Time Series Forecasting

Coded Project

SRINIVASAN T

Table of Content:

List of Tables:

List of Figures:

Scoring guide (Rubric) - Some Rubric

Criteria	Points
Define the problem and perform Exploratory Data Analysis - Read the data as an appropriate time series data - Plot the data - Perform EDA - Perform Decomposition	9
Data Pre-processing - Missing value treatment - Visualize the processed data - Train-test split	4
Model Building - Original Data - Build forecasting models - Linear regression - Simple Average - Moving Average - Exponential Models (Single, Double, Triple) - Check the performance of the models built	15
Check for Stationarity - Check for stationarity - Make the data stationary (if needed)	4
Model Building - Stationary Data - Generate ACF & PACF Plot and find the AR, MA values. - Build different ARIMA models - Auto ARIMA - Manual ARIMA - Build different SARIMA models - Auto SARIMA - Manual SARIMA - Check the performance of the models built	12
Compare the performance of the models - Compare the performance of all the models built - Choose the best model with proper rationale - Rebuild the best model using the entire data - Make a forecast for the next 12 months	6
Actionable Insights & Recommendations - Conclude with the key takeaways (actionable insights and recommendations) for the business	4
Business Report Quality - Adhere to the business report checklist	6
Points	60

1. Define the problem and perform Exploratory Data Analysis

Problem Statement - TSF Project

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.

1.1 Read the data as an appropriate time series data

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

Table 1.1: Reading the dataset (top 5 rows)

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

Table 1.2: Reading the dataset (bottom 5 rows)

Checking the Datatypes:

```
YearMonth    object
Sparkling     int64
dtype: object
```

Need to change the datatype to datatime64[ns] format using the parse_date function

```
YearMonth    datetime64[ns]
Sparkling             int64
dtype: object
```

YearMonth	Sparkling
1980-01-01	1686
1980-02-01	1591
1980-03-01	2304
1980-04-01	1712
1980-05-01	1471

Table 1.3: Final dataset with appropriate time series format

1.2 Plot the data

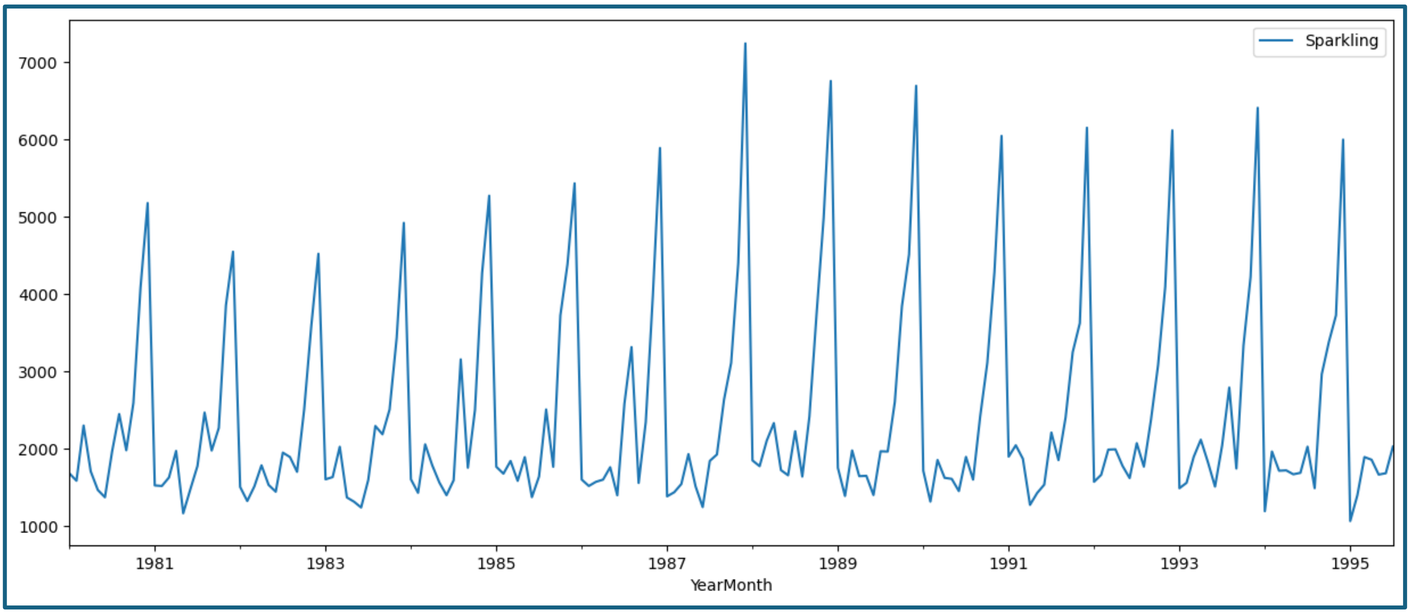


Figure 1.1: Sparkling Dataset plot

The data seems to have seasonality but not trend. 1987 - 19990 seems to have the highest sales across the given date range.

1.3 Perform EDA

1.3.1 Yearly Boxplot:

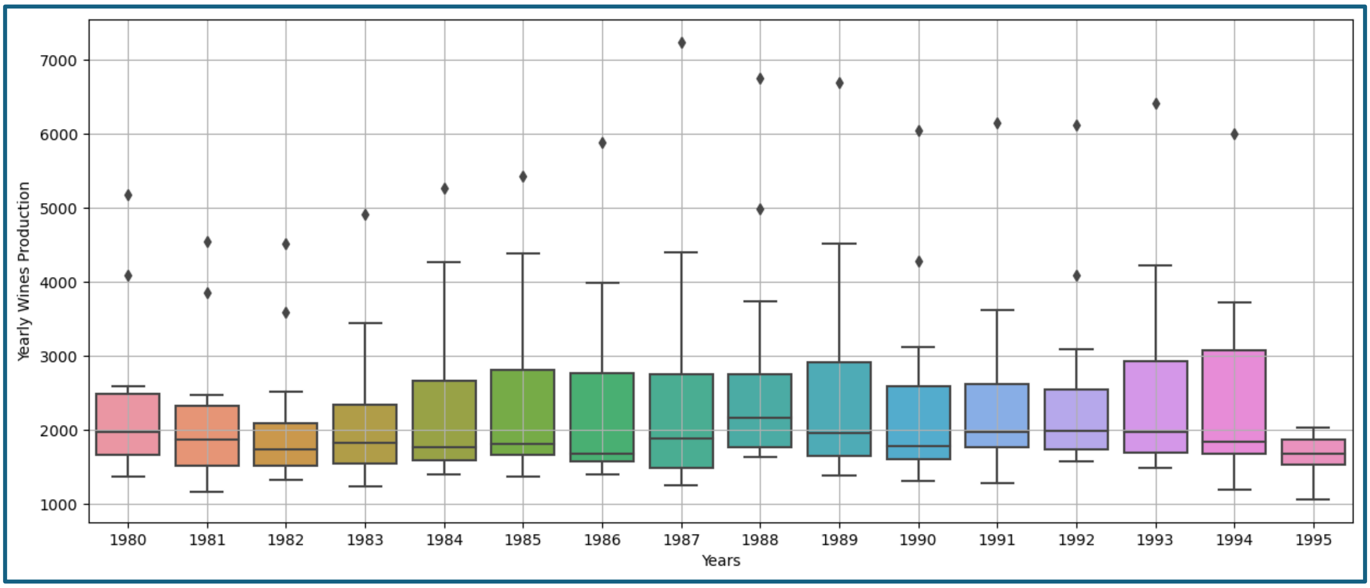


Figure 1.2: Yearly Boxplot

The year 1987 seems to have the highest sales across all years. But the year 1988 has the highest average sales. The seems to be no trend in the dataset.

1.3.2 Monthly Boxplot:

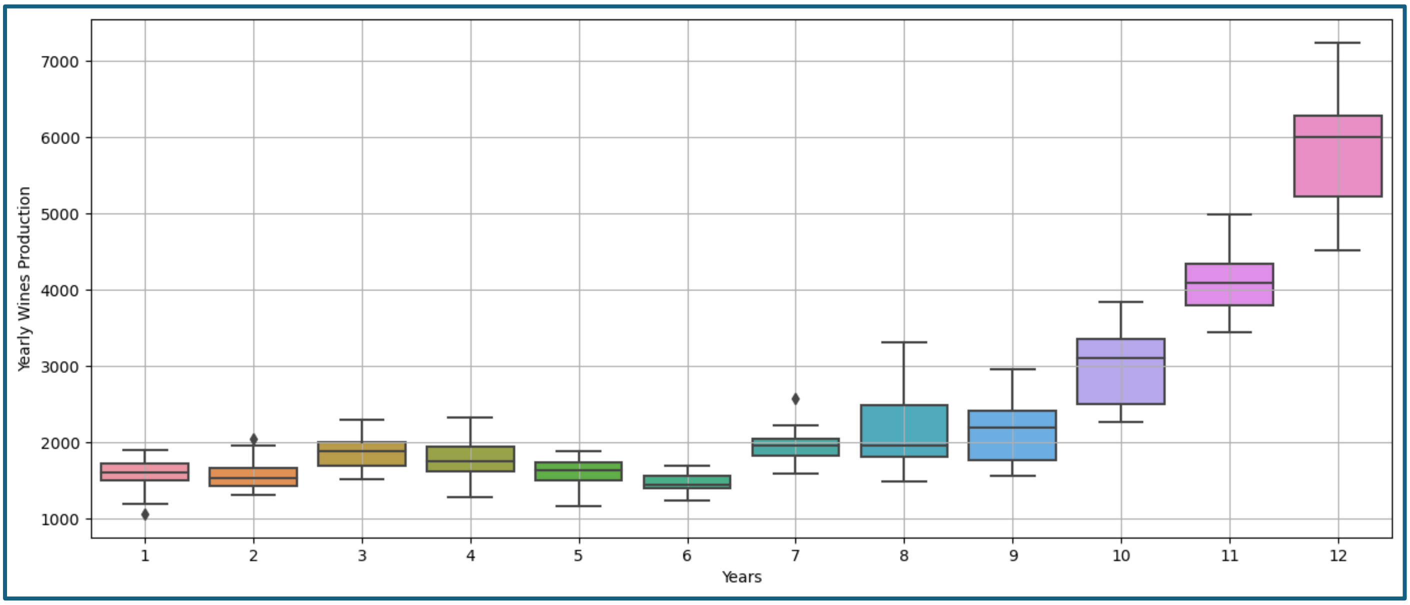


Figure 1.3: Monthly Boxplot

The monthly boxplot clearly shows that there is a trend between months decreasing trend from January till June and then a clear increasing trend from July till December. December has the highest sales across all the years.

1.3.3 Quarterly Line plot:

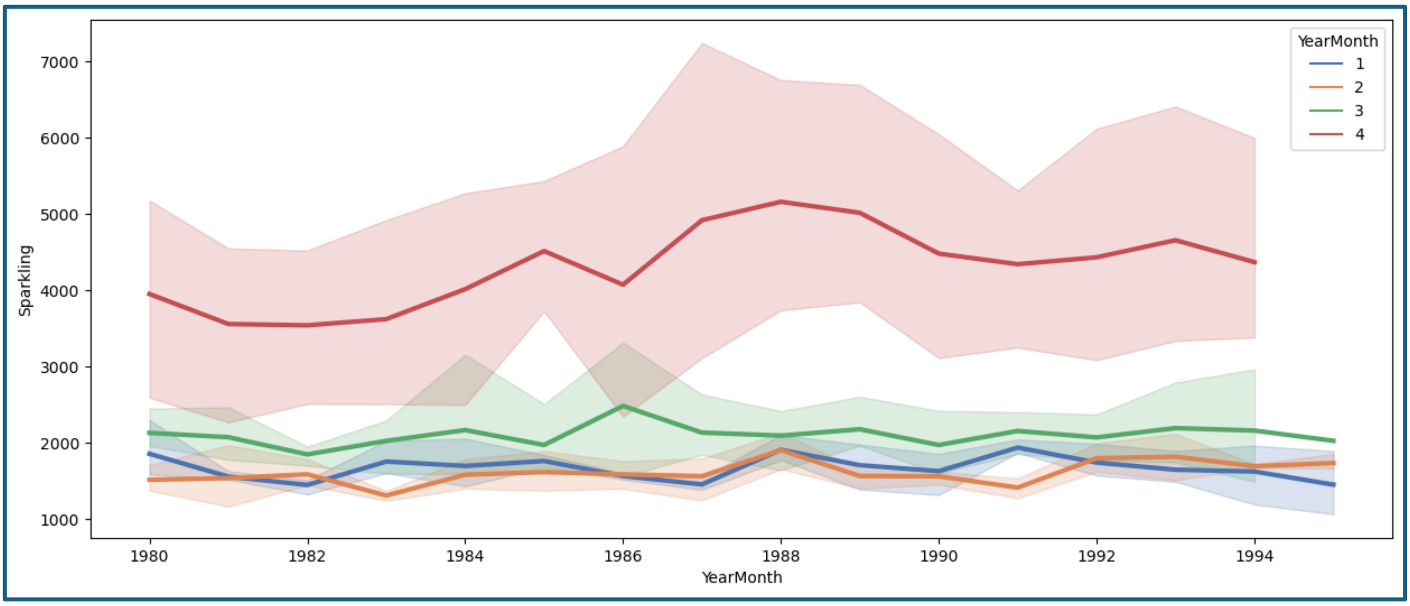


Figure 1.4: Monthly Line plot

The Q4 having the highest sales across all the months.

1.3.4 Descriptive plot:

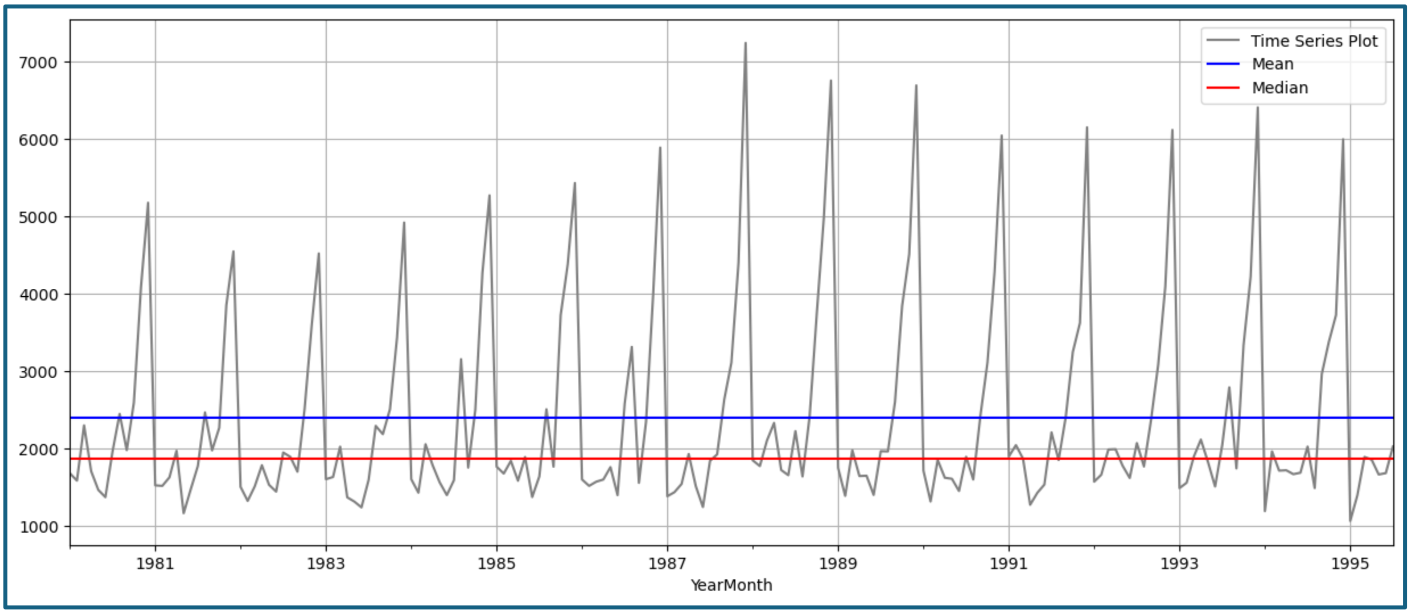


Figure 1.5: Descriptive plot

The mean is greater than median, obviously shows a right skewed plot.

1.4 Perform Decomposition

1.4.1 Additive Model

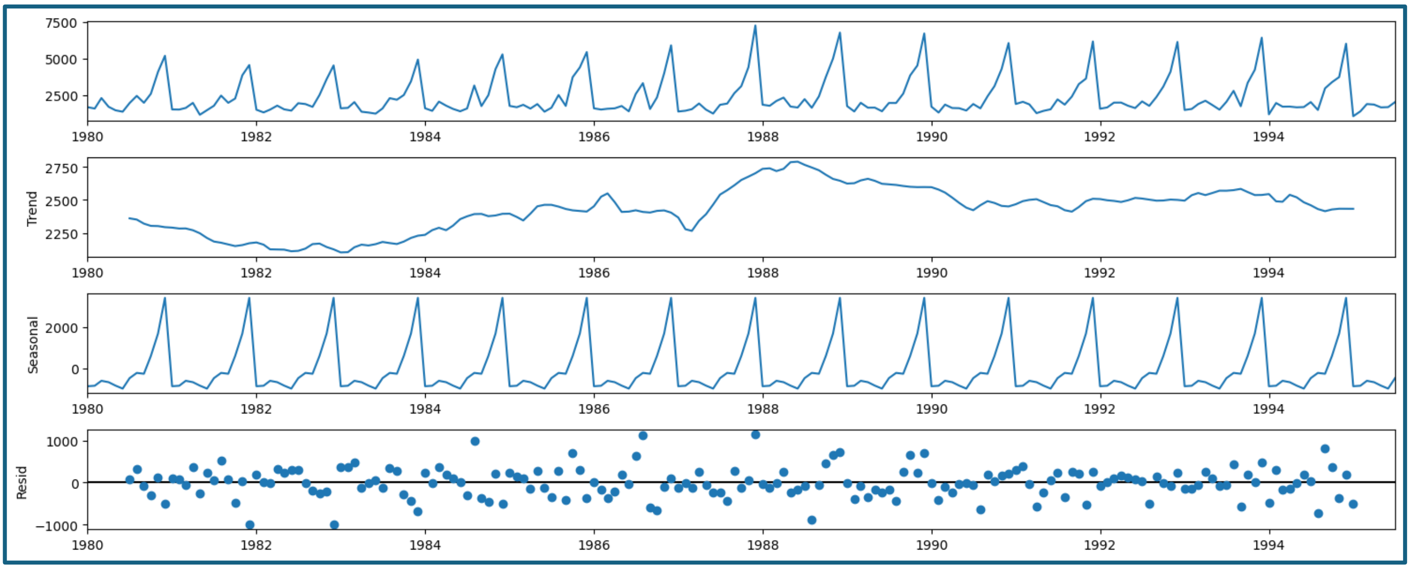


Figure 1.6: Additive Decomposition Model

The additive model shows that the original data does not have trend and have seasonality and have residual around 0 with some pattern in it.

1.4.2 Multiplicative Model

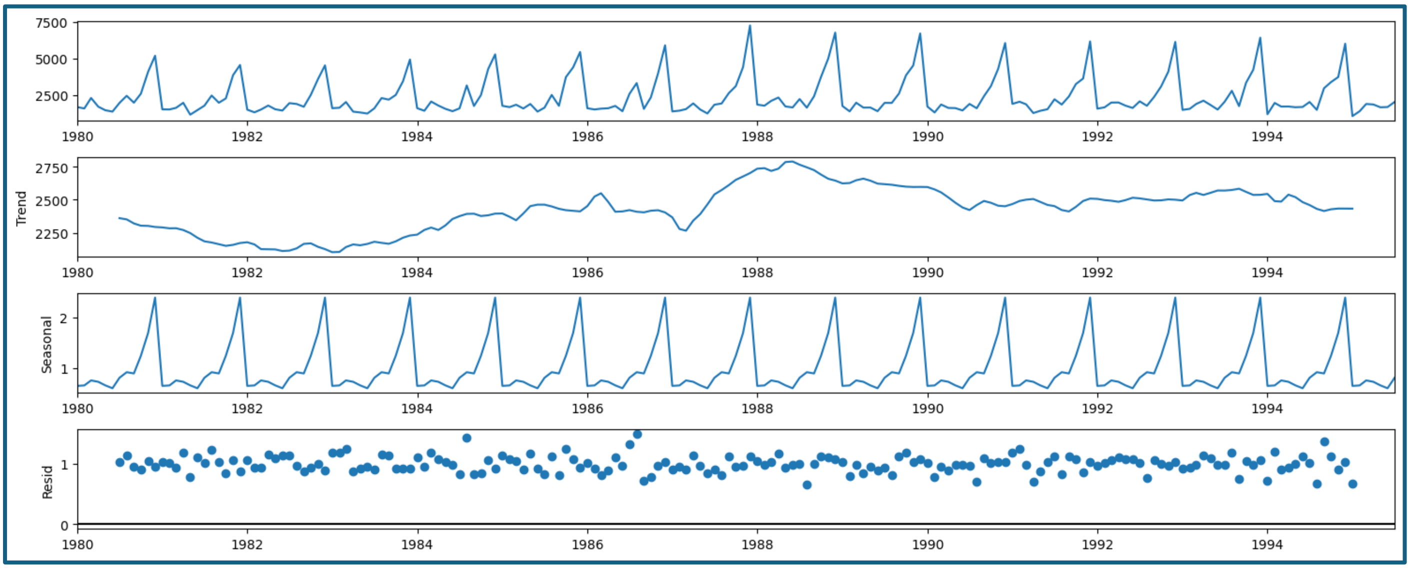


Figure 1.7: Multiplicative Decomposition Model

The multiplicative model shows that the original data does not have trend and have seasonality and have residual around 1 without any pattern in it.

2. Data Pre-processing

2.1 Missing value treatment

The missing values can be identified using the info function on the dataset.

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

There are no missing values in the dataset.

2.2 Visualize the processed data

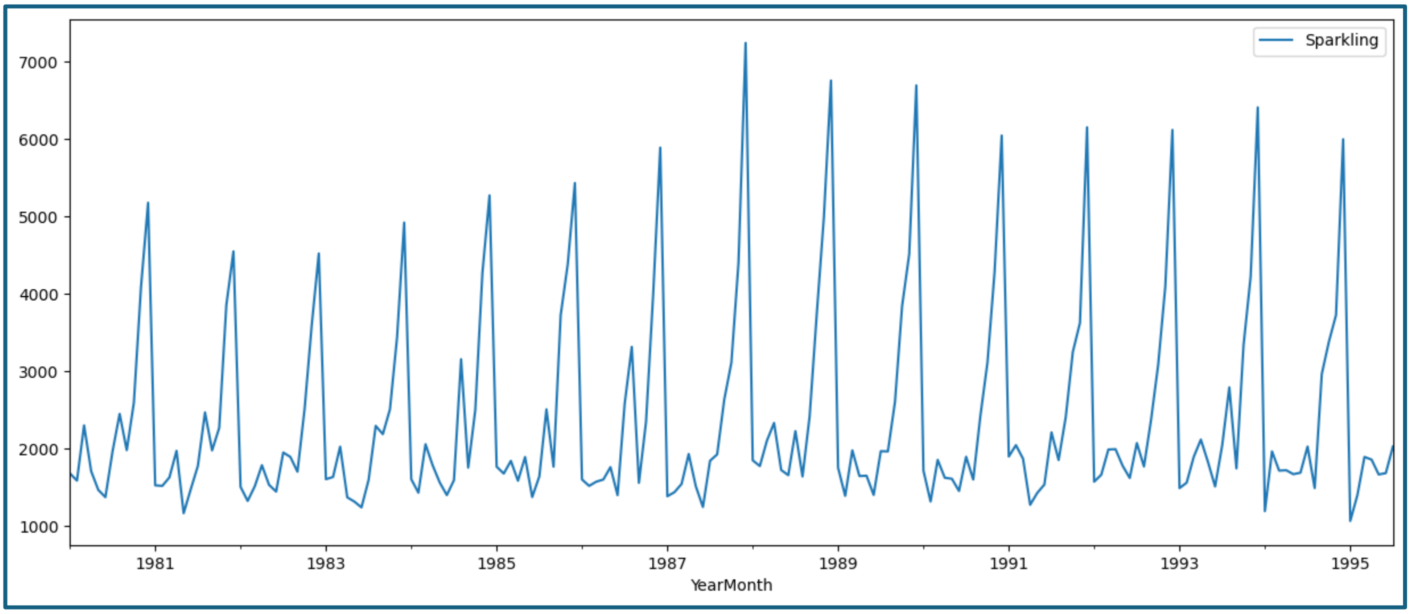


Figure 1.8: Processed Data

2.3 Train-test split

The train-test split happens in sequential manner, such that the first 70% of the data is considered as train in the same times series order and the last 30% again in the same sequential times series order is considered and test data. The train-test split cannot random as it is a time series format.

```
Train Shape - (130, 1)  [train has 130 rows]
Test Shape - (57, 1)    [test has 57 rows]
```

2.3.1 Train-test split

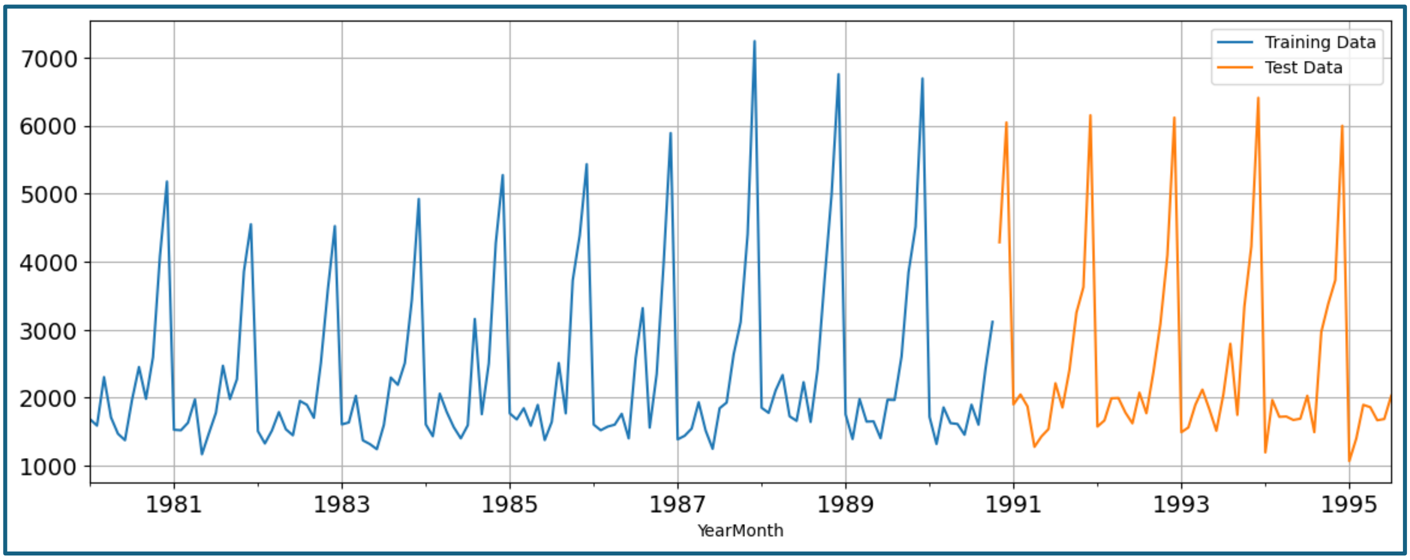


Figure 1.9: Train-test Data Plot

The plot clearly shows that the time series train-train split in a sequential order.

3. Model Building - Original Data

3.1 Build forecasting models

3.1.1 Linear regression

Before Linear Regression we will fitting the numerical time instance of the sales (sparkling) and adding the same to the DataFrame.

Training Time instance

```
 [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23
, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 4
4, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64,
65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85,
86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 10
5, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121,
122, 123, 124, 125, 126, 127, 128, 129, 130]
```

Test Time instance

```
 [131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 1
47, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163,
164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180
, 181, 182, 183, 184, 185, 186, 187]
```

```
First few rows of Training Data
              Sparkling  time
YearMonth
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
              Sparkling  time
YearMonth
1990-06-01         1457   126
1990-07-01         1899   127
1990-08-01         1605   128
1990-09-01         2424   129
1990-10-01         3116   130

First few rows of Test Data
              Sparkling  time
YearMonth
1990-11-01         4286   131
1990-12-01         6047   132
1991-01-01         1902   133
1991-02-01         2049   134
1991-03-01         1874   135

Last few rows of Test Data
              Sparkling  time
YearMonth
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 1.4: Time instance for Regression

Running the Linear Regression Model:

```
▾ LinearRegression
LinearRegression()
```

Figure 1.10: Linear Regression

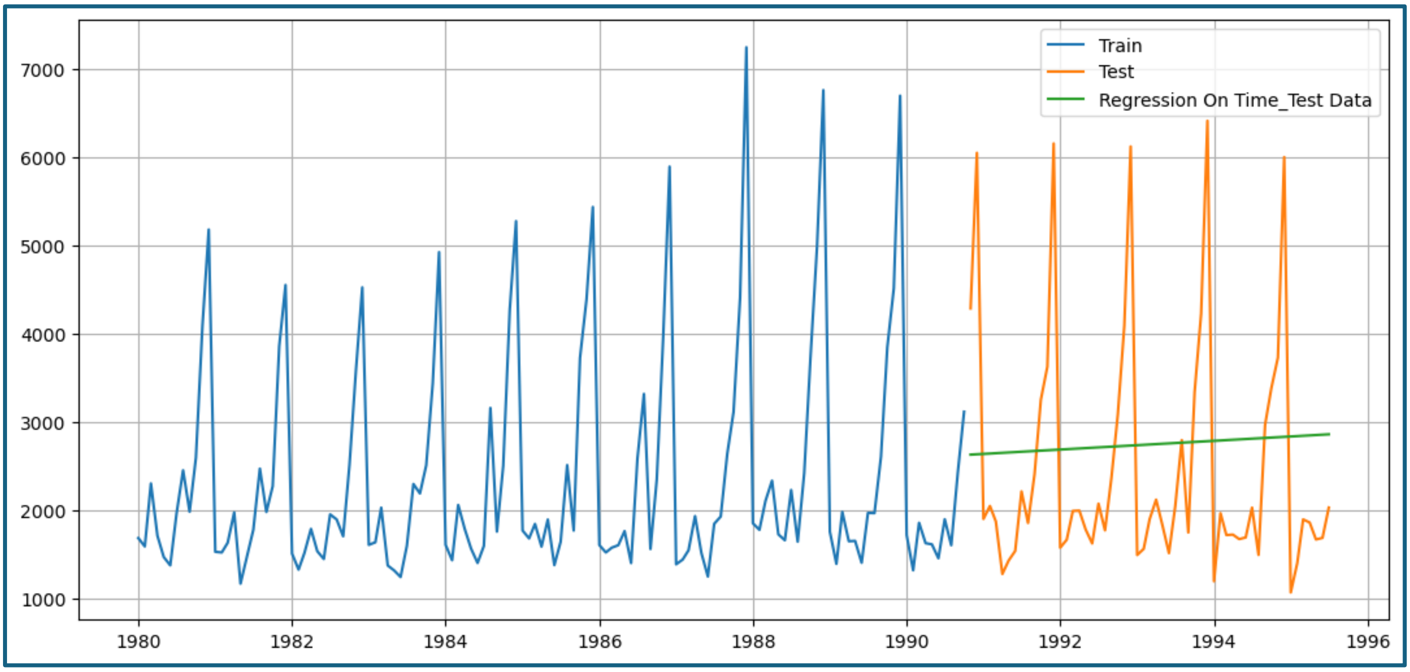


Figure 1.11: Linear Regression Plot

The green line on the test data represents the regression on time.

Model Evaluation:

The model is evaluated using the Root Mean Squared Error value on the regression model and added to a new DataFrame for comparison purpose of different models.

	Test RMSE
RegressionOnTime	1392.438305

Table 1.5: Linear Regression

3.3 Simple Average

YearMonth	Sparkling	forecast
1990-11-01	4286	2361.276923
1990-12-01	6047	2361.276923
1991-01-01	1902	2361.276923
1991-02-01	2049	2361.276923
1991-03-01	1874	2361.276923

Table 1.5: Simple Average

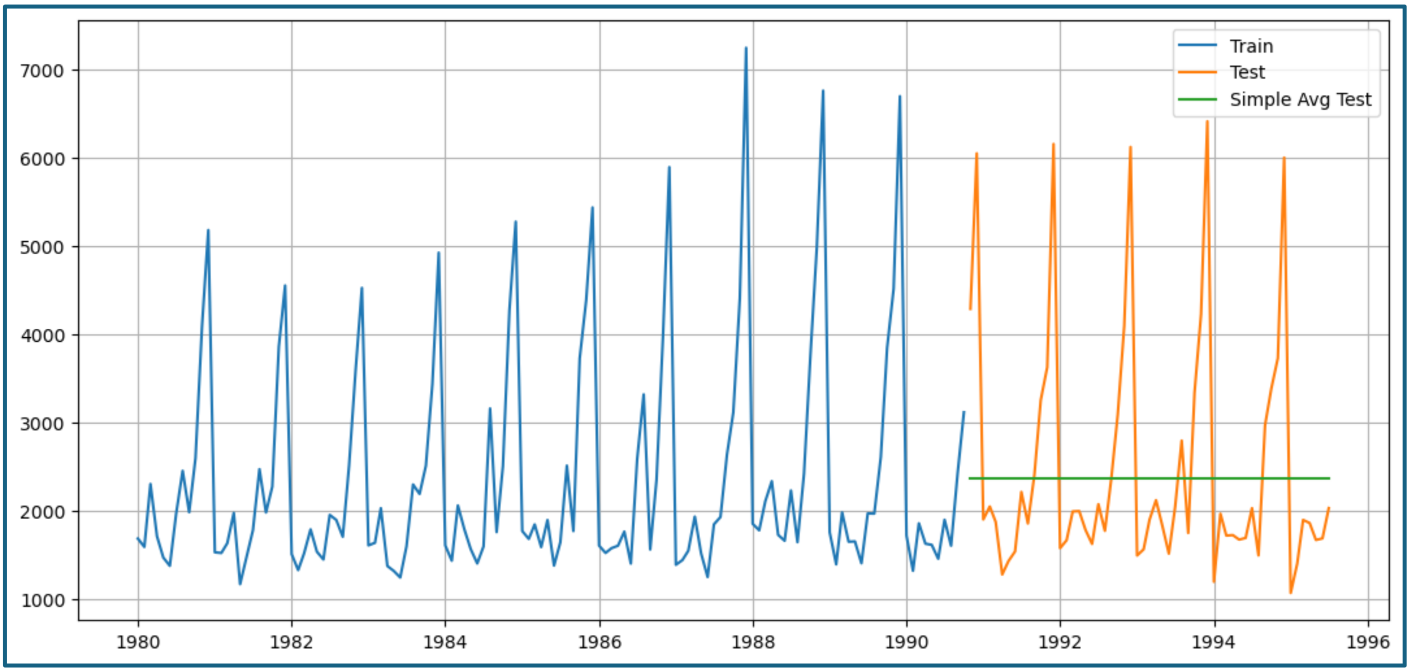


Figure 1.11: Simple Average Plot

3.4 Moving Average

For the moving average model rolling means (or moving averages) are calculated for different intervals. The best average can be determined by min. error value.

The trailing values for the moving average can be considered as 2, 4, 6 & 9

YearMonth	Sparkling	Trailing_2	Trailing_4	Trailing_6	Trailing_9
1980-01-01	1686	NaN	NaN	NaN	NaN
1980-02-01	1591	1638.5	NaN	NaN	NaN
1980-03-01	2304	1947.5	NaN	NaN	NaN
1980-04-01	1712	2008.0	1823.25	NaN	NaN
1980-05-01	1471	1591.5	1769.50	NaN	NaN
...	...	...	...	...	...
1995-03-01	1897	1649.5	2592.00	2913.666667	2664.000000
1995-04-01	1862	1879.5	1557.75	2659.833333	2645.222222
1995-05-01	1670	1766.0	1707.75	2316.666667	2664.666667
1995-06-01	1688	1679.0	1779.25	1598.166667	2522.444444
1995-07-01	2031	1859.5	1812.75	1758.333333	2372.000000

Table 1.6: Moving Average

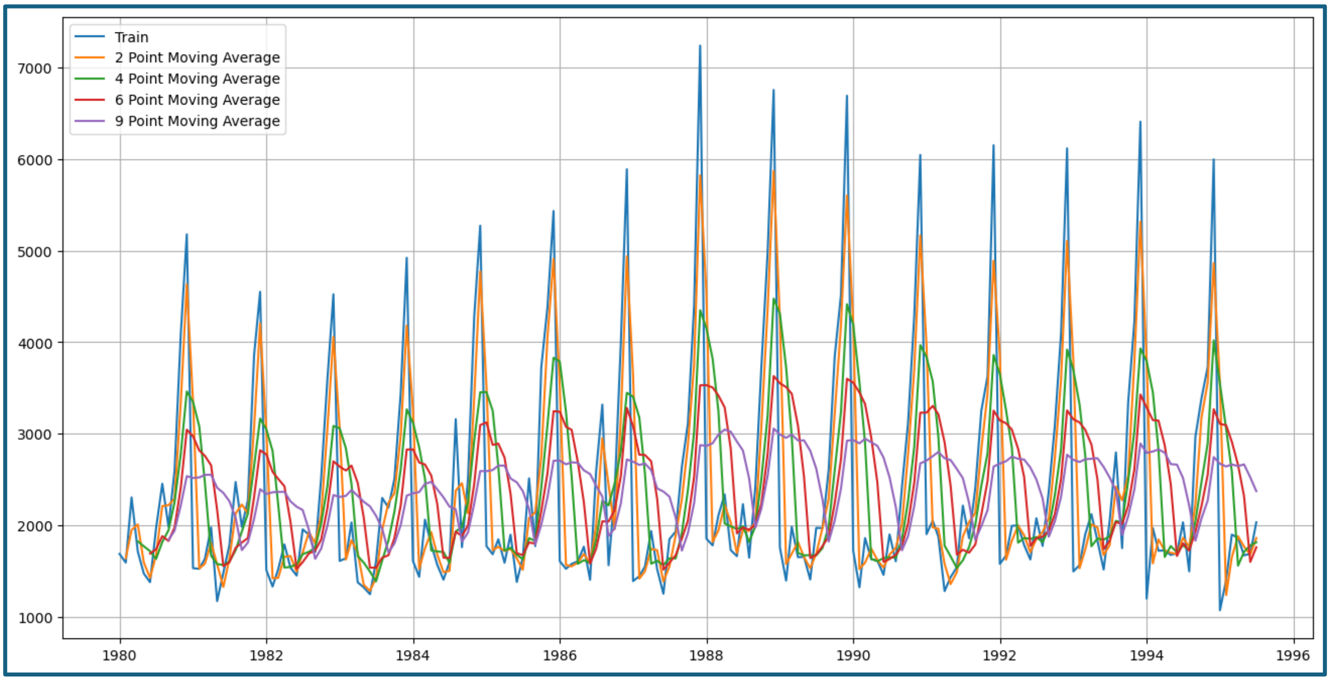


Figure 1.12: Moving Average Plot

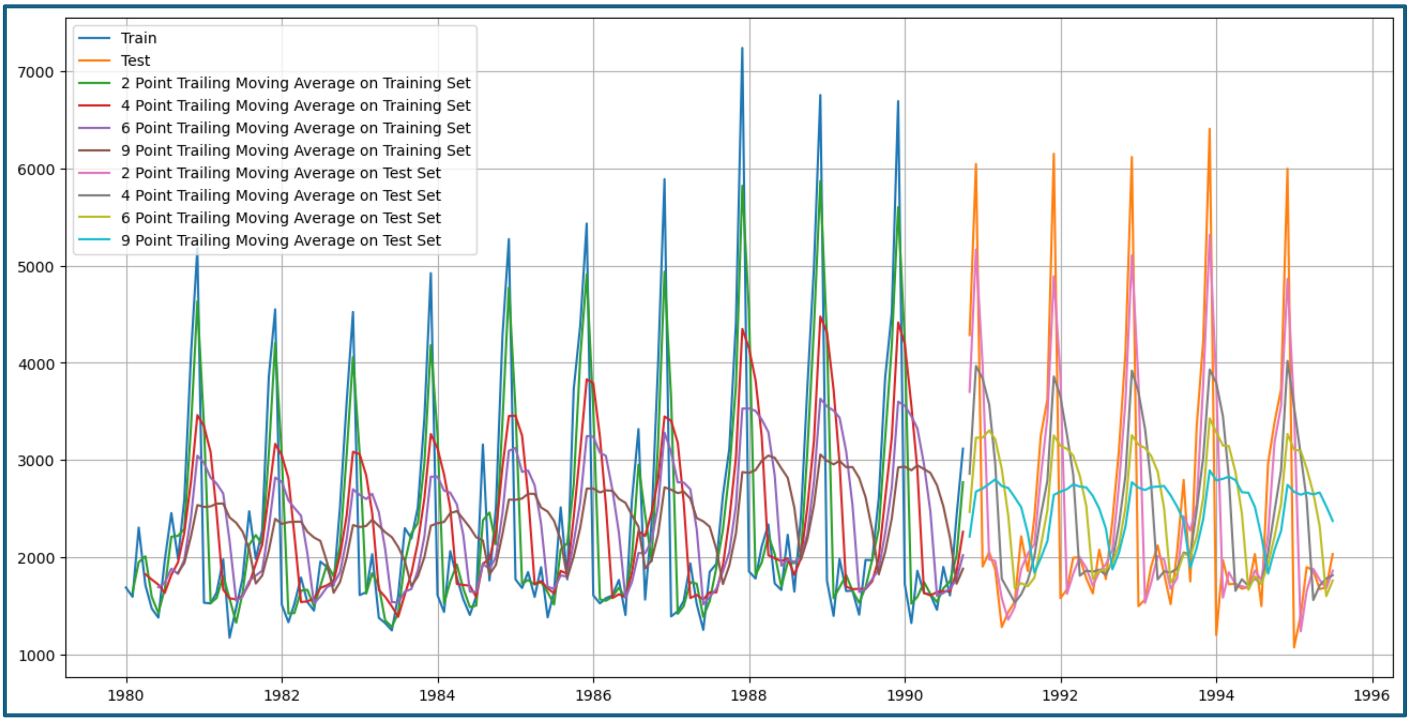


Figure 1.13: Moving Average Train-Test Split Plot

Model Evaluation:

```
For 2 point Moving Average Model forecast on the Training Data,  RMSE is 811.179
For 4 point Moving Average Model forecast on the Training Data,  RMSE is 1184.213
For 6 point Moving Average Model forecast on the Training Data,  RMSE is 1337.201
For 9 point Moving Average Model forecast on the Training Data,  RMSE is 1422.653
```

The 2 point moving average model has the lowest RMSE score compared to others.

	Test RMSE
RegressionOnTime	1392.438305
2pointTrailingMovingAverage	811.178937
4pointTrailingMovingAverage	1184.213295
6pointTrailingMovingAverage	1337.200524
9pointTrailingMovingAverage	1422.653281

Table 1.7: RMSE of different models

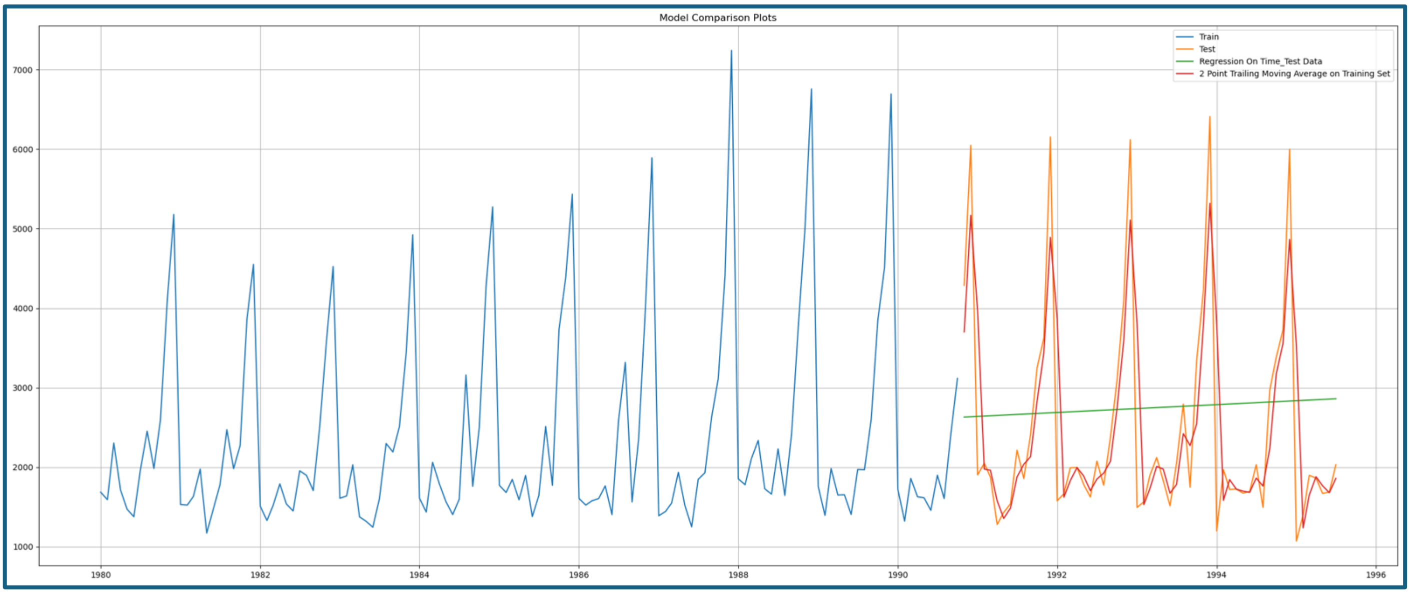


Figure 1.14: Train-Test Split – Moving Average - Regression Plots

3.5 Exponential Models (Single, Double, Triple)

3.5.1 Single Exponential Smoothing

Running the single exponential smoothing using the SimpleExpSmoothing function with train data and fitting it.

Model Parameters:

```
{'smoothing_level': 0.038003579704776386,
 'smoothing_trend': nan,
 'smoothing_seasonal': nan,
 'damping_trend': nan,
 'initial_level': 2173.2611215945826,
 'initial_trend': nan,
 'initial_seasons': array([], dtype=float64),
 'use_boxcox': False,
 'lamda': None,
 'remove_bias': False}
```

Prediction:

	Sparkling	predict
YearMonth		
1990-11-01	4286	2468.64942
1990-12-01	6047	2468.64942
1991-01-01	1902	2468.64942
1991-02-01	2049	2468.64942
1991-03-01	1874	2468.64942

Table 1.8: Single Exponential Smoothing Predictions

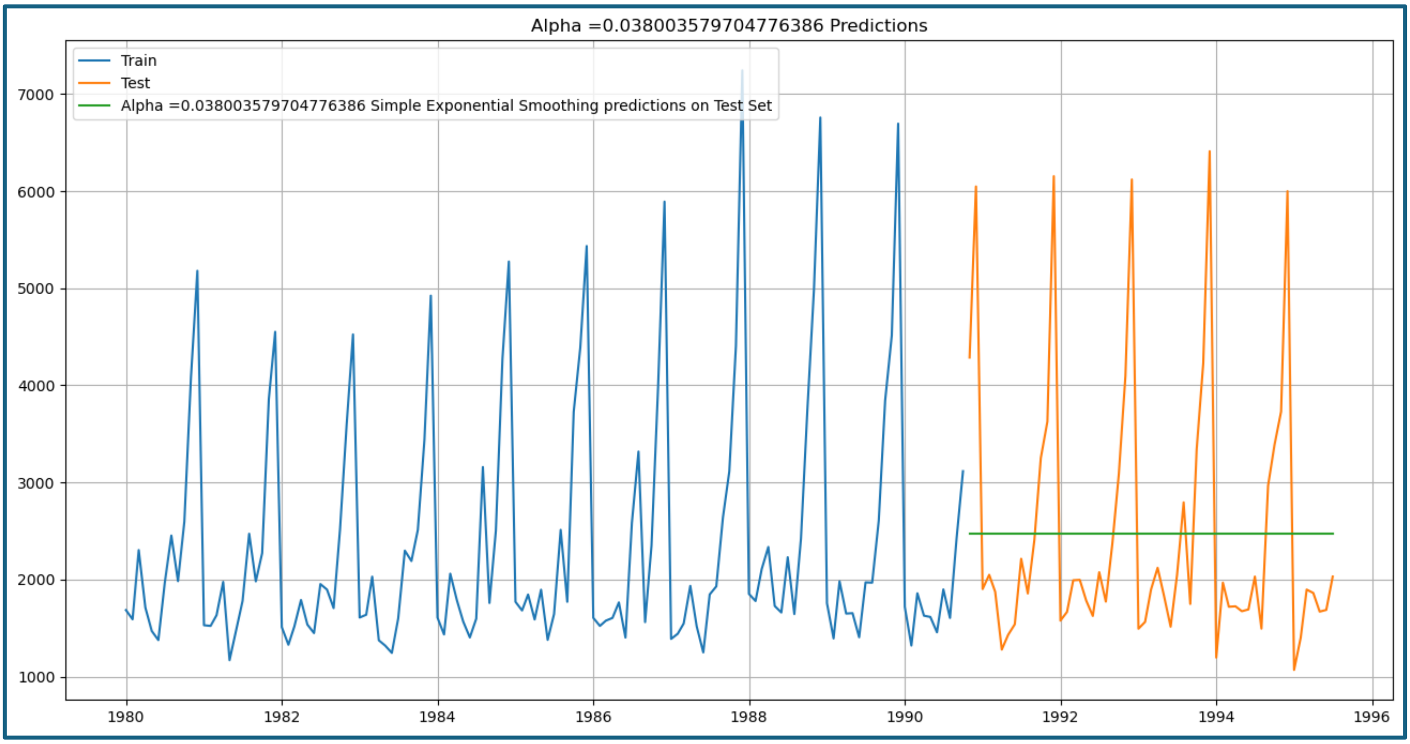


Figure 1.15: Single Exponential Smoothing with Alpha – 0.038 value Plot

Model Evaluation for α = 0.038003579704776386 : Simple Exponential Smoothing

	Test RMSE
RegressionOnTime	1392.438305
2pointTrailingMovingAverage	811.178937
4pointTrailingMovingAverage	1184.213295
6pointTrailingMovingAverage	1337.200524
9pointTrailingMovingAverage	1422.653281
Alpha=0.038003579704776386,SimpleExponentialSmoothing	1362.355525

Table 1.9: RMSE after Single Exponential Smoothing Model

Still the RMSE of Single Exponential Smoothing Model is higher than the 2 Point Moving Average Value

Different RMSE values for different alpha values range from 0.1 to 1

	Alpha Values	Train RMSE	Test RMSE
3	0.4	1329.814823	1363.037803
4	0.5	1326.403864	1364.863549
0	0.1	1298.211536	1367.395642
2	0.3	1331.102204	1372.323705
1	0.2	1322.658289	1378.320562
5	0.6	1325.588422	1379.988733
6	0.7	1329.257530	1404.659104
7	0.8	1337.879425	1434.578214
8	0.9	1351.645478	1466.179706

Table 1.10: Different RMSE for Alpha 0.1 - 1

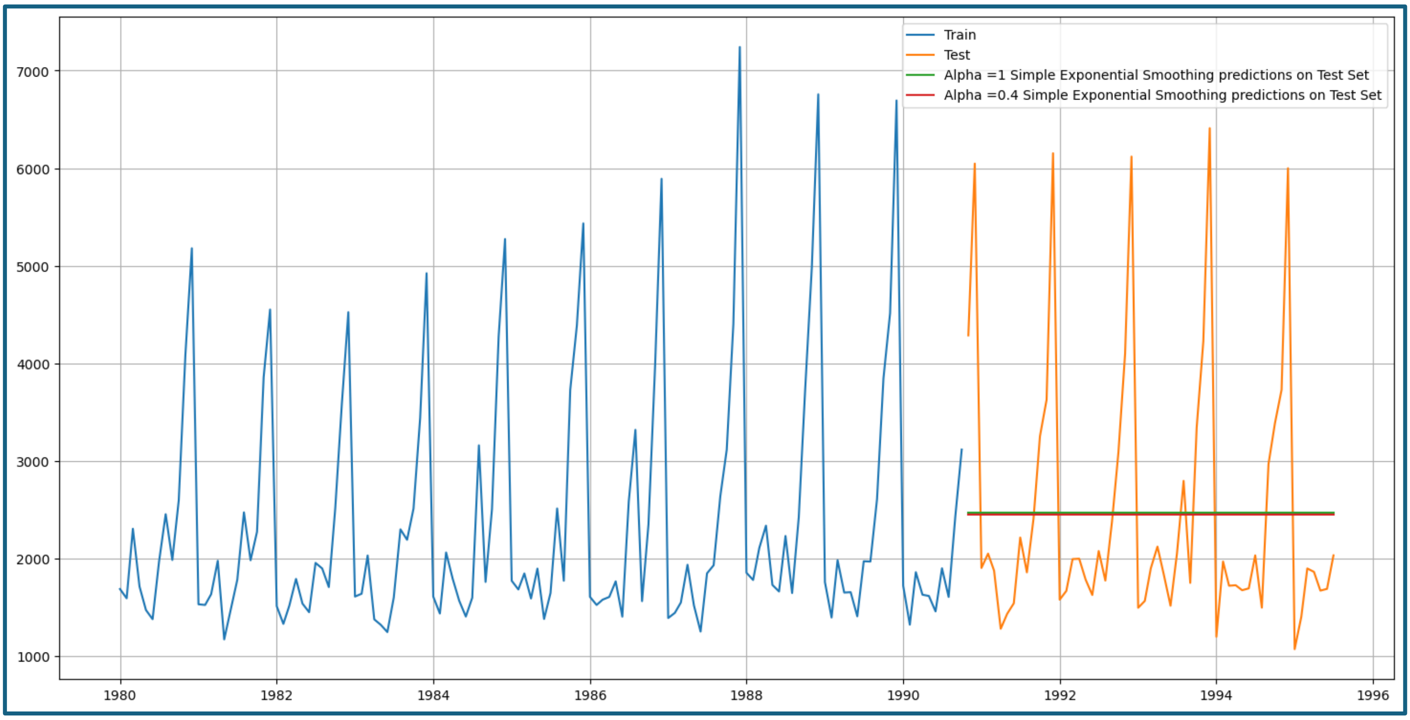


Figure 1.16: Alpha – 0.4 value Plot on Single Exponential Smoothing

	Test RMSE
RegressionOnTime	1392.438305
2pointTrailingMovingAverage	811.178937
4pointTrailingMovingAverage	1184.213295
6pointTrailingMovingAverage	1337.200524
9pointTrailingMovingAverage	1422.653281
Alpha=0.038003579704776386,SimpleExponentialSmoothing	1362.355525
Alpha=0.4,SimpleExponentialSmoothing	1363.037803

Table 1.11: Different RMSE for different models

3.5.2 Double Exponential Smoothing

Running the single exponential smoothing using the Holt function with train data and fitting it.

	Alpha Values	Beta Values	Train RMSE	Test RMSE
0	0.3	0.3	1567.524066	1597.853999
1	0.3	0.4	1662.549225	4023.672164
2	0.3	0.5	1758.543876	8879.172380
3	0.3	0.6	1843.560670	15645.080035
4	0.3	0.7	1902.735965	23205.442323
...	...	...	...	...
59	1.0	0.6	1764.658812	20558.025827
60	1.0	0.7	1837.425218	22155.074151
61	1.0	0.8	1915.148280	23241.839479
62	1.0	0.9	1999.362743	23787.747852
63	1.0	1.0	2092.531564	23712.944127

Table 1.12: Different RMSE for different models

	Alpha Values	Beta Values	Train RMSE	Test RMSE
0	0.3	0.3	1567.524066	1597.853999
1	0.3	0.4	1662.549225	4023.672164
8	0.4	0.3	1556.795694	5049.478887
16	0.5	0.3	1525.615506	7817.569799
2	0.3	0.5	1758.543876	8879.172380

Table 1.13: Different RMSE for different alpha & beta sorted

The Model with Alpha – 0.3 & Beta – 0.3 has the lowest RMSE value in both train and test data

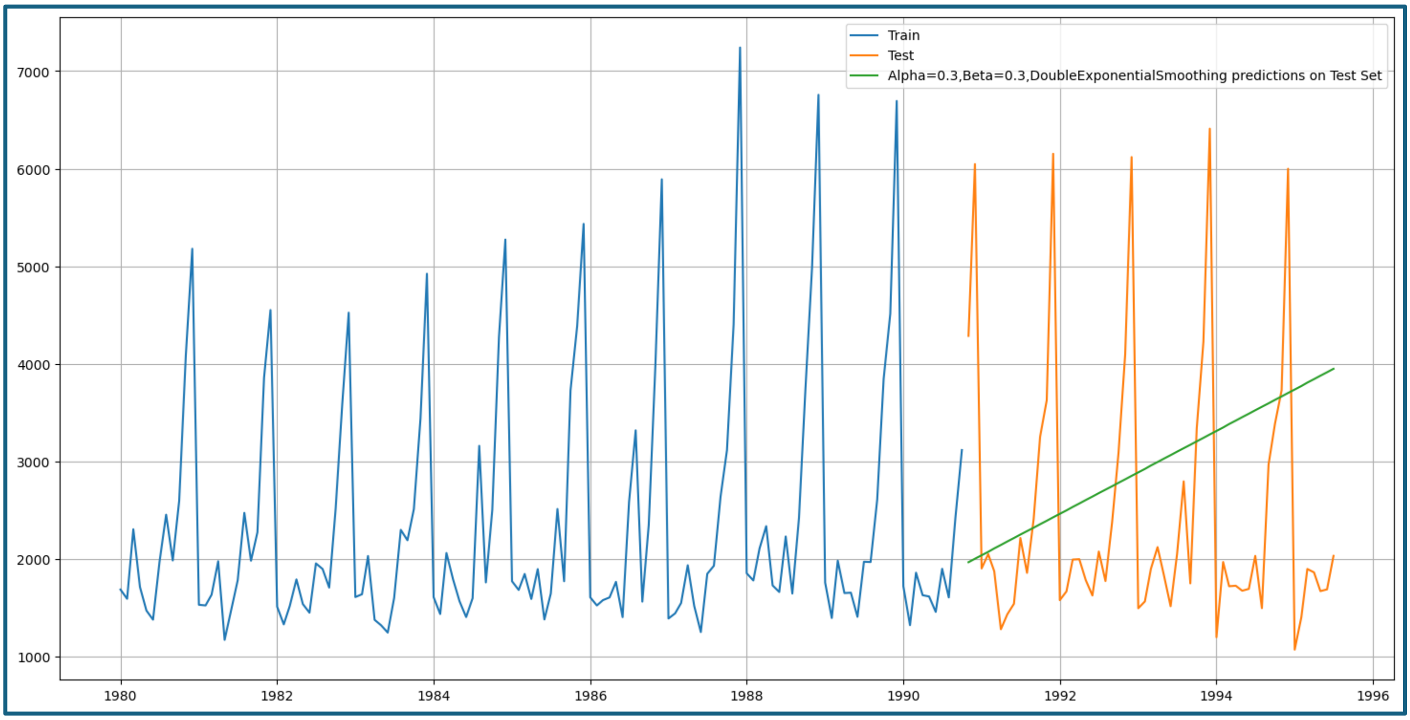


Figure 1.17: Alpha – 0.4 & Beta – 0.3 value Plot on Double Exponential Smoothing

	Test RMSE
RegressionOnTime	1392.438305
2pointTrailingMovingAverage	811.178937
4pointTrailingMovingAverage	1184.213295
6pointTrailingMovingAverage	1337.200524
9pointTrailingMovingAverage	1422.653281
Alpha=0.038003579704776386,SimpleExponentialSmoothing	1362.355525
Alpha=0.4,SimpleExponentialSmoothing	1363.037803
Alpha=0.3,Beta=0.3,DoubleExponentialSmoothing	1597.853999

Table 1.14: Different RMSE for different models

Till the double exponential smoothing the 2Point MA has the lowest RMSE value.

3.5.3 Triple Exponential Smoothing

Three parameters α , β and γ are estimated in this model. Level, Trend and Seasonality are accounted for in this model.

YearMonth	Sparkling	auto_predict
1990-11-01	4286	4327.609727
1990-12-01	6047	6208.854280
1991-01-01	1902	1621.601290
1991-02-01	2049	1379.862158
1991-03-01	1874	1791.912018

Table 1.15: Prediction Triple Exponential Smoothing model

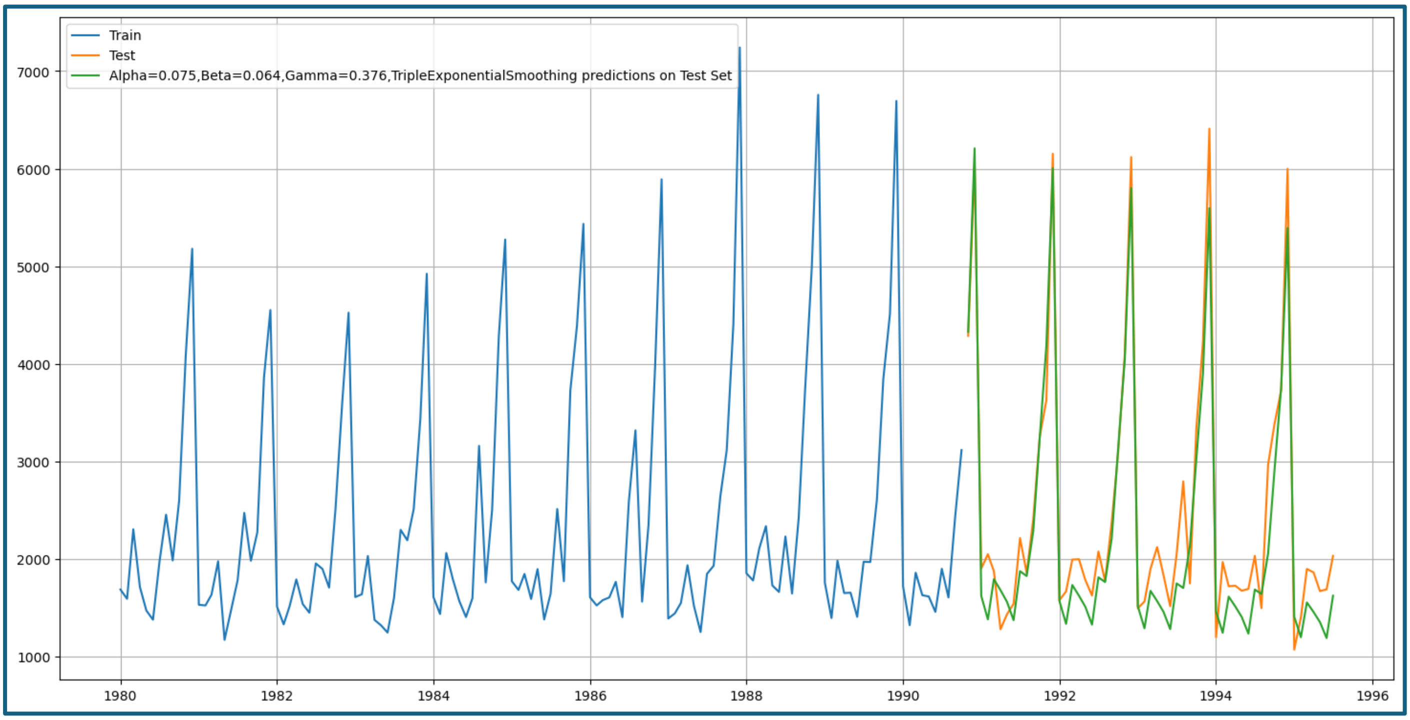


Figure 1.18: Alpha – 0.075, Beta – 0.064 & Gamma – 0.376 value Plot on Triple Exponential Smoothing

The predicted values are almost similar to the test data in the plot for the Triple Exponential Smoothing Model with Alpha – 0.075, Beta – 0.064 & Gamma – 0.376

	Test RMSE
RegressionOnTime	1392.438305
2pointTrailingMovingAverage	811.178937
4pointTrailingMovingAverage	1184.213295
6pointTrailingMovingAverage	1337.200524
9pointTrailingMovingAverage	1422.653281
Alpha=0.038003579704776386,SimpleExponentialSmoothing	1362.355525
Alpha=0.4,SimpleExponentialSmoothing	1363.037803
Alpha=0.3,Beta=0.3,DoubleExponentialSmoothing	1597.853999
Alpha=0.075,Beta=0.064,Gamma=0.376,TripleExponentialSmoothing	381.657232

Table 1.16: Different RMSE for different models

After all the Models the TES model has the lowest RMSE value.

	Alpha Values	Beta Values	Gamma Values	Train RMSE	Test RMSE
0	0.1	0.1	0.1	3.809925e+02	7.185399e+02
1	0.1	0.1	0.2	3.693765e+02	6.996973e+02
2	0.1	0.1	0.3	3.638351e+02	6.799473e+02
3	0.1	0.1	0.4	3.633589e+02	6.571580e+02
4	0.1	0.1	0.5	3.671489e+02	6.335335e+02
...	...	...	...	...	...
995	1.0	1.0	0.6	1.897409e+05	5.807913e+03
996	1.0	1.0	0.7	2.680381e+05	3.585079e+05
997	1.0	1.0	0.8	1.110365e+06	2.855947e+06
998	1.0	1.0	0.9	7.751842e+04	7.857850e+04
999	1.0	1.0	1.0	2.466982e+04	1.354272e+05

Table 1.17: Different RMSE for different α , β and γ values

	Alpha Values	Beta Values	Gamma Values	Train RMSE	Test RMSE
801	0.9	0.1	0.2	463.624419	342.322388
500	0.6	0.1	0.1	410.031940	345.280719
611	0.7	0.2	0.2	443.170290	356.766667
901	1.0	0.1	0.2	487.030825	380.676161
128	0.2	0.3	0.9	481.094409	385.715679

Table 1.18: Different RMSE for different α , β and γ values sorted

While sorting the different values of RMSE on train and test data of TES model, the α – 0.9 , β – 0.1 and γ – 0.2 has the lowest RMSE values across all the values

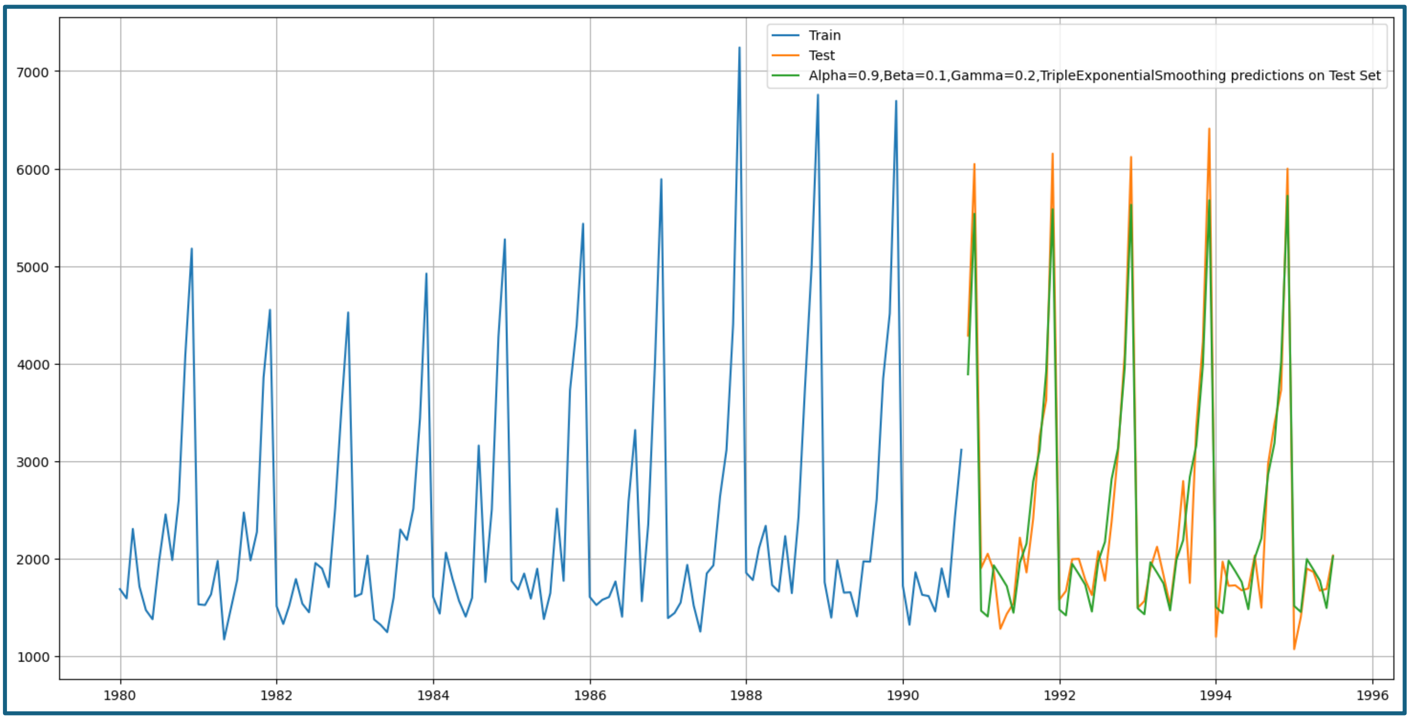


Figure 1.19: Alpha – 0.9, Beta – 0.1 & Gamma – 0.2 value Plot on Triple Exponential Smoothing

RMSE

	Test RMSE
RegressionOnTime	1392.438305
2pointTrailingMovingAverage	811.178937
4pointTrailingMovingAverage	1184.213295
6pointTrailingMovingAverage	1337.200524
9pointTrailingMovingAverage	1422.653281
Alpha=0.038003579704776386,SimpleExponentialSmoothing	1362.355525
Alpha=0.4,SimpleExponentialSmoothing	1363.037803
Alpha=0.3,Beta=0.3,DoubleExponentialSmoothing	1597.853999
Alpha=0.075,Beta=0.064,Gamma=0.376,TripleExponentialSmoothing	381.657232
Alpha=0.9,Beta=0.1,Gamma=0.2,TripleExponentialSmoothing	342.322388

Table 1.19: Different RMSE for different models

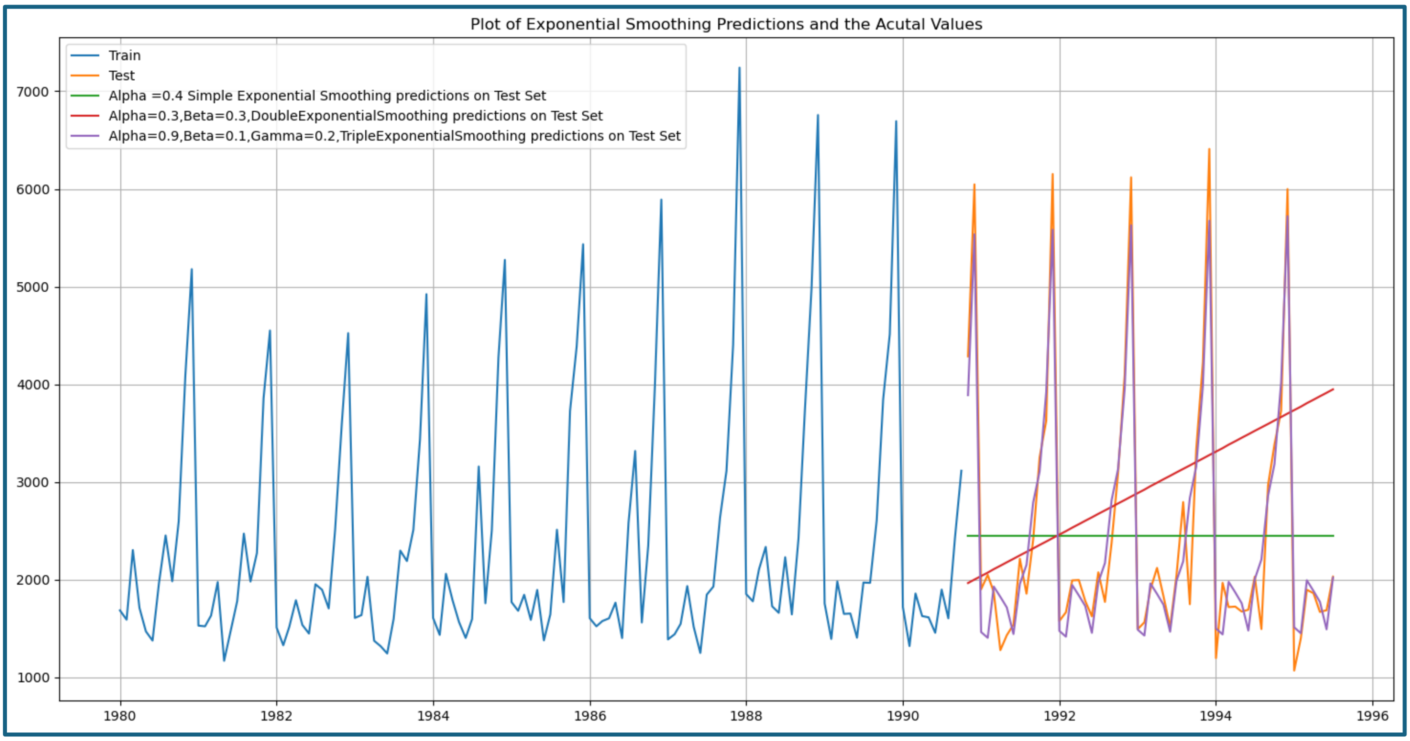


Figure 1.20: Exponential Smoothing predictions plot and Actual values

Full Model

The predictions on the full model values are based on the TES model with $\alpha - 0.9$, $\beta - 0.1$ and $\gamma - 0.2$ and RMSE of 342.322

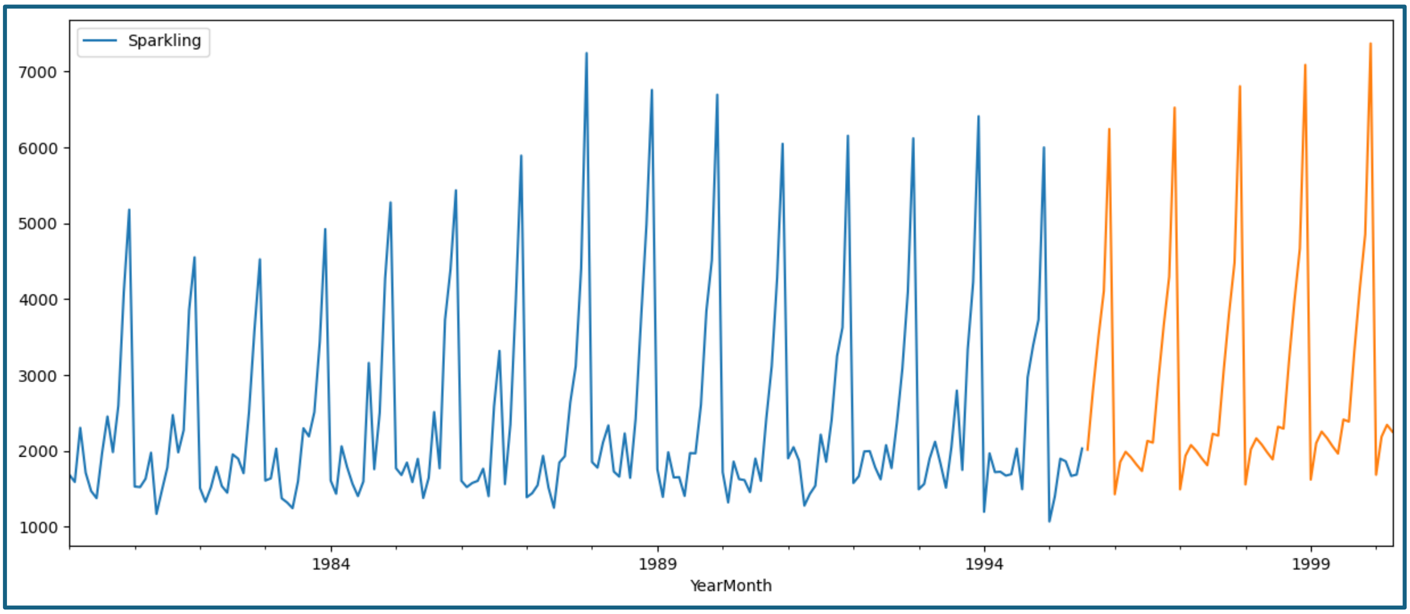


Figure 1.21: Predictions plot and Actual values on the full model

	lower_CI	prediction	upper_ci
1995-08-01	1072.699497	2015.899185	2959.098874
1995-09-01	1852.821704	2796.021393	3739.221081
1995-10-01	2552.622810	3495.822498	4439.022187
1995-11-01	3165.384431	4108.584120	5051.783808
1995-12-01	5298.253285	6241.452973	7184.652661

Table 1.20: First 5 rows of the predicted values on the full model

	lower_CI	prediction	upper_ci
1999-12-01	6425.807446	7369.007135	8312.206823
2000-01-01	742.501587	1685.701275	2628.900964
2000-02-01	1243.091326	2186.291014	3129.490703
2000-03-01	1399.805964	2343.005652	3286.205341
2000-04-01	1309.705321	2252.905009	3196.104698

Table 1.21: Last 5 rows of the predicted values on the full model

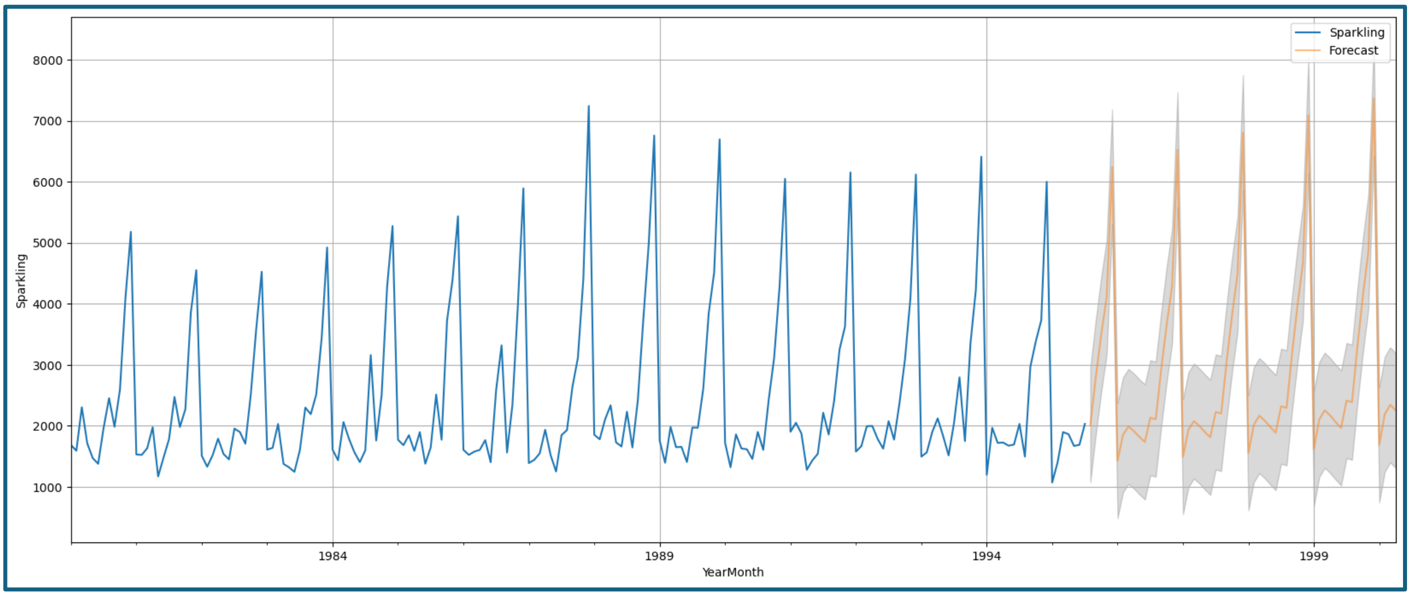


Figure 1.22: Forecast with Lower and Upper Limit and Actual values on the full model

The final plot shows the forecast values using TES model with 95% interval level.

3.6 Check the performance of the models built

	Test RMSE
Alpha=0.9,Beta=0.1,Gamma=0.2,TripleExponentialSmoothing	342.322388
Alpha=0.075,Beta=0.064,Gamma=0.376,TripleExponentialSmoothing	381.657232
2pointTrailingMovingAverage	811.178937
4pointTrailingMovingAverage	1184.213295
6pointTrailingMovingAverage	1337.200524
Alpha=0.038003579704776386,SimpleExponentialSmoothing	1362.355525
Alpha=0.4,SimpleExponentialSmoothing	1363.037803
RegressionOnTime	1392.438305
9pointTrailingMovingAverage	1422.653281
Alpha=0.3,Beta=0.3,DoubleExponentialSmoothing	1597.853999

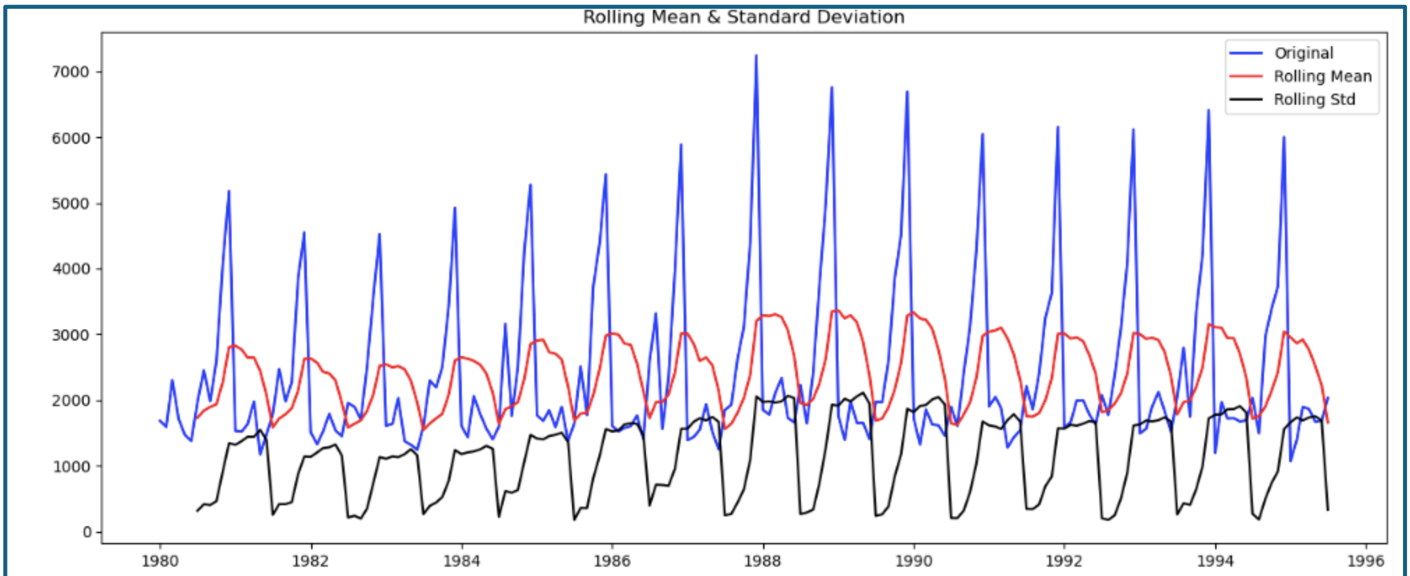
Table 1.22: Different RMSE for different models

Across all the models built the TES with best params Alpha – 0.9, Beta – 0.1 & Gamma – 0.2 value has the lowest RMSE value

4. Check for Stationarity

4.1 Check for stationarity

AD Fuller Test if p value > 0.05 then We fail to reject null hypothesis (H0 is true). We go for differencing if p value < 0.05 then the data is stationary

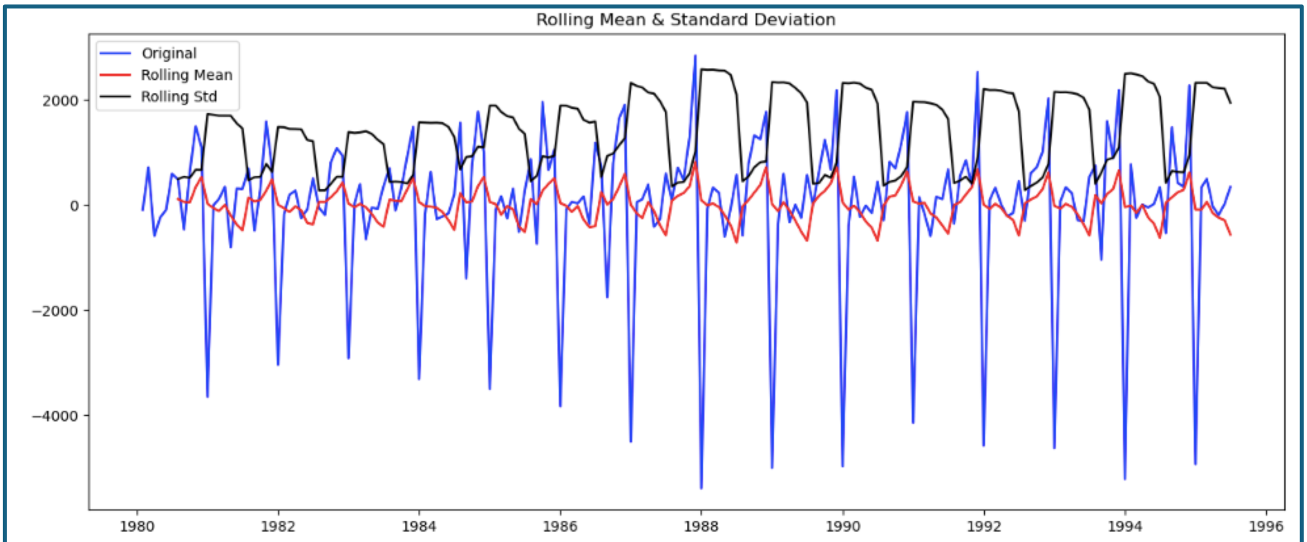



```
Results of Dickey-Fuller Test:
Test Statistic                  -1.360497
p-value                          0.601061
#Lags Used                      11.000000
Number of Observations Used    175.000000
Critical Value (1%)             -3.468280
Critical Value (5%)             -2.878202
Critical Value (10%)            -2.575653
dtype: float64
```

Figure 1.23: AD Fuller Test

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

4.2 Make the data stationary (if needed)




```
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
```

Figure 1.24: AD Fuller Test after differencing of order 1

The p-value is less than 0.05 and hence the data is stationary

5. Model Building - Stationary Data

5.1 Generate ACF & PACF Plot and find the AR, MA values

ACF

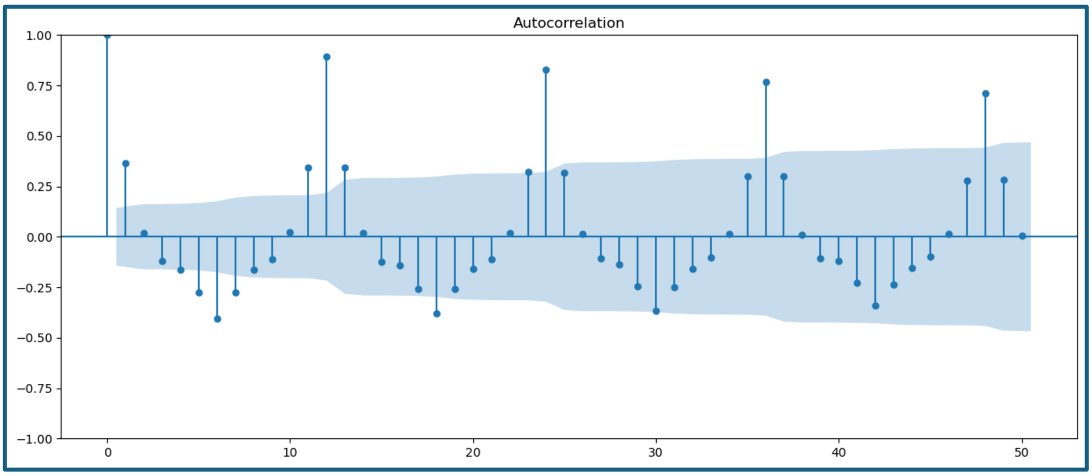


Figure 1.25: ACF

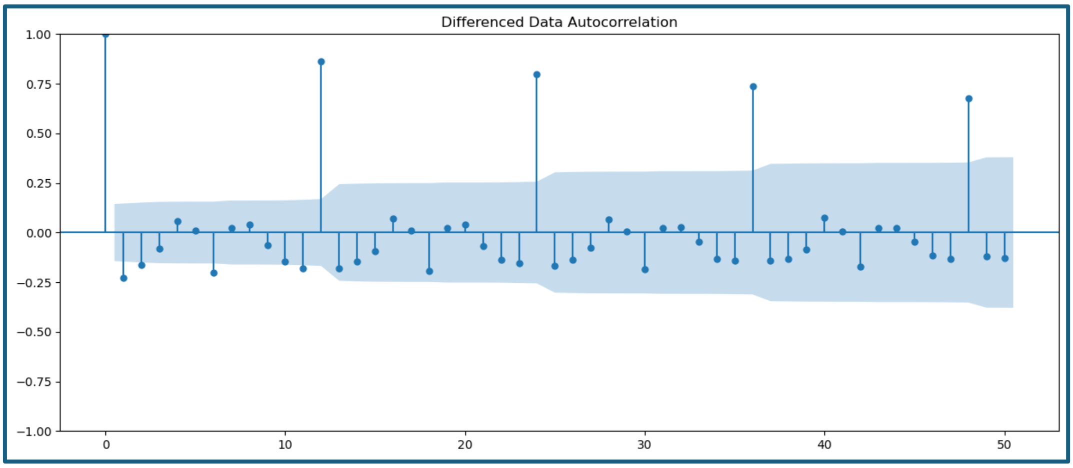


Figure 1.26: ACF with order 1 differencing

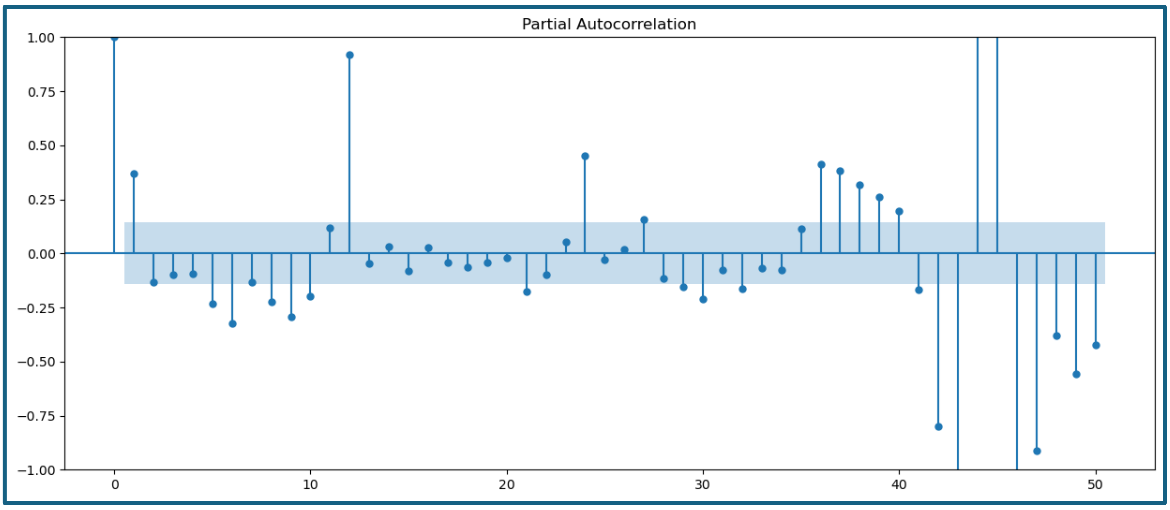


Figure 1.27: PACF

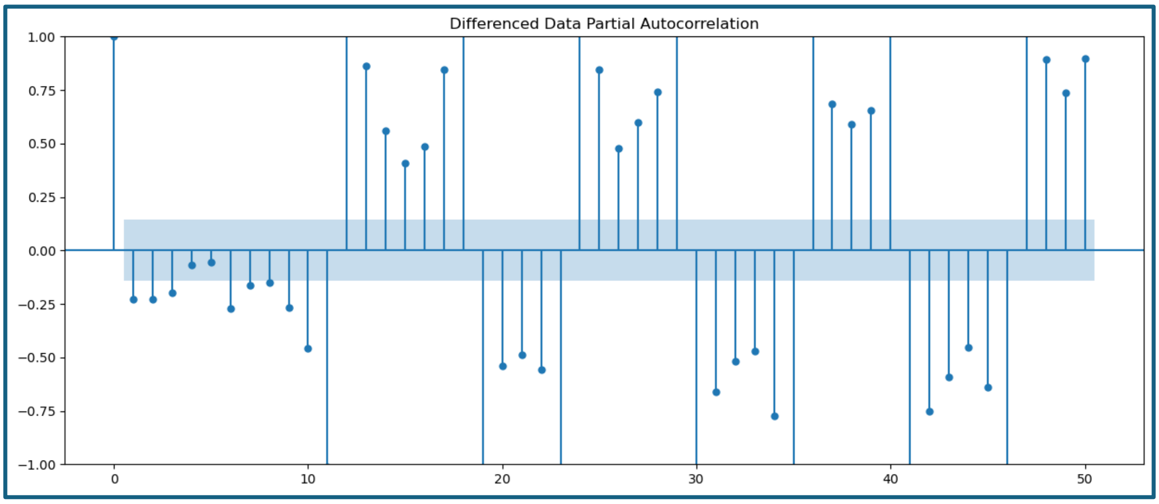


Figure 1.28: PACF with order 1 differencing

5.2 Build different ARIMA models

5.2.1 Auto ARIMA

Getting a combination of different parameters of p and q in the range of 0 and 2 the value of d as 1 as we need to take a difference of the series to make it stationary.

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

Sorted AIC values for different (p,d,q) combinations are derived using the ARIMA function and fitting the same.

	param	AIC
8	(2, 1, 2)	2178.109741
7	(2, 1, 1)	2193.974962
2	(0, 1, 2)	2194.034361
5	(1, 1, 2)	2194.959653
4	(1, 1, 1)	2196.050086
1	(0, 1, 1)	2217.939217
6	(2, 1, 0)	2223.899470
3	(1, 1, 0)	2231.137663
0	(0, 1, 0)	2232.719438

Table 1.23: Sorted AIC values in ARIMA Model

Getting the SARIMAX results by using the ARIMA function with the best params of p,d,q of (2,1,2)

```
                               SARIMAX Results
==============================================================================
Dep. Variable:              Sparkling   No. Observations:                  130
Model:                 ARIMA(2, 1, 2)   Log Likelihood               -1084.055
Date:                Fri, 12 Apr 2024   AIC                           2178.110
Time:                        20:13:40   BIC                           2192.409
Sample:                    01-01-1980   HQIC                          2183.920
                         - 10-01-1990
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          1.3020      0.046     28.555      0.000       1.213       1.391
ar.L2         -0.5360      0.079     -6.744      0.000      -0.692      -0.380
ma.L1         -1.9915      0.110    -18.146      0.000      -2.207      -1.776
ma.L2          0.9997      0.110      9.070      0.000       0.784       1.216
sigma2      1.085e+06   2.04e-07   5.31e+12      0.000    1.08e+06    1.08e+06
===================================================================================
Ljung-Box (L1) (Q):                   0.10   Jarque-Bera (JB):                19.53
Prob(Q):                              0.75   Prob(JB):                         0.00
Heteroskedasticity (H):               2.30   Skew:                             0.71
Prob(H) (two-sided):                  0.01   Kurtosis:                         4.27
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
[2] Covariance matrix is singular or near-singular, with condition number 4.55e+27. Standard errors may be unstabl
e.
```

Table 1.24: SARIMAX results in ARIMA Model

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

```
1990-11-01    3752.200234
1990-12-01    3824.775166
1991-01-01    3578.276765
1991-02-01    3218.439812
1991-03-01    2882.053991
```

RMSE value is calculated using the best params of (2,1,2) and added to a results DataFrame.

	RMSE
ARIMA(2,1,2)	1325.166412

Table 1.25: RMSE of ARIMA(2,1,2)

5.2.2 Manual ARIMA

Differentiated ACF & PACF plot to get the p & q value for the model building

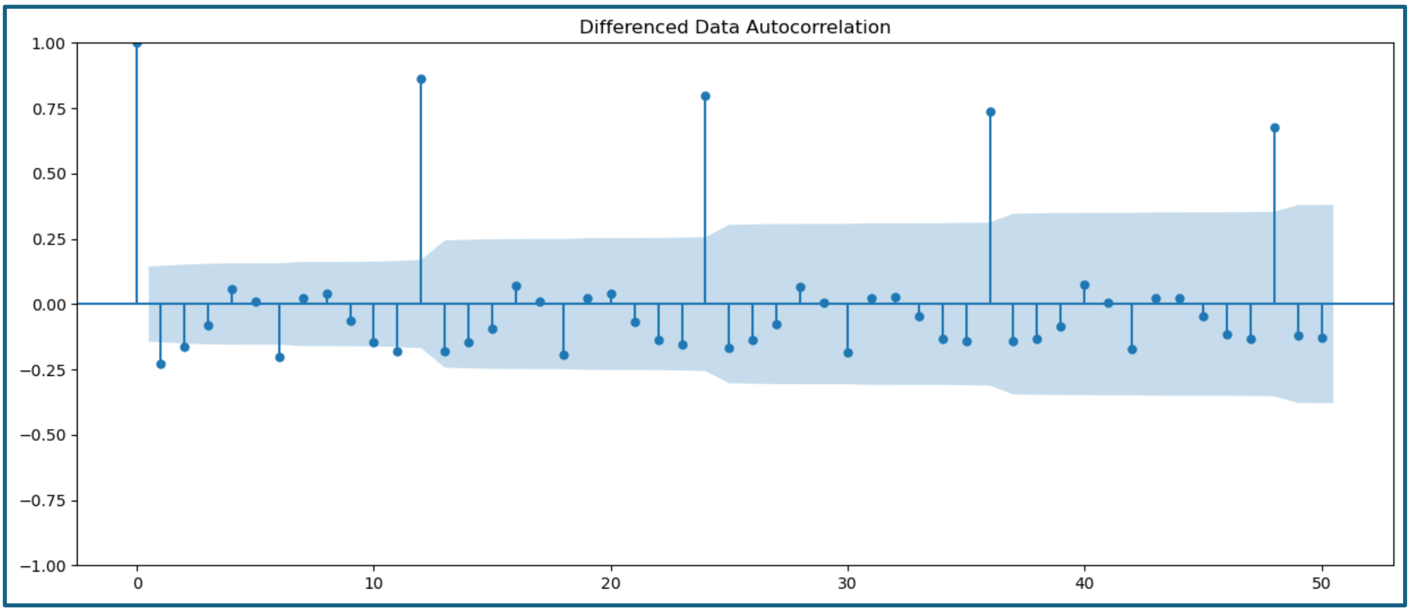


Figure 1.29: Differentiated ACF plot

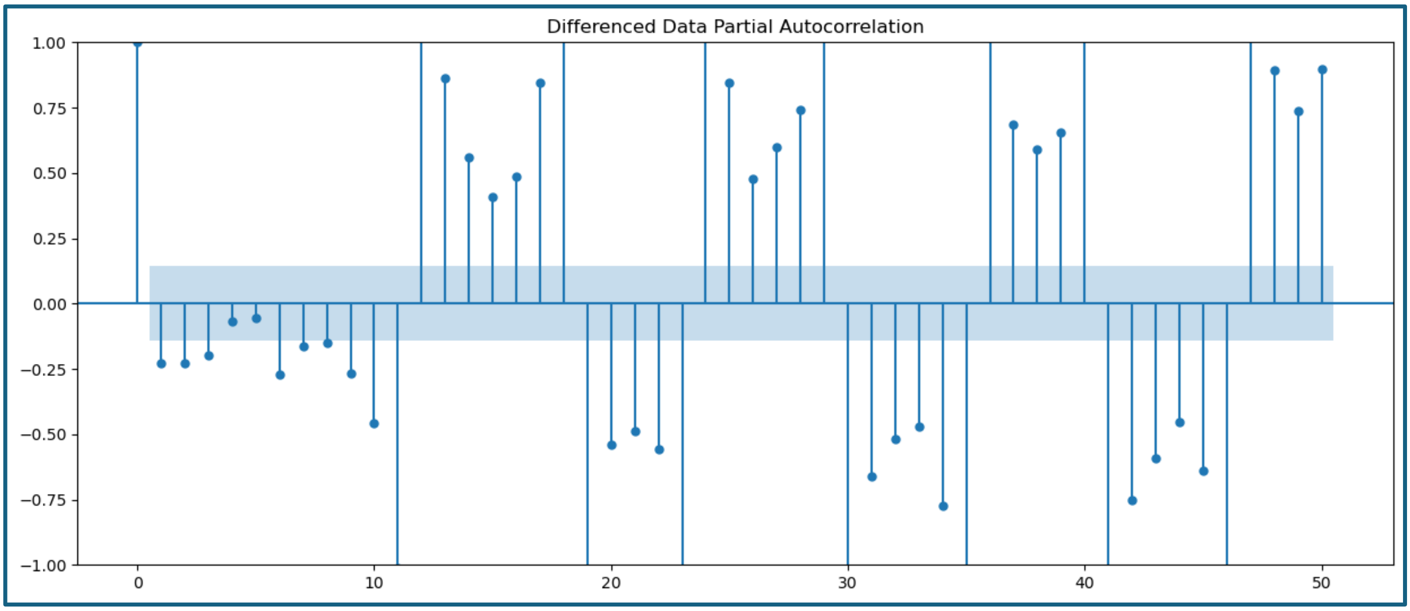


Figure 1.30: Differentiated PACF plot

Here, we have taken alpha=0.05.

The Auto-Regressive parameter in an ARIMA model is 'p' which comes from the significant lag before which the PACF plot cuts-off to 3.

The Moving-Average parameter in an ARIMA model is 'q' which comes from the significant lag before the ACF plot cuts-off to 2.

By looking at the above plots, we can say that both the PACF and ACF plot cuts-off at lag 0.

```
                                SARIMAX Results
==============================================================================
Dep. Variable:              Sparkling   No. Observations:                  130
Model:                 ARIMA(3, 1, 2)   Log Likelihood               -1089.814
Date:                Sun, 14 Apr 2024   AIC                           2191.628
Time:                        10:32:01   BIC                           2208.787
Sample:                    01-01-1980   HQIC                          2198.600
                         - 10-01-1990
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.4601      0.099     -4.644      0.000      -0.654      -0.266
ar.L2          0.3090      0.092      3.371      0.001       0.129       0.489
ar.L3         -0.2309      0.148     -1.559      0.119      -0.521       0.059
ma.L1         -0.0002     10.839  -2.16e-05      1.000     -21.244      21.244
ma.L2         -0.9998      0.148     -6.761      0.000      -1.290      -0.710
sigma2      1.205e+06   8.95e-06   1.35e+11      0.000      1.2e+06     1.2e+06
===================================================================================
Ljung-Box (L1) (Q):                   0.02   Jarque-Bera (JB):                 9.66
Prob(Q):                              0.90   Prob(JB):                         0.01
Heteroskedasticity (H):               2.45   Skew:                             0.54
Prob(H) (two-sided):                  0.00   Kurtosis:                         3.80
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
[2] Covariance matrix is singular or near-singular, with condition number 2.09e+28. Standard errors may be unstabl
e.
```

Table 1.26: SARIMAX results in Manual ARIMA Model

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

RMSE value is calculated using the best params of (3,1,2) and added to a results DataFrame.

	RMSE
ARIMA(2,1,2)	1325.166412
ARIMA(3,1,2)	1341.095895

Table 1.27: RMSE of both Auto & Manual ARIMA models

The Auto ARIMA(2,1,2) has the lowest RMSE compared to the Manual ARIMA(3,1,2)

5.3 Build different SARIMA models

5.3.1 Auto SARIMA

Checking the ACF plot once again

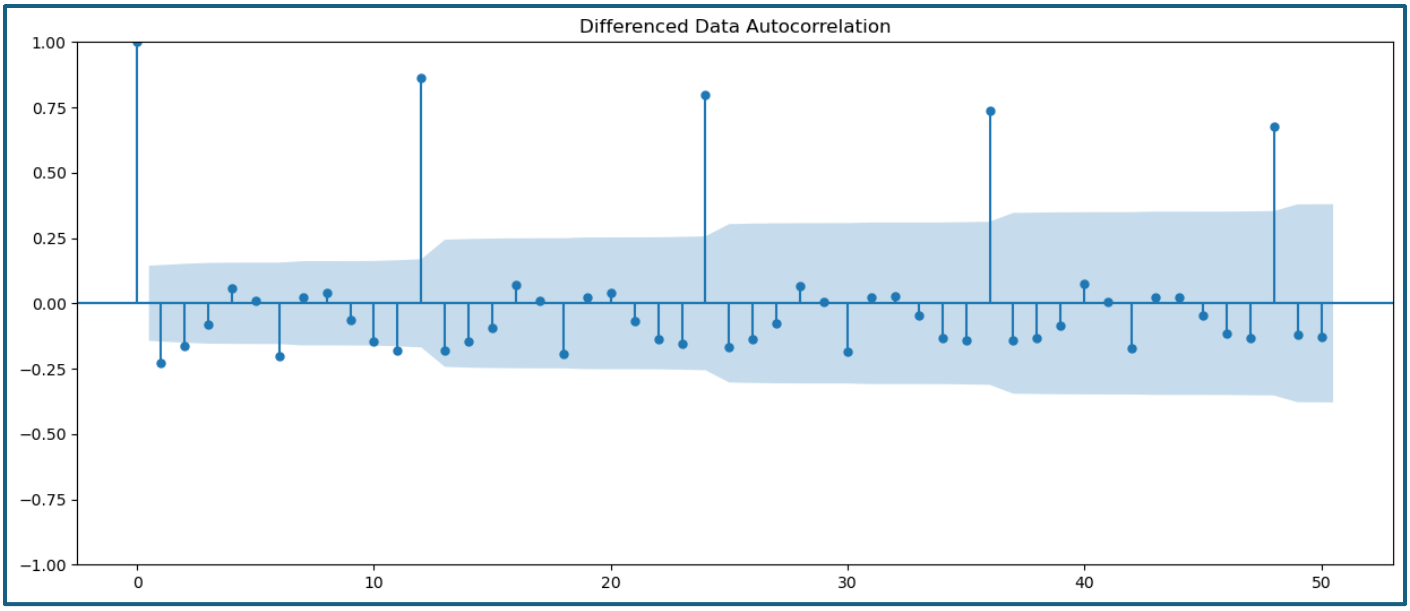


Figure 1.31: Differentiated ACF plot

We see that there can be a seasonality of 6 as well as 12 . We will run our auto SARIMA models by setting seasonality both as 6 and 12.

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

Getting a combination of different parameters of p and q in the range of 0 and 2 the value of d as 1 & 2 as we need to take a difference of the series to make it stationary with P,D,Q and seasonality as 6

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

Running the SARIMA model for all the combination of (p,d,q) & (P,D,Q,S)

	param	seasonal	AIC
53	(1, 1, 2)	(2, 0, 2, 6)	1694.342838
26	(0, 1, 2)	(2, 0, 2, 6)	1694.839211
80	(2, 1, 2)	(2, 0, 2, 6)	1695.565323
17	(0, 1, 1)	(2, 0, 2, 6)	1708.125767
44	(1, 1, 1)	(2, 0, 2, 6)	1710.045544

Table 1.28: AIC value of SARIMA model sorted

Building the model again with the best params (1,1,2) (2,0,2,6)

```
                                      SARIMAX Results
==========================================================================================
Dep. Variable:                                  y   No. Observations:                  130
Model:             SARIMAX(1, 1, 2)x(2, 0, 2, 6)   Log Likelihood                -839.171
Date:                            Fri, 12 Apr 2024   AIC                           1694.343
Time:                                    20:14:19   BIC                           1716.232
Sample:                                         0   HQIC                          1703.227
                                            - 130
Covariance Type:                              opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.6519      0.278     -2.349      0.019      -1.196      -0.108
ma.L1         -0.3008      0.240     -1.254      0.210      -0.771       0.169
ma.L2         -0.8551      0.260     -3.283      0.001      -1.366      -0.345
ar.S.L6       -0.0001      0.025     -0.005      0.996      -0.048       0.048
ar.S.L12       1.0526      0.018     57.014      0.000       1.016       1.089
ma.S.L6        0.0636      0.152      0.419      0.675      -0.234       0.361
ma.S.L12      -0.6546      0.087     -7.510      0.000      -0.825      -0.484
sigma2      1.177e+05   1.66e+04      7.075      0.000    8.51e+04     1.5e+05
===================================================================================
Ljung-Box (L1) (Q):                   0.26   Jarque-Bera (JB):                21.01
Prob(Q):                              0.61   Prob(JB):                         0.00
Heteroskedasticity (H):               2.75   Skew:                             0.37
Prob(H) (two-sided):                  0.00   Kurtosis:                         4.97
===================================================================================

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

Table 1.29: SARIMAX results in Manual SARIMA Model with best params

Model Diagnostic Plot

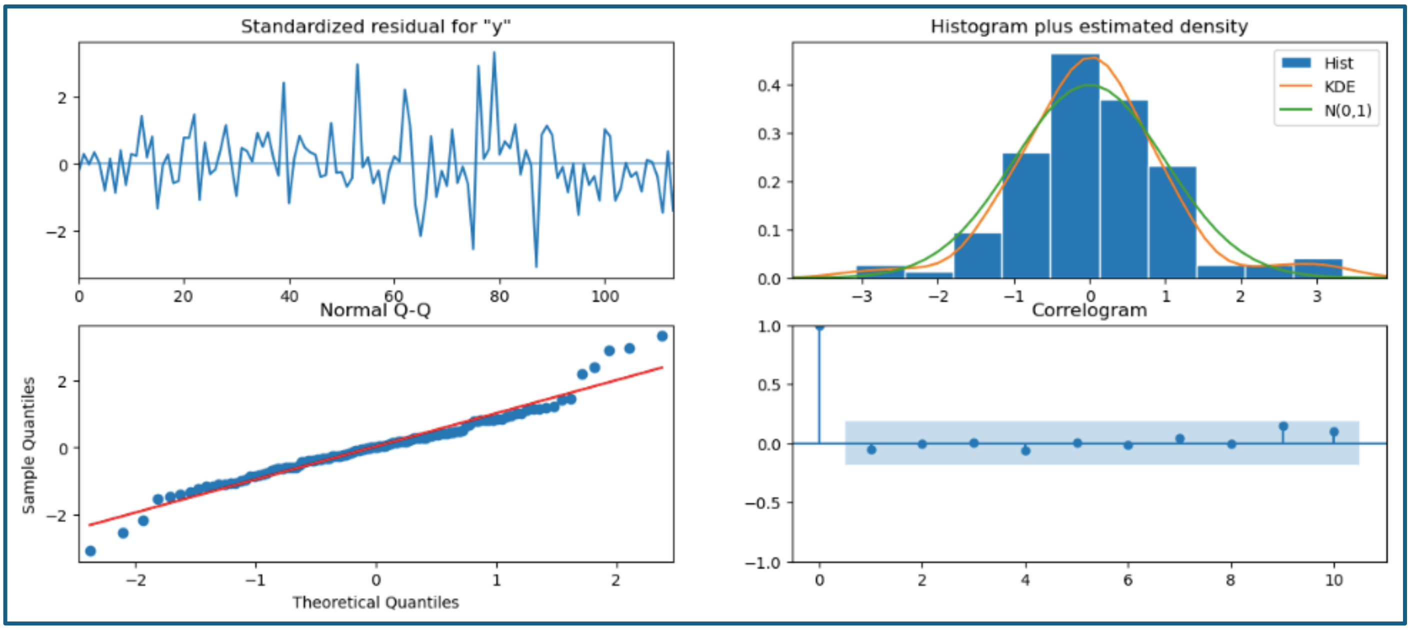


Figure 1.32: SARIMA Model Diagnostic Plot

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

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

y	mean	mean_se	mean_ci_lower	mean_ci_upper
0	4672.717101	373.045100	3941.562141	5403.872062
1	7093.584066	381.551723	6345.756431	7841.411701
2	1539.812232	381.553795	791.980535	2287.643929
3	1257.084753	385.165282	502.174673	2011.994833
4	1806.805787	385.639395	1050.966461	2562.645112

Table 1.30: Prediction with SARIMA Model of seasonality 6

	RMSE
ARIMA(2,1,2)	1325.166412
ARIMA(2,1,3)	1348.711388
SARIMA(1,1,2)(2,0,2,6)	642.829279

Table 1.31: Different model RSME value

The Auto SARIMA model with seasonality 6 has the lowest RSME value till now.

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

Getting a combination of different parameters of p and q in the range of 0 and 2 the value of d as 1 & 2 as we need to take a difference of the series to make it stationary with P,D,Q and seasonality as 12

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

Running the SARIMA model for all the combination of (p,d,q) & (P,D,Q,S)

	param	seasonal	AIC
53	(1, 1, 2)	(2, 0, 2, 12)	1521.737957
50	(1, 1, 2)	(1, 0, 2, 12)	1521.949483
80	(2, 1, 2)	(2, 0, 2, 12)	1523.208356
77	(2, 1, 2)	(1, 0, 2, 12)	1523.524949
26	(0, 1, 2)	(2, 0, 2, 12)	1523.707297

Table 1.32: AIC value of SARIMA model sorted with seasonality 12

Building the model again with the best params (1,1,2) (2,0,2,12)

```
                                     SARIMAX Results
==========================================================================================
Dep. Variable:                                 y   No. Observations:                  130
Model:             SARIMAX(1, 1, 2)x(2, 0, 2, 12)   Log Likelihood                -752.869
Date:                            Fri, 12 Apr 2024   AIC                           1521.738
Time:                                    20:15:32   BIC                           1542.738
Sample:                                         0   HQIC                          1530.241
                                            - 130
Covariance Type:                              opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.6465      0.268     -2.414      0.016      -1.171      -0.122
ma.L1          0.2836      0.300      0.945      0.345      -0.305       0.872
ma.L2         -1.1683      0.331     -3.529      0.000      -1.817      -0.519
ar.S.L12       0.7532      0.508      1.483      0.138      -0.243       1.749
ar.S.L24       0.3249      0.541      0.601      0.548      -0.735       1.385
ma.S.L12      -0.9793      0.490     -1.997      0.046      -1.941      -0.018
ma.S.L24      -0.5626      0.670     -0.840      0.401      -1.875       0.750
sigma2      4.951e+04   2.51e+04      1.975      0.048     381.919    9.86e+04
===================================================================================
Ljung-Box (L1) (Q):                   0.16   Jarque-Bera (JB):                 8.05
Prob(Q):                              0.69   Prob(JB):                         0.02
Heteroskedasticity (H):               1.46   Skew:                             0.21
Prob(H) (two-sided):                  0.27   Kurtosis:                         4.31
===================================================================================

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

Table 1.33: SARIMAX results in Manual SARIMA Model with best params of seasonality 12

Model Diagnostic Plot

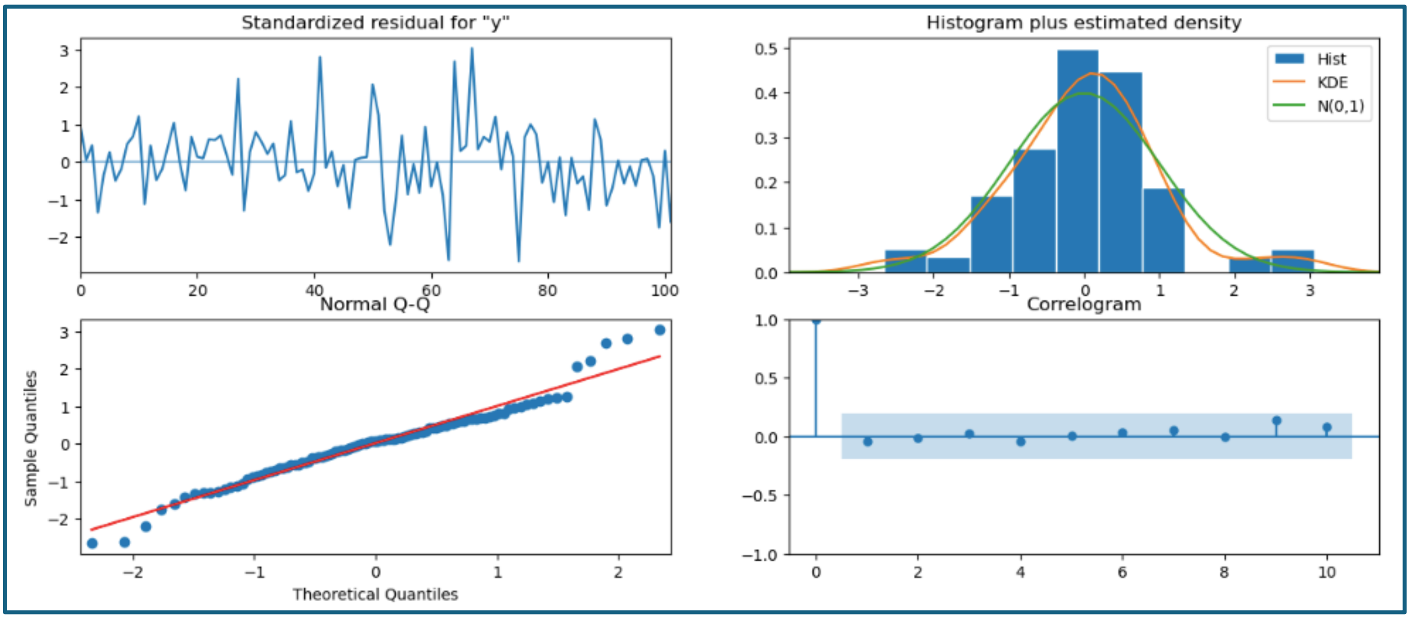


Figure 1.33: SARIMA Model Diagnostic Plot with seasonality 12

Similar to the last iteration of the model where the seasonality parameter was taken as 12, here also we see that the model diagnostics plot does not indicate any remaining information that we can get.

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

y	mean	mean_se	mean_ci_lower	mean_ci_upper
0	4695.832826	380.005189	3951.036343	5440.629310
1	7226.352997	388.862347	6464.196802	7988.509192
2	1584.409010	389.251758	821.489584	2347.328436
3	1417.334905	392.093979	648.844829	2185.824982
4	1828.712536	392.130443	1060.150991	2597.274081

Table 1.34: Prediction with SARIMA Model of seasonality 12

	RMSE
ARIMA(2,1,2)	1325.166412
ARIMA(2,1,3)	1348.711388
SARIMA(1,1,2)(2,0,2,6)	642.829279
SARIMA(1,1,2)(2,0,2,12)	712.724260

Table 1.35: Different model RSME value

Still Auto SARIMA model with seasonality 6 has the lowest RSME value till now.

5.3.2 Manual SARIMA

Differentiated ACF & PACF plot to get the p & q value for the model building

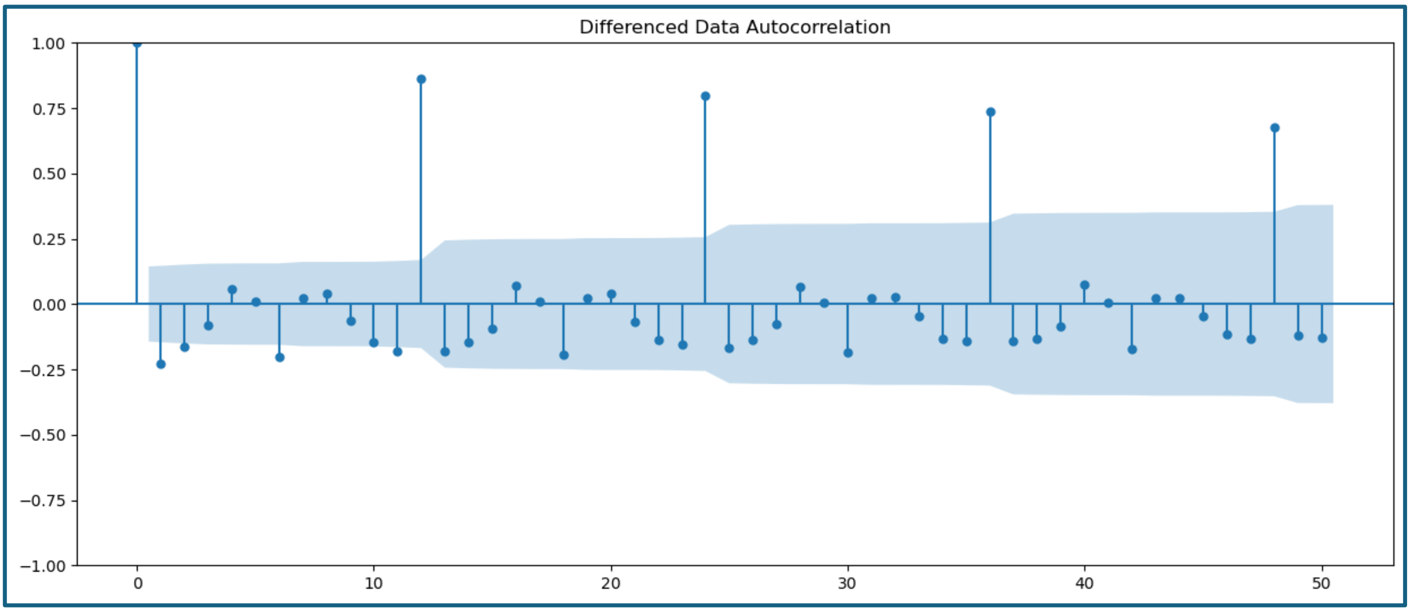


Figure 1.34: Differentiated ACF plot

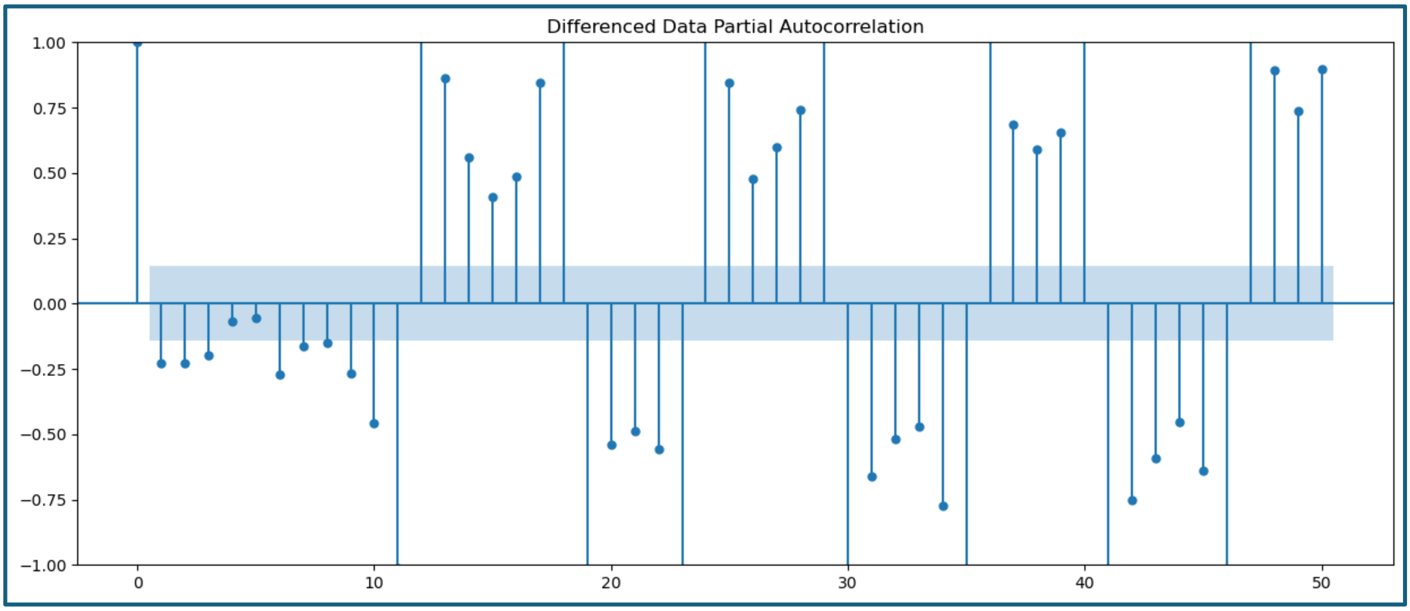


Figure 1.35: Differentiated PACF plot

We see that our ACF plot at the seasonal interval (6) does not taper off. So, we go ahead and take a seasonal differencing of the original series.

Original Data

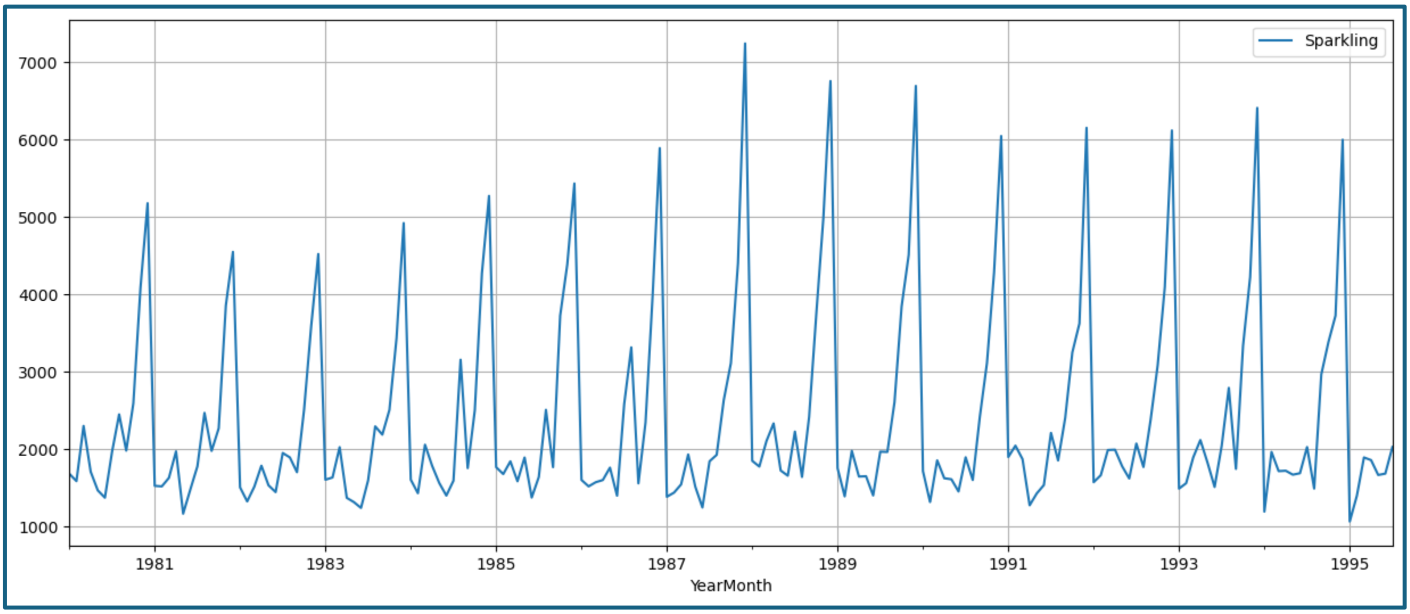


Figure 1.36: Original Data Plot

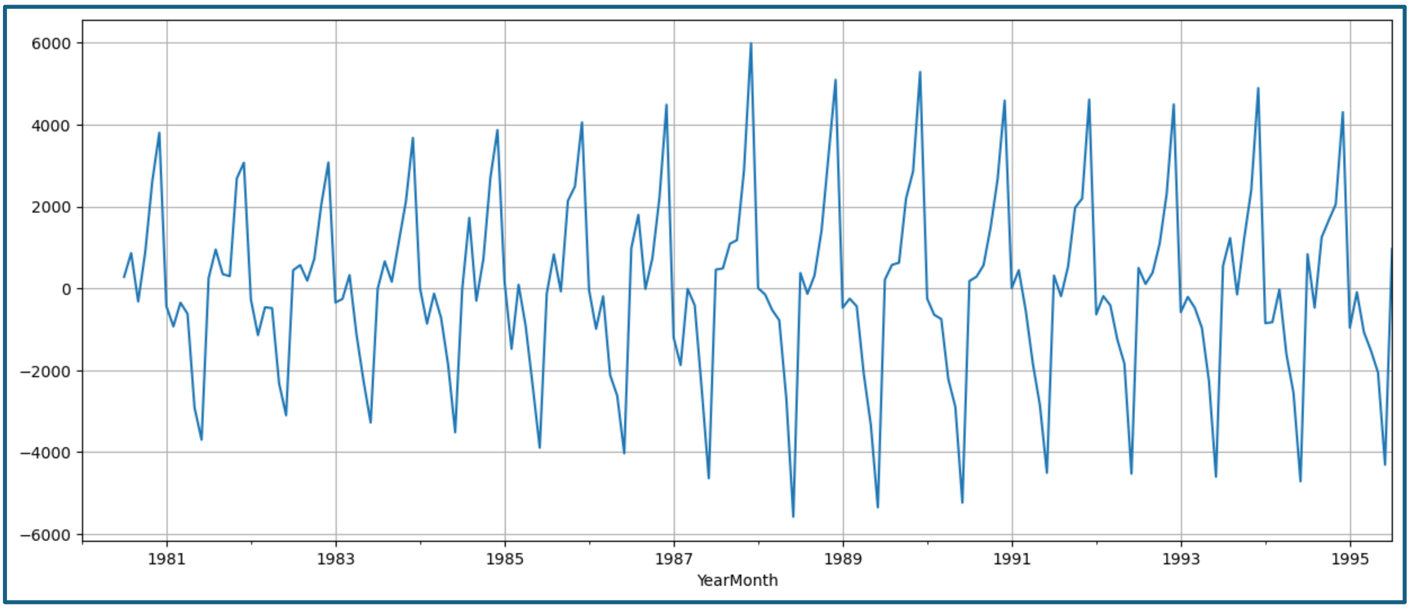


Figure 1.37: Differentiated Plot with Seasonality 6

We see that there might be a slight trend which can be noticed in the data. So we take a differencing of first order on the seasonally differenced series.

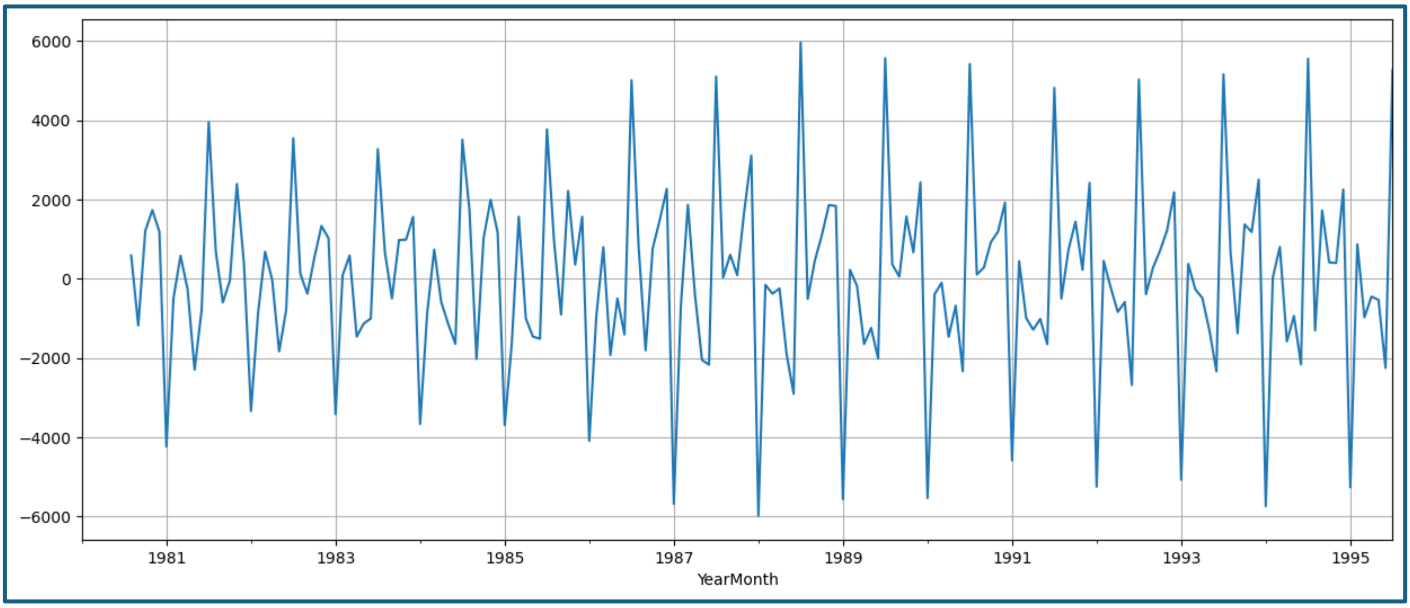
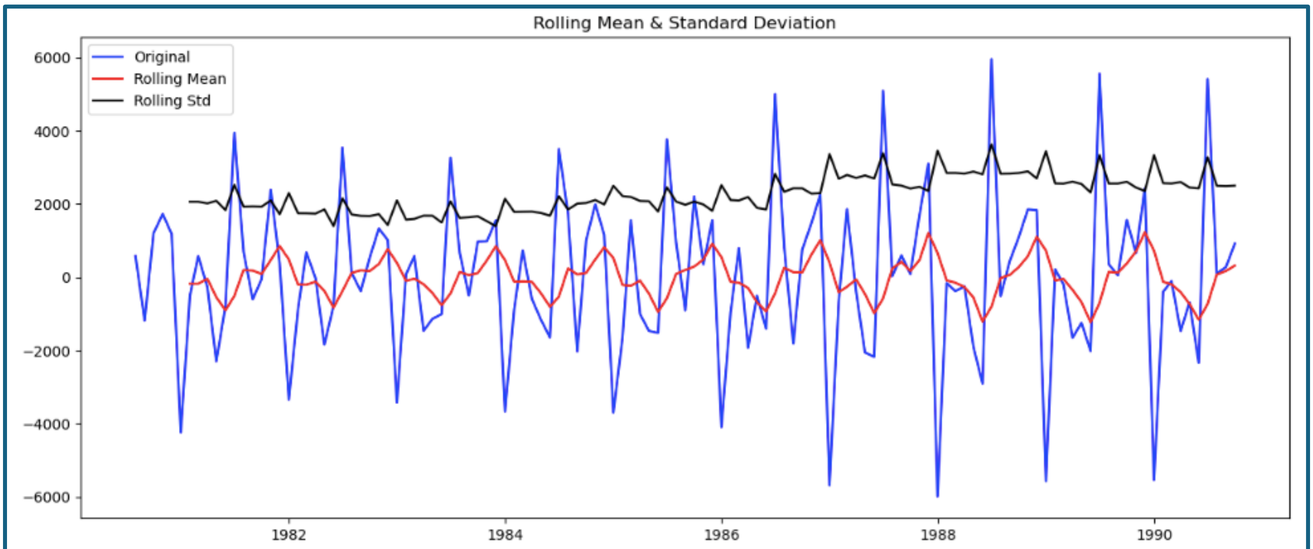


Figure 1.38: Differentiated 1 order after seasonality 6

Now we see that there is almost no trend present in the data. Seasonality is only present in the data. Let us go ahead and check the stationarity of the above series before fitting the SARIMA model.




```
Results of Dickey-Fuller Test:
Test Statistic                 -7.009360e+00
p-value                         6.986500e-10
#Lags Used                      1.300000e+01
Number of Observations Used     1.090000e+02
Critical Value (1%)            -3.491818e+00
Critical Value (5%)            -2.888444e+00
Critical Value (10%)           -2.581120e+00
dtype: float64
```

Figure 1.39: Differentiated PACF plot

The p-value is less than 0.05 and the data is stationary

Checking the ACF and the PACF plots for the new modified Time Series.

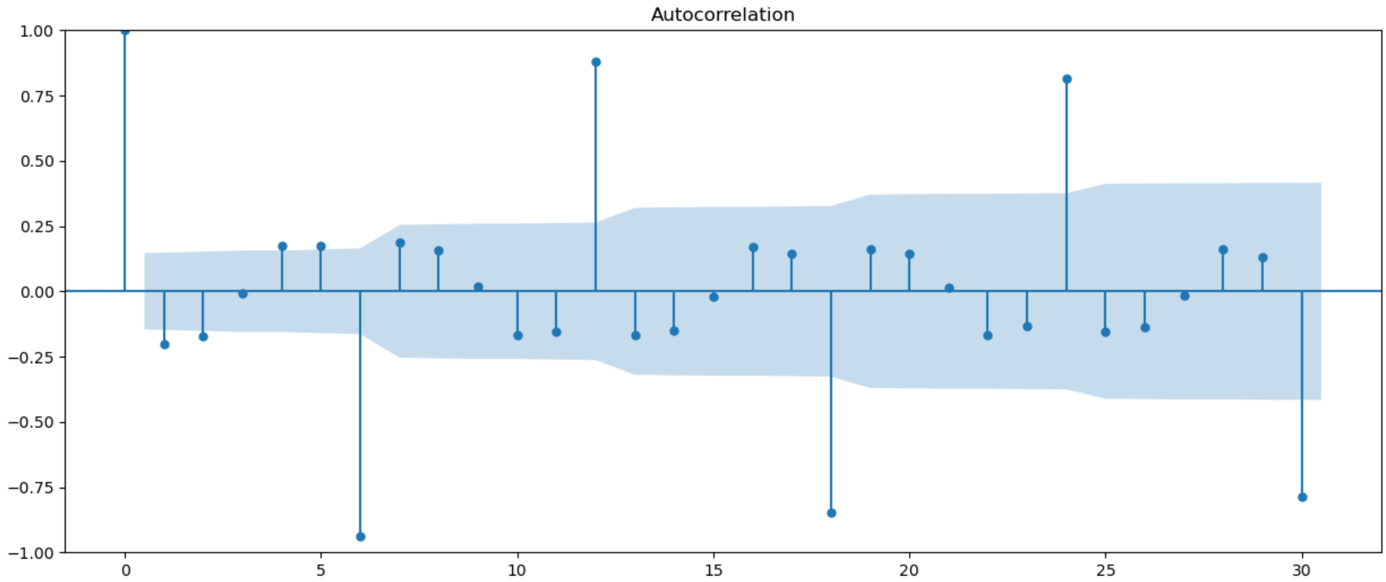


Figure 1.40: Modified ACF plot

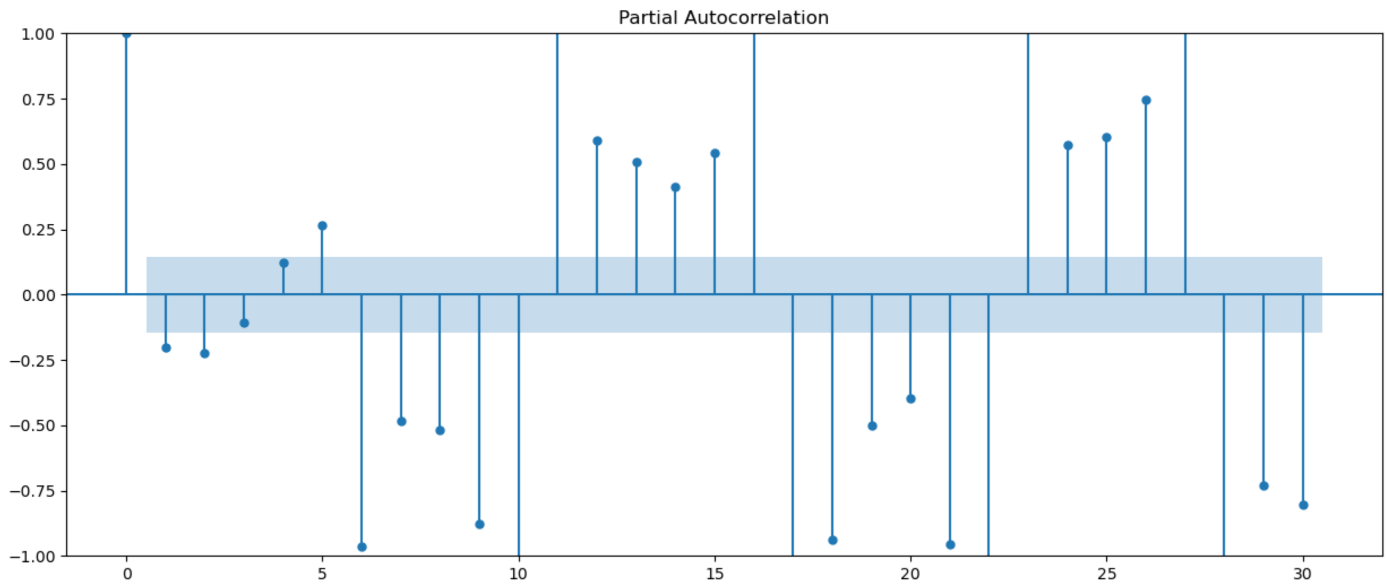


Figure 1.41: Modified PACF plot

Here, we have taken alpha=0.05.

We are going to take the seasonal period as 6. We will keep the p(1) and q(1) parameters same as the ARIMA model.

The Auto-Regressive parameter in an SARIMA model is 'P' which comes from the significant lag after which the PACF plot cuts-off to 2.

The Moving-Average parameter in an SARIMA model is 'q' which comes from the significant lag after which the ACF plot cuts-off to 2.

Model Diagnostic Plot

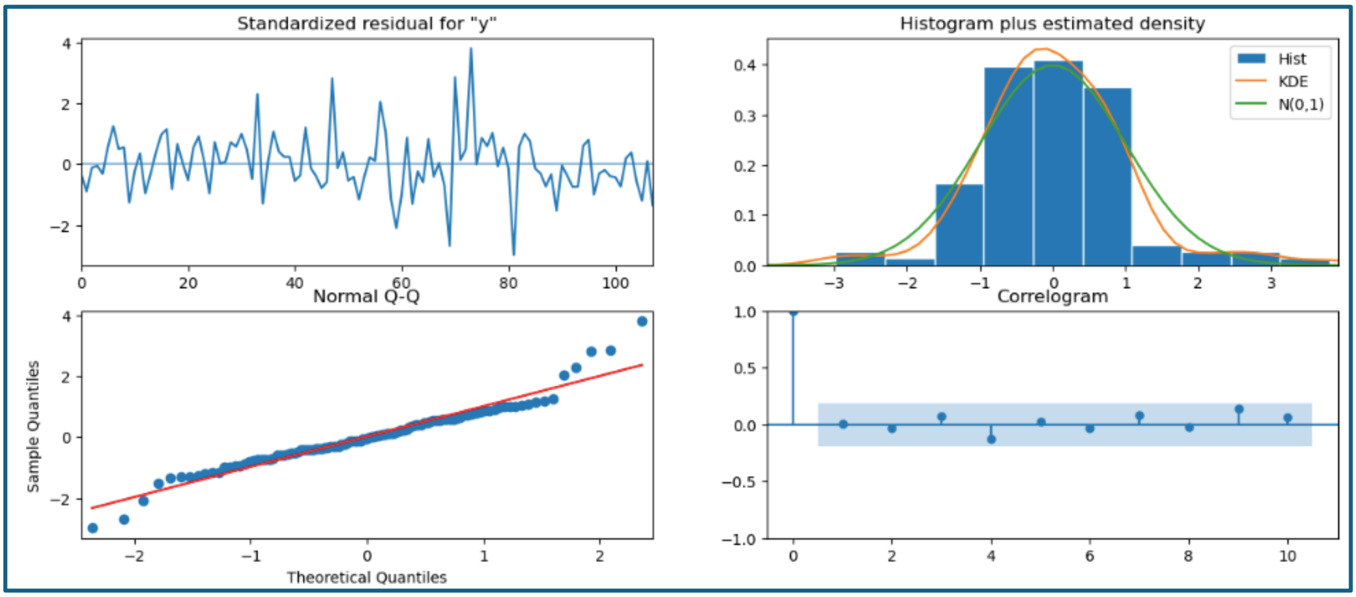


Figure 1.42: Manual SARIMA Model Diagnostic Plot with seasonality 6

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

y	mean	mean_se	mean_ci_lower	mean_ci_upper
0	4628.987382	389.485940	3865.608967	5392.365797
1	6888.685089	396.768602	6111.032918	7666.337260
2	1643.295018	396.772175	865.635845	2420.954190
3	1232.288954	397.204240	453.782950	2010.794958
4	1846.040814	398.715461	1064.572869	2627.508758

Table 1.36: Prediction with Manual SARIMA Model of seasonality 6

	RMSE
ARIMA(2,1,2)	1325.166412
ARIMA(2,1,3)	1348.711388
SARIMA(1,1,2)(2,0,2,6)	642.829279
SARIMA(1,1,2)(2,0,2,12)	712.724260
SARIMA(2,1,2)(2,1,2,6)	587.720803

Table 1.37: Different model RSME value

The Manual SARIMA model with seasonality 6 has the lowest RSME value till now.

Manual SARIMA with seasonality 12

Building the model again with the best params (1,1,2) (2,0,2,12)

```
                                     SARIMAX Results
==========================================================================================
Dep. Variable:                                  y   No. Observations:                  130
Model:             SARIMAX(2, 1, 2)x(2, 1, 2, 12)   Log Likelihood                -668.833
Date:                            Fri, 12 Apr 2024   AIC                           1355.665
Time:                                    20:20:59   BIC                           1378.163
Sample:                                         0   HQIC                          1364.738
                                            - 130
Covariance Type:                              opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.5699      0.276     -2.063      0.039      -1.111      -0.029
ar.L2          0.0136      0.147      0.092      0.926      -0.275       0.302
ma.L1         -0.1961      0.314     -0.624      0.532      -0.812       0.419
ma.L2         -0.8039      0.256     -3.139      0.002      -1.306      -0.302
ar.S.L12      -0.2824      0.472     -0.599      0.549      -1.206       0.642
ar.S.L24      -0.2641      0.218     -1.209      0.227      -0.692       0.164
ma.S.L12      -0.1330      0.475     -0.280      0.779      -1.064       0.798
ma.S.L24       0.1185      0.344      0.344      0.731      -0.556       0.793
sigma2       1.62e+05   1.66e-06   9.78e+10      0.000    1.62e+05    1.62e+05
===================================================================================
Ljung-Box (L1) (Q):                   0.01   Jarque-Bera (JB):                10.53
Prob(Q):                              0.92   Prob(JB):                         0.01
Heteroskedasticity (H):               0.89   Skew:                             0.50
Prob(H) (two-sided):                  0.76   Kurtosis:                         4.35
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
[2] Covariance matrix is singular or near-singular, with condition number 1.45e+27. Standard errors may be unstabl
e.
```

Table 1.38: SARIMAX results Manual SARIMA with seasonality 12

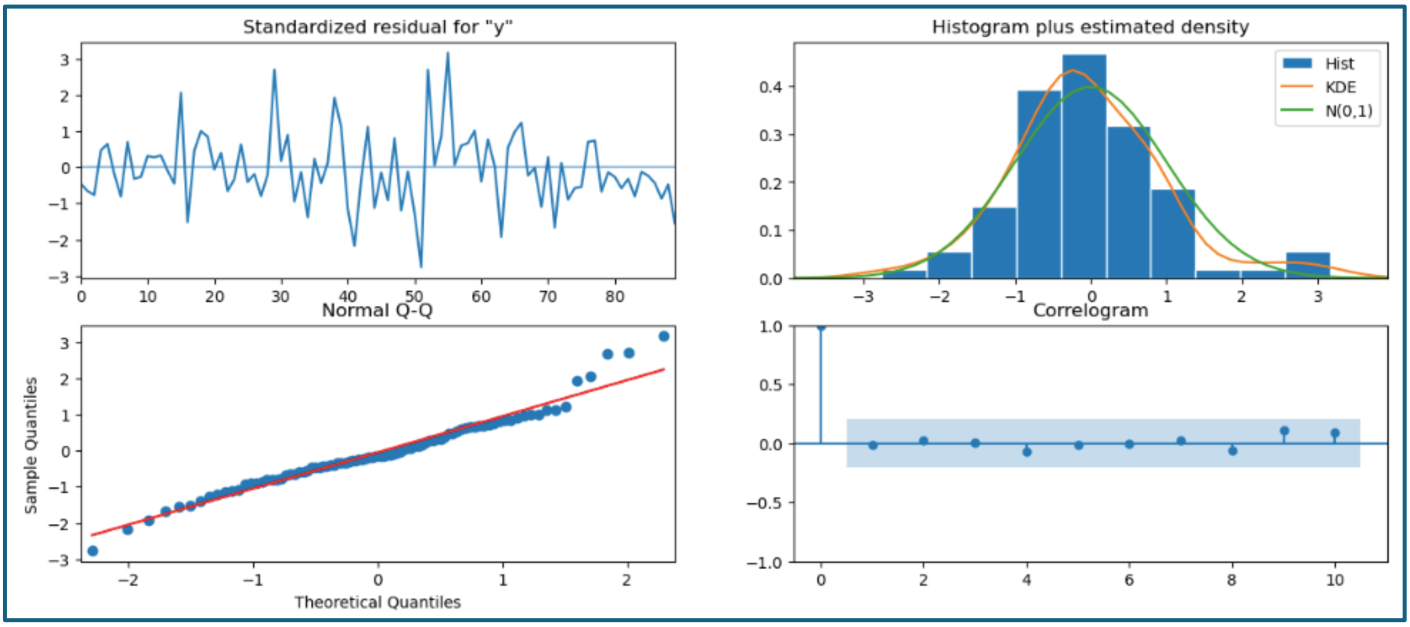


Figure 1.43: Manual SARIMA Model Diagnostic Plot with seasonality 12

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

y	mean	mean_se	mean_ci_lower	mean_ci_upper
0	4531.666718	404.548854	3738.765533	5324.567902
1	7018.733809	416.400310	6202.604199	7834.863420
2	1800.710841	418.633102	980.205039	2621.216643
3	1545.417534	419.938830	722.352551	2368.482516
4	2018.328957	420.092912	1194.961980	2841.695935

Table 1.39: Prediction with Manual SARIMA Model of seasonality 12

5.4 Check the performance of the models built

	RMSE
ARIMA(2,1,2)	1325.166412
ARIMA(2,1,3)	1348.711388
SARIMA(1,1,2)(2,0,2,6)	642.829279
SARIMA(1,1,2)(2,0,2,12)	712.724260
SARIMA(2,1,2)(2,1,2,6)	587.720803
SARIMA(2,1,2)(2,1,2,12)	486.396758

Table 1.40: Different model RSME value

In the models build using ARIMA & SARIMA the best model is Manual SARIMA with params of (2,1,2) (2,1,2,12) if seasonality 12 has the lowest RMSE among the other models.

6. Compare the performance of the models

6.1 Compare the performance of all the models built

The models with RMSE values are

	RMSE
RegressionOnTime	1392.438305
2pointTrailingMovingAverage	811.178937
4pointTrailingMovingAverage	1184.213295
6pointTrailingMovingAverage	1337.200524
9pointTrailingMovingAverage	1422.653281
Alpha=0.038003579704776386,SimpleExponentialSmoothing	1362.355525
Alpha=0.4,SimpleExponentialSmoothing	1363.037803
Alpha=0.3,Beta=0.3,DoubleExponentialSmoothing	1597.853999

	RMSE
Alpha=0.075,Beta=0.064,Gamma=0.376,TripleExponentialSmoothing	381.657232
Alpha=0.9,Beta=0.1,Gamma=0.2,TripleExponentialSmoothing	342.322388
ARIMA(2,1,2)	1325.166412
ARIMA(2,1,3)	1348.711388
SARIMA(1,1,2)(2,0,2,6)	642.829279
SARIMA(1,1,2)(2,0,2,12)	712.724260
SARIMA(2,1,2)(2,1,2,6)	587.720803
SARIMA(2,1,2)(2,1,2,12)	486.396758

Table 1.41: All model RSME value

6.2 Choose the best model with proper rationale

The best model to choose is the Manual SARIMA model with best params of (2,1,2) (2,1,2,12) and seasonality 12 has the lowest RMSE values across all the model.

We can build optimum model on the full data with this params.

6.3 Rebuild the best model using the entire data

Building the most optimum model on the Full Data.

```
                                      SARIMAX Results
==========================================================================================
Dep. Variable:                            Sparkling   No. Observations:               187
Model:             SARIMAX(2, 1, 2)x(2, 1, 2, 12)   Log Likelihood            -1085.726
Date:                            Fri, 12 Apr 2024   AIC                        2189.453
Time:                                    20:23:42   BIC                        2216.367
Sample:                                01-01-1980   HQIC                       2200.388
                                     - 07-01-1995
Covariance Type:                              opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.6655      0.275     -2.421      0.015      -1.204      -0.127
ar.L2         -0.0111      0.113     -0.099      0.921      -0.232       0.210
ma.L1         -0.1696      0.260     -0.652      0.515      -0.680       0.340
ma.L2         -0.7336      0.242     -3.032      0.002      -1.208      -0.259
ar.S.L12       0.0659      0.370      0.178      0.859      -0.659       0.791
ar.S.L24       0.2880      0.140      2.060      0.039       0.014       0.562
ma.S.L12      -0.7496      0.997     -0.752      0.452      -2.704       1.205
ma.S.L24      -0.2228      0.327     -0.682      0.495      -0.863       0.418
sigma2      1.332e+05   9.92e+04      1.343      0.179   -6.12e+04    3.28e+05
===================================================================================
Ljung-Box (L1) (Q):                   0.00   Jarque-Bera (JB):                26.11
Prob(Q):                              0.95   Prob(JB):                         0.00
Heteroskedasticity (H):               0.89   Skew:                             0.56
Prob(H) (two-sided):                  0.68   Kurtosis:                         4.74
===================================================================================

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

Table 1.42: SARIMAX results Manual SARIMA with seasonality 12

The SARIMAX results of SARIMA model with seasonality 12 along with the best params (2,1,2) (2,1,2,12)

6.4 Make a forecast for the next 12 months

Sparkling	mean	mean_se	mean_ci_lower	mean_ci_upper
1995-08-31	2132.78	373.93	1399.88	2865.68
1995-09-30	2359.40	379.07	1616.42	3102.37
1995-10-31	3266.88	379.15	2523.76	4010.00
1995-11-30	4040.25	381.49	3292.55	4787.95
1995-12-31	6101.12	381.59	5353.21	6849.03
1996-01-31	1339.52	382.79	589.26	2089.78
1996-02-29	1627.01	382.91	876.52	2377.51
1996-03-31	1852.84	383.72	1100.76	2604.92
1996-04-30	1794.72	383.98	1042.13	2547.30
1996-05-31	1662.21	384.68	908.26	2416.17
1996-06-30	1616.36	385.23	861.33	2371.39
1996-07-31	2025.59	385.89	1269.27	2781.92

Table 1.43: Predicted values for 12 months

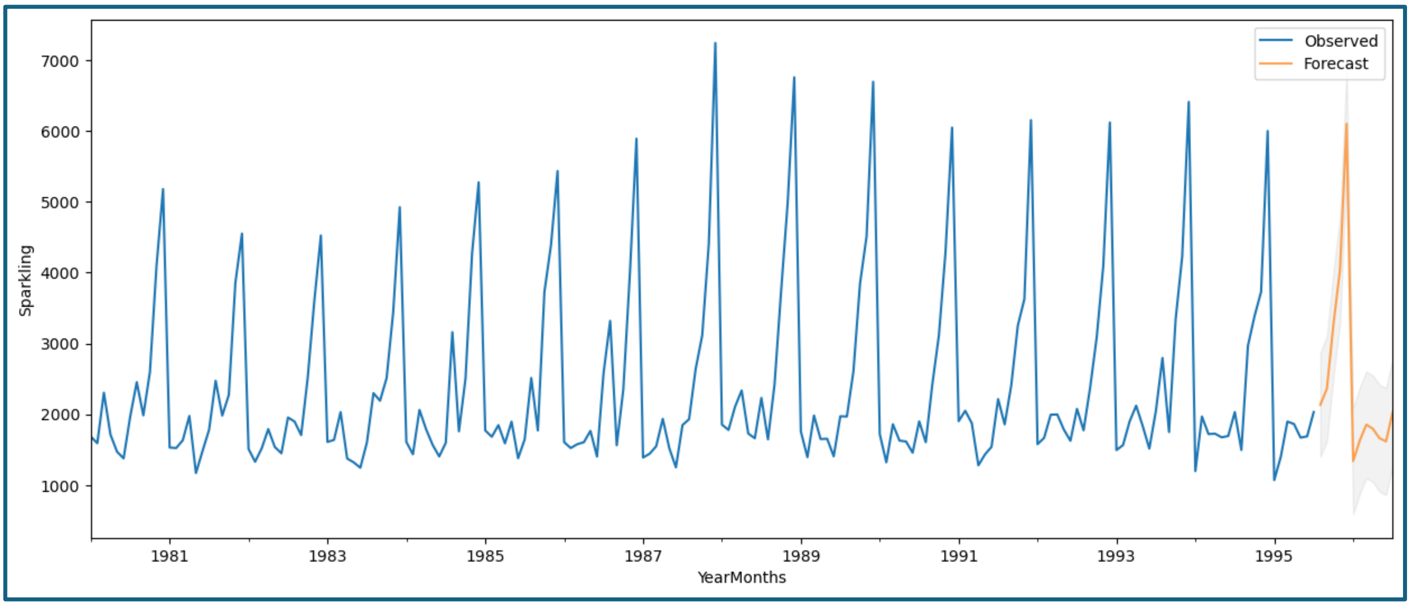


Figure 1.44: Final forecasted plot on the entire data

7. Actionable Insights & Recommendations

7.1 Key takeaways (actionable insights and recommendations) for the business

Insights:

- There is no certain trend in the data series
- There are outliers across all years data
- In monthly trend there was decrease trend from March to June and from July there was an increasing trend
- December month across years has the highest sales
- On quarterly basis the 4^{th} quarter has the highest sales
- The data is right skewed as the mean is greater than median
- In decomposition, multiplicative model is the suited model as there no specific pattern in the residual
- The data has no defined trend and has seasonality
- Across LR, SA, MA, SES, DES & TES – TES has the lowest RMSE value
- The DataFrame is not stationary and AD fuller method is used to make it stationary
- Among Auto ARIMA & SARIMA, Manual ARIMA & SARIMA – Manual SARIMA with seasonality 12 has the lowest RMSE value

Recommendations:

- The ABC Estate Wines company must ensure the stock availability of Sparkling type wines on a higher during the month of OCT, NOV & DEC on all years
- The marketing spend can be reduced during these high sales months
- July, Aug & Sept seems to be a avg sales months and the company can Optimize the production and marketing spend on these months
- January month can be ignored for marketing & stock availability as it contributes to the lowest sales due the highest sales in December
- Winter season is the influencing factor on the wine sales