<u>Project: Forecasting Sales</u>

Complete each section. When you are ready, save your file as a PDF document and submit it here: https://classroom.udacity.com/nanodegrees/nd008/parts/edd0e8e8-158f-4044-9468-3e08fd08cbf8/project

Step 1: Plan Your Analysis

Look at your data set and determine whether the data is appropriate to use time series models. Determine which records should be held for validation later on (250 word limit).

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.

    Yes, the dataset monthly-sales.xlsx meets the below criteria
    Time series data key characteristics:
    1. It's over a continuous time interval
    2. There are sequential measurements across that interval
    3. There is equal spacing between every two consecutive measurements
    4. Each time unit within the time interval has at most one data point
2. Which records should be used as the holdout sample?

    Since we are forecasting 4 months into the future, we should us a 4 month holdout sample.

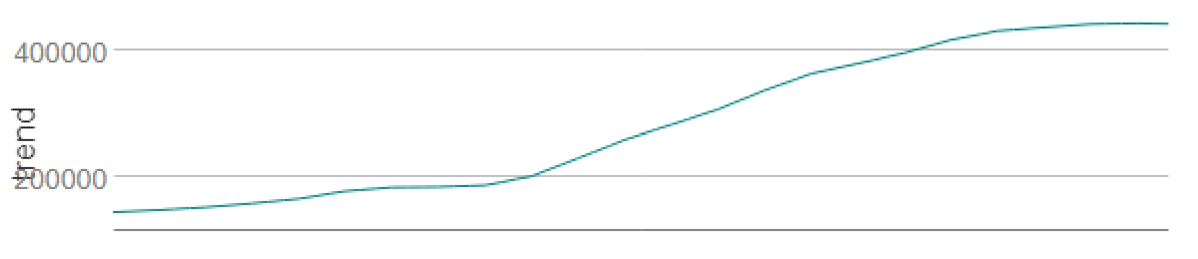
Step 2: Determine Trend, Seasonal, and Error components

Graph the data set and decompose the time series into its three main components: trend, seasonality, and error. *(250 word limit)*

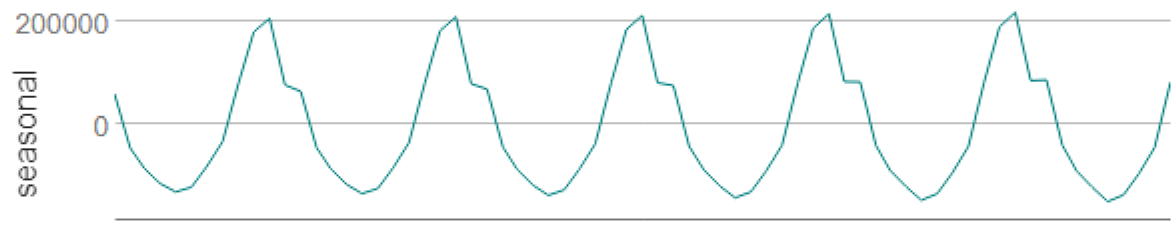
Answer this question:

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.

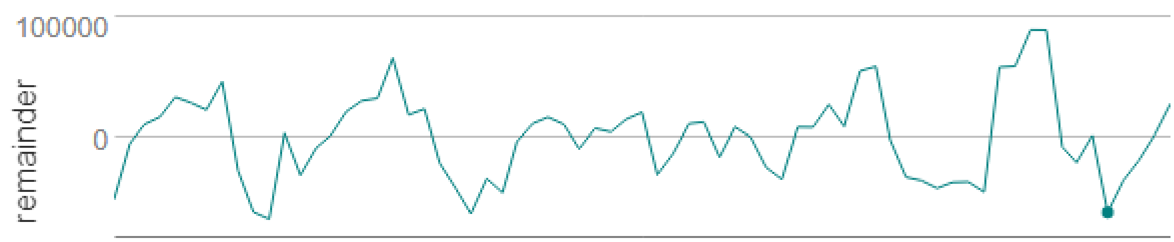
    Trend: Graph is linear. Additively.



Seasonality: Seasonality increases in volume each seasonal period. Multiplicative.



Error: Graph changes variance as the time series moves along. Multiplicative.



Step 3: Build your Models

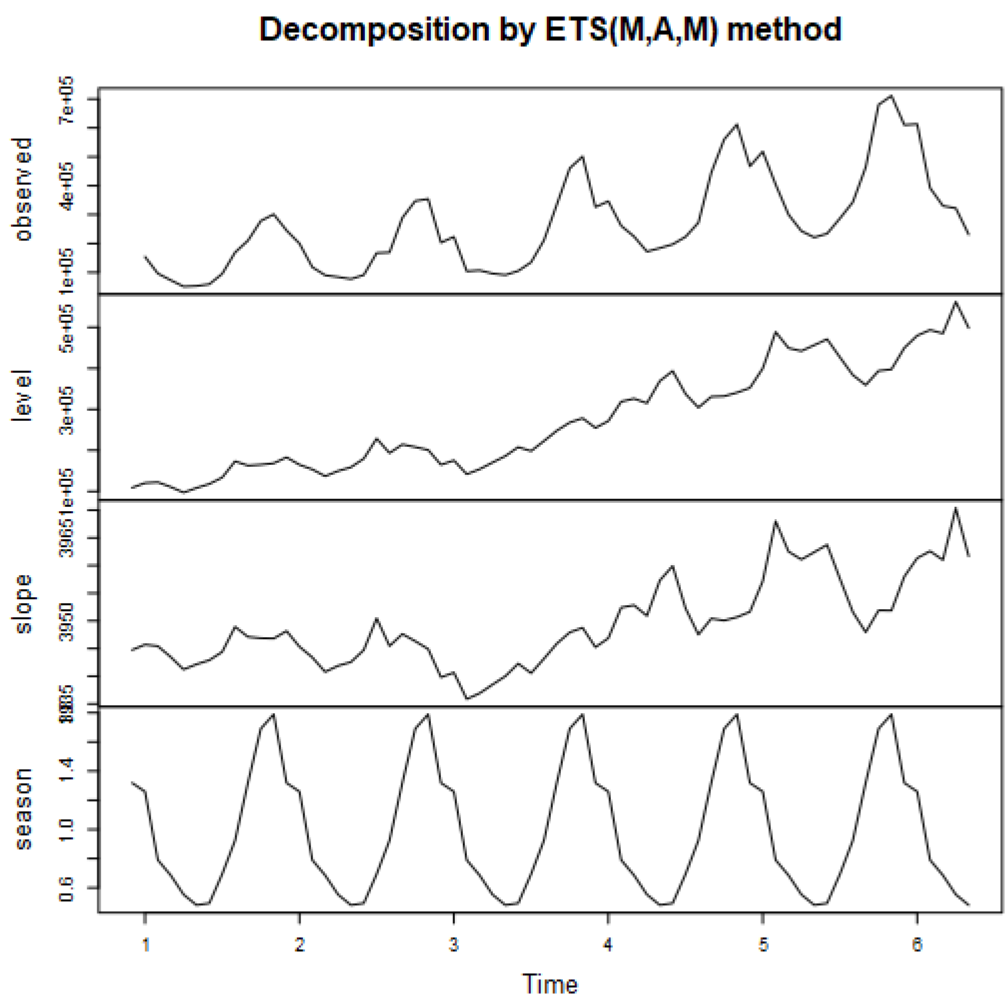
Analyze your graphs and determine the appropriate measurements to apply to your ARIMA and ETS models and describe the errors for both models. (500 word limit)

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

        I will compare values on ETS with and without dampening and then compare actual vs forecasted future values.

ETS without dampening: RMSE is 32992.73, MASE is .372685 and AIC is 1639.74



Summary of Time Series Exponential Smoothing Model MAM

Method:
ETS(M,A,M)

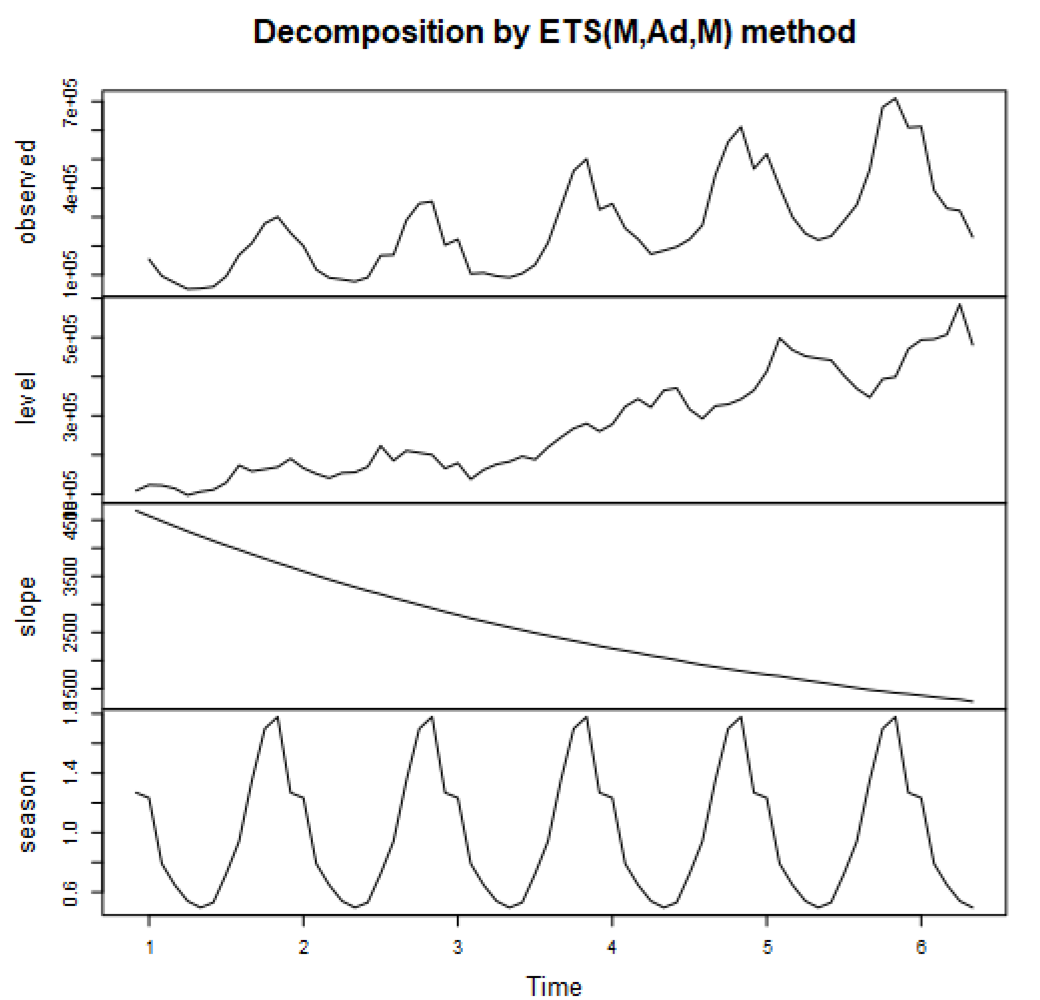
In-sample error measures:

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

ETS with dampening: RMSE is 33153.53, MASE is .3675478 and AIC is 1639.47



Summary of Time Series Exponential Smoothing Model MAM_Dampened

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

ETS without dampering:

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

ETS with dampering

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

:

Based on these results I choose ETS with dampening as the best model, it has a higher RMSE, but lower MASE and AIC. Also, looking at the actual vs forecasted values, the dampened model has a more accurate prediction and lower MASE/RMSE.

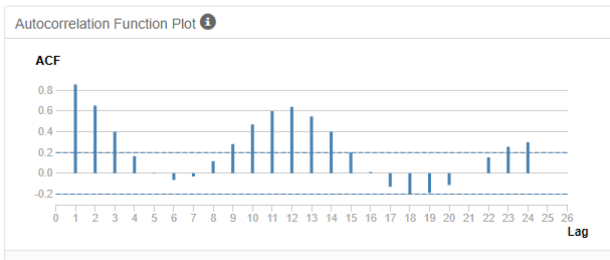
ETS without dampening: RMSE is 32992.73, MASE is .372685 and AIC is 1639.74

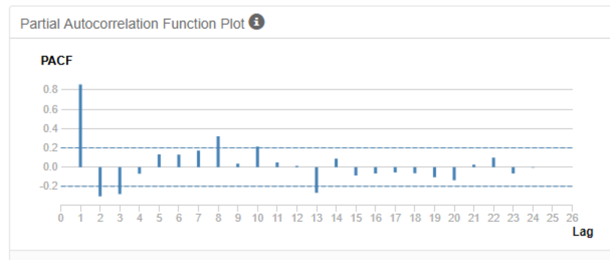
ETS with dampening: RMSE is 33153.53, MASE is .3675478 and AIC is 1639.47

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.

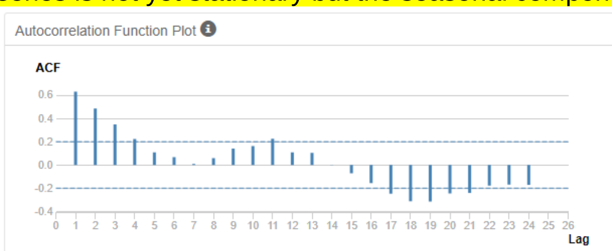
ARIMA stands for Auto Regressive Integrated Moving Average model. It uses either a seasonal or a non-seasonal model; we need to look at the ACF and PACF graphs top determine which model to use in our analysis.

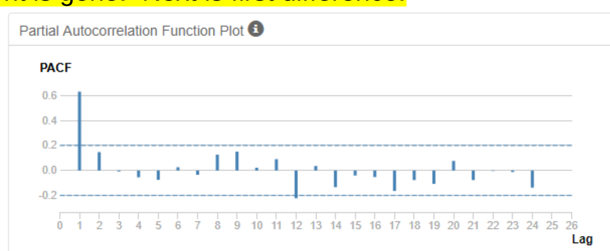
1): This is the ACF/PACF plots before any differencing, series is not stationary. Model is seasonal. Next step is to apply seasonal differencing.



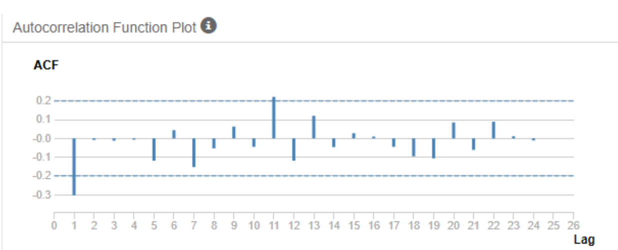


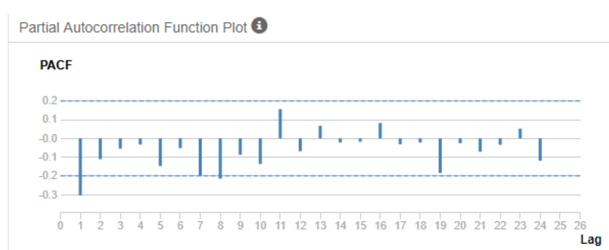
2): Taking the seasonal difference using [Monthly Sales]-[Row-12: Monthly Sales], we see that series is not yet stationary but the seasonal component is gone. Next is first difference.





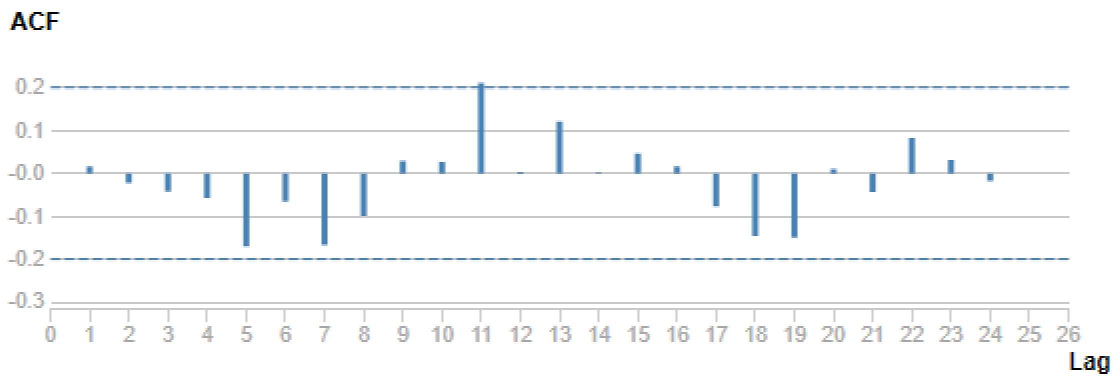
3): Taking the first difference using [Seasonal Difference]-[Row-1:Seasonal Difference], we see that series is now stationary excepts for lag 1. Next is to add MA term in ARIMA tool.

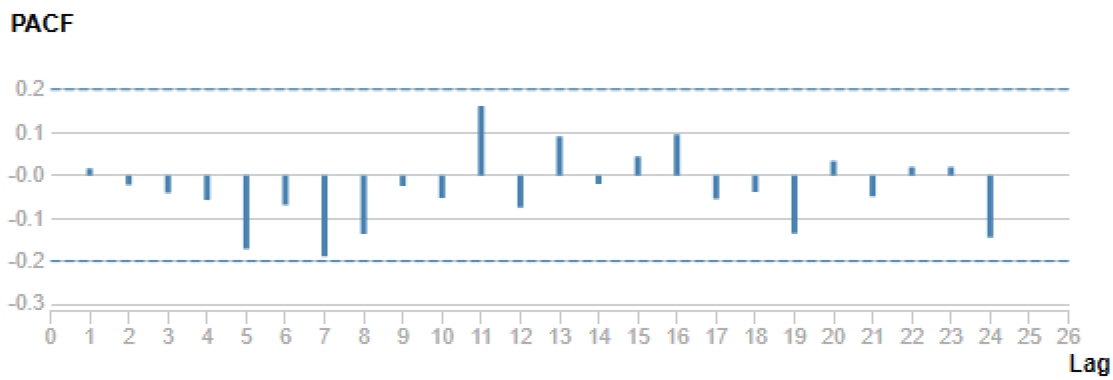




4): ACF/PACF plots after MA term is added, after taking all the steps above the series is now stationary. Seasonal ARIMA models are denoted (p,d,q)(P,D,Q)m
From the graphs we need to use a ARIMA(0,1,1)(0,1,0)12 model.

Autocorrelation Function Plot





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

Summary of ARIMA Model ARIMA

Method: ARIMA(0,1,1)(0,1,0)[12]

Call:
Arima(Monthly.Sales, order = c(0, 1, 1), seasonal = list(order = c(0, 1, 0), period = 12))

Coefficients:

	ma1
Value	-0.378032
Std Err	0.146228

sigma^2 estimated as 1722385234.94439: log likelihood = -626.29834

Information Criteria:

AIC	AICc	BIC
1256.5967	1256.8416	1260.4992

In-sample error measures:

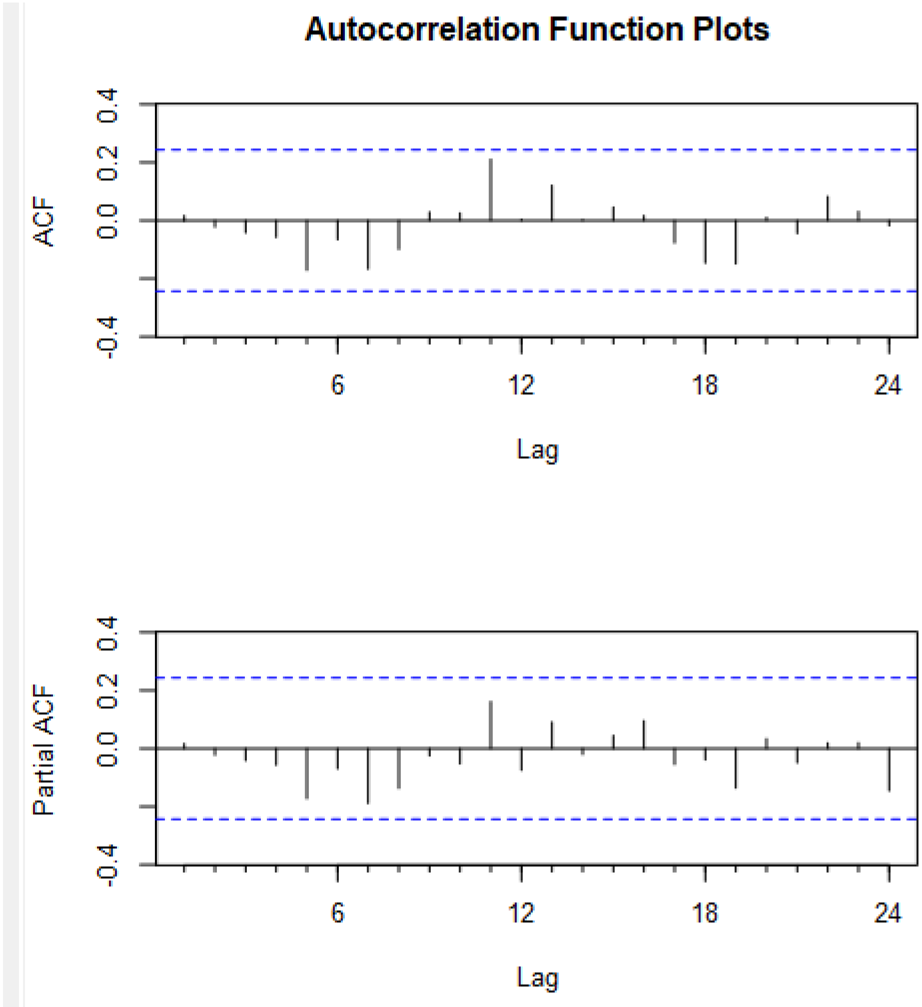
ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
-356.2665104	36761.5281724	24993.041976	-1.8021372	9.824411	0.3646109	0.0164145

Ljung-Box test of the model residuals:
Chi-squared = 16.4458, df = 23, p-value = 0.83553

RMSE is 36761.528, MASE is 0.365 and AIC is 1256.597

RMSE=Root Mean Squared Error, is the standard deviation of the differences between an actual and a forecasted value.
MASE=Mean Absolute Scaled Error, a low value represents a good model.
AIC=Akaike information criterion, a low value represents a good model.

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



No more correlation

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)

Answer these questions.

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

ETS Dampened:

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

ARIMA:

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

I Choose the ARIMA model: The projected values are close to actual values with the ARIMA model.

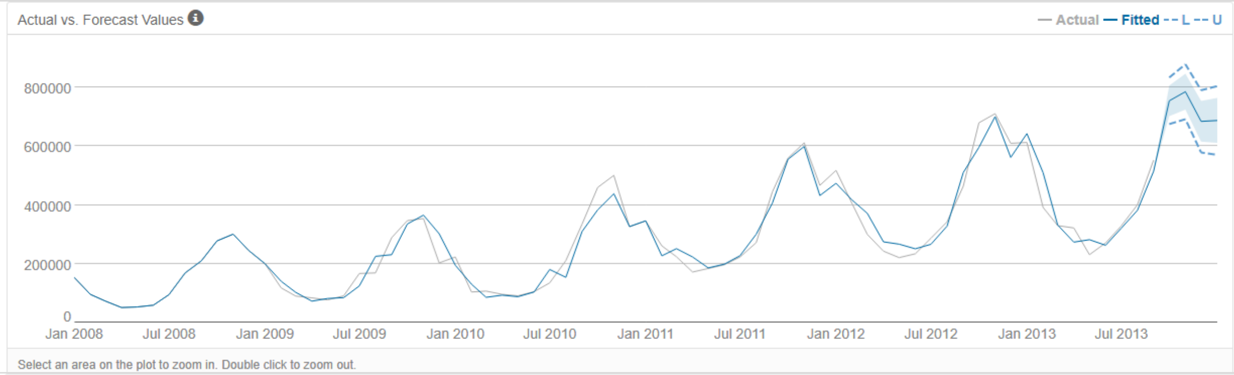
Also, RMSE & MASE is lower than in ETS model:
RMSE=33999.79 vs 60176.47 & MASE=0.04532 vs 0.8116

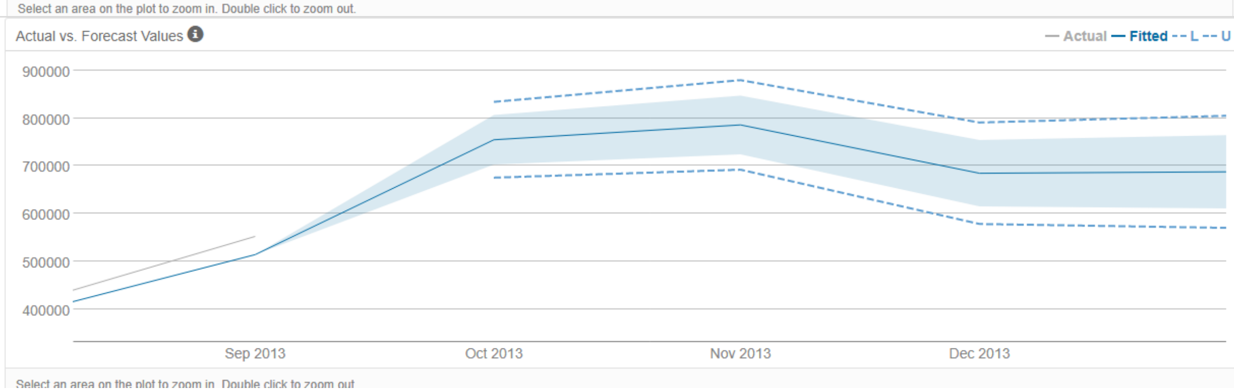
2. What is the forecast for the next four periods? Graph the results using 95% and 80% confidence intervals.

ARIMA Forecast:

7 of 7 Fields | Cell Viewer | 4 records displayed, 1743 bytes

Record #	Period	Sub_Period	forecast	forecast_high_95	forecast_high_80	forecast_low_80	forecast_low_95
1	2013	10	754854.460048	834046.21595	806635.165997	703073.754099	675662.704146
2	2013	11	785854.460048	879377.753117	847006.054462	724702.865635	692331.166979
3	2013	12	684854.460048	790787.828211	754120.566407	615588.35369	578921.091886
4	2014	1	687854.460048	804889.286634	764379.419903	611329.500193	570819.633462





Before you Submit

Please check your answers against the requirements of the project dictated by the rubric here. Reviewers will use this rubric to grade your project.