Project: Forecasting Sales

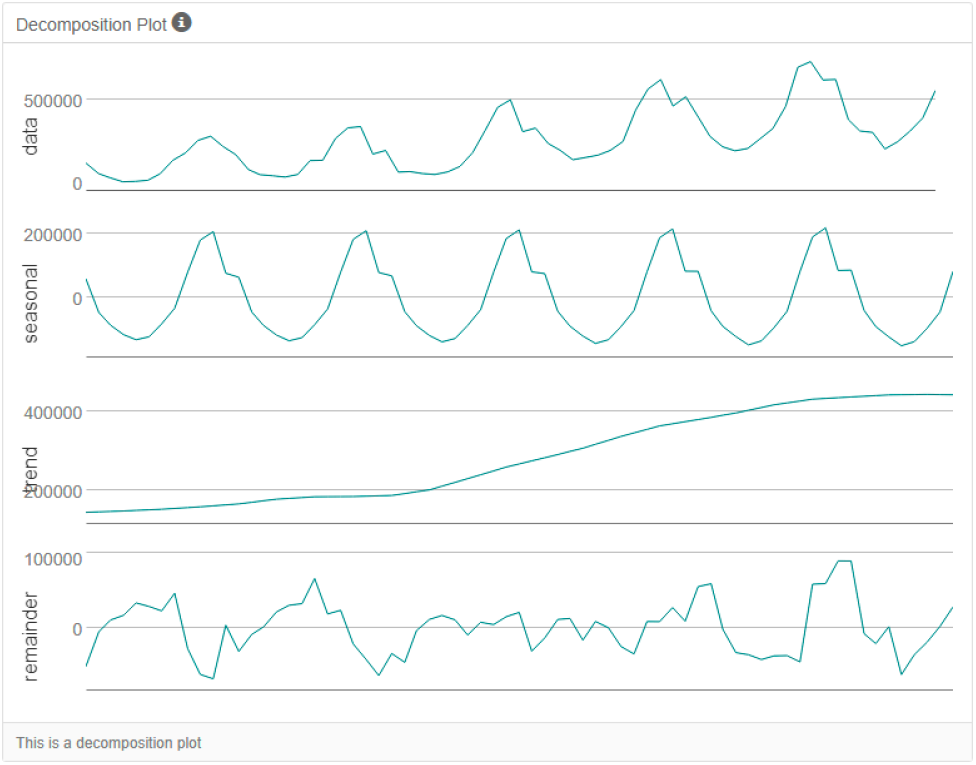
Step 1: Plan Your Analysis

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.
    a. The main four attributes of time series are:
        i. The data should be over a continuous time interval
        ii. There should be sequential measurements across that time interval
        iii. There should be equal spacing between every two consecutive measurements
        iv. Finally, each time unit within the time interval has at most one data point.
    b. The dataset "monthly-sales" meets all the time series requirements.
2. Which records should be used as the holdout sample?
    a. Since we are trying to forecast the next four months of sales, the holdout sample should be from June 2013 to September 2013.

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.

- The trend plot is linear, so we apply it additively (A).
- The fluctuations in seasonality plot tend to increase over time, so we apply it multiplicatively (M).
- The fluctuations in error plot have large and small errors over time, so we apply it multiplicatively (M).



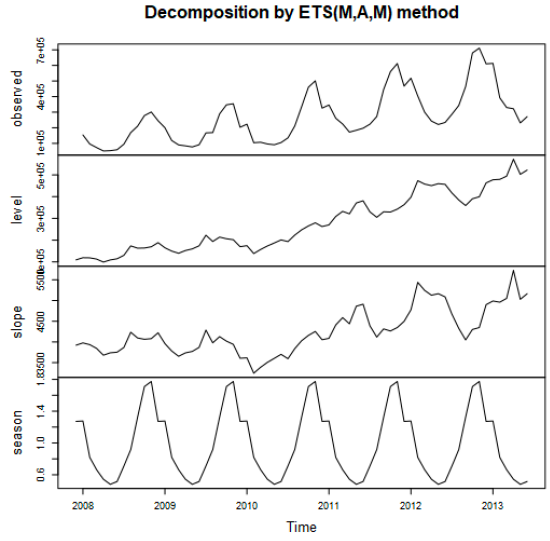
Step 3: Build your Models

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

- Based on the decomposition graph, the terms for ETS model are: ETS(MAM).
- E(rror) has large and small fluctuations, hence multiplicatively (M).
- T(rend) plot is linear, hence additively (A).
- S(easonality) plot has increasing fluctuations over time, hence multiplicatively (M).

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

- Non-Dampening ETS:



Method:
ETS(M,A,M)

In-sample error measures:

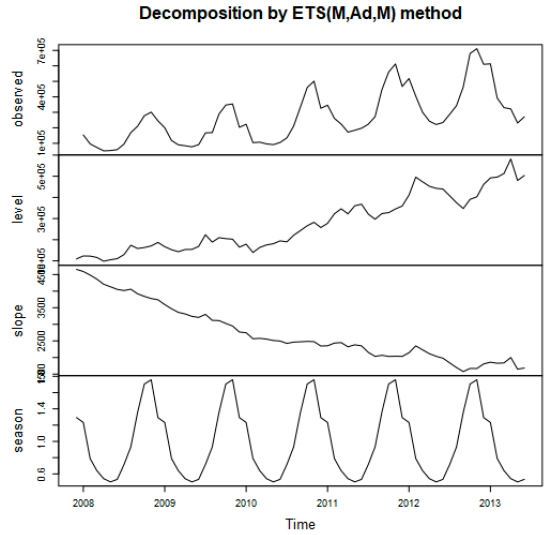
ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
3043.2978292	31029.1817279	23919.1914869	-0.2970111	10.1745464	0.3519445	0.0777025

Information criteria:

AIC	AICc	BIC
1661.6964	1674.4464	1698.9206

- The AIC value is 1661.6964
- The RMSE value is 31029.1817279
- The MASE value is 0.3519445

- Dampened ETS:



Method:
ETS(M,Ad,M)

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
  - The RMSE value is 32950.2581028
  - The MASE value is 0.3741767
- Comparing Dampened model and Non-Dampened model:

Actual and Forecast Values:

Actual	MAM
329000	374661.17407
401000	487854.47713
553000	711534.68825

Actual and Forecast Values:

Actual	MAM_Dampening
329000	361677.89227
401000	472365.14677
553000	689750.36646

Accuracy Measures:

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

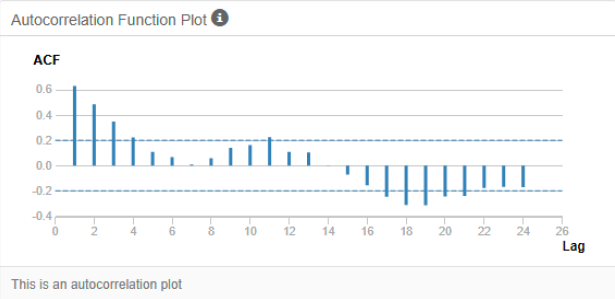
Accuracy Measures:

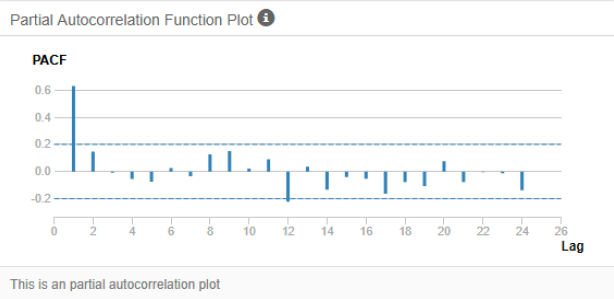
Model	ME	RMSE	MAE	MPE	MAPE	MASE	NA
MAM_Dampening	-80264.47	91033.86	80264.47	-17.486	17.486	1.3408	NA

- Picked the Non-dampened model since it has the low AIC value of 1661.6964

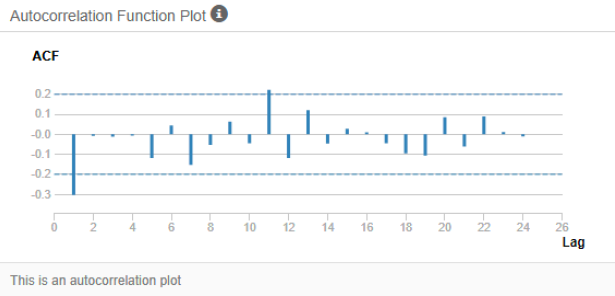
2.1 What are the model terms for ARIMA? Explain why you chose those terms. Graph the Auto-Correlation Function (ACF) and Partial Autocorrelation Function Plots (PACF) for the time series and seasonal component and use these graphs to justify choosing your model terms.

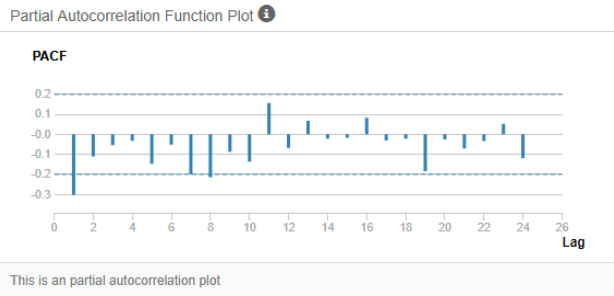
- ARIMA (0,1,1) (0,1,0)12 is used as there is a negative at lag-1 and the number of periods is 12 months.
- Seasonal difference is used to stationarize the time series. However, the ACF shows some correlation while the PACF doesn't show any correlation.





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





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

Information Criteria:

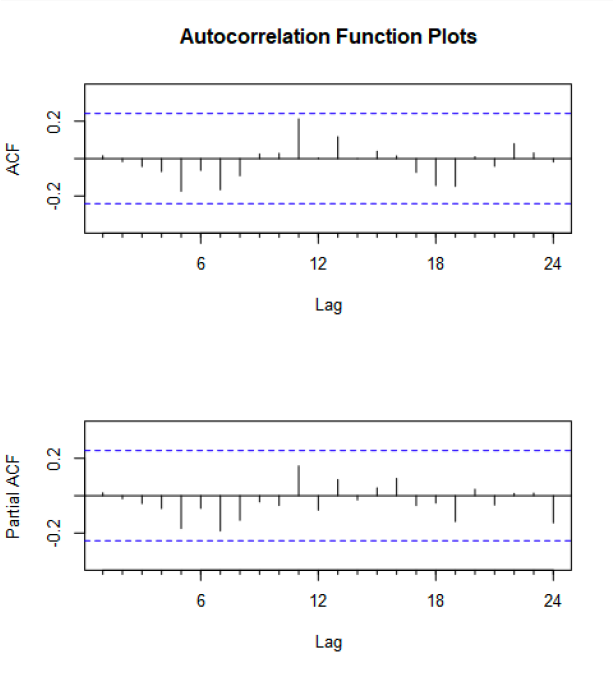
AIC	AICc	BIC
1279.7082	1279.9482	1283.6488

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

- Picked the lowest AIC value for the ARIMA model.
- The AIC value is 1279.7082
- The RMSE value is 36493.027463
- The MASE value is 0.364064

2.3 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 show much correlation.

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.

Actual and Forecast Values:

Actual	MAM
329000	374661.17407
401000	487854.47713
553000	711534.68825

Actual and Forecast Values:

Actual	ARIMA
329000	321091.5723
401000	377091.5723
553000	498091.5723

Accuracy Measures:

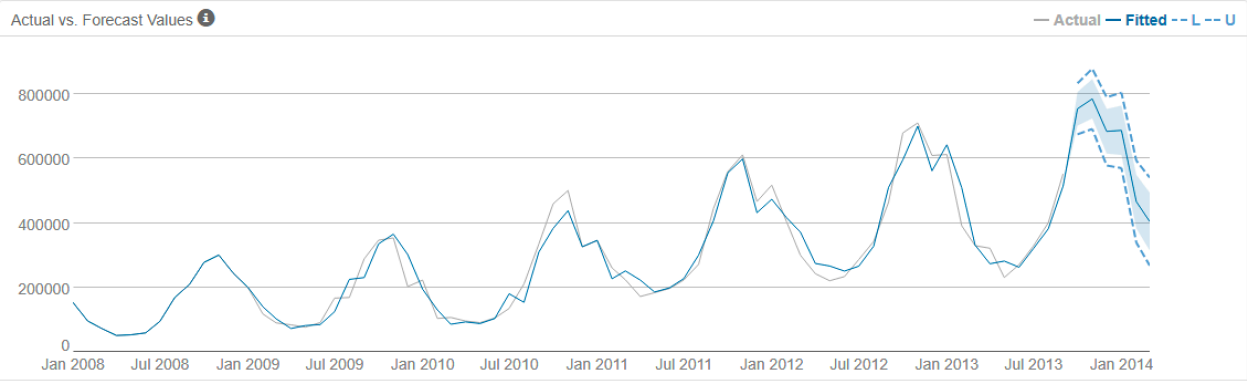
Model	ME	RMSE	MAE	MPE	MAPE	MASE	NA
MAM	-97016.78	107644.3	97016.78	-21.4021	21.4021	1.6207	NA

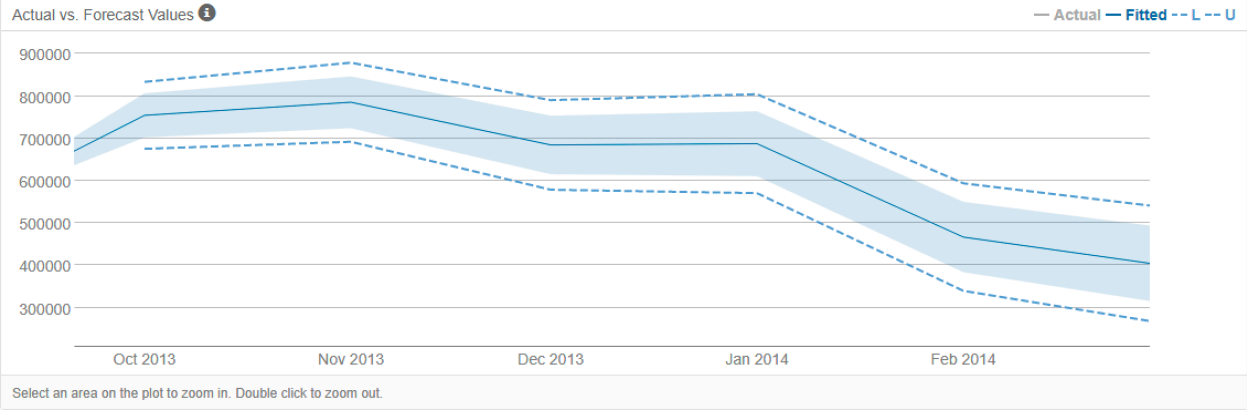
Accuracy Measures:

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

- Out of the two models (ETS and ARIMA), picked the one with the low AIC, which is ARIMA (1279.7082).
- The RMSE value (34876.41) and MASE value (0.4829) is also low compared the ETS model.

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 (October 2013 – January 2014):
  - October: 754854.46
  - November: 785854.46
  - December: 684854.46
  - January: 687854.46