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.**
   To meet the criteria of a time series dataset:
   i)   Each measurement of data taken across a continuous time interval is sequential and of equal intervals.
   ii)   Each time unit having at most one data point.
   iii) Ordering matters in the list of observations.
   iv)   Dependency of time.
2. **Which records should be used as the holdout sample?**
   i)   Holdout sample size depends on how far the prediction is.
   ii)   Since we need to predict the sales for the next 4 months.
   iii) A 4-month long holdout sample from Jun-13 till Sept-13 should be used.

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.**
   i)   Using TS plot function, the time series and decomposition plots are generated.
   ii)   Seasonality and trend show increasing trends, thus multiplication and addition should be applied respectively.
   iii) There isn't a trend for error plot but rather fluctuations and thus should be applied multiplicatively as well.
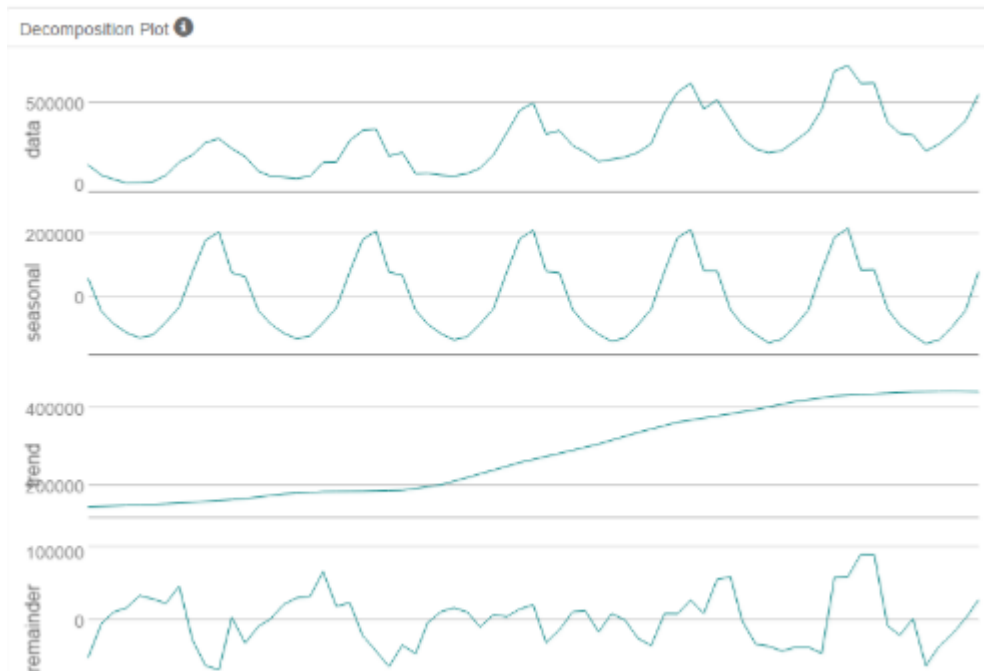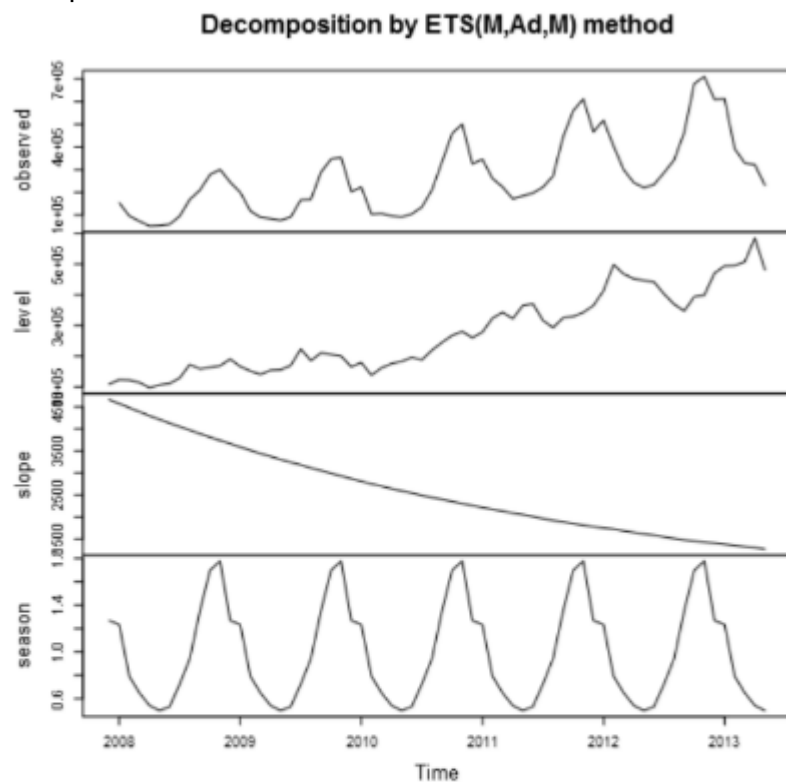

Time Series Plot

This is a time series plot

Seasonplot

This is a season plot
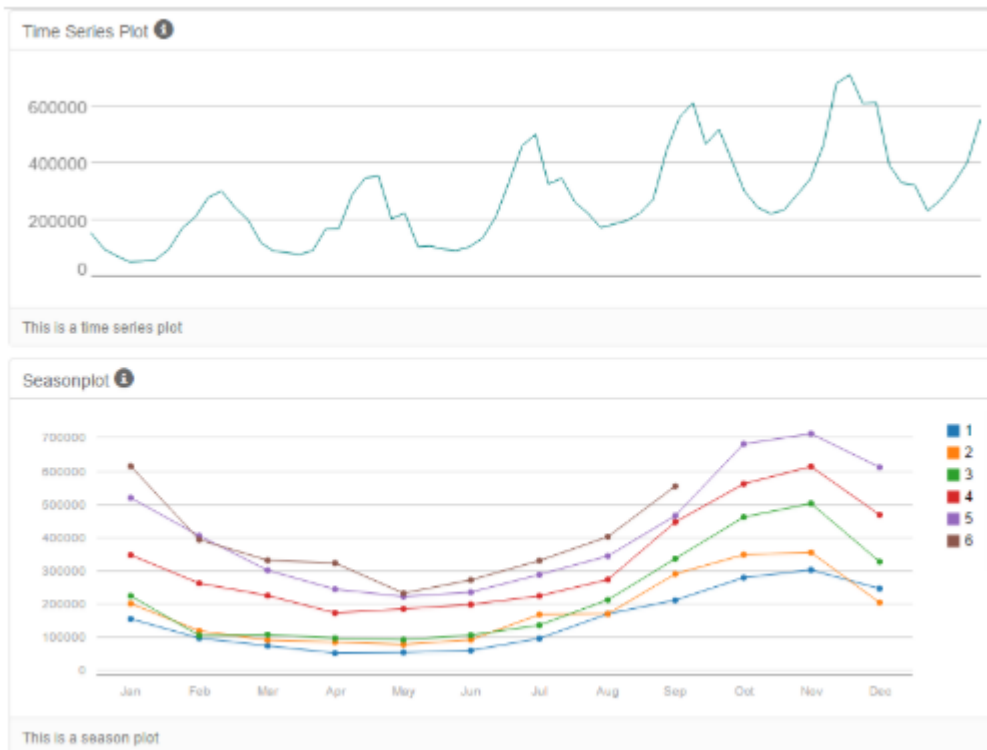
Fig: Time series, season and decomposition plots.

Step 3: Build your Models

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

   Dampened ETS Model:

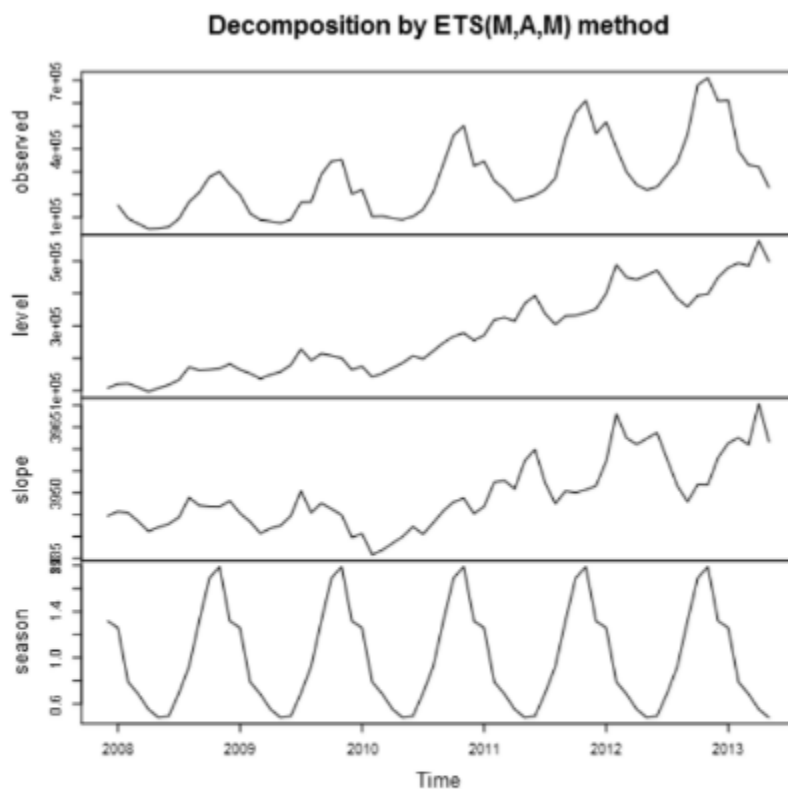Decomposition by ETS(M,Ad,M) method

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

The AIC value is **1639.47**, RMSE is **33153.53** and MASE is **0.3675**.
Non-Dampened ETS Model:

Decomposition by ETS(M,A,M) method

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

The AIC value is 1639.74, RMSE (Moot Mean Square Error) is 32992.73 and MASE (Mean Absolute Percentage Error) is 0.3727.

Dampened:

Actual and Forecast Values:

Actual	ETS_damped
271000	255966.17855
329000	350001.90227
401000	456886.11249
553000	656414.09775

Accuracy Measures:

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

Non-Dampened:

Actual and Forecast Values:

Actual	ETS
271000	248063.01908
329000	351306.93837
401000	471888.58168
553000	679154.7895

Accuracy Measures:

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

Dampened model is chosen by comparing the forecast and actual results due to its higher accuracy.
The dampened model's RMSE & MASE are lower and could offset its marginally lower AIC.

2. **What are the model terms for ARIMA? Explain why you chose those terms.**
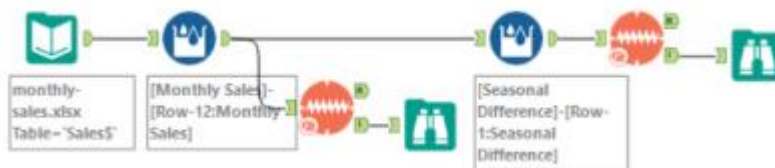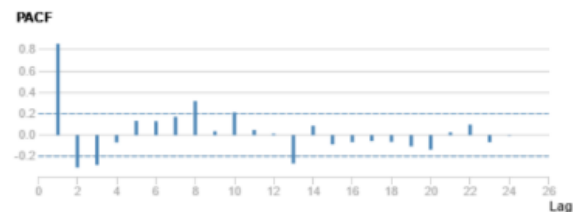

Without differencing, the time series and seasonal component's Auto-Correlation Function (ACF) shows high correlation and the Partial Autocorrelation Function (PACF) shows a significant lag at period 13 which is due to seasonal effect.

Then a seasonal difference is taken.
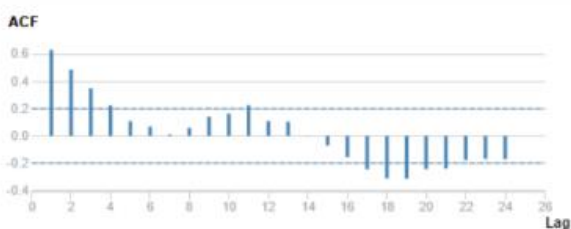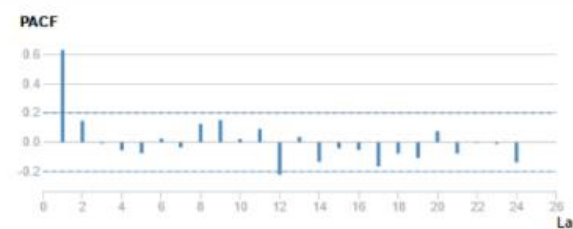However, the ACF still shows high correlation while the data doesn't have strong correlation in PACF after a seasonal difference is applied.

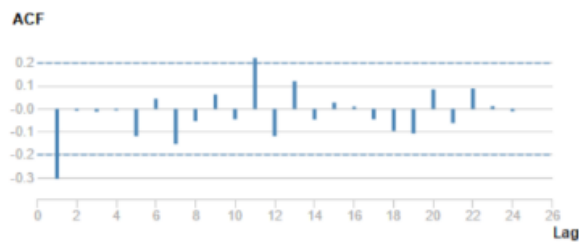Seasonal first difference is performed and ACF plot doesn't show strong correlation anymore.

ACF

PACF
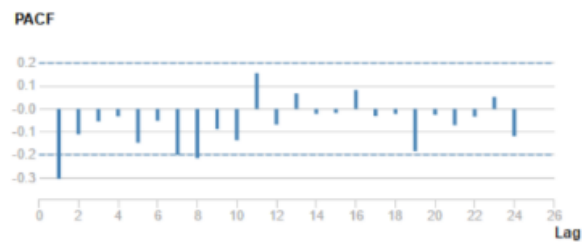
The number of period is 12 months and ARIMA(0,1,1)(0,1,0)12 is used as lag-1 is negative.

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

The AIC is 1256.60, RMSE is 36761.53 and MASE is 0.3646.

Autocorrelation Function Plots
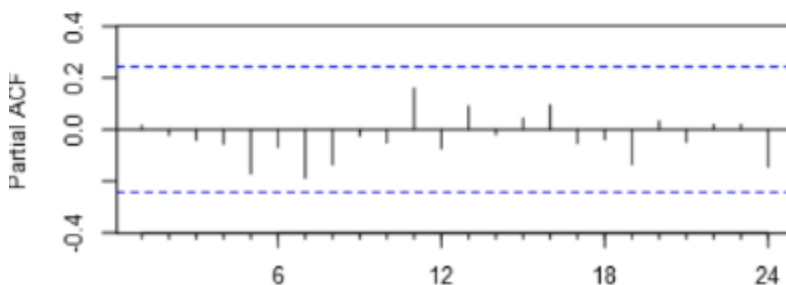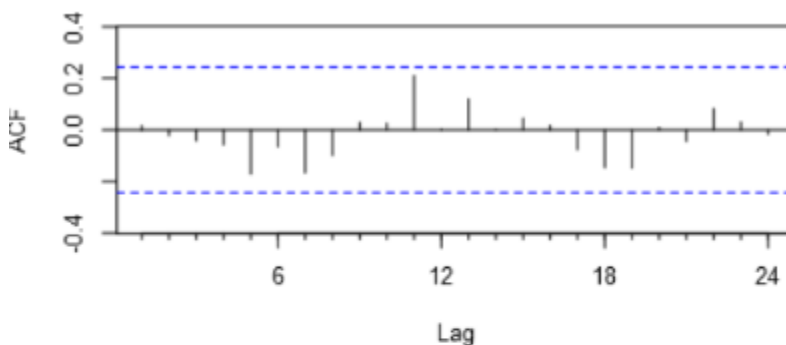
Significant correlation is not shown by both ACF and PACF and no additional AR or MA terms needed.

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.**

   At forecasting sales ARIMA model is better using holdout sample as validation data as the MAPE and ME value are lower than ETS model.

   For ARIMA the RMSE is 33999.79 compared to ETS' RMSE at 60176.47.

ARIMA's MASE value of 0.4532 is also lower than ETS' MASE value of 0.8116.
It is clear that ARIMA model is better since its in-sample error measurements and forecast error measurements are smaller.

ETS Model:

Actual and Forecast Values:

Actual	ETS_damped
271000	255966.17855
329000	350001.90227
401000	456886.11249
553000	656414.09775

Accuracy Measures:

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

ARIMA Model:

Actual and Forecast Values:

Actual	ARIMA
271000	263228.48013
329000	316228.48013
401000	372228.48013
553000	493228.48013

Accuracy Measures:

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

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

Actual vs. Forecast Values ❶   — Actual — Fitted -- L -- U
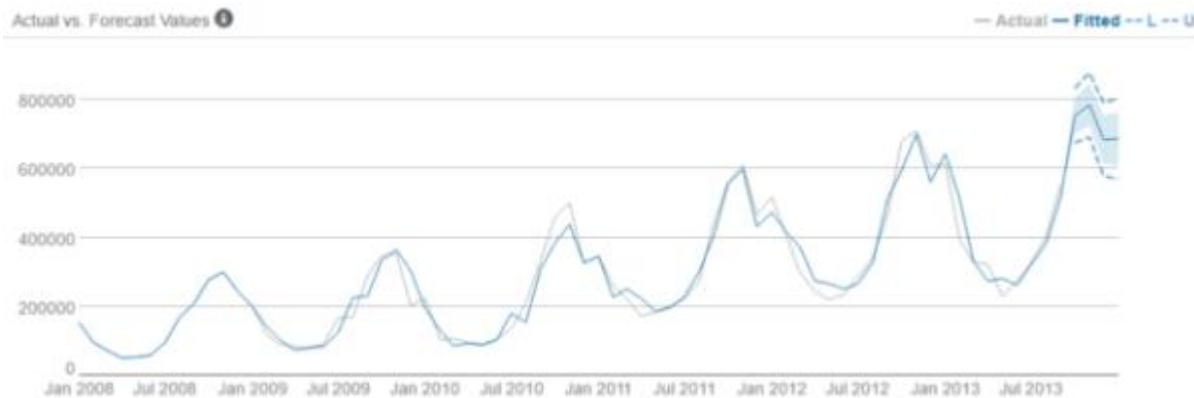
For the next 4 periods (Oct-13 till Jan-14) the forecasts are **754,854**, **785,854**, **684,654** and **687,854**
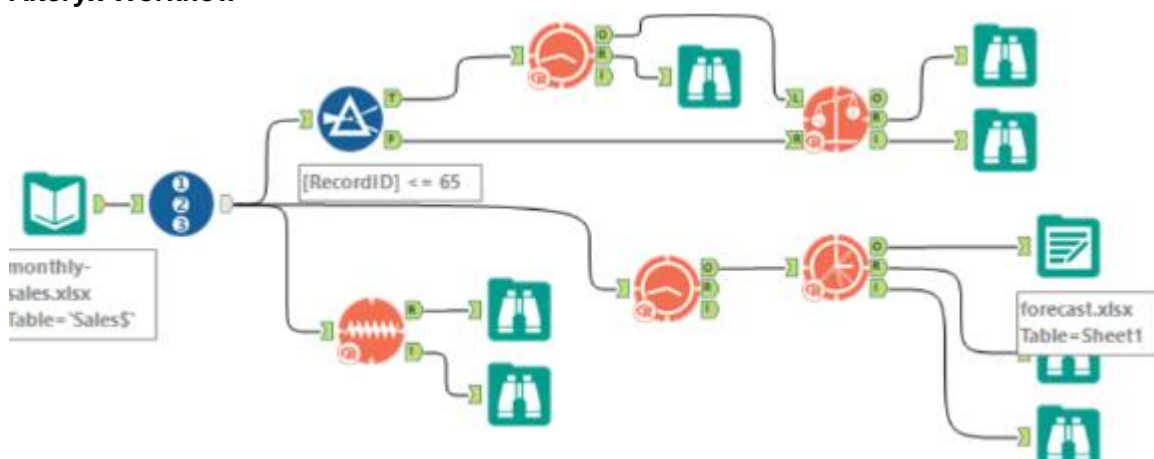
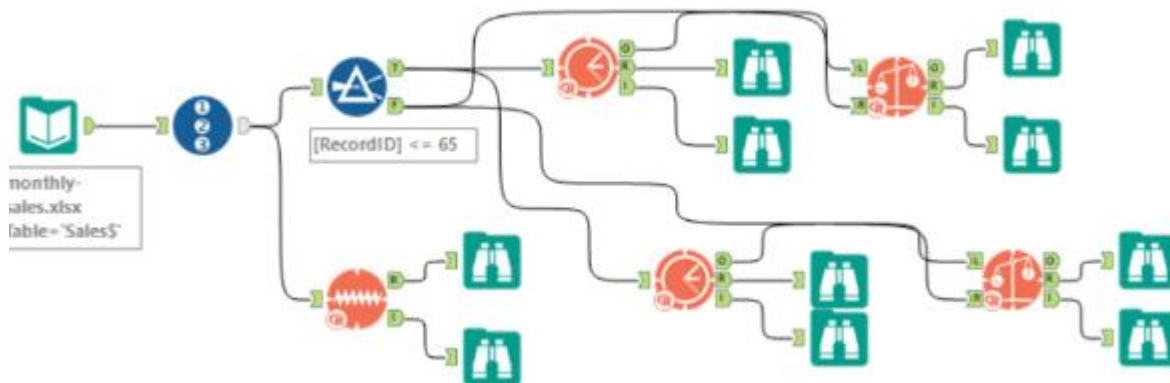**Alteryx Workflow**

Fig: ARIMA workflow

Fig: ETS workflow