9/25/2022

Time Series Forecasting

ROSE.CSV

Name – Vivek Augustine

Table of Contents:

Problem:

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

Data set for the Problem: Rose.csv

Please do perform the following questions on this set.

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

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

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

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

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

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

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

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

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

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

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

TABLE :

YearMonth	*Rose*
1980-01	112.
1980-02	118.
1980-03	129.
1980-04	99.
1980-05	116

Description :

	Rose
count	185.000000
mean	90.394595
std	39.175344
min	28.000000
25%	63.000000
50%	86.000000
75%	112.000000
max	267.000000

SHAPE :

```
(187, 2)
```

Duplicates :

```
0
```

Null Values :

```
YearMonth    0
Rose         2
```

Table :

	YearMonth	Rose	Date
0	1980-01	112.0	1980-01-31
1	1980-02	118.0	1980-02-29
2	1980-03	129.0	1980-03-31
3	1980-04	99.0	1980-04-30
4	1980-05	116.0	1980-05-31

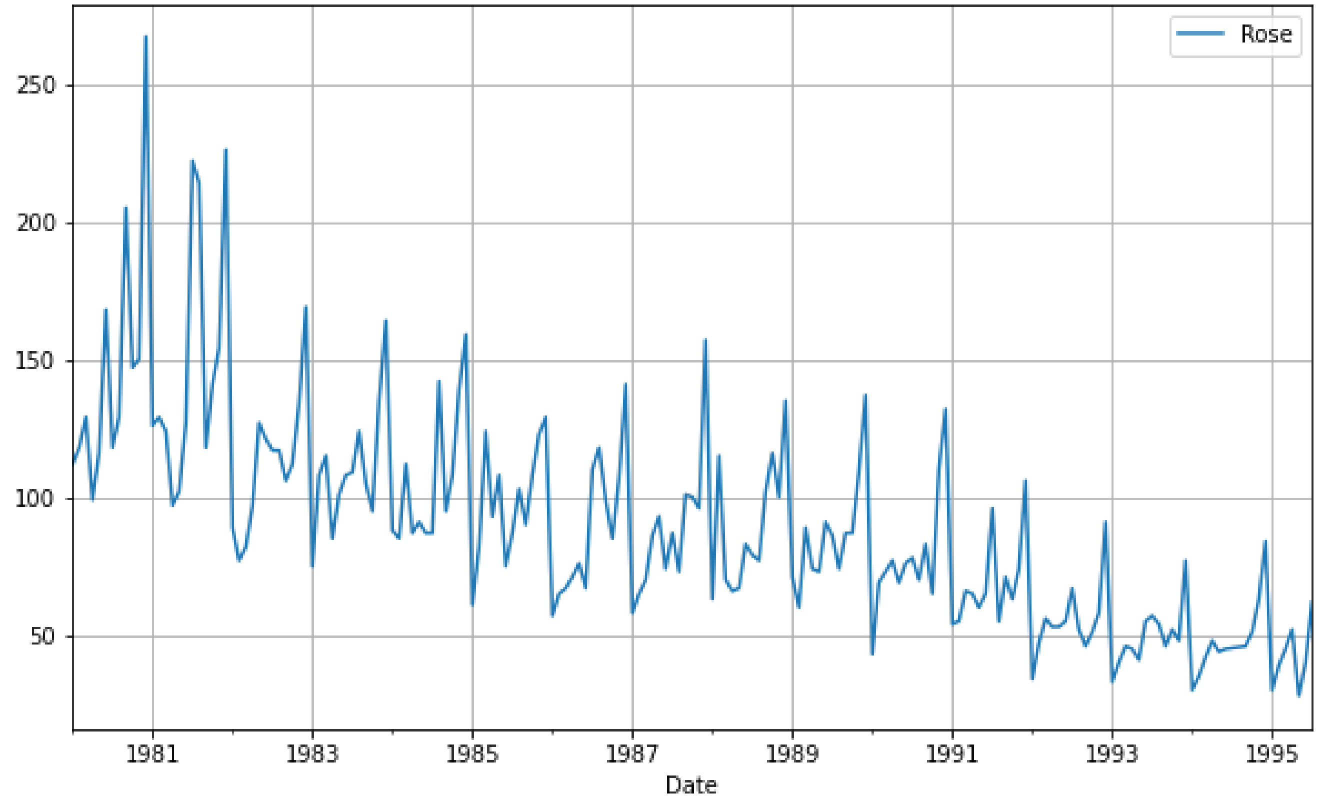
Information:

```
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   Rose    187 non-null    float64
 1   Date    187 non-null    datetime64[ns]
```

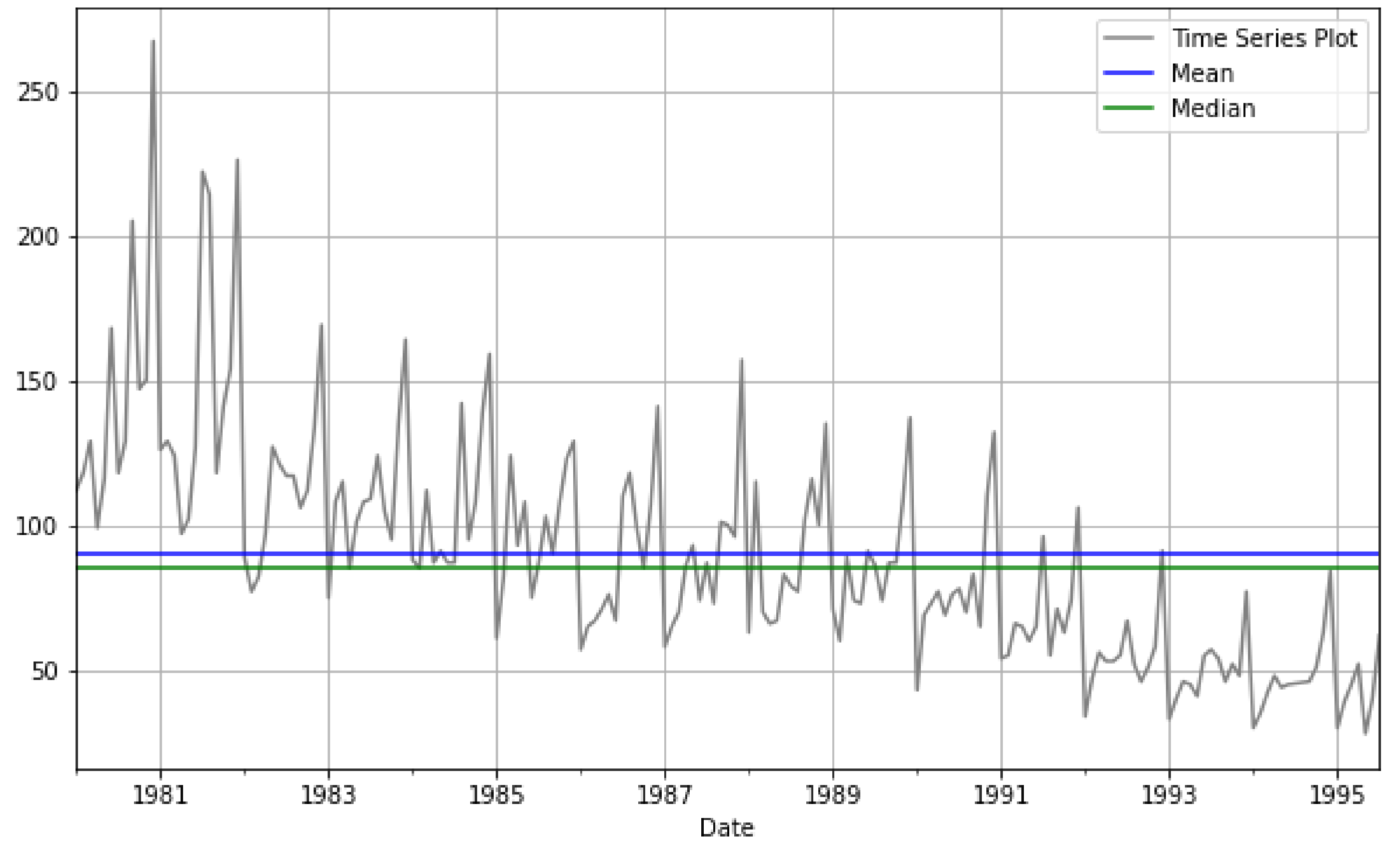
Setting Index :

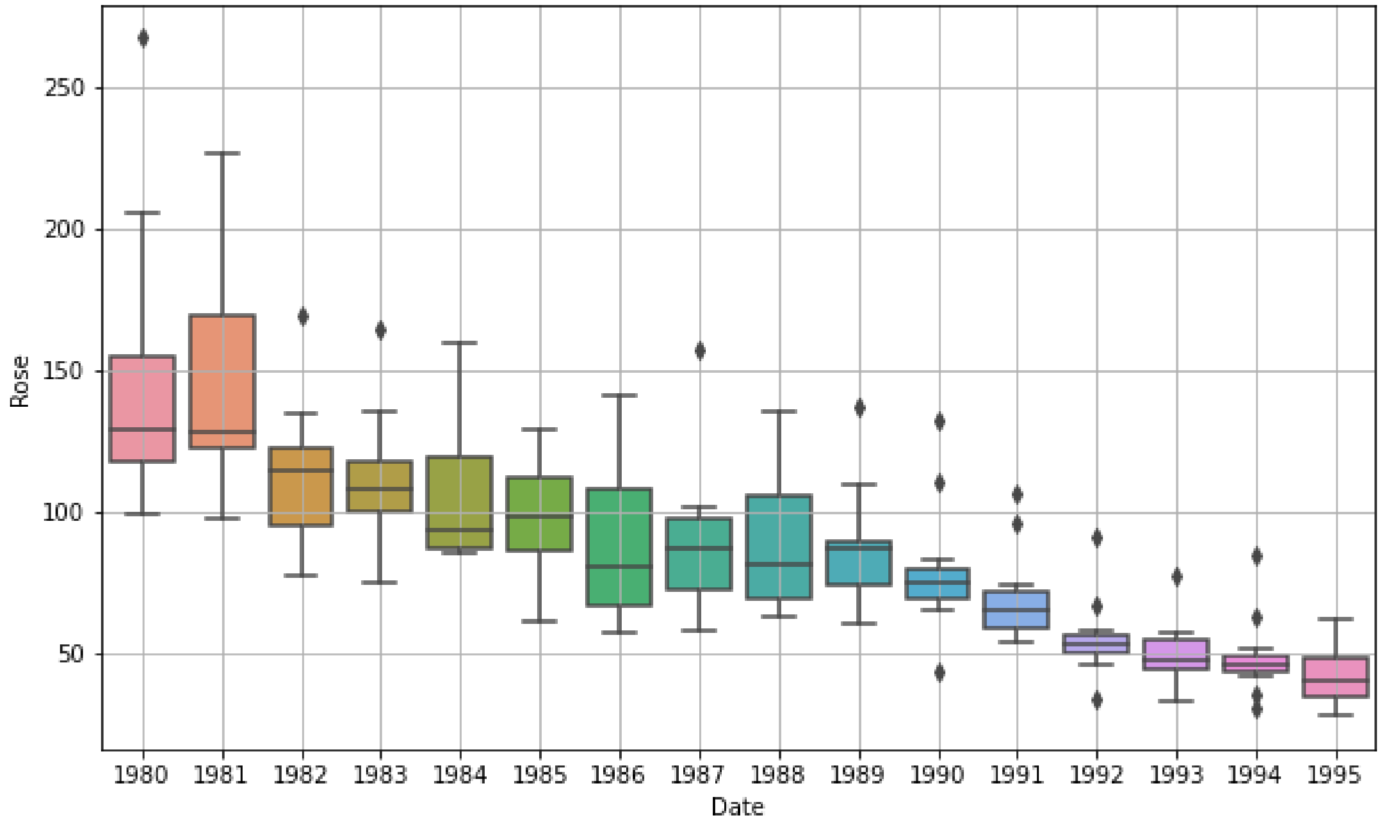
	Rose
Date	
1980-01-31	112.0
1980-02-29	118.0
1980-03-31	129.0
1980-04-30	99.0
1980-05-31	116.0

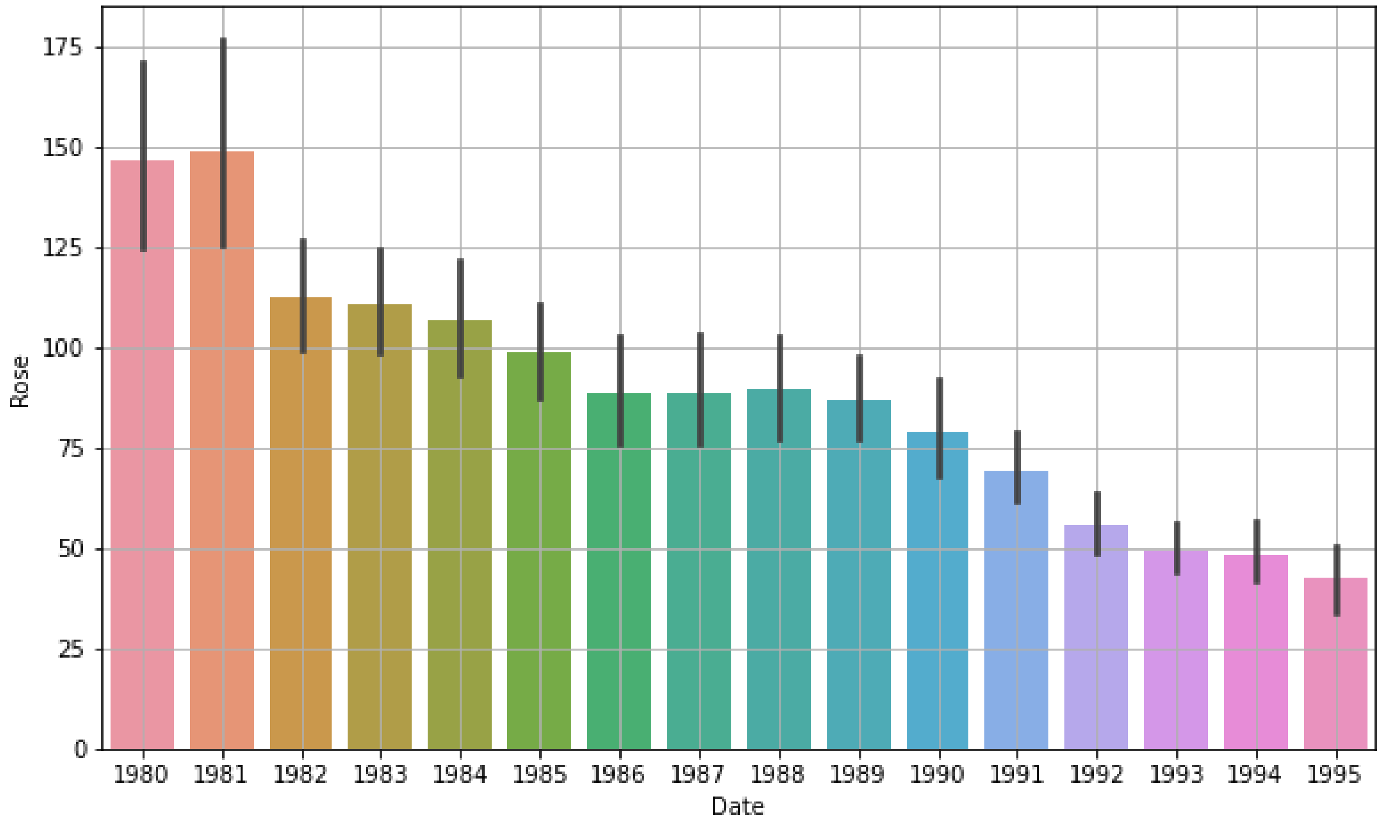
Plotting Data :



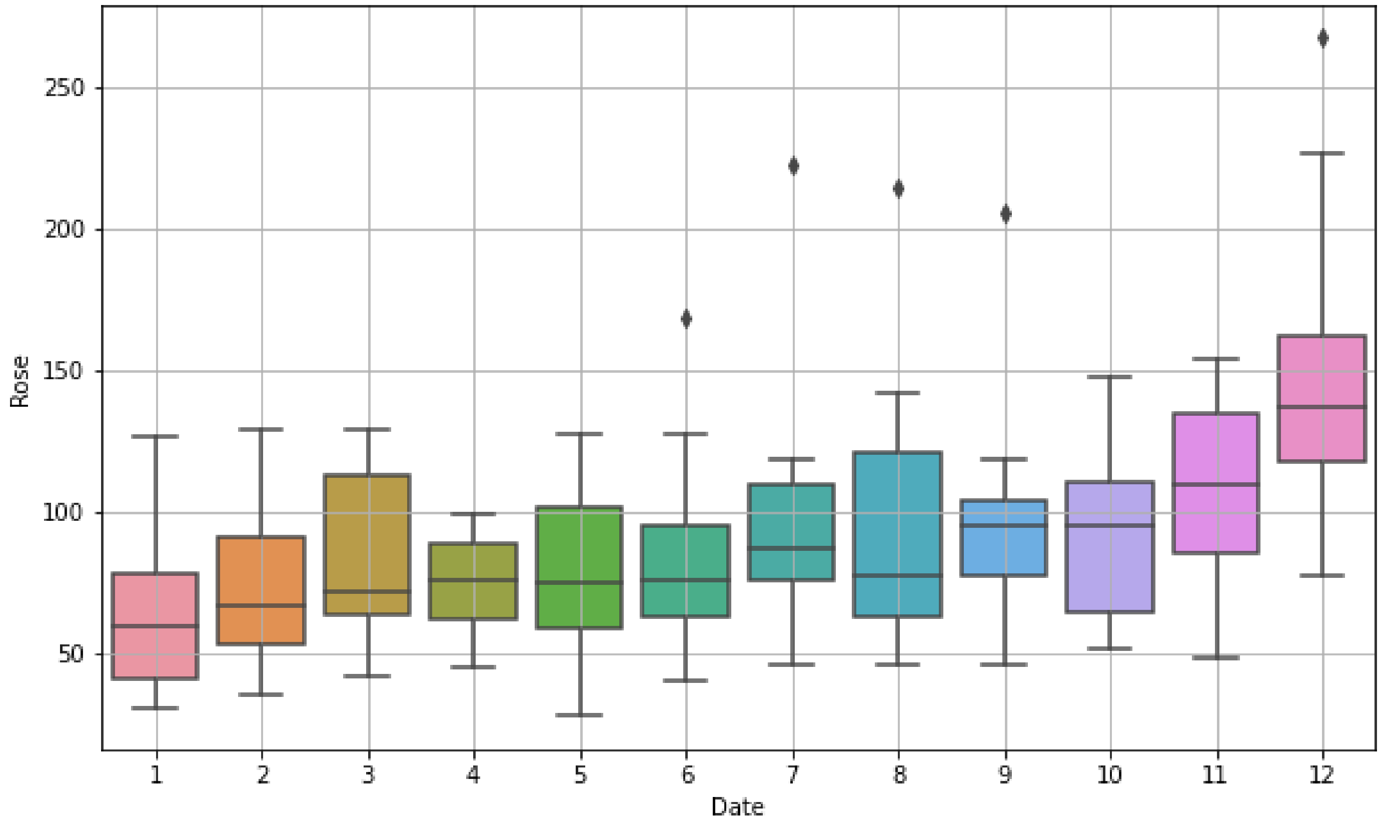
Q.2 Perform appropriate Exploratory Data Analysis to understand the data and also perform decomposition.

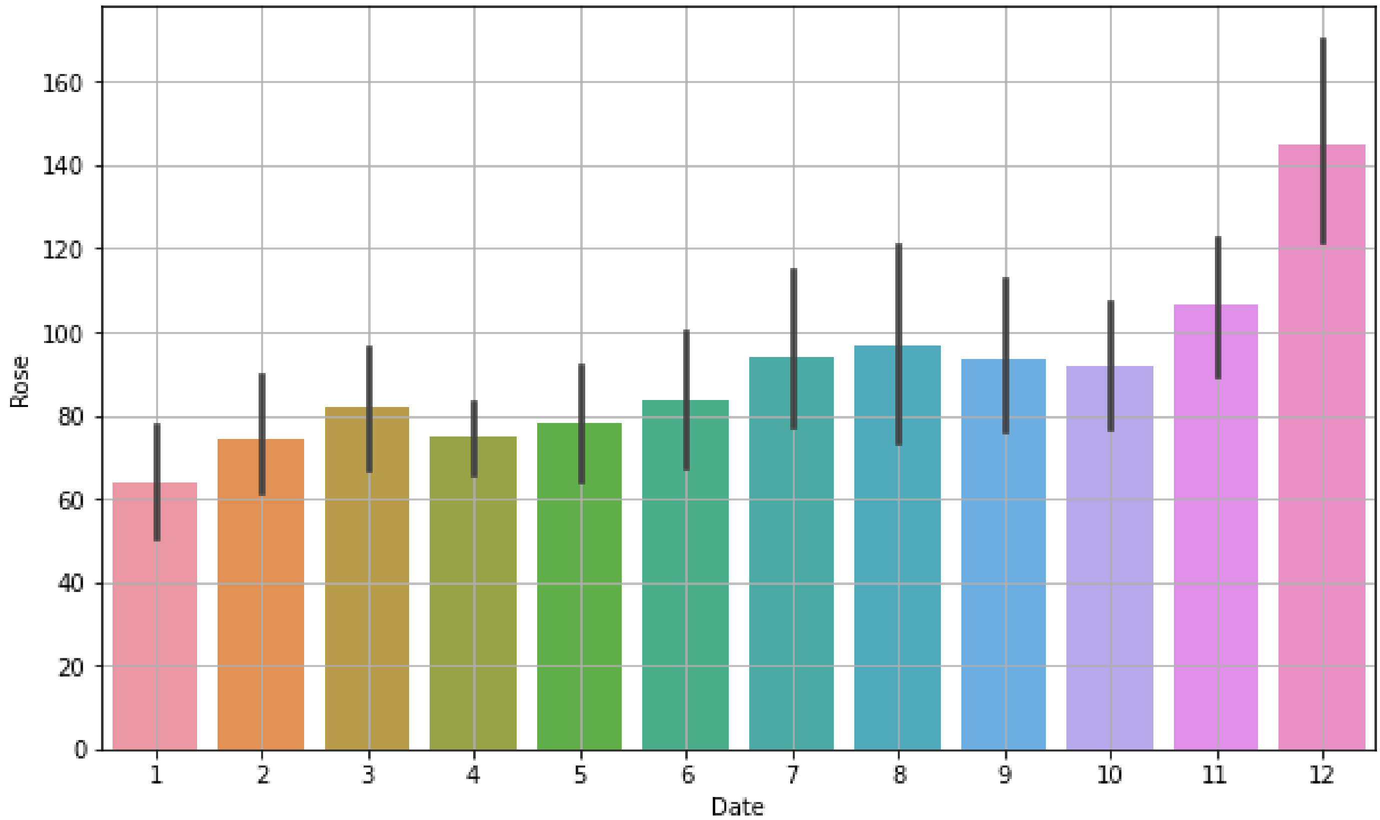




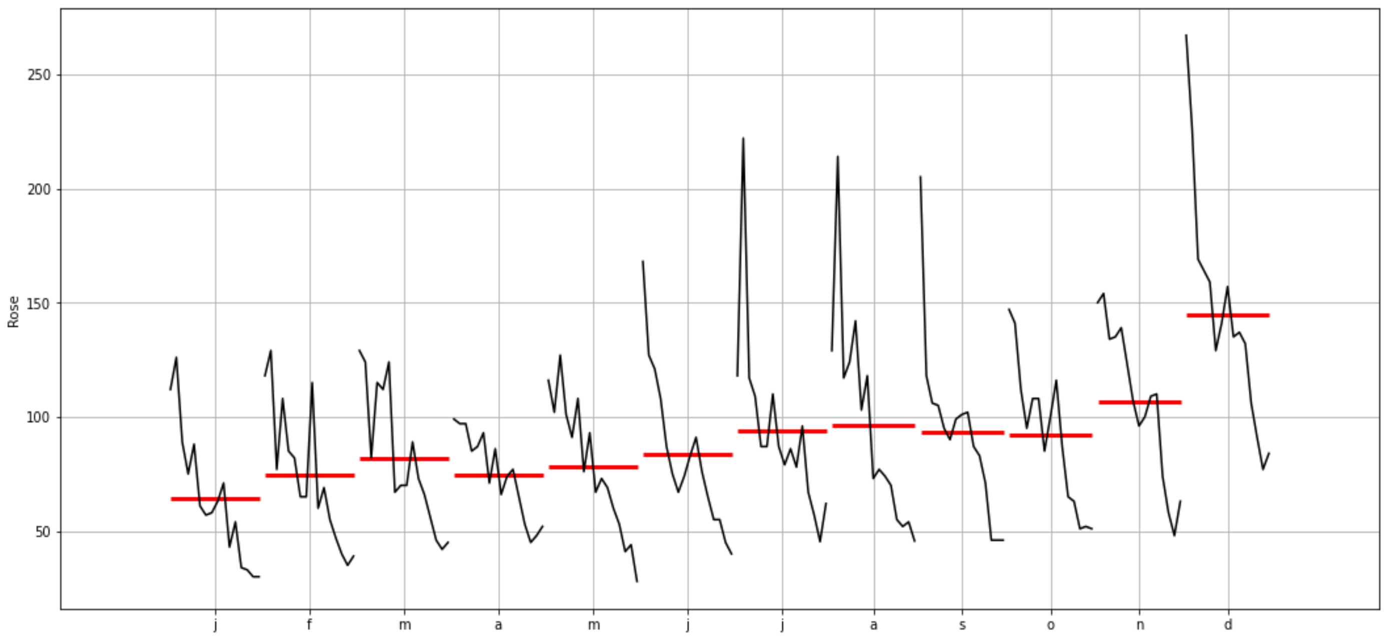


Here, we can see the year of 1981 is recording the highest sale in the year and after that we can see that the sales kept on decreasing the year of 1986, 1987 & 1988 also did pretty decently in terms of sales but if we see the overall trend of the sales of wine it's not good as it kept on decreasing with passing time and need some measures in order to keep up in the market.

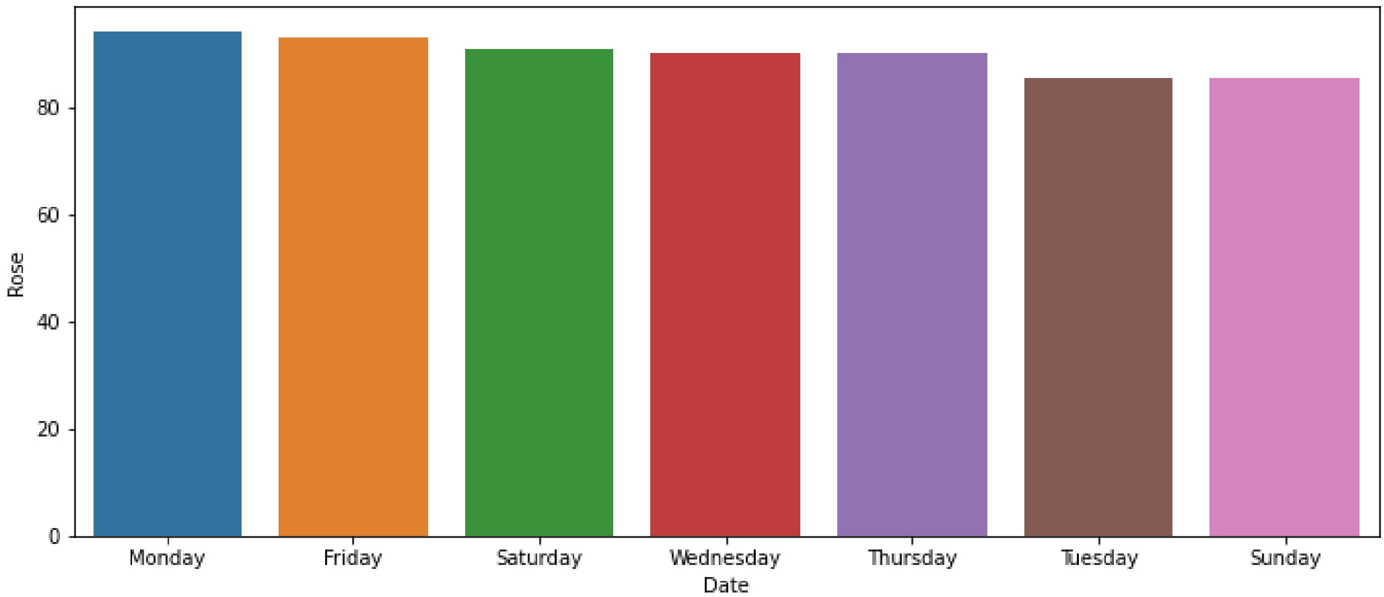




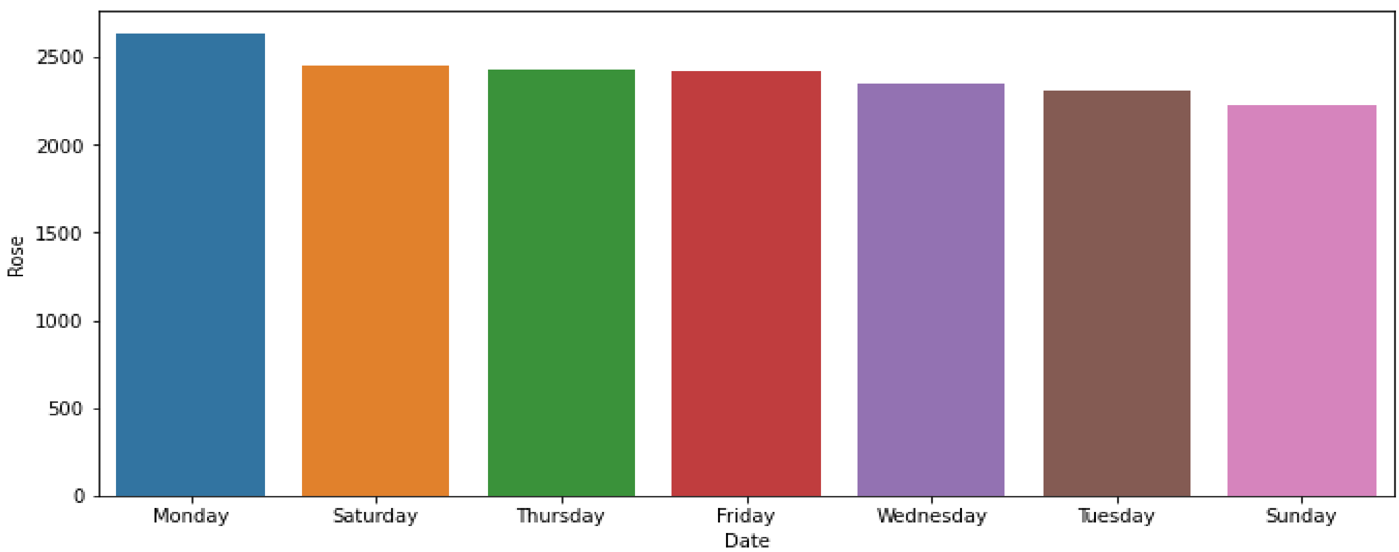
Here, we can see monthly sales from this graph we can make out that the month of december is recording the highest sales followed by November and the month of july, august, september is doing pretty decent in terms of sale. the month january is recorded as the month with the lowest of the sale for this we have to do some reaserch as to why the first 5 months are recording very low selling of wine.



sales of beer throughout the whole week.



highest sales overall

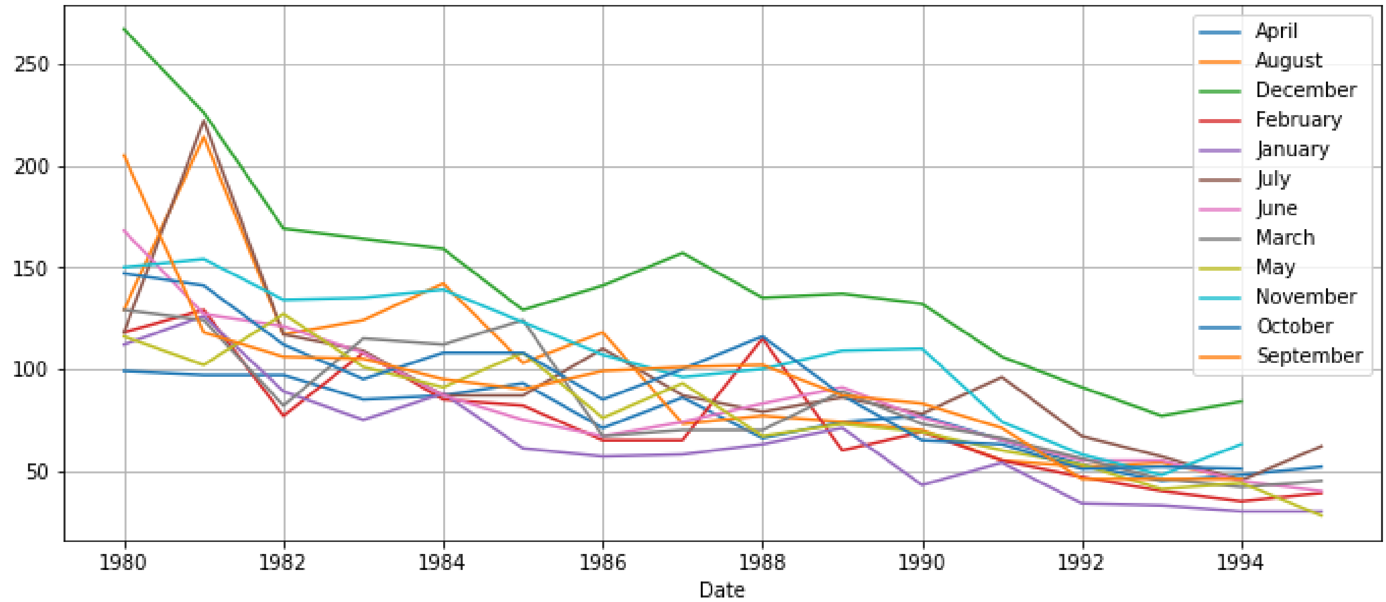


Pivot Table:

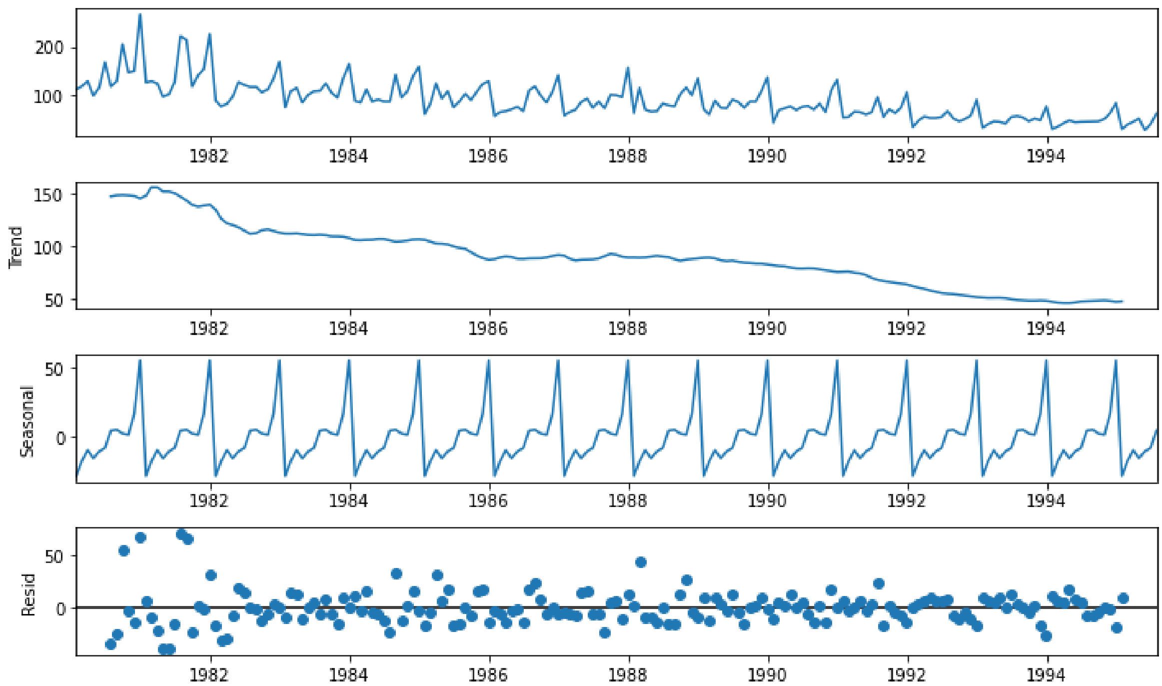
Date	April	August	December	February	January	July	June	March	May	November	October	September
Date												
1980	99.0	129.000000	267.0	118.0	112.0	118.000000	168.0	129.0	116.0	150.0	147.0	205.0
1981	97.0	214.000000	226.0	129.0	126.0	222.000000	127.0	124.0	102.0	154.0	141.0	118.0
1982	97.0	117.000000	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

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

Monthly sales across years



Decompostion (Additive)



Trend, seasonality,residual :

Trend

```
 Date
1980-01-31           NaN
1980-02-29           NaN
1980-03-31           NaN
1980-04-30           NaN
1980-05-31           NaN
1980-06-30           NaN
1980-07-31    147.083333
1980-08-31    148.125000
1980-09-30    148.375000
1980-10-31    148.083333
1980-11-30    147.416667
1980-12-31    145.125000
```

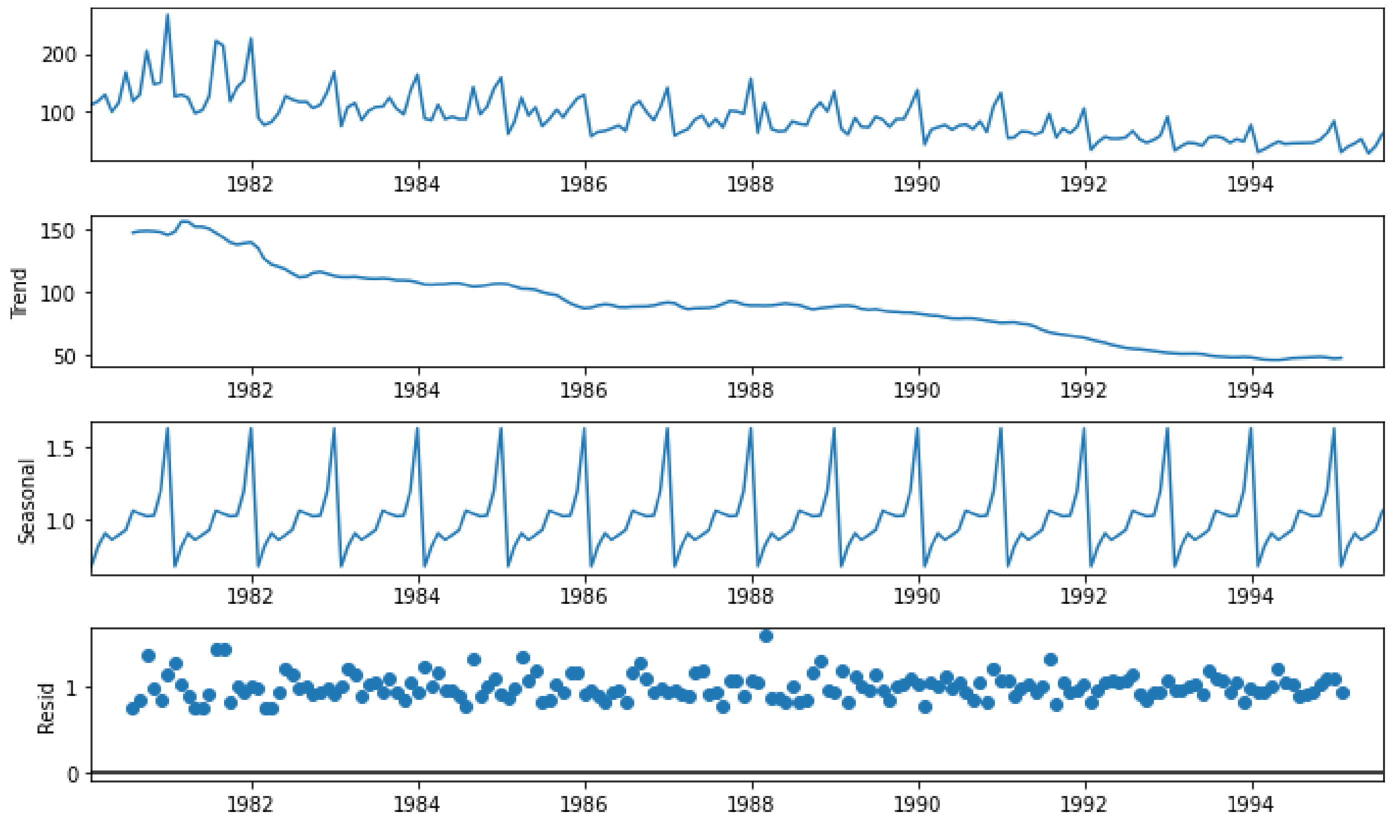
Seasonality

```
 Date
1980-01-31   -27.908647
1980-02-29   -17.435632
1980-03-31    -9.285830
1980-04-30   -15.098330
1980-05-31   -10.196544
1980-06-30    -7.678687
1980-07-31     4.896908
1980-08-31     5.499686
1980-09-30     2.774686
1980-10-31     1.871908
1980-11-30    16.846908
1980-12-31    55.713575
```

Residual

```
 Date
1980-01-31          NaN
1980-02-29          NaN
1980-03-31          NaN
1980-04-30          NaN
1980-05-31          NaN
1980-06-30          NaN
1980-07-31   -33.980241
1980-08-31   -24.624686
1980-09-30    53.850314
1980-10-31    -2.955241
1980-11-30   -14.263575
1980-12-31    66.161425
```

Decomposition(Multiplicative)



Trend, seasonality,residual :

```
Trend
 Date
1980-01-31          NaN
1980-02-29          NaN
1980-03-31          NaN
1980-04-30          NaN
1980-05-31          NaN
1980-06-30          NaN
1980-07-31   147.083333
1980-08-31   148.125000
1980-09-30   148.375000
1980-10-31   148.083333
1980-11-30   147.416667
1980-12-31   145.125000
```

```
Seasonality
 Date
1980-01-31   0.670111
1980-02-29   0.806163
1980-03-31   0.901164
1980-04-30   0.854024
1980-05-31   0.889415
1980-06-30   0.923985
1980-07-31   1.058038
1980-08-31   1.035881
1980-09-30   1.017648
1980-10-31   1.022573
1980-11-30   1.192349
```

```
1980-12-31    1.628646
```

Residual

```
 Date
1980-01-31         NaN
1980-02-29         NaN
1980-03-31         NaN
1980-04-30         NaN
1980-05-31         NaN
1980-06-30         NaN
1980-07-31    0.758258
1980-08-31    0.840720
1980-09-30    1.357674
1980-10-31    0.970771
1980-11-30    0.853378
1980-12-31    1.129646
```

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

Shape :

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

Rows:

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

	Rose
Date	
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

```
Last few rows of Training Data
```

Date	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

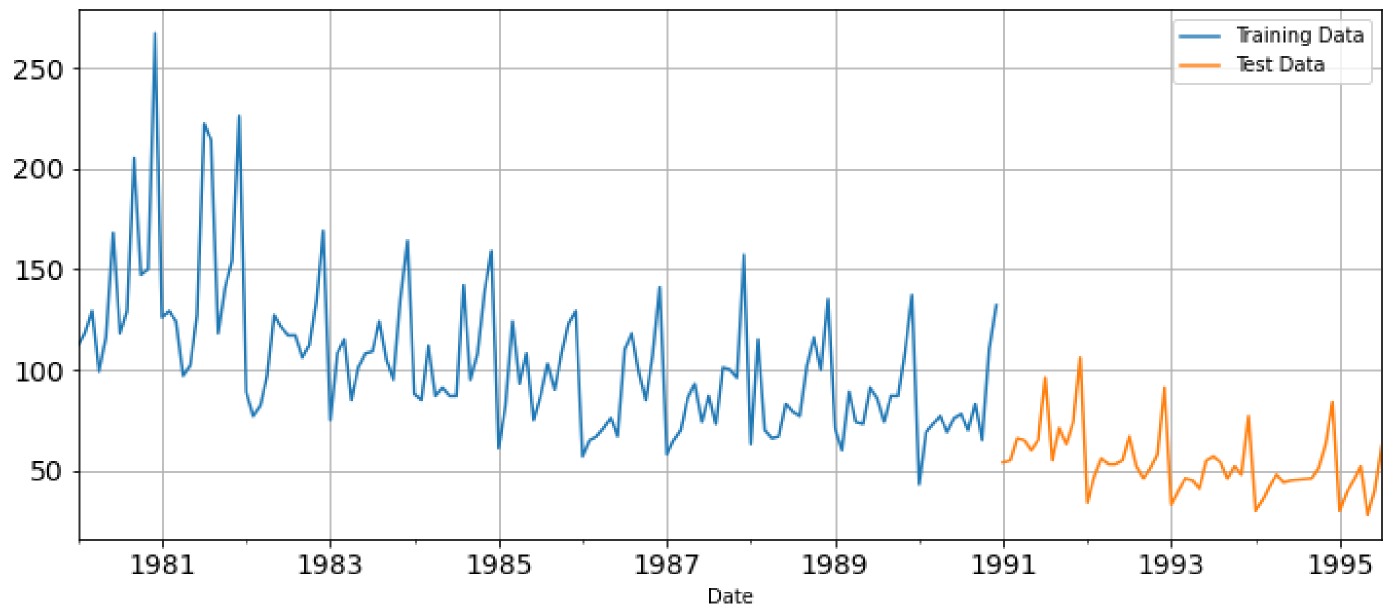
First few rows of Test Data

Date	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

Last few rows of Test Data

Date	Rose
1995-03-31	45.0
1995-04-30	52.0
1995-05-31	28.0
1995-06-30	40.0
1995-07-31	62.0

Plotting Training - Testing Data



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

Model 1: Linear Regression

Dividing the data:

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

LinearRegression train & test:

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

	Rose	time
Date		
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

```
Last few rows of Training Data
```

	Rose	Time
Date		
1990-08-31	70.0	128
1990-09-30	83.0	129
1990-10-31	65.0	130
1990-11-30	110.0	131
1990-12-31	132.0	132

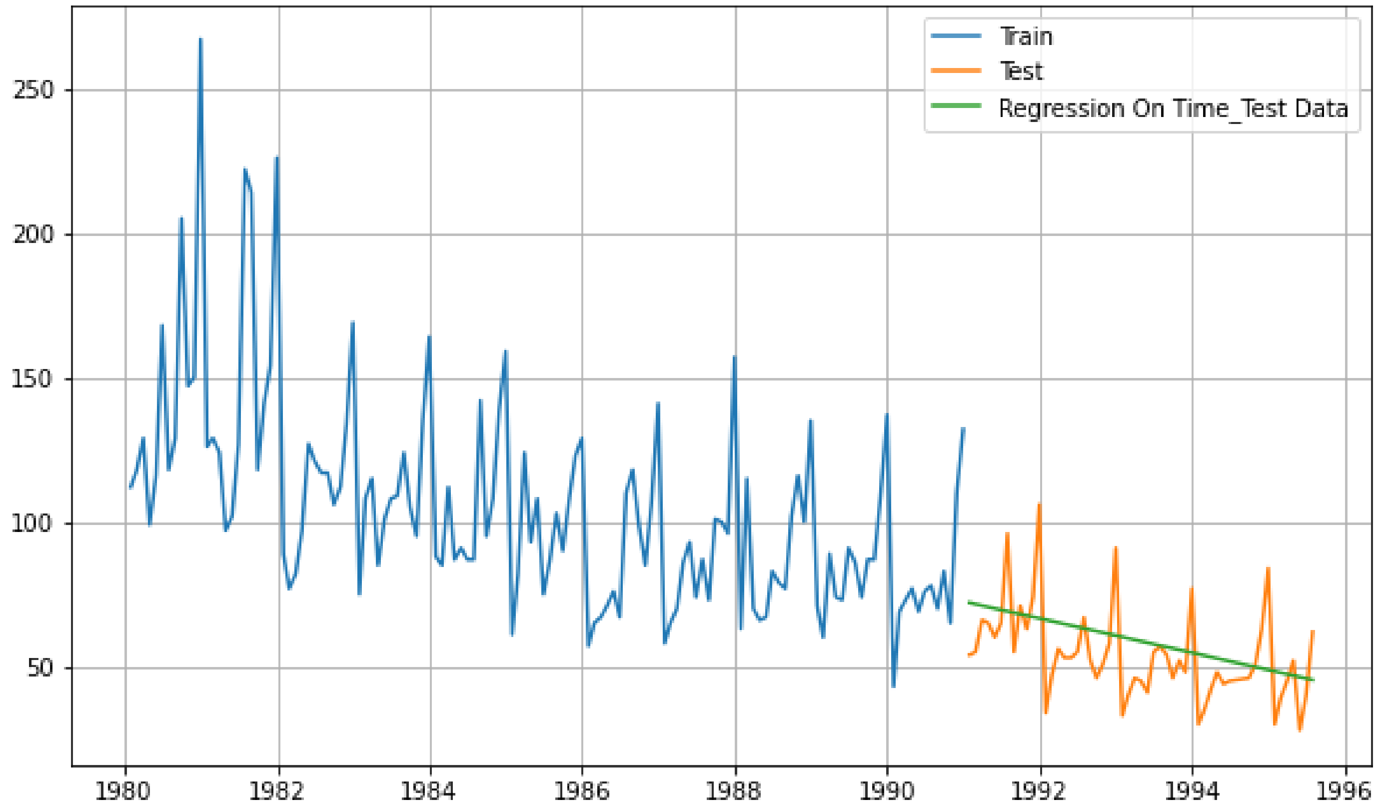
```
First few rows of Test Data
```

Date	Rose	time
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

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

Date	Rose	time
1995-03-31	45.0	183
1995-04-30	52.0	184
1995-05-31	28.0	185
1995-06-30	40.0	186
1995-07-31	62.0	187

Regression On Time Test Data



Model Evaluation

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

Model 2: Naive Approach

Train Head

	Rose
Date	
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

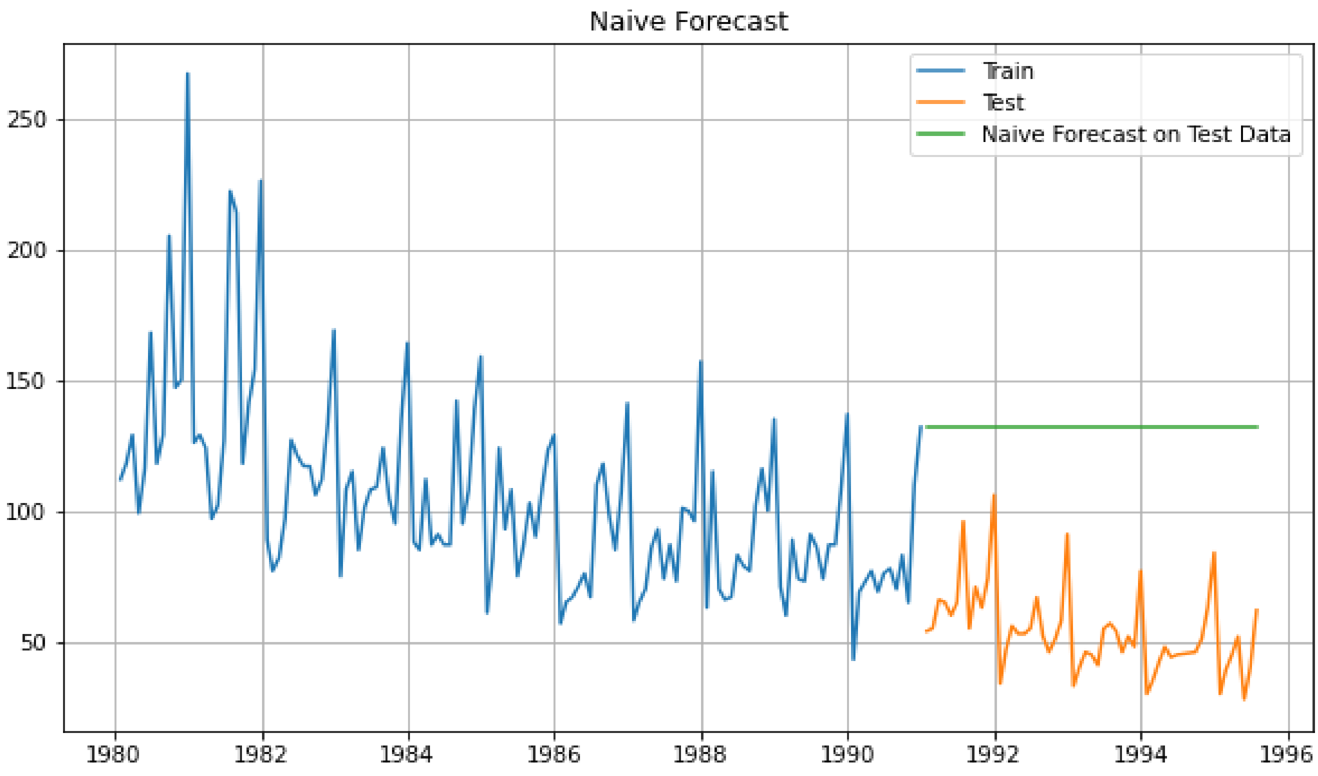
Test Head

	Rose
Date	
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

Naive Model Test

```
Date
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
```

Naive Forecast on Test Data



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

Model Evaluation

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

```
Table
```

	Test RMSE
RegressionOnTime	15.268955
NaiveModel	79.718773

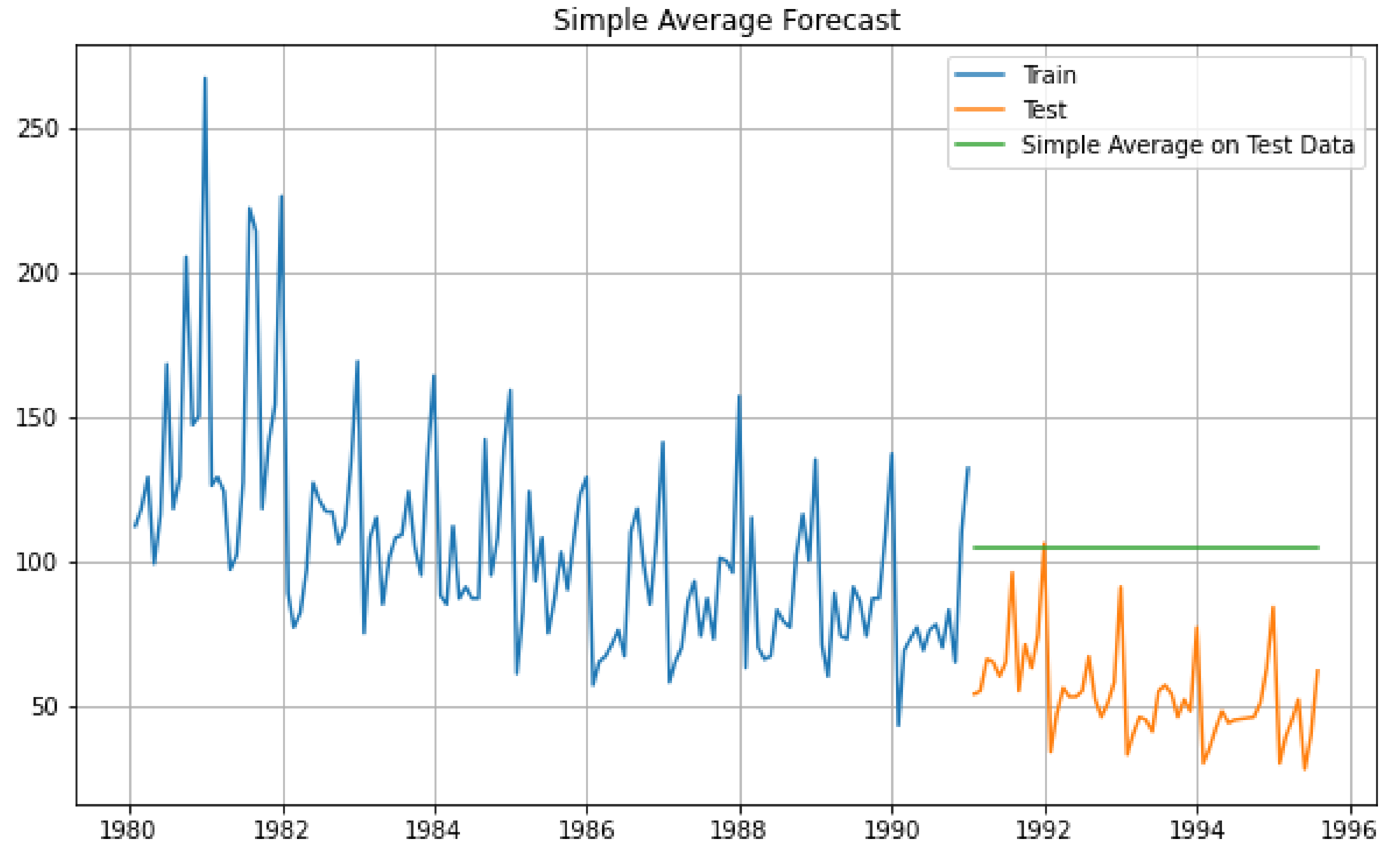
Method 3: Simple Average

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

Table

	Births	mean_forecast
Time_Stamp		
1959-09-13	42	40.803922
1959-09-14	34	40.803922
1959-09-15	40	40.803922
1959-09-16	56	40.803922
1959-09-17	44	40.803922

Simple Average on test data



Model Evaluation

For Simple Average forecast on the Test Data, RMSE is 53.461

Table

	Test RMSE
RegressionOnTime	15.268955
NaiveModel	79.718773
SimpleAverageModel	53.460570

Method 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.

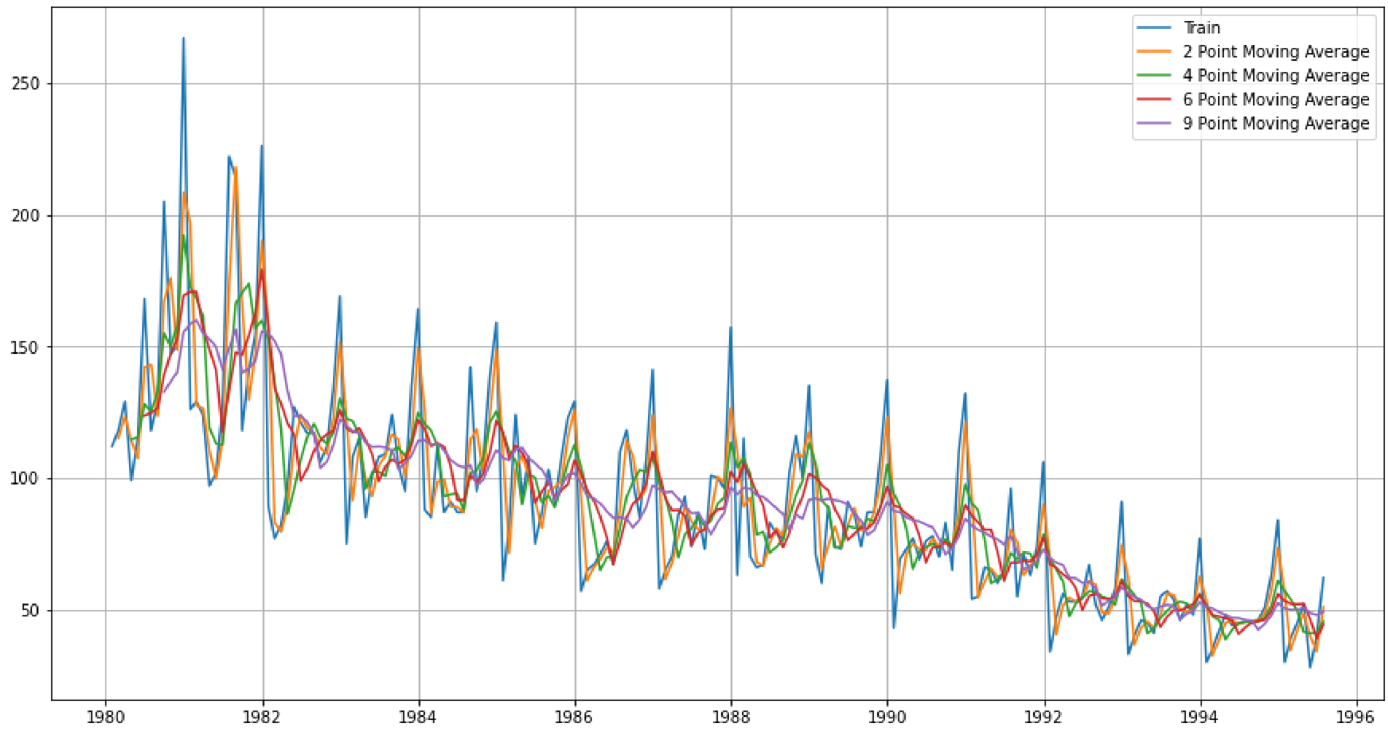
Table

	Rose
Date	
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

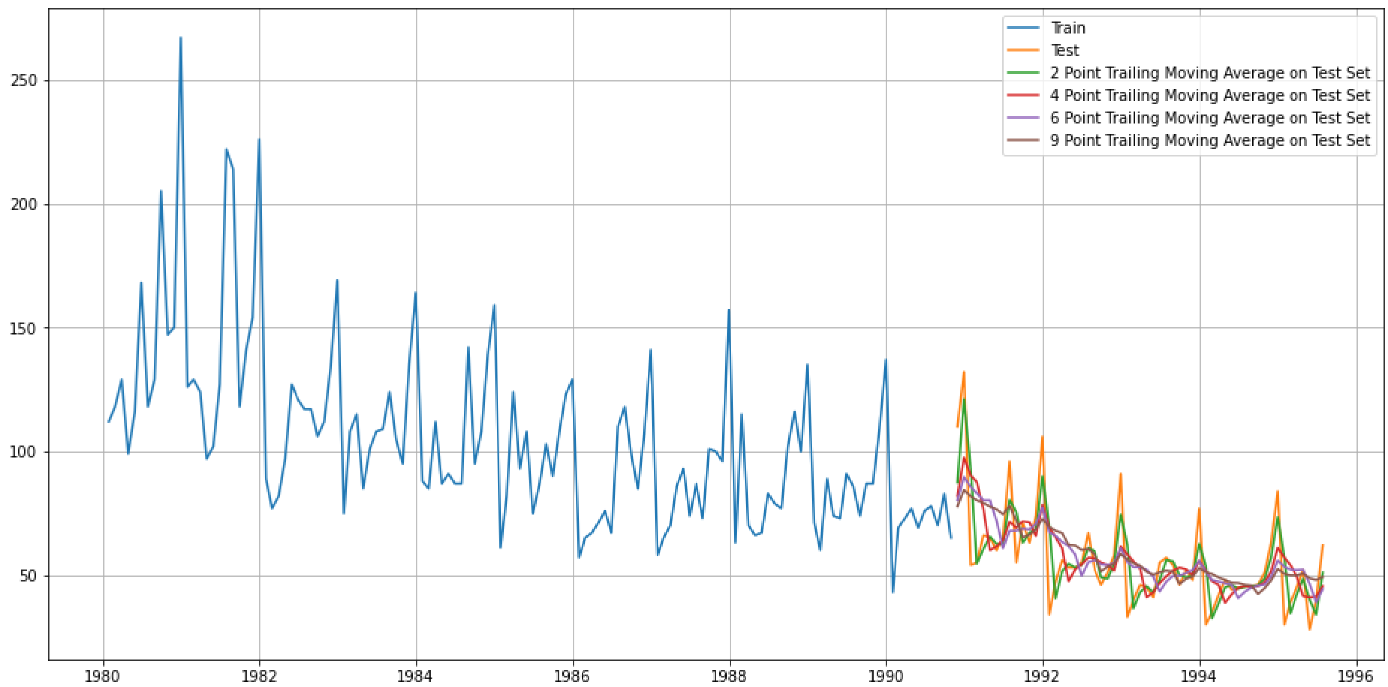
Trailing moving averages

	Rose	Trailing_2	Trailing_4	Trailing_6	Trailing_9
Date					
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

Moving Averages



Plotting on both the Training and Test data



Trailing Moving test head

	Rose	Trailing_2	Trailing_4	Trailing_6	Trailing_9
Date					
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

Trailing Moving Average test Shape

```
(57, 5)
```

Test Head

	Rose
Date	
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

Test Head

```
(55, 1)
```

Model Evaluation

Done only on the test data.

```
For 2 point Moving Average Model forecast on the Training Data,  RMSE is
11.529
For 4 point Moving Average Model forecast on the Training Data,  RMSE is
14.451
For 6 point Moving Average Model forecast on the Training Data,  RMSE is
14.566
For 9 point Moving Average Model forecast on the Training Data,  RMSE is
14.728
```

Table

	Test RMSE
RegressionOnTime	15.268955
NaiveModel	79.718773
SimpleAverageModel	53.460570
2pointTrailingMovingAverage	11.529278
4pointTrailingMovingAverage	14.451403
6pointTrailingMovingAverage	14.566327
9pointTrailingMovingAverage	14.727630

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

Exponential Smoothing methods

Exponential smoothing methods consist of flattening time series data.

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

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

These parameters have values between 0 and 1

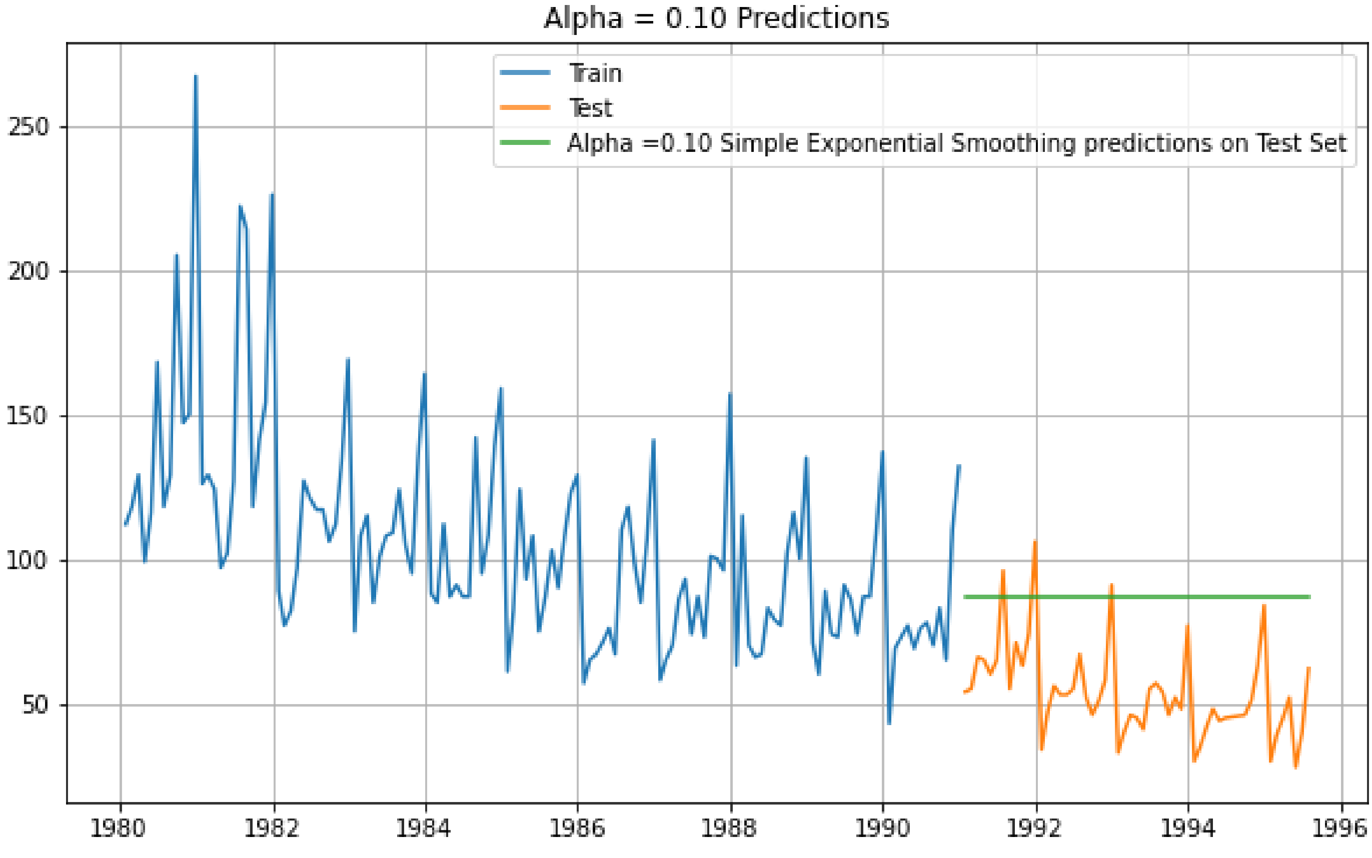
Parameters

```
{'smoothing_level': 0.09874983698117956,
 'smoothing_trend': nan,
 'smoothing_seasonal': nan,
 'damping_trend': nan,
 'initial_level': 134.38702481818487,
 'initial_trend': nan,
 'initial_seasons': array([], dtype=float64),
 'use_boxcox': False,
 'lamda': None,
 'remove_bias': False}
```

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

```
1991-01-31     87.104997
1991-02-28     87.104997
1991-03-31     87.104997
1991-04-30     87.104997
1991-05-31     87.104997
```

Alpha =0.10 Simple Exponential Smoothing predictions o n Test Set



calculate RMSE

```
SES RMSE: 36.79624054770398
SES RMSE (calculated using statsmodels): 36.796240547703974
```

Table

	Test RMSE
RegressionOnTime	15.268955
NaiveModel	79.718773
SimpleAverageModel	53.460570
2pointTrailingMovingAverage	11.529278
4pointTrailingMovingAverage	14.451403
6pointTrailingMovingAverage	14.566327
9pointTrailingMovingAverage	14.727630
Alpha=0.10,SimpleExponentialSmoothing	36.796241

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

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

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

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

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

Double Exponential Smoothing

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

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

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

Intercept or Level equation, Lt is given by: $Lt=\alpha Yt+(1-\alpha)Ft$ Trend equation is given by $Tt=\beta(Lt-Lt-1)+(1-\beta)Tt-1$ Here, α and β are the smoothing constants for level and trend, respectively,

$0 < \alpha < 1$ and $0 < \beta < 1$. The forecast at time t + 1 is given by

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

Parameters

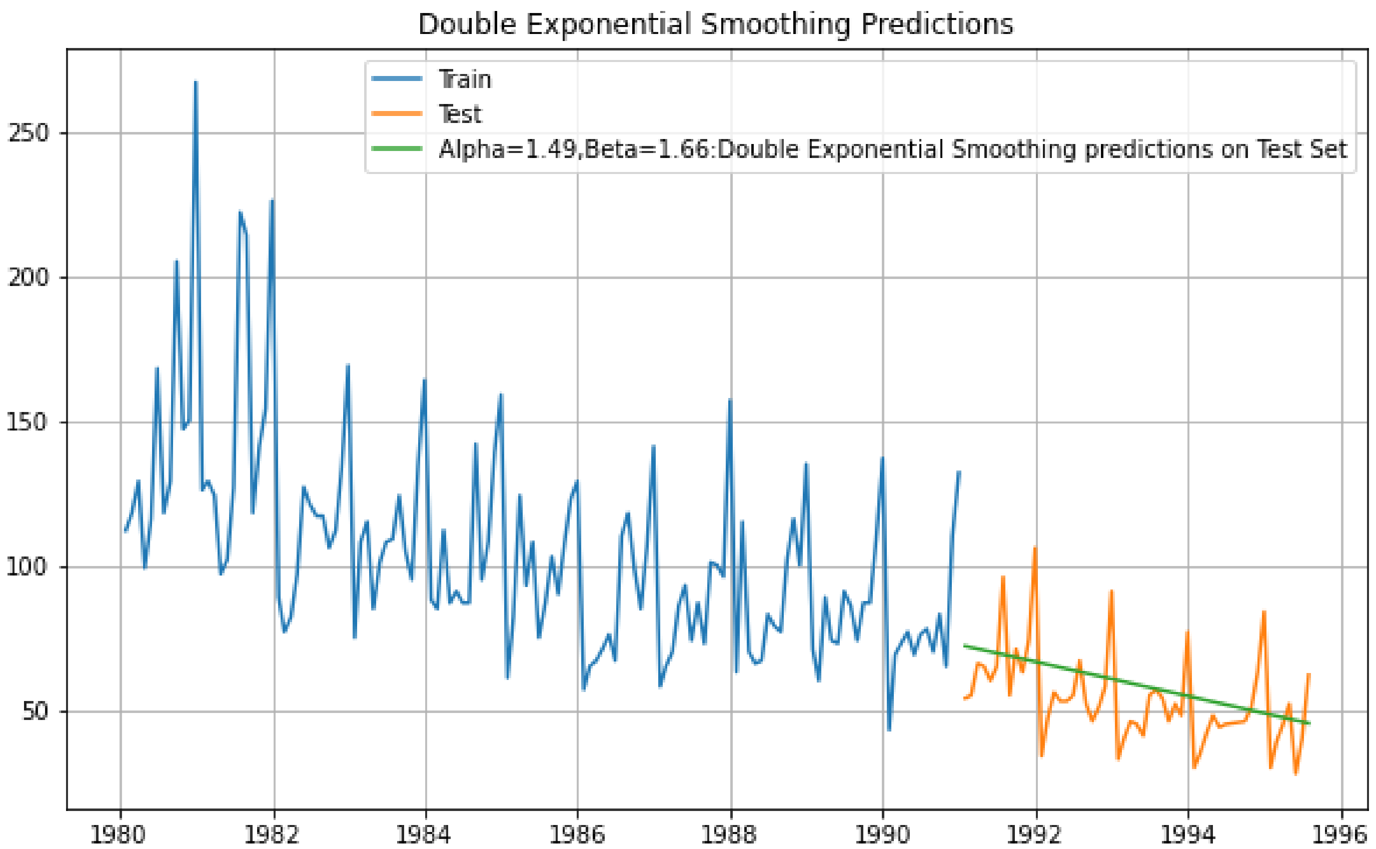
```
==Holt model Exponential Smoothing Estimated Parameters ==

{'smoothing_level': 1.4901161193847656e-08, 'smoothing_trend':
1.6610391146660035e-10, 'smoothing_seasonal': nan, 'damping_tr
end': nan, 'initial_level': 137.81553690867275, 'initial_trend
': -0.4943781897068274, 'initial_seasons': array([], dtype=flo
at64), 'use_boxcox': False, 'lamda': None, 'remove_bias': Fals
e}
```

Forecasting using this model for the duration of the test set

```
1991-01-31    72.063238
1991-02-28    71.568859
1991-03-31    71.074481
1991-04-30    70.580103
1991-05-31    70.085725
```

Plotting the Training data, Test data and the forecasted values



RMSE

DES RMSE: 15.268943764436564

Table

	Test RMSE
RegressionOnTime	15.268955
NaiveModel	79.718773
SimpleAverageModel	53.460570
2pointTrailingMovingAverage	11.529278
4pointTrailingMovingAverage	14.451403

	Test RMSE
6pointTrailingMovingAverage	14.566327
9pointTrailingMovingAverage	14.727630
Alpha=0.10,SimpleExponentialSmoothing	36.796241
Alpha=1.49,Beta=1.66:DoubleExponentialSmoothing	15.268944

Inference

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

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

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

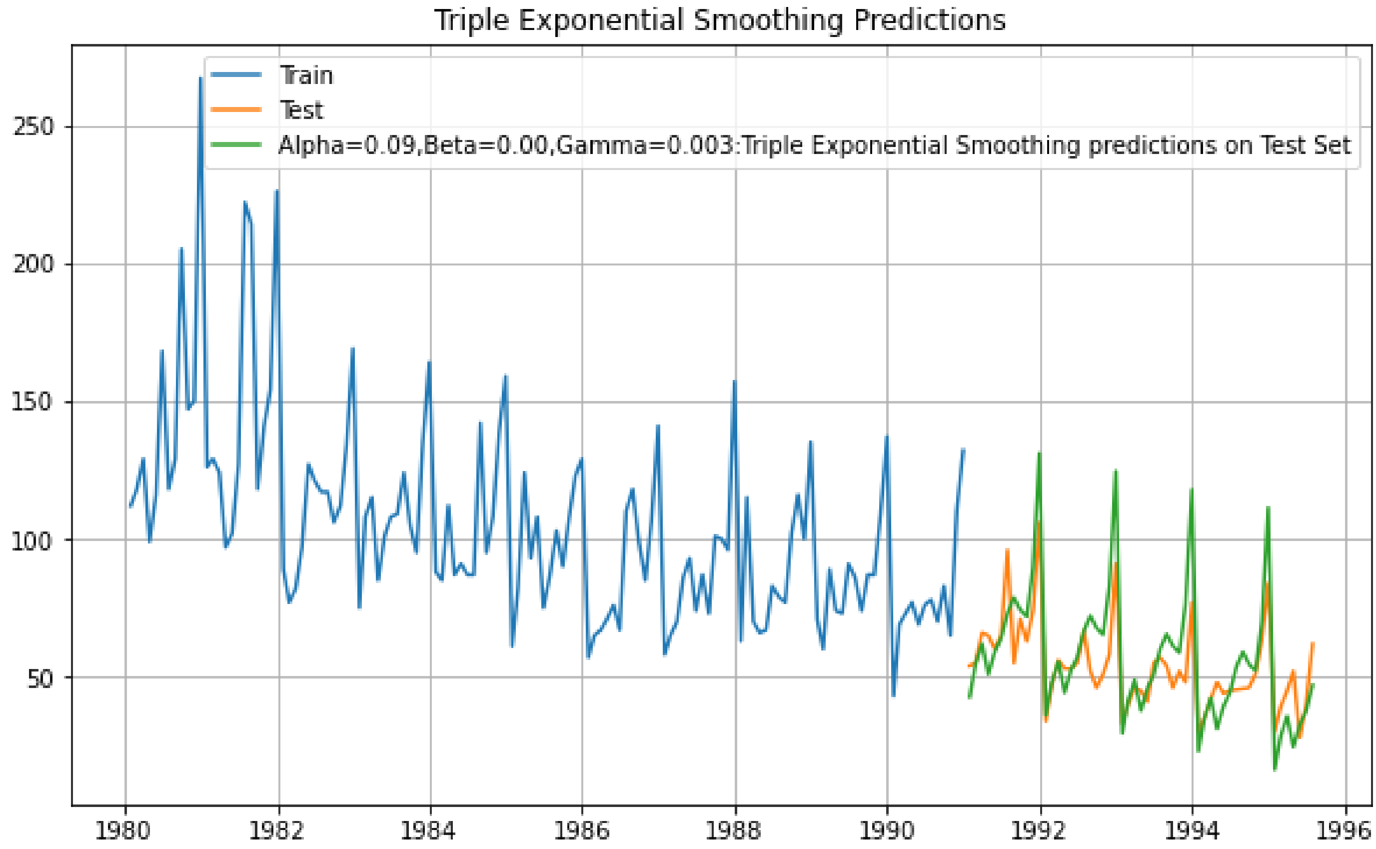
```
==Holt Winters model Exponential Smoothing Estimated Parameters ==

{'smoothing_level': 0.08954054664605082, 'smoothing_trend': 0.0002400108
693915795, 'smoothing_seasonal': 0.003466872515750747, 'damping_trend':
nan, 'initial_level': 146.5570157826235, 'initial_trend': -0.54719698350
9005, 'initial_seasons': array([-31.17478463, -18.74839869, -10.76961776
, -21.36741017,
       -12.63775539,  -7.27430333,   2.61279801,   8.69603625,
         4.79381122,   2.96110122,  21.05738849,  63.18279918]), 'use_bo
xcox': False, 'lamda': None, 'remove_bias': False}
```

Forecasting using this model for the duration of the test set

```
1991-01-31    42.684928
1991-02-28    54.564005
1991-03-31    61.995209
1991-04-30    50.852018
1991-05-31    59.034271
```

Triple Exponential Smoothing Predictions



RMSE

```
TES RMSE: 14.249660750768982
```

Table

	Test RMSE
RegressionOnTime	15.268955
NaiveModel	79.718773
SimpleAverageModel	53.460570
2pointTrailingMovingAverage	11.529278
4pointTrailingMovingAverage	14.451403
6pointTrailingMovingAverage	14.566327
9pointTrailingMovingAverage	14.727630
Alpha=0.10,SimpleExponentialSmoothing	36.796241
Alpha=1.49,Beta=1.66:DoubleExponentialSmoothing	15.268944

	Test RMSE
Alpha=0.09,Beta=0.00,Gamma=0.003:Triple Exponential Smoothing	14.249661

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

Parameters

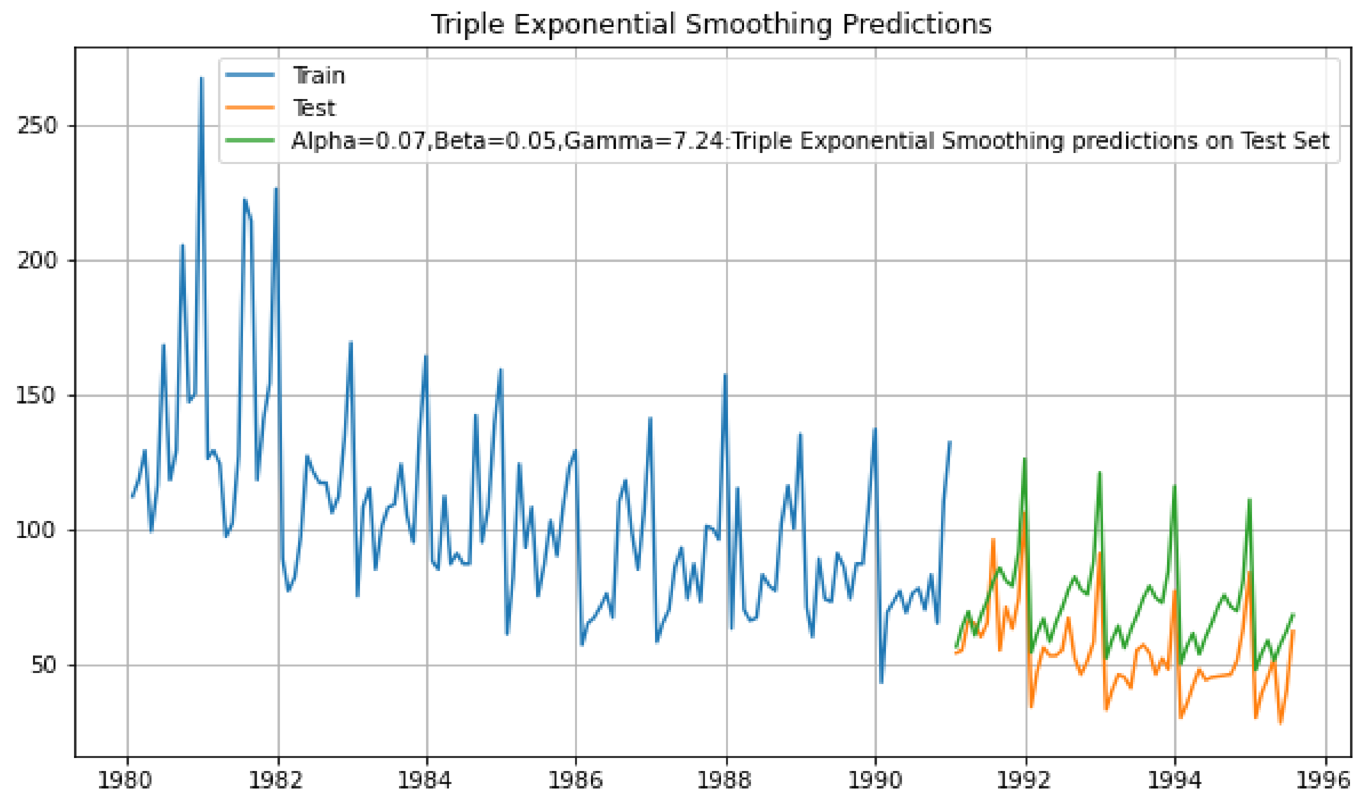
```
==Holt Winters model Exponential Smoothing Estimated Parameters ==

{'smoothing_level': 0.0715106306609405, 'smoothing_trend': 0.04529179757
535142, 'smoothing_seasonal': 7.244325029450242e-05, 'damping_trend': na
n, 'initial_level': 130.40839142502193, 'initial_trend': -0.779857431793
86, 'initial_seasons': array([0.86218996, 0.977675  , 1.0687727 , 0.9340
3881, 1.050625  ,
       1.14410977, 1.25836944, 1.33937772, 1.26778766, 1.24131254,
       1.44724625, 1.99553681]), 'use_boxcox': False, 'lamda': None, 're
move_bias': False}
```

Forecasting using this model for the duration of the test set

```
1991-01-31    56.321655
1991-02-28    63.664690
1991-03-31    69.374024
1991-04-30    60.435528
1991-05-31    67.758341
```

Triple Exponential Smoothing predictions on Test Set



Rmse

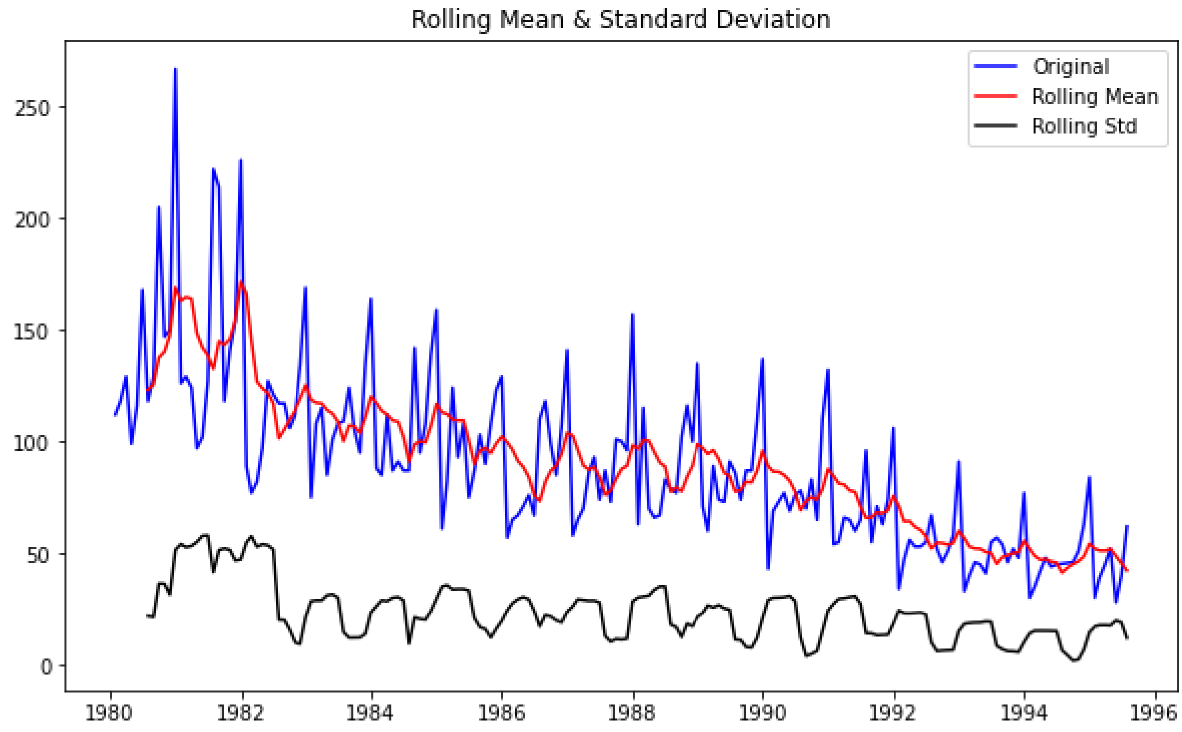
```
TES_am RMSE: 20.156762582665337
```

Table

	Test RMSE
RegressionOnTime	15.268955
NaiveModel	79.718773
SimpleAverageModel	53.460570
2pointTrailingMovingAverage	11.529278
4pointTrailingMovingAverage	14.451403
6pointTrailingMovingAverage	14.566327
9pointTrailingMovingAverage	14.727630
Alpha=0.10,SimpleExponentialSmoothing	36.796241
Alpha=1.49,Beta=1.66:DoubleExponentialSmoothing	15.268944
Alpha=0.09,Beta=0.00,Gamma=0.003:Triple Exponential Smoothing	14.249661
Alpha=0.07,Beta=0.05,Gamma=7.24:Triple Exponential Smoothing 2	20.156763

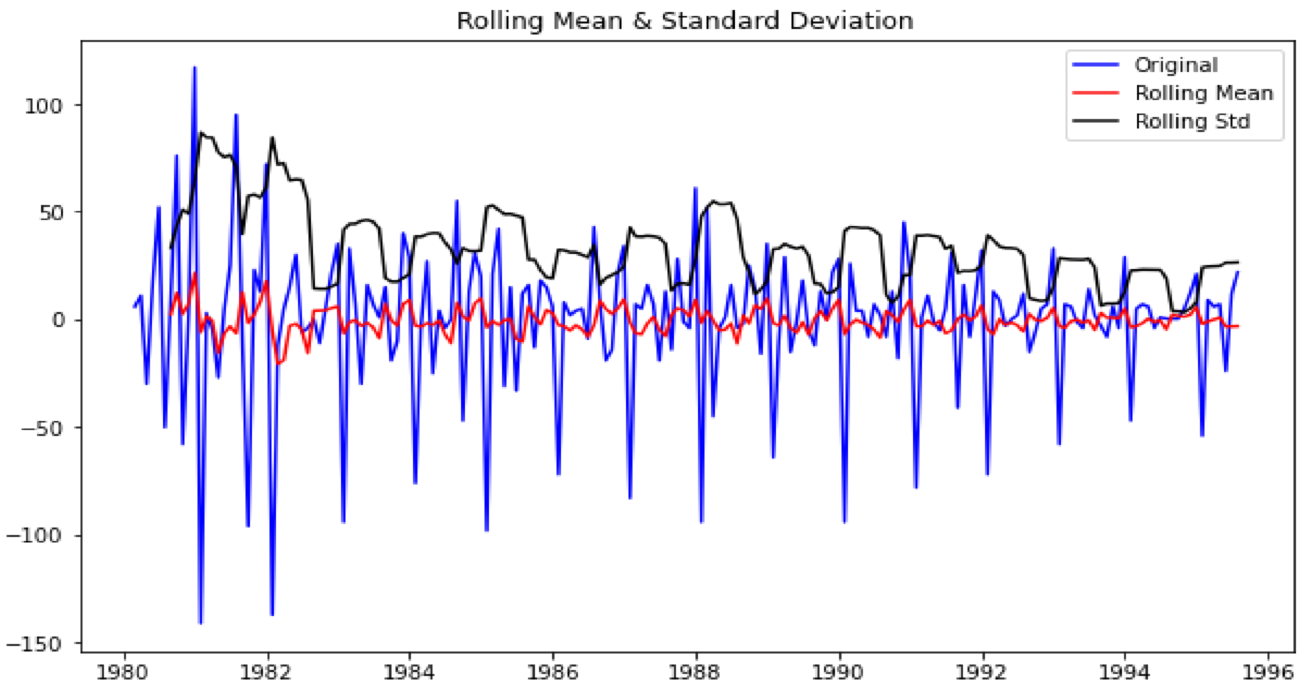
Q.5 Check for the stationarity of the data on which the model is being built on using appropriate statistical tests and also mention the hypothesis for the statistical test. If the data is found to be non-stationary, take appropriate steps to make it stationary. Check the new data for stationarity and comment.

Note: Stationarity should be checked at alpha = 0.05.




```
Results of Dickey-Fuller Test:
Test Statistic                  -1.876699
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
```

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

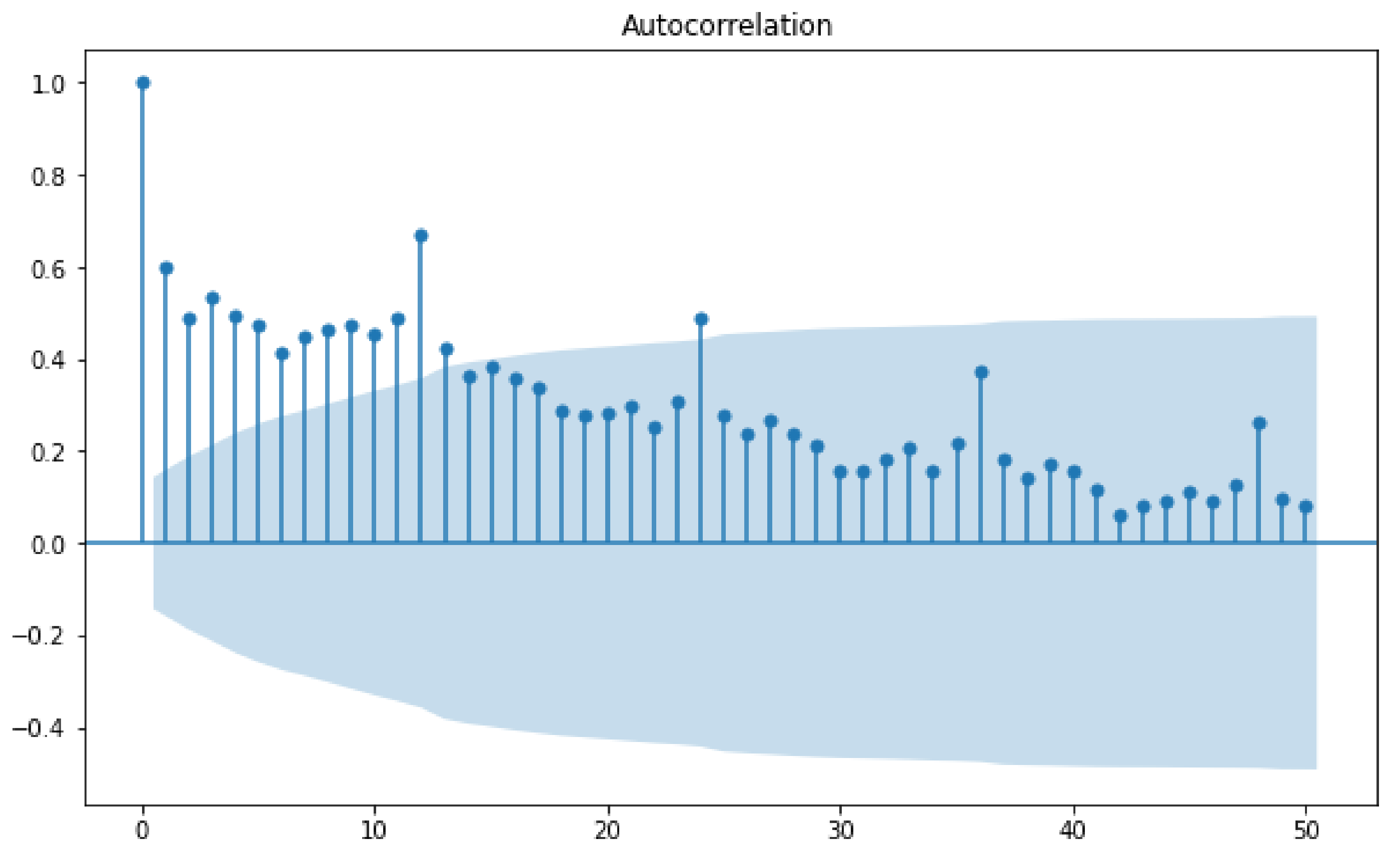


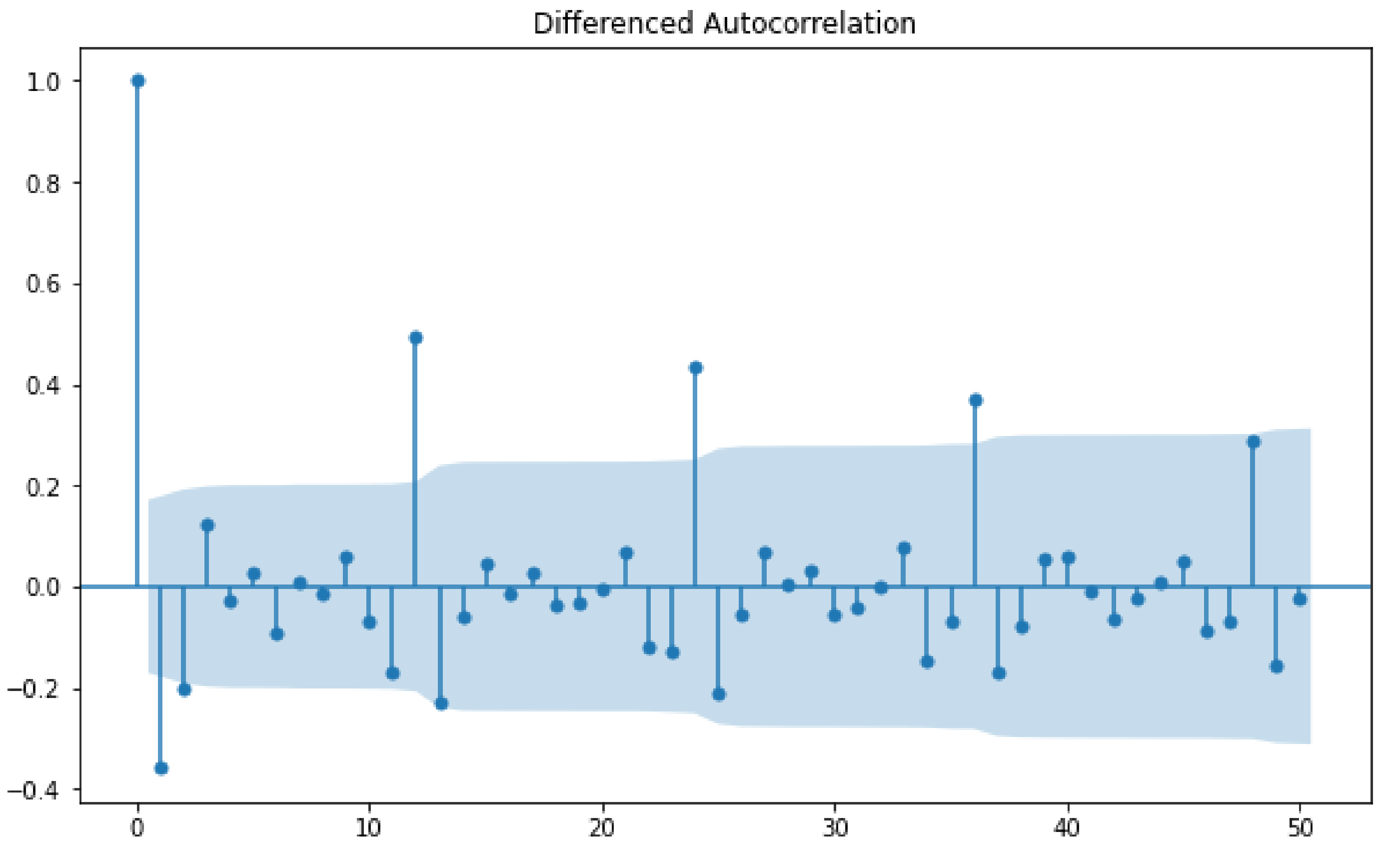
```
Results of Dickey-Fuller Test:
Test Statistic                  -8.044392e+00
p-value                          1.810895e-12
#Lags Used                       1.200000e+01
Number of Observations Used      1.730000e+02
Critical Value (1%)             -3.468726e+00
Critical Value (5%)             -2.878396e+00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

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

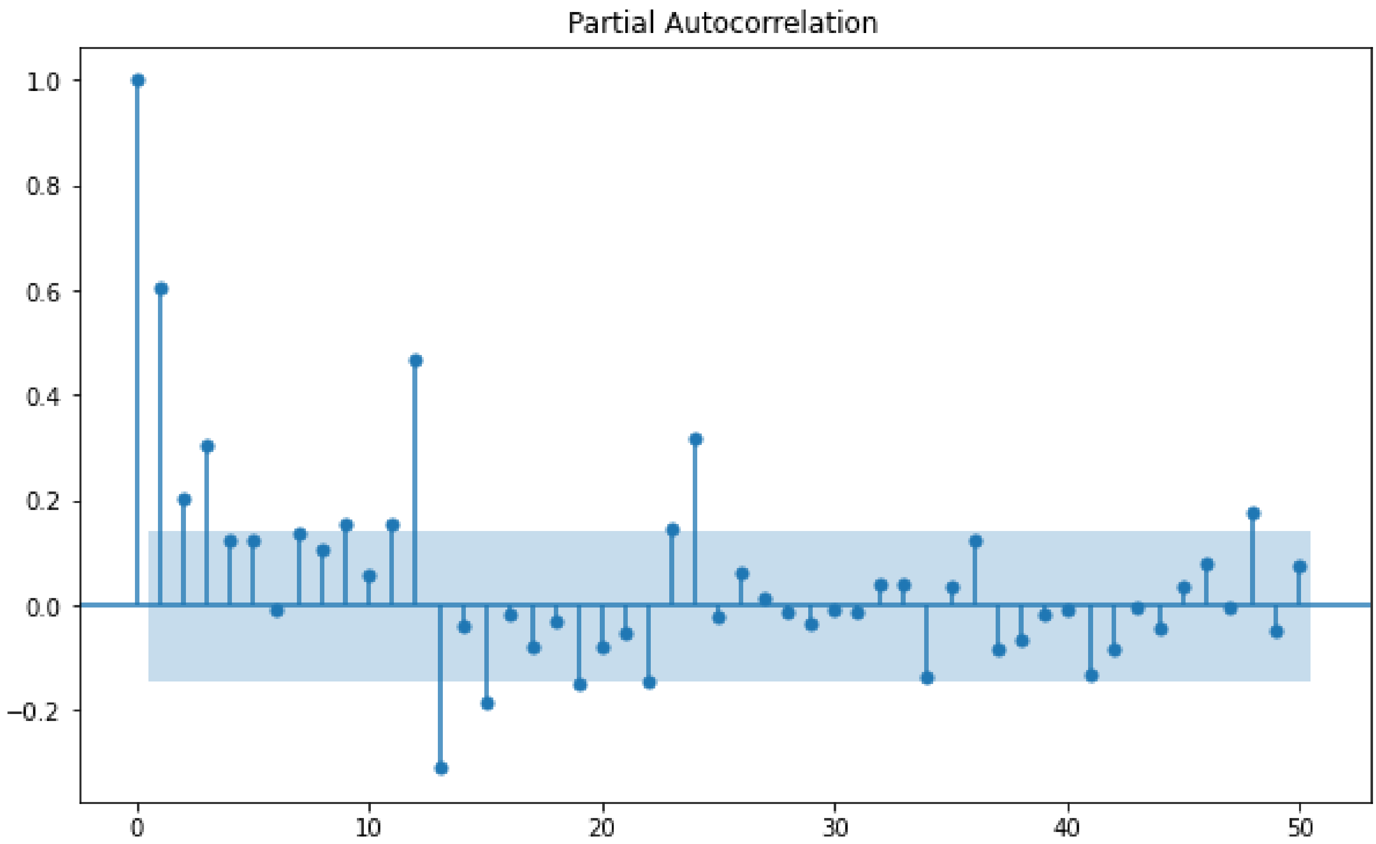
Plot the Autocorrelation and the Partial Autocorrelation function plots on the whole data.

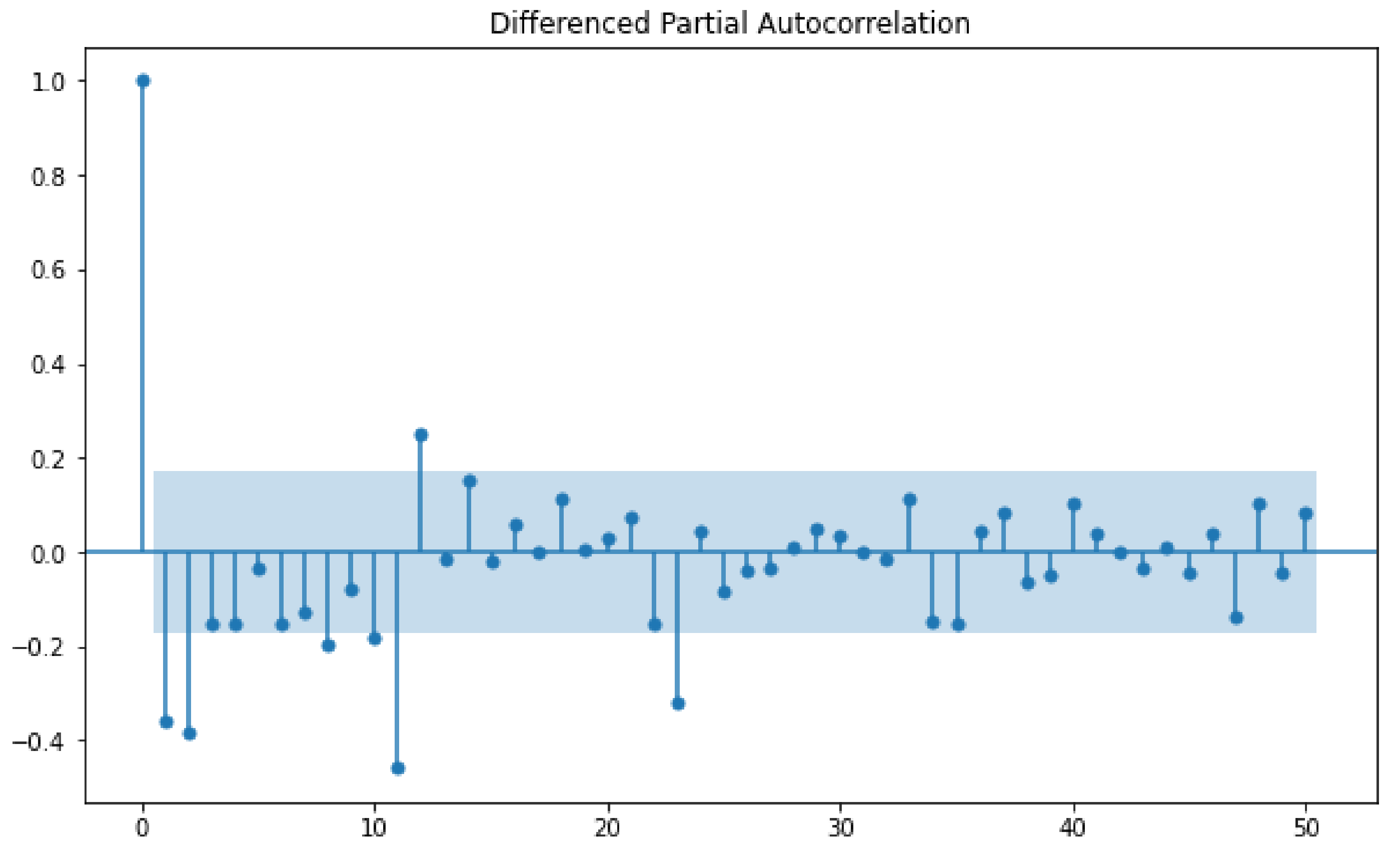
Autocorrelation



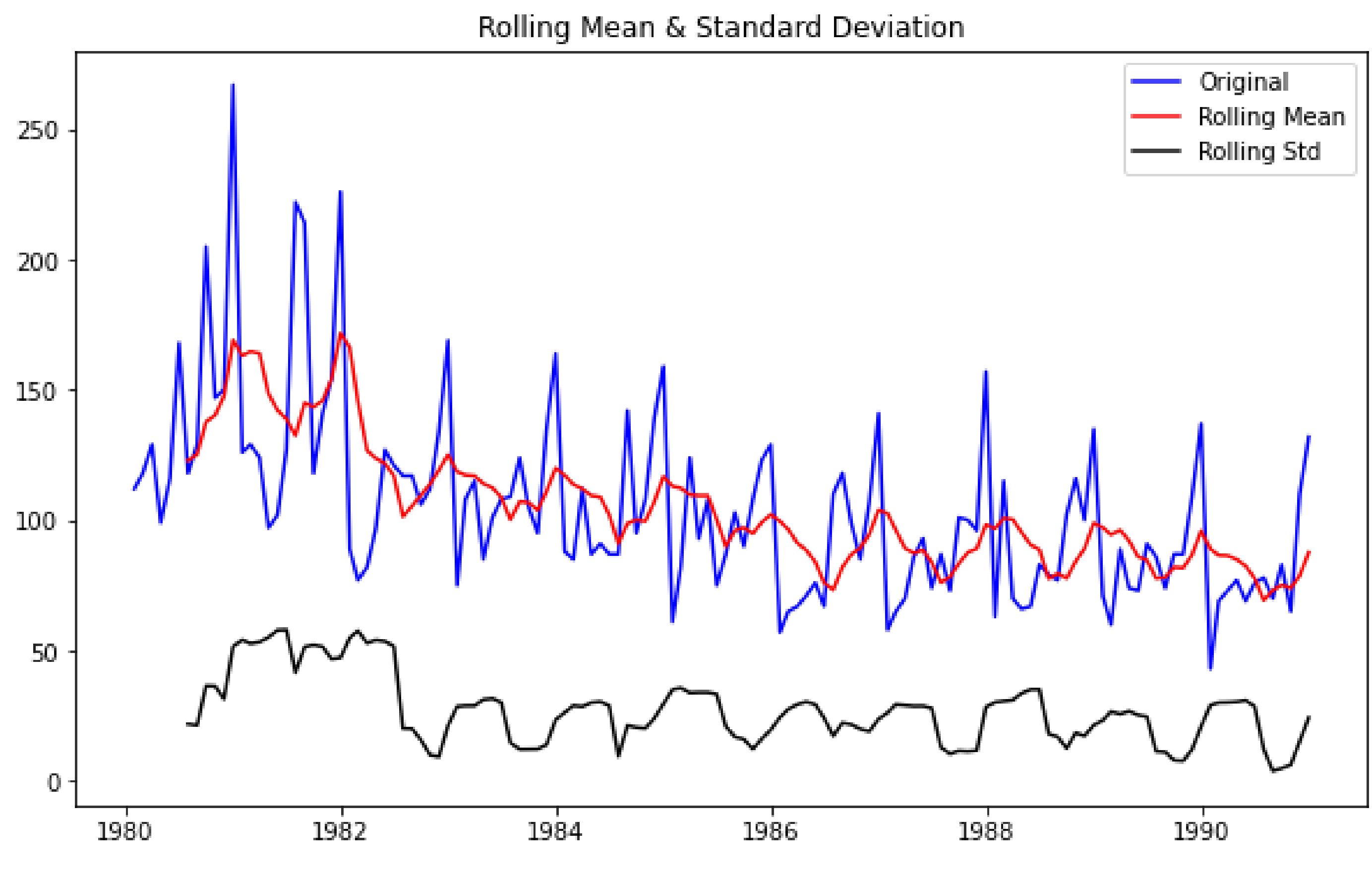


Partial Autocorrelation function



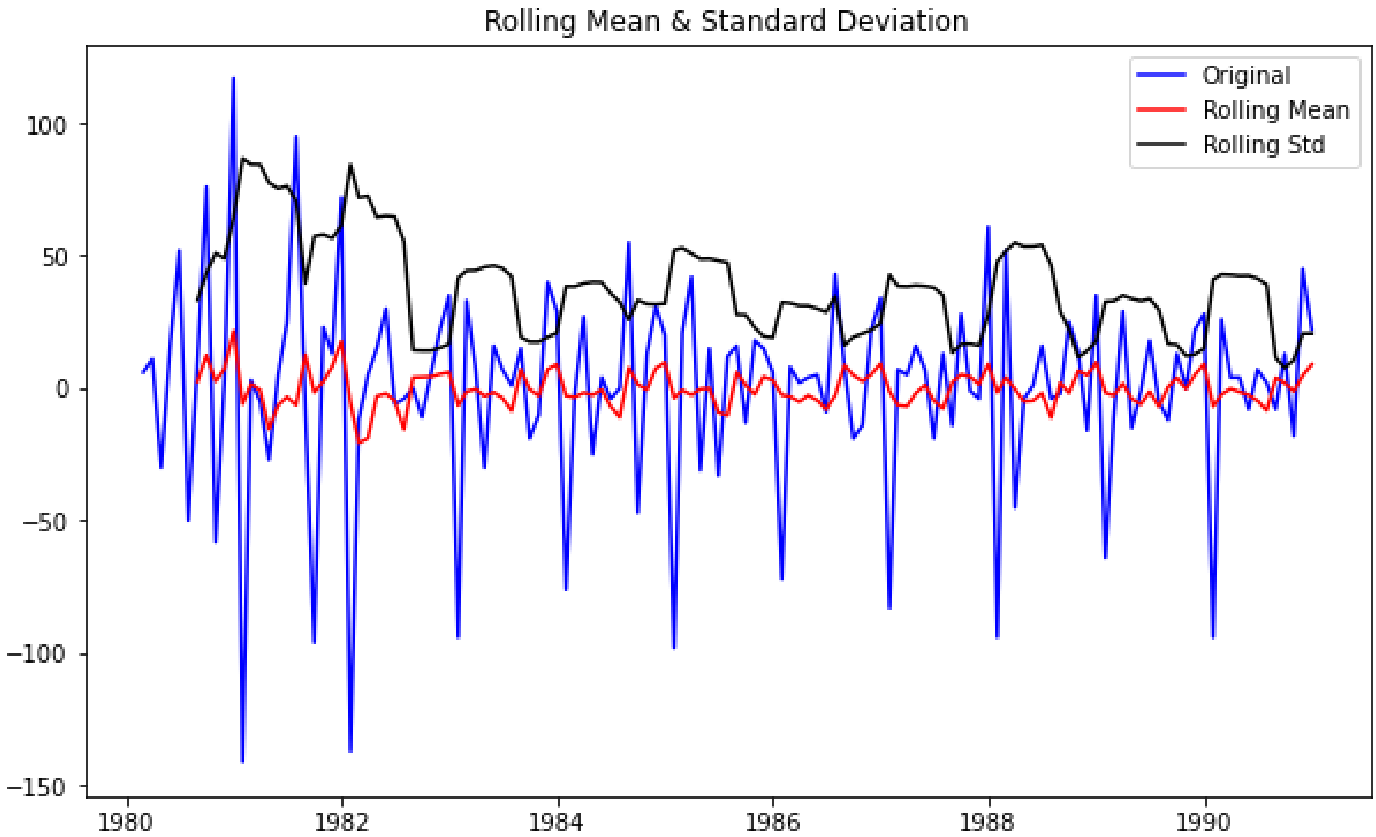


Check for stationarity of the Training Data Time Series.



```
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
```

Train at difference of 1




```
Results of Dickey-Fuller Test:
Test Statistic                 -6.592372e+00
p-value                         7.061944e-09
#Lags Used                      1.200000e+01
Number of Observations Used     1.180000e+02
Critical Value (1%)            -3.487022e+00
Critical Value (5%)            -2.886363e+00
Critical Value (10%)           -2.580009e+00
```

The Data have become stationary at the difference of 1.

Q.6 Build an automated version of the ARIMA/SARIMA model in which the parameters are selected using the

lowest Akaike Information Criteria (AIC) on the training data and evaluate this model on the test data using RMSE.

Some parameter combinations for the Model...

```
Model: (0, 1, 1)
Model: (0, 1, 2)
Model: (0, 1, 3)
Model: (0, 1, 4)
Model: (1, 1, 0)
Model: (1, 1, 1)
Model: (1, 1, 2)
Model: (1, 1, 3)
Model: (1, 1, 4)
Model: (2, 1, 0)
Model: (2, 1, 1)
Model: (2, 1, 2)
Model: (2, 1, 3)
Model: (2, 1, 4)
Model: (3, 1, 0)
Model: (3, 1, 1)
Model: (3, 1, 2)
Model: (3, 1, 3)
Model: (3, 1, 4)
Model: (4, 1, 0)
Model: (4, 1, 1)
Model: (4, 1, 2)
Model: (4, 1, 3)
Model: (4, 1, 4)
```

AIC

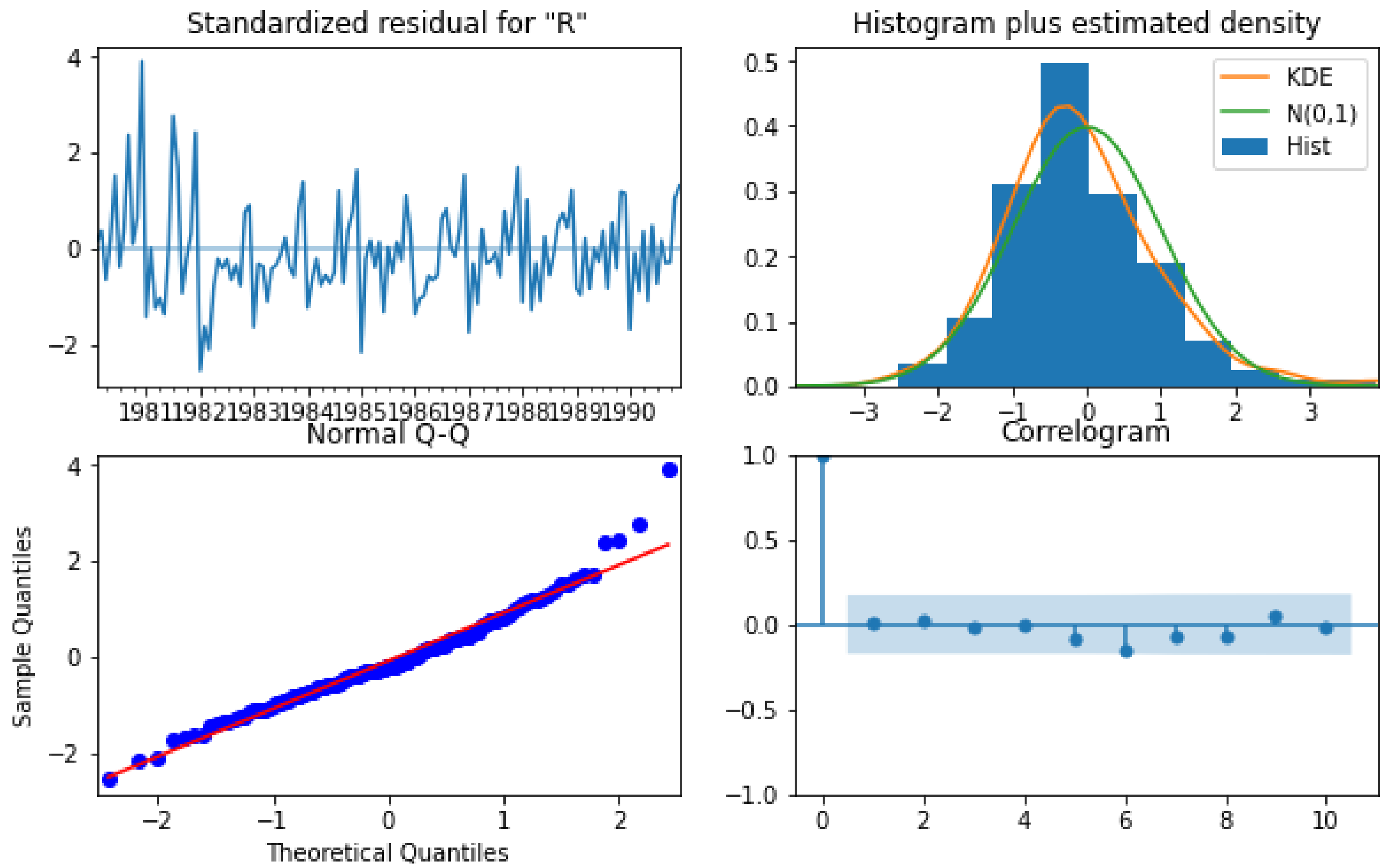
```
ARIMA(0, 1, 0) - AIC:1333.1546729124348
ARIMA(0, 1, 1) - AIC:1282.3098319748312
ARIMA(0, 1, 2) - AIC:1279.6715288535818
ARIMA(0, 1, 3) - AIC:1280.5453761734655
ARIMA(0, 1, 4) - AIC:1281.676698214394
ARIMA(1, 1, 0) - AIC:1317.3503105381492
ARIMA(1, 1, 1) - AIC:1280.5742295380032
ARIMA(1, 1, 2) - AIC:1279.870723423191
ARIMA(1, 1, 3) - AIC:1281.8707223310003
ARIMA(1, 1, 4) - AIC:1279.6052625434186
ARIMA(2, 1, 0) - AIC:1298.6110341605004
ARIMA(2, 1, 1) - AIC:1281.5078621868474
ARIMA(2, 1, 2) - AIC:1281.8707222264284
ARIMA(2, 1, 3) - AIC:1274.6949119626274
ARIMA(2, 1, 4) - AIC:1278.772249045519
ARIMA(3, 1, 0) - AIC:1297.48109172717
ARIMA(3, 1, 1) - AIC:1282.4192776271977
ARIMA(3, 1, 2) - AIC:1283.720740597716
ARIMA(3, 1, 3) - AIC:1278.6588655941036
ARIMA(3, 1, 4) - AIC:1287.7190768443138
ARIMA(4, 1, 0) - AIC:1296.32665690046
ARIMA(4, 1, 1) - AIC:1283.7931715123075
ARIMA(4, 1, 2) - AIC:1285.7182485626197
ARIMA(4, 1, 3) - AIC:1278.4514105832604
ARIMA(4, 1, 4) - AIC:1282.3776177189604
```

Table (First 5)

	param	AIC
13	(2, 1, 3)	1274.694912
23	(4, 1, 3)	1278.451411
18	(3, 1, 3)	1278.658866
14	(2, 1, 4)	1278.772249
9	(1, 1, 4)	1279.605263

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                   Rose   No. Observations:                  132
Model:                 ARIMA(2, 1, 3)   Log Likelihood                -631.347
Date:                Sat, 24 Sep 2022   AIC                           1274.695
Time:                        18:11:31   BIC                           1291.946
Sample:                    01-31-1980   HQIC                          1281.705
                         - 12-31-1990
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -1.6781      0.084    -20.035      0.000      -1.842      -1.514
ar.L2         -0.7289      0.084     -8.703      0.000      -0.893      -0.565
ma.L1          1.0450      0.685      1.527      0.127      -0.297       2.387
ma.L2         -0.7716      0.137     -5.636      0.000      -1.040      -0.503
ma.L3         -0.9046      0.622     -1.455      0.146      -2.123       0.314
sigma2       858.3595    576.845      1.488      0.137    -272.237    1988.956
===================================================================================
Ljung-Box (L1) (Q):                   0.02   Jarque-Bera (JB):                24.45
Prob(Q):                              0.88   Prob(JB):                         0.00
Heteroskedasticity (H):               0.40   Skew:                             0.71
Prob(H) (two-sided):                  0.00   Kurtosis:                         4.57
===================================================================================
```

Result For ARIMA



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

RMSE

```
RMSE: 36.81742323289034
MAPE: 75.84837827903644
```

Table

	Test RMSE	MAPE
RegressionOnTime	15.268955	NaN
NaiveModel	79.718773	NaN
SimpleAverageModel	53.460570	NaN
2pointTrailingMovingAverage	11.529278	NaN

	Test RMSE	MAPE
4pointTrailingMovingAverage	14.451403	NaN
6pointTrailingMovingAverage	14.566327	NaN
9pointTrailingMovingAverage	14.727630	NaN
Alpha=0.10,SimpleExponentialSmoothing	36.796241	NaN
Alpha=1.49,Beta=1.66:DoubleExponentialSmoothing	15.268944	NaN
Alpha=0.09,Beta=0.00,Gamma=0.003:Triple Exponential Smoothing	14.249661	NaN
Alpha=0.07,Beta=0.05,Gamma=7.24:Triple Exponential Smoothing 2	20.156763	NaN
Arima 2,1,3	36.817423	75.848378

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

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

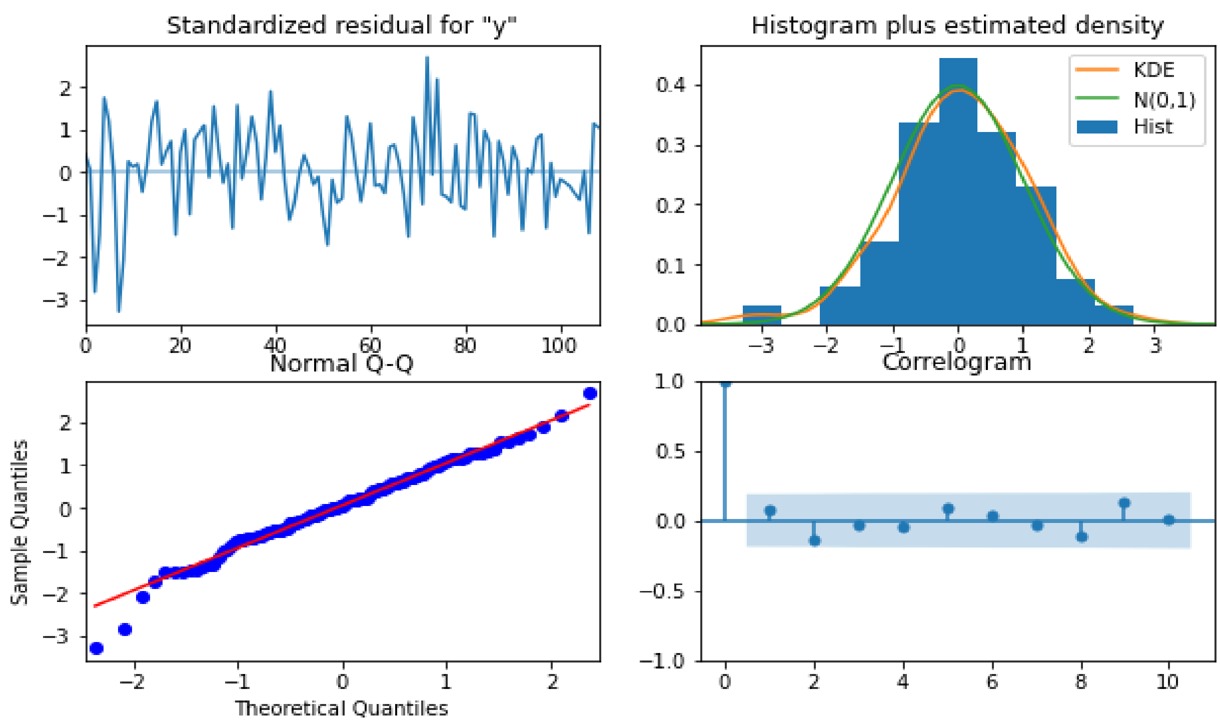
```
Examples of some parameter combinations for Model...
Model: (0, 1, 1)(0, 0, 1, 6)
Model: (0, 1, 2)(0, 0, 2, 6)
Model: (0, 1, 3)(0, 0, 3, 6)
Model: (1, 1, 0)(1, 0, 0, 6)
Model: (1, 1, 1)(1, 0, 1, 6)
Model: (1, 1, 2)(1, 0, 2, 6)
Model: (1, 1, 3)(1, 0, 3, 6)
Model: (2, 1, 0)(2, 0, 0, 6)
Model: (2, 1, 1)(2, 0, 1, 6)
Model: (2, 1, 2)(2, 0, 2, 6)
Model: (2, 1, 3)(2, 0, 3, 6)
Model: (3, 1, 0)(3, 0, 0, 6)
Model: (3, 1, 1)(3, 0, 1, 6)
Model: (3, 1, 2)(3, 0, 2, 6)
Model: (3, 1, 3)(3, 0, 3, 6)

Optimum AIC
```

	param	seasonal	AIC
187	(2, 1, 3)	(2, 0, 3, 6)	951.744297
59	(0, 1, 3)	(2, 0, 3, 6)	952.073632
251	(3, 1, 3)	(2, 0, 3, 6)	952.582106

	param	seasonal	AIC
191	(2, 1, 3)	(3, 0, 3, 6)	953.205635
123	(1, 1, 3)	(2, 0, 3, 6)	953.684951

```
                                     SARIMAX Results
==========================================================================================
Dep. Variable:                                   y   No. Observations:                 132
Model:             SARIMAX(2, 1, 3)x(2, 0, 3, 6)   Log Likelihood                -464.872
Date:                             Sun, 25 Sep 2022   AIC                            951.744
Time:                                     12:32:13   BIC                            981.349
Sample:                                          0   HQIC                           963.750
                                             - 132
Covariance Type:                               opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.5027      0.083     -6.082      0.000      -0.665      -0.341
ar.L2         -0.6628      0.084     -7.919      0.000      -0.827      -0.499
ma.L1         -0.3714    229.835     -0.002      0.999    -450.839     450.096
ma.L2          0.2033    144.466      0.001      0.999    -282.945     283.351
ma.L3         -0.8320    191.169     -0.004      0.997    -375.516     373.852
ar.S.L6       -0.0838      0.049     -1.720      0.085      -0.179       0.012
ar.S.L12       0.8099      0.052     15.465      0.000       0.707       0.913
ma.S.L6        0.1701      0.248      0.687      0.492      -0.315       0.656
ma.S.L12      -0.5645      0.199     -2.836      0.005      -0.955      -0.174
ma.S.L18       0.1710      0.143      1.198      0.231      -0.109       0.451
sigma2       260.7941   5.99e+04      0.004      0.997   -1.17e+05    1.18e+05
===================================================================================
Ljung-Box (L1) (Q):                   0.72   Jarque-Bera (JB):                 4.77
Prob(Q):                              0.40   Prob(JB):                         0.09
Heteroskedasticity (H):               0.54   Skew:                            -0.36
Prob(H) (two-sided):                  0.06   Kurtosis:                         3.73
===================================================================================
```



Summary frame

y	mean	mean_se	mean_ci_lower	mean_ci_upper
0	66.899868	16.350418	34.853637	98.946099
1	65.988383	16.481660	33.684922	98.291843
2	74.437591	16.587582	41.926527	106.948655
3	76.041031	16.710175	43.289690	108.792373
4	78.413894	16.710789	45.661351	111.166438

RMSE & MAPE

```
RMSE: 27.12407595995941
MAPE: 55.238878434557925
```

Table

	Test RMSE	MAPE
RegressionOnTime	15.268955	NaN
NaiveModel	79.718773	NaN
SimpleAverageModel	53.460570	NaN
2pointTrailingMovingAverage	11.529278	NaN
4pointTrailingMovingAverage	14.451403	NaN
6pointTrailingMovingAverage	14.566327	NaN
9pointTrailingMovingAverage	14.727630	NaN
Alpha=0.10,SimpleExponentialSmoothing	36.796241	NaN
Alpha=1.49,Beta=1.66:DoubleExponentialSmoothing	15.268944	NaN
Alpha=0.09,Beta=0.00,Gamma=0.003:Triple Exponential Smoothing	14.249661	NaN
Alpha=0.07,Beta=0.05,Gamma=7.24:Triple Exponential Smoothing 2	20.156763	NaN
Arima (2,1,3)	36.817423	75.848378
Sarima (1,1,2)(2,0,2,6)	27.124076	55.238878

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

```
Examples of some parameter combinations for Model...
```

```
Model: (0, 1, 1)(0, 0, 1, 12)
Model: (0, 1, 2)(0, 0, 2, 12)
Model: (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)
```

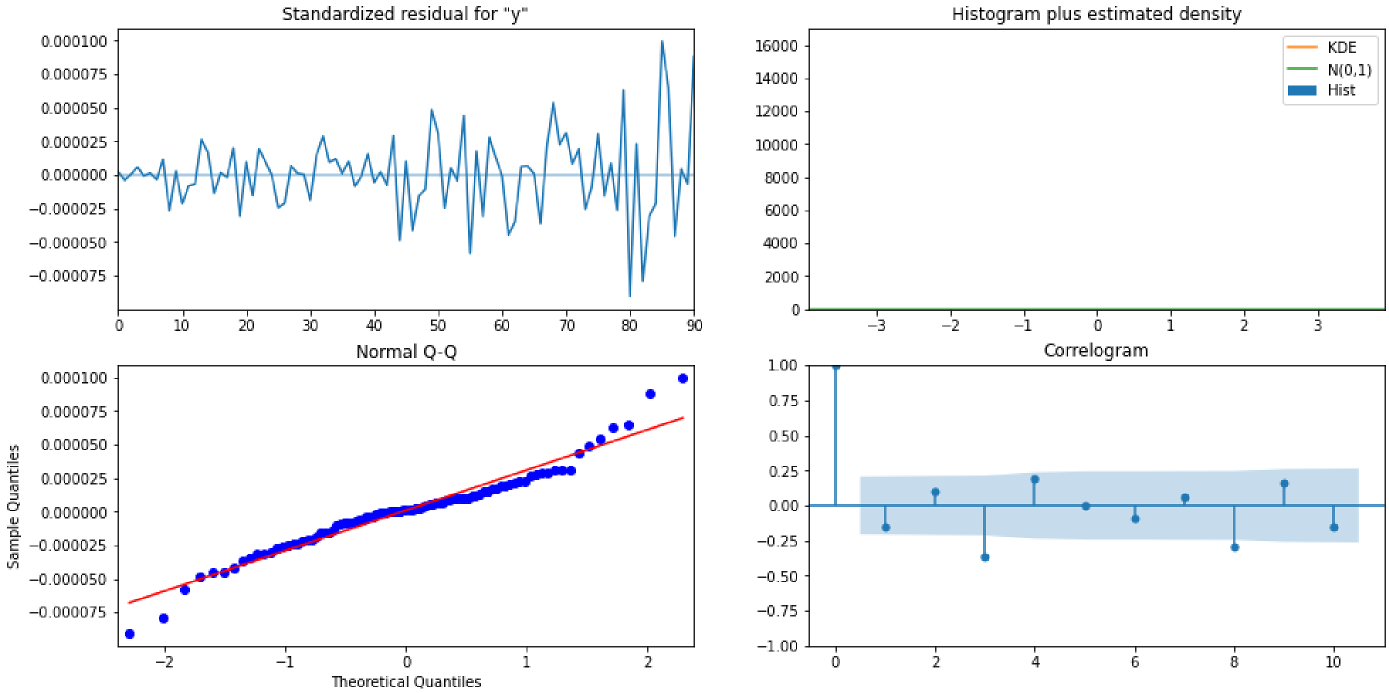
Optimum parameter

	param	seasonal	AIC
222	(3, 1, 1)	(3, 0, 2, 12)	774.400285
238	(3, 1, 2)	(3, 0, 2, 12)	774.880935
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)	775.561018

```
                                     SARIMAX Results
==========================================================================================
============
Dep. Variable:                                  Rose   No. Observations:          132
Model:             SARIMAX(3, 1, 1)x(3, 0, [1, 2], 12)   Log Likelihood        -377.200
Date:                               Sat, 24 Sep 2022   AIC                    774.400
Time:                                       18:44:28   BIC                    799.618
Sample:                                   01-31-1980   HQIC                   784.578
                                        - 12-31-1990
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     -1.837      0.066      -0.374       0.012
ma.L1         -0.9370      0.067    -13.904      0.000      -1.069      -0.805
ar.S.L12       0.7639      0.165      4.639      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      -0.4968      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       192.1628     39.631      4.849      0.000     114.488     269.838
==========================================================================================
```

```
Ljung-Box (L1) (Q):                   0.30   Jarque-Bera (JB):                 1.64
Prob(Q):                              0.58   Prob(JB):                         0.44
Heteroskedasticity (H):               1.11   Skew:                             0.33
Prob(H) (two-sided):                  0.77   Kurtosis:                         3.03
===================================================================================
```

Result SARIMA



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

Summary frame

Rose	mean	mean_se	mean_ci_lower	mean_ci_upper
1991-01-31	55.235183	13.907850	27.976299	82.494068
1991-02-28	68.123102	13.991296	40.700666	95.545538
1991-03-31	67.908695	14.012349	40.444996	95.372394
1991-04-30	66.786183	14.099653	39.151370	94.420996

Rose	mean	mean_se	mean_ci_lower	mean_ci_upper
1991-05-31	69.760050	14.109013	42.106893	97.413206

RMSE & MAPE

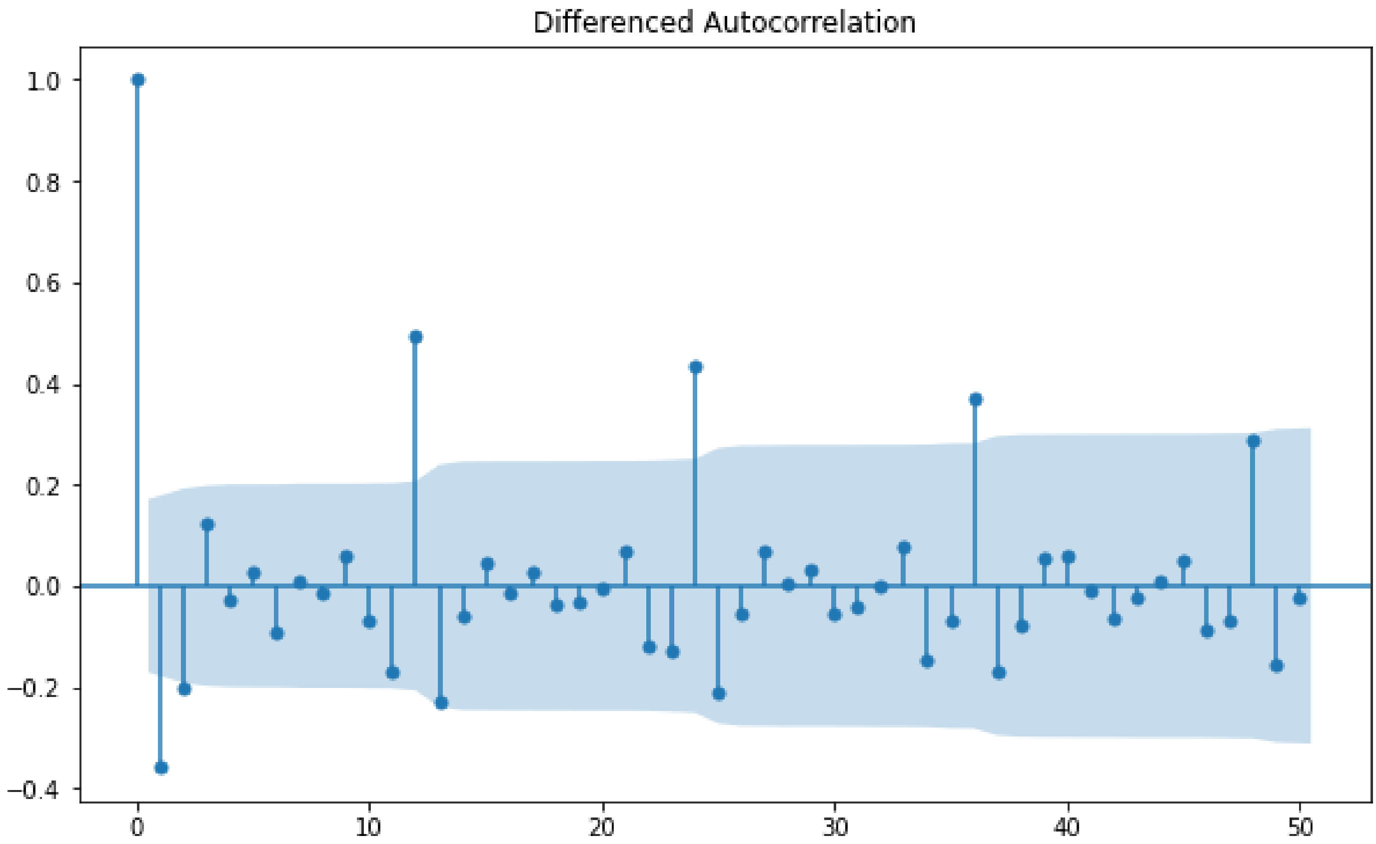
```
RMSE: 18.881945915548297
MAPE: 36.37549759795202
```

Table :

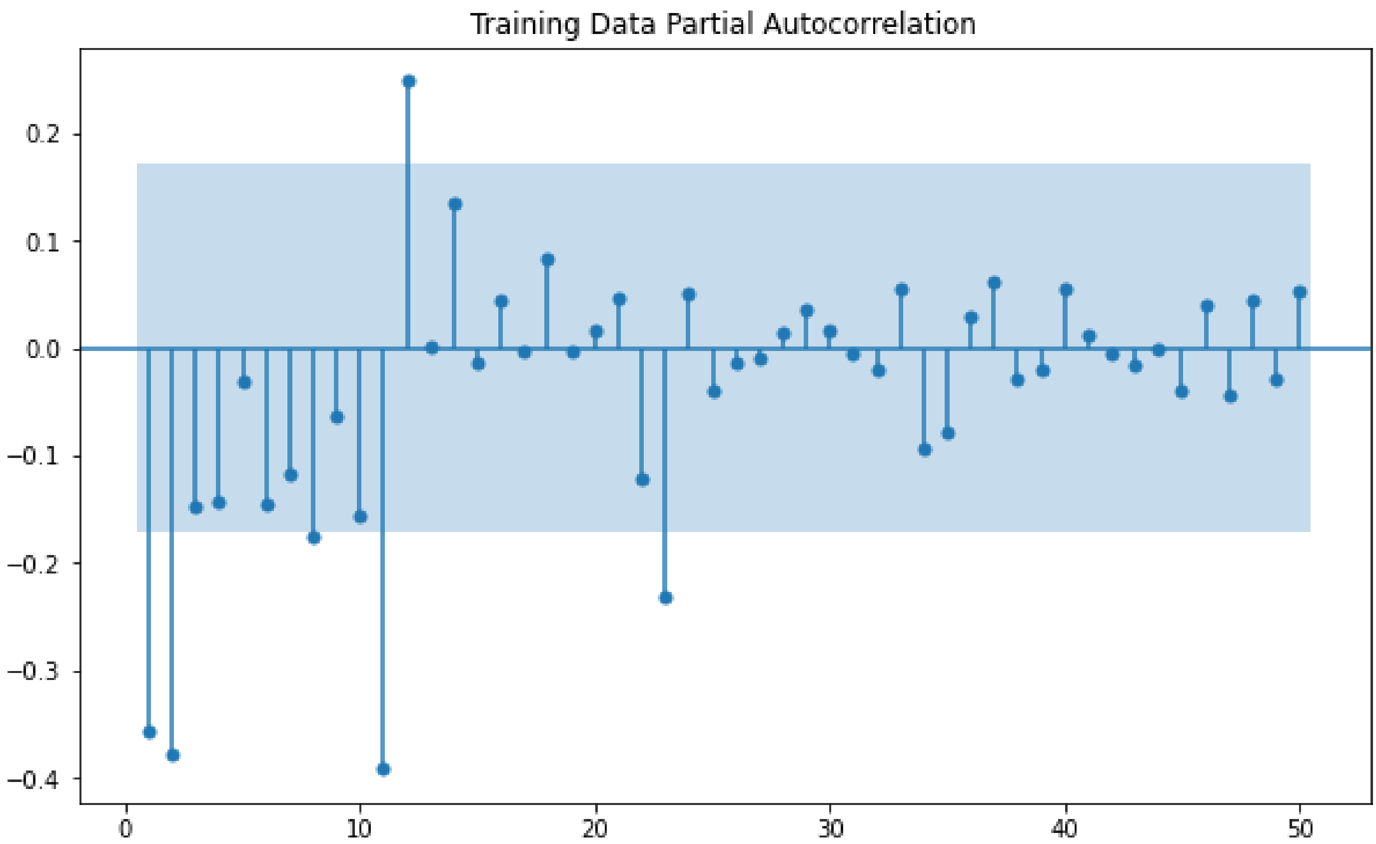
	Test RMSE	MAPE
RegressionOnTime	15.268955	NaN
NaiveModel	79.718773	NaN
SimpleAverageModel	53.460570	NaN
2pointTrailingMovingAverage	11.529278	NaN
4pointTrailingMovingAverage	14.451403	NaN
6pointTrailingMovingAverage	14.566327	NaN
9pointTrailingMovingAverage	14.727630	NaN
Alpha=0.10,SimpleExponentialSmoothing	36.796241	NaN
Alpha=1.49,Beta=1.66:DoubleExponentialSmoothing	15.268944	NaN
Alpha=0.09,Beta=0.00,Gamma=0.003:Triple Exponential Smoothing	14.249661	NaN
Alpha=0.07,Beta=0.05,Gamma=7.24:Triple Exponential Smoothing 2	20.156763	NaN
Arima (2,1,3)	36.817423	75.848378
Sarima (1,1,2)(2,0,2,6)	27.124076	55.238878
SARIMA(3,1,1)(3,0,2,12)	18.881946	36.375498

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

Training Data Autocorrelation



Training Data Partial Autocorrelation



Here, we have taken alpha=0.05.

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

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

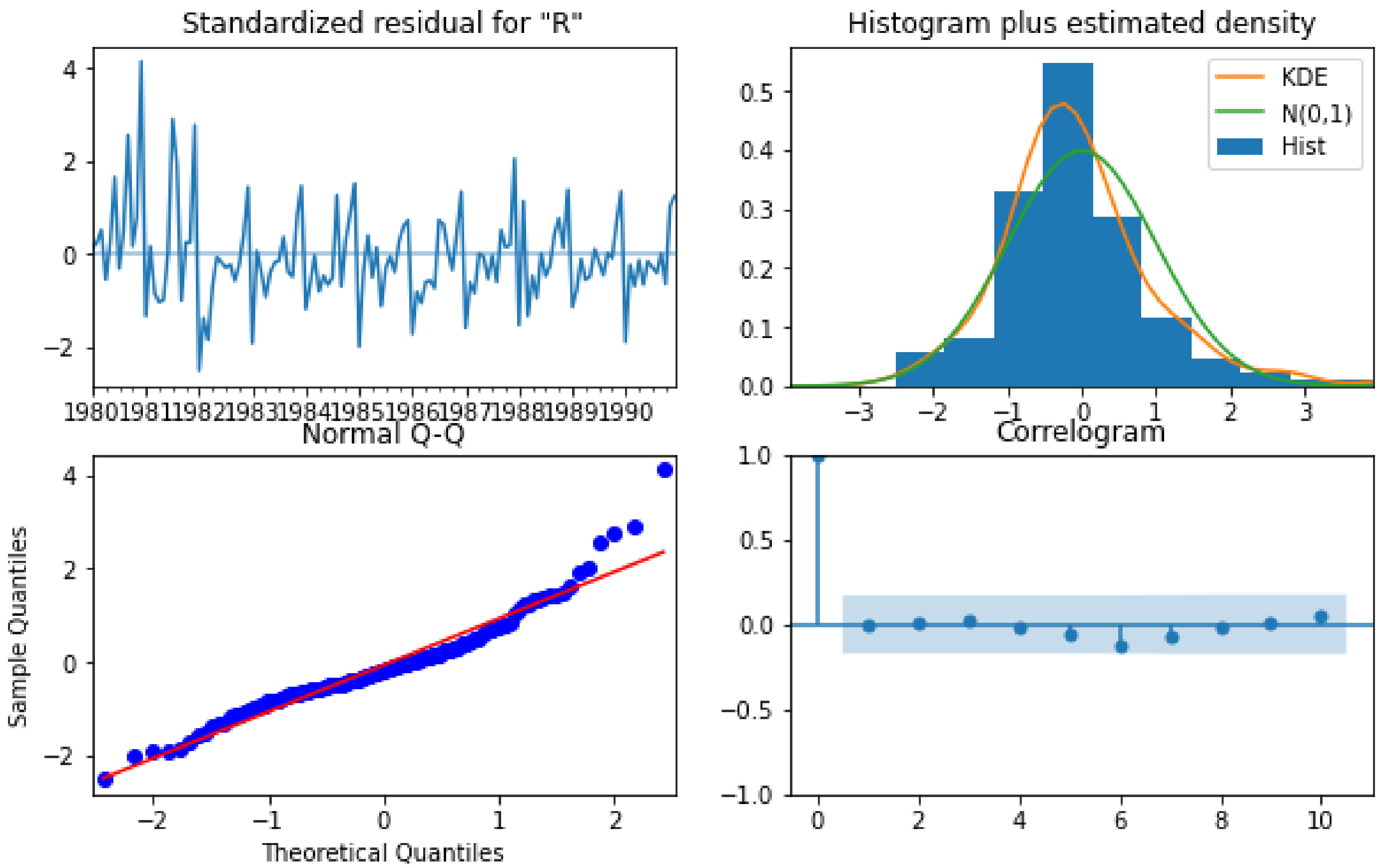
This is a common problem while building models by looking at the ACF and the PACF plots. But we are able to explain the model.

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

ARIMA

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                   Rose   No. Observations:                  132
Model:                 ARIMA(2, 0, 2)   Log Likelihood                -640.124
Date:                Sun, 25 Sep 2022   AIC                           1292.248
Time:                        19:26:54   BIC                           1309.545
Sample:                    01-31-1980   HQIC                          1299.277
                         - 12-31-1990
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const        107.7361     24.545      4.389      0.000      59.628     155.844
ar.L1          0.5278      0.293      1.804      0.071      -0.046       1.101
ar.L2          0.4511      0.278      1.624      0.104      -0.093       0.996
ma.L1         -0.2392      0.272     -0.880      0.379      -0.772       0.294
ma.L2         -0.5910      0.208     -2.847      0.004      -0.998      -0.184
sigma2       947.3279     98.828      9.586      0.000     753.628    1141.028
===================================================================================
Ljung-Box (L1) (Q):                   0.00   Jarque-Bera (JB):                50.35
Prob(Q):                              0.98   Prob(JB):                         0.00
Heteroskedasticity (H):               0.38   Skew:                             0.94
Prob(H) (two-sided):                  0.00   Kurtosis:                         5.37
===================================================================================
```

Results auto ARIMA



Table

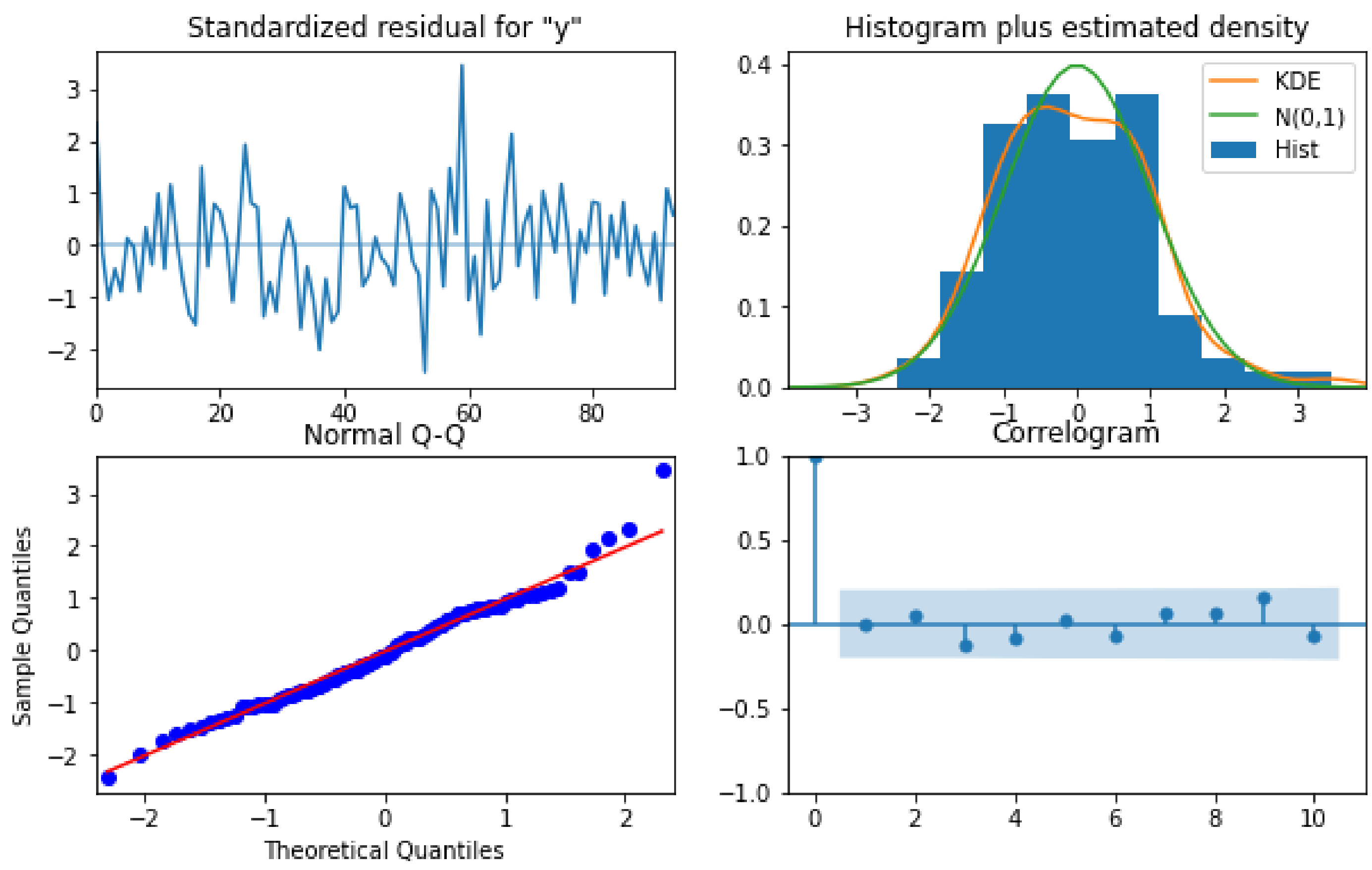
	Test RMSE	MAPE
RegressionOnTime	15.268955	NaN
NaiveModel	79.718773	NaN
SimpleAverageModel	53.460570	NaN
2pointTrailingMovingAverage	11.529278	NaN
4pointTrailingMovingAverage	14.451403	NaN
6pointTrailingMovingAverage	14.566327	NaN
9pointTrailingMovingAverage	14.727630	NaN
Alpha=0.10,SimpleExponentialSmoothing	36.796241	NaN
Alpha=1.49,Beta=1.66:DoubleExponentialSmoothing	15.268944	NaN

	Test RMSE	MAPE
Alpha=0.09,Beta=0.00,Gamma=0.003:Triple Exponential Smoothing	14.249661	NaN
Alpha=0.07,Beta=0.05,Gamma=7.24:Triple Exponential Smoothing 2	20.156763	NaN
Arima (2,1,3)	36.817423	75.848378
Sarima (1,1,2)(2,0,2,6)	27.124076	55.238878
SARIMA(3,1,1)(3,0,2,12)	18.881946	36.375498
Arima (2,0,2)	45.224558	93.349757

SARIMA

```
                                    SARIMAX Results
==========================================================================================
=======
Dep. Variable:                                  y   No. Observations:                  132
Model:             SARIMAX(2, 0, 2)x(3, 0, 2, 12)   Log Likelihood                -386.084
Date:                            Sun, 25 Sep 2022   AIC                            792.168
Time:                                    19:27:01   BIC                            817.601
Sample:                                         0   HQIC                           802.441
                                            - 132
Covariance Type:                              opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -1.0161      0.636     -1.597      0.110      -2.263       0.231
ar.L2         -0.1892      0.580     -0.326      0.744      -1.326       0.948
ma.L1          1.0892      0.635      1.715      0.086      -0.156       2.334
ma.L2          0.1855      0.608      0.305      0.760      -1.006       1.377
ar.S.L12       0.7937      0.117      6.770      0.000       0.564       1.023
ar.S.L24       0.1189      0.131      0.908      0.364      -0.138       0.375
ar.S.L36       0.0318      0.078      0.406      0.685      -0.122       0.186
ma.S.L12      -0.6147      0.367     -1.674      0.094      -1.335       0.105
ma.S.L24      -0.2486      0.243     -1.021      0.307      -0.726       0.229
sigma2       183.5680     58.477      3.139      0.002      68.956     298.180
===================================================================================
Ljung-Box (L1) (Q):                   0.00   Jarque-Bera (JB):                 4.04
Prob(Q):                              0.97   Prob(JB):                         0.13
Heteroskedasticity (H):               0.70   Skew:                             0.41
Prob(H) (two-sided):                  0.32   Kurtosis:                         3.60
===================================================================================
```

Results SARIMA New



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

RMSE & MAPE

```
RMSE: 16.947204698440167
MAPE: 31.29841985287618
```

Table

	Test RMSE	MAPE
RegressionOnTime	15.268955	NaN
NaiveModel	79.718773	NaN
SimpleAverageModel	53.460570	NaN
2pointTrailingMovingAverage	11.529278	NaN
4pointTrailingMovingAverage	14.451403	NaN

	Test RMSE	MAPE
6pointTrailingMovingAverage	14.566327	NaN
9pointTrailingMovingAverage	14.727630	NaN
Alpha=0.10,SimpleExponentialSmoothing	36.796241	NaN
Alpha=1.49,Beta=1.66:DoubleExponentialSmoothing	15.268944	NaN
Alpha=0.09,Beta=0.00,Gamma=0.003:Triple Exponential Smoothing	14.249661	NaN
Alpha=0.07,Beta=0.05,Gamma=7.24:Triple Exponential Smoothing 2	20.156763	NaN
Arima (2,1,3)	36.817423	75.848378
Sarima (1,1,2)(2,0,2,6)	27.124076	55.238878
SARIMA(3,1,1)(3,0,2,12)	18.881946	36.375498
Arima (2,0,2)	45.224558	93.349757
SARIMA(2, 0, 2)(3, 0, 2, 12)	16.947205	31.298420

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

```
Sorted by RMSE values on the Test Data:
                                                     Test RMSE       MAPE
2pointTrailingMovingAverage                         11.529278        NaN
Alpha=0.09,Beta=0.00,Gamma=0.003:Triple Exponen...  14.249661        NaN
4pointTrailingMovingAverage                         14.451403        NaN
6pointTrailingMovingAverage                         14.566327        NaN
9pointTrailingMovingAverage                         14.727630        NaN
Alpha=1.49,Beta=1.66:DoubleExponentialSmoothing     15.268944        NaN
RegressionOnTime                                    15.268955        NaN
SARIMA(2, 0, 2)(3, 0, 2, 12)                        16.947205  31.298420
SARIMA(3,1,1)(3,0,2,12)                             18.881946  36.375498
Alpha=0.07,Beta=0.05,Gamma=7.24:Triple Exponent...  20.156763        NaN
Sarima (1,1,2)(2,0,2,6)                             27.124076  55.238878
Alpha=0.10,SimpleExponentialSmoothing               36.796241        NaN
Arima (2,1,3)                                       36.817423  75.848378
Arima (2,0,2)                                       45.224558  93.349757
```

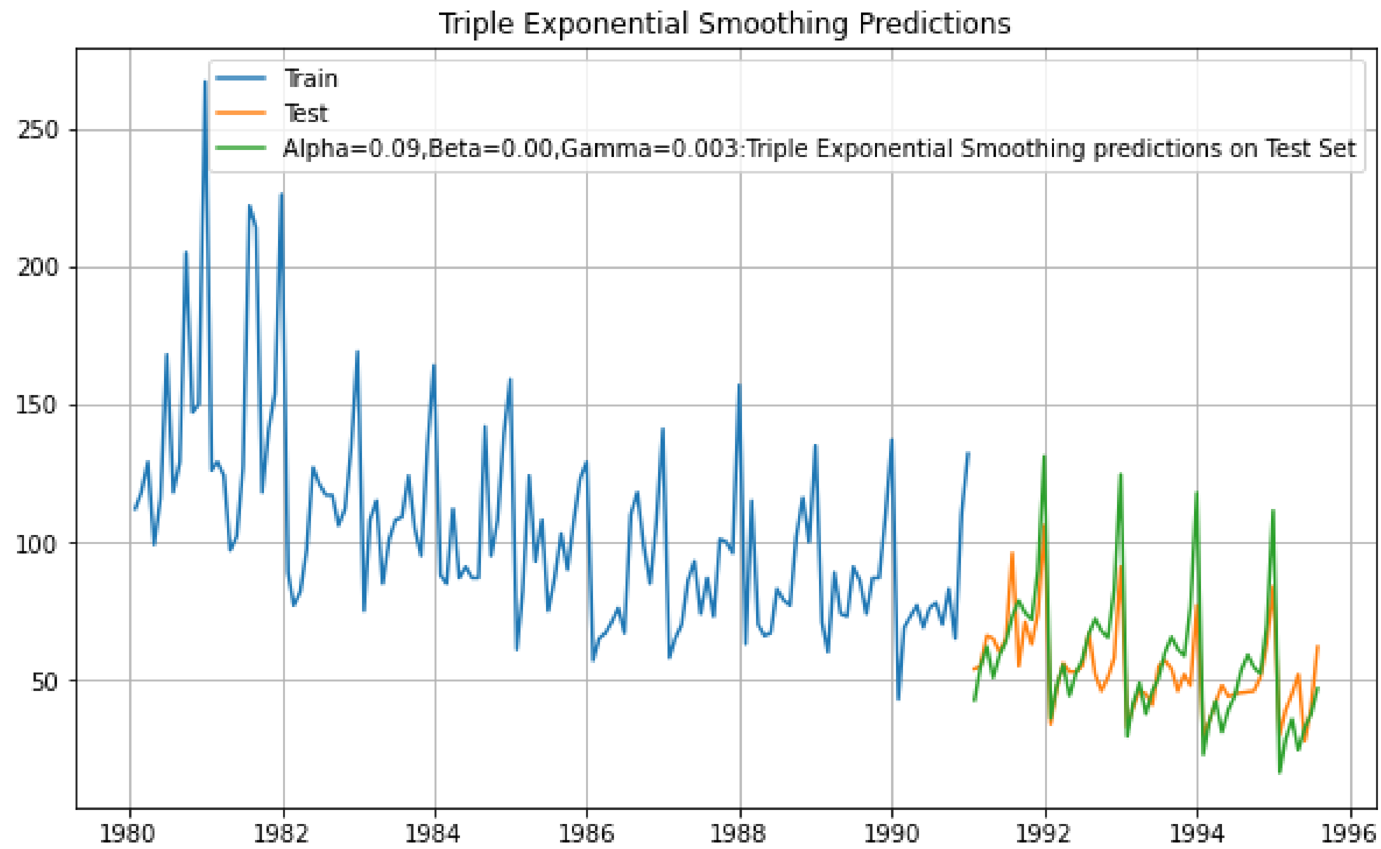
```
SimpleAverageModel                                        53.460570        NaN
NaiveModel                                                79.718773        NaN
```

```
Sorted by MAPE values on the Test Data:
                                                           Test RMSE       MAPE
SARIMA(2, 0, 2)(3, 0, 2, 12)                              16.947205  31.298420
SARIMA(3,1,1)(3,0,2,12)                                   18.881946  36.375498
Sarima (1,1,2)(2,0,2,6)                                   27.124076  55.238878
Arima (2,1,3)                                             36.817423  75.848378
Arima (2,0,2)                                             45.224558  93.349757
RegressionOnTime                                          15.268955        NaN
NaiveModel                                                79.718773        NaN
SimpleAverageModel                                        53.460570        NaN
2pointTrailingMovingAverage                               11.529278        NaN
4pointTrailingMovingAverage                               14.451403        NaN
6pointTrailingMovingAverage                               14.566327        NaN
9pointTrailingMovingAverage                               14.727630        NaN
Alpha=0.10,SimpleExponentialSmoothing                     36.796241        NaN
Alpha=1.49,Beta=1.66:DoubleExponentialSmoothing           15.268944        NaN
Alpha=0.09,Beta=0.00,Gamma=0.003:Triple Exponen...        14.249661        NaN
Alpha=0.07,Beta=0.05,Gamma=7.24:Triple Exponent...        20.156763        NaN
```

Although we are seeing that the best model as per RMSE is the 2pointTrailingMovingAverage 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 2pointTrailingMovingAverage value hence we can choose this model as well.

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

Triple Exponential Smoothing Predictions



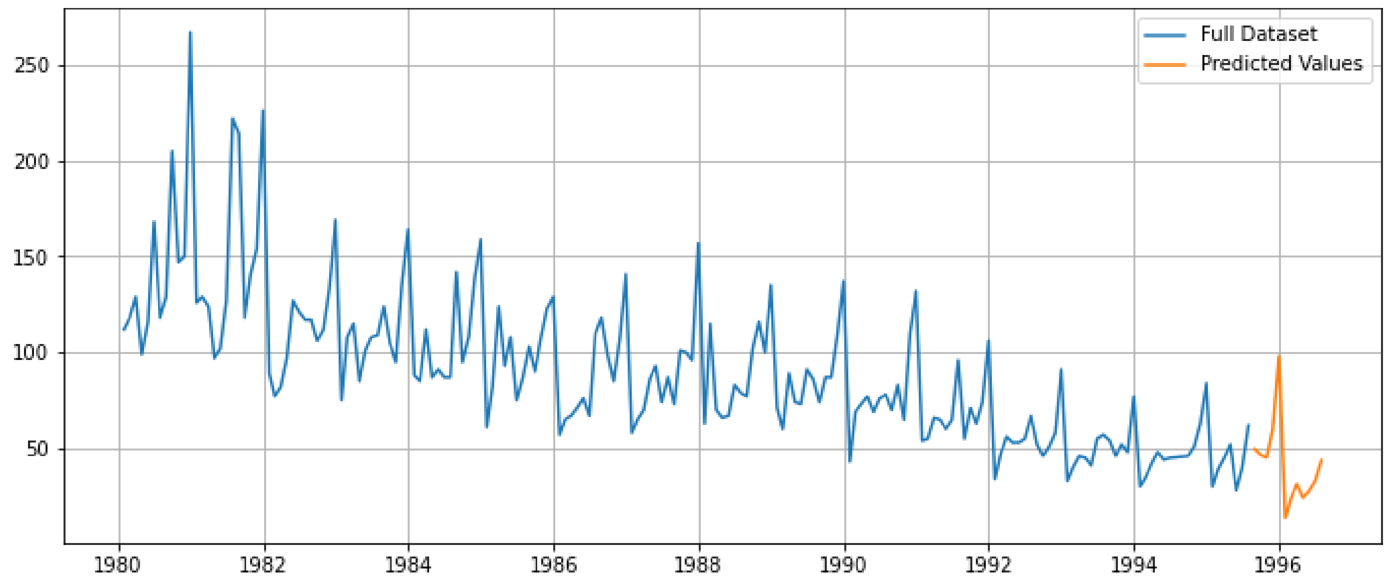
RMSE

RMSE of the Full Model 17.68567622997536

Getting the predictions for the 12 months

1995-08-31	49.598298
1995-09-30	46.440145
1995-10-31	45.178368
1995-11-30	59.780864
1995-12-31	98.069768
1996-01-31	13.600846
1996-02-29	23.900520
1996-03-31	31.450408
1996-04-30	24.243478
1996-05-31	27.593761
1996-06-30	33.089608
1996-07-31	43.695300

Predicted Values



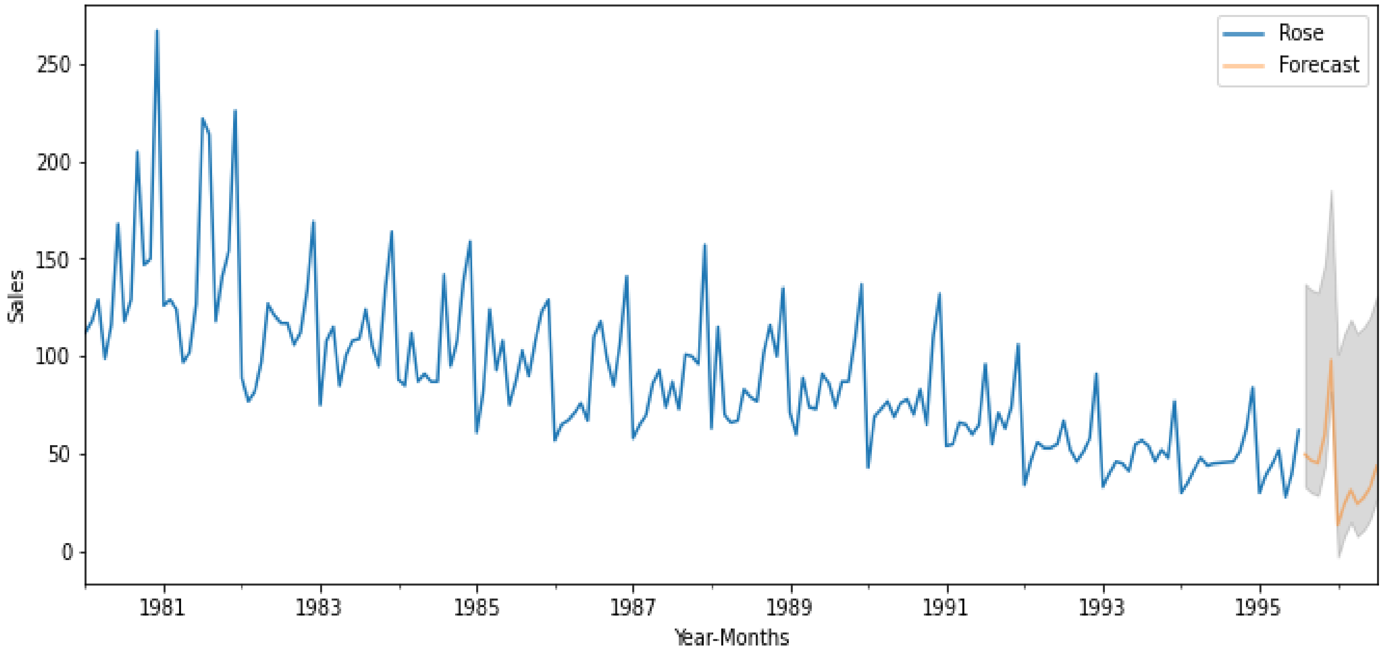
In the below code, we have calculated the upper and lower confidence bands at 95% confidence level

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

This gives us the necessary confidence bands for the predictions

	lower_CI	prediction	upper_ci
1995-08-31	33.165042	49.598298	137.138617
1995-09-30	30.006889	46.440145	133.980465
1995-10-31	28.745112	45.178368	132.718688
1995-11-30	43.347608	59.780864	147.321184
1995-12-31	81.636512	98.069768	185.610087

plotting the forecast along with the confidence band:



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

We can see here that the sales are going down as we have to do something about it because it shouldn't be continue like that here are my some suggestion to solve the problems.

1. If the wine is not tasty enough for the consumer we should perform some test and make the taste better so that the sales can increase as more number of consumer will start consuming it.

2. We can also supply or can say do partnership with hotels so that they start giving these wines at their hotel so the reviews can increse also it will be advertised and also will be consumed more and hence the people will start to like it and the sales can increase.

3. We should establish schemes and discount and different types of offers so that we can attract the customers ans then they buy the wine.

4. We can see that in the last months the sales of the wine is high so we need to address other months as well and find out what the problem is. is it the taste of the wine, the quality of the wine or the price of the wine which can be playing a role in the sales of the wine.