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.

Yes it does, it has the following four key characteristics (taken from Udacity):

The series is over a continuous time interval,
It is of sequential measurements across that interval,
It is equally spaced between two measurements and;
No time unit has more than one data point.

2. Which records should be used as the holdout sample?

Since the project requires me to predict 4 intervals into the future. I have filtered out the most recent 4 records (06-2013 to 09-2013) to test how good my model is on unseen data.

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)*

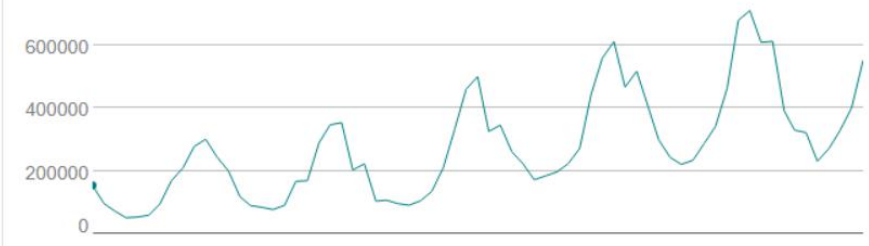
Answer this question:

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.

Looking at the TS Plot output, the trend is positive showing an increase over the time period, the seasonality shows a peak in November and a trough in May and an overall positive increasing trend and the error is distributed around 0.

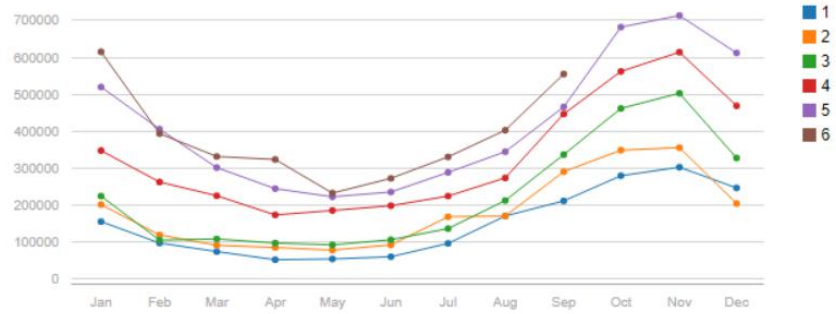
Time Series Plot

Jan, 1: **V1**: 154000



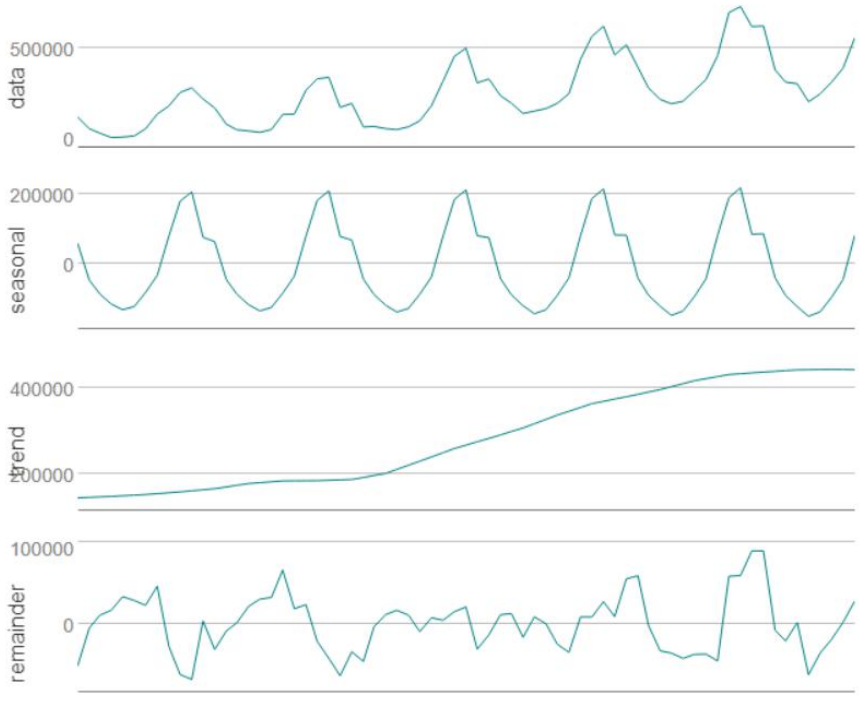
This is a time series plot

Seasonplot



This is a season plot

Decomposition Plot



This is a decomposition plot

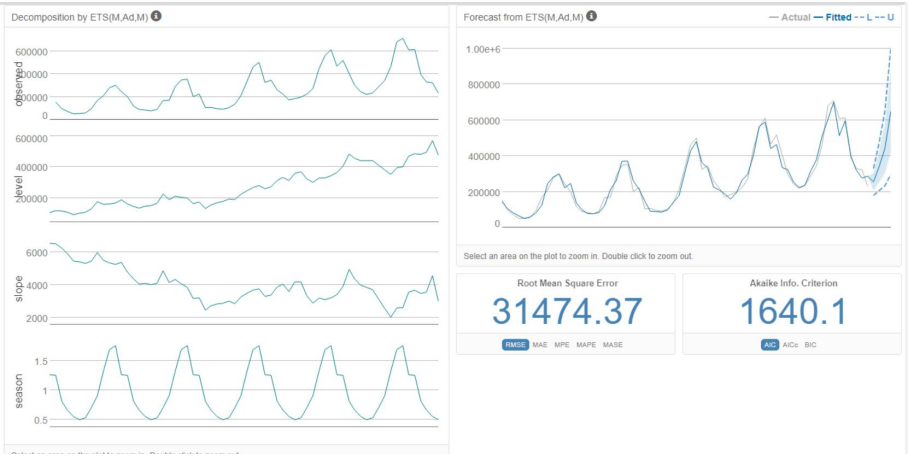
Step 3: Build your Models

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)

1. What are the model terms for ETS? Explain why you chose those terms.

The model terms chosen are Multiplicative, Additive and Multiplicative: ETS(M,A,M).
This decision was based on the series decomposition plot in Step 2.

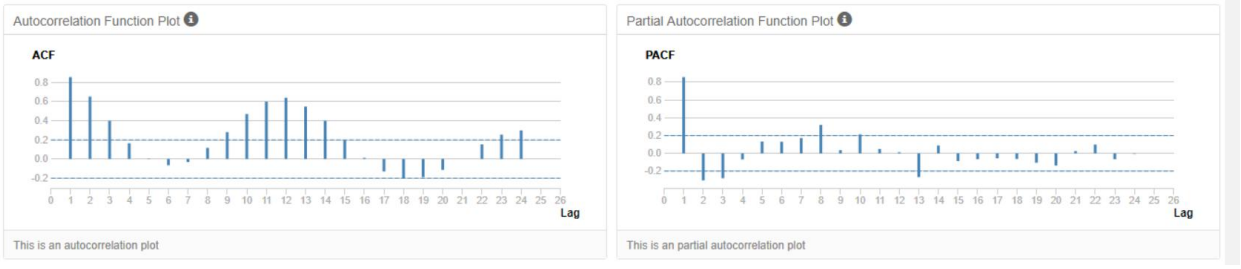
   a. Describe the in-sample errors. Use at least RMSE and MASE when examining results.



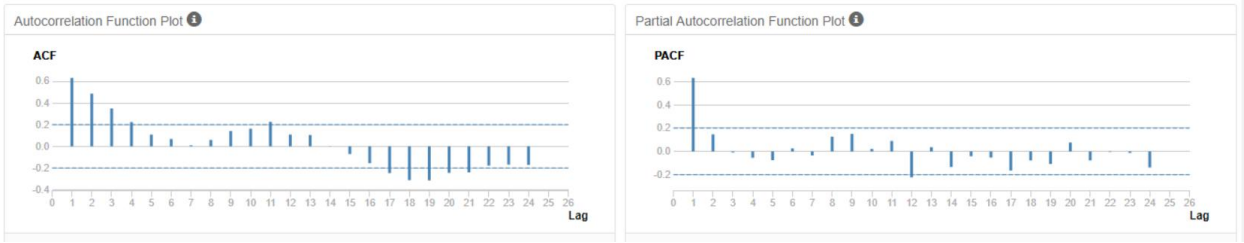
The RMSE is 31,474.37 and the MASE is 0.35. The AIC is 1640.1.

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.

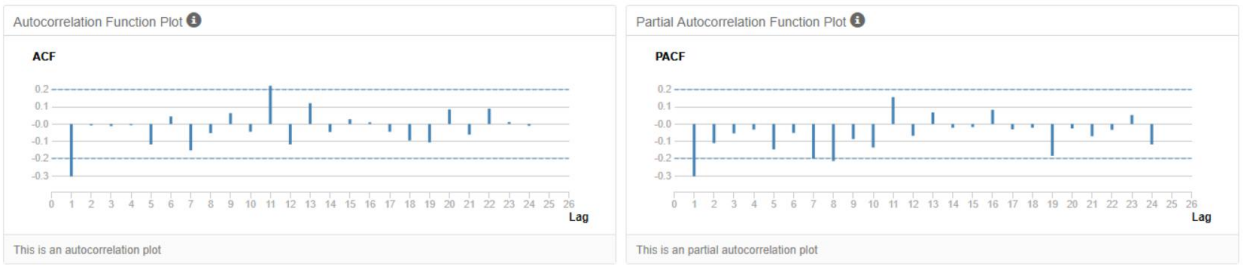
Before starting the model, a TS Plot was created to analyse the monthly sales.This shows a high correlation in the ACF and also a large negative lag at period 13 which is seasonal (due to monthly data).



There was a high seasonal correlation, so a seasonal difference was taken to try make the data stationary. However the data is still highly correlated after this so a seasonal first difference is taken.



The ACF plot shows the data is now stationary. Given the fact that the data is negatively auto-correlated at lag 1, an MA term will be needed in the ARIMA model.



So, with a seasonal difference and a seasonal first difference, an MA term and a 12 month period, an ARIMA(0,1,1)(0,1,0)[12] model will be used.

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

RMSE is 36,761.53 and MASE is 0.36. AIC is 1256.6.

b. 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)

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

I chose the ARIMA model the ME was 27271.52 better than the ETS models -33,469.61, the RMSE was lower, 33999.79 vs 53,828.48 and the MASE was lower 0.4532 vs 0.6904.

ETS:

Actual and Forecast Values:

Actual	ETS
271000	254853.70905
329000	340280.41766
401000	442291.20116
553000	650453.11029

Accuracy Measures:

Model	ME	RMSE	MAE	MPE	MAPE	MASE
ETS	-33469.61	53828.48	41542.75	-6.3476	9.3266	0.6904

ARIMA:

Actual and Forecast Values:

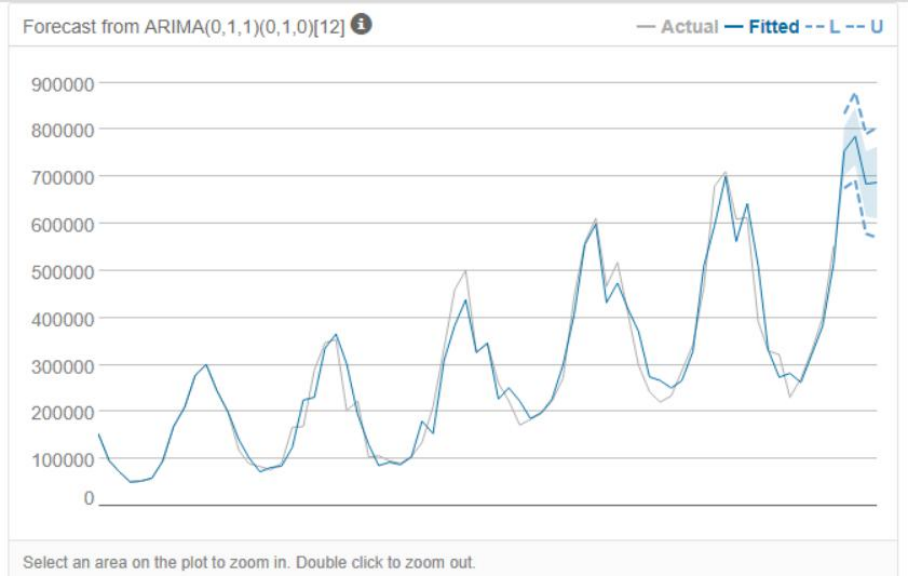
Actual	ARIMA
271000	263228.48013
329000	316228.48013
401000	372228.48013
553000	493228.48013

Accuracy Measures:

Model	ME	RMSE	MAE	MPE	MAPE	MASE
ARIMA	27271.52	33999.79	27271.52	6.1833	6.1833	0.4532

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

The forecasts for the next four periods are 754,854.46, 785,854.46, 684,854.46 and 687,854.46.



Alteryx Workflow: