Project: Forecasting Sales

Step 1: Plan Your Analysis

Answer the following questions to help you plan out 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.
   The dataset is in continuous time interval which conclude that dataset meet criteria, as well as each measurement of dataset is sequential with an even interval. Each time unit of dataset has most one data points. The ordering of measurement matters & it is dependent on time.
2. Which records should be used as the holdout sample?
   As we are predicting sales for upcoming 4 months, hence holdout sample should be at least 4-month worth of data. So the last 4 data points must be separated from whole dataset.

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.

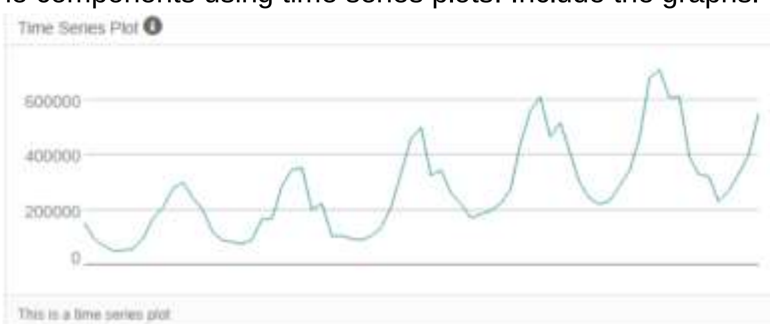


Fig. 2.1 Time series plot

Fig. 2.2 Season plot

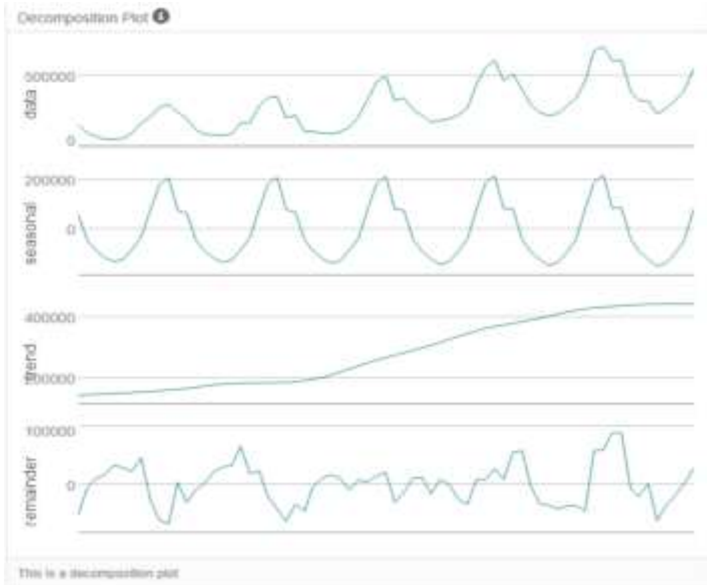


Fig. 2.3 Decomposition plot

As per decomposition plot in figure 2.3, we can state below conclusions

- Seasonal pot we can see a slight increase in magnitude every month which depicts that use of multiplicative for this component.
- Trend plot seems linear hence additive should be used for this component.
- Remainder plot shows variation over time which depicts the error should be used multiplicatively.

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

       The terms for both dampened and non-dampened ETS models are multiplicative, multiplicative and addictively for error, season, and trend respectively. The effect from dampening tested and each module is evaluated for its accuracy by fitting hold-out sample from last 4 months. The notation is hereby ETS(M.A.M). Then the results compare as follows.

Fig. 3.1 Workflow for dampened ETS and non-dampened ETS model.

Dampened ETS Model

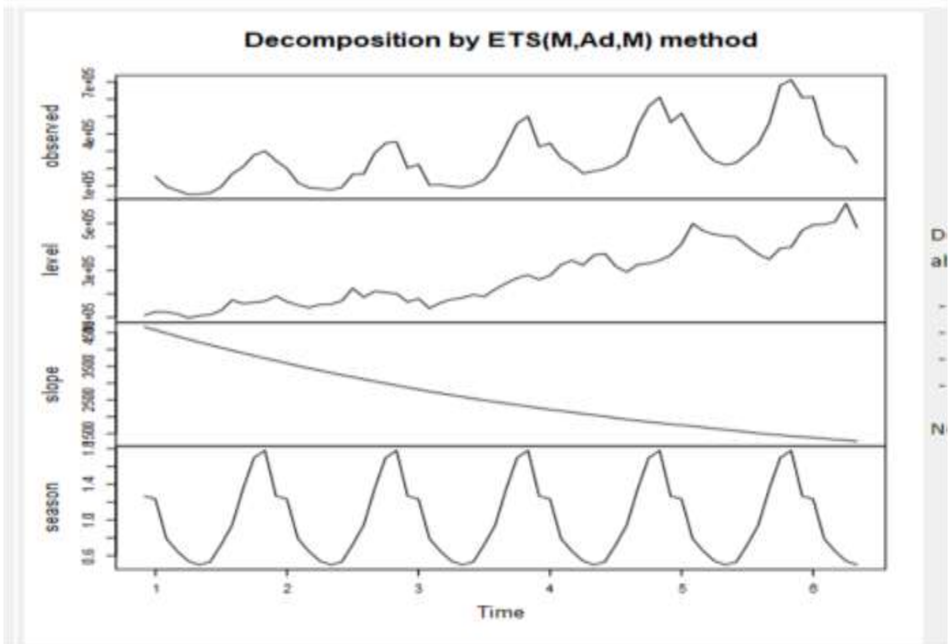


Fig. 3.2 Decompostion plot by ETS dampened model

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

Fig. 3.3 in-sample error measures and information criteria reports for dampened ETS

Actual and Forecast Values:

Actual	ETS_dampening
271000	255966.17855
329000	350001.90227
401000	456886.11249
553000	656414.09775

Accuracy Measures:

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

Fig 3.4 Upper, Actual vs forecast values using hold-out sample for evaluating the accuracy of dampened ETS model. Bottom, Accuracy measures for dampened ETS model in terms of mean error, root mean squared error, absolute mean error, average percentage error, average absolute percentage error and mean absolute scaled error.

Non-dampened ETS Model

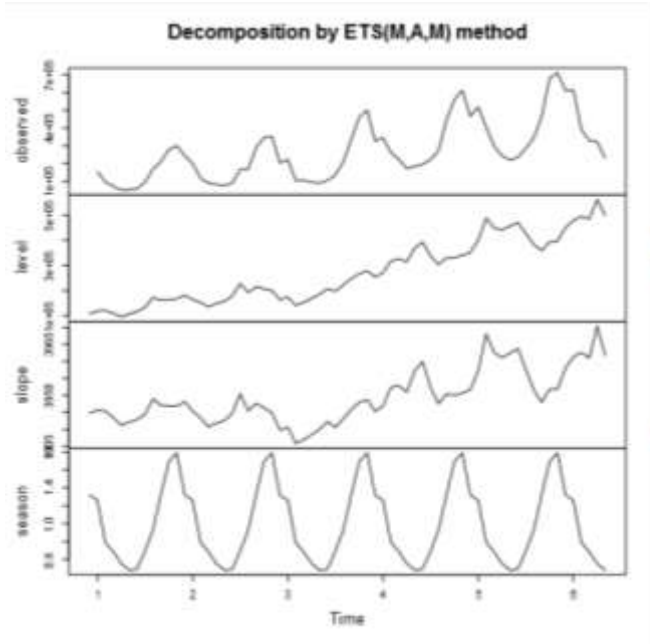


Fig. 3.5 Decomposition plot by ETS non-dampened model

Method:
ETS(M,A,M)

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

Fig. 3.6 In-sample error measures and information criteria reports for non-dampened ETS

Actual and Forecast Values:

Actual	ETS_nodampening
271000	248063.01908
329000	351306.93837
401000	471888.58168
553000	679154.7895

Accuracy Measures:

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

Fig. 3.7 Upper, Actual and forecast values using the hold-out sample for evaluating the accuracy of non-dampened ETS model. Bottom, Accuracy measures for no-dampened ETS model in terms of mean error, root mean squared error, absolute mean error, average percentage error, average absolute percentage error and mean absolute scaled error.

Comparison of dampened ETS & non-dampened ETS

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

Actual and Forecast Values:

Actual	ETS_dampening
271000	255966.17855
329000	350001.90227
401000	456886.11249
553000	656414.09775

Accuracy Measures:

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

Method:
ETS(M,A,M)

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	ETS_nodampening
271000	248063.01908
329000	351306.93837
401000	471888.58168
553000	679154.7895

Accuracy Measures:

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

From above comparison between Dampened ETS model and Non-dampened ETS model, we can see that dampened ETS model gives better forecast results using hold-out sample. All Accuracy measures except ME and MPE of non-dampened ETS model are higher than those of dampened ETS model. RMSE measure of obtained from in-sample error measures of dampened ETS model is higher than that of non-dampened ETS model. MASE from in-sample error measures of dampened ETS model is lower than non-dampened ETS model.

RMSE represents the standard deviation of difference between predicted values and observed values. MASE is mean absolute scaled error which is used to measure the relative reduction in error across the model. MASE is metrics for evaluating performance of a time series model. As MASE value of Dampened ETS is lower than non-dampened model, The Dampened ETS model should be chose.

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.

Non seasonal component

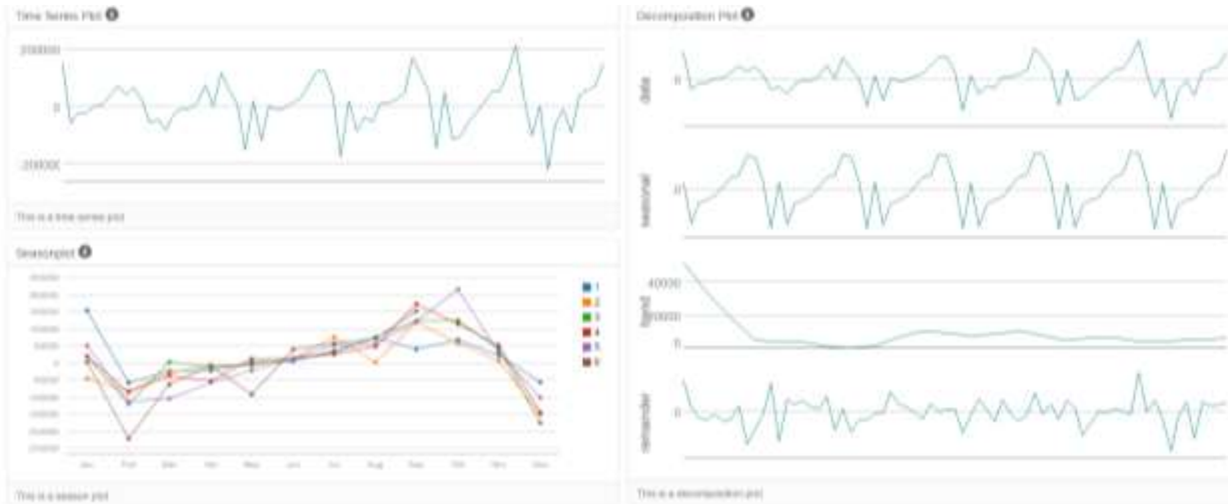


Figure 3.8 Time series plot and decomposition plot after the first non-seasonal differencing

From figure 2.1,2.2,2.3, we can say that series is not stationary. non-seasonal differencing is needed. After first non-seasonal differencing, the time series plot seems stationary as per figure 3.8. Hence integrated part(I) should be 1 for non-seasonal component of ARIMA model.

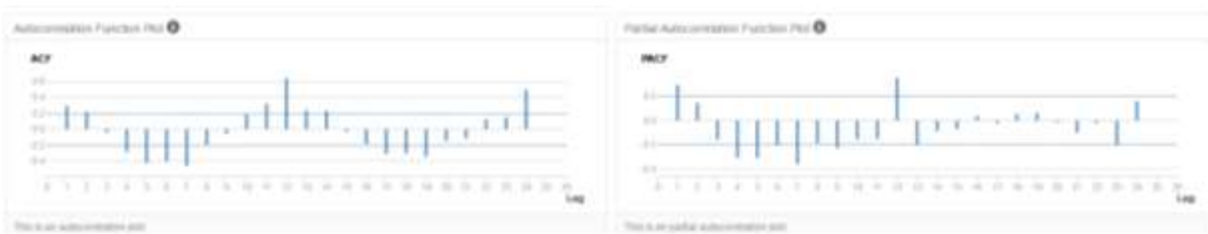


Figure 3.9 autocorrelation function plot (ACF) without differencing(LEFT); partial autocorrelation function plot (PACF) after the first non-seasonal differencing(RIGHT).

As per figure 3.9 ACF plot generated after first non-seasonal differencing shows shart cut-off after a few lags whereas the PACF plot gradual decat towards 0. This shows use of MA. Therefore, non-seasonal component of ARIMA model should be (0,1,1).

Seasonal Component

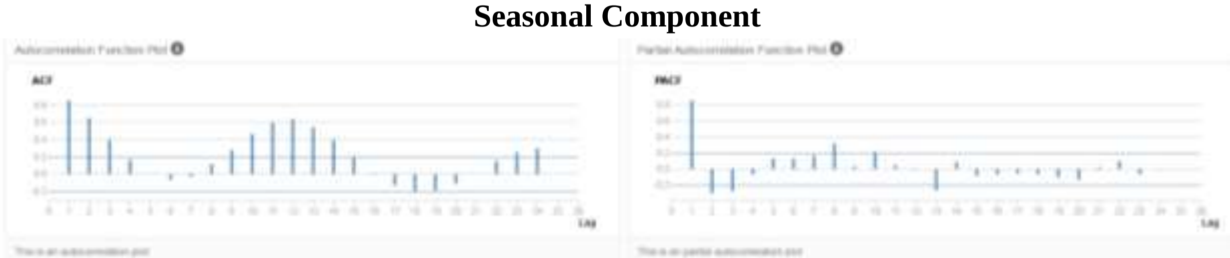


Fig. 3.10 Autocorrelation function plot (ACF without differencing(LEFT), Partial autocorrelation function plot PACF without any seasonal differencing.

From figure 3.10, Without any seasonal differencing ACF & PACF plots shows a high correlation of series at lag 1. Correlation coefficient is on vertical axis and lag value is on horizontal axis.

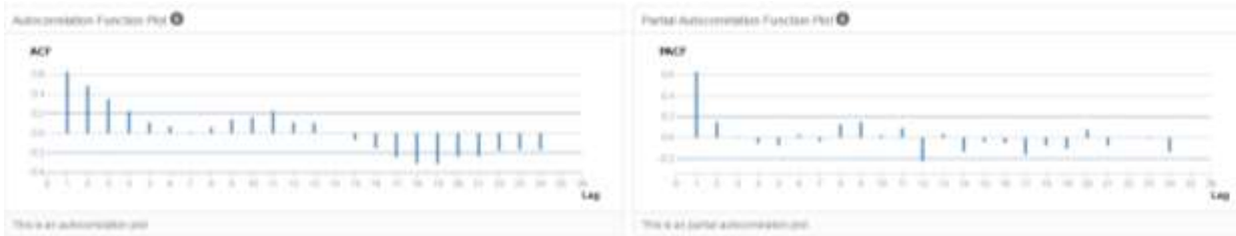


Fig. 3.11 Autocorrelation function plot ACF without differencing(LEFT), Partial autocorrelation function plot PACF after taking seasonal differencing(RIGHT)

Fig. 3.12 Time series plot after taking the seasonal differencing

As per fig 3.11, we can see high correlation of series after seasonal differencing is performed. High correlation at lag 1 in both the ACF and PACF plots in fig 3.11 confirm this fact. As per fig 3.12, time series plot after taking seasonal difference is not stationary. Therefore, the first seasonal differencing is needed.

Fig 3.13 Autocorrelation function plot ACF without differencing(LEFT), Partial autocorrelation function plot PACF after taking first seasonal differencing(RIGHT)

Fig. 3.14 Time series plot after taking first seasonal differencing

As per fig 3.14 after taking first seasonal differencing, we cannot see any high correlation shown in both ACF and PACF plots. Time series plot seems stationary which means series is now stationary. As ACF and PACF plot for this stationary dataset shows any signature of AR or MA term, neither AR or MA terms preferred. ARIMA model would follow (0,1,1) for non-seasonal part and (0,1,0) for seasonal part with a period of 12. Hence the notation should be ARIMA (0,1,1)(0,1,0)12. This analysis is consistent with what is chosen automatically by alteryx.

ARIMA (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

Fig. 3.15 Information criteria & in-sample error measures for (0,1,1)(0,1,0)12

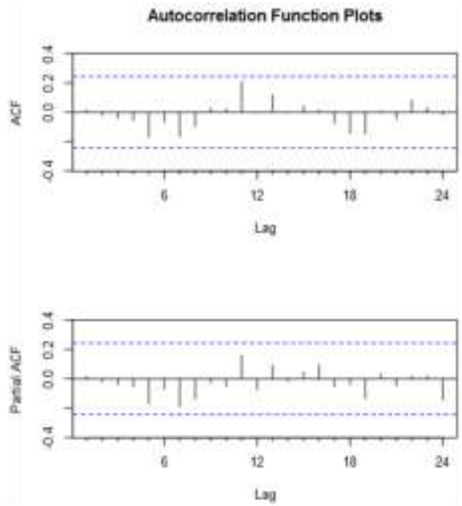


Fig. 3.16 ACF & PACF plots for ARIMA (0,1,1)(0,1,0)12

Actual and Forecast Values:

Actual	ARIMA_noMA
271000	263228.48013
320000	316228.48013
401000	372228.48013
553000	493228.48013

Accuracy Measures:

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

Fig 3.17 Forecast values & accuracy measures for ARIMA (0,1,1)(0,1,0)12

The RMSE and MASE values generated from in-sample error measures for ARIMA (0,1,1)(0,1,0)12 are 36761.5 and 0.365 respectively.

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.

Actual and Forecast Values:

Actual	ETS_dampening	ARIMA_noMA
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
ETS_dampening	-41317.07	60176.47	48833.98	-8.3683	11.1421	0.8116	NA
ARIMA_noMA	27271.52	33999.79	27271.52	6.1833	6.1833	0.4532	NA

Fig 4.1 Accuracy measures and in-smaple error measures of dampened ETS vs ARIMA model using hold-out sample

ARIMA model is chosen over dampened ETS due to its higher accuracy in predicting the monthly sales for hand-out samples. RMSE value of dampened ETS is 60176.47, which is higher than ARIMA model which is 33999.79. Also MASE value of ARIMA model which is 0.4532 is much lower than ETS model 0.8116.

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

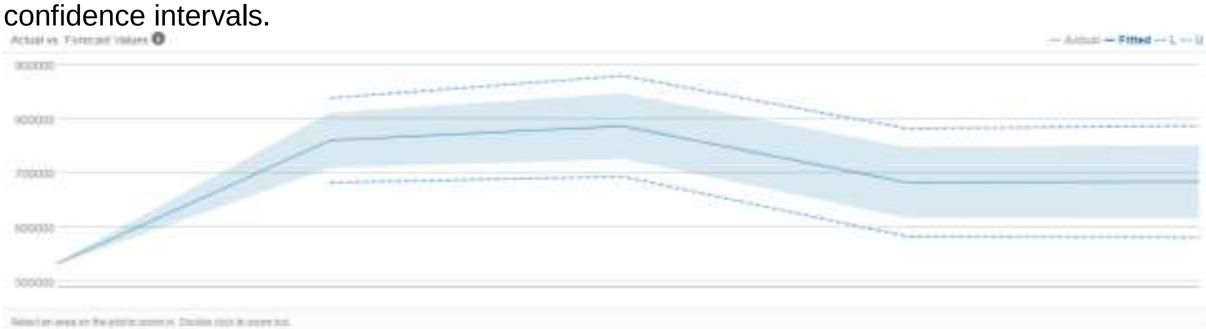
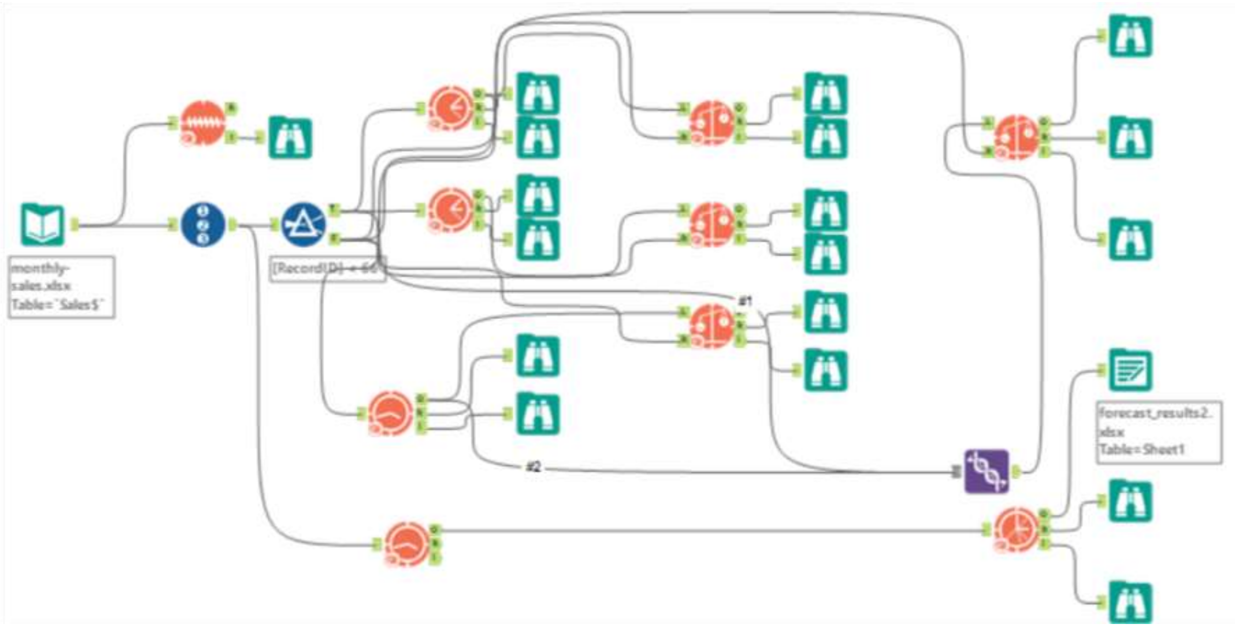


Fig 4.2 Forecast values for next 4 months using 95% & 80% CI

Fig. 4.3 Forecast values for next 4 months using 95% & 80% CI(Zoom out)

Forecast values for next 4 months are follows:

OCT-13	760617.2
NOV-13	786812.7
DEC-13	683059.1
JAN-14	684481



Alteryx workflow