Project: Forecasting Sales

By: Alhanoof Alyabes

Date 28/4/2018

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.

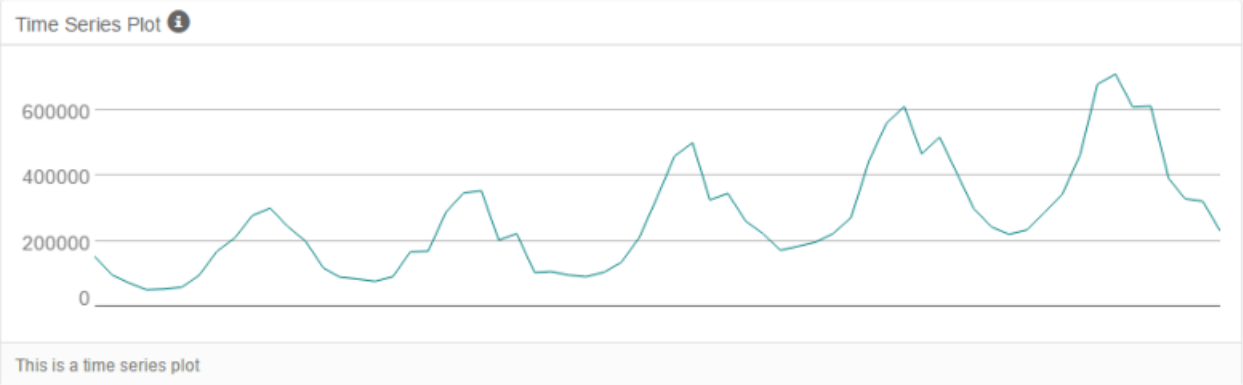
This data set meet the criteria of time series dataset. Where the ordering is matters in time series, this dataset has continued time interval, The measurement of data taken across sequential and equal spacing intervals, there is at most one data point for each time unit within the time interval.

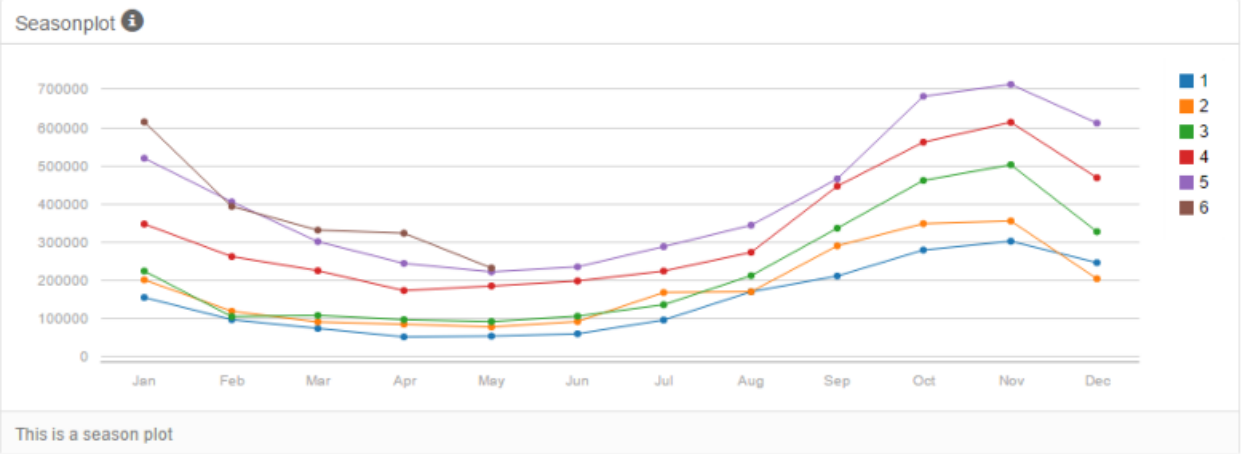
2. Which records should be used as the holdout sample?

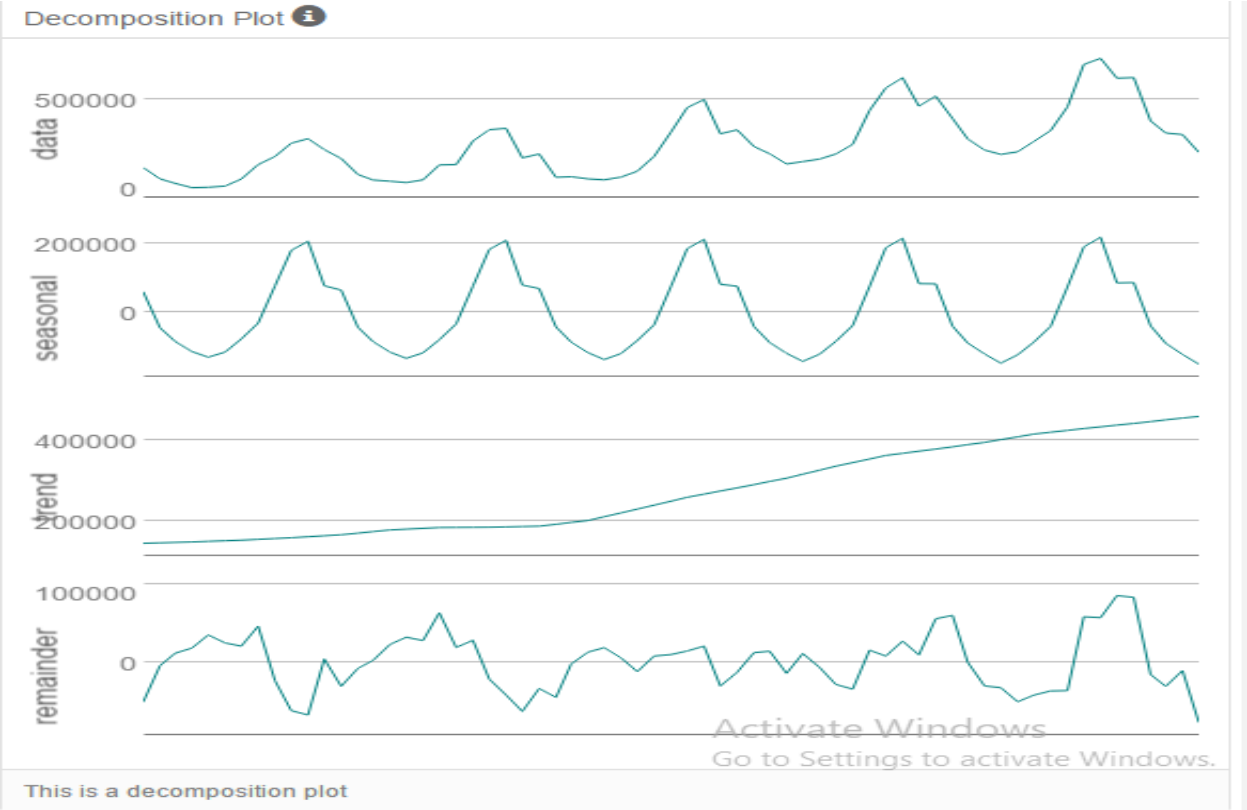
It is depending on the data that we have and the how far we need to forecast. Since we need to forecast monthly sales data in order to synchronize supply with demand, and we need to predict the next four months, the last four month will be used from 2013-6 to 2013-9 as holdout sample.

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.







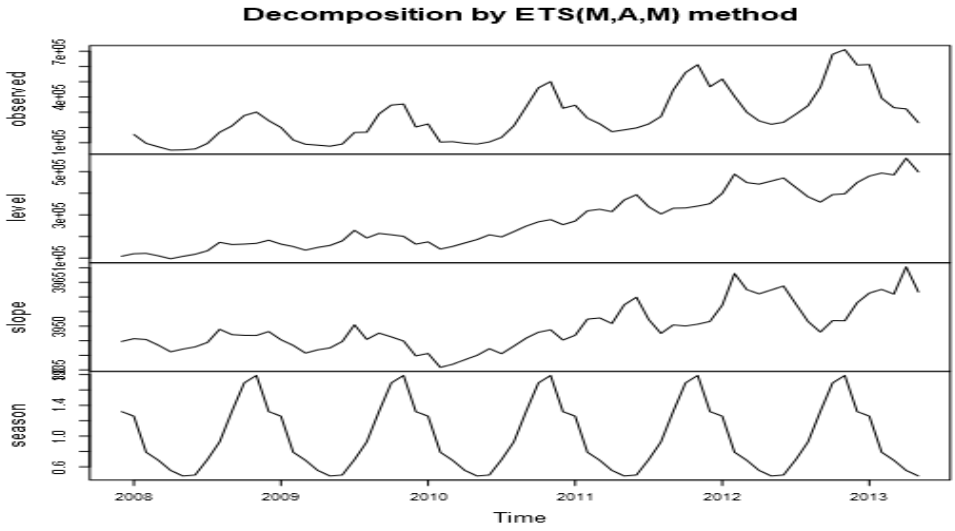
season and series plot are generated by TS plot.at decomposition plot, trend moves in leaner fashion which suggesting applying it additively while seasonal portion looks be about the same every year and increase in volume in each seasonal period, so it applied in multiplicative manner. Error are more fluctuations so thus it is applied multiplicatively as well.

Step 3: Build your Models

1. What are the model terms for ETS? Explain why you chose those terms.
    a. Describe the in-sample errors. Use at least RMSE and MASE when examining results

As shown in decomposition above, we choose ETS (M, A, M), now we will compare dampened and non-dampened ETS model with holdout sample 4 months.

Non-dampened ETS



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

Information criteria:

AIC	AICc	BIC
1639.7367	1652.7579	1676.7012

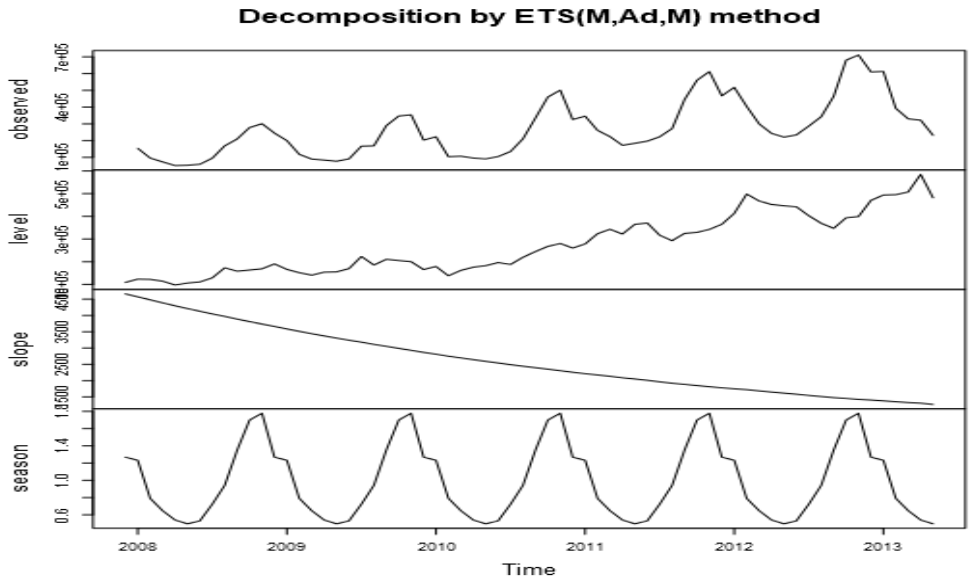
Actual and Forecast Values:

Actual	MAM
271000	248063.01908
329000	351306.93837
401000	471888.58168
553000	679154.7895

Accuracy Measures:

Model	ME	RMSE	MAE	MPE	MAPE	MASE	NA
MAM	-49103.33	74101.16	60571.82	-9.7018	13.9337	1.0066	NA

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

Information criteria:

AIC	AICc	BIC
1639.465	1654.3346	1678.604

Actual and Forecast Values:

Actual	X_MAM_damp
271000	255966.17855
329000	350001.90227
401000	456886.11249
553000	656414.09775

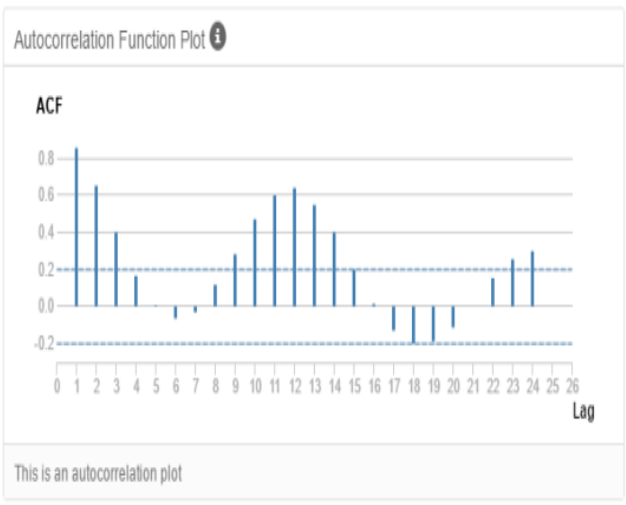
Accuracy Measures:

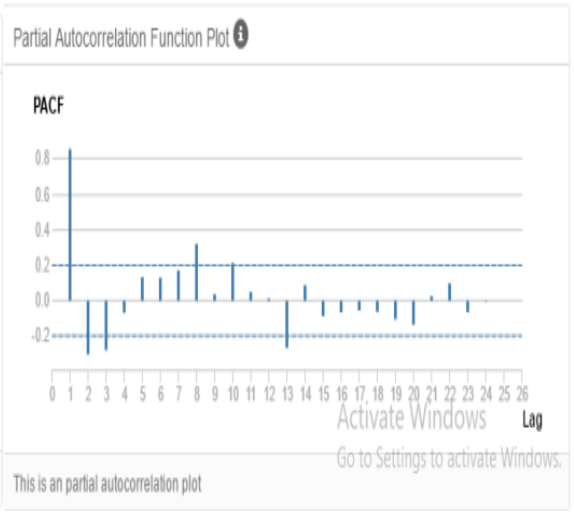
	Model	ME	RMSE	MAE	MPE	MAPE	MASE	NA
X_MAM_damp		-41317.07	60176.47	48833.98	-8.3683	11.1421	0.8116	NA

when compare the two models, we can find that RMSE for non-dampened is **32992** which is lower than dampened **33153.** while the MASE for non-dampened .**37** and dampened.**36**.the AIC for non-dampened is higher than dampened.
when compare each model by forecast and actual result, we noticed that dampened model is higher accuracy than non-dampened, so it will be chosen.

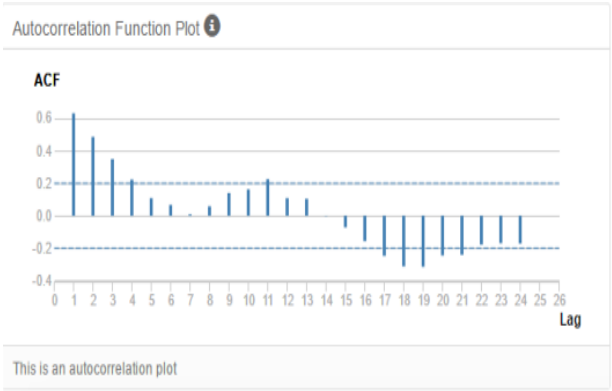
2. What are the model terms for ARIMA? Explain why you chose those terms. Graph the Auto-Correlation Function (ACF) and Partial Autocorrelation Function Plots (PACF) for the time series and seasonal component and use these graphs to justify choosing your model terms.
    a. Describe the in-sample errors. Use at least RMSE and MASE when examining results
    b. Regraph ACF and PACF for both the Time Series and Seasonal Difference and include these graphs in your answer.

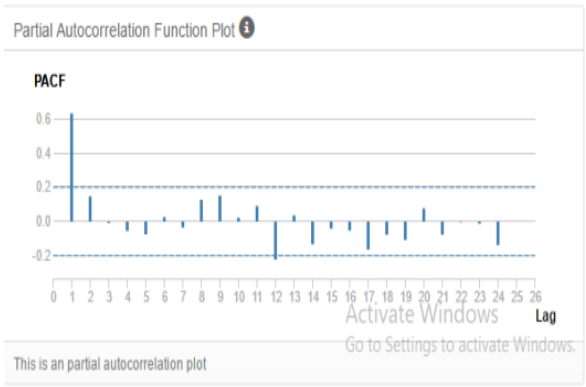
For ARIMA, the series need to be stationary. here, without differencing, the ACF shows high correlation and PACF has seasonal lag at 13.



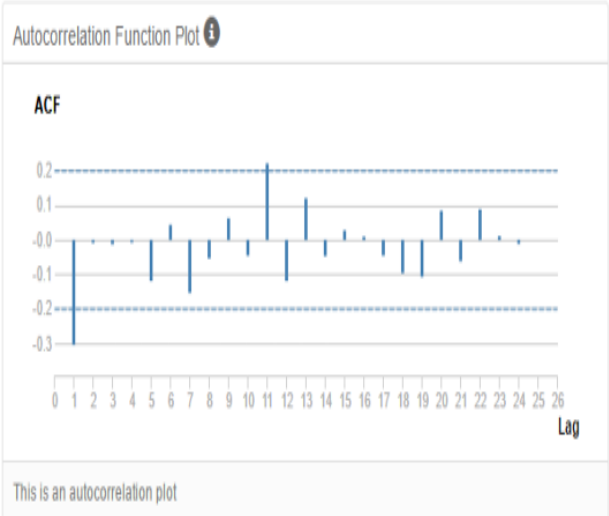


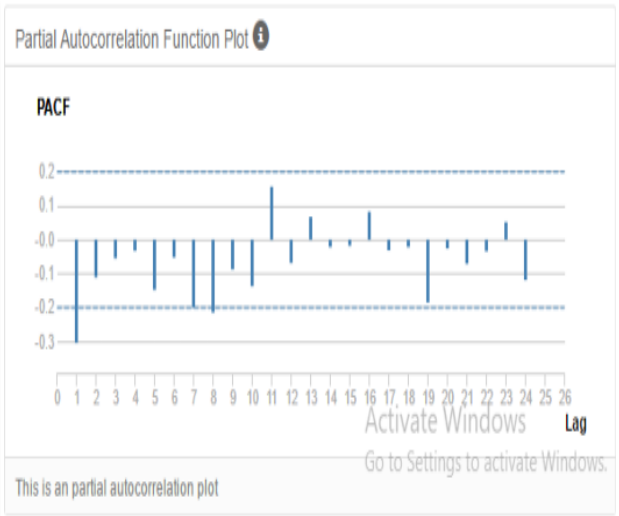
the non-stationary can be corrected by taking the differencing. seasonal differences taken, ACF still have high correlation, while PACF does not have strong correlation





first difference for time series plot blew has been stationaries.





we can observe here that seasonal first difference, there is no peaks accrued at 12, 24 lag, while the non-seasonal shows that spike lag 1 which is referred to non-seasonal MA term.

Therefore, the ARIMA model fit (0,1,1) (0,1,0) 12

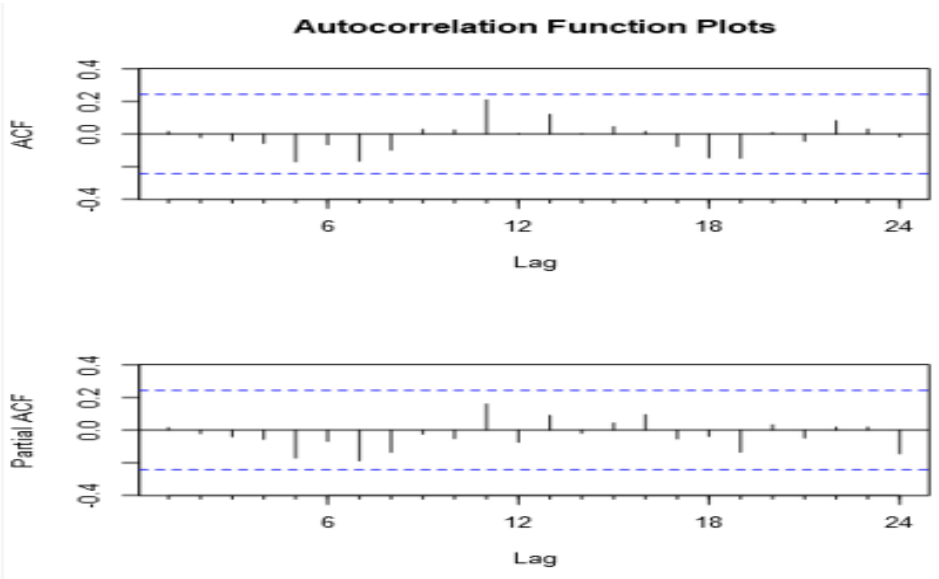
Information Criteria:

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

as shown above, the RMSE **36761**, MASE .**36** and the AIC **1256**, which is tower than ETS model.



there is no significant correlation for ACF and PACF, and no need for more AR or MA terms.

Step 4: Forecast

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

In sample error

ETS

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

ARIMA

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

at in sample error, ETS model RMSE **33153.52** and MASE **.367**, while ARIMA model RMSE **46761.52** and MASE **.364**. the RMSE of ETS model lower than ARIMA.

Forecast error

ETS

Actual and Forecast Values:

Actual	X_MAM_damp
271000	255966.17855
329000	350001.90227
401000	456886.11249
553000	656414.09775

Accuracy Measures:

Model	ME	RMSE	MAE	MPE	MAPE	MASE	NA
X_MAM_damp	-41317.07	60176.47	48833.98	-8.3683	11.1421	0.8116	NA

ARIMA

Actual and Forecast Values:

Actual	damp
271000	263228.48013
329000	316228.48013
401000	372228.48013
553000	493228.48013

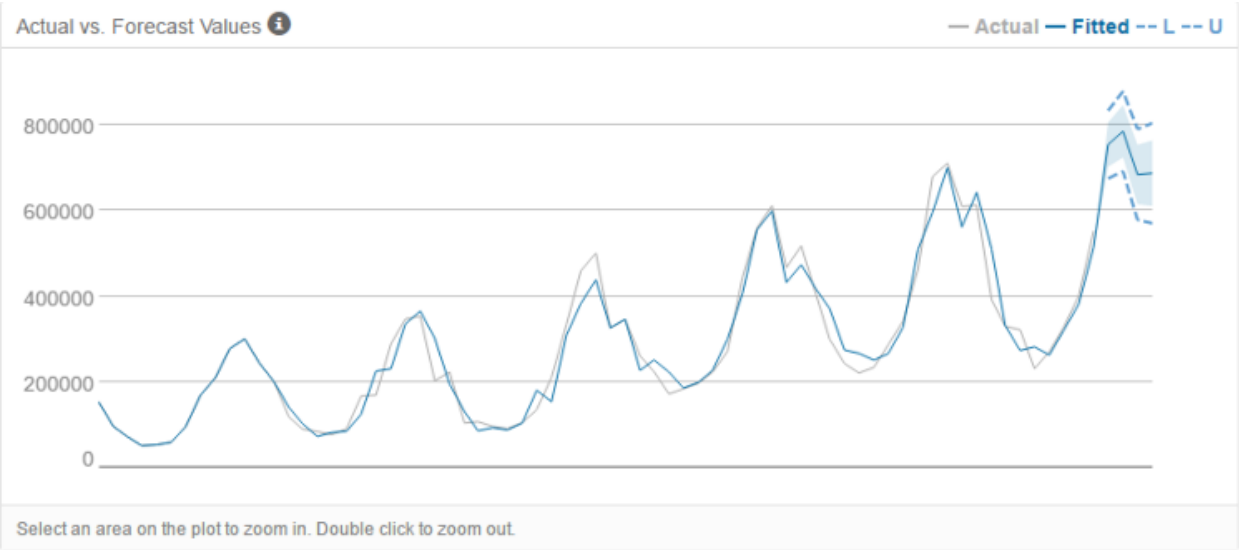
Accuracy Measures:

Model	ME	RMSE	MAE	MPE	MAPE	MASE	NA
damp	27271.52	33999.79	27271.52	6.1833	6.1833	0.4532	NA

at the forecast measurement, we consider that ARIMA better using holdout sample as validation data. The ARIMA model RMSE **33999.79** MASE **.4532** lower than ETS model which have RMSE **60167.47** and MASE **.8116**. Then the ARIMA chosen because the in-sample error and forecast error are smaller than ETS Model.

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

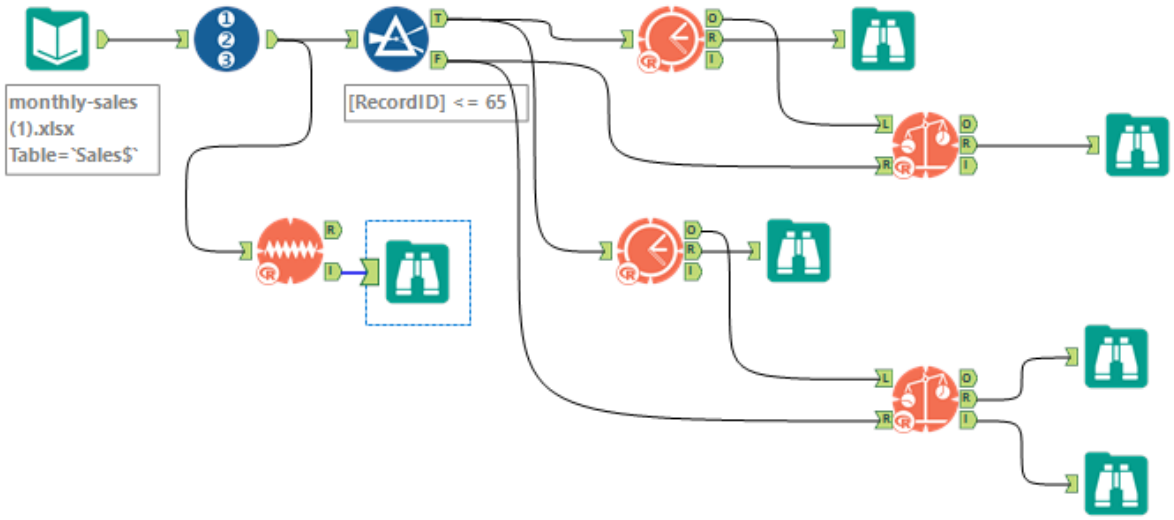
By using TS forecast to predict the next four period which is from 2013/10 to 2014/1 shown in plot blew



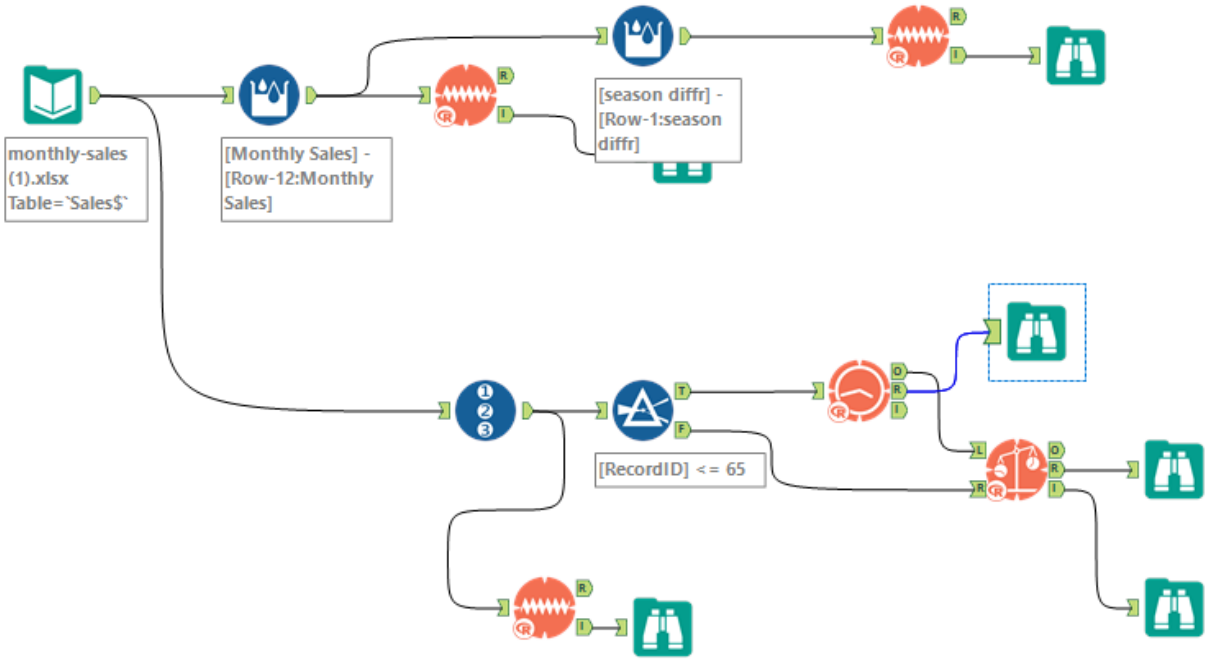
the forecast for four period are **754854 ,785854, 684845** and **687854**

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

Alteryx workflow



ETS model



ARIMA model