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).

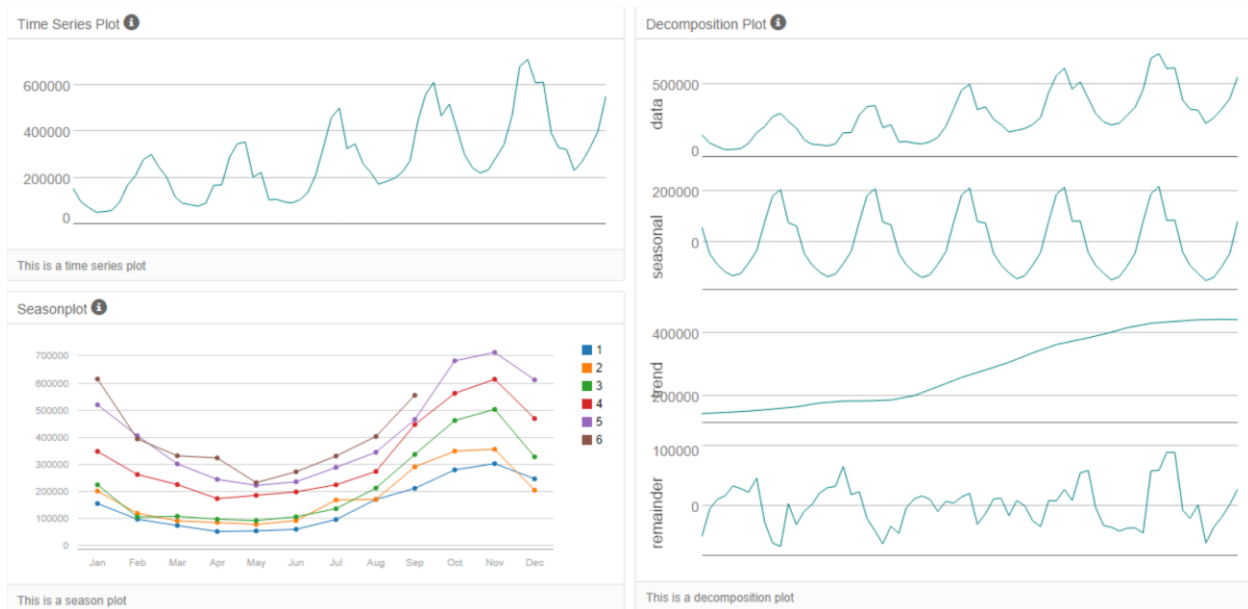*Answer the following questions to help you plan out your analysis:*
1. Does the dataset meet the criteria of a time series dataset? Make sure to explore all four key characteristics of a time series data.
   The dataset meets the criteria of a time series dataset
   a. It's over a continuous time interval 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 upward linear trend, seasonality with slightly increasing magnitude, and inconsistenly fluctuating error.

Time Series Plot ⓘ

This is a time series plot

Seasonplot ⓘ

This is a season plot

Decomposition Plot ⓘ

This is a decomposition plot

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
   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:

   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

   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.

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.
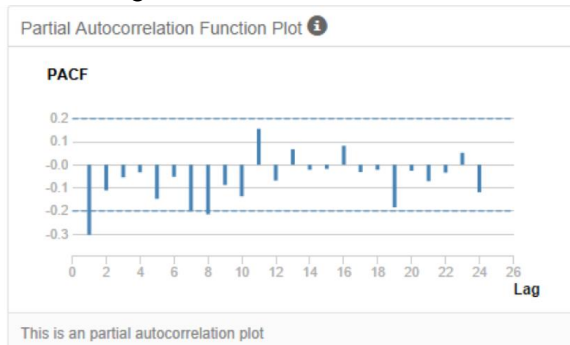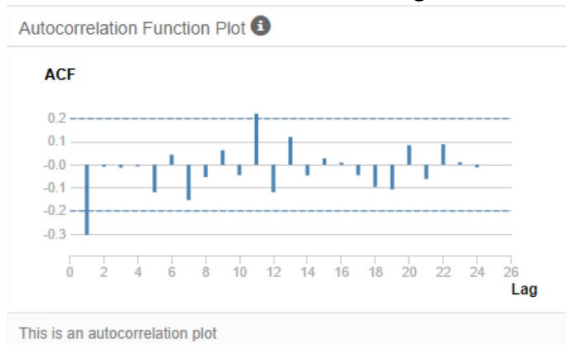
ACF and PACF for original time series: we shall consider the seasonal effect.
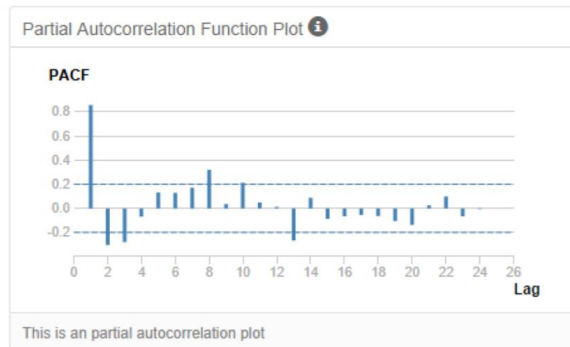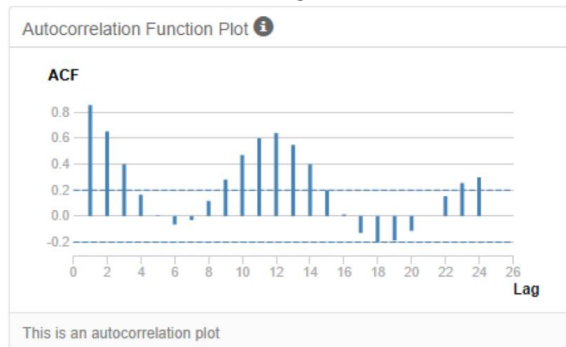
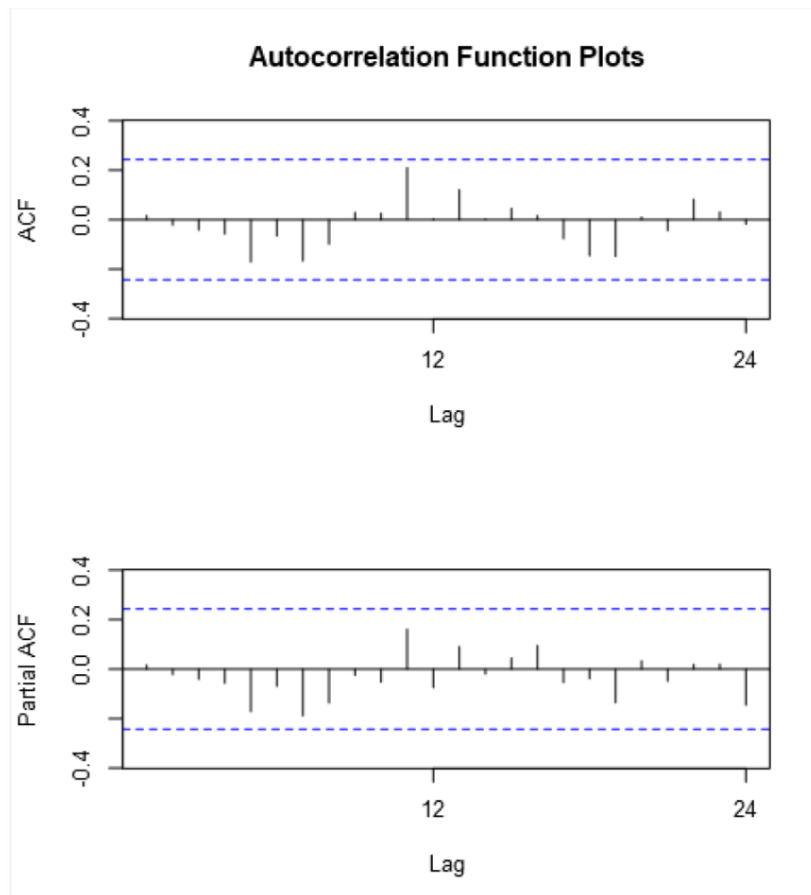ACF and PACF after seasonal differencing:
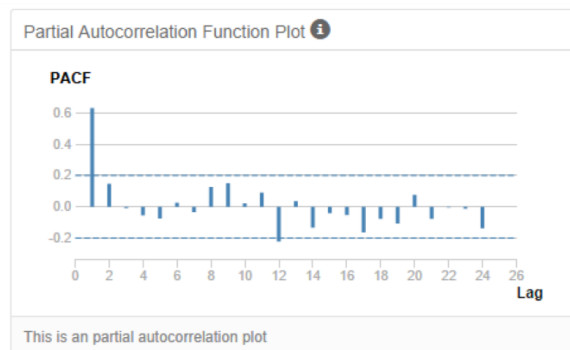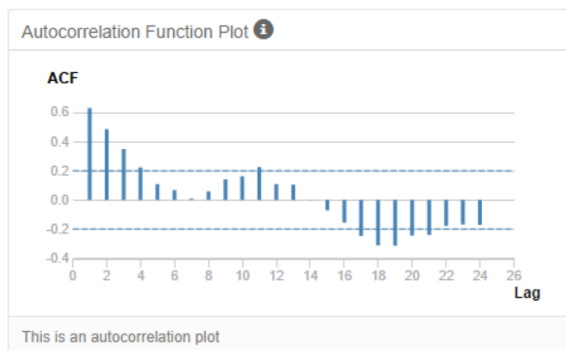
ACF and PACF after subtracting seasonality and taking the first difference:

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:

Autocorrelation Function Plots

In-sample error measures are given bellow:

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

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 error measures:

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

In-sample errors for ARIMA:

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

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	27271.52	33999.79	27271.52	6.1833	6.1833	0.4532	NA

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.

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	754854.460048	833335.856133	806170.686679	703538.233418	676373.063963
2013	11	785854.460048	878538.837645	846457.517118	725251.402978	693170.082452
2013	12	684854.460048	789837.592834	753499.24089	616209.679206	579871.327263
2014	1	687854.460048	803839.469806	763692.981576	612015.938521	571869.450291

Actual vs. Forecast Values ⓘ — Actual — Fitted -- L -- U
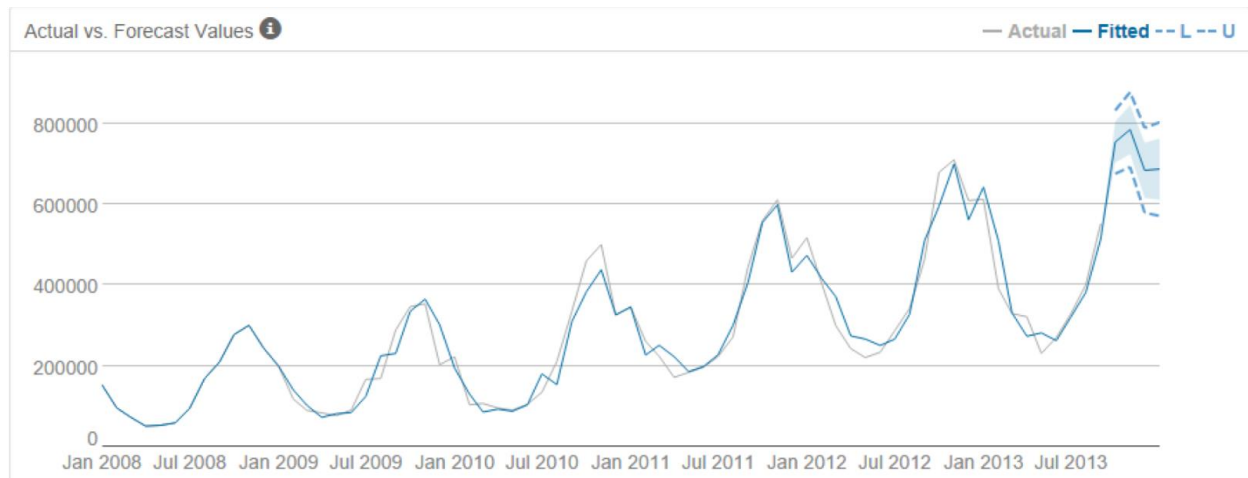
Before you Submit

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