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

Fig 1.1 Dataset Sample

## Exploratory Data Analysis

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

```
YearMonth     object
Rose         float64
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):
Rose    187 non-null float64
dtypes: float64(1)
memory usage: 7.9 KB
```

Fig- 1.3.1. Check null values

| YearMonth | Rose |
|---|---|
| 1994-07-31 | NaN |
| 1994-08-31 | NaN |

Fig- 1.3.2. Checking NaN values

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

| YearMonth | Rose |
|---|---|
| 1980-01-31 | 112.0 |
| 1980-02-29 | 118.0 |
| 1980-03-31 | 129.0 |
| 1980-04-30 | 99.0 |
| 1980-05-31 | 116.0 |

Fig- 1.4. Initialising Date as index Column

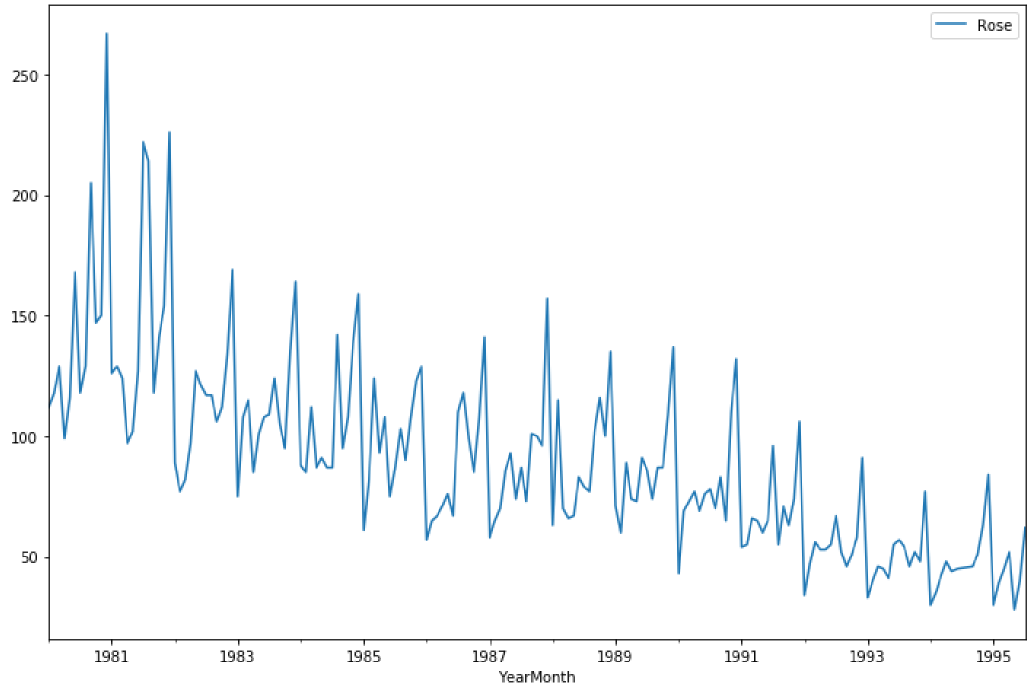



Fig – 1.5 Plotting Rose wine data

The Rose 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.

| | Rose |
|---|---|
| count | 187.000000 |
| mean | 89.914439 |
| std | 39.238325 |
| min | 28.000000 |
| 25% | 62.500000 |
| 50% | 85.000000 |
| 75% | 111.000000 |
| max | 267.000000 |

Fig – 1.6 Rose wine sales data spread.

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

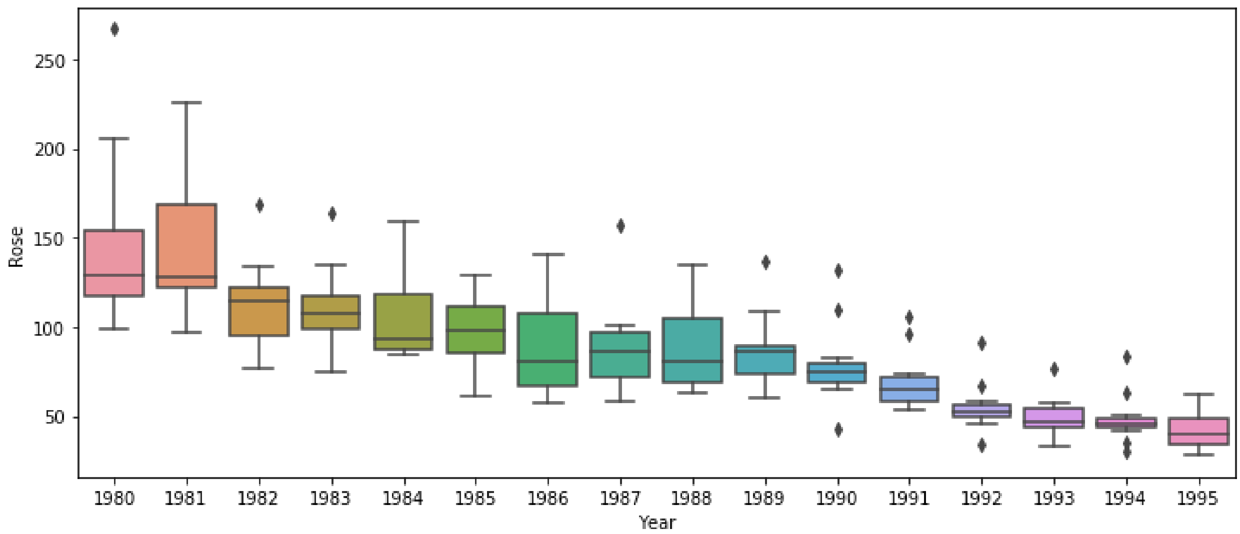



Fig – 1.7 Rose wine sales across years

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

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

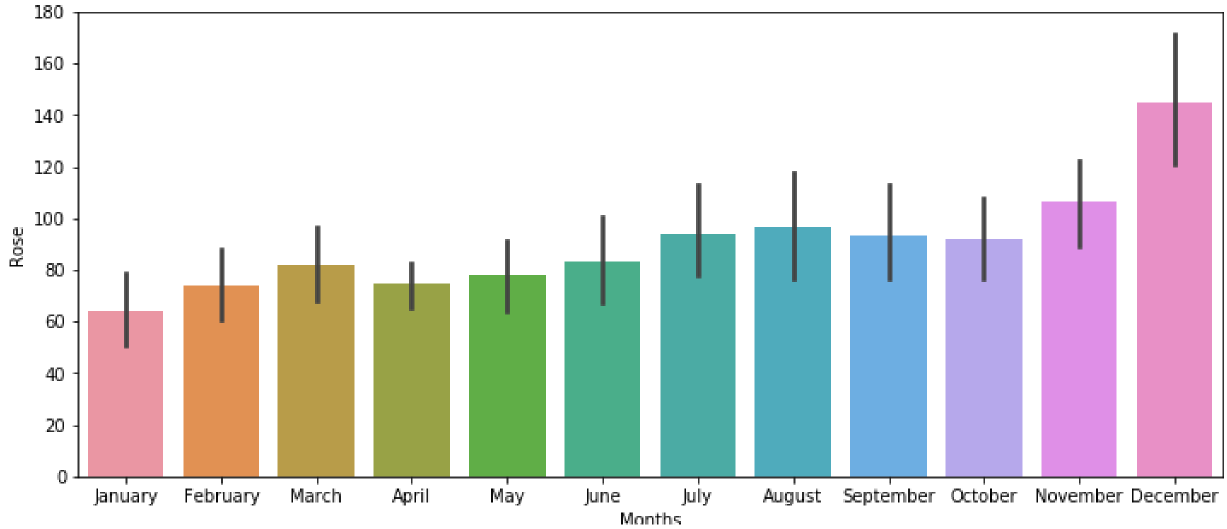



Fig – 1.8 Rose wine sales across months

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

| YearMonth | April | August | December | February | January | July | June | March | May | November | October | September |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **YearMonth** | | | | | | | | | | | | |
| **1980** | 99.0 | 129.000000 | 267.0 | 118.0 | 112.0 | 118.000000 | 168.0 | 129.0 | 116.0 | 150.0 | 147.0 | 205.0 |
| **1981** | 97.0 | 214.000000 | 226.0 | 129.0 | 126.0 | 222.000000 | 127.0 | 124.0 | 102.0 | 154.0 | 141.0 | 118.0 |
| **1982** | 97.0 | 117.000000 | 169.0 | 77.0 | 89.0 | 117.000000 | 121.0 | 82.0 | 127.0 | 134.0 | 112.0 | 106.0 |
| **1983** | 85.0 | 124.000000 | 164.0 | 108.0 | 75.0 | 109.000000 | 108.0 | 115.0 | 101.0 | 135.0 | 95.0 | 105.0 |
| **1984** | 87.0 | 142.000000 | 159.0 | 85.0 | 88.0 | 87.000000 | 87.0 | 112.0 | 91.0 | 139.0 | 108.0 | 95.0 |
| **1985** | 93.0 | 103.000000 | 129.0 | 82.0 | 61.0 | 87.000000 | 75.0 | 124.0 | 108.0 | 123.0 | 108.0 | 90.0 |
| **1986** | 71.0 | 118.000000 | 141.0 | 65.0 | 57.0 | 110.000000 | 67.0 | 67.0 | 76.0 | 107.0 | 85.0 | 99.0 |
| **1987** | 86.0 | 73.000000 | 157.0 | 65.0 | 58.0 | 87.000000 | 74.0 | 70.0 | 93.0 | 96.0 | 100.0 | 101.0 |
| **1988** | 66.0 | 77.000000 | 135.0 | 115.0 | 63.0 | 79.000000 | 83.0 | 70.0 | 67.0 | 100.0 | 116.0 | 102.0 |
| **1989** | 74.0 | 74.000000 | 137.0 | 60.0 | 71.0 | 86.000000 | 91.0 | 89.0 | 73.0 | 109.0 | 87.0 | 87.0 |
| **1990** | 77.0 | 70.000000 | 132.0 | 69.0 | 43.0 | 78.000000 | 76.0 | 73.0 | 69.0 | 110.0 | 65.0 | 83.0 |
| **1991** | 65.0 | 55.000000 | 106.0 | 55.0 | 54.0 | 96.000000 | 65.0 | 66.0 | 60.0 | 74.0 | 63.0 | 71.0 |
| **1992** | 53.0 | 52.000000 | 91.0 | 47.0 | 34.0 | 67.000000 | 55.0 | 56.0 | 53.0 | 58.0 | 51.0 | 46.0 |
| **1993** | 45.0 | 54.000000 | 77.0 | 40.0 | 33.0 | 57.000000 | 55.0 | 46.0 | 41.0 | 48.0 | 52.0 | 46.0 |
| **1994** | 48.0 | 45.666667 | 84.0 | 35.0 | 30.0 | 45.333333 | 45.0 | 42.0 | 44.0 | 63.0 | 51.0 | 46.0 |
| **1995** | 52.0 | NaN | NaN | 39.0 | 30.0 | 62.000000 | 40.0 | 45.0 | 28.0 | NaN | NaN | NaN |

Fig – 1.9 Monthwise wine sales across years

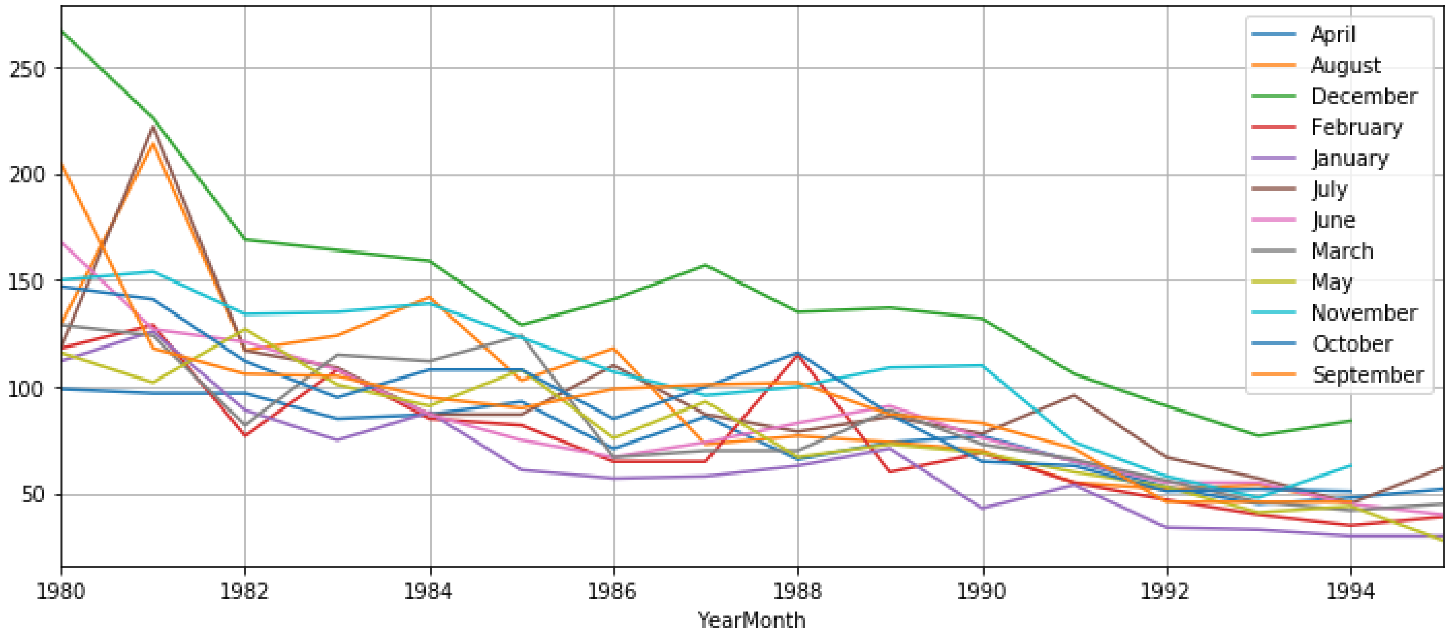



Fig – 1.10 Monthwise wine sales across years

***Additive Model:***

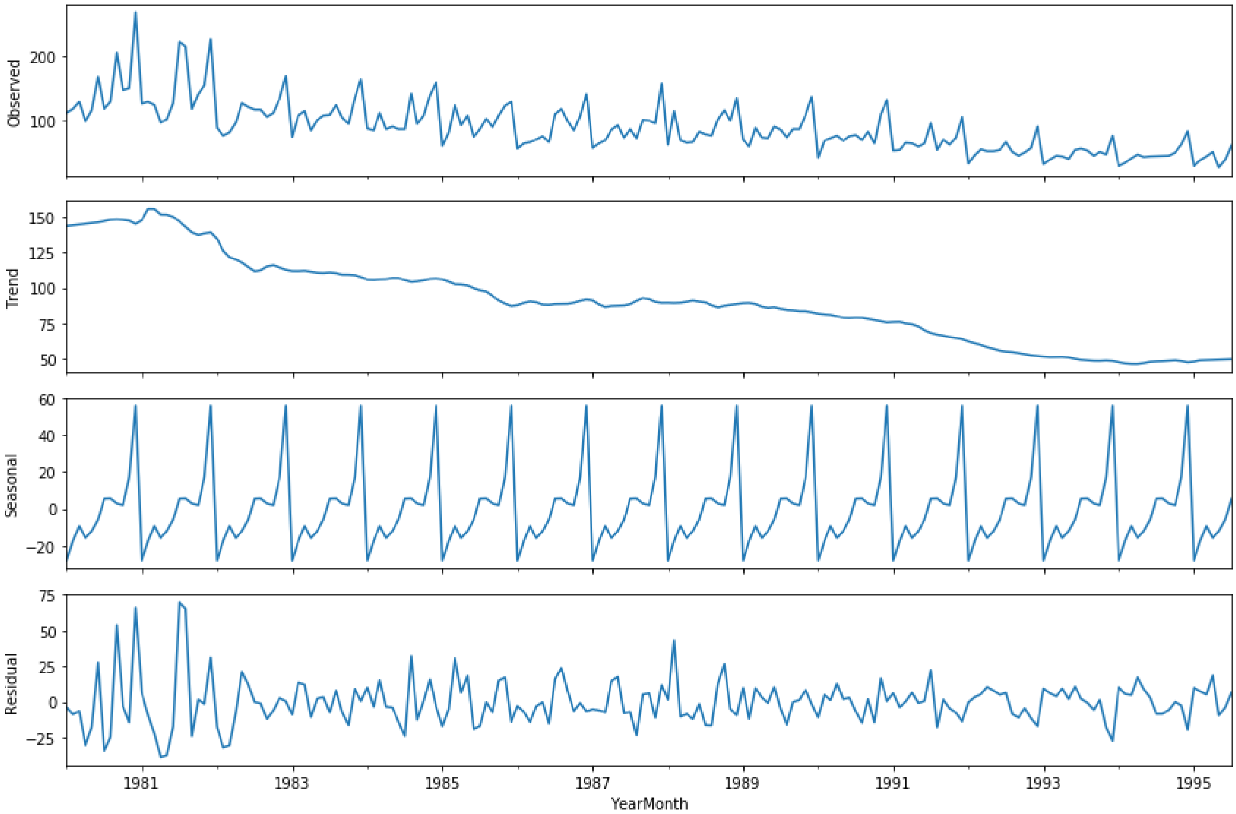



Fig – 1.11 Decompose data form the original dataset for Rose wines (Additive model)

```
Trend
                   Rose
YearMonth
1980-01-31  143.619658
1980-02-29  144.148504
1980-03-31  144.677350
1980-04-30  145.206197
1980-05-31  145.735043

Seasonality
                  Rose
YearMonth
1980-01-31 -28.058855
1980-02-29 -17.428254
1980-03-31  -9.278095
1980-04-30 -15.844951
1980-05-31 -12.036806

Residual
                  Rose
YearMonth
1980-01-31  -3.560804
1980-02-29  -8.720251
1980-03-31  -6.399255
1980-04-30 -30.361246
1980-05-31 -17.698237
```

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

## ***Multiplicative Model:***

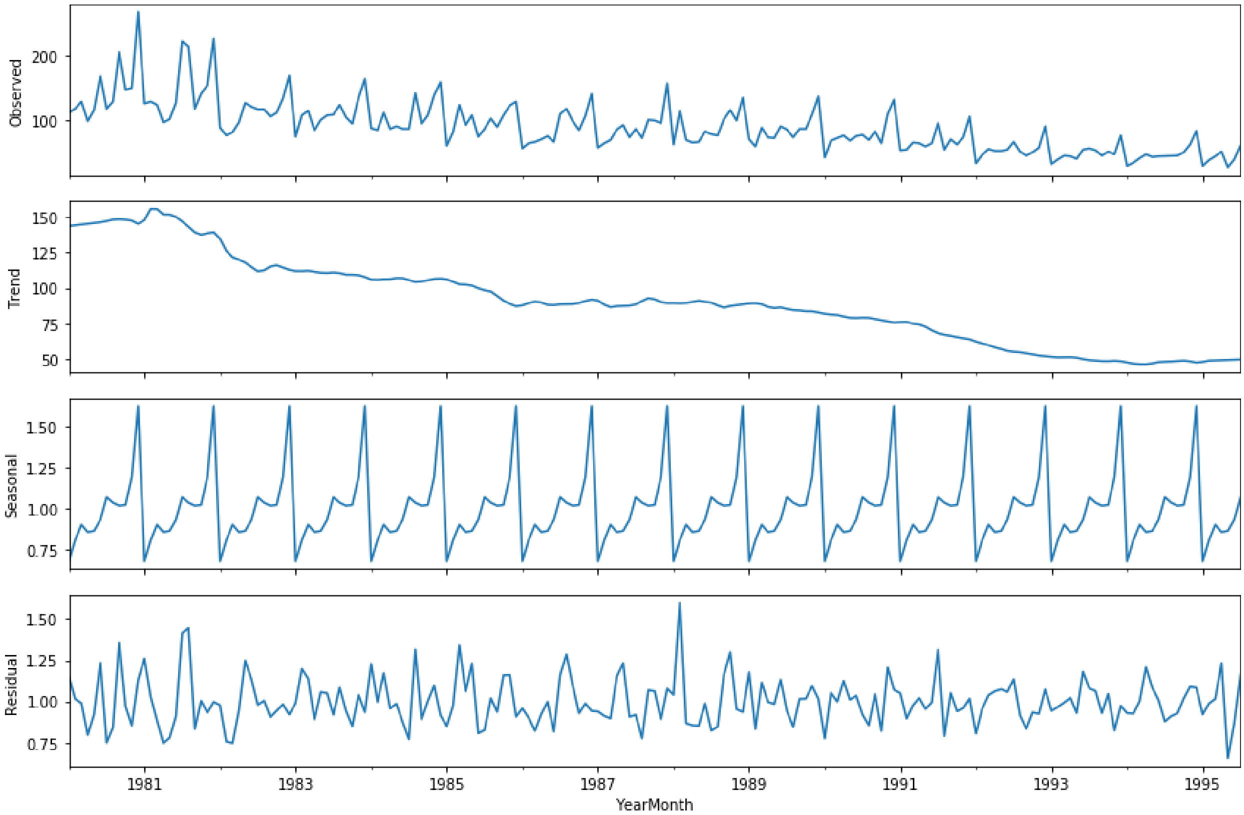



Fig – 1.13 Decompose data form the original dataset for Rose wines (Multiplicative model)

```
Trend
                  Rose
YearMonth
1980-01-31  143.619658
1980-02-29  144.148504
1980-03-31  144.677350
1980-04-30  145.206197
1980-05-31  145.735043

Seasonality
                Rose
YearMonth
1980-01-31  0.676904
1980-02-29  0.806254
1980-03-31  0.901399
1980-04-30  0.855717
1980-05-31  0.863276

Residual
                Rose
YearMonth
1980-01-31  1.152065
1980-02-29  1.015313
1980-03-31  0.989173
1980-04-30  0.796746
1980-05-31  0.922028
```

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

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

| YearMonth | Rose |
|---|---|
| 1990-08-31 | 70.0 |
| 1990-09-30 | 83.0 |
| 1990-10-31 | 65.0 |
| 1990-11-30 | 110.0 |
| 1990-12-31 | 132.0 |

Fig – 1.15 Last 5 values for Training data

| YearMonth | Rose |
|---|---|
| 1991-01-31 | 54.0 |
| 1991-02-28 | 55.0 |
| 1991-03-31 | 66.0 |
| 1991-04-30 | 65.0 |
| 1991-05-31 | 60.0 |

Fig – 1.16 First 5 values for testing data

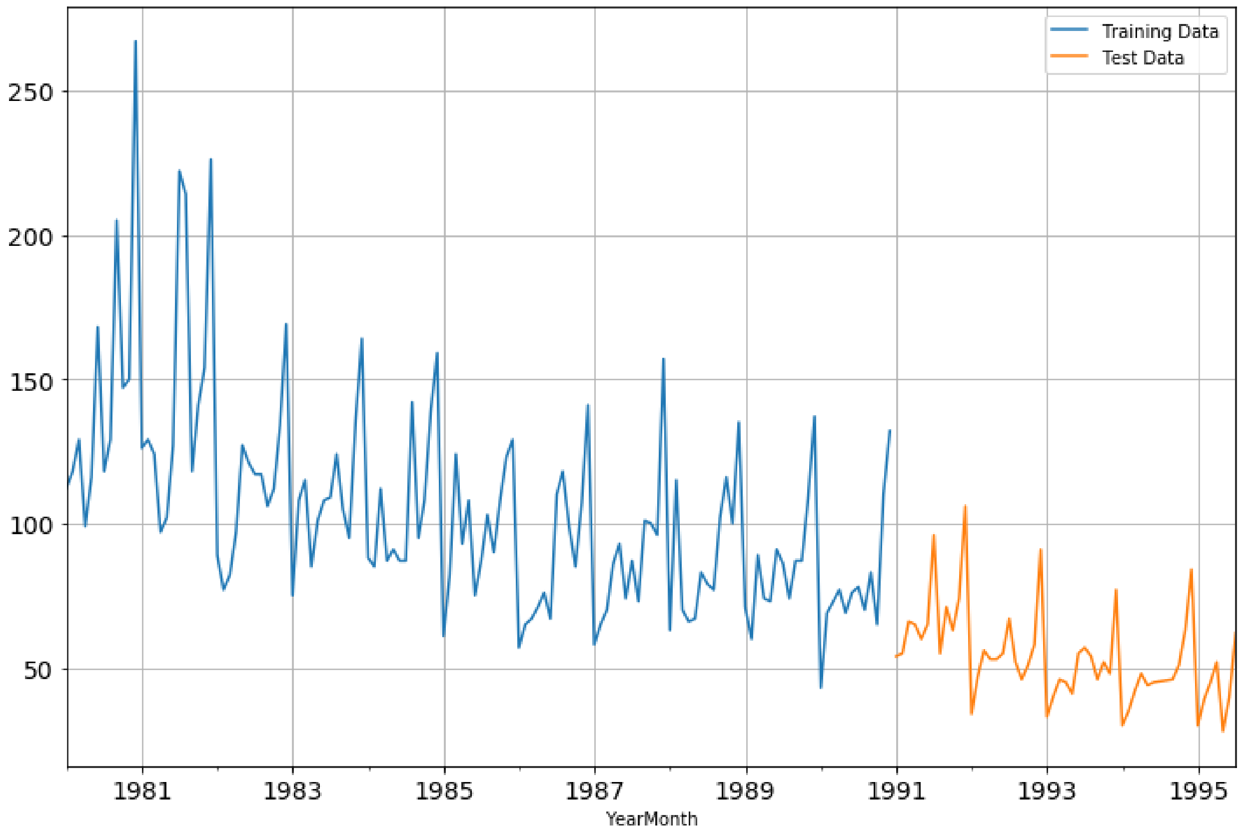


Fig – 1.17 Plotting Training and test dataset for Rose wines

**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:***

```
The train time are  [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, 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, 12
0, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132]

The test time are  [133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 15
4, 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, 183, 184, 185, 186, 187]
```

Fig – 1.18.1 Generating numerical time instance for both training and test dataset for Rose 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.

```
             Rose  train_time
YearMonth
1980-01-31  112.0           1
1980-02-29  118.0           2
1980-03-31  129.0           3
1980-04-30   99.0           4
1980-05-31  116.0           5
            Rose  test_time
YearMonth
1991-01-31  54.0        133
1991-02-28  55.0        134
1991-03-31  66.0        135
1991-04-30  65.0        136
1991-05-31  60.0        137
```

Fig – 1.18.2 Loading numerical time into the dataframe

```
LinearRegression()
```

Fig – 1.19 Initializing Linear Regression method

| YearMonth | Rose | train_time | RegOnTime |
|---|---|---|---|
| 1980-01-31 | 112.0 | 1 | 137.321144 |
| 1980-02-29 | 118.0 | 2 | 136.826766 |
| 1980-03-31 | 129.0 | 3 | 136.332388 |
| 1980-04-30 | 99.0 | 4 | 135.838010 |
| 1980-05-31 | 116.0 | 5 | 135.343632 |

Fig – 1.20 Predicting for training dataset

| YearMonth | Rose | test_time | RegOnTime |
|---|---|---|---|
| 1991-01-31 | 54.0 | 133 | 72.063266 |
| 1991-02-28 | 55.0 | 134 | 71.568888 |
| 1991-03-31 | 66.0 | 135 | 71.074511 |
| 1991-04-30 | 65.0 | 136 | 70.580133 |
| 1991-05-31 | 60.0 | 137 | 70.085755 |

Fig – 1.21 Predicting for test dataset

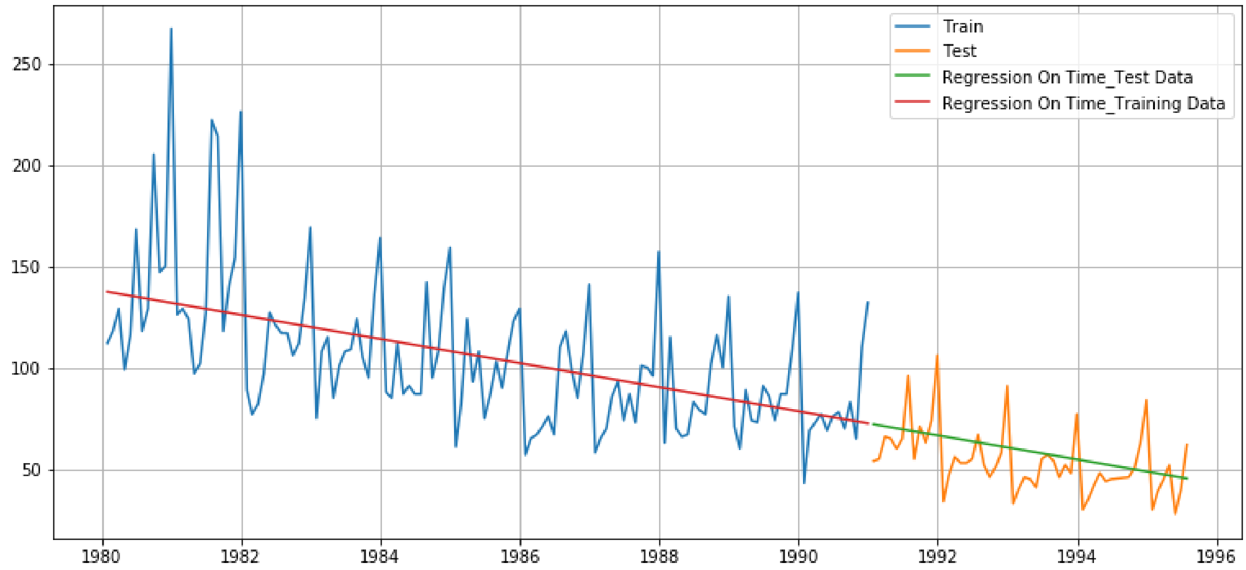



Fig – 1.22 Ploting original and predicted train and test datas using linear regression

```
For RegressionOnTime forecast on the Training Data,  RMSE is 30.718 and MAPE is 21.22
```

Fig – 1.23 RMSE and MAPE value Training data

```
For RegressionOnTime forecast on the Test Data,  RMSE is 15.269 and MAPE is 22.82
```

Fig – 1.24  RMSE and MAPE value Test data

| | Test RMSE | Test MAPE |
|---|---|---|
| RegressionOnTime | 15.268955 | 22.82 |

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    132.0
1980-02-29    132.0
1980-03-31    132.0
1980-04-30    132.0
1980-05-31    132.0
Name: naive, dtype: float64
```

Fig – 1.26 Predicting values for training data using naïve approach

```
YearMonth
1991-01-31    132.0
1991-02-28    132.0
1991-03-31    132.0
1991-04-30    132.0
1991-05-31    132.0
Name: naive, dtype: float64
```

Fig – 1.27 Predicting values for test data using naïve approach

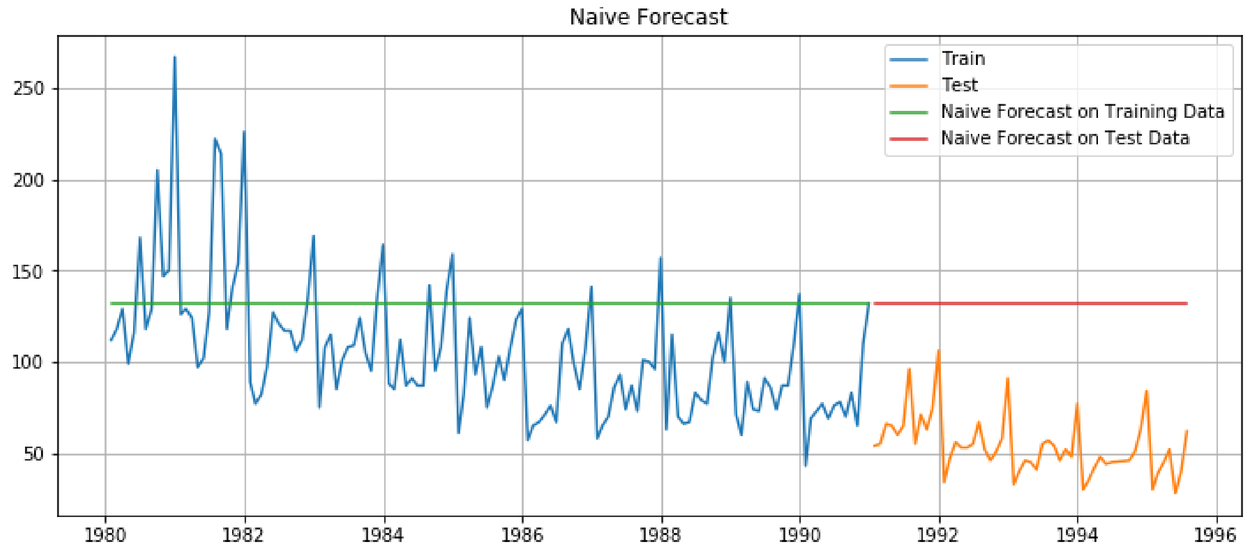



Fig – 1.28 Plotting the predicted values for train and test data using naïve approach

```
For Naive Model forecast on the Training Data,  RMSE is 45.064 MAPE is 36.38
```

Fig – 1.29 RMSE and MAPE value Training data

```
For RegressionOnTime forecast on the Test Data,  RMSE is 79.719 and MAPE is 145.10
```

Fig – 1.30 RMSE and MAPE value Test data

| | Test RMSE | Test MAPE |
|---|---|---|
| **RegressionOnTime** | 15.268955 | 22.82 |
| **NaiveModel** | 79.718773 | 145.10 |

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 | Rose | mean_forecast |
|---|---|---|
| **1980-01-31** | 112.0 | 104.939394 |
| **1980-02-29** | 118.0 | 104.939394 |
| **1980-03-31** | 129.0 | 104.939394 |
| **1980-04-30** | 99.0 | 104.939394 |
| **1980-05-31** | 116.0 | 104.939394 |

Fig – 1.32 Taking mean of Rose wine sales training data

| YearMonth | Rose | mean_forecast |
|---|---|---|
| **1991-01-31** | 54.0 | 53.854545 |
| **1991-02-28** | 55.0 | 53.854545 |
| **1991-03-31** | 66.0 | 53.854545 |
| **1991-04-30** | 65.0 | 53.854545 |
| **1991-05-31** | 60.0 | 53.854545 |

Fig – 1.33 Taking mean of Rose wine sales test data

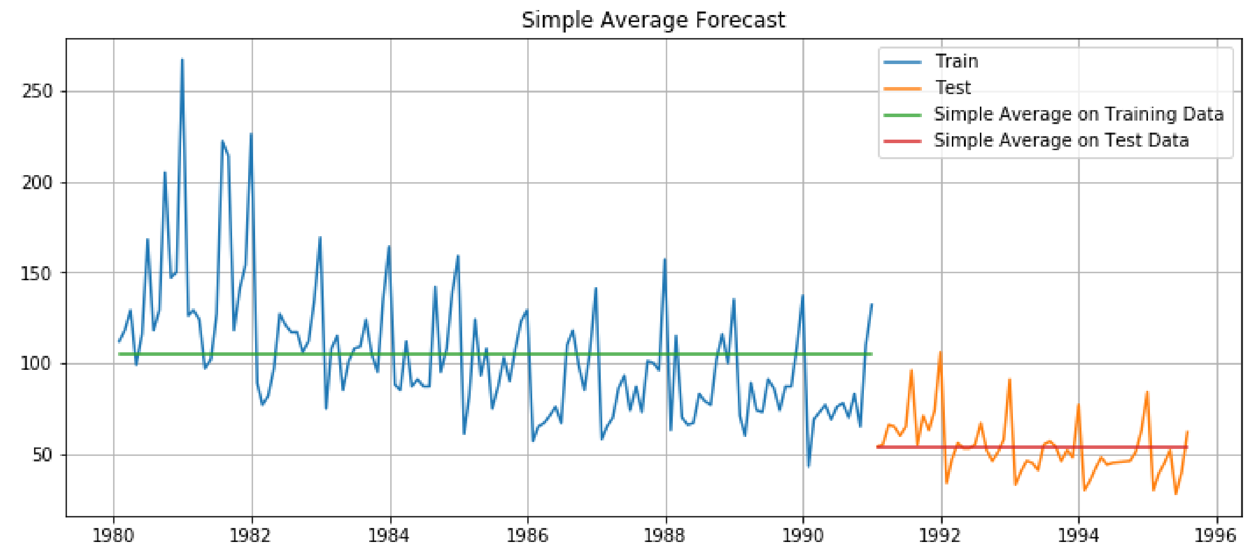



Fig – 1.34 Plotting the Simple Average data, train and test data

```
For Simple Average Model forecast on the Training Data,  RMSE is 36.034 MAPE is 25.39
```

Fig – 1.35 RMSE and MAPE value Training data

```
For Simple Average forecast on the Test Data,  RMSE is 15.760 MAPE is 21.37
```

Fig – 1.36 RMSE and MAPE value Test data

| | Test RMSE | Test MAPE |
|---|---|---|
| **RegressionOnTime** | 15.268955 | 22.82 |
| **NaiveModel** | 79.718773 | 145.10 |
| **SimpleAverageModel** | 15.759783 | 21.37 |

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

## ***<u>Model – 4 Moving Average(MA):</u>***

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.

| | Rose |
|---|---|
| **YearMonth** | |
| **1980-01-31** | 112.0 |
| **1980-02-29** | 118.0 |
| **1980-03-31** | 129.0 |
| **1980-04-30** | 99.0 |
| **1980-05-31** | 116.0 |

Fig – 1.38 training data

| | Rose | Trailing_2 | Trailing_4 | Trailing_6 | Trailing_9 |
|---|---|---|---|---|---|
| **YearMonth** | | | | | |
| **1980-01-31** | 112.0 | NaN | NaN | NaN | NaN |
| **1980-02-29** | 118.0 | 115.0 | NaN | NaN | NaN |
| **1980-03-31** | 129.0 | 123.5 | NaN | NaN | NaN |
| **1980-04-30** | 99.0 | 114.0 | 114.5 | NaN | NaN |
| **1980-05-31** | 116.0 | 107.5 | 115.5 | NaN | NaN |

Fig – 1.39 Making data from training data for moving average

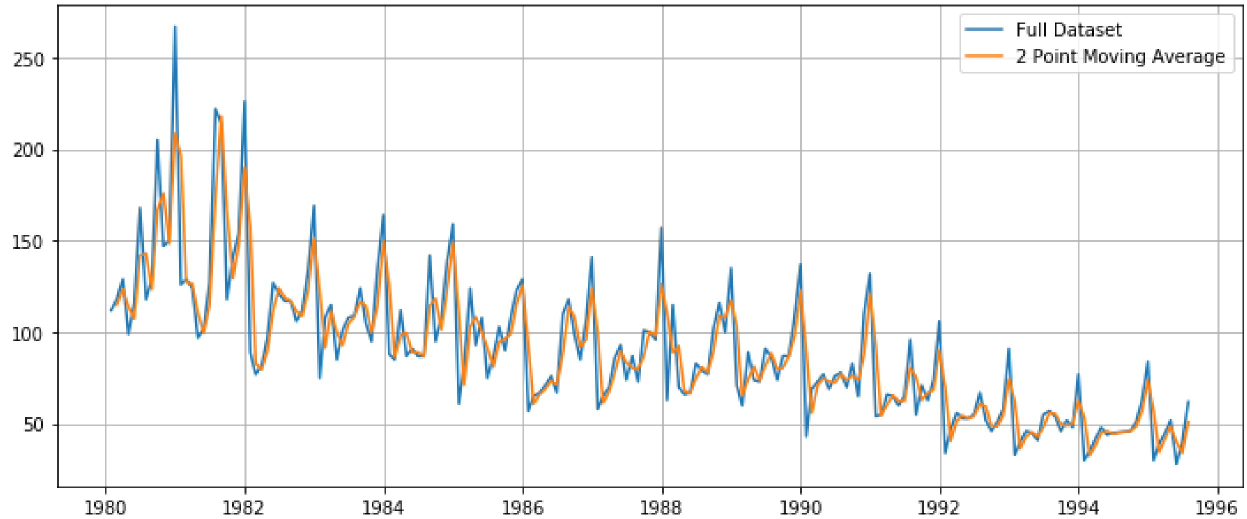



Fig – 1.40 Plotting Moving average data for rolling 2 point

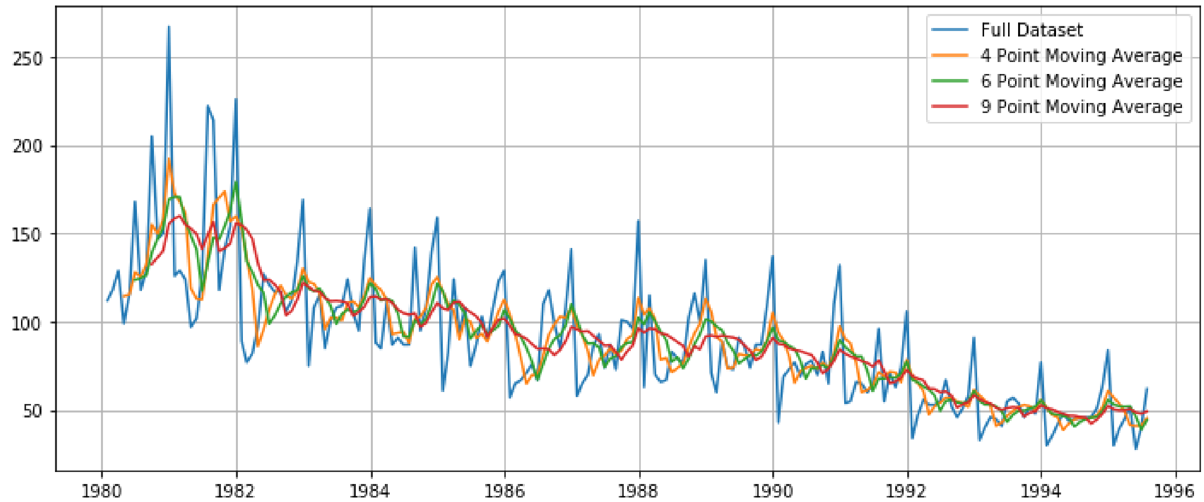



Fig – 1.41 Plotting Moving average data for rolling 2,4,6,9 point

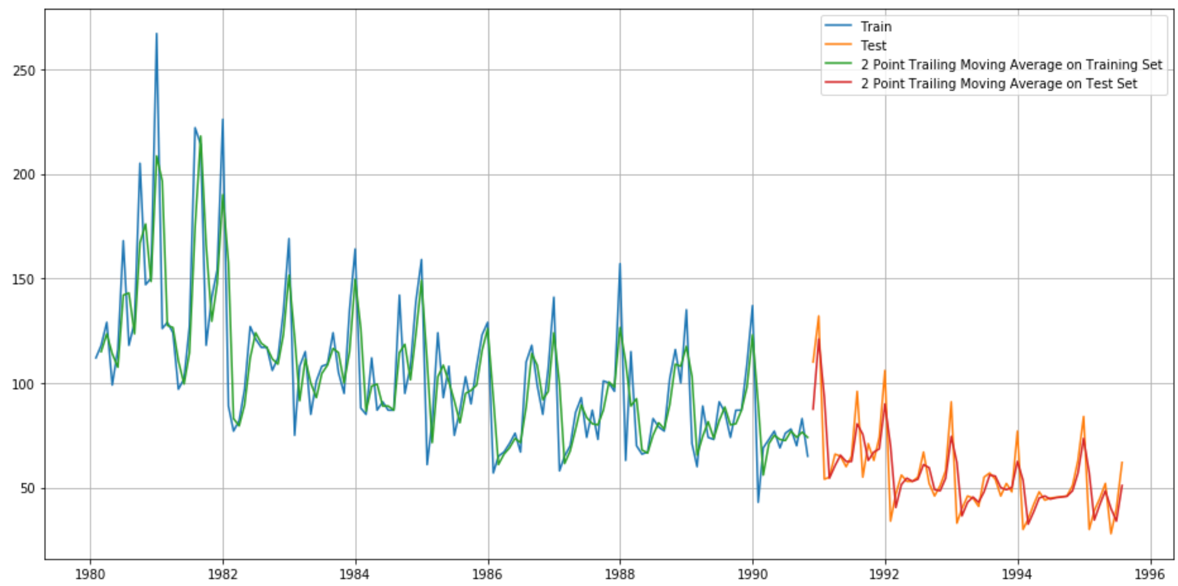



Fig – 1.42 Plotting Moving average data for rolling 2 point for training and test data

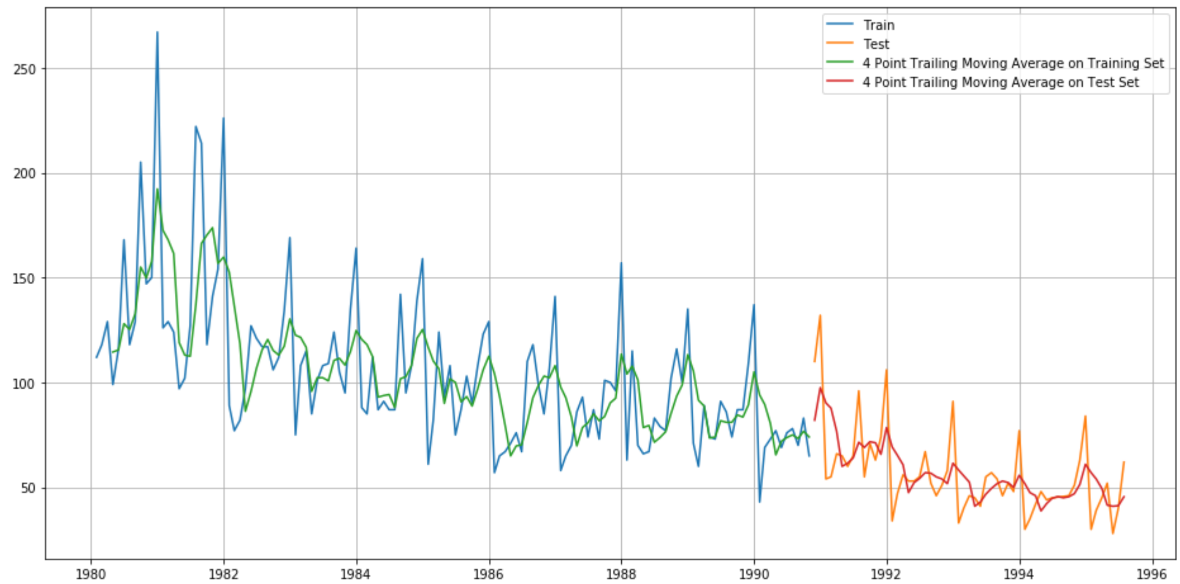



Fig – 1.43 Plotting Moving average data for rolling 4 point for training and test data

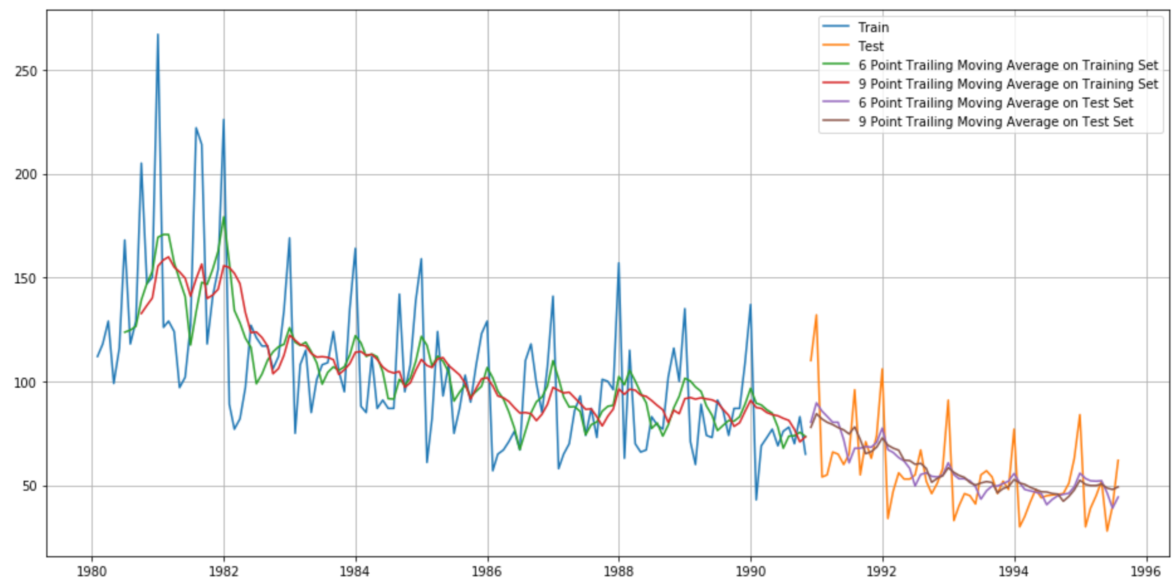



Fig – 1.44 Plotting Moving average data for rolling 6,9 point for training and test data

| | Rose | Trailing_2 | Trailing_4 | Trailing_6 | Trailing_9 |
|---|---|---|---|---|---|
| **YearMonth** | | | | | |
| **1990-11-30** | 110.0 | 87.5 | 82.00 | 80.333333 | 77.888889 |
| **1990-12-31** | 132.0 | 121.0 | 97.50 | 89.666667 | 84.444444 |
| **1991-01-31** | 54.0 | 93.0 | 90.25 | 85.666667 | 81.888889 |
| **1991-02-28** | 55.0 | 54.5 | 87.75 | 83.166667 | 80.333333 |
| **1991-03-31** | 66.0 | 60.5 | 76.75 | 80.333333 | 79.222222 |

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 11.529 and MAPE is 13.54
For 4 point Moving Average Model forecast on the Testing Data,  RMSE is 14.451 and MAPE is 19.49
For 6 point Moving Average Model forecast on the Testing Data,  RMSE is 14.566 and MAPE is 20.82
For 9 point Moving Average Model forecast on the Testing Data,  RMSE is 14.728 and MAPE is 21.01
```

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

| | Test RMSE | Test MAPE |
|---|---|---|
| **RegressionOnTime** | 15.268955 | 22.82 |
| **NaiveModel** | 79.718773 | 145.10 |
| **SimpleAverageModel** | 15.759783 | 21.37 |
| **2pointTrailingMovingAverage** | 11.529278 | 13.54 |
| **4pointTrailingMovingAverage** | 14.451403 | 19.49 |
| **6pointTrailingMovingAverage** | 14.566327 | 20.82 |
| **9pointTrailingMovingAverage** | 14.727630 | 21.01 |

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

***<u>Model – 5 Simple Exponential Smoothing:</u>***

```
{'smoothing_level': 0.09874989028077343,
 'smoothing_slope': nan,
 'smoothing_seasonal': nan,
 'damping_slope': nan,
 'initial_level': 134.3870166871304,
 '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 0.0987
Initial Level 134.387
```

| | 0 |
|---|---|
| **1991-01-31** | 87.104999 |
| **1991-02-28** | 87.104999 |
| **1991-03-31** | 87.104999 |

Fig – 1.49 predicting values for the Simple Exponential Smoothing

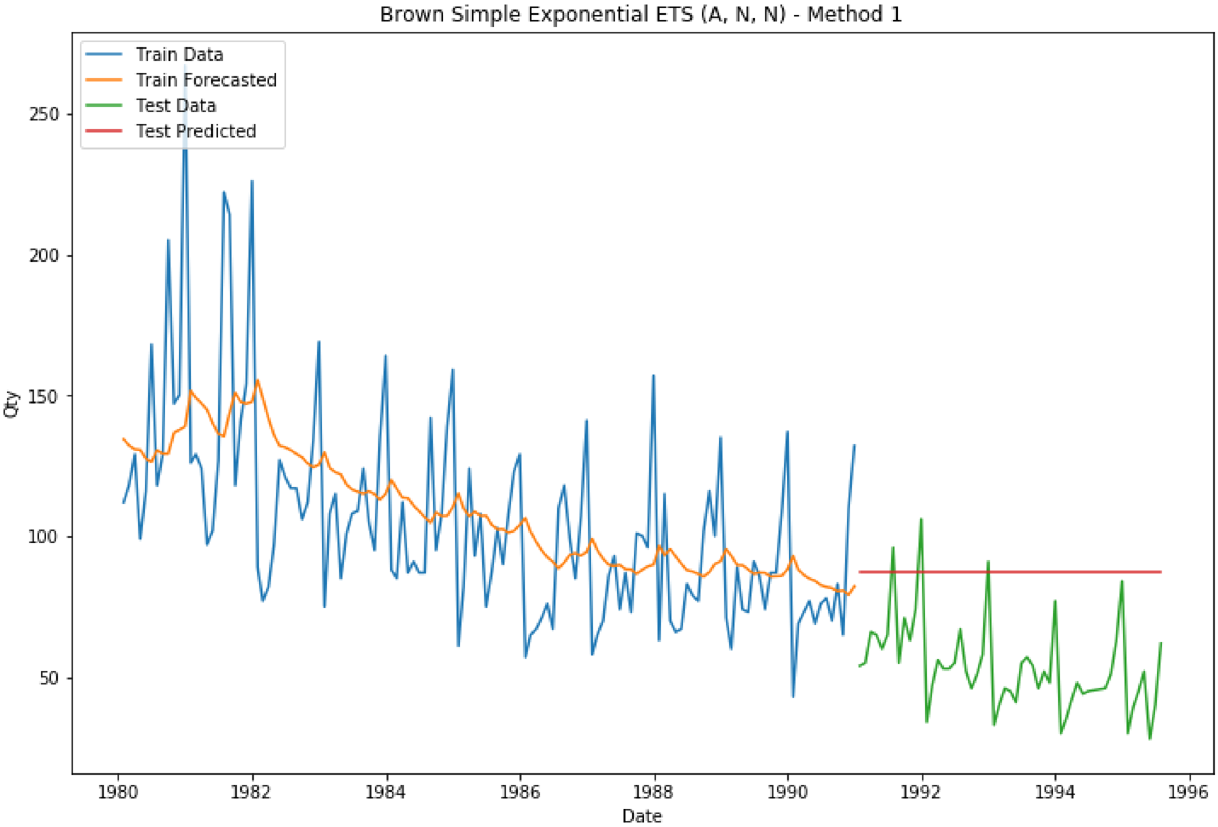



Fig – 1.50 Plotting predicted values for the Simple Exponential Smoothing

```
For Alpha =0.0987 Simple Exponential Smoothing Model forecast on the Training Data,  RMSE is 31.501 MAPE is 22.73
```

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

```
For Alpha =0.0987 Simple Exponential Smoothing Model forecast on the Testing Data,  RMSE is 36.796 MAPE is 63.88
```

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

| | Test RMSE | Test MAPE |
|---|---|---|
| **RegressionOnTime** | 15.268955 | 22.82 |
| **NaiveModel** | 79.718773 | 145.10 |
| **SimpleAverageModel** | 15.759783 | 21.37 |
| **2pointTrailingMovingAverage** | 11.529278 | 13.54 |
| **4pointTrailingMovingAverage** | 14.451403 | 19.49 |
| **6pointTrailingMovingAverage** | 14.566327 | 20.82 |
| **9pointTrailingMovingAverage** | 14.727630 | 21.01 |
| **Alpha=0.0987,SimpleExponentialSmoothing** | 36.796242 | 63.88 |

Fig – 1.53 RMSE and MAPE value for the test data 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 – 1.54 Initializing the Double Exponential Smoothing

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

Smoothing Level 0.01
Initial Level 116.4741
```

| YearMonth | Rose | predict |
|---|---|---|
| 1991-01-31 | 54.0 | 100.012001 |
| 1991-02-28 | 55.0 | 99.873957 |
| 1991-03-31 | 66.0 | 99.735913 |
| 1991-04-30 | 65.0 | 99.597869 |
| 1991-05-31 | 60.0 | 99.459825 |

Fig – 1.55 Predicting the values using Double Exponential Smoothing

```
Train RMSE is -> 35.797908987769944
Train MAPE is -> 26.73
```

Fig – 1.56 RMSE and MAPE for training data

```
Test RMSE is -> 45.00503693213322
Test MAPE is -> 79.29
```

Fig – 1.57 RMSE and MAPE for test data

| | Test RMSE | Test MAPE |
|---|---|---|
| RegressionOnTime | 15.268955 | 22.82 |
| NaiveModel | 79.718773 | 145.10 |
| SimpleAverageModel | 15.759783 | 21.37 |
| 2pointTrailingMovingAverage | 11.529278 | 13.54 |
| 4pointTrailingMovingAverage | 14.451403 | 19.49 |
| 6pointTrailingMovingAverage | 14.566327 | 20.82 |
| 9pointTrailingMovingAverage | 14.727630 | 21.01 |
| Alpha=0.0987,SimpleExponentialSmoothing | 36.796242 | 63.88 |
| Alpha=0.01,beta=0.01,DoubleExponentialSmoothing | 45.005037 | 79.29 |

Fig – 1.58 RMSE and MAPE for test data into the dataframe

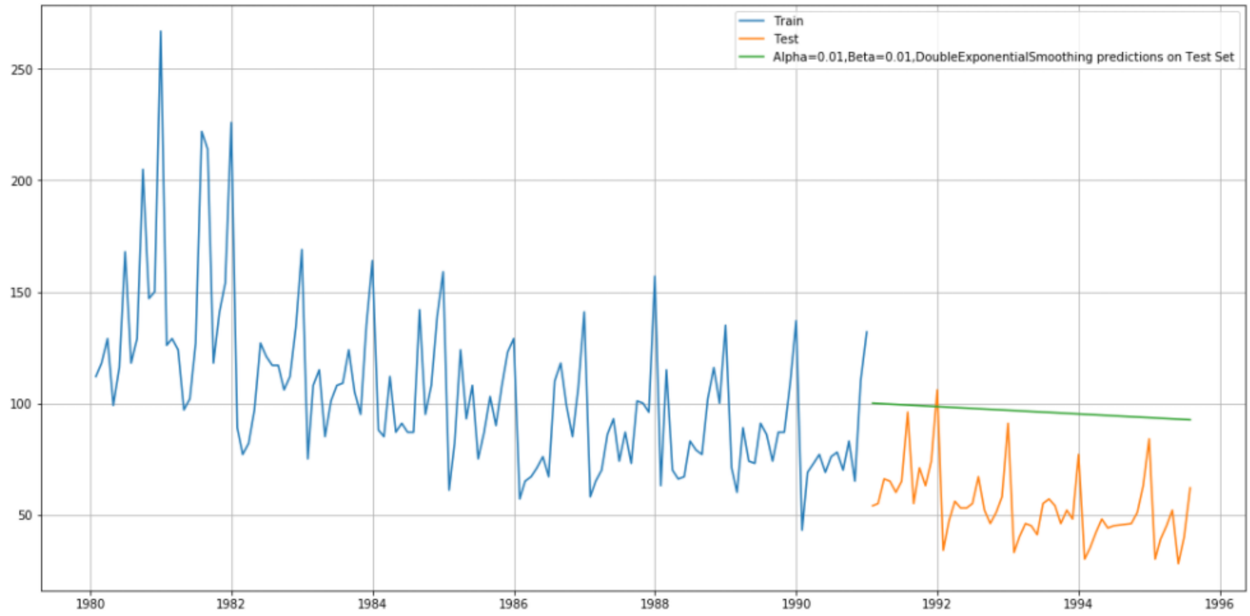



Fig – 1.59 plotting DES predicted output

| | Alpha Values | Beta Values | Train RMSE | Test RMSE | Test MAPE | Train MAPE |
|---|---|---|---|---|---|---|
| **0** | 0.10 | 0.1 | 34.439111 | 36.923416 | 63.78 | 24.83 |
| **10** | 0.11 | 0.1 | 34.000195 | 39.062023 | 67.48 | 24.57 |
| **20** | 0.12 | 0.1 | 33.684824 | 41.420568 | 71.51 | 24.36 |
| **30** | 0.13 | 0.1 | 33.459597 | 44.010923 | 76.01 | 24.18 |
| **40** | 0.14 | 0.1 | 33.301637 | 46.802589 | 80.87 | 24.03 |

Fig – 1.60 Finding RMSE and MAPE for different Alpha and beta values

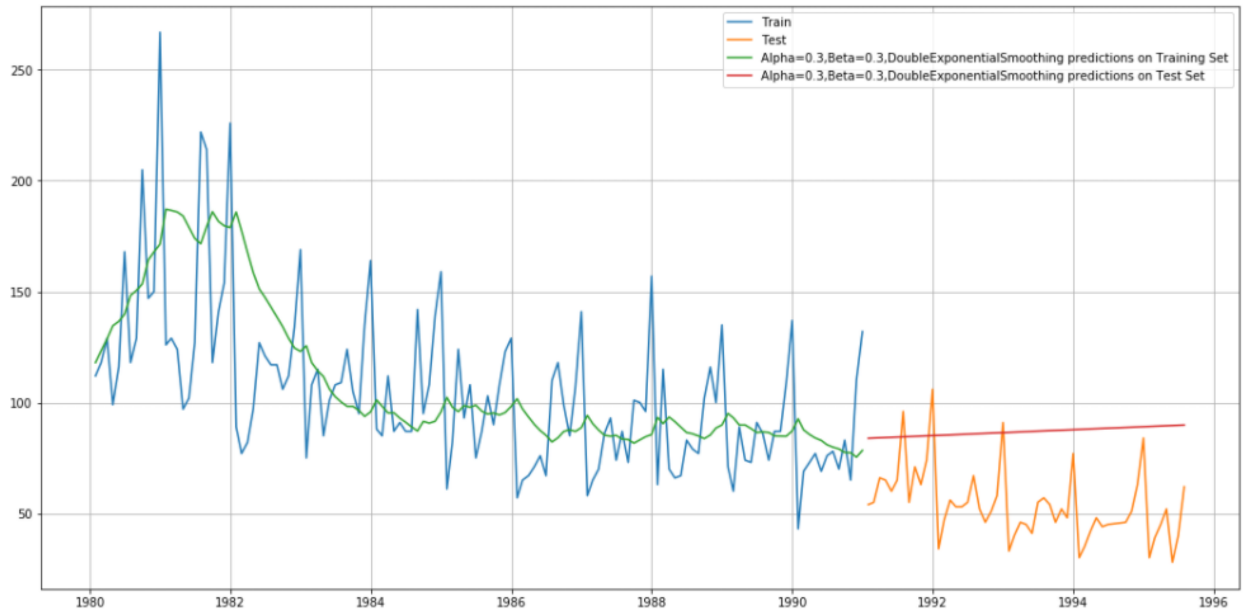


Fig – 1.61 plotting for least RMSE value of different alpha n beta values

| | Test RMSE | Test MAPE |
|---|---|---|
| RegressionOnTime | 15.268955 | 22.82 |
| NaiveModel | 79.718773 | 145.10 |
| SimpleAverageModel | 15.759783 | 21.37 |
| 2pointTrailingMovingAverage | 11.529278 | 13.54 |
| 4pointTrailingMovingAverage | 14.451403 | 19.49 |
| 6pointTrailingMovingAverage | 14.566327 | 20.82 |
| 9pointTrailingMovingAverage | 14.727630 | 21.01 |
| Alpha=0.0987,SimpleExponentialSmoothing | 36.796242 | 63.88 |
| Alpha=0.01,beta=0.01,DoubleExponentialSmoothing | 45.005037 | 79.29 |
| Alpha=0.10,Beta=0.1,DoubleExponentialSmoothing | 36.923416 | 63.78 |

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

***<u>Model 7: Triple Exponential Smoothing (Holt - Winter's Model)</u>***

Three parameters $\alpha$ , $\beta$ and $\gamma$ are estimated in this model. Level, Trend and Seasonality are accounted for in this model.

```
{'smoothing_level': 0.99,
 'smoothing_slope': 0.01,
 'smoothing_seasonal': 0.01,
 'damping_slope': nan,
 'initial_level': 93.7006343484885,
 'initial_slope': 0.0,
 'initial_seasons': array([1.19877733, 1.39685338, 1.58600046]),
 'use_boxcox': False,
 'lamda': None,
 'remove_bias': False}
```

Fig 1.63 Initializing the TES

| YearMonth | Rose | auto_predict |
|---|---|---|
| 1980-01-31 | 112.0 | 112.326196 |
| 1980-02-29 | 118.0 | 130.505991 |
| 1980-03-31 | 129.0 | 133.975457 |
| 1980-04-30 | 99.0 | 97.392625 |
| 1980-05-31 | 116.0 | 115.079766 |

Fig 1.64 Predicting the values from training data

| YearMonth | Rose | auto_predict |
|---|---|---|
| 1991-01-31 | 54.0 | 105.568220 |
| 1991-02-28 | 55.0 | 116.810432 |
| 1991-03-31 | 66.0 | 132.260129 |
| 1991-04-30 | 65.0 | 105.834395 |
| 1991-05-31 | 60.0 | 117.104706 |

Fig 1.65 Predicting the values from test data

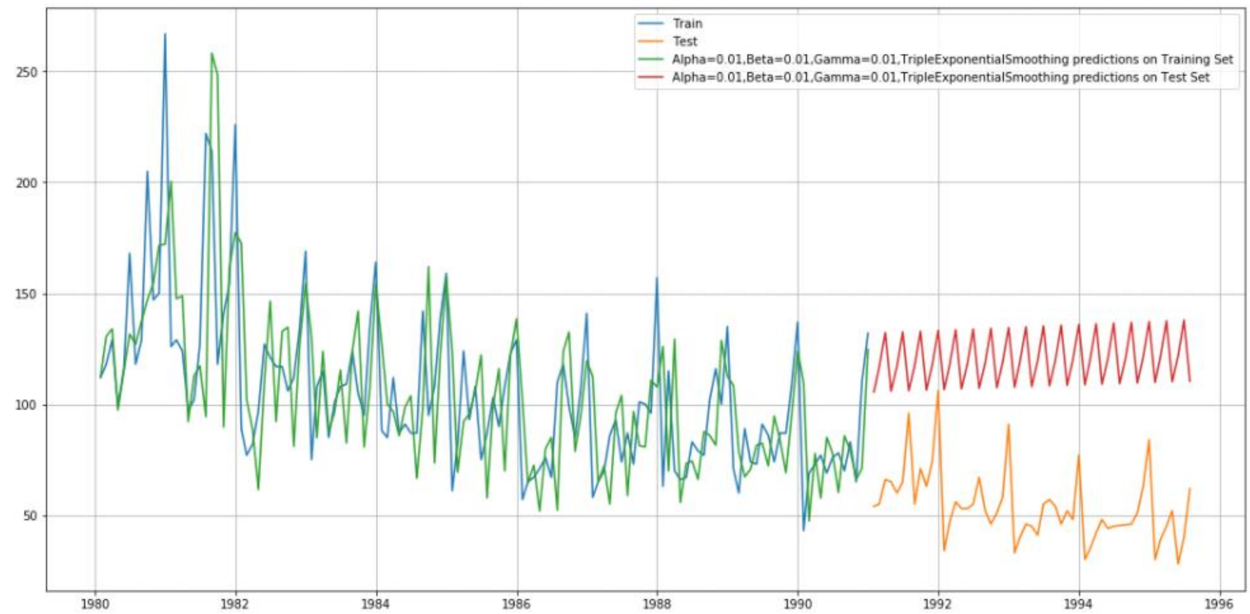



Fig 1.66 Plotting predicted train and test values

```
For Alpha: 0.99,Beta: 0.01 and Gamma:0.01, Triple Exponential Smoothing Model forecast on the Training Data,  RMSE is 33.018 MA
PE is 22.41
```

Fig 1.67 RMSE and MAPE values from train data

```
For Alpha: 0.99,Beta: 0.01 and Gamma:0.01,Triple Exponential Smoothing Model forecast on the Test Data,  RMSE is 69.035 MAPE is
123.85
```

Fig 1.68 RMSE and MAPE values from test data

| | Test RMSE | Test MAPE |
|---|---|---|
| **RegressionOnTime** | 15.268955 | 22.82 |
| **NaiveModel** | 79.718773 | 145.10 |
| **SimpleAverageModel** | 15.759783 | 21.37 |
| **2pointTrailingMovingAverage** | 11.529278 | 13.54 |
| **4pointTrailingMovingAverage** | 14.451403 | 19.49 |
| **6pointTrailingMovingAverage** | 14.566327 | 20.82 |
| **9pointTrailingMovingAverage** | 14.727630 | 21.01 |
| **Alpha=0.0987,SimpleExponentialSmoothing** | 36.796242 | 63.88 |
| **Alpha=0.01,beta=0.01,DoubleExponentialSmoothing** | 45.005037 | 79.29 |
| **Alpha=0.10,Beta=0.1,DoubleExponentialSmoothing** | 36.923416 | 63.78 |
| **Alpha: 0.99,Beta: 0.01 and Gamma:0.01,TripleExponentialSmoothing** | 69.035167 | 123.85 |

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 |
|---|---|---|---|---|---|---|---|
| **301** | 0.01 | 0.76 | 0.06 | 36.786490 | 26.95 | 14.949445 | 23.54 |
| **281** | 0.01 | 0.71 | 0.06 | 37.027698 | 27.09 | 15.115942 | 21.57 |
| **61** | 0.01 | 0.16 | 0.06 | 47.975604 | 36.42 | 15.505402 | 19.69 |
| **320** | 0.01 | 0.81 | 0.01 | 41.297574 | 31.49 | 16.045656 | 23.41 |
| **62** | 0.01 | 0.16 | 0.11 | 41.530087 | 30.42 | 16.131210 | 26.18 |

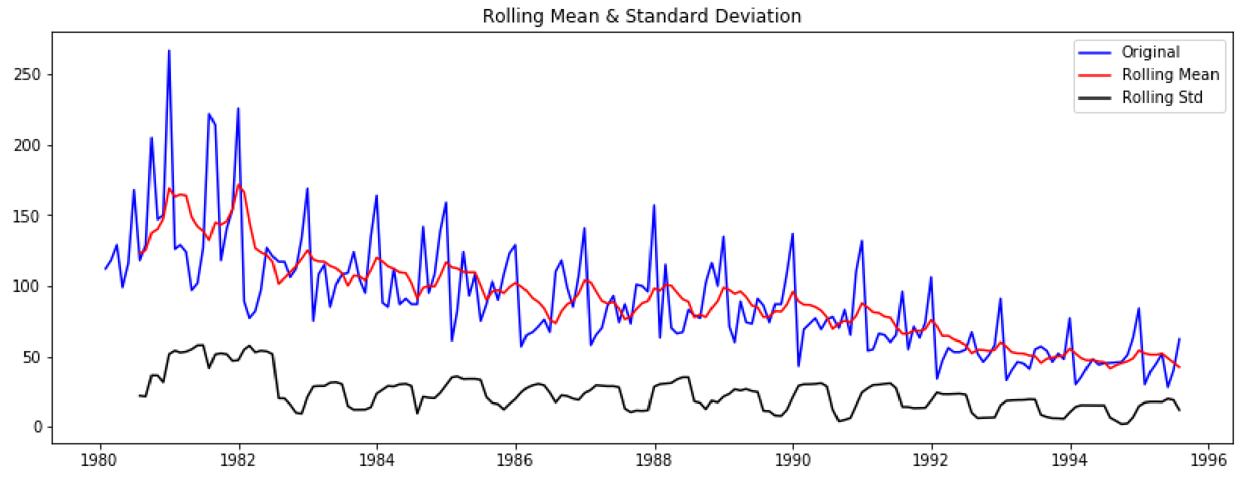
Fig 1.70 RMSE and MAPE for different alpha, beta, and gamma values into the dataframe

| | Test RMSE | Test MAPE |
|---|---|---|
| RegressionOnTime | 15.268955 | 22.82 |
| NaiveModel | 79.718773 | 145.10 |
| SimpleAverageModel | 15.759783 | 21.37 |
| 2pointTrailingMovingAverage | 11.529278 | 13.54 |
| 4pointTrailingMovingAverage | 14.451403 | 19.49 |
| 6pointTrailingMovingAverage | 14.566327 | 20.82 |
| 9pointTrailingMovingAverage | 14.727630 | 21.01 |
| Alpha=0.0987,SimpleExponentialSmoothing | 36.796242 | 63.88 |
| Alpha=0.01,beta=0.01,DoubleExponentialSmoothing | 45.005037 | 79.29 |
| Alpha=0.10,Beta=0.1,DoubleExponentialSmoothing | 36.923416 | 63.78 |
| Alpha: 0.99,Beta: 0.01 and Gamma:0.01,TripleExponentialSmoothing | 69.035167 | 123.85 |
| Alpha=0.01,Beta=0.76,Gamma=0.06,TripleExponentialSmoothingWithGrid | 14.949445 | 19.69 |

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.

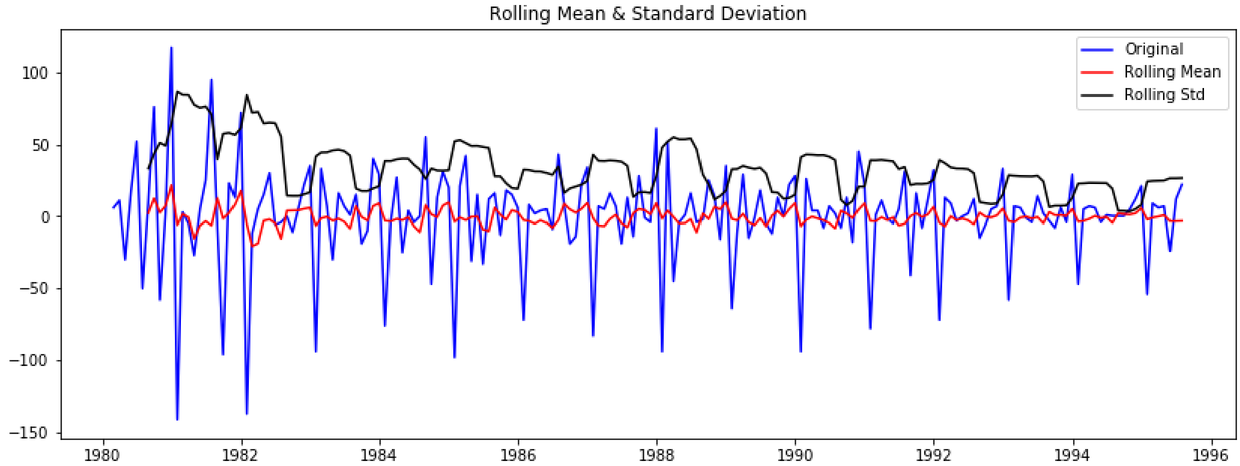
***Note:*** Stationarity should be checked at alpha($\alpha$) = 0.05.
**Null Hypothesis:** p-value > $\alpha$ (alpha value) - then the data is not stationary
**Alternate Hypothesis:** p-value < $\alpha$ (alpha value) - then the data is having stationarity




```
Results of Dickey-Fuller Test:
Test Statistic                  -1.876699
p-value                          0.343101
#Lags Used                      13.000000
Number of Observations Used    173.000000
Critical Value (1%)             -3.468726
Critical Value (5%)             -2.878396
Critical Value (10%)            -2.575756
dtype: float64
```

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                -8.044392e+00
p-value                        1.810895e-12
#Lags Used                     1.200000e+01
Number of Observations Used    1.730000e+02
Critical Value (1%)           -3.468726e+00
Critical Value (5%)           -2.878396e+00
Critical Value (10%)          -2.575756e+00
dtype: float64
```

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.

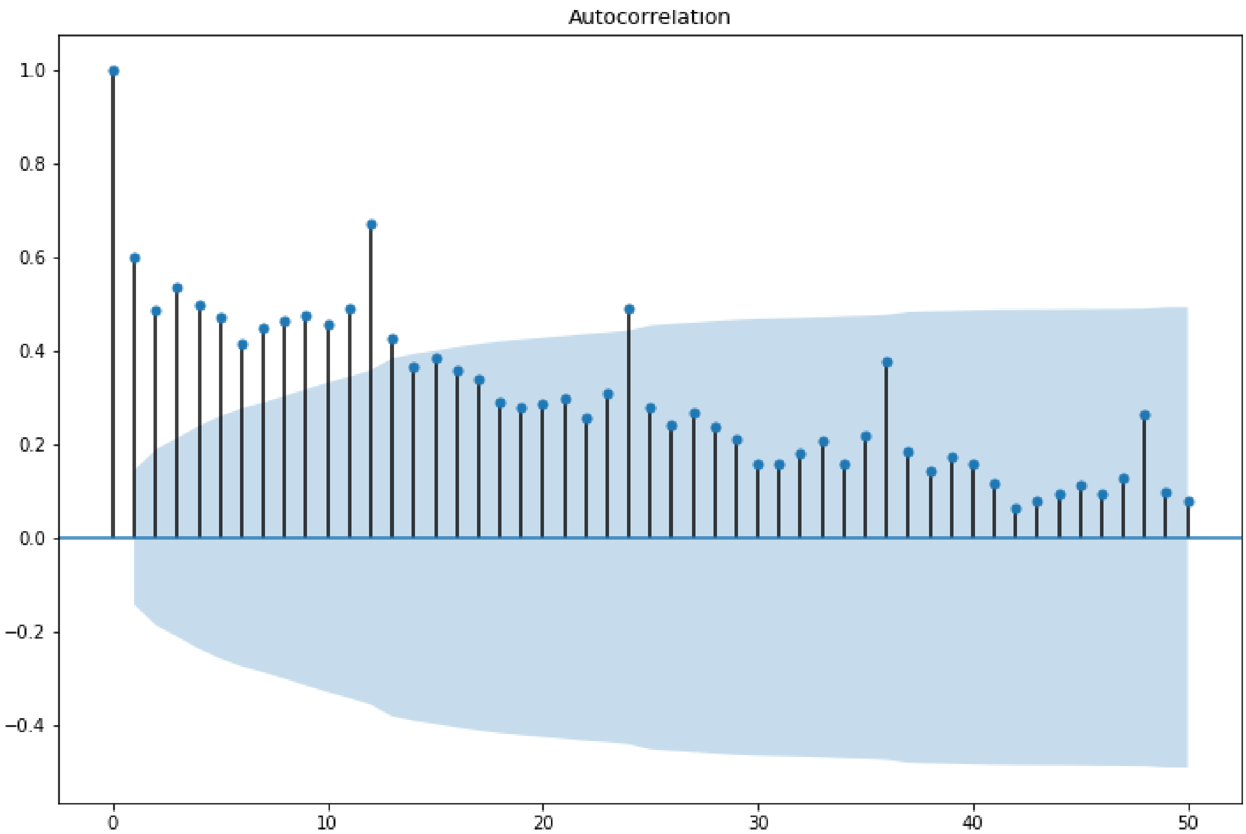



Fig 1.74 ACF plot for original data

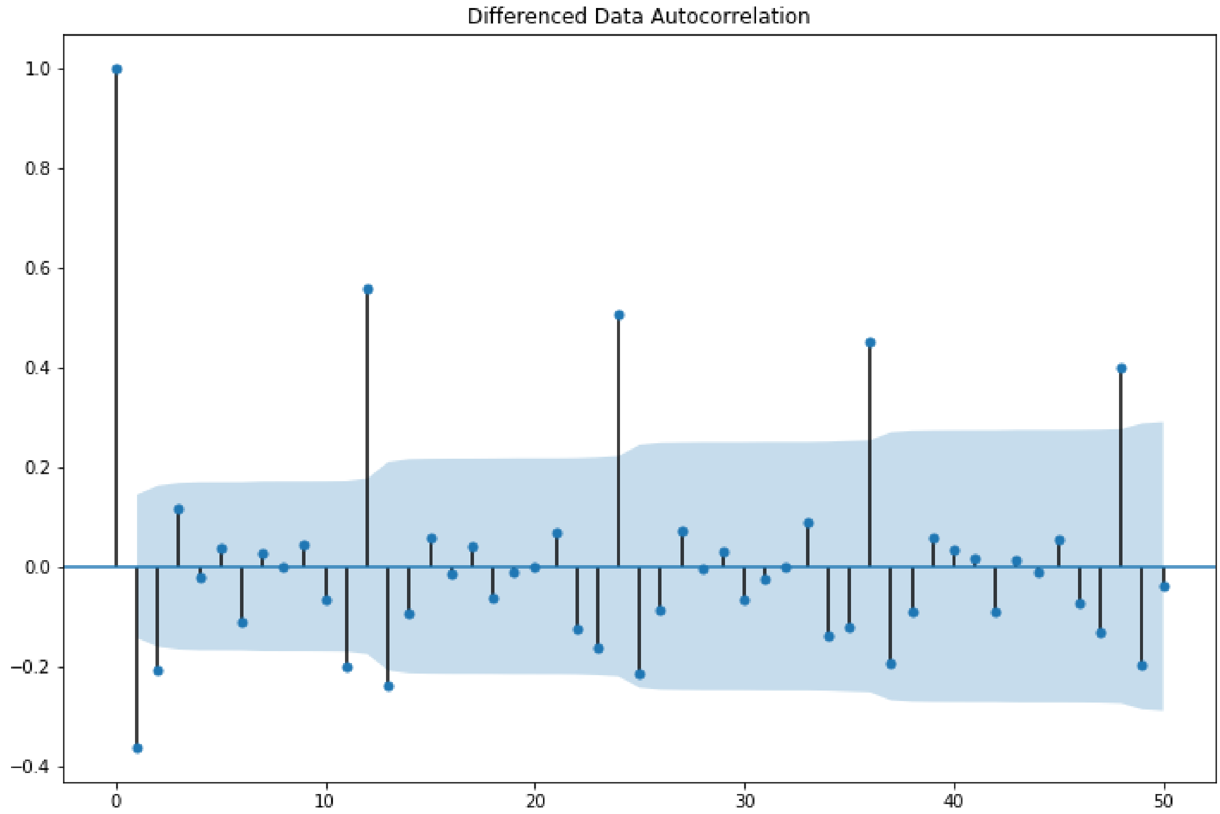



Fig 1.75 ACF plot for 1st difference data

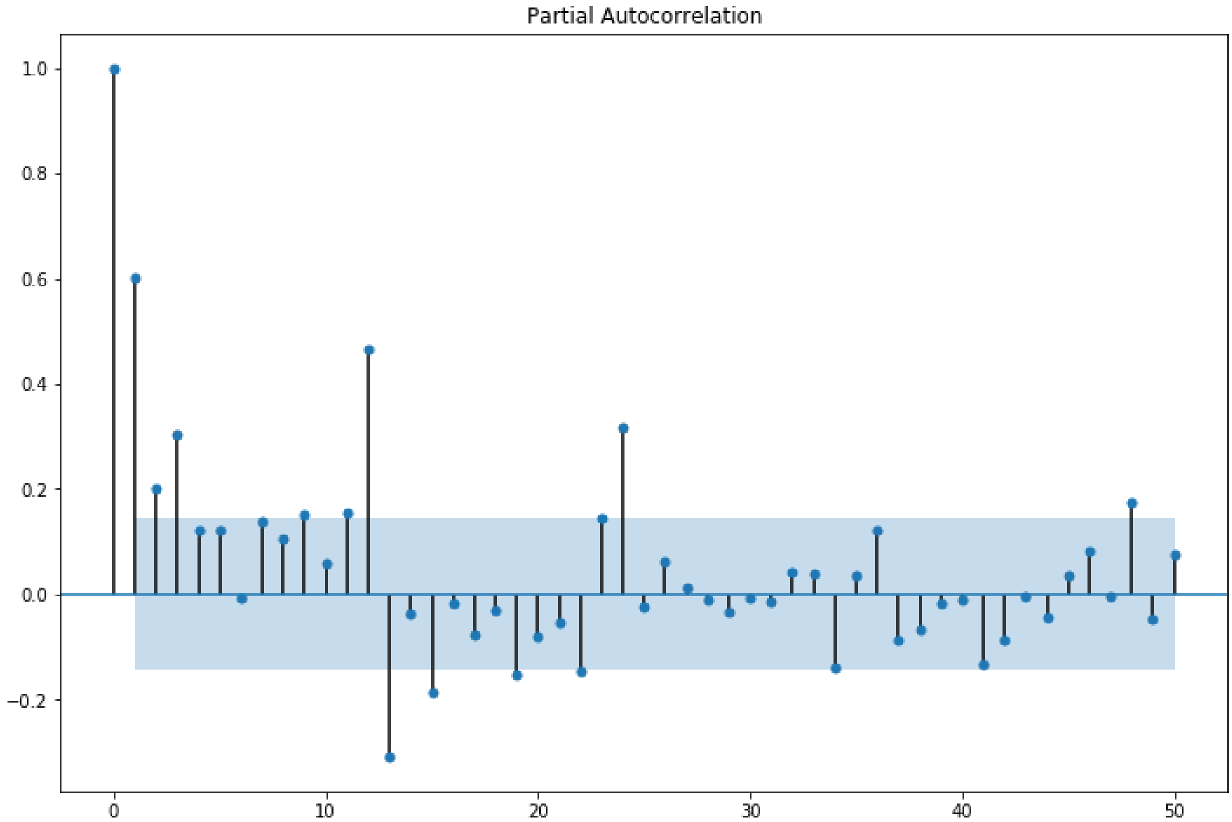



Fig 1.76 PACF plot for original data

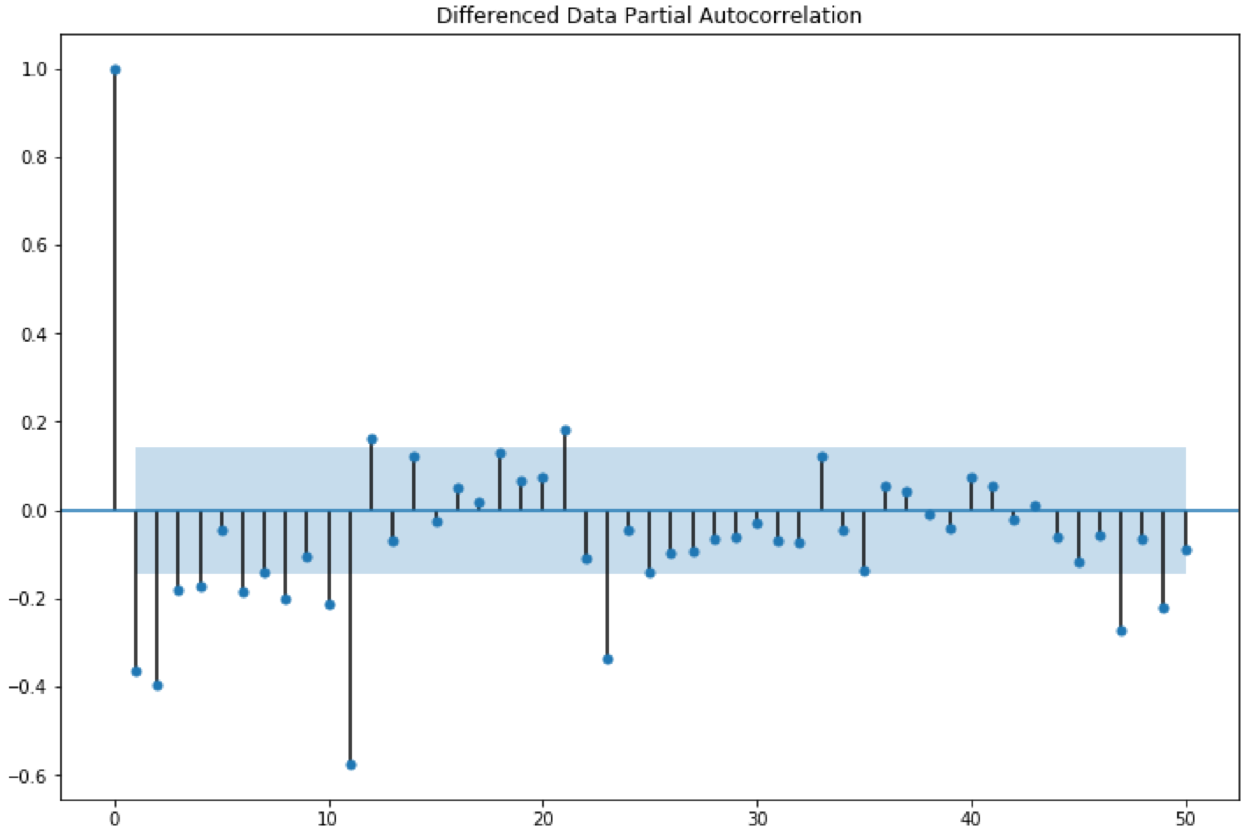
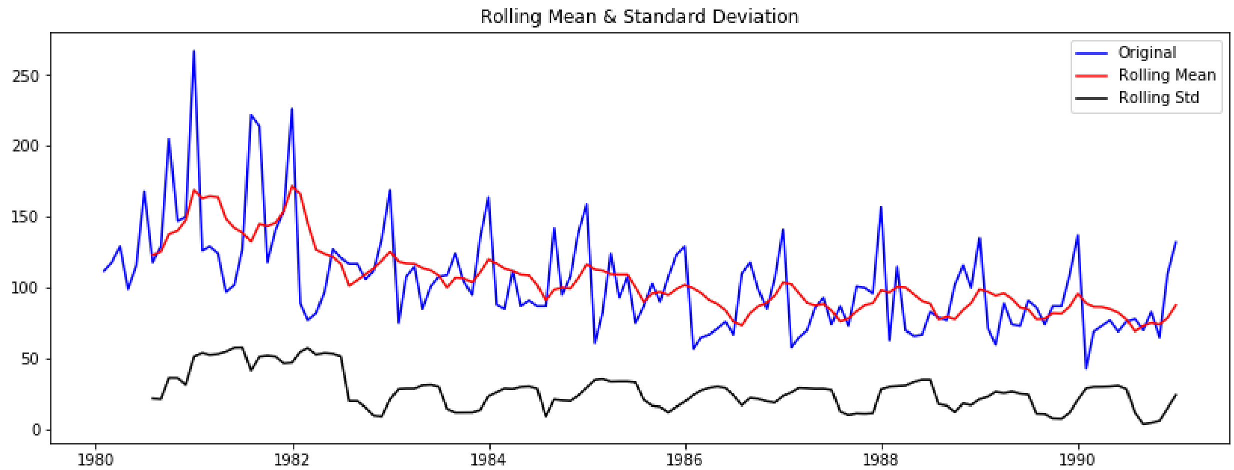


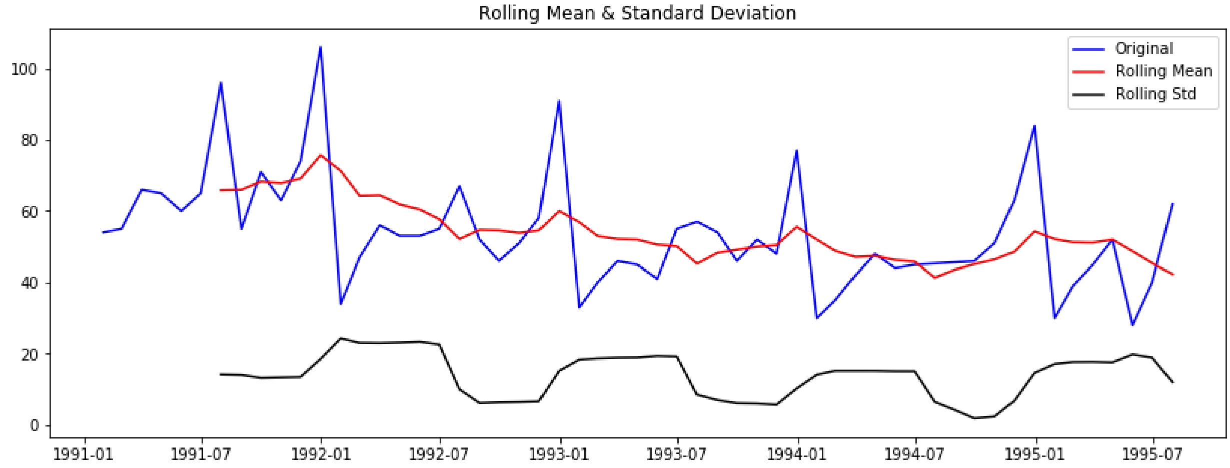

Fig 1.77 PACF plot for 1st difference data




```
Results of Dickey-Fuller Test:
Test Statistic                  -2.164250
p-value                          0.219476
#Lags Used                      13.000000
Number of Observations Used    118.000000
Critical Value (1%)             -3.487022
Critical Value (5%)             -2.886363
Critical Value (10%)            -2.580009
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                 -4.432301
p-value                         0.000260
#Lags Used                     11.000000
Number of Observations Used    43.000000
Critical Value (1%)            -3.592504
Critical Value (5%)            -2.931550
Critical Value (10%)           -2.604066
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:1335.1526583086775
ARIMA(0, 1, 1) - AIC:1280.7261830464574
ARIMA(0, 1, 2) - AIC:1276.8353792448133
ARIMA(1, 1, 0) - AIC:1319.3483105803643
ARIMA(1, 1, 1) - AIC:1277.7757551519576
ARIMA(1, 1, 2) - AIC:1277.3592239364807
ARIMA(1, 1, 3) - AIC:1279.3126365443986
```

Fig 1.81 Sample AIC value for ARIMA Model

| | param | AIC |
|---|---|---|
| **16** | (3, 1, 3) | 1273.194207 |
| **17** | (3, 1, 4) | 1274.335727 |
| **2** | (0, 1, 2) | 1276.835379 |
| **5** | (1, 1, 2) | 1277.359224 |
| **4** | (1, 1, 1) | 1277.775755 |

Fig 1.82 Sorting least AIC value for ARIMA Model

```
                             ARIMA Model Results
==============================================================================
Dep. Variable:                 D.Rose   No. Observations:                  131
Model:                 ARIMA(3, 1, 3)   Log Likelihood                -628.597
Method:                       css-mle   S.D. of innovations             28.356
Date:                Sun, 20 Feb 2022   AIC                           1273.194
Time:                        23:02:05   BIC                           1296.196
Sample:                    02-29-1980   HQIC                          1282.541
                         - 12-31-1990
================================================================================
                   coef    std err          z      P>|z|      [0.025      0.975]
--------------------------------------------------------------------------------
const           -0.4906      0.088     -5.546      0.000      -0.664      -0.317
ar.L1.D.Rose    -0.7241      0.086     -8.385      0.000      -0.893      -0.555
ar.L2.D.Rose    -0.7215      0.087     -8.314      0.000      -0.892      -0.551
ar.L3.D.Rose     0.2765      0.086      3.228      0.002       0.109       0.444
ma.L1.D.Rose    -0.0154      0.045     -0.344      0.731      -0.103       0.072
ma.L2.D.Rose     0.0154      0.044      0.346      0.730      -0.072       0.103
ma.L3.D.Rose    -1.0000      0.046    -21.820      0.000      -1.090      -0.910
                                    Roots
=============================================================================
                  Real          Imaginary           Modulus         Frequency
-----------------------------------------------------------------------------
AR.1           -0.5011           -0.8661j            1.0006           -0.3335
AR.2           -0.5011           +0.8661j            1.0006            0.3335
AR.3            3.6113           -0.0000j            3.6113           -0.0000
MA.1            1.0000           -0.0000j            1.0000           -0.0000
MA.2           -0.4923           -0.8704j            1.0000           -0.3319
MA.3           -0.4923           +0.8704j            1.0000            0.3319
-----------------------------------------------------------------------------
```

Fig 1.83 Summary report for ARIMA Model

```
Test rmse for arima is  15.98895461779234
Test mape for arima is  26.09
```

Fig 1.84 RMSE and MAPE for ARIMA Model

| | Test RMSE | Test MAPE |
|---|---|---|
| RegressionOnTime | 15.268955 | 22.82 |
| NaiveModel | 79.718773 | 145.10 |
| SimpleAverageModel | 15.759783 | 21.37 |
| 2pointTrailingMovingAverage | 11.529278 | 13.54 |
| 4pointTrailingMovingAverage | 14.451403 | 19.49 |
| 6pointTrailingMovingAverage | 14.566327 | 20.82 |
| 9pointTrailingMovingAverage | 14.727630 | 21.01 |
| Alpha=0.0987,SimpleExponentialSmoothing | 36.796242 | 63.88 |
| Alpha=0.01,beta=0.01,DoubleExponentialSmoothing | 45.005037 | 79.29 |
| Alpha=0.10,Beta=0.1,DoubleExponentialSmoothing | 36.923416 | 63.78 |
| Alpha: 0.99,Beta: 0.01 and Gamma:0.01,TripleExponentialSmoothing | 69.035167 | 123.85 |
| Alpha=0.01,Beta=0.76,Gamma=0.06,TripleExponentialSmoothingWithGrid | 14.949445 | 19.69 |
| Arima 3,1,3 | 15.988955 | 26.09 |

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:1323.9657875279158
SARIMA(0, 1, 0)x(0, 0, 1, 12)7 - AIC:1145.4230827207164
SARIMA(0, 1, 0)x(0, 0, 2, 12)7 - AIC:976.4375296380889
```

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

```
SARIMA(0, 1, 0)x(0, 0, 3, 12)7 - AIC:3942.6338417126094
SARIMA(0, 1, 0)x(1, 0, 0, 12)7 - AIC:1139.921738995602
SARIMA(0, 1, 0)x(1, 0, 1, 12)7 - AIC:1116.0207869386063
SARIMA(0, 1, 0)x(1, 0, 2, 12)7 - AIC:969.6913635753214
SARIMA(0, 1, 0)x(1, 0, 3, 12)7 - AIC:3910.656358528058
SARIMA(0, 1, 0)x(2, 0, 0, 12)7 - AIC:960.8812220353041
SARIMA(0, 1, 0)x(2, 0, 1, 12)7 - AIC:962.8794540697533
SARIMA(0, 1, 0)x(2, 0, 2, 12)7 - AIC:955.5735408945662
```

Fig 1.87 Sample AIC  for SARIMA Model

| | param | seasonal | AIC |
|---|---|---|---|
| **222** | (3, 1, 1) | (3, 0, 2, 12) | 774.400286 |
| **238** | (3, 1, 2) | (3, 0, 2, 12) | 774.880940 |
| **220** | (3, 1, 1) | (3, 0, 0, 12) | 775.426699 |
| **221** | (3, 1, 1) | (3, 0, 1, 12) | 775.495330 |
| **252** | (3, 1, 3) | (3, 0, 0, 12) | 775.561019 |

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, 2, 12)   Log Likelihood                -377.200
Date:                            Sun, 20 Feb 2022   AIC                            774.400
Time:                                    23:13:42   BIC                            799.618
Sample:                                         0   HQIC                           784.578
                                            - 132
Covariance Type:                              opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.0464      0.126      0.367      0.714      -0.202       0.294
ar.L2         -0.0060      0.120     -0.050      0.960      -0.241       0.229
ar.L3         -0.1808      0.098     -1.838      0.066      -0.374       0.012
ma.L1         -0.9370      0.067    -13.905      0.000      -1.069      -0.805
ar.S.L12       0.7639      0.165      4.640      0.000       0.441       1.087
ar.S.L24       0.0840      0.159      0.527      0.598      -0.229       0.397
ar.S.L36       0.0727      0.095      0.764      0.445      -0.114       0.259
ma.S.L12      -0.4969      0.250     -1.988      0.047      -0.987      -0.007
ma.S.L24      -0.2191      0.210     -1.044      0.296      -0.630       0.192
sigma2       192.1502     39.626      4.849      0.000     114.484     269.817
===================================================================================
Ljung-Box (Q):                       34.23   Jarque-Bera (JB):                 1.64
Prob(Q):                              0.73   Prob(JB):                         0.44
Heteroskedasticity (H):               1.11   Skew:                             0.33
Prob(H) (two-sided):                  0.77   Kurtosis:                         3.03
===================================================================================

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

Fig 1.89 Summary report for SARIMA Model

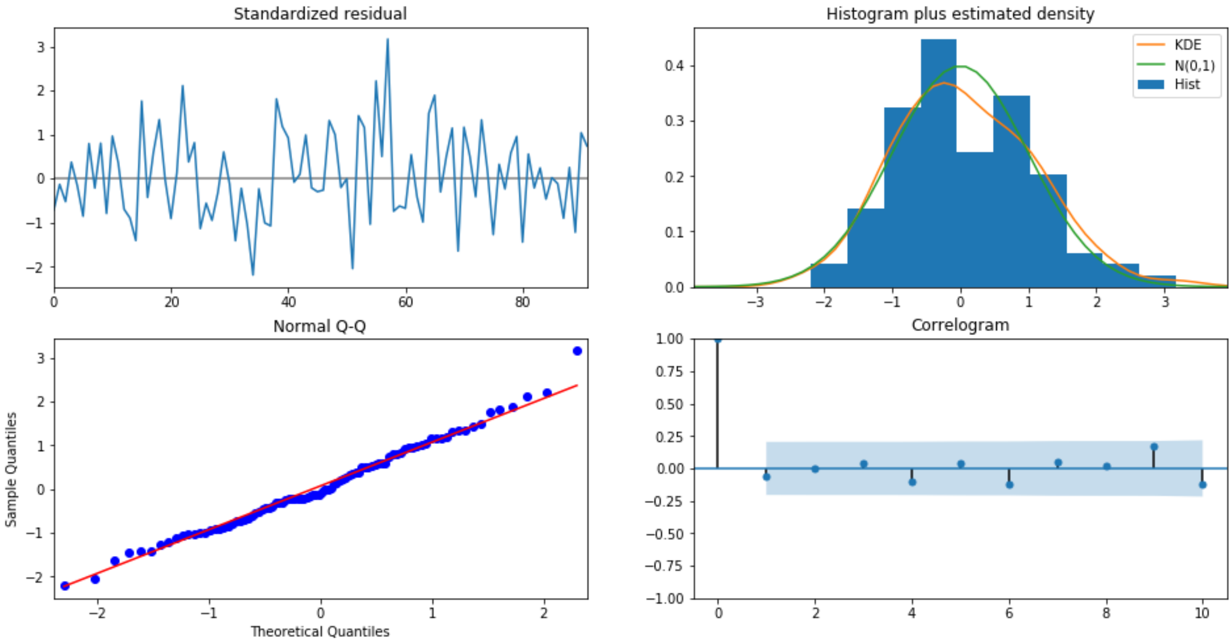



Fig 1.90 plotting diagnostic for SARIMA Model

```
array([ 55.23575577,  68.12261097,  67.90878015,  66.78622912,
        69.76047168,  70.3289832 ,  75.35954438,  76.49217979,
        78.97132197,  76.53862693,  93.24904879, 116.28317142,
        55.20248119,  64.44406372,  68.54777414,  63.87232536,
        67.70018699,  68.44357115,  72.97210368,  74.32550441,
        75.31784431,  76.04683595,  87.42127189, 109.80717304,
        51.29829317,  62.61706238,  65.91208553,  62.26431815,
        64.61216749,  65.74752277,  69.82629533,  70.42010232,
        72.33171215,  71.3657079 ,  84.72221947, 105.35778471,
        49.94997509,  60.31173938,  63.15821132,  59.89686303,
        62.22798764,  63.19905767,  67.06122628,  67.71097974,
        69.43493489,  68.58134086,  80.95530204, 100.27484656,
        48.59076447,  58.12924132,  60.87894865,  57.74110091,
        59.99748694,  60.88874594,  64.51109411])
```

Fig 1.91 Predicting values for testing data

```
Test rmse for sarima is  18.882003816724545
Test mape for sarima is  32.02
```

Fig 1.92 RMSE and MAPE SARIMA Model

| | Test RMSE | Test MAPE |
|---|---|---|
| RegressionOnTime | 15.268955 | 22.82 |
| NaiveModel | 79.718773 | 145.10 |
| SimpleAverageModel | 15.759783 | 21.37 |
| 2pointTrailingMovingAverage | 11.529278 | 13.54 |
| 4pointTrailingMovingAverage | 14.451403 | 19.49 |
| 6pointTrailingMovingAverage | 14.566327 | 20.82 |
| 9pointTrailingMovingAverage | 14.727630 | 21.01 |
| Alpha=0.0987,SimpleExponentialSmoothing | 36.796242 | 63.88 |
| Alpha=0.01,beta=0.01,DoubleExponentialSmoothing | 45.005037 | 79.29 |
| Alpha=0.10,Beta=0.1,DoubleExponentialSmoothing | 36.923416 | 63.78 |
| Alpha: 0.99,Beta: 0.01 and Gamma:0.01,TripleExponentialSmoothing | 69.035167 | 123.85 |
| Alpha=0.01,Beta=0.76,Gamma=0.06,TripleExponentialSmoothingWithGrid | 14.949445 | 19.69 |
| Arima 3,1,3 | 15.988955 | 26.09 |
| SARIMA(3, 1, 1)(3, 0, 2, 12) | 18.882004 | 32.02 |

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.

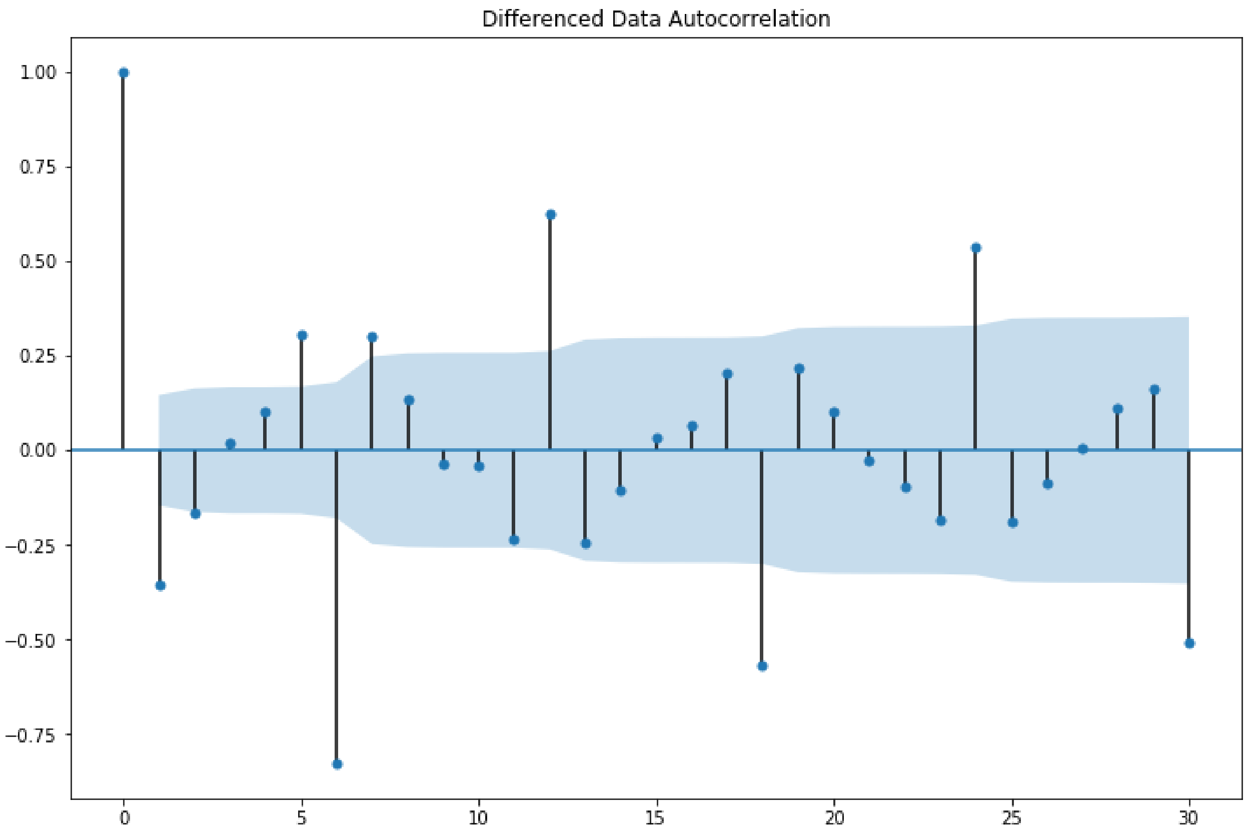



Fig 1.94 Plotting 2nd Difference after taking cut-off points in ACF plot

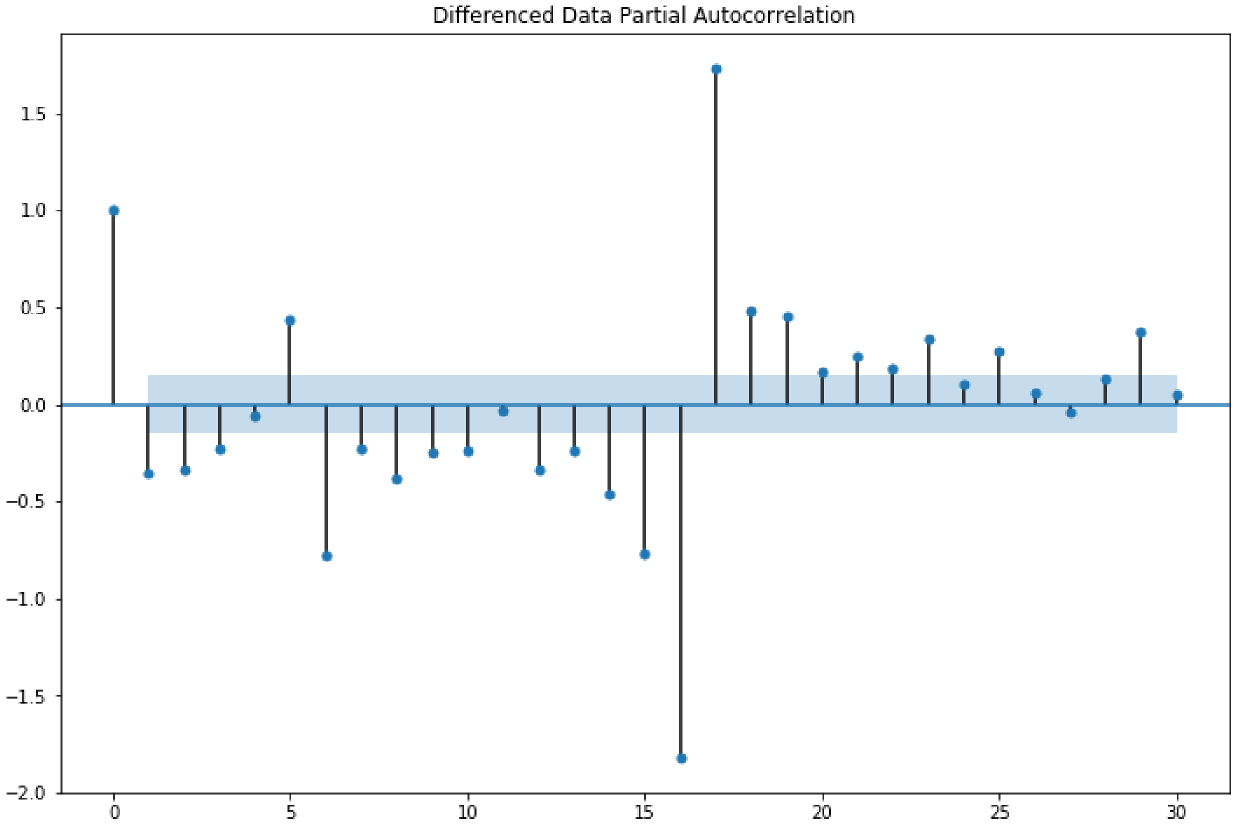



Fig 1.95 Plotting 2nd Difference after taking cut-off points in PACF plot

```
                             ARIMA Model Results
==============================================================================
Dep. Variable:                 D.Rose   No. Observations:                  131
Model:                 ARIMA(2, 1, 3)   Log Likelihood                -633.598
Method:                       css-mle   S.D. of innovations             29.971
Date:                Sun, 20 Feb 2022   AIC                           1281.196
Time:                        23:13:44   BIC                           1301.323
Sample:                    02-29-1980   HQIC                          1289.374
                         - 12-31-1990
================================================================================
                   coef    std err          z      P>|z|      [0.025      0.975]
--------------------------------------------------------------------------------
const           -0.4893      0.084     -5.849      0.000      -0.653      -0.325
ar.L1.D.Rose    -0.0605      0.833     -0.073      0.942      -1.694       1.573
ar.L2.D.Rose     0.2075      0.360      0.577      0.565      -0.497       0.912
ma.L1.D.Rose    -0.7052      0.833     -0.847      0.399      -2.337       0.927
ma.L2.D.Rose    -0.5471      0.381     -1.436      0.153      -1.294       0.200
ma.L3.D.Rose     0.2522      0.523      0.482      0.631      -0.774       1.278
                                    Roots
=============================================================================
                  Real          Imaginary           Modulus         Frequency
-----------------------------------------------------------------------------
AR.1           -2.0544           +0.0000j            2.0544            0.5000
AR.2            2.3458           +0.0000j            2.3458            0.0000
MA.1           -1.4907           +0.0000j            1.4907            0.5000
MA.2            1.0000           +0.0000j            1.0000            0.0000
MA.3            2.6596           +0.0000j            2.6596            0.0000
-----------------------------------------------------------------------------
```

Fig 1.96 Cut-off point summary report for ARIMA model

```
array([80.29349704, 70.94695747, 72.73026545, 70.26558252, 70.36730806,
       69.43233762, 69.09260729, 68.50175729, 68.04960992, 67.53696527,
       67.05676089, 66.56204106, 66.07493062, 65.58434794, 65.09555424,
       64.60593184, 64.11673078, 63.62733227, 63.13803314, 62.64868703,
       62.15936438, 61.67003056, 61.18070229, 60.69137136, 60.20204175,
       59.7127115 , 59.22338157, 58.73405148, 58.24472147, 57.75539143,
       57.2660614 , 56.77673136, 56.28740133, 55.79807129, 55.30874126,
       54.81941122, 54.33008119, 53.84075115, 53.35142112, 52.86209108,
       52.37276105, 51.88343101, 51.39410098, 50.90477094, 50.41544091,
       49.92611087, 49.43678084, 48.9474508 , 48.45812077, 47.96879073,
       47.4794607 , 46.99013066, 46.50080063, 46.01147059, 45.52214056])
```

Fig 1.97 predicting values for ARIMA model after taking cutoff points

```
Test rmse for arima is  15.48824414782164
Test mape for arima is  23.15
```

Fig 1.98 RMSE and MAPE value for cutoff point ARIMA model

| | Test RMSE | Test MAPE |
|---|---|---|
| RegressionOnTime | 15.268955 | 22.82 |
| NaiveModel | 79.718773 | 145.10 |
| SimpleAverageModel | 15.759783 | 21.37 |
| 2pointTrailingMovingAverage | 11.529278 | 13.54 |
| 4pointTrailingMovingAverage | 14.451403 | 19.49 |
| 6pointTrailingMovingAverage | 14.566327 | 20.82 |
| 9pointTrailingMovingAverage | 14.727630 | 21.01 |
| Alpha=0.0987,SimpleExponentialSmoothing | 36.796242 | 63.88 |
| Alpha=0.01,beta=0.01,DoubleExponentialSmoothing | 45.005037 | 79.29 |
| Alpha=0.10,Beta=0.1,DoubleExponentialSmoothing | 36.923416 | 63.78 |
| Alpha: 0.99,Beta: 0.01 and Gamma:0.01,TripleExponentialSmoothing | 69.035167 | 123.85 |
| Alpha=0.01,Beta=0.76,Gamma=0.06,TripleExponentialSmoothingWithGrid | 14.949445 | 19.69 |
| Arima 3,1,3 | 15.988955 | 26.09 |
| SARIMA(3, 1, 1)(3, 0, 2, 12) | 18.882004 | 32.02 |
| Arima 2,1,3 | 15.488244 | 23.15 |

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

```
                                      Statespace Model Results
==========================================================================================
Dep. Variable:                                  y   No. Observations:                  132
Model:             SARIMAX(4, 1, 4)x(4, 0, 2, 12)   Log Likelihood                -321.321
Date:                            Sun, 20 Feb 2022   AIC                            672.641
Time:                                    23:14:10   BIC                            708.183
Sample:                                         0   HQIC                           686.880
                                            - 132
Covariance Type:                              opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.1232      0.223      0.552      0.581      -0.314       0.561
ar.L2          0.0844      0.248      0.340      0.734      -0.402       0.571
ar.L3         -0.6638      0.228     -2.913      0.004      -1.110      -0.217
ar.L4         -0.2091      0.119     -1.753      0.080      -0.443       0.025
ma.L1         -1.1735     47.833     -0.025      0.980     -94.924      92.577
ma.L2          0.2007     83.335      0.002      0.998    -163.134     163.535
ma.L3          0.9311     37.707      0.025      0.980     -72.973      74.836
ma.L4         -0.8167     67.345     -0.012      0.990    -132.810     131.176
ar.S.L12       0.2239      0.176      1.270      0.204      -0.122       0.569
ar.S.L24       0.6525      0.131      4.986      0.000       0.396       0.909
ar.S.L36       0.1138      0.093      1.219      0.223      -0.069       0.297
ar.S.L48      -0.0318      0.093     -0.343      0.732      -0.214       0.150
ma.S.L12       0.1108     38.624      0.003      0.998     -75.590      75.812
ma.S.L24      -0.8922     34.201     -0.026      0.979     -67.925      66.140
sigma2       131.3087   1.22e+04      0.011      0.991   -2.37e+04     2.4e+04
===================================================================================
Ljung-Box (Q):                       47.59   Jarque-Bera (JB):                     0.73
Prob(Q):                              0.19   Prob(JB):                             0.69
Heteroskedasticity (H):               0.73   Skew:                                 0.12
Prob(H) (two-sided):                  0.42   Kurtosis:                             3.40
===================================================================================

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

Fig 1.100 Cut-off point for SARIMA model summary report

```
Test rmse for sarima is  17.28613245134336
Test mape for sarima is  27.69
```

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

| | Test RMSE | Test MAPE |
|---|---|---|
| RegressionOnTime | 15.268955 | 22.82 |
| NaiveModel | 79.718773 | 145.10 |
| SimpleAverageModel | 15.759783 | 21.37 |
| 2pointTrailingMovingAverage | 11.529278 | 13.54 |
| 4pointTrailingMovingAverage | 14.451403 | 19.49 |
| 6pointTrailingMovingAverage | 14.566327 | 20.82 |
| 9pointTrailingMovingAverage | 14.727630 | 21.01 |
| Alpha=0.0987,SimpleExponentialSmoothing | 36.796242 | 63.88 |
| Alpha=0.01,beta=0.01,DoubleExponentialSmoothing | 45.005037 | 79.29 |
| Alpha=0.10,Beta=0.1,DoubleExponentialSmoothing | 36.923416 | 63.78 |
| Alpha: 0.99,Beta: 0.01 and Gamma:0.01,TripleExponentialSmoothing | 69.035167 | 123.85 |
| Alpha=0.01,Beta=0.76,Gamma=0.06,TripleExponentialSmoothingWithGrid | 14.949445 | 19.69 |
| Arima 3,1,3 | 15.988955 | 26.09 |
| SARIMA(3, 1, 1)(3, 0, 2, 12) | 18.882004 | 32.02 |
| Arima 2,1,3 | 15.488244 | 23.15 |
| SARIMA(4, 1, 4)(4, 0, 2, 12) | 17.286132 | 27.69 |

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 |
|---|---|---|
| 2pointTrailingMovingAverage | 11.529278 | 13.54 |
| 4pointTrailingMovingAverage | 14.451403 | 19.49 |
| 6pointTrailingMovingAverage | 14.566327 | 20.82 |
| 9pointTrailingMovingAverage | 14.727630 | 21.01 |
| Alpha=0.01,Beta=0.76,Gamma=0.06,TripleExponentialSmoothingWithGrid | 14.949445 | 19.69 |
| RegressionOnTime | 15.268955 | 22.82 |
| Arima 2,1,3 | 15.488244 | 23.15 |
| SimpleAverageModel | 15.759783 | 21.37 |
| Arima 3,1,3 | 15.988955 | 26.09 |
| SARIMA(4, 1, 4)(4, 0, 2, 12) | 17.286132 | 27.69 |
| SARIMA(3, 1, 1)(3, 0, 2, 12) | 18.882004 | 32.02 |
| Alpha=0.0987,SimpleExponentialSmoothing | 36.796242 | 63.88 |
| Alpha=0.10,Beta=0.1,DoubleExponentialSmoothing | 36.923416 | 63.78 |
| Alpha=0.01,beta=0.01,DoubleExponentialSmoothing | 45.005037 | 79.29 |
| Alpha: 0.99,Beta: 0.01 and Gamma:0.01,TripleExponentialSmoothing | 69.035167 | 123.85 |
| NaiveModel | 79.718773 | 145.10 |

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.

Although we are seeing that the best model as per RMSE is the 2pointTrailingMovingAverage, 4pointTrailingMovingAverage, 6pointTrailingMovingAverage, 9pointTrailingMovingAverage as it is giving us the least RMSE value. But the moving average models are actually quite a naive model and assumes that the trend and seasonality components of the time series have already been removed or adjusted for. Hence we will going to choose the second best model which comes out to be Triple Exponential Model. Its RMSE value is very close to the 2point Trailing Moving Average , 4point Trailing Moving Average, 6point Trailing Moving Average, 9point Trailing Moving Average value hence we can choose this model as well.

```
RMSE: 29.39707075749173
```

Fig 1.104 Summary report for SARIMA model

```
1995-08-31     34.206001
1995-09-30     64.244438
1995-10-31     42.369170
1995-11-30     52.886688
1995-12-31    121.321299
1996-01-31     14.026713
1996-02-29     23.987492
1996-03-31     33.740397
1996-04-30     20.696789
1996-05-31     28.215274
1996-06-30     51.634478
1996-07-31     38.941272
Freq: M, dtype: float64
```

Fig 1.105 plotting diagnostic report for SARIMA model

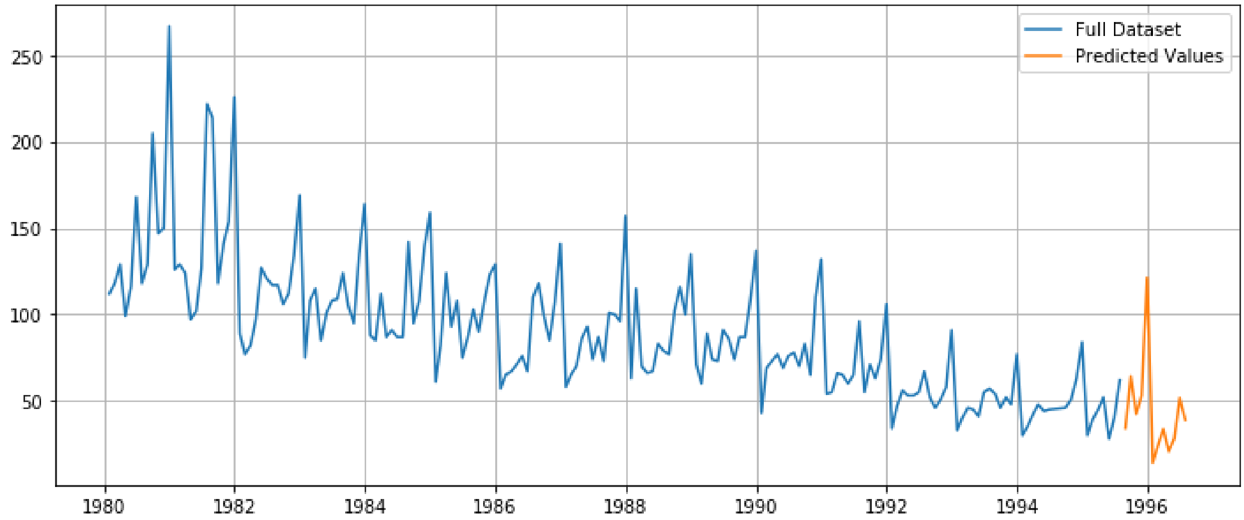



Fig 1.106 plotting diagnostic report for SARIMA model

***Predicting the values for the next 12 months in future***

| | lower_CI | prediction | upper_ci |
|---|---|---|---|
| **1995-08-31** | 18.345018 | 34.206001 | 139.831163 |
| **1995-09-30** | 48.383455 | 64.244438 | 169.869601 |
| **1995-10-31** | 26.508186 | 42.369170 | 147.994332 |
| **1995-11-30** | 37.025705 | 52.886688 | 158.511851 |
| **1995-12-31** | 105.460316 | 121.321299 | 226.946462 |

Fig 1.107 Predicting the values

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

Fig 1.108 RMSE value for future data.

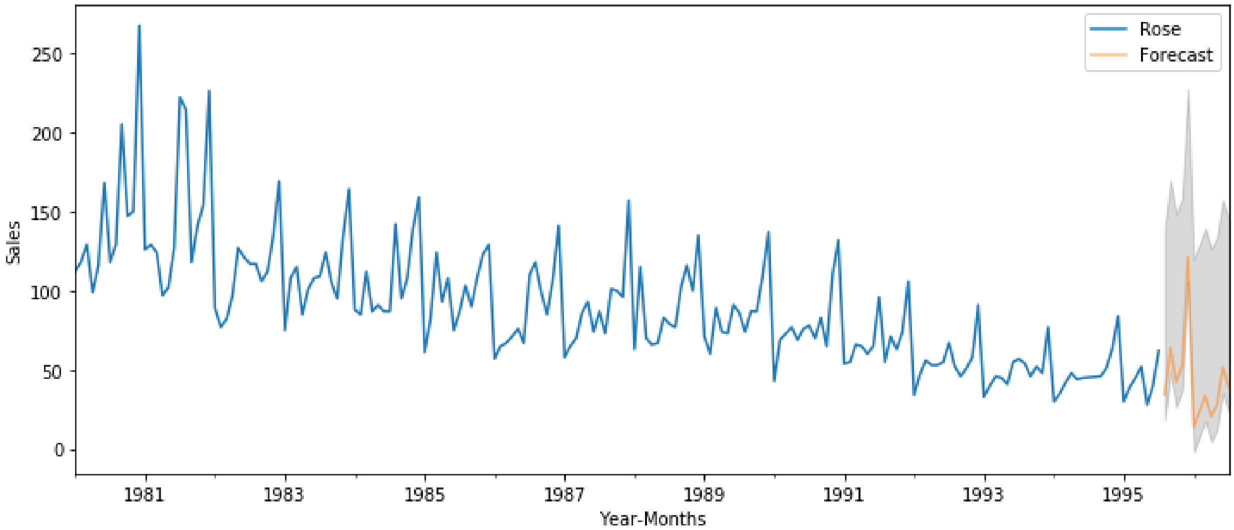



Fig 1.109 Plotting the future data in graph.

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

Triple Exponential 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 this insights to increase our sales further. We can introduce certain offers in other months to attract more customers.

Year 1981 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 are declining each and every year.The reason behind this may be either deterioration in quality or arrival of any rival product in the market.Hence some important measures have to be taken to identify the reason and tale appropriate measures to increase the trend.