Project: Forecasting Sales

Step 1: Plan Your Analysis

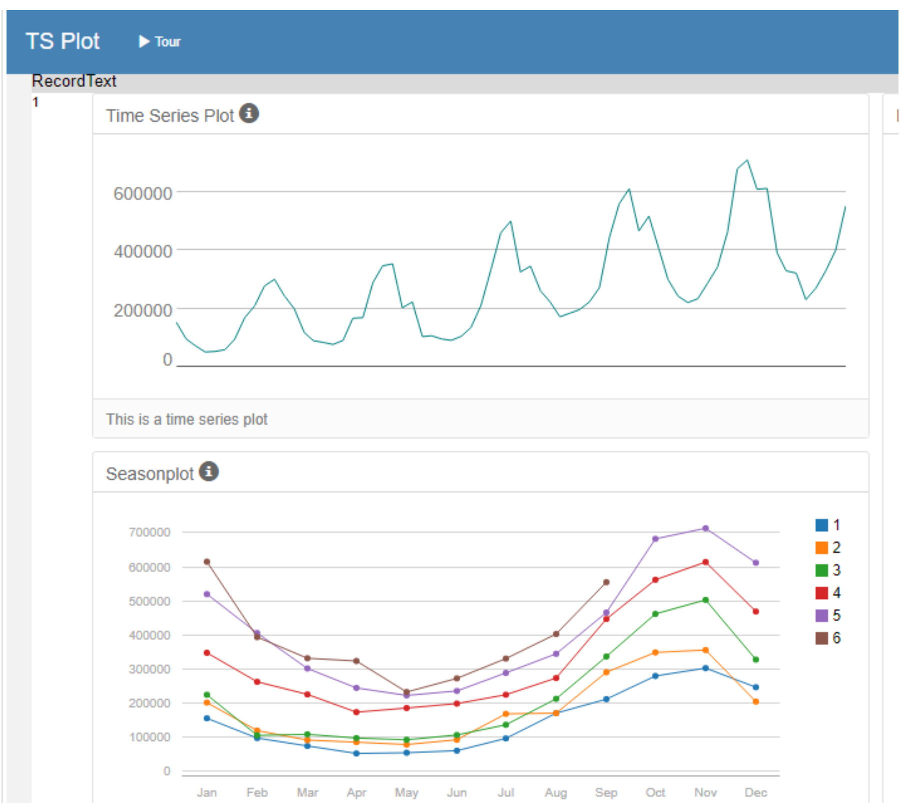
1. Does the dataset meet the criteria of a time series dataset? Make sure to explore all four key characteristics of a time series data.
   - ➔ Yes, the given dataset meets the criteria of a time series dataset., as :
     - The month column contains continuous data from Jan 2008 to Sep 2013.
     - Each time unit has at most one data point, i.e., monthly sales.
     - There is equal spacing of time intervals between two values, i.e., of one month.
     - Sequential measurements are available – in this case monthly sales are available.

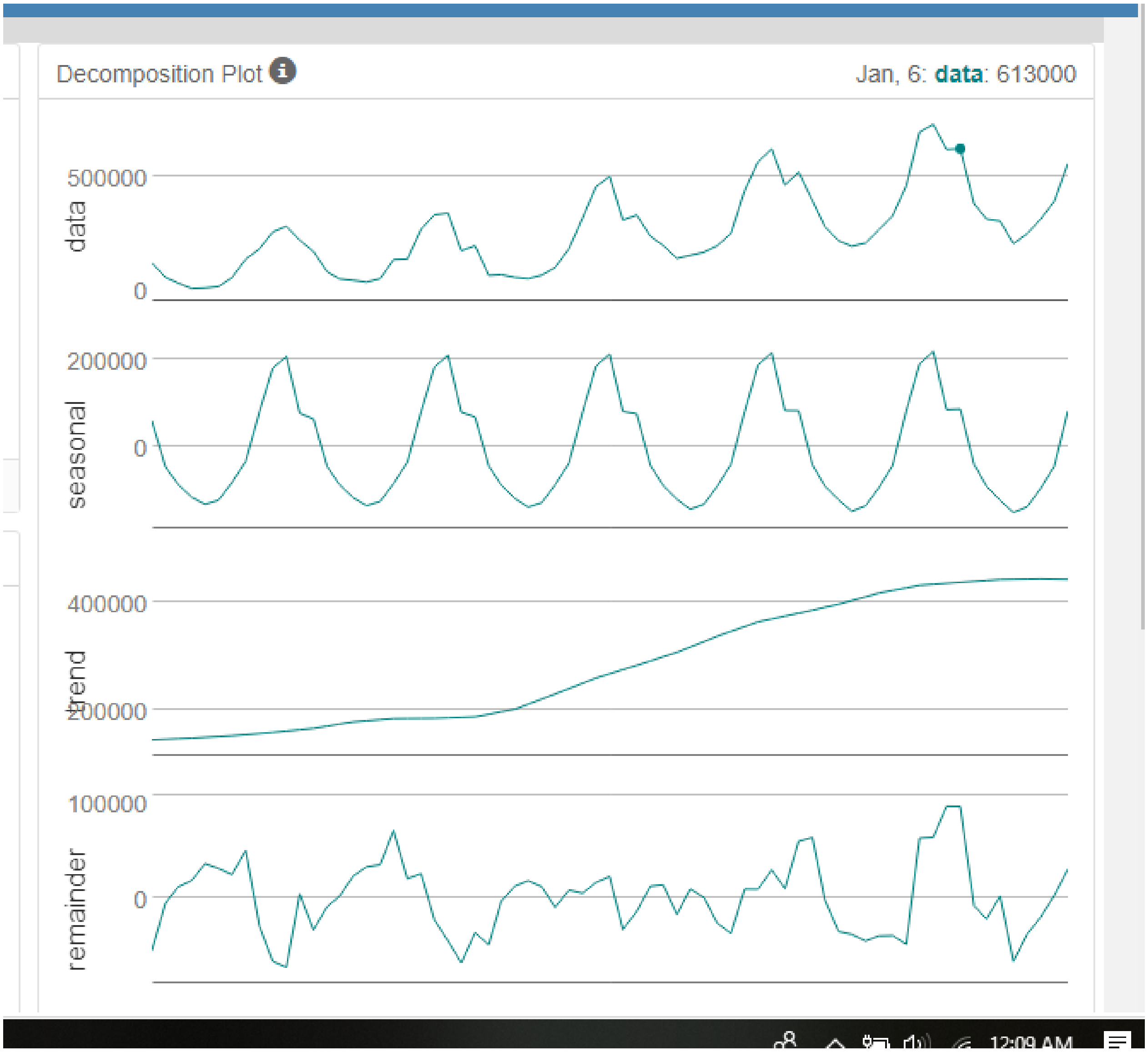
2. Which records should be used as the holdout sample?
   - ➔ For the holdout sample, we will use the recent four months data i.e., from June 2013 to September 2013,as we need to predict the next four months.

Step 2: Determine Trend, Seasonal, and Error components

Answer this question:

1. What are the trend, seasonality, and error of the time series? Show how you were able to determine the components using time series plots. Include the graphs.





➔ **From the above graphs , we can see that**
- Trend is linear and increasing with time.
- Seasonality exists with small amount of increase in its peaks.
- Error is varying with respect to time.

Step 3: Build your Models

1. What are the model terms for ETS? Explain why you chose those terms.
   ➔ As the trend is linear, I have used it as additive.
   ➔ As the seasonality exists, but the peaks are increasing even though at a small rate, have considered it as exponential.
   ➔ The error is not constant with respect to time , so it is also considered as exponential.

a. Describe the in-sample errors. Use at least RMSE and MASE when examining results

➔ **For the dampened ETS Model**

In-sample error measures:

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
5597.130809	33153.5267713	25194.3638912	0.1087234	10.3793021	0.3675478	0.0456277

➔ **Root Mean Squared Error (RMSE)** represents the sample standard deviation of the differences between predicted values and observed values.This is a great measurement to use when comparing models as it shows how many deviations from the mean the forecasted values fall.

- In this case its value is 33153.52

➔ **Mean Absolute Scaled Error (MASE)** is another relative measure of error that is applicable only to time series data. It is the mean absolute error of the model divided by the the mean absolute value of the first difference of the series. It measures the relative reduction in error compared to a naive model. Ideally its value will be significantly less than 1 but is relative to comparison across other models for the same series.

- In this case it is 0.36 which is less than 1 and acceptable.

➔ **For the non dampened ETS Model**

In-sample error measures:

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
2818.2731122	32992.7261011	25546.503798	-0.3778444	10.9094683	0.372685	0.0661496

➔ **Root Mean Squared Error (RMSE)**

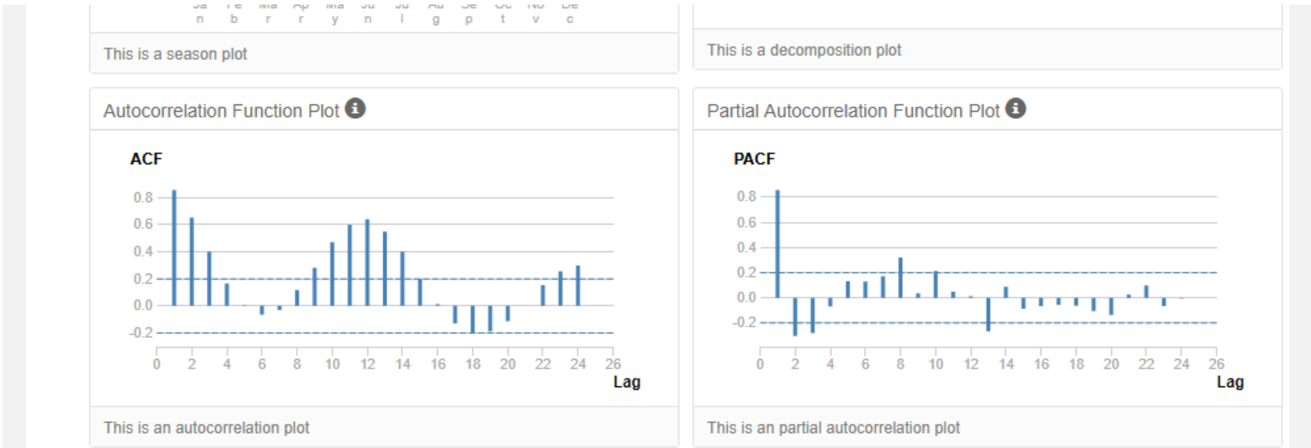
- In this case its value is 32992.72

➔ **Mean Absolute Scaled Error (MASE)**

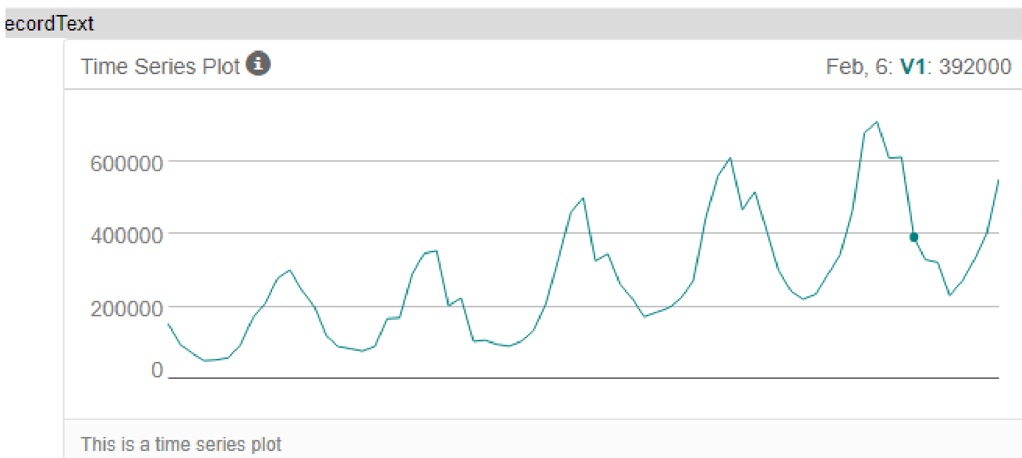
- In this case it is 0.37 which is less than 1 and acceptable.

2. What are the model terms for ARIMA? Explain why you chose those terms. Graph the Auto-Correlation Function (ACF) and Partial Autocorrelation Function Plots (PACF) for the time series and seasonal component and use these graphs to justify choosing your model terms.

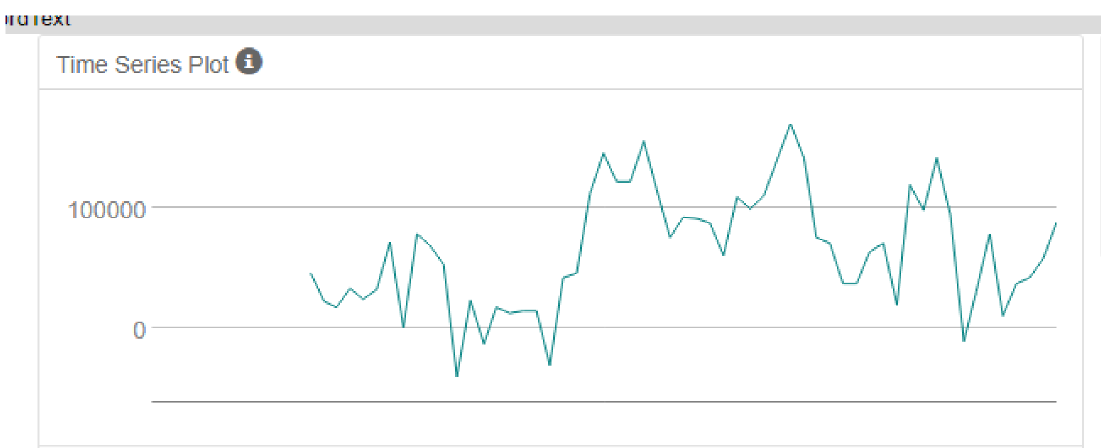
➔ The ACF and PACF graphs for the time series are as follows:-



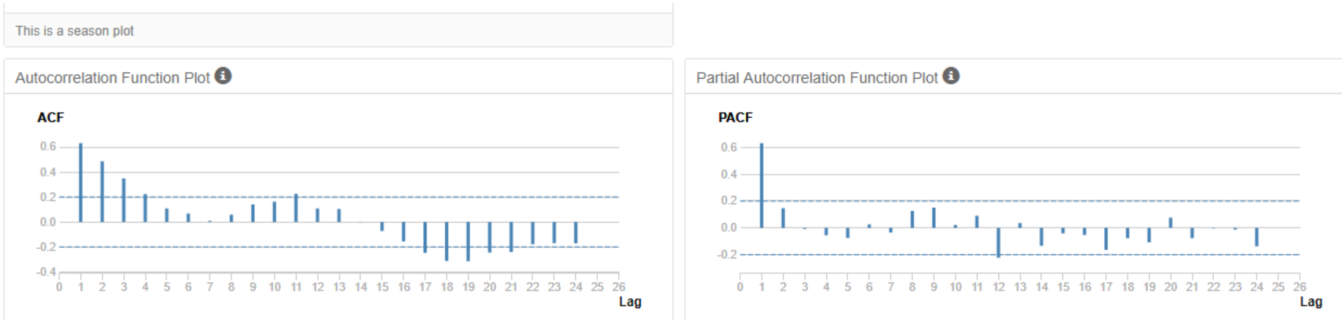
➔ As our time series is seasonal, we have to remove the seasonality.



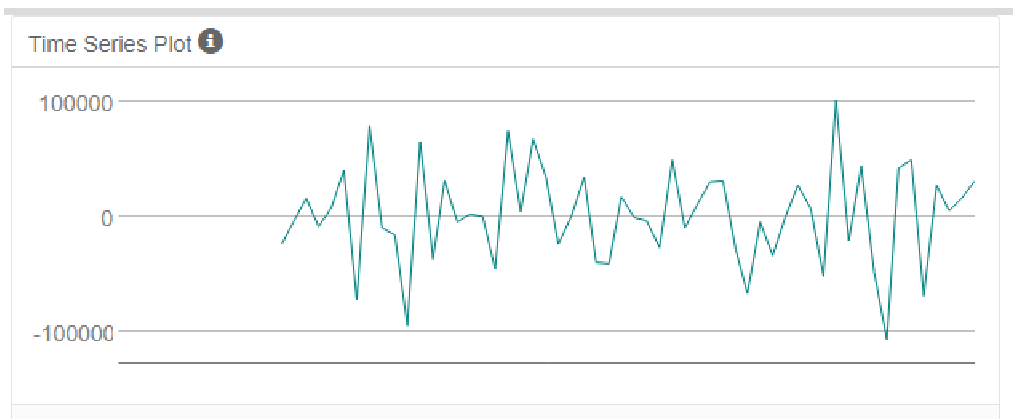
➔ **Seasonal differencing** :- After the seasonal differencing we can still see that the we still have correlation in the lags and all the lags are significant.



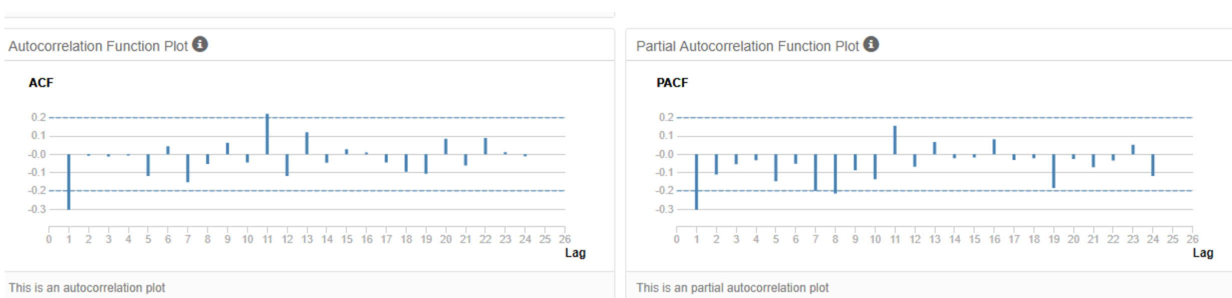
➔ The ACF and PACF graphs for the seasonal difference are as follows:-



➔ **Seasonal first differencing** :- All the significant lags are smoothened and this can be seen in the ACF and PACF plots, so there is no need for further differencing   We have D (1) and d (1).



➔ The ACF and PACF plots show that there is a strong negative correlation at lag 1, this suggests that the MA (1) for the non-seasonal component.



➔ ARIMA (p,d,q) (P,D,Q) m Non-seasonal (p,d,q) is (0,1,1) .

➔ As our seasonal ACF and PACF are not much significant, we will consider the Seasonal (P, D, Q) as (0,1,0) m is 12.

a. Describe the in-sample errors. Use at least RMSE and MASE when examining results.

sigma^2 estimated as 1722385234.94439: log likelihood = -626.29834

6 Information Criteria:

AIC	AICc	BIC
1256.5967	1256.8416	1260.4992

7 In-sample error measures:

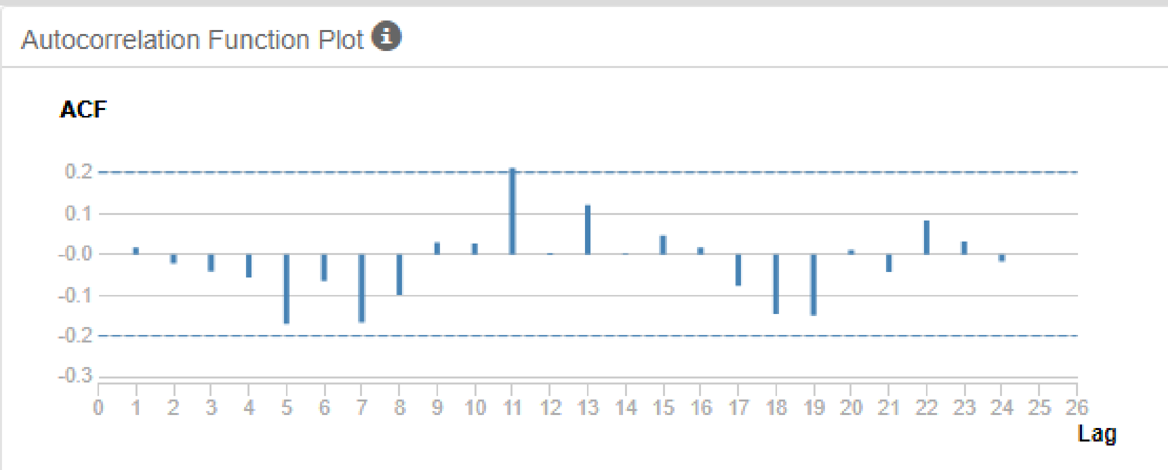
ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
-356.2665104	36761.5281724	24993.041976	-1.8021372	9.824411	0.3646109	0.0164145

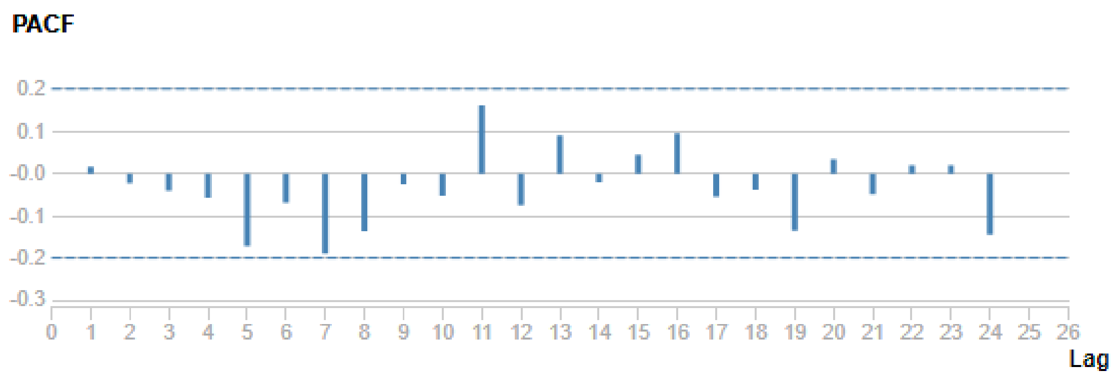
8

- ➔ The mean error is -356.266.
- ➔ The RMSE is 36761.52
- ➔ The MASE value is 0.3646 , which is less than 1 and is significant.

b. Regraph ACF and PACF for both the Time Series and Seasonal Difference and include these graphs in your answer.

- ➔ **The ACF and PACF graphs are redrawn by the ARIMA model as derived as follows:-**





Step 4: Forecast

Compare the in-sample error measurements to both models and compare error measurements for the holdout sample in your forecast. Choose the best fitting model and forecast the next four periods. (250 words limit)

Answer these questions.

1. Which model did you choose? Justify your answer by showing: in-sample error measurements and forecast error measurements against the holdout sample.

➔ Summary of ARIMA Model Arima_1

4 Coefficients:

	ma1
Value	-0.378032
Std Err	0.146228

5 sigma^2 estimated as 1722385234.94439: log likelihood = -626.29834

6 Information Criteria:

AIC	AICc	BIC
1256.5967	1256.8416	1260.4992

7 In-sample error measures:

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
-356.2665104	36761.5281724	24993.041976	-1.8021372	9.824411	0.3646109	0.0164145

8

➔ Summary of Time Series Exponential Smoothing Model ETS_Dampened

5 Method:
ETS(M,Ad,M)

6 In-sample error measures:

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
5597.130809	33153.5267713	25194.3638912	0.1087234	10.3793021	0.3675478	0.0456277

7 Information criteria:

AIC	AICc	BIC
1639.465	1654.3346	1678.604

8 Smoothing parameters:

Parameter	Value
alpha	0.870454
beta	0.000104
gamma	0.000112
phi	0.979992

➔ Summary of Time Series Exponential Smoothing Model ETS_non_dampened

5 Method:
ETS(M,A,M)

6 In-sample error measures:

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
2818.2731122	32992.7261011	25546.503798	-0.3778444	10.9094683	0.372685	0.0661496

7 Information criteria:

AIC	AICc	BIC
1639.7367	1652.7579	1676.7012

8 Smoothing parameters:

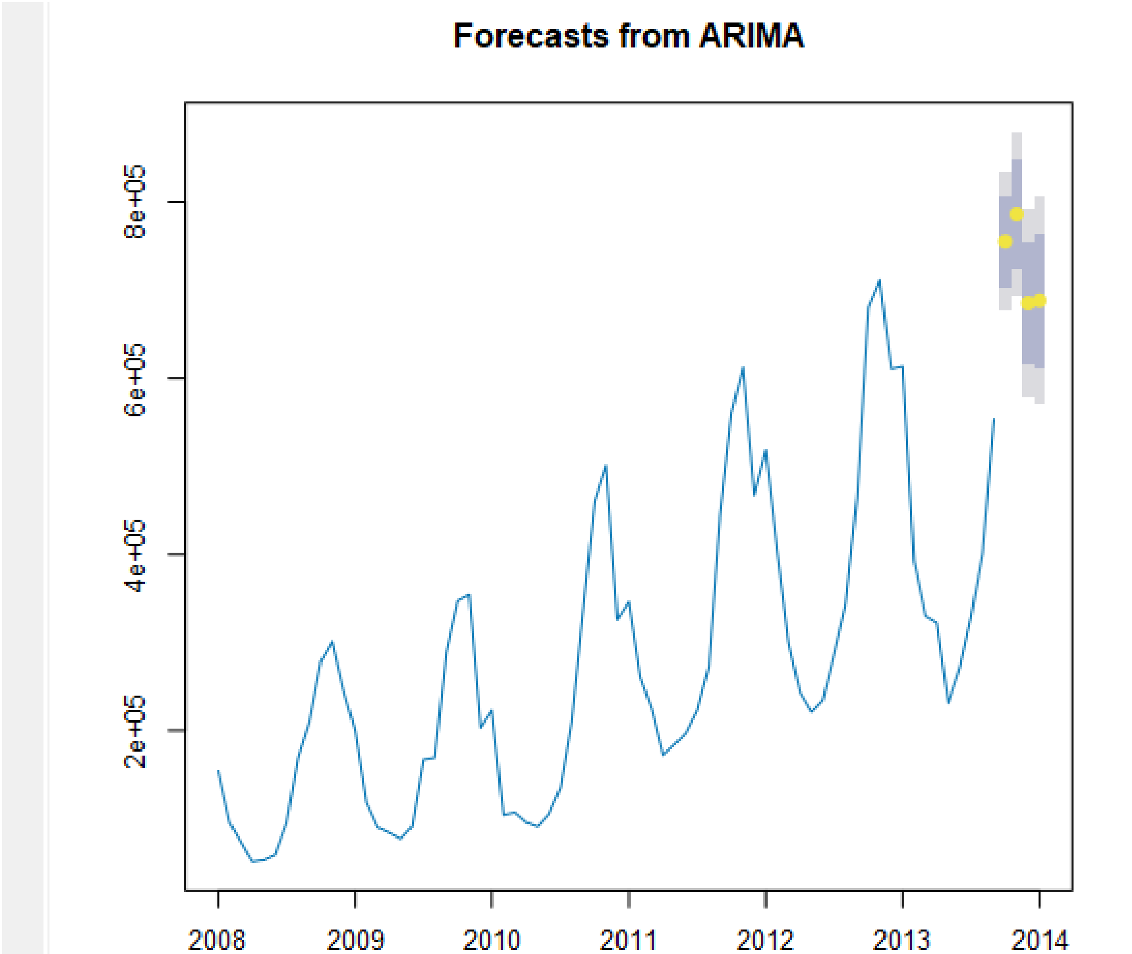
Parameter	Value
alpha	0.787787
beta	1e-04
gamma	0.000522

9

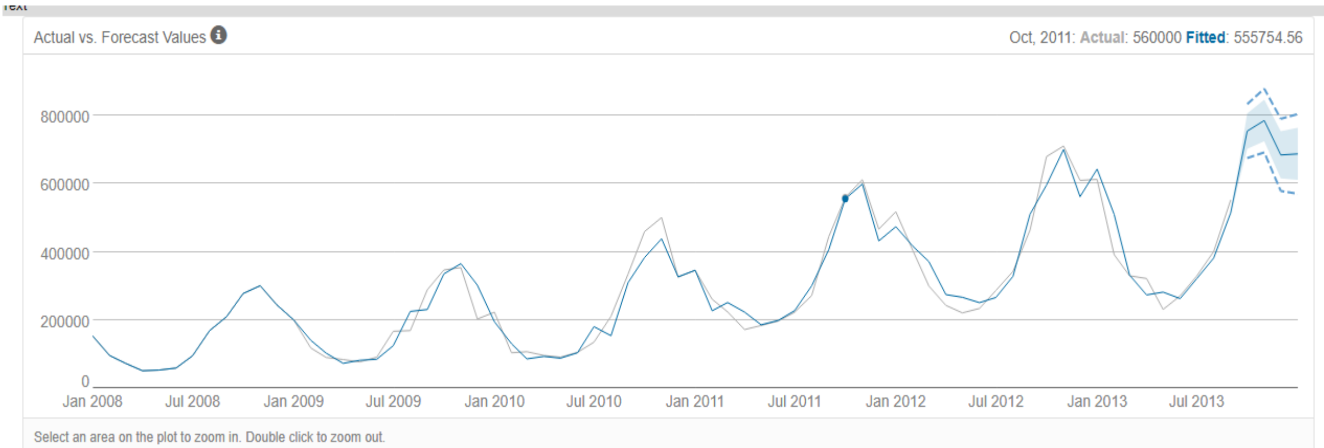
➔ From the above info , we can see that the ARIMA Model has less AIC , Mean Error(ME) and Mean Absolute Error(MAE).

➔ Therefore, ARIMA model will be used for forecasting the next four periods.

2. What is the forecast for the next four periods? Graph the results using 95% and 80% confidence intervals.
   ➔ The forecast of the next four months is as follows:-



Period	Sub_Period	forecast	forecast_high_95	forecast_high_80	forecast_low_80	forecast_low_95
2013	10	754854.460048	834046.21595	806635.165997	703073.754099	675662.704146
2013	11	785854.460048	879377.753117	847006.054462	724702.865635	692331.166979
2013	12	684854.460048	790787.828211	754120.566407	615588.35369	578921.091886
2014	1	687854.460048	804889.286634	764379.419903	611329.500193	570819.633462


Text
Actual vs. Forecast Values
Oct, 2011: Actual: 560000 Fitted: 555754.56
800000
600000
400000
200000
0
Jan 2008
Jul 2008
Jan 2009
Jul 2009
Jan 2010
Jul 2010
Jan 2011
Jul 2011
Jan 2012
Jul 2012
Jan 2013
Jul 2013
Select an area on the plot to zoom in. Double click to zoom out.