Time Series Analysis Business Report

DSBA

‘SPARKLING’

Vikash Kumar

Aug’22 Batch

Contents

Table Content

Figures Content

Problem Statement:

For this particular assignment, the data of different types of wine sales in the 20th century is to be analysed. Both of these data are from the same company but of different wines. As an analyst in the ABC Estate Wines, you are tasked to analyse and forecast Wine Sales in the 20th century.

Data set for the Problem: Sparkling.csv and Rose.csv

Time Series Analysis & Forecasting of Sparkling Wine Sales

QN1 Read the data as an appropriate Time Series data and plot the data.¶

- To enable us to address the business issue, we have imported necessary libraries:
  - Basic - from Numpy and Pandas
  - Data visualization - Matplotlib, Seaborn,
  - General – Python standard warnings
  - Statistics - scipy and scikit learn for time series modelling and plotting time series specific plots
- The dataset in csv format 'Sparkling' was loaded in python as time series data and we have verified that the same is properly loaded. The Header and Tail of dataset is verified for completeness.

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

TABLE 1 - HEADER OF 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
```

Shape of Sparkling a TS Data : (187 rows and 1 variable column)
While Year Month has now as Index feature for TS dataset
No null value appears in this table.

TABLE 2: SHAPE & INFORMATION OF THE DATASET

- The shape of the dataset and the information of number of columns, number of records as well as data type, was verified

TABLE 3: SHAPE & INFORMATION OF THE DATASET

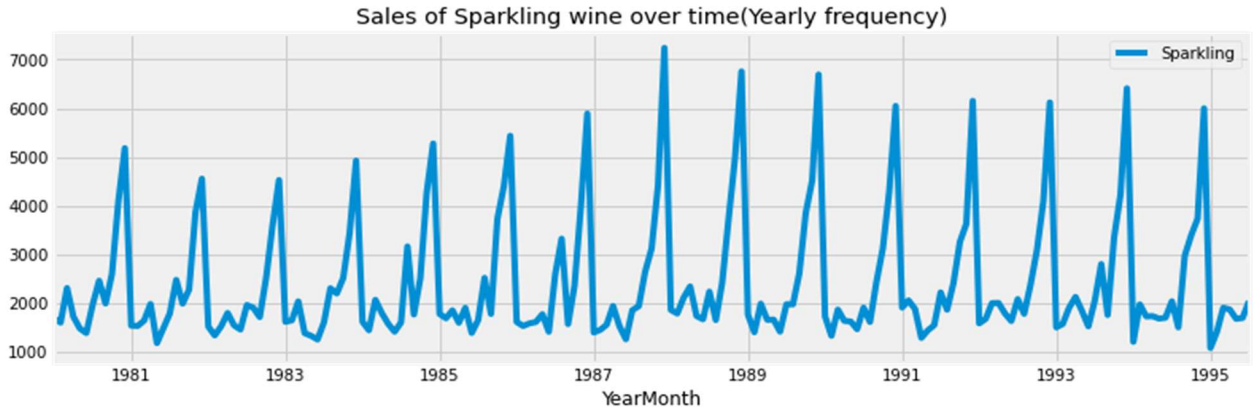


FIGURE 1 YEARLY SALES PATTERN

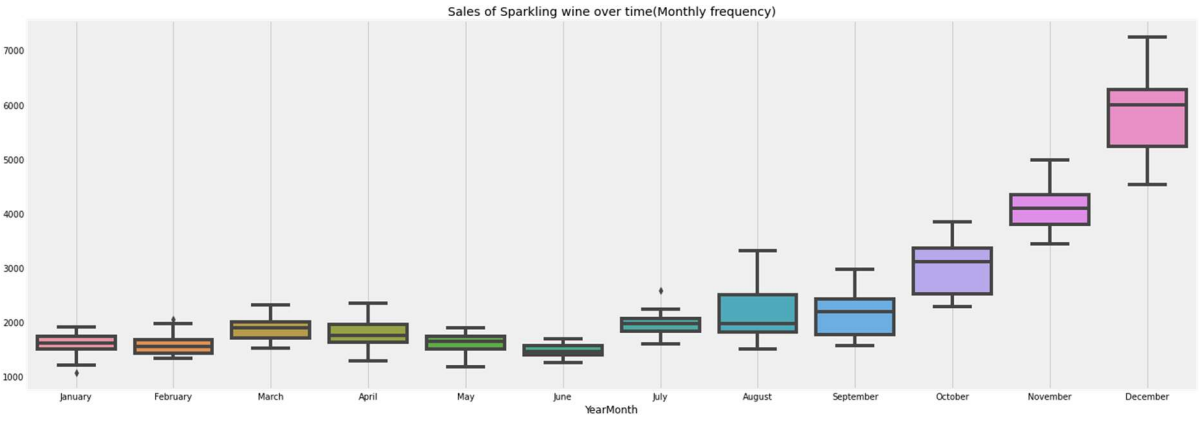


FIGURE 2 MONTHLY SALES PATTERN

Plot Insights:

1.In Yearly Box plot there is not fix type of trend can be observed while in Monthly plot the Seasonality can be seen clearly.
2.There in sharp increase in sale from month of August every year which touches the peak in December.
3.Where as a normal jump is sale also can be observed in month of March and April.
4.Month June can be considered as month with very low sale within year of seasonality.
5.As per sales perspective the minimum sales was in year 1983 was less than 2000 while the highest sales are recorded in year 1988

QN2 Perform appropriate Exploratory Data Analysis to understand the data and also perform decomposition.

- The descriptive statistics showing 5 point summary is verified

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

TABLE 4 DATA DESCRIPTION

Insights:

1.There is difference in mean and median hence data seems as right skewed.

2.There is large space between minimum sales value and maximum sales values.

- Checking The Null Values in TS Dataset

```
Sparkling    0
dtype: int64
```

Insight:

There is NO any missing value in Sparkling time series dataset

- Decomposition of Wine Sales

These are the components of a time series

Trend - Consistent upwards or downwards slope of a time series
Seasonality - Clear periodic pattern of a time series (like sine function)
Noise/Error - Outliers or missing values White noise has...

Constant mean Constant variance Zero auto-correlation at all lags

Multiplicative Decomposition

1.A Multiplicative model suggests that the components are multiplied together.
2.A Multiplicative model is non-linear such as quadratic or exponential.
3.Changes increase or decrease over time.
4.A non-linear seasonality has an increasing or decreasing frequency (width of the cycles) and / or amplitude (height of the cycles) over time.

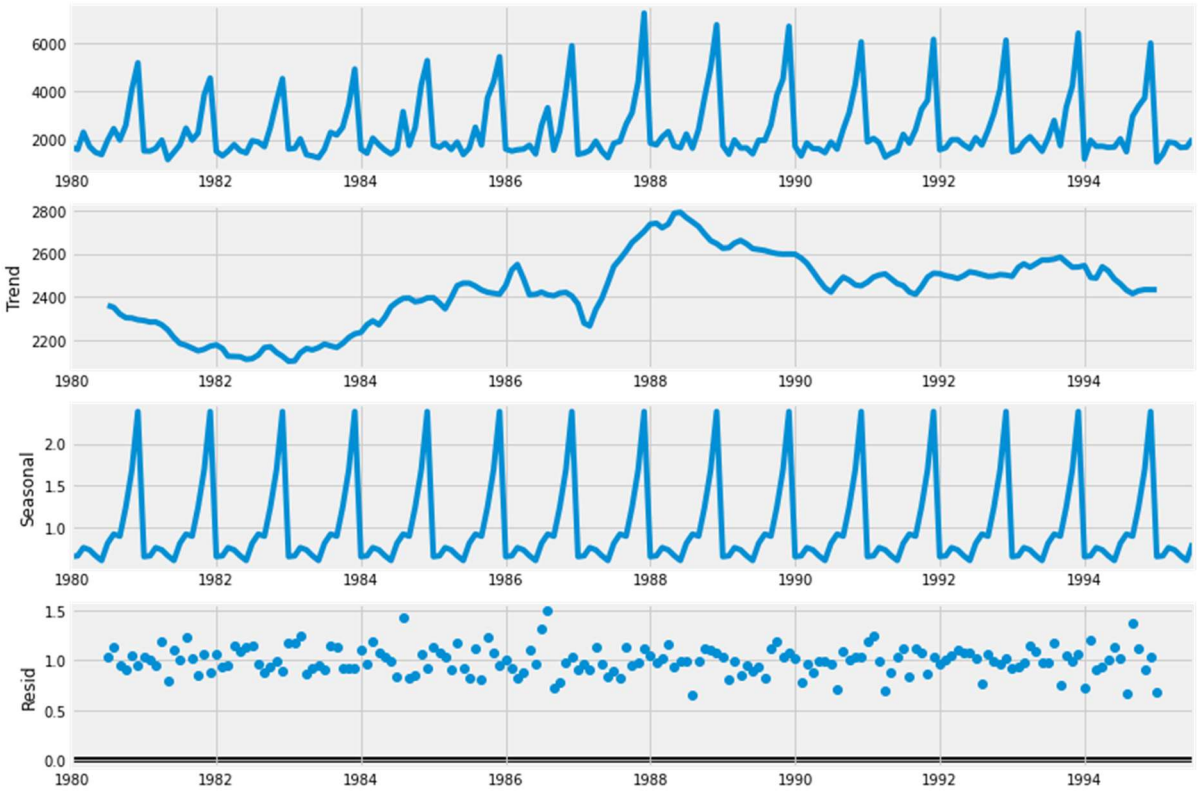


FIGURE 3 SEASONAL DECOMPOSITION MULTIPLICATIVE

Additive Decomposition

An additive model suggests that the components are added together.
An additive model is linear where changes over time are consistently made by the same amount.
A linear seasonality has the same frequency (width of the cycles) and amplitude (height of the cycles).

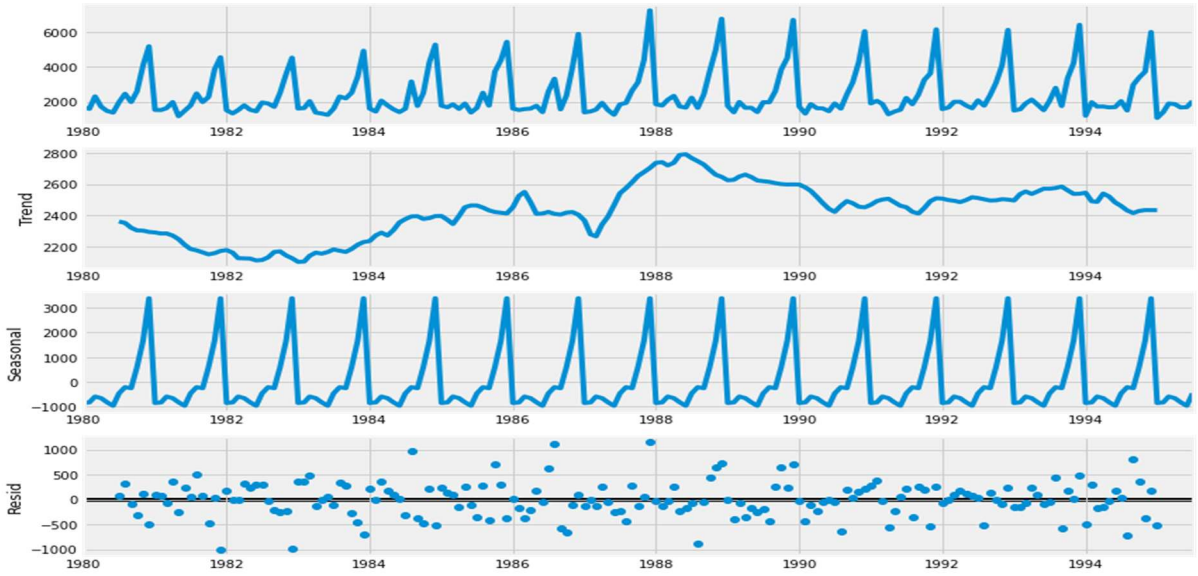


FIGURE 4 SEASONAL DECOMPOSITION ADDITIVE

Insights:

Some of our key observations from Decomposition plots:

Running the above code performs the decomposition, and plots the 4 resulting series.
1.We observe that the trend and seasonality are clearly separated.
2.Trend: The year wise data in plot don't follow the increasing or decreasing trend. But Tell the story of Growth and degrowth of Sparkling sales.
3.Seasonality: Seasonal plot displays a fairly consistent month-on-month pattern.
4.The residuals are also present and more or less align to centre line.
5.The Multiplicative model are performing better in decomposition because the residual are more align to centre line.

QN3.Split the data into training and test. The test data should start in 1991.

Train Test Split

Before a forecast method is proposed, the method needs to be validated. For that purpose, data has to be split into two sets i.e. training and testing.
Training data helps in identifying and fitting right model(s) and test data is used to validate the same.
In case of time series data, the test data is the most recent part of the series so that the ordering in the data is preserved.

Training Data

	Sparkling
YearMonth	
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

YearMonth	Sparkling
1990-12-01	6047

132 rows × 1 columns

Test Data

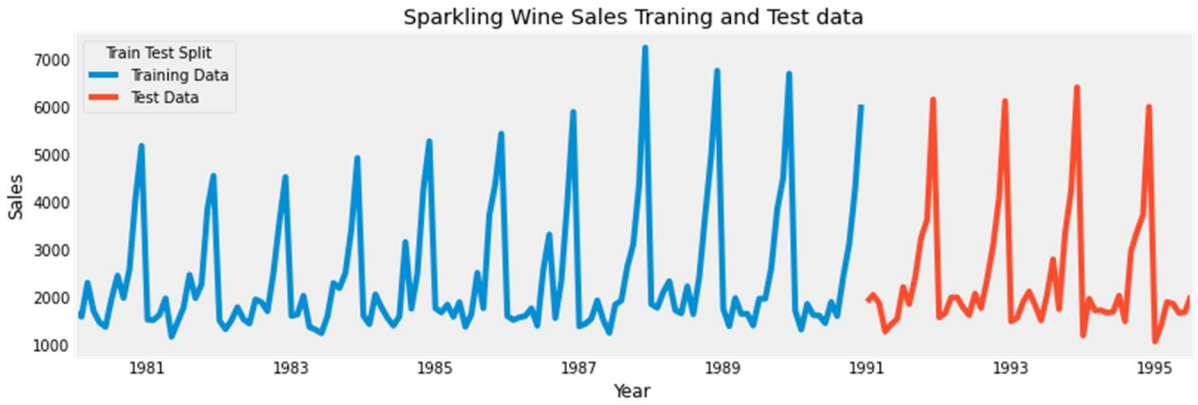
YearMonth	Sparkling
1991-01-01	1902
1991-02-01	2049
1991-03-01	1874
1991-04-01	1279
1991-05-01	1432
1991-06-01	1540
1991-07-01	2214
1991-08-01	1857
1991-09-01	2408
1991-10-01	3252
1991-11-01	3627
1991-12-01	6153
1992-01-01	1577
1992-02-01	1667
1992-03-01	1993
1992-04-01	1997
1992-05-01	1783
1992-06-01	1625
1992-07-01	2076
1992-08-01	1773
1992-09-01	2377

YearMonth	Sparkling
1992-10-01	3088
1992-11-01	4096
1992-12-01	6119
1993-01-01	1494
1993-02-01	1564
1993-03-01	1898
1993-04-01	2121
1993-05-01	1831
1993-06-01	1515
1993-07-01	2048
1993-08-01	2795
1993-09-01	1749
1993-10-01	3339
1993-11-01	4227
1993-12-01	6410
1994-01-01	1197
1994-02-01	1968
1994-03-01	1720
1994-04-01	1725
1994-05-01	1674
1994-06-01	1693
1994-07-01	2031
1994-08-01	1495
1994-09-01	2968
1994-10-01	3385
1994-11-01	3729

YearMonth	Sparkling
1994-12-01	5999
1995-01-01	1070
1995-02-01	1402
1995-03-01	1897
1995-04-01	1862
1995-05-01	1670
1995-06-01	1688
1995-07-01	2031

```
Shape of Sparkling of Train Data: (132, 1)
Shape of Sparkling of Test Data : (55, 1)
```

TABLE 5 TRAIN TEST SPLIT DATA



The Disjunction at 1991 showing the split of train and test data in Time series.

FIGURE 5 TRAIN TEST SPLIT PLOT

QN4. 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.

Exponential Smoothing Method

Smoothing is a technique applied to time series to remove the fine-grained variation between time steps.

Exponential smoothing is a time series forecasting method for univariate data. Exponential smoothing forecasting methods consist of flattening time series data and are similar in that a prediction is a weighted sum of past observations, Exponential smoothing averages or exponentially weighted moving averages consist of forecast based on previous periods data with exponentially declining influence 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).

There are three main types of exponential smoothing time series forecasting methods.

SES - ETS(A, N, N) - Simple Exponential Smoothing with additive errors

A simple method that assumes no systematic structure, an extension that explicitly handles trends, and the most advanced approach that add support for seasonality.

Single Exponential Smoothing, SES for short, also called Simple Exponential Smoothing, is a time series forecasting method for univariate data without a trend or seasonality.

$$\hat{Y}t+1 = \alpha Yt+\alpha(1-\alpha)Yt-1+\alpha(1-\alpha)2Yt-1+\cdots,\ 0<\alpha<1$$

It requires a single parameter, called alpha (a), also called the smoothing factor or smoothing coefficient.

This parameter controls the rate at which the influence of the observations at prior time steps decay exponentially.
Alpha is often set to a value between 0 and 1.
Large values mean that the model pays attention mainly to the most recent past observations, whereas smaller values mean more of the history is taken into account when making a prediction.

 # created class

 # Fitting the Simple Exponential Smoothing model and asking python to choose the optimal parameters

 # Check the parameters

```
{'smoothing_level': 0.07028442075641193,
 'smoothing_trend': nan,
 'smoothing_seasonal': nan,
 'damping_trend': nan,
 'initial_level': 1763.8402828521703,
 'initial_trend': nan,
 'initial_seasons': array([], dtype=float64),
 'use_boxcox': False,
 'lamda': None,
 'remove_bias': False}
```

FIGURE 6 SIMPLE EXPO PARAMETER

```
Plotting the Training data, Test data and the forecasted values
```

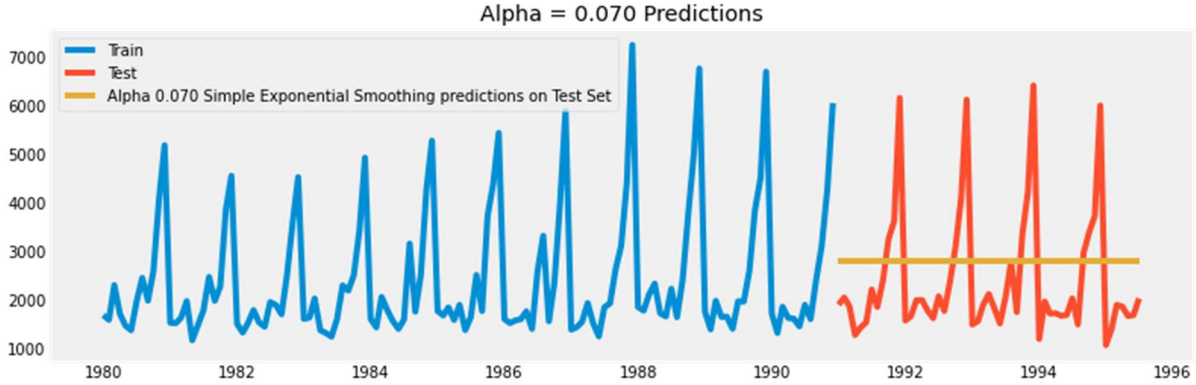


FIGURE 7 SIMPLE EXP SMOOTHING PLOT

Insights:

The simple Exponential model predicting the straight line, which does not support the level trend and seasonality complete missing. So this model would not work.

SES Model Accuracy Check by generating RMSE values

	Test RMSE
Alpha=0.70,	**SES1338.000861**

Double Exponential Smoothing or Holt - ETS(A, A, N) - Holt's linear method with additive errors

Double Exponential Smoothing is an extension to Exponential Smoothing that explicitly adds support for trends in the univariate time series.

Forecast Equation: $\hat{Y}t+1=lt+bt$

Level Equation: $lt=\alpha Yt+\alpha(1-\alpha)Yt-1$, $0 < \alpha < 1$

Trend Equation: $bt=\beta(lt-lt-1)+(1-\beta)bt-1$, $0 < \beta < 1$

where, lt is the estimate of level and bt is the trend estimate.

$\alpha$is the smoothing parameter for the level and β is the smoothing parameter for trend.

In addition to the alpha parameter for controlling smoothing factor for the level, an additional smoothing factor is added to control the decay of the influence of the change in trend called beta (b).

The method supports trends that change in different ways: an additive and a multiplicative, depending on whether the trend is linear or exponential respectively.

Double Exponential Smoothing with an additive trend is classically referred to as Holt's linear trend model, named for the developer of the method Charles Holt.

Additive Trend: Double Exponential Smoothing with a linear trend. Multiplicative Trend: Double Exponential Smoothing with an exponential trend. For longer range (multi-step) forecasts, the trend may continue on unrealistically. As such, it can be useful to dampen the trend over time.

Hyperparameters:

Alpha: Smoothing factor for the level.

Beta: Smoothing factor for the trend.

Trend Type: Additive or multiplicative.

Dampen Type: Additive or multiplicative.

Phi: Damping coefficient.

Initializing the Double Exponential Smoothing Model

```
-----Holt model Exponential Smoothing Estimated Parameters ---

{'smoothing_level': 0.6649999999999999, 'smoothing_trend': 0.0001, '
smoothing_seasonal': nan, 'damping_trend': nan, 'initial_level': 150
2.1999999999991, 'initial_trend': 74.87272727272739, 'initial_season
s': array([], dtype=float64), 'use_boxcox': False, 'lamda': None, 'r
emove_bias': False}
```

TABLE 6 HOLT MODEL PARAMETERS

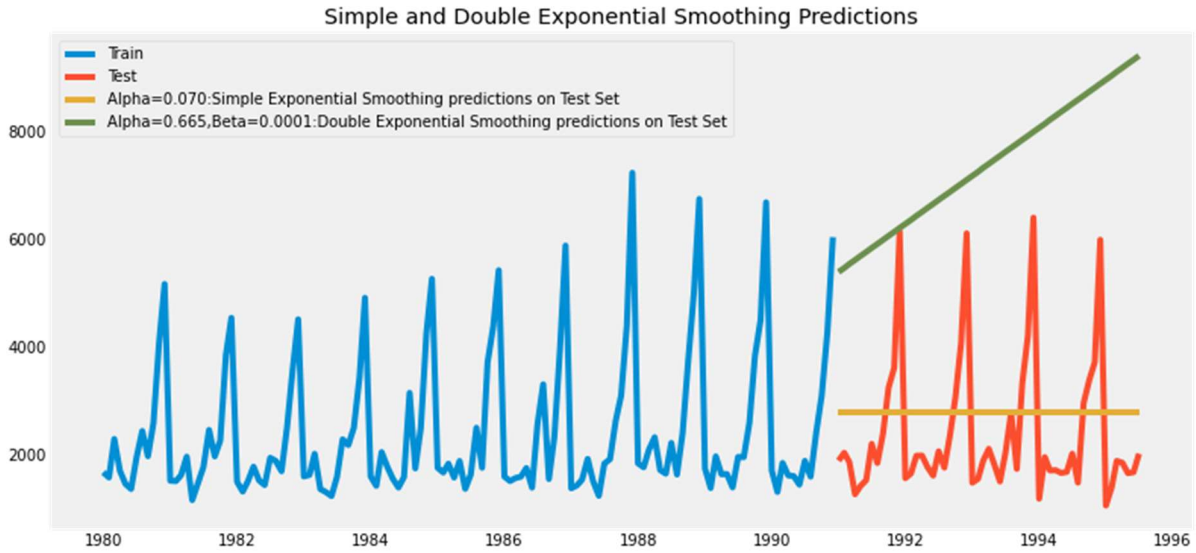


FIGURE 8 DOUBLE EXPO SMOOTHING PLOT

DES Model Accuracy

```
SES RMSE (calculated using statsmodels): 3262.10733251485
```

RMSE Summary of Models

	Test RMSE
Alpha=0.70,SES	1338.000861
Alpha=.665,Beta=0.0001:DES	5291.879833

TABLE 7 RMSE SUMMARY

Inferences:

1.Double Exponential model with low AIC value of(Alpha=.665,Beta=0.0001) giving the high to RSME 5291.87.
2.The plot trend is also not supporting the desired outcome.

Triple Exponential Smoothing or Holt-Winters - ETS(A, A, A) - **Holt Winter's** linear method with additive errors

Triple Exponential Smoothing is an extension of Exponential Smoothing that explicitly adds support for seasonality to the univariate time series.

This is an extension of Holt's method when seasonality is found in the data.

$$\text{Forecast equation: } Yt+1 = lt + bt + st - m(k+1)$$

$$\text{Level Equation: } lt = \alpha(Yt - st - m) + \alpha(1-\alpha)Yt-1, \quad 0 < \alpha < 1$$

$$\text{Trend Equation: } bt = \beta(lt - lt-1) + (1-\beta)bt-1, \quad 0 < \beta < 1$$

$$\text{Seasonal Equation: } \gamma(Yt - lt-1 - bt-1) + (1-\gamma)st-m, \quad 0 < \gamma < 1$$

This is also known as three parameters exponential or triple expone ntial because of the three smoothing parameters α, β and γ. This i s a general method and a true multi-step ahead forecast.

In addition to the alpha and beta smoothing factors, a new parameter is added called gamma (g) that controls the influence on the seasonal component.

As with the trend, the seasonality may be modelled as either an additive or multiplicative process for a linear or exponential change in the seasonality.

Additive Seasonality: Triple Exponential Smoothing with a linear seasonality. Multiplicative Seasonality: Triple Exponential Smoothing with an exponential seasonality. Triple exponential smoothing is the most advanced variation of exponential smoothing and through configuration, it can also develop double and single exponential smoothing models.

This method is sometimes called Holt-Winters Exponential Smoothing, named for two contributors to the method: Charles Holt and Peter Winters.

Hyperparameters:

Alpha: Smoothing factor for the level.

Beta: Smoothing factor for the trend.

Gamma: Smoothing factor for the seasonality.

Trend Type: Additive or multiplicative.

Dampen Type: Additive or multiplicative.

Phi: Damping coefficient.

Seasonality Type: Additive or multiplicative.

Period: Time steps in seasonal period.

Initializing the Triple Exponential Smoothing Model

---Holt Winters model Exponential Smoothing Estimated Parameters----

{'smoothing_level': 0.11127217693511166, 'smoothing_trend': 0.012360783126182025, 'sm oothing_seasonal': 0.4607177659431463, 'damping_trend': nan, 'initial_level': 2356.578307 8812697, 'initial_trend': -0.018442178724720648, 'initial_seasons': array([-636.23349205, -722.98346399, -398.64349841, -473.43073157,
-808.42502897, -815.35019273, -384.23061339, 72.99513671,
-237.44278517, 272.32607144, 1541.37826596, 2590.07759442]), 'use_boxcox': False, ' lamda': None, 'remove_bias': False}

TABLE 8 TES EXPO SMOOTHING PARAMETERS

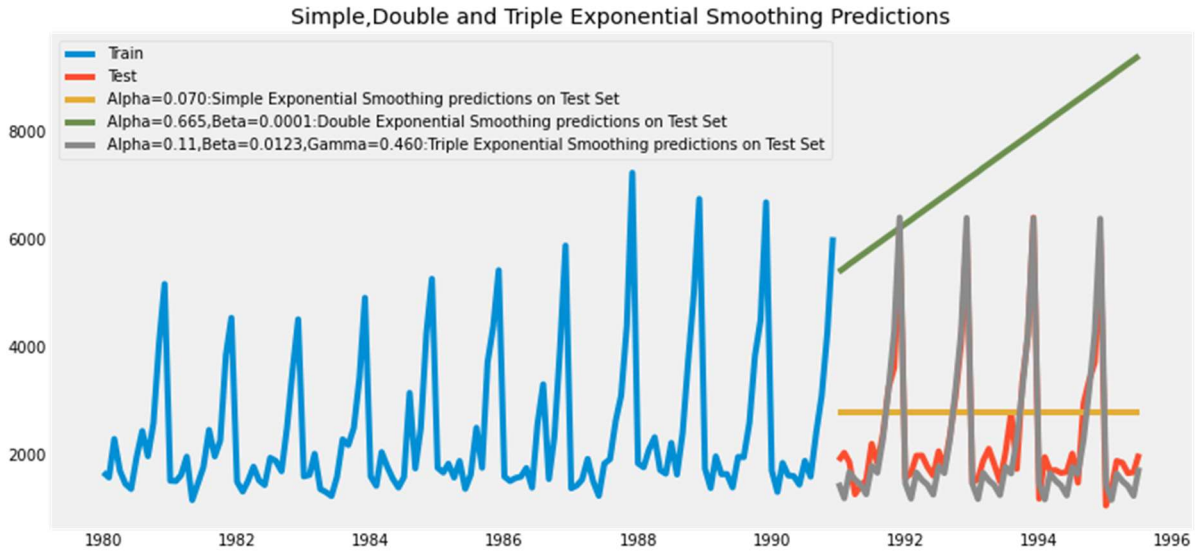


FIGURE 9 TRIPLE EXP SMOOTHING PLOT

Inferences:

1.Triple Exponential Smoothing with Lowest AIC(Alpha=0.11,Beta=0.012,Gamma=0.460) has given the Low RMSE value as 378.62.

2.Which is comparatively better over last two models RSME score.

3.Lower the Error model better.

4.In plot the clear trend and seasonality cane be defined by this model.

RMSE Comparison Table

	Test RMSE
Alpha=0.70,SES	**1338.000861**
Alpha=.665,Beta=0.0001:DES	**5291.879833**
Alpha=0.11,Beta=0.012,Gamma=0.460:TES	**378.625883**

TABLE 9 RMSE COMPARISON TABLE

Holt-Winters - ETS(A, A, M) - Holt Winter's linear method

ETS(A, A, M) model

Initializing the Double Exponential Smoothing Model

```
----Holt Winters model Exponential Smoothing Estimated Paramet
ers---

{'smoothing_level': 0.11101471561088701, 'smoothing_trend': 0.049314
5907614654, 'smoothing_seasonal': 0.36244934537370843, 'damping_tren
d': nan, 'initial_level': 2356.496908624238, 'initial_trend': -9.809
526161838415, 'initial_seasons': array([0.713711  , 0.68278724, 0.90
458411, 0.8053878 , 0.65571739,
       0.65388935, 0.88616088, 1.13350811, 0.91894498, 1.21186447,
       1.87099202, 2.37505867]), 'use_boxcox': False, 'lamda': None,
 'remove_bias': False}
```

TABLE 10 HOLT WINTERS MODEL PARAMETERS

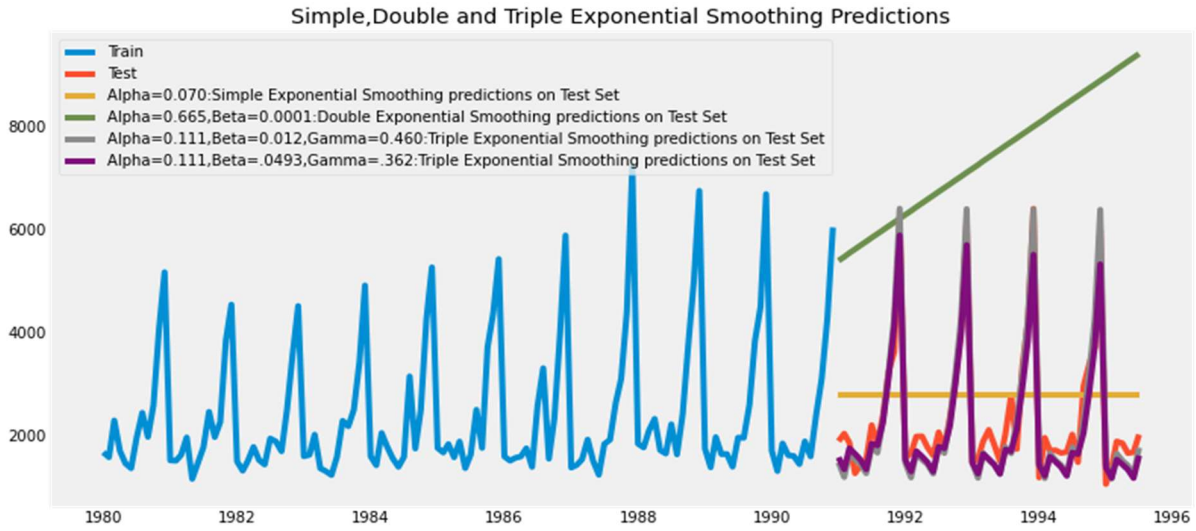


FIGURE 10 HOLT WINTER PLOT

Inference:

Among all the Triple Exponential Smoothing has performed the best on the test as expected since the data had both trend and seasonality.

RMSE Model Comparison	Test RMSE
Alpha=0.70, SES	1338.000861
Alpha=.665, Beta=0.0001:DES	5291.879833
Alpha=0.11, Beta=0.012, Gamma=0.460: TES	378.625883
Alpha=0.111, Beta=.0493, Gamma=.362: TES	402.936179

TABLE 11 RMSE SUMMARY

MODEL 1: LINEAR REGRESSION

A time series regression forecasts a time series as a linear relationship with the independent variables. The linear regression model assumes there is a linear relationship between the forecast variable and the predictor variables

For this particular linear regression, we are going to regress the 'Sparkling' variable against the order of the occurrence. For this we need to modify our training data before fitting it into a linear regression.

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	43
1991-02-01	2049	44
1991-03-01	1874	45
1991-04-01	1279	46
1991-05-01	1432	47

Last few rows of Test Data

YearMonth	Sparkling	time
1995-03-01	1897	93
1995-04-01	1862	94
1995-05-01	1670	95
1995-06-01	1688	96
1995-07-01	2031	97

TABLE 12 LR TRAIN TEST DATA

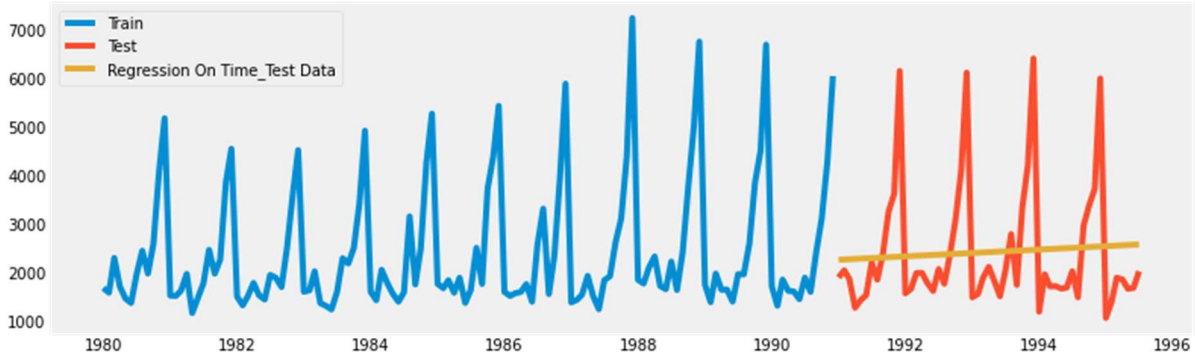


FIGURE 11 LINEAR REGRESSION PLOT

For RegressionOnTime forecast on the Test Data, RMSE is 1275.867

Inference:

The Linear Regression models giving the RMSE 1275.86 which is much higher than the TES Model.

RMSE Comparison Table

	Test RMSE
ALPHA=0.70,SES	1338.000861
ALPHA=.665, BETA=0.0001:DES	5291.879833
ALPHA=0.11, BETA=0.012,GAMMA=0.460:TES	378.625883
ALPHA=0.111, BETA=.0493,GAMMA=.362:TES	402.936179
REGRESSIONONTIME	**1275.867052**

TABLE 13 RMSE COMPARISON SUMMARY

Model 2: Naive Approach: y^t+1=yt

The Naive Bayes method is a classification algorithm that uses Bayes' theorem to predict the probability of a class given a set of features. However, there is a method called Naive Method which uses the most recent value as the forecasted value for the next time step. The assumption followed by this method is that its value tomorrow is equal to its value today

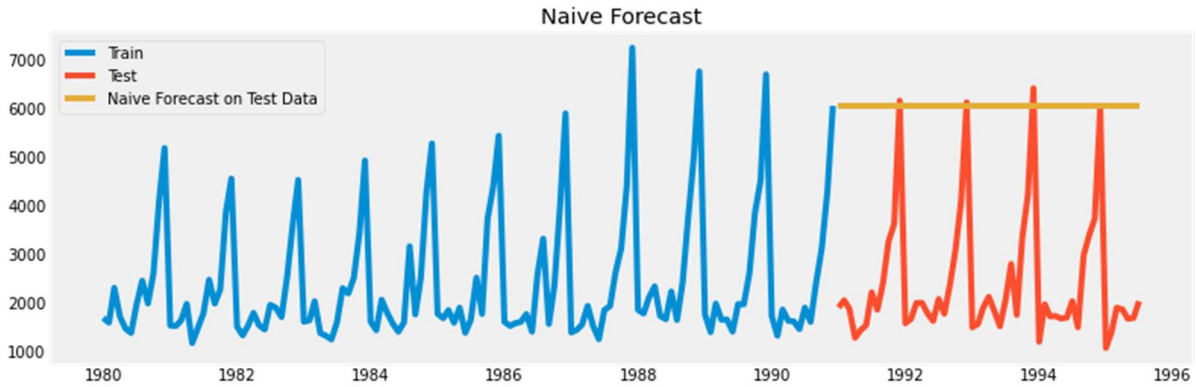


FIGURE 12 NAIVE BAYES METHOD PLOT

```
For RegressionOnTime forecast on the Test Data,  RMSE is 3864.279
```

Inferences:
The Naive model uses the last value for forecasting hence the RSME value are too hight to 3 864.27.

RMSE Summary of Models

	Test RMSE
ALPHA=0.70,SES	1338.000861
ALPHA=.665, BETA=0.0001:DES	5291.879833
ALPHA=0.11, BETA=0.012,GAMMA=0.460:TES	378.625883
ALPHA=0.111, BETA=.0493, GAMMA=.362:TES	402.936179
REGRESSIONONTIME	**1275.867052**
Naive Model	**3864.279352**

TABLE 14 RMSE SUMMARY

Simple Average Model

The method is very simple: average the data by months or quarters or years and them calculate the average for the period. Then find out, what percentage it is to the grand average.

	Sparkling	mean_forecast
YearMonth		
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

TABLE 15 SIMPLE AVERAGE

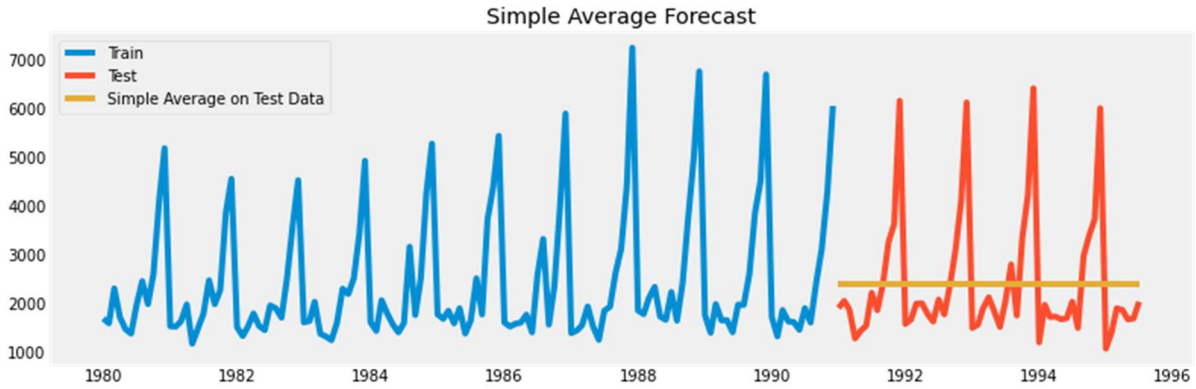


FIGURE 13 SIMPLE AVERAGE PLOT

For Simple Average forecast on the Test Data, RMSE is 1275.082

RMSE Summary of Models

	Test RMSE
ALPHA=0.70,SES	1338.000861
ALPHA=.665, BETA=0.0001: DES	5291.879833
ALPHA=0.11, BETA=0.012, GAMMA=0.460:TES	378.625883
ALPHA=0.111, BETA=.0493, GAMMA=.362:TES	402.936179
REGRESSIONONTIME	**1275.867052**
Naive Model	**3864.279352**
Simple Average	**1275.081804**

TABLE 16 RMSE SUMMARY OF MODELS

Moving Average Forecast

The moving average is a statistical method used for forecasting long-term trends. The technique represents taking an average of a set of numbers in a given range while moving the range. In simple terms, a moving average plot takes the average of several different points in the data set and then plots it over time.

Two main types of moving averages:

1) Centred moving average - calculated as the average of raw observations at, before and after time, t.
2) Trailing moving average - uses historical observations and is used on time series forecasting.

The rolling () function on the Series Pandas object will automatically group observations into a window.

The main advantage of the moving average method is that it takes into account all previous values when predicting future values. This helps to reduce the effect of outliers when making predictions and also makes it easier to identify seasonal patterns in a time-series data set. The moving average method is an effective tool for short-term forecasting due to its flexibility and ease of use. Its ability to take into account all past values when making predictions ensures accuracy while its ability to identify seasonal patterns means that it can be used effectively for long-term forecasting too

This algorithm helps us to forecast new observations based on a time series. This algorithm uses smoothing methods. The moving average algorithm is used only on time series that DOESN'T have a trend. This method is by far the easiest and it consists of making the arithmethic mean of the last ,,n'' observations contained by the time series to forecast the next observation. We use the following

$$formula: MA\_(t+1) = (\sum\_(i=t-n)^t x\_i)/n$$

We need to find the optimal number,, n" of observation to be used in the forecast. We can find it by checking the square error mean of multiple ,,n'' observations. We should start at 3 observations and we can go up to half of the data set size + 1.

Moving Average train Data

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

TABLE 17 MOVING AVERAGE TOP 5 DATA ON TRAIN

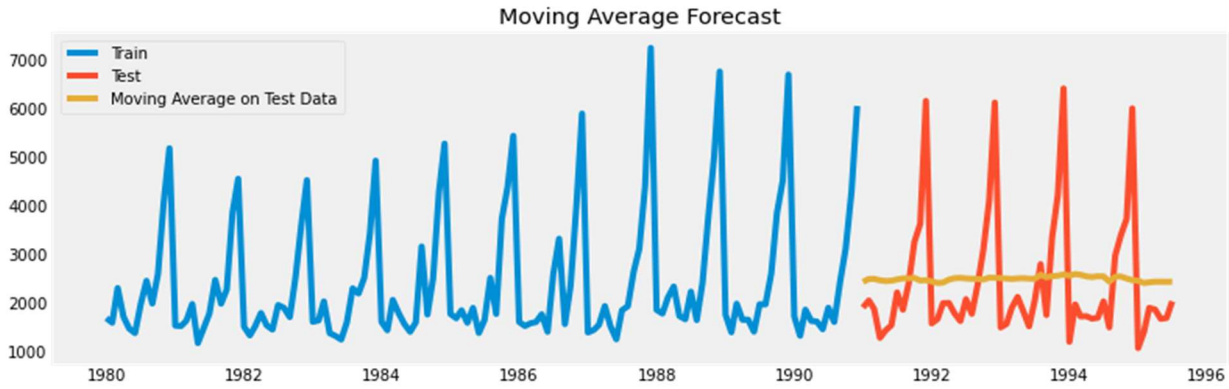


FIGURE 14 MOVING AVERAGE PLOT

For Moving Average forecast on the Test Data, RMSE is 1275.082

RMSE Summary of Models

	Test RMSE
ALPHA=0.70,SES	1338.000861
ALPHA=.665,BETA=0.0001:DES	5291.879833
ALPHA=0.11,BETA=0.012,GAMMA=0.460:TES	378.625883
ALPHA=0.111,BETA=.0493,GAMMA=.362:TES	402.936179
REGRESSIONONTIME	**1275.867052**
Naive Model	**3864.279352**
Simple Average	**1275.081804**
Moving Average	**1267.92533**

TABLE 18 RMSE SUMMARY OF MODELS

Trailing Moving Average

	Sparkling	Trailing_2	Trailing_4	Trailing_6	Trailing_9
YearMonth					
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

[187 rows x 5 columns]

(None, 3)

TABLE 19 TRAILING MOVING AVERAGE

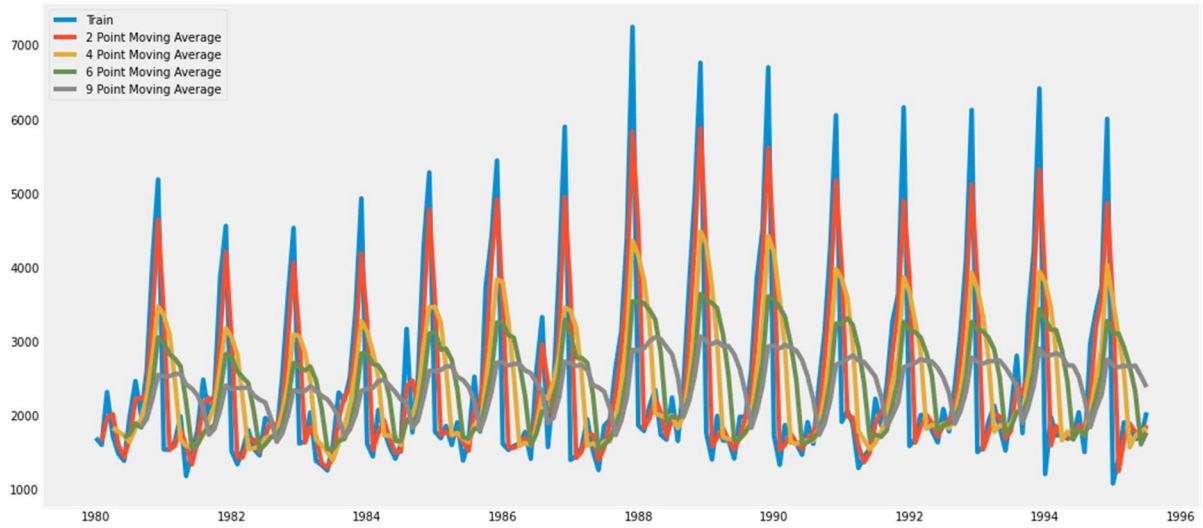


FIGURE 15 TRAILING MOVING AVERAGE FORECAST PLOT

```
Moving Average of train 5166.5
              and test  1859.5
```

Last five Data of Trailing Moving average

```
YearMonth
1990-08-01    1752.0
1990-09-01    2014.5
1990-10-01    2770.0
1990-11-01    3701.0
1990-12-01    5166.5
Name: Trailing_2, dtype: float64
```

TABLE 20 LAST FIVE DATA OF TRAILING MA

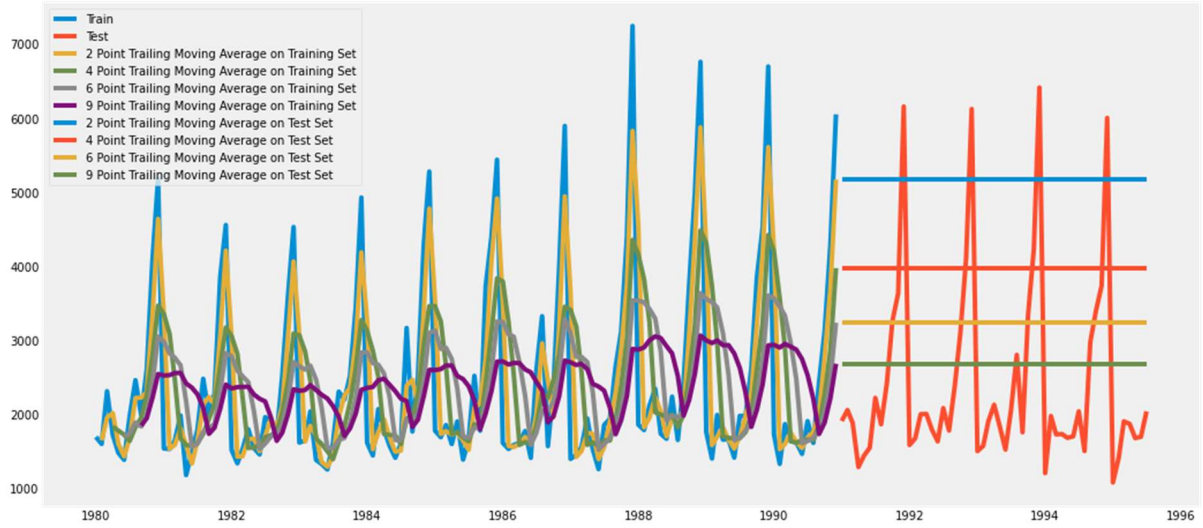


FIGURE 16 TRAILING MA ON POINTS PLOT

The Tabulating the RMSE Values of Trailing Average point wise forecast models

```
For 2 point Moving Average Model forecast on the Test Data,RMSE is 3046
.976
For 4 point Moving Average Model forecast on the Test Data,RMSE is 2021
.856
For 6 point Moving Average Model forecast on the Test Data,RMSE is 1521
.611
For 9 point Moving Average Model forecast on the Test Data,RMSE is 1304
.619
```

TABLE 21 RMSE TRAILING AVERAGE POINT WISE PLOT

Inference:

1.For the MA model the RMSE is lowest at 9 point, The Moving Average Model forecast RMSE is 1304.619
2.Which still higher than the other TES models.

RMSE Summary of Models

	Test RMSE
ALPHA=0.70,SES	1338.000861
ALPHA=.665, BETA=0.0001:DES	5291.879833
ALPHA=0.11, BETA=0.012,GAMMA=0.460:TES	378.625883
ALPHA=0.111, BETA=.0493, GAMMA=.362: TES	402.936179
REGRESSIONONTIME	1275.867052
Naive Model	3864.279352
Simple Average	1275.081804
Moving Average	1267.92533
2pointTrailingMovingAverage	3046.976092
4pointTrailingMovingAverage	2021.85588
6pointTrailingMovingAverage	1521.61125
9pointTrailingMovingAverage	1304.618912

TABLE 22 RMSE SUMMARY

Q5.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.

Checking for stationarity of the whole Time Series data.

Stationarity

A stationary time series is one whose statistical properties such as mean, variance, autocorrelation, etc. are all constant over time.
Stationarity means that the autocorrelation of lag 'k' depends on k, but not on time t.
Let Xt denote the time series at time t.
Autocorrelation of lag k is the correlation between Xt and X(t−k)

Strong stationarity: is a stochastic process whose unconditional joint probability distribution does not change when shifted in time. Consequently, parameters such as mean and variance also do not change over time.
Weak stationarity: is a process where mean, variance, autocorrelation are constant throughout the time Stationarity is important as non-stationary series that depend on time have too many parameters to account for when modelling the time series. diff() method can easily convert a non-stationary series to a stationary series.

We will try to decompose seasonal component of the above decomposed time series.

Check for Stationarity

There are multiple tests that can be used to check stationarity.

ADF(Augmented Dicky Fuller Test)

KPSS

PP (Phillips-Perron test)

Let's just perform the ADF which is the most commonly used one

Dickey-Fuller Test - Dicky Fuller Test on the timeseries is run to check for stationarity of data.

Hypothesis

Null Hypothesis H0: Time Series is non-stationary.
Alternate Hypothesis Ha: Time Series is stationary.

So Ideally if p-value < 0.05 then null hypothesis: TS is non-stationary is rejected else the TS is non-stationary is failed to be rejected

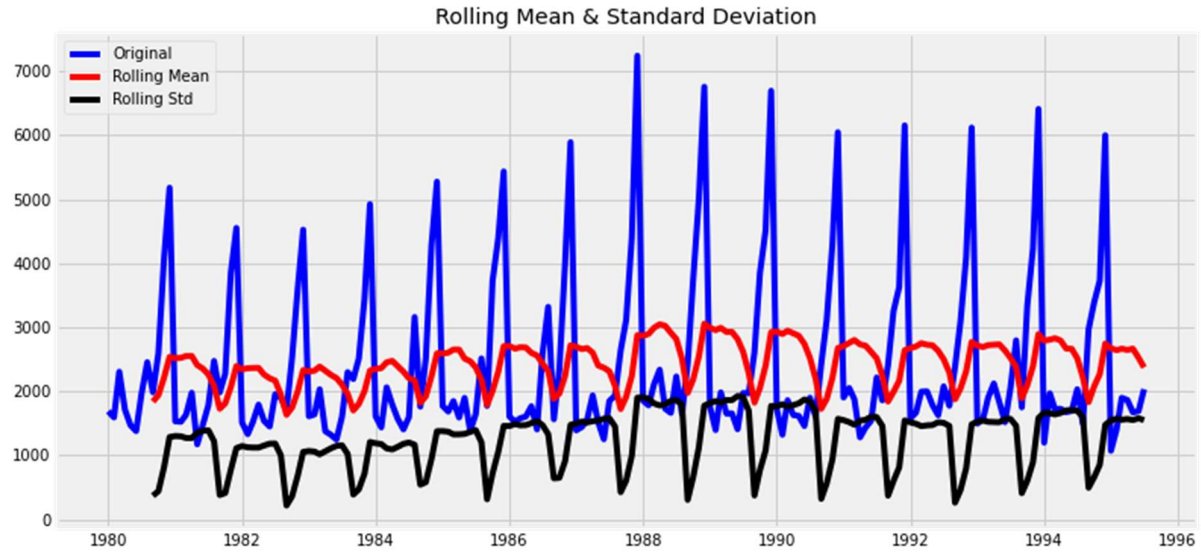


FIGURE 17 DATA STATIONARY PLOT

```
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
```

TABLE 23 DICKEY FULLER TEST

Insight:

The p-value 0.60 is very large, and not smaller than 0.05 and thus we fail to challenge the Null hypothesis so the TS data of Sparkling is not a stationary in nature.

Making a Time Series Stationary by - Differencing 'd'

Differencing 'd' is done on a non-stationary time series data one or more times to convert it into stationary. (d=1) 1st order differencing is done where the difference between the current and previous (1 lag before) series is taken and then checked for stationarity using the ADF(Augmented Dicky Fueller) test. If differenced time series is stationary, we proceed with AR modelling. Else we do (d=2) 2nd order differencing, and this process repeats till we get a stationary time series

1st order differencing equation is : $y_t = y_t - y_{t-1}$
2nd order differencing equation is : $y_t = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2})$ and so on…
The variance of a time series may also not be the same over time. To remove this kind of non-stationarity, we can transform the data. If the variance is increasing over time, then a log transformation can stabilize the variance.
Let us take a difference of order 1 and check whether the Time Series is stationary or not.

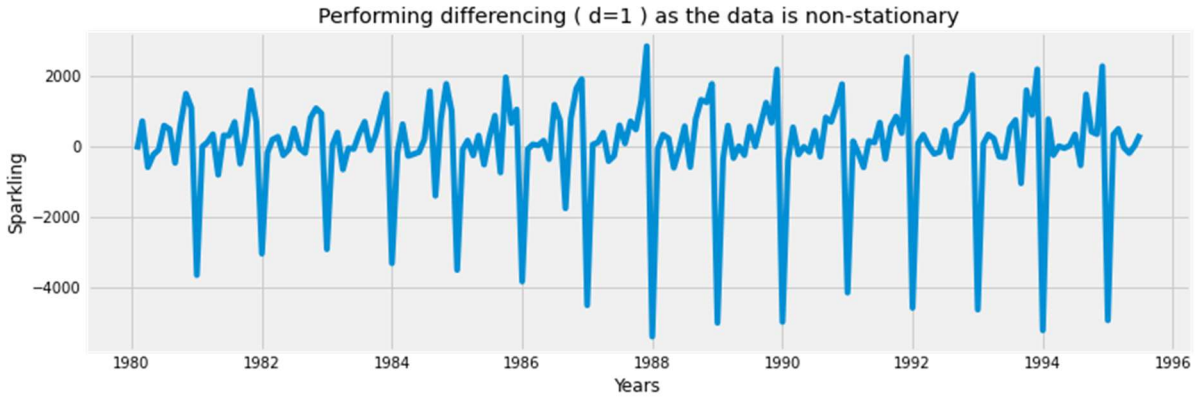


FIGURE 18 DICKEY FULLER TEST

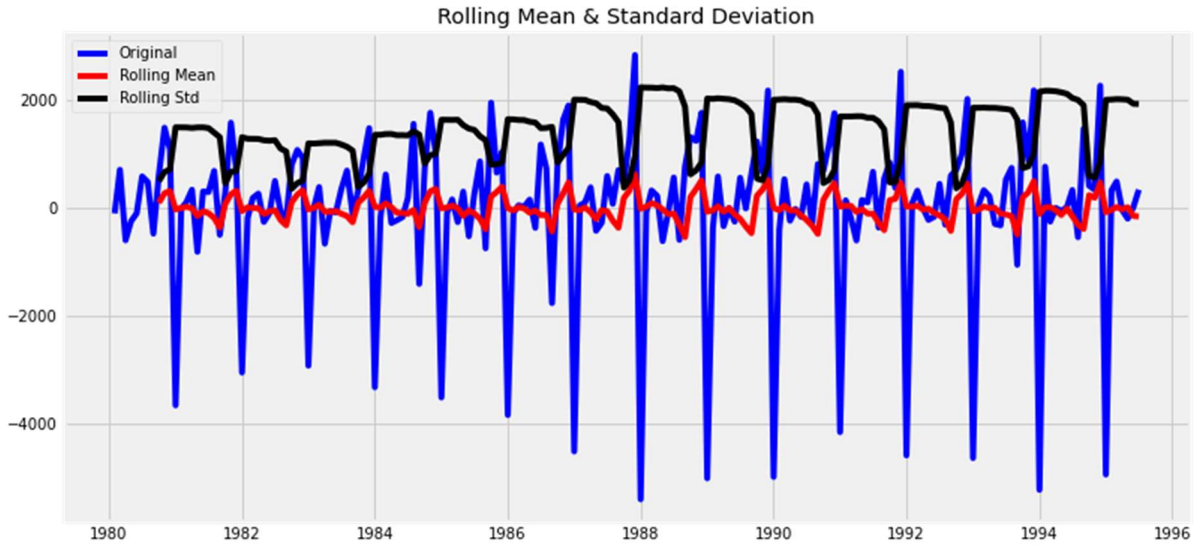


FIGURE 19 AFTER DICKEY FULLER CORRECTION PLOT

```
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
```

TABLE 24 AFTER DICKEY FULLER CORRECTION

Insight : Now, We can see that at α= 0.05 the Time Series is indeed stationary.

Plotting the Autocorrelation function plots on the whole data.

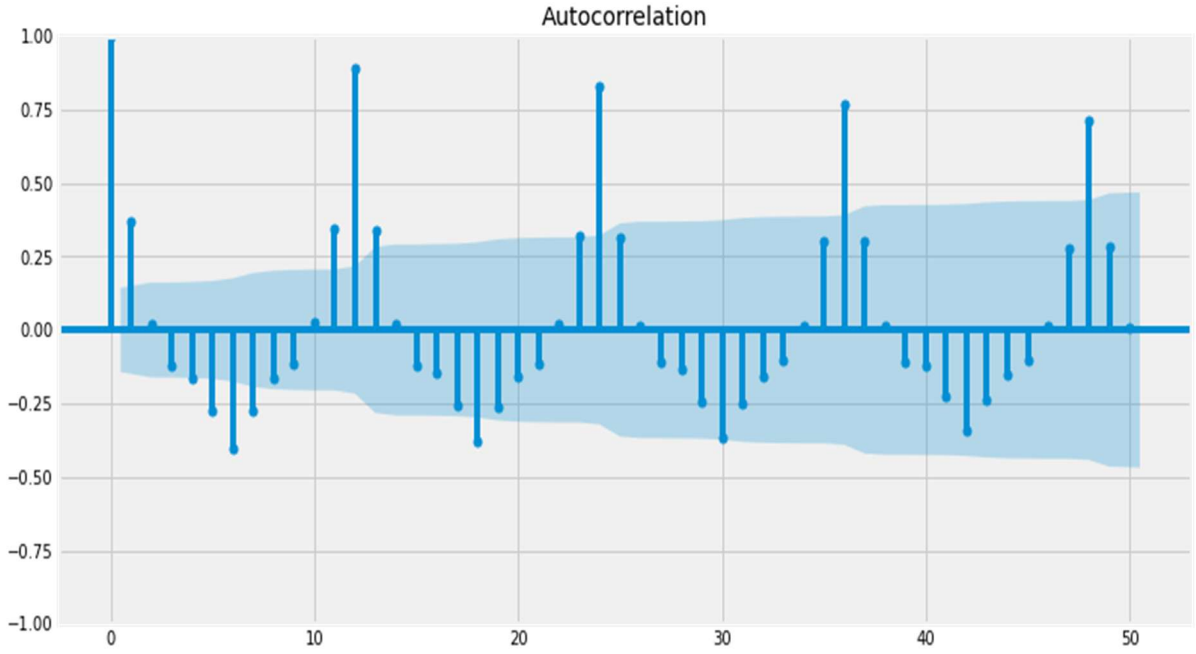


FIGURE 20 AUTOCORRELATION

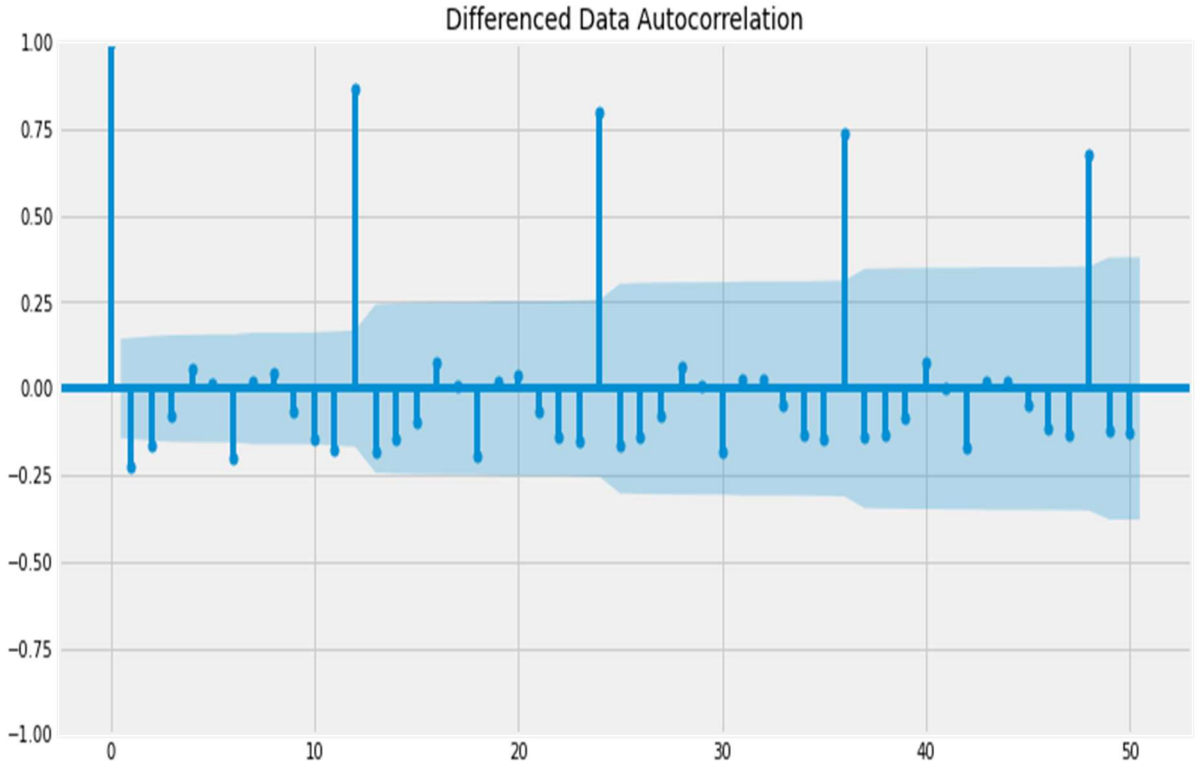


FIGURE 21 DIFFERENCED AUTO CORRELATION

QN 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.

Auto Regressive (AR) Models

Autoregression means regression of a variable on itself which means Autoregressive models use previous time period values to predict the current time period values.

An autoregressive (AR) model is a representation of a type of random process; as such, it is used to describe certain time-varying processes in nature, economics, etc

The autoregressive model specifies that the output variable depends linearly on its own previous values and on a stochastic term (an imperfectly predictable term); thus the model is in the form of a stochastic difference equation.\

One of the fundamental assumptions of an AR model is that the time series is assumed to be a stationary process. An AR(p) model (Auto-Regressive model of order p) can be written as:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots\ldots + \phi_p y_{t-p} + \varepsilon_t$$

ε_t is an error term which is an independent and identically distributed random variable (or in other words, a white noise) with the parameters mean = 0 and standard deviation = σ

The ϕ are regression coefficients multiplied by lagged time series variable, which captures the effect of the input variable on the output, provided intermediate values do not change.

Choose the order 'p' of AR model

We look at the Partial Autocorrelations of a stationary Time Series to understand the order of Auto-Regressive models. For an AR model, 2 ways to identify order of 'p':

1) PACF Approach: the PACF method where the (Partial Auto Correlation Function) values cut off and become zero after a certain lag. PACF vanishes if there is no regression coefficient that far back. The cut-off value where this happens can be taken as the order of AR as ‘p’. This can be seen from a PACF plot.

If the 2nd PACF vanishes (cut off in PACF) then the 2nd coefficient is not considered and thus ‘p’ is 1.

If the 3rd PACF vanishes (cuts off in PACF) then the 3rd coefficient is not considered and thus ‘p’ is 2 and so on…

Partial Autocorrelation of order 2 = Partial autocorrelation of lag 2 = Correlation between Xt and Xt−2 holding Xt−1 fixed.

2) Lowest AIC Approach: the lowest Akaike Information Criteria (AIC) value compared among different orders of ‘p’ is considered.

Moving Average (MA) Models

Moving average model considers past residual values to predict the current time period values. These past residuals are past prediction errors. For a MA model, the residual or error component is modeled. The moving average model MA(q) of order q can be represented as:

$$y_t=\varepsilon_t+\theta_1\varepsilon_{t-1}+\ldots\ldots+\theta_q\varepsilon_{t-q}$$

Where yt time series variable, θ are numeric coefficients multiplied to lagged residuals and ε is the residual term considered as a purely random process with mean 0, variance σ2 and $Cov(\varepsilon_{t-1},\varepsilon_{t-q})= 0$.

Choose the order 'q' of MA model

We look at the Autocorrelations of a stationary Time Series to understand the order of Moving Average models. For a MA model,

1) ACF Approach: ACF (Autocorrelation Function) values cut off at a certain lag. ACF vanishes, and there are no coefficients that far back; thus, the cut-off value where this happens is taken as the order of MA as ‘q’. This can be seen from the ACF plot.

2) Lowest AIC Approach: the lowest Akaike Information Criteria (AIC) value compared among different orders of ‘q’ is considered

ARIMA model¶

Auto Regressive Integrated Moving Average is a way of modelling time series data for forecasting or predicting future data point An autoregressive integrated moving average (ARIMA) model is a generalization of an autoregressive moving average (ARMA) model. Both of these models are fitted to time series data either to better understand the data or to predict future points in the series (forecasting). ARIMA models are applied in some cases where data show evidence of non-stationarity, where an initial differencing step (corresponding to the "integrated" part of the model) can be applied one or more times to eliminate the non-stationarity. ARIMA model is of the form: ARIMA(p,d,q): p is AR parameter, d is differential parameter, q is MA parameter

ARIMA(1,0,0)

$$yt = a1yt-1 + \epsilon t$$

ARIMA(1,0,1)

$$yt = a1yt-1 + \epsilon t + b1\epsilon t-1$$

ARIMA(1,1,1)

$$\Delta yt = a1\Delta yt-1 + \epsilon t + b1\epsilon t-1 \text{ where } \Delta yt = yt - yt-1$$

Improving AR Models by making Time Series stationary through Moving AVerage Forecasts
ARIMA models consist of 3 components: -
AR model: The data is modelled based on past observations.
Integrated component: Whether the data needs to be differenced/transformed.
MA model: Previous forecast errors are incorporated into the model.

Building an Automated version of an ARIMA model for which the best parameters are selected in accordance with the lowest Akaike Information Criteria (AIC

Getting a combination of different parameters of p and q in the range of 0 and 2, We have kept 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)
```

TABLE 25 ARIMA BEST COMBINATION

```
ARIMA(0, 1, 0) - AIC:2267.6630357855465
ARIMA(0, 1, 1) - AIC:2263.0600155918555
ARIMA(0, 1, 2) - AIC:2234.408323131469
ARIMA(1, 1, 0) - AIC:2266.6085393190097
ARIMA(1, 1, 1) - AIC:2235.7550946707897
```

```
ARIMA(1, 1, 2) - AIC:2234.5272004520434
ARIMA(2, 1, 0) - AIC:2260.36574396809
ARIMA(2, 1, 1) - AIC:2233.7776262944917
ARIMA(2, 1, 2) - AIC:2213.5092144007594
```

TABLE 26 AIC VALUE WITH COMBINATIONS

AIC in Descending Order for get the AIC value minimum.

	param	AIC
8	(2, 1, 2)	2213.509214
7	(2, 1, 1)	2233.777626
2	(0, 1, 2)	2234.408323
5	(1, 1, 2)	2234.527200
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 27 AIC VALUES IN DESCENDING ORDER

Modelling the ARIMA on lowest AIC with combination (2,1,2)

```
SARIMAX Results
==========================================================================
=======
Dep. Variable:              Sparkling   No. Observations:
    132
Model:                 ARIMA(2, 1, 2)   Log Likelihood               -1
101.755
Date:                Sat, 08 Apr 2023   AIC                          2
213.509
Time:                        13:00:57   BIC                          2
227.885
Sample:                    01-01-1980   HQIC                         2
219.351
                         - 12-01-1990

Covariance Type:                 opg

==========================================================================
=======
                 coef    std err          z      P>|z|      [0.025
 0.975]
--------------------------------------------------------------------------
-------
ar.L1          1.3121      0.046     28.775      0.000       1.223
  1.401
ar.L2         -0.5593      0.072     -7.738      0.000      -0.701
 -0.418
ma.L1         -1.9917      0.109    -18.212      0.000      -2.206
 -1.777
```

```
ma.L2          0.9999      0.110      9.106      0.000       0.785
  1.215
sigma2      1.099e+06      2e-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
==============================================================================
============

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (
complex-step).
[2] Covariance matrix is singular or near-singular, with condition numb
er 3.68e+27. Standard errors may be unstable.
```

TABLE 28 ARIMA MODEL RESULTS

RMSE for the Model is 1299.9792795539977

Inferences:

Criteria to choose the best fit model is the lowest/minimum AIC value For ARIMA (p, d, q) we got (2, 1, 2),model with the least AIC of 1299.97.

Here,

p = non-seasonal AR order = 2,
d = non-seasonal differencing = 1,
q = non-seasonal MA order = 2,

S = time span of repeating seasonal pattern = 12
The ARIMA model with order 2,1,2 has AIC of 1299.97.50 which is pretty higher than few models.

RMSE Summary of Models

	Test RMSE
ALPHA=0.70,SES	1338.000861
ALPHA=.665,BETA=0.0001:DES	5291.879833
ALPHA=0.11,BETA=0.012,GAMMA=0.460:TES	378.625883
ALPHA=0.111,BETA=.0493,GAMMA=.362:TES	402.936179
REGRESSIONONTIME	1275.867052
Naive Model	3864.279352
Simple Average	1275.081804
Moving Average	1267.92533
2pointTrailingMovingAverage	3046.976092
4pointTrailingMovingAverage	2021.85588
6pointTrailingMovingAverage	1521.61125
9pointTrailingMovingAverage	1304.618912
ARIMA(2,1,2)	1299.97928

TABLE 29 RMSE SUMMARY OF MODELS

Building an Automated version of a SARIMA model for which the best parameters are selected in accordance with the lowest Akaike Information Criteria (AIC)

SAMRIMA Model

SARIMA stands for Seasonal Autoregressive Integrated Moving Average. It is a statistical analysis model that uses time-series data to either better understand the data set or to predict future trends. SARIMA is an extension of ARIMA (Autoregressive Integrated Moving Average) model and is used when the time series data has seasonal frequency yet also by non-seasonal differencing in Univariate data. It adds three new hyperparameters to specify the autoregression (AR), differencing (I) and moving average (MA) for the seasonal component of the series, as well as an additional parameter for the period of the seasonality.

A typical SARIMA model equation looks like the following –

SARIMA(p,d,q)x(P,D,Q)lag

The parameters for these types of models are as follows:

p and seasonal P: indicate the number of AR terms (lags of the stationary series)

d and seasonal D: indicate differencing that must be done to stationary series

q and seasonal Q: indicate the number of MA terms (lags of the forecast errors)

lag: indicates the seasonal length in the data

Seasonality S(P, D, Q, s), where s is simply the season's length. This component requires the parameters P and Q which are the same as p and q, but for the seasonal component. Finally, D is the order of seasonal integration representing the number of differences required to remove seasonality from the series.

Combining all, we get the SARIMA (p, d, q)(P, D, Q, s) model.
Let us look at the ACF plot once more to understand the seasonal parameter for the SARIMA model.

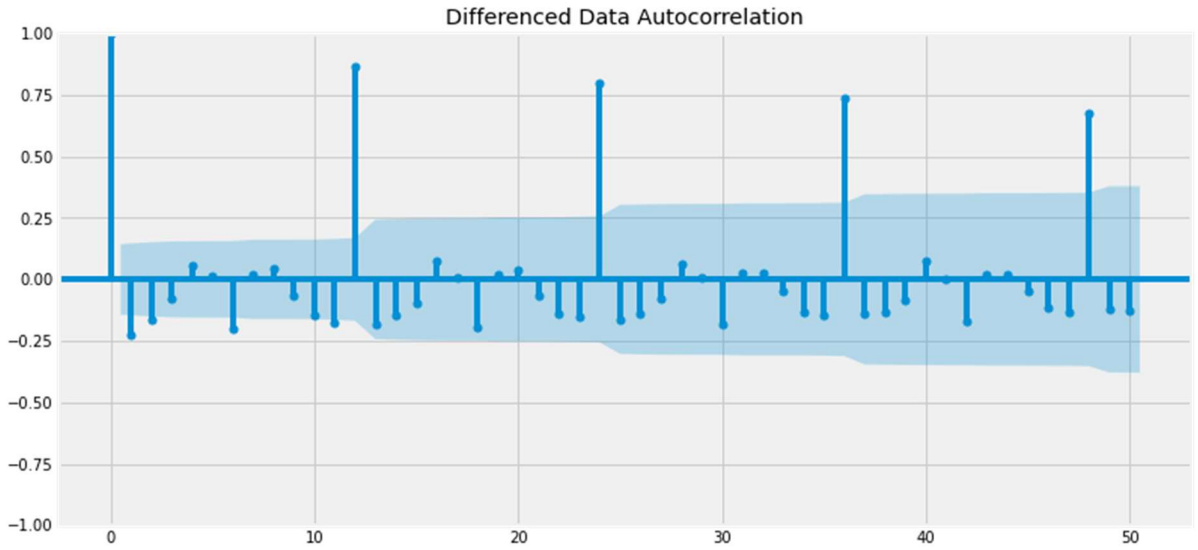


FIGURE 22 SEASONALITY FOR SRAIMA

INSIGHT:

We see that there can be a seasonality at 12. But from the decomposition at the start we ascertained that visually it looks like the seasonality =12 and thus using the same

Setting the seasonality as 6 to estimate parameters using auto SARIMA model

```
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)
```

TABLE 30 PARAMETER COMBINATIONS

Getting the lowest AIC by The best combination using param by a loop method.

The best 5 are as follows:

	Param	seasonal	AIC
50	(1, 1, 2)	(1, 0, 2, 12)	1555.584248
53	(1, 1, 2)	(2, 0, 2, 12)	1555.934568
26	(0, 1, 2)	(2, 0, 2, 12)	1557.121563
23	(0, 1, 2)	(1, 0, 2, 12)	1557.160507
77	(2, 1, 2)	(1, 0, 2, 12)	1557.340404

TABLE 31 BEST AIC FOR SARIMA

```
SARIMAX Results
==========================================================================
==================
Dep. Variable:  y                              No. Observations:  132

Model: SARIMAX(1, 1, 2)x(1, 0, 2, 12)      Log Likelihood  -770.792

Date: Sat, 08 Apr 2023                              AIC   1555.584

Time: 12:50:42                                      BIC   1574.095

Sample:  0                                          HQIC   1563.083

                                                  - 132

Covariance Type:                                  opg

==========================================================================
=======
                 coef    std err          z      P>|z|      [0.025
 0.975]
--------------------------------------------------------------------------
-------
ar.L1         -0.6281      0.255     -2.463      0.014      -1.128
 -0.128
ma.L1         -0.1041      0.225     -0.463      0.643      -0.545
  0.337
ma.L2         -0.7276      0.154     -4.734      0.000      -1.029
 -0.426
ar.S.L12       1.0439      0.014     72.842      0.000       1.016
  1.072
ma.S.L12      -0.5551      0.098     -5.663      0.000      -0.747
 -0.363
ma.S.L24      -0.1355      0.120     -1.133      0.257      -0.370
  0.099
sigma2      1.506e+05   2.03e+04      7.400      0.000    1.11e+05
1.9e+05
==========================================================================
============
Ljung-Box (L1) (Q):                   0.04   Jarque-Bera (JB): 11.72
```

```
Prob(Q):                                  0.84   Prob(JB):         0.00
Heteroskedasticity (H):                   1.47   Skew:             0.36
Prob(H) (two-sided):                      0.26   Kurtosis:         4.48
===================================================================================
Warnings:
[1] Covariance matrix calculated using the outer product of gradients (
complex-step).
```

TABLE 32 SARIMA MODEL SUMMARY

Plotting the assumption of Models

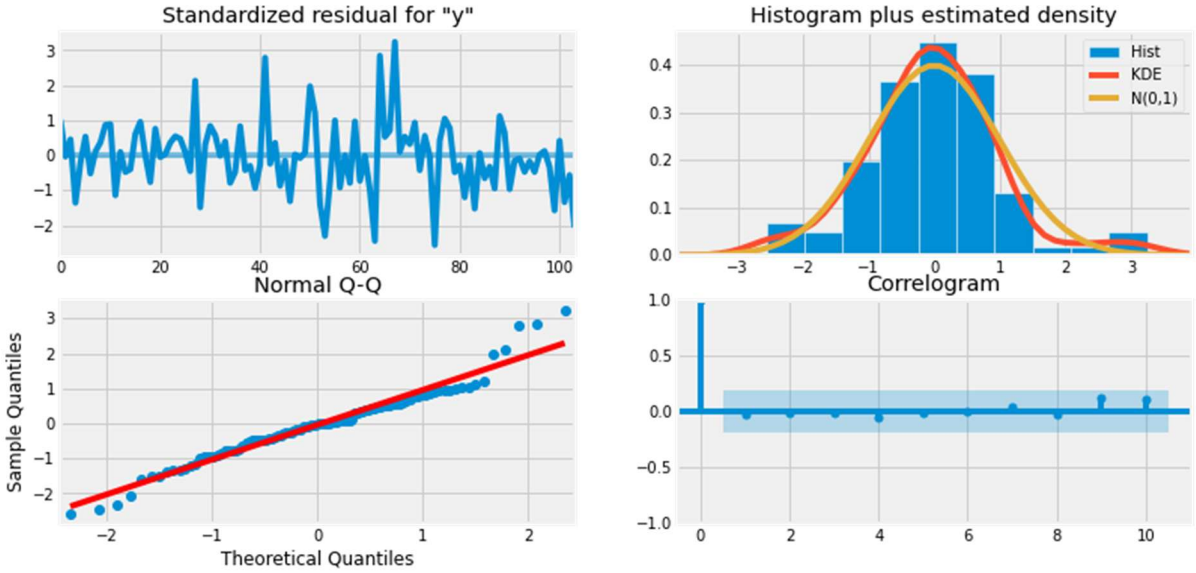


FIGURE 23 ASSUMPTION OF MODELS

y	mean	mean_se	mean_ci_lower	mean_ci_upper
0	1327.356639	388.344284	566.215828	2088.497449
1	1315.107471	402.012226	527.177986	2103.036955
2	1621.581317	402.005826	833.664377	2409.498256
3	1598.852273	407.244684	800.667360	2397.037186

y	mean	mean_se	mean_ci_lower	mean_ci_upper
4	1392.680152	407.975275	593.063307	2192.296997

TABLE 33 CI OF SARIMA MODEL

The RMSE of SARIMA model with lowest AIC is 528.6063958242969

Inferences:

Criteria to choose the best fit model is the lowest/minimum AIC value For ARIMA(p, d, q) × (P, D, Q)S, we got SARIMAX(1, 1, 2)x(1, 0, 2, 12)model with the least AIC of 1555.58248

Here,

```
p = non-seasonal AR order = 1,

d = non-seasonal differencing = 1,

q = non-seasonal MA order = 2,

P = seasonal AR order = 1,

D = seasonal differencing = 0,

Q = seasonal MA order = 2,

S = time span of repeating seasonal pattern = 12
```

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. In SARIMA model we fail to reject the null hypothesis because some p values are above the 0.05 in ma.L1 and ma.S.L24. RMSE wise its stand next to TES.

Building the most optimum model on the Full Data.

```
SARIMAX Results
==========================================================================
==================
Dep. Variable:      Sparkling   No. Observations:                 187
Model: SARIMAX(1, 1, 2)x(1, 0, 2, 12)   Log Likelihood       -1173.413
Date:      Sat, 08 Apr 2023                         AIC         2360.827
Time: 12:50:46                                      BIC         2382.309
Sample:  01-01-1980                                 HQIC        2369.551
       - 07-01-1995
Covariance Type:                                opg

==========================================================================
=======
                 coef    std err          z      P>|z|      [0.025
 0.975]
--------------------------------------------------------------------------
-------
ar.L1         -0.6609      0.242     -2.733      0.006      -1.135
 -0.187
ma.L1         -0.2740      0.200     -1.368      0.171      -0.666
  0.118
ma.L2         -0.8111      0.227     -3.576      0.000      -1.256
 -0.367
ar.S.L12       1.0157      0.012     84.454      0.000       0.992
  1.039
ma.S.L12      -1.3874      0.338     -4.101      0.000      -2.050
 -0.724
ma.S.L24      -0.1461      0.146     -1.001      0.317      -0.432
  0.140
sigma2      5.947e+04   1.84e+04      3.231      0.001    2.34e+04    9
.56e+04
==========================================================================
============
Ljung-Box (L1) (Q):                   0.00   Jarque-Bera (JB):
       27.47
Prob(Q):                              0.96   Prob(JB):
        0.00
Heteroskedasticity (H):               1.03   Skew:
        0.52
Prob(H) (two-sided):                  0.93   Kurtosis:
        4.76
==========================================================================
============

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

TABLE 34 SARIMA MODEL SUMMARY

Plotting the assumption of Models

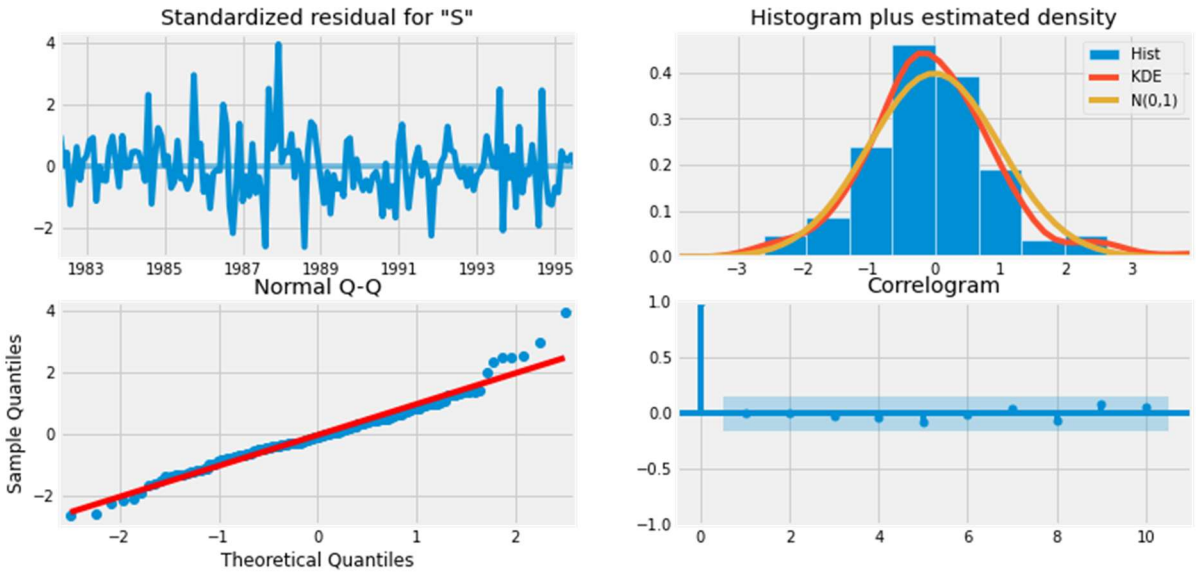


FIGURE 24 ASSUMPTION OF SARIMA

Inference:

In this case, our model diagnostics suggests that the model residuals are normally distributed based on the following:

1.The KDE plot of the residuals on the top right is almost similar with the normal distribution.
2.The Q-Q-plot on the bottom left shows that the ordered distribution of residuals (blue dots) follows the linear trend of the samples taken from a standard normal distribution with N(0, 1). Again, this is a strong indication that the residuals are normally distributed.
3.The residuals over time (top left plot) don't display any obvious seasonality and appear to be white noise. This is confirmed by the autocorrelation (i.e. correlogram) plot on the bottom right, which shows that the time series residuals have low correlation with lagged versions of itself.
4.Those observations coupled with the fact that there are no spikes outside the insignificant zone for both ACF and PACF plots lead us to conclude that that residuals are random with no information or juice in them and our model produces a satisfactory fit that could help us understand our time series data and forecast future values.

It seems that our ARIMA model is working fine.

Evaluate the model on the whole and predict 12 months into the future

Time Series Assumptions Table of Contents

Some of the most common assumptions made for time series are based on the common sense. But always Keep in mind one thing

1.Forecast is done by keeping in mind that the market and the other conditions are not going to change in the future.
2.There will be not any change in the market.
3.But the change is gradual and not a drastic change.
4.Situations like recession in 2008 US market will send the forecasts into a tizzy.
5.Events like demonetization would throw the forecasts into disarray Based on the data available ,
we should not try to forecast for more than a few periods ahead.

Building the most optimum model on the Full Data.

We have used the number of methods to get the model having minimum RMSE.In This on the basis of RSME table we find Triple Exponential Smoothing Modelling is perfect for this solution with the Triple Exponential Smoothing with multiplicative seasonality with the following parameters:

α = 0.11,
β = 0.12 and
γ = 0.460

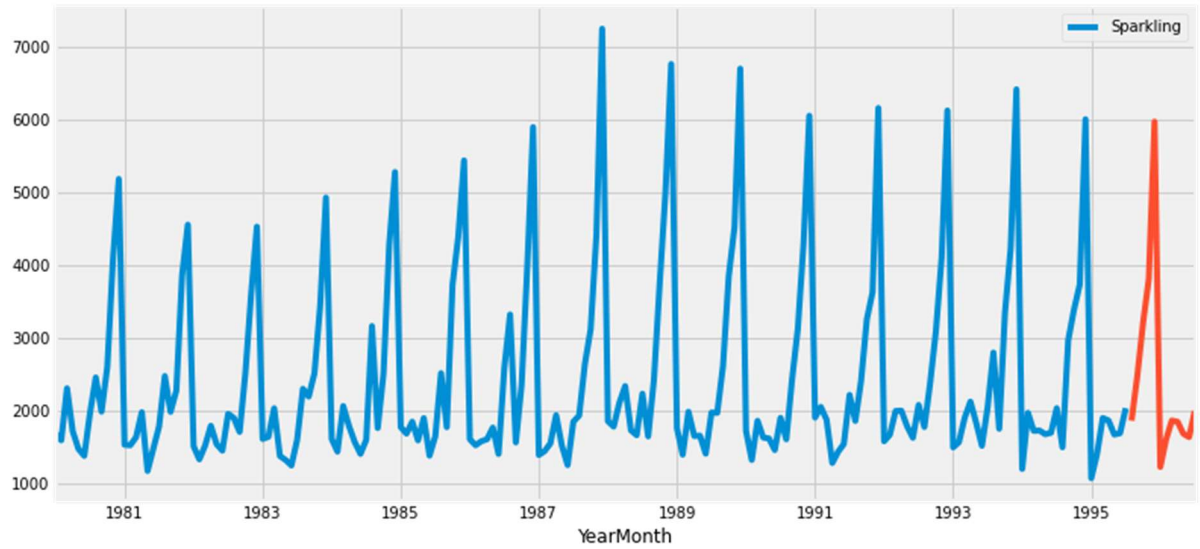


FIGURE 25 TES BEST COMBINATION ON FULL DATA

One assumption that we have made over here while calculating the confidence bands is that the standard deviation of the forecast distribution is almost equal to the residual standard deviation.

RMSE of the TES Full Model 356.9910609869544

	LOWER_CI	PREDICTION	UPPER_CI
1995-08-01	1159.064911	1860.575075	2562.085240
1995-09-01	1768.945346	2470.455510	3171.965675
1995-10-01	2497.912104	3199.422268	3900.932432
1995-11-01	3104.277507	3805.787671	4507.297835
1995-12-01	5264.864501	5966.374666	6667.884830

TABLE 35 CI INTERVAL

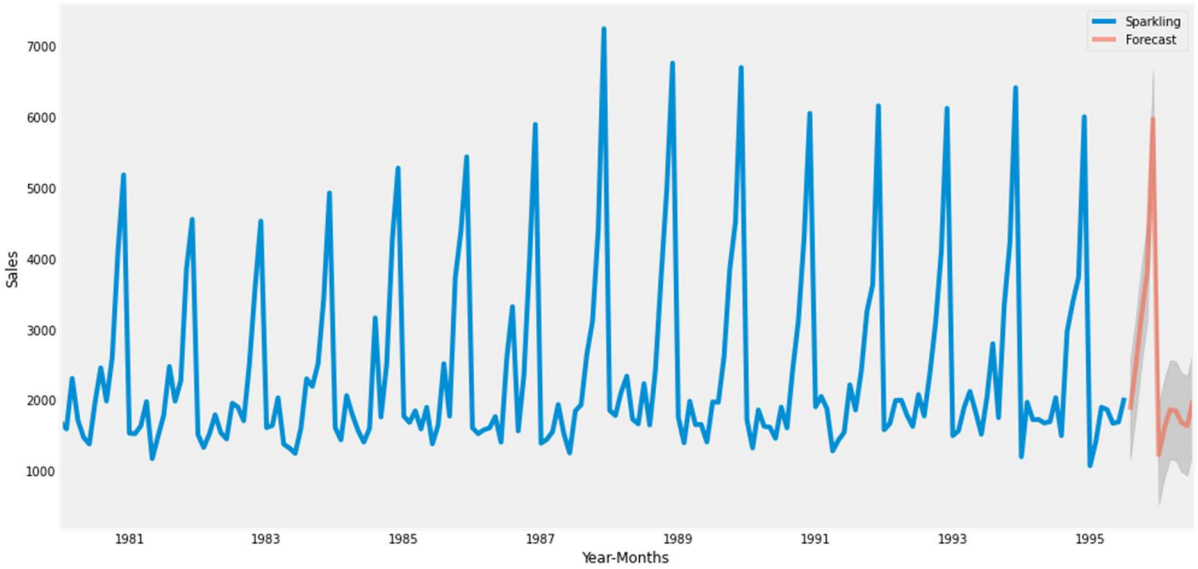


FIGURE 26 FORECAST WITH CI

Inferences:

TES stands as best model to determine a good forecast on the full Sparkling TS dataset.356.99 is the RMSE in on th full dataset. Which is lower than the most of training and test models performed here.
CI Intervals : The lower_CI and upper_CI columns represent the lower and upper bounds of the confidence interval for the prediction column respectively. A confidence interval is a range of values that we can be confident contains the true population parameter with a certain degree of certainty. Our case we have considered the confidence level is 95%. Which is plotted as well.\

So, for example, if we take the first row of your data (1995-08-01), we can say that we are 95% confident that the true value of prediction lies between 1159.064911 and 2562.085240.

RMSE Tabular Summary of Models

	Test RMSE
Alpha=0.70,SES	1338.000861
Alpha=.665,Beta=0.0001:DES	5291.879833
Alpha=0.11,Beta=0.012,Gamma=0.460:TES	378.625883
Alpha=0.111,Beta=0.0493,Gamma=0.362:TES	402.936179
RegressionOnTime	1275.867052
Naive Model	3864.279352
Simple Average	1275.081804
Moving Average	1267.925330
2pointTrailingMovingAverage	3046.976092
4pointTrailingMovingAverage	2021.855880
6pointTrailingMovingAverage	1521.611250
9pointTrailingMovingAverage	1304.618912
ARIMA(2,1,2)	1299.979280
SARIMA(1,1,2)(1,0,2,12)	528.606396
RMSE of the TES Full Model	356.991061

Table 36 RMSE SUMMARY OF MODELS

QN-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.

Sorting the RMSE Values of all models in Descending Order

	Test RMSE
RMSE of the TES Full Model	356.991061
Alpha=0.11,Beta=0.012,Gamma=0.460:TES	378.625883
Alpha=0.111,Beta=0.0493,Gamma=0.362:TES	402.936179
SARIMA(1,1,2)(1,0,2,12)	528.606396
Moving Average	1267.925330
Simple Average	1275.081804
RegressionOnTime	1275.867052
ARIMA(2,1,2)	1299.979280
9pointTrailingMovingAverage	1304.618912
Alpha=0.70,SES	1338.000861
6pointTrailingMovingAverage	1521.611250
4pointTrailingMovingAverage	2021.855880
2pointTrailingMovingAverage	3046.976092
Naive Model	3864.279352
Alpha=.665,Beta=0.0001:DES	5291.879833

TABLE 37 RMSE VALUE IN DESCENDING ORDER

Insights:

1.The TES model on full dataset is having the lowest RMSE(356.99) Value with parameter Alpha=0.11,Beta=0.012,Gamma=0.460.
2.While the highest RMSE(2562.085240) value is outcome of DES Model with parameter Alpha=.665,Beta=0.0001.

QN 9 Please explain and summarise the various steps performed in this project. There should be proper business interpretation and actionable insights present.(for Both Sparkling and Rose TS dataset)¶

1.The three fundamental steps to model a time series are building a model for time series, validating the model and using the model to forecast future values/impute missing values.

2.The first step in time series modelling is to read the data ,account for existing seasons (a recurring pattern over a given period of time) and trends (upward or downward movement in the data). Accounting for these embedded patterns is what we call making the data stationary.

Steps Performed :

a. Collecting the data and cleaning it

b. Preparing Visualization with respect to time vs key feature

c. Normal check of data shape, data Information, Description and plot (Yearly and Monthly),Missing values (For Rose Data missing value correction also)

d. Performed and plot Decomposition by both additive and multiplicative methods and log transformed.

e. Compared the normal plot with log plot

f. identified the patterns in time series data

3.Check for stationary of data and make it stationary by prescribe common method. Once the data is stationary, the next step is modelling to establish a base level forecast. This can be done using various techniques such as exponential smoothing methods as Simple, Double and Triple exponential(moving average),Linear regression, Naive Model, Simple Average, Moving average.

4.For the modelling of ARIMA/SARIMA model in which the parameters are selected using the lowest Akaike Information Criteria (AIC) on the training data. Considered the best param combination to get the lowest RMSE.

a. Data brakeup into Train and test as per desired year

b. All types of Exponential methods performed

c. Tuned the models on hypermeter by Param search

d. hyperparameters:

Alpha: Smoothing factor for the level.

Beta: Smoothing factor for the trend.

Gamma: Smoothing factor for the seasonality.

Trend Type: Additive or multiplicative.

Dampen Type: Additive or multiplicative.

Phi: Damping coefficient.

Seasonality Type: Additive or multiplicative.

Period: Time steps in seasonal period.

e. Performed the Models and Evaluate them individually and plotting

ARIMA and SARIMA model performed on low AIC value

Uses the combination parameter to get the best combination

Improving AR Models by making Time Series stationary through Moving AVerage Forecasts

ARIMA models consist of 3 components: -

AR model: The data is modelled based on past observations.

Integrated component: Whether the data needs to be differenced/transformed.

MA model: Previous forecast errors are incorporated into the model.

for SARIMA, we first need to define a few parameters and a range of values for other parameters to generate

a list of all possible combinations of p, q, d, P, Q, D, s.

we get the SARIMA (p, d, q)(P, D, Q, s) model.

5.The final step in modelling is evaluating model accuracy. This has been done by using various statistical measures root mean squared error (RMSE).Generated a table of RMSE for the whole modelling outcome and sorted out in descending order to get the lowest RMSE value at top.

6.Than Run the best model on test data set than on the whole dataset to predict 12 months into the future (till the end of next year) as desired.

Business Interpretations and Actionable Insights

Business Interpretations of Sparkling

1.In Yearly Box plot there is not fix growth and degrowth trend whereas trend can be observed while in Monthly sales analysis by plotting the data and the Seasonality can be seen clearly.
2. There are sharp increase in sale from month of August every year which touches the peak in December. While the sales peak is in month of December, which might be due to Holidays, some strong festival and new year celebration.
3.We can also observe a bit jump in sales in month of March and April the actual reason not mentioned in data record.
4.Month June can be considered as poor month in year for Sparkling sales view point.
5.As per sales perspective the minimum sales was in year 1983 was less than 2000 while the highest sales are recorded in year 1988 Dec.
6.After 1983 there was increasing trend where as sales has started shooting from 1987 which cannot be accidental.

Business Interpretations of Rose

1.In Yearly Box plot there is degrowth trend whereas trend can be observed while in Monthly sales analysis by plotting the data and the Seasonality can be seen clearly.
2. There are normal increase in sale in month of August every year and hiting the peak in month of December. 3.We can also observe lowest sales in Month January and a little jump in sales in month of March and dip in April the actual reason of fluctuation in sales has not mentioned in data record.
4.The Rose wine is showing continuous downward trend in sales throughout the time period. After year 1988 the sales are in sharp decline trend which is cause of worry and need address immediately. 5.The Rose sales are going down year by year this means the brand has not having a minimum loyal customer base.

Over all Actionable Insights

A. For this we have to go through the factors apart from data prevailing in market, like Competition, Demand Trend, Pricing, Lucrative Schemes.
B.As per data modelling and projection the Sparkling sales will not going to record the degrowth in sales but not showing the healthy growth too.
C. Company have to find the actual factor(s) behind this sales jump from 1987 and sales dip of June 1983.
D. Company should analyse the physical market conditions of 1987 to 1988, because that was neither their entrance period nor their stage of wrapping up.

E. Market survey and competitor pricing or other relevant lever has to find to use to incorporate in planning. These are not insight of analysis but there was a trend for two year which should be analysed or added for further forecast.
F .In year 1988 both brands(Sparkling and Rose) has generated good number of sales, but after 1988 the Rose decline and in case of Sparkling after 1990 sales has showing the good increasing trend.
G. After analysis of Both the brand it can be infer as company is supporting more to brand Sparkling more than Rose. Which may be one of reason for continuous downward trend.
H. Need to analyse the not picking of seasonal sales from Aug to Dec in every year of Rose as per Sparkling pattern.
I. The forecast has been shared for the both brand on the basis of past data modelling. Best Model suggesting that next 12-month sales are on growth pattern. Which would be helpful analyse the other physical factors also for further growth of both the Brands.

Note: ***The Jupyter notebook has been attached here which contains the all type of calculations, process and libraries formulas used, Assumptions considered, Verification, Model evaluation, RMSE generation and Model comparison, AIC value extraction. Model performance. On these vary basis the Inferences has been made to translate the data into actions***.

----------THE END----------