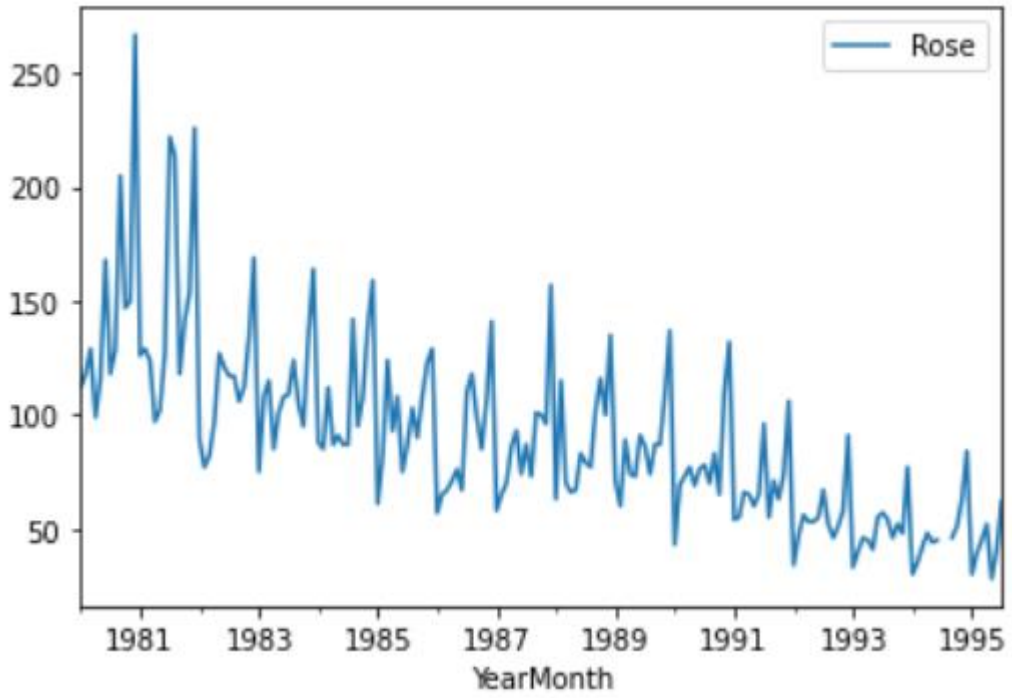
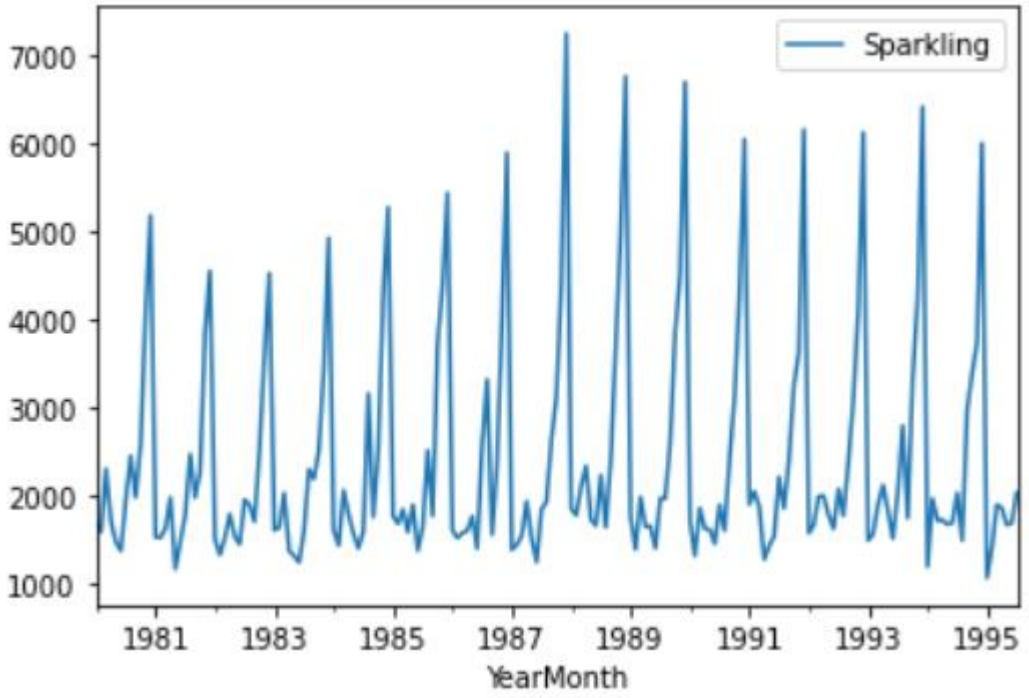
An Analysis of Sales of Rose and Sparkling Wines

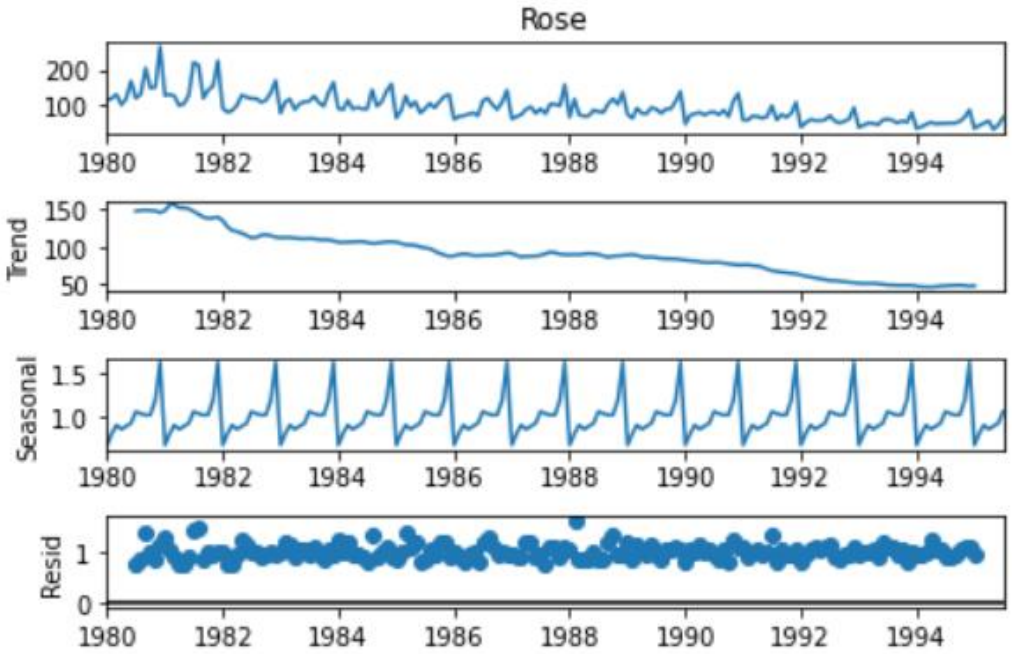
DIVE INTO THE DATA:

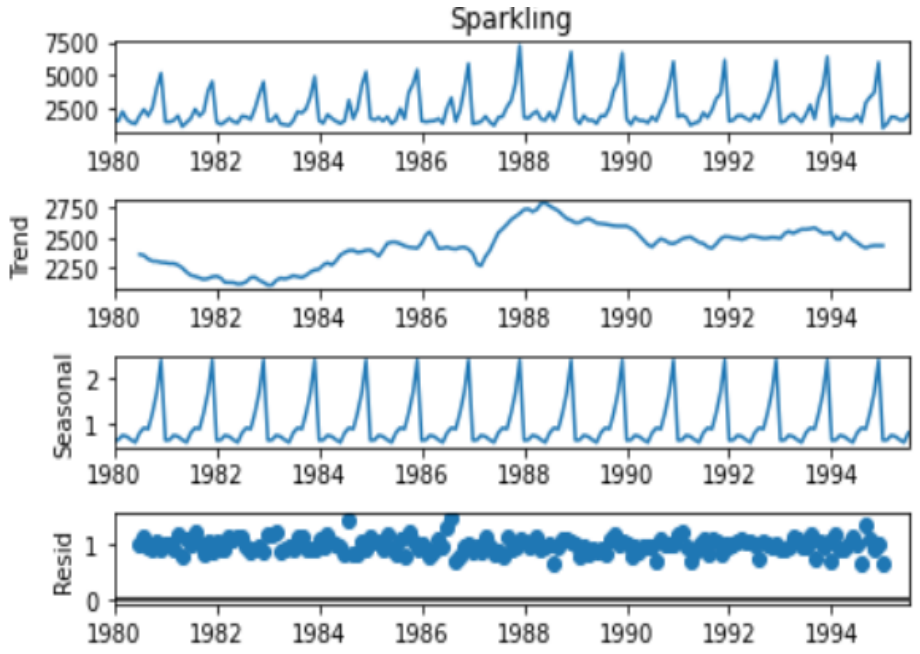
- ‘Rose’ wine prices shows downward trend as time passes while no generic trend observed in ‘Sparkling’ wine prices observed:



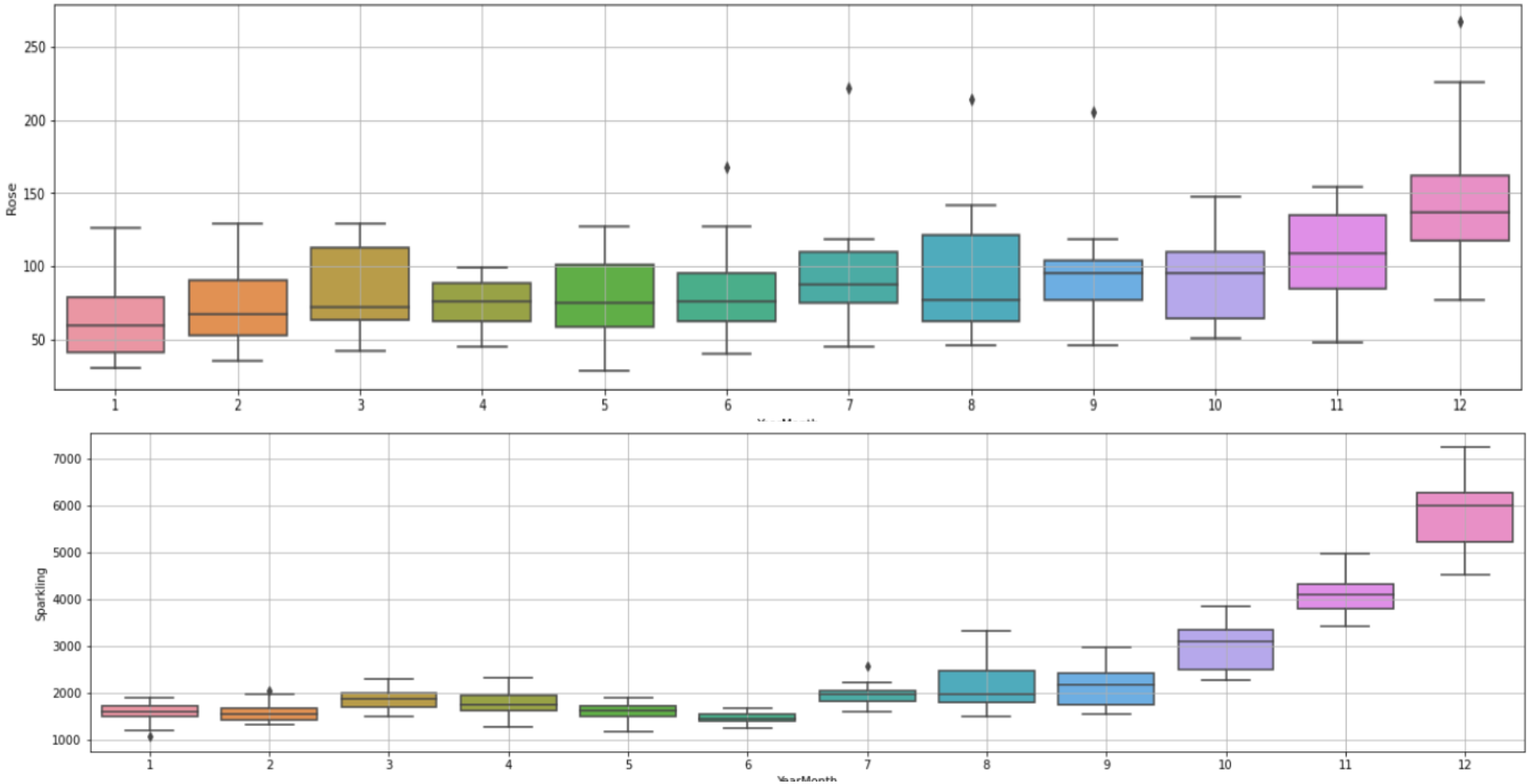


- ‘Rose’ wine prices shows multiplicative seasonality component whereas ‘Sparkling’ wine prices shows neither obvious trend nor seasonality component.

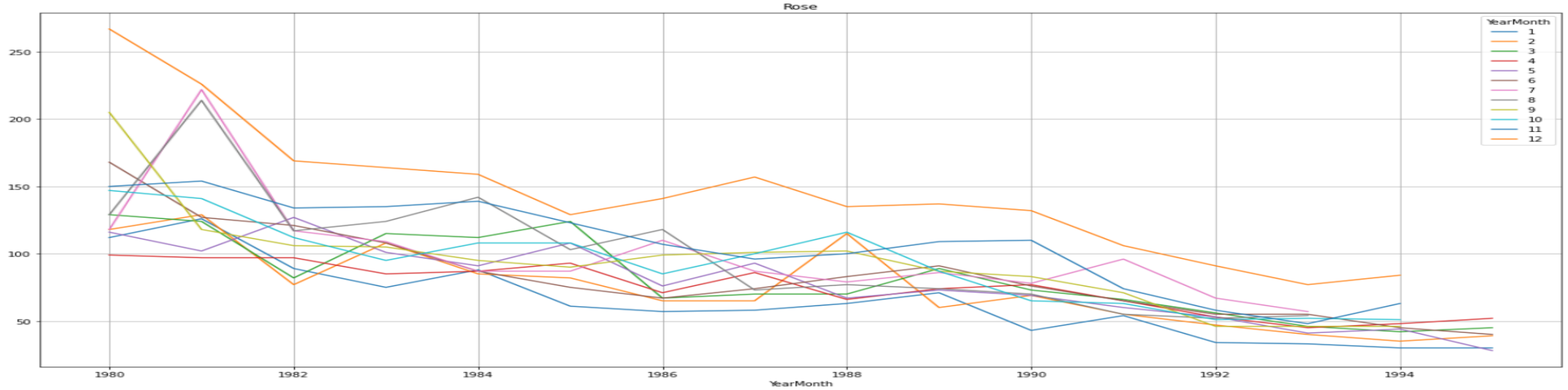


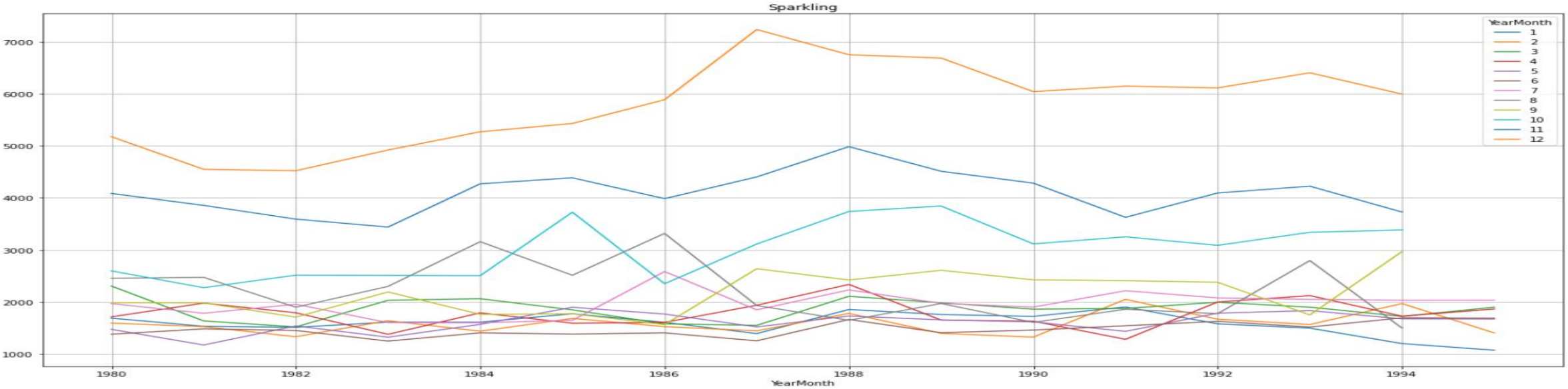


- For both Rose and Sparkling wine prices data , average price of the 12th month for each year is higher than all of the months



- We could see average price of Rose wine of a month for each year is higher for the year 1980 while there is no generic trend observed in Sparkling wine



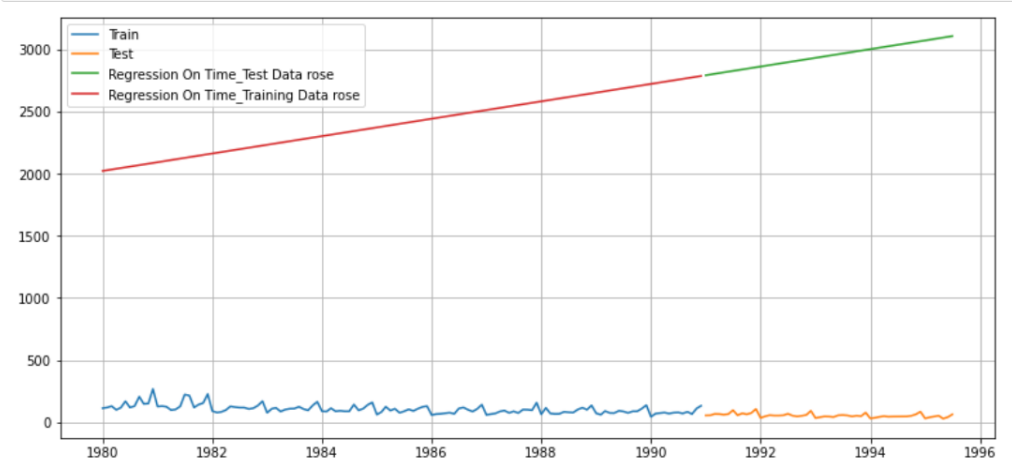


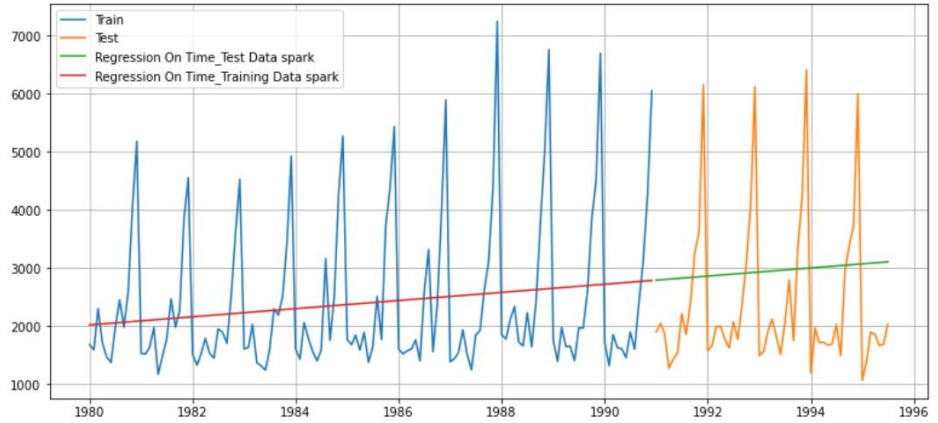
Applying Various models and RMSE Evaluation

- We could see RMSE for Linear Regression below for both Rose and Sparkle data:

	Test RMSE-Rose Data
RegressionOnTime	2897.007078

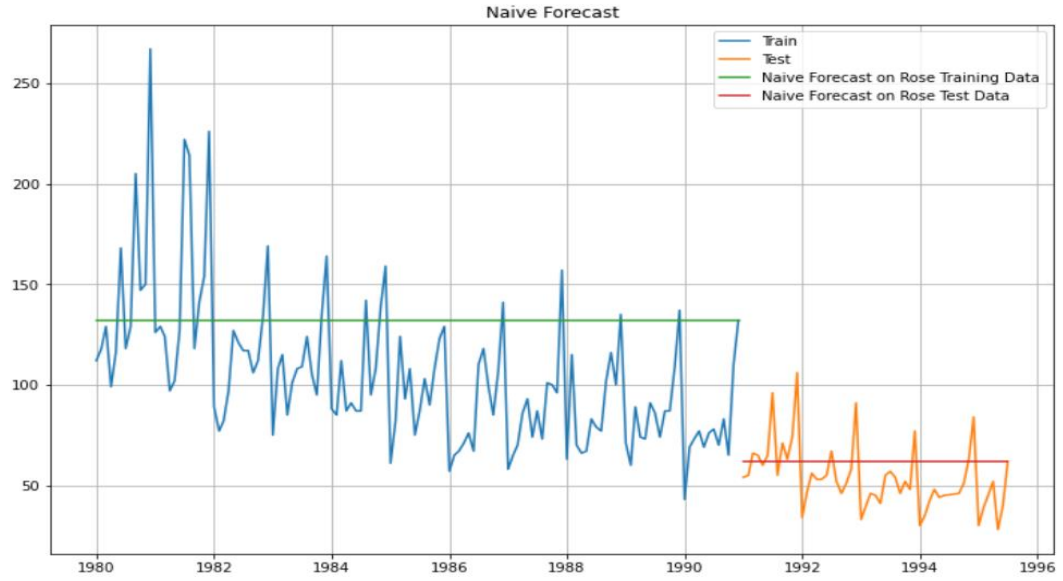
	Test RMSE-Spark Data
RegressionOnTime	1389.135175



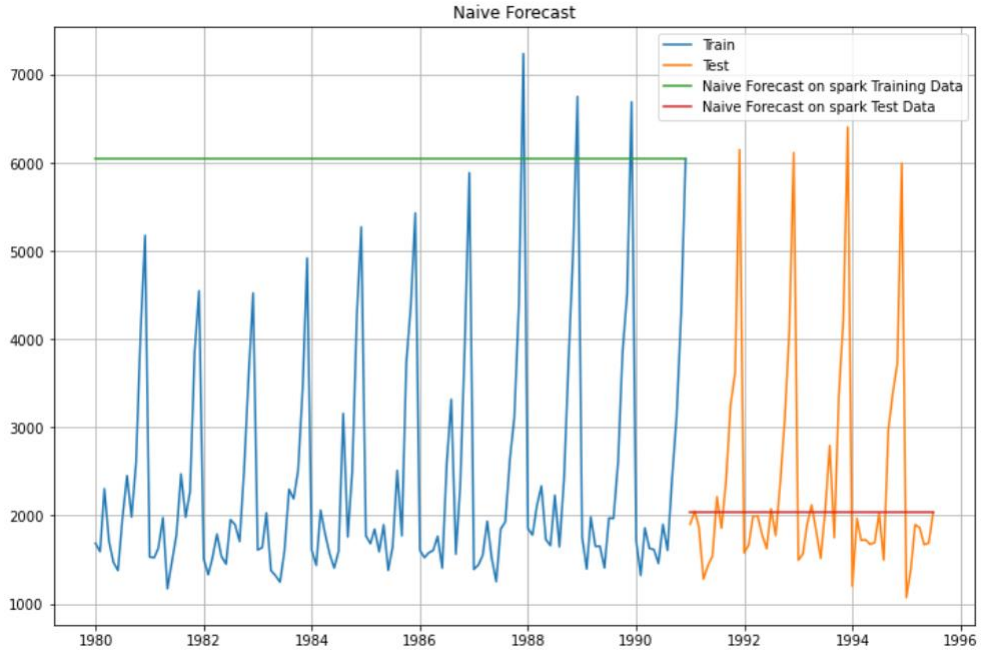


- We could see RMSE using Naïve approach is greatly reduced for Rose but not so much for Sparkling:

	Test RMSE-Rose Data
RegressionOnTime	2897.007078
Naive	17.740511



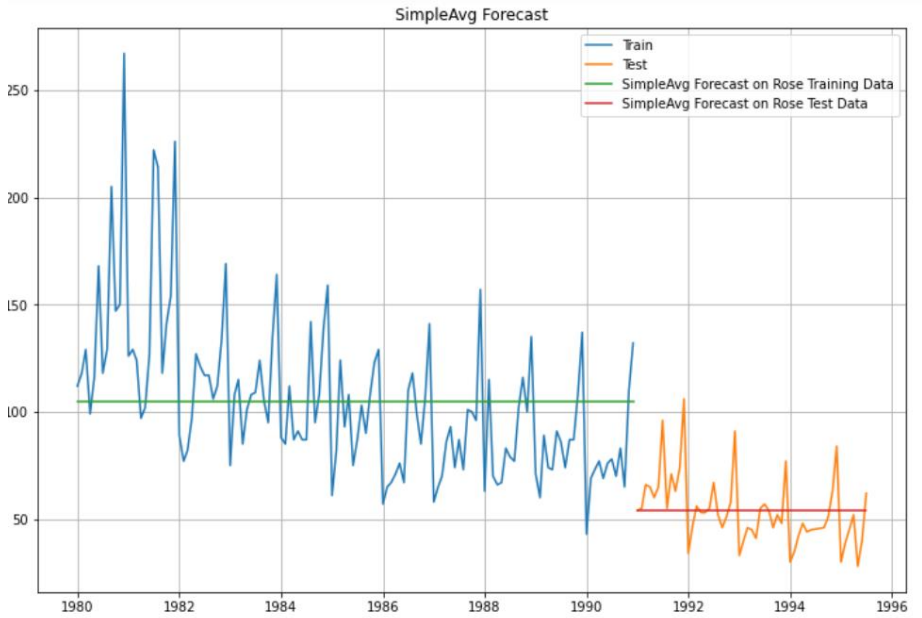
	Test RMSE-Spark Data
RegressionOnTime	1389.135175
Naive	1327.156057

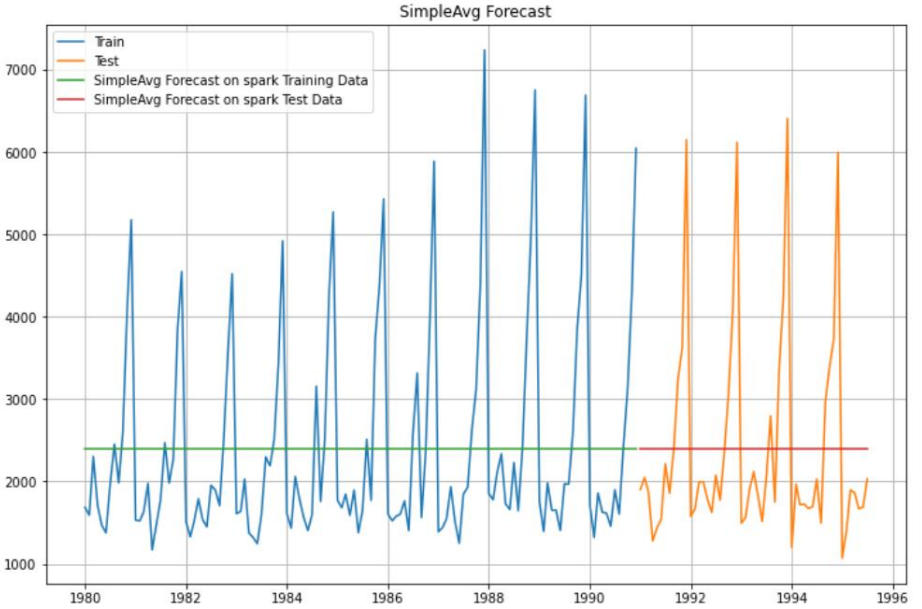


- We could see RMSE using SimpleAvg approach reduced for Rose and Sparkling:

	Test RMSE-Rose Data
RegressionOnTime	2897.007078
Naive	17.740511
SimpleAvg	15.759889

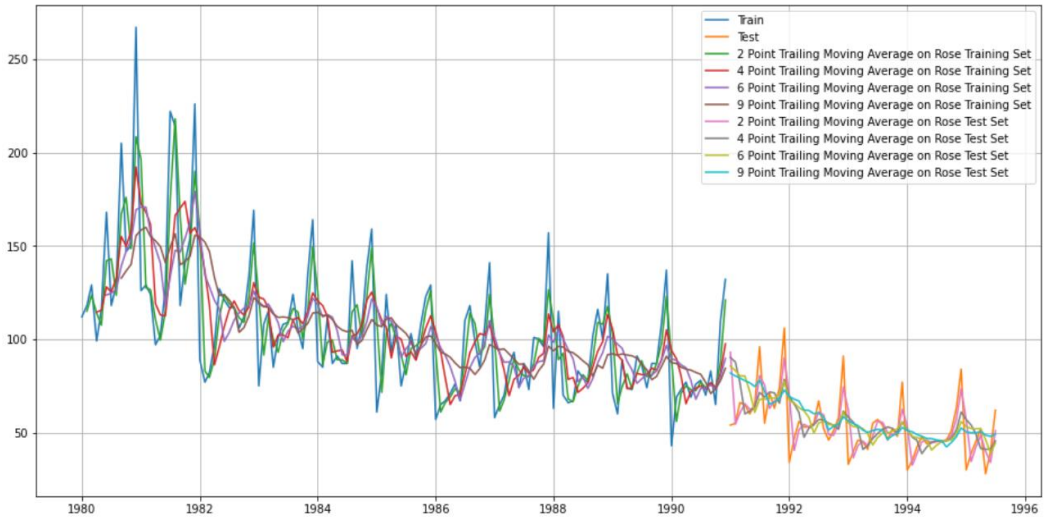
	Test RMSE-Spark Data
RegressionOnTime	1389.135175
Naive	1327.156057
SimpleAvg	1275.073380



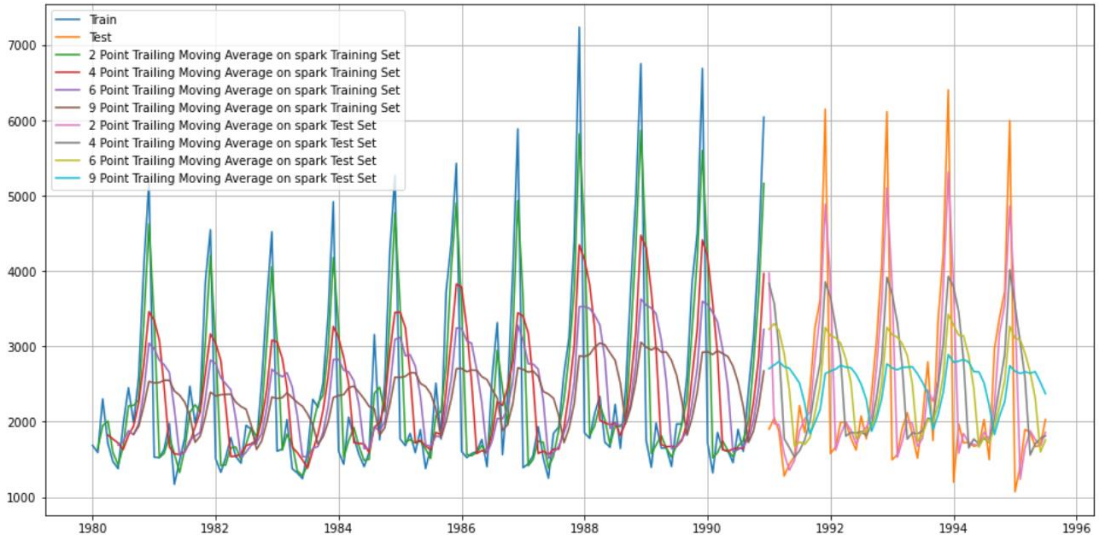


- We could see RMSE using MovingAvg approach for Rose and Sparkling(Both have min at 2 point avg):

	Test RMSE-Rose Data
RegressionOnTime	2897.007078
Naive	17.740511
SimpleAvg	15.759889
2pointTrailingMovingAverageRose	11.529278
4pointTrailingMovingAverageRose	14.451433
6pointTrailingMovingAverageRose	14.566399
9pointTrailingMovingAverageRose	14.727667



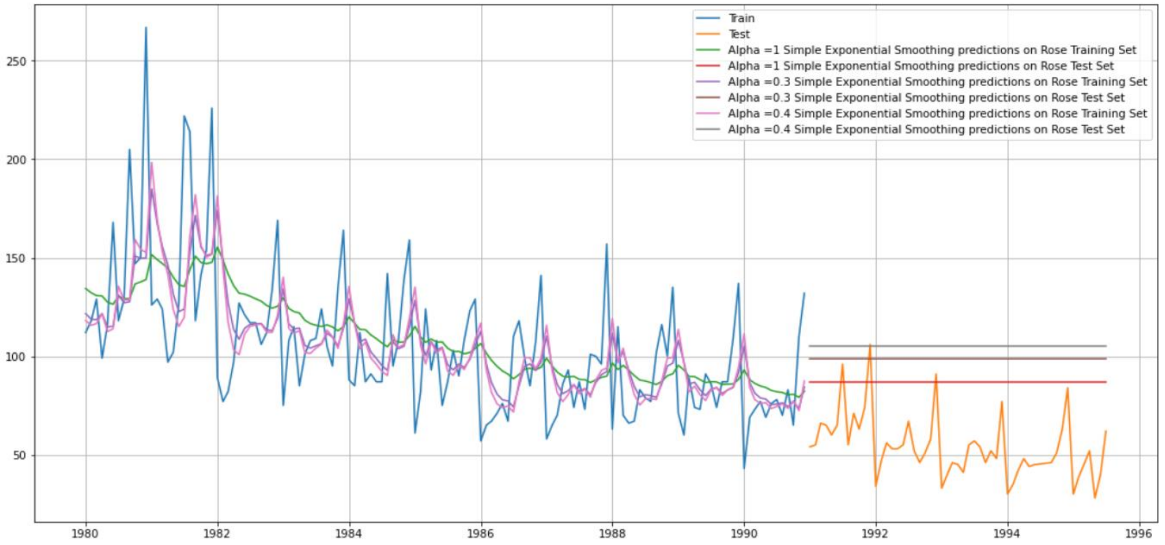
	Test RMSE-Spark Data
RegressionOnTime	1389.135175
Naive	1327.156057
SimpleAvg	1275.073380
2pointTrailingMovingAveragespark	813.400684
4pointTrailingMovingAveragespark	1156.589694
6pointTrailingMovingAveragespark	1283.927428
9pointTrailingMovingAveragespark	1346.278315

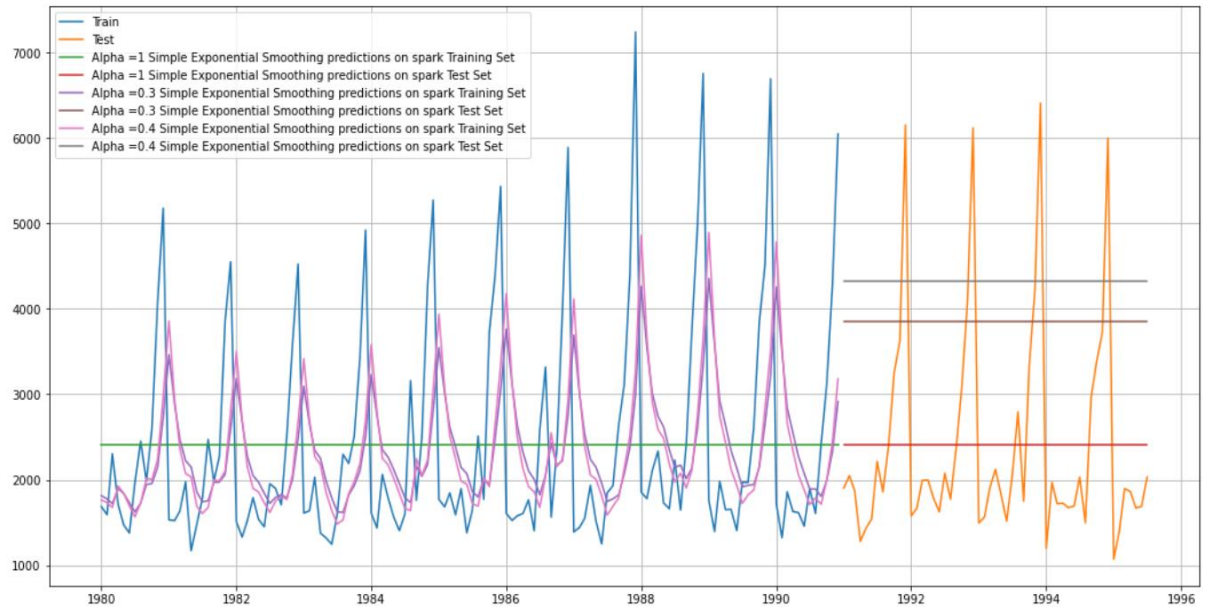


- We could see RMSE using SES for various values of alpha for Rose and Sparkling(min at alpha=1,while 2 point MovAvg still has lower RMSE):

	Test RMSE-Rose Data
RegressionOnTime	2897.007078
Naive	17.740511
SimpleAvg	15.759889
2pointTrailingMovingAverageRose	11.529278
4pointTrailingMovingAverageRose	14.451433
6pointTrailingMovingAverageRose	14.566399
9pointTrailingMovingAverageRose	14.727667
Alpha=1,SimpleExponentialSmoothing	36.796465
Alpha=0.3,SimpleExponentialSmoothingRose	77.139491
Alpha=0.4,SimpleExponentialSmoothingRose	77.139491

	Test RMSE-Spark Data
RegressionOnTime	1389.135175
Naive	1327.156057
SimpleAvg	1275.073380
2pointTrailingMovingAveragespark	813.400684
4pointTrailingMovingAveragespark	1156.589694
6pointTrailingMovingAveragespark	1283.927428
9pointTrailingMovingAveragespark	1346.278315
Alpha=1,SimpleExponentialSmoothing	1275.081839
Alpha=0.3,SimpleExponentialSmoothingspark	3686.794285
Alpha=0.4,SimpleExponentialSmoothingspark	3686.794285

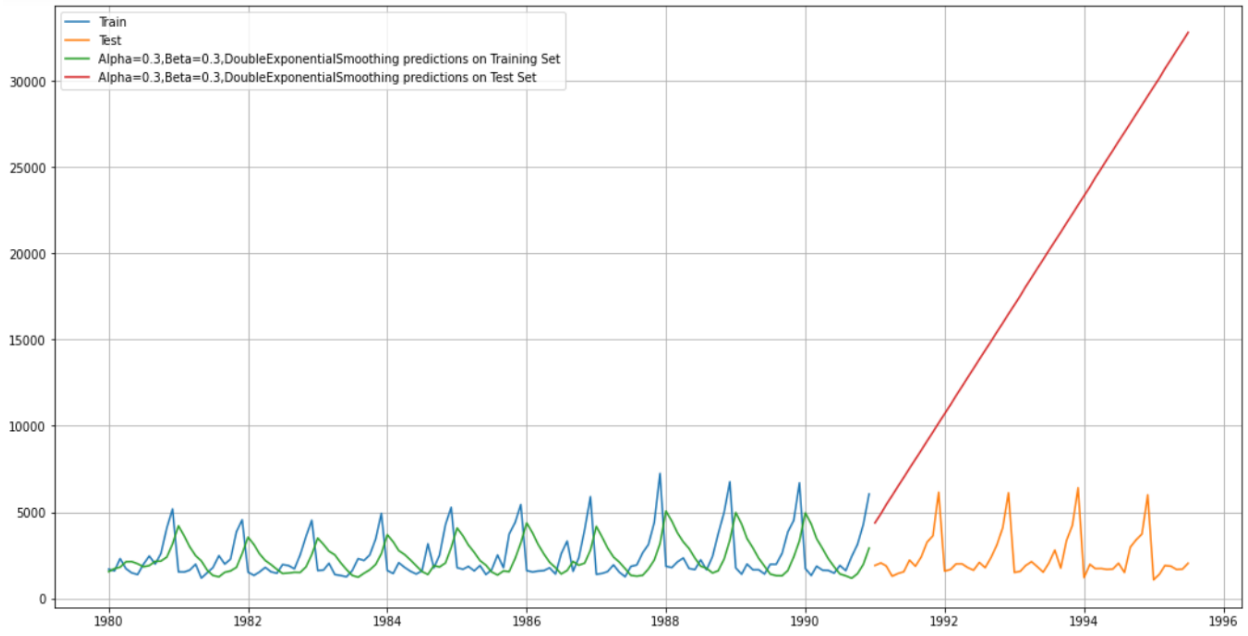


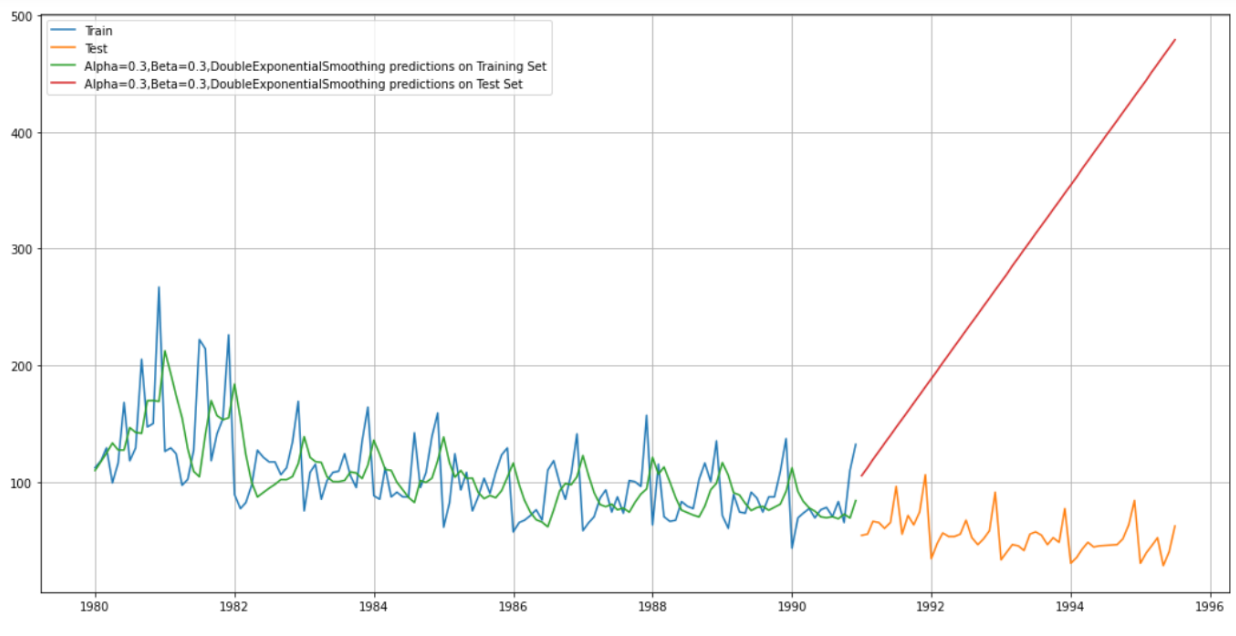


- We could see RMSE using Holtz's Model for various values of alpha and beta for Rose and Sparkling(it is found RMSE is minimum at alpha=0.3 and beta=0.3,please find calculations in code notebook):

	Test RMSE-Rose Data
RegressionOnTime	2897.007078
Naive	17.740511
SimpleAvg	15.759889
2pointTrailingMovingAverageRose	11.529278
4pointTrailingMovingAverageRose	14.451433
6pointTrailingMovingAverageRose	14.566399
9pointTrailingMovingAverageRose	14.727667
Alpha=1,SimpleExponentialSmoothing	36.796465
Alpha=0.3,SimpleExponentialSmoothingRose	77.139491
Alpha=0.4,SimpleExponentialSmoothingRose	77.139491
Alpha=0.3,Beta=0.3HoltzModelRose	265.567857

	Test RMSE-Spark Data
RegressionOnTime	1389.135175
Naive	1327.156057
SimpleAvg	1275.073380
2pointTrailingMovingAveragespark	813.400684
4pointTrailingMovingAveragespark	1156.589694
6pointTrailingMovingAveragespark	1283.927428
9pointTrailingMovingAveragespark	1346.278315
Alpha=1,SimpleExponentialSmoothing	1275.081839
Alpha=0.3,SimpleExponentialSmoothingspark	3686.794285
Alpha=0.4,SimpleExponentialSmoothingspark	3686.794285
Alpha=0.3,Beta=0.3HoltzModelspark	18259.110704

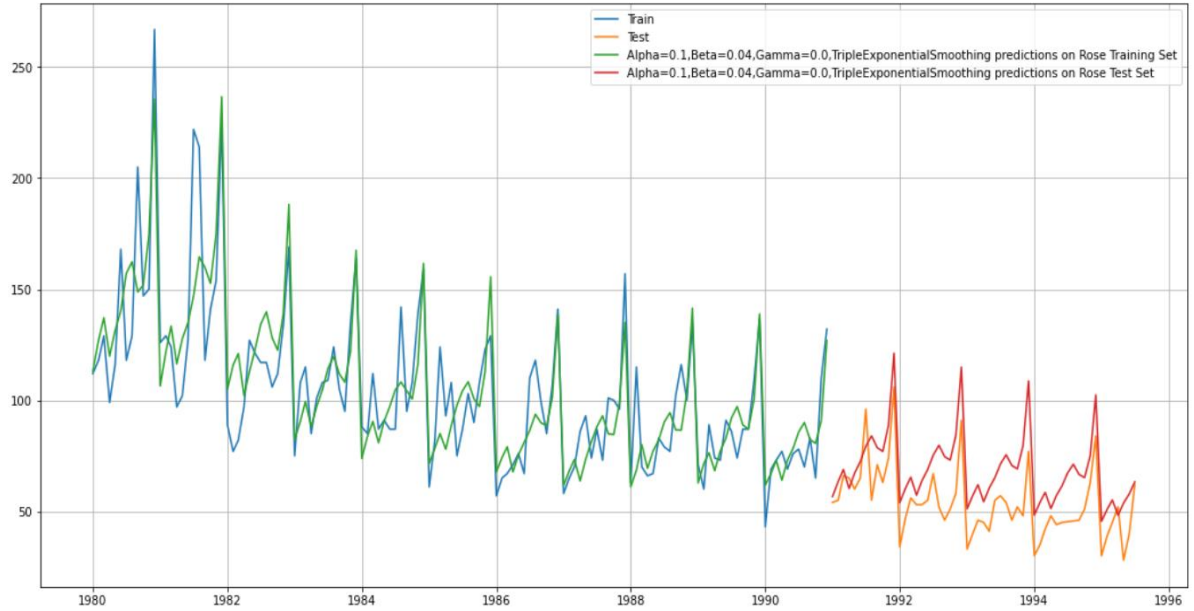


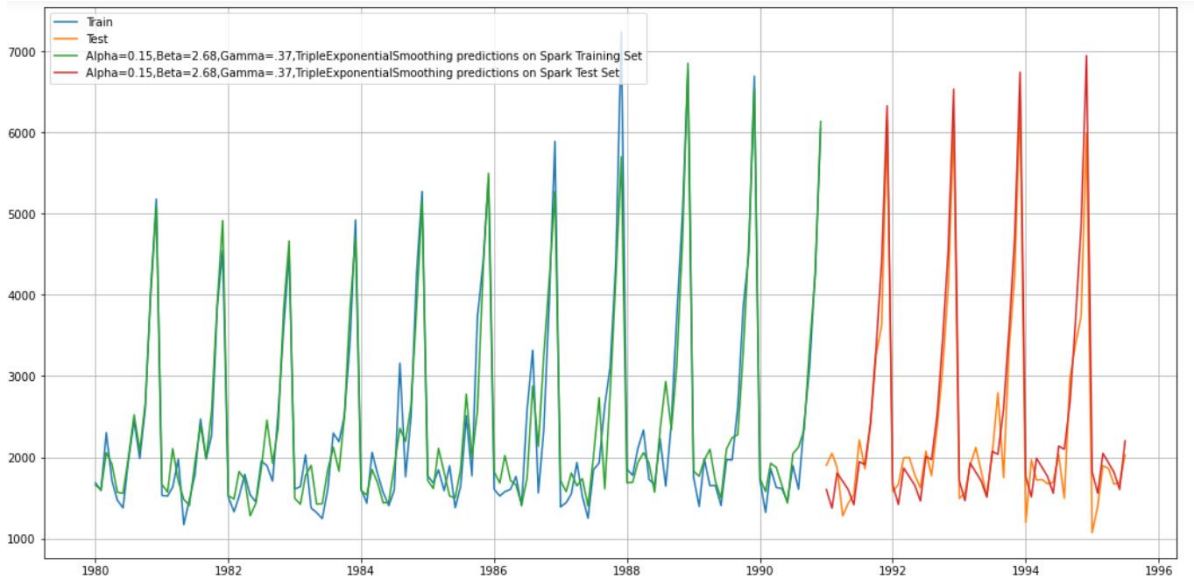


- We could see RMSE using Holtz's Winter Model for various values of alpha,beta and gamma for Rose and Sparkling(it is found RMSE is minimum for the below values and surprisingly we are able to find 2 day Movavg performs well on Rose while Holtz Winter model performs well on Sparkling wine data):

	Test RMSE-Rose Data
RegressionOnTime	1267.516575
Naive	17.740511
SimpleAvg	15.759889
2pointTrailingMovingAverageRose	11.529278
4pointTrailingMovingAverageRose	14.451433
6pointTrailingMovingAverageRose	14.566399
9pointTrailingMovingAverageRose	14.727667
Alpha=1,SimpleExponentialSmoothing	36.796465
Alpha=0.3,SimpleExponentialSmoothingRose	77.139491
Alpha=0.4,SimpleExponentialSmoothingRose	77.139491
Alpha=0.3,Beta=0.3HoltzModelRose	265.567857
Alpha=0.684,Beta=0.052,Gamma=0.315,TripleExponentialSmoothing	17.369752

	Test RMSE-Spark Data
RegressionOnTime	1389.249047
Naive	1327.156057
SimpleAvg	1275.073380
2pointTrailingMovingAveragespark	813.400684
4pointTrailingMovingAveragespark	1156.589694
6pointTrailingMovingAveragespark	1283.927428
9pointTrailingMovingAveragespark	1346.278315
Alpha=1,SimpleExponentialSmoothing	1275.081839
Alpha=0.3,SimpleExponentialSmoothingspark	3686.794285
Alpha=0.4,SimpleExponentialSmoothingspark	3686.794285
Alpha=0.3,Beta=0.3HoltzModelspark	18259.110704
Alpha=0.15,Beta=2.68,Gamma=.37,TripleExponentialSmoothing	383.192343

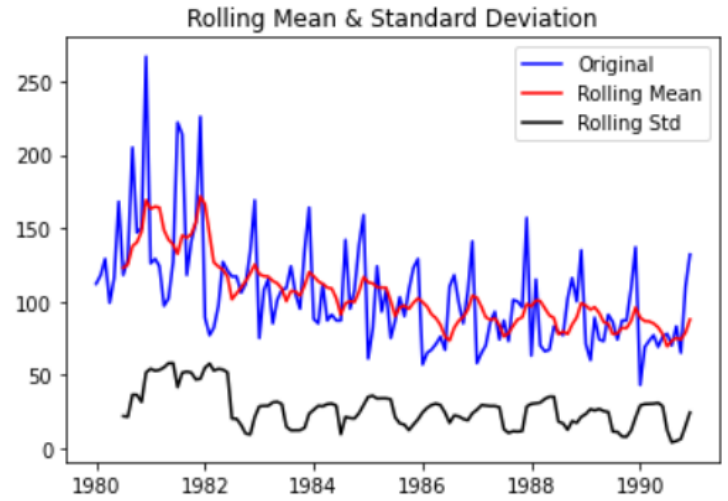




Check For Stationarity:

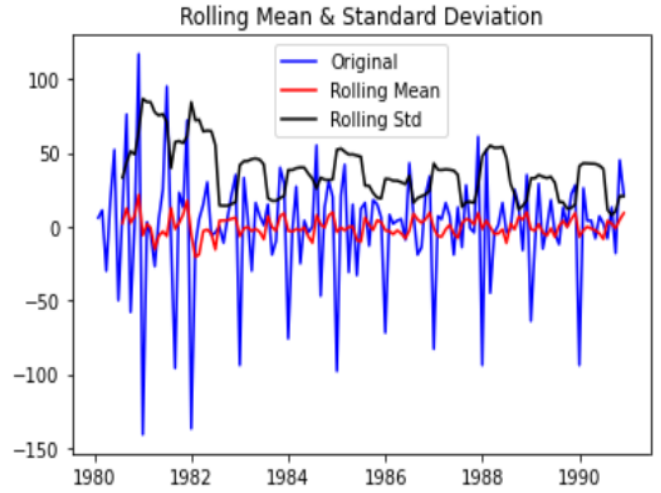
- We could see data as it is , is not stationary so we have taken first order differencing to make the data stationary for both Rose and Sparkling wine Dataset:

```
test_stationarity(train_rose['Rose'])
```




```
Results of Dickey-Fuller Test:
Test Statistic                  -2.164250
p-value                          0.219476
#Lags Used                      13.000000
Number of Observations Used    118.000000
Critical Value (1%)             -3.487022
Critical Value (5%)             -2.886363
Critical Value (10%)            -2.580009
```

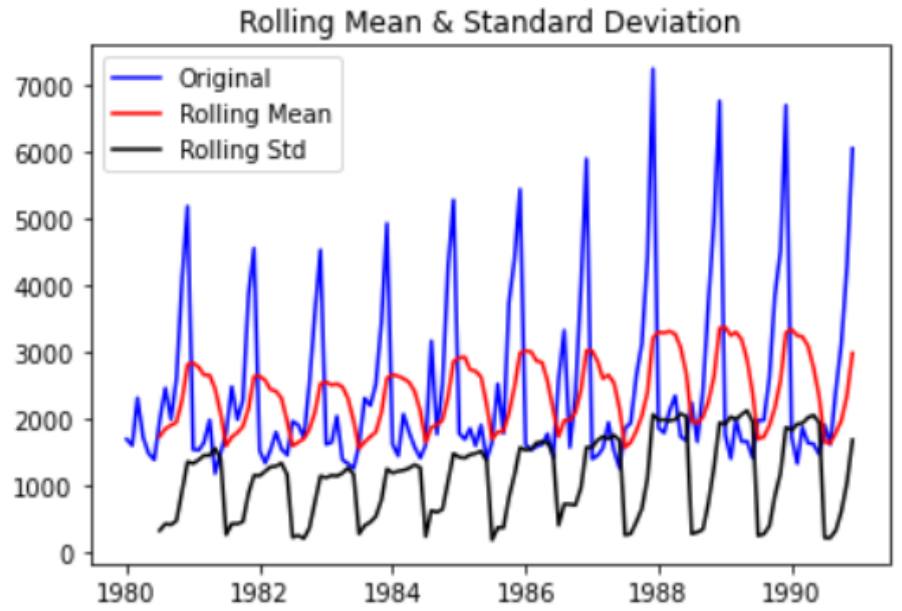
```
test_stationarity(train_rose['Rose'].diff().dropna())
```




```
Results of Dickey-Fuller Test:
Test Statistic                -6.592372e+00
p-value                        7.061944e-09
#Lags Used                     1.200000e+01
Number of Observations Used    1.180000e+02
Critical Value (1%)           -3.487022e+00
Critical Value (5%)           -2.886363e+00
Critical Value (10%)          -2.580009e+00
```

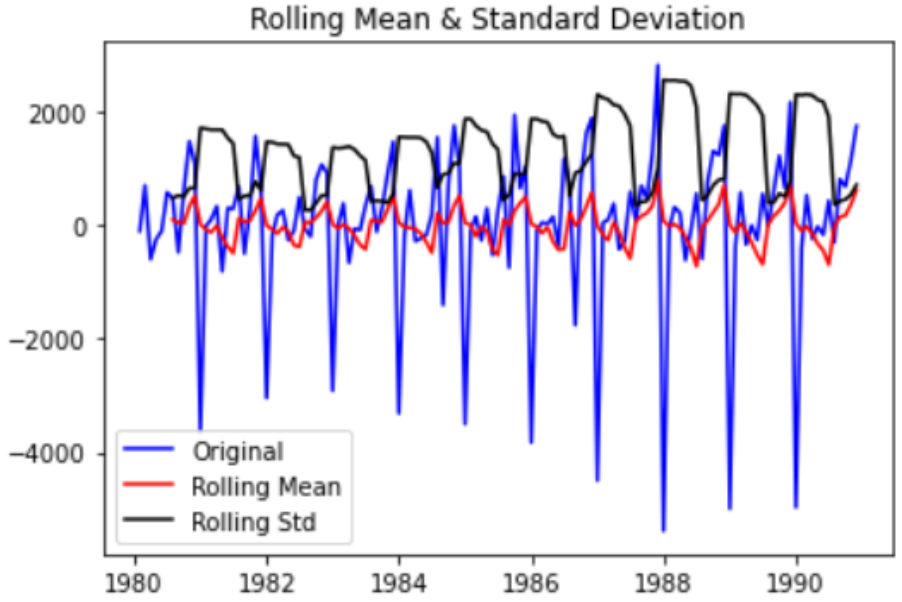
- Below visual plot shows differencing required for Sparkling wine dataset too:

```
test_stationarity(train_spark['Sparkling'])
```




```
Results of Dickey-Fuller Test:
Test Statistic                  -1.208926
p-value                          0.669744
#Lags Used                      12.000000
Number of Observations Used    119.000000
Critical Value (1%)             -3.486535
Critical Value (5%)             -2.886151
Critical Value (10%)            -2.579896
```

```
test_stationarity(train_spark['Sparkling'].diff().dropna())
```




```
Results of Dickey-Fuller Test:
Test Statistic                  -8.005007e+00
p-value                          2.280104e-12
#Lags Used                       1.100000e+01
Number of Observations Used      1.190000e+02
Critical Value (1%)             -3.486535e+00
Critical Value (5%)             -2.886151e+00
Critical Value (10%)            -2.579896e+00
```

ARIMA model-Automated:

- For various values of p,q,d of ARIMA model we could select one with lowest AIC in this case (0,0,2) for Rose wine dataset and (2,0,2) for Sparkling wine data:

	param_rose	AIC
11	(0, 0, 2)	1276.835372
14	(1, 0, 2)	1277.359222
13	(1, 0, 1)	1277.775748
16	(2, 0, 1)	1279.045689
17	(2, 0, 2)	1279.298694
10	(0, 0, 1)	1280.726183
5	(1, 0, 2)	1292.053210
8	(2, 0, 2)	1292.248055
7	(2, 0, 1)	1292.937195
4	(1, 0, 1)	1294.510585

	param_spark	AIC
8	(2, 0, 2)	2201.092654
17	(2, 0, 2)	2210.623067
16	(2, 0, 1)	2232.360490
11	(0, 0, 2)	2232.783098
14	(1, 0, 2)	2233.597647
13	(1, 0, 1)	2235.013945
7	(2, 0, 1)	2236.590818
6	(2, 0, 0)	2244.799915
1	(0, 0, 1)	2245.268851

- Rose:

ARMA Model Results

Dep. Variable:	Rose	No. Observations:	131
Model:	ARMA(0, 2)	Log Likelihood	-634.418
Method:	css-mle	S.D. of innovations	30.167
Date:	Sun, 13 Sep 2020	AIC	1276.835
Time:	15:07:46	BIC	1288.336
Sample:	02-01-1980 - 12-01-1990	HQIC	1281.509

	coef	std err	z	P>|z|	[0.025	0.975]
const	-0.4885	0.085	-5.742	0.000	-0.655	-0.322
ma.L1.Rose	-0.7601	0.101	-7.499	0.000	-0.959	-0.561
ma.L2.Rose	-0.2398	0.095	-2.518	0.012	-0.427	-0.053

Roots

	Real	Imaginary	Modulus	Frequency
MA.1	1.0000	+0.0000j	1.0000	0.0000
MA.2	-4.1695	+0.0000j	4.1695	0.5000

- Sparkling:

ARMA Model Results

Dep. Variable:	Sparkling	No. Observations:	131
Model:	ARMA(2, 2)	Log Likelihood	-1099.312
Method:	css-mle	S.D. of innovations	1013.526
Date:	Sun, 13 Sep 2020	AIC	2210.623
Time:	15:08:14	BIC	2227.874
Sample:	02-01-1980 - 12-01-1990	HQIC	2217.633

	coef	std err	z	P>|z|	[0.025	0.975]
const	5.5837	0.519	10.757	0.000	4.566	6.601
ar.L1.Sparkling	1.2699	0.075	17.042	0.000	1.124	1.416
ar.L2.Sparkling	-0.5602	0.074	-7.618	0.000	-0.704	-0.416
ma.L1.Sparkling	-1.9961	0.043	-46.886	0.000	-2.080	-1.913
ma.L2.Sparkling	0.9961	0.043	23.340	0.000	0.912	1.080

Roots

	Real	Imaginary	Modulus	Frequency
AR.1	1.1334	-0.7074j	1.3361	-0.0888
AR.2	1.1334	+0.7074j	1.3361	0.0888
MA.1	1.0003	+0.0000j	1.0003	0.0000
MA.2	1.0036	+0.0000j	1.0036	0.0000

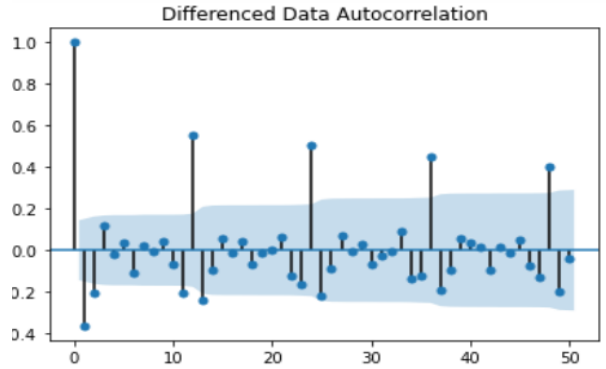
- RMSE for Rose dataset looks ok but it did not perform well when compared with Triple Exponential Smoothing, same is the case for Sparkling wine data.

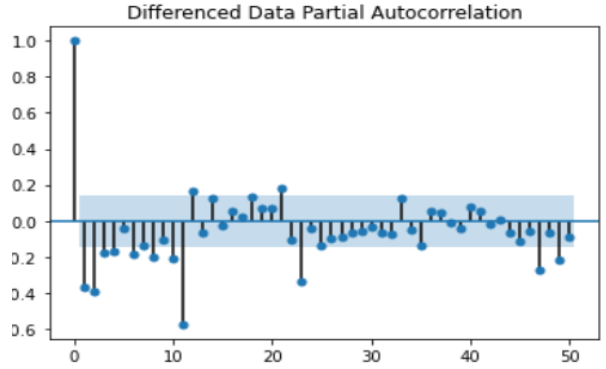
	Test RMSE-Rose Data
RegressionOnTime	1267.516575
Naive	17.740511
SimpleAvg	15.759889
2pointTrailingMovingAverageRose	11.529278
4pointTrailingMovingAverageRose	14.451433
6pointTrailingMovingAverageRose	14.566399
9pointTrailingMovingAverageRose	14.727667
Alpha=1,SimpleExponentialSmoothing	36.796465
Alpha=0.3,SimpleExponentialSmoothingRose	77.139491
Alpha=0.4,SimpleExponentialSmoothingRose	77.139491
Alpha=0.3,Beta=0.3HoltzModelRose	265.567857
Alpha=0.684,Beta=0.052,Gamma=0.315,TripleExponentialSmoothing	17.369752
ARIMA(0,0,2)	57.997243

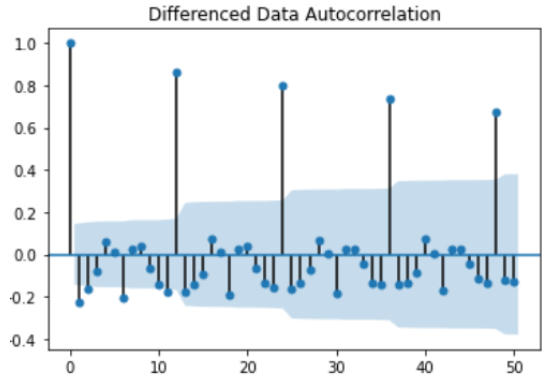
	Test RMSE-Spark Data
RegressionOnTime	1389.249047
Naive	1327.156057
SimpleAvg	1275.073380
2pointTrailingMovingAveragespark	813.400684
4pointTrailingMovingAveragespark	1156.589694
6pointTrailingMovingAveragespark	1283.927428
9pointTrailingMovingAveragespark	1346.278315
Alpha=1,SimpleExponentialSmoothing	1275.081839
Alpha=0.3,SimpleExponentialSmoothingspark	3686.794285
Alpha=0.4,SimpleExponentialSmoothingspark	3686.794285
Alpha=0.3,Beta=0.3HoltzModelspark	18259.110704
Alpha=0.15,Beta=2.68,Gamma=.37,TripleExponentialSmoothing	383.192343
ARIMA(2,0,2)	2777.091535

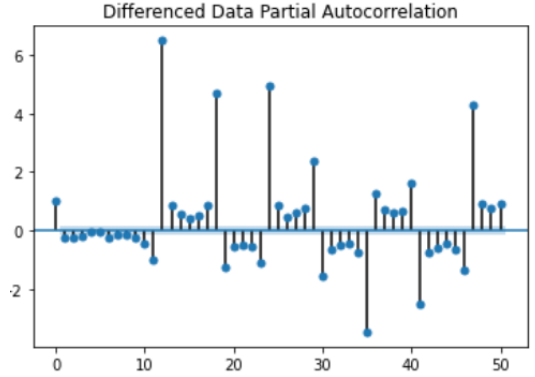
ARIMA model-Manual:

- For various values of p,q,d of ARIMA model we could select the parameters from PACF-ACF plot,in this case (5,1,2) for Rose wine(left) dataset and (1,1,2) for Sparkling wine data(right):









- Rose:

ARIMA Model Results

Dep. Variable:	D.Rose	No. Observations:	131
Model:	ARIMA(5, 1, 2)	Log Likelihood	-633.567
Method:	css-mle	S.D. of innovations	29.962
Date:	Sun, 13 Sep 2020	AIC	1285.134
Time:	15:27:18	BIC	1311.011
Sample:	02-01-1980 - 12-01-1990	HQIC	1295.649

	coef	std err	z	P>|z|	[0.025	0.975]
const	-0.4903	0.084	-5.841	0.000	-0.655	-0.326
ar.L1.D.Rose	-0.7739	0.088	-8.840	0.000	-0.945	-0.602
ar.L2.D.Rose	0.1243	0.102	1.215	0.224	-0.076	0.325
ar.L3.D.Rose	0.0154	0.103	0.149	0.882	-0.187	0.217
ar.L4.D.Rose	0.0552	0.141	0.391	0.696	-0.222	0.332
ar.L5.D.Rose	-0.0620	0.104	-0.597	0.551	-0.266	0.142
ma.L1.D.Rose	-3.298e-08	0.041	-8.06e-07	1.000	-0.080	0.080
ma.L2.D.Rose	-1.0000	0.041	-24.449	0.000	-1.080	-0.920

Roots

	Real	Imaginary	Modulus	Frequency
AR.1	-1.0001	-0.0000j	1.0001	-0.5000
AR.2	-0.8896	-1.5957j	1.8269	-0.3309
AR.3	-0.8896	+1.5957j	1.8269	0.3309
AR.4	1.8349	-1.2113j	2.1986	-0.0929
AR.5	1.8349	+1.2113j	2.1986	0.0929
MA.1	1.0000	+0.0000j	1.0000	0.0000
MA.2	-1.0000	+0.0000j	1.0000	0.5000

- Sparkling:

ARIMA Model Results

Dep. Variable:	D.Sparkling	No. Observations:	131
Model:	ARIMA(1, 1, 2)	Log Likelihood	-1111.799
Method:	css-mle	S.D. of innovations	1155.290
Date:	Sun, 13 Sep 2020	AIC	2233.598
Time:	15:29:52	BIC	2247.974
Sample:	02-01-1980 - 12-01-1990	HQIC	2239.439

	coef	std err	z	P>|z|	[0.025	0.975]
const	6.4579	4.211	1.534	0.125	-1.795	14.711
ar.L1.D.Sparkling	0.1896	0.166	1.143	0.253	-0.135	0.515
ma.L1.D.Sparkling	-0.6951	0.153	-4.548	0.000	-0.995	-0.396
ma.L2.D.Sparkling	-0.3049	0.152	-2.009	0.045	-0.602	-0.007

Roots

	Real	Imaginary	Modulus	Frequency
AR.1	5.2740	+0.0000j	5.2740	0.0000
MA.1	1.0000	+0.0000j	1.0000	0.0000
MA.2	-3.2802	+0.0000j	3.2802	0.5000

- Manual Arima performed well for both the dataset compared to Automated Arima, in fact RMSE for rose dataset greatly reduced..

	Test RMSE-Rose Data
RegressionOnTime	1267.516575
Naive	17.740511
SimpleAvg	15.759889
2pointTrailingMovingAverageRose	11.529278
4pointTrailingMovingAverageRose	14.451433
6pointTrailingMovingAverageRose	14.566399
9pointTrailingMovingAverageRose	14.727667
Alpha=1,SimpleExponentialSmoothing	36.796465
Alpha=0.3,SimpleExponentialSmoothingRose	77.139491
Alpha=0.4,SimpleExponentialSmoothingRose	77.139491
Alpha=0.3,Beta=0.3HoltzModelRose	265.567857
Alpha=0.684,Beta=0.052,Gamma=0.315,TripleExponentialSmoothing	17.369752
ARIMA(0,0,2)	57.997243
Manual ARIMA(5,1,2)	15.422461

	Test RMSE-Spark Data
RegressionOnTime	1389.249047
Naive	1327.156057
SimpleAvg	1275.073380
2pointTrailingMovingAveragespark	813.400684
4pointTrailingMovingAveragespark	1156.589694
6pointTrailingMovingAveragespark	1283.927428
9pointTrailingMovingAveragespark	1346.278315
Alpha=1,SimpleExponentialSmoothing	1275.081839
Alpha=0.3,SimpleExponentialSmoothingspark	3686.794285
Alpha=0.4,SimpleExponentialSmoothingspark	3686.794285
Alpha=0.3,Beta=0.3HoltzModelspark	18259.110704
Alpha=0.15,Beta=2.68,Gamma=.37,TripleExponentialSmoothing	383.192343
ARIMA(2,0,2)	2777.091535
Manual ARIMA(1,1,2)	1436.723674

SARIMA model-Automated:

- For various values of p,q,d and P,D,Q,S of SARIMA model we could select one with lowest AIC in this case (1,1,2),(2,0,2,6) for Rose wine dataset(left) and (1,1,2), (2,0,2,6) for Sparkling wine data(right):

	param	seasonal	AIC
53	(1, 1, 2)	(2, 0, 2, 6)	1041.655817
26	(0, 1, 2)	(2, 0, 2, 6)	1043.600261
80	(2, 1, 2)	(2, 0, 2, 6)	1045.288112
71	(2, 1, 1)	(2, 0, 2, 6)	1051.673461
44	(1, 1, 1)	(2, 0, 2, 6)	1052.778469

	param	seasonal	AIC
53	(1, 1, 2)	(2, 0, 2, 6)	1727.510410
26	(0, 1, 2)	(2, 0, 2, 6)	1727.887985
17	(0, 1, 1)	(2, 0, 2, 6)	1741.647352
71	(2, 1, 1)	(2, 0, 2, 6)	1744.040750
80	(2, 1, 2)	(2, 0, 2, 6)	1752.673959

- Rose:

SARIMAX Results

Dep. Variable:	y	No. Observations:	132
Model:	SARIMAX(1, 1, 2)x(2, 0, 2, 6)	Log Likelihood	-512.828
Date:	Sun, 13 Sep 2020	AIC	1041.656
Time:	15:47:55	BIC	1063.685
Sample:	0	HQIC	1050.598
	- 132		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	-0.5939	0.152	-3.899	0.000	-0.892	-0.295
ma.L1	-0.1954	238.206	-0.001	0.999	-467.071	466.680
ma.L2	-0.8046	191.705	-0.004	0.997	-376.540	374.931
ar.S.L6	-0.0625	0.035	-1.763	0.078	-0.132	0.007
ar.S.L12	0.8451	0.039	21.884	0.000	0.769	0.921
ma.S.L6	0.2225	481.498	0.000	1.000	-943.497	943.942
ma.S.L12	-0.7774	374.375	-0.002	0.998	-734.540	732.985
sigma2	335.2087	1.86e+05	0.002	0.999	-3.65e+05	3.66e+05

Ljung-Box (Q):	15.89	Jarque-Bera (JB):	56.68
Prob(Q):	1.00	Prob(JB):	0.00
Heteroskedasticity (H):	0.47	Skew:	0.52
Prob(H) (two-sided):	0.02	Kurtosis:	6.26

- Sparkling:

SARIMAX Results

Dep. Variable:	y	No. Observations:	132
Model:	SARIMAX(1, 1, 2)x(2, 0, 2, 6)	Log Likelihood	-855.755
Date:	Sun, 13 Sep 2020	AIC	1727.510
Time:	15:48:18	BIC	1749.539
Sample:	0	HQIC	1736.453
	- 132		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	-0.6456	0.287	-2.250	0.024	-1.208	-0.083
ma.L1	-0.1063	0.251	-0.423	0.672	-0.599	0.386
ma.L2	-0.7008	0.203	-3.459	0.001	-1.098	-0.304
ar.S.L6	-0.0047	0.027	-0.176	0.861	-0.058	0.048
ar.S.L12	1.0362	0.018	56.111	0.000	1.000	1.072
ma.S.L6	0.4750	0.143	3.321	0.001	0.195	0.755
ma.S.L12	-0.9177	0.180	-5.098	0.000	-1.271	-0.565
sigma2	9.656e+04	2.3e+04	4.193	0.000	5.14e+04	1.42e+05

Ljung-Box (Q):	28.98	Jarque-Bera (JB):	25.23
Prob(Q):	0.90	Prob(JB):	0.00
Heteroskedasticity (H):	2.65	Skew:	0.46
Prob(H) (two-sided):	0.00	Kurtosis:	5.09

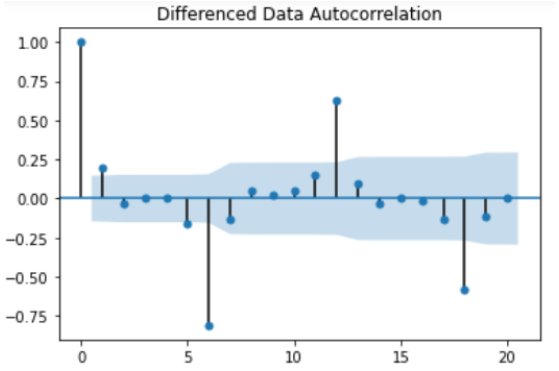
- Automated SARIMA performed well for both the dataset, in case of Rose dataset Manual ARIMA is still best performer, we will be not taking into consideration of Moving Average model since it does not take into consideration of any seasonal or trend parameter. In case of Sparkling wine dataset SARIMA performed, but still the winner is TES.

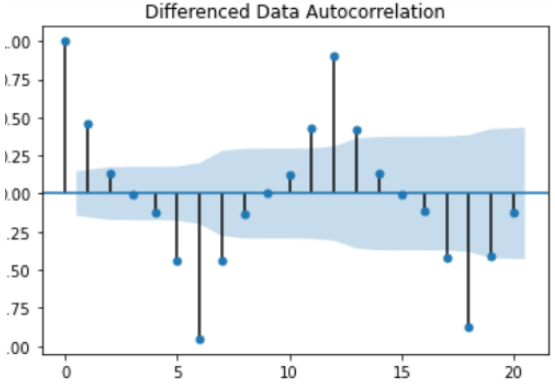
	Test RMSE-Rose Data
RegressionOnTime	1267.516575
Naive	17.740511
SimpleAvg	15.759889
2pointTrailingMovingAverageRose	11.529278
4pointTrailingMovingAverageRose	14.451433
6pointTrailingMovingAverageRose	14.566399
9pointTrailingMovingAverageRose	14.727667
Alpha=1,SimpleExponentialSmoothing	36.796465
Alpha=0.3,SimpleExponentialSmoothingRose	77.139491
Alpha=0.4,SimpleExponentialSmoothingRose	77.139491
Alpha=0.3,Beta=0.3HoltzModelRose	265.567857
Alpha=0.684,Beta=0.052,Gamma=0.315,TripleExponentialSmoothing	17.369752
ARIMA(0,0,2)	57.997243
Manual ARIMA(5,1,2)	15.422461
SARIMA(0,1,2)(2,0,2,6)	26.134000

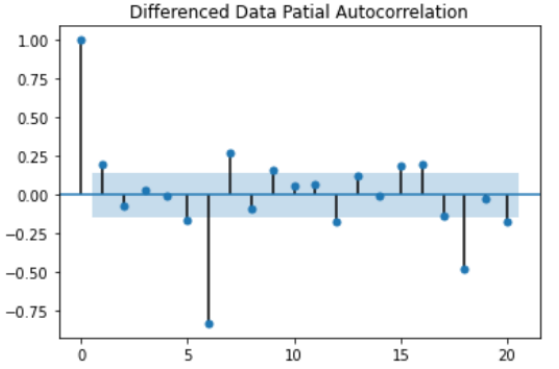
	Test RMSE-Spark Data
RegressionOnTime	1389.249047
Naive	1327.156057
SimpleAvg	1275.073380
2pointTrailingMovingAveragespark	813.400684
4pointTrailingMovingAveragespark	1156.589694
6pointTrailingMovingAveragespark	1283.927428
9pointTrailingMovingAveragespark	1346.278315
Alpha=1,SimpleExponentialSmoothing	1275.081839
Alpha=0.3,SimpleExponentialSmoothingspark	3686.794285
Alpha=0.4,SimpleExponentialSmoothingspark	3686.794285
Alpha=0.3,Beta=0.3HoltzModelspark	18259.110704
Alpha=0.15,Beta=2.68,Gamma=.37,TripleExponentialSmoothing	383.192343
ARIMA(2,0,2)	2777.091535
Manual ARIMA(1,1,2)	1436.723674
SARIMA(0,1,2)(2,0,2,6)	629.280166

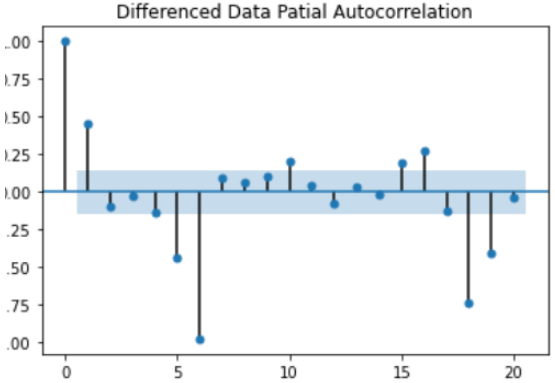
SARIMA model-Manual:

- For various values of p,q,d and P,D,Q,S of SARIMA model we could based on the PACF,ACF plot p,q,d derived from ACF plot we saw in during ARIMA model building .i.e (5,1,2) while Seasonal part will be (2,1,1,6):









- Rose:

SARIMAX Results

Dep. Variable:	y	No. Observations:	132
Model:	SARIMAX(5, 1, 2)x(2, 1, [1], 6)	Log Likelihood	-474.187
Date:	Sun, 13 Sep 2020	AIC	970.373
Time:	16:31:56	BIC	999.877
Sample:	0	HQIC	982.336
	- 132		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	-2.1969	0.182	-12.081	0.000	-2.553	-1.840
ar.L2	-1.9398	0.322	-6.021	0.000	-2.571	-1.308
ar.L3	-1.4611	0.309	-4.727	0.000	-2.067	-0.855
ar.L4	-1.1413	0.239	-4.778	0.000	-1.609	-0.673
ar.L5	-0.4052	0.099	-4.109	0.000	-0.598	-0.212
ma.L1	2.1872	0.504	4.343	0.000	1.200	3.174
ma.L2	1.1861	0.551	2.152	0.031	0.106	2.267
ar.S.L6	-0.6429	0.075	-8.519	0.000	-0.791	-0.495
ar.S.L12	0.0711	0.029	2.421	0.015	0.014	0.129
ma.S.L6	-0.3735	0.149	-2.499	0.012	-0.667	-0.081
sigma2	262.9410	175.111	1.502	0.133	-80.271	606.153

Ljung-Box (Q):	37.13	Jarque-Bera (JB):	3.40
Prob(Q):	0.60	Prob(JB):	0.18
Heteroskedasticity (H):	0.94	Skew:	0.42
Prob(H) (two-sided):	0.84	Kurtosis:	3.25

- Sparkling:

SARIMAX Results

Dep. Variable:	y	No. Observations:	132
Model:	SARIMAX(5, 1, 2)x(2, 1, [1], 6)	Log Likelihood	-806.603
Date:	Sun, 13 Sep 2020	AIC	1635.206
Time:	16:32:04	BIC	1664.710
Sample:	0	HQIC	1647.169
	- 132		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	-0.8685	0.159	-5.465	0.000	-1.180	-0.557
ar.L2	-0.8977	0.172	-5.232	0.000	-1.234	-0.561
ar.L3	-0.7449	0.187	-3.978	0.000	-1.112	-0.378
ar.L4	-0.7453	0.182	-4.096	0.000	-1.102	-0.389
ar.L5	-0.6272	0.190	-3.309	0.001	-0.999	-0.256
ma.L1	0.1678	0.168	1.001	0.317	-0.161	0.496
ma.L2	-0.0314	0.145	-0.216	0.829	-0.316	0.253
ar.S.L6	-0.8416	0.197	-4.282	0.000	-1.227	-0.456
ar.S.L12	0.1515	0.202	0.749	0.454	-0.245	0.548
ma.S.L6	-0.9848	0.957	-1.029	0.303	-2.861	0.891
sigma2	1.624e+05	1.46e+05	1.109	0.267	-1.25e+05	4.49e+05

Ljung-Box (Q):	24.48	Jarque-Bera (JB):	5.30
Prob(Q):	0.97	Prob(JB):	0.07
Heteroskedasticity (H):	0.88	Skew:	0.29
Prob(H) (two-sided):	0.71	Kurtosis:	3.92

- Manual SARIMA performed well for both the dataset, in case of Rose dataset Manual ARIMA is still best performer,In case of Sparkling wine dataset SARIMA performed well.

	Test RMSE-Rose Data
RegressionOnTime	1267.516575
Naive	17.740511
SimpleAvg	15.759889
2pointTrailingMovingAverageRose	11.529278
4pointTrailingMovingAverageRose	14.451433
6pointTrailingMovingAverageRose	14.566399
9pointTrailingMovingAverageRose	14.727667
Alpha=1,SimpleExponentialSmoothing	36.796465
Alpha=0.3,SimpleExponentialSmoothingRose	77.139491
Alpha=0.4,SimpleExponentialSmoothingRose	77.139491
Alpha=0.3,Beta=0.3HoltzModelRose	265.567857
Alpha=0.684,Beta=0.052,Gamma=0.315,TripleExponentialSmoothing	17.369752
ARIMA(0,0,2)	57.997243
Manual ARIMA(5,1,2)	15.422461
SARIMA(0,1,2)(2,0,2,6)	26.134000
Manual SARIMA(5,1,2)(2,1,1,6)	27.782109

	Test RMSE-Spark Data
RegressionOnTime	1389.249047
Naive	1327.156057
SimpleAvg	1275.073380
2pointTrailingMovingAveragespark	813.400684
4pointTrailingMovingAveragespark	1156.589694
6pointTrailingMovingAveragespark	1283.927428
9pointTrailingMovingAveragespark	1346.278315
Alpha=1,SimpleExponentialSmoothing	1275.081839
Alpha=0.3,SimpleExponentialSmoothingspark	3686.794285
Alpha=0.4,SimpleExponentialSmoothingspark	3686.794285
Alpha=0.3,Beta=0.3HoltzModelspark	18259.110704
Alpha=0.15,Beta=2.68,Gamma=.37,TripleExponentialSmoothing	383.192343
ARIMA(2,0,2)	2777.091535
Manual ARIMA(1,1,2)	1436.723674
SARIMA(0,1,2)(2,0,2,6)	629.280166
Manual SARIMA(1,1,2)(2,1,1,6)	305.533637

Best Optimum Model:

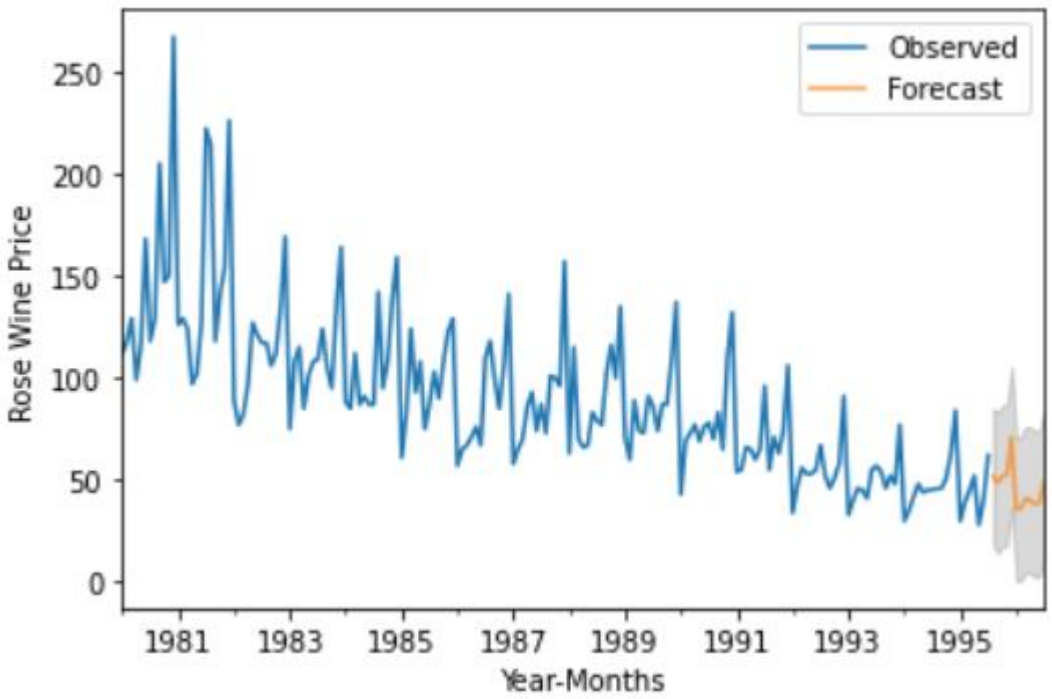
- From below summary we could choose Automated SARIMA for Rose Dataset(even though Manual Arima is a clear choice because of the business objective and data under consideration is ‘Wine’ we are going with SARIMA for the Rose data set , additionally there is no gradual difference between RMSE of Manual ARIMA and Auto SARIMA)and Manual SARIMA for Sparkling dataset for forecasting based on RMSE score:

	Test RMSE-Rose Data
RegressionOnTime	1267.516575
Naive	17.740511
SimpleAvg	15.759889
2pointTrailingMovingAverageRose	11.529278
4pointTrailingMovingAverageRose	14.451433
6pointTrailingMovingAverageRose	14.566399
9pointTrailingMovingAverageRose	14.727667
Alpha=1,SimpleExponentialSmoothing	36.796465
Alpha=0.3,SimpleExponentialSmoothingRose	77.139491
Alpha=0.4,SimpleExponentialSmoothingRose	77.139491
Alpha=0.3,Beta=0.3HoltzModelRose	265.567857
Alpha=0.684,Beta=0.052,Gamma=0.315,TripleExponentialSmoothing	17.369752
ARIMA(0,0,2)	57.997243
Manual ARIMA(5,1,2)	15.422461
SARIMA(0,1,2)(2,0,2,6)	26.134000
Manual SARIMA(5,1,2)(2,1,1,6)	27.782109

	Test RMSE-Spark Data
RegressionOnTime	1389.249047
Naive	1327.156057
SimpleAvg	1275.073380
2pointTrailingMovingAveragespark	813.400684
4pointTrailingMovingAveragespark	1156.589694
6pointTrailingMovingAveragespark	1283.927428
9pointTrailingMovingAveragespark	1346.278315
Alpha=1,SimpleExponentialSmoothing	1275.081839
Alpha=0.3,SimpleExponentialSmoothingspark	3686.794285
Alpha=0.4,SimpleExponentialSmoothingspark	3686.794285
Alpha=0.3,Beta=0.3HoltzModelspark	18259.110704
Alpha=0.15,Beta=2.68,Gamma=.37,TripleExponentialSmoothing	383.192343
ARIMA(2,0,2)	2777.091535
Manual ARIMA(1,1,2)	1436.723674
SARIMA(0,1,2)(2,0,2,6)	629.280166
Manual SARIMA(1,1,2)(2,1,1,6)	305.533637

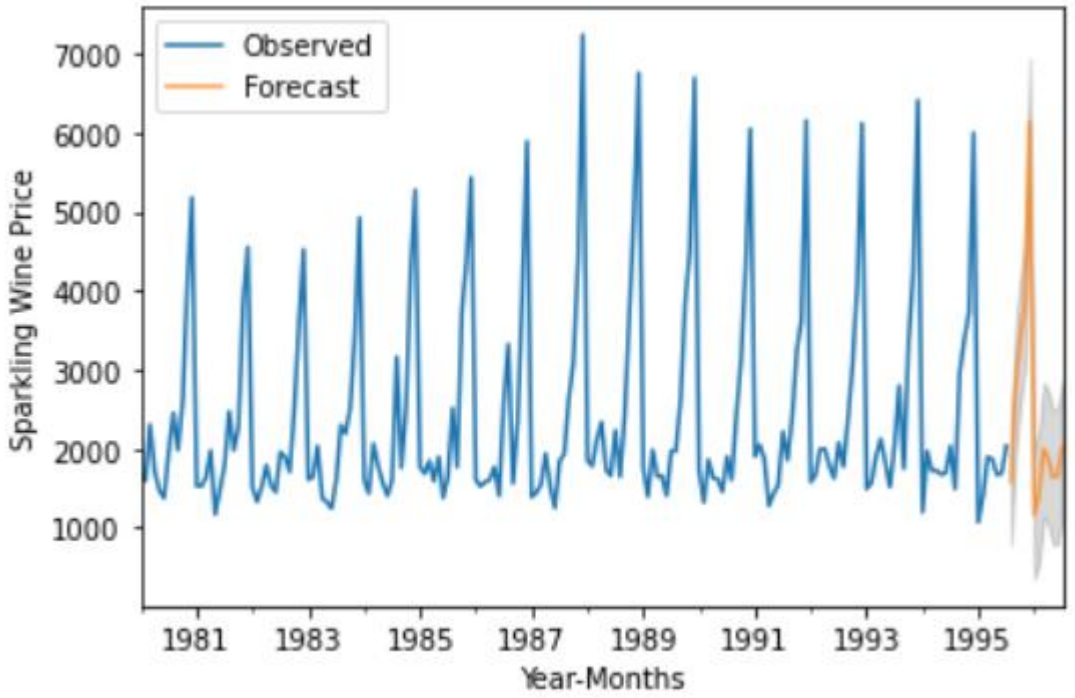
- Rose wine prices forecasted from 1995-08-01 to 1996-08-01:

RMSE of the Final Model 28.051000974535224



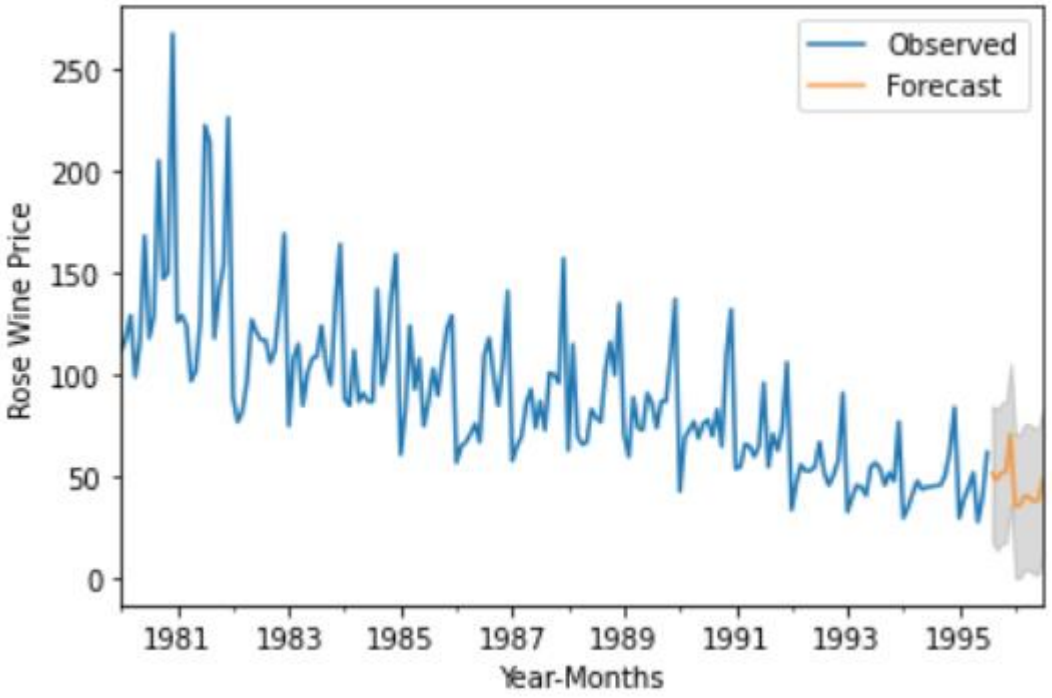
- Sparkling wine prices forecasted from 1995-08-01 to 1996-08-01:

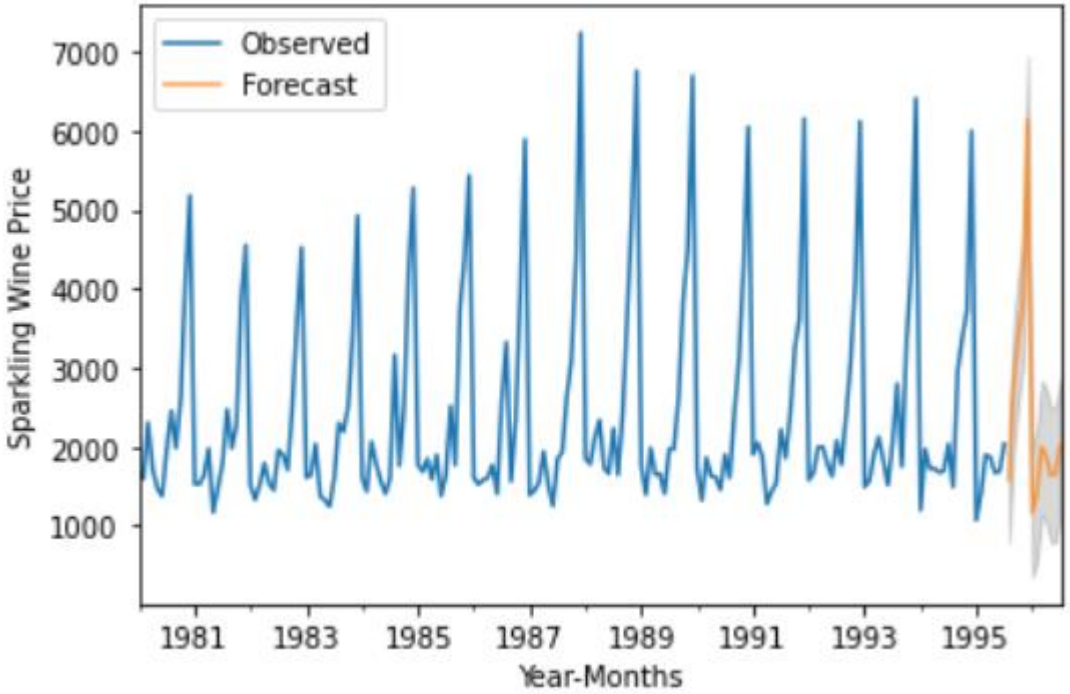
RMSE of the Final Model 600.572522314169



Conclusion:

- We could see for Rose wine price forecast, prices are still dwindling down. It is high time Manufacture process and Quality is put under strict supervision, maybe People's taste may have changed , it is a must do list to obtain information on customer's developed interest and cater to their needs.





- As for Sparkling wine prices do not show any obvious trend, it does show seasonality component. Highlight is it has maintained constant average Price over the years, with the projection just 12 months we are unable confidently say that its prices are dropping , we might need 2 or 3 years of projection to give a confident insight.

THE END