

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.

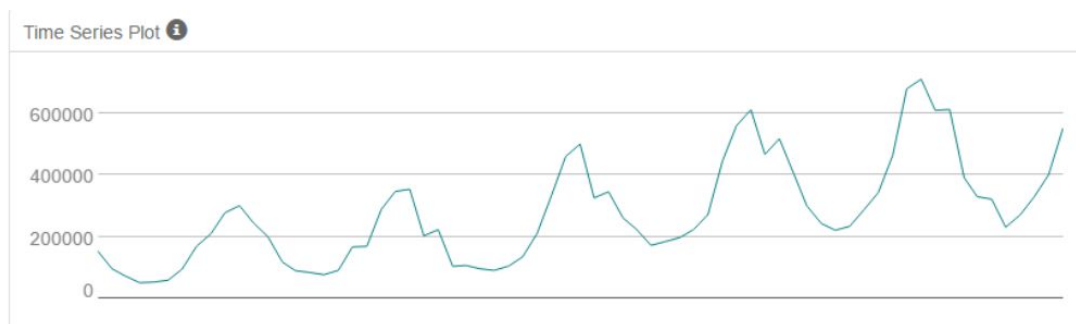
The dataset meets the criteria of a times series dataset:

- A. The data is represented over a continuous time interval
 - B. There is a sequential measurement across that interval
 - C. There is equal spacing between two-time points
 - D. Each time point has at most one data post.
2. Which records should be used as the holdout sample?

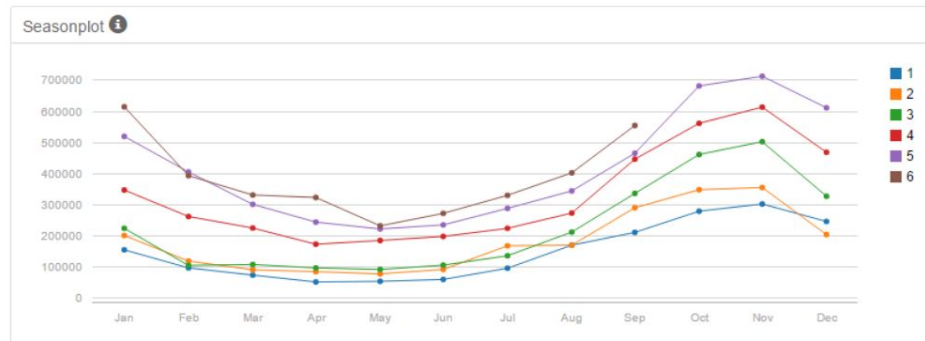
Holdout sample size depends on how far the prediction is. Since we need to predict the sales for the next 4 months, a 4-month long holdout sample from Jun-13 till Sept-13 should be used.

Step 2: Determine Trend, Seasonal, and Error components

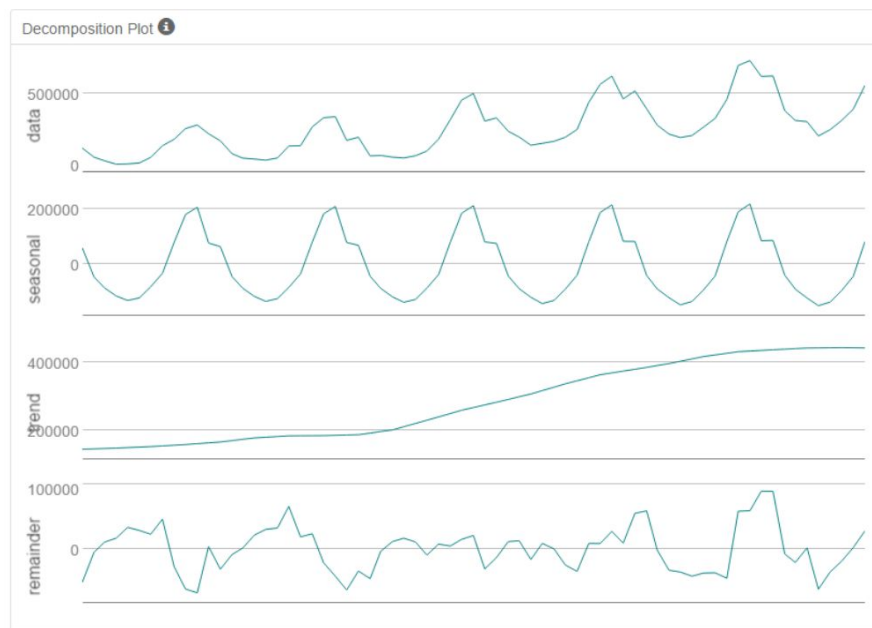
- What are the trend, seasonality, and error of the time series? Show how you were able to determine the components using time series plots. Include the graphs.



-Times Series Plot-



- Season Plot-



-Decomposition Plot-

From the graphs we can see that :

The trend plot is linear, so we apply it additively but the seasonality plot tends to increase over time, thus multiplication should be applied. When it comes to the error plot have large and small errors over time, so multiplication should be applied.

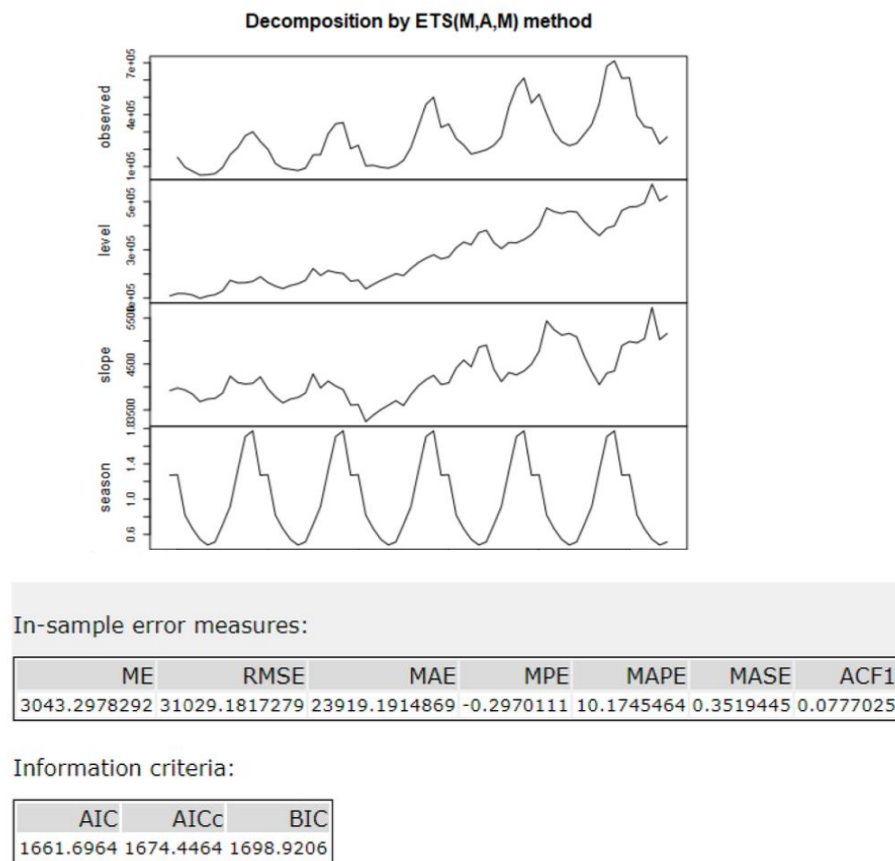
Step 3: Build your Models

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

The model terms for ETS model are ETS(MAM) and this was decided based on the previous decomposition graphs.

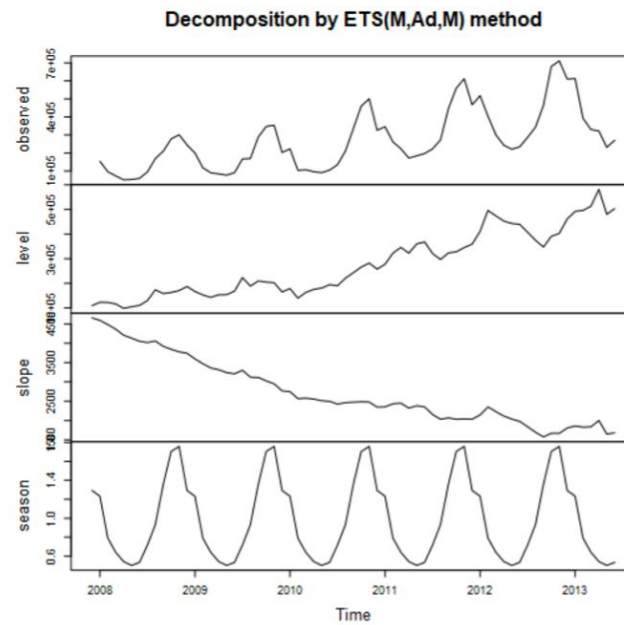
2. Describe the in-sample errors. Use at least RMSE and MASE when examining results

Non-dampened ETS model:



The AIC value is 1661.6964, RMSE (Root Mean Square Error) is 31029.1817279 and MASE (Mean Absolute Percentage Error) is 0.3519445.

Dampened ETS Model:



In-sample error measures:

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
5779.4518261	32950.2581028	25430.1562379	0.1454236	10.3589682	0.3741767	0.075948

Information criteria:

AIC	AICc	BIC
1665.5653	1680.1185	1704.9791

The AIC value is 1665.5653, RMSE is 32950.2581 and MASE is 0.37418.

Comparison between the two models:

Non Damped Model

Actual	MAM
329000	374661.17407
401000	487854.47713
553000	711534.68825

Damped Model

Actual	MAM_Dampening
329000	361677.89227
401000	472365.14677
553000	689750.36646

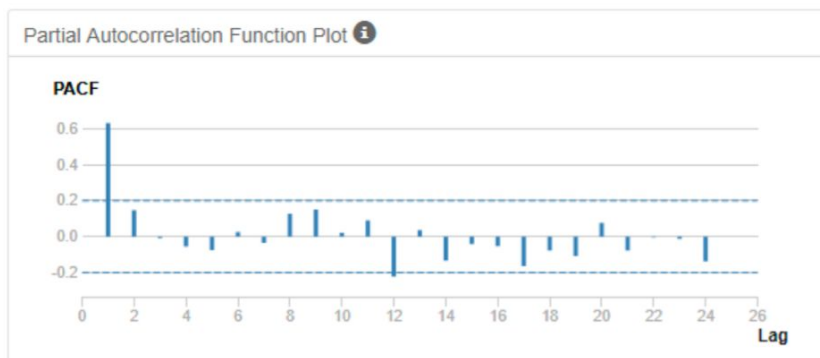
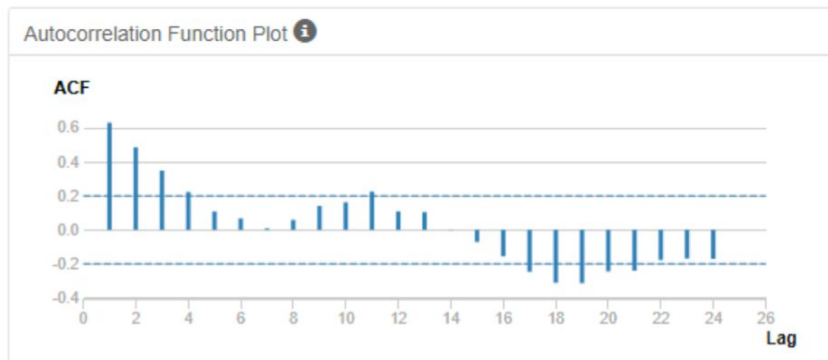
Model	ME	RMSE	MAE	MPE	MAPE	MASE	NA
MAM	-97016.78	107644.3	97016.78	-21.4021	21.4021	1.6207	NA

Model	ME	RMSE	MAE	MPE	MAPE	MASE	NA
MAM_Dampening	-80264.47	91033.86	80264.47	-17.486	17.486	1.3408	NA

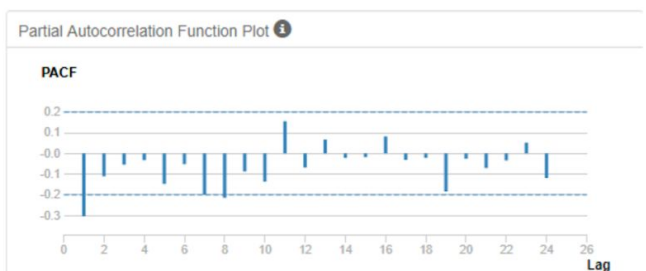
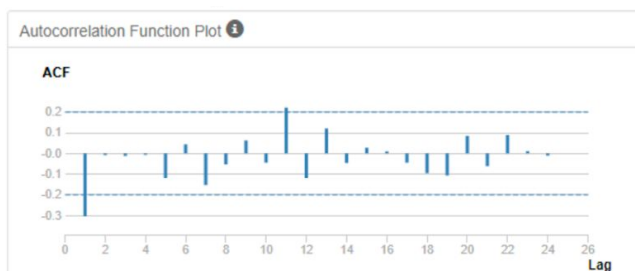
The choice went to the Non-Damped model because of its low AIC value.

- 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.

Because of the negative at lag-1 and the 12 period month, the appropriate ARIMA is (0,1,1) (0,1,0)₁. The seasonal difference is used to rationalize the time series. However, the ACF shows some correlation while the PACF doesn't show any correlation.



Another seasonal difference is taken (S first difference) to stationarize the time series again. Here, both ACF and PACF doesn't show much correlation.



- Describe the in-sample errors. Use at least RMSE and MASE when examining results

Information Criteria:

AIC	AICc	BIC
1279.7082	1279.9482	1283.6488

In-sample error measures:

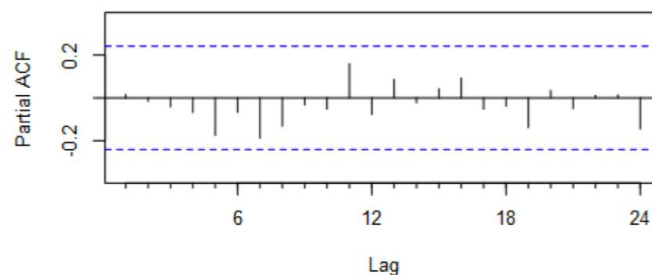
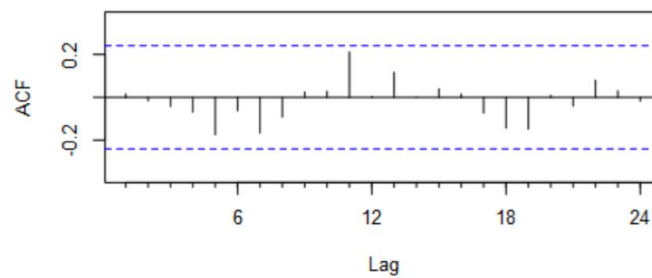
ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
-232.6526865	36493.0207463	24742.8684783	-1.7390789	9.7203024	0.364064	0.0154848

As shown above, the AIC is **1279.7082**, RMSE is **36493.0207** and **MASE is 0.3641**.

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

Both ACF and PACF doesn't shows significant correlation and no additional AR or MA terms needed.

Autocorrelation Function Plots



Step 4: Forecast

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

ETS (MAM) Model:

Actual and Forecast Values:

Actual	MAM
329000	374661.17407
401000	487854.47713
553000	711534.68825

Accuracy Measures:

Model	ME	RMSE	MAE	MPE	MAPE	MASE	NA
MAM	-97016.78	107644.3	97016.78	-21.4021	21.4021	1.6207	NA

ARIMA model:

Actual and Forecast Values:

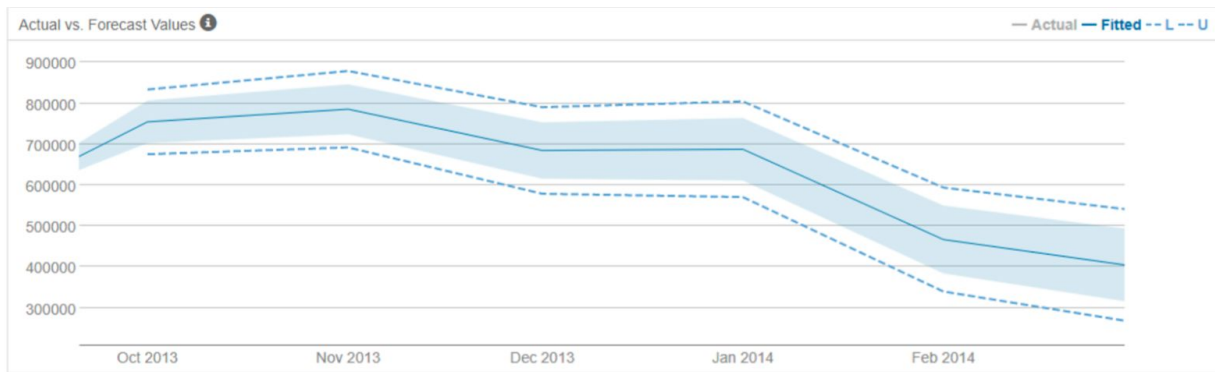
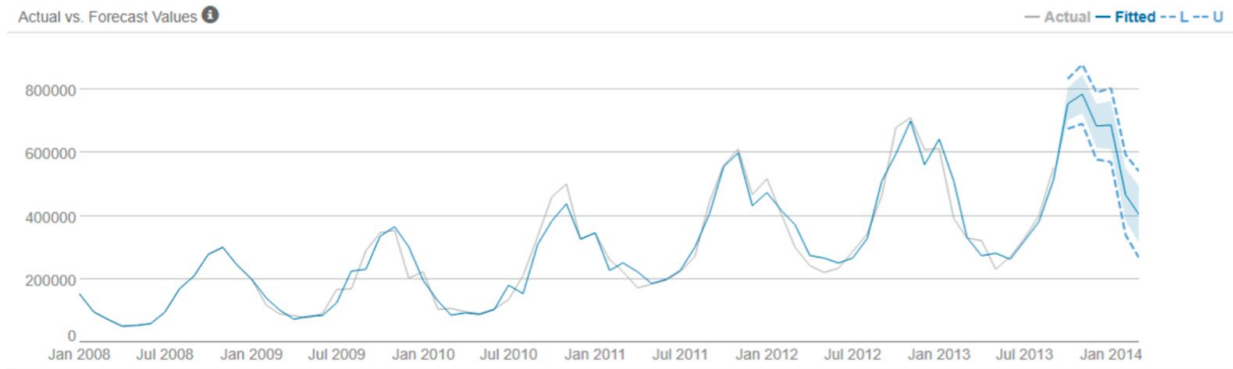
Actual	ARIMA
329000	321091.5723
401000	377091.5723
553000	498091.5723

Accuracy Measures:

Model	ME	RMSE	MAE	MPE	MAPE	MASE	NA
ARIMA	28908.43	34876.41	28908.43	6.0984	6.0984	0.4829	NA

The chosen model is ARIMA, because it had the lowest AIC(1279.7082) and also the lowest RMSE and MASE values. So ARIME is the most appropriate model for forecasting sales.

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



The forecast for the next 4 periods (Oct-13 till Jan-14) :

Month	Forecast
October'13	754854.46
November'13	785854.46
December'13	684854.46
January'14	687854.46