Project: Forecasting Sales

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 dataset meets the criteria of a time series dataset as shown below:

1. Over a continuous time interval
    - The dataset is over a continuous period and in YYYY-MM format.
2. Of sequential measurements across the interval
    - The dataset is ordered chronologically by month and year.
3. Using equal spacing between every two consecutive measurements
    - Each data point is separated by monthly intervals each year.
4. With each time unit within the time interval having at most one data point
    - There are data point for each time interval.

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

Size of the holdout sample will depend on how long the time series is and how far we would like to forecast. Since we need to predict the sales for the next 4 months, the holdout sample will be the final 4 months of the dataset from Jun-2013 to Sep-2013.

Determine Trend, Seasonal, and Error components

1. **What are the trend, seasonality, and error of the time series?**

The time series decomposition plot shown below allows us to observe the seasonality, trend and error/remainder terms of a time series. The trend plot is linear so we apply it additively (A). The size of the seasonal fluctuations tends to increase and decrease with the level of time series so we will apply it multiplicatively (M). The error plot is fluctuating between large and small errors over time so we will apply it multiplicatively (M).

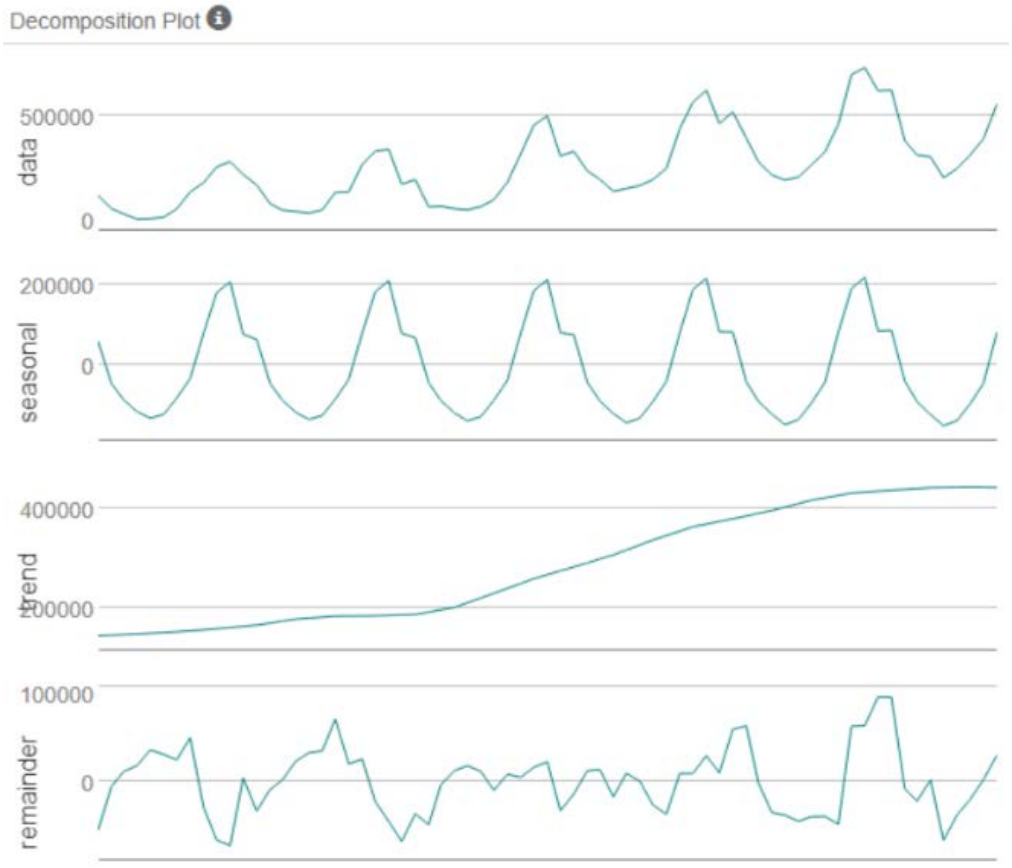


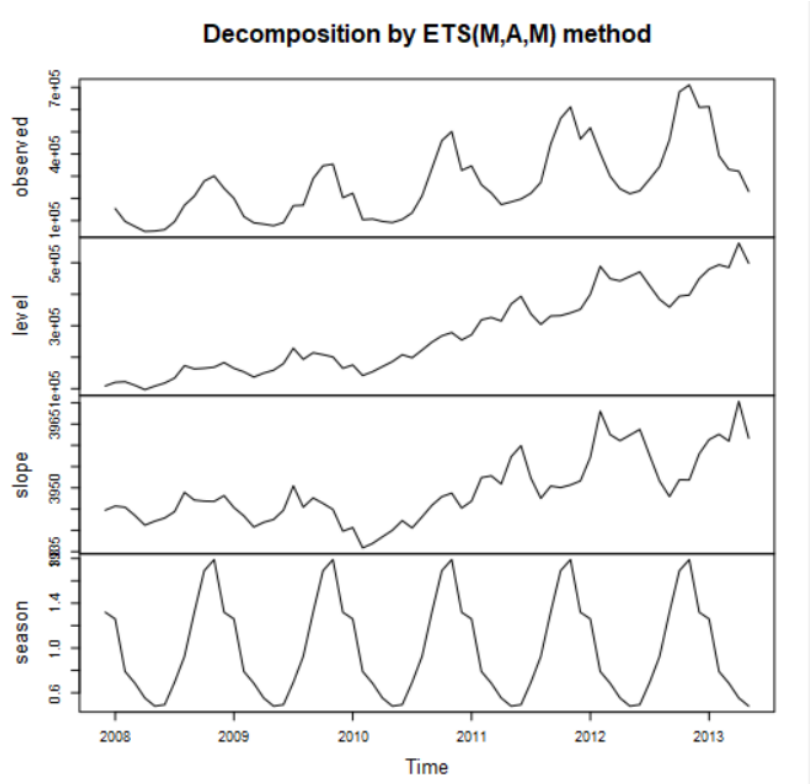
Figure 1 Decomposition Plot

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

ETS(M,A,M) is chosen based on the decomposition plot shown above. Dampened and non-dampened ETS models are run with a holdout sample of 4 months.

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

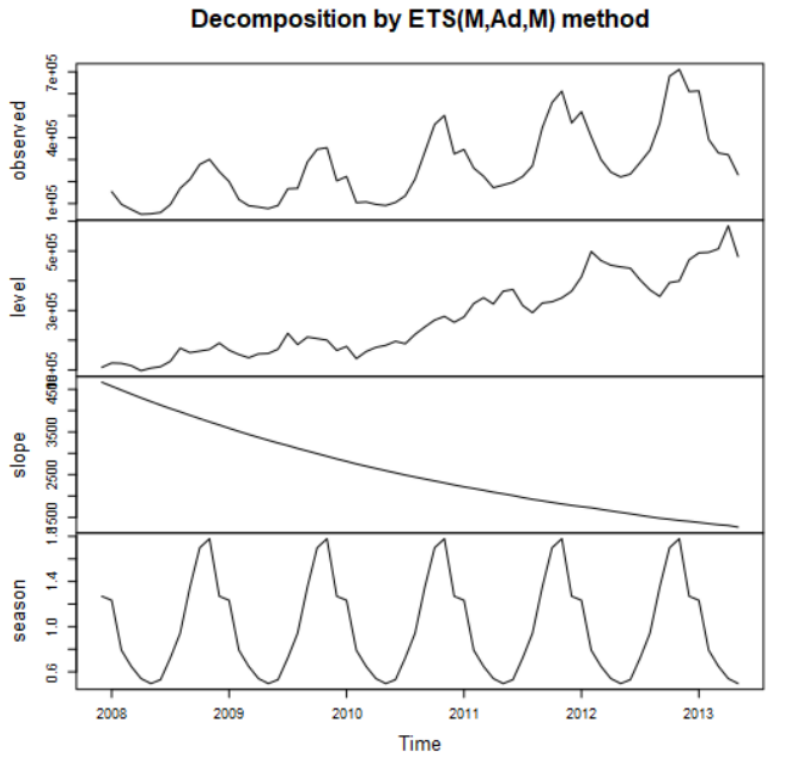
Information criteria:

AIC	AICc	BIC
1639.7367	1652.7579	1676.7012

Figure 2 Undampened ETS Report

The AIC value is 1639.74, RMSE (Root Mean Square Error) is 32992.73 and MASE (Mean Absolute Scaled Error) is 0.37.

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

Figure 3 Dampened ETS Report

The AIC value is 1639.47, RMSE (Root Mean Square Error) is 33153.53 and MASE (Mean Absolute Scaled Error) is 0.37.

Based on the ETS result shown above, we can see that the non-dampened ETS model actually gives better model that dampened ETS model with overall less errors.

TS Comparison Result:

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
MAM	-49103.33	74101.16	60571.82	-9.7018	13.9337	1.0066

Figure 4 TS Comparison Result (Non-dampened ETS model)

Actual and Forecast Values:

Actual	MAM_dampened
271000	255966.17855
329000	350001.90227
401000	456886.11249
553000	656414.09775

Accuracy Measures:

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

Figure 5 TS Comparison Result (Dampened ETS model)

By comparing the forecast and actual results, the dampened model is chosen due to its higher accuracy. The dampened model's RMSE (60176.47) and MASE (0.8116) are lower and could offset its marginally lower AIC.

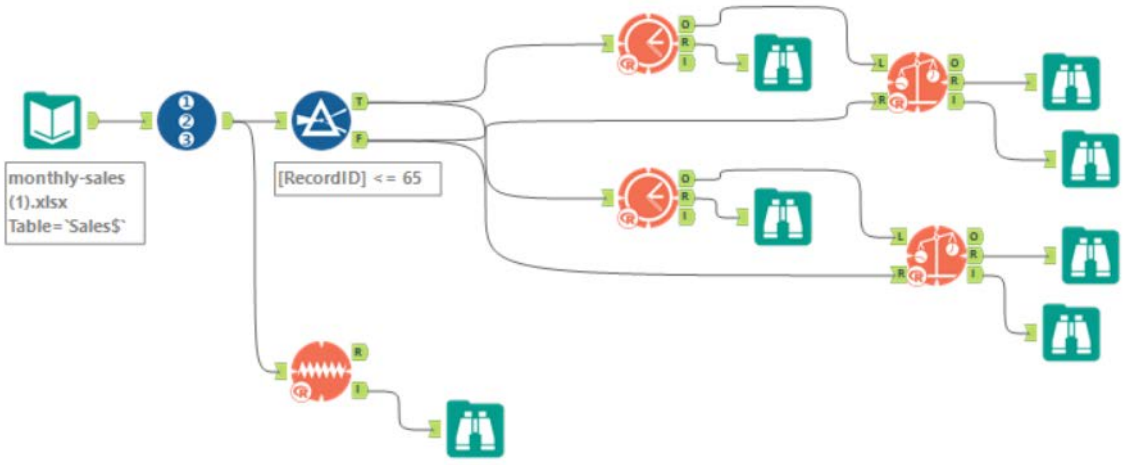


Figure 6 Alteryx Flow (ETS)

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.

Without differencing, we can see that the Auto-Correlation Function (ACF) shows slowly decaying serial correlation towards 0 with increases at the season lags.

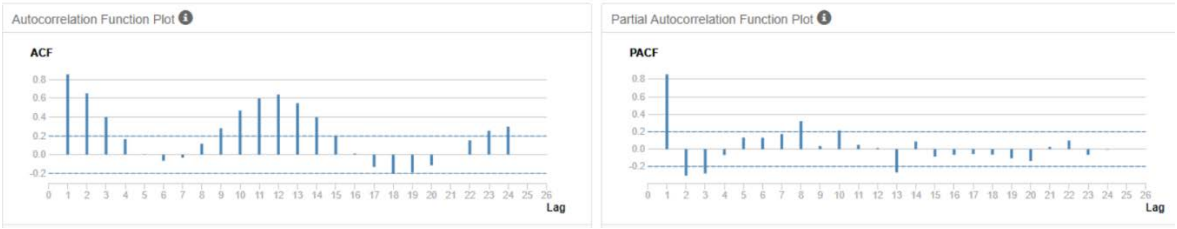


Figure 7 TS Plot (without differencing)

A seasonal difference is taken but the ACF still shows high correlation while the data doesn't have strong correlation in PACF after a seasonal difference is applied. The remaining correlation can be accounted for using autoregressive and moving average terms.

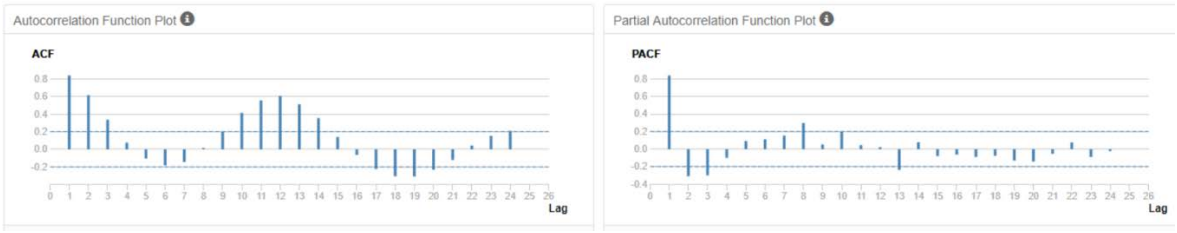


Figure 8 TS Plot (seasonal difference)

A sesonal first difference is performed and has removed most of the significant lags from the ACF and PACF so there is no meed for futher differencing.

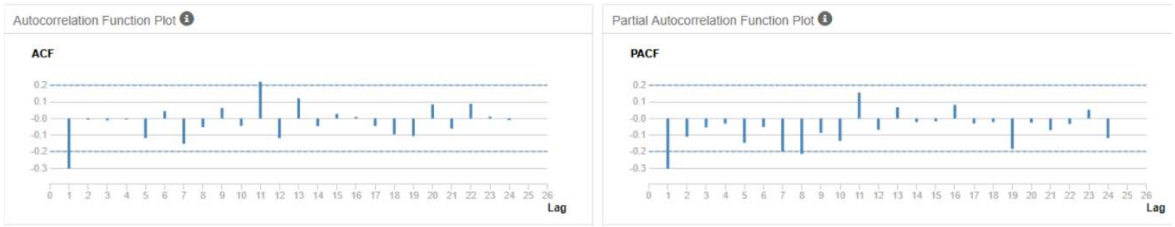


Figure 9 TS Plot (seasonal first difference)

The ACF plot shows a strong negative correlation at lag-1 which is confirmed in the PACF plot. This suggest an MA(1). There isn't significant correlation in the ACF and PACF thus there will not need for seasonal autoregressive or moving average terms, which means it would be 0 for both P & Q. The forecast is monthly thus m will be 12. Therefore, we can define the ARIMA(p,d,q)(P,D,Q)m will be ARIMA(0,1,1) (0,10)12.

ARIMA (0,1,1) (0,1,0) 12 Result:

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

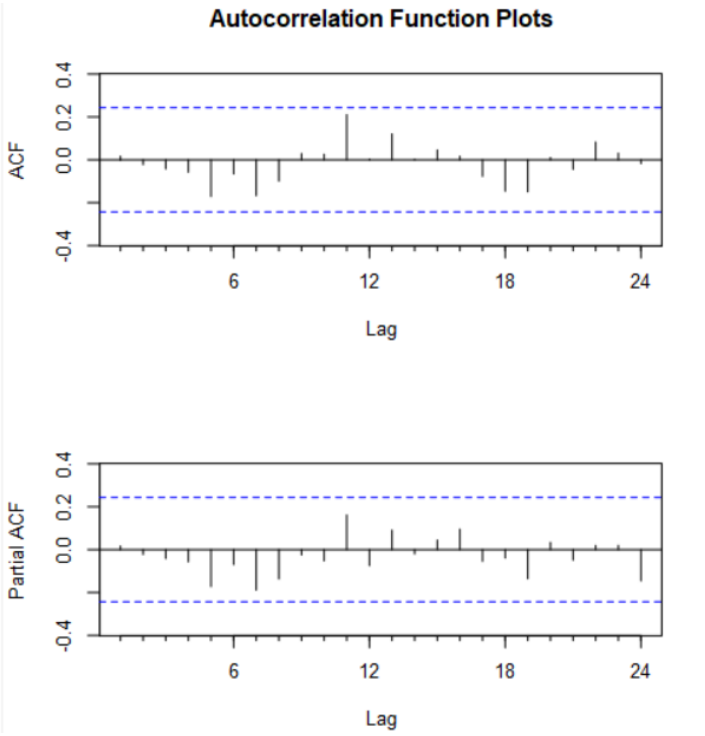


Figure 10 ARIMA Result

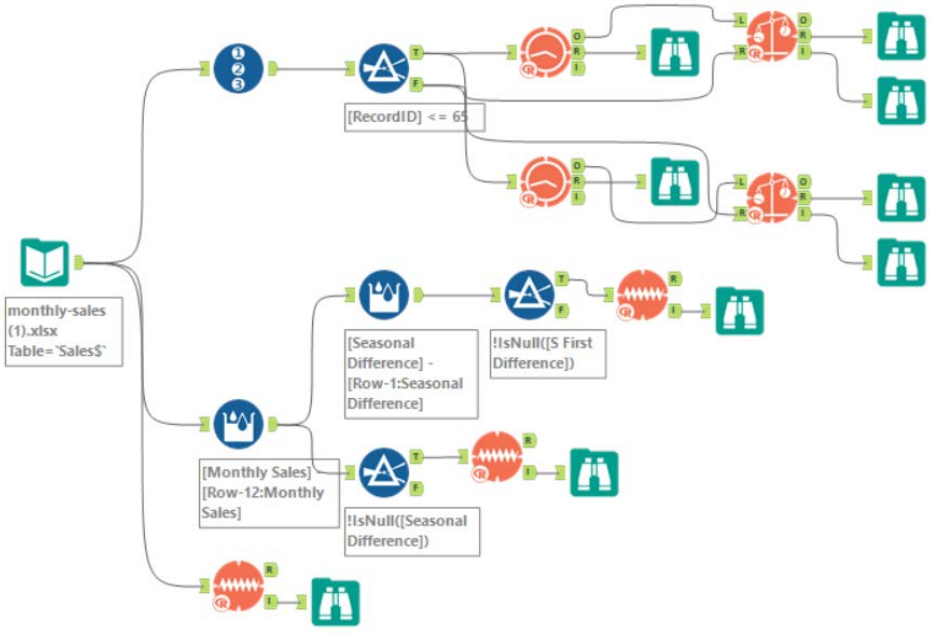


Figure 11 Alteryx Flow (ARIMA)

Forecast

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

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
ARIMA	27271.52	33999.79	27271.52	6.1833	6.1833	0.4532

ETS Model:

Actual and Forecast Values:

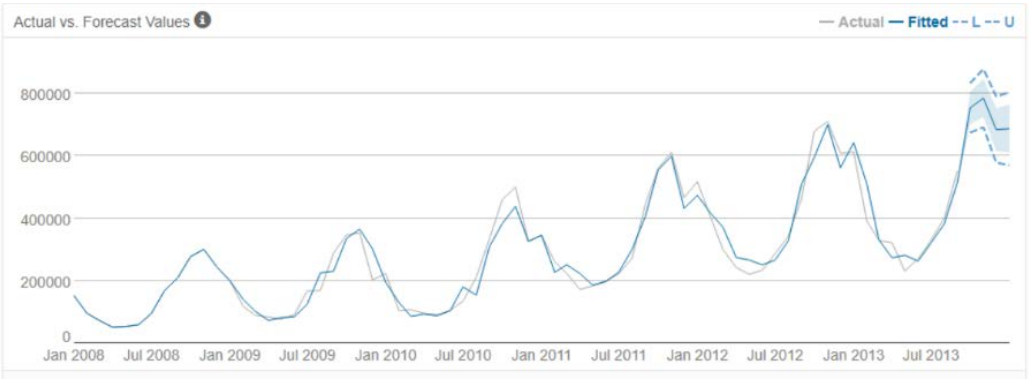
Actual	MAM_dampened
271000	255966.17855
329000	350001.90227
401000	456886.11249
553000	656414.09775

Accuracy Measures:

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

The RMSE for ARIMA 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. Hence ARIMA model is chosen to forecast the result.

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



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

Figure 12 ARIMA Forecast

The forecast for next 4 periods (Oct-2013 till Jan-2014) are 754,854, 785,854, 684,654 and 687,854.

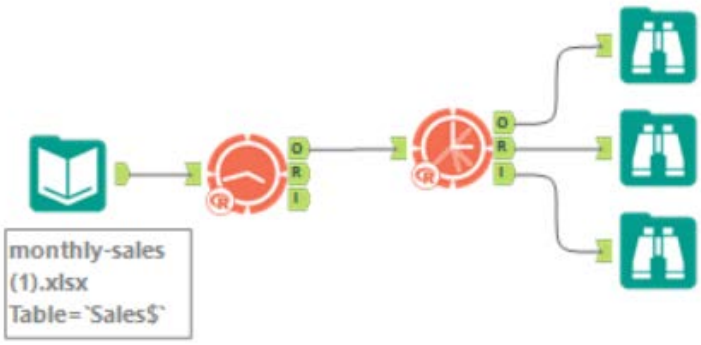


Figure 13 Alteryx (ARIMA Forecast)