

Sparkling Wine Sales

Forecast

Presented by :

Sanjay Rajan J

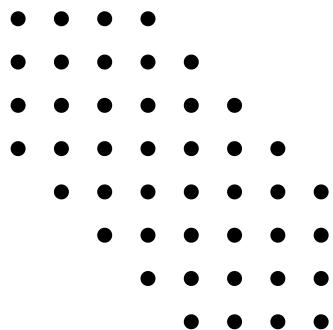


TABLE OF CONTENTS

CHAPTER NO	CONTENT	PAGE NO
	LIST OF FIGURES	4
1.	DATA OVERVIEW	6
2.	EXPLORATORY DATA ANALYSIS	8
	2.1 TIME SERIES DATA PLOTS	8
	2.1 TIME SERIES DECOMPOSITION	12
3.	DATA PRE-PROCESSING	14
4.	MODEL BUILDING – ORIGINAL DATA	15
	4.1 LINEAR REGRESSION	15
	4.2 SIMPLE AVERAGE	16
	4.3 MOVING AVERAGE	16
	4.4 SIMPLE EXPONENTIAL MODEL	17
	4.5 DOUBLE EXPONENTIAL MODEL	18
	4.6 TRIPLE EXPONENTIAL MODEL	18
	4.7 MODEL PERFORMANCE EVALUATION	20
5.	CHECK FOR STATIONARITY	22
6.	MODEL BUILDING – STATIONARY DATA	25
	6.1 ACF AND PACF PLOTS	25
	6.2 AUTO ARIMA	26

	6.3 MANUAL ARIMA MODEL	27
	6.4 AUTO SARIMA MODEL	28
	6.5 MANUAL SARIMA MODEL	30
	6.6 MODEL PERFORMANCE EVALUATION	32
7.	MODEL PERFORMANCE COMPARISON AND FINAL MODEL SELECTION FOR FORECAST	33
8.	ACTIONABLE INSIGHTS & RECOMMENDATIONS	36

LIST OF FIGURES

FIG NO.	NAME	PAGE
1.	Basic info	7
2.	First 5 rows	7
3.	Time Series Plot	7
4.	Numerical Statistics	8
5.	Sales Pivot Table	8
6.	Yearly boxplot	9
7.	Monthly boxplot	9
8.	Month plot	10
9.	Monthly sales across years (plot)	10
10.	Average Sales and Percent change	11
11.	Empirical Cumulative Distribution	11
12.	Additive Decomposition Plot	12
13.	Multiplicative Decomposition Plot	12
14.	Train-Test Split	14
15.	Visualisation of train-test split	14
16.	Regression Model Performance	15
17.	Simple Average Model Performance	16
18.	Moving Average Models Performance	16
19.	SES model Performance	17
20.	DES model Performance	18
21.	TES model 1 Performance	18
22.	TES model 2 Performance	19
23.	Model Comparison plot (Regression, Simple average, Moving average)	20
24.	Model Comparison plot (Exponential Smoothing Models)	20

25.	RMSE Scores of all models	21
26.	Stationarity check (whole data)	22
27.	Stationarity check (Training data)	23
28.	ACF, PACF plots (Training data)	25
29.	Auto ARIMA Results	26
30.	Auto ARIMA model plot	26
31.	Manual ARIMA Results	27
32.	Manual ARIMA plot	27
33.	Auto SARIMA results	28
34.	Auto SARIMA diagnostics	29
35.	Auto SARIMA plot	29
36.	Manual SARIMA Results	30
37.	Manual SARIMA diagnostics	31
38.	Manual SARIMA plot	31
39.	ARIMA & SARIMA model performances	32
40.	ARIMA & SARIMA model predictions	32
41.	All model performances	33
42.	Triple Exponential Smoothing ($\alpha = 0.4$, $\beta = 0.1$ and $\gamma = 0.2$)	33
43.	Forecasted Results	34
44.	Forecast of the best model	35
45.	Plotting only Forecast results	35
46.	Time Series trend	36
47.	Monthly Sales	36

1. DATA OVERVIEW

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, analyse trends, patterns, and factors influencing wine sales over the course of 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 analyse 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.

- Shape: This dataset has 187 entries with dates ranging from Jan 1980 to July 1995
- Basic Info:

Fig.1 Basic info

```
<class 'pandas.core.frame.DataFrame'>
DatetimeIndex: 187 entries, 1980-01-31 to 1995-07-31
Freq: M
Data columns (total 1 columns):
 #   Column      Non-Null Count  Dtype  
--- 
 0   Sparkling    187 non-null    int64  
dtypes: int64(1)
memory usage: 2.9 KB
```

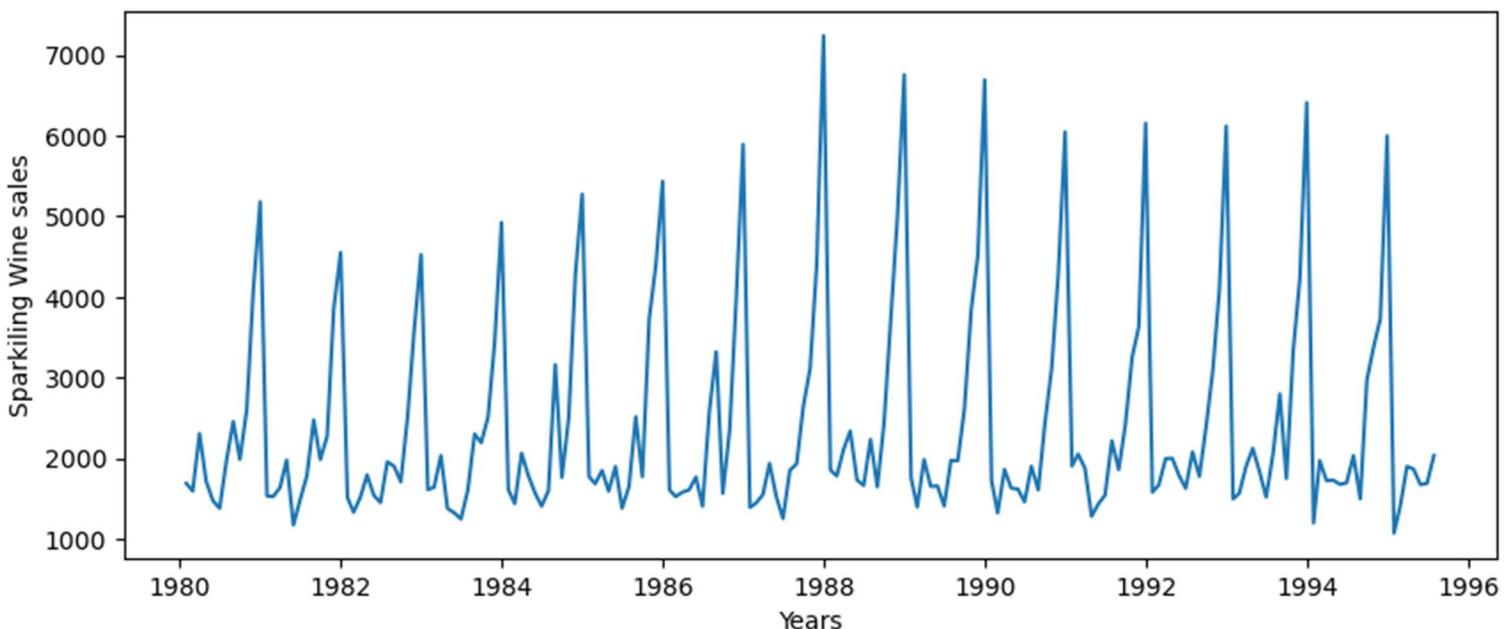
- First Few rows:

Fig.2 First 5 rows

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

- Plotting the Time Series:

Fig.3 Time Series Plot



2. EXPLORATORY DATA ANALYSIS

2.1 TIME SERIES DATA PLOTS:

Fig.4 Numerical Statistics

Sparkling	
count	187.000000
mean	2402.417112
std	1295.111540
min	1070.000000
25%	1605.000000
50%	1874.000000
75%	2549.000000
max	7242.000000

Fig.5 Sales Pivot Table

	Apr	Aug	Dec	Feb	Jan	Jul	Jun	Mar	May	Nov	Oct	Sep
1980	1712.0	2453.0	5179.0	1591.0	1686.0	1966.0	1377.0	2304.0	1471.0	4087.0	2596.0	1984.0
1981	1976.0	2472.0	4551.0	1523.0	1530.0	1781.0	1480.0	1633.0	1170.0	3857.0	2273.0	1981.0
1982	1790.0	1897.0	4524.0	1329.0	1510.0	1954.0	1449.0	1518.0	1537.0	3593.0	2514.0	1706.0
1983	1375.0	2298.0	4923.0	1638.0	1609.0	1600.0	1245.0	2030.0	1320.0	3440.0	2511.0	2191.0
1984	1789.0	3159.0	5274.0	1435.0	1609.0	1597.0	1404.0	2061.0	1567.0	4273.0	2504.0	1759.0
1985	1589.0	2512.0	5434.0	1682.0	1771.0	1645.0	1379.0	1846.0	1896.0	4388.0	3727.0	1771.0
1986	1605.0	3318.0	5891.0	1523.0	1606.0	2584.0	1403.0	1577.0	1765.0	3987.0	2349.0	1562.0
1987	1935.0	1930.0	7242.0	1442.0	1389.0	1847.0	1250.0	1548.0	1518.0	4405.0	3114.0	2638.0
1988	2336.0	1645.0	6757.0	1779.0	1853.0	2230.0	1661.0	2108.0	1728.0	4988.0	3740.0	2421.0
1989	1650.0	1968.0	6694.0	1394.0	1757.0	1971.0	1406.0	1982.0	1654.0	4514.0	3845.0	2608.0
1990	1628.0	1605.0	6047.0	1321.0	1720.0	1899.0	1457.0	1859.0	1615.0	4286.0	3116.0	2424.0
1991	1279.0	1857.0	6153.0	2049.0	1902.0	2214.0	1540.0	1874.0	1432.0	3627.0	3252.0	2408.0
1992	1997.0	1773.0	6119.0	1667.0	1577.0	2076.0	1625.0	1993.0	1783.0	4096.0	3088.0	2377.0
1993	2121.0	2795.0	6410.0	1564.0	1494.0	2048.0	1515.0	1898.0	1831.0	4227.0	3339.0	1749.0
1994	1725.0	1495.0	5999.0	1968.0	1197.0	2031.0	1693.0	1720.0	1674.0	3729.0	3385.0	2968.0
1995	1862.0	NaN	NaN	1402.0	1070.0	2031.0	1688.0	1897.0	1670.0	NaN	NaN	NaN

Fig.6 Yearly boxplot

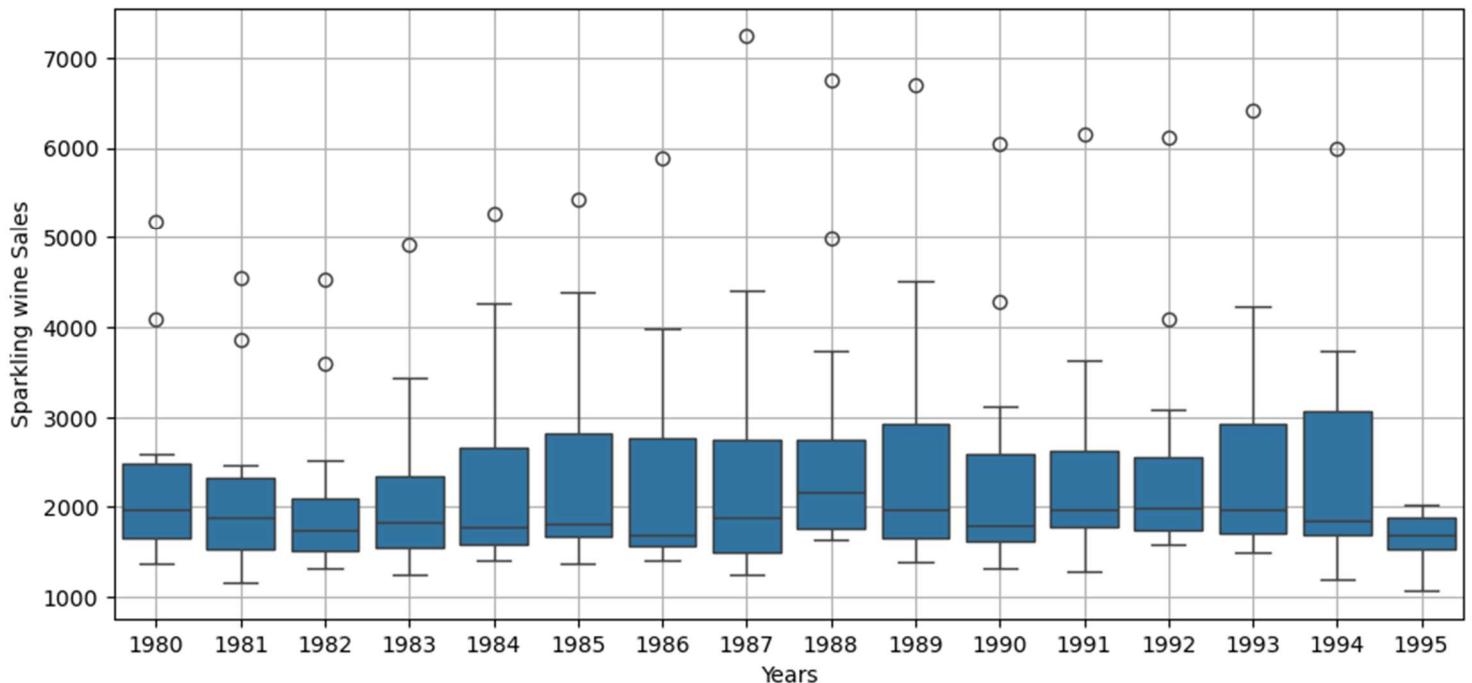


Fig.7 Monthly boxplot

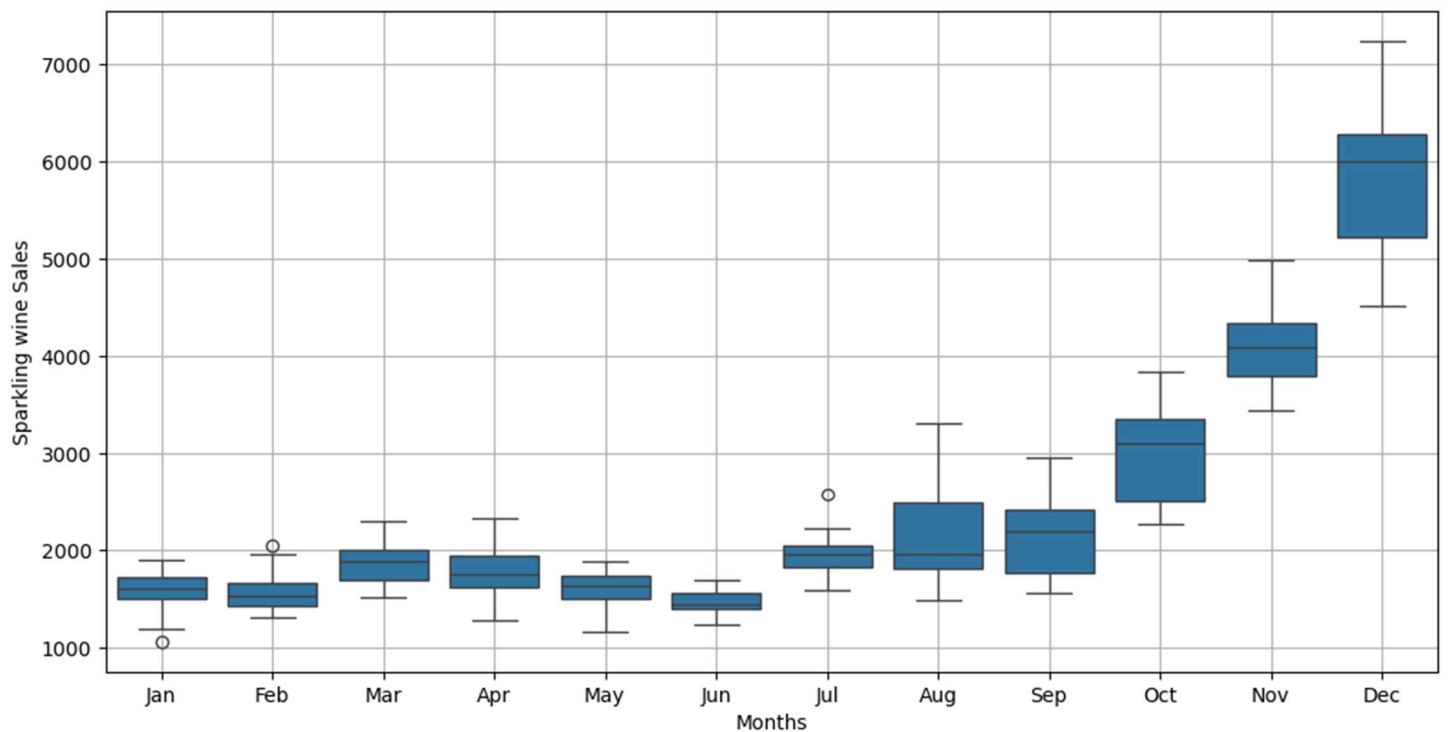


Fig.8 Month plot

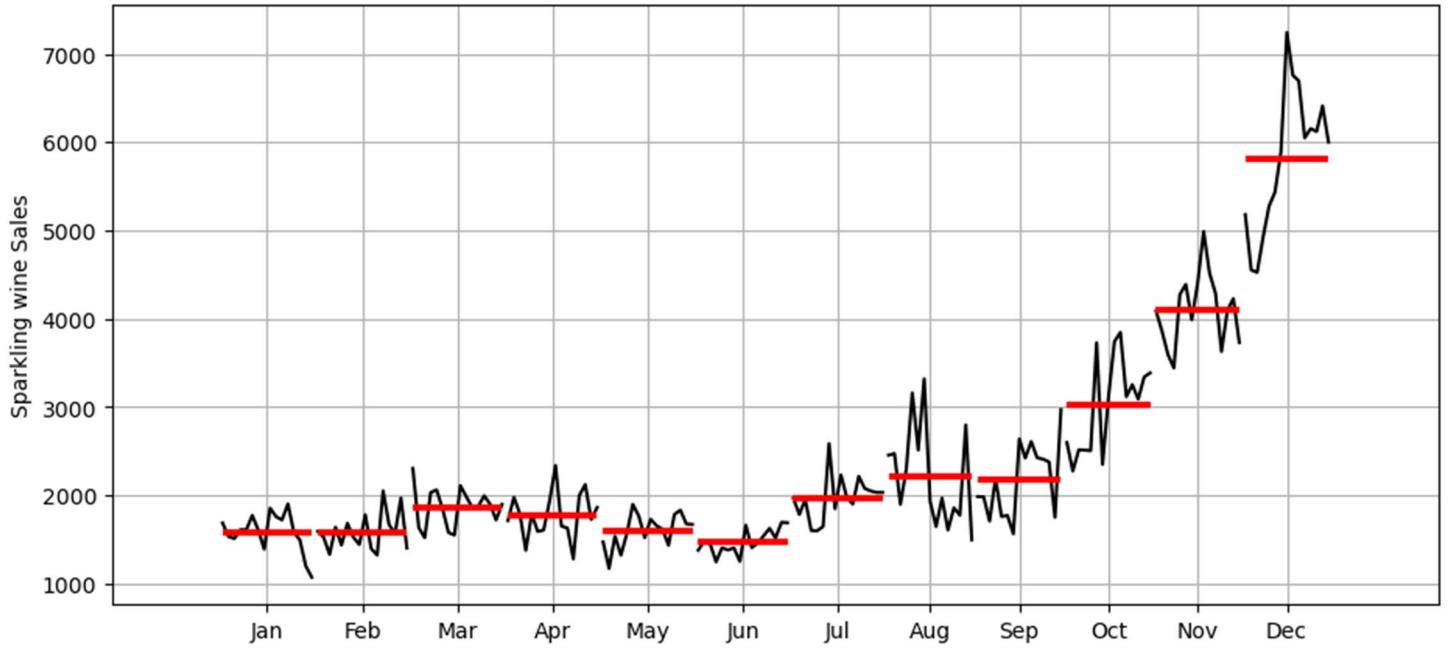


Fig.9 Monthly sales across years (plot)

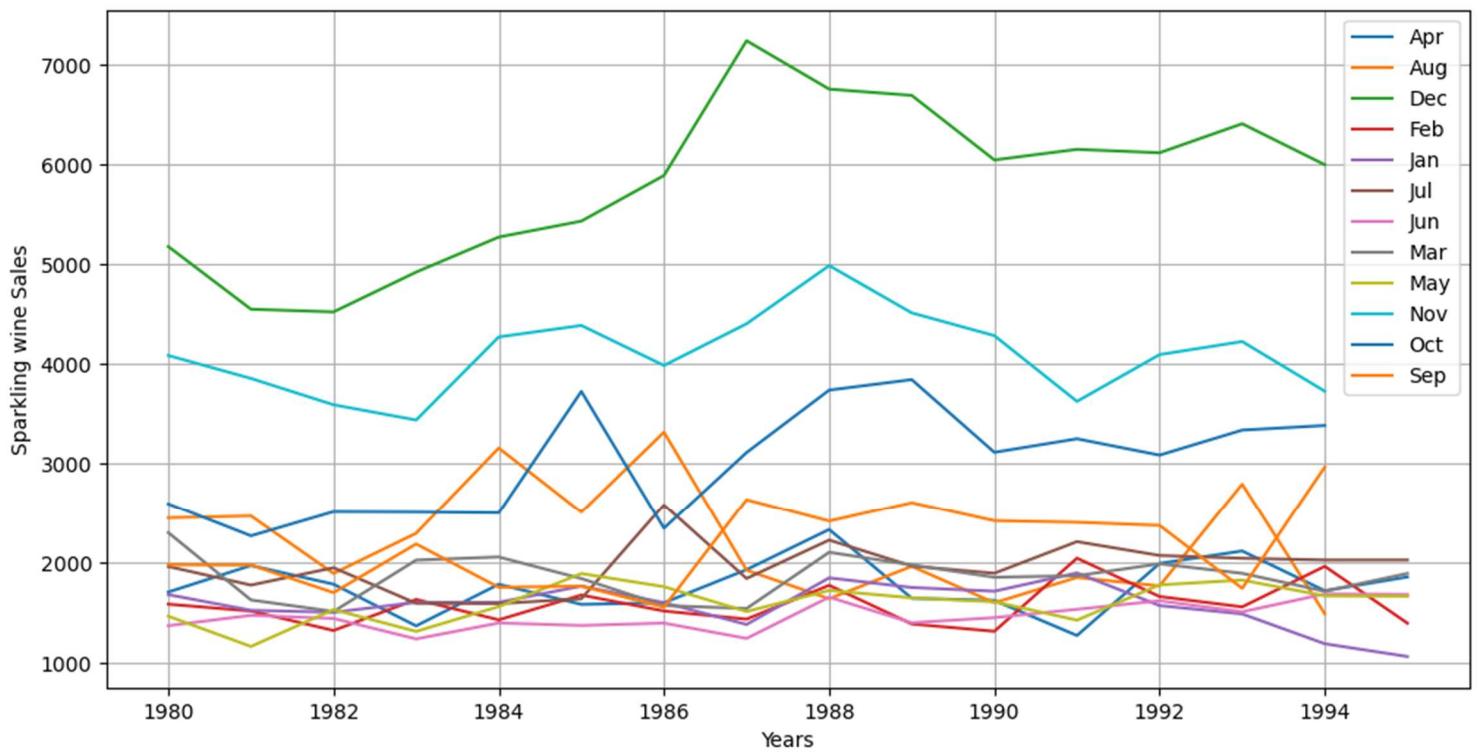


Fig.10 Average Sales and Percent change

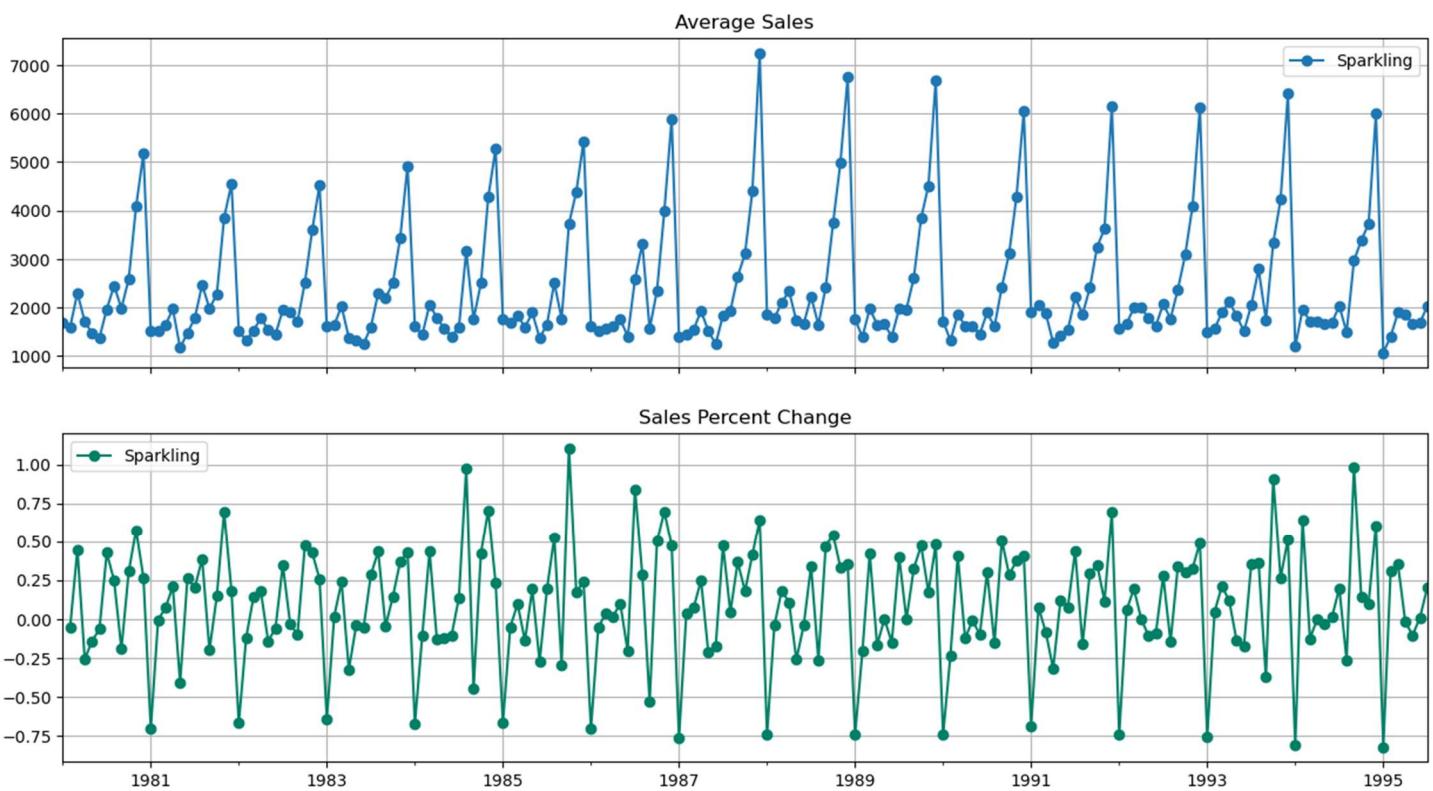
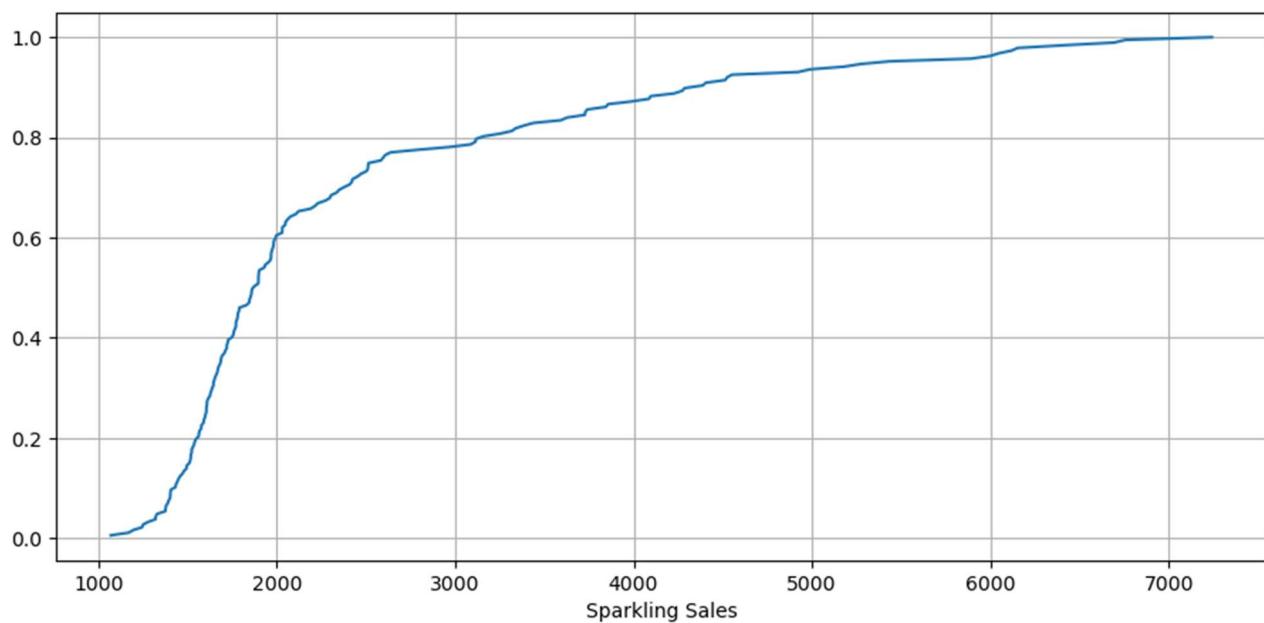


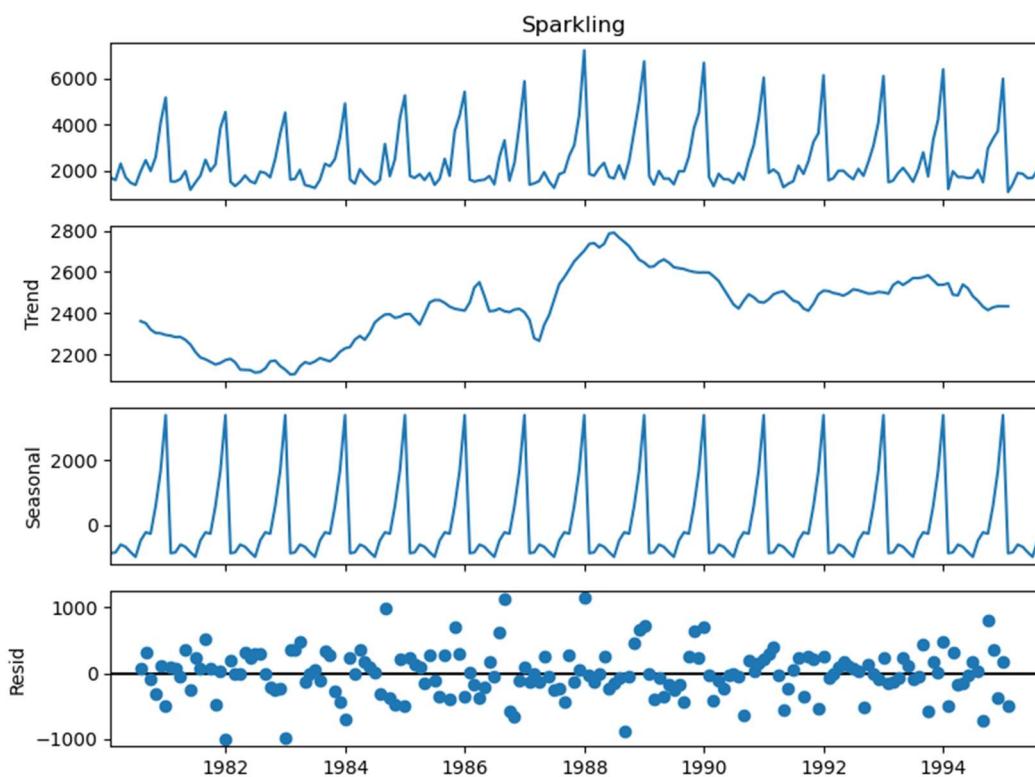
Fig.11 Empirical Cumulative Distribution



2.2 TIME SERIES DECOMPOSITION

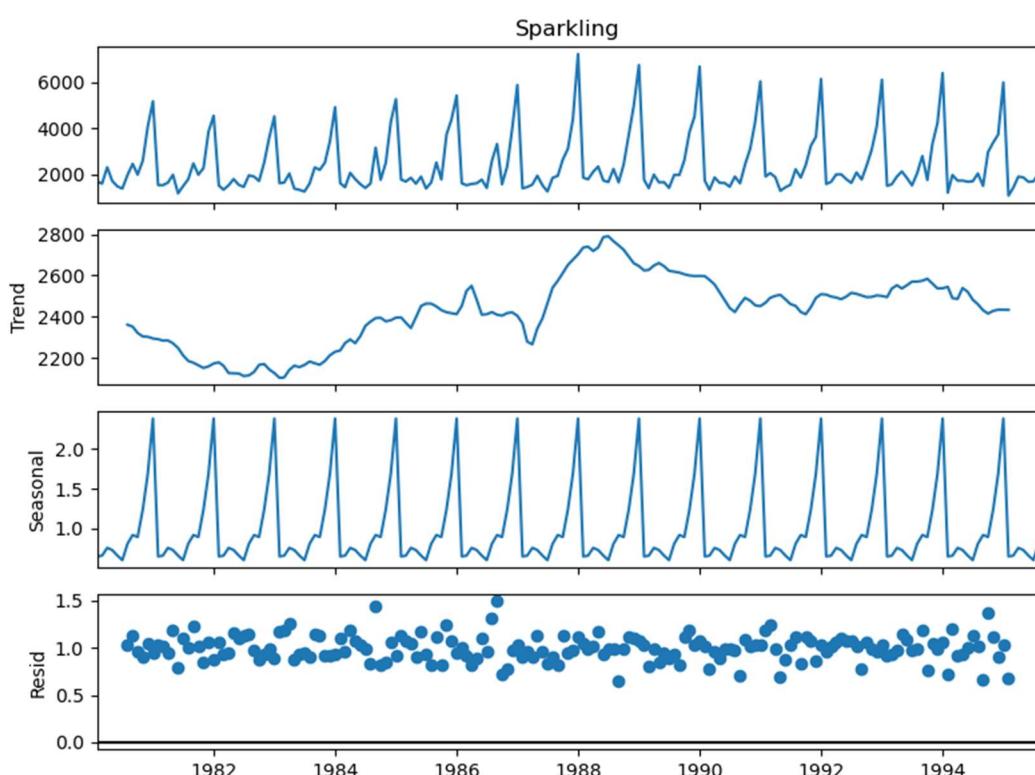
Additive Decomposition:

Fig.12 Additive Decomposition Plot



Multiplicative Decomposition:

Fig.13 Multiplicative Decomposition Plot



INSIGHTS:

- The Sparkling Wine Sales dataset slightly shows an increasing trend in the sales from 1980 to 1995.
- Sparkling Wine Sales dataset shows a clear seasonality pattern.
- The highest median sales are achieved in the month of December which indicates a huge demand during the holiday season.
- In every year, there is a significant increase in the wine sales after the month of August and reaches to peak in the December month.
- January, May and June months have the lowest sales compared to other months.
- Even the dataset slightly shows an upward trend, The seasonal pattern remains same in every year.
- From the multiplicative decomposition plot, we see that a lot of residuals are located around 1. So, the Multiplicative Decomposition is the right way to decompose the time series.
- Seasonality in the dataset indicates a demand in wine during holiday seasons.

3. DATA PRE-PROCESSING

Missing Value Treatment:

- There are no null values or missing values in this dataset.

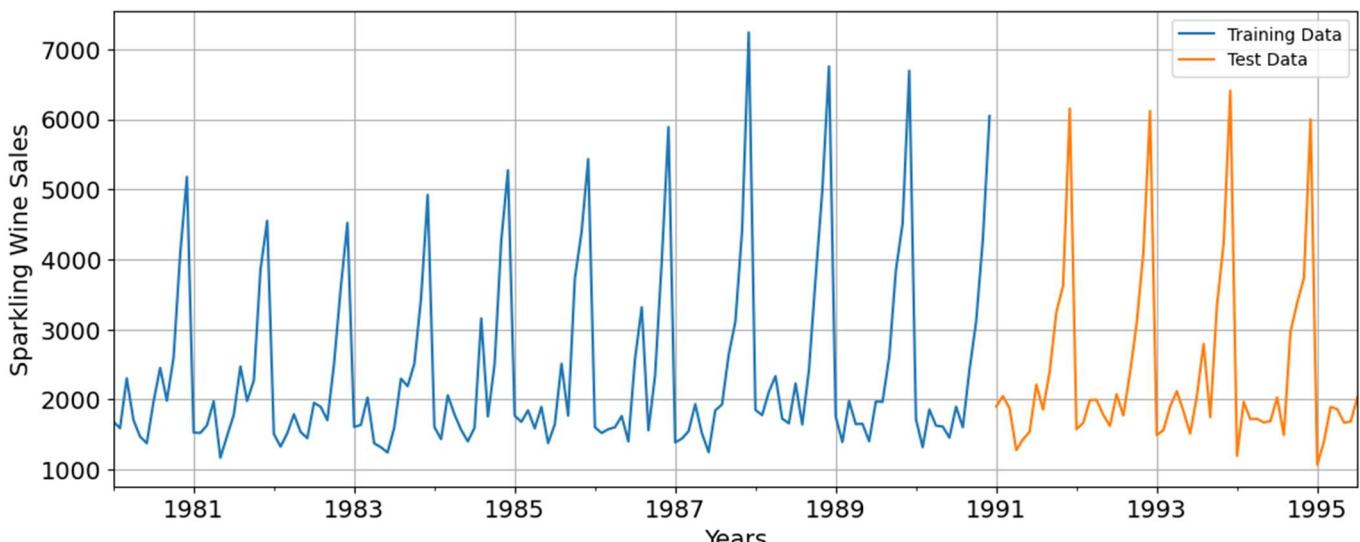
Train-Test Split:

- The time series dataset was split, using data before 1991 for training set and data after 1991 for testing.

Fig.14 Train-Test Split

First few rows of Training Data		First few rows of Test Data	
Sparkling		Sparkling	
1980-01-31	1686	1991-01-31	1902
1980-02-29	1591	1991-02-28	2049
1980-03-31	2304	1991-03-31	1874
1980-04-30	1712	1991-04-30	1279
1980-05-31	1471	1991-05-31	1432
Last few rows of Training Data		Last few rows of Test Data	
Sparkling		Sparkling	
1990-08-31	1605	1995-03-31	1897
1990-09-30	2424	1995-04-30	1862
1990-10-31	3116	1995-05-31	1670
1990-11-30	4286	1995-06-30	1688
1990-12-31	6047	1995-07-31	2031

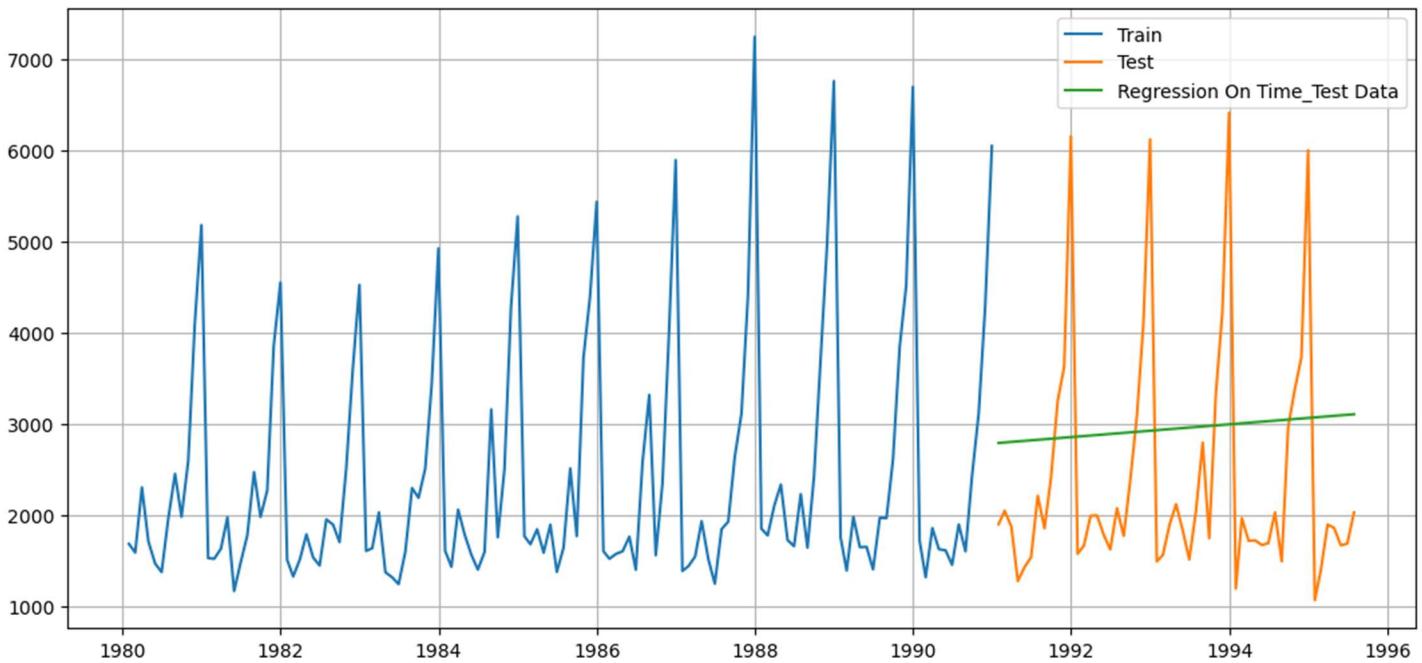
Fig.15 Visualisation of train-test split



4. MODEL BUILDING – ORIGINAL DATA

4.1 LINEAR REGRESSION:

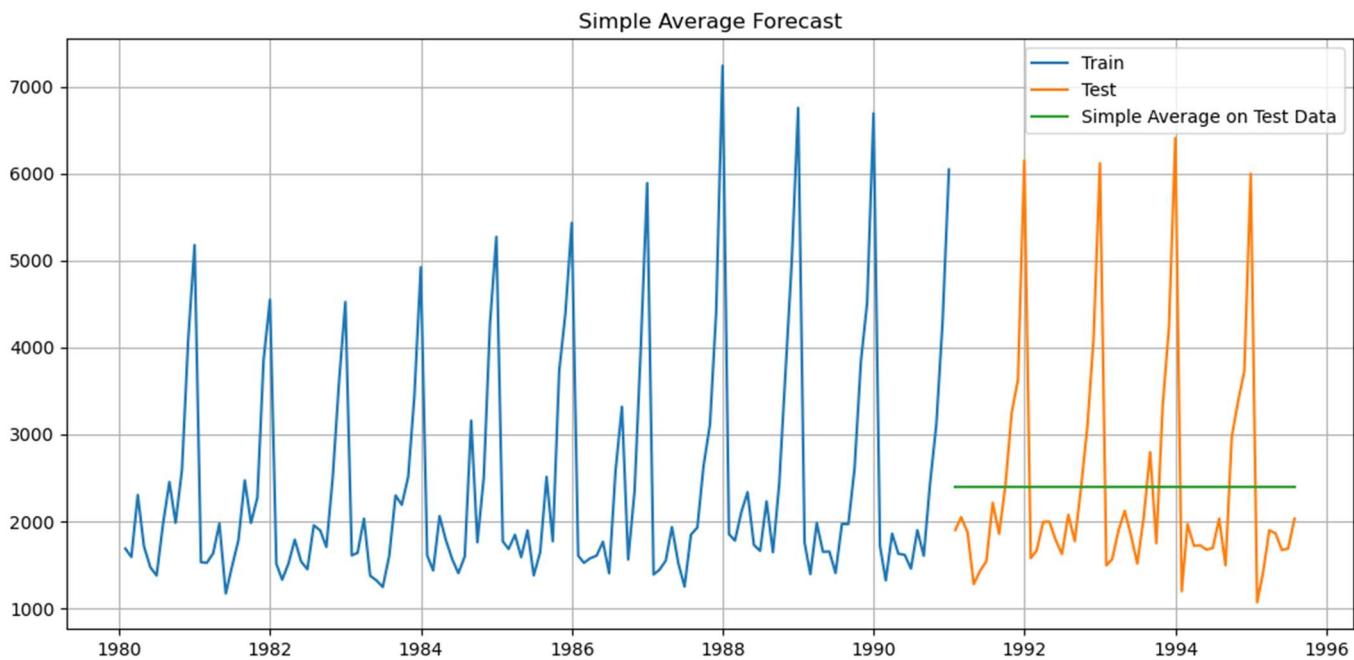
Fig.16 Regression Model Performance



- The **Linear Regression** model shows a slight upward trend in the predictions.
- This model has an RMSE score of **1389.135**
- Regression Model doesn't capture the seasonality in the Sparkling dataset.

4.2 SIMPLE AVERAGE:

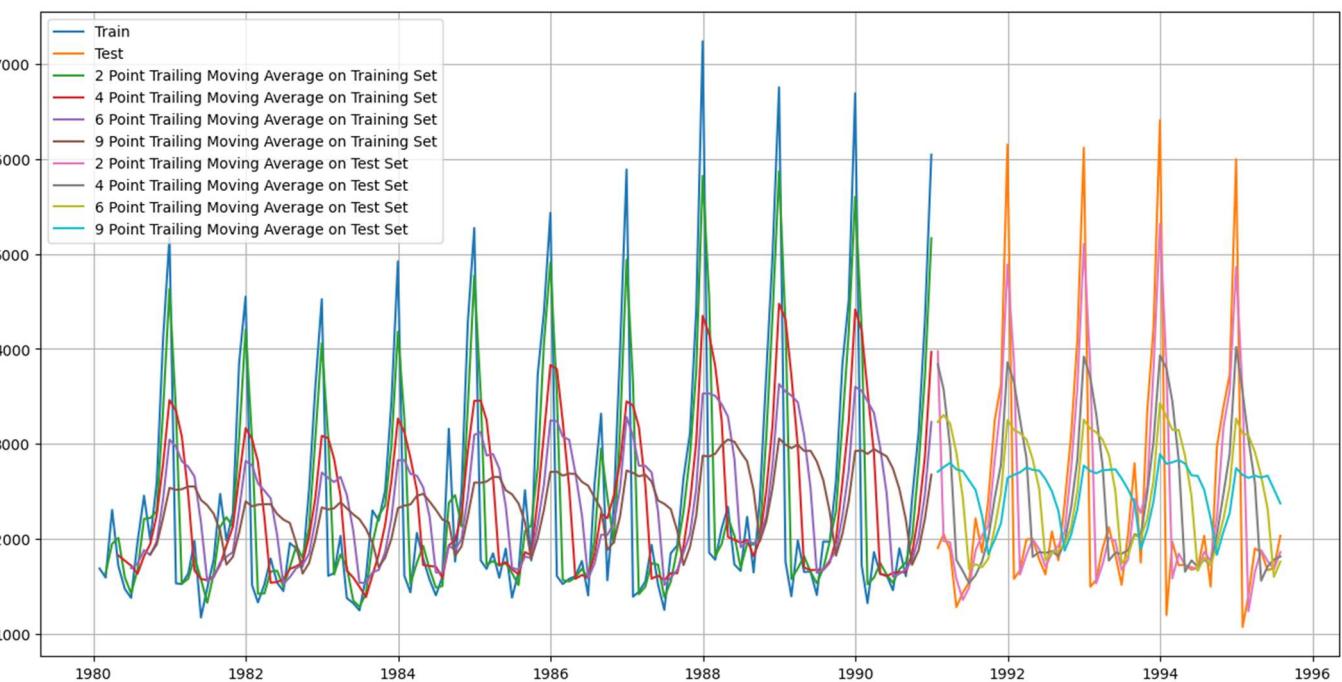
Fig.17 Simple Average Model Performance



- The **Simple Average** forecast remains constant over the test dataset and doesn't capture any trend or seasonality.
- This model has an RMSE score of **1275.08**

4.3 MOVING AVERAGE:

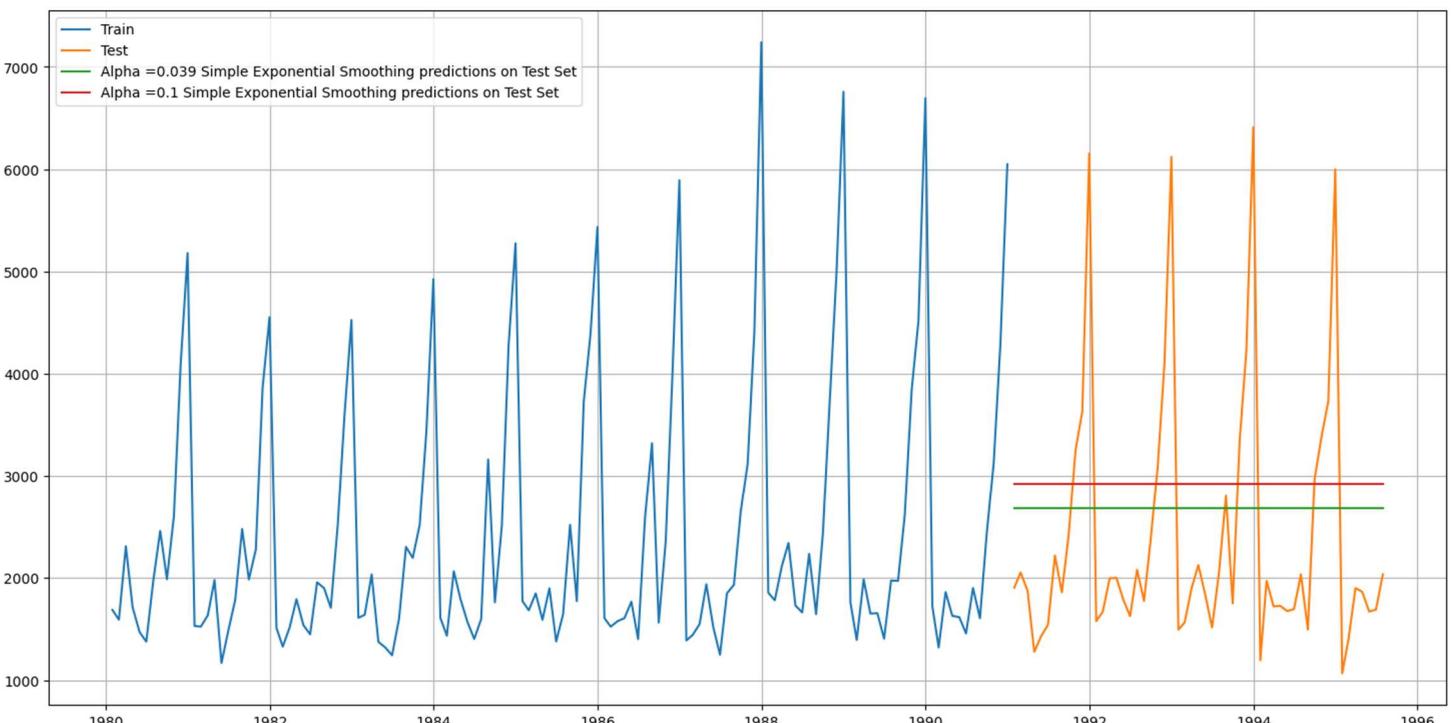
Fig.18 Moving Average Models Performance



- The **Moving Average** models captures trend and seasonality.
- The **2-point trailing** moving average model have the lowest RMSE score of **813.40** comparing to other moving average models
- The **9-point trailing** moving average model have the highest RMSE score of **1346.27**
- Increasing in window size, results in a smoother curve with low accuracy.

4.4 SIMPLE EXPONENTIAL SMOOTHING MODEL:

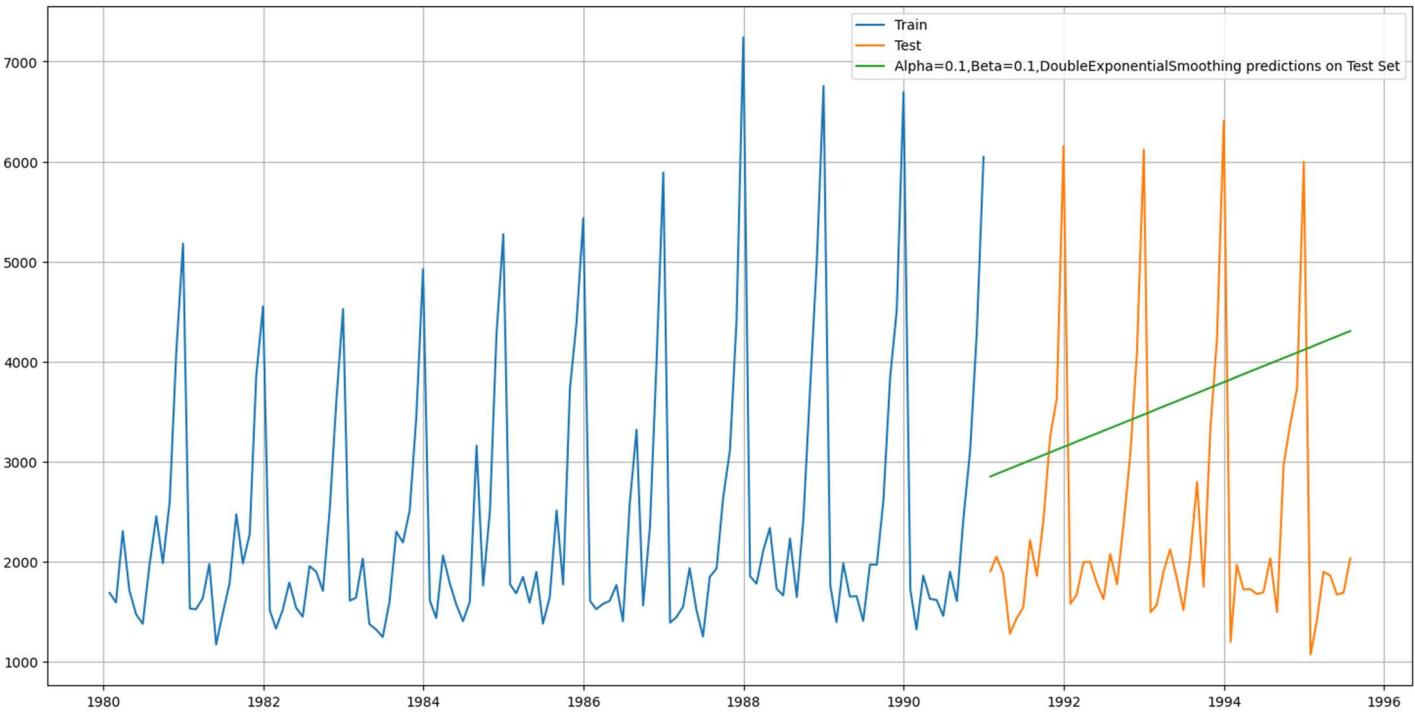
Fig.19 SES model Performance



- The **Simple Exponential Smoothing** Model performs better than simple average, but it still doesn't capture the seasonality and almost remains constant.
- RMSE score of SES ($\alpha = 0.039$) : **1304.92**
- RMSE score of SES ($\alpha = 1$) : **1375.39**

4.5 DOUBLE EXPONENTIAL SMOOTHING MODEL:

Fig.20 DES model Performance



- The Double Exponential Smoothing model ($\alpha=0.1, \beta=0.1$) only captures trend and doesn't capture seasonality with RMSE of **1778.5**.

4.6 TRIPLE EXPONENTIAL SMOOTHING MODEL:

Fig.21 TES model 1 Performance

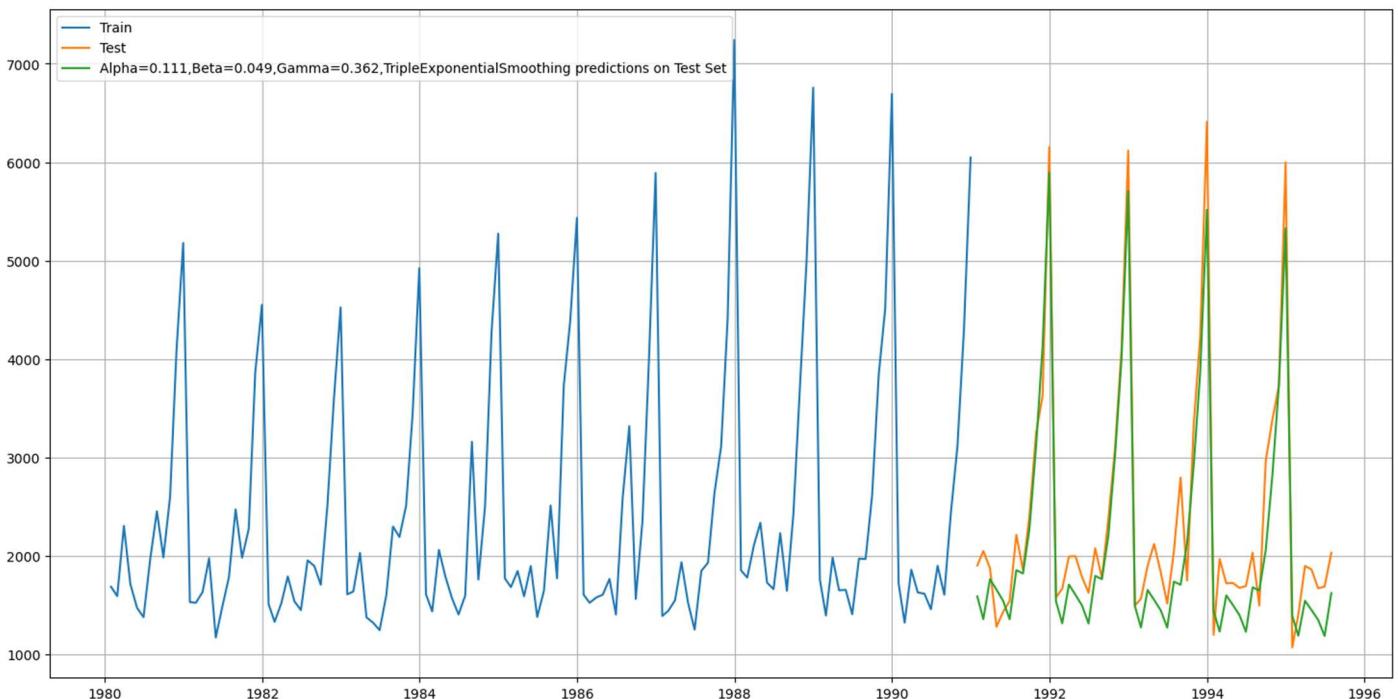
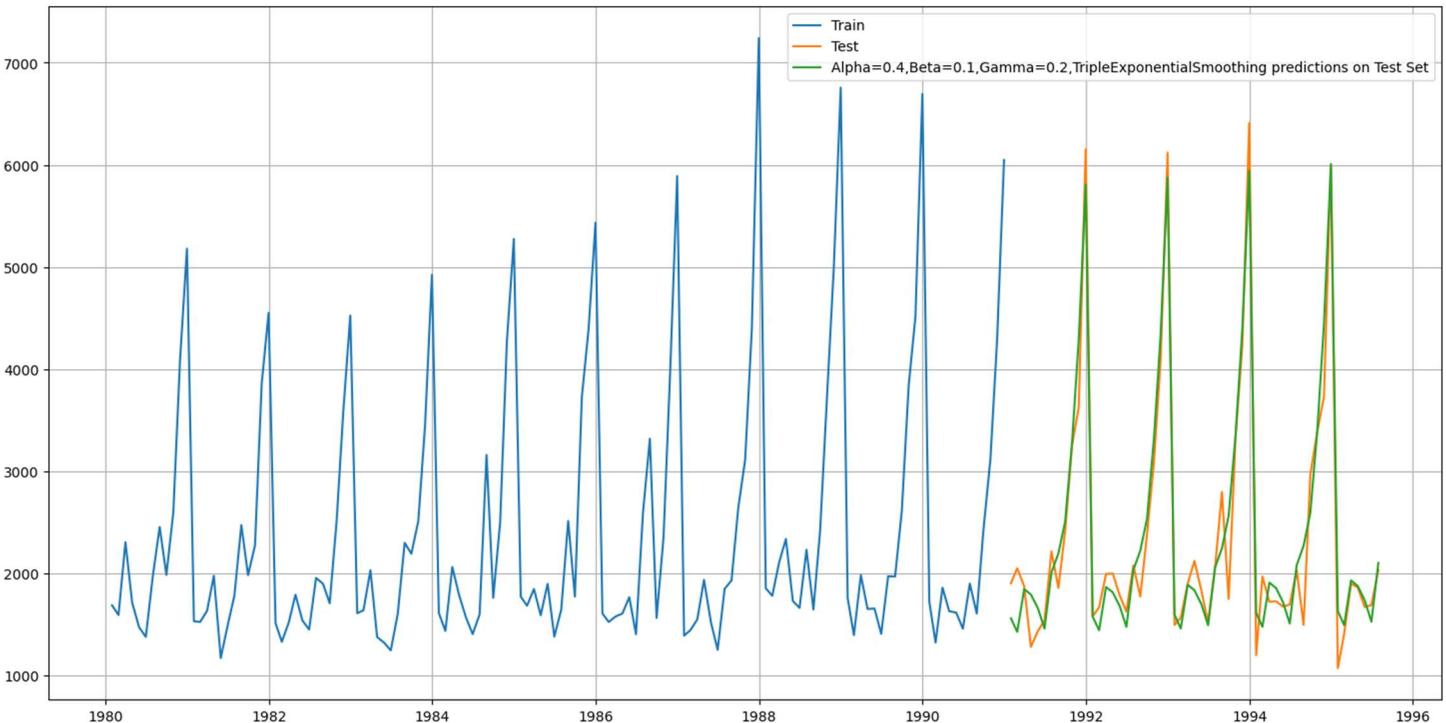


Fig.22 TES model 2 Performance



- The **Triple Exponential Smoothing** model performs better than other models and have the least RMSE scores.
- As we can clearly observe that, the second TES model performs better than the autofit TES model.
- RMSE score of TES ($\alpha = 0.4$, $\beta = 0.1$, $\gamma = 0.2$): **317.434**
- RMSE score of TES ($\alpha = 0.111$, $\beta = 0.049$, $\gamma = 0.362$): **403.706**

4.7 MODEL PERFORMANCE EVALUATION:

Fig.23 Model Comparison plot (Regression, Simple average, Moving average)

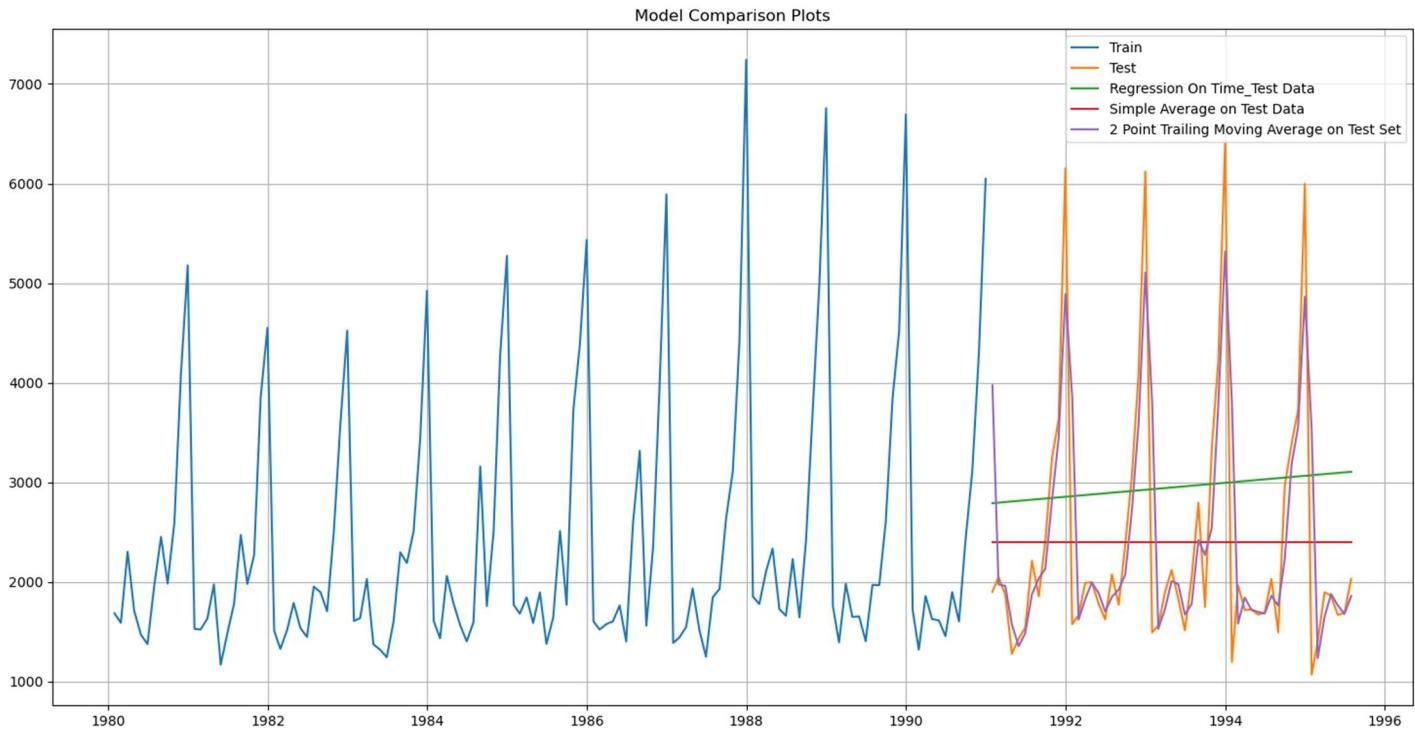


Fig.24 Model Comparison plot (Exponential Smoothing Models)

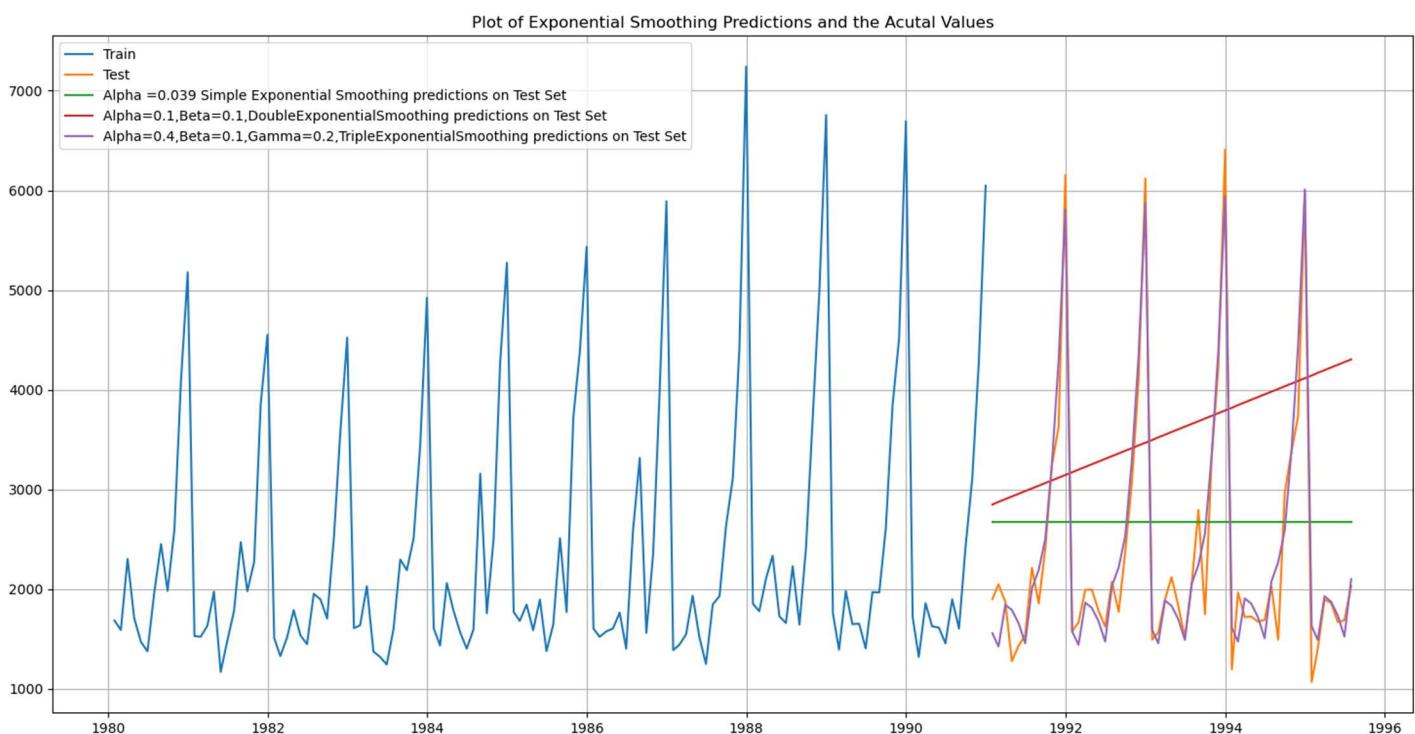


Fig.25 RMSE Scores of all models

	Test RMSE
Alpha=0.4,Beta=0.1,Gamma=0.2,TripleExponentialSmoothing	317.434302
Alpha=0.111,Beta=0.049,Gamma=0.362,TripleExponentialSmoothing	403.706228
2-point Trailing MovingAverage	813.400684
4-point Trailing MovingAverage	1156.589694
SimpleAverageModel	1275.081804
6-point Trailing MovingAverage	1283.927428
Alpha=0.039,SimpleExponentialSmoothing	1304.927405
9-point Trailing MovingAverage	1346.278315
Alpha=0.1,SimpleExponentialSmoothing	1375.393398
Linear Regression Model	1389.135175
Alpha=0.1,Beta=0.1,DoubleExponentialSmoothing	1778.564670

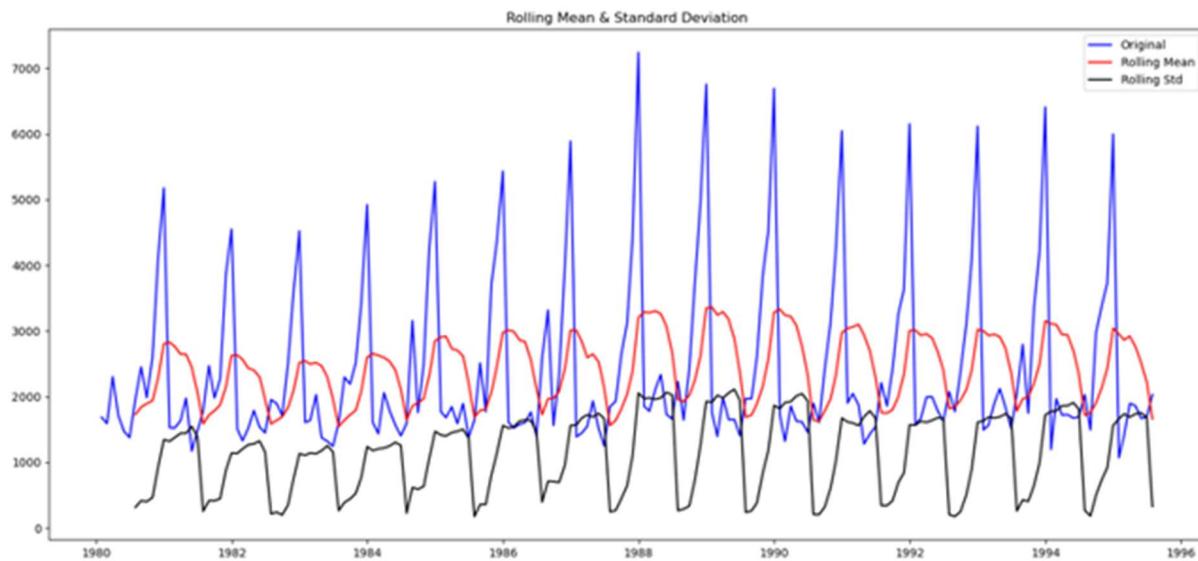
INSIGHTS:

- The **Triple Exponential Smoothing model ($\alpha = 0.4$, $\beta = 0.1$, $\gamma = 0.2$)** has the lowest RMSE of 317.43 and performs better than other models.
- Both TES models outperforms other models by capturing trend and seasonality in forecasting.
- The Double Exponential Smoothing model has the highest RMSE score of 1778.564
- In Moving Average models, the 2-point Trailing moving average model has the lowest RMSE of 813.4
- The Simple average model, Regression model and SES models doesn't capture seasonality which leads to poor performance.

5. CHECK FOR STATIONARITY

Checking for Stationarity on the whole Time series data:

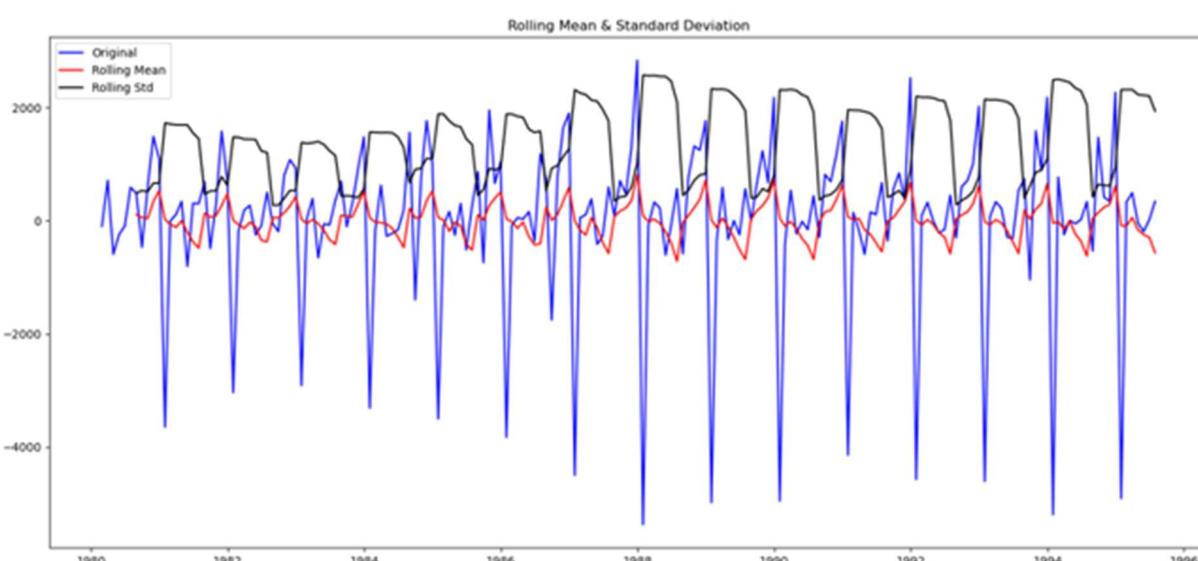
Fig.26 Stationarity check (whole data)



Results of Dickey-Fuller Test:

Test Statistic	-1.360497
p-value	0.601061
#Lags Used	11.000000
Number of Observations Used	175.000000
Critical Value (1%)	-3.468280
Critical Value (5%)	-2.878202
Critical Value (10%)	-2.575653

dtype: float64



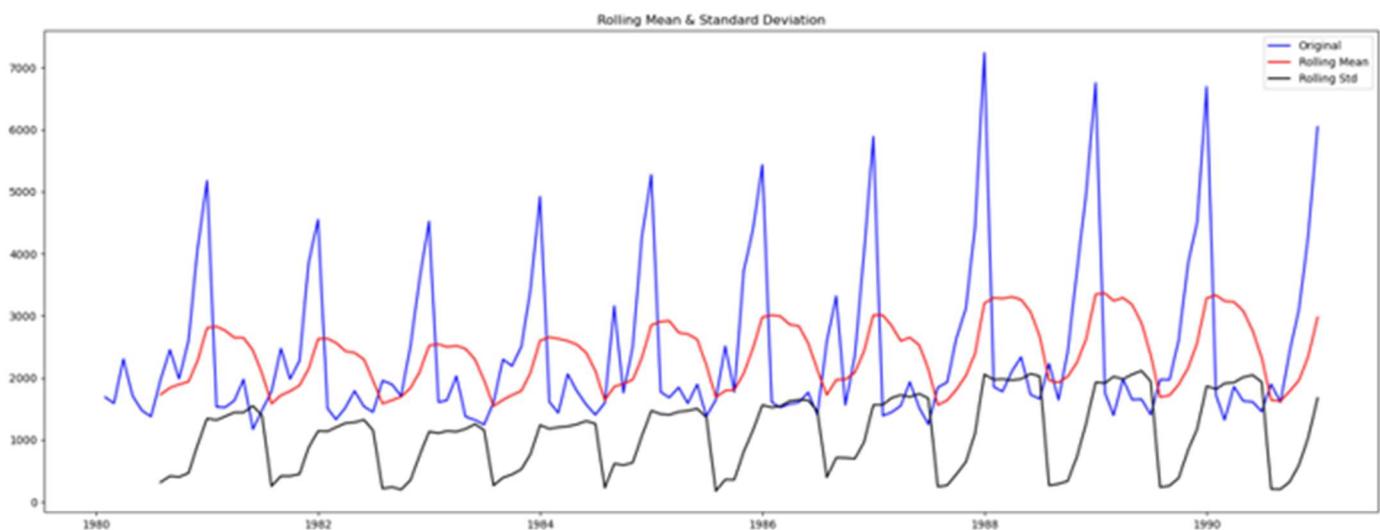
Results of Dickey-Fuller Test:

Test Statistic	-45.050301
p-value	0.000000
#Lags Used	10.000000
Number of Observations Used	175.000000
Critical Value (1%)	-3.468280
Critical Value (5%)	-2.878202
Critical Value (10%)	-2.575653

dtype: float64

Checking for Stationarity on the Training data:

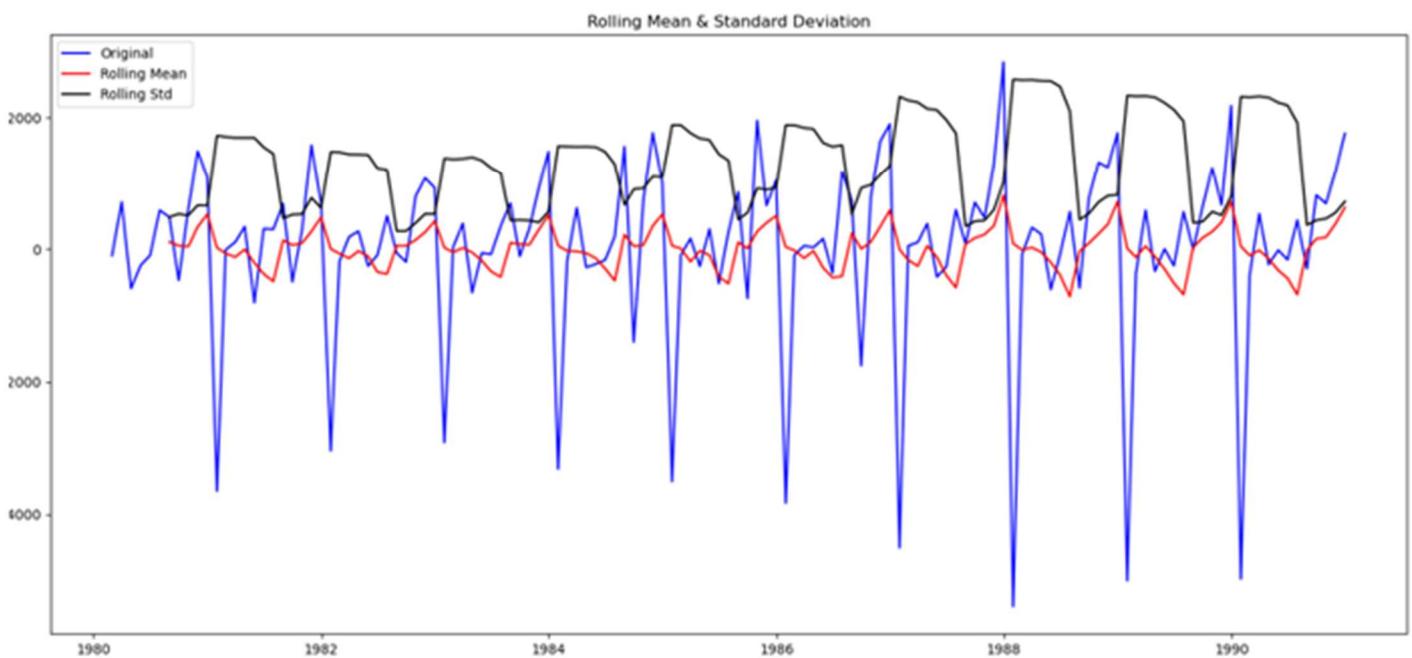
Fig.27 Stationarity check (Training data)



Results of Dickey-Fuller Test:

Test Statistic	-1.208926
p-value	0.669744
#Lags Used	12.000000
Number of Observations Used	119.000000
Critical Value (1%)	-3.486535
Critical Value (5%)	-2.886151
Critical Value (10%)	-2.579896

dtype: float64



Results of Dickey-Fuller Test:

Test Statistic	-8.005007e+00
p-value	2.280104e-12
#Lags Used	1.100000e+01
Number of Observations Used	1.190000e+02
Critical Value (1%)	-3.486535e+00
Critical Value (5%)	-2.886151e+00
Critical Value (10%)	-2.579896e+00

dtype: float64

Whole Time series Dataset:

- From the Dickey-Fuller Test in whole time series, we can clearly observe that the p-value is 0.601, which is greater than 0.05 at 5% significance level. The Time series is non-stationary.
- After taking a difference of order 1, we see that at 5% significance level, the Time Series data is stationary.

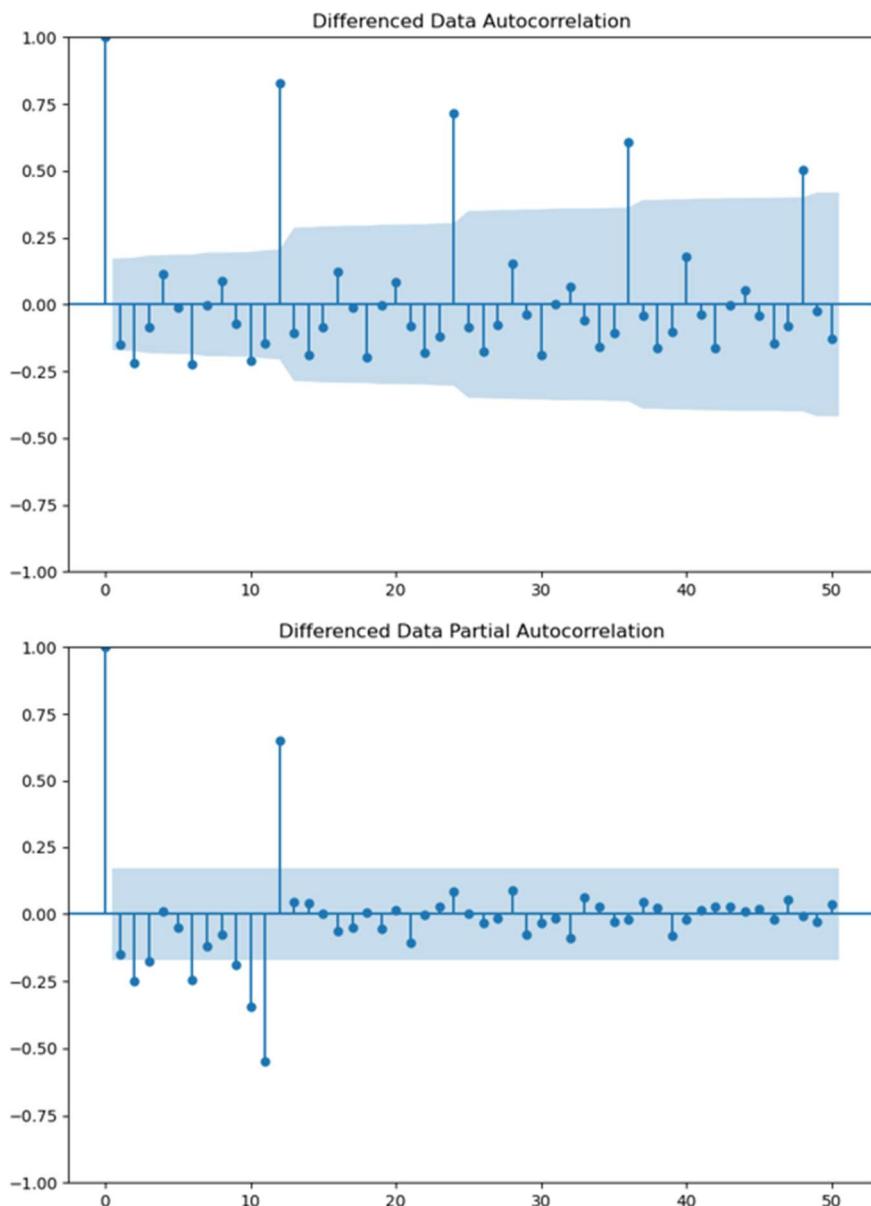
Training Dataset:

- From the Dickey-Fuller Test in training set, we can clearly observe that the p-value is 0.66, which is greater than 0.05 at 5% significance level. The Time series is non-stationary.
- After taking a difference of order 1, we see that at 5% significance level, the Time Series data is stationary.

6. MODEL BUILDING – STATIONARY DATA

6.1 ACF AND PACF PLOTS:

Fig.28 ACF, PACF plots (Training data)



- By observing the PACF plot, the Auto-regressive (AR) parameter p value is taken as 0.
- By observing the ACF plot, the Moving-Average (MA) parameter q value is taken as 0.
- By using these values, we can set a range and iterate with different values for manual ARIMA and SARIMA models.

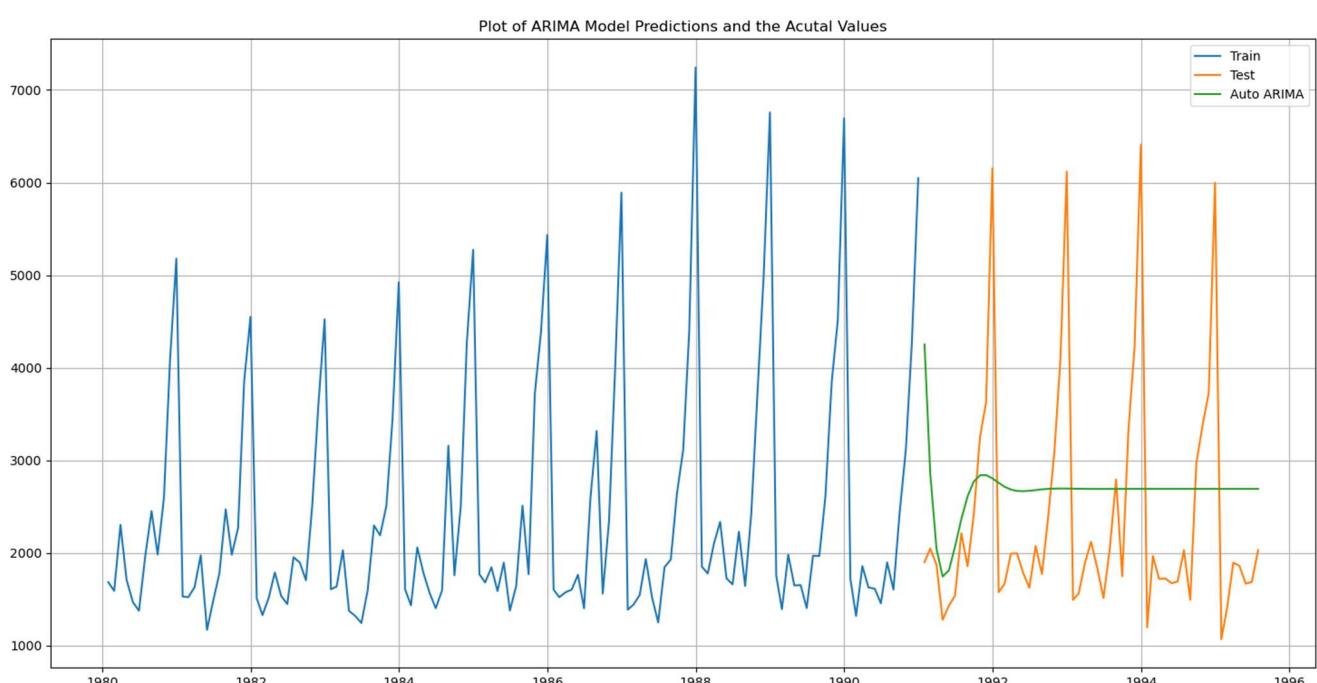
6.2 AUTO ARIMA:

- An Auto ARIMA model was built on the training data in which the parameters are selecting using the lowest Akaike Information Criteria (AIC) values by iterating different p , q values.
- This ARIMA model was built on parameters (2,1,2)
- Test RMSE: 1299.98

Fig.29 Auto ARIMA Results

```
SARIMAX Results
=====
Dep. Variable: Sparkling    No. Observations: 132
Model: ARIMA(2, 1, 2)    Log Likelihood: -1101.755
Date: Sun, 09 Mar 2025   AIC: 2213.509
Time: 19:33:13            BIC: 2227.885
Sample: 01-31-1980        HQIC: 2219.351
                           - 12-31-1990
Covariance Type: opg
=====
              coef    std err      z    P>|z|    [0.025]    [0.975]
-----+
ar.L1      1.3121    0.046   28.781    0.000     1.223     1.401
ar.L2     -0.5593    0.072   -7.740    0.000    -0.701    -0.418
ma.L1     -1.9917    0.109   -18.217    0.000    -2.206    -1.777
ma.L2      0.9999    0.110     9.109    0.000     0.785     1.215
sigma2    1.099e+06  1.99e-07  5.51e+12    0.000    1.1e+06   1.1e+06
=====
Ljung-Box (L1) (Q):      0.19    Jarque-Bera (JB): 14.46
Prob(Q):                0.67    Prob(JB):          0.00
Heteroskedasticity (H): 2.43    Skew:             0.61
Prob(H) (two-sided):    0.00    Kurtosis:         4.08
=====
```

Fig.30 Auto ARIMA model plot



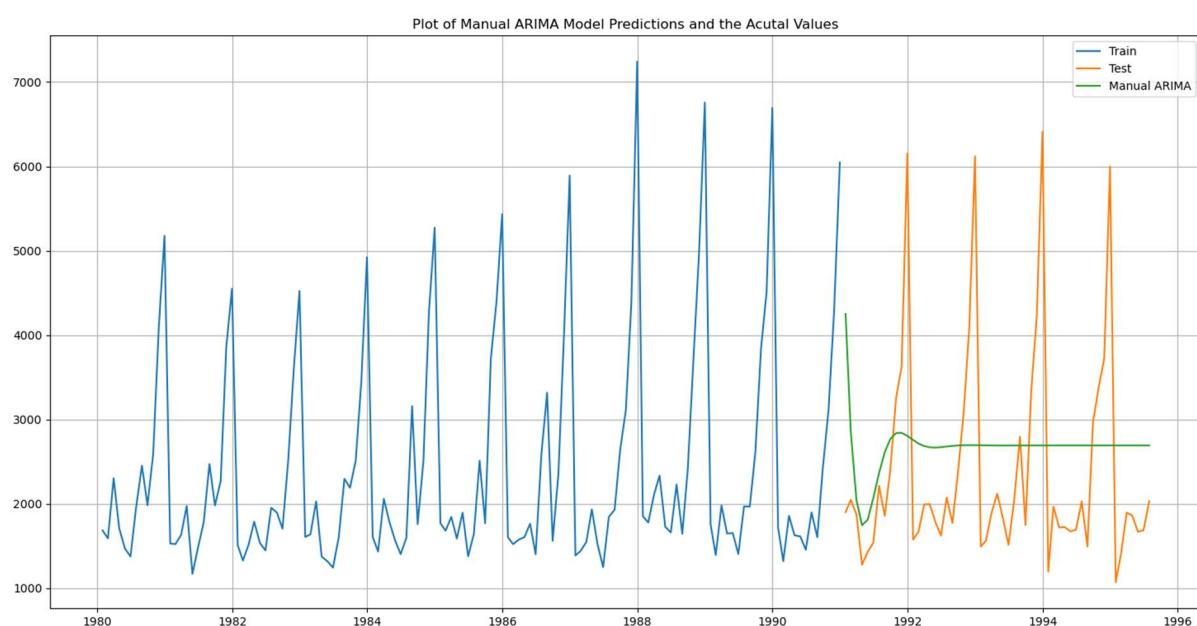
6.3 MANUAL ARIMA MODEL:

- From observing the PACF and ACF plots, p and q values are selected from a range of (1-4)
- Data is made stationary after differencing of order 1 (So, d=1)
- After iterating all the values, the same set of parameters (2,1,2) yield the lowest AIC score, with test RMSE of 1299.98

Fig.31 Manual ARIMA Results

```
SARIMAX Results
=====
Dep. Variable: Sparkling   No. Observations: 132
Model: ARIMA(2, 1, 2)   Log Likelihood: -1101.755
Date: Sun, 09 Mar 2025   AIC: 2213.509
Time: 19:34:02           BIC: 2227.885
Sample: 01-31-1980       HQIC: 2219.351
                           - 12-31-1990
Covariance Type: opg
=====
              coef    std err      z   P>|z|   [0.025]   [0.975]
-----
ar.L1      1.3121    0.046   28.781   0.000     1.223    1.401
ar.L2     -0.5593    0.072   -7.740   0.000    -0.701   -0.418
ma.L1     -1.9917    0.109   -18.217   0.000    -2.206   -1.777
ma.L2      0.9999    0.110     9.109   0.000     0.785    1.215
sigma2    1.099e+06  1.99e-07  5.51e+12  0.000    1.1e+06  1.1e+06
-----
Ljung-Box (L1) (Q):      0.19   Jarque-Bera (JB):      14.46
Prob(Q):                0.67   Prob(JB):                 0.00
Heteroskedasticity (H):  2.43   Skew:                   0.61
Prob(H) (two-sided):    0.00   Kurtosis:                4.08
=====
```

Fig.32 Manual ARIMA plot



6.4 AUTO SARIMA MODEL:

- From observing the ACF plots, we can clearly observe that seasonality s = 12
- An Auto SARIMA model was built on the training data in which the parameters are selecting using the lowest Akaike Information Criteria (AIC) values by iterating different parameters.
- This SARIMA model was built on parameters (1,1,2) (1,0,2,12)
- Test RMSE: 528.5

Fig.33 Auto SARIMA results

```
SARIMAX Results
=====
Dep. Variable:                      y    No. Observations:                 132
Model:             SARIMAX(1, 1, 2)x(1, 0, 2, 12)   Log Likelihood:            -770.792
Date:                Sun, 09 Mar 2025   AIC:                         1555.584
Time:                    19:34:40     BIC:                         1574.095
Sample:                   0 - 132   HQIC:                        1563.083
Covariance Type:                  opg
=====
              coef    std err        z     P>|z|      [0.025      0.975]
-----
ar.L1       -0.6281    0.255   -2.463     0.014    -1.128    -0.128
ma.L1       -0.1041    0.225   -0.463     0.643    -0.545     0.337
ma.L2       -0.7276    0.154   -4.734     0.000    -1.029    -0.426
ar.S.L12      1.0439    0.014   72.842     0.000     1.016     1.072
ma.S.L12     -0.5550    0.098   -5.663     0.000    -0.747    -0.363
ma.S.L24     -0.1355    0.120   -1.133     0.257    -0.370     0.099
sigma2      1.506e+05  2.03e+04    7.401     0.000   1.11e+05   1.9e+05
=====
Ljung-Box (L1) (Q):                  0.04  Jarque-Bera (JB):            11.72
Prob(Q):                           0.84  Prob(JB):                     0.00
Heteroskedasticity (H):               1.47  Skew:                         0.36
Prob(H) (two-sided):                0.26  Kurtosis:                     4.48
=====
```

Fig.34 Auto SARIMA diagnostics

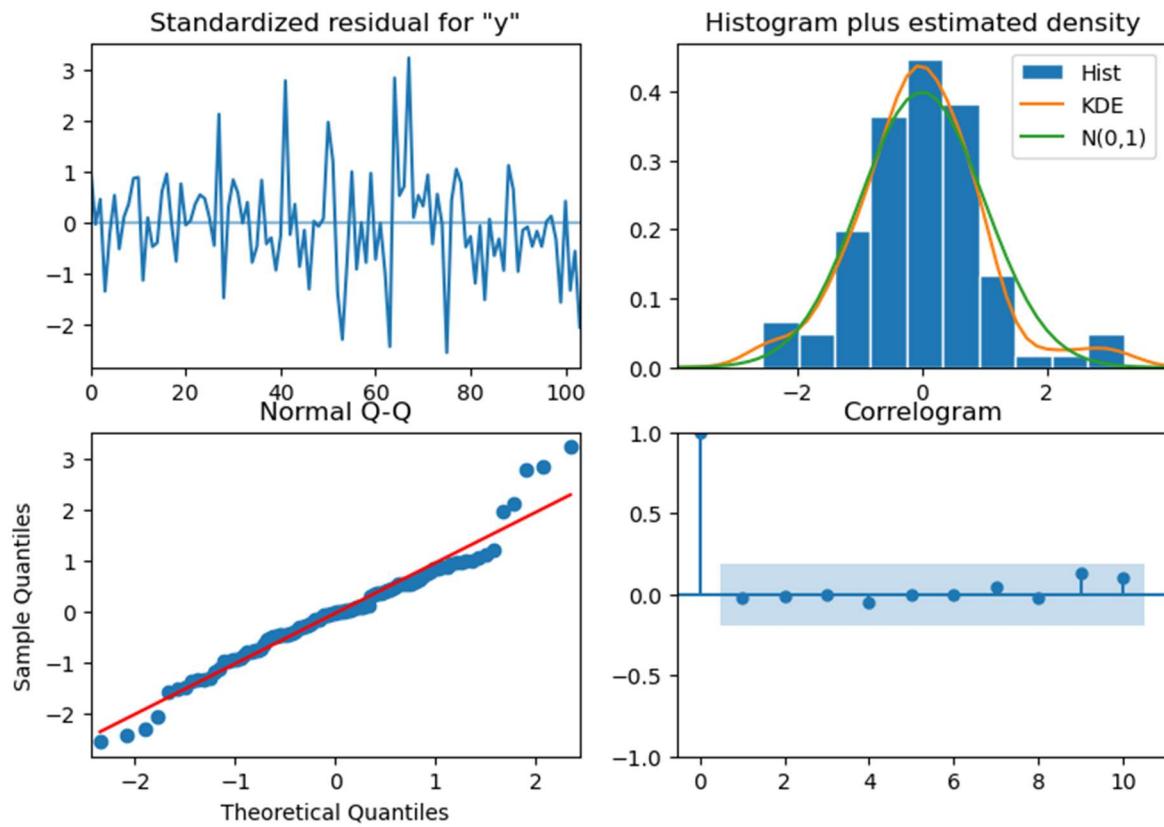
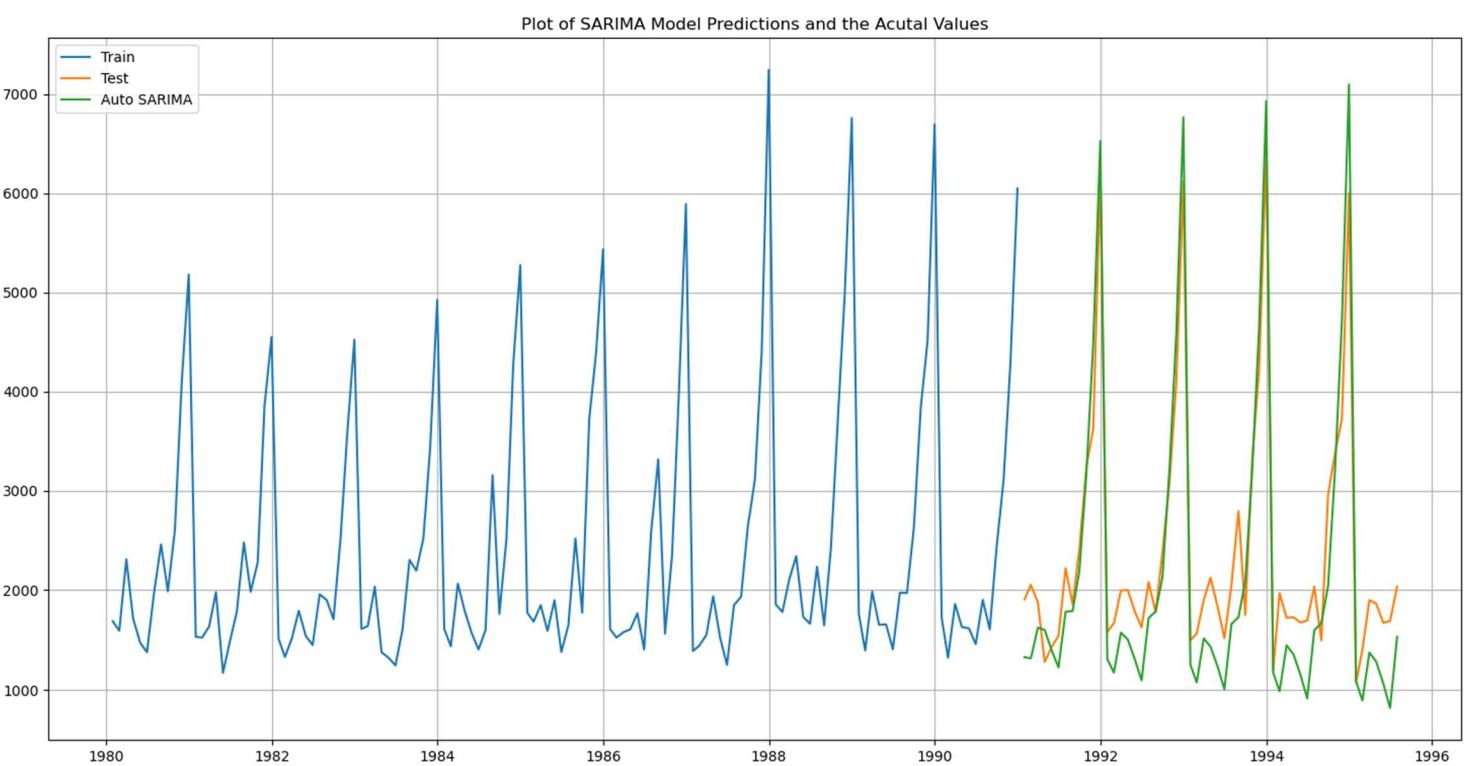


Fig.35 Auto SARIMA plot



6.5 MANUAL SARIMA MODEL:

- From observing the PACF and ACF plots, p and q values are selected from a range of (0-4)
- Data is made stationary after differencing of order 1 (So, d=1)
- Since there are strong seasonal patterns in ACF, seasonal differencing can be done. (So, D=1)
- After iterating different values, the SARIMA model was built with parameters (1,1,2) (0,1,2,12) which yield the lowest AIC score, with test RMSE of 382.576

Fig.36 Manual SARIMA Results

```
SARIMAX Results
=====
Dep. Variable:                      y      No. Observations:                 132
Model:             SARIMAX(1, 1, 2)x(0, 1, 2, 12)   Log Likelihood:            -685.174
Date:                Sun, 09 Mar 2025      AIC:                         1382.348
Time:                    19:35:59        BIC:                         1397.479
Sample:                   0 - 132       HQIC:                        1388.455
Covariance Type:            opg
=====
              coef    std err      z   P>|z|      [0.025]     [0.975]
-----
ar.L1     -0.5507    0.287   -1.922     0.055     -1.112     0.011
ma.L1     -0.1612    0.235   -0.687     0.492     -0.621     0.299
ma.L2     -0.7218    0.175   -4.132     0.000     -1.064     -0.379
ma.S.L12   -0.4062    0.092   -4.401     0.000     -0.587     -0.225
ma.S.L24   -0.0274    0.138   -0.198     0.843     -0.298     0.243
sigma2    1.705e+05  2.45e+04    6.956     0.000    1.22e+05   2.19e+05
=====
Ljung-Box (L1) (Q):                  0.00  Jarque-Bera (JB):           13.48
Prob(Q):                           0.95  Prob(JB):                  0.00
Heteroskedasticity (H):               0.89  Skew:                      0.60
Prob(H) (two-sided):                0.75  Kurtosis:                  4.44
=====
```

Fig.37 Manual SARIMA diagnostics

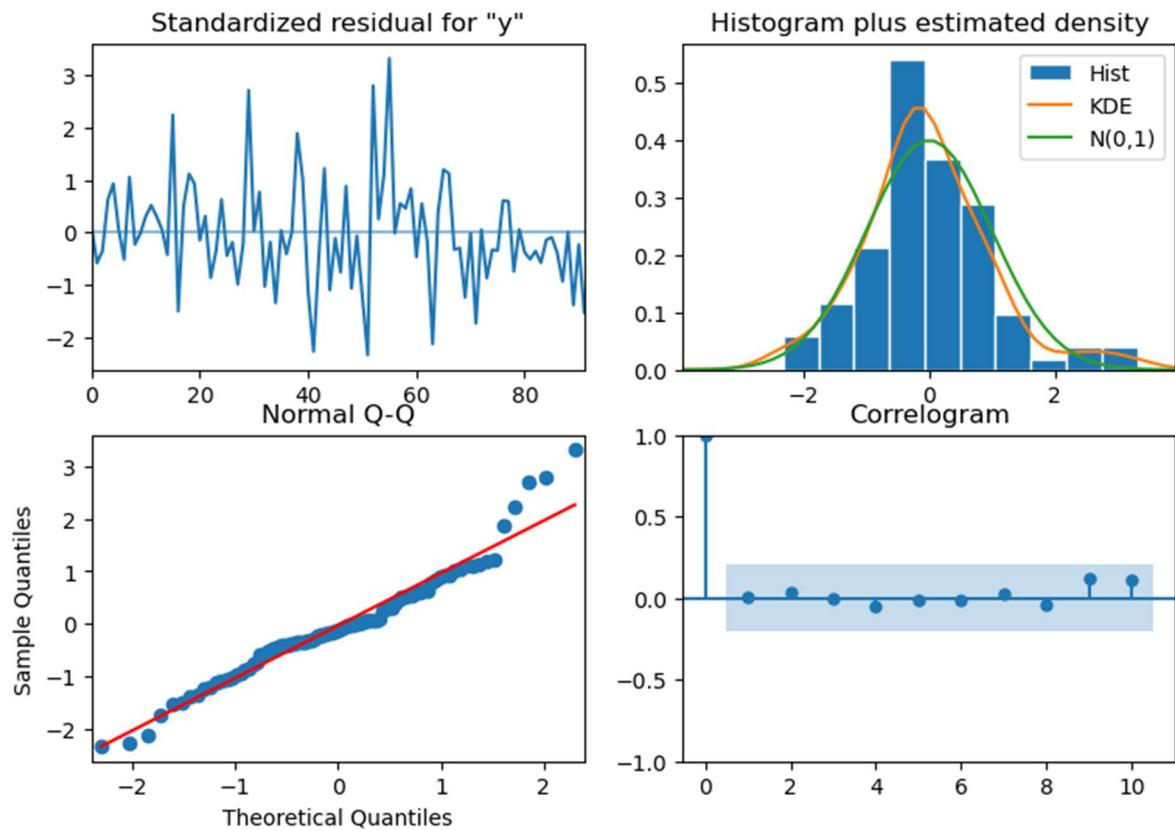
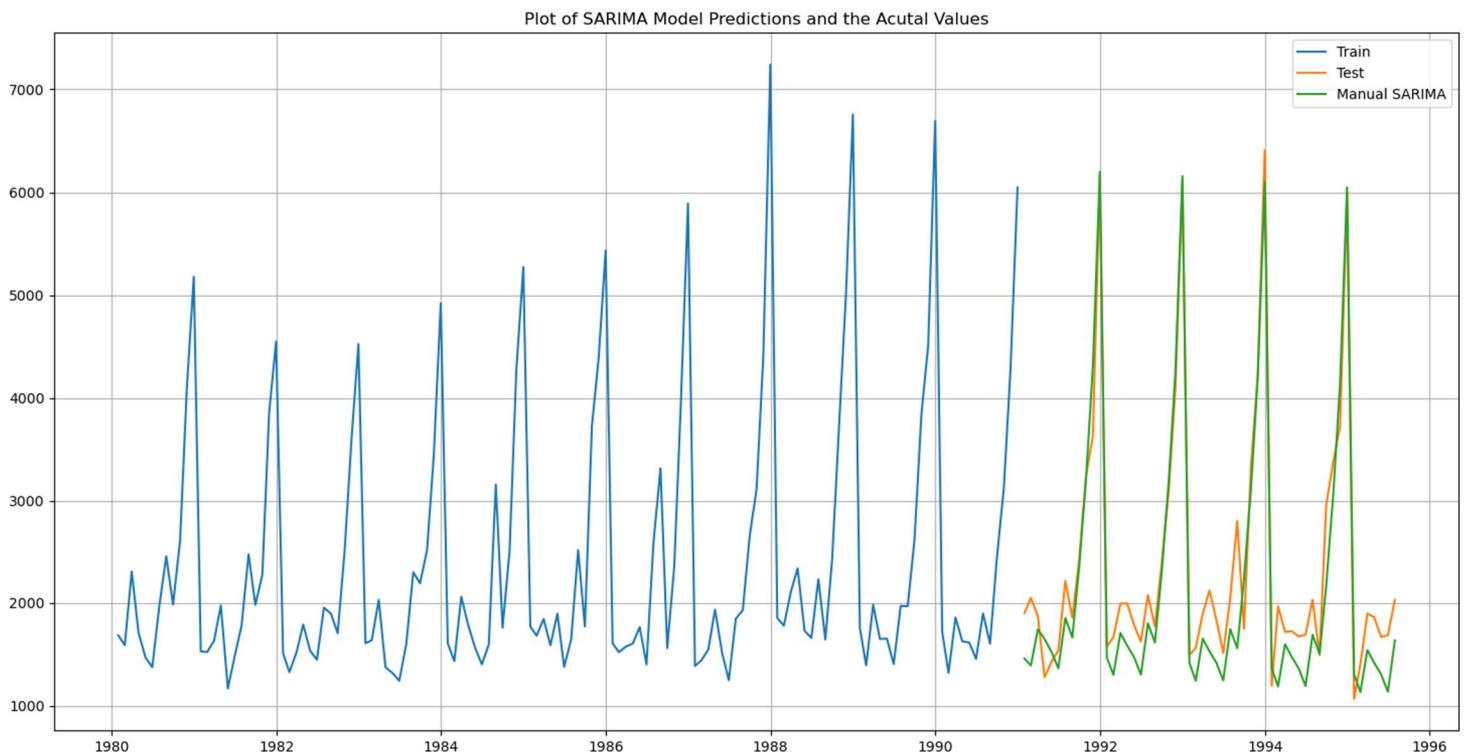


Fig.38 Manual SARIMA plot

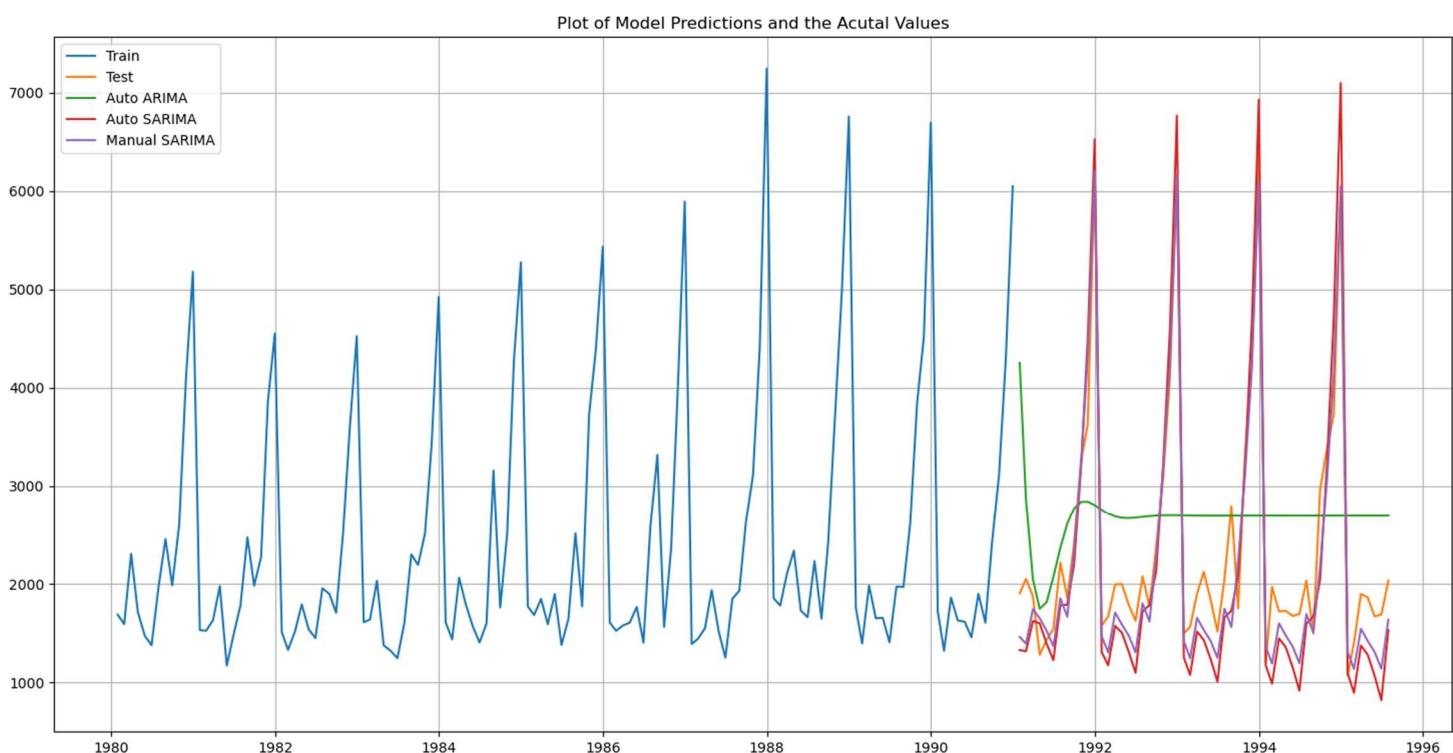


6.6 MODEL PERFORMANCE EVALUATION:

Fig.39 ARIMA & SARIMA model performances

Auto ARIMA(2,1,2)	1299.980107
Manual ARIMA(2,1,2)	1299.980107
Auto SARIMA(1,1,2)(1,0,2,12)	528.586898
Manual SARIMA(1,1,2)(0,1,2,12)	382.576727

Fig.40 ARIMA & SARIMA model predictions



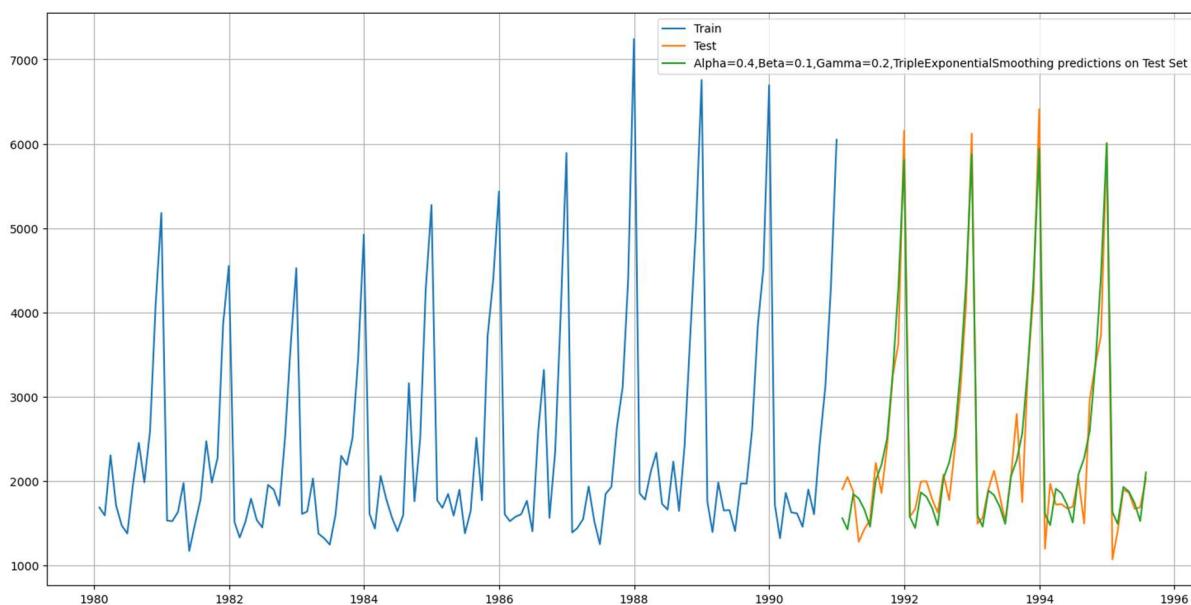
- Auto ARIMA and Manual ARIMA has the same parameters of (2,1,2) with Test RMSE of 1299.98.
- These ARIMA models underperforms when compared to other models.
- Auto SARIMA model performs slightly better than the ARIMA models with the Test RMSE of 528.5
- Manual SARIMA with parameters as (1,1,2)(0,1,2,12) has the lowest RMSE score of 382.57 when compared to other ARIMA/SARIMA models

7. MODEL PERFORMANCE COMPARISON AND FINAL MODEL SELECTION FOR FORECAST

Fig.41 All model performances

	Test RMSE
Alpha=0.4,Beta=0.1,Gamma=0.2,TripleExponentialSmoothing	317.434302
Manual SARIMA(1,1,2)(0,1,2,12)	382.576727
Alpha=0.111,Beta=0.049,Gamma=0.362,TripleExponentialSmoothing	403.706228
Auto SARIMA(1,1,2)(1,0,2,12)	528.586898
2-point Trailing MovingAverage	813.400684
4-point Trailing MovingAverage	1156.589694
SimpleAverageModel	1275.081804
6-point Trailing MovingAverage	1283.927428
Auto ARIMA(2,1,2)	1299.980107
Manual ARIMA(2,1,2)	1299.980107
Alpha=0.039,SimpleExponentialSmoothing	1304.927405
9-point Trailing MovingAverage	1346.278315
Alpha=0.1,SimpleExponentialSmoothing	1375.393398
Linear Regression Model	1389.135175
Alpha=0.1,Beta=0.1,DoubleExponentialSmoothing	1778.564670

Fig.42 Triple Exponential Smoothing ($\alpha = 0.4$, $\beta = 0.1$ and $\gamma = 0.2$)



- The **Triple Exponential Smoothing** with the parameters $\alpha = 0.4$, $\beta = 0.1$ and $\gamma = 0.2$ has the lowest Test RMSE score of **317.434** and is considered as the **best forecasting model** for this Sparkling wine dataset.
- Manual SARIMA with parameters as $(1,1,2)(0,1,2,12)$ has the RMSE score of 382.57 and performs better compared to other ARIMA/SARIMA models.
- The Double Exponential Smoothing has the highest Test RMSE score of 1778.56 and is considered as the worst performing model.

Forecasted Results of the best model (For next 12 months):

Fig.43 Forecasted Results

```

1995-08-31      2088.425429
1995-09-30      2614.039802
1995-10-31      3457.976195
1995-11-30      4330.695877
1995-12-31      6595.144301
1996-01-31      1550.768494
1996-02-29      1832.282175
1996-03-31      2074.922505
1996-04-30      2002.415693
1996-05-31      1821.629000
1996-06-30      1708.486723
1996-07-31      2190.179438
Freq: M, dtype: float64

```

Fig.44 Forecast of the best model

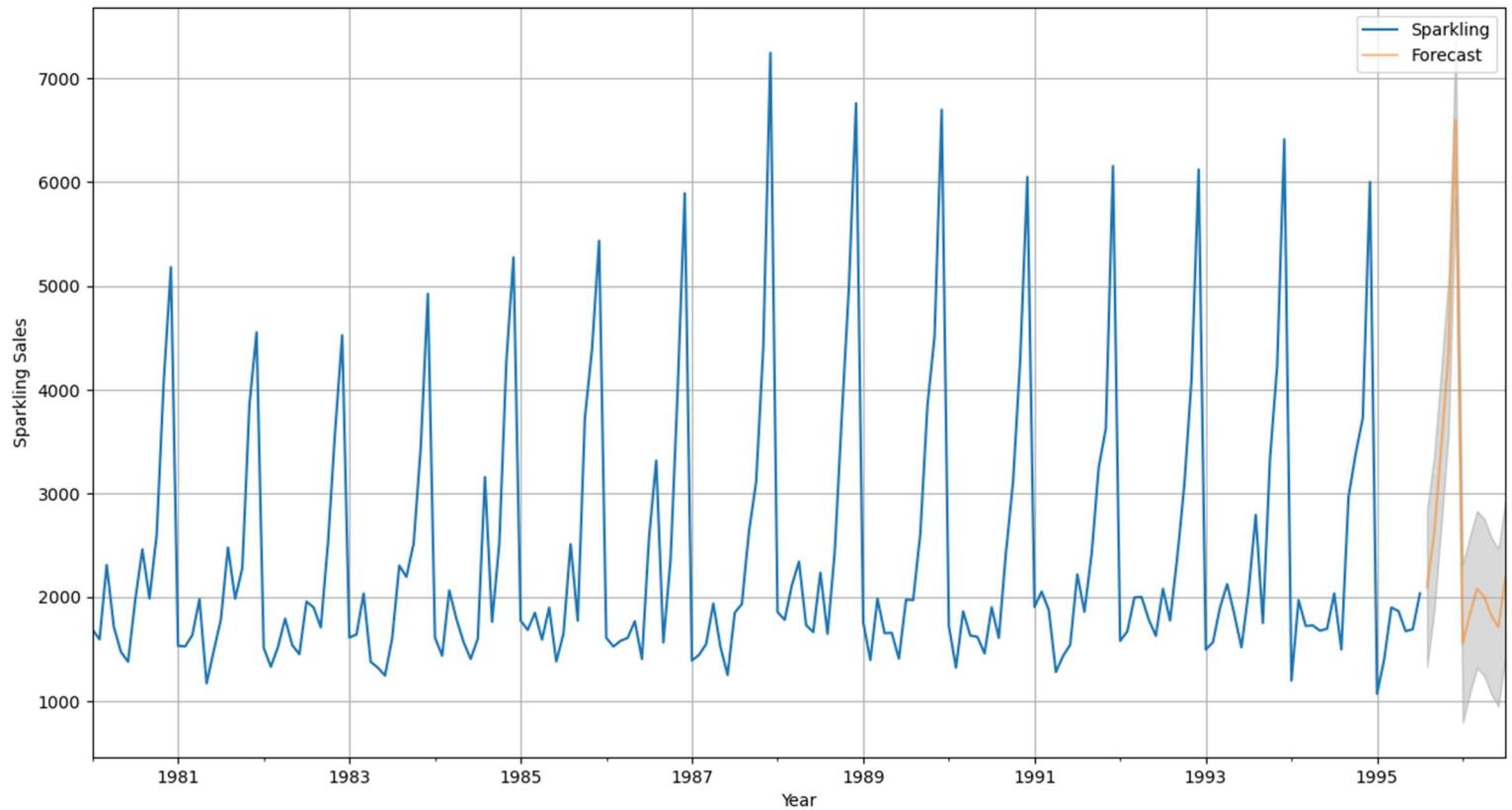
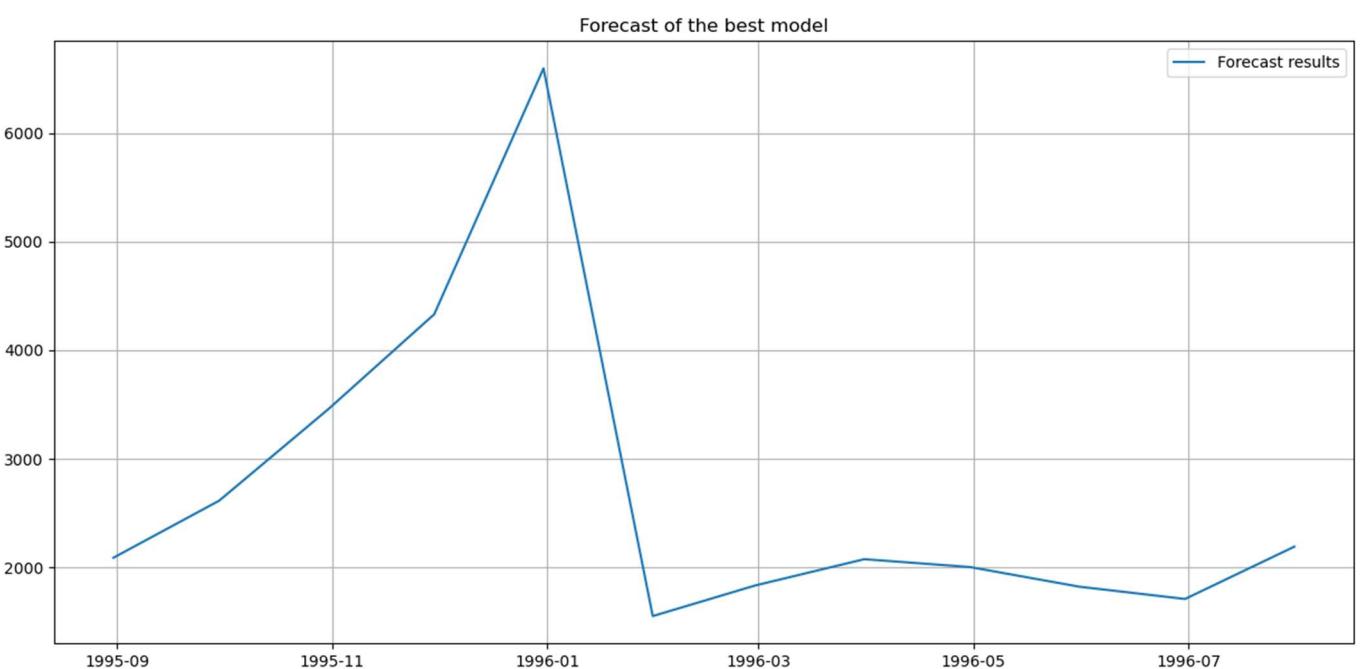


Fig.45 Plotting only Forecast results



8. ACTIONABLE INSIGHTS & RECOMMENDATIONS

- **December** month have the highest sales compared to other months due to holiday and festival season demands.

Fig.46 Time Series Trend

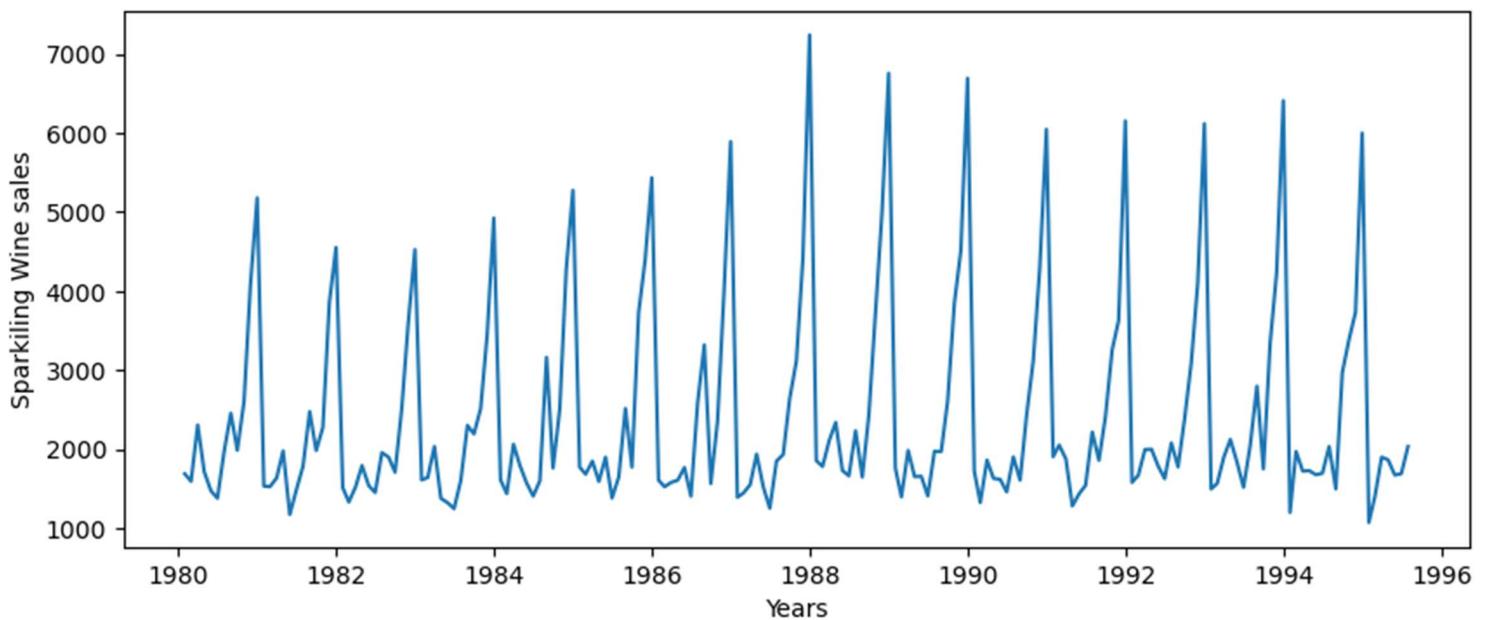
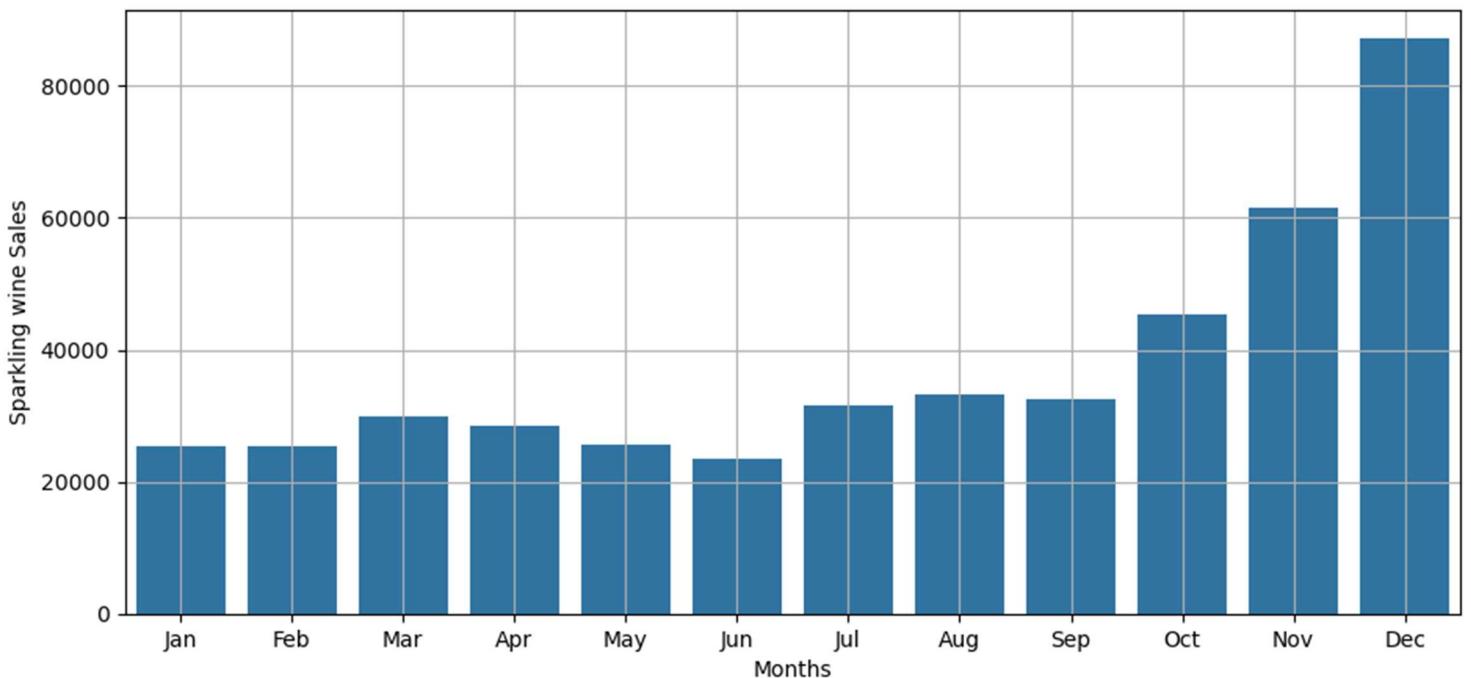


Fig.47 Monthly Sales



- **January, February and June** months have the lowest sales and can be boosted by giving limited time discounts after holiday season.
- The Sparkling Wine Sales dataset **slightly shows an increasing trend** in the sales from 1980 to 1995.
- Peak sales can be improved by restocking and optimizing supply chain department.
- By introducing summer campaigns and special offers, sales can be boosted in the initial six months.
- By using data driven analysis and forecasting methods, production and logistics can be improved efficiently.
- The upward trend can be further improved by obtaining frequent customer feedbacks at regular intervals and addressing their requests.
- Growing and emerging markets should be partnered where wine consumption is growing.
- Product packages with eco-friendly measures can introduce new set of customers.
- The Factors behind the **stable trend** can be studied by obtaining frequent customer feedbacks at regular intervals and addressing their requests.
- Discounts and offers can be provided for bulk orders in the initial six months.