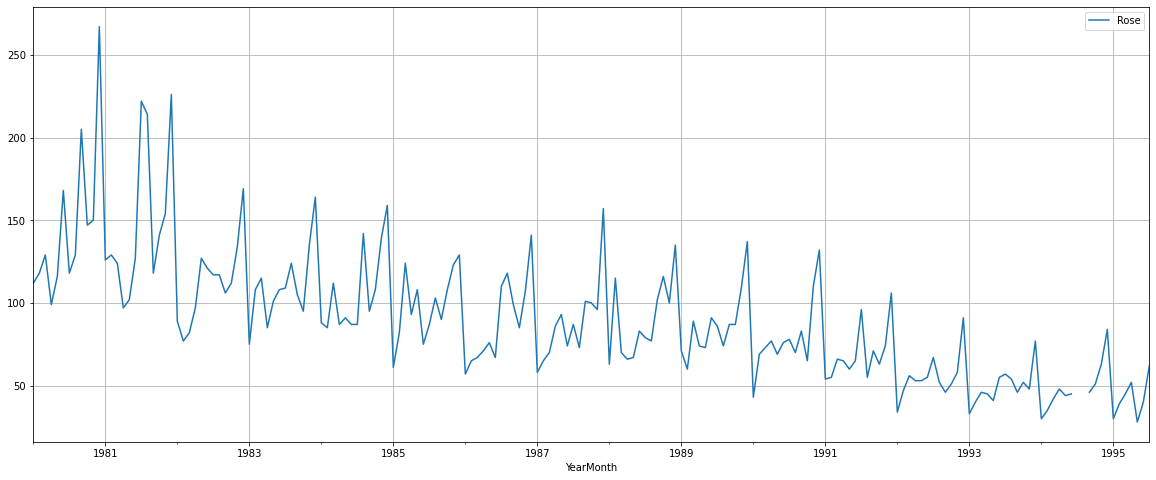
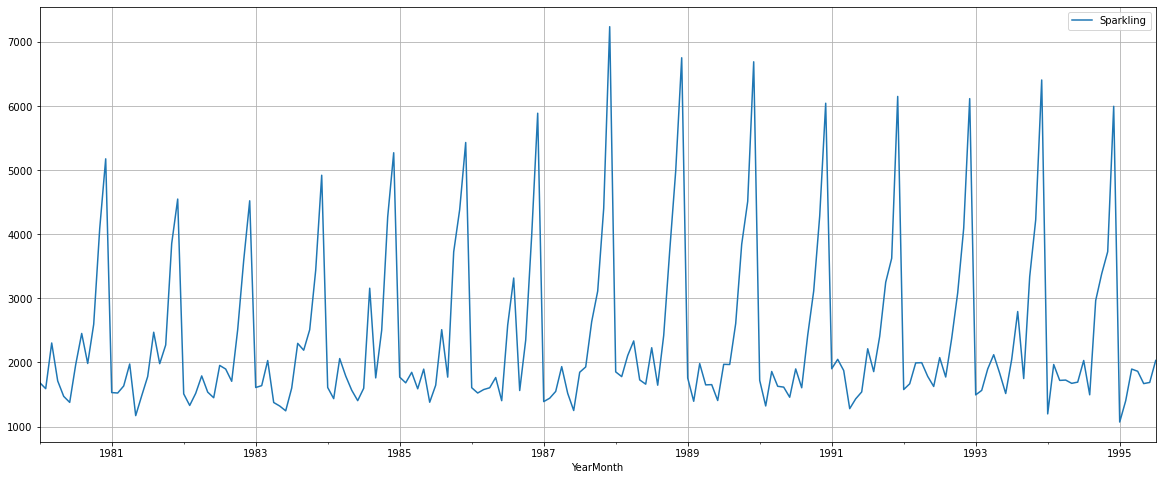
1. **Read the data as an appropriate Time Series data and plot the data.**

**Below is the plot for Rose dataset**

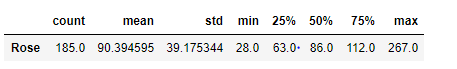


Below is the plot for Sparkling dataset



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

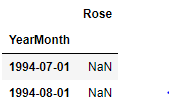
Below is the five number summary for the rose data set.



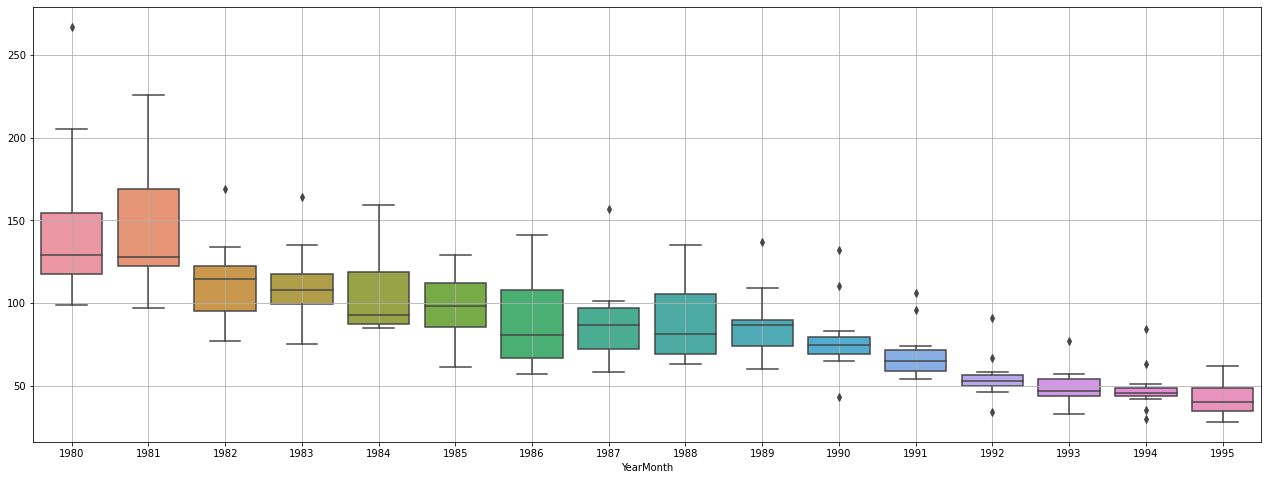
Below is the five number summary for the sparkling data set.



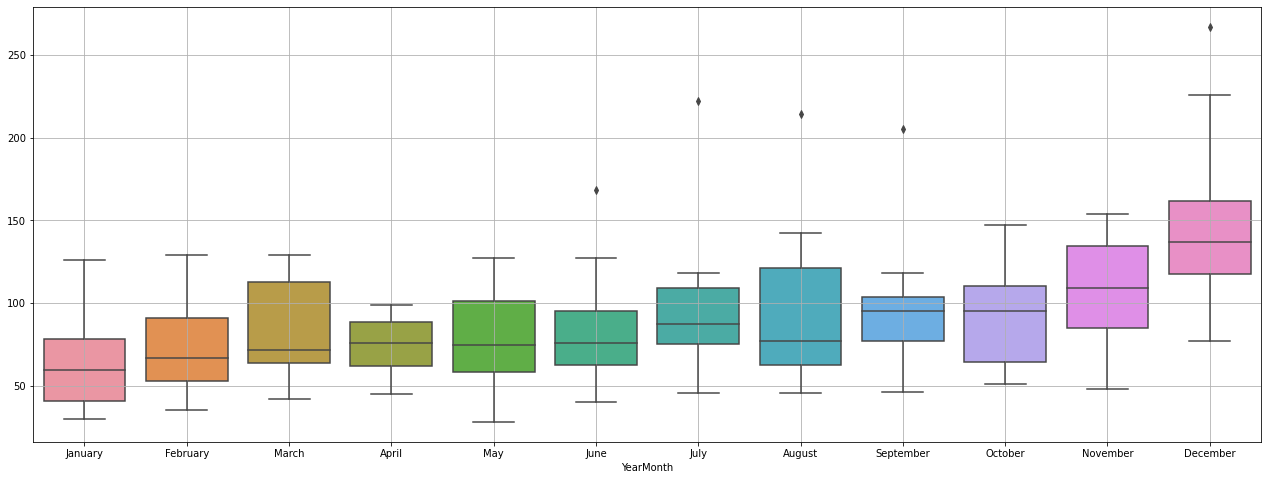
Rose data sets has two months with null information as below based on the null check. Sparkling data set is free of null values.



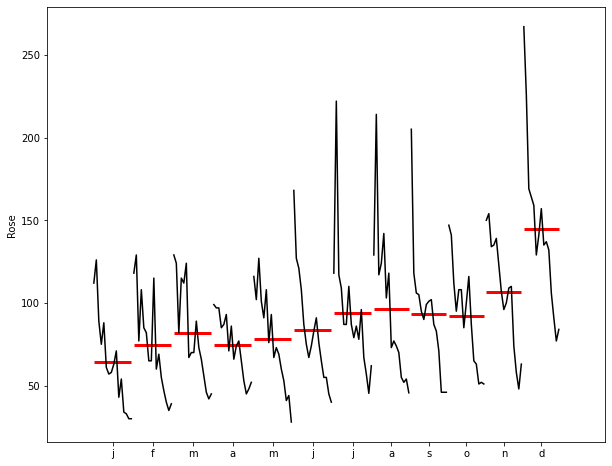
Below is the box plot depicting spread of sales for Rose wine across years for the entire data set. This depicts presence of trend in the Rose wine sales and there is decreasing trend in the sales for Rose wine across the years. Also depicts that later years has had comparatively more outliers in the sales compared to earlier years.



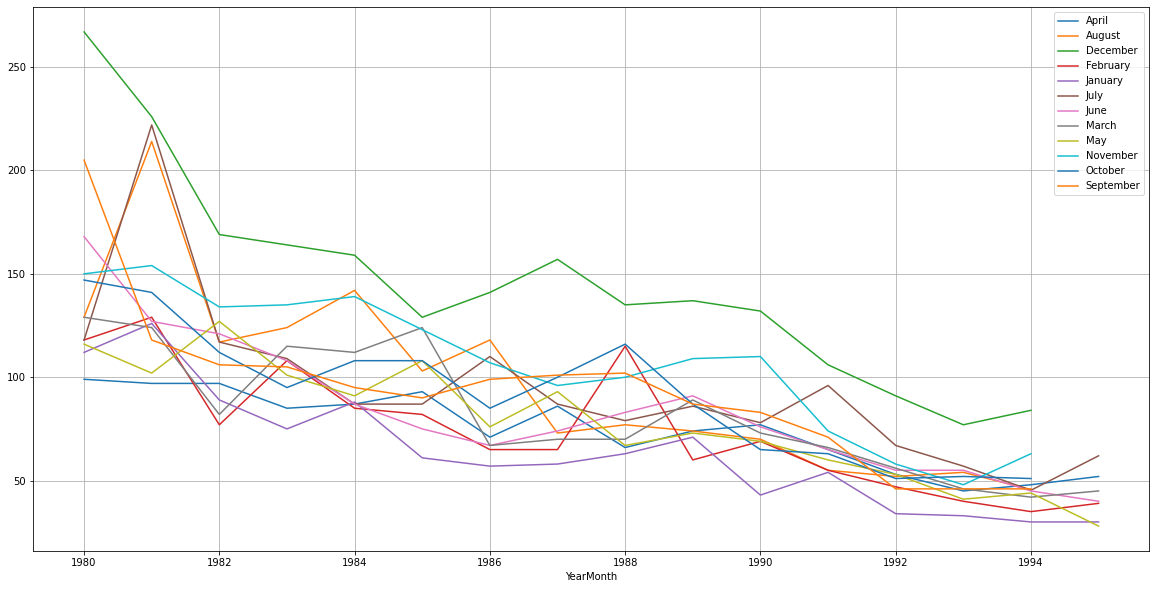
Below is the box plot depicting spread of sales for Rose wine by month across the years for the entire data set. It can be noticed that December has had the highest sales looking compared to other months. Also very minimal outliers noticed during the second half of the year generally.



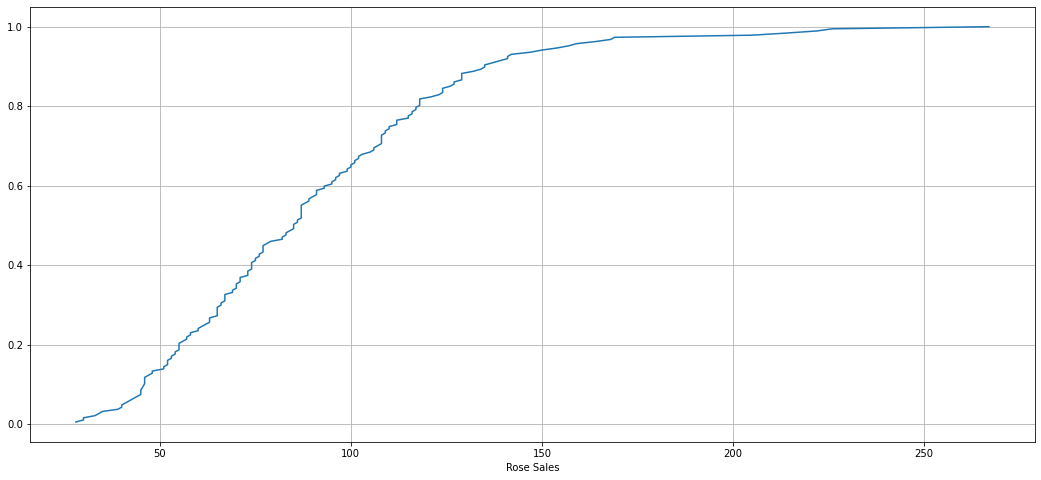
The below plot depicts the spread of Rose wine sales by each month putting all the years together. This also reveals December had a wider range of transaction compared to other months with highest mean sales too.



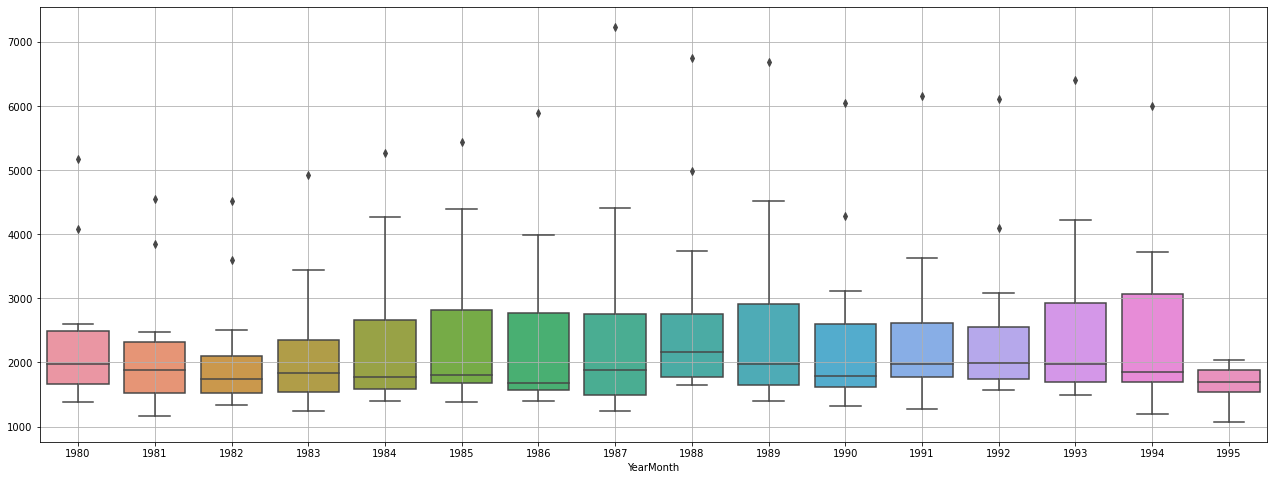
The below plot depicts the trend of monthly sales across years for Rose wine. This depicts that while the Rose wine sales are decreasing across years, the December month has had the peak sales during this period and January has the lowest.



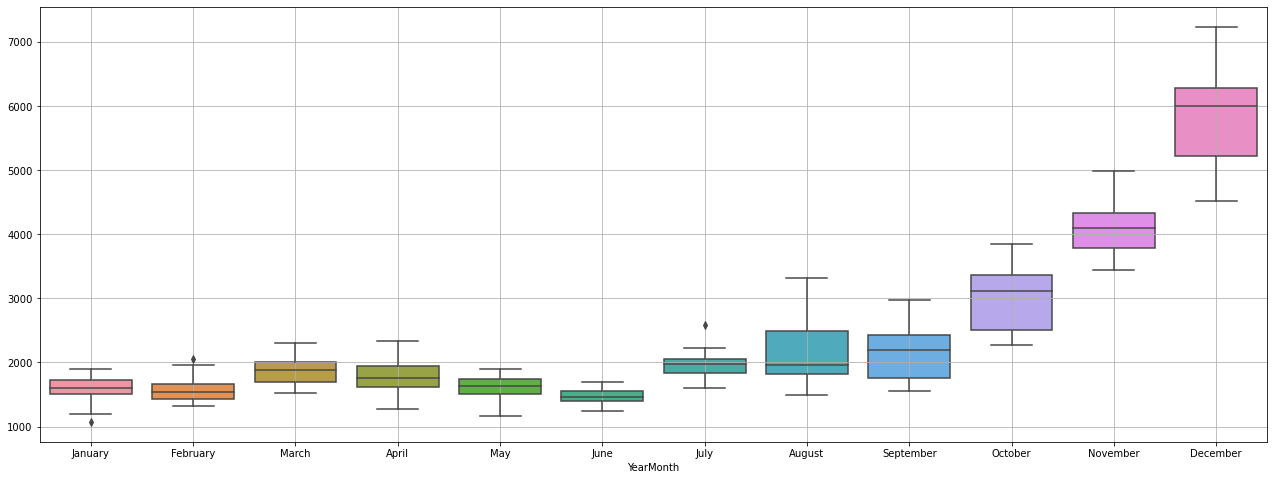
Below plot depicts the empirical cumulative distribution of the sales for Rose wine sales.



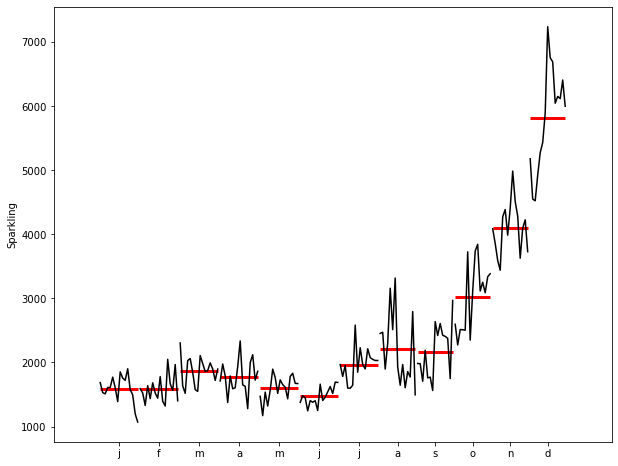
Below plot depicts the pattern of sales across years for the sparkling wine. This depicts that the sales for the sparkling sales across years more or less tends to approximately same and there isn’t a significant trend. However in this case outliers are spread across almost all years except 1995.



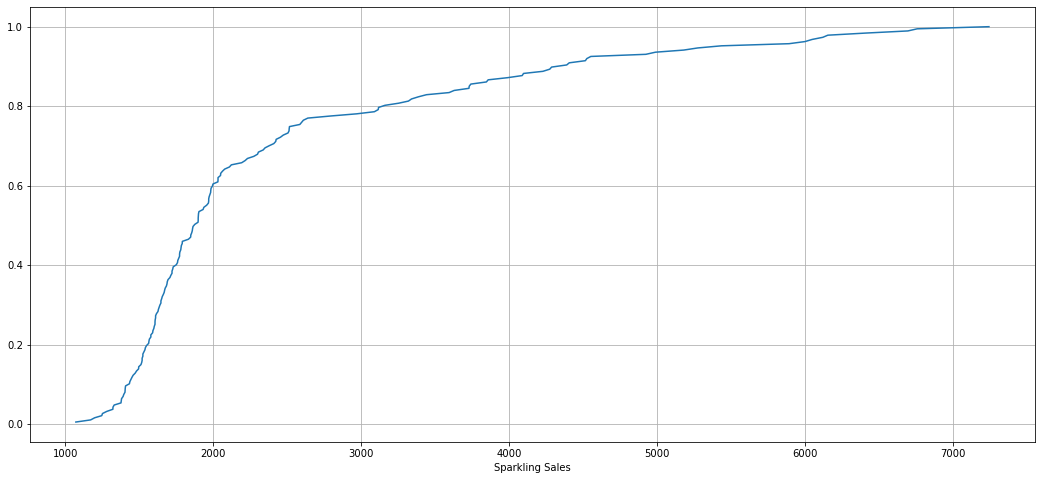
Below plot depicts sales pattern across months depicts an increasing trend towards the later half of the year with December showing highest sales.



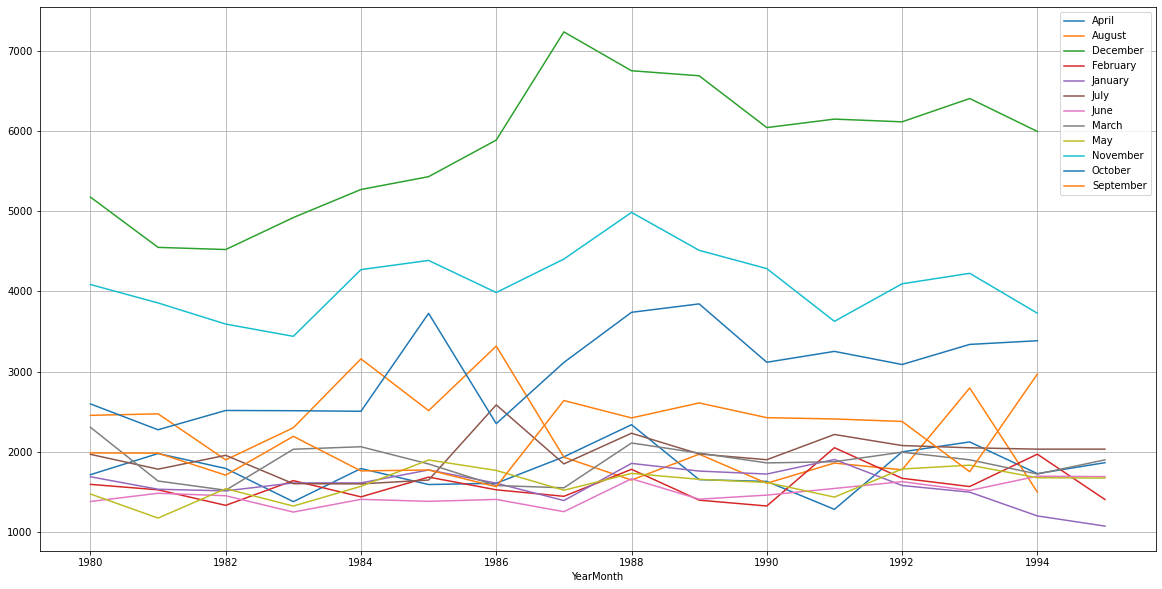
The below plot depicts the spread of sparkling wine sales by each month putting all the years together. Here also its noticed that December has had wider range of sales comparatively than the other months.



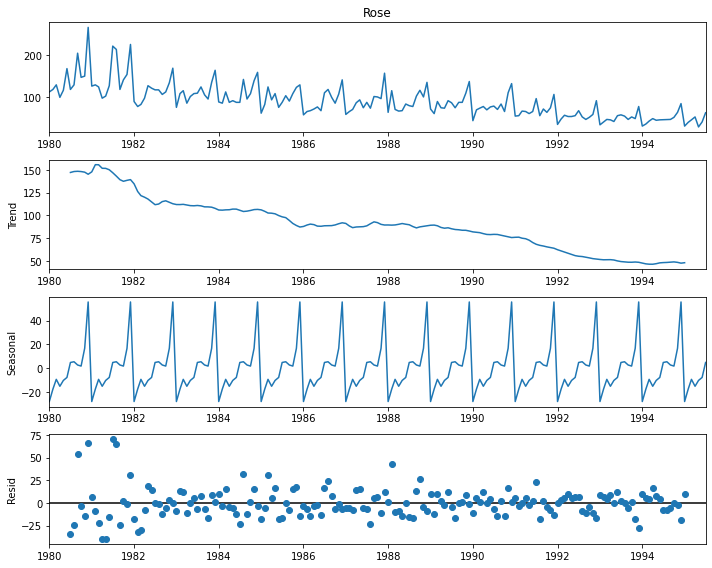
Below depicts the empirical cumulative distribution of sparkling sales



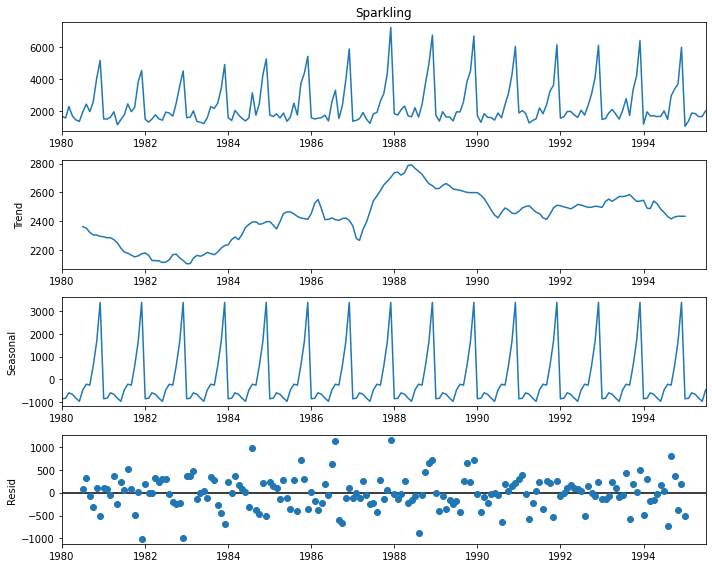
The below plot depicts the trend of monthly sales across years for Sparkling wine. The plot depicts trend across the months with December showing the highest sales across all the years while there are many months cluttering around the lower side of the sales trend such as January, February, July.



Below is the seasonal decomposition of the Rose data set depicting level, trend, seasonality and residual.

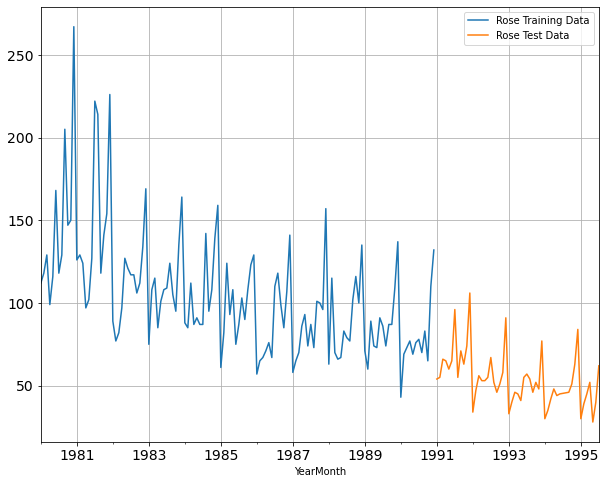


Below is the seasonal decomposition of the sparkling data set depicting level, trend, seasonality and residual.

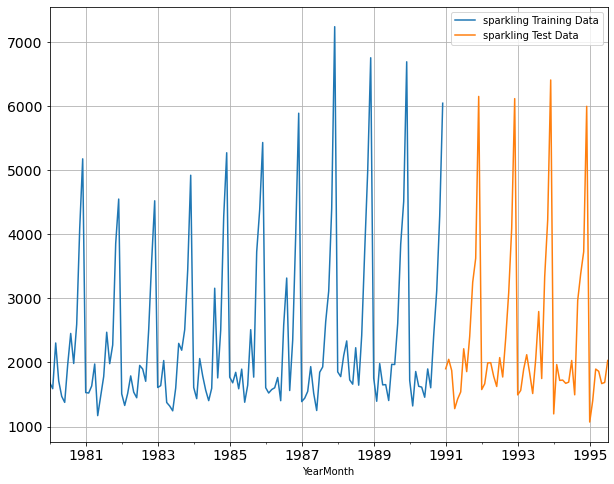


1. **Split the data into training and test. The test data should start in 1991.**

Please find below the plot of training and testing data set for the Rose wine after the data prior to 1991 has been split into the training while from 1991 and beyond splits into test data.



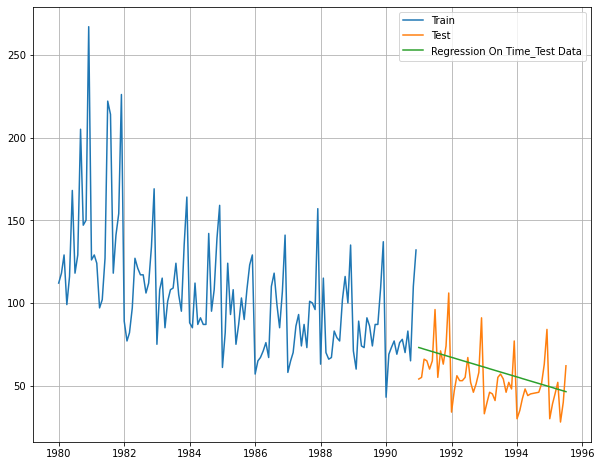
Please find below the plot of training and testing data set for the Sparkling wine after the data prior to 1991 has been split into the training while from 1991 and beyond splits into test data.



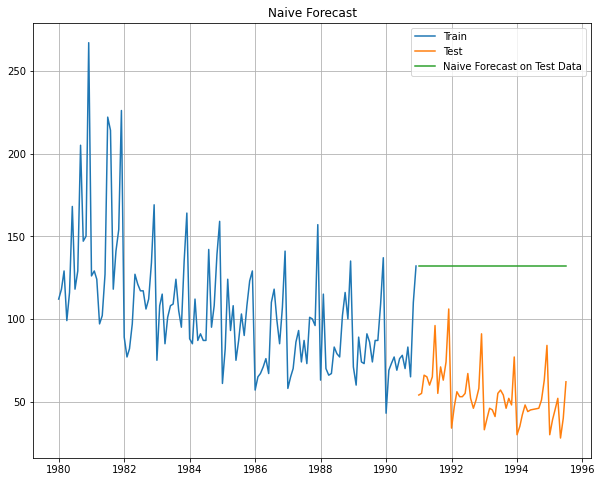
1. **Build various exponential smoothing models on the training data and evaluate the model using RMSE on the test data. Other models such as regression,naïve forecast models, simple average models etc. should also be built on the training data and check the performance on the test data using RMSE.**

**MODELS FOR ROSE WINE DATA SET**

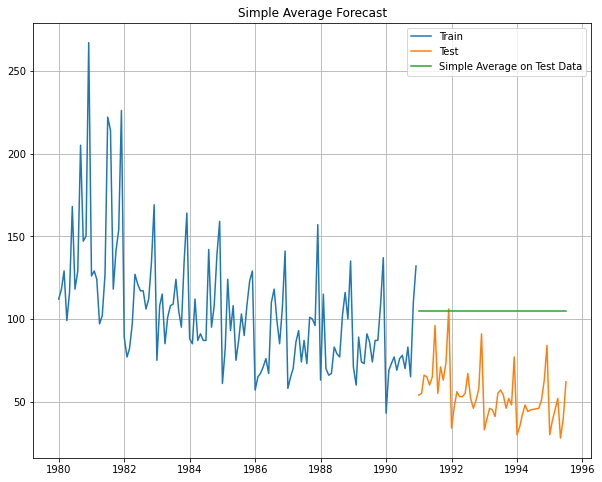
Below plot from Linear Regression model depicts the prediction on the test data overlaps with the trend from the actuals however being smoothened to the straight line with just the trend without any seasonality.



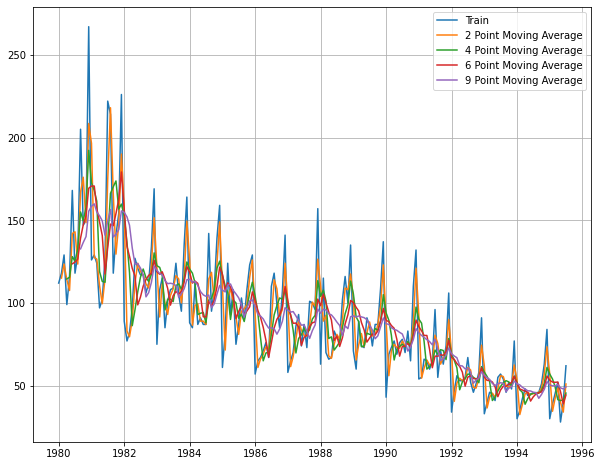
Below plot from Naive model depicts the prediction on the test data being away from the actuals and being constant throughout while having values higher than the originals.



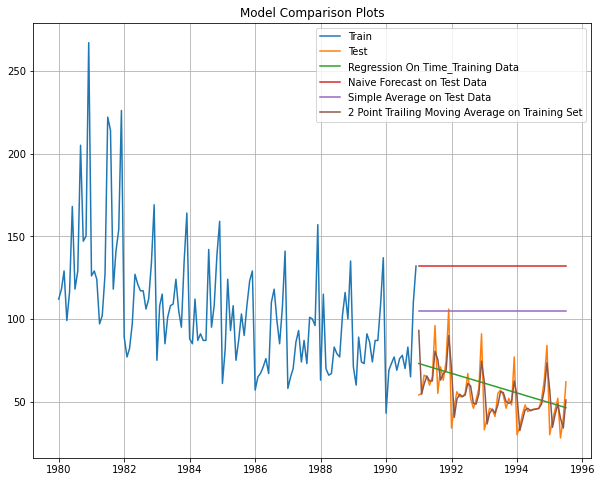
Below plot from simple Average model depicts the prediction on the test data being away from the actuals and being constant throughout while having values higher than the originals.



Plot below depicting the outcome of forecast against the test data for the moving average model with 2 point moving average showing up more closer to the original values however higher point moving average gets smoothened more and more as the points increase.



Below chart provides comparative chart of prediction trend after adopting respective straight forward models against the time series with 2 point moving average model catching up much closer to the trend and the seasonality



We can now move to 3 types of smoothing model across single, double and triple exponential smoothing models.

Exponential smoothing models apply weighted averages of past observations with weights decaying as observations gets older. Especially recent observations has significant weightage. Based on the smoothing models one or more parameters control how weights decay and each of these parameters will have its values ranging from 0 to 1.

Below are the comparative error measures (root mean squared error) across the above models with moving average 2 point being the best of all. However considering the objective of taking the model application to the dataset beyond this to consider exponential smoothing and ARIMA/SARIMA models we are to compare the performance of other models alone for the given data sets for Rose and Sparkling wines towards narrowing down the best performing model among those based on the RMSE values and predict next 12 months of sales for each of the wines accordingly. Below is the just a summary of performance by models so far worked upon with the stated observation above.

|  |  |
| --- | --- |
| **Model** | **RMSE** |
| **Linear Regression** | 15.61187 |
| **Naive Model** | 79.71877 |
| **Simple Average** | 53.46057 |
| **Moving Average - rolling 2** | 11.52928 |
| **Movinbg Average - rolling 4** | 14.4514 |
| **Moving Average - rolling 6** | 14.56633 |
| **Moving Average - rolling 9** | 14.72763 |

Single exponential smoothing model:

If the time series neither has a noticeable trend or seasonality it is understood to retain only the level and its related parameter that controls the decay of weight across the observations in time series referred to as alpha will range from 0 to 1 with 1 meaning the forecast is closely following the actuals whereas on the converse 0 refers to forecasts are farther from the actual observations with the forecast being more smooth. Please note level refers to the local mean. Alpha smoothens the level.

After auto fitment of single exponential smoothening model against the training data for Rose wine dataset below are the parameters that could be derived that depicts the smoothening level.

{'smoothing\_level': 0.09874933517484011,

'smoothing\_trend': nan,

'smoothing\_seasonal': nan,

'damping\_trend': nan,

'initial\_level': 134.38703609891138,

'initial\_trend': nan,

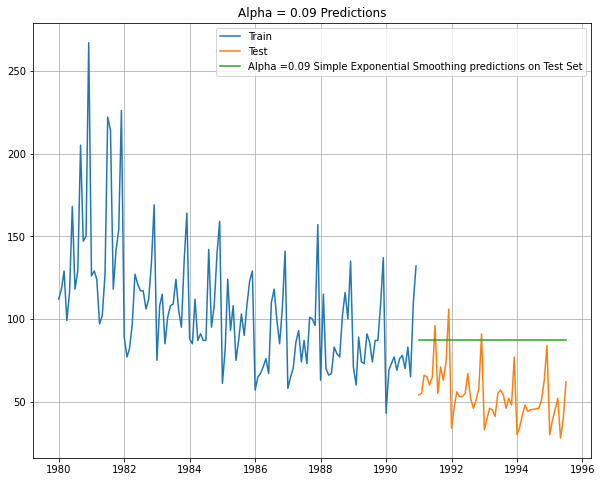
'initial\_seasons': array([], dtype=float64),

'use\_boxcox': False,

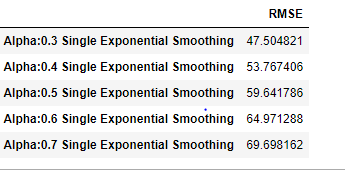
'lamda': None,

'remove\_bias': False}

Below plot depicts the results of the model on the test data for the auto fitted alpha value of 0.09 for the single exponential smoothing. The respective RMSE value is 36.79.



Also please note below are the manual fitment done for this SES model across range of values for the alpha and it has been noticed that their RMSE value is much more than the auto fitted models resulting in autofitted alpha value being considered for this model. The below list depicts top 5 lowest RMSE values among the permutations and combinations across the parameter range.

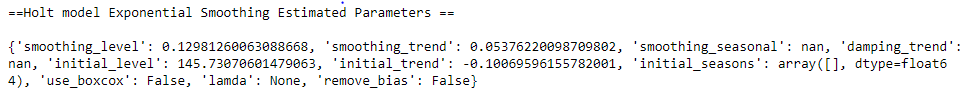


Double exponential smoothing model:

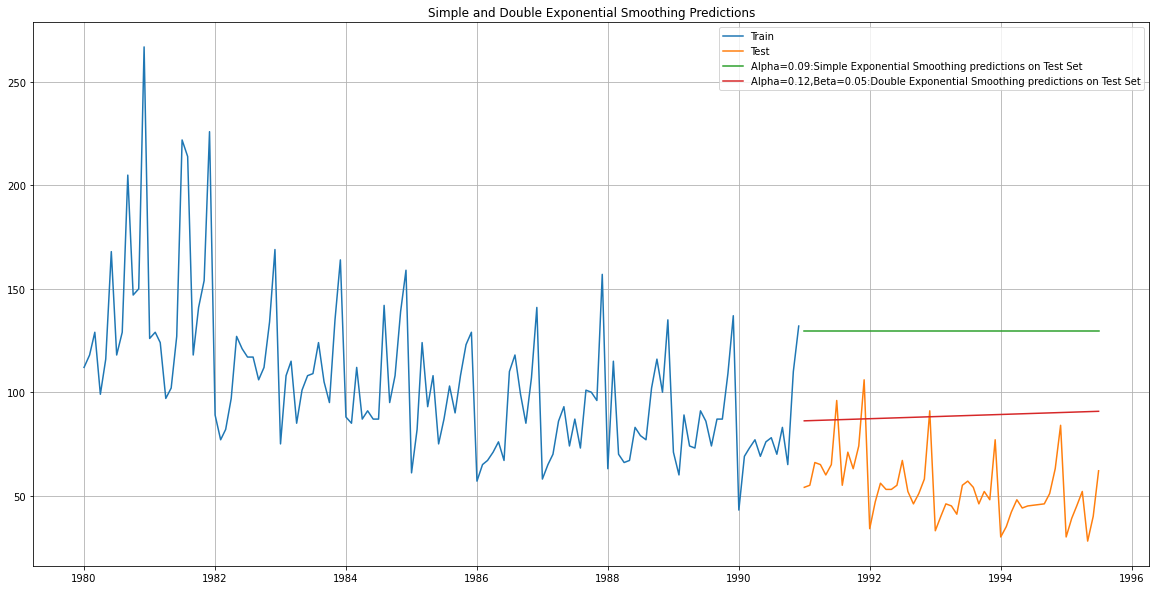
This is a further extension to single exponential smoothing model and applies to datasets there is a trend but not seasonality. Hence both level and trend are considered towards weighing the observations from the time series. Apart from the parameter alpha for the level , this introduces beta parameter for the trend. Beta smoothens the trend.

This model is also known as Holts model.

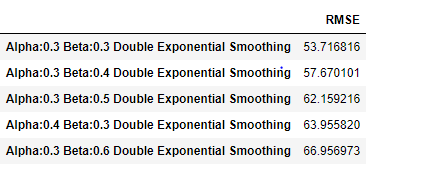
After auto fitment of double exponential smoothening model against the training data for Rose wine dataset below are the parameters that could be derived that depicts the smoothening level and the trend. Accordingly alpha value is 0.12 and the beta value is 0.05. RMSE value for the auto fitted model is 38.28.



Below plot depicts the results of the model on the test data for the auto fitted alpha value of 0.12, beta value of 0.05 for the double exponential smoothing.



Also please note below are the manual fitment done for this DES model across range of values for the alpha and it has been noticed that their RMSE value is much more than the auto fitted models resulting in autofitted alpha and gamma value mentioned above being considered for this model. The below list depicts top 5 lowest RMSE values among the permutations and combinations across the parameter range.

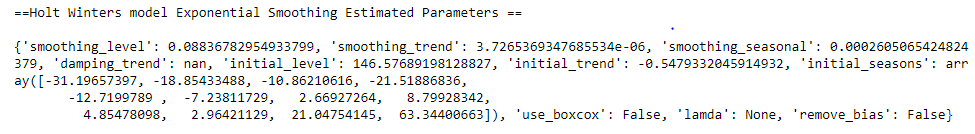


Triple exponential or Holt Winters model

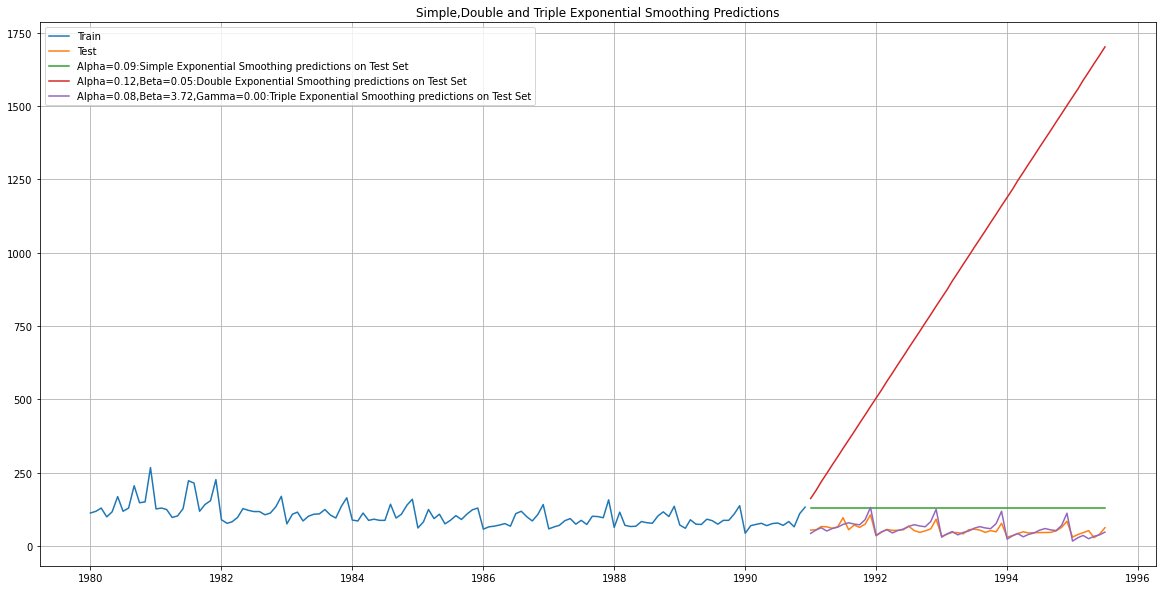
Apart from level (alpha), trend (beta) this model applies to the data with seasonality and hence introduces new parameter called gamma. Gamma smoothens the seasonality.

Beta value of 0 refers to insignificant year over year movement with 1 being the converse. A higher value of gamma attributes most of the data fluctuations in the observations across time series to seasonality.

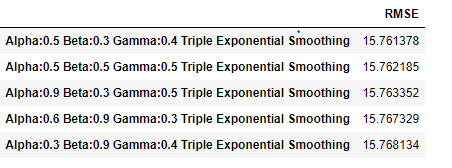
After auto fitment of triple exponential smoothening model against the training data for Rose wine dataset below are the parameters that could be derived that depicts the smoothening level, trend and the seasonality. Accordingly alpha value is 0.08, beta value is 3.72 and the gamma value is 0.00. RMSE value for the auto fitted model is 14.28.



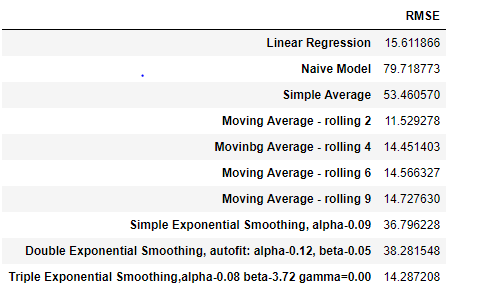
Below plot depicts the results of the model on the test data for the auto fitted alpha value of 0.08, beta value of 3.72 and the gamma value of 0.00 for the triple exponential smoothing.



Also please note below are the manual fitment done for this TES model across range of values for the alpha and it has been noticed that their RMSE value is much more than the auto fitted models resulting in autofitted alpha, beta and gamma values mentioned above being considered for this model. The below list depicts top 5 lowest RMSE values among the permutations and combinations across the parameter range.

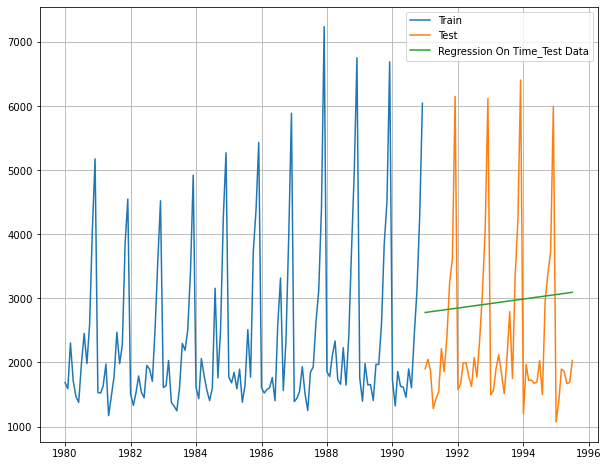


Below is the summary of RMSE values across all the models run so far and within the scope of models considered for assignment triple exponential model has performed much better with lowest RMSE value. Please note that the moving average model is not considered within the scope of models for arriving at better alternates,

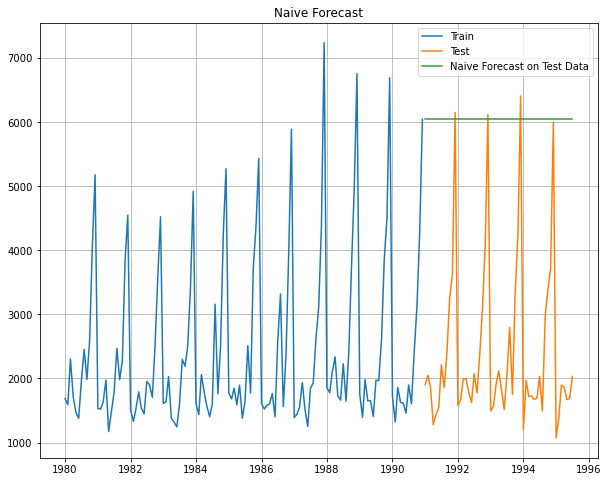


**MODELS FOR SPARKLING WINE DATA SET**

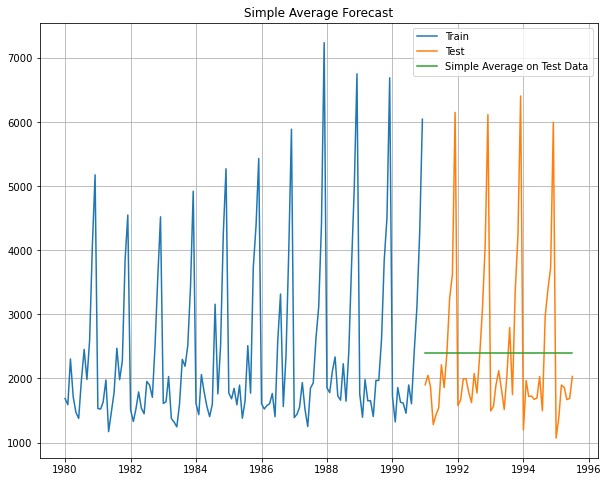
Below plot from Linear Regression model depicts the prediction on the test data overlaps with the trend from the actuals however being smoothened to the straight line with just the trend without any seasonality.

****

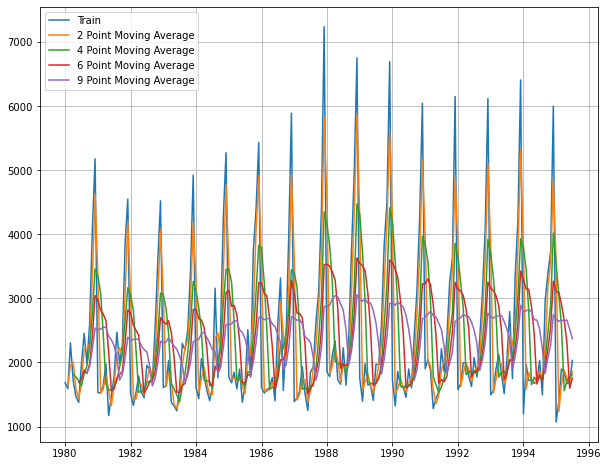
Below plot from Naive model depicts the prediction on the test data overlapping with the peaks of the actuals and being constant throughout.

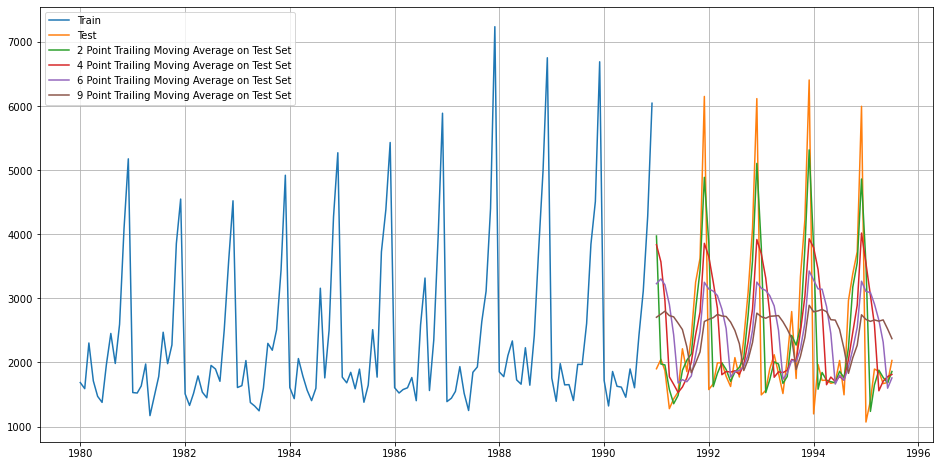


Below plot from simple Average model depicts the prediction on the test data overlapping by dissecting the spikes in the actuals and being constant throughout.

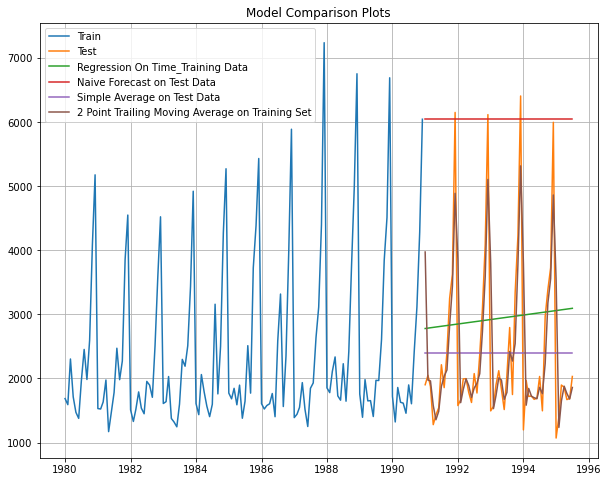


Plot below depicting the outcome of forecast against the test data for the moving average model with 2 point moving average showing up more closer to the original values however higher point moving average gets smoothened more and more as the points increase.



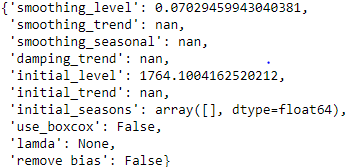


Below chart provides comparative chart of prediction trend after adopting respective straight forward models against the time series with 2 point moving average model catching up much closer to the trend and the seasonality

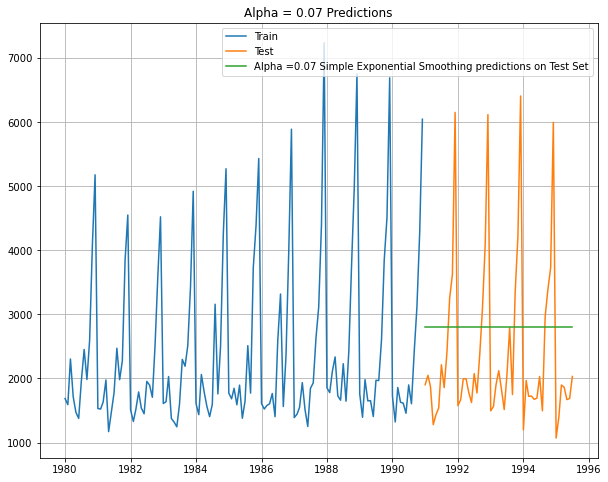


Single exponential smoothing:

After auto fitment of single exponential smoothening model against the training data for Sparkling wine dataset below are the parameters that could be derived that depicts the smoothening level. Accordingly the alpha parameter is 0.07 indicating the weightage of observations for the level.



Below plot depicts the results of the model on the test data for the auto fitted alpha value of 0.07 for the single exponential smoothing. The respective RMSE value is 1338.01

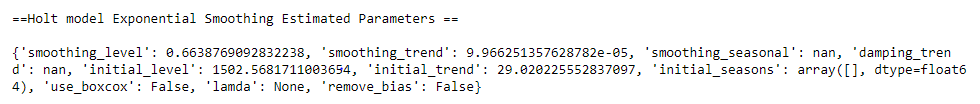


Also please note below are the manual fitment done for this SES model across range of values for the alpha and it has been noticed that their RMSE value is much more than the auto fitted models resulting in autofitted alpha value being considered for this model. The below list depicts top 5 lowest RMSE values among the permutations and combinations across the parameter range.

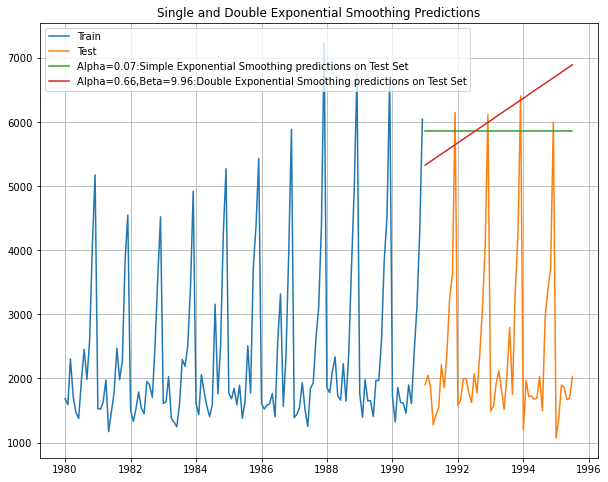


Double exponential smoothing(DES)/ Holt model:

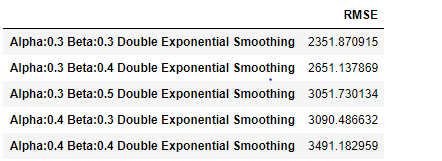
After auto fitment of double exponential smoothening model against the training data for Sparkling wine dataset below are the parameters that could be derived that depicts the smoothening level and the trend. Accordingly alpha value is 0.66 and the beta value is 9.96. RMSE value for the auto fitted model is 3949.99.



Below plot depicts the results of the model on the test data for the auto fitted alpha value of 0.66 and beta value of 9.95 for the double exponential smoothing. The respective RMSE value is 3949.99.

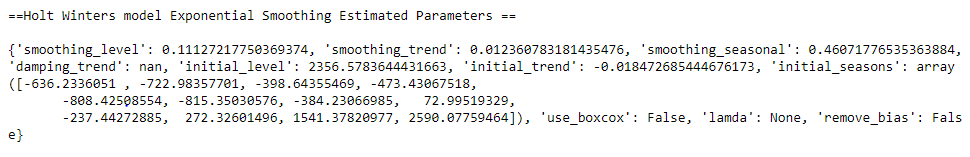


Also please note below are the manual fitment done for this DES model across range of parameter values for the alpha and the beta parameter and it has been noticed that their RMSE value is much more than the auto fitted models resulting in autofitted alpha value being considered for this model. The below list depicts top 5 lowest RMSE values among the permutations and combinations across the parameter range.

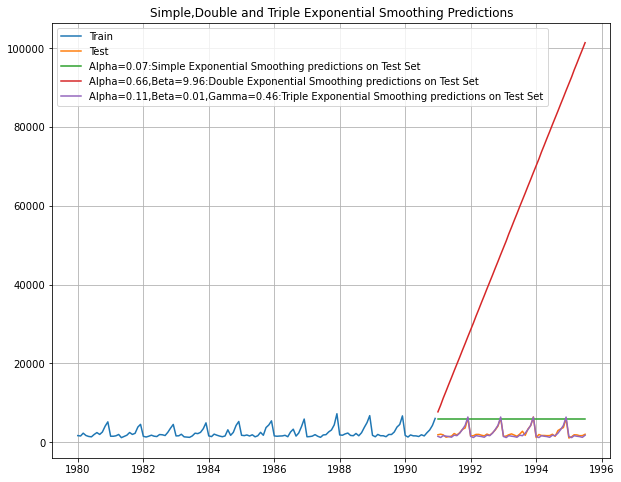


Triple exponential or Holt winters model:

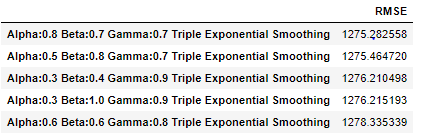
After auto fitment of triple exponential smoothening model against the training data for Sparkling wine dataset below are the parameters that could be derived that depicts the smoothening level, trend and the seasonality. Accordingly alpha value is 0.11, beta value is 0.01 and the gamma value is 0.46. RMSE value for the auto fitted model is 378.62.



Below plot depicts the results of the model on the test data for the auto fitted alpha value of 0.11, beta value of 0.01 and the gamma value of 0.46 for the triple exponential smoothing.



Also please note below are the manual fitment done for this TES model across range of parameter values for the alpha and the beta parameter and it has been noticed that their RMSE value is much more than the auto fitted models resulting in autofitted alpha value being considered for this model. The below list depicts top 5 lowest RMSE values among the permutations and combinations across the parameter range.



Based on the above comparison of smoothing models it is evident that triple exponential smoothing otherwise Holt winters model is able to forecast values much closer to the actuals after considering level, trend and the seasonality parameters based on the auto fitted parameter values depicted above.

1. **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.**

ARIMA model requires the time series to be stationary towards being able to proceed with the model building. Data is defined to be stationary if its mean and variance are constant over period of time and the correlation between the two time periods depends only on the distance or lag between the two period.

Hence a formal stationarity (hypothesis) test needs to be applied on the time series data to check whether it follows stationary process for which the hypothesis definition given as below:

Null Hypothesis: Time series is non stationary

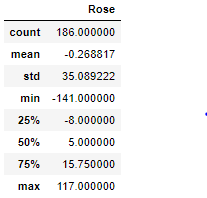
Alternate hypothesis: Time series is stationary

Based on the hypothesis done for the Rose Wine time series, it is understood that the p\_value for the test supports null hypothesis and hence the series is not stationary. Please find below the test statistics and the p value for the same.

DF test statistic is -1.877

DF test p-value is 0.3431

Please find below the 5 number summary after applying one level of difference to the data set.



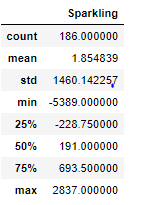
Accordingly after applying the difference of

Based on the hypothesis done for the Sparkling Wine time series, it is understood that the p\_value for the test supports null hypothesis and hence the series is not stationary. Please find below the test statistics and the p value for the same.

DF test statistic is -1.360

DF test p-value is 0.6011

Please find below the 5 number summary after applying one of level of difference to the Sparkling data set.



Also during ARIMA model building exercise, differencing parameter could be derived using auto ARIMA to be the required level of differencing to make the series stationary towards optimal AIC.

1. **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 and SARIMA for Sparkling wine dataset.**

Auto correlation function (ACF): ACF of order (p) measures the strength of dependency of current observations on past observations whereas PACF provides correlation value between current and (k) lagged series by removing the influence of all other observations that exists in between. ACF and PACF acts together to be considered for identification of order of autoregression.

In general, ARIMA models are defined by 3 parameters

p: No of autoregressive terms

𝑑: No of differencing to stationarize the series

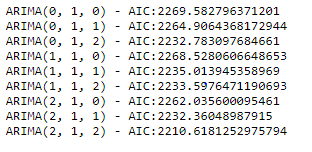
𝑞: No of moving average terms

Parameter p can be utilized towards representing current value of a variable in the time series as a linear function of its past values through auto regression (AR) process. PACF can be used for identifying the value of p.

Parameter q can be utilized towards representing current value of the series as a function of past forecast errors through moving average (MA) model. ACF can be used to identify the value of q.

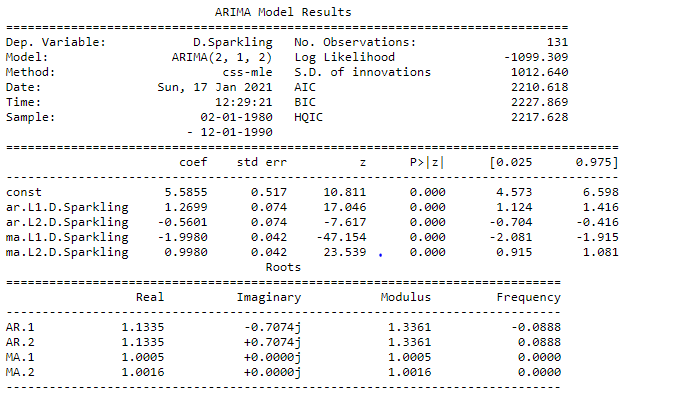
ARIMA model is an advance version of ARMA model where 𝑑 > 0 indicates that the original series is non-stationary and d differencing is required to make is stationary.

In order to arrive optimal combination of p, d, q to apply to the ARIMA model what we do here is to build auto ARIMA process towards narrowing down the best values for these parameters by comparing AIC value of the ARIMA model output for each p,d,q combination accordingly to narrow down the combination with the lowest AIC score. Based on the range of p and q to iterate through 0, 1 and 2 along with d (differencing parameter for applying stationarity) to iterate for 0 and 1 for various combination of p,d,q across permutations and combinations, we arrived at following AIC scores after applying ARIMA model.



Accordingly ARIMA model with the order of (2,1,2) has performed the best with lowest AIC score. Lower the AIC score better the model performance.

Below is the statistical summary of the ARIMA result for the narrowed down order of (2,1,2).



Accordingly based on the predictions applied on the test series based on the above model the resulting RMSE (root mean square error value) is 1374.6492.

Also, now lets look into SARIMA model.

On top of ARIMA model, additionally we could apply seasonal adjustments to process SARIMA model.

The most general form of seasonal ARIMA is ,

𝐴𝑅𝐼𝑀𝐴(𝑝,𝑑,𝑞)∗𝐴𝑅𝐼𝑀𝐴(𝑃,𝐷,𝑄)[m], where P, D, Q are defined as seasonal AR component, seasonal difference and seasonal MA component respectively. And, ‘𝑚’ represents the frequency (time interval) at which the data is observed.

In this case we are going to apply 6 and 12 for SARIMA to compare model outcome. We use the same p, d, q ranges as before for ARIMA while retaining P=p and Q=q with D as 0 for the seasonal part.

Accordingly below are the lists of auto SARIMA model outcome along with its AIC score for m=6.

SARIMA(0, 1, 0)x(0, 0, 0, 6)6 - AIC:2251.3597196862966

SARIMA(0, 1, 0)x(0, 0, 1, 6)6 - AIC:2152.3780761716284

SARIMA(0, 1, 0)x(0, 0, 2, 6)6 - AIC:1955.6355536890933

SARIMA(0, 1, 0)x(1, 0, 0, 6)6 - AIC:2164.4097581959904

SARIMA(0, 1, 0)x(1, 0, 1, 6)6 - AIC:2079.559984442563

SARIMA(0, 1, 0)x(1, 0, 2, 6)6 - AIC:1926.9360111185642

SARIMA(0, 1, 0)x(2, 0, 0, 6)6 - AIC:1839.4012986872267

SARIMA(0, 1, 0)x(2, 0, 1, 6)6 - AIC:1841.199361751051

SARIMA(0, 1, 0)x(2, 0, 2, 6)6 - AIC:1810.9177805657487

SARIMA(0, 1, 1)x(0, 0, 0, 6)6 - AIC:2230.1629078505825

SARIMA(0, 1, 1)x(0, 0, 1, 6)6 - AIC:2130.5652859082847

SARIMA(0, 1, 1)x(0, 0, 2, 6)6 - AIC:1918.1876339543767

SARIMA(0, 1, 1)x(1, 0, 0, 6)6 - AIC:2139.573242878454

SARIMA(0, 1, 1)x(1, 0, 1, 6)6 - AIC:2006.5174298135796

SARIMA(0, 1, 1)x(1, 0, 2, 6)6 - AIC:1855.7093274084523

SARIMA(0, 1, 1)x(2, 0, 0, 6)6 - AIC:1798.7885104034895

SARIMA(0, 1, 1)x(2, 0, 1, 6)6 - AIC:1800.77179337265

SARIMA(0, 1, 1)x(2, 0, 2, 6)6 - AIC:1741.7036712072565

SARIMA(0, 1, 2)x(0, 0, 0, 6)6 - AIC:2187.4410101687026

SARIMA(0, 1, 2)x(0, 0, 1, 6)6 - AIC:2087.6843840215897

SARIMA(0, 1, 2)x(0, 0, 2, 6)6 - AIC:1886.1151457125393

SARIMA(0, 1, 2)x(1, 0, 0, 6)6 - AIC:2129.7395689235523

SARIMA(0, 1, 2)x(1, 0, 1, 6)6 - AIC:1988.4215580217822

SARIMA(0, 1, 2)x(1, 0, 2, 6)6 - AIC:1839.6963217060352

SARIMA(0, 1, 2)x(2, 0, 0, 6)6 - AIC:1791.6537079050197

SARIMA(0, 1, 2)x(2, 0, 1, 6)6 - AIC:1793.6190999290054

SARIMA(0, 1, 2)x(2, 0, 2, 6)6 - AIC:1727.8888035990105

SARIMA(1, 1, 0)x(0, 0, 0, 6)6 - AIC:2250.3181267386713

SARIMA(1, 1, 0)x(0, 0, 1, 6)6 - AIC:2151.0782683083426

SARIMA(1, 1, 0)x(0, 0, 2, 6)6 - AIC:1953.3652245477829

SARIMA(1, 1, 0)x(1, 0, 0, 6)6 - AIC:2146.1836648562185

SARIMA(1, 1, 0)x(1, 0, 1, 6)6 - AIC:2073.981368525952

SARIMA(1, 1, 0)x(1, 0, 2, 6)6 - AIC:1917.5889468384437

SARIMA(1, 1, 0)x(2, 0, 0, 6)6 - AIC:1813.2423977989204

SARIMA(1, 1, 0)x(2, 0, 1, 6)6 - AIC:1814.8301602829295

SARIMA(1, 1, 0)x(2, 0, 2, 6)6 - AIC:1791.3715264887628

SARIMA(1, 1, 1)x(0, 0, 0, 6)6 - AIC:2204.9340491545727

SARIMA(1, 1, 1)x(0, 0, 1, 6)6 - AIC:2103.2471520742797

SARIMA(1, 1, 1)x(0, 0, 2, 6)6 - AIC:1906.3976381402158

SARIMA(1, 1, 1)x(1, 0, 0, 6)6 - AIC:2109.667120973266

SARIMA(1, 1, 1)x(1, 0, 1, 6)6 - AIC:2005.612566393106

SARIMA(1, 1, 1)x(1, 0, 2, 6)6 - AIC:1856.0775241006636

SARIMA(1, 1, 1)x(2, 0, 0, 6)6 - AIC:1776.9417670618593

SARIMA(1, 1, 1)x(2, 0, 1, 6)6 - AIC:1778.8222557528889

SARIMA(1, 1, 1)x(2, 0, 2, 6)6 - AIC:1743.3797826615128

SARIMA(1, 1, 2)x(0, 0, 0, 6)6 - AIC:2188.463345050498

SARIMA(1, 1, 2)x(0, 0, 1, 6)6 - AIC:2089.132092446418

SARIMA(1, 1, 2)x(0, 0, 2, 6)6 - AIC:1908.3347898062104

SARIMA(1, 1, 2)x(1, 0, 0, 6)6 - AIC:2108.5645510270633

SARIMA(1, 1, 2)x(1, 0, 1, 6)6 - AIC:1987.1476984956794

SARIMA(1, 1, 2)x(1, 0, 2, 6)6 - AIC:1838.9472829343013

SARIMA(1, 1, 2)x(2, 0, 0, 6)6 - AIC:1773.4229389294615

SARIMA(1, 1, 2)x(2, 0, 1, 6)6 - AIC:1775.2584010169296

SARIMA(1, 1, 2)x(2, 0, 2, 6)6 - AIC:1730.1038790837822

SARIMA(2, 1, 0)x(0, 0, 0, 6)6 - AIC:2227.302761872421

SARIMA(2, 1, 0)x(0, 0, 1, 6)6 - AIC:2145.3576991201103

SARIMA(2, 1, 0)x(0, 0, 2, 6)6 - AIC:1945.1561426085786

SARIMA(2, 1, 0)x(1, 0, 0, 6)6 - AIC:2124.9071786318195

SARIMA(2, 1, 0)x(1, 0, 1, 6)6 - AIC:2054.170071228054

SARIMA(2, 1, 0)x(1, 0, 2, 6)6 - AIC:1915.633692249958

SARIMA(2, 1, 0)x(2, 0, 0, 6)6 - AIC:1782.735782105787

SARIMA(2, 1, 0)x(2, 0, 1, 6)6 - AIC:1782.359816938853

SARIMA(2, 1, 0)x(2, 0, 2, 6)6 - AIC:1760.342670823393

SARIMA(2, 1, 1)x(0, 0, 0, 6)6 - AIC:2199.8586131454495

SARIMA(2, 1, 1)x(0, 0, 1, 6)6 - AIC:2103.0859058223045

SARIMA(2, 1, 1)x(0, 0, 2, 6)6 - AIC:1903.0416542490439

SARIMA(2, 1, 1)x(1, 0, 0, 6)6 - AIC:2088.133636367894

SARIMA(2, 1, 1)x(1, 0, 1, 6)6 - AIC:1997.3692882582268

SARIMA(2, 1, 1)x(1, 0, 2, 6)6 - AIC:1852.7863840750376

SARIMA(2, 1, 1)x(2, 0, 0, 6)6 - AIC:1794.8112219321938

SARIMA(2, 1, 1)x(2, 0, 1, 6)6 - AIC:1763.1914589558787

SARIMA(2, 1, 1)x(2, 0, 2, 6)6 - AIC:1743.8742069625332

SARIMA(2, 1, 2)x(0, 0, 0, 6)6 - AIC:2176.868114688141

SARIMA(2, 1, 2)x(0, 0, 1, 6)6 - AIC:2068.7780944519573

SARIMA(2, 1, 2)x(0, 0, 2, 6)6 - AIC:1889.7875404654014

SARIMA(2, 1, 2)x(1, 0, 0, 6)6 - AIC:2074.1102217500147

SARIMA(2, 1, 2)x(1, 0, 1, 6)6 - AIC:1955.605896297144

SARIMA(2, 1, 2)x(1, 0, 2, 6)6 - AIC:1826.0433954137

SARIMA(2, 1, 2)x(2, 0, 0, 6)6 - AIC:1763.274775274491

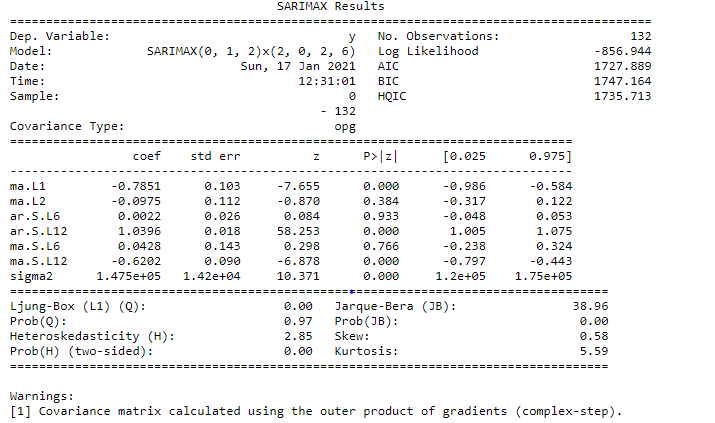
SARIMA(2, 1, 2)x(2, 0, 1, 6)6 - AIC:1760.8267450584606

SARIMA(2, 1, 2)x(2, 0, 2, 6)6 - AIC:1752.2758266821847

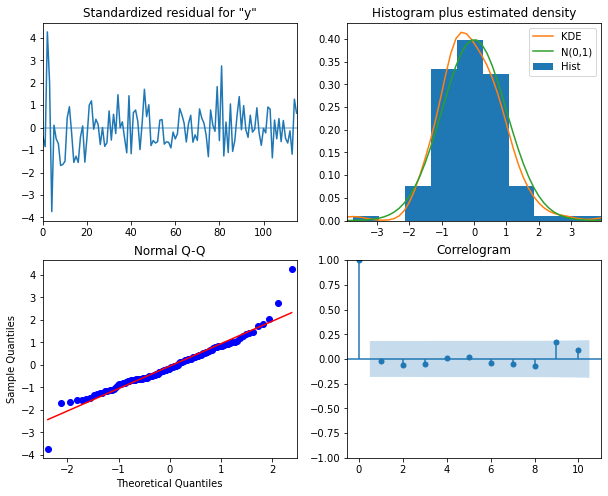
Accordingly following parameters performed well with lowest AIC score for SARIMA 6



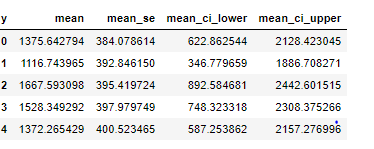
Please find below the statistical result summary after running the model for the above parameters.



Following plot for the residuals on the SARIMA model built based on the above optimized parameter values of (0,1,2) (2,0,2,6) that depicts near normality of the residuals.



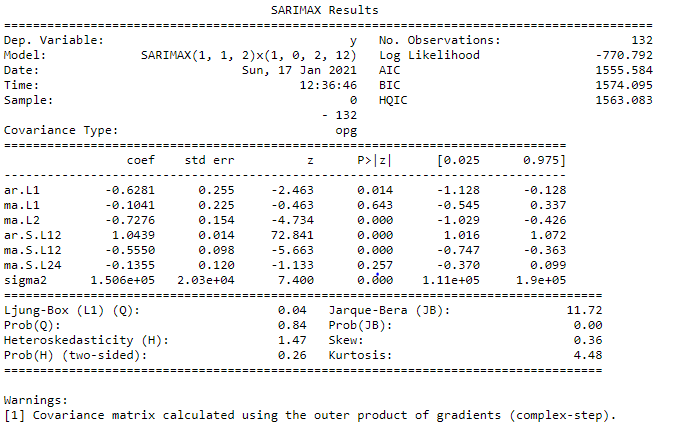
Based on the SARIMA model built above below are the predictions for the test data along with its confidence intervals with RMSE value of 601.2547122351654.



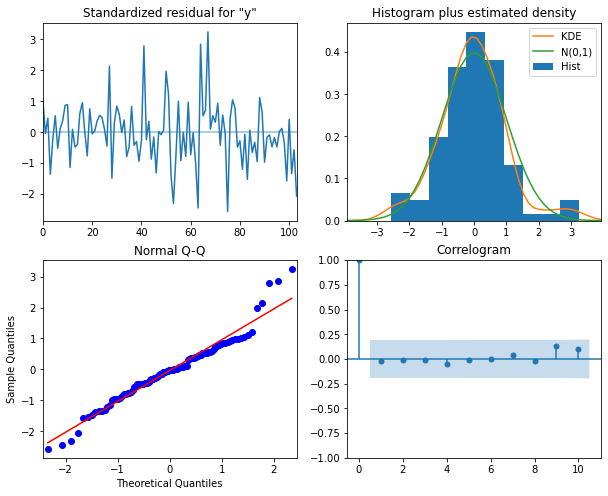
Now we retain the same parameters for p,d,q,P,D,Q but run the model for m=12 to apply 12 month seasonal component. Accordingly the following optimal parameters have been arrived after auto SARIMA process.



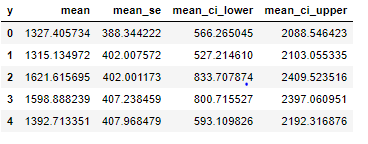
Please find below the statistical result summary after running the model for the above parameters.



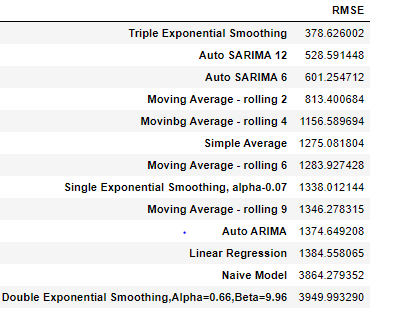
Following plot for the residuals on the SARIMA model built based on the above optimized parameter values of (1,1,2) (1,0,2,12) that depicts near normality of the residuals.



Based on the SARIMA model built above below are the predictions for the test data along with its confidence intervals with RMSE value of 528.5914475977731.



Below is the comparison of performance across all the model built so far for Sparkling data set. Based on the model in scope the Triple exponential smoothing / Holt Winters model has performed the best.



**ARIMA and SARIMA for Rose wine dataset.**

Based on the auto ARIMA process below are the parameter values and the respective AIC scores for the Sparkling dataset. We are retaining the p, d, q ranges aligning with what was done for Sparkling dataset.

ARIMA(0, 1, 0) - AIC:1335.1526583086775

ARIMA(0, 1, 1) - AIC:1280.7261830464035

ARIMA(0, 1, 2) - AIC:1276.8353726229147

ARIMA(1, 1, 0) - AIC:1319.348310580781

ARIMA(1, 1, 1) - AIC:1277.775749172235

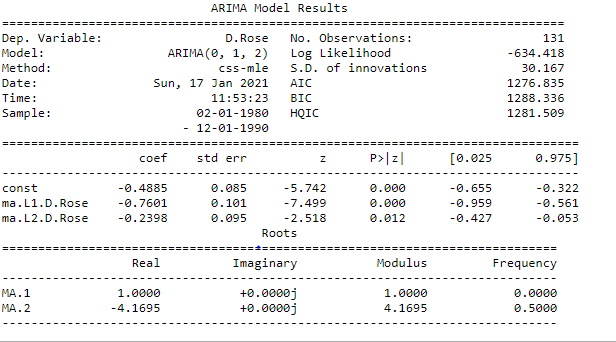
ARIMA(1, 1, 2) - AIC:1277.3592262067277

ARIMA(2, 1, 0) - AIC:1300.6092611745498

ARIMA(2, 1, 1) - AIC:1279.0456894093218

ARIMA(2, 1, 2) - AIC:1279.2986939364814

Accordingly the parameter with the order of (0,1,2) for (p,d,q) has performed better with the least AIC Score of 1276.835 for us to proceed with applying those parameter for the ARIMA model to check the statistical result on the output.



Based on the above following are the predictions on the test time series with the RMSE score of 15.61800957004907.

array([83.95205681, 71.47835275, 70.98981208, 70.50127142, 70.01273076,

69.5241901 , 69.03564944, 68.54710878, 68.05856812, 67.57002746,

67.0814868 , 66.59294614, 66.10440548, 65.61586481, 65.12732415,

64.63878349, 64.15024283, 63.66170217, 63.17316151, 62.68462085,

62.19608019, 61.70753953, 61.21899887, 60.7304582 , 60.24191754,

59.75337688, 59.26483622, 58.77629556, 58.2877549 , 57.79921424,

57.31067358, 56.82213292, 56.33359226, 55.8450516 , 55.35651093,

54.86797027, 54.37942961, 53.89088895, 53.40234829, 52.91380763,

52.42526697, 51.93672631, 51.44818565, 50.95964499, 50.47110433,

49.98256366, 49.494023 , 49.00548234, 48.51694168, 48.02840102,

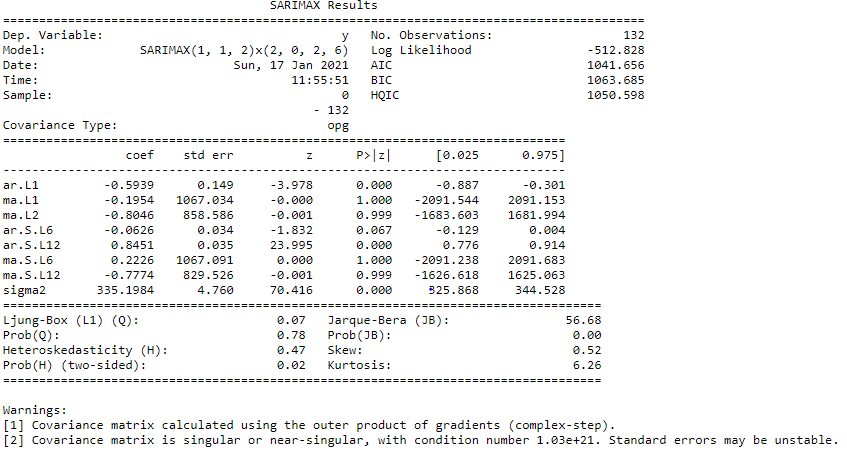
47.53986036, 47.0513197 , 46.56277904, 46.07423838, 45.58569772])

Now let’s build the SARIMA model with m = 6 with same p,d,q,P,D,Q parameter ranges for auto SARIMA process.

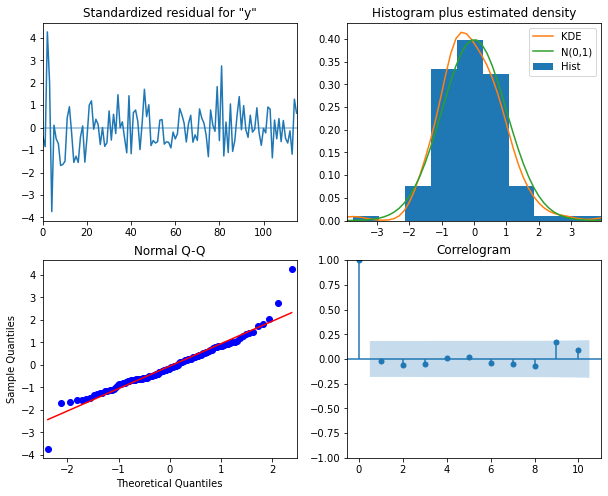
Based on the auto SARIMA outcome below is the optimal parameters narrowed down with the least AIC score to proceed further in building this model.



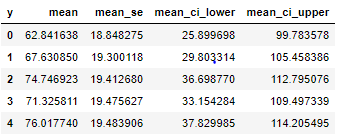
Below is the statistical summary post the model training.



Below plot diagnostics on the residuals depict normality.



Based on the trained model, following are the forecasts on the test series along with the confidence interval with the RMSE score of 26.1352027416264.

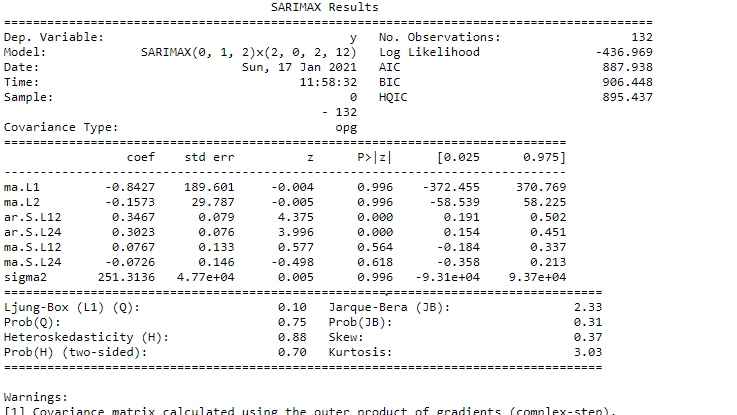


Likewise below is the model build for SARIMA with m=12.

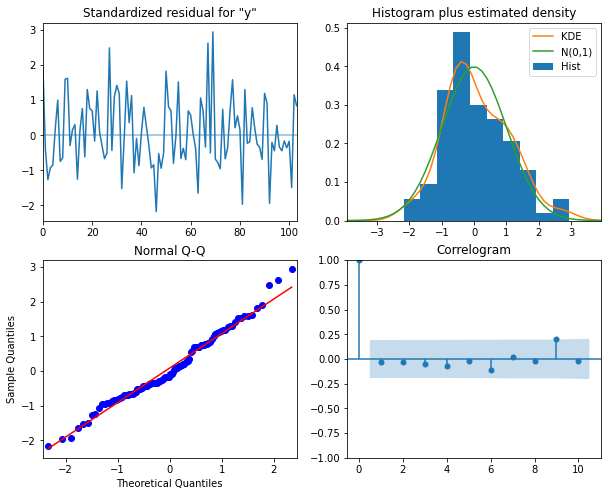
Optimal parameter post auto SARIMA process is as below:



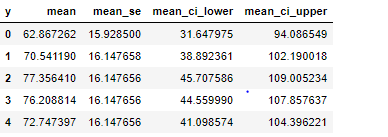
Statistical summary for the above parameters after building the SARIMA model for it is as below.



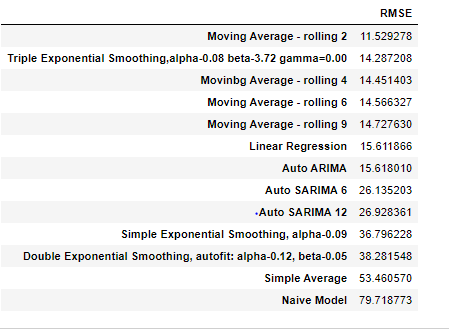
Below is the plot of residuals depicting normality of the residuals.



Based on the trained model below are the forecasts against the test series with the RMSE score of 26.928361253034.



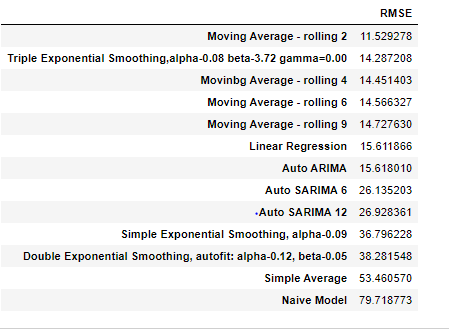
Below is the comparison of performance across all the model built so far for Rose data set. Based on the model in scope the Triple exponential smoothing model has performed the best not counting moving average approach.



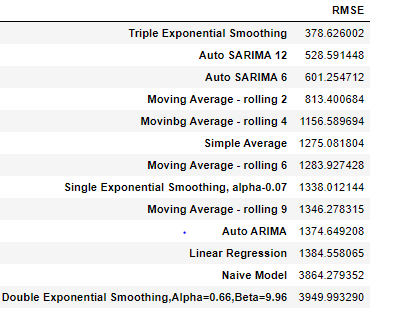
1. **Build ARIMA/SARIMA models based on the cut-off points of ACF and PACF on the training data and evaluate this model on the test data using RMSE.**
2. **Build a table (create a data frame) with all the models built along with their corresponding parameters and the respective RMSE values on the test data.**

For Rose data set

Below is the comparison of performance across all the model built so far for Rose data set. Based on the model in scope the Triple exponential smoothing model has performed the best not counting moving average approach.



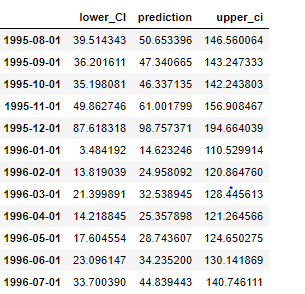
Below is the comparison of performance across all the models built so far for Sparkling data set. Based on the model in scope the Triple exponential smoothing / Holt Winters model has performed the best.



1. **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.**

With Triple exponential model being the best across both Rose and Sparkling wine data sets we are going ahead predicting for next 12 months using the full data volume based on the Holt Winters model as below.

For Rose wine data set below are the predictions of sales for next 12 months along with the respective confidence factor.



For Sparkling wine data set below are the predictions of sales for next 12 months along with the respective confidence factor.

