Business Report

SMDM Project Business Report DSBA

Sanjay Srinivasan
PGP-DSBA Online
JULY' 21 Batch
Date: 20-02-2022

INDEX

List Of Figures

Problem - 1

Summary

The data is gathered from ABC Estate wines, for Sparkling wine sales data from this ABC Estate wine company. An analyst for the company needs to analyze the wine sales in the 19th century and forecast the wine sales for the 20th century.

Introduction

The purpose of this exercise is to explore the dataset and make the analyze the wine sales in the 19th Century, based on the sales data we need to forecast for the wine sales data for next 12 months.

Sample of the dataset:

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

Fig 1.1 Dataset Sample

Exploratory Data Analysis

Let us check the types of variables in the data frame.

```
YearMonth    object
Sparkling     int64
dtype: object
```

Fig- 1.2. Datatypes of the variable

There are total 187 rows and 2 columns in the dataset. 1 columns are object and 1 columns are int64

Check for missing values in the dataset:

```
<class 'pandas.core.frame.DataFrame'>
DatetimeIndex: 187 entries, 1980-01-31 to 1995-07-31
Data columns (total 1 columns):
Sparkling    187 non-null int64
dtypes: int64(1)
memory usage: 7.9 KB
```

Fig- 1.3. Check null values

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

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

Fig- 1.4. Initialising Date as index Column

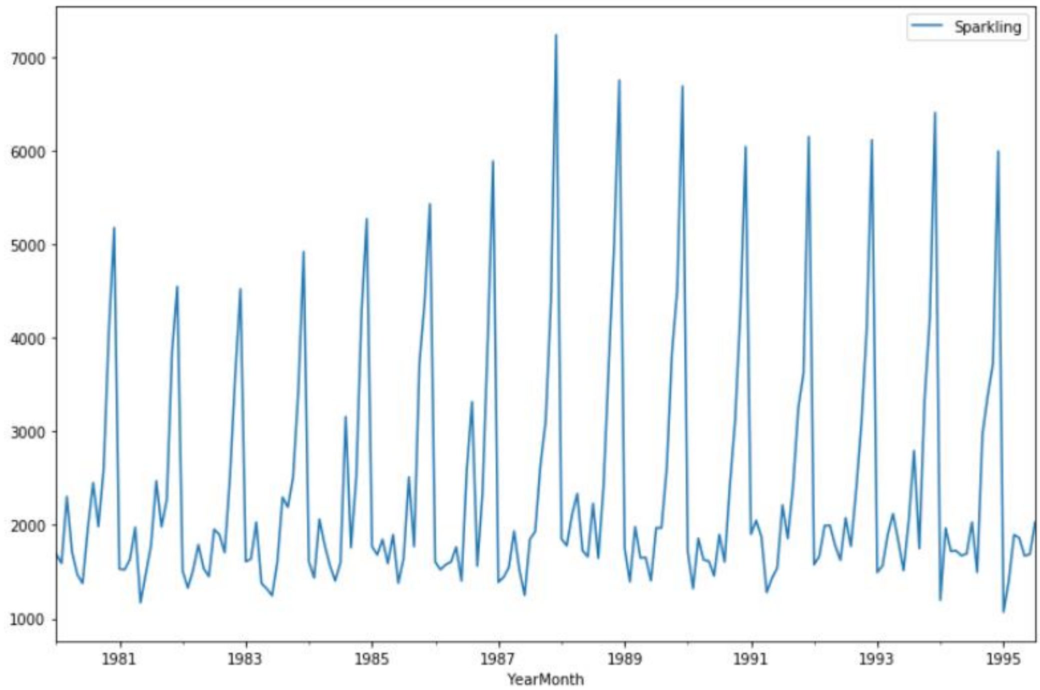


Fig – 1.5 Plotting Sparkling wine data

The sparkling wine sales data has been plotted against the year of sales.

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

	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

Fig – 1.6 Sparkling wine sales data spread.

```
Text(0.5, 0, 'Year')
```

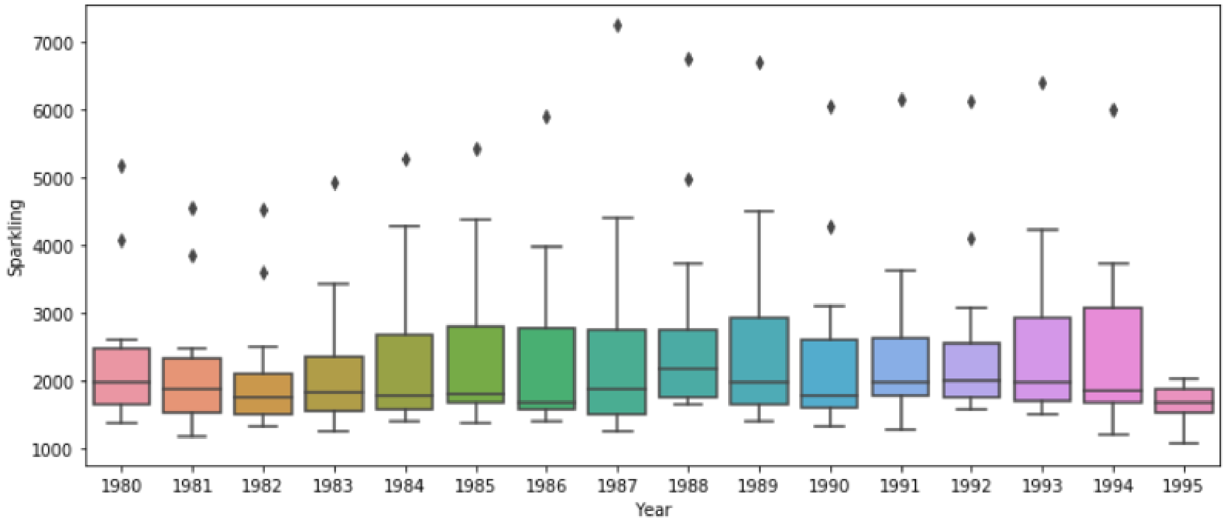


Fig – 1.7 Sparkling wine sales across years

From the above chart (boxplot), there are outliers present in the data and we can observe that there was good sales record for sparkling wine from 1980-1994 and wine sales has been decreased in the year 1995.

```
Text(0.5, 0, 'Months')
```

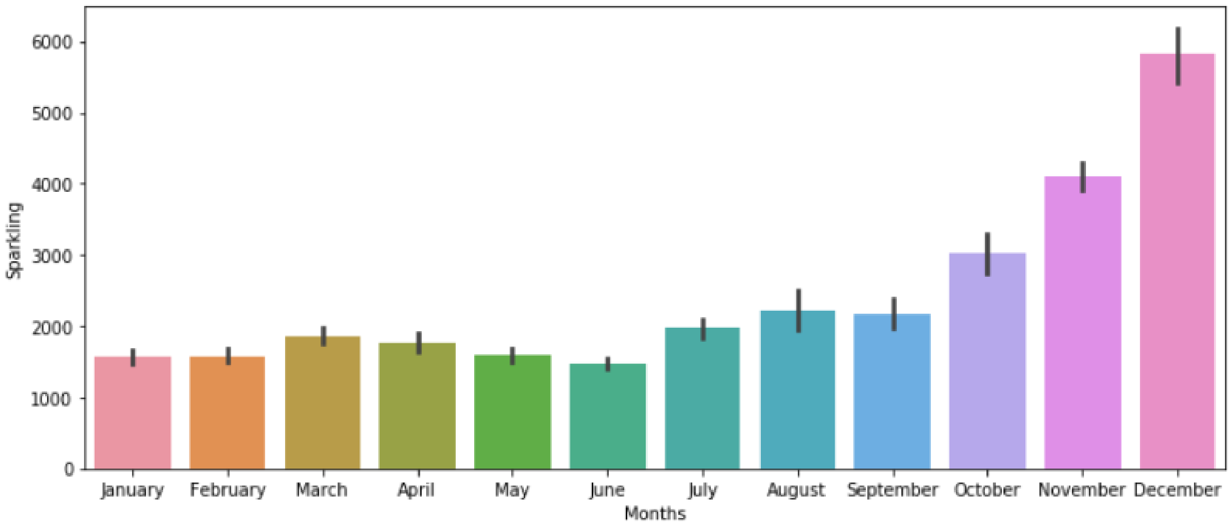


Fig – 1.8 Sparkling wine sales across months

From the above chart (boxplot), the December month has highest number of sparkling wine sales when compared with other months.

YearMonth	April	August	December	February	January	July	June	March	May	November	October	September
YearMonth												
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
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

Fig – 1.9 Monthwise wine sales across years

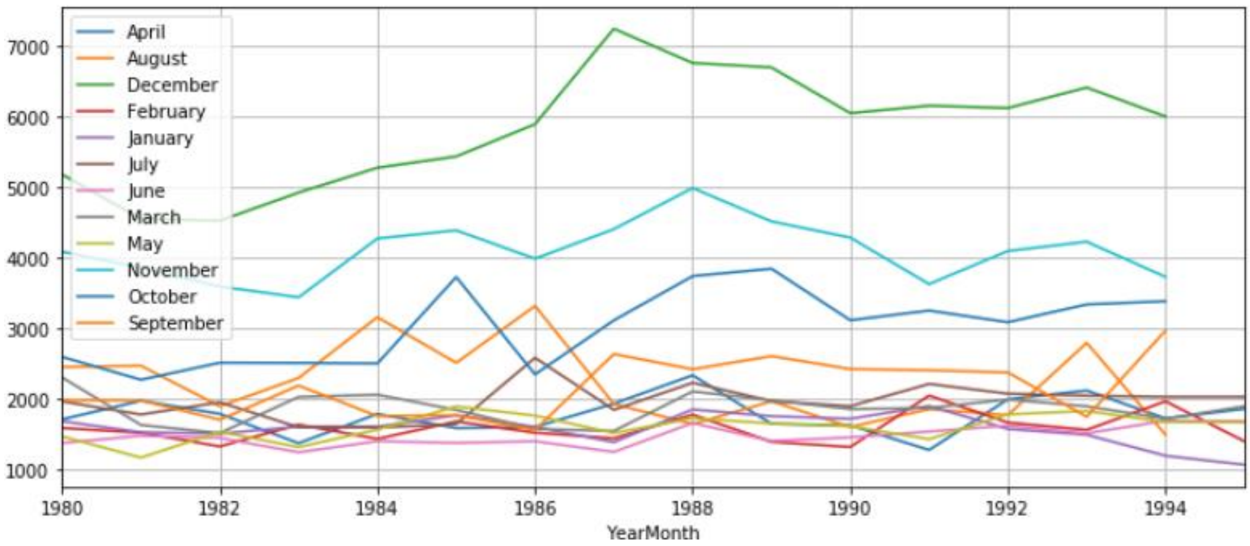


Fig – 1.10 Monthwise wine sales across years

Additive Model:

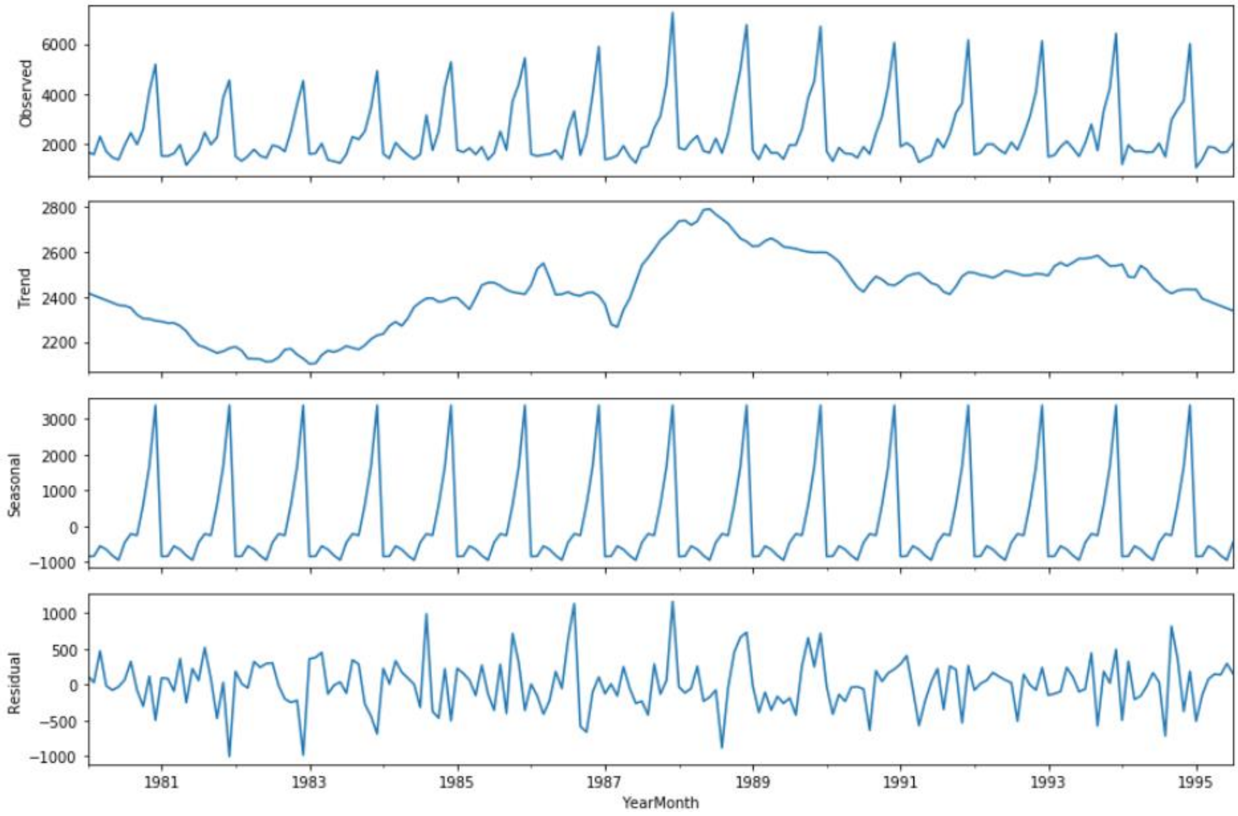


Fig – 1.11 Decompose data form the original dataset for sparkling wines (Additive model)

```
Trend
               Sparkling
YearMonth
1980-01-31  2417.116647
1980-02-29  2406.350767
1980-03-31  2395.584887
1980-04-30  2384.819007
1980-05-31  2374.053127

Seasonality
              Sparkling
YearMonth
1980-01-31 -852.939513
1980-02-29 -845.687698
1980-03-31 -560.602944
1980-04-30 -656.317669
1980-05-31 -827.225103

Residual
              Sparkling
YearMonth
1980-01-31  121.822865
1980-02-29   30.336930
1980-03-31  469.018056
1980-04-30  -16.501338
1980-05-31  -75.828025
```

Fig – 1.12 Trend, Seasonality and residual values after decomposing the original data for sparkling wines (Additive model)

Multiplicative Model:

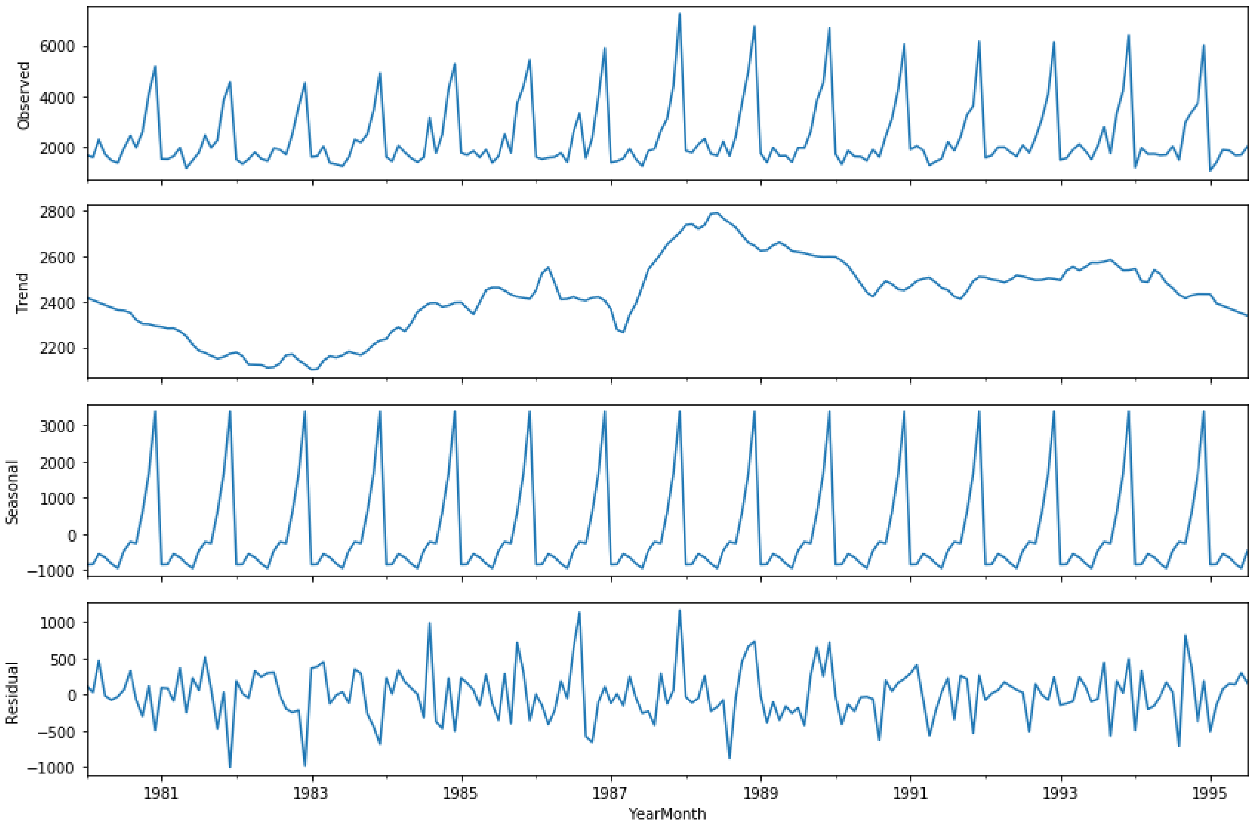


Fig – 1.13 Decompose data form the original dataset for sparkling wines (Multiplicative model)

```
Trend
              Sparkling
YearMonth
1980-01-31  2417.116647
1980-02-29  2406.350767
1980-03-31  2395.584887
1980-04-30  2384.819007
1980-05-31  2374.053127

Seasonality
             Sparkling
YearMonth
1980-01-31 -852.939513
1980-02-29 -845.687698
1980-03-31 -560.602944
1980-04-30 -656.317669
1980-05-31 -827.225103

Residual
             Sparkling
YearMonth
1980-01-31  121.822865
1980-02-29   30.336930
1980-03-31  469.018056
1980-04-30  -16.501338
1980-05-31  -75.828025
```

Fig – 1.14 Trend, Seasonality and residual values after decomposing the original data for sparkling wines (Multiplicative model)

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

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

Fig – 1.15 Last 5 values for Training data

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

Fig – 1.16 First 5 values for testing data

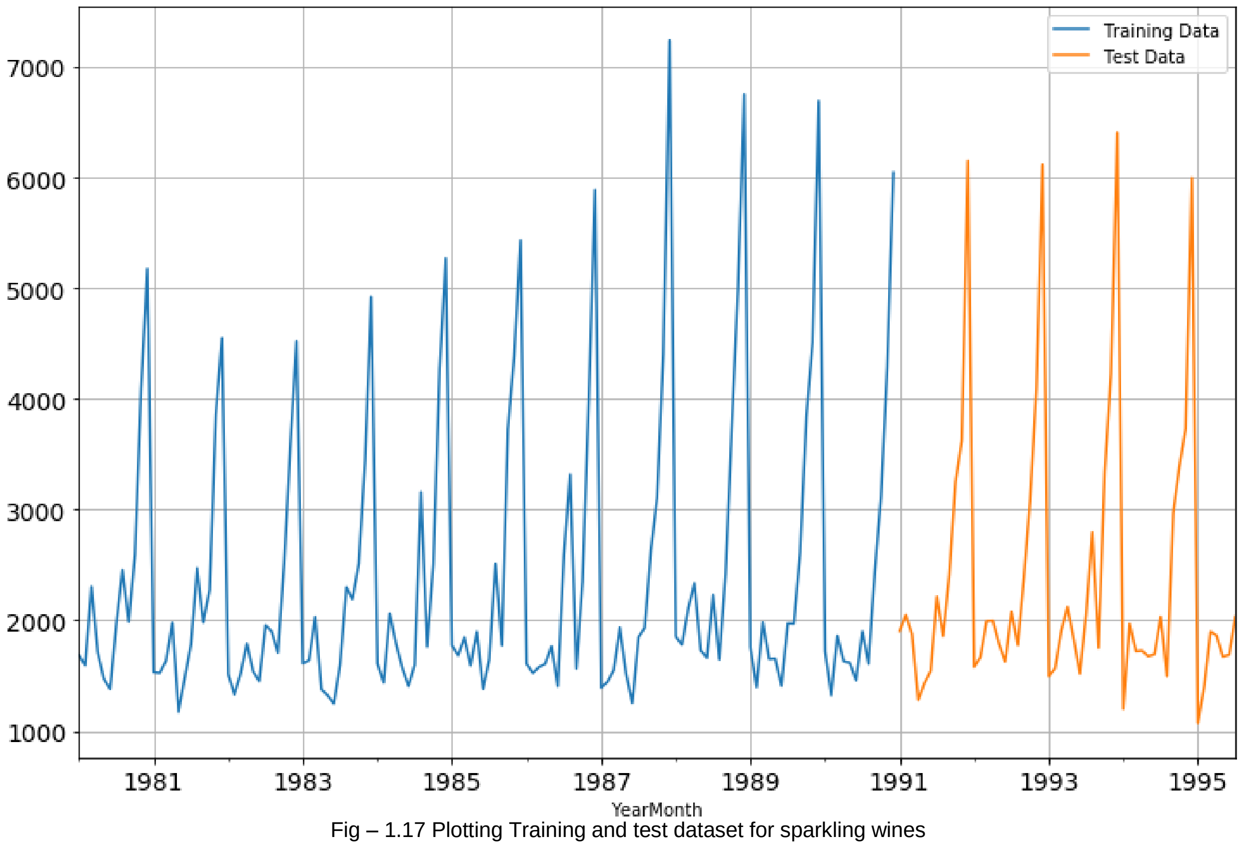


Fig – 1.17 Plotting Training and test dataset for sparkling wines

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:

```
Training Time instance for Sparkling dataset

[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, 3
4, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 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, 105, 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 for Sparkling dataset

[133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157,
158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 18
3, 184, 185, 186, 187]
```

Fig – 1.18 Generating numerical time instance for both training and test dataset for sparkling wines

We see that we have successfully the generated the numerical time instance order for both the training and test set. Now we will add these values in the training and test set.

```
LinearRegression()
```

Fig – 1.19 Initializing Linear Regression method

YearMonth	Sparkling	time	RegOnTime
1980-01-31	1686	1	2021.741171
1980-02-29	1591	2	2027.573830
1980-03-31	2304	3	2033.406488
1980-04-30	1712	4	2039.239147
1980-05-31	1471	5	2045.071805

Fig – 1.20 Predicting for training dataset

YearMonth	Sparkling	time	RegOnTime
1991-01-31	1902	133	2791.652093
1991-02-28	2049	134	2797.484752
1991-03-31	1874	135	2803.317410
1991-04-30	1279	136	2809.150069
1991-05-31	1432	137	2814.982727

Fig – 1.21 Predicting for test dataset

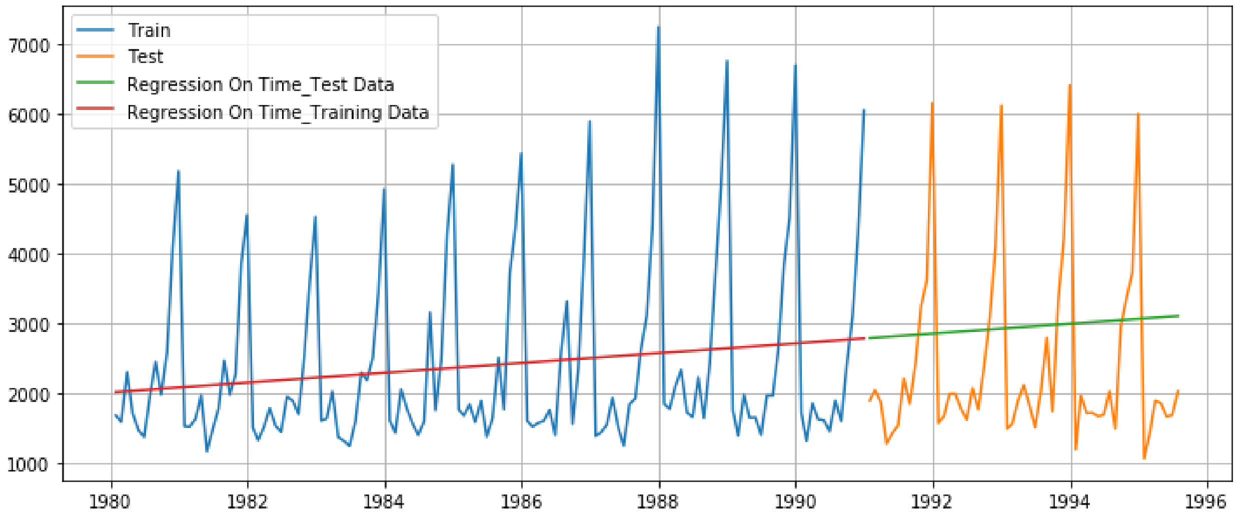


Fig – 1.22 Ploting original and predicted train and test datas using linear regression

```
For RegressionOnTime forecast on the Training Data,  RMSE is 1279.322 and MAPE is 40.05
```

Fig – 1.23 RMSE and MAPE value Training data

```
For RegressionOnTime forecast on the Test Data,  RMSE is 1389.135 and MAPE is 50.15
```

Fig – 1.24 RMSE and MAPE value Test data

	Test RMSE	Test MAPE
Regression On Time	1389.135175	50.15

Fig – 1.25 Loading RMSE and MAPE value Test data into dataframe

Model - 2: Naive Approach: ($\hat{y}_{t+1} = y_t$)

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 prediction of tomorrow is same as today,therefore the prediction for day after tomorrow is also today.

```
YearMonth
1980-01-31    6047
1980-02-29    6047
1980-03-31    6047
1980-04-30    6047
1980-05-31    6047
Name: naive, dtype: int64
```

Fig – 1.26 Predicting values for training data using naïve approach

```
YearMonth
1991-01-31    2031
1991-02-28    2031
1991-03-31    2031
1991-04-30    2031
1991-05-31    2031
Name: naive, dtype: int64
```

Fig – 1.27 Predicting values for test data using naïve approach

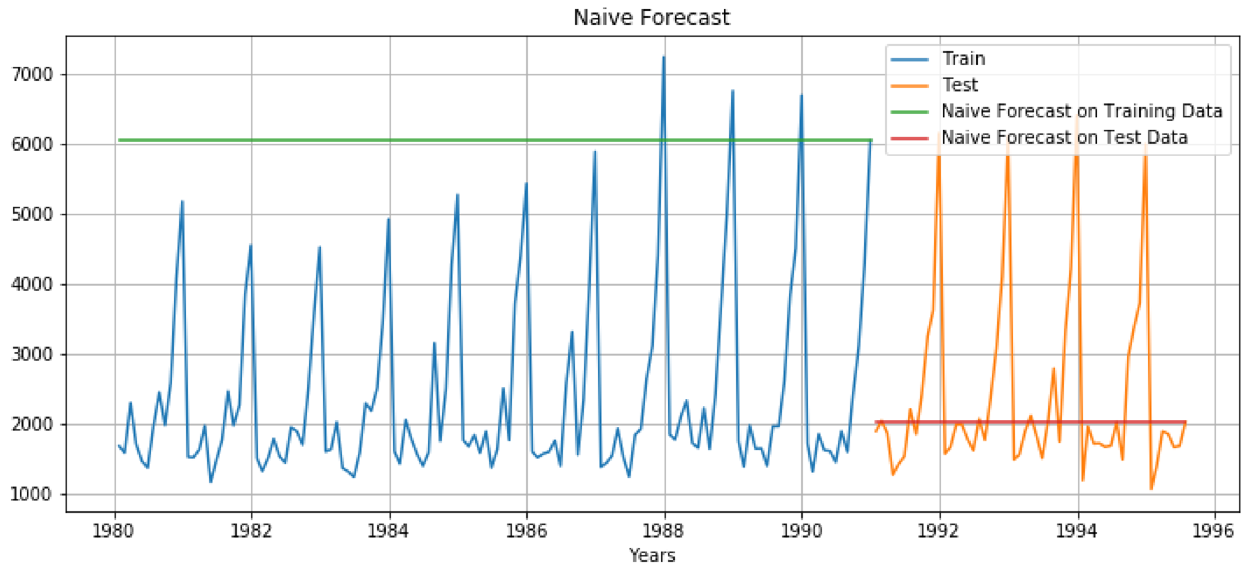


Fig – 1.28 Plotting the predicted values for train and test data using naïve approach

```
For Naive Model forecast on the Training Data,  RMSE is 3867.701 and MAPE is 153.17
```

Fig – 1.29 RMSE and MAPE value Training data

```
For RegressionOnTime forecast on the Test Data,  RMSE is 1327.156 and MAPE is 32.90
```

Fig – 1.30 RMSE and MAPE value Test data

	Test RMSE	Test MAPE
Regression On Time	1389.135175	50.15
NaiveModel	1327.156057	32.90

Fig – 1.31 Loading RMSE and MAPE value Test data into dataframe

Model – 3 Simple Average:

For this particular simple average method, we will forecast by using the average of the training values

YearMonth	Sparkling	mean_forecast
1980-01-31	1686	2403.780303
1980-02-29	1591	2403.780303
1980-03-31	2304	2403.780303
1980-04-30	1712	2403.780303
1980-05-31	1471	2403.780303

Fig – 1.32 Taking mean of sparkling wine sales training data

YearMonth	Sparkling	mean_forecast
1980-01-31	1686	2403.780303
1980-02-29	1591	2403.780303
1980-03-31	2304	2403.780303
1980-04-30	1712	2403.780303
1980-05-31	1471	2403.780303

Fig – 1.33 Taking mean of sparkling wine sales test data

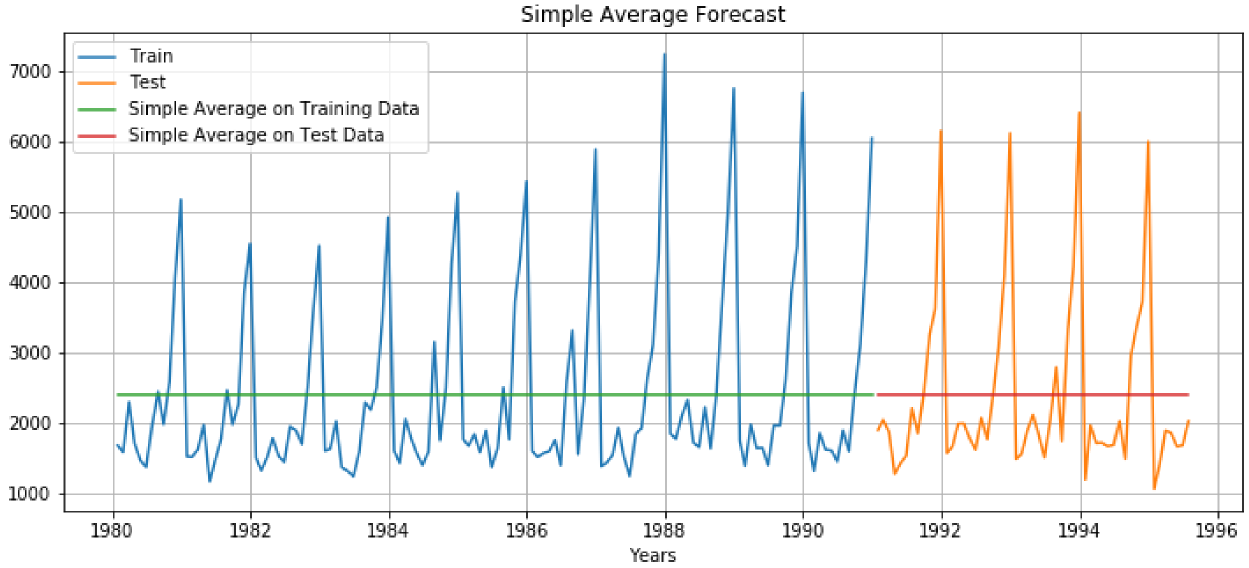


Fig – 1.34 Plotting the Simple Average data, train and test data

```
For Simple Average Model forecast on the Training Data,  RMSE is 1298.484 and MAPE is 40.36
```

Fig – 1.35 RMSE and MAPE value Training data

```
For Simple Average forecast on the Test Data,  RMSE is 1275.073 and MAPE is 38.81
```

Fig – 1.36 RMSE and MAPE value Test data

	Test RMSE	Test MAPE
Regression On Time	1389.135175	50.15
NaiveModel	1327.156057	32.90
SimpleAverageModel	1275.073380	38.81

Fig – 1.37 Loading RMSE and MAPE value Test data into dataframe

Model – 4 Moving Average(MA):

For the moving average model, we are going to calculate rolling means (or moving averages) for different intervals. The best interval can be determined by the maximum accuracy (or the minimum error) over here.

For Moving Average, we are going to average over the entire data.

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

Fig – 1.38 training data

	Sparkling	Trailing_2	Trailing_4	Trailing_6	Trailing_9
YearMonth					
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

Fig – 1.39 Making data from training data for moving average

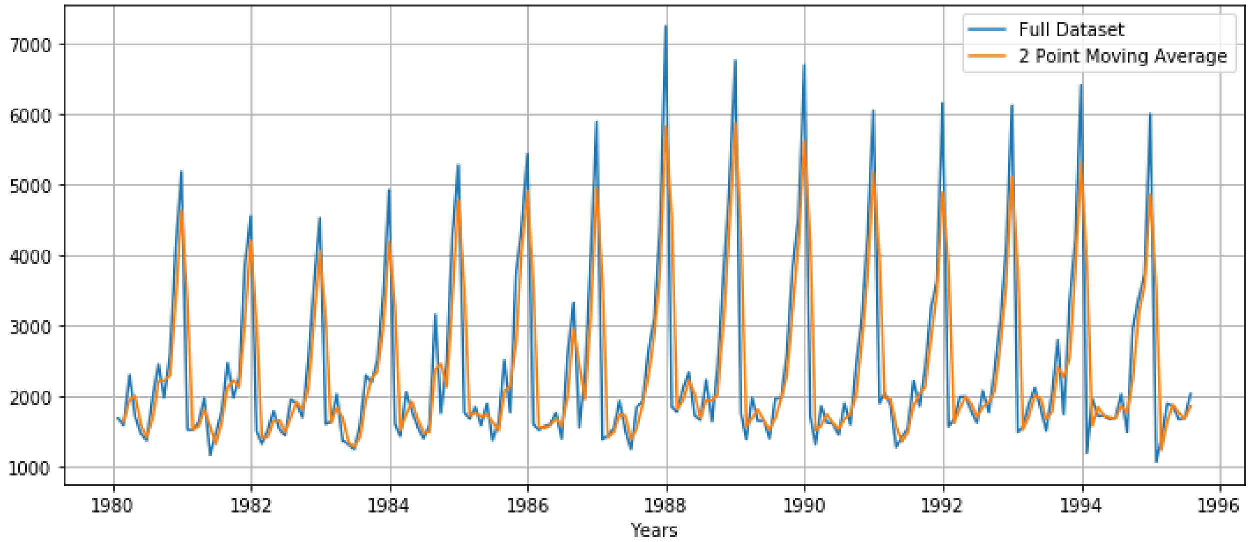


Fig – 1.40 Plotting Moving average data for rolling 2 point

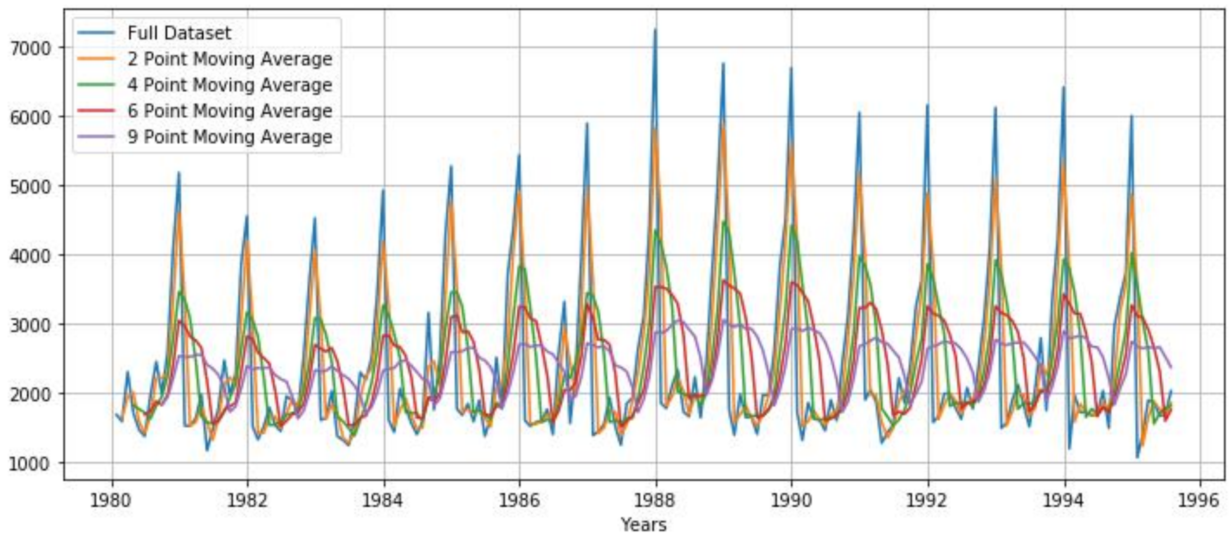


Fig – 1.41 Plotting Moving average data for rolling 2,4,6,9 point

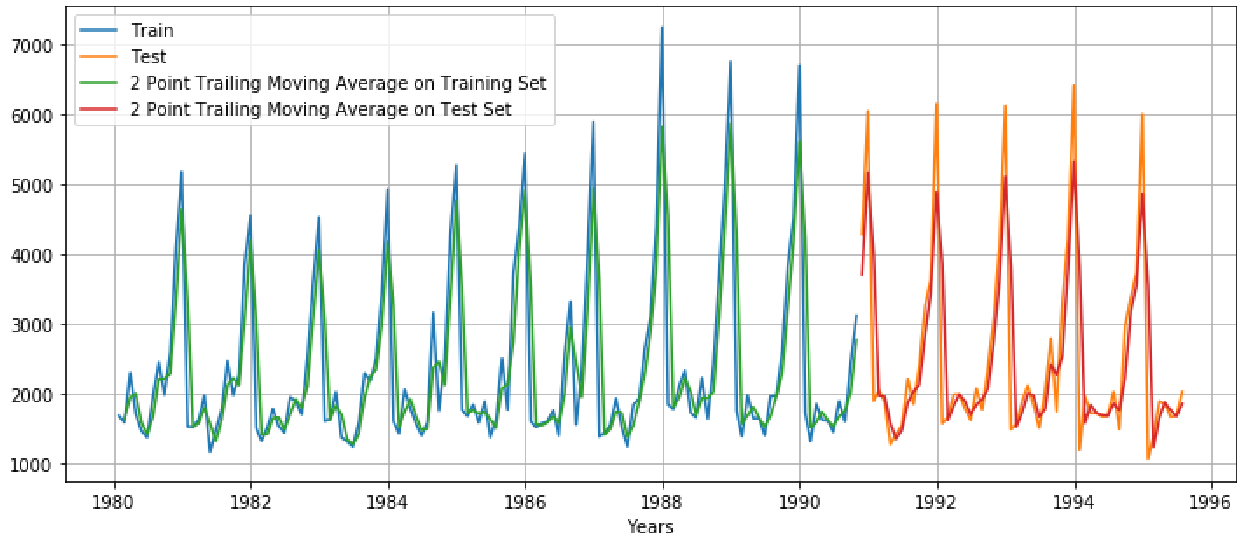


Fig – 1.42 Plotting Moving average data for rolling 2 point for training and test data

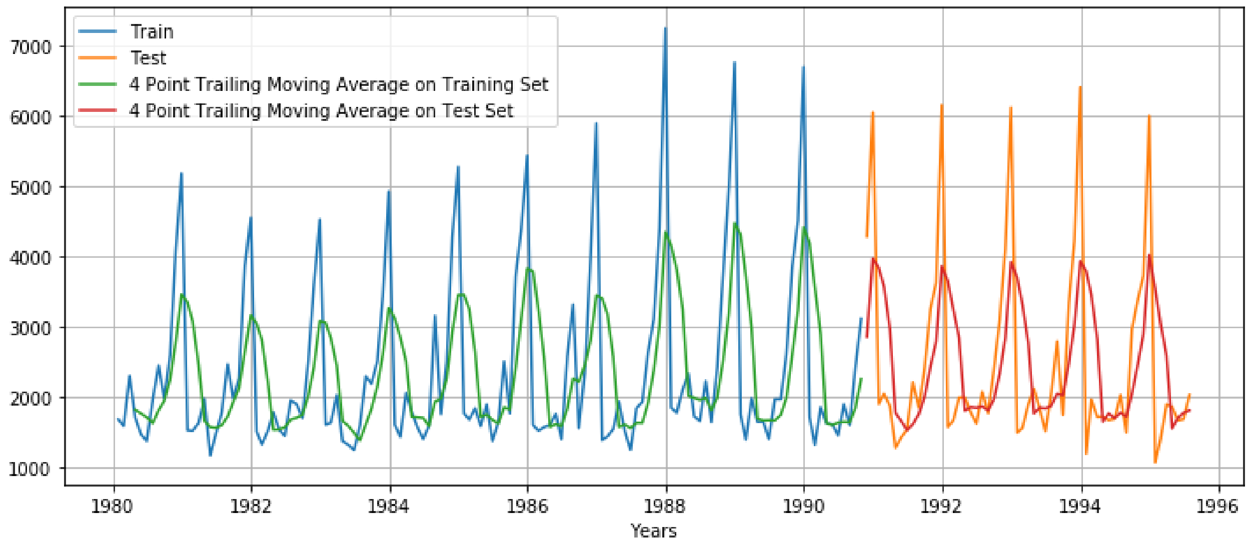


Fig – 1.43 Plotting Moving average data for rolling 4 point for training and test data

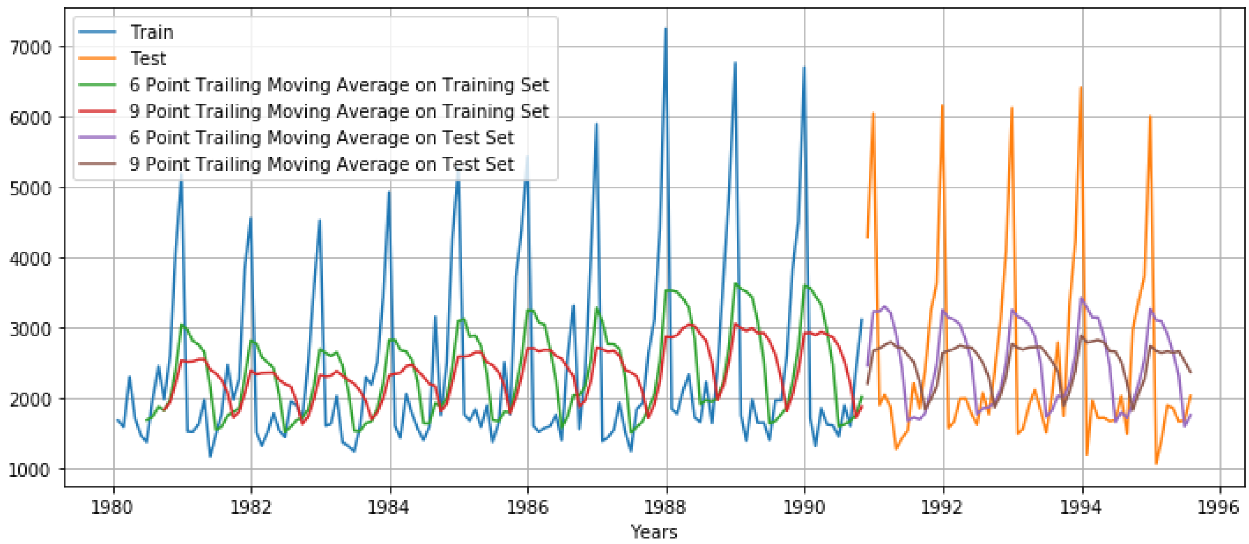


Fig – 1.44 Plotting Moving average data for rolling 6,9 point for training and test data

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

Fig – 1.45 load Moving average data for rolling 2,4,6,9 into the dataframe

```
For 2 point Moving Average Model forecast on the Testing Data,  RMSE is 813.401 and MAPE is 19.70
For 4 point Moving Average Model forecast on the Testing Data,  RMSE is 1156.590 and MAPE is 35.96
For 6 point Moving Average Model forecast on the Testing Data,  RMSE is 1283.927 and MAPE is 43.86
For 9 point Moving Average Model forecast on the Testing Data,  RMSE is 1346.278 and MAPE is 46.86
```

Fig – 1.46 RMSE and MAPE Moving average data for rolling 2,4,6,9

	Test RMSE	Test MAPE
Regression On Time	1389.135175	50.15
NaiveModel	1327.156057	32.90
SimpleAverageModel	1275.073380	38.81
2pointTrailingMovingAverage	813.400684	19.70
4pointTrailingMovingAverage	1156.589694	35.96
6pointTrailingMovingAverage	1283.927428	43.86
9pointTrailingMovingAverage	1346.278315	46.86

Fig – 1.47 Loading MA data for rolling 2,4,6,9 of RMSE and MAPE value Test data into dataframe

Model – 5 Simple Exponential Smoothing:

```
{'smoothing_level': 0.0,
 'smoothing_slope': nan,
 'smoothing_seasonal': nan,
 'damping_slope': nan,
 'initial_level': 2403.7786728672204,
 'initial_slope': nan,
 'initial_seasons': array([], dtype=float64),
 'use_boxcox': False,
 'lamda': None,
 'remove_bias': False}
```

Fig – 1.48 Intializing the Simple Exponential Smoothing

```
== Brown Simple Exponential Smoothing ETS (A, N, N) Parameters ==

Smoothing Level 0.0
Initial Level 2403.7787
```

YearMonth	Sparkling	predict
1991-01-31	1902	2403.778673
1991-02-28	2049	2403.778673
1991-03-31	1874	2403.778673
1991-04-30	1279	2403.778673
1991-05-31	1432	2403.778673

Fig – 1.49 predicting values for the Simple Exponential Smoothing

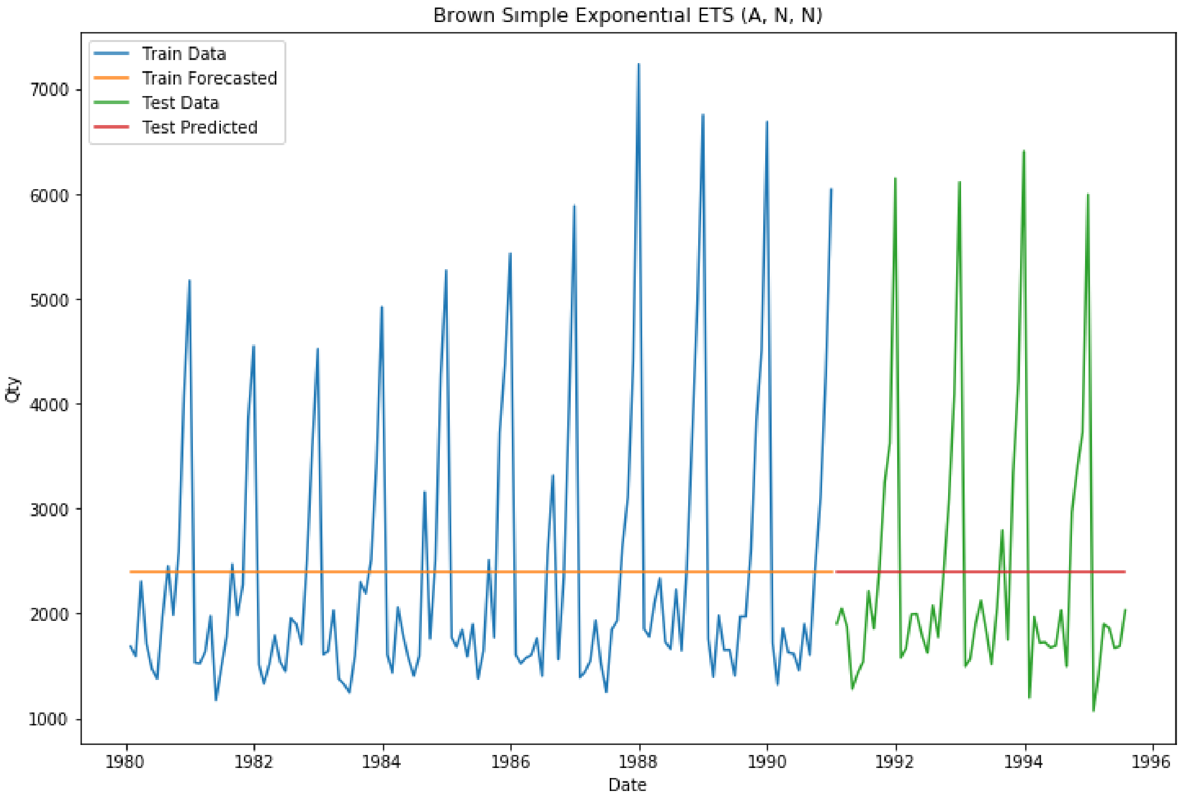


Fig – 1.50 Plotting predicted values for the Simple Exponential Smoothing

```
For Alpha =0 Simple Exponential Smoothing Model forecast on the Training Data,  RMSE is 1298.484 and MAPE is 40.36
```

Fig – 1.51 RMSE and MAPE value for the training data using Simple Exponential Smoothing

```
For Alpha =0 Simple Exponential Smoothing Model forecast on the Testing Data,  RMSE is 1275.082 and MAPE is 38.90
```

Fig – 1.52 RMSE and MAPE value for the test data using Simple Exponential Smoothing

	Test RMSE	Test MAPE
Regression On Time	1389.135175	50.15
NaiveModel	1327.156057	32.90
SimpleAverageModel	1275.073380	38.81
2pointTrailingMovingAverage	813.400684	19.70
4pointTrailingMovingAverage	1156.589694	35.96
6pointTrailingMovingAverage	1283.927428	43.86
9pointTrailingMovingAverage	1346.278315	46.86
Alpha=0,SimpleExponentialSmoothing	1275.081798	38.90

Fig – 1.53 RMSE and MAPE value for the test data using Simple Exponential Smoothing

Setting different alpha values.

Remember, the higher the alpha value more weightage is given to the more recent observation. That means, what happened recently will happen again. We will run a loop with different alpha values to understand which particular value works best for alpha on the test set.

	Alpha Values	Train RMSE	Test RMSE
0	0.3	1.848272e+06	1.625834e+06
1	0.4	1.829497e+06	1.625834e+06
2	0.5	1.806348e+06	1.625834e+06
3	0.6	1.792400e+06	1.625834e+06
4	0.7	1.792504e+06	1.625834e+06
5	0.8	1.807578e+06	1.625834e+06
6	0.9	1.837986e+06	1.625834e+06

Fig – 1.54 RMSE and MAPE value for different Alpha value using Simple Exponential Smoothing

Model 6: Double Exponential Smoothing (Holt's Model):

```
{'smoothing_level': 0.01,
 'smoothing_slope': 0.01,
 'smoothing_seasonal': nan,
 'damping_slope': nan,
 'initial_level': 2015.010536631007,
 'initial_slope': 5.807032214943793,
 'initial_seasons': array([], dtype=float64),
 'use_boxcox': False,
 'lamda': None,
 'remove_bias': False}
```

Fig – 1.55 Initializing the Double Exponential Smoothing

```
== Brown Double Exponential Smoothing ETS (A, A, N) Parameters ==

Smoothing Level 0.01
Initial Level 2015.0105
```

YearMonth	Sparkling	predict
1991-01-31	1902	2789.014686
1991-02-28	2049	2794.850613
1991-03-31	1874	2800.686540
1991-04-30	1279	2806.522467
1991-05-31	1432	2812.358394

Fig – 1.56 Predicting the values using Double Exponential Smoothing

```
Train RMSE is -> 1285.9448360849685
Train MAPE is -> 40.38
```

Fig – 1.57 RMSE and MAPE for training data

```
Test RMSE is -> 1388.1298948092965
Test MAPE is -> 50.09
```

Fig – 1.58 RMSE and MAPE for test data

	Test RMSE	Test MAPE
Regression On Time	1389.135175	50.15
NaiveModel	1327.156057	32.90
SimpleAverageModel	1275.073380	38.81
2pointTrailingMovingAverage	813.400684	19.70
4pointTrailingMovingAverage	1156.589694	35.96
6pointTrailingMovingAverage	1283.927428	43.86
9pointTrailingMovingAverage	1346.278315	46.86
Alpha=0,SimpleExponentialSmoothing	1275.081798	38.90
Alpha=0.99,beta=0.01,DoubleExponentialSmoothing	1388.129895	50.09

Fig – 1.59 RMSE and MAPE for test data into the dataframe

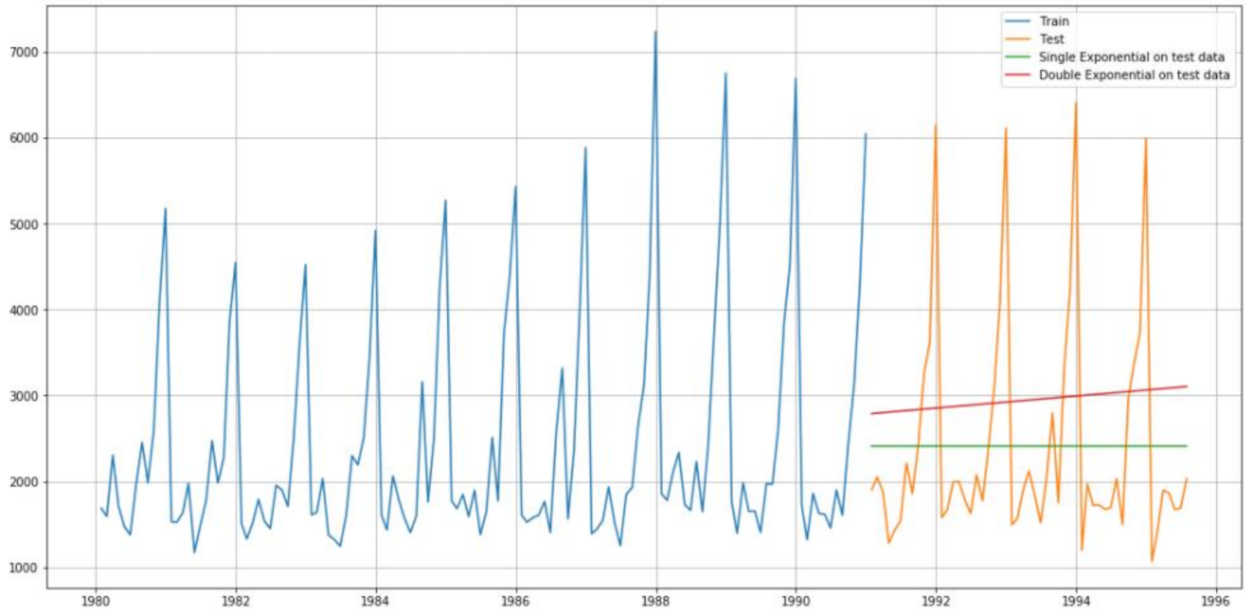


Fig – 1.60 plotting DES predicted output

	Alpha Values	Beta Values	Train RMSE	Test RMSE	Test MAPE	Train MAPE
11	0.1	0.1	1382.520870	1778.564670	67.20	44.37
60	0.6	0.0	1353.081728	2004.479267	68.02	40.13
50	0.5	0.0	1363.627617	2046.410581	68.38	41.54
70	0.7	0.0	1349.696523	2024.922482	69.00	39.15
80	0.8	0.0	1352.968870	2080.557083	70.59	38.48

Fig – 1.61 Finding RMSE and MAPE for different Alpha and beta values

	Test RMSE	Test MAPE
Regression On Time	1389.135175	50.15
NaiveModel	1327.156057	32.90
SimpleAverageModel	1275.073380	38.81
2pointTrailingMovingAverage	813.400684	19.70
4pointTrailingMovingAverage	1156.589694	35.96
6pointTrailingMovingAverage	1283.927428	43.86
9pointTrailingMovingAverage	1346.278315	46.86
Alpha=0,SimpleExponentialSmoothing	1275.081798	38.90
Alpha=0.99,beta=0.01,DoubleExponentialSmoothing	1388.129895	50.09
Alpha=0.1,Beta=0.1,DoubleExponentialSmoothingWithGrid	1778.564670	67.20

Fig 1.62 Finding the least RMSE and MAPE value from the different ALPHA and BETA values

Model 7: Triple Exponential Smoothing (Holt - Winter's Model)

Three parameters α , β and γ are estimated in this model. Level, Trend and Seasonality are accounted for in this model.

```
{'smoothing_level': 0.01,
 'smoothing_slope': 0.01,
 'smoothing_seasonal': 0.01,
 'damping_slope': nan,
 'initial_level': 2063.3631779423836,
 'initial_slope': 5.779628125188699,
 'initial_seasons': array([0.8356073 , 0.96222868, 1.1301695 ]),
 'use_boxcox': False,
 'lamda': None,
 'remove_bias': False}
```

Fig 1.63 Initializing the TES

YearMonth	Sparkling	predict
1980-01-31	1686	1728.990835
1980-02-29	1591	1996.049879
1980-03-31	2304	2346.147908
1980-04-30	1712	1738.701518
1980-05-31	1471	2003.795846

Fig 1.64 Predicting the values from training data

YearMonth	Sparkling	predict
1991-01-31	1902	2387.844924
1991-02-28	2049	2755.798116
1991-03-31	1874	3195.403080
1991-04-30	1279	2402.611531
1991-05-31	1432	2772.805115

Fig 1.65 Predicting the values from test data

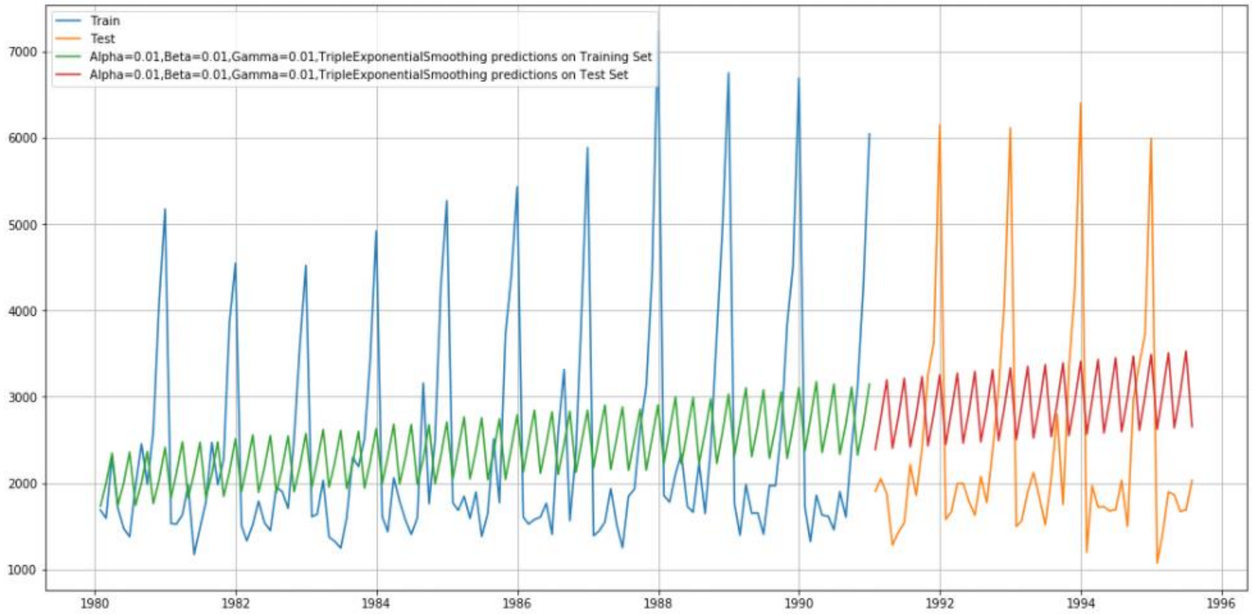


Fig 1.66 Plotting predicted train and test values

```
For Alpha: 0.01,Beta: 0.01 and Gamma:0.01, Triple Exponential Smoothing Model forecast on the Training Data,  RMSE is 1257.887
MAPE is 40.80
```

Fig 1.67 RMSE and MAPE values from train data

```
For Alpha: 0.01,Beta: 0.01 and Gamma:0.01,Triple Exponential Smoothing Model forecast on the Test Data,  RMSE is 1339.779 MAPE
is 49.34
```

Fig 1.68 RMSE and MAPE values from test data

	Test RMSE	Test MAPE
Regression On Time	1389.135175	50.15
NaiveModel	1327.156057	32.90
SimpleAverageModel	1275.073380	38.81
2pointTrailingMovingAverage	813.400684	19.70
4pointTrailingMovingAverage	1156.589694	35.96
6pointTrailingMovingAverage	1283.927428	43.86
9pointTrailingMovingAverage	1346.278315	46.86
Alpha=0,SimpleExponentialSmoothing	1275.081798	38.90
Alpha=0.99,beta=0.01,DoubleExponentialSmoothing	1388.129895	50.09
Alpha=0.1,Beta=0.1,DoubleExponentialSmoothingWithGrid	1778.564670	67.20
Alpha=0.01,Beta=0.01,Gamma=0.01,TripleExponentialSmoothing	1339.779367	49.34

Fig 1.69 RMSE and MAPE for test data into the dataframe

	Alpha Values	Beta Values	Gamma Values	Train RMSE	Train MAPE	Test RMSE	Test MAPE
41	0.01	0.41	0.11	1482.180431	47.82	1285.793066	45.65
70	0.01	0.71	0.01	1429.984221	46.58	1291.252161	45.55
91	0.01	0.91	0.11	1402.523316	45.67	1297.968447	46.32
200	0.21	0.01	0.01	1317.798492	42.15	1305.466942	44.92
81	0.01	0.81	0.11	1402.417016	45.57	1306.915859	47.08

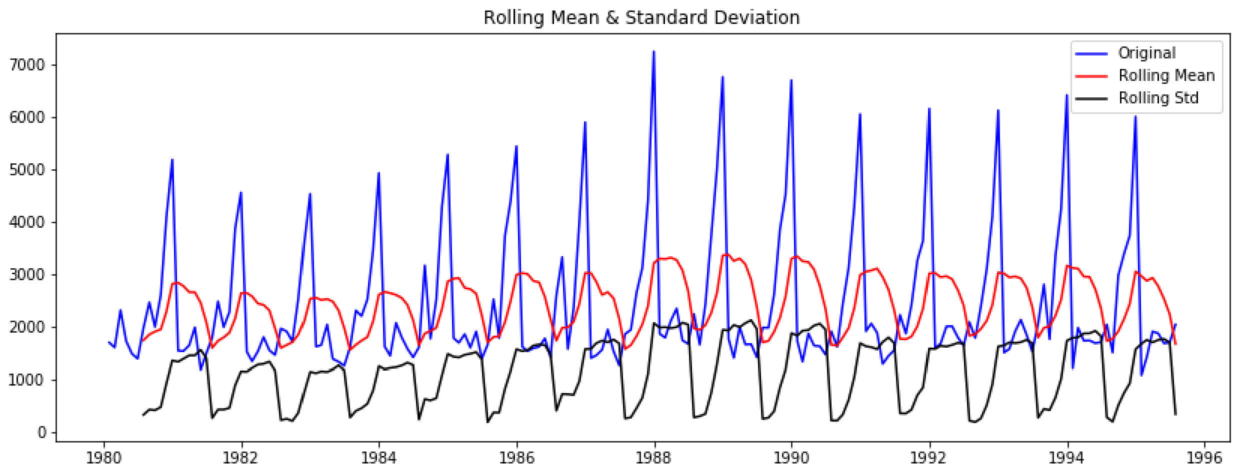
Fig 1.70 RMSE and MAPE for different alpha, beta, and gamma values into the dataframe

	Test RMSE	Test MAPE
Regression On Time	1389.135175	50.15
NaiveModel	1327.156057	32.90
SimpleAverageModel	1275.073380	38.81
2pointTrailingMovingAverage	813.400684	19.70
4pointTrailingMovingAverage	1156.589694	35.96
6pointTrailingMovingAverage	1283.927428	43.86
9pointTrailingMovingAverage	1346.278315	46.86
Alpha=0,SimpleExponentialSmoothing	1275.081798	38.90
Alpha=0.99,beta=0.01,DoubleExponentialSmoothing	1388.129895	50.09
Alpha=0.1,Beta=0.1,DoubleExponentialSmoothingWithGrid	1778.564670	67.20
Alpha=0.01,Beta=0.01,Gamma=0.01,TripleExponentialSmoothing	1339.779367	49.34
Alpha=0.03,Beta=0.03,Gamma=0.03,TripleExponentialSmoothingWithGrid	1285.793066	34.90

Fig 1.71 Finding least RMSE and MAPE for different alpha, beta, and gamma values and loading into the dataframe

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.

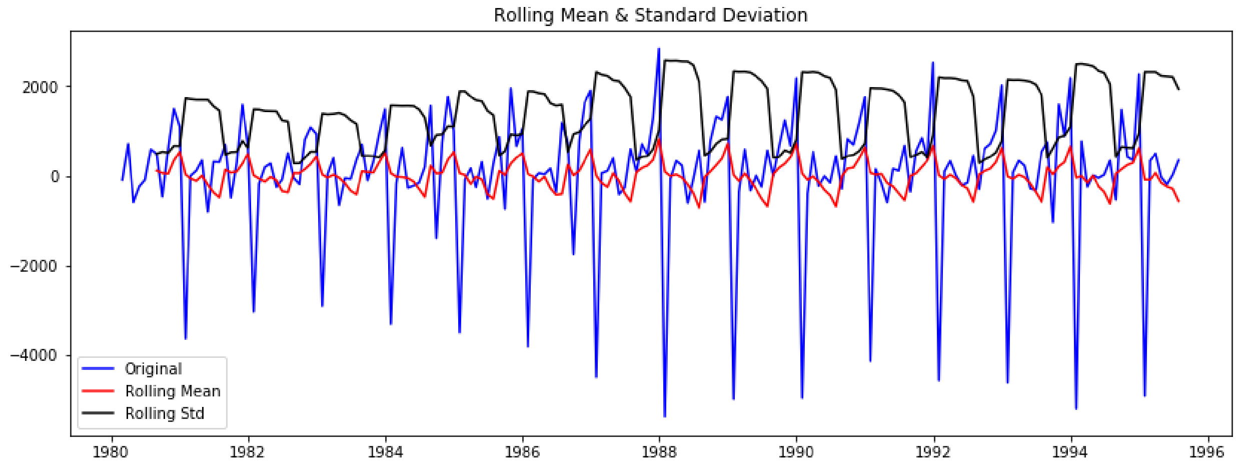
Note: Stationarity should be checked at alpha(α) = 0.05.
Null Hypothesis: p-value > α (alpha value) - then the data is not stationary
Alternate Hypothesis: p-value < α (alpha value) - then the data is having stationarity




```
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
```

Fig 1.72 Finding the p-value and plotting Rolling mean and standard deviation with original data.

The p-value is greater than the alpha value so the data is not stationarity and alternate hypothesis is rejected. To find the stationarity we need to take the 1st difference and plotting the graph and finding the p-value.




```
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
```

Fig 1.73 Finding the p-value and plotting Rolling mean and standard deviation with original data after taking 1st difference data.

Now p-value is less than alpha value. Therefore null hypothesis is rejected.

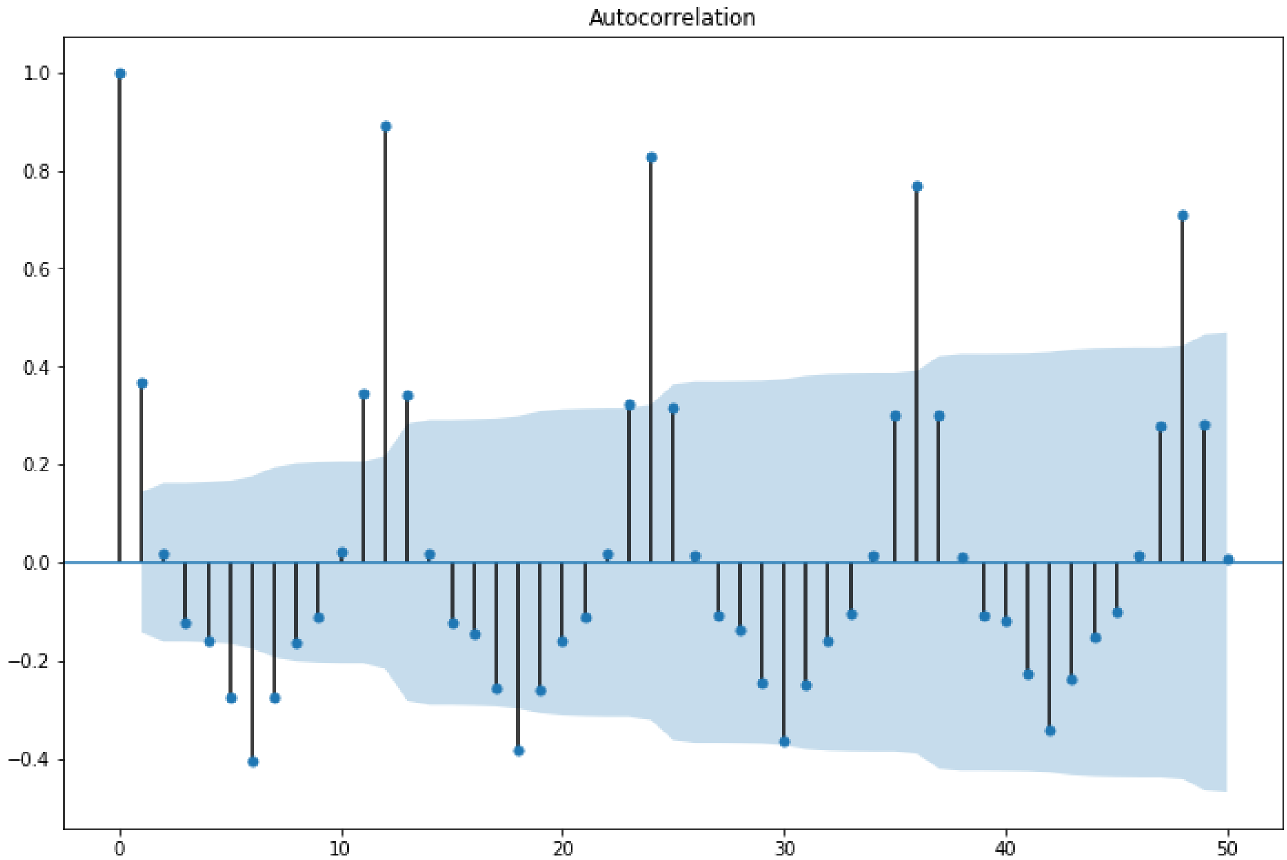


Fig 1.74 ACF plot for original data

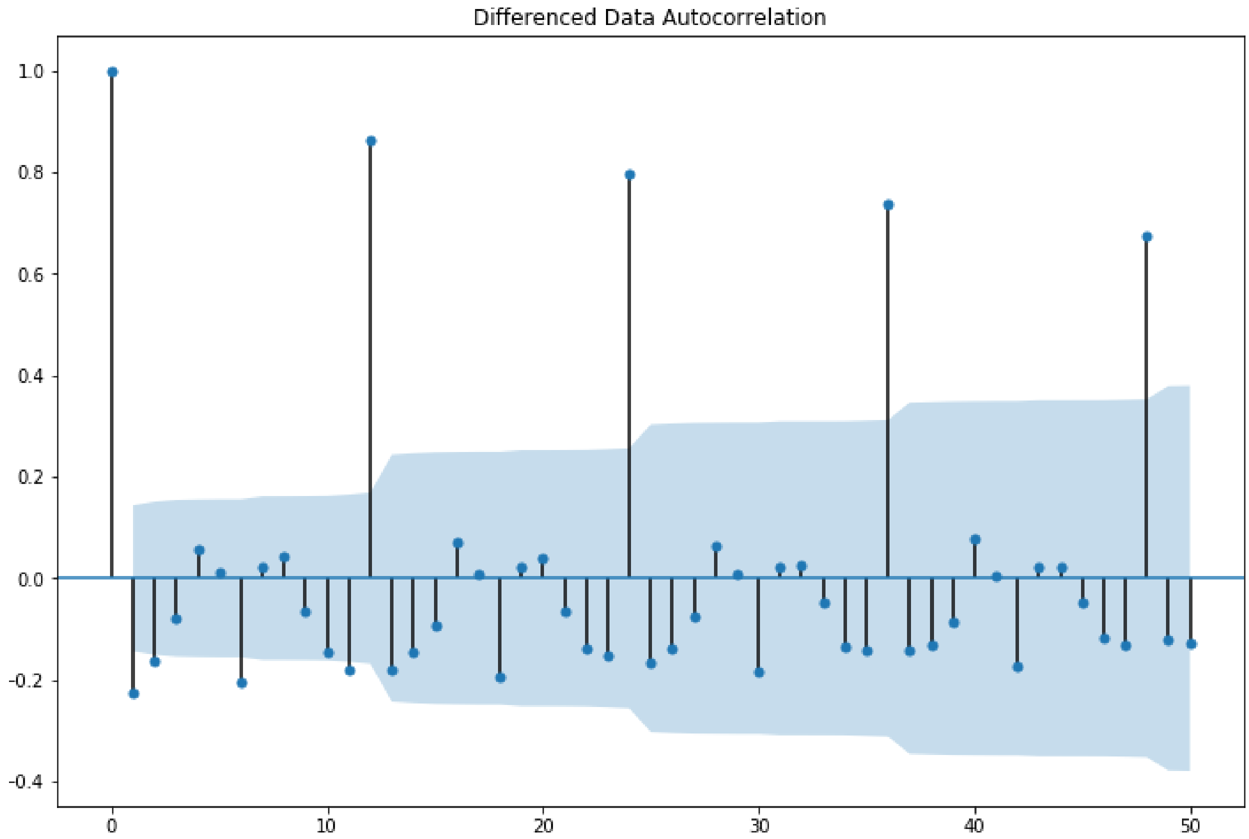


Fig 1.75 ACF plot for 1st difference data

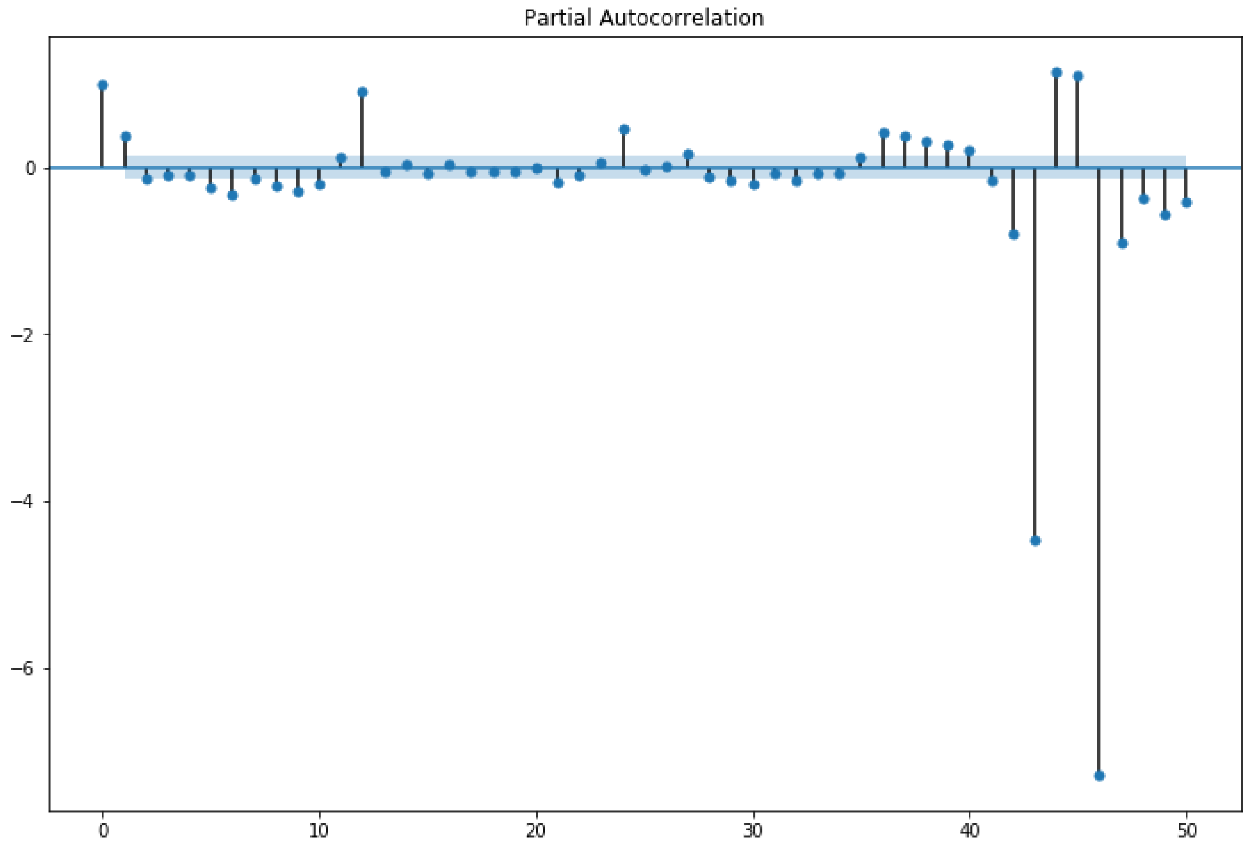


Fig 1.76 PACF plot for original data

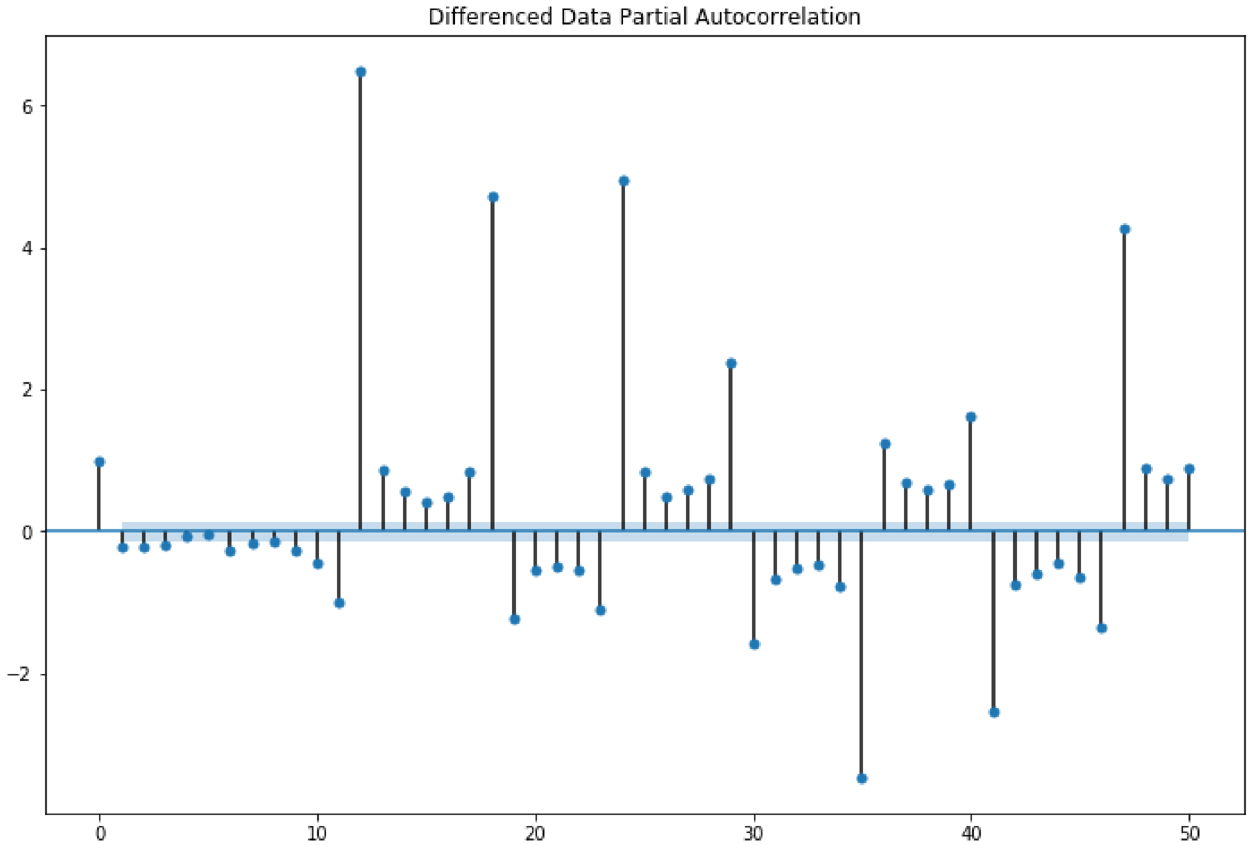
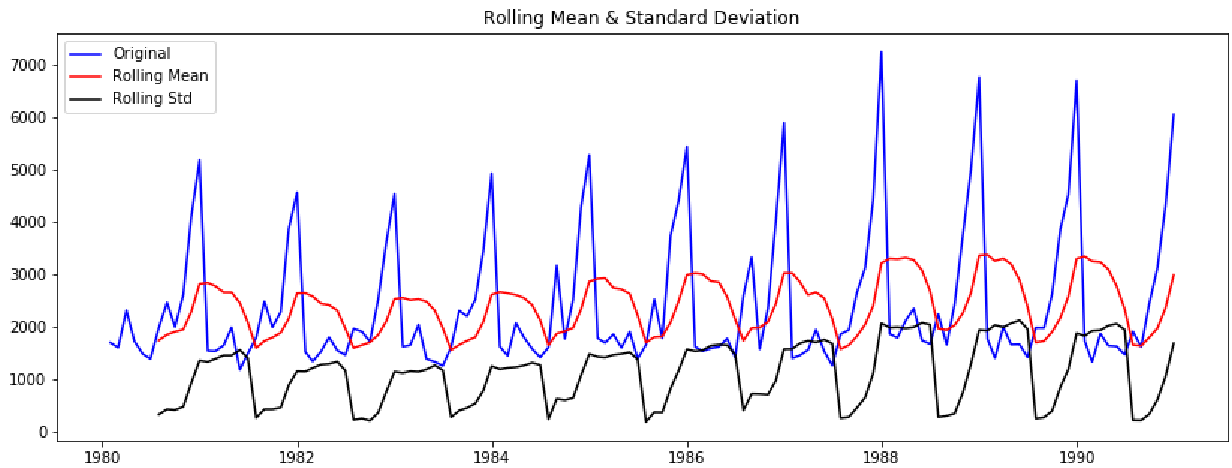


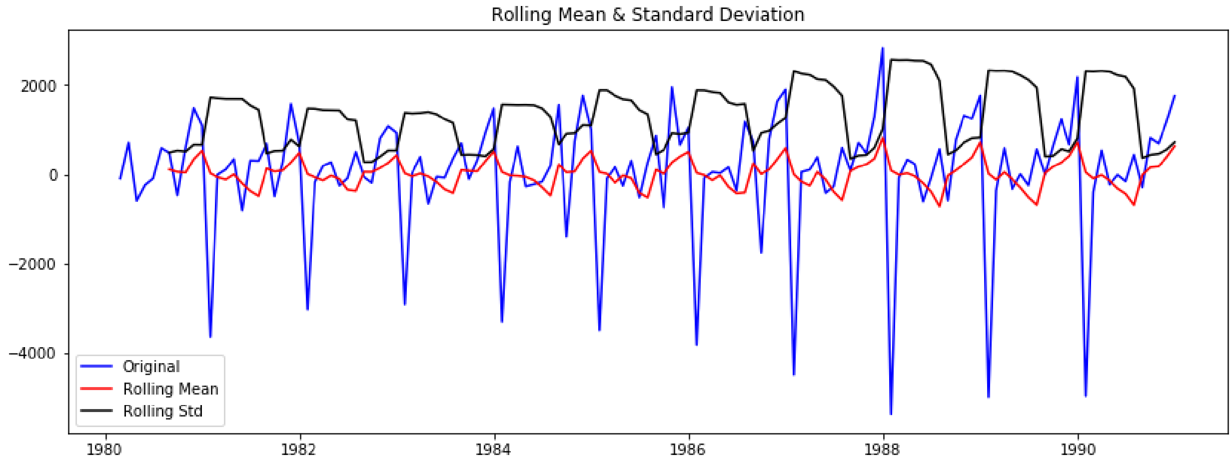
Fig 1.77 PACF plot for 1st difference data




```
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
dtype: float64
```

Fig 1.78 Finding stationarity and p-value for train data

The p-value is greater than the alpha value so the data is not stationarity and alternate hypothesis is rejected. To find the stationarity we need to take the 1st difference and plotting the graph and finding the p-value.




```
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
dtype: float64
```

Fig 1.79 Finding stationarity and p-value for train data after taking 1st difference

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)
```

Fig 1.80 Finding combination for the model

```
ARIMA(0, 1, 0) - AIC:2269.582796371201
ARIMA(0, 1, 1) - AIC:2264.906436793315
ARIMA(0, 1, 2) - AIC:2232.7830976841055
ARIMA(0, 1, 3) - AIC:2233.0166051371466
ARIMA(0, 1, 4) - AIC:2233.8017181420228
ARIMA(1, 1, 0) - AIC:2268.5280606338
ARIMA(1, 1, 1) - AIC:2235.0139453521197
ARIMA(1, 1, 2) - AIC:2233.597647119164
ARIMA(1, 1, 3) - AIC:2234.5741415468874
ARIMA(2, 1, 0) - AIC:2262.0356001527553
ARIMA(2, 1, 1) - AIC:2232.3604898985964
```

Fig 1.81 Sample AIC value for ARIMA Model

	param	AIC
11	(2, 1, 2)	2210.619331
16	(3, 1, 3)	2225.661559
17	(3, 1, 4)	2226.054856
15	(3, 1, 2)	2228.927287
12	(2, 1, 3)	2229.358094

Fig 1.82 Sorting least AIC value for ARIMA Model

```
                             ARIMA Model Results
==============================================================================
Dep. Variable:            D.Sparkling   No. Observations:                  131
Model:                 ARIMA(2, 1, 2)   Log Likelihood               -1099.310
Method:                       css-mle   S.D. of innovations           1012.858
Date:                Sat, 19 Feb 2022   AIC                           2210.619
Time:                        23:57:41   BIC                           2227.871
Sample:                    02-29-1980   HQIC                          2217.629
                         - 12-31-1990
=====================================================================================
                        coef    std err          z      P>|z|      [0.025      0.975]
-------------------------------------------------------------------------------------
const                 5.5854      0.517     10.803      0.000       4.572       6.599
ar.L1.D.Sparkling     1.2699      0.075     17.045      0.000       1.124       1.416
ar.L2.D.Sparkling    -0.5602      0.074     -7.618      0.000      -0.704      -0.416
ma.L1.D.Sparkling    -1.9975      0.042    -47.122      0.000      -2.081      -1.914
ma.L2.D.Sparkling     0.9975      0.042     23.507      0.000       0.914       1.081
                                    Roots
=============================================================================
                  Real          Imaginary           Modulus         Frequency
-----------------------------------------------------------------------------
AR.1            1.1335           -0.7074j            1.3361           -0.0888
AR.2            1.1335           +0.7074j            1.3361            0.0888
MA.1            1.0002           +0.0000j            1.0002            0.0000
MA.2            1.0023           +0.0000j            1.0023            0.0000
-----------------------------------------------------------------------------
```

Fig 1.83 Summary report for ARIMA Model

```
Test rmse for arima is  1374.5293608813515
Test mape for arima is  48.36
```

Fig 1.84 RMSE and MAPE for ARIMA Model

	Test RMSE	Test MAPE
Regression On Time	1389.135175	50.15
NaiveModel	1327.156057	32.90
SimpleAverageModel	1275.073380	38.81
2pointTrailingMovingAverage	813.400684	19.70
4pointTrailingMovingAverage	1156.589694	35.96
6pointTrailingMovingAverage	1283.927428	43.86
9pointTrailingMovingAverage	1346.278315	46.86
Alpha=0,SimpleExponentialSmoothing	1275.081798	38.90
Alpha=0.99,beta=0.01,DoubleExponentialSmoothing	1388.129895	50.09
Alpha=0.1,Beta=0.1,DoubleExponentialSmoothingWithGrid	1778.564670	67.20
Alpha=0.01,Beta=0.01,Gamma=0.01,TripleExponentialSmoothing	1339.779367	49.34
Alpha=0.03,Beta=0.03,Gamma=0.03,TripleExponentialSmoothingWithGrid	1285.793066	34.90
Arima 2,1,2	1374.529361	48.36

Fig 1.85 RMSE and MAPE for ARIMA Model into the dataframe

```
Examples of some parameter combinations for Model...
Model: (0, 1, 1)(0, 0, 1, 12)
Model: (0, 1, 2)(0, 0, 2, 12)
Model: (0, 1, 3)(0, 0, 3, 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: (1, 1, 3)(1, 0, 3, 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)
Model: (2, 1, 3)(2, 0, 3, 12)
Model: (3, 1, 0)(3, 0, 0, 12)
Model: (3, 1, 1)(3, 0, 1, 12)
Model: (3, 1, 2)(3, 0, 2, 12)
Model: (3, 1, 3)(3, 0, 3, 12)
```

Fig 1.86 combination parameters for SARIMA Model

```
SARIMA(0, 1, 0)x(0, 0, 0, 12)7 - AIC:2251.3597196862966
SARIMA(0, 1, 0)x(0, 0, 1, 12)7 - AIC:1956.2614616843744
SARIMA(0, 1, 0)x(0, 0, 2, 12)7 - AIC:1723.1533640237703
```

```
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\b
failed to converge. Check mle_retvals
  "Check mle_retvals", ConvergenceWarning)
```

```
SARIMA(0, 1, 0)x(0, 0, 3, 12)7 - AIC:3825.745688723207
SARIMA(0, 1, 0)x(1, 0, 0, 12)7 - AIC:1837.4366022456675
SARIMA(0, 1, 0)x(1, 0, 1, 12)7 - AIC:1806.99053013884
SARIMA(0, 1, 0)x(1, 0, 2, 12)7 - AIC:1633.2108735792217
```

Fig 1.87 Sample AIC for SARIMA Model

	param	seasonal	AIC
220	(3, 1, 1)	(3, 0, 0, 12)	1387.788331
237	(3, 1, 2)	(3, 0, 1, 12)	1388.602622
221	(3, 1, 1)	(3, 0, 1, 12)	1388.681484
222	(3, 1, 1)	(3, 0, 2, 12)	1389.195910
238	(3, 1, 2)	(3, 0, 2, 12)	1389.701996

Fig 1.88 Sample AIC for SARIMA Model

```
                                     Statespace Model Results
==========================================================================================
Dep. Variable:                                  y   No. Observations:                  132
Model:             SARIMAX(3, 1, 1)x(3, 0, 0, 12)   Log Likelihood                -685.894
Date:                            Sun, 20 Feb 2022   AIC                           1387.788
Time:                                    00:15:19   BIC                           1407.963
Sample:                                         0   HQIC                          1395.931
                                            - 132
Covariance Type:                              opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.1615      0.150      1.075      0.282      -0.133       0.456
ar.L2         -0.0928      0.150     -0.618      0.537      -0.388       0.202
ar.L3          0.0916      0.136      0.676      0.499      -0.174       0.357
ma.L1         -0.9195      0.092    -10.033      0.000      -1.099      -0.740
ar.S.L12       0.5805      0.104      5.575      0.000       0.376       0.785
ar.S.L24       0.2559      0.119      2.159      0.031       0.024       0.488
ar.S.L36       0.2132      0.121      1.761      0.078      -0.024       0.451
sigma2      1.729e+05   2.18e+04      7.940      0.000     1.3e+05    2.16e+05
===================================================================================
Ljung-Box (Q):                       31.39   Jarque-Bera (JB):                18.78
Prob(Q):                              0.83   Prob(JB):                         0.00
Heteroskedasticity (H):               1.08   Skew:                             0.47
Prob(H) (two-sided):                  0.84   Kurtosis:                         5.00
===================================================================================

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

Fig 1.89 Summary report for SARIMA Model

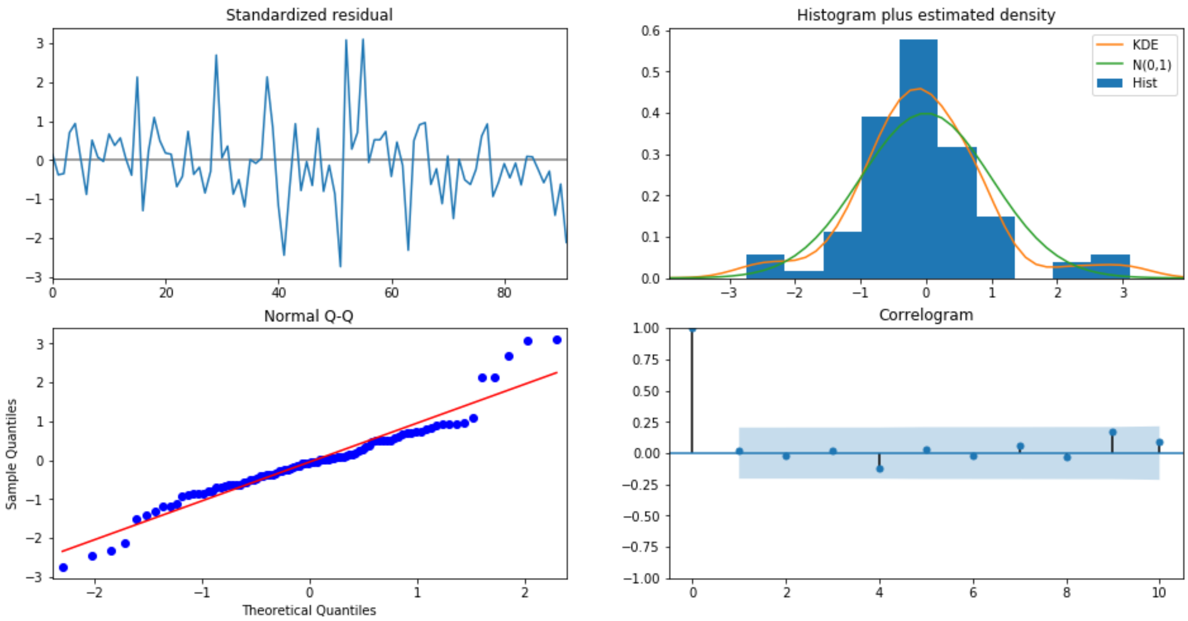


Fig 1.90 plotting diagnostic for SARIMA Model

```
array([1389.35414447, 1224.67522784, 1673.34099324, 1533.30689195,
       1425.95174345, 1250.61621936, 1772.2263109 , 1479.1141852 ,
       2283.75791693, 3282.83196143, 4399.46839163, 6356.70652653,
       1314.29569047, 1039.22158027, 1562.6974022 , 1351.51410197,
       1286.72525869, 1091.64429923, 1627.98242717, 1381.98114896,
       2195.05961235, 3215.80153732, 4305.96847966, 6357.47526074,
       1178.25441918,  891.37159589, 1424.74088645, 1217.0725402 ,
       1149.22551055,  957.43755915, 1496.46611318, 1215.98706344,
       2068.45691365, 3064.13249883, 4232.11118064, 6299.20175084,
       1009.579547  ,  737.56140227, 1276.76541294, 1072.32991558,
        993.4792034 ,  794.85408648, 1356.18814625, 1067.93952173,
       1942.37115052, 2994.51917732, 4189.51258734, 6331.61300314,
        860.85887789,  570.90769989, 1131.98182785,  915.15083558,
        838.20706093,  632.24546072, 1210.35571855])
```

Fig 1.91 Predicting values for testing data

```
Test rmse for sarima is  601.2393388459468
Test mape for sarima is  21.17
```

Fig 1.92 RMSE and MAPE SARIMA Model

	Test RMSE	Test MAPE
Regression On Time	1389.135175	50.15
NaiveModel	1327.156057	32.90
SimpleAverageModel	1275.073380	38.81
2pointTrailingMovingAverage	813.400684	19.70
4pointTrailingMovingAverage	1156.589694	35.96
6pointTrailingMovingAverage	1283.927428	43.86
9pointTrailingMovingAverage	1346.278315	46.86
Alpha=0,SimpleExponentialSmoothing	1275.081798	38.90
Alpha=0.99,beta=0.01,DoubleExponentialSmoothing	1388.129895	50.09
Alpha=0.1,Beta=0.1,DoubleExponentialSmoothingWithGrid	1778.564670	67.20
Alpha=0.01,Beta=0.01,Gamma=0.01,TripleExponentialSmoothing	1339.779367	49.34
Alpha=0.03,Beta=0.03,Gamma=0.03,TripleExponentialSmoothingWithGrid	1285.793066	34.90
Arima 2,1,2	1374.529361	48.36
SARIMA(3, 1, 1)(3, 0, 0, 12)	601.239339	21.17

Fig 1.93 Loading RMSE and MAPE SARIMA Model into dataframe

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.

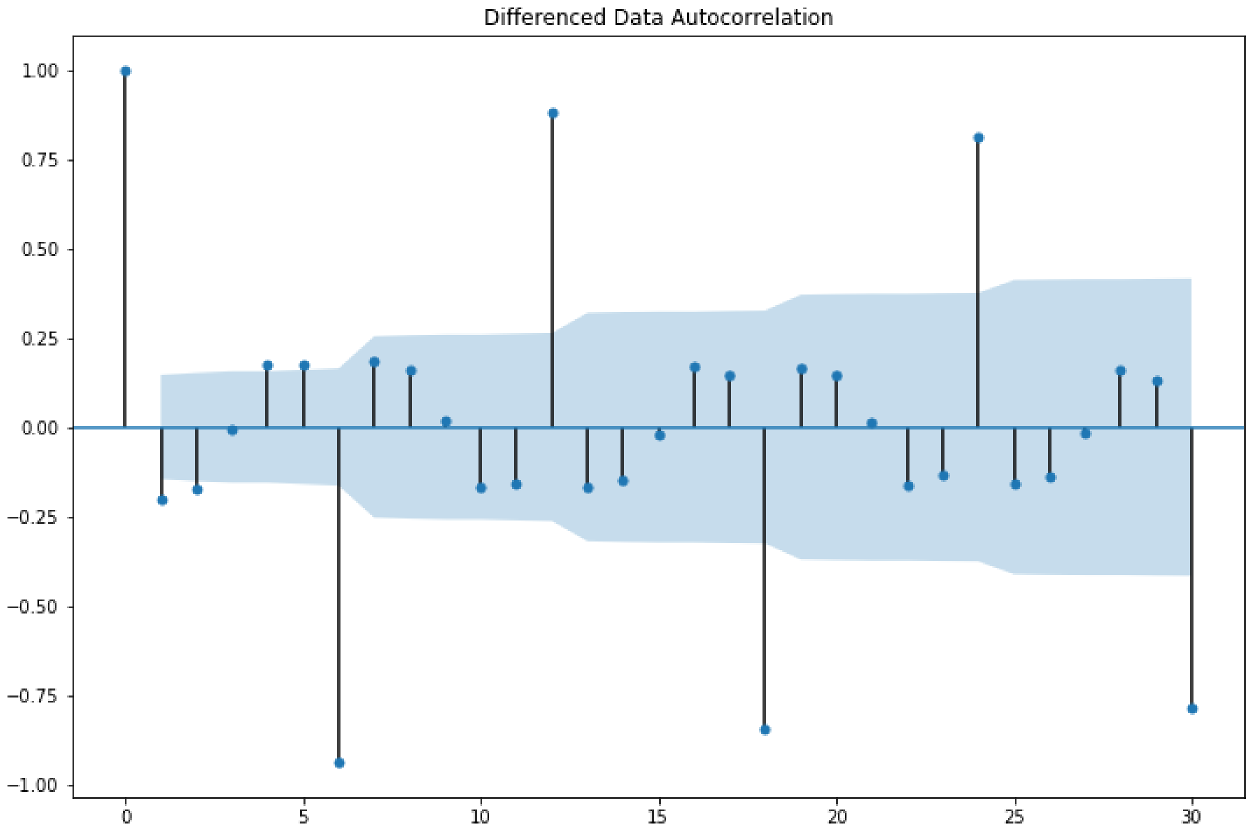


Fig 1.94 Plotting 2nd Difference after taking cut-off points in ACF plot

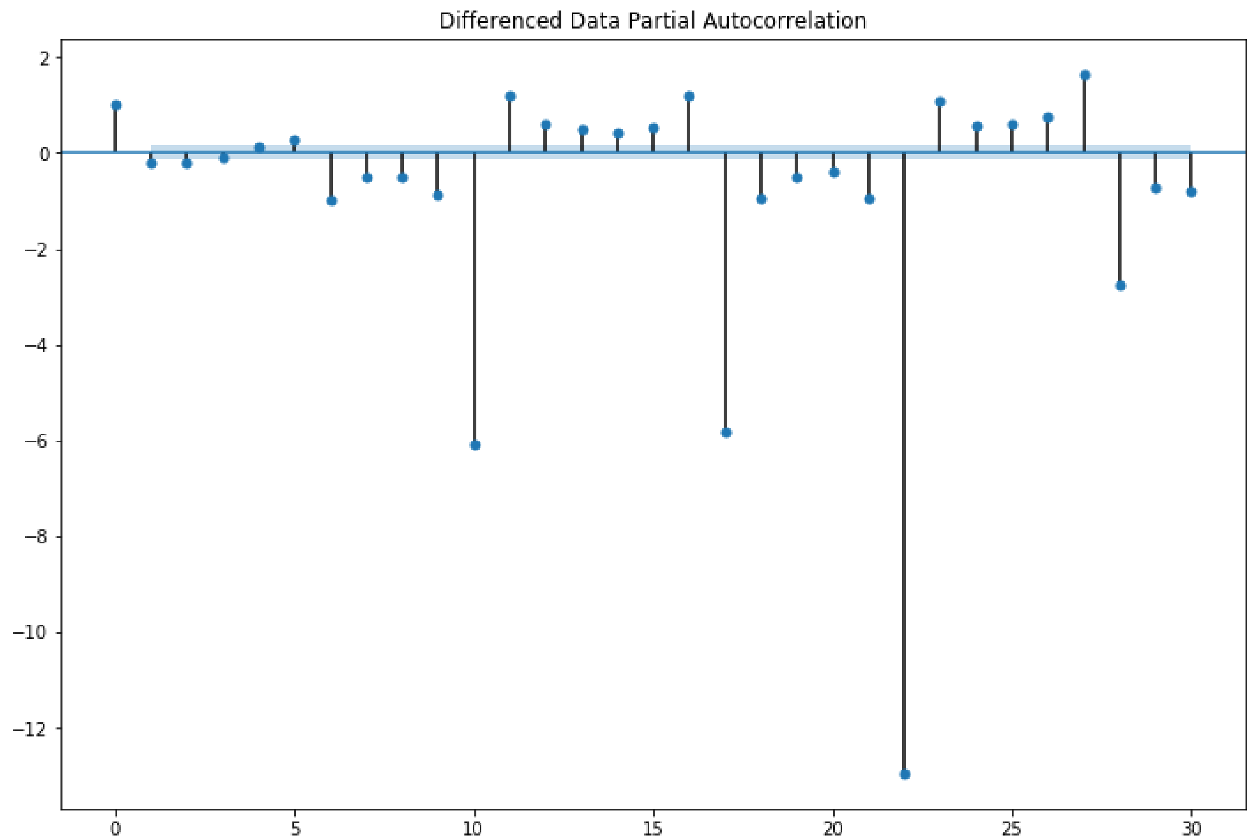


Fig 1.95 Plotting 2nd Difference after taking cut-off points in PACF plot

```
                                     Statespace Model Results
==========================================================================================
Dep. Variable:                                 y   No. Observations:                  132
Model:             SARIMAX(2, 1, 2)x(2, 1, 2, 12)   Log Likelihood                -684.814
Date:                            Sun, 20 Feb 2022   AIC                           1387.628
Time:                                    00:15:27   BIC                           1410.324
Sample:                                         0   HQIC                          1396.788
                                            - 132
Covariance Type:                              opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.5623      0.332     -1.695      0.090      -1.212       0.088
ar.L2          0.0325      0.167      0.195      0.845      -0.294       0.359
ma.L1         -0.1506      0.293     -0.513      0.608      -0.725       0.424
ma.L2         -0.7626      0.259     -2.942      0.003      -1.271      -0.254
ar.S.L12      -0.2455      0.977     -0.251      0.802      -2.161       1.670
ar.S.L24      -0.1093      0.215     -0.508      0.612      -0.532       0.313
ma.S.L12      -0.1621      0.976     -0.166      0.868      -2.076       1.752
ma.S.L24      -0.0210      0.440     -0.048      0.962      -0.884       0.842
sigma2         1.7e+05   2.53e+04      6.731      0.000    1.21e+05     2.2e+05
===================================================================================
Ljung-Box (Q):                       22.34   Jarque-Bera (JB):                 13.51
Prob(Q):                              0.99   Prob(JB):                          0.00
Heteroskedasticity (H):               0.95   Skew:                              0.59
Prob(H) (two-sided):                  0.88   Kurtosis:                          4.46
===================================================================================

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

Fig 1.96 Cut-off point summary report for SARIMA model

```
array([1486.06575329, 1445.53344308, 1799.46261867, 1729.60673392,
       1561.67206174, 1427.44832941, 1904.33414197, 1662.16163031,
       2400.26361382, 3306.87496771, 4386.88262233, 6270.81556737,
       1530.99233787, 1364.77872122, 1790.28242024, 1658.3768708 ,
       1535.5845103 , 1378.97491418, 1867.11698994, 1646.46993087,
       2381.36758199, 3303.22575612, 4345.76908492, 6253.74691244,
       1503.27926337, 1322.66405518, 1753.82565795, 1617.59341923,
       1501.84962867, 1347.40194385, 1829.41893605, 1597.54273625,
       2342.24196208, 3236.79030416, 4298.43490541, 6187.03027829,
       1458.75652912, 1295.40156289, 1717.35883801, 1588.96639931,
       1466.56058802, 1314.02711119, 1796.3188189 , 1564.84578747,
       2307.48920921, 3207.07591929, 4268.1262771 , 6158.85199692,
       1426.29264018, 1260.27404814, 1683.87279367, 1554.02840391,
       1432.48780344, 1279.24798468, 1762.14182877])
```

Fig 1.97 predicting values for SARIMA model after taking cutoff points

```
Test rmse for sarima is  345.93884642937763
Test mape for sarima is  11.21
```

Fig 1.98 RMSE and MAPE value for cutoff point SARIMA model

	Test RMSE	Test MAPE
Regression On Time	1389.135175	50.15
NaiveModel	1327.156057	32.90
SimpleAverageModel	1275.073380	38.81
2pointTrailingMovingAverage	813.400684	19.70
4pointTrailingMovingAverage	1156.589694	35.96
6pointTrailingMovingAverage	1283.927428	43.86
9pointTrailingMovingAverage	1346.278315	46.86
Alpha=0,SimpleExponentialSmoothing	1275.081798	38.90
Alpha=0.99,beta=0.01,DoubleExponentialSmoothing	1388.129895	50.09
Alpha=0.1,Beta=0.1,DoubleExponentialSmoothingWithGrid	1778.564670	67.20
Alpha=0.01,Beta=0.01,Gamma=0.01,TripleExponentialSmoothing	1339.779367	49.34
Alpha=0.03,Beta=0.03,Gamma=0.03,TripleExponentialSmoothingWithGrid	1285.793066	34.90
Arima 2,1,2	1374.529361	48.36
SARIMA(3, 1, 1)(3, 0, 0, 12)	601.239339	21.17
SARIMA(2, 1, 2)(2, 1, 2, 12)	345.938846	11.21

Fig 1.99 Cut-off point for SARIMA model RMSE and MAPE into Dataframe

```
                             ARIMA Model Results
==============================================================================
Dep. Variable:            D.Sparkling   No. Observations:                  131
Model:                 ARIMA(2, 1, 3)   Log Likelihood               -1107.679
Method:                       css-mle   S.D. of innovations           1093.029
Date:                Sun, 20 Feb 2022   AIC                           2229.358
Time:                        00:15:29   BIC                           2249.484
Sample:                    02-29-1980   HQIC                          2237.536
                         - 12-31-1990
=====================================================================================
                        coef    std err          z      P>|z|      [0.025      0.975]
-------------------------------------------------------------------------------------
const                 5.8832      3.397      1.732      0.086      -0.775      12.542
ar.L1.D.Sparkling    -0.8782      0.077    -11.438      0.000      -1.029      -0.728
ar.L2.D.Sparkling    -0.5714      0.077     -7.430      0.000      -0.722      -0.421
ma.L1.D.Sparkling     0.3445      0.028     12.396      0.000       0.290       0.399
ma.L2.D.Sparkling    -0.3445      0.037     -9.401      0.000      -0.416      -0.273
ma.L3.D.Sparkling    -1.0000        nan        nan        nan         nan         nan
                                    Roots
=============================================================================
                  Real          Imaginary           Modulus         Frequency
-----------------------------------------------------------------------------
AR.1           -0.7685           -1.0768j            1.3229           -0.3487
AR.2           -0.7685           +1.0768j            1.3229            0.3487
MA.1            1.0000           -0.0000j            1.0000           -0.0000
MA.2           -0.6723           -0.7403j            1.0000           -0.3673
MA.3           -0.6723           +0.7403j            1.0000            0.3673
-----------------------------------------------------------------------------
```

Fig 1.100 Cut-off point for ARIMA model summary report

```
Test rmse for arima is  1393.6347572218879
Test mape for arima is  49.8
```

Fig 1.101 Cut-off point for ARIMA model RMSE and MAPE value

	Test RMSE	Test MAPE
Regression On Time	1389.135175	50.15
NaiveModel	1327.156057	32.90
SimpleAverageModel	1275.073380	38.81
2pointTrailingMovingAverage	813.400684	19.70
4pointTrailingMovingAverage	1156.589694	35.96
6pointTrailingMovingAverage	1283.927428	43.86
9pointTrailingMovingAverage	1346.278315	46.86
Alpha=0,SimpleExponentialSmoothing	1275.081798	38.90
Alpha=0.99,beta=0.01,DoubleExponentialSmoothing	1388.129895	50.09
Alpha=0.1,Beta=0.1,DoubleExponentialSmoothingWithGrid	1778.564670	67.20
Alpha=0.01,Beta=0.01,Gamma=0.01,TripleExponentialSmoothing	1339.779367	49.34
Alpha=0.03,Beta=0.03,Gamma=0.03,TripleExponentialSmoothingWithGrid	1285.793066	34.90
Arima 2,1,2	1374.529361	48.36
SARIMA(3, 1, 1)(3, 0, 0, 12)	601.239339	21.17
SARIMA(2, 1, 2)(2, 1, 2, 12)	345.938846	11.21
ARIMA(2,1,3)	1393.634757	49.80

Fig 1.102 Cut-off point for ARIMA model RMSE and MAPE into Dataframe

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

	Test RMSE	Test MAPE
SARIMA(2, 1, 2)(2, 1, 2, 12)	345.938846	11.21
SARIMA(3, 1, 1)(3, 0, 0, 12)	601.239339	21.17
2pointTrailingMovingAverage	813.400684	19.70
4pointTrailingMovingAverage	1156.589694	35.96
SimpleAverageModel	1275.073380	38.81
Alpha=0,SimpleExponentialSmoothing	1275.081798	38.90
6pointTrailingMovingAverage	1283.927428	43.86
Alpha=0.03,Beta=0.03,Gamma=0.03,TripleExponentialSmoothingWithGrid	1285.793066	34.90
NaiveModel	1327.156057	32.90
Alpha=0.01,Beta=0.01,Gamma=0.01,TripleExponentialSmoothing	1339.779367	49.34
9pointTrailingMovingAverage	1346.278315	46.86
Arima 2,1,2	1374.529361	48.36
Alpha=0.99,beta=0.01,DoubleExponentialSmoothing	1388.129895	50.09
Regression On Time	1389.135175	50.15
ARIMA(2,1,3)	1393.634757	49.80
Alpha=0.1,Beta=0.1,DoubleExponentialSmoothingWithGrid	1778.564670	67.20

Fig 1.103 Sorting the RMSE value and finding best model

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.

```
                                     Statespace Model Results
==========================================================================================
Dep. Variable:                            Sparkling   No. Observations:                187
Model:             SARIMAX(2, 1, 2)x(2, 1, 2, 12)   Log Likelihood               -1085.726
Date:                              Sun, 20 Feb 2022   AIC                           2189.453
Time:                                      00:16:00   BIC                           2216.367
Sample:                                  01-31-1980   HQIC                          2200.388
                                       - 07-31-1995
Covariance Type:                                opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.6655      0.275     -2.421      0.015      -1.204      -0.127
ar.L2         -0.0111      0.113     -0.099      0.921      -0.232       0.210
ma.L1         -0.1696      0.260     -0.652      0.515      -0.680       0.340
ma.L2         -0.7336      0.242     -3.032      0.002      -1.208      -0.259
ar.S.L12       0.0659      0.370      0.178      0.859      -0.659       0.791
ar.S.L24       0.2880      0.140      2.060      0.039       0.014       0.562
ma.S.L12      -0.7496      0.997     -0.752      0.452      -2.704       1.204
ma.S.L24      -0.2228      0.327     -0.682      0.495      -0.863       0.418
sigma2       1.332e+05   9.91e+04      1.344      0.179   -6.11e+04    3.27e+05
===================================================================================
Ljung-Box (Q):                       17.23   Jarque-Bera (JB):                26.11
Prob(Q):                              1.00   Prob(JB):                         0.00
Heteroskedasticity (H):               0.89   Skew:                             0.56
Prob(H) (two-sided):                  0.68   Kurtosis:                         4.74
===================================================================================

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

Fig 1.104 Summary report for SARIMA model

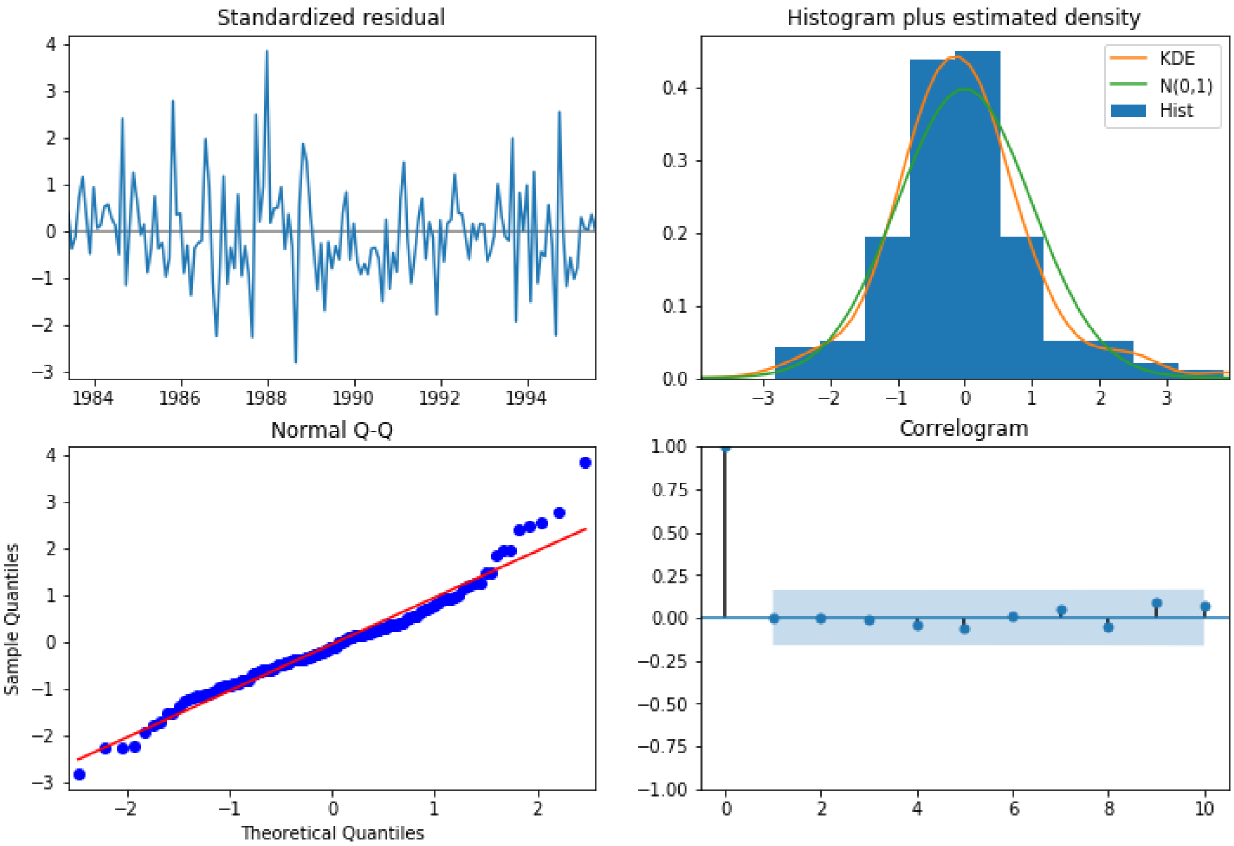


Fig 1.105 plotting diagnostic report for SARIMA model

Predicting the values for the next 12 months in future

Sparkling	mean	mean_se	mean_ci_lower	mean_ci_upper
1995-08-31	2132.776769	373.934654	1399.878314	2865.675223
1995-09-30	2359.398133	379.075631	1616.423549	3102.372717
1995-10-31	3266.882782	379.150939	2523.760596	4010.004967
1995-11-30	4040.248120	381.487465	3292.546427	4787.949812
1995-12-31	6101.119618	381.594674	5353.207802	6849.031435
1996-01-31	1339.521607	382.793219	589.260685	2089.782529
1996-02-29	1627.011523	382.914157	876.513566	2377.509481
1996-03-31	1852.837623	383.722533	1100.755278	2604.919968
1996-04-30	1794.715870	383.981625	1042.125714	2547.306026
1996-05-31	1662.211602	384.678945	908.254724	2416.168481
1996-06-30	1616.363993	385.226882	861.333179	2371.394807
1996-07-31	2025.592032	385.887735	1269.265970	2781.918094

Fig 1.106 Predicting the values

```
RMSE of the Full Model 568.78862913566
```

Fig 1.107 RMSE value for future data.

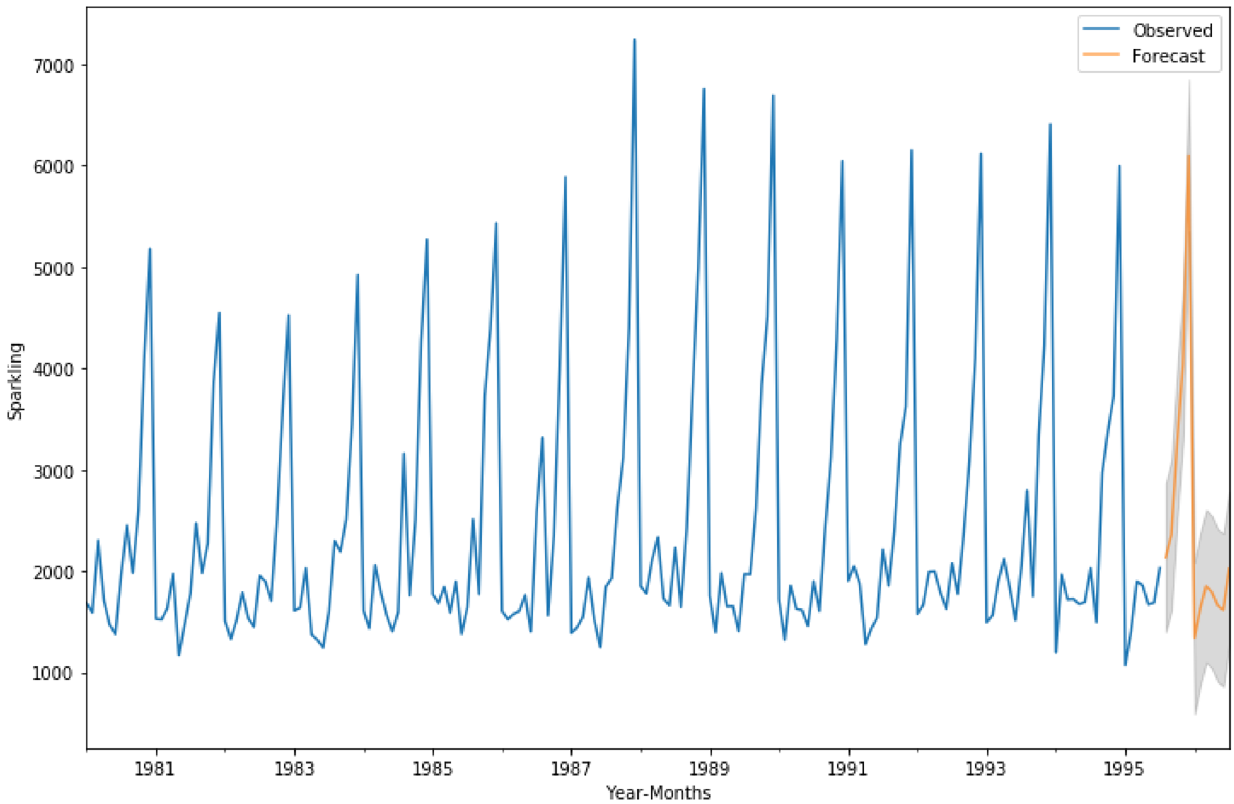


Fig 1.108 Plotting the future data in graph.

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

SARIMA Model is performing best in this case giving us the least error.

Looking at the bar plot, we can see that on December months the sales are highest. We can use these insights to increase our sales further.

We can introduce certain offers in November, December months to attract more customers.

Year 1988 has the highest sales recorded till data. We can go back to find out the reasons to which pushed the sales so much. Looking at the prediction, we can say that the sales figure will be more or less same as that of previous year. Hence some important measures have to be taken to increase the trend. As the trend has been more or less constant throughout the years.