<u>Project: Forecasting Sales</u>

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 dats is over a continuous time interval (from 01.2008 to 09.2013)
- There are sequential measurements across that interval
- There is equal spacing between every two consecutive measurements (monthly data)
- Each time unit within the time interval has at most one data point

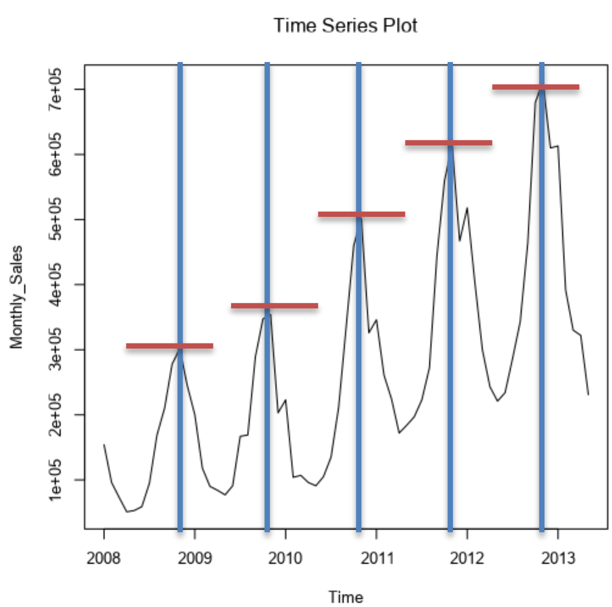
2. **Which records should be used as the holdout sample?**

I used the last 6 months of data as a training set, which is almost 9% of data sample.

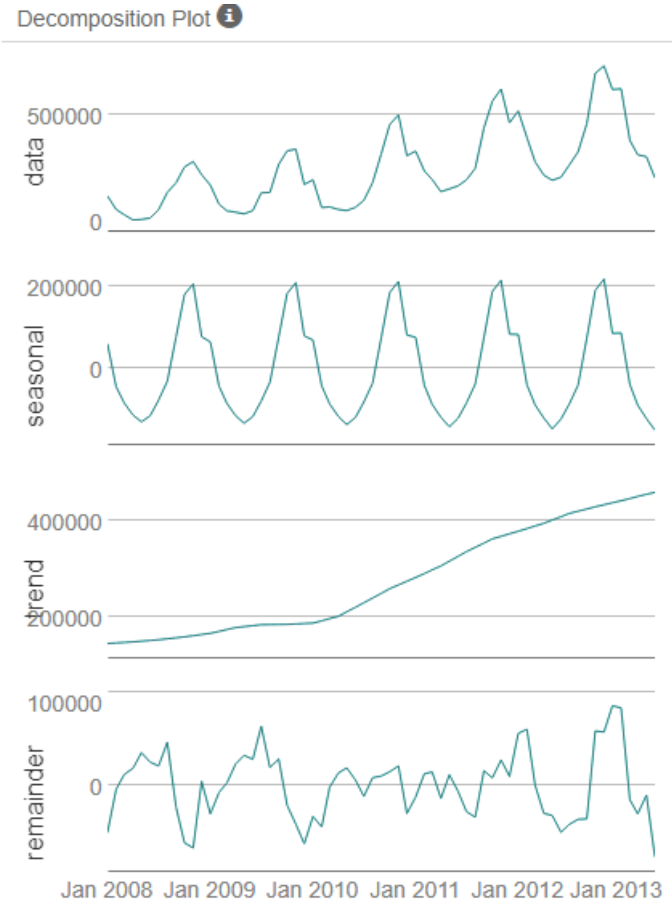
Step 2: Determine Trend, Seasonal, and Error components

Answer this question:

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.**



The Time Series Plot depicts regularly occurring spikes that confirm seasonality in the time series. The data is not cyclic and has upward trend which is raising over time.



The Decomposition Plot confirms the seasonality and increasing trend of the time series.

The seasonality of the data suggest us to use seasonal ARIMA model.

The change in magnitude of seasonality suggest to use **multiplicative method** in the seasonal component of EST model.

Step 3: Build your Models

Answer these questions:

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 model

The decomposition plot shows:

- E - Remainder plot (error) changes along with data and suggest to use a **Multiplicative method.**
- T - Trend increases over time with linear behaviour which suggest to use an **Additive method**.
- S – Seasonality changes in magnitude each year, so a **Multiplicative method** is suggested**.**

ETS (M,A,M) model with trend damping is suggested, *ETS Error Terms:*

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.

ETS (M,A,M) with trend damping

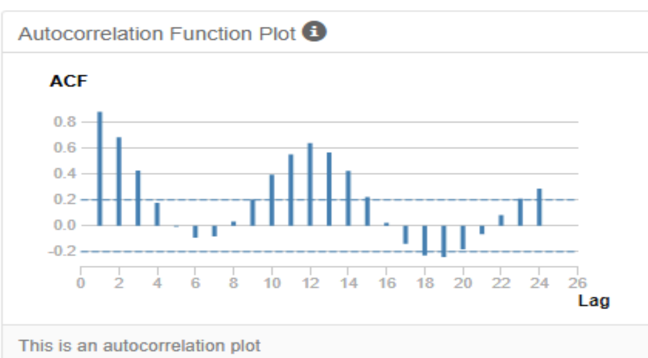
ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
3243.4703524	31474.3668886	24188.2167878	-0.572395	10.3052041	0.3528697	0.0087402

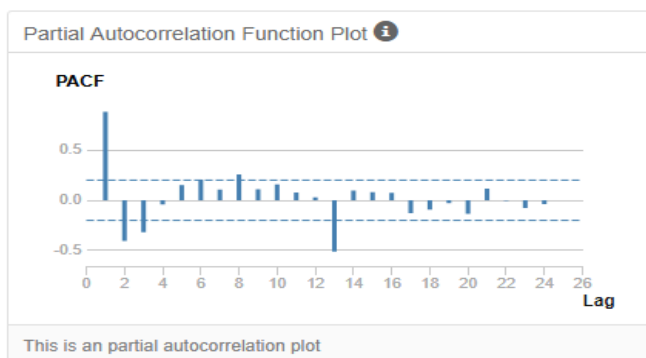
ARIMA model

The decomposition plot shows seasonal components in the time series **ARIMA(p, d, q)(P, D, Q) S model** should be used for forecasting.

Time Series ACF and PACF

The ACF presents slowly decaying serial correlations towards 0 with increases at the seasonal lags. Since serial correlation is high I will need to seasonally difference the series.

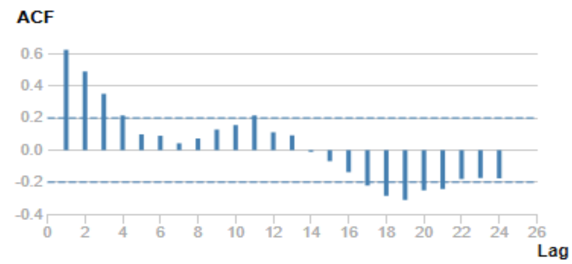




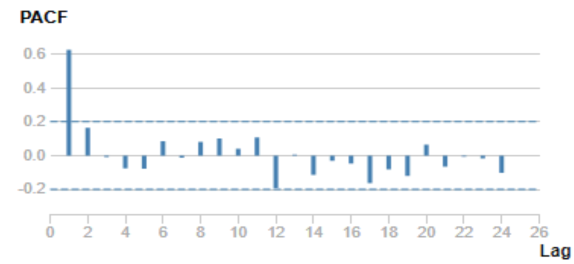
Seasonal Difference ACF and PACF

The seasonal difference presents similar ACF and PACF results as the initial plots without differencing, only slightly less correlated. In order to remove correlation we will need to difference further.

Autocorrelation Function Plot



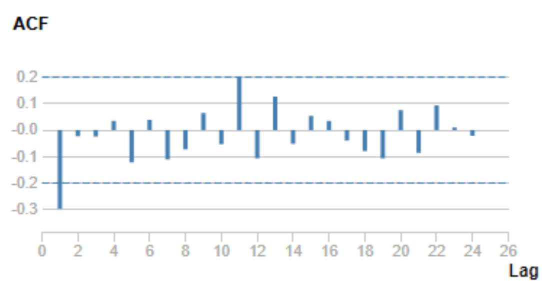
Partial Autocorrelation Function Plot



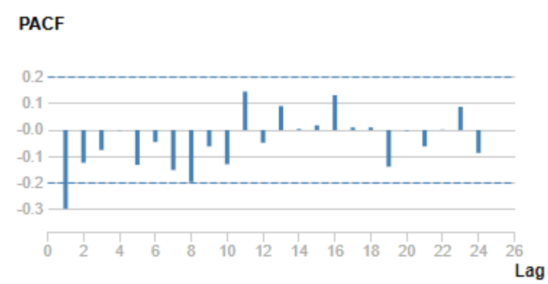
Seasonal First Difference:

The seasonal first difference of the series has removed most of the significant lags from the ACF and PACF so there is no need for further differencing. The remaining correlation can be accounted for using autoregressive and moving average terms and the differencing terms will be d(1) and D(1).

Autocorrelation Function Plot



Partial Autocorrelation Function Plot



The ACF plot shows a strong negative correlation at lag 1 which is confirmed in the PACF. This suggests an MA(1) model since there is only 1 significant lag. The seasonal lags (lag 12, 24, etc.) in the ACF and PACF do not have any significant correlation so there will be no need for seasonal autoregressive or moving average terms.

Therefore the model terms for my ARIMA model are: **ARIMA(0, 1, 1)(0, 1, 0)[12]**

ARIMA Error Terms:

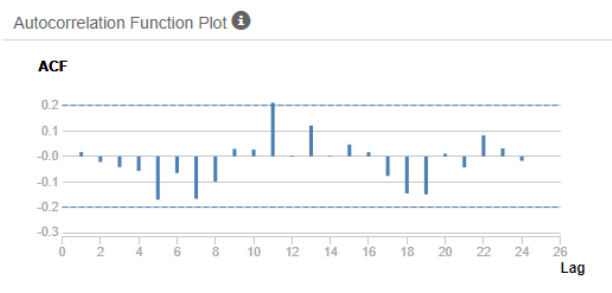
1. What are the model terms for ARIMA? Explain why you chose those terms. Graph the Auto-Correlation Function (ACF) and Partial Autocorrelation Function Plots (PACF) for the time series and seasonal component and use these graphs to justify choosing your model terms.
   a. Describe the in-sample errors.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

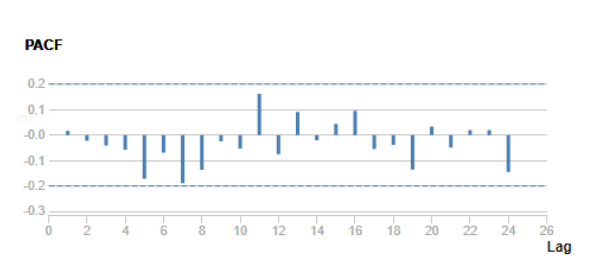
ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
-356.2665104	36761.5281724	24993.041976	-1.8021372	9.824411	0.3646109	0.0164145

The **RMSE** shows the in-sample standard deviation and its value is 36 762 units around the mean.
The **MASE** shows the mean absolute error is 0.36 that is significantly less 1.00.

b. Regraph ACF and PACF for both the Time Series and Seasonal Difference and include these graphs in your answer.

ACF and PACF does not have significantly correlated lags, justifying the **ARIMA(0, 1, 1)(0, 1, 0)[12]** model





Step 4: Forecast

Compare the in-sample error measurements to both models and compare error measurements for the holdout sample in your forecast. Choose the best fitting model and forecast the next four periods. (250 words limit)

1. Which model did you choose? Justify your answer by showing: in-sample error measurements and forecast error measurements against the holdout sample.
2. What is the forecast for the next four periods? Graph the results using 95% and 80% confidence intervals.

ETS in-sample error measures:

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
3243.4703524	31474.3668886	24188.2167878	-0.572395	10.3052041	0.3528697	0.0087402

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

Comparison RMSE and MASE: Both models have similar results. The ETS model has RMSE with less standard deviation by about 5 000 units, and slightly smaller MASE value than ARIMA model.

Comparison MAPE and ME: ARIMA model has smaller values for both MAPE and ME value and more precise with its forecast.

Accuracy Measures:

Model	ME	RMSE	MAE	MPE	MAPE	MASE
ARIMA_011_010_12	27271.52	33999.79	27271.52	6.1833	6.1833	0.4532
ETS_MAM_trend_damp	-33469.61	53828.48	41542.75	-6.3476	9.3266	0.6904

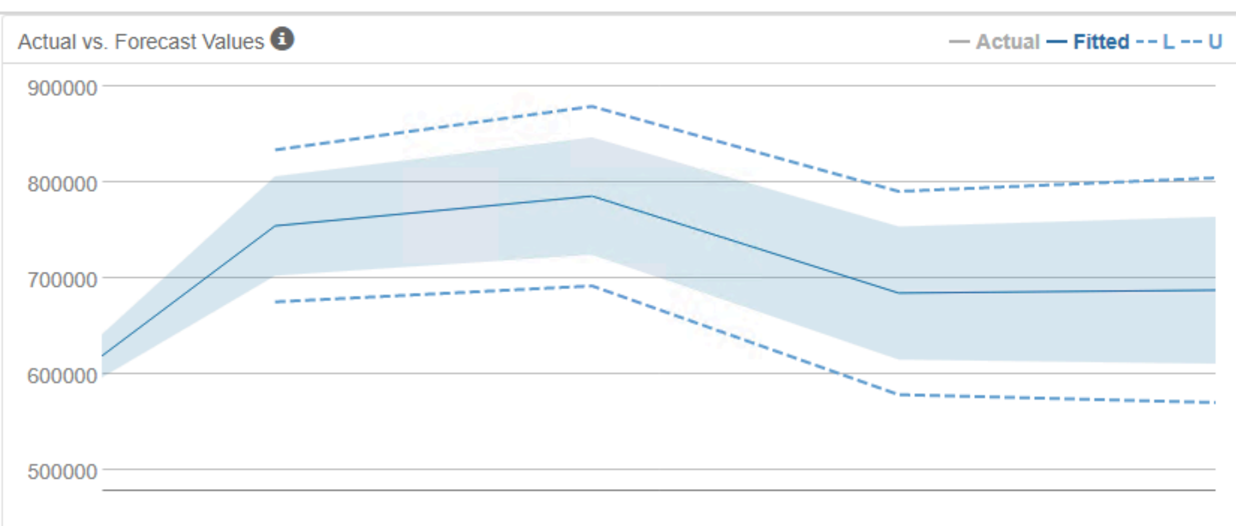
Both models have similar magnitude of bias, where ARIMA underestimates while ETS overestimates forecast data in comparison with validation set.

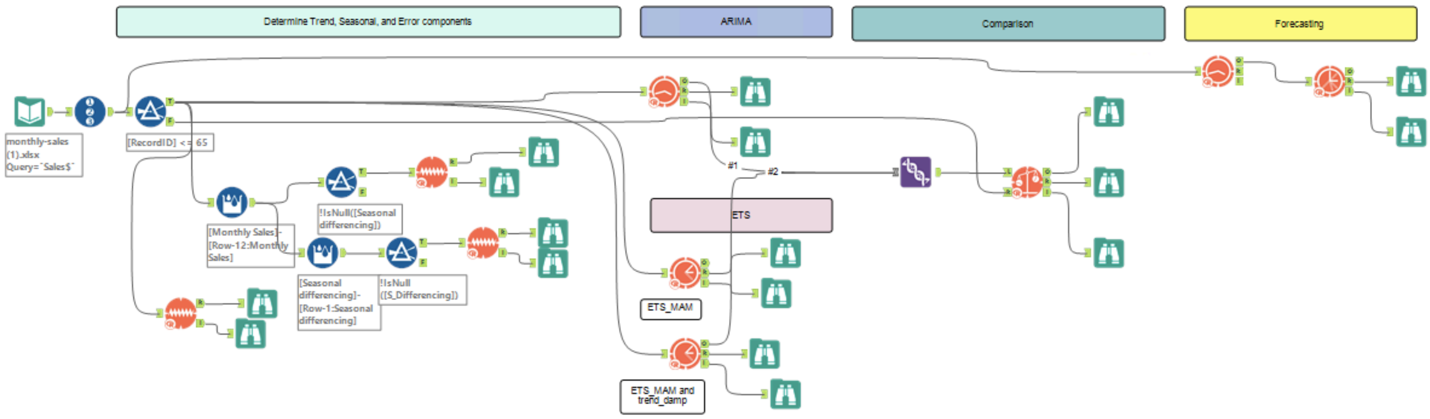
ARIMA model has lower values for RMSE, MASE, MAPE, MAE that tells ARIMA is more precise in forecasting and we should use it to predict.

The table below depicts forecast for the next four periods:

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

The graphical presentation of data:




Determine Trend, Seasonal, and Error components
ARIMA
Comparison
Forecasting
monthly-sales (1).xlsx Query=`Sales$`
[RecordID] <= 65
[Monthly Sales]-[Row-12:Monthly Sales]
!IsNull([Seasonal differencing])
[Seasonal differencing]-[Row-1:Seasonal differencing]
!IsNull ([S_Differencing])
ETS
ETS_MAM
ETS_MAM and trend_damp
#1
#2