Time Series Analysis

Steps to perform Time Series Analysis

- Importing the data.
- Cleaning(if required) and performing EDA on the data.
- Splitting the data into train(70%) and test(30%).
- Checking for trend and seasonality using seasonal decompose.
- Checking for stationarity using ADF and KPSS test.
- Detrending and de-seasonalizing the data if required.
- Plotting the ACF and PACF plots to get the lag values.
- Model Building(ARIMA(M1), Auto-ARIMA(M2), HOLT/SimpleExpSmoothing/ExponentialSmooting(M3)).
- Evaluating the models using AIC and RMSE values.

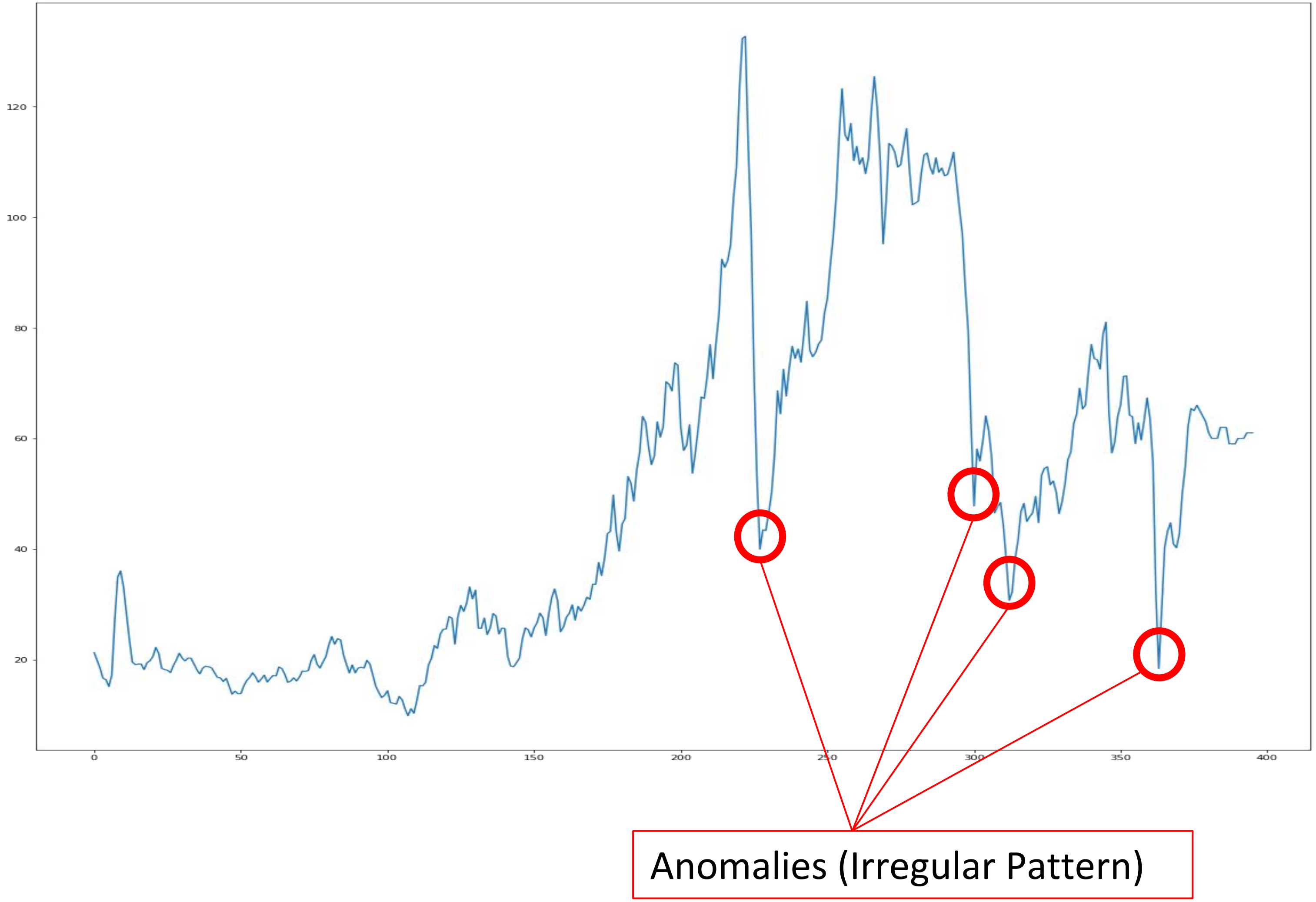
Datasets

- Brent Spot Prices
- Cali Emissions
- Coal Power
- HH Spot Price
- Imports Crude Oil

Brent Spot Prices

Problem Statement

- Using the data of Brent Spot Prices from 1990 to 2012, we need to perform a time series analysis to understand the changes in the price of Brent (Oil) and build a model to forecast the price for last 10 years to check the accuracy of the model.

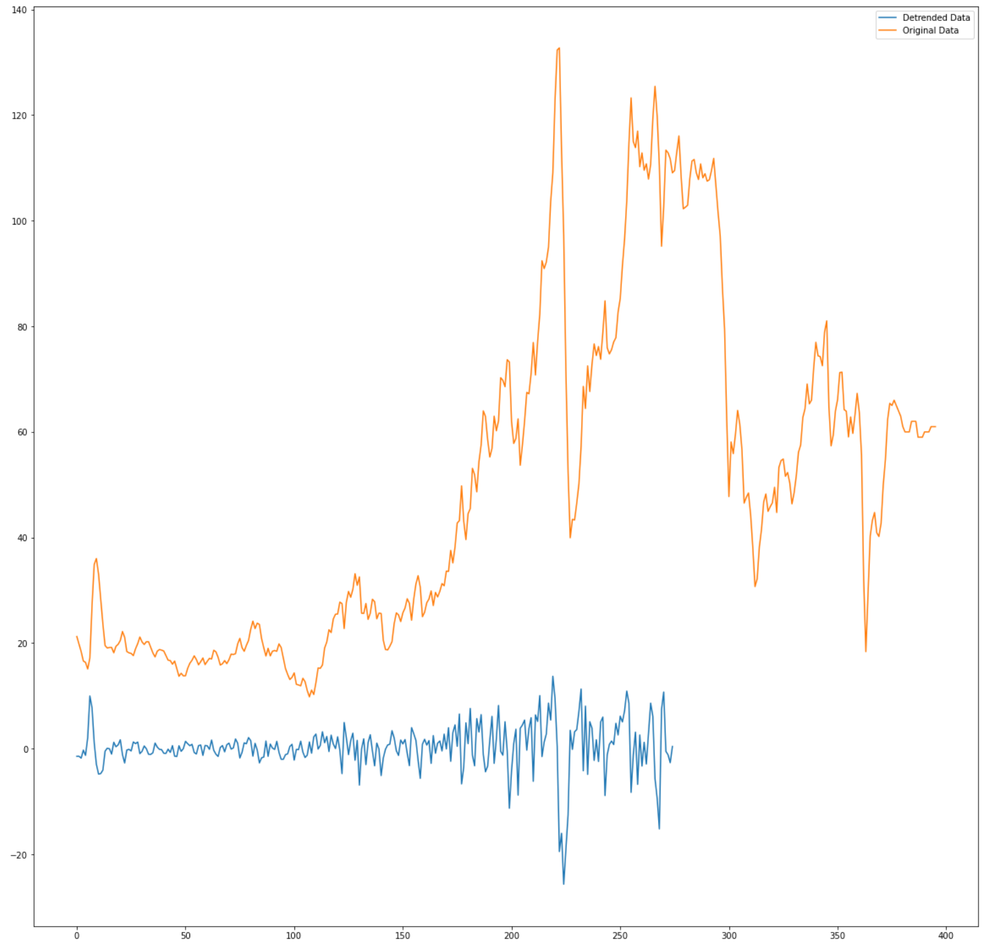


Observations Based on Raw Data

- Trend: - Data follows a stochastic trend i.e. the rise and drop in the values is highly dependent on several other factors other than time.
- Seasonality: - The data shows seasonality with a periodicity of (12 months)
- Cyclicity:- Due to the presence of anomalies there seems to be no cyclicity.
- Variance:- The Variance Keeps on increasing as the data moves ahead in time due to presence of anomalies and other factors as well.

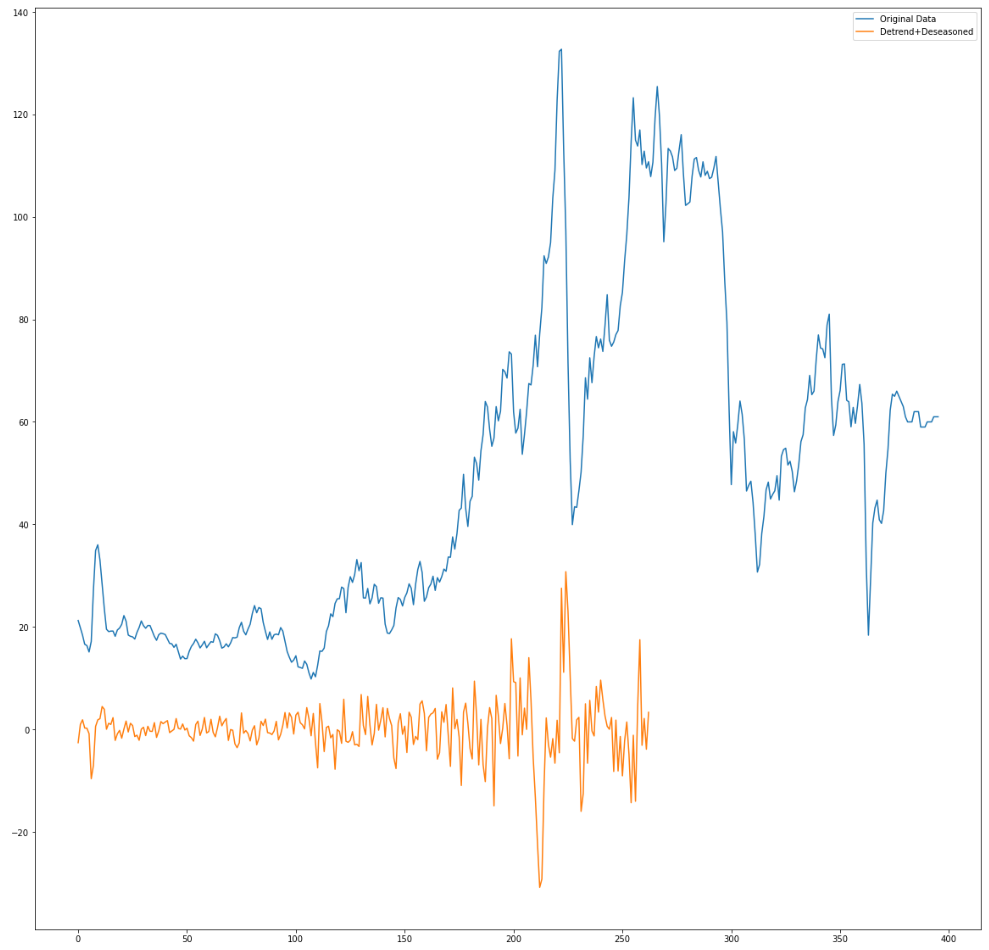
Observations Based on Detrending of data.

- The primary ADF test on the raw data shows that the data is not stationary.
- Since there is a multiplicative trend in data we need to detrend it first and check for stationarity, however in this case after detrending the data, it shows stationarity but the ACF and PACF plots don't generate valid results to move ahead.



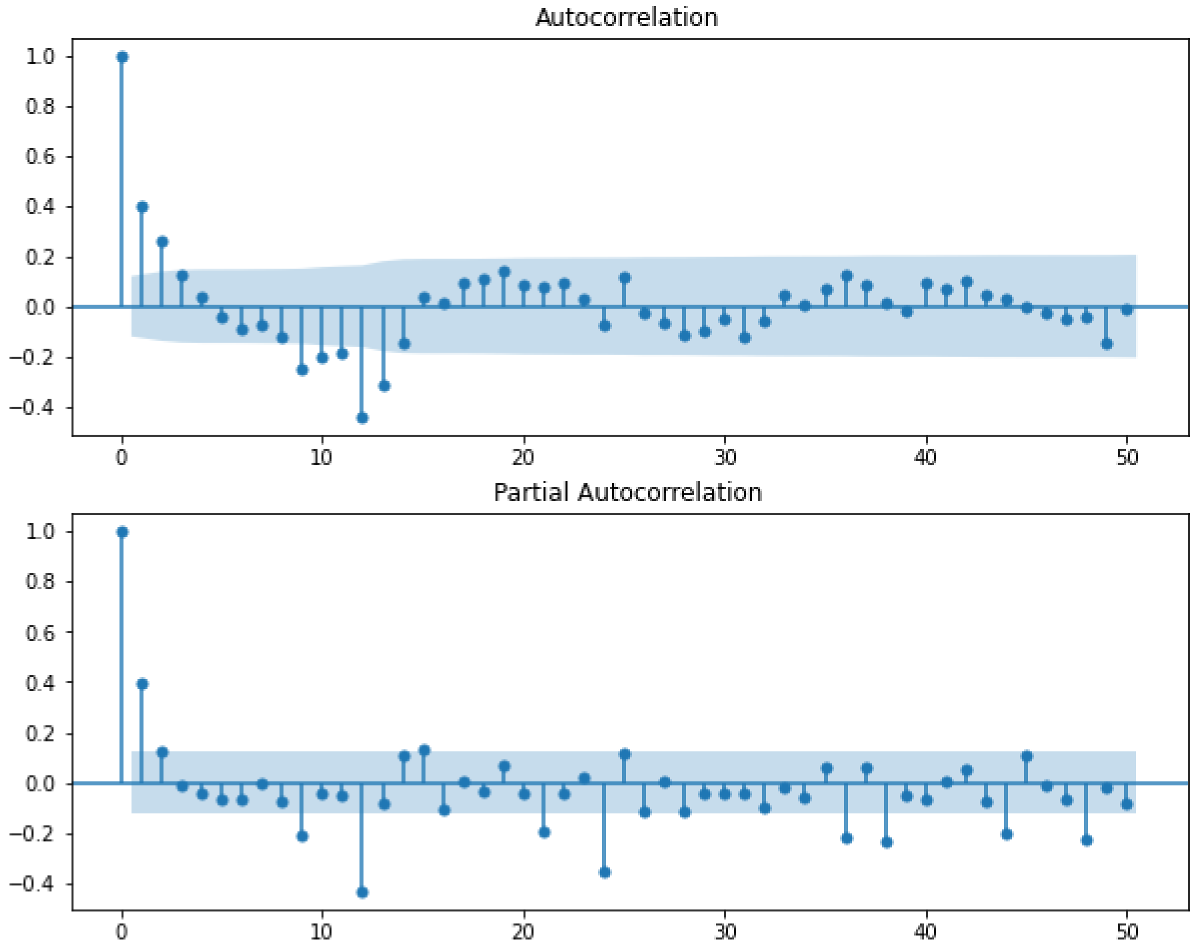
Observations Based on De-seasonalizing of data.

- We de-seasonalize the detrended data once and run the stationarity test, and found the data to be stationary.
- Hence further plot the ACF and PACF plot for the same data.



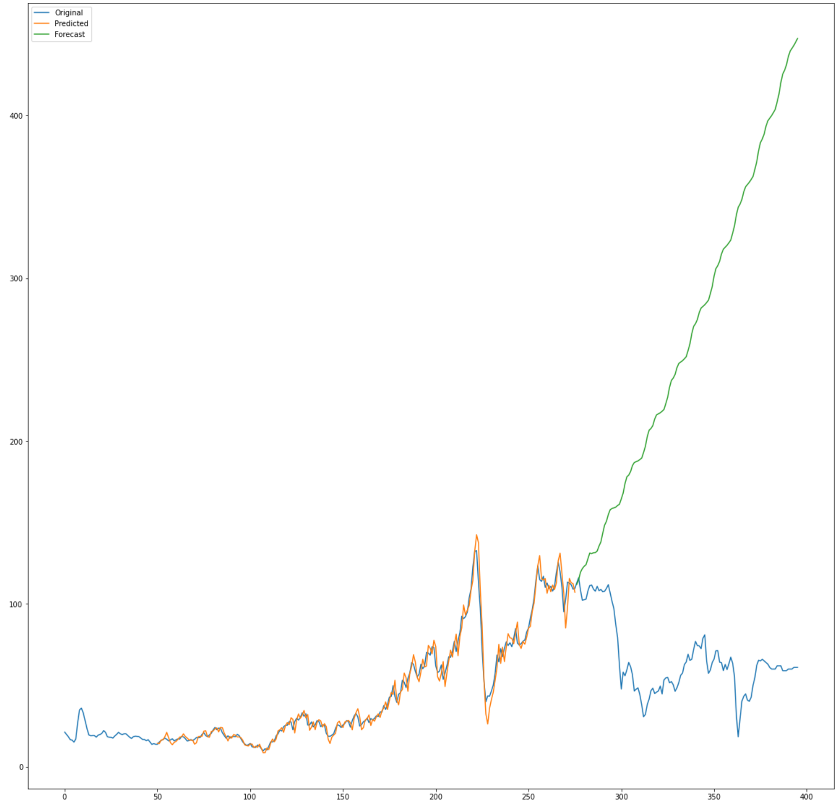
ACF and PACF Plots:-

- Further we plot the ACF and PACF where it is observed that
- (The ACF is gradual and PACF is sudden.)
- Hence it is AR(p) model with p = 1 from PACF plot .

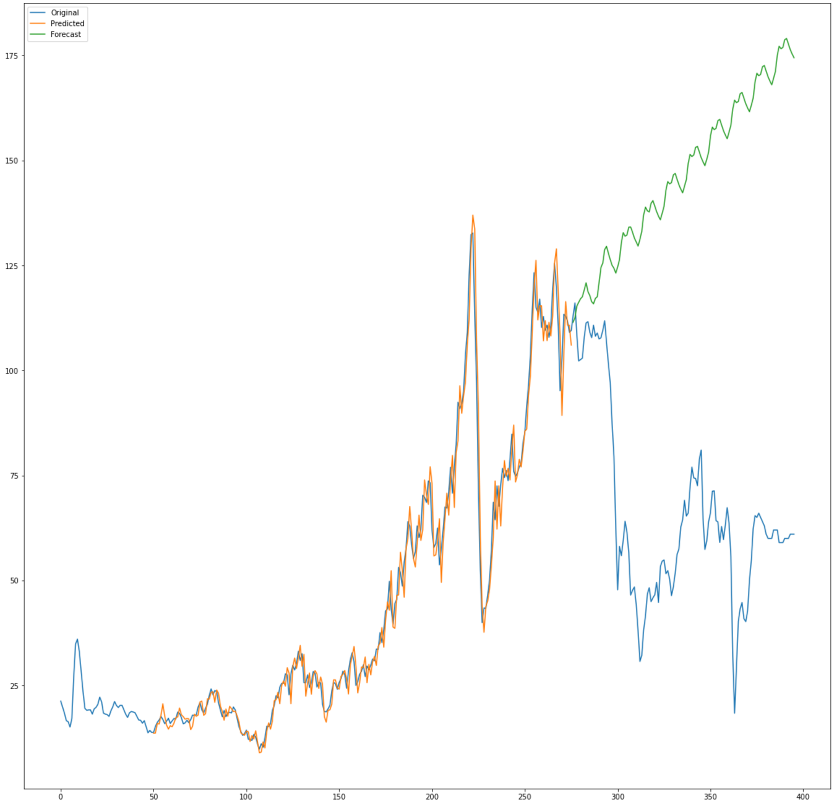


Results Of M1,M2 and M3 processes :-

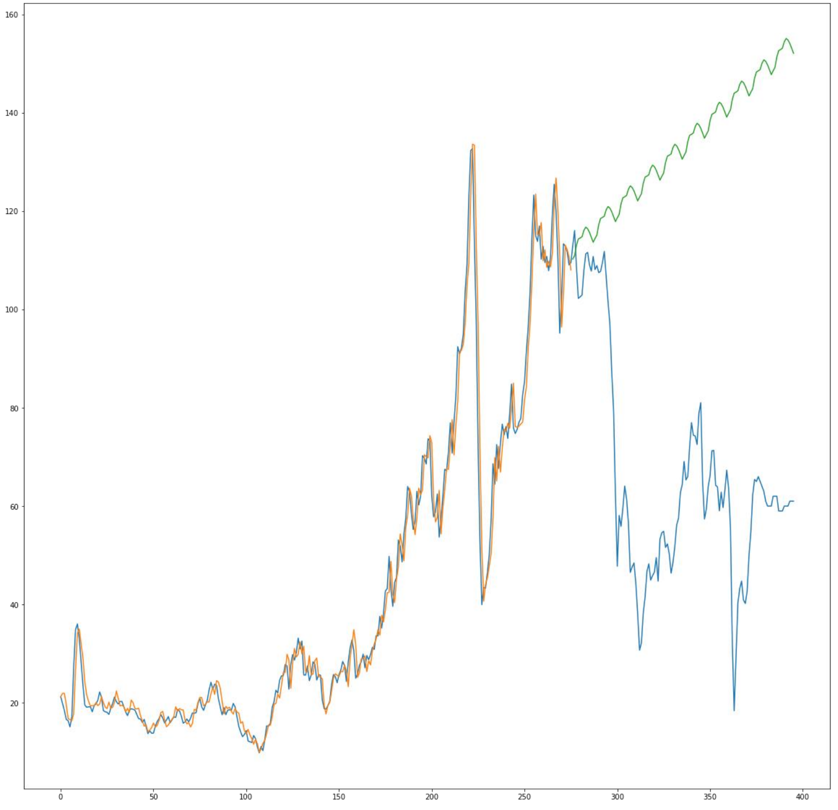
- In the below figure M1 we can see that for manual ARIMA the predicted values overlap on the original values but the forecast for the test values is naive and follows an upward trend which shows that the forecast is not correct and the model fails to give adequate results.
- Similarly in the figure M2 for Auto ARIMA as well the predicted values overlap on the original values but the forecast is naive and shows and upward trend which is also same for the M3 i.e.(Holt winters Exponential Smoothening).
- The overall observation from these three models M1,M2 & M3 is that they predict the values correctly for the train data but when it comes to forecast of the test data they fail to give adequate results.



M1



M2



M3

Model Analysis

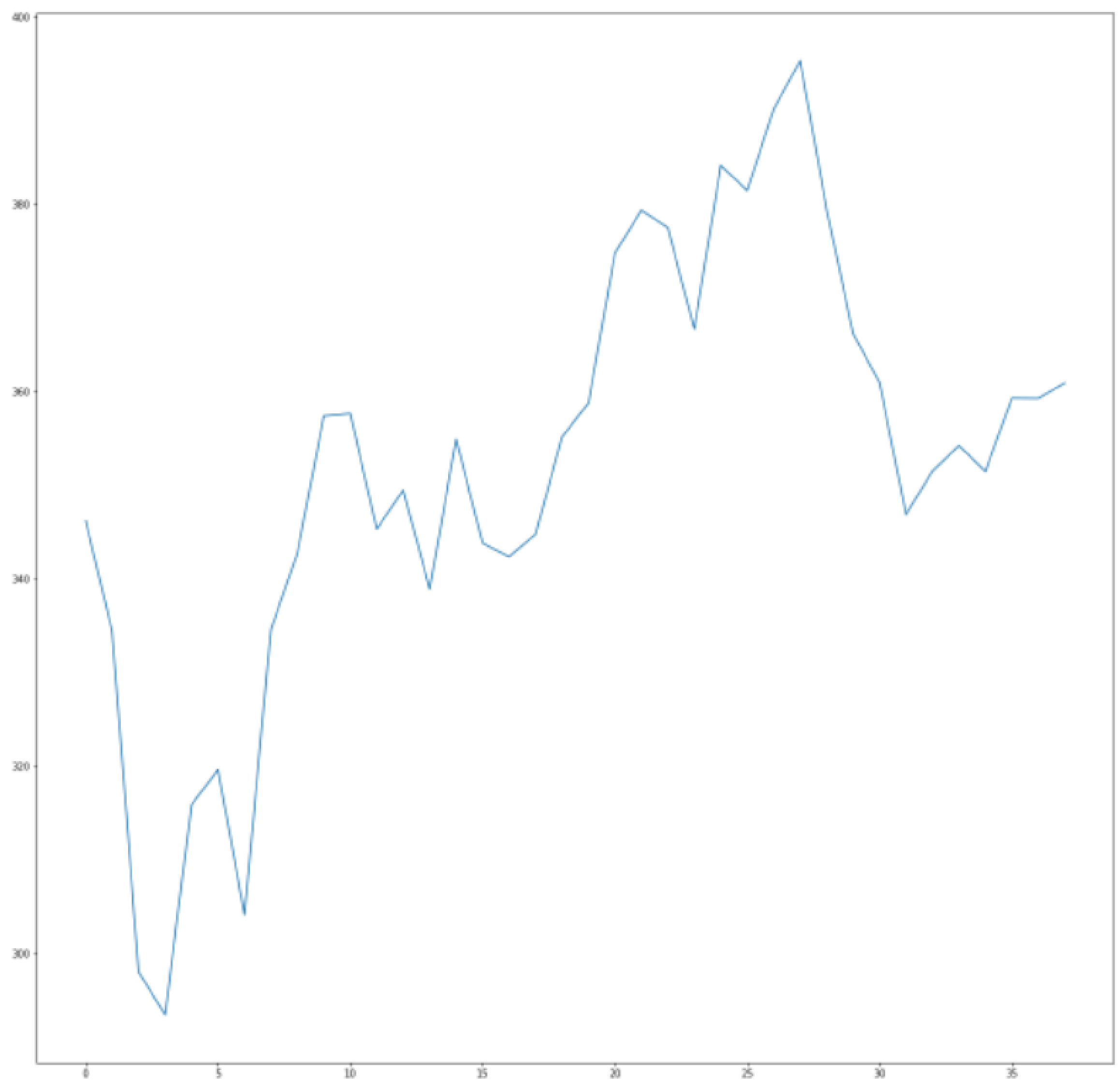
- Based on the model parameters mentioned in the below table for M1,M2 and M3 process by observing the RMSE and AIC values for each model It is clear that the M3 process for Holt Winters Exponential Smoothing gives the optimum parameters for the model and is the best model amongst the three models with RMSE value of 73.642 and AIC value of 852.002.
- After which the Auto Arima i.e.(pmdarima) gives the better results followed by the manual ARIMA process.
- Here we can see that though the RMSE value of manual ARIMA process is the lowest of all but the AIC values for the same are the highest ones which is why we focus on the AIC values to evaluate the goodness of the model.

	Final Model Parameters							**RMSE**	**AIC**
	p	**d**	**q**	**P**	**D**	**Q**	**Period**		
M1	1	1	0	0	1	0	12	73.316	1664.555
M2	1	1	0	2	1	1	12	88.48	1540.916
M3 (Exponential Smoothing)	-	-	-	-	-	-	12	73.642	852.002

Cali Emissions

Problem Statement

- Using the yearly(1980 to 2005) data of carbon dioxide emissions in California, we need to perform a time series analysis to understand the changes in the emission and build a model to forecast the emission for the years 2006-2017.

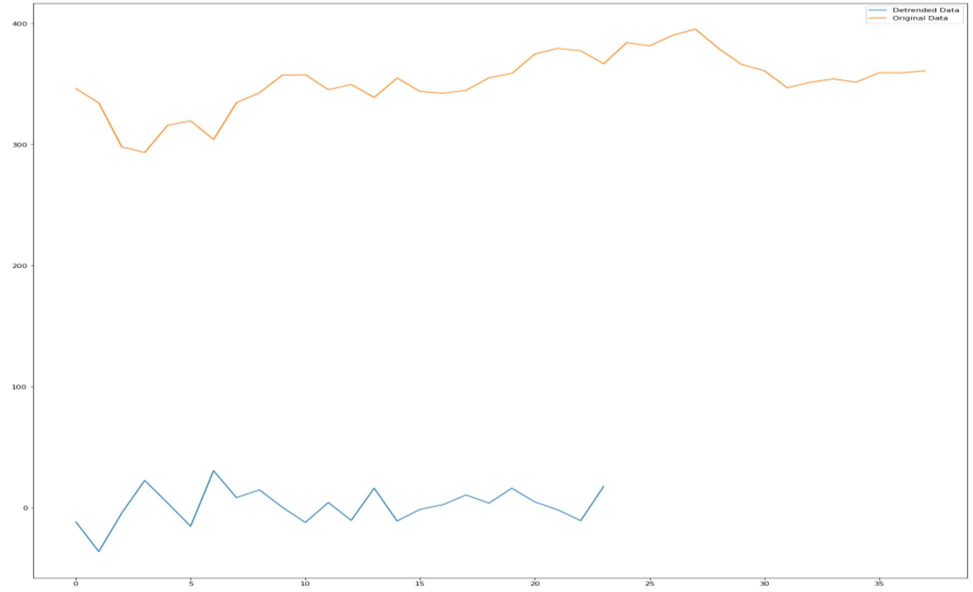


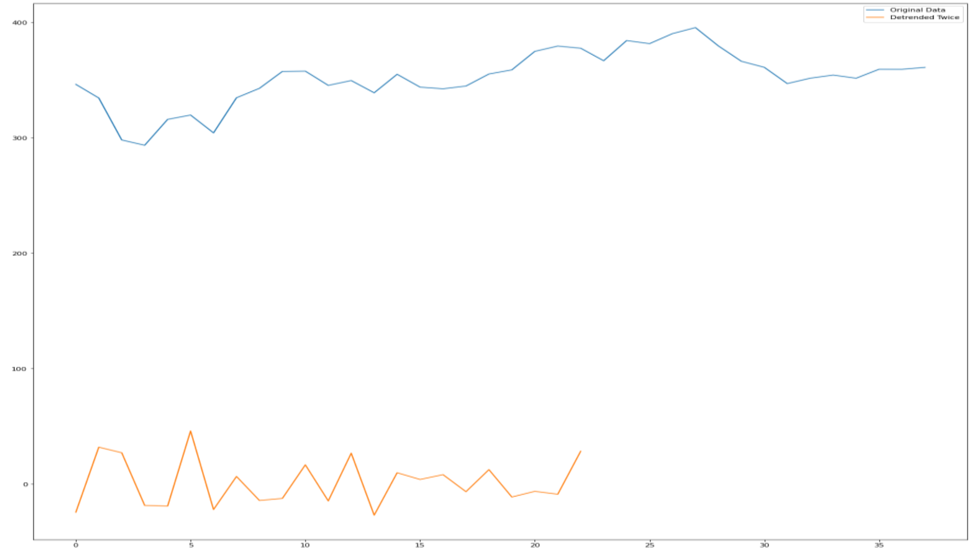
Observations Based on Raw Data

- Trend: - Data follows a rising and additive trend with respect to time
- Seasonality: - The raw data doesn’t show seasonality hence we assume there to be no seasonality.
- Cyclicity:- Since we assume that there is no seasonality therefore cyclicity is also absent.
- Variance:- There is more or less no variance in the data.

Observations Based on Detrending of data.

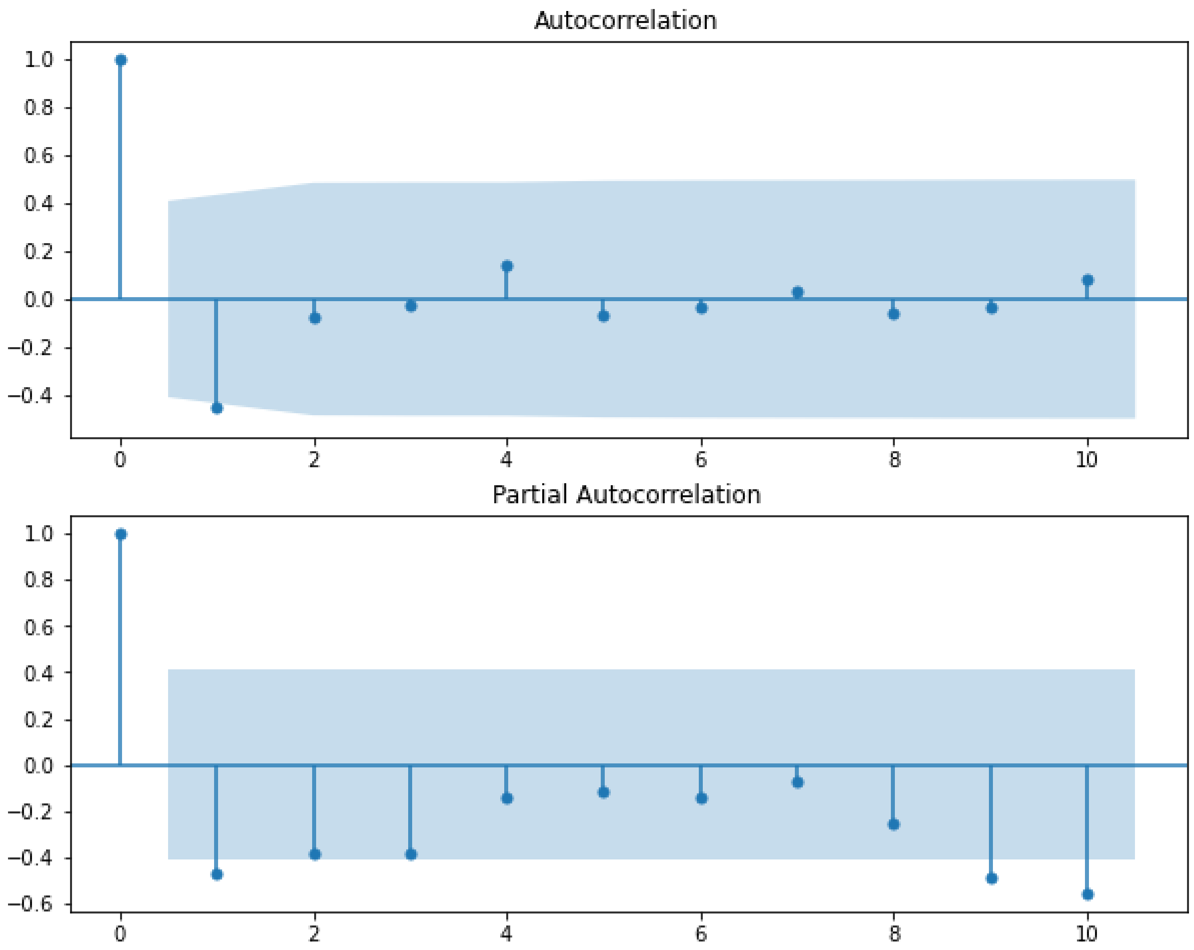
- The primary ADF test on the raw data shows that the data is not stationary.
- Since there is a additive trend in data we need to detrend it first and check for stationarity, however in this case after detrending the data, it shows stationarity but the ACF and PACF plots don't generate valid results to move ahead.
- So we detrend the detrended data once again and then by carrying out the stationarity test (i.e ADF test) plot the ACF & PACF .





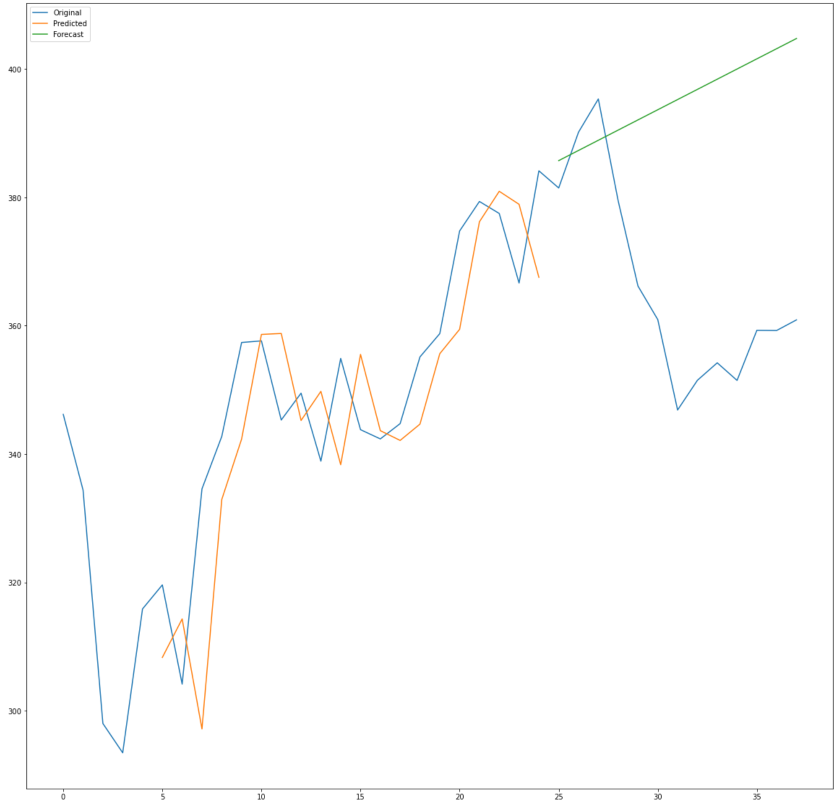
ACF and PACF Plots:-

- Based on the ACF & PACF plot we can see that the PACF is gradual and ACF is sudden.
- Hence it is Ma(q) model with q = 1 from ACF plot .

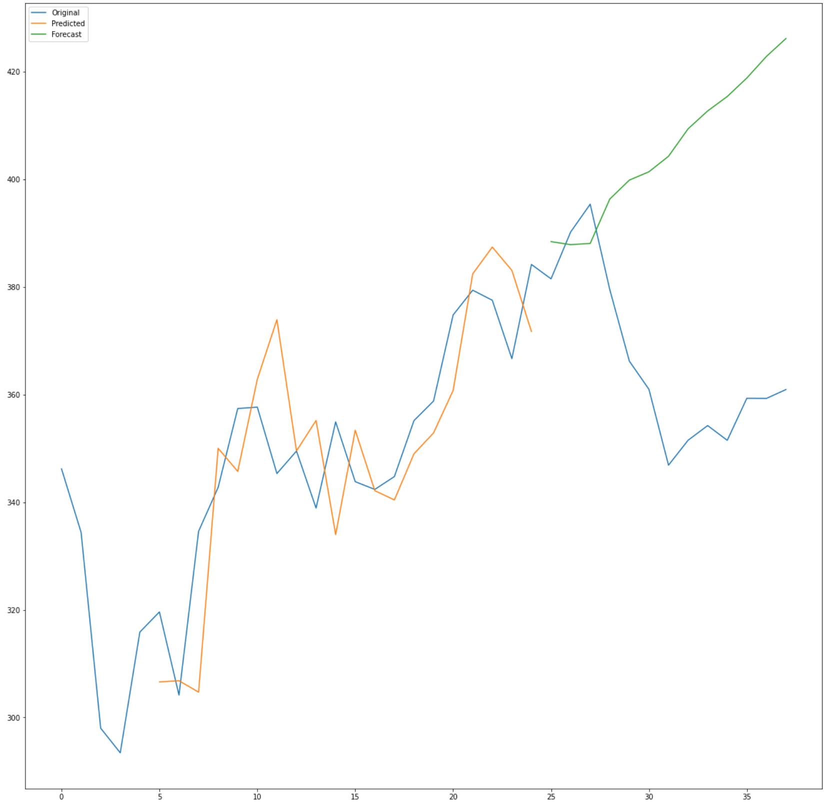


Results Of M1,M2 and M3 processes :-

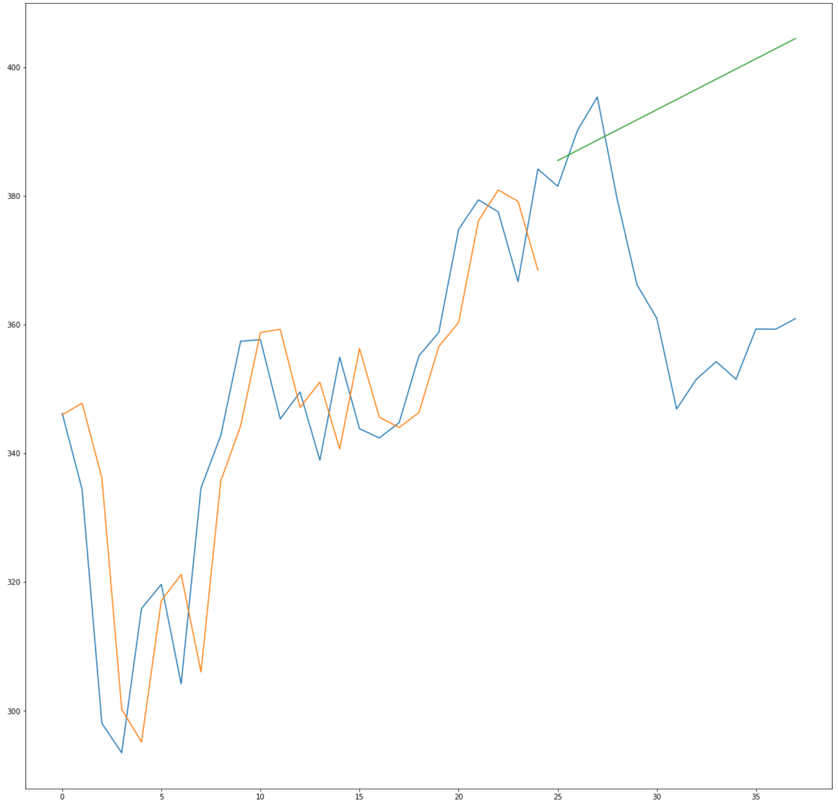
- In the below figure M1 we can see that for manual ARIMA the predicted values more or less overlap on the original values but the forecast for the test values is naive and follows an upward trend which shows that the forecast is not correct and the model fails to give adequate results.
- Similarly in the figure M2 for Auto ARIMA as well the predicted values more or less overlap on the original values but the forecast is naive and shows and upward trend which is also same for the M3 i.e.(Holt winters Model).
- The overall observation from these three models M1,M2 & M3 is that they predict the values correctly for the train data but when it comes to forecast of the test data they fail to give adequate results.



M1



M2



M3

Model Analysis

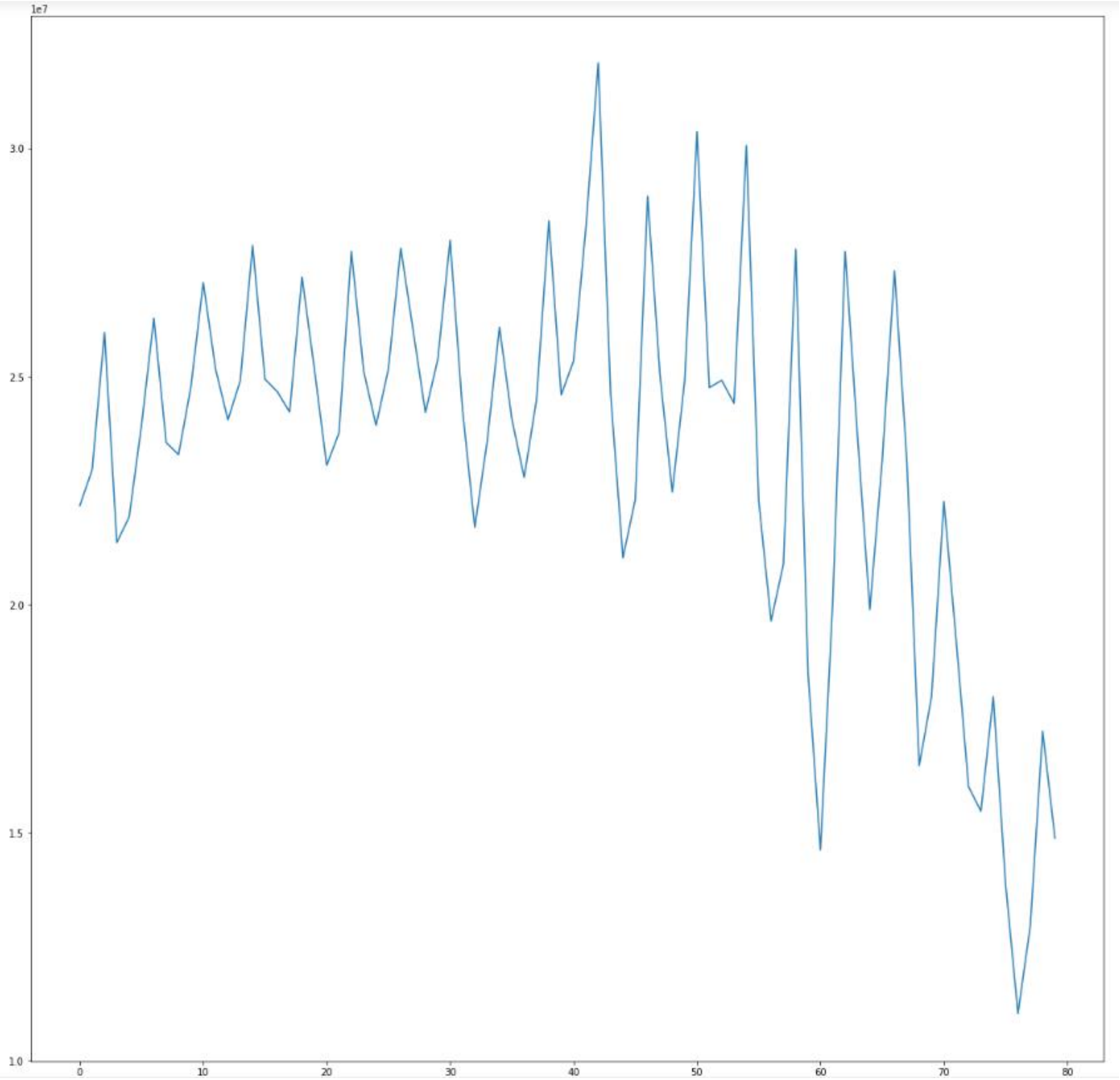
- Based on the model parameters mentioned in the below table for M1,M2 and M3 process by observing the RMSE and AIC values for each model It is clear that the M3 process for Holt Winters Model gives the optimum parameters for the model and is the best model amongst the three models with RMSE value of 35.128 and AIC value of 139.851.
- After which the Manual Arima gives the better results followed by the Auto ARIMA i.e.(pmdarima) process.

	Final Model Parameters							RMSE	AIC
	p	d	q	P	D	Q	Period		
M1	0	2	1	-	-	-	0	35.386	195.664
M2	3	2	0	-	-	-	0	47.322	200.126
M3 (Holt)	-	-	-	-	-	-	0	35.128	139.851

Coal Power

Problem Statement

- Using the quarterly data of Coal Consumption to generate electric power, we need to perform a time series analysis to understand the rate of consumption of coal and build a model to forecast the demand for the years 2015-2020.

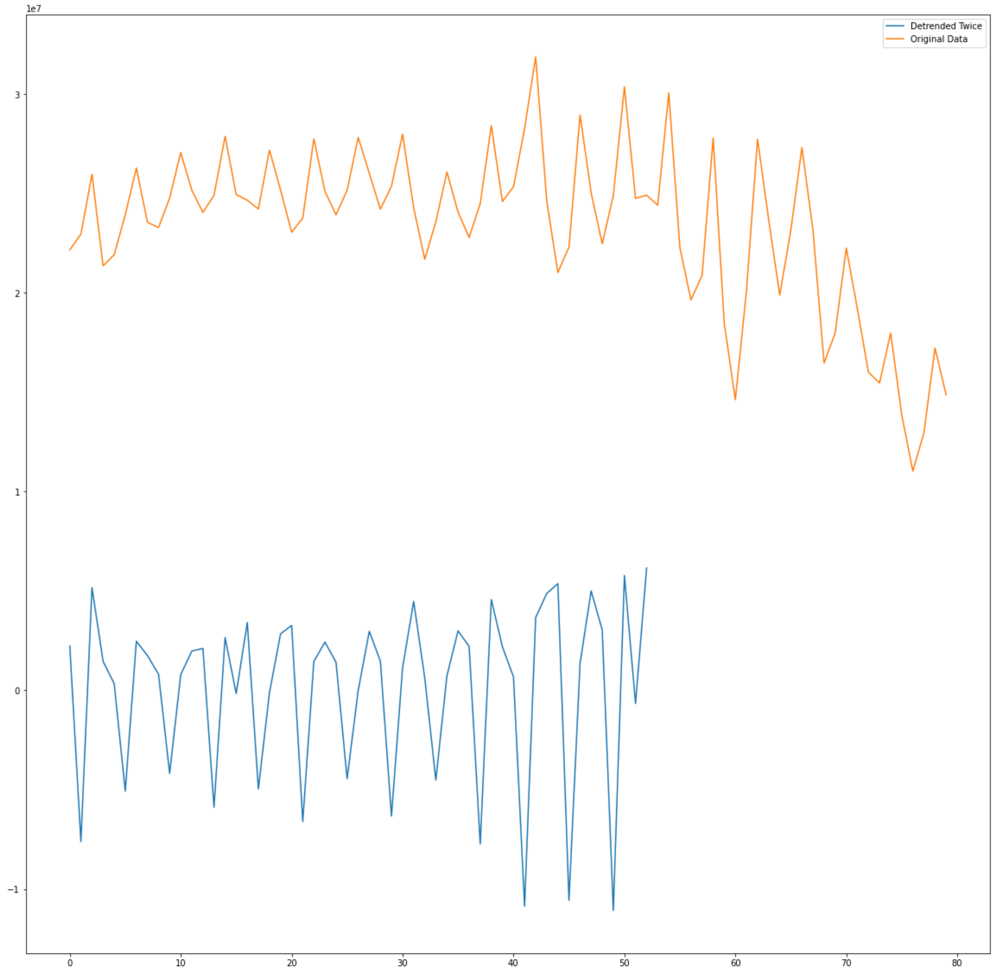


Observations Based on Raw Data

- Trend: - Data follows a exponential downward trend with respect to time
- Seasonality: - The data shows additive seasonality with periodicity of 4 (Since the data is quarterly)
- Cyclicity:- Since there is seasonality the data also shows cyclic patterns hence the cyclicity is present.
- Variance:- There seems to be no variance in the data

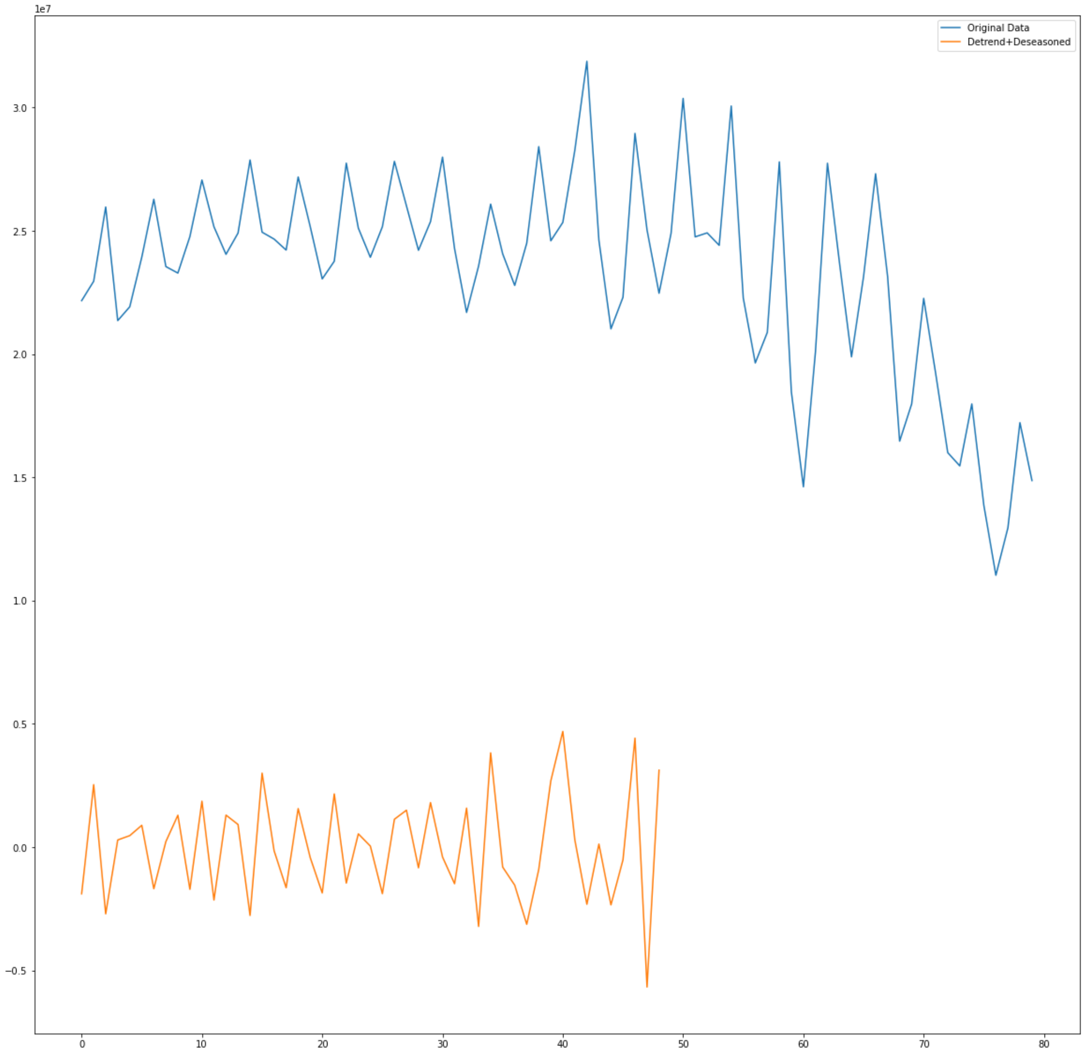
Observations Based on Detrending of data.

- The primary ADF test on the raw data shows that the data is not stationary.
- Since there is a downward trend in data we need to detrend it first and check for stationarity, however in this case after detrending the data, it doesn't show stationarity, Hence we detrend it twice.



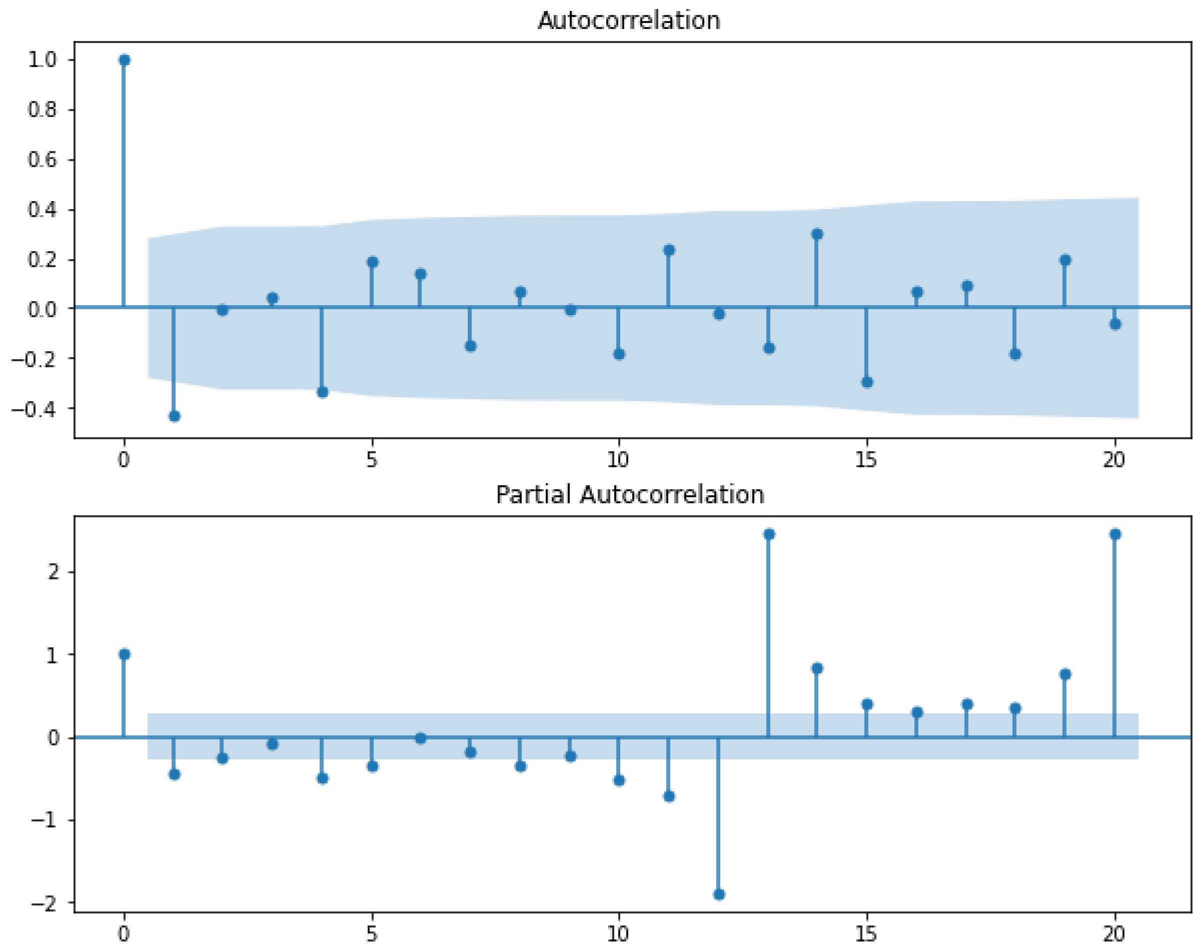
Observations Based on De-seasonalizing of data.

- We de-seasonalize the detrended data once since there is seasonality in the data and then run the stationarity test, and found the data to be stationary.
- Hence further plot the ACF and PACF plot for the same data



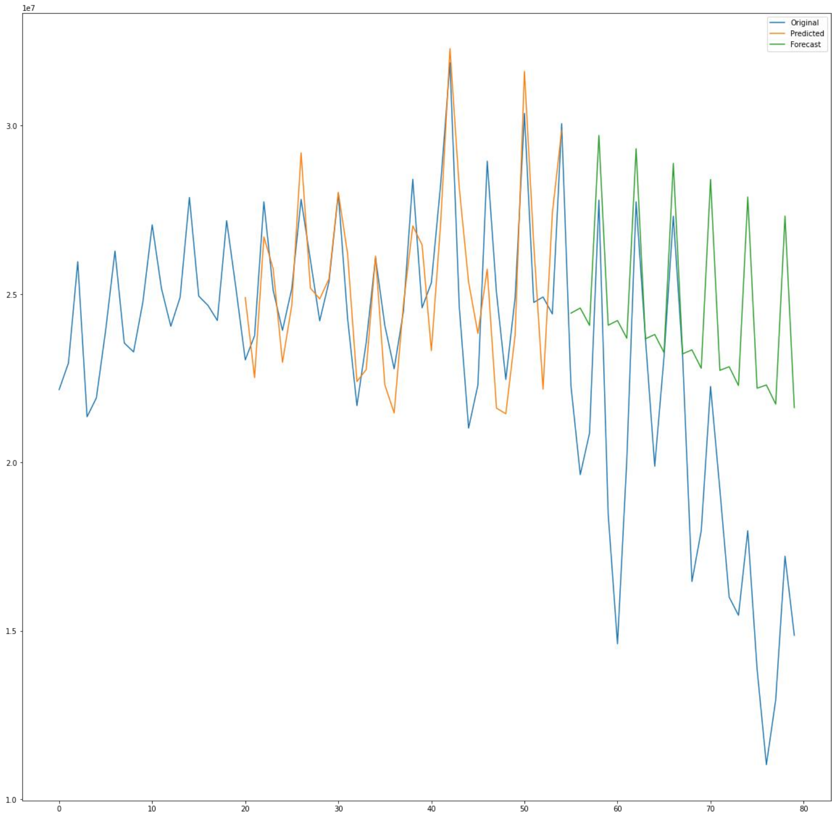
ACF and PACF Plots:-

- Further we plot the ACF and PACF where it is observed that
- (The PACF is gradual and ACF is sudden.)
- Hence it is MA(q) model with q = 1 from ACF plot .

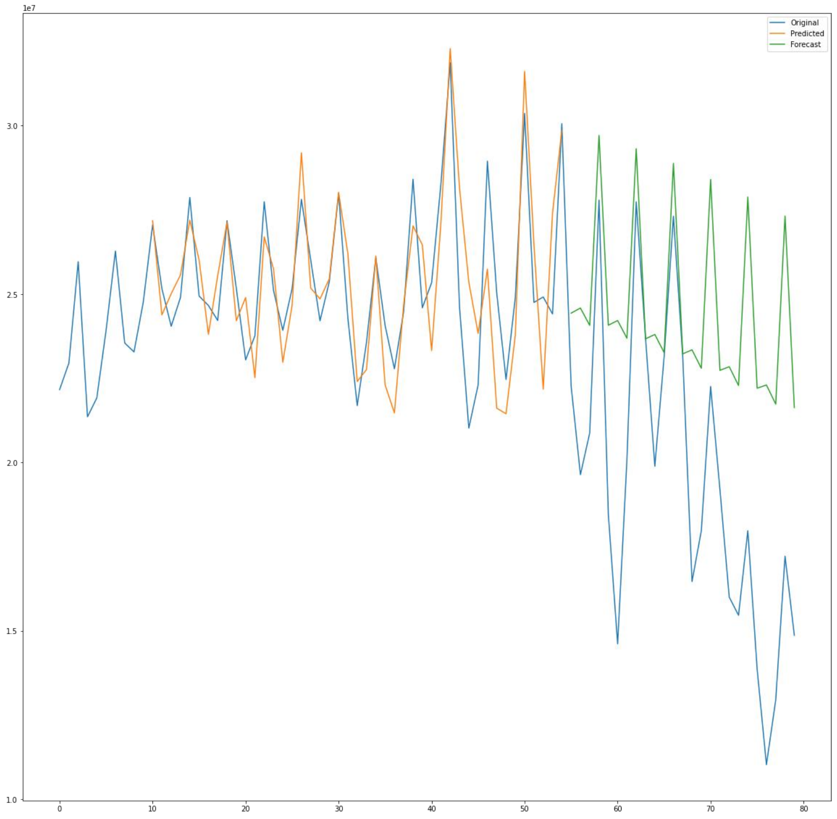


Results Of M1,M2 and M3 processes

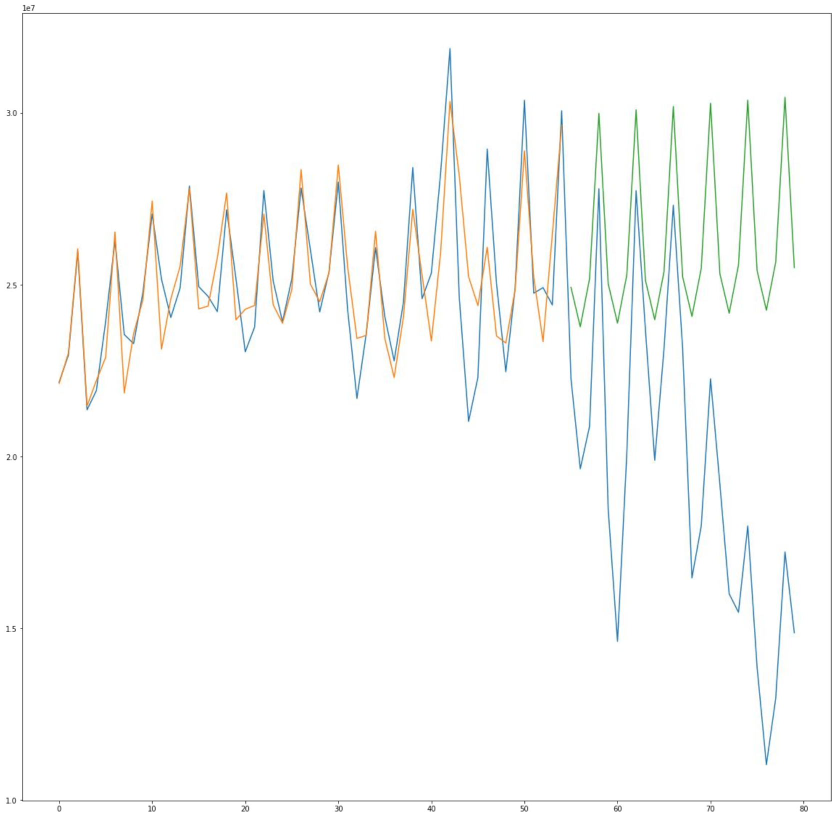
- In the below figure M1 we can see that for manual ARIMA the predicted values more or less overlap on the original values and the forecast for the test values follows a downward trend same as the test data with seasonality which is good and shows that the manual ARIMA model is showing better results as in comparison with the Auto ARIMA and Holt Winters Exponential Smoothening Process.
- Similarly in the figure M2 for Auto ARIMA as well the predicted values more or less overlap on the original values and the forecast for test data is similar to as that of Manual ARIMA. Whereas in M3 though the predictions for train data is accurate but the forecast for test data follows an upward trend. Which is why to evaluate the best model among these three we need to look into the RMSE and AIC values of each.



M1



M2



M3

Model Analysis

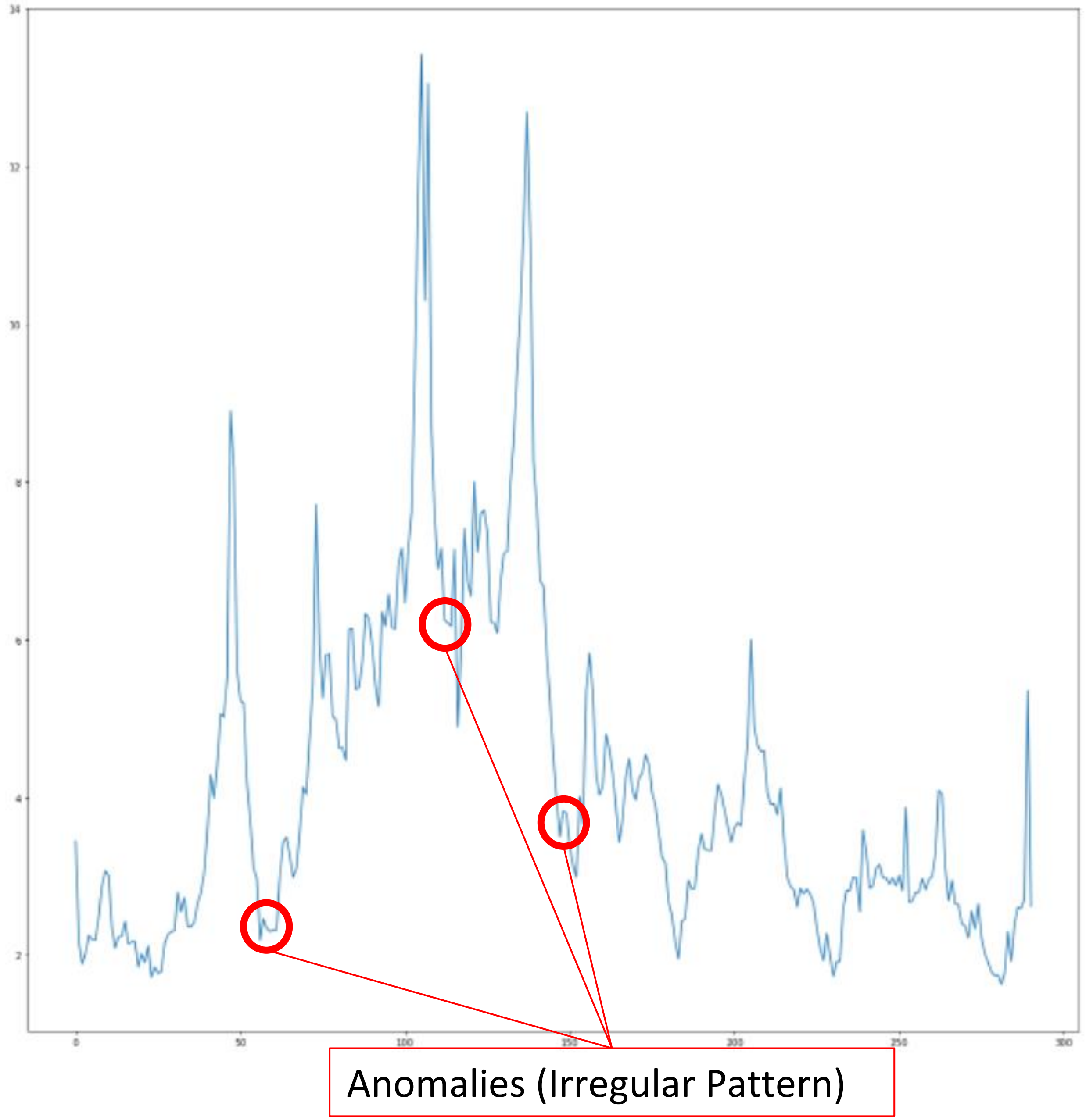
- Based on the model parameters mentioned in the below table for M1,M2 and M3 process by observing the RMSE and AIC values for each model It is clear that the M1 and M2 process of Manual ARIMA and Auto ARIMA gives the optimum parameters for the model and is the best model amongst the three models with RMSE values of 6111549.003 each and AIC value of 1545.793 each.
- Here the Holt Winters Exponential Smoothing model has the highest RMSE and AIC values which means it doesn’t give optimum results for this model which also broke our myth that Holt winters model would give best results in every circumstance.

	Final Model Parameters							**RMSE**	**AIC**
	p	**d**	**q**	**P**	**D**	**Q**	**Period**		
M1	0	2	1	0	1	0	4	6111549.003	1545.793
M2	0	2	1	0	1	0	4	6111549.003	1545.793
M3 (Exponential Smoothing)	-	-	-	-	-	-	4	9034610.355	1567.361

HH Spot Price

Problem Statement

- Using the monthly data of Henry Hub Natural Gas Spot Prices from 1997 to 2013, we need to perform a time series analysis to understand the changes in the price of Natural Gas and build a model to forecast the price for the years 2014 - 2021 to check the accuracy of the model.

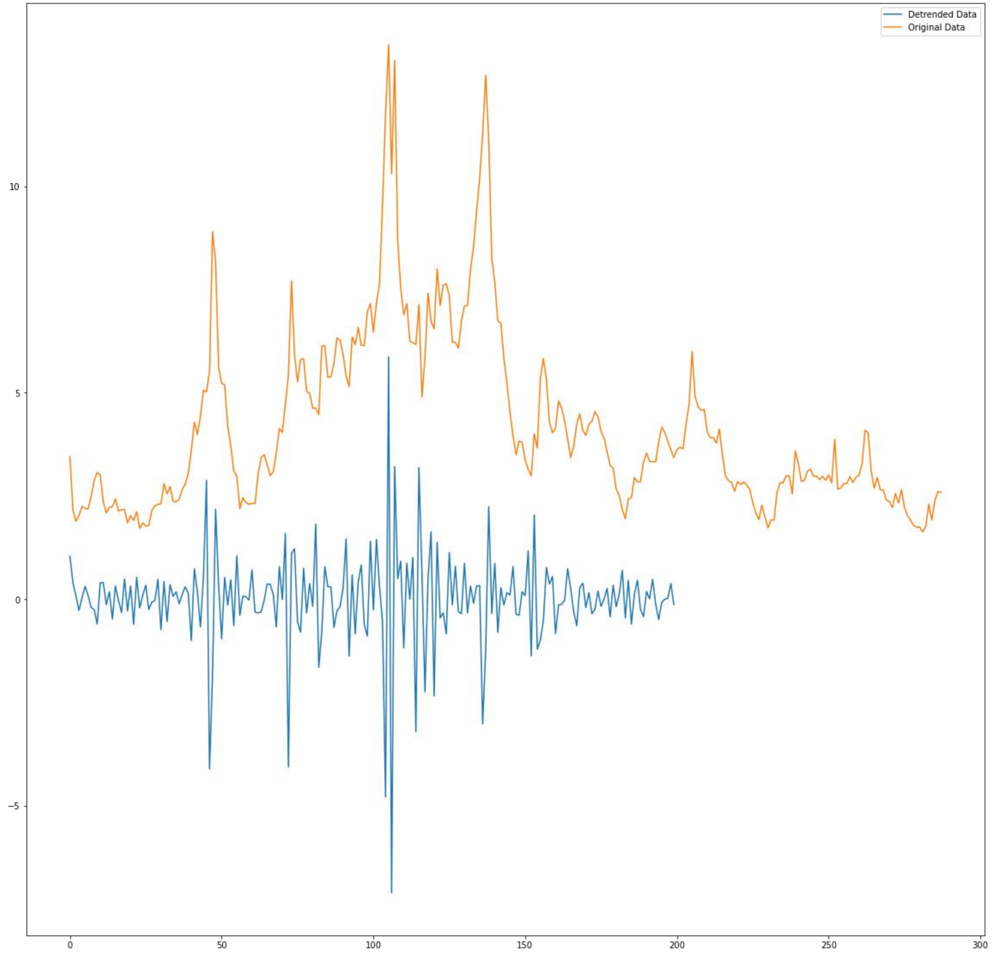


Observations Based on Raw Data

- Trend: - There is an irregular trend in the data due to the presence of anomalies.
- Seasonality: - The seasonality is present in the data with period= 12.
- Cyclicity:- The data is seasonal but there is no cyclicity
- Variance:- There is variance in the data due to the presence of anomalies.

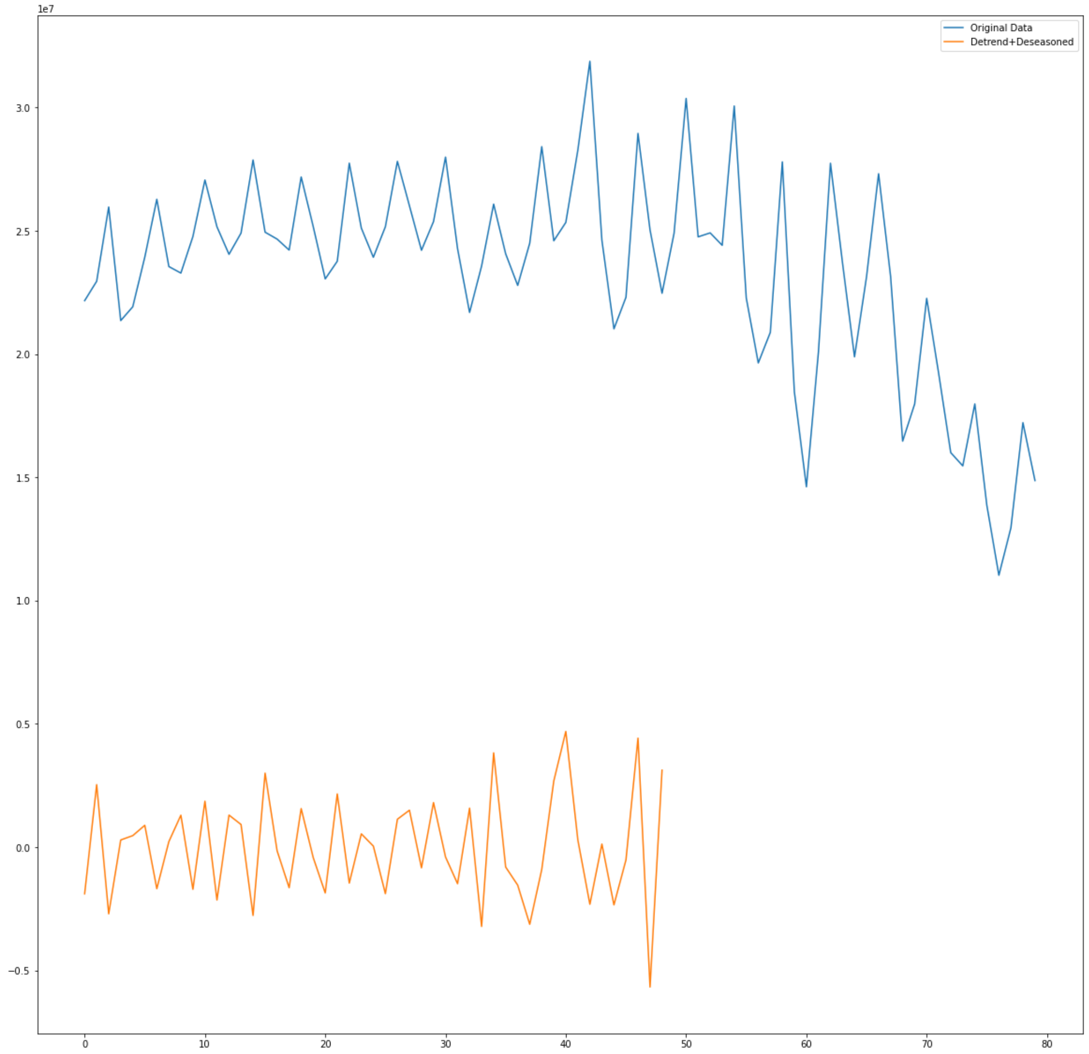
Observations Based on Detrending of data: -

- The primary ADF test on the raw data shows that the data is not stationary.
- Since there is a downward trend in data we need to detrend it first and check for stationarity, however in this case after detrending the data, though it becomes stationary but since we have considered it to be seasonal data we need to de-seasonalize it further.



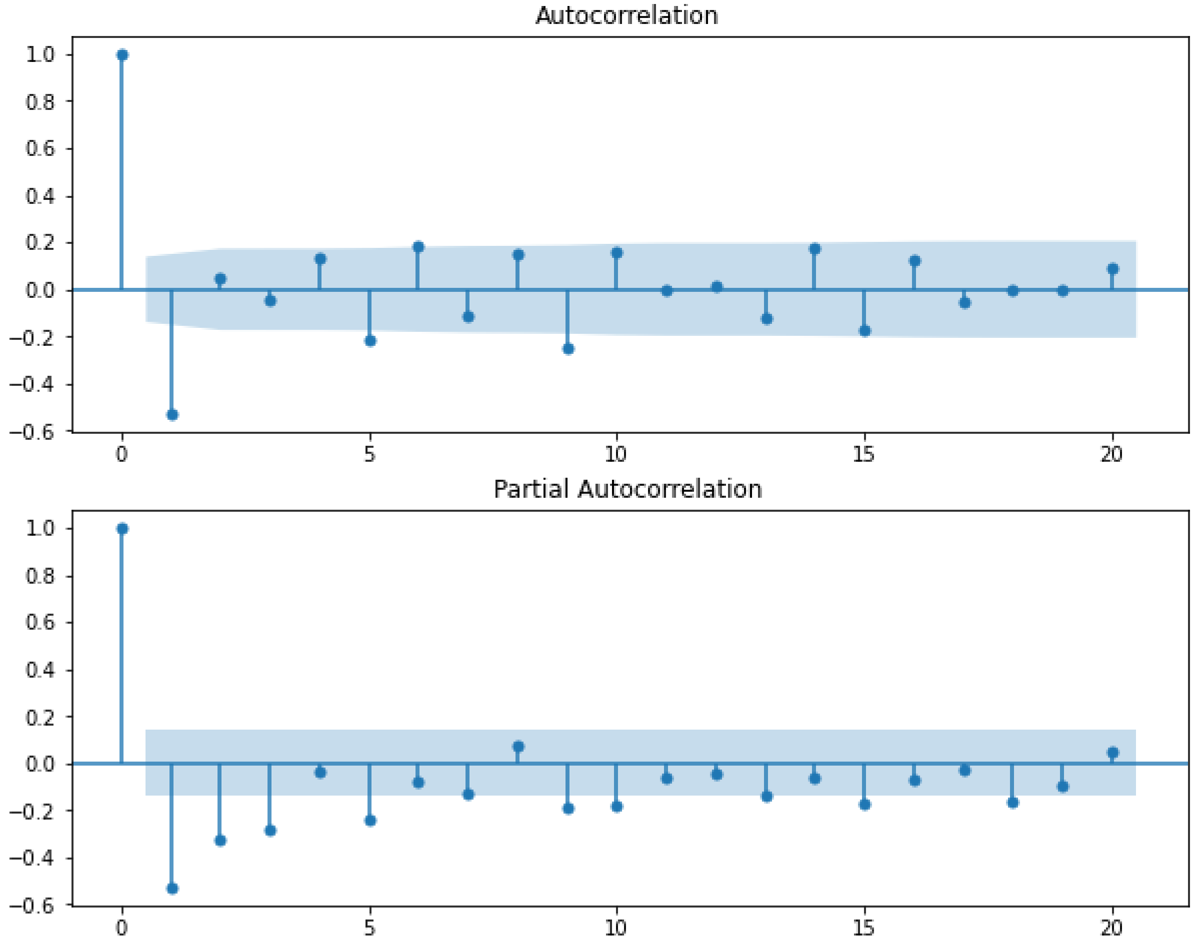
Observations Based on De-seasonalizing of data.

- We de-seasonalize the detrended data once since there is seasonality in the data and then run the stationarity test, and found the data to be stationary.
- Hence further plot the ACF and PACF plot for the same data



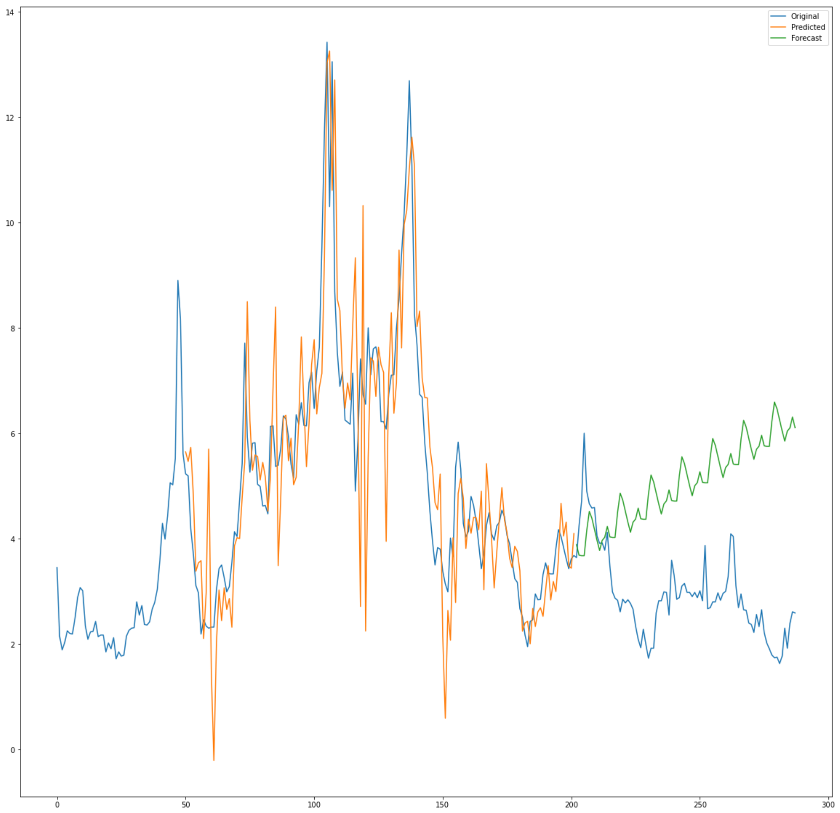
ACF and PACF Plots:-

- Further we plot the ACF and PACF where it is observed that
- (The PACF is gradual and ACF is sudden.)
- Hence it is MA(q) model with q = 1 from ACF plot .

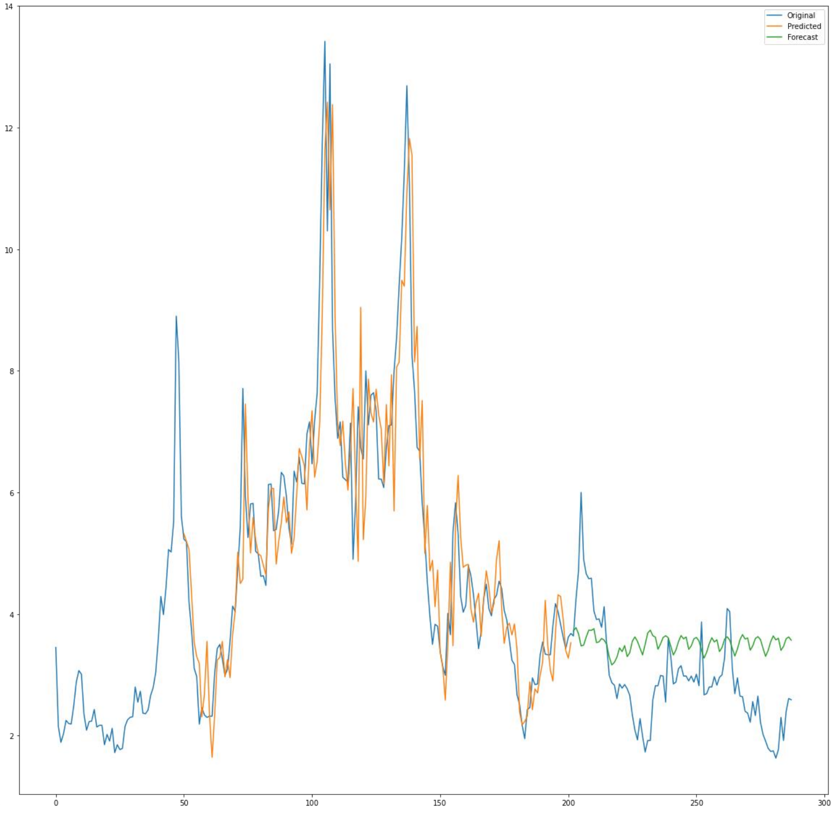


Results Of M1,M2 and M3 processes :-

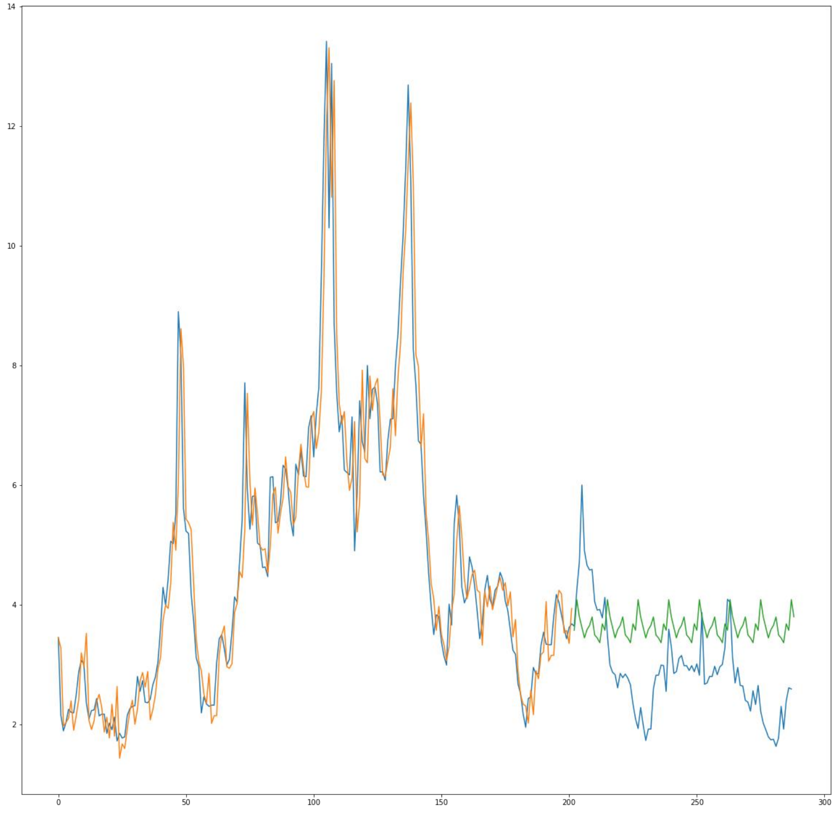
- In the below figure M1 we can see that for manual ARIMA the predicted values more or less overlap on the original values and the forecast for the test values follows an upward trend which is not correct because the test data here follows a downward trend.
- Similarly in the figure M2 for Auto ARIMA as well the predicted values more or less overlap on the original values and the forecast for test data shows that the forecast doesn't show any trend and also the forecast is naive and doesn't give expected output. Similarly for the M3 i.e.(Holt winters smoothing) the forecast shows similar trend as of M2 process. Which is why to evaluate the best model among these three we need to look into the RMSE and AIC values of each.



M1



M2



M3

Model Analysis

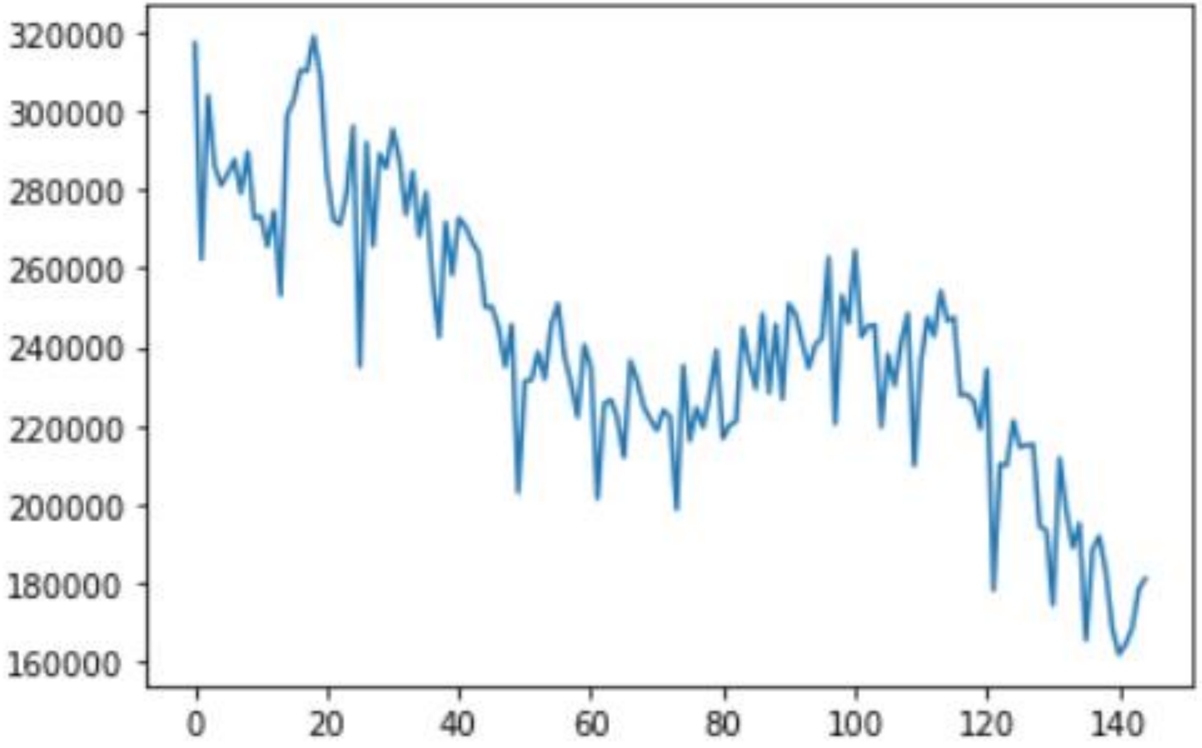
- Based on the model parameters mentioned in the below table for M1,M2 and M3 process by observing the RMSE and AIC values for each model It is clear that the M3 process of (Holt Winters Exponential Smoothing) gives the optimum parameters for the model and is the best model amongst the three models with RMSE values of 1.08 and AIC value of -60.219.
- Here the Manual Arima has the highest RMSE and AIC values which means it doesn't give optimum results for this data.

	Final Model Parameters							**RMSE**	**AIC**
	p	**d**	**q**	**P**	**D**	**Q**	**Period**		
M1	0	1	1	0	1	0	12	2.552	623.031
M2	1	0	0	2	1	0	12	1.00	546.958
M3 **(Exponential Smoothing)**	-	-	-	-	-	-	12	1.08	-60.219

Imports Crude Oil

Problem Statement

- Using the data of the imports of Crude Oil in the US, we need to perform a time series analysis to understand the quantity of oil imported and build a model to forecast the demand for the years 2018-2020.

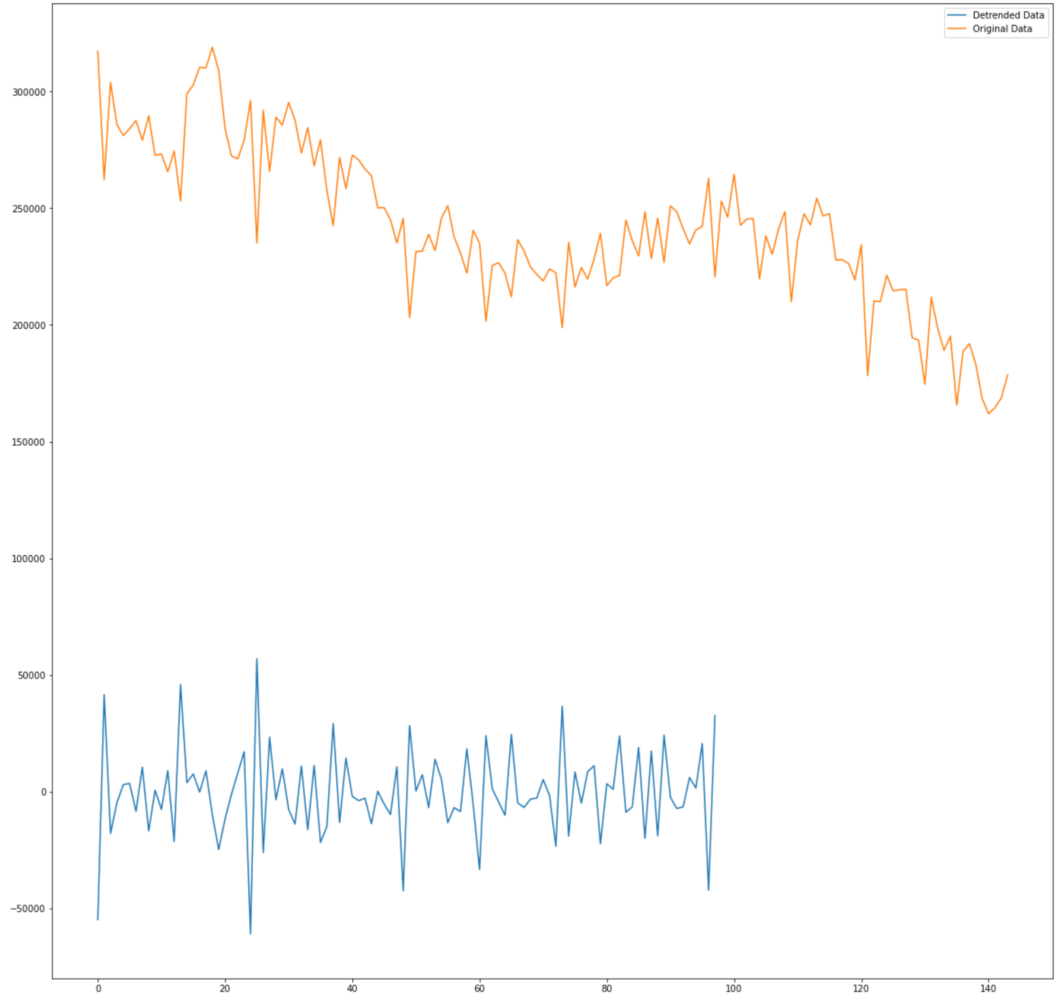


Observations Based on Raw Data

- Trend: - The data has a linear downward trend.
- Seasonality: - The seasonality is present in the data with period= 12.
- Cyclicity:- The data is seasonal but there is no cyclicity.
- Variance:- There is no variance in the data it remains more or less the same.

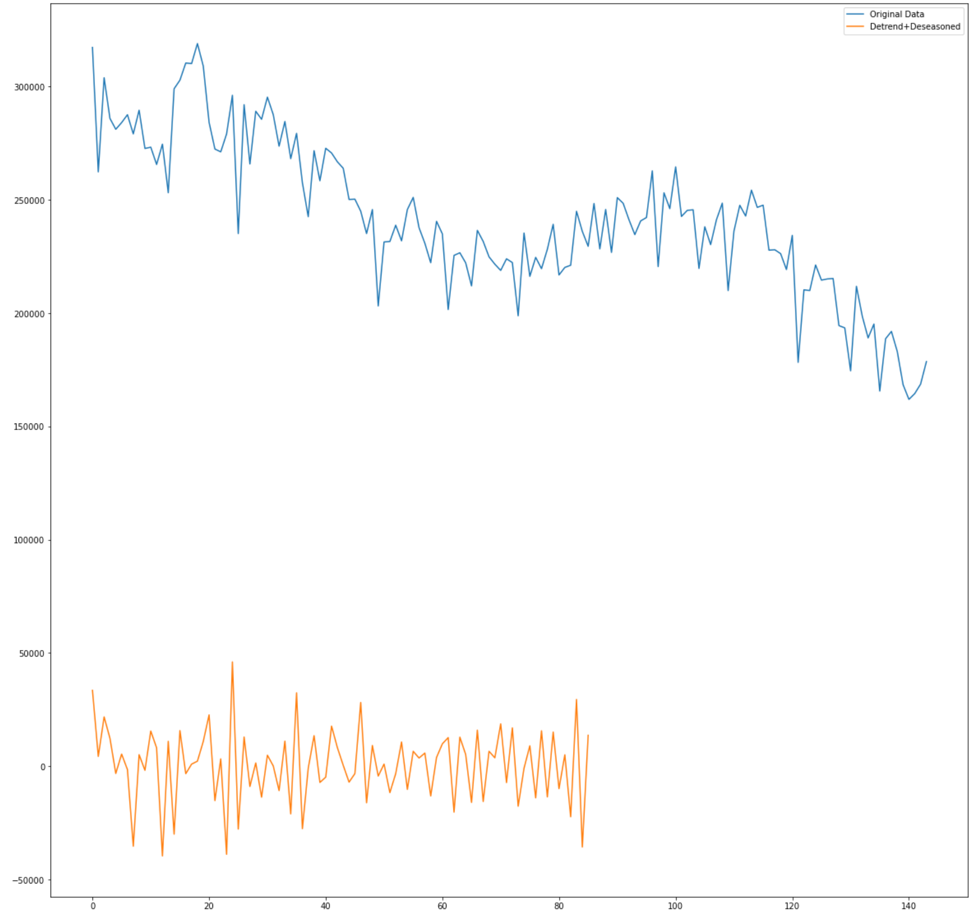
Observations Based on Detrending of data.

- The primary ADF test on the data shows that it is not stationary.
- Since there is a downward trend in data we need to detrend it first and check for stationarity, however in this case after detrending the data, it doesn't show stationarity, Hence we de-seasonalize it further.



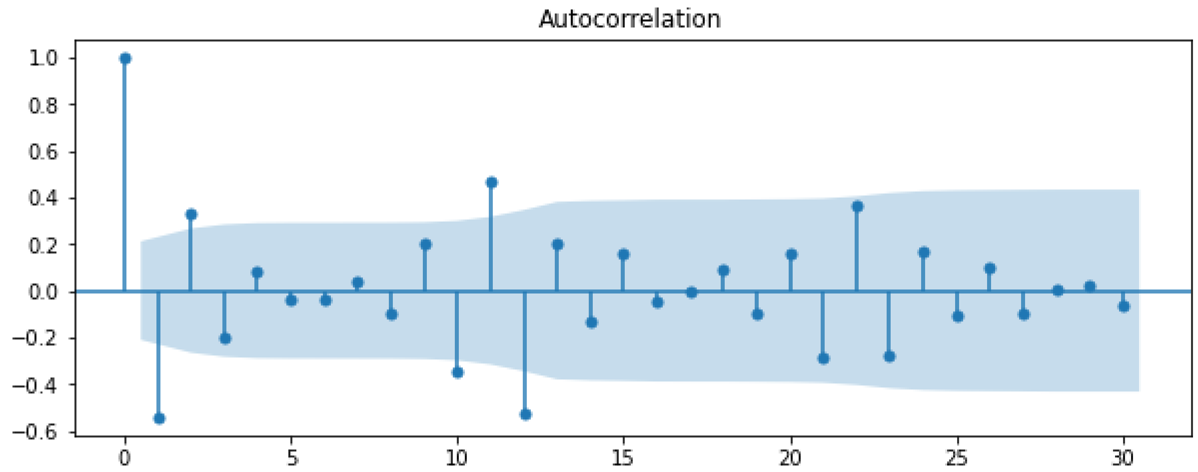
Observations Based on De-seasonalizing of data.

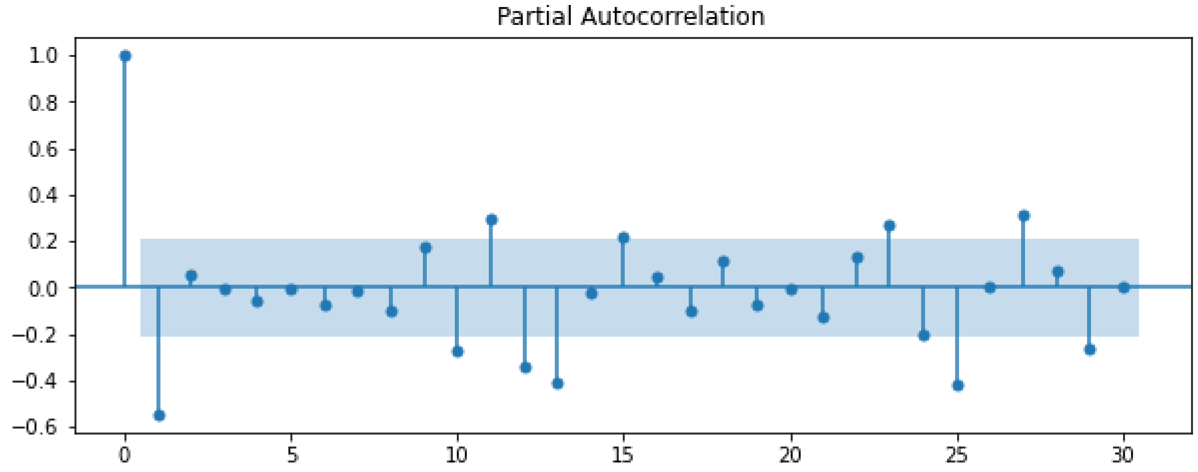
- We de-seasonalize the detrended data once since there is seasonality in the data and then run the stationarity test, and found the data to be stationary.
- Hence further plot the ACF and PACF plot for the same data.



ACF and PACF Plots:-

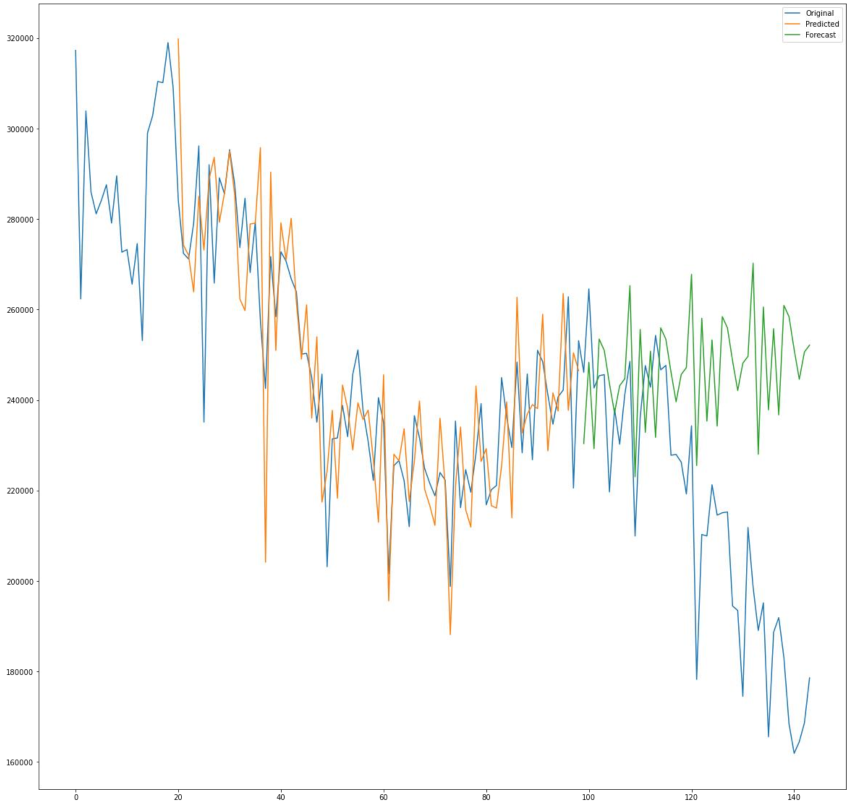
- Further in the ACF and PACF plot it is observed that
- (The ACF is gradual and PACF is sudden.)
- Hence it is AR(p) model with p = 1 from PACF plot .



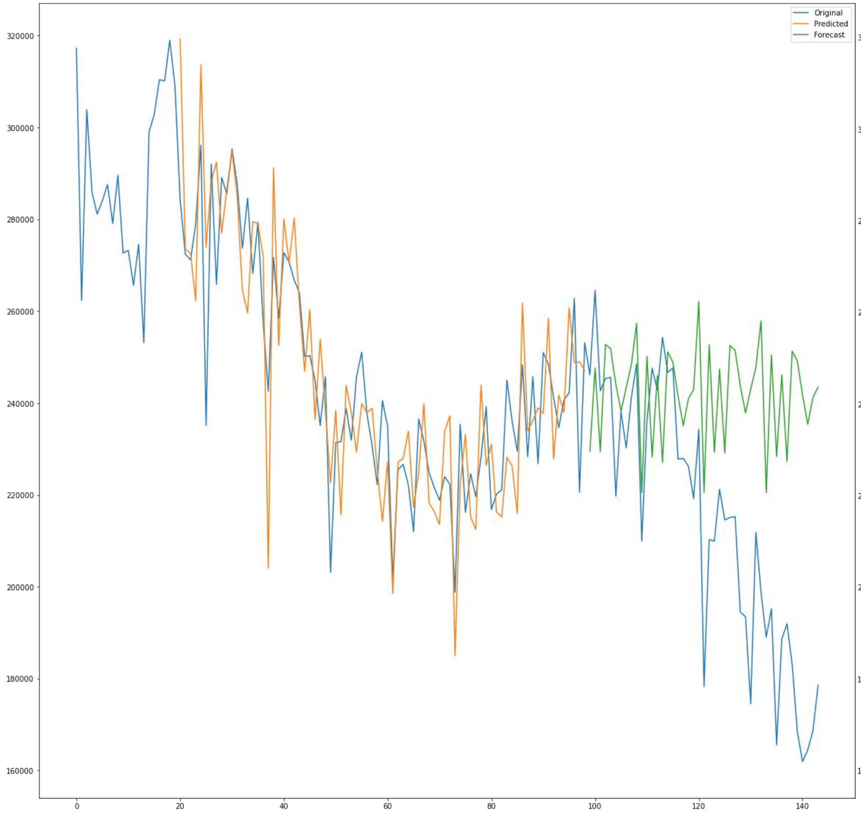


Results Of M1,M2 and M3 processes :-

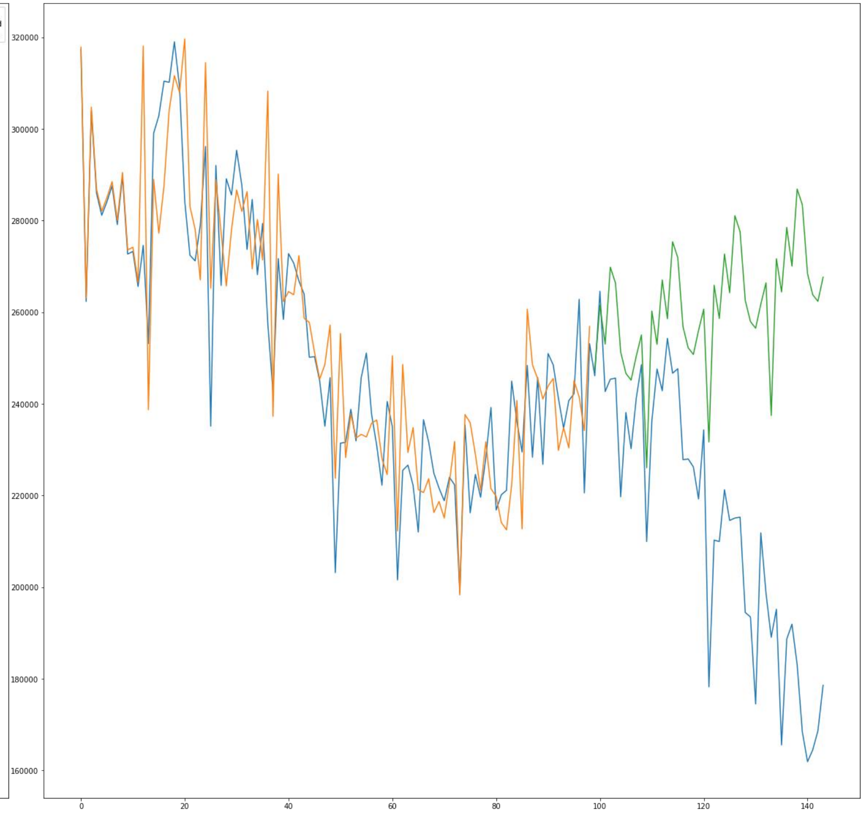
- In the below figure M1 we can see that for manual ARIMA the predicted values overlap on the original values more or less and the forecast for the test values is naive yet seems fine for a small interval of time and follows an upward trend which shows that the forecast is not correct and the model fails to give adequate results.
- Similarly in the figure M2 for Auto ARIMA as well the predicted values overlap on the original values, the forecast is naive and shows no trend whereas for the M3 i.e.(Holt winters Exponential Smoothing) the forecast again shows an upward trend for the test data.
- The overall observation from these three models M1,M2 & M3 is that they predict the values correctly for the train data but when it comes to forecast of the test data from the plot it is observed that for a very small interval they give results which look fine but thenthey fail to give adequate results.



M1



M2



M3

Model Analysis

- Based on the model parameters mentioned in the below table for M1,M2 and M3 process by observing the RMSE and AIC values for each model It is clear that the M2 process of Auto ARIMA gives the optimum parameters for the model and is the best model amongst the three models with RMSE values of 39719.881 and AIC value of 1890.671.
- Here the Holt Winters Exponential Smoothing model has the highest RMSE and AIC values which means it doesn't give optimum results for this model.

	Final Model Parameters							RMSE	AIC
	p	d	q	P	D	Q	Period		
M1	1	1	0	0	1	0	12	45113.227	1902.175
M2	1	1	0	1	1	1	12	39719.881	1890.671
M3 (Exponential Smoothing)	-	-	-	-	-	-	12	57311.833	1915.297

Conclusion

- Time Series Models that we have generated using the ARIMA process work well on the linear data that follow a regular seasonality and cyclicity whereas for the data with anomalies and irregular patterns they fail to give adequate results.
- The forecasts generated for such datasets are naive and not adequate, at times fail to give adequate results, which is why we need to use other techniques to solve these problems which could be using neural networks etc.
- These techniques of Arima modelling are good enough for learners to understand the working of time series models but when it comes to complex datasets they are not every time give better results.
- The holt winters process gives better results in some cases and is also easy to use and gives in depth results but again it has its limitations and doesn’t work well for the irregular data with anomalies.
- Overall the assignment was a good practice to understand the various techniques used in time series modelling and the outputs that can be expected through each technique.

THANK YOU