# 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.**

   The dataset meets the criteria of a of a time series dataset.
   The dataset is taken across a continuous time interval. ✓
   The dataset is sequential. ✓
   The dataset is also of equal monthly intervals. ✓
   Lastly, for the dataset, each time unit has only one data point. ✓

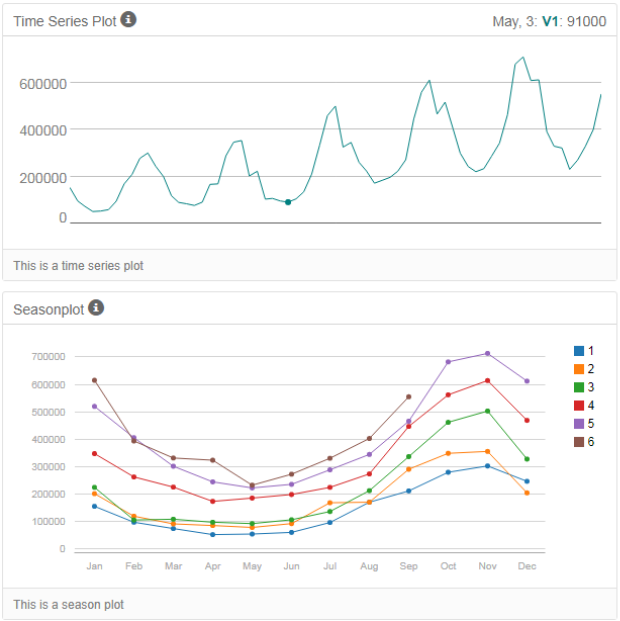
2. **Which records should be used as the holdout sample?**

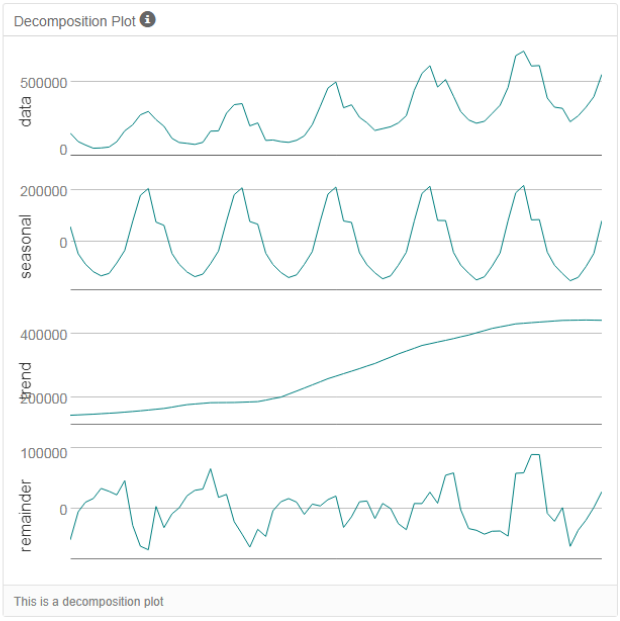
   Since we need to forecast 4 months of sales, records from the last 4 periods which is 2013-06 to 2013-09 should be used.

## 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.**

   To observe the seasonality, trend, and error/remainder terms of a time series, we use a time series decomposition plot. For Error, we see the remainder plot fluctuating between large and small errors over time. For Trend, we see a trend line that rather seems linear. For Seasonal, we see the seasonal fluctuations which tends to increase with the level of time series.



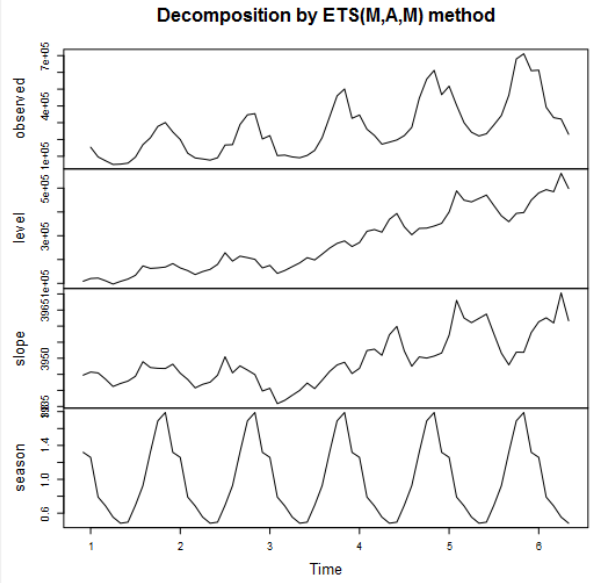



# Build your Models

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

   Based on the time series decomposition plot above, ETS(M,A,M) is chosen. For Error, we see the plot fluctuating between large and small errors over time, so we apply multiplicatively (M). For Trend, the trend line is linear, so we apply additively (A). For Seasonal, size of the seasonal fluctuations tends to increase with the level of time series, so we apply it multiplicatively (M). Also, dampened and non-dampened ETS models are run with 4 months of holdout samples.

### Non-dampened ETS Model




Method:
ETS(M,A,M)

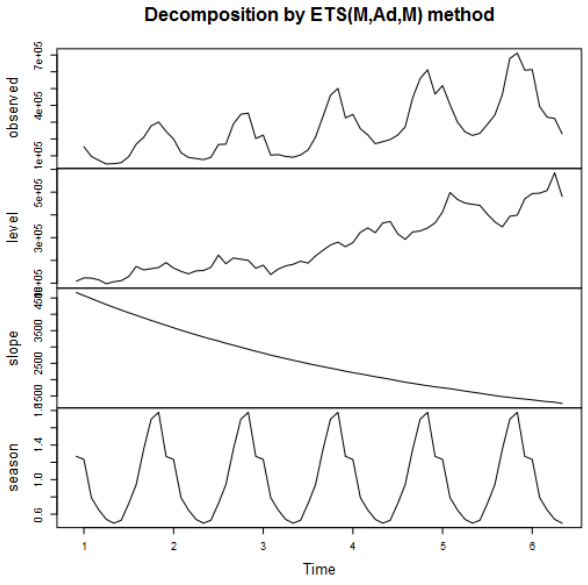
In-sample error measures:

| 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 |

### Dampened ETS Model




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 |

Looking at in-sample error measures, RMSE and MASE are 32992.73 and 0.3727 for the non-dampened ETS model and 33153.53 and 0.3675 for the dampened ETS model.
Also, the AIC value is 1639.74 for the non-dampened ETS model and 1639.47 for the dampened ETS model.

**Non-dampened ETS Model**

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 |
|---|---|---|---|---|---|---|
| ETS | -49103.33 | 74101.16 | 60571.82 | -9.7018 | 13.9337 | 1.0066 |

**Dampened ETS Model**

Actual and Forecast Values:

| Actual | ETS_dampened |
|---|---|
| 271000 | 255966.17855 |
| 329000 | 350001.90227 |
| 401000 | 456886.11249 |
| 553000 | 656414.09775 |

Accuracy Measures:

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

Overall, looking at the accuracy measures by comparing the forecasted values with the holdout sample, the dampened ETS model is closer to the actual value. Therefore, dampened ETS model is chosen due to its higher accuracy.

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.**

Looking at the in-sample error measures, RMSE is 36761.53 and MASE is 0.36. Also, the AIC value is 1256.60.

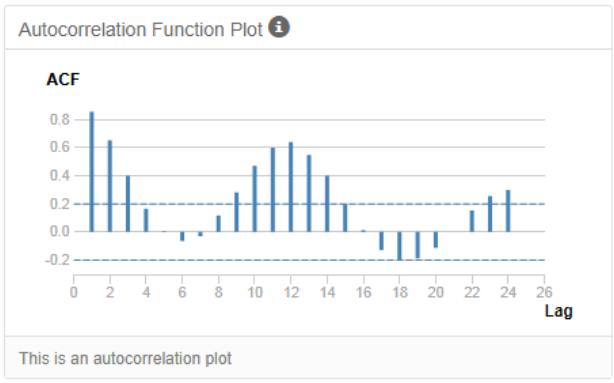
Information Criteria:

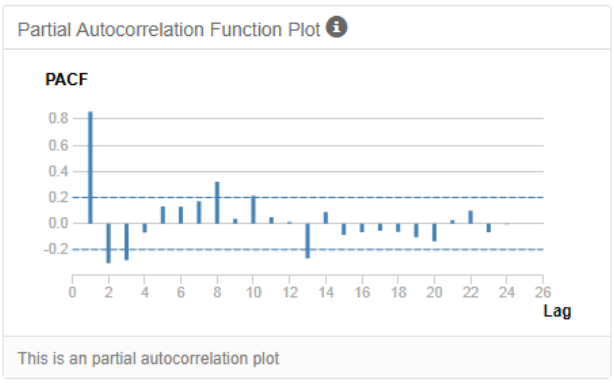
| AIC | AICc | BIC |
|---|---|---|
| 1256.5967 | 1256.8416 | 1260.4992 |

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 |

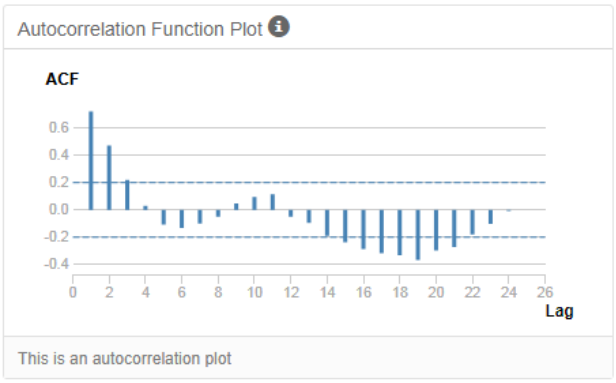
**Original time series plot without differencing**

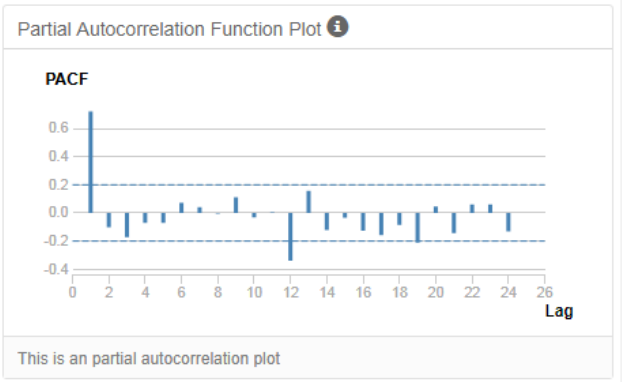



The ACF and PACF plot show positive high correlation at lag 1.

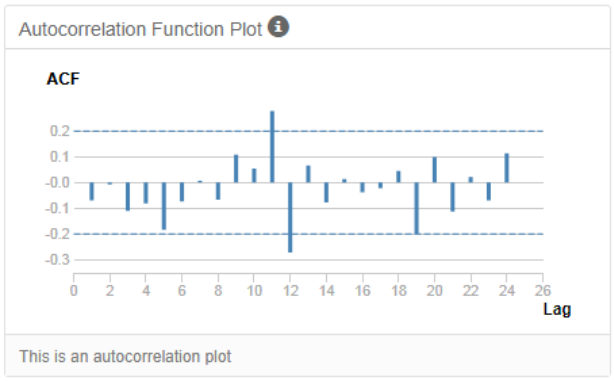
**Seasonal difference**

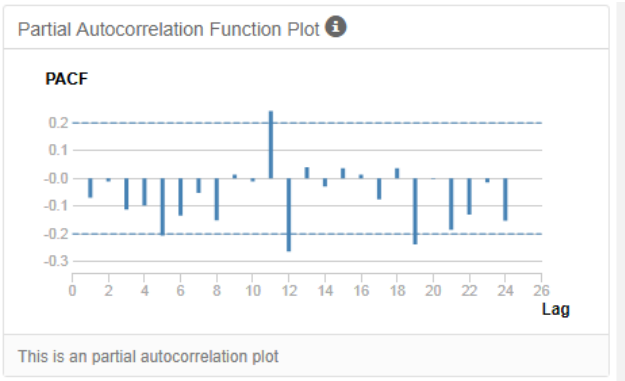


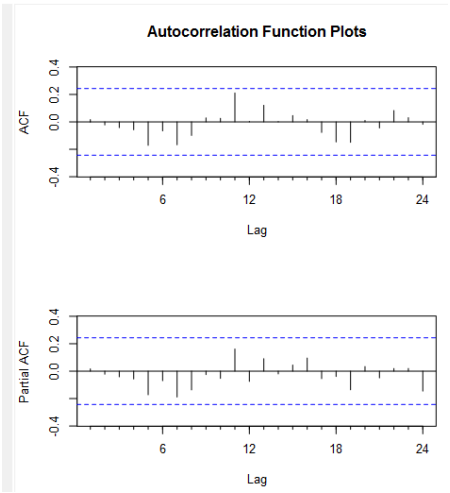

After applying seasonal difference, ACF still shows high correlation at lag 1, yet the data does not show a strong correlation in PACF after applying seasonal difference.

**Seasonal first difference**




After applying the seasonal first difference, ACF plot no longer shows a strong correlation. Since lag 1 is negative, ARIMA model (0,1,1) (0,1,0)12 is used.




The Autocorrelation Function plots for ACF and PACF does not display a AR or MA behavior so no other AR or MA terms are necessary.

# Forecast

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

   Looking at the in-sample error measures, the RMSE for ARIMA model is 33999.79 while for the dampened ETS model is 60176.47.
   The MASE for ARIMA model is 0.4532 while for the dampened ETS model is 0.8116.
   Also, looking at the accuracy measures by comparing the forecasted values with the holdout sample, ARIMA model's forecasted values are closer to the actual values.
   Therefore, **ARIMA model** is better since its in-sample error measurements and forecast error measurements are smaller than the ETS model.

Actual and Forecast Values:

| Actual | ARIMA |
|---|---|
| 271000 | 263228.48013 |
| 329000 | 316228.48013 |
| 401000 | 372228.48013 |
| 553000 | 493228.48013 |

Actual and Forecast Values:

| Actual | ETS_dampened |
|---|---|
| 271000 | 255966.17855 |
| 329000 | 350001.90227 |
| 401000 | 456886.11249 |
| 553000 | 656414.09775 |

Accuracy Measures:

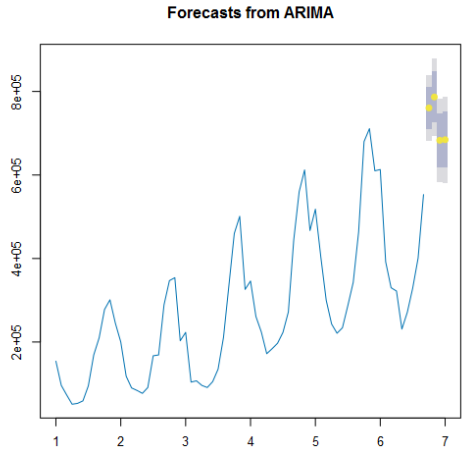
| Model | ME | RMSE | MAE | MPE | MAPE | MASE |
|---|---|---|---|---|---|---|
| ARIMA | 27271.52 | 33999.79 | 27271.52 | 6.1833 | 6.1833 | 0.4532 |

Accuracy Measures:

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

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

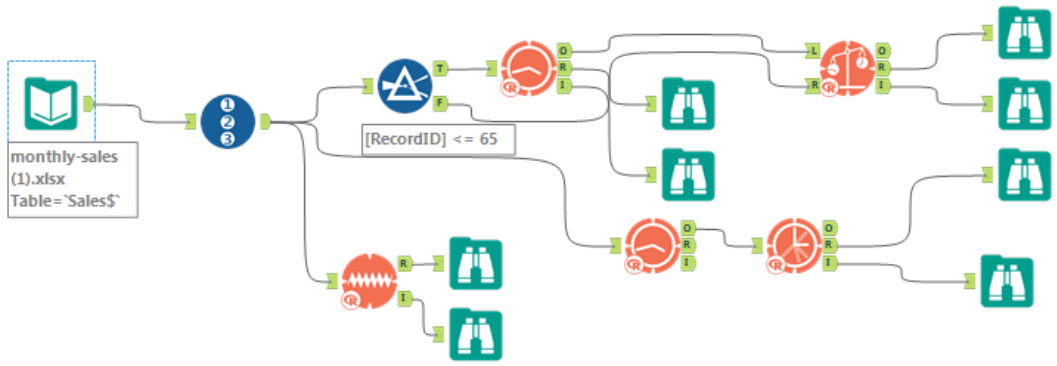
   The forecast for the next 4 months of sales are 760617.15, 786812.70, 683059.13 and 684480.98.



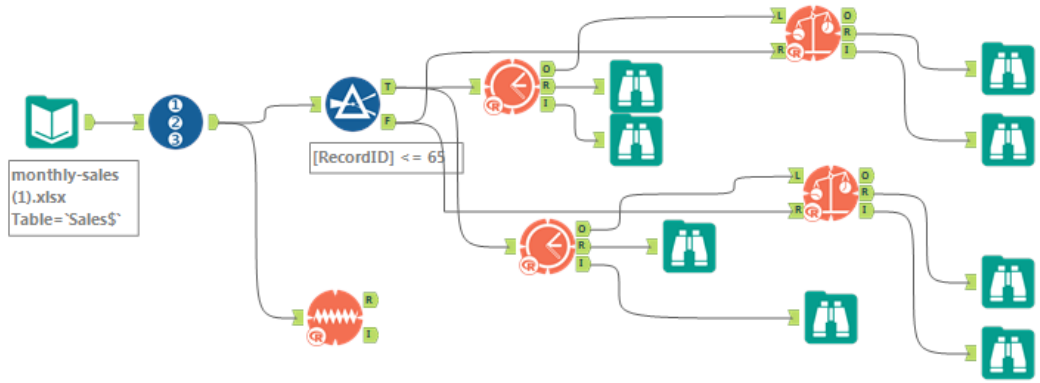

| Period | Sub_Period | forecast | forecast_high_95 | forecast_high_80 | forecast_low_80 | forecast_low_95 |
|---|---|---|---|---|---|---|
| 6 | 10 | 760617.152585 | 838430.096697 | 811496.301964 | 709738.003206 | 682804.208473 |
| 6 | 11 | 786812.700678 | 879783.548852 | 847603.070791 | 726022.330566 | 693841.852505 |
| 6 | 12 | 683059.130563 | 782556.813242 | 748117.168369 | 618001.092757 | 583561.447885 |
| 7 | 1 | 684480.980021 | 787471.896139 | 751823.120402 | 617138.839641 | 581490.063904 |

# Alteryx Workflow

## ARIMA Model




## ETS Model




## Time Series and Seasonal Component