**Deepak Singh Manola**

**26 OCT 2023**

**Great Learning – PGP – DSBA**

**Time Series Forecasting Project**

INTRODUCTION

This report consists of Time Series analysis and forecasting of 2 datasets -

* DATASET 1 - Sales data of Shoe Sales
* DATASET 2 - Sales data of Soft Drink

Please find the Jupyter Code Notebook. Analysis code is in Python.

# Problem 1 for the Data Set: [Shoesales.csv](https://olympus.mygreatlearning.com/courses/61442/files/5344205/download?wrap=1)

You are an analyst in the IJK shoe company, and you are expected to forecast the sales of the pairs of shoes for the upcoming 12 months from where the data ends. The data for the pair of shoe sales have been given to you from January 1980 to July 1995.

# Problem 2 for the Data Set [SoftDrink.csv](https://olympus.mygreatlearning.com/courses/61442/files/5344204/download?wrap=1):

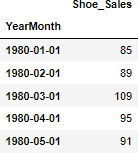
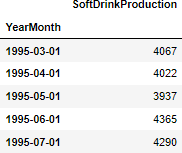
## You are an analyst in the RST soft drink company, and you are expected to forecast the sales of the production of the soft drink for the upcoming 12 months from where the data ends. The data for the production of soft drinks has been given to you from January 1980 to July 1995.

SYNOPSIS

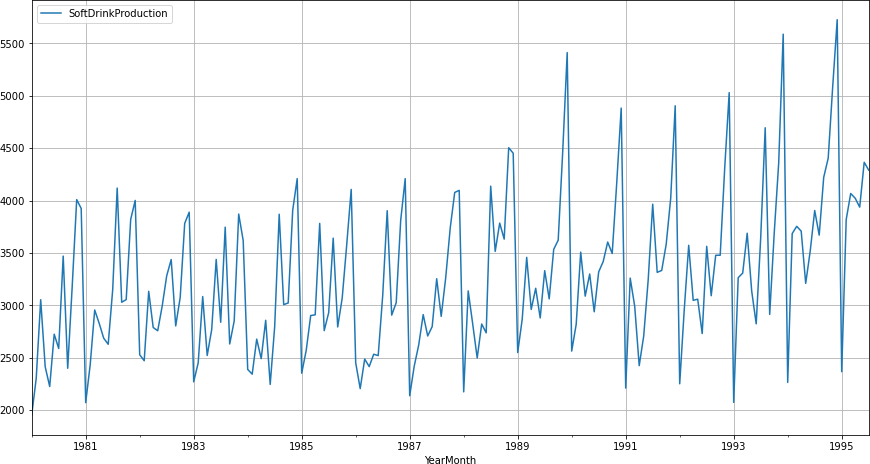
1. Total No. Of Shoe Sales Data Entries:187 Total No. Of Soft Drink Data Entries: 187
2. No. Of Missing Values in both data = 0
3. No. Of Duplicate entries in Shoe Sales data = 0 No. Of Duplicate entries in Soft Drink data = 0
4. Both datasets are split in Train : Test at year 1991 - Test data starts at 1991
5. Various forecasting models applied are -
   1. Linear Regression
   2. Naive Bayes
   3. Simple Average
   4. 2-pt Moving Average
   5. 4-pt Moving Average
   6. 6-pt Moving Average
   7. 9-pt Moving Average
   8. Single Exponential Smoothing
   9. Double Exponential Smoothing (Holt’s Model)
   10. Triple Exponential Smoothing (Holt-Winter Model)
   11. ARIMA / SARIMA (Auto fitted)
   12. ARIMA / SARIMA (Manually fitted)

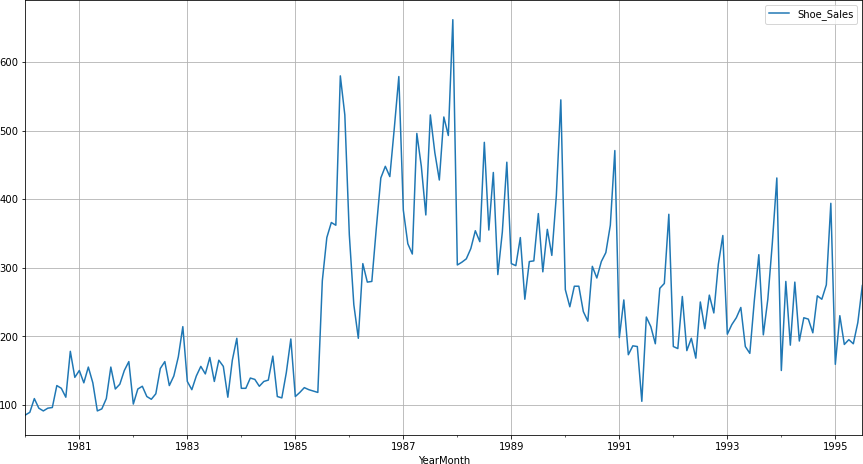
# [Q 1] Read the data as an appropriate Time Series data and plot the data.

* Both Datasets are read and stored as Pandas Data Frames for analysis
* First 5 rows of both the data are given below -



* Soft Drink Data plot -



* Shoe Sales Data plot -

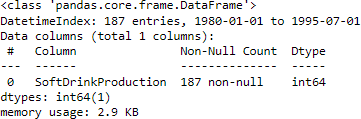
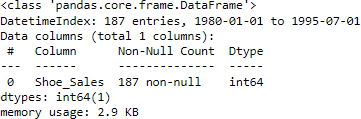
# [Q 2] Perform appropriate Exploratory Data Analysis to understand the data and also perform decomposition.

### Exploratory Data Analysis -



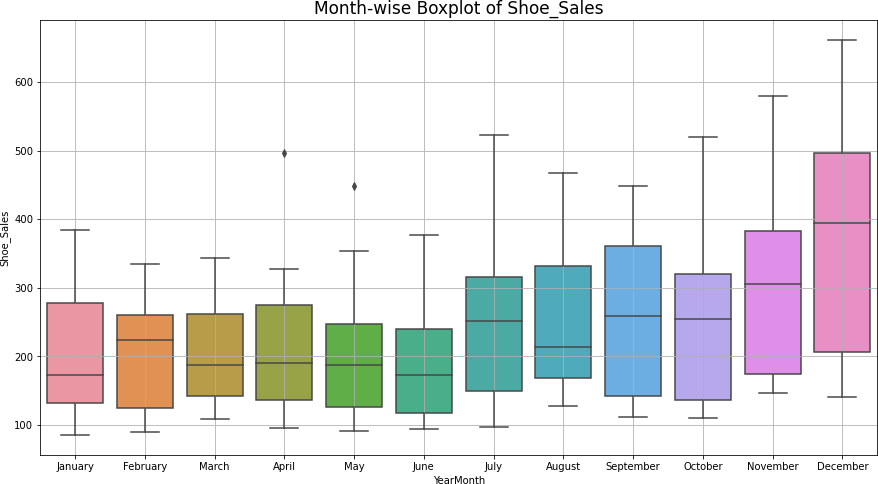


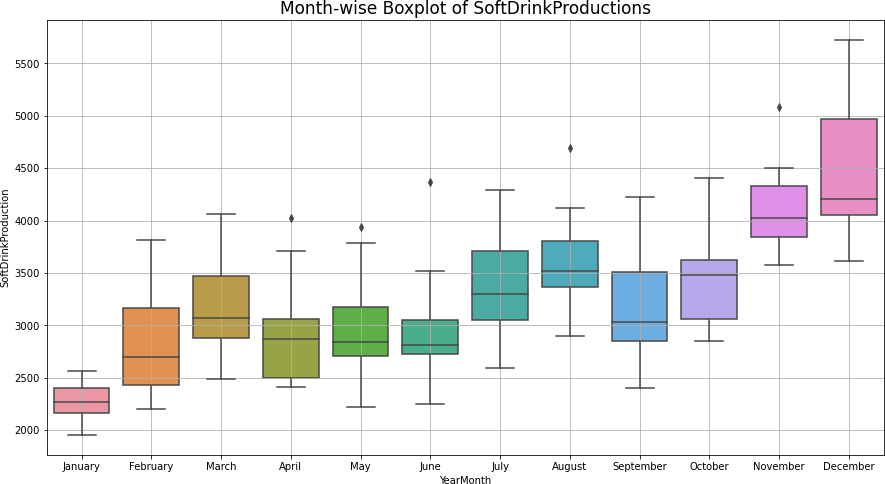
Descriptive Stats of Shoe Sales and Soft Drink datasets



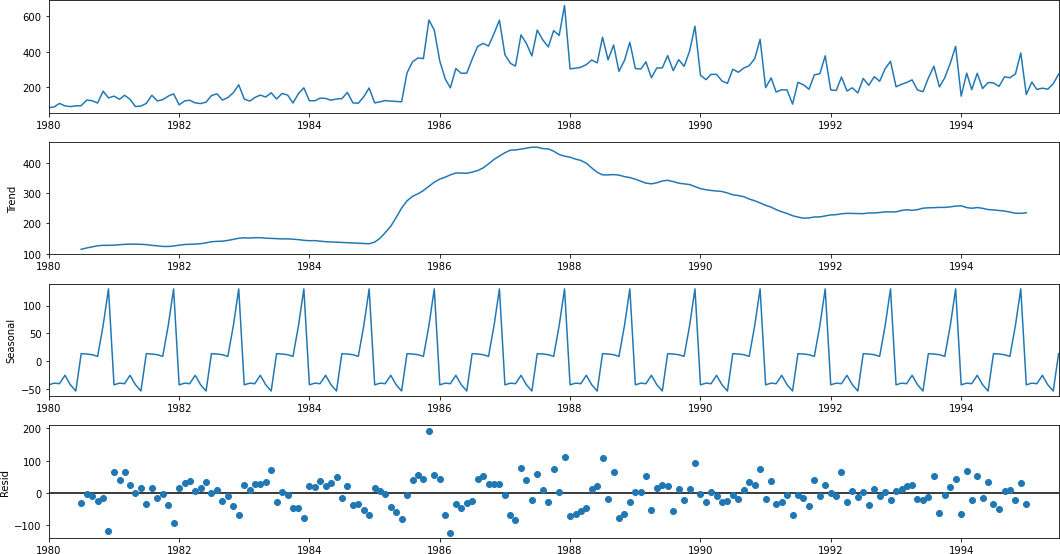
Info – Shoe Sales data Info – Soft Drink data

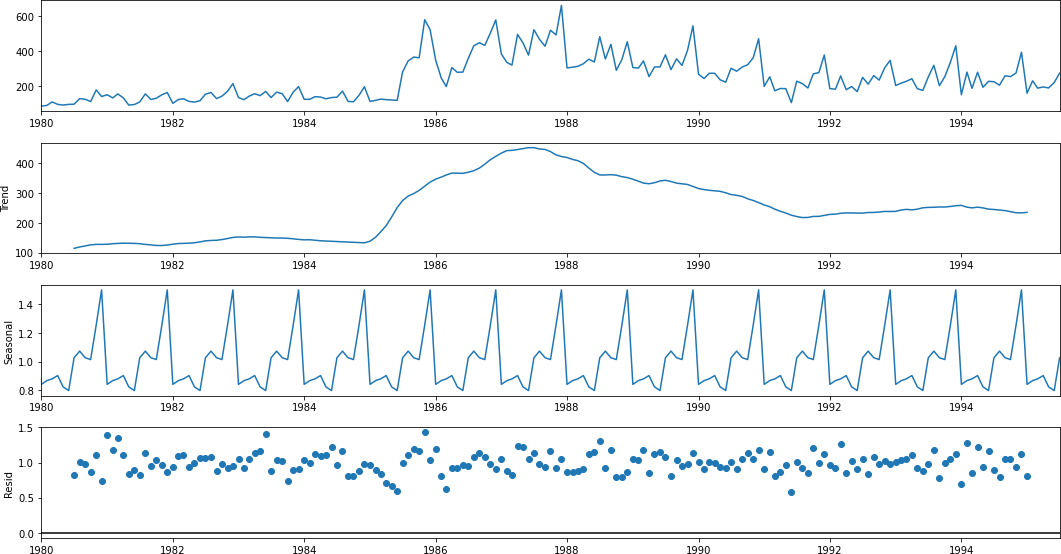
* Month-wise Boxplot of Shoe Sales -



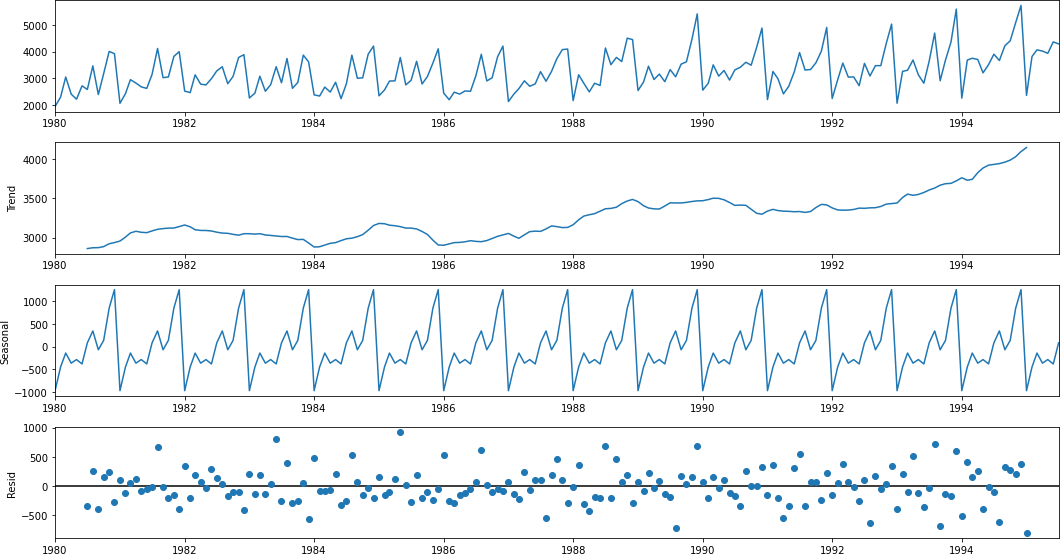
* Month-wise Boxplot of Soft Drink-
  + Sales of both show a spike in the last quarter of Oct- Dec
  + Spike is much more accentuated in Soft Drink sales.
  + This spike may be due to the Holiday season starting in Oct.

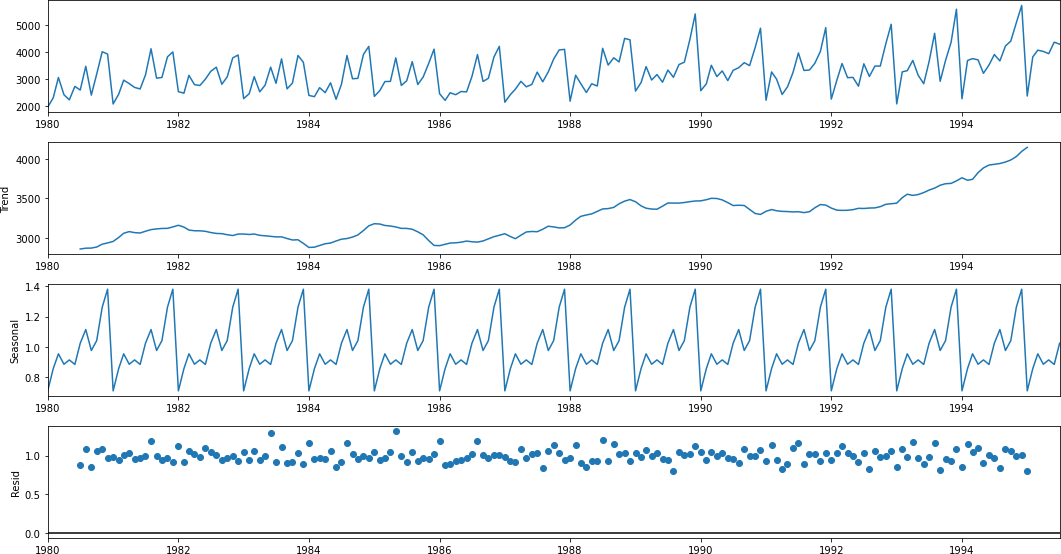
### Additive Decomposition of Shoe Sales:-



Multiplicative Decomposition of Shoe Sales:-

### Additive Decomposition of Soft Drink -

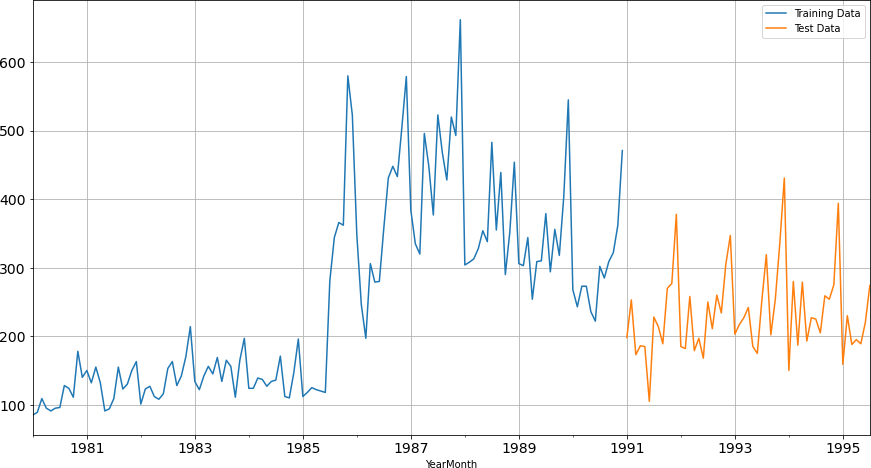




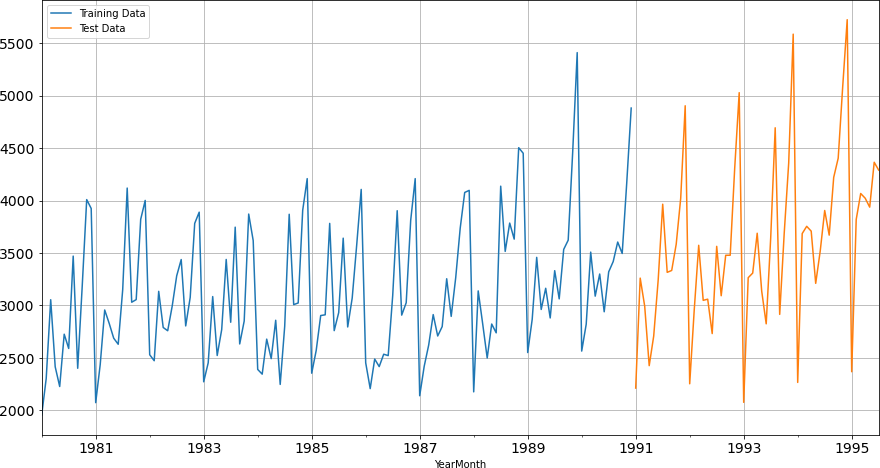
* Additive Models -
  + The seasonality is relatively constant over time
  + *yt* = *Trend* + *Seasonalit y* + *Residual*
* Multiplicative Models -
  + The seasonality increases or decreases over time. It is proportionate to the trend
  + *yt* = *Trend* \* *Seasonalit y* \* *Residual*
* Here by just observing the Residual patterns of Additive and Multiplicativemodels of both datasets. It seems that -

# [Q 3 ] Split the data into training and test. The test data should start in 1991.

* Both datasets are split at the year 1991
* Test datasets start at 1991

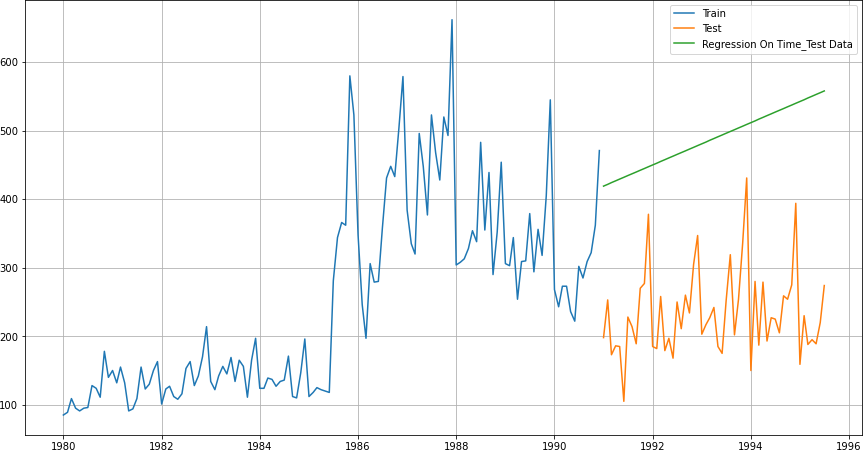


Soft Drink:

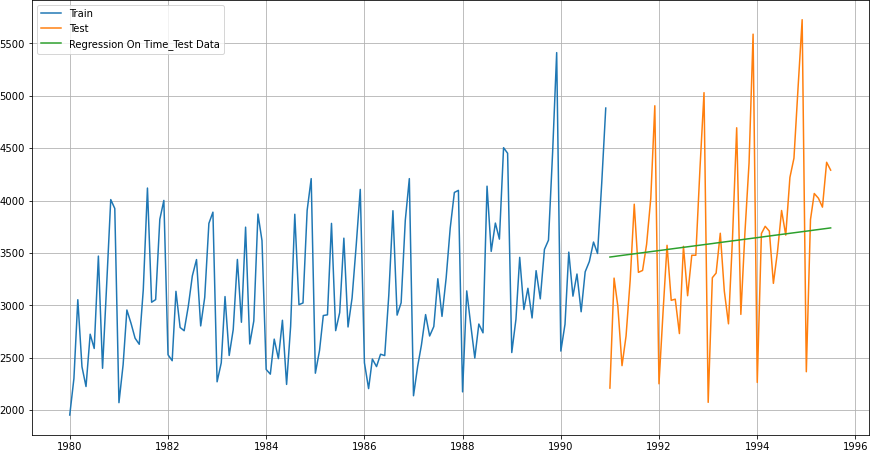


# [Q 4 ] Build various exponential smoothing models on the training data and evaluate the model using RMSE on the test data. Other models such as regression, naïve forecast models, simple average models etc. should also be built on the training data and check the performance on the test data using RMSE.

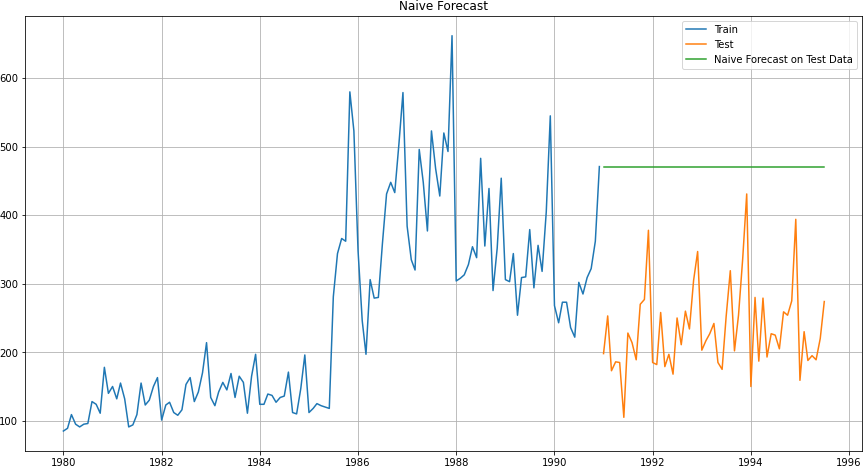
### Model 1 - Linear Regression

Shoe Sales:

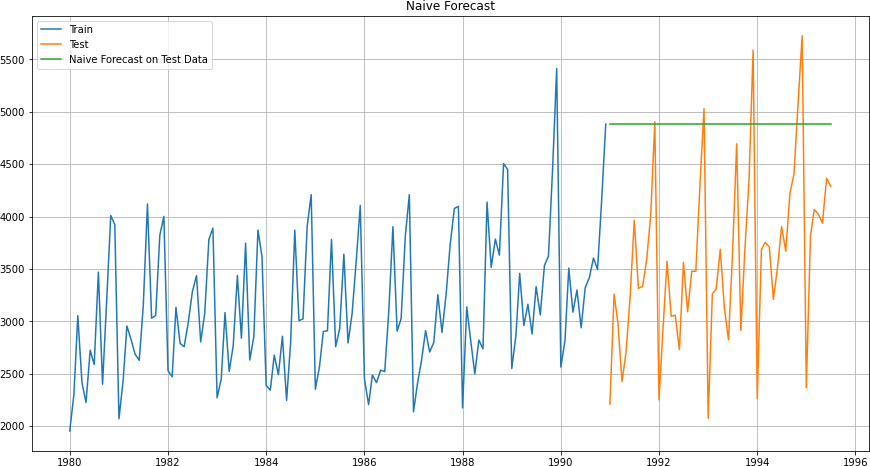
Soft Drink:



|  |  |  |
| --- | --- | --- |
|  | Test RMSE Shoe Sales | **Test RMSE Sparkling** |
| **RegressionOnTime** | 263.79 | 775.75 |

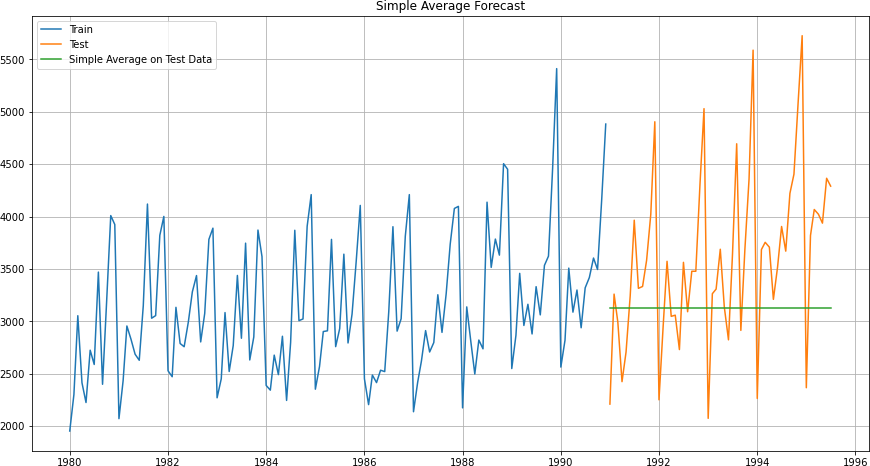
Model 2 - Naive Bayes: Shoe Sales:

Soft Drink:



|  |  |  |
| --- | --- | --- |
|  | Test RMSE Shoe Sales | **Test RMSE Soft Drink** |
| **RegressionOnTime** | 263.79 | 775.75 |
| **Naïve Model** | 245.12 | 1519.25 |

### C:\Users\Arun Sivaji\Downloads\download (97).pngModel 3 - Simple Average Shoe Sales:

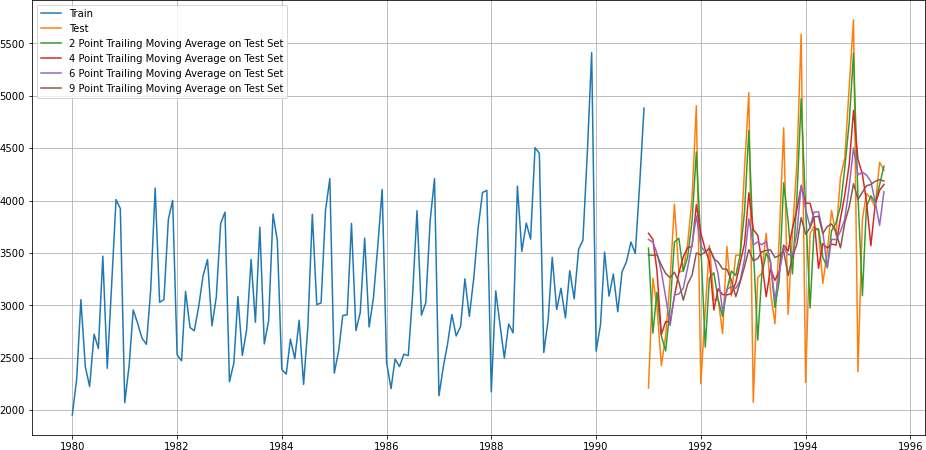
Soft Drink:

|  |  |  |
| --- | --- | --- |
|  | Test RMSE Shoe Sales | Test RMSE Soft Drink |
| **RegressionOnTime** | 263.79 | 775.75 |
| **NaiveModel** | 245.12 | 1519.25 |
| **SimpleAverageModel** | 63.98 | 934.35 |

### C:\Users\Arun Sivaji\Downloads\download (99).pngModel 4.A - Moving Average Shoe Sales:

|  |  |
| --- | --- |
|  | Test RMSE Shoe Sales |
| **2pointTrailingMovingAverage** | 45.94 |
| **4pointTrailingMovingAverage** | 57.87 |
| **6pointTrailingMovingAverage** | 63.45 |
| **9pointTrailingMovingAverage** | 67.72 |

Model 4.B - Moving Average Soft Drink:

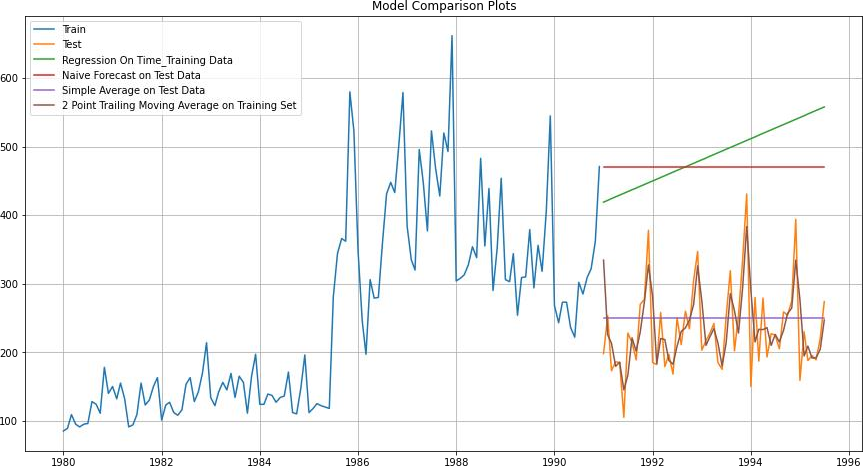


|  |  |
| --- | --- |
|  | Test RMSE Soft Drink |
| **2pointTrailingMovingAverage** | 556.72 |
| **4pointTrailingMovingAverage** | 687.18 |
| **6pointTrailingMovingAverage** | 710.51 |
| **9pointTrailingMovingAverage** | 735.88 |

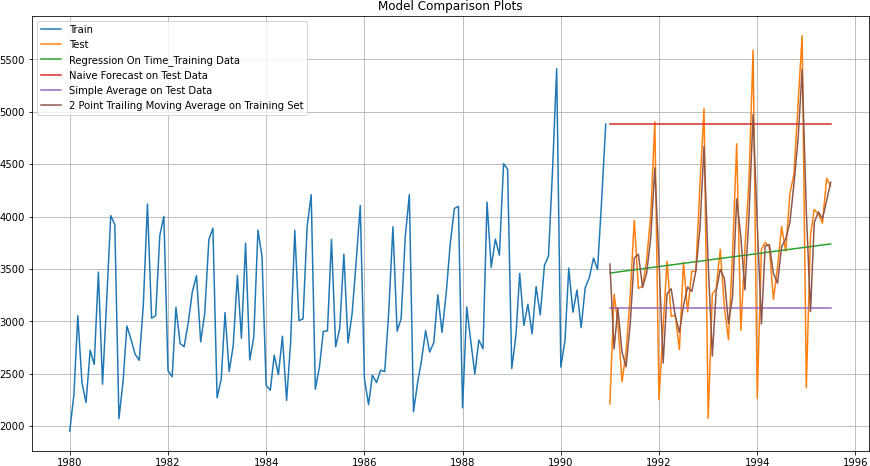
### NOTE -

* We have built 4 models till now for both datasets
* We fitted various models to the Train split and Tested it on Test split. Accuracymetrics used is Root Mean Squared Error (RMSE) on Test data

**Shoe Sales:**



**Soft Drink:**

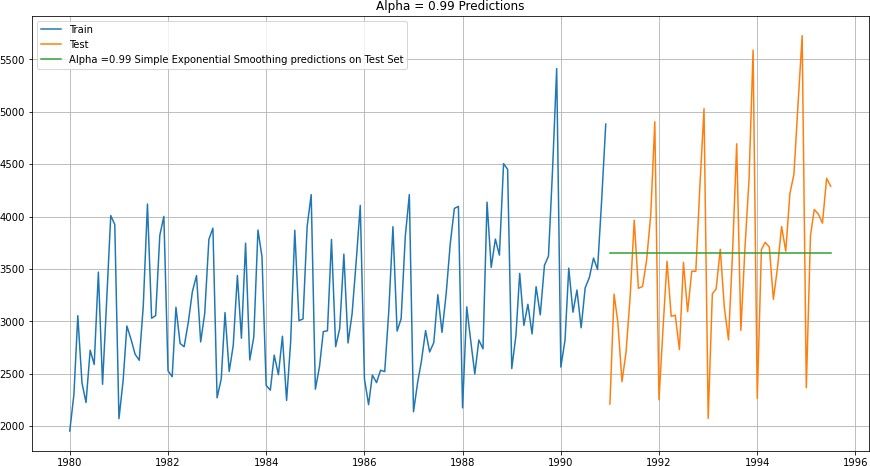


|  |  |  |
| --- | --- | --- |
|  | Test RMSE Shoe sales | **Test RMSE Soft Drink** |
| **Regression On Time** | 263.79 | 775.75 |
| **Naïve Model** | 245.12 | 1519.25 |
| **Simple Average Model** | 63.98 | 934.35 |
| **2pointTrailingMovingAverage** | **45.94** | **556.72** |
| **4pointTrailingMovingAverage** | 57.87 | 687.18 |
| **6pointTrailingMovingAverage** | 63.45 | 710.51 |
| **9pointTrailingMovingAverage** | 67.72 | 735.88 |

Consolidated Scores of Regression, Naive, Simple Average & Moving Average

* Till now, Best Model which gives lowest RMSE score for both ——> 2 Pt Moving Average Model
* We’ll continue to forecast using Exponential Smoothing Models for both datasets.
* Exponential smoothing averages or exponentially weighted moving averages consist of forecast based on previous periods data with exponentially declininginfluence on the older observations
* Exponential smoothing methods consist of special case exponential moving with notation ETS (Error, Trend, Seasonality) where each can be None(N), Additive (N),Additive damped (Ad), Multiplicative (M) or Multiplicative damped (Md)
* One or more parameters control how fast the weights decay. The values of theparameters lie between 0 and 1
* We’ll build following Exponential Smoothing Models -
  + Single Exponential Smoothing with Additive Errors - ETS(A, N, N)
  + Double Exponential Smoothing with Additive Errors, Additive Trends - ETS(A, A, N)
  + Triple Exponential Smoothing with Additive Errors, Additive Trends, Additive Seasonality - ETS(A, A, A)
  + Triple Exponential Smoothing with Additive Errors, Additive Trends, Multiplicative Seasonality - ETS(A, A, M)
  + Triple Exponential Smoothing with Additive Errors, Additive DAMPED Trends, Additive Seasonality - ETS(A, Ad, A)
  + Triple Exponential Smoothing with Additive Errors, Additive DAMPED Trends, Multiplicative Seasonality - ETS(A, Ad, M)

### C:\Users\Arun Sivaji\Downloads\download - 2022-06-20T190911.122.pngSingle Exponential Smoothing with Additive Errors - ETS(A, N, N) Shoe Sales:

Soft Drink:

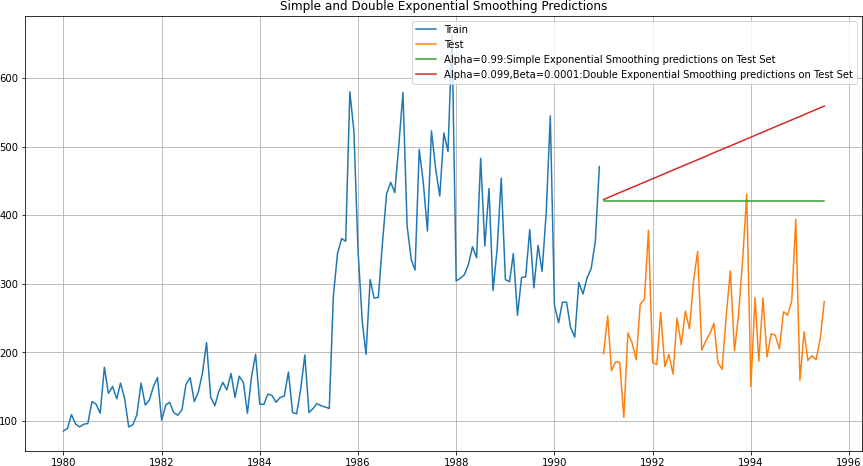
* + - For Shoe Sales - Level Parameter, Alpha = 0.99
    - For Soft Drink - Level Parameter, Alpha = 0.99

|  |  |  |
| --- | --- | --- |
|  | Test RMSE Rose | **Test RMSE Sparkling** |
| **RegressionOnTime** | 263.79 | 775.75 |
| **Naive Model** | 245.12 | 1519.25 |
| **Simple Average Model** | 63.98 | 934.35 |
| **2pointTrailingMovingAverage** | **45.94** | **556.72** |
| **4pointTrailingMovingAverage** | 57.87 | 687.18 |
| **6pointTrailingMovingAverage** | 63.45 | 710.51 |
| **9pointTrailingMovingAverage** | 67.72 | 735.88 |
| **Simple Exponential Smoothing** | 196.40 | 809.50 |

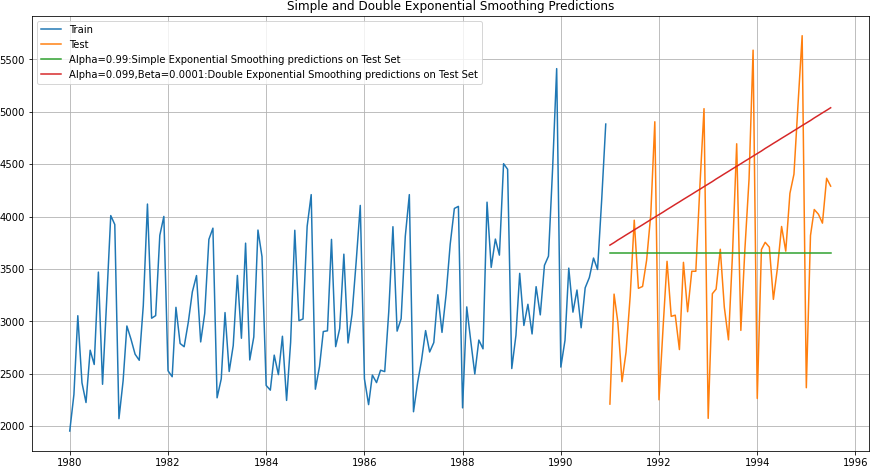
* Best Model till now for Both dataset ——- > 2 Pt Moving Average Model

### Double Exponential Smoothing with Additive Errors, Additive Trends - ETS(A, A, N)

Shoe Sales:



Soft Drink:



* In Shoe Sales - DES has no trend and seasonalitywell.
* In Soft Drink: DES has shown increasing trend and seasonality.
* Shoe Sales: Alpha = 0.099

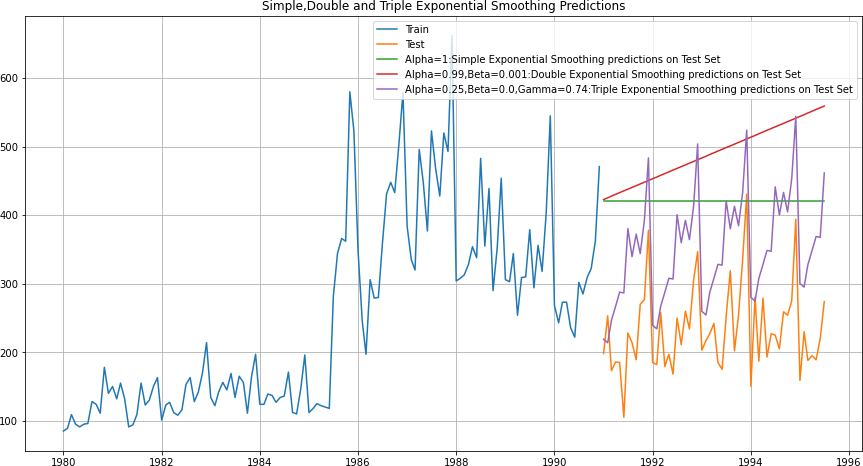
&Beta= 0.0001

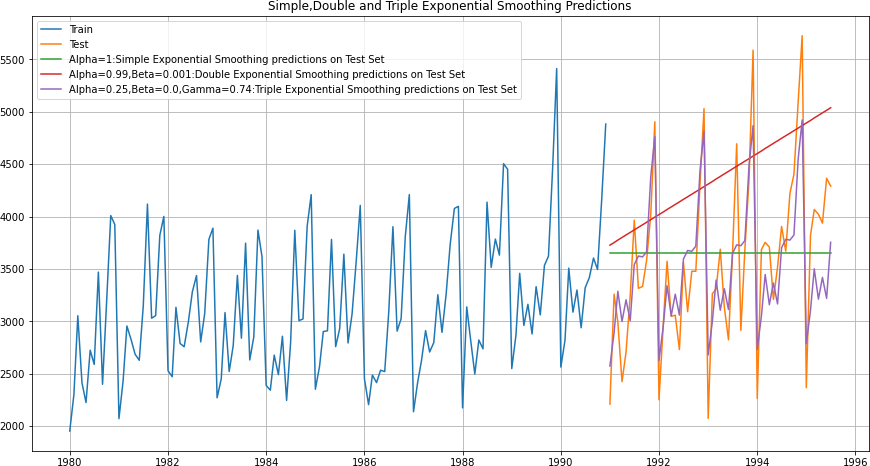
* Soft Drink - Level parameter, Alpha = 0.099 Trend parameter, Beta = 0.0001

|  |  |  |
| --- | --- | --- |
|  | Test RMSE Rose | **Test RMSE Sparkling** |
| **RegressionOnTime** | 263.79 | 775.75 |
| **Naive Model** | 245.12 | 1519.25 |
| **Simple Average Model** | 63.98 | 934.35 |
| **2pointTrailingMovingAverage** | **45.94** | **556.72** |
| **4pointTrailingMovingAverage** | 57.87 | 687.18 |
| **6pointTrailingMovingAverage** | 63.45 | 710.51 |
| **9pointTrailingMovingAverage** | 67.72 | 735.88 |
| **Simple Exponential Smoothing** | 196.40 | 809.50 |
| **Double Exponential Smoothing** | 266.16 | 1074.32 |

* + Best Model till now for both dataset——- > 2 Pt Moving Average Model

### Triple Exponential Smoothing with Additive Errors, Additive Trends, Additive Seasonality - ETS(A, A, A)

Shoe Sales:

Soft Drink:

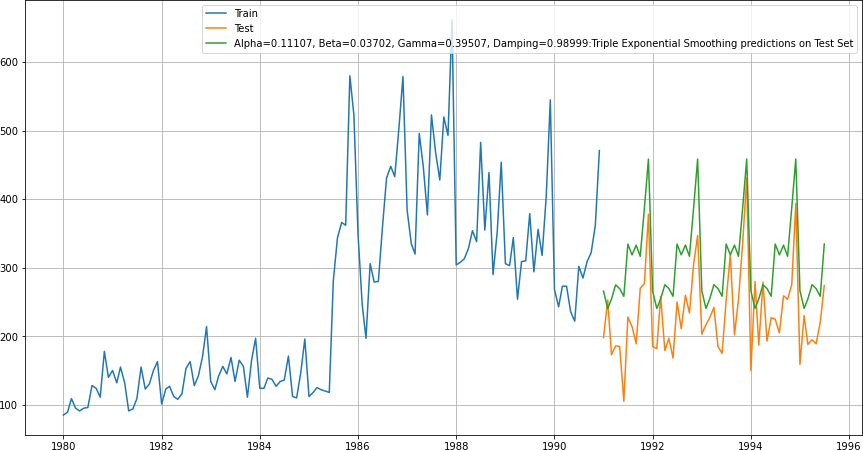
* + In both dataset - TES has picked up the trend very well
* Shoe sales-
* Level parameter, Alpha = 0.25
* Trend parameter, Beta = 0 Seasonality parameter, Gamma = 0.74
* Soft Drink:
* Level parameter, Alpha = 0.25
* Trend parameter, Beta = 0 Seasonality parameter, Gamma = 0.74

|  |  |  |
| --- | --- | --- |
|  | Test RMSE Rose | **Test RMSE Sparkling** |
| **RegressionOnTime** | 263.79 | 775.75 |
| **Naive Model** | 245.12 | 1519.25 |
| **Simple Average Model** | 63.98 | 934.35 |
| **2pointTrailingMovingAverage** | **45.94** | **556.72** |
| **4pointTrailingMovingAverage** | 57.87 | 687.18 |
| **6pointTrailingMovingAverage** | 63.45 | 710.51 |
| **9pointTrailingMovingAverage** | 67.72 | 735.88 |
| **Simple Exponential Smoothing** | 196.40 | 809.50 |
| **Double Exponential Smoothing** | 266.16 | 1074.32 |
| **Triple Exponential Smoothing** | 128.99 | **458.96** |

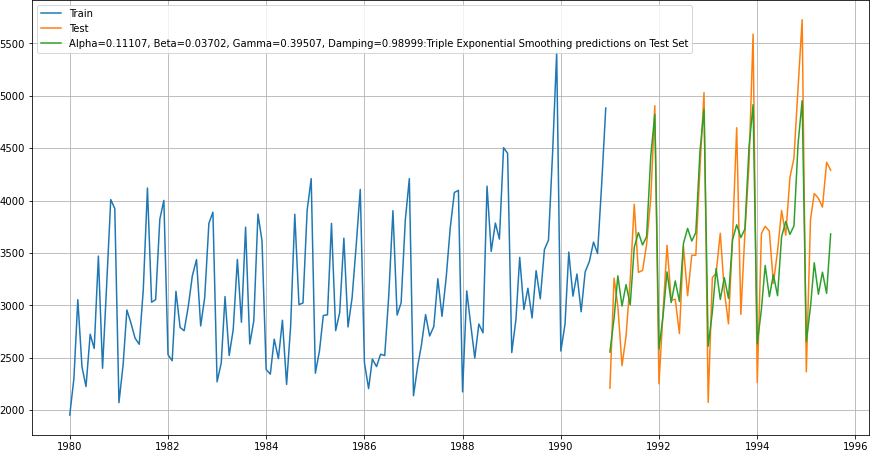
* + Till now, Best Model for Shoe sales ——> 2 Pt Moving Average Best Model for Soft drink ——> Holt-Winter - ETS (A, A, A)

### Triple Exponential Smoothing with Additive Errors, Additive Trends, Multiplicative Seasonality (Damped) - ETS(A, A, M)

Shoe Sales:



Soft Drink:



* **Shoe Sales-**

Level parameter, Alpha = 0.11107 Trend parameter, Beta = 0.03702

Seasonality parameter, Gamma = 0.39507

* Soft Drink

Level parameter, Alpha = 0.11107 Trend parameter, Beta = 0.03702

Seasonality parameter, Gamma = 0.39507

|  |  |  |
| --- | --- | --- |
|  | Test RMSE Rose | **Test RMSE Sparkling** |
| **Regression On Time** | 263.79 | 775.75 |
| **Naive Model** | 245.12 | 1519.25 |
| **Simple Average Model** | 63.98 | 934.35 |
| **2pointTrailingMovingAverage** | **45.94** | 556.72 |
| **4pointTrailingMovingAverage** | 57.87 | 687.18 |
| **6pointTrailingMovingAverage** | 63.45 | 710.51 |
| **9pointTrailingMovingAverage** | 67.72 | 735.88 |
| **Simple Exponential Smoothing** | 196.40 | 809.50 |
| **Double Exponential Smoothing** | 266.16 | 1074.32 |
| **Triple Exponential Smoothing** | 128.99 | **458.96** |
| **Triple Exponential Smoothing (Multiplicative Season, Damped)** | 80.92 | 477.00 |

We conclude that models with least RMSE,

**Best Model for Shoe Sales——> 2 Pt Moving Average Best Model for Soft Drink ——> Holt-Winter - ETS (A, A, A)**

# [Q 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.

### To Check Stationarity of Data -

* + We use Augmented Dicky - Fuller (ADF) Test to check the Stationarity of Data
  + Hypotheses of ADF Test :

*H*0 *Time Series is not stationary*

*Ha Time Series is Stationary*

* + So for Industry standard (also given for this problem), the Confidence Interval is95%
  + Hence, alpha = 0.05
  + So in ADF Test, if p-value < alpha ===> We reject the Null Hypothesis and hence

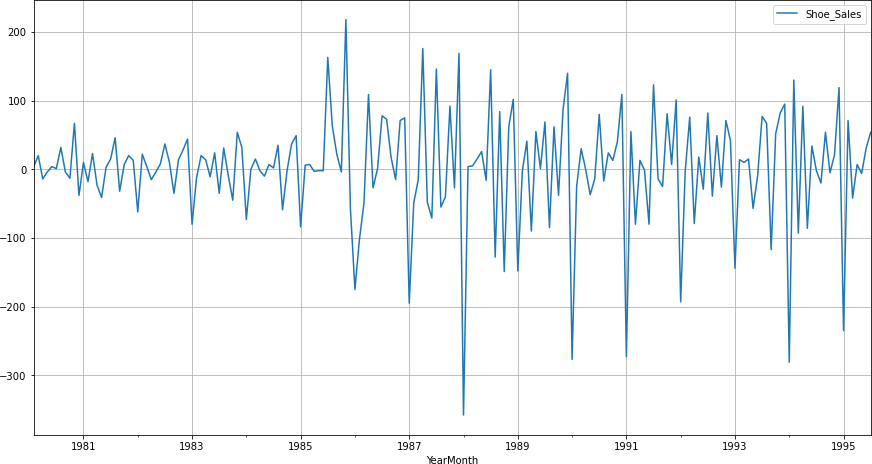
Conclude that given Time Series isStationary

* + So in ADF Test, if p-value > alpha ===> We fail to reject the Null Hypothesis and

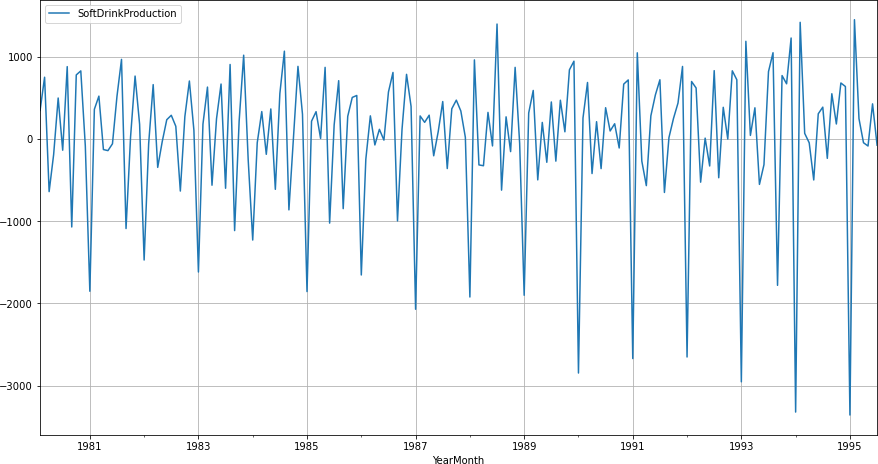
Hence conclude that given Time Series isNot Stationary

* + If Time Series is not stationary then we apply one level of diﬀerencing and checkfor Stationarity again.
  + Again, if the Time Series is still not Stationary, we apply one more level ofdiﬀerencing and check for Stationarity again
  + Once the Time Series is Stationary then we are ready to apply ARIMA / SARIMAmodels

### Stationarity of Shoe Sales Dataset -

* + Augmented Dicky-Fuller Test was applied to the whole Rose dataset
  + We found, p-value = 0.80
  + Here, p-value > alpha=0.05
  + We fail to reject the Null Hypothesis and hence conclude that Rose Wine TimeSeries is Not Stationary
  + We take 1 level of diﬀerencing and check again for Stationarity
  + Now, p-value = 0.036
  + Now, p-value < alpha=0.05
  + Now, we reject the Null Hypothesis and conclude that Rose Time Series isStationary with a lag of 1

### Stationarity of Soft Drink Dataset -

* + Augmented Dicky-Fuller Test was applied to the whole Sparkling dataset
  + We found, p-value = 0.98
  + Here, p-value > alpha=0.05
  + We fail to reject the Null Hypothesis and hence conclude that Sparkling WineTime Series is Not Stationary
  + We take 1 level of diﬀerencing and check again for Stationarity
  + Now, p-value = 3.05x10^-14
  + Now, p-value < alpha=0.05
  + Now, we reject the Null Hypothesis and conclude that Sparkling Time Series isStationary with a lag of 1

# [ Q 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 / SARIMA Models -

* + ARIMA is an acronym for Auto-Regressive Integrated Moving Average
  + SARIMA stands for Seasonal ARIMA, when the TS has seasonality.

ARIMA / SARIMA are forecasting models on Stationary Time Series

### ARIMA / SARIMA Modelling on Train Shoe sales and Soft Drink Data -

* + We check for stationarity of Train both data by using AugmentedDicky Fuller Test
  + We take a diﬀerence of 1 and make both these datasets Stationary
  + We apply the following iterations to both these datasets -

1. ARIMA Automated
2. SARIMA Automated

### ARIMA Automated -

* + We create a grid of all possible combinations of (p, d, q)
  + Range of p = Range of q = 0 to 3, Constant d = 1
  + Few Examples of the grid -

Model: (0, 1, 2)

Model: (0, 1, 3)

Model: (1, 1, 0)

Model: (1, 1, 1)

Model: (1, 1, 2)

Model: (1, 1, 3)

Model: (2, 1, 0)

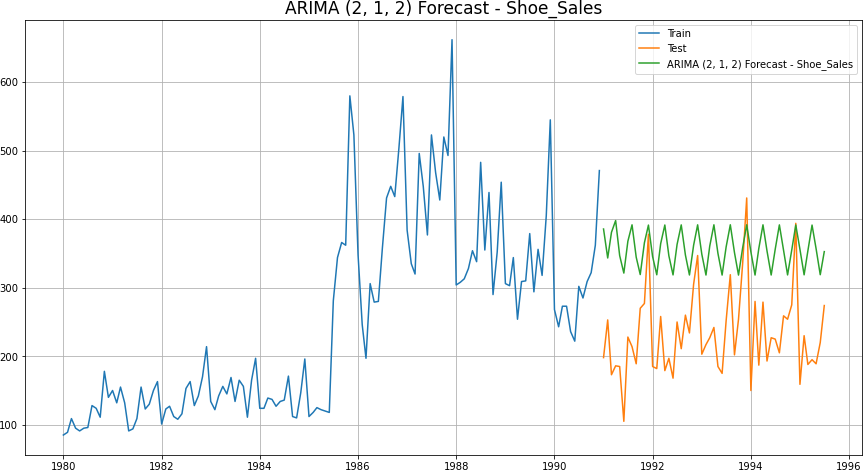
Model: (2, 1, 1)

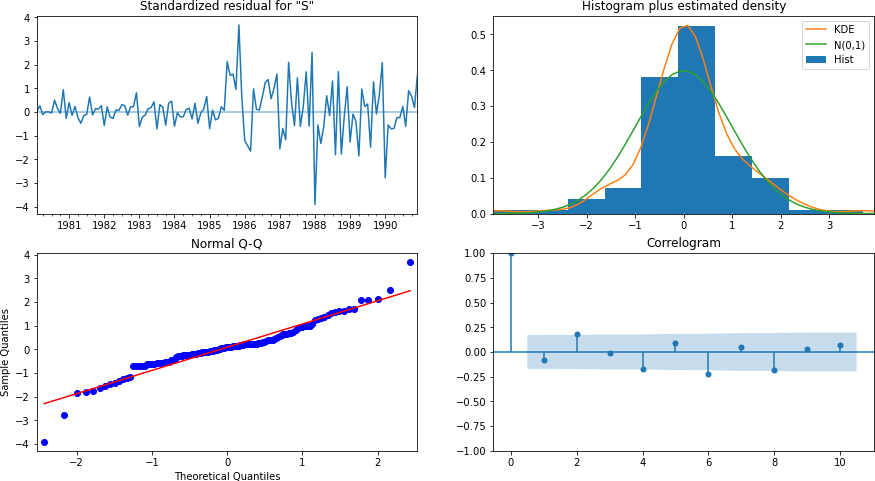
Model: (2, 1, 2)

Model: (2, 1, 3)

Model: (3, 1, 0)

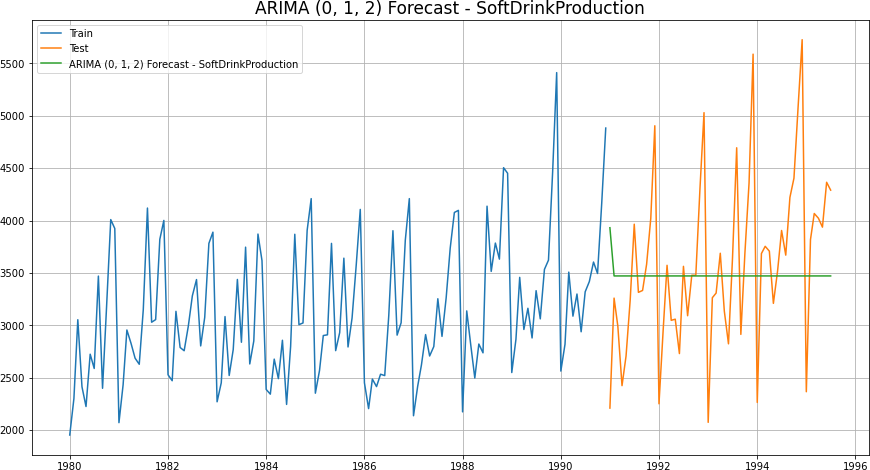
Model: (3, 1, 1)

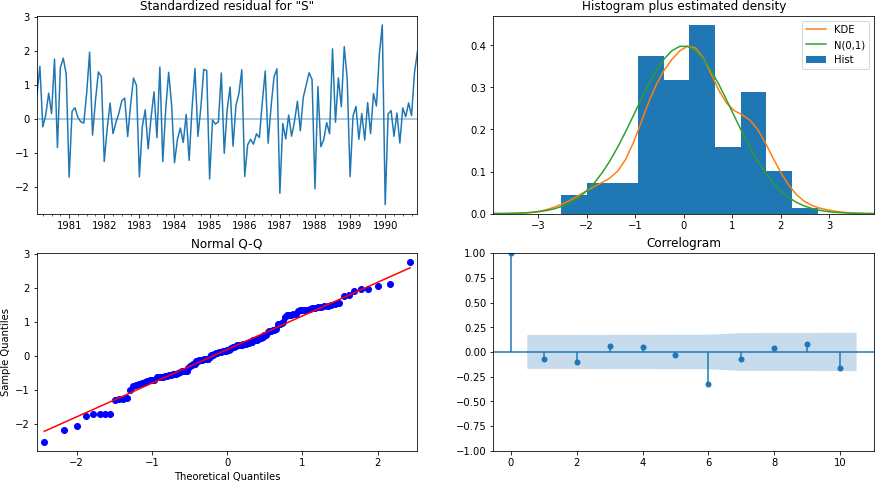
* + We fit ARIMA models to each of these combinations for both datasets
  + We choose the combination with the least Akaike Information Criteria (AIC)
  + We fit ARIMA to this combination of (p, d, q) to the Train set and forecast on the Test set
  + Finally, we check the accuracy of this model by checking RMSE of Test set
  + For Rose, Best Combination with Least AIC is - (p, d, q) —-> (2, 1, 2)

ARIMA (2, 1, 2) Diagnostic Plot

|  |  |  |
| --- | --- | --- |
|  | Test RMSE Shoe Sales | **Test MAPE Shoe sales** |
| **ARIMA(2,1,2)** | 135.80 | 61.98 |

* + For Sparkling, Best Combination with Least AIC is - (p, d, q) —-> (0, 1, 2)

•

ARIMA (0, 1, 2) Diagnostic Plot

|  |  |  |
| --- | --- | --- |
|  | Test RMSE Soft Drink | **Test MAPE** Soft Drink |
| **ARIMA(0,1,2)** | 831..6 | 18.49 |

### SARIMA Automated -

* + We create a grid of all possible combinations of (p, d, q) along with Seasonal (P, D, Q) & Seasonality of 12 (for both datasets)
  + Range of p = Range of q = 0 to 3, Constant d = 1
  + Range of Seasonal P = Range of Seasonal Q = 0 to 3, Constant D = 1, Seasonality m = 12
  + Few Examples of the grid (p, d, q) (P, D, Q, m) -

Model: (0, 1, 2) (0, 0, 2, 12)

Model: (0, 1, 3) (0, 0, 3, 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: (1, 1, 3) (1, 0, 3, 12)

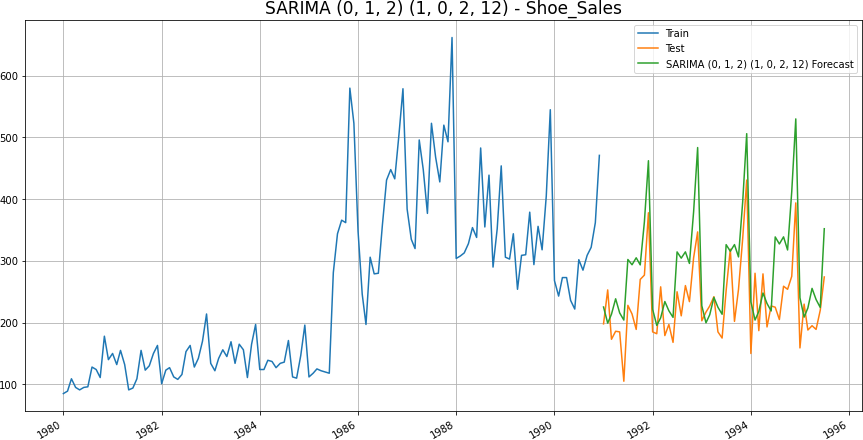
Model: (2, 1, 0) (2, 0, 0, 12)

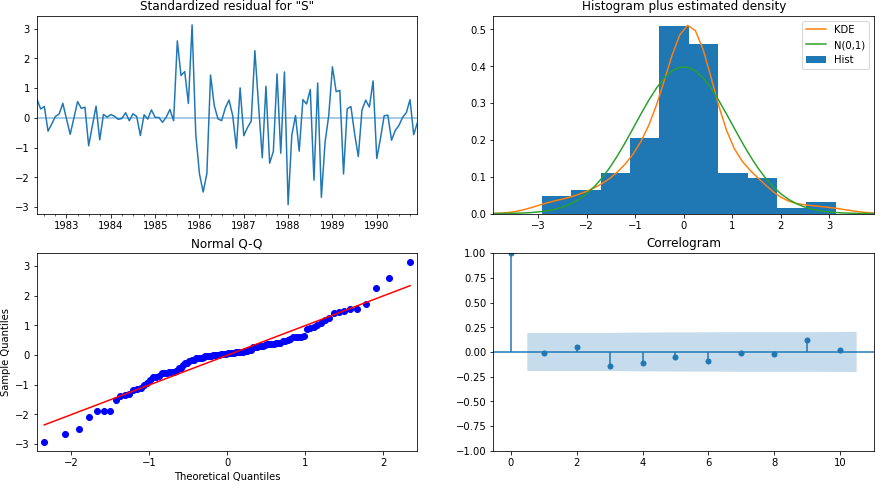
Model: (2, 1, 1) (2, 0, 1, 12)

Model: (2, 1, 2) (2, 0, 2, 12)

Model: (2, 1, 3) (2, 0, 3, 12)

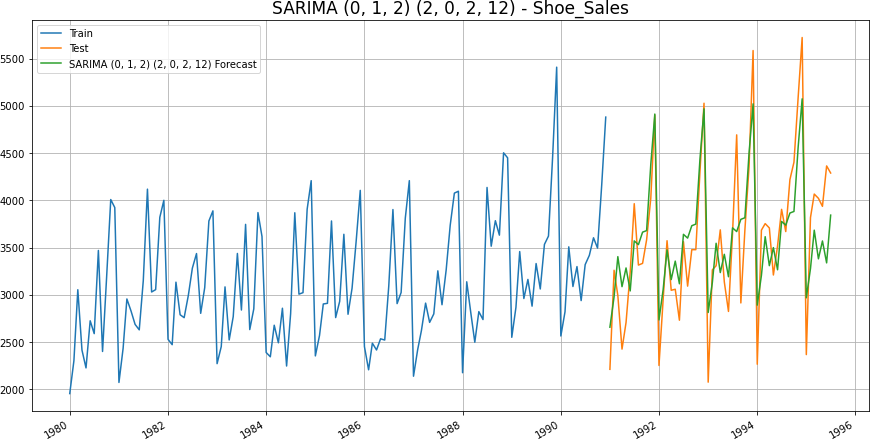
Model: (3, 1, 0) (3, 0, 0, 12)

* + We fit SARIMA models to each of these combinations and select with least AIC
  + We fit SARIMA to this best combination of (p, d, q) (P, D, Q, m) to the Train set and forecast on the Test set. Then, we check accuracy using RMSE on Test set
  + Best Combination with Least AIC is - (0, 1, 2) (1, 0, 2, 12)

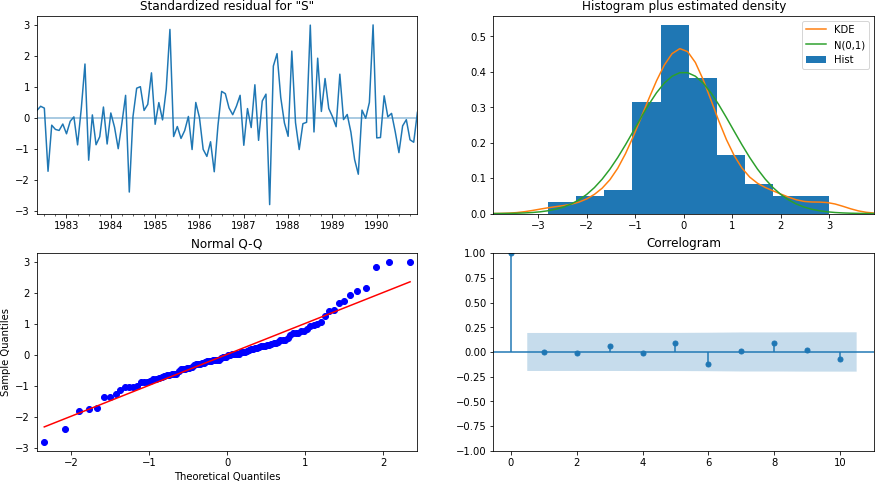


|  |  |  |
| --- | --- | --- |
|  | Test RMSE Shoe Sales | **Test MAPE** Shoe Sales |
| **ARIMA(2,1,2)** | 135.80 | 61.98 |
| **SARIMA (0, 1, 2) (1, 0, 2, 12)** | 69.03 | 26.45 |

* + Best Combination with low AIC and low Test RMSE is - (0, 1, 2) (2, 0, 2, 12)



SARIMA (0, 1, 2) (2, 0, 2, 12) Diagnostic Plot



|  |  |  |
| --- | --- | --- |
|  | Test RMSE Soft Drink | **Test MAPE** Soft Drink |
| **ARIMA(0,1,2)** | 831.16 | 18.49 |
| **SARIMA (0, 1, 2) (2, 0, 2, 12)** | 437.70 | 11.04 |

Best Model for Shoe Sales with Least RMSE: **SARIMA (0, 1, 2) (1, 0, 2, 12)**

Best Model for Soft Drink with Least RMSE: **SARIMA (0, 1, 2) (2, 0, 2, 12)**

# [Q 7] Build ARIMA/SARIMA models based on the cut-off points of ACF and PACF on the training data and evaluate this model on the test data using RMSE.

### Auto-Correlation Function (ACF) -

* + Autocorrelation refers to how correlated a time series is with its past values. e.g.

*yt* with *yt*−1 also *yt*+1 with *yt* and so on.

* + ‘Auto’ part of Autocorrelation refers to Correlation of any time instance with its previous time instance in the SAME Time Series
  + ACF is the plot used to see the correlation between the points, up to and including the lag unit
  + ACF indicates the value of ‘q’ - which is the Moving Average parameter in ARIMA / SARIMA models

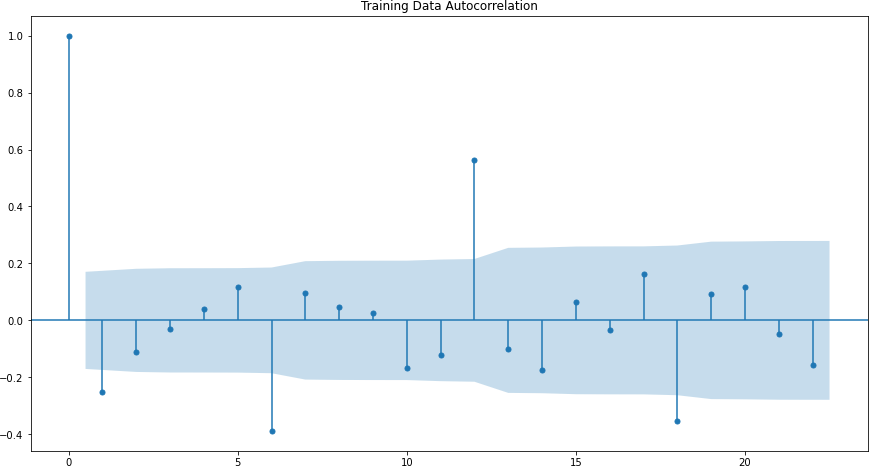
### Partial Auto-Correlation Function (PACF) -

* + Partial Autocorrelation refers to how correlated a time series is with its past lag values.
  + For example, let lag=k, then Partial Autocorrelation is Correlation of *yt* with *yt*−*k*, ignoring the eﬀects of all the instances between *yt* and *yt*−*k*
  + PACF is the plot used to see the correlation between the lag points
  + PACF indicates the value of ‘p’ - which is the Auto-Regressive parameter in ARIMA / SARIMA models

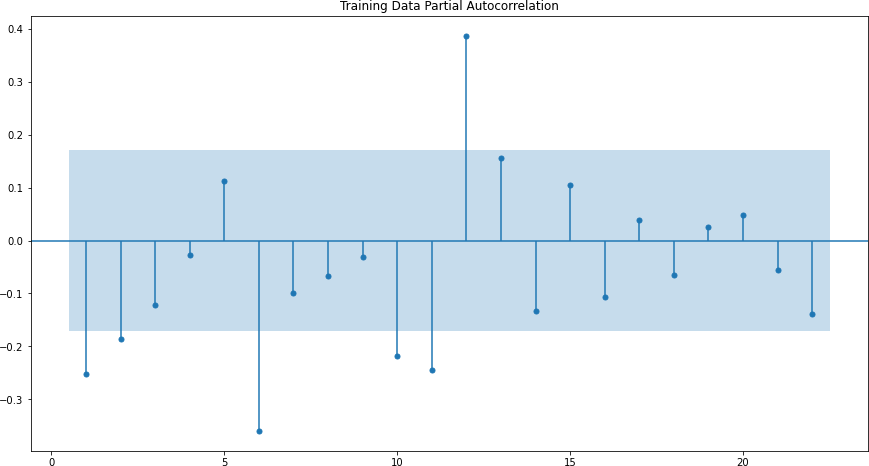
### ACF & PACF of Shoe Sales -

* + Observing the cutoﬀs in ACF and PACF plots for Shoe Sales dataset, - FOR ARIMA —-> p = 2, q = 2 and diﬀerence d = 1

FOR SARIMA —-> p = 2, q = 2, d = 1 and P = 2, D = 1, Q = 2, Seasonality=12

**ACF:-**

PACF:-

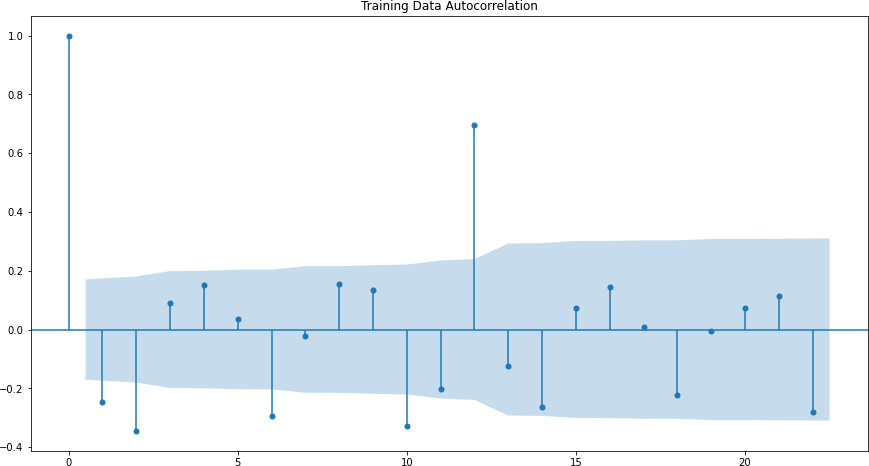


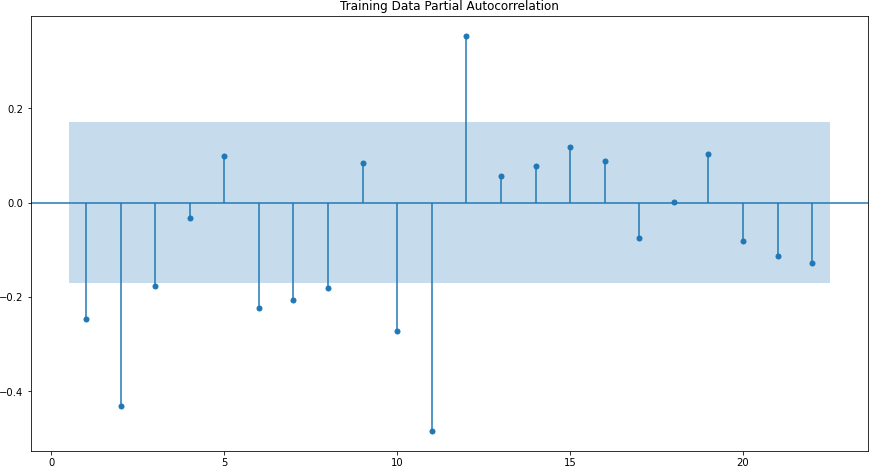
### ACF & PACF of Soft Drink -

* + Observing the cutoﬀs in ACF and PACF plots for Sparkling dataset, we get - FOR ARIMA —-> p = 3, q = 3 and diﬀerence d = 1

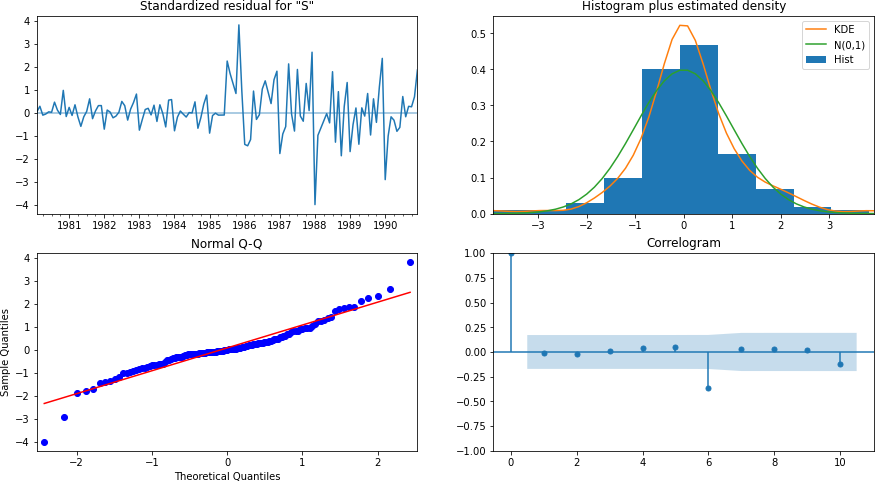
ACF:

FOR SARIMA —-> p = 3, q = 3, d = 1 and P = 0, 1, 2, 3 | D = 0, Q = 1, 2, 3



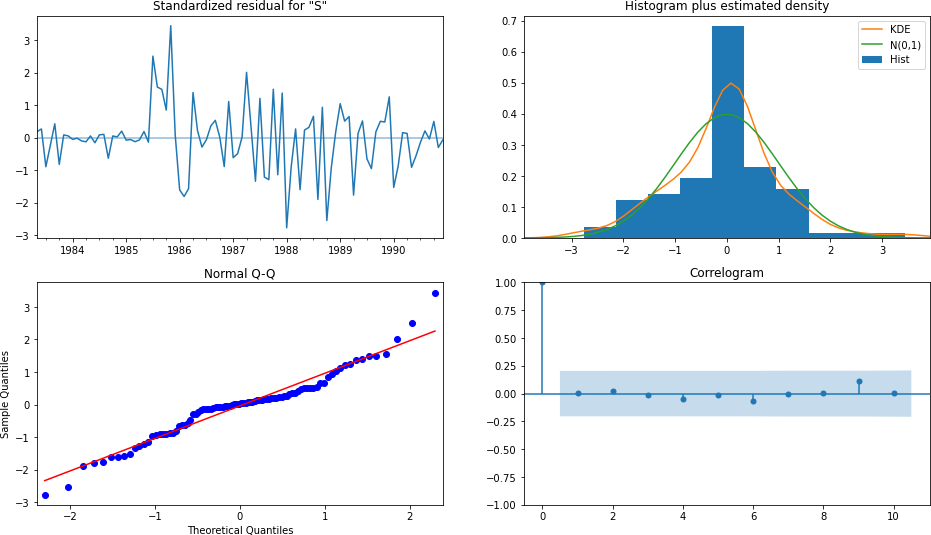
PACF:

### C:\Users\Arun Sivaji\Downloads\download - 2022-06-20T204550.980.pngARIMA Manual –Shoe Sales - (2, 1, 2)

**ARIMA (2, 1, 2) Diagnostic Plot**

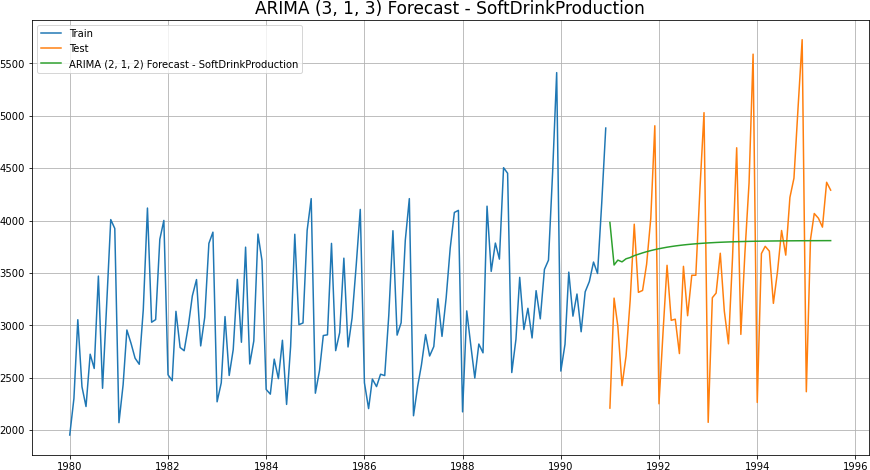
|  |  |  |
| --- | --- | --- |
|  | Test RMSE Shoe Sales | **Test MAPE Shoe Sales** |
| **ARIMA( 2, 1, 2)** | 143.5 | 66.6 |

### C:\Users\Arun Sivaji\Downloads\download - 2022-06-20T210739.366.pngSARIMA Manual – Shoe Sales- (2, 1, 2) (2, 1, 2, 12)

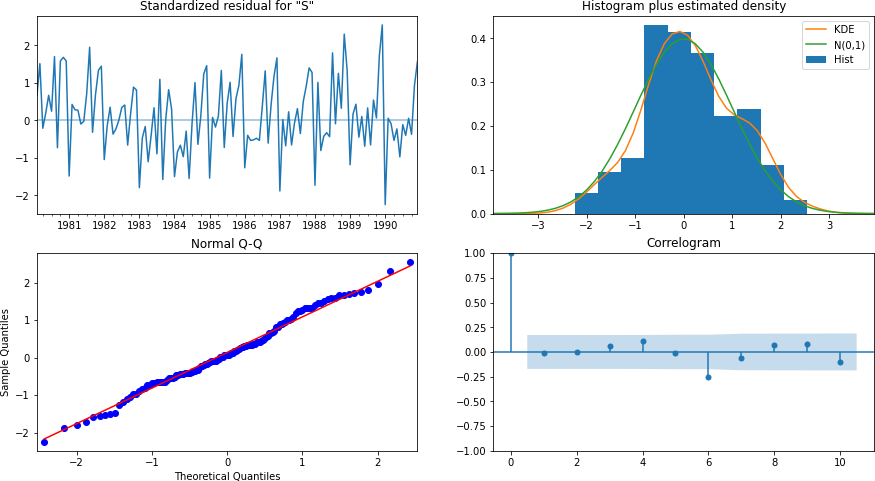


|  |  |  |
| --- | --- | --- |
|  | Test RMSE Shoe sales | **Test MAPE Shoe Sales** |
| **ARIMA( 2, 1, 2)** | 143.5 | 66.6 |
| **SARIMA (2, 1, 2) (2, 1, 2, 12)** | 68.48 | 28.45 |

ARIMA Manual – Soft Drink - (3, 1, 3)

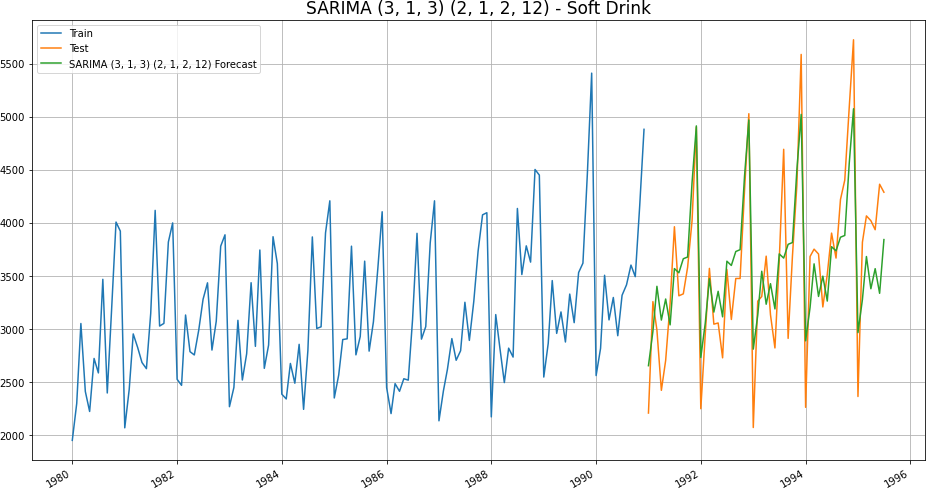


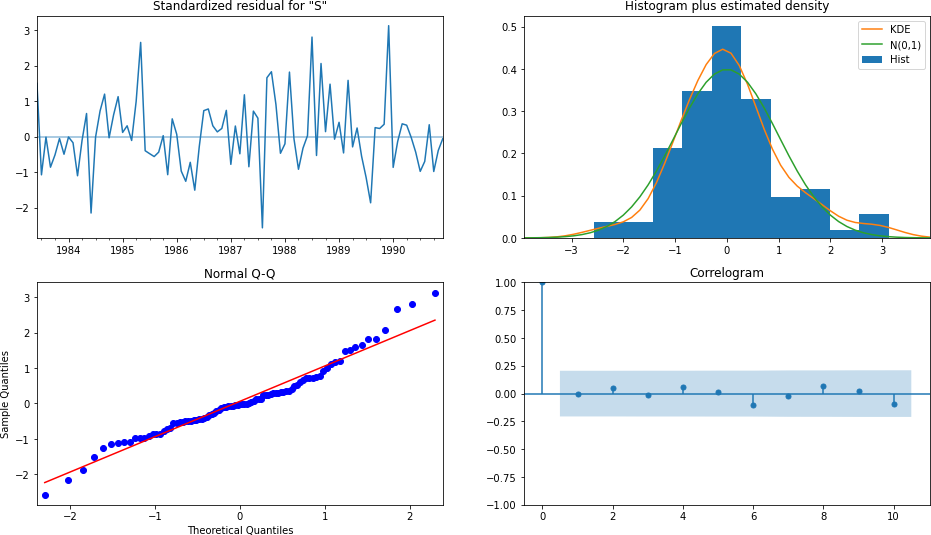
**ARIMA (3, 1, 3) Diagnostic Plot – Soft Drink**



|  |  |  |
| --- | --- | --- |
|  | Test RMSE Sparkling | **Test MAPE Sparkling** |
| **ARIMA( 3, 1, 3)** | 816.7 | 19.9 |

### SARIMA Manual - SoftDrink - (3, 1, 3) (2, 1, 2, 12)



**SARIMA (3, 1, 3) (2, 1, 2, 12) Diagnostic Plot -**

|  |  |  |
| --- | --- | --- |
|  | Test RMSE Sparkling | **Test MAPE Sparkling** |
| **ARIMA( 3, 1, 3)** | 816.7 | 19.9 |
| **SARIMA (3, 1,3) (2, 1, 2, 12)** | 431.8 | 11.04 |

* + In all Manual methods, Best Model for Shoe Sales with Least RMSE

**—-> SARIMA (2, 1, 2) (2, 1, 2, 12)**

* + In all Manual methods, Best Model for Soft Drink with Least RMSE

—-> **SARIMA (3, 1, 3) (2, 1, 2, 12)**

# [Q 8] Build a table (create a data frame) with all the models built along with their corresponding parameters and the respective RMSE values on the test data.

All Models built with Shoe Sales and Soft Drinks (sorted by RMSE) –

|  |  |  |
| --- | --- | --- |
|  | Test RMSE Rose | **Test RMSE Sparkling** |
| **Regression On Time** | 263.79 | 775.75 |
| **Naive Model** | 245.12 | 1519.25 |
| **Simple Average Model** | 63.98 | 934.35 |
| **2pointTrailingMovingAverage** | **45.94** | 556.72 |
| **4pointTrailingMovingAverage** | 57.87 | 687.18 |
| **6pointTrailingMovingAverage** | 63.45 | 710.51 |
| **9pointTrailingMovingAverage** | 67.72 | 735.88 |
| **Simple Exponential Smoothing** | 196.40 | 809.50 |
| **Double Exponential Smoothing** | 266.16 | 1074.32 |
| **Triple Exponential Smoothing** | 128.99 | **458.96** |
| **Triple Exponential Smoothing (Multiplicative Season, Damped)** | 80.92 | 477.00 |
| **ARIMA( 2, 1, 2)- Automated** | 143.5 | 66.6 |
| **SARIMA (2, 1, 2) (2, 1, 2, 12) -**  **Automated** | 68.48 | 28.45 |
| **ARIMA( 3, 1, 3) - Manual** | 816.7 | 19.9 |
| **SARIMA (3, 1,3) (2, 1, 2, 12) -**  **Manual** | 431.8 | **11.04** |

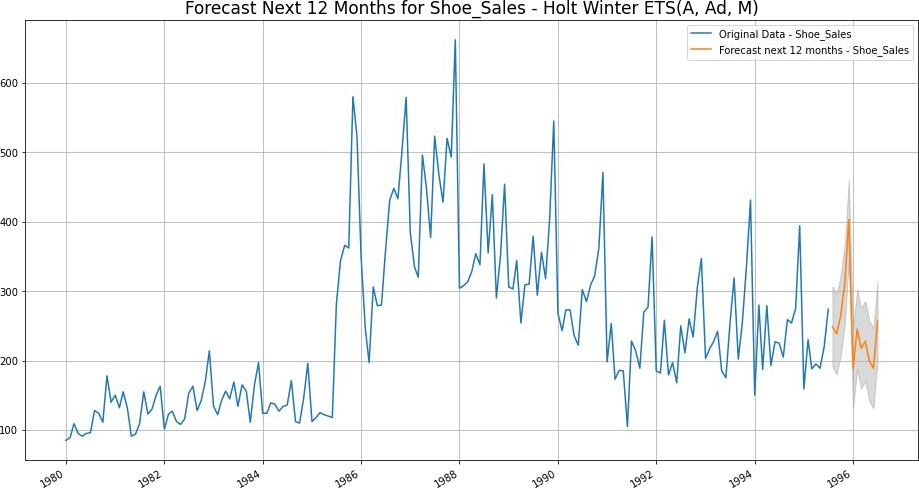
# [Q 9]Based on the model-building exercise, build the most optimum model(s) on the complete data and predict 12 months into the future with appropriate confidence intervals/bands.

Best Models as per the Least RMSE on Shoe Sales Test set ——>

- 2 Pt Trailing Moving Average Best Model as per the Least RMSE on Soft Drink Test set ——>

Triple Exponential Smoothing (Multiplicative Season, Damped Trend)

Forecasting on the best model - Triple Exponential SmoothingETS(A, A, A) – Multiplicative Seasonal.



### Sparkling Forecast Next 12 months - Triple Exponential Smoothing ETS (A, Ad, M) - Damped Trend, Multiplicative Seasonality

