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:

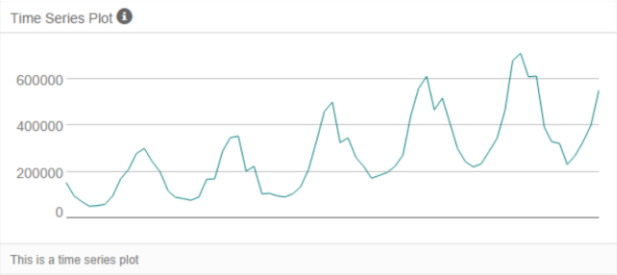
1. Does the dataset meet the criteria of a time series dataset? Make sure to explore all four key characteristics of a time series data.
   The dataset meets the criteria of a time series dataset
   a. It's over a continuous time interval from Jan, 2008 to Sept, 2013.
   b. Sequential measurements of monthly sales are recorded across the interval.
   c. Consecutive measurements are taken with an equal spacing of one month.
   d. Each time unit (a month) only has one data point (the monthly sale)
2. 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.

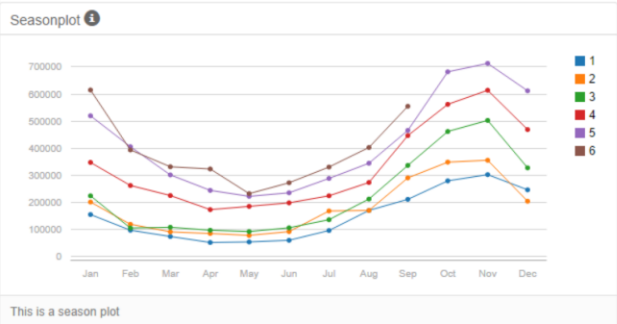
Step 2: Determine Trend, Seasonal, and Error components

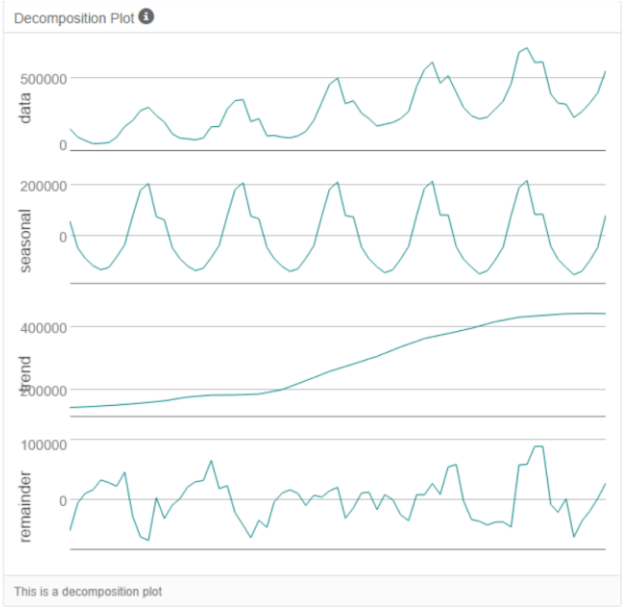
Graph the data set and decompose the time series into its three main components: trend, seasonality, and error. *(250 word limit)*

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 linear trend, slightly increasing seasonality, and multiplicative 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.
    a. Describe the in-sample errors. Use at least RMSE and MASE when examining results

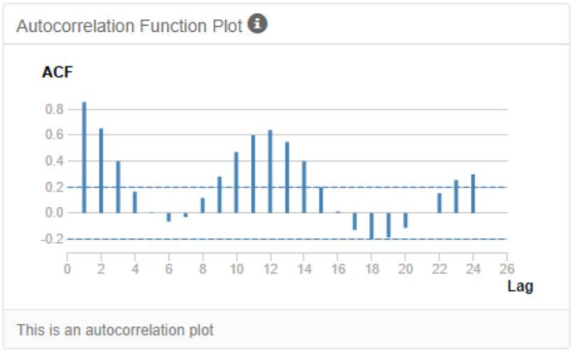
    As described earlier and shown in the graph, we shall choose multiplicative error, additive trend, multiplicative seasonal, and after trying, no trend dampening. In-sample error measures are shown below:

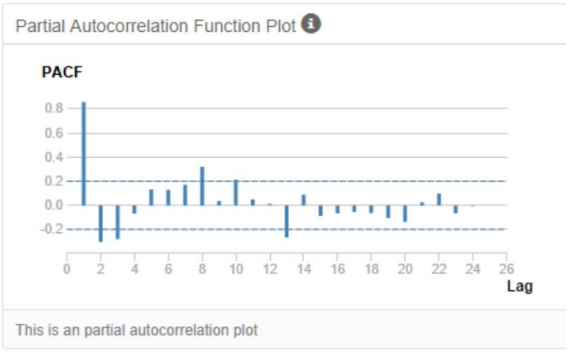
In-sample error measures:

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
3729.2947922	32883.8331471	24917.2814212	-0.9481496	10.2264109	0.3635056	0.1436491

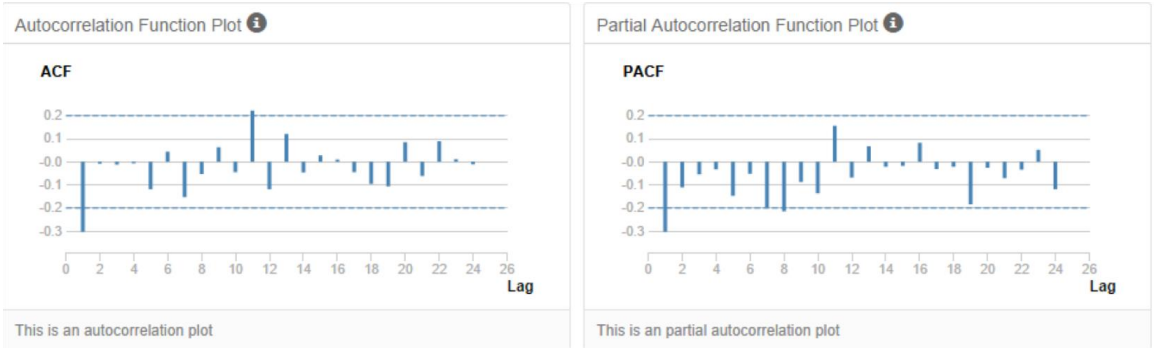
2. What are the model terms for ARIMA? Explain why you chose those terms. Graph the Auto-Correlation Function (ACF) and Partial Autocorrelation Function Plots (PACF) for the time series and seasonal component and use these graphs to justify choosing your model terms.
    a. Describe the in-sample errors.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

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





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=1).
In-sample error measures are given bellow:

In-sample error measures:

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
-358.1274828	36758.4027043	24996.5435416	-1.800917	9.8272386	0.3646619	0.0166958

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.

Forecast error measurements for ETS:

Accuracy Measures:

Model	ME	RMSE	MAE	MPE	MAPE	MASE	NA
AAM	-68257.47	85623.18	69392.72	-15.2446	15.6635	1.1532	NA

Forecast error measurements for ARIMA:

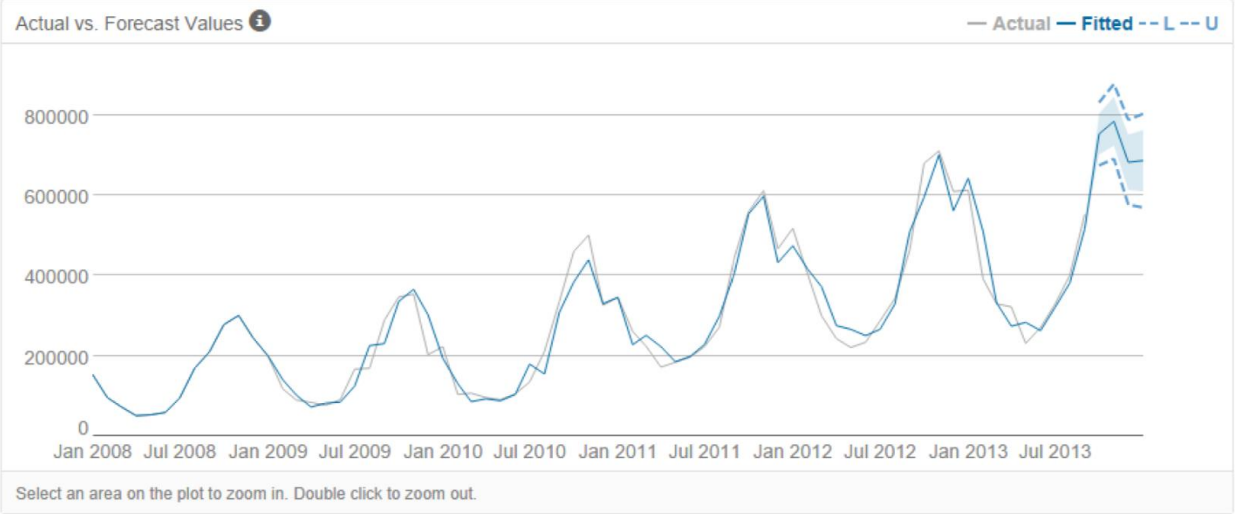
Accuracy Measures:

Model	ME	RMSE	MAE	MPE	MAPE	MASE	NA
arima	27184.45	34010.92	27184.45	6.1547	6.1547	0.4518	NA

So although the ETS model has slightly lower in-sample errors, the ARIMA achieves much lower forecast errors than the ETS model. So we shall use the ARIMA model.

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

Period	Sub_Period	forecast	forecast_high_95	forecast_high_80	forecast_low_80	forecast_low_95
2013	10	753163.3696	831632.418741	804471.522998	701855.216202	674694.32046
2013	11	784679.954091	877726.277336	845519.6747	723840.233482	691633.630845
2013	12	682563.879889	788194.580284	751632.082432	613495.677346	576933.179494
2014	1	686742.165916	803609.917937	763157.881585	610326.450247	569874.413895



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.