9/25/2022

Time Series Forecasting

Sparkling.CSV

Name – Vivek Augustine

Table of Contents:

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.

Data set for the Problem: Rose.csv

Please do perform the following questions on this set.

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

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

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

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

5. Check for the stationarity of the data on which the model is being built on using appropriate statistical tests and 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.

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

7. Build 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.

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

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.

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

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

Table

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

Description

	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

Shape

```
(187, 2)
```

Duplicates

0

Null Values

```
YearMonth    0
Sparkling    0
```

Table

	YearMonth	Sparkling	Date
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

Information

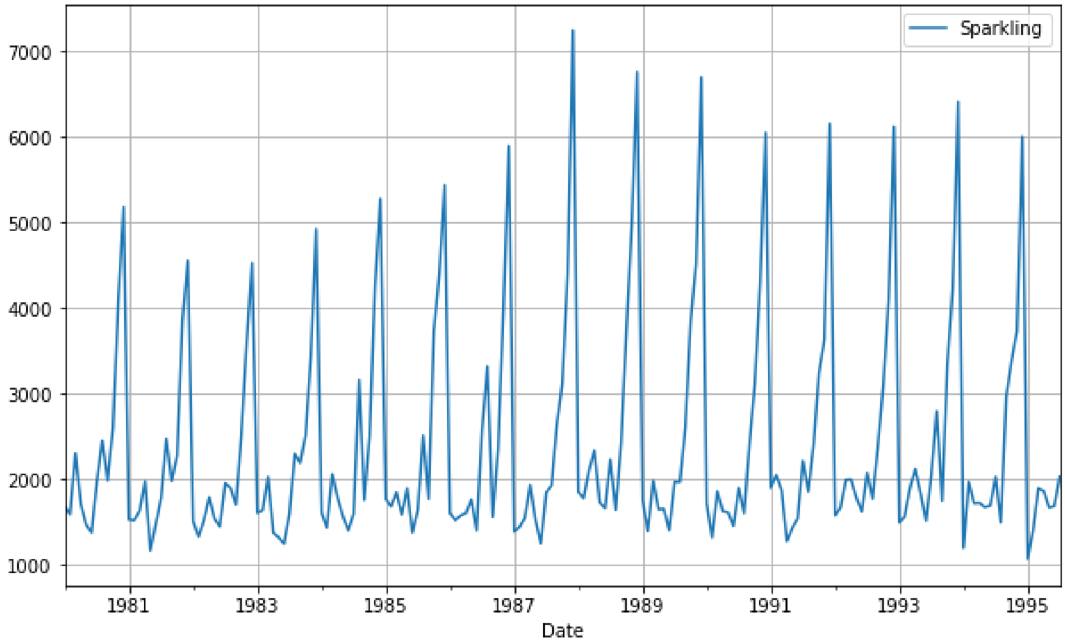
```
#   Column     Non-Null Count  Dtype
---  ------     --------------  -----
 0   Sparkling  187 non-null    int64
 1   Date       187 non-null    datetime64[ns]
```

Making Index

	Sparkling
Date	
1980-01-31	1686
1980-02-29	1591
1980-03-31	2304
1980-04-30	1712

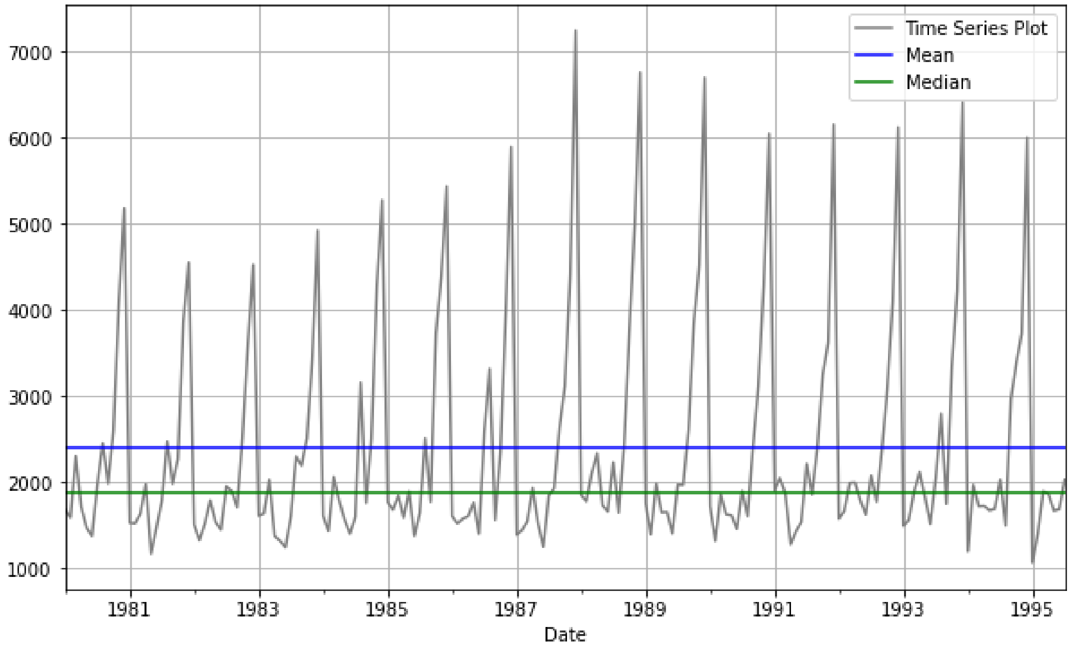
	Sparkling
Date	
1980-05-31	1471

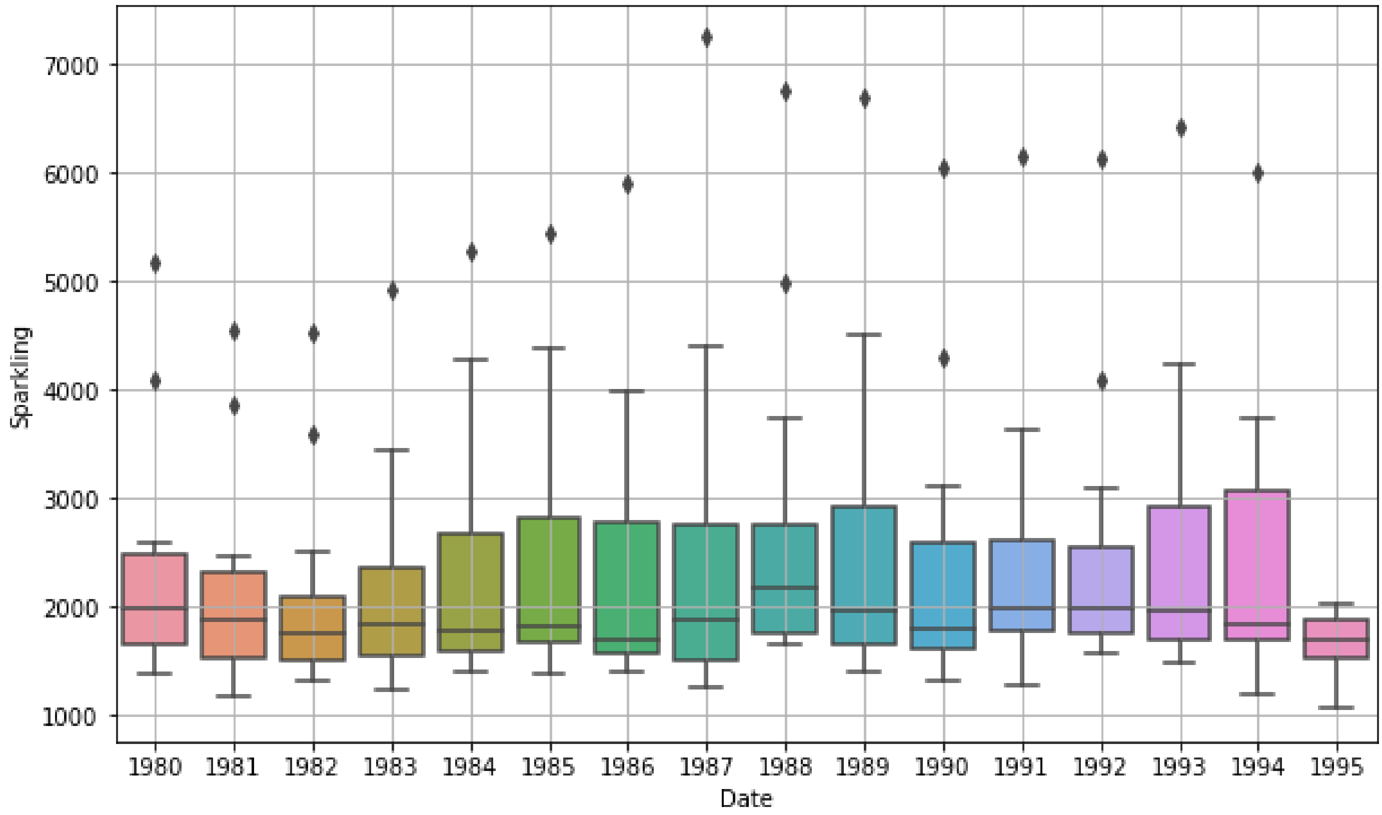
Plotting data



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

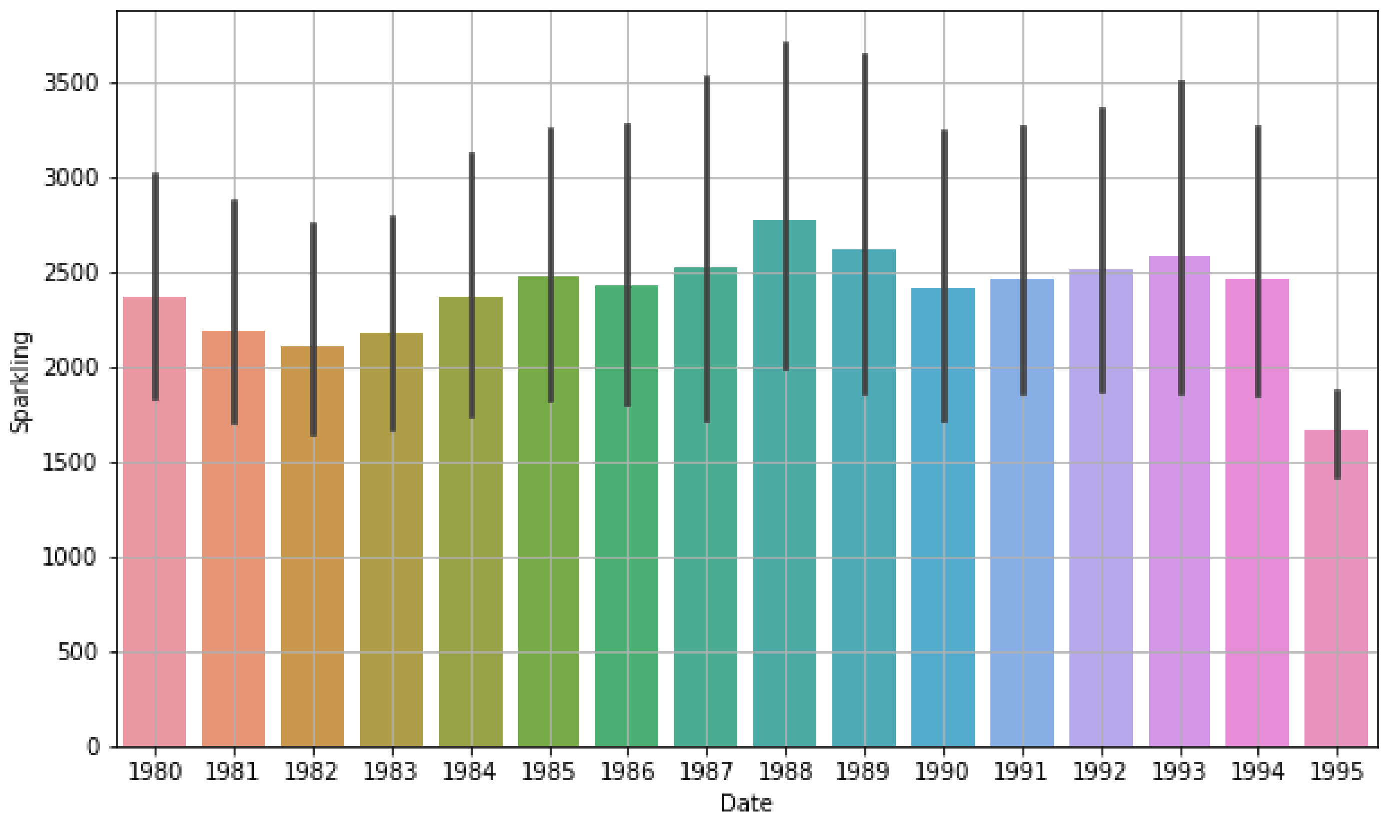
'Time Series Plot','Mean','Median'





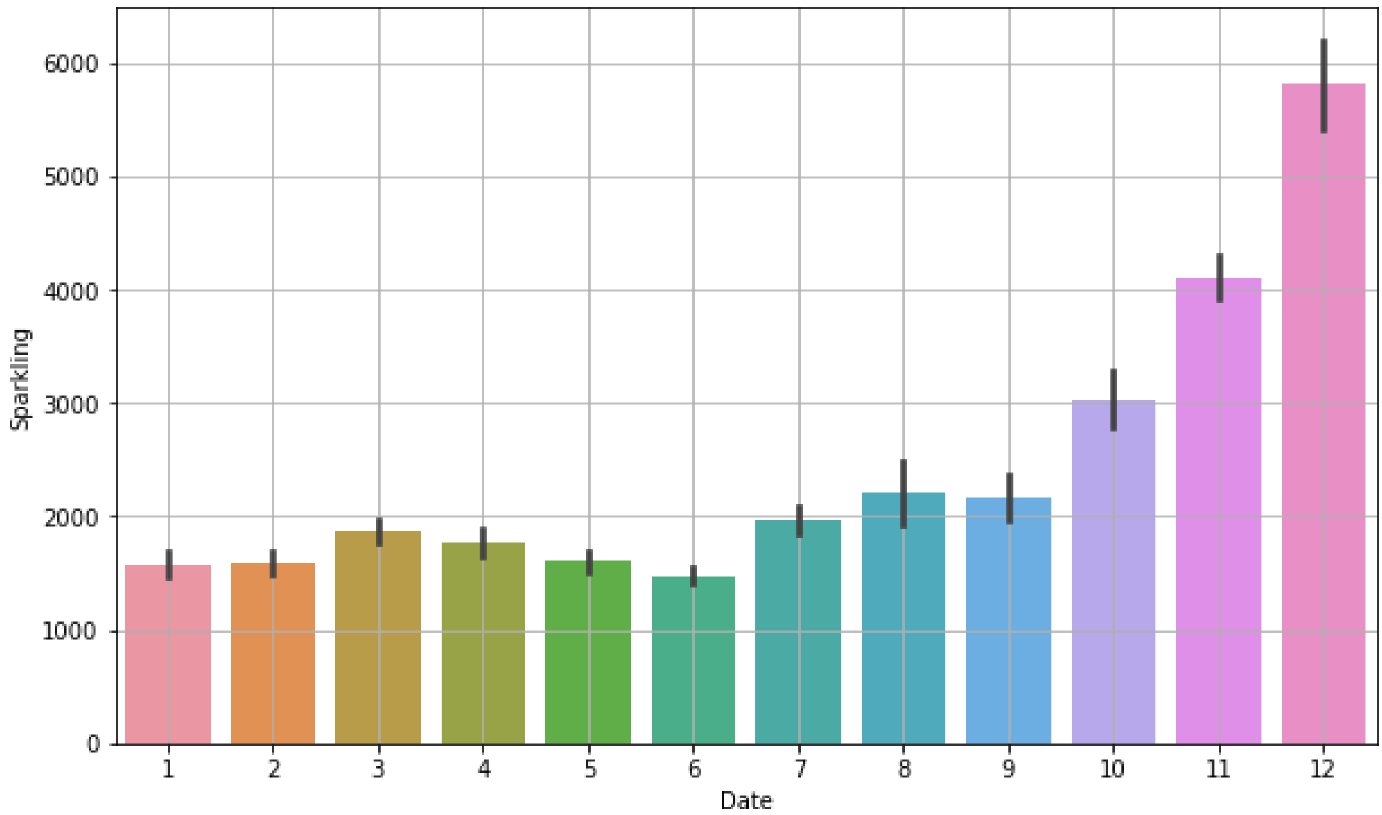
Here, we can see that in 1988 sales are at its peak and in 1995 the sales are significantly decreased and we can also see an overall yearly trend which shows us that the sale started at a decent rate and then began to increse and then by the year 1995 it decreased.

Barplot

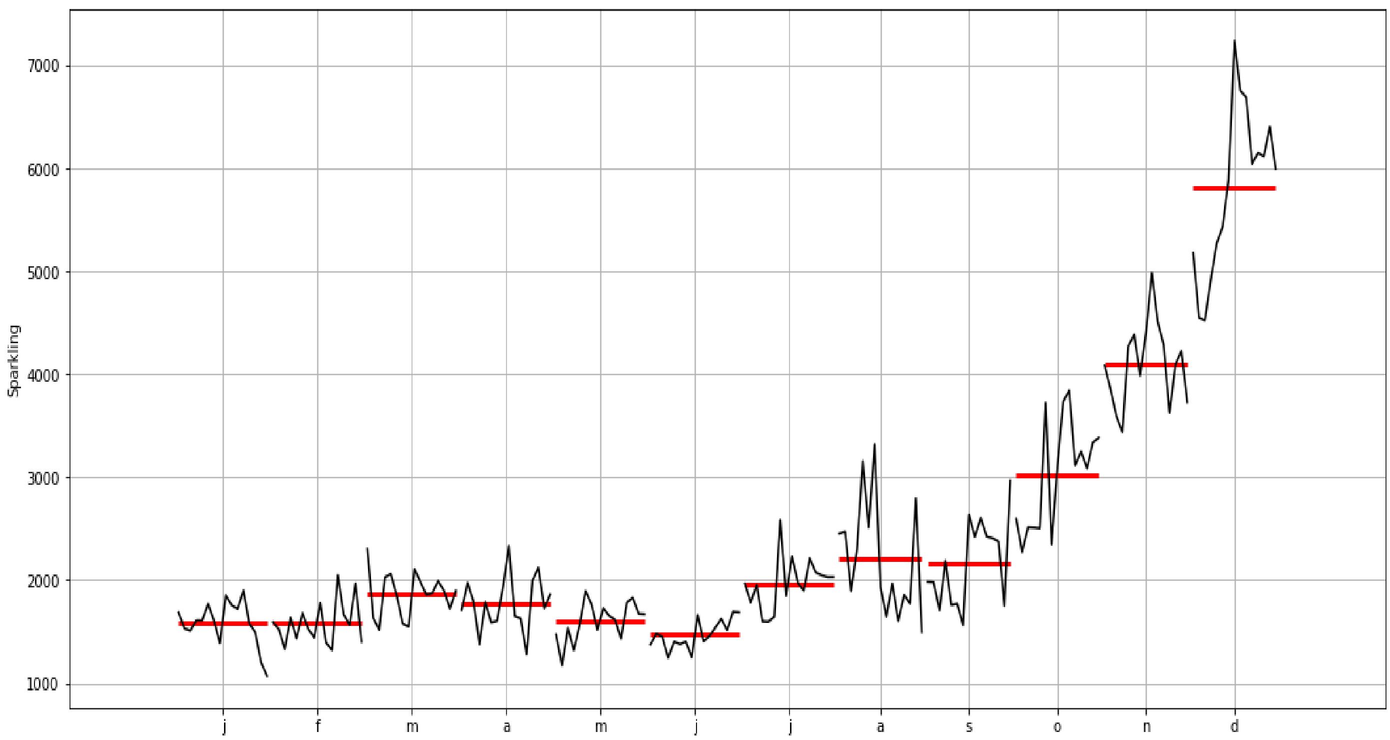


Here we can see the sales and the trend according to month we can see that in the month of dec the sales are at its peak and a trend which can be seen increasing. In the month of jan the company didn't started well and was running low on sales but we can see decent amount of sales in the month of August and september and then from their it increased.

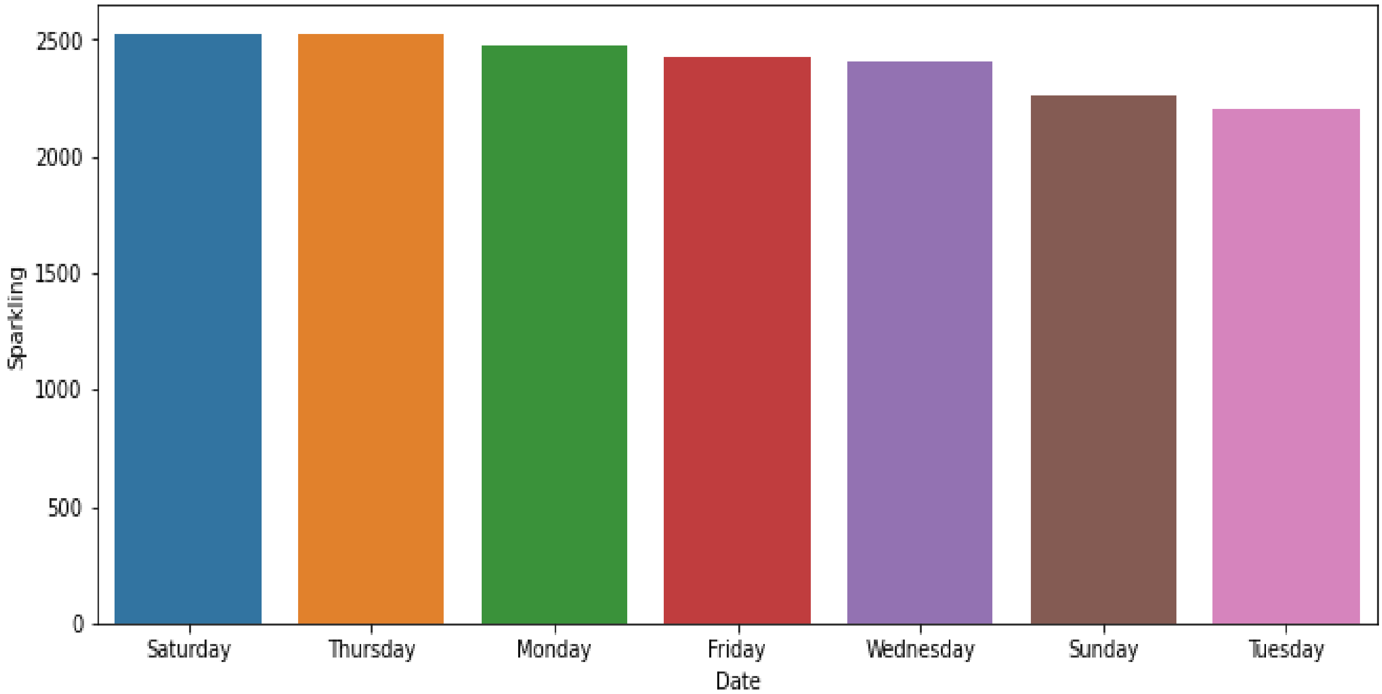
Barplot



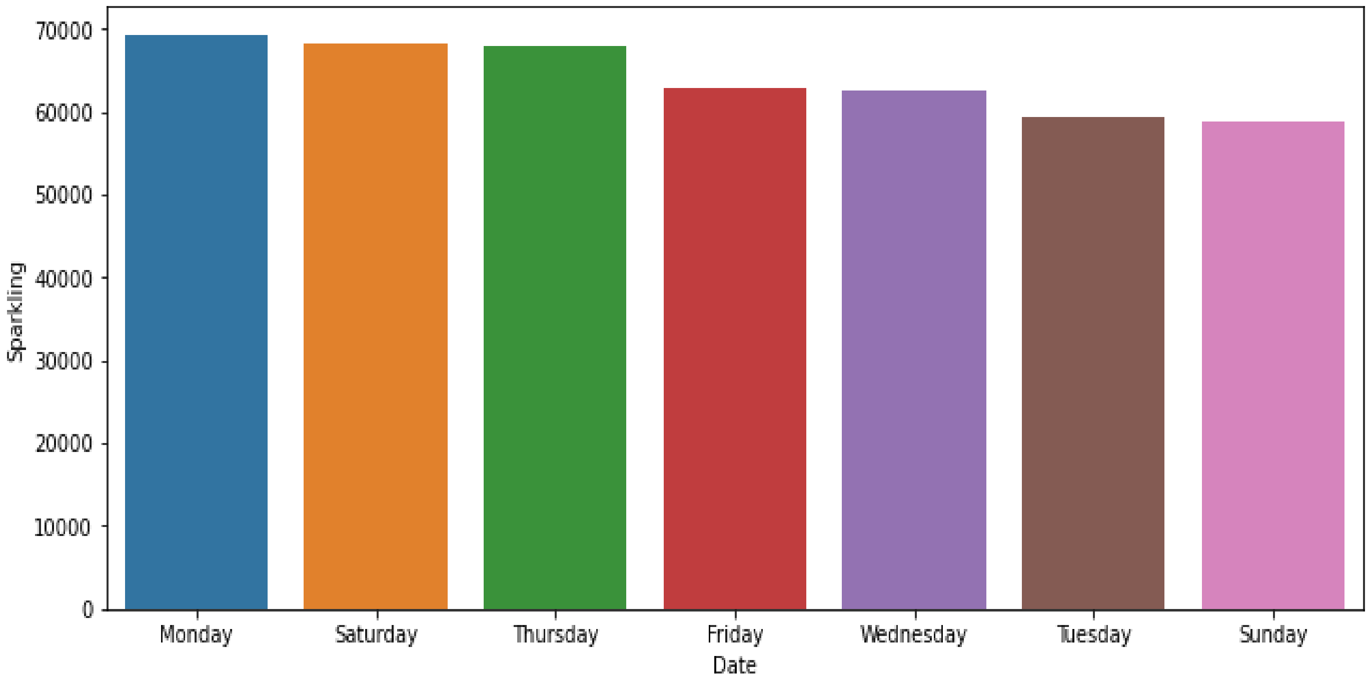
Month plot



Average sales of beer throughout the whole week



highest sales overall

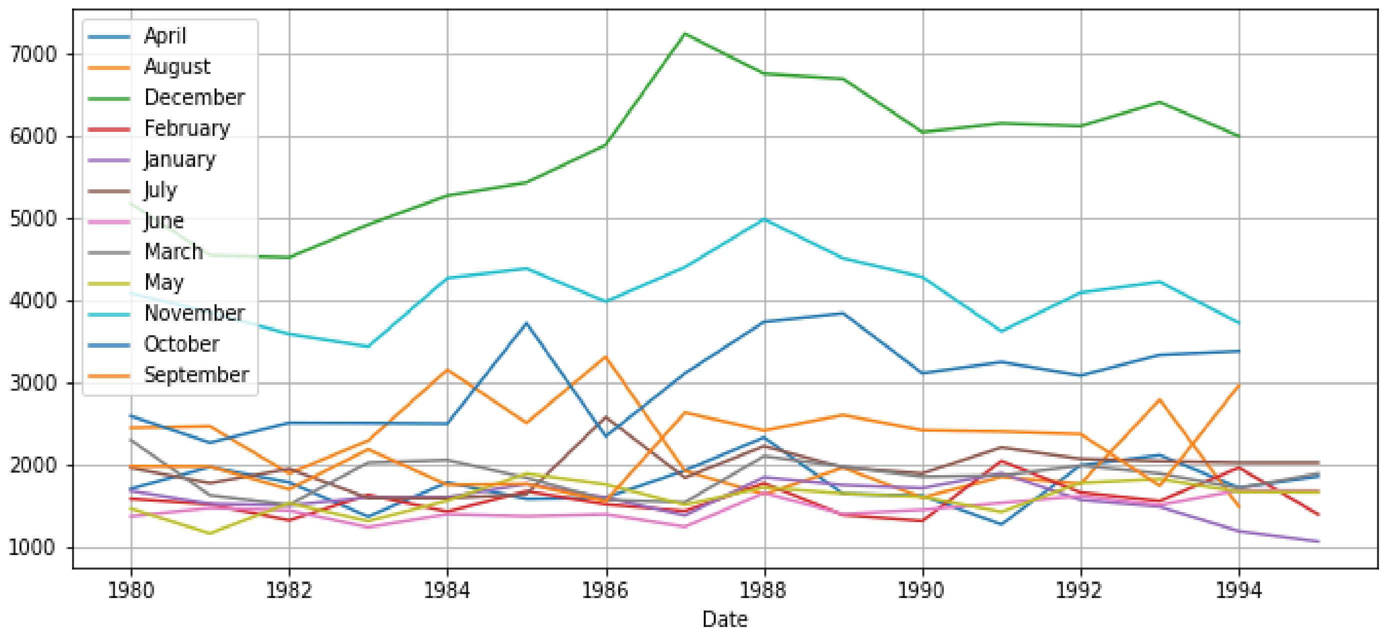


Pivot Table

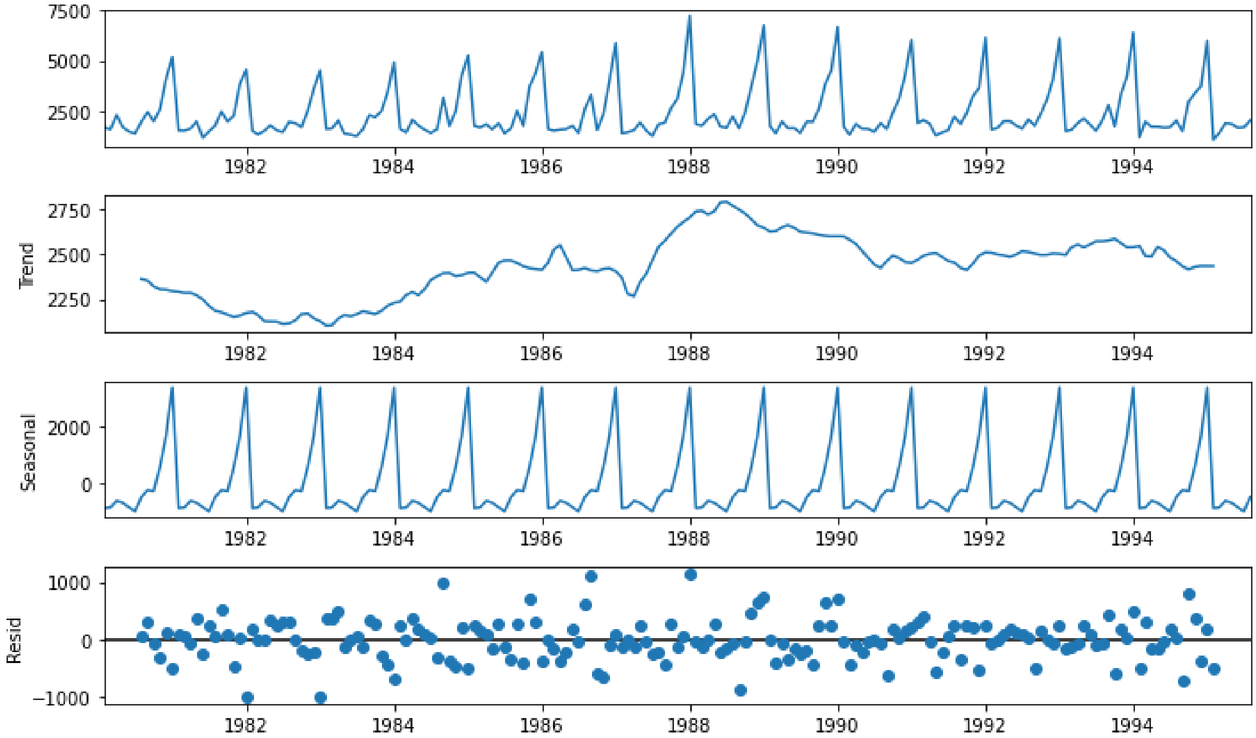
Date	April	August	December	February	January	July	June	March	May	November	October	September
Date												
1980	1712.0	2453.0	5179.0	1591.0	1686.0	1966.0	1377.0	2304.0	1471.0	4087.0	2596.0	1984.0
1981	1976.0	2472.0	4551.0	1523.0	1530.0	1781.0	1480.0	1633.0	1170.0	3857.0	2273.0	1981.0
1982	1790.0	1897.0	4524.0	1329.0	1510.0	1954.0	1449.0	1518.0	1537.0	3593.0	2514.0	1706.0
1983	1375.0	2298.0	4923.0	1638.0	1609.0	1600.0	1245.0	2030.0	1320.0	3440.0	2511.0	2191.0
1984	1789.0	3159.0	5274.0	1435.0	1609.0	1597.0	1404.0	2061.0	1567.0	4273.0	2504.0	1759.0
1985	1589.0	2512.0	5434.0	1682.0	1771.0	1645.0	1379.0	1846.0	1896.0	4388.0	3727.0	1771.0
1986	1605.0	3318.0	5891.0	1523.0	1606.0	2584.0	1403.0	1577.0	1765.0	3987.0	2349.0	1562.0
1987	1935.0	1930.0	7242.0	1442.0	1389.0	1847.0	1250.0	1548.0	1518.0	4405.0	3114.0	2638.0
1988	2336.0	1645.0	6757.0	1779.0	1853.0	2230.0	1661.0	2108.0	1728.0	4988.0	3740.0	2421.0
1989	1650.0	1968.0	6694.0	1394.0	1757.0	1971.0	1406.0	1982.0	1654.0	4514.0	3845.0	2608.0
1990	1628.0	1605.0	6047.0	1321.0	1720.0	1899.0	1457.0	1859.0	1615.0	4286.0	3116.0	2424.0
1991	1279.0	1857.0	6153.0	2049.0	1902.0	2214.0	1540.0	1874.0	1432.0	3627.0	3252.0	2408.0
1992	1997.0	1773.0	6119.0	1667.0	1577.0	2076.0	1625.0	1993.0	1783.0	4096.0	3088.0	2377.0
1993	2121.0	2795.0	6410.0	1564.0	1494.0	2048.0	1515.0	1898.0	1831.0	4227.0	3339.0	1749.0

Date	April	August	December	February	January	July	June	March	May	November	October	September
Date												
1994	1725.0	1495.0	5999.0	1968.0	1197.0	2031.0	1693.0	1720.0	1674.0	3729.0	3385.0	2968.0
1995	1862.0	NaN	NaN	1402.0	1070.0	2031.0	1688.0	1897.0	1670.0	NaN	NaN	NaN

Amount of sales



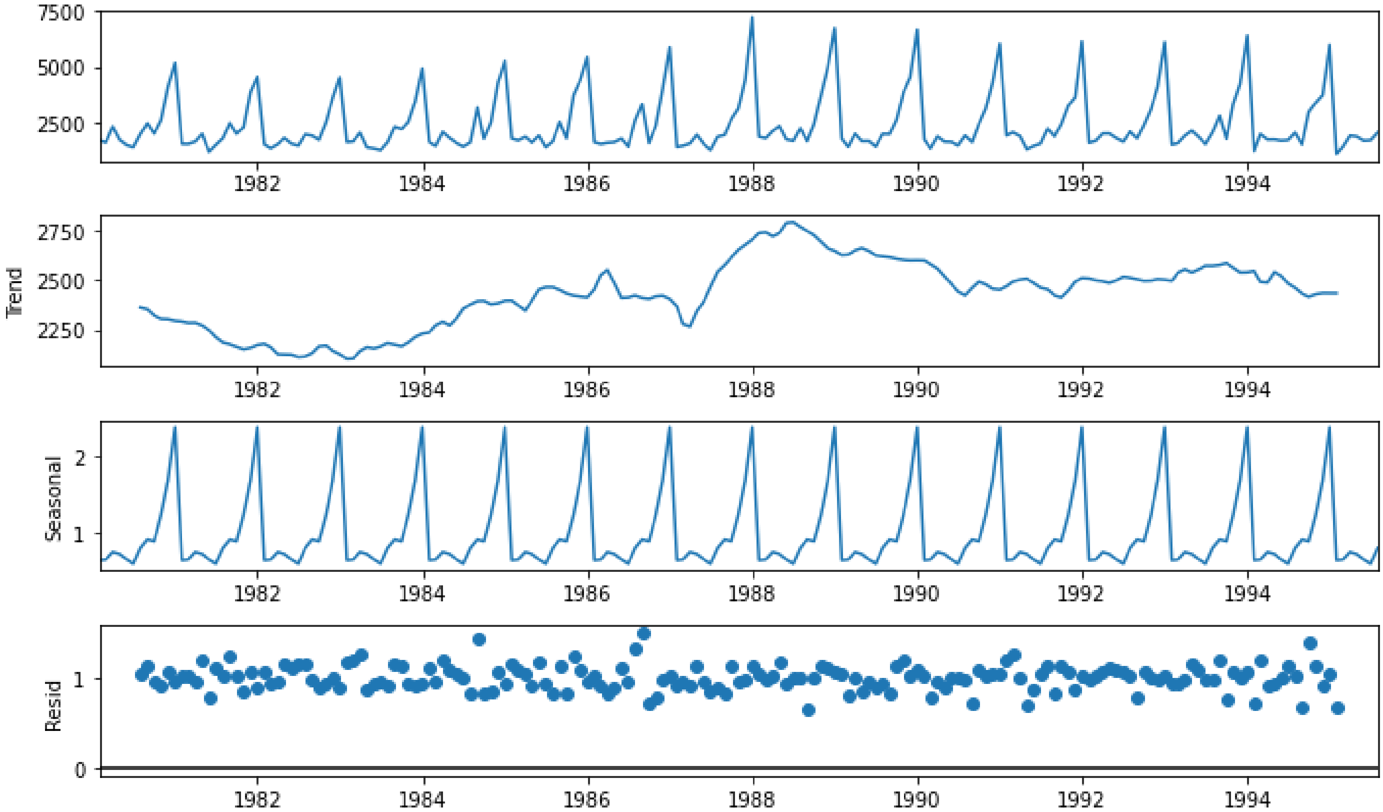
Decompostion(Additive)



Trend, seasonality and Residual

```
Trend
 Date
1980-01-31            NaN
1980-02-29            NaN
1980-03-31            NaN
1980-04-30            NaN
1980-05-31            NaN
1980-06-30            NaN
1980-07-31    2360.666667
1980-08-31    2351.333333
1980-09-30    2320.541667
1980-10-31    2303.583333
1980-11-30    2302.041667
1980-12-31    2293.791667

Seasonality
 Date
1980-01-31    -854.260599
1980-02-29    -830.350678
1980-03-31    -592.356630
1980-04-30    -658.490559
1980-05-31    -824.416154
1980-06-30    -967.434011
1980-07-31    -465.502265
1980-08-31    -214.332821
1980-09-30    -254.677265
1980-10-31     599.769957
1980-11-30    1675.067179
1980-12-31    3386.983846

Residual
 Date
1980-01-31           NaN
1980-02-29           NaN
1980-03-31           NaN
1980-04-30           NaN
1980-05-31           NaN
1980-06-30           NaN
1980-07-31     70.835599
1980-08-31    315.999487
1980-09-30    -81.864401
1980-10-31   -307.353290
1980-11-30    109.891154
1980-12-31   -501.775513
```

Decomposition(multiplicative)



Trend, seasonality and Residual

```
Trend
 Date
1980-01-31            NaN
1980-02-29            NaN
1980-03-31            NaN
1980-04-30            NaN
1980-05-31            NaN
1980-06-30            NaN
1980-07-31    2360.666667
1980-08-31    2351.333333
1980-09-30    2320.541667
1980-10-31    2303.583333
1980-11-30    2302.041667
1980-12-31    2293.791667

Seasonality
 Date
1980-01-31    0.649843
1980-02-29    0.659214
1980-03-31    0.757440
1980-04-30    0.730351
1980-05-31    0.660609
1980-06-30    0.603468
1980-07-31    0.809164
```

```
1980-08-31    0.918822
1980-09-30    0.894367
1980-10-31    1.241789
1980-11-30    1.690158
1980-12-31    2.384776

Residual
 Date
1980-01-31         NaN
1980-02-29         NaN
1980-03-31         NaN
1980-04-30         NaN
1980-05-31         NaN
1980-06-30         NaN
1980-07-31    1.029230
1980-08-31    1.135407
1980-09-30    0.955954
1980-10-31    0.907513
1980-11-30    1.050423
1980-12-31    0.946770
```

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

Shape

```
(132, 1)
(55, 1)
```

Train and test data

```
First few rows of Training Data
```

	Sparkling
Date	
1980-01-31	1686
1980-02-29	1591
1980-03-31	2304

Date	Sparkling
1980-04-30	1712
1980-05-31	1471

Last few rows of Training Data

Date	Sparkling
1990-08-31	1605
1990-09-30	2424
1990-10-31	3116
1990-11-30	4286
1990-12-31	6047

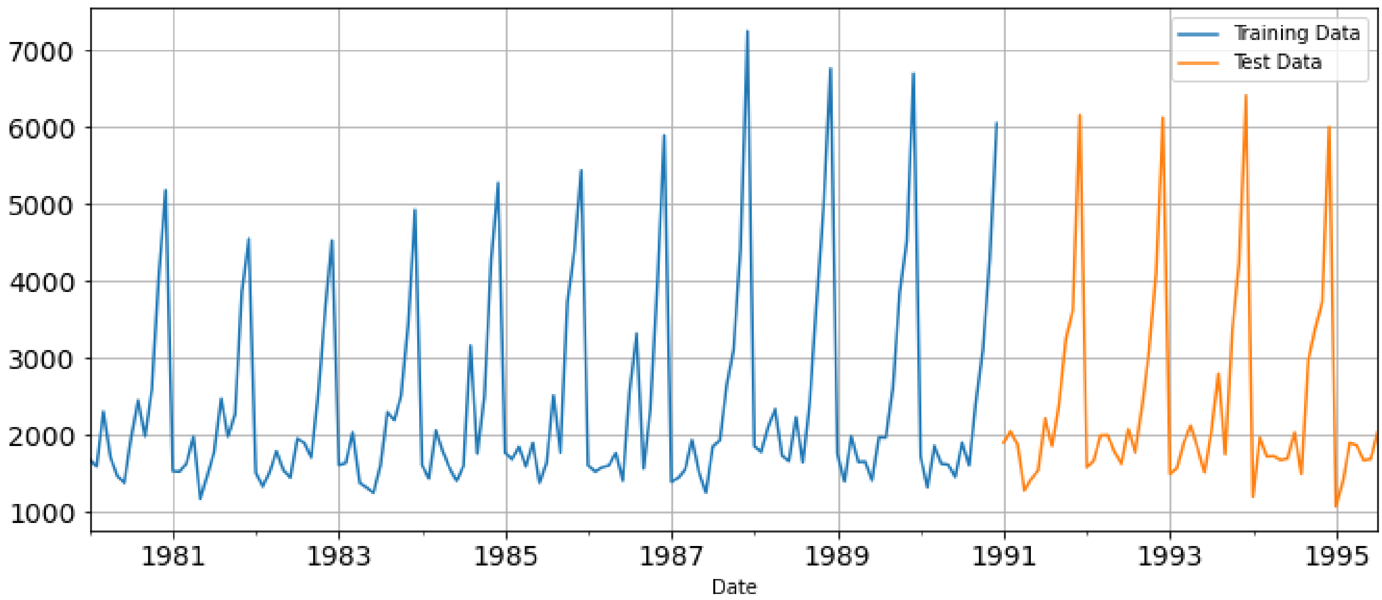
First few rows of Test Data

Date	Sparkling
1991-01-31	1902
1991-02-28	2049
1991-03-31	1874
1991-04-30	1279
1991-05-31	1432

```
Last few rows of Test Data
```

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

Plotting Training - Testing Data



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

Model 1: Linear Regression

Dividing

```
Training Time instance
 [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23
, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 4
4, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64,
65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85,
86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 10
5, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121,
122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132]
Test Time instance
 [133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 1
49, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 1
```

Train and Test

```
First few rows of Training Data
```

	Sparkling	time
Date		
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
Date		
1990-08-31	1605	128
1990-09-30	2424	129
1990-10-31	3116	130

Date	Sparkling	time
1990-11-30	4286	131
1990-12-31	6047	132

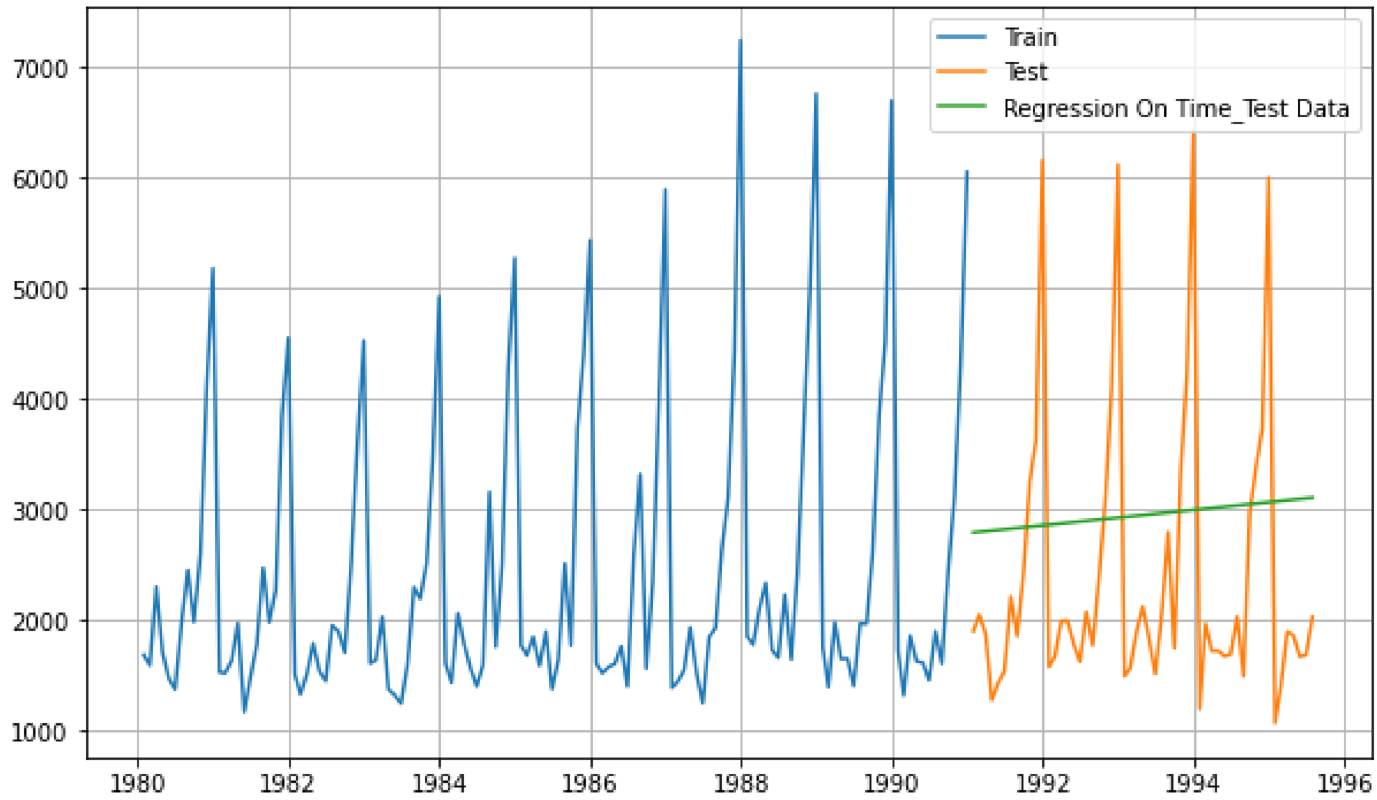
```
First few rows of Test Data
```

Date	Sparkling	time
1991-01-31	1902	133
1991-02-28	2049	134
1991-03-31	1874	135
1991-04-30	1279	136
1991-05-31	1432	137

```
Last few rows of Test Data
```

Date	Sparkling	time
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

Regression On Time Test Data



Model Evaluation

```
For RegressionOnTime forecast on the Test Data,RMSE is 1389.135
```

Table

	Test RMSE
RegressionOnTime	1389.135175

Model 2: Naive Approach

Train

	Sparkling
Date	
1980-01-31	1686

	Sparkling
Date	
1980-02-29	1591
1980-03-31	2304
1980-04-30	1712
1980-05-31	1471

Test

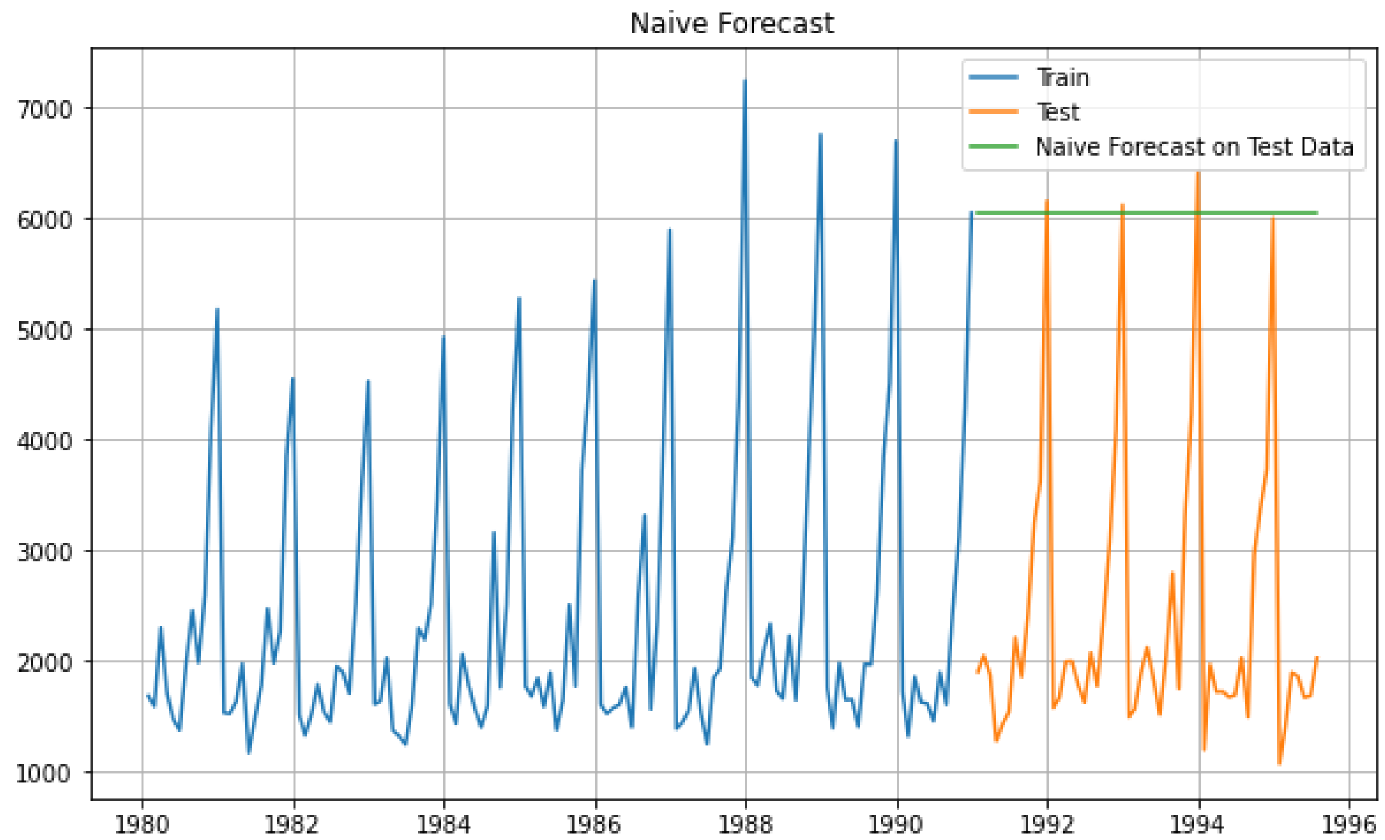
	Sparkling
Date	
1991-01-31	1902
1991-02-28	2049
1991-03-31	1874
1991-04-30	1279
1991-05-31	1432

NaiveModel_test

```
Date
1991-01-31    6047
1991-02-28    6047
1991-03-31    6047
1991-04-30    6047
```

Naive Forecast on Test Data

For this particular naive model, we say that the prediction for tomorrow is the same as today and the prediction for day after tomorrow is tomorrow and since the predicti on of tomorrow is same as today, therefore the prediction for day after tomorrow is also today.



Model Evaluation

```
For RegressionOnTime forecast on the Test Data,  RMSE is 3864.279
```

Table

	Test RMSE
RegressionOnTime	1389.135175
NaiveModel	3864.279352

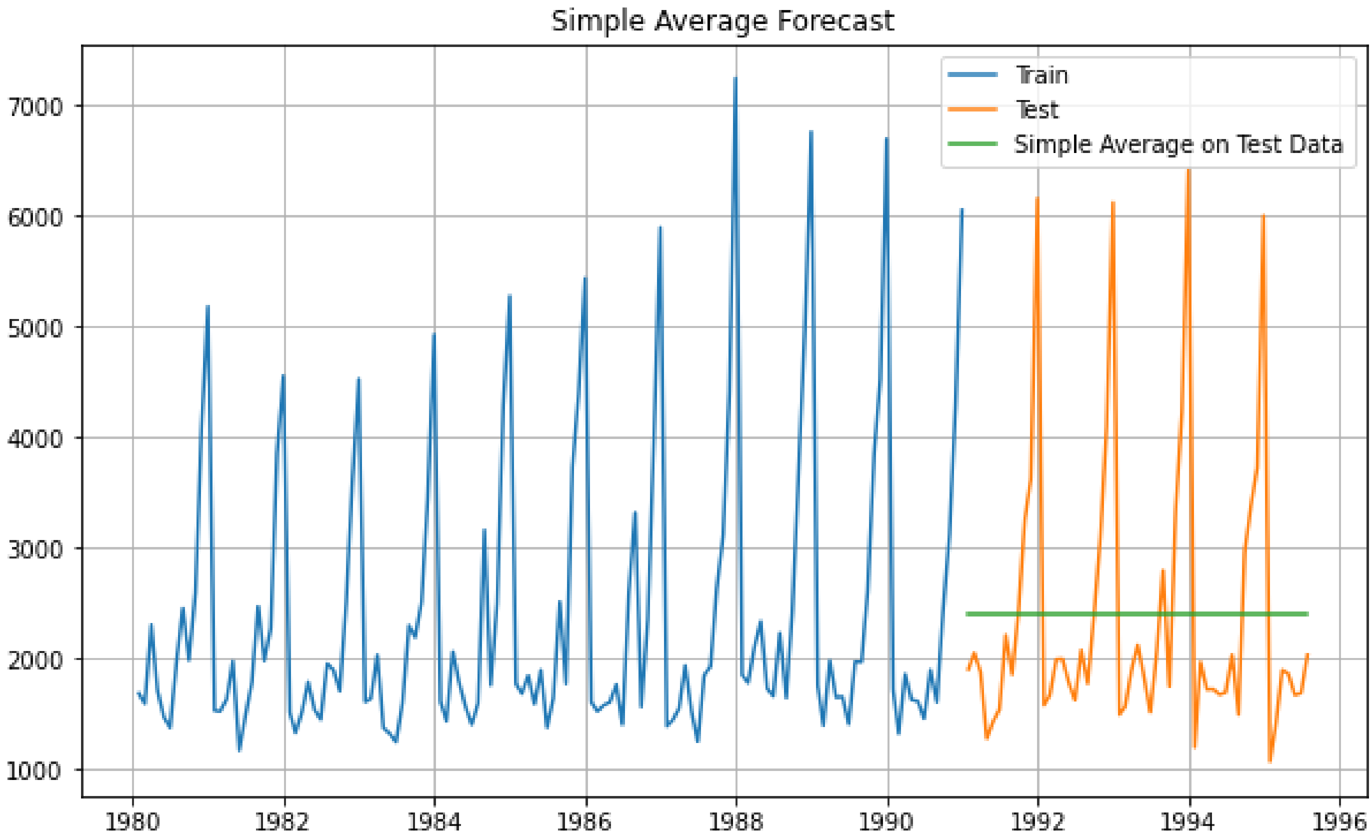
Method 3: Simple Average

Mean Forecast

	Sparkling	mean_forecast
Date		
1991-01-31	1902	2403.780303
1991-02-28	2049	2403.780303

	Sparkling	mean_forecast
Date		
1991-03-31	1874	2403.780303
1991-04-30	1279	2403.780303
1991-05-31	1432	2403.780303

Simple Average on Test Data



Model Evaluation

```
For Simple Average forecast on the Test Data,  RMSE is 1275.082
```

Table

	Test RMSE
RegressionOnTime	1389.135175
NaiveModel	3864.279352

	Test RMSE
SimpleAverageModel	1275.081804

Method 4: Moving Average(MA)

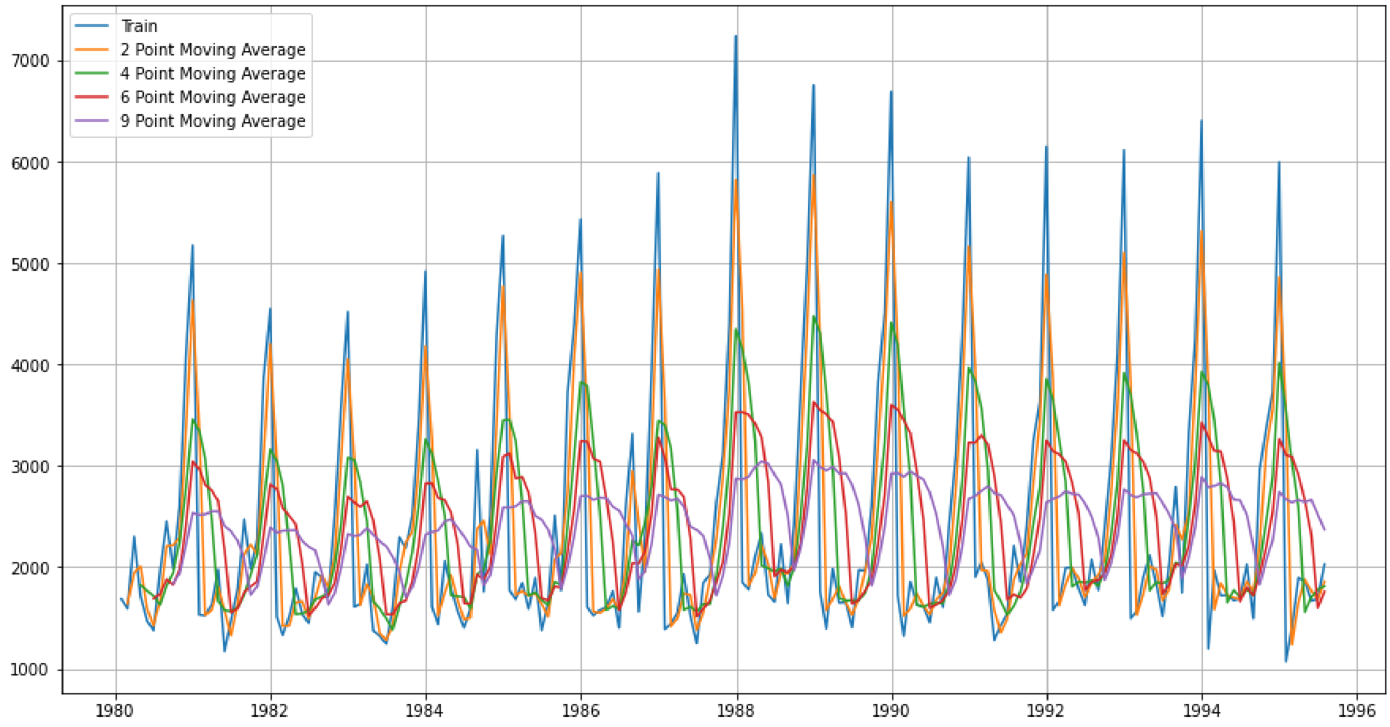
Table

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

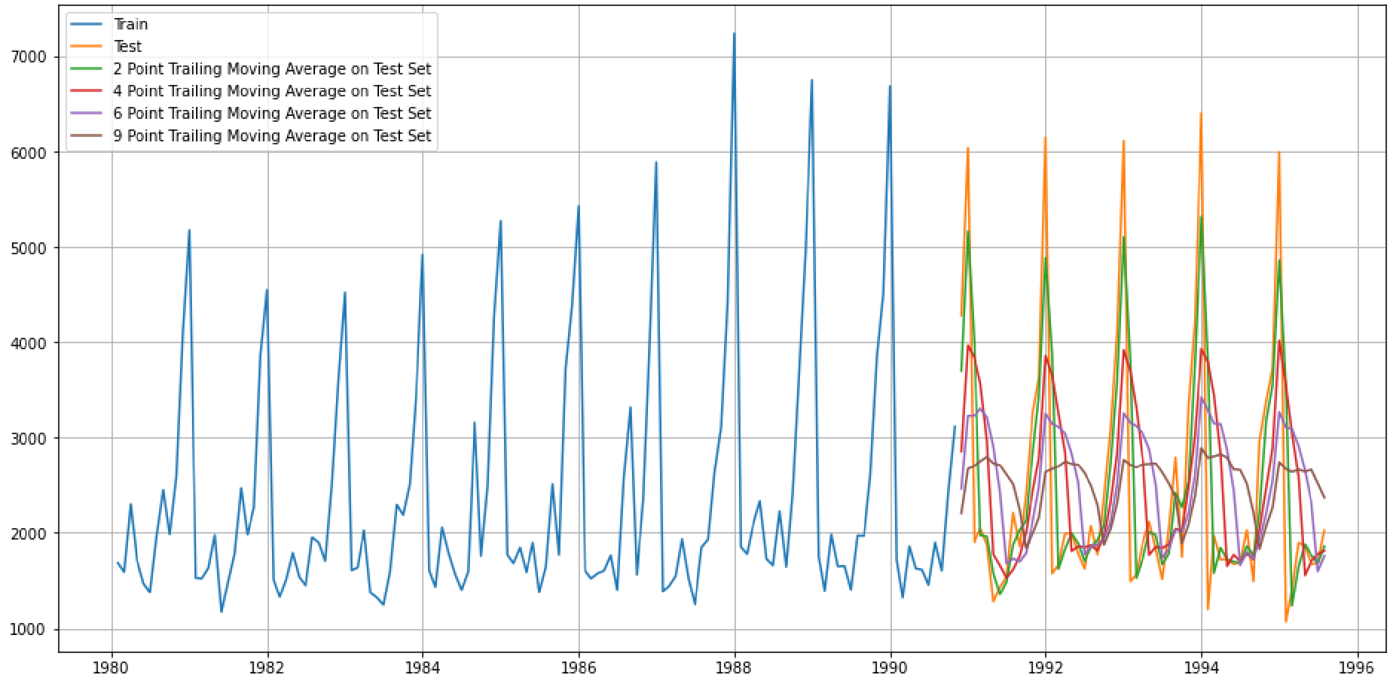
Trailing moving averages

	Sparkling	Trailing_2	Trailing_4	Trailing_6	Trailing_9
Date					
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

Plotting on the whole data



Plotting on both the Training and Test data



Trailing Moving Average

	Sparkling	Trailing_2	Trailing_4	Trailing_6	Trailing_9
Date					
1990-11-30	4286	3701.0	2857.75	2464.500000	2209.888889

	Sparkling	Trailing_2	Trailing_4	Trailing_6	Trailing_9
Date					
1990-12-31	6047	5166.5	3968.25	3229.500000	2675.222222
1991-01-31	1902	3974.5	3837.75	3230.000000	2705.666667
1991-02-28	2049	1975.5	3571.00	3304.000000	2753.888889
1991-03-31	1874	1961.5	2968.00	3212.333333	2800.222222

Shape

```
(57, 5)
```

```
Table
```

	Sparkling
Date	
1991-01-31	1902
1991-02-28	2049
1991-03-31	1874
1991-04-30	1279
1991-05-31	1432

Test shape

```
(55, 1)
```

Model Evaluation

```
For 2 point Moving Average Model forecast on the Training Data,  RMSE is
813.401
For 4 point Moving Average Model forecast on the Training Data,  RMSE is
1156.590
For 6 point Moving Average Model forecast on the Training Data,  RMSE is
1283.927
For 9 point Moving Average Model forecast on the Training Data,  RMSE is
1346.278
```

Table

	Test RMSE
RegressionOnTime	1389.135175
NaiveModel	3864.279352
SimpleAverageModel	1275.081804
2pointTrailingMovingAverage	813.400684
4pointTrailingMovingAverage	1156.589694
6pointTrailingMovingAverage	1283.927428
9pointTrailingMovingAverage	1346.278315

SES - ETS(A, N, N) - Simple Exponential Smoothing with additive errors

Exponential Smoothing methods

Exponential smoothing methods consist of flattening time series data.

Exponential smoothing averages or exponentially weighted moving averages consist of forecast based on previous periods data with exponentially declining influence on the older observations.

Exponential smoothing methods consist of special case exponential moving with notation ETS (Error, Trend, Seasonality) where each can be none(N), additive (N), additive damped (Ad), Multiplicative (M) or multiplicative damped (Md). One or more parameters control how fast the weights decay.

These parameters have values between 0 and 1

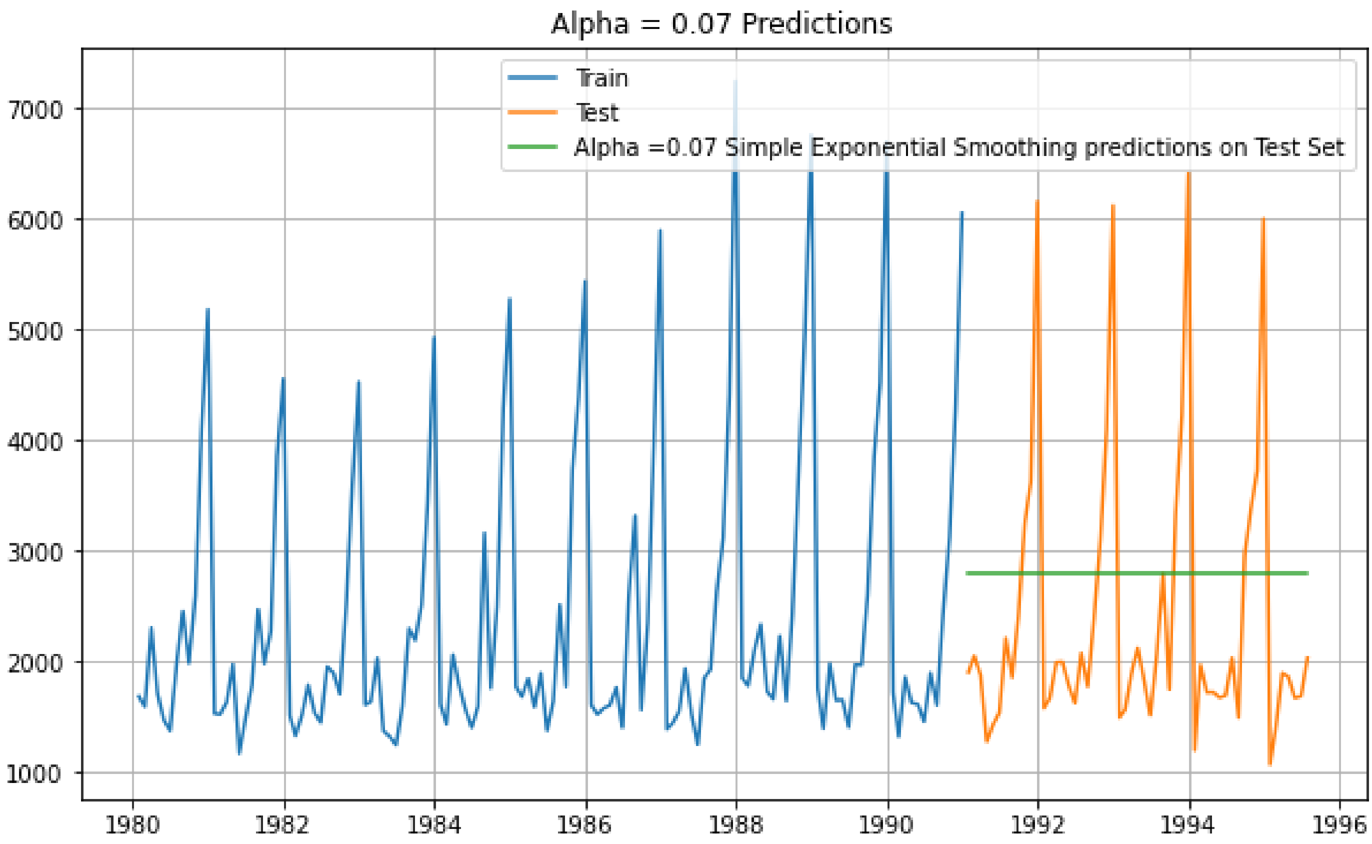
Parameters

```
{'smoothing_level': 0.07029120765764557,
 'smoothing_trend': nan,
 'smoothing_seasonal': nan,
 'damping_trend': nan,
 'initial_level': 1764.0137060346985,
 'initial_trend': nan,
 'initial_seasons': array([], dtype=float64),
 'use_boxcox': False,
 'lamda': None,
 'remove_bias': False}
```

Using the fitted model on the training set to forecast on the test set

```
1991-01-31    2804.675124
1991-02-28    2804.675124
1991-03-31    2804.675124
1991-04-30    2804.675124
1991-05-31    2804.675124
```

Simple Exponential Smoothing predictions



RMSE

```
SES RMSE: 1338.0083844916467
SES RMSE (calculated using statsmodels): 1338.0083844916464
```

Table

	Test RMSE
RegressionOnTime	1389.135175
NaiveModel	3864.279352
SimpleAverageModel	1275.081804
2pointTrailingMovingAverage	813.400684

	Test RMSE
4pointTrailingMovingAverage	1156.589694
6pointTrailingMovingAverage	1283.927428
9pointTrailingMovingAverage	1346.278315
Alpha=0.07,SimpleExponentialSmoothing	1338.008384

The simplest of the exponentially smoothing methods is naturally called simple exponential smoothing (SES). This method is suitable for forecasting data with no clear trend or seasonal pattern. In Single ES, the forecast at time (t + 1) is given by Winters,1960

$F_{t+1}=\alpha Y_t+(1-\alpha)F_t$ Parameter α is called the smoothing constant and its value lies between 0 and 1. Since the model uses only one smoothing constant, it is called Single Exponential Smoothing.

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

The fit() function is then called providing the fit configuration, the alpha value, smoothing_level. If this is omitted or set to None, the model will automatically optimize the value.

Holt - ETS(A, A, N) - Holt's linear method with additive errors

Double Exponential Smoothing

One of the drawbacks of the simple exponential smoothing is that the model does not do well in the presence of the trend. This model is an extension of SES known as Double Exponential model which estimates two smoothing parameters.

Applicable when data has Trend but no seasonality. Two separate components are considered: Level and Trend. Level is the local mean.

One smoothing parameter α corresponds to the level series A second smoothing parameter β corresponds to the trend series. Double Exponential Smoothing uses two equations to forecast future values of the time series, one for forecating the short term avarage value or level and the other for capturing the trend.

Intercept or Level equation, Lt is given by: Lt=αYt+(1−α)Ft Trend equation is given by Tt=β(Lt−Lt−1)+(1−β)Tt−1 Here, α and β are the smoothing constants for level and trend, respectively,

0 <α < 1 and 0 < β < 1. The forecast at time t + 1 is given by

Ft+1=Lt+Tt Ft+n=Lt+nTt

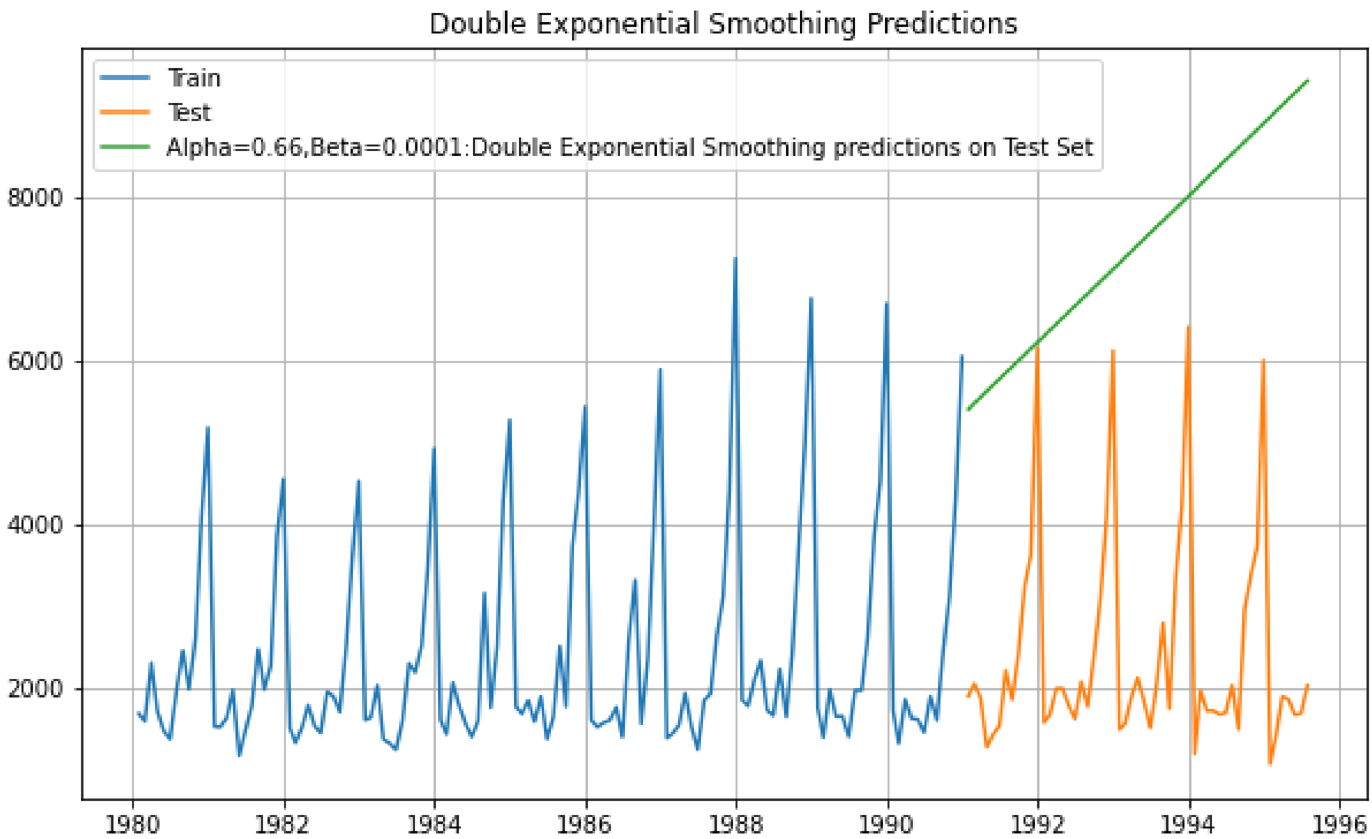
Parameters

```
==Holt model Exponential Smoothing Estimated Parameters ==

{'smoothing_level': 0.6649999999999999, 'smoothing_trend': 0.0
001, 'smoothing_seasonal': nan, 'damping_trend': nan, 'initial
_level': 1502.1999999999991, 'initial_trend': 74.8727272727273
9, 'initial_seasons': array([], dtype=float64), 'use_boxcox':
False, 'lamda': None, 'remove_bias': False}
```

Forecasting using this model for the duration of the test set

```
1991-01-31    5401.733026
1991-02-28    5476.005230
1991-03-31    5550.277433
1991-04-30    5624.549637
1991-05-31    5698.821840
```

Plotting the Training data, Test data and the forecasted value s



```
RMSE
DES RMSE: 5291.8798332269125
```

```
Table
```

	Test RMSE
RegressionOnTime	1389.135175
NaiveModel	3864.279352
SimpleAverageModel	1275.081804
2pointTrailingMovingAverage	813.400684
4pointTrailingMovingAverage	1156.589694
6pointTrailingMovingAverage	1283.927428
9pointTrailingMovingAverage	1346.278315
Alpha=0.07,SimpleExponentialSmoothing	1338.008384
Alpha=0.66,Beta=0.0001:DoubleExponentialSmoothing	5291.879833

Inference

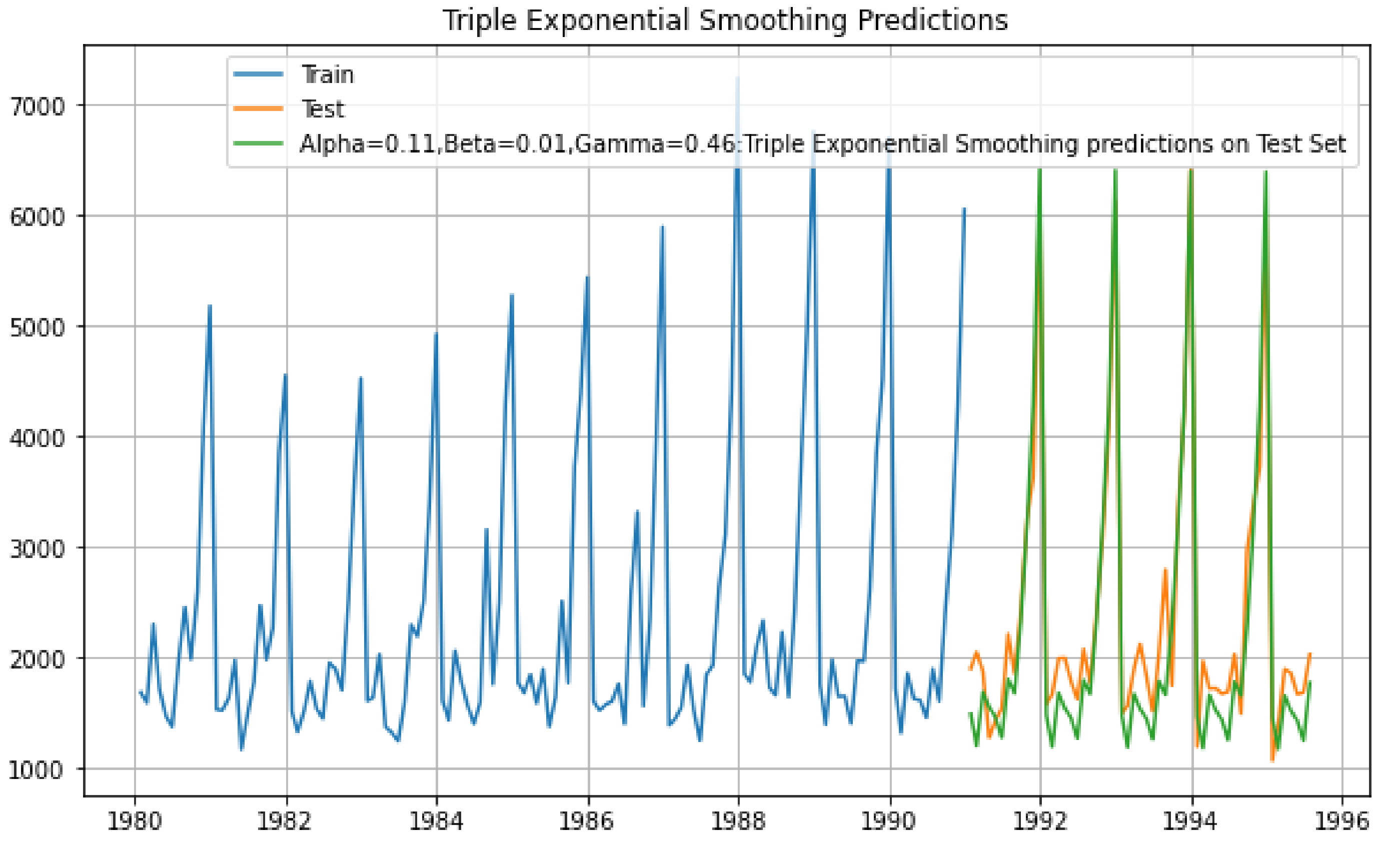
Here, we see that the Double Exponential Smoothing has actually done well when compared to the Simple Exponential Smoothing. This is because of the fact that the Double Exponential Smoothing model has picked up the trend component as well.

The Holt's model in Python has certain other options of exponential trends or whether the smoothing parameters should be damped. You can try these out later to check whether you get a better forecast.

Holt-Winters - ETS(A, A, A) - Holt Winter's linear method with additive errors

```
==Holt Winters model Exponential Smoothing Estimated Parameters ==

{'smoothing_level': 0.11127227248079453, 'smoothing_trend': 0.0123608043
05088534, 'smoothing_seasonal': 0.46071766688111543, 'damping_trend': na
n, 'initial_level': 2356.577980956387, 'initial_trend': -0.1024367553302
```

```
1725, 'initial_seasons': array([-636.23319334, -722.9832009 , -398.64410
813, -473.43045416,
       -808.42473284, -815.34991402, -384.23065038,   72.99484403,
       -237.44226045,  272.32608272, 1541.37737052, 2590.07692296]), 'us
e_boxcox': False, 'lamda': None, 'remove_bias': False}

Forecasting using this model for the duration of the test set
1991-01-31    1490.402890
1991-02-28    1204.525152
1991-03-31    1688.734182
1991-04-30    1551.226125
1991-05-31    1461.197883

Triple Exponential Smoothing Predictions
```




```
RMSE
TES RMSE: 378.95102286703

Table
```

	Test RMSE
RegressionOnTime	1389.135175
NaiveModel	3864.279352
SimpleAverageModel	1275.081804

	Test RMSE
2pointTrailingMovingAverage	813.400684
4pointTrailingMovingAverage	1156.589694
6pointTrailingMovingAverage	1283.927428
9pointTrailingMovingAverage	1346.278315
Alpha=0.07,SimpleExponentialSmoothing	1338.008384
Alpha=0.66,Beta=0.0001:DoubleExponentialSmoothing	5291.879833
Alpha=0.11,Beta=0.001,Gamma=0.46:Triple Exponential Smoothing	378.951023

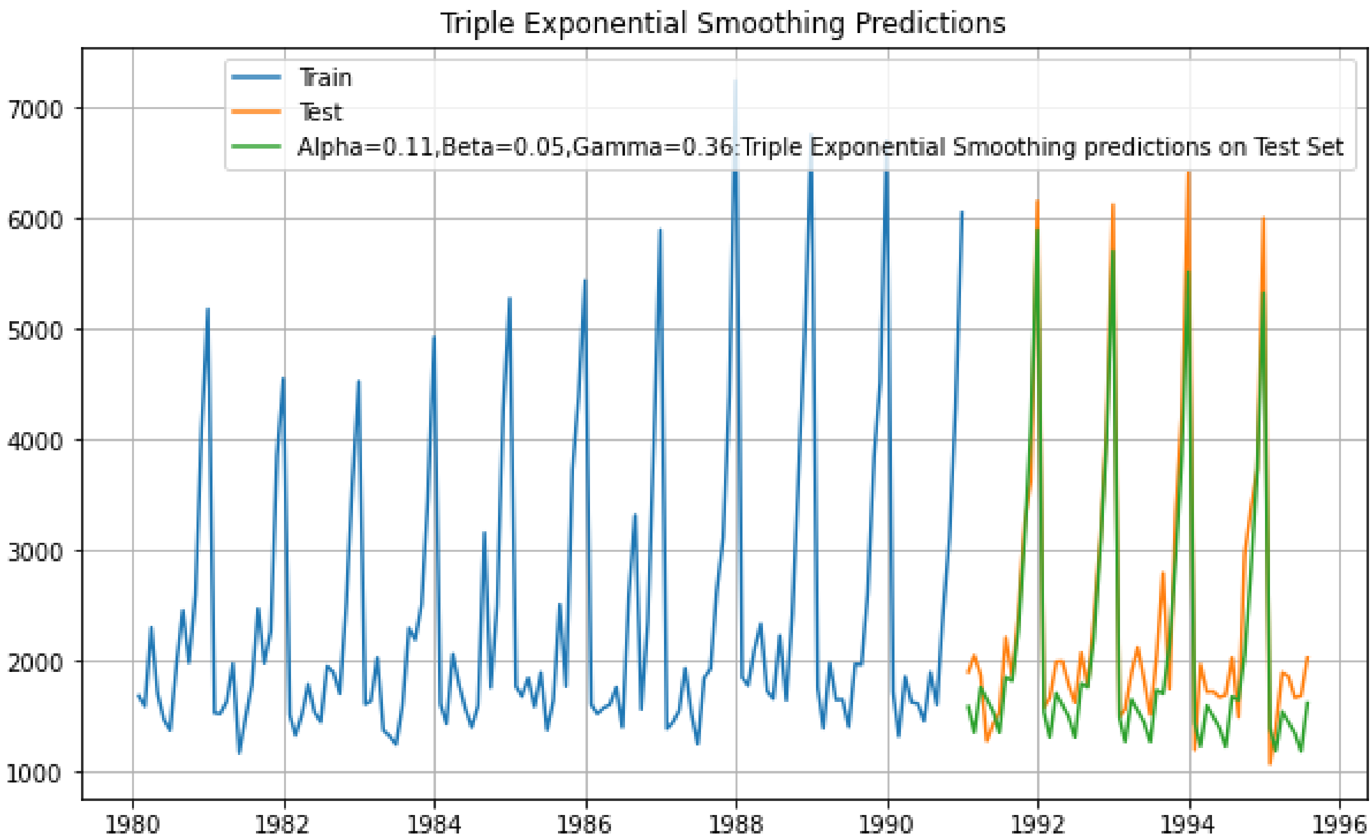
Holt-Winters - ETS(A, A, M) - Holt Winter's linear method

```
Parameters

==Holt Winters model Exponential Smoothing Estimated Parameters ==

{'smoothing_level': 0.11133818361298699, 'smoothing_trend': 0.049505131019509915, '
smoothing_seasonal': 0.3620795793580111, 'damping_trend': nan, 'initial_level': 235
6.4967888704355, 'initial_trend': -10.187944726007238, 'initial_seasons': array([0.
71296382, 0.68242226, 0.90755008, 0.80515228, 0.65597218,
       0.65414505, 0.88617935, 1.13345121, 0.92046306, 1.21337874,
       1.87340336, 2.37811768]), 'use_boxcox': False, 'lamda': None, 'remove_bias':
False}
```

Forecasting using this model for the duration of the test set

```
1991-01-31    1587.497468
1991-02-28    1356.394925
1991-03-31    1762.929755
1991-04-30    1656.165933
1991-05-31    1542.002730
```

Triple Exponential Smoothing Predictions



Report model accuracy

RMSE

TES_am RMSE: 404.286809456071

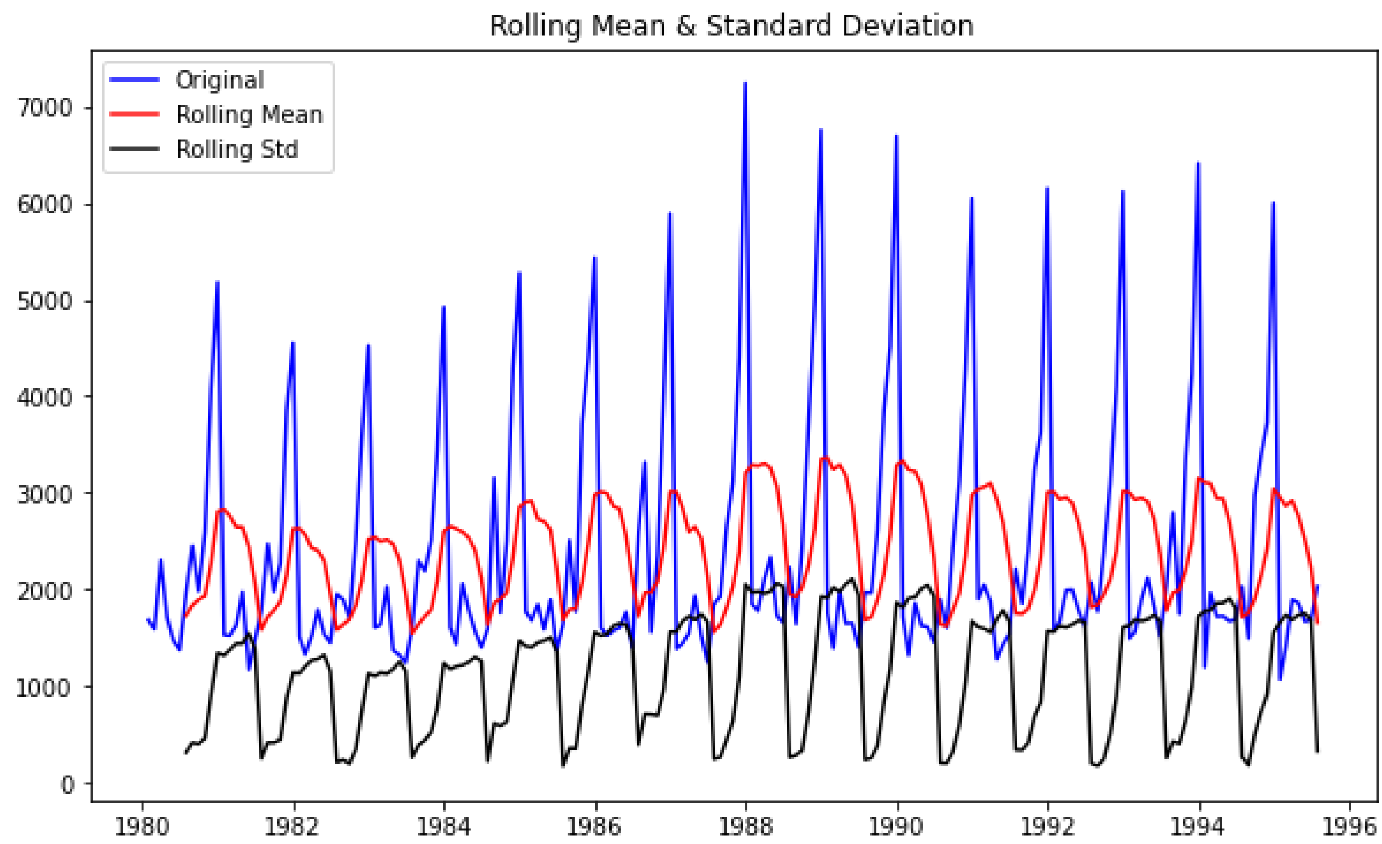
Table

	Test RMSE
RegressionOnTime	1389.135175
NaiveModel	3864.279352
SimpleAverageModel	1275.081804
2pointTrailingMovingAverage	813.400684
4pointTrailingMovingAverage	1156.589694
6pointTrailingMovingAverage	1283.927428
9pointTrailingMovingAverage	1346.278315
Alpha=0.07,SimpleExponentialSmoothing	1338.008384

	Test RMSE
Alpha=0.66,Beta=0.0001:DoubleExponentialSmoothing	5291.879833
Alpha=0.11,Beta=0.001,Gamma=0.46:Triple Exponential Smoothing	378.951023
Alpha=0.11,Beta=0.05,Gamma=0.36:Triple Exponential Smoothing 2	404.286809

Q.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.

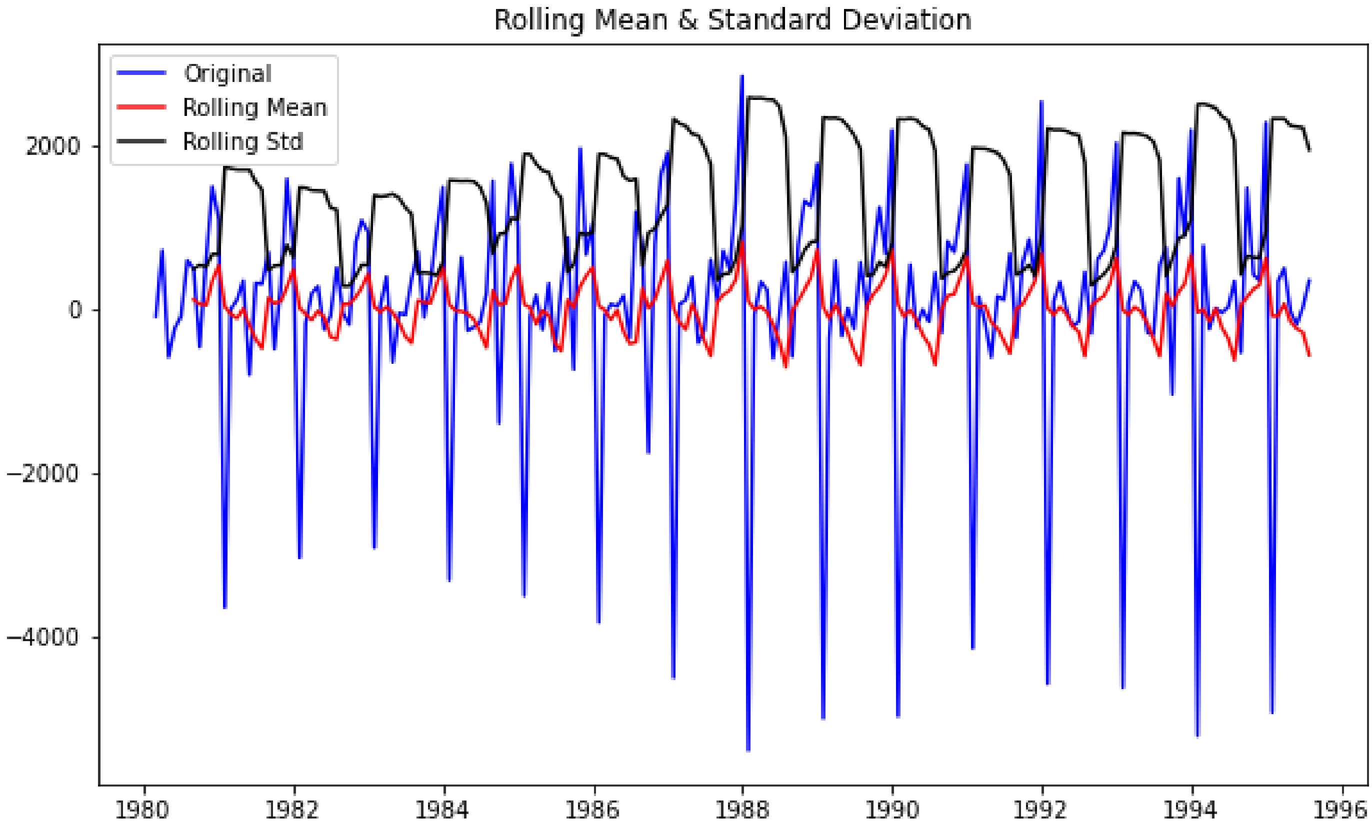
Note: Stationarity should be checked at alpha = 0.05.




```
Results of Dickey-Fuller Test:
Test Statistic                -1.360497
p-value                        0.601061
#Lags Used                    11.000000
```

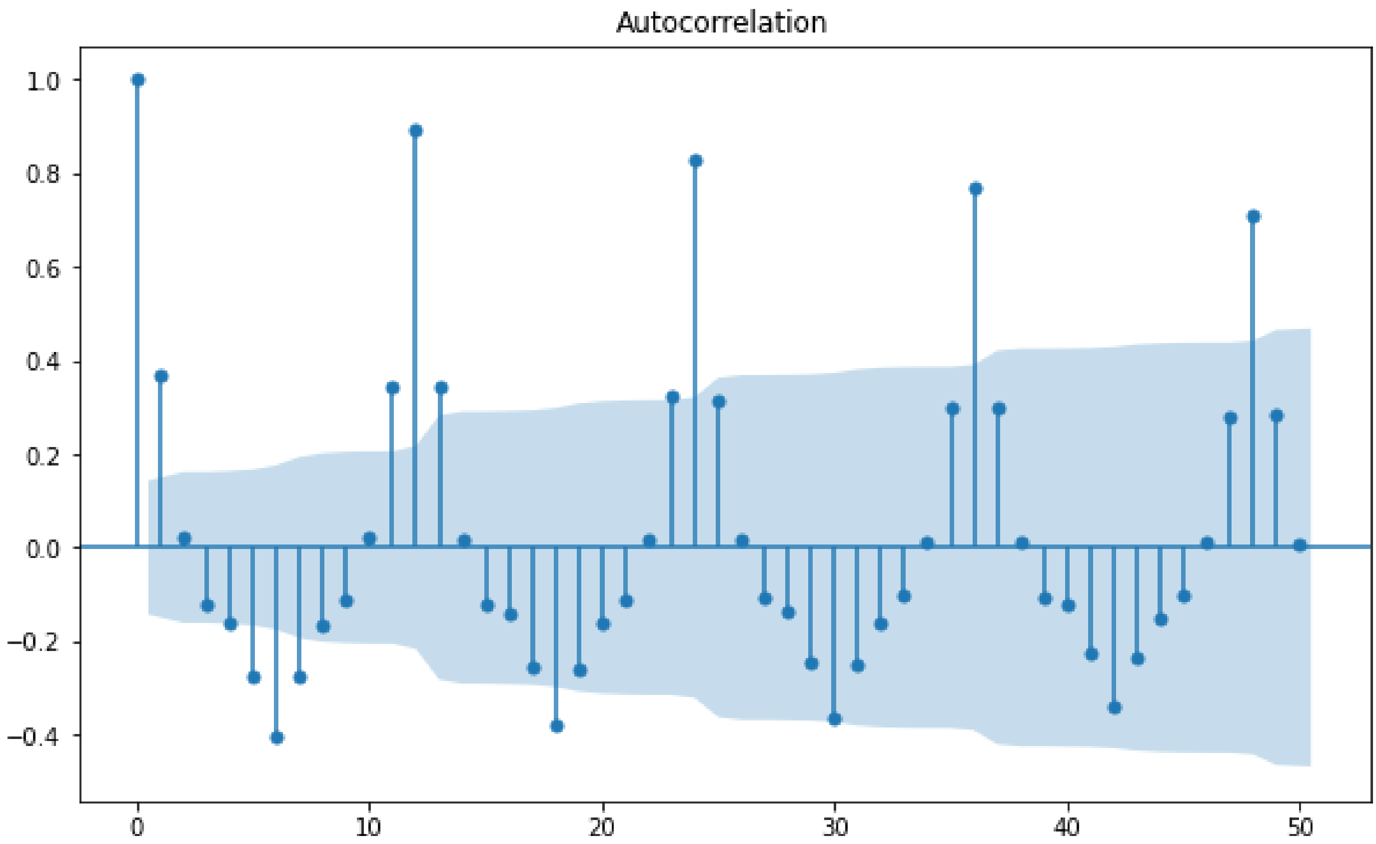
```
Number of Observations Used    175.000000
Critical Value (1%)             -3.468280
Critical Value (5%)             -2.878202
Critical Value (10%)            -2.575653
```

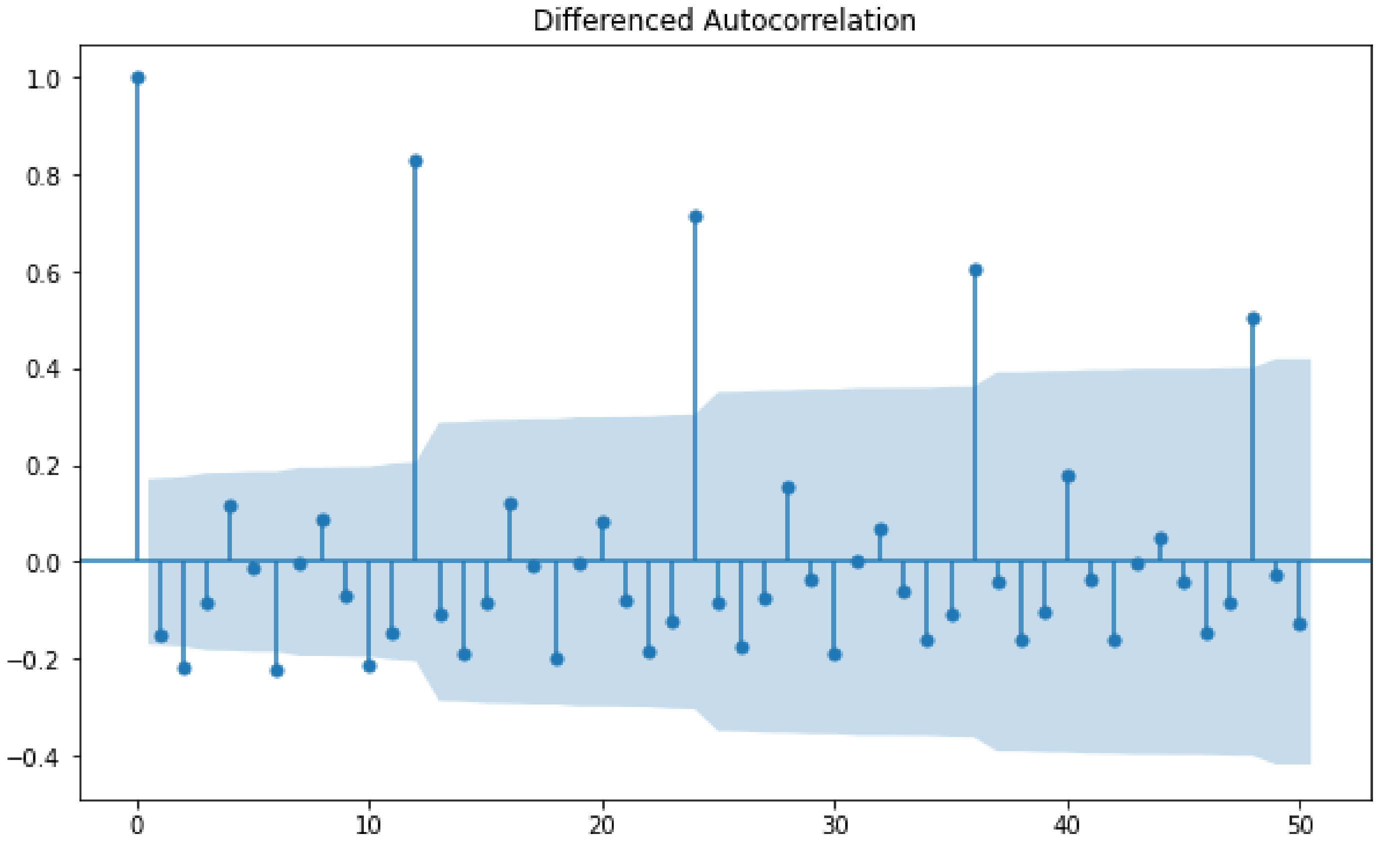
Plotting with the difference of 1



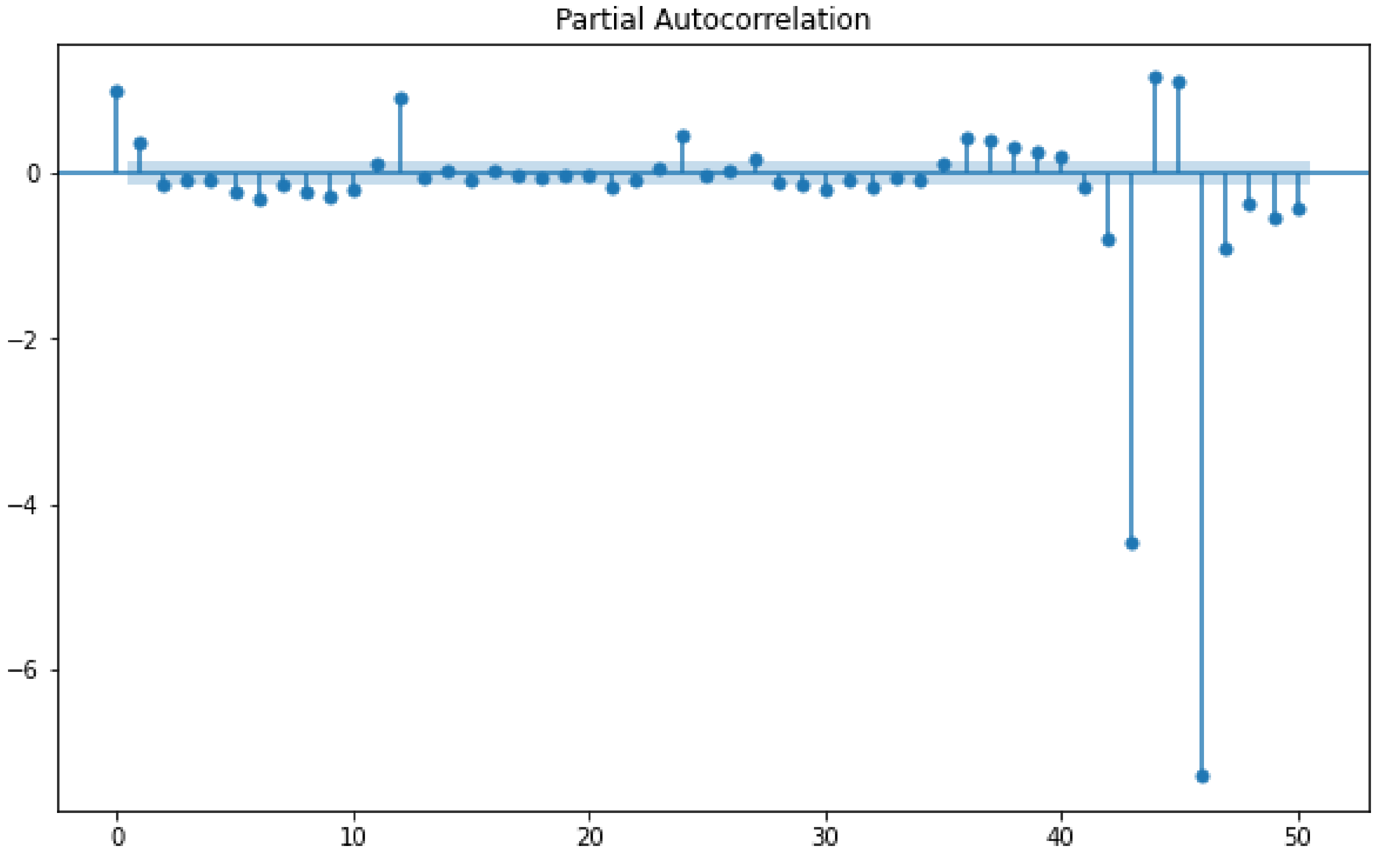

```
Results of Dickey-Fuller Test:
Test Statistic                 -45.050301
p-value                          0.000000
#Lags Used                      10.000000
Number of Observations Used    175.000000
Critical Value (1%)             -3.468280
Critical Value (5%)             -2.878202
Critical Value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

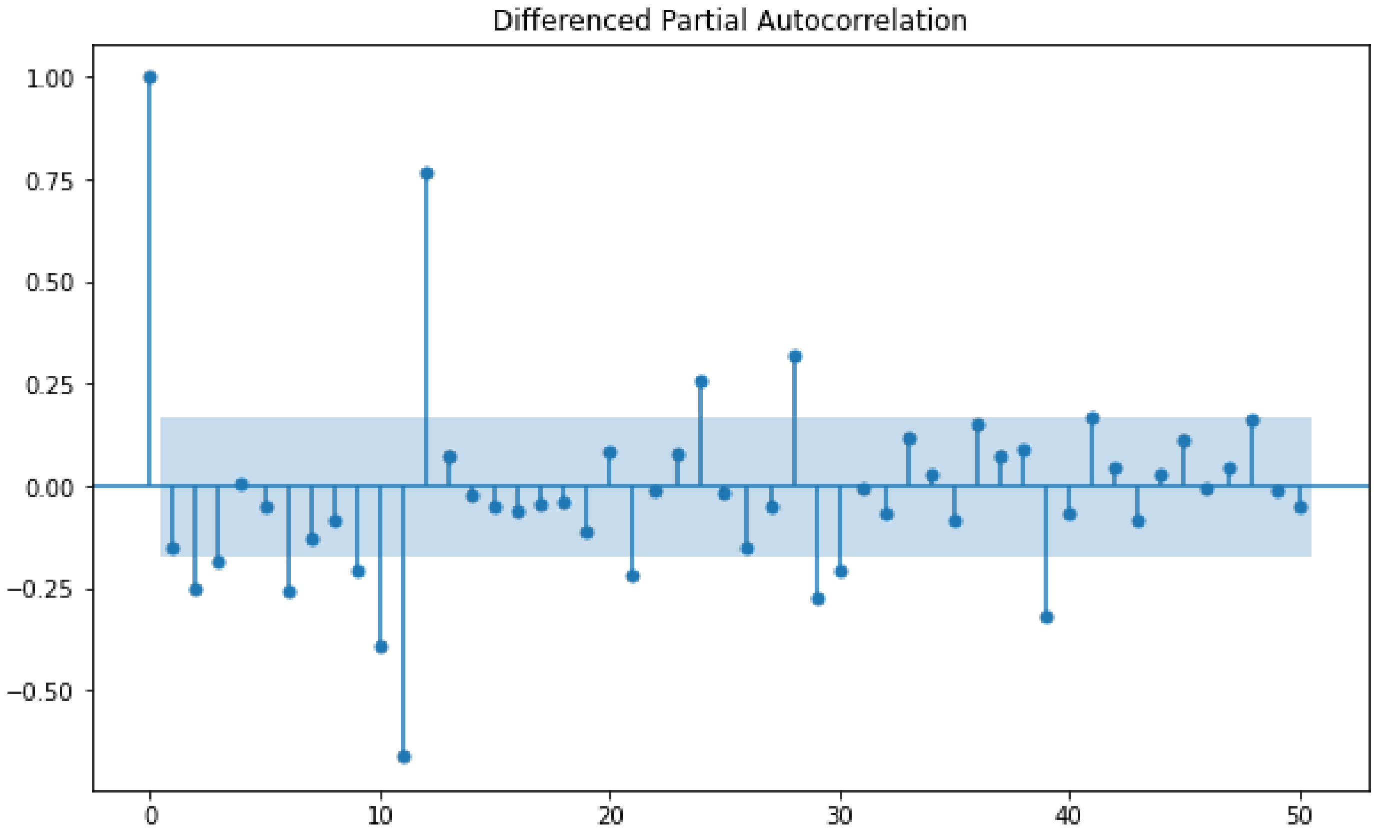
Plot the Autocorrelation and the Partial Autocorrelation function plots on the whole data.


Autocorrelation
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
0
10
20
30
40
50

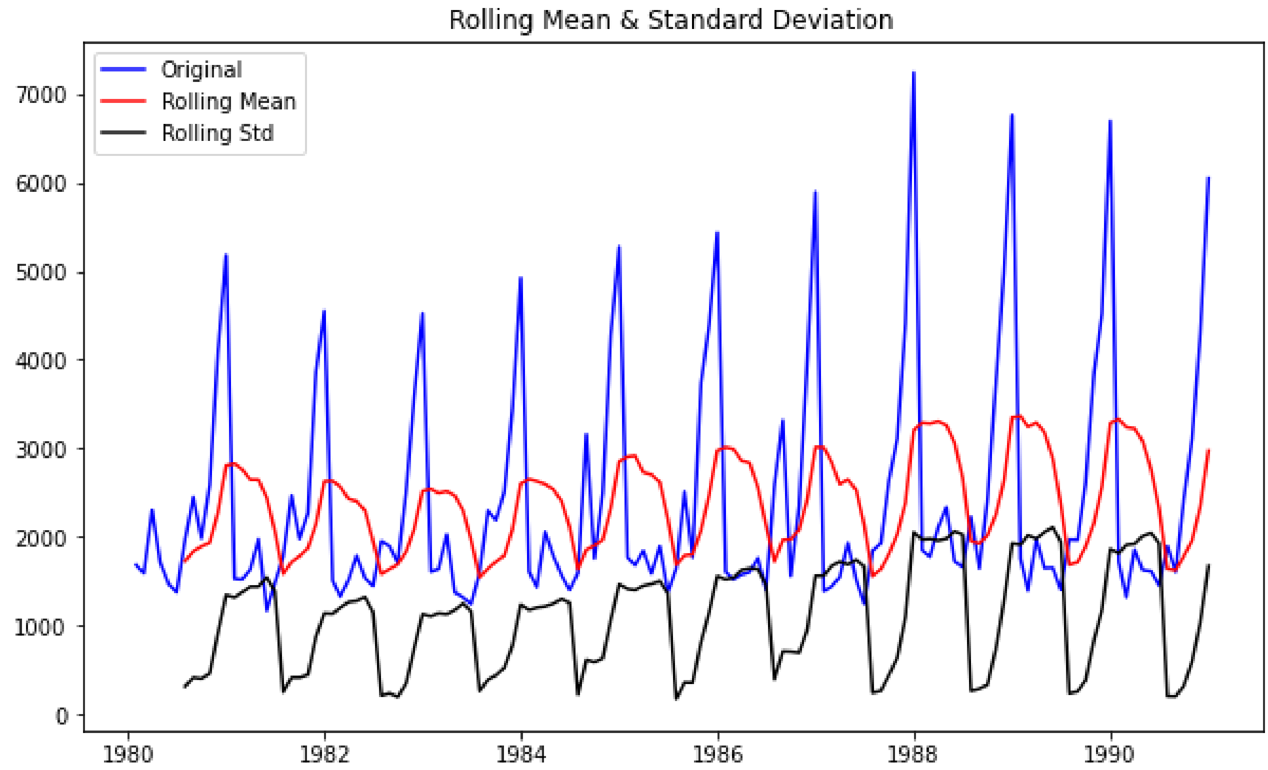

Differenced Autocorrelation
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
0
10
20
30
40
50

PACF PLOT

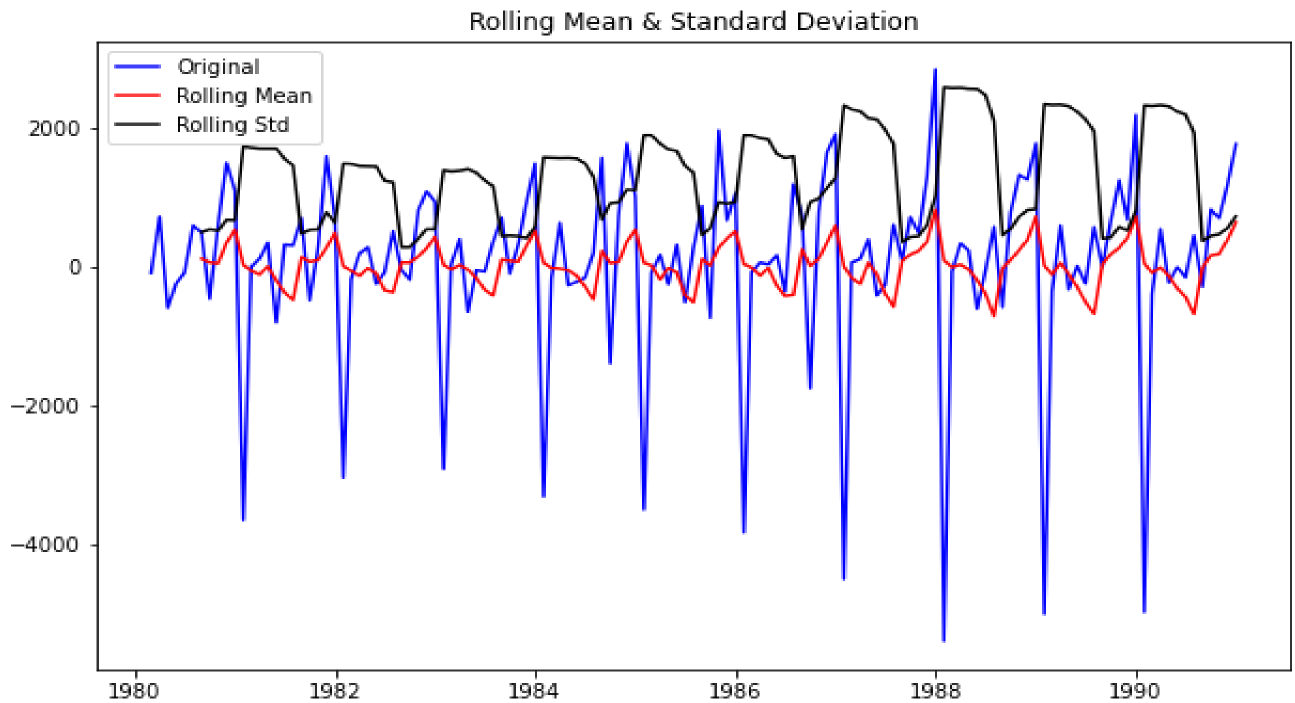

Partial Autocorrelation
0
−2
−4
−6
0
10
20
30
40
50


Differenced Partial Autocorrelation
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
0
10
20
30
40
50

Check for stationarity of the Training Data Time Series.




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



```
Results of Dickey-Fuller Test:
Test Statistic                -8.005007e+00
p-value                        2.280104e-12
#Lags Used                     1.100000e+01
Number of Observations Used    1.190000e+02
Critical Value (1%)           -3.486535e+00
Critical Value (5%)           -2.886151e+00
Critical Value (10%)          -2.579896e+00
```

Q.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.

```
Some parameter combinations for the Model...
Model: (0, 1, 1)
Model: (0, 1, 2)
Model: (0, 1, 3)
Model: (0, 1, 4)
Model: (1, 1, 0)
Model: (1, 1, 1)
Model: (1, 1, 2)
Model: (1, 1, 3)
Model: (1, 1, 4)
Model: (2, 1, 0)
Model: (2, 1, 1)
Model: (2, 1, 2)
Model: (2, 1, 3)
Model: (2, 1, 4)
Model: (3, 1, 0)
Model: (3, 1, 1)
Model: (3, 1, 2)
Model: (3, 1, 3)
Model: (3, 1, 4)
Model: (4, 1, 0)
Model: (4, 1, 1)
Model: (4, 1, 2)
Model: (4, 1, 3)
Model: (4, 1, 4)
```

AIC

```
ARIMA(0, 1, 0) - AIC:2267.6630357855465
ARIMA(0, 1, 1) - AIC:2263.060015591336
ARIMA(0, 1, 2) - AIC:2234.4083231283275
ARIMA(0, 1, 3) - AIC:2233.994857753515
ARIMA(0, 1, 4) - AIC:2235.173736469558
ARIMA(1, 1, 0) - AIC:2266.6085393190087
ARIMA(1, 1, 1) - AIC:2235.755094673383
ARIMA(1, 1, 2) - AIC:2234.527200452466
ARIMA(1, 1, 3) - AIC:2235.607816390617
ARIMA(1, 1, 4) - AIC:2227.736977676672
ARIMA(2, 1, 0) - AIC:2260.365743968086
ARIMA(2, 1, 1) - AIC:2233.777626239922
ARIMA(2, 1, 2) - AIC:2213.509212306332
```

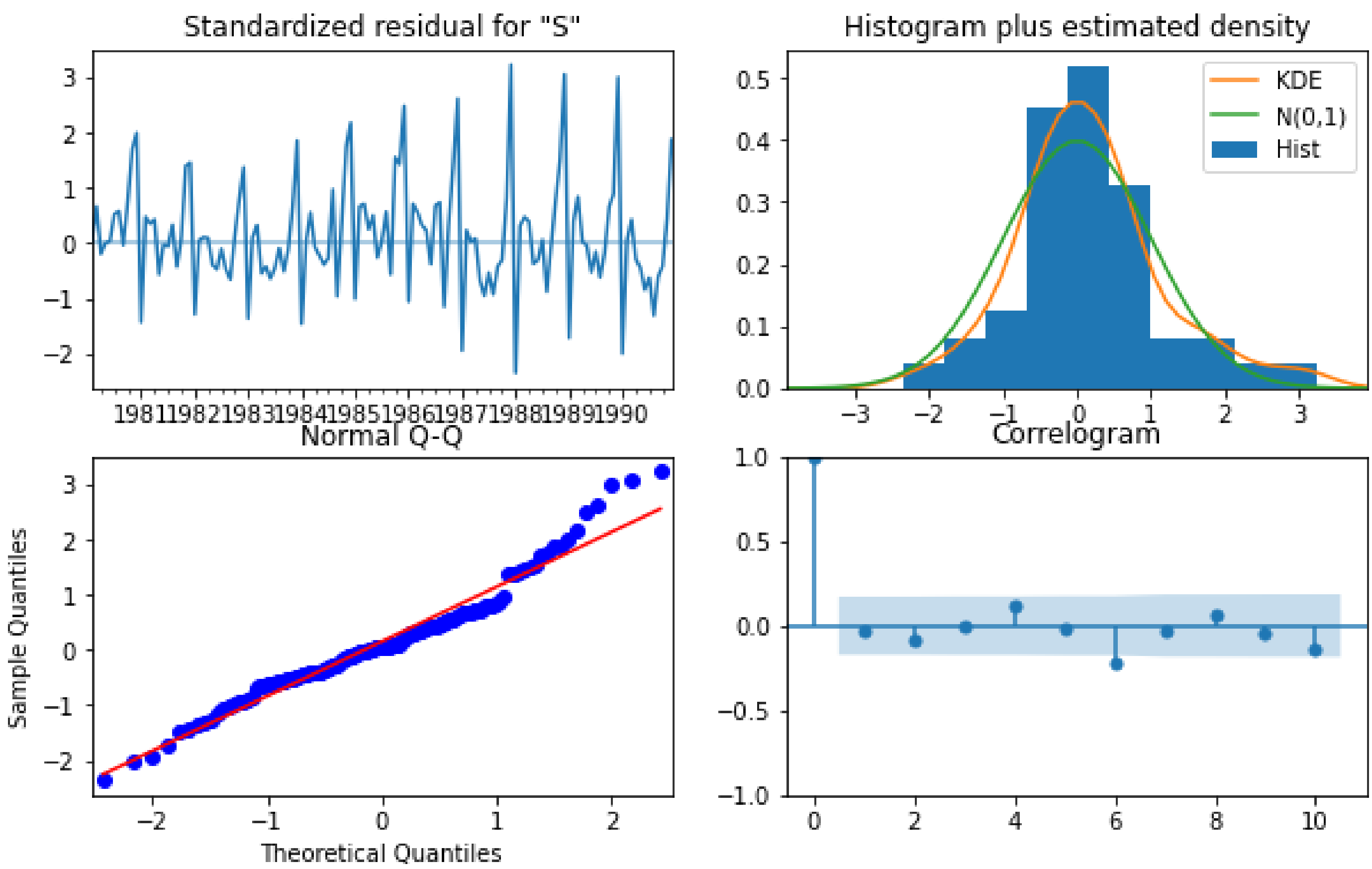
```
ARIMA(2, 1, 3) - AIC:2232.921136688177
ARIMA(2, 1, 4) - AIC:2222.921832384166
ARIMA(3, 1, 0) - AIC:2257.72337899794
ARIMA(3, 1, 1) - AIC:2235.498878057432
ARIMA(3, 1, 2) - AIC:2230.759636959836
ARIMA(3, 1, 3) - AIC:2221.4566102276085
ARIMA(3, 1, 4) - AIC:2219.8923646545354
ARIMA(4, 1, 0) - AIC:2259.741841399269
ARIMA(4, 1, 1) - AIC:2237.073047636303
ARIMA(4, 1, 2) - AIC:2233.049523102294
ARIMA(4, 1, 3) - AIC:2222.9040959025133
ARIMA(4, 1, 4) - AIC:2213.5641907793433
```

Optimum AIC

	param	AIC
12	(2, 1, 2)	2213.509212
24	(4, 1, 4)	2213.564191
19	(3, 1, 4)	2219.892365
18	(3, 1, 3)	2221.456610
23	(4, 1, 3)	2222.904096

```
                               SARIMAX Results                                
==============================================================================
Dep. Variable:              Sparkling   No. Observations:                  132
Model:                 ARIMA(2, 1, 2)   Log Likelihood               -1101.755
Date:                Sun, 25 Sep 2022   AIC                           2213.509
Time:                        12:34:57   BIC                           2227.885
Sample:                    01-31-1980   HQIC                          2219.351
                         - 12-31-1990                                         
Covariance Type:                  opg                                         
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          1.3121      0.046     28.781      0.000       1.223       1.401
ar.L2         -0.5593      0.072     -7.741      0.000      -0.701      -0.418
ma.L1         -1.9917      0.109    -18.217      0.000      -2.206      -1.777
ma.L2          0.9999      0.110      9.109      0.000       0.785       1.215
sigma2      1.099e+06   1.99e-07   5.51e+12      0.000     1.1e+06     1.1e+06
===================================================================================
Ljung-Box (L1) (Q):                   0.19   Jarque-Bera (JB):                14.46
Prob(Q):                              0.67   Prob(JB):                         0.00
Heteroskedasticity (H):               2.43   Skew:                             0.61
Prob(H) (two-sided):                  0.00   Kurtosis:                         4.08
===================================================================================
```

Results auto ARIMA



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

RMSE & MAPE

```
RMSE: 1299.9795689481477
MAPE: 47.099932436388684
```

Table

	Test RMSE	MAPE
RegressionOnTime	1389.135175	NaN
NaiveModel	3864.279352	NaN
SimpleAverageModel	1275.081804	NaN
2pointTrailingMovingAverage	813.400684	NaN
4pointTrailingMovingAverage	1156.589694	NaN
6pointTrailingMovingAverage	1283.927428	NaN

	Test RMSE	MAPE
9pointTrailingMovingAverage	1346.278315	NaN
Alpha=0.07,SimpleExponentialSmoothing	1338.008384	NaN
Alpha=0.66,Beta=0.0001:DoubleExponentialSmoothing	5291.879833	NaN
Alpha=0.11,Beta=0.001,Gamma=0.46:Triple Exponential Smoothing	378.951023	NaN
Alpha=0.11,Beta=0.05,Gamma=0.36:Triple Exponential Smoothing 2	404.286809	NaN
Arima 2,1,2	1299.979569	47.099932

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

Setting the seasonality as 6 for the first iteration of the auto SARIMA model.

```
Examples of some parameter combinations for Model...
Model: (0, 1, 1)(0, 0, 1, 6)
Model: (0, 1, 2)(0, 0, 2, 6)
Model: (1, 1, 0)(1, 0, 0, 6)
Model: (1, 1, 1)(1, 0, 1, 6)
Model: (1, 1, 2)(1, 0, 2, 6)
Model: (2, 1, 0)(2, 0, 0, 6)
Model: (2, 1, 1)(2, 0, 1, 6)
Model: (2, 1, 2)(2, 0, 2, 6)
```

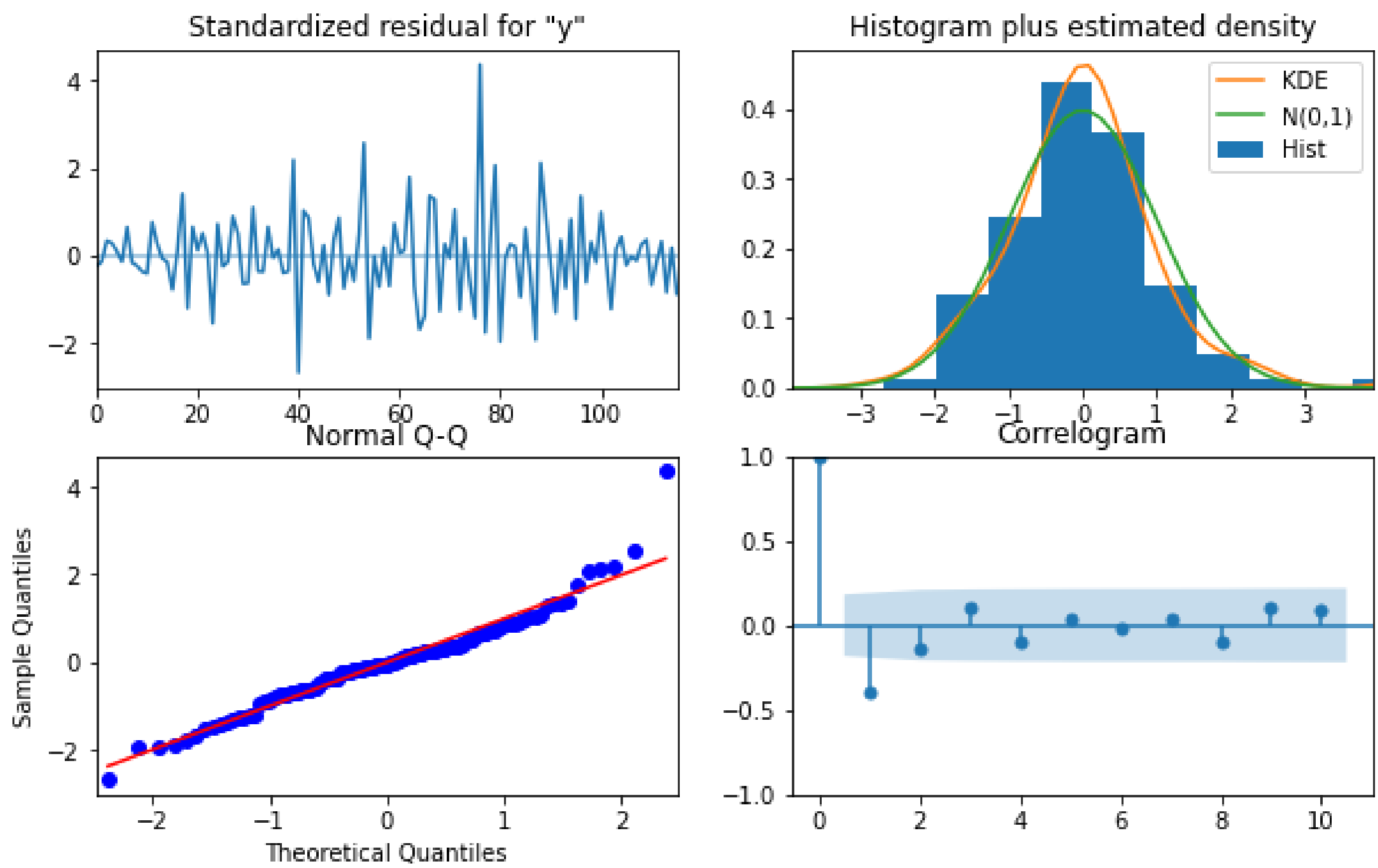
Optimum AIC

	param	seasonal	AIC
0	(2, 1, 2)	(2, 0, 2, 6)	1782.342657

```
                                     SARIMAX Results
==========================================================================================
======
Dep. Variable:                                  y   No. Observations:                  132
Model:             SARIMAX(2, 1, 2)x(2, 0, 2, 6)   Log Likelihood                -882.171
Date:                            Sun, 25 Sep 2022   AIC                           1782.343
Time:                                    12:36:14   BIC                           1807.125
Sample:                                         0   HQIC                          1792.403
                                            - 132
```

```
Covariance Type:                                    opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -1.8430      0.035    -52.843      0.000      -1.911      -1.775
ar.L2         -0.8754      0.039    -22.702      0.000      -0.951      -0.800
ma.L1          1.9729      0.106     18.641      0.000       1.765       2.180
ma.L2          0.9990      0.106      9.466      0.000       0.792       1.206
ar.S.L6       -0.0014      0.031     -0.045      0.964      -0.061       0.059
ar.S.L12       1.0337      0.025     41.709      0.000       0.985       1.082
ma.S.L6        0.0351      0.142      0.247      0.805      -0.243       0.313
ma.S.L12      -0.4658      0.080     -5.823      0.000      -0.623      -0.309
sigma2      2.172e+05   9.74e-07   2.23e+11      0.000    2.17e+05    2.17e+05
===================================================================================
Ljung-Box (L1) (Q):                  18.76   Jarque-Bera (JB):                43.54
Prob(Q):                              0.00   Prob(JB):                         0.00
Heteroskedasticity (H):               2.55   Skew:                             0.67
Prob(H) (two-sided):                  0.00   Kurtosis:                         5.69
===================================================================================
```

Results auto SARIMA 6



summary_frame

y	mean	mean_se	mean_ci_lower	mean_ci_upper
0	778.004948	469.563739	-142.323069	1698.332966
1	441.314773	703.296508	-937.121053	1819.750599

y	mean	mean_se	mean_ci_lower	mean_ci_upper
2	968.981356	848.689857	-694.420197	2632.382909
3	796.130852	993.887443	-1151.852741	2744.114445
4	667.442810	1104.575506	-1497.485401	2832.371020

RMSE & MAPE

```
RMSE: 2117.224709406799
MAPE: 100.28095751809441
```

Table

	Test RMSE	MAPE
RegressionOnTime	1389.135175	NaN
NaiveModel	3864.279352	NaN
SimpleAverageModel	1275.081804	NaN
2pointTrailingMovingAverage	813.400684	NaN
4pointTrailingMovingAverage	1156.589694	NaN
6pointTrailingMovingAverage	1283.927428	NaN
9pointTrailingMovingAverage	1346.278315	NaN
Alpha=0.07,SimpleExponentialSmoothing	1338.008384	NaN
Alpha=0.66,Beta=0.0001:DoubleExponentialSmoothing	5291.879833	NaN
Alpha=0.11,Beta=0.001,Gamma=0.46:Triple Exponential Smoothing	378.951023	NaN
Alpha=0.11,Beta=0.05,Gamma=0.36:Triple Exponential Smoothing 2	404.286809	NaN
Arima 2,1,2	1299.979569	47.099932
Sarima (1,1,2)(2,0,2,6)	2117.224709	100.280958

Setting the seasonality as 12 for the second iteration of the auto SARIMA model.

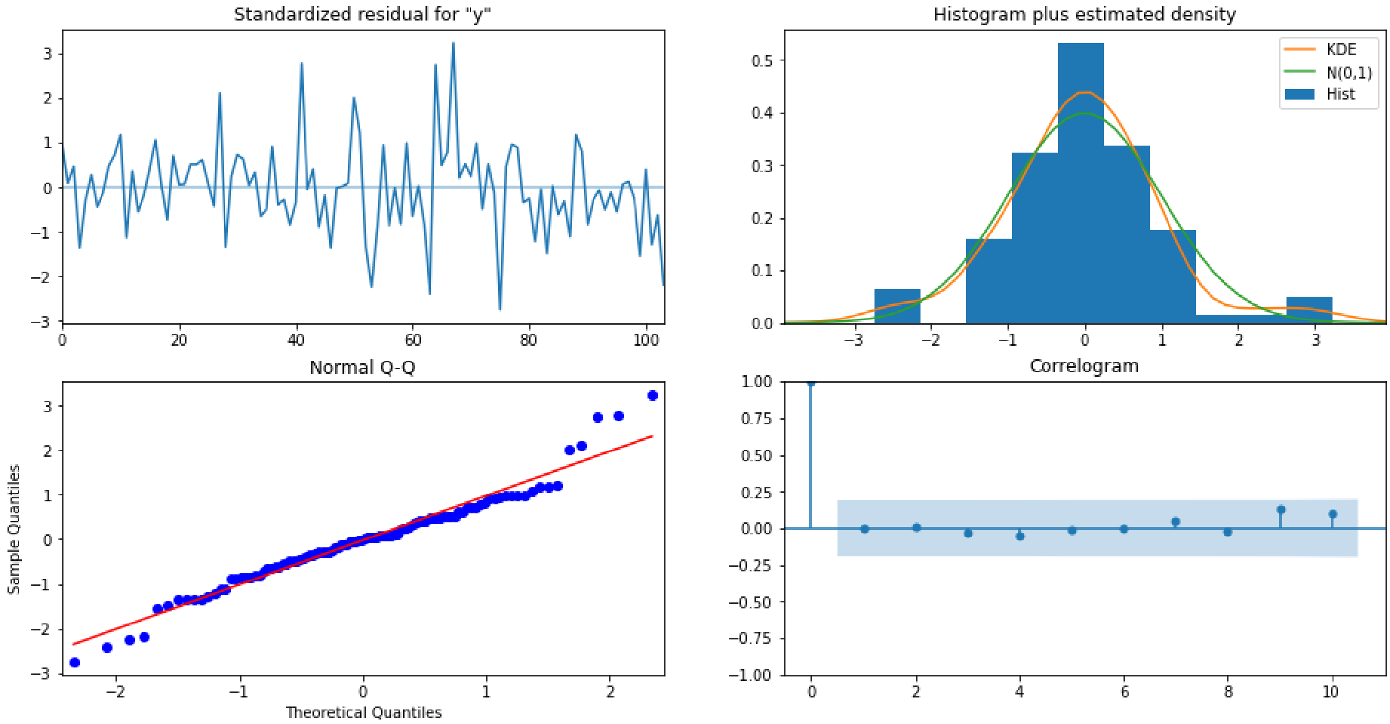
```
Examples of some parameter combinations for Model...
Model: (0, 1, 1)(0, 0, 1, 12)
Model: (0, 1, 2)(0, 0, 2, 12)
Model: (1, 1, 0)(1, 0, 0, 12)
Model: (1, 1, 1)(1, 0, 1, 12)
Model: (1, 1, 2)(1, 0, 2, 12)
Model: (2, 1, 0)(2, 0, 0, 12)
Model: (2, 1, 1)(2, 0, 1, 12)
Model: (2, 1, 2)(2, 0, 2, 12)
```

```
Optimum AIC
```

	param	seasonal	AIC
50	(1, 1, 2)	(1, 0, 2, 12)	1555.584247
53	(1, 1, 2)	(2, 0, 2, 12)	1555.934563
26	(0, 1, 2)	(2, 0, 2, 12)	1557.121584
23	(0, 1, 2)	(1, 0, 2, 12)	1557.160507
77	(2, 1, 2)	(1, 0, 2, 12)	1557.340403

```
                                     SARIMAX Results
=====================================================================================
=======
Dep. Variable:                          Sparkling   No. Observations:             132
Model:             SARIMAX(1, 1, 2)x(1, 0, 2, 12)   Log Likelihood           -770.792
Date:                            Sun, 25 Sep 2022   AIC                      1555.584
Time:                                    12:37:53   BIC                      1574.095
Sample:                                01-31-1980   HQIC                     1563.083
                                     - 12-31-1990
Covariance Type:                              opg
==================================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
----------------------------------------------------------------------------------
ar.L1         -0.6281      0.255     -2.463      0.014      -1.128      -0.128
ma.L1         -0.1041      0.225     -0.463      0.643      -0.545       0.337
ma.L2         -0.7276      0.154     -4.734      0.000      -1.029      -0.426
ar.S.L12       1.0439      0.014     72.844      0.000       1.016       1.072
ma.S.L12      -0.5550      0.098     -5.663      0.000      -0.747      -0.363
ma.S.L24      -0.1355      0.120     -1.134      0.257      -0.370       0.099
sigma2      1.506e+05   2.03e+04      7.400      0.000    1.11e+05     1.9e+05
=====================================================================================
Ljung-Box (L1) (Q):                   0.04   Jarque-Bera (JB):                11.72
Prob(Q):                              0.84   Prob(JB):                         0.00
Heteroskedasticity (H):               1.47   Skew:                             0.36
Prob(H) (two-sided):                  0.26   Kurtosis:                         4.48
=====================================================================================
```

```
Results SARIMA
```



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

Predicted Auto Sarima

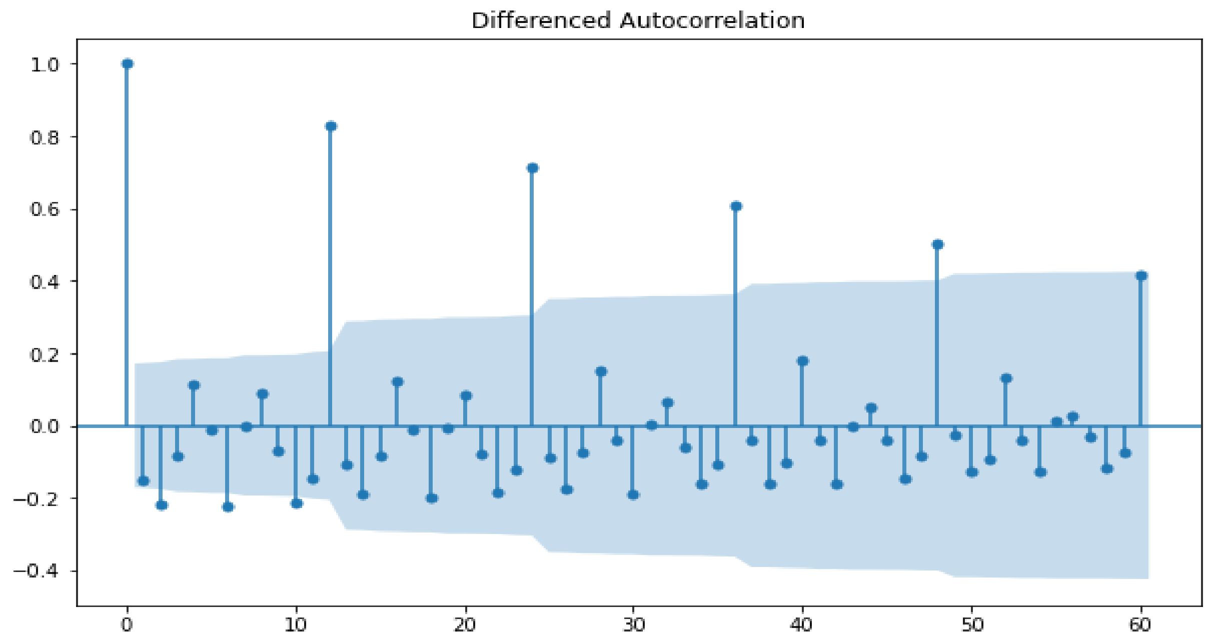
Sparkling	mean	mean_se	mean_ci_lower	mean_ci_upper
1991-01-31	1327.397751	388.342695	566.260054	2088.535448
1991-02-28	1315.126244	402.007271	527.206471	2103.046017
1991-03-31	1621.613165	402.000869	833.705940	2409.520389
1991-04-30	1598.878659	407.238428	800.706007	2397.051312
1991-05-31	1392.706939	407.968694	593.102993	2192.310886

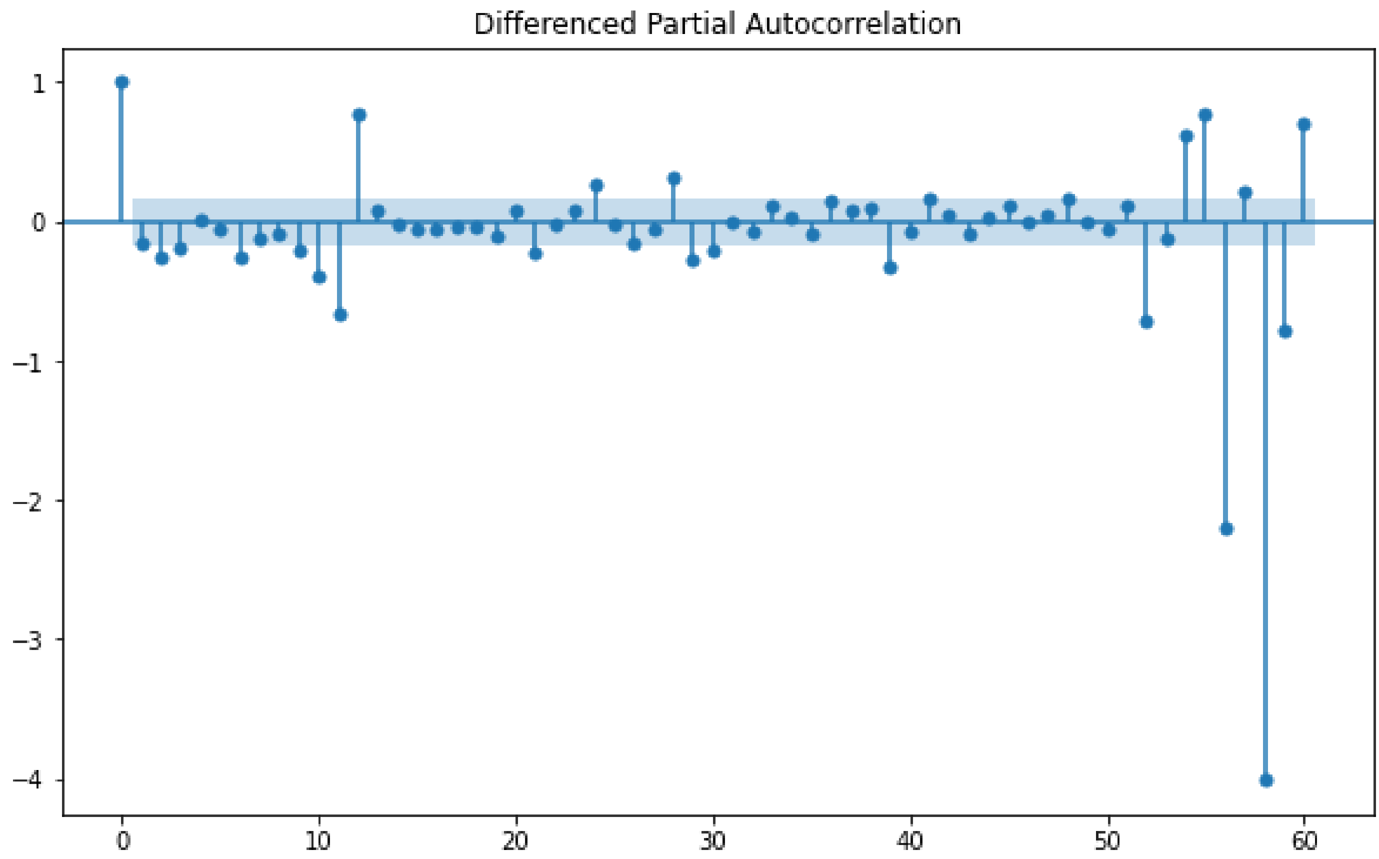
TABLE

	Test RMSE	MAPE
RegressionOnTime	1389.135175	NaN
NaiveModel	3864.279352	NaN

	Test RMSE	MAPE
SimpleAverageModel	1275.081804	NaN
2pointTrailingMovingAverage	813.400684	NaN
4pointTrailingMovingAverage	1156.589694	NaN
6pointTrailingMovingAverage	1283.927428	NaN
9pointTrailingMovingAverage	1346.278315	NaN
Alpha=0.07,SimpleExponentialSmoothing	1338.008384	NaN
Alpha=0.66,Beta=0.0001:DoubleExponentialSmoothing	5291.879833	NaN
Alpha=0.11,Beta=0.001,Gamma=0.46:Triple Exponential Smoothing	378.951023	NaN
Alpha=0.11,Beta=0.05,Gamma=0.36:Triple Exponential Smoothing 2	404.286809	NaN
Arima 2,1,2	1299.979569	47.099932
Sarima (1,1,2)(2,0,2,6)	2117.224709	100.280958
SARIMA(1,1,1)(1,0,2,12)	528.592450	20.955012

Q.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.





Here, we have taken alpha=0.05.

We are going to take the seasonal period as 6. We will keep the p and q parameters same as the ARIMA model.

The Auto-Regressive parameter in an SARIMA model is 'P' which comes from the significant lag after which the PACF plot cuts-off to 0. The Moving-Average parameter in an SARIMA model is 'q' which comes from the significant lag after which the ACF plot cuts-off to 0. Remember to check the ACF and the PACF plots only at multiples of 6 (since 6 is the seasonal period). By looking at the plots we see that the ACF and the PACF do not directly cut-off to 0.

This is a common problem while building models by looking at the ACF and the PACF plots. But we are able to explain the model.

Please do play around with the data and try out different kinds of transformations and different levels of differencing on this data. We have not taken the logarithm of the series and then trying it out.

ARIMA MODEL

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

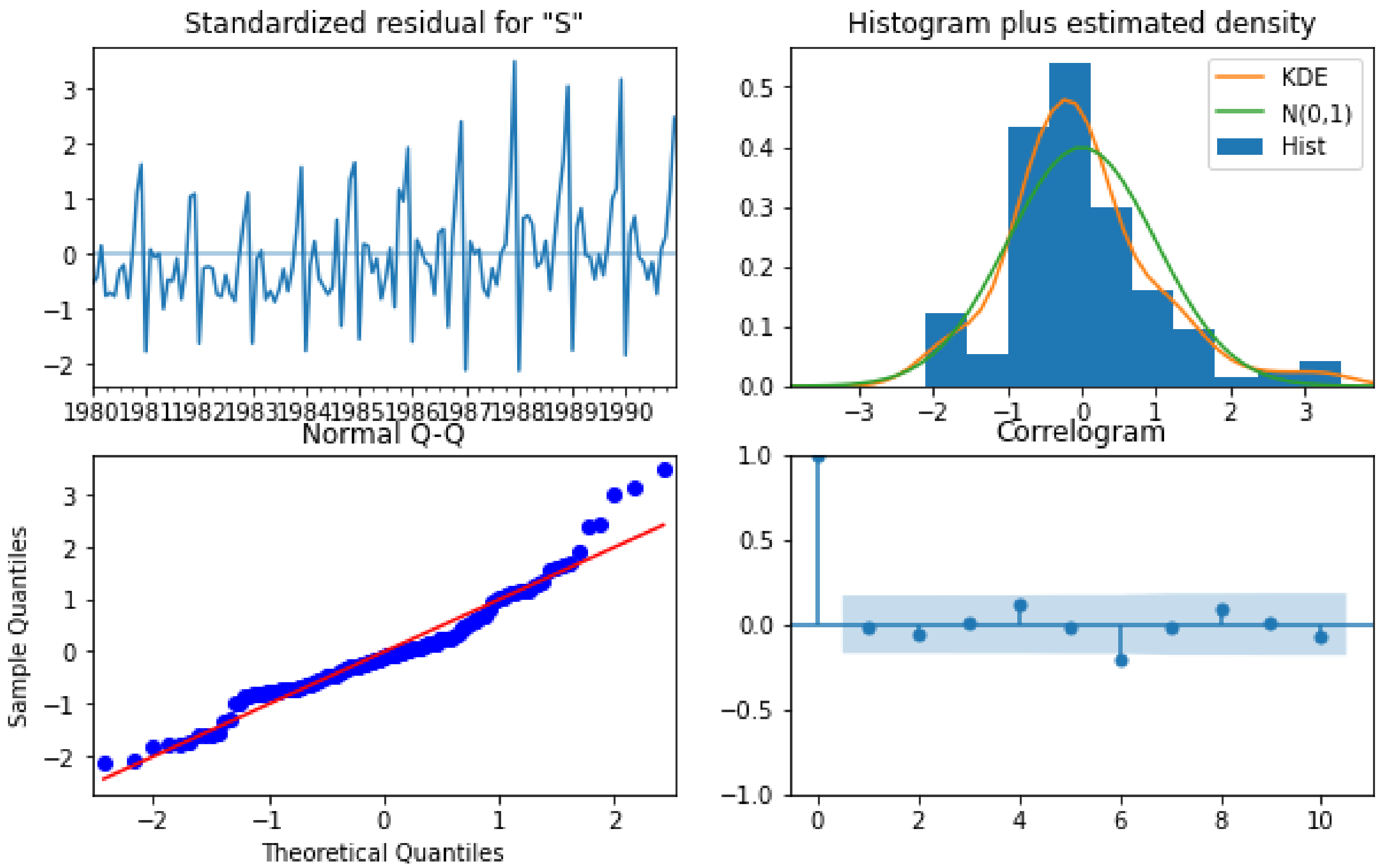
```
Model: (1, 1, 0)
Model: (1, 1, 1)
Model: (1, 1, 2)
Model: (2, 1, 0)
Model: (2, 1, 1)
Model: (2, 1, 2)
```

Optimum AIC

	param	AIC
8	(2, 1, 2)	2213.509212
7	(2, 1, 1)	2233.777626
2	(0, 1, 2)	2234.408323
5	(1, 1, 2)	2234.527200
4	(1, 1, 1)	2235.755095

```
                               SARIMAX Results
==============================================================================
Dep. Variable:              Sparkling   No. Observations:                  132
Model:                 ARIMA(2, 0, 1)   Log Likelihood               -1113.295
Date:                Sun, 25 Sep 2022   AIC                           2236.591
Time:                        12:37:57   BIC                           2251.005
Sample:                    01-31-1980   HQIC                          2242.448
                         - 12-31-1990
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const       2399.4586    118.215     20.297      0.000    2167.762    2631.155
ar.L1          1.2375      0.138      8.938      0.000       0.966       1.509
ar.L2         -0.5293      0.124     -4.266      0.000      -0.772      -0.286
ma.L1         -0.8080      0.156     -5.174      0.000      -1.114      -0.502
sigma2      1.233e+06   1.37e+05      9.016      0.000    9.65e+05     1.5e+06
===================================================================================
Ljung-Box (L1) (Q):                   0.03   Jarque-Bera (JB):                26.42
Prob(Q):                              0.86   Prob(JB):                         0.00
Heteroskedasticity (H):               2.40   Skew:                             0.80
Prob(H) (two-sided):                  0.00   Kurtosis:                         4.49
===================================================================================
```

Results auto ARIMA



TABLE

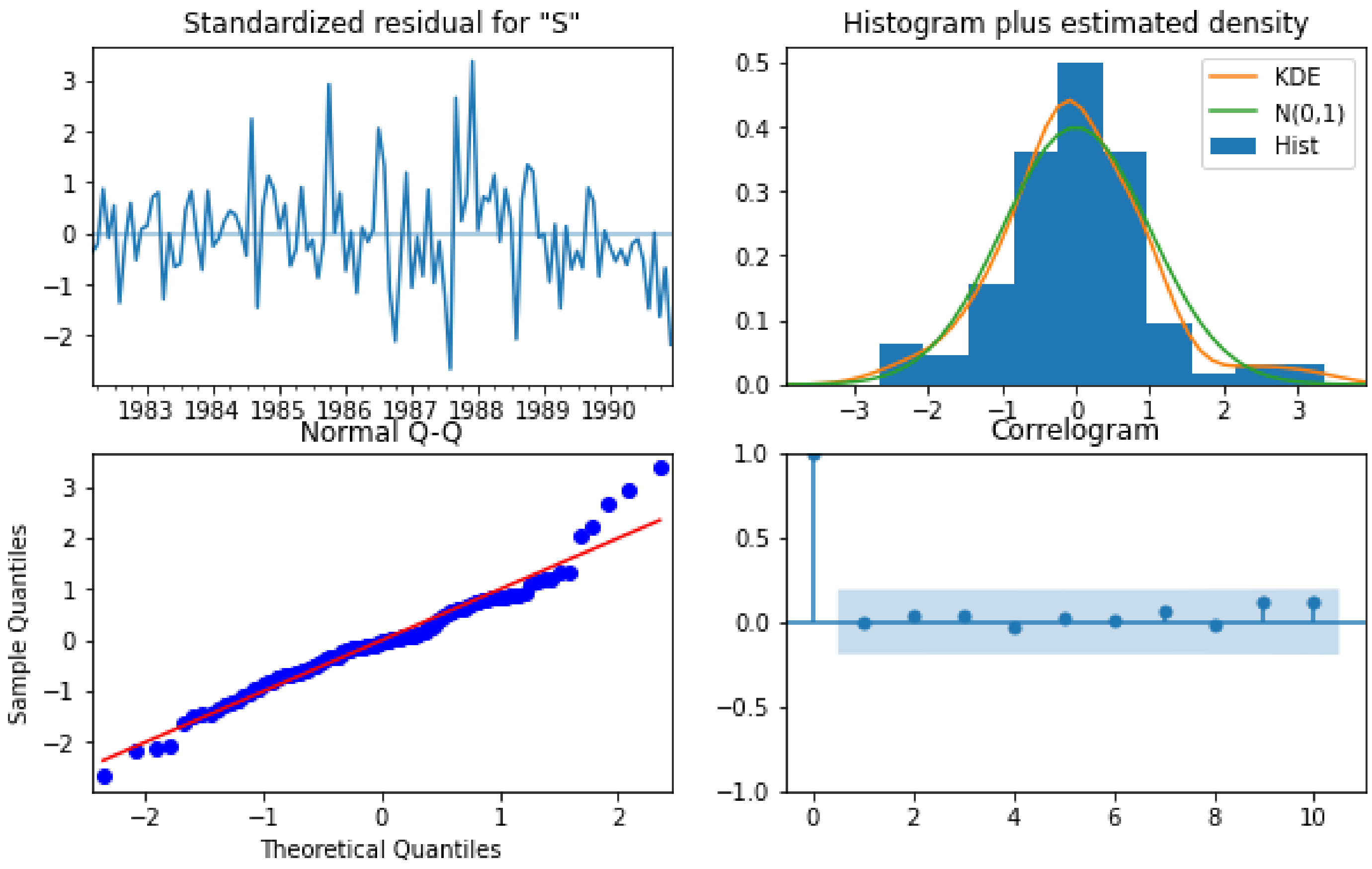
	Test RMSE	MAPE
RegressionOnTime	1389.135175	NaN
NaiveModel	3864.279352	NaN
SimpleAverageModel	1275.081804	NaN
2pointTrailingMovingAverage	813.400684	NaN
4pointTrailingMovingAverage	1156.589694	NaN
6pointTrailingMovingAverage	1283.927428	NaN
9pointTrailingMovingAverage	1346.278315	NaN
Alpha=0.07,SimpleExponentialSmoothing	1338.008384	NaN
Alpha=0.66,Beta=0.0001:DoubleExponentialSmoothing	5291.879833	NaN

	Test RMSE	MAPE
Alpha=0.11,Beta=0.001,Gamma=0.46:Triple Exponential Smoothing	378.951023	NaN
Alpha=0.11,Beta=0.05,Gamma=0.36:Triple Exponential Smoothing 2	404.286809	NaN
Arima 2,1,2	1299.979569	47.099932
Sarima (1,1,2)(2,0,2,6)	2117.224709	100.280958
SARIMA(1,1,1)(1,0,2,12)	528.592450	20.955012
Arima (2,0,1)	1269.345658	36.690219

SARIMA MODEL

```
                                     SARIMAX Results
==========================================================================================
============
Dep. Variable:                               Sparkling   No. Observations:          132
Model:             SARIMAX(2, 0, 1)x(1, 0, [1, 2], 12)   Log Likelihood        -783.994
Date:                                 Sun, 25 Sep 2022   AIC                   1581.989
Time:                                         12:38:03   BIC                   1600.633
Sample:                                     01-31-1980   HQIC                  1589.546
                                          - 12-31-1990
Covariance Type:                                   opg
=================================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
---------------------------------------------------------------------------------
ar.L1         -0.5530      0.237     -2.336      0.019      -1.017      -0.089
ar.L2          0.0422      0.133      0.318      0.751      -0.218       0.302
ma.L1          0.7914      0.217      3.653      0.000       0.367       1.216
ar.S.L12       1.0250      0.009    112.167      0.000       1.007       1.043
ma.S.L12      -0.4792      0.096     -4.976      0.000      -0.668      -0.290
ma.S.L24      -0.1047      0.123     -0.848      0.396      -0.347       0.137
sigma2      1.487e+05   1.93e+04      7.685      0.000    1.11e+05    1.87e+05
======================================================================================
Ljung-Box (L1) (Q):                   0.00   Jarque-Bera (JB):                    12.04
Prob(Q):                              0.98   Prob(JB):                             0.00
Heteroskedasticity (H):               1.46   Skew:                                 0.43
Prob(H) (two-sided):                  0.27   Kurtosis:                             4.41
======================================================================================
```

Results SARIMA New



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

Table

	Test RMSE	MAPE
RegressionOnTime	1389.135175	NaN
NaiveModel	3864.279352	NaN
SimpleAverageModel	1275.081804	NaN
2pointTrailingMovingAverage	813.400684	NaN
4pointTrailingMovingAverage	1156.589694	NaN
6pointTrailingMovingAverage	1283.927428	NaN
9pointTrailingMovingAverage	1346.278315	NaN
Alpha=0.07,SimpleExponentialSmoothing	1338.008384	NaN
Alpha=0.66,Beta=0.0001:DoubleExponentialSmoothing	5291.879833	NaN

	Test RMSE	MAPE
Alpha=0.11,Beta=0.001,Gamma=0.46:Triple Exponential Smoothing	378.951023	NaN
Alpha=0.11,Beta=0.05,Gamma=0.36:Triple Exponential Smoothing 2	404.286809	NaN
Arima 2,1,2	1299.979569	47.099932
Sarima (1,1,2)(2,0,2,6)	2117.224709	100.280958
SARIMA(1,1,1)(1,0,2,12)	528.592450	20.955012
Arima (2,0,1)	1269.345658	36.690219
SARIMA(2, 0, 1)(1, 0, 2, 12)	460.609889	15.298293

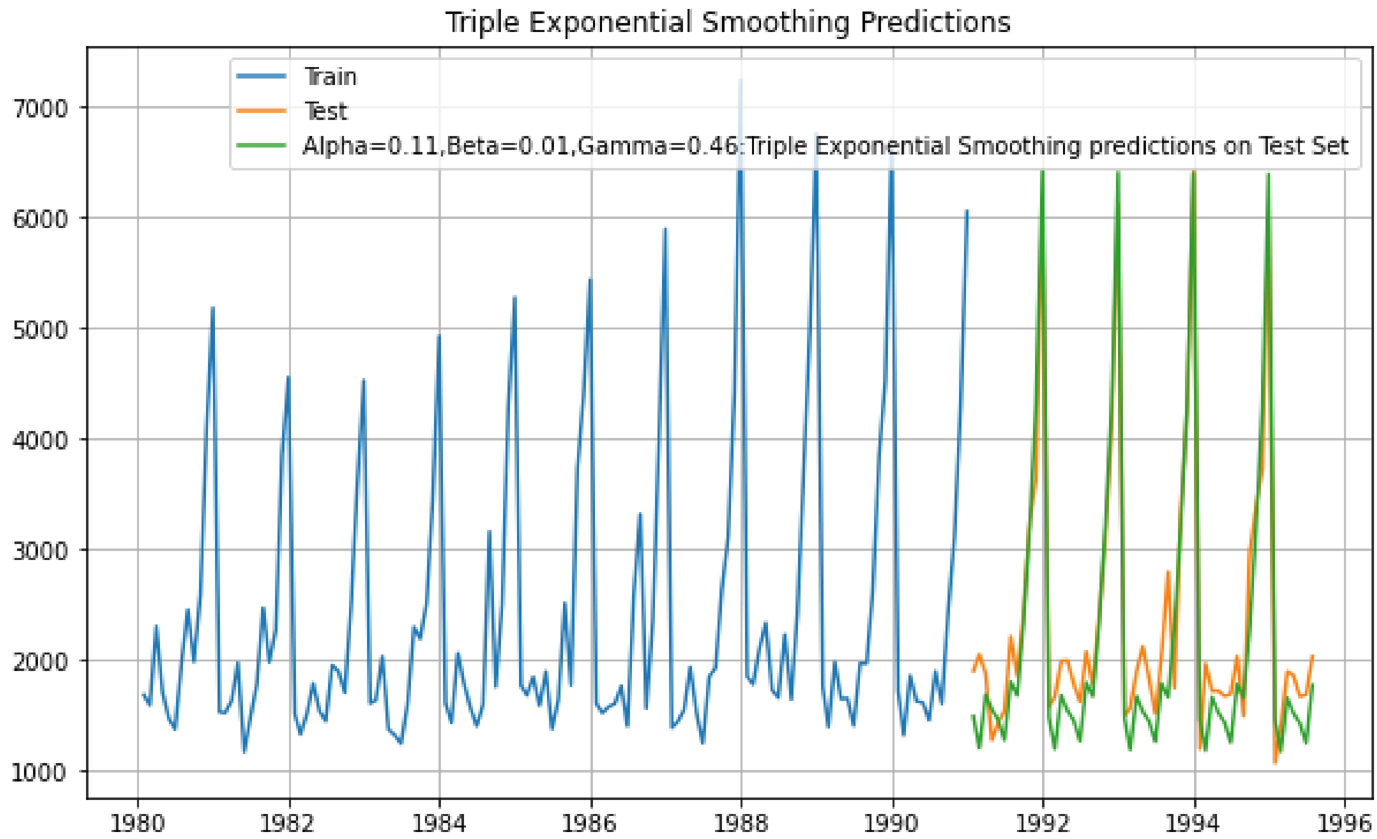
Q.8 Build a table with all the models built along with their corresponding parameters and the respective RMSE values on the test data.

```
Sorted by RMSE values on the Test Data:
                                                      Test RMSE        MAPE
Alpha=0.11,Beta=0.001,Gamma=0.46:Triple Exponen...   378.951023         NaN
Alpha=0.11,Beta=0.05,Gamma=0.36:Triple Exponent...   404.286809         NaN
SARIMA(2, 0, 1)(1, 0, 2, 12)                         460.609889   15.298293
SARIMA(1,1,1)(1,0,2,12)                              528.592450   20.955012
2pointTrailingMovingAverage                          813.400684         NaN
4pointTrailingMovingAverage                         1156.589694         NaN
Arima (2,0,1)                                       1269.345658   36.690219
SimpleAverageModel                                  1275.081804         NaN
6pointTrailingMovingAverage                         1283.927428         NaN
Arima 2,1,2                                         1299.979569   47.099932
Alpha=0.07,SimpleExponentialSmoothing               1338.008384         NaN
9pointTrailingMovingAverage                         1346.278315         NaN
RegressionOnTime                                    1389.135175         NaN
Sarima (1,1,2)(2,0,2,6)                             2117.224709  100.280958
NaiveModel                                          3864.279352         NaN
Alpha=0.66,Beta=0.0001:DoubleExponentialSmoothing   5291.879833         NaN


Sorted by MAPE values on the Test Data:
                                                      Test RMSE        MAPE
SARIMA(2, 0, 1)(1, 0, 2, 12)                         460.609889   15.298293
SARIMA(1,1,1)(1,0,2,12)                              528.592450   20.955012
Arima (2,0,1)                                       1269.345658   36.690219
Arima 2,1,2                                         1299.979569   47.099932
Sarima (1,1,2)(2,0,2,6)                             2117.224709  100.280958
RegressionOnTime                                    1389.135175         NaN
NaiveModel                                          3864.279352         NaN
SimpleAverageModel                                  1275.081804         NaN
```

2pointTrailingMovingAverage	813.400684	NaN
4pointTrailingMovingAverage	1156.589694	NaN
6pointTrailingMovingAverage	1283.927428	NaN
9pointTrailingMovingAverage	1346.278315	NaN
Alpha=0.07,SimpleExponentialSmoothing	1338.008384	NaN
Alpha=0.66,Beta=0.0001:DoubleExponentialSmoothing	5291.879833	NaN
Alpha=0.11,Beta=0.001,Gamma=0.46:Triple Exponen...	378.951023	NaN
Alpha=0.11,Beta=0.05,Gamma=0.36:Triple Exponent...	404.286809	NaN

Q.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.

TRIPLE EXPONENTIAL



FULL MODEL RMSE

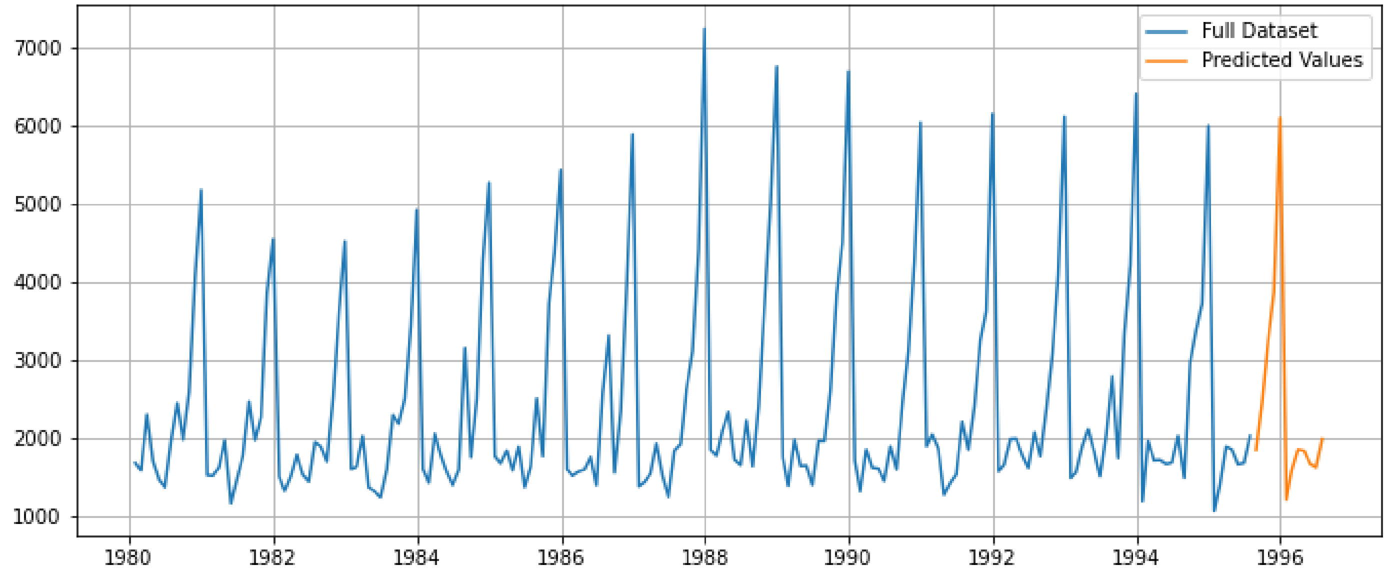
RMSE of the Full Model 367.8420808631773

Getting the predictions for the 12 months

```
1995-08-31    1852.150187
1995-09-30    2456.335190
1995-10-31    3247.505332
1995-11-30    3875.303642
```

```
1995-12-31    6104.225647
1996-01-31    1217.141787
1996-02-29    1601.821047
1996-03-31    1859.066850
1996-04-30    1841.588172
1996-05-31    1680.129986
1996-06-30    1630.752897
1996-07-31    1992.614818
```

Predicted Values



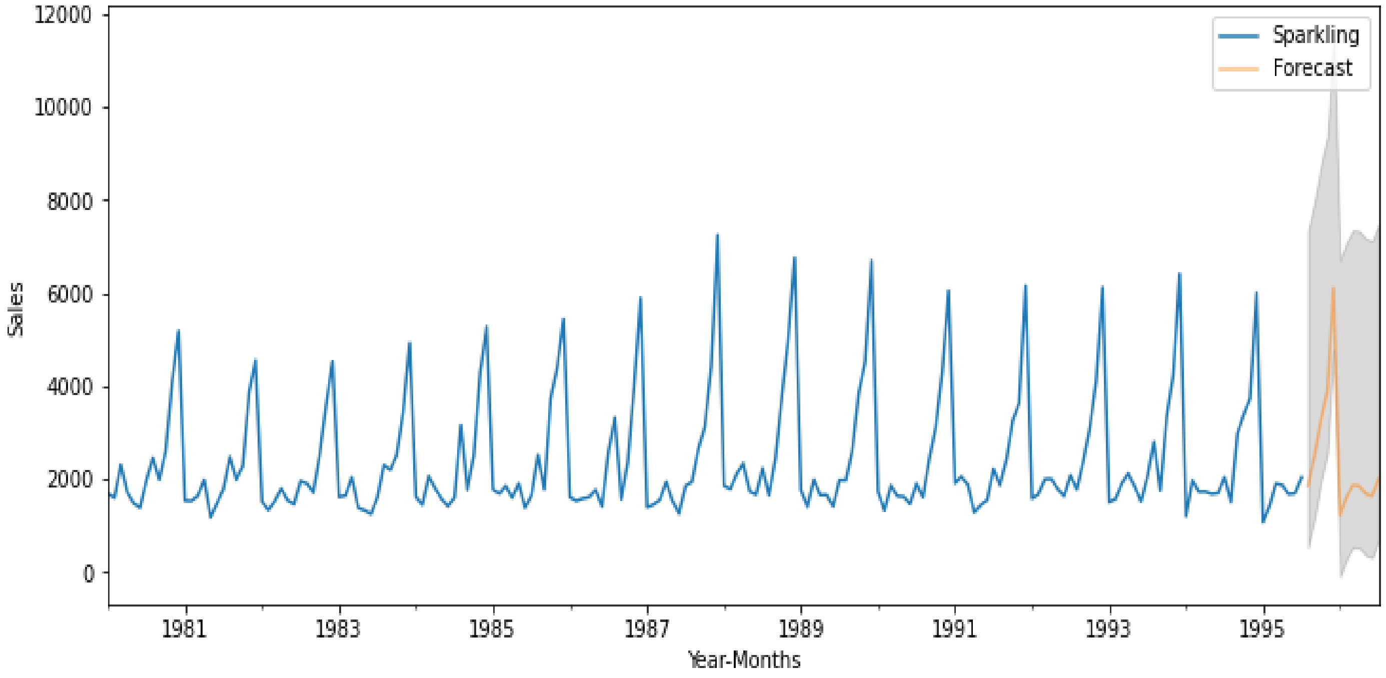
we have calculated the upper and lower confidence bands at 95% confidence level

The percentile function under numpy lets us calculate these and adding and subtracting from the predictions

This gives us the necessary confidence bands for the predictions

	lower_CI	prediction	upper_ci
1995-08-31	529.221603	1852.150187	7343.422282
1995-09-30	1133.406607	2456.335190	7947.607286
1995-10-31	1924.576749	3247.505332	8738.777427
1995-11-30	2552.375059	3875.303642	9366.575738
1995-12-31	4781.297063	6104.225647	11595.497742

Plot the forecast along with the confidence band



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

Here, we can see that the sale of the wine is pretty good the highest sale recorded was in 1988 after that there is a decent amount of decrease in the sales but the sales is quite constant but we in order to increase the sale we have to take certain steps

1. We can see what our competitors pricing are if our customer are buying different company's wine due to pricing.

2. We can also pay attention to our services as well what kind of services are we giving to our customers.

3. We can go by the motto of less price but quality wine as people would love to buy good heavy wine but in less price we have to adjust according to the market running and also have to keep an eye on our competitors.

4. Wine can be a thing which can be a seasonal thing as when the time of festivals comes around we tend to buy more wine so we can offer some kind of coupons or a buy 1 get 1 free scheme or we can see while buying

wine what other things customer tend to purchase so that we can add that item in our schemes.

5.Since it can be highly seasonal thing we have to make some kind of strategic adjustment so that we may be able to sell more wines in off season as well and we should keep our customer happy by our services and quality of wine of course with a attractive price.