Project: Forecasting Sales

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 fulfills the Time Series criteria. The dataset shows be a sorted sequentially list of data taken at same intervals of time and with at least one data point taking monthly from January 2008 to September 2014; satisfying the Time Series criteria

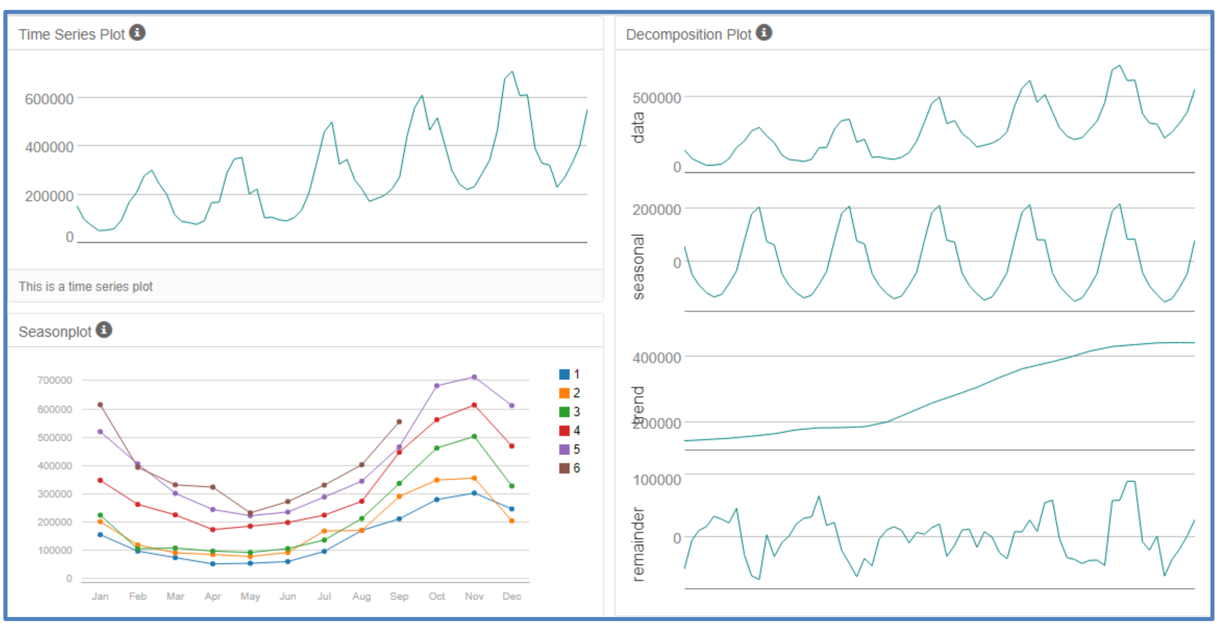
2. Which records should be used as the holdout sample?
   The holdout sample should be considered taking data from the last 4 months of the dataset from Jun 2013 to September 2013, considering the request: 4 months of sales prediction

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 plot of the dataset shows an upward trend (a general trend up to the right) and seasonality pattern, while error plot shows fluctuation



Step 3: Build your Models

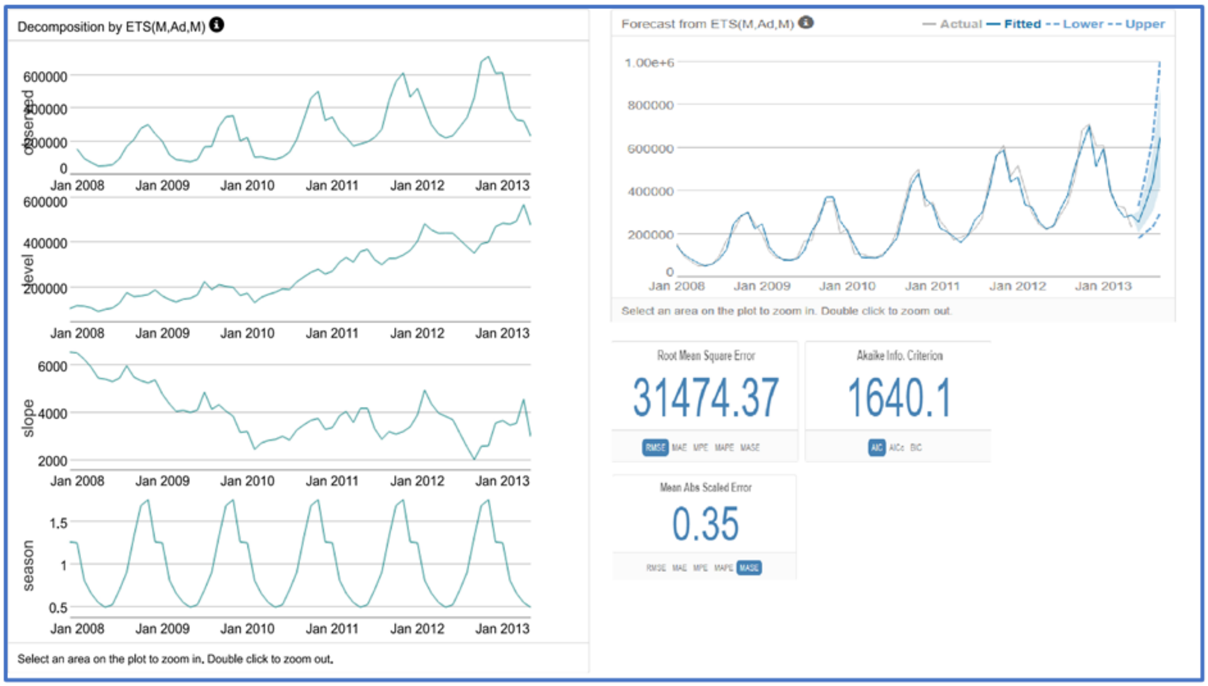
Analyze your graphs and determine the appropriate measurements to apply to your ARIMA and ETS models and describe the errors for both models. (500 word limit)

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

The model terms for the ETS model chosen were:

- Multiplicative term for error or remainder, because it shows variation over time
- Addictive damped term for the trend because it shows a linear increasing variation in magnitude over time
- Multiplicative term for seasonality because it shows an increment over time



The RMSE or Root Mean Square Error is relatively low (31,474.37) and the MASE or Mean Absolute Scaled Error is less than 1 (0.35) implicating that the model is a good predictive model

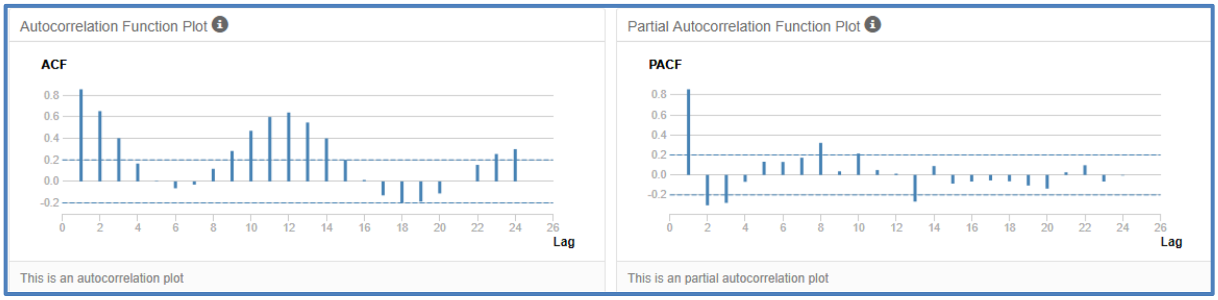
In-sample error measures:

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
3243.4703524	31474.3668886	24188.2167878	-0.572395	10.3052041	0.3528697	0.0087402

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

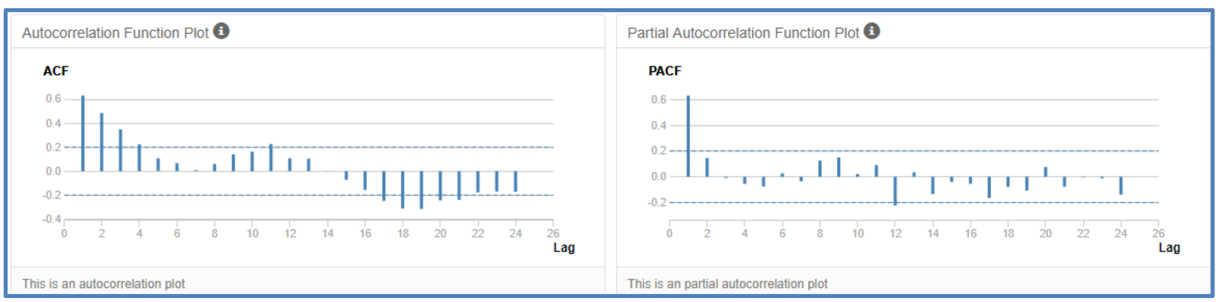
Because the data set shows seasonal components in the time series the predictive model to use was seasonal ARIMA model.

ACF-PACF Time Series:



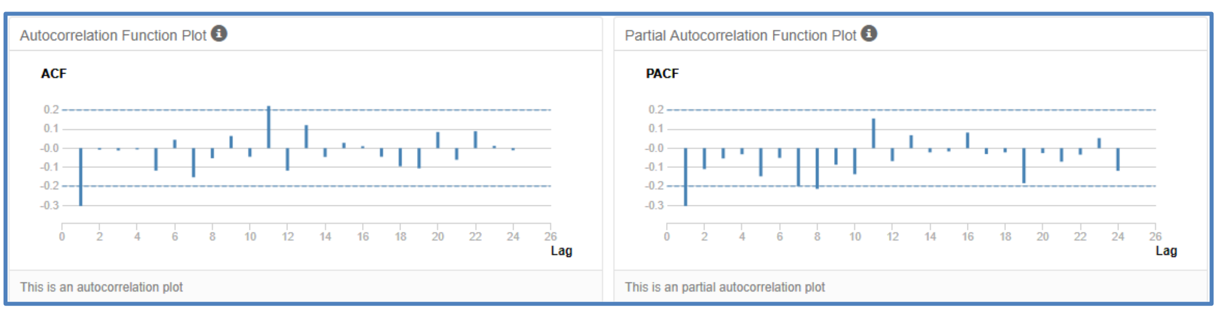
ACF shows a slowly decaying serial correlation to 0, with increase at seasonal lags. Because the correlation is high, is needed to seasonally the series.

ACF-PACF First Seasonal Difference:



The seasonal difference ACF and PACF plot shows similarity with the previous plot, so it is needed toa second difference

ACF-PACF Second Seasonal Difference:



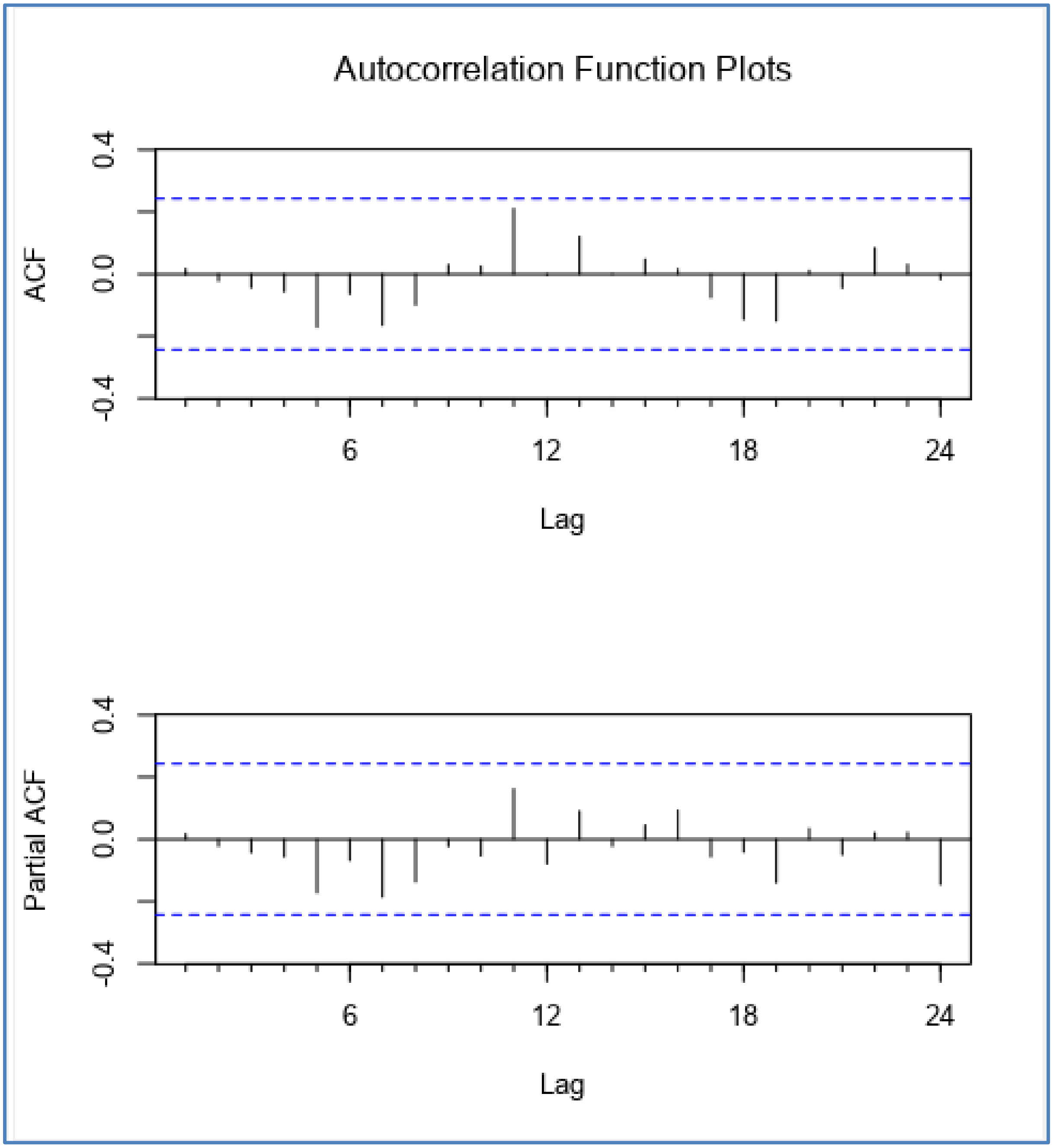
This second seasonal difference removed nearly all of the significant lags form ACF and PACF, so there is not needed future differencing

The model terms for this model chosen were:

- For AR or p and P term 0 because in second differentiation at lag-1 ACF is less than 0
- For I or d and D term 1 because the remaining correlation can be accounted for using autoregressive and moving average terms
- For MA or q and Q term 1 because the ACF plot shows strong negative correlation al lag 1 (since there is only 1 significant lag)
- m term 12 because is the cycle in moths

Then, the model terms for Seasonal ARIMA model are: ARIMA (0,1,1) (0,1,1) [12]

Finally, the Seasonal ARIMA models ACF and PACF show no significantly correlate lags suggesting no to be necessary additional AR() or MA() terms.



The RMSE is relatively low and MASE less than 1 (0.36) showing a strong actuary for forecast

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

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

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.
   Comparing the RMSE and MASE of booth models (ETS and ARIMA) they are extremely similar

In-sample error measures	ETS Model	ARMIA Model
RMSE	31474.3668886	36758.4027043
MASE	0.3528697	0.3646619

The model comparation results shows that the ARIMA model has a lower MASE and RMSE

Accuracy Measures:

Model	ME	RMSE	MAE	MPE	MAPE	MASE
ETSM	-33469.61	53828.48	41542.75	-6.3476	9.3266	0.6904
ARIMAM	27184.45	34010.92	27184.45	6.1547	6.1547	0.4518

Suggesting that ARIMA model is a better predictive model to use

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