TIME SERIES FORECASTING

Project Report

Part - 1

Submitted by,

Sindhu R Udupa.

BATCH: PGPDSBA.O. NOV22.B

Contents

List of Figures

List of Tables

Part 1: Report on the sales of the Sparkling wine.

The data contains the information on the sales of a wine called ‘Sparkling’ from January 1980 to July 1995.

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

The data contains two columns namely ‘Year Month’ and ‘Sparkling’. The data has been read as an appropriate time series by using the ‘parse date’ parameter in the read function of the pandas, which will help us identify the column ‘Year Month’ as date. We have set it as the index of our data frame, which makes it easier to analyze the data. The column ‘Sparkling’ contains information about the sales of that wine in that particular month.

The plot the time series is as given below.

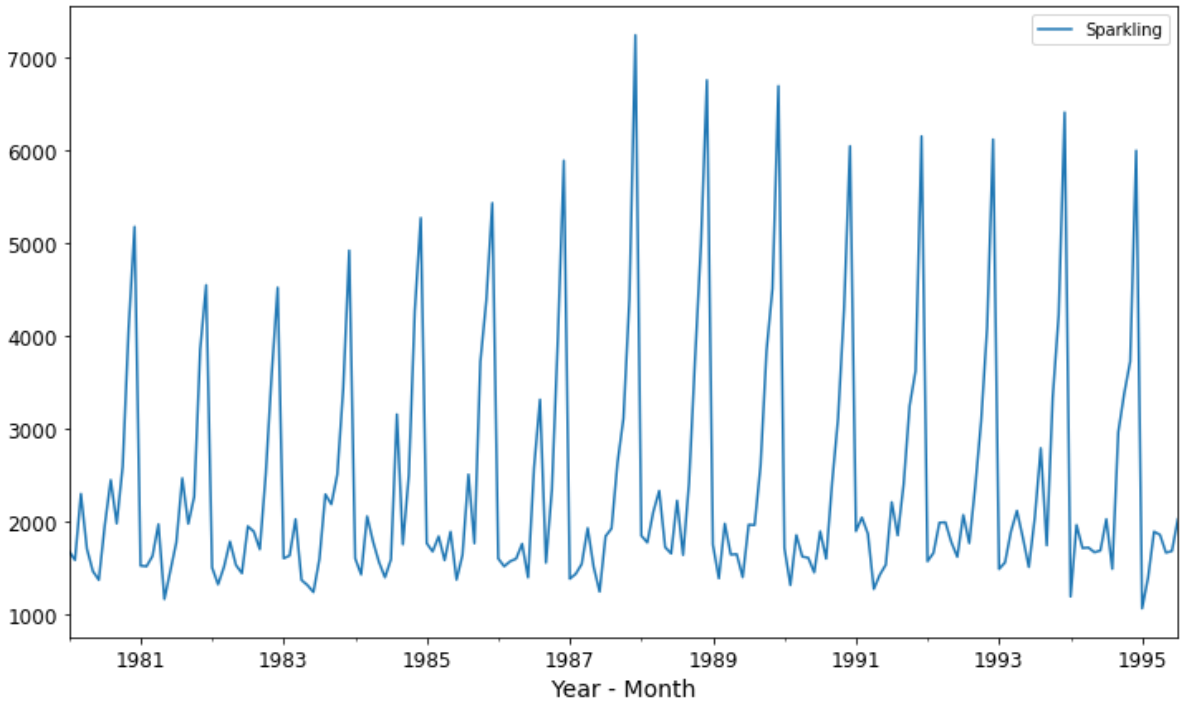

Fig. 1: Plot of the sales of the wine Sparkling.

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

Exploratory Data Analysis:

a. After making the Year Month column as index, there is only one column named ‘Sparkling’ of integer data type, which depicts the number of sales of the wine in that particular month.
b. There are no duplicate values present in the data.
c. The data starts from 1980 Jan and ends in 1995 July.

d. There are no missing values in the data.
e. The monthly sales, yearly sales of the wine are as follows:

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

Table 1: Monthly Sales.

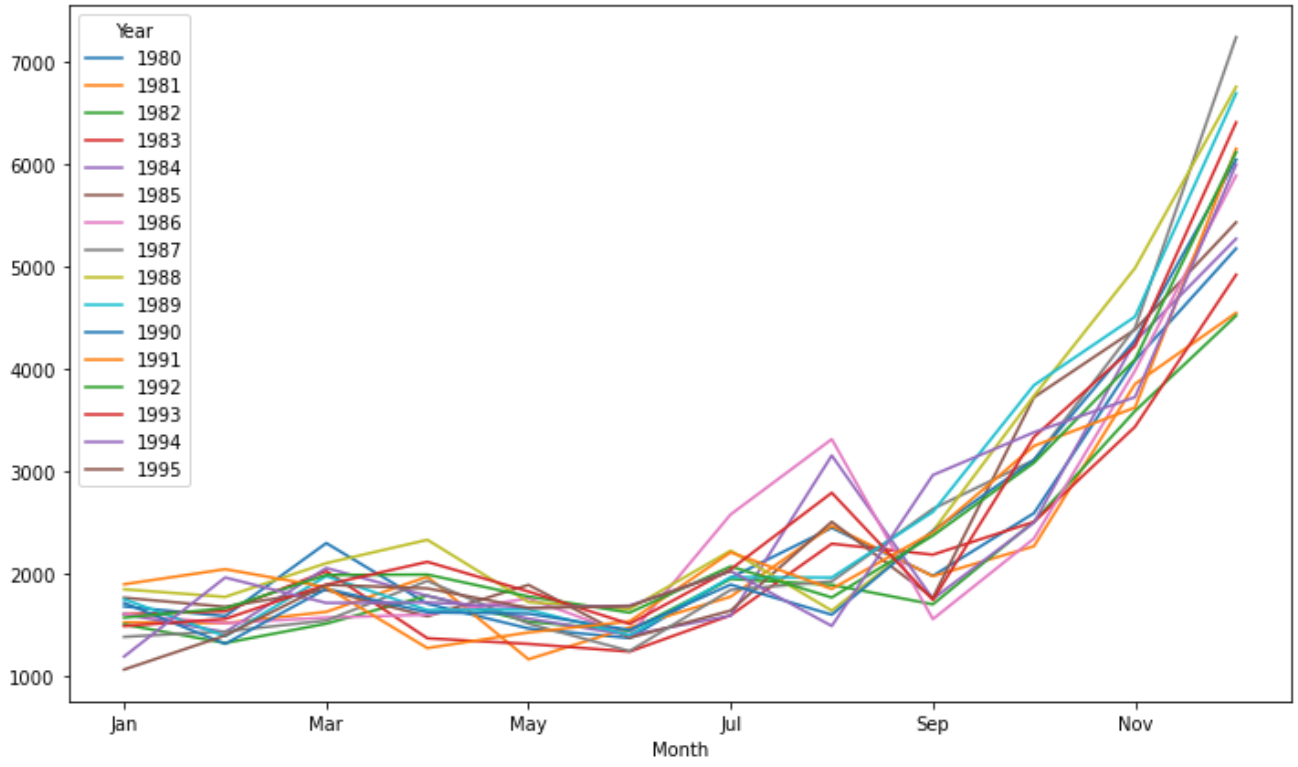

Fig. 2: Plot of the monthly sales.

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

Table 2: Yearly Sales.

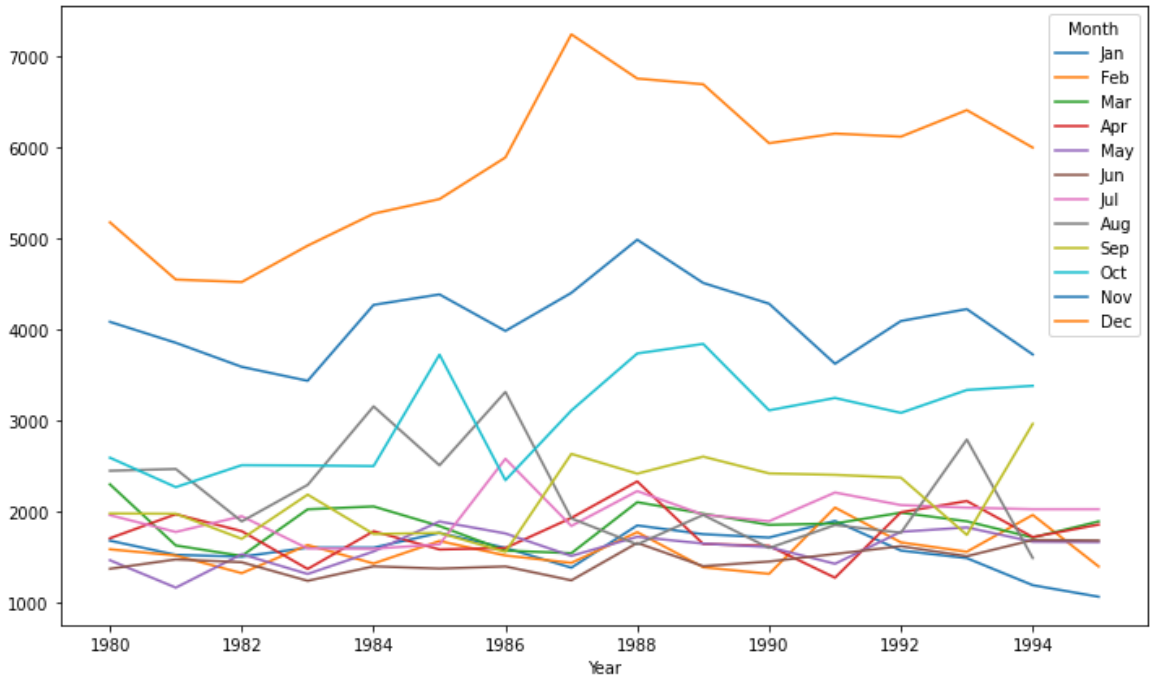

Fig. 3: Plot of the yearly sales.

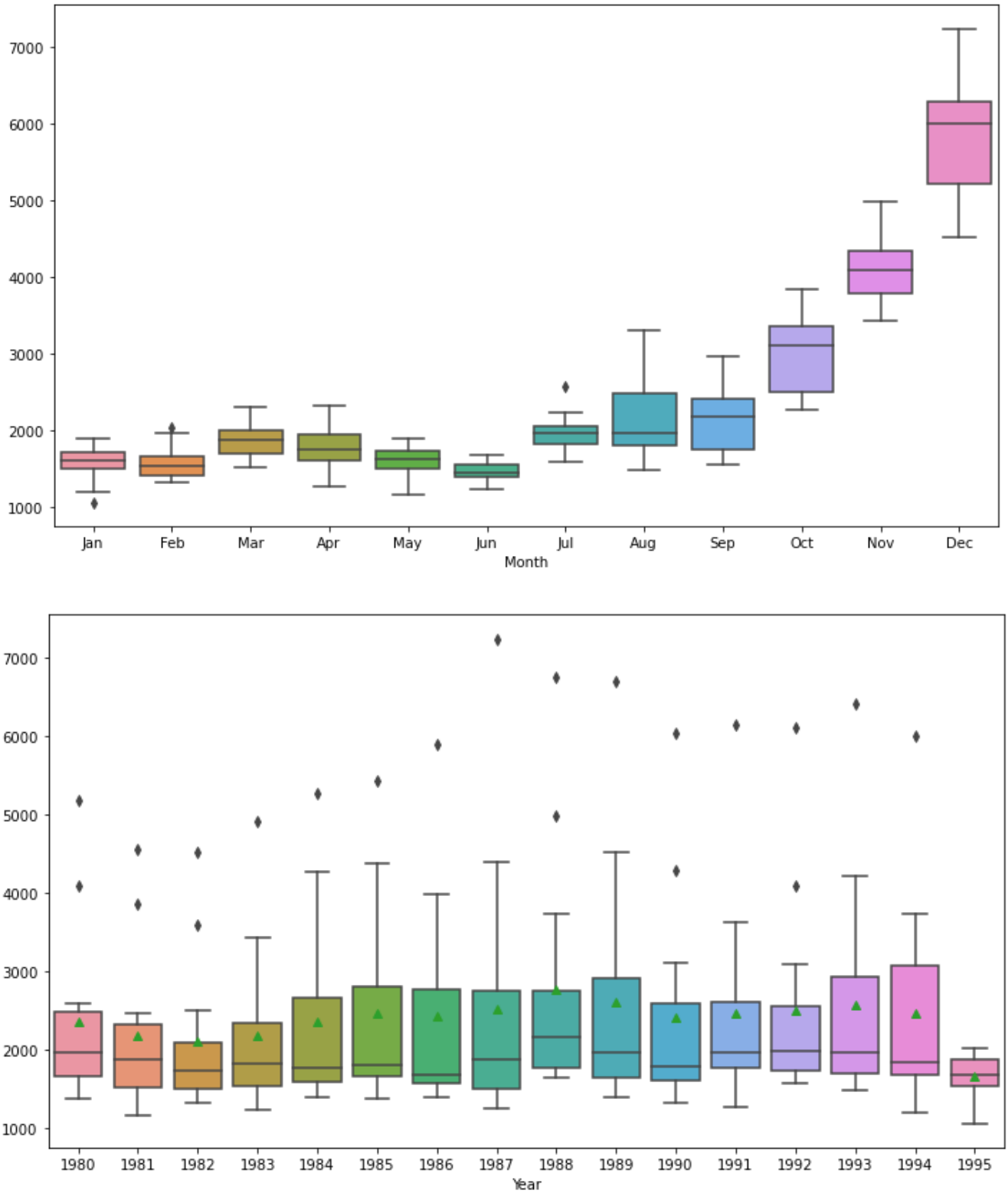

Fig. 4: Box plot of the monthly and yearly sales.

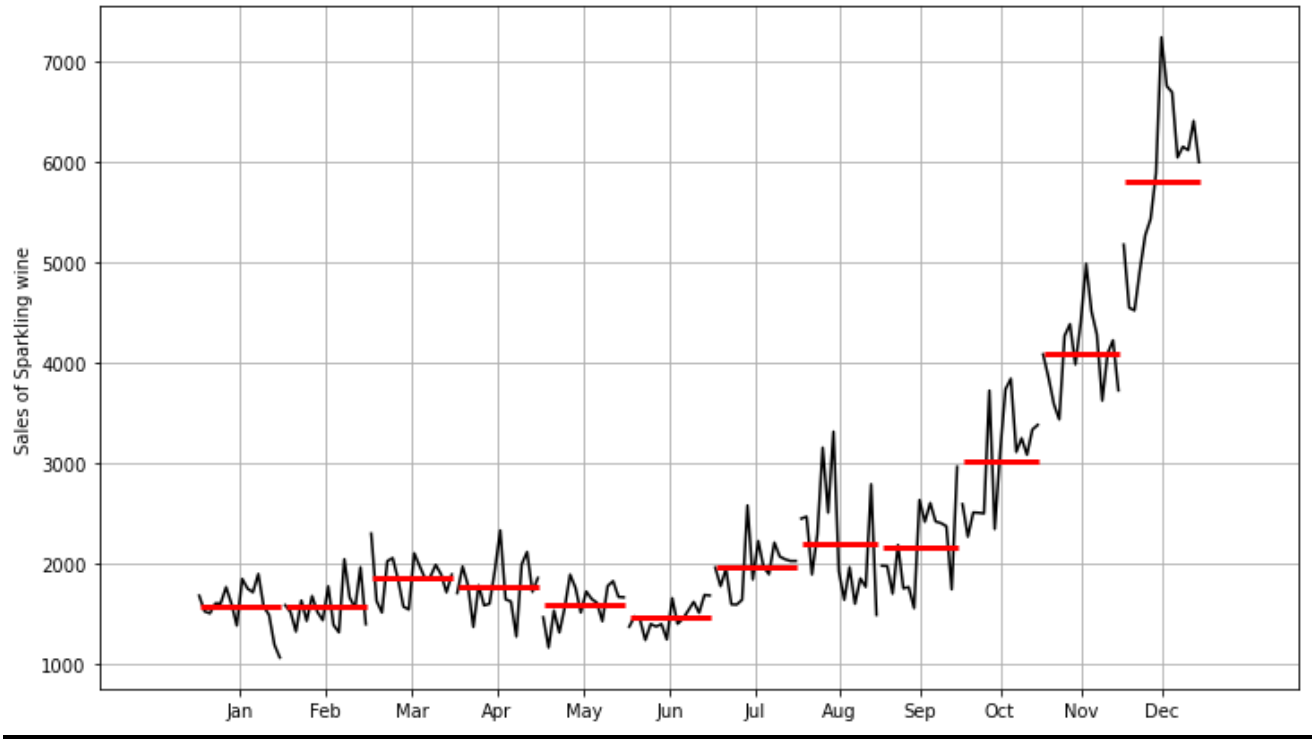

Fig. 5: Month plot of the sales.

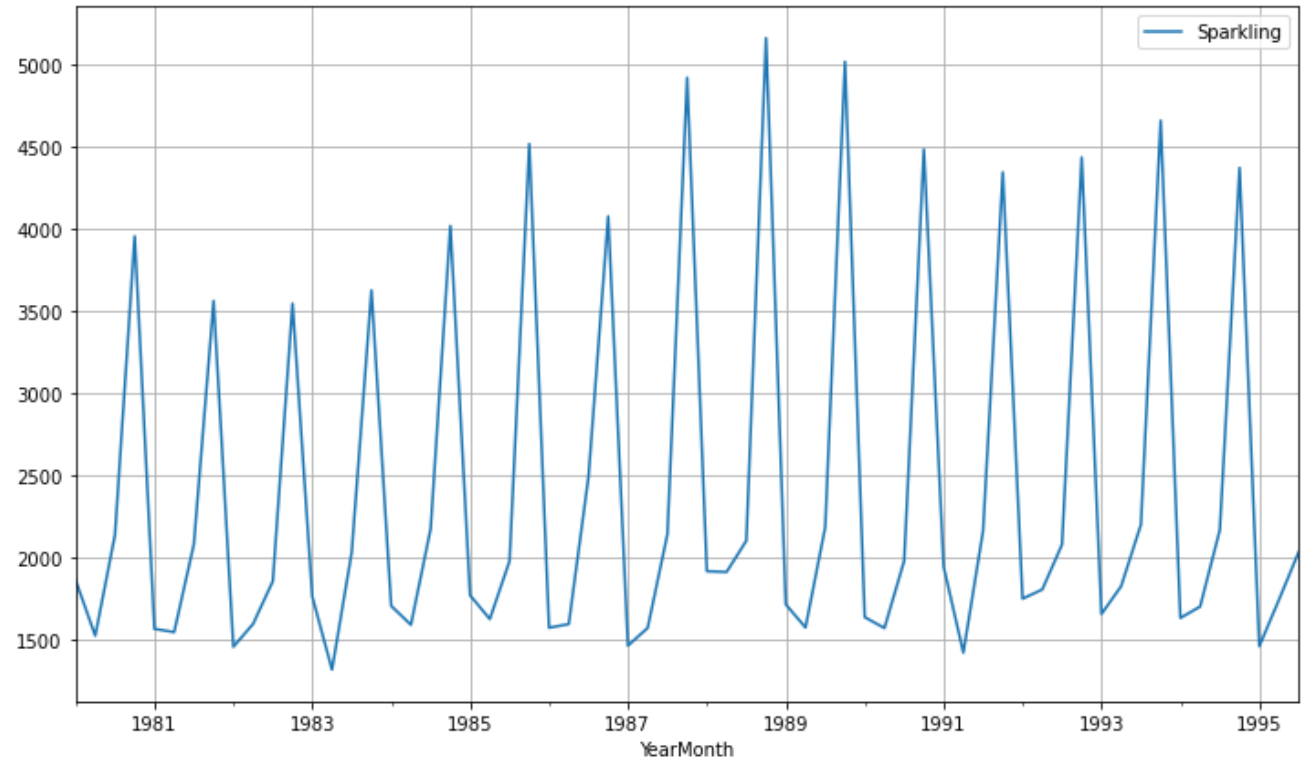

Fig. 6: Quarterly plot of the sales.

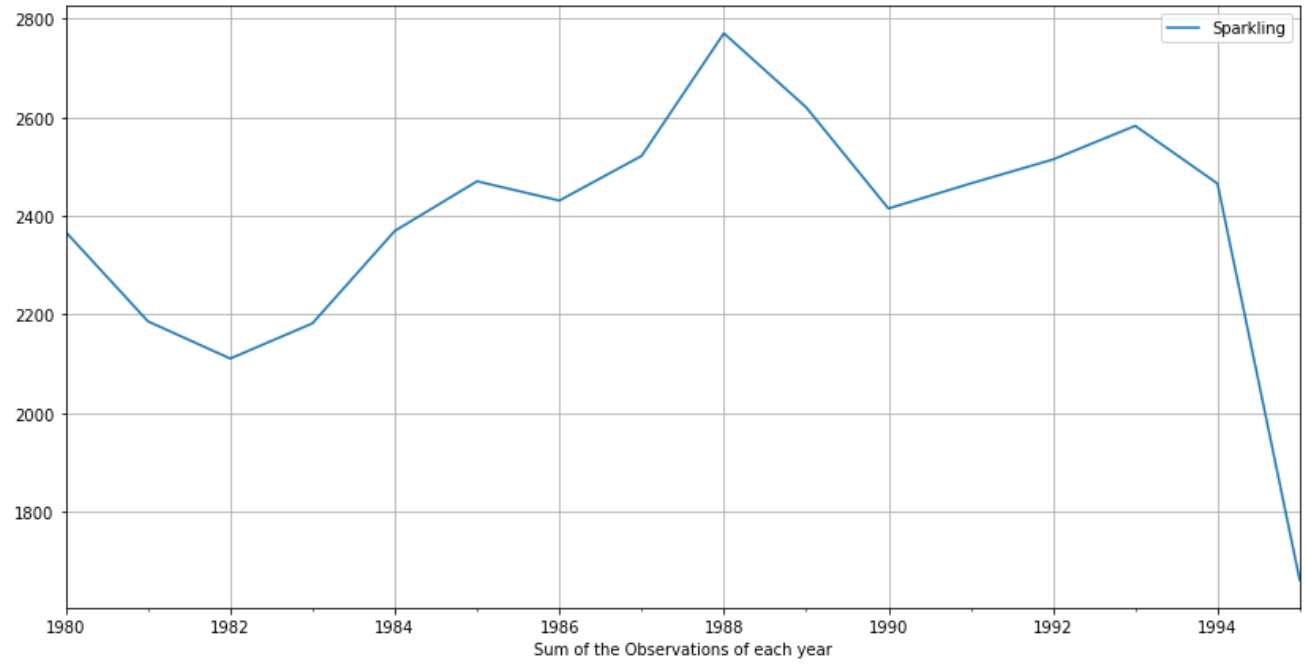

Fig. 7: Yearly plot of the sales.

. From the above tables and graphs we can see that

a. The sales of the wine have been highest in the month of December in each year.
b. The average sales of all the year are in the same range between 2200 – 2800.
c. In the box plot we can see that the average sale of wine in the year 1995 is less than the other years. But since we have data only till 1995 July, this is obvious.
d. From the quarterly plot, we can see that sales are at a peak in the last quarter of each year.
e. From the yearly plot, the highest sale has happened in the year 1988 and the lowest sale happened in the year 1982.

Additive Decomposition:

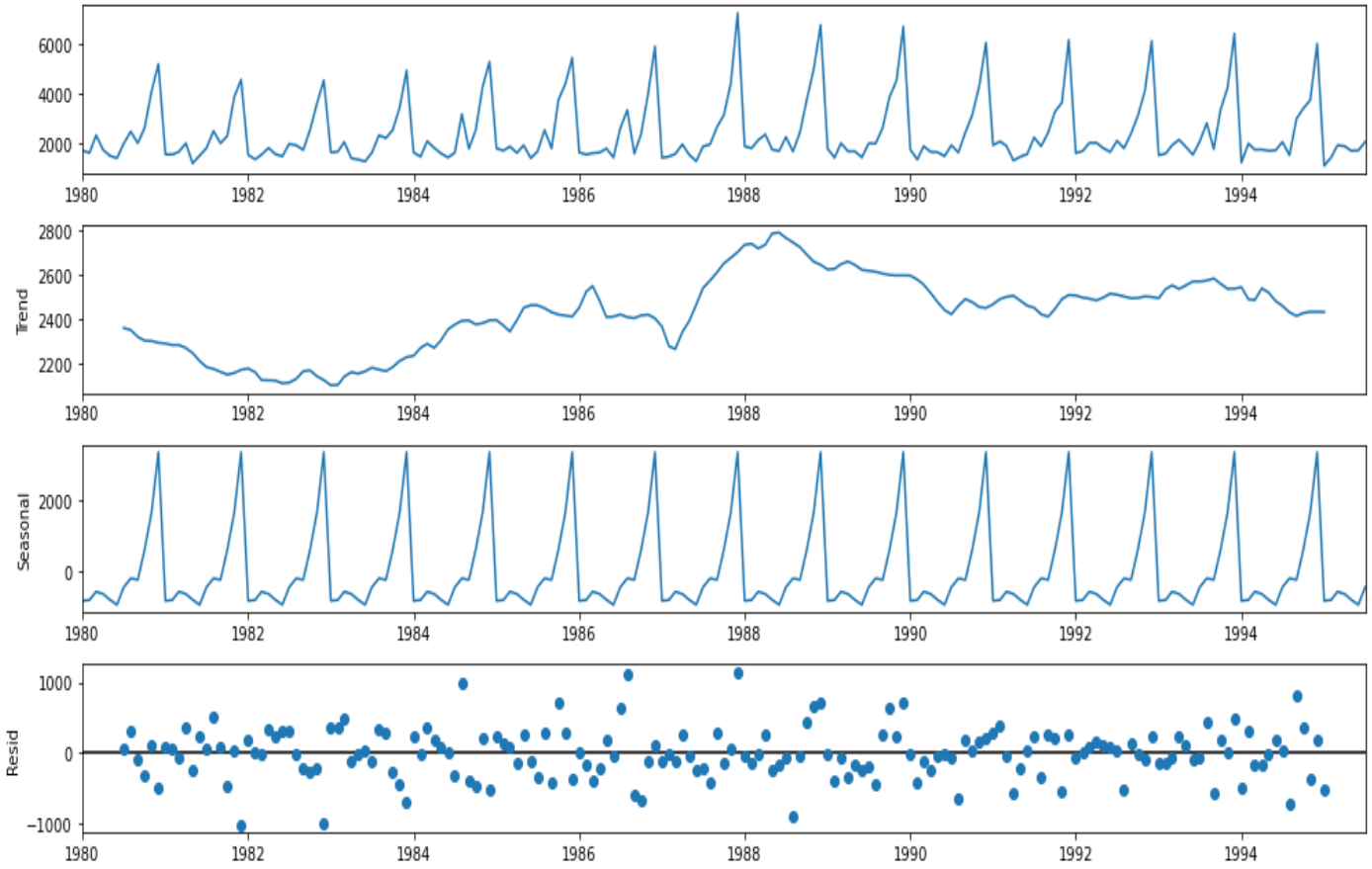

Fig. 8: Additive Decomposition.

The time series is decomposed into three components.

1. Trend

2. Seasonality and

3. Residual Component.

In an additive decomposition, the three components are added to get the time series.

Multiplicative Decomposition:

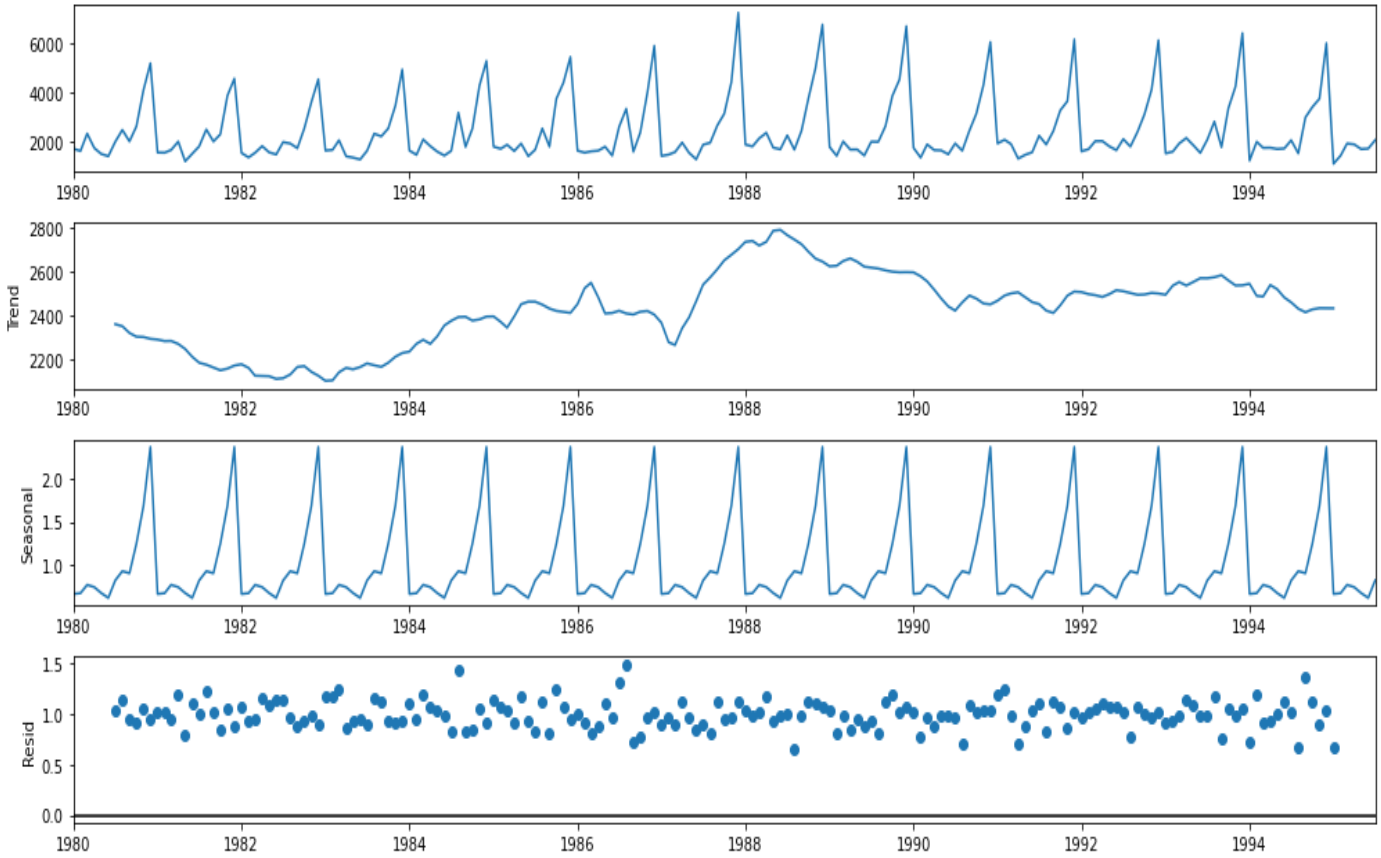

Fig. 9: Multiplicative Decomposition.

In the multiplicative decomposition, the three components – trend, seasonality and the residuals are multiplied to get the time series. The random component or the residual component should not follow any pattern, we see they do not follow any pattern in both the additive and the multiplicative decomposition. The trend is showing an increasing pattern in both the decompositions.

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

- The data has been split into train and test.
- The train data contains data from 1980 Jan to 1989 Dec.
- There are total of 132 data points in the train set.
- The test data contains data from 1991 Jan to 1995 July.
- There are total of 55 data points in the test set.

YearMonth	Sparkling
1980-01-01	1686
1980-02-01	1591
1980-03-01	2304
1980-04-01	1712
1980-05-01	1471
...	...
1990-08-01	1605
1990-09-01	2424
1990-10-01	3116
1990-11-01	4286
1990-12-01	6047

Table 3: Train Set

YearMonth	Sparkling
1991-01-01	1902
1991-02-01	2049
1991-03-01	1874
1991-04-01	1279
1991-05-01	1432
...	...
1995-03-01	1897
1995-04-01	1862
1995-05-01	1670
1995-06-01	1688
1995-07-01	2031

Table 4: Test Set

4. Build all the 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 and simple average models. should also be built on the training data and check the performance on the test data using RMSE.

Regression Model

We are building a linear regression model using the sklearn library. First, we assigned integers to each data point in order to build the model. The sample is as follows.

First few rows of Training Data

YearMonth	Sparkling	time
1980-01-01	1686	1
1980-02-01	1591	2
1980-03-01	2304	3
1980-04-01	1712	4
1980-05-01	1471	5

Last few rows of Training Data

YearMonth	Sparkling	time
1990-08-01	1605	128
1990-09-01	2424	129
1990-10-01	3116	130
1990-11-01	4286	131
1990-12-01	6047	132

First few rows of Test Data

YearMonth	Sparkling	time
1991-01-01	1902	133
1991-02-01	2049	134
1991-03-01	1874	135
1991-04-01	1279	136
1991-05-01	1432	137

Last few rows of Test Data

YearMonth	Sparkling	time
1995-03-01	1897	183
1995-04-01	1862	184
1995-05-01	1670	185
1995-06-01	1688	186
1995-07-01	2031	187

Table 5: Time points in train and test data.

Predictions on the test set:

The below table gives the predicted values of the Regression model on the first ten data points of the test set.

YearMonth	Actual Values	Prediction
1991-01-01	1902	2791.652093
1991-02-01	2049	2797.484752
1991-03-01	1874	2803.317410
1991-04-01	1279	2809.150069
1991-05-01	1432	2814.982727
1991-06-01	1540	2820.815386
1991-07-01	2214	2826.648044
1991-08-01	1857	2832.480703
1991-09-01	2408	2838.313361
1991-10-01	3252	2844.146020

Table 6: Predictions of the Linear Regression model.

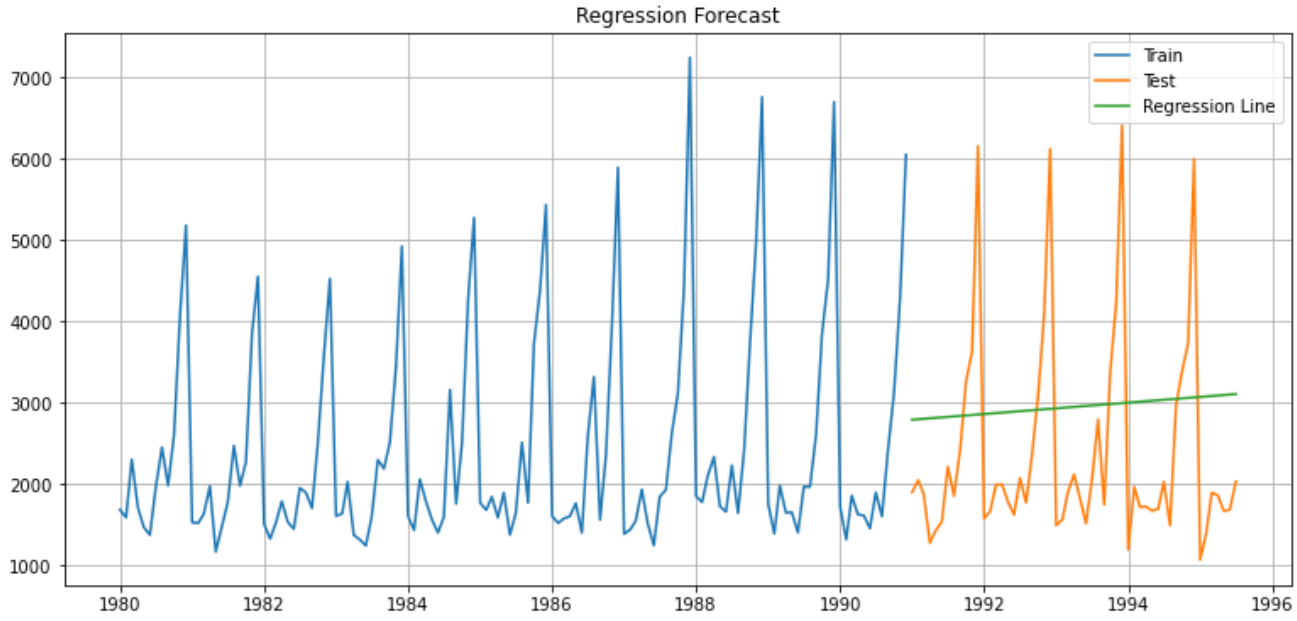

Fig. 10: Predictions of the Linear Regression model.

Model Evaluation:

The RMSE on the test set comes out to be 1389.135.

Naïve Forecast Model

This model naively forecasts the last recorded data to the future. So, the prediction of this model on the test data is the number of sales happened in Dec. 1990.

Predictions on the test set:

The below table gives the predicted values of the Naïve Forecast model on the first ten data points of the test set.

YearMonth	Actual Values	Prediction
1991-01-01	1902	6047
1991-02-01	2049	6047
1991-03-01	1874	6047
1991-04-01	1279	6047
1991-05-01	1432	6047
1991-06-01	1540	6047
1991-07-01	2214	6047
1991-08-01	1857	6047
1991-09-01	2408	6047
1991-10-01	3252	6047

Table 7: Predictions of the Naïve Forecast model.

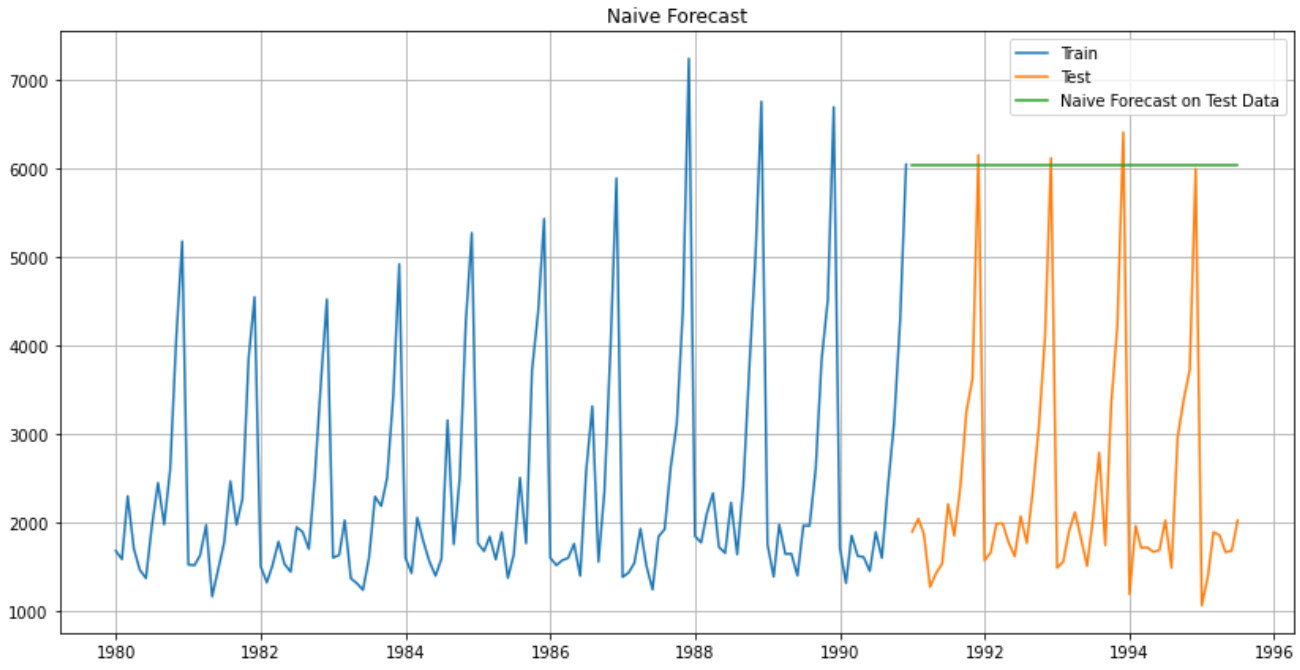

Fig. 11: Predictions of the Naïve Forecast model.

Model Evaluation:

The RMSE on the test set comes out to be 3864.279.

Simple Average Model

This model forecasts the average of the previous values to the future.

Predictions on the test set:

The below table gives the predicted values of the Simple Average model on the first ten data points of the test set.

YearMonth	Actual Values	Prediction
1991-01-01	1902	2403.780303
1991-02-01	2049	2403.780303
1991-03-01	1874	2403.780303
1991-04-01	1279	2403.780303
1991-05-01	1432	2403.780303
1991-06-01	1540	2403.780303
1991-07-01	2214	2403.780303
1991-08-01	1857	2403.780303
1991-09-01	2408	2403.780303
1991-10-01	3252	2403.780303

Table 8: Predictions of the Simple Average model.

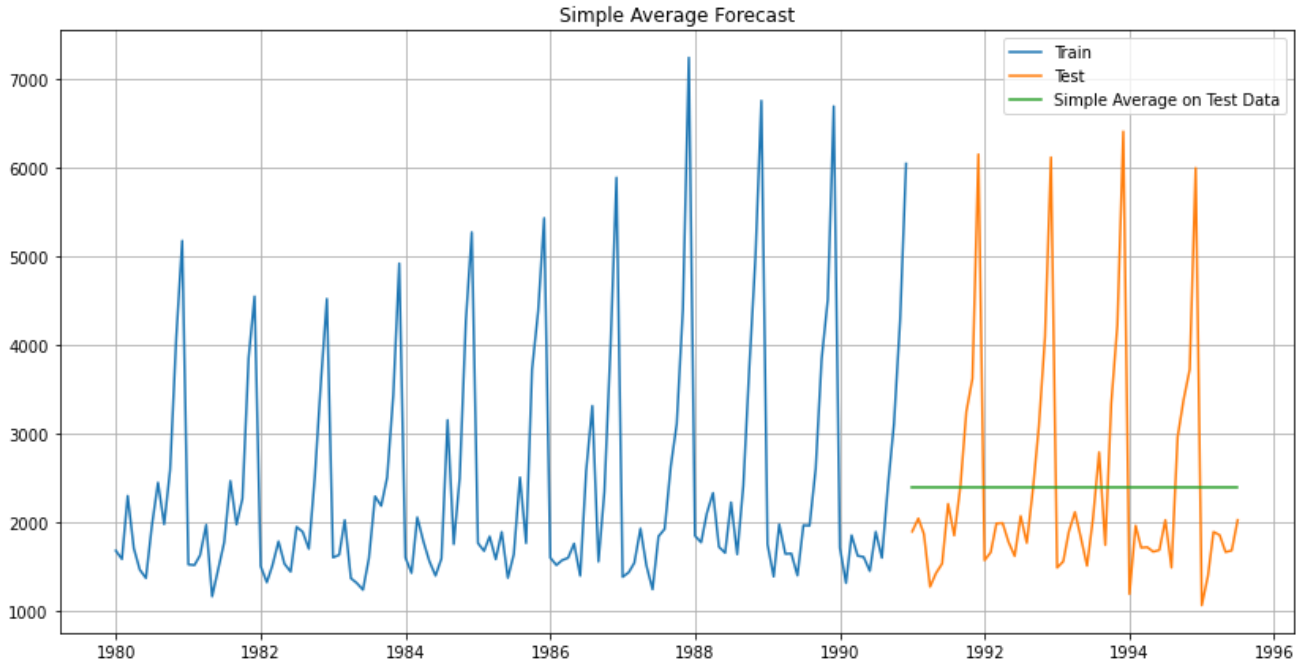

Fig. 12: Predictions of the Simple Average model.

Model Evaluation:

The RMSE on the test set comes out to be 1275.082.

Moving Average Model

For the moving average model, we are going to calculate rolling means for different intervals. The best interval can be determined by the minimum error.

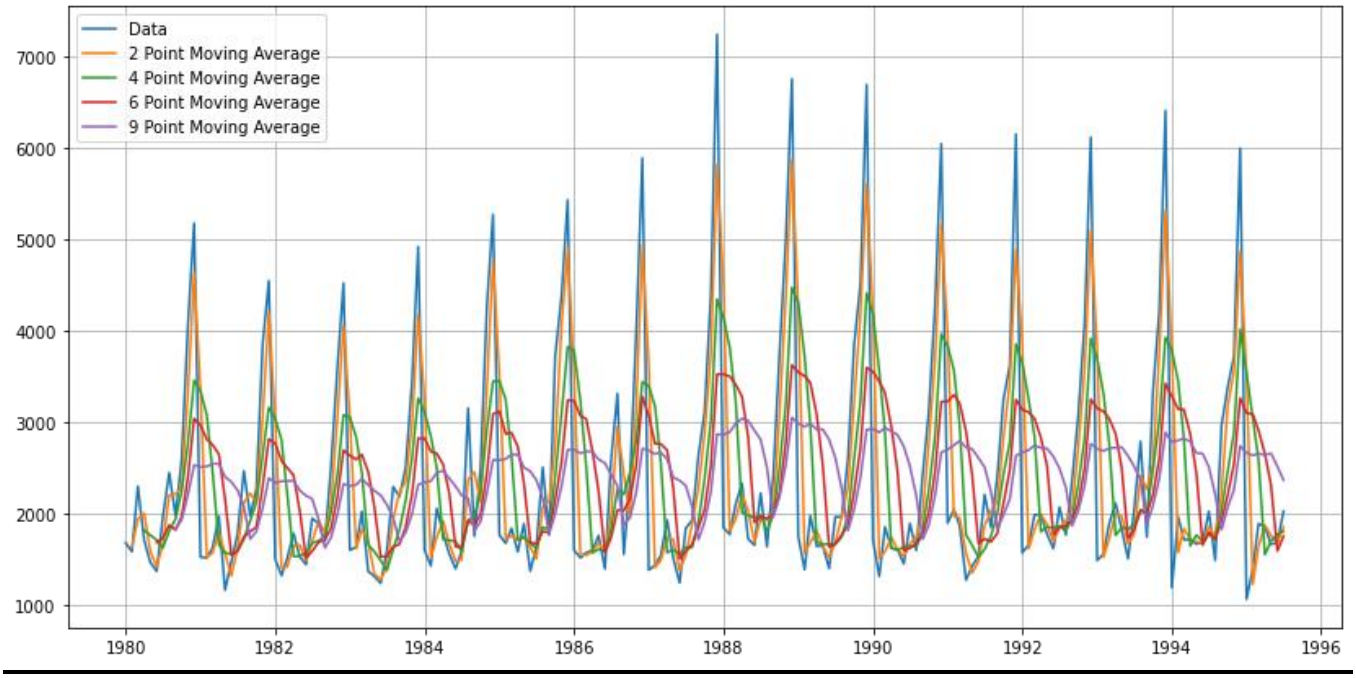

Fig. 13: Moving Average on the whole data:

Predictions on the test set:

YearMonth	Sparkling	Trailing_2	Trailing_4	Trailing_6	Trailing_9
1991-01-01	1902	3974.5	3837.75	3230.000000	2705.666667
1991-02-01	2049	1975.5	3571.00	3304.000000	2753.888889
1991-03-01	1874	1961.5	2968.00	3212.333333	2800.222222
1991-04-01	1279	1576.5	1776.00	2906.166667	2731.333333
1991-05-01	1432	1355.5	1658.50	2430.500000	2712.111111
1991-06-01	1540	1486.0	1531.25	1679.333333	2613.888889
1991-07-01	2214	1877.0	1616.25	1731.333333	2513.666667
1991-08-01	1857	2035.5	1760.75	1699.333333	2243.777778
1991-09-01	2408	2132.5	2004.75	1788.333333	1839.444444
1991-10-01	3252	2830.0	2432.75	2117.166667	1989.444444

Table 9: Predictions of the Moving Average model.

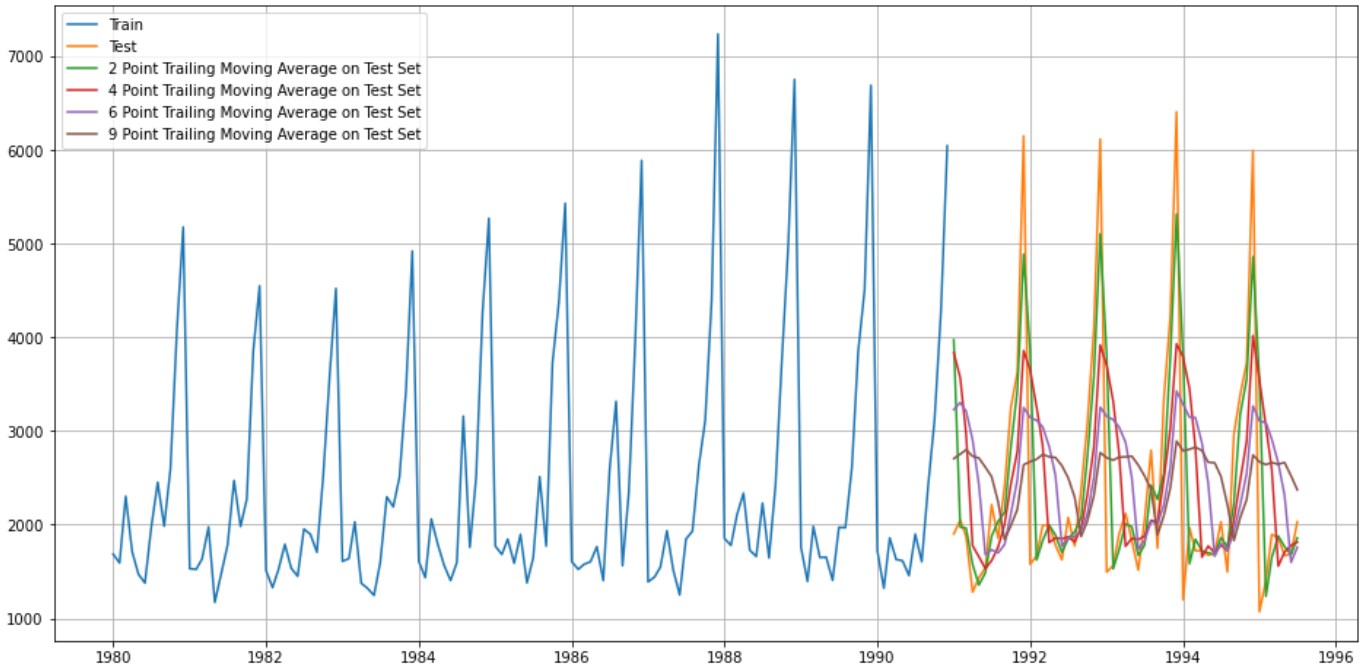

Fig. 14: Predictions of the Moving Average model.

Model Evaluation:

For 2 point Moving Average Model forecast on the Training Data, RMSE is 813.401
For 4 point Moving Average Model forecast on the Training Data, RMSE is 1156.590
For 6 point Moving Average Model forecast on the Training Data, RMSE is 1283.927
For 9 point Moving Average Model forecast on the Training Data, RMSE is 1346.278

Simple Exponential Smoothing Model

The level is taken into account while building the simple exponential smoothing model. The parameter considered here is alpha.

Predictions on the test set:

The below table gives the predicted values of the Simple Average model on the first ten data points of the test set.

YearMonth	Actual Values	Prediction
1991-01-01	1902	2804.675124
1991-02-01	2049	2804.675124
1991-03-01	1874	2804.675124
1991-04-01	1279	2804.675124
1991-05-01	1432	2804.675124
1991-06-01	1540	2804.675124
1991-07-01	2214	2804.675124
1991-08-01	1857	2804.675124
1991-09-01	2408	2804.675124
1991-10-01	3252	2804.675124

Table 10: Predictions of the Simple Exponential Smoothing model.

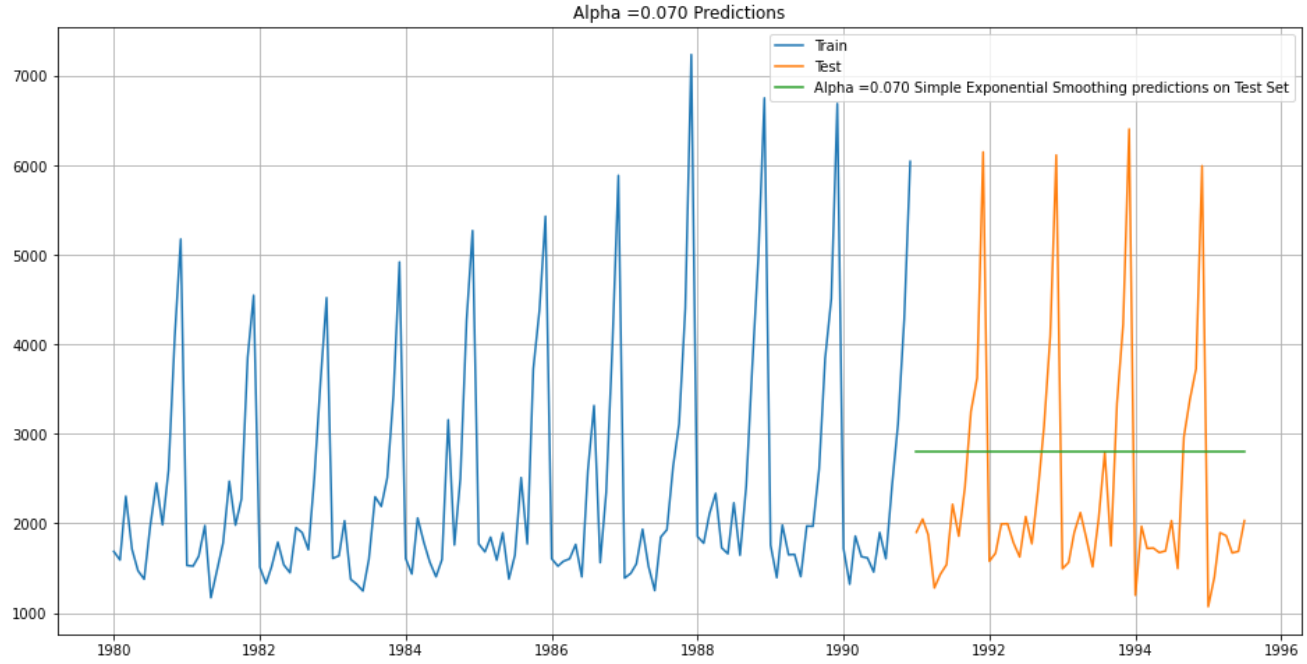

Fig. 15: Predictions of the Simple Exponential Smoothing model.

Model Evaluation:

The RMSE on the test set comes out to be 1338.008.

Double Exponential Smoothing – Holt Model

The level and trend are taken into account while building the double exponential smoothing model. The parameter considered here is alpha and beta.

Predictions on the test set:

The below table gives the predicted values of the Holt model on the first ten data points of the test set.

YearMonth	Actual Values	Prediction
1991-01-01	1902	5401.733026
1991-02-01	2049	5476.005230
1991-03-01	1874	5550.277433
1991-04-01	1279	5624.549637
1991-05-01	1432	5698.821840
1991-06-01	1540	5773.094044
1991-07-01	2214	5847.366248
1991-08-01	1857	5921.638451
1991-09-01	2408	5995.910655
1991-10-01	3252	6070.182858

Table 11: Predictions of the Holt model.

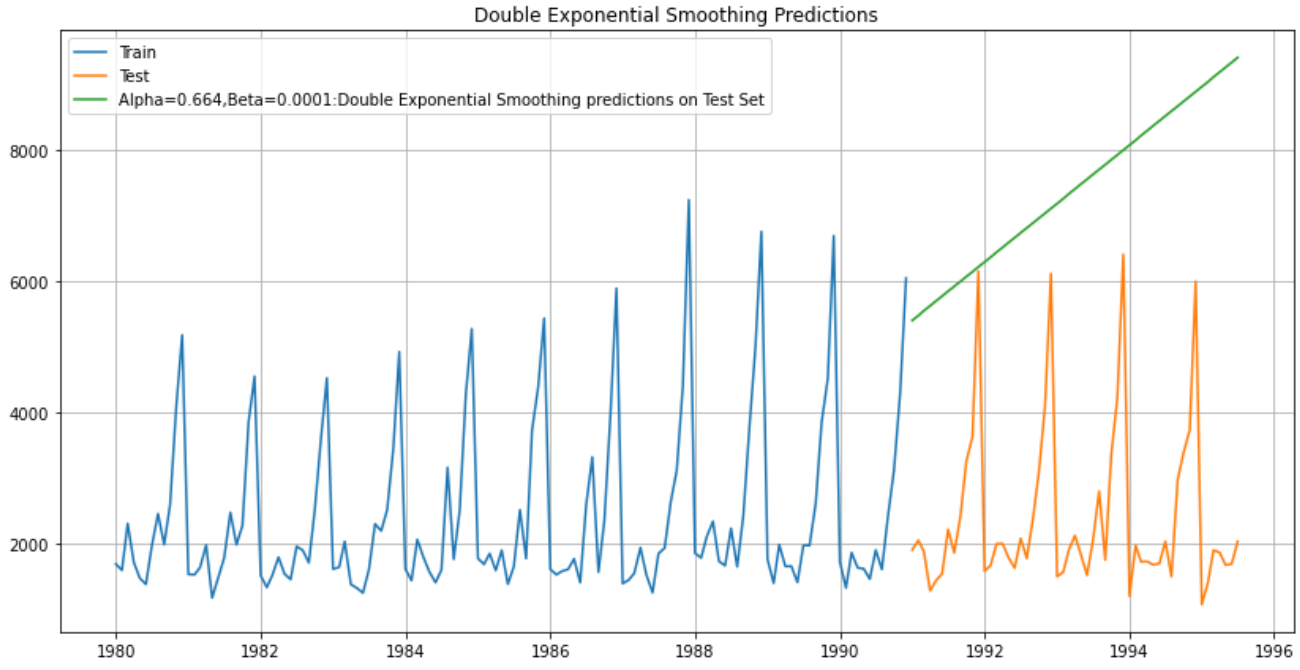

Fig. 16: Predictions of the Holt model.

Model Evaluation:

The RMSE on the test set comes out to be 5291.879.

Triple Exponential Smoothing – Holt Winter Model

Multiplicative seasonality

The level, trend and seasonality are taken into account while building the triple exponential smoothing model. The parameter considered here is alpha, beta and gamma.

Predictions on the test set:

The below table gives the predicted values of the Holt Winter model on the first ten data points of the test set.

YearMonth	Actual Values	Prediction
1991-01-01	1902	1587.497468
1991-02-01	2049	1356.394925
1991-03-01	1874	1762.929755
1991-04-01	1279	1656.165933
1991-05-01	1432	1542.002730
1991-06-01	1540	1355.102435
1991-07-01	2214	1854.197719
1991-08-01	1857	1820.513188
1991-09-01	2408	2276.971718
1991-10-01	3252	3122.024202

Table 12: Predictions of the Holt Winter model with multiplicative seasonality.

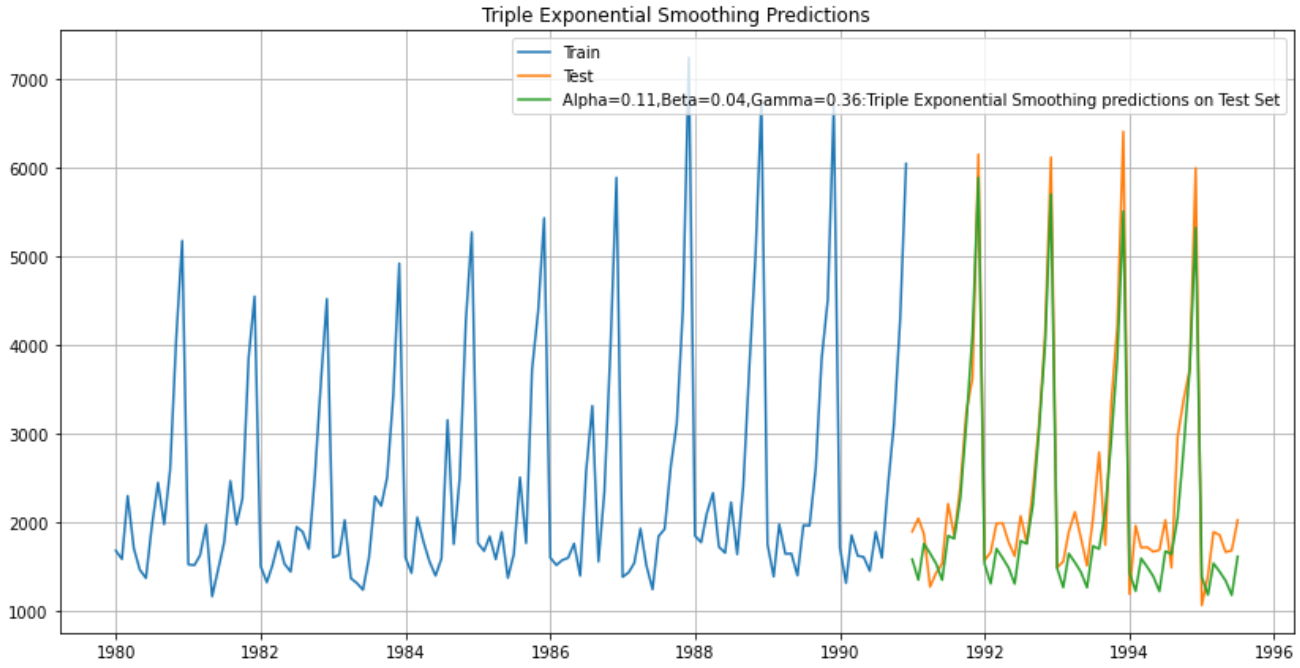

Fig. 17: Predictions of the Holt Winter model with multiplicative seasonality

Model Evaluation:

The RMSE on the test set comes out to be 404.286.

Triple Exponential Smoothing – Holt Winter Model

Additive seasonality

The level, trend and seasonality are taken into account while building the triple exponential smoothing model. The parameter considered here is alpha, beta and gamma.

Predictions on the test set:

The below table gives the predicted values of the Holt Winter model on the first ten data points of the test set.

YearMonth	Actual Values	Prediction
1991-01-01	1902	1490.402890
1991-02-01	2049	1204.525152
1991-03-01	1874	1688.734182
1991-04-01	1279	1551.226125
1991-05-01	1432	1461.197883
1991-06-01	1540	1278.646707
1991-07-01	2214	1804.885616
1991-08-01	1857	1678.955032
1991-09-01	2408	2315.373126
1991-10-01	3252	3224.976222

Table 13: Predictions of the Holt Winter model with additive seasonality.

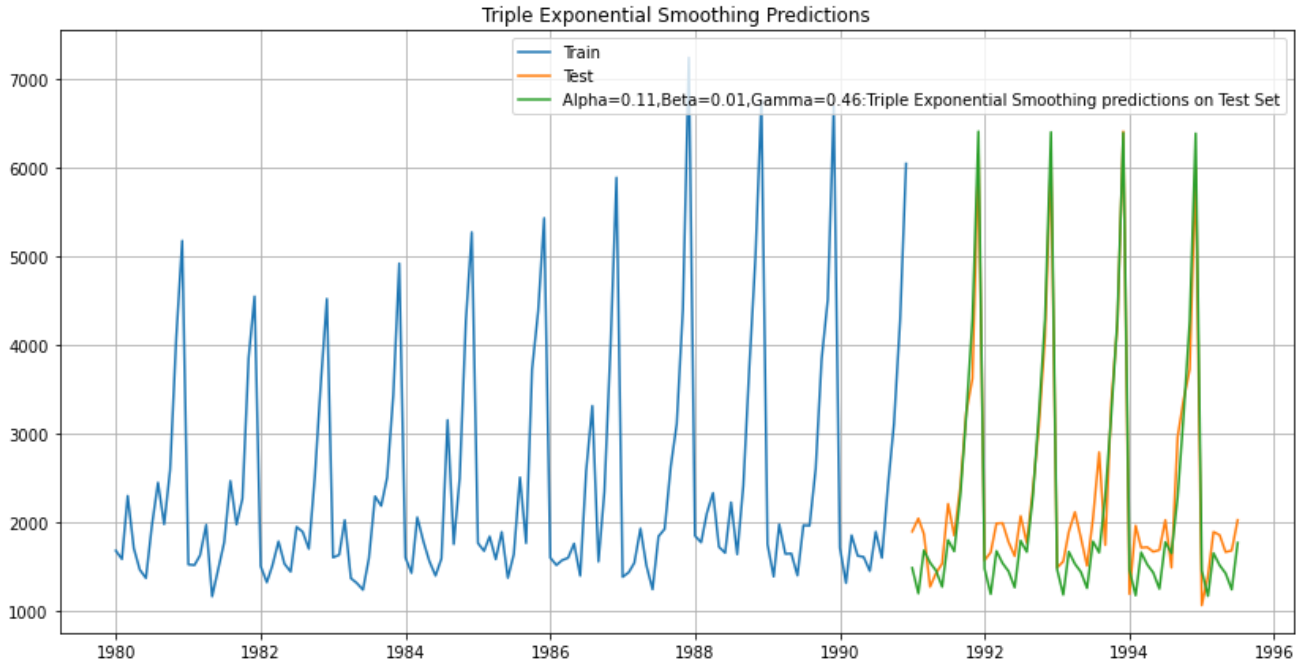

Fig. 18: Predictions of the Holt Winter model with additive seasonality

Model Evaluation:

The RMSE on the test set comes out to be 378.951.

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.

Stationarity Check:

Dickey-Fuller Test – is a statistical test on the timeseries to check for stationarity of data.

- Null Hypothesis - H_o: Time Series is non-stationary.
- Alternate Hypothesis – H_a: Time Series is stationary.

If p-value < alpha = 0.05 then null hypothesis is rejected else we fail to reject the null hypothesis.

When we run the Dickey-Fuller Test on our time series, we find the p-value to be 0.6011, which is greater than 0.05. Hence we fail to reject the null hypothesis. That is the time series is not stationary.

To make the series stationary, we shall difference the series once, i.e., d=1. Then the time series looks as below:

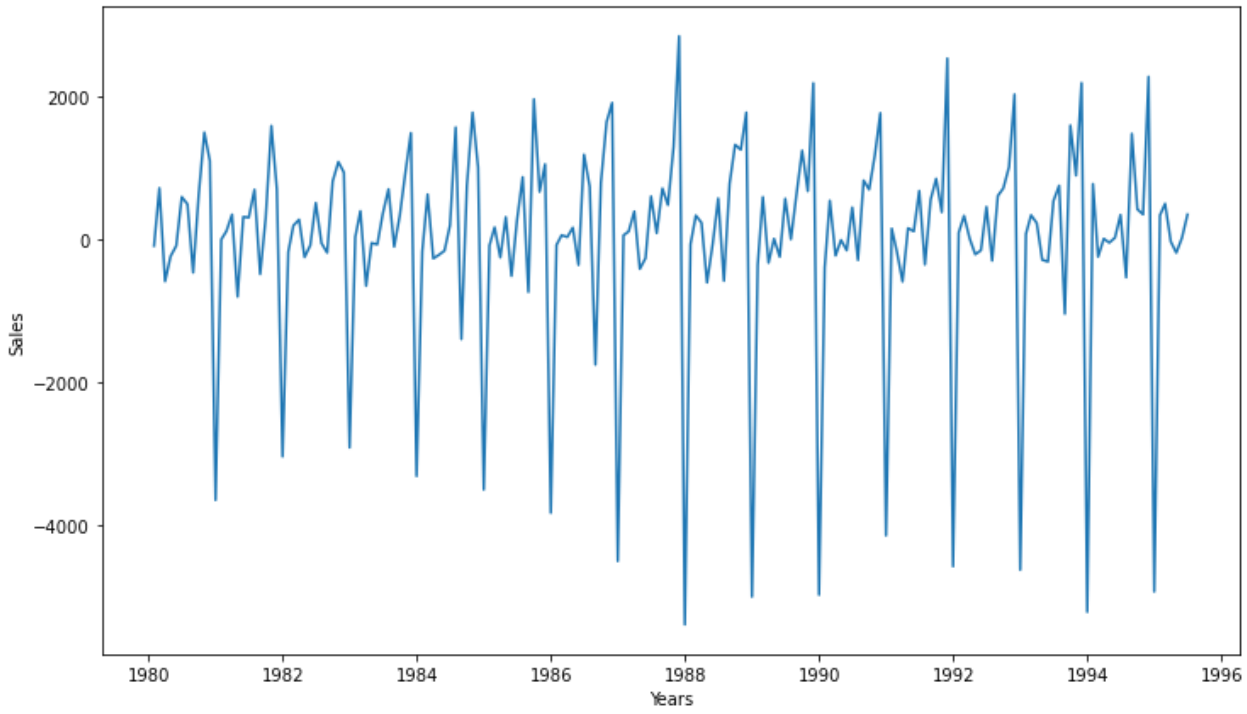

Fig. 19: First order differentiated time series.

To check if the differentiated new series is stationary or not, we run the Dickey Fuller test on the new series again and find that p-value is 0.000, which is less than 0.05. Hence, we can say that the new series is stationary.

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.

ARIMA

The ARIMA model takes three parameters into account. The parameters are p, d and q. The parameter d depicts the order of the differencing that makes the series stationary. The parameter p is the order of the auto regression model and the parameter q is the order of the moving average model. Different combinations of p, d and q are considered and the one with the lowest AIC value is considered to build the model.

Different parameter combinations and AIC:

	param	AIC
8	(2, 1, 2)	2213.509212
7	(2, 1, 1)	2233.777626
2	(0, 1, 2)	2234.408323
5	(1, 1, 2)	2234.5272
4	(1, 1, 1)	2235.755095
6	(2, 1, 0)	2260.365744
1	(0, 1, 1)	2263.060016
3	(1, 1, 0)	2266.608539
0	(0, 1, 0)	2267.663036

Table 14: Parameter combinations and AIC values.

We shall build an ARIMA model with p=2, d=1, q=2, since this gives the least AIC.

SARIMAX Results

Dep. Variable:	Sparkling	No. Observations:	132
Model:	ARIMA(2, 1, 2)	Log Likelihood	-1101.755
Date:	Tue, 08 Aug 2023	AIC	2213.509
Time:	04:47:12	BIC	2227.885
Sample:	01-01-1980 - 12-01-1990	HQIC	2219.351
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	-7.741	0.000	-0.701	-0.418
ma.L1	-1.9917	0.109	-18.218	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

Table 15: Summary of the ARIMA model.

Predictions on the test set:

The below table gives the predicted values of the ARIMA model on the first ten data points of the test set.

YearMonth	Actual Values	Prediction
1991-01-01	1902	4252.347924
1991-02-01	2049	2863.090133
1991-03-01	1874	2043.973171
1991-04-01	1279	1746.207340
1991-05-01	1432	1813.633672
1991-06-01	1540	2068.642722
1991-07-01	2214	2365.530665
1991-08-01	1857	2612.454374
1991-09-01	2408	2770.397088
1991-10-01	3252	2839.531923

Table 16: Predictions of the ARIMA model.

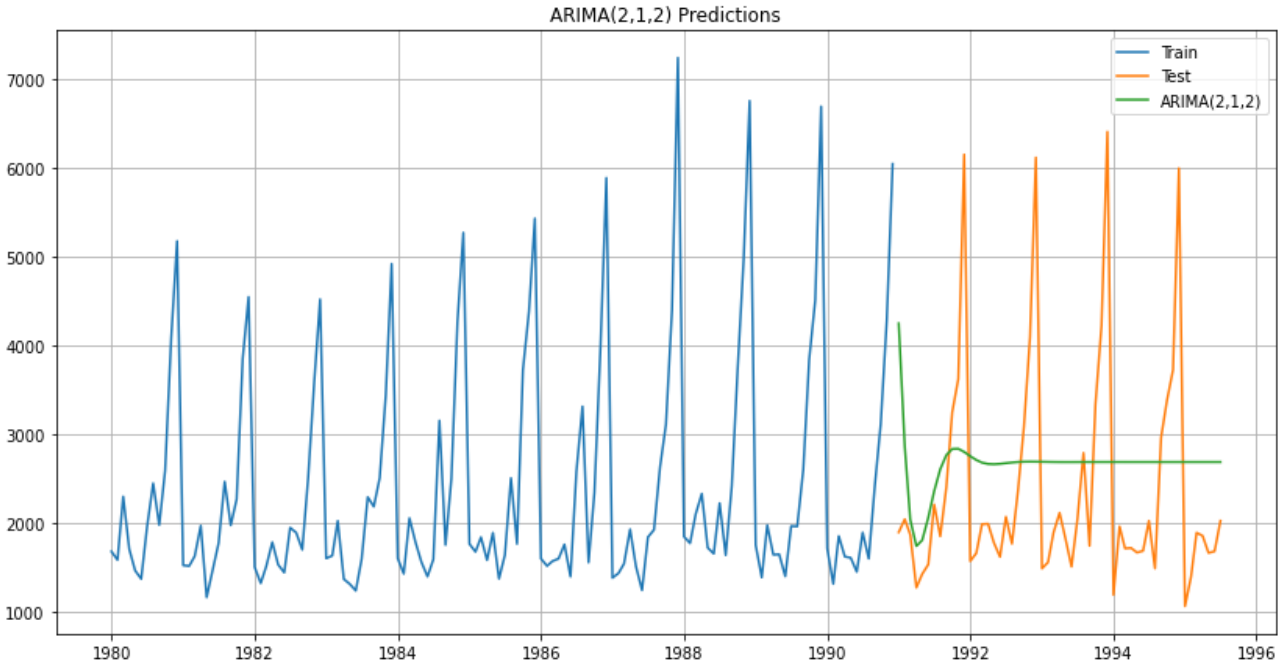

Fig. 20: Predictions of the ARIMA model.

Plot Diagnostics:

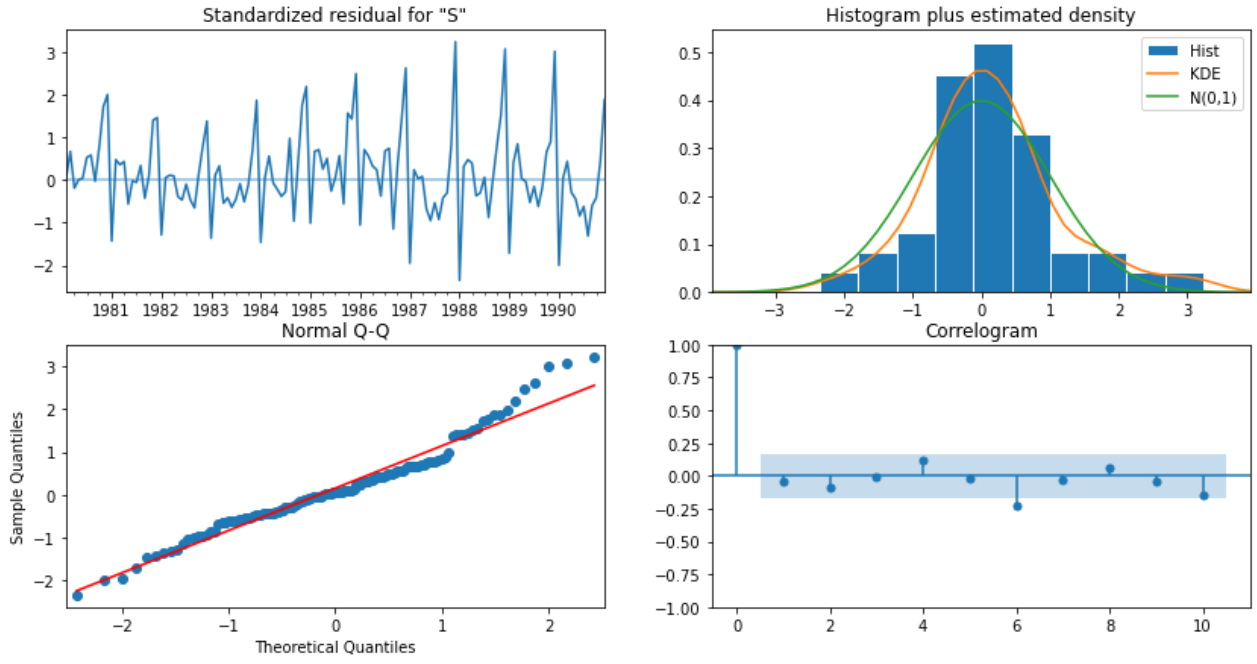

Fig. 21: Plot diagnostics of the ARIMA model.

Model Evaluation:

The RMSE on the test set comes out to be 1299.979.

SARIMA

The SARIMA model, in addition to the p, d, q parameters, it takes into account the seasonal parameters – P, D, Q and S. The parameter S is determined by looking at the auto correlation plot.

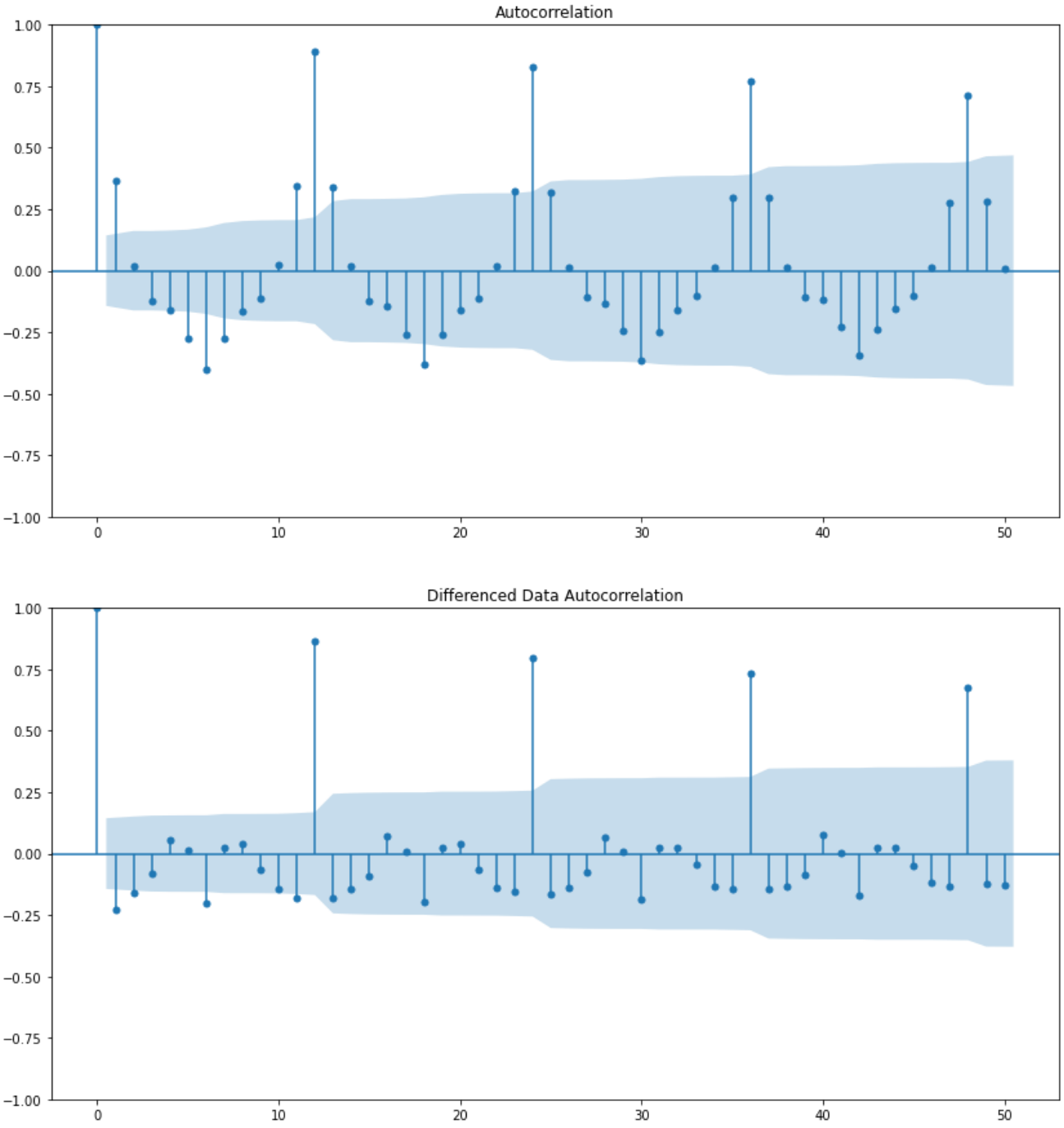

Fig. 22: Auto correlation of the original and differenced series.

It is clear that the seasonal component is 12 as we see a spike in every 12th data point.

Different parameter combinations and AIC:

	param	seasonal	AIC
50	(1, 1, 2)	(1, 0, 2, 12)	1555.584247
53	(1, 1, 2)	(2, 0, 2, 12)	1555.934563
26	(0, 1, 2)	(2, 0, 2, 12)	1557.121564
23	(0, 1, 2)	(1, 0, 2, 12)	1557.160507
77	(2, 1, 2)	(1, 0, 2, 12)	1557.340403

Table 17: Parameter combinations and AIC values.

We shall build a SARIMA model with p=1, d=1, q=2, P=1, D=0, Q=2 and S=12 since this gives the least AIC.

SARIMAX Results

Dep. Variable:	Sparkling	No. Observations:	132
Model:	SARIMAX(1, 1, 2)x(1, 0, 2, 12)	Log Likelihood	-774.780
Date:	Tue, 08 Aug 2023	AIC	1563.560
Time:	04:48:07	BIC	1582.071
Sample:	01-01-1980	HQIC	1571.059
	- 12-01-1990		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	-0.8698	0.119	-7.324	0.000	-1.103	-0.637
ma.L1	-0.0012	3.878	-0.000	1.000	-7.602	7.600
ma.L2	-0.9988	0.161	-6.187	0.000	-1.315	-0.682
ar.S.L12	1.0395	0.011	91.533	0.000	1.017	1.062
ma.S.L12	-0.5552	0.105	-5.280	0.000	-0.761	-0.349
ma.S.L24	-0.1782	0.113	-1.581	0.114	-0.399	0.043
sigma2	1.48e+05	2.54e-05	5.82e+09	0.000	1.48e+05	1.48e+05

Ljung-Box (L1) (Q):	0.87	Jarque-Bera (JB):	9.79
Prob(Q):	0.35	Prob(JB):	0.01
Heteroskedasticity (H):	1.73	Skew:	0.39
Prob(H) (two-sided):	0.11	Kurtosis:	4.29

Table 18: Summary of the SARIMA model.

Predictions on the test set:

The below table gives the predicted values of the SARIMA model on the first ten data points of the test set.

YearMonth	Actual Values	Prediction
1991-01-01	1902	1562.067235
1991-02-01	2049	1652.122228
1991-03-01	1874	1796.845713
1991-04-01	1279	1942.934706
1991-05-01	1432	1565.746841
1991-06-01	1540	1562.905850
1991-07-01	2214	1963.324086
1991-08-01	1857	2142.358416
1991-09-01	2408	2361.195512
1991-10-01	3252	3433.573753

Table 19: Predictions of the SARIMA model.

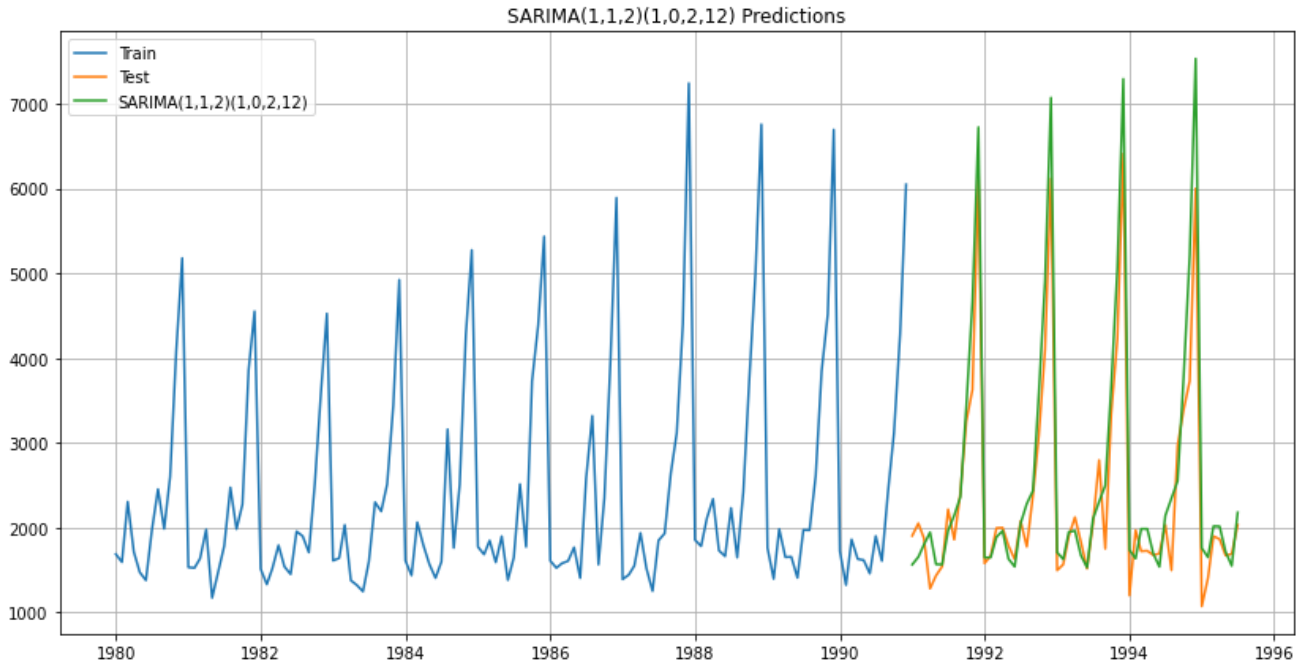

Fig. 23: Predictions of the SARIMA model.

Plot Diagnostics:

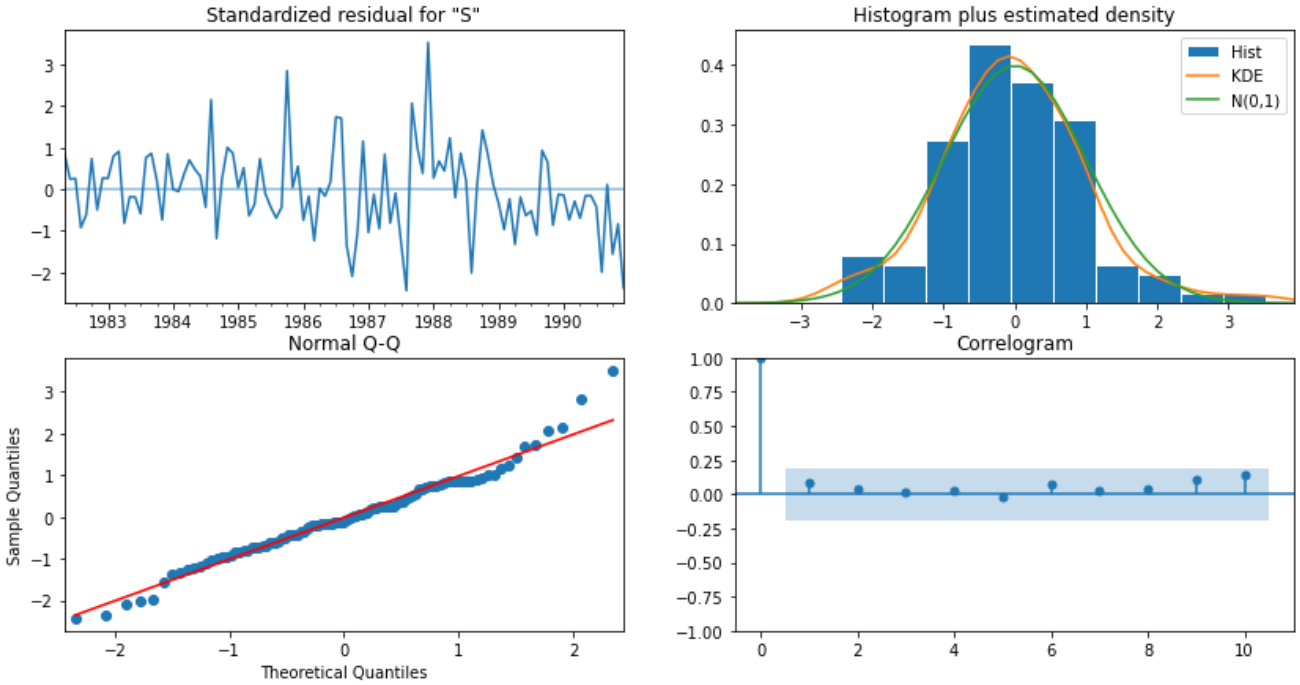

Fig. 24: Plot diagnostics of the SARIMA model.

Model Evaluation:

The RMSE on the test set comes out to be 499.047.

7. Build a table (create a data frame) with all the models built along with their corresponding parameters and the respective RMSE values on the test data.

	Test RMSE	Parameters
TES(A,A,A)	378.951023	Alpha=0.11, Beta=0.01, Gamma=0.46
TES(A,A,M)	404.286809	Alpha=0.11, Beta=0.04, Gamma=0.36
SARIMA	499.047848	(1,1,2)(1,0,2,12)
Moving Average-1	813.400684	2 point
Moving Average-2	1156.589694	4 point
Simple Average Model	1275.081804	
Moving Average-3	1283.927428	6 point
ARIMA	1299.979640	(2,1,2)
SES	1338.008384	Alpha=0.07
Moving Average-4	1346.278315	9 point
Linear Regression	1389.135000	
Naive Model	3864.279352	
DES	5291.879833	Alpha = 0.66, Beta=0.0001

Table 20: Different models and their RMSE values on test data

8. 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.

From the above table, we see that the triple exponential models have the lowest RMSE scores. Hence, we shall build these models on the whole data to forecast 12 months into future with 95% confidence interval.

Predictions of Triple Exponential Smoothing – with Additive Seasonality:

	lower_CI	prediction	upper_CI
1995-08-01	1127.039982	1850.631377	2574.222772
1995-09-01	1731.890798	2455.482192	3179.073587
1995-10-01	2522.093951	3245.685345	3969.276740
1995-11-01	3149.232196	3872.823591	4596.414986
1995-12-01	5378.462112	6102.053507	6825.644902
1996-01-01	491.178830	1214.770224	1938.361619
1996-02-01	876.474187	1600.065582	2323.656977
1996-03-01	1133.945467	1857.536862	2581.128257
1996-04-01	1116.396154	1839.987548	2563.578943
1996-05-01	954.726212	1678.317607	2401.909002
1996-06-01	905.321639	1628.913033	2352.504428
1996-07-01	1266.071449	1989.662844	2713.254239

Table 21: Future Prediction with 95% confidence interval from TES (A,A,A) Model.

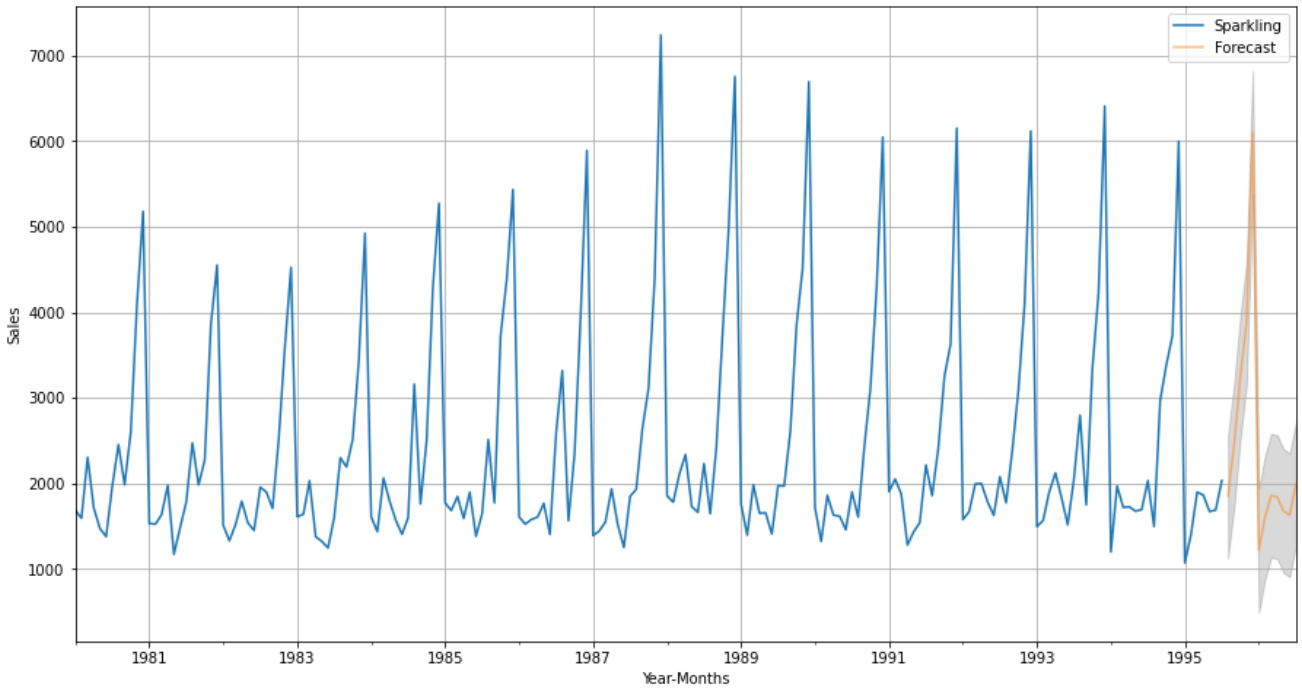

Fig. 25: Future Prediction with 95% confidence interval from TES (A,A,A) Model.

Predictions of Triple Exponential Smoothing – with Multiplicative Seasonality:

	lower_CI	prediction	upper_CI
1995-08-01	1185.271242	1875.803268	2566.335294
1995-09-01	1703.972697	2394.504723	3085.036749
1995-10-01	2478.542558	3169.074584	3859.606610
1995-11-01	3138.187343	3828.719369	4519.251395
1995-12-01	5251.087626	5941.619652	6632.151678
1996-01-01	590.001582	1280.533608	1971.065634
1996-02-01	907.504335	1598.036361	2288.568387
1996-03-01	1146.327163	1836.859189	2527.391215
1996-04-01	1119.472374	1810.004400	2500.536426
1996-05-01	961.852034	1652.384060	2342.916086
1996-06-01	894.965059	1585.497085	2276.029111
1996-07-01	1265.021550	1955.553576	2646.085602

Table 22: Future Prediction with 95% confidence interval from TES (A,A,M) Model.

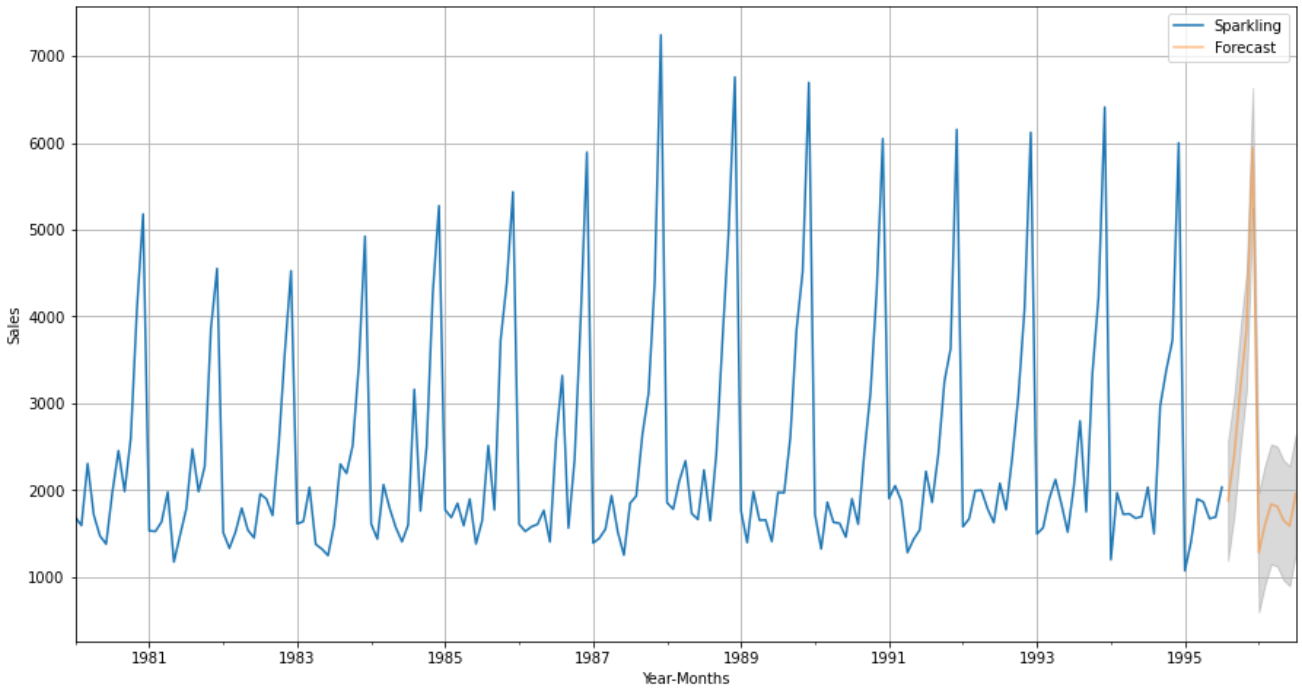

Fig. 26: Future Prediction with 95% confidence interval from TES (A,A,M) Model.

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

Inference:

- Triple Exponential Smoothing model is used to forecast into the future for next 12 months.
- The forecast tells that the sale will be more in the month of October, November and December 1995.
- The company should make sure the stocks are not emptied in these months.
- The sale will be lowest in the first and second quarter of the next year. The company may consider giving some offers to lure the customers in.
- The sale starts to pick up in the month of July.