DSBA-DEC20

Assignment

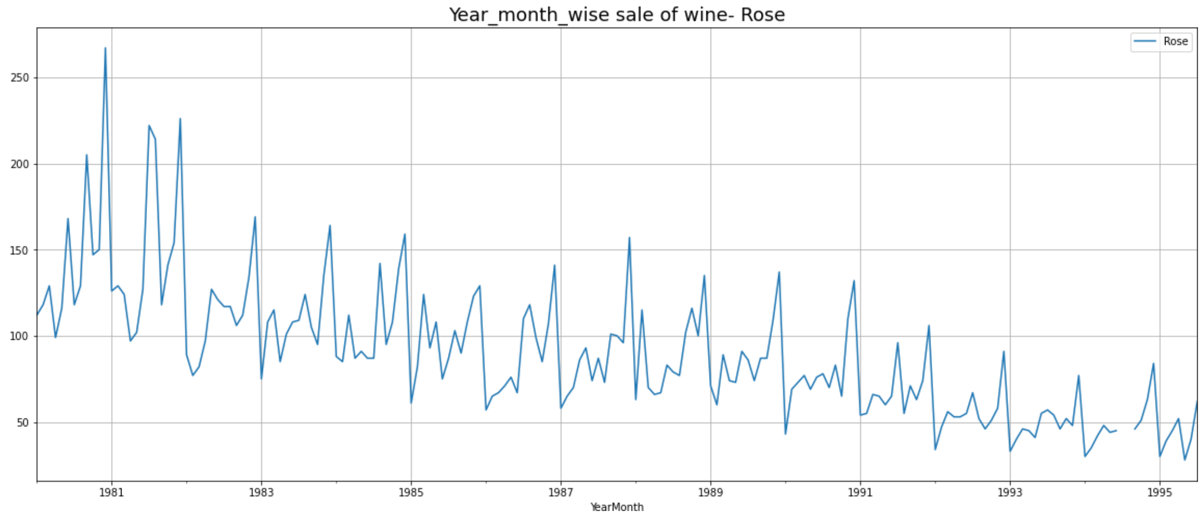
- Time Series Forecasting

Karthik Vadlamudi

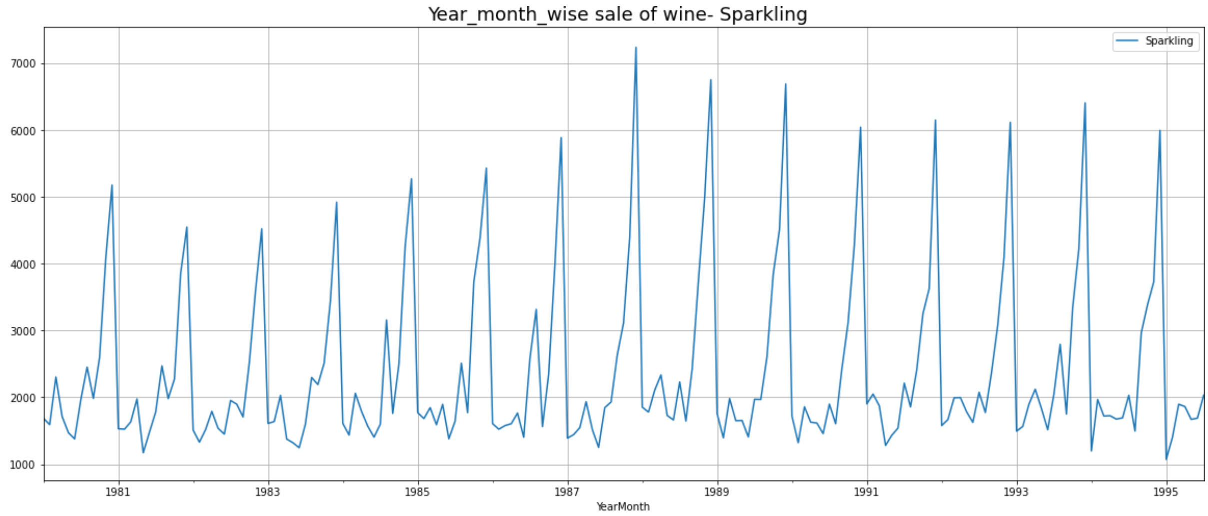
For this particular assignment, the data of different types of wine sales in the 20th century is to be analysed. Both of these data are from the same company but of different wines. As an analyst in the ABC Estate Wines, you are tasked to analyse and forecast Wine Sales in the 20th century.

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

Rose wine sales



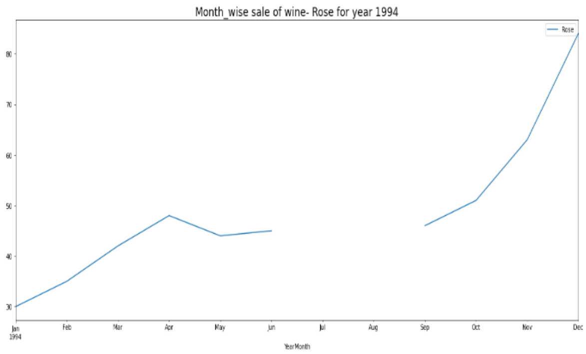
Sparkling -wine sale

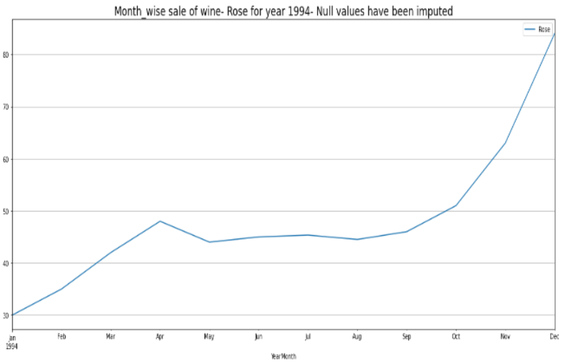


2. Perform appropriate Exploratory Data Analysis to understand the data and also perform decomposition.

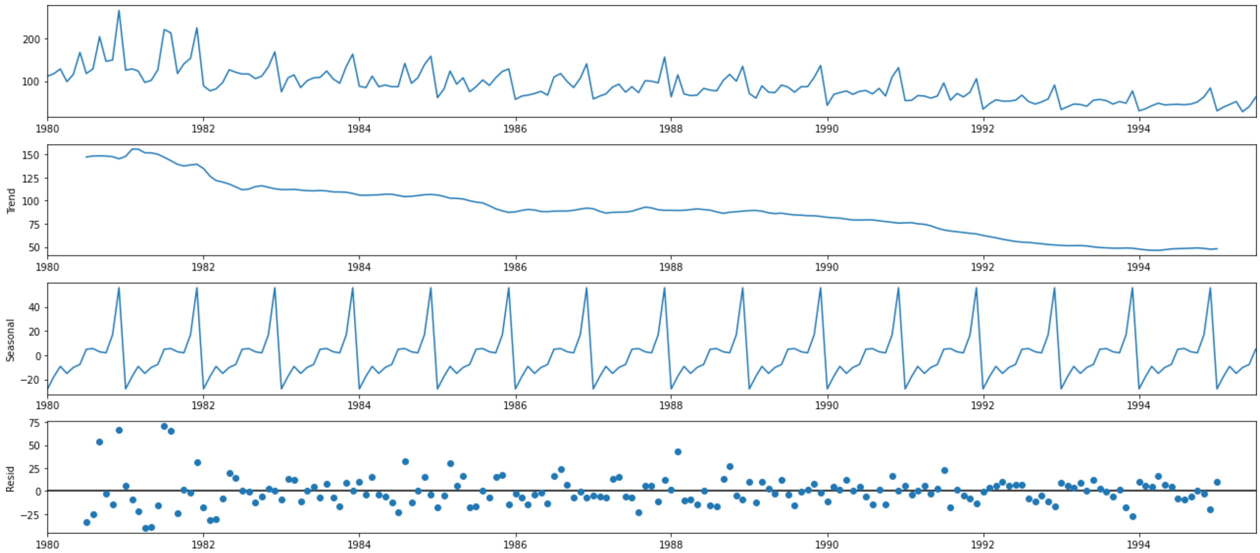
Rose wine sales

Data set has 187 records and two null values. Null values have been imputed

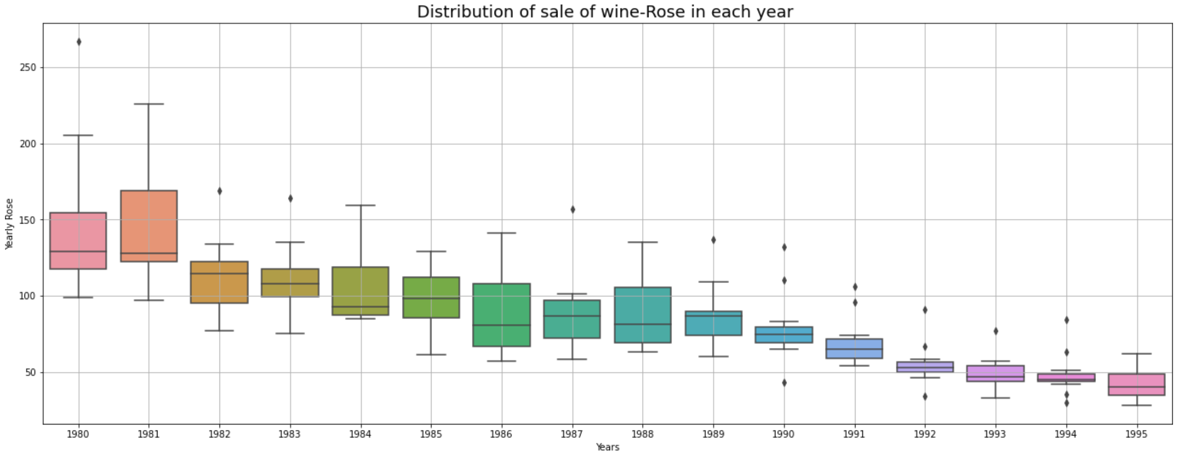




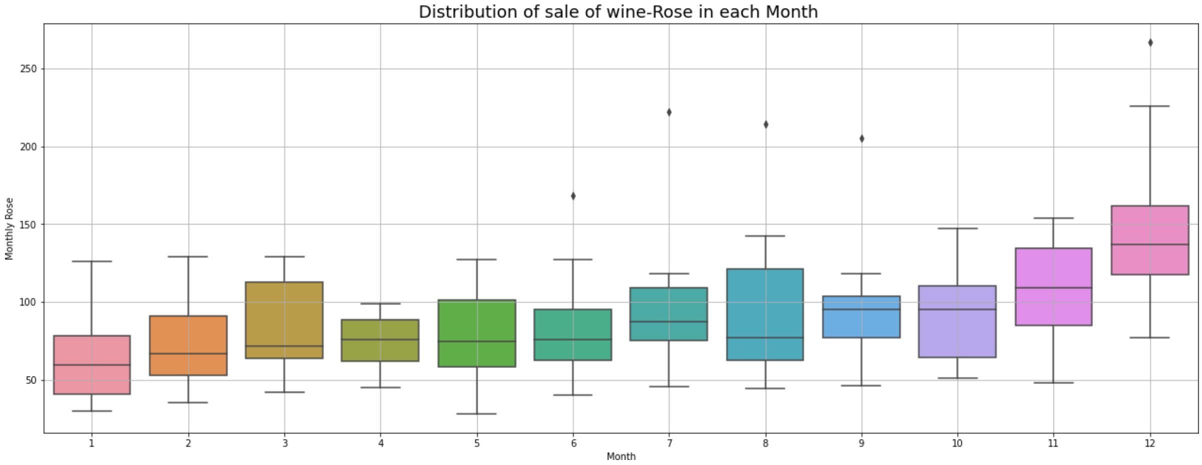
Decomposition of Rose-wine sales into Trend, Seasonal and Residual



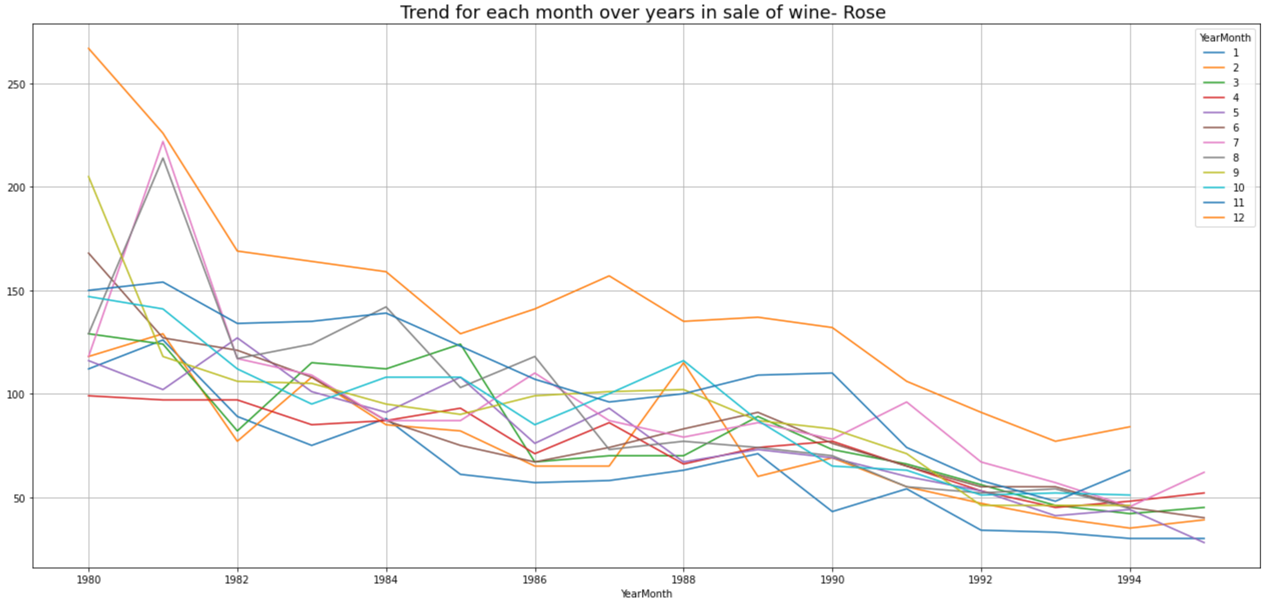
Distribution of sale of wine-Rose in each year



Distribution of sale of wine-Rose in each Month



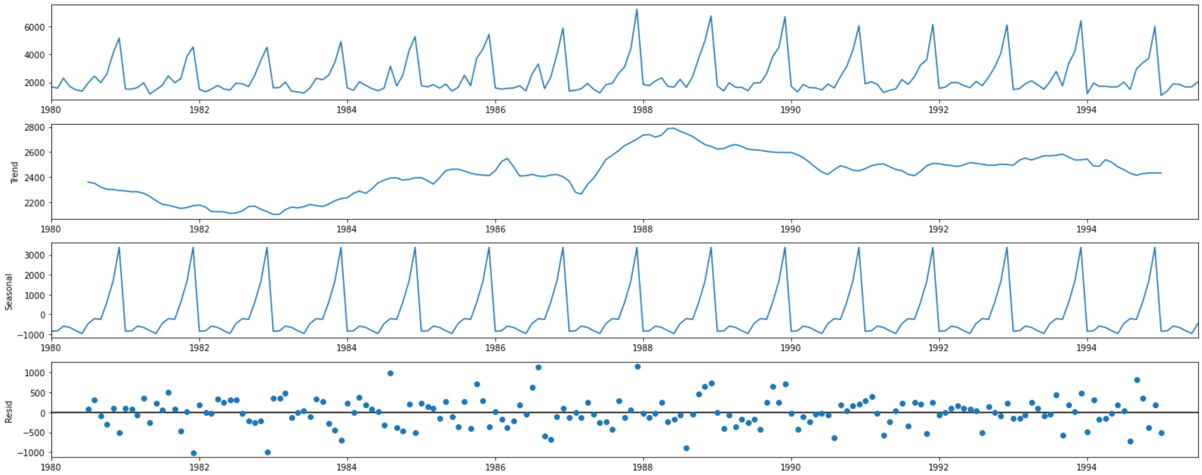
Trend for each month over the years in sale of wine-Rose



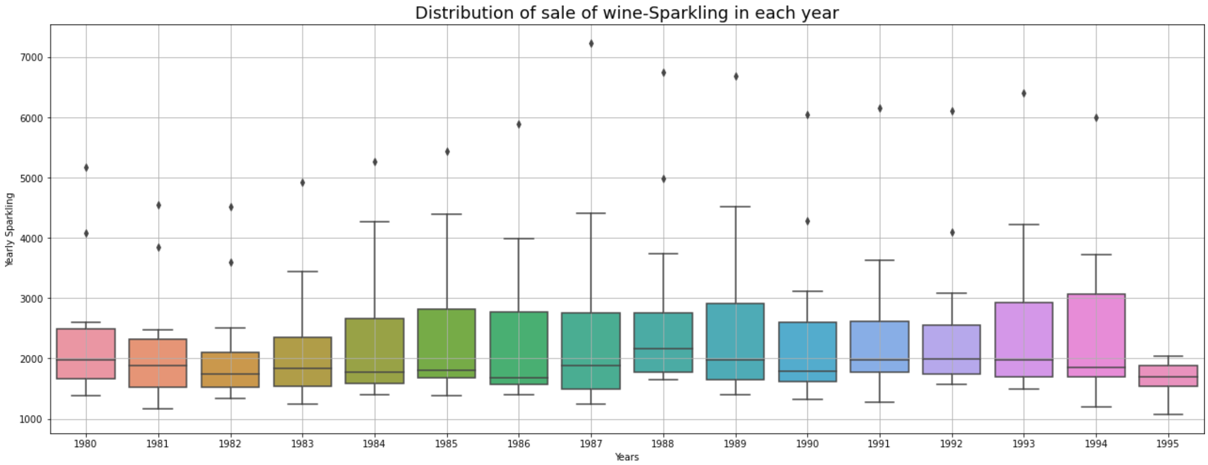
Sparkling wine sales

Data set has 187 records and there are no null values in data set Sparkling

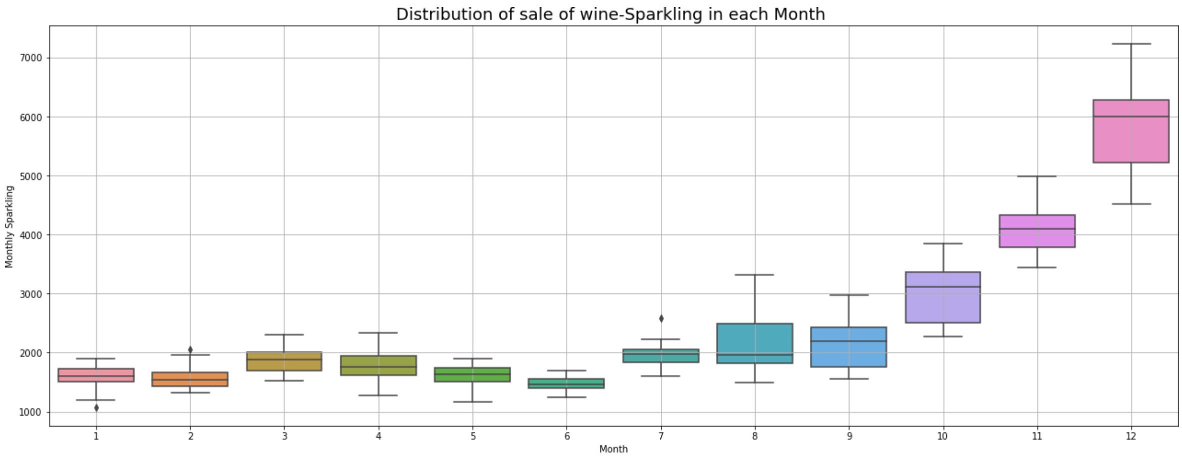
Decomposition of Sparkling-wine sales into Trend, Seasonal and Residual



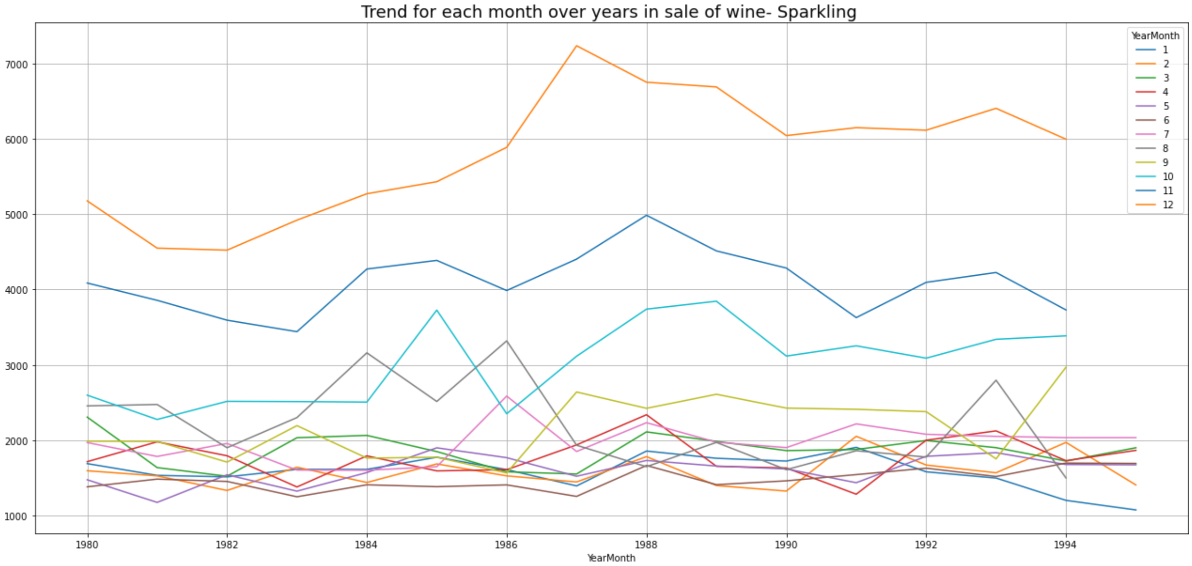
Distribution of sale of wine-Sparkling in each year



Distribution of sale of wine-Sparkling in each Month

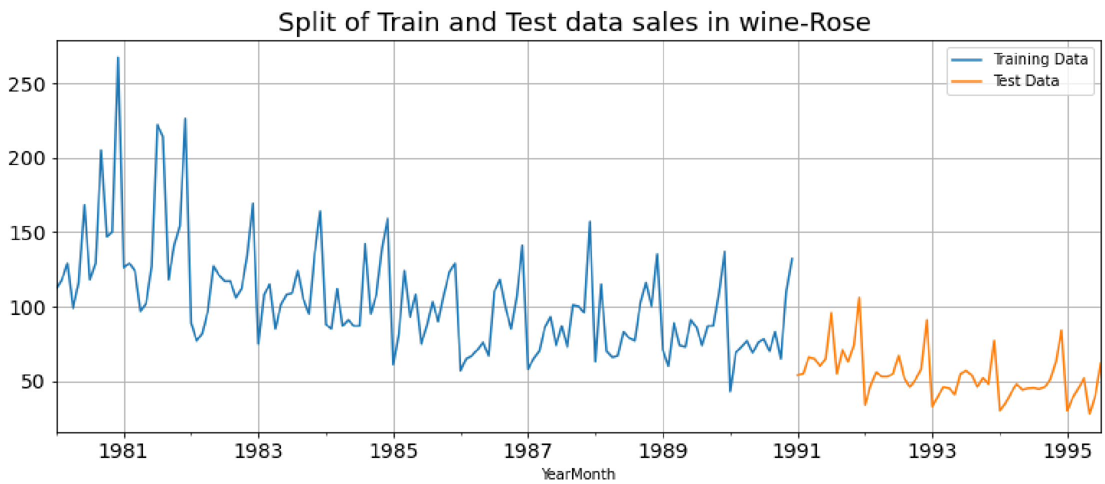


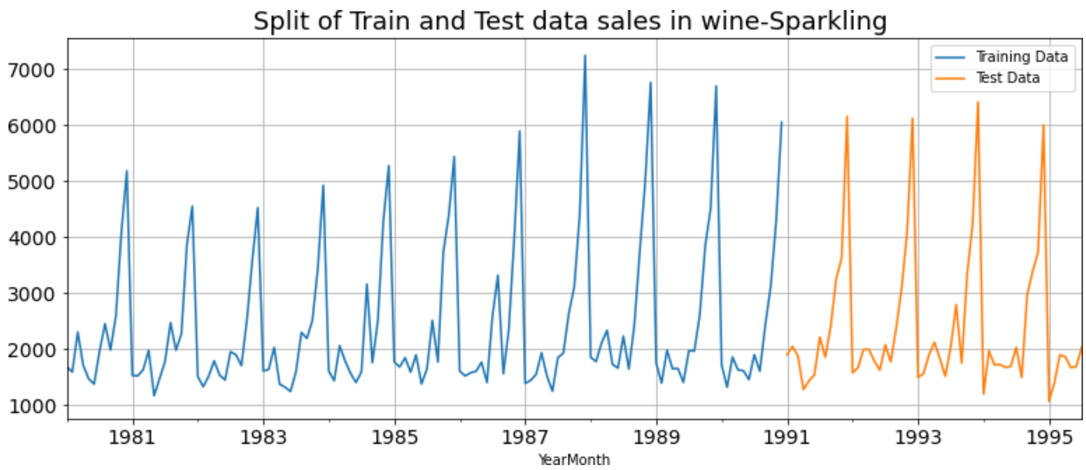
Trend for each month over the years in sale of wine-Sparkling



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

After splitting train and test data set , train data set has 132 records and test has 55 records



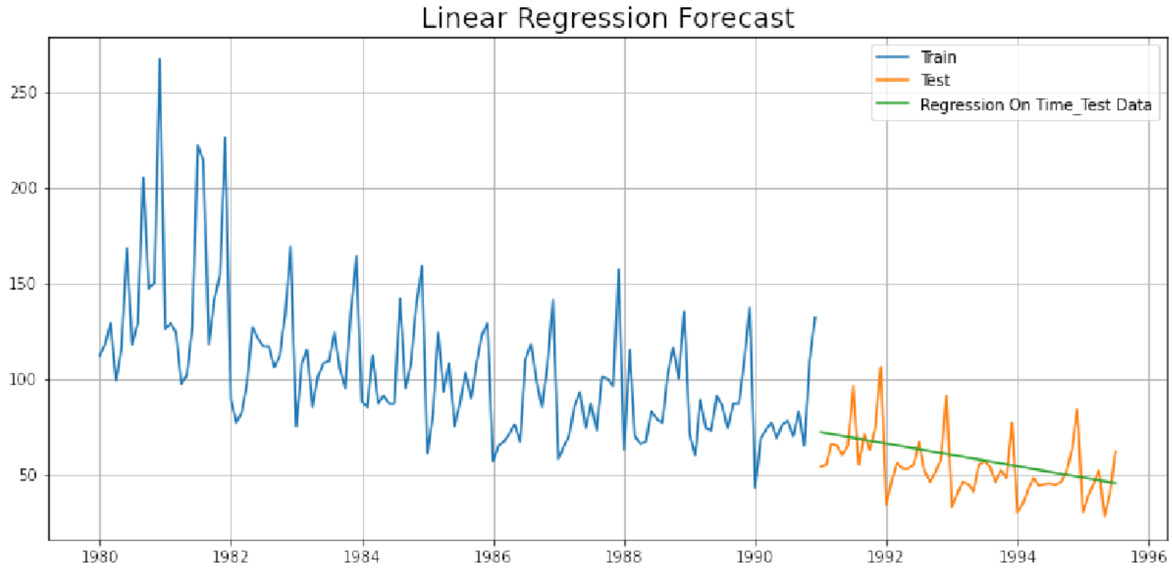


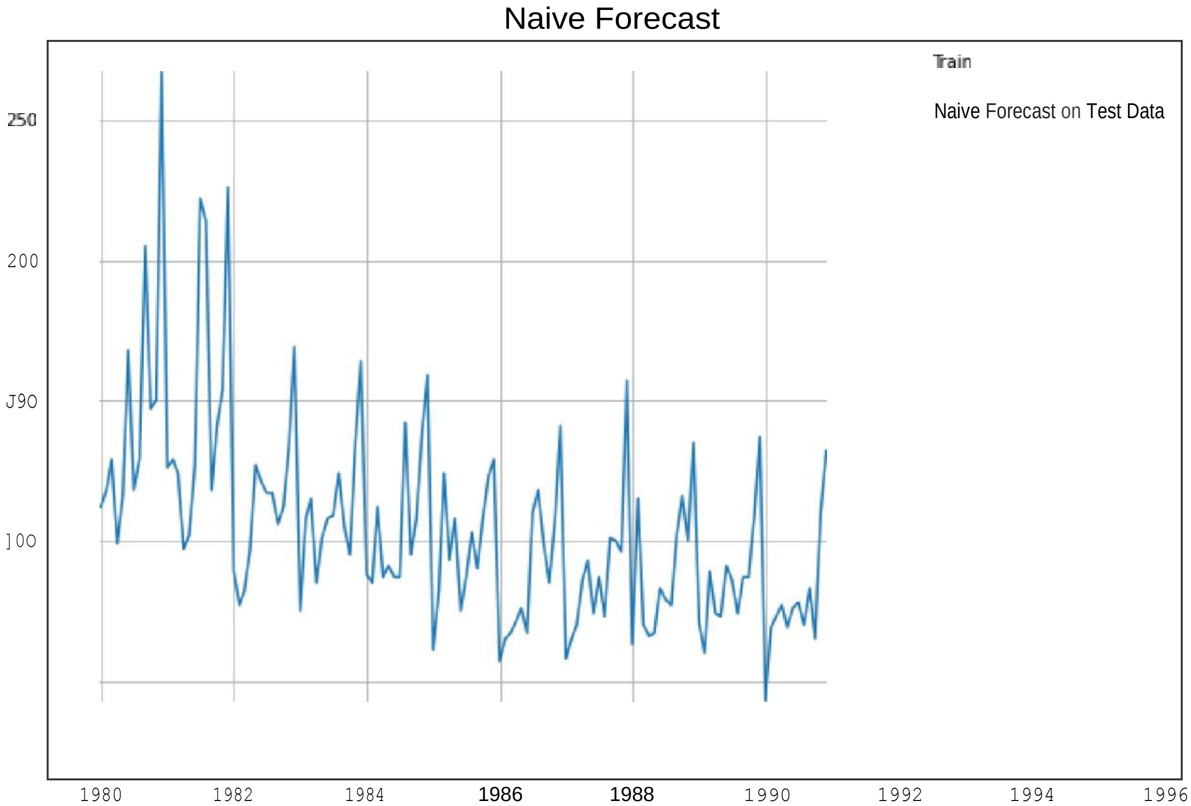
4. Build various exponential smoothing models on the training data and evaluate the model using RMSE on the test data. Other models such as regression,naïve forecast models, simple average models etc. should also be built on the training data and check the performance on the test data using RMSE.

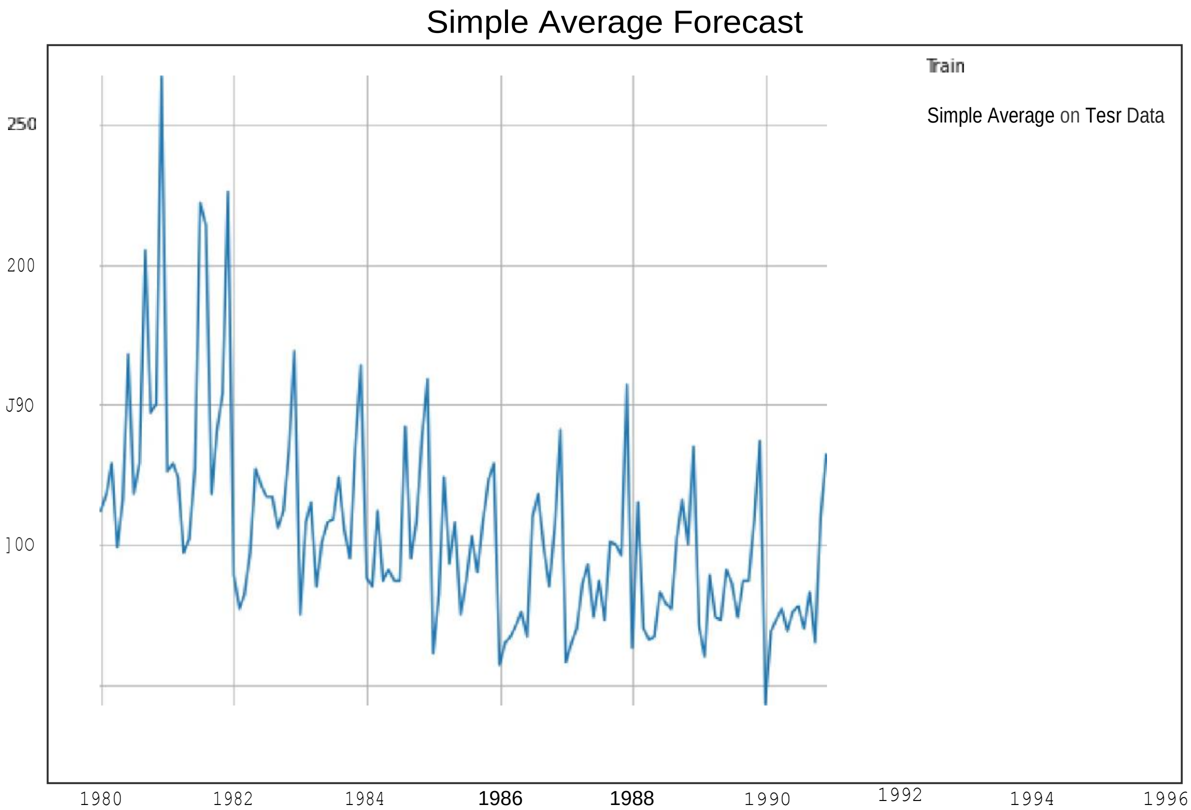
Please do try to build as many models as possible and as many iterations of models as possible with different parameters.

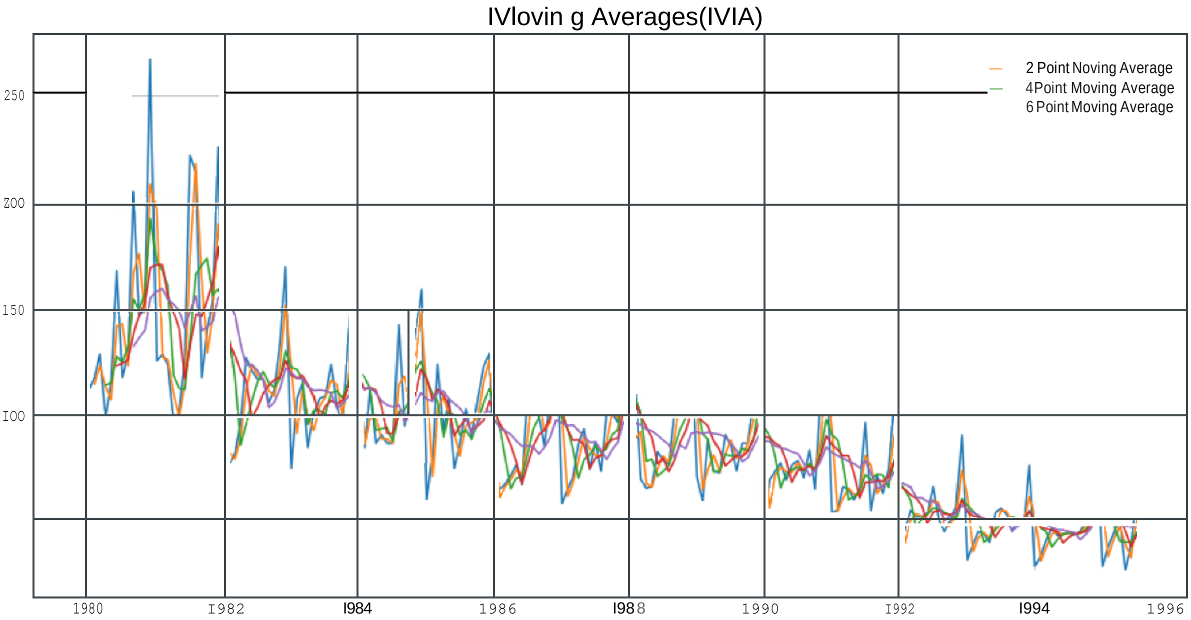
Various models and forecast as below.

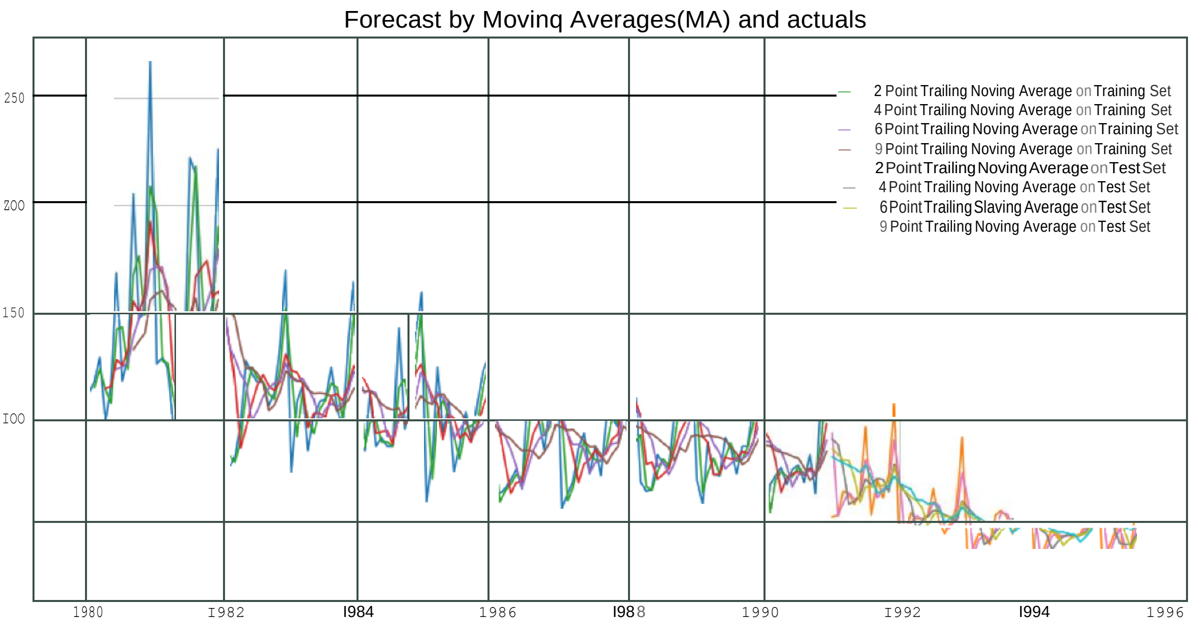
Rose wine sales

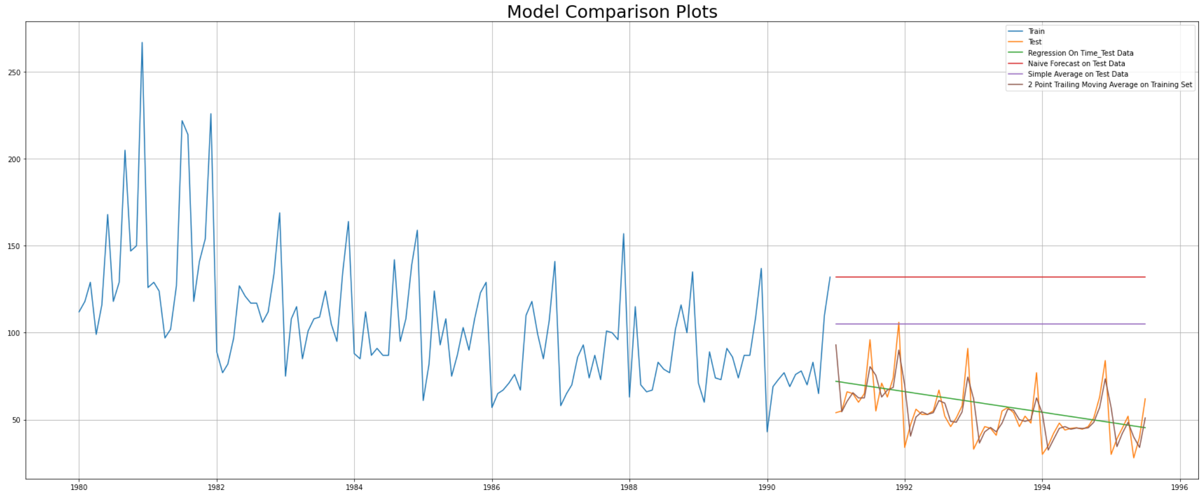


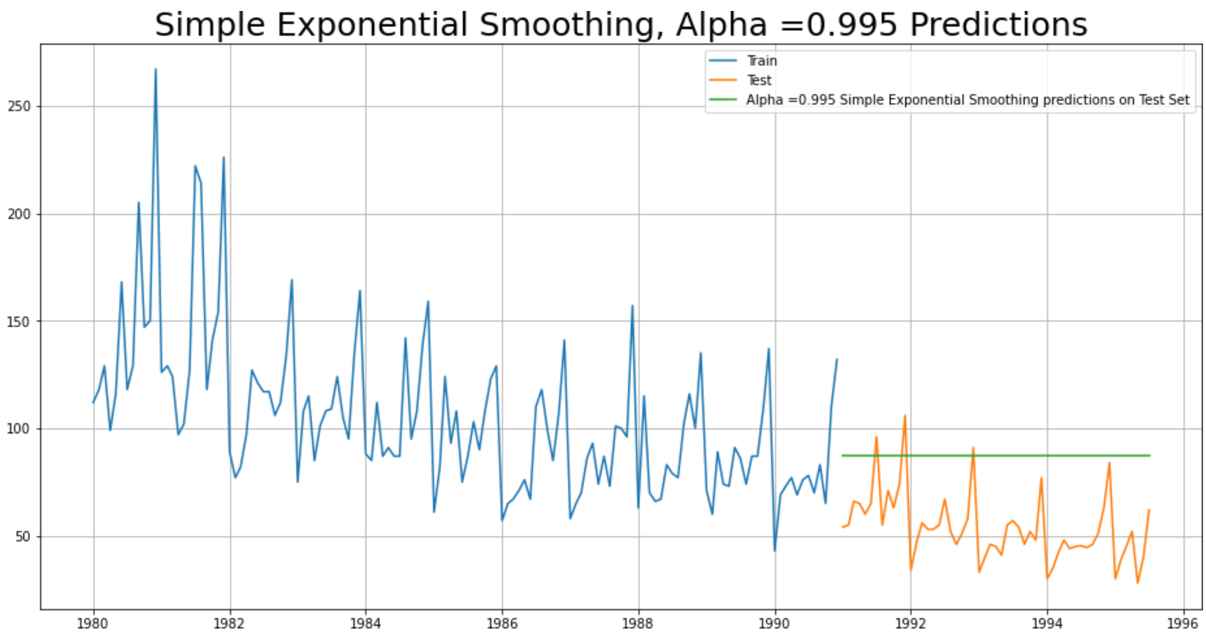

Naive Forecast
Train
Naive Forecast on Test Data
250
200
J90
]00
1980
1982
1984
1986
1988
1990
1992
1994
1996


Simple Average Forecast
Train
Simple Average on Tesr Data
250
200
J90
]00
1980
1982
1984
1986
1988
1990
1992
1994
1996


IVlovin g Averages(IVIA)
2 Point Noving Average
4Point Moving Average
6 Point Moving Average
250
Z00
150
I00
1980
I982
I984
1986
I988
1990
I992
I994
1996


Forecast by Movinq Averages(MA) and actuals
2 Point Trailing Noving Average on Training Set
4 Point Trailing Noving Average on Training Set
6 Point Trailing Noving Average on Training Set
9 Point Trailing Noving Average on Training Set
2 Point Trailing Noving Average on Test Set
4 Point Trailing Noving Average on Test Set
6 Point Trailing Slaving Average on Test Set
9 Point Trailing Noving Average on Test Set
250
Z00
150
I00
1980
I982
I984
1986
I988
1990
1992
I994
1996



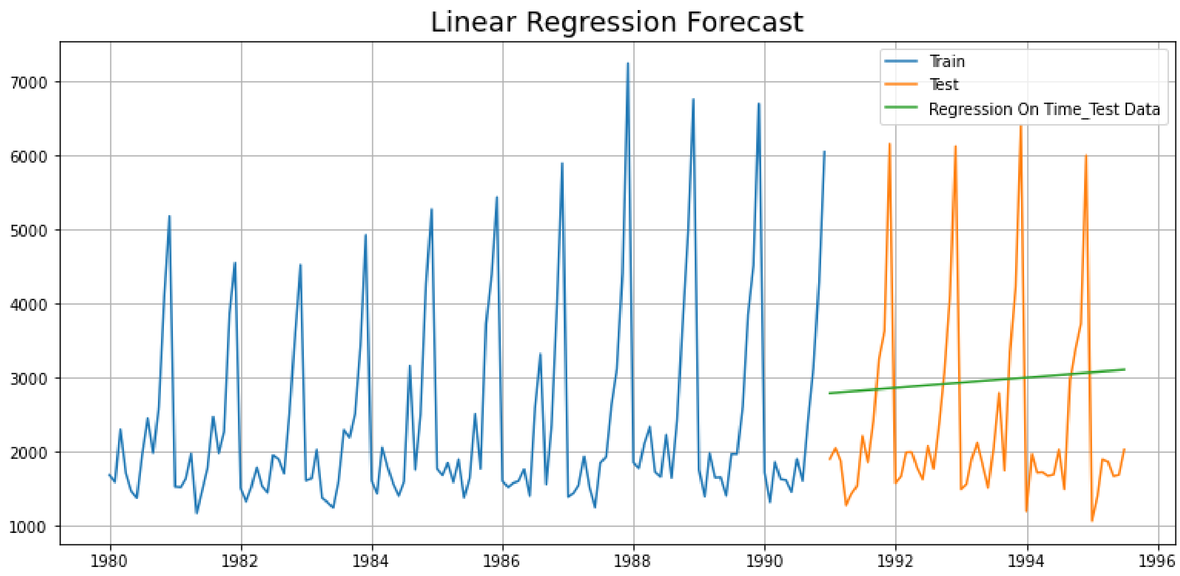


Various test results and RMSE are as under.

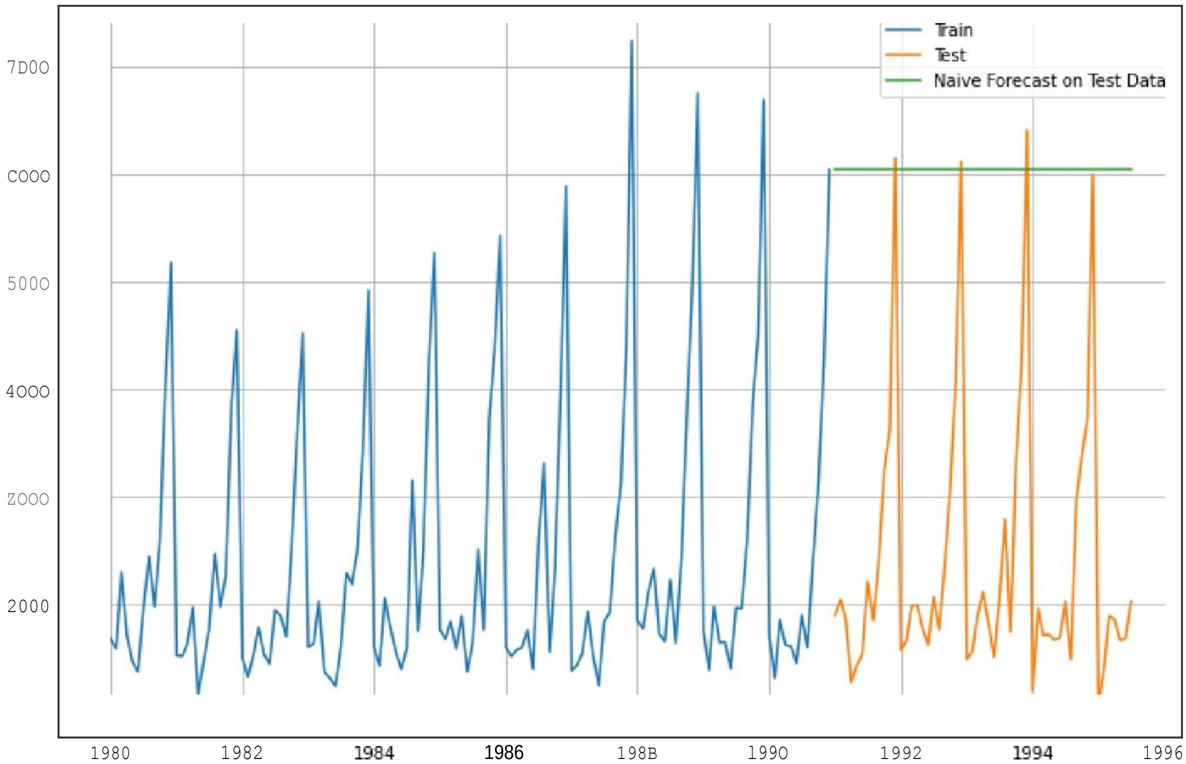
Test	RMSE
RegressionOnTime	15.280000
NaiveModel	79.741326
SimpleAverageModel	53.483727
2pointTrailingMovingAverage	11.529811
4pointTrailingMovingAverage	14.457115
6pointTrailingMovingAverage	14.571789
9pointTrailingMovingAverage	14.731914

Alpha=0.995,SimpleExponentialSmoothing	36.819844
Alpha=0.995,Beta=0.995:DoubleExponentialSmoothing	15.276679
Alpha=0.99,Beta=0.0001,Gamma=0.005:DoubleExponentialSmoothing	20.962011
Alpha=0.02,SimpleExponentialSmoothing	36.459396

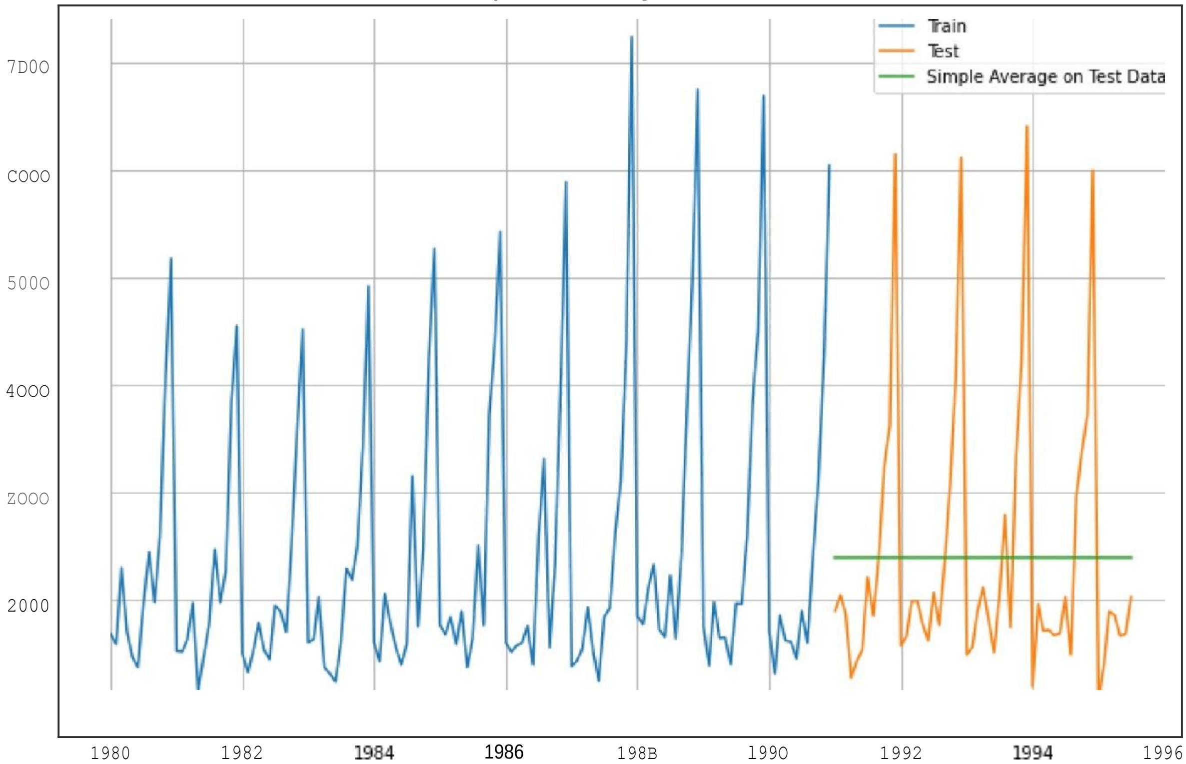
Sparkling wine sales



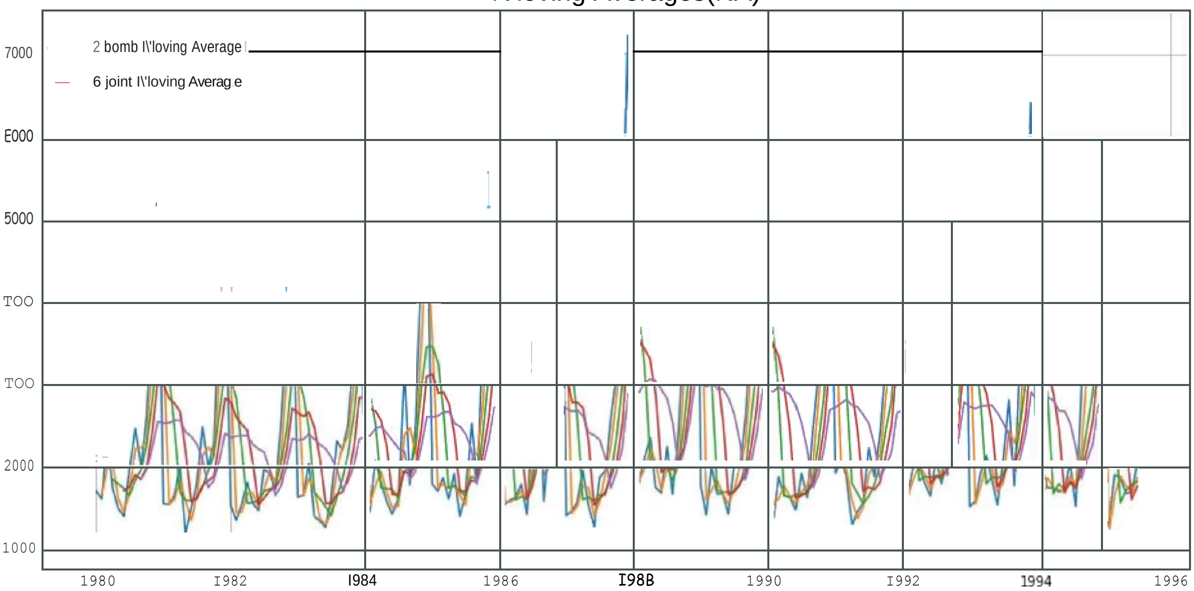
Naive Forecast



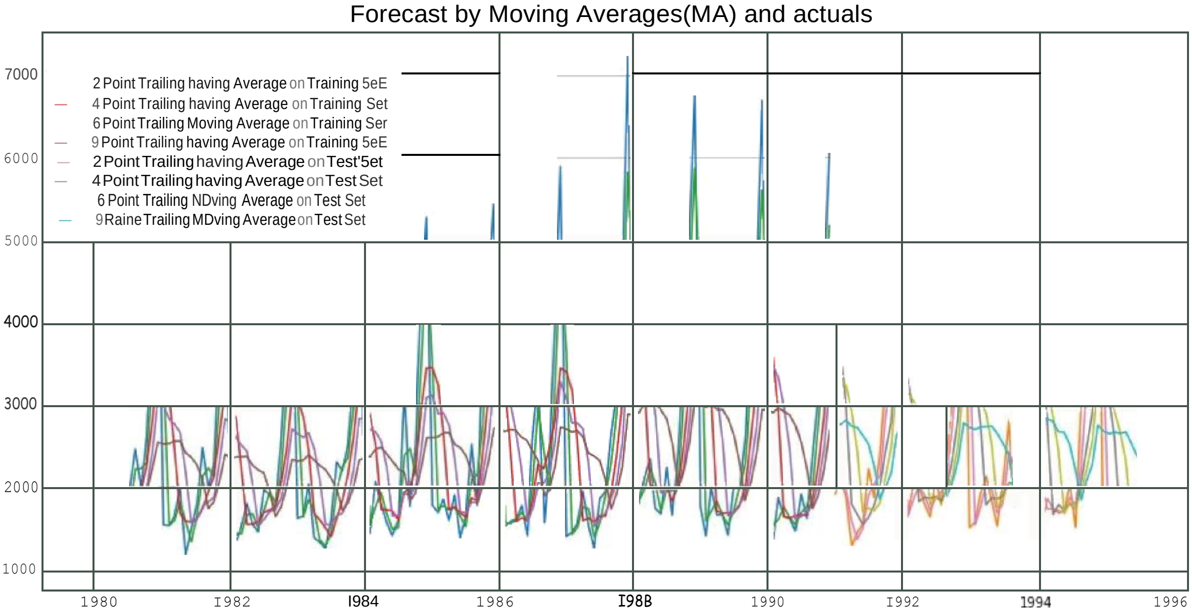
Simple Average Forecast

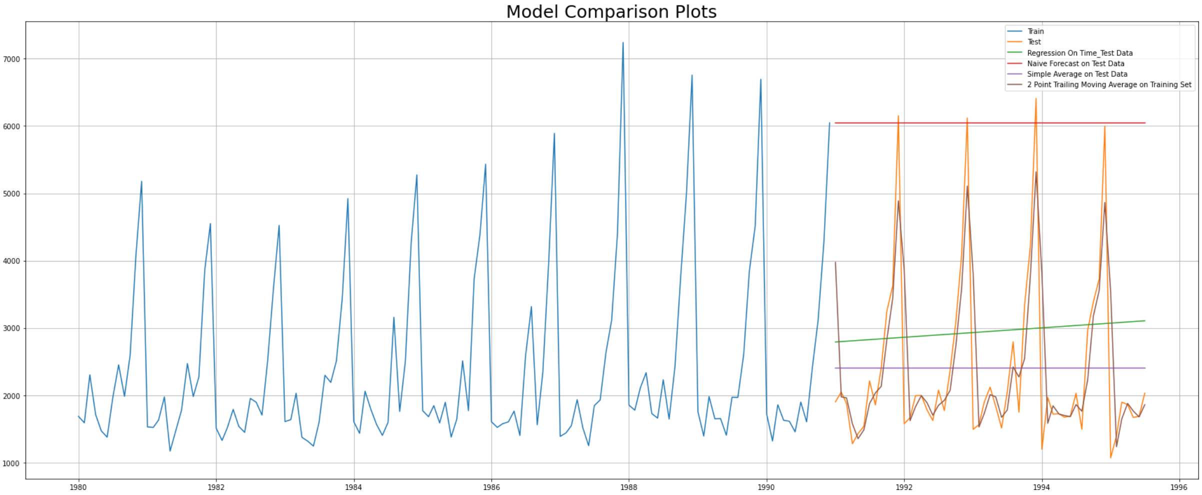


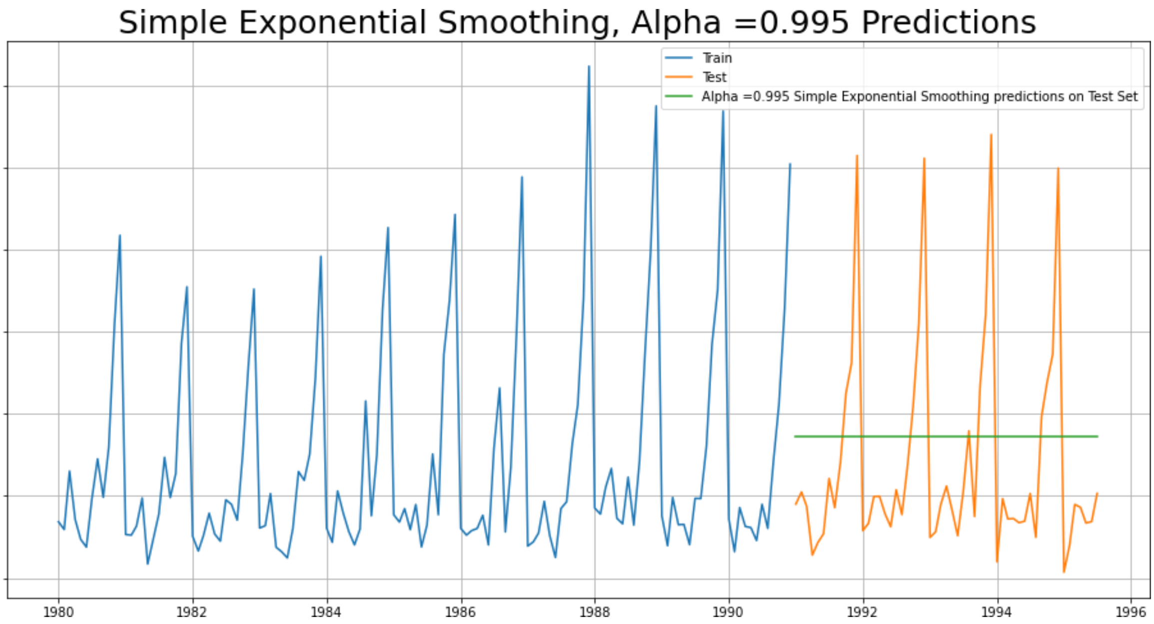
IVloving Averages(NA)



Forecast by Moving Averages(MA) and actuals







RMSE for models are as below.

	Test RMSE
RegressionOnTime	1389.140000
NaiveModel	3864.279352
SimpleAverageModel	1275.081804
2pointTrailingMovingAverage	813.400684
4pointTrailingMovingAverage	1156.589694
6pointTrailingMovingAverage	1283.927428

	Test RMSE
9pointTrailingMovingAverage	1346.278315
Alpha=0.995,SimpleExponentialSmoothing	1316.034674
Alpha=0.995,Beta=0.995:DoubleExponentialSmoothing	2007.238526
Alpha=0.99,Beta=0.0001,Gamma=0.005:DoubleExponentialSmoothing	469.591976
Alpha=0.02,SimpleExponentialSmoothing	1279.495201

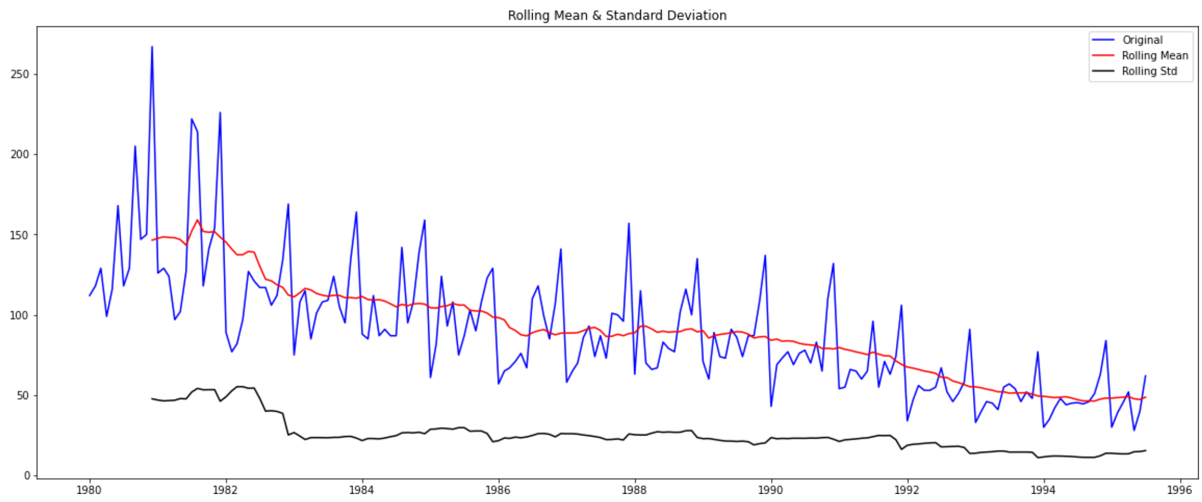
5. Check for the stationarity of the data on which the model is being built on using appropriate statistical tests and also mention the hypothesis for the statistical test. If the data is found to be non-stationary, take appropriate steps to make it stationary. Check the new data for stationarity and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

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

•Null Hypothesis H_0: Time Series is non-stationary

•Alternate Hypothesis H_a: Time Series is stationary

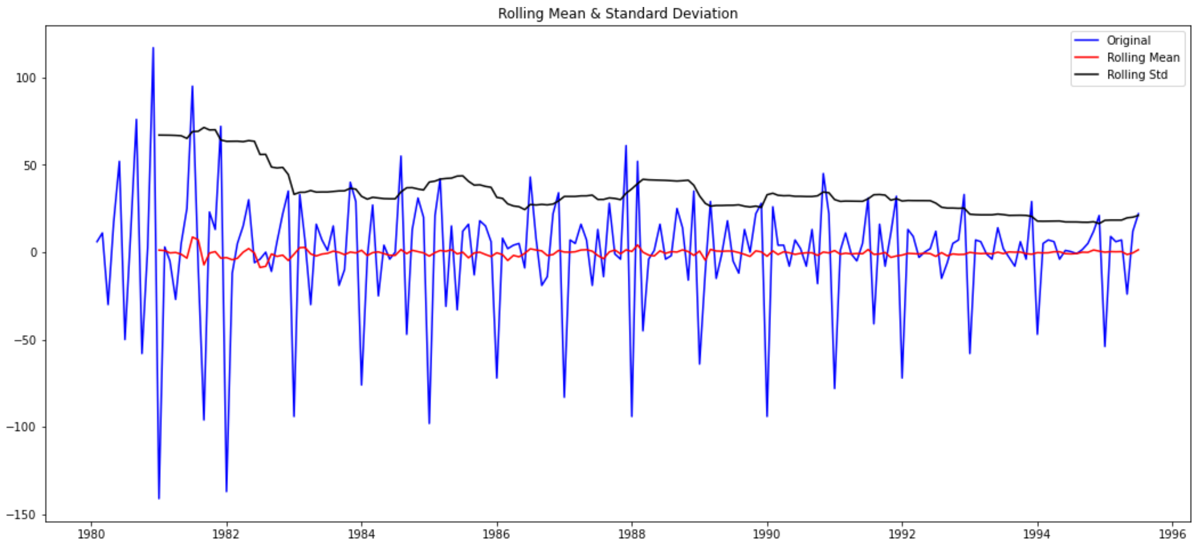
Rose wine sales




```
Results of Dickey-Fuller Test:
Test Statistic                  -1.873307
p-value                          0.344721
#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
```

since p-value> 0.05, at alpha 0.05, time series is not stationary.

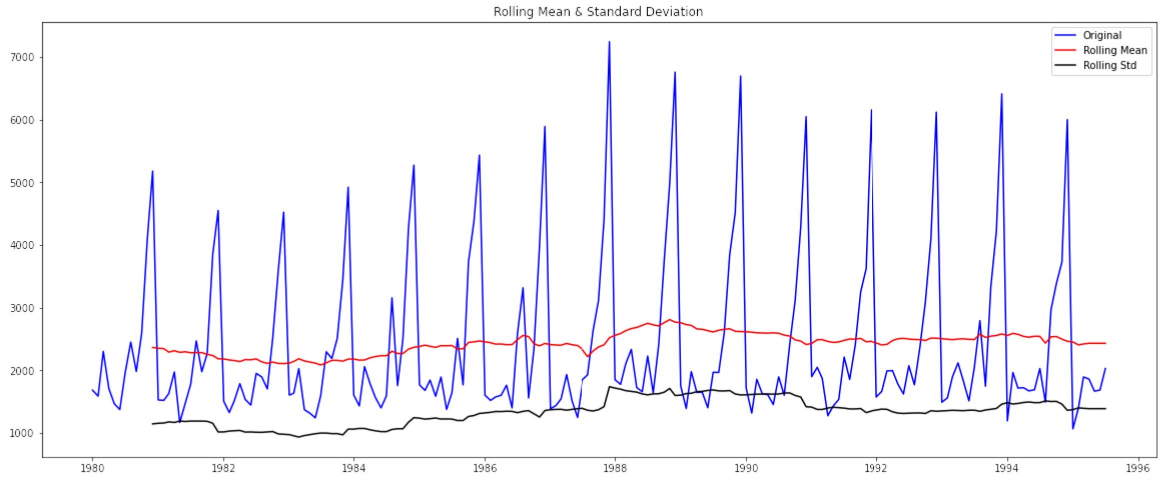
We can take next levels of differencing to make a Time Series stationary.




```
Results of Dickey-Fuller Test:
Test Statistic                -8.044136e+00
p-value                        1.813615e-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
```

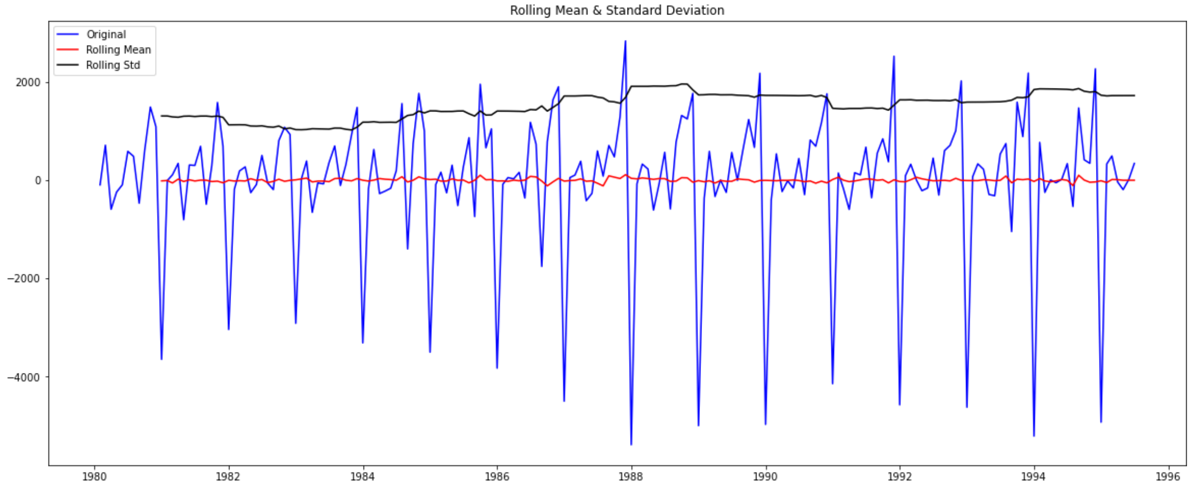
After next level of levels of differencing p-value <0.05 therefore series is stationary.

Sparkling wine sales




```
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
```

since p-value> 0.05, at alpha 0.05, time series is not stationary.
We can take next levels of differencing to make a Time Series 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
```

After next level of levels of differencing p-value <0.05 therefore series is stationary.

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.

Rose wine sales

ARIMA Model
AIC values in descending order

	param	AIC
17	(3, 1, 3)	1273.194108
4	(1, 1, 2)	1277.359223
3	(1, 1, 1)	1277.775747

	param	AIC
9	(2, 1, 1)	1279.045689
10	(2, 1, 2)	1279.298694
5	(1, 1, 3)	1279.312635
15	(3, 1, 1)	1279.605966
16	(3, 1, 2)	1280.969245
11	(2, 1, 3)	1281.196226
1	(1, 0, 2)	1292.053210
7	(2, 0, 2)	1292.248055
2	(1, 0, 3)	1292.929011
6	(2, 0, 1)	1292.937195
14	(3, 0, 3)	1293.042709
8	(2, 0, 3)	1294.247938
0	(1, 0, 1)	1294.510585
12	(3, 0, 1)	1333.933193
13	(3, 0, 2)	1355.403813

Lowest AIC is: 1273.194 with **param (3,1,3)**

```
                              ARIMA Model Results
==============================================================================
Dep. Variable:                 D.Rose   No. Observations:                  131
Model:                 ARIMA(3, 1, 3)   Log Likelihood                -628.597
Method:                       css-mle   S.D. of innovations             28.355
Date:                Sat, 06 Mar 2021   AIC                           1273.194
```

```
Time:                        14:46:16   BIC                           1296.196
Sample:                    02-01-1980   HQIC                          1282.541
                         - 12-01-1990
======================== ==================================================
                     coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const             -0.4906      0.088     -5.548      0.000      -0.664      -0.317
ar.L1.D.Rose      -0.7244      0.086     -8.417      0.000      -0.893      -0.556
ar.L2.D.Rose      -0.7218      0.086     -8.349      0.000      -0.891      -0.552
ar.L3.D.Rose       0.2763      0.085      3.236      0.001       0.109       0.444
ma.L1.D.Rose      -0.0150      0.044     -0.338      0.735      -0.102       0.072
ma.L2.D.Rose       0.0150      0.044      0.339      0.734      -0.072       0.102
ma.L3.D.Rose      -1.0000      0.046    -21.918      0.000      -1.089      -0.911
                                    Roots
=============================================================================
                  Real          Imaginary           Modulus         Frequency
-----------------------------------------------------------------------------
AR.1           -0.5011            -0.8661j            1.0006           -0.3335
AR.2           -0.5011            +0.8661j            1.0006            0.3335
AR.3            3.6147            -0.0000j            3.6147           -0.0000
MA.1            1.0000            -0.0000j            1.0000           -0.0000
MA.2           -0.4925            -0.8703j            1.0000           -0.3320
MA.3           -0.4925            +0.8703j            1.0000            0.3320
-----------------------------------------------------------------------------
```

Evaluation of model using RMSE: RMSE from ARIMA model 15.99

SARIMA model

AIC values in descending order (lowest AIC 10 records)

	param	seasonal	AIC
107	(0, 1, 2)	(2, 1, 2, 12)	774.969119
215	(1, 1, 2)	(2, 1, 2, 12)	776.940114
323	(2, 1, 2)	(2, 1, 2, 12)	776.996102
269	(2, 0, 2)	(2, 1, 2, 12)	780.716942
161	(1, 0, 2)	(2, 1, 2, 12)	780.992971
89	(0, 1, 1)	(2, 1, 2, 12)	782.153872
322	(2, 1, 2)	(2, 1, 1, 12)	783.703652
197	(1, 1, 1)	(2, 1, 2, 12)	783.899095

	param	seasonal	AIC
95	(0, 1, 2)	(0, 1, 2, 12)	784.014096
311	(2, 1, 2)	(0, 1, 2, 12)	784.140949

Lowest AIC is : 774.969 with **param (0,1,2) and seasonal (2,1,2,12)**

```
                                     SARIMAX Results
==========================================================================================
Dep. Variable:                              y   No. Observations:                   132
Model:       ARIMAX(0, 1, 2)x(2, 1, 2, 12)    Log Likelihood                   -380.485
Date:                        Sat, 06 Mar 2021   AIC                               774.969
Time:                                13:41:44   BIC                               792.622
Sample:                                     0   HQIC                              782.094
                                        - 132
Covariance Type:                          opg
===============================================================================
                  coef    std err          z      P>|z|      [0.025      0.975]
-------------------------------------------------------------------------------
ma.L1          -0.9524      0.184     -5.166      0.000      -1.314      -0.591
ma.L2          -0.0764      0.126     -0.605      0.545      -0.324       0.171
ar.S.L12        0.0480      0.177      0.271      0.786      -0.299       0.395
ar.S.L24       -0.0419      0.028     -1.513      0.130      -0.096       0.012
ma.S.L12       -0.7526      0.301     -2.503      0.012      -1.342      -0.163
ma.S.L24       -0.0721      0.204     -0.354      0.723      -0.471       0.327
sigma2        187.8679     45.274      4.150      0.000      99.132     276.604
==========================================================================================
Ljung-Box (L1) (Q):                    0.06   Jarque-Bera (JB):                      4.86
Prob(Q):                               0.81   Prob(JB):                              0.09
Heteroskedasticity (H):                0.91   Skew:                                  0.41
Prob(H) (two-sided):                   0.79   Kurtosis:                              3.77
==========================================================================================
```

Evaluation of model using RMSE : RMSE from SARIMA model 16.52

Sparkling wine sales

ARIMA Model

AIC values in descending order

	param	AIC
8	(2, 1, 2)	2210.616954
7	(2, 1, 1)	2232.360490

	param	AIC
2	(0, 1, 2)	2232.783098
5	(1, 1, 2)	2233.597647
4	(1, 1, 1)	2235.013945
6	(2, 1, 0)	2262.035601
1	(0, 1, 1)	2264.906439
3	(1, 1, 0)	2268.528061
0	(0, 1, 0)	2269.582796

Lowest AIC is 2232.783 with **param (2,1,2)**

```
                              ARIMA Model Results
==============================================================================
Dep. Variable:           D.Sparkling   No. Observations:                  131
Model:                ARIMA(2, 1, 2)   Log Likelihood               -1099.308
Method:                      css-mle   S.D. of innovations           1011.985
Date:               Sun, 28 Feb 2021   AIC                           2210.617
Time:                       18:02:49   BIC                           2227.868
Sample:                   02-01-1980   HQIC                          2217.627
                        - 12-01-1990
=====================================================================================
                        coef    std err          z      P>|z|      [0.025      0.975]
-------------------------------------------------------------------------------------
const                 5.5860      0.516     10.825      0.000       4.575       6.597
ar.L1.D.Sparkling     1.2698      0.074     17.045      0.000       1.124       1.416
ar.L2.D.Sparkling    -0.5601      0.074     -7.617      0.000      -0.704      -0.416
ma.L1.D.Sparkling    -1.9993      0.042    -47.149      0.000      -2.082      -1.916
ma.L2.D.Sparkling     0.9993      0.042     23.584      0.000       0.916       1.082
                                    Roots
=============================================================================
                  Real          Imaginary           Modulus         Frequency
-----------------------------------------------------------------------------
AR.1            1.1335           -0.7075j            1.3361           -0.0888
AR.2            1.1335           +0.7075j            1.3361            0.0888
MA.1            1.0002           +0.0000j            1.0002            0.0000
MA.2            1.0006           +0.0000j            1.0006            0.0000
-----------------------------------------------------------------------------
```

Evaluation of model using RMSE: RBSE for ARIMA model 1375.03

SARIMA model

AIC values in descending order (lowest AIC 10 records)

	param	seasonal	AIC
95	(1, 1, 2)	(0, 1, 2, 12)	1382.347780
41	(0, 1, 2)	(0, 1, 2, 12)	1382.484254
101	(1, 1, 2)	(1, 1, 2, 12)	1384.137874
149	(2, 1, 2)	(0, 1, 2, 12)	1384.317618
47	(0, 1, 2)	(1, 1, 2, 12)	1384.398867
107	(1, 1, 2)	(2, 1, 2, 12)	1385.688721
53	(0, 1, 2)	(2, 1, 2, 12)	1386.023734
155	(2, 1, 2)	(1, 1, 2, 12)	1386.097242
161	(2, 1, 2)	(2, 1, 2, 12)	1387.627785
77	(1, 1, 1)	(0, 1, 2, 12)	1398.756167

Lowest AIC is : 1382.347780 with **param (1,1,2) and seasonal 0,1,2,12)**

```
                                     SARIMAX Results
==================================================================================
Dep. Variable:                                 y   No. Observations:             132
Model:         SARIMAX(0, 1, 2)x(0, 1, 2, 12)   Log Likelihood            -686.242
Date:                        Sun, 28 Feb 2021   AIC                       1382.484
Time:                                18:06:38   BIC                       1395.093
Sample:                                     0   HQIC                      1387.573
                                        - 132
Covariance Type:                          opg
==================================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
----------------------------------------------------------------------------------
ma.L1         -0.7223      0.107     -6.752      0.000      -0.932      -0.513
ma.L2         -0.2292      0.137     -1.671      0.095      -0.498       0.040
ma.S.L12      -0.4113      0.087     -4.743      0.000      -0.581      -0.241
ma.S.L24      -0.0419      0.138     -0.304      0.761      -0.312       0.228
sigma2      1.736e+05   2.06e+04      8.425      0.000    1.33e+05    2.14e+05
==================================================================================
```

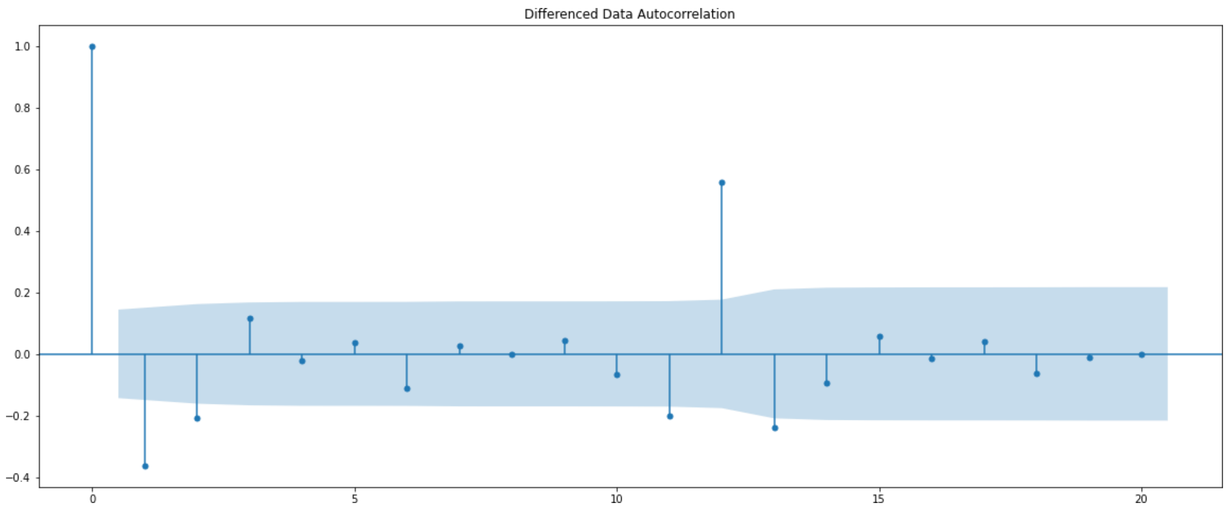
```
Ljung-Box (L1) (Q):                   0.02   Jarque-Bera (JB):                27.42
Prob(Q):                              0.88   Prob(JB):                         0.00
Heteroskedasticity (H):               0.84   Skew:                             0.80
Prob(H) (two-sided):                  0.62   Kurtosis:                         5.15
===================================================================================
```

Evaluation of model using RMSE: RBSE for ARIMA model `321.48`

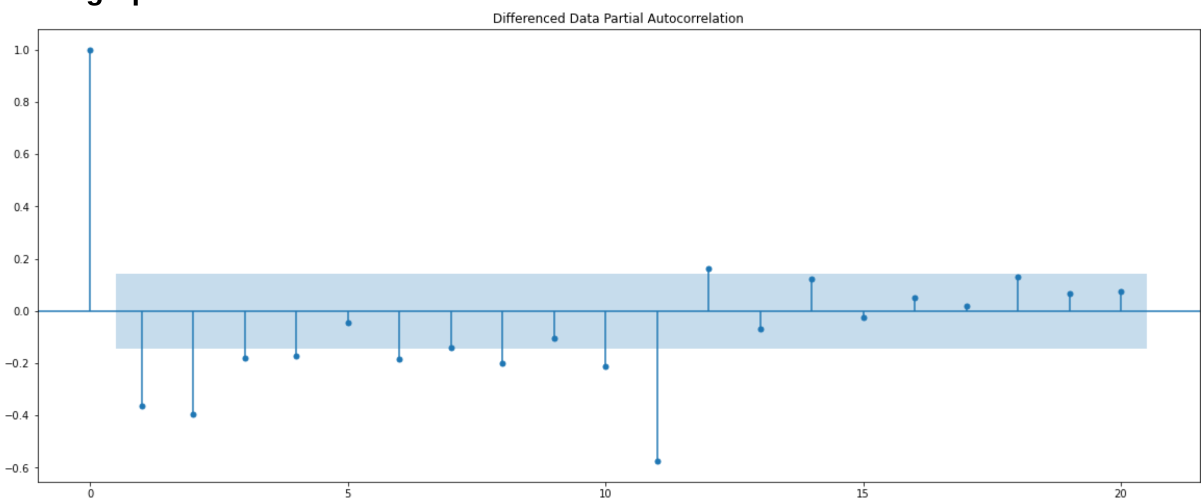
7. Build ARIMA/SARIMA models based on the cut-off points of ACF and PACF on the training data and evaluate this model on the test data using RMSE.

Rose wine sales

ACF graph for Rose wine sales as under



PACF graph for Rose wine sales as under



It can be observed from above the ACF and PACF plots, the cut off points for p and q for ARIMA model is 4 and 2 respectively.

By taking these parameters (4,1,2) ARIMA results are as under

```
                             ARIMA Model Results
==============================================================================
Dep. Variable:                 D.Rose   No. Observations:                  131
Model:                 ARIMA(4, 1, 2)   Log Likelihood                -633.876
Method:                       css-mle   S.D. of innovations             29.793
Date:                Sun, 07 Mar 2021   AIC                           1283.753
Time:                        13:23:57   BIC                           1306.754
Sample:                    02-01-1980   HQIC                          1293.099
                         - 12-01-1990
=================================================================================
                    coef    std err          z      P>|z|      [0.025      0.975]
---------------------------------------------------------------------------------
const            -0.1905      0.576     -0.331      0.741      -1.319       0.938
ar.L1.D.Rose      1.1685      0.087     13.391      0.000       0.997       1.340
ar.L2.D.Rose     -0.3562      0.132     -2.693      0.007      -0.616      -0.097
ar.L3.D.Rose      0.1855      0.132      1.402      0.161      -0.074       0.445
ar.L4.D.Rose     -0.2227      0.091     -2.443      0.015      -0.401      -0.044
ma.L1.D.Rose     -1.9506        nan        nan        nan         nan         nan
ma.L2.D.Rose      1.0000        nan        nan        nan         nan         nan
                                    Roots
=============================================================================
                  Real          Imaginary           Modulus         Frequency
-----------------------------------------------------------------------------
AR.1            1.1027           -0.4116j            1.1770           -0.0569
AR.2            1.1027           +0.4116j            1.1770            0.0569
AR.3           -0.6863           -1.6643j            1.8003           -0.3122
AR.4           -0.6863           +1.6643j            1.8003            0.3122
MA.1            0.9753           -0.2209j            1.0000           -0.0355
MA.2            0.9753           +0.2209j            1.0000            0.0355
-----------------------------------------------------------------------------
```

RMSE from ARIMA model is 33.97

By taking these parameters (4,1,2) and (4,1,2,12) SARIMA results are as under.

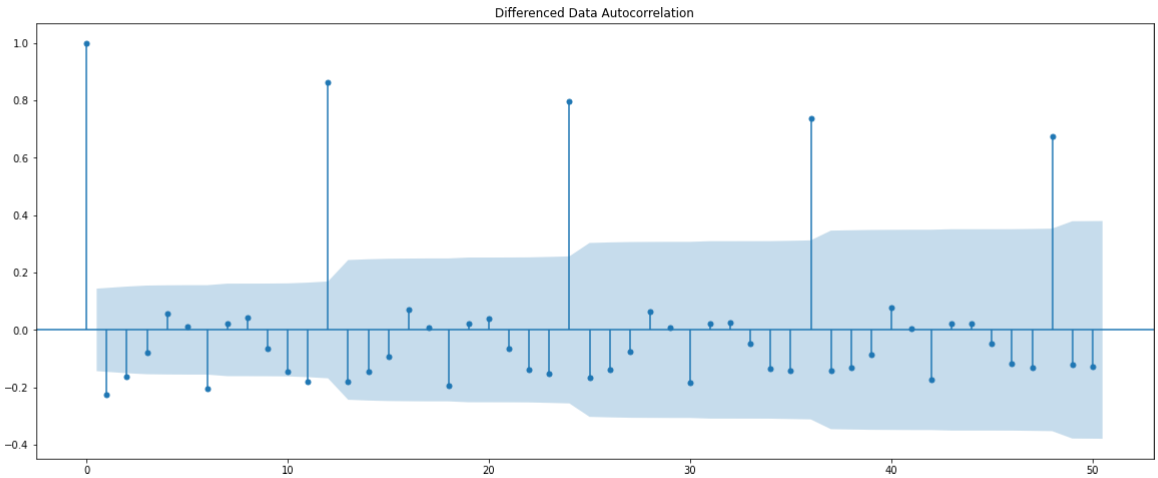
```
                                    SARIMAX Results
=================================================================================
Dep. Variable:                                 y   No. Observations:          132
Model:          SARIMAX(4, 1, 2)x(4, 1, 2, 12)   Log Likelihood            -277.661
Date:                         Sun, 07 Mar 2021   AIC                        581.322
Time:                                 14:22:21   BIC                        609.983
Sample:                                      0   HQIC                       592.663
                                         - 132
Covariance Type:                           opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.9742      0.199     -4.899      0.000      -1.364      -0.584
```

```
ar.L2          -0.1122      0.285     -0.394      0.694      -0.670       0.446
ar.L3          -0.1044      0.277     -0.377      0.706      -0.647       0.438
ar.L4          -0.1285      0.162     -0.794      0.427      -0.446       0.189
ma.L1           0.1605    328.137      0.000      1.000    -642.976     643.297
ma.L2          -0.8395    275.462     -0.003      0.998    -540.734     539.055
ar.S.L12       -0.1441      0.364     -0.396      0.692      -0.858       0.569
ar.S.L24       -0.3597      0.227     -1.587      0.113      -0.804       0.085
ar.S.L36       -0.2153      0.106     -2.039      0.041      -0.422      -0.008
ar.S.L48       -0.1195      0.093     -1.281      0.200      -0.302       0.063
ma.S.L12       -0.5159      0.343     -1.503      0.133      -1.189       0.157
ma.S.L24        0.2086      0.373      0.559      0.576      -0.523       0.940
sigma2        215.3512   7.07e+04      0.003      0.998   -1.38e+05    1.39e+05
===================================================================================
Ljung-Box (L1) (Q):                   0.03   Jarque-Bera (JB):                 2.41
Prob(Q):                              0.86   Prob(JB):                         0.30
Heteroskedasticity (H):               0.49   Skew:                             0.32
Prob(H) (two-sided):                  0.10   Kurtosis:                         3.68
===================================================================================
```

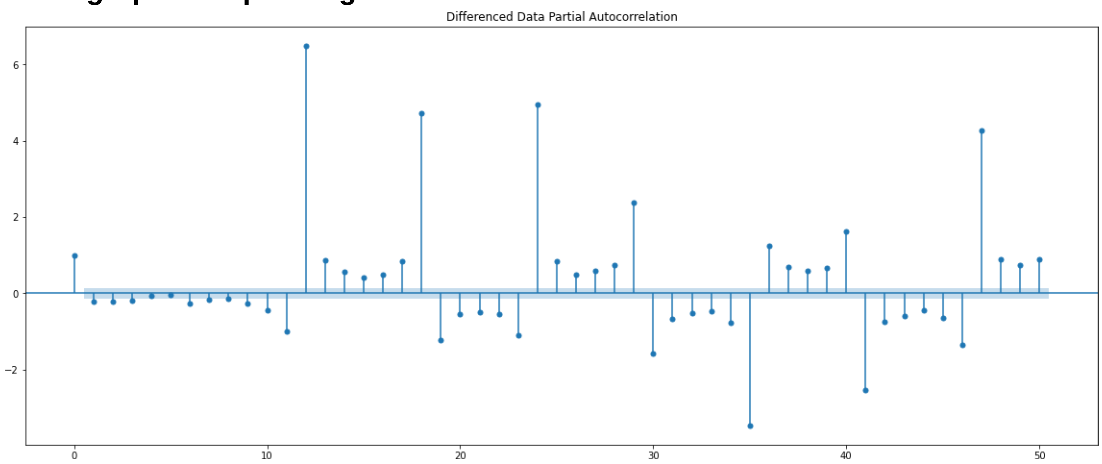
RMSE from SARIMA model is 17.54

Sparkling wine sales

ACF graph for Sparkling wine sales as under



PACF graph for Sparkling wine sales as under



It can be observed from above ACF and PACF plots, the cut off points for p and q for ARIMA model is 3 and 2 respectively.

By taking these parameters (3,1,2) ARIMA results are as under

```
                             ARIMA Model Results
==============================================================================
Dep. Variable:           D.Sparkling   No. Observations:                  131
Model:                ARIMA(3, 1, 2)   Log Likelihood               -1107.464
Method:                      css-mle   S.D. of innovations           1106.033
Date:               Sun, 07 Mar 2021   AIC                            2228.928
Time:                       14:58:55   BIC                            2249.054
Sample:                   02-01-1980   HQIC                           2237.106
                        - 12-01-1990
=====================================================================================
                        coef    std err          z      P>|z|      [0.025      0.975]
-------------------------------------------------------------------------------------
const                 5.8816        nan        nan        nan         nan         nan
ar.L1.D.Sparkling    -0.4422        nan        nan        nan         nan         nan
ar.L2.D.Sparkling     0.3075   7.77e-06   3.96e+04      0.000       0.308       0.308
ar.L3.D.Sparkling    -0.2503        nan        nan        nan         nan         nan
ma.L1.D.Sparkling    -0.0004      0.028     -0.013      0.990      -0.055       0.054
ma.L2.D.Sparkling    -0.9996      0.028    -36.010      0.000      -1.054      -0.945
                                    Roots
=============================================================================
                  Real          Imaginary           Modulus         Frequency
-----------------------------------------------------------------------------
AR.1           -1.0000           -0.0000j            1.0000           -0.5000
AR.2            1.1145           -1.6595j            1.9990           -0.1559
AR.3            1.1145           +1.6595j            1.9990            0.1559
MA.1            1.0000           +0.0000j            1.0000            0.0000
MA.2           -1.0004           +0.0000j            1.0004            0.5000
-----------------------------------------------------------------------------
```

RMSE from ARIMA model is 1375.10

By taking these parameters (3,1,2) and (3,1,2,12) SARIMA results are as under.

```
                                     SARIMAX Results
=========================================================================================
Dep. Variable:                                 y   No. Observations:                  132
Model:        SARIMAX(3, 1, 2)x(3, 1, 2, 12)   Log Likelihood                -598.630
Date:                       Sun, 07 Mar 2021   AIC                            1219.260
Time:                               15:01:53   BIC                            1245.462
Sample:                                    0   HQIC                           1229.765
                                       - 132
Covariance Type:                         opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.7556      0.151     -5.013      0.000      -1.051      -0.460
ar.L2          0.1169      0.185      0.633      0.527      -0.245       0.479
ar.L3         -0.0520      0.143     -0.365      0.715      -0.332       0.228
ma.L1          0.0330      0.191      0.173      0.863      -0.341       0.407
ma.L2         -0.9670      0.156     -6.197      0.000      -1.273      -0.661
ar.S.L12      -0.7538      0.496     -1.520      0.128      -1.725       0.218
ar.S.L24      -0.6371      0.351     -1.818      0.069      -1.324       0.050
ar.S.L36      -0.2469      0.151     -1.641      0.101      -0.542       0.048
ma.S.L12       0.3719      0.491      0.758      0.448      -0.590       1.334
ma.S.L24       0.3466      0.365      0.949      0.343      -0.370       1.063
```

```
sigma2          1.79e+05   1.67e-06   1.07e+11      0.000    1.79e+05    1.79e+05
===================================================================================
Ljung-Box (L1) (Q):                   0.01   Jarque-Bera (JB):                13.16
Prob(Q):                              0.93   Prob(JB):                         0.00
Heteroskedasticity (H):               0.66   Skew:                             0.62
Prob(H) (two-sided):                  0.29   Kurtosis:                         4.55
===================================================================================
```
RMSE from SARIMA model is 329.53

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

RMSE for Rose for Exponential Smoothing Models

	Test RMSE
RegressionOnTime	15.280000
NaiveModel	79.741326
SimpleAverageModel	53.483727
2pointTrailingMovingAverage	11.529811
4pointTrailingMovingAverage	14.457115
6pointTrailingMovingAverage	14.571789
9pointTrailingMovingAverage	14.731914
Alpha=0.995,SimpleExponentialSmoothing	36.819844
Alpha=0.995,Beta=0.995:DoubleExponentialSmoothing	15.276679
Alpha=0.99,Beta=0.0001,Gamma=0.005:DoubleExponentialSmoothing	20.962011
Alpha=0.02,SimpleExponentialSmoothing	36.459396

RMSE for Rose for ARIMA/SARIMA

	Test RMSE
ARIMA(3, 1, 3)	15.99
SARIMA(0, 1, 2)x(2, 1, 2, 12)	16.52
ARIMA(4, 1, 2)	33.97
SARIMA(4, 1, 2)x(4, 1, 2, 12)	17.54

RMSE for Sparkling for Exponential Smoothing Models

	Test RMSE
RegressionOnTime	1389.140000
NaiveModel	3864.279352
SimpleAverageModel	1275.081804

	Test RMSE
2pointTrailingMovingAverage	813.400684
4pointTrailingMovingAverage	1156.589694
6pointTrailingMovingAverage	1283.927428
9pointTrailingMovingAverage	1346.278315
Alpha=0.995,SimpleExponentialSmoothing	1316.034674
Alpha=0.995,Beta=0.995:DoubleExponentialSmoothing	2007.238526
Alpha=0.99,Beta=0.0001,Gamma=0.005:DoubleExponentialSmoothing	469.591976
Alpha=0.02,SimpleExponentialSmoothing	1279.495201

RMSE for ARIMA Models for Sparkling

	Test RMSE
ARIMA(2, 1, 2)	1375.03
SARIMA(0, 1, 2)x (0, 1, 2, 12)	321.48
ARIMA(3, 1, 2)	1375.10
SARIMA(3, 1, 2)x(3, 1, 2, 12)	329.53

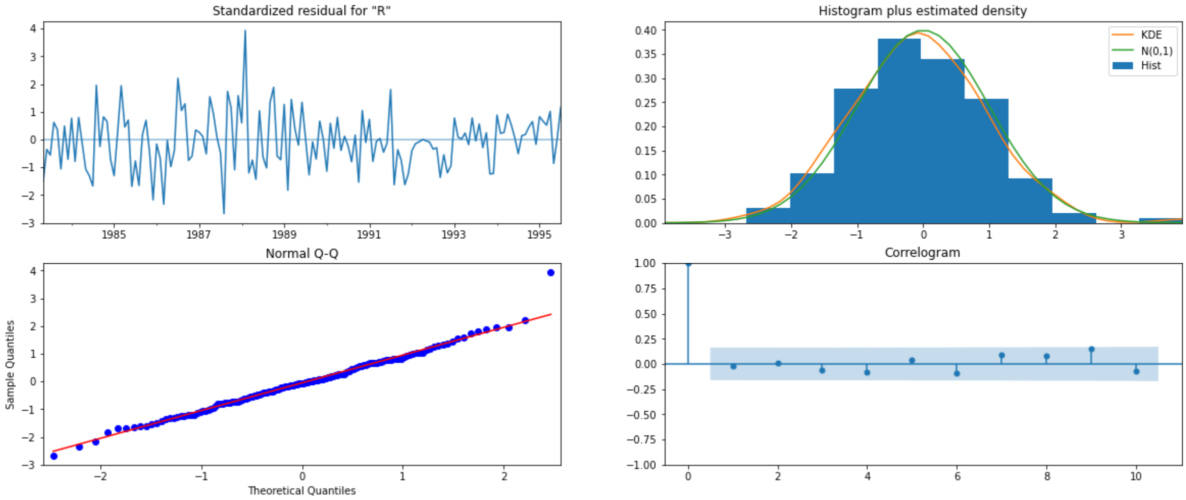
9. Based on the model-building exercise, build the most optimum model(s) on the complete data and predict 12 months into the future with appropriate confidence intervals/bands.

Since Rose data set has clear component of seasonality SARIM. Therefore, SARIMA model with parameters (0,1,2)x(2,1,2,12) is selected for forecasting time line series and model details are as under.

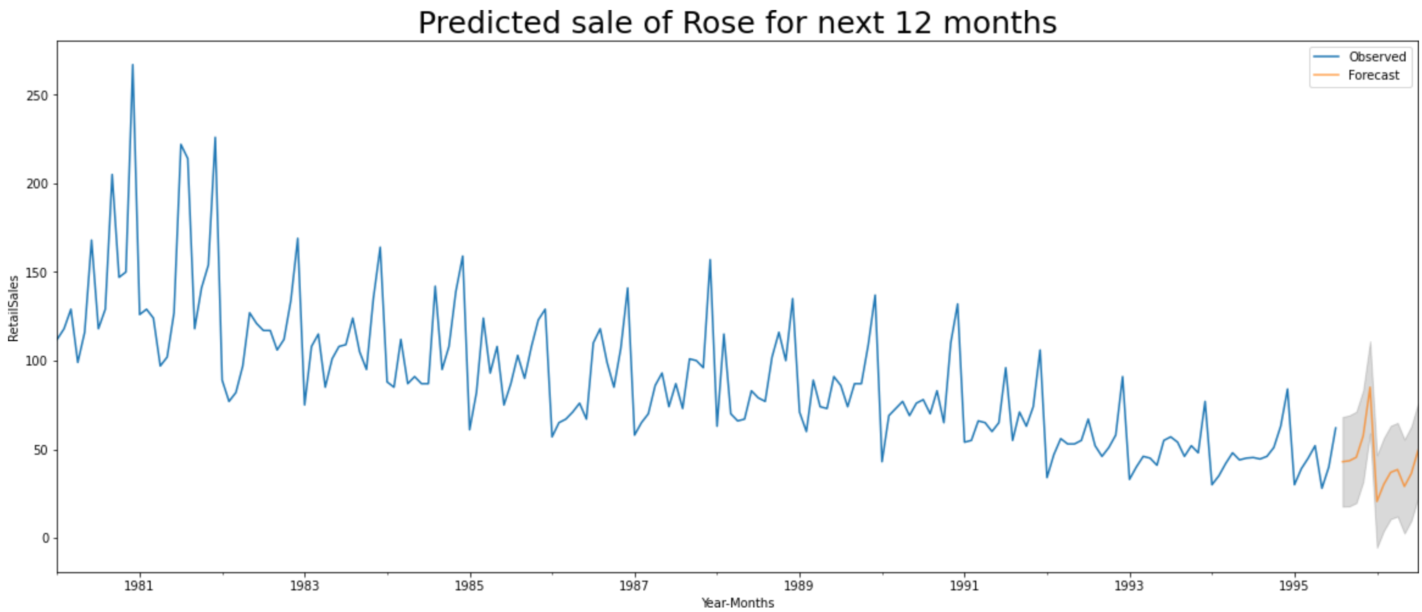
```
                                     SARIMAX Results
==========================================================================================
Dep. Variable:                          Rose   No. Observations:                     187
Model:      SARIMAX(0, 1, 2)x(2, 1, 2, 12)    Log Likelihood                   -588.604
Date:                       Sat, 06 Mar 2021   AIC                              1191.208
Time:                               13:41:48   BIC                              1212.142
Sample:                           01-01-1980   HQIC                             1199.714
                                - 07-01-1995
Covariance Type:                         opg
==================================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
----------------------------------------------------------------------------------
ma.L1         -0.8254      0.080    -10.334      0.000      -0.982      -0.669
ma.L2         -0.0807      0.086     -0.934      0.350      -0.250       0.089
ar.S.L12       0.0635      0.160      0.398      0.691      -0.249       0.376
ar.S.L24      -0.0340      0.019     -1.790      0.074      -0.071       0.003
ma.S.L12      -0.6953      0.207     -3.360      0.001      -1.101      -0.290
ma.S.L24      -0.0547      0.150     -0.365      0.715      -0.348       0.239
sigma2       166.0900     17.899      9.279      0.000     131.008     201.172
==========================================================================================
Ljung-Box (L1) (Q):                   0.07   Jarque-Bera (JB):                         8.28
```

```
Prob(Q):                              0.79   Prob(JB):                         0.02
Heteroskedasticity (H):               0.51   Skew:                             0.33
Prob(H) (two-sided):                  0.02   Kurtosis:                         3.95
```

Diagnostics of the Final Model for Rose



RMSE of full model is 33.47



The predicted sales of Rose for next 12 months is as below

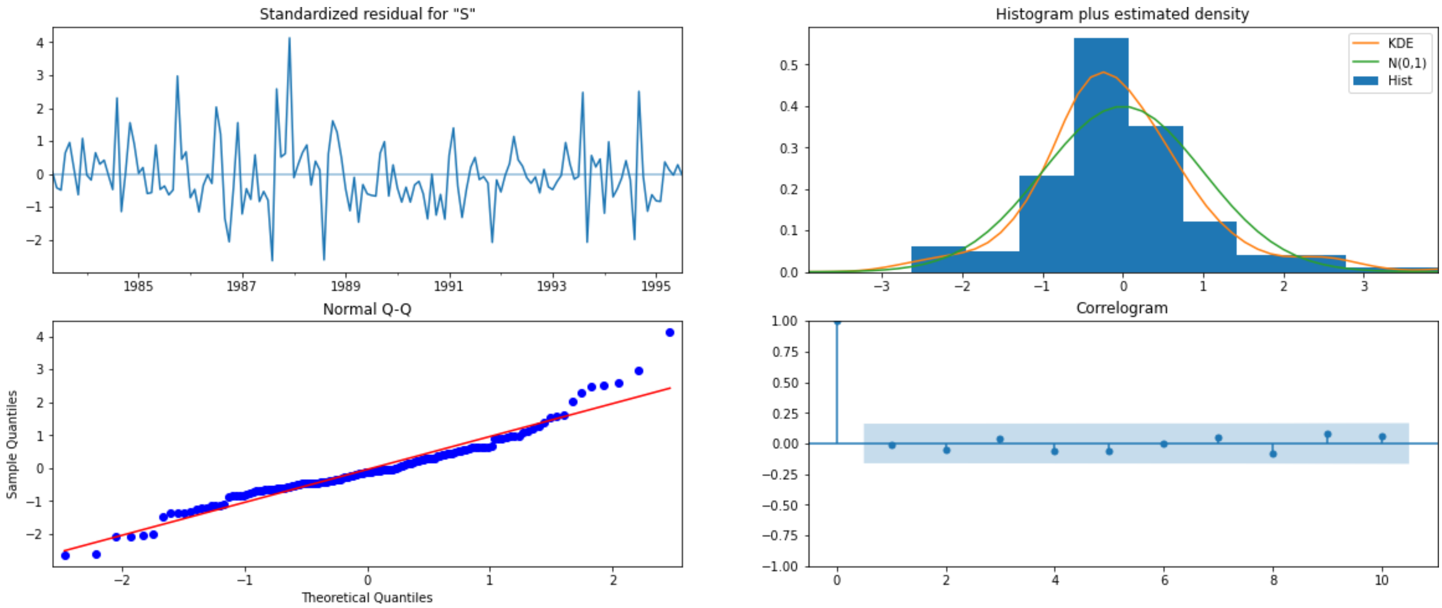
Rose	mean	mean_se	mean_ci_lower	mean_ci_upper
1995-08-01	42.984338	12.890006	17.720391	68.248285
1995-09-01	43.513258	13.085191	17.866755	69.159761
1995-10-01	45.491994	13.141097	19.735918	71.248071
1995-11-01	57.520151	13.196775	31.654948	83.385355
1995-12-01	84.989586	13.252239	59.015674	110.963498
1996-01-01	20.575007	13.307260	-5.506743	46.656757

Rose	mean	mean_se	mean_ci_lower	mean_ci_upper
1996-02-01	30.224797	13.362216	4.035335	56.414260
1996-03-01	36.974058	13.416865	10.677486	63.270631
1996-04-01	38.520738	13.471317	12.117442	64.924035
1996-05-01	29.043623	13.525605	2.533925	55.553321
1996-06-01	36.323188	13.579682	9.707500	62.938875
1996-07-01	49.477037	13.633557	22.755757	76.198318

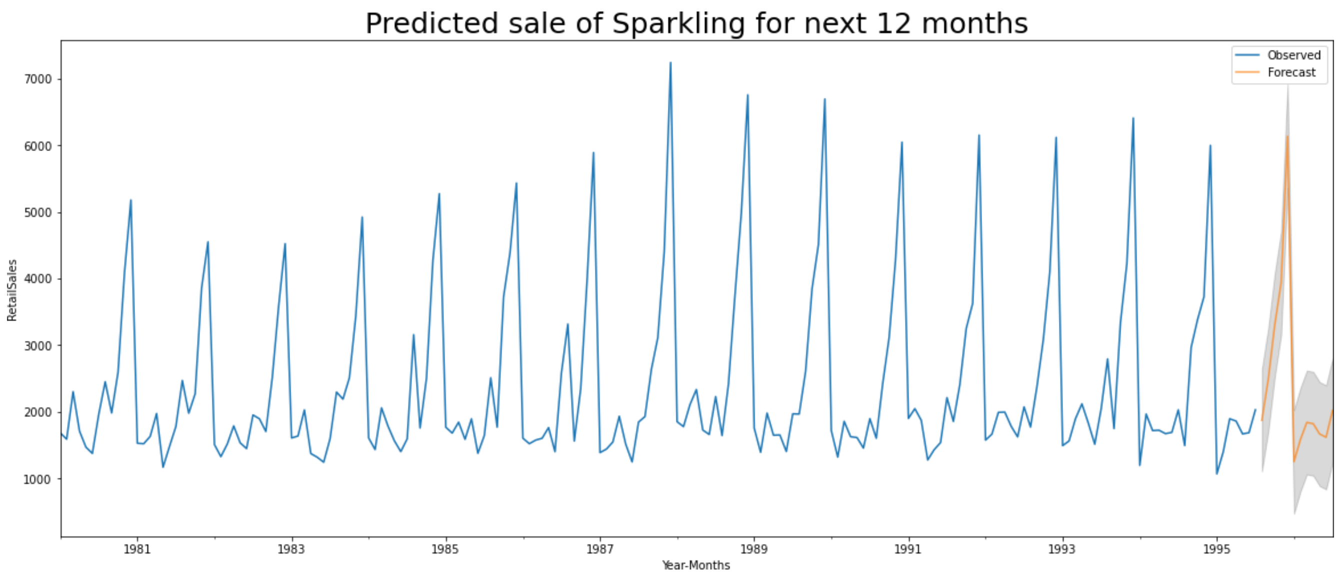
Since Sparkling sales data has component of seasonality. Therefore, SARIMA model with para meters (0,1,2)(0, 1, 2, 12) is proposed to used for forecast for next 12 months using full data. Details of model are as under.

```
                                    SARIMAX Results
==========================================================================================
Dep. Variable:                            Sparkling   No. Observations:              187
Model:      SARIMAX(0, 1, 2)x(0, 1, 2, 12)   Log Likelihood             -1087.003
Date:                          Sun, 28 Feb 2021   AIC                         2184.006
Time:                                  18:26:36   BIC                         2198.958
Sample:                              01-01-1980   HQIC                        2190.081
                                   - 07-01-1995
Covariance Type:                            opg
==================================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
----------------------------------------------------------------------------------
ma.L1         -0.9094      0.104     -8.713      0.000      -1.114      -0.705
ma.L2         -0.1316      0.087     -1.507      0.132      -0.303       0.040
ma.S.L12      -0.5456      0.065     -8.393      0.000      -0.673      -0.418
ma.S.L24      -0.0202      0.084     -0.241      0.810      -0.185       0.145
sigma2      1.419e+05   1.32e+04     10.755      0.000    1.16e+05    1.68e+05
==========================================================================================
Ljung-Box (L1) (Q):                   0.01   Jarque-Bera (JB):                  49.28
Prob(Q):                              0.91   Prob(JB):                           0.00
Heteroskedasticity (H):               0.79   Skew:                               0.74
Prob(H) (two-sided):                  0.42   Kurtosis:                           5.41
==========================================================================================
```

Diagnostics of the Final Model for Sparkling



RMSE of the Full Model is 550.06



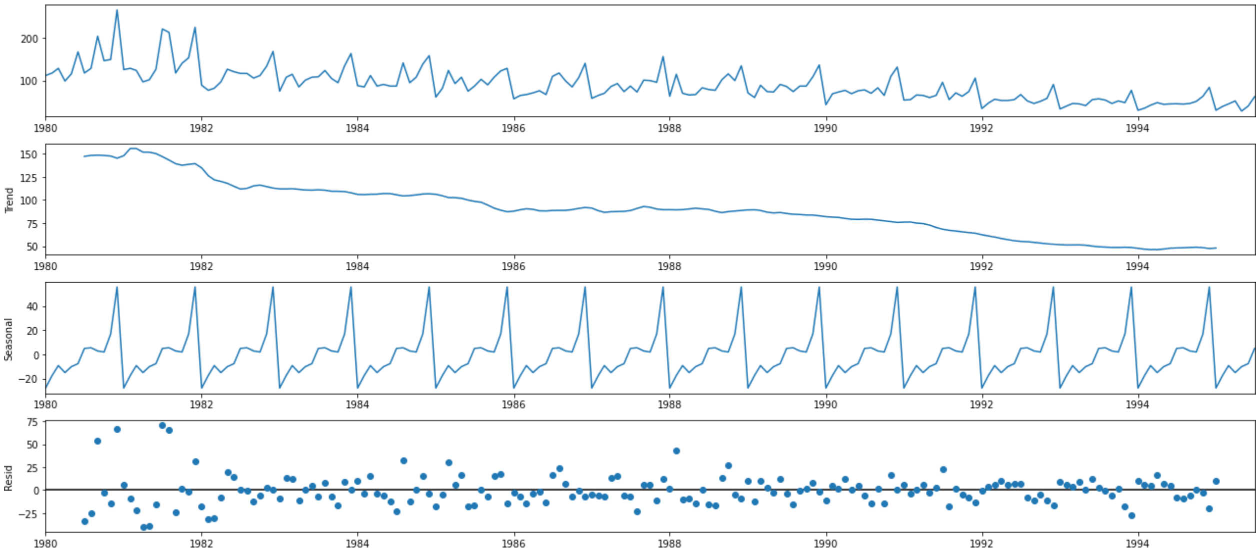
The predicted sales of Sparkling for next 12 months is as below

Sparkling	mean	mean_se	mean_ci_lower	mean_ci_upper
1995-08-01	1874.535574	390.423218	1109.320128	2639.751020
1995-09-01	2487.730812	395.512564	1712.540432	3262.921192
1995-10-01	3299.133285	395.812900	2523.354256	4074.912313
1995-11-01	3937.427323	396.113009	3161.060093	4713.794554
1995-12-01	6136.305467	396.412891	5359.350478	6913.260456
1996-01-01	1251.541103	396.712549	473.998795	2029.083412
1996-02-01	1583.924557	397.012002	805.795332	2362.053782

Sparkling	mean	mean_se	mean_ci_lower	mean_ci_upper
1996-03-01	1842.202832	397.311309	1063.486976	2620.918689
1996-04-01	1822.837693	397.610284	1043.535857	2602.139529
1996-05-01	1668.252244	397.909036	888.364864	2448.139623
1996-06-01	1619.268043	398.207564	838.795558	2399.740528
1996-07-01	2021.196789	398.505870	1240.139637	2802.253941

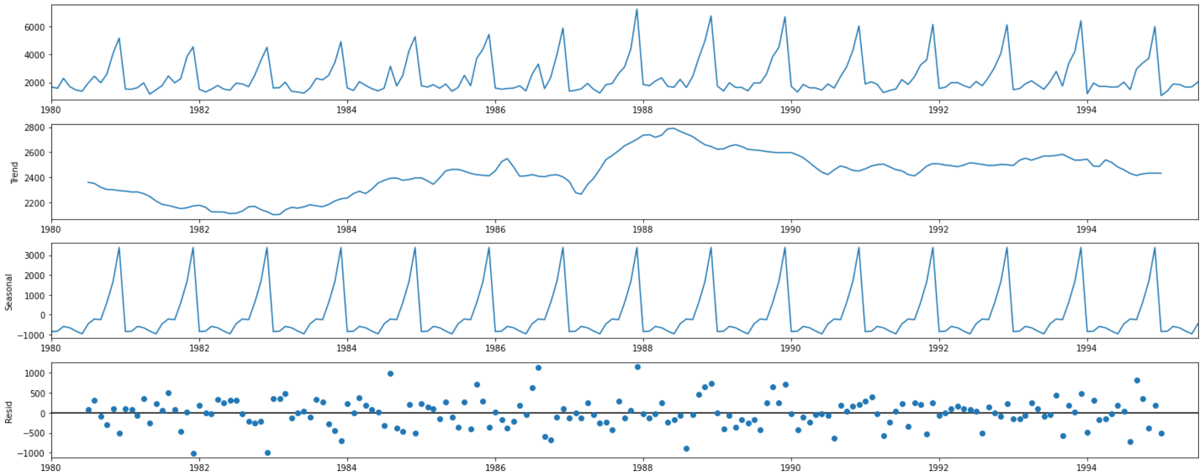
10. Comment on the model thus built and report your findings and suggest the measures that the company should be taking for future sales.

The decomposition of the time series trend of Rose wine is as below



Trend in sales of Rose is continuously decreasing over the period. Detailed study may be required to see whether decreasing trend is due to change in customer preference or due to substitution. Seasonality of sales is observed, and higher sales is maintained in the end of the year. Some promotion schemes and improvement / quality enhancers in the product can be examined so as to attract new young generation customers.

The decomposition of the sales of Sparkling is as below.



Sales in Sparking does not have uniform trend but increased in some years and decreased later. Business study may be done to find why sales are not increasing and what the contributing factors. Study can also include to see which wine product has substituted/ had higher sales in the years of low sales of Sparkling. With promotion and focussed effort with micro detailing it may be feasible to increase the sales. Sales of Sparkling wine higher in the later part of the year. This may be due to climatic condition of the geography under study.