Project

Karthiheswar

Table of Contents

1 Project Objective

The objective of the report is to explore the data of different types of wine sales in the 20th century in ABC Estate Wines. The project data sets are analyzed in Python and generate insights and forecast about the data set. This exploration report will consist of the following:

- Importing the dataset in Python
- Understanding the structure of dataset
- Checking null values
- Univariate Analysis
- Decomposition on data sets
- Applying Exponential Smoothing methods and ARIMA/SARIMA models
- Calculating RMSE values on test data
- Forecast the data
- Insights from the dataset

2 Time Series Forecasting on Sparkling and Rose Wine Sales

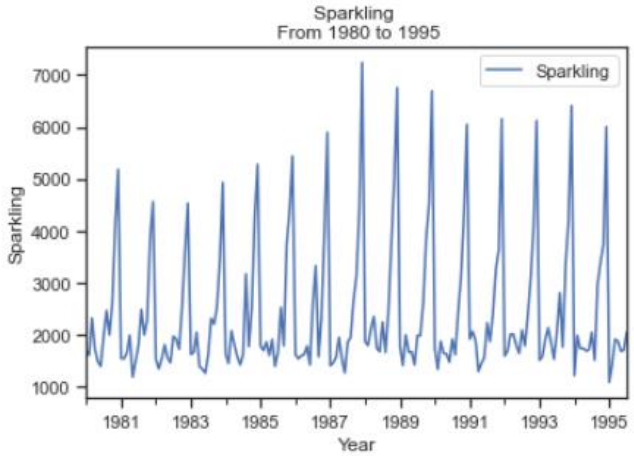
2.1 Reading the data as an appropriate Time Series data and plotting the data

SPARKLING

Reading the dataset (head)

YearMonth	Sparkling
1980-01-31	1686
1980-02-29	1591
1980-03-31	2304
1980-04-30	1712
1980-05-31	1471

Plotting the data

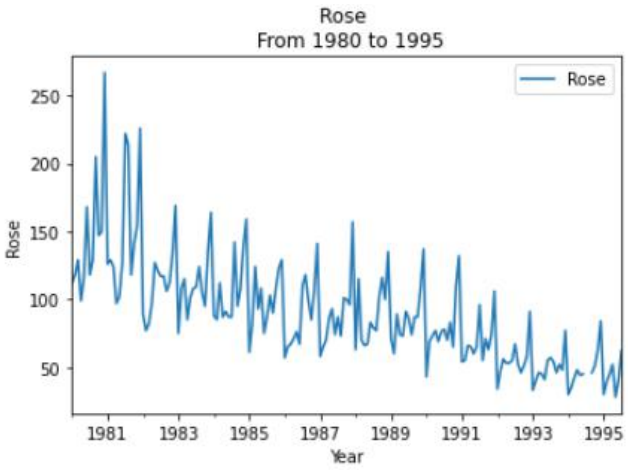


ROSE

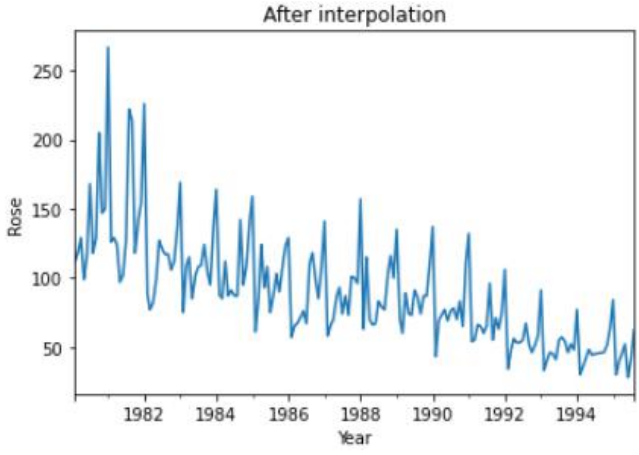
Reading the dataset (head)

YearMonth	Rose
1980-01-31	112.0
1980-02-29	118.0
1980-03-31	129.0
1980-04-30	99.0
1980-05-31	116.0

Plotting the data



Plotting the data after interpolating the missing values



2.2 Performing appropriate Exploratory Data Analysis to understand the data and also performing decomposition

SPARKLING EDA

Checking the null values

There are no null values present

Checking the Info

```
DatetimeIndex: 187 entries, 1980-01-31 to 1995-07-31
Data columns (total 1 columns):
 #   Column     Non-Null Count  Dtype
---  ------     --------------  -----
 0   Sparkling  187 non-null    int64
dtypes: int64(1)
memory usage: 2.9 KB
```

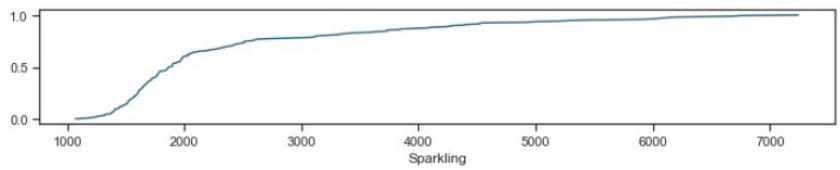
Checking the Shape

The number of rows: 187
The number of columns: 1

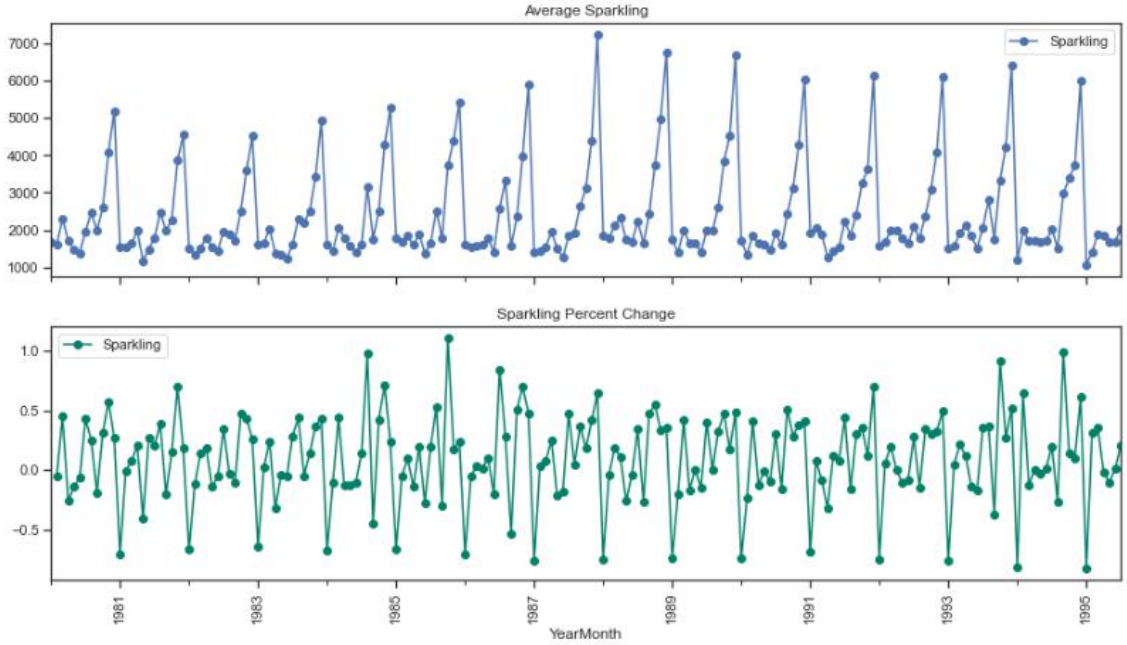
Describing the data

	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

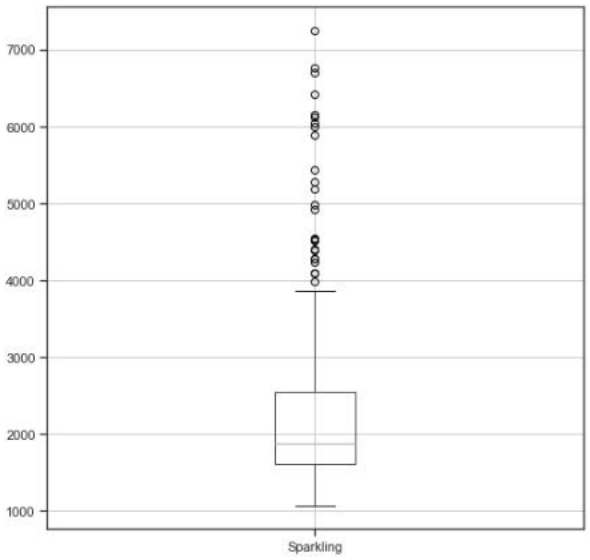
Empirical Cumulative Distribution Function



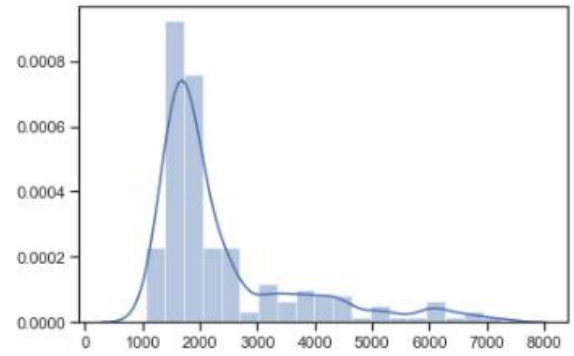
Average Sparkling and its Percent change



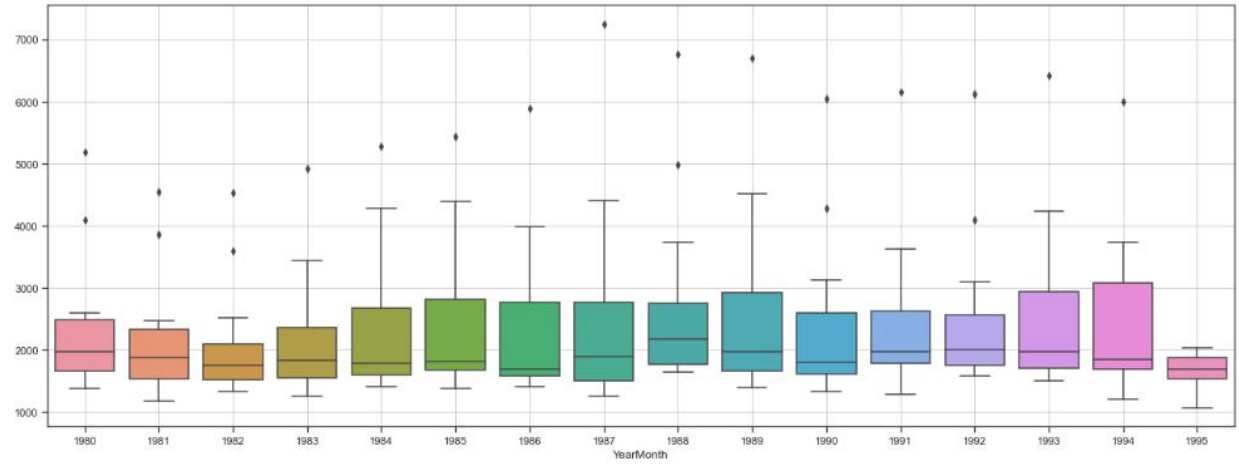
Box plot



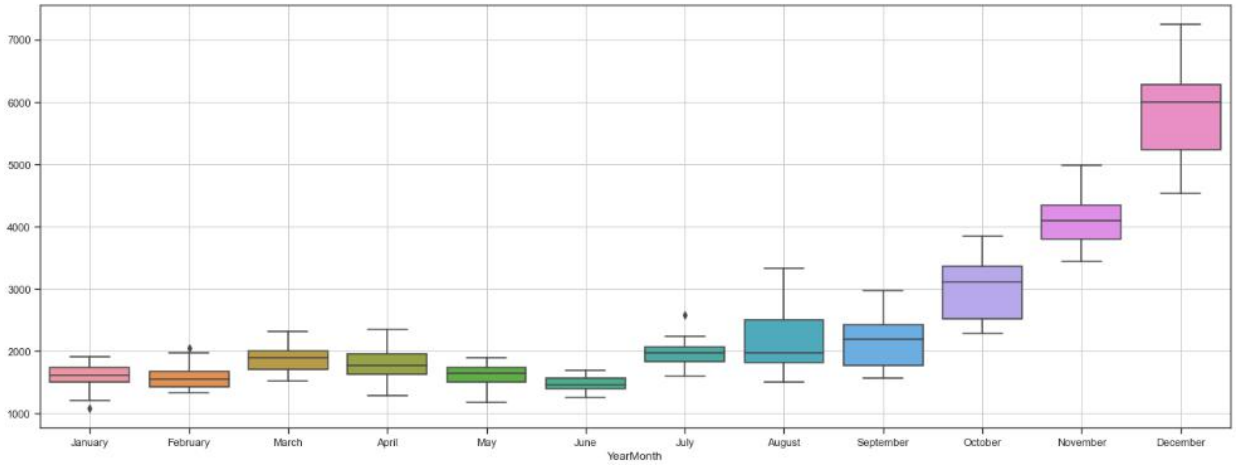
Histogram



Yearly box plot



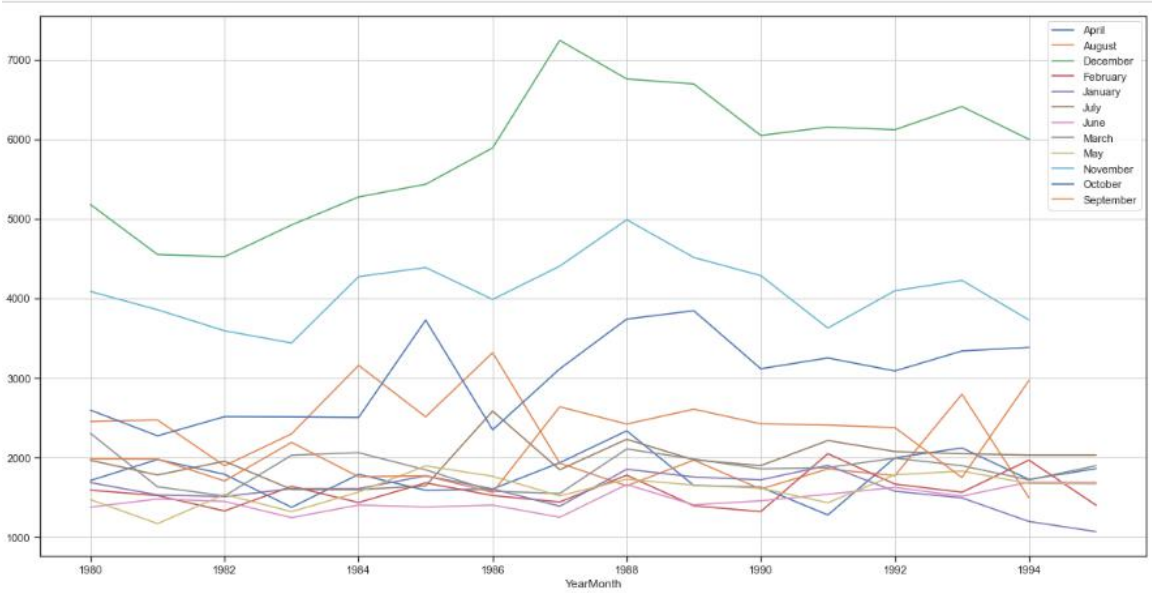
Monthly box plot



Pivot table

YearMonth	April	August	December	February	January	July	June	March	May	November	October	September
YearMonth												
1980	1712.0	2453.0	5179.0	1591.0	1686.0	1966.0	1377.0	2304.0	1471.0	4087.0	2596.0	1984.0
1981	1976.0	2472.0	4551.0	1523.0	1530.0	1781.0	1480.0	1633.0	1170.0	3857.0	2273.0	1981.0
1982	1790.0	1897.0	4524.0	1329.0	1510.0	1954.0	1449.0	1518.0	1537.0	3593.0	2514.0	1706.0
1983	1375.0	2298.0	4923.0	1638.0	1609.0	1600.0	1245.0	2030.0	1320.0	3440.0	2511.0	2191.0
1984	1789.0	3159.0	5274.0	1435.0	1609.0	1597.0	1404.0	2061.0	1567.0	4273.0	2504.0	1759.0
1985	1589.0	2512.0	5434.0	1682.0	1771.0	1645.0	1379.0	1846.0	1896.0	4388.0	3727.0	1771.0
1986	1605.0	3318.0	5891.0	1523.0	1606.0	2584.0	1403.0	1577.0	1765.0	3987.0	2349.0	1562.0
1987	1935.0	1930.0	7242.0	1442.0	1389.0	1847.0	1250.0	1548.0	1518.0	4405.0	3114.0	2638.0
1988	2336.0	1645.0	6757.0	1779.0	1853.0	2230.0	1661.0	2108.0	1728.0	4988.0	3740.0	2421.0
1989	1650.0	1968.0	6694.0	1394.0	1757.0	1971.0	1406.0	1982.0	1654.0	4514.0	3845.0	2608.0
1990	1628.0	1605.0	6047.0	1321.0	1720.0	1899.0	1457.0	1859.0	1615.0	4286.0	3116.0	2424.0
1991	1279.0	1857.0	6153.0	2049.0	1902.0	2214.0	1540.0	1874.0	1432.0	3627.0	3252.0	2408.0
1992	1997.0	1773.0	6119.0	1667.0	1577.0	2076.0	1625.0	1993.0	1783.0	4096.0	3088.0	2377.0
1993	2121.0	2795.0	6410.0	1564.0	1494.0	2048.0	1515.0	1898.0	1831.0	4227.0	3339.0	1749.0
1994	1725.0	1495.0	5999.0	1968.0	1197.0	2031.0	1693.0	1720.0	1674.0	3729.0	3385.0	2968.0
1995	1862.0	NaN	NaN	1402.0	1070.0	2031.0	1688.0	1897.0	1670.0	NaN	NaN	NaN

Monthly Sparkling across Years



ROSE EDA

Checking the null values

There are 2 null values present, so the null values are imputed with linear interpolation

```
Imputed value
YearMonth
1994-07-01    45.010870
1994-07-02    45.021739
1994-07-03    45.032609
1994-07-04    45.043478
1994-07-05    45.054348
Freq: D, Name: Rose, dtype: float64
Imputed value
YearMonth
1994-08-01    45.347826
1994-08-02    45.358696
1994-08-03    45.369565
1994-08-04    45.380435
1994-08-05    45.391304
Freq: D, Name: Rose, dtype: float64
```

Checking the Info

```
DatetimeIndex: 187 entries, 1980-01-31 to 1995-07-31
Data columns (total 1 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   Rose    187 non-null    float64
dtypes: float64(1)
memory usage: 7.9 KB
```

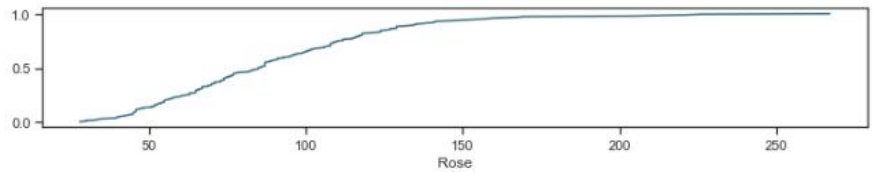
Checking the Shape

The number of rows: 187
The number of columns: 1

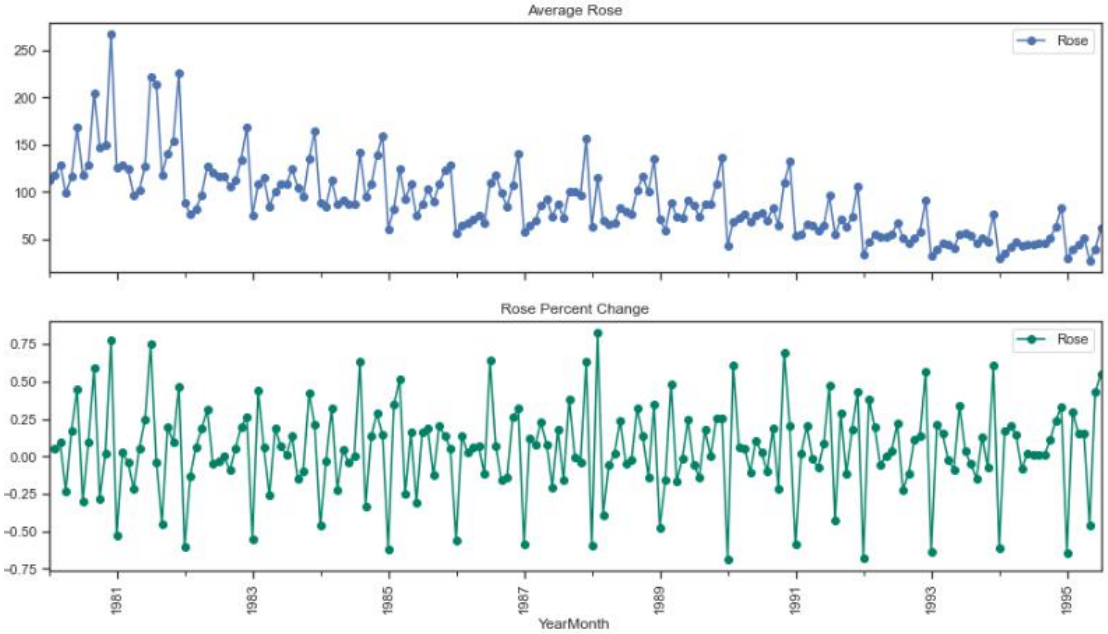
Describing the data

	Rose
count	187.000000
mean	89.914497
std	39.238259
min	28.000000
25%	62.500000
50%	85.000000
75%	111.000000
max	267.000000

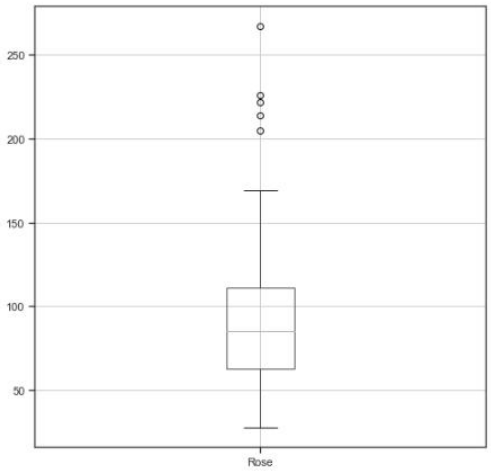
Empirical Cumulative Distribution Function



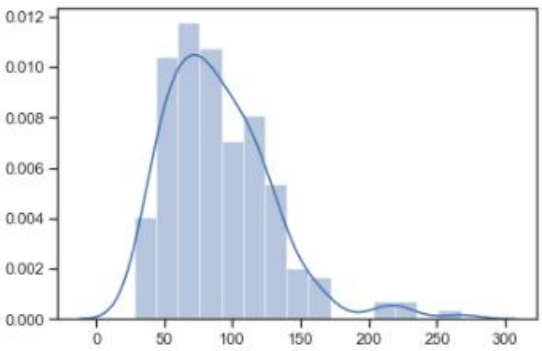
Average Sparkling and its Percent change



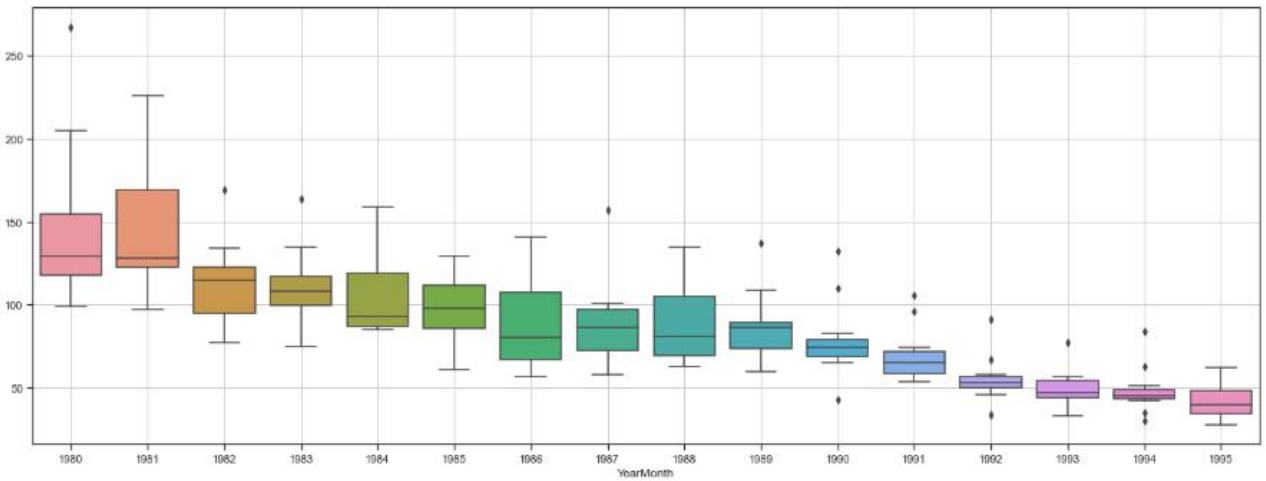
Box plot



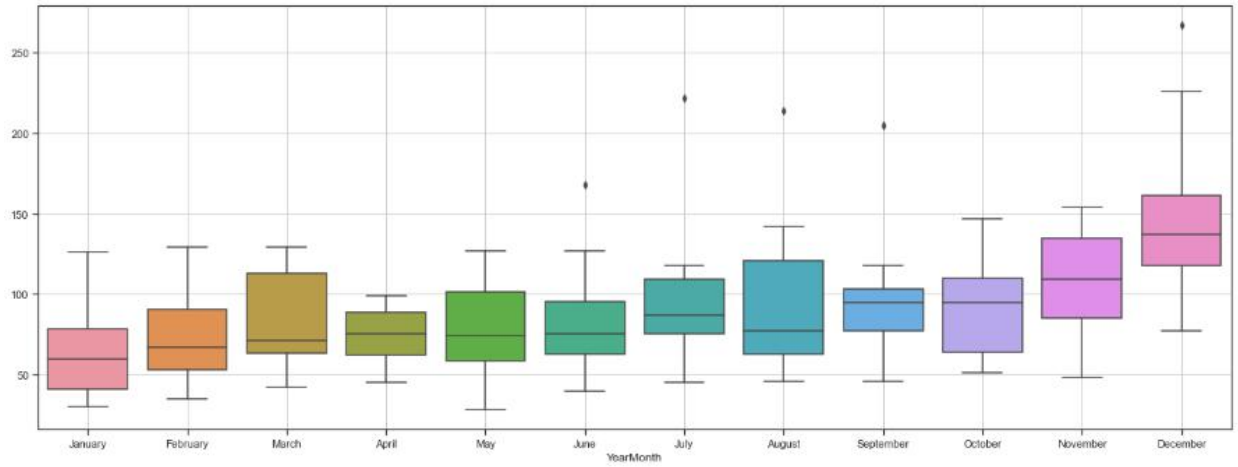
Histogram



Yearly box plot



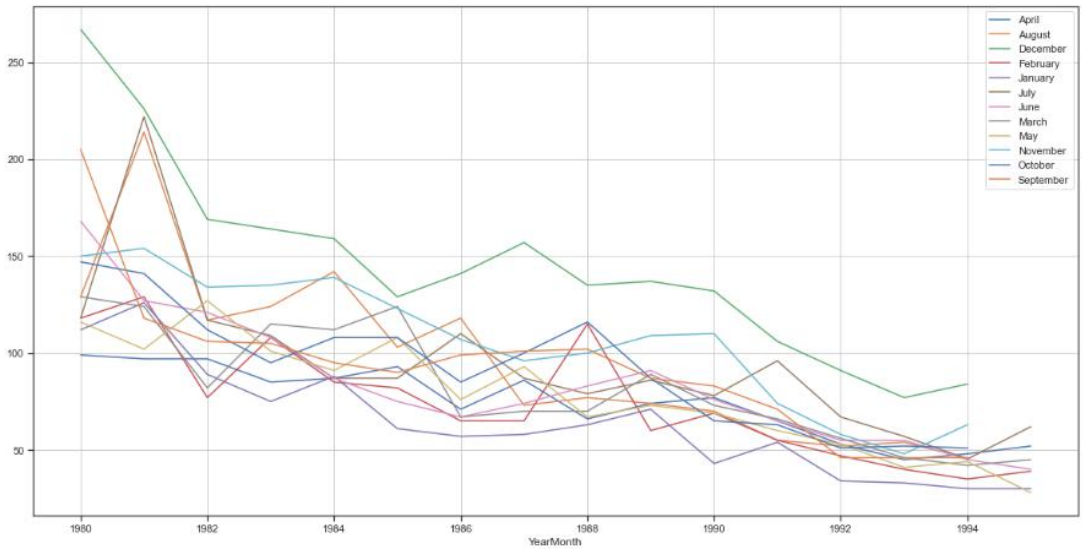
Monthly box plot



Pivot table

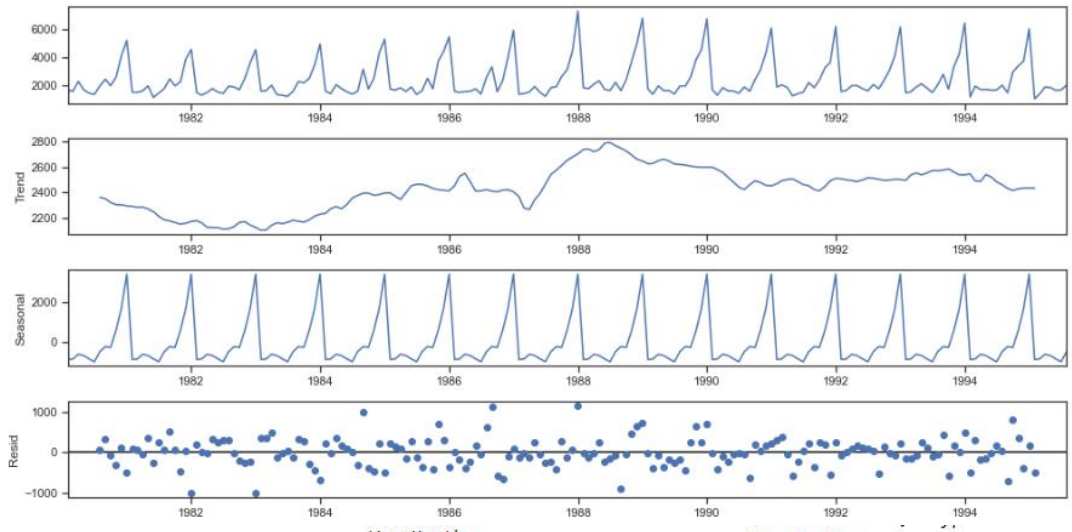
YearMonth	April	August	December	February	January	July	June	March	May	November	October	September
YearMonth												
1980	99.0	129.000000	267.0	118.0	112.0	118.000000	168.0	129.0	116.0	150.0	147.0	205.0
1981	97.0	214.000000	226.0	129.0	126.0	222.000000	127.0	124.0	102.0	154.0	141.0	118.0
1982	97.0	117.000000	169.0	77.0	89.0	117.000000	121.0	82.0	127.0	134.0	112.0	106.0
1983	85.0	124.000000	164.0	108.0	75.0	109.000000	108.0	115.0	101.0	135.0	95.0	105.0
1984	87.0	142.000000	159.0	85.0	88.0	87.000000	87.0	112.0	91.0	139.0	108.0	95.0
1985	93.0	103.000000	129.0	82.0	61.0	87.000000	75.0	124.0	108.0	123.0	108.0	90.0
1986	71.0	118.000000	141.0	65.0	57.0	110.000000	67.0	67.0	76.0	107.0	85.0	99.0
1987	86.0	73.000000	157.0	65.0	58.0	87.000000	74.0	70.0	93.0	96.0	100.0	101.0
1988	66.0	77.000000	135.0	115.0	63.0	79.000000	83.0	70.0	67.0	100.0	116.0	102.0
1989	74.0	74.000000	137.0	60.0	71.0	86.000000	91.0	89.0	73.0	109.0	87.0	87.0
1990	77.0	70.000000	132.0	69.0	43.0	78.000000	76.0	73.0	69.0	110.0	65.0	83.0
1991	65.0	55.000000	106.0	55.0	54.0	96.000000	65.0	66.0	60.0	74.0	63.0	71.0
1992	53.0	52.000000	91.0	47.0	34.0	67.000000	55.0	56.0	53.0	58.0	51.0	46.0
1993	45.0	54.000000	77.0	40.0	33.0	57.000000	55.0	46.0	41.0	48.0	52.0	46.0
1994	48.0	45.673913	84.0	35.0	30.0	45.336957	45.0	42.0	44.0	63.0	51.0	46.0
1995	52.0	NaN	NaN	39.0	30.0	62.000000	40.0	45.0	28.0	NaN	NaN	NaN

Monthly Sparkling across Years



SPARKLING DECOMPOSITION

Additive decomposition

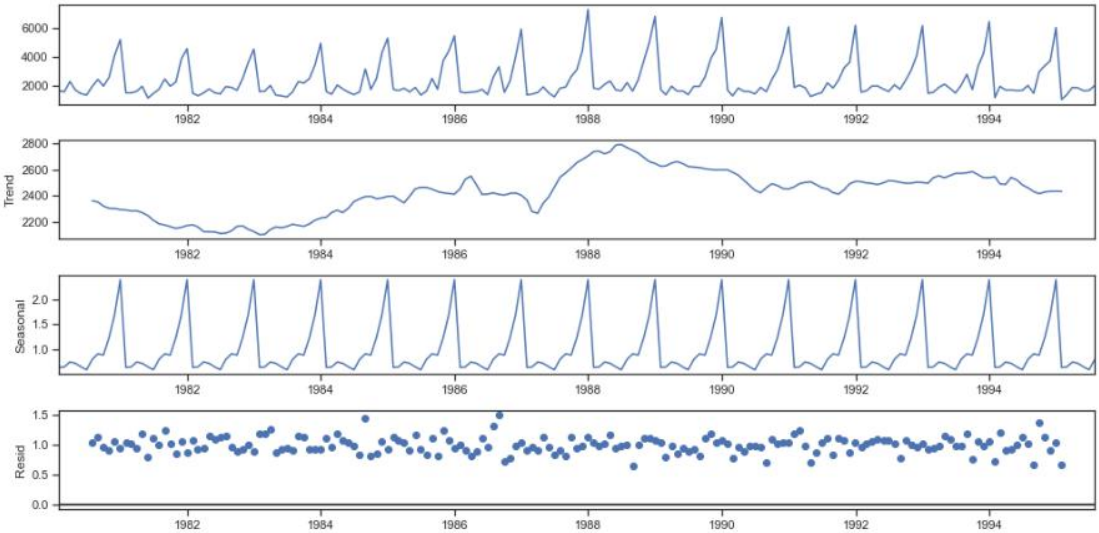



```
YearMonth
1980-01-31           NaN
1980-02-29           NaN
1980-03-31           NaN
1980-04-30           NaN
1980-05-31           NaN
1980-06-30           NaN
1980-07-31    2360.666667
1980-08-31    2351.333333
1980-09-30    2320.541667
1980-10-31    2303.583333
1980-11-30    2302.041667
1980-12-31    2293.791667
Name: trend, dtype: float64
```

```
YearMonth
1980-01-31    -854.260599
1980-02-29    -830.350678
1980-03-31    -592.356630
1980-04-30    -658.490559
1980-05-31    -824.416154
1980-06-30    -967.434011
1980-07-31    -465.502265
1980-08-31    -214.332821
1980-09-30    -254.677265
1980-10-31     599.769957
1980-11-30    1675.067179
1980-12-31    3386.983846
Name: seasonal, dtype: float64
```

```
YearMonth
1980-01-31           NaN
1980-02-29           NaN
1980-03-31           NaN
1980-04-30           NaN
1980-05-31           NaN
1980-06-30           NaN
1980-07-31     70.835599
1980-08-31    315.999487
1980-09-30    -81.864401
1980-10-31   -307.353290
1980-11-30    109.891154
1980-12-31   -501.775513
Name: resid, dtype: float64
```

Multiplicative decomposition



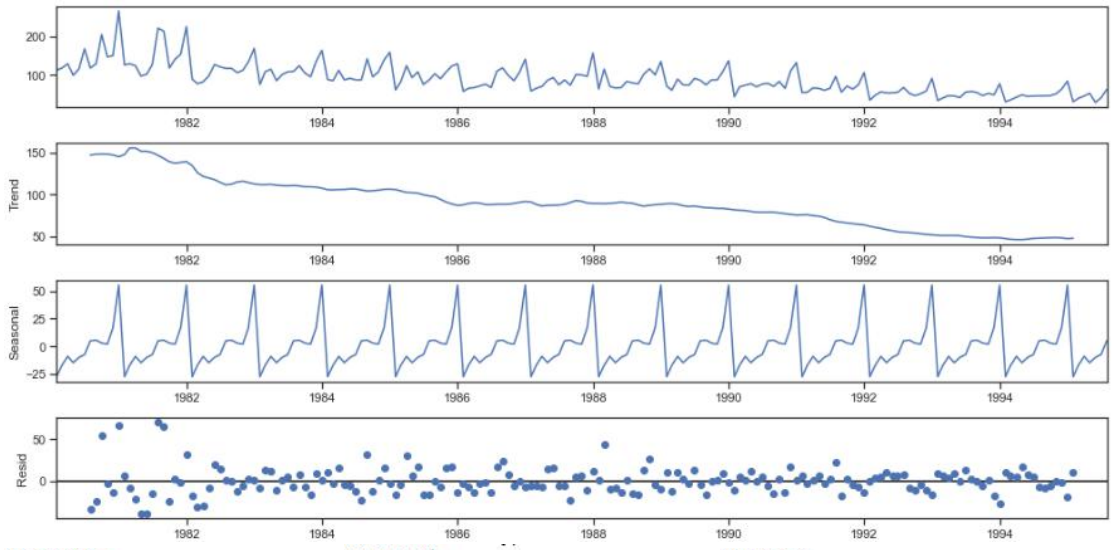

```
YearMonth
1980-01-31           NaN
1980-02-29           NaN
1980-03-31           NaN
1980-04-30           NaN
1980-05-31           NaN
1980-06-30           NaN
1980-07-31    2360.666667
1980-08-31    2351.333333
1980-09-30    2320.541667
1980-10-31    2303.583333
1980-11-30    2302.041667
1980-12-31    2293.791667
Name: trend, dtype: float64
```

```
YearMonth
1980-01-31    0.649843
1980-02-29    0.659214
1980-03-31    0.757440
1980-04-30    0.730351
1980-05-31    0.660609
1980-06-30    0.603468
1980-07-31    0.809164
1980-08-31    0.918822
1980-09-30    0.894367
1980-10-31    1.241789
1980-11-30    1.690158
1980-12-31    2.384776
Name: seasonal, dtype: float64
```

```
YearMonth
1980-01-31         NaN
1980-02-29         NaN
1980-03-31         NaN
1980-04-30         NaN
1980-05-31         NaN
1980-06-30         NaN
1980-07-31    1.029230
1980-08-31    1.135407
1980-09-30    0.955954
1980-10-31    0.907513
1980-11-30    1.050423
1980-12-31    0.946770
Name: resid, dtype: float64
```

ROSE DECOMPOSITION

Additive decomposition



YearMonth	
1980-01-31	NaN
1980-02-29	NaN
1980-03-31	NaN
1980-04-30	NaN
1980-05-31	NaN
1980-06-30	NaN
1980-07-31	147.083333
1980-08-31	148.125000
1980-09-30	148.375000
1980-10-31	148.083333
1980-11-30	147.416667
1980-12-31	145.125000

Name: trend, dtype: float64

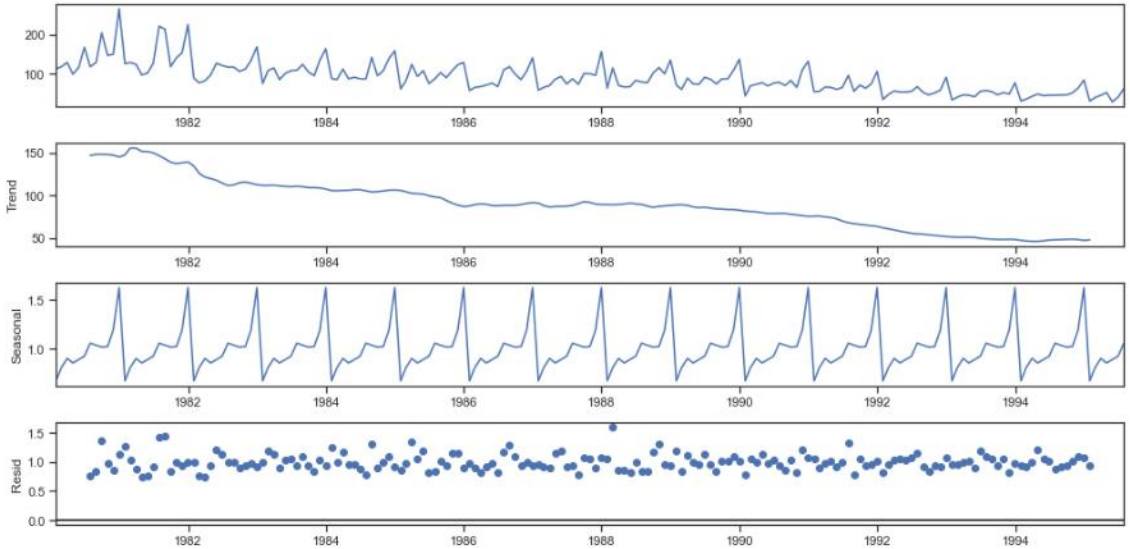
YearMonth	
1980-01-31	-27.908708
1980-02-29	-17.435675
1980-03-31	-9.285895
1980-04-30	-15.098395
1980-05-31	-10.196609
1980-06-30	-7.678752
1980-07-31	4.897089
1980-08-31	5.500109
1980-09-30	2.774625
1980-10-31	1.871848
1980-11-30	16.846848
1980-12-31	55.713514

Name: seasonal, dtype: float64

YearMonth	
1980-01-31	NaN
1980-02-29	NaN
1980-03-31	NaN
1980-04-30	NaN
1980-05-31	NaN
1980-06-30	NaN
1980-07-31	-33.980423
1980-08-31	-24.625109
1980-09-30	53.850375
1980-10-31	-2.955181
1980-11-30	-14.263514
1980-12-31	66.161486

Name: resid, dtype: float64

Multiplicative decomposition



YearMonth	
1980-01-31	NaN
1980-02-29	NaN
1980-03-31	NaN
1980-04-30	NaN
1980-05-31	NaN
1980-06-30	NaN
1980-07-31	147.083333
1980-08-31	148.125000
1980-09-30	148.375000
1980-10-31	148.083333
1980-11-30	147.416667
1980-12-31	145.125000

Name: trend, dtype: float64

YearMonth	
1980-01-31	0.670111
1980-02-29	0.806163
1980-03-31	0.901163
1980-04-30	0.854023
1980-05-31	0.889414
1980-06-30	0.923984
1980-07-31	1.058042
1980-08-31	1.035890
1980-09-30	1.017647
1980-10-31	1.022572
1980-11-30	1.192347
1980-12-31	1.628644

Name: seasonal, dtype: float64

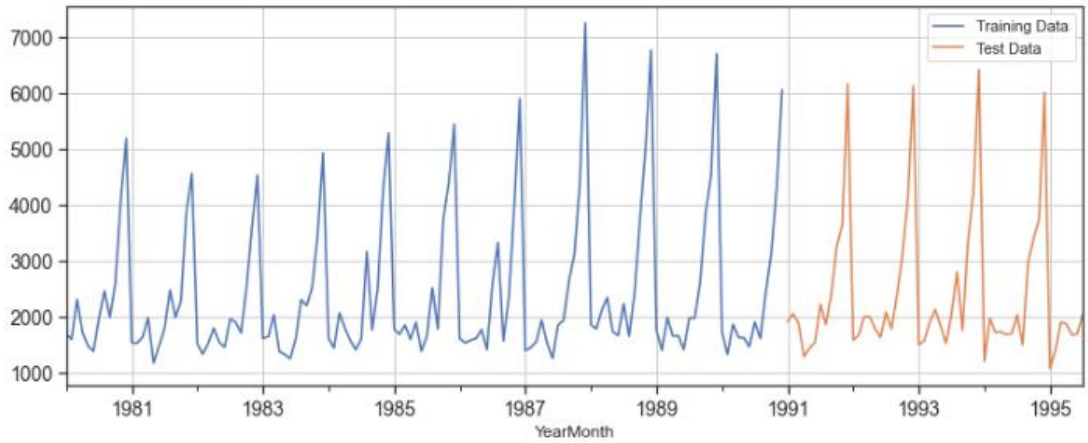
YearMonth	
1980-01-31	NaN
1980-02-29	NaN
1980-03-31	NaN
1980-04-30	NaN
1980-05-31	NaN
1980-06-30	NaN
1980-07-31	0.758256
1980-08-31	0.840713
1980-09-30	1.357675
1980-10-31	0.970772
1980-11-30	0.853379
1980-12-31	1.129647

Name: resid, dtype: float64

- Since both the data sets seasonality varies with time, both the data sets are considered as multiplicative seasonality

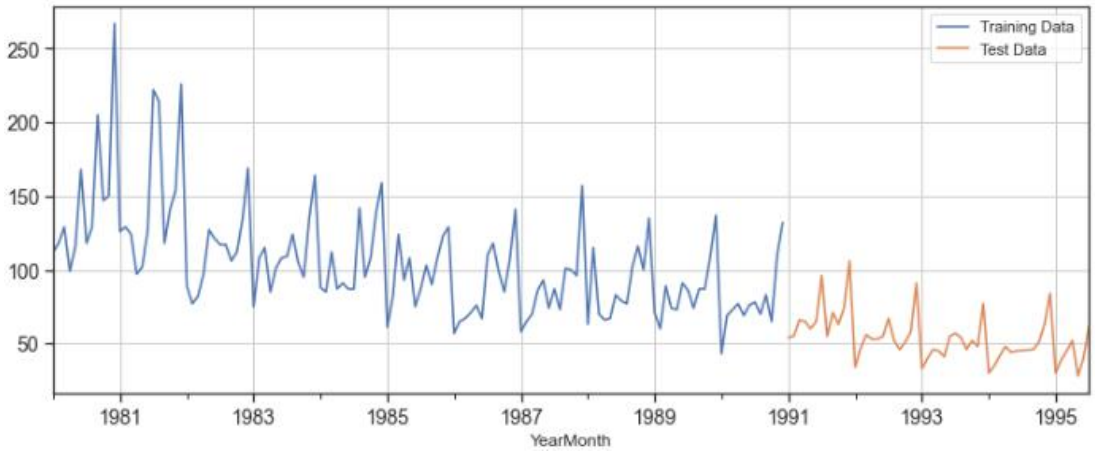
2.3 Spliting the data into training and test. The test data starts from 1991

SPARKLING



Shape of train and test set are (132, 1) and (55, 1)

ROSE



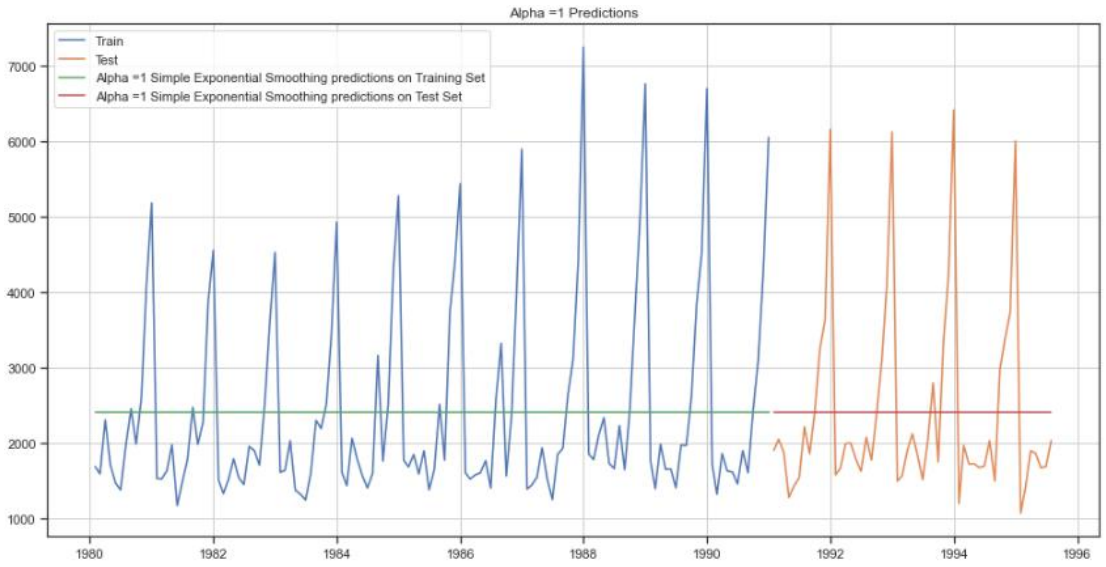
Shape of train and test set are (132, 1) and (55, 1)

2.4 Building various models on training data and evaluating model using RMSE on test data

SPARKLING

Simple Exponential Smoothing on auto parameters

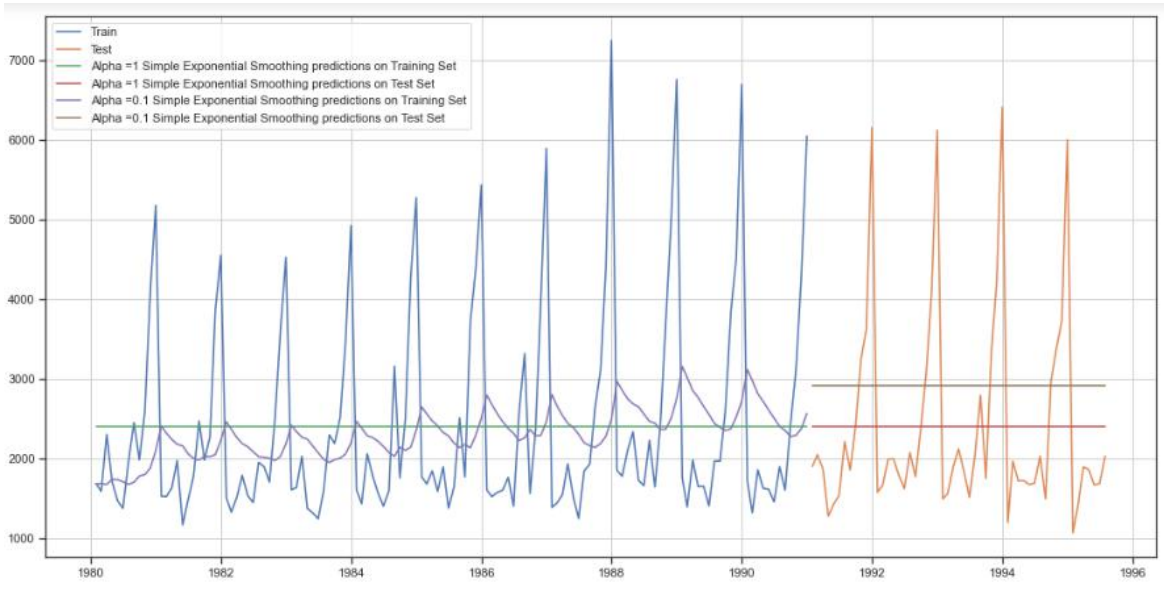
	name	param	optimized
smoothing_level	alpha	0.000000	True
initial_level	l.0	2403.785622	True



For Alpha =1 Simple Exponential Smoothing Model forecast on Training Data, RMSE is 1298.48 MAPE is 40.36
For Alpha =1 Simple Exponential Smoothing Model forecast on Testing Data, RMSE is 1275.08 MAPE is 38.90

Simple Exponential Smoothing on iteration

	Alpha Values	Train RMSE	Train MAPE	Test RMSE	Test MAPE
0	0.1	1333.873836	42.03	1375.393398	49.53
1	0.2	1356.042987	43.79	1595.206839	60.46
2	0.3	1359.511747	43.73	1935.507132	75.66
3	0.4	1352.588879	42.75	2311.919615	91.55
4	0.5	1344.004369	41.16	2666.351413	106.27
5	0.6	1338.805381	39.80	2979.204388	118.77
6	0.7	1338.844308	38.55	3249.944092	129.34
7	0.8	1344.462091	37.60	3483.801006	138.34
8	0.9	1355.723518	36.79	3686.794285	146.08



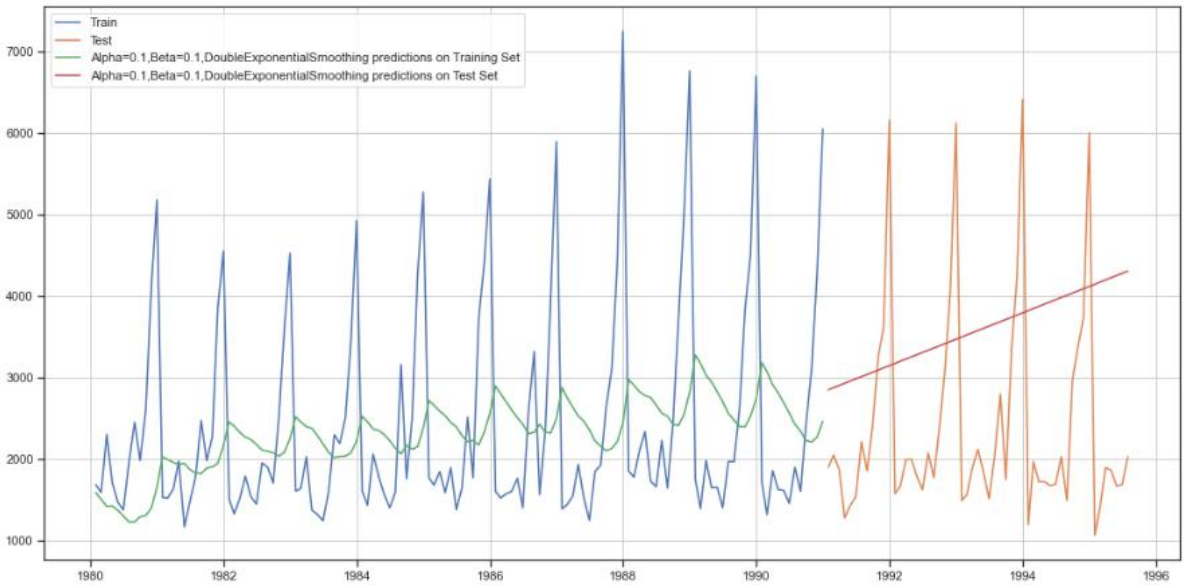
For Alpha =0.1 Simple Exponential Smoothing Model forecast on Testing Data, RMSE is 1375.39 MAPE is 49.53

Double Exponential Smoothing on auto parameters

	name	param	optimized
smoothing_level	alpha	0.647792	True
smoothing_slope	beta	0.000000	True
initial_level	l.0	1686.083777	True
initial_slope	b.0	27.059653	True

Double Exponential Smoothing on iteration

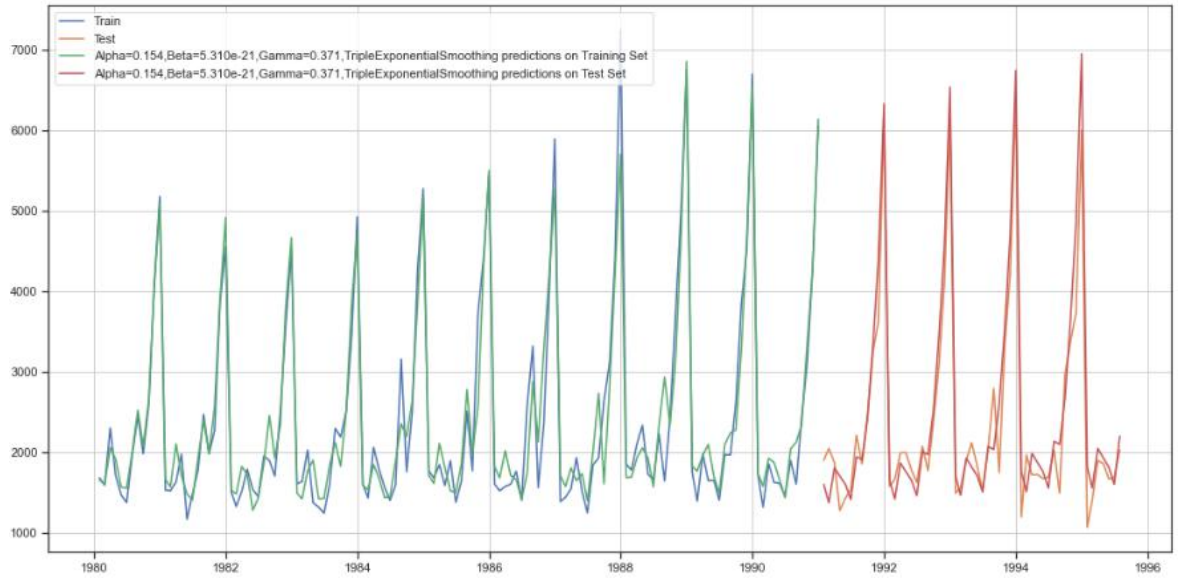
	Alpha Values	Beta Values	Train RMSE	Train MAPE	Test RMSE	Test MAPE
0	0.1	0.1	1382.520870	44.37	1778.564670	67.20
1	0.1	0.2	1413.598835	46.14	2599.439986	95.43
10	0.2	0.1	1418.041591	46.65	3611.763322	135.41
2	0.1	0.3	1445.762015	47.43	4293.084674	155.43
20	0.3	0.1	1431.169601	46.99	5908.185554	223.50



For Alpha=0.1, Beta=0.1 DES Model forecast on Testing Data, RMSE is 1778.56 MAPE is 67.20

Triple Exponential Smoothing on auto parameters

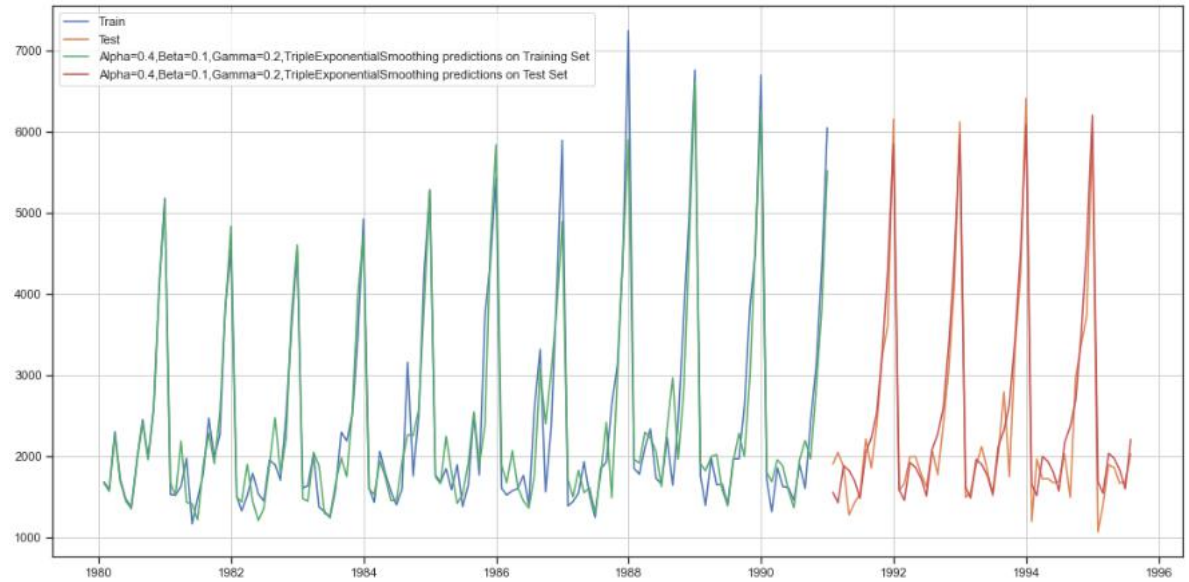
{'smoothing_level': 0.15420626100710186,
'smoothing_slope': 5.31004588730174e-21,
'smoothing_seasonal': 0.3713228125445634,
'damping_slope': nan,
'initial_level': 1639.99933972618,
'initial_slope': 4.847695527861546,
'initial_seasons': array([1.00842978, 0.96899514, 1.24181168, 1.13206716, 0.93979295,
0.93811177, 1.22457951, 1.54428109, 1.27335495, 1.631977 ,
2.48292059, 3.11861503]),
'use_boxcox': False,
'lamda': None,
'remove_bias': False}



For Alpha=0.154,Beta=5.310e-21,Gamma=0.371 TES Model forecast on the Training Data, RMSE is 353.37 MAPE is 10.17
For Alpha=0.154,Beta=5.310e-21,Gamma=0.371 TES Model forecast on the Testing Data, RMSE is 383.15 MAPE is 11.91

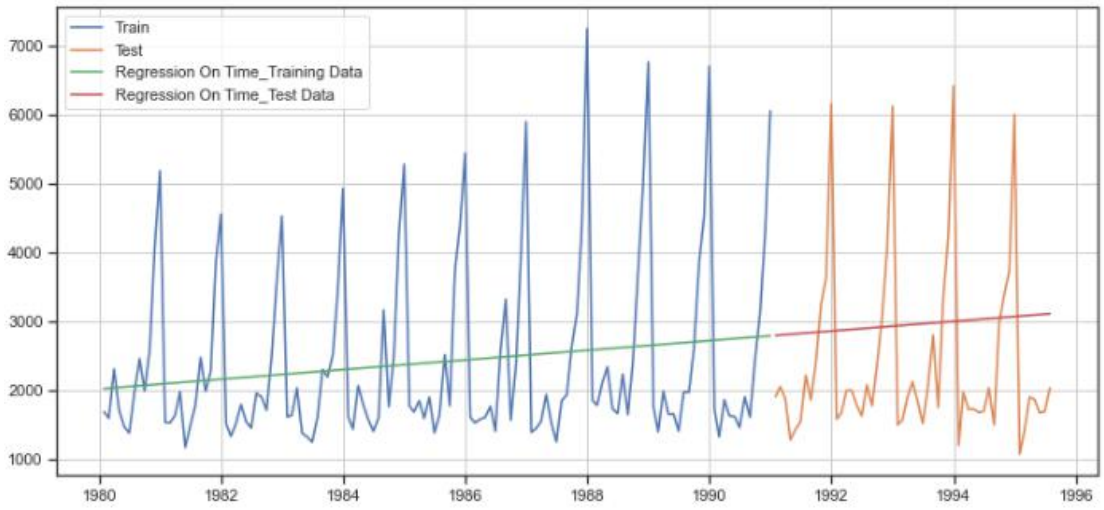
Triple Exponential Smoothing on iteration

	Alpha Values	Beta Values	Gamma Values	Train RMSE	Train MAPE	Test RMSE	Test MAPE
301	0.4	0.1	0.2	389.772245	11.00	336.715250	10.56
211	0.3	0.2	0.2	395.529174	11.09	350.145204	11.08
110	0.2	0.2	0.1	405.333164	11.47	352.571689	11.28
200	0.3	0.1	0.1	394.630053	11.32	352.607849	11.11
20	0.1	0.3	0.1	414.423963	11.48	354.534561	11.77



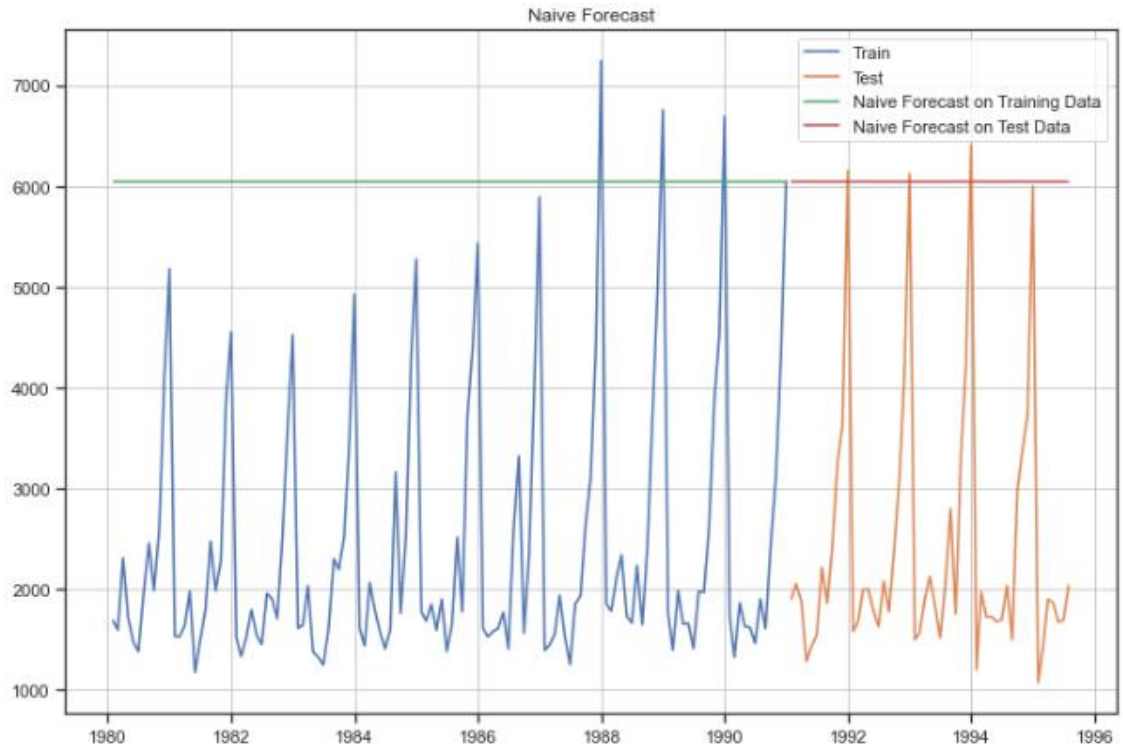
For Alpha=0.4,Beta=0.1,Gamma=0.2 TES Model forecast on the Testing Data, RMSE is 336.71 MAPE is 10.56

Linear Regression



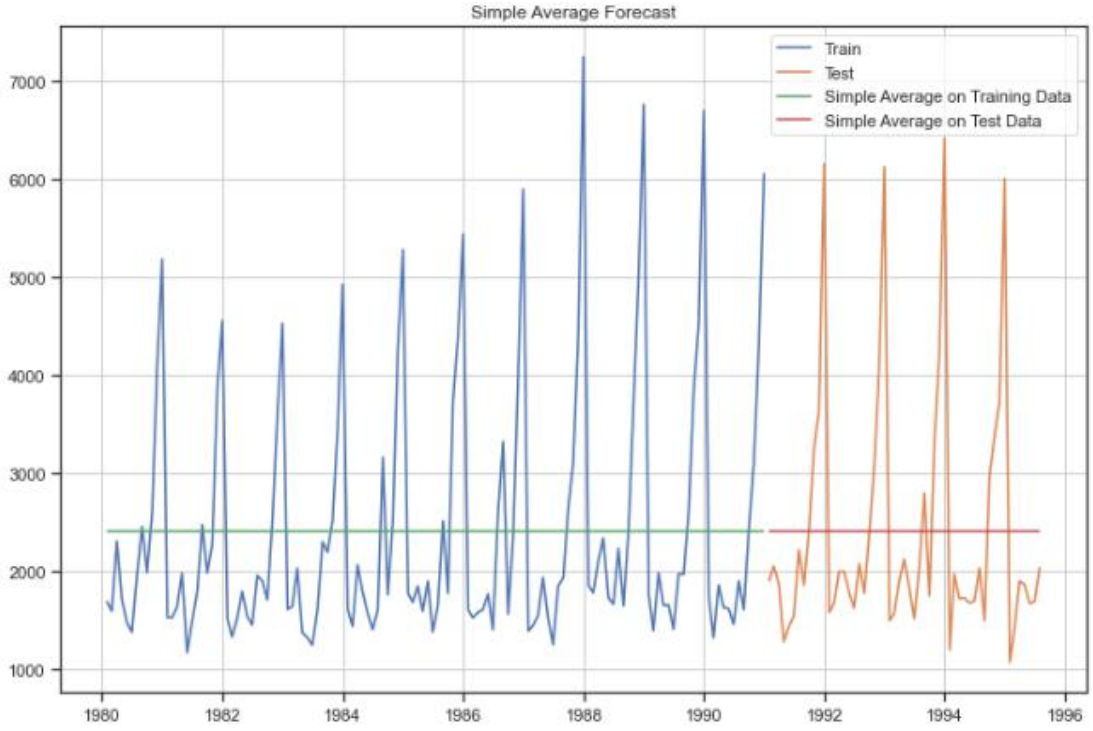
For Regression on Time forecast on the Training Data, RMSE is 1279.322 MAPE is 40.05
For Regression on Time forecast on the Test Data, RMSE is 1389.135 MAPE is 50.15

Naive method



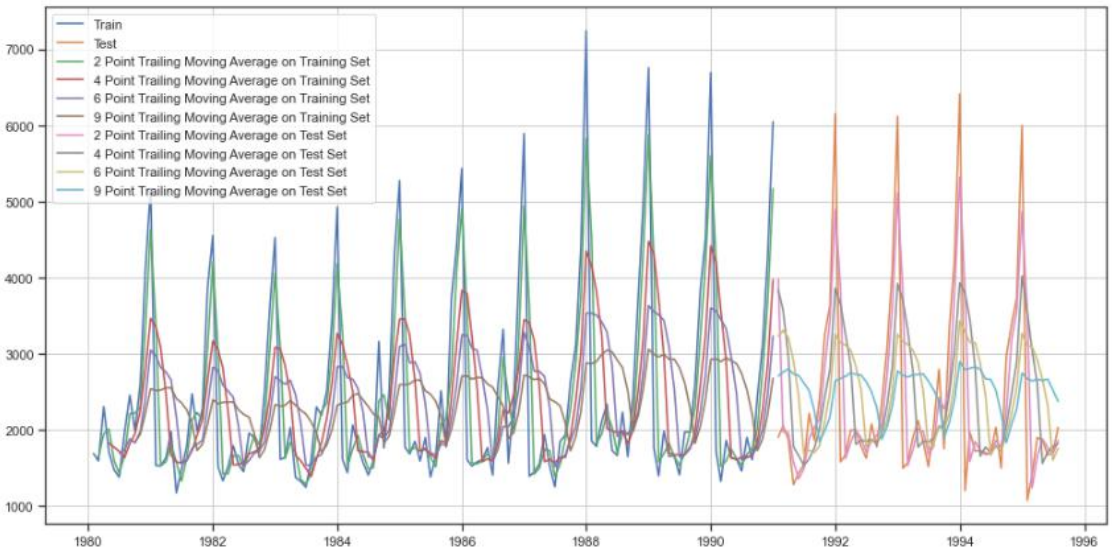
For Naive Model forecast on the Training Data, RMSE is 3867.701 MAPE is 153.17
For Naive Model forecast on the Test Data, RMSE is 3864.279 MAPE is 152.87

Simple Average



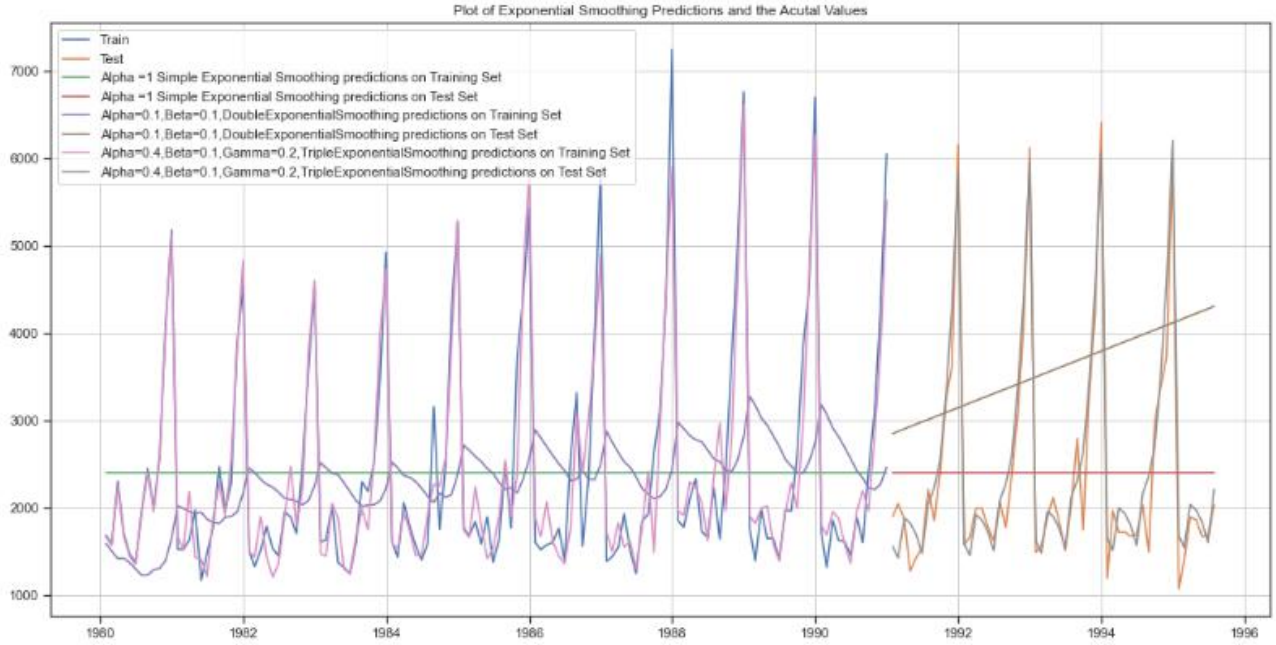
For Simple Average Model forecast on the Training Data, RMSE is 1298.484 MAPE is 40.36
For Simple Average forecast on the Test Data, RMSE is 1275.082 MAPE is 38.90

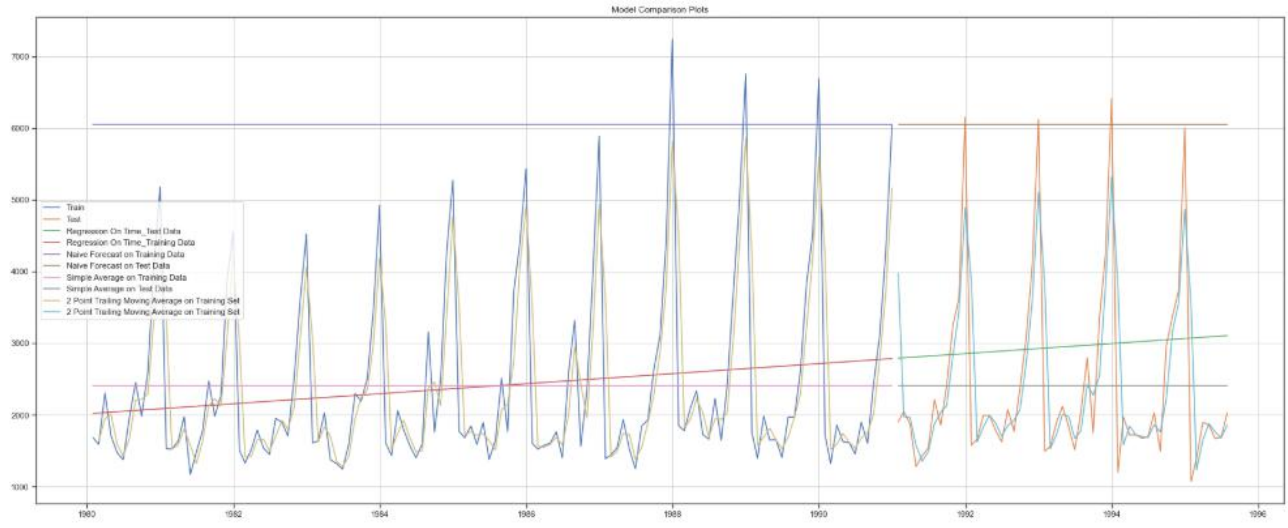
Moving Average



For 2 point Moving Average Model forecast on the Test Data, RMSE is 813.401 MAPE is 19.70
For 4 point Moving Average Model forecast on the Test Data, RMSE is 1156.590 MAPE is 35.96
For 6 point Moving Average Model forecast on the Test Data, RMSE is 1283.927 MAPE is 43.86
For 9 point Moving Average Model forecast on the Test Data, RMSE is 1346.278 MAPE is 46.86

Overall plot on all methods:

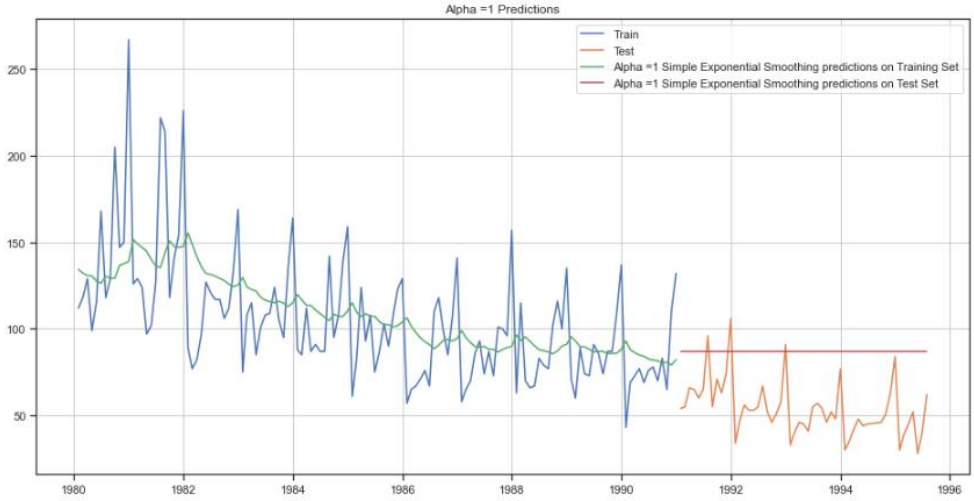




ROSE

Simple Exponential Smoothing on auto parameters

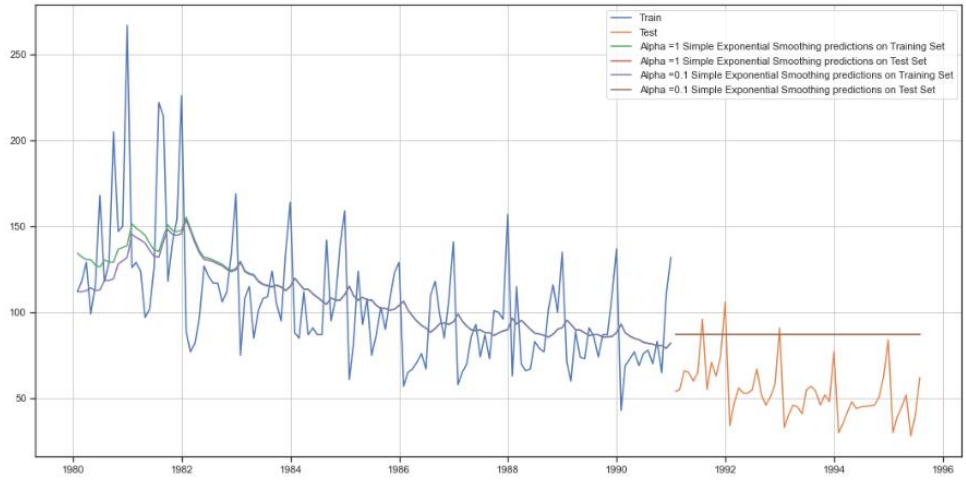
	name	param	optimized
smoothing_level	alpha	0.098750	True
initial_level	l.0	134.386956	True



For Alpha =1 Simple Exponential Smoothing Model forecast on Training Data, RMSE is 31.50 MAPE is 22.73
For Alpha =1 Simple Exponential Smoothing Model forecast on Testing Data, RMSE is 36.79 MAPE is 63.88

Simple Exponential Smoothing on iteration

	Alpha Values	Train RMSE	Train MAPE	Test RMSE	Test MAPE
0	0.1	31.815610	22.47	36.827810	63.94
1	0.2	31.979391	22.20	41.361654	72.21
2	0.3	32.470164	22.31	47.504600	83.71
3	0.4	33.035130	22.50	53.767186	95.50
4	0.5	33.682839	22.68	59.641567	106.81
5	0.6	34.441171	22.88	64.971071	117.04
6	0.7	35.323261	23.08	69.697946	126.07
7	0.8	36.334596	23.41	73.773777	133.82
8	0.9	37.482782	23.93	77.139061	140.21



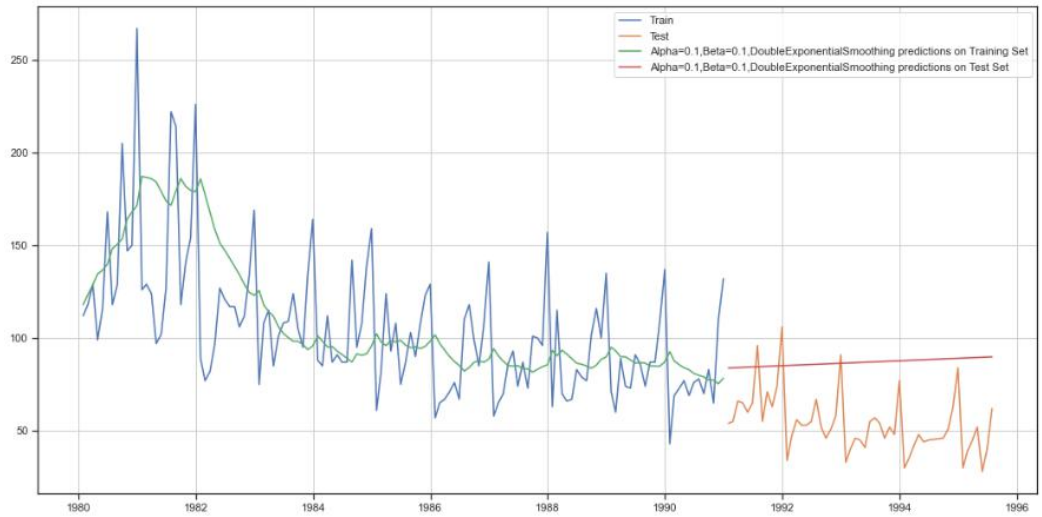
For Alpha =0.1 Simple Exponential Smoothing Model forecast on Testing Data, RMSE is 36.82 MAPE is 63.94

Double Exponential Smoothing on auto parameters

	name	param	optimized
smoothing_level	alpha	0.157895	True
smoothing_slope	beta	0.157895	True
initial_level	l.0	112.000000	True
initial_slope	b.0	6.000000	True

Double Exponential Smoothing on iteration

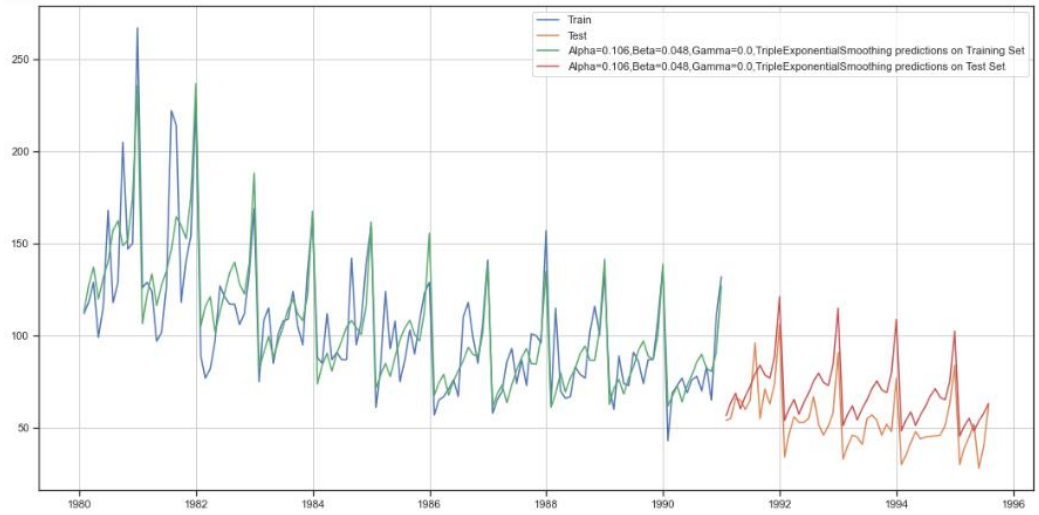
	Alpha Values	Beta Values	Train RMSE	Train MAPE	Test RMSE	Test MAPE
0	0.1	0.1	34.439111	24.83	36.923185	63.78
1	0.1	0.2	33.450729	24.45	48.688399	83.09
10	0.2	0.1	33.097427	23.69	65.731452	113.20
2	0.1	0.3	33.145789	24.46	78.156381	131.24
20	0.3	0.1	33.611269	23.80	98.653063	170.12



For Alpha=0.1, Beta=0.1 DES Model forecast on Testing Data, RMSE is 36.92 MAPE is 63.78

Triple Exponential Smoothing on auto parameters

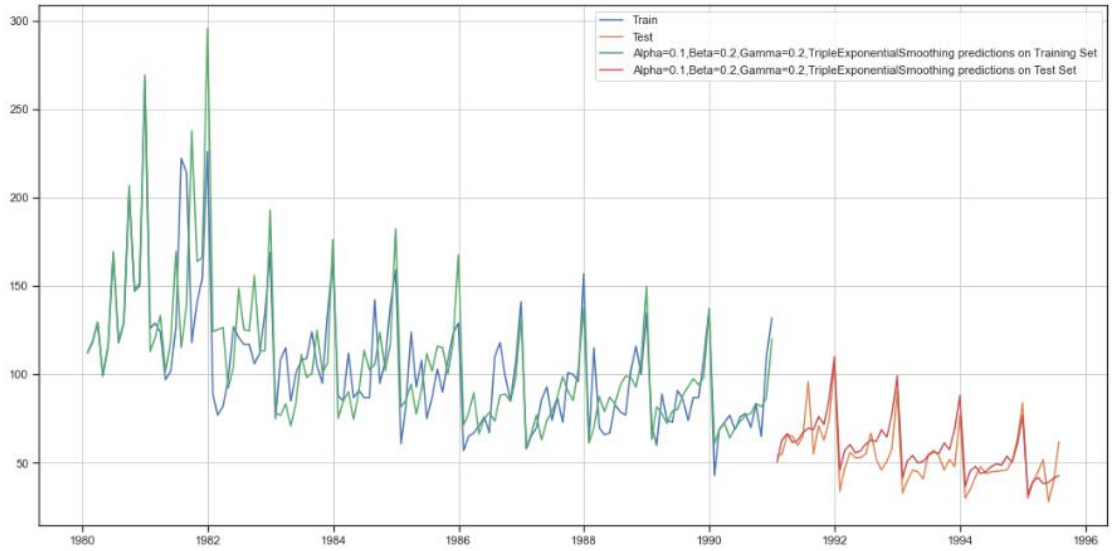
{'smoothing_level': 0.10609628211044227,
'smoothing_slope': 0.04843853483296353,
'smoothing_seasonal': 0.0,
'damping_slope': nan,
'initial_level': 76.65565233828687,
'initial_slope': 0.0,
'initial_seasons': array([1.47550285, 1.65927166, 1.80572675, 1.58888842, 1.77822733,
1.92604398, 2.11649492, 2.25135237, 2.11690636, 2.08112883,
2.40927327, 3.30448211]),
'use_boxcox': False,
'lamda': None,
'remove_bias': False}



For Alpha=0.106,Beta=0.048,Gamma=0.0 TES Model forecast on Training Data, RMSE is 18.57 MAPE is 13.21
For Alpha=0.106,Beta=0.048,Gamma=0.0 TES Model forecast on Testing Data, RMSE is 17.36 MAPE is 28.88

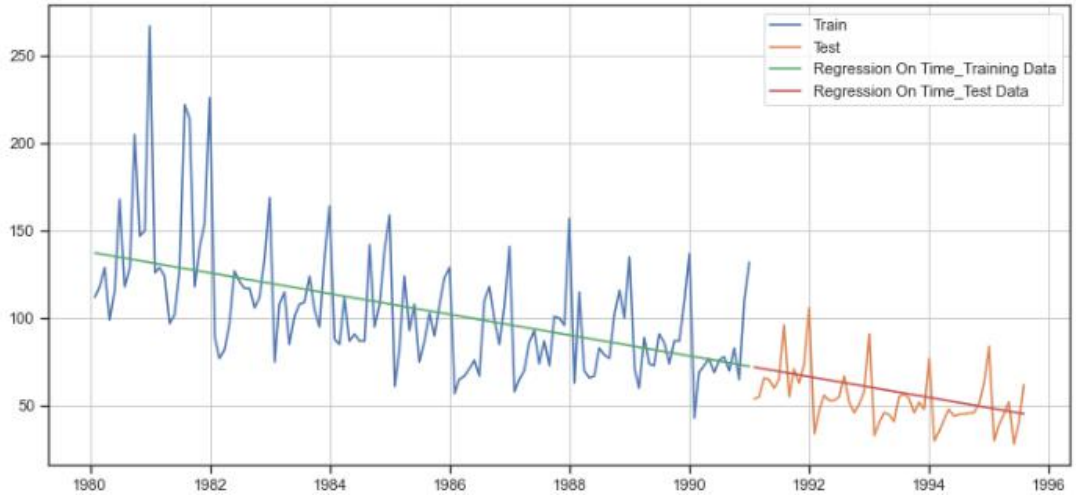
Triple Exponential Smoothing on iteration

	Alpha Values	Beta Values	Gamma Values	Train RMSE	Train MAPE	Test RMSE	Test MAPE
10	0.1	0.2	0.2	24.365597	15.36	9.640616	13.96
11	0.1	0.2	0.3	23.969166	15.13	9.935672	14.21
9	0.1	0.2	0.1	25.529854	16.06	9.943512	14.39
119	0.2	0.5	0.3	27.631767	17.87	10.026322	14.34
127	0.2	0.6	0.2	28.289836	18.09	10.031754	13.62



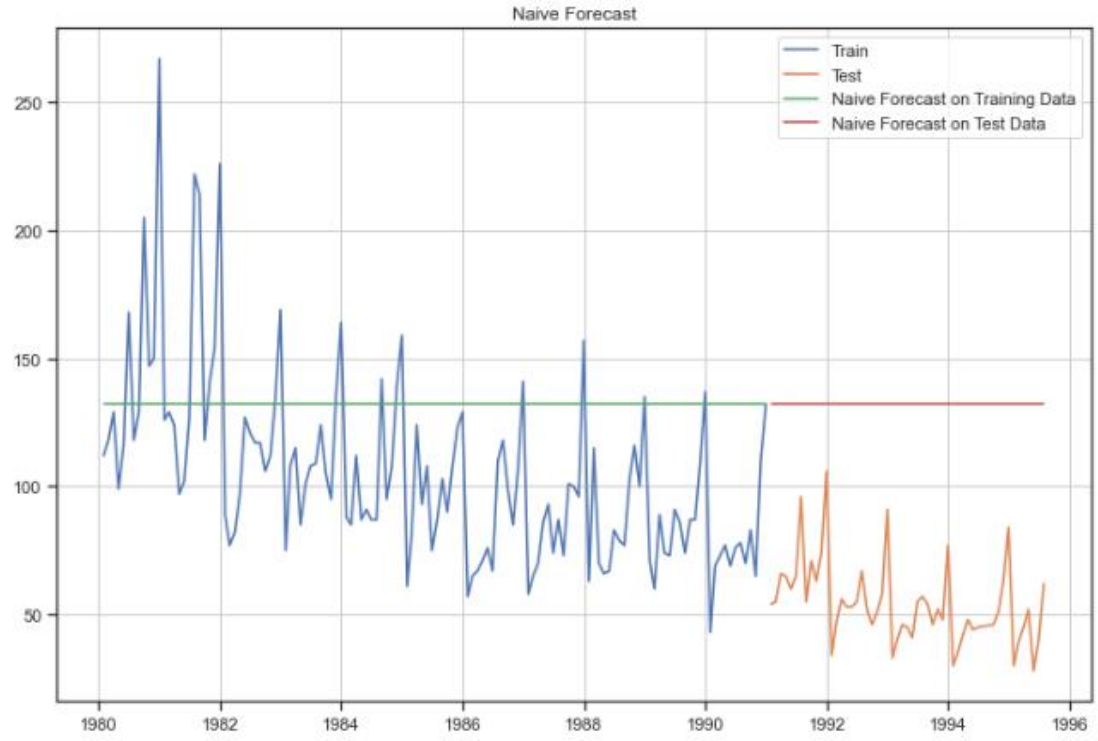
For Alpha=0.1,Beta=0.2,Gamma=0.2 TES Model forecast on the Testing Data, RMSE is 9.64 MAPE is 13.96

Linear Regression



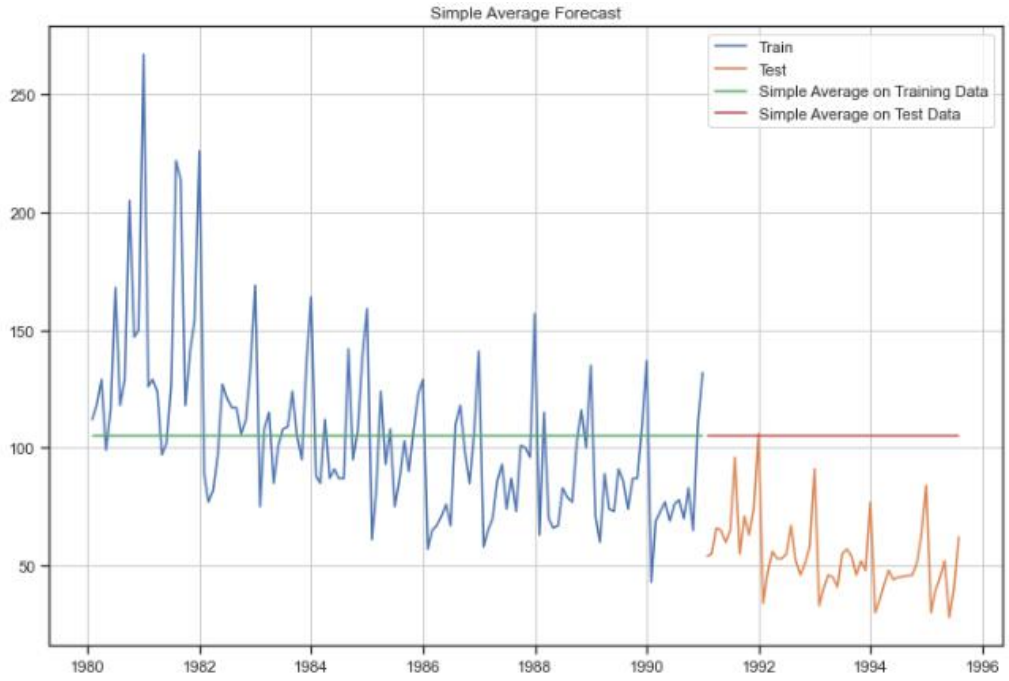
For Regression on Time forecast on the Training Data, RMSE is 30.71 MAPE is 21.22
For Regression on Time forecast on the Test Data, RMSE is 15.26 MAPE is 22.82

Naive method



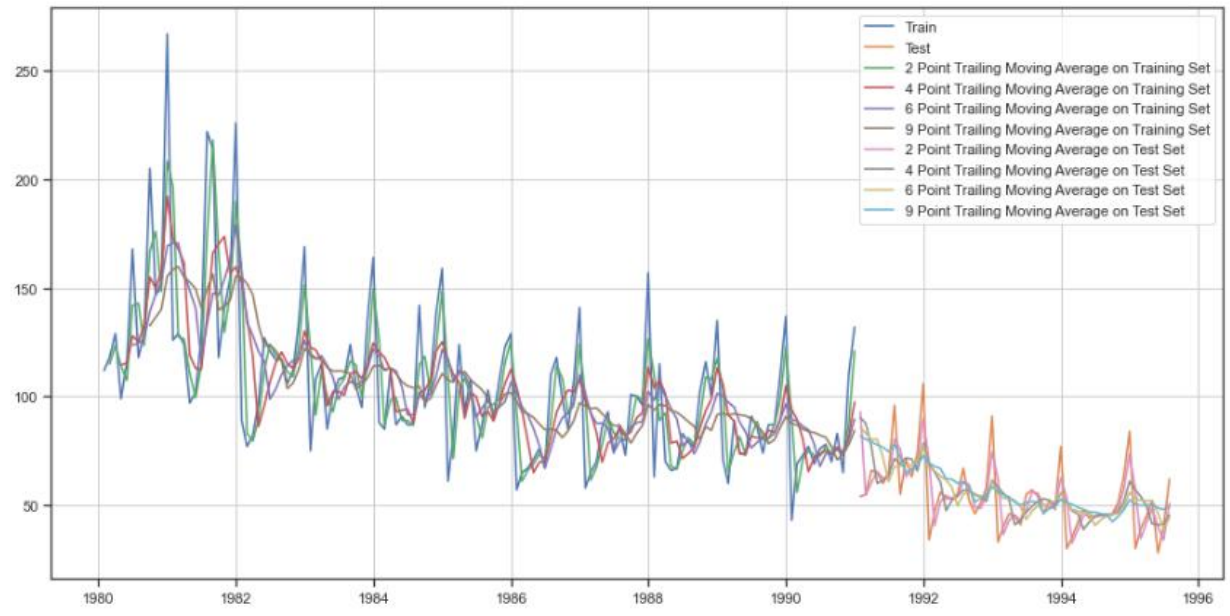
For Naive Model forecast on the Training Data, RMSE is 45.06 MAPE is 36.38
For Naive Model forecast on the Test Data, RMSE is 79.71 MAPE is 145.10

Simple Average



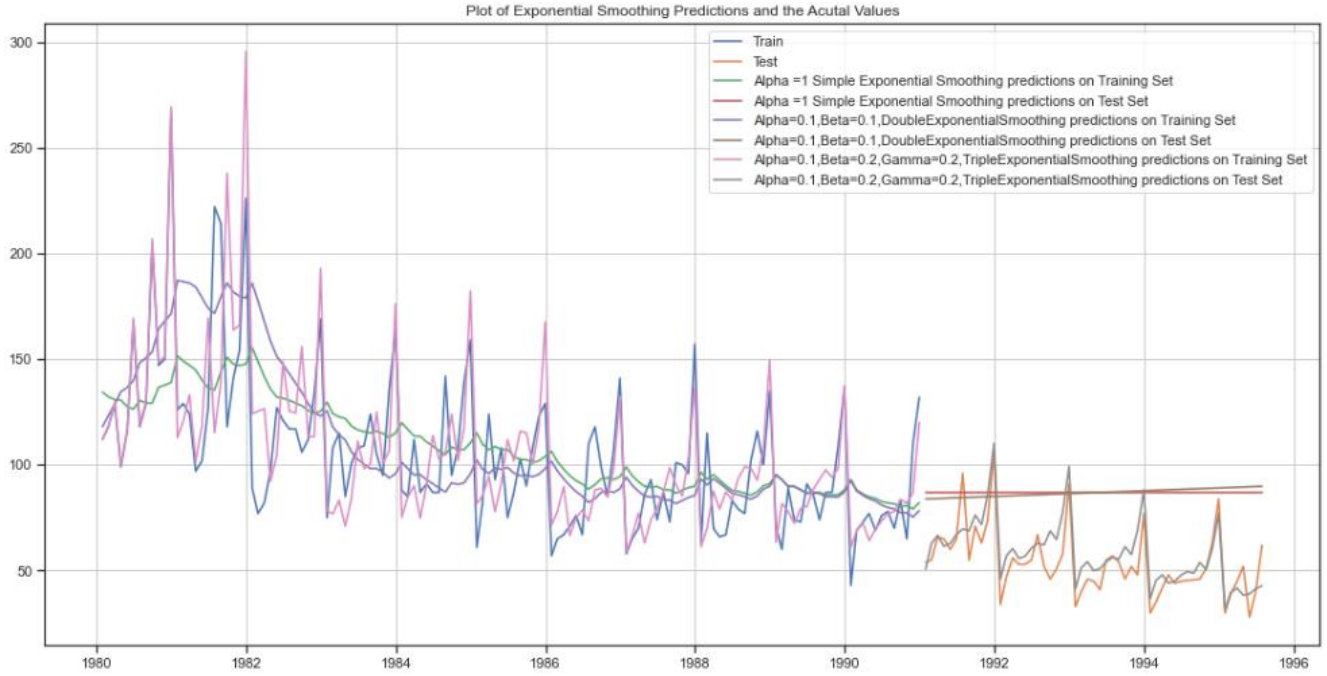
For Simple Average Model forecast on the Training Data, RMSE is 36.03 MAPE is 25.39
For Simple Average forecast on the Test Data, RMSE is 53.46 MAPE is 94.93

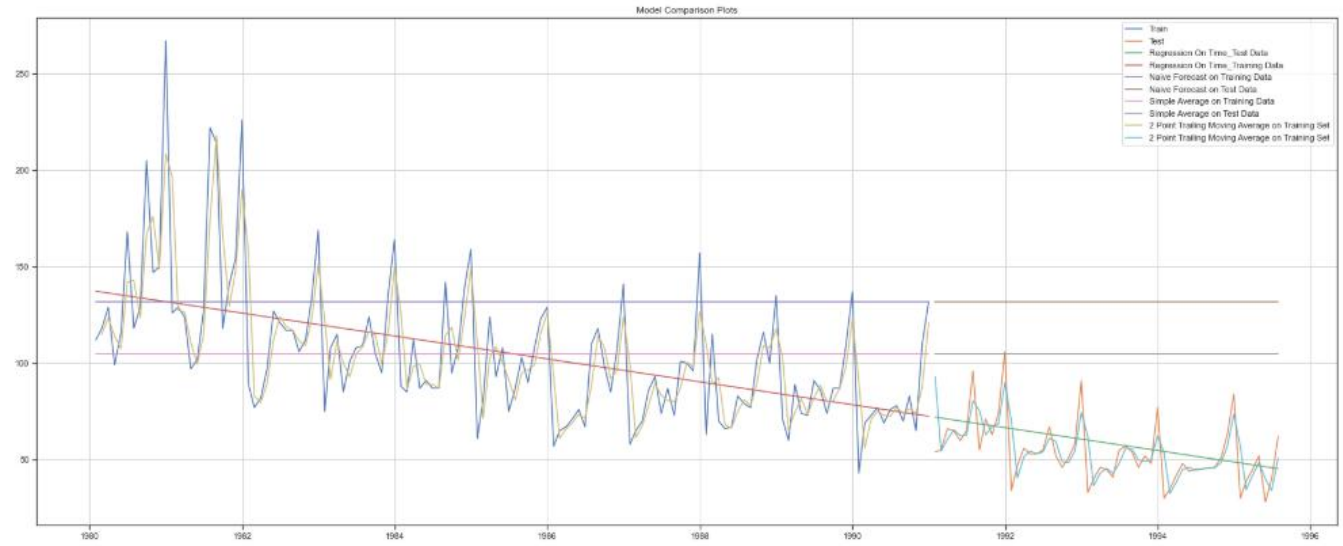
Moving Average



For 2 point Moving Average Model forecast on the Test Data, RMSE is 11.529 MAPE is 13.54
For 4 point Moving Average Model forecast on the Test Data, RMSE is 14.451 MAPE is 19.49
For 6 point Moving Average Model forecast on the Test Data, RMSE is 14.566 MAPE is 20.82
For 9 point Moving Average Model forecast on the Test Data, RMSE is 14.728 MAPE is 21.01

Overall plot on all methods:

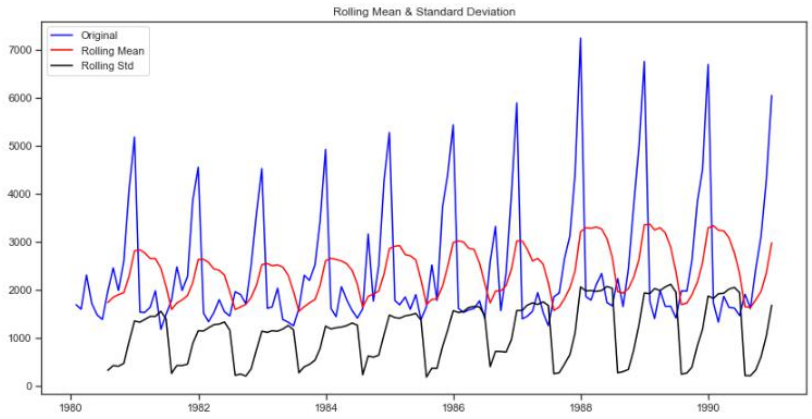




2.5 Checking for the stationarity of data

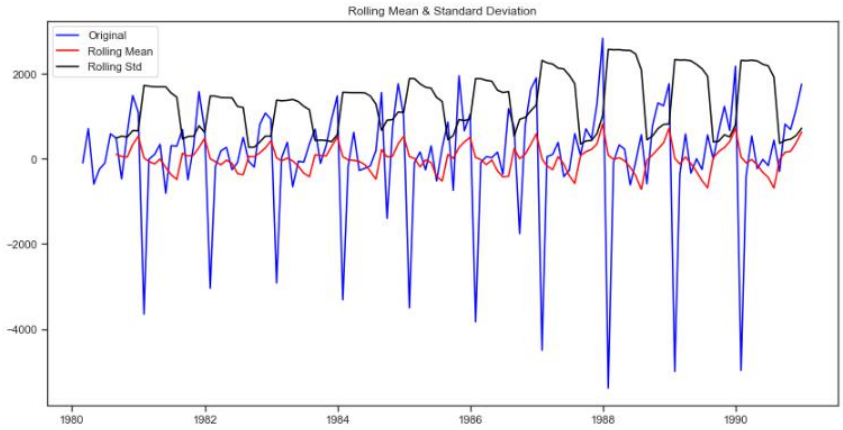
SPARKLING

Performing Dickey-Fuller test



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

The p-value>0.05, so its not stationary and Null hypothesis is accepted. So 1st order difference is done

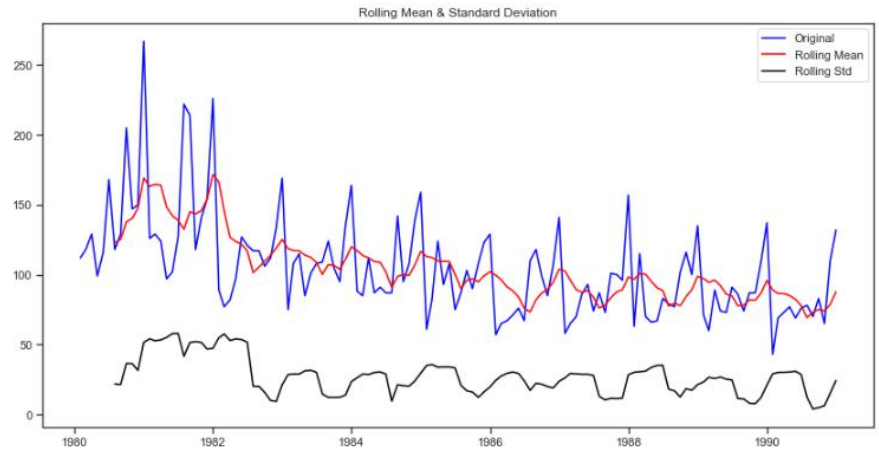


Results of Dickey-Fuller Test:
Test Statistic -8.005007e+00
p-value 2.280104e-12
Lags Used 1.100000e+01
Number of Observations Used 1.190000e+02
Critical Value (1%) -3.486535e+00
Critical Value (5%) -2.886151e+00
Critical Value (10%) -2.579896e+00

Here the p-value<0.05 so it is stationary and Null hypothesis is rejected

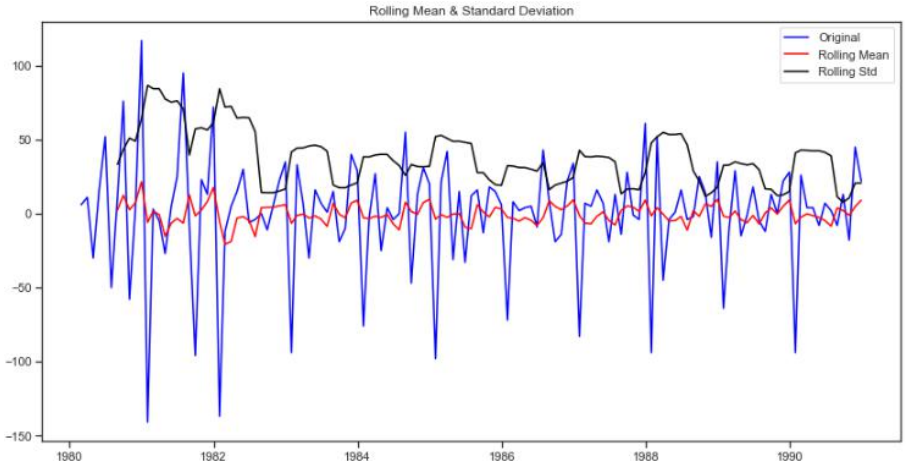
ROSE

Performing Dickey-Fuller test



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

The p-value>0.05, so its not stationary and Null hypothesis is accepted. So 1st order difference is done



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

Here the p-value<0.05 so it is stationary and Null hypothesis is rejected

2.6 Building 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

SPARKLING - ARIMA

	param	AIC
8	(2, 1, 2)	2210.621369
7	(2, 1, 1)	2232.360490
2	(0, 1, 2)	2232.783098
5	(1, 1, 2)	2233.597647
4	(1, 1, 1)	2235.013945
6	(2, 1, 0)	2262.035600
1	(0, 1, 1)	2264.906439
3	(1, 1, 0)	2268.528061
0	(0, 1, 0)	2269.582796

Summary of ARIMA on training set of Sparkling:

ARIMA Model Results

Dep. Variable:	D.Sparkling	No. Observations:	131
Model:	ARIMA(2, 1, 2)	Log Likelihood	-1099.311
Method:	css-mle	S.D. of innovations	1013.283
Date:	Thu, 23 Jul 2020	AIC	2210.621
Time:	20:02:20	BIC	2227.873
Sample:	02-29-1980 - 12-31-1990	HQIC	2217.631

	coef	std err	z	P>|z|	[0.025	0.975]
const	5.5852	0.518	10.789	0.000	4.571	6.600
ar.L1.D.Sparkling	1.2699	0.075	17.043	0.000	1.124	1.416
ar.L2.D.Sparkling	-0.5601	0.074	-7.617	0.000	-0.704	-0.416
ma.L1.D.Sparkling	-1.9966	0.042	-47.026	0.000	-2.080	-1.913
ma.L2.D.Sparkling	0.9966	0.043	23.428	0.000	0.913	1.080

Roots

	Real	Imaginary	Modulus	Frequency
AR.1	1.1335	-0.7074j	1.3361	-0.0888
AR.2	1.1335	+0.7074j	1.3361	0.0888
MA.1	1.0003	+0.0000j	1.0003	0.0000
MA.2	1.0030	+0.0000j	1.0030	0.0000

The RMSE value of ARIMA on training set of Sparkling is 1374.29

SPARKLING - SARIMA

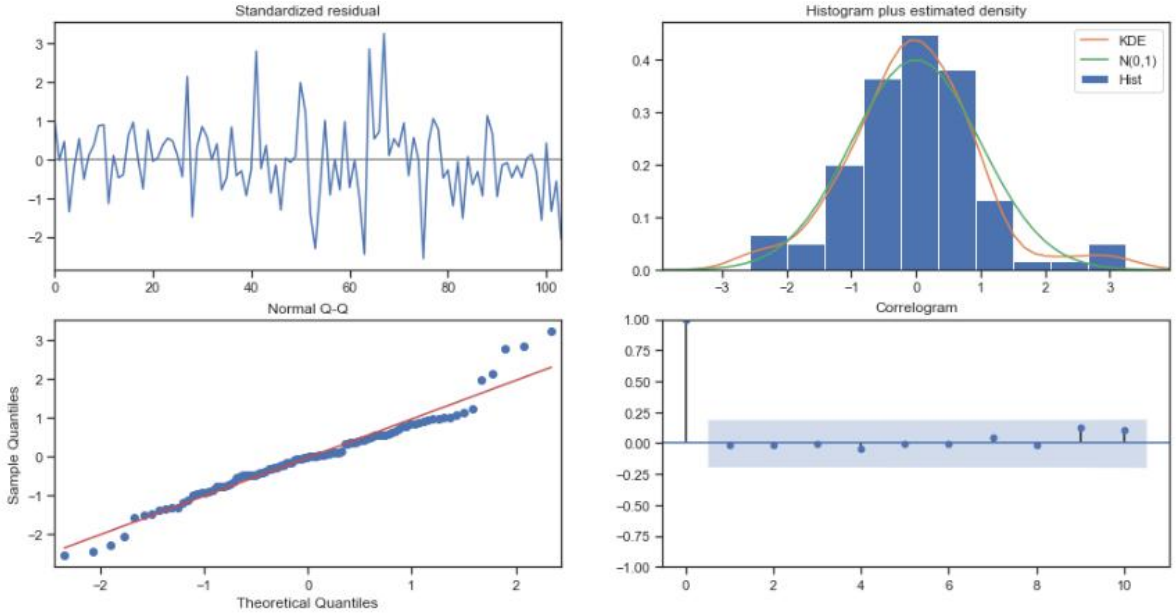
	param	seasonal	AIC
50	(1, 1, 2)	(1, 0, 2, 12)	1555.584247
53	(1, 1, 2)	(2, 0, 2, 12)	1555.934564
26	(0, 1, 2)	(2, 0, 2, 12)	1557.121563
23	(0, 1, 2)	(1, 0, 2, 12)	1557.160507
77	(2, 1, 2)	(1, 0, 2, 12)	1557.340405

Summary of SARIMA on training set of Sparkling:

```
                                 SARIMAX Results
==========================================================================================
Dep. Variable:                                 y   No. Observations:                  132
Model:             SARIMAX(1, 1, 2)x(1, 0, 2, 12)   Log Likelihood                -770.792
Date:                            Thu, 23 Jul 2020   AIC                           1555.584
Time:                                    20:08:03   BIC                           1574.095
Sample:                                         0   HQIC                          1563.083
                                              - 132
Covariance Type:                              opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.6282      0.255     -2.464      0.014      -1.128      -0.128
ma.L1         -0.1040      0.225     -0.463      0.644      -0.545       0.337
ma.L2         -0.7276      0.154     -4.736      0.000      -1.029      -0.426
ar.S.L12       1.0439      0.014     72.838      0.000       1.016       1.072
ma.S.L12      -0.5550      0.098     -5.663      0.000      -0.747      -0.363
ma.S.L24      -0.1354      0.120     -1.133      0.257      -0.370       0.099
sigma2     1.506e+05   2.03e+04      7.401      0.000    1.11e+05    1.9e+05
===================================================================================
Ljung-Box (Q):                       23.02   Jarque-Bera (JB):                11.72
Prob(Q):                              0.99   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).
```

The RMSE value of SARIMA on training set of Sparkling is 528.65



ROSE - ARIMA

	param	AIC
2	(0, 1, 2)	1276.835382
5	(1, 1, 2)	1277.359225
4	(1, 1, 1)	1277.775750
7	(2, 1, 1)	1279.045689
8	(2, 1, 2)	1279.298694
1	(0, 1, 1)	1280.726183
6	(2, 1, 0)	1300.609261
3	(1, 1, 0)	1319.348311
0	(0, 1, 0)	1335.152658

Summary of ARIMA on training set of Rose:

ARIMA Model Results

Dep. Variable:	D.Rose	No. Observations:	131
Model:	ARIMA(0, 1, 2)	Log Likelihood	-634.418
Method:	css-mle	S.D. of innovations	30.168
Date:	Thu, 23 Jul 2020	AIC	1276.835
Time:	14:41:22	BIC	1288.336
Sample:	02-29-1980 - 12-31-1990	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.D.Rose	-0.7600	0.101	-7.500	0.000	-0.959	-0.561
ma.L2.D.Rose	-0.2398	0.095	-2.518	0.012	-0.427	-0.053

Roots

	Real	Imaginary	Modulus	Frequency
MA.1	1.0001	+0.0000j	1.0001	0.0000
MA.2	-4.1694	+0.0000j	4.1694	0.5000

The RMSE value of ARIMA on training set of Rose is 15.61

ROSE - SARIMA

	param	seasonal	AIC
26	(0, 1, 2)	(2, 0, 2, 12)	887.937509
53	(1, 1, 2)	(2, 0, 2, 12)	889.871767
80	(2, 1, 2)	(2, 0, 2, 12)	890.668798
69	(2, 1, 1)	(2, 0, 0, 12)	896.518161
78	(2, 1, 2)	(2, 0, 0, 12)	897.346444

Summary of SARIMA on training set of Rose:

SARIMAX Results

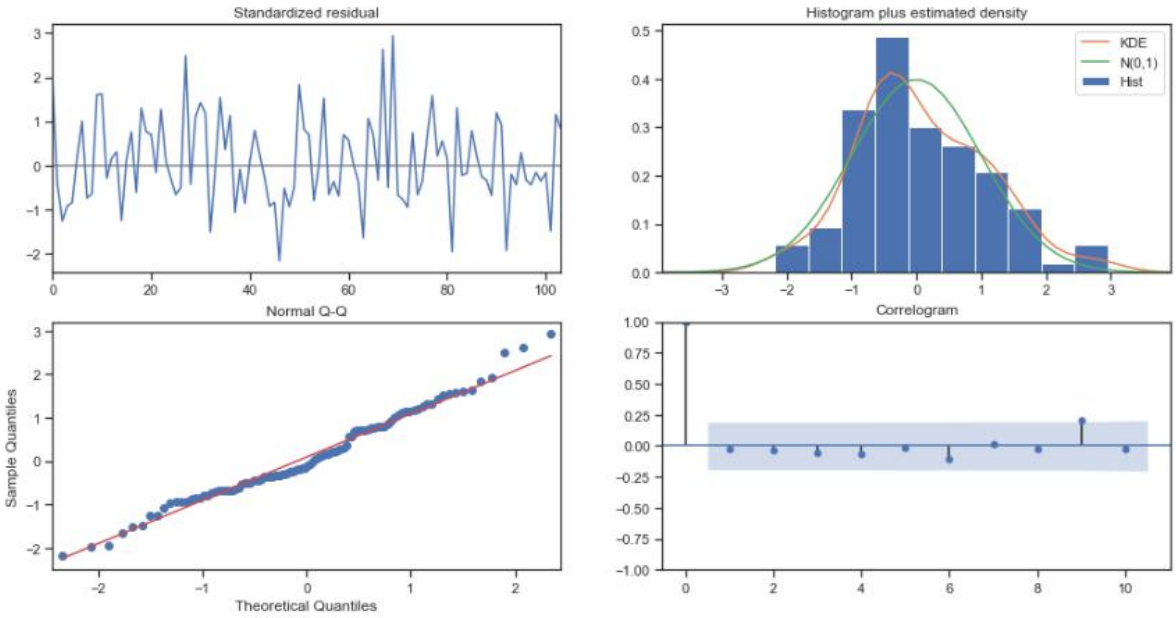
Dep. Variable:	y	No. Observations:	132
Model:	SARIMAX(0, 1, 2)x(2, 0, 2, 12)	Log Likelihood	-436.969
Date:	Thu, 23 Jul 2020	AIC	887.938
Time:	15:27:59	BIC	906.448
Sample:	0 - 132	HQIC	895.437
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ma.L1	-0.8427	190.102	-0.004	0.996	-373.435	371.750
ma.L2	-0.1573	29.866	-0.005	0.996	-58.694	58.379
ar.S.L12	0.3467	0.079	4.375	0.000	0.191	0.502
ar.S.L24	0.3023	0.076	3.996	0.000	0.154	0.451
ma.S.L12	0.0767	0.133	0.577	0.564	-0.184	0.337
ma.S.L24	-0.0726	0.146	-0.498	0.618	-0.358	0.213
sigma2	251.3137	4.78e+04	0.005	0.996	-9.34e+04	9.39e+04

Ljung-Box (Q):	24.56	Jarque-Bera (JB):	2.33
Prob(Q):	0.97	Prob(JB):	0.31
Heteroskedasticity (H):	0.88	Skew:	0.37
Prob(H) (two-sided):	0.70	Kurtosis:	3.03

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

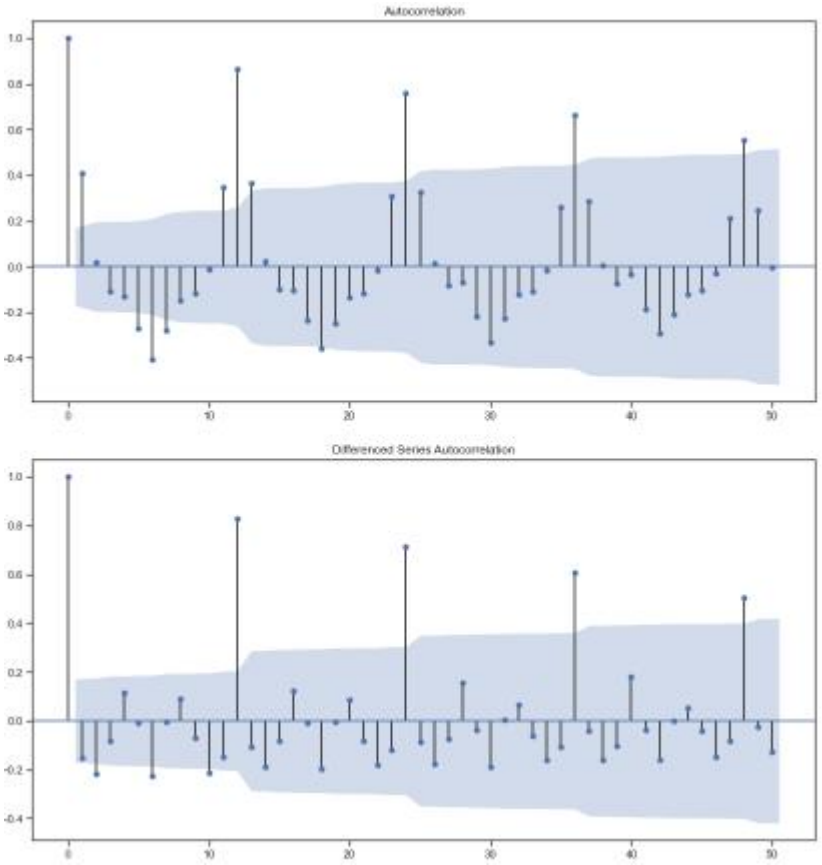
The RMSE value of SARIMA on training set of Rose is 26.92



2.7 Building ARIMA/SARIMA models based on the cut-off points of ACF and PACF on the training data and evaluating this model on the test data using RMSE

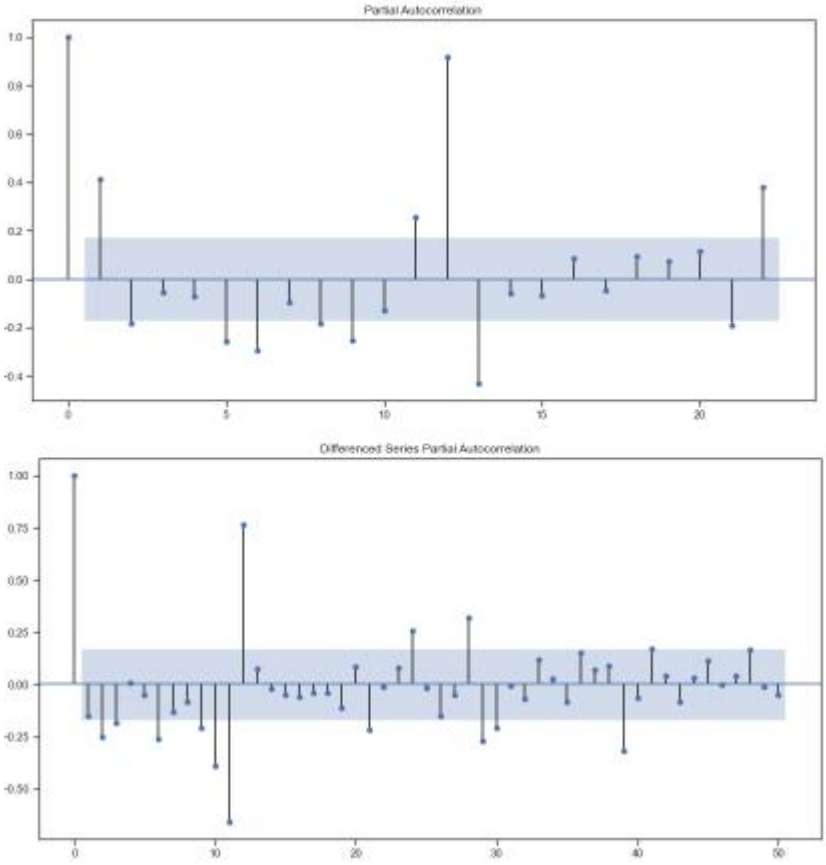
SPARKLING - ACF

ACF plot with original data set and differenced data set:



SPARKLING - PACF

PACF plot with original data set and differenced data set:



Summary of ARIMA of ACF/PACF on training set of Sparkling:

ARIMA Model Results

Dep. Variable:	D.y	No. Observations:	131
Model:	ARIMA(1, 1, 1)	Log Likelihood	-1113.507
Method:	css-mle	S.D. of innovations	1171.378
Date:	Thu, 23 Jul 2020	AIC	2235.014
Time:	20:08:24	BIC	2246.515
Sample:	1	HQIC	2239.687

	coef	std err	z	P>|z|	[0.025	0.975]
const	6.7489	4.616	1.462	0.144	-2.299	15.797
ar.L1.D.y	0.4289	0.082	5.221	0.000	0.268	0.590
ma.L1.D.y	-1.0000	0.019	-51.962	0.000	-1.038	-0.962

Roots

	Real	Imaginary	Modulus	Frequency
AR.1	2.3313	+0.0000j	2.3313	0.0000
MA.1	1.0000	+0.0000j	1.0000	0.0000

The RMSE value of ARIMA of ACF/PACF on training set of Sparkling is 1461.66

Summary of SARIMA of ACF/PACF on training set of Sparkling:

SARIMAX Results

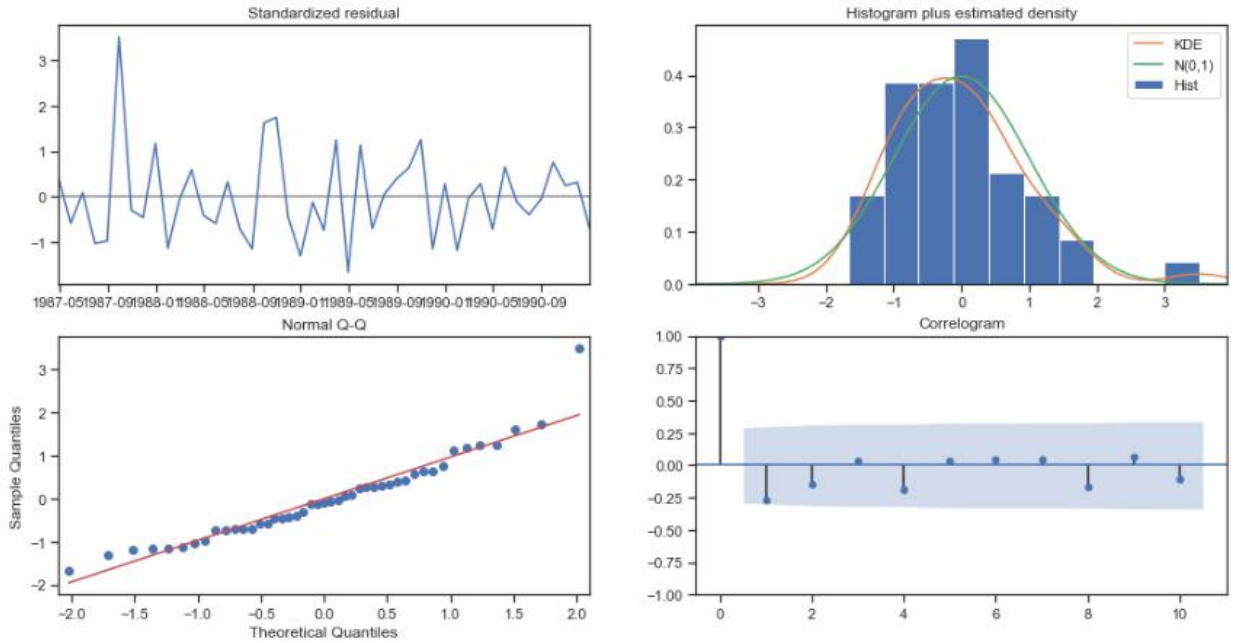
Dep. Variable:	Sparkling	No. Observations:	132
Model:	SARIMAX(1, 1, 1)x(2, 0, [1, 2, 3, 4, 5, 6, 7], 12)	Log Likelihood	-342.505
Date:	Thu, 23 Jul 2020	AIC	709.010
Time:	20:10:55	BIC	730.690
Sample:	01-31-1980 - 12-31-1990	HQIC	717.092
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	-0.7670	0.150	-5.130	0.000	-1.060	-0.474
ma.L1	0.9996	17.586	0.057	0.955	-33.469	35.468
ar.S.L12	1.4380	0.155	9.304	0.000	1.135	1.741
ar.S.L24	-0.4954	0.158	-3.135	0.002	-0.805	-0.186
ma.S.L12	-1.5552	18.536	-0.084	0.933	-37.885	34.775
ma.S.L24	1.1753	14.323	0.082	0.935	-26.897	29.248
ma.S.L36	0.0172	3.750	0.005	0.996	-7.333	7.368
ma.S.L48	-0.1688	6.593	-0.026	0.980	-13.090	12.753
ma.S.L60	-0.2969	5.010	-0.059	0.953	-10.115	9.522
ma.S.L72	0.5903	9.357	0.063	0.950	-17.749	18.930
ma.S.L84	-0.0059	1.419	-0.004	0.997	-2.787	2.775
sigma2	9.251e+04	0.000	4.49e+08	0.000	9.25e+04	9.25e+04

Ljung-Box (Q):	40.77	Jarque-Bera (JB):	16.30
Prob(Q):	0.44	Prob(JB):	0.00
Heteroskedasticity (H):	0.36	Skew:	1.07
Prob(H) (two-sided):	0.06	Kurtosis:	5.03

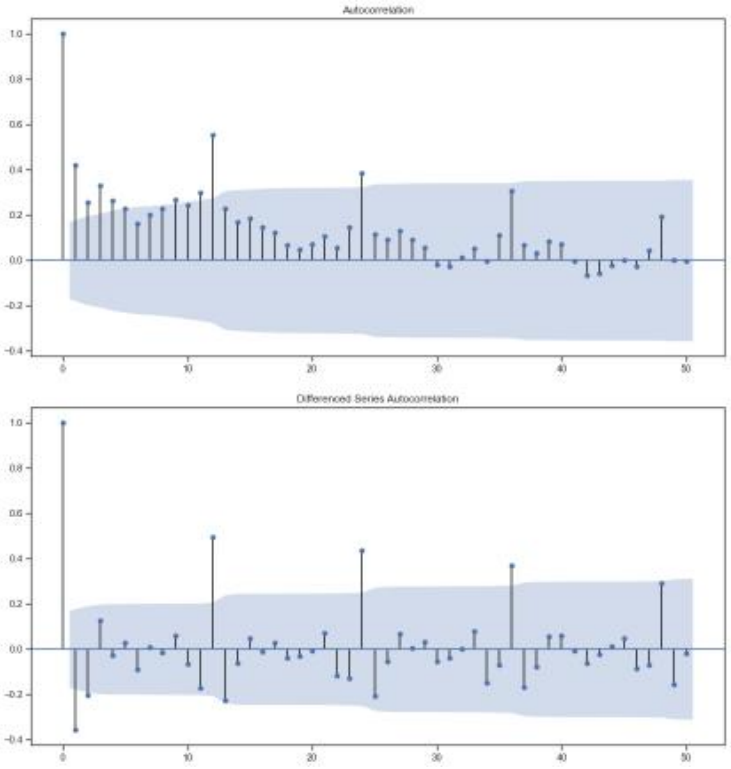
Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
[2] Covariance matrix is singular or near-singular, with condition number 3.25e+26. Standard errors may be unstable.

The RMSE value of SARIMA of ACF/PACF on training set of Sparkling is 455.49



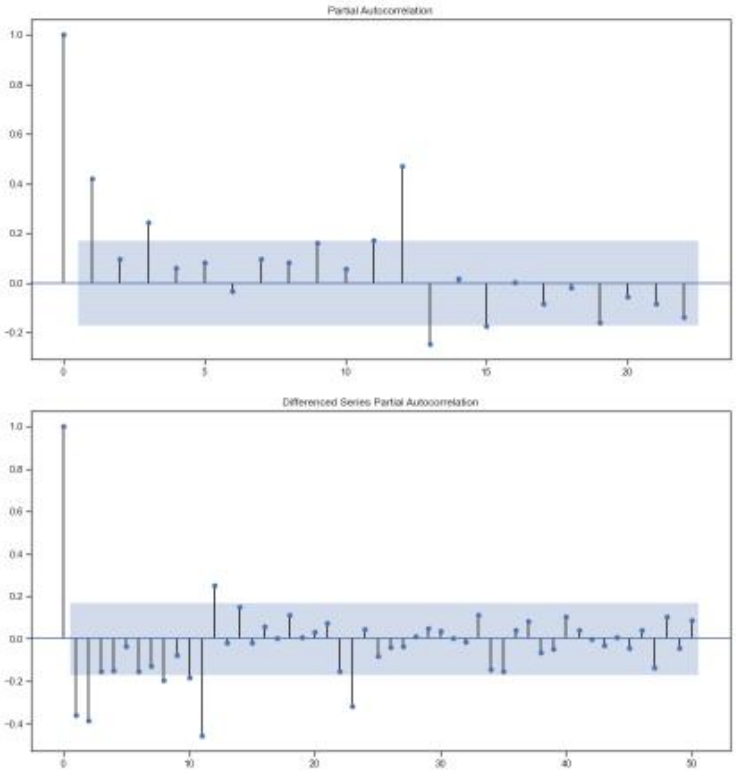
ROSE - ACF

ACF plot with original data set and differenced data set:



ROSE - PACF

PACF plot with original data set and differenced data set:



Summary of ARIMA of ACF/PACF on training set of Rose:

ARIMA Model Results

Dep. Variable:	D.y	No. Observations:	131
Model:	ARIMA(1, 1, 1)	Log Likelihood	-634.888
Method:	css-mle	S.D. of innovations	30.279
Date:	Thu, 23 Jul 2020	AIC	1277.776
Time:	15:31:57	BIC	1289.277
Sample:	1	HQIC	1282.449

	coef	std err	z	P>|z|	[0.025	0.975]
const	-0.4871	0.086	-5.656	0.000	-0.656	-0.318
ar.L1.D.y	0.2006	0.087	2.293	0.022	0.029	0.372
ma.L1.D.y	-0.9999	0.035	-28.646	0.000	-1.068	-0.932

Roots

	Real	Imaginary	Modulus	Frequency
AR.1	4.9856	+0.0000j	4.9856	0.0000
MA.1	1.0001	+0.0000j	1.0001	0.0000

The RMSE value of ARIMA of ACF/PACF on training set of Rose is 15.73

Summary of SARIMA of ACF/PACF on training set of Rose:

SARIMAX Results

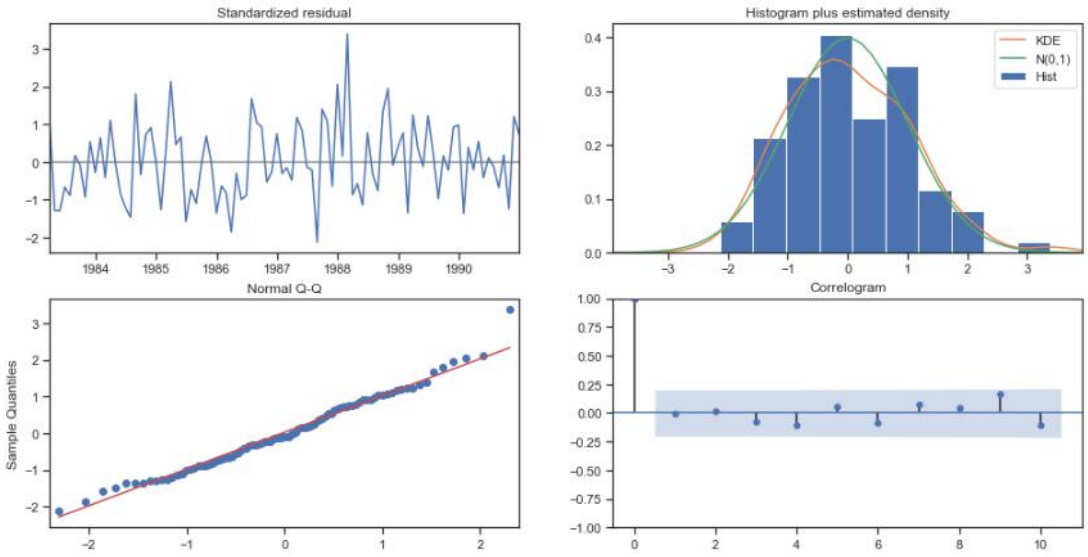
Dep. Variable:	Rose	No. Observations:	132
Model:	SARIMAX(1, 1, 1)x(3, 0, [1, 2], 12)	Log Likelihood	-387.406
Date:	Thu, 23 Jul 2020	AIC	790.811
Time:	15:32:27	BIC	811.158
Sample:	01-31-1980	HQIC	799.030
	- 12-31-1990		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	0.0716	0.116	0.617	0.537	-0.156	0.299
ma.L1	-0.9802	0.104	-9.382	0.000	-1.185	-0.775
ar.S.L12	0.7482	0.139	5.385	0.000	0.476	1.021
ar.S.L24	0.1199	0.126	0.953	0.340	-0.127	0.366
ar.S.L36	0.0457	0.082	0.560	0.575	-0.114	0.206
ma.S.L12	-0.4830	0.219	-2.201	0.028	-0.913	-0.053
ma.S.L24	-0.1920	0.188	-1.020	0.308	-0.561	0.177
sigma2	200.7421	36.128	5.556	0.000	129.932	271.552

Ljung-Box (Q):	33.09	Jarque-Bera (JB):	3.05
Prob(Q):	0.77	Prob(JB):	0.22
Heteroskedasticity (H):	0.77	Skew:	0.43
Prob(H) (two-sided):	0.46	Kurtosis:	3.21

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

The RMSE value of SARIMA of ACF/PACF on training set of Rose is 18.95



2.8 Building a table (data frame) with all the models built along with their corresponding parameters and the respective RMSE values on the test data

SPARKLING

	Test RMSE	Test MAPE
Alpha=1,SimpleExponentialSmoothing	1275.081823	38.90
Alpha=0.1,SimpleExponentialSmoothing	1375.393398	49.53
Alpha=0.1,Beta=0.1,DoubleExponentialSmoothing	1778.564670	67.20
Alpha=0.154,Beta=5.310e-21,Gamma=0.371,TripleExponentialSmoothing	383.157627	11.91
Alpha=0.4,Beta=0.1,Gamma=0.2,TripleExponentialSmoothing	336.715250	10.56
RegressionOnTime	1389.135175	50.15
NaiveModel	3864.279352	152.87
SimpleAverageModel	1275.081804	38.90
2pointTrailingMovingAverage	813.400684	19.70
4pointTrailingMovingAverage	1156.589694	35.96
6pointTrailingMovingAverage	1283.927428	43.86
9pointTrailingMovingAverage	1346.278315	46.86
ARIMA(2,1,2)	1374.297411	NaN
SARIMA(1, 1, 2)(1, 0, 2, 12)	528.655337	NaN
ARIMA_CF(1, 1, 1)	1461.662603	NaN
SARIMA_CF(1, 1, 1)(2, 0, 7, 12)	455.498645	NaN

ROSE

	Test RMSE	Test MAPE
Alpha=1,SimpleExponentialSmoothing	36.796019	63.88
Alpha=0.1,SimpleExponentialSmoothing	36.827810	63.94
Alpha=0.1,Beta=0.1,DoubleExponentialSmoothing	36.923185	63.78
Alpha=0.106,Beta=0.048,Gamma=0.0,TripleExponentialSmoothing	17.369210	28.88
Alpha=0.1,Beta=0.2,Gamma=0.2,TripleExponentialSmoothing	9.640616	13.96
RegressionOnTime	15.268885	22.82
NaiveModel	79.718559	145.10
SimpleAverageModel	53.460350	94.93
2pointTrailingMovingAverage	11.529278	13.54
4pointTrailingMovingAverage	14.451364	19.49
6pointTrailingMovingAverage	14.566269	20.82
9pointTrailingMovingAverage	14.727594	21.01
ARIMA(2,1,2)	15.619588	NaN
SARIMA(0, 1, 2)(2, 0, 2, 12)	26.928140	NaN
ARIMA_CF(1, 1, 1)	15.733476	NaN
SARIMA_CF(1, 1, 1)(3, 0, 2, 12)	18.959324	NaN

2.9 Based on the model-building exercise, building the most optimum model(s) on the complete data and predicting 12 months into the future with appropriate confidence intervals/bands

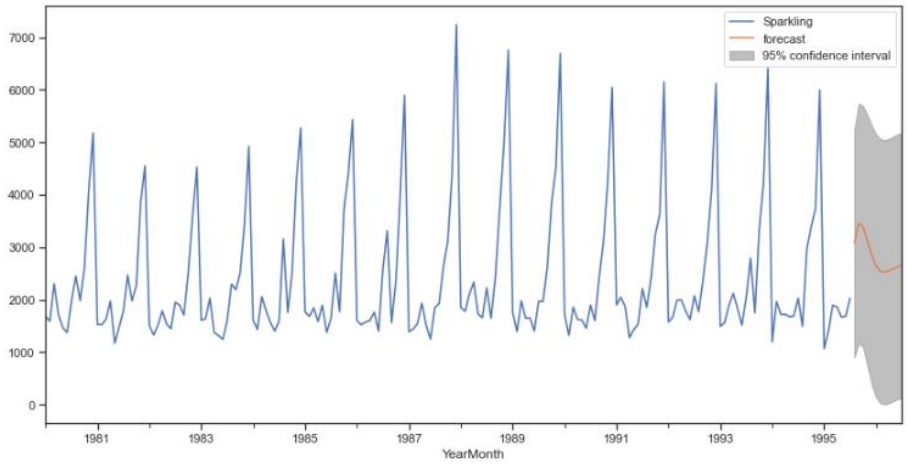
SPARKLING

ARIMA forecast on auto parameters

Summary:

```
                             ARIMA Model Results
==============================================================================
Dep. Variable:            D.Sparkling   No. Observations:                  186
Model:                 ARIMA(2, 1, 2)   Log Likelihood               -1572.156
Method:                       css-mle   S.D. of innovations           1110.608
Date:                Thu, 23 Jul 2020   AIC                           3156.313
Time:                        20:02:22   BIC                           3175.667
Sample:                  02-29-1980   HQIC                          3164.156
                       - 07-31-1995
==========================================================================================
                             coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------------------
const                     2.2957      0.694      3.310      0.001       0.936       3.655
ar.L1.D.Sparkling         1.1903      0.068     17.545      0.000       1.057       1.323
ar.L2.D.Sparkling        -0.4928      0.064     -7.724      0.000      -0.618      -0.368
ma.L1.D.Sparkling        -1.8747        nan        nan        nan         nan         nan
ma.L2.D.Sparkling         0.8747      0.015     58.268      0.000       0.845       0.904
                                    Roots
=============================================================================
                  Real          Imaginary           Modulus         Frequency
-----------------------------------------------------------------------------
AR.1            1.2078           -0.7554j            1.4245           -0.0890
AR.2            1.2078           +0.7554j            1.4245            0.0890
MA.1            1.0000           +0.0000j            1.0000            0.0000
MA.2            1.1432           +0.0000j            1.1432            0.0000
-----------------------------------------------------------------------------
```

Forecast:



SARIMA forecast on auto parameters

Summary:

SARIMAX Results

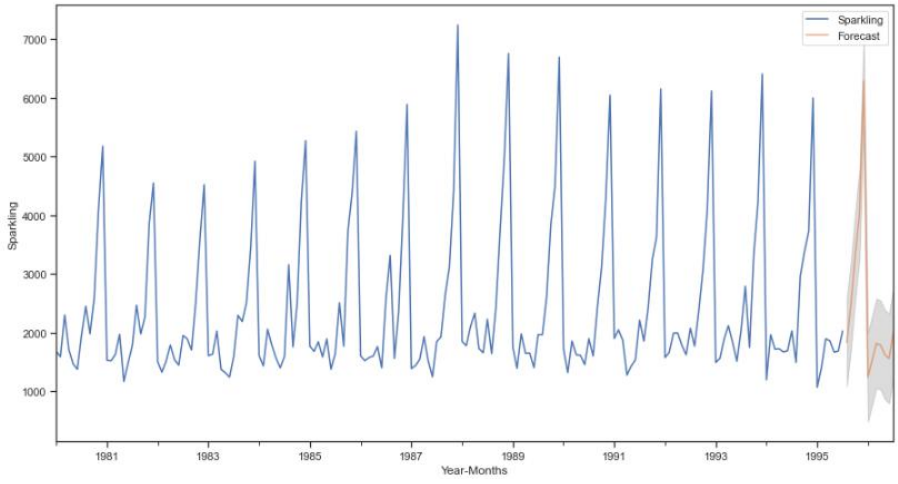
Dep. Variable:	Sparkling	No. Observations:	187
Model:	SARIMAX(1, 1, 2)x(1, 0, 2, 12)	Log Likelihood	-1173.413
Date:	Thu, 23 Jul 2020	AIC	2360.826
Time:	20:08:17	BIC	2382.308
Sample:	01-31-1980 - 07-31-1995	HQIC	2369.549
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	-0.6610	0.242	-2.733	0.006	-1.135	-0.187
ma.L1	-0.1801	0.219	-0.822	0.411	-0.609	0.249
ma.L2	-0.7387	0.192	-3.846	0.000	-1.115	-0.362
ar.S.L12	1.0157	0.012	84.460	0.000	0.992	1.039
ma.S.L12	-1.3874	0.338	-4.101	0.000	-2.051	-0.724
ma.S.L24	-0.1460	0.146	-1.000	0.317	-0.432	0.140
sigma2	6.532e+04	2.08e+04	3.136	0.002	2.45e+04	1.06e+05

Ljung-Box (Q):	18.00	Jarque-Bera (JB):	27.47
Prob(Q):	1.00	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).

Forecast:



ARIMA forecast on ACF/PACF parameters

Summary:

ARIMA Model Results

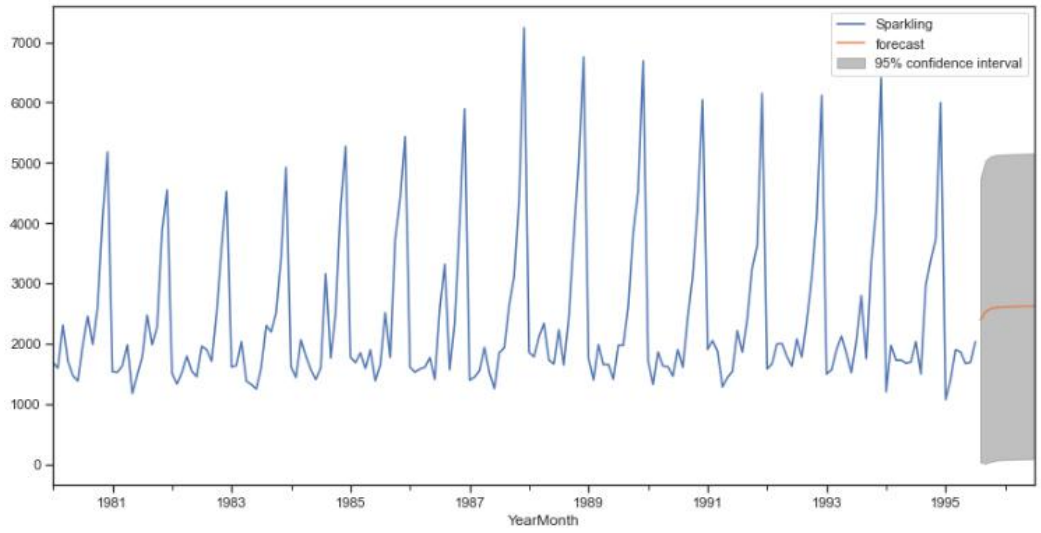
Dep. Variable:	D.Sparkling	No. Observations:	186
Model:	ARIMA(1, 1, 1)	Log Likelihood	-1585.295
Method:	css-mle	S.D. of innovations	1202.511
Date:	Thu, 23 Jul 2020	AIC	3178.590
Time:	20:08:25	BIC	3191.493
Sample:	02-29-1980 - 07-31-1995	HQIC	3183.819

	coef	std err	z	P>|z|	[0.025	0.975]
const	2.0785	2.554	0.814	0.416	-2.928	7.085
ar.L1.D.Sparkling	0.3681	0.069	5.370	0.000	0.234	0.503
ma.L1.D.Sparkling	-1.0000	0.014	-70.574	0.000	-1.028	-0.972

Roots

	Real	Imaginary	Modulus	Frequency
AR.1	2.7163	+0.0000j	2.7163	0.0000
MA.1	1.0000	+0.0000j	1.0000	0.0000

Forecast:



SARIMA forecast on ACF/PACF parameters

Summary:

SARIMAX Results

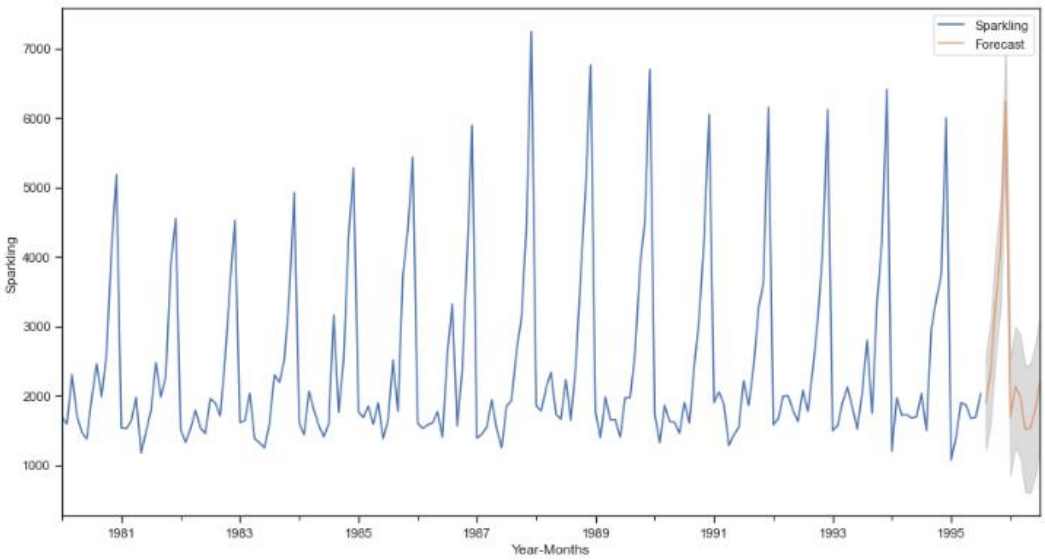
Dep. Variable:	Sparkling	No. Observations:	187
Model:	SARIMAX(1, 1, 1)x(2, 0, [1, 2, 3, 4, 5, 6, 7], 12)	Log Likelihood	-756.831
Date:	Thu, 23 Jul 2020	AIC	1537.663
Time:	20:13:26	BIC	1568.925
Sample:	01-31-1980 - 07-31-1995	HQIC	1550.315
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	0.2072	0.172	1.203	0.229	-0.130	0.545
ma.L1	-1.2880	0.259	-4.979	0.000	-1.795	-0.781
ar.S.L12	0.9041	0.149	6.072	0.000	0.612	1.196
ar.S.L24	0.0793	0.149	0.531	0.595	-0.213	0.372
ma.S.L12	-1.0457	0.889	-1.177	0.239	-2.788	0.696
ma.S.L24	0.5704	0.422	1.352	0.176	-0.257	1.397
ma.S.L36	-1.5694	1.455	-1.079	0.281	-4.421	1.282
ma.S.L48	1.4401	0.966	1.490	0.136	-0.454	3.334
ma.S.L60	-0.1133	0.416	-0.273	0.785	-0.928	0.701
ma.S.L72	0.7538	0.645	1.169	0.242	-0.510	2.017
ma.S.L84	-1.2643	0.952	-1.328	0.184	-3.130	0.602
sigma2	3.025e+04	4.33e+04	0.699	0.484	-5.45e+04	1.15e+05

Ljung-Box (Q):	75.76	Jarque-Bera (JB):	4.47
Prob(Q):	0.00	Prob(JB):	0.11
Heteroskedasticity (H):	1.03	Skew:	0.31
Prob(H) (two-sided):	0.94	Kurtosis:	3.83

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

Forecast:



ROSE

ARIMA forecast on auto parameters

Summary:

ARIMA Model Results

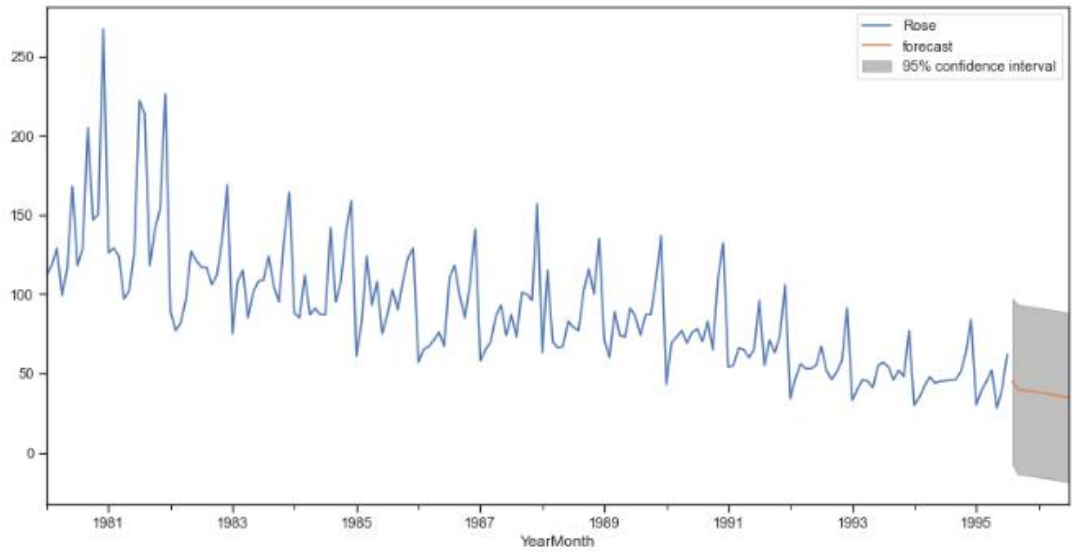
Dep. Variable:	D.Rose	No. Observations:	186
Model:	ARIMA(0, 1, 2)	Log Likelihood	-876.961
Method:	css-mle	S.D. of innovations	26.649
Date:	Thu, 23 Jul 2020	AIC	1761.922
Time:	14:41:28	BIC	1774.825
Sample:	02-29-1980 - 07-31-1995	HQIC	1767.151

	coef	std err	z	P>|z|	[0.025	0.975]
const	-0.5229	0.043	-12.026	0.000	-0.608	-0.438
ma.L1.D.Rose	-0.7923	0.082	-9.618	0.000	-0.954	-0.631
ma.L2.D.Rose	-0.2076	0.081	-2.572	0.010	-0.366	-0.049

Roots

	Real	Imaginary	Modulus	Frequency
MA.1	1.0000	+0.0000j	1.0000	0.0000
MA.2	-4.8166	+0.0000j	4.8166	0.5000

Forecast:



SARIMA forecast on auto parameters

Summary:

SARIMAX Results

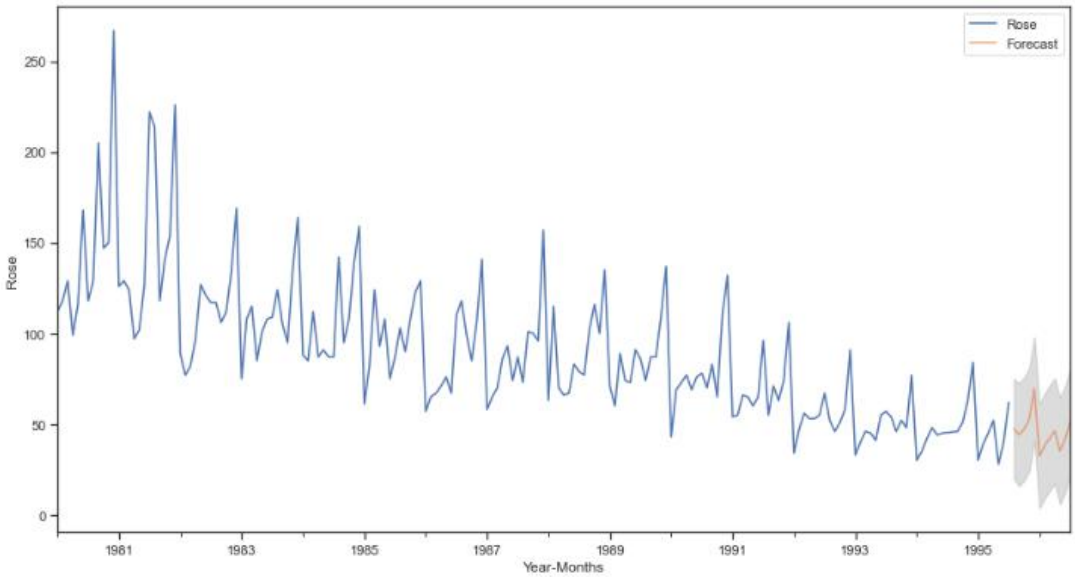
Dep. Variable:	Rose	No. Observations:	187
Model:	SARIMAX(0, 1, 2)x(2, 0, 2, 12)	Log Likelihood	-647.342
Date:	Thu, 23 Jul 2020	AIC	1308.684
Time:	15:28:44	BIC	1330.166
Sample:	01-31-1980	HQIC	1317.408
	- 07-31-1995		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ma.L1	-0.7659	0.088	-8.744	0.000	-0.938	-0.594
ma.L2	-0.1386	0.081	-1.713	0.087	-0.297	0.020
ar.S.L12	0.3978	0.052	7.622	0.000	0.295	0.500
ar.S.L24	0.3367	0.049	6.827	0.000	0.240	0.433
ma.S.L12	0.0126	0.089	0.141	0.888	-0.162	0.187
ma.S.L24	-0.1469	0.099	-1.482	0.138	-0.341	0.047
sigma2	199.2096	21.416	9.302	0.000	157.235	241.184

Ljung-Box (Q):	27.28	Jarque-Bera (JB):	8.75
Prob(Q):	0.94	Prob(JB):	0.01
Heteroskedasticity (H):	0.26	Skew:	0.50
Prob(H) (two-sided):	0.00	Kurtosis:	3.58

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

Forecast:



ARIMA forecast on ACF/PACF parameters

Summary:

ARIMA Model Results

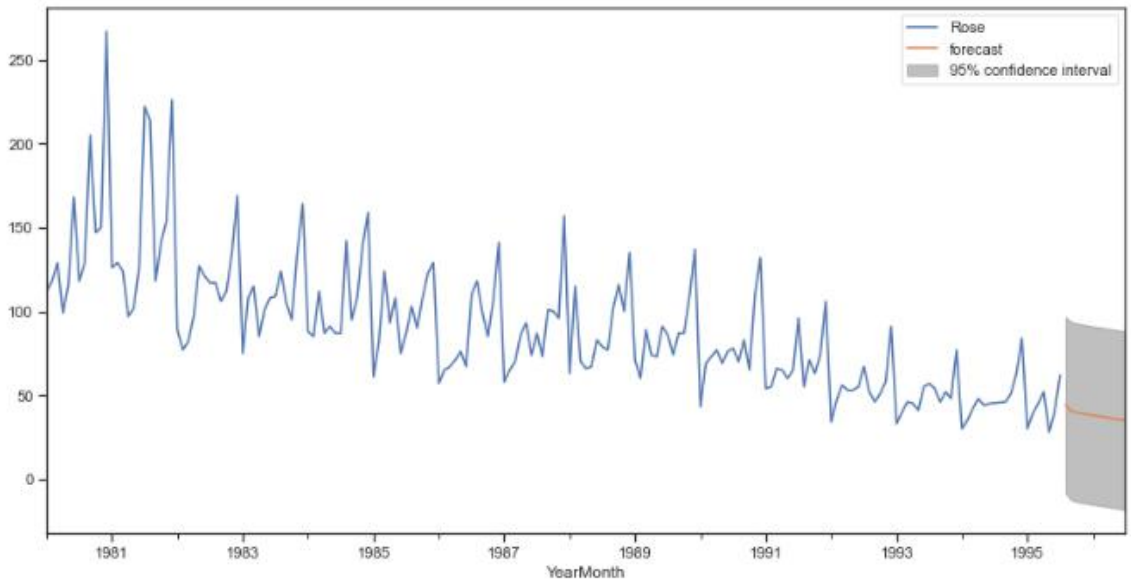
Dep. Variable:	D.Rose	No. Observations:	186
Model:	ARIMA(1, 1, 1)	Log Likelihood	-877.543
Method:	css-mle	S.D. of innovations	26.734
Date:	Thu, 23 Jul 2020	AIC	1763.086
Time:	15:32:08	BIC	1775.989
Sample:	02-29-1980 - 07-31-1995	HQIC	1768.315

	coef	std err	z	P>|z|	[0.025	0.975]
const	-0.5228	0.043	-12.044	0.000	-0.608	-0.438
ar.L1.D.Rose	0.1682	0.073	2.313	0.021	0.026	0.311
ma.L1.D.Rose	-0.9999	0.017	-59.434	0.000	-1.033	-0.967

Roots

	Real	Imaginary	Modulus	Frequency
AR.1	5.9441	+0.0000j	5.9441	0.0000
MA.1	1.0001	+0.0000j	1.0001	0.0000

Forecast:



SARIMA forecast on ACF/PACF parameters

Summary:

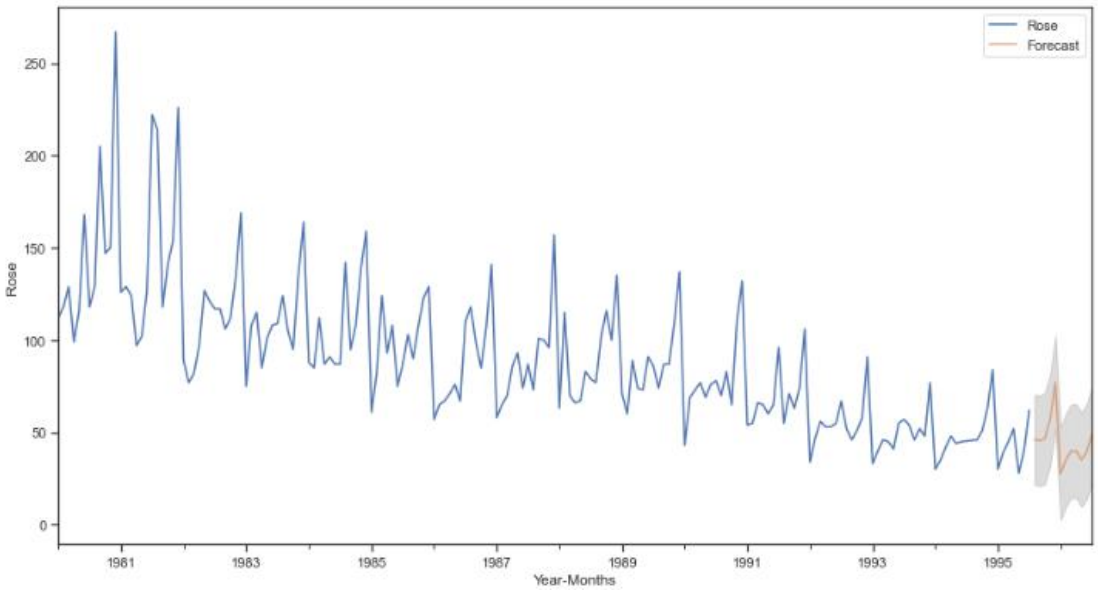
SARIMAX Results

Dep. Variable:	Rose	No. Observations:	187
Model:	SARIMAX(1, 1, 1)x(3, 0, [1, 2], 12)	Log Likelihood	-591.807
Date:	Thu, 23 Jul 2020	AIC	1199.614
Time:	15:32:51	BIC	1223.646
Sample:	01-31-1980	HQIC	1209.378
	- 07-31-1995		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	0.0554	0.085	0.654	0.513	-0.111	0.221
ma.L1	-1.1104	0.060	-18.571	0.000	-1.228	-0.993
ar.S.L12	0.7913	0.098	8.064	0.000	0.599	0.984
ar.S.L24	0.1186	0.097	1.224	0.221	-0.071	0.309
ar.S.L36	0.0205	0.057	0.362	0.718	-0.091	0.132
ma.S.L12	-0.5665	0.154	-3.668	0.000	-0.869	-0.264
ma.S.L24	-0.2142	0.125	-1.712	0.087	-0.459	0.031
sigma2	124.2743	19.240	6.459	0.000	86.565	161.983

Ljung-Box (Q):	39.63	Jarque-Bera (JB):	11.47
Prob(Q):	0.49	Prob(JB):	0.00
Heteroskedasticity (H):	0.39	Skew:	0.53
Prob(H) (two-sided):	0.00	Kurtosis:	3.85

Forecast:



2.10 Commenting on the model thus built and reporting the findings

1. Sparkling data set shows clear seasonality and a kind of constant trend following through out the data set.
2. Rose data set shows clear decreasing trend and with descent seasonal patterns.
3. Sparkling and Rose sells best on December.
4. Triple Exponential Smoothing and Moving Average methods performs good on both Sparkling and Rose testing data, as it captures its both trend and seasonality.
5. SARIMA model performs good on forecasting the models than ARIMA models on both Sparkling and Rose as it contains seasonality.
6. For Sparkling, Triple Exponential Smoothing model on iterated method with Alpha=0.4, Beta=0.1 and Gamma=0.2 has least RMSE value on test data with 336.71 followed by SARIMA model with ACF/PACF of (1,1,1)(2,0,7,12) has 455.49 RMSE value.
7. For Rose, Triple Exponential Smoothing model on iterated method with Alpha=0.1, Beta=0.2 and Gamma=0.2 has least RMSE value on test data with 9.64 followed by 2 point Trailing Moving Average with 11.529 RMSE value. SARIMA model with ACF/PACF of (1,1,1)(3,0,2,12) has 18.959 RMSE value.
8. Since 2 point Trailing Moving Average is not much reliable for longer forecast, its better to go with Triple Exponential Smoothing model and SARIMA model with ACF/PACF for Sparkling and Rose.
9. From the above forecast models, even though the sales will be better in Decembers, the sales of Rose wine will be poor where the sales of Sparkling wine will be constant as before with much more increased sales in December.
10. ABC Estate Wines can concentrate more on Sparkling wines than Rose wine for increase in sales.

3 Appendix A – Source Code

Karthiheswar_26Jul_Sparkling.ipynb

Karthiheswar_26Jul_Rose.ipynb