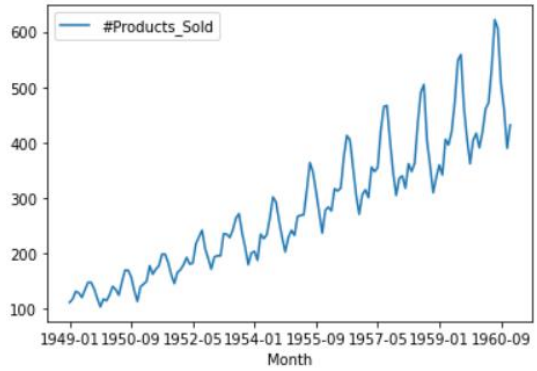
TIME SERIES FORECASTING

DATA SCIENCE

MAY 22, 2022

ASSIGNMENT – 3

Plotting



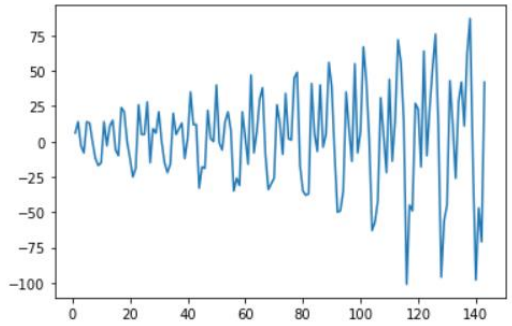
By seeing the graph, we get the idea that the about the hyper parameters p,d,q

p lies somewhere around 3

Differentiating:

1st Order:

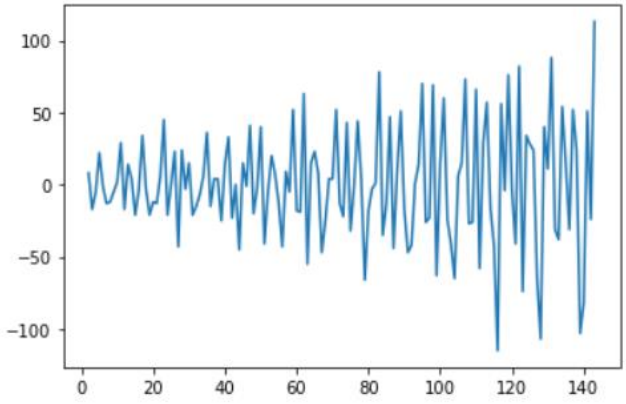
```
diff_1 = sr["#Products_Sold"].diff()
diff_1.plot()
pyplot.show()
```



We got promising results with first order; however, we will just check whether it can be improved.

2nd Order:

```
diff_2 = diff_1.diff()
pyplot.plot(diff_2)
pyplot.show()
```



We can still check further, however, the results indicated we can stop at 1.

Results of RMSE with ARIMA:

With (0,1,0):

Train set: 0.0

Testing set: 53.152

With (3,1,0):

Train set: 12.55

Testing set: 49.499

As the above model clearly indicates overfitting, the previous better hyper parameters gave better results as forecasted timeseries were predicted much accurately.

Results of RMSE with MLP_regressor:

With random_state=42, learning_rate="invscaling", solver="lbfgs"; window_size = 3

Train set: 35.79

Testing set: 40.281

With random_state=42, learning_rate="invscaling", solver="lbfgs"; window_size = 1

Train set: 35.57

Testing set: 53.789

As the above model clearly indicates overfitting, the previous better hyper parameters gave better results as forecasted timeseries were predicted much accurately.Verdict:

The results indicate ARIMA is a better model for time series data