<u>Project: Forecasting Sales</u>

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.

<u>Answer</u>:
Yes, the dataset meets the criteria of a time series dataset.

a. It is over a continuous time interval from Jan'08 to Sept'13.
b. There are sequential measurements across the intervals.
c. There is equal spacing between every two consecutive measurements.
d. Each time unit within the time interval has at most one data point.

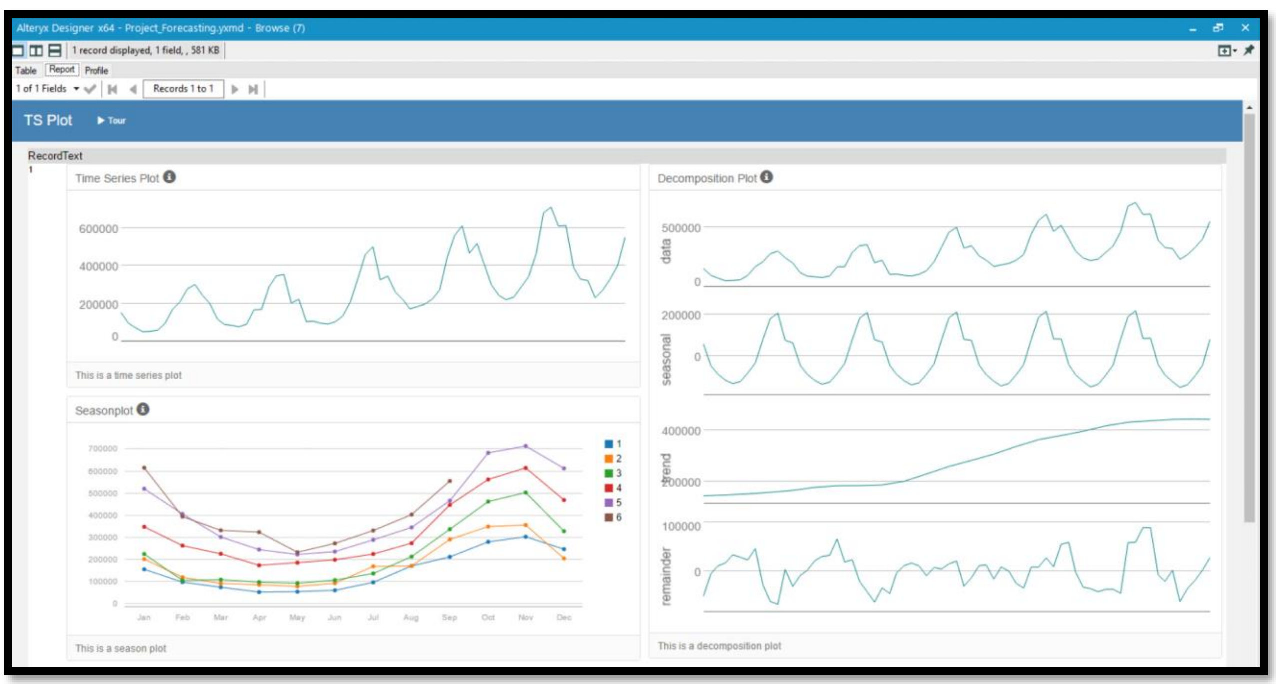
2. Which records should be used as the holdout sample?

<u>Answer</u>:
It is a good practice to holdout data of at least same period or more (most recent). Hence, I have used last 4 months of data in the holdout sample (June'13 – Sept'13).

Step 2: Determine Trend, Seasonal, and Error components

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.

<u>Answer</u>:



After plotting the data in the TS Plot tool, we can see the decomposition plot to understand the behavior of trend, seasonality and error.

It is evident from the above image that trend is behaving in a linear fashion, seasonality's magnitude is increasing with time interval and magnitude of error also increases with magnitude of seasonality.

This gives us the type of ETS model to be used for our forecasting. (ETS – M,A,M).

Step 3: Build your Models

1. What are the model terms for ETS? Explain why you chose those terms.

Answer:
Model terms for ETS are M,A,M. This is because:

Error: Increases with increase in magnitude of seasonality, so multiplicative.
Trend: Increases linearly (positive) , hence additive.
Seasonality: Magnitude increases with increase in lag, hence, multiplicative.

Also, I used dampening ETS model as it has low in-sample errors and lower AIC as compared to that of ETS model without dampening.

In-sample error measures:

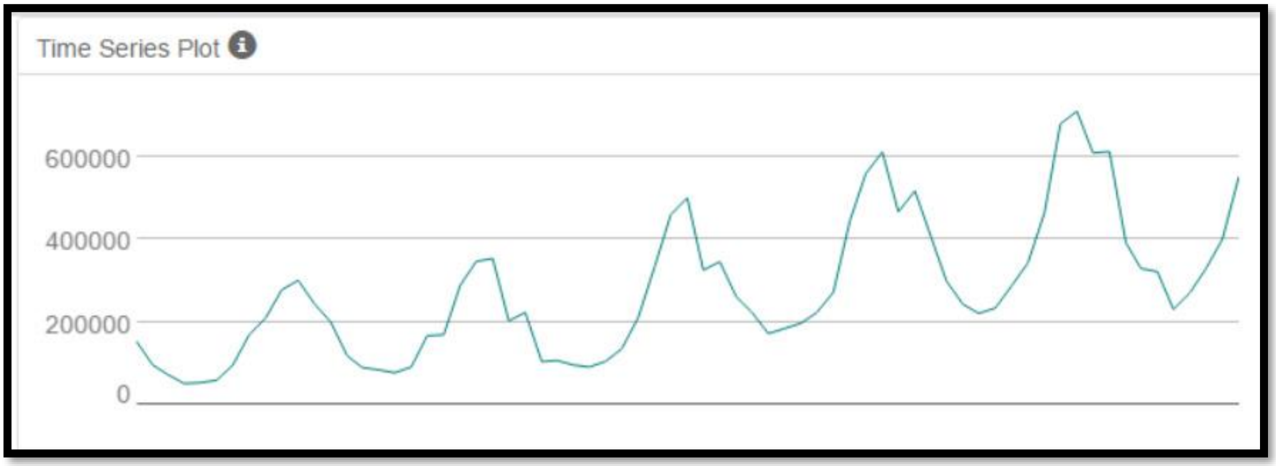
ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
5597.130809	33153.5267713	25194.3638912	0.1087234	10.3793021	0.3675478	0.0456277

RMSE is 33,153.52 for this model, which shows that there is a deviation of this much value of the forecasted value from the actual value. Also, MASE is 0.367, it doesn't cross the threshold of 1, which basically points that it is a good model.

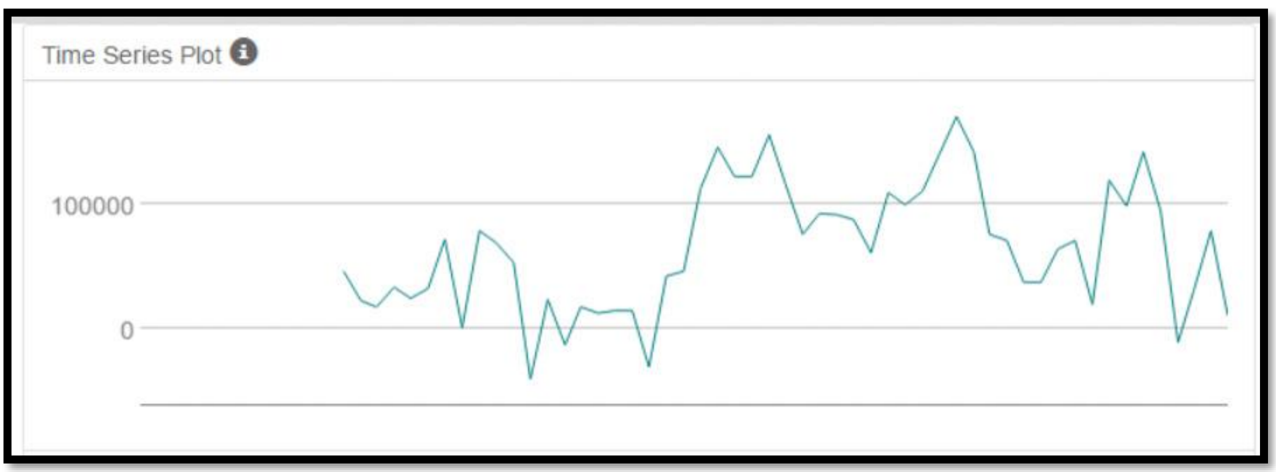
2. What are the model terms for ARIMA? Explain why you chose those terms. Graph the Auto-Correlation Function (ACF) and Partial Autocorrelation Function Plots (PACF) for the time series and seasonal component and use these graphs to justify choosing your model terms.

Answer:
Model terms for ARIMA Model are (0,1,1)(0,1,0)12. This is because:

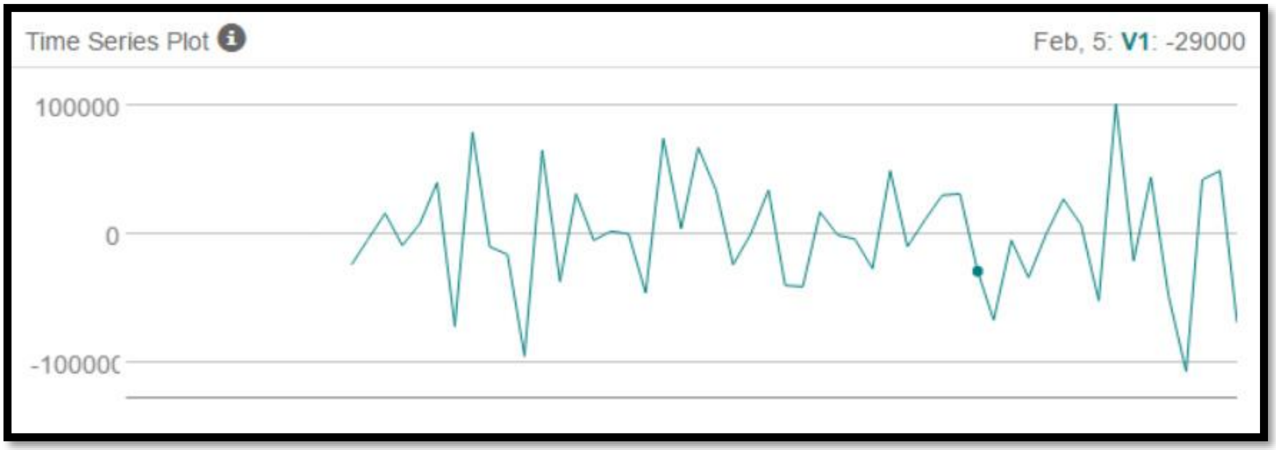
Indifference:



We can see that the Time series plot is not stationarised.
To stationarise data, we used ‘Multi-Row Formula’ tool to get a seasonal difference. After this we plot Time series again to see if the series is stationarised.



Again, the series is not stationarised.
So, we used ‘Multi-row Formula’ tool again to take first difference of seasonal difference. Then we get a stationarised series.



Hence, D=1 and d=1.

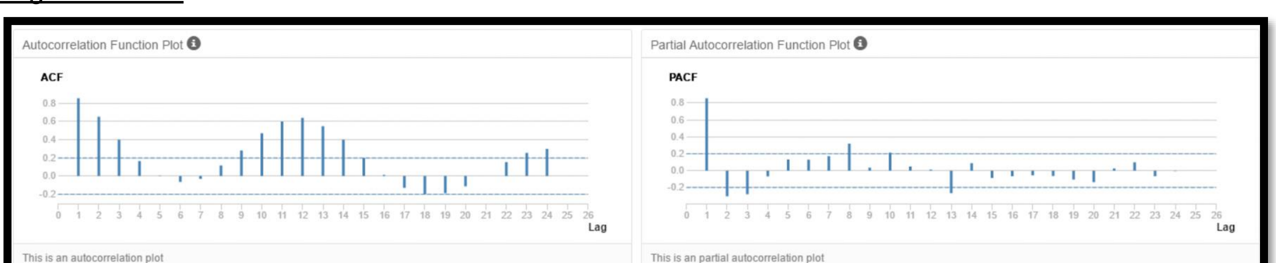
In-sample error measures:

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

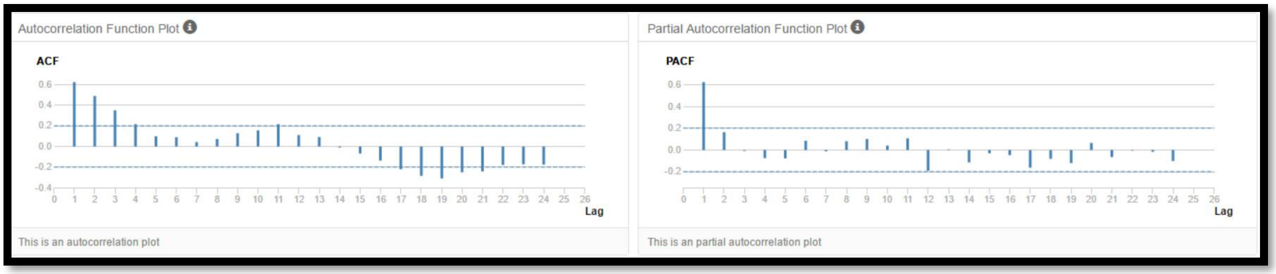
ME or Mean Average shows that the forecasted terms on an average are bigger than actual terms as it is negative.
RMSE is again showing standard deviation of forecasted values from actual values, in this case it is higher in comparison to ETS Model.
MASE is 0.36 and less than 1, hence implies a good model.
The MAE value of ARIMA Model is less than that of ETS Model, it implies that in terms of magnitude, there is less difference in terms on an average.

Auto Regression and Moving Average terms:
By looking at the stationarised ACF and PACF graphs we understand that there are no AR terms in the series.

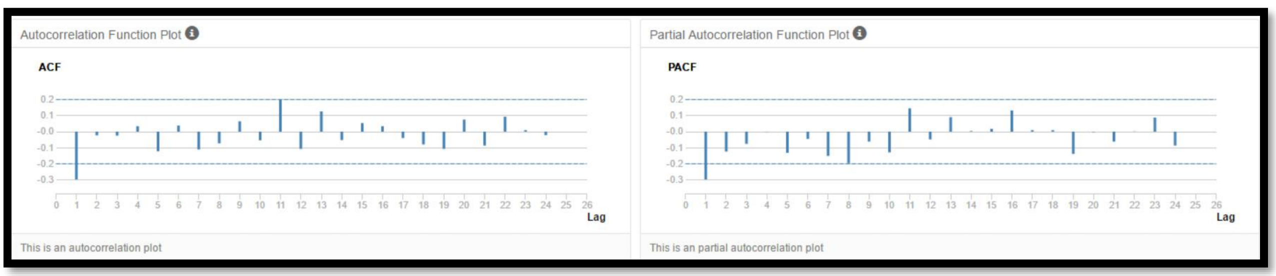
<u>Orginial series</u>



<u>Seasonal difference</u>

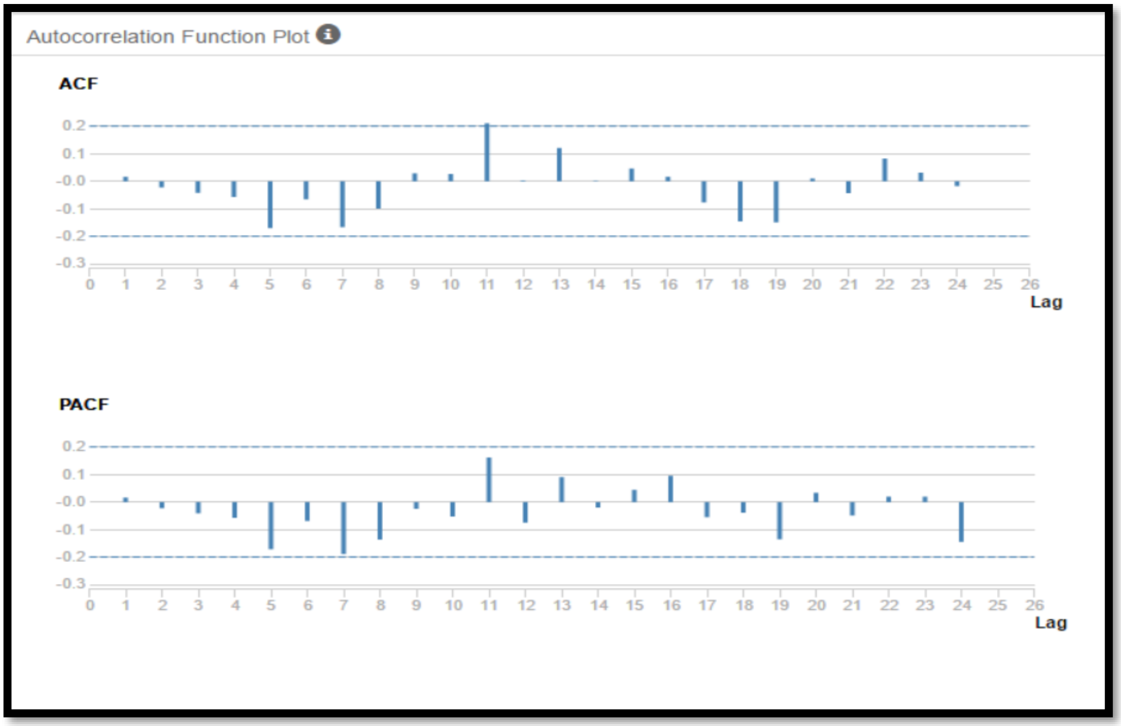


<u>First difference of seasonal difference or Stationarised series</u>



In the ACF plot, as the first lag is significant and negative we can say MA(1). Since, there is no seasonal significance (at lag 12, 24, etc) hence MA term for seasonal is 0.
In the PACF plot, we can see that it gradually falls to 0 again confirming MA terms in the model.

There are no AR terms in the model. Also, m = 12.
Hence, we conclude that ARIMA Model terms are **(0,1,1)(0,1,0)12**.
Also, the time series is stationarised after applying ARIMA Model. This can be established from the ACF and PACF plots below (next page).



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)

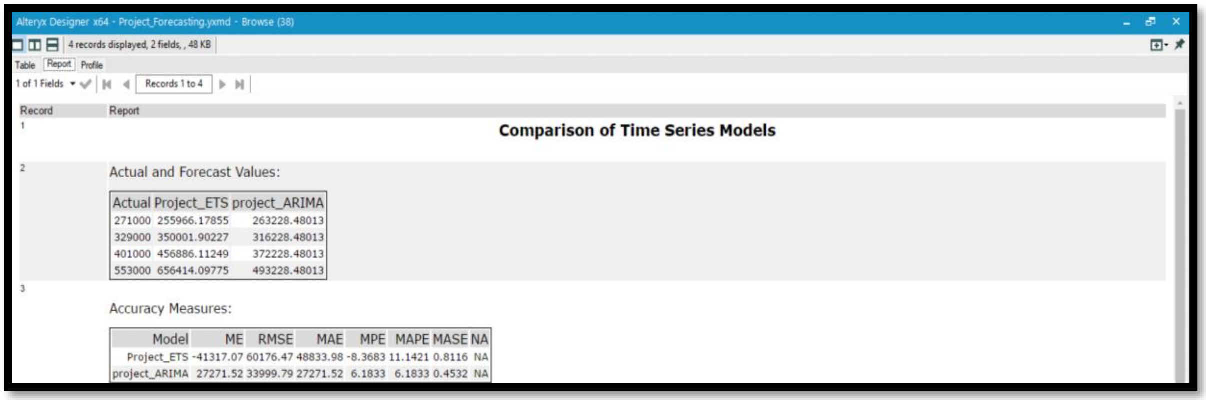
Answer these questions.

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

Answer:
I chose **ARIMA** Model for forecasting. I chose this model as it had low AIC value (**1256.5967**) as compared to ETS Model (with dampening) (AIC = **1639.465**).
Initially in ETS Model forecasting, I also compared ETS Models (with and without dampening), which showed that ETS Model with dampening performs better.

It was said in the lesson that AIC criterion has more priority in deciding better model than in-sample error measures, still I compared these errors to confirm my decision.

Looking at images below, it is quite evident that majority of ARIMA Model errors are less than that of ETS Model errors; MAE, MAPE, MASE. Hence, this confirms our decision to use ARIMA Model to forecast sales.



Alteryx Designer x64 - Project_Forecasting.yxmd - Browse (38)

4 records displayed, 2 fields, , 48 KB

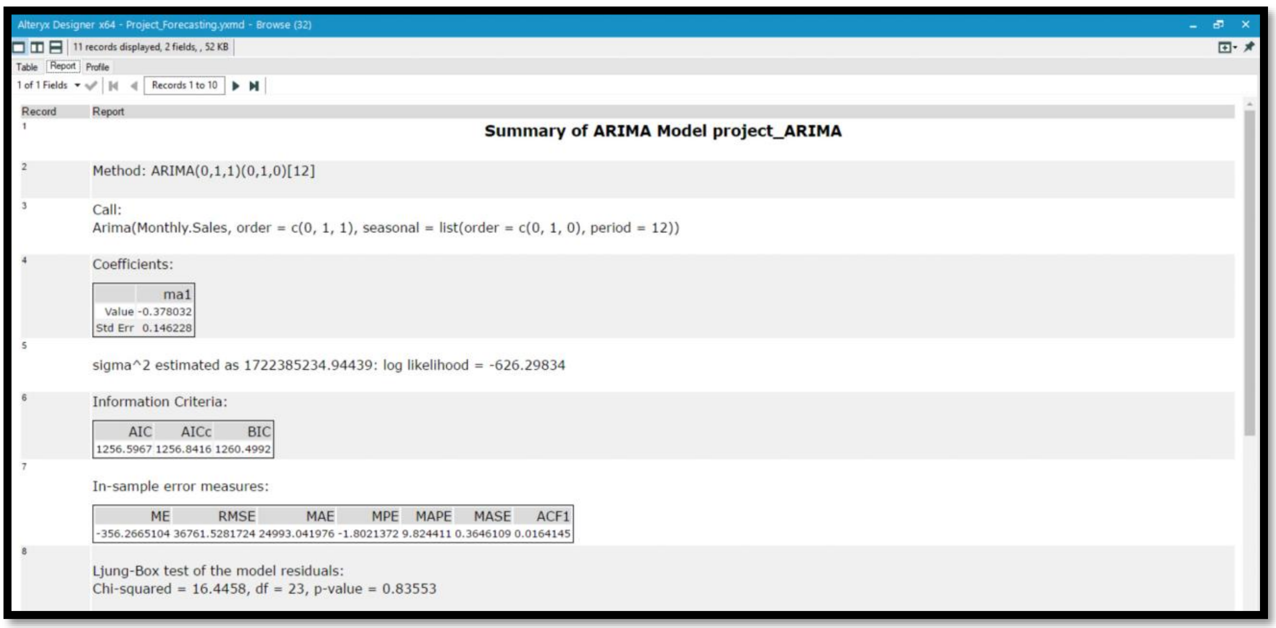
Comparison of Time Series Models

Actual and Forecast Values:

Actual	Project_ETS	project_ARIMA
271000	255966.17855	263228.48013
329000	350001.90227	316228.48013
401000	456886.11249	372228.48013
553000	656414.09775	493228.48013

Accuracy Measures:

Model	ME	RMSE	MAE	MPE	MAPE	MASE	NA
Project_ETS	-41317.07	60176.47	48833.98	-8.3683	11.1421	0.8116	NA
project_ARIMA	27271.52	33999.79	27271.52	6.1833	6.1833	0.4532	NA



Alteryx Designer x64 - Project_Forecasting.yxmd - Browse (32)

11 records displayed, 2 fields, , 52 KB

Summary of ARIMA Model project_ARIMA

Method: ARIMA(0,1,1)(0,1,0)[12]

Call:
Arima(Monthly.Sales, order = c(0, 1, 1), seasonal = list(order = c(0, 1, 0), period = 12))

Coefficients:

	ma1
Value	-0.378032
Std Err	0.146228

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

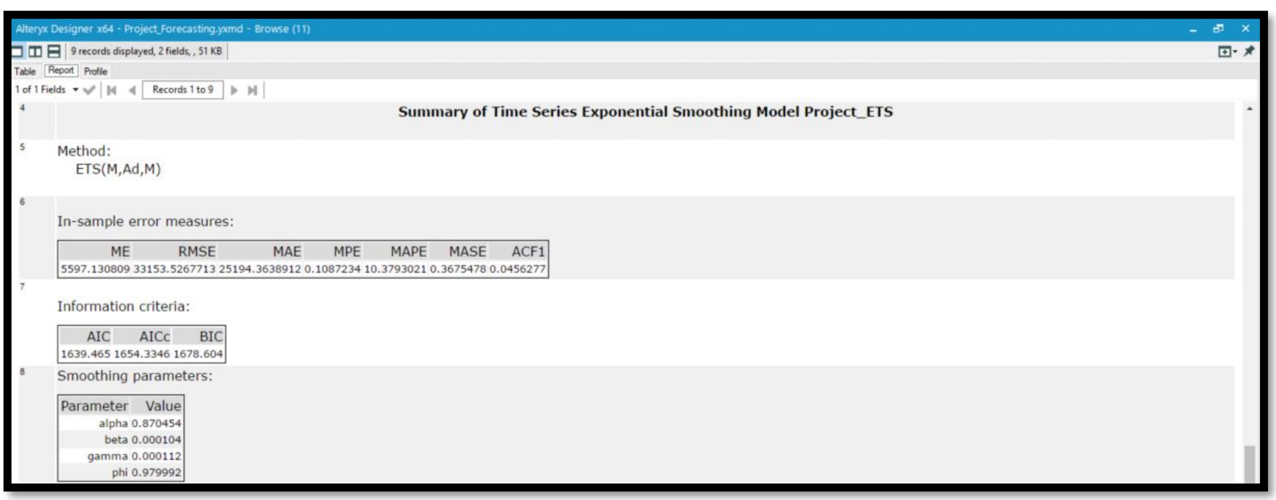
Information Criteria:

AIC	AICc	BIC
1256.5967	1256.8416	1260.4992

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

Ljung-Box test of the model residuals:
Chi-squared = 16.4458, df = 23, p-value = 0.83553



Alteryx Designer x64 - Project_Forecasting.yxmd - Browse (11)

9 records displayed, 2 fields, , 51 KB

Summary of Time Series Exponential Smoothing Model Project_ETS

Method:
ETS(M,Ad,M)

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

Information criteria:

AIC	AICc	BIC
1639.465	1654.3346	1678.604

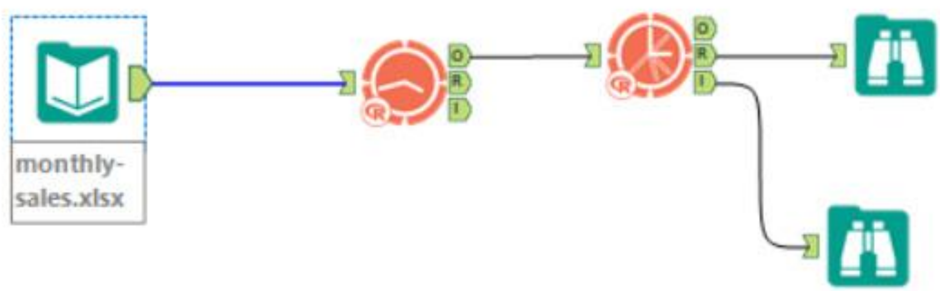
Smoothing parameters:

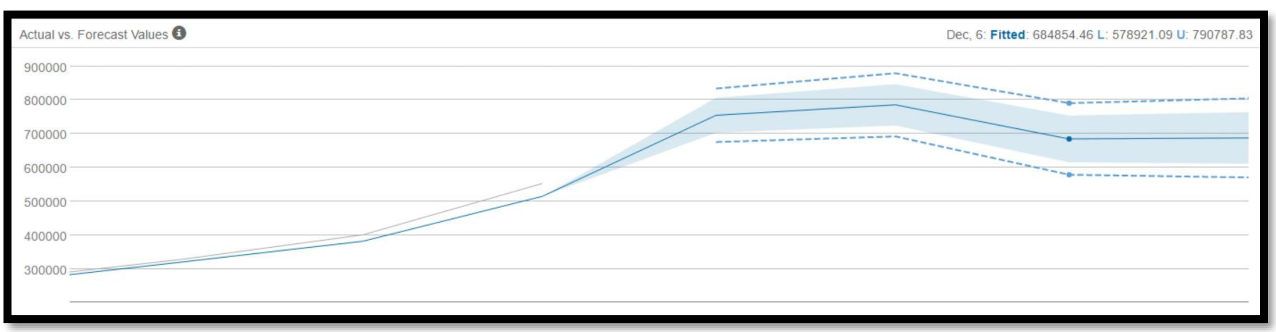
Parameter	Value
alpha	0.870454
beta	0.000104
gamma	0.000112
phi	0.979992

2. What is the forecast for the next four periods? Graph the results using 95% and 80% confidence intervals.

<u>Answer</u>:
To forecast for the next four periods, I used all the data (69 rows) in TS Forecast tool. Configured Forecast tool to forecast for next four months.

Months	**Sales**
October'13	754,854.46
November'13	785,854.46
December'13	684,854.46
January'14	687,854.46





Period	Sub_Period	forecast	forecast_high_95	forecast_high_80	forecast_low_80	forecast_low_95
6	10	754854.460048	834046.21595	806635.165997	703073.754099	675662.704146
6	11	785854.460048	879377.753117	847006.054462	724702.865635	692331.166979
6	12	684854.460048	790787.828211	754120.566407	615588.35369	578921.091886
7	1	687854.460048	804889.286634	764379.419903	611329.500193	570819.633462