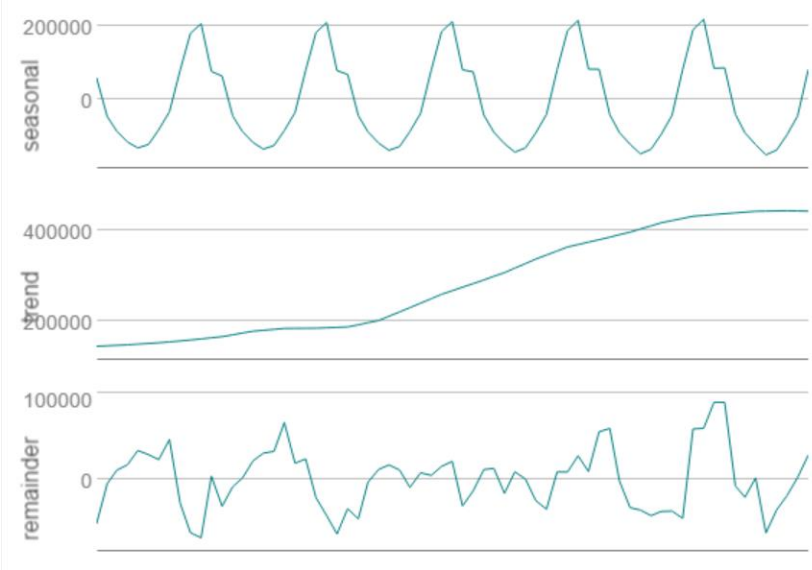
Wilson

Project 6 **Time Series Forecasting**

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.
    a. Yes, it has all 4 attributes of a time series data set. It has continuous time interval, sequential measurements across intervals, equal spacing between every 2 consecutive measurements and each time unit within the time interval has at most one data point.
2. Which records should be used as the holdout sample?
    a. The last 4 records 2013-06, 2013-07, 2013-08, 2013-09 are used as validation.

    I use TS Plot to graph and gather more information about the data.



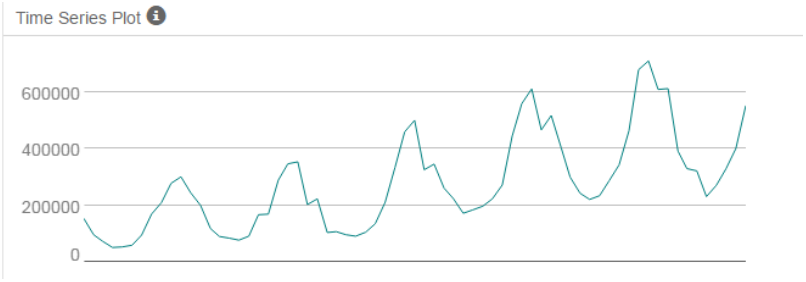
In this graph, seasonality is multiplicative and trend is addictive. Seasonality peaks are growing over time and trend stays relatively constant. Trend is growing upwards linearly over time. Error is multiplicative because its peaks are shrinking and growing.

Model terms for ETS is (M,Ad,M). I chose those terms based on my conclusion for seasonality, trend and error listed above.

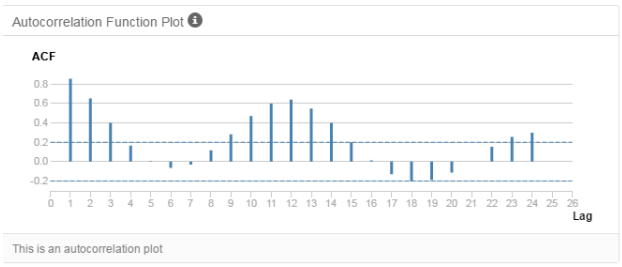
ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
5597.130809	33153.5267713	25194.3638912	0.1087234	10.3793021	0.3675478	0.0456277

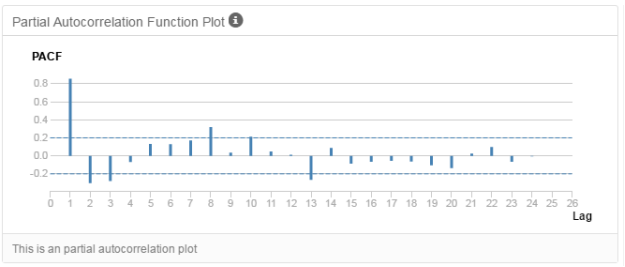
For ETS model RMSE is 33153 and MASE is 0.367. This shows that deviation from the mean of our predicted and observed values is 33153. MASE .367 is <1 which indicates goodness of model.

This is how the time series plot looks before differencing. The data needs to be stationary (mean and variance are constant over time) before it can be used in ARIMA modeling. To determine the ARIMA model terms we need to observe ACF and PACF plots. If ACF plot is positive at lag1 then we use AR terms otherwise MA. If PACF plot drops off quickly then AR, gradual drop off suggest MA terms.

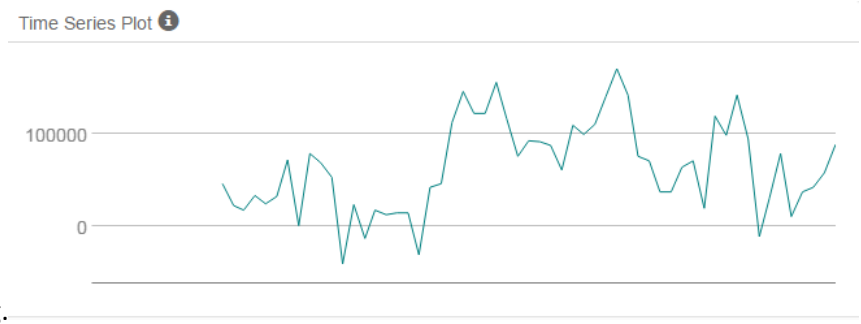


ACF and PACF Plots before differencing.

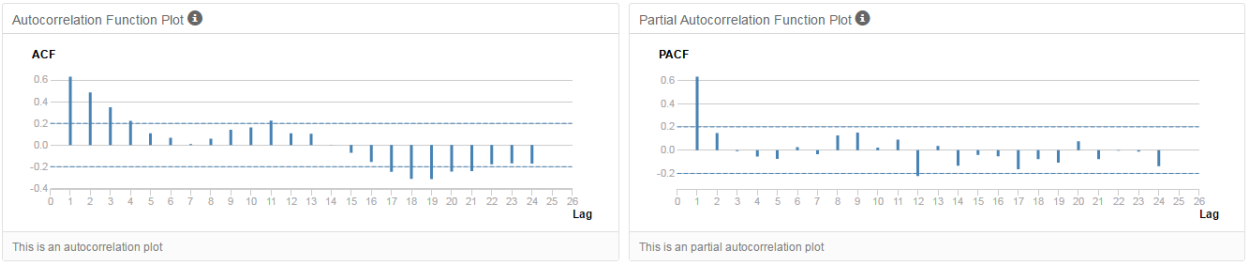




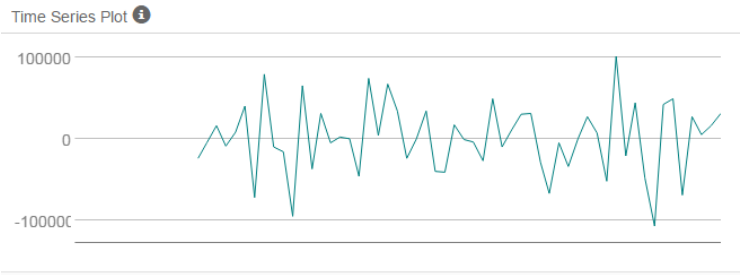
After seasonal differencing time series plot isn't stationary yet. Therefore we need to apply a first differencing.



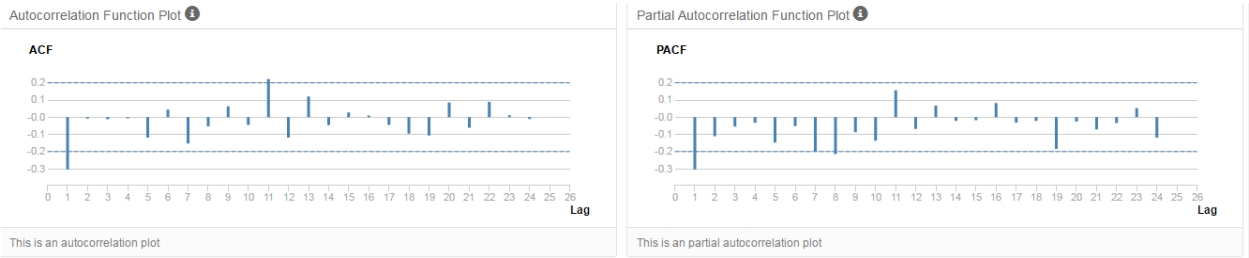
ACF and PACF plots after seasonal differencing.



After first seasonal differencing time series is stationary.



ACF and PACF plots after first seasonal differencing.



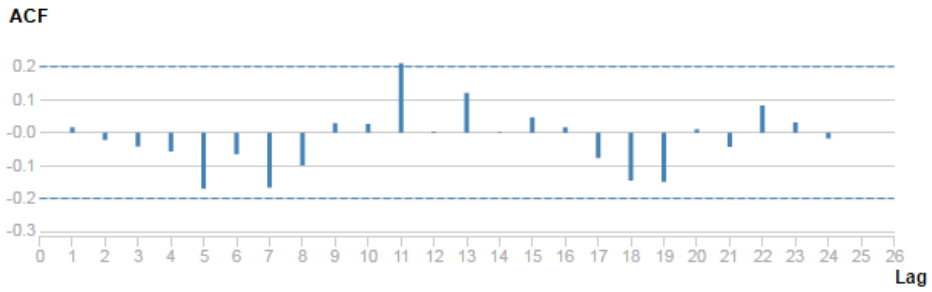
Since our data is seasonal we're going to use ARIMA(p,d,q)(P,D,Q)m. The terms for this model is ARIMA(0,1,1)(0,1,0)12. q is 1 because ACF/PACF is negative at Lag 1. Q is 0 because there's no significant negative correlation at Lag12. d (non-seasonal difference) and D (seasonal difference) is 1 because there's seasonal differencing and first differencing.

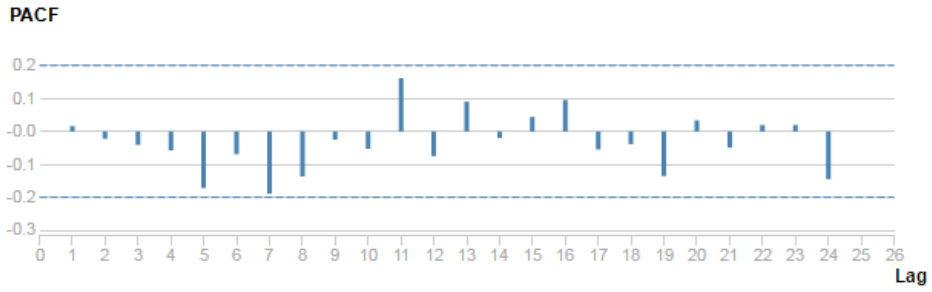
In-sample error measures:

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
-356.2665104	36761.5281724	24993.041976	-1.8021372	9.824411	0.3646109	0.0164145

RMSE is 36761 and MASE is 0.364. This shows that deviation from the mean of our predicted and observed values is 36761. MASE .364 is <1 which indicates goodness of model

ACF and PACF plots for the final ARIMA model. Lack of significant correlation/drop off at lag 1 suggest no MA/AR needed.





In-sample error measures: ETS

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
-356.2665104	36761.5281724	24993.041976	-1.8021372	9.824411	0.3646109	0.0164145

ARIMA

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
5597.130809	33153.5267713	25194.3638912	0.1087234	10.3793021	0.3675478	0.0456277

Accuracy Measures:

Model	ME	RMSE	MAE	MPE	MAPE	MASE	NA
ARIMA65	27271.52	33999.79	27271.52	6.1833	6.1833	0.4532	NA

Accuracy Measures:

Model	ME	RMSE	MAE	MPE	MAPE	MASE	NA
ETS65	-41317.07	60176.47	48833.98	-8.3683	11.1421	0.8116	NA

Pictures above show in sample error measurements and forecast error measurements for our ETS and ARIMA model. ARIMA validation comparison has a lower accuracy measure RMSE (33999) and MASE (.4532) which indicates less standard deviation from mean and goodness than ETS validation comparison. ARIMA estimation model has 1256.5967 in sample AIC which suggest a better fit and less model complexity compared to ETS 1744.6597 in-sample AIC. Therefore ARIMA Model is best used to forecast than ETS Model.

Here's the forecast and graph for the next 4 periods using 95% and 80% confidence intervals.

Record #	Period	Sub_Period	forecast	forecast_high_95	forecast_high_80	forecast_low_80	forecast_low_95
1	6	10	754854.460048	834046.21595	806635.165997	703073.754099	675662.704146
2	6	11	785854.460048	879377.753117	847006.054462	724702.865635	692331.166979
3	6	12	684854.460048	790787.828211	754120.566407	615588.35369	578921.091886
4	7	1	687854.460048	804889.286634	764379.419903	611329.500193	570819.633462