Project Time series forecasting

Parnoshree Chatterjee

A. **Rose.csv Dataset:**

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

Answer:

Table 1. Head 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

Table 2. Tail of the dataset

	YearMonth	Rose
182	1995-03	45.0
183	1995-04	52.0
184	1995-05	28.0
185	1995-06	40.0
186	1995-07	62.0

Table 3. Creating the Time Stamps and adding to the data frame to make it a Time Series Data

```
DatetimeIndex(['1980-01-31', '1980-02-29', '1980-03-31', '1980-04-30',
               '1980-05-31', '1980-06-30', '1980-07-31', '1980-08-31',
               '1980-09-30', '1980-10-31',
               ...
               '1994-10-31', '1994-11-30', '1994-12-31', '1995-01-31',
               '1995-02-28', '1995-03-31', '1995-04-30', '1995-05-31',
               '1995-06-30', '1995-07-31'],
              dtype='datetime64[ns]', length=187, freq='M')
```

Table 4. Adding Time stamp

	YearMonth	Rose	Time_Stamp
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

Time_Stamp	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

Plotting the same graph from the second data frame with the date-time modifications

- In the first question we are reading the date set and checking the head and tail.
- It is to understand what is the start and end date of the time series.
- The above Sparkling data set starts from 1980-1991.
- We have two variables with name YearMonth and Sparkling
- Sparkling is the target variable.
- As we are trying to find the sales for Sparkling wine it is important to check and understand what are the variables, we will be working with

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

Answer:

- First we check the table for monthly sparks throughout the year using monthly data.

- There are no duplicates in the data set.
- We do have 2 null values in the data set which we will be using Median imputation
- We describe the function to check the description of the data set.
- We have standard deviation of 38.967 and mean of 90.348
- The shape of the data set is 187,1

Table 5. For monthly Rose across years

Time_Stamp	April	August	December	February	January	July	June	March	May	November	October	September
Time_Stamp												
1980	99.0	129.0	267.0	118.0	112.0	118.0	168.0	129.0	116.0	150.0	147.0	205.0
1981	97.0	214.0	226.0	129.0	126.0	222.0	127.0	124.0	102.0	154.0	141.0	118.0
1982	97.0	117.0	169.0	77.0	89.0	117.0	121.0	82.0	127.0	134.0	112.0	106.0
1983	85.0	124.0	164.0	108.0	75.0	109.0	108.0	115.0	101.0	135.0	95.0	105.0
1984	87.0	142.0	159.0	85.0	88.0	87.0	87.0	112.0	91.0	139.0	108.0	95.0
1985	93.0	103.0	129.0	82.0	61.0	87.0	75.0	124.0	108.0	123.0	108.0	90.0
1986	71.0	118.0	141.0	65.0	57.0	110.0	67.0	67.0	76.0	107.0	85.0	99.0
1987	86.0	73.0	157.0	65.0	58.0	87.0	74.0	70.0	93.0	96.0	100.0	101.0
1988	66.0	77.0	135.0	115.0	63.0	79.0	83.0	70.0	67.0	100.0	116.0	102.0
1989	74.0	74.0	137.0	60.0	71.0	86.0	91.0	89.0	73.0	109.0	87.0	87.0
1990	77.0	70.0	132.0	69.0	43.0	78.0	76.0	73.0	69.0	110.0	65.0	83.0
1991	65.0	55.0	106.0	55.0	54.0	96.0	65.0	66.0	60.0	74.0	63.0	71.0
1992	53.0	52.0	91.0	47.0	34.0	67.0	55.0	56.0	53.0	58.0	51.0	46.0
1993	45.0	54.0	77.0	40.0	33.0	57.0	55.0	46.0	41.0	48.0	52.0	46.0
1994	48.0	NaN	84.0	35.0	30.0	NaN	45.0	42.0	44.0	63.0	51.0	46.0
1995	52.0	NaN	NaN	39.0	30.0	62.0	40.0	45.0	28.0	NaN	NaN	NaN

Checking for Null values:

```
Rose    2
dtype: int64
```

Shape of the data set:

```
(187, 1)
```

Info of the data set:

```
<class 'pandas.core.frame.DataFrame'>
DatetimeIndex: 187 entries, 1980-01-31 to 1995-07-31
Data columns (total 1 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   Rose    187 non-null    float64
dtypes: float64(1)
memory usage: 2.9 KB
```

Table 6 . Describe function

	Rose
count	187.000
mean	90.348
std	38.967
min	28.000
25%	63.000
50%	86.000
75%	111.000
max	267.000

Fig 2. Plot for Monthly Sparks throughout year

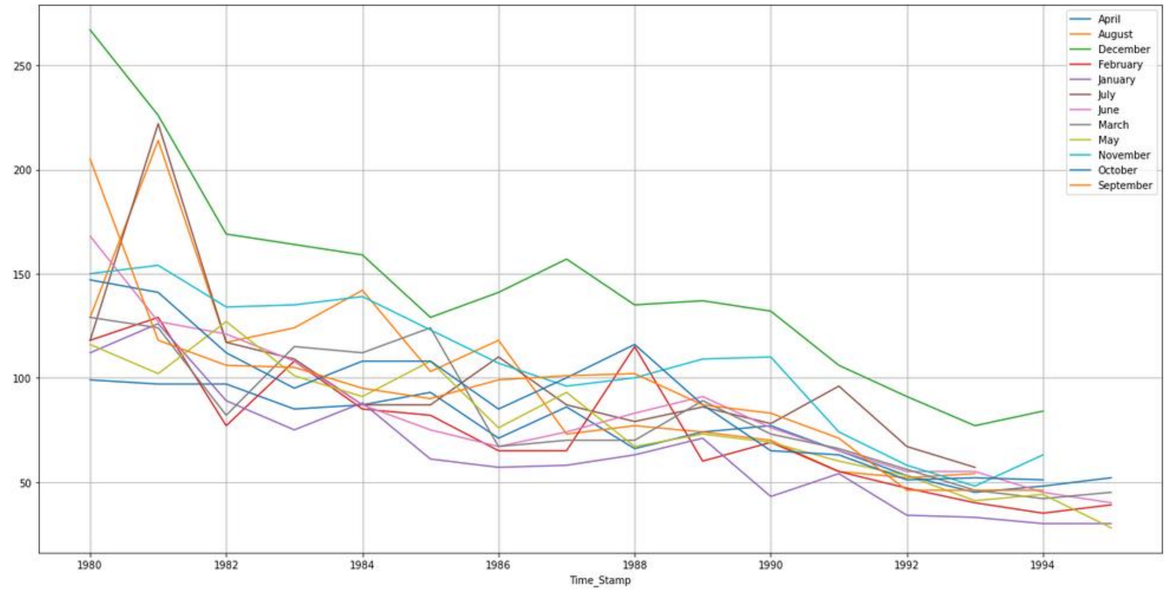


Fig 3. Plot for Yearly box plot

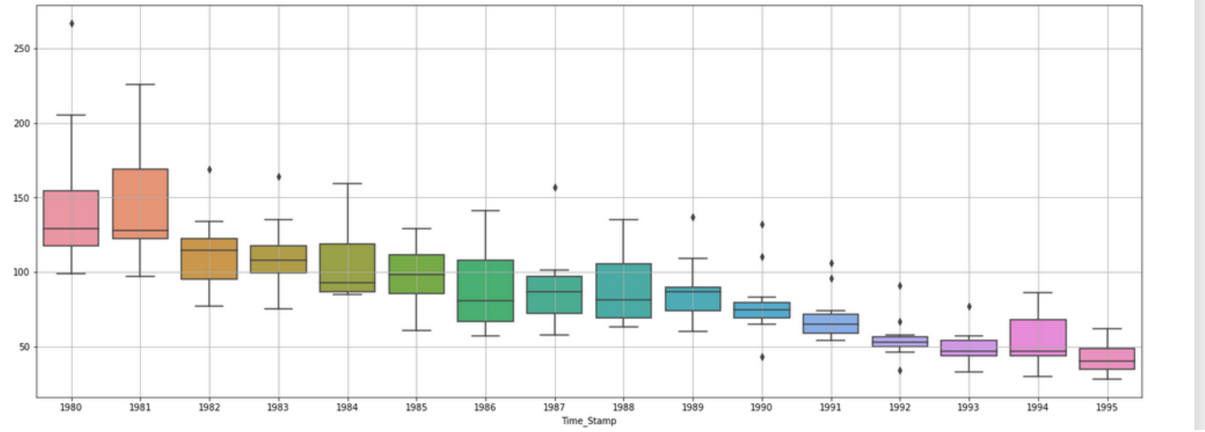


Fig 4. Plot for Monthly Boxplot

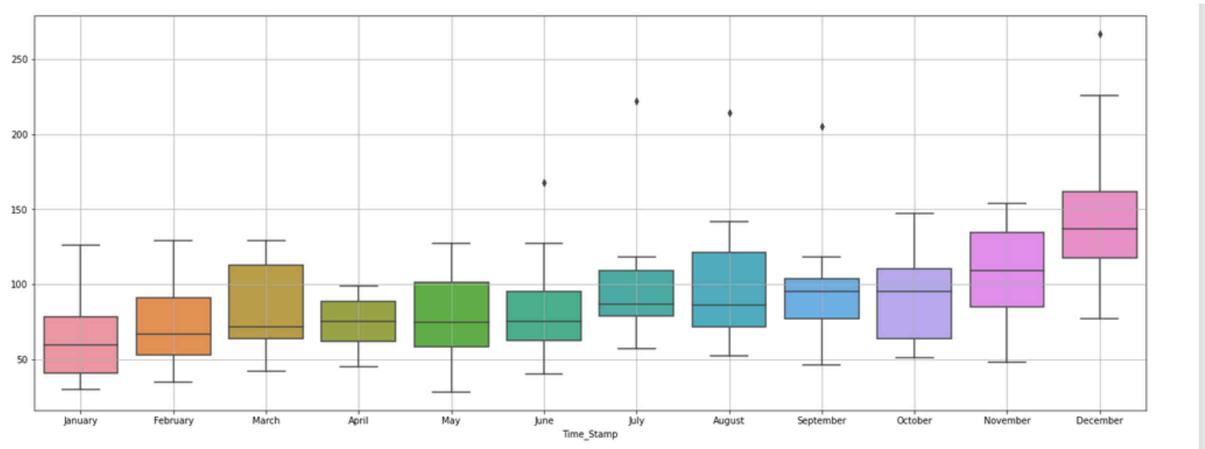
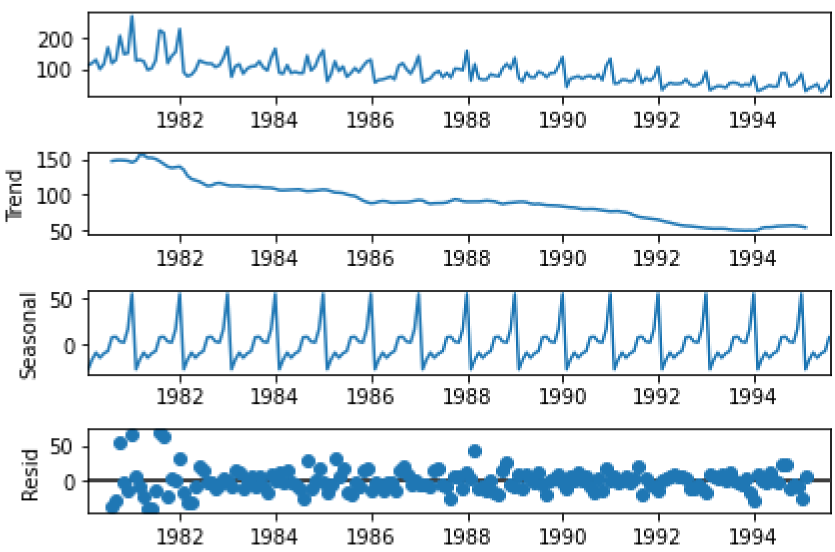


Fig 5. Plot for Decomposition



Decomposition Trend, Seasonality and Residual:

```
Trend
 Time_Stamp
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
Name: trend, dtype: float64

Seasonality
 Time_Stamp
1980-01-31   -28.355258
1980-02-29   -17.794345
1980-03-31    -9.764583
1980-04-30   -15.577083
1980-05-31   -10.675298
1980-06-30    -8.157440
1980-07-31     7.161409
1980-08-31     7.741964
1980-09-30     2.328075
1980-10-31     1.425298
1980-11-30    16.400298
1980-12-31    55.266964
Name: seasonal, dtype: float64

Residual
 Time_Stamp
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   -36.244742
1980-08-31   -26.866964
1980-09-30    54.296925
1980-10-31    -2.508631
1980-11-30   -13.816964
1980-12-31    66.608036
```

- Next we plot a graph to check monthly and yearly Rose.
- We have also used box plot to show the monthly and yearly Rose.
- As we see there are no outliers in the data set.

- We have plotted a decomposition graph.
- A **decomposition** of a **graph** is a collection of edge-disjoint subgraphs of such that every edge of belongs to exactly one . If each is a path or a cycle in , then is called a path **decomposition** of . If each is a path in , then is called an acyclic path **decomposition.**
- We can see the trend, seasonality and Residual in the above graph of Fig 5.
- We can also see the values for trend, seasonality and Residual in Table 7.

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

Answer:

<u>Shape of train and test data set is</u>

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

- We have split the data into Train and Test the above table 8 show the first and last few rows of train and test data.
- The shape of the dataset has changed to 132,1 and 55,1
- We have made sure the split for test data starts from 1991 as per the requirements.

Table 7. Table showing Trend, Seasonality and Residual for Rose

```
                 Rose
Time_Stamp
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
                 Rose
Time_Stamp
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
                Rose
Time_Stamp
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
                Rose
Time_Stamp
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
```

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

Model 1: Linear Regression

Table 9. For Training Time Instance and Test Time Instance

```
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, 3
3, 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]
Test Time instance
 [133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 1
57, 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]
```

- Here our first model is Linear regression I will be briefly explaining what are the steps used during the modelling process.
- In statistics, linear regression is a linear approach to modelling the relationship between a scalar response and one or more explanatory variables. The case of one explanatory variable is called simple linear regression; for more than one, the process is called multiple linear regression
- The above table 9 shows us the training and testing time instance.
- The below table 10 talks about the first and last few rows of training and testing data set. It is important to check the values to understand the difference made to the dataset.

Table 10. First few and Last few rows of training and testing dataset:

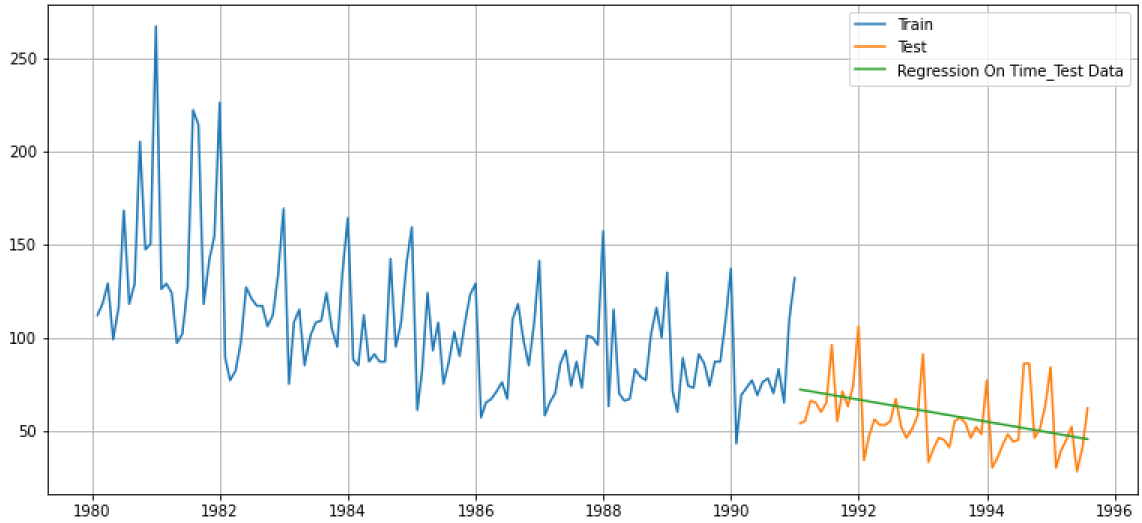
```
First few rows of Training Data
               Rose  time
Time_Stamp
1980-01-31  112.0     1
1980-02-29  118.0     2
1980-03-31  129.0     3
1980-04-30   99.0     4
1980-05-31  116.0     5

Last few rows of Training Data
               Rose  time
Time_Stamp
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
              Rose  time
Time_Stamp
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
              Rose  time
Time_Stamp
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
```

Fig 6. Test_Predictions_Model1



Test RMSE for Linear Regression

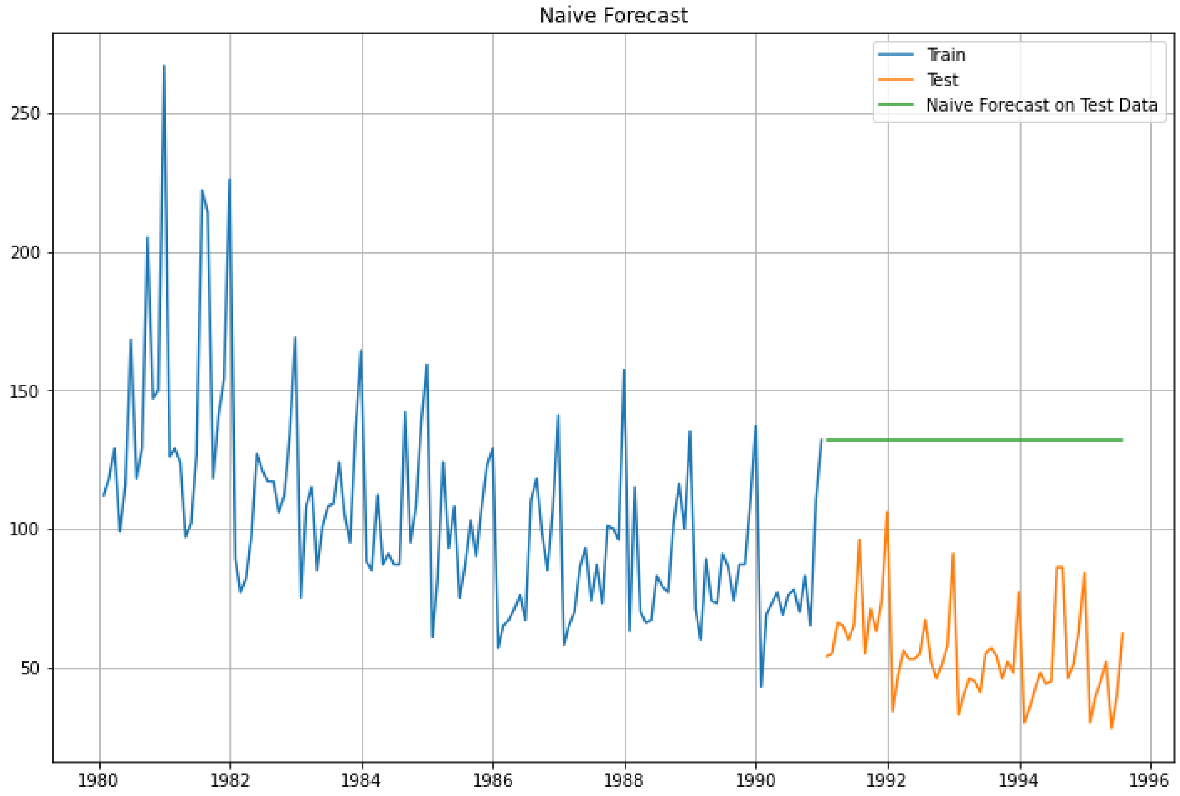
	Test RMSE
RegressionOnTime	16.626144

Model 2: Naive Approach

Head of the data set

```
Time_Stamp
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 7.* Plot for Naive Approach**



Model evaluation using RMSE

```
For RegressionOnTime forecast on the Test Data,  RMSE is 78.485
```

Test RMSE:

	Test RMSE
RegressionOnTime	16.626144
NaiveModel	78.485320

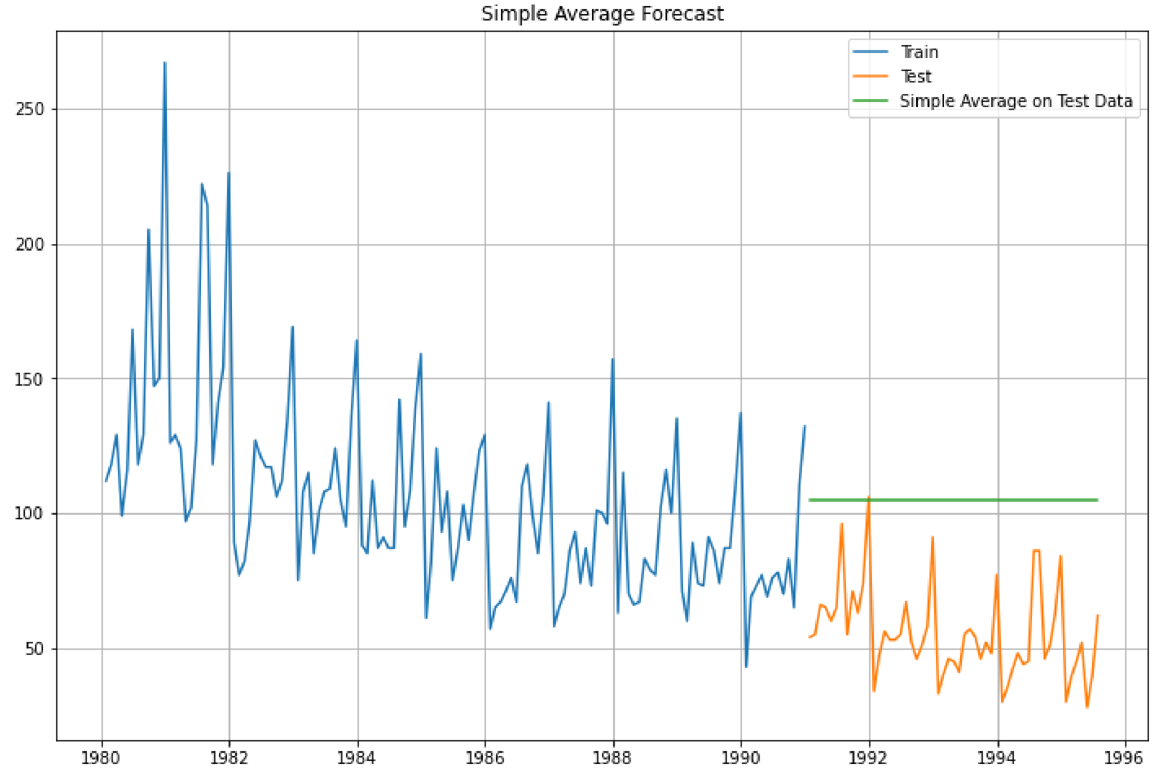
- A **model** in which minimum amounts of effort and manipulation of data are used to prepare a forecast. Most often naïve models used are random walk (current value as a forecast of the next period) and seasonal random walk (value from the same period of prior year as a forecast for the same period of forecasted year.)
- We have read the data head to check the values.
- Fig 7 shows the trend on the test and train data.
- The RMSE for Naïve Model is 78 which is higher than the LR model.
- We have noted the value for both the models above for a better comparison.

Method 3: Simple Average

Head of the data set

Time_Stamp	Rose	mean_forecast
1991-01-31	54.0	104.939394
1991-02-28	55.0	104.939394
1991-03-31	66.0	104.939394
1991-04-30	65.0	104.939394
1991-05-31	60.0	104.939394

Fig 8. Plot for Simple Average



Model evaluation using RMSE

```
For Simple Average forecast on the Test Data,  RMSE is 52.370
```

RMSE for Simple Average:

	Test RMSE
RegressionOnTime	16.626144
NaiveModel	78.485320
SimpleAverageModel	52.369847

- The **simple average** of a set of observations is computed as the sum of the individual observations divided by the number of observations in the set.
- We have read the data head to check the values.
- Fig 8. shows the trend on the test and train data.
- The RMSE for Simple average is 52 which is lower than the Naïve model and higher than LR
- We have noted the value for all three models above for a better comparison.

Method 4: Moving Average (MA)

Head of the data set

Time_Stamp	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

Head after adding Trailing_2 to Trailing_9

Time_Stamp	Rose	Trailing_2	Trailing_4	Trailing_6	Trailing_9
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 9. Plot for Moving Average

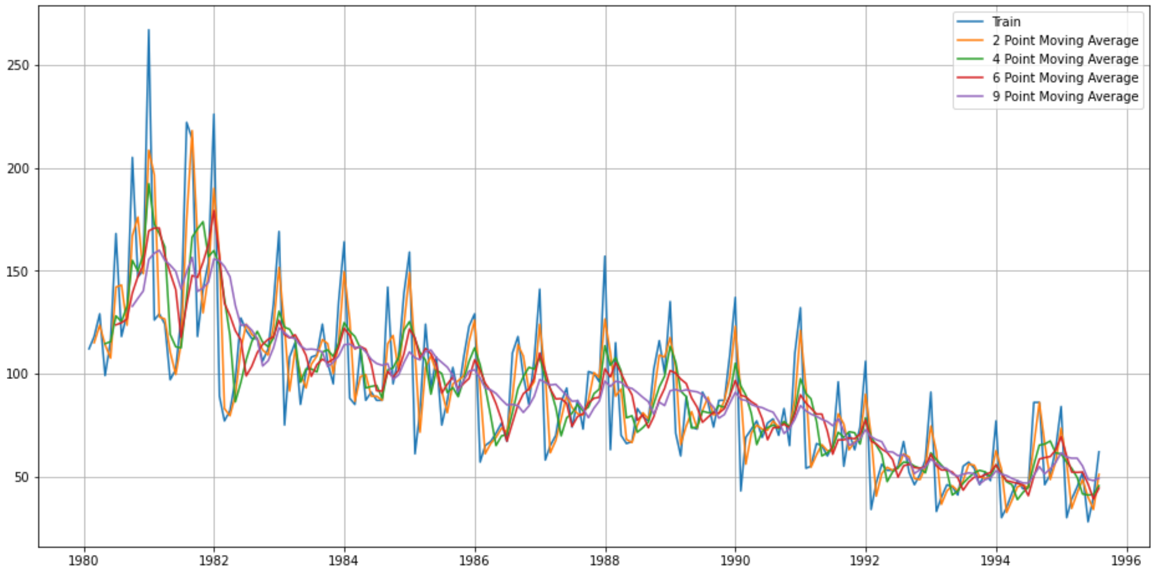
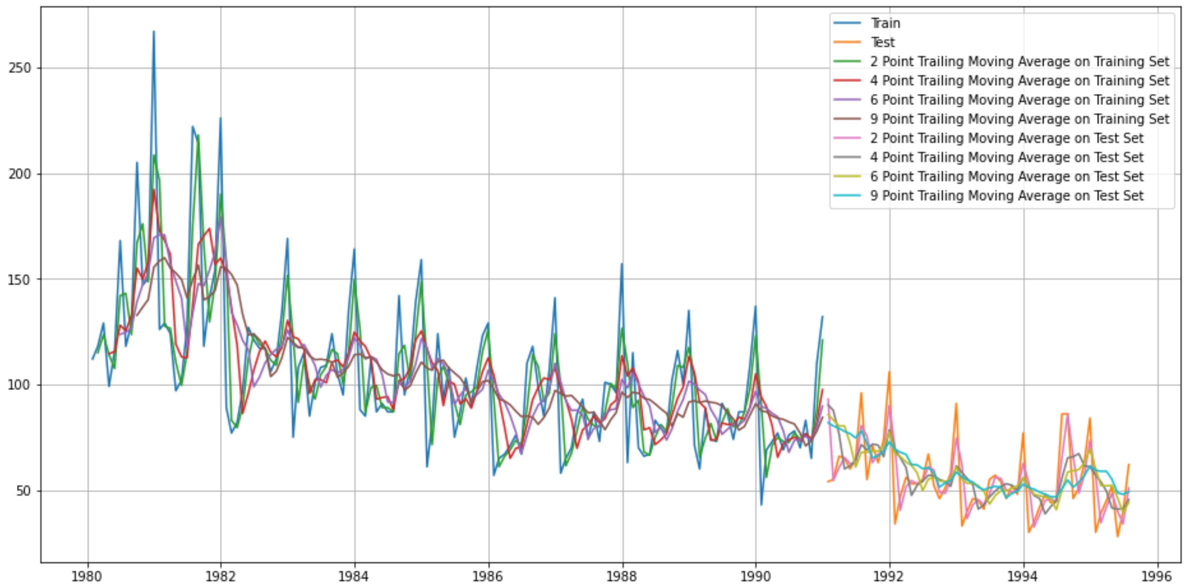


Fig 10. Plotting on both Train and Test



Model Evaluation Done only on the test data.

```
For 2 point Moving Average Model forecast on the Training Data,  RMSE is 12.159
For 4 point Moving Average Model forecast on the Training Data,  RMSE is 15.572
For 6 point Moving Average Model forecast on the Training Data,  RMSE is 15.687
For 9 point Moving Average Model forecast on the Training Data,  RMSE is 16.161
```

Test RMSE:

	Test RMSE
RegressionOnTime	16.626144
NaiveModel	78.485320
SimpleAverageModel	52.369847
2pointTrailingMovingAverage	12.158798
4pointTrailingMovingAverage	15.572375
6pointTrailingMovingAverage	15.687446
9pointTrailingMovingAverage	16.161176

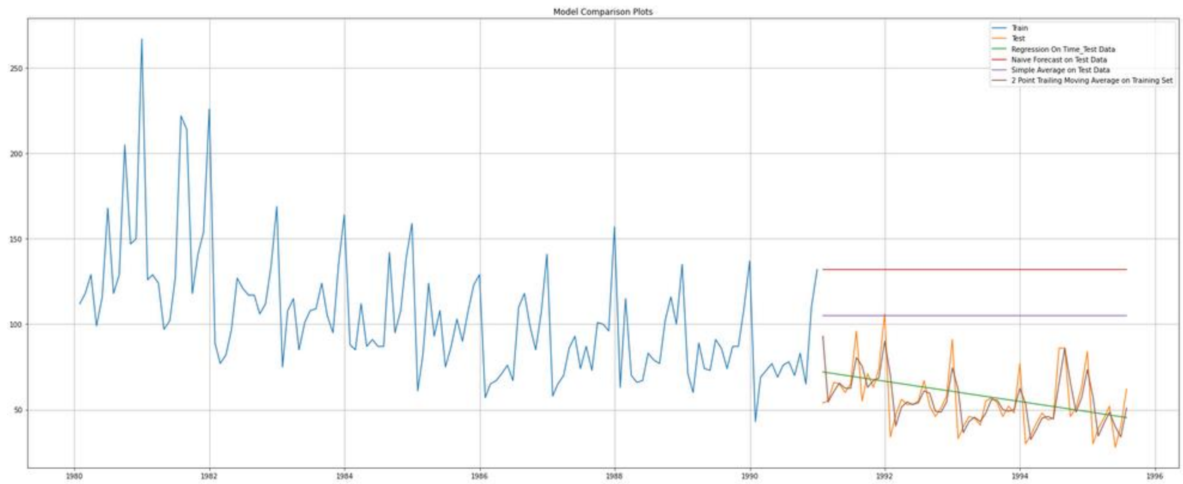
Creating train and test set

- In statistics, a moving average is a calculation to analyze data points by creating a series of averages of different subsets of the full data set. It is also called a moving mean or rolling mean and is a type of finite impulse response filter. Variations include: simple, and cumulative, or weighted forms.

- We have read the data head to check the values.
- Fig 9. shows the trend on the test and train data.
- For 2 point moving average model forecast on the training data RMSE is 12.19
- For 4 point moving average model forecast on the training data RMSE is 15.57
- For 6 point moving average model forecast on the training data RMSE is 15.68
- For 9 point moving average model forecast on the training data RMSE is 16.16
- We have noted the value for all the models above in the Table 11 for a better comparison.
- We have also plotted all the models so far to show the comparison as of now the best model performance is for 2 point moving average model forecast on the training data RMSE is 12.15

Before we go on to build the various Exponential Smoothing models, let us plot all the models and compare the Time Series plots.

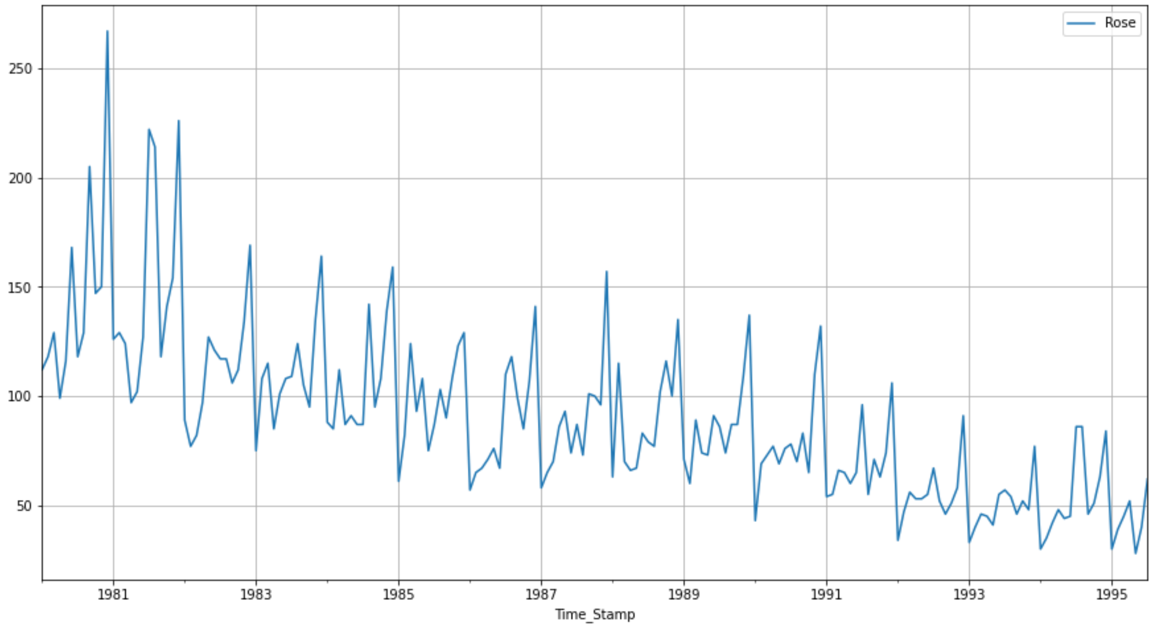
Fig 11. All the above models



Method 5: Simple Exponential Smoothing

- Exponential smoothing is a rule of thumb technique for smoothing time series data using the exponential window function. Whereas in the simple moving average the past observations are weighted equally, exponential functions are used to assign exponentially decreasing weights over time. It is an easily learned and easily applied procedure for making some determination based on prior assumptions by the user, such as seasonality. Exponential smoothing is often used for analysis of time-series data. We have read the data head to check the values.
- Fig 13. shows the trend on the test and train data for SES.
- The RMSE for SES is 35 which is the lower than the Naïve model
- We have noted the value for all models above for a better comparison.

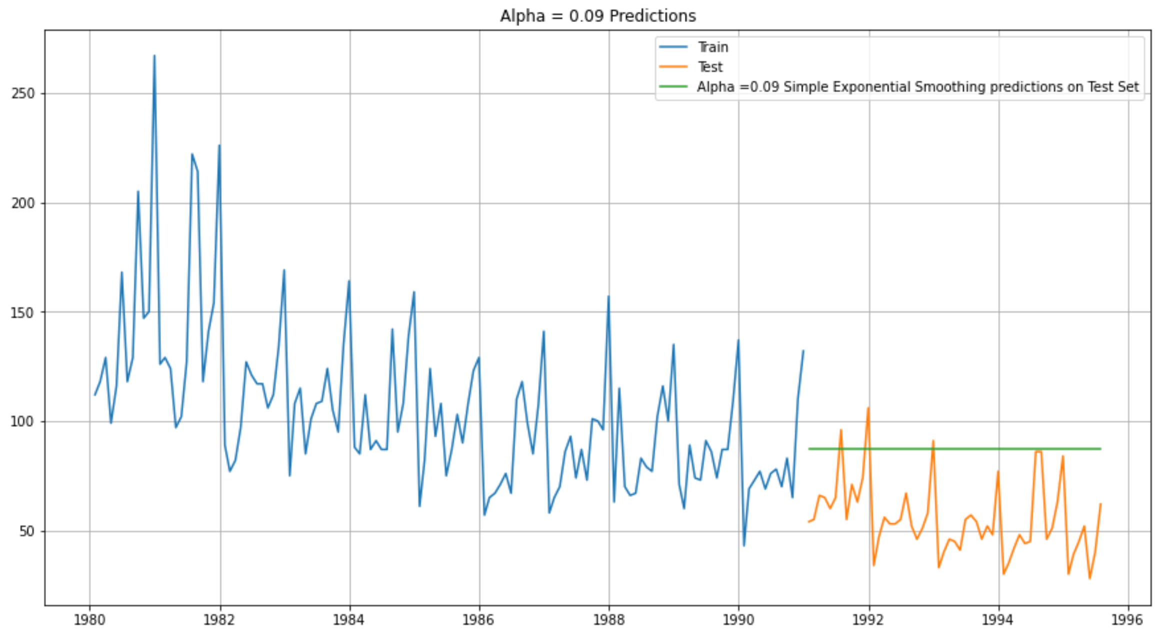
Fig 12. Plot for SES



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

```
{'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}
```

Plotting the Training data, Test data and the forecasted values



Mean Absolute Percentage Error (MAPE)

```
SES RMSE: 35.931353079283994
SES RMSE (calculated using statsmodels): 35.93135307928399
```

Test RMSE

	Test RMSE
Alpha=0.09,SES	35.931353

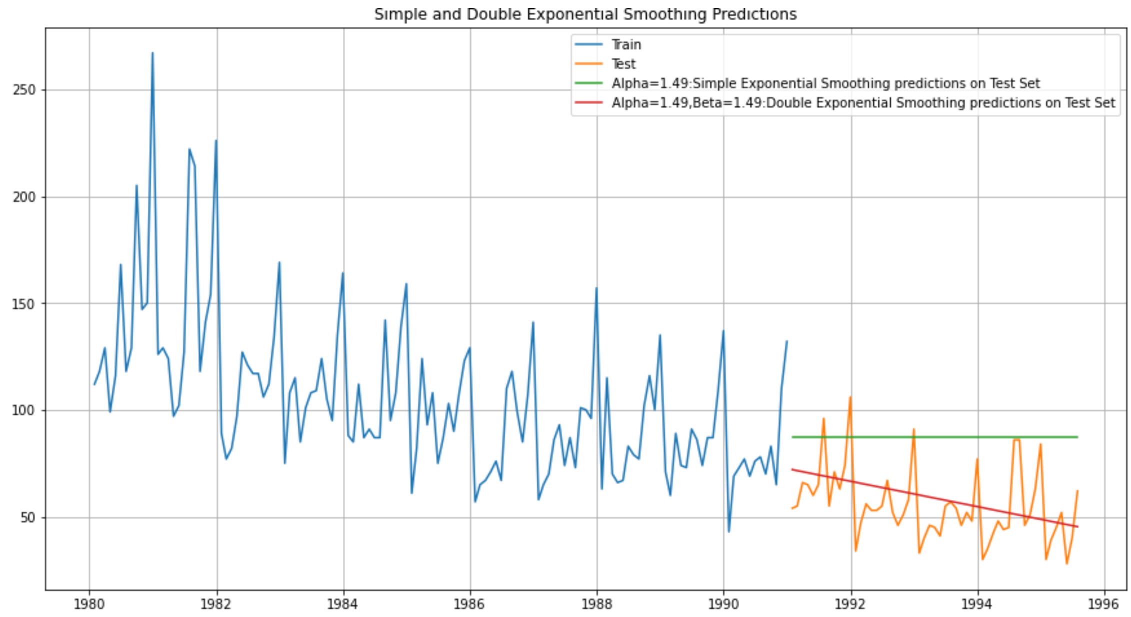
	Test RMSE
RegressionOnTime	16.626144
NaiveModel	78.485320
SimpleAverageModel	52.369847
2pointTrailingMovingAverage	12.158798
4pointTrailingMovingAverage	15.572375
6pointTrailingMovingAverage	15.687446
9pointTrailingMovingAverage	16.161176
Alpha=0.09,SES	35.931353

- 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.*

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

{'smoothing_level': 1.4901161193847656e-08, 'smoothing_trend': 1.4901161193847656e-08, 'smoothing_seasonal': nan, 'damping_
trend': nan, 'initial_level': 137.81550342222744, 'initial_trend': -0.4943776987426454, 'initial_seasons': array([], dtype=
float64), 'use_boxcox': False, 'lamda': None, 'remove_bias': False}
```

<u>Fig 13. Plot for Simple and Double Exponential Smoothing Predictions</u>



TEST RMSE:

```
DES RMSE: 16.62614518641044
```

- Fig 13. shows the trend on the test and train data for SES.
- The RMSE for DES is 16.62 which is higher than all the models.
- We have noted the value for all models above for a better comparison.

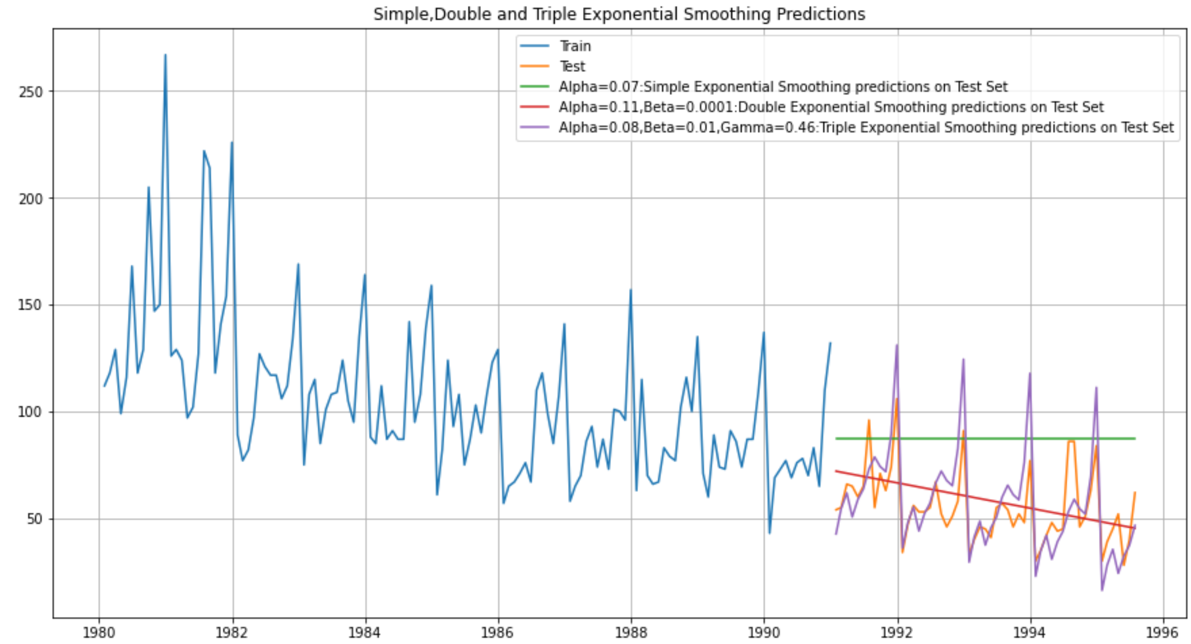
Holt-Winters - ETS(A, A, A) - Holt Winter's linear method with additive errors

Holt Winters model Exponential Smoothing Estimated Parameters:

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

{'smoothing_level': 0.08869629918171464, 'smoothing_trend': 0.0, 'smoothing_seasonal': 0.0, 'damping_trend': nan, 'initial_
level': 147.10516908592166, 'initial_trend': -0.5506122197227136, 'initial_seasons': array([-31.05717358, -18.70319747, -1
0.73802295, -21.4285425 ,
       -12.7031475 ,  -7.27605026,   2.65721611,   8.81731169,
         4.93599591,   3.06596287,  21.13082476,  63.40417384]), 'use_boxcox': False, 'lamda': None, 'remove_bias': False}
```

<u>Fig 14. Plot for Simple, Double and Triple Exponential Smoothing Predictions</u>



Test RMSE:

```
TES RMSE: 15.230029281691053
```

- We see that the Triple Exponential Smoothing is picking up the seasonal component as well.¶
- The RMSE for TES is 15.23 which is lowest RMSE when compared to all the models.
- We have noted the value for all models above for a better comparison.
- Fig 14 shows the graph for SES, DES and TES.

Table for all Models TEST RMSE:

	Test RMSE
RegressionOnTime	16.626144
NaiveModel	78.485320
SimpleAverageModel	52.369847
2pointTrailingMovingAverage	12.158798
4pointTrailingMovingAverage	15.572375
6pointTrailingMovingAverage	15.687446
9pointTrailingMovingAverage	16.161176
Alpha=0.09,SES	35.931353
Alpha=0.08,Beta=0.00:DES	16.626145
Alpha=0.11,Beta=0.01,Gamma=0.46:TES	15.230029

Inference

Triple Exponential Smoothing has performed the best on the test as expected since the data had both trend and seasonality.

But we see that our triple exponential smoothing is under forecasting. Let us try to tweak some of the parameters in order to get a better forecast on the test set.

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

ETS(A, A, M) model

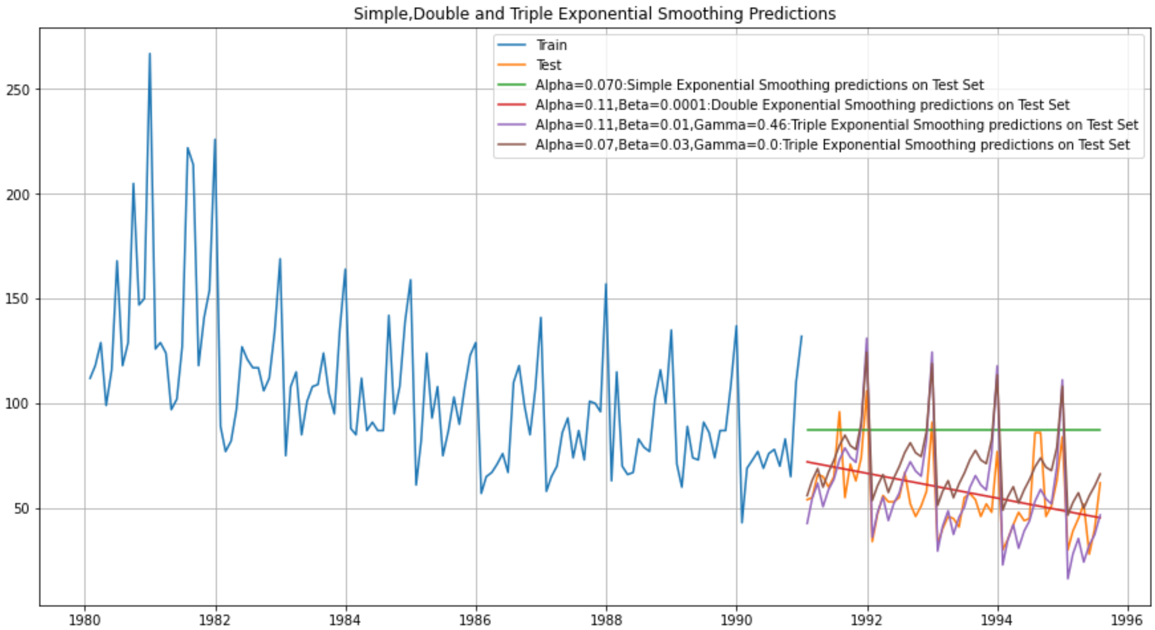
Table for smoothing level:

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

{'smoothing_level': 0.07774504348137964, 'smoothing_trend': 0.03906206880299537, 'smoothing_seasonal': 0.000324409492392027
5, 'damping_trend': nan, 'initial_level': 121.67618105631797, 'initial_trend': -0.7018504179529415, 'initial_seasons': arra
y([0.92092023, 1.04380587, 1.14154799, 0.99750231, 1.12221009,
       1.22075564, 1.34309679, 1.43113635, 1.35137029, 1.32464229,
       1.54437336, 2.12872752]), 'use_boxcox': False, 'lamda': None, 'remove_bias': False}
```

Fig 15. Plot for Simple, Double and Triple Exponential Smoothing Predictions

Plotting the Training data, Test data and the forecasted values



Report model accuracy

```
TES_am RMSE: 18.583117087942373
```

Table 14. Report model accuracy

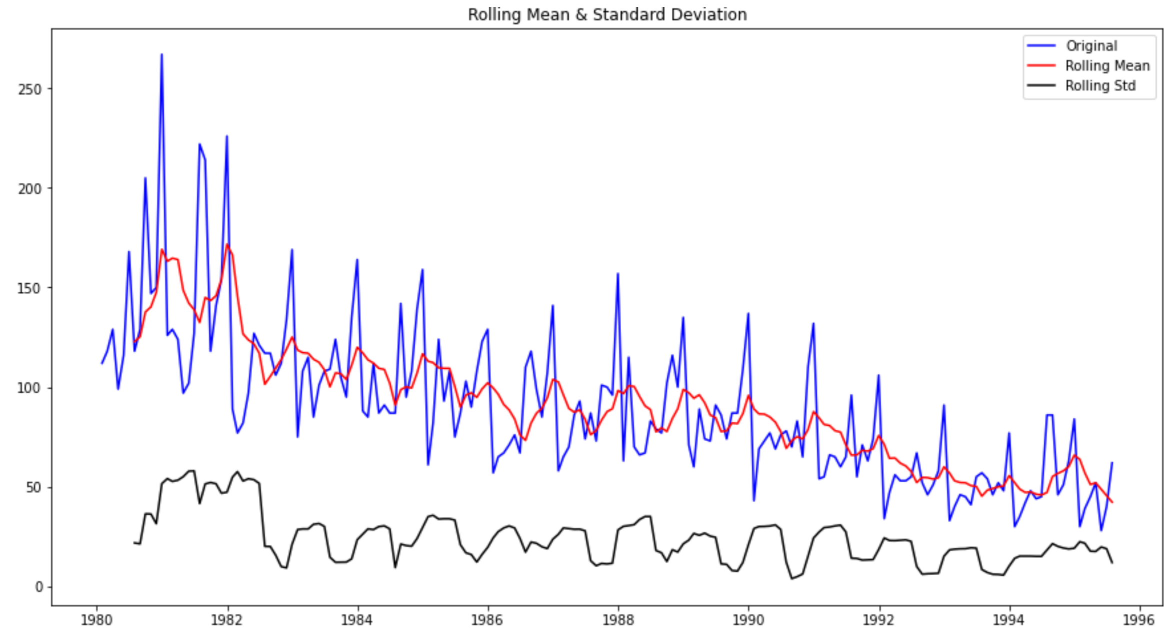
	Test RMSE
RegressionOnTime	16.626144
NaiveModel	78.485320
SimpleAverageModel	52.369847
2pointTrailingMovingAverage	12.158798
4pointTrailingMovingAverage	15.572375
6pointTrailingMovingAverage	15.687446
9pointTrailingMovingAverage	16.161176
Alpha=0.09,SES	35.931353
Alpha=0.08,Beta=0.00:DES	16.626145
Alpha=0.11,Beta=0.01,Gamma=0.46:TES	15.230029
Alpha=0.74,Beta=2.73e-06,Gamma=5.2e-07,Gamma=0:TES	18.583117

We see that the multiplicative seasonality model has not done that well when compared to the additive seasonality Triple Exponential Smoothing model.

- The model accuracy for TES_am RMSE is 18 which is higher than the TES.
- Fig 14 shows the graph for SES, DES and TES predictions.

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.

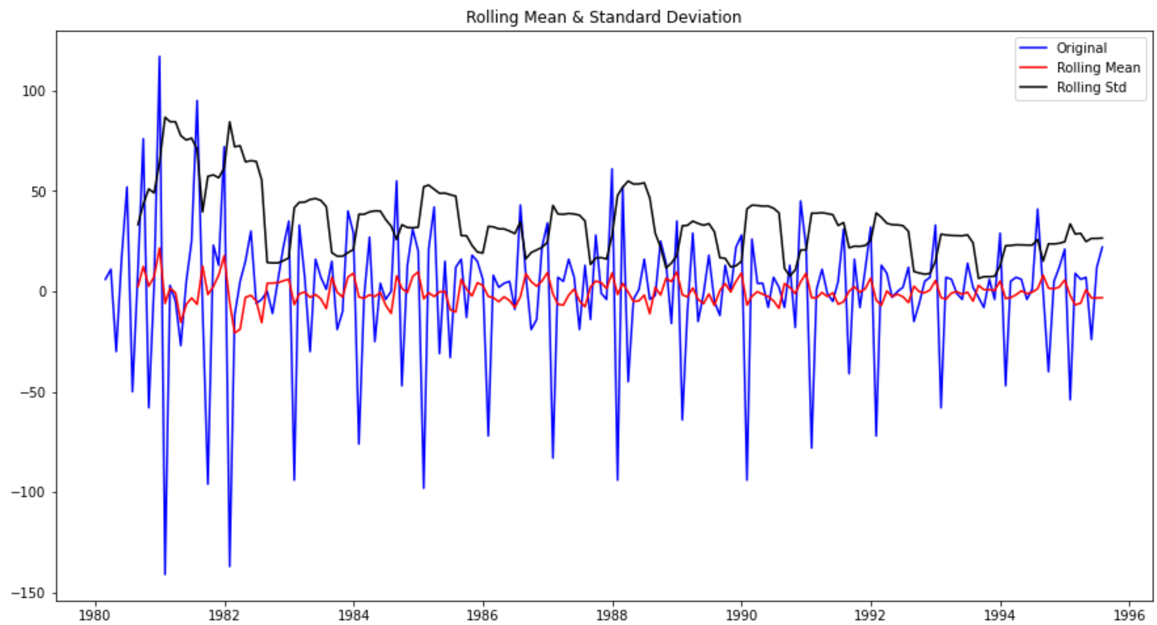
Fig g 15. Test for stationarity of the series - Dicky Fuller test




```
Results of Dickey-Fuller Test:
Test Statistic                  -1.933053
p-value                          0.316675
#Lags Used                      13.000000
Number of Observations Used    173.000000
Critical Value (1%)             -3.468726
Critical Value (5%)             -2.878396
Critical Value (10%)            -2.575756
dtype: float64
```

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

Fig 16. Plot for difference of Rolling mean and standard mean



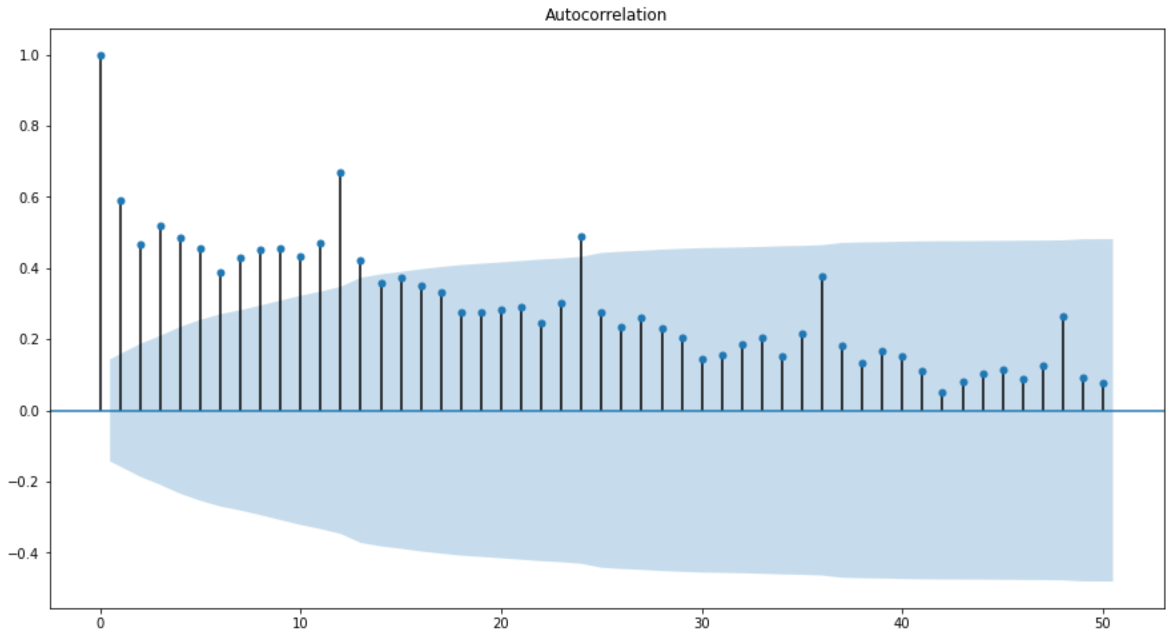

```
Results of Dickey-Fuller Test:
Test Statistic                 -7.890753e+00
p-value                         4.443288e-12
#Lags Used                      1.200000e+01
Number of Observations Used     1.730000e+02
Critical Value (1%)            -3.468726e+00
Critical Value (5%)            -2.878396e+00
Critical Value (10%)           -2.575756e+00
dtype: float64
```

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

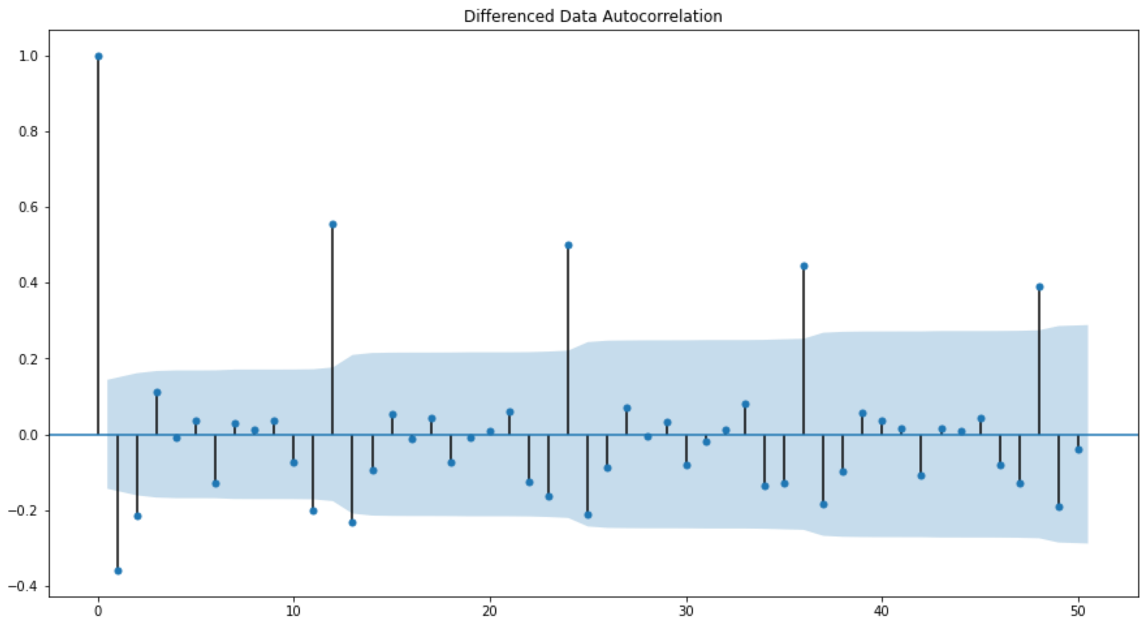
- We see that at α= 0.05 the Time Series is indeed stationary.
- Plot the Autocorrelation and the Partial Autocorrelation function plots on the whole data.

Fig 17. Auto correlation:

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



<u>Fig 18. Differenced data Autocorrelation:</u>



<u>Fig 19. Plot for Partial Autocorrelation</u>

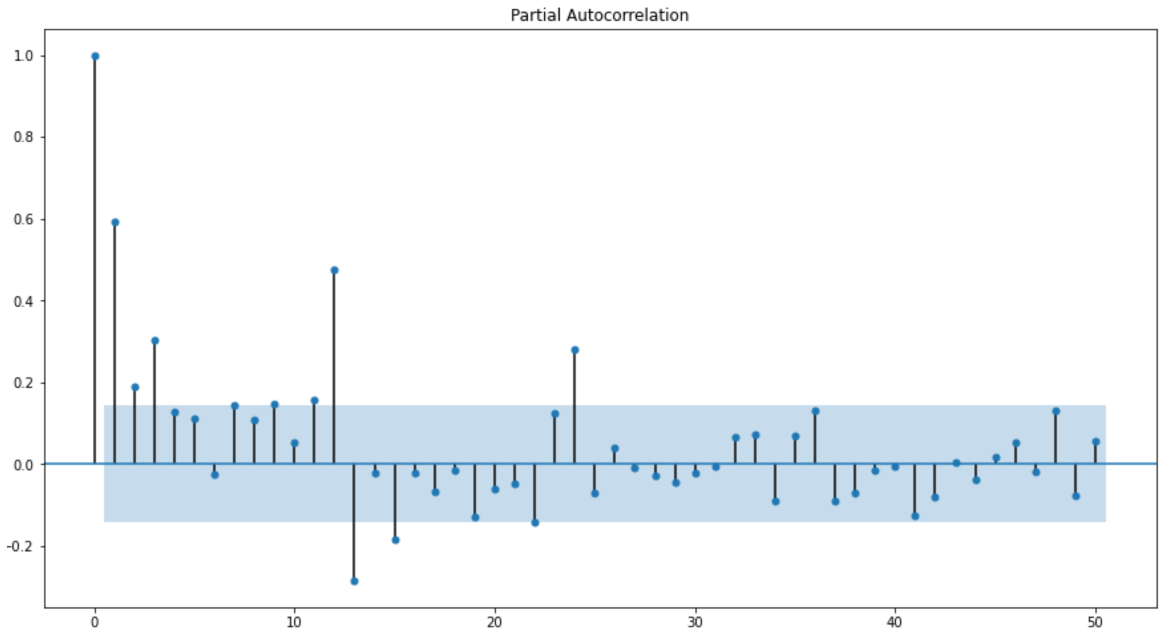
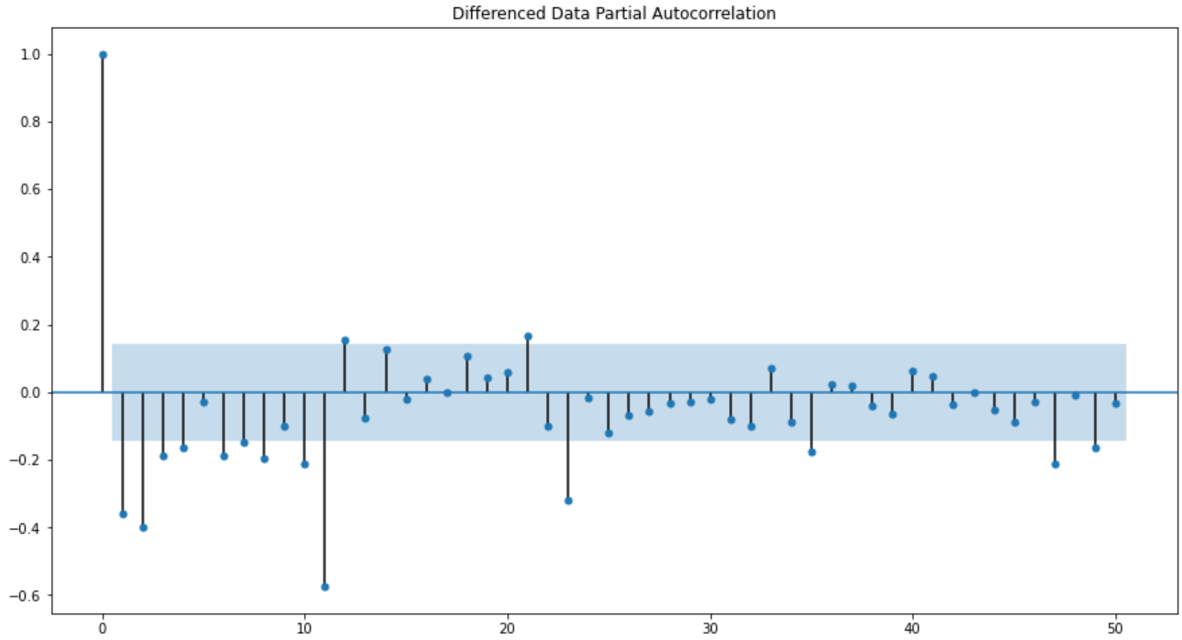


Fig 20. Differenced Data Partial Autocorrelation:



From the above plots, we can say that there seems to be no seasonality in the data.

- We have plotted different graphs for Rolling and difference mean, Autocorrelation and partial auto correlation.
- The graphs show no trend and seasonality.
- However, the correlation graphs show trend but no seasonality.

Table 14. First few and Last few rows of training and testing dataset:

```
First few rows of Training Da
```

Time_Stamp	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

```
Last few rows of Training Dat
```

Time_Stamp	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
```

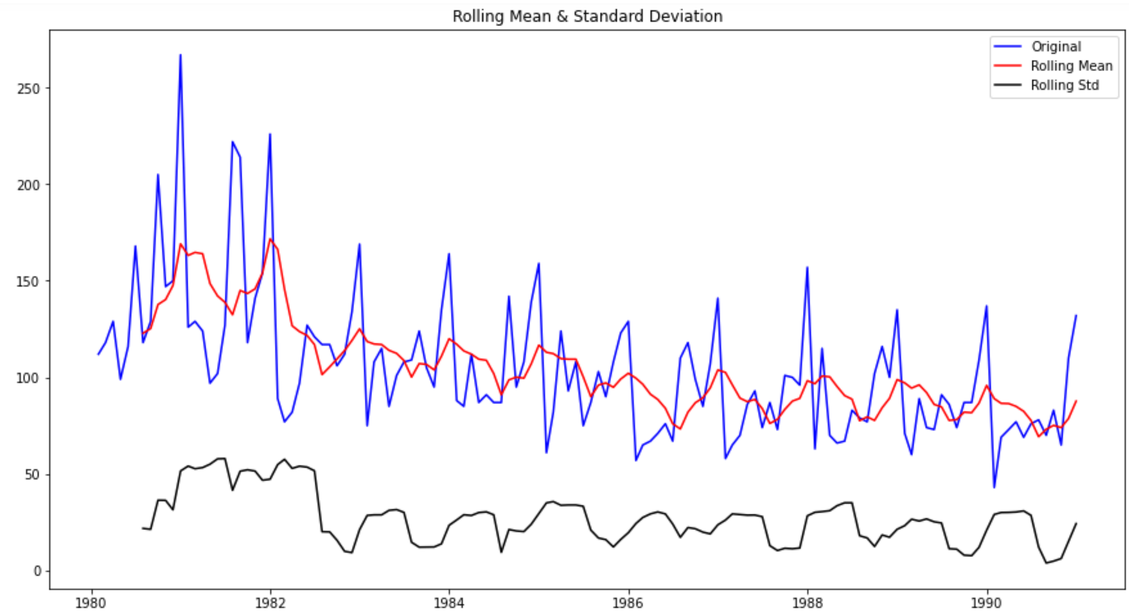
Time_Stamp	Rose
1991-01-31	54.0
1991-02-28	55.0
1991-03-31	66.0
1991-04-30	65.0
1991-05-31	60.0

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

Test and Train data set shape:

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

Fig 20. Plot for of Rolling mean and standard deviation

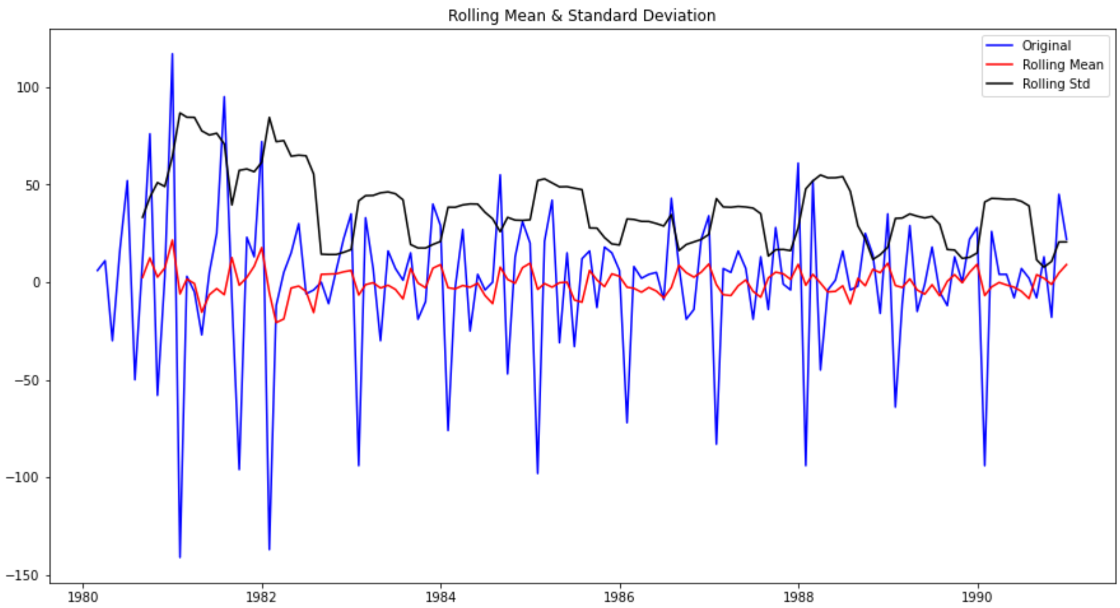



```
Results of Dickey-Fuller Test:
Test Statistic                  -2.164250
p-value                          0.219476
#Lags Used                      13.000000
Number of Observations Used    118.000000
Critical Value (1%)             -3.487022
Critical Value (5%)             -2.886363
Critical Value (10%)            -2.580009
dtype: float64
```

We see that the series is not stationary at α = 0.05

- We can see the rolling mean and standard deviation show trend in the data but no seasonality.

<u>Fig 21. Plot for difference of Rolling mean and standard deviation</u>




```
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
dtype: float64
```

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.

Answer:

- The following loop helps us in getting a combination of different parameters of p and q in the range of 0 and 2
- We have kept the value of d as 1 as we need to take a difference of the series to make it stationary.
- The highest AIC score is for 0 1324.
- The Akaike information criterion is an estimator of prediction error and thereby relative quality of statistical models for a given set of data. Given a collection of models for the data, AIC estimates the quality of each model, relative to each of the other models. Thus, AIC provides a means for model selection.

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

ARIMA Models

```
ARIMA(0, 0, 0) - AIC:1324.8997029577333
ARIMA(0, 0, 1) - AIC:1305.4684057684467
ARIMA(0, 0, 2) - AIC:1306.5866794772203
ARIMA(1, 0, 0) - AIC:1301.5463044353148
```

```
ARIMA(1, 0, 1) - AIC:1294.510585182307
ARIMA(1, 0, 2) - AIC:1292.0532102443954
ARIMA(2, 0, 0) - AIC:1302.3460741784133
ARIMA(2, 0, 1) - AIC:1292.937194561076
ARIMA(2, 0, 2) - AIC:1292.248055329439
```

ARMIA AIC:

	param	AIC
5	(1, 0, 2)	1292.053210
8	(2, 0, 2)	1292.248055
7	(2, 0, 1)	1292.937195
4	(1, 0, 1)	1294.510585
3	(1, 0, 0)	1301.546304
6	(2, 0, 0)	1302.346074
1	(0, 0, 1)	1305.468406
2	(0, 0, 2)	1306.586679
0	(0, 0, 0)	1324.899703

ARIMA Model Result:

```
                             ARIMA Model Results
==============================================================================
Dep. Variable:                 D.Rose   No. Observations:                  131
Model:                 ARIMA(2, 1, 1)   Log Likelihood                -634.523
Method:                       css-mle   S.D. of innovations             30.176
Date:                Tue, 06 Apr 2021   AIC                           1279.046
Time:                        12:34:10   BIC                           1293.422
Sample:                    02-29-1980   HQIC                          1284.887
                         - 12-31-1990
================================================================================
                   coef    std err          z      P>|z|      [0.025      0.975]
--------------------------------------------------------------------------------
const           -0.4915      0.080     -6.148      0.000      -0.648      -0.335
ar.L1.D.Rose     0.2127      0.088      2.409      0.016       0.040       0.386
ar.L2.D.Rose    -0.0759      0.089     -0.856      0.392      -0.250       0.098
ma.L1.D.Rose    -1.0000      0.044    -22.574      0.000      -1.087      -0.913
                                    Roots
=============================================================================
                  Real          Imaginary           Modulus         Frequency
-----------------------------------------------------------------------------
AR.1            1.4008           -3.3477j            3.6289           -0.1869
AR.2            1.4008           +3.3477j            3.6289            0.1869
MA.1            1.0000           +0.0000j            1.0000            0.0000
-----------------------------------------------------------------------------
```

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

```
16.788317607460872
```

RMSE:

	Test RMSE
ARIMA(2,1,1)	16.788318

Build a version of the ARIMA model for which the best parameters are selected by looking at the ACF and the PACF plots.

Fig 21. Differenced Data Autocorrelation ARIMA

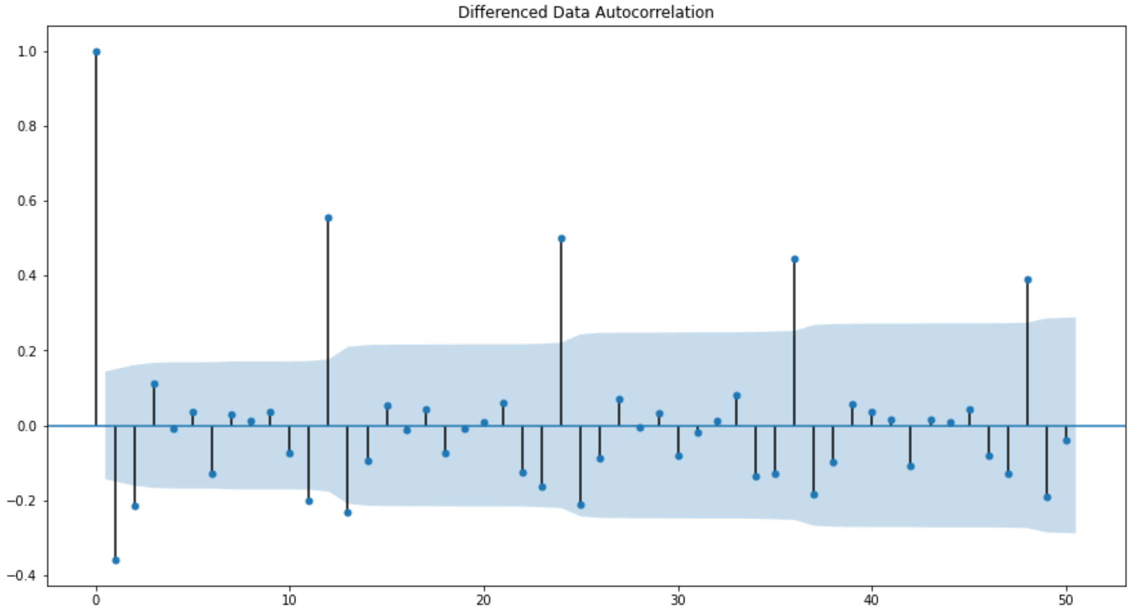
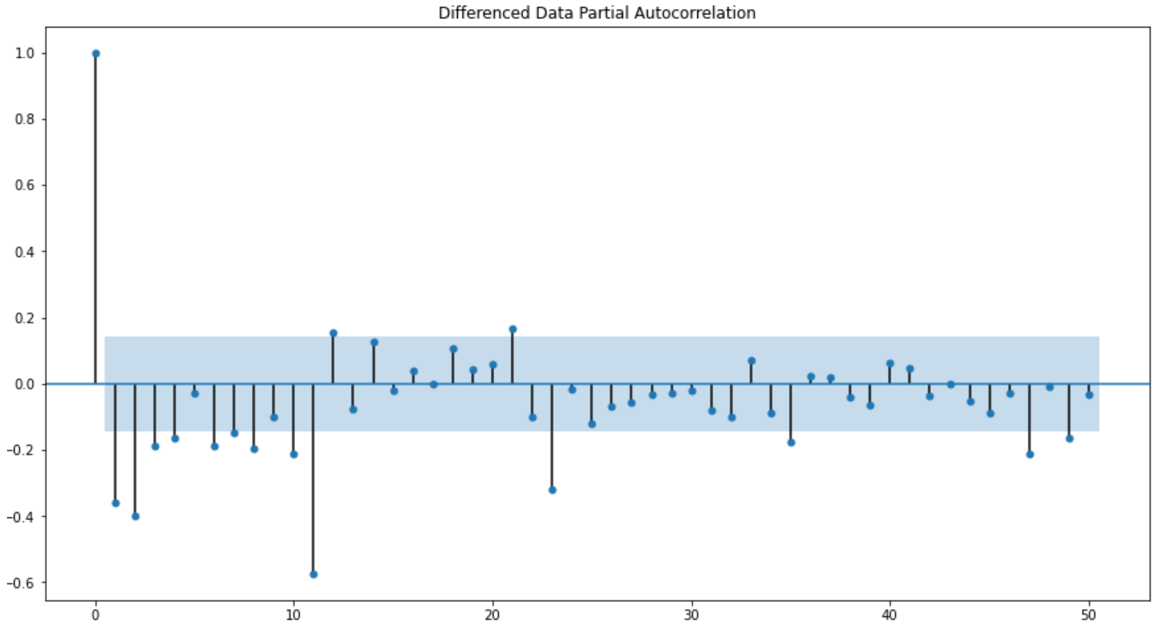


Fig 22. Differenced Data Partial Autocorrelation ARIMA



- Here, we have taken alpha=0.05.
- The Auto-Regressive parameter in an ARIMA model is 'p' which comes from the significant lag before which the PACF plot cuts-off to 0.
- The Moving-Average parameter in an ARIMA model is 'q' which comes from the significant lag before the ACF plot cuts-off to 0.
- By looking at the above plots, we can say that both the PACF and ACF plot cuts-off at lag 0.

- We can see the AIC of ARIMA models are in thousands which Is good however we need to make sure the values are consistent.4

- Hence we will be considering Manual ARIMA where the RMSE value is 4779 which is very high which shows the model is not stable.

ARIMA Model Result:

```
                             ARIMA Model Results
==============================================================================
Dep. Variable:                 D.Rose   No. Observations:                  131
Model:                 ARIMA(0, 1, 0)   Log Likelihood                -665.576
Method:                           css   S.D. of innovations             38.931
Date:                Tue, 06 Apr 2021   AIC                           1335.153
Time:                        12:34:11   BIC                           1340.903
Sample:                    02-29-1980   HQIC                          1337.489
                         - 12-31-1990
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.1527      3.401      0.045      0.964      -6.514       6.819
==============================================================================
```

TEST RMSE :

```
82.84724346992752
```

	Test RMSE
ARIMA(2,1,1)	16.788318
ARIMA(0,1,0)	82.847243

We see that there is difference in the RMSE values for both the models, but remember that the second model is a much simpler model.

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

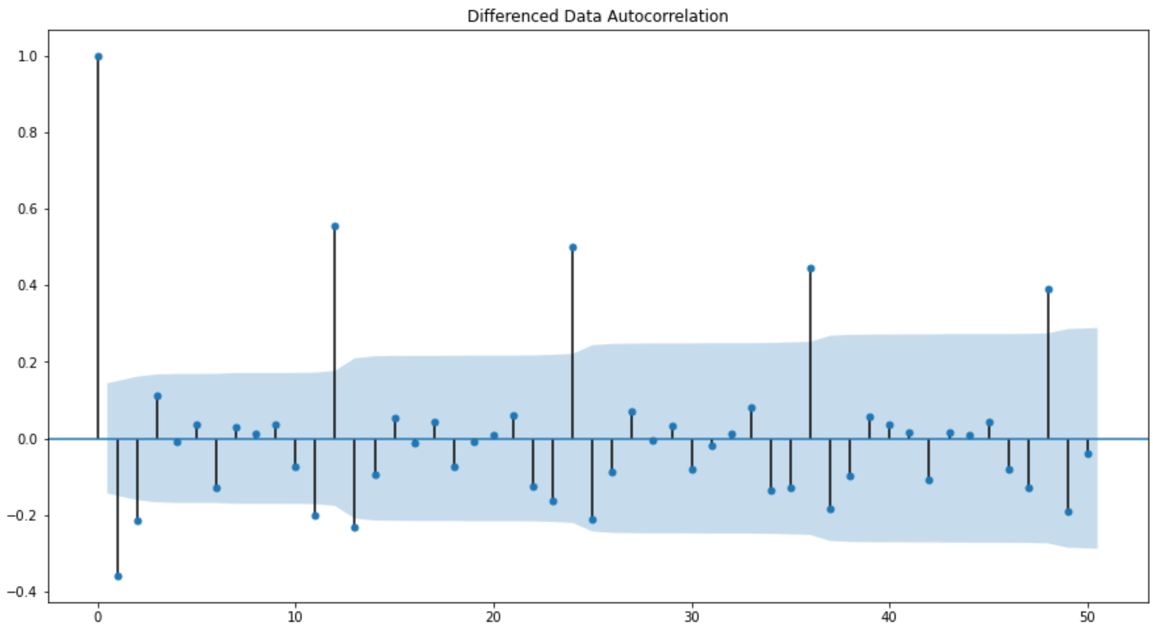
ACF plot

- We see that there is difference in the RMSE values for both the models, but remember that the second model is a much simpler model.

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

- Autoregressive integrated moving average In statistics and econometrics, and in particular in time series analysis, an autoregressive integrated moving average model is a generalization of an autoregressive moving average model. Both of these models are fitted to time series data either to better understand the data or to predict future points in the series.

<u>Fig 23. Differenced Data Partial Autocorrelation SARIMA</u>



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

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

SARIMA AIC:

	param	seasonal	AIC
26	(0, 1, 2)	(2, 0, 2, 12)	887.937509
80	(2, 1, 2)	(2, 0, 2, 12)	890.668848
69	(2, 1, 1)	(2, 0, 0, 12)	896.518161
78	(2, 1, 2)	(2, 0, 0, 12)	897.346498
70	(2, 1, 1)	(2, 0, 1, 12)	897.639957

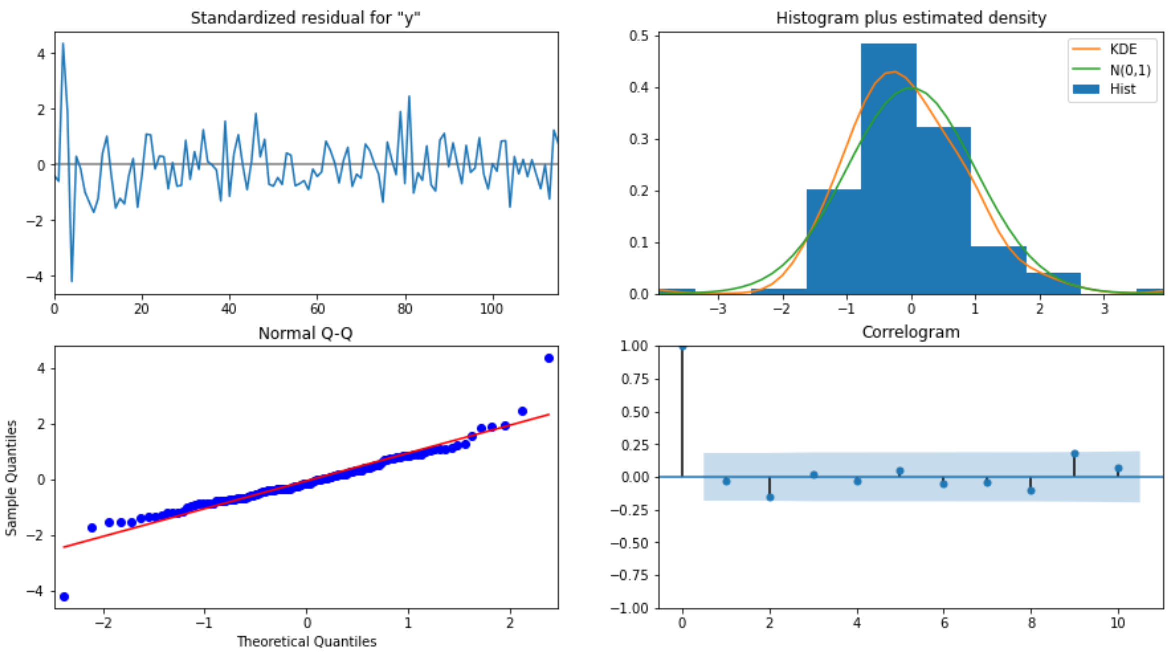
SARIMA Model Results:

```
                              SARIMAX Results
==========================================================================================
Dep. Variable:                              y   No. Observations:                  132
Model:             SARIMAX(0, 1, 2)x(2, 0, 2, 6)   Log Likelihood                -514.800
Date:                        Tue, 06 Apr 2021   AIC                           1043.600
Time:                                12:34:35   BIC                           1062.875
Sample:                                     0   HQIC                          1051.425
                                        - 132
Covariance Type:                          opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ma.L1         -0.7884    781.883     -0.001      0.999   -1533.252    1531.675
ma.L2         -0.2116    165.510     -0.001      0.999    -324.605     324.181
ar.S.L6       -0.0727      0.037     -1.991      0.047      -0.144      -0.001
ar.S.L12       0.8368      0.042     19.890      0.000       0.754       0.919
ma.S.L6        0.2238    781.873      0.000      1.000   -1532.220    1532.667
ma.S.L12      -0.7762    606.912     -0.001      0.999   -1190.302    1188.750
sigma2       347.5239      2.872    120.995      0.000     341.894     353.153
===================================================================================
Ljung-Box (L1) (Q):                   0.14   Jarque-Bera (JB):                90.77
Prob(Q):                              0.71   Prob(JB):                         0.00
Heteroskedasticity (H):               0.42   Skew:                             0.37
Prob(H) (two-sided):                  0.01   Kurtosis:                         7.27
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
[2] Covariance matrix is singular or near-singular, with condition number 8.43e+20. Standard errors may be unstable.
```

Plot to results auto SARIMA

Fig 24. Different graphs for standard residual for Y



From the model diagnostics plot, we can see that all the individual diagnostics plots almost follow the theoretical numbers and thus we cannot develop any pattern from these plots.

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

y	mean	mean_se	mean_ci_lower	mean_ci_upper
0	69.066606	19.191877	31.451218	106.681994
1	67.813722	19.662985	29.274980	106.352465
2	76.132471	19.654377	37.610600	114.654342
3	71.773849	19.654375	33.251982	110.295716
4	76.562527	19.654374	38.040661	115.084392

RMSE

```
26.446047882541137
```

TEST RMSE

	Test RMSE
ARIMA(2,1,1)	16.788318
ARIMA(0,1,0)	82.847243
SARIMA(0,1,2)(2,0,2,6)	26.446048

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

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

SARIMA AIC

```
SARIMA(0, 0, 0)x(0, 1, 0, 6) - AIC:1320.0985789105328
SARIMA(0, 0, 0)x(0, 1, 1, 6) - AIC:1166.6525964306707
SARIMA(0, 0, 0)x(0, 1, 2, 6) - AIC:1069.740706856482
SARIMA(0, 0, 0)x(1, 1, 0, 6) - AIC:1128.961846229569
SARIMA(0, 0, 0)x(1, 1, 1, 6) - AIC:1122.145339069296
```

SARIMA AIC Short values:

	param	seasonal	AIC
50	(1, 0, 2)	(1, 1, 2, 6)	961.730761
53	(1, 0, 2)	(2, 1, 2, 6)	963.506578
77	(2, 0, 2)	(1, 1, 2, 6)	965.172463
80	(2, 0, 2)	(2, 1, 2, 6)	966.549892
23	(0, 0, 2)	(1, 1, 2, 6)	978.778467

SARIMA Results:

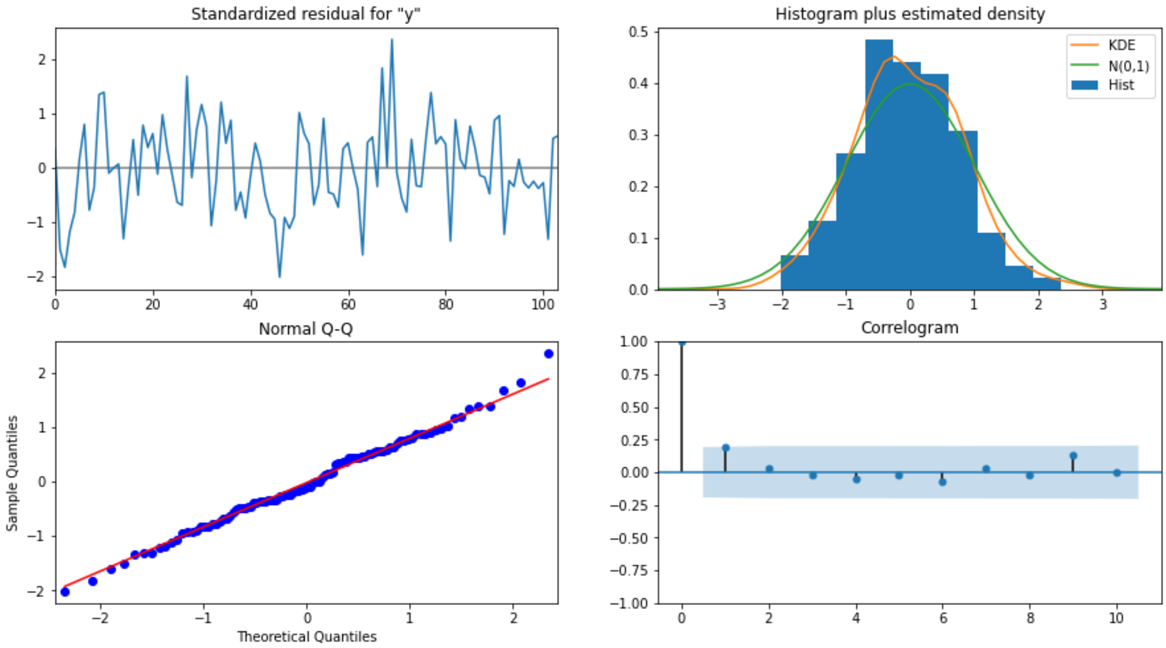
```
                                     SARIMAX Results
==========================================================================================
Dep. Variable:                                 y   No. Observations:                  132
Model:             SARIMAX(1, 1, 2)x(2, 0, 2, 12)   Log Likelihood                -444.580
Date:                           Tue, 06 Apr 2021   AIC                            905.160
Time:                                   12:35:11   BIC                            926.315
Sample:                                        0   HQIC                           913.731
                                           - 132
Covariance Type:                             opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.8945      0.091      9.778      0.000       0.715       1.074
ma.L1         -1.9657      0.258     -7.611      0.000      -2.472      -1.459
ma.L2          0.9669      0.245      3.941      0.000       0.486       1.448
ar.S.L12       0.3722      0.087      4.256      0.000       0.201       0.544
ar.S.L24       0.3447      0.087      3.946      0.000       0.174       0.516
ma.S.L12       0.0489      0.179      0.273      0.785      -0.302       0.400
ma.S.L24      -0.1588      0.226     -0.702      0.482      -0.602       0.284
sigma2       380.7292    136.412      2.791      0.005     113.366     648.092
===================================================================================
Ljung-Box (L1) (Q):                   4.02   Jarque-Bera (JB):                 0.21
Prob(Q):                              0.04   Prob(JB):                         0.90
Heteroskedasticity (H):               0.75   Skew:                             0.11
Prob(H) (two-sided):                  0.41   Kurtosis:                         2.97
===================================================================================

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

Auto SARIMA:

- We can see the AIC for SARIMA model is also performing in 1656 and increasing with the Param and seasonal.
- If the predictors consist only of lagged values of Y, it is a pure autoregressive ("self-regressed") model, which is just a special case of a regression model and which could be fitted with standard regression software.
- For example, a first-order autoregressive ("AR(1)") model for Y is a simple regression model in which the independent variable is just Y lagged by one period (LAG(Y,1) in Stat graphics or Y_LAG1 in Regress
- If some of the predictors are lags of the errors, an ARIMA model it is NOT a linear regression model, because there is no way to specify "last period's error" as an independent variable: the errors must be computed on a period-to-period basis when the model is fitted to the data.

Fig 25. Different graphs for standard residual for Y



Similar to the last iteration of the model where the seasonality parameter was taken as 5, here also we see that the model diagnostics plot does not indicate any remaining information that we can get.

Predicted Auto SARIMA:

y	mean	mean_se	mean_ci_lower	mean_ci_upper
0	60.797474	19.534099	22.511343	99.083605
1	73.172941	19.579108	34.798595	111.547287
2	77.205566	19.613563	38.763690	115.647442
3	76.336856	19.639818	37.843520	114.830193
4	73.045183	19.659694	34.512891	111.577474

RMSE:

```
26.391283583414577
```

TEST RMSE:

	Test RMSE
ARIMA(2,1,1)	16.788318
ARIMA(0,1,0)	82.847243
SARIMA(0,1,2)(2,0,2,6)	26.446048
SARIMA(1,1,2)(2,0,2,6)	26.391284

Fig 26. Plot for Differenced Data Autocorrelation

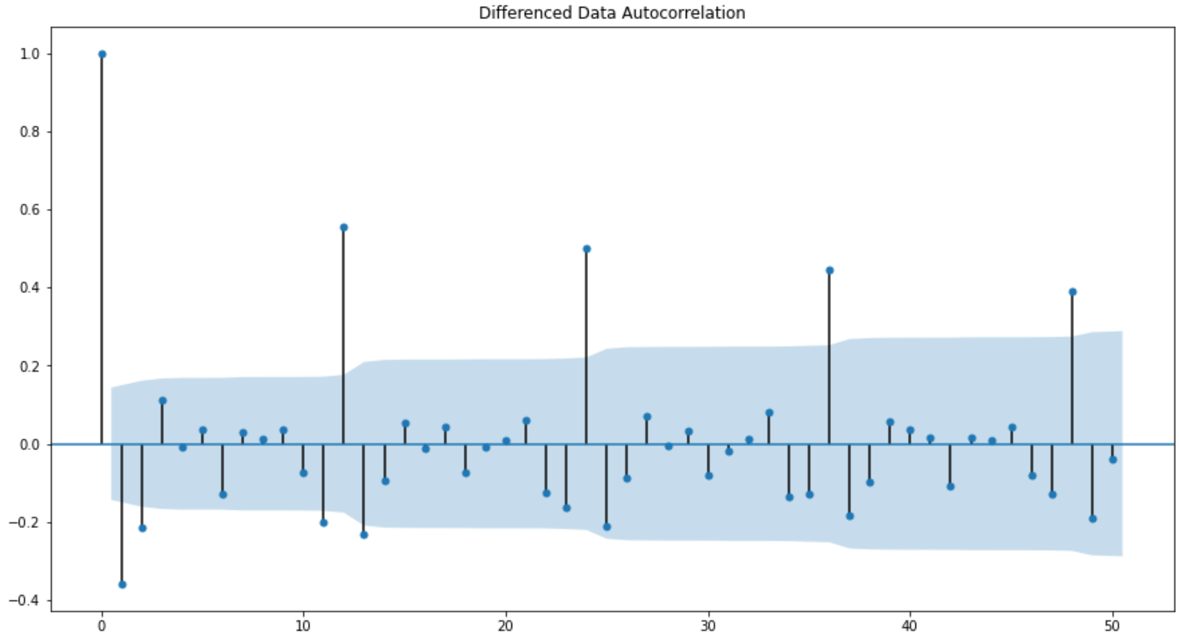


Fig 27. Plot for Differenced Data Partial Autocorrelation

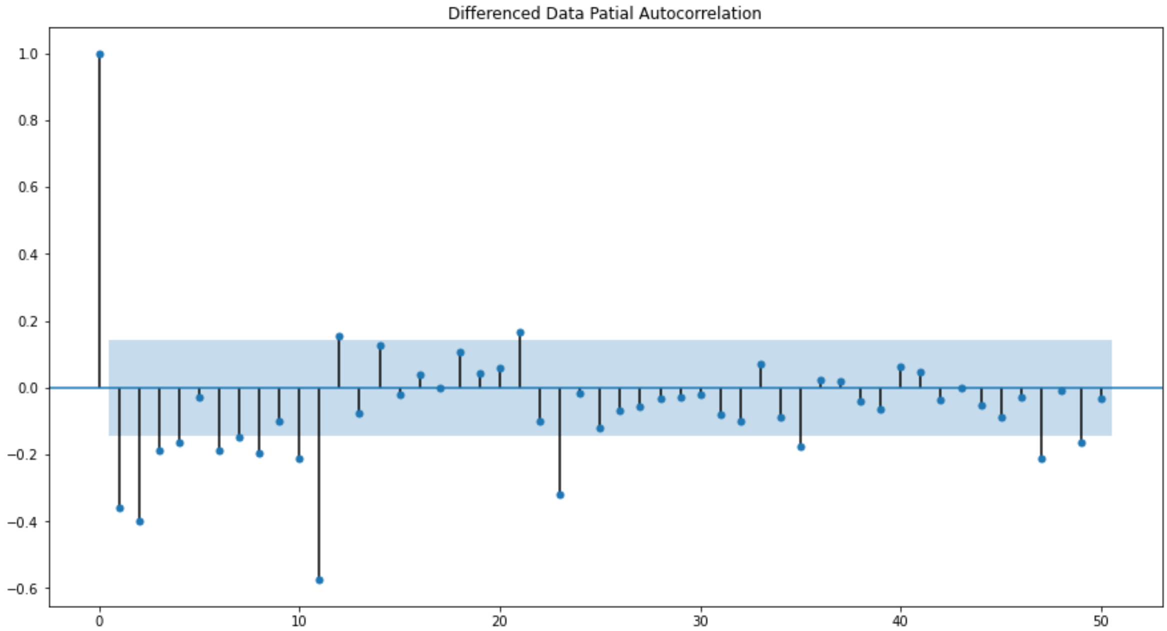
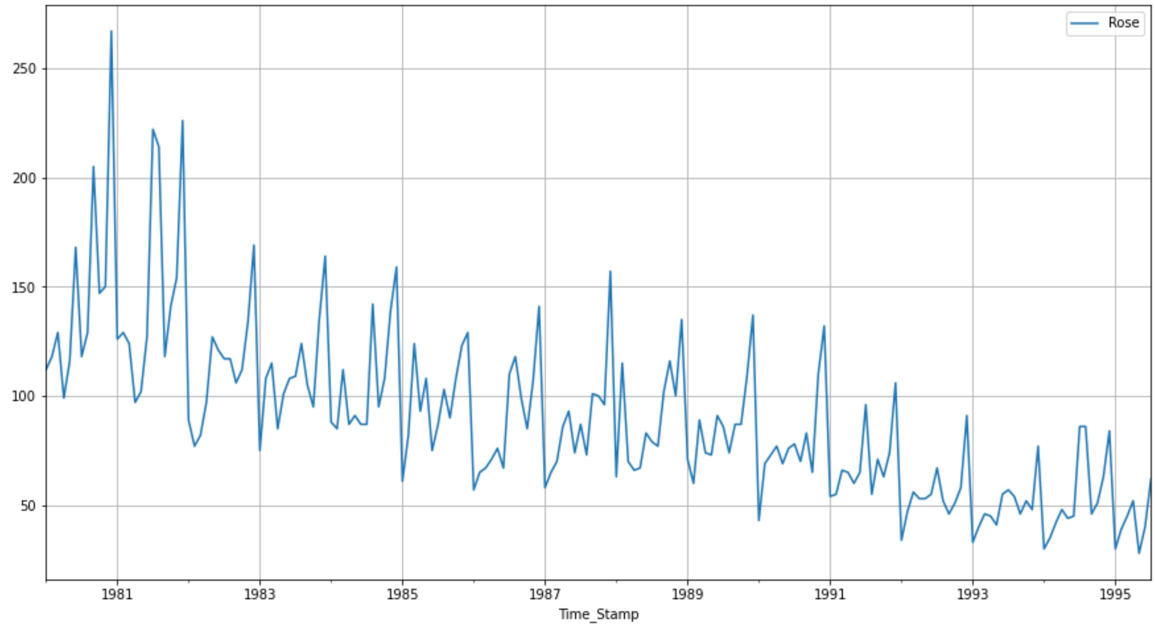
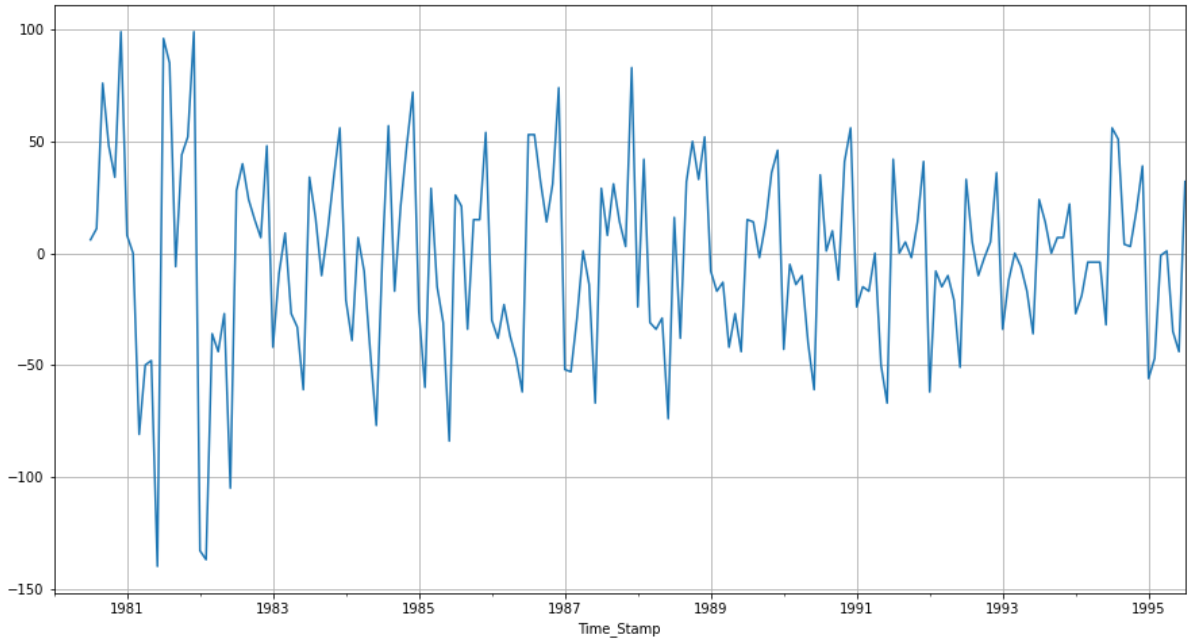


Fig 26. Plot for Test stamp and Time stamp



We see that there is a trend and a seasonality. So, now we take a seasonal differencing and check the series.



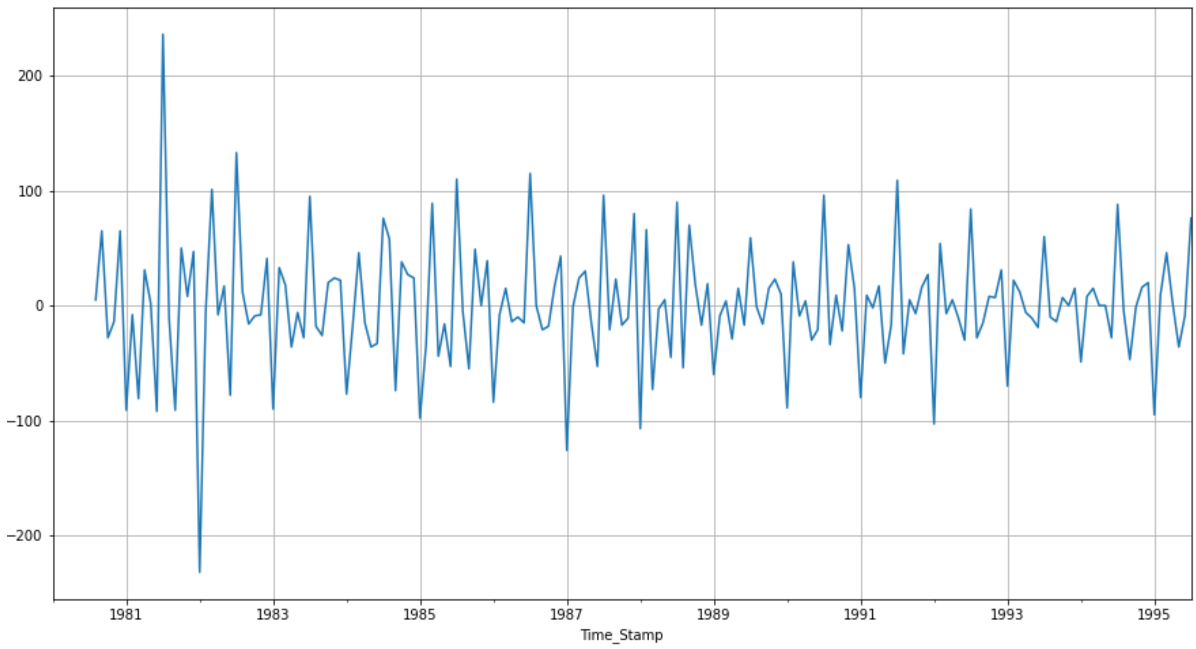
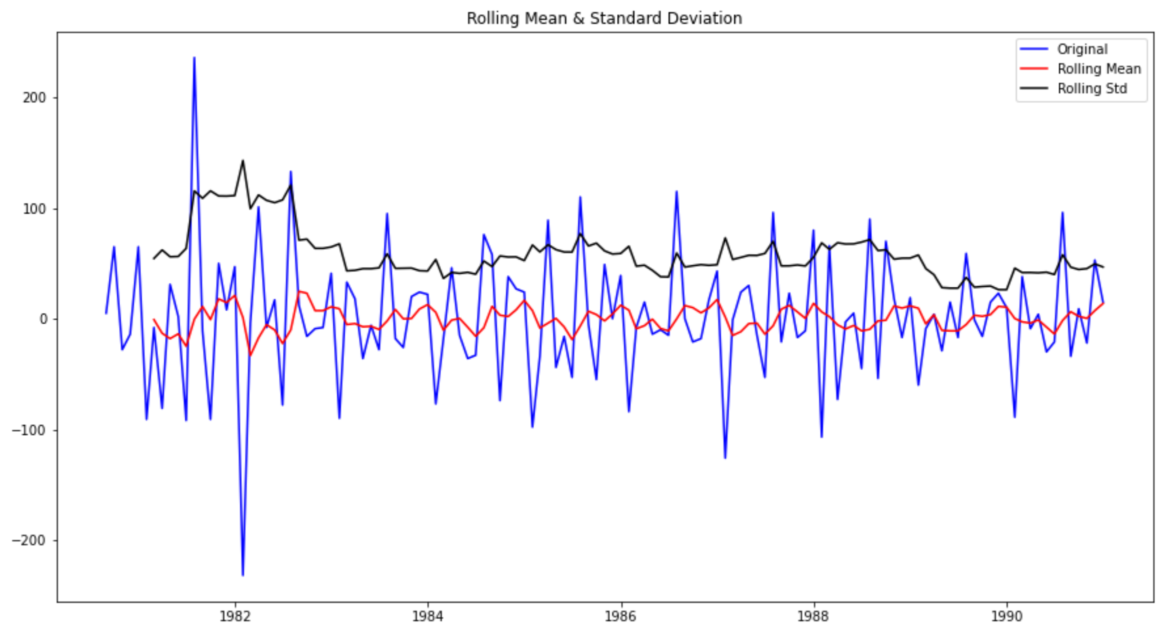


Fig 28. Rolling Mean & Standard Deviation




```
Results of Dickey-Fuller Test:
Test Statistic                -6.882869e+00
p-value                        1.418693e-09
#Lags Used                     1.300000e+01
Number of Observations Used    1.110000e+02
Critical Value (1%)           -3.490683e+00
Critical Value (5%)           -2.887952e+00
Critical Value (10%)          -2.580857e+00
dtype: float64
```

Fig 29. Plot for Autocorrelation

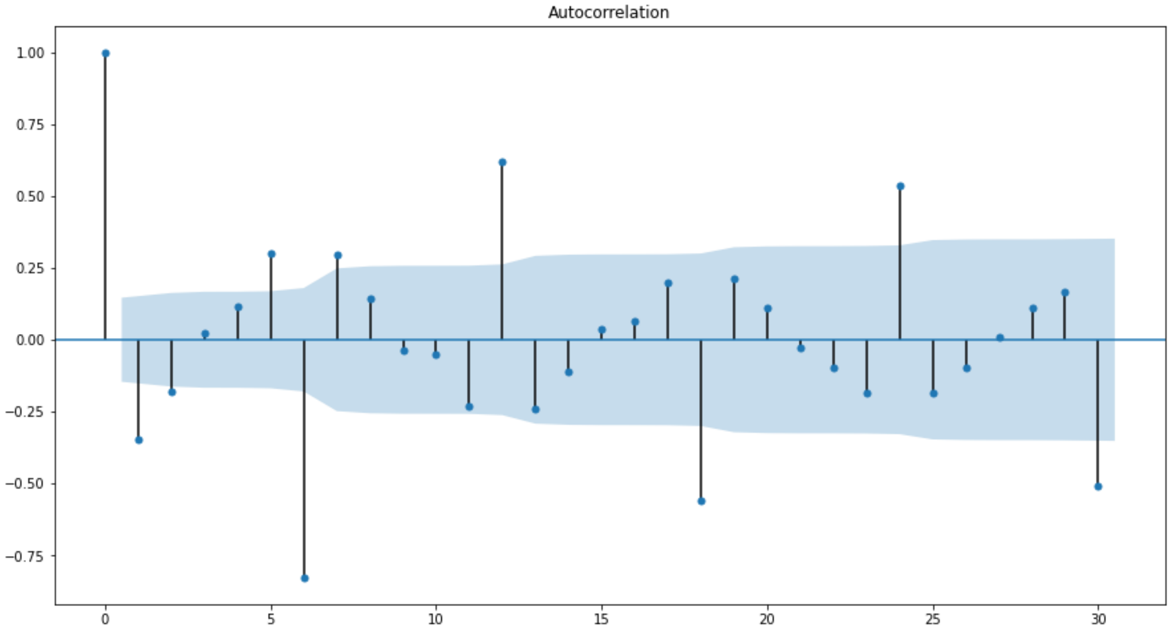
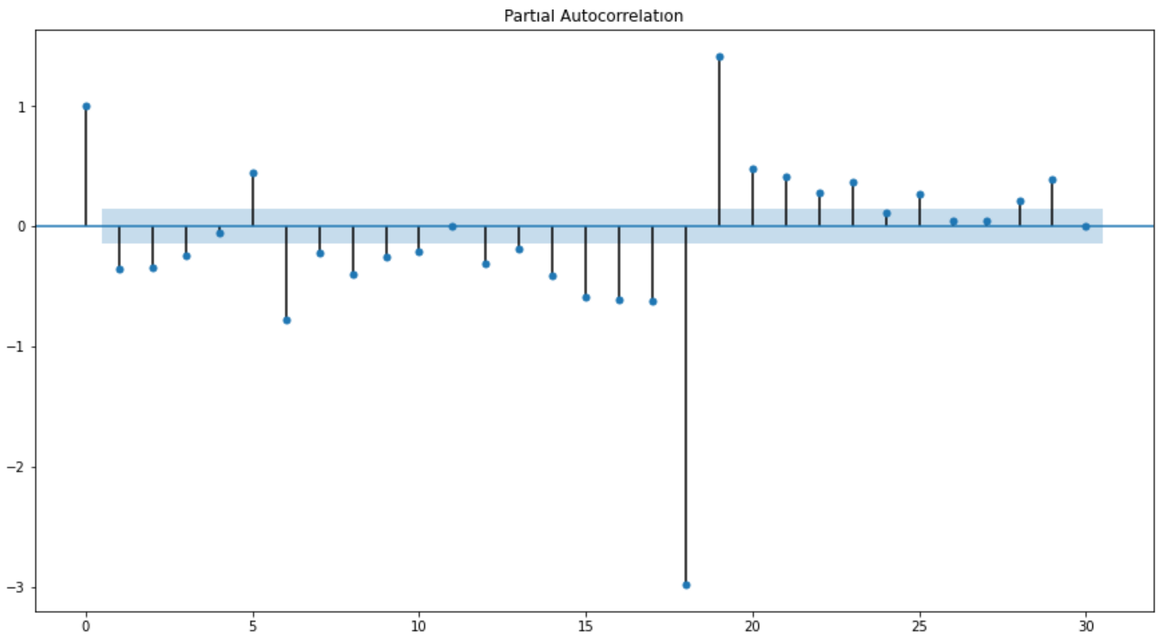


Fig 30. Plot for Partial Autocorrelation:

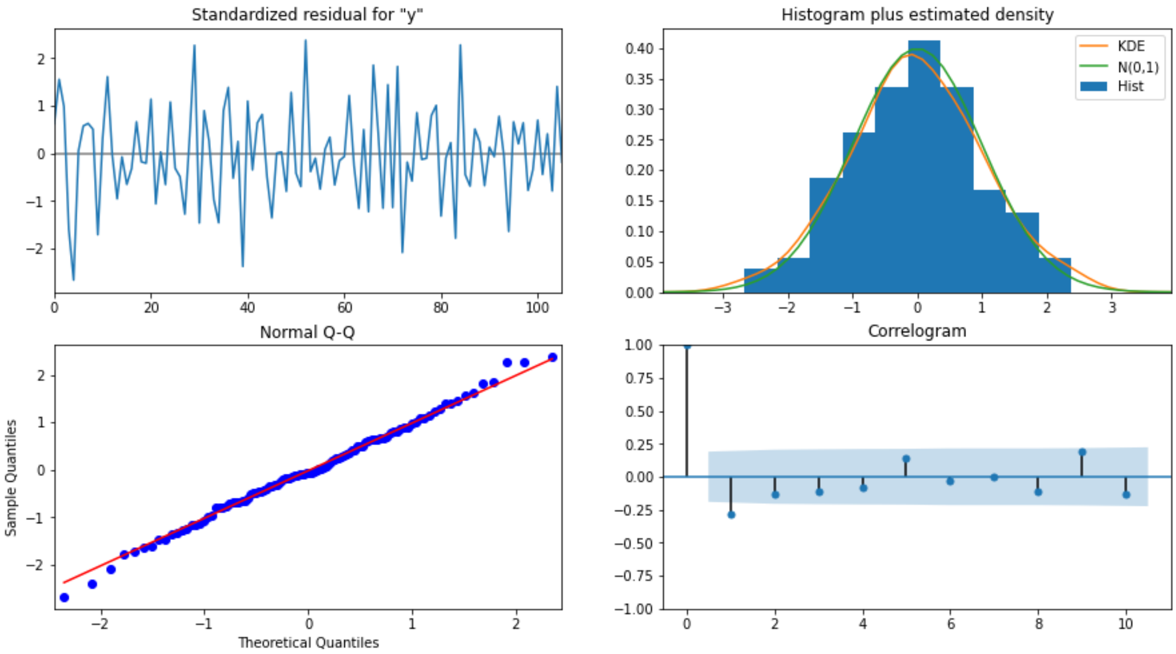


SARIMAX Model Results

```
                                      SARIMAX Results
==========================================================================================
Dep. Variable:                                  y   No. Observations:                  132
Model:             SARIMAX(0, 1, 0)x(1, 1, [1, 2, 3], 6)   Log Likelihood            -478.459
Date:                            Tue, 06 Apr 2021   AIC                            966.918
Time:                                    12:35:13   BIC                            980.235
Sample:                                         0   HQIC                           972.315
                                            - 132
Covariance Type:                              opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.S.L6       -0.8506      0.039    -22.079      0.000      -0.926      -0.775
ma.S.L6       -0.2404      0.119     -2.025      0.043      -0.473      -0.008
ma.S.L12      -0.5019      0.127     -3.961      0.000      -0.750      -0.254
ma.S.L18      -0.1041      0.104     -0.998      0.318      -0.308       0.100
sigma2       464.4148     69.417      6.690      0.000     328.360     600.469
===================================================================================
Ljung-Box (L1) (Q):                   8.65   Jarque-Bera (JB):                 0.00
Prob(Q):                              0.00   Prob(JB):                         1.00
Heteroskedasticity (H):               0.77   Skew:                            -0.01
Prob(H) (two-sided):                  0.45   Kurtosis:                         2.97
===================================================================================

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

Fig 40. Standard Results Manual SARIMA



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

y	mean	mean_se	mean_ci_lower	mean_ci_upper
0	65.934366	21.600170	23.598811	108.269921
1	83.523763	30.541319	23.663877	143.383649
2	84.349397	37.402901	11.041057	157.657736
3	82.359645	43.187751	-2.286792	167.006083
4	81.766414	48.284436	-12.869341	176.402168

RMSE:

```
37.25211539240966
```

TEST RMSE:

	Test RMSE
ARIMA(2,1,1)	16.788318
ARIMA(0,1,0)	82.847243
SARIMA(0,1,2)(2,0,2,6)	26.446048
SARIMA(1,1,2)(2,0,2,6)	26.391284
SARIMA(0,1,0)(1,1,3,6)	37.252115

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.

Answer:

Test RMSE

	Test RMSE
RegressionOnTime	16.626144
NaiveModel	78.485320
SimpleAverageModel	52.369847
2pointTrailingMovingAverage	12.158798
4pointTrailingMovingAverage	15.572375
6pointTrailingMovingAverage	15.687446
9pointTrailingMovingAverage	16.161176
Alpha=0.09,SES	35.931353
Alpha=0.08,Beta=0.00:DES	16.626145
Alpha=0.11,Beta=0.01,Gamma=0.46:TES	15.230029
Alpha=0.74,Beta=2.73e-06,Gamma=5.2e-07,Gamma=0:TES	18.583117

ARIMA(2,1,1)	16.788318
SARIMA(1,1,2)(2,0,2,6)	26.391284
SARIMA(0,1,2)(2,0,2,6)	26.446048
SARIMA(0,1,0)(1,1,3,6)	37.252115
ARIMA(0,1,0)	82.847243

- We have plotted a model for comparasion the above table shows the best performing model is 2poing Trailling Moving average at 12.15.
- There are few models which are performing at different levels.
- We can see the RMSE score for the Rose data set is giving output within 100 hence the data set is stable.
- The model with highest root mean square is 82/ 84 which is ARIMA at seasonality 0.
- Which concludes the data set doesn't have seasonality, but trend can be seen in few graphs above.

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.

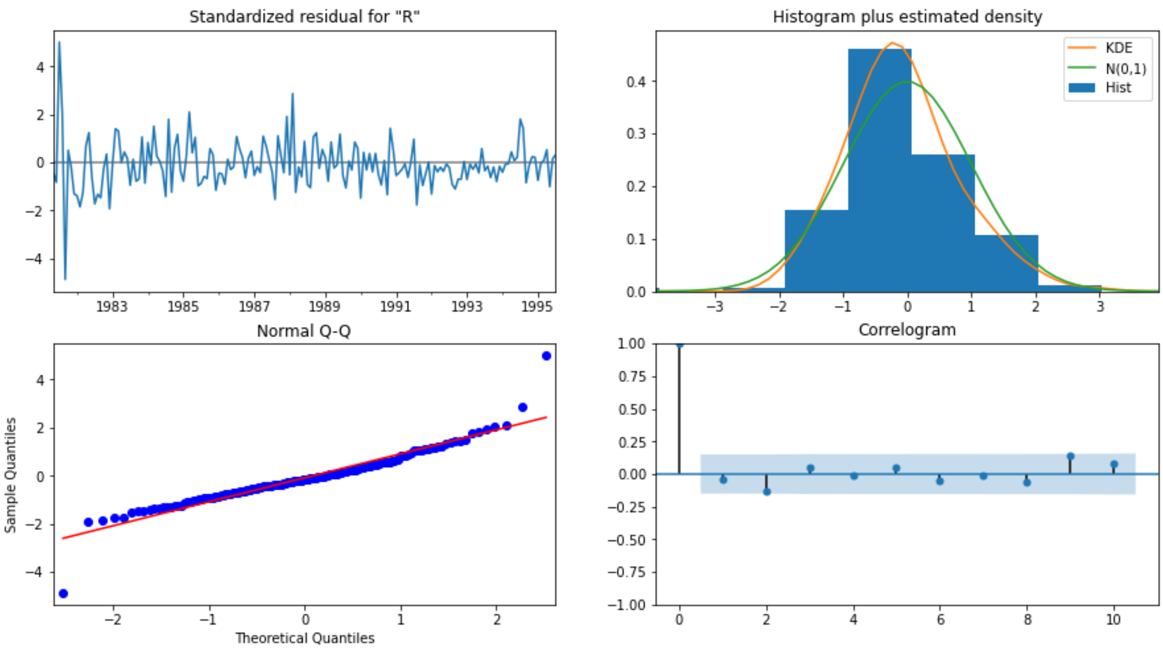
Answer:

SARIMAX Results

```
                                     SARIMAX Results
==========================================================================================
Dep. Variable:                             Rose   No. Observations:                   187
Model:             SARIMAX(0, 1, 2)x(2, 0, 2, 6)   Log Likelihood                -737.093
Date:                          Tue, 06 Apr 2021   AIC                           1488.186
Time:                                  12:35:14   BIC                           1510.178
Sample:                              01-31-1980   HQIC                          1497.109
                                   - 07-31-1995
Covariance Type:                            opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ma.L1         -0.7061      0.075     -9.449      0.000      -0.853      -0.560
ma.L2         -0.2183      0.069     -3.157      0.002      -0.354      -0.083
ar.S.L6       -0.0545      0.029     -1.902      0.057      -0.111       0.002
ar.S.L12       0.8787      0.030     29.327      0.000       0.820       0.937
ma.S.L6        0.2130      0.219      0.974      0.330      -0.215       0.641
ma.S.L12      -0.8291      0.191     -4.342      0.000      -1.203      -0.455
sigma2       279.5702     66.058      4.232      0.000     150.099     409.042
===================================================================================
Ljung-Box (L1) (Q):                   0.23   Jarque-Bera (JB):               248.21
Prob(Q):                              0.63   Prob(JB):                         0.00
Heteroskedasticity (H):               0.23   Skew:                             0.44
Prob(H) (two-sided):                  0.00   Kurtosis:                         8.84
===================================================================================

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

Standard Results Auto SARIMA:



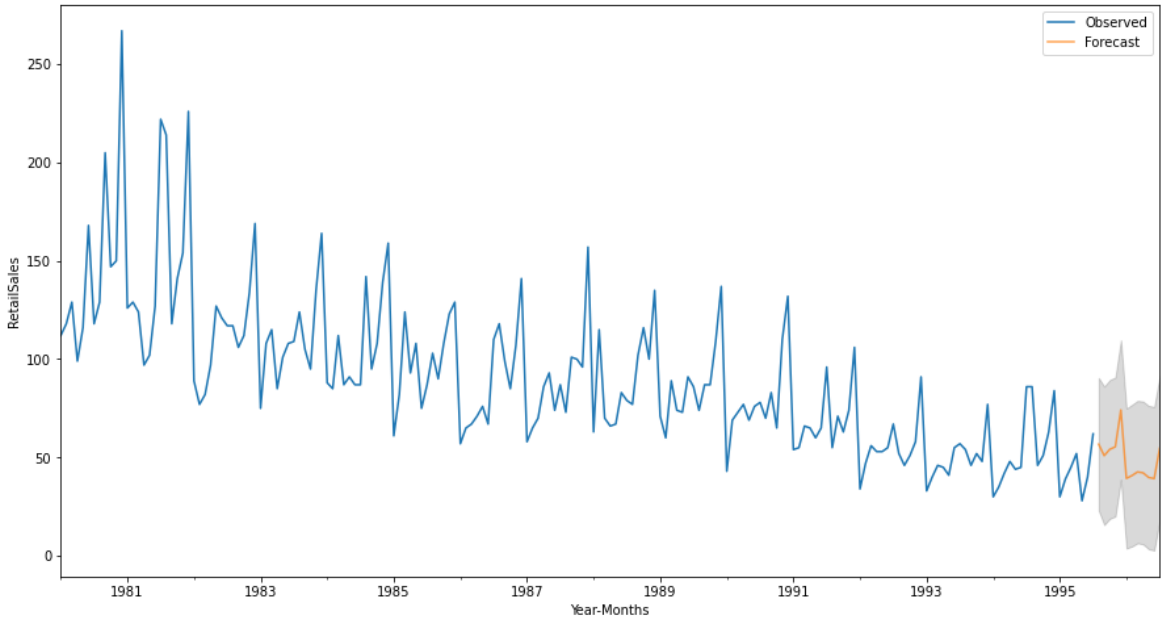
- We can see the full model RMSE is 12 for Trailling MA which is a very good RMSE we can understand the model is performing at its best and we can consider it to check the Rose wine sales.

Predicted manual SARIMA

Rose	mean	mean_se	mean_ci_lower	mean_ci_upper
1995-08-31	56.749588	17.237994	22.963741	90.535435
1995-09-30	50.921455	17.959972	15.720557	86.122353
1995-10-31	54.150666	18.007130	18.857339	89.443992
1995-11-30	55.419384	18.054165	20.033871	90.804898
1995-12-31	74.150422	18.101078	38.672961	109.627884

RMSE:

```
RMSE of the Full Model 28.242114247529067
```



TEST RMSE

	Test RMSE
ARIMA(2,1,1)	16.788318
ARIMA(0,1,0)	82.847243
SARIMA(0,1,2)(2,0,2,6)	26.446048
SARIMA(1,1,2)(2,0,2,6)	26.391284
SARIMA(0,1,0)(1,1,3,6)	37.252115

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

Answer:

- The company should be using the 2-point trailing Moving average to check the sales of the Rose wine.
- The best measures company should be considering is it to make sure the sales are published by different sources of media options.
- We know in order to get the customer base we need to be manufacturing the best quality wines with affordable prices.
- Having expensive wine reduces the sale as not everyone can afford such expensive drinks.
- The quality of the wines should be maintained.
- As we know the older the drink the expensive it is. We need to make sure there is enough stock which is saved in the backend to supply when necessary.
- Proper testing and tasting should happen to understand the likes and dislikes of the drink.
- We can use the above Trailing moving average model and focus on the customer base depending on the seasonal data.
- When we compare both the wines as they are manufactured by same company.
- We need to make sure the taste varies a little and quality is different for consumers to understand the difference to purchase it.
- We can use the moving average model to predict what the trend is and proceed accordingly.
- In order to achieve better sales in need to focus in all the fields like Sales, Quality, Manufacturing etc.
- As we have plotted lots of graph including AIC to understand if the sales are based on the season or a trend.
- The Rose data set speaks more about trend and not seasonality.