Name: Kuntumalla Jayashree

Date: 24/12/2022

TIME SERIES FORECASTING

WINES DATA

TABLE OF CONTENTS

List of Tables

List of figures

Questions

Problem 1:

Wines Data Analysis:

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: Sparkling.csv and Rose.csv.

Data Dictionary of Sparkling and Rose datasets:
YearMonth *: Year and month for which sales count is calculated*
Sparkling *: Sales of Sparkling wine*
Rose *: Sales of Rose wine.*

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

- Both datasets are read and stored in the pandas dataframes (df_rose and df_spar) for the purpose of analysis.
- Datasets are loaded as time series data with parse_date as true and "YearMonth" as index.
- There is total of 187 records from 1980 to 1995 of wine types Rose and sparkling.
- There are no duplicates in both datasets.
- There are 2 null values in the Rose wine dataset while the sparkling dataset has no null values in it.

	Sparkling
YearMonth	
1980-01-01	1686
1980-02-01	1591
1980-03-01	2304
1980-04-01	1712
1980-05-01	1471

Fig1: Sparkling data

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

Fig2: Rose data

- We have imputed the null values in the Rose dataset with forward values using ffill() method of python. "ffill()" method is used to fill the missing values in the dataframe. Ffill stands for forward fill.
- Data Description is as below:

	Rose
count	187.000000
mean	89.909091
std	39.244440
min	28.000000
25%	62.500000
50%	85.000000
75%	111.000000
max	267.000000

Fig3: Rose describe

	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

Fig4: Sparkling describe

- Minimum number of sales of Rose wine type is 28 while the maximum sales count is 267.Average count of sales of Rose wine is nearly 90.
- We see maximum sales of Rose wine type happened in Dec 1980 while minimum sales happened in May-1995.
- Minimum number of sales of Sparkling wine type is 1070 while the maximum sales count is 7242.Average count of sales of Rose wine is nearly 2402.
- We see maximum sales of Sparkling wine type happened in Dec 1987 while minimum sales happened in Jan-1995.

Rose Data Plot:

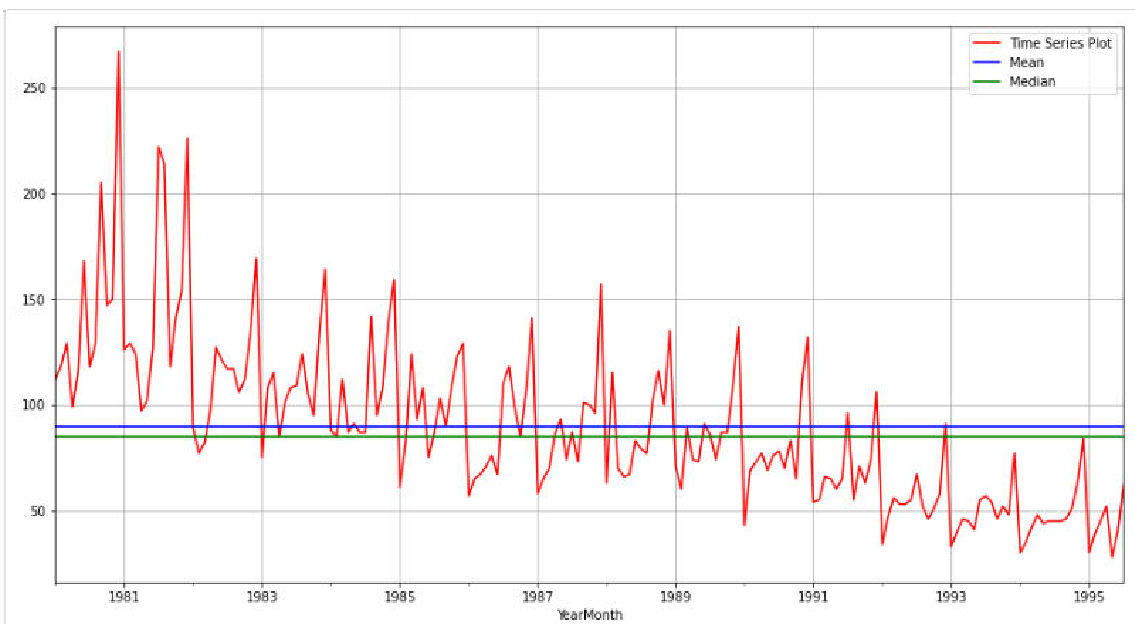


Fig5: Rose data plot

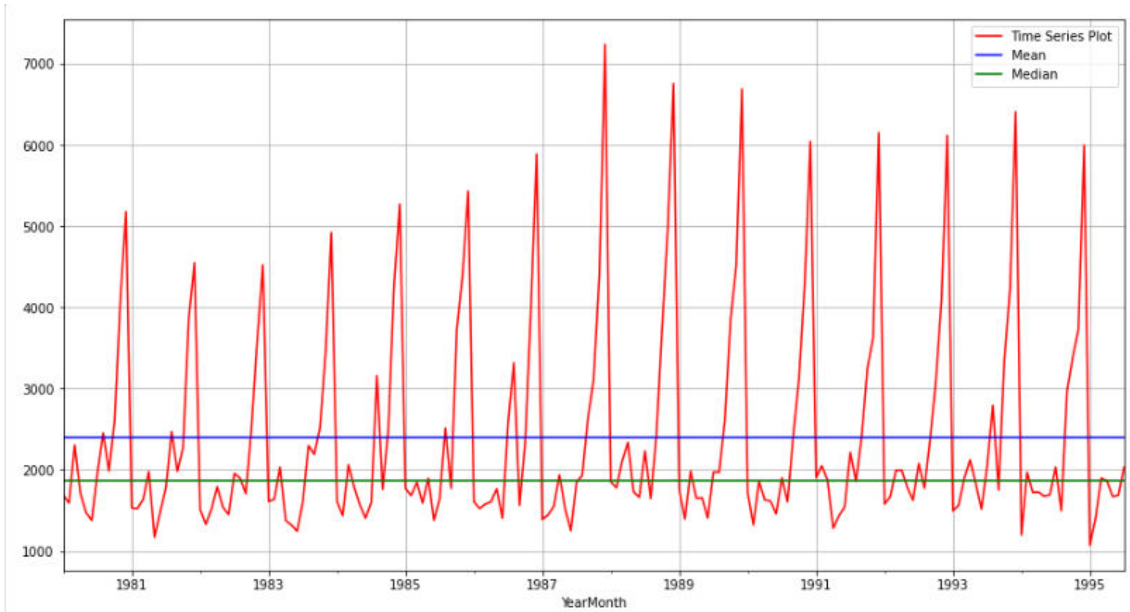


Fig6: Sparkling data plot

Observations:

- There is a slight downward trend with seasonality associated. Average sales and most of the sales are almost same .i.e Mean and median are almost near to each other.
- There is a some upward and downward trend with some seasonality associated. Most of the sales count is around 1990.

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

Descriptive statistics of both the datasets:

	Rose
count	187.000000
mean	89.909091
std	39.244440
min	28.000000
25%	62.500000
50%	85.000000
75%	111.000000
max	267.000000

	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

Fig7: Descriptive stats

Info:

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

Fig8: Info-Rose wine

```
<class 'pandas.core.frame.DataFrame'>
DatetimeIndex: 187 entries, 1980-01-01 to 1995-07-01
Data columns (total 1 columns):
 #   Column     Non-Null Count  Dtype
---  ------     --------------  -----
 0   Sparkling  187 non-null    int64
dtypes: int64(1)
memory usage: 2.9 KB
```

Fig9: Info-Sparkling wine

Yearly Box Plot:

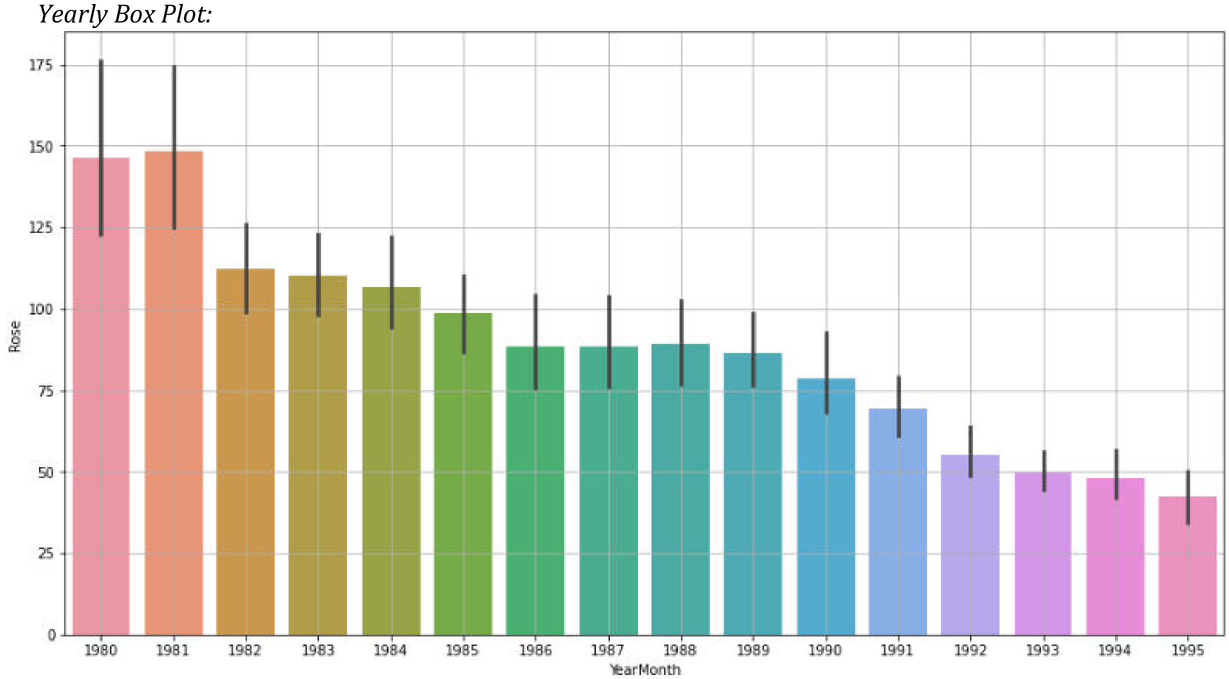


Fig10: Rose Yearly Boxplot

- Downward trend in the sales of the Rose wine from 1980 to 1995
- Highest number of sales got recorded in the year 1981 and least in the year 1995

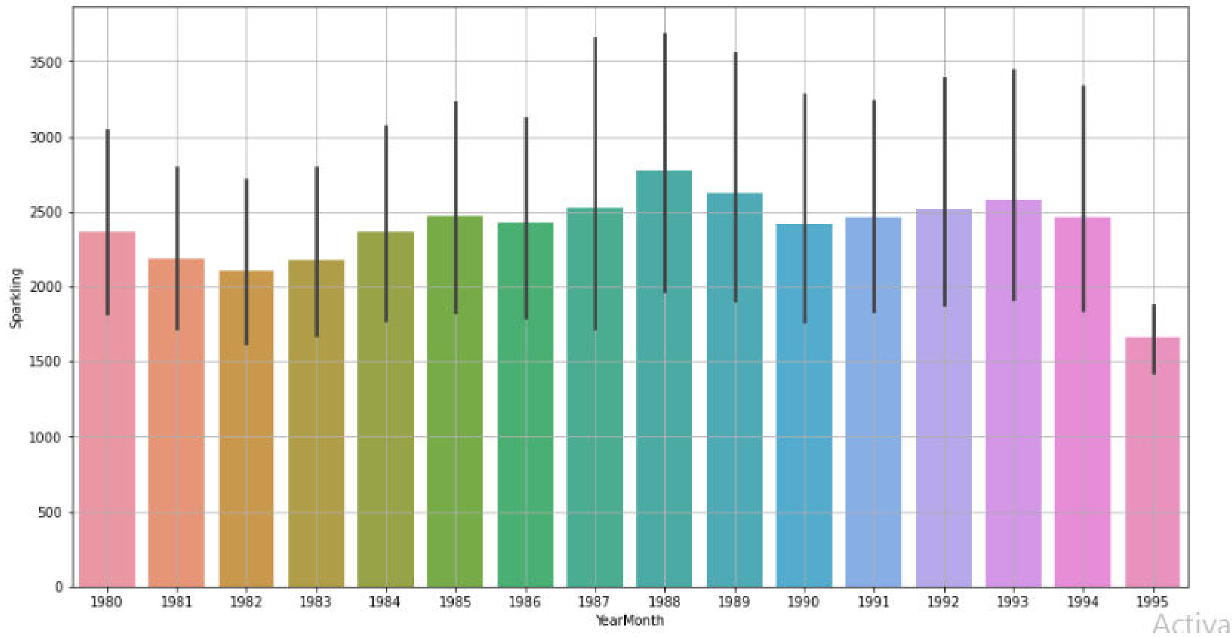


Fig11: Sparkling Yearly Box plot

- Polynomial trend in the sales count of the Sparkling wine type.
- Max sales count is in the year 1988 while the least is in the year 1995

Monthly Box Plot:

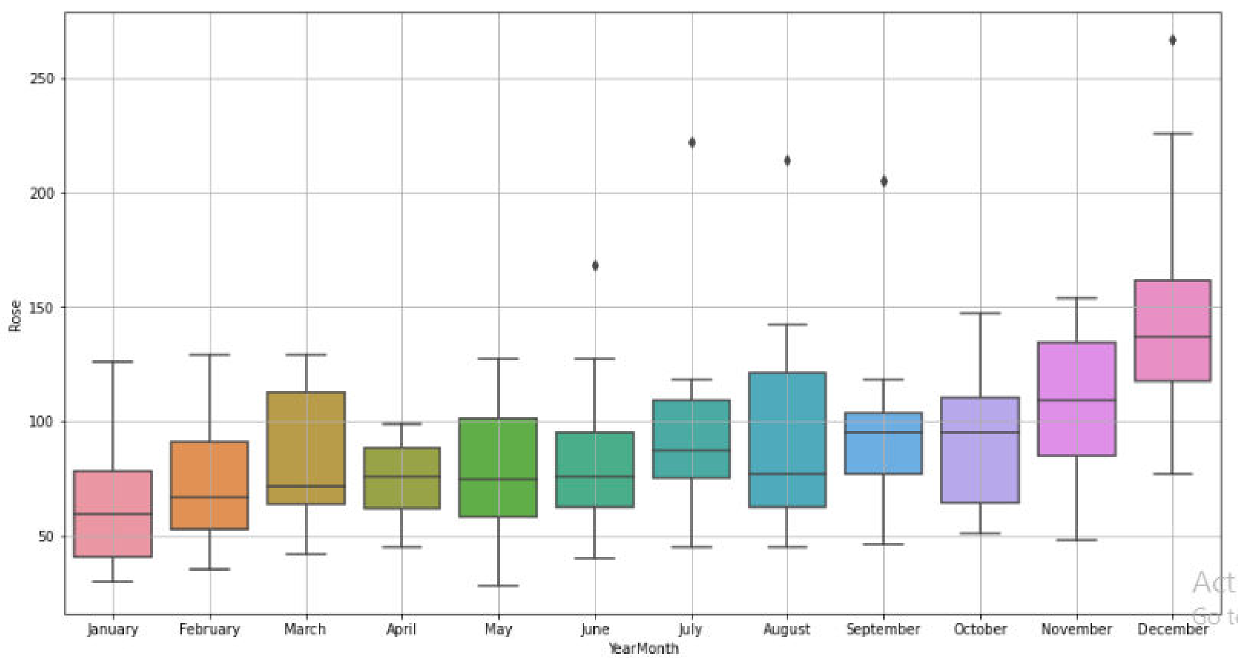


Fig12: Rose Monthly Box plot

- We can observe there are some outliers in the month of June,July,August,September and December in Rose wine type.
- Highest sales happened in the month of December while the least in January across various years.

- Sales got increased in the 4th quarter and decreased in quarter1 starting.

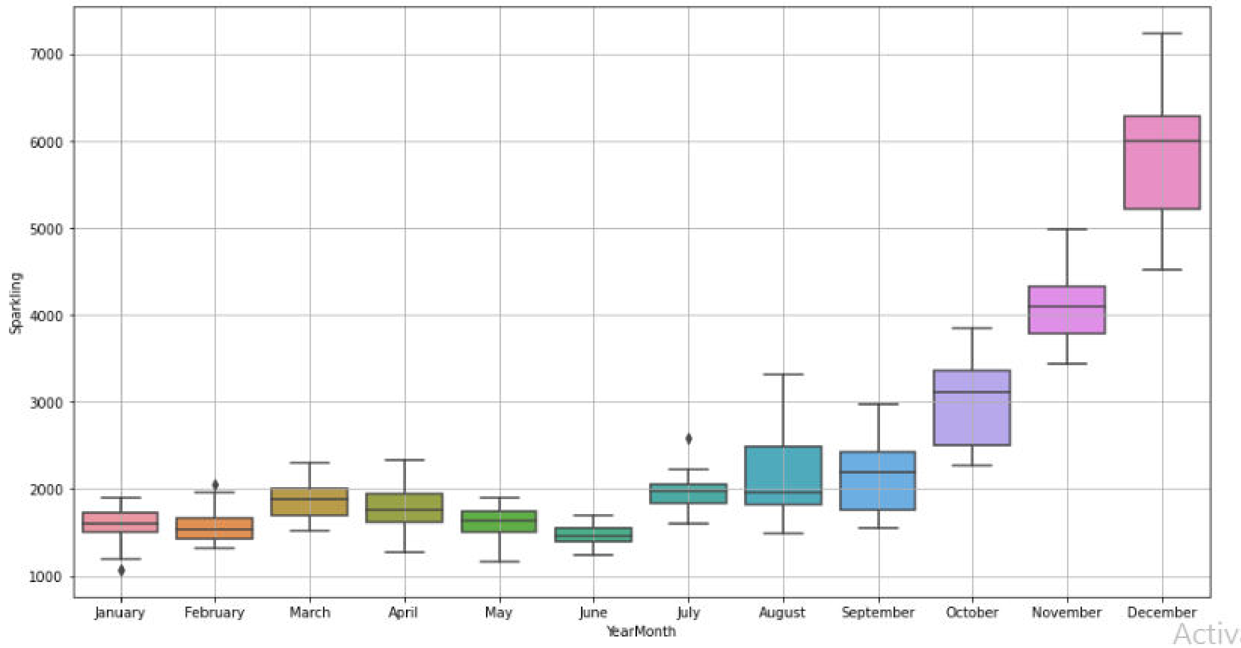


Fig13: Sparkling Monthly Box plot

- We can observe there are some outliers in the month of January, February and July in Sparkling wine type.
- Highest sales happened in the month of December while the least in June across various years.
- Sales increased in the quarter4 and drastically decreased by quarter1 starting.

Sales of both the wines increased in quarter4 due to holiday season.

<u>Month plot across different years and within different months across years</u>:

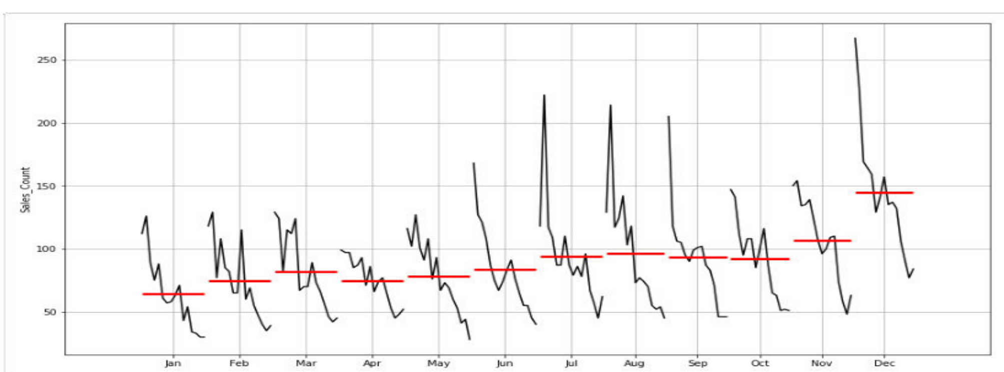


Fig14: Rose Month plot

- This plot shows us the behavior of the Time Series ('Rose sales' in this case) across various months. The red line is the median value.

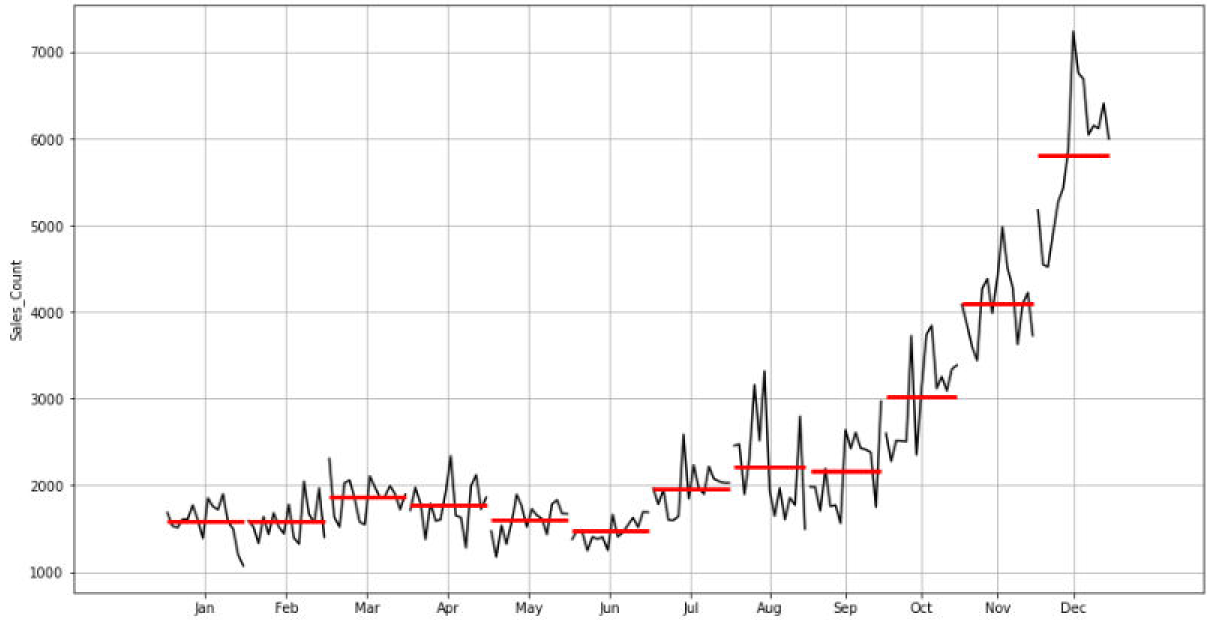
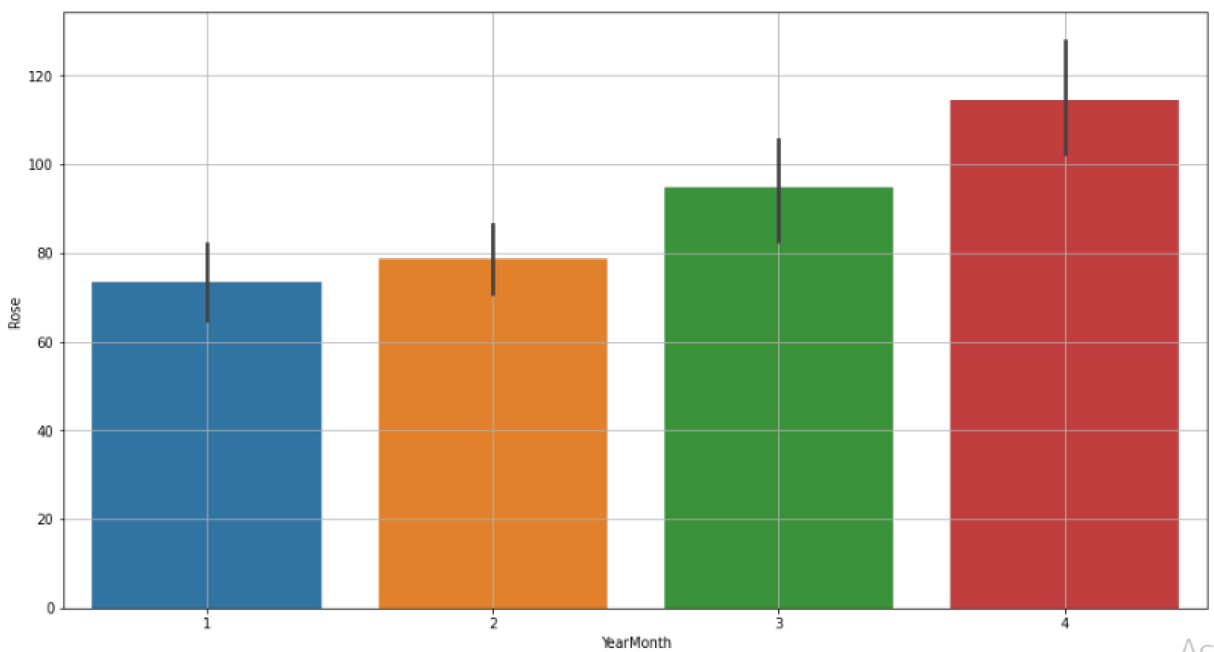


Fig15: Sparkling Month plot

- This plot shows us the behavior of the Time Series (Sparkling sales' in this case) across various months. The red line is the median value.

Quarterly Box Plot:



Activ

Fig16: Rose Quarterly Box plot

- Most of the sales is observed in the 4th quarter.
- Less number of sales is observed in the 1st quarter.

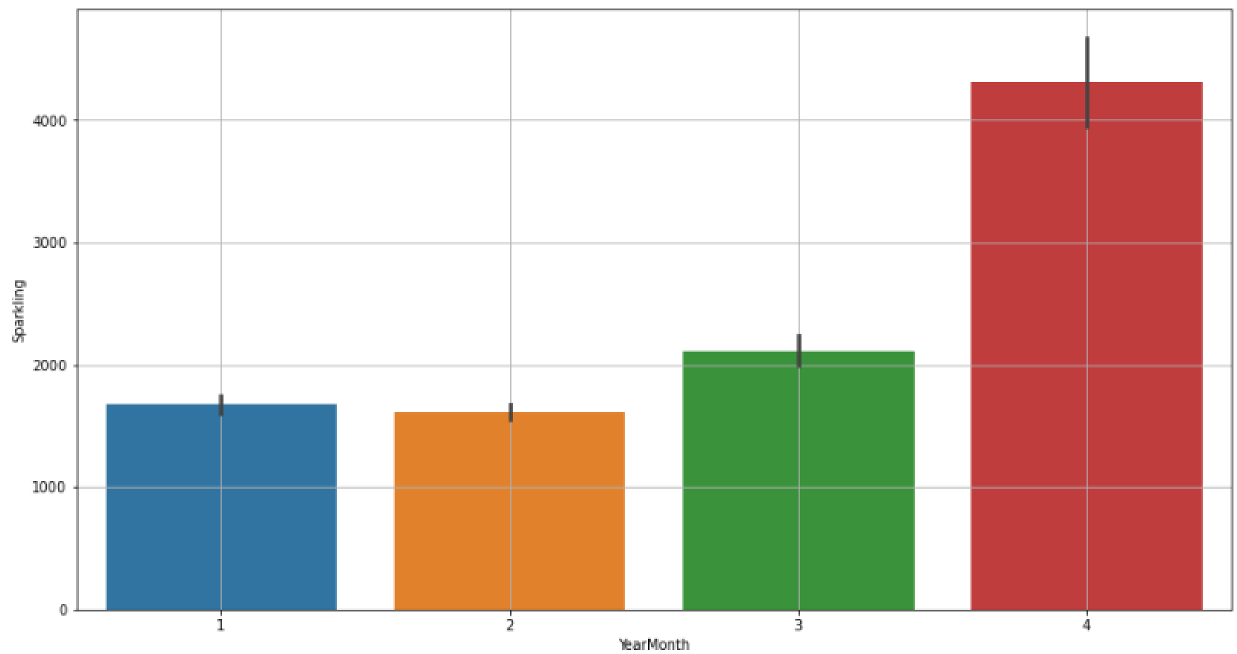


Fig17: Sparkling Quarterly Box plot

- Most of the sales is observed in the 4th quarter.
- Less number of sales is observed in the 2nd quarter.

Monthly sales graph across years:

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

Fig18: Pivot Monthly sales-Rose

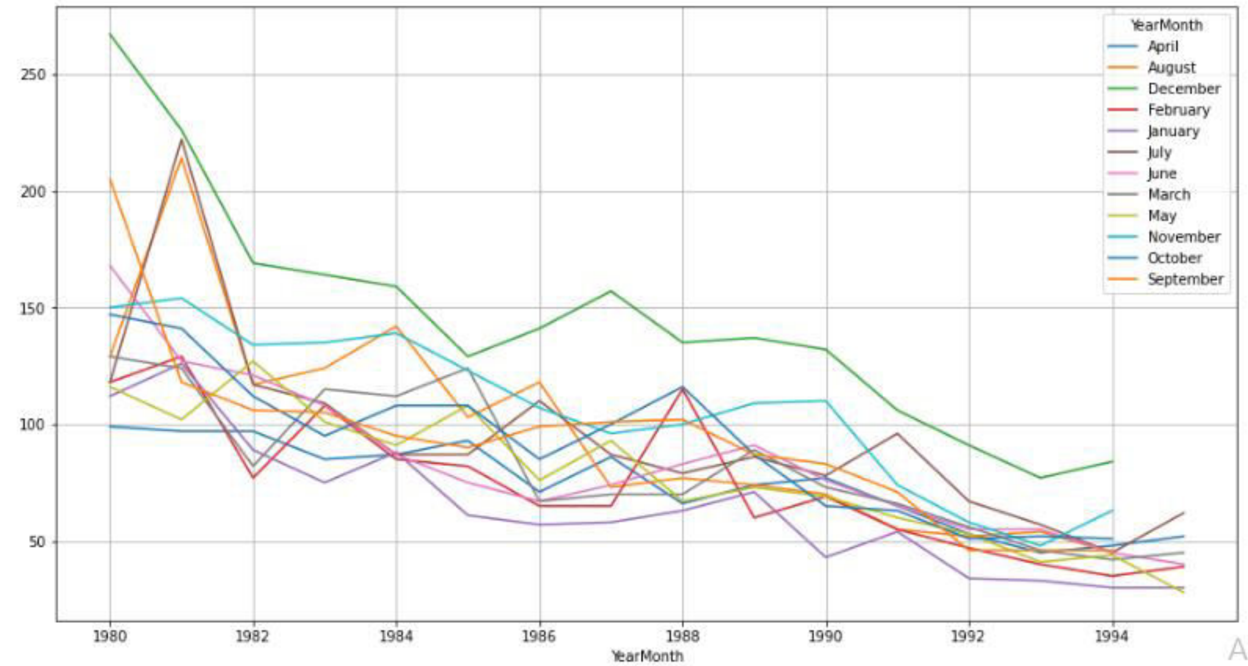


Fig19: Monthly sales Graph-Rose

YearMonth	April	August	December	February	January	July	June	March	May	November	October	September
YearMonth												
1980	1712.0	2453.0	5179.0	1591.0	1686.0	1966.0	1377.0	2304.0	1471.0	4087.0	2596.0	1984.0
1981	1976.0	2472.0	4551.0	1523.0	1530.0	1781.0	1480.0	1633.0	1170.0	3857.0	2273.0	1981.0
1982	1790.0	1897.0	4524.0	1329.0	1510.0	1954.0	1449.0	1518.0	1537.0	3593.0	2514.0	1706.0
1983	1375.0	2298.0	4923.0	1638.0	1609.0	1600.0	1245.0	2030.0	1320.0	3440.0	2511.0	2191.0
1984	1789.0	3159.0	5274.0	1435.0	1609.0	1597.0	1404.0	2061.0	1567.0	4273.0	2504.0	1759.0
1985	1589.0	2512.0	5434.0	1682.0	1771.0	1645.0	1379.0	1846.0	1896.0	4388.0	3727.0	1771.0
1986	1605.0	3318.0	5891.0	1523.0	1606.0	2584.0	1403.0	1577.0	1765.0	3987.0	2349.0	1562.0
1987	1935.0	1930.0	7242.0	1442.0	1389.0	1847.0	1250.0	1548.0	1518.0	4405.0	3114.0	2638.0
1988	2336.0	1645.0	6757.0	1779.0	1853.0	2230.0	1661.0	2108.0	1728.0	4988.0	3740.0	2421.0
1989	1650.0	1968.0	6694.0	1394.0	1757.0	1971.0	1406.0	1982.0	1654.0	4514.0	3845.0	2608.0
1990	1628.0	1605.0	6047.0	1321.0	1720.0	1899.0	1457.0	1859.0	1615.0	4286.0	3116.0	2424.0
1991	1279.0	1857.0	6153.0	2049.0	1902.0	2214.0	1540.0	1874.0	1432.0	3627.0	3252.0	2408.0
1992	1997.0	1773.0	6119.0	1667.0	1577.0	2076.0	1625.0	1993.0	1783.0	4096.0	3088.0	2377.0
1993	2121.0	2795.0	6410.0	1564.0	1494.0	2048.0	1515.0	1898.0	1831.0	4227.0	3339.0	1749.0
1994	1725.0	1495.0	5999.0	1968.0	1197.0	2031.0	1693.0	1720.0	1674.0	3729.0	3385.0	2968.0
1995	1862.0	NaN	NaN	1402.0	1070.0	2031.0	1688.0	1897.0	1670.0	NaN	NaN	NaN

Fig20: Pivot Monthly sales-Sparkling

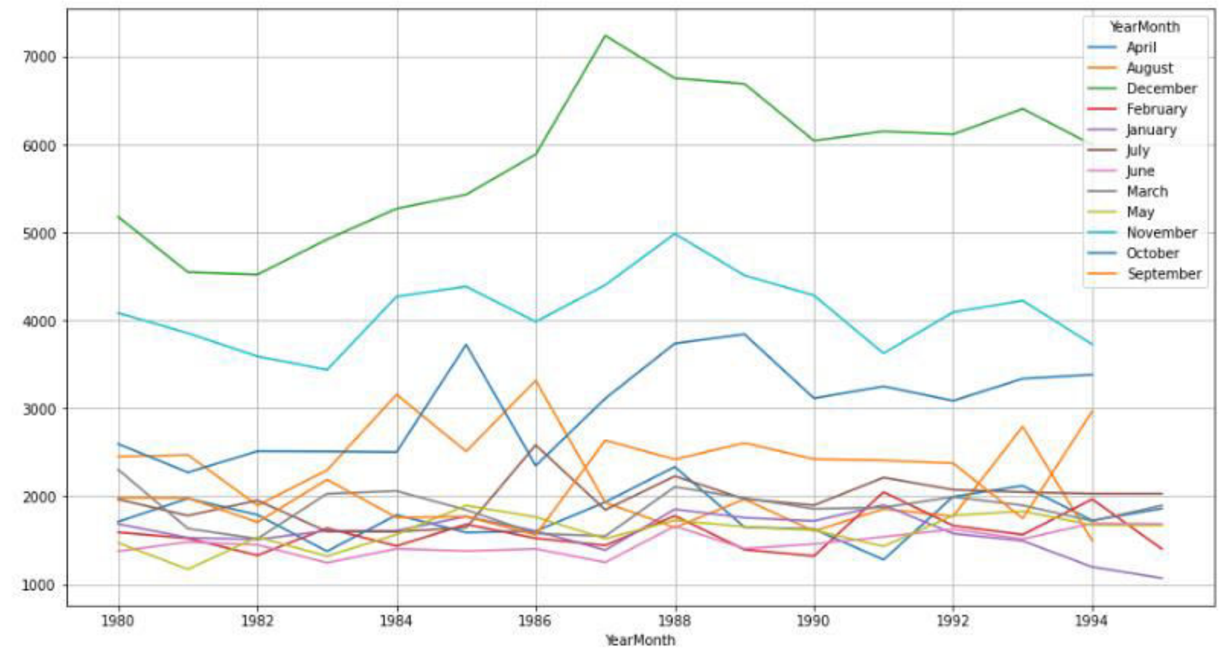


Fig21: Monthly sales Graph-Sparkling

<u>Additive Decomposition of Rose:</u>

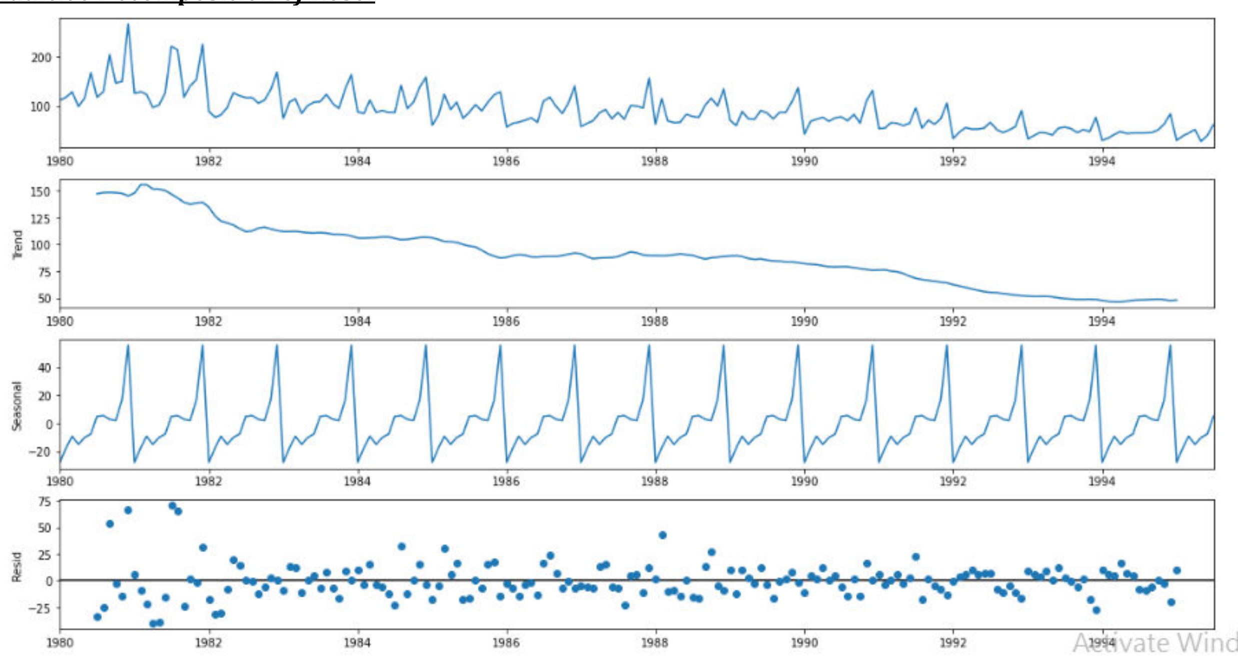


Fig22: Additive decomposition-Rose

<u>Multiplicative Decomposition of Rose:</u>

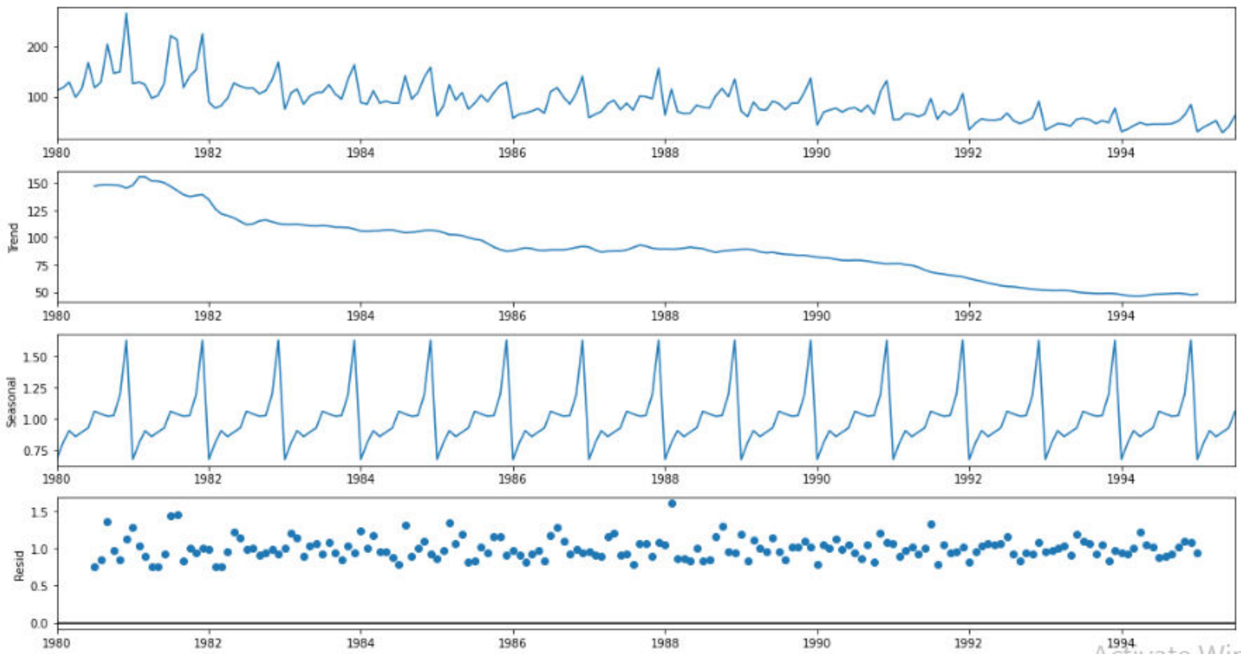


Fig23: Multiplicative decompostion-Rose

Additive Decompostion							
YearMonth	Rose Trend		YearMonth	Rose Seasonality		YearMonth	Rose Residual
1980-1-1	NaN		1980-1-1	-27.903092		1980-1-1	NaN
1980-2-1	NaN		1980-2-1	-17.431663		1980-2-1	NaN
1980-3-1	NaN		1980-3-1	-9.279878		1980-3-1	NaN
1980-4-1	NaN		1980-4-1	-15.092378		1980-4-1	NaN
1980-5-1	NaN		1980-5-1	-10.190592		1980-5-1	NaN
1980-6-1	NaN		1980-6-1	-7.672735		1980-6-1	NaN
1980-7-1	147.083333		1980-7-1	4.880241		1980-7-1	-33.963575
1980-8-1	148.125		1980-8-1	5.460797		1980-8-1	-24.585797
1980-9-1	148.375		1980-9-1	2.780241		1980-9-1	53.844759
1980-10-1	148.083333		1980-10-1	1.877464		1980-10-1	-2.960797
1980-11-1	147.416667		1980-11-1	16.852464		1980-11-1	-14.26913
1980-12-1	145.125		1980-12-1	55.71913		1980-12-1	66.15587

Multiplicative Decompostion							
YearMonth	Rose Trend		YearMonth	Rose Seasonality		YearMonth	Rose Residual
1980-1-1	NaN		1980-1-1	0.670182		1980-1-1	NaN
1980-2-1	NaN		1980-2-1	0.806224		1980-2-1	NaN
1980-3-1	NaN		1980-3-1	0.901278		1980-3-1	NaN
1980-4-1	NaN		1980-4-1	0.854154		1980-4-1	NaN
1980-5-1	NaN		1980-5-1	0.889531		1980-5-1	NaN
1980-6-1	NaN		1980-6-1	0.924099		1980-6-1	NaN
1980-7-1	147.083333		1980-7-1	1.057682		1980-7-1	0.758514
1980-8-1	148.125		1980-8-1	1.035066		1980-8-1	0.841382
1980-9-1	148.375		1980-9-1	1.017753		1980-9-1	1.357534
1980-10-1	148.083333		1980-10-1	1.022688		1980-10-1	0.970661
1980-11-1	147.416667		1980-11-1	1.192494		1980-11-1	0.853274
1980-12-1	145.125		1980-12-1	1.628848		1980-12-1	1.129506

Table1: Trend-seasonal-residual-Rose

Additive Decomposition of Sparkling:

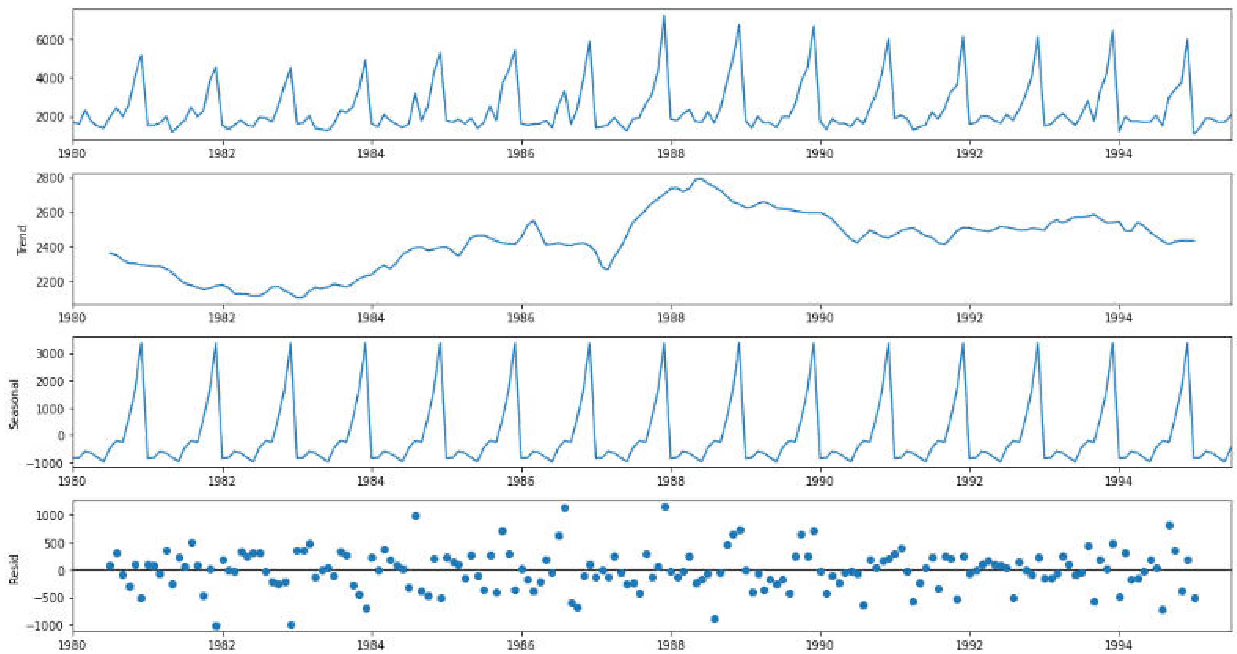


Fig24: Additive decomposition-Sparkling

Multiplicative Decomposition of Sparkling:

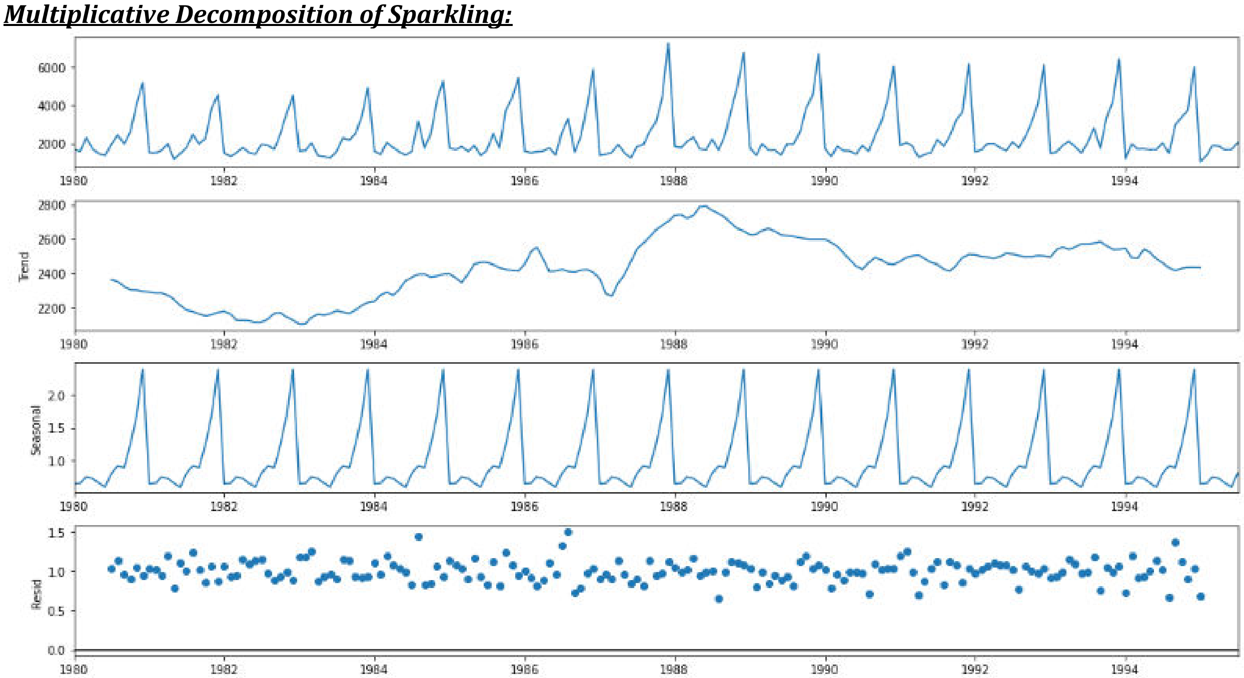


Fig25: Multiplicative decomposition-Sparkling

Additive Decompostion							
YearMonth	Sparkling Trend		YearMonth	rkling Seasonality		YearMonth	rkling Residual
1980-1-1	NaN		1980-1-1	-854.260599		1980-1-1	NaN
1980-2-1	NaN		1980-2-1	-830.350678		1980-2-1	NaN
1980-3-1	NaN		1980-3-1	-592.35663		1980-3-1	NaN
1980-4-1	NaN		1980-4-1	-658.490559		1980-4-1	NaN
1980-5-1	NaN		1980-5-1	-824.416154		1980-5-1	NaN
1980-6-1	NaN		1980-6-1	-967.434011		1980-6-1	NaN
1980-7-1	2360.666667		1980-7-1	-465.502265		1980-7-1	70.835599
1980-8-1	2351.333333		1980-8-1	-214.332821		1980-8-1	315.999487
1980-9-1	2320.541667		1980-9-1	-254.677265		1980-9-1	-81.864401
1980-10-1	2303.583333		1980-10-1	599.769957		1980-10-1	-307.35329
1980-11-1	2302.041667		1980-11-1	1675.067179		1980-11-1	109.891154
1980-12-1	2293.791667		1980-12-1	3386.983846		1980-12-1	-501.775513

Multiplicative Decompostion							
YearMonth	Sparkling Trend		YearMonth	rkling Seasonality		YearMonth	rkling Residual
1980-1-1	NaN		1980-1-1	0.649843		1980-1-1	NaN
1980-2-1	NaN		1980-2-1	0.659214		1980-2-1	NaN
1980-3-1	NaN		1980-3-1	0.75744		1980-3-1	NaN
1980-4-1	NaN		1980-4-1	0.730351		1980-4-1	NaN
1980-5-1	NaN		1980-5-1	0.660609		1980-5-1	NaN
1980-6-1	NaN		1980-6-1	0.603468		1980-6-1	NaN
1980-7-1	2360.666667		1980-7-1	0.809164		1980-7-1	1.02923
1980-8-1	2351.333333		1980-8-1	0.918822		1980-8-1	1.135407
1980-9-1	2320.541667		1980-9-1	0.894367		1980-9-1	0.955954
1980-10-1	2303.583333		1980-10-1	1.241789		1980-10-1	0.907513
1980-11-1	2302.041667		1980-11-1	1.690158		1980-11-1	1.050423
1980-12-1	2293.791667		1980-12-1	2.384776		1980-12-1	0.94677

Table2: Trend-seasonal-residual-Sparkling

Additive Model:

Seasonality remains constant over time

yt = Trend + Seasonalit y + Residual

Multiplicative Model:

Seasonality changes (increases or decreases) over time

yt = Trend * Seasonalit y * Residual

Observations:

From above graphs, we can say that Rose is Multiplicative while sparkling is Additive.

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

Both datasets start split at 1991.

Training Data is till the end of 1990. Test Data is from the beginning of 1991 to the last time stamp provided.

Rose wine:
First few rows of Training Data

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

First few rows of Test Data

YearMonth	Rose
1991-01-01	54.0
1991-02-01	55.0
1991-03-01	66.0
1991-04-01	65.0
1991-05-01	60.0

Last few rows of Training Data

YearMonth	Rose
1990-08-01	70.0
1990-09-01	83.0
1990-10-01	65.0
1990-11-01	110.0
1990-12-01	132.0

Last few rows of Test Data

YearMonth	Rose
1995-03-01	45.0
1995-04-01	52.0
1995-05-01	28.0
1995-06-01	40.0
1995-07-01	62.0

Sparkling wine:
First few rows of Training Data

YearMonth	Sparkling
1980-01-01	1686
1980-02-01	1591
1980-03-01	2304
1980-04-01	1712
1980-05-01	1471

First few rows of Test Data

YearMonth	Sparkling
1991-01-01	1902
1991-02-01	2049
1991-03-01	1874
1991-04-01	1279
1991-05-01	1432

Last few rows of Training Data

YearMonth	Sparkling
1990-08-01	1605
1990-09-01	2424
1990-10-01	3116
1990-11-01	4286
1990-12-01	6047

Last few rows of Test Data

YearMonth	Sparkling
1995-03-01	1897
1995-04-01	1862
1995-05-01	1670
1995-06-01	1688
1995-07-01	2031

Fig26: Training and Testing dataset of Rose and Sparkling

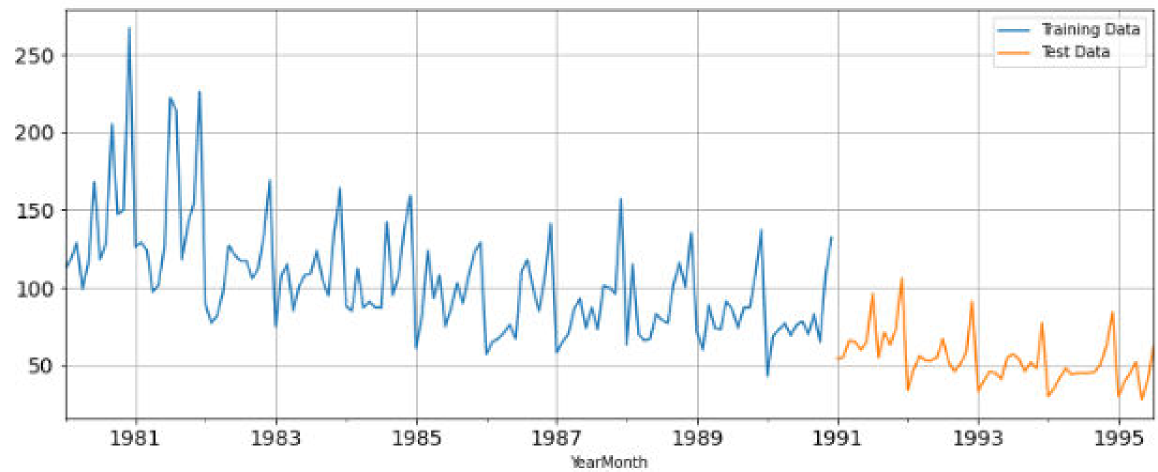


Fig27: Train and Test split-Rose wine

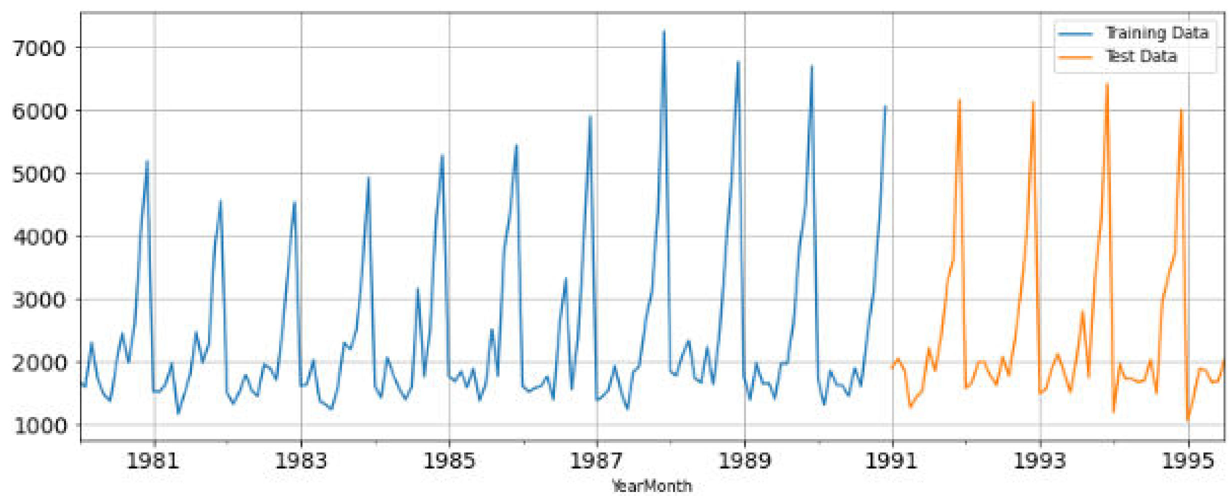


Fig28: Train and Test split-Sparkling wine

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, naive forecast models and simple average models. Should also be built on the training data and check the performance on the test data using RMSE.**

Various Forecasting models applied on the data are as below:

- Linear Regression
- Naïve Forecasting
- Simple Average model
- Moving Average model
- Exponential Smoothing Techniques(Single,Double and Triple exponential smoothing techniques)

Accuracy metric considered to validate performance is RMSE-Root Mean Square Error.

Linear Regression:

Linear Regression
We have applied linear regression on both the datasets (Rose and Sparkling) by modifying the datasets and tagged sales to their individual time.

LR-ROSE

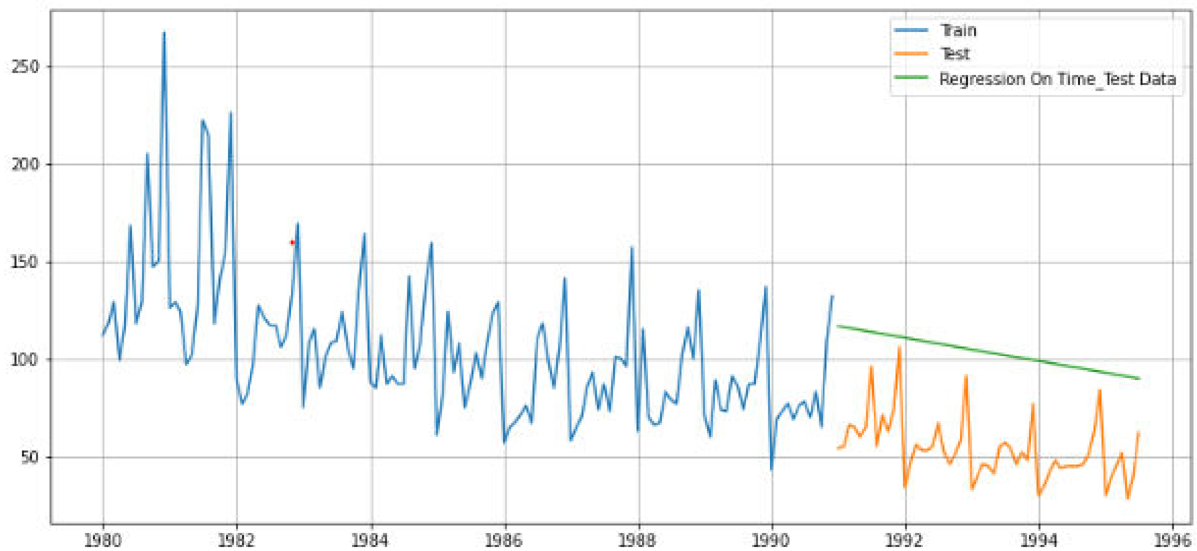


Fig29: Linear Regression-Rose wine

LR-Sparkling

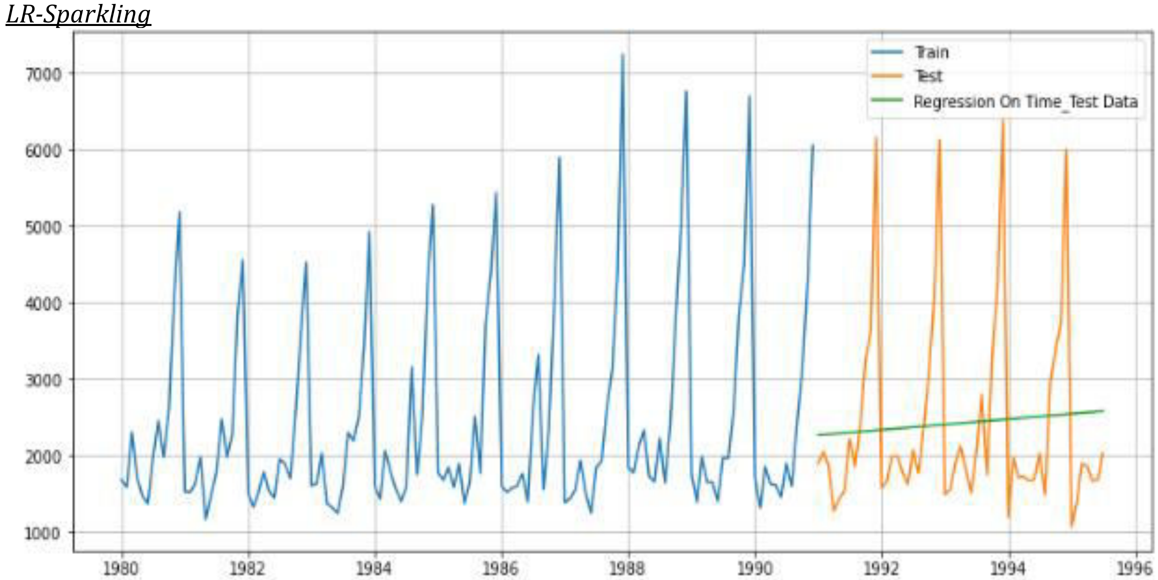


Fig30: Linear Regression-Sparkling wine

RMSE post Regression

	Test RMSE-Rose	Test RMSE-Sparkling
RegressionOnTime	51.45105	1275.867052

Fig31: RMSE-LR

Naïve Forecasting:

Estimating technique in which the last period's actual s are used as this period's forecast, without adjusting them or attempting to establish causal factors

Naïve Forecast -Rose

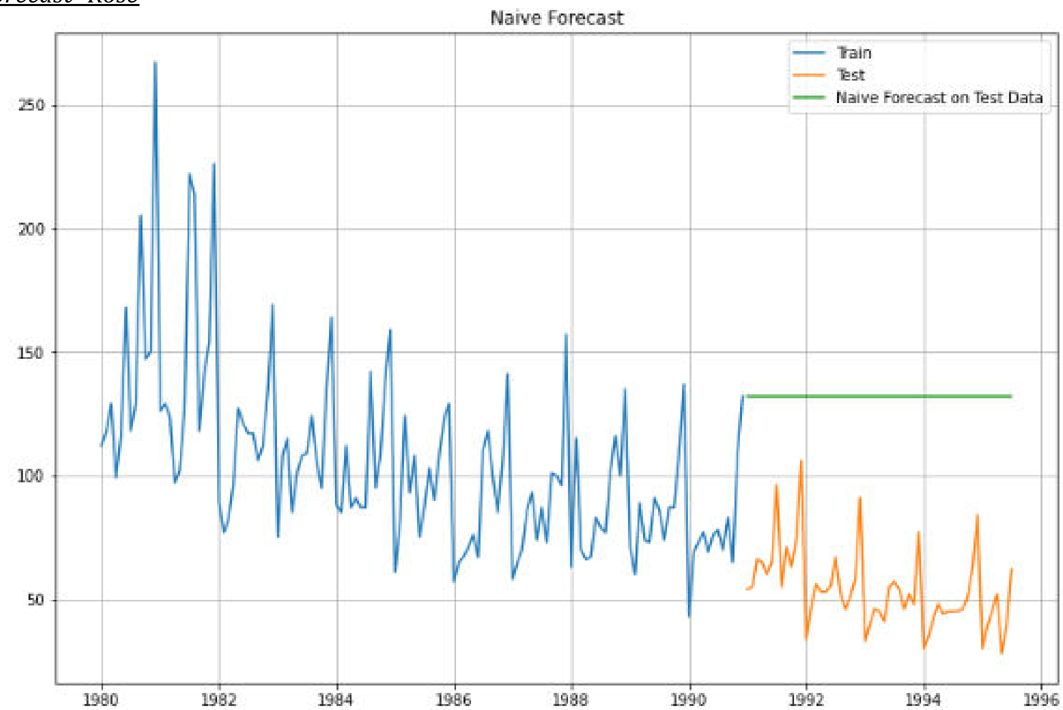


Fig32: Naïve-Rose

Naïve Forecast -Sparkling

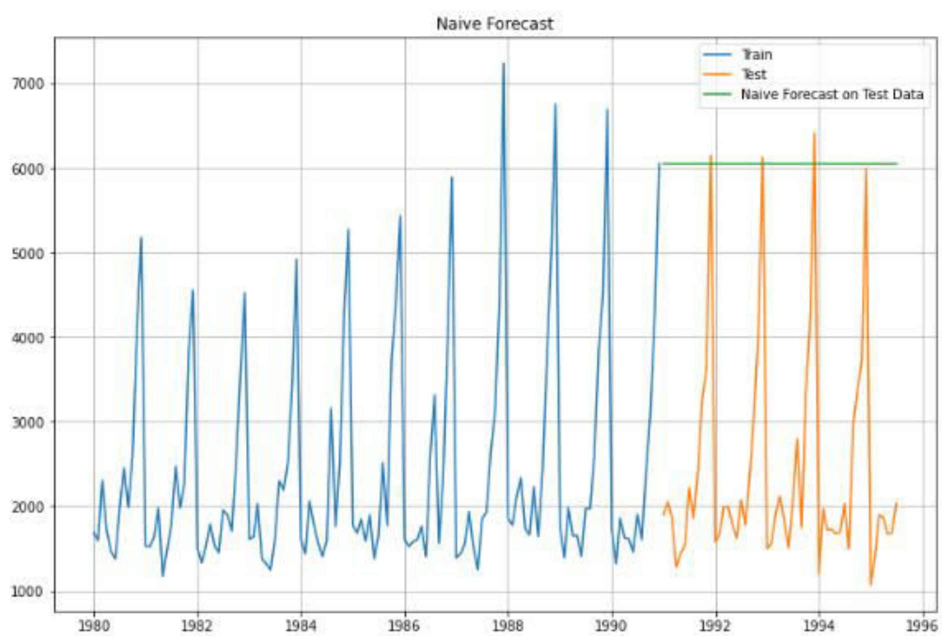


Fig33: Naïve-Sparkling

__Naïve Forecast –RMSE:__

	Test RMSE-Rose	Test RMSE-Sparkling
RegressionOnTime	51.45105	1275.867052
NaiveModel	79.73855	3864.279352

Fig34: Naïve-RMSE

__Simple Average:__

Forecast the expected value equal to the average of all previously observed points.

__Simple Average Forecast -Rose__

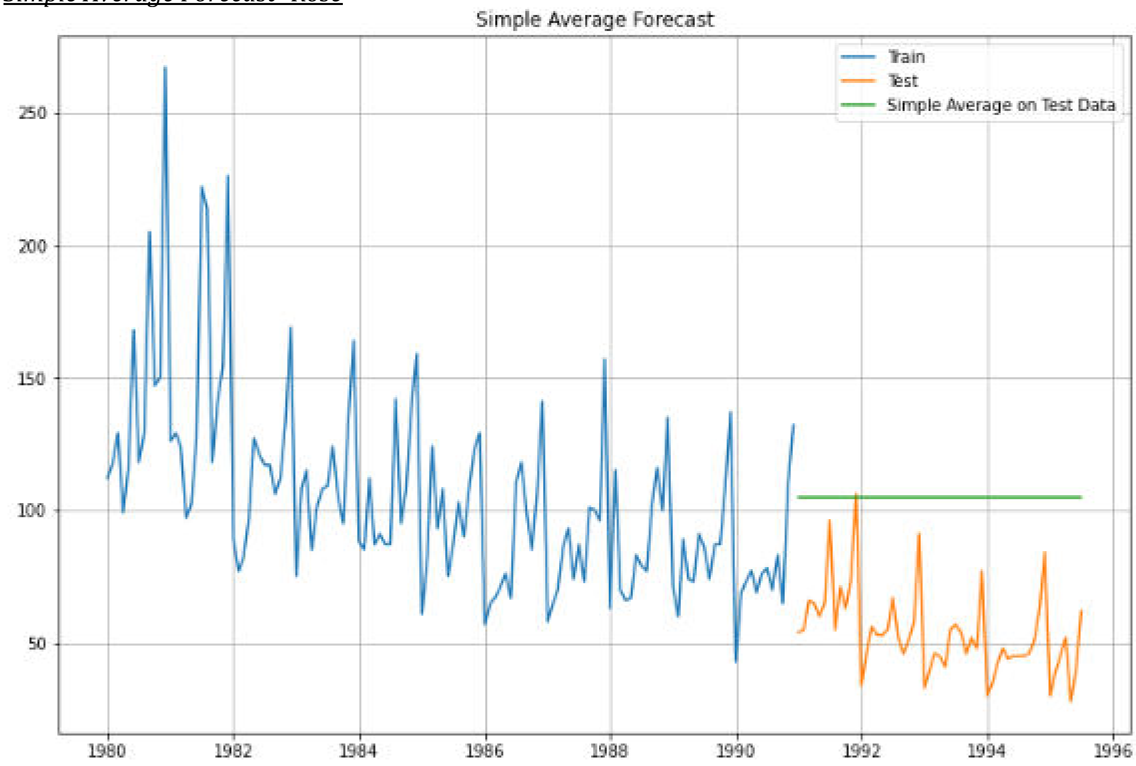


Fig35: Simple Average-Rose

__Simple Average Forecast -Sparkling__

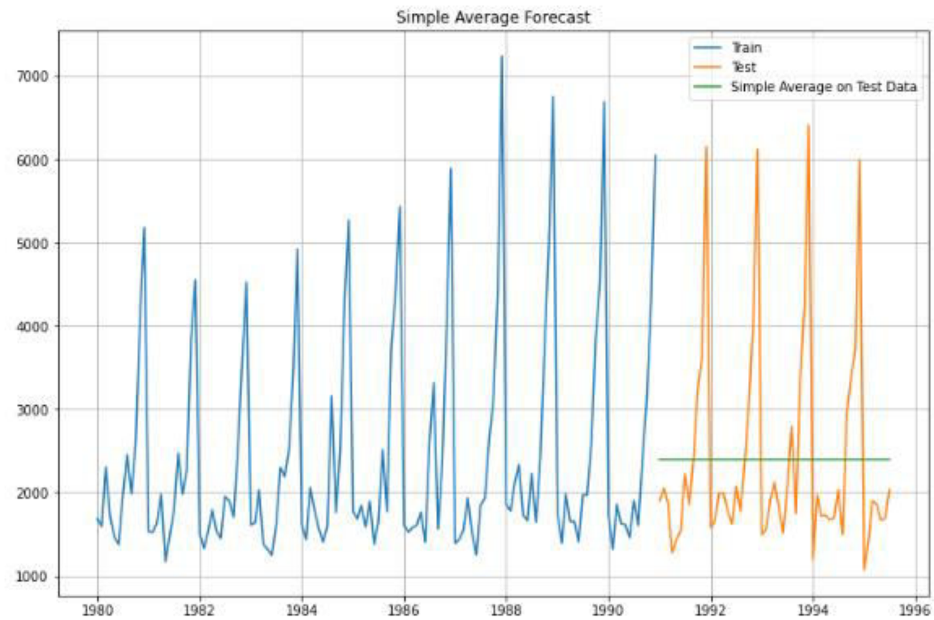


Fig36: Simple Average-Sparkling

Simple Average Forecast -RMSE

	Test RMSE-Rose	Test RMSE-Sparkling
RegressionOnTime	51.45105	1275.867052
NaiveModel	79.73855	3864.279352
SimpleAverageModel	79.73855	1275.081804

Fig37: Simple Average-RMSE

Moving Average:

The technique represents taking an average of a set of numbers in a given range while moving the range. Rolling method from python is used to shift the range .
Here we have taken 2point ,4point,6point and 9pointMoving Average.

2-point MA: Considering 1st and 2nd values to predict the 3rd value. Same way considering the 2nd and 3rd to predict the 4th value and so on.

4-point MA: Considering 1st , 2nd ,3rd and 4th values to predict the 5th value. Same way considering the 2nd, 3rd ,4th and 5th to predict the 6th value and so on.

6-point MA: Considering 1st , 2nd ,3rd , 4th ,5th and 6th values to predict the 7th value. Same way considering the 2nd, 3rd ,4th , 5th ,6th and 7th to predict the 8th value and so on.

9-point MA: Considering 1st , 2nd ,3rd , 4th ,5th , 6th ,7th ,8th and 9thvalues to predict the 10th value. Same way considering the 2nd, 3rd ,4th , 5th ,6th ,7th ,8th ,9th and 10th , to predict the 11th value and so on

Moving Average Forecast –Whole Rose data

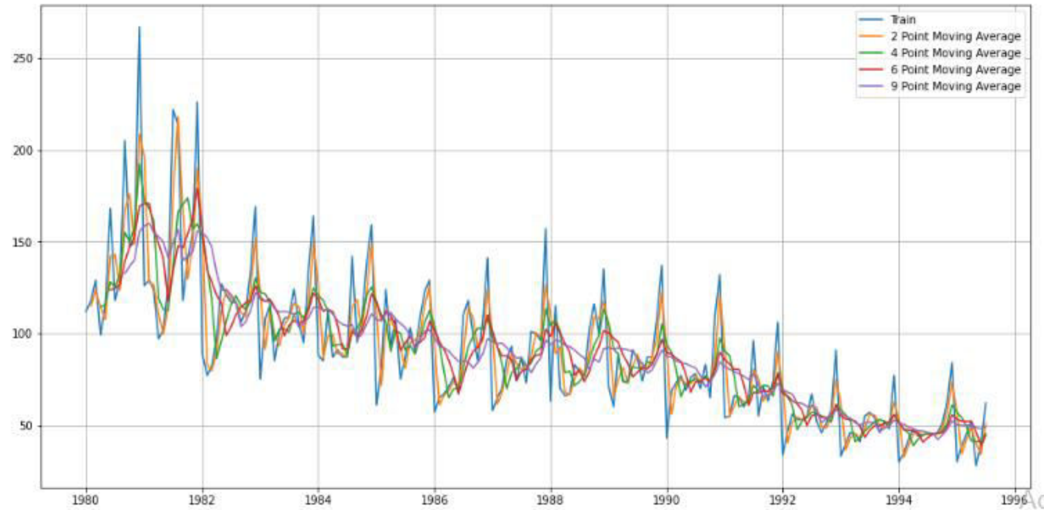


Fig38: Moving Average-Rolling means on whole data-Rose

Moving Average Forecast –Whole Sparkling data

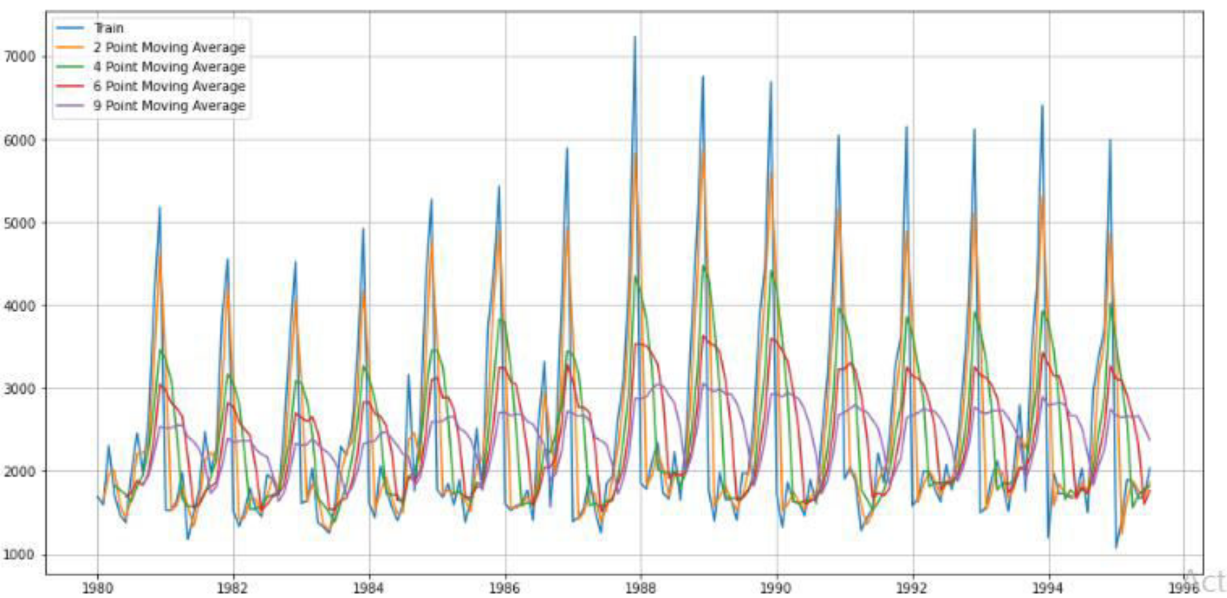
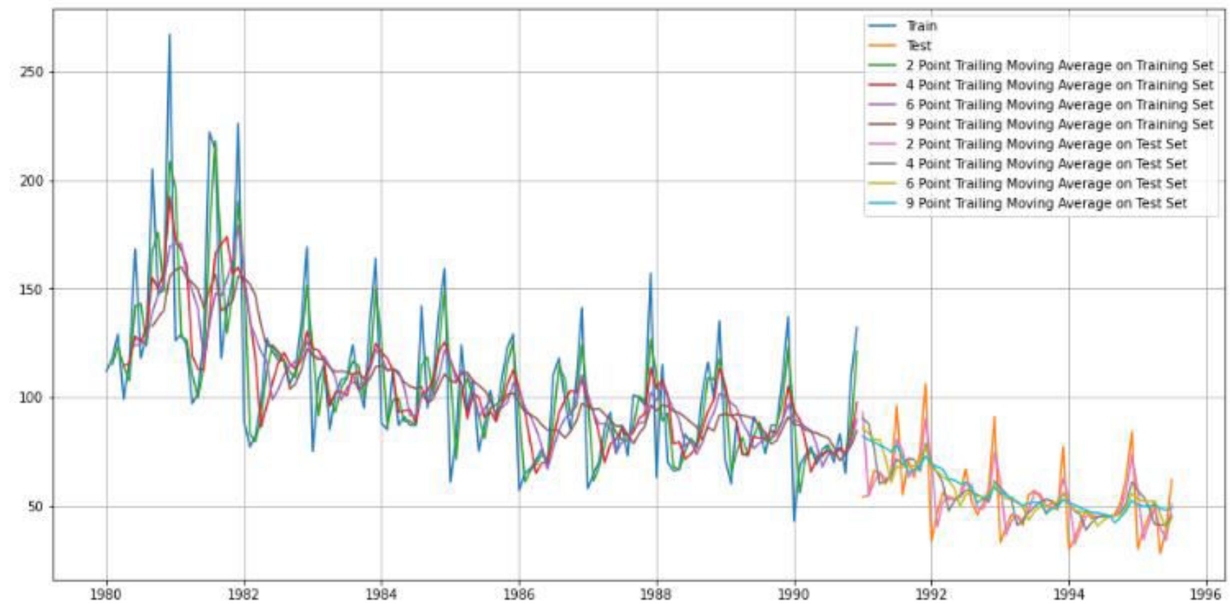


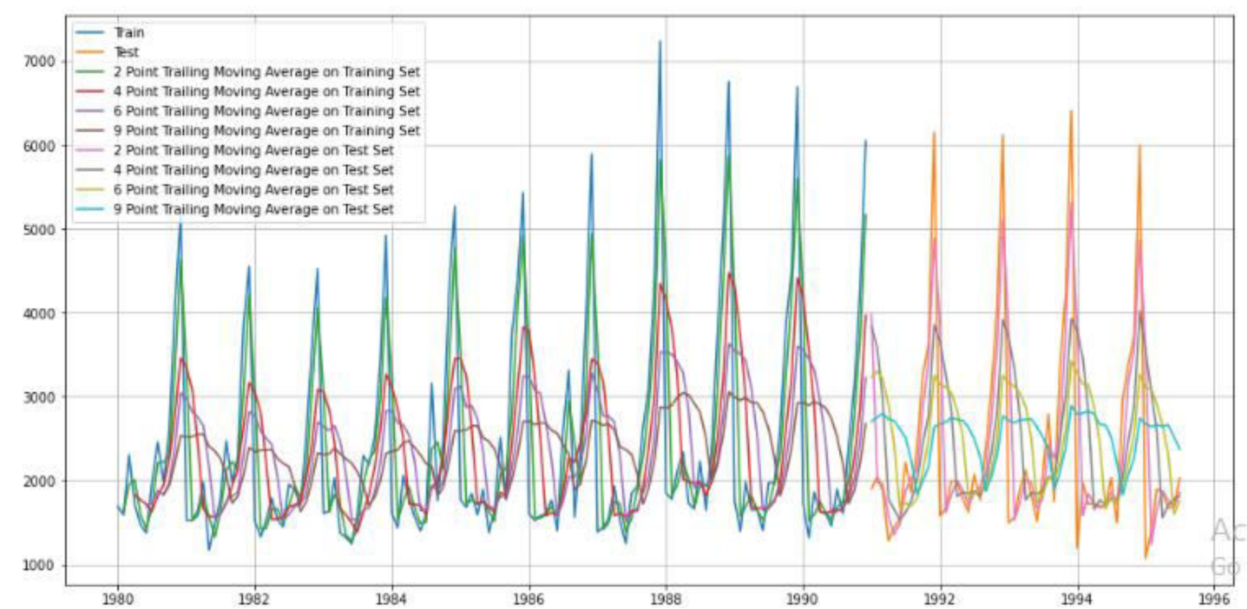
Fig39: Moving Average-Rolling means on Whole data-Sparkling

Moving Average Forecast –Train and Test Rose data



`Fig40: Moving Average-Rolling means on Train & Test data-Rose`

Moving Average Forecast –Train and Test Sparkling data



`Fig41: Moving Average-Rolling means on Train & Test data-Sparkling`

Moving Average Forecast -RMSE

	Test RMSE-Rose	Test RMSE-Sparkling
RegressionOnTime	51.451050	1275.867052
NaiveModel	79.738550	3864.279352
SimpleAverageModel	79.738550	1275.081804
2pointTrailingMovingAverage	11.529409	813.400684
4pointTrailingMovingAverage	14.455221	1156.589694
6pointTrailingMovingAverage	14.572009	1283.927428
9pointTrailingMovingAverage	14.731209	1346.278315

Fig42: RMSE after Moving average

From above RMSE data, we see 2 point trailing moving average is giving the best results with low RMSE values for both Rose and Sparkling datasets

Model comparison plots-Rose Data:

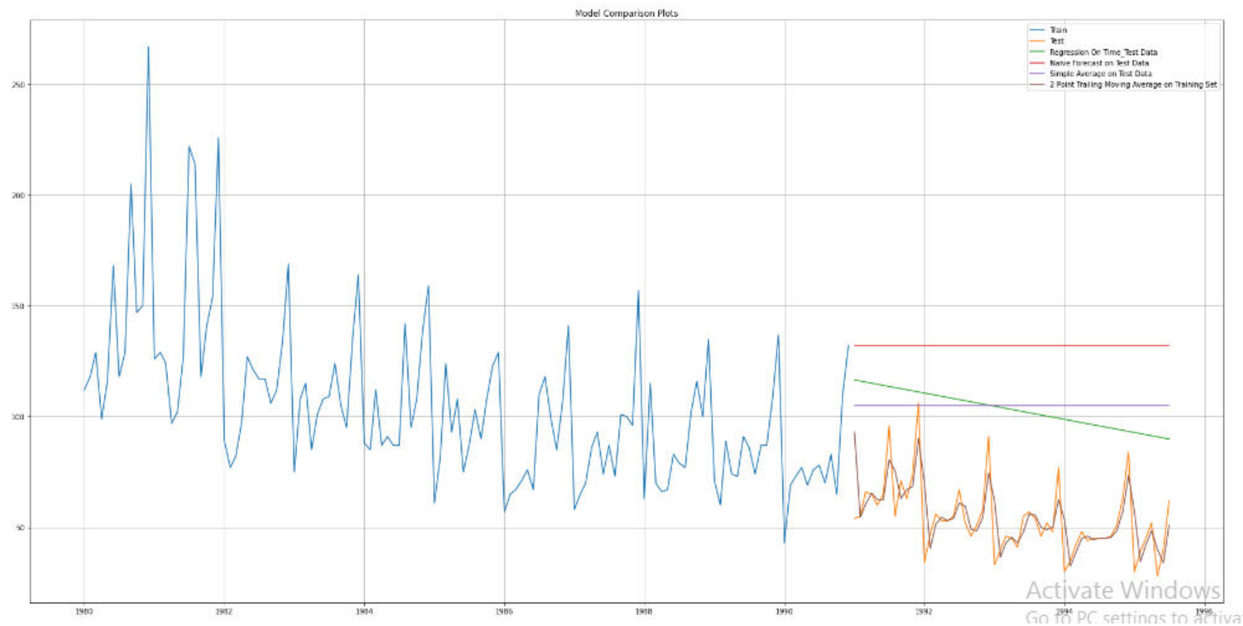


Fig43: Models Comparison1-Rose

Model comparison plots-Sparkling Data:

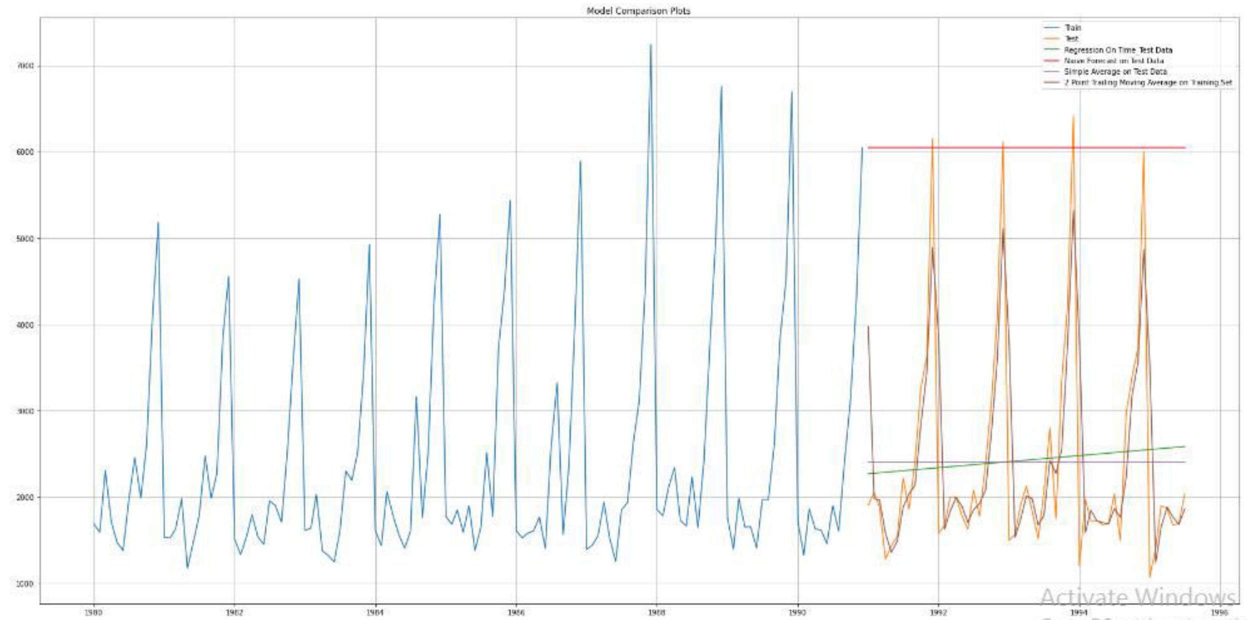


Fig44: Models Comparison1-Sparkling

Simple Exponential Smoothing:

SES is a time series forecasting method with only single parameter alpha which is called as smoothing factor, without trend and seasonality. This method uses weighted moving averages with exponentially decreasing weights.

For Rose ,level parameter (alpha) is 0.0987

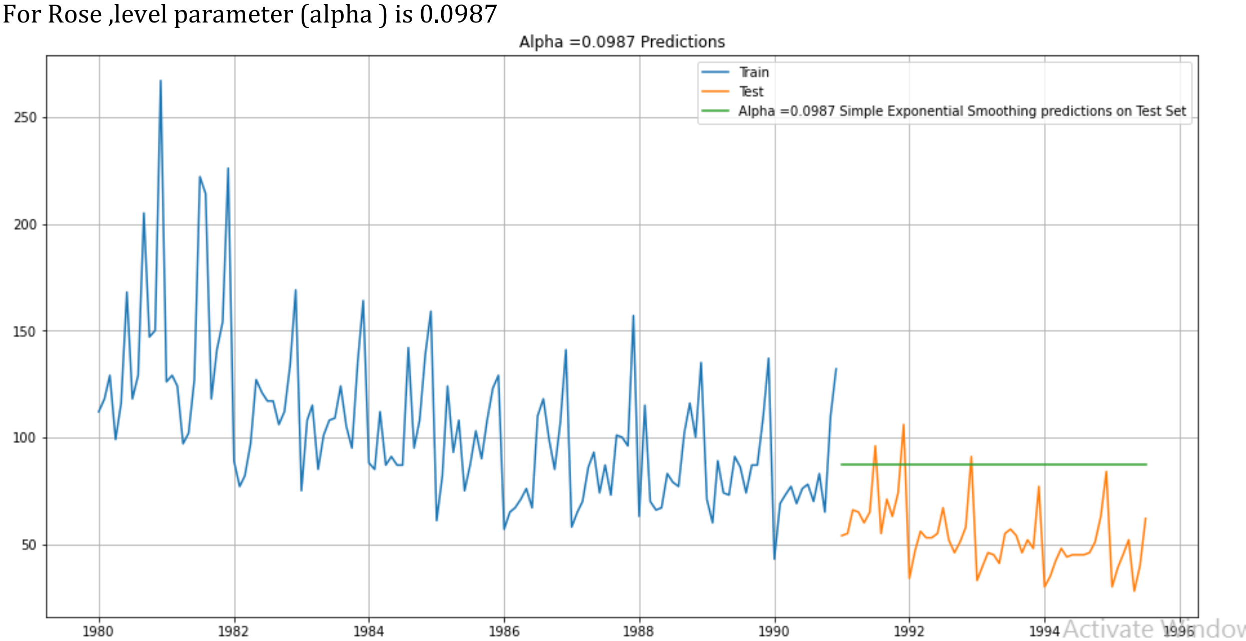


Fig45: Simple Exponential Smoothing-Rose

For Sparkling ,level parameter (alpha) is 0.0.0496

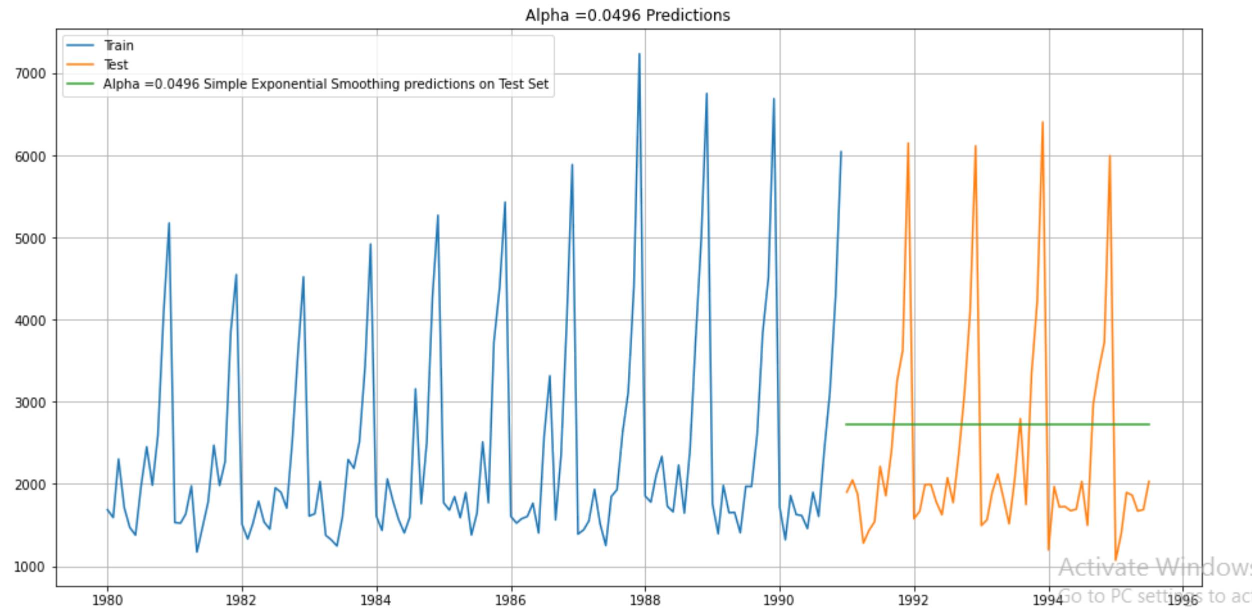


Fig46: Simple Exponential Smoothing-Sparkling

RMSE values post the SES:

	Test RMSE-Rose	Test RMSE-Sparkling
RegressionOnTime	51.451050	1275.867052
NaiveModel	79.738550	3864.279352
SimpleAverageModel	79.738550	1275.081804
2pointTrailingMovingAverage	11.529409	813.400684
4pointTrailingMovingAverage	14.455221	1156.589694
6pointTrailingMovingAverage	14.572009	1283.927428
9pointTrailingMovingAverage	14.731209	1346.278315
Alpha=0.0987,SimpleExponentialSmoothing	36.816905	NaN
Alpha=0.0496,SimpleExponentialSmoothing	NaN	1316.034674

Fig47: Models-RMSE

For now 2 point Moving average has low RMSE score for both Rose and Sparkling Wine.

RMSE for different alpha values:

Performed validation with different Alpha values. Below is the result .

	Alpha Values	Train RMSE	Test RMSE
0	0.1	31.815610	36.848694
1	0.2	31.979391	41.382452
2	0.3	32.470164	47.525251
3	0.4	33.035130	53.787686
4	0.5	33.682839	59.661932
5	0.6	34.441171	64.991324
6	0.7	35.323261	69.718108
7	0.8	36.334596	73.793865
8	0.9	37.482782	77.159094

Fig48: Different Alpha-RMSE-Rose

	Alpha Values	Train RMSE	Test RMSE
0	0.1	1333.873836	1375.393398
1	0.2	1356.042987	1595.206839
2	0.3	1359.511747	1935.507132
3	0.4	1352.588879	2311.919615
4	0.5	1344.004369	2666.351413
5	0.6	1338.805381	2979.204388
6	0.7	1338.844308	3249.944092
7	0.8	1344.462091	3483.801006
8	0.9	1355.723518	3686.794285

Fig49: Different Alpha-RMSE-Sparkling

For both Rose and Sparkling wines, Alpha=0.1 gave the low RMSE value.

Below are the graphs :

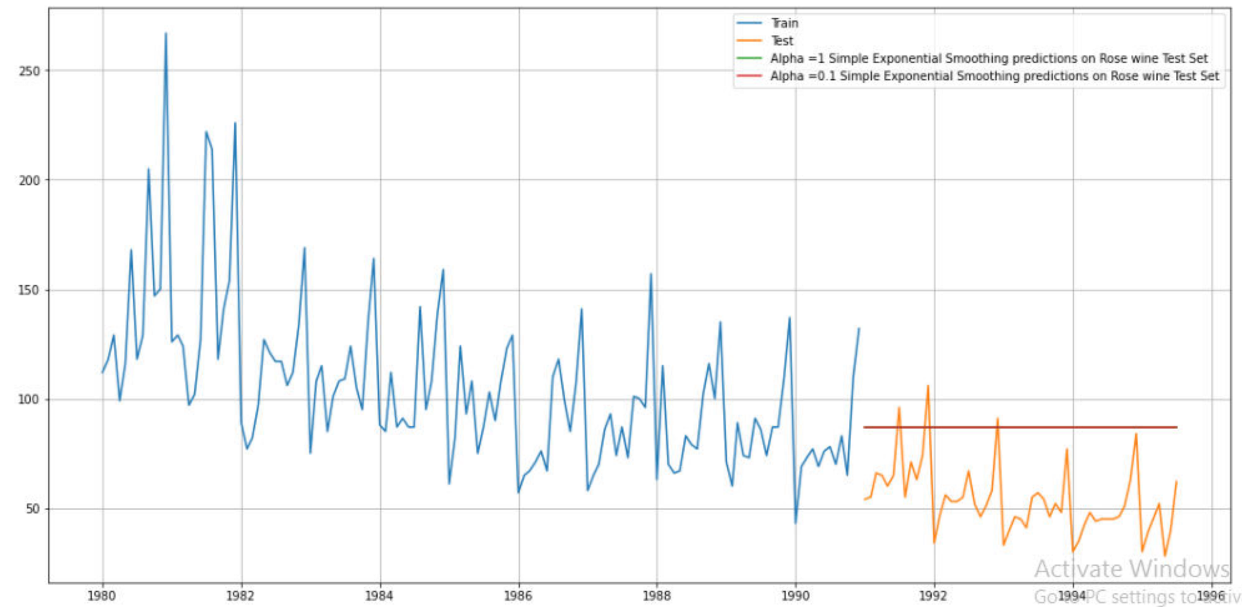


Fig50: SES-Rose-Aplha-1,0.1

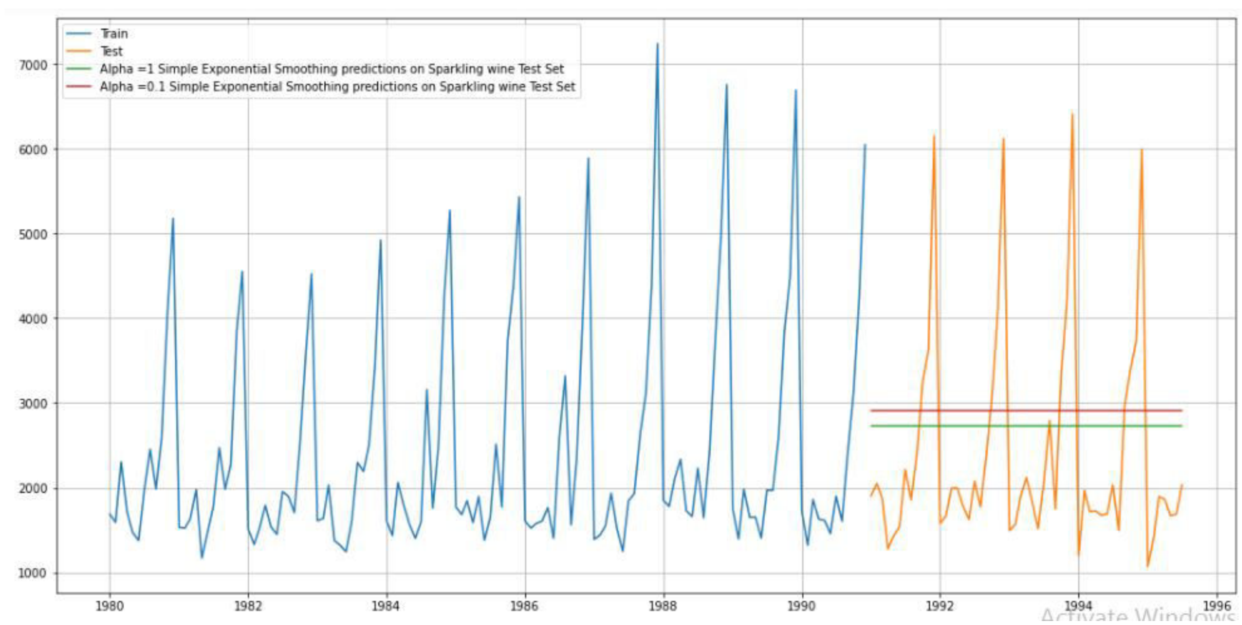


Fig51: SES-Sparkling-Aplha-1,0.1

RMSE:

	Test RMSE-Rose	Test RMSE-Sparkling
RegressionOnTime	51.451050	1275.867052
NaiveModel	79.738550	3864.279352
SimpleAverageModel	79.738550	1275.081804
2pointTrailingMovingAverage	11.529409	813.400684
4pointTrailingMovingAverage	14.455221	1156.589694
6pointTrailingMovingAverage	14.572009	1283.927428
9pointTrailingMovingAverage	14.731209	1346.278315
Alpha=0.0987,SimpleExponentialSmoothing	36.816905	NaN
Alpha=0.0496,SimpleExponentialSmoothing	NaN	1316.034674
Alpha=0.1,SimpleExponentialSmoothing	36.848694	1375.393398
Alpha=0.1,SimpleExponentialSmoothing	36.848694	1375.393398

Fig51.1: RMSE-SES

Even now the 2-point moving average is having low RMSE.

Double Exponential Smoothing (Holt's Model):
DES is a time series forecasting method with 2 parameters alpha which is called as smoothing factor and beta which is called as trend without seasonality.

RMSE for different alpha and beta values:

Performed validation with different Alpha and beta values. Below is the result .

	Alpha Values	Beta Values	Train RMSE	Test RMSE
0	0.3	0.3	35.944983	265.591922
8	0.4	0.3	36.749123	339.330850
1	0.3	0.4	37.393239	358.775361
16	0.5	0.3	37.433314	394.296935
24	0.6	0.3	38.348984	439.320331

Fig52: RMSE-DES-Rose

	Alpha Values	Beta Values	Train RMSE	Test RMSE
0	0.3	0.3	1592.292788	18259.110704
8	0.4	0.3	1569.338606	23878.496940
1	0.3	0.4	1682.573828	26069.841401
16	0.5	0.3	1530.575845	27095.532414
24	0.6	0.3	1506.449870	29070.722592

Fig53: RMSE-DES-Sparkling

For both Rose and Sparkling wines, Alpha=0.3 and beta=0.3 gave the low RMSE value.Below are the graphs :

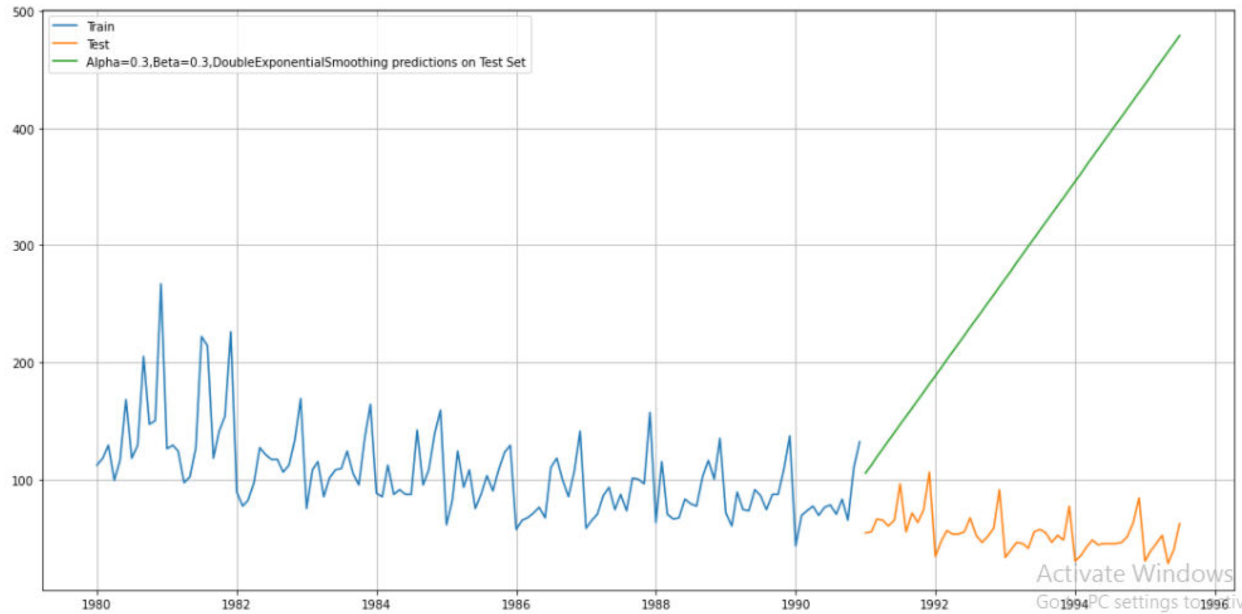


Fig54: DES-Rose

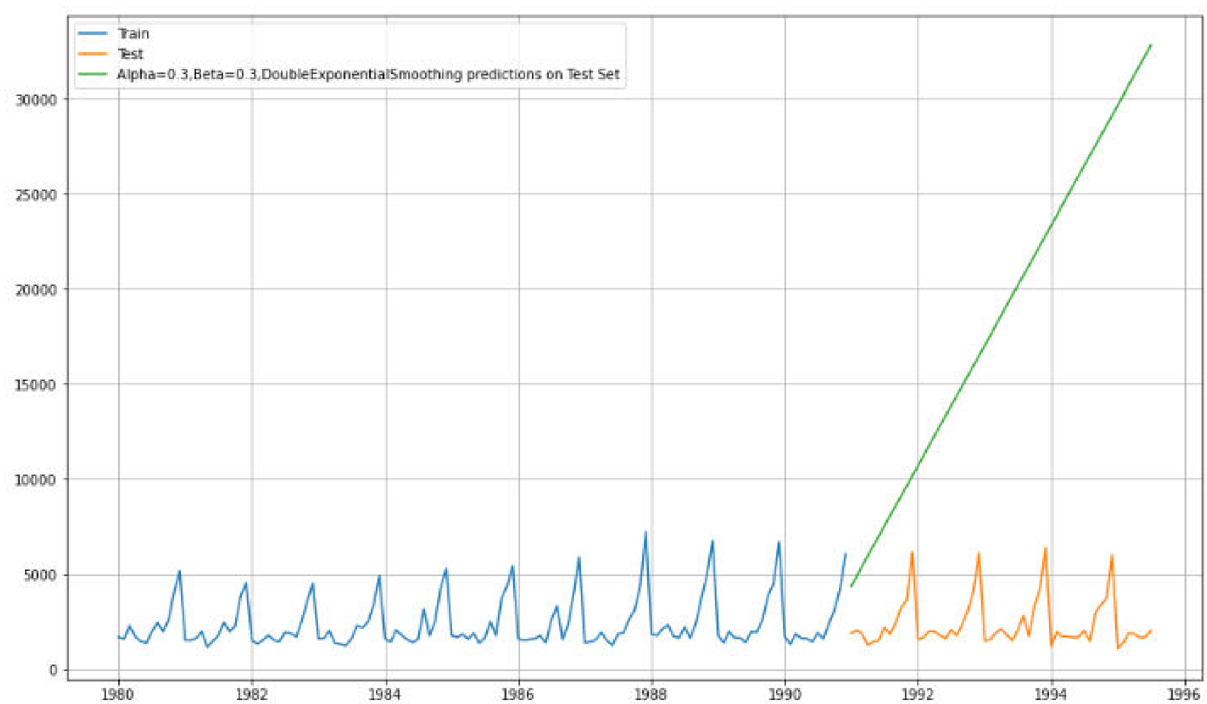


Fig54.1: DES-Sparkling

RMSE:

	Test RMSE-Rose	Test RMSE-Sparkling
RegressionOnTime	51.451050	1275.867052
NaiveModel	79.738550	3864.279352
SimpleAverageModel	79.738550	1275.081804
2pointTrailingMovingAverage	11.529409	813.400684
4pointTrailingMovingAverage	14.455221	1156.589694
6pointTrailingMovingAverage	14.572009	1283.927428
9pointTrailingMovingAverage	14.731209	1346.278315
Alpha=0.0987,SimpleExponentialSmoothing	36.816905	NaN
Alpha=0.0496,SimpleExponentialSmoothing	NaN	1316.034674
Alpha=0.1,SimpleExponentialSmoothing	36.848694	1375.393398
Alpha=0.1,SimpleExponentialSmoothing	36.848694	1375.393398
Alpha=0.3,Beta=0.3,DoubleExponentialSmoothing	265.591922	1375.393398

Fig55: RMSE-DES

2point moving average model is considered best one till now with low RMSE values for both Rose and Sparkling wine

Triple Exponential Smoothing (Holt - Winter's Model):

TES is a time series forecasting method with 3 parameters alpha which is called as smoothing factor and beta which is called as trend and gamma which is seasonality.

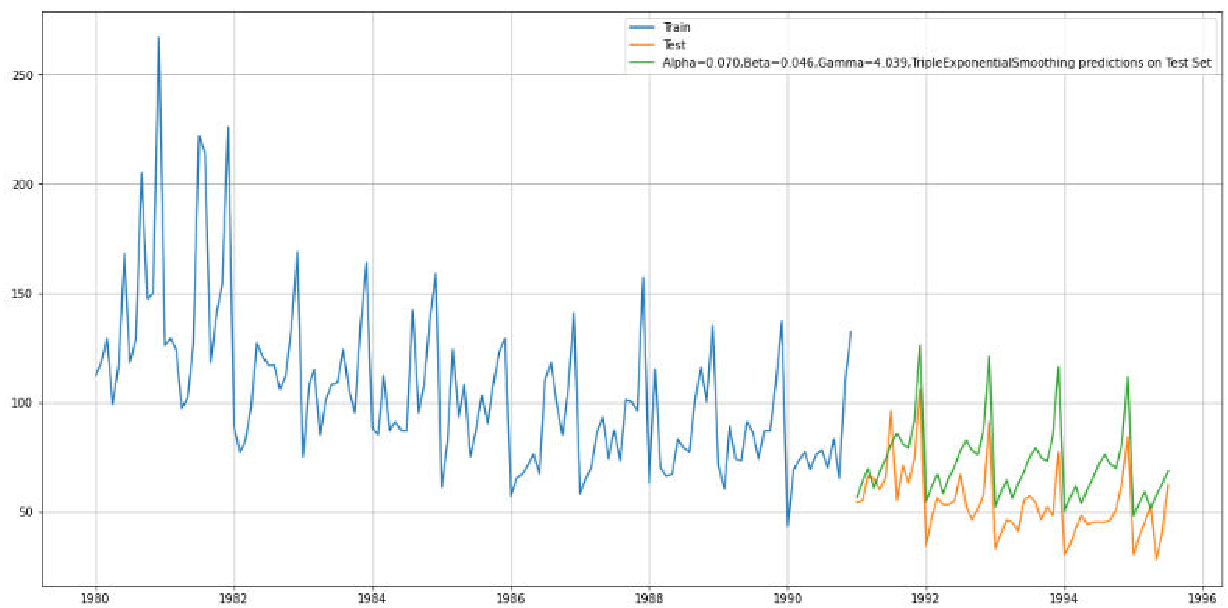


Fig56: TES-Rose1

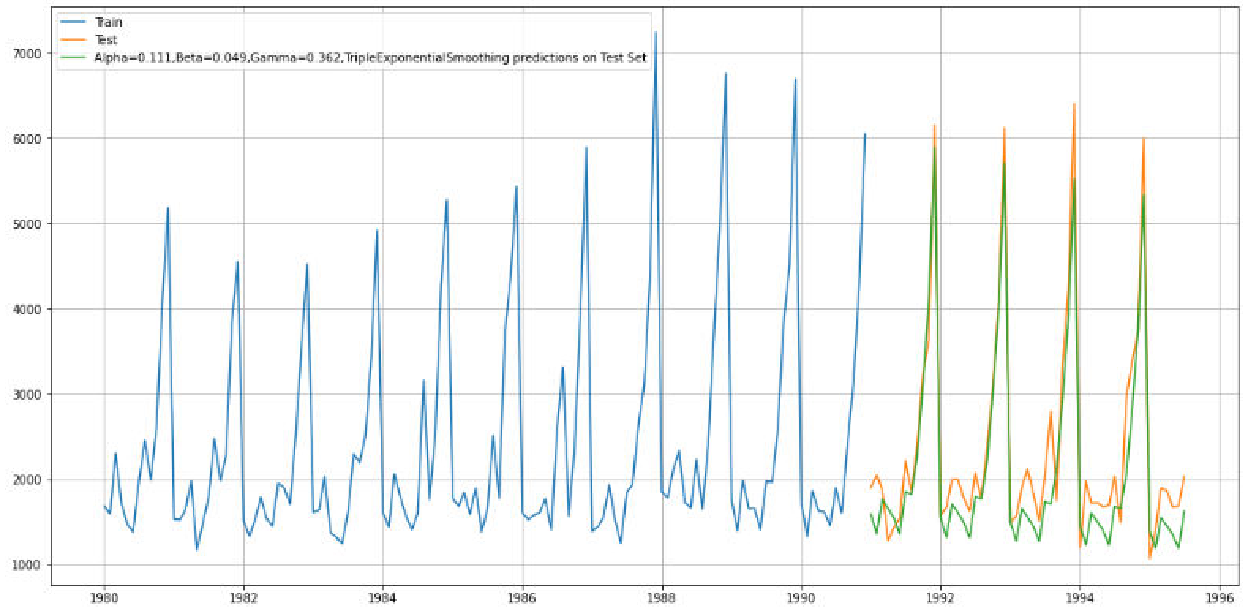


Fig57: TES-Sparkling1

<u>*RMSE for different alpha , beta and gamma values:*</u>

Performed validation with different Alpha , beta and gamma values. Below is the result .

	Alpha Values	Beta Values	Gamma Values	Train RMSE	Test RMSE
1	0.3	0.3	0.4	24.588120	10.158543
9	0.3	0.4	0.4	25.599445	10.361475
80	0.4	0.5	0.3	26.917917	13.375197
24	0.3	0.6	0.3	25.815213	15.497246
194	0.6	0.3	0.5	31.758130	17.249825

Fig58: TES-different values of alpha,beta,gamma-Rose

	Alpha Values	Beta Values	Gamma Values	Train RMSE	Test RMSE
0	0.3	0.3	0.3	397.797318	381.397300
17	0.3	0.5	0.4	452.801424	512.542557
376	0.8	1.0	0.3	790.740655	580.266110
66	0.4	0.3	0.5	448.661280	592.153132
8	0.3	0.4	0.3	415.172097	605.110479

Fig58.1: TES-different values of alpha,beta,gamma-Sparkling

For Rose wines, Alpha=0.3 , beta=0.3 and gamma=0.4 gave the low RMSE value.
For Sparkling wines,Alpha=0.3 , beta=0.3 and gamma=0.3 gave the low RMSE value.
Below are the graphs :

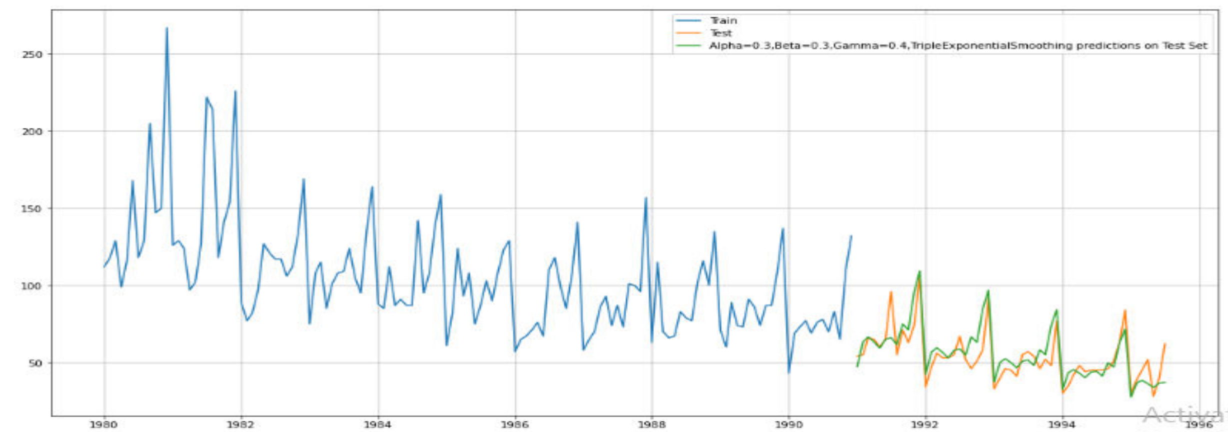


Fig59: TES-with optimal of alpha,beta,gamma-Rose

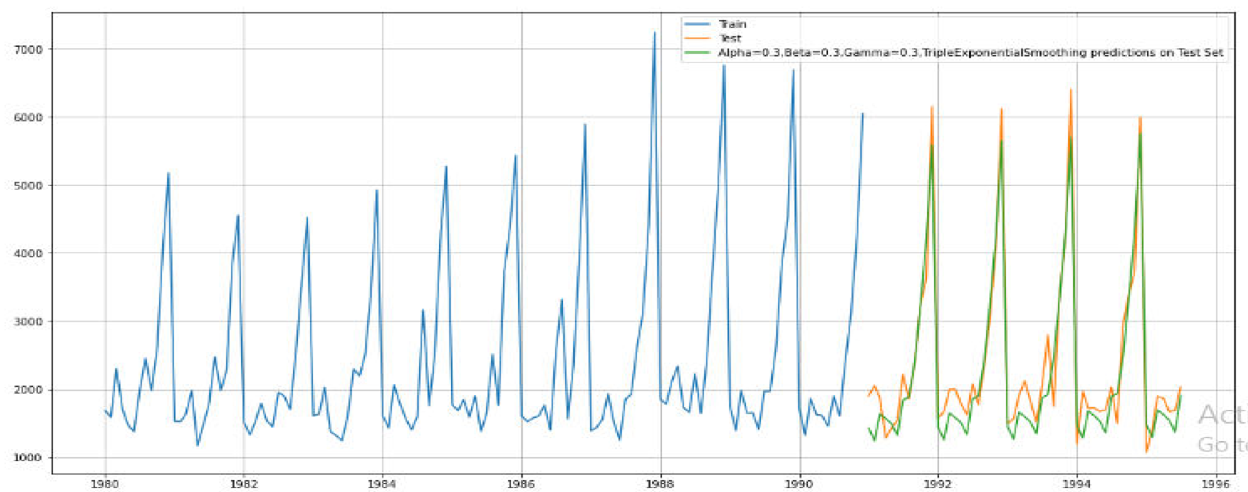


Fig59.1: TES-with optimal of alpha,beta,gamma-Sparkling

<u>RMSE</u>:

	Test RMSE-Rose	Test RMSE-Sparkling
RegressionOnTime	51.451050	1275.867052
NaiveModel	79.738550	3864.279352
SimpleAverageModel	79.738550	1275.081804
2pointTrailingMovingAverage	11.529409	813.400684
4pointTrailingMovingAverage	14.455221	1156.589694
6pointTrailingMovingAverage	14.572009	1283.927428
9pointTrailingMovingAverage	14.731209	1346.278315
Alpha=0.0987,SimpleExponentialSmoothing	36.816905	NaN
Alpha=0.0496,SimpleExponentialSmoothing	NaN	1316.034674
Alpha=0.1,SimpleExponentialSmoothing	36.848694	1375.393398
Alpha=0.3,Beta=0.3,DoubleExponentialSmoothing	265.591922	1375.393398
Alpha=0.070,Beta=0.046,Gamma=4.039,TripleExponentialSmoothing	20.359346	NaN
Alpha=0.111,Beta=0.049,Gamma=0.362,TripleExponentialSmoothing	NaN	402.946854
Alpha=0.3,Beta=0.3,Gamma=0.4,TripleExponentialSmoothing	10.158543	NaN
Alpha=0.3,Beta=0.3,Gamma=0.3,TripleExponentialSmoothing	NaN	361.397300

Fig59.2: TES-Rmse

Alpha=0.3,Beta=0.3,Gamma=0.4,TripleExponentialSmoothing is considered best one till now with low RMSE values for Rose and Alpha=0.3,Beta=0.3,Gamma=0.3,TripleExponentialSmoothing is considered best for Sparkling wine

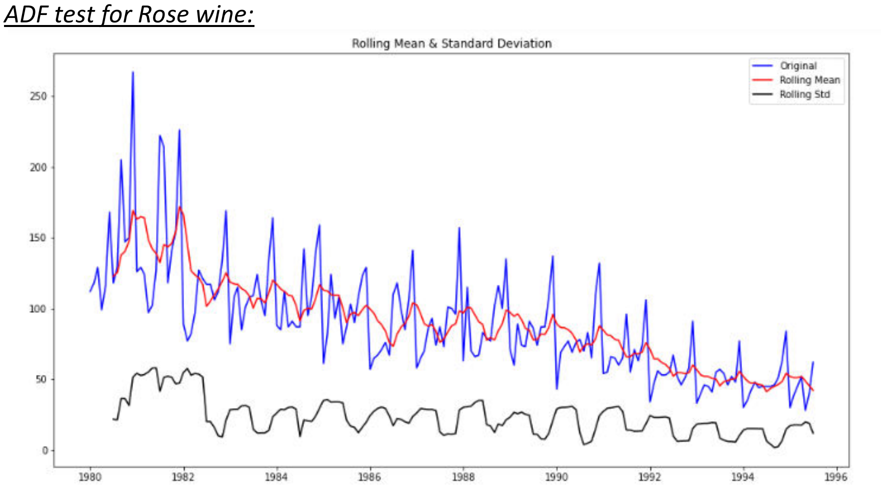
5) Check for the stationarity of the data on which the model is being built on using appropriate statistical tests and also mention the hypothesis for the statistical test. If the data is found to be non-stationary, take appropriate steps to make it stationary. Check the new data for stationarity and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

To check whether the series is stationary, we use the Augmented Dickey Fuller (ADF)test whose null and alternate hypothesis can be simplified to

- Null Hypothesis H0: Time Series is non-stationary
- Alternate Hypothesis Ha: Time Series is stationary

At our desired level of significance (chosen alpha value), we can test for stationary using the ADF test.
Given That alpha to be considered is 0.05(Confidence interval to be 95%)
If p-value from ADF test is less than alpha then reject the null hypothesis and hence data is said to be stationary.
If p-value from ADF test is greater than alpha then accept the null hypothesis and hence data is said to be non-stationary.
If data is non-stationary then we take appropriate levels of differencing to make a Time Series stationary
We can try various mathematical transformations to make the series stationary.
•Apply transformation and/or differencing.
•Check for stationarity.
•If the time series is not stationary repeat the process of differencing
•Remember, complicated transformations might give us a stationary series very easily but after the forecast values are obtained we need to get back to the original series by tracing back the transformation steps.

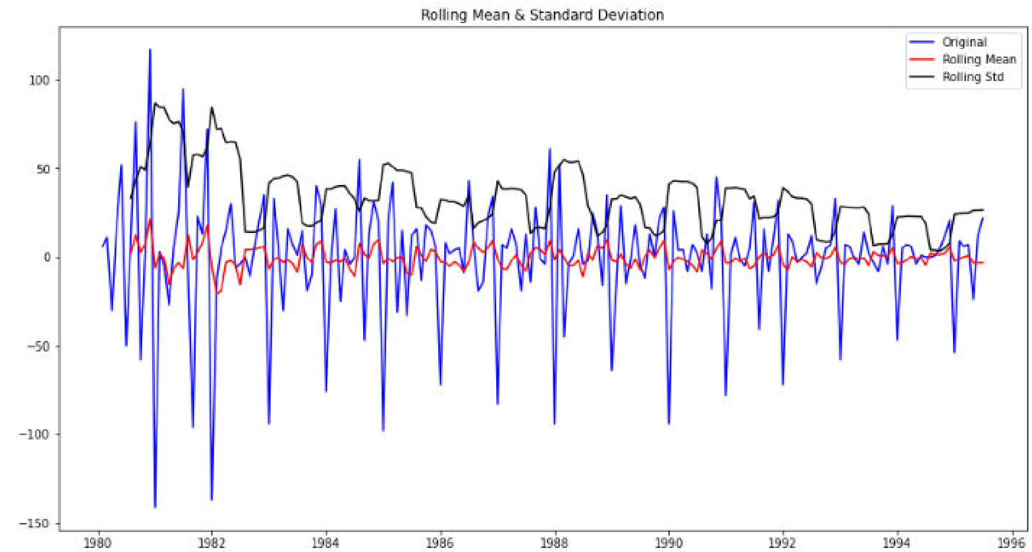
<u>ADF test for Rose wine:</u>




```
Results of Dickey-Fuller Test:
Test Statistic                  -1.874856
p-value                          0.343981
#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
```

Fig60: ADF-Rose

We see that at 5% significant level the Time Series is non-stationary.
Let us take a difference of order 1 and check whether the Time Series is stationary or not.

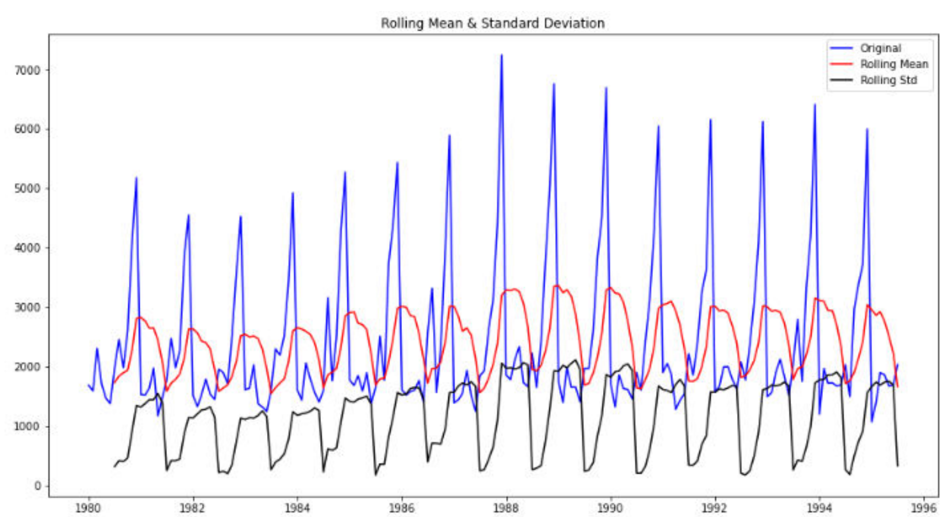



```
Results of Dickey-Fuller Test:
Test Statistic                -8.044139e+00
p-value                        1.813580e-12
#Lags Used                     1.200000e+01
Number of Observations Used    1.730000e+02
Critical Value (1%)           -3.468726e+00
Critical Value (5%)           -2.878396e+00
Critical Value (10%)          -2.575756e+00
dtype: float64
```

Fig61: ADF-Rose-after differencing-1

From results, we see differential data of order 1 is stationary since p-value is less than 0.05.

ADF test for Sparkling wine:

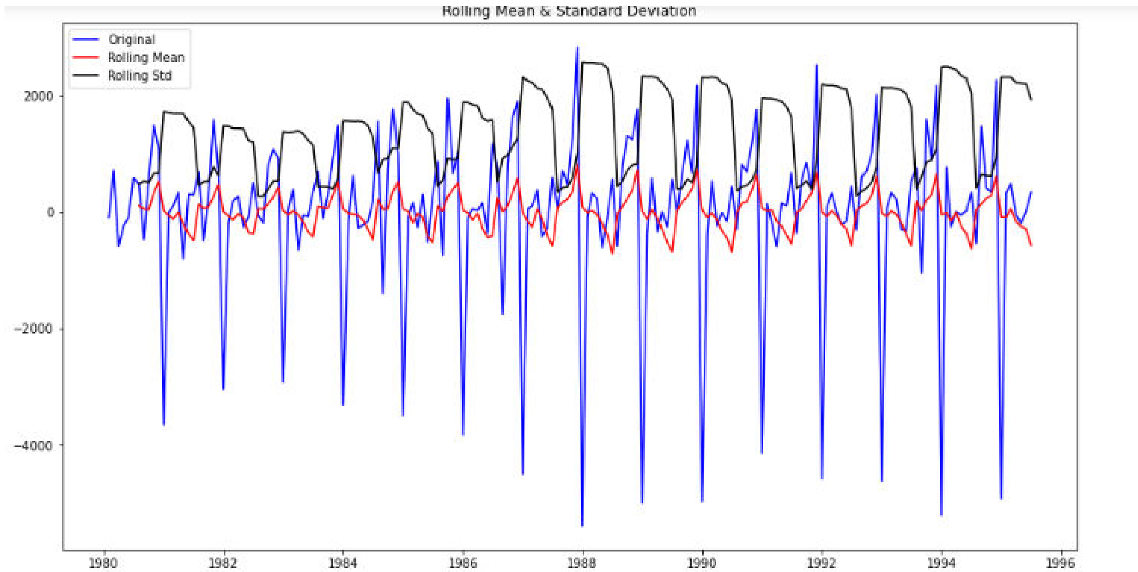



```
Results of Dickey-Fuller Test:
Test Statistic                  -1.360497
p-value                          0.601061
#Lags Used                      11.000000
Number of Observations Used    175.000000
Critical Value (1%)             -3.468280
Critical Value (5%)             -2.878202
Critical Value (10%)            -2.575653
dtype: float64
```

Fig62: ADF-Sparkling

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

Let us take a difference of order 1 and check whether the Time Series is stationary or not.




```
Results of Dickey-Fuller Test:
Test Statistic                 -45.050301
p-value                          0.000000
#Lags Used                      10.000000
Number of Observations Used    175.000000
Critical Value (1%)             -3.468280
Critical Value (5%)             -2.878202
Critical Value (10%)            -2.575653
dtype: float64
```


Fig63: ADF-Sparkling-after differencing-1

From results, we see differential data of order 1 is stationary since p-value is less than 0.05.

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.***

ARIMA and SRIMA are time series forecasting models where ARIMA stands for Auto Regressive Integrated Moving Average while the SARIMA is Seasonal ARIMA

Before building ARIMA model, data stationary is checked. Since non-stationary data is differenced with 1.

Building an ARIMA model: (Automated):

A grid of (p,d,q) is created with all possible combinations.

P and q ranges from 0 to 3 while d from 0,1

Some parameter combinations for the Model...

Model: (0, 1, 1)
Model: (0, 1, 2)
Model: (1, 1, 0)
Model: (1, 1, 1)
Model: (1, 1, 2)
Model: (2, 1, 0)
Model: (2, 1, 1)
Model: (2, 1, 2)

ARIMA model is built on the train data and fit to forecast on test data.

Parameter considered for evaluation is RMSE(Root Mean Square Error).

p,d,q combination that results in less AIC value is chosen as the best parameter values for building model.

	param	AIC
2	(0, 1, 2)	1279.671529
5	(1, 1, 2)	1279.870723
4	(1, 1, 1)	1280.57423
7	(2, 1, 1)	1281.507862
8	(2, 1, 2)	1281.870722
1	(0, 1, 1)	1282.309832
6	(2, 1, 0)	1298.611034
3	(1, 1, 0)	1317.350311
0	(0, 1, 0)	1333.154673

Fig64: AIC Rose

	param	AIC
0	(0, 1, 0)	1281.870722
1	(0, 1, 1)	1281.870722
2	(0, 1, 2)	1281.870722
3	(1, 1, 0)	1281.870722
4	(1, 1, 1)	1281.870722
5	(1, 1, 2)	1281.870722
6	(2, 1, 0)	1281.870722
7	(2, 1, 1)	1281.870722
8	(2, 1, 2)	1281.870722

Fig65: AIC Sparkling

For Rose, pdq values with low AIC value is (0.1.2) while for Sparkling its (0,1,0)
Models are built using the above p,d,q values.

Model Summary-Rose:

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                   Rose   No. Observations:                  132
Model:                 ARIMA(0, 1, 2)   Log Likelihood                -636.836
Date:                Sat, 24 Dec 2022   AIC                           1279.672
Time:                        22:27:36   BIC                           1288.297
Sample:                    01-01-1980   HQIC                          1283.176
                         - 12-01-1990
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ma.L1         -0.6970      0.072     -9.689      0.000      -0.838      -0.556
ma.L2         -0.2042      0.073     -2.794      0.005      -0.347      -0.061
sigma2       965.8407     88.305     10.938      0.000     792.766    1138.915
===================================================================================
Ljung-Box (L1) (Q):                   0.14   Jarque-Bera (JB):                39.24
Prob(Q):                              0.71   Prob(JB):                         0.00
Heteroskedasticity (H):               0.36   Skew:                             0.82
Prob(H) (two-sided):                  0.00   Kurtosis:                         5.13
===================================================================================

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

Fig64.1: Rose summary(0,1,2)

Model Summary-Sparkling:

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                   Rose   No. Observations:                  132
Model:                 ARIMA(0, 1, 2)   Log Likelihood                -636.836
Date:                Sat, 24 Dec 2022   AIC                           1279.672
Time:                        22:27:36   BIC                           1288.297
Sample:                    01-01-1980   HQIC                          1283.176
                         - 12-01-1990
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ma.L1         -0.6970      0.072     -9.689      0.000      -0.838      -0.556
ma.L2         -0.2042      0.073     -2.794      0.005      -0.347      -0.061
sigma2       965.8407     88.305     10.938      0.000     792.766    1138.915
===================================================================================
Ljung-Box (L1) (Q):                   0.14   Jarque-Bera (JB):                39.24
Prob(Q):                              0.71   Prob(JB):                         0.00
Heteroskedasticity (H):               0.36   Skew:                             0.82
Prob(H) (two-sided):                  0.00   Kurtosis:                         5.13
===================================================================================

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

Fig65.1: Sparkling summary(0,1,0)

RMSE post the Automated ARIMA MODEL:

	Test RMSE-Rose	Test RMSE-Sparkling
RegressionOnTime	51.451050	1275.867052
NaiveModel	79.738550	3864.279352
SimpleAverageModel	79.738550	1275.081804
2pointTrailingMovingAverage	11.529409	813.400684
4pointTrailingMovingAverage	14.455221	1156.589694
6pointTrailingMovingAverage	14.572009	1283.927428
9pointTrailingMovingAverage	14.731209	1346.278315
Alpha=0.0987,SimpleExponentialSmoothing	36.816905	NaN
Alpha=0.0496,SimpleExponentialSmoothing	NaN	1316.034674
Alpha=0.1,SimpleExponentialSmoothing	36.848694	1375.393398
Alpha=0.1,SimpleExponentialSmoothing	36.848694	1375.393398
Alpha=0.3,Beta=0.3,DoubleExponentialSmoothing	265.591922	1375.393398
Alpha=0.070,Beta=0.046,Gamma=4.039,TripleExponentialSmoothing	20.359346	NaN
Alpha=0.111,Beta=0.049,Gamma=0.362,TripleExponentialSmoothing	NaN	402.946854
Alpha=0.3,Beta=0.3,Gamma=0.4,TripleExponentialSmoothing	10.158543	NaN
Alpha=0.3,Beta=0.3,Gamma=0.3,TripleExponentialSmoothing	NaN	361.397300
ARIMA_R(0,1,2)	37.327049	NaN
ARIMA_R(0,1,0)	NaN	3864.279352

Fig66: ARIMA-RMSE

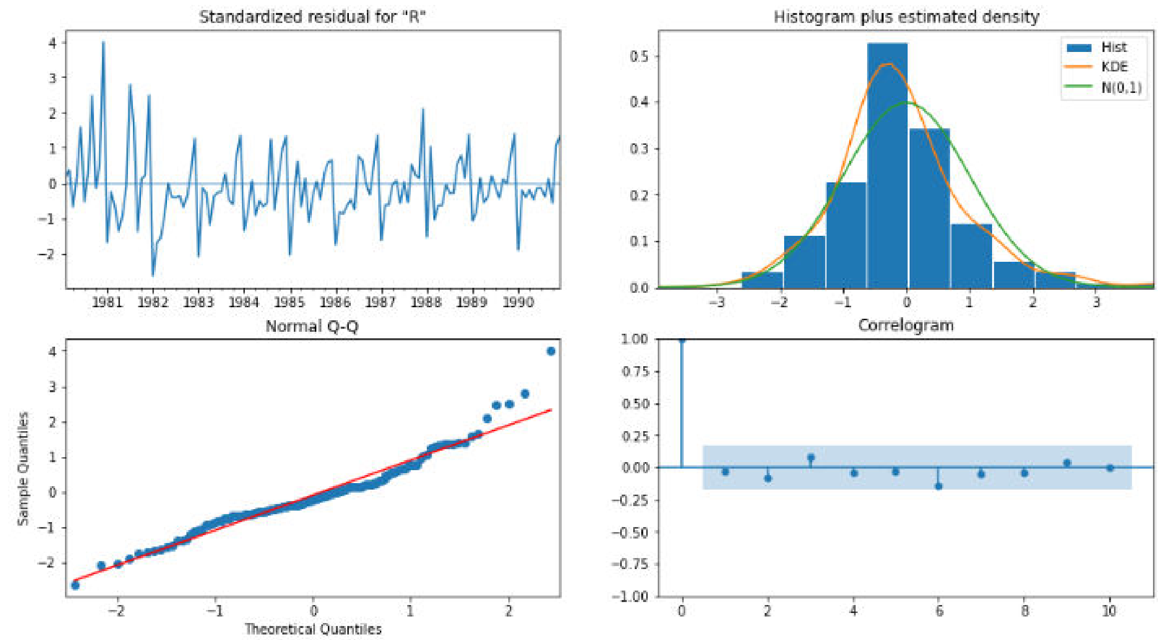


Fig67: ARIMA-Diagnostic-Rose

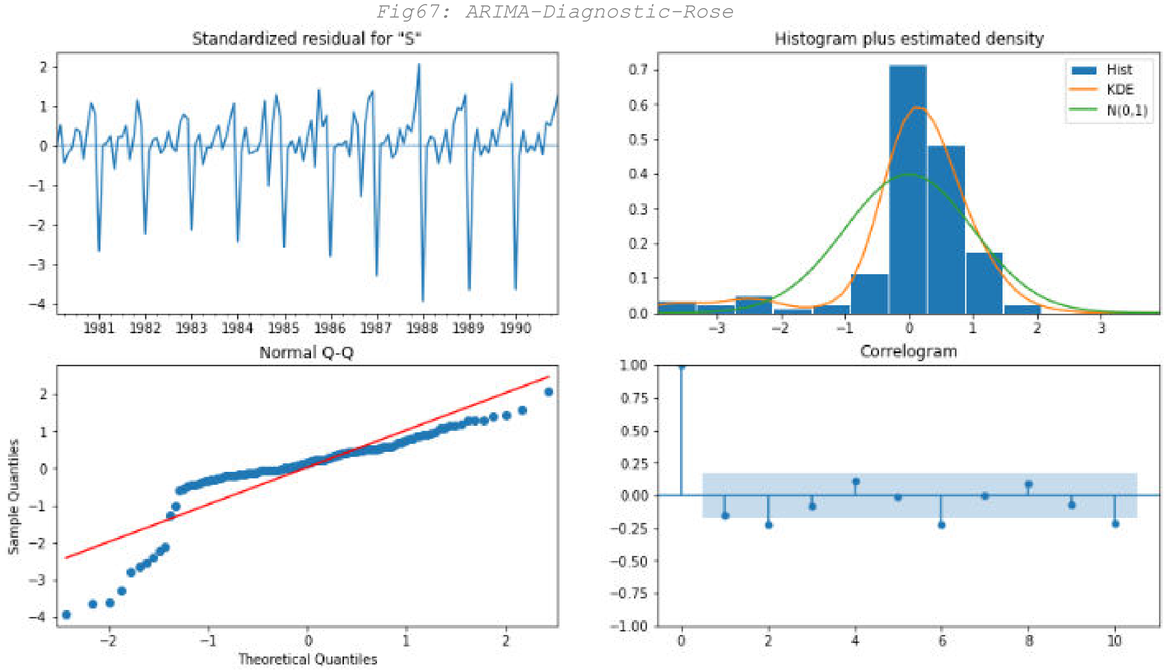


Fig68: ARIMA-Diagnostic-Sparkling

ACF Plot:

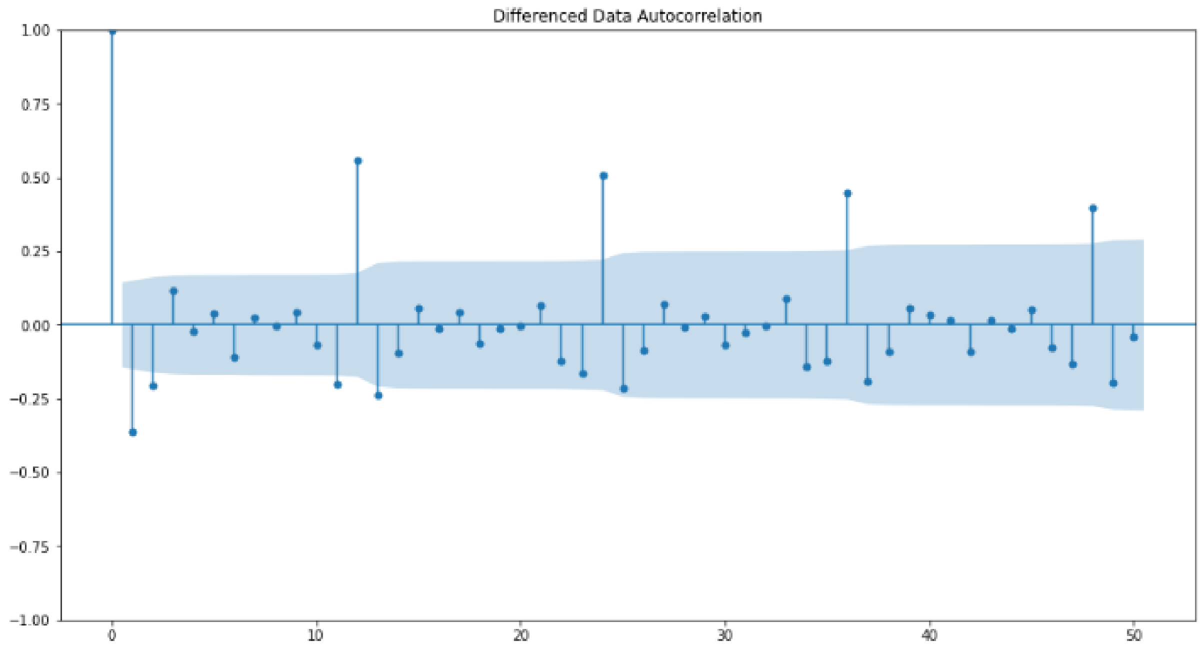


Fig69: ACF plot-rose

From above plot, we can see some seasonality at 1 and 12. We will run our auto SARIMA models by setting seasonality both as 1 and 12

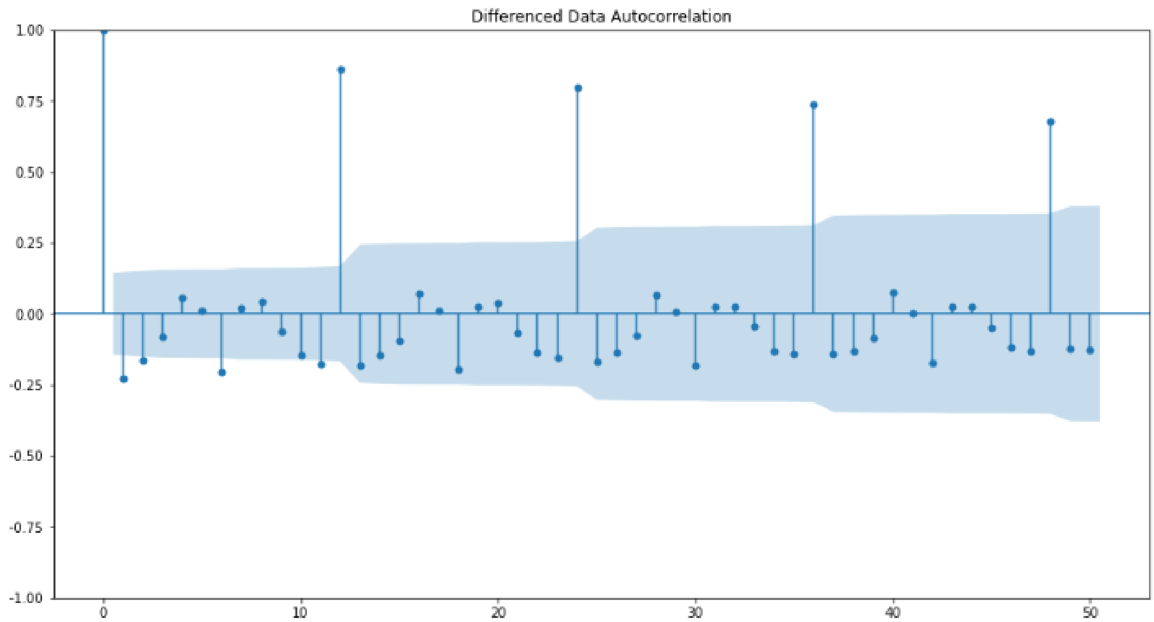


Fig70: ACF plot-Sparkling

From above plot, we can see some seasonality at 1 and 12. We will run our auto SARIMA models by setting seasonality both as 1 and 12
Observation:
p value from PACF plot is 0 as we can see there is sharp decline from the original to lag 1
q value from ACF plot is 0 as we can see there is sharp decline from the original to lag 1
Examples of some parameter combinations for Model...
Model: (0, 0, 1)(0, 0, 1, 6)

Model: (0, 0, 2)(0, 0, 2, 6)
Model: (0, 1, 0)(0, 1, 0, 6)
Model: (0, 1, 1)(0, 1, 1, 6)
Model: (0, 1, 2)(0, 1, 2, 6)
Model: (1, 0, 0)(1, 0, 0, 6)
Model: (1, 0, 1)(1, 0, 1, 6)
Model: (1, 0, 2)(1, 0, 2, 6)
Model: (1, 1, 0)(1, 1, 0, 6)
Model: (1, 1, 1)(1, 1, 1, 6)
Model: (1, 1, 2)(1, 1, 2, 6)
Model: (2, 0, 0)(2, 0, 0, 6)
Model: (2, 0, 1)(2, 0, 1, 6)
Model: (2, 0, 2)(2, 0, 2, 6)
Model: (2, 1, 0)(2, 1, 0, 6)
Model: (2, 1, 1)(2, 1, 1, 6)
Model: (2, 1, 2)(2, 1, 2, 6)

<u>SARIMA-ROSE</u>

	param	seasonal	AIC
107	(0, 1, 2)	(2, 1, 2, 12)	774.969119
215	(1, 1, 2)	(2, 1, 2, 12)	776.940108
323	(2, 1, 2)	(2, 1, 2, 12)	776.996101
269	(2, 0, 2)	(2, 1, 2, 12)	780.716945
161	(1, 0, 2)	(2, 1, 2, 12)	780.992967

<u>SARIMA-SPARKLING</u>

	param	seasonal	AIC
203	(1, 1, 2)	(0, 1, 2, 12)	1382.34778
95	(0, 1, 2)	(0, 1, 2, 12)	1382.484254
209	(1, 1, 2)	(1, 1, 2, 12)	1384.137874
311	(2, 1, 2)	(0, 1, 2, 12)	1384.317618
101	(0, 1, 2)	(1, 1, 2, 12)	1384.398867

For rose, the least AIC is for combination- (0, 1, 2) (1, 1, 2, 6) and (0, 1, 2) (2, 1, 2, 12)

For Sparking, the AIC value is least for combination - (0, 1, 2)(1,1,2,6) and (1, 1, 2) (0, 1, 2, 12)

<u>SARIMA-Rose:</u>

```
                                     SARIMAX Results
==========================================================================================
Dep. Variable:                                  y   No. Observations:                  132
Model:             SARIMAX(0, 1, 2)x(1, 1, 2, 6)   Log Likelihood                -472.310
Date:                           Sat, 24 Dec 2022   AIC                            956.620
Time:                                   23:55:35   BIC                            972.823
Sample:                                        0   HQIC                           963.192
                                           - 132
Covariance Type:                             opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ma.L1         -0.8657      0.129     -6.711      0.000      -1.119      -0.613
ma.L2         -0.2372      0.105     -2.259      0.024      -0.443      -0.031
ar.S.L6       -0.9513      0.015    -61.761      0.000      -0.982      -0.921
ma.S.L6        0.3785      0.154      2.457      0.014       0.077       0.680
ma.S.L12      -0.8505      0.112     -7.590      0.000      -1.070      -0.631
sigma2       197.7246     53.589      3.690      0.000      92.691     302.758
===================================================================================
Ljung-Box (L1) (Q):                   0.00   Jarque-Bera (JB):                 1.96
Prob(Q):                              0.96   Prob(JB):                         0.38
Heteroskedasticity (H):               0.62   Skew:                             0.32
Prob(H) (two-sided):                  0.16   Kurtosis:                         3.09
===================================================================================

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

Fig71: SARIMA-Rose-6

<u>SARIMA-Sparkling:</u>

```
                                     SARIMAX Results
==========================================================================================
Dep. Variable:                                  y   No. Observations:                  132
Model:             SARIMAX(0, 1, 2)x(1, 1, 2, 6)   Log Likelihood                -814.465
Date:                           Sat, 24 Dec 2022   AIC                           1640.931
Time:                                   23:55:37   BIC                           1657.134
Sample:                                        0   HQIC                          1647.503
                                           - 132
Covariance Type:                             opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ma.L1         -0.7629      0.107     -7.153      0.000      -0.972      -0.554
ma.L2         -0.1424      0.113     -1.262      0.207      -0.364       0.079
ar.S.L6       -1.0186      0.008   -119.905      0.000      -1.035      -1.002
ma.S.L6        0.1051      0.149      0.708      0.479      -0.186       0.396
ma.S.L12      -0.5578      0.083     -6.733      0.000      -0.720      -0.395
sigma2      1.556e+05   1.57e+04      9.898      0.000    1.25e+05    1.86e+05
===================================================================================
Ljung-Box (L1) (Q):                   0.02   Jarque-Bera (JB):                34.51
Prob(Q):                              0.90   Prob(JB):                         0.00
Heteroskedasticity (H):               1.82   Skew:                             0.61
Prob(H) (two-sided):                  0.07   Kurtosis:                         5.46
===================================================================================

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

Fig72: SARIMA-Sparkling-6

SARIMA Rose-12:

```
                                      SARIMAX Results
==========================================================================================
Dep. Variable:                                     y   No. Observations:                  132
Model:             SARIMAX(0, 1, 2)x(2, 1, 2, 12)   Log Likelihood                -380.485
Date:                            Sun, 25 Dec 2022   AIC                            774.969
Time:                                    00:16:43   BIC                            792.622
Sample:                                         0   HQIC                           782.094
                                            - 132
Covariance Type:                              opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ma.L1         -0.9524      0.184     -5.166      0.000      -1.314      -0.591
ma.L2         -0.0764      0.126     -0.605      0.545      -0.324       0.171
ar.S.L12       0.0480      0.177      0.271      0.786      -0.299       0.395
ar.S.L24      -0.0419      0.028     -1.513      0.130      -0.096       0.012
ma.S.L12      -0.7526      0.301     -2.503      0.012      -1.342      -0.163
ma.S.L24      -0.0721      0.204     -0.354      0.723      -0.472       0.327
sigma2       187.8702     45.278      4.149      0.000      99.127     276.613
===================================================================================
Ljung-Box (L1) (Q):                   0.06   Jarque-Bera (JB):                 4.86
Prob(Q):                              0.81   Prob(JB):                         0.09
Heteroskedasticity (H):               0.91   Skew:                             0.41
Prob(H) (two-sided):                  0.79   Kurtosis:                         3.77
===================================================================================

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

Fig73: SARIMA-Rose-12

SARIMA-Sparkling-12:

```
                                      SARIMAX Results
==========================================================================================
Dep. Variable:                                     y   No. Observations:                  132
Model:             SARIMAX(1, 1, 2)x(0, 1, 2, 12)   Log Likelihood                -685.174
Date:                            Sun, 25 Dec 2022   AIC                           1382.348
Time:                                    00:16:45   BIC                           1397.479
Sample:                                         0   HQIC                          1388.455
                                            - 132
Covariance Type:                              opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.5507      0.287     -1.922      0.055      -1.112       0.011
ma.L1         -0.1612      0.235     -0.687      0.492      -0.621       0.299
ma.L2         -0.7218      0.175     -4.132      0.000      -1.064      -0.379
ma.S.L12      -0.4062      0.092     -4.401      0.000      -0.587      -0.225
ma.S.L24      -0.0274      0.138     -0.198      0.843      -0.298       0.243
sigma2      1.705e+05   2.45e+04      6.956      0.000    1.22e+05    2.19e+05
===================================================================================
Ljung-Box (L1) (Q):                   0.00   Jarque-Bera (JB):                13.48
Prob(Q):                              0.95   Prob(JB):                         0.00
Heteroskedasticity (H):               0.89   Skew:                             0.60
Prob(H) (two-sided):                  0.75   Kurtosis:                         4.44
===================================================================================

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

Fig74: SARIMA-Sparkling-12

RMSE after SARIMA:

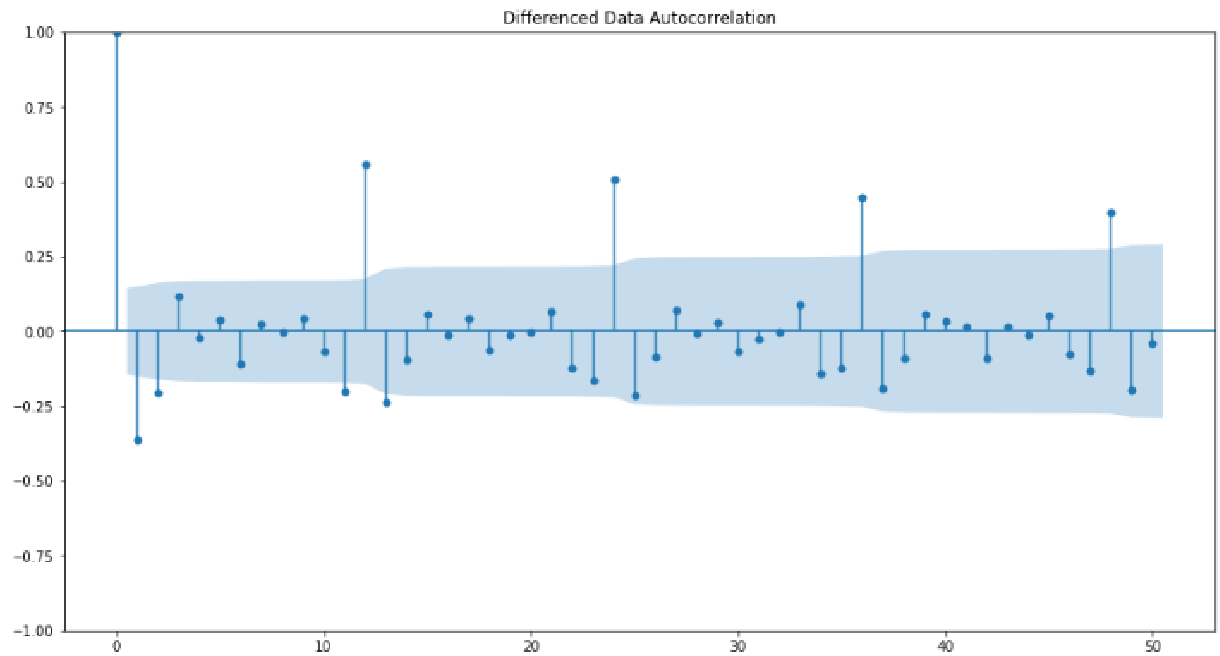
	Test RMSE-Rose	Test RMSE-Sparkling
RegressionOnTime	51.451050	1275.867052
NaiveModel	79.738550	3864.279352
SimpleAverageModel	79.738550	1275.081804
2pointTrailingMovingAverage	11.529409	813.400684
4pointTrailingMovingAverage	14.455221	1156.589694
6pointTrailingMovingAverage	14.572009	1283.927428
9pointTrailingMovingAverage	14.731209	1346.278315
Alpha=0.0987,SimpleExponentialSmoothing	36.816905	NaN
Alpha=0.0496,SimpleExponentialSmoothing	NaN	1316.034674
Alpha=0.1,SimpleExponentialSmoothing	36.848694	1375.393398
Alpha=0.1,SimpleExponentialSmoothing	36.848694	1375.393398
Alpha=0.3,Beta=0.3,DoubleExponentialSmoothing	265.591922	1375.393398
Alpha=0.070,Beta=0.046,Gamma=4.039,TripleExponentialSmoothing	20.359346	NaN
Alpha=0.111,Beta=0.049,Gamma=0.362,TripleExponentialSmoothing	NaN	402.946854
Alpha=0.3,Beta=0.3,Gamma=0.4,TripleExponentialSmoothing	10.158543	NaN
Alpha=0.3,Beta=0.3,Gamma=0.3,TripleExponentialSmoothing	NaN	361.397300
ARIMA_R(0,1,0)	NaN	3864.279352
SARIMA(0,1,2)(1,1,2,6)	18.444903	558.345168
ARIMA_R(0,1,2)	37.327049	NaN
SARIMA(0,1,2)(2,1,2,12)	16.519152	NaN
SARIMA(1,1,2)(0,1,2,12)	NaN	382.576754

Fig75: RMSE-SARIMA

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***

ACF Plot: Auto Correlation Function is the correlation between a time series with the lagged version of itself. The autocorrelation function (ACF) evaluates the correlation between observations in a time series over a given range of lags. It is used to determine a time series' randomness and stationarity
In an ACF plot, each bar represents the size and direction of the connection. Bars that cross the red line are statistically significant.
PACF plot: Partial Auto correlation Function gives the partial correlation of a stationary time series with its own lagged values regressed the values of the time series at all shorter lag. It contrasts with the autocorrelation function, which does not control for other lags.

ACF plot and PACF plot for Rose wine data:



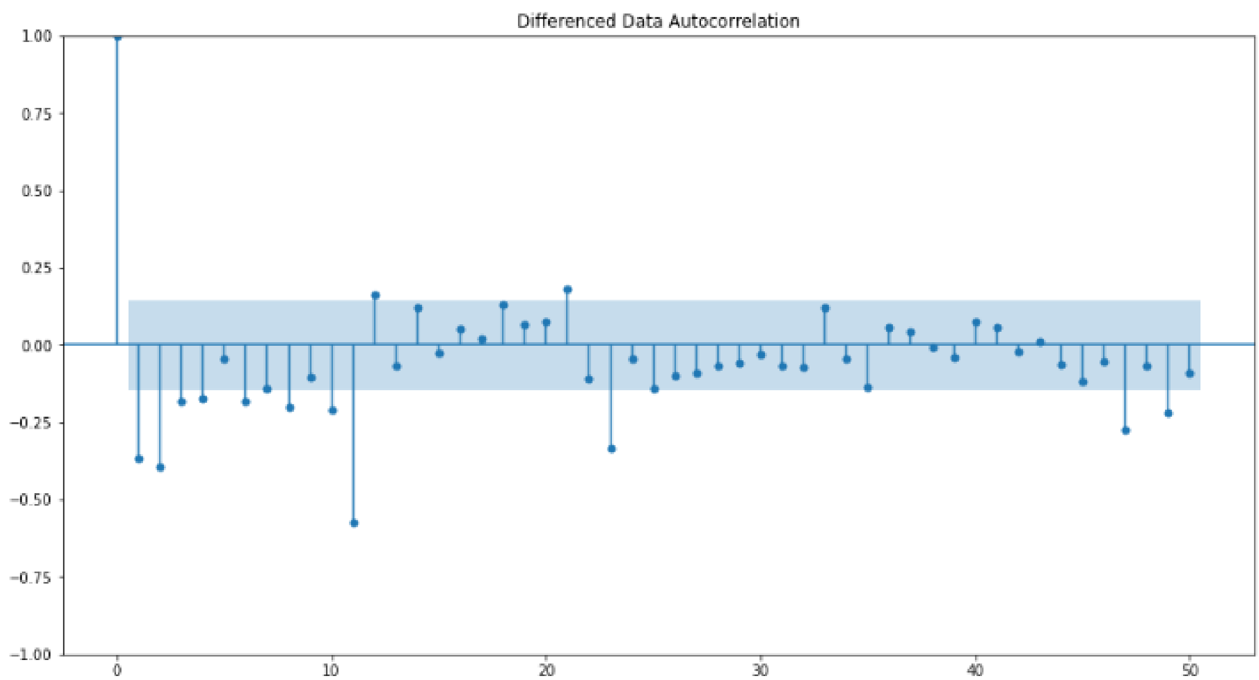


Fig76: ACF-PACF-ROSE

Here, we have taken alpha=0.05.
The Auto-Regressive parameter in an ARIMA model is 'p' which comes from the significant lag before which the PACF plot cuts-off to 2.
The Moving-Average parameter in an ARIMA model is 'q' which comes from the significant lag before the ACF plot cuts-off to 2.
By looking at the above plots, we can say that PACF plot cuts-off at lag 2 and ACF plot cuts-off at lag 2
Seasonality is observed at 6 and 12 and taking D as 6 and 12.

ARIMA model for Rose wine data with best parameters that are selected by looking at the ACF and the PACF plots:

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                   Rose   No. Observations:                  132
Model:                 ARIMA(2, 1, 2)   Log Likelihood                -635.935
Date:                Sun, 25 Dec 2022   AIC                           1281.871
Time:                        07:13:19   BIC                           1296.247
Sample:                    01-01-1980   HQIC                          1287.712
                         - 12-01-1990
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.4540      0.469     -0.969      0.333      -1.372       0.464
ar.L2          0.0001      0.170      0.001      0.999      -0.334       0.334
ma.L1         -0.2541      0.459     -0.554      0.580      -1.154       0.646
ma.L2         -0.5984      0.430     -1.390      0.164      -1.442       0.245
sigma2       952.1601     91.424     10.415      0.000     772.973    1131.347
===================================================================================
Ljung-Box (L1) (Q):                   0.02   Jarque-Bera (JB):                34.16
Prob(Q):                              0.88   Prob(JB):                         0.00
Heteroskedasticity (H):               0.37   Skew:                             0.79
Prob(H) (two-sided):                  0.00   Kurtosis:                         4.94
===================================================================================

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

Fig77: Rose-ARIMA-plot params

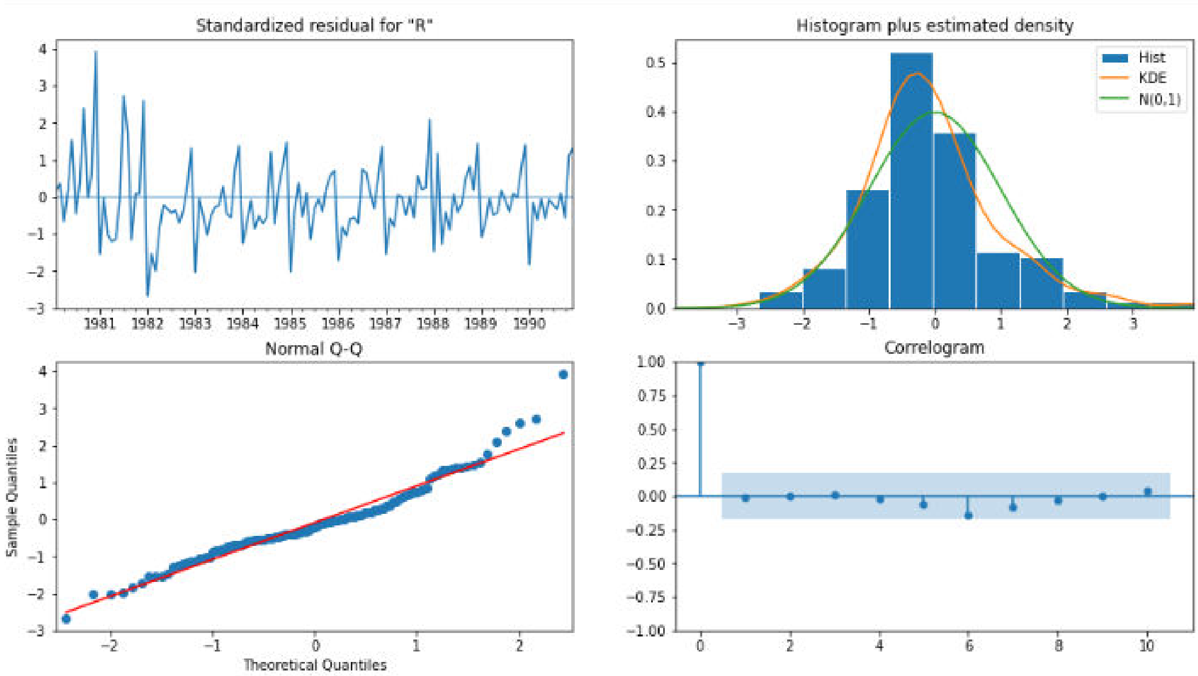
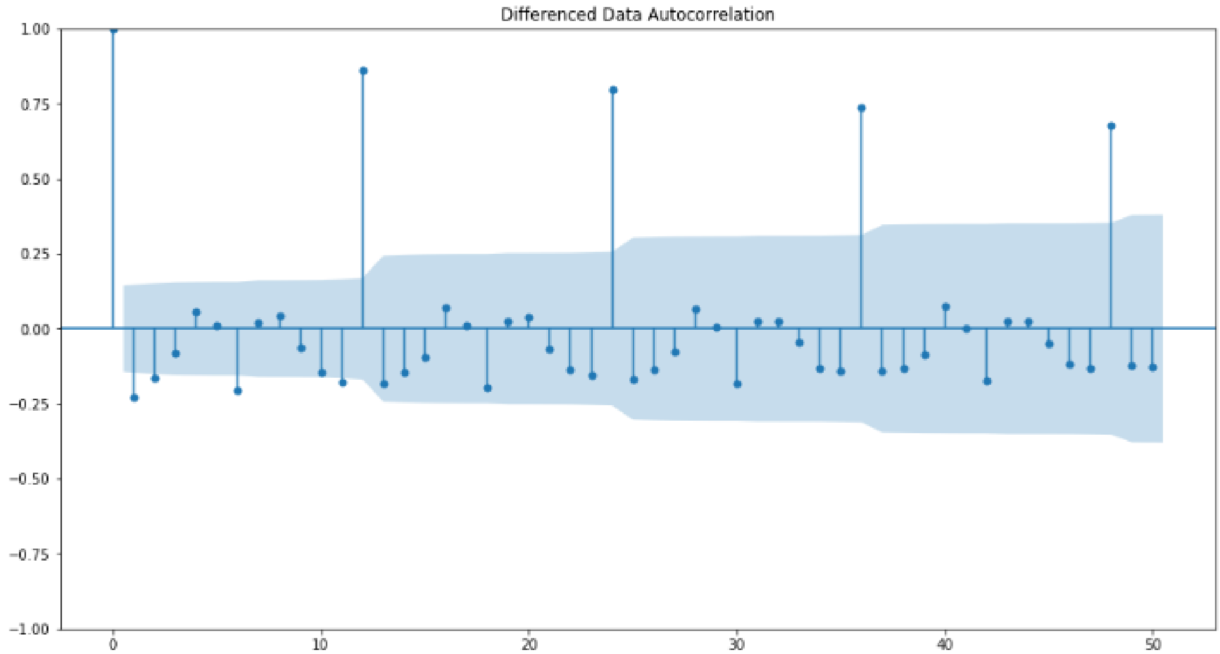


Fig78: Rose-Diagnostics-plot params

Sparkling data:



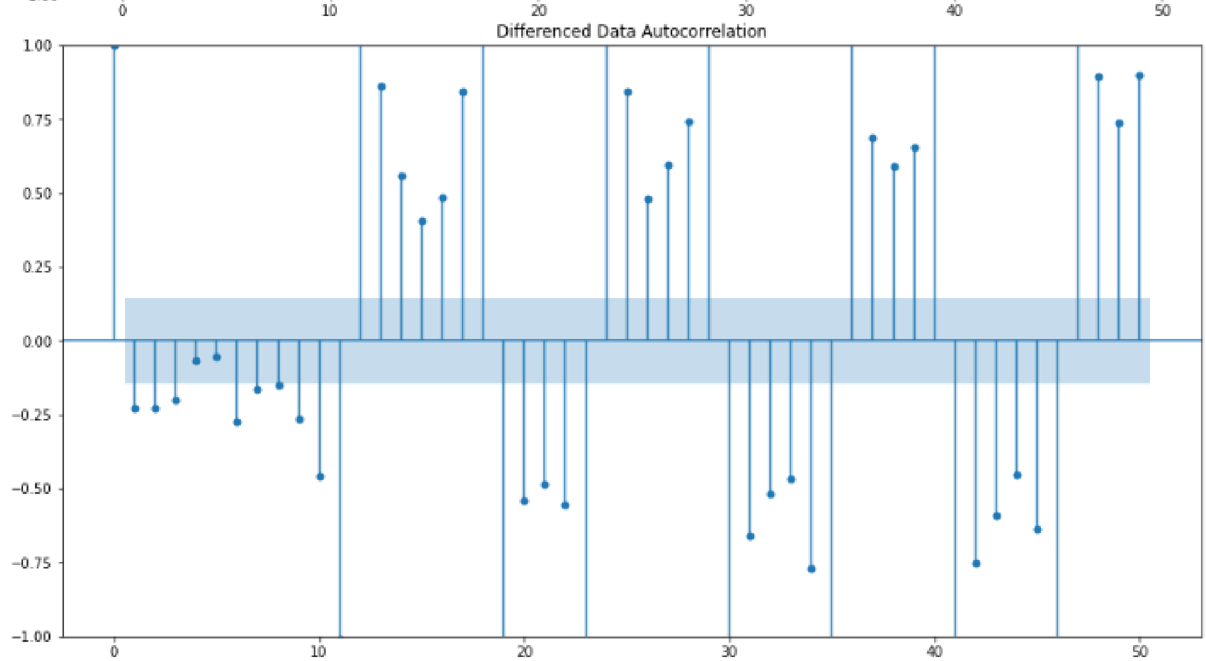


Fig79: ACF-PACF-sparkling

p value from PACF plot is 1 as we can see there is sharp decline from the original to lag 1
q value from ACF plot is 2 as we can see there is sharp decline from the original to lag 2
Seasonality value D is 6 and 12.

```
                               SARIMAX Results
==============================================================================
Dep. Variable:              Sparkling   No. Observations:                  132
Model:                 ARIMA(1, 1, 2)   Log Likelihood               -1113.264
Date:                Sun, 25 Dec 2022   AIC                           2234.527
Time:                        07:22:07   BIC                           2246.028
Sample:                    01-01-1980   HQIC                          2239.200
                         - 12-01-1990
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.2171      0.316      0.688      0.492      -0.402       0.836
ma.L1         -0.6943      0.385     -1.803      0.071      -1.449       0.060
ma.L2         -0.2852      0.372     -0.767      0.443      -1.014       0.443
sigma2      1.378e+06   1.34e+05     10.284      0.000    1.12e+06    1.64e+06
===================================================================================
Ljung-Box (L1) (Q):                   0.01   Jarque-Bera (JB):                11.16
Prob(Q):                              0.93   Prob(JB):                         0.00
Heteroskedasticity (H):               2.72   Skew:                             0.44
Prob(H) (two-sided):                  0.00   Kurtosis:                         4.12
===================================================================================

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

Fig80: Sparkling- ARIMA-plot params

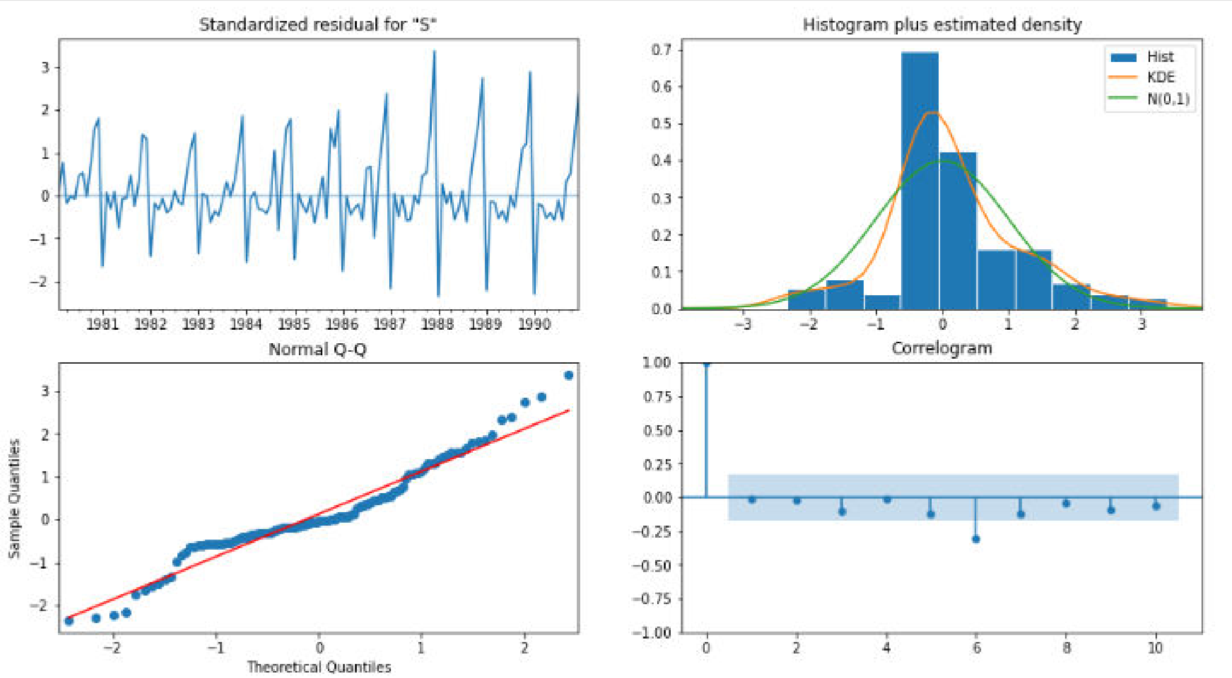


Fig81: Sparkling-Diagnostics-plot params

RMSE:

	Test RMSE-Rose	Test RMSE-Sparkling
RegressionOnTime	51.451050	1275.867052
NaiveModel	79.738550	3864.279352
SimpleAverageModel	79.738550	1275.081804
2pointTrailingMovingAverage	11.529409	813.400684
4pointTrailingMovingAverage	14.455221	1156.589694
6pointTrailingMovingAverage	14.572009	1283.927428
9pointTrailingMovingAverage	14.731209	1346.278315
Alpha=0.0987,SimpleExponentialSmoothing	36.816905	NaN
Alpha=0.0496,SimpleExponentialSmoothing	NaN	1316.034674
Alpha=0.1,SimpleExponentialSmoothing	36.848694	1375.393398
Alpha=0.3,Beta=0.3,DoubleExponentialSmoothing	265.591922	1375.393398
Alpha=0.070,Beta=0.046,Gamma=4.039,TripleExponentialSmoothing	20.359346	NaN
Alpha=0.111,Beta=0.049,Gamma=0.362,TripleExponentialSmoothing	NaN	402.946854
Alpha=0.3,Beta=0.3,Gamma=0.4,TripleExponentialSmoothing	10.158543	NaN
Alpha=0.3,Beta=0.3,Gamma=0.3,TripleExponentialSmoothing	NaN	361.397300
ARIMA_R(0,1,2)	37.327049	NaN
ARIMA_S(0,1,0)	NaN	3864.279352
ARIMA_R(2,1,2)	36.891832	NaN
ARIMA_S(1,1,2)	NaN	1316.597320

Fig82: RMSE -plot params

SARIMA-Plot Parameters with Seasonality at 6:

```
                                     SARIMAX Results
==========================================================================================
Dep. Variable:                                   y   No. Observations:                  132
Model:             SARIMAX(2, 1, 2)x(2, 1, 2, 6)   Log Likelihood                -470.903
Date:                             Sun, 25 Dec 2022   AIC                            959.806
Time:                                     08:30:20   BIC                            984.110
Sample:                                          0   HQIC                           969.664
                                             - 132
Covariance Type:                               opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.7002      0.120     -5.832      0.000      -0.936      -0.465
ar.L2          0.1834      0.105      1.745      0.081      -0.023       0.389
ma.L1          0.0658    424.138      0.000      1.000    -831.229     831.360
ma.L2         -0.9342    396.270     -0.002      0.998    -777.608     775.740
ar.S.L6       -0.9504      0.075    -12.710      0.000      -1.097      -0.804
ar.S.L12       0.0001      0.076      0.001      0.999      -0.148       0.148
ma.S.L6        0.3895      0.170      2.288      0.022       0.056       0.723
ma.S.L12      -0.8156      0.142     -5.756      0.000      -1.093      -0.538
sigma2       223.4544   9.48e+04      0.002      0.998   -1.85e+05    1.86e+05
===================================================================================
Ljung-Box (L1) (Q):                   0.01   Jarque-Bera (JB):                 2.24
Prob(Q):                              0.93   Prob(JB):                         0.33
Heteroskedasticity (H):               0.64   Skew:                             0.34
Prob(H) (two-sided):                  0.17   Kurtosis:                         3.15
===================================================================================

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

Fig83: SARIMA -Rose plot params1-6

```
                                      SARIMAX Results
==========================================================================================
Dep. Variable:                                 y   No. Observations:                  132
Model:             SARIMAX(1, 1, 2)x(1, 1, 2, 6)   Log Likelihood                -824.123
Date:                           Sun, 25 Dec 2022   AIC                           1662.247
Time:                                   08:30:41   BIC                           1681.150
Sample:                                        0   HQIC                          1669.914
                                           - 132
Covariance Type:                             opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.6035      0.155     -3.885      0.000      -0.908      -0.299
ma.L1         -0.1382      0.603     -0.229      0.819      -1.321       1.044
ma.L2         -0.8637      0.557     -1.551      0.121      -1.955       0.228
ar.S.L6       -0.9975      0.018    -55.530      0.000      -1.033      -0.962
ma.S.L6      893.7927    445.081      2.008      0.045      21.450    1766.135
ma.S.L12    4023.3187     98.844     40.704      0.000    3829.588    4217.050
sigma2         0.0114      0.007      1.616      0.106      -0.002       0.025
===================================================================================
Ljung-Box (L1) (Q):                   0.15   Jarque-Bera (JB):                 3.12
Prob(Q):                              0.70   Prob(JB):                         0.21
Heteroskedasticity (H):               1.21   Skew:                             0.02
Prob(H) (two-sided):                  0.56   Kurtosis:                         3.82
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
[2] Covariance matrix is singular or near-singular, with condition number 3.68e+17. Standard errors may be unstable.
```

Fig83.1: SARIMA -Sparkling plot params1-6

RMSE:

	Test RMSE-Rose	Test RMSE-Sparkling
RegressionOnTime	51.451050	1275.867052
NaiveModel	79.738550	3864.279352
SimpleAverageModel	79.738550	1275.081804
2pointTrailingMovingAverage	11.529409	813.400684
4pointTrailingMovingAverage	14.455221	1156.589694
6pointTrailingMovingAverage	14.572009	1283.927428
9pointTrailingMovingAverage	14.731209	1346.278315
Alpha=0.0987,SimpleExponentialSmoothing	36.816905	NaN
Alpha=0.0496,SimpleExponentialSmoothing	NaN	1316.034674
Alpha=0.1,SimpleExponentialSmoothing	36.848694	1375.393398
Alpha=0.3,Beta=0.3,DoubleExponentialSmoothing	265.591922	1375.393398
Alpha=0.070,Beta=0.046,Gamma=4.039,TripleExponentialSmoothing	20.359346	NaN
Alpha=0.111,Beta=0.049,Gamma=0.362,TripleExponentialSmoothing	NaN	402.946854
Alpha=0.3,Beta=0.3,Gamma=0.4,TripleExponentialSmoothing	10.158543	NaN
Alpha=0.3,Beta=0.3,Gamma=0.3,TripleExponentialSmoothing	NaN	361.397300
ARIMA_R(0,1,2)	37.327049	NaN
ARIMA_S(0,1,0)	NaN	3864.279352
ARIMA_R(2,1,2)	36.891832	NaN
ARIMA_S(1,1,2)	NaN	1316.597320
SARIMA(0,1,2)(1,1,2,6)	18.444903	558.345168
SARIMA(0,1,2)(2,1,2,12)	16.519152	NaN
SARIMA(1,1,2)(0,1,2,12)	NaN	382.576754
SARIMA(2,1,2)(2,1,2,6)	18.649293	NaN
SARIMA(1,1,2)(1,1,2,6)	NaN	334.815310

Fig84: RMSE with plot params-6

Seasonality with 12:

```
                                     SARIMAX Results
==========================================================================================
Dep. Variable:                                  y   No. Observations:                  132
Model:             SARIMAX(2, 1, 2)x(2, 1, 2, 12)   Log Likelihood                -379.498
Date:                            Sun, 25 Dec 2022   AIC                            776.996
Time:                                    08:35:06   BIC                            799.692
Sample:                                         0   HQIC                           786.156
                                            - 132
Covariance Type:                              opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.8551      0.146     -5.837      0.000      -1.142      -0.568
ar.L2         -0.0022      0.125     -0.017      0.986      -0.247       0.242
ma.L1          0.0120      0.184      0.066      0.948      -0.348       0.372
ma.L2         -0.9435      0.150     -6.291      0.000      -1.237      -0.650
ar.S.L12       0.0348      0.185      0.188      0.851      -0.328       0.398
ar.S.L24      -0.0459      0.029     -1.599      0.110      -0.102       0.010
ma.S.L12      -0.7224      0.333     -2.172      0.030      -1.374      -0.071
ma.S.L24      -0.0771      0.212     -0.363      0.716      -0.493       0.339
sigma2       192.1955     39.484      4.868      0.000     114.809     269.582
===================================================================================
Ljung-Box (L1) (Q):                   0.03   Jarque-Bera (JB):                 7.06
Prob(Q):                              0.86   Prob(JB):                         0.03
Heteroskedasticity (H):               0.87   Skew:                             0.45
Prob(H) (two-sided):                  0.71   Kurtosis:                         4.01
===================================================================================

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

Fig85: SARIMA-Rose-Seasonality12-plot params

```
                                     SARIMAX Results
==========================================================================================
Dep. Variable:                                  y   No. Observations:                  132
Model:             SARIMAX(1, 1, 2)x(1, 1, 2, 12)   Log Likelihood                -685.069
Date:                            Sun, 25 Dec 2022   AIC                           1384.138
Time:                                    08:35:30   BIC                           1401.790
Sample:                                         0   HQIC                          1391.263
                                            - 132
Covariance Type:                              opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.5751      0.285     -2.017      0.044      -1.134      -0.016
ma.L1         -0.1375      0.238     -0.578      0.563      -0.603       0.328
ma.L2         -0.7335      0.169     -4.346      0.000      -1.064      -0.403
ar.S.L12      -0.1807      1.545     -0.117      0.907      -3.209       2.847
ma.S.L12      -0.2324      1.552     -0.150      0.881      -3.274       2.809
ma.S.L24      -0.1009      0.662     -0.152      0.879      -1.399       1.197
sigma2      1.706e+05   2.46e+04      6.939      0.000    1.22e+05    2.19e+05
===================================================================================
Ljung-Box (L1) (Q):                   0.01   Jarque-Bera (JB):                13.27
Prob(Q):                              0.92   Prob(JB):                         0.00
Heteroskedasticity (H):               0.88   Skew:                             0.60
Prob(H) (two-sided):                  0.73   Kurtosis:                         4.43
===================================================================================

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

Fig85.1: SARIMA-Sparkling-Seasonality12-plot params

RMSE:

	Test RMSE-Rose	Test RMSE-Sparkling
RegressionOnTime	51.451050	1275.867052
NaiveModel	79.738550	3864.279352
SimpleAverageModel	79.738550	1275.081804
2pointTrailingMovingAverage	11.529409	813.400684
4pointTrailingMovingAverage	14.455221	1156.589694
6pointTrailingMovingAverage	14.572009	1283.927428
9pointTrailingMovingAverage	14.731209	1346.278315
Alpha=0.0987,SimpleExponentialSmoothing	36.816905	NaN
Alpha=0.0496,SimpleExponentialSmoothing	NaN	1316.034674
Alpha=0.1,SimpleExponentialSmoothing	36.848694	1375.393398
Alpha=0.3,Beta=0.3,DoubleExponentialSmoothing	265.591922	1375.393398
Alpha=0.070,Beta=0.046,Gamma=4.039,TripleExponentialSmoothing	20.359346	NaN
Alpha=0.111,Beta=0.049,Gamma=0.362,TripleExponentialSmoothing	NaN	402.946854
Alpha=0.3,Beta=0.3,Gamma=0.4,TripleExponentialSmoothing	10.158543	NaN
Alpha=0.3,Beta=0.3,Gamma=0.3,TripleExponentialSmoothing	NaN	361.397300
ARIMA_R(0,1,2)	37.327049	NaN
ARIMA_S(0,1,0)	NaN	3864.279352
ARIMA_R(2,1,2)	36.891832	NaN
ARIMA_S(1,1,2)	NaN	1316.597320
SARIMA(0,1,2)(1,1,2,6)	18.444903	558.345168
SARIMA(0,1,2)(2,1,2,12)	16.519152	NaN
SARIMA(1,1,2)(0,1,2,12)	NaN	382.576754
SARIMA(2,1,2)(2,1,2,6)	18.649293	NaN
SARIMA(1,1,2)(1,1,2,6)	NaN	334.815310
SARIMA(2,1,2)(2,1,2,12)	16.569775	NaN
SARIMA(1,1,1)(1,1,2,12)	NaN	401.515730

Fig86: RMSE-Seasonality12

Observations:
We can see from the above RMSE values, SARIMA(0,1,2)(2,1,2,12) is the optimal model for Rose wine type while SARIMA(1,1,2)(0,1,2,12) for Sparkling wine type.

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

Below is the RMSE values test results dataframe for all models:

	Test RMSE-Rose	Test RMSE-Sparkling
RegressionOnTime	51.451050	1275.867052
NaiveModel	79.738550	3864.279352
SimpleAverageModel	79.738550	1275.081804
2pointTrailingMovingAverage	11.529409	813.400684
4pointTrailingMovingAverage	14.455221	1156.589694
6pointTrailingMovingAverage	14.572009	1283.927428
9pointTrailingMovingAverage	14.731209	1346.278315
Alpha=0.0987,SimpleExponentialSmoothing	36.816905	NaN
Alpha=0.0496,SimpleExponentialSmoothing	NaN	1316.034674
Alpha=0.1,SimpleExponentialSmoothing	36.848694	1375.393398
Alpha=0.3,Beta=0.3,DoubleExponentialSmoothing	265.591922	1375.393398
Alpha=0.070,Beta=0.046,Gamma=4.039,TripleExponentialSmoothing	20.359346	NaN
Alpha=0.111,Beta=0.049,Gamma=0.362,TripleExponentialSmoothing	NaN	402.946854
Alpha=0.3,Beta=0.3,Gamma=0.4,TripleExponentialSmoothing	10.158543	NaN
Alpha=0.3,Beta=0.3,Gamma=0.3,TripleExponentialSmoothing	NaN	361.397300
ARIMA_R(0,1,2)	37.327049	NaN
ARIMA_S(0,1,0)	NaN	3864.279352
ARIMA_R(2,1,2)	36.891832	NaN
ARIMA_S(1,1,2)	NaN	1316.597320
SARIMA(0,1,2)(1,1,2,6)	18.444903	558.345168
SARIMA(0,1,2)(2,1,2,12)	16.519152	NaN
SARIMA(1,1,2)(0,1,2,12)	NaN	382.576754
SARIMA(2,1,2)(2,1,2,6)	18.649293	NaN
SARIMA(1,1,2)(1,1,2,6)	NaN	334.815310
SARIMA(2,1,2)(2,1,2,12)	16.569775	NaN
SARIMA(1,1,1)(1,1,2,12)	NaN	401.515730

Fig86.1: RMSE-Test data.

From above RMSE values,

- Alpha=0.3, Beta=0.3, Gamma=0.4, TripleExponentialSmoothing is the optimal model for Rose wine dataset as it resulted in low RMSE value.
- SARIMA(1,1,2)(1,1,2,6) is the optimal model for Sparkling wine dataset with low RMSE results

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.

Building TripleExponentialSmoothing model with Alpha=0.3, Beta=0.3, Gamma=0.4 parameters for whole Rose wine dataset as it resulted in low RMSE compared to other models

Predicted Values as below with seasonality as Multiplicative:

```
1995-08-01    47.582258
1995-09-01    44.331039
1995-10-01    43.260058
1995-11-01    49.520605
1995-12-01    67.320363
1996-01-01    28.997325
1996-02-01    32.482033
1996-03-01    35.320631
1996-04-01    31.107448
1996-05-01    33.483025
1996-06-01    35.728600
1996-07-01    39.489194
```

Fig87: Rose-Predicted values

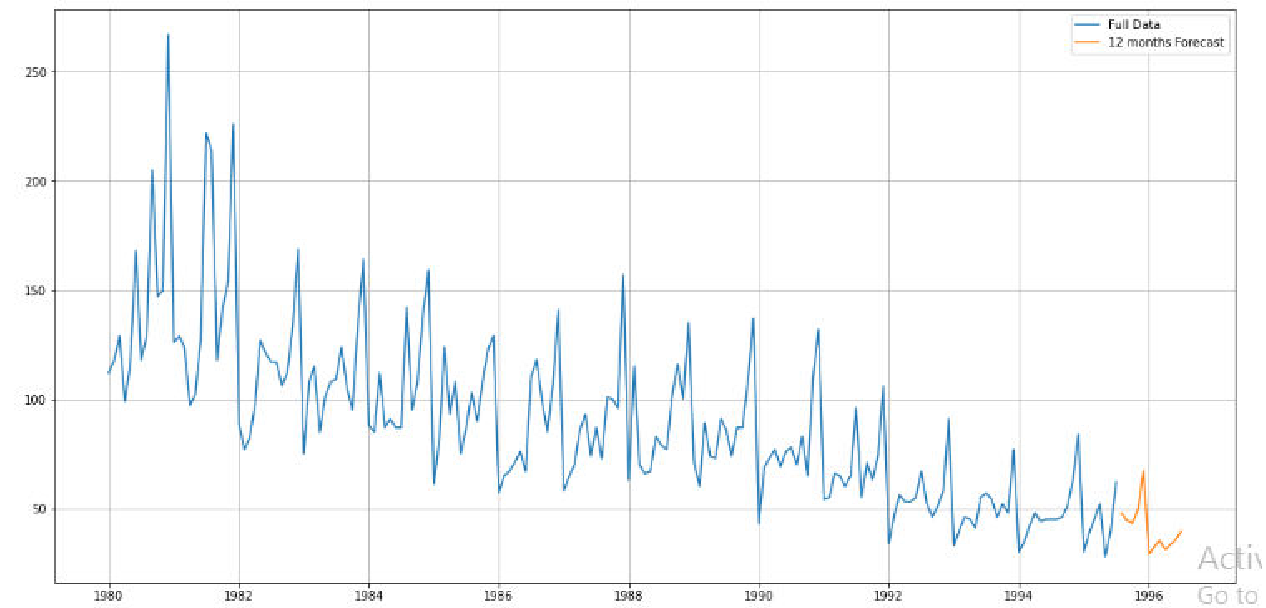


Fig88: Rose-Forecast 12 months

Predicted Values as below with seasonality as Additive:

```
1995-08-01    50.132262
1995-09-01    46.986428
1995-10-01    45.701327
1995-11-01    60.288510
1995-12-01    98.547665
1996-01-01    14.070885
1996-02-01    24.380181
1996-03-01    31.945905
1996-04-01    24.733174
1996-05-01    28.088762
1996-06-01    33.534757
1996-07-01    44.233175
```

Fig89: Rose-Predicted values1

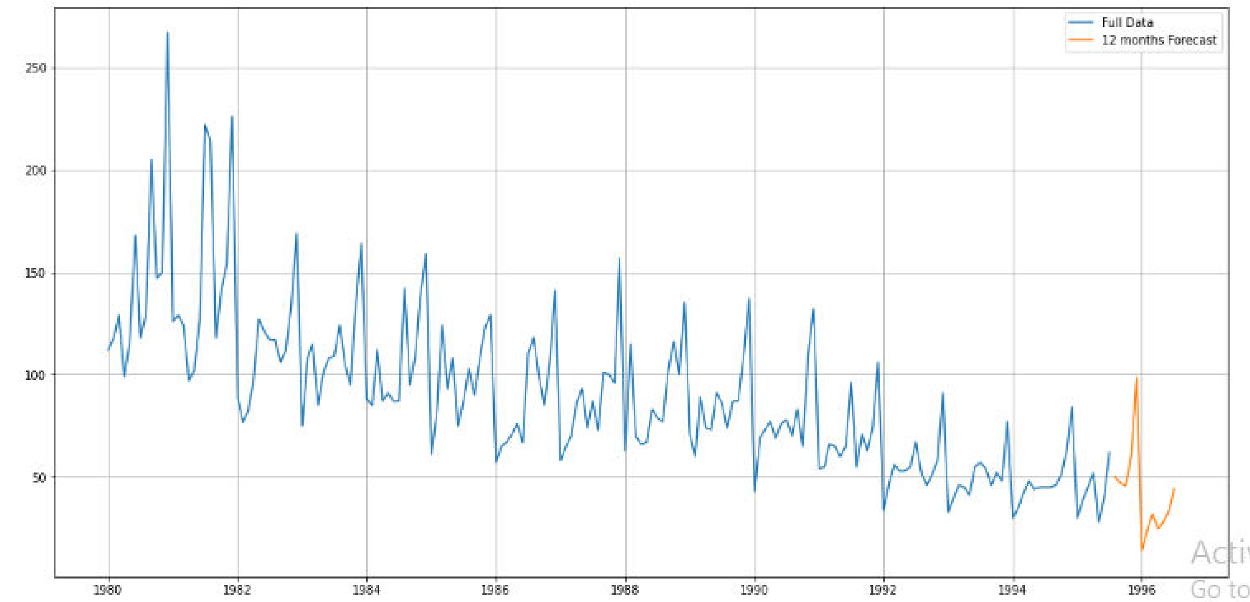


Fig90: Rose-Forecast 12 months-1

Building SARIMA(1,1,2)(1,1,2,6) for whole Sparkling wine dataset as it resulted in low RMSE compared to other models

Predicted Values as below:

Sparkling	mean	mean_se	mean_ci_lower	mean_ci_upper
1995-07-01	1876.060647	389.457904	1112.737182	2639.384113
1995-08-01	2478.217110	394.168624	1705.660804	3250.773416
1995-09-01	3293.555584	394.318119	2520.706272	4066.404896
1995-10-01	3933.325337	395.602520	3157.958646	4708.692027
1995-11-01	6132.783764	395.653854	5357.316460	6908.251067
1995-12-01	1244.081710	396.241414	467.462809	2020.700610
1996-01-01	1582.434484	396.438804	805.428707	2359.440261
1996-02-01	1836.383355	396.843319	1058.584743	2614.181968
1996-03-01	1819.012343	397.115073	1040.681102	2597.343583
1996-04-01	1664.539905	397.462382	885.527950	2443.551859
1996-05-01	1615.814946	397.763140	836.213517	2395.416375
1996-06-01	2016.543329	398.090836	1236.299627	2796.787031

Fig91: Sparkling 12months forecast values

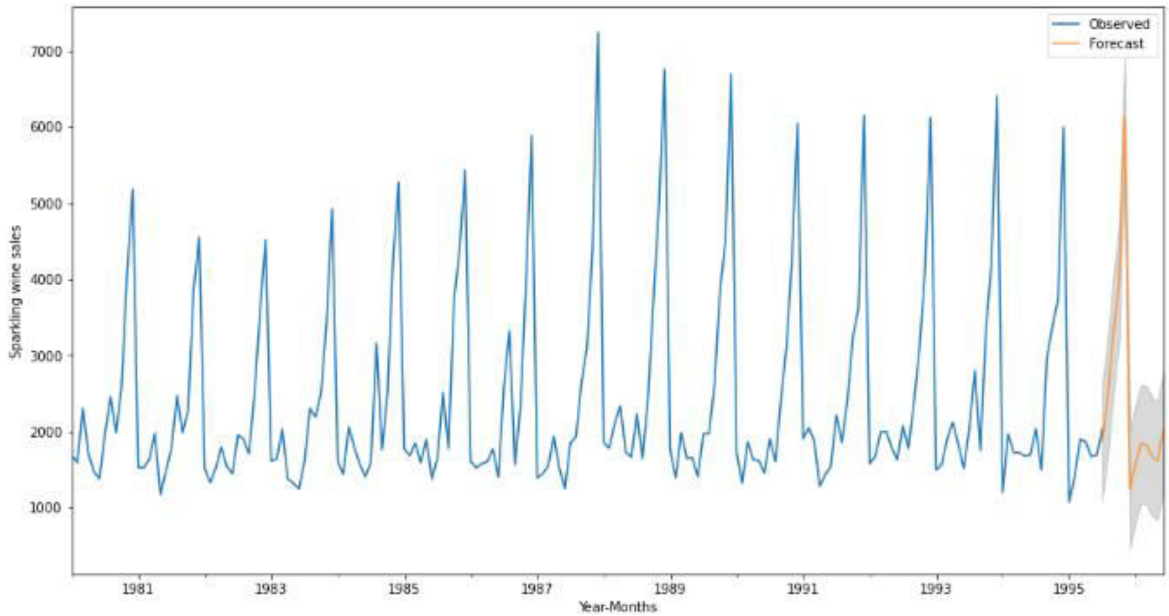


Fig92: Sparkling 12months forecast plot

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

- We see there is significant decrease in the sale of rose wine thru years from 1980 to 1995.
- There is spike in the rose wine sales in quarter4 due to holiday season.
- There seems to be no trend in the sales of Sparkling wine with high sales in the quarter 4.
- Sudden fall in the sales of Wines in the Jan month with slow sales improve from July for sparkling wine sales.
- After June Sales of Rose wine slowly increased

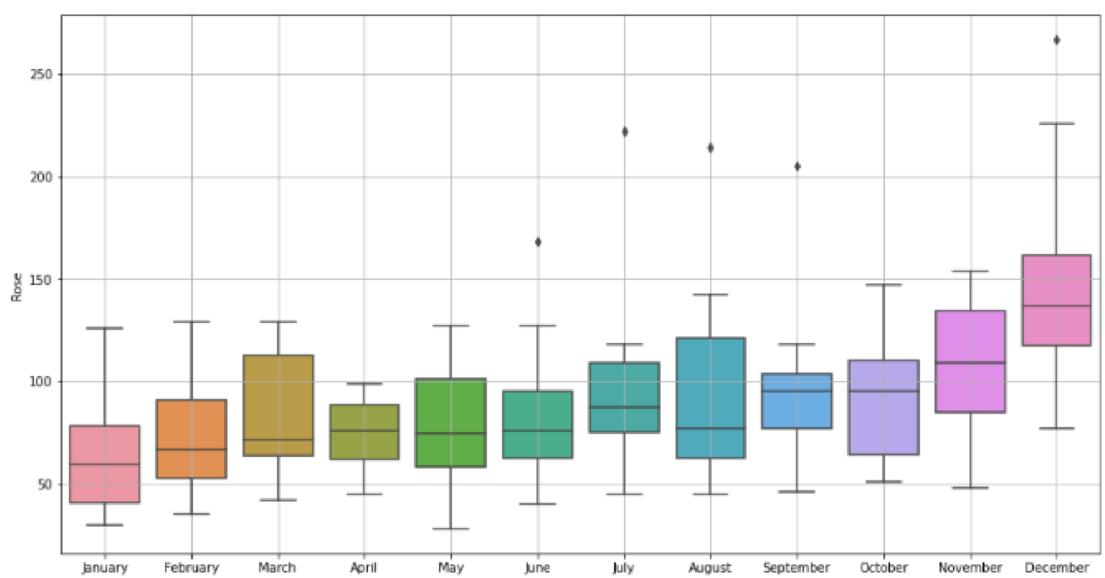


Fig93: Rose Box plot1

- After June, Sales of Sparkling also increased slowly

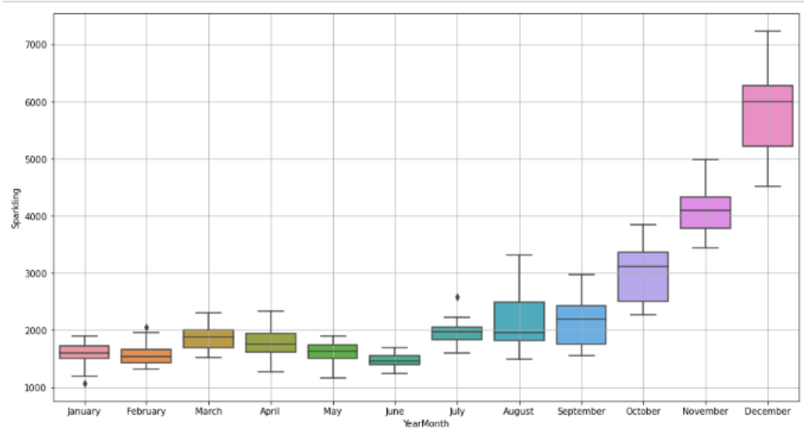


Fig94: Sparkling Box plot1

- December month sales of Sparkling wine are 3 times more than the June month Sales.
- TripleExponentialSmoothing gave the low RMSE values on the test data for Rose wine and so final model is built using TripleExponentialSmoothing for consistent forecast with respect to data.
- SARIMA model is choose for building final model of Sparkling wine as it resulted in low RMSE value.
- For Rose wine, year on year there is a decline in the sales where as for sales of Sparkling there is no significant increase or decrease.
- Special offers and ads to be introduced by the company to improve the sales and if sales didn't improve company has to investigate in depth about the cause or drop the variant and introduce new upgraded one.
- Holiday season is more important attract customers with different offers and company need to benefited from holiday season. Need to investigate in depth about the sharp sales fall in order to improve sales.
- Sparkling wine has more popularity.
- Offering the add-ons along with the Wine may lead to increase the sales.