Time Series Forecasting Report - Coded Project

Table of Contents

Table of Figures

Problem Statement - TSF Project

Context

As an analyst at ABC Estate Wines, we are presented with historical data encompassing the sales of different types of wines throughout the 20th century. These datasets originate from the same company but represent sales figures for distinct wine varieties. Our objective is to delve into the data, analyze trends, patterns, and factors influencing wine sales over the course of the century. By leveraging data analytics and forecasting techniques, we aim to gain actionable insights that can inform strategic decision-making and optimize sales strategies for the future.

Objective

The primary objective of this project is to analyze and forecast wine sales trends for the 20th century based on historical data provided by ABC Estate Wines. We aim to equip ABC Estate Wines with the necessary insights and foresight to enhance sales performance, capitalize on emerging market opportunities, and maintain a competitive edge in the wine industry.

Rose Dataset:

Define the problem and perform Exploratory Data Analysis

- Read the data as an appropriate time series data - Plot the data - Perform EDA - Perform Decomposition

Head of top 5 series in the Rose dataset

Out[275]:

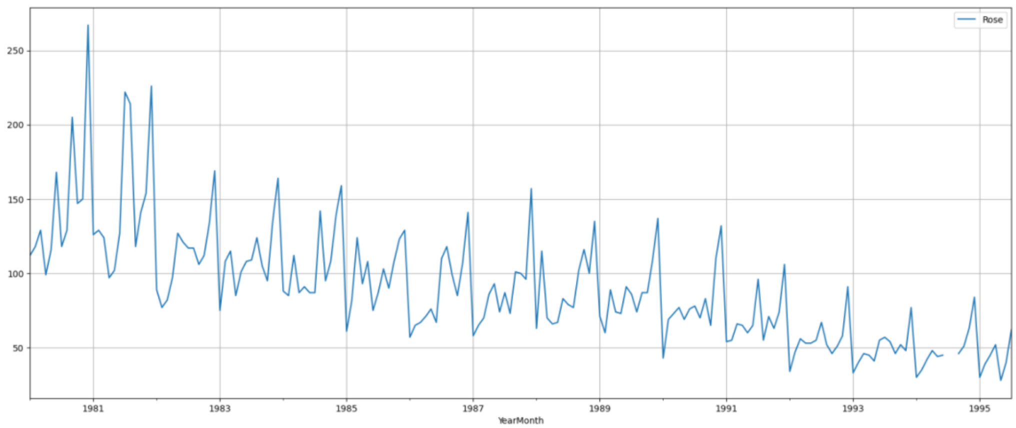
YearMonth	Rose
1980-01-01	112.0
1980-02-01	118.0
1980-03-01	129.0
1980-04-01	99.0
1980-05-01	116.0

Tail of the last 5 series in the Rose dataset

Out[276]:

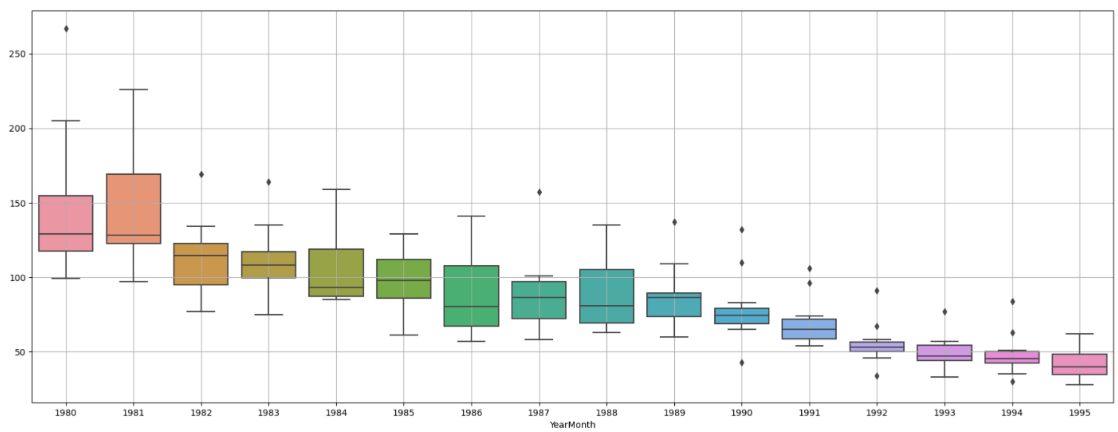
YearMonth	Rose
1995-03-01	45.0
1995-04-01	52.0
1995-05-01	28.0
1995-06-01	40.0
1995-07-01	62.0

Plotting the Rose dataset



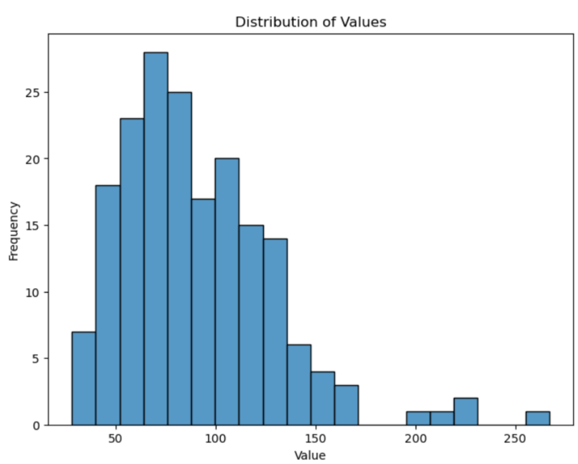
1. Plotting the Rose dataset

Visualize the Boxplot in Rose dataset on yearly basis



2. the Boxplot in Rose dataset on yearly basis

. Visualize data distribution



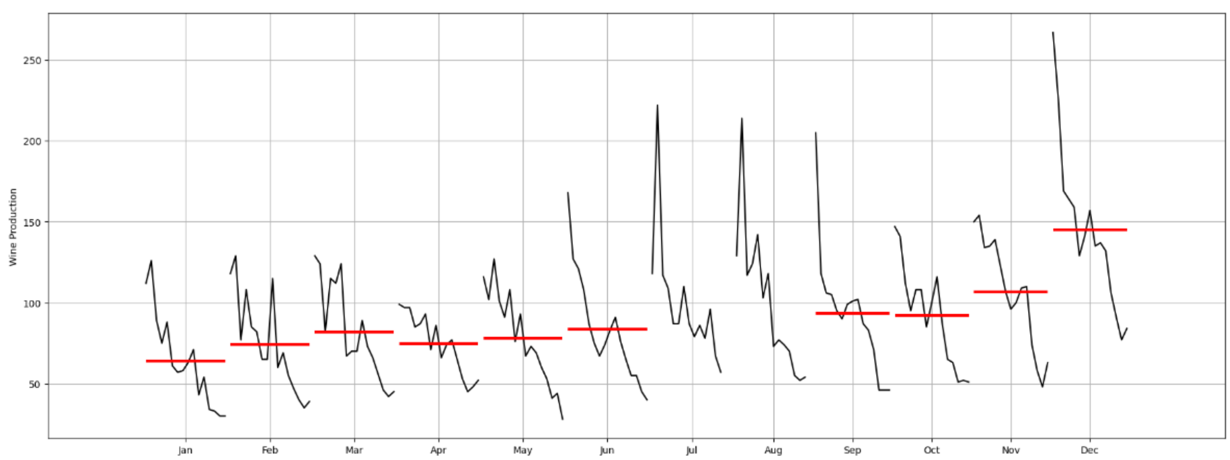
3. Visualize data distribution

```
             Rose
count  185.000000
mean    90.394595
std     39.175344
min     28.000000
25%     63.000000
50%     86.000000
75%    112.000000
max    267.000000
<class 'pandas.core.frame.DataFrame'>
DatetimeIndex: 187 entries, 1980-01-01 to 1995-07-01
Data columns (total 1 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   Rose    185 non-null    float64
dtypes: float64(1)
memory usage: 2.9 KB
None
```

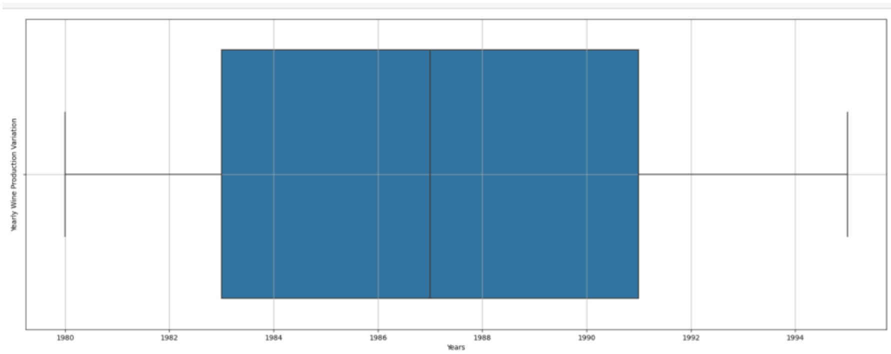
4. Describe the rose dataset

```
Rose     2
dtype: int64
```

There are two missing values in the dataset.



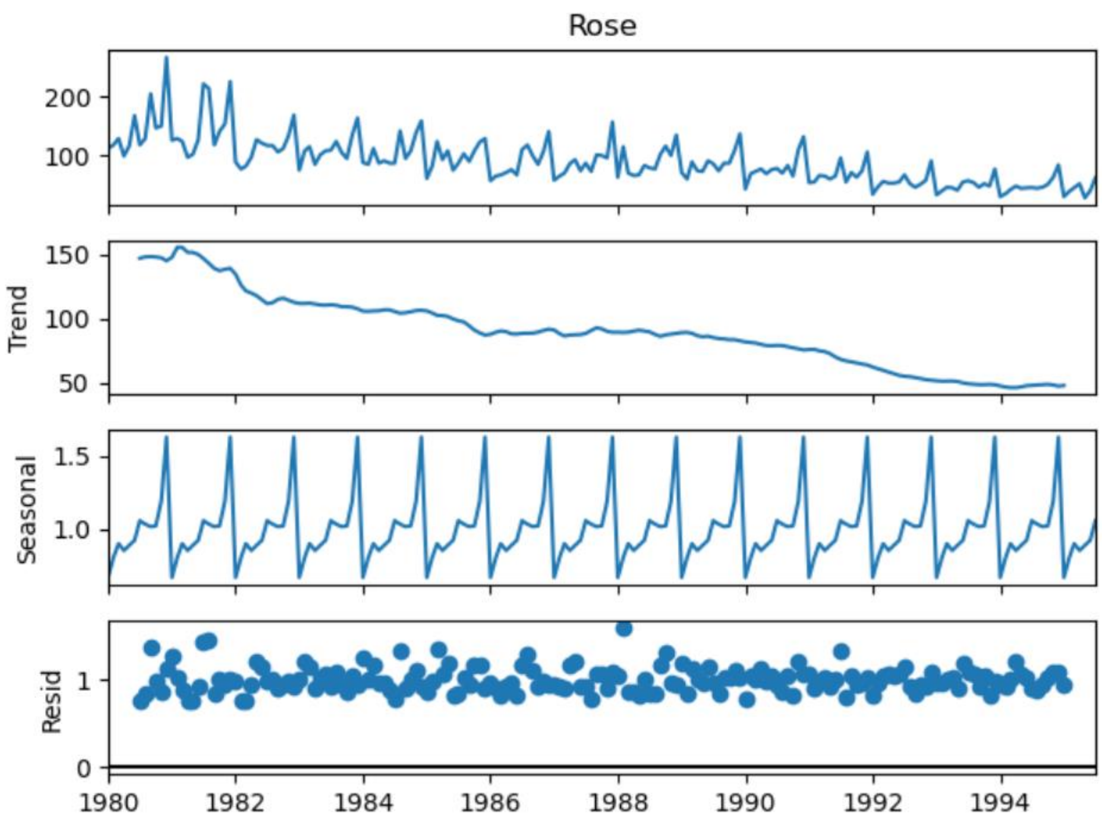
5. Plot a monthplot of the give Time Series



A decreasing Trend could be observed with a multiplicative seasonality present. The Null values could be observed as a break in the plot for the observed timestamps.

Data Pre-processing

- Missing value treatment - Visualize the processed data - Train-test split



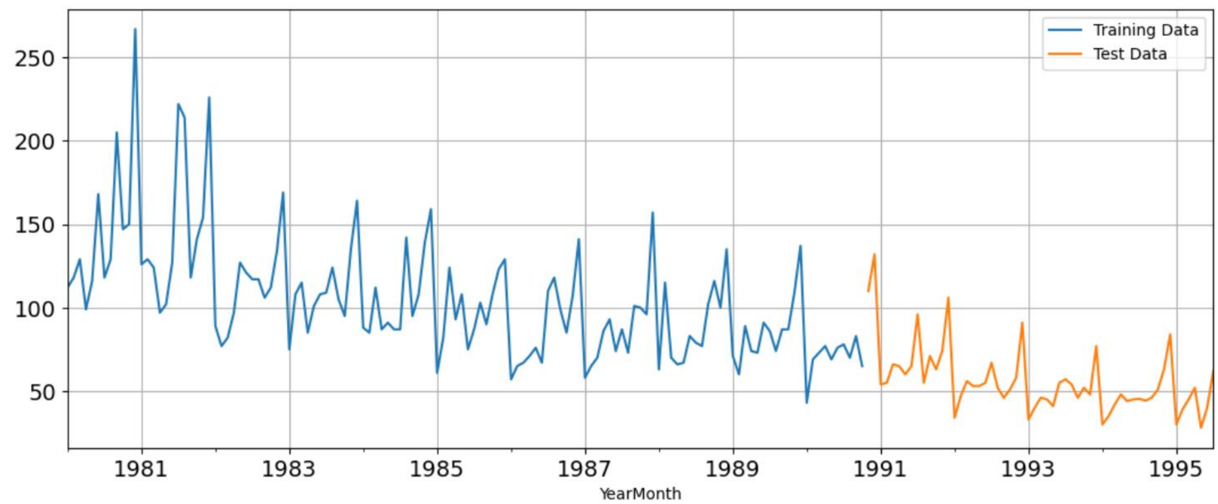
6. Perform decomposition

Train- test split dataset

```
(130, 1)
(57, 1)
```

Train dataset contains 130 rows and test dataset contains 57 rows.

```
Last few rows of Training Data
             Rose
YearMonth
1990-06-01  76.0
1990-07-01  78.0
1990-08-01  70.0
1990-09-01  83.0
1990-10-01  65.0

First few rows of Test Data
              Rose
YearMonth
1990-11-01  110.0
1990-12-01  132.0
1991-01-01   54.0
1991-02-01   55.0
1991-03-01   66.0

Last few rows of Test Data
             Rose
YearMonth
1995-03-01  45.0
1995-04-01  52.0
1995-05-01  28.0
1995-06-01  40.0
1995-07-01  62.0
```



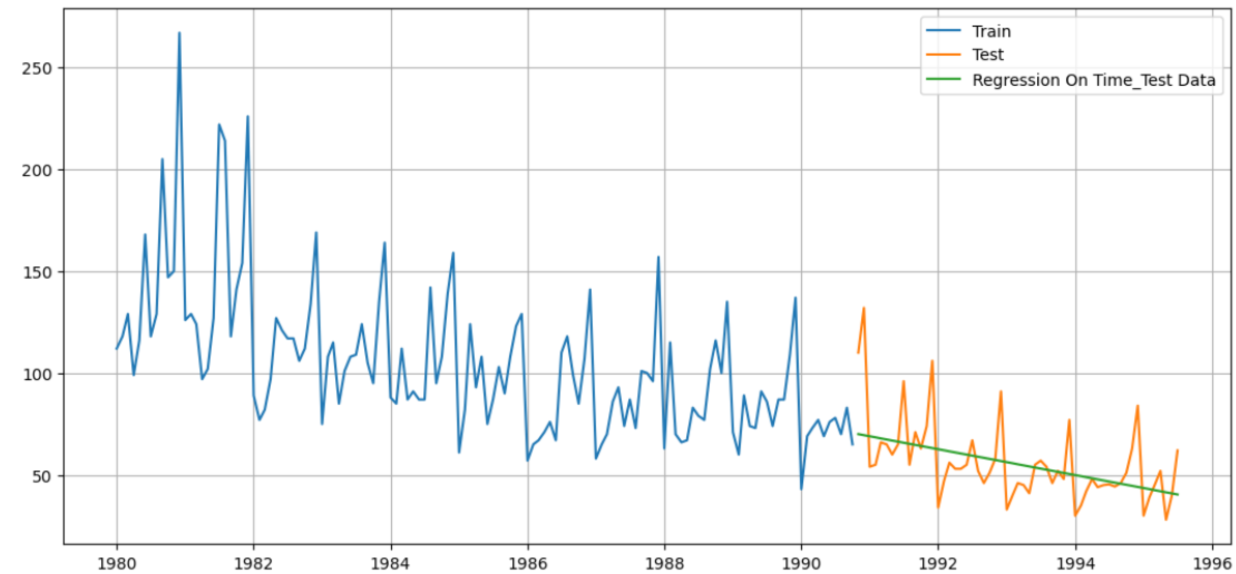
7. Train - Test plot

Model Building - Original Data

- Build forecasting models - Linear regression - Simple Average - Moving Average - Exponential Models (Single, Double, Triple) - Check the performance of the models built

Model 1: Linear Regression

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

Last few rows of Training Data
             Rose  time
YearMonth
1990-06-01  76.0   126
1990-07-01  78.0   127
1990-08-01  70.0   128
1990-09-01  83.0   129
1990-10-01  65.0   130

First few rows of Test Data
              Rose  time
YearMonth
1990-11-01  110.0   131
1990-12-01  132.0   132
1991-01-01   54.0   133
1991-02-01   55.0   134
1991-03-01   66.0   135

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



8. Linear Regression on test data

For RegressionOnTime forecast on the Test Data, RMSE is 17.36

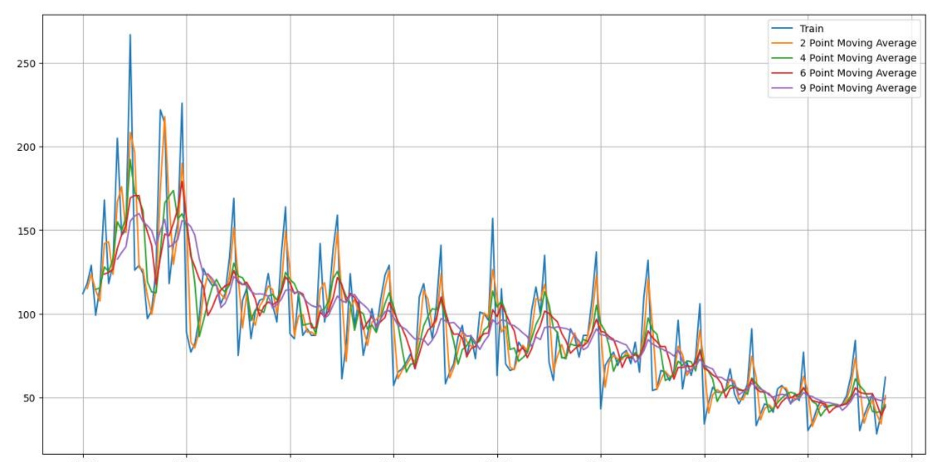
Out[297]:

	Test RMSE
RegressionOnTime	17.357497

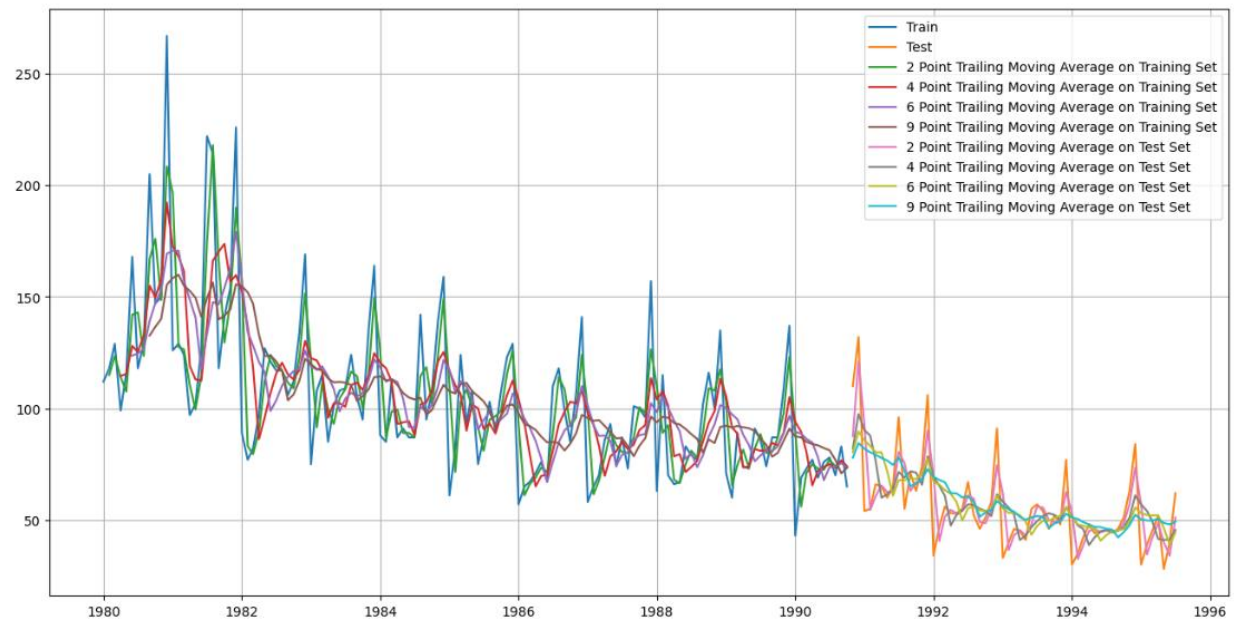
Method 2: Moving Average(MA)

Out[299]:

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



9. Moving Average



RMSE Calculation as below,

For 2 point Moving Average Model forecast on the Training Data, RMSE is 11.802

For 4 point Moving Average Model forecast on the Training Data, RMSE is 15.374

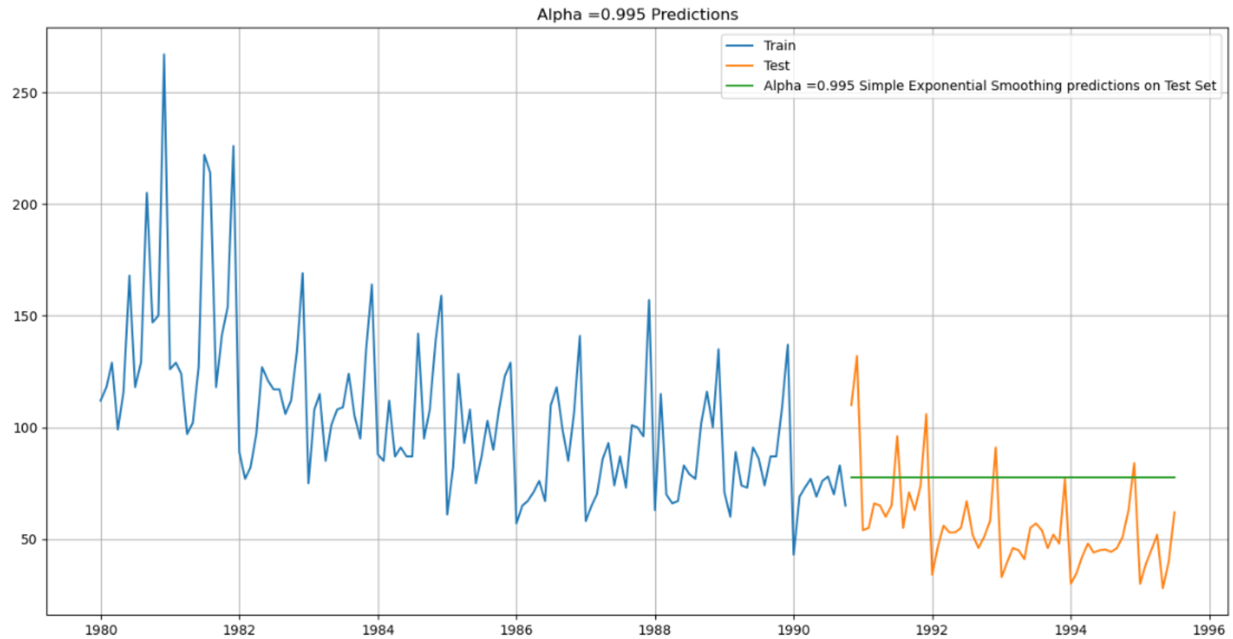
For 6 point Moving Average Model forecast on the Training Data, RMSE is 15.868

For 9 point Moving Average Model forecast on the Training Data, RMSE is 16.347

Out[304]:

	Test RMSE
RegressionOnTime	17.357497
2pointTrailingMovingAverage	11.801775
4pointTrailingMovingAverage	15.373563
6pointTrailingMovingAverage	15.868241
9pointTrailingMovingAverage	16.346517

Method 3: Simple Exponential Smoothing



10. Alpha =0.995 Simple Exponential Smoothing predictions on Test Set

Model Evaluation for α = 0.995 : Simple Exponential Smoothing

or Alpha =0.995 Simple Exponential Smoothing Model forecast on the Test Data, RMSE is 29.250

Out[313]:

	Test RMSE
RegressionOnTime	17.357497
2pointTrailingMovingAverage	11.801775
4pointTrailingMovingAverage	15.373563
6pointTrailingMovingAverage	15.868241
9pointTrailingMovingAverage	16.346517
Alpha=0.995,SimpleExponentialSmoothing	29.250243

Method 4: Double Exponential Smoothing (Holt's Model)

Train dataset

Out[318]:

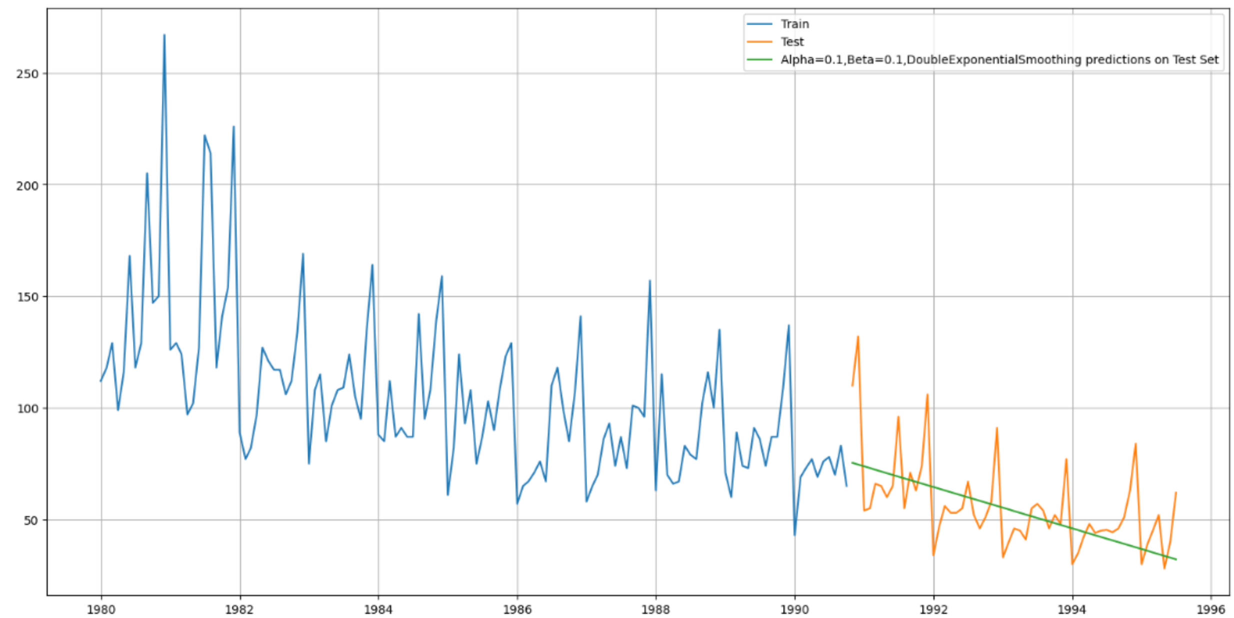
	Alpha Values	Beta Values	Train RMSE	Test RMSE
0	0.1	0.1	34.25	17.68
1	0.1	0.2	33.16	29.59
2	0.1	0.3	32.85	30.26
3	0.1	0.4	32.95	39.64
4	0.1	0.5	33.36	49.70
...	...	...	...	...
95	1.0	0.6	52.02	264.82
96	1.0	0.7	54.70	331.59
97	1.0	0.8	57.58	406.47
98	1.0	0.9	60.69	491.21
99	1.0	1.0	64.09	587.86

100 rows × 4 columns

Test dataset

Out[319]:

	Alpha Values	Beta Values	Train RMSE	Test RMSE
0	0.1	0.1	34.25	17.68
16	0.2	0.7	40.35	17.74
23	0.3	0.4	37.29	18.34
36	0.4	0.7	40.74	18.99
33	0.4	0.4	37.99	19.15



11. Alpha=0.1,Beta=0.1,DoubleExponentialSmoothing predictions on Test Set

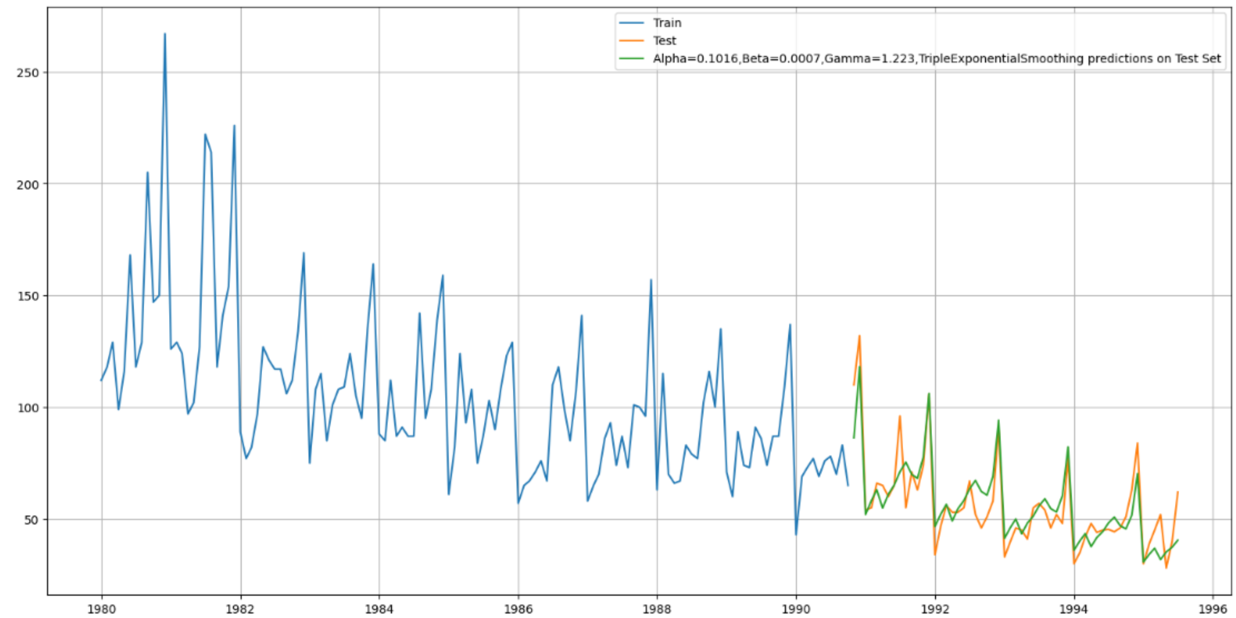
Out[321]:

	Test RMSE
RegressionOnTime	17.357497
2pointTrailingMovingAverage	11.801775
4pointTrailingMovingAverage	15.373563
6pointTrailingMovingAverage	15.868241
9pointTrailingMovingAverage	16.346517
Alpha=0.995,SimpleExponentialSmoothing	29.250243
Alpha=0.3,Beta=0.3,DoubleExponentialSmoothing	17.680000

Method 5: Triple Exponential Smoothing (Holt - Winter's Model)

Out[326]:

	Rose	auto_predict
YearMonth		
1990-11-01	110.0	86.341086
1990-12-01	132.0	118.068900
1991-01-01	54.0	51.939367
1991-02-01	55.0	58.209268
1991-03-01	66.0	63.101160



12. Alpha=0.1016,Beta=0.0007,Gamma=1.223,TripleExponentialSmoothing predictions on Test Set

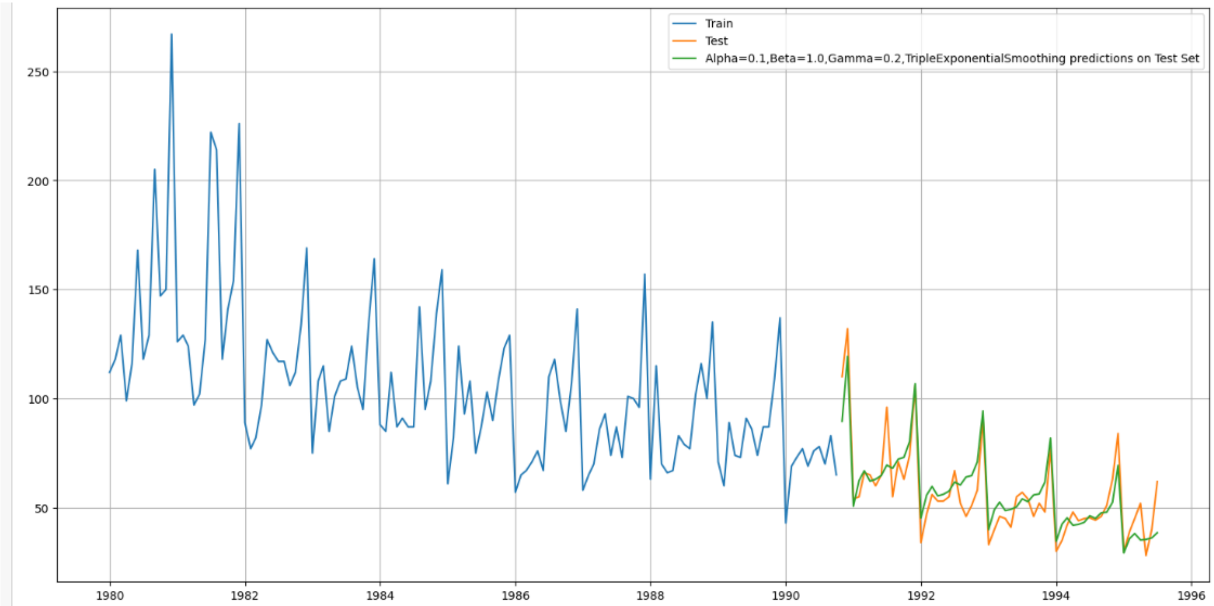
For Alpha=0.1016,Beta=0.0007,Gamma=1.223, Triple Exponential Smoothing Model forecast on the Test Data, RMSE is 9.338

Out[329]:

	Test RMSE
RegressionOnTime	17.357497
2pointTrailingMovingAverage	11.801775
4pointTrailingMovingAverage	15.373563
6pointTrailingMovingAverage	15.868241
9pointTrailingMovingAverage	16.346517
Alpha=0.995,SimpleExponentialSmoothing	29.250243
Alpha=0.3,Beta=0.3,DoubleExponentialSmoothing	17.680000
Alpha=0.1016,Beta=0.0007,Gamma=1.223,TripleExponentialSmoothing	9.337808

Out[332]:

	Alpha Values	Beta Values	Gamma Values	Train RMSE	Test RMSE
91	0.1	1.0	0.2	23.14	9.13
4	0.1	0.1	0.5	22.42	9.20
5	0.1	0.1	0.6	23.40	9.27
3	0.1	0.1	0.4	21.55	9.28
106	0.2	0.1	0.7	25.62	9.32



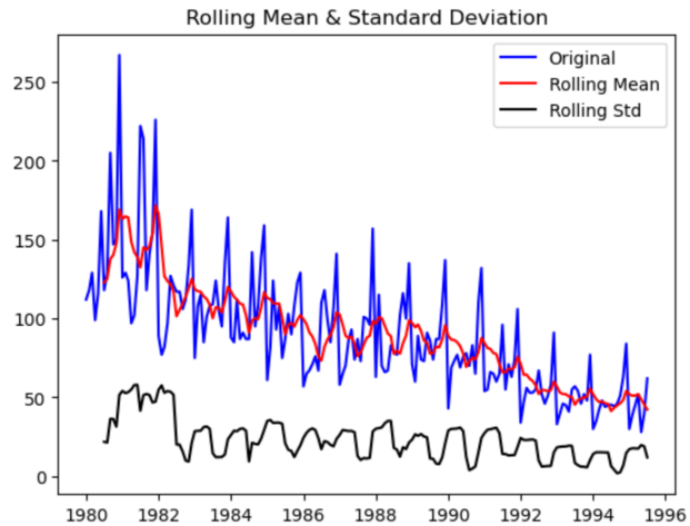
13. Alpha=0.1,Beta=1.0,Gamma=0.2,TripleExponentialSmoothing predictions on Test Set

Out[334]:

	Test RMSE
RegressionOnTime	17.357497
2pointTrailingMovingAverage	11.801775
4pointTrailingMovingAverage	15.373563
6pointTrailingMovingAverage	15.868241
9pointTrailingMovingAverage	16.346517
Alpha=0.995,SimpleExponentialSmoothing	29.250243
Alpha=0.3,Beta=0.3,DoubleExponentialSmoothing	17.680000
Alpha=0.1016,Beta=0.0007,Gamma=1.223,TripleExponentialSmoothing	9.337808
Alpha=0.1,Beta=1.0,Gamma=0.2,TripleExponentialSmoothing	9.130000

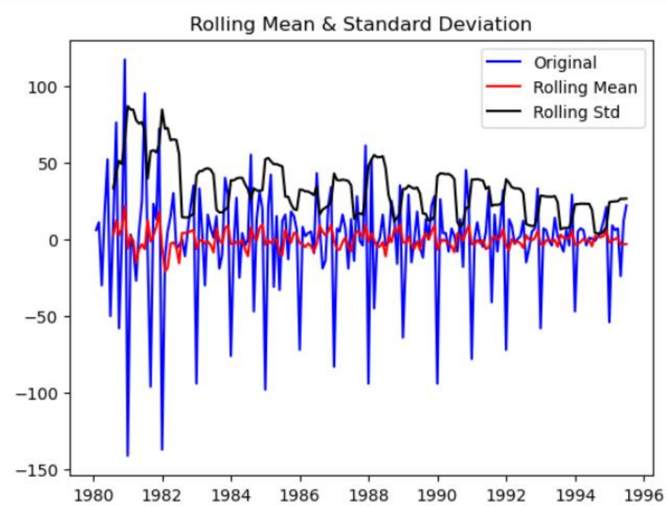
Check for Stationarity

- Check for stationarity - Make the data stationary (if needed)




```
Results of Dickey-Fuller Test:
Test Statistic                  -1.872615
p-value                          0.345051
#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
```

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

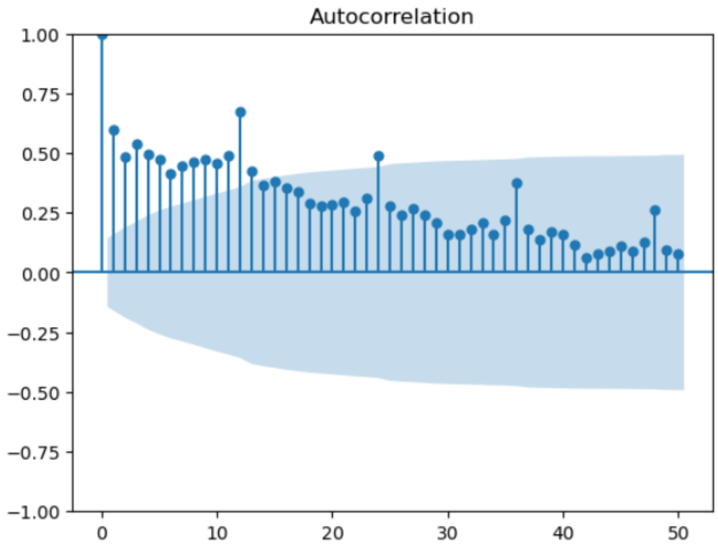


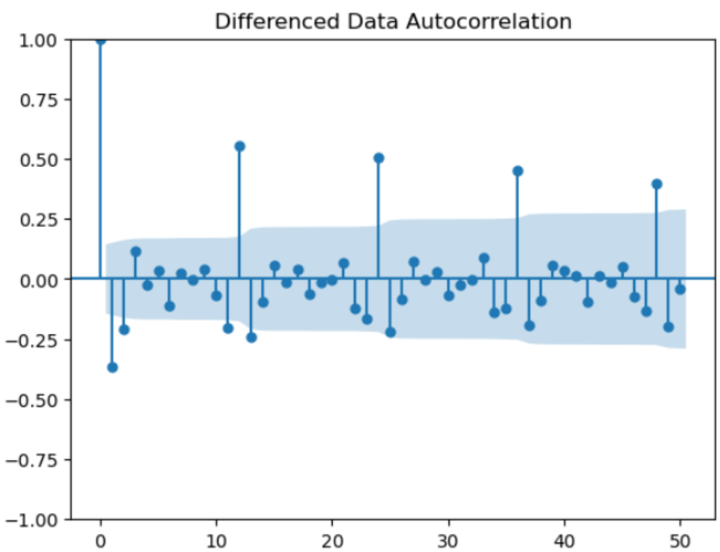

```
Results of Dickey-Fuller Test:
Test Statistic                -8.044081e+00
p-value                        1.814191e-12
#Lags Used                     1.200000e+01
Number of Observations Used    1.730000e+02
Critical Value (1%)           -3.468726e+00
Critical Value (5%)           -2.878396e+00
Critical Value (10%)          -2.575756e+00
dtype: float64
```

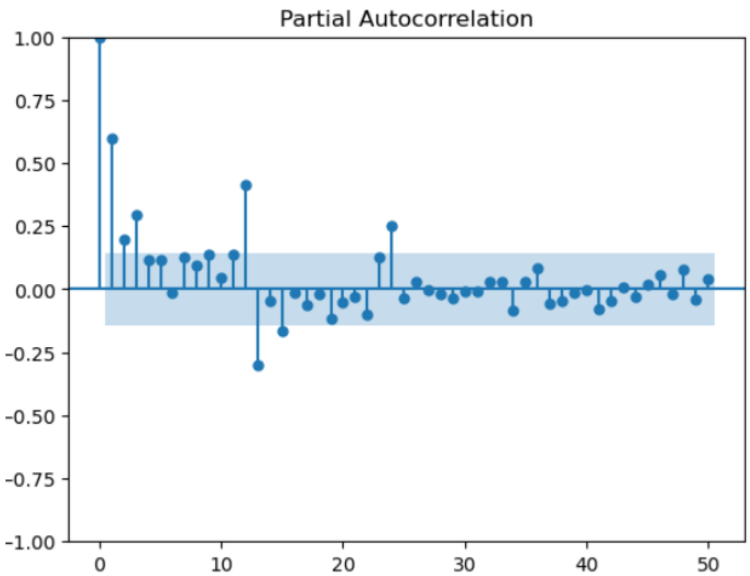
We see that at α = 0.05 the Time Series is indeed stationary.

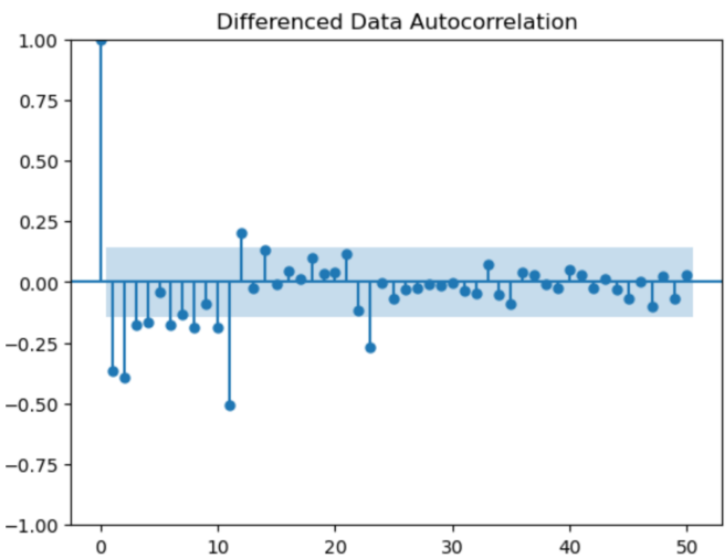
Model Building - Stationary Data

- Generate ACF & PACF Plot and find the AR, MA values. - Build different ARIMA models - Auto ARIMA - Manual ARIMA - Build different SARIMA models - Auto SARIMA - Manual SARIMA - Check the performance of the models built









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

```
Some parameter combinations for the Model...
Model: (0, 0, 1)
Model: (0, 0, 2)
Model: (1, 0, 0)
Model: (1, 0, 1)
Model: (1, 0, 2)
Model: (2, 0, 0)
Model: (2, 0, 1)
Model: (2, 0, 2)
```

Out[344]:

	param	AIC
5	(1, 0, 2)	1272.008961
8	(2, 0, 2)	1272.230529
7	(2, 0, 1)	1272.784560
4	(1, 0, 1)	1273.969672
3	(1, 0, 0)	1282.884243
6	(2, 0, 0)	1283.462128
1	(0, 0, 1)	1287.093068
2	(0, 0, 2)	1288.116816
0	(0, 0, 0)	1306.288690

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                   Rose   No. Observations:                  130
Model:                 ARIMA(1, 0, 2)   Log Likelihood                -631.004
Date:                Fri, 13 Sep 2024   AIC                           1272.009
Time:                        20:38:16   BIC                           1286.347
Sample:                    01-01-1980   HQIC                          1277.835
                         - 10-01-1990
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const        104.6906     31.457      3.328      0.001      43.036     166.345
ar.L1          0.9886      0.024     41.384      0.000       0.942       1.035
ma.L1         -0.6983      0.089     -7.839      0.000      -0.873      -0.524
ma.L2         -0.1894      0.091     -2.090      0.037      -0.367      -0.012
sigma2       955.1353     99.717      9.578      0.000     759.694    1150.576
===================================================================================
Ljung-Box (L1) (Q):                   0.05   Jarque-Bera (JB):                62.77
Prob(Q):                              0.82   Prob(JB):                         0.00
Heteroskedasticity (H):               0.32   Skew:                             1.01
Prob(H) (two-sided):                  0.00   Kurtosis:                         5.74
===================================================================================

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

For Auto-ARIMA Model forecast accuracy_score on the Test Data, RMSE is 38.675

Setting the seasonality as 12 to estimate parametrs using auto SARIMA model.

```
Examples of some parameter combinations for Model...
Model: (0, 1, 1)(0, 1, 1, 12)
Model: (0, 1, 2)(0, 1, 2, 12)
Model: (1, 1, 0)(1, 1, 0, 12)
Model: (1, 1, 1)(1, 1, 1, 12)
Model: (1, 1, 2)(1, 1, 2, 12)
Model: (2, 1, 0)(2, 1, 0, 12)
Model: (2, 1, 1)(2, 1, 1, 12)
Model: (2, 1, 2)(2, 1, 2, 12)
```

SARIMA(0, 1, 0)x(0, 1, 0, 12)7 - AIC:1152.8944112896427
SARIMA(0, 1, 0)x(0, 1, 1, 12)7 - AIC:946.5255706383725
SARIMA(0, 1, 0)x(0, 1, 2, 12)7 - AIC:827.4161882799052
SARIMA(0, 1, 0)x(1, 1, 0, 12)7 - AIC:963.3074810361367
SARIMA(0, 1, 0)x(1, 1, 1, 12)7 - AIC:955.90241535815
SARIMA(0, 1, 0)x(1, 1, 2, 12)7 - AIC:829.2186210967307
SARIMA(0, 1, 0)x(2, 1, 0, 12)7 - AIC:839.1329917213063
SARIMA(0, 1, 0)x(2, 1, 1, 12)7 - AIC:834.4921533619314
SARIMA(0, 1, 0)x(2, 1, 2, 12)7 - AIC:824.3830509688956
SARIMA(0, 1, 1)x(0, 1, 0, 12)7 - AIC:1096.7564633976488
SARIMA(0, 1, 1)x(0, 1, 1, 12)7 - AIC:899.1279003224759
SARIMA(0, 1, 1)x(0, 1, 2, 12)7 - AIC:775.3675849137794
SARIMA(0, 1, 1)x(1, 1, 0, 12)7 - AIC:928.1728057664981
SARIMA(0, 1, 1)x(1, 1, 1, 12)7 - AIC:913.2749250995581
SARIMA(0, 1, 1)x(1, 1, 2, 12)7 - AIC:777.1680791501379
SARIMA(0, 1, 1)x(2, 1, 0, 12)7 - AIC:794.878534195182
SARIMA(0, 1, 1)x(2, 1, 1, 12)7 - AIC:788.3653415950496
SARIMA(0, 1, 1)x(2, 1, 2, 12)7 - AIC:767.3752584863522
SARIMA(0, 1, 2)x(0, 1, 0, 12)7 - AIC:1081.5930721187483
SARIMA(0, 1, 2)x(0, 1, 1, 12)7 - AIC:888.1453938346057
SARIMA(0, 1, 2)x(0, 1, 2, 12)7 - AIC:768.4215063916055
SARIMA(0, 1, 2)x(1, 1, 0, 12)7 - AIC:925.1239001312563
SARIMA(0, 1, 2)x(1, 1, 1, 12)7 - AIC:898.8756745213946
SARIMA(0, 1, 2)x(1, 1, 2, 12)7 - AIC:770.2181164945283
SARIMA(0, 1, 2)x(2, 1, 0, 12)7 - AIC:795.8587520859533
SARIMA(0, 1, 2)x(2, 1, 1, 12)7 - AIC:789.6477608525199
SARIMA(0, 1, 2)x(2, 1, 2, 12)7 - AIC:759.854456337735
SARIMA(1, 1, 0)x(0, 1, 0, 12)7 - AIC:1146.5245639869129
SARIMA(1, 1, 0)x(0, 1, 1, 12)7 - AIC:936.1793735397487
SARIMA(1, 1, 0)x(0, 1, 2, 12)7 - AIC:811.6876559416335
SARIMA(1, 1, 0)x(1, 1, 0, 12)7 - AIC:944.9341691555421
SARIMA(1, 1, 0)x(1, 1, 1, 12)7 - AIC:946.6249192435952
SARIMA(1, 1, 0)x(1, 1, 2, 12)7 - AIC:813.0825861131768
SARIMA(1, 1, 0)x(2, 1, 0, 12)7 - AIC:809.341284916012

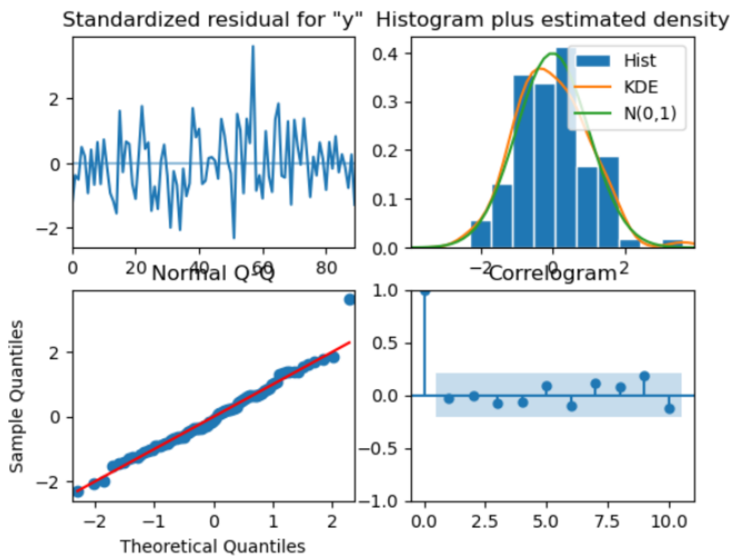
SARIMA(1, 1, 0)x(2, 1, 1, 12)7 - AIC:806.8254045230403
SARIMA(1, 1, 0)x(2, 1, 2, 12)7 - AIC:808.4299128187591
SARIMA(1, 1, 1)x(0, 1, 0, 12)7 - AIC:1093.5988179239696
SARIMA(1, 1, 1)x(0, 1, 1, 12)7 - AIC:897.7667931010029
SARIMA(1, 1, 1)x(0, 1, 2, 12)7 - AIC:776.6744753192086
SARIMA(1, 1, 1)x(1, 1, 0, 12)7 - AIC:916.9084995137714
SARIMA(1, 1, 1)x(1, 1, 1, 12)7 - AIC:910.8884983743482
SARIMA(1, 1, 1)x(1, 1, 2, 12)7 - AIC:778.512074478587
SARIMA(1, 1, 1)x(2, 1, 0, 12)7 - AIC:784.2808874065543
SARIMA(1, 1, 1)x(2, 1, 1, 12)7 - AIC:776.709647706393
SARIMA(1, 1, 1)x(2, 1, 2, 12)7 - AIC:768.9979136123605
SARIMA(1, 1, 2)x(0, 1, 0, 12)7 - AIC:1079.4524066139743
SARIMA(1, 1, 2)x(0, 1, 1, 12)7 - AIC:890.1378220931629
SARIMA(1, 1, 2)x(0, 1, 2, 12)7 - AIC:770.3906403523735
SARIMA(1, 1, 2)x(1, 1, 0, 12)7 - AIC:918.4763772990403
SARIMA(1, 1, 2)x(1, 1, 1, 12)7 - AIC:900.8377043226517
SARIMA(1, 1, 2)x(1, 1, 2, 12)7 - AIC:772.1781817850084
SARIMA(1, 1, 2)x(2, 1, 0, 12)7 - AIC:786.2709969925108
SARIMA(1, 1, 2)x(2, 1, 1, 12)7 - AIC:778.5728925206383
SARIMA(1, 1, 2)x(2, 1, 2, 12)7 - AIC:761.8340847942384
SARIMA(2, 1, 0)x(0, 1, 0, 12)7 - AIC:1115.6453843716572
SARIMA(2, 1, 0)x(0, 1, 1, 12)7 - AIC:931.975620270612
SARIMA(2, 1, 0)x(0, 1, 2, 12)7 - AIC:810.2393911009702
SARIMA(2, 1, 0)x(1, 1, 0, 12)7 - AIC:932.6053635140116
SARIMA(2, 1, 0)x(1, 1, 1, 12)7 - AIC:934.5783634759562
SARIMA(2, 1, 0)x(1, 1, 2, 12)7 - AIC:811.7123326713876
SARIMA(2, 1, 0)x(2, 1, 0, 12)7 - AIC:797.2689249311138
SARIMA(2, 1, 0)x(2, 1, 1, 12)7 - AIC:791.6332001791104
SARIMA(2, 1, 0)x(2, 1, 2, 12)7 - AIC:793.3001859578382
SARIMA(2, 1, 1)x(0, 1, 0, 12)7 - AIC:1090.142414920315
SARIMA(2, 1, 1)x(0, 1, 1, 12)7 - AIC:898.7844234538268
SARIMA(2, 1, 1)x(0, 1, 2, 12)7 - AIC:778.6275923519522
SARIMA(2, 1, 1)x(1, 1, 0, 12)7 - AIC:909.896906649278
SARIMA(2, 1, 1)x(1, 1, 1, 12)7 - AIC:911.8731618382475
SARIMA(2, 1, 1)x(1, 1, 2, 12)7 - AIC:780.4623774343154
SARIMA(2, 1, 1)x(2, 1, 0, 12)7 - AIC:777.2798270113633
SARIMA(2, 1, 1)x(2, 1, 1, 12)7 - AIC:769.0961329445562
SARIMA(2, 1, 1)x(2, 1, 2, 12)7 - AIC:770.8360215597357
SARIMA(2, 1, 2)x(0, 1, 0, 12)7 - AIC:1082.4796249535534
SARIMA(2, 1, 2)x(0, 1, 1, 12)7 - AIC:888.4121912955013
SARIMA(2, 1, 2)x(0, 1, 2, 12)7 - AIC:768.8907825473697
SARIMA(2, 1, 2)x(1, 1, 0, 12)7 - AIC:908.7507874392862
SARIMA(2, 1, 2)x(1, 1, 1, 12)7 - AIC:898.9034728197205
SARIMA(2, 1, 2)x(1, 1, 2, 12)7 - AIC:772.7278700485623

SARIMA(2, 1, 2)x(2, 1, 0, 12)7 - AIC:779.1656800773759
SARIMA(2, 1, 2)x(2, 1, 1, 12)7 - AIC:768.7498409183947
SARIMA(2, 1, 2)x(2, 1, 2, 12)7 - AIC:763.8208489210297

Out[354]:

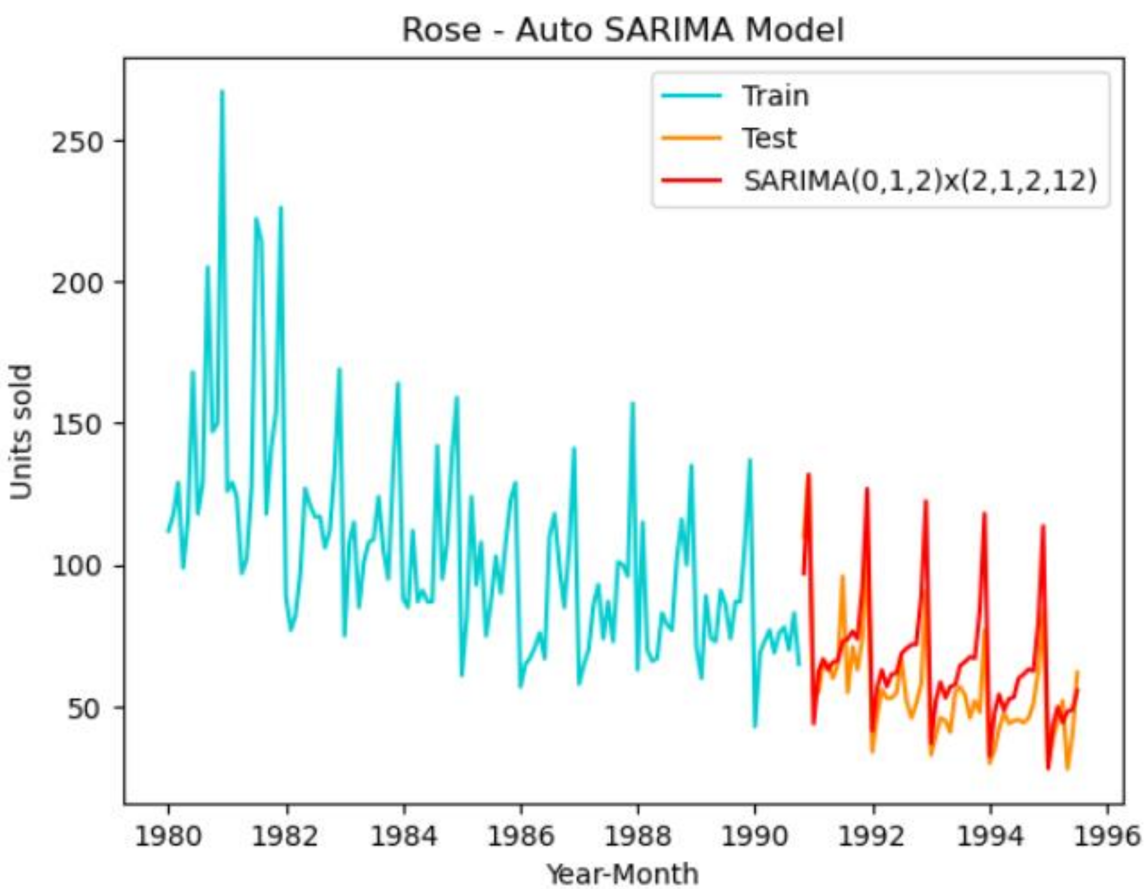
	param	seasonal	AIC
26	(0, 1, 2)	(2, 1, 2, 12)	759.854456
53	(1, 1, 2)	(2, 1, 2, 12)	761.834085
80	(2, 1, 2)	(2, 1, 2, 12)	763.820849
17	(0, 1, 1)	(2, 1, 2, 12)	767.375258
20	(0, 1, 2)	(0, 1, 2, 12)	768.421506

```
                                      SARIMAX Results                                       
==========================================================================================
Dep. Variable:                                   y   No. Observations:                  130
Model:             SARIMAX(0, 1, 2)x(2, 1, 2, 12)   Log Likelihood                -372.927
Date:                             Fri, 13 Sep 2024   AIC                            759.854
Time:                                     20:41:38   BIC                            777.353
Sample:                                          0   HQIC                           766.911
                                             - 130                                         
Covariance Type:                               opg                                         
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ma.L1         -0.9328      0.191     -4.878      0.000      -1.308      -0.558
ma.L2         -0.0923      0.126     -0.733      0.463      -0.339       0.154
ar.S.L12       0.0367      0.187      0.197      0.844      -0.329       0.402
ar.S.L24      -0.0421      0.030     -1.427      0.154      -0.100       0.016
ma.S.L12      -0.7249      0.294     -2.469      0.014      -1.300      -0.149
ma.S.L24      -0.0677      0.204     -0.331      0.740      -0.468       0.333
sigma2       194.3968     46.227      4.205      0.000     103.793     285.001
===================================================================================
Ljung-Box (L1) (Q):                   0.05   Jarque-Bera (JB):                 4.90
Prob(Q):                              0.82   Prob(JB):                         0.09
Heteroskedasticity (H):               0.93   Skew:                             0.43
Prob(H) (two-sided):                  0.84   Kurtosis:                         3.75
===================================================================================

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



Out[362]:

	Test RMSE
SARIMA(0, 1, 2)*(2, 1, 2, 12)	15.433926



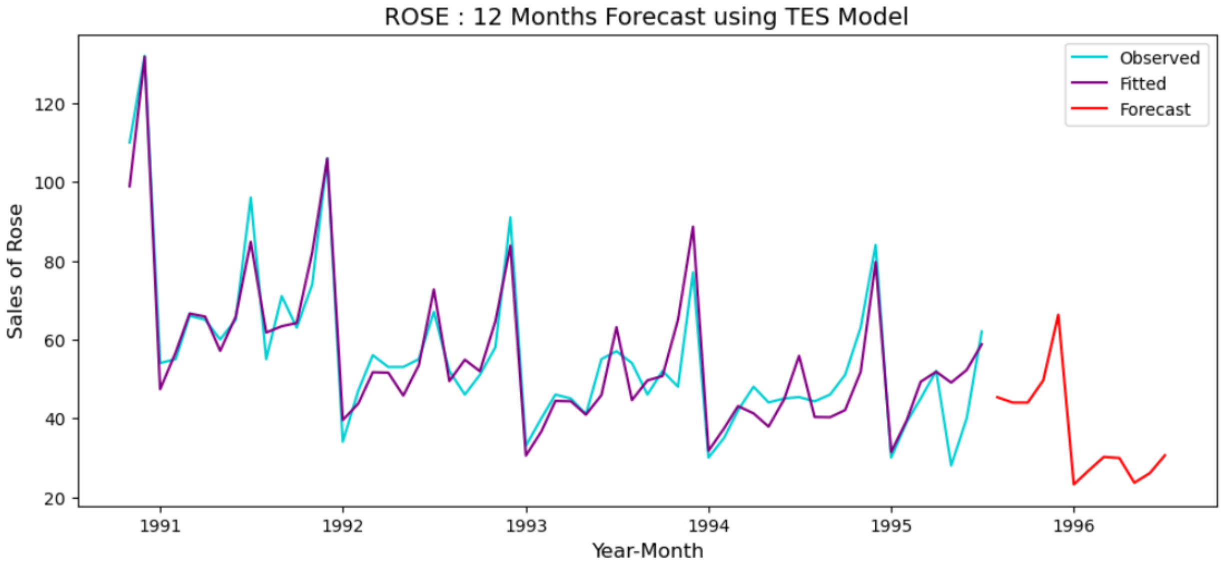
14. Rose - Auto SARIMA Model

Actionable Insights & Recommendations

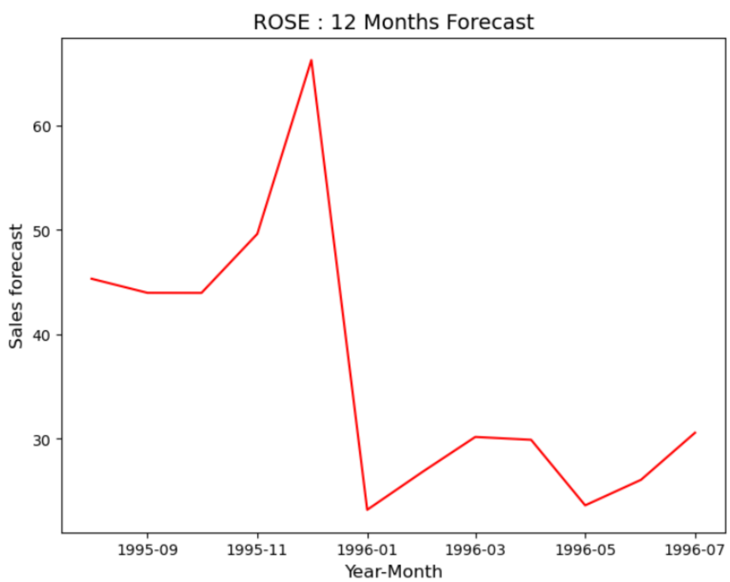
- Conclude with the key takeaways (actionable insights and recommendations) for the business

Based on the results, Triple exponential smoorthing has lower RMSE of 9.130000 (Alpha=0.1,Beta=1.0,Gamma=0.2,TripleExponentialSmoothing).

For Triple Exponential Smoothing Model forecast on the Entire Data, RMSE is 6.725



15. ROSE : 12 Months Forecast using TES Model



16.ROSE : 12 Months Forecast

```
Out[393]: count    12.000000
          mean     36.636415
          std      13.181135
          min      23.202954
          25%      26.670772
          50%      30.381838
          75%      44.319365
          max      66.279325
          dtype: float64
```

There is a declining trend in the forecasted dataset, and the maximum sales forecast is expected to be 66 units.

Sparkling Dataset:

Define the problem and perform Exploratory Data Analysis of sparkling dataset

- Read the data as an appropriate time series data - Plot the data - Perform EDA - Perform Decomposition

Top 5 of the head in the sparkling series

Out[181]:

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

Last 5 of the tail in the sparkling series

Out[182]:

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

Adding the time stamp to the data frame

Out[184]:

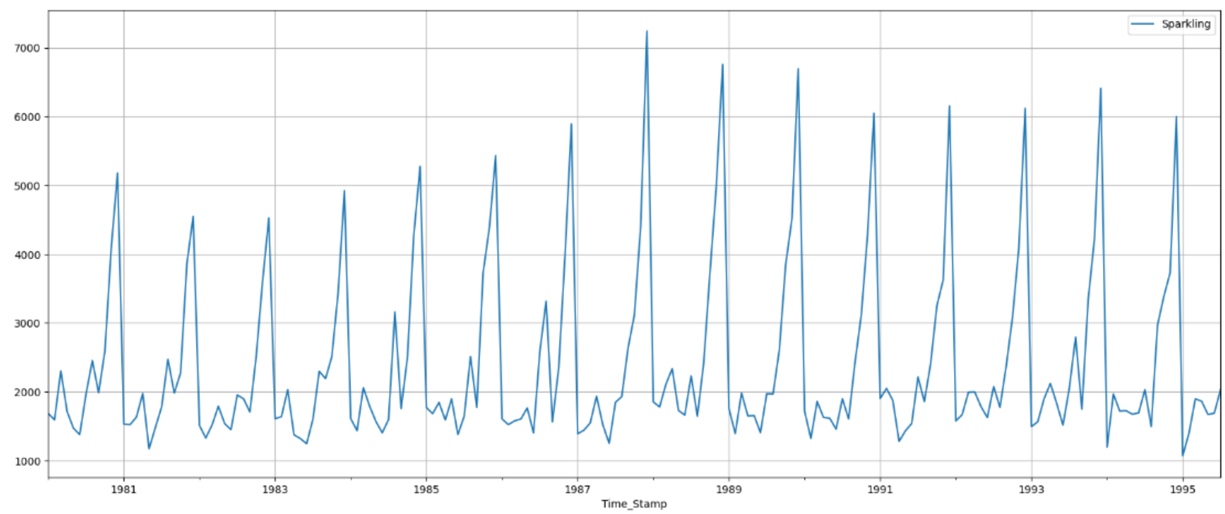
	YearMonth	Sparkling	Time_Stamp
0	1980-01	1686	1980-01-31
1	1980-02	1591	1980-02-29
2	1980-03	2304	1980-03-31
3	1980-04	1712	1980-04-30
4	1980-05	1471	1980-05-31

Out[186]:

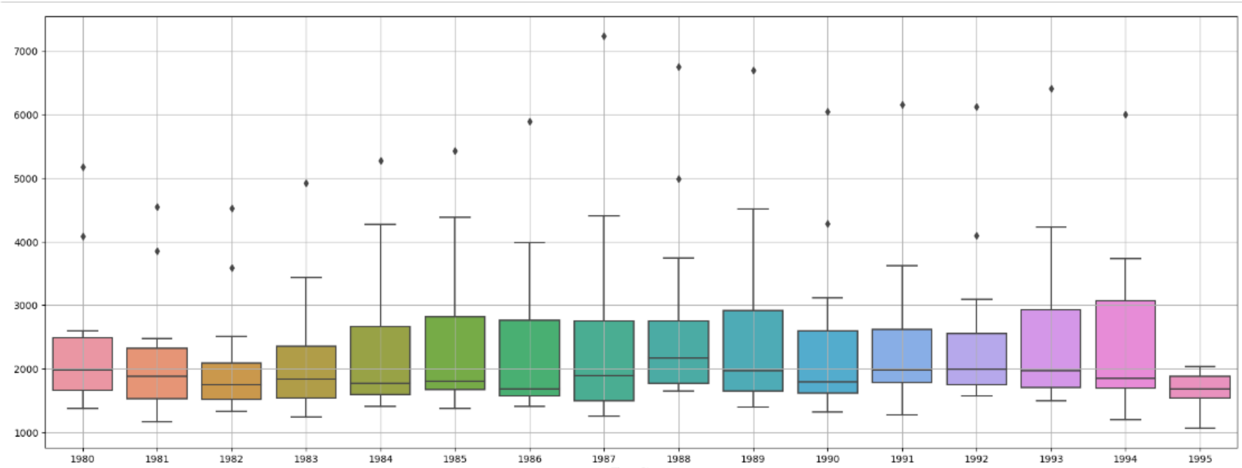
Time_Stamp	Sparkling
1980-01-31	1686
1980-02-29	1591
1980-03-31	2304
1980-04-30	1712
1980-05-31	1471

Out[187]:

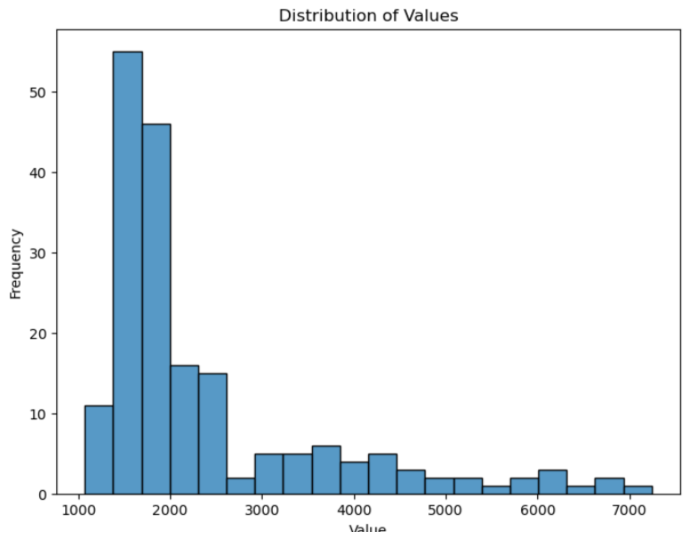
Time_Stamp	Sparkling
1995-03-31	1897
1995-04-30	1862
1995-05-31	1670
1995-06-30	1688
1995-07-31	2031



17. Plotting the timeseries - Sparkling dataset



18. Boxplot on sparkling dataset – yearly



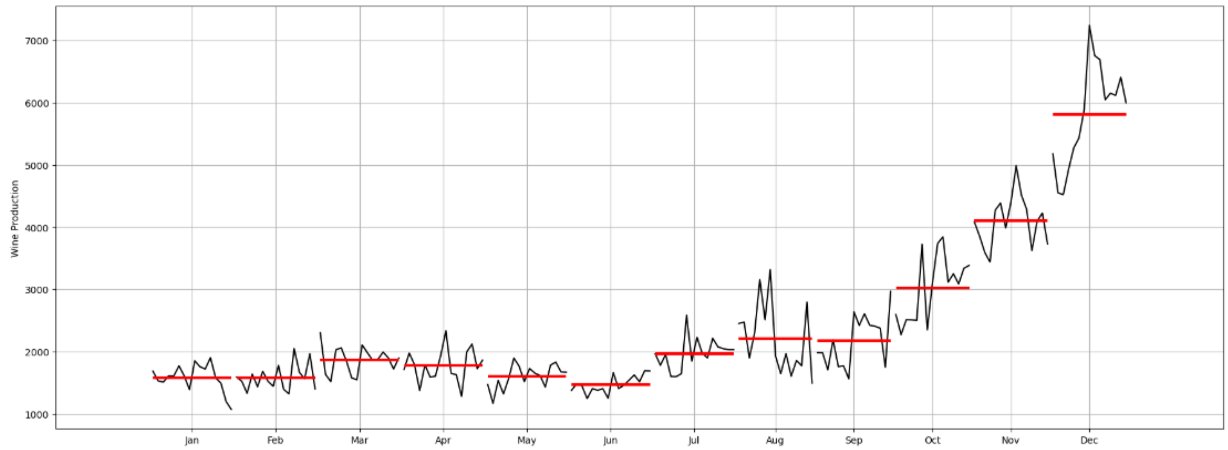
19. Distribution of sprakling dataset

Describe on the sparkling dataset

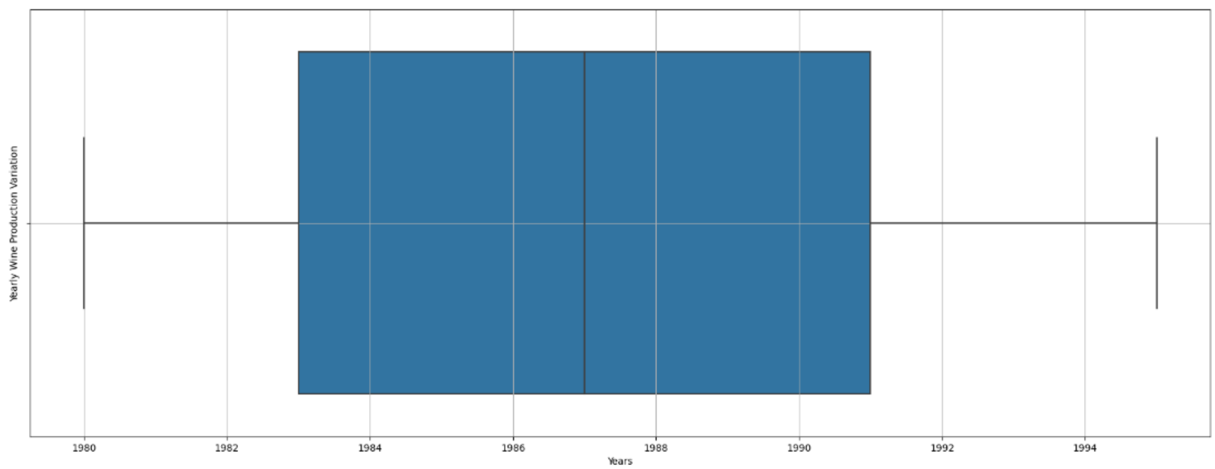
```
          Sparkling
count   187.000000
mean   2402.417112
std    1295.111540
min    1070.000000
25%    1605.000000
50%    1874.000000
75%    2549.000000
max    7242.000000
<class 'pandas.core.frame.DataFrame'>
DatetimeIndex: 187 entries, 1980-01-31 to 1995-07-31
Data columns (total 1 columns):
 #   Column     Non-Null Count  Dtype
---  ------     --------------  -----
 0   Sparkling  187 non-null    int64
dtypes: int64(1)
memory usage: 2.9 KB
None
```

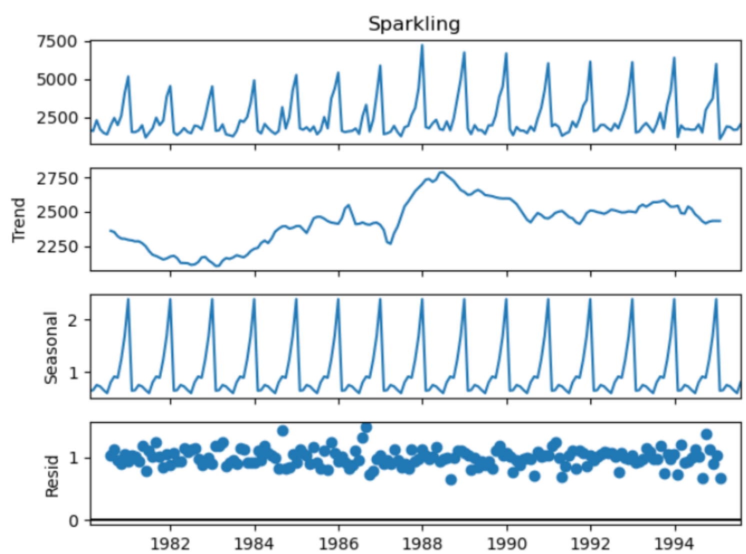
```
Sparkling    0
dtype: int64
```

There are 0 null values found in the sparkling dataset.

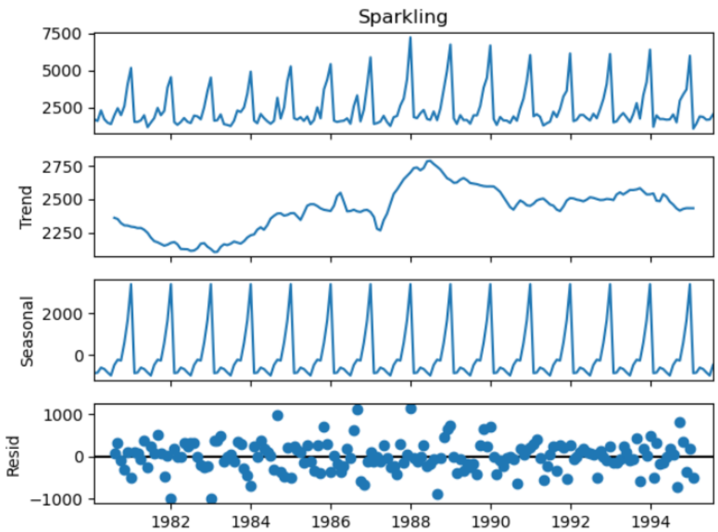


20. Plot a monthplot of the give Time Series.





21. Seasonal decompose - multiplicative method



22. Seasonal decompose - additive method

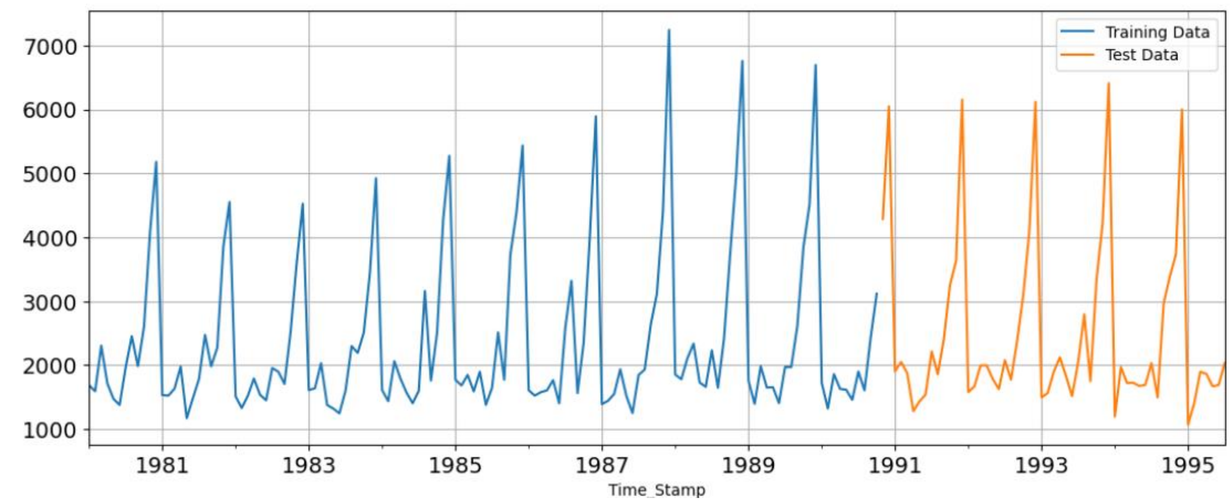
Train – test split

(130, 1)

(57, 1)

There are 130 data in the training dataset and 57 data in the test dataset.

```
First few rows of Training Data
             Sparkling
Time_Stamp
1980-01-31        1686
1980-02-29        1591
1980-03-31        2304
1980-04-30        1712
1980-05-31        1471

Last few rows of Training Data
             Sparkling
Time_Stamp
1990-06-30        1457
1990-07-31        1899
1990-08-31        1605
1990-09-30        2424
1990-10-31        3116

First few rows of Test Data
             Sparkling
Time_Stamp
1990-11-30        4286
1990-12-31        6047
1991-01-31        1902
1991-02-28        2049
1991-03-31        1874

Last few rows of Test Data
             Sparkling
Time_Stamp
1995-03-31        1897
1995-04-30        1862
1995-05-31        1670
1995-06-30        1688
1995-07-31        2031
```



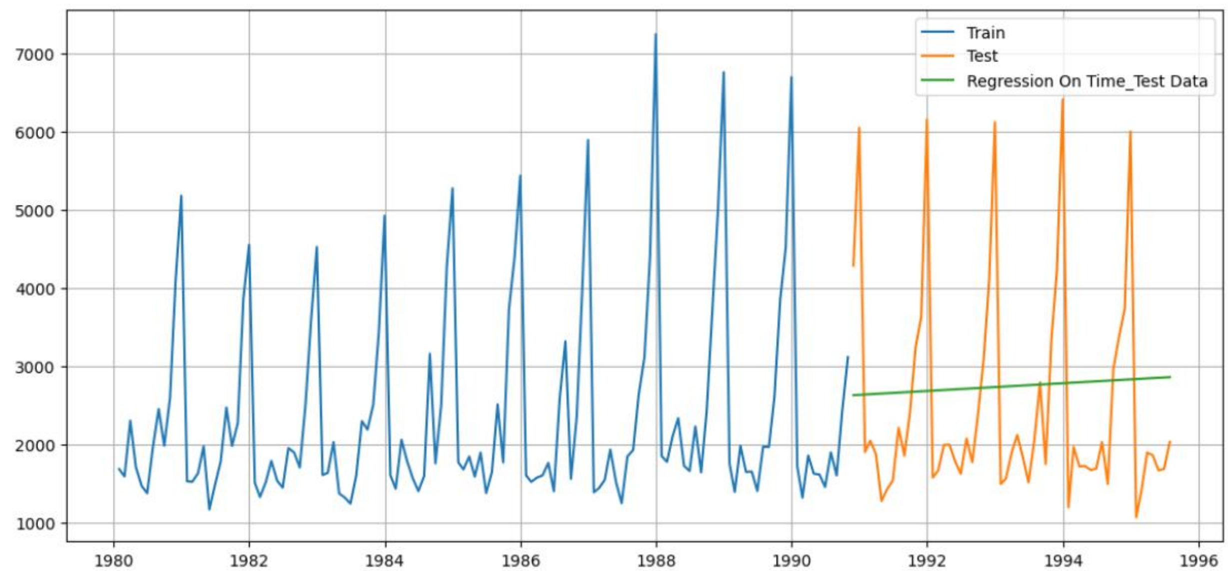
23. Plotting the train-test dataset

Model Building - Original Data

- Build forecasting models - Linear regression - Simple Average - Moving Average - Exponential Models (Single, Double, Triple) - Check the performance of the models built

Model 1: Linear Regression

```
First few rows of Training Data
             Sparkling  time
Time_Stamp
1980-01-31        1686     1
1980-02-29        1591     2
1980-03-31        2304     3
1980-04-30        1712     4
1980-05-31        1471     5

Last few rows of Training Data
             Sparkling  time
Time_Stamp
1990-06-30        1457   126
1990-07-31        1899   127
1990-08-31        1605   128
1990-09-30        2424   129
1990-10-31        3116   130

First few rows of Test Data
             Sparkling  time
Time_Stamp
1990-11-30        4286   131
1990-12-31        6047   132
1991-01-31        1902   133
1991-02-28        2049   134
1991-03-31        1874   135

Last few rows of Test Data
             Sparkling  time
Time_Stamp
1995-03-31        1897   183
1995-04-30        1862   184
1995-05-31        1670   185
1995-06-30        1688   186
1995-07-31        2031   187
```



24. Linear Regression - sparkling dataset

Model Evaluation

For RegressionOnTime forecast on the Test Data, RMSE is 1392.44

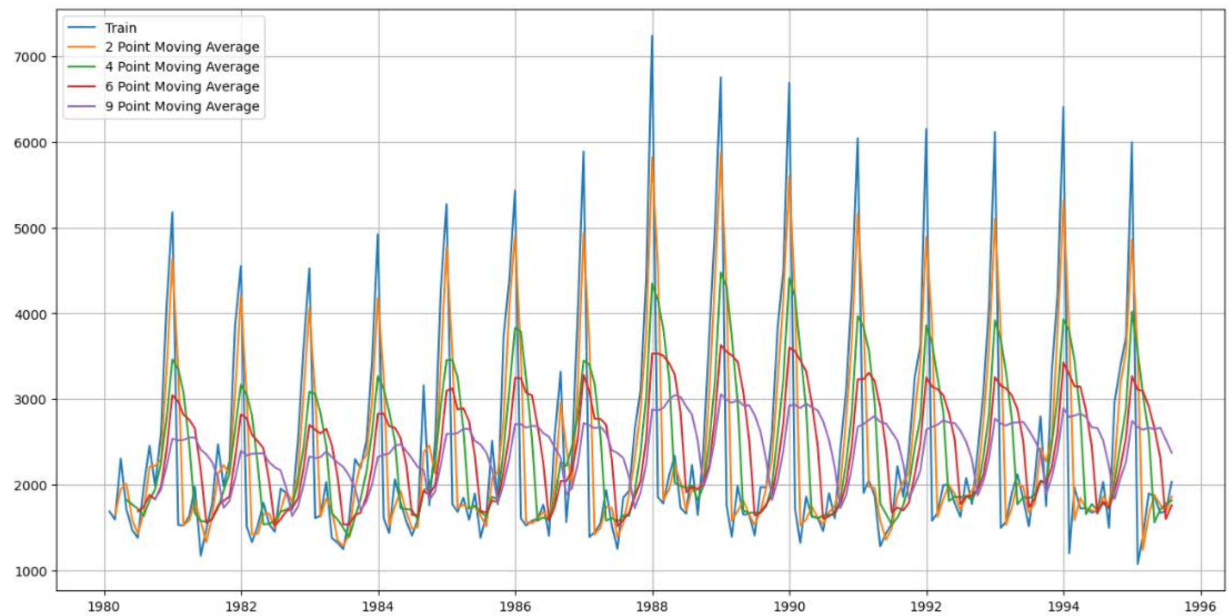
Out[208]:

	Test RMSE
RegressionOnTime	1392.438305

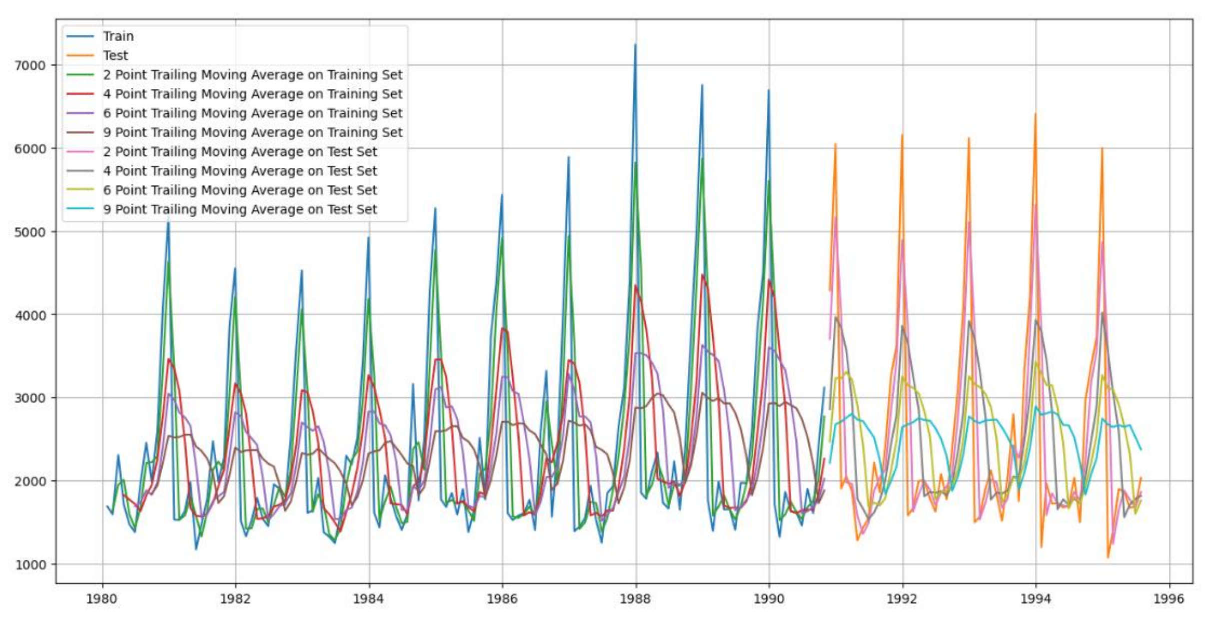
Method 2: Moving Average(MA)

Out[210]:

Time_Stamp	Sparkling	Trailing_2	Trailing_4	Trailing_6	Trailing_9
1980-01-31	1686	NaN	NaN	NaN	NaN
1980-02-29	1591	1638.5	NaN	NaN	NaN
1980-03-31	2304	1947.5	NaN	NaN	NaN
1980-04-30	1712	2008.0	1823.25	NaN	NaN
1980-05-31	1471	1591.5	1769.50	NaN	NaN



25. Plotting on the train - moving average



26. Plotting on the train-test - moving average

For 2 point Moving Average Model forecast on the Training Data, RMSE is 811.179

For 4 point Moving Average Model forecast on the Training Data, RMSE is 1184.213

For 6 point Moving Average Model forecast on the Training Data, RMSE is 1337.201

For 9 point Moving Average Model forecast on the Training Data, RMSE is 1422.653

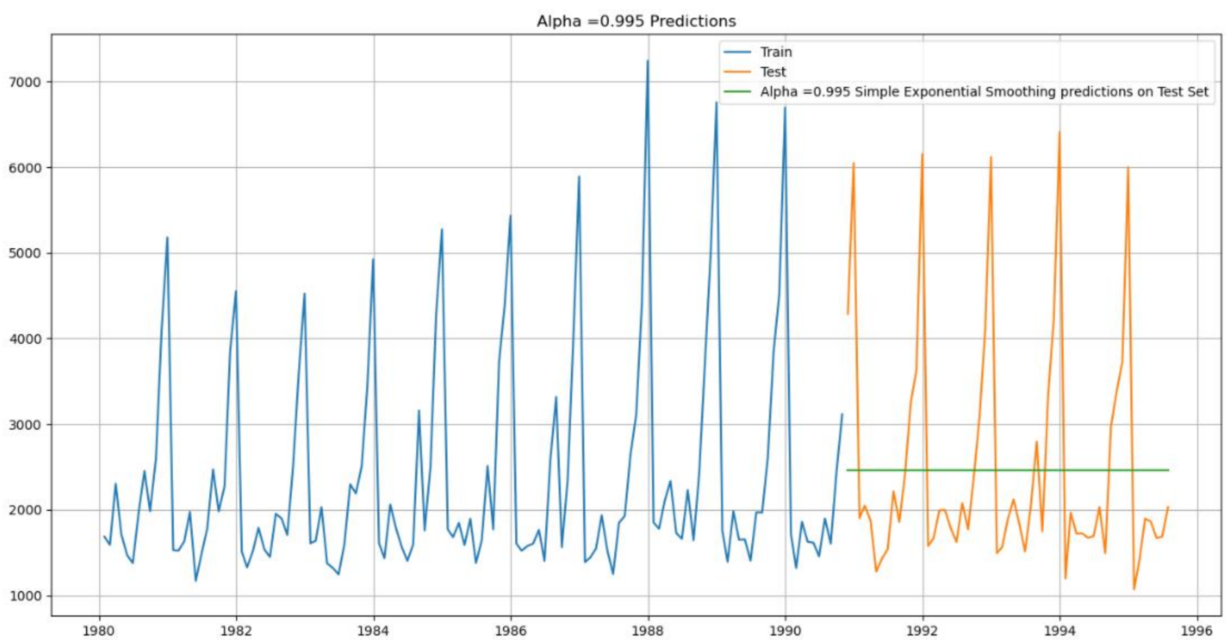
Out[215]:

	Test RMSE
RegressionOnTime	1392.438305
2pointTrailingMovingAverage	811.178937
4pointTrailingMovingAverage	1184.213295
6pointTrailingMovingAverage	1337.200524
9pointTrailingMovingAverage	1422.653281

Method 3: Simple Exponential Smoothing

Out[221]:

	Sparkling	predict
Time_Stamp		
1990-11-30	4286	2465.235699
1990-12-31	6047	2465.235699
1991-01-31	1902	2465.235699
1991-02-28	2049	2465.235699
1991-03-31	1874	2465.235699



27. Alpha =0.995 Simple Exponential Smoothing predictions on Test Set

For Alpha =0.995 Simple Exponential Smoothing Model forecast on the Test Data, RMSE is 1362.429

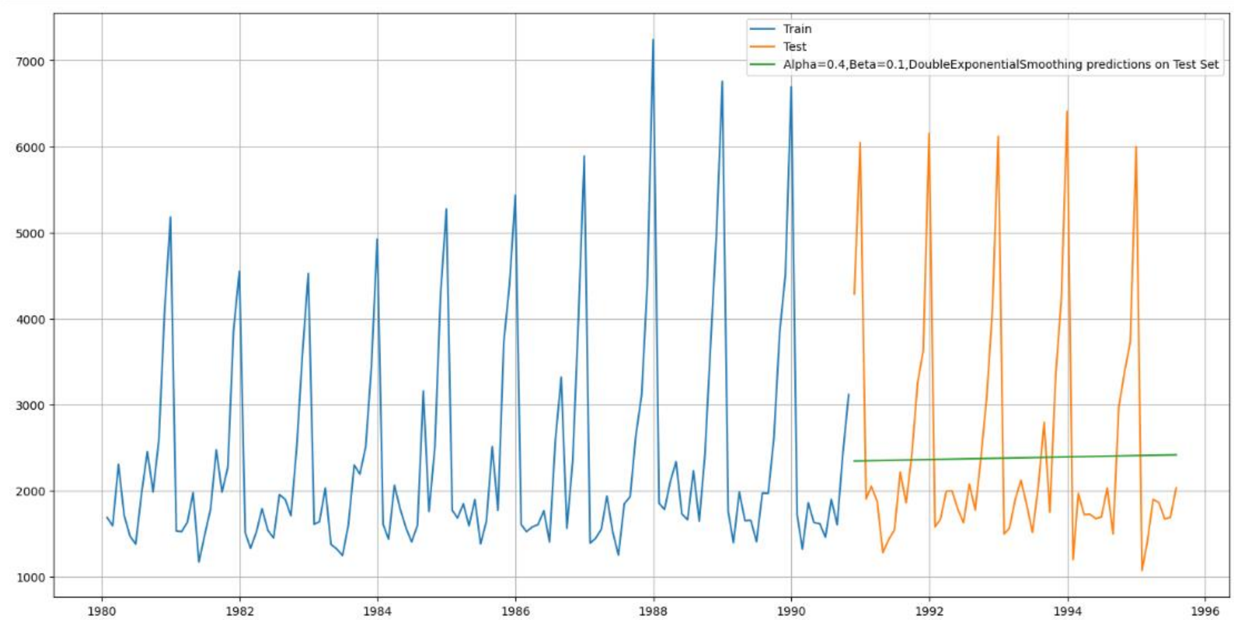
Out[224]:

	Test RMSE
RegressionOnTime	1392.438305
2pointTrailingMovingAverage	811.178937
4pointTrailingMovingAverage	1184.213295
6pointTrailingMovingAverage	1337.200524
9pointTrailingMovingAverage	1422.653281
Alpha=0.995,SimpleExponentialSmoothing	1362.428949

Method 4: Double Exponential Smoothing (Holt's Model)

Out[230]:

	Alpha Values	Beta Values	Train RMSE	Test RMSE
30	0.4	0.1	1402.99	1369.14
21	0.3	0.2	1479.94	1516.79
15	0.2	0.6	1798.09	1540.32
40	0.5	0.1	1396.30	1585.87
22	0.3	0.3	1567.52	1597.85



28. Alpha=0.4,Beta=0.1,DoubleExponentialSmoothing predictions on Test Set

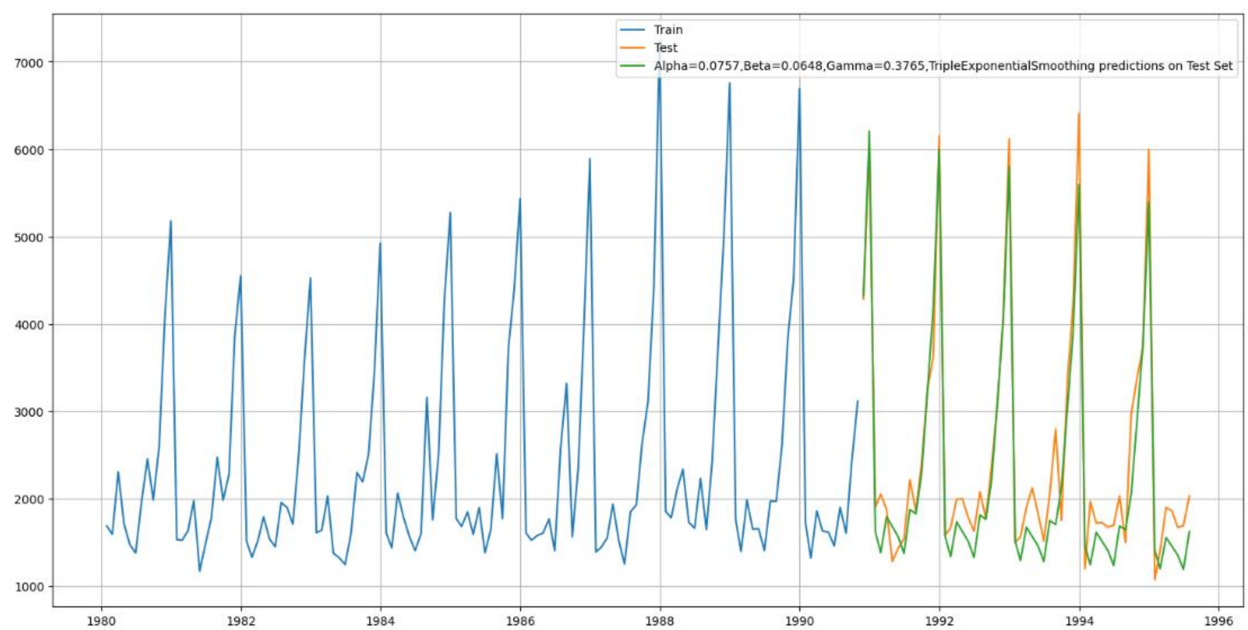
Out[232]:

	Test RMSE
RegressionOnTime	1392.438305
2pointTrailingMovingAverage	811.178937
4pointTrailingMovingAverage	1184.213295
6pointTrailingMovingAverage	1337.200524
9pointTrailingMovingAverage	1422.653281
Alpha=0.995,SimpleExponentialSmoothing	1362.428949
Alpha=0.4,Beta=0.1,DoubleExponentialSmoothing	1369.140000

Method 5: Triple Exponential Smoothing (Holt - Winter's Model)

Out[237]:

Time_Stamp	Sparkling	auto_predict
1990-11-30	4286	4327.597955
1990-12-31	6047	6208.850701
1991-01-31	1902	1621.603392
1991-02-28	2049	1379.864103
1991-03-31	1874	1791.912187



29. Alpha=0.0757,Beta=0.0648,Gamma=0.3765,TripleExponentialSmoothing predictions on Test Set

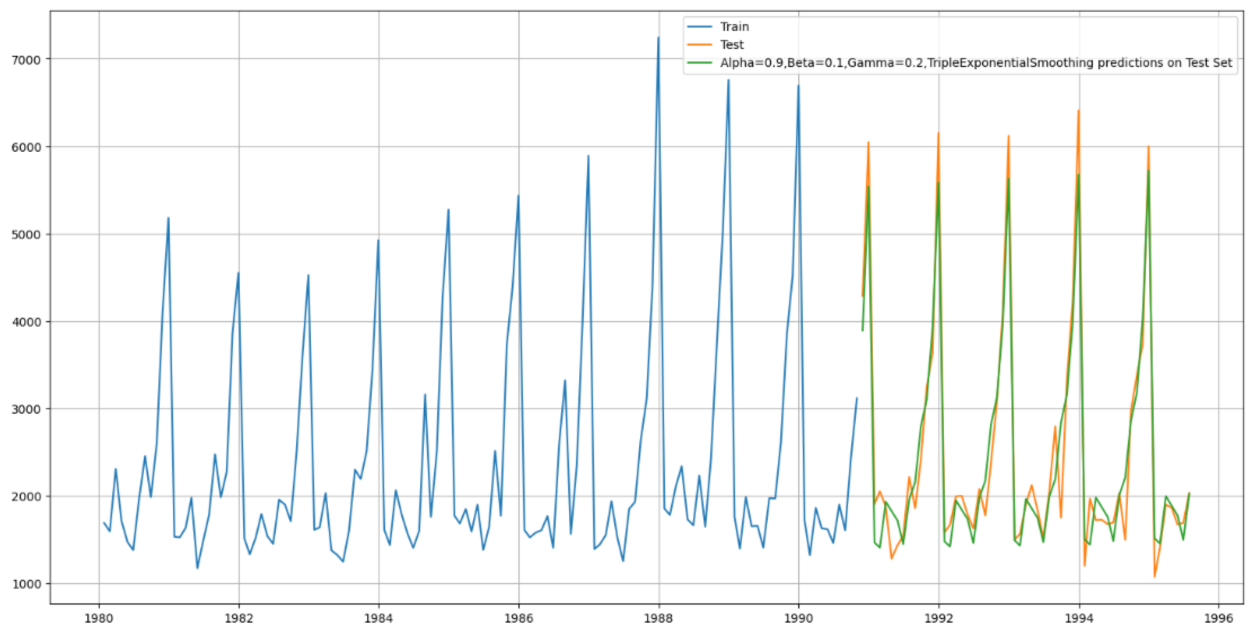
For Alpha=0.0757,Beta=0.0648,Gamma=0.3765, Triple Exponential Smoothing Model forecast on the Test Data, RMSE is 381.656

Out[240]:

	Test RMSE
RegressionOnTime	1392.438305
2pointTrailingMovingAverage	811.178937
4pointTrailingMovingAverage	1184.213295
6pointTrailingMovingAverage	1337.200524
9pointTrailingMovingAverage	1422.653281
Alpha=0.995,SimpleExponentialSmoothing	1362.428949
Alpha=0.4,Beta=0.1,DoubleExponentialSmoothing	1369.140000
Alpha=0.0757,Beta=0.0648,Gamma=0.3765,TripleExponentialSmoothing	381.656471

Out[243]:

	Alpha Values	Beta Values	Gamma Values	Train RMSE	Test RMSE
801	0.9	0.1	0.2	463.62	342.32
500	0.6	0.1	0.1	410.03	345.28
611	0.7	0.2	0.2	443.17	356.77
901	1.0	0.1	0.2	487.03	380.68
128	0.2	0.3	0.9	481.09	385.72

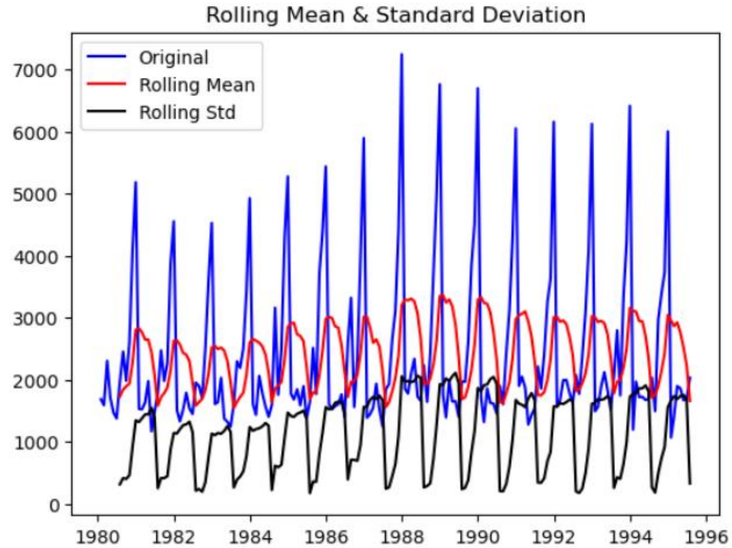


30. Alpha=0.9,Beta=0.1,Gamma=0.2,TripleExponentialSmoothing predictions on Test Set

Out[245]:

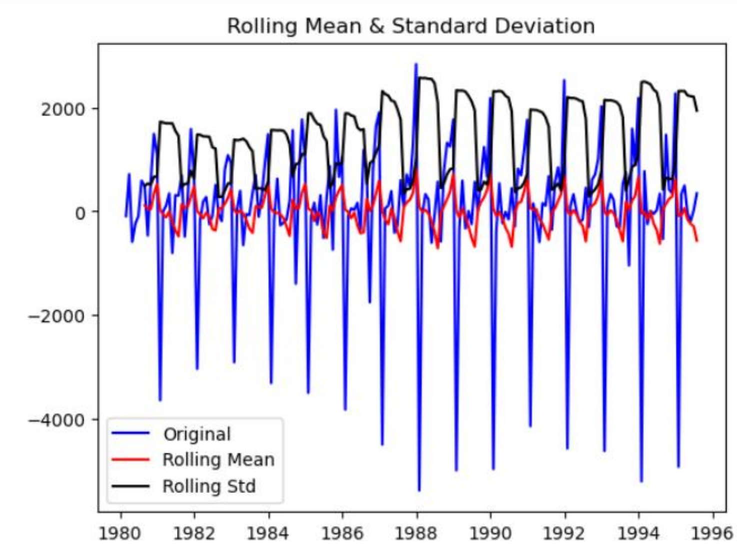
	Test RMSE
RegressionOnTime	1392.438305
2pointTrailingMovingAverage	811.178937
4pointTrailingMovingAverage	1184.213295
6pointTrailingMovingAverage	1337.200524
9pointTrailingMovingAverage	1422.653281
Alpha=0.995,SimpleExponentialSmoothing	1362.428949
Alpha=0.4,Beta=0.1,DoubleExponentialSmoothing	1369.140000
Alpha=0.0757,Beta=0.0648,Gamma=0.3765,TripleExponentialSmoothing	381.656471
Alpha=0.9,Beta=0.1,Gamma=0.2,TripleExponentialSmoothing	342.320000

Check for stationarity of the whole Time Series data




```
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
```

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



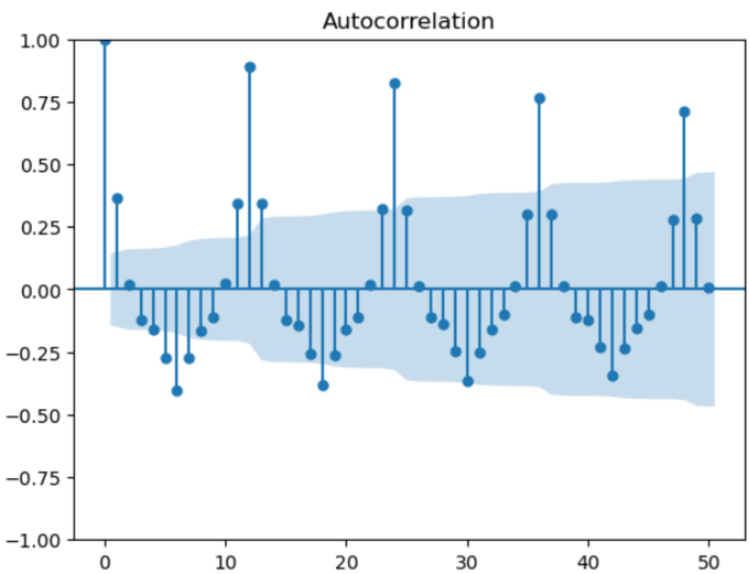

```
Results of Dickey-Fuller Test:
Test Statistic                 -45.050301
p-value                          0.000000
#Lags Used                      10.000000
Number of Observations Used    175.000000
Critical Value (1%)             -3.468280
Critical Value (5%)             -2.878202
Critical Value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

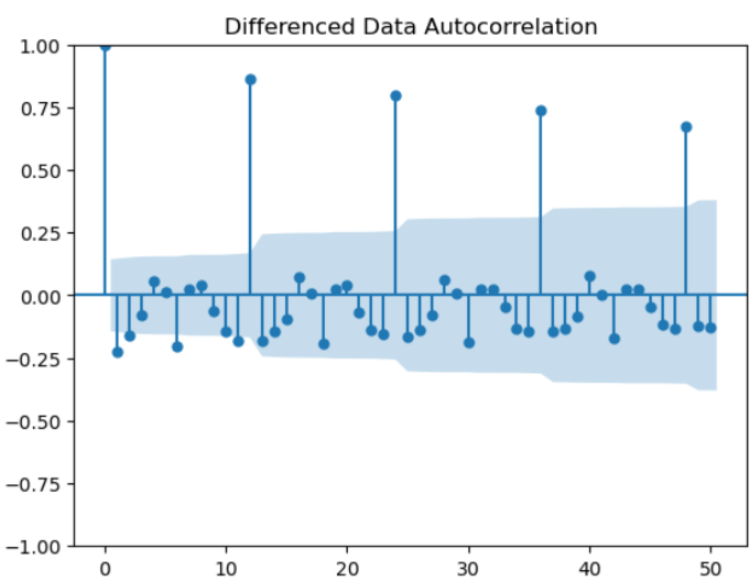
We see that at α = 0.05 the Time Series is indeed stationary.

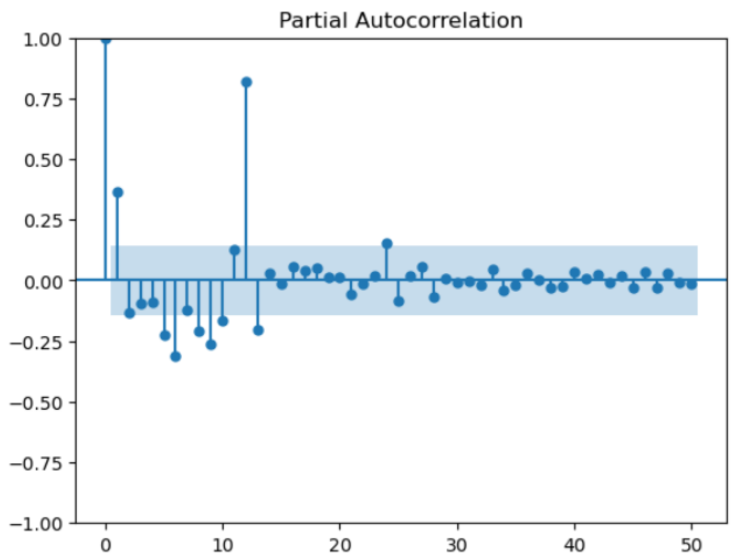
Model Building - Stationary Data

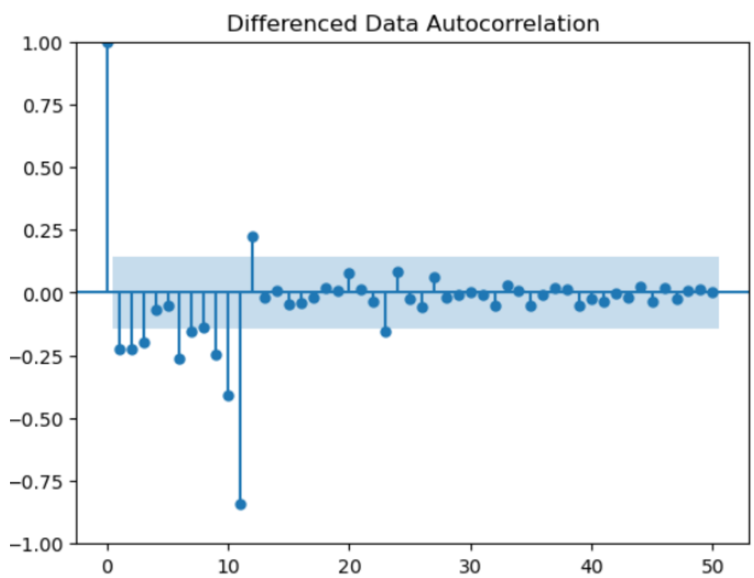
- Generate ACF & PACF Plot and find the AR, MA values. - Build different ARIMA models - Auto ARIMA - Manual ARIMA - Build different SARIMA models - Auto SARIMA - Manual SARIMA - Check the performance of the models built

Plot the Autocorrelation function plots on the whole data.









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

```
Some parameter combinations for the Model...
Model: (0, 0, 1)
Model: (0, 0, 2)
Model: (1, 0, 0)
Model: (1, 0, 1)
Model: (1, 0, 2)
Model: (2, 0, 0)
Model: (2, 0, 1)
Model: (2, 0, 2)
```

Sort the above AIC values in the ascending order to get the parameters for the minimum AIC value

Out[255]:

	param	AIC
7	(2, 0, 1)	2197.084442
1	(0, 0, 1)	2204.869799
6	(2, 0, 0)	2204.880722
2	(0, 0, 2)	2206.111207
4	(1, 0, 1)	2206.142158
5	(1, 0, 2)	2207.163048
3	(1, 0, 0)	2207.502101
8	(2, 0, 2)	2208.120889
0	(0, 0, 0)	2228.483660

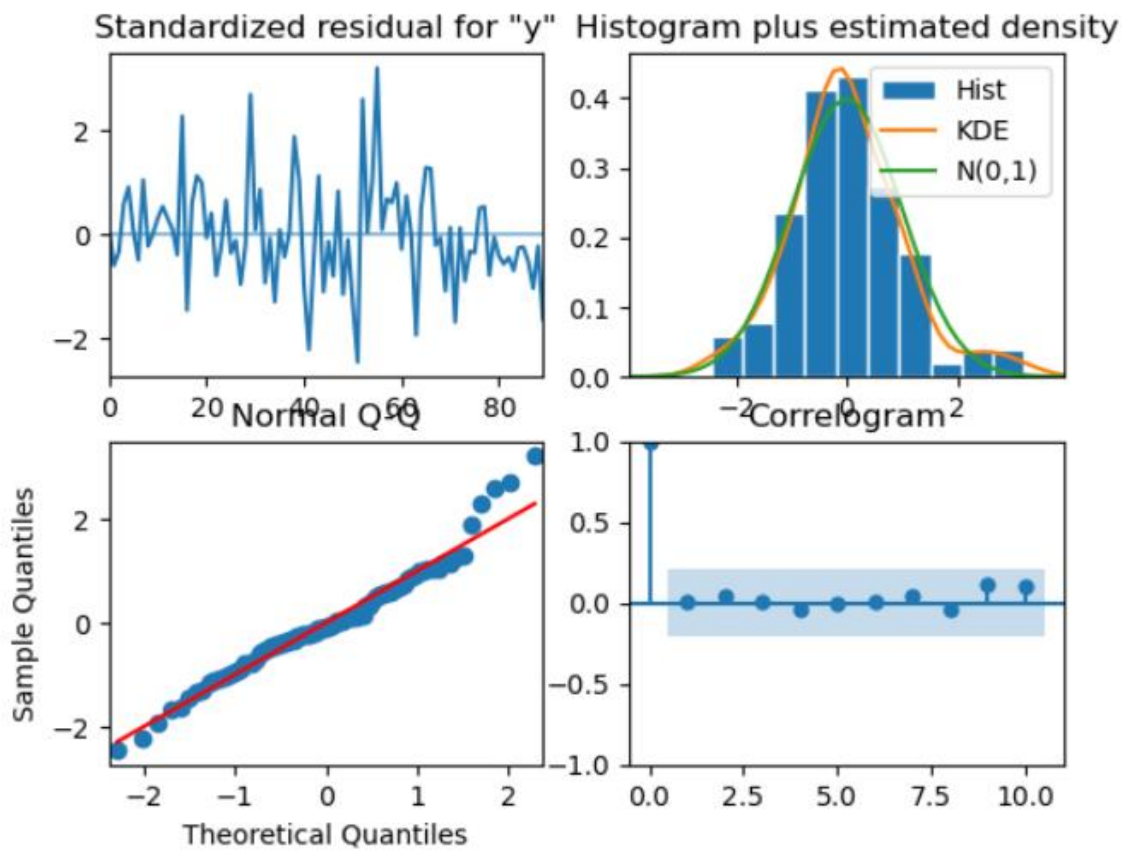
```
                               SARIMAX Results
==============================================================================
Dep. Variable:              Sparkling   No. Observations:                  130
Model:                 ARIMA(2, 0, 1)   Log Likelihood               -1093.542
Date:                Sat, 14 Sep 2024   AIC                           2197.084
Time:                        16:11:13   BIC                           2211.422
Sample:                    01-31-1980   HQIC                          2202.910
                         - 10-31-1990
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const       2379.9376    112.866     21.086      0.000    2158.724    2601.151
ar.L1          1.2114      0.135      8.991      0.000       0.947       1.475
ar.L2         -0.4998      0.124     -4.046      0.000      -0.742      -0.258
ma.L1         -0.8128      0.152     -5.349      0.000      -1.111      -0.515
sigma2      1.182e+06   1.28e+05      9.214      0.000    9.31e+05    1.43e+06
===================================================================================
Ljung-Box (L1) (Q):                   0.02   Jarque-Bera (JB):                39.43
Prob(Q):                              0.90   Prob(JB):                         0.00
Heteroskedasticity (H):               2.19   Skew:                             0.96
Prob(H) (two-sided):                  0.01   Kurtosis:                         4.90
===================================================================================

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

Out[259]:

	Test RMSE
RegressionOnTime	1392.438305
2pointTrailingMovingAverage	811.178937
4pointTrailingMovingAverage	1184.213295
6pointTrailingMovingAverage	1337.200524
9pointTrailingMovingAverage	1422.653281
Alpha=0.995,SimpleExponentialSmoothing	1362.428949
Alpha=0.4,Beta=0.1,DoubleExponentialSmoothing	1369.140000
Alpha=0.0757,Beta=0.0648,Gamma=0.3765,TripleExponentialSmoothing	381.656471
Alpha=0.9,Beta=0.1,Gamma=0.2,TripleExponentialSmoothing	342.320000
ARIMA(2,0,1)	1338.139913

For Auto-ARIMA Model forecast accuracy_score on the Test Data, RMSE is 1338.140

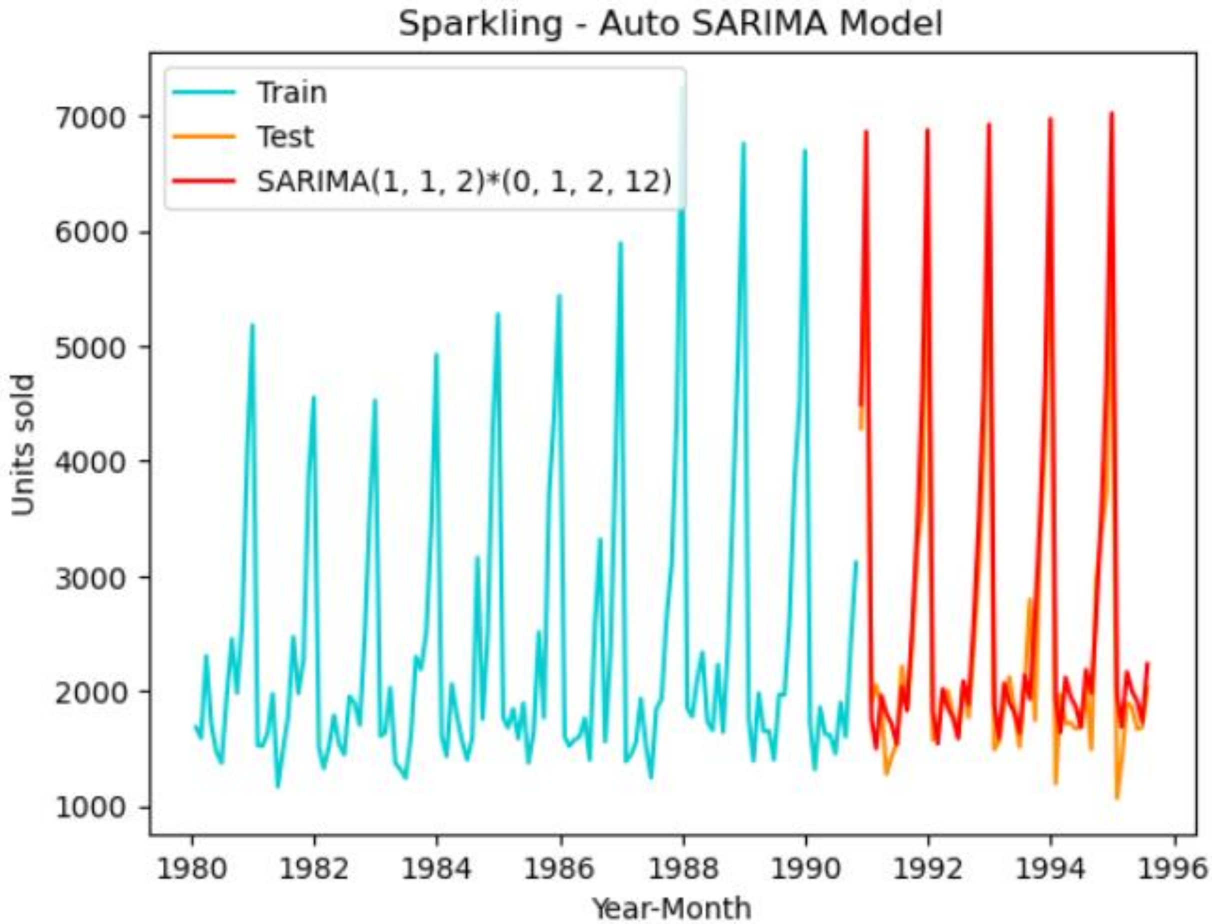
```
                                     SARIMAX Results                                      
==========================================================================================
Dep. Variable:                                  y   No. Observations:                  130
Model:             SARIMAX(1, 1, 2)x(0, 1, 2, 12)   Log Likelihood                -669.850
Date:                            Sat, 14 Sep 2024   AIC                           1351.700
Time:                                    16:15:30   BIC                           1366.699
Sample:                                         0   HQIC                          1357.749
                                            - 130                                         
Covariance Type:                              opg                                         
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.5355      0.225     -2.380      0.017      -0.977      -0.094
ma.L1         -0.2232      0.259     -0.862      0.389      -0.731       0.284
ma.L2         -0.7768      0.162     -4.802      0.000      -1.094      -0.460
ma.S.L12      -0.3970      0.094     -4.231      0.000      -0.581      -0.213
ma.S.L24      -0.0105      0.137     -0.077      0.938      -0.278       0.257
sigma2      1.649e+05   1.47e-06   1.12e+11      0.000    1.65e+05    1.65e+05
===================================================================================
Ljung-Box (L1) (Q):                   0.01   Jarque-Bera (JB):                 8.58
Prob(Q):                              0.92   Prob(JB):                         0.01
Heteroskedasticity (H):               0.83   Skew:                             0.50
Prob(H) (two-sided):                  0.62   Kurtosis:                         4.14
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
[2] Covariance matrix is singular or near-singular, with condition number 2.03e+26. Standard errors may be unstable.
```



Predict on the Test Set using this model and evaluate the model.

Out[272]:

	Test RMSE
SARIMA(1, 1, 2)*(0, 1, 2, 12)	440.186889

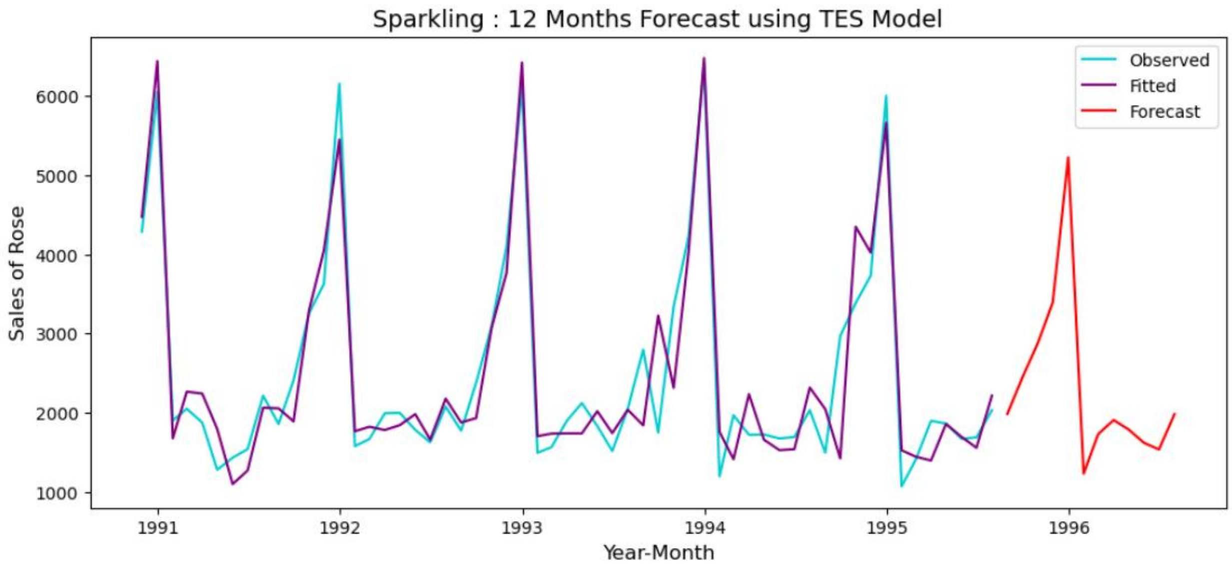


Actionable Insights & Recommendations

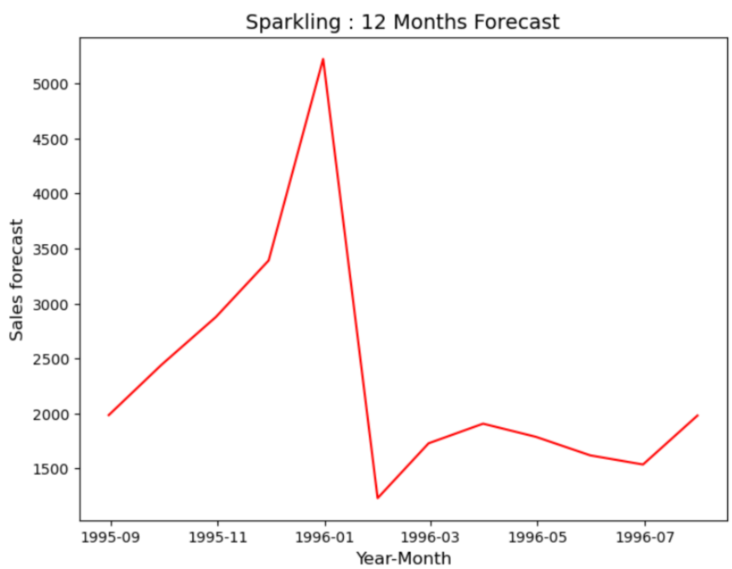
- Conclude with the key takeaways (actionable insights and recommendations) for the business

Based on the results, Triple exponential smoorthing has lower RMSE of 342.320000 Alpha=0.9,Beta=0.1,Gamma=0.2,TripleExponentialSmoothing

For Triple Exponential Smoothing Model forecast on the Entire Data, RMSE is 465.106



31. Sparkling : 12 Months Forecast using TES Model



32. Sparkling : 12 Months Forecast

```
Out[279]: count       12.000000
          mean      2308.663554
          std       1096.319790
          min       1229.894929
          25%       1700.977209
          50%       1943.818134
          75%       2550.539170
          max       5222.324482
          dtype: float64
```

The maximum sales forecast is expected to be 5222 units in the starting month of 1996 and the sales will tend to decrease to 1229.89 between the first month and the third month of 1996.