TIME SERIES FORECASTING

[Document subtitle]

SEPTEMBER 24, 2021

VARSHA BALAJI

Problem:

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.

First we import the required libraries and load the dataset

```
dfs.head()
```

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

```
dfs.tail()
```

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

```
dfr.head()
```

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

```
dfr.tail()
```

	YearMonth	Rose
182	1995-03	45.0
183	1995-04	52.0
184	1995-05	28.0
185	1995-06	40.0
186	1995-07	62.0

Creating time stamp and adding to the dataframe to make it a time series data

```
DatetimeIndex(['1980-01-31', '1980-02-29', '1980-03-31', '1980-04-30',
               '1980-05-31', '1980-06-30', '1980-07-31', '1980-08-31',
               '1980-09-30', '1980-10-31',
               ...
               '1994-10-31', '1994-11-30', '1994-12-31', '1995-01-31',
               '1995-02-28', '1995-03-31', '1995-04-30', '1995-05-31',
               '1995-06-30', '1995-07-31'],
              dtype='datetime64[ns]', length=187, freq='M')
```

Creating a combined dataframe

	Sparkling	Rose
YearMonth		
1980-01-31	1686	112.0
1980-02-29	1591	118.0
1980-03-31	2304	129.0
1980-04-30	1712	99.0
1980-05-31	1471	116.0

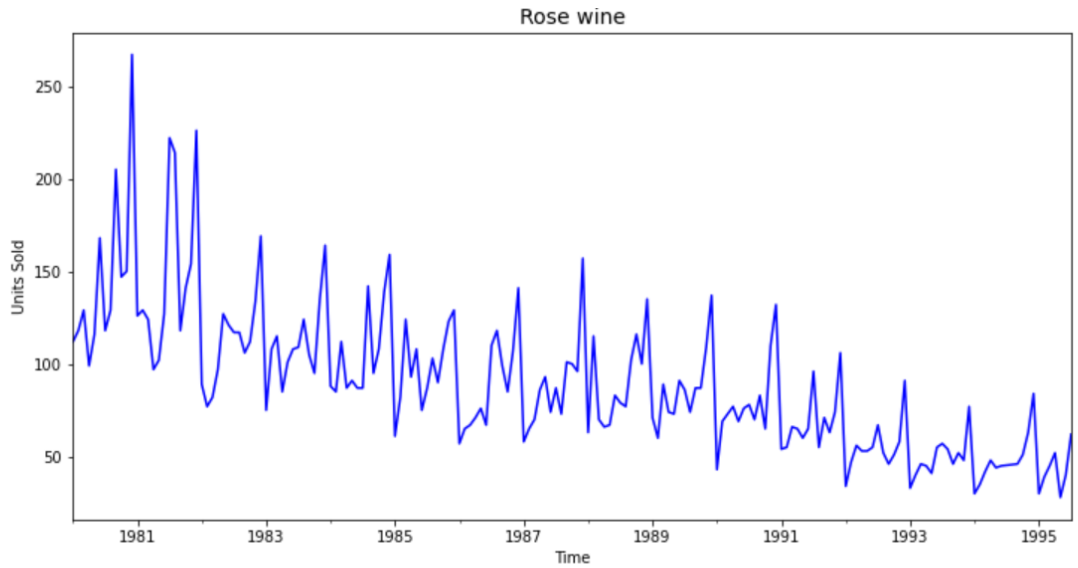
Checking for null values

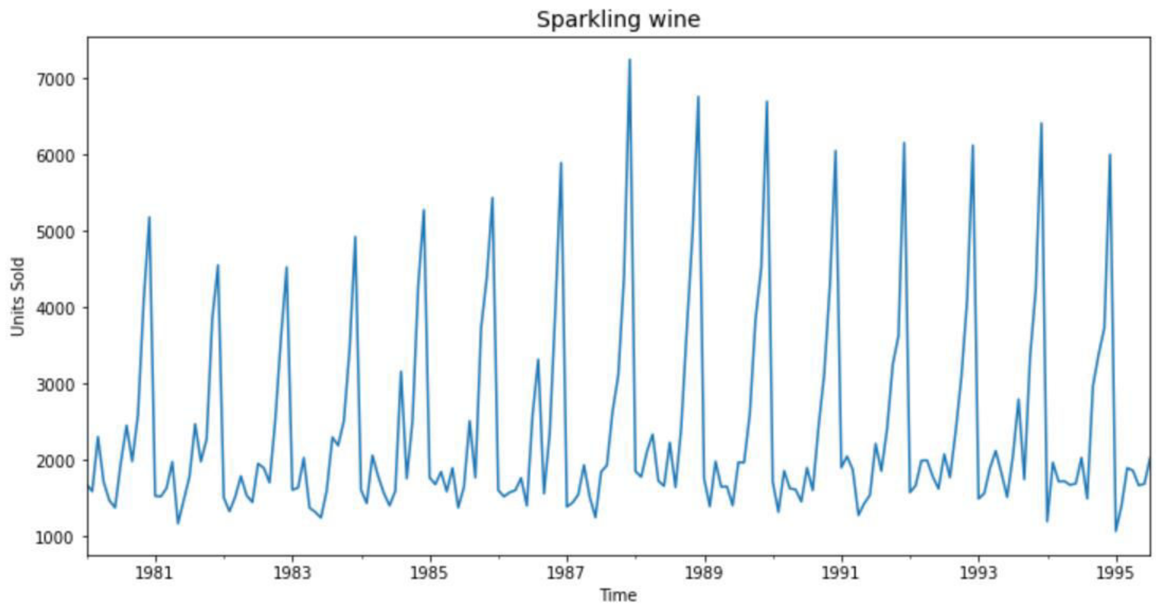
```
Sparkling    0
Rose         2
dtype: int64
```

Imputing the missing values

```
YearMonth
1994-01-31    30.000000
1994-02-28    35.000000
1994-03-31    42.000000
1994-04-30    48.000000
1994-05-31    44.000000
1994-06-30    45.000000
1994-07-31    45.336957
1994-08-31    45.673913
1994-09-30    46.000000
1994-10-31    51.000000
1994-11-30    63.000000
1994-12-31    84.000000
Name: Rose, dtype: float64
```

Plot for rose and sparkling wine





Both the wines show seasonality but rose wine shows a downward trend.

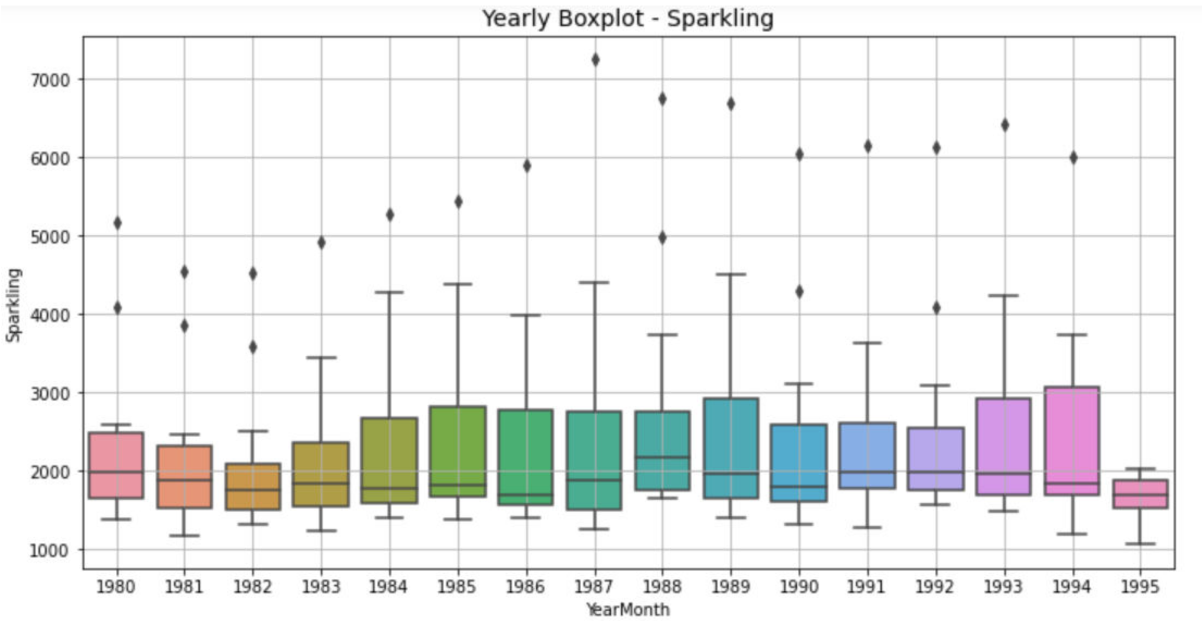
While sparkling wine has been consistently favoured over the years, the demand for rose has been out-of-favour over the years.

<u>Exploratory data analysis</u>

	Sparkling	Rose
count	187.000000	185.000000
mean	2402.417112	90.394595
std	1295.111540	39.175344
min	1070.000000	28.000000
25%	1605.000000	63.000000
50%	1874.000000	86.000000
75%	2549.000000	112.000000
max	7242.000000	267.000000

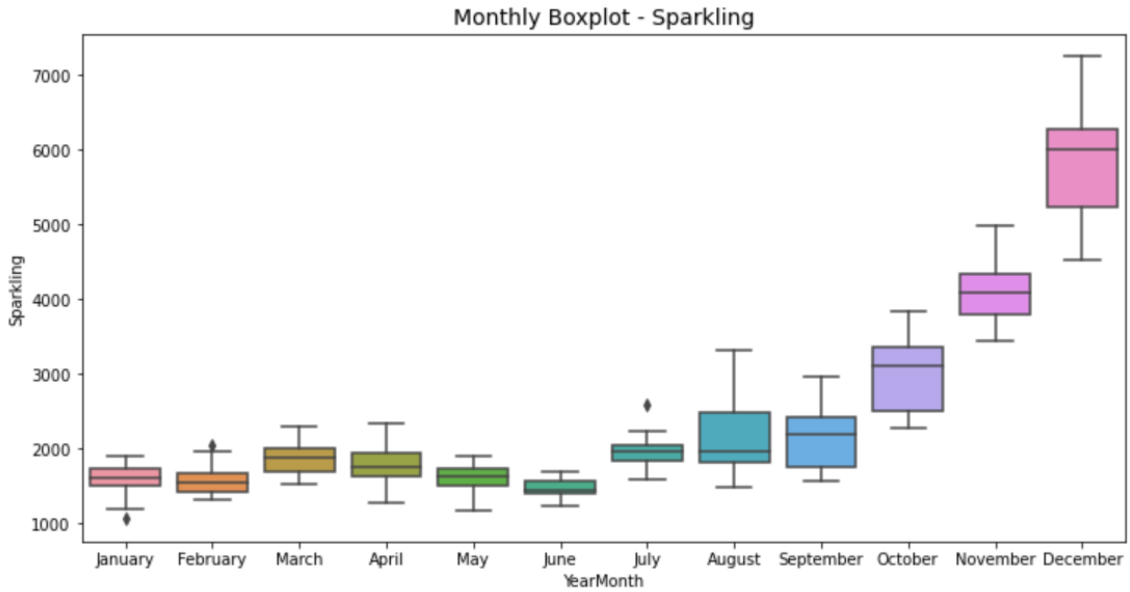
The descriptive summary of the data shows that on an average 2402 units of Sparkling wines were sold each month on the given period. 50% of months sales varied from 1605 units to 2549 units. Maximum sale reported in a month is 7242 units.

SPARKLING

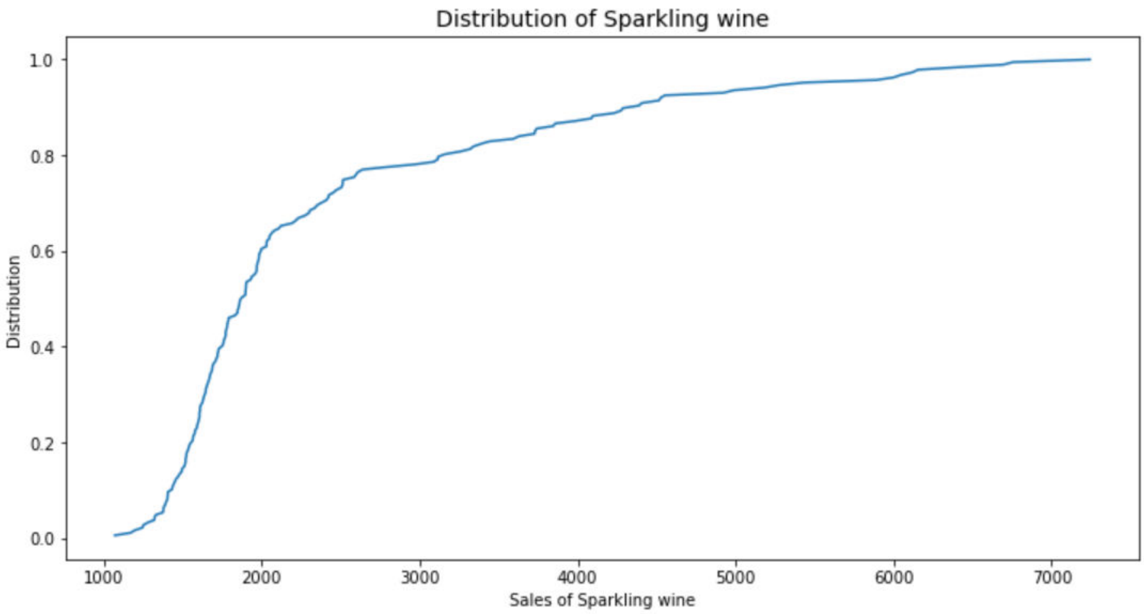


The yearly-boxplot, shows that the average sale of Sparkling has been more or less consistent across the period, at or a little below 2000 units.

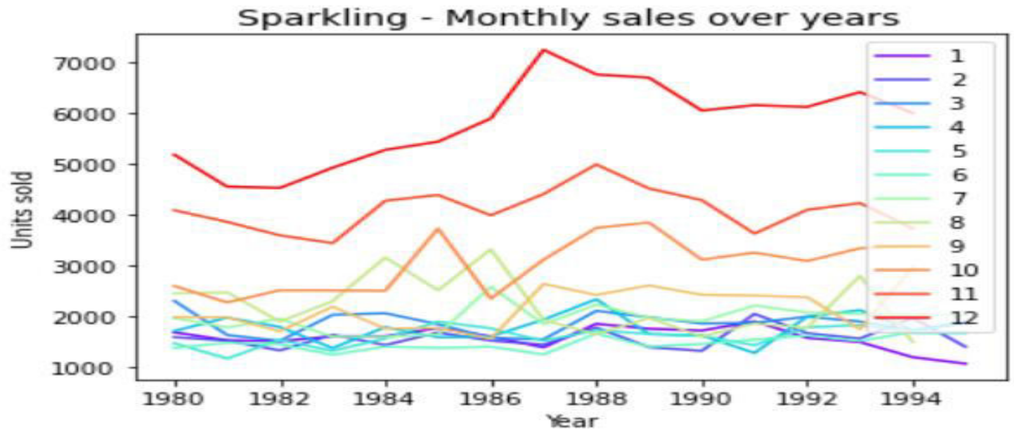
The outliers in the yearly-boxplot most probably represent the seasonal sale during the seasonal months



The monthly-box-plot shows a clear seasonality during the festive seasonal months of October, November and December, which peaks in December. The sale tanks in the month of June.



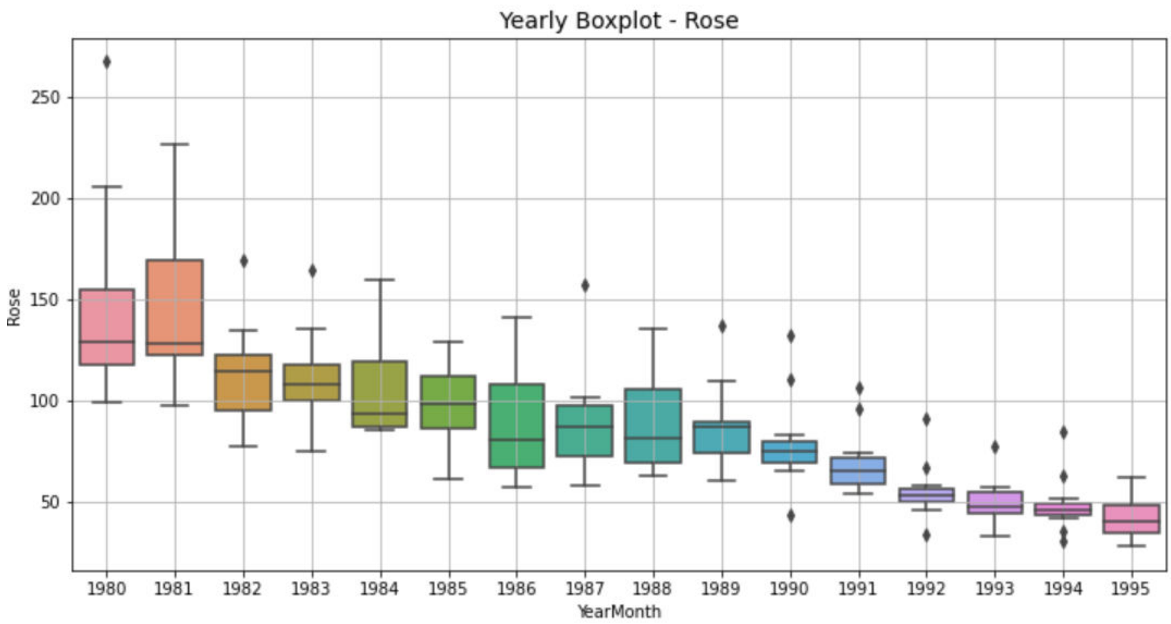
The plot shows that, in 80% of months, at least 3000 units of Sparkling wine were sold.



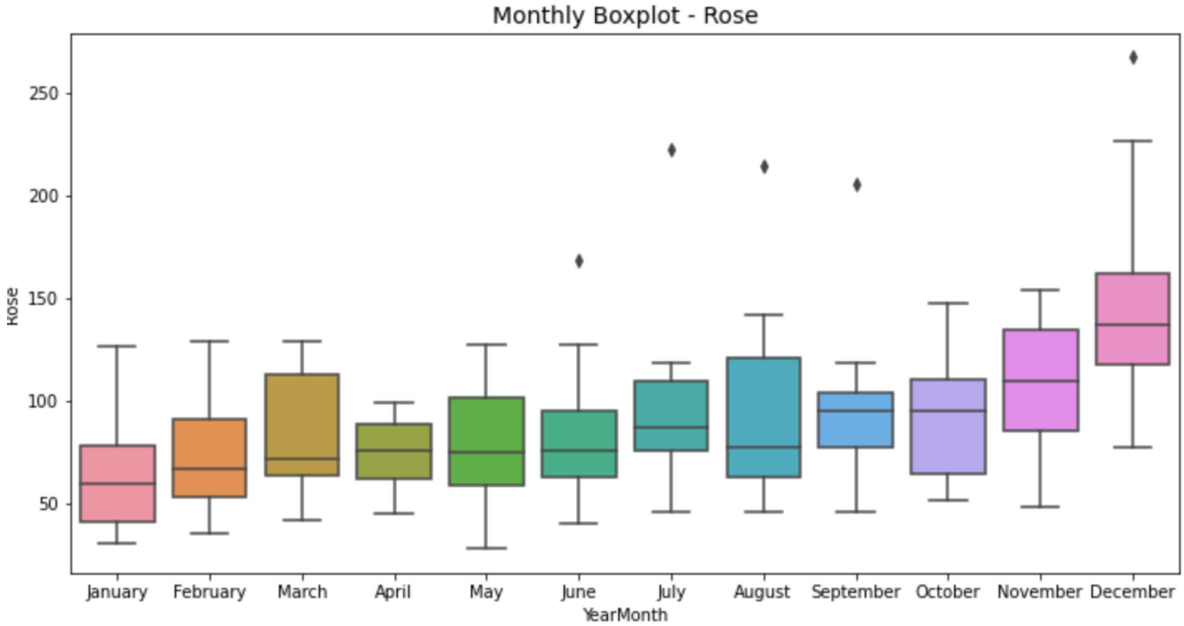
The plot of monthly sale over the years also shows the seasonality component of the time-series, with October November and December selling exponentially higher volumes.

The highest volume of Sparkling wines was sold in December,1987 and the least of December sale was in 1981. Post 1987 December sales is around an average 6500 units, which was around 5000 in early 80’s

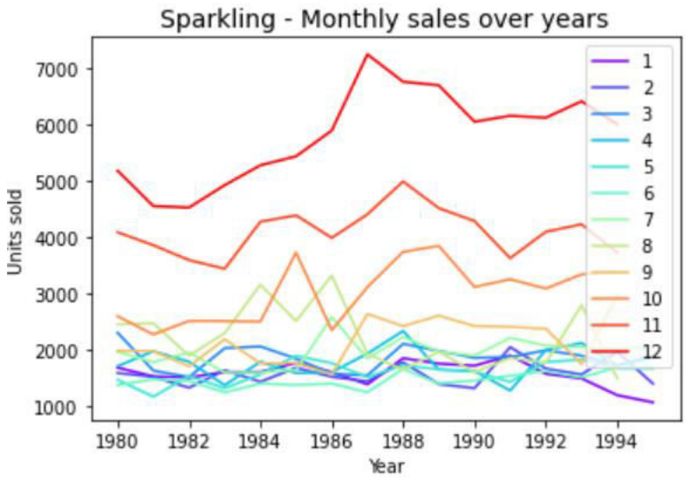
ROSE



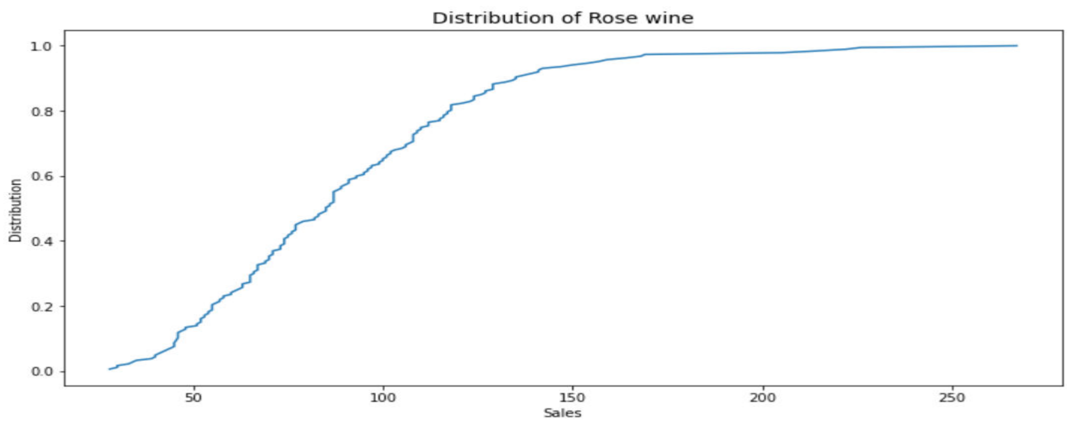
The yearly-boxplot, shows that the average sale of Rose wine moving according to the downward trend in sales over the years. The outliers over upperbound in the yearly-boxplot most probably represent the seasonal sale during the seasonal months.



The monthly-box-plot shows a clear seasonality during the seasonal months of November and December. Though the sale tanks in the month of January, it picks up in the due course of the year.



The plot of monthly sale over the years also shows the seasonality component of the time-series, with November and December selling exponentially higher volumes than other months.

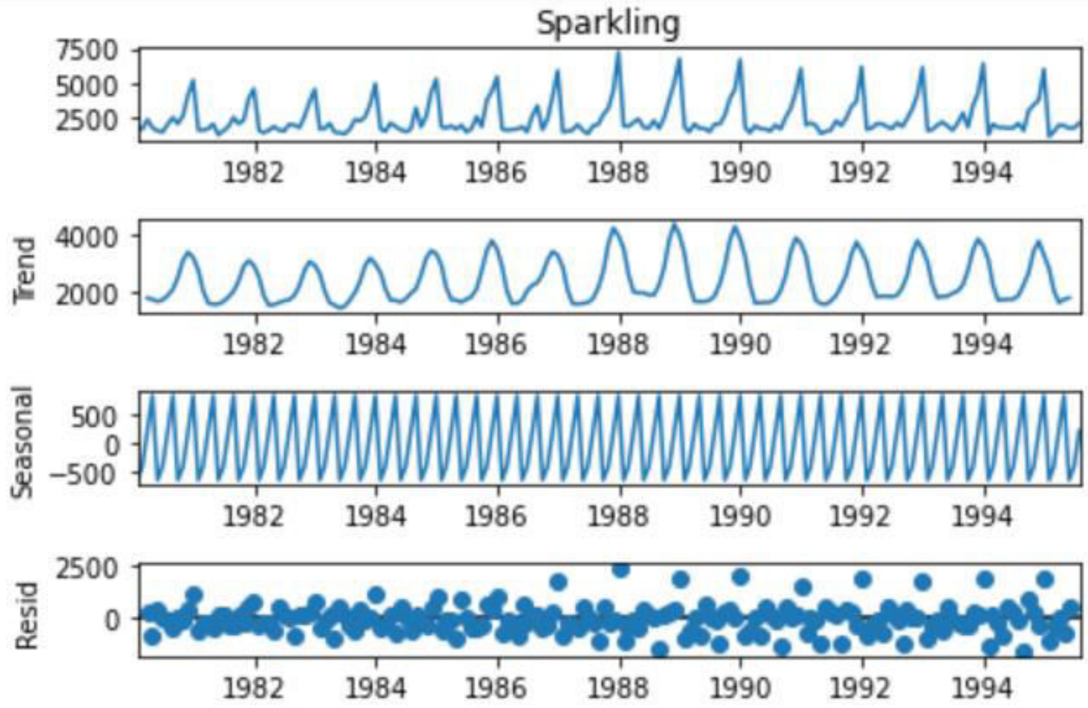


The plot shows that, in 80% of months, at least 120 units of Rose wine were sold.

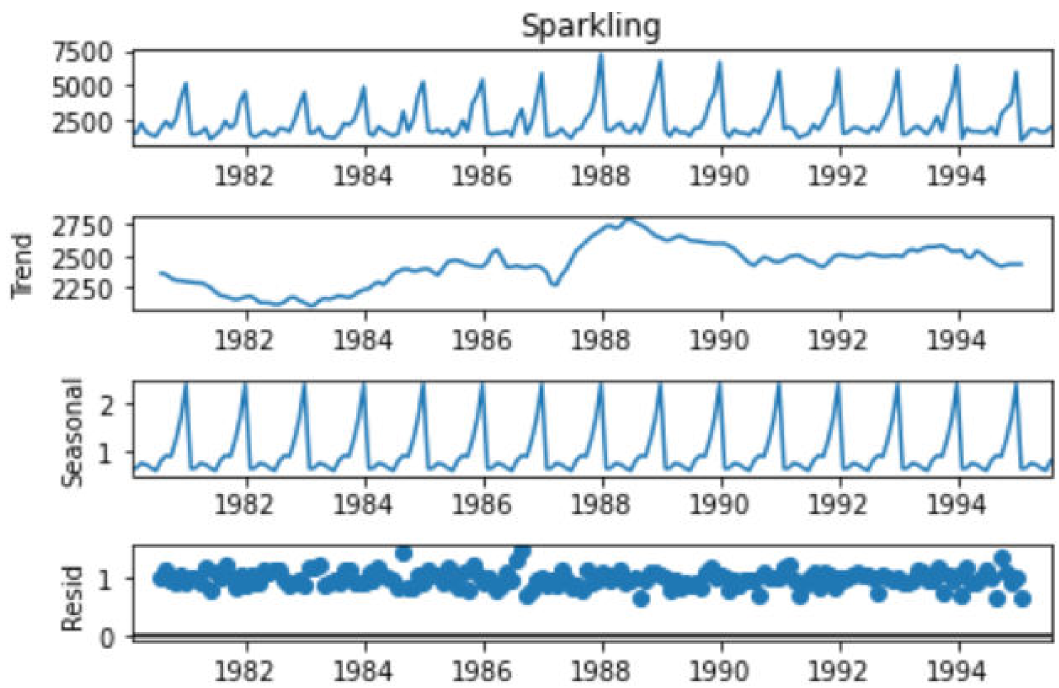
Time series decomposition

Sparkling

Additive decomposition



Multiplicative model



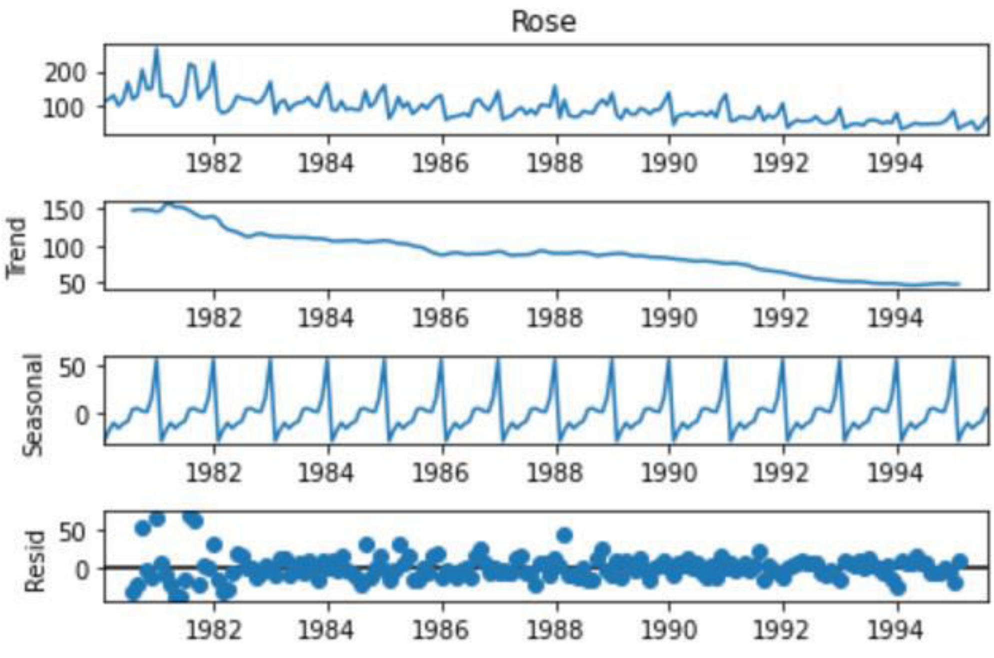
As the altitude of the seasonal peaks in the observed plot is changing according to the change in trend, the time-series is assumed to be ‘multiplicative’

The plot of the trend component does not show a consistent trend, but an intermediary period shows an upward slope which gets consistent on the late half of time-series

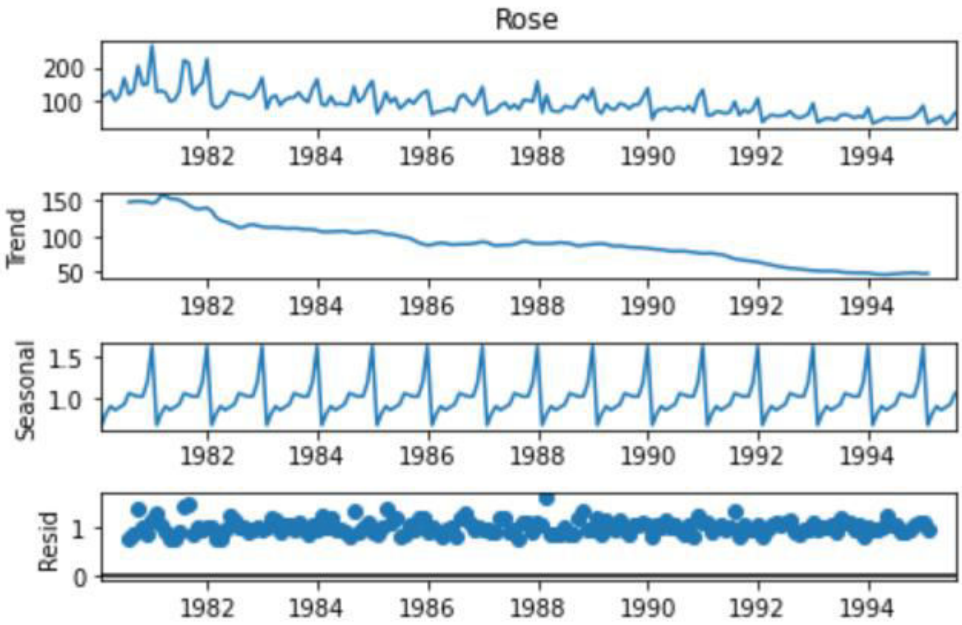
The residual shows a pattern of high variability across the period of time-series, which is more or less consistent in both additive and multiplicative decompositions

ROSE

Additive



Multiplicative



The observed plot of the decomposition diagrams shows visible annual seasonality and a downward trend. The early period of the plot shows higher variation than in the later periods

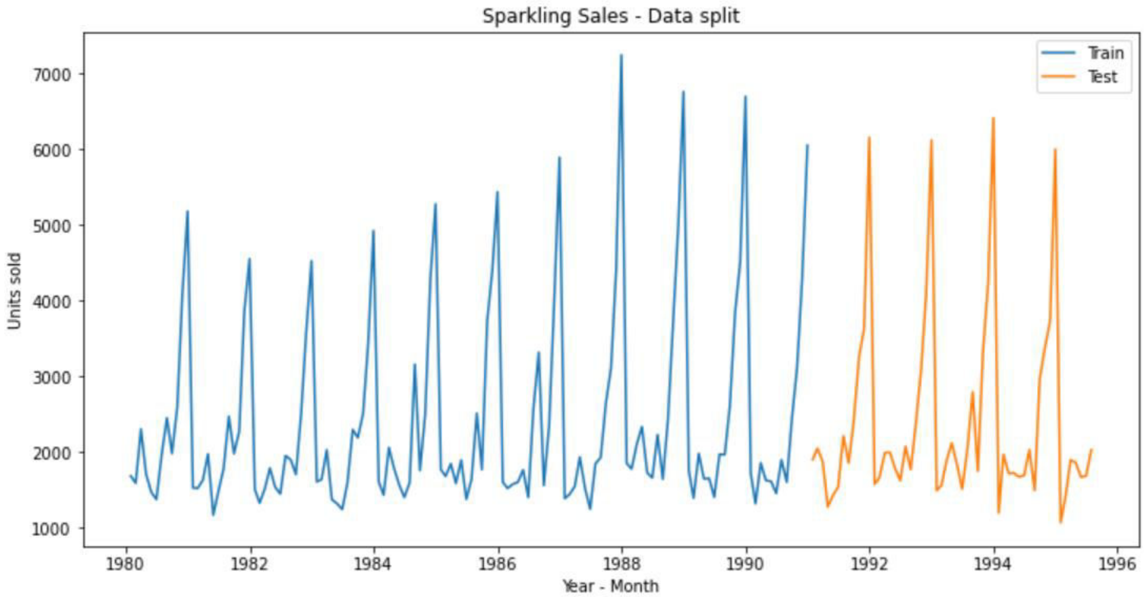
The trend diagram shows a downward trend overall. Exponential dips can be seen between 1981 and 1983 and later from 1991 to 1993

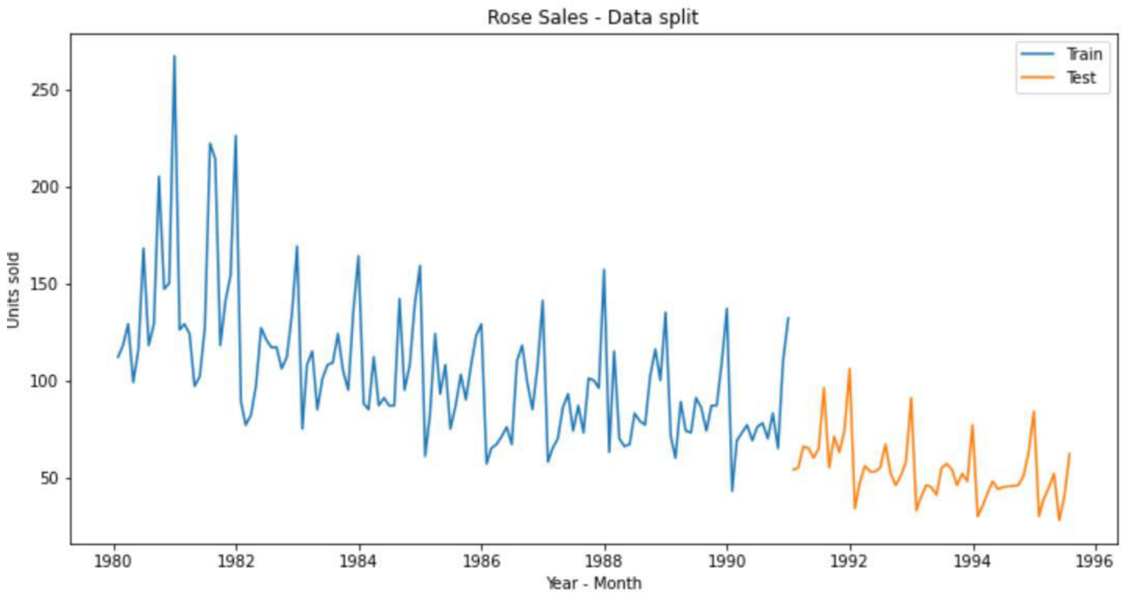
The residuals show a pattern of high variability across the period of time-series, which is more or less consistent in both additive and multiplicative decompositions

As the seasonality peaks are consistently reducing its altitude in consistent with trend, the series can be treated as multiplicative in model building.

<u>Splitting the data into train and test data</u>

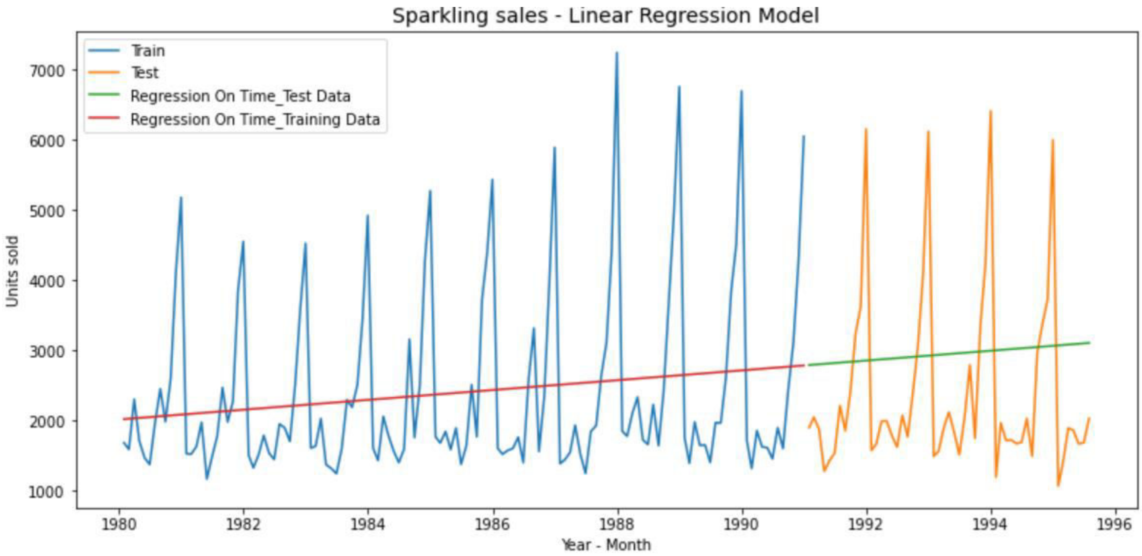
The train data is created with data before the year 1991 and the train dataset after the year 1991.





<u>Models</u>

<u>Model1 -Linear regression</u>

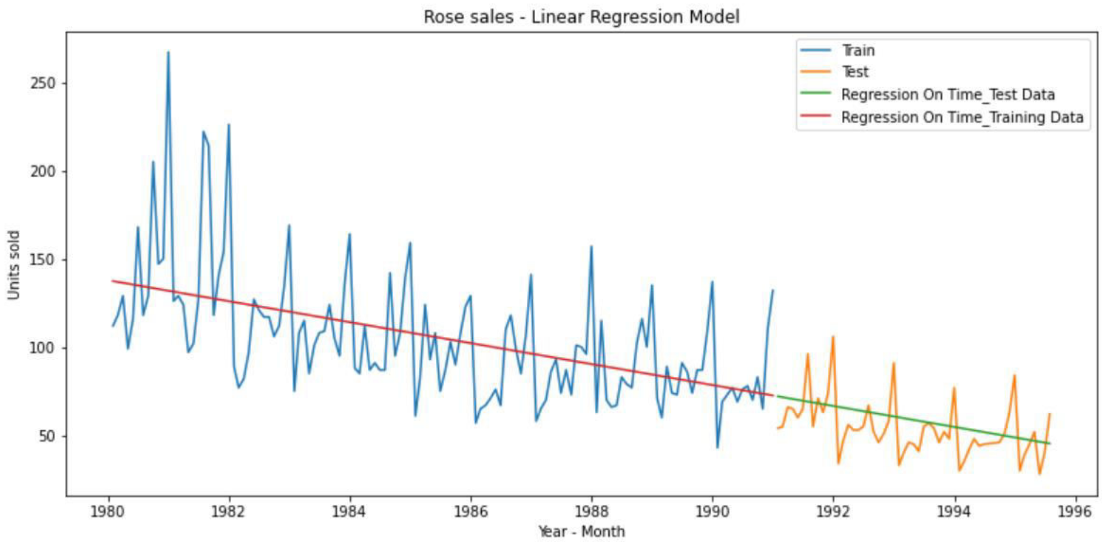


The liner regression plot shows a gradual upward trend in forecast of the sparkling wine .

	RMSE	MAPE
TRAIN	1279.322346	40.050000
TEST	1389.1315	50.15

The MAPE values is 50.15 for the test data set. We can say that the forecast is 50% accurate.

ROSE



The linear model shows a downward trend as observed with the original data.

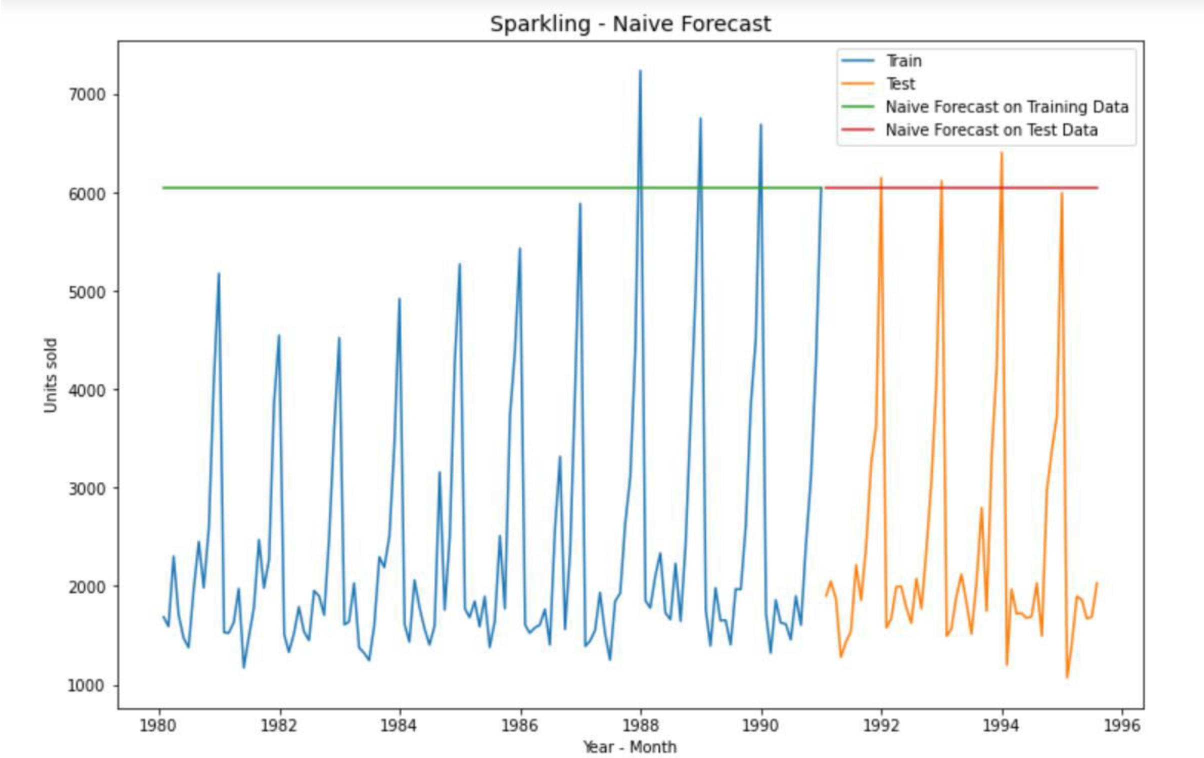
	RMSE	MAPE

TRAIN	30.718	21.2200
TEST	15.2688	22.8200

The model has successfully captured the trend of the series but not the seasonality.

Model2- naïve forecasting

SPARKLING

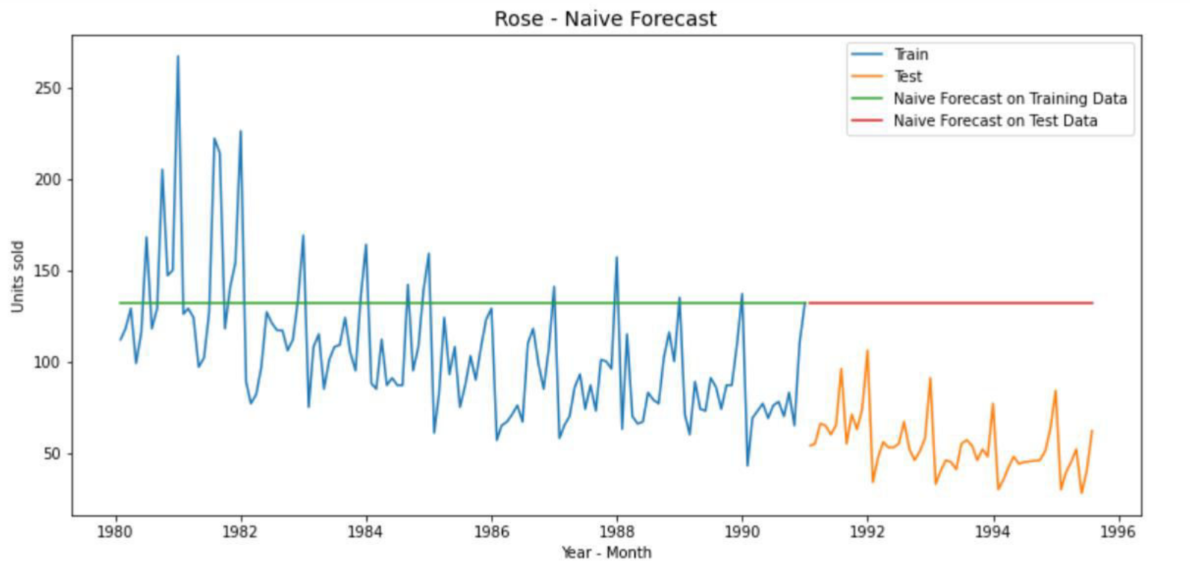


The model has taken the last value from the test data and fitted it in the rest of the train.

	RMSE	MAPE
TRAIN	3867.700802	153.170000
TEST	3864.279352	152.870000

The performance matrix is very poor with high % of error

ROSE

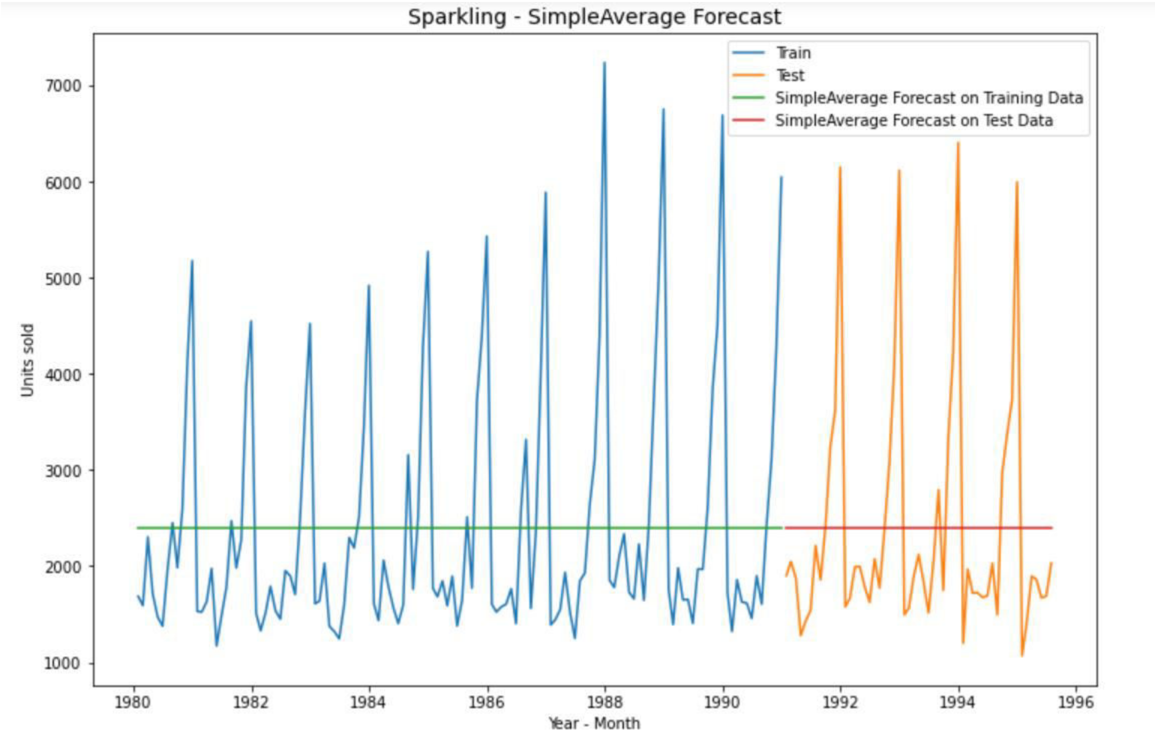


The model has neither captured the trend nor the seasonality of the data.

	RMSE	MAPE
TRAIN	45.063	36.38000
TEST	79.719	145.10

The error for the test is higher than train.

<u>Model3- Simple average forecasting</u>

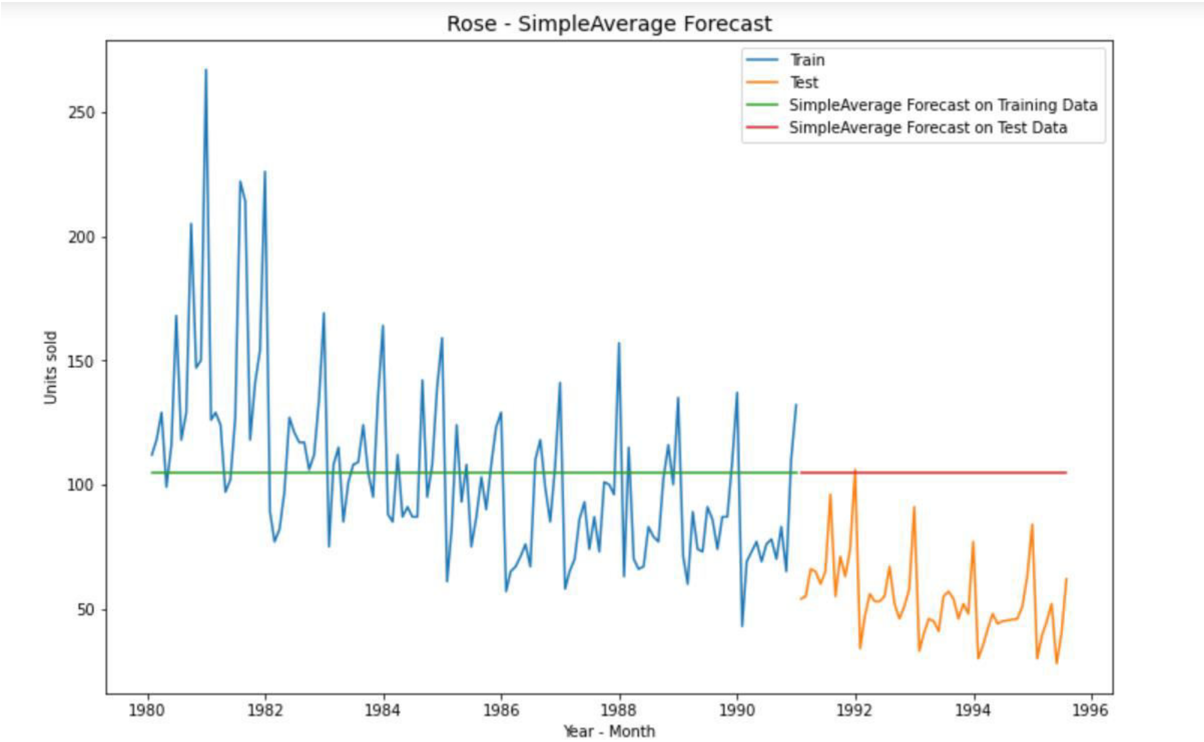


The model has not captured the trend or the seasonality of the data.

	RMSE	MAPE
TRAIN	1298.483628	40.3600
TEST	1275.082	38.90

The RMSE and MAPE are consistent on both train and test data.

ROSE



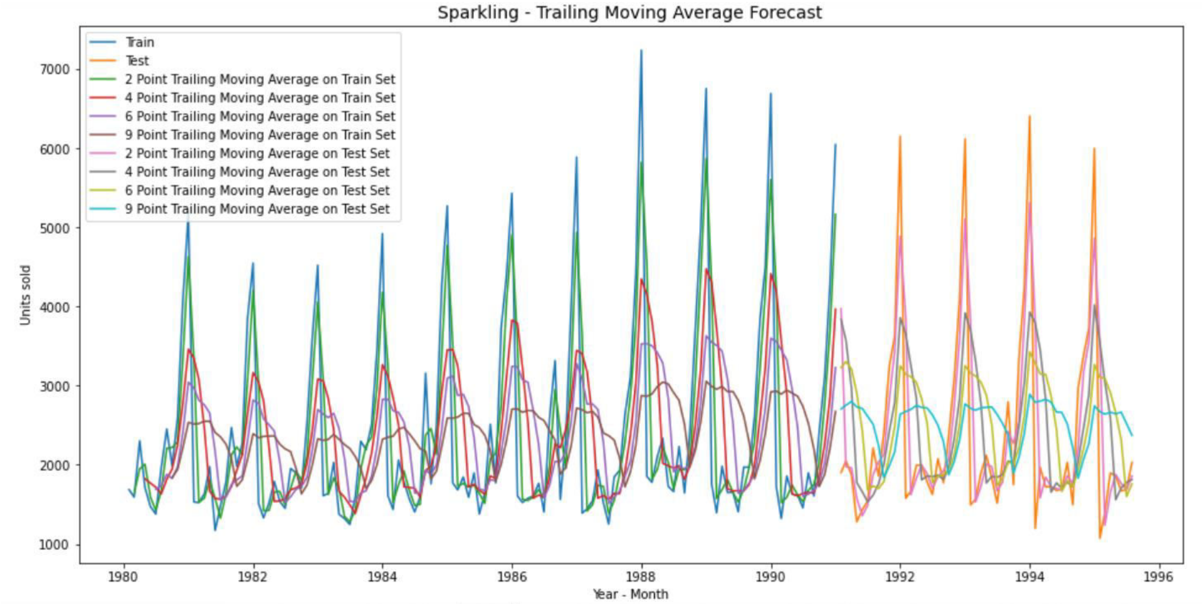
Due to the downward trend the performance of the train dataset is comparatively better than the test dataset

	RMSE	MAPE
TRAIN	36.034	25.39
TEST	53.460	94.93

The forecast has higher error than the train dataset

Model4 – moving average

SPARKLING



In the moving average model, we calculate rolling means for different intervals. The nest interval can be determined by the maximum accuracy.

The model is built for 2,4,6,9 moving points.

The accuracy is found better in the lower rolling points.

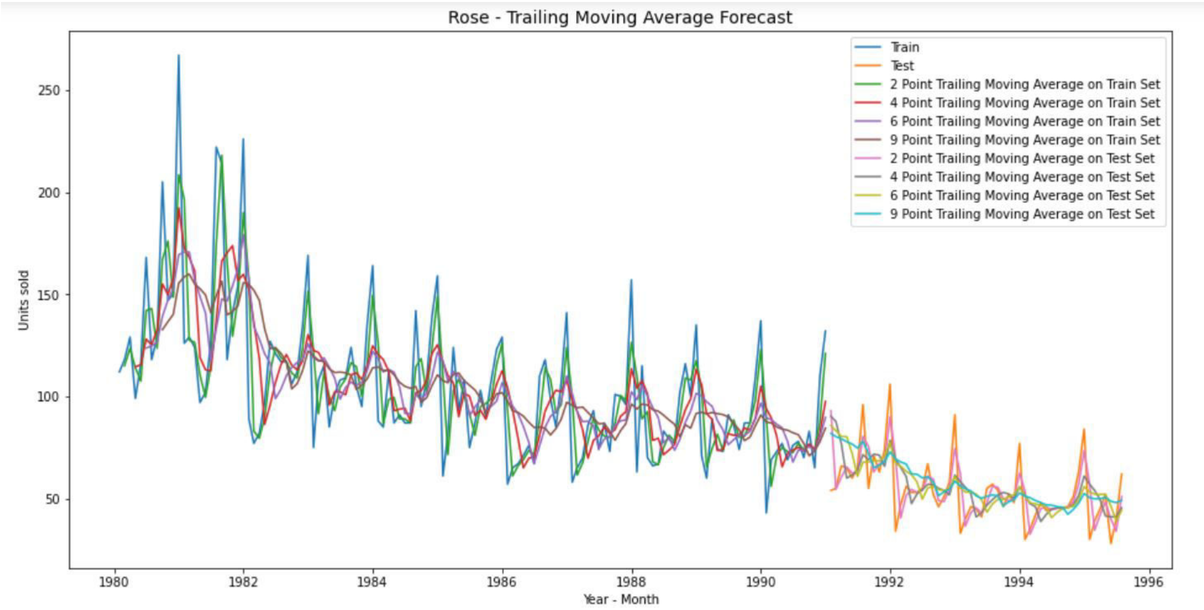
For 2 point Moving Average Model forecast on the Training Data, rmse_spark is 813.400684 mape_spark is 19.700000

For 4 point Moving Average Model forecast on the Training Data, rmse_spark is 1156.589694 mape_spark is 35.960000

For 6 point Moving Average Model forecast on the Training Data, rmse_spark is 1283.927428 mape_spark is 43.860000

For 9 point Moving Average Model forecast on the Training Data, rmse_spark is 1346.278315 mape_spark is 46.860000

ROSE



For 2 point Moving Average Model forecast on the Training Data, rmse_rose is 11.529 mape_rose is 13.54

For 4 point Moving Average Model forecast on the Training Data, rmse_rose is 14.451 mape_rose is 19.49

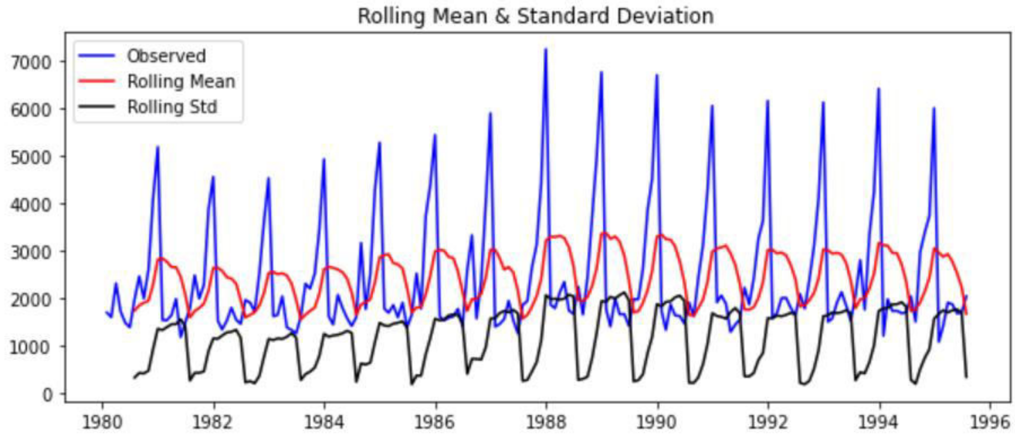
For 6 point Moving Average Model forecast on the Training Data, rmse_rose is 14.566 mape_rose is 20.82

For 9 point Moving Average Model forecast on the Training Data, rmse_rose is 14.728 mape_rose is 21.01

The best interval of moving average from the model is 2 point.

Checking stationary of data

SPARKLING




```
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
```

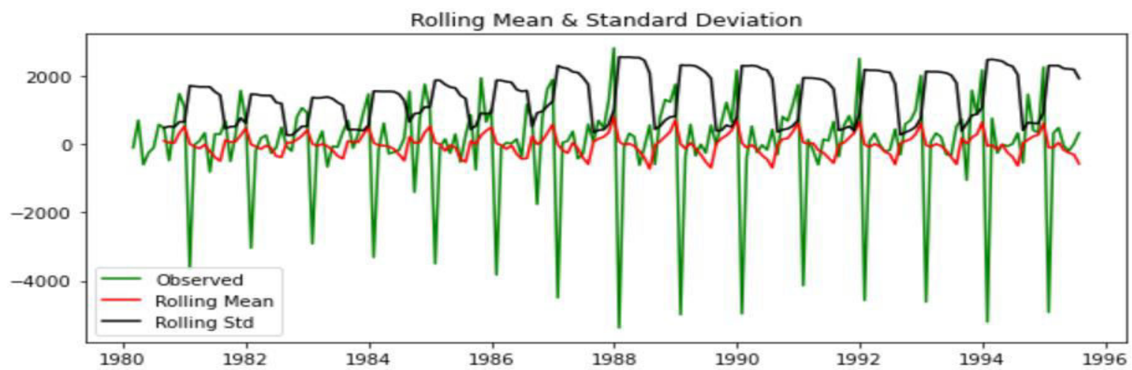
Augmented dicky fuller test is a test to check for stationarity.

Null hypothesis H0: the series is non-stationary.

Alternate hypothesis H1: the series is stationary.

The p value is greater than 0.05 hence we fail to reject the null hypothesis. Therefore the series is non stationary.

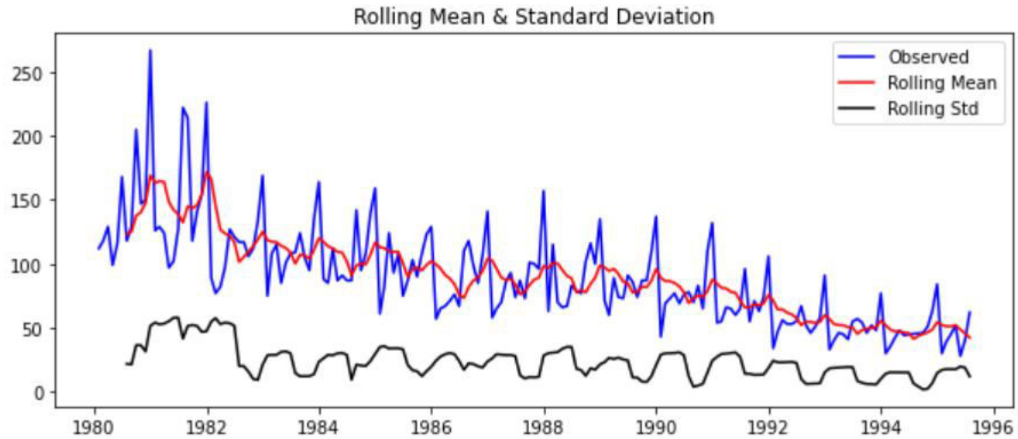
Differencing of order is applied for the sparkling dataset. Now, let's check if its stationary.




```
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
```

P value is less than 0.05 . we can reject the null hypothesis. That means it is stationary.

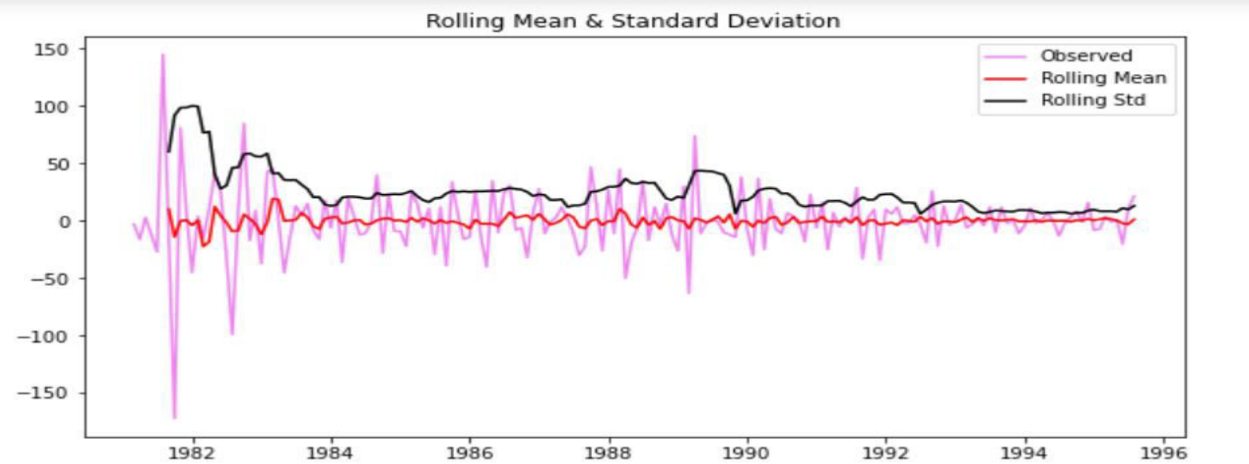
ROSE




```
Results of Dickey-Fuller Test:
Test Statistic                  -1.876719
p-value                          0.343091
#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
```

Similar to sparkling, the p value is greater than 0.05 , hence we fail to reject the null hypothesis.

Now, we apply differencing of one is applied.




```
Results of Dickey-Fuller Test:
Test Statistic                  -4.605732
p-value                          0.000126
#Lags Used                      11.000000
Number of Observations Used    162.000000
Critical Value (1%)             -3.471374
Critical Value (5%)             -2.879552
Critical Value (10%)            -2.576373
dtype: float64
```

The p value is less than 0.05 hence we reject the null hypothesis. We can say it is stationary.

Model 5- SARIMA

Auto SARIMA

SPARKLING

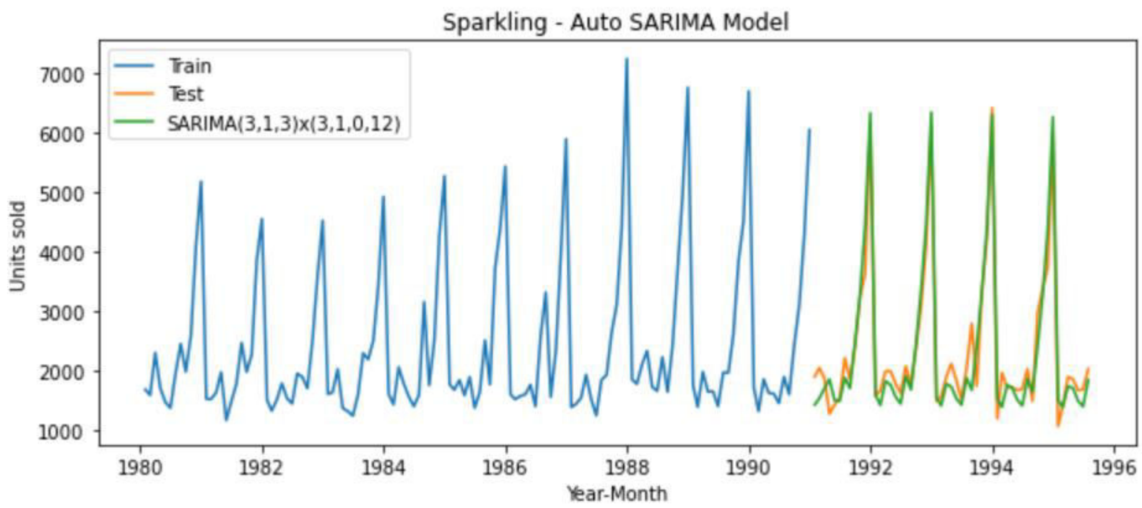
Since the data has seasonality, we use SARIMA instead of ARIMA.

	param	seasonal	AIC
220	(3, 1, 1)	(3, 1, 0, 12)	1215.898777
236	(3, 1, 2)	(3, 1, 0, 12)	1216.859173
221	(3, 1, 1)	(3, 1, 1, 12)	1217.713892
222	(3, 1, 1)	(3, 1, 2, 12)	1218.416045
237	(3, 1, 2)	(3, 1, 1, 12)	1218.991387

```
                                     SARIMAX Results
==========================================================================================
Dep. Variable:                                   y   No. Observations:                  132
Model:             SARIMAX(3, 1, 3)x(3, 1, [], 12)   Log Likelihood                -596.641
Date:                             Fri, 24 Sep 2021   AIC                           1213.283
Time:                                     22:44:17   BIC                           1237.103
Sample:                                          0   HQIC                          1222.833
                                             - 132
Covariance Type:                               opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -1.6142      0.176     -9.170      0.000      -1.959      -1.269
ar.L2         -0.6125      0.299     -2.047      0.041      -1.199      -0.026
ar.L3          0.0859      0.161      0.535      0.593      -0.229       0.401
ma.L1          0.9858      0.466      2.118      0.034       0.073       1.898
ma.L2         -0.8731      0.166     -5.261      0.000      -1.198      -0.548
ma.L3         -0.9460      0.483     -1.960      0.050      -1.892    -4.7e-05
ar.S.L12      -0.4521      0.142     -3.193      0.001      -0.730      -0.175
ar.S.L24      -0.2340      0.144     -1.621      0.105      -0.517       0.049
ar.S.L36      -0.1006      0.122     -0.827      0.408      -0.339       0.138
sigma2      1.839e+05   8.86e+04      2.076      0.038    1.03e+04    3.57e+05
===================================================================================
Ljung-Box (Q):                       23.19   Jarque-Bera (JB):                 4.07
Prob(Q):                              0.98   Prob(JB):                         0.13
Heteroskedasticity (H):               0.73   Skew:                             0.48
Prob(H) (two-sided):                  0.41   Kurtosis:                         3.54
===================================================================================

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

The top three models with lowest AIC values are as given. As per the AIC criteria, the optimum values for final SARIMA model selected is (3, 1, 1)x(3, 1, 0, 12).



The model captures both seasonality and trend of the data.

The rmse and mape value is given below

For SARIMA forecast on the Sparkling Testing Data: RMSE is 324.198 and MAPE is 9.48.

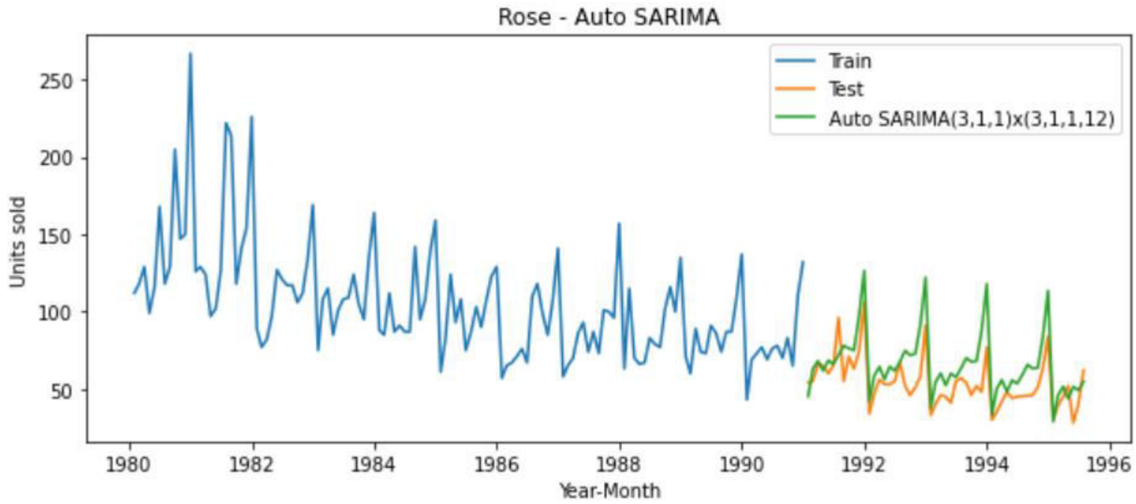
The error rate is much lesser than the other models

From the p-values it can be inferred that terms AR(1), AR(2), MA(1), MA(2), MA(3) and seasonal AR(1) are significant terms, as their values are below 0.05

ROSE

```
                                     SARIMAX Results
==========================================================================================
Dep. Variable:                                   y   No. Observations:                 132
Model:             SARIMAX(3, 1, 1)x(3, 1, 1, 12)   Log Likelihood                -331.681
Date:                             Sat, 25 Sep 2021   AIC                            681.363
Time:                                     00:10:24   BIC                            702.801
Sample:                                          0   HQIC                           689.958
                                             - 132
Covariance Type:                               opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.0171      0.151      0.113      0.910      -0.279       0.313
ar.L2         -0.0427      0.141     -0.302      0.762      -0.319       0.234
ar.L3         -0.0574      0.119     -0.484      0.629      -0.290       0.175
ma.L1         -0.9387      0.085    -11.085      0.000      -1.105      -0.773
ar.S.L12       0.0907      0.126      0.719      0.472      -0.156       0.338
ar.S.L24      -0.0438      0.107     -0.407      0.684      -0.254       0.167
ar.S.L36   -2.657e-05      0.054     -0.000      1.000      -0.106       0.106
ma.S.L12      -1.0001    424.198     -0.002      0.998    -832.413     830.413
sigma2       185.3701   7.86e+04      0.002      0.998   -1.54e+05    1.54e+05
===================================================================================
Ljung-Box (Q):                       42.97   Jarque-Bera (JB):                 2.56
Prob(Q):                              0.35   Prob(JB):                         0.28
Heteroskedasticity (H):               0.56   Skew:                             0.42
Prob(H) (two-sided):                  0.13   Kurtosis:                         3.22
===================================================================================

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



The model captures both seasonality and trend of the rose data.

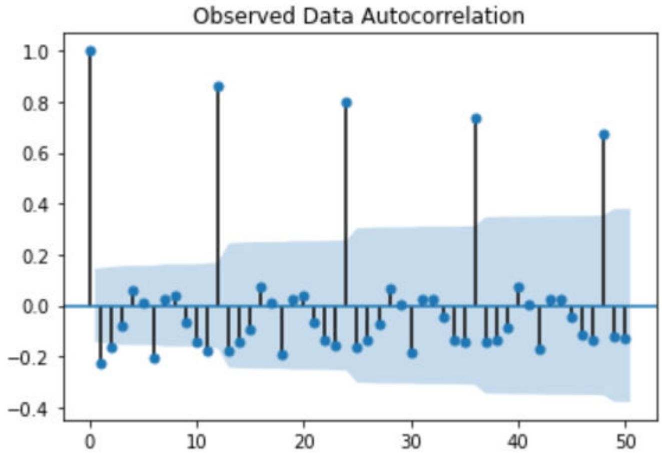
The rmse and mape value are given below

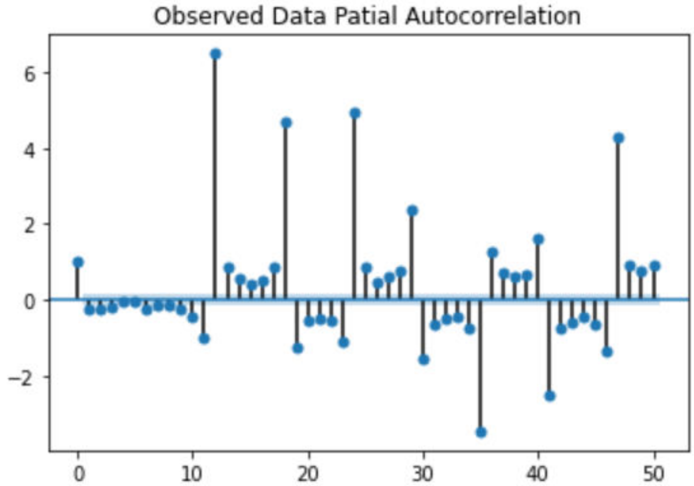
For SARIMA forecast on the SRose Testing Data: RMSE is 16.823 and MAPE is 25.48

From the below model summary it can be inferred that seasonal AR(2) term has the highest weightage, followed by seasonal MA(2)

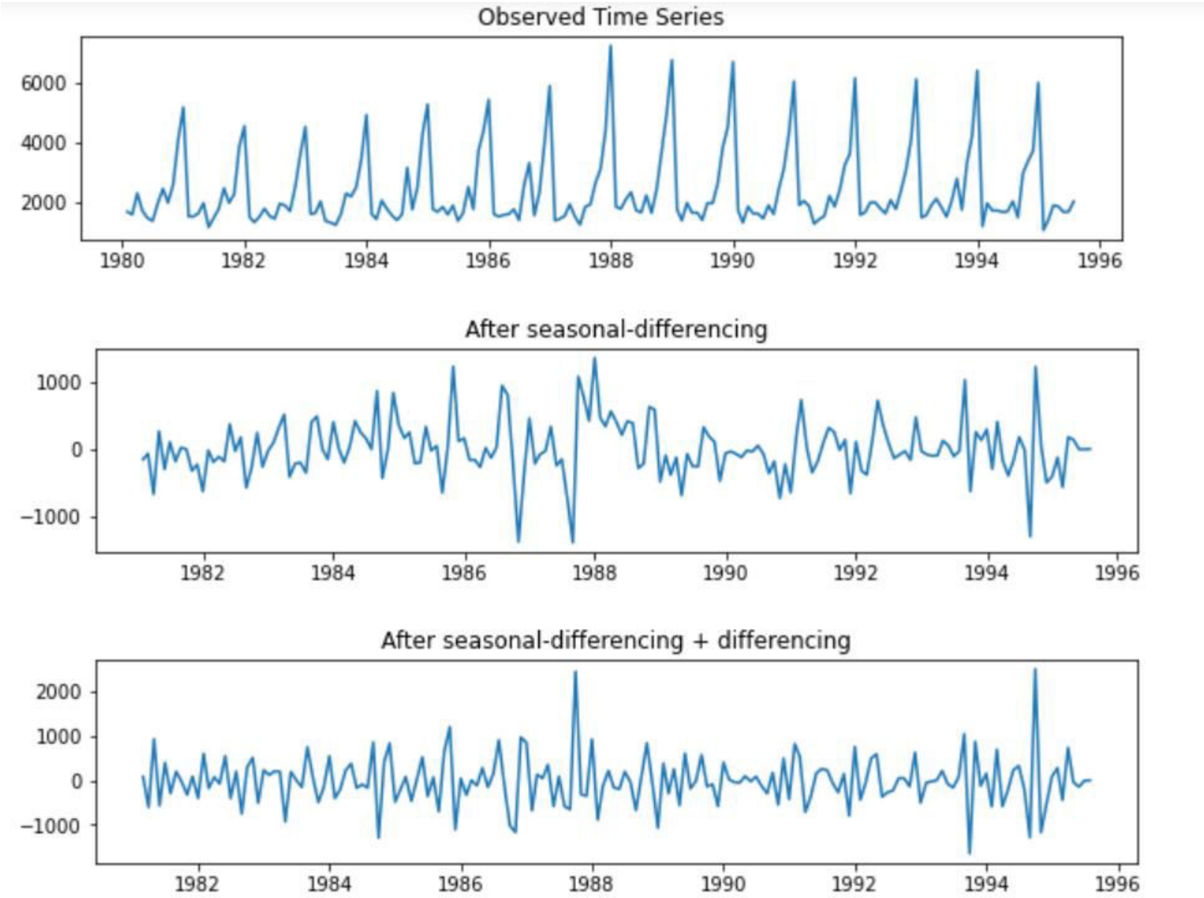
From the p-values it can be inferred that all the AR and MA terms are significant as the values are below .05.

<u>Manual SARIMA</u>

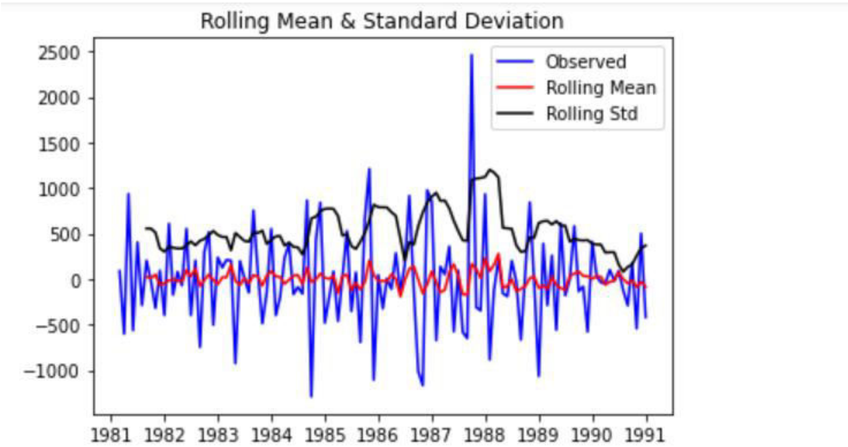




From the ACF plot of the observed/ train data, it can be inferred that at seasonal interval of 12, the plot is not quickly tapering off. So a seasonal differencing of 12 has to be taken.



From the plots above we can slight trend in the sparkling data. We still have to do stationarity test using ADF.




```
Results of Dickey-Fuller Test:
Test Statistic                  -3.342905
p-value                          0.013066
#Lags Used                      10.000000
Number of Observations Used    108.000000
Critical Value (1%)             -3.492401
Critical Value (5%)             -2.888697
Critical Value (10%)            -2.581255
dtype: float64
```

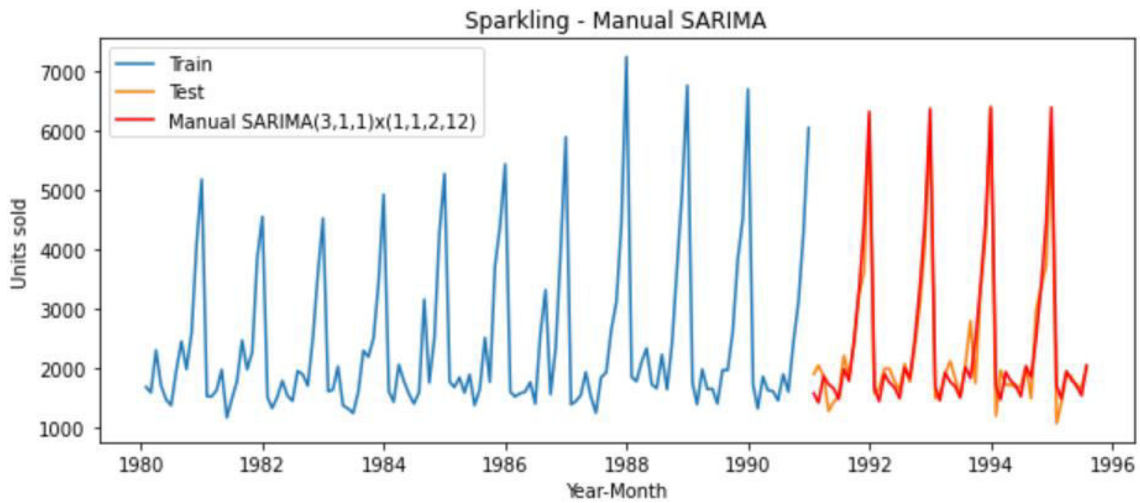
The p value is less than 0.05 , hence we can confirm that the data is stationary.

```
                                     SARIMAX Results
==========================================================================================
Dep. Variable:                                   y   No. Observations:                  132
Model:             SARIMAX(3, 1, 1)x(1, 1, [1, 2], 12)   Log Likelihood                -693.697
Date:                             Fri, 24 Sep 2021   AIC                           1403.394
Time:                                     23:56:14   BIC                           1423.654
Sample:                                          0   HQIC                          1411.574
                                             - 132
Covariance Type:                               opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.2229      0.130      1.714      0.086      -0.032       0.478
ar.L2         -0.0798      0.131     -0.607      0.544      -0.337       0.178
ar.L3          0.0922      0.122      0.756      0.450      -0.147       0.331
ma.L1         -1.0240      0.094    -10.903      0.000      -1.208      -0.840
ar.S.L12      -0.1992      0.866     -0.230      0.818      -1.896       1.498
ma.S.L12      -0.2108      0.880     -0.239      0.811      -1.936       1.515
ma.S.L24      -0.1299      0.381     -0.341      0.733      -0.876       0.616
sigma2      1.654e+05   2.63e+04      6.297      0.000    1.14e+05    2.17e+05
===================================================================================
Ljung-Box (Q):                       24.16   Jarque-Bera (JB):                19.65
Prob(Q):                              0.98   Prob(JB):                         0.00
Heteroskedasticity (H):               0.81   Skew:                             0.69
Prob(H) (two-sided):                  0.56   Kurtosis:                         4.78
===================================================================================

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

The model summary indicates that that only MA(1) term used in the model is significant in terms of p-values.

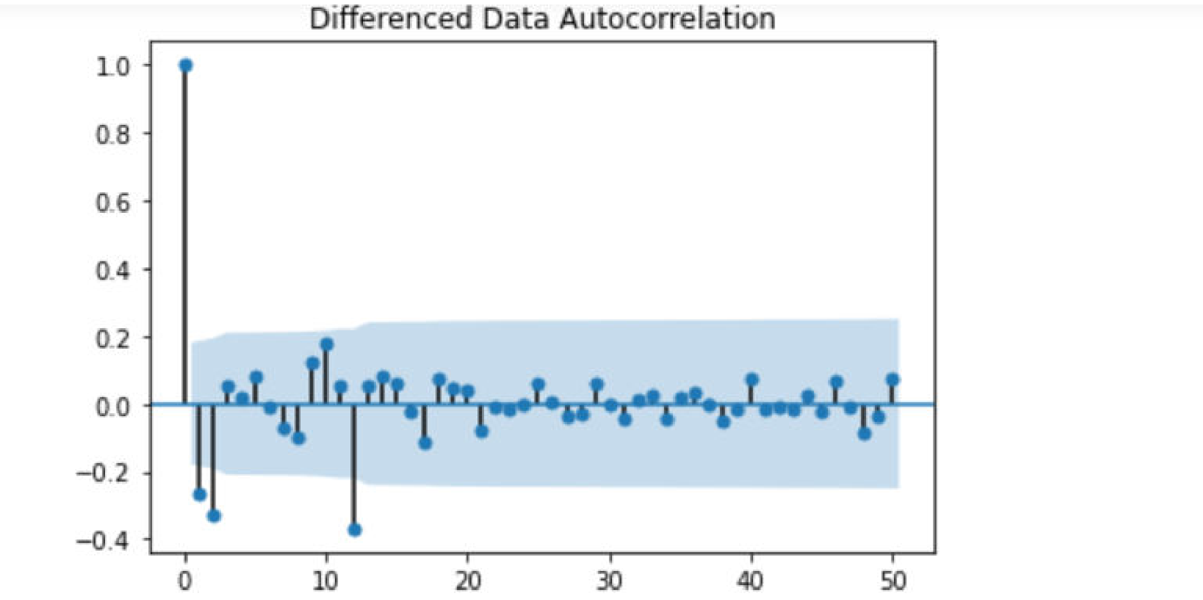
For SARIMA forecast on the Sparkling Testing Data: RMSE is 324.198 and MAPE is 9.48

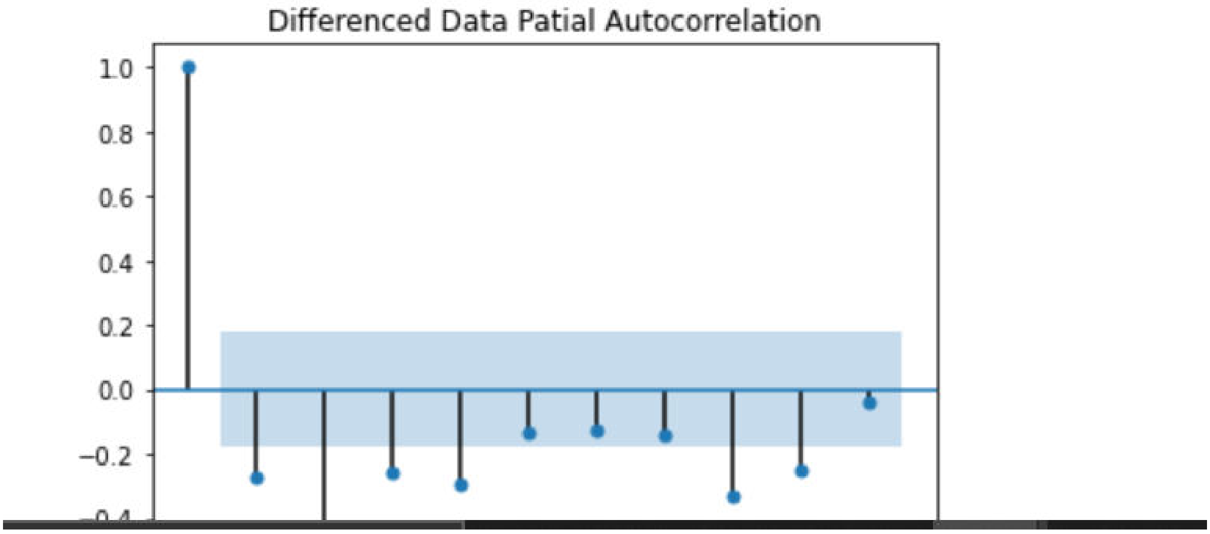


and MAPE.

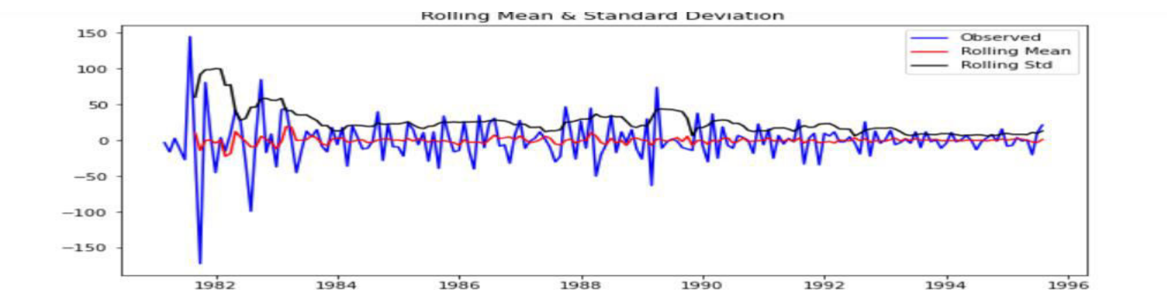
The trend and seasonality have been captured by the model.

ROSE





From the ACF plot of the log transformed data, it can be seen that at seasonal interval of 12, the plot is not quickly tapering off. So we need to take a seasonal differencing of 12



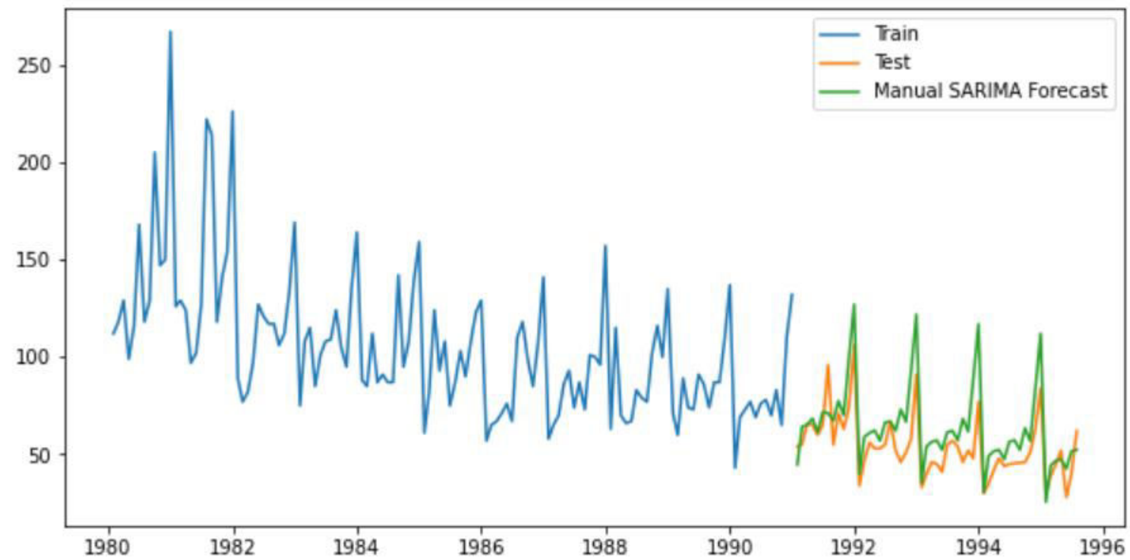

```
Results of Dickey-Fuller Test:
Test Statistic                  -4.605732
p-value                          0.000126
#Lags Used                      11.000000
Number of Observations Used    162.000000
Critical Value (1%)             -3.471374
Critical Value (5%)             -2.879552
Critical Value (10%)            -2.576373
dtype: float64
```

P value is less than 0.05 , he we can confirm that the data is stationary.

```
                                      SARIMAX Results
==========================================================================================
Dep. Variable:                                   y   No. Observations:                  132
Model:             SARIMAX(4, 1, 2)x(0, 1, 2, 12)   Log Likelihood                -384.369
Date:                            Sat, 25 Sep 2021   AIC                            786.737
Time:                                    00:13:24   BIC                            809.433
Sample:                                         0   HQIC                           795.898
                                            - 132
Covariance Type:                              opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.8966      0.134     -6.707      0.000      -1.159      -0.635
ar.L2          0.0167      0.171      0.097      0.922      -0.319       0.352
ar.L3         -0.1130      0.174     -0.649      0.516      -0.454       0.228
ar.L4         -0.1597      0.120     -1.329      0.184      -0.395       0.076
ma.L1          0.1507    590.502      0.000      1.000   -1157.212    1157.514
ma.L2         -0.8493    501.528     -0.002      0.999    -983.826     982.127
ma.S.L12      -0.3907      0.103     -3.780      0.000      -0.593      -0.188
ma.S.L24      -0.0887      0.091     -0.970      0.332      -0.268       0.090
sigma2       238.9655   1.41e+05      0.002      0.999   -2.76e+05    2.77e+05
===================================================================================
Ljung-Box (Q):                       27.59   Jarque-Bera (JB):                 0.01
Prob(Q):                              0.93   Prob(JB):                         0.99
Heteroskedasticity (H):               0.76   Skew:                            -0.01
Prob(H) (two-sided):                  0.46   Kurtosis:                         3.06
===================================================================================
```

The model summary indicates that that none of the terms used in the model are significant in terms of pvalues

For SARIMA forecast on the Rose Testing Data: RMSE is 15.378 and MAPE is 22.16



The model captures both seasonality and trend.

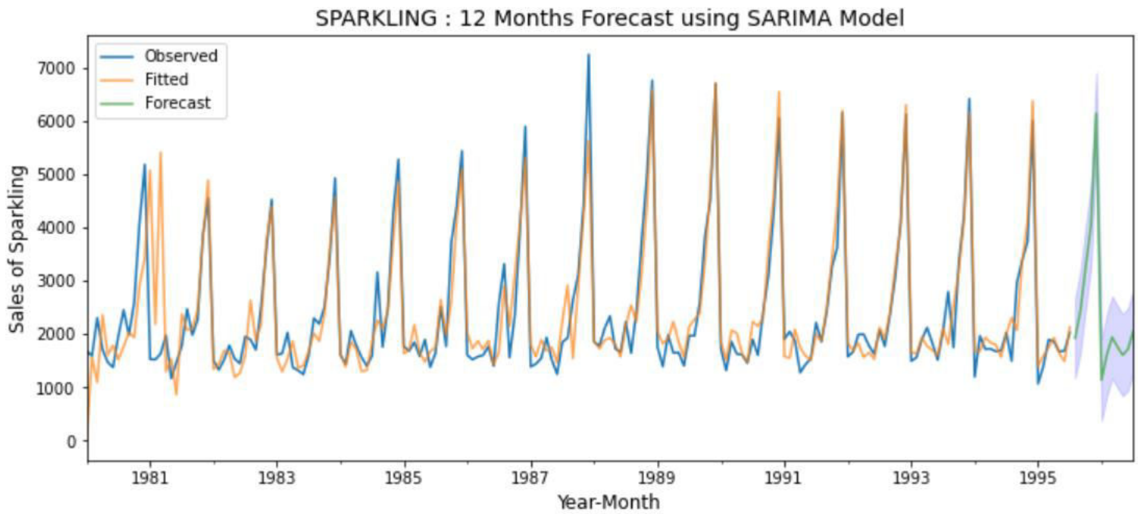
Model comparison

	Test RMSE	Test MAPE
Auto SARIMA(3,1,3)x(3,1,0,12)	324.198328	9.48
Manual SARIMA(3,1,1)x(1,1,2,12)	324.198328	9.48
2 point TMA	813.400684	19.70
4 point TMA	1156.589694	35.96
SimpleAverage	1275.081804	38.90
6 point TMA	1283.927428	43.86
9 point TMA	1346.278315	46.86
RegressionOnTime	1389.135175	50.15
NaiveModel	3864.279352	152.87
NaiveModel	3864.279352	152.87

Auto SARIMA looks like the optimal model

Performing Auto SARIMA for both sparkling and rose .

Forecasting for the next 12 months.



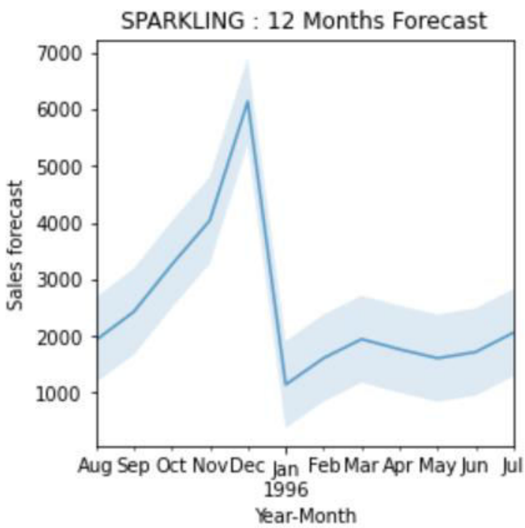
SARIMA model has reflected the trend and seasonality of the series continuing into the future year as well.

SARIMA model is seen to have better fitment with the most recent observed data and shows high variations in the farthest periods of observations, which explains the high RMSE and MAPE values.

```
                                      SARIMAX Results
==========================================================================================
Dep. Variable:                                Rose   No. Observations:                  187
Model:             SARIMAX(4, 1, 1)x(0, 1, 1, 12)   Log Likelihood                -664.135
Date:                             Sat, 25 Sep 2021   AIC                           1342.270
Time:                                     00:26:36   BIC                           1363.796
Sample:                                 01-31-1980   HQIC                          1351.011
                                      - 07-31-1995
Covariance Type:                               opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.0913      0.084      1.093      0.274      -0.072       0.255
ar.L2         -0.1077      0.077     -1.393      0.164      -0.259       0.044
ar.L3         -0.1315      0.076     -1.729      0.084      -0.280       0.018
ar.L4         -0.1071      0.078     -1.375      0.169      -0.260       0.046
ma.L1         -0.8270      0.056    -14.901      0.000      -0.936      -0.718
ma.S.L12      -0.5963      0.059    -10.122      0.000      -0.712      -0.481
sigma2       232.4258     24.359      9.542      0.000     184.683     280.169
===================================================================================
Ljung-Box (Q):                       35.39   Jarque-Bera (JB):                 5.30
Prob(Q):                              0.68   Prob(JB):                         0.07
Heteroskedasticity (H):               0.22   Skew:                             0.04
Prob(H) (two-sided):                  0.00   Kurtosis:                         3.89
===================================================================================

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

The model summary also provides valuable insights in the model. From the snapshot of summary below it can be understood that AR(2), MA(3) terms has the highest absolute weightage. The p-values indicates that the terms AR(1), AR(2), MA(1), MA(2) and MA(3) are the most significant terms .



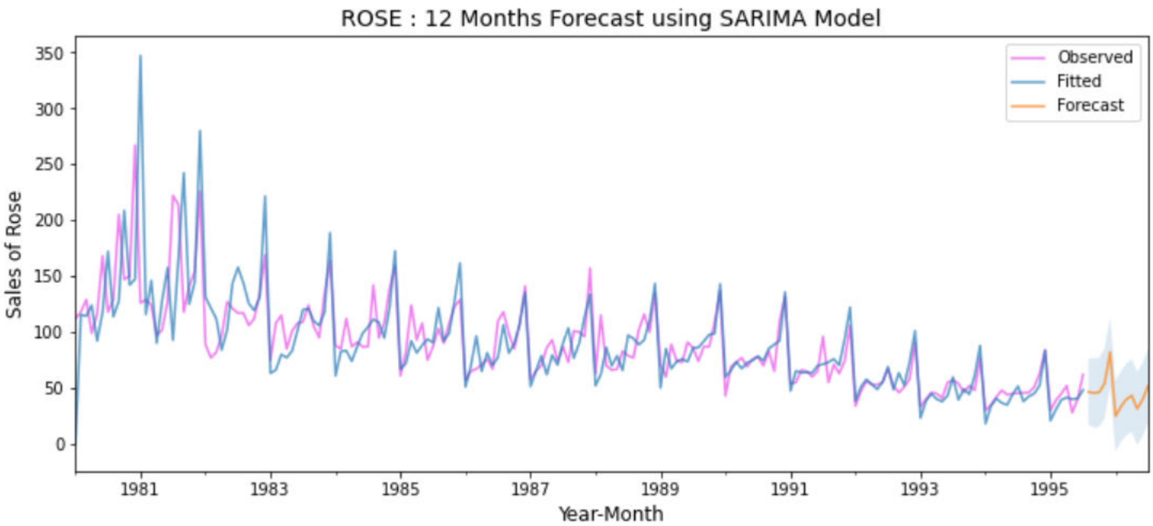
	Sparkling
1995-08-31	1925.82
1995-09-30	2427.05
1995-10-31	3258.60
1995-11-30	4037.14
1995-12-31	6137.71
1996-01-31	1140.13
1996-02-29	1609.25
1996-03-31	1943.26
1996-04-30	1764.12
1996-05-31	1607.08
1996-06-30	1715.31
1996-07-31	2057.31

The model forecasts average sale of 2459 units per month.

The seasonal sale in December 1995 will hit a maximum of 6084 units, before it drops to the lowest sale in January 1996.

The forecast also indicates that the year-on-year sale of sparkling wine is not showing an upward trend. The winery must adopt innovative marketing skills to improve the sale compared to previous years

ROSE

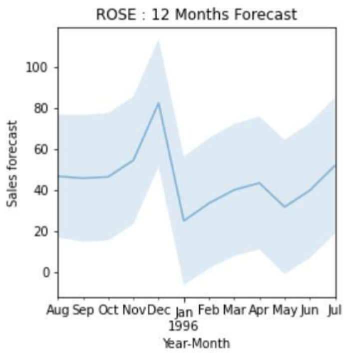


```
                                      SARIMAX Results
==========================================================================================
Dep. Variable:                                 Rose   No. Observations:                  187
Model:             SARIMAX(4, 1, 1)x(0, 1, 1, 12)   Log Likelihood                -664.135
Date:                              Sat, 25 Sep 2021   AIC                           1342.270
Time:                                      00:26:36   BIC                           1363.796
Sample:                                  01-31-1980   HQIC                          1351.011
                                       - 07-31-1995
Covariance Type:                                opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.0913      0.084      1.093      0.274      -0.072       0.255
ar.L2         -0.1077      0.077     -1.393      0.164      -0.259       0.044
ar.L3         -0.1315      0.076     -1.729      0.084      -0.280       0.018
ar.L4         -0.1071      0.078     -1.375      0.169      -0.260       0.046
ma.L1         -0.8270      0.056    -14.901      0.000      -0.936      -0.718
ma.S.L12      -0.5963      0.059    -10.122      0.000      -0.712      -0.481
sigma2       232.4258     24.359      9.542      0.000     184.683     280.169
===================================================================================
Ljung-Box (Q):                       35.39   Jarque-Bera (JB):                 5.30
Prob(Q):                              0.68   Prob(JB):                         0.07
Heteroskedasticity (H):               0.22   Skew:                             0.04
Prob(H) (two-sided):                  0.00   Kurtosis:                         3.89
===================================================================================

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

The model summary also provides valuable insights in the model. From the snapshot of summary below it can be understood that MA(1) and seasonal MA(1) term has the highest weightage. The p-values indicates that the terms MA(1) and Seasonal MA(1) are the most significant terms.

Forecasted plot



	ROSE
1995-08-31	46.54
1995-09-30	45.51
1995-10-31	46.23
1995-11-30	54.32
1995-12-31	82.21
1996-01-31	24.81
1996-02-29	33.36
1996-03-31	39.87
1996-04-30	43.23
1996-05-31	31.53
1996-06-30	39.56
1996-07-31	51.70

The model forecasts average sale of 45 units per month.

The seasonal sale in December 1995 will reach a maximum of 82 units, before it drops to the lowest sale in January 1996; at 25 units.

Apart from higher sale in November and December months, Rose sales will be above average in the summer months of July and August