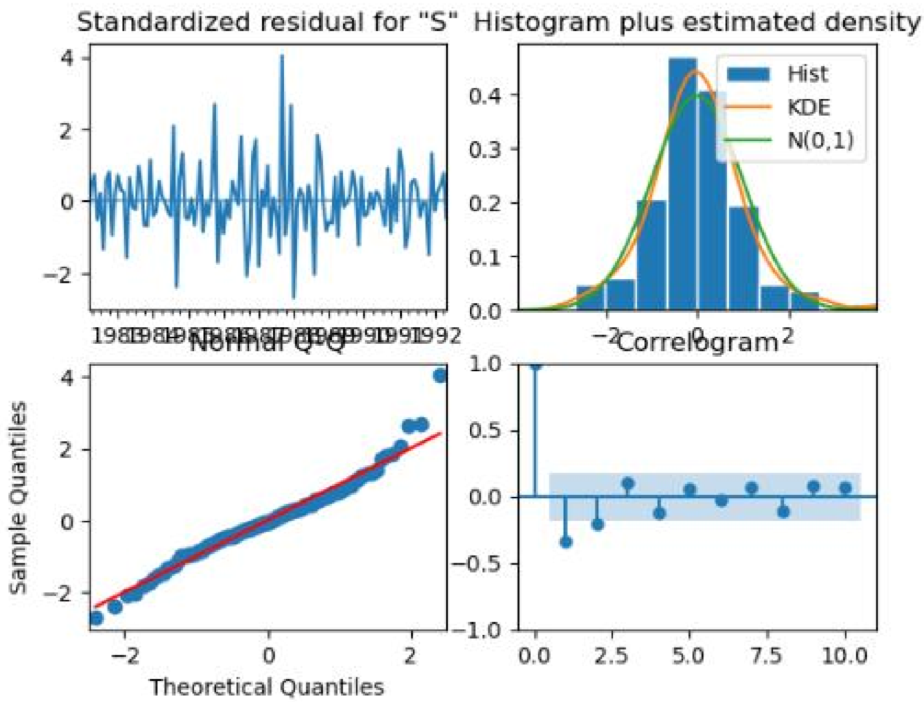




	RMSE	MAPE
ARIMA(2,1,2)	1326.623799	40.736520
ARIMA(0,1,0)	1490.207448	25.510451
SARIMA(3,1,1)(3,0,2,12)	294.687559	10.510252

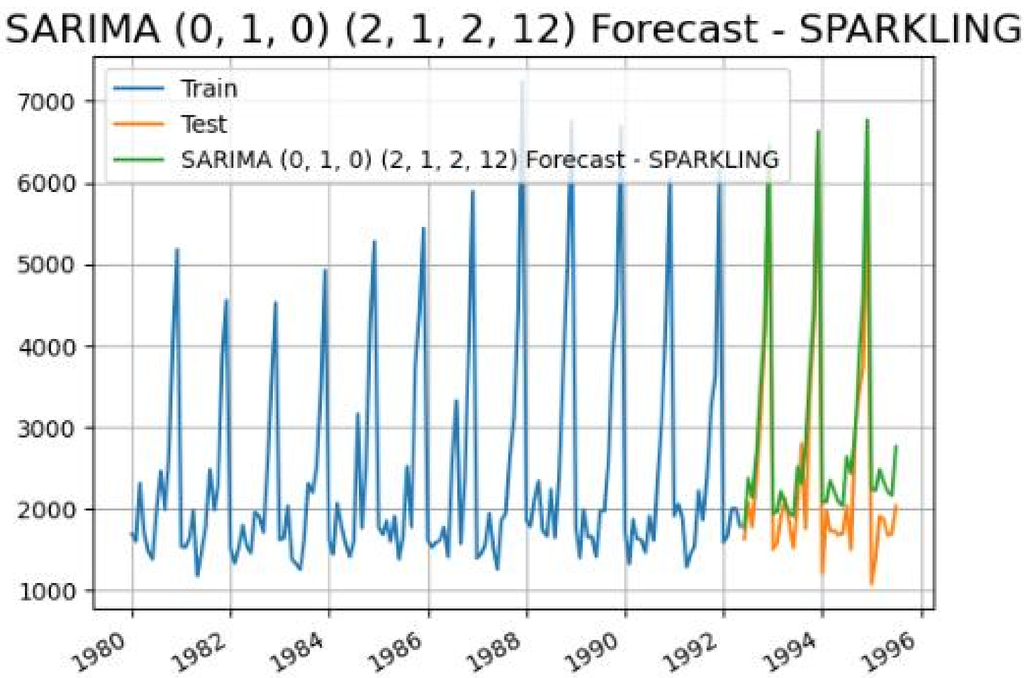
4. SARIMA Manual- Sparkling (0,1,0)(1,1,1,12):-

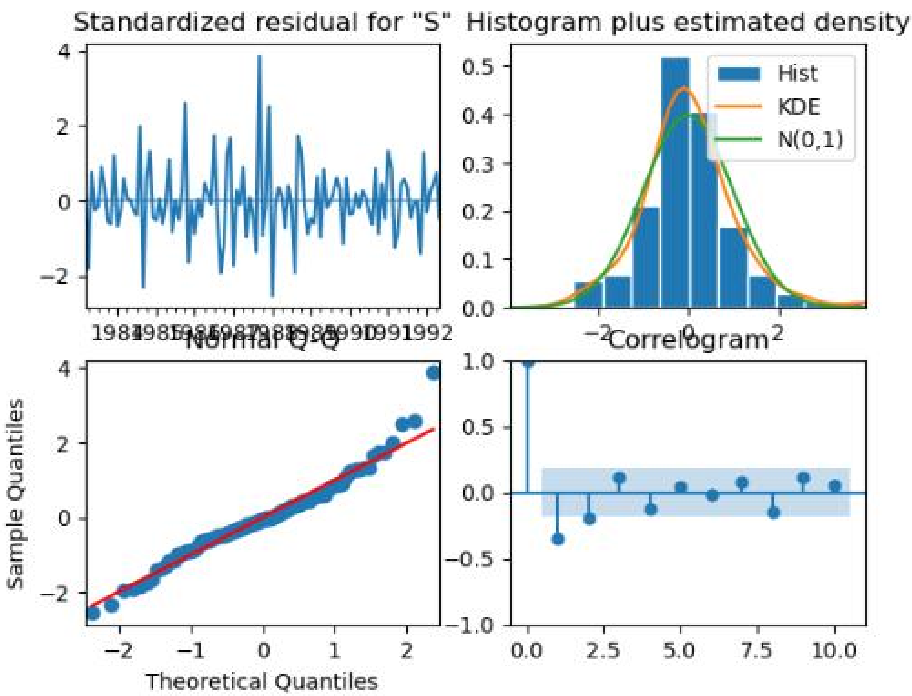




RMSE: 526.1637123348268
MAPE: 24.241683473179016

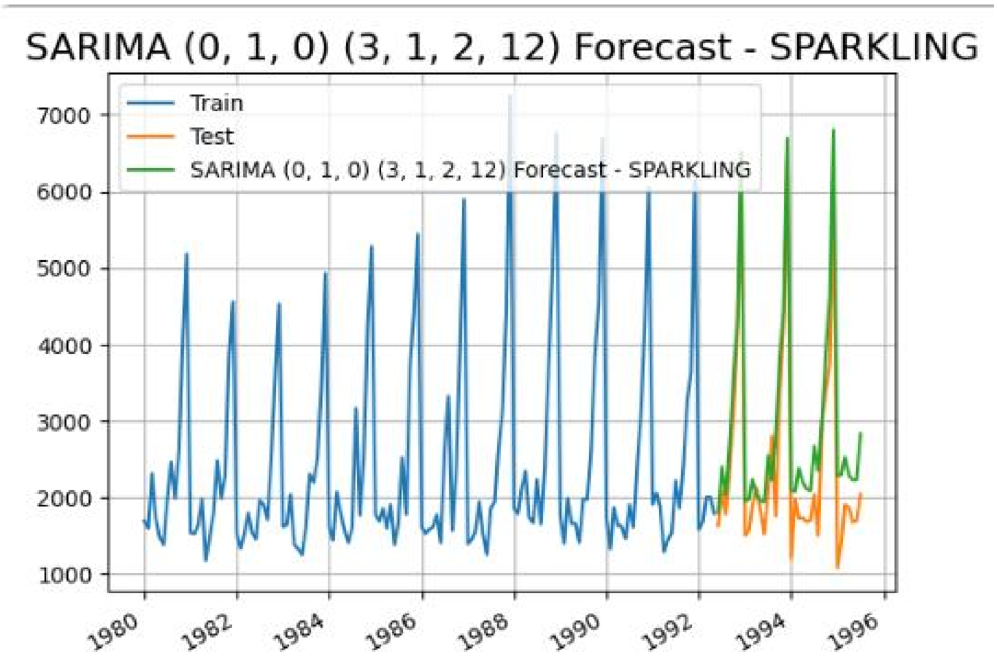
5. SARIMA Manual- Sparkling (0,1,0)(2,1,1,12):-

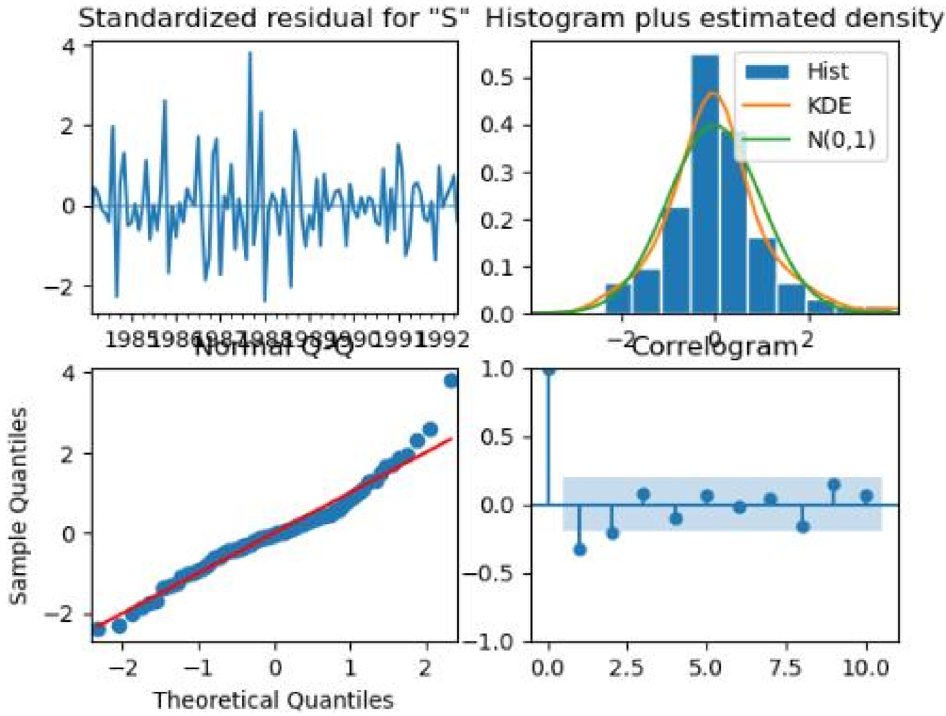




RMSE: 538.4523638412815
MAPE: 24.935115778469214

6. SARIMA Manual- Sparkling (0,1,0)(3,1,1,12):-





	RMSE	MAPE
ARIMA(2,1,2)	1326.623799	40.736520
ARIMA(0,1,0)	1490.207448	25.510451
SARIMA(3,1,1)(3,0,2,12)	294.687559	10.510252
SARIMA(0,1,0)(3,1,2,12)	526.163712	24.241683
SARIMA(0,1,0)(2,1,2,12)	538.452364	24.935116
SARIMA(0,1,0)(3,1,2,12)	565.521029	26.131922

ACCORDING TO THE DATA OF ARIMA / SARIMA

BEST MODEL FOR ROSE:- SARIMA(3,1,1)(3,0,2,12)
BEST MODEL FOR SPARKLING:- SARIMA (3,1,1)(3,0,2,12)

F. Compare the performance of the models:-

Compare the performance of all the models built - Choose the best model with proper rationale - Rebuild the best model using the entire data - Make a forecast for the next 12 months

F.1 Compare the performance of all the models built:-

	Test RMSE Rose	Test RMSE Sparkling	RMSE	MAPE
SimpleAverageModel	52.184392	2705.474937	NaN	NaN
2pointTrailingMovingAverage	11.529409	813.400684	NaN	NaN
4pointTrailingMovingAverage	14.448930	1156.589694	NaN	NaN
6pointTrailingMovingAverage	14.560046	1283.927428	NaN	NaN
9pointTrailingMovingAverage	14.724503	1346.278315	NaN	NaN
RegressionOnTime	17.475054	1349.042457	NaN	NaN
Simple Exponential Smoothing	20.263217	1329.402416	NaN	NaN
Double Exponential Smoothing	13.723188	1340.452773	NaN	NaN
Triple Exponential Smoothing (Additive Season)	13.814672	304.269498	NaN	NaN
Triple Exponential Smoothing (Multiplicative Season)	8.392115	318.448069	NaN	NaN
Triple Exponential Smoothing (Additive Season, Damped Trend)	19.869175	304.269498	NaN	NaN
Triple Exponential Smoothing (Multiplicative Season, Damped Trend)	21.959062	318.393160	NaN	NaN
ARIMA(2,1,2)	NaN	NaN	1326.623799	40.736520
ARIMA(0,1,0)	NaN	NaN	1490.207448	25.510451
SARIMA(3,1,1)(3,0,2,12)	NaN	NaN	294.687559	10.510252
SARIMA(0,1,0)(3,1,2,12)	NaN	NaN	526.163712	24.241683
SARIMA(0,1,0)(2,1,2,12)	NaN	NaN	538.452364	24.935116
SARIMA(0,1,0)(3,1,2,12)	NaN	NaN	565.521029	26.131922

F.2 Choose the best model with proper rationale:-

Best model:-

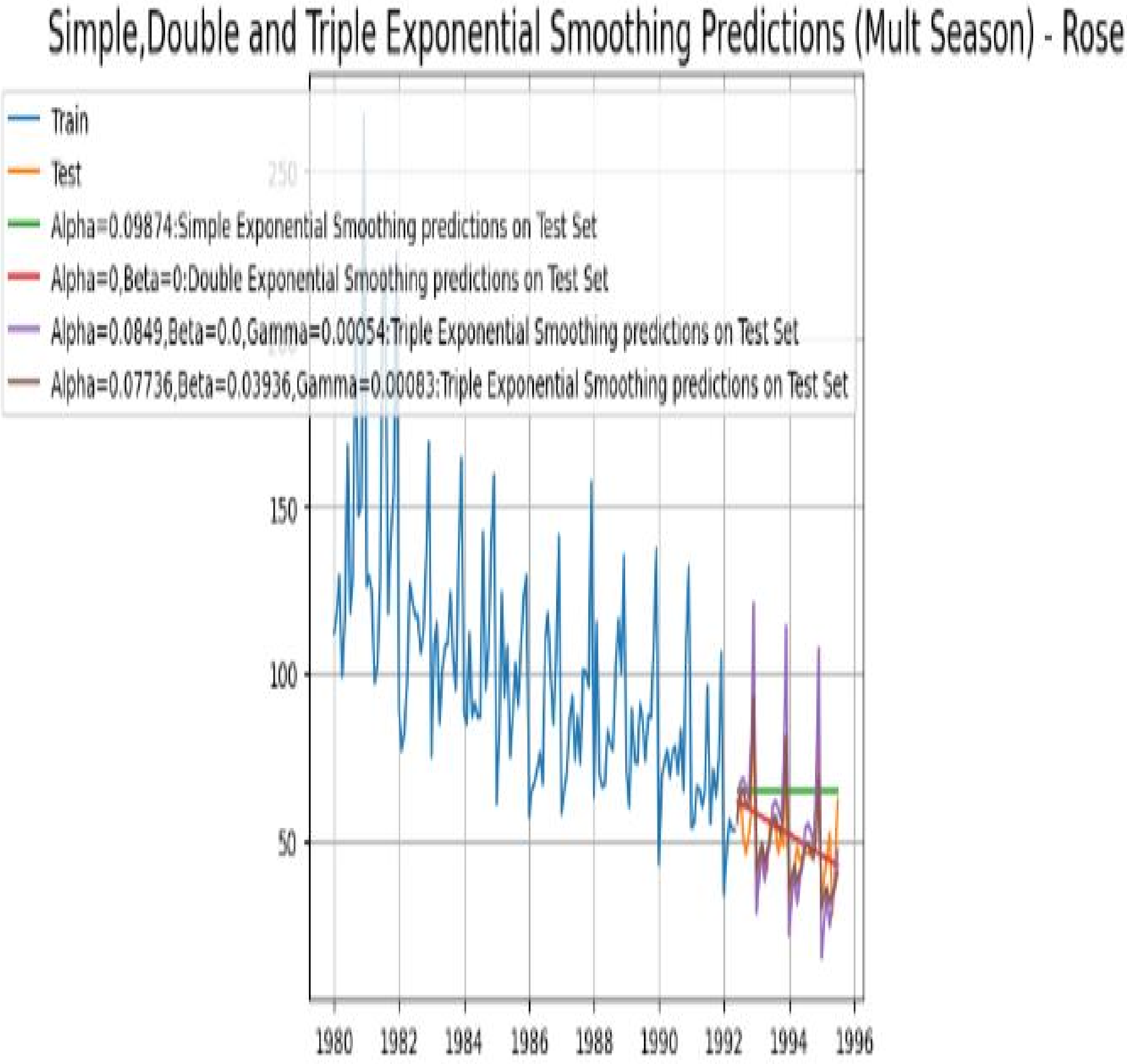
- **Rose** — Triple Exponential Smoothing (Multiplicative Season)
- **Sparkling** — Triple Exponential Smoothing (Additive Season)

ACCORDING TO THE DATA OF ARIMA / SARIMA:-

- **ROSE:- SARIMA(3,1,1)(3,0,2,12)**
- **SPARKLING:- SARIMA (3,1,1)(3,0,2,12)**

F.3 Rebuild the best model using the entire data:-

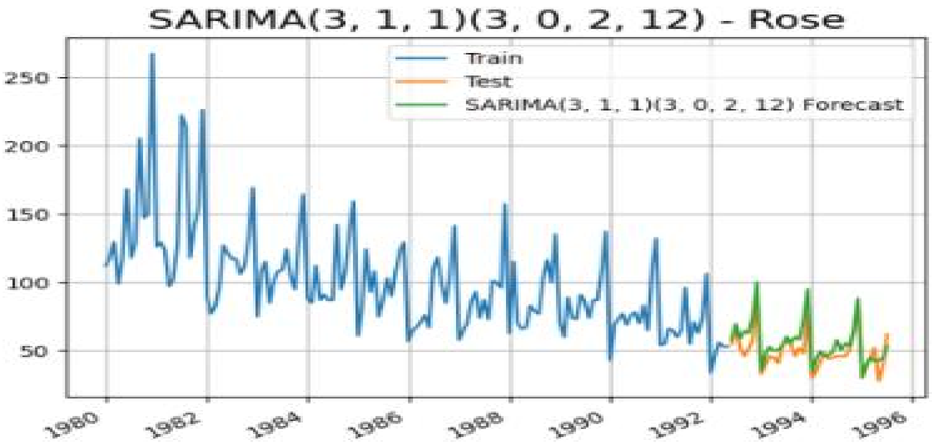
FOR ROSE:- Triple Exponential Smoothing (Multiplicative Season)

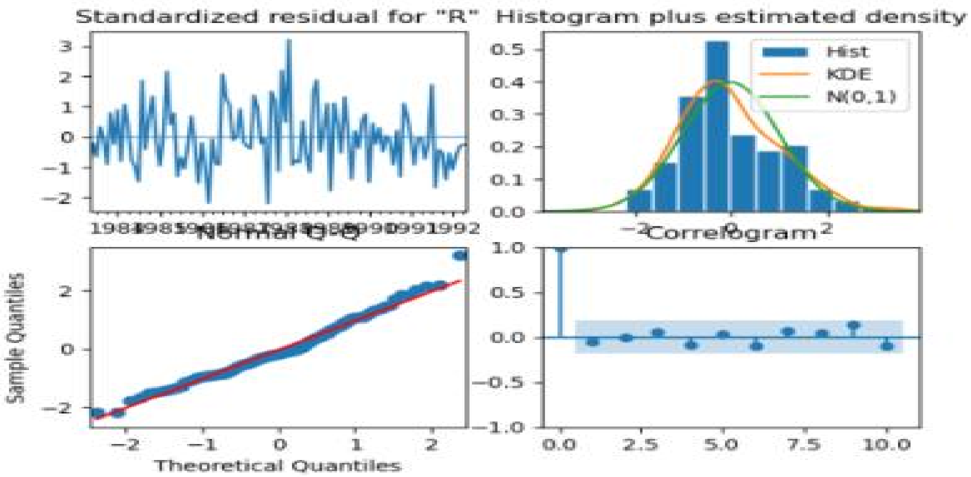


• **Rose** - Alpha = 0.07736, Beta = 0.03936, Gamma = 0.00083

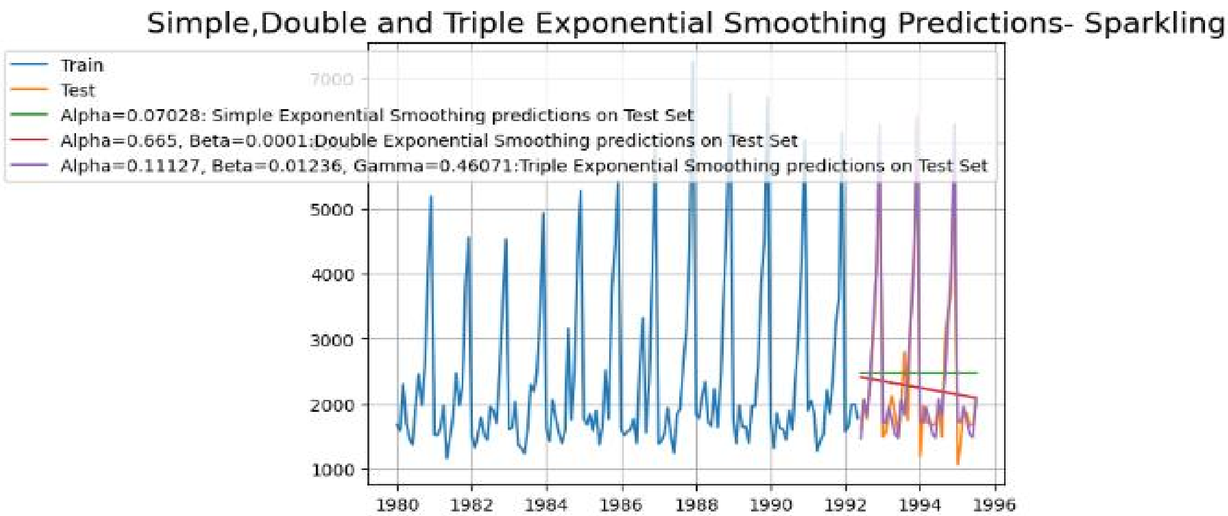
ACCORDING TO THE DATA OF ARIMA / SARIMA:-

- **ROSE:- SARIMA(3,1,1)(3,0,2,12)**





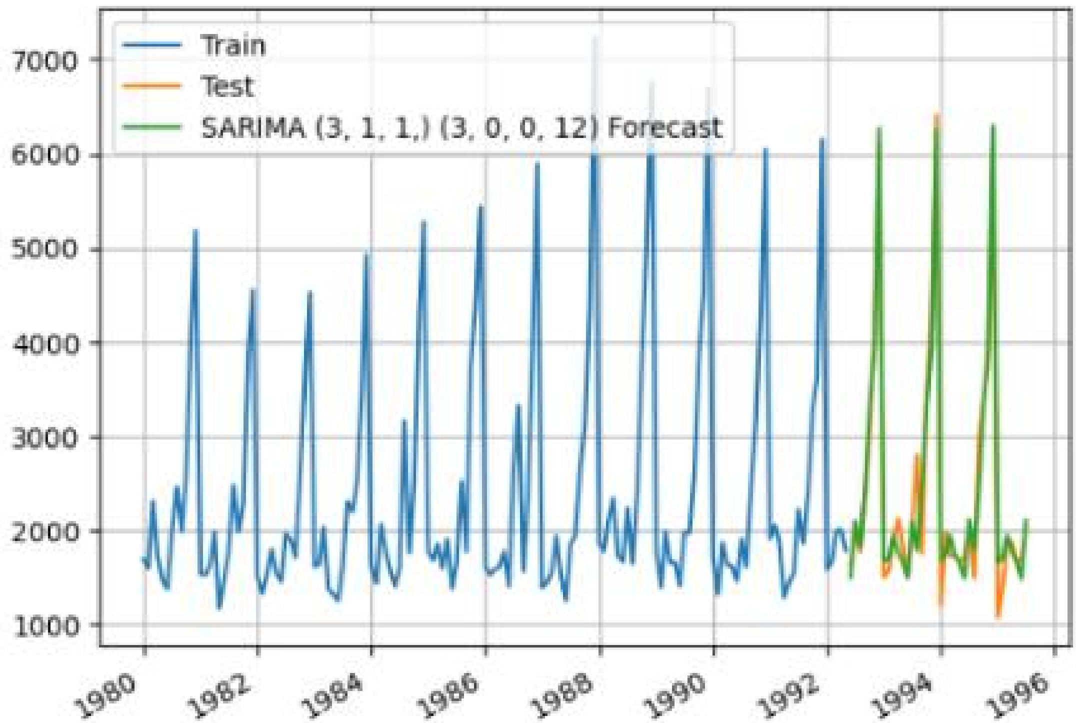
FOR SPARKLING:- Triple Exponential Smoothing (Additive Season)

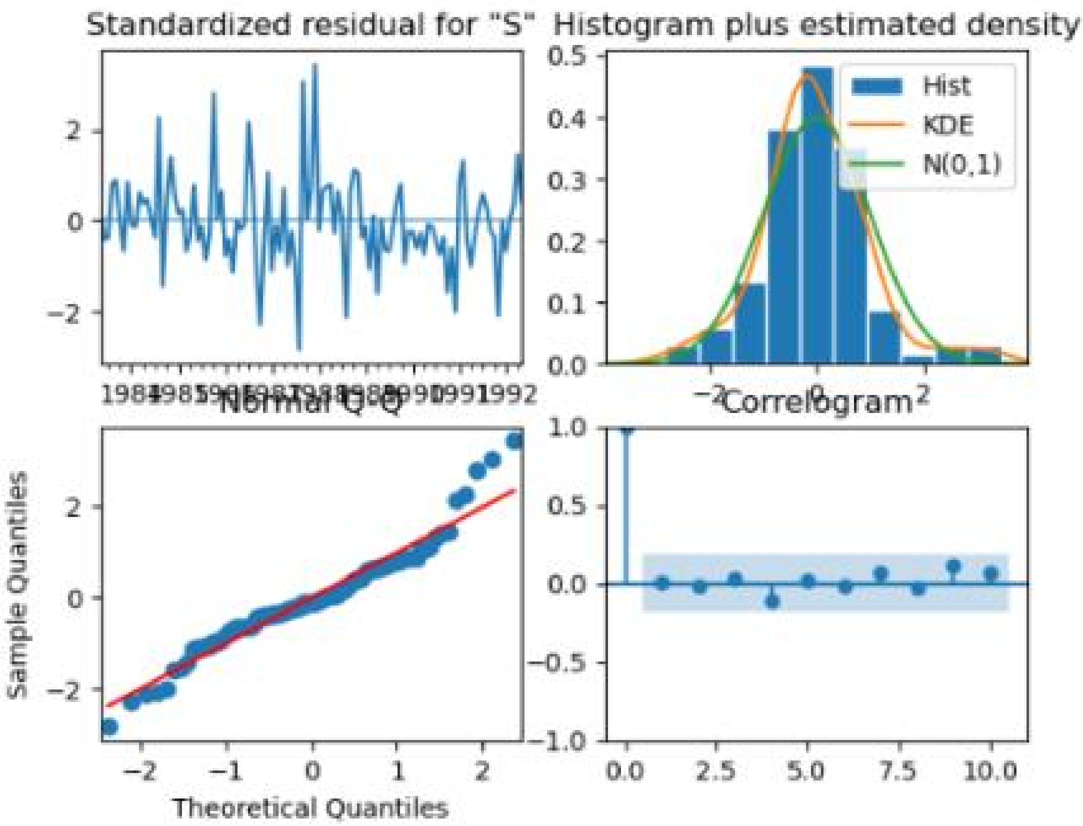


- **Sparkling** - Alpha = 0.11127 ; Beta = 0.01236 ; Gamma = 0.46071

ACCORDING TO THE DATA OF ARIMA / SARIMA:-

- **SPARKLING:- SARIMA(3,1,1)(3,0,2,12)**

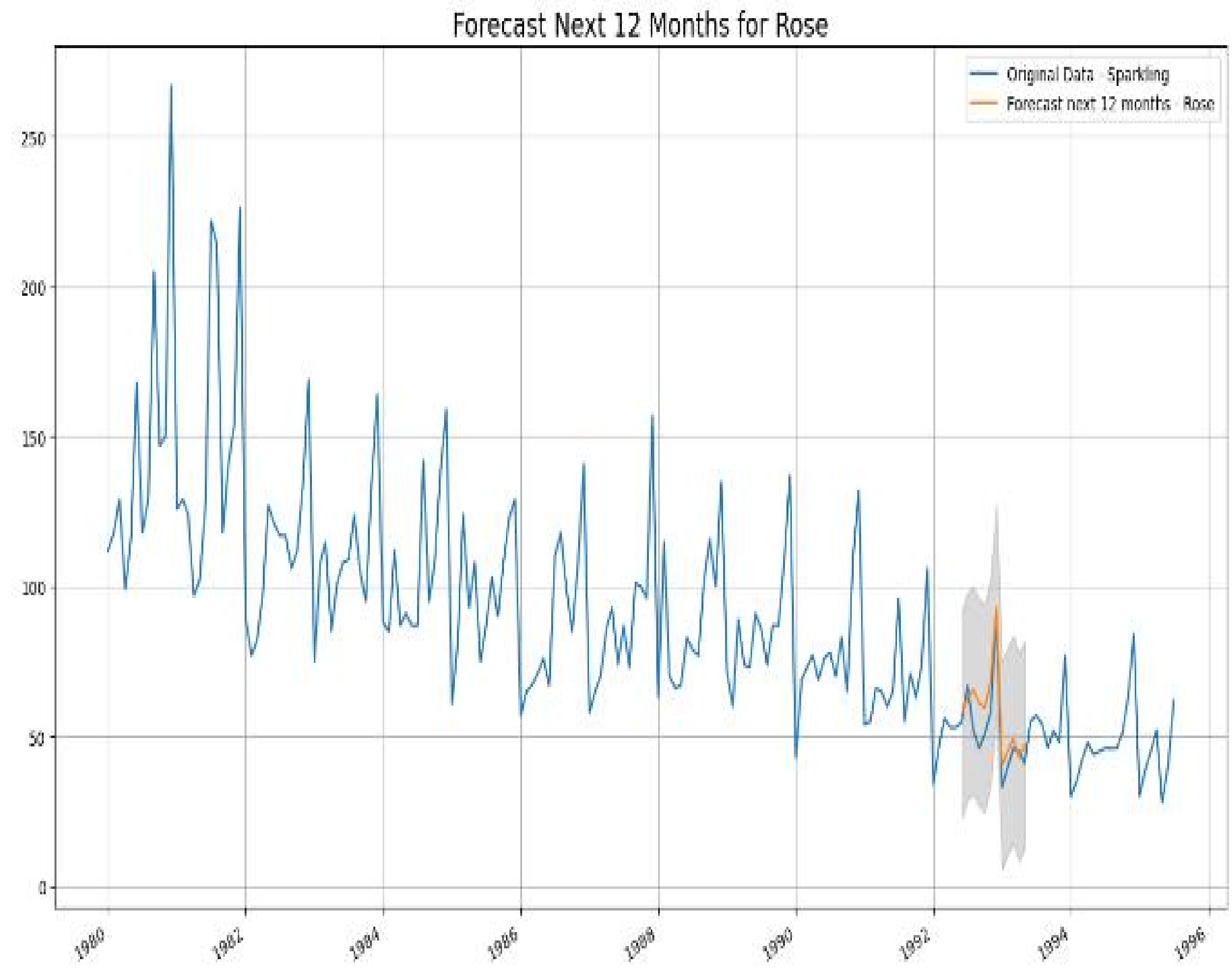




F.4 Make a forecast for the next 12 months:-

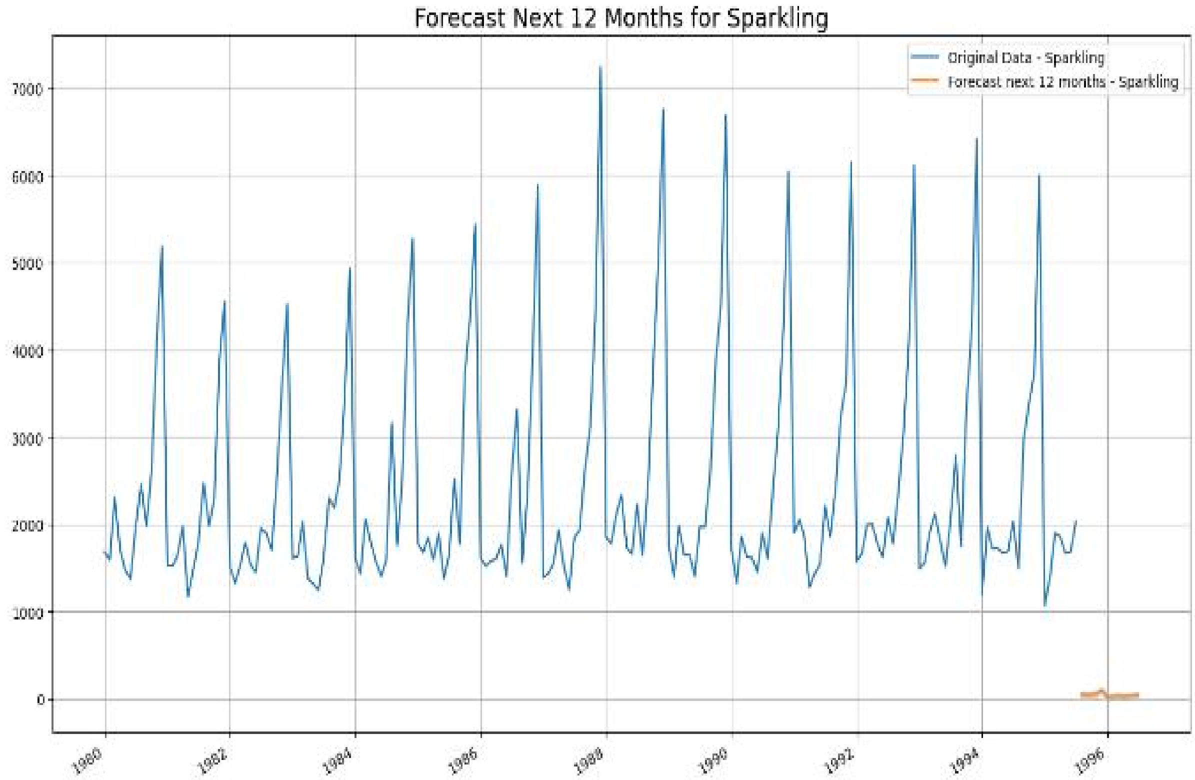
- **Rose** — Triple Exponential Smoothing (Multiplicative Season)

```
1992-06-01    57.261352
1992-07-01    63.262227
1992-08-01    65.404902
1992-09-01    61.240479
1992-10-01    59.431138
1992-11-01    68.471884
1992-12-01    92.957663
1993-01-01    40.386411
1993-02-01    45.204826
1993-03-01    49.182585
1993-04-01    42.966883
1993-05-01    47.213414
Freq: MS, dtype: float64
```



- **Sparkling** — Triple Exponential Smoothing (Additive Season)

```
1995-08-01    50.097242
1995-09-01    46.897041
1995-10-01    45.635882
1995-11-01    60.244340
1995-12-01    98.468088
1996-01-01    14.018996
1996-02-01    24.307667
1996-03-01    31.849735
1996-04-01    24.654896
1996-05-01    28.034295
1996-06-01    33.536602
1996-07-01    44.225345
Freq: MS, dtype: float64
```

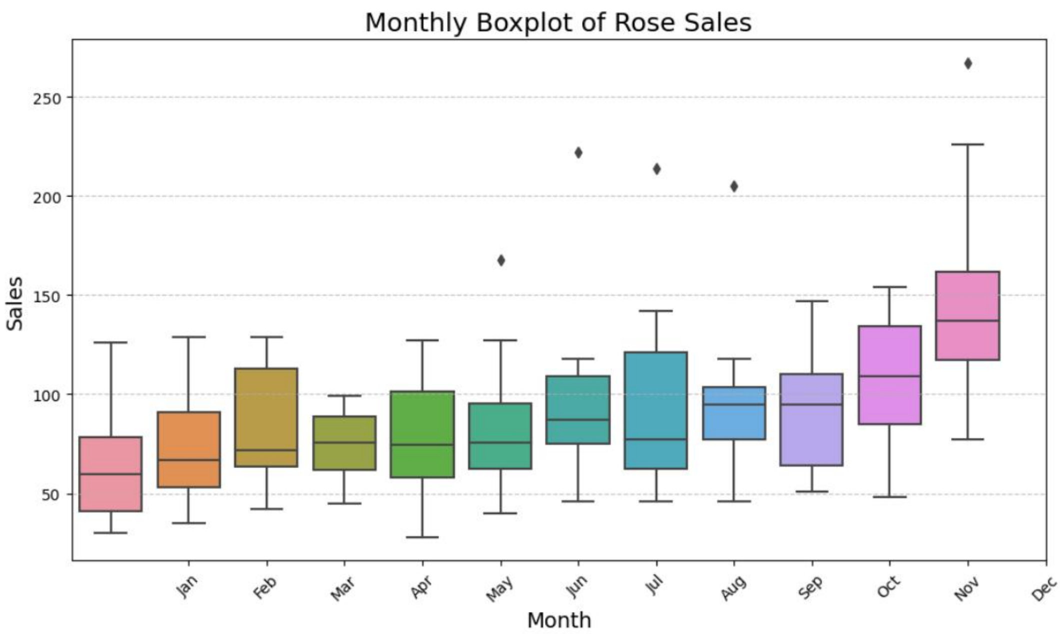


G. Actionable Insights & Recommendations:-

Conclude with the key takeaways (actionable insights and recommendations) for the business:-

ROSE WINE:-

- **Long-term decline:** Sales have been dropping since 1980, indicating a decrease in popularity.
- **Seasonal spike:** Sales rise significantly during the holiday season (Oct-Dec), peaking in December. (Likely due to holiday celebrations)
- **Post-holiday slump:** Sales sharply decline in the first quarter (Jan-Mar), possibly reflecting a post-holiday slowdown.
- **Gradual recovery:** Sales slowly pick up again by May-June.



Rose Wine Sales - Action Plan:

Capitalize on the Holiday Season:

- Increase Inventory: Stock up on rose wine in anticipation of the rising sales and December peak (based on forecast).

Address Long-Term Decline:

- Data Analysis: Conduct further data analysis to understand the reasons behind the long-term decline in sales.

Rebranding & Innovation:

- Consider Rebranding: Explore rebranding the existing rose wine with a fresh image, potentially alongside a new winemaker.

Marketing & Promotions:

- Pre-Holiday Push (Aug-Oct): Launch targeted marketing campaigns and special offers to attract new customers, particularly first-time wine drinkers and those open to different brands.

Decision Point (Post-Holiday Season):

- Evaluate Sales Performance: Assess the overall sales trend after the December peak.
- Positive Trend: Continue with the existing rose wine variant.

SPARKLING WINE:-

Flat Trend, Seasonal Fluctuations:

- Unlike rose wine, sparkling wine sales show no long-term upward or downward trend, indicating a stable but stagnant market.

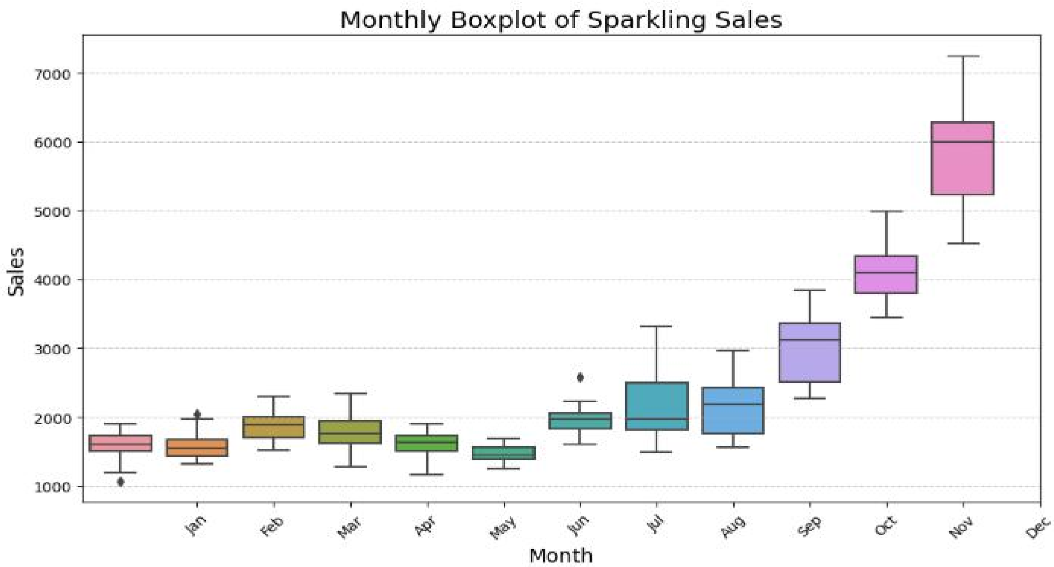
Holiday Boom & Post-Holiday Slump:

- Similar to rose, sparkling wine experiences a significant seasonal spike during the holiday season (Oct-Dec), with sales in December reaching nearly triple the volume of September. This surge likely aligns with holiday celebrations.
- Following the December peak, sales plummet in the first quarter (Jan-Mar), mirroring the post-holiday slump observed in rose.

Recovery and Recommendations:

- Sales gradually recover by July-August, suggesting a potential opportunity to:

1. Target summer celebrations: Develop marketing campaigns promoting sparkling wine for summer gatherings (e.g., picnics, barbecues).
2. Maintain steady inventory: While December brings a significant sales increase, a consistent inventory level throughout the year might help capture potential customers who enjoy sparkling wine outside the holiday season.



Sparkling Wine Sales - Action Plan:

Capitalize on the Holiday Season:

- Increase Inventory: Build stock in anticipation of rising sales and the December peak (based on forecast).
- Targeted Advertising (Oct-Dec): Launch focused advertising campaigns during the holiday season (Oct-Dec) to leverage the existing buying trend and potentially boost sales further.

Product Innovation & Marketing:

- Celebration-Themed Design: Consider introducing a special, lower-priced bottle design specifically for celebratory purposes (e.g., bottle designed for popping).
- Summer Marketing: Explore marketing campaigns promoting sparkling wine for summer gatherings (e.g., picnics, barbecues) to capitalize on the sales recovery period (Jul-Aug).

Investigate Flat Sales Trend:

- Deep Sales Dive (Jan-Mar): Utilize the first quarter (Jan-Mar) slowdown to conduct a thorough analysis of year-over-year sales data to understand the stagnant sales trend. This will help identify potential areas for improvement outside the holiday season.