TIME SERIES FORESCASTING PROJECT

Pooja Kabadi
PGP-DSBA Online
Batch- A4
20-02-2022

Table of Contents:

List of Figures:

List of Tables:

Problem Statement:

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

Objective of the report:
The Objective of this report is to check for sales of two distinct types of wines (Rose and Sparkling) for ABC Estate Wines throughout the twentieth century for 15 years. We have to analyse and to read the data from time series perspective and apply different exponential smoothing models, ARIMA and SARIMA model (auto/manual) to forecast the future wine sales. We'll also examine the outcomes of the various models and create a final model to forecast future sales based on the model that produces the greatest results.

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

The data consist of monthly wine sales for ABC Estate wines from year 1980 to 1995. It is recommended that we make our time series reference as the index while reading the data.

Reading the dataset from the .csv file and checking the head () of the dataset i.e., the first 5 rows of the dataset.

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

Figure 2. Sparkling Wine dataset

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

Figure 1. Rose Wine dataset

<u>Plotting the Time Series to understand the behaviour of the Sparkling data:</u>

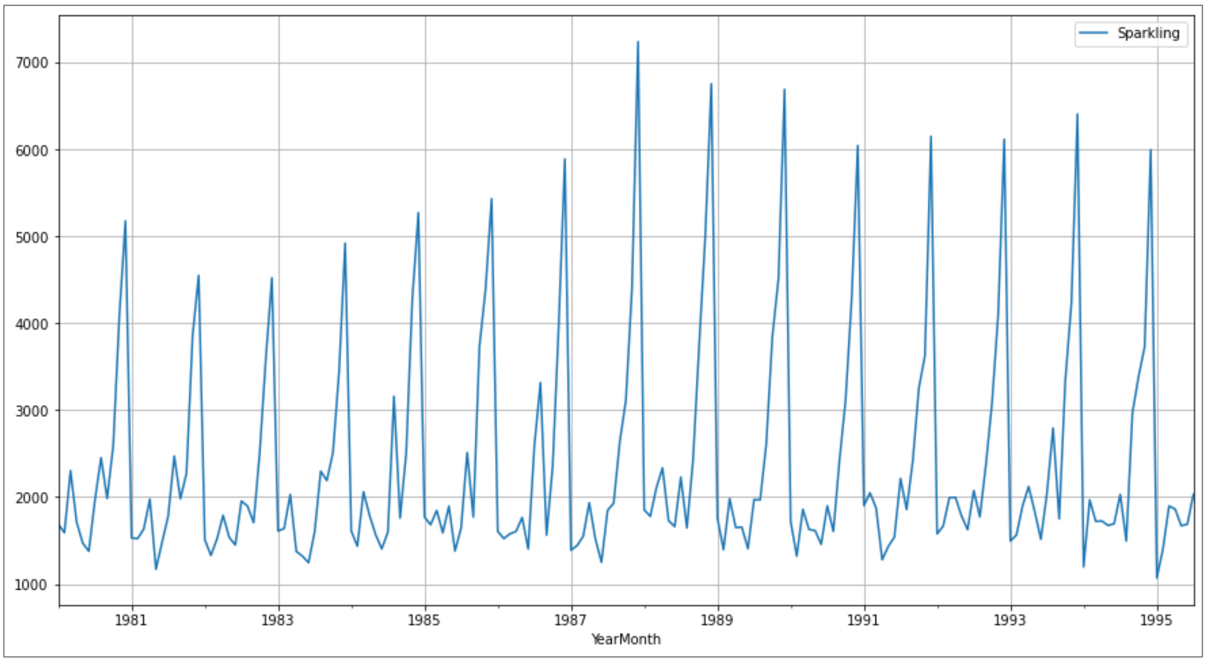

Figure 3. Plot of Sparkling dataset.

Observations:

- We notice that there is not much trend in the plot.
- we can clearly say that there is a seasonality factor throughout the period. The seasonality seems to have a pattern on yearly basis.
- No missing values are seen in the plot.

Plotting the Time Series to understand the behaviour of the Rose data:

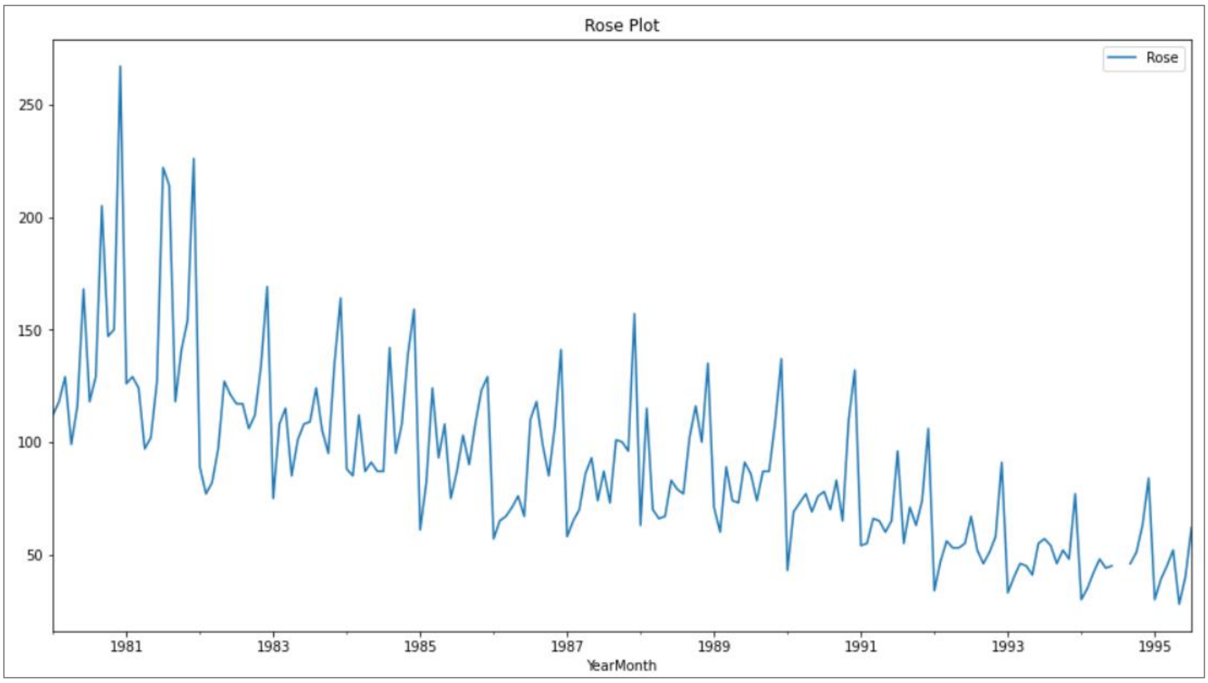

Figure 4. Plot of Rose dataset.

Observations:

- The above plot shows the sales of Rose wines from 1980 to 1995. We can see a downward trend in sales.
- We notice that there is a decreasing trend in the initial years which stabilizes after few years and again shows a decreasing trend.
- We also observe seasonality in the data trend and pattern seem to repeat on yearly basis.
- We can also see from this graph that there are missing values in data in 1994. The missing values should be imputed.

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

Checking the basic measures of descriptive statistics for Sparkling:

The basic measures of descriptive statistics tell us how the Sales have varied across years. But, for this measure of descriptive statistics we have averaged over the whole data without taking the time component into account.

	count	mean	std	min	25%	50%	75%	max
Sparkling	187.0	2402.417112	1295.11154	1070.0	1605.0	1874.0	2549.0	7242.0

Figure 5. Descriptive statistics for Sparkling

Observation:
The data consist of 187 monthly sales of Sparkling wine from year 1980 to 1995. The mean sale of wine across different months of year is ~ 2402 and the max sale is 7242 and min sale is 1070.

Checking the basic measures of descriptive statistics for Rose:

	count	mean	std	min	25%	50%	75%	max
Rose	185.0	90.394595	39.175344	28.0	63.0	86.0	112.0	267.0

Figure 6. Descriptive statistics for Rose.

Observation:
The data consist of 187 monthly sales of Rose wine from year 1980 to 1995. The mean sale of wine across different months of year is ~ 89.9 and the max sale is 267 and min sale is 28.

Checking data type of data features for Sparkling:

```
<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
```

Figure 7. Sparkling data information.

The Sparkling dataset consists of 187 observations and there are no missing values.

Checking data type of data features for Rose

```
<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
```

Figure 8. Rose data information.

The Rose dataset consists of 187 observations, we can see the missing values in dataset which should be imputed.

Imputing the missing values of Rose dataset:

Checking the data where null values are present. Below is the snippet showing the presence of null values in the year 1994, which was also seen in the graph *[figure 4]* above.

YearMonth	Rose
1994-01-01	30.0
1994-02-01	35.0
1994-03-01	42.0
1994-04-01	48.0
1994-05-01	44.0
1994-06-01	45.0
1994-07-01	NaN
1994-08-01	NaN
1994-09-01	46.0
1994-10-01	51.0
1994-11-01	63.0
1994-12-01	84.0

YearMonth	Rose
1994-01-01	30.000000
1994-02-01	35.000000
1994-03-01	42.000000
1994-04-01	48.000000
1994-05-01	44.000000
1994-06-01	45.000000
1994-07-01	45.333333
1994-08-01	45.666667
1994-09-01	46.000000
1994-10-01	51.000000
1994-11-01	63.000000
1994-12-01	84.000000

- We can observe from the adjacent table that there are missing values for the months of July and August in 1994.
- To fill in the missing data, we'll use pandas' interpolate() method.
- Interpolate() is a very powerful function and is basically used to fill NA values using different algorithms as required.
- We see that the values for July and August are now imputed with 45.33 and 45.67.

Figure 9. Snippet of checking NAN values.

<u>Plotting the Time Series again of the Rose data after imputation:</u>

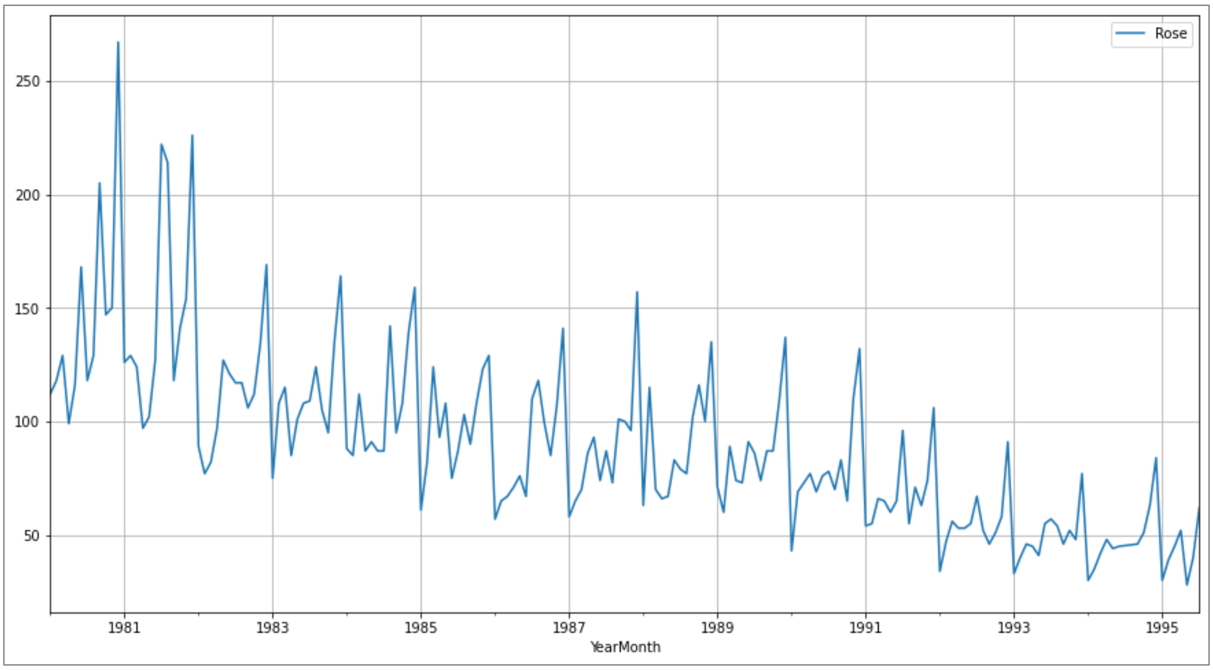

Figure 10. Plot of Rose data after imputation.

The Rose time series graph is now continuous, we can see that the missing values are interpolated in such a way the linear aggression is applied to the previous value to get the next value.

<u>Plotting a boxplot to understand the spread of sales across different years and within different months across years.</u>

<u>Yearly Sale Boxplot for Sparkling wine.</u>

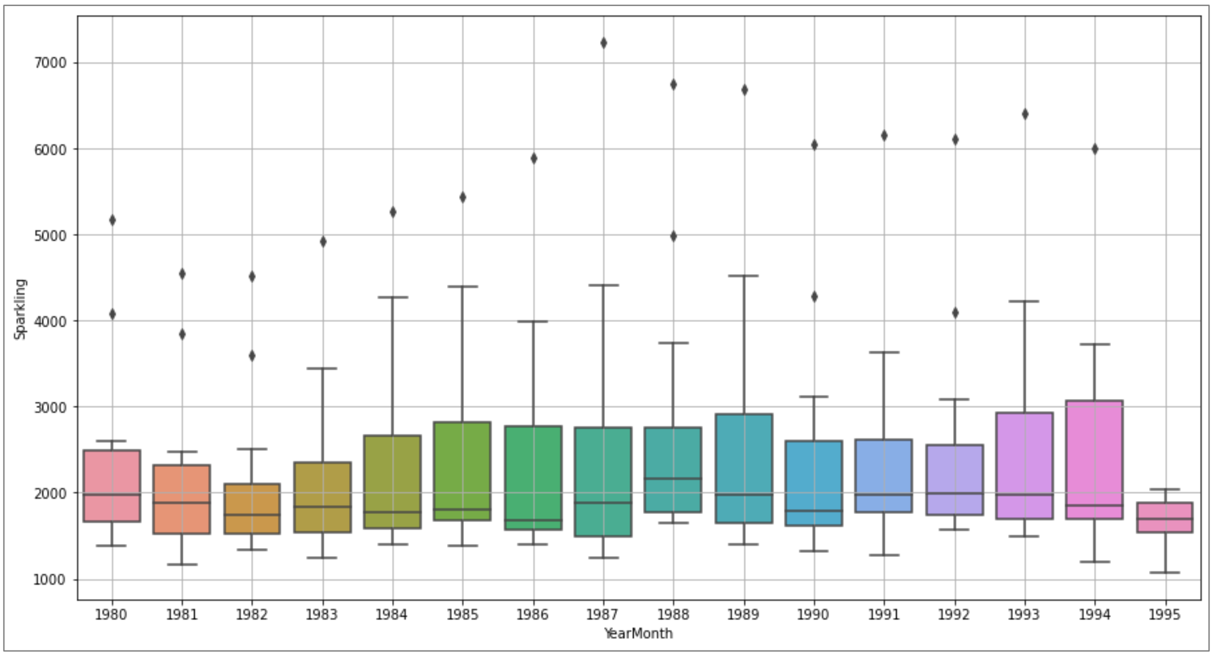

Figure 11. Yearly Sale Boxplot for Sparkling wine.

Observations:

- From the above yearly plot, we see that the box plots do not indicate any trend.
- We also observe that the sale of Sparkling wine has outliers for almost all the years except 1955.
- The highest mean sale for sprinkling is shown in year 1988 and the lowest sales are in the year of 1995.
- The sales remain stagnant between years 1991-1993 and then starts decreasing slightly.
- There is no increasing or decreasing trend in sale throughout the given time series.
- The sale of Sparkling is quite stagnant and management should definitely work on why the sales are not showing any trend.

<u>Yearly Sale Boxplot for Rose wine</u>

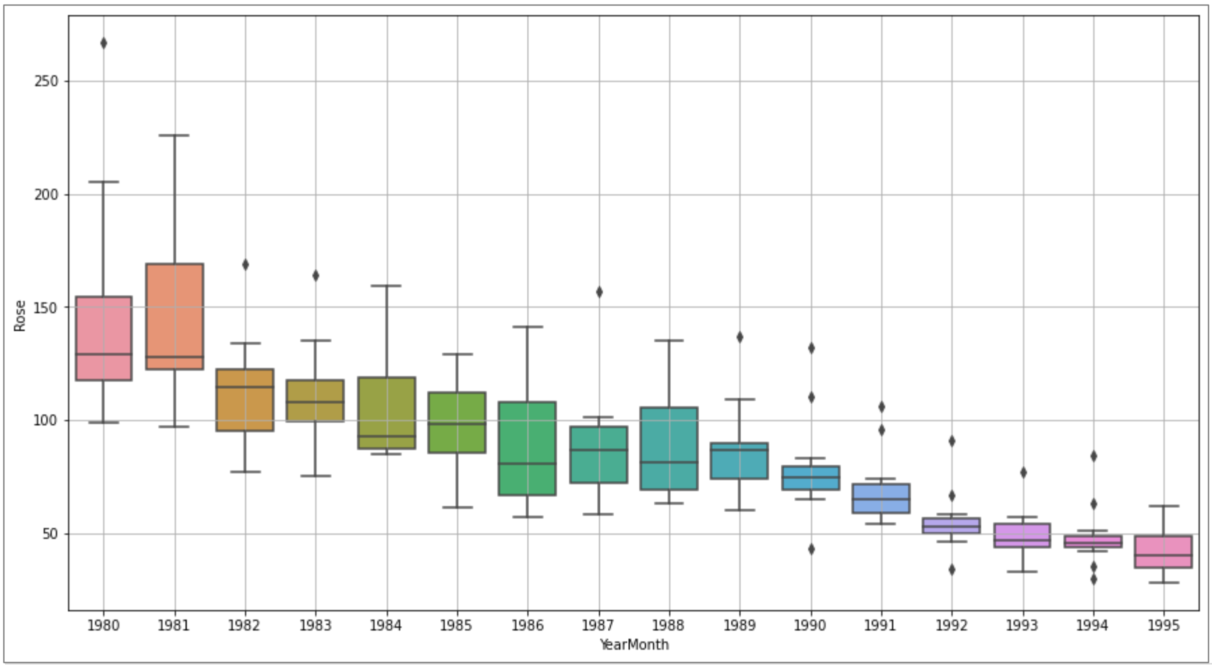

Figure 12. Yearly Sale Boxplot for Rose wine

Observations:

- From the yearly above plot, we see that the box plots indicate a downward trend.
- We also observe that there are few outliers present in the Rose wine sales plot.

<u>Monthly Sale Boxplot for Sparkling wine:</u>

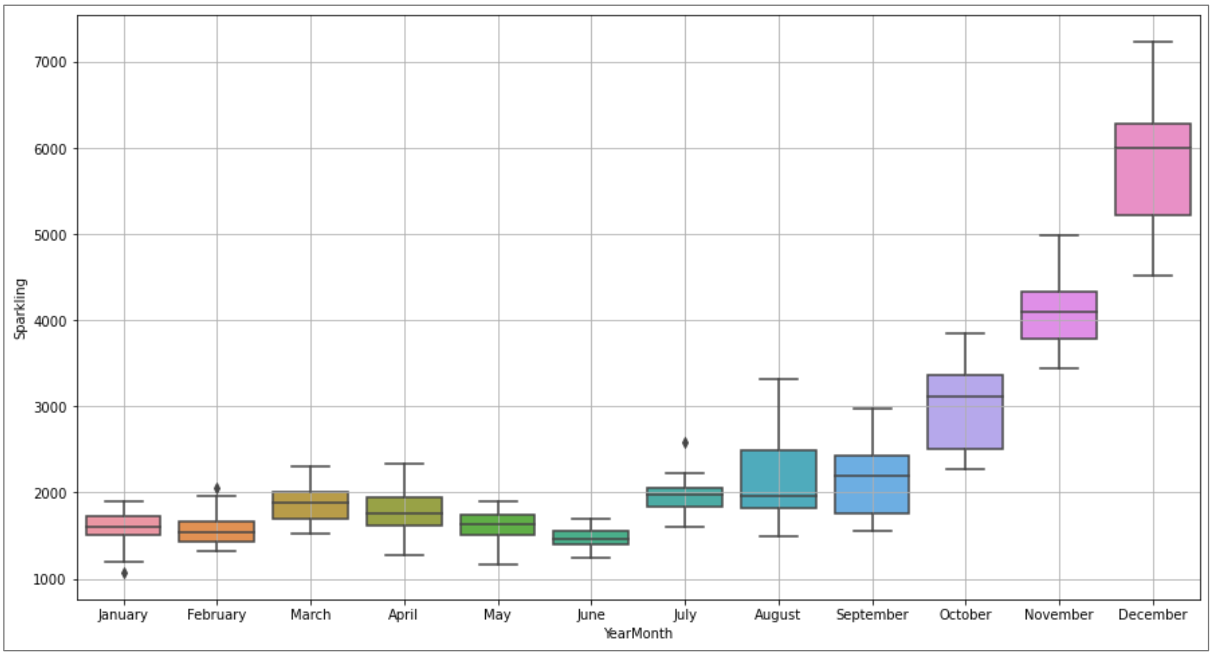

Figure 13. Monthly Sale Boxplot for Sparkling wine.

Observations:

- The highest monthly sale of wine is in month of Nov-Dec, this may be due to winter and festival season.
- The sale is quite stagnant in starting months till May and then there is increasing trend from July.
- The above plot concludes that people love drinking wine in winter.

<u>Monthly Sale Boxplot for Rose wine:</u>

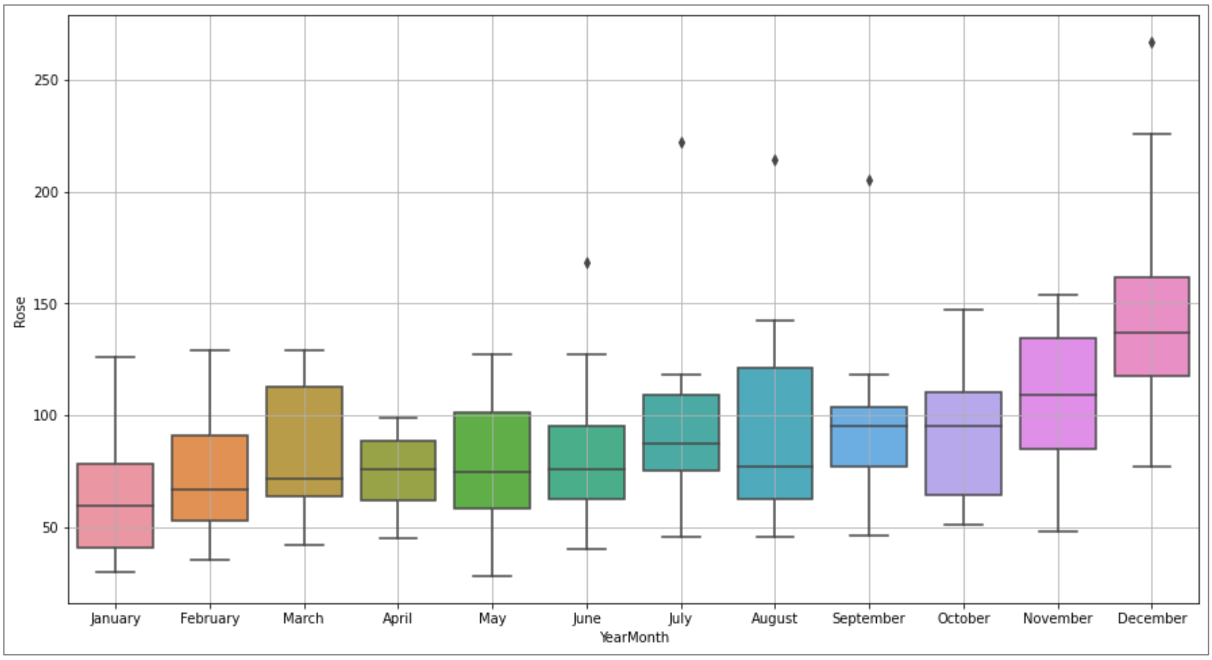

Figure 14. Monthly Sale Boxplot for Rose wine.

Observations:

- The highest monthly sale of wine is in month of Nov-Dec, this may be due to winter and festival season.
- From the above plot, we see that December month has the highest sales of wine.
- There are also outliers present in June, July, August and September months.

<u>Plotting a graph of monthly Sales of Sparkling wine across years:</u>

YearMonth	1	2	3	4	5	6	7	8	9	10	11	12
YearMonth												
1980	1686.0	1591.0	2304.0	1712.0	1471.0	1377.0	1966.0	2453.0	1984.0	2596.0	4087.0	5179.0
1981	1530.0	1523.0	1633.0	1976.0	1170.0	1480.0	1781.0	2472.0	1981.0	2273.0	3857.0	4551.0
1982	1510.0	1329.0	1518.0	1790.0	1537.0	1449.0	1954.0	1897.0	1706.0	2514.0	3593.0	4524.0
1983	1609.0	1638.0	2030.0	1375.0	1320.0	1245.0	1600.0	2298.0	2191.0	2511.0	3440.0	4923.0
1984	1609.0	1435.0	2061.0	1789.0	1567.0	1404.0	1597.0	3159.0	1759.0	2504.0	4273.0	5274.0
1985	1771.0	1682.0	1846.0	1589.0	1896.0	1379.0	1645.0	2512.0	1771.0	3727.0	4388.0	5434.0
1986	1606.0	1523.0	1577.0	1605.0	1765.0	1403.0	2584.0	3318.0	1562.0	2349.0	3987.0	5891.0
1987	1389.0	1442.0	1548.0	1935.0	1518.0	1250.0	1847.0	1930.0	2638.0	3114.0	4405.0	7242.0
1988	1853.0	1779.0	2108.0	2336.0	1728.0	1661.0	2230.0	1645.0	2421.0	3740.0	4988.0	6757.0
1989	1757.0	1394.0	1982.0	1650.0	1654.0	1406.0	1971.0	1968.0	2608.0	3845.0	4514.0	6694.0
1990	1720.0	1321.0	1859.0	1628.0	1615.0	1457.0	1899.0	1605.0	2424.0	3116.0	4286.0	6047.0
1991	1902.0	2049.0	1874.0	1279.0	1432.0	1540.0	2214.0	1857.0	2408.0	3252.0	3627.0	6153.0
1992	1577.0	1667.0	1993.0	1997.0	1783.0	1625.0	2076.0	1773.0	2377.0	3088.0	4096.0	6119.0
1993	1494.0	1564.0	1898.0	2121.0	1831.0	1515.0	2048.0	2795.0	1749.0	3339.0	4227.0	6410.0
1994	1197.0	1968.0	1720.0	1725.0	1674.0	1693.0	2031.0	1495.0	2968.0	3385.0	3729.0	5999.0
1995	1070.0	1402.0	1897.0	1862.0	1670.0	1688.0	2031.0	NaN	NaN	NaN	NaN	NaN

Table 1. Sparkling Monthly sales trend across years

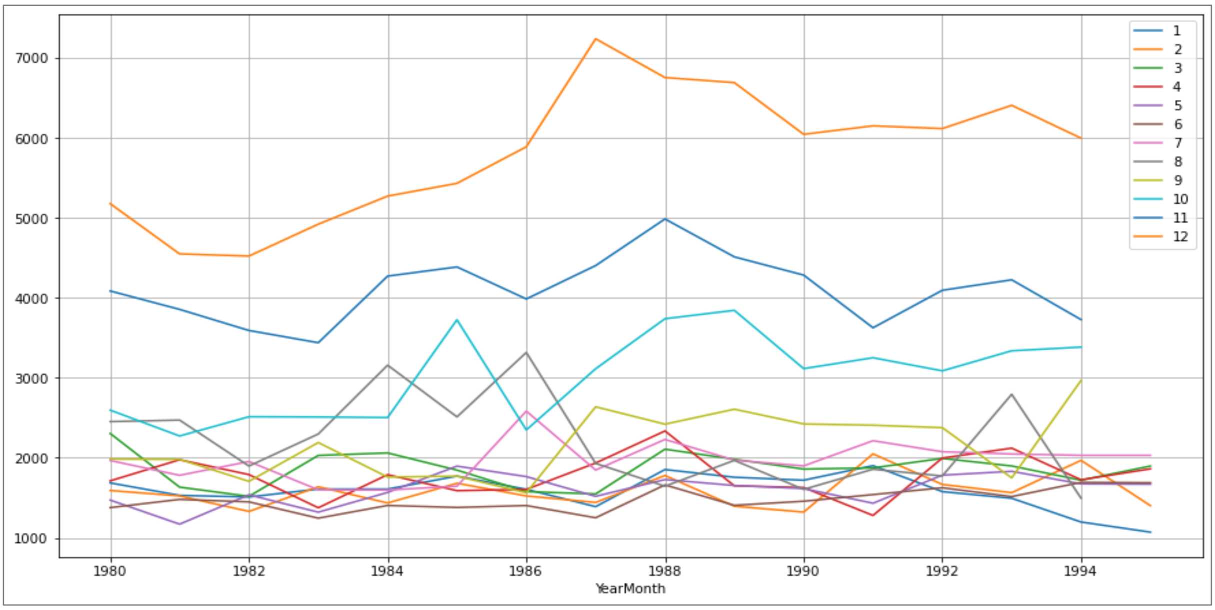

Figure 15. Plot of monthly Sales of Sparkling wine across years.

Observations:

- The number 1-12 indicates the month of years from January to December. The highest sales of sparkling wine is in the month of December and the lowest with very stagnant sales is in the month of June for all the years.
- The sale for wine in month of August is good till 1986 and then there is steep decrease in later years.

<u>Potting a graph of monthly Sales of Rose wine across years:</u>

YearMonth	1	2	3	4	5	6	7	8	9	10	11	12
YearMonth												
1980	112.0	118.0	129.0	99.0	116.0	168.0	118.000000	129.000000	205.0	147.0	150.0	267.0
1981	126.0	129.0	124.0	97.0	102.0	127.0	222.000000	214.000000	118.0	141.0	154.0	226.0
1982	89.0	77.0	82.0	97.0	127.0	121.0	117.000000	117.000000	106.0	112.0	134.0	169.0
1983	75.0	108.0	115.0	85.0	101.0	108.0	109.000000	124.000000	105.0	95.0	135.0	164.0
1984	88.0	85.0	112.0	87.0	91.0	87.0	87.000000	142.000000	95.0	108.0	139.0	159.0
1985	61.0	82.0	124.0	93.0	108.0	75.0	87.000000	103.000000	90.0	108.0	123.0	129.0
1986	57.0	65.0	67.0	71.0	76.0	67.0	110.000000	118.000000	99.0	85.0	107.0	141.0
1987	58.0	65.0	70.0	86.0	93.0	74.0	87.000000	73.000000	101.0	100.0	96.0	157.0
1988	63.0	115.0	70.0	66.0	67.0	83.0	79.000000	77.000000	102.0	116.0	100.0	135.0
1989	71.0	60.0	89.0	74.0	73.0	91.0	86.000000	74.000000	87.0	87.0	109.0	137.0
1990	43.0	69.0	73.0	77.0	69.0	76.0	78.000000	70.000000	83.0	65.0	110.0	132.0
1991	54.0	55.0	66.0	65.0	60.0	65.0	96.000000	55.000000	71.0	63.0	74.0	106.0
1992	34.0	47.0	56.0	53.0	53.0	55.0	67.000000	52.000000	46.0	51.0	58.0	91.0
1993	33.0	40.0	46.0	45.0	41.0	55.0	57.000000	54.000000	46.0	52.0	48.0	77.0
1994	30.0	35.0	42.0	48.0	44.0	45.0	45.333333	45.666667	46.0	51.0	63.0	84.0
1995	30.0	39.0	45.0	52.0	28.0	40.0	62.000000	NaN	NaN	NaN	NaN	NaN

Table 2. Rose Monthly sales trend across years

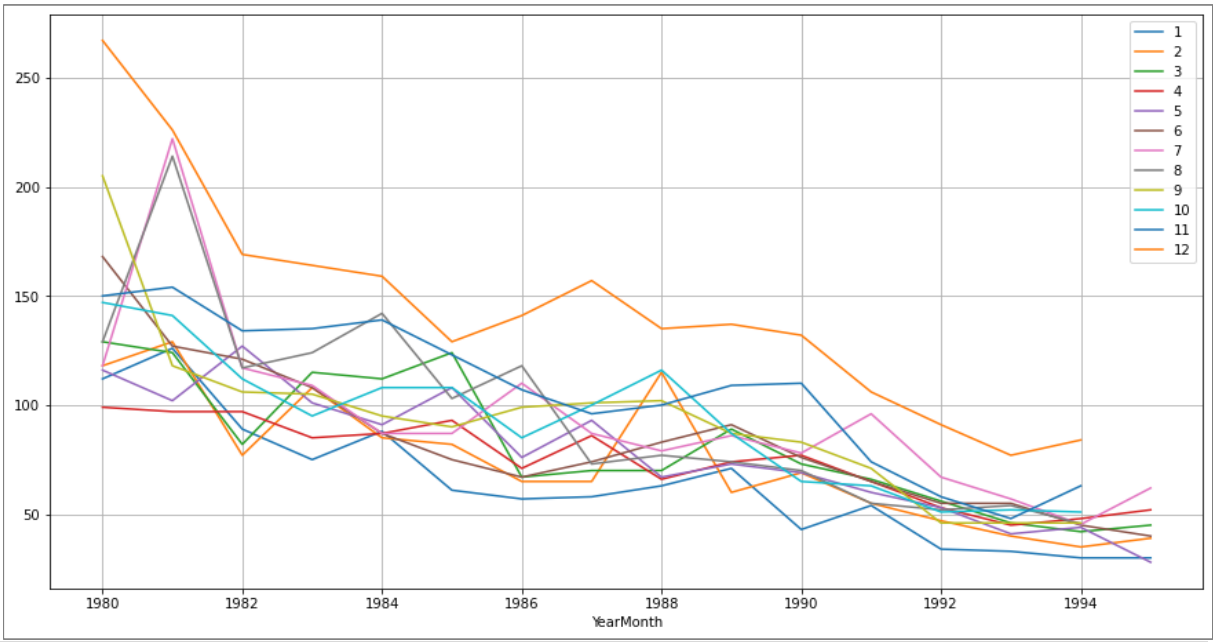

Figure 16. Plot of monthly Sales of Rose wine across years.

Observations:

- The number 1-12 indicates the month of years from January to December. The highest sales of Rose wine is in the month of December and January, February and May show relatively lower sales.
- Company can start promoting and advertising during winter to increase its sales.

Plotting the Empirical Cumulative Distribution for Sparkling and Rose wine:

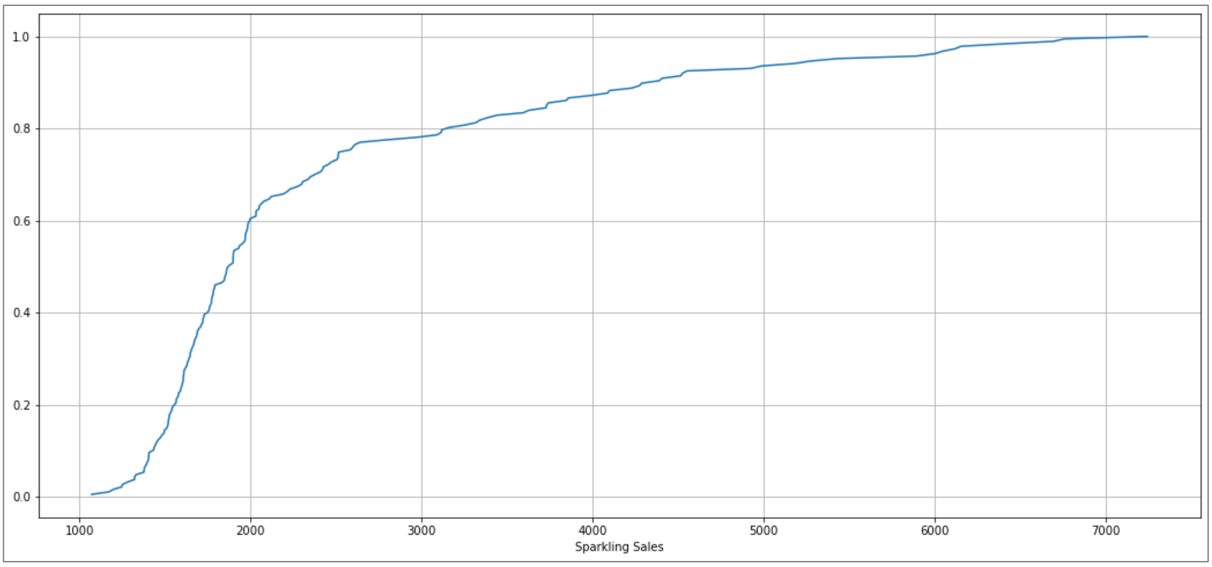

Figure 17. Plot of Empirical Cumulative Distribution for Sparkling

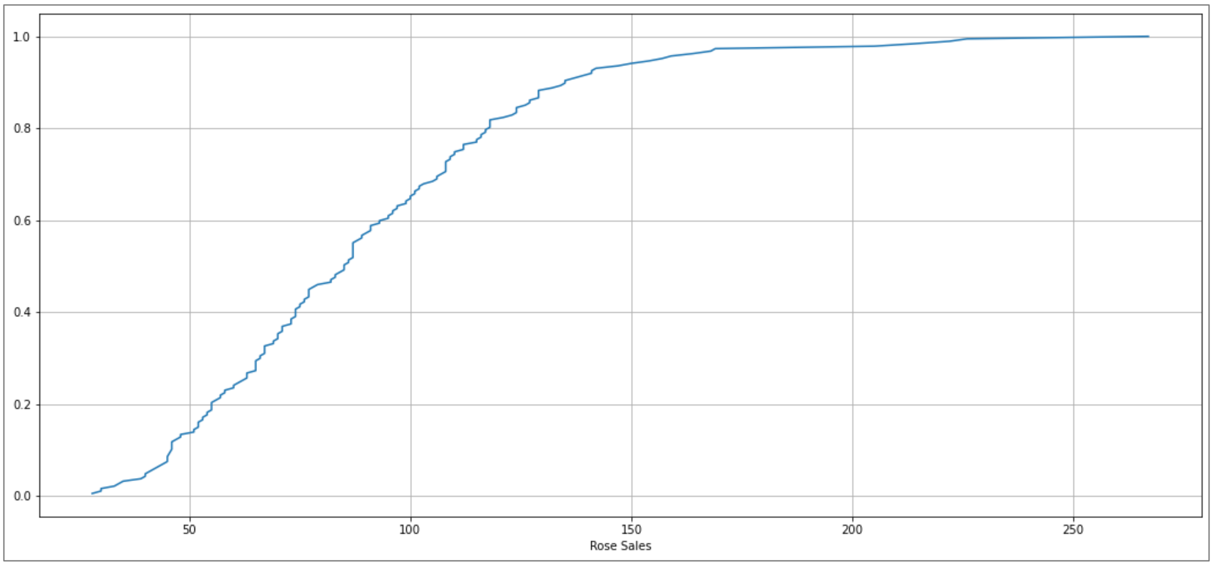

Figure 18. Plot of Empirical Cumulative Distribution for Rose.

- This particular graph tells us what percentage of data points refer to what number of Sales.

Sum of sales of Sparkling across different year:

Let us try to resample or aggregate the Time Series from an annual perspective and sum up the observations of each month.

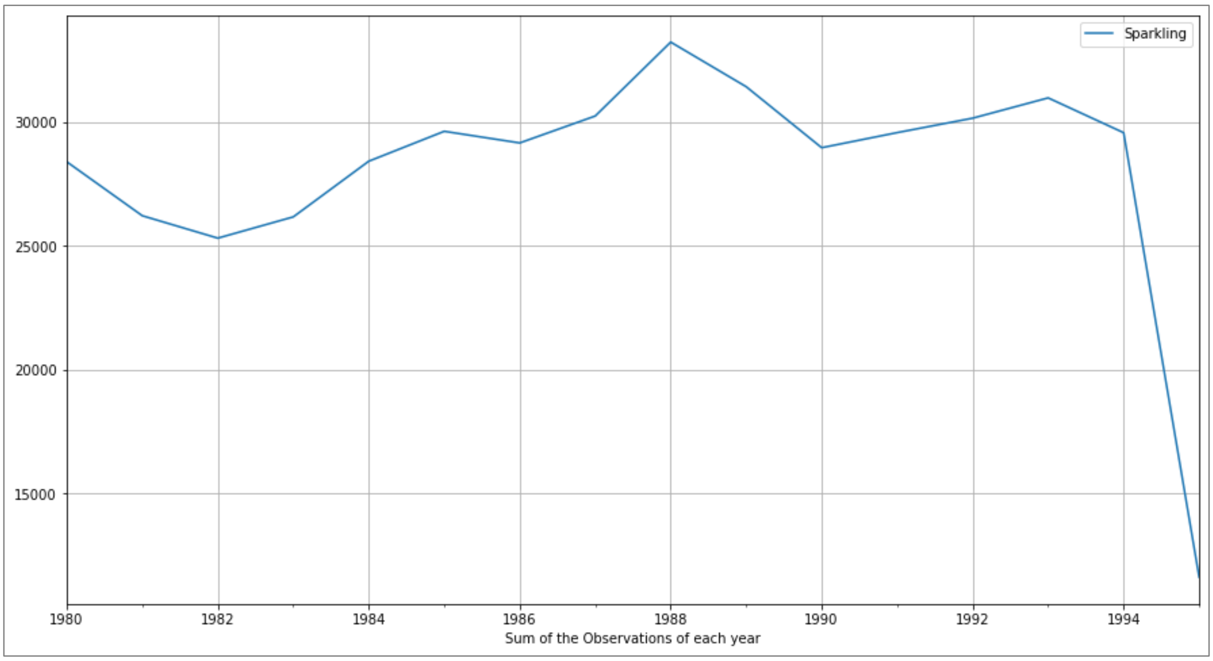

Figure 19. Plot of Sum of sales of Sparkling across different year

Observations:

- Resample or aggregate the Time Series from an annual perspective and sum up the observations.
- We can see that, the sale of Sparkling wine decreases from 1980-1982 and then increases till 1988. There is steep decrease in sale from 1994

Sum of sales of Rose across different year:

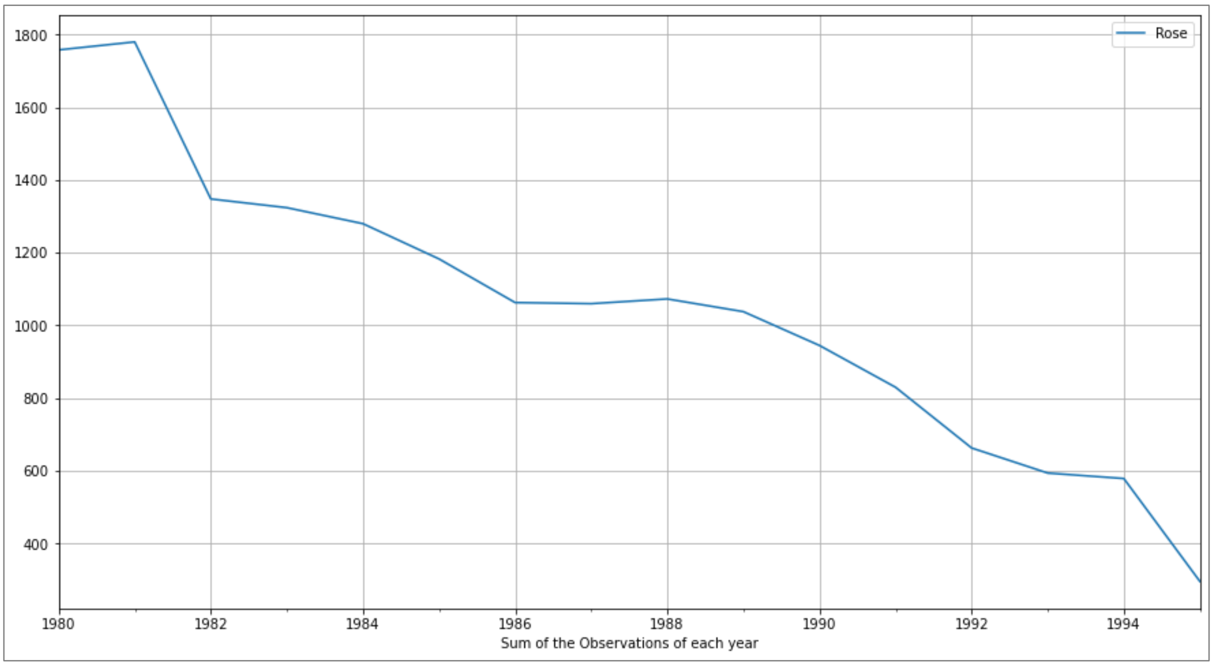

Figure 20, Plot of Sum of sales of Rose across different year

- We can see that; the sale of Rose wine has steep decreases from 1980-1982 and then there is gradual decreasing trend.

Plot the average Sales per month and the month-on-month percentage change of Sales of Sparkling wine:

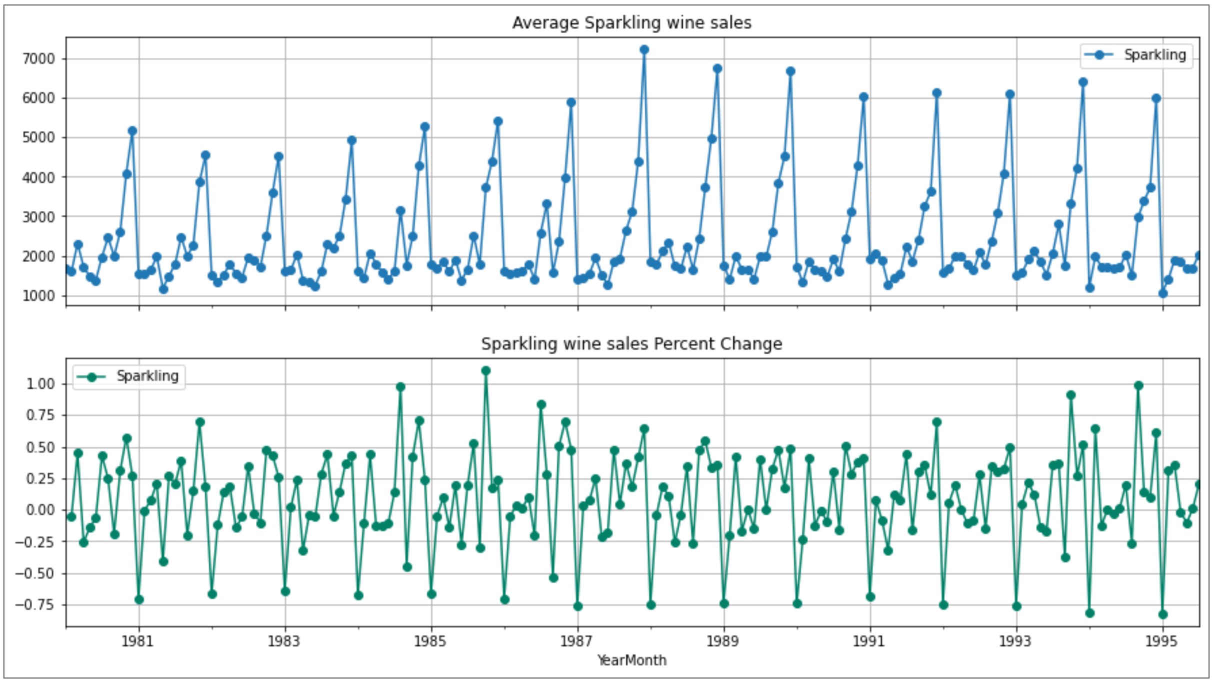

Figure 21. Sparkling Average sales & percentage change sales.

Observations:

- The above two graphs tell us the Average Sales and the Percentage change of Sales with respect to the time for Sparkling wine.
- The Average Sales value does not show any trend.

Plot the average Sales per month and the month-on-month percentage change of Sales of Rose wine:

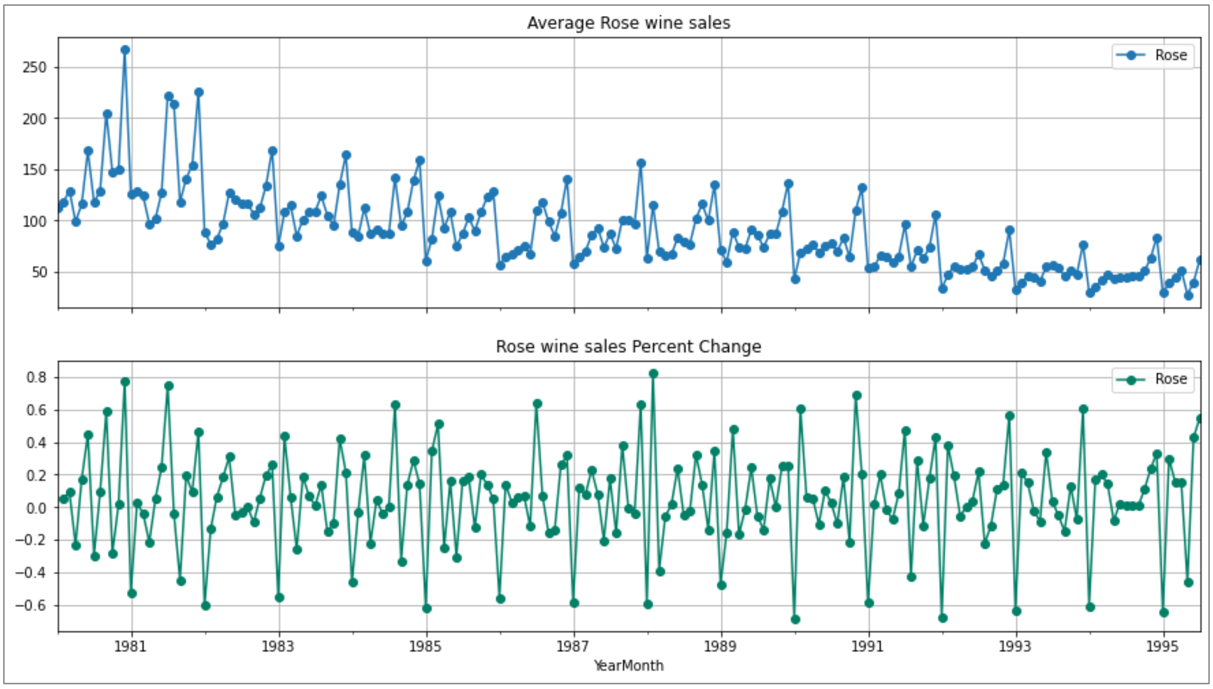

Figure 22. Rose Average sales & percentage change sales.

Observations:

- The above two graphs tell us the Average Sales and the Percentage change of Sales with respect to the time for Rose wine.
- The Average Sales value shows decreasing trend.

<u>Decompose the Time Series and plot the different components for Sparkling.</u>

<u>Additive Decomposition for Sparkling:</u>

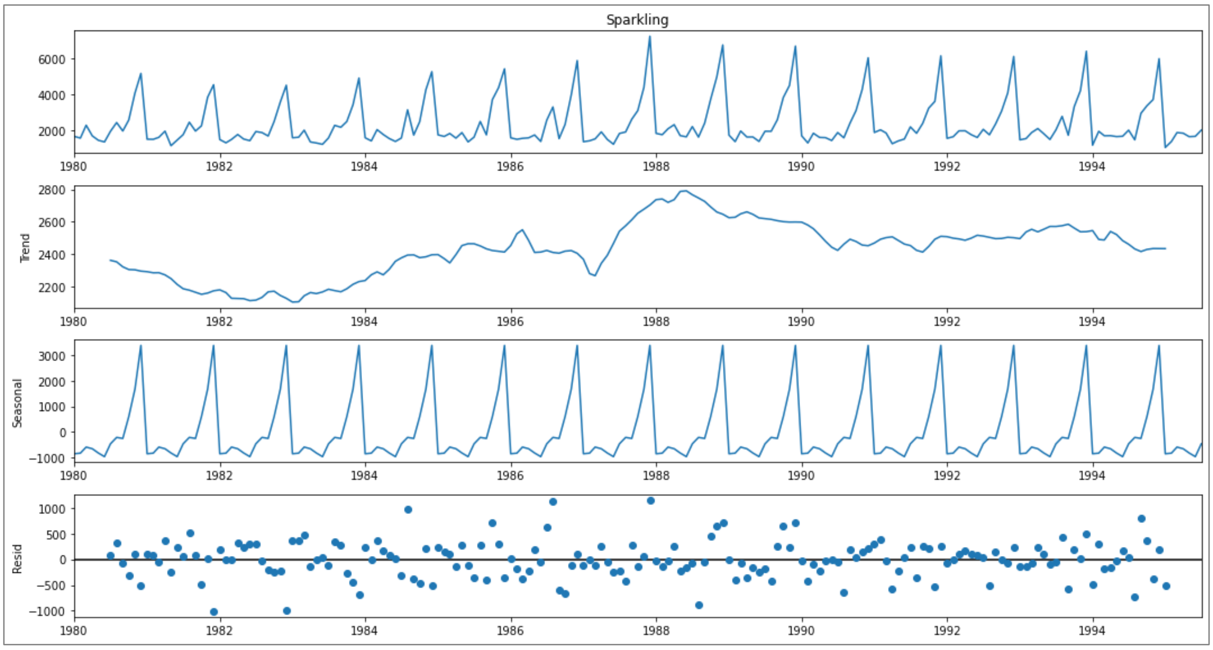

Figure 23. Plot of Additive Decomposition for Sparkling.

Observations:

- We see that the residuals are located around 0 from the plot of the residuals in the decomposition.
- The residuals are ranging from -1000 to +1000. we can see some pattern in the residual so further decomposing to multiplicative model to minimize the residuals.
- There is seasonality and we don't observe pronounced trend.

The first 12 months trend, seasonality and residual values of Sparkling wine sales dataset:

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

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

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

Figure 24. Sparkling Additive decomposition description

Multiplicative Decomposition for Sparkling:

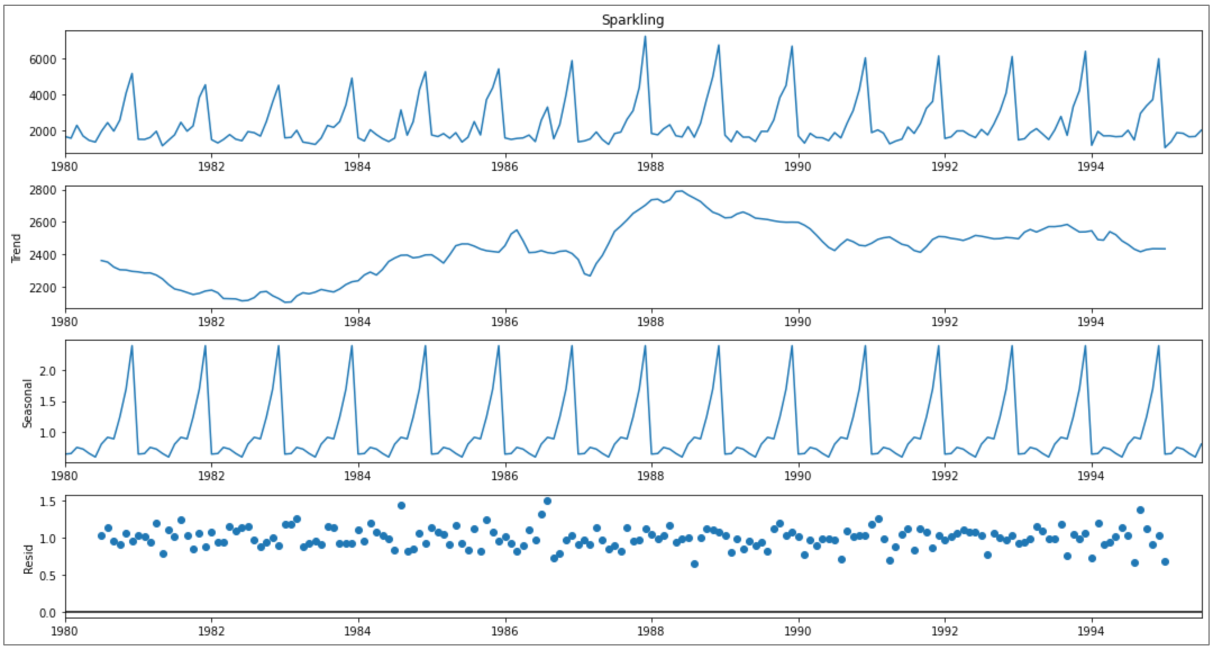

Figure 25. Plot of Multiplicative Decomposition for Sparkling.

Observations:

- For the multiplicative series, we see that a lot of residuals are located around 1. The residuals are minimized and multiplicative model is best fit for decomposition.
- There is seasonality and we don't observe pronounced trend.

The first 12 months multiplicative trend, seasonality and residual values of Sparkling wine sales dataset:

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

Seasonality YearMonth	
1980-01-01	0.649843
1980-02-01	0.659214
1980-03-01	0.757440
1980-04-01	0.730351
1980-05-01	0.660609
1980-06-01	0.603468
1980-07-01	0.809164
1980-08-01	0.918822
1980-09-01	0.894367
1980-10-01	1.241789
1980-11-01	1.690158
1980-12-01	2.384776

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	1.029230
1980-08-01	1.135407
1980-09-01	0.955954
1980-10-01	0.907513
1980-11-01	1.050423
1980-12-01	0.946770

Figure 26. Sparkling Multiplicative decomposition description

Decompose the Time Series and plot the different components for Rose.

Additive Decomposition for Rose:

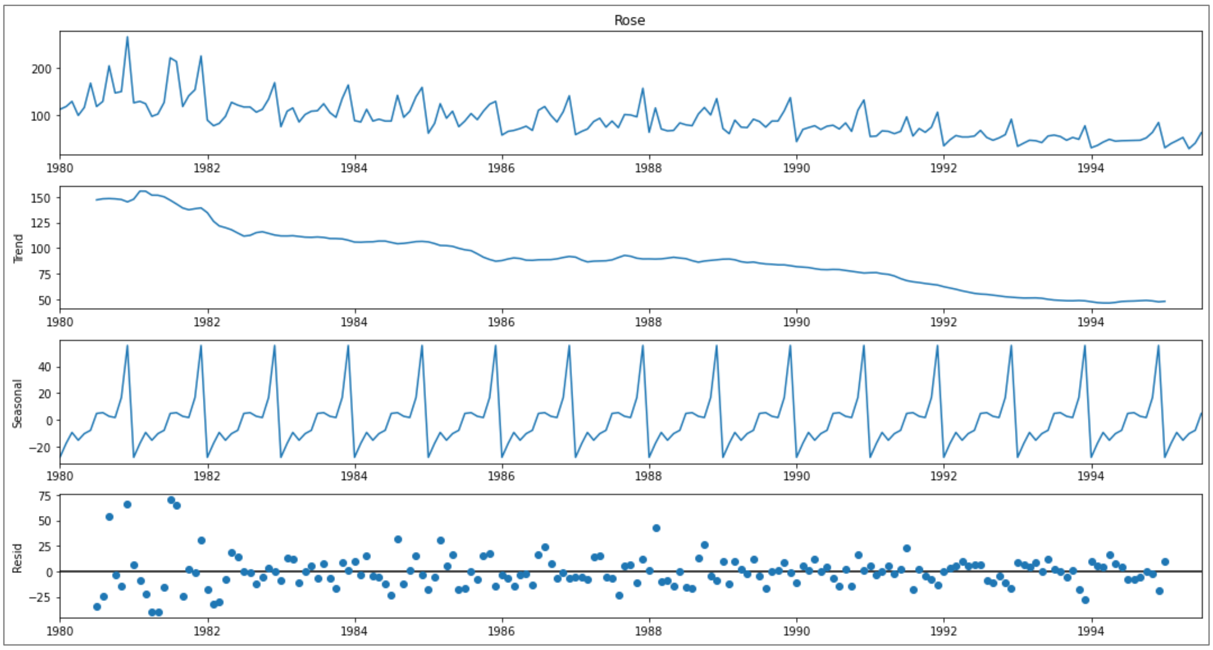

Figure 27. Plot of Additive Decomposition for Rose.

Observations:

- We see that the residuals are located around 0 from the plot of the residuals in the decomposition.
- The residuals are ranging from -25 to +75. we can see some pattern in the residual so further decomposing to multiplicative model to minimize the residuals.
- There is seasonality and we observe decreasing trend.

The first 12 months Additive trend, seasonality and residual values of Rose wine sales dataset:

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

Seasonality YearMonth	
1980-01-01	-27.908647
1980-02-01	-17.435632
1980-03-01	-9.285830
1980-04-01	-15.098330
1980-05-01	-10.196544
1980-06-01	-7.678687
1980-07-01	4.896908
1980-08-01	5.499686
1980-09-01	2.774686
1980-10-01	1.871908
1980-11-01	16.846908
1980-12-01	55.713575

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	-33.980241
1980-08-01	-24.624686
1980-09-01	53.850314
1980-10-01	-2.955241
1980-11-01	-14.263575
1980-12-01	66.161425

Figure 28. Rose Additive decomposition description

Multiplicative Decomposition for Rose:

Observations:

- For the multiplicative series, we see that a lot of residuals are located around 1. The residuals are minimized and multiplicative model is best fit for decomposition.
- There is seasonality and we observe decreasing trend.

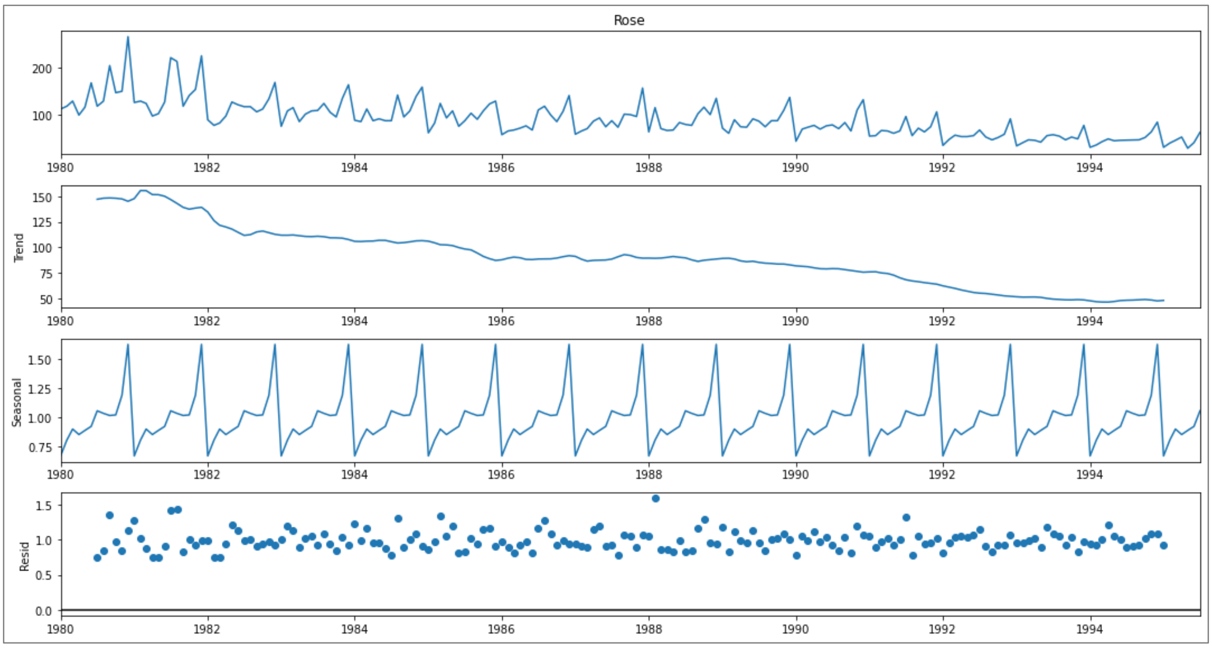

Figure 29. Plot of Multiplicative Decomposition for Rose.

The first 12 months multiplicative trend, seasonality and residual values of Rose wine sales dataset:

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

Seasonality YearMonth	
1980-01-01	0.670111
1980-02-01	0.806163
1980-03-01	0.901164
1980-04-01	0.854024
1980-05-01	0.889415
1980-06-01	0.923985
1980-07-01	1.058038
1980-08-01	1.035881
1980-09-01	1.017648
1980-10-01	1.022573
1980-11-01	1.192349
1980-12-01	1.628646

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	0.758258
1980-08-01	0.840720
1980-09-01	1.357674
1980-10-01	0.970771
1980-11-01	0.853378
1980-12-01	1.129646

Figure 30. Rose Multiplicative decomposition description

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

Splitting the time series data into train and test. Training Data is till the end of 1990. Test Data is from the beginning of 1991 to the last time stamp provided.

<u>Splitting Sparkling dataset:</u>

Train data: Data < 1991, which contains (132, 1) observations.

Test data: Data > = 1991, which contains (55,1) observations.

<u>Plotting the Training data, Testing data for Sparkling data:</u>

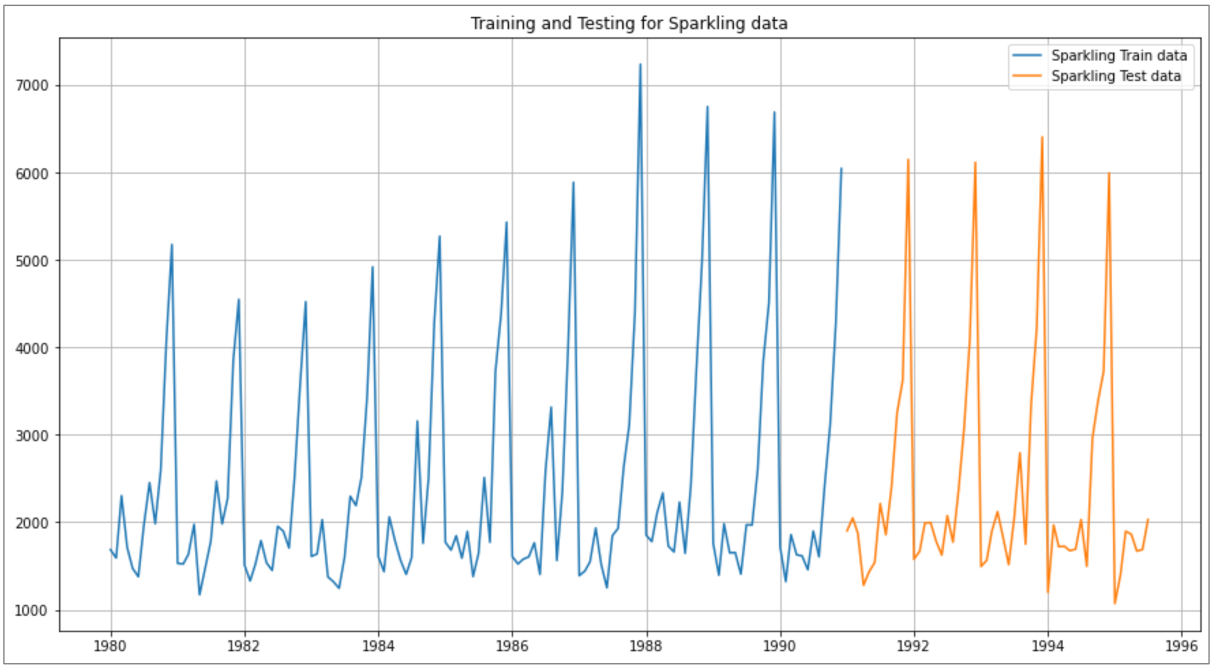

Figure 31. Plot of train and test data for Sparkling wine.

Splitting Rose dataset:

Train data: Data < 1991, which contains (132, 1) observations.

Test data: Data > = 1991, which contains (55,1) observations.

Plotting the Training data, Testing data for Rose data:

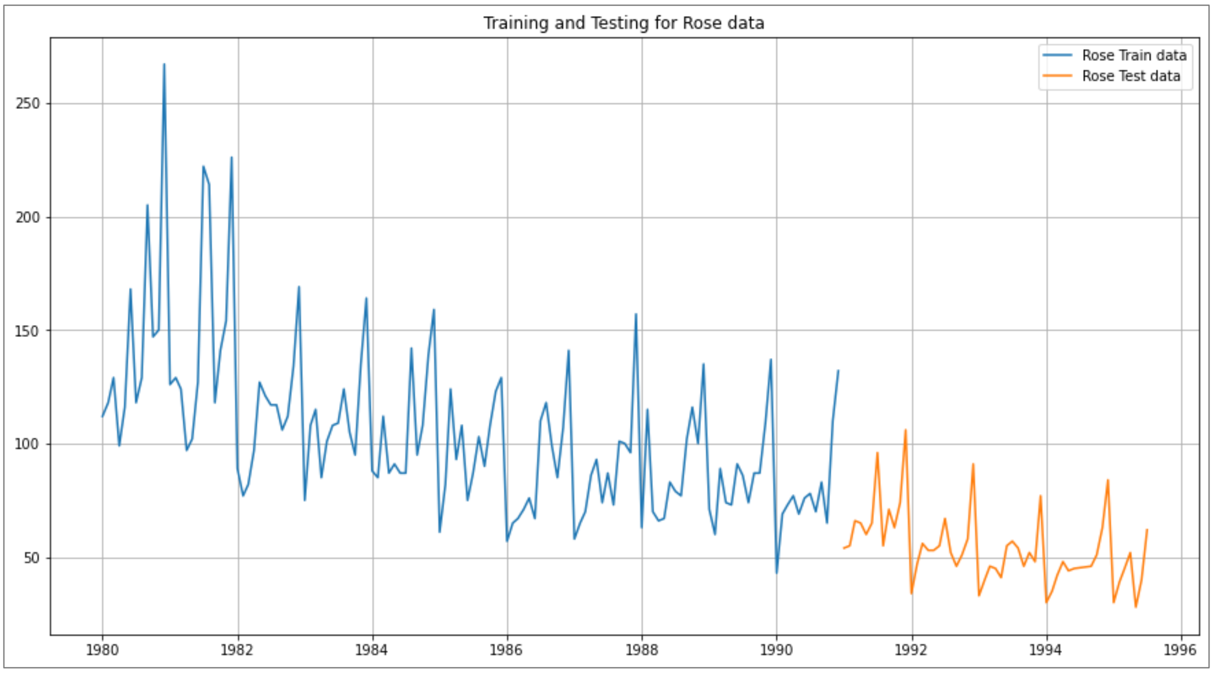

Figure 32. Plot of train and test data for Rose wine

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

Building different models on the training data and forecast the values on test data and check performance of each model using root mean squared error (RMSE) values. The model with the minimum RMSE value would be our best fit model.

The different models that we are going to build are as follows:

1. Linear Regression model
2. Naïve Approach
3. Simple Average
4. Moving Average
5. Simple Exponential Smoothing
6. Double Exponential Smoothing (Holt's model)
7. Triple Exponential Smoothing (Holt Winter's model)

Let's take a look at each model for both Rose and Sparkling wine sales and check for their performance.

Model 1 – Linear Regression model.

Regression algorithms try to find the line of best fit for a given dataset. The linear regression algorithm tries to minimize the value of the sum of the squares of the differences between the observed value and predicted value.

In the simplest case, the regression model allows for a linear relationship between the forecast variable y and a single predictor variable x.

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$$

For this particular linear regression, we are going to regress the 'Sales' variable against the order of the occurrence. For this we need to modify our training data before fitting it into a linear regression.

Target = weight 1 * feature 1 + weight 2 * feature 2 + bias

The regression algorithm discovers the best values for the parameters weight 1, weight 2, and bias during training. Then use this model to forecast on future values.

Linear Regression for Sparkling wine:

Firstly, creating training and testing time instances which is appended to the time series dataset.

```
Training Time instance for Sparkling
 [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 3
4, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65,
66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97,
98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123,
124, 125, 126, 127, 128, 129, 130, 131, 132]
Test Time instance for Sparkling
 [133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157,
158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 18
3, 184, 185, 186, 187]
```

Figure 33. Time Instances for Sparkling data.

Looking at the first and last few data for training and testing for Sparkling after appending the time instances.

First few rows of Training Data for Sparkling

YearMonth	Sparkling	time
1980-01-01	1686	1
1980-02-01	1591	2
1980-03-01	2304	3
1980-04-01	1712	4
1980-05-01	1471	5

Last few rows of Training Data for Sparkling

YearMonth	Sparkling	time
1990-08-01	1605	128
1990-09-01	2424	129
1990-10-01	3116	130
1990-11-01	4286	131
1990-12-01	6047	132

Figure 35. Sparkling train data sample for regression.

First few rows of Test Data for Sparkling

YearMonth	Sparkling	time
1991-01-01	1902	133
1991-02-01	2049	134
1991-03-01	1874	135
1991-04-01	1279	136
1991-05-01	1432	137

Last few rows of Test Data for Sparkling

YearMonth	Sparkling	time
1995-03-01	1897	183
1995-04-01	1862	184
1995-05-01	1670	185
1995-06-01	1688	186
1995-07-01	2031	187

Figure 34. Sparkling test data sample for regression.

Now that our training and test data has been modified, let us go ahead use *LinearRegression* to build the model on the training data and test the model on the test data.

To build a Linear Regression time series model for Sparkling:

- Fitting the Linear Regression model on train data which is imported from Sklearn linear model
- Linear Regression uses the Gradient descent method for finding the best fit model, which uses partial derivatives on the parameters to minimize the sum of squared errors.
- Predict the values for test data and plot the graph.
- Checking the performance of the model by calculating the RMSE value.

<u>Plotting Linear Regression for Sparkling data:</u>

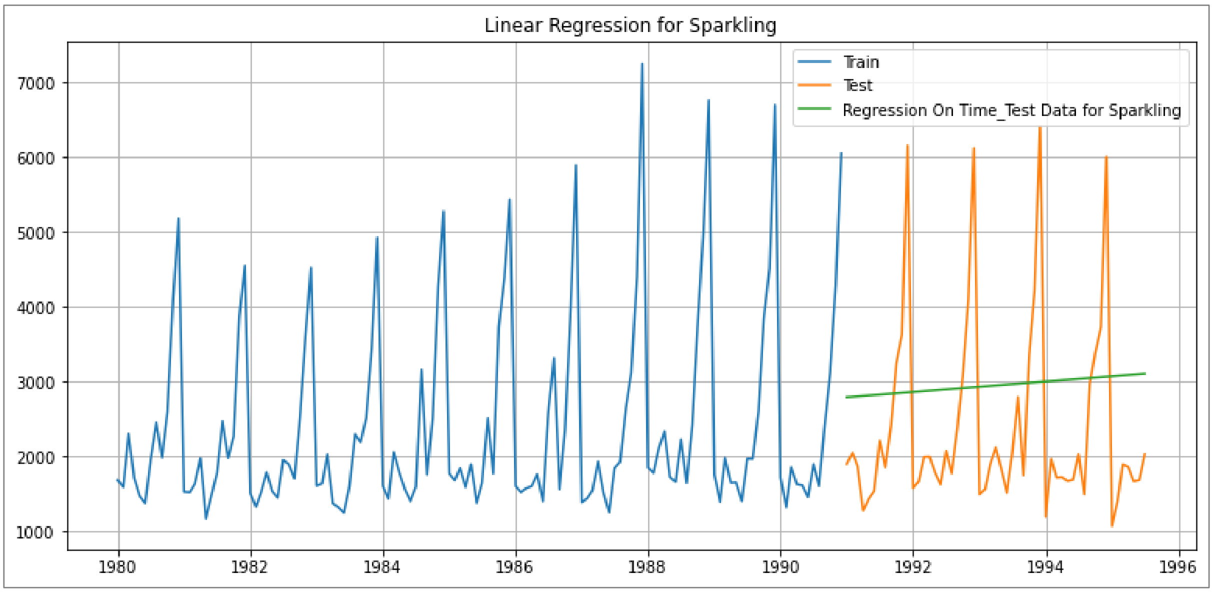

Figure 36. Plot of Linear Regression model for Sparkling.

Checking the performance of the model by calculating the RMSE value:

For Regression on time forecast on the Test Data for Sparkling, **RMSE is 1389.135.**

Linear Regression for Rose wine:

Firstly, creating training and testing time instances which is appended to the time series dataset:

```
Training Time instance for Rose
 [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 3
4, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65,
66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97,
98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123,
124, 125, 126, 127, 128, 129, 130, 131, 132]
Test Time instance for Rose
 [133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157,
158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 18
3, 184, 185, 186, 187]
```

Figure 37. Time Instances for Rose data

Looking at the first and last few data for training and testing for Rose data after appending the time instances.

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

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

Figure 39. Rose train data sample for regression.

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

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

Figure 38. Rose test data sample for regression.

Now that our training and test data has been modified, let us go ahead use *LinearRegression* to build the model on the training data and test the model on the test data.

To build a Linear Regression time series model for Sparkling:

- Fitting the Linear Regression model on train data which is imported from Sklearn linear model.
- Linear Regression uses the Gradient descent method for finding the best fit model, which uses partial derivatives on the parameters to minimize the sum of squared errors.
- Predict the values for test data and plot the graph.
- Checking the performance of the model by calculating the RMSE value.

Plotting Linear Regression for Rose data:

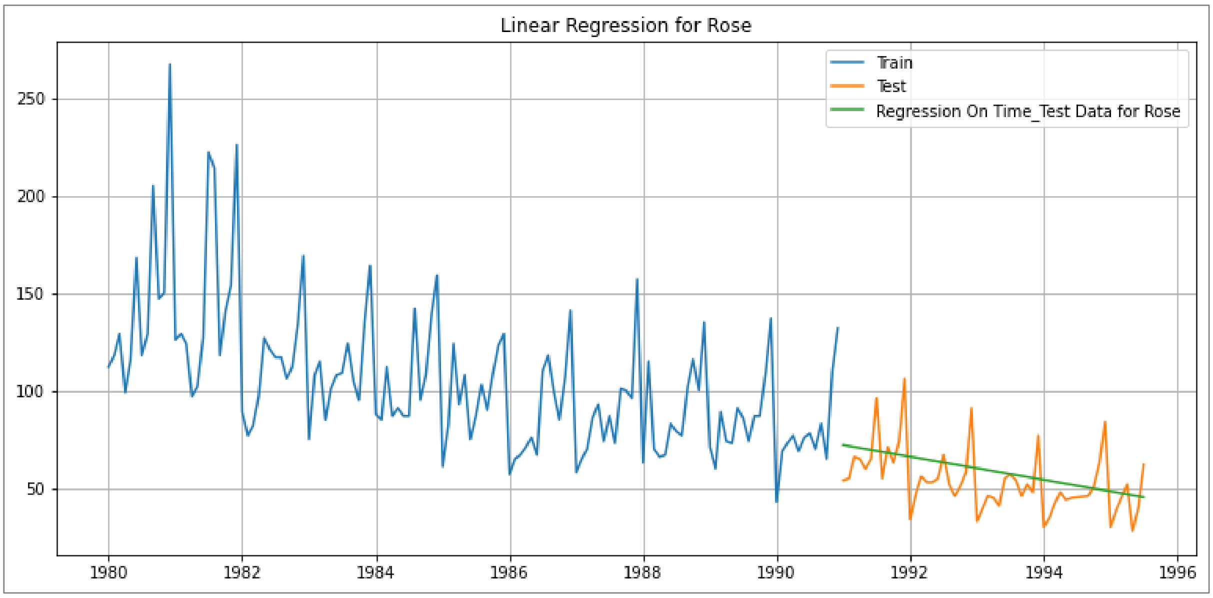

Figure 40. Plot of Linear Regression model for Rose.

Checking the performance of the model by calculating the RMSE value:

For Regression on time forecast on the Test Data for Rose, **RMSE is 15.268.**

Model 2 – Naïve Approach

Naive Methods such as assuming the predicted value at time ‘t’ to be the actual value of the variable at time ‘t-1’ or rolling mean of series, are used to weigh how well do the statistical models and machine learning models can perform and emphasize their need.

$$\hat{y}_{t+1} = y_t$$

A naive forecast involves using the previous observation directly as the forecast without any change. It is often called the persistence forecast as the prior observation is persisted.

Naive Approach for Sparkling:

Naïve forecasting is the technique in which the last period's sales are used for the next period's forecast without predictions or adjusting the factors. Forecasts produced using a naïve approach are equal to the final observed value.

In sparkling wine data, the last value of train dataset was 6047. Thus, our model will predict 6047 for every instance of the test data.

```
YearMonth
1991-01-01    6047
1991-02-01    6047
1991-03-01    6047
1991-04-01    6047
1991-05-01    6047
Name: naive, dtype: int64
```

Figure 41. Sparkling naive forecast values on test.

Plotting Graph for Naïve approach for Sparkling:

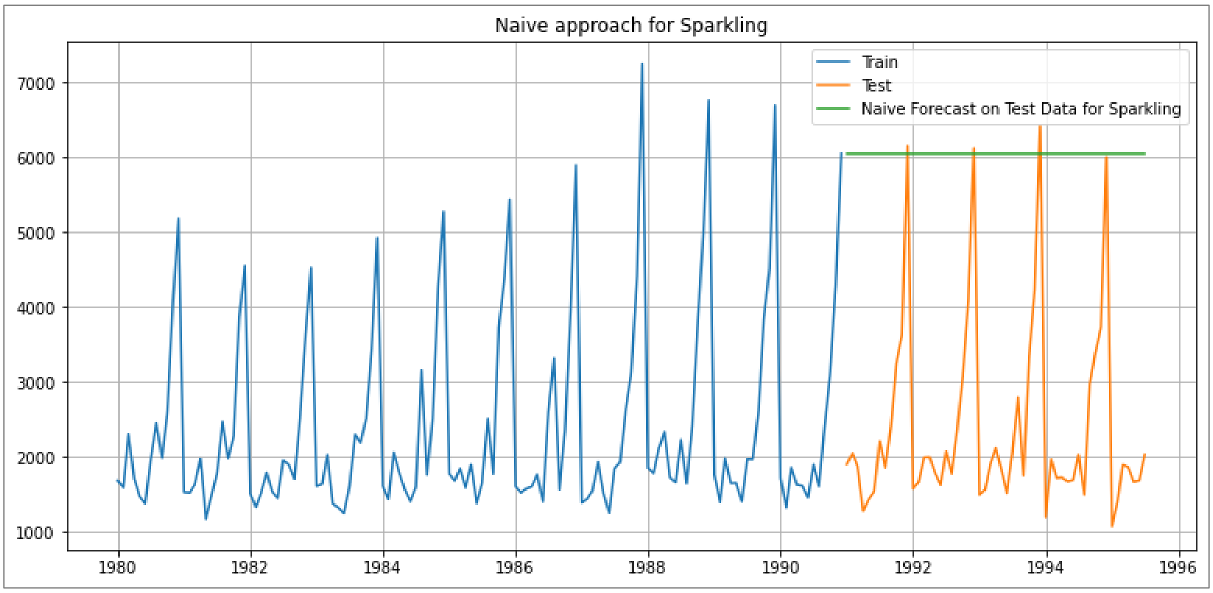

Figure 42. Plot of Naive approach for Sparkling.

Checking the performance of the model by calculating the RMSE value:

For Naïve Approach on time forecast on the Test Data for Sparkling, **RMSE is 3864.297.**

Naive Approach for Rose:

Naïve forecasting is the technique in which the last period's sales are used for the next period's forecast without predictions or adjusting the factors. Forecasts produced using a naïve approach are equal to the final observed value.

So, in this model we will take the last value of train dataset as our forecast value for the entire test data. The last value of train dataset was 132. Thus, our model will predict 132 for every instance of the test data

Plotting Graph for Naïve approach for Rose:

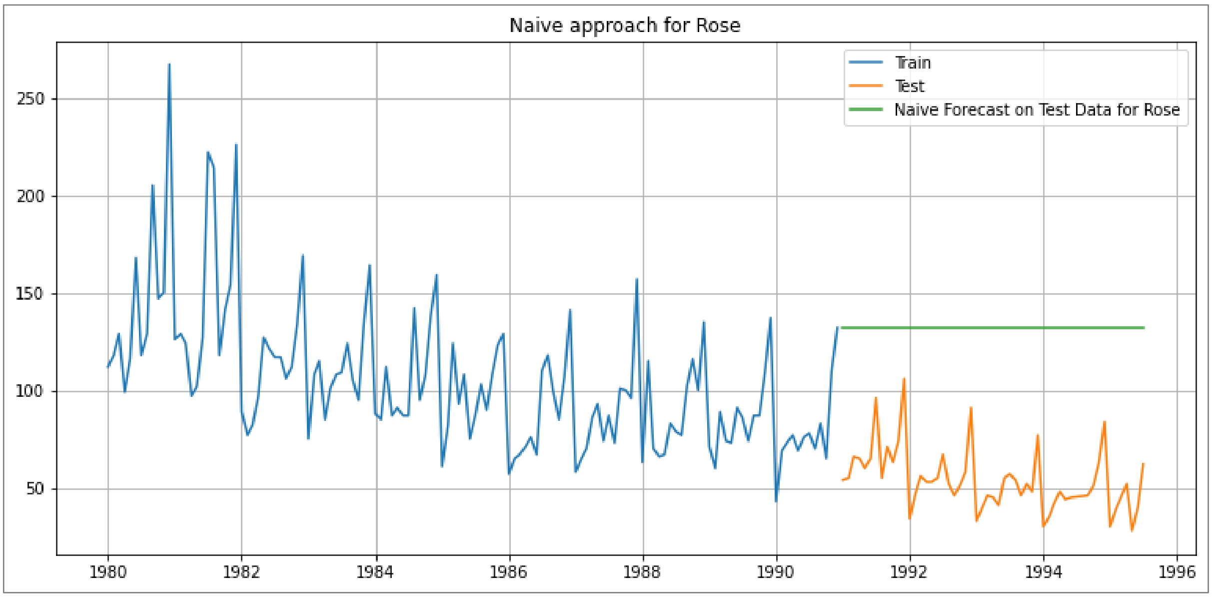

Figure 43. Plot of Naive approach for Rose.

Checking the performance of the model by calculating the RMSE value:

For Naïve Approach on time forecast on the Test Data for Rose, **RMSE is 79.719.**

Model 3: Simple Average

The simple average approach of time series forecasting is a very simple method of forecasting the values. We average the data by months or quarters or years and then calculate the average for the period.

In our case, the average of training data by months is considered for forecasting for the testing data.

Simple Average for Sparkling:

The average of sales of training data for sparkling is calculated which is the predicted forecast vale for the test data. In this case, the average of training data is 2403.780. Thus, our predicted values for the entire test data will be 2403.780

YearMonth	Sparkling	mean_forecast
1991-01-01	1902	2403.780303
1991-02-01	2049	2403.780303
1991-03-01	1874	2403.780303
1991-04-01	1279	2403.780303
1991-05-01	1432	2403.780303

Figure 44. Simple average of Sparkling

Plotting Graph for Simple Average for Sparkling:

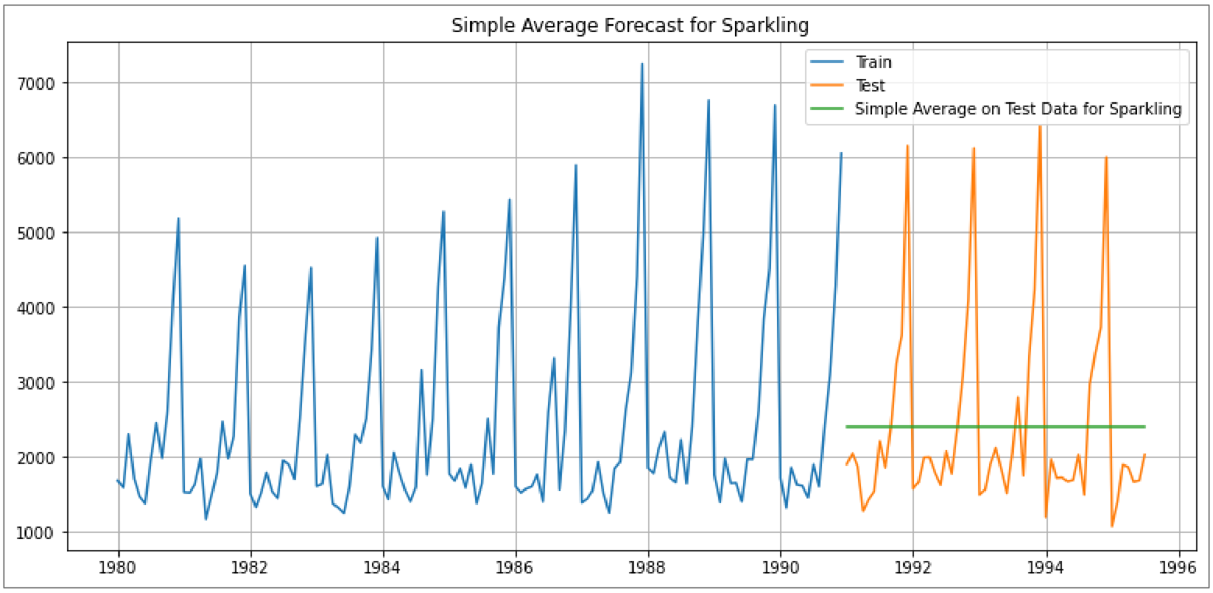

Figure 45. Plot of Simple Average for Sparkling.

From the plot above we can see that the predictions on test were made of the same value as the average of train data which is 2403.780.

Checking the root mean square error **(RMSE)** for the same, we get **1275.081.**

Simple Average for Rose:

The average of sales of training data for Rose is calculated which is the predicted forecast vale for the test data. In this case, the average of training data is 104.93. Thus, our predicted values for the entire test data will be 104.93.

YearMonth	Rose	mean_forecast
1991-01-01	54.0	104.939394
1991-02-01	55.0	104.939394
1991-03-01	66.0	104.939394
1991-04-01	65.0	104.939394
1991-05-01	60.0	104.939394

Figure 46. Simple average of Rose

Plotting Graph for Simple Average for Rose:

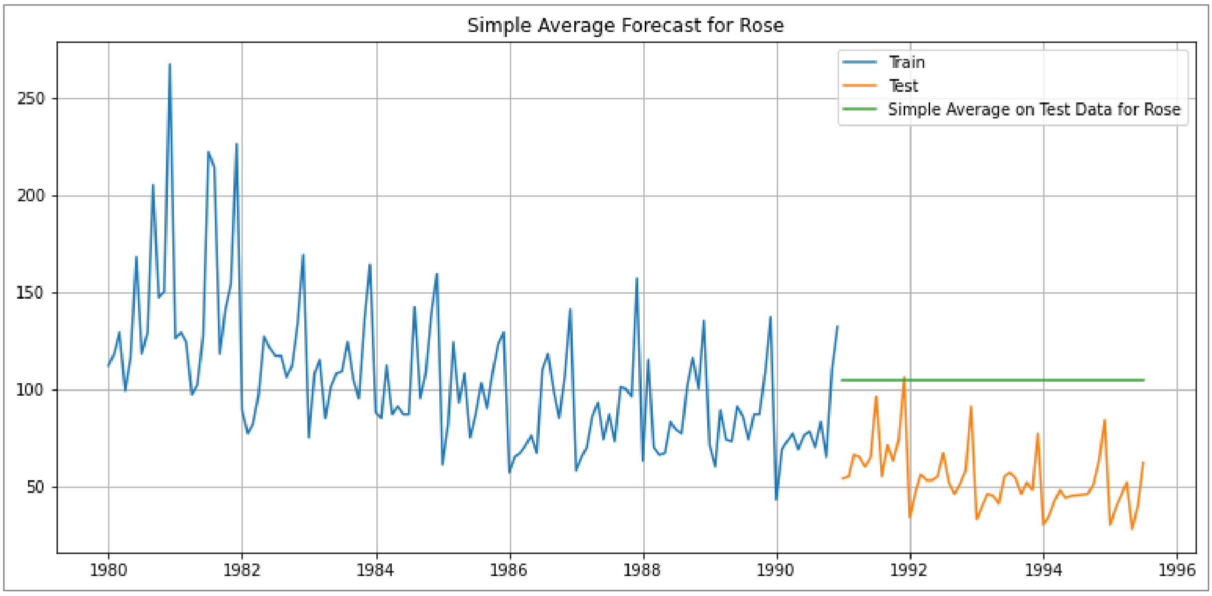

Figure 47. Plot of Simple Average for Rose.

From the plot above we can see that the predictions on test were made of the same value as the average of train data which is 104.93.

Checking the root mean square error **(RMSE)** for the same, we get **53.46.**

Model 4: Moving Average method

A moving average is defined as an average of fixed number of items in the time series which move through the series by dropping the top items of the previous averaged group and adding the next in each successive average.

For the moving average model, we are going to calculate rolling means (or moving averages) for different intervals. The best interval can be determined by the maximum accuracy (or the minimum error) over here.

For Moving Average, we are going to average over the entire data. we will be taking 4 different groupings. We will be considering 2, 4, 6 and 9 rolling time periods and take average of values.

Moving Average for Sparkling:

Firstly, considering the entire Sparkling dataset, and calculating the rolling averages of 2,4,6 and 9.

Checking the head of the moving averages for all the 4 rolling intervals.

YearMonth	Sparkling	Trailing_2	Trailing_4	Trailing_6	Trailing_9
1980-01-01	1686	NaN	NaN	NaN	NaN
1980-02-01	1591	1638.5	NaN	NaN	NaN
1980-03-01	2304	1947.5	NaN	NaN	NaN
1980-04-01	1712	2008.0	1823.25	NaN	NaN
1980-05-01	1471	1591.5	1769.50	NaN	NaN

Figure 48. Moving Average for Sparkling

Plotting the Graph for Moving Average for entire Sparkling data:

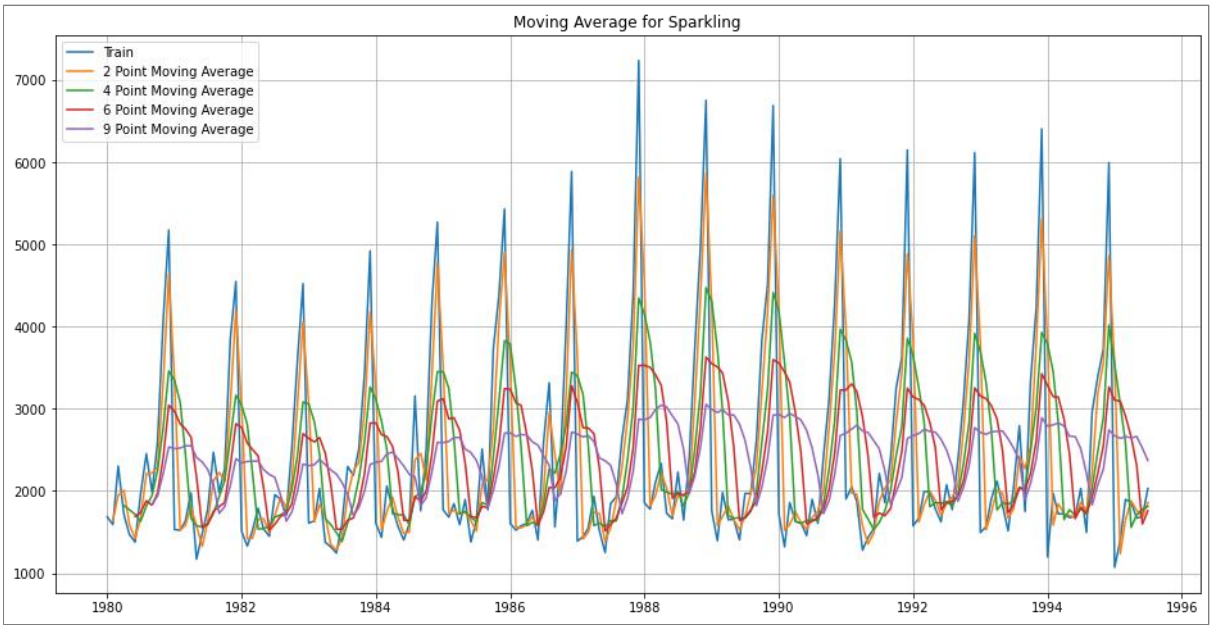

Figure 49. Plot of Moving average for entire Sparkling data.

Now creating a train and test data and calculating the moving averages on train and forecasting on train for all the 4 groups of rolling averages of 2, 4, 6 and 8 and calculating the RMSE on the test data for the same.

For 2 point Moving Average Model forecast on the Training Data for Sparkling, RMSE **is 813.401**

For 4 point Moving Average Model forecast on the Training Data for Sparkling, RMSE **is 1156.590**

For 6 point Moving Average Model forecast on the Training Data for Sparkling, RMSE **is 1283.927**

For 9 point Moving Average Model forecast on the Training Data for Sparkling, RMSE **is 1346.278**

Plotting the Graph for Moving Average for Sparkling of both train and test data:

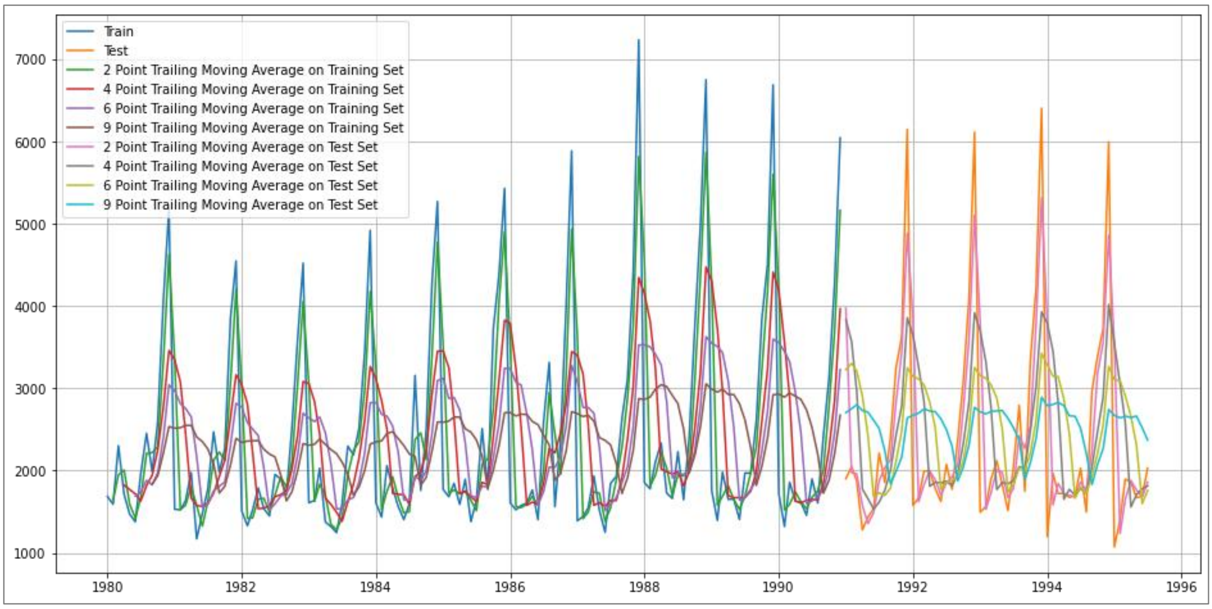

Figure 50. Plot of Moving average for Sparkling for train and test data.

We can see from the graphs that as we increase the rolling period from 2 to 4 to 6 to 9, the noise in the data keeps on decreasing and the graph gets smoother and smoother. Evidently, taking moving averages helps in reducing the noise in the data and makes the graph smoother.

The RMSE for moving average with **2-point rolling is performing better**. \

Moving Average for Rose:

Firstly, considering the entire Rose dataset, and calculating the rolling averages of 2,4,6 and 9.

Checking the head of the moving averages for all the 4 rolling intervals.

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

Figure 51. Moving Average for Rose.

Plotting the Graph for Moving Average for entire Rose data:

Plotting the graph for entire Rose dataset, we can see from the graphs that as we increase the rolling period from 2 to 4 to 6 to 9, the noise in the data keeps on decreasing and the graph gets smoother and smoother. Evidently, taking moving averages helps in reducing the noise in the data and makes the graph smoother.

Smaller the rolling point of moving average, lesser the RMSE value and more accurate the model performance.

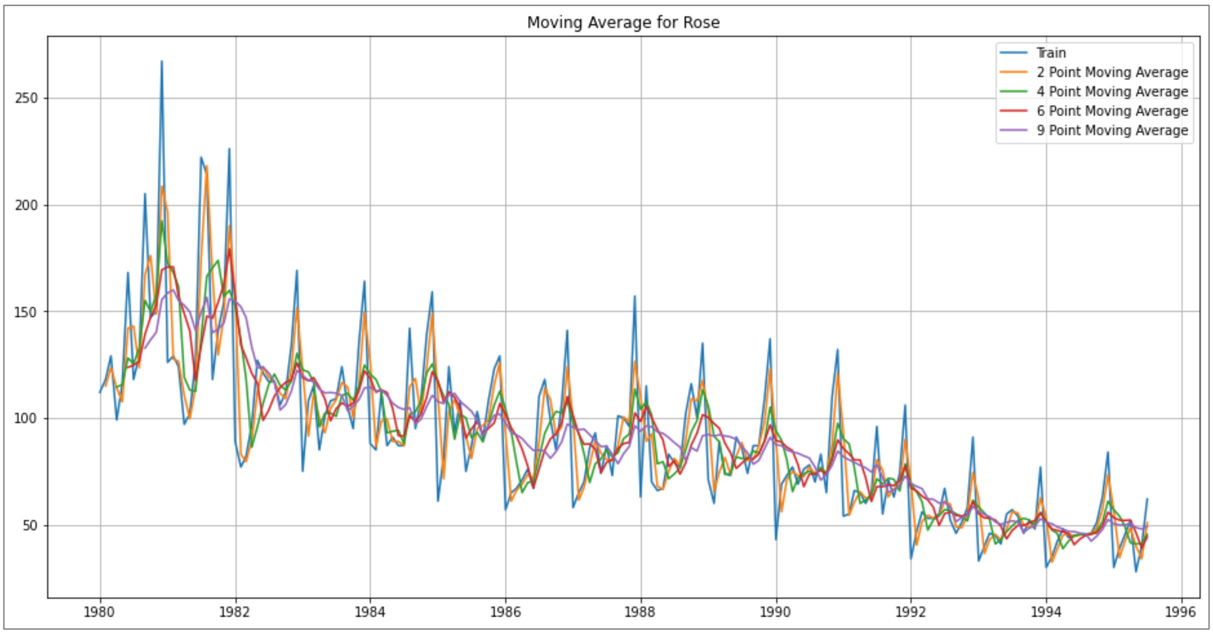

Figure 52. Plot of Moving average for entire Rose data

Now creating a train and test data and calculating the moving averages on train and forecasting on train for all the 4 groups of rolling averages of 2, 4, 6 and 8 and calculating the RMSE on the test data for the same.

Plotting the Graph for Moving Average for Rose of both train and test data:

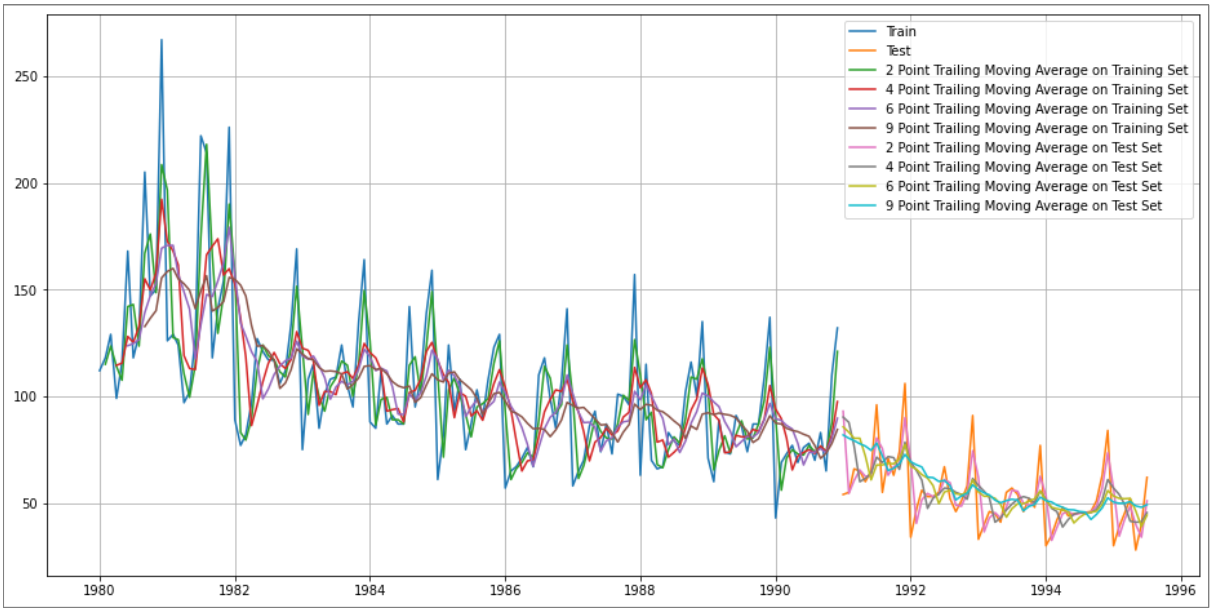

Figure 53. Plot of Moving average for Rose of train and test data

Checking the RMSE for test data:

For 2 point Moving Average Model forecast on the Training Data for Rose, **RMSE is 11.529**

For 4 point Moving Average Model forecast on the Training Data for Rose, **RMSE is 14.451**

For 6 point Moving Average Model forecast on the Training Data for Rose, **RMSE is 14.566**

For 9 point Moving Average Model forecast on the Training Data for Rose, **RMSE is 14.728**

The RMSE for moving average with **2-point rolling is performing better**.

<u>Exponential smoothing model</u>

- Exponential smoothing methods consist of flattening time series data.
- Exponential smoothing averages or exponentially weighted moving averages consist of forecast based on previous periods data with exponentially declining influence on the older observations.
- Exponential smoothing methods consist of special case exponential moving with notation ETS (Error, Trend, Seasonality) where each can be none(N), additive (N), additive damped (Ad), Multiplicative (M) or multiplicative damped (Md).
- One or more parameters control how fast the weights decay.
- These parameters have values between 0 and 1

<u>Model 5 - Simple Exponential Smoothing (SES) - ETS(A, N, N)</u>

The simplest of the exponentially smoothing methods is naturally called simple exponential smoothing (SES). This method is suitable for forecasting data with no clear trend or seasonal pattern.

In Single ES, the forecast at time (t + 1) is given by Winters,1960

$$\boldsymbol{Ft+1=\alpha Yt+(1-\alpha)Ft}\mathbf{Ft+1=\alpha Yt+(1-\alpha)Ft}$$

Parameter αα is called the smoothing constant and its value lies between 0 and 1. Since the model uses only one smoothing constant, it is called Single Exponential Smoothing.

Here, there is both trend and seasonality in the data. So, we should have directly gone for the Triple Exponential Smoothing but Simple Exponential Smoothing and the Double Exponential Smoothing models are built over here to get an idea of how the three types of models compare in this case.

Simple Exp Smoothing class must be instantiated and passed the training data. The fit() function is then called providing the fit configuration, the alpha value, smoothing_level. If this is omitted or set to None, the model will automatically optimize the value.

<u>Simple Exponential Smoothing for Sparkling:</u>

To build a SES time series model for Sparkling:

- Fitting the Simple Exponential smoothing model on train data which is imported from Sklearn linear model
- Getting the autofit parameters with optimized = True.
- Forecasting on the test data using the autofit parameters.
- Checking the performance of the model by calculating the RMSE value.

```
{'smoothing_level': 0.049607360581862936,
 'smoothing_trend': nan,
 'smoothing_seasonal': nan,
 'damping_trend': nan,
 'initial_level': 1818.535750008871,
 'initial_trend': nan,
 'initial_seasons': array([], dtype=float64),
 'use_boxcox': False,
 'lamda': None,
 'remove_bias': False}
```

Figure 54. Sparkling SES parameters.

As Simple exponential smoothing model only works with the level, trend and seasonality parameters are blank. **Smoothing level parameter (Alpha) = 0.0496.**

Getting the test predictions on the autofit parameters:

YearMonth	Sparkling	predict
1991-01-01	1902	2724.932624
1991-02-01	2049	2724.932624
1991-03-01	1874	2724.932624
1991-04-01	1279	2724.932624
1991-05-01	1432	2724.932624

Figure 55. Sparking SES test predictions.

Plotting the graph of forecast on the test data and look at the predictions:

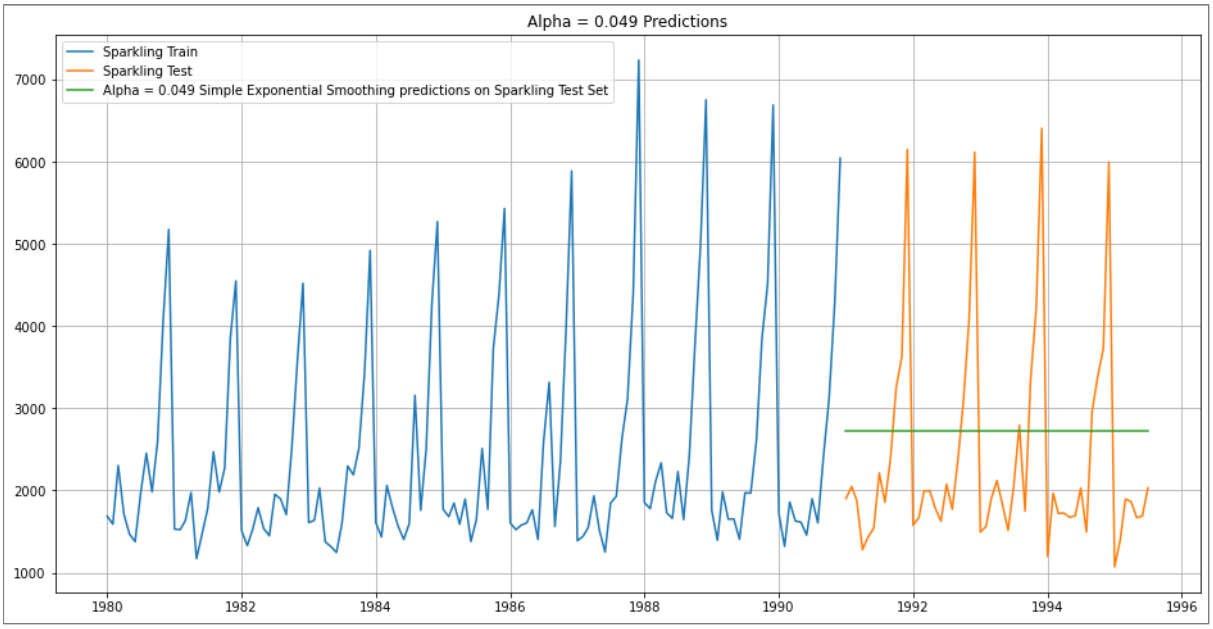

Figure 56. Plot of Sparkling SES.

We get a predicted value of 2724.932 for the entire test data. If we check the **RMSE** score for this mode, it comes out to be **1316.035.**

Tuned Simple Exponential Smoothing for Sparkling:

Setting different alpha values. We know that, the higher the alpha value more weightage is given to the more recent observation. That means, what happened recently will happen again**.** We will run a loop with different alpha values to understand which particular value works best for alpha on the test set.

We can also tune this model by checking RMSE score for different values of alpha and pick the best fit model which gives us the minimum RMSE. We run a loop in range from 0.01 to 0.1 with 0.01 difference and consider each value as alpha. We create a table with these alpha values and sort the RMSE values in descending to look for the best alpha value.

	Alpha Values	Train RMSE	Test RMSE
1	0.02	1328.406554	1279.495201
0	0.01	1361.997529	1281.032699
2	0.03	1318.846031	1293.110073
3	0.04	1317.138929	1305.462953
4	0.05	1318.429335	1316.411742

Figure 57. Sparkling SES RMSE Scores.

From the above table, it is evident that **alpha = 0.02 gives the minimum RMSE score of 1279.495.**

Plotting the graph for Tuned Simple Exponential Smoothing Predictions for Sparkling:

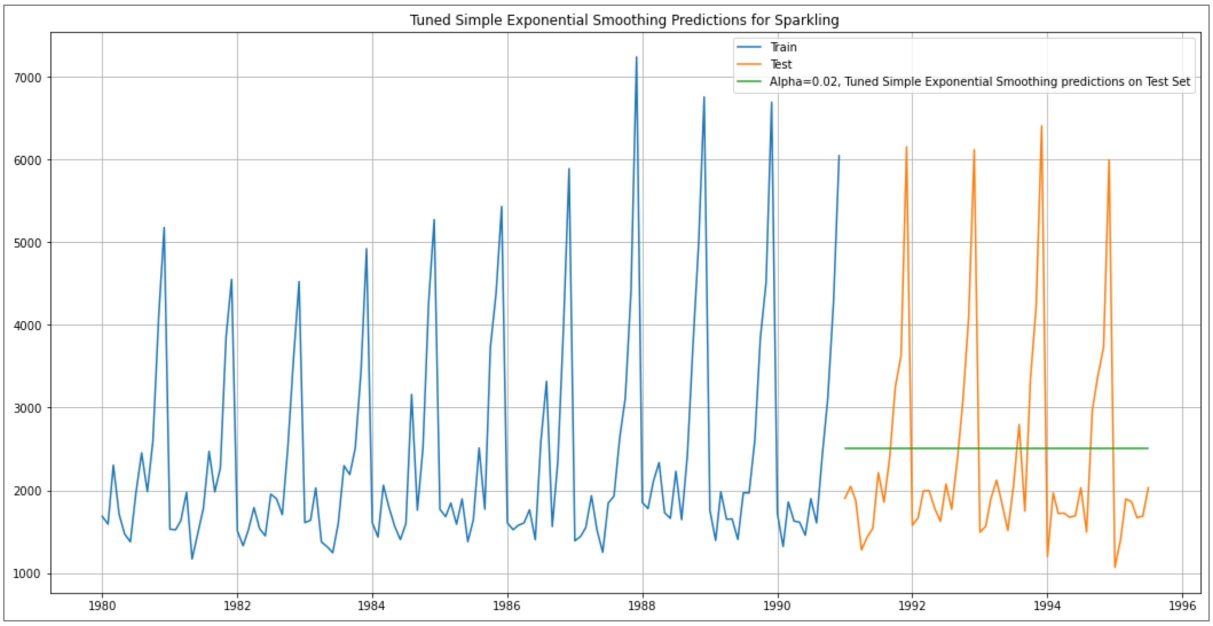

Figure 58. Plot of Tuned SES predictions for Sparkling.

A prediction of approximately 2500 is made for the test data using the tuned Simple exponential smoothing model. The RMSE for test is reduced from **1316.035 to 1279.495.**

Simple Exponential Smoothing for Rose:

To build a SES time series model for Rose:

- Fitting the Simple Exponential smoothing model on train data which is imported from Sklearn linear model
- Getting the autofit parameters with optimized = True.
- Forecasting on the test data using the autofit parameters.
- Checking the performance of the model by calculating the RMSE value.

```
{'smoothing_level': 0.0987493111726833,
 'smoothing_trend': nan,
 'smoothing_seasonal': nan,
 'damping_trend': nan,
 'initial_level': 134.38720226208358,
 'initial_trend': nan,
 'initial_seasons': array([], dtype=float64),
 'use_boxcox': False,
 'lamda': None,
 'remove_bias': False}
```

Figure 59. Rose SES parameters.

As Simple exponential smoothing model only works with the level, trend and seasonality parameters are blank. **Smoothing level parameter (Alpha) = 0.098.**

Getting the test predictions on the autofit parameters:

YearMonth	Rose	predict
1991-01-01	54.0	87.104983
1991-02-01	55.0	87.104983
1991-03-01	66.0	87.104983
1991-04-01	65.0	87.104983
1991-05-01	60.0	87.104983

Figure 60. Rose SES test predictions.

Plotting the graph of forecast on the test data and look at the predictions foe Rose:

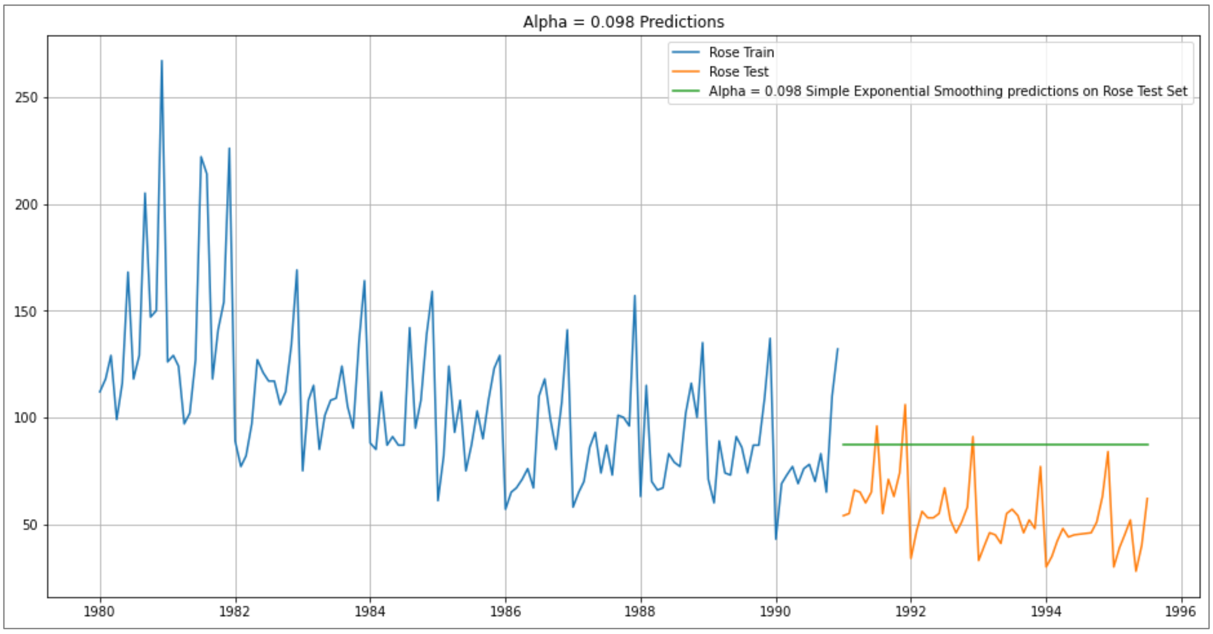

Figure 61. Plot of Rose SES.

We get a predicted value of 87.104 for the entire test data. If we check the **RMSE** score for this model, it comes out to be **36.796.**

Tuned Simple Exponential Smoothing for Rose:

Setting different alpha values. We know that, the higher the alpha value more weightage is given to the more recent observation. That means, what happened recently will happen again**.** We will run a loop with different alpha values to understand which particular value works best for alpha on the test set.

We can also tune this model by checking RMSE score for different values of alpha and pick the best fit model which gives us the minimum RMSE. We run a loop in range from 0.01 to 0.1 with 0.01 difference and consider each value as alpha. We create a table with these alpha values and sort the RMSE values in descending to look for the best alpha value.

	Alpha Values	Train RMSE	Test RMSE
6	0.07	32.046904	36.435772
7	0.08	31.936243	36.462965
5	0.06	32.209657	36.580469
8	0.09	31.862435	36.604118
9	0.10	31.815610	36.828033

Figure 62. Rose SES RMSE Scores

From the above table, it is evident that **alpha = 0.07 gives the minimum RMSE score of 36.435.**

Plotting the graph for Tuned Simple Exponential Smoothing Predictions for Rose:

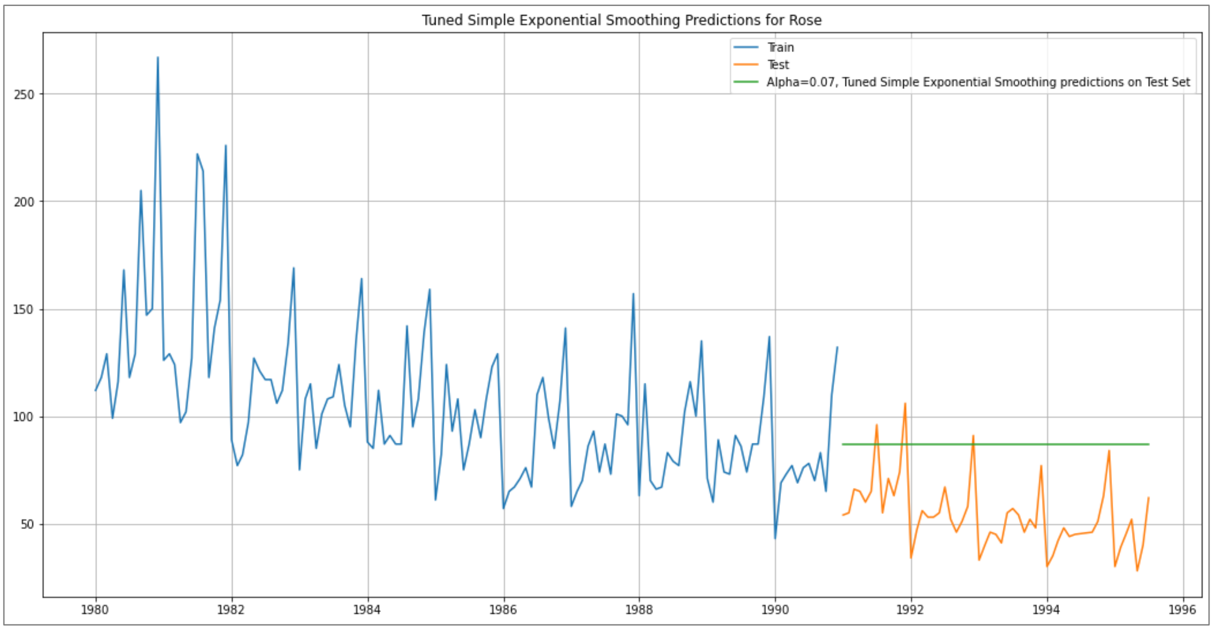

Figure 63. Plot of Tuned SES predictions for Rose.

A prediction of approximately 85 is made for the test data using the tuned Simple exponential smoothing model. The RMSE for test is reduced from **36.796 to 36.435.** No much difference.

Model- 6 Double Exponential Smoothing (DES) Holt - ETS(A, A, N)

- One of the drawbacks of the simple exponential smoothing is that the model does not do well in the presence of the trend.
- This model is an extension of SES known as Double Exponential model which estimates two smoothing parameters.
- Applicable when data has Trend but no seasonality.
- Two separate components are considered: Level and Trend.
- Level is the local mean.
- One smoothing parameter α corresponds to the level series
- A second smoothing parameter β corresponds to the trend series.

Double Exponential Smoothing uses two equations to forecast future values of the time series, one for forecasting the short term average value or level and the other for capturing the trend.

- Intercept or Level equation, L_tLt is given by: $L_t=\alpha Y_t+(1-\alpha)F_t$Lt=αYt+(1−α)Ft
- Trend equation is given by $T_t=\beta(L_t-L_{t-1})+(1-\beta)T_{t-1}$Tt=β(Lt−Lt−1)+(1−β)Tt−1

Here, αα and ββ are the smoothing constants for level and trend, respectively,

- 0 <αα < 1 and 0 < ββ < 1.

The forecast at time t + 1 is given by

- $F_{t+1}=L_t+T_t$Ft+1=Lt+Tt
- $F_{t+n}=L_t+nT_t$

Double Exponential Smoothing for Sparkling (Holt's Model):

To build a DES time series model for Sparkling:

- Initializing the Double Exponential Smoothing Model
- Fitting the Double Exponential smoothing model on train data which is imported from Sklearn linear model
- Getting the autofit parameters with optimized = True.
- Forecasting on the test data using the autofit parameters.
- Checking the performance of the model by calculating the RMSE value.

Checking the parameters after fitting the DES model for training data:

{'smoothing_level': 0.6649999999999999,
'smoothing_trend': 0.0001,
'smoothing_seasonal': nan,
'damping_trend': nan,
'initial_level': 1502.1999999999991,
'initial_trend': 74.87272727272739,
'initial_seasons': array([],
dtype=float64), 'use_boxcox': False,
'lamda': None,
'remove_bias': False}

As Double exponential smoothing model only works with the level and trend; seasonality parameter is blank.

- Smoothing level parameter (Alpha) = 0.6649
- Smoothing trend parameter (Beta) = 0.0001

Getting the test predictions from the parameters:
Below is the output of head (10) of test predictions forecasted using this model for the duration of the test set.

```
1991-01-01    5401.733026
1991-02-01    5476.005230
1991-03-01    5550.277433
1991-04-01    5624.549637
1991-05-01    5698.821840
1991-06-01    5773.094044
1991-07-01    5847.366248
1991-08-01    5921.638451
1991-09-01    5995.910655
1991-10-01    6070.182858
```

<u>Plotting the graph of forecast on the test data and look at the predictions for Sparkling:</u>

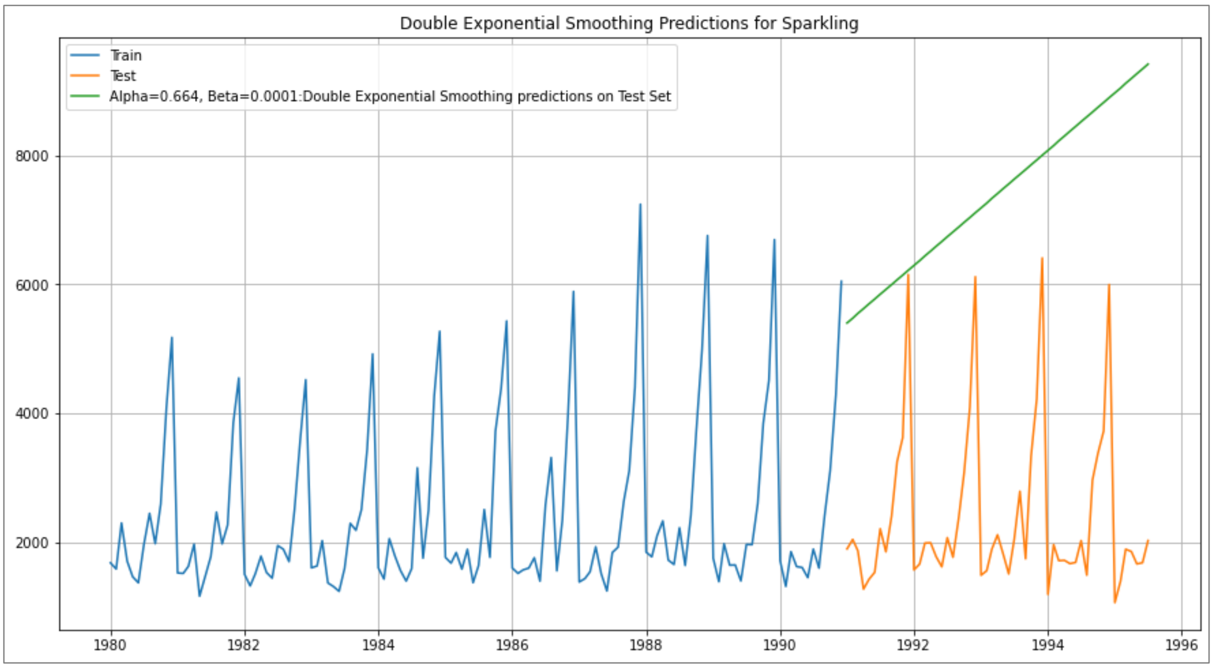

Figure 64. Plot for Sparkling DES

We see that the double exponential smoothing is picking up the trend component along with the level component as well. The **RMSE** score for this model, it comes out to be **5291.879**

<u>Tuned Double Exponential Smoothing for Sparkling:</u>
We can also tune this model by checking RMSE score for different values of alpha & Beta and pick the best fit model which gives us the minimum RMSE. We run two loops in range from 0.1 to 1.0 with 0.1 difference and consider each value as alpha from one loop and Beta from other loop. We create a table of different combinations of Alpha and Beta and look for the top 5 kept in the descending order of their RMSE score.

	Alpha Values	Beta Values	Train RMSE	Test RMSE
0	0.1	0.1	1382.520870	1778.564670
1	0.1	0.2	1413.598835	2599.439986
10	0.2	0.1	1418.041591	3611.763322
2	0.1	0.3	1445.762015	4293.084674
20	0.3	0.1	1431.169601	5908.185554

Figure 65. Sparkling DES Scores

Now we take **Alpha = 0.1** and **Beta = 0.1** and forecast the values on test data and plot these predictions on the graph and **gives the minimum RMSE score of 1778.56**

<u>Plotting the graph for Tuned Double Exponential Smoothing Predictions for Sparkling:</u>

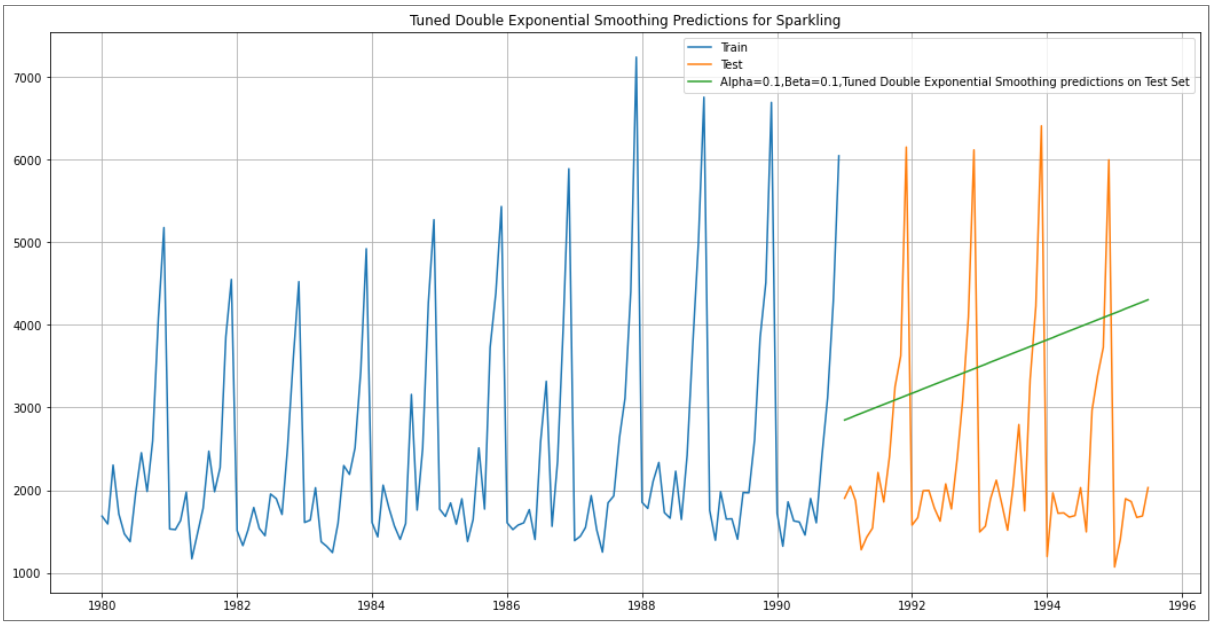

Figure 66. Plot of Tuned DES predictions for Sparkling

The tuned model gives us a better prediction than the original model we built. If we check the RMSE score has reduced from **5291.879 to 1778.56.**

<u>Double Exponential Smoothing for Rose (Holt's Model)</u>

To build a DES time series model for Sparkling:

- Initializing the Double Exponential Smoothing Model
- Fitting the Double Exponential smoothing model on train data which is imported from Sklearn linear model
- Getting the autofit parameters with optimized = True.
- Forecasting on the test data using the autofit parameters.
- Checking the performance of the model by calculating the RMSE value.

Checking the parameters after fitting the DES model for training data:

{'smoothing_level': 1.4901161193847656e-08,
'smoothing_trend': 1.6610391146660035e-10,
'smoothing_seasonal': nan,
'damping_trend': nan,
'initial_level': 137.81553690867275,
'initial_trend': -0.4943781897068274,
'initial_seasons': array([],
dtype=float64),
'use_boxcox': False,
'lamda': None,
'remove_bias': False}

As Double exponential smoothing model only works with the level and trend; seasonality parameter is blank.

- Smoothing level parameter (Alpha) = 1.490-08
- Smoothing trend parameter (Beta) = 1.661-10

Getting the test predictions from the parameters:
Below is the output of head (10) of test predictions forecasted using this model for the duration of the test set.

```
1991-01-01    72.063238
1991-02-01    71.568859
1991-03-01    71.074481
1991-04-01    70.580103
1991-05-01    70.085725
1991-06-01    69.591347
1991-07-01    69.096969
1991-08-01    68.602590
1991-09-01    68.108212
1991-10-01    67.613834
```

<u>Plotting the graph of forecast on the test data and look at the predictions for Rose:</u>

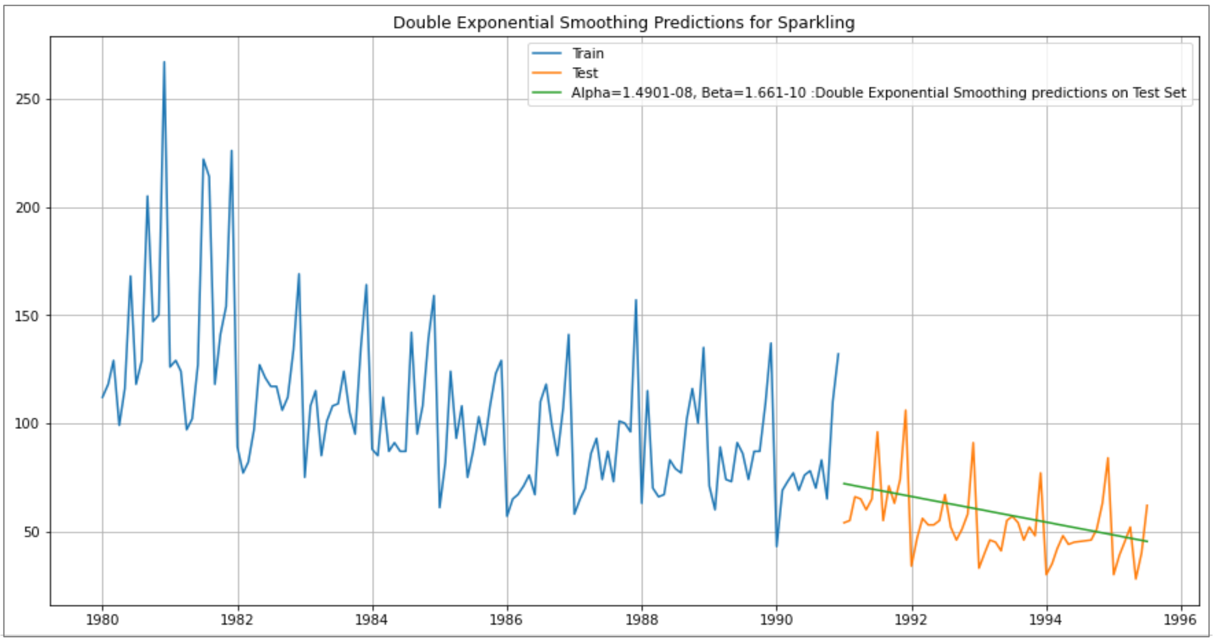

Figure 67. Plot for Rose DES

We see that the double exponential smoothing is picking up the trend component along with the level component as well. The **RMSE** score for this model, it comes out to be **15.268.**

<u>Tuned Double Exponential Smoothing for Sparkling:</u>

We can also tune this model by checking RMSE score for different values of alpha & Beta and pick the best fit model which gives us the minimum RMSE. We run two loops in range from 0.1 to 1.0 with 0.1 difference and consider each value as alpha from one loop and Beta from other loop. We create a table of different combinations of Alpha and Beta and look for the top 5 kept in the descending order of their RMSE score.

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

Figure 68. Rose DES Scores

Now we take **Alpha = 0.1** and **Beta = 0.1** and forecast the values on test data and plot these predictions on the graph and **gives the minimum RMSE score of 36.923**

<u>Plotting the graph for Tuned Double Exponential Smoothing Predictions for Rose:</u>

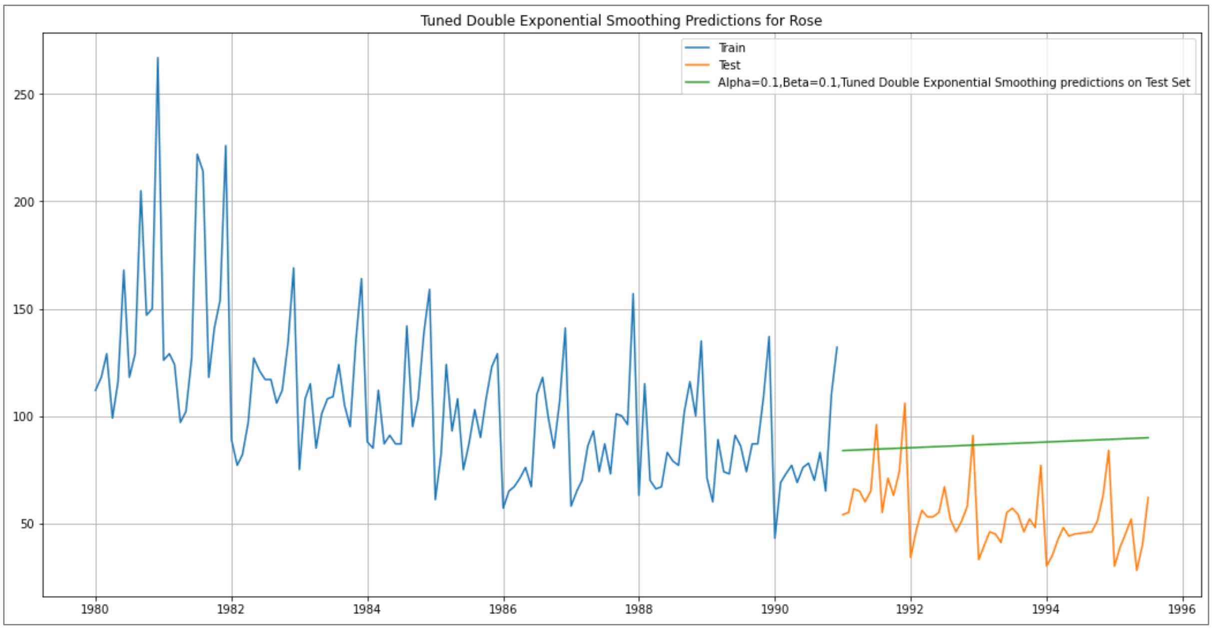

Figure 69. Plot of Tuned DES predictions for Rose.

Here, we see that the Double Exponential Smoothing has actually done well when compared to the Simple Exponential Smoothing. This is because of the fact that the Double Exponential Smoothing model has picked up the trend component as well.

Model- 3 Triple Exponential Smoothing (TES) Holt Winter's - ETS(A, A, M)

Triple exponential smoothing is used to handle the time series data containing a seasonal component. This method is based on three smoothing equations: stationary component, trend, and seasonal. Both seasonal and trend can be additive or multiplicative.

Triple Exponential Smoothing for Sparkling (Holt Winter's Model)

To build a TES time series model for Sparkling:

- Initializing the Triple Exponential Smoothing Model
- Fitting the Triple Exponential smoothing model on train data which is imported from Sklearn linear model
- Getting the autofit parameters with optimized = True.
- Forecasting on the test data using the autofit parameters.
- Checking the performance of the model by calculating the RMSE value.

Checking the parameters after fitting the TES model for training data:

```
{'smoothing_level': 0.11106668752955826,
 'smoothing_trend': 0.04936072355729082,
 'smoothing_seasonal': 0.3621821387810734,
 'damping_trend': nan,
 'initial_level': 2360.4089797373545,
 'initial_trend': 0.9992288111047797,
 'initial_seasons': array([0.71936124, 0.6984697 , 0.90024844, 0.80991063, 0.66820986,
        0.66898271, 0.87875613, 1.11648842, 0.90067181, 1.17297733,
        1.82687893, 2.27815792]),
 'use_boxcox': False,
 'lamda': None,
 'remove_bias': False}
```

Figure 70. Sparkling TES parameters.

Triple exponential smoothing model works with smoothing level, trend as well as seasonality.

- Smoothing level parameter (Alpha) = 0.111
- Smoothing trend parameter (Beta) = 0.0493
- Smoothing seasonal parameter (Gamma) = 0.3621

Getting the test predictions from the parameters:

Below is the output of head (first 5 observations) of test predictions forecasted using this model for the duration of the test set.

YearMonth	Sparkling	auto_predict
1991-01-01	1902	1591.299973
1991-02-01	2049	1360.408886
1991-03-01	1874	1767.949510
1991-04-01	1279	1661.619432
1991-05-01	1432	1547.414170

Figure 71. Sparkling TES test predictions.

Plotting the graph of forecast on the test data and look at the predictions for Sparkling:

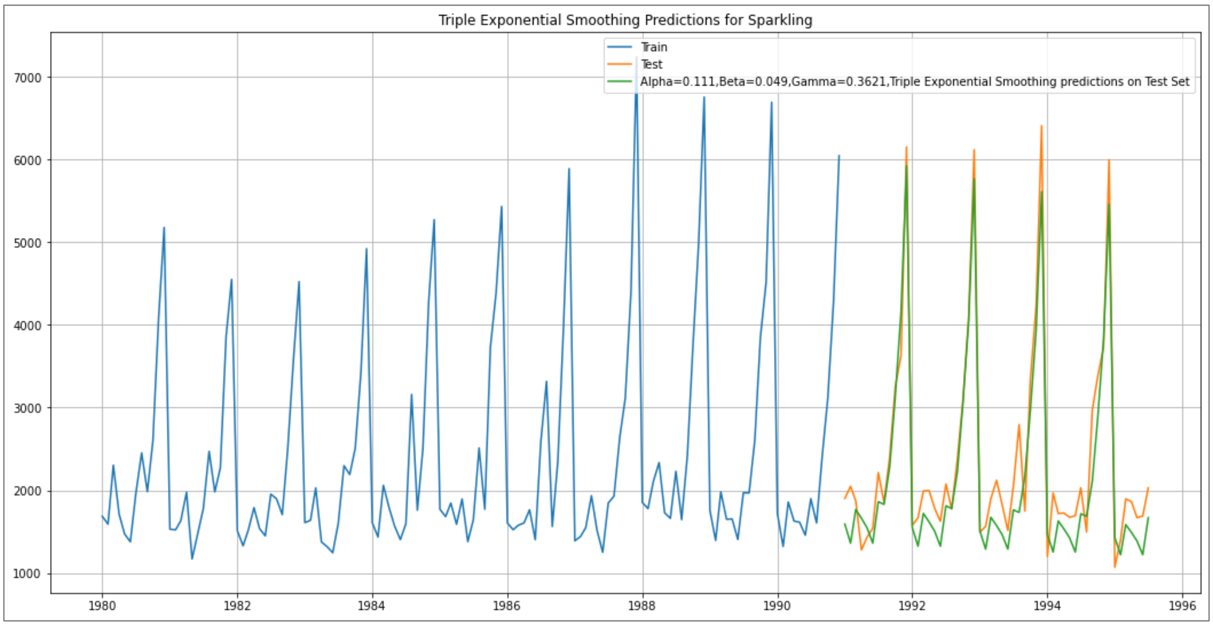

Figure 72. Plot for Sparkling TES

We check for the **RMSE** score for this model and it comes out to be **380.398**. The predictions are very aligned with the original test data as soon as seasonality is also added in the smoothing mode.

Tuning Triple Exponential Smoothing for Sparkling (Holt Winter's Model)

We can also tune this model by checking RMSE score for different values of alpha, Beta & Gamma and pick the best fit model which gives us the minimum RMSE.

We run three loops in range from 0.1 to 1.0 with 0.1 difference and consider each value as alpha from one loop and Beta from other loop and Gamma from the third loop. We create a table of different combinations of Alpha, Beta & Gamma and look for the top 5 kept in the descending order of their RMSE score.

	Alpha Values	Beta Values	Gamma Values	Train RMSE	Test RMSE
302	0.4	0.1	0.3	381.106645	326.579641
201	0.3	0.1	0.2	375.956510	342.464413
110	0.2	0.2	0.1	395.987244	345.931571
131	0.2	0.4	0.2	401.704682	349.425739
222	0.3	0.3	0.3	396.692796	353.602587

Figure 73. Sparkling TES Scores

From the above table, it is evident that alpha = 0.4, Beta = 0.1 & Gamma = 0.3 gives the minimum RMSE score of 326.579.

Till now, this model has been the best with the minimum RMSE score.

Now we take Alpha = 0.4, Beta = 0.1 & Gamma = 0.3 and forecast the values on test data and plot these predictions on the graph.

Plotting the graph for Tuned Triple Exponential Smoothing Predictions for Sparkling:

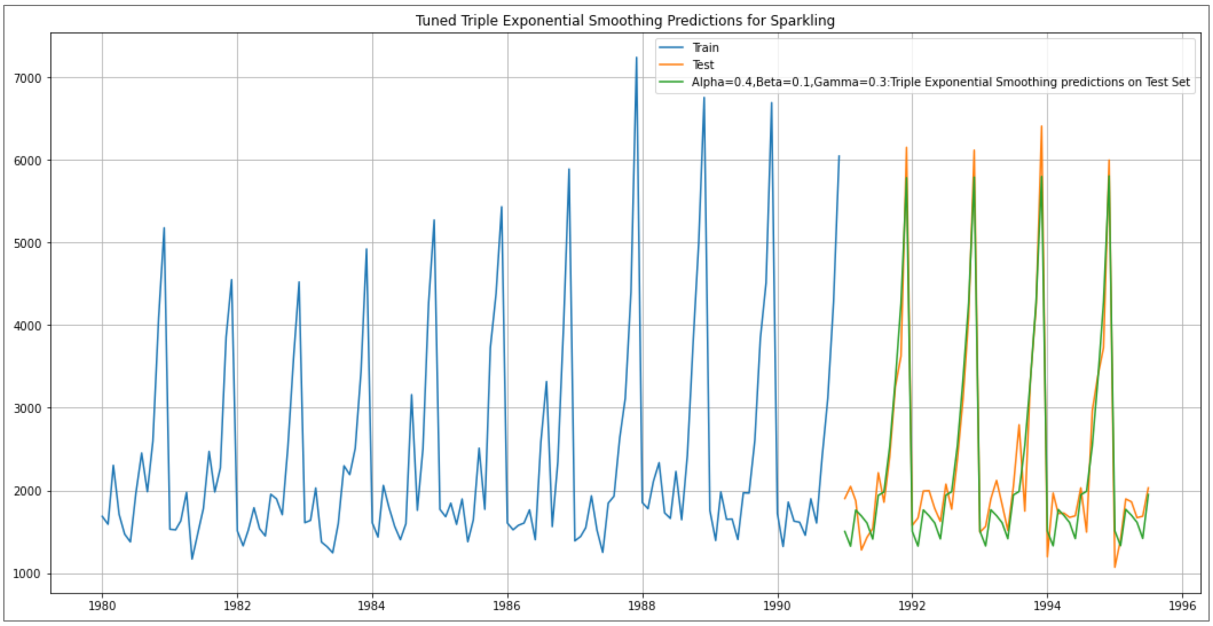

Figure 74. Plot of Tuned TES predictions for Sparkling,

The predictions have become much better on test model than our original model as we can see from the graph.

Triple Exponential Smoothing for Rose (Holt Winter's Model)

To build a TES time series model for Rose:

- Initializing the Triple Exponential Smoothing Model
- Fitting the Triple Exponential smoothing model on train data which is imported from Sklearn linear model
- Getting the autofit parameters with optimized = True.
- Forecasting on the test data using the autofit parameters.
- Checking the performance of the model by calculating the RMSE value.

Checking the parameters after fitting the TES model for training data of Rose:

```
{'smoothing_level': 0.05509258651447915,
 'smoothing_trend': 0.03163443011388579,
 'smoothing_seasonal': 0.00033441920536960617,
 'damping_trend': nan,
 'initial_level': 162.24448448772696,
 'initial_trend': 0.9924159109944972,
 'initial_seasons': array([0.69939026, 0.79380649, 0.86893412, 0.75865299, 0.85377453,
        0.9282575 , 1.02003364, 1.08767274, 1.03068915, 1.00761385,
        1.17626069, 1.61916255]),
 'use_boxcox': False,
 'lamda': None,
 'remove_bias': False}
```

Figure 75. Rose TES parameters

Triple exponential smoothing model works with smoothing level, trend as well as seasonality.

- Smoothing level parameter (Alpha) = 0.055
- Smoothing trend parameter (Beta) = 0.0316
- Smoothing seasonal parameter (Gamma) = 0.000334

Getting the test predictions from the parameters:

Below is the output of head (first 5 observations) of test predictions forecasted using this model for the duration of the test set.

YearMonth	Rose	auto_predict
1991-01-01	54.0	55.663816
1991-02-01	55.0	62.993228
1991-03-01	66.0	68.738503
1991-04-01	65.0	59.835212
1991-05-01	60.0	67.118704

Figure 76. Rose TES test predictions.

<u>Plotting the graph of forecast on the test data and look at the predictions for Rose:</u>

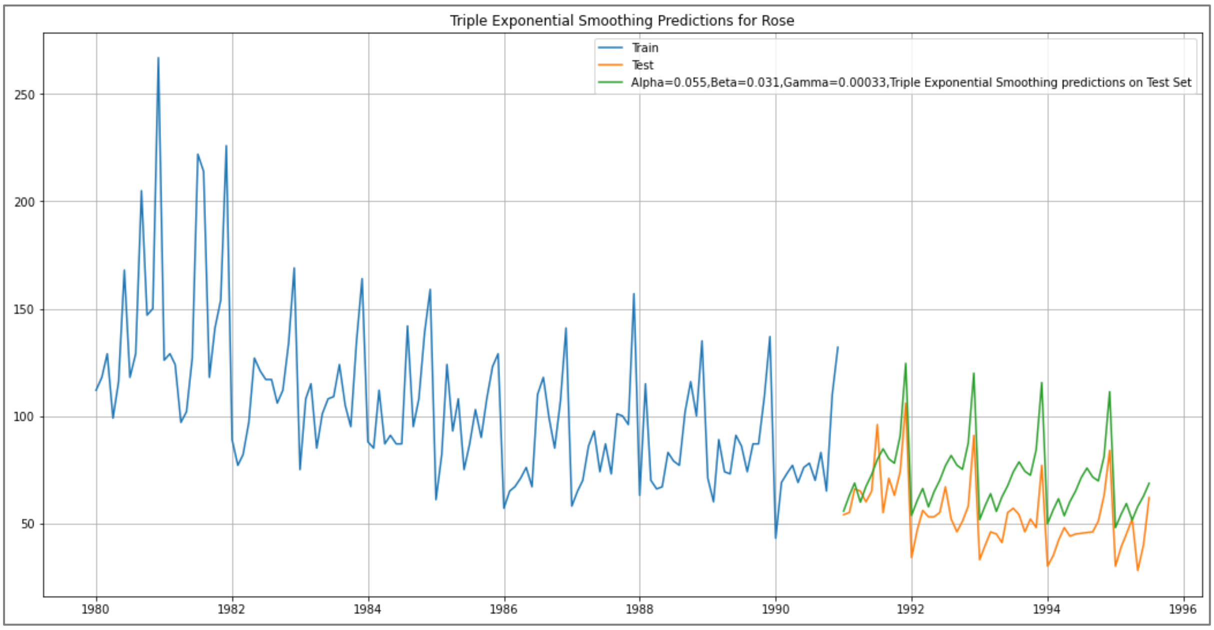

Figure 77. Plot for Rose TES

We check for the **RMSE** score for this model and it comes out to be **19.987**. The predictions are very aligned with the original test data as soon as seasonality is also added in the smoothing mode.

<u>Tuning Triple Exponential Smoothing for Rose (Holt Winter's Model)</u>

We can also tune this model by checking RMSE score for different values of alpha, Beta & Gamma and pick the best fit model which gives us the minimum RMSE.

We run three loops in range from 0.1 to 1.0 with 0.1 difference and consider each value as alpha from one loop and Beta from other loop and Gamma from the third loop. We create a table of different combinations of Alpha, Beta & Gamma and look for the top 5 kept in the descending order of their RMSE score.

	Alpha Values	Beta Values	Gamma Values	Train RMSE	Test RMSE
1080	0.2	0.7	0.2	24.042290	8.702460
161	0.2	0.7	0.2	24.042290	8.702460
215	0.3	0.2	0.6	26.940472	10.206702
1134	0.3	0.2	0.6	26.940472	10.206702
10	0.1	0.2	0.1	19.647823	11.133402

Figure 78. Rose TES Scores

From the above table, it is evident that alpha = 0.2, Beta = 0.7 & Gamma = 0.2 gives the minimum RMSE score of 8.7.

Till now, this model has been the best with the minimum RMSE score. Now we take Alpha = 0.2, Beta = 0.7 & Gamma = 0.2 and forecast the values on test data and plot these predictions on the graph

<u>Plotting the graph for Tuned Triple Exponential Smoothing Predictions for Rose:</u>

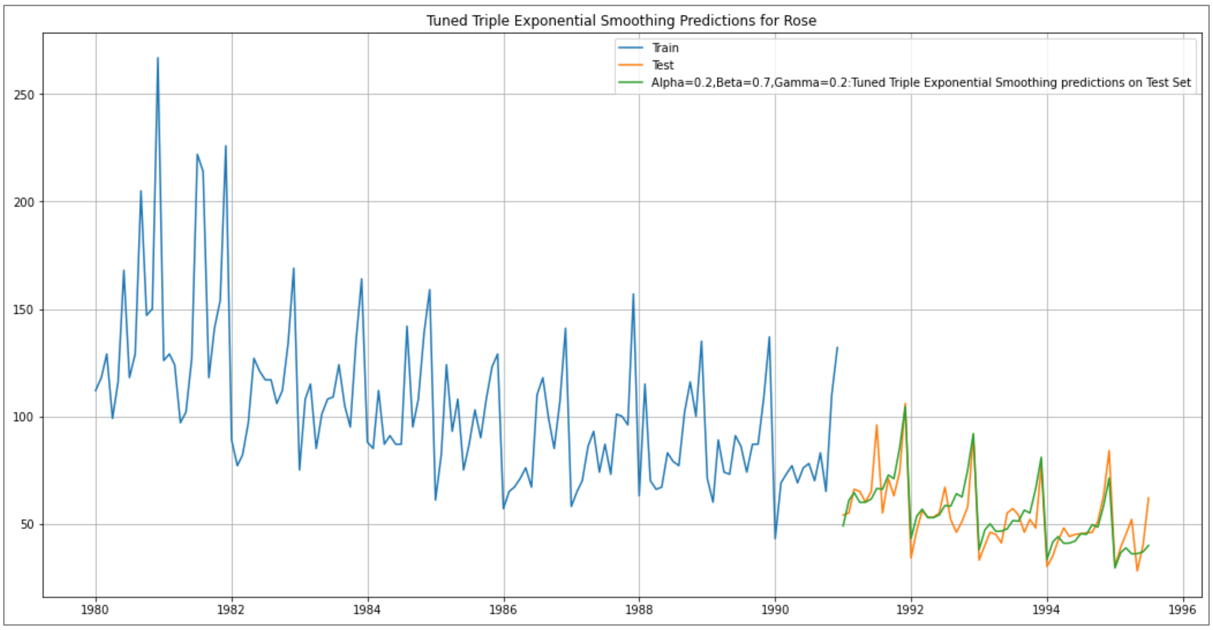

Figure 79. Plot of Tuned TES predictions for Rose.

The predictions have become much better on test model than our original model as we can see from the graph. The **RMSE** for test of this model is **8.702** which is so far the best performance.

We are now done with creating all smoothening models and additional models. Let us compare the performance of each model in a tabular form to check for the lowest RMSE score and thus the best model built.

Sparkling Wine

	Test RMSE-Sparkling
Linear Regression	1389.135175
Naive Approach	3864.279352
Simple Average	1275.081804
2point Trailing Moving Average	813.400684
4point Trailing Moving Average	1156.589694
6point Trailing Moving Average	1283.927428
9point Trailing Moving Average	1346.278315
Alpha=0.049:Simple Exponential Smoothing	1316.035487
Alpha=0.02:Tuned Simple Exponential Smoothing	1279.495201
Alpha=0.66,Beta=0.0001:DoubleExponentialSmoothing	5291.879833
Alpha=0.1,Beta=0.1:Tuned Double Exponential Smoothing	1778.564670
Alpha=0.111,Beta=0.049,Gamma=0.362:Triple Exponential Smoothing	380.398478
Alpha=0.4,Beta=0.01,Gamma=0.3:Tuned Triple Exponential Smoothing	326.579641

Table 3. Sparkling RMSE scores [4th Ques]

From the above table it is evident that Tuned Triple Exponential Smoothening model is the best model so far with the lowest **RMSE score of 326.579**

Rose Wine:

	Test RMSE-Rose
Linear Regression	15.268955
Naive Approach	79.718773
Simple Average	53.460570
2point Trailing Moving Average	11.529278
4point Trailing Moving Average	14.451403
6point Trailing Moving Average	14.566327
9point Trailing Moving Average	14.727630
Alpha= 0.098:Simple Exponential Smoothing	36.796227
Alpha=0.07:Tuned Simple Exponential Smoothing	36.435772
Alpha=1.4901-08,,Beta=1.661-10:Double Exponential Smoothing	15.268944
Alpha=0.1,,Beta=0.1:Tuned Double Exponential Smoothing	36.923416
Alpha=0.055,Beta=0.031,Gamma=0.00033:Triple Exponential Smoothing	19.987449
Alpha=0.2,Beta=0.7,Gamma=0.2:Tuned Triple Exponential Smoothing	8.702460

Table 4. Rose RMSE scores [4th Ques]

From the above table it is evident that Tuned Triple Exponential Smoothening model is the best model so far with the lowest **RMSE score of 8.702**

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.

A common assumption in many time series techniques is that the data are stationary.

A stationary process has the property that the mean, variance and autocorrelation structure do not change over time. Stationarity can be defined in precise mathematical terms, but for our purpose we mean a flat looking series, without trend, constant variance over time, a constant autocorrelation structure over time and no periodic fluctuations (seasonality).

A Time Series with a noticeable trend will not have the same mean over the whole observed time frame. This indicates that the full Time Series is not included in the same sample, necessitating the use of a regression model.

The Augmented Dickey Fuller Test (ADF) is unit root test for stationarity. Unit roots can cause unpredictable results in your time series analysis.

The Augmented Dickey-Fuller test can be used with serial correlation. The ADF test can handle more complex models than the Dickey-Fuller test, and it is also more powerful. That said, it should be used with caution because - like most unit root tests- it has a relatively high Type I error rate.

Hypothesis for Augmented Dickey Fuller (ADF) :

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

Since the null hypothesis assumes the presence of a unit root, the p-value obtained by the test should be less than the significance level (say 0.05) to reject the null hypothesis. Thereby, inferring that the series is stationary.

To perform the ADF test in any time series package, statsmodel provides the implementation function adfuller().

Function adfuller() provides the following information.

- p-value
- Value of the test statistic
- Number of lags for testing consideration
- The critical values

At the chosen alpha value (0.05), we can test for stationarity using the ADF test.

<u>Test for stationarity of the series for Sparkling data:</u>

Checking for stationarity of the whole Time Series data for Sparkling. Firstly, we create a plot having original series along with rolling mean and rolling standard deviation.

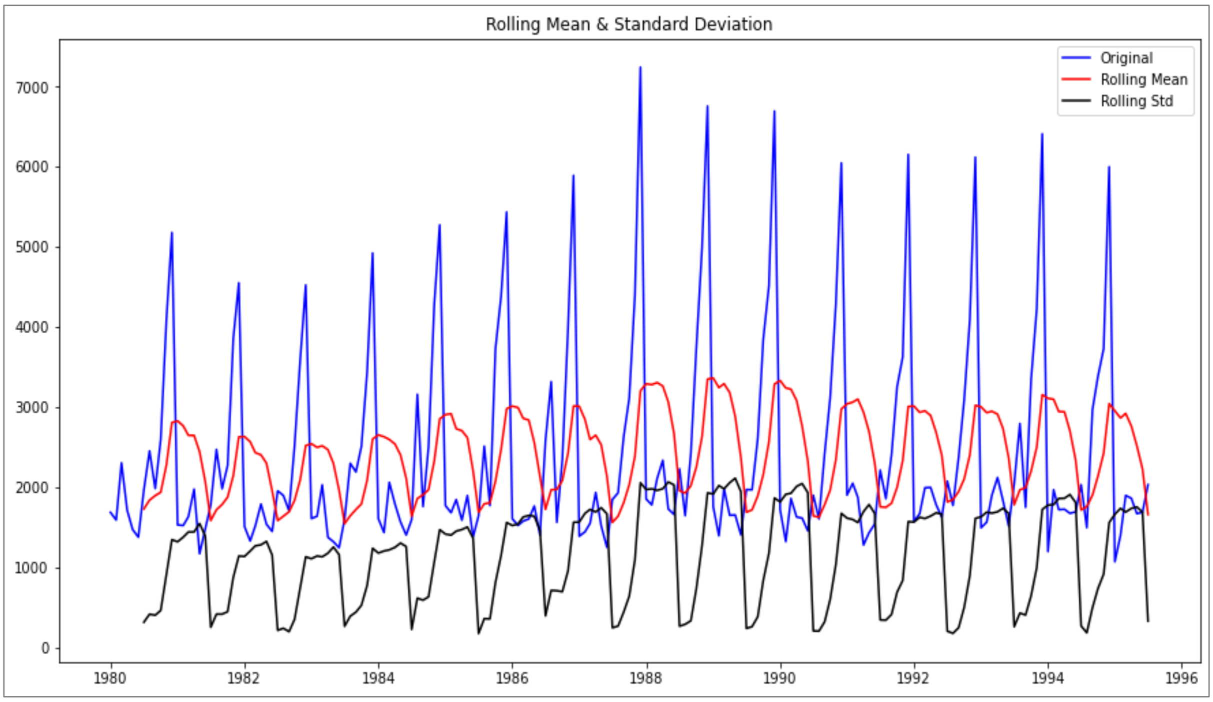

Figure 80. Stationarity check on Sparkling data.

The statistical features of variance and deviation, as seen in the graph, do not remain constant with time in this series. As a result, the data isn't stationary. We can also notice a pattern in the series, indicating that the data is not stagnant.

Crossing checking the same by Augmented Dickey-Fuller test:

```
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
```

Figure 81. Sparkling ADF test.

- Although from the graphs, it looks like the series is stationary but p-value (0.6) is greater than alpha (0.05)
- Therefore, we fail to reject Null hypothesis. The series is non-stationary
- Let us take a difference of order 1 and check whether the Time Series is stationary or not.

Checking the Plot and Stationary of time series with difference of order of one and also cross checking with Augmented Dickey Fuller test.

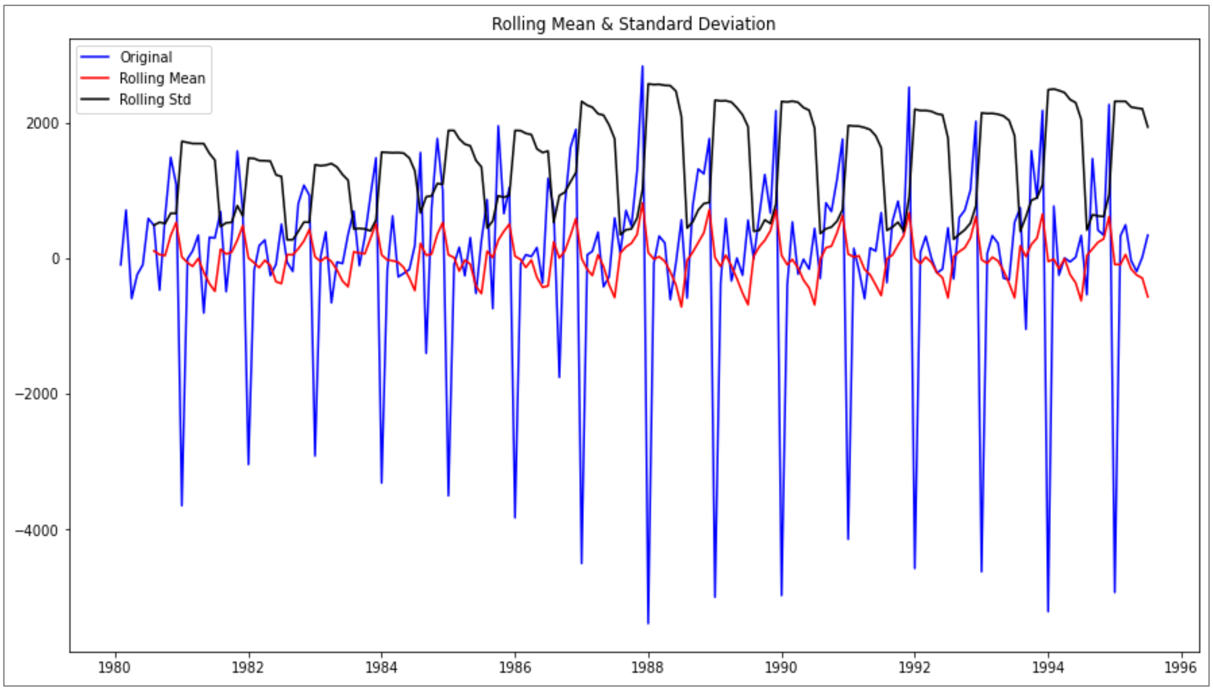

Figure 82. Stationarity check on Sparkling data with difference of order 1.

The statistical parameters of variance and deviation are constant throughout time in this series, as seen in the graph. As a result, the data is stagnant. There is now a constant graph with no trend, indicating that the data is stagnant.

Crossing checking the same by Augmented Dickey-Fuller test:

```
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
```

Figure 83. Sparkling difference 1 ADF test

For this series, the p-value (0.0000) is less than alpha (0.05), therefore we can reject the null hypothesis. **Therefore, the series at difference of order 1 is stationary for Sparkling dataset.**

<u>Test for stationarity of the series for Rose data:</u>

Checking for stationarity of the whole Time Series data for Rose dataset. Firstly, we create a plot having original series along with rolling mean and rolling standard deviation. And also find the results of Augmented Dickey-Fuller test for stationarity.

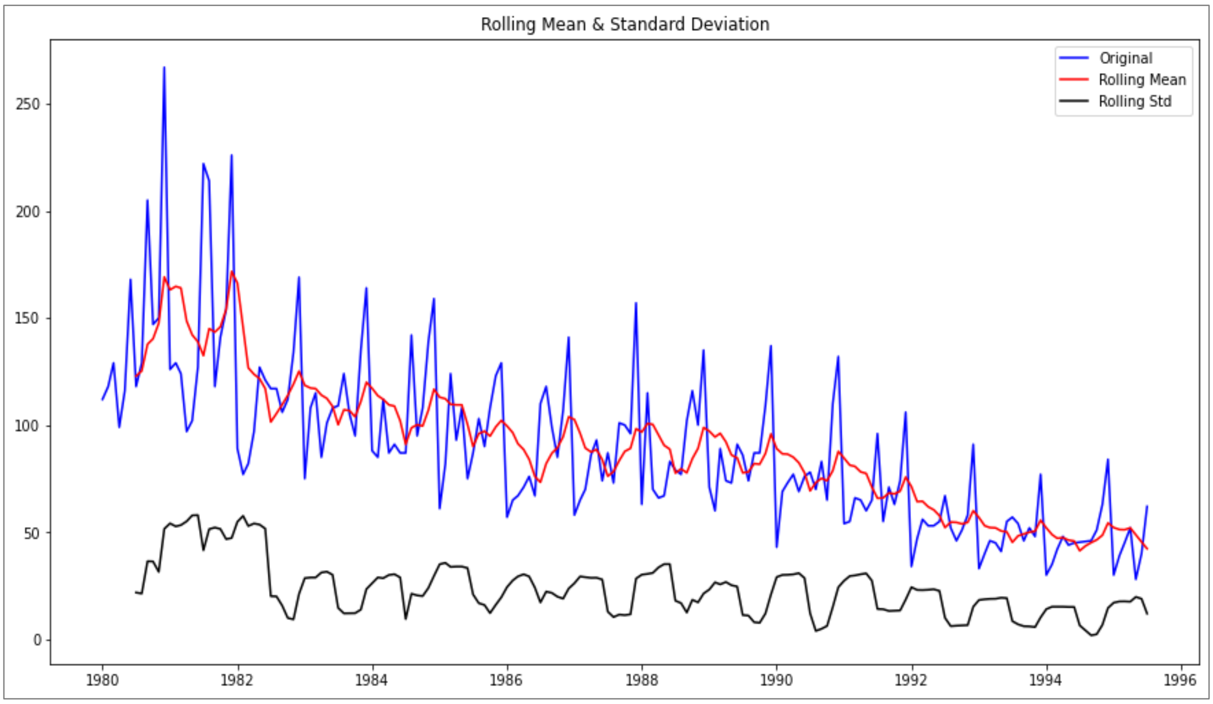

Figure 84. Stationarity check on Rose data

The statistical features of variance and deviation, as seen in the graph, do not remain constant with time in this series. As a result, the data isn't stationary. A distinct decreasing trend is also evident, indicating that the data is not stagnant.

Crossing checking the same by Augmented Dickey-Fuller test:

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

Figure 85. Rose ADF test.

- Since p-value (0.34) is greater than alpha (0.05), we fail to reject the Null Hypothesis.
- Therefore, Rose time series is non-stationary
- Let us take a difference of order 1 and check whether the Time Series is stationary or not.

Checking the Plot and Stationary of time series with difference of order of one and also cross checking with Augmented Dickey Fuller test.

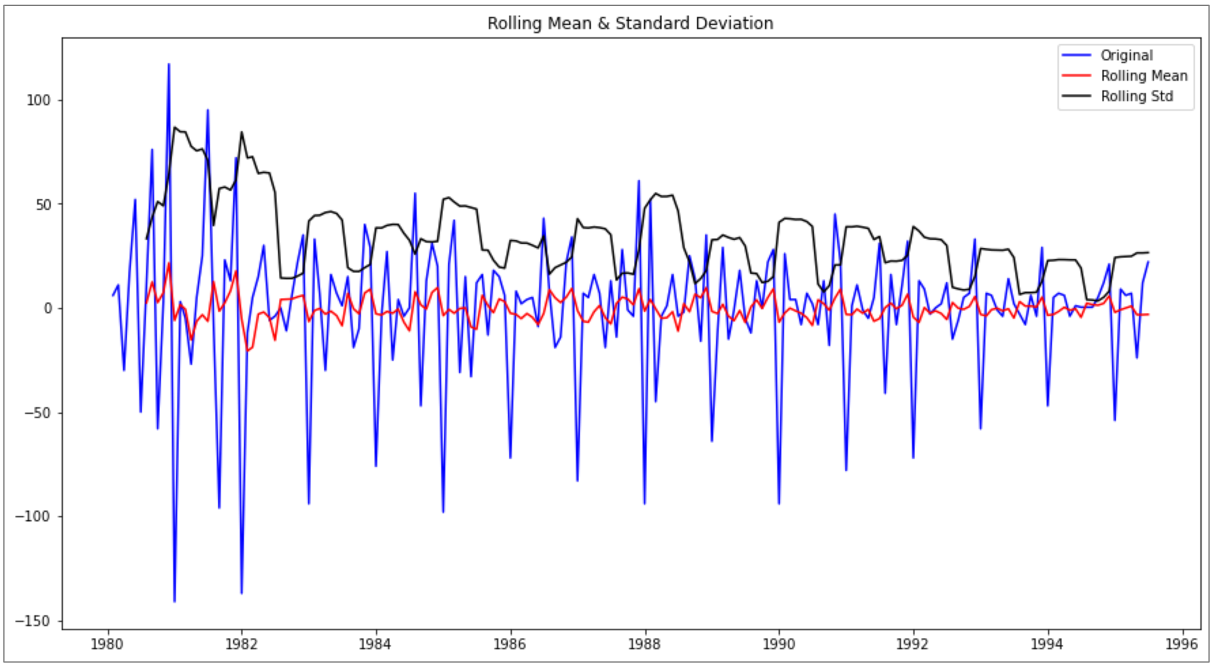

Figure 86. Stationarity check on Rose data with difference of order 1.

The statistical parameters of variance and deviation are constant throughout time in this series, as seen in the graph. As a result, the data is stagnant. A distinct downward trend that was previously visible has now vanished. Now there is a constant graph with no trend, indicating that the data is stationary.

Crossing checking the same by Augmented Dickey-Fuller test:

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

Figure 87. Rose difference 1 ADF test

For this series, the p-value (1.8e-12) is less than alpha (0.05), therefore we can reject the null hypothesis. **Therefore, the series at difference of order 1 is stationary for Rose dataset.**

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.

An autoregressive integrated moving average model is a form of regression analysis that gauges the strength of one dependent variable relative to other changing variables. The model's goal is to predict future securities or financial market moves by examining the differences between values in the series instead of through actual values.

An ARIMA model can be understood by outlining each of its components as follows:

Autoregression (AR): refers to a model that shows a changing variable that regresses on its own lagged, or prior, values.

Integrated (I): represents the differencing of raw observations to allow for the time series to become stationary (i.e., data values are replaced by the difference between the data values and the previous values).

Moving average (MA): incorporates the dependency between an observation and a residual error from a moving average model applied to lagged observations.

Each component in ARIMA functions as a parameter with a standard notation. For ARIMA models, a standard notation would be ARIMA with p, d, and q, where integer values substitute for the parameters to indicate the type of ARIMA model used. The parameters can be defined as:

- ***p*:** the number of lag observations in the model; also known as the lag order.
- ***d*:** the number of times that the raw observations are differenced; also known as the degree of differencing.
- **q**: the size of the moving average window; also known as the order of the moving average

Before starting with the model building, we first plot the ACF and PACF plots.

ACF: The autocorrelation coefficient function, define how the data points in a time series are related to the preceding data points.

PACF: The partial autocorrelation coefficient function, like the autocorrelation function, conveys vital information regarding the dependence structure of a stationary process.

PACF gives the AR component (p) and ACF plot gives the MA component (q).

Automate ARIMA model for Sparkling:

The model will be built using training data, and the values will be forecasted using test data. To do so, we'll first examine the train dataset's stationarity.

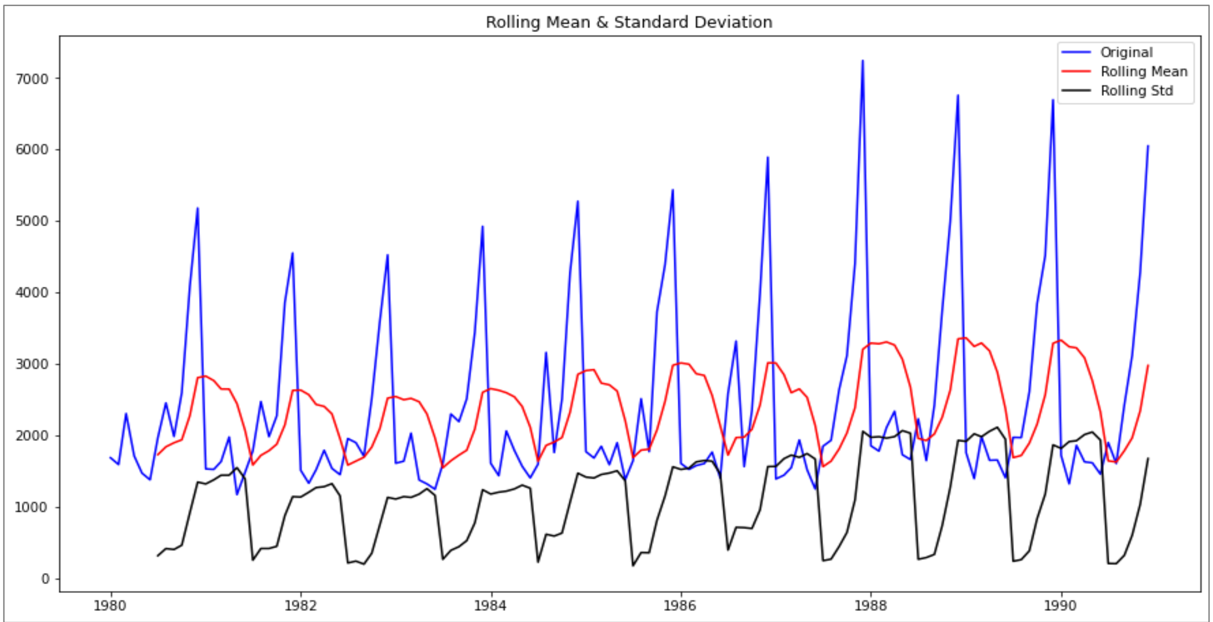

Figure 88. Stationarity check for Sparkling train.

```
Results of Dickey-Fuller Test:
Test Statistic                  -1.208926
p-value                          0.669744
#Lags Used                      12.000000
Number of Observations Used    119.000000
Critical Value (1%)             -3.486535
Critical Value (5%)             -2.886151
Critical Value (10%)            -2.579896
dtype: float64
```

Figure 89. - Sparkling train ADF test

The p value is greater than alpha (0.05), therefore, we fail to reject Null Hypothesis. Therefore, this series is not stationary.

We now know that the series is becoming stationary at differencing level 1. As a result, the 'd' (order of differencing) value is 1.

We'll start by looking at the ACF and PACF plots to gain a sense of the 'p' and 'q' values. These plots will be built on the differenced train data series i.e. Stationary data.

<u>Plotting the Autocorrelation and the Partial Autocorrelation function plots for Sparkling:</u>

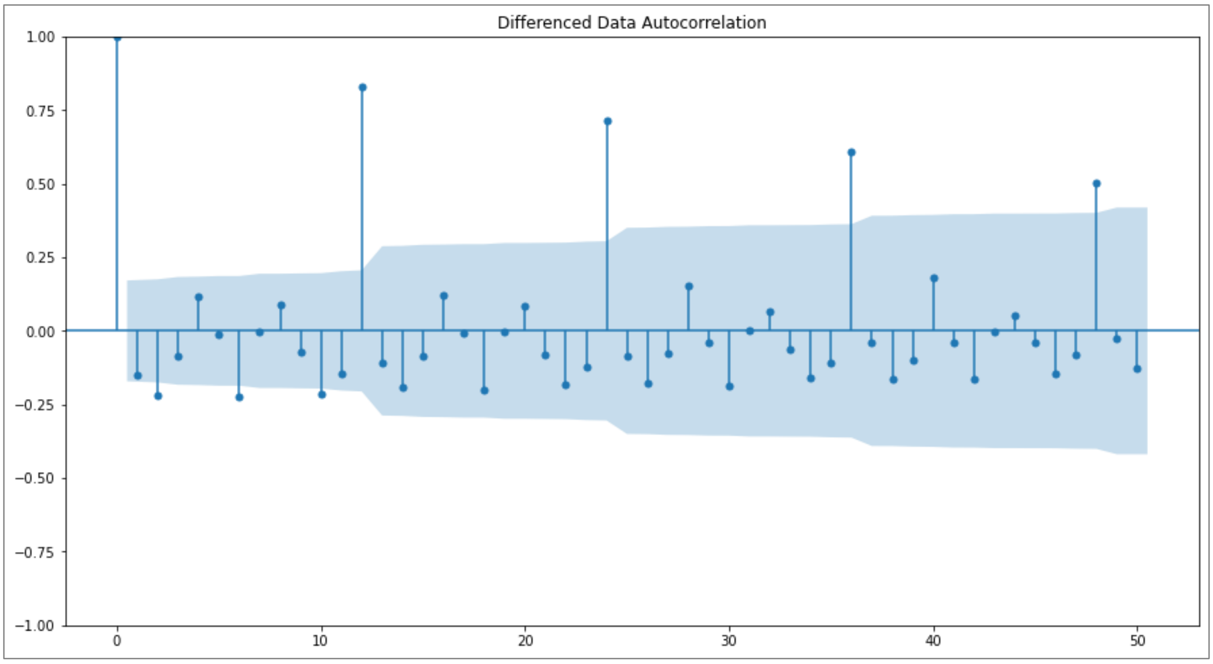

Figure 90. ACF plot of Sparkling

By looking at the above plots, we can say that of ACF, we can say that the Moving-Average parameter in an ARIMA model is 'q' which comes from the significant lag before the ACF plot cuts-off to 0.

Next plotting PACF (Partial autocorrelation coefficient function) on the difference train stationary dataset for sparkling.

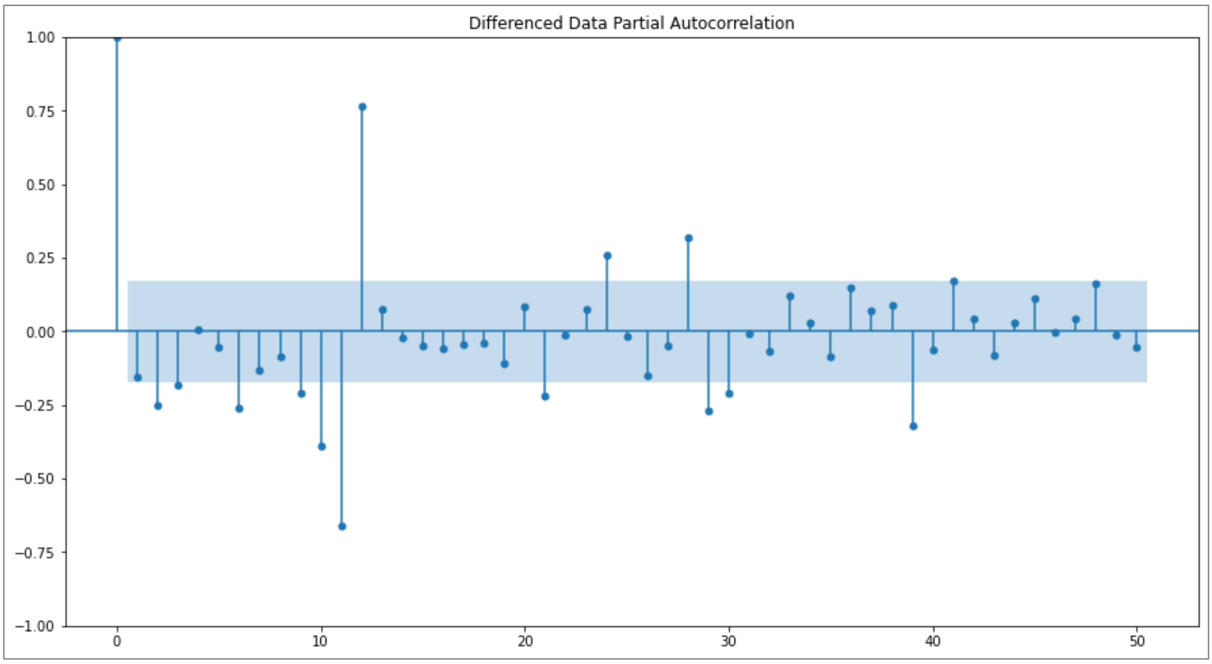

Figure 91. PACF plot for Sparkling.

By looking at the above plots, we can say that of ACF, we can say that the Auto-Regressive parameter in an ARIMA model is 'p' which comes from the significant lag before which the PACF plot cuts-off to 0.

Build an Automated version of an ARIMA model for which the best parameters are selected in accordance with the lowest Akaike Information Criteria (AIC).

Combining different parameters of p and q in the range of 0 and 3 and keeping d as 1 as we already know that at difference of order 1 the time series data is stationary.

Let us take a range from 0 to 2 and use itertools to generate all combinations. The combinations are as below after using itertools products:

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)
```

- Creating an empty Data frame with column names parameters and AIC values for all the combinations.
- Fitting ARIMA model on the train data with order = parameters of all the combinations and running a loop to calculate the AIC values for all.
- The combination with least AIC is considered and summary is checked.

	param	AIC
8	(2, 1, 2)	2213.509212
7	(2, 1, 1)	2233.777626
2	(0, 1, 2)	2234.408323
5	(1, 1, 2)	2234.527200
4	(1, 1, 1)	2235.755095
6	(2, 1, 0)	2260.365744
1	(0, 1, 1)	2263.060016
3	(1, 1, 0)	2266.608539
0	(0, 1, 0)	2267.663036

Figure 92. Sparkling AIC

From the above table we can see that the lowest AIC value is obtained from the combination where **p = 2, d = 1, q = 1.**

Considering this order, the ARIMA model is built on the train data and the summary is checked.

```
                               SARIMAX Results
==============================================================================
Dep. Variable:              Sparkling   No. Observations:                  132
Model:                 ARIMA(2, 1, 2)   Log Likelihood               -1101.755
Date:                Wed, 23 Mar 2022   AIC                           2213.509
Time:                        00:46:21   BIC                           2227.885
Sample:                    01-01-1980   HQIC                          2219.351
                         - 12-01-1990
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          1.3121      0.046     28.781      0.000       1.223       1.401
ar.L2         -0.5593      0.072     -7.741      0.000      -0.701      -0.418
ma.L1         -1.9917      0.109    -18.217      0.000      -2.206      -1.777
ma.L2          0.9999      0.110      9.109      0.000       0.785       1.215
sigma2      1.099e+06   1.99e-07   5.51e+12      0.000     1.1e+06     1.1e+06
===================================================================================
Ljung-Box (L1) (Q):                   0.19   Jarque-Bera (JB):                14.46
Prob(Q):                              0.67   Prob(JB):                         0.00
Heteroskedasticity (H):               2.43   Skew:                             0.61
Prob(H) (two-sided):                  0.00   Kurtosis:                         4.08
===================================================================================
```

Figure 93. Summary of ARIMA AIC criteria Sparkling.

The p value for all parameters is less than alpha (0.05). AIC is 2213.509. All looks good. We will now forecast values for test data set using this model built and check for its performance with the help of RMSE score.

The RMSE score for Sparkling with AIC criteria is 1299.979

<u>Automate ARIMA model for Rose:</u>

The model will be built using training data, and the values will be forecasted using test data. To do so, we'll first examine the train dataset's stationarity.

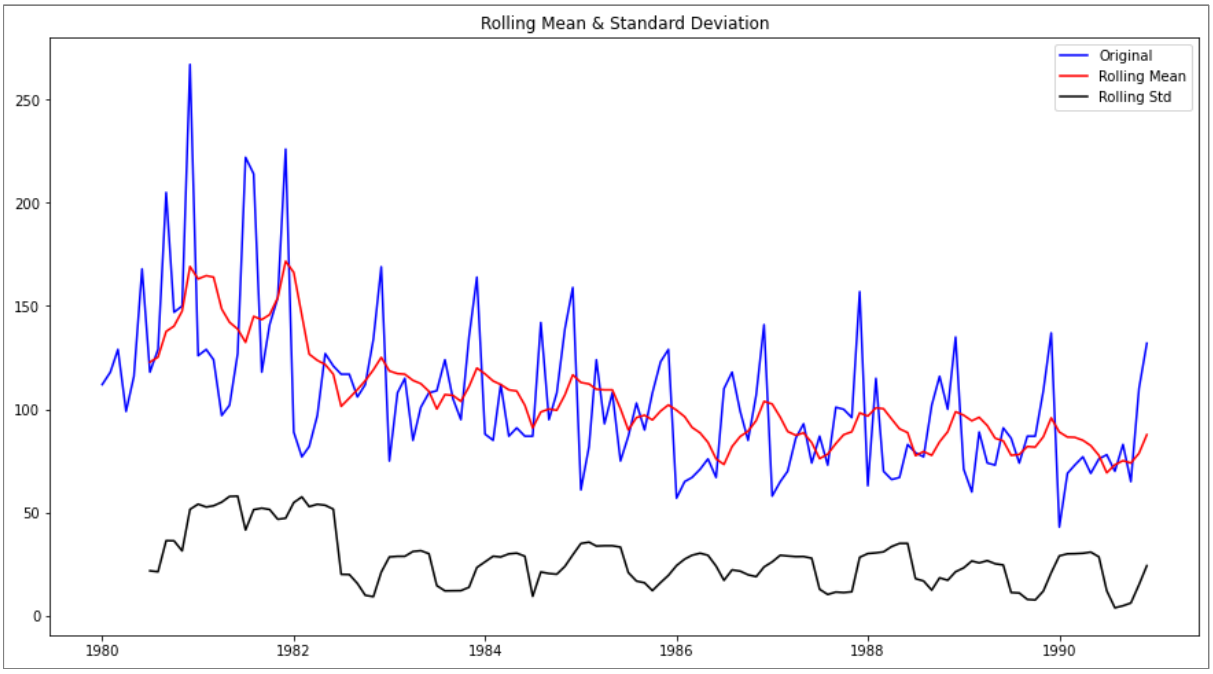

Figure 94. Stationarity check for Rose train.

```
Results of Dickey-Fuller Test:
Test Statistic                  -2.164250
p-value                          0.219476
#Lags Used                      13.000000
Number of Observations Used    118.000000
Critical Value (1%)             -3.487022
Critical Value (5%)             -2.886363
Critical Value (10%)            -2.580009
dtype: float64
```

Figure 95. Rose train ADF test

The p value is greater than alpha (0.05), therefore, we fail to reject Null Hypothesis. Therefore, this series is not stationary.

We now know that the series is becoming stationary at differencing level 1. As a result, the 'd' (order of differencing) value is 1.

To find out the ‘p’ and ‘q’ values, we can look at the ACF and PACF plots and try to figure out the values normally. The other way is to use the itertools package to find out different combinations from a range of values given for ‘p’ and ‘q’. For all combinations, we can check the AIC value and the combination of ‘p’, ‘q’ and ‘d’ which gives the lowest AIC is used to build the ARIMA model.

The Akaike information criterion (AIC) is an estimator of out-of-sample prediction error and thereby relative quality of statistical models for a given set of data. Given a collection of models for the data, AIC estimates the quality of each model, relative to each of the other models. Lower the AIC, better the model.

Now, we can check the ACF and PACF plots first to get an idea of 'p' and 'q' values. These plots will be built on the differenced train data series.

Plotting the Autocorrelation and the Partial Autocorrelation function plots for Rose:

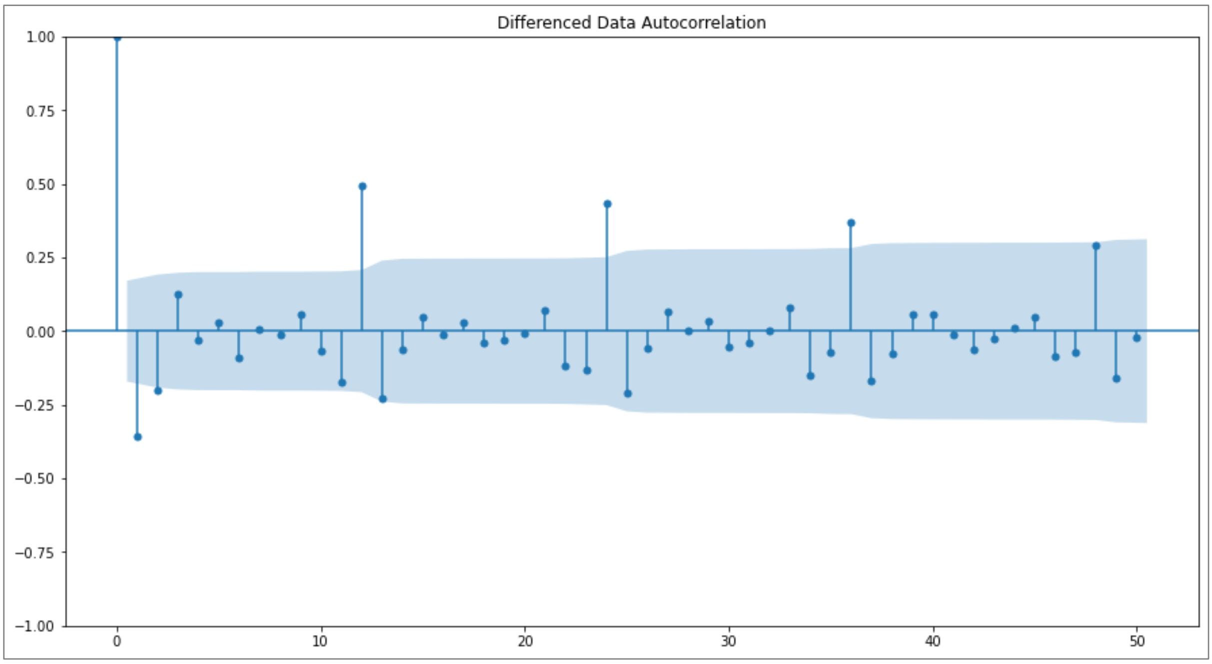

Figure 97. ACF plot of Rose.

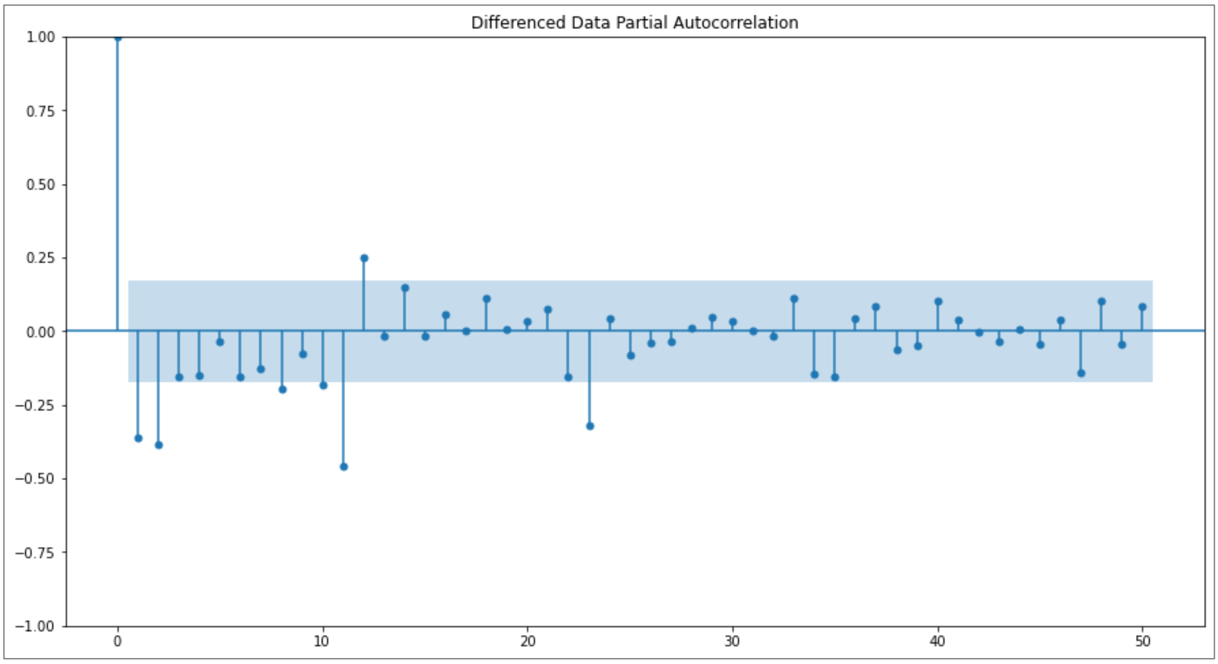

Figure 96. PACF plot of Rose.

From the above plots, we can say that p and q both values could be equal to 2.

Build an Automated version of an ARIMA model for which the best parameters are selected in accordance with the lowest Akaike Information Criteria (AIC).

Combining different parameters of p and q in the range of 0 and 3 and keeping d as 1 as we already know that at difference of order 1 the time series data is stationary.

Let us take a range from 0 to 2 and use itertools to generate all combinations. The combinations are as below after using itertools products:

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)
```

- Creating an empty Data frame with column names parameters and AIC values for all the combinations.
- Fitting ARIMA model on the train data with order = parameters of all the combinations and running a loop to calculate the AIC values for all.
- The combination with least AIC is considered and summary is checked.

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

Figure 98. Rose AIC

From the above table we can see that the lowest AIC value is obtained from the combination where **p = 0, d = 1, q = 2.**

Considering this order, the ARIMA model is built on the train data and the summary is checked.

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                   Rose   No. Observations:                  132
Model:                 ARIMA(0, 1, 2)   Log Likelihood                -636.836
Date:                Wed, 23 Mar 2022   AIC                           1279.672
Time:                        00:55:39   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
===================================================================================
```

Figure 99. Summary of ARIMA AIC criteria Rose.

The p value for all parameters is less than alpha (0.05). AIC is 1279.672. All looks good. We will now forecast values for test data set using this model built and check for its performance with the help of RMSE score.

The RMSE score for Sparkling with AIC criteria is 37.306

SARIMA Model:

Seasonal Autoregressive Integrated Moving Average, SARIMA or Seasonal ARIMA, is an extension of ARIMA that explicitly supports univariate time series data with a seasonal component.

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

Trend Elements

There are three trend elements that require configuration.

They are the same as the ARIMA model; specifically:

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

Seasonal Elements

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

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

SARIMA model for Sparkling:

Building an Automated version of a SARIMA model for which the best parameters are selected in accordance with the lowest Akaike Information Criteria (AIC). Let us look at the ACF plot on train data once more to understand the seasonal parameter for the SARIMA model.

Plotting the Autocorrelation function plots for Sparkling:

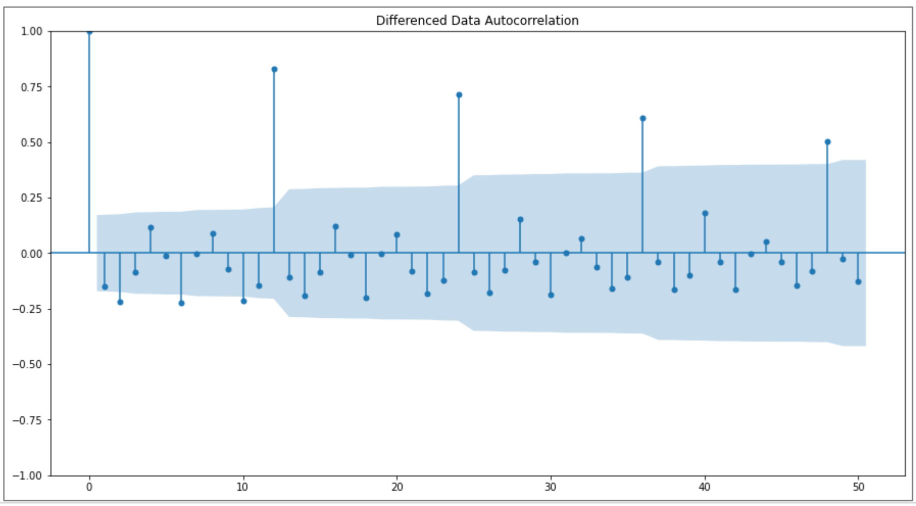

Figure 100. Sparkling SARIMA ACF plot.

Its evident from the plot that there is seasonality of 12, setting the seasonality of 12 in the itertools loop to find out different combinations from range of values for ‘p’ and ‘q’, ‘P’ and ‘Q’. For all these combinations, we will check the AIC values for which lowest AIC parameters are considered to build SARIMA model.

We already have checked the ‘p’, ‘q’ and ‘d’ values manually while creating ARIMA model. p = 0, q=0, d=1. So, we will take a range from 0 to 2 while using itertools. The value of D, we will be taking as 0 as we have already considered differencing with d=1. For ‘P’ and ‘Q’ values, we will take a range from 0 to 2 and proceed with the itertools. product function to get the best parameters.

Some parameter combinations for Model

```
Model: (0, 1, 1)(0, 0, 1, 12)
Model: (0, 1, 2)(0, 0, 2, 12)
Model: (1, 1, 0)(1, 0, 0, 12)
Model: (1, 1, 1)(1, 0, 1, 12)
Model: (1, 1, 2)(1, 0, 2, 12)
Model: (2, 1, 0)(2, 0, 0, 12)
Model: (2, 1, 1)(2, 0, 1, 12)
Model: (2, 1, 2)(2, 0, 2, 12)
```

- Creating an empty Data frame with column names parameters and AIC values for all the combinations.
- Fitting SARIMA model on the train data with order = parameters of all the combinations and running a loop to calculate the AIC values for all.
- The combination with least AIC is considered and summary is checked.

	param	seasonal	AIC
50	(1, 1, 2)	(1, 0, 2, 12)	1555.584248
53	(1, 1, 2)	(2, 0, 2, 12)	1556.076790
26	(0, 1, 2)	(2, 0, 2, 12)	1557.121579
23	(0, 1, 2)	(1, 0, 2, 12)	1557.160507
77	(2, 1, 2)	(1, 0, 2, 12)	1557.340402

Figure 101.Sparkling SARIMA AIC

From the above table we can see that the lowest AIC value is obtained from the combination where **p = 1, d = 1, q = 2, P = 1, D = 0, Q = 2** with seasonality of **12**.

Considering this order, the SARIMA model is built on the train data and the summary is checked with the best parameters with least AIC score.

```
                                      SARIMAX Results
==========================================================================================
Dep. Variable:                                 y   No. Observations:                  132
Model:             SARIMAX(1, 1, 2)x(1, 0, 2, 12)   Log Likelihood                -770.792
Date:                            Wed, 23 Mar 2022   AIC                           1555.584
Time:                                    00:51:14   BIC                           1574.095
Sample:                                         0   HQIC                          1563.083
                                            - 132
Covariance Type:                              opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.6283      0.255     -2.464      0.014      -1.128      -0.128
ma.L1         -0.1040      0.225     -0.463      0.644      -0.545       0.337
ma.L2         -0.7277      0.154     -4.736      0.000      -1.029      -0.427
ar.S.L12       1.0439      0.014     72.834      0.000       1.016       1.072
ma.S.L12      -0.5550      0.098     -5.663      0.000      -0.747      -0.363
ma.S.L24      -0.1354      0.120     -1.133      0.257      -0.370       0.099
sigma2      1.506e+05   2.03e+04      7.401      0.000    1.11e+05     1.9e+05
===================================================================================
Ljung-Box (L1) (Q):                   0.04   Jarque-Bera (JB):                11.72
Prob(Q):                              0.84   Prob(JB):                         0.00
Heteroskedasticity (H):               1.47   Skew:                             0.36
Prob(H) (two-sided):                  0.26   Kurtosis:                         4.48
===================================================================================
```

Figure 102. Summary of SARIMA AIC criteria Sparkling.

.AIC is 1555.584. The model looks good.

<u>Checking a diagnostic on the residual:</u>

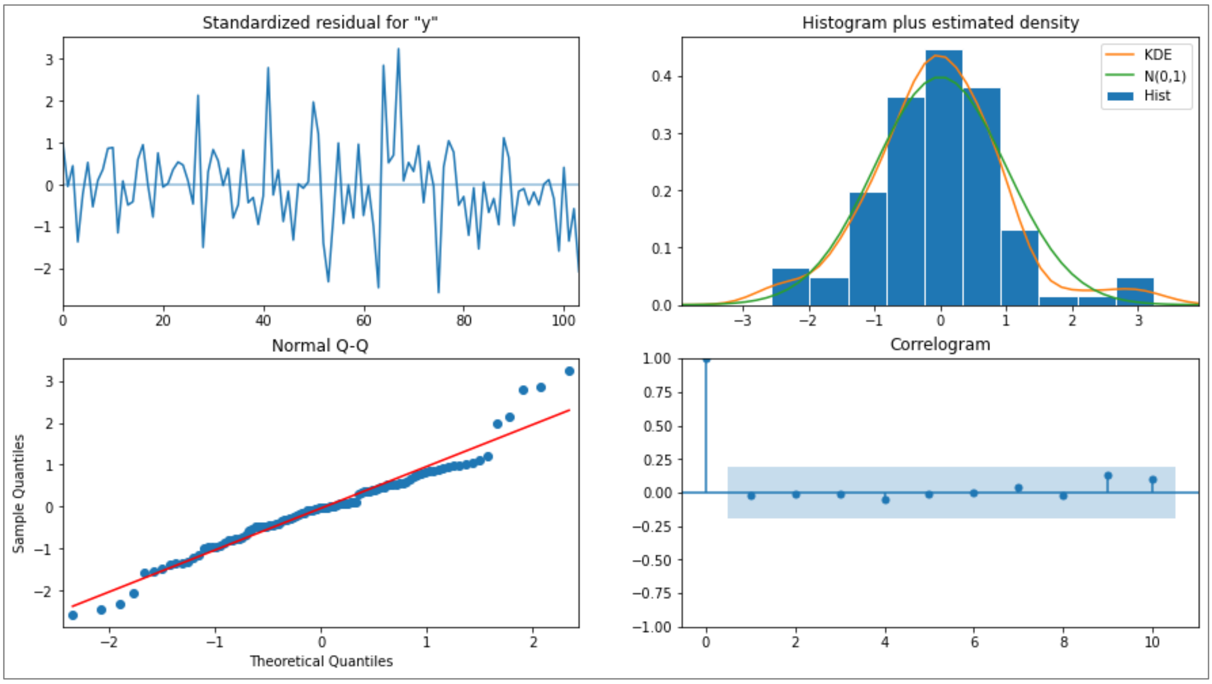

Figure 103. Sparkling residual diagnostic

We will now forecast values for test data set using this model built and check for its performance with the help of RMSE score.

The RMSE score for Sparkling with AIC criteria is 528.611.

SARIMA model for Rose:

Building an Automated version of a SARIMA model for which the best parameters are selected in accordance with the lowest Akaike Information Criteria (AIC). Let us look at the ACF plot on train data once more to understand the seasonal parameter for the SARIMA model.

Plotting the Autocorrelation function plots for Sparkling:

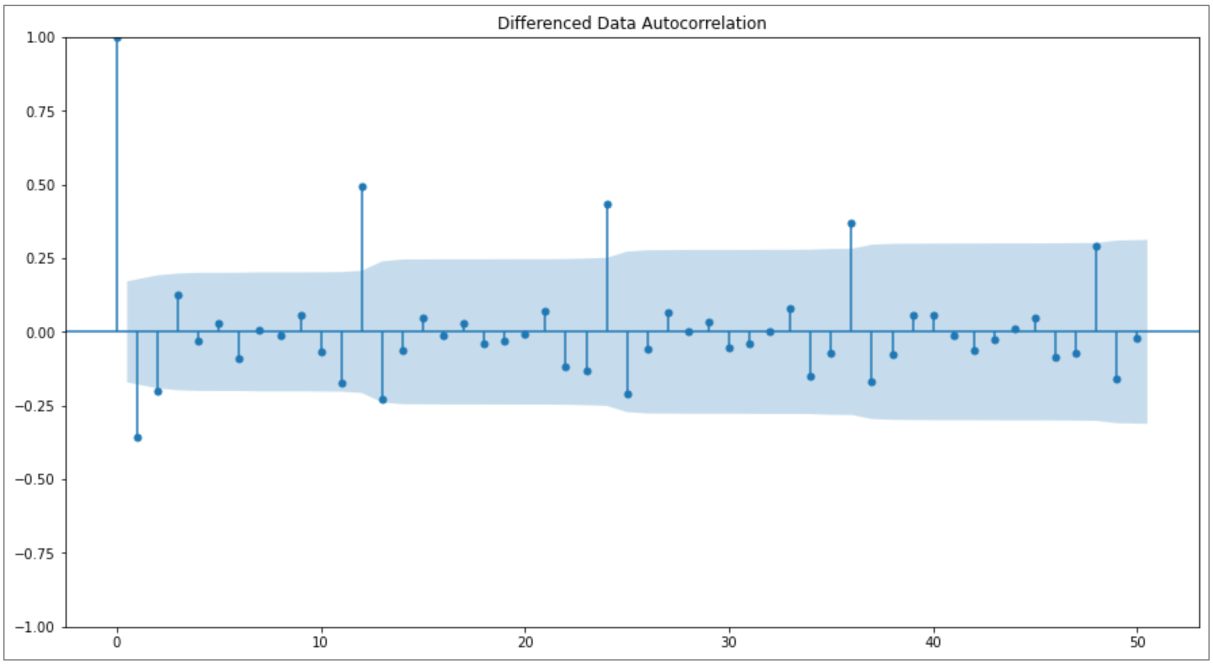

Figure 104. Rose SARIMA ACF plot

Its evident from the plot that there is seasonality of 12, setting the seasonality of 12 in the itertools loop to find out different combinations from range of values for ‘p’ and ‘q’, ‘P’ and ‘Q’. For all these combinations, we will check the AIC values for which lowest AIC parameters are considered to build SARIMA model.

We already have checked the ‘p’, ‘q’ and ‘d’ values manually while creating ARIMA model. p = 0, q=0, d=1. So, we will take a range from 0 to 2 while using itertools. The value of D, we will be taking as 0 as we have already considered differencing with d=1. For ‘P’ and ‘Q’ values, we will take a range from 0 to 2 and proceed with the itertools. product function to get the best parameters.

Some parameter combinations for Model

```
Model: (0, 1, 1)(0, 0, 1, 12)
Model: (0, 1, 2)(0, 0, 2, 12)
Model: (1, 1, 0)(1, 0, 0, 12)
Model: (1, 1, 1)(1, 0, 1, 12)
Model: (1, 1, 2)(1, 0, 2, 12)
Model: (2, 1, 0)(2, 0, 0, 12)
Model: (2, 1, 1)(2, 0, 1, 12)
Model: (2, 1, 2)(2, 0, 2, 12)
```

- Creating an empty Data frame with column names parameters and AIC values for all the combinations.
- Fitting SARIMA model on the train data with order = parameters of all the combinations and running a loop to calculate the AIC values for all.

- The combination with least AIC is considered and summary is checked

	param	seasonal	AIC
26	(0, 1, 2)	(2, 0, 2, 12)	887.937509
53	(1, 1, 2)	(2, 0, 2, 12)	889.903048
80	(2, 1, 2)	(2, 0, 2, 12)	890.668798
69	(2, 1, 1)	(2, 0, 0, 12)	896.518161
78	(2, 1, 2)	(2, 0, 0, 12)	897.346444

Figure 105. Rose SARIMA AIC.

From the above table we can see that the lowest AIC value is obtained from the combination where **p = 0, d = 1, q = 2, P = 2, D = 0, Q = 2** with seasonality of **12**.

Considering this order, the SARIMA model is built on the train data and the summary is checked with the best parameters with least AIC score.

```
                                      SARIMAX Results
==========================================================================================
Dep. Variable:                                  y   No. Observations:                  132
Model:             SARIMAX(0, 1, 2)x(2, 0, 2, 12)   Log Likelihood                -436.969
Date:                            Wed, 23 Mar 2022   AIC                            887.938
Time:                                    00:59:41   BIC                            906.448
Sample:                                         0   HQIC                           895.437
                                            - 132
Covariance Type:                              opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ma.L1         -0.8427    189.943     -0.004      0.996    -373.124     371.439
ma.L2         -0.1573     29.841     -0.005      0.996     -58.645      58.330
ar.S.L12       0.3467      0.079      4.375      0.000       0.191       0.502
ar.S.L24       0.3023      0.076      3.996      0.000       0.154       0.451
ma.S.L12       0.0767      0.133      0.577      0.564      -0.184       0.337
ma.S.L24      -0.0726      0.146     -0.498      0.618      -0.358       0.213
sigma2       251.3137   4.77e+04      0.005      0.996   -9.33e+04    9.38e+04
===================================================================================
Ljung-Box (L1) (Q):                   0.10   Jarque-Bera (JB):                 2.33
Prob(Q):                              0.75   Prob(JB):                         0.31
Heteroskedasticity (H):               0.88   Skew:                             0.37
Prob(H) (two-sided):                  0.70   Kurtosis:                         3.03
===================================================================================
```

Figure 106. Summary of SARIMA AIC criteria Rose.

AIC is 887.938. Jarque-Bura p value is greater than 0.05. We have coefficients for 7 parameters. All looks good.

<u>Checking a diagnostic on the residual:</u>

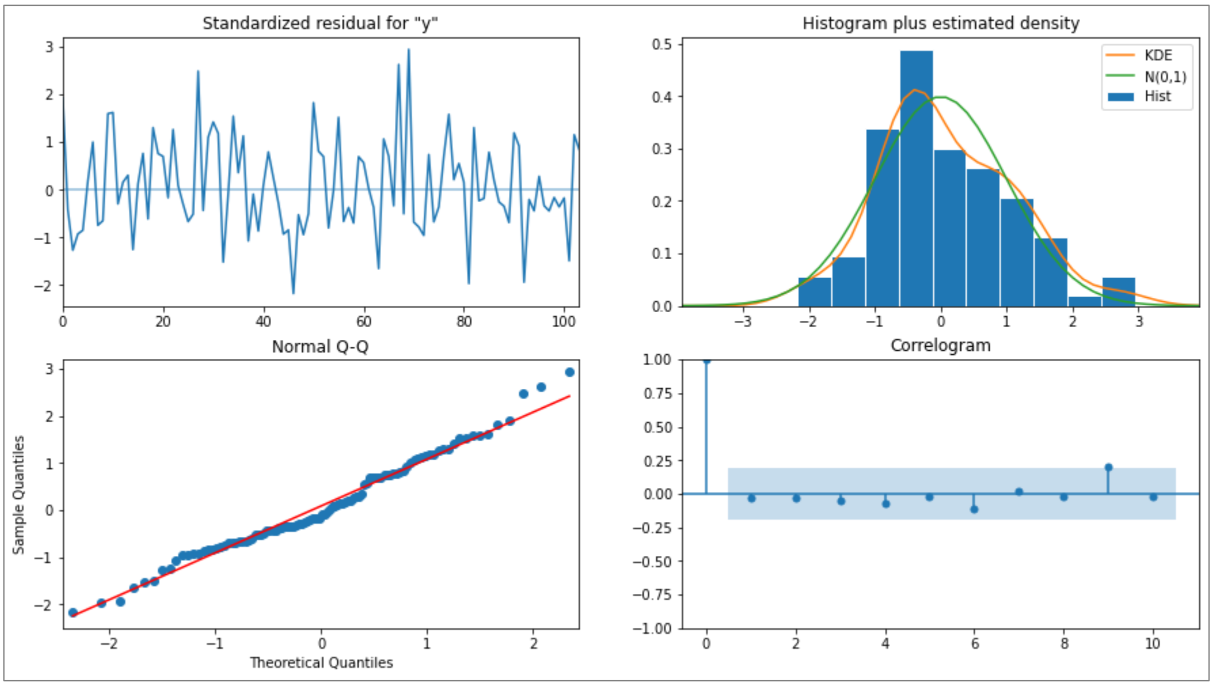

Figure 107. Rose residual diagnostic

We will now forecast values for test data set using this model built and check for its performance with the help of RMSE score.

The RMSE score for Sparkling with AIC criteria is 26.928.

The ARIMA and SARIMA models based on AIC criteria are now complete. Let's look at the RMSE scores of all the models we've built so far and compare them to find the best fit model.

Sparkling Wine RMSE Scores:

	Test RMSE-Sparkling
Linear Regression	1389.135175
Naive Approach	3864.279352
Simple Average	1275.081804
2point Trailing Moving Average	813.400684
4point Trailing Moving Average	1156.589694
6point Trailing Moving Average	1283.927428
9point Trailing Moving Average	1346.278315
Alpha=0.049:Simple Exponential Smoothing	1316.035487
Alpha=0.02:Tuned Simple Exponential Smoothing	1279.495201
Alpha=0.66,Beta=0.0001:DoubleExponentialSmoothing	5291.879833
Alpha=0.1,Beta=0.1:Tuned Double Exponential Smoothing	1778.564670
Alpha=0.111,Beta=0.049,Gamma=0.362:Triple Exponential Smoothing	380.398478
Alpha=0.4,Beta=0.01,Gamma=0.3:Tuned Triple Exponential Smoothing	326.579641
ARIMA(2,1,2) AIC criteria	1299.979569
SARIMA(1,1,2)(1,0,2,12) AIC criteria	528.611364

Table 5. Sparkling RMSE compare [Ques 6]

Rose Wine RMSE Scores:

	Test RMSE-Rose
Linear Regression	15.268955
Naive Approach	79.718773
Simple Average	53.460570
2point Trailing Moving Average	11.529278
4point Trailing Moving Average	14.451403
6point Trailing Moving Average	14.566327
9point Trailing Moving Average	14.727630
Alpha= 0.098:Simple Exponential Smoothing	36.796227
Alpha=0.07:Tuned Simple Exponential Smoothing	36.435772
Alpha=1.4901-08,,Beta=1.661-10:Double Exponential Smoothing	15.268944
Alpha=0.1,,Beta=0.1:Tuned Double Exponential Smoothing	36.923416
Alpha=0.055,Beta=0.031,Gamma=0.00033:Triple Exponential Smoothing	19.987449
Alpha=0.2,Beta=0.7,Gamma=0.2:Tuned Triple Exponential Smoothing	8.702460
ARIMA(0,1,2) AIC criteria	15.619203
SARIMA(0,1,2)(2,0,2,12) AIC criteria	26.928362

Table 6. Rose RMSE compare [Ques 6]

We can see that, for both Sparkling and Rose wine data set, still Tuned Triple Exponential Smoothing model is thus far the best fit.

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.

We'll now build ARIMA and SARIMA models based on the ACF and PACF plots of training data's cut-off points, and then forecast on test data. Finally, we'll use RMSE values to assess the model's performance.

ARIMA model for Sparkling:

Start with building the ACF and PACF plots for train data with difference 1 (to make the dataset stationary). We will check for the values of p and q by looking at the plots and where the cut-off points are located in the plot.

We can select the order **p** for AR model based on significant spikes from the PACF plot. One more indication of the AR process is that the ACF plot decays more slowly.

In contrast to the AR model, we can select the order **q** for model MA from ACF if this plot has a sharp cut-off after lag.

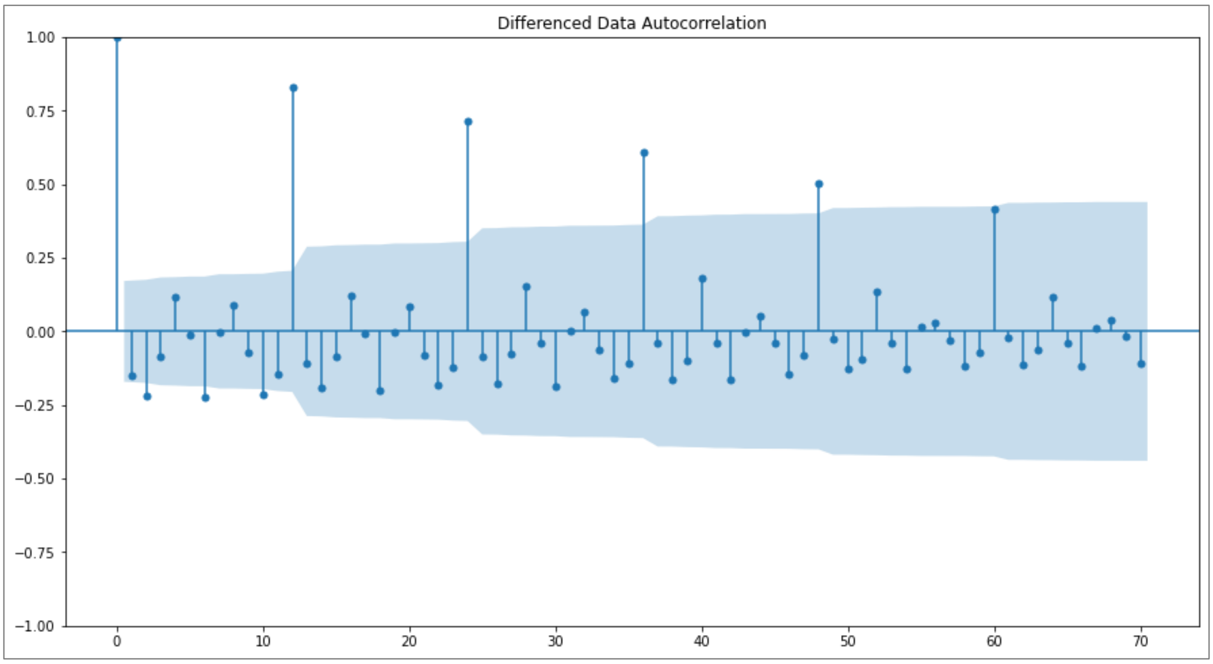

Figure 108. Sparkling ACF

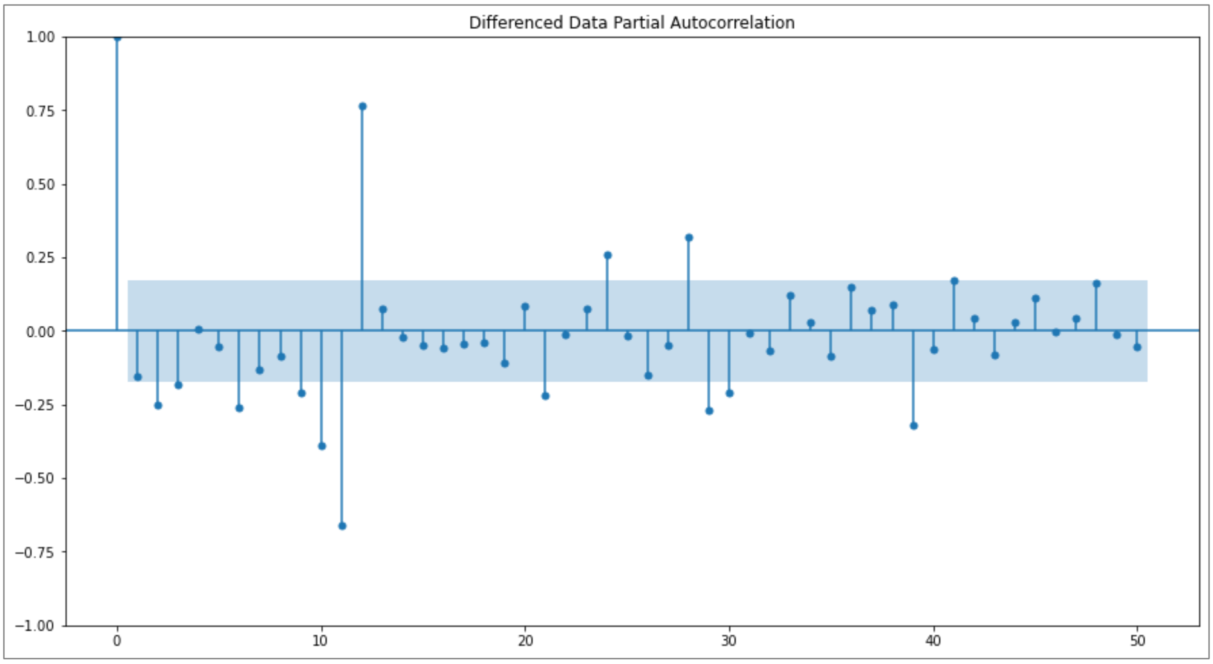

Figure 109. Sparkling PACF.

To find the value of **p,** we look into the **Partial Autocorrelation plot**. We see that the cut-off point is the first point. Therefore, we will take p where the last point was not yet cut by confidence interval. Hence, **p = 0**

To find the value of **q**, we look into the **Autocorrelation plot**. We see that first point cuts the confidence interval, therefore**, q = 0.**

So, our order would be, **p = 0, d = 1 and q = 0.** We take this order and build an ARIMA model on the training data.

Checking for summary of manual ARIMA model:

```
                               SARIMAX Results
==============================================================================
Dep. Variable:              Sparkling   No. Observations:                  132
Model:                 ARIMA(0, 1, 0)   Log Likelihood               -1132.832
Date:                Wed, 23 Mar 2022   AIC                           2267.663
Time:                        03:12:27   BIC                           2270.538
Sample:                    01-01-1980   HQIC                          2268.831
                         - 12-01-1990
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
sigma2      1.885e+06   1.29e+05     14.658      0.000    1.63e+06    2.14e+06
===================================================================================
Ljung-Box (L1) (Q):                   3.07   Jarque-Bera (JB):               198.83
Prob(Q):                              0.08   Prob(JB):                         0.00
Heteroskedasticity (H):               2.46   Skew:                            -1.92
Prob(H) (two-sided):                  0.00   Kurtosis:                         7.65
===================================================================================
```

Figure 110. Summary of Manual ARIMA for Sparkling.

AIC for this model is 2267.663. Overall, the model looks good.

We will now forecast values for test data set using this model built and check for its performance with the help of RMSE score. With this ARIMA model.

The RMSE score on test data set is 3864.279.

ARIMA model for Rose:

Start with building the ACF and PACF plots for train data with difference 1 (to make the dataset stationary). We will check for the values of p and q by looking at the plots and where the cut-off points are located in the plot.

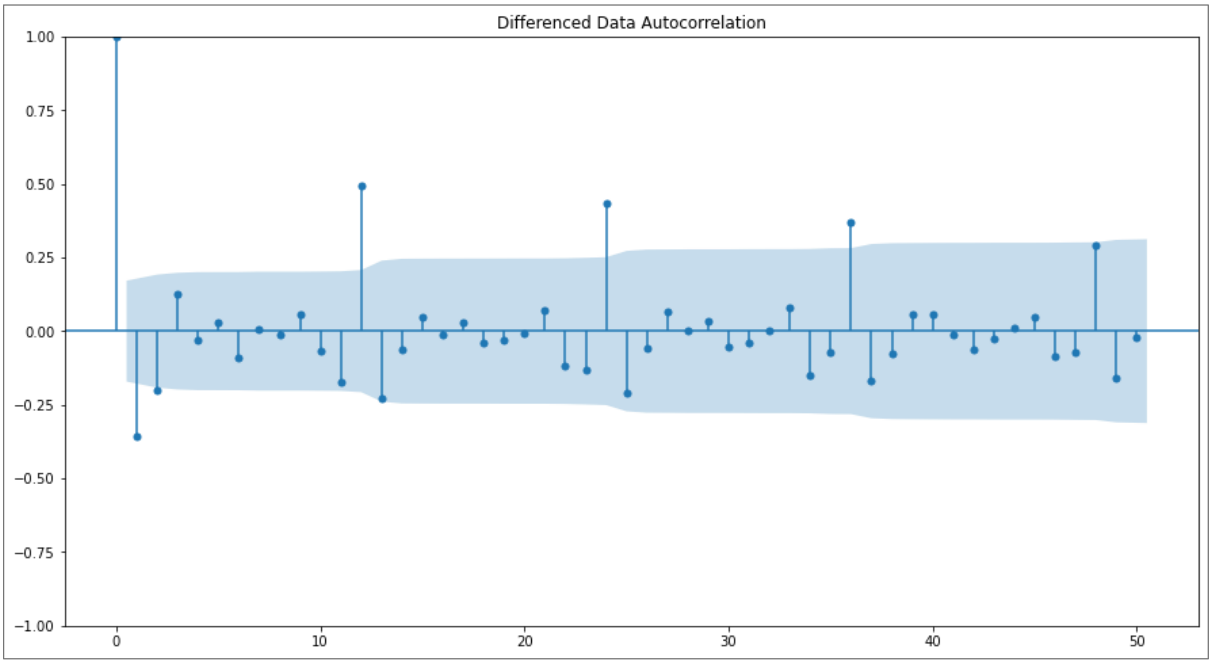

Figure 111. Rose ACF

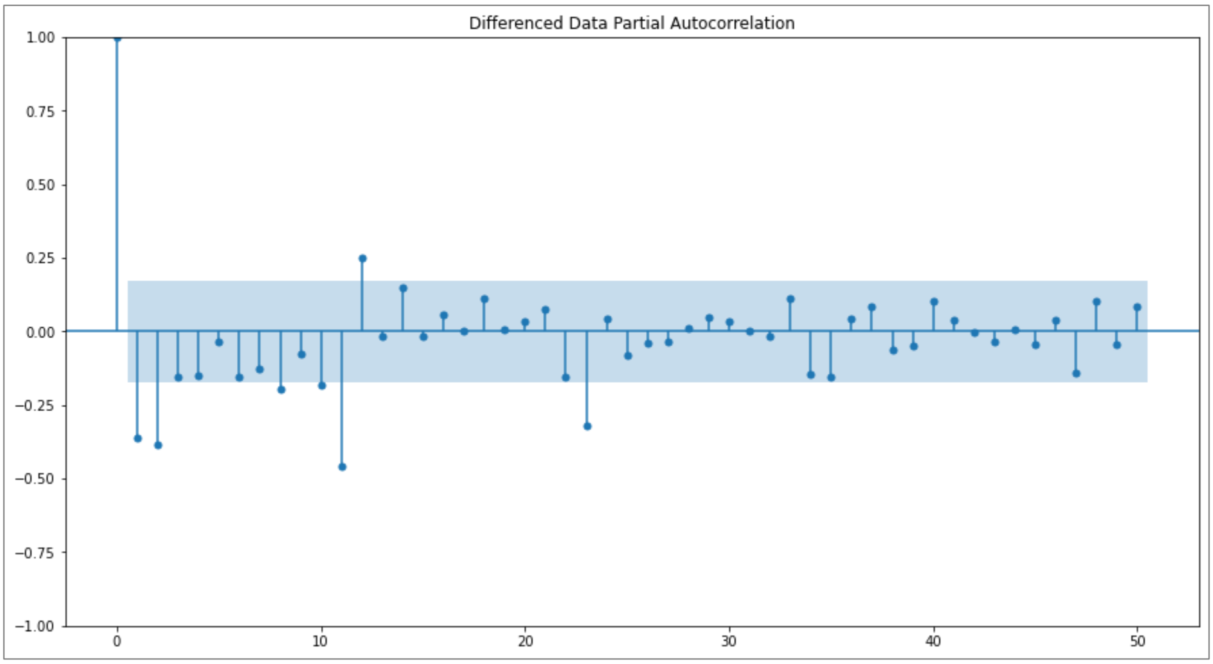

Figure 112. Sparkling PACF.

To find the value of **p**, we look into the **Partial Autocorrelation plot**. We see that the cut-off point is the third point. Therefore, we will take p where the last point was not yet cut by confidence interval. Hence, **p = 2**

To find the value of **q**, we look into the Autocorrelation plot. We see that third point cuts the confidence interval, Hence, **q = 2**

So, our order would be, **p = 2, d = 1 and q = 2.** We take this order and create an ARIMA model on the training data.

Checking the summary for manual ARIMA for Rose:

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                   Rose   No. Observations:                  132
Model:                 ARIMA(2, 1, 2)   Log Likelihood                -635.935
Date:                Wed, 23 Mar 2022   AIC                           1281.871
Time:                        03:15:55   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
===================================================================================
```

Figure 113. Summary of Manual ARIMA for Rose

AIC is 1281.247. The model looks good with all the coefficients and p values.

We will now forecast values for test data set using this model built and check for its performance with the help of RMSE score. With this ARIMA model.

The RMSE score on test data set is 36.871.

SARIMA Models:

We'll now build SARIMA models based on the ACF and PACF plots of training data's cut-off points, and then forecast on test data. Finally, we'll use RMSE values to assess the model's performance.

SARIMA model for Sparkling:

We have already seen from the ACF plot that the seasonality for Sparkling wine train dataset after one difference is 12. So, we will be taking Seasonality = 12.

p and q values from ACF and PACF plots are 0 and 0.

D = 1 (Differencing of 1) and finally P and Q values from the ACF and PACF plots are 2 and 4 respectively.

So, the final order will be **(0, 1, 0) (2, 1, 4, 12)** We take this order and build a SARIMA model on the training data.

Checking the summary for Sparkling for manual SARIMA model:

```
                                       SARIMAX Results
==========================================================================================================
Dep. Variable:                                         y   No. Observations:                        132
Model:             SARIMAX(0, 1, 0)x(2, 1, [1, 2, 3, 4], 12)   Log Likelihood                      -538.663
Date:                                   Wed, 23 Mar 2022   AIC                                  1091.326
Time:                                           03:14:08   BIC                                  1107.066
Sample:                                                0   HQIC                                 1097.578
                                                   - 132
Covariance Type:                                     opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.S.L12      -0.5734      0.253     -2.266      0.023      -1.070      -0.077
ar.S.L24      -0.5548      0.108     -5.147      0.000      -0.766      -0.344
ma.S.L12       0.3449      0.391      0.882      0.378      -0.422       1.111
ma.S.L24       0.5798      0.191      3.040      0.002       0.206       0.954
ma.S.L36      -0.5033      0.117     -4.306      0.000      -0.732      -0.274
ma.S.L48      -0.0809      0.349     -0.232      0.816      -0.764       0.602
sigma2      2.044e+05   1.02e-06      2e+11      0.000    2.04e+05    2.04e+05
===================================================================================
Ljung-Box (L1) (Q):                   7.81   Jarque-Bera (JB):                32.02
Prob(Q):                              0.01   Prob(JB):                         0.00
Heteroskedasticity (H):               0.32   Skew:                             0.95
Prob(H) (two-sided):                  0.01   Kurtosis:                         5.72
===================================================================================
```

Figure 114. Summary of manual SARIMA for Sparkling.

AIC for this model is 1091.326. Overall, the model looks good.

Checking a diagnostic on the residual:

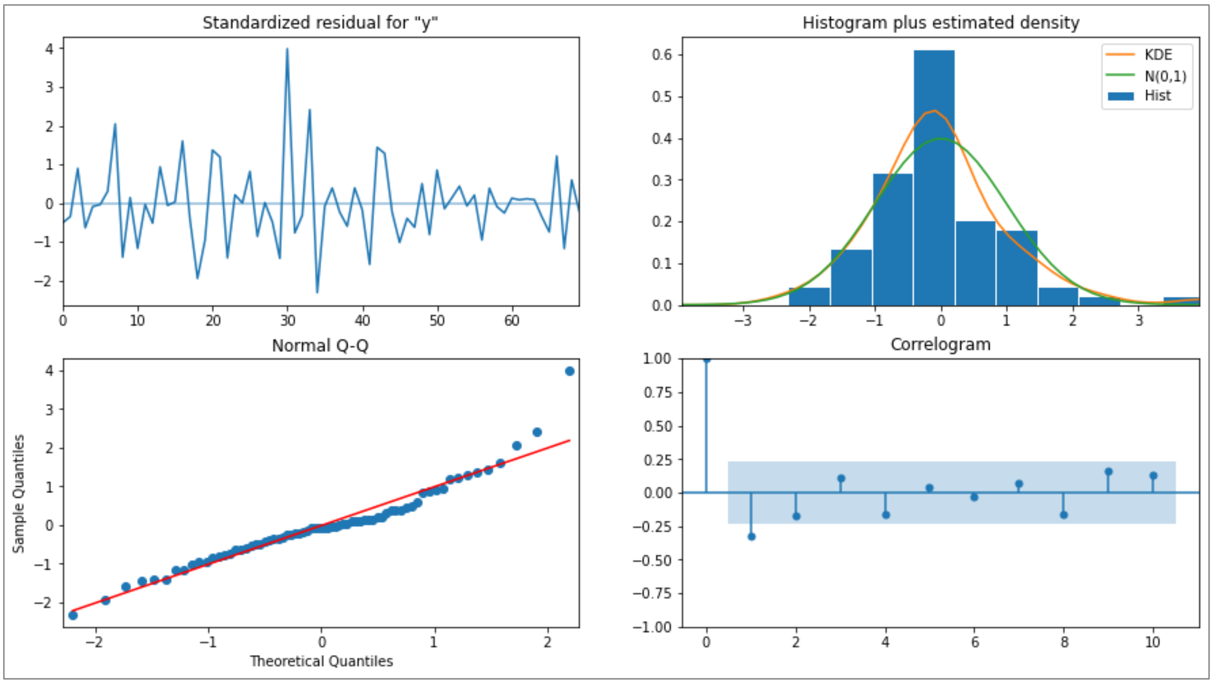

Figure 115. Sparkling residual diagnostic of manual SARIMA

Everything appears to be OK, including the residual component. The model is now good to go.

We'll now use this model to forecast values for the test data set and evaluate its performance using the RMSE score.

The RMSE score on the test data set with this SARIMA model is 937.445.

SARIMA model for Rose:

We have already seen from the ACF plot that the seasonality for Sparkling wine train dataset after one difference is 12. So, we will be taking Seasonality = 12.

p and q values from ACF and PACF plots are 2 and 2.

D = 1 (Differencing of 1) and finally P and Q values from the ACF and PACF plots are 2 and 4 respectively.

So, the final order will be **(2, 1, 2) (2, 1, 4, 12)** We take this order and build a SARIMA model on the training data.

Checking the summary for Rose for manual SARIMA model:

```
                                        SARIMAX Results
==========================================================================================
Dep. Variable:                                      y   No. Observations:              132
Model:             SARIMAX(2, 1, 2)x(2, 1, [1, 2, 3, 4], 12)   Log Likelihood          -276.118
Date:                                Wed, 23 Mar 2022   AIC                        574.236
Time:                                        03:17:47   BIC                        598.651
Sample:                                             0   HQIC                       583.910
                                                - 132
Covariance Type:                                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -1.0068      0.182     -5.525      0.000      -1.364      -0.650
ar.L2         -0.1190      0.157     -0.759      0.448      -0.426       0.188
ma.L1          0.1071   4633.725   2.31e-05      1.000   -9081.828    9082.042
ma.L2         -0.8929   4137.264     -0.000      1.000   -8109.780    8107.995
ar.S.L12      -1.1228      0.136     -8.263      0.000      -1.389      -0.856
ar.S.L24      -0.4381      0.119     -3.682      0.000      -0.671      -0.205
ma.S.L12       0.6868   3672.518      0.000      1.000   -7197.317    7198.691
ma.S.L24      -0.3282   1333.671     -0.000      1.000   -2614.275    2613.619
ma.S.L36      -0.7460   3809.820     -0.000      1.000   -7467.855    7466.363
ma.S.L48       0.0961    445.613      0.000      1.000    -873.289     873.482
sigma2       134.4710     45.874      2.931      0.003      44.560     224.382
===================================================================================
Ljung-Box (L1) (Q):                   0.04   Jarque-Bera (JB):                 6.49
Prob(Q):                              0.84   Prob(JB):                         0.04
Heteroskedasticity (H):               0.69   Skew:                             0.58
Prob(H) (two-sided):                  0.39   Kurtosis:                         3.97
===================================================================================
```

Figure 116. Summary of manual SARIMA for Rose.

Checking a diagnostic on the residual:

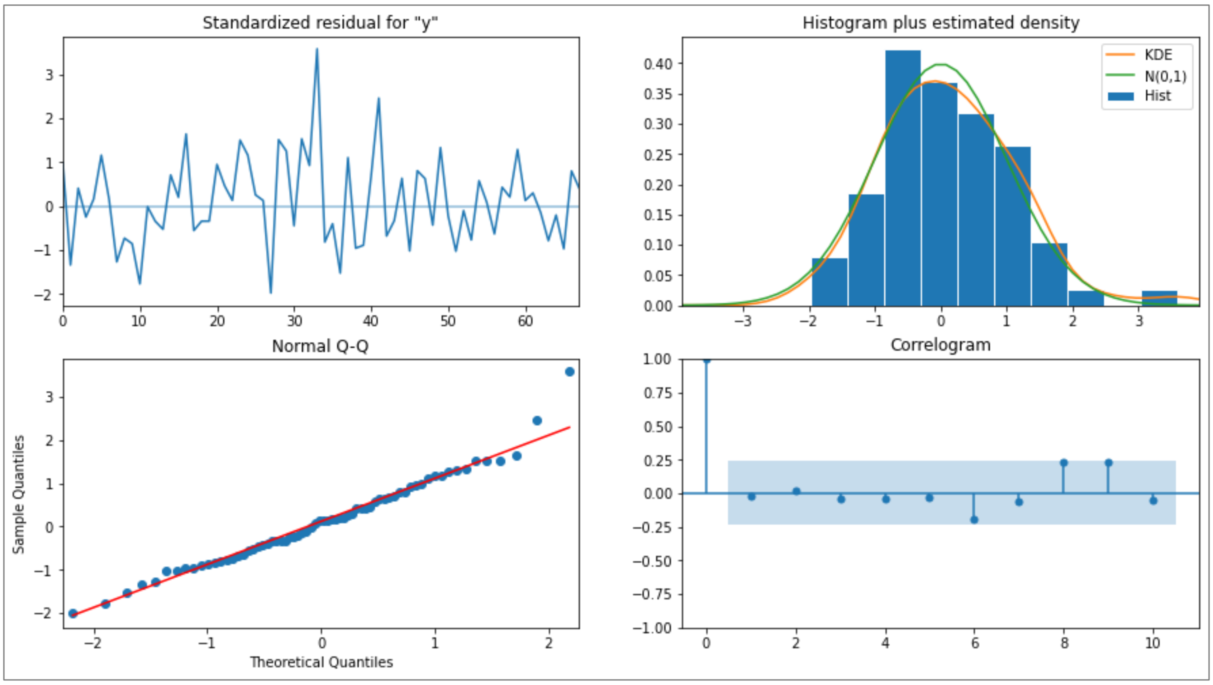

Figure 117. Rose residual diagnostic of manual SARIMA

The residual component is normal, everything looks good. The model is good to go. We will now forecast values for test data set using this model built and check for its performance with the help of RMSE score. With this SARIMA model.

The RMSE score on test data set is 16.931.

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

All of the models have now been created. All models have their performance evaluated using RMSE values. Now we'll compare the different models based on their RMSE values to see which one is the best fit.

Sparkling Wine:

	Test RMSE-Sparkling
Alpha=0.4,Beta=0.01,Gamma=0.3:Tuned Triple Exponential Smoothing	326.579641
Alpha=0.111,Beta=0.049,Gamma=0.362:Triple Exponential Smoothing	380.398478
SARIMA(1,1,2)(1,0,2,12) AIC criteria	528.611364
2point Trailing Moving Average	813.400684
SARIMA(0,1,0)(2,1,4,12) Manual plot	937.540131
4point Trailing Moving Average	1156.589694
Simple Average	1275.081804
Alpha=0.02:Tuned Simple Exponential Smoothing	1279.495201
6point Trailing Moving Average	1283.927428
ARIMA(2,1,2) AIC criteria	1299.979569
Alpha=0.049:Simple Exponential Smoothing	1316.035487
9point Trailing Moving Average	1346.278315
Linear Regression	1389.135175
Alpha=0.1,Beta=0.1:Tuned Double Exponential Smoothing	1778.564670
ARIMA(0,1,0) Manual plot	3864.279352
Naive Approach	3864.279352
Alpha=0.66,Beta=0.0001:DoubleExponentialSmoothing	5291.879833

Table 7. Sparkling models RMSE comparison.

The RMSE Scores of all the models performed above are listed in ascending order, with least RMSE as the best performing model. We can see that for Sparkling wine sales data **Tuned Triple Exponential Smoothing with Alpha = 0.4, Beta = 0.1, Gamma = 0.3** is best fit model which can be used to forecast the sales with least **RMSE score of 326.57.**

Rose Wine:

	Test RMSE-Rose
Alpha=0.2,Beta=0.7,Gamma=0.2:Tuned Triple Exponential Smoothing	8.702460
2point Trailing Moving Average	11.529278
4point Trailing Moving Average	14.451403
6point Trailing Moving Average	14.566327
9point Trailing Moving Average	14.727630
Alpha=1.4901-08,,Beta=1.661-10:Double Exponential Smoothing	15.268944
Linear Regression	15.268955
ARIMA(0,1,2) AIC criteria	15.619203
SARIMA(2,1,2)(2,1,4,12) Manual plot	16.931818
Alpha=0.055,Beta=0.031,Gamma=0.00033:Triple Exponential Smoothing	19.987449
SARIMA(0,1,2)(2,0,2,12) AIC criteria	26.928362
Alpha=0.07:Tuned Simple Exponential Smoothing	36.435772
Alpha= 0.098:Simple Exponential Smoothing	36.796227
ARIMA(2,1,2) Manual plot	36.871197
Alpha=0.1,,Beta=0.1:Tuned Double Exponential Smoothing	36.923416
Simple Average	53.460570
Naive Approach	79.718773

Table 8. Rose models RMSE comparison.

The RMSE Scores of all the models performed above are listed in ascending order, with least RMSE as the best performing model. We can see that for Rose wine sales data **Tuned Triple Exponential Smoothing with Alpha = 0.2, Beta = 0.7, Gamma = 0.2** is best fit model which can be used to forecast the sales with least **RMSE score of 8.702**

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.

We've built all of the models so far and got the best fit for both datasets. Now, for both the Rose and Sparkling wine full datasets, we'll use the best fit models to forecast values for the following 12 months with appropriate confidence intervals.

Sparkling Wine:

Sparkling best model is Triple Exponential smoothing with Alpha = 0.4 , Beta = 0.1, Gamma = 0.3

We will build the most optimum Triple Exponential Smoothing model with the parameters Alpha = 0.4, Beta = 0.1, and Gamma = 0.3 on the Full Sparkling Dataset. **The RMSE value for the model on original dataset is 391.513**. And getting the predictions for 12 months into the future.

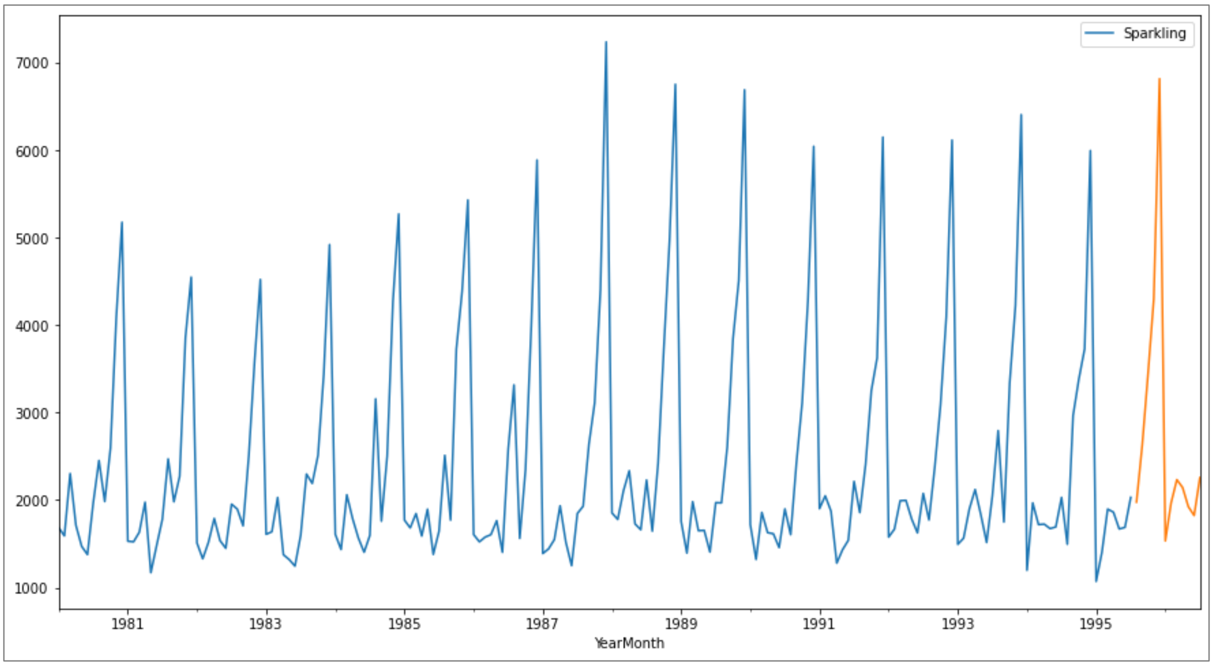

Figure 118. Plot of Sparkling data and forecast of 12months into future.

This model is used to forecast for the next 12 months. We'll further create an adequate confidence interval for the predictions on the graph. At a 95% confidence level, we construct the upper and lower confidence bands. We use 1.96 as the multiplier because we want to plot with 95 percent confidence intervals.

The predictions and lower and upper confidence intervals for 12 months in future is as below:

	lower_CI	prediction	upper_ci
1995-08-01	-0.538512	39.978825	80.496161
1995-09-01	-0.744898	39.772439	80.289776
1995-10-01	0.034640	40.551977	81.069314
1995-11-01	5.969809	46.487146	87.004483
1995-12-01	23.535285	64.052622	104.569959
1996-01-01	-16.118004	24.399333	64.916669
1996-02-01	-10.029194	30.488143	71.005479
1996-03-01	-5.152905	35.364432	75.881768
1996-04-01	-4.223762	36.293575	76.810911
1996-05-01	-9.811917	30.705419	71.222756
1996-06-01	-5.845705	34.671631	75.188968
1996-07-01	-3.264476	37.252860	77.770197

Table 9. Sparkling model prediction for future 12 months

Plotting the graph for Sparkling data and forecast along with the confidence band:

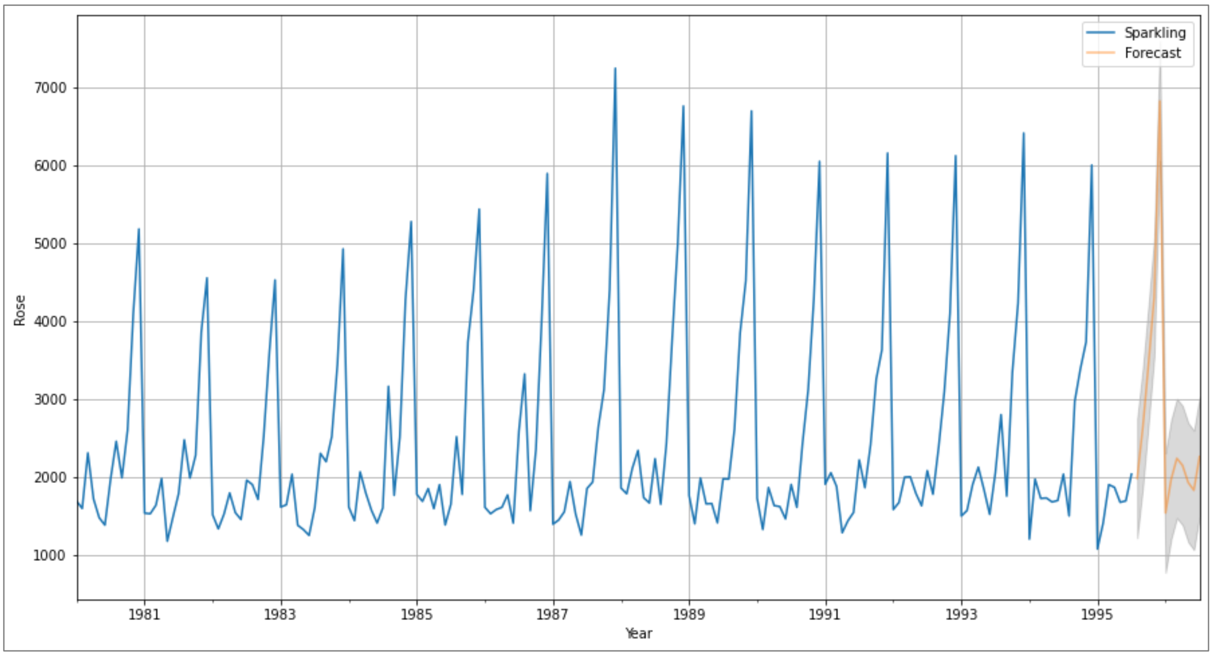

Figure 119. Plot of Sparkling data and forecast along with the confidence band

The predictions are consistent with the original dataset, as seen in the graph above. The seasonality and trend are maintained.

Rose Wine:

Rose best model is Triple Exponential smoothing with Alpha = 0.2, Beta = 0.7, Gamma = 0.2. We will build the most optimum Triple Exponential Smoothing model with the parameters Alpha = 0.2, Beta = 0.7, and Gamma = 0.2 on the Full Rose Dataset. **The RMSE value for the model on original dataset is 20.681**. And getting the predictions for 12 months into the future.

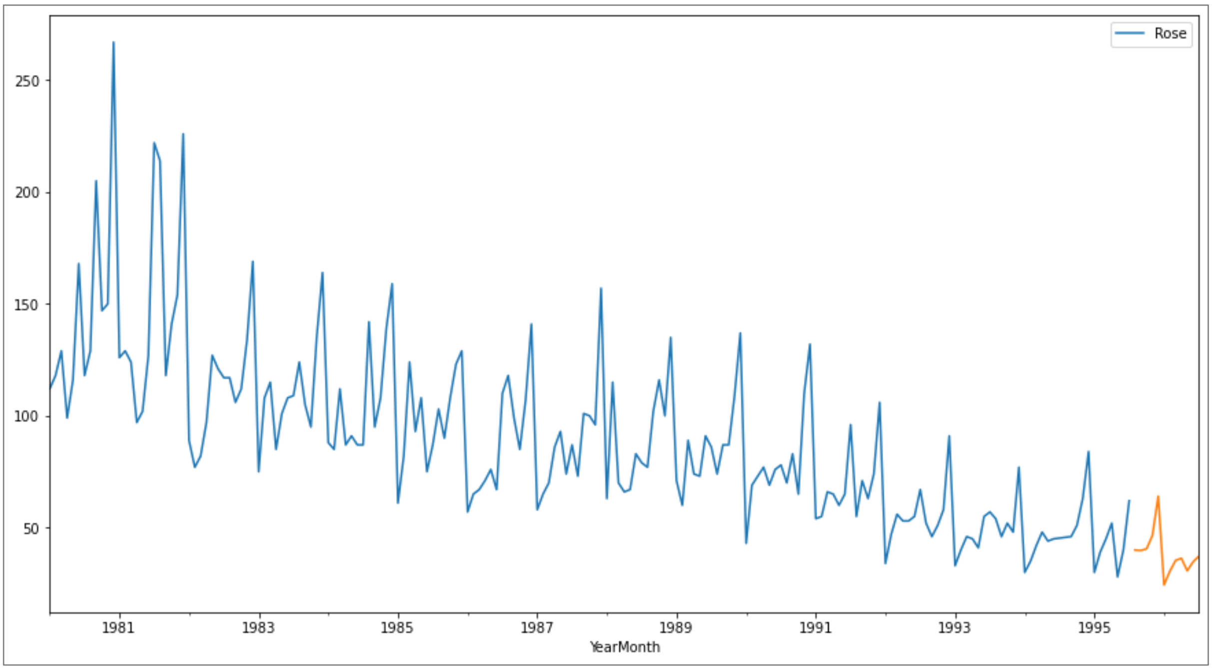

Figure 120. Plot of Rose data and forecast of 12months into future

This model is used to forecast for the next 12 months. We'll further create an adequate confidence interval for the predictions on the graph. At a 95% confidence level, we construct the upper and lower confidence bands. We use 1.96 as the multiplier because we want to plot with 95 percent confidence intervals.

	lower_CI	prediction	upper_ci
1995-08-01	-0.538512	39.978825	80.496161
1995-09-01	-0.744898	39.772439	80.289776
1995-10-01	0.034640	40.551977	81.069314
1995-11-01	5.969809	46.487146	87.004483
1995-12-01	23.535285	64.052622	104.569959
1996-01-01	-16.118004	24.399333	64.916669
1996-02-01	-10.029194	30.488143	71.005479
1996-03-01	-5.152905	35.364432	75.881768
1996-04-01	-4.223762	36.293575	76.810911
1996-05-01	-9.811917	30.705419	71.222756
1996-06-01	-5.845705	34.671631	75.188968
1996-07-01	-3.264476	37.252860	77.770197

Table 10. Rose model prediction for future 12 months

<u>Plotting the graph for Rose data and forecast along with the confidence band:</u>

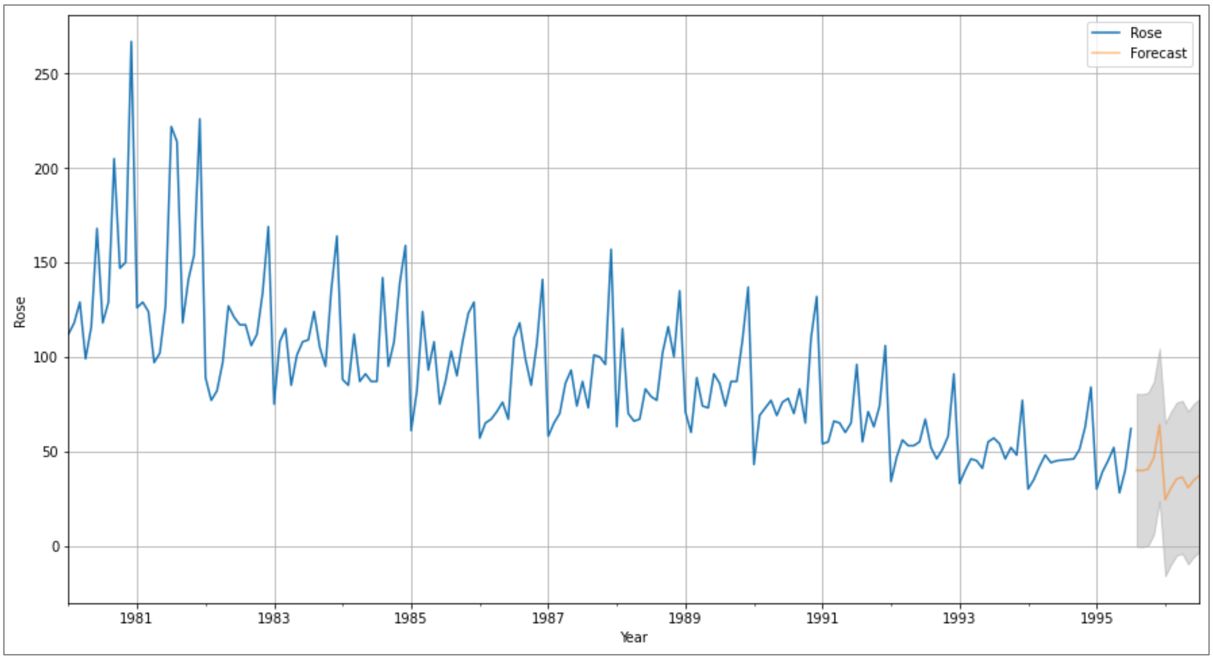

Figure 121. Plot of Rose data and forecast along with the confidence band

The predictions are consistent with the original dataset, as seen in the graph above. The seasonality and trend are maintained.

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

The problem statement states that we must analyse past sales of two distinct types of wines (Rose and Sparkling) for ABC Estate Wines throughout the twentieth century for 15 years. build several models based on the analysis, and choose the best fit model to forecast future sales.

1. Let's take a look at a quick rundown of the stages involved in accomplishing this.
2. First, we read the data as a time series and plotted it on a graph to show how sales for Sparkling and Rose wines have changed over the years.
3. We performed some exploratory data analyses on the data sets, creating various types of charts for better understanding of sales.
4. Next, we moved on to the model building section, where we began by splitting the data into test and train, treating all data less than 1991 as train and all data after and equal to 1991 as test.
5. Started with Basic models, Built the models:
    - Linear Regression Model
    - Naïve Approach
    - Simple Average Model
    - Moving Average Model
6. Then we moved on to create Exponential Smoothing models. We built the following:
    - Simple Exponential Smoothing Model
    - Tuned Simple Exponential Smoothing Model
    - Double Exponential Smoothing Model
    - Tuned Double Exponential Smoothing Model
    - Triple Exponential Smoothing Model
    - Tuned Triple Exponential Smoothing Model.
7. We calculate the RMSE values for all of these models to see how well they perform. The model with the lowest RMSE value is the most accurate.
8. From here, we build ARIMA and SARIMA models, but first we examine the dataset for stationarity. By stating hypothesis for statistical testing, we employ the ADF test. If the series is not stationary, we use the difference of the series.
9. The ARIMA/SARIMA models are now being built. Using AIC scores, we first create auto predictive models. Then we build the ones when the parameters are entered manually by the ACF and PACF graphs visually.
10. We check the performance of all these models and compare the RMSE values in the tabular form sorted in the ascending order, where least RMSE is on top and the best fit model.
11. Finally, we take the best fit model with minimum RMSE value and build that most optimum model on the complete data and then predict the sales for the next 12 months in future with appropriate confidence intervals
12. All these steps are performed for both Rose and Sparkling wine sales separately.
13. Finally, the suggestions and recommendations that the company must can take for future sales.

Sparkling wine: Insights and Recommendation

Let us have a look at Monthly Sale Boxplot for Sparkling wine throughout the years of given dataset:

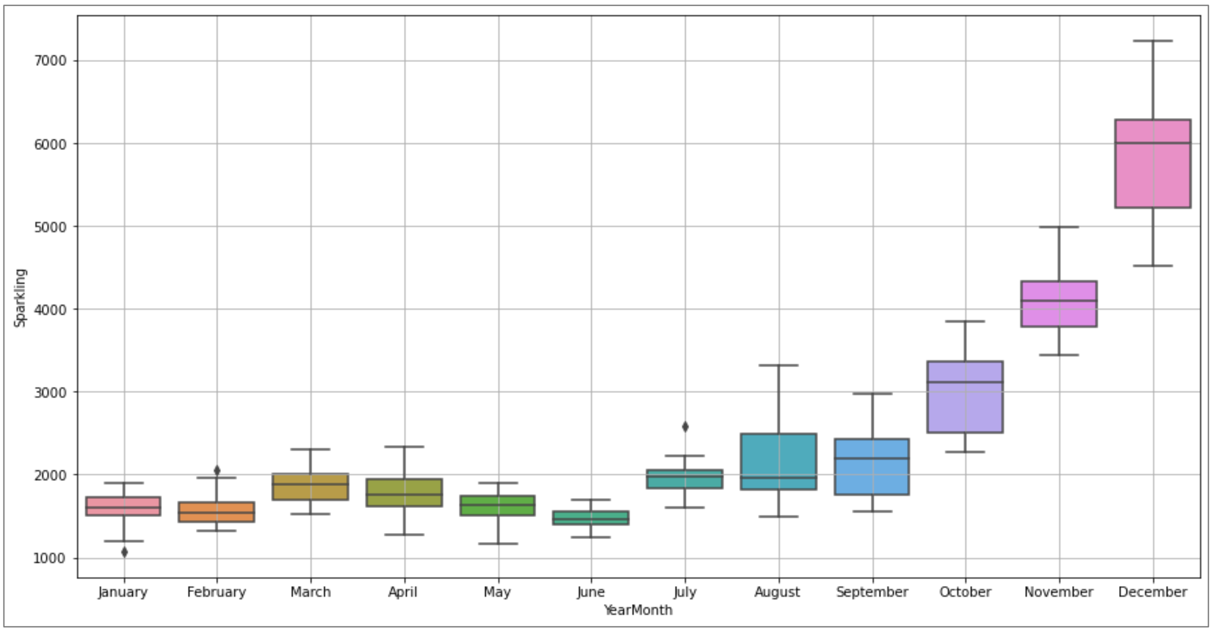

Figure 122. Monthly sale boxplot for Sparkling.

The first and second quarters indicate low sales of sparkling wines, whereas the third and fourth quarters show extraordinarily strong increasing trend in sales. This frothy champagne wine is especially popular around the holidays. The wine is mainly have high sales or sold-out during Thanksgiving, Halloween party, Christmas, and New Year's Eve. So, this is evidently seen in the boxplot above.

Looking at the average Sales per month and the month-on-month percentage change of Sales of Sparkling wine:

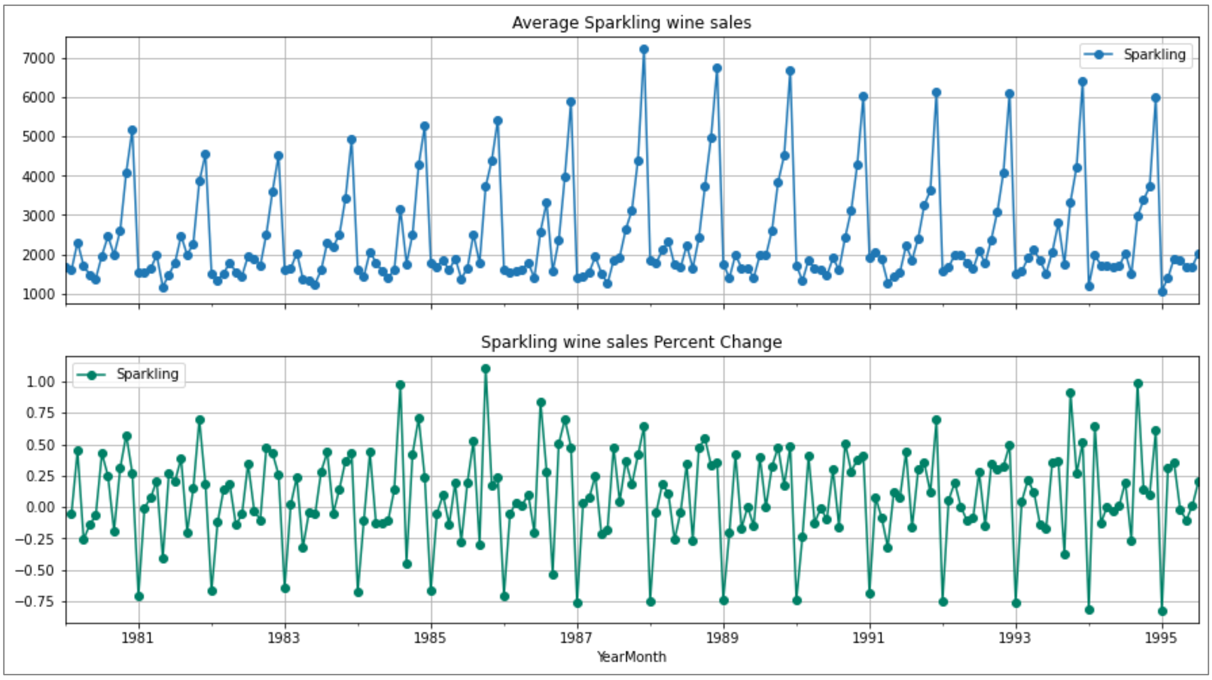

Figure 123. Avg sales and percentage sales for Sparkling.

By learning the little history about the Sparkling wine, we can say that, back in 1980's the majority of individuals only drank sparkling wine on exceptional occasions. It was royal to drink because of the fuzzy sparkle caused by the carbon di-oxide in it.

The Sparkling wine was considered to be royal drink and there were effective means of enticing target population to purchase the Sparkling wine on the basis of taste, convenience and novelty as compared to other types of wine, because of these qualities the wine always maintained its place in royalty drink.

We can see the same in average sales plot that the sales of sparkling wines over different years month-on-month is almost consistent with seasonality within the year.

The forecast for future 12 months in done on Tuned Triple Exponenetial model which is optimal model with least RMSE value.

<u>Recommendations to the Sparkling wine company:</u>

- For forecasting the future sales, using the Tuned Triple Exponential Smoothing model with Alpha = 0.4, Beta = 0.1, Gamma = 0.3 since it gives the lowest RMSE value as compared to other time series models built and gives the best prediction.
- **Social Media**: social media, if used correctly with dedicated time given to it and quality content to share can most certainly at the very least increase your brand profile thus doing what one of your main tasks should be in marketing, which may cost financially less to achieve, but will save you money and time in the long run by achieving a bigger reach and getting in front of more active audiences.
- **Wine Tasting:** Levering a big event**,** plan for a promotion around a big event nearby, like a local sporting event, concert, or festival. This helps drive business and generate awareness of Sparkling wine to a large audience.
- **Press / Media / Magazines**: Most people, within the industry or not, will know the likes of the Decanter magazine so it is without question that a feature / mention / praise within its pages is a good plus for marketing in itself, the audience you can reach far outweighs the possible results of your own endeavours or that of PR company.
- **Branding and USP** is also important.
- **Your wines**. To produce the most unique wine ever and this alone carries it out to the whole industry and great reviews breed themselves on the back of what your wines taste like – The ideal situation for any wine maker / wine house.
- In order to be ready for the strong demand in the last quarter, the company should focus on inventories in the first two quarters.
- New deals and discounts may be made available on various occasions to ensure that sales are available all year long, not only during the holiday season.
- Obtain further information to determine which sector and location of the population consumes the most sparkling wine. Those locations or groups of individuals are the ones you can't change. Continue in the same direction. Bring something fresh to the areas with poor sales. Experiment with special deals and discounts, as well as a fresh marketing strategy. People's behaviour changes as a result of tactical adjustments, according to research.

Rose wine: Insights and Recommendation:

Let us have a look at Monthly Sale Boxplot for Sparkling wine throughout the years of given dataset:

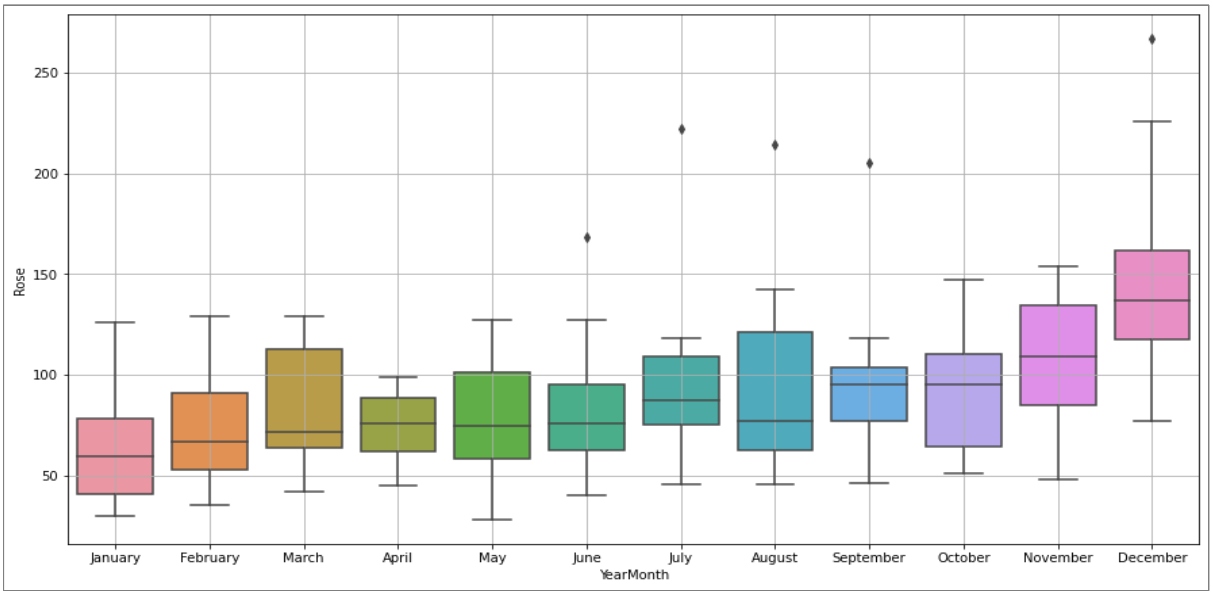

Figure 124. Monthly sale boxplot for Rose.

We can observe that sales of Rose wines begin to rise in August and continue to rise until December. The month of December has the most sales of any other month this can be due to Thanksgiving, Halloween party, Christmas, and New Year's Eve [festive season] which is evidently seen in the boxplot above.

The first quarter has the lowest sales, while the fourth quarter has the largest. The second and third quarters are reasonable, not too high or too low.

Looking at the average Sales per month and the month-on-month percentage change of Sales of Rose wine:

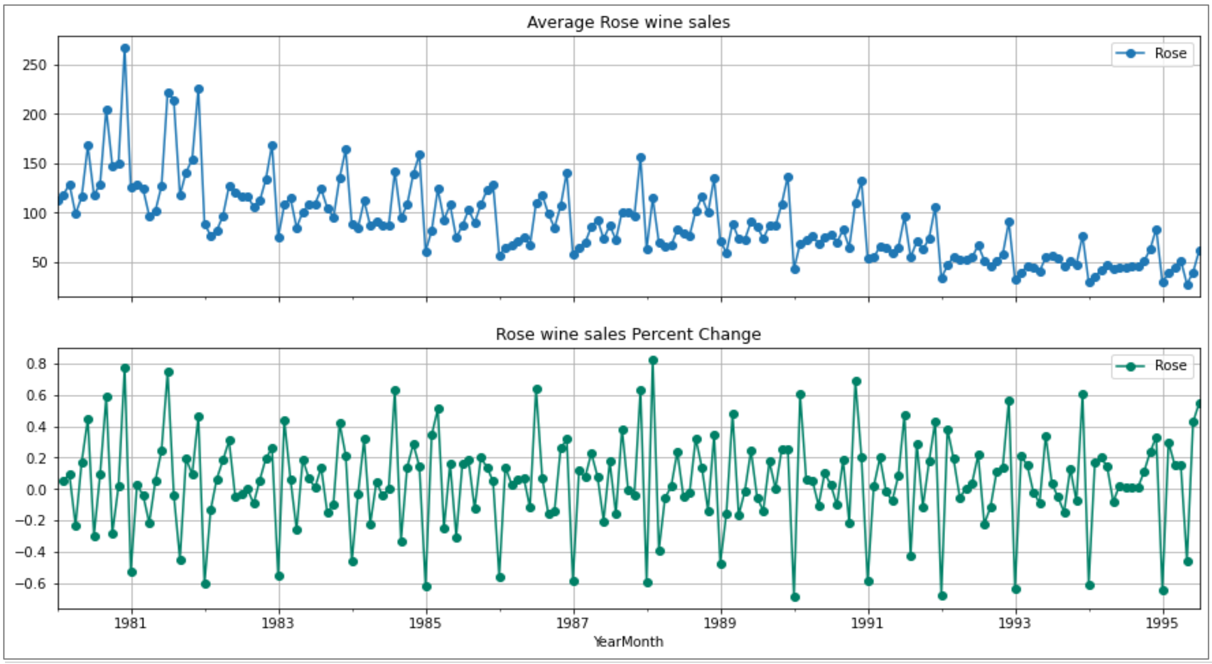

Figure 125. Avg sales and percentage sales for Rose.

By learning the little history about the Rose wine, we can say that, back in 1980's the Rose wine is the primary product with skin contact method. In which the black skinned grapes were crushed and the skins was allowed to remain in contact with juice for certain period of time.

Because of its uniqueness and new product in market the Rose wine sales were initially high, which can be seen in the plot above.

The Rose wine started experiencing downfall in quality of wine, while at the same time other wines with the same properties came in market at that time which gradually started decreasing the sales of wine.

Even in average sales of rose wine, we can evidently see the downward trend in sales.

The forecast for future 12 months in done on Tuned Triple Exponenetial model which is optimal model with least RMSE value.

<u>Recommendations to the Rose wine company:</u>

- For predicting the future sales, use the Triple Exponential Smoothing model with Alpha = 0.2, Beta = 0.7, Gamma = 0.2 because it gives the lowest RMSE value as compared to other Time series models and would give the best prediction.
- **Social Media**: social media, if used correctly with dedicated time given to it and quality content to share can most certainly at the very least increase your brand profile thus doing what one of your main tasks should be in marketing, which may cost financially less to achieve, but will save you money and time in the long run by achieving a bigger reach and getting in front of more active audiences.
- **Wine Tasting:** Levering a big event**,** plan for a promotion around a big event nearby, like a local sporting event, concert, or festival. This helps drive business and generate awareness of Sparkling wine to a large audience.
- **Press / Media / Magazines**: Most people, within the industry or not, will know the likes of the Decanter magazine so it is without question that a feature / mention / praise within its pages is a good plus for marketing in itself, the audience you can reach far outweighs the possible results of your own endeavours or that of PR company.
- **Branding and USP** is also important.
- **Your wines**. To produce the most unique wine ever and this alone carries it out to the whole industry and great reviews breed themselves on the back of what your wines taste like – The ideal situation for any wine maker / wine house.
- We can observe from our graph that sales increased in the fourth quarter. As a result, the company should arrange their inventory from the start to deal with the high sales time.
- The company should review customer feedback over time to determine why rose wine was not a favour of theirs, and seek to improve the quality.
- Once the quality is up to par, many various methods of advertising the wine should be implemented so that people are aware of the improvements and may sample the improved product.
- Bringing something fresh to the areas with poor sales. Experiment with special deals and discounts, as well as a fresh marketing strategy. People's behaviour changes as a result of tactical adjustments, according to research.
