TIME SERIES FORECASTING BUSINESS REPORT

-APOORVA P

Contents

PROBLEM 1 : SPARKLING.CSV

CONTEXT

As an analyst at ABC Estate Wines, we are presented with historical data encompassing the sales of different types of wines throughout the 20th century. These datasets originate from the same company but represent sales figures for distinct wine varieties. Our objective is to delve into the data, and analyze trends, patterns, and factors influencing wine sales of sparkling and rose wine over the century. By leveraging data analytics and forecasting techniques, we aim to gain actionable insights that can inform strategic decision-making and optimize sales strategies for the future.

OBJECTIVE

The primary objective of this project is to analyze and forecast wine sales trends for the 20th century based on historical data provided by ABC Estate Wines. We aim to equip ABC Estate Wines with the necessary insights and foresight to enhance sales performance, capitalize on emerging market opportunities, and maintain a competitive edge in the wine industry.

DEFINE THE PROBLEM AND PERFORM EXPLORATORY DATA ANALYSIS

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

```
df.tail()
```

	YearMonth	Sparkling
182	1995-03	1897
183	1995-04	1862
184	1995-05	1670
185	1995-06	1688
186	1995-07	2031

Data INFO

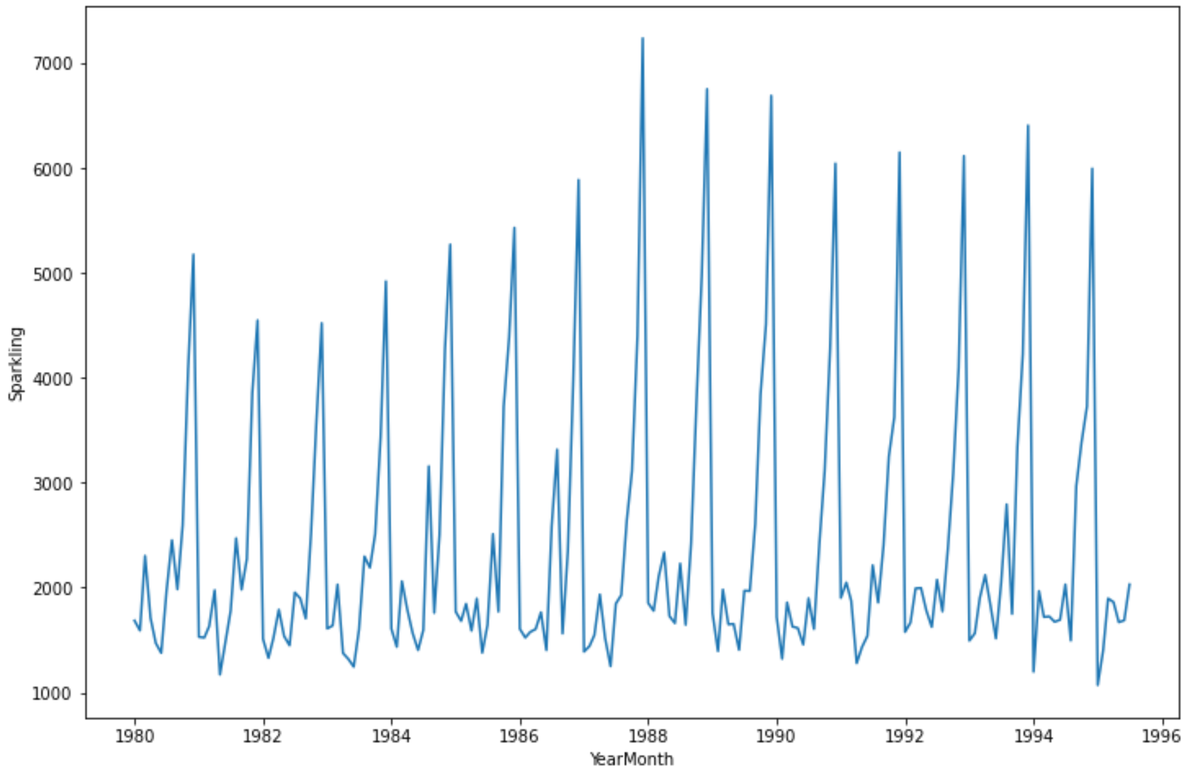
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 187 entries, 0 to 186
Data columns (total 2 columns):
 #   Column     Non-Null Count  Dtype 
---  ------     --------------  ----- 
 0   YearMonth  187 non-null    object
 1   Sparkling  187 non-null    int64 
dtypes: int64(1), object(1)
memory usage: 3.0+ KB
```

DESCRIPTION:

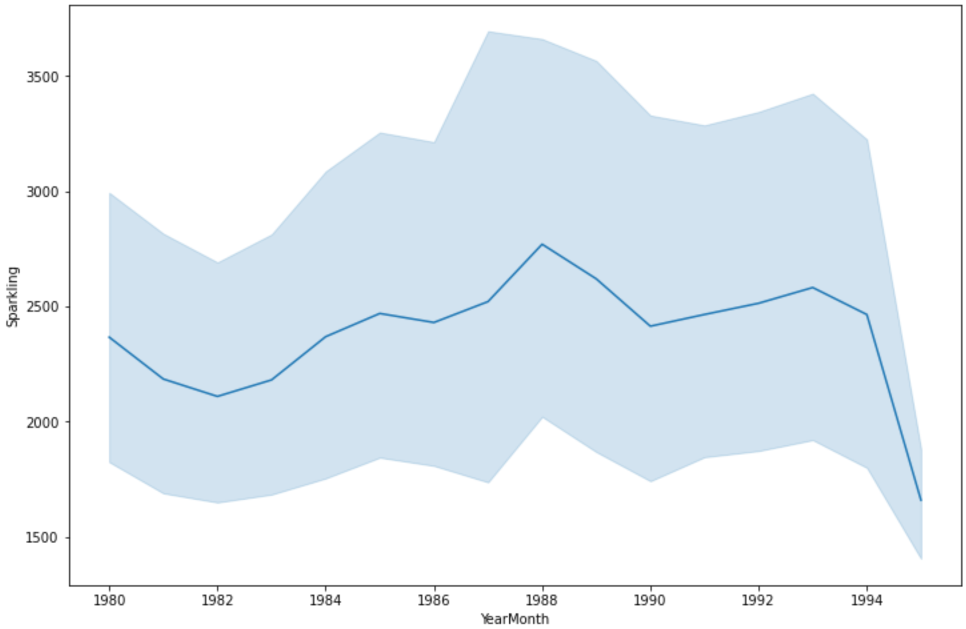
	Sparkling
Count	*187.000*
Mean.	*2402.417*
Std.	*1295.112*
Min.	*1070.000.*

25%.	*1605.000*
50%.	*1874.000*
75%.	*2549.000*
Max.	*7242.000*

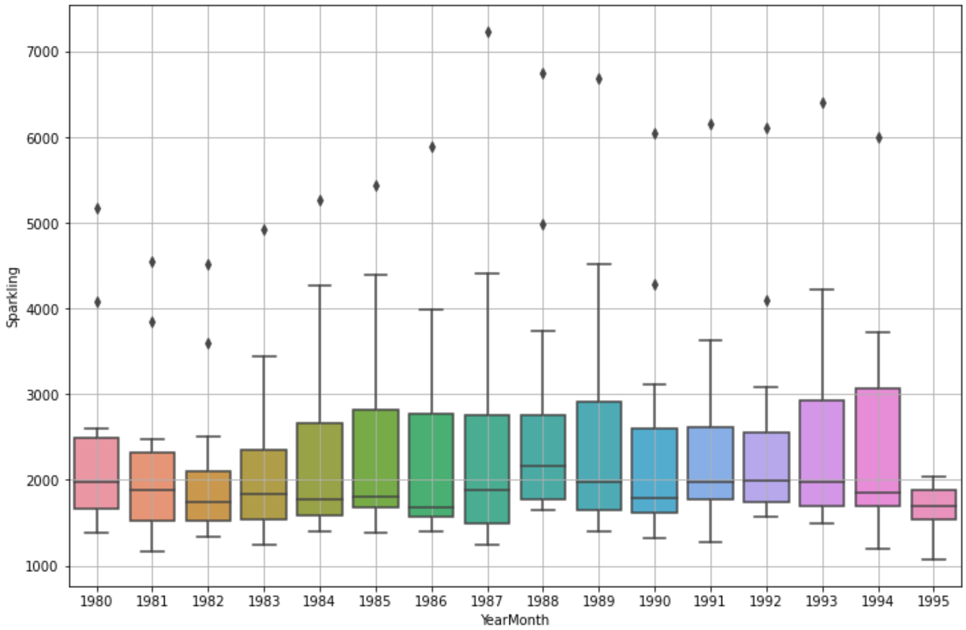
The Sparkling dataset has 187 rows and no null values the yearMonth column was originally saved as object datatype has been converted to date time data type and set as index



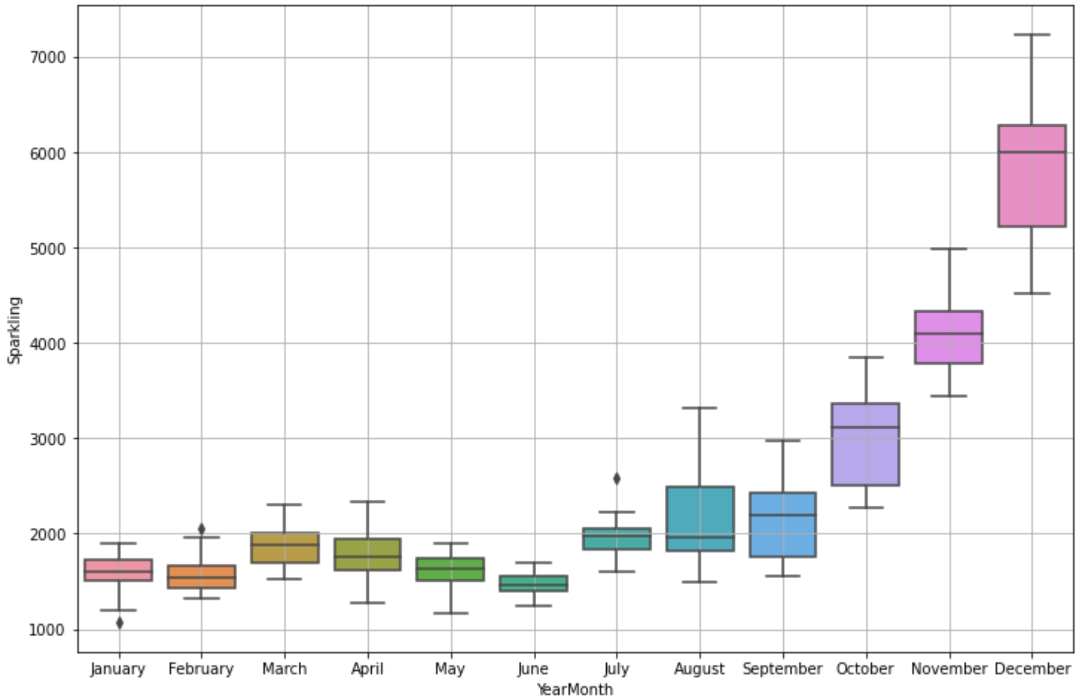
THE GRAPH SHOWS THAT THERE EXISTS SOME SEASONALITY AND NO SIGNS OF TREND



THE LINE PLOT SHOWS A CLEAR NON EXSISTING TREND

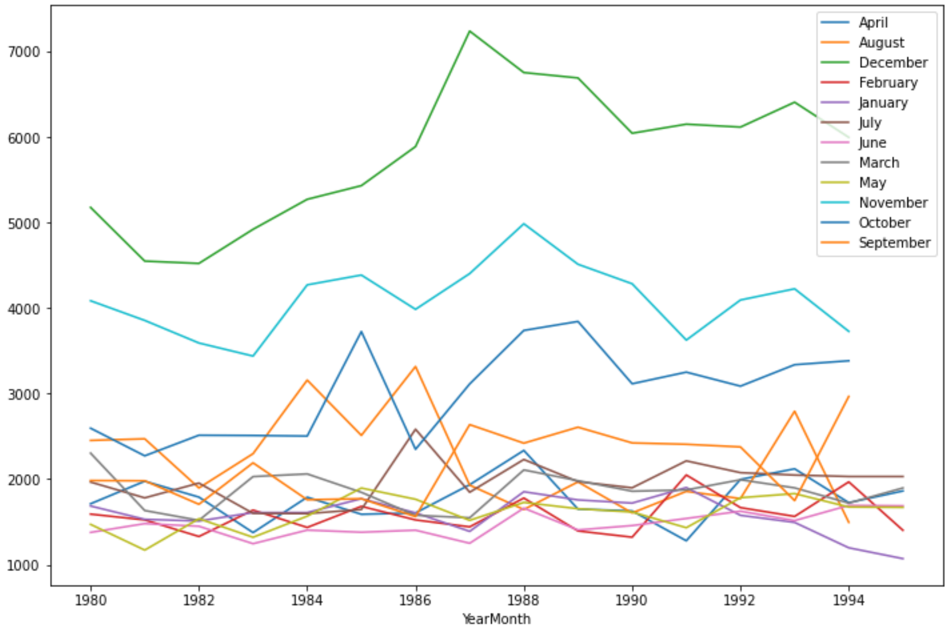


THE BOX PLOT OF YEAR WISE SALE SHOWS THAT THE SALE HAS MORE OVER BEEN A CONSTANT



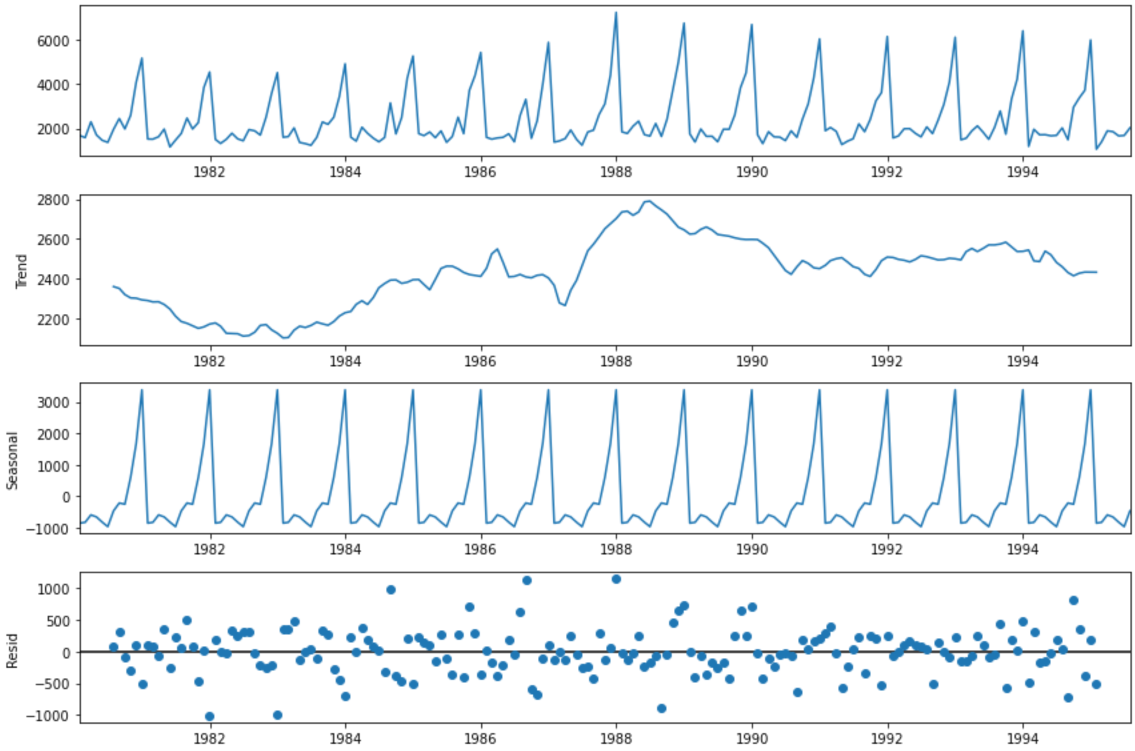
THE MONTH LEVEL BOX PLOT SHOWS THAT THE SALES HAVE BEEN HIGHEST IN DECEMBER.

IT SHOWS A GRADUAL INCREASE IN SALES AS WE GO FROM JAN TO DEC

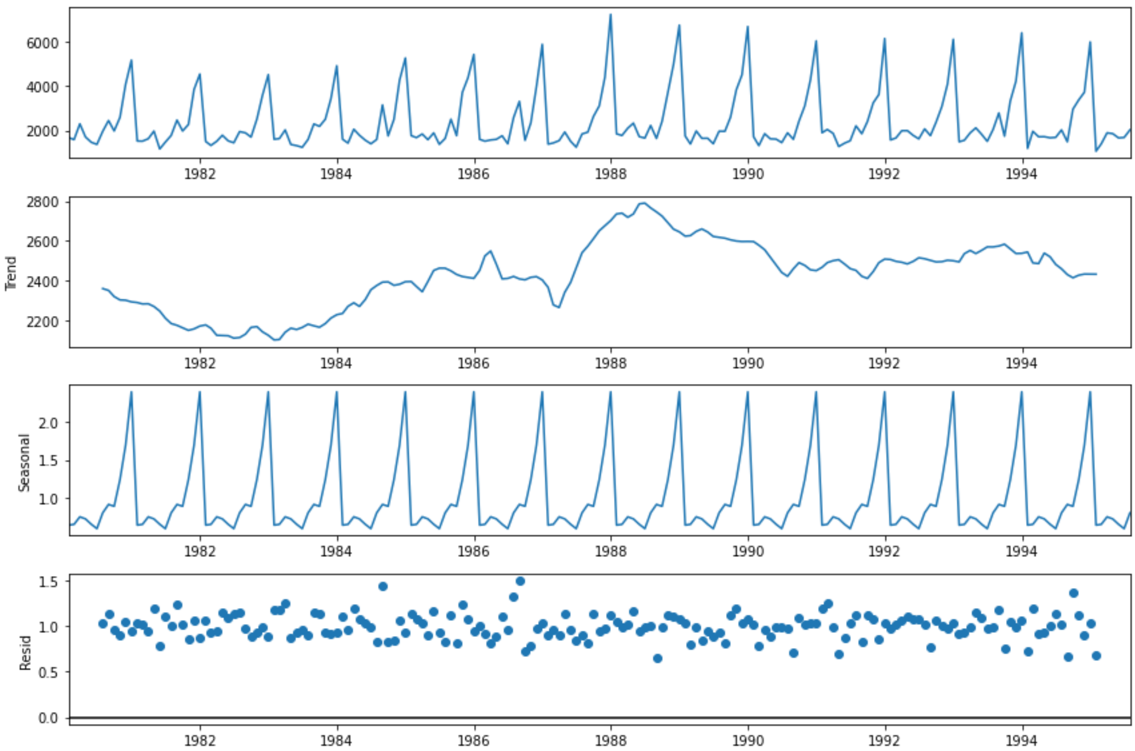


THE LINE PLOT CLEARLY SHOWS THE HIGHEST SALES ARE IN DECEMBER FOLLOWED BY NOV AND THEN OCT

AS THERE IS SEASONALITY AND NO TREAND WE CAN DECOMPOSE ADDITIVELY



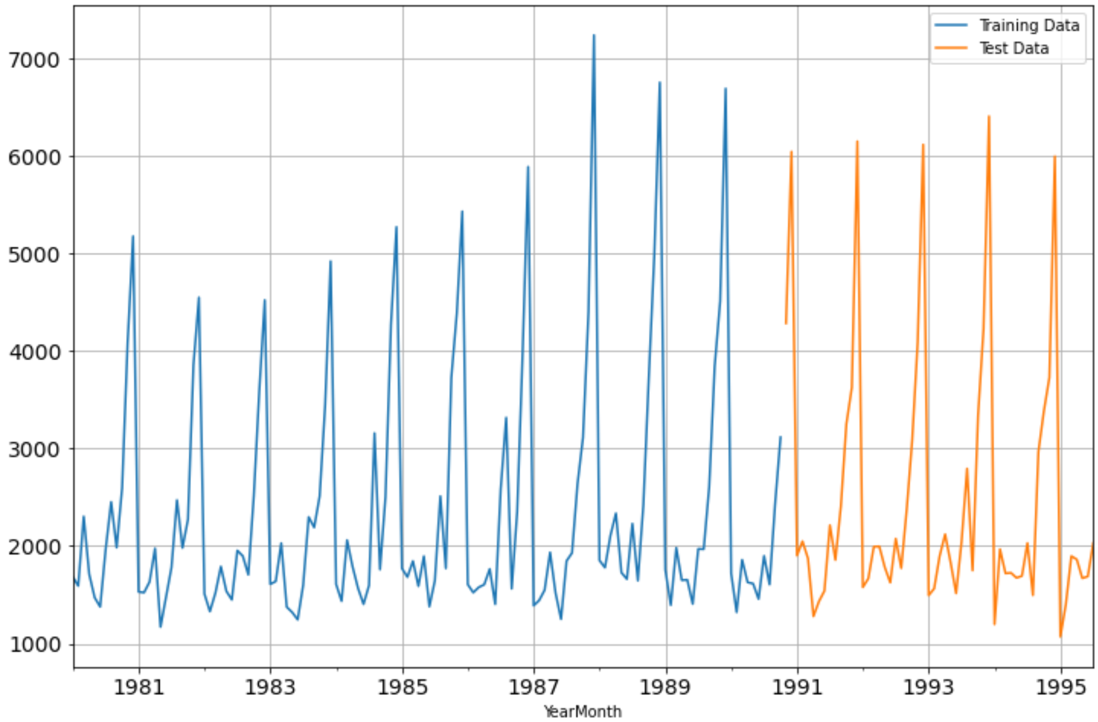
THE RESID IS RANDOM WHICH SHOWS THE DECOMPOSITION FITS
THE MULTIPLICATIVE DECOMPOSTIONS GIVES A SIMILAR GRAPH IN TERMS OF TREAND AND SEASONALITY



DATA PRE-PROCESSING

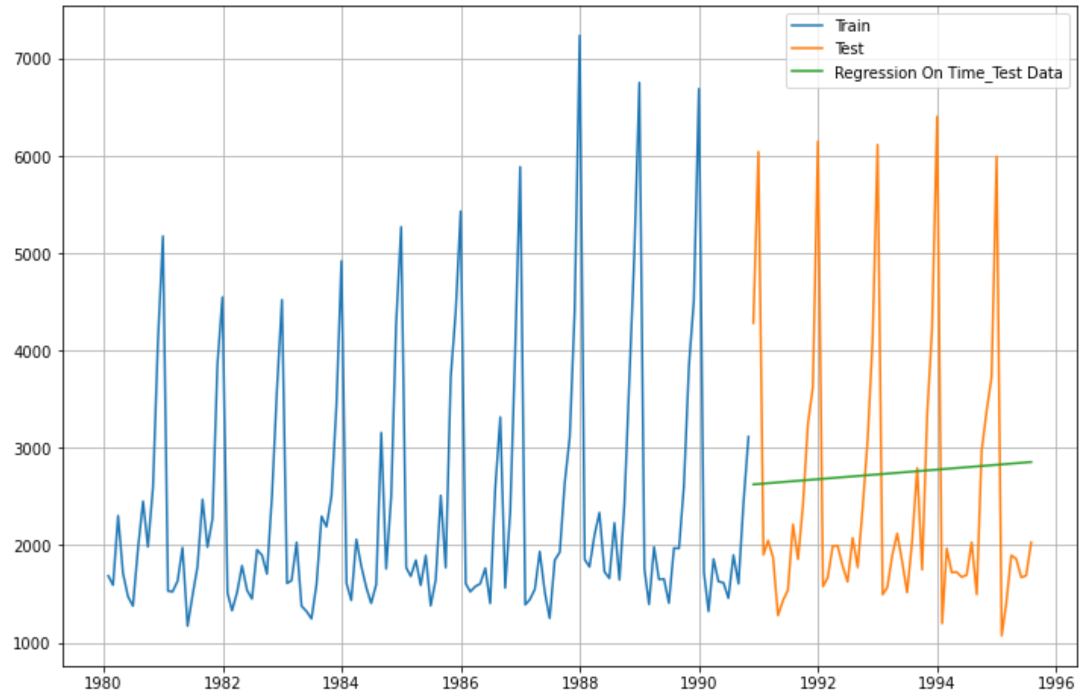
WE WILL SPLIT THE DATA AT 70-30 RATIO
THE FIRST 70% WILL BE TRAIN DATA AND REST 30% WILL BE TEST D
ATA

TRAIN DATA (130, 1)
TEST DATA (57, 1)



MODEL BUILDING - ORIGINAL DATA

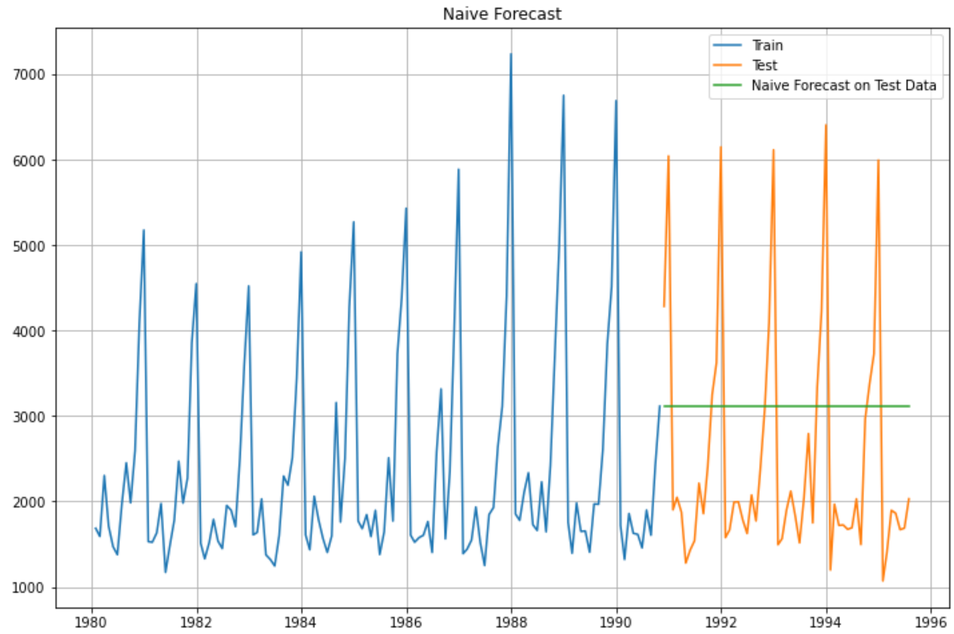
LINEAR REGRESSION




```
For RegressionOnTime forecast on the Test Data,  RMSE is 1391.709
```

NAIVE APPROACH

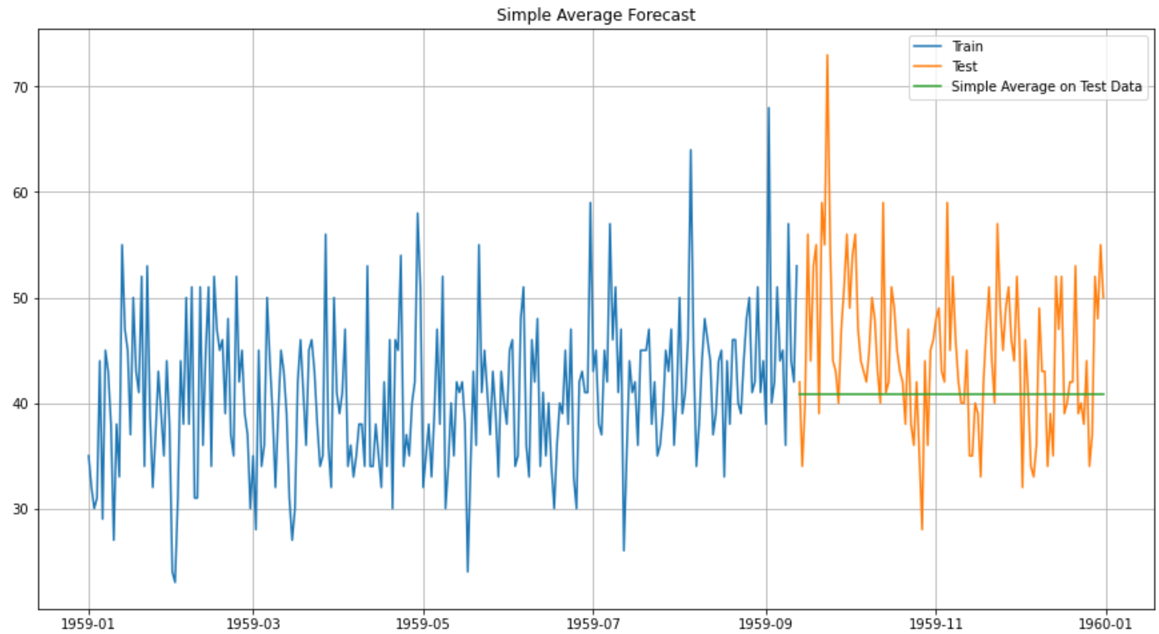
For this particular naive model, we say that the prediction for tomorrow is the same as today and the prediction for day after tomorrow is tomorrow and since the prediction of tomorrow is same as today, therefore the prediction for day after tomorrow is also today



For RegressionOnTime forecast on the Test Data, RMSE is 1496.445

SIMPLE AVERAGE

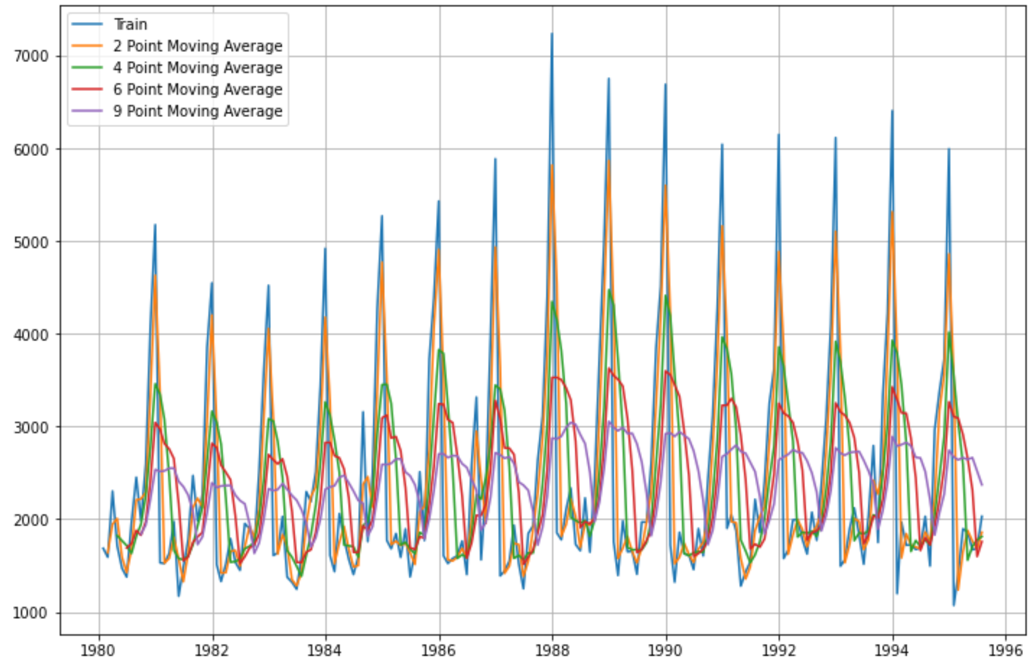
For this particular simple average method, we will forecast by using the average of the training values.

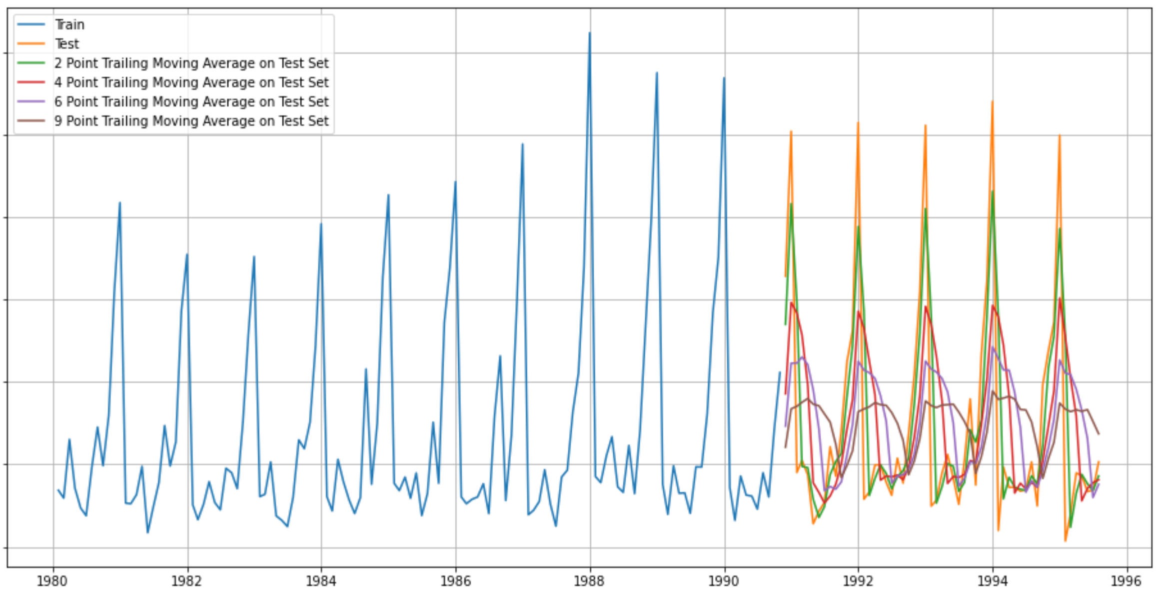


For Simple Average forecast on the Test Data, RMSE is 8.195

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.

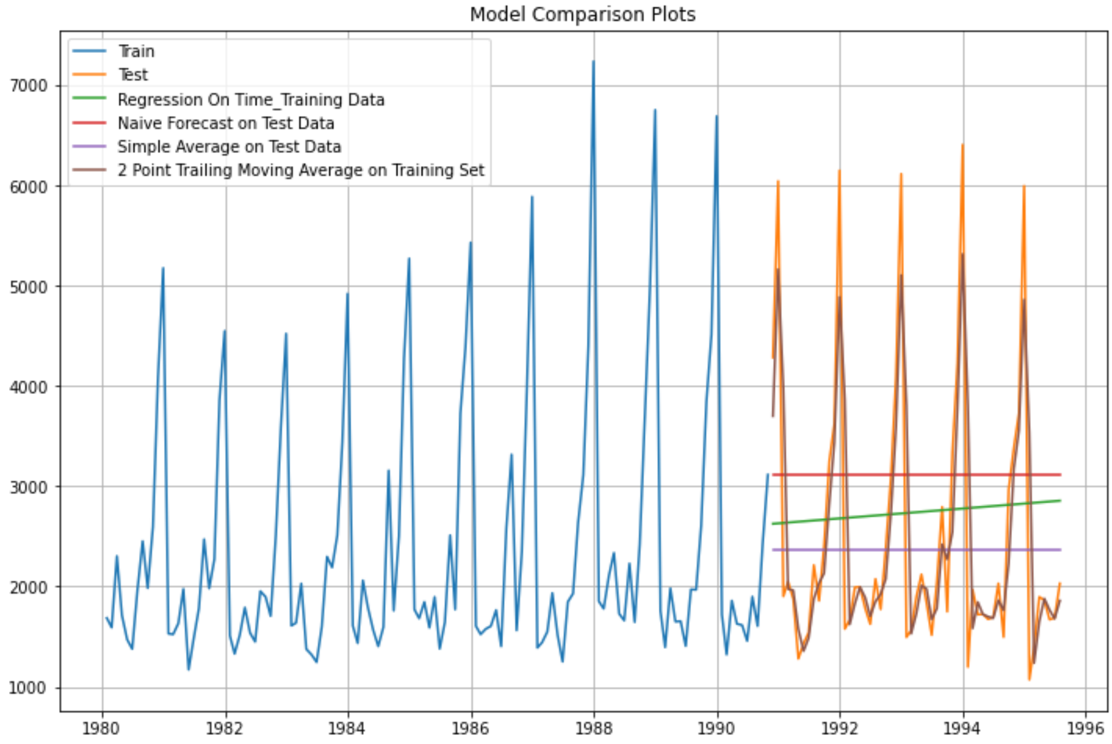





```
For 2 point Moving Average Model forecast on the Training Data,  RMSE i
s 811.179
For 4 point Moving Average Model forecast on the Training Data,  RMSE i
s 1184.213
```

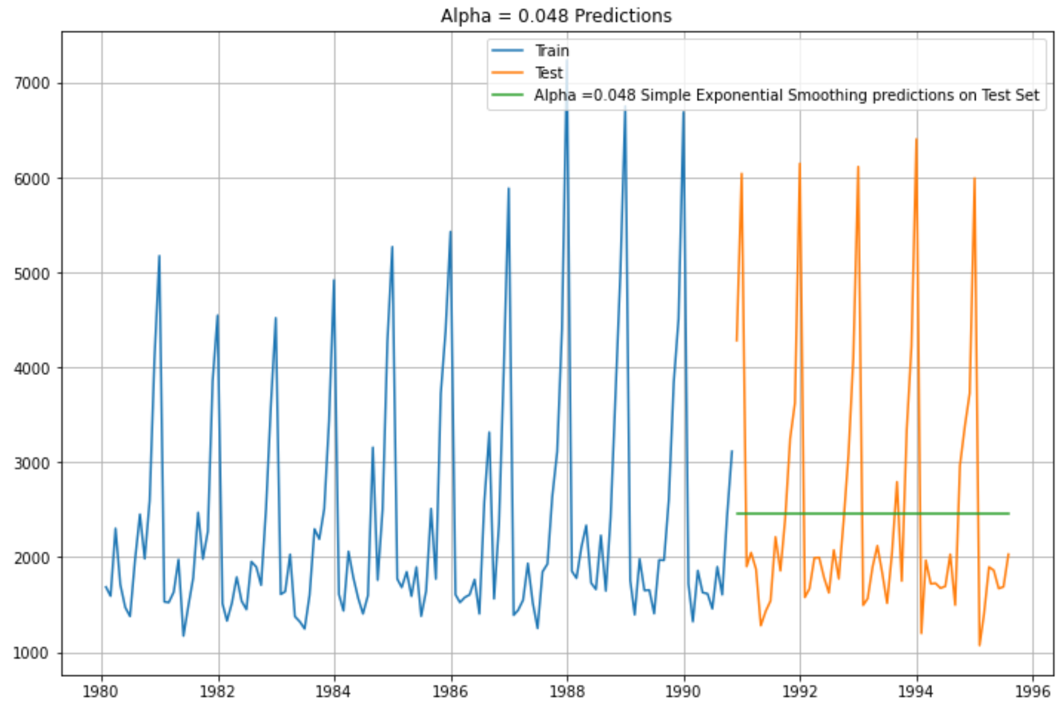
```
For 6 point Moving Average Model forecast on the Training Data,  RMSE i
s 1337.201
For 9 point Moving Average Model forecast on the Training Data,  RMSE i
s 1422.653
```

As you can see 2 point gives the best predictions compared to others



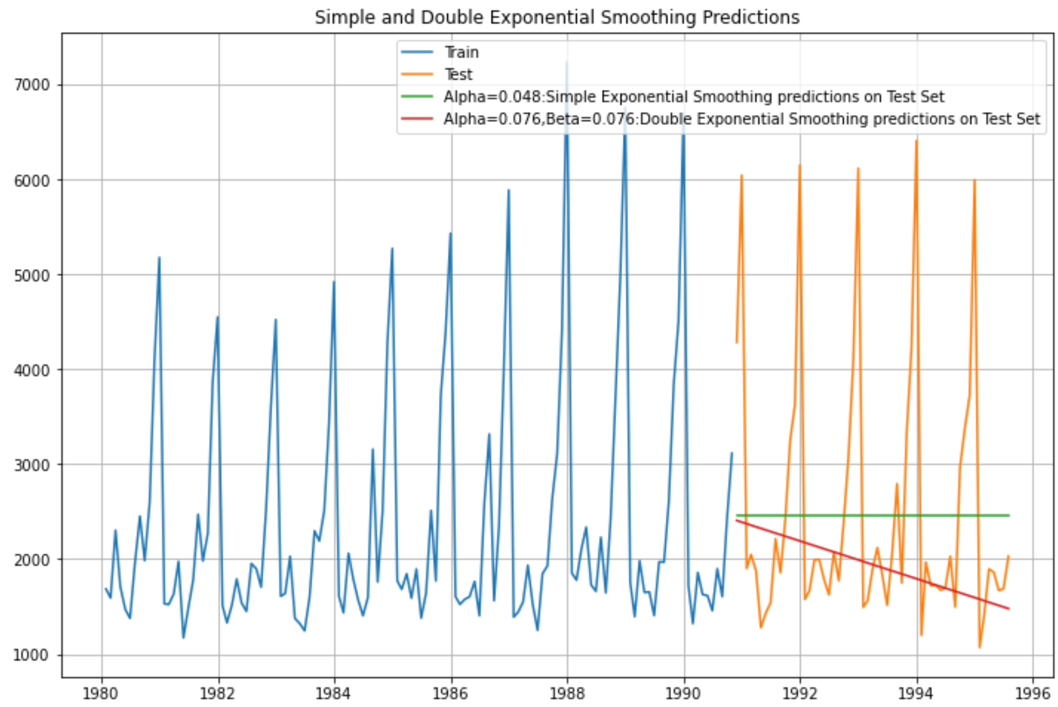
SIMPLE EXPONENTIAL SMOOTHENING MODELS

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 PARAMETER ALPHA 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: THERE IS SEASONALITY BUT NO TREND IN THIS DATA***



SES RMSE: 1362.488304892539
SES RMSE (calculated using statsmodels): 1362.488304892539

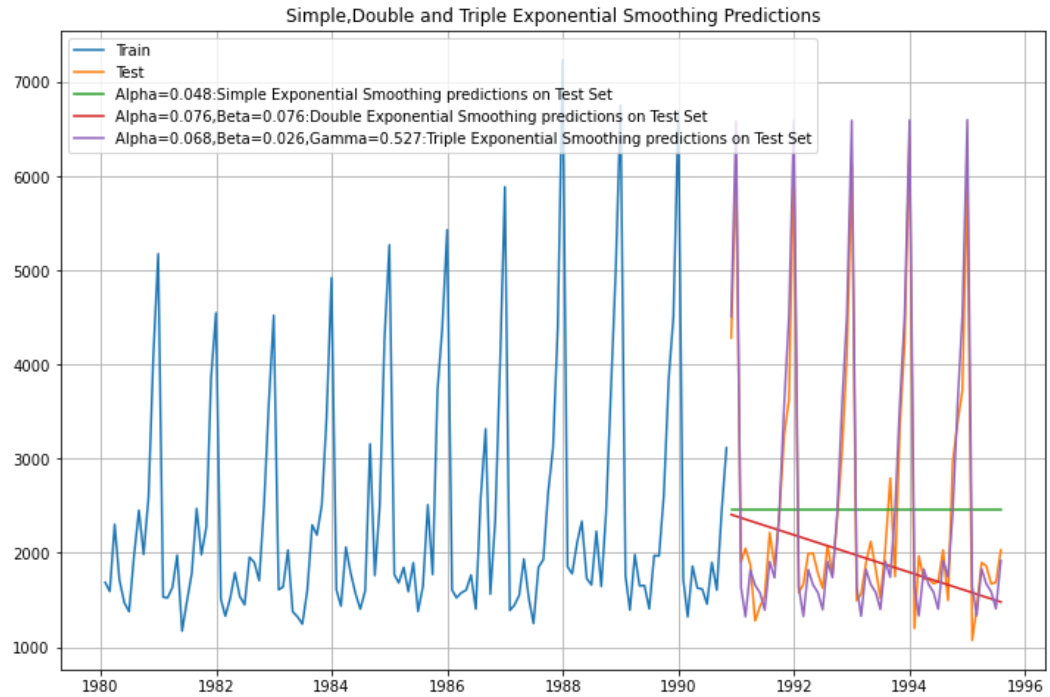
DOUBLE EXPONENTIAL SMOOTHING



DES RMSE: 1472.2536598820277

The DES tries to capture a trend

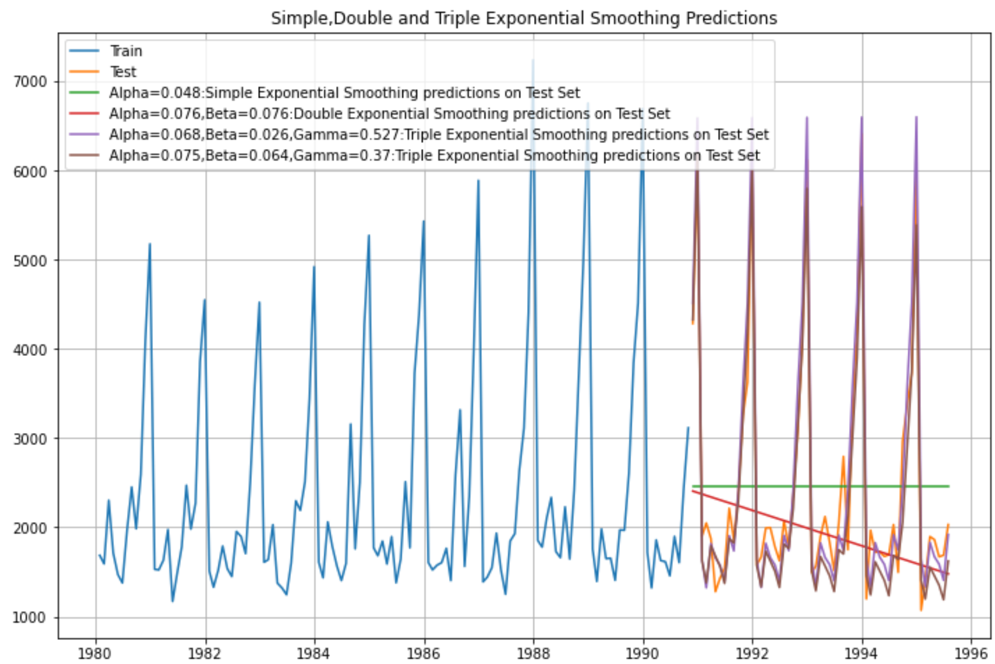
TRIPLE EXPONENTIAL SMOOTHING




```
TES RMSE: 377.3607522826444
```

TES captures and best predicts the future

HOLT WINTER'S LINEAR METHOD



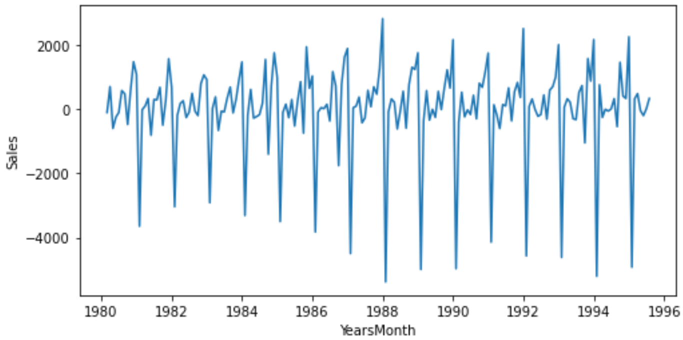

```
TES_am RMSE: 381.6553661142565

This is TES but by defining the parameters we can get a better predicti
on
```

CHECK FOR STATIONARITY

```
DF test statistic is -1.360
DF test p-value is 0.6011
```

The p value is greater than 0.5 hence it is not stationary
To make it stationary we need to decompose it

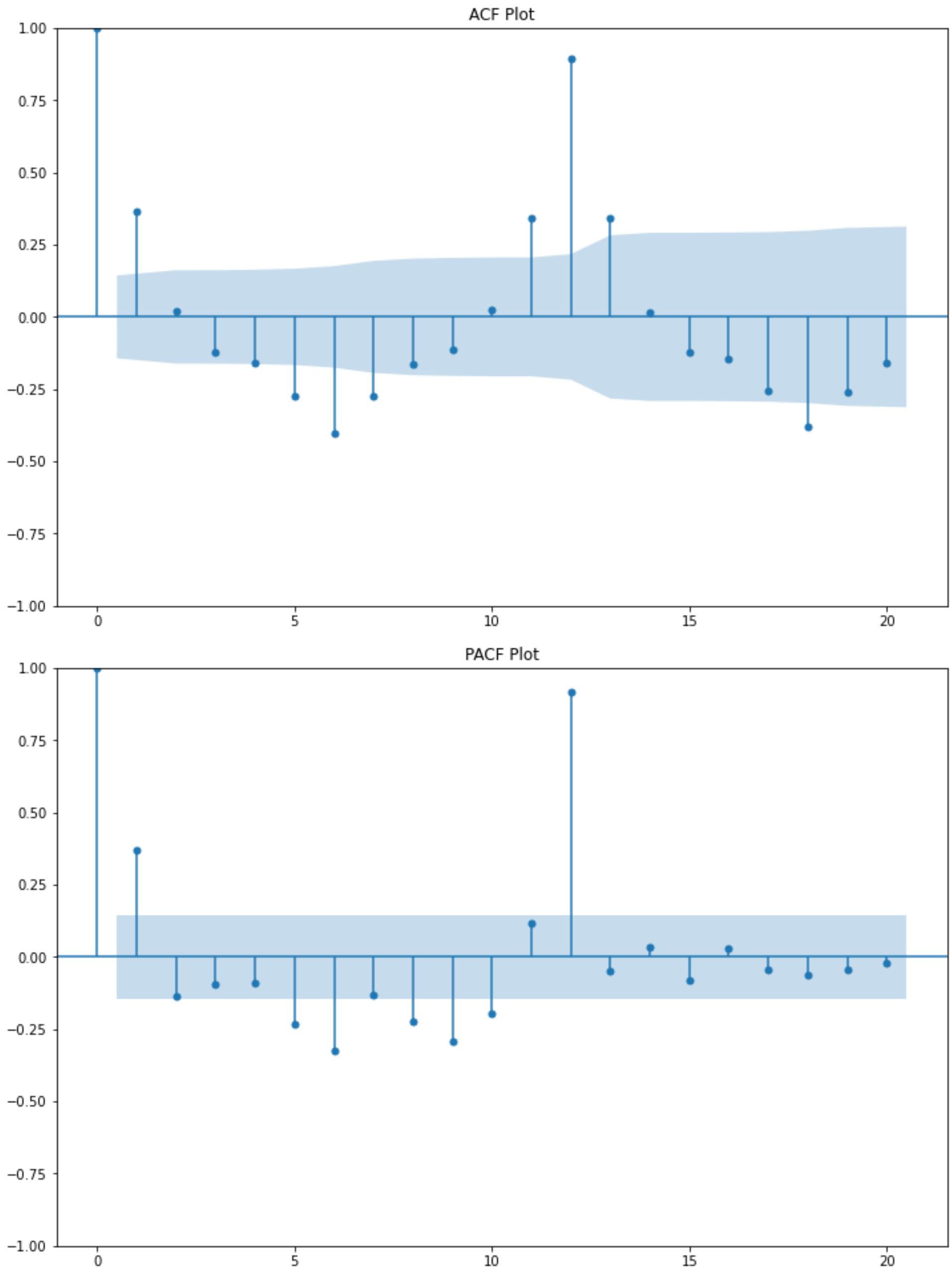
```
ADF Statistic: -1.3604974548123385
p-value: 0.6010608871634847
Critical Values:
	1%: -3.4682803641749267
	5%: -2.8782017240816327
	10%: -2.5756525795918366
Data made stationary.
```



d=1

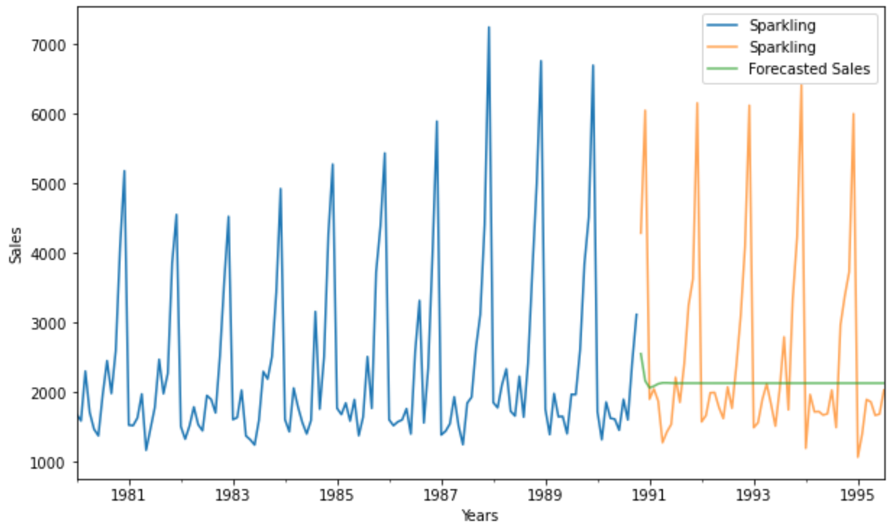
MODEL BUILDING - STATIONARY DATA

ACF AND PACF MODEL:



AR MODEL

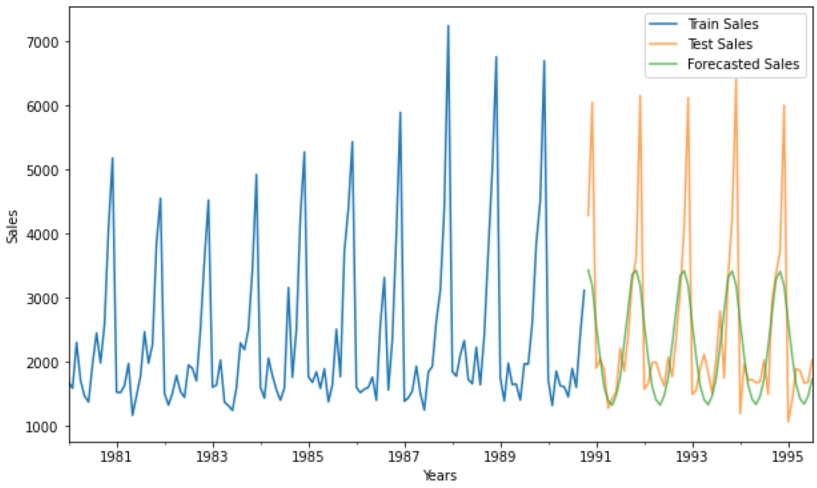
```
SARIMAX Results
==============================================================================
Dep. Variable:              Sparkling   No. Observations:                  130
Model:                 ARIMA(2, 0, 0)   Log Likelihood                  53.843
Date:                Sun, 17 Mar 2024   AIC                            -99.687
Time:                        20:50:21   BIC                            -88.216
Sample:                    01-31-1980   HQIC                           -95.026
                         - 10-31-1990
Covariance Type:                  opg
==============================================================================
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const          3.3287      0.030    110.321      0.000       3.270       3.388
ar.L1          0.5503      0.125      4.406      0.000       0.306       0.795
ar.L2         -0.2206      0.153     -1.440      0.150      -0.521       0.080
sigma2         0.0255      0.004      6.974      0.000       0.018       0.033
==============================================================================
```



Best AR Model : ARIMA(2,0,0) RSME : 1398.059012

ARMA MODEL

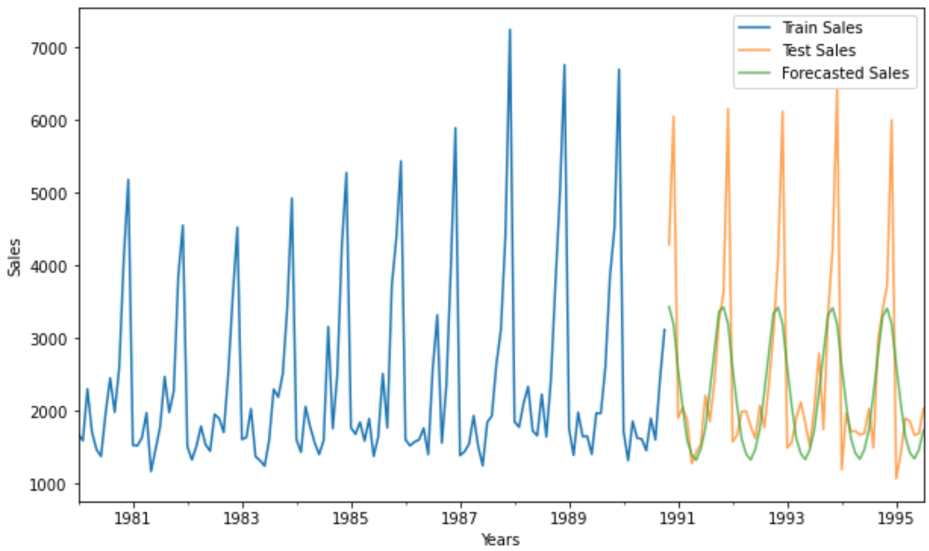
```
SARIMAX Results
==============================================================================
Dep. Variable:              Sparkling   No. Observations:                  130
Model:                 ARIMA(3, 0, 3)   Log Likelihood                  76.073
Date:                Sun, 17 Mar 2024   AIC                           -136.146
Time:                        21:16:38   BIC                           -113.206
Sample:                    01-31-1980   HQIC                          -126.824
                         - 10-31-1990
Covariance Type:                  opg
==============================================================================
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const          3.3330      0.012    281.782      0.000       3.310       3.356
ar.L1          0.7438      0.157      4.731      0.000       0.436       1.052
ar.L2          0.7133      0.274      2.605      0.009       0.177       1.250
ar.L3         -0.9872      0.158     -6.243      0.000      -1.297      -0.677
ma.L1         -0.7658      0.201     -3.804      0.000      -1.160      -0.371
ma.L2         -0.7357      0.357     -2.063      0.039      -1.435      -0.037
ma.L3          0.9615      0.253      3.794      0.000       0.465       1.458
sigma2         0.0176      0.003      5.127      0.000       0.011       0.024
==============================================================================
```



Best ARMA Model : ARIMA(3,0,3) RMSE : 1021.304310

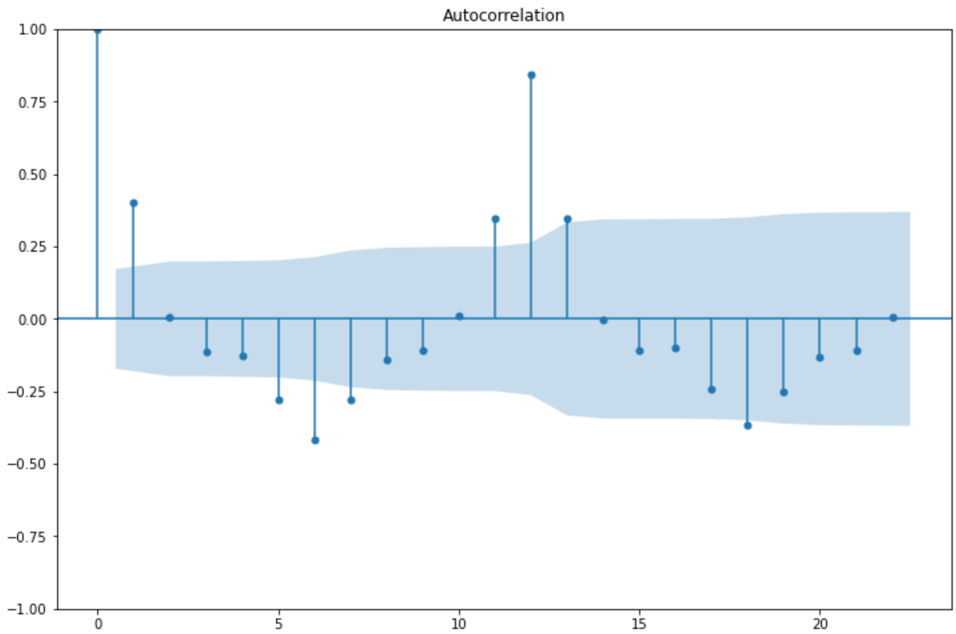
ARIMA MODEL

```
SARIMAX Results
==============================================================================
Dep. Variable:              Sparkling   No. Observations:                  130
Model:                 ARIMA(3, 0, 3)   Log Likelihood                  76.073
Date:                Sun, 17 Mar 2024   AIC                           -136.146
Time:                        21:23:03   BIC                           -113.206
Sample:                    01-31-1980   HQIC                          -126.824
                         - 10-31-1990
Covariance Type:                  opg
==============================================================================
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const          3.3330      0.012    281.782      0.000       3.310       3.356
ar.L1          0.7438      0.157      4.731      0.000       0.436       1.052
ar.L2          0.7133      0.274      2.605      0.009       0.177       1.250
ar.L3         -0.9872      0.158     -6.243      0.000      -1.297      -0.677
ma.L1         -0.7658      0.201     -3.804      0.000      -1.160      -0.371
ma.L2         -0.7357      0.357     -2.063      0.039      -1.435      -0.037
ma.L3          0.9615      0.253      3.794      0.000       0.465       1.458
sigma2         0.0176      0.003      5.127      0.000       0.011       0.024
==============================================================================
```



Best ARIMA Model : ARIMA(3,0,3)RMSE : 1021.304310

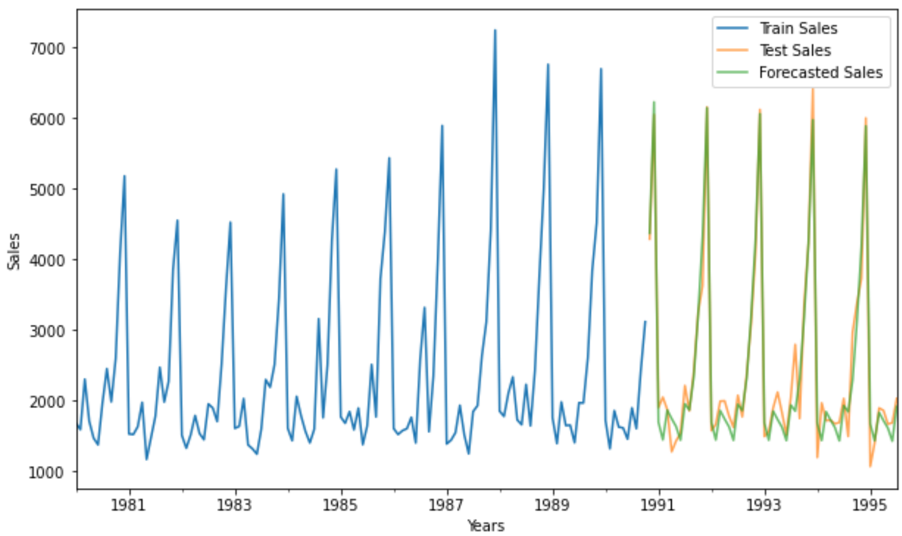
SARIMA MODEL



	param	seasonal	AIC
0	(1, 0, 1)	(1, 0, 1, 12)	-284.255373
156	(1, 0, 1)	(1, 0, 1, 12)	-284.255373

	param	seasonal	AIC
108	(2, 0, 1)	(1, 0, 1, 12)	-282.384005
18	(1, 0, 2)	(1, 0, 1, 12)	-280.896437
126	(2, 0, 2)	(1, 0, 1, 12)	-280.126867

```
                                     SARIMAX Results
==========================================================================================
Dep. Variable:                            Sparkling   No. Observations:                130
Model:             SARIMAX(1, 0, 1)x(1, 0, 1, 12)   Log Likelihood                146.198
Date:                              Sun, 17 Mar 2024   AIC                          -282.396
Time:                                      22:21:52   BIC                          -268.058
Sample:                                  01-31-1980   HQIC                         -276.570
                                       - 10-31-1990
Covariance Type:                                opg
==========================================================================================
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.9991      0.005    209.467      0.000       0.990       1.008
ma.L1         -0.9301      0.041    -22.890      0.000      -1.010      -0.850
ar.S.L12       0.9917      0.006    168.411      0.000       0.980       1.003
ma.S.L12      -0.6331      0.079     -7.992      0.000      -0.788      -0.478
sigma2         0.0045      0.001      8.066      0.000       0.003       0.006
==============================================================================
```



Best SARIMA Model : SARIMAX(1, 0, 1)x(1, 0, 1, 12) RMSE: 301.293319

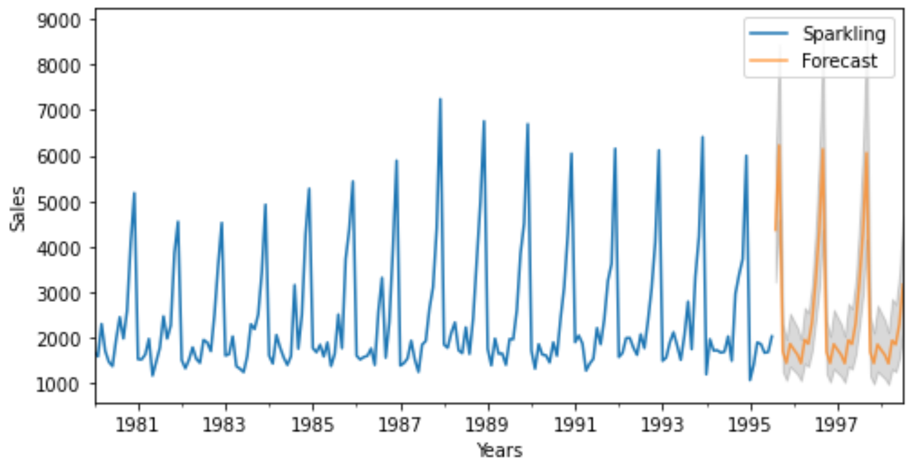
COMPARE THE PERFORMANCE OF THE MODELS

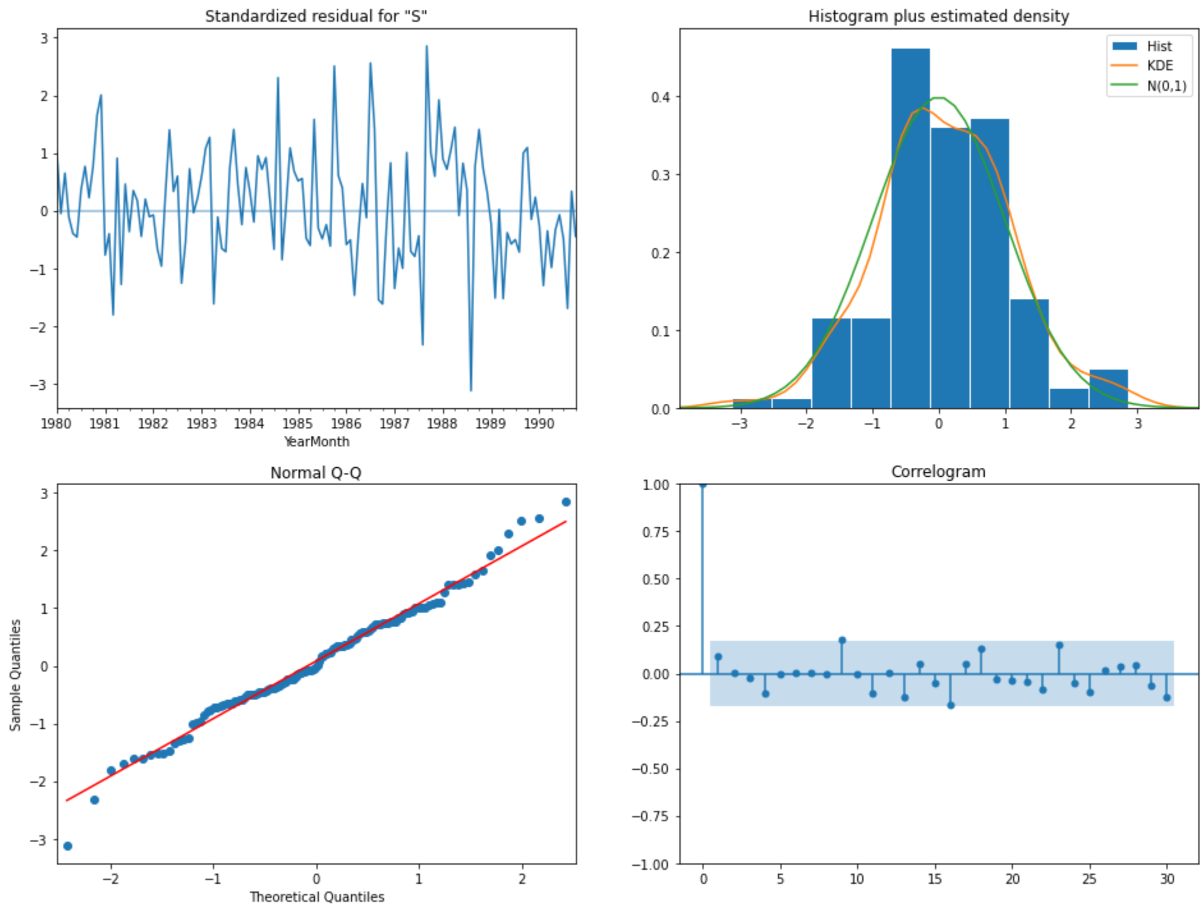
REGRESSIONONTIME	1391.708631
SIMPLEAVERAGEMODE	1368.746 77
2POINTTRAILINGMOVINGAVERAGE	811.1789 37
4POINTTRAILINGMOVINGAVERAGE	1184.213 25
6POINTTRAILINGMOVINGAVERAGE	1337.200524
9POINTTRAILINGMOVINGAVERA GE	1422.653281
SES	1362.488305
DES	1472.253 60
TES	377.360752
TES_AM	381.6553 66
BEST AR MODEL : ARIMA(2,0,0)	1398.059012
BEST ARMA MODEL : ARIMA(3,0,3)	1021.304310
BEST ARIMA MODEL : ARIMA(3,0,3)	1021.304310
BEST SARIMA MODEL :SARIMAX(1,0,1)X(1,0,1,12)	301.293319

This show the results of all the models and based on this

The best fit model for original data is TES
And the best fit model for the stationary data is SARIMA MODEL

Forecast for the next 3 years:





ACF and PACF for residuals of ARIMA model to ensure no more information is left for extraction

ACTIONABLE INSIGHTS & RECOMMENDATIONS

Sales in Sparking does not have uniform trend but increased in some years and decreased later. Business study may be done to find why sales are not increasing and what the contributing factors. Study can also include to see which wine product has substituted/ had higher sales in the years of low sales of Sparkling.

With promotion and focussed effort with micro detailing it may be feasible to increase the sales. Sales of Sparkling wine higher in the later part of the year. This may be due to climatic condition of the geography under study.

The sales for Sparkling wine for the company are predicted to be at least the same as last year, if not more, with peak sales for next year potentially higher than this year.

Sparkling wine has been a consistently popular wine among customers with only a very marginal decline in sales, despite reaching its peak popularity in the late 1980s.

Seasonality has a significant impact on the sales of Sparkling wine, with sales being slow in the first half of the year and picking up from August to December.

It is recommended for the company to run campaigns in the first half of the year when sales are slow, particularly in the months of March to July.

Combining promotions where Sparkling wine is paired with a less popular wine such as "Rose wine" under a special offer may encourage customers to try the underperforming wine, which could potentially boost its sales and benefit the company.

PROBLEM 2 : ROSE.CSV

CONTEXT

As an analyst at ABC Estate Wines, we are presented with historical data encompassing the sales of different types of wines throughout the 20th century. These datasets originate from the same company but represent sales figures for distinct wine varieties. Our objective is to delve into the data, and analyze trends, patterns, and factors influencing wine sales of sparkling and rose wine over the century. By leveraging data analytics and forecasting techniques, we aim to gain actionable insights that can inform strategic decision-making and optimize sales strategies for the future.

OBJECTIVE

The primary objective of this project is to analyze and forecast wine sales trends for the 20th century based on historical data provided by ABC Estate Wines. We aim to equip ABC Estate Wines with the necessary insights and foresight to enhance sales performance, capitalize on emerging market opportunities, and maintain a competitive edge in the wine industry.

DEFINE THE PROBLEM AND PERFORM EXPLORATORY DATA ANALYSIS

```
df.head()
```

	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

```
df.tail()
```

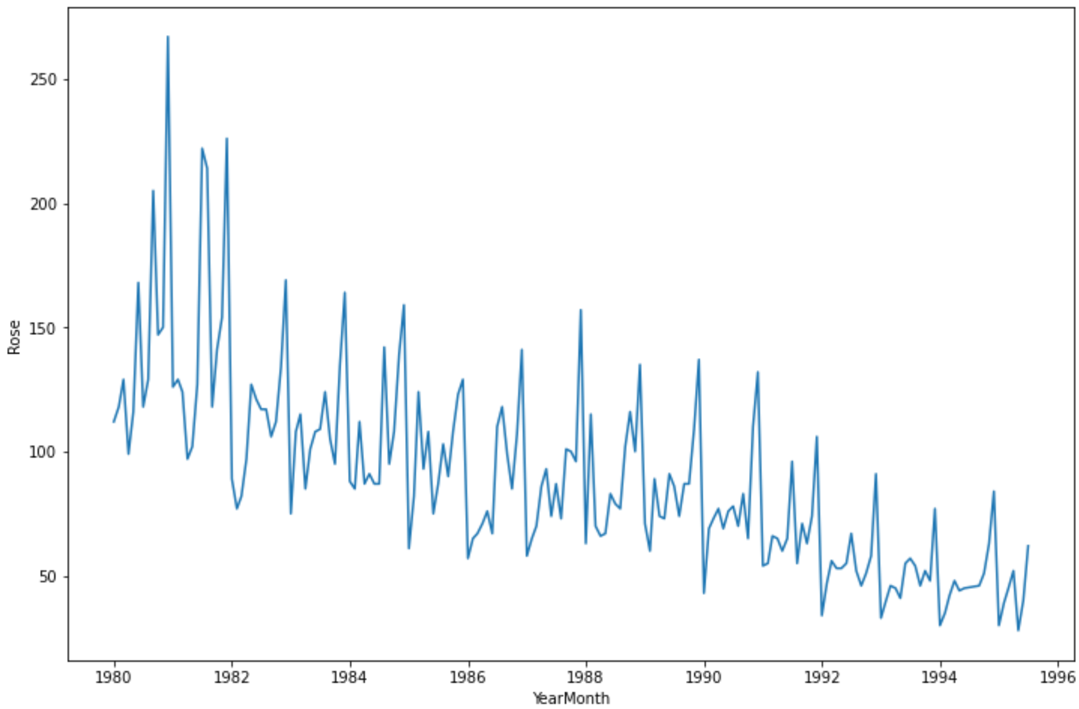
	YearMonth	Rose
182	1995-03	45.0
183	1995-04	52.0
184	1995-05	28.0
185	1995-06	40.0
186	1995-07	62.0

The Rose dataset has 187 rows and 2 null values. The yearMonth column was originally saved as object datatype has been converted to date time data type and set as index.

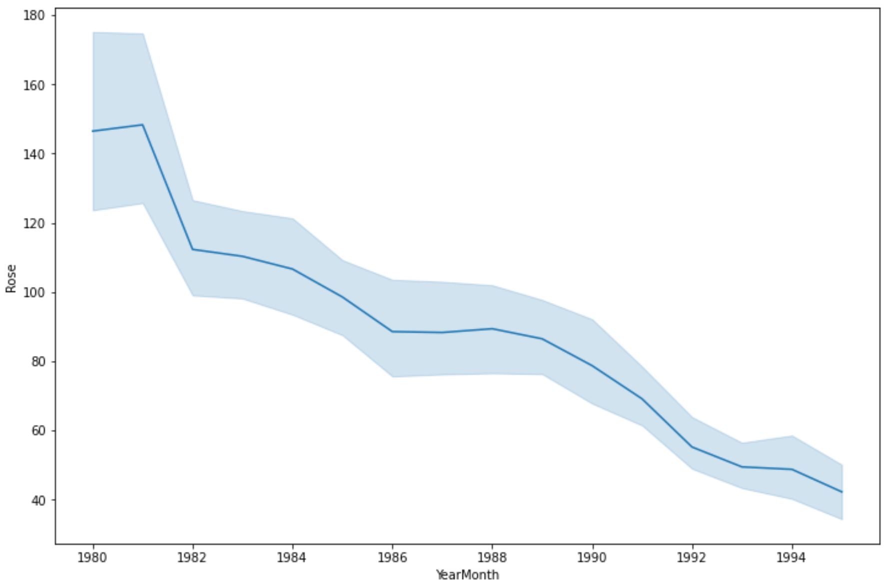
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 187 entries, 0 to 186
Data columns (total 2 columns):
 #   Column     Non-Null Count  Dtype  
---  ------     --------------  -----  
 0   YearMonth  187 non-null    object 
 1   Rose       185 non-null    float64
dtypes: float64(1), object(1)
memory usage: 3.0+ KB
```

The null values were treated by taking the average of the month over the years

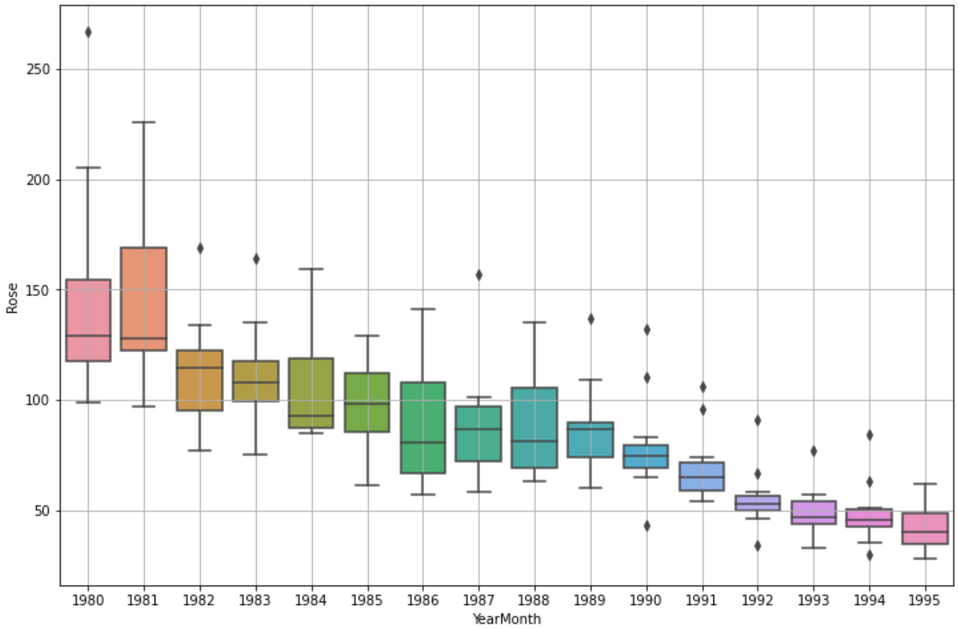
	Rose
count	185.000
mean	90.395
std	39.175
min	28.000
25%	63.000
50%	86.000
75%	112.000
max	267.000



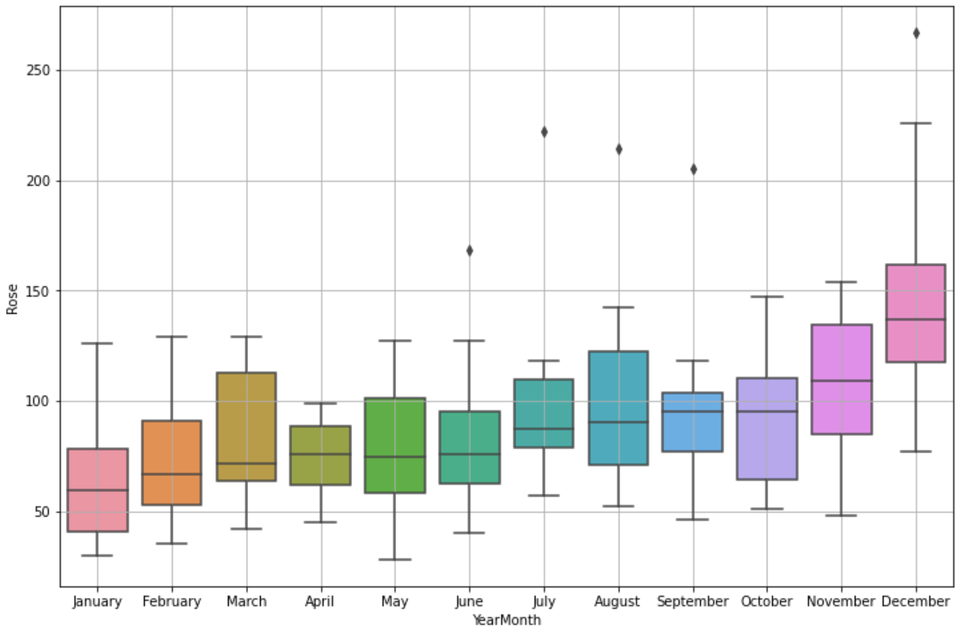
The trend seems to be declining. There is seasonality it seems to be dependent on the trend



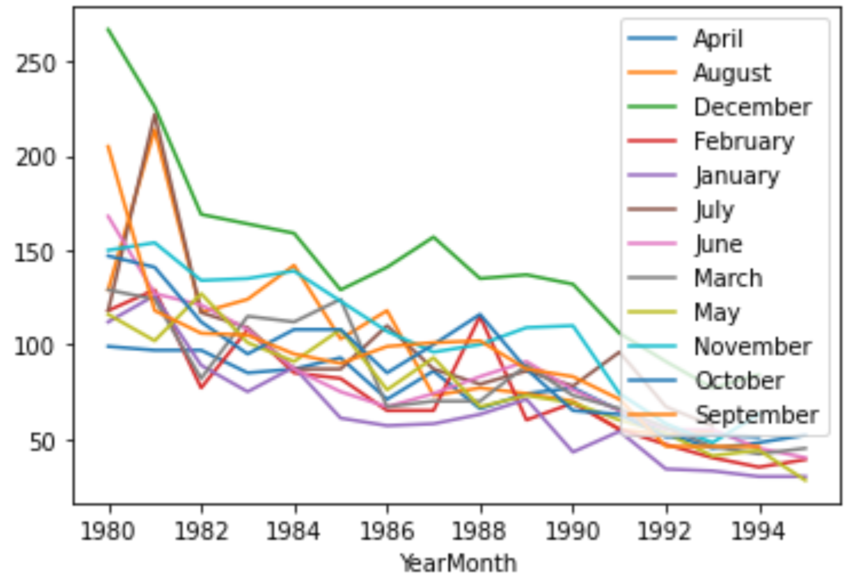
Decreasing trend



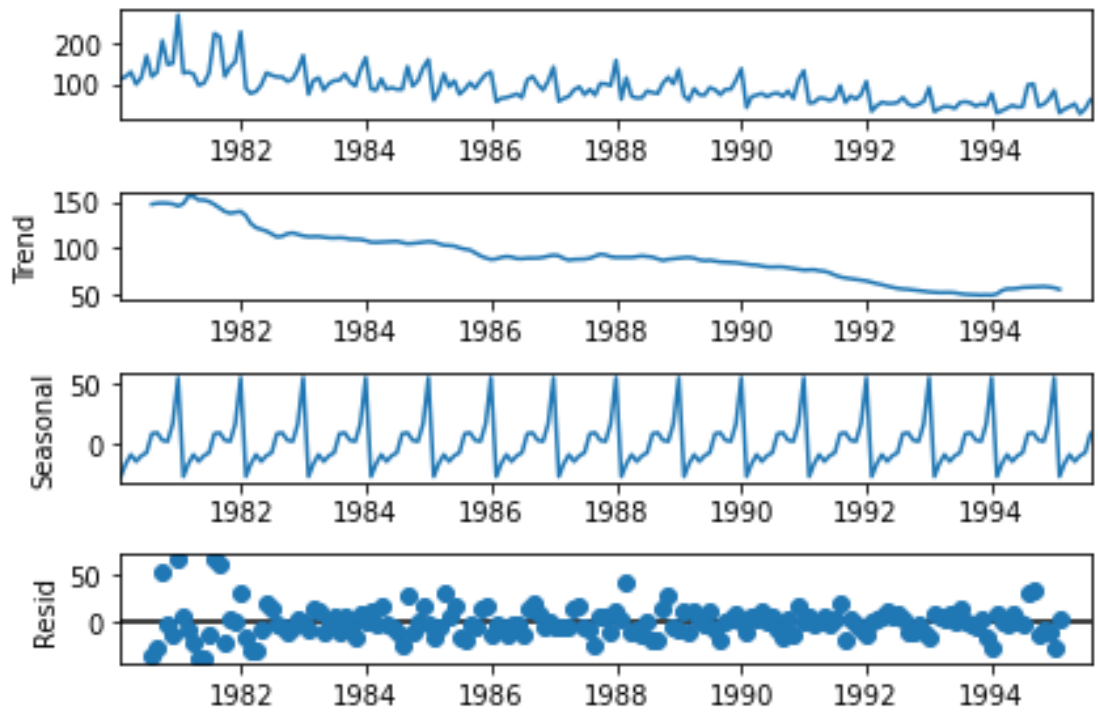
The year wise box plot shows the trend diminishing



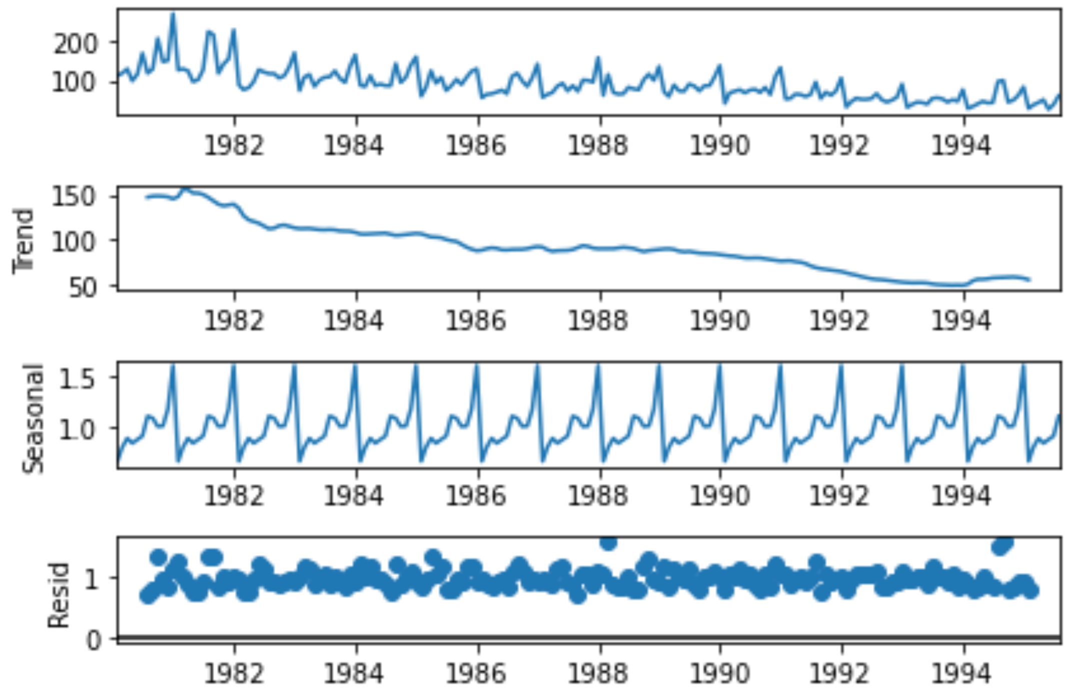
Month wise shows highest sale in December



October and January are at least sales



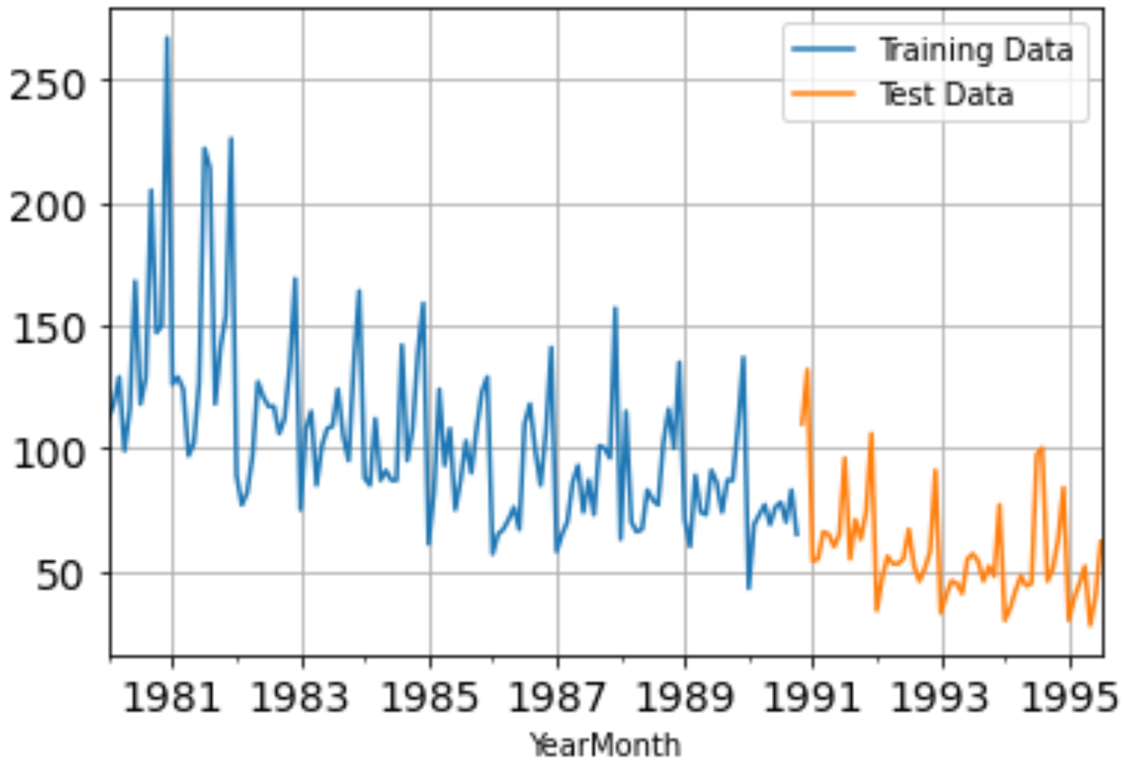
Additive decomposition shows and resid around the center line . it looks random



TREND IN SALES OF ROSE IS CONTINUOUSLY DECREASING OVER THE PERIOD. DETAILED STUDY MAY BE REQUIRED TO SEE WHETHER DECREASING TREND IS DUE TO CHANGE IN CUSTOMER PREFERENCE OR DUE TO SUBSTITUTION. SEASONALITY OF SALES IS OBSERVED, AND HIGHER SALES IS MAINTAINED IN THE END OF THE YEAR. SOME PROMOTION SCHEMES AND IMPROVEMENT / QUALITY ENHANCERS IN THE PRODUCT CAN BE EXAMINED SO AS TO ATTRACT NEW YOUNG GENERATION CUSTOMERS

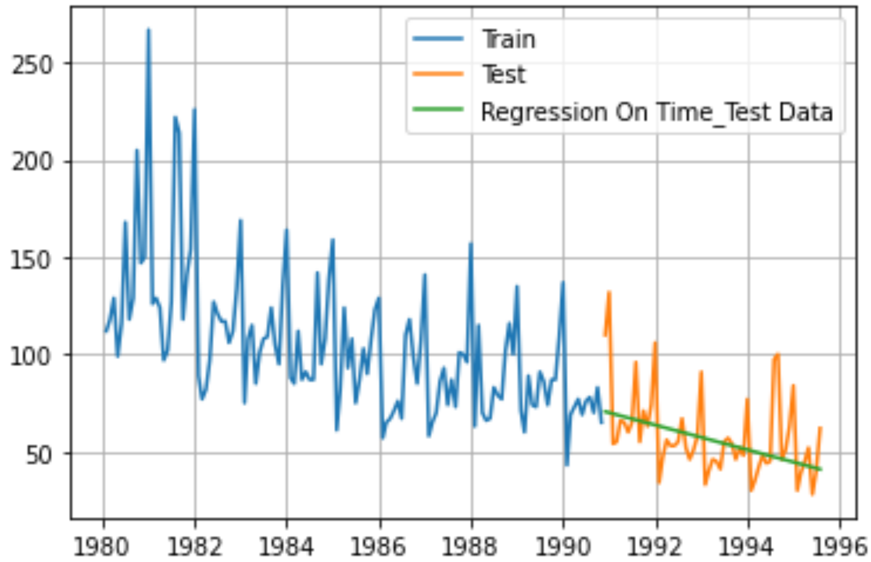
DATA PRE-PROCESSING

DIVINDING THE DATA INTO 70% TRAIN DATA AND 30% TEST DATA
TRAIN DATA : (130, 1)
TEST DATA (57, 1)



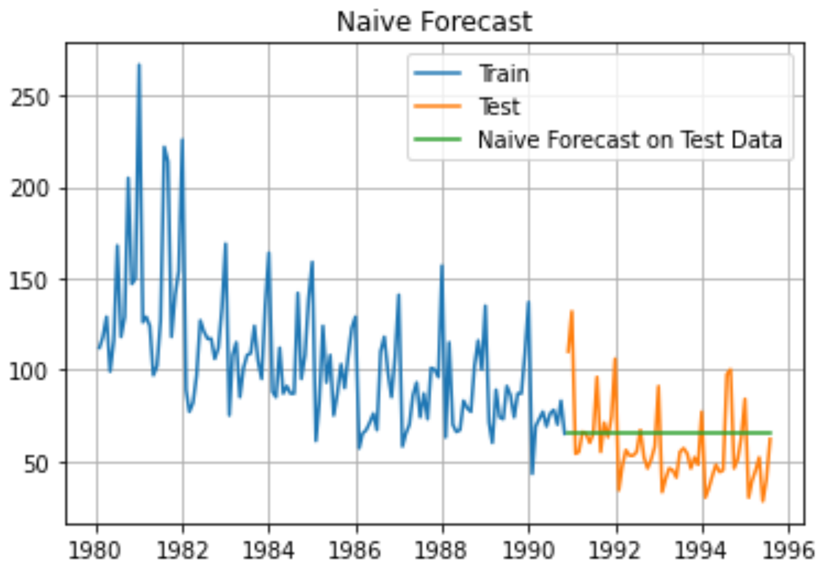
MODEL BUILDING - ORIGINAL DATA

LINEAR REGRESSION



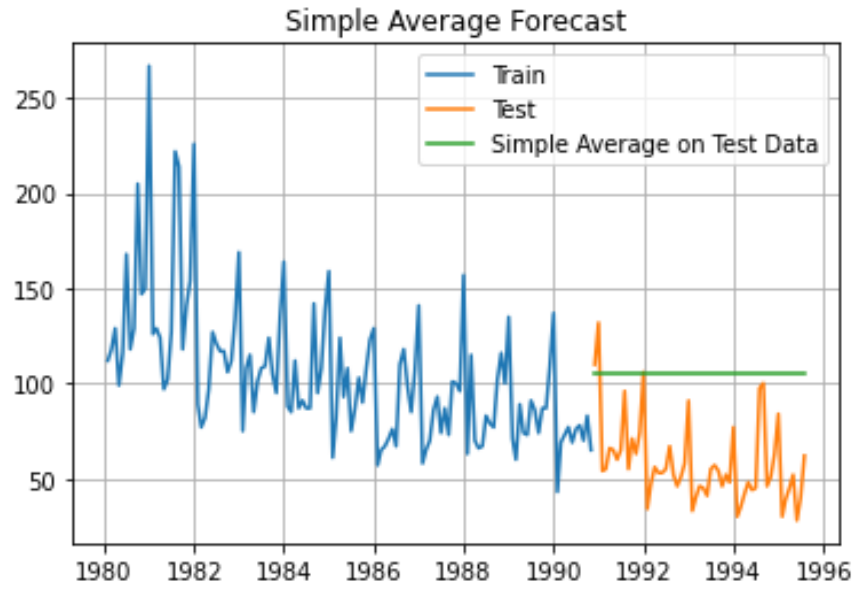
For RegressionOnTime forecast on the Test Data, RMSE is 19.864

NAIVE APPROACH



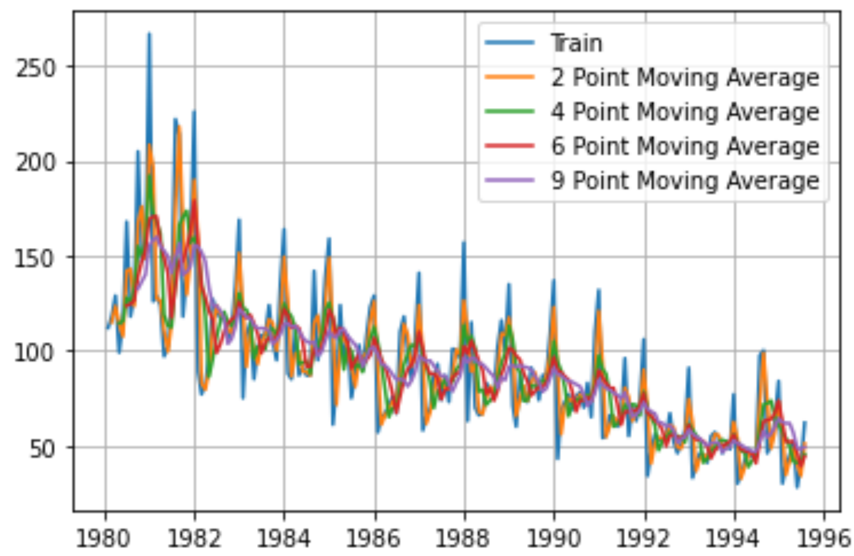
For RegressionOnTime forecast on the Test Data, RMSE is 22.374

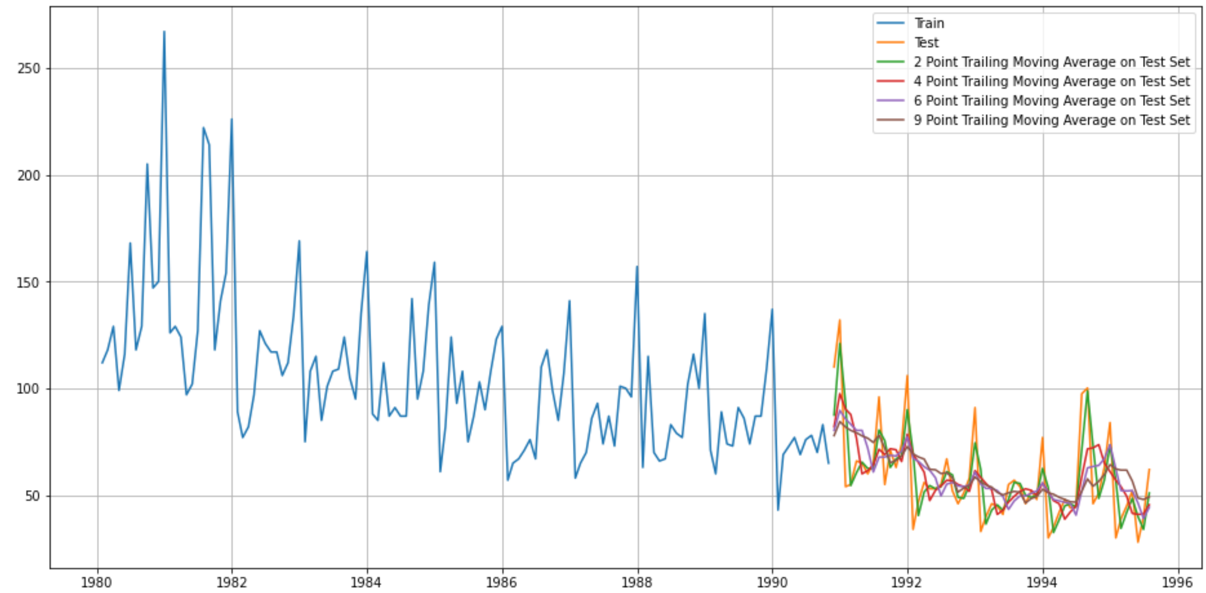
SIMPLE AVERAGE



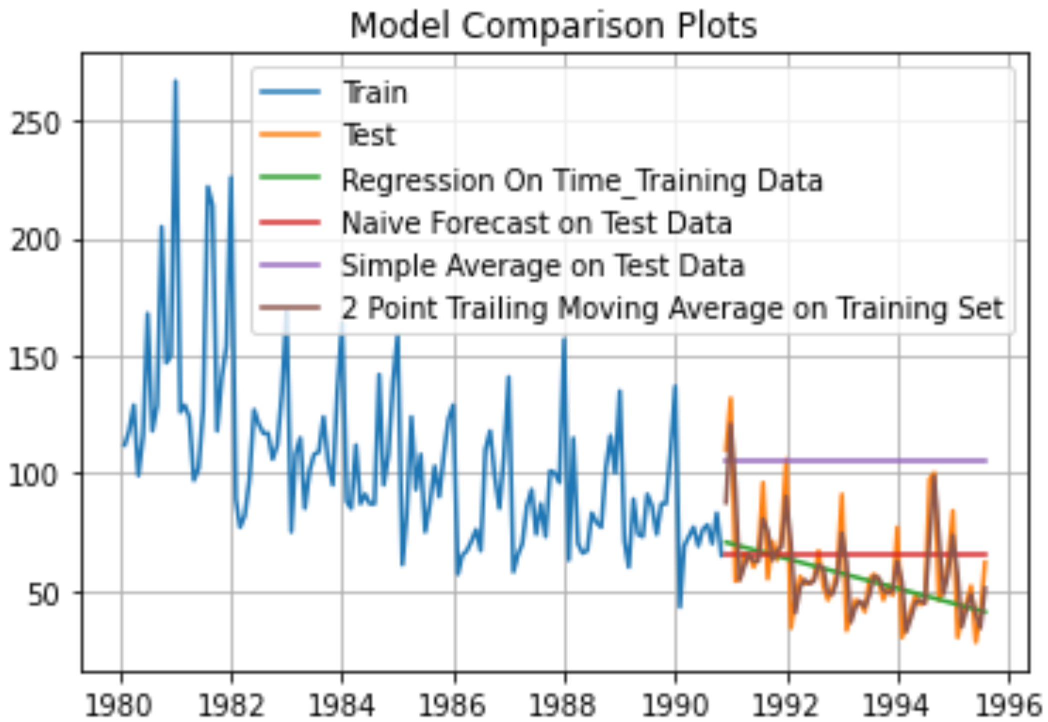

```
For Simple Average forecast on the Test Data,  RMSE is 51.238
```

MOVING AVERAGE(MA)





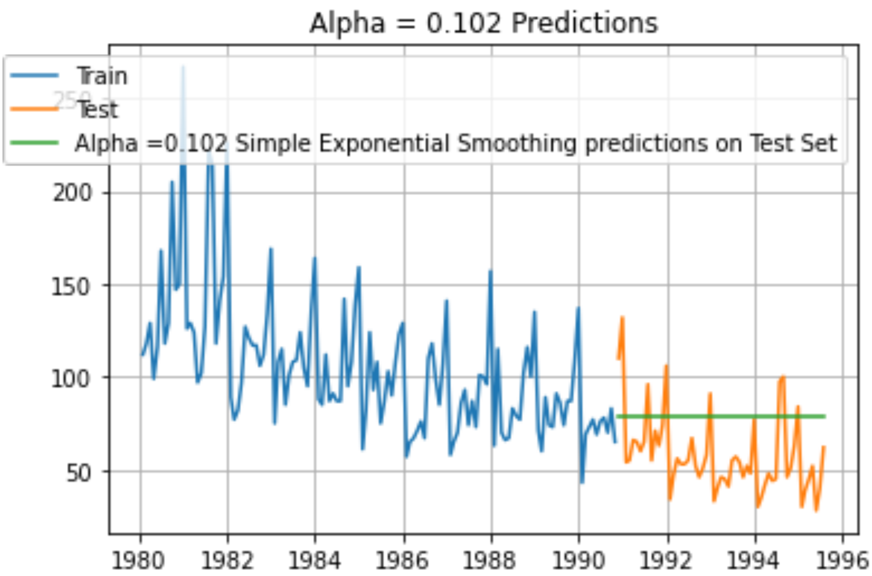

```
For 2 point Moving Average Model forecast on the Training Data,  RMSE is 12.815
For 4 point Moving Average Model forecast on the Training Data,  RMSE is 17.184
For 6 point Moving Average Model forecast on the Training Data,  RMSE is 17.725
For 9 point Moving Average Model forecast on the Training Data,  RMSE is 18.557
```



The 2 point moving average model gives the best fit predictions.

SIMPLE EXPONENTIAL SMOOTHENING MODELS

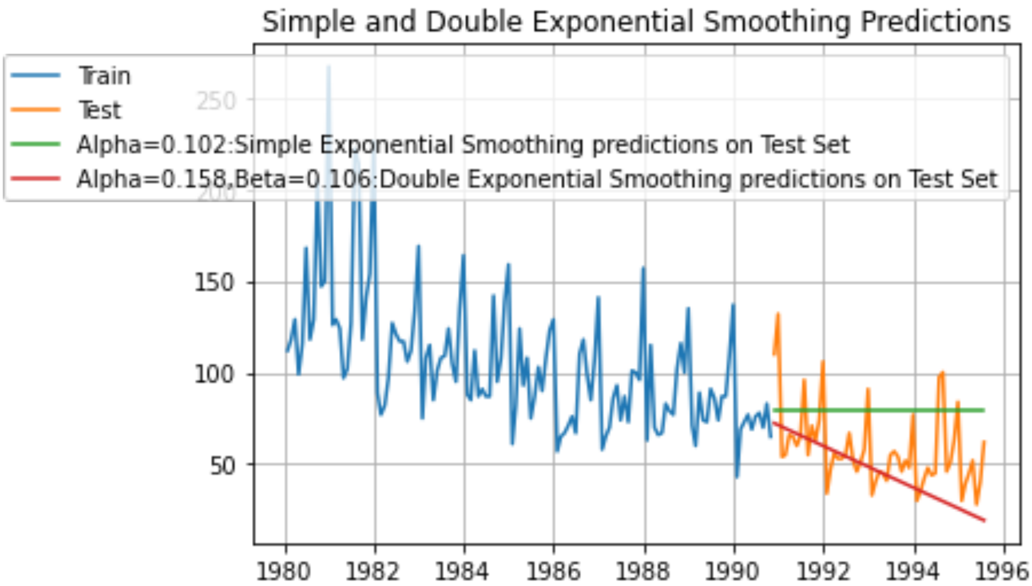
```
'smoothing_level': 0.10272100493683696,
 'smoothing_trend': nan,
 'smoothing_seasonal': nan,
 'damping_trend': nan,
 'initial_level': 134.26284082377956,
 'initial_trend': nan,
 'initial_seasons': array([], dtype=float64),
 'use_boxcox': False,
 'lamda': None,
 'remove_bias': False}
```




```
SES RMSE: 29.768404800540043
SES RMSE (calculated using statsmodels): 29.768404800540043
```

DOUBLE EXPONENTIAL SMOOTHING

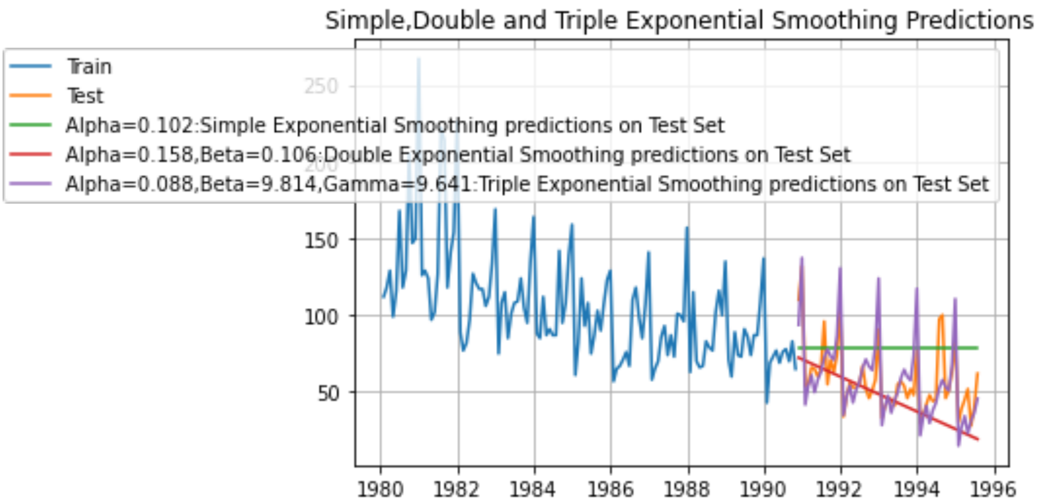
```
{'smoothing_level': 0.15849768299408634,
'smoothing_trend': 0.10653336272408515,
'smoothing_seasonal': nan,
'damping_trend': nan,
'initial_level': 131.3618884097747,
'initial_trend': 1.4806378622676117,
'initial_seasons': array([], dtype=float64),
'use_boxcox': False,
'lamda': None,
'remove_bias': False}
```




```
DES RMSE: 24.39684843587642
```

HOLT-WINTERS - ETS(A, A, A) - HOLT WINTER'S LINEAR METHOD WITH ADDITIVE ERRORS

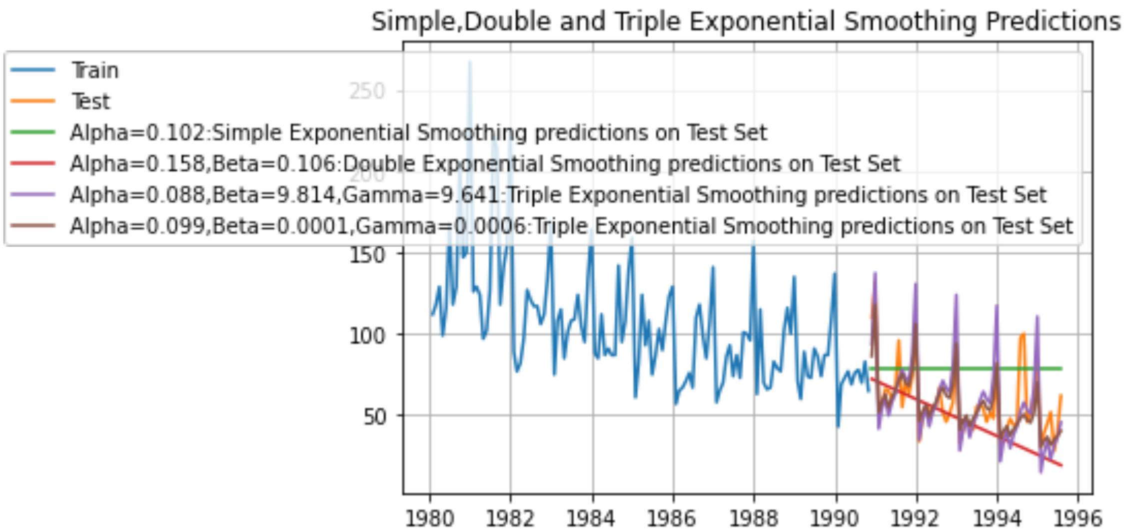
```
{'smoothing_level': 0.08829506527539115,
'smoothing_trend': 9.814104658031581e-07,
'smoothing_seasonal': 9.641663991608711e-05,
'damping_trend': nan,
'initial_level': 146.93438103038488,
'initial_trend': -0.5564380117261886,
'initial_seasons': array([-31.14676036, -18.77142217, -10.75975746, -21
.38605299,
       -12.55715163,  -7.08670312,   2.83633991,   8.92770254,
         4.94979556,   3.04596856,  19.68667018,  63.90465032]),
'use_boxcox': False,
'lamda': None,
'remove_bias': False}
```



```
TES RMSE: 16.074613084967055
```

HOLT-WINTERS - ETS(A, A, M) - HOLT WINTER'S LINEAR METHOD

```
{'smoothing_level': 0.09929582516403294,
'smoothing_trend': 0.00011579899094879335,
'smoothing_seasonal': 0.0006694995284830673,
'damping_trend': nan,
'initial_level': 144.6150833355538,
'initial_trend': -0.5843194599617058,
'initial_seasons': array([0.75819946, 0.85683872, 0.93677882, 0.8214941
2, 0.922274  ,
       0.99357229, 1.09383301, 1.16982102, 1.09518559, 1.07713919,
       1.23853049, 1.70793685]),
'use_boxcox': False,
'lamda': None,
'remove_bias': False}
```



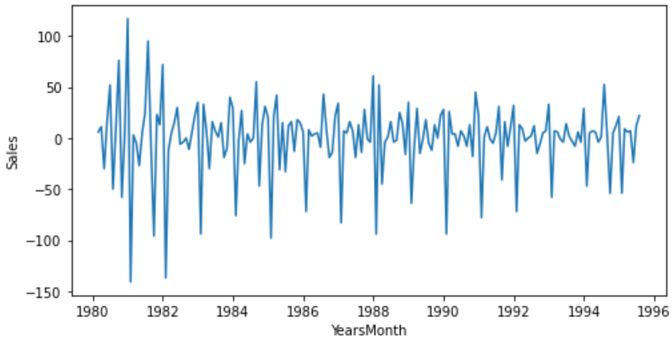

```
TES_am RMSE: 13.113063349263944
```

This method gives the best fit model to predict the data

CHECK FOR STATIONARITY

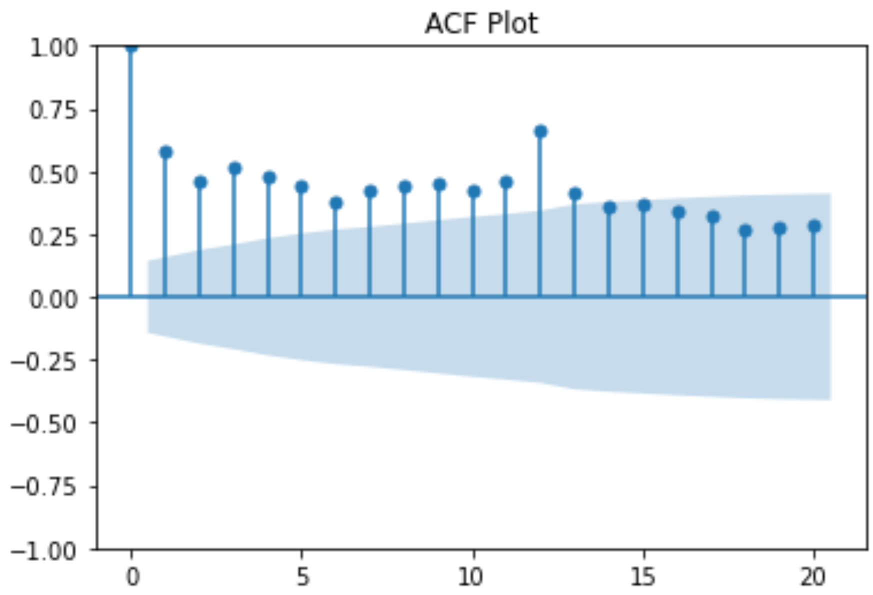
```
ADF Statistic: -1.9375726417334478
p-value: 0.314597330866447523
Critical Values:
	1%: -3.4687256239864017
	5%: -2.8783961376954363
	10%: -2.57575634100705
Data made stationary.
```

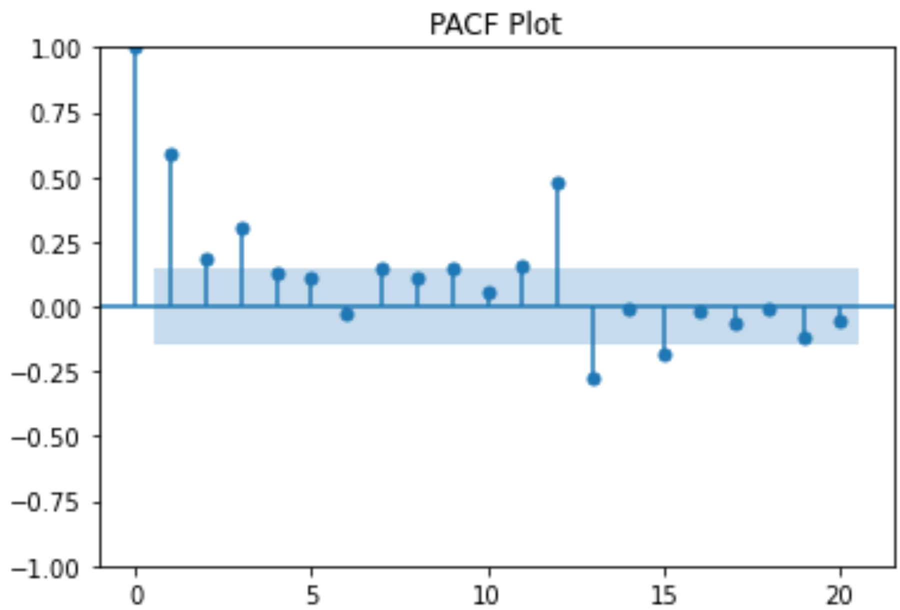
THE DATA WAS NOT STATIONARY
WE DID SO BY DECOMPOSING THE DATA TO MAKE IT SO



D=1
DECOMPOSED DATA

MODEL BUILDING - STATIONARY DATA

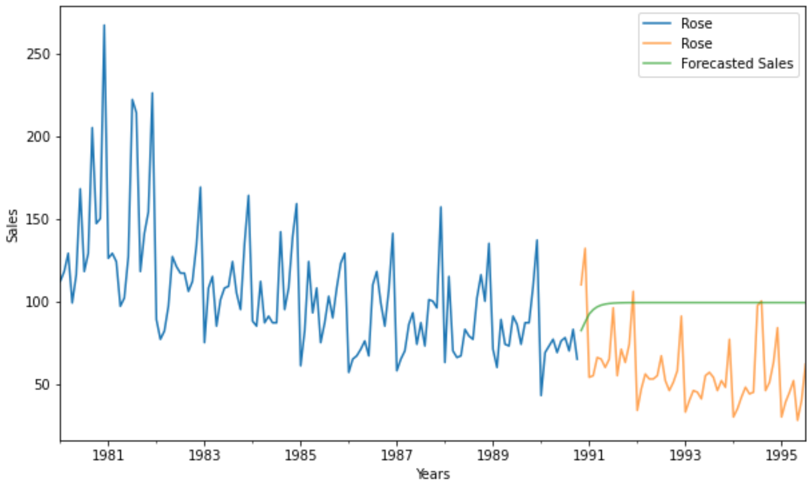




AR MODEL

```
ARIMA(1, 0, 0) - AIC:-172.2170806650721
ARIMA(2, 0, 0) - AIC:-172.32137910312673
ARIMA(3, 0, 0) - AIC:-176.7712183546709
```

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                   Rose   No. Observations:                  130
Model:                 ARIMA(2, 0, 0)   Log Likelihood                  90.161
Date:                Sun, 17 Mar 2024   AIC                           -172.321
Time:                        22:08:45   BIC                           -160.851
Sample:                    01-31-1980   HQIC                          -167.661
                         - 10-31-1990
Covariance Type:                  opg
==============================================================================
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const          1.9965      0.027     73.544      0.000       1.943       2.050
ar.L1          0.3879      0.096      4.054      0.000       0.200       0.575
ar.L2          0.1268      0.113      1.124      0.261      -0.094       0.348
sigma2         0.0146      0.001      9.930      0.000       0.012       0.017
```



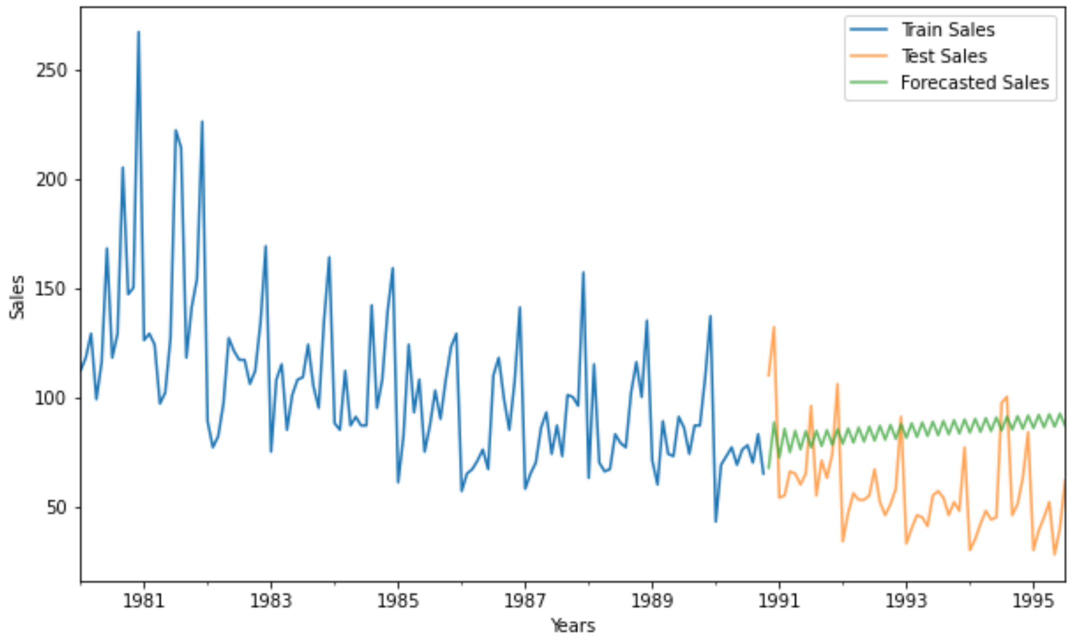
Best AR Model : ARIMA(2,0,0) RMSE 46.338452

ARMA MODEL

```
ARIMA(1, 0, 1) - AIC:-182.5073594926111
ARIMA(1, 0, 2) - AIC:-184.5584641108209
ARIMA(1, 0, 3) - AIC:-183.11119873629906
ARIMA(2, 0, 1) - AIC:-183.97687528082542
ARIMA(2, 0, 2) - AIC:-183.88805168984663
ARIMA(2, 0, 3) - AIC:-181.97663823278842
ARIMA(3, 0, 1) - AIC:-182.62430161349243
```

```
ARIMA(3, 0, 2) - AIC:-181.98502313762043
ARIMA(3, 0, 3) - AIC:-187.94781349342395
```

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                   Rose   No. Observations:                  130
Model:                 ARIMA(3, 0, 3)   Log Likelihood                 101.974
Date:                Sun, 17 Mar 2024   AIC                           -187.948
Time:                        22:10:41   BIC                           -165.008
Sample:                    01-31-1980   HQIC                          -178.626
                         - 10-31-1990
Covariance Type:                  opg
==============================================================================
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const          2.0085      0.093     21.570      0.000       1.826       2.191
ar.L1         -0.6625      0.132     -5.008      0.000      -0.922      -0.403
ar.L2          0.9775      0.037     26.747      0.000       0.906       1.049
ar.L3          0.6440      0.123      5.227      0.000       0.403       0.885
ma.L1          1.0333      0.110      9.422      0.000       0.818       1.248
ma.L2         -0.7871      0.144     -5.459      0.000      -1.070      -0.504
ma.L3         -0.8309      0.076    -10.977      0.000      -0.979      -0.683
sigma2         0.0120      0.002      6.955      0.000       0.009       0.015
==============================================================================
```



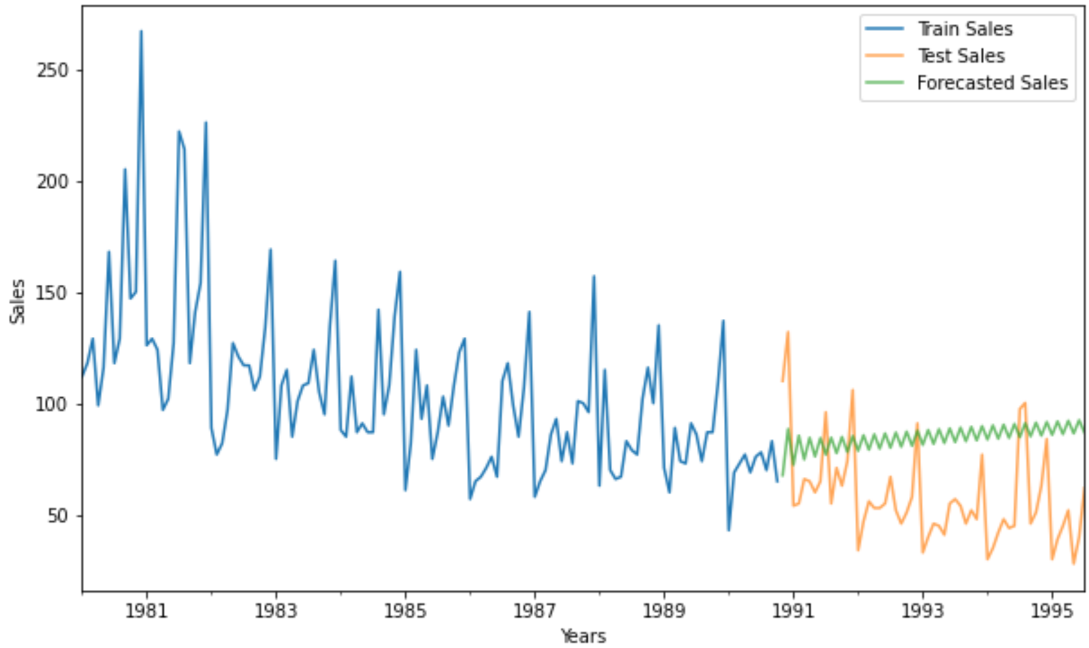
Best ARMA Model : ARIMA(3,0,3) RMSE : 34.892298

ARIMA MODEL

```
ARIMA(1, 0, 1) - AIC:-182.5073594926111
ARIMA(1, 0, 2) - AIC:-184.5584641108209
ARIMA(1, 0, 3) - AIC:-183.11119873629906
ARIMA(1, 1, 1) - AIC:-185.8390579435197
ARIMA(1, 1, 2) - AIC:-185.77925788107757
ARIMA(1, 1, 3) - AIC:-183.86300088764386
```

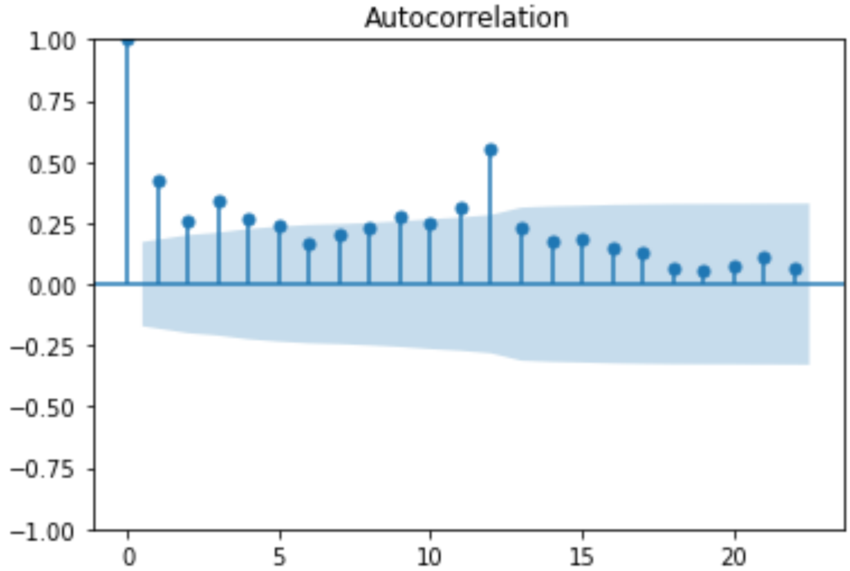
```
ARIMA(2, 0, 1) - AIC:-183.97687528082542
ARIMA(2, 0, 2) - AIC:-183.88805168984663
ARIMA(2, 0, 3) - AIC:-181.97663823278842
ARIMA(2, 1, 1) - AIC:-184.5190201812606
ARIMA(2, 1, 2) - AIC:-183.8724738652746
ARIMA(2, 1, 3) - AIC:-181.884622561631
ARIMA(3, 0, 1) - AIC:-182.62430161349243
ARIMA(3, 0, 2) - AIC:-181.98502313762043
ARIMA(3, 0, 3) - AIC:-187.94781349342395
ARIMA(3, 1, 1) - AIC:-182.97828646851366
ARIMA(3, 1, 2) - AIC:-181.93030788403888
ARIMA(3, 1, 3) - AIC:-184.1021669626872
```

```
SARIMAX Results
==============================================================================
Dep. Variable:                   Rose   No. Observations:                  130
Model:                 ARIMA(3, 0, 3)   Log Likelihood                 101.974
Date:                Sun, 17 Mar 2024   AIC                           -187.948
Time:                        22:11:18   BIC                           -165.008
Sample:                    01-31-1980   HQIC                          -178.626
                         - 10-31-1990
Covariance Type:                  opg
==============================================================================
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const          2.0085      0.093     21.570      0.000       1.826       2.191
ar.L1         -0.6625      0.132     -5.008      0.000      -0.922      -0.403
ar.L2          0.9775      0.037     26.747      0.000       0.906       1.049
ar.L3          0.6440      0.123      5.227      0.000       0.403       0.885
ma.L1          1.0333      0.110      9.422      0.000       0.818       1.248
ma.L2         -0.7871      0.144     -5.459      0.000      -1.070      -0.504
ma.L3         -0.8309      0.076    -10.977      0.000      -0.979      -0.683
sigma2         0.0120      0.002      6.955      0.000       0.009       0.015
```

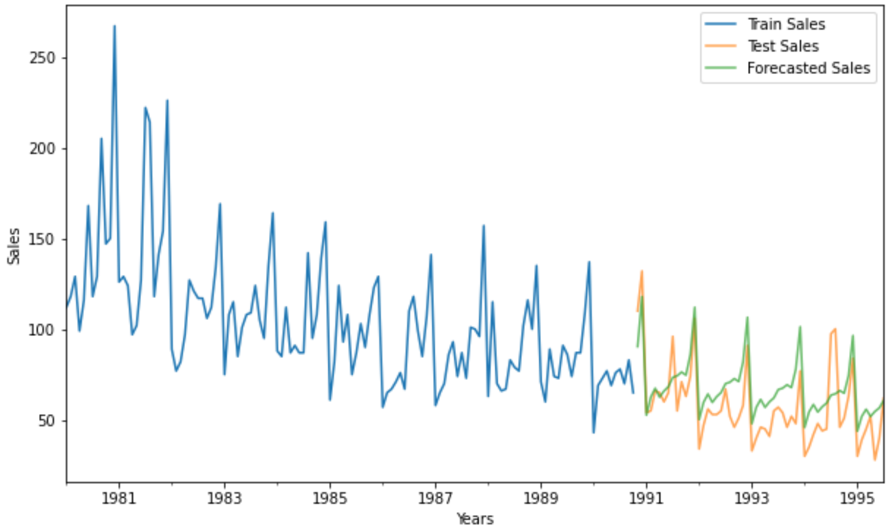


Best ARIMA Model : ARIMA(3,0,3) RMSE: 34.892298

SARIMA Model




```
                                      SARIMAX Results
==========================================================================================
Dep. Variable:                                 Rose   No. Observations:                130
Model:             SARIMAX(1, 0, 1)x(1, 0, 1, 12)   Log Likelihood                 135.117
Date:                              Sun, 17 Mar 2024   AIC                           -260.235
Time:                                      22:29:25   BIC                           -245.897
Sample:                                  01-31-1980   HQIC                          -254.409
                                       - 10-31-1990
Covariance Type:                                opg
==========================================================================================
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.9988      0.004    253.573      0.000       0.991       1.006
ma.L1         -0.8947      0.049    -18.180      0.000      -0.991      -0.798
ar.S.L12       0.9943      0.014     71.273      0.000       0.967       1.022
ma.S.L12      -0.8643      0.162     -5.341      0.000      -1.181      -0.547
sigma2         0.0058      0.001      6.201      0.000       0.004       0.008
```

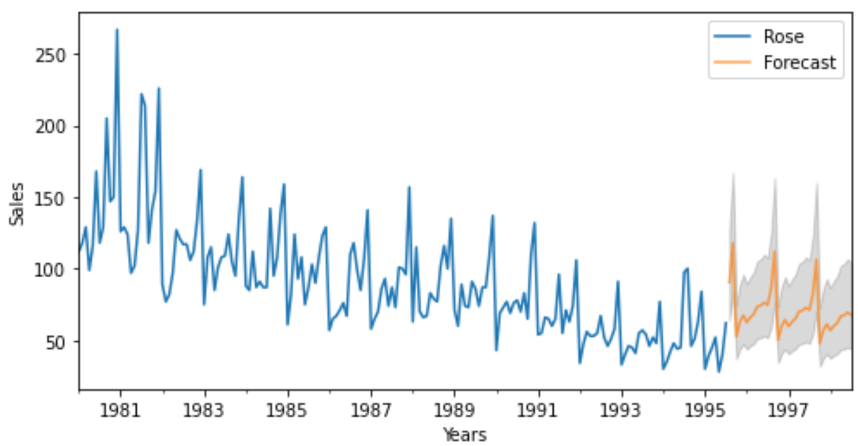


Best SARIMA Model : SARIMAX(1, 0, 1)x(1, 0, 1, 12) RSME: 16.172475

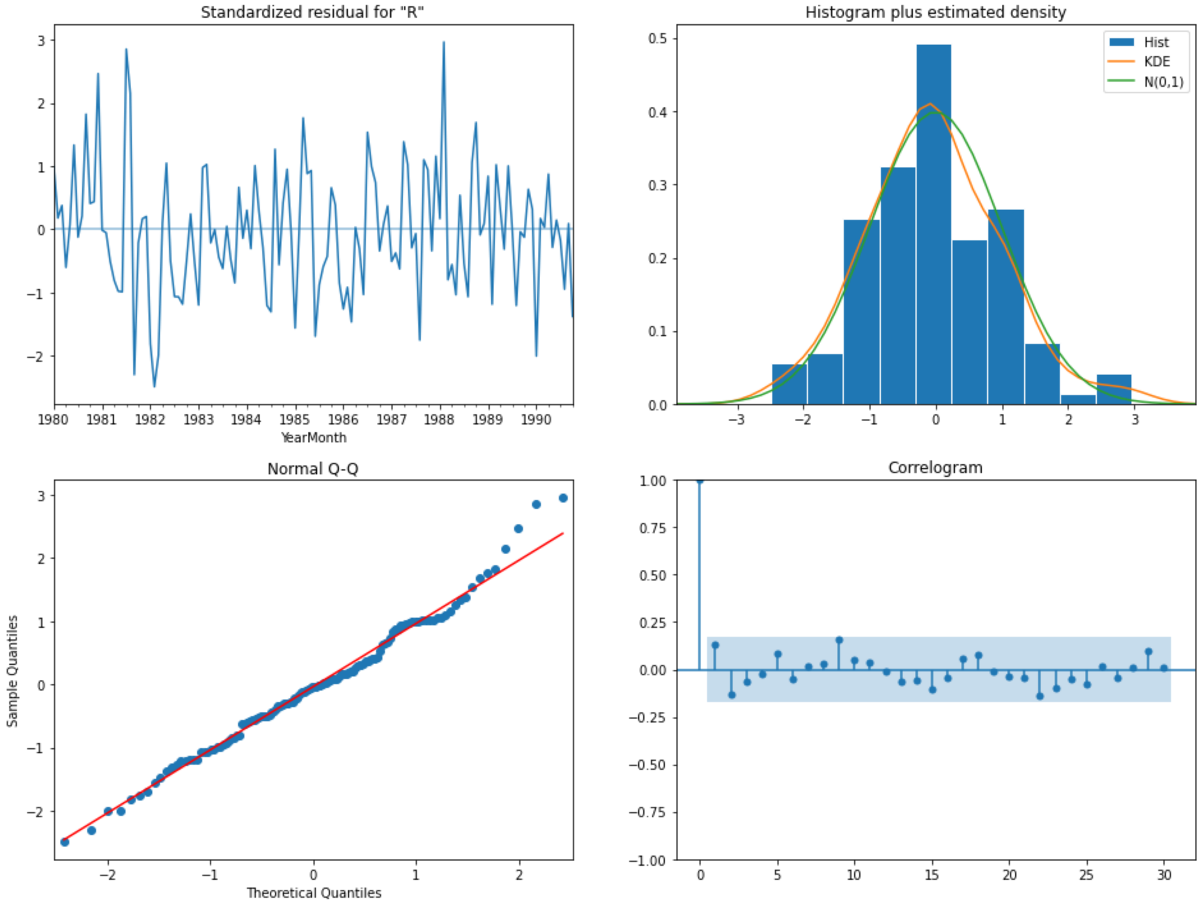
COMPARE THE PERFORMANCE OF THE MODELS

	Test RMSE
RegressionOnTime	19.864009
NaiveModel	22.374447
SimpleAverageModel	51.238342
2pointTrailingMovingAverage	12.815041
4pointTrailingMovingAverage	17.184213
6pointTrailingMovingAverage	17.724516
9pointTrailingMovingAverage	18.556574
Alpha=0.102,SES	29.768405
Alpha=0.158,Beta=0.106:DES	24.396848
Alpha=0.088,Beta=9.814,Gamma=9.641:TES	16.074613
Alpha=0.099,Beta=0.0001,Gamma=0.0006,Gamma=0:TES	13.113063
Best AR Model : ARIMA(2,0,0)	46.338452
Best ARMA Model : ARIMA(3,0,3)	34.892298
Best ARIMA Model : ARIMA(3,0,3)	34.892298
Best SARIMA Model : SARIMAX(1, 0, 1)x(1, 0, 1, 12)	16.172475

The above results show that in original data modelling 2 point moving average is the best fit
TES_am is the best exponential smoothing model
And with stationary data SARIMA model fits best



ACF and PACF for residuals of ARIMA model to ensure no more information is left for extraction



ACTIONABLE INSIGHTS & RECOMMENDATIONS

The analysis of the wine sales data indicates a clear downward trend for the Rose wine variety for the company, which has been declining in popularity for more than a decade.

This trend is expected to continue in the future as well, based on the predictions of the most optimal model.

Wine sales are highly influenced by seasonal changes, with sales increasing during festival season and dropping during peak winter time i.e. January.

The company should consider running campaigns to boost the consumption of the wine during the rest of the year, as sales are subdued during this period.

Campaigns during the lean period (April to June) might yield maximum results for the company, as sales are low during this period, and boosting them would increase the overall performance of the wine in the market across the year.

Running campaigns during peak periods (such as during festivals) might not generate significant impact on sales, as they are already high during this time of the year.

Campaigns during peak winter time (January) are not recommended as people are less likely to purchase wine due to climatic reasons, and running campaigns during this period may not change people's opinion.

The company should also consider exploring reasons behind the decline in popularity of the Rose wine variety, and if needed, revamp its production and marketing strategies to regain the market share.