

# **Time Series Forecasting Business Report**

**By: Ambrish Verma**

**Date: 25-Feb-2024**

## Contents

Date: 25-Feb-2024 .....	1
<b>1. Read the data as an appropriate Time Series data and plot the data. ....</b>	<b>5</b>
<b>2. Perform appropriate Exploratory Data Analysis to understand the data and perform decomposition. ....</b>	<b>6</b>
<b>3. Split the data into training and test. The test data should start in 1991. ....</b>	<b>12</b>
<b>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. ....</b>	<b>13</b>
<b>5. Check for the stationarity of the data on which the model is being built on using appropriate statistical tests and 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 <math>\alpha = 0.05</math>. ....</b>	<b>16</b>
<b>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. ....</b>	<b>20</b>
<b>7. Build a table with all the models built along with their corresponding parameters and the respective RMSE values on the test data. ....</b>	<b>26</b>
<b>8. 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. ....</b>	<b>27</b>
<b>9. Comment on the model thus built and report your findings and suggest the measures that the company should be taking for future sales. ....</b>	<b>29</b>
Table 1 Interpolation - Original versus Interpolated Values.....	5
Table 2 Sparkling - train data – head.....	12
Table 3 Sparkling - test data – head .....	13
Table 4 Rose - train data - head.....	13
Table 5 Rose - test data - head .....	13
Table 6 Sales forecast - Sparkling Wine - multiple models.....	14
Table 7 Sales Forecast - Rose wine - multiple models.....	14
Table 8 RMSE values - sparkling wine – multiple models .....	15
Table 9 RMSE values - sparkling wine - multiple models .....	15
Table 10 Lowest AIC for ARIMA - sparkling wine .....	20
Table 11 Lowest AIC for ARIMA - rose wine .....	23
Table 12 Lowest AIC for SARIMA - rose wine .....	24
Table 13 RMSE values for models - sparkling wine .....	26
Table 14 RMSE values for models - rose wine.....	26
Table 15 Sales prediction values - Sparkling wine.....	28
Table 16 Sales prediction values - Rose wine.....	29

Figure 1: Wines Sale volume over the years .....	5
Figure 2 Wines Sale Volume - Post Interpolation.....	6
Figure 3 Box-plot of Monthly sparkling wine sales .....	7
Figure 4 Box-plot of monthly Rose Wine sales.....	7
Figure 5 Plot of sparkling wine sales - monthly trend over the years.....	8
Figure 6 Plot of rose wine sales - monthly trend over the years .....	8
Figure 7 Sparkling Wine sales - for individual months .....	9
Figure 8 Rose Wine sales - for individual months .....	9
Figure 9 Seasonal Decompose - Additive .....	10
Figure 10 Multiplicative Decomposition .....	10
Figure 11 Residuals - Additive decompose .....	11
Figure 12 Residuals - Multiplicative decompose.....	12
Figure 13 Stationarity test - Sparkling wine .....	16
Figure 14 Stationarity test - Rose wine .....	17
Figure 15 Stationarity achieved - sparkling wine .....	18
Figure 16 Stationarity achieved - Rose wine .....	19
Figure 17 Differenced auto-correlation of Sparkling Wine sales .....	20
Figure 18 Lowest AIC for SARIMA - sparkling wine .....	21
Figure 19 Differenced auto-correlation of Rose Wine sales.....	23
Figure 20 Sales prediction - sparkling wine.....	27
Figure 21 Sales prediction - rose wine .....	28

### Problem Statement:

The data of different types of wine sales in the 20th century is to be analyzed. Both data are from the same company but of different wines. As an analyst in the ABC Estate Wines, you are tasked to analyze and forecast Wine Sales in the 20th century.

### Questions:

1. Read the data as an appropriate Time Series data and plot the data.
2. Perform appropriate Exploratory Data Analysis to understand the data and perform decomposition.
3. Split the data into training and test. The test data should start in 1991.
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.
5. Check for the stationarity of the data on which the model is being built on using appropriate statistical tests and 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  $\alpha = 0.05$ .
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.
7. Build a table with all the models built along with their corresponding parameters and the respective RMSE values on the test data.
8. 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.
9. Comment on the model thus built and report your findings and suggest the measures that the company should be taking for future sales.

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

- There are 187 entries made in both the datasets.
- The datasets are both time series data with the sale volumes of the 2 wines measured over months.
- The start date for the captured sale volumes is: January-1980
- The end date for the captured sale volumes is: July-1995
- The sale volume is plotted in the below graph:

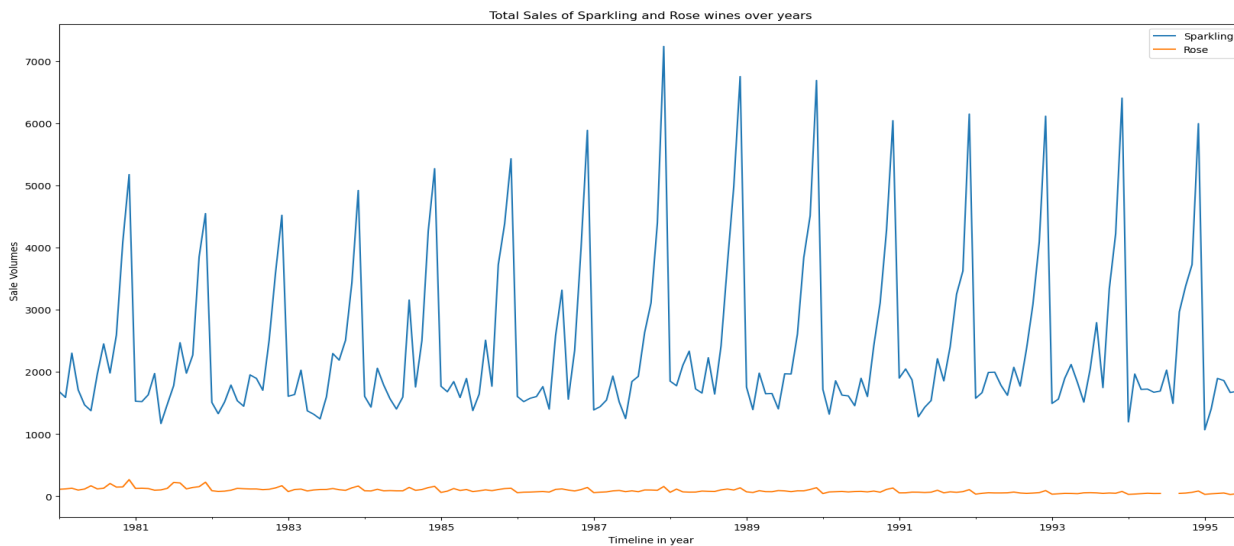


Figure 1: Wines Sale volume over the years

- As is evident in the above graph, there are missing values for the Rose wine in the year 1994. Hence, Interpolation will be done to fill the missing values.
- The 2 missing values in the Rose sale volume, were resampled using the adjoining values. The Original values and replaced values are mentioned in the table below:

TimeSeries	Old Value	Replaced Value
1994-06	45	45
1994-07	NaN	45.33333333
1994-08	NaN	45.66666667
1994-09	46	46

Table 1 Interpolation - Original versus Interpolated Values

- After resampling, the plot for the updated sale volumes is shown below:

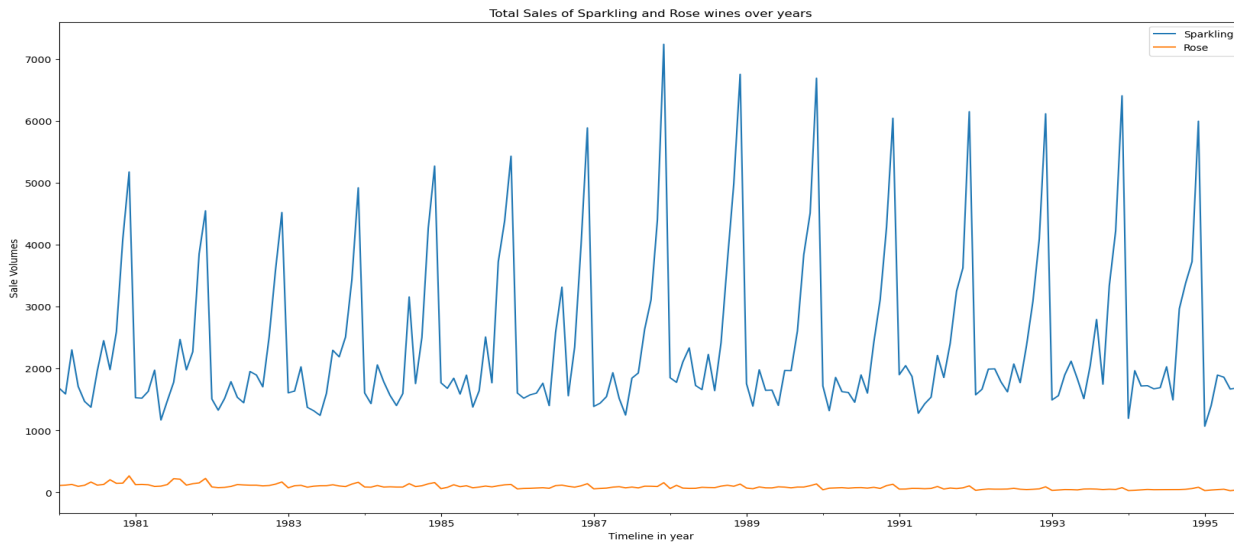
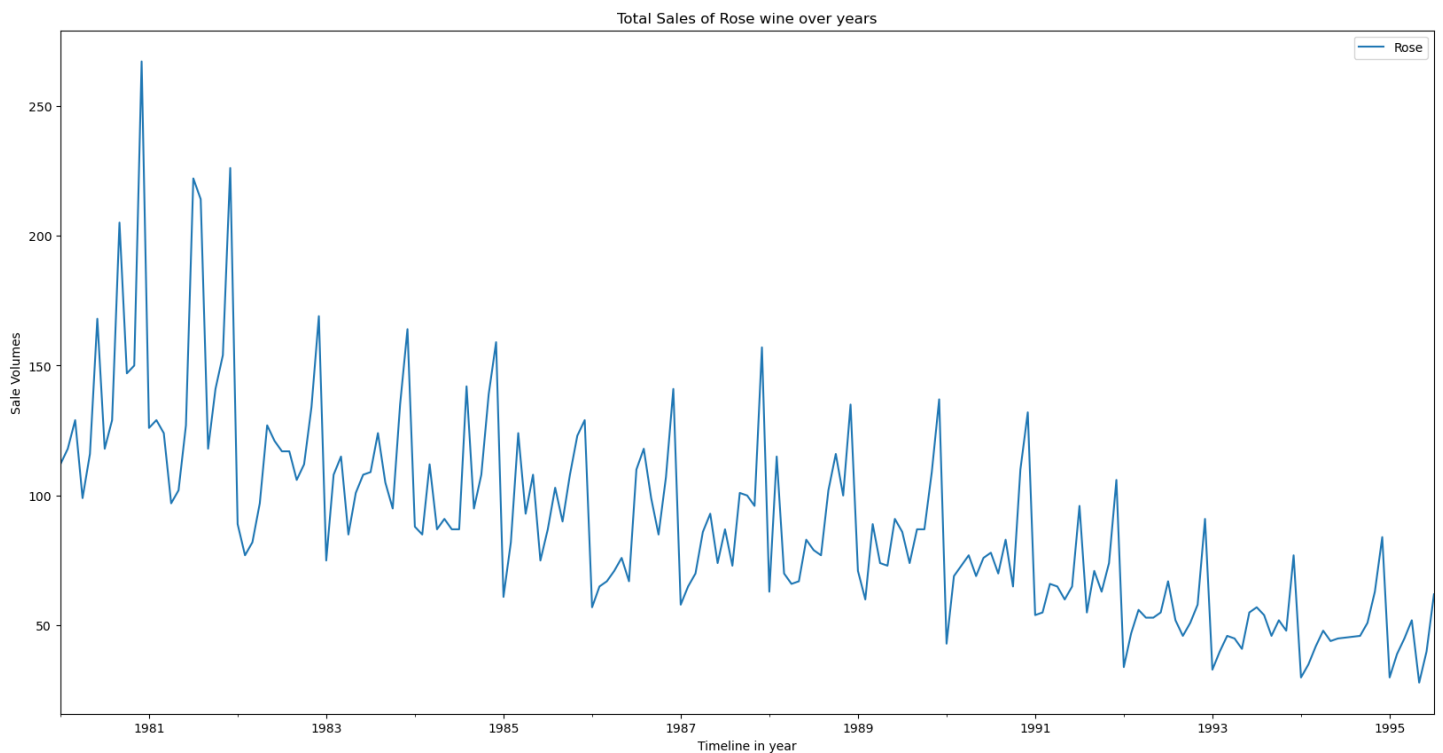


Figure 2 Wines Sale Volume - Post Interpolation

- Since the volume of Rose wine is much smaller, the trend or seasonality in this data is not clear, hence splitting the graph and plotting the Sale volume of Rose Wine again:



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

- Box plot of Sparkling Wine sales over the years split month-wise.

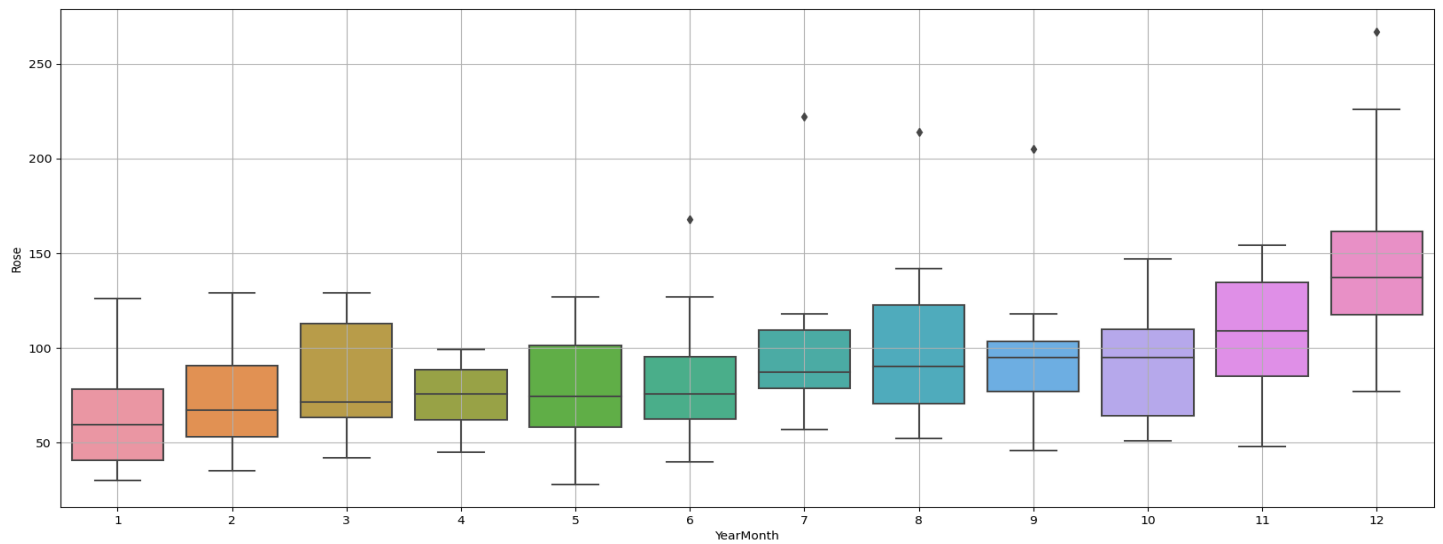


Figure 3 Box-plot of Monthly sparkling wine sales

- Box-plot of Rose wine sales over the years split month-wise:

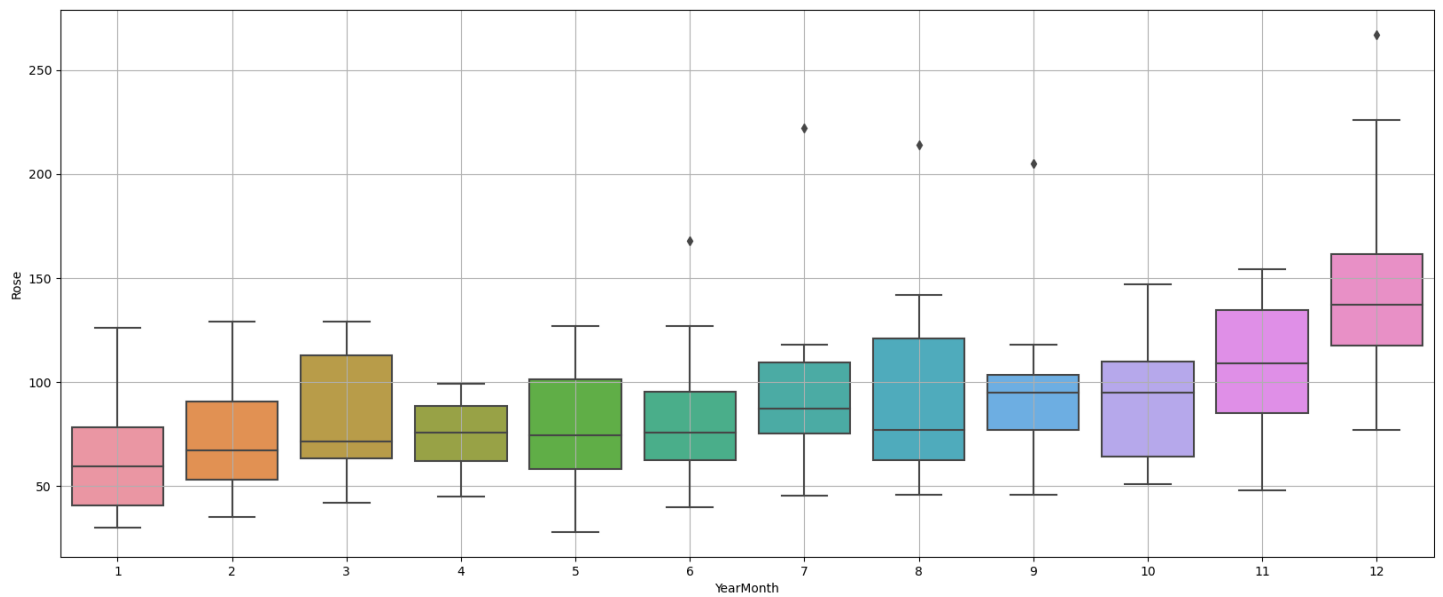


Figure 4 Box-plot of monthly Rose Wine sales

- Month-plot of the wines sales of sparkling wine(to show the trend of the monthly sales)

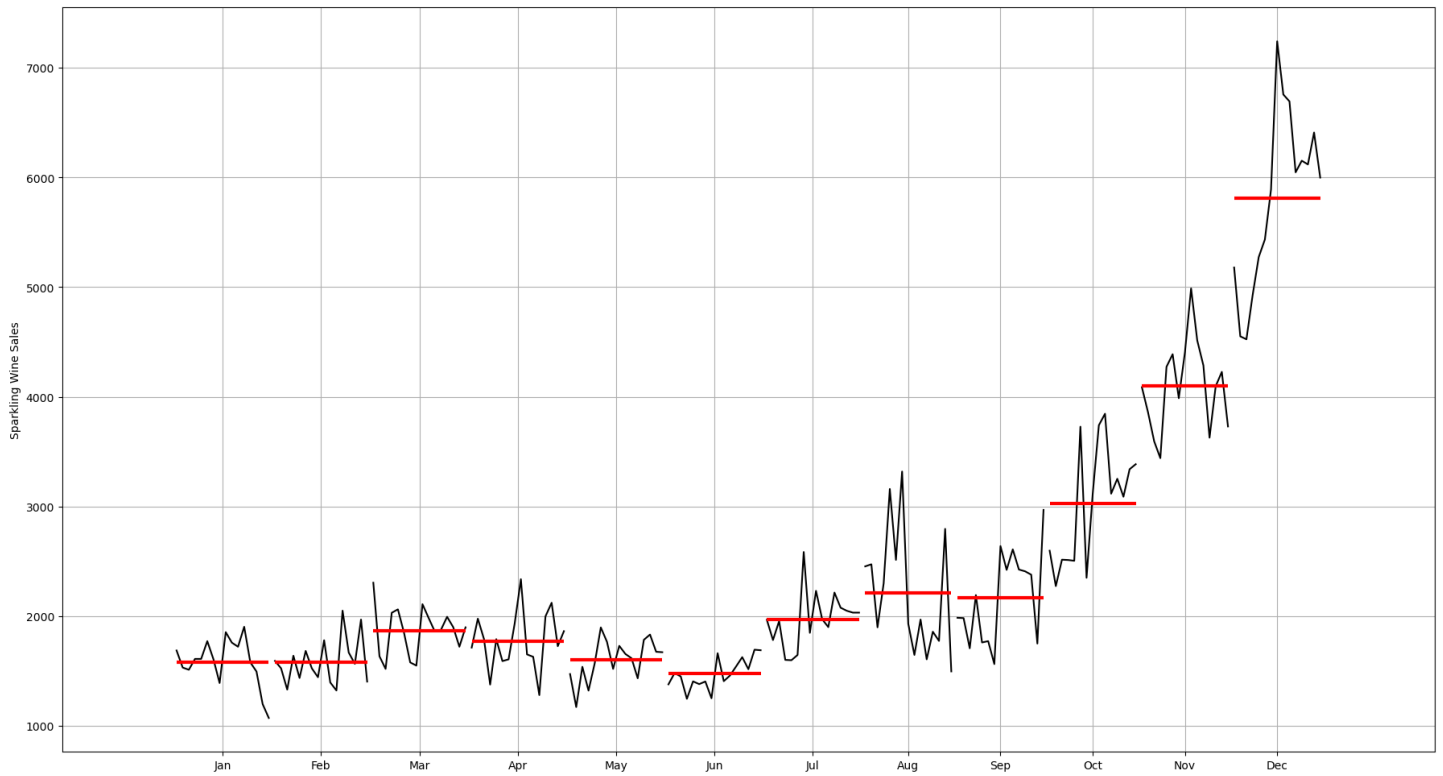


Figure 5 Plot of sparkling wine sales - monthly trend over the years

- Month-plot of the wines sales of Rose wine(to show the trend of the monthly sales)

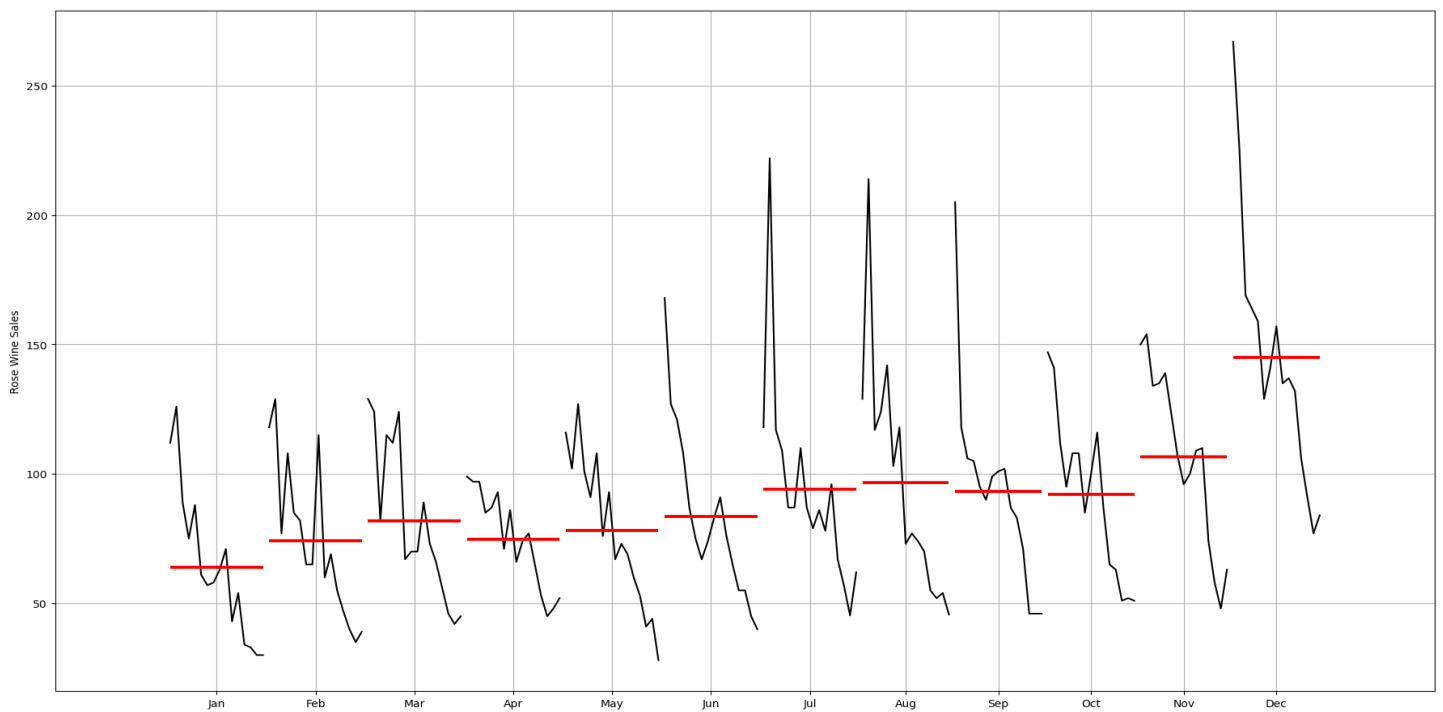


Figure 6 Plot of rose wine sales - monthly trend over the years

- Sparkling Wines sales over the years spread across months:



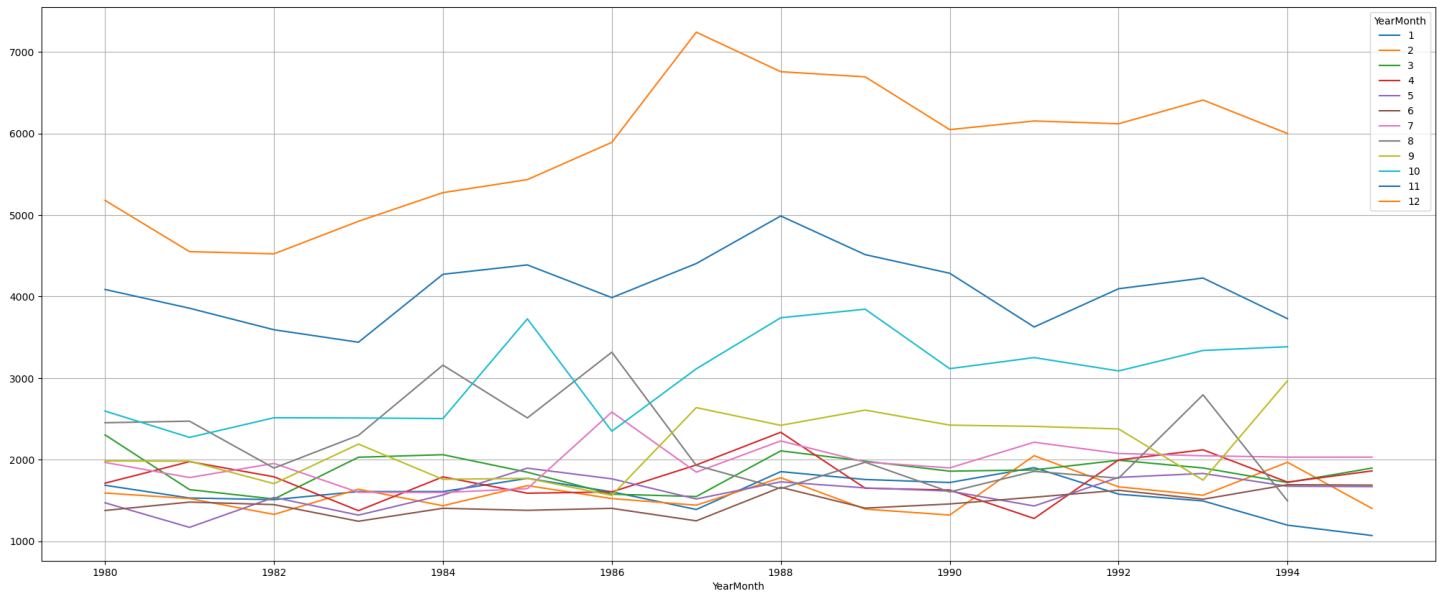


Figure 7 Sparkling Wine sales - for individual months

- Rose wines sales over the years spread across months:

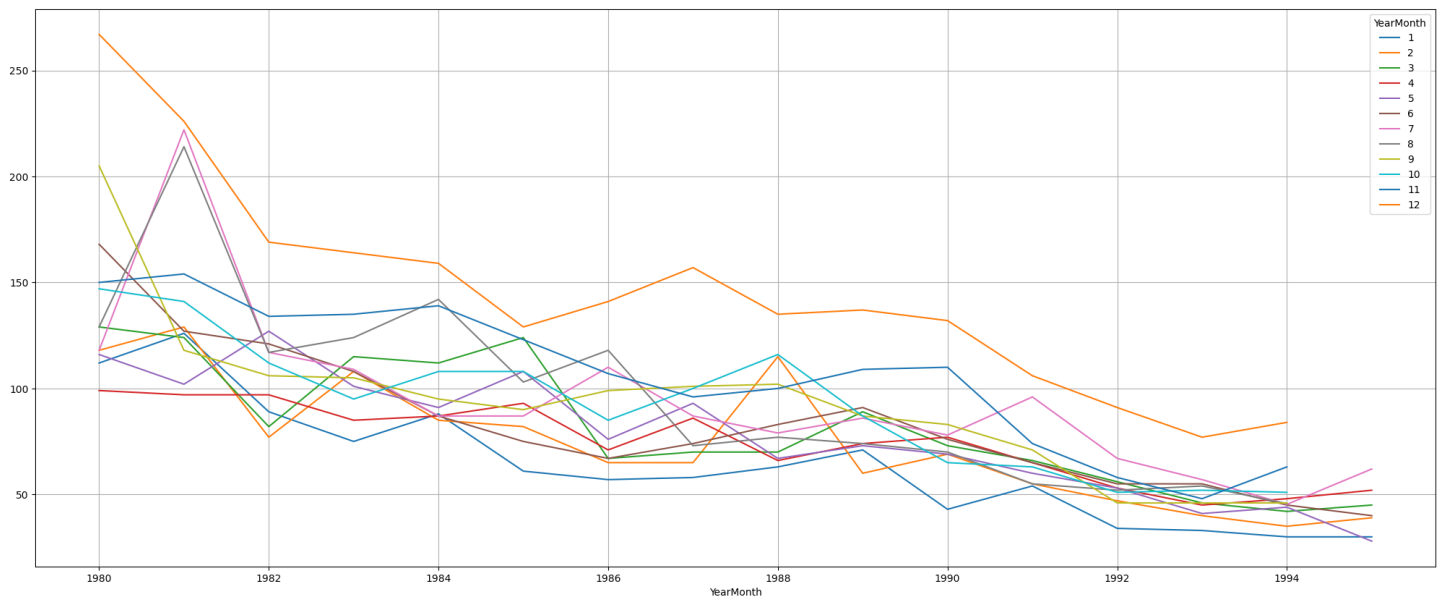


Figure 8 Rose Wine sales - for individual months

- It is evident in the above graphs that there is a downward trend for Rose wine sales across all the months.
- Also, there is slight seasonality in the sale volume of Rose wines as well. The sales volume shows a slight uptick towards second half of a year.
- In the case of sparkling wines, there does not seem to be an upward/downward trend.

- But there is seasonality in the sale volume of sparkling wine, wherein, the sales spike in the last quarter (October-December).
- Performing additive decomposition:

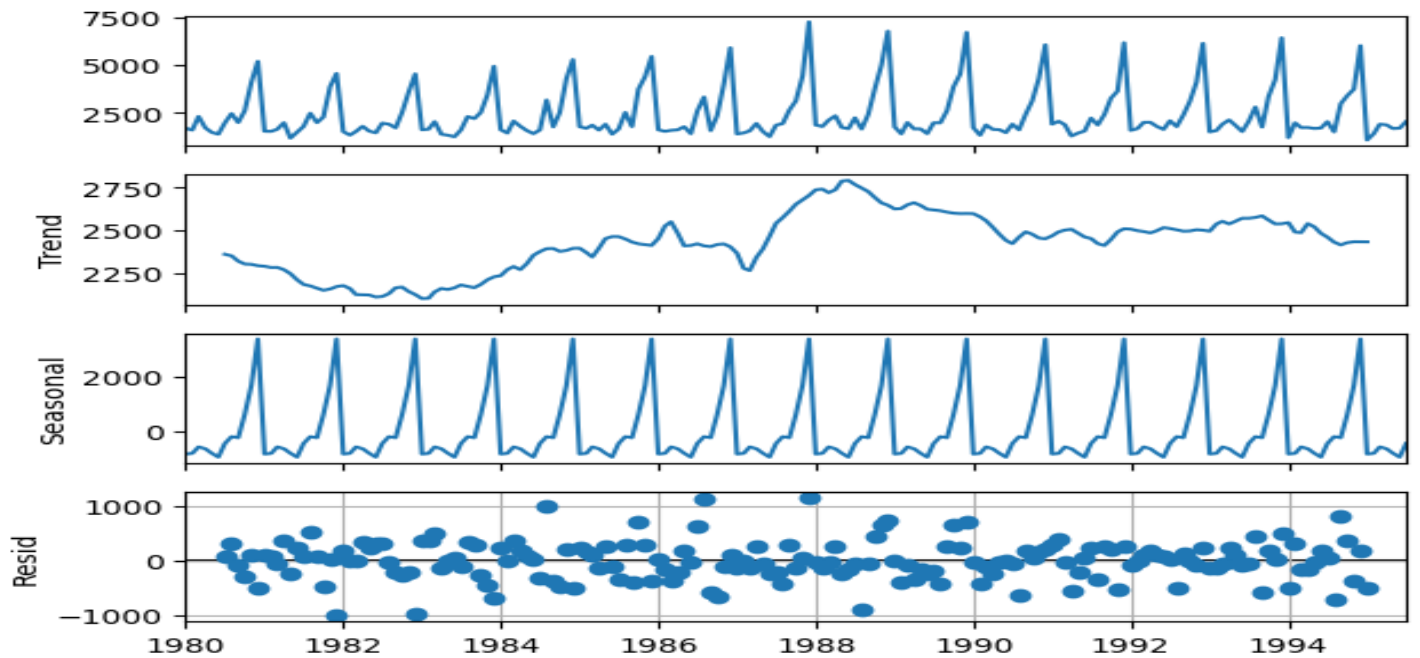


Figure 9 Seasonal Decompose - Additive

- Performing Multiplicative Seasonal Decompose:

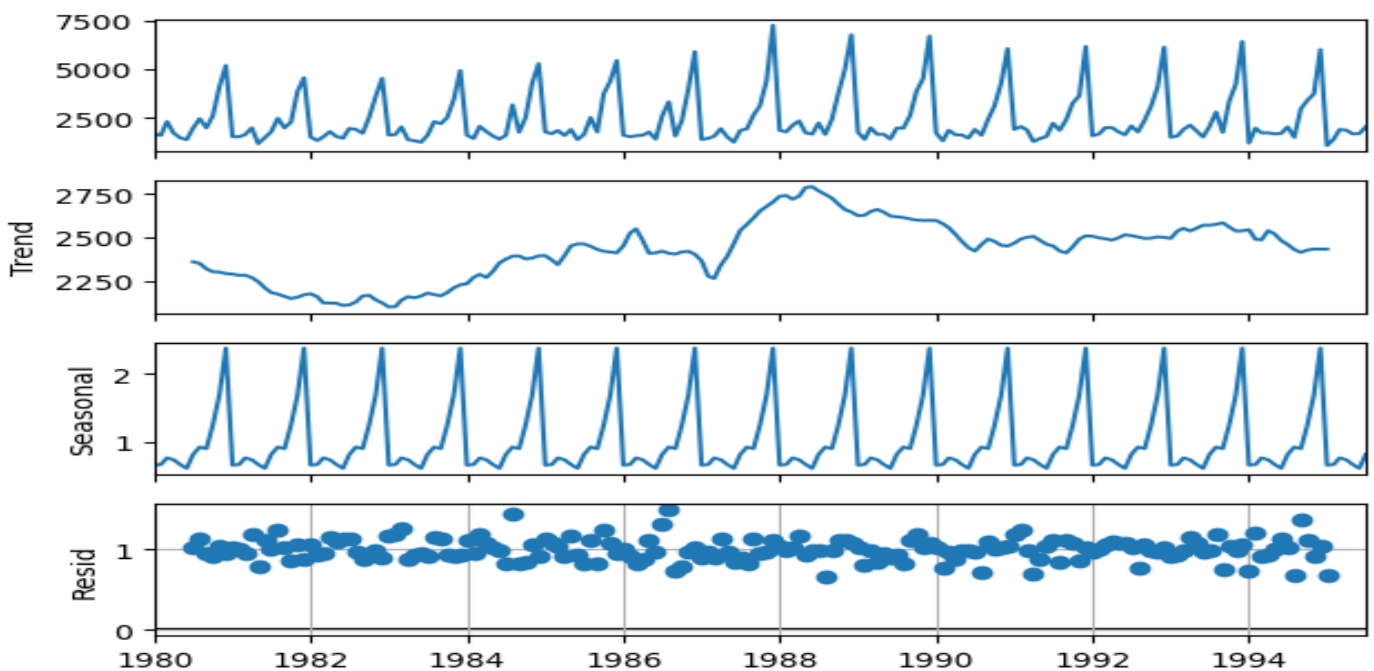


Figure 10 Multiplicative Decomposition

- Since the patten of decomposition is not clear in additive versus multiplicative, plotting the residuals separately:

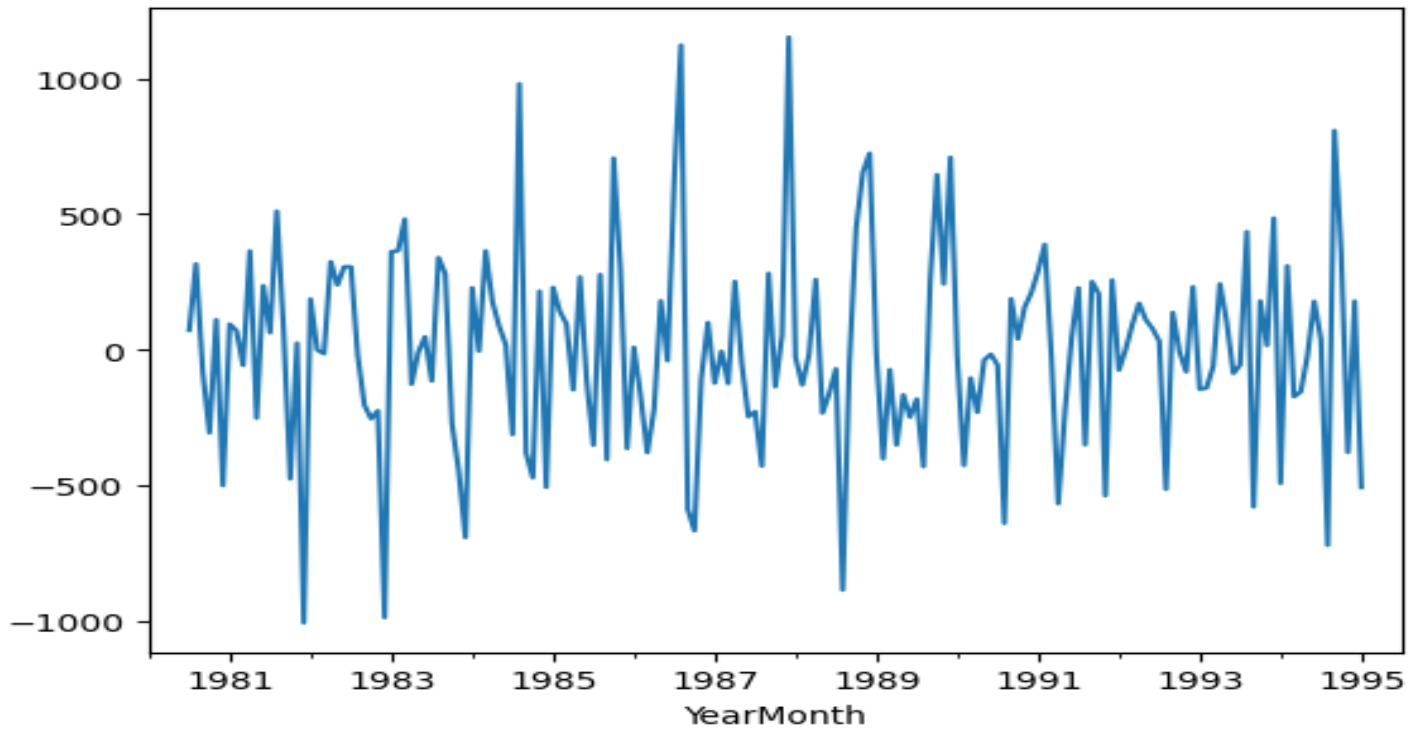


Figure 11 Residuals - Additive decompose

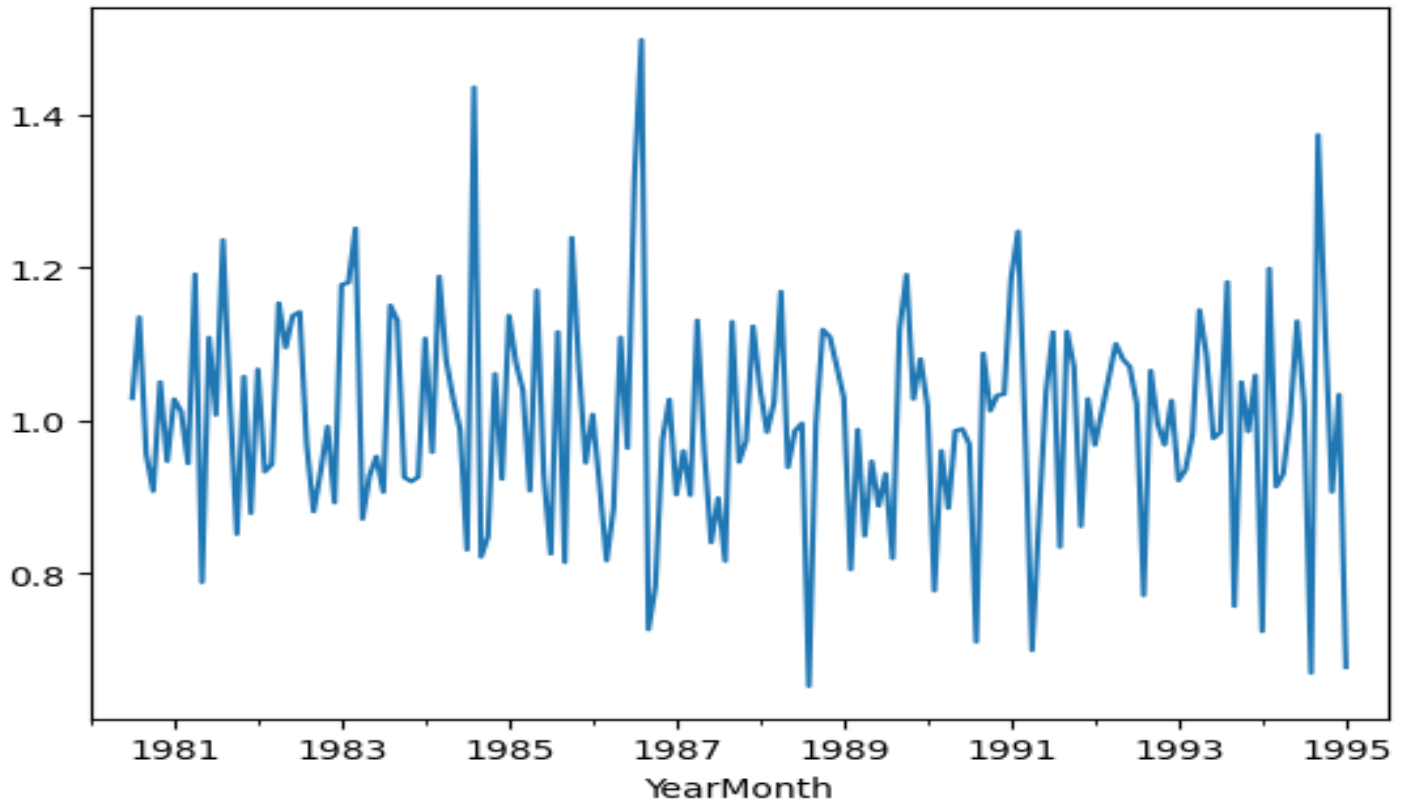


Figure 12 Residuals - Multiplicative decompose

- Since there is no clear pattern in residuals of additive versus multiplicative, additive decomposition is selected.

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

The dataframes are split in the following ways:

Sale volumes for year '1990' and earlier are kept in training dataset and sale volume for year '1991' and later are kept in the test dataset, as is required in the problem.

The same is illustrated in the 4 tables below:

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

Table 2 Sparkling - train data – head

YearMonth	Sparkling
1991-01-01	1902
1991-02-01	2049

<b>1991-03-01</b>	1874
<b>1991-04-01</b>	1279
<b>1991-05-01</b>	1432

Table 3 Sparkling - test data – head

YearMonth	Sparkling
<b>1980-01-01</b>	112
<b>1980-02-01</b>	118
<b>1980-03-01</b>	129
<b>1980-04-01</b>	99
<b>1980-05-01</b>	116

Table 4 Rose - train data - head

YearMonth	Sparkling
<b>1991-01-01</b>	54
<b>1991-02-01</b>	55
<b>1991-03-01</b>	66
<b>1991-04-01</b>	65
<b>1991-05-01</b>	60

Table 5 Rose - test data - head

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.

- As part of the above questions the models built for both the sale data are:
  - Simple Exponential Smoothing
  - Double Exponential Smoothing
  - Triple Exponential Smoothing
  - Linear Regression
  - Naïve Forecast
  - Simple Average
  - Moving Average (2-point, 4-point, 6-point, 9-point)
- RMSE for all these models have been calculated and the forecasts plotted on graphs.
- For the sake of better comparison, graphs from all the models have been consolidated.
- Also, the RMSE values for all the models have been added to a table which is also attached below.
- Plot of sparkling wine sales forecast:

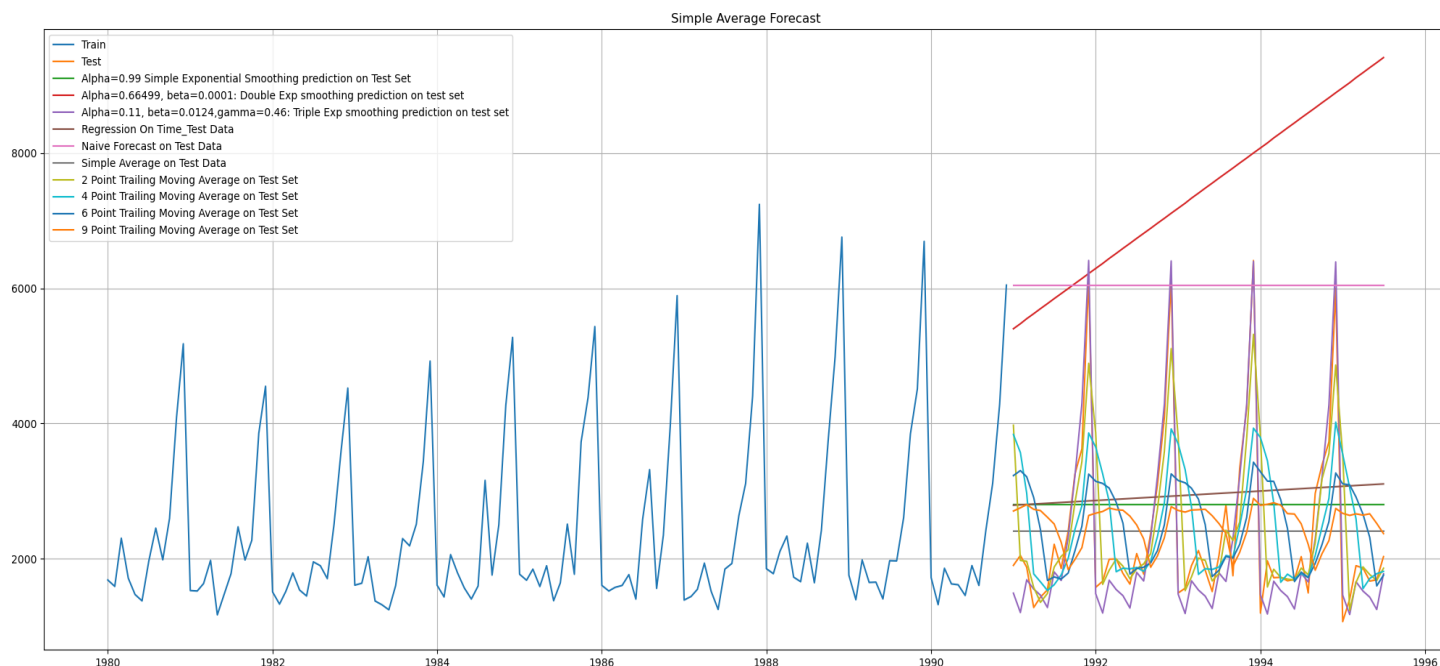


Table 6 Sales forecast - Sparkling Wine - multiple models

- Plot of Rose wine sales forecast:

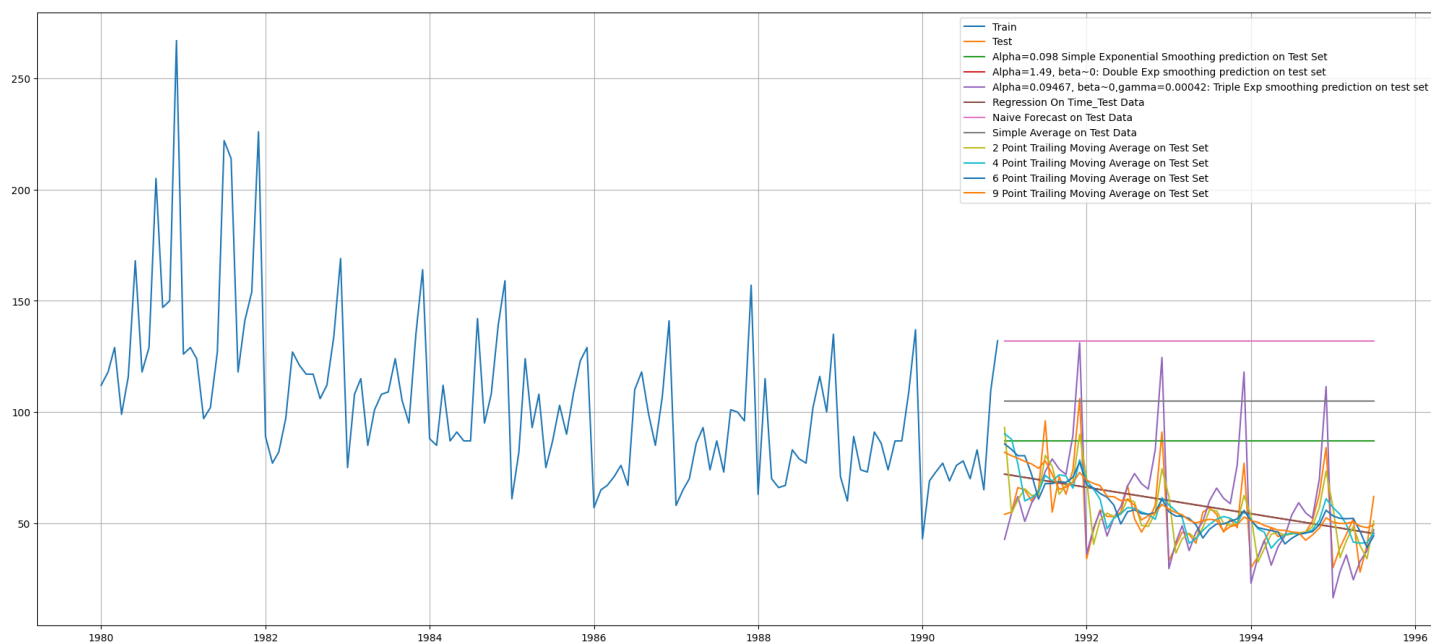


Table 7 Sales Forecast - Rose wine - multiple models

- RMSE values for all the models for Sparkling Wines sales forecast:

<b>Alpha = 0.07028, SES</b>	1338.000861
<b>Alpha=0.66499, beta=0.0001, DES</b>	3262.107333
<b>Alpha=0.11, beta=0.0124, gamma=0.46 TES</b>	1565.666979
<b>LinearRegression</b>	1389.135175
<b>NaiveForecast</b>	3864.279352
<b>SimpleAverage</b>	1275.081804
<b>2pointTrailingMovingAverage</b>	813.400684
<b>4pointTrailingMovingAverage</b>	1156.589694
<b>6pointTrailingMovingAverage</b>	1283.927428
<b>9pointTrailingMovingAverage</b>	1346.278315

*Table 8 RMSE values - sparkling wine – multiple models*

- RMSE values for all the models for Rose Wines sales forecast:

<b>Alpha = 0.0987, SES</b>	36.796236
<b>Alpha=0.66499, beta=0.0001, DES</b>	24.081702
<b>Alpha=0.11, beta=0.0124, gamma=0.46 TES</b>	19.298584
<b>LinearRegression</b>	15.268955
<b>NaiveForecast</b>	79.718773
<b>SimpleAverage</b>	53.46057
<b>2pointTrailingMovingAverage</b>	11.529278
<b>4pointTrailingMovingAverage</b>	14.451403
<b>6pointTrailingMovingAverage</b>	14.566327
<b>9pointTrailingMovingAverage</b>	14.72763

*Table 9 RMSE values - sparkling wine - multiple models*

- In the above tables, the term SES stands for Simple Exponential Smoothing, DES is for Double Exponential Smoothing and TES is for Triple Exponential Smoothing.

5. Check for the stationarity of the data on which the model is being built on using appropriate statistical tests and 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  $\alpha = 0.05$ .

- Graph of Original sale, Rolling mean and rolling standard deviation over a window of 7 months are plotted.

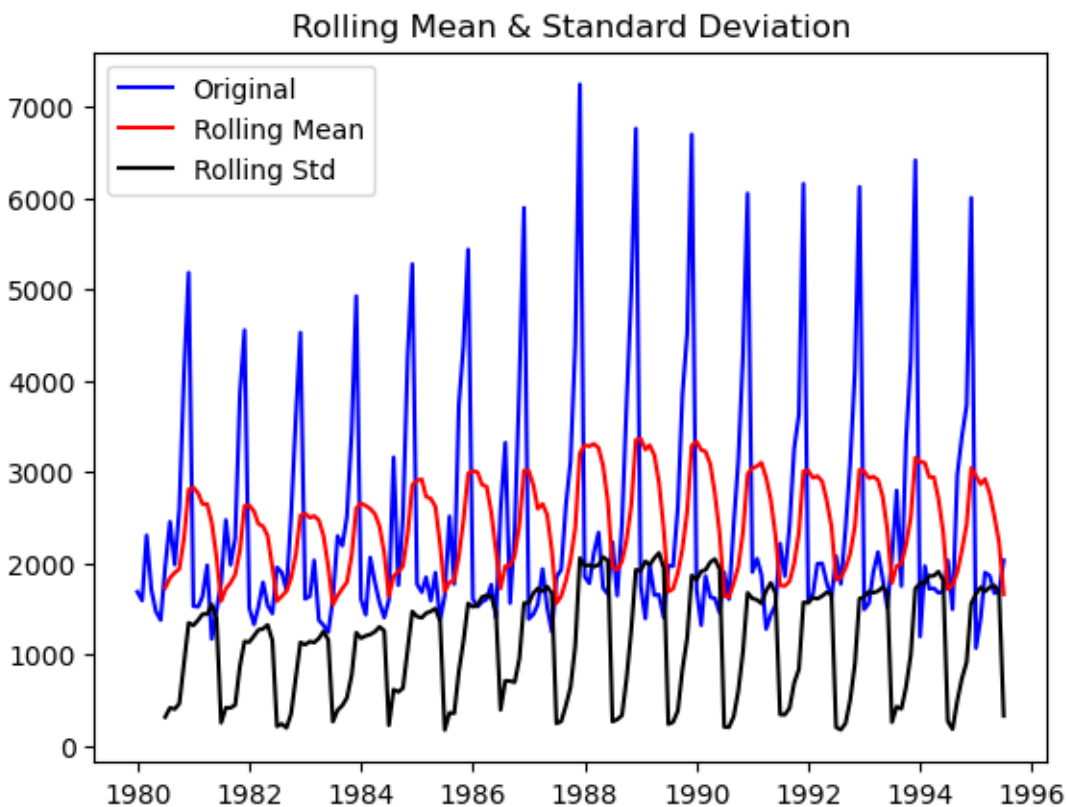


Figure 13 Stationarity test - Sparkling wine

- Results of Dickey-Fuller Test:  
Test Statistic -1.360497  
p-value 0.601061  
#Lags Used 11.000000  
Number of Observations Used 175.000000  
Critical Value (1%) -3.468280  
Critical Value (5%) -2.878202  
Critical Value (10%) -2.575653  
dtype: float64
- In the above, the p-value(0.601) is greater than the expected 'alpha' of 0.05. Hence the null hypothesis of Stationarity is rejected.



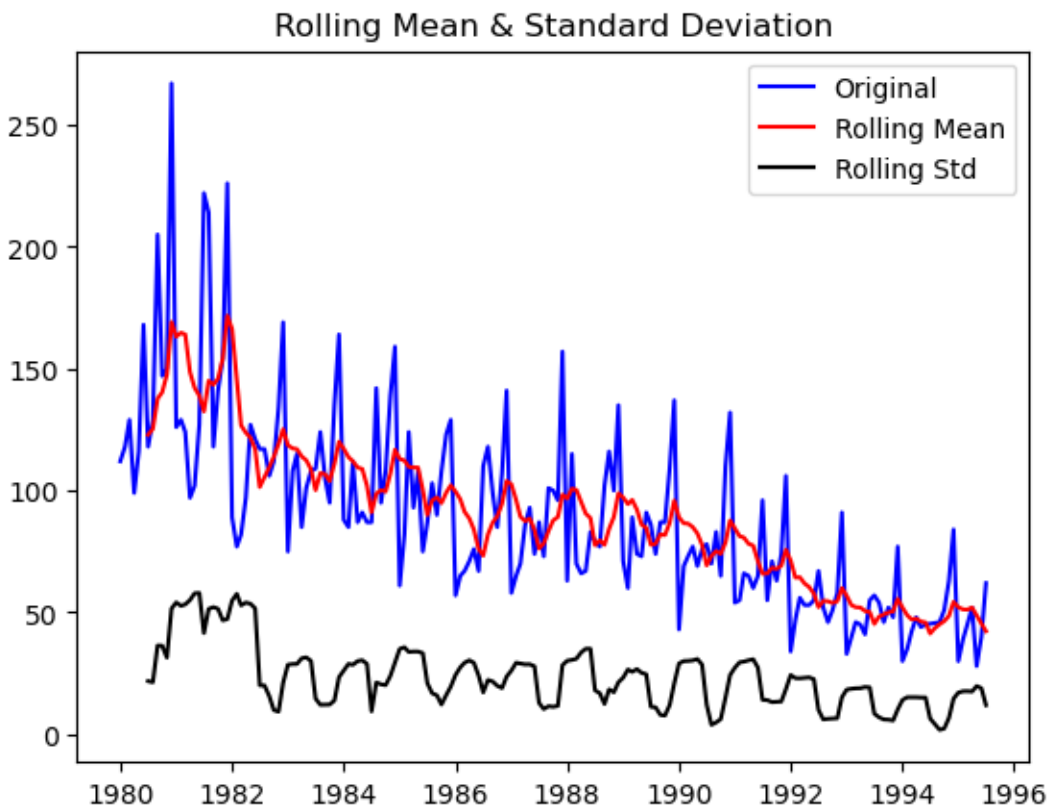


Figure 14 Stationarity test - Rose wine

- 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

- In the above, the p-value (0.343) is greater than the expected 'alpha' of 0.05. Hence the null hypothesis of Stationarity is rejected.
- By Using differencing of the order 1, The stationarity is achieved, as shown below:

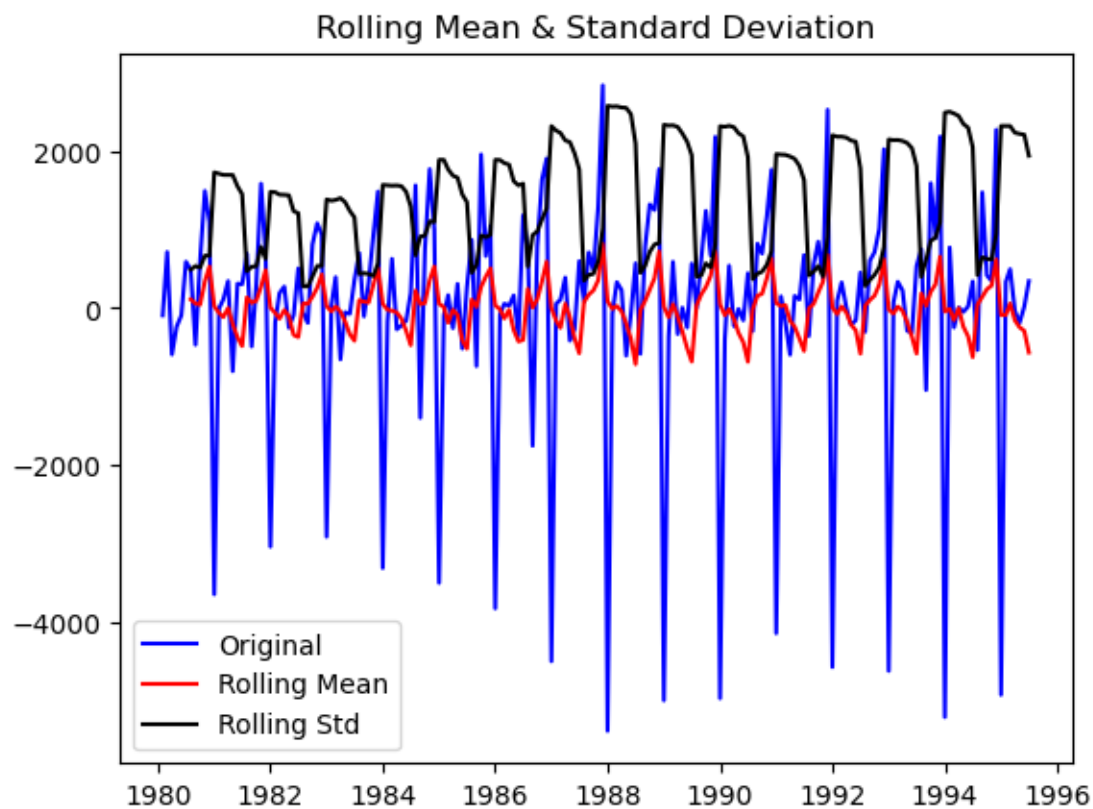


Figure 15 Stationarity achieved - sparkling wine

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

- The p-value is now 0, hence stationarity is achieved.

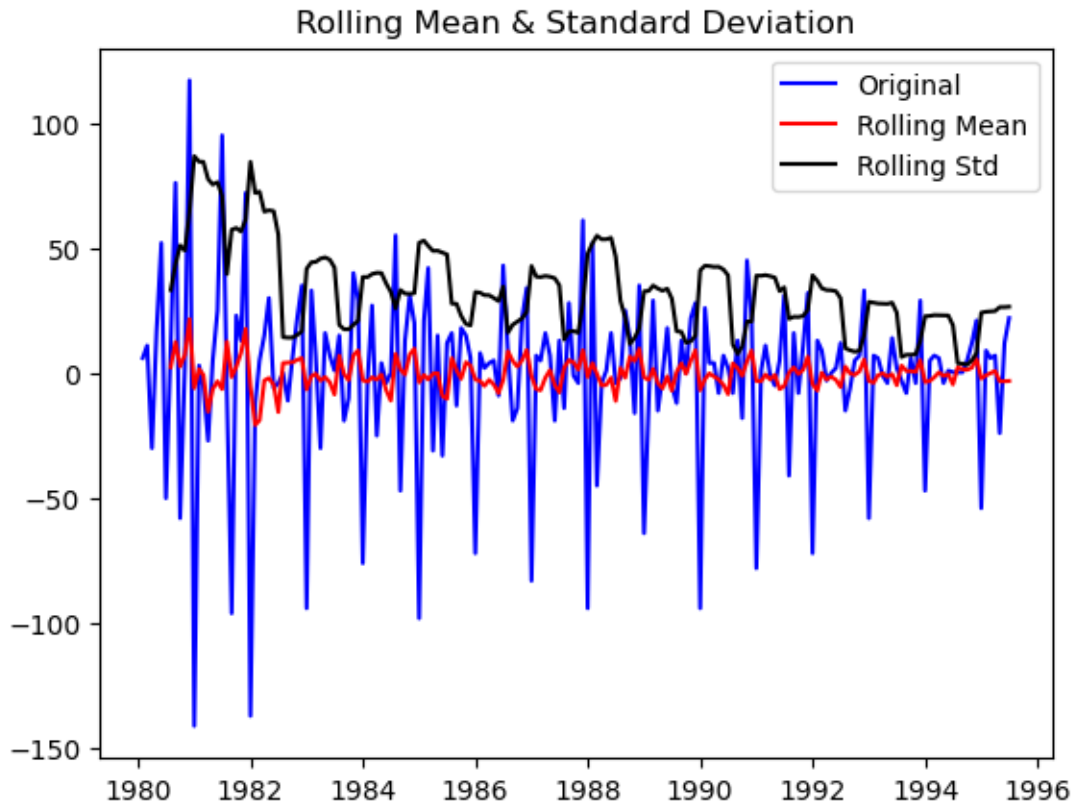


Figure 16 Stationarity achieved - Rose wine

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

- The stationarity is achieved as p value is very close to 0, hence lesser than the alpha value of 0.05.

## 6. Build an automated version of the ARIMA/SARIMA model in which the parameters are selected using the lowest Akaike Information Criteria (AIC) on the training data and evaluate this model on the test data using RMSE.

- ARIMA and SARIMA models were built on both the data sets of Rose wine sales and Sparkling Wine sales. They have been illustrated below:

### SPARKLING WINE:

- The ACF plot for Sparkling wine sales(differentiated):

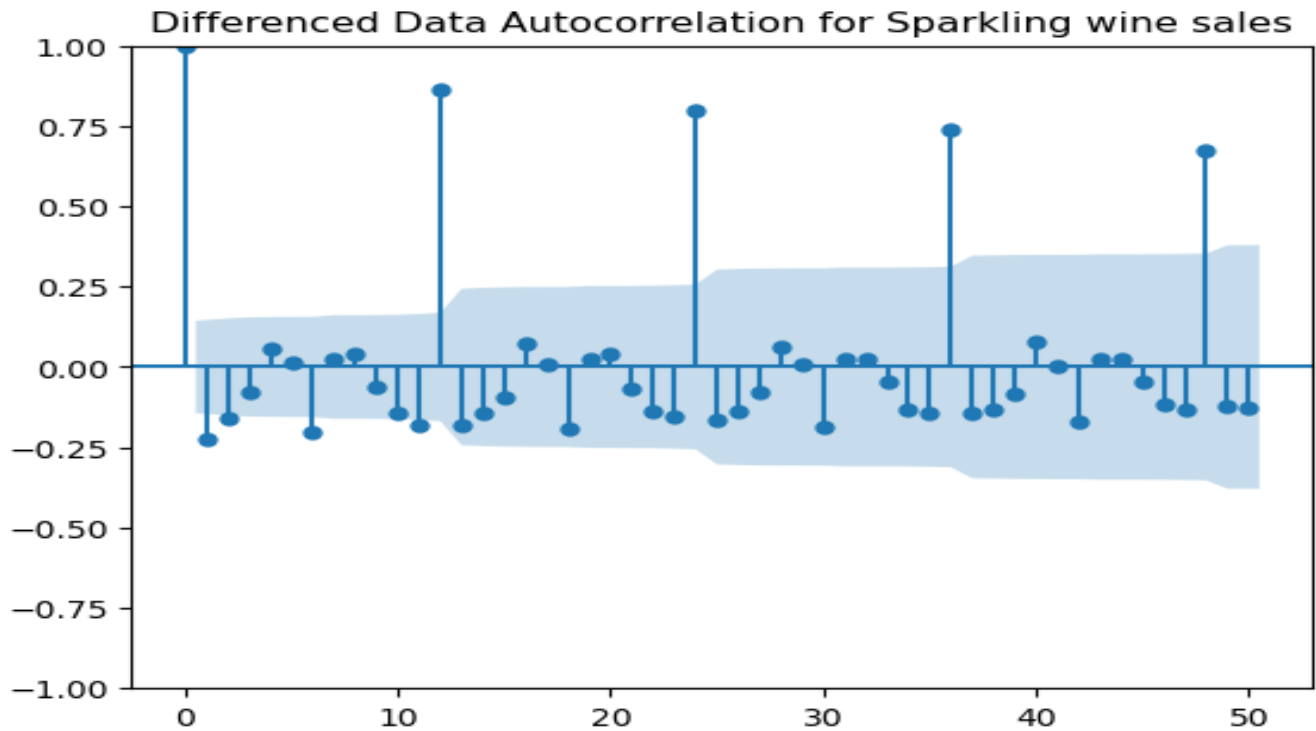


Figure 17 Differenced auto-correlation of Sparkling Wine sales

The AIC values for different parameters are calculated. Following is the list of AIC (top 10 in the ascending order of magnitude of AIC values) for the parameters as part of ARIMA model:

param	AIC
(2, 1, 2)	2213.509212
(2, 2, 2)	2227.341449
(1, 2, 2)	2228.042217
(2, 1, 1)	2233.777626
(0, 1, 2)	2234.408323
(1, 1, 2)	2234.5272
(1, 1, 1)	2235.755095
(2, 0, 1)	2236.59086
(2, 0, 0)	2244.811782
(0, 0, 1)	2245.312136

Table 10 Lowest AIC for ARIMA - sparkling wine

- The summary of the ARIMA model built for sparkling wine:

```

=====
SARIMAX Results
=====
Dep. Variable:          Sparkling      No. Observations:          132
Model:                ARIMA(2, 1, 2)   Log Likelihood            -1101.755
Date:                 Sun, 03 Mar 2024   AIC                      2213.509
Time:                  01:53:04          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      -7.740      0.000       -0.701       -0.418
ma.L1         -1.9917      0.109     -18.216      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
=====

```

- The AIC values for different parameters are calculated. Following is the list of AIC (top 10 in the ascending order of magnitude of AIC values) for the parameters as part of SARIMA model:

param	seasonal	AIC
(0, 2, 2)	(0, 2, 2, 12)	1211.69578
(1, 2, 2)	(0, 2, 2, 12)	1212.339659
(2, 2, 2)	(0, 2, 2, 12)	1214.131024
(1, 2, 2)	(1, 2, 2, 12)	1214.336006
(2, 2, 2)	(1, 2, 2, 12)	1216.086726
(2, 2, 2)	(2, 2, 2, 12)	1218.010733
(0, 2, 2)	(1, 2, 2, 12)	1218.673957
(0, 2, 2)	(2, 2, 2, 12)	1219.651642
(1, 2, 2)	(2, 2, 2, 12)	1220.672175
(0, 1, 2)	(0, 2, 2, 12)	1224.12496

Figure 18 Lowest AIC for SARIMA - sparkling wine

- The summary of the SARIMA model built for sparkling wine:

```

=====
SARIMAX Results
=====
=====
Dep. Variable:          y    No. Observations:
132
Model:          SARIMAX(0, 2, 2)x(0, 2, 2, 12)    Log Likelihood          -60
0.848
Date:              Sun, 03 Mar 2024    AIC          121
1.696
Time:              13:36:52    BIC          122
3.543
Sample:              0    HQIC          121
6.442
                                - 132
Covariance Type:          opg
=====
              coef      std err          z      P>|z|      [0.025      0.975]
-----
ma.L1          -2.0719        1.947        -1.064        0.287        -5.889         1.745
ma.L2           1.0720        2.086         0.514        0.607        -3.016         5.160
ma.S.L12        -2.3924        2.006        -1.193        0.233        -6.324         1.539
ma.S.L24         1.3446        2.412         0.557        0.577        -3.384         6.073
sigma2        6.951e+04    5.96e-05    1.17e+09    0.000        6.95e+04    6.95e+04
=====
Ljung-Box (L1) (Q):          0.74    Jarque-Bera (JB):          10.99
Prob(Q):          0.39    Prob(JB):          0.00
Heteroskedasticity (H):      0.61    Skew:          0.49
Prob(H) (two-sided):      0.22    Kurtosis:          4.54
=====

```

- The RMSE values obtained for the ARIMA and SARIMA models built are:

RMSE\_ARIMA = 1299.9796

RMSE\_SARIMA= 2197.0996

### ROSE WINE:

- ACF plot for ROSE wine sales(differentiated):

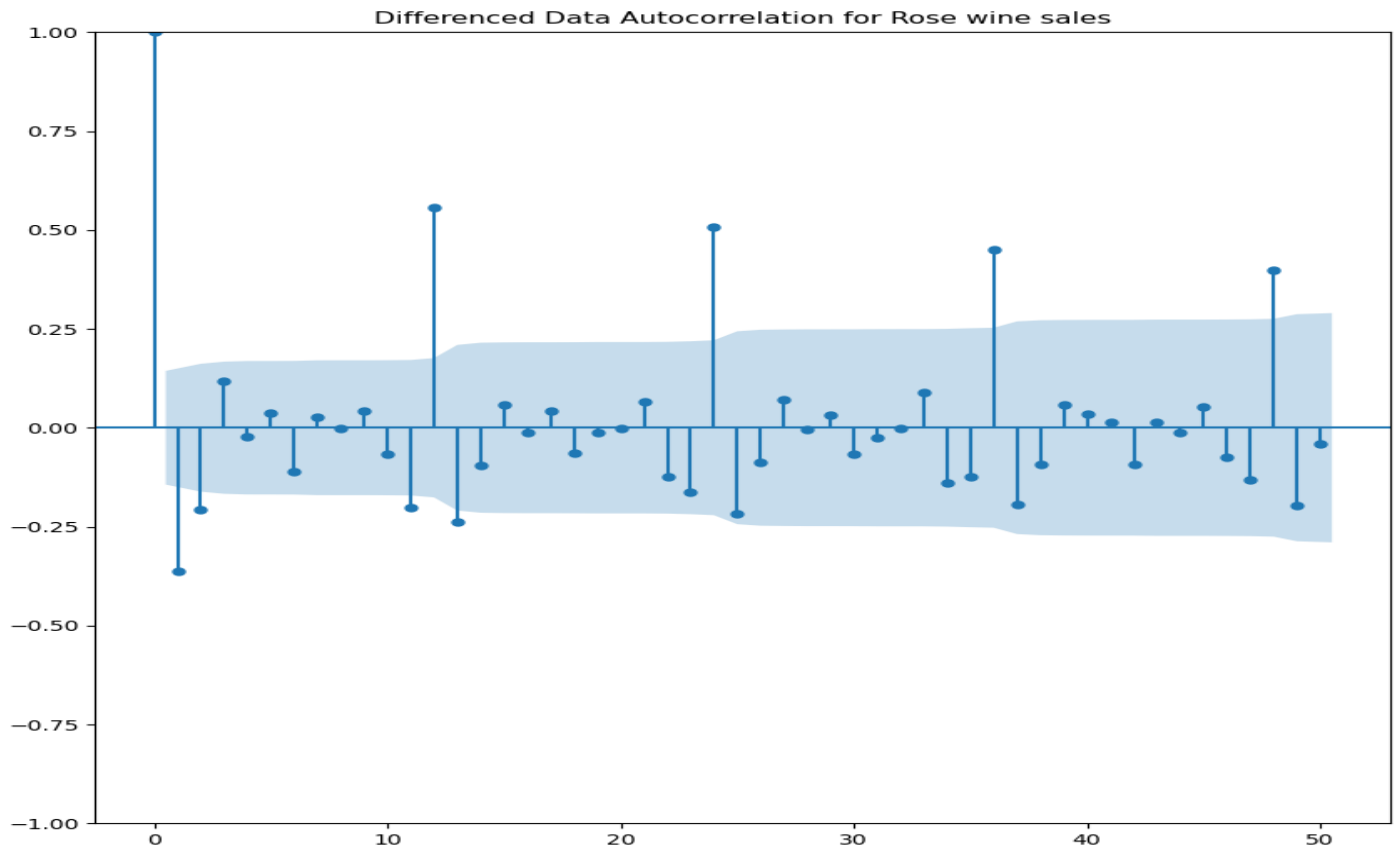


Figure 19 Differenced auto-correlation of Rose Wine sales

- Seasonality=12
- The AIC values for different parameters are calculated. Following is the list of AIC (top 10 in the ascending order of magnitude of AIC values) for the parameters as part of ARIMA model:

param	AIC
(1, 0, 2)	1292.053213
(2, 0, 2)	1292.248056
(2, 0, 1)	1292.937195
(1, 0, 1)	1294.510585
(1, 0, 0)	1301.546304
(2, 0, 0)	1302.347685
(0, 0, 1)	1305.468406
(0, 0, 2)	1306.587015
(0, 0, 0)	1324.899703

Table 11 Lowest AIC for ARIMA - rose wine

- The summary of the ARIMA model built for sparkling wine:

```

=====
SARIMAX Results
=====
Dep. Variable:          Rose      No. Observations:          132
Model:                ARIMA(1, 0, 2)  Log Likelihood          -641.027
Date:                Sun, 03 Mar 2024  AIC              1292.053
Time:                13:59:43      BIC              1306.467
Sample:              01-01-1980    HQIC             1297.910
                  - 12-01-1990
Covariance Type:          opg
=====
              coef      std err          z      P>|z|      [0.025      0.975]
-----
const          107.8405      24.779      4.352      0.000      59.275     156.406
ar.L1           0.9861       0.027     36.818      0.000       0.934       1.039
ma.L1          -0.6874       0.090     -7.622      0.000      -0.864      -0.511
ma.L2          -0.2007       0.093     -2.148     0.032      -0.384      -0.018
sigma2         960.8592     100.353      9.575      0.000     764.172    1157.547
=====
Ljung-Box (L1) (Q):                0.05      Jarque-Bera (JB):                58.48
Prob(Q):                0.82      Prob(JB):                0.00
Heteroskedasticity (H):            0.36      Skew:                0.98
Prob(H) (two-sided):            0.00      Kurtosis:              5.61
=====

```

- The AIC values for different parameters are calculated. Following is the list of AIC (top 10 in the ascending order of magnitude of AIC values) for the parameters as part of SARIMA model:

param	seasonal	AIC
(0, 1, 2)	(2, 2, 2, 12)	716.792983
(0, 2, 2)	(1, 2, 2, 12)	717.90783
(0, 1, 2)	(0, 2, 2, 12)	718.35098
(1, 1, 2)	(2, 2, 2, 12)	718.768293
(1, 2, 2)	(2, 2, 2, 12)	719.164782
(1, 2, 2)	(1, 2, 2, 12)	719.411002
(1, 1, 2)	(0, 2, 2, 12)	720.276054
(0, 1, 2)	(1, 2, 2, 12)	720.342735
(2, 1, 2)	(0, 2, 2, 12)	720.662535
(2, 2, 2)	(2, 2, 2, 12)	721.013902

Table 12 Lowest AIC for SARIMA - rose wine

- The summary of the SARIMA model built for sparkling wine:



# SARIMAX Results

```

=====
=====
Dep. Variable:          y      No. Observations:
132
Model:                SARIMAX(0, 1, 2)x(2, 2, 2, 12)      Log Likelihood      -35
1.396
Date:                  Sun, 03 Mar 2024      AIC      71
6.793
Time:                  14:19:53      BIC      73
3.467
Sample:                0      HQIC      72
3.478
                                - 132
Covariance Type:      opg
=====
=====
              coef      std err          z      P>|z|      [0.025      0.975]
-----
ma.L1          -0.9186    1374.651     -0.001     0.999    -2695.185    2693.347
ma.L2          -0.0814     111.877     -0.001     0.999    -219.356     219.193
ar.S.L12       -0.4301       0.197     -2.180     0.029      -0.817     -0.043
ar.S.L24       -0.2034       0.097     -2.099     0.036      -0.393     -0.013
ma.S.L12       -0.9798    1374.663     -0.001     0.999    -2695.271    2693.311
ma.S.L24       -0.0202      27.717     -0.001     0.999      -54.344     54.304
sigma2         274.1713      5.510     49.759     0.000      263.372    284.971
=====
Ljung-Box (L1) (Q):          0.01      Jarque-Bera (JB):          5.21
Prob(Q):                    0.92      Prob(JB):          0.07
Heteroskedasticity (H):      0.67      Skew:          -0.23
Prob(H) (two-sided):        0.30      Kurtosis:          4.16
=====

```

- The RMSE values obtained for the ARIMA and SARIMA models built are:

RMSE\_ARIMA = 45.438584

RMSE\_SARIMA= 34.134463

- For both the models: ARIMA and SARIMA on the datasets: Sparkling wine sales and Rose wine sales have lower RMSE values than the **2pointTrailingMovingAverage** model built.

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

- RMSE values for different models built on sparkling wine sales data:

	Test RMSE
<b>Alpha = 0.07028, SES</b>	1338.000861
<b>Alpha=0.66499, beta=0.0001, DES</b>	3262.107333
<b>Alpha=0.11, beta=0.0124,gamma=0.46 TES</b>	1565.666979
<b>LinearRegression</b>	1389.135175
<b>NaiveForecast</b>	3864.279352
<b>SimpleAverage</b>	1275.081804
<b>2pointTrailingMovingAverage</b>	813.400684
<b>4pointTrailingMovingAverage</b>	1156.589694
<b>6pointTrailingMovingAverage</b>	1283.927428
<b>9pointTrailingMovingAverage</b>	1346.278315
<b>ARMA</b>	1269.345658
<b>ARIMA</b>	1299.979592
<b>SARIMAX(0, 2, 2)x(0, 2, 2, 12)</b>	2197.099578

Table 13 RMSE values for models - sparkling wine

- RMSE values for different models built on rose wine sales data:

	Test RMSE
<b>Alpha = 0.0987, SES</b>	36.796236
<b>Alpha=0.66499, beta=0.0001, DES</b>	24.081702
<b>Alpha=0.11, beta=0.0124,gamma=0.46 TES</b>	19.298584
<b>LinearRegression</b>	15.268955
<b>NaiveForecast</b>	79.718773
<b>SimpleAverage</b>	53.46057
<b>2pointTrailingMovingAverage</b>	11.529278
<b>4pointTrailingMovingAverage</b>	14.451403
<b>6pointTrailingMovingAverage</b>	14.566327
<b>9pointTrailingMovingAverage</b>	14.72763
<b>ARIMA</b>	45.438584
<b>SARIMAX(0, 1, 2)x(2, 2, 2, 12)</b>	34.134463

Table 14 RMSE values for models - rose wine

## 8. 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.

- For both Sparkling and Rose wine sales prediction, SARIMAX models are built.
- Steps of 12, one for each monthly sales prediction is made. The same is displayed in sales predictions tables below in the current section.
- The RMSE values for full data set obtained are:
  - RMSE\_sparkling\_wine = 764.844
  - RMSE\_rose\_wine = 51.54
- Sparkling wine sales predictions plot 12 months in future:

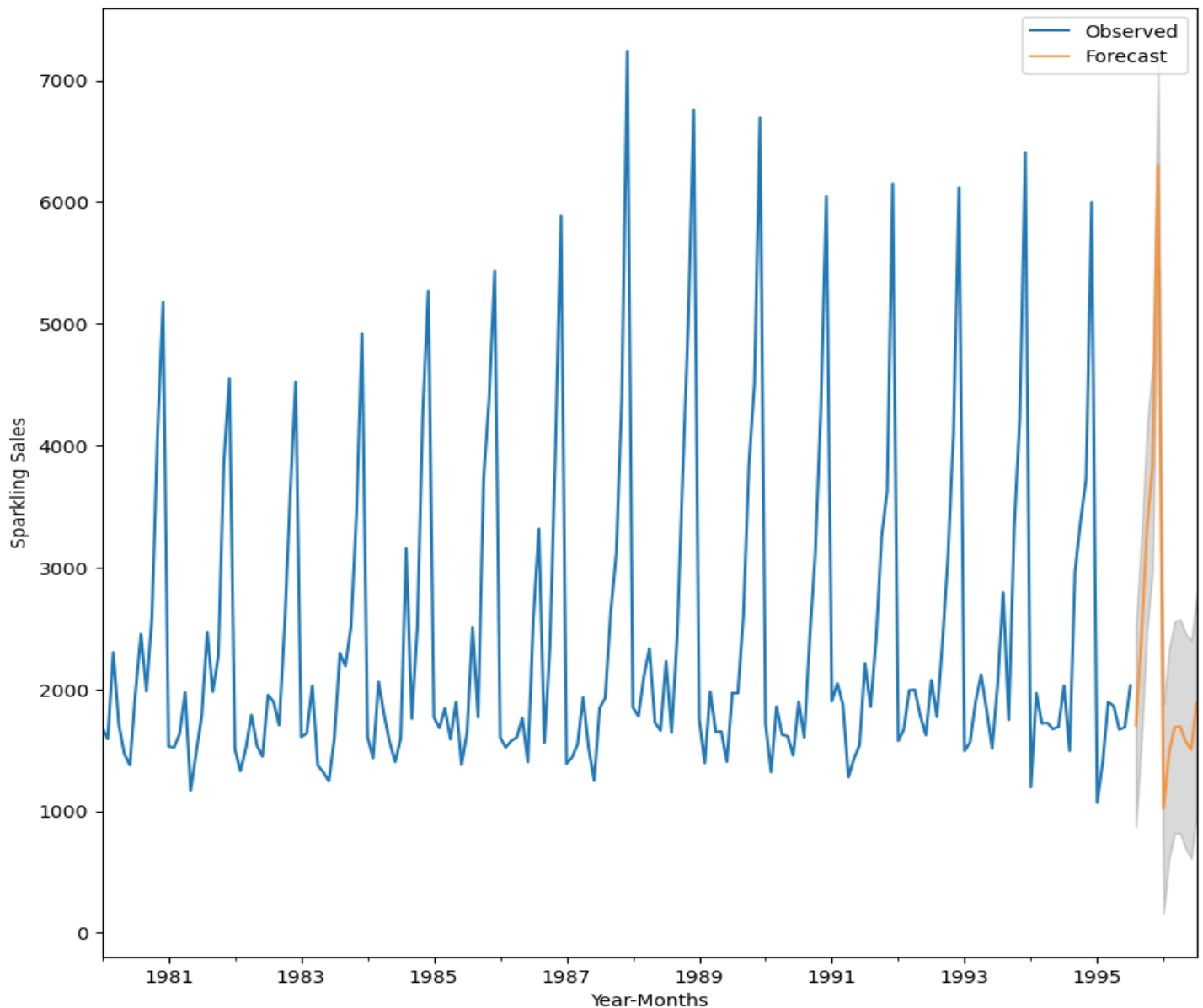


Figure 20 Sales prediction - sparkling wine

- Sparkling Wine sales prediction values:

Year-Month	Sales
1995-08-01	1702.187078
1995-09-01	2436.456159
1995-10-01	3319.777504
1995-11-01	3829.67655
1995-12-01	6302.520094
1996-01-01	1019.743677
1996-02-01	1477.793476
1996-03-01	1689.56455
1996-04-01	1696.212868
1996-05-01	1571.851554
1996-06-01	1505.378744
1996-07-01	1892.590185

Table 15 Sales prediction values - Sparkling wine

- Rose wine sales predictions 12 months in future:

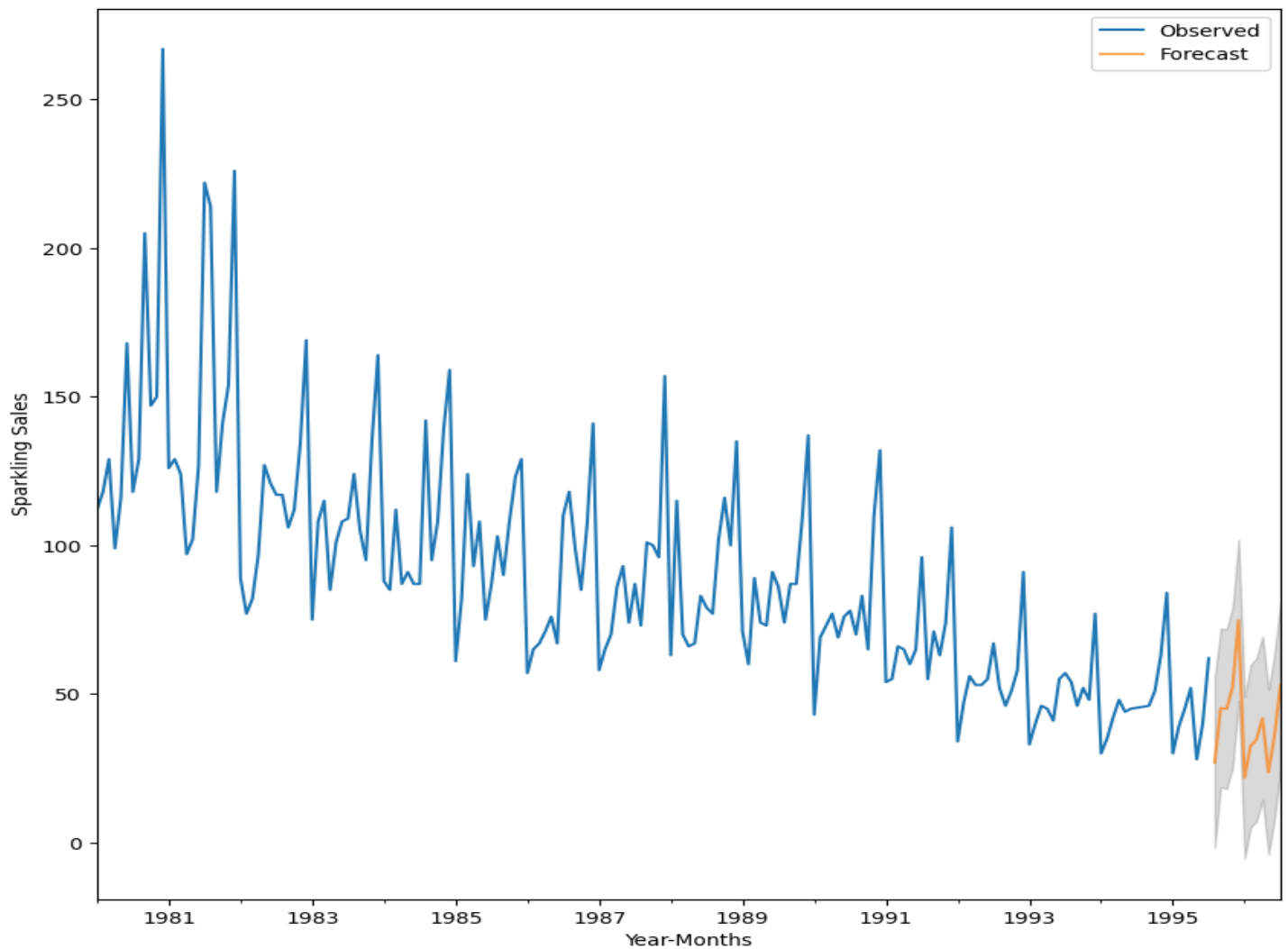


Figure 21 Sales prediction - rose wine

- Rose Wine sales prediction values:

Year-Month	Sales
1995-08-01	26.861759
1995-09-01	45.258635
1995-10-01	44.916695
1995-11-01	52.200499
1995-12-01	74.819353
1996-01-01	21.79635
1996-02-01	32.435196
1996-03-01	34.616086
1996-04-01	41.801689
1996-05-01	23.655514
1996-06-01	35.724278
1996-07-01	52.792439

Table 16 Sales prediction values - Rose wine

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

- Monthly sales figures for sparkling wine and rose wine from the year 1980 till the mid of 1995 were provided in 2 data sets.
- Exploratory data analysis on both the data sets were performed. Interpolation was performed wherever necessary to complete the data when missing.
- The sales figures for both the wines were then plotted separately year-wise and month-wise, to identify the pattern of sales.
- Moreover, trend, seasonality and residue, if any, were also plotted in separate graphs to identify these attributes, along with identification of stationarity and order of differentiation.
- Multiple models were built, and their respective accuracies were then listed in tables to show the accuracy of each model that would aid in deciding which model would be the best fit.
- Even though the 2-point trailing average was found to be having the least value of RMSE, since it is a trailing average, SARIMA model was chosen above it to make the predictions.
- This was found to be the correct decision, as the RMSE using SARIMA for the whole dataset was then found to be better than observed earlier.
- There was trend and seasonality present in both the data sets.

**Sparkling Wine:**

- It was observed that for sparkling wine, the sales spike up in the last quarter of each year.
- In the months: November and December, the recent years have seen a decline in monthly sales. Since these are the 2 most important months in terms of sale of sparkling wine, proper investigation is required to identify the downfall in sales in the last 2 years.
- In the early months of the year, the sales figure is traditionally the lowest whereas the previous year-end has been the best. Hence, running marketing campaign for sparkling wine should continue to the next year as well.
- Measures need to be taken to ensure adequate supply of sparkling wine in view of the expected increased demand of the wine in the last quarter, as shown in the sales predictions.

**Rose Wine:**

- The sale of Rose wine has seen a consistent decline over the years.
- The decline has been sharper in the last quarter of the years.
- However, the sale has seen a slight uptick in the last 2 years over several months. Hence, it is important to ensure the steps that have been taken recently to improve the sale of rose wine are continued.
- Sale prediction for January – 1996 is expected to drop to one-third from that of December -1995. Hence, measures need to be taken to arrest this steep fall: marketing campaign for rose wine needs to be continued even after the year-end to help better sales.