Project: Forecasting Sales

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.

The dataset meets the criteria of a time series dataset

1. It’s over a continuous time interval from Jan, 2008 to Sept, 2013.
2. Sequential measurements of monthly sales are recorded across the interval.
3. Consecutive measurements are taken with an equal spacing of one month.
4. Each time unit (a month) only has one data point (the monthly sale)
5. Which records should be used as the holdout sample?

Since we are going to predict 4 month sales. We can hold the last 4 month as our holdout samples, so 2013-06, -07, -08, -09.

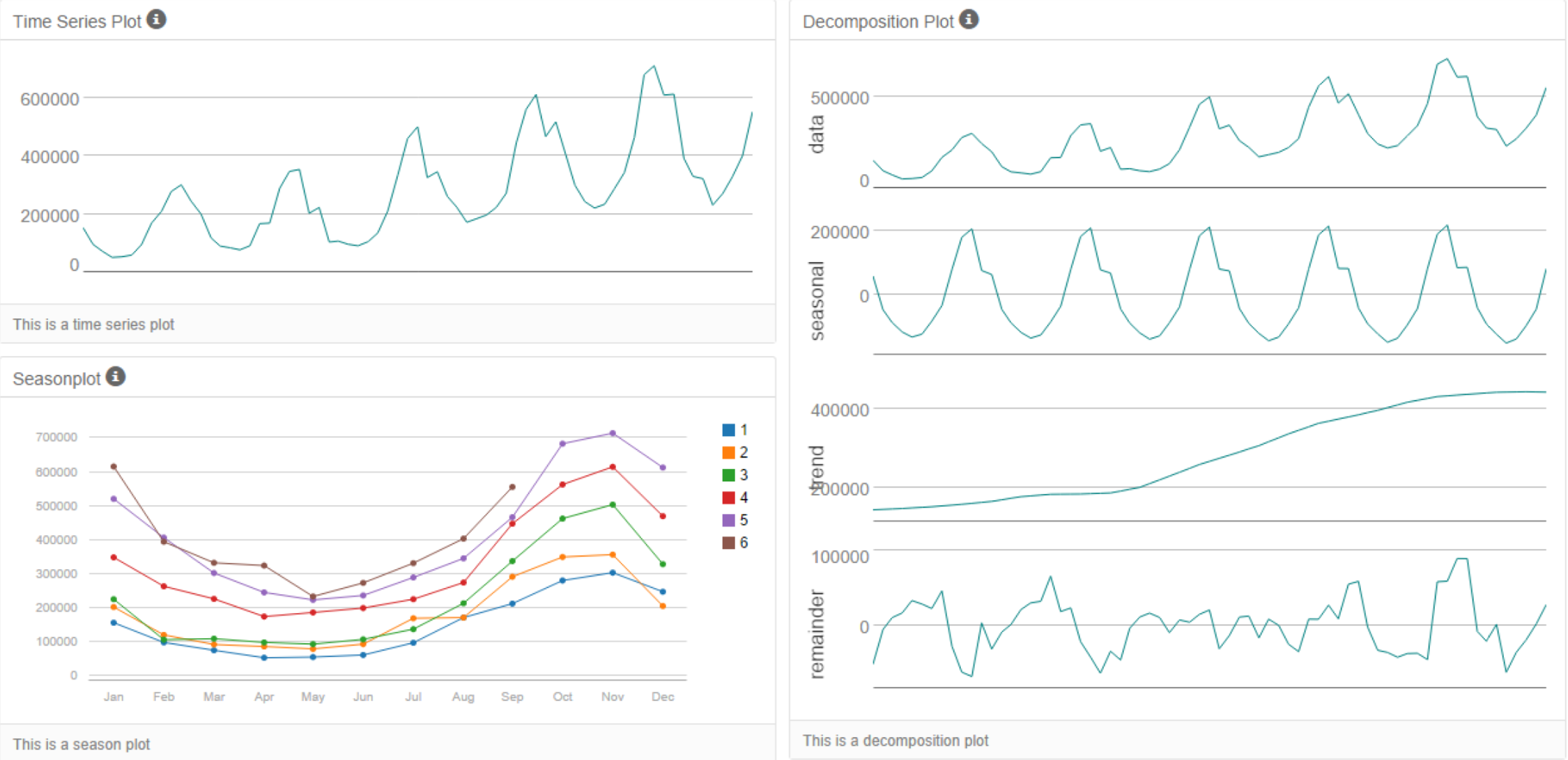
## Step 2: Determine Trend, Seasonal, and Error components

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.

*The time series has a upward linear trend, seasonality with slightly increasing magnitude, and inconsistenly fluctuating error.*

**

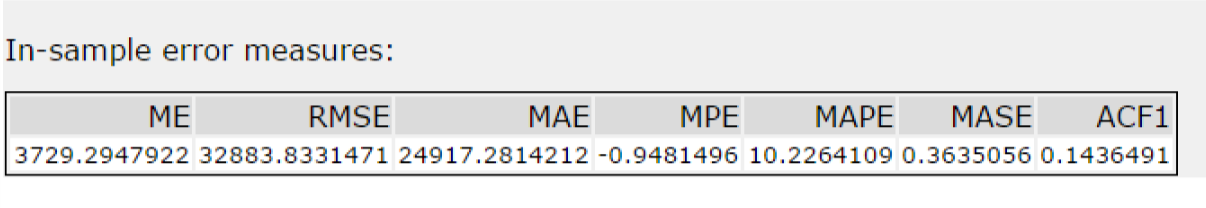
Step 3: Build your Models

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

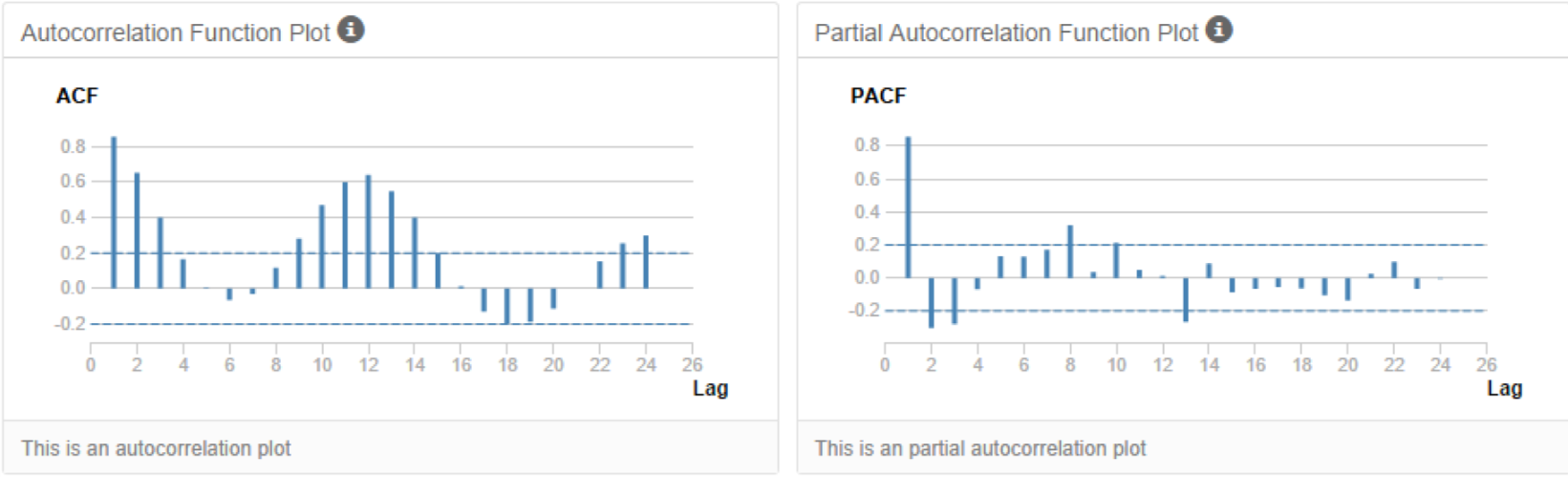
We shall use ETA(MAM) model, i.e., additive for trend since it exhibits linear behavior, multiplicative for seasonality and error since they change in magnitude. In-sample error measures are shown below:



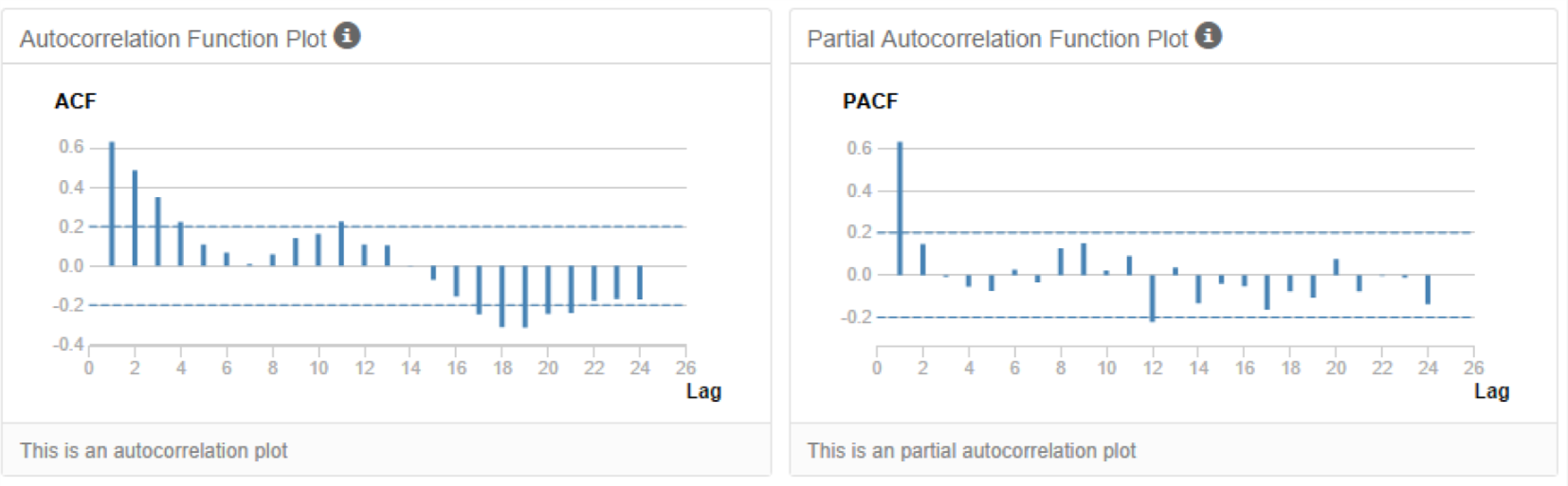
RMSE has the same order of magnitude as the data (~1e5), and MASE is significantly less than 1, this supports our choice of the ETS model.

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.
   1. Describe the in-sample errors. Use at least RMSE and MASE when examining results
   2. Regraph ACF and PACF for both the Time Series and Seasonal Difference and include these graphs in your answer.

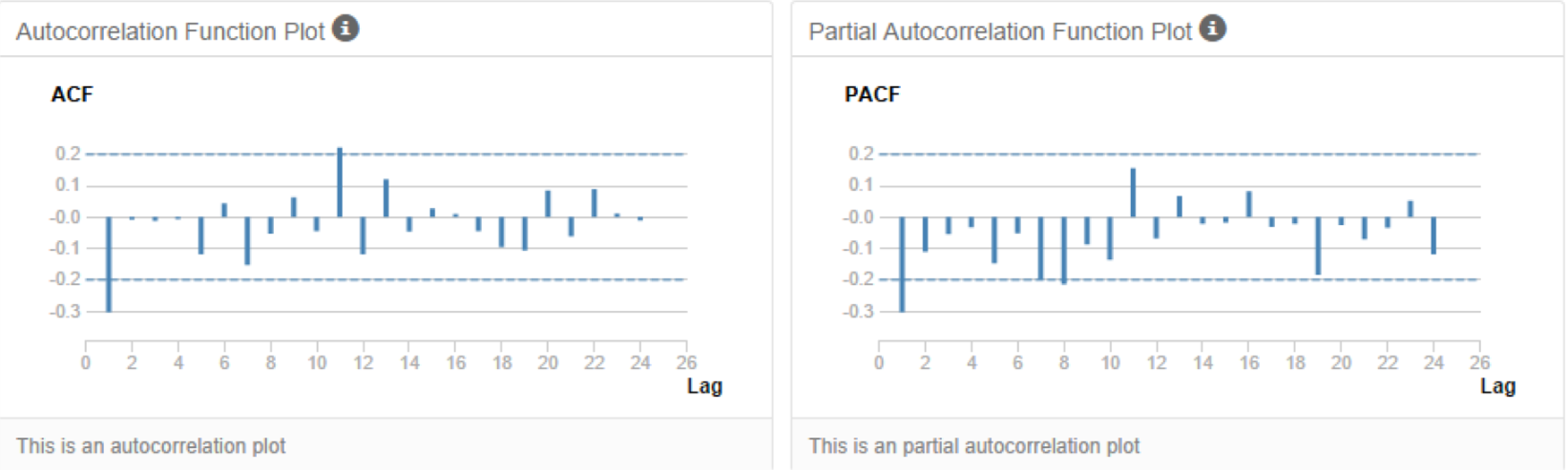
ACF and PACF for original time series: we shall consider the seasonal effect.



ACF and PACF after seasonal differencing:

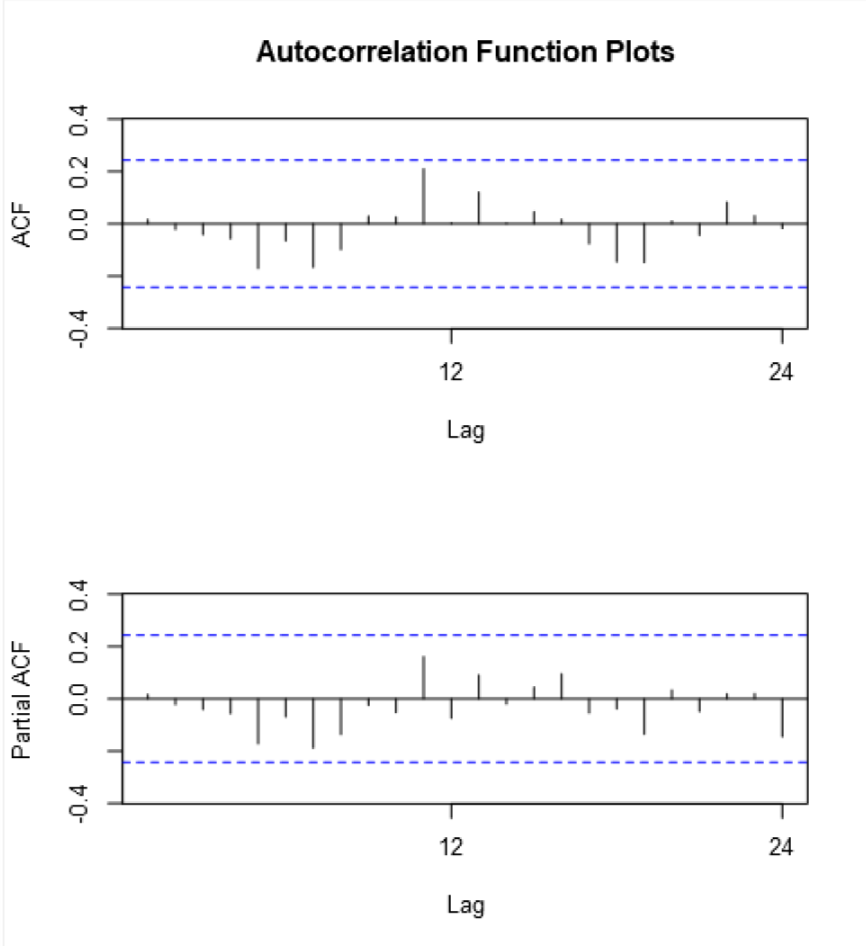


ACF and PACF after subtracting seasonality and taking the first difference:

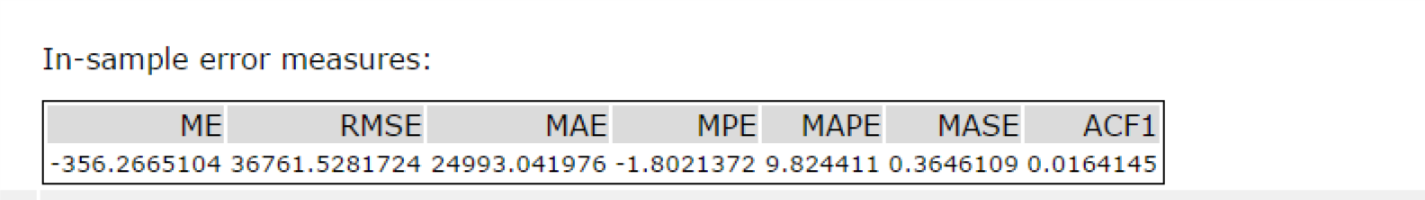


In summary, since ACF lag-1 is negative, ACF drops suddenly while PACF drops more gradually, we shall use MA. So ARIMA(p=0, d=1, q=1)(P=0, D=1, Q=0).

ACF and PACF after applying the ARIMA model:



In-sample error measures are given bellow:



RMSE has the same order of magnitude as the time series (~1e5), MASE (0.3646) is significantly less than 1. So our model is well-established.

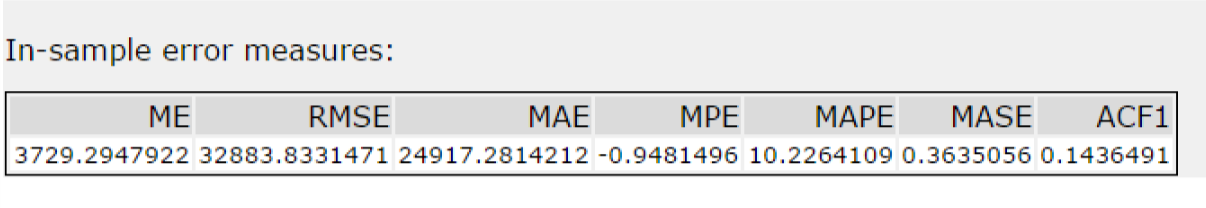
## Step 4: Forecast

*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)*

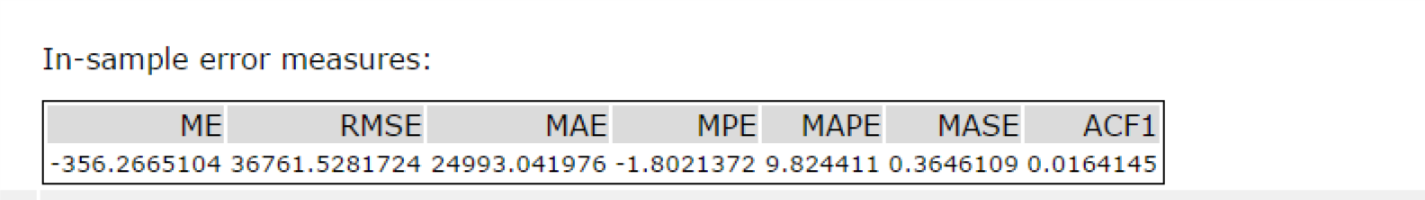
*Answer these questions.*

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

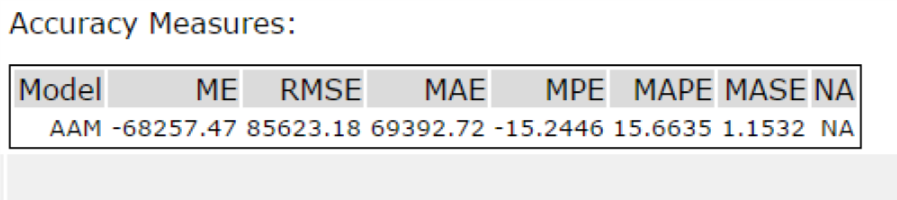
In-sample errors for ETS:



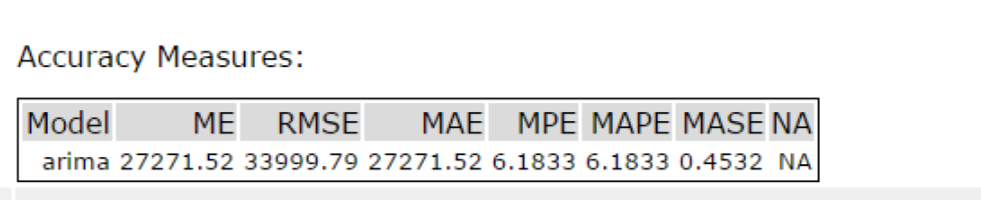
In-sample errors for ARIMA:



Forecast error measurements for ETS:

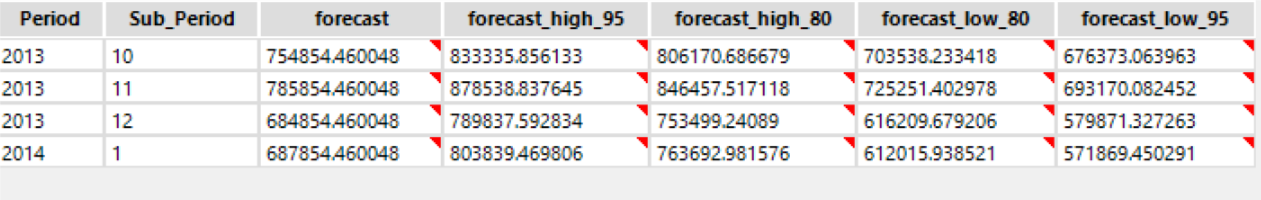


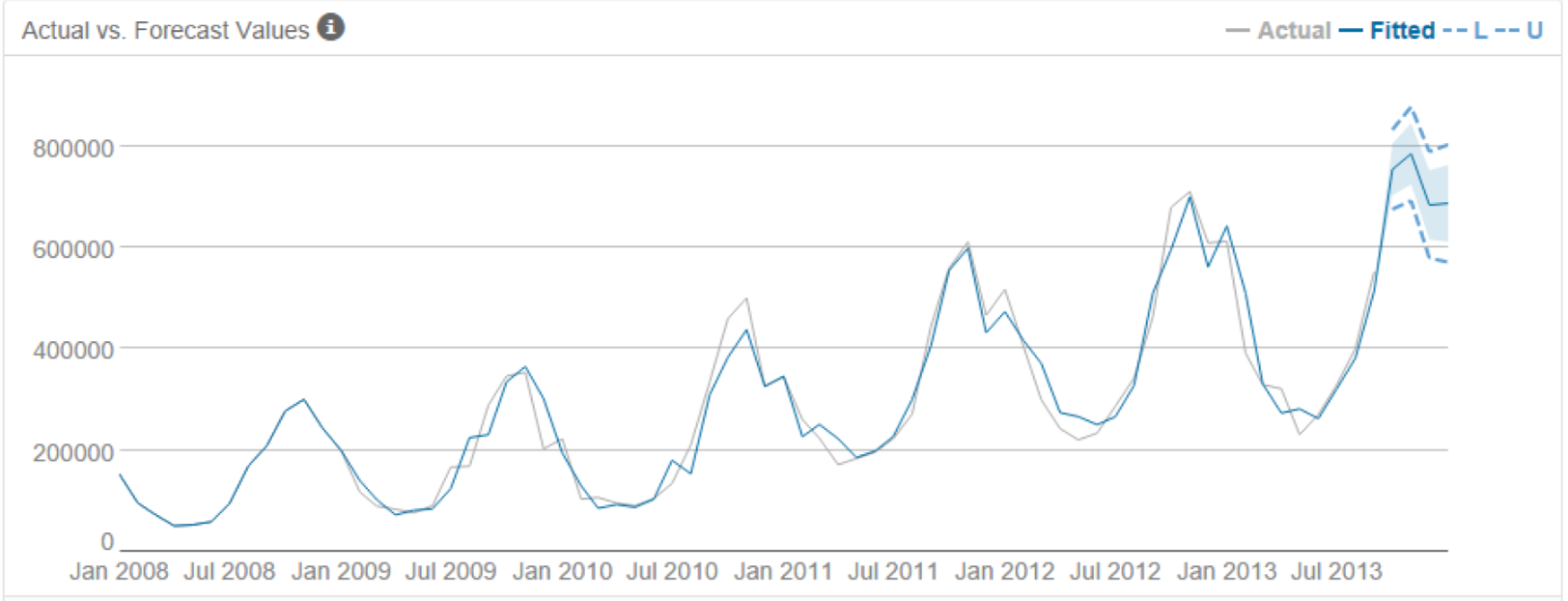
Forecast error measurements for ARIMA:



So although the ETS model has slightly lower in-sample MASE (0.3635 vs 0.3646), the ARIMA achieves much lower forecast MASE than the ETS model (0.4532 vs 1.1532). So we shall use the ARIMA model.

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





Before you Submit

Please check your answers against the requirements of the project dictated by the [rubric](https://review.udacity.com/#!/rubrics/302/view) here. Reviewers will use this rubric to grade your project.