Project: Forecasting Sales

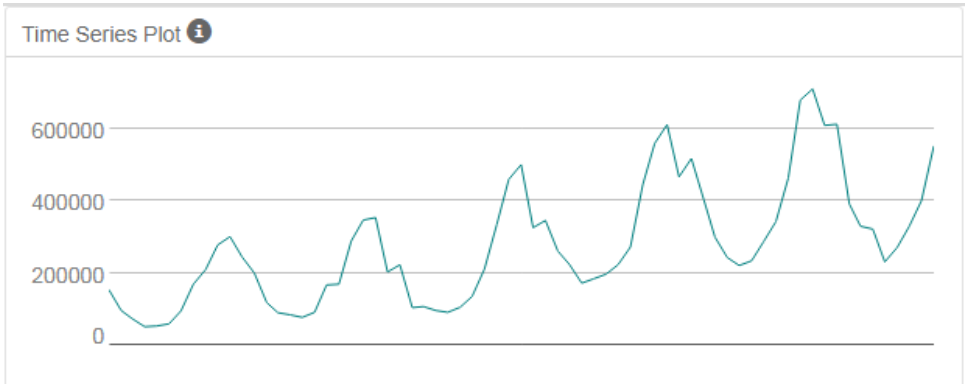
Step 1: Plan Your Analysis

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 data set is cleaned and a new field record id is added in order to sequence the dataset.
   The four key characteristics of time series data i.e.
   The data set has continuous time interval,
   Each time unit in the interval has one data point,
   The data in the dataset is sequential by months,
   There is equal spacing between two consecutive intervals i.e. the spacing is by month

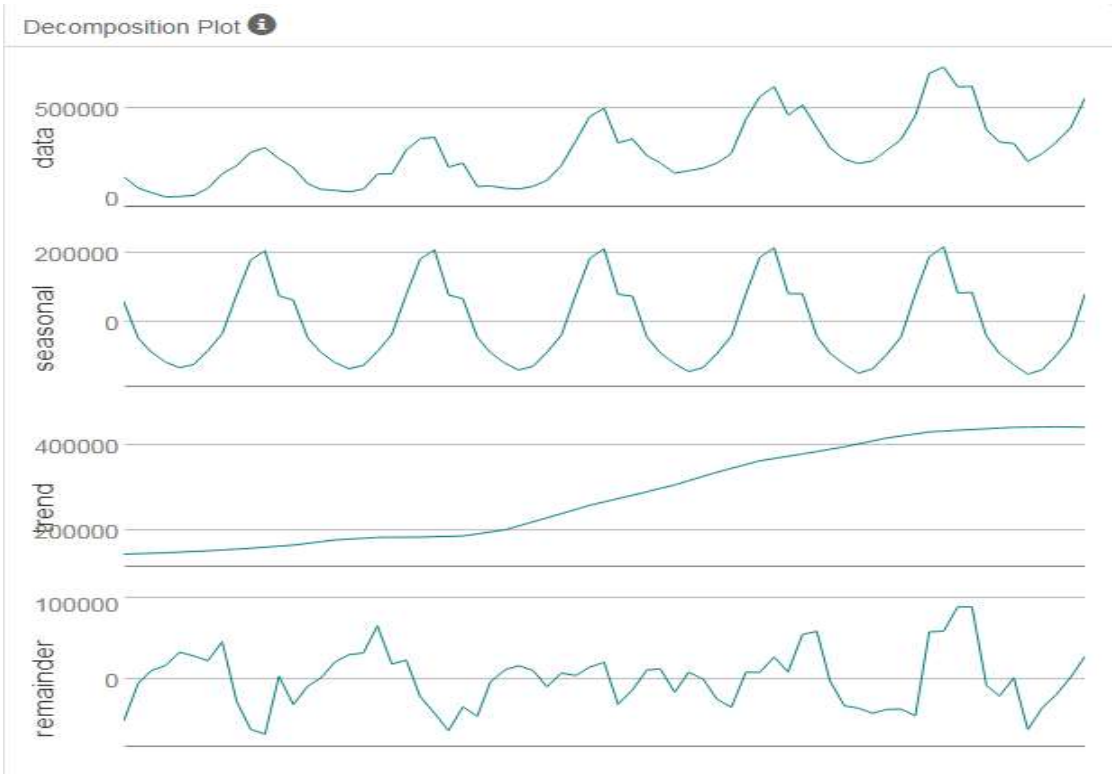
2. Which records should be used as the holdout sample?
   Since we want to forecast four months of data , we holdout a sample of four months i.e. from June 2013 to September 2013

Step 2: Determine Trend, Seasonal, and Error components

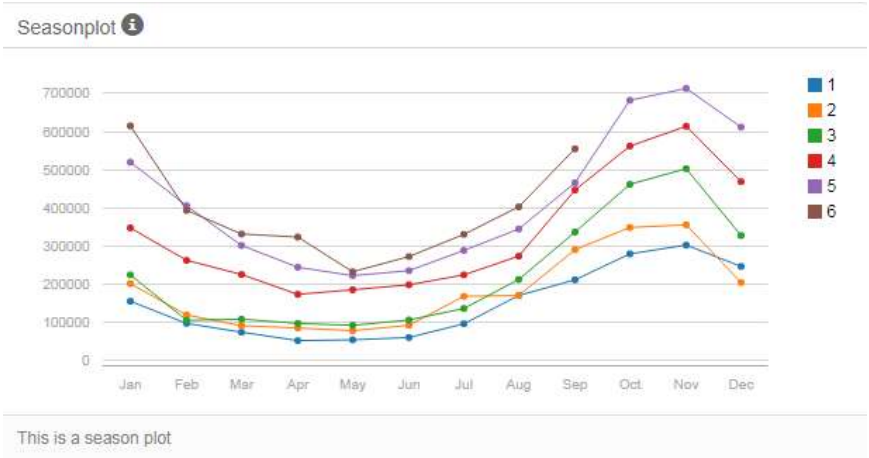
1. What are the trend, seasonality, and error of the time series? Show how you were able to determine the components using time series plots. Include the graphs.



Time Series Plot



Decomposition plot of data, season, trend & remainder



Season plot

Since the seasonality and trends are increasing and showing upward trends we will apply multiplication for seasonality and addition for trend.

Also for error plot we will apply multiplication since there is no trend but fluctuations.

Step 3: Build your Models

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 ETS model are Error, Trend & Seasonality.
For this dataset we chose
Error – Multiplicative because there is no trend rather fluctuations
Trend – Additive because it is showing an increasing trend
Seasonality – Multiplicative
So the ETS(M,A,M) is chosen. The dampened and non-dampened ETS models are executed with the holdout sample.

ETS dampened model

In-sample error measures:

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

The RMSE is 33153.53 & MASE is 0.3675

ETS non-dampened model

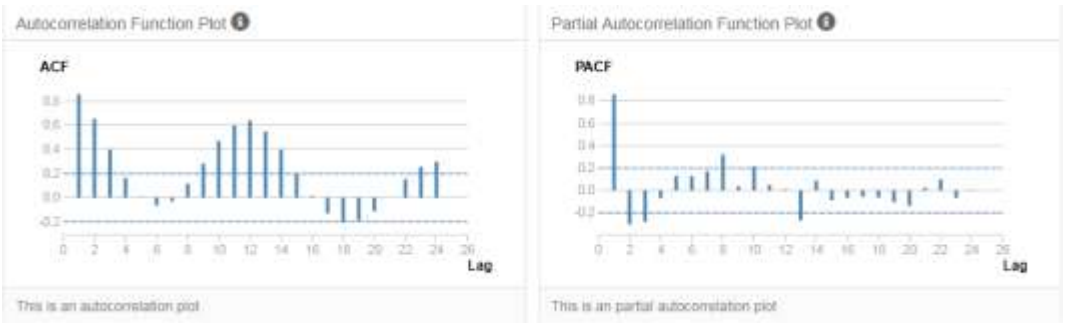
In-sample error measures:

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
2818.2731122	32992.7261011	25546.503798	-0.3778444	10.9094683	0.372685	0.0661496

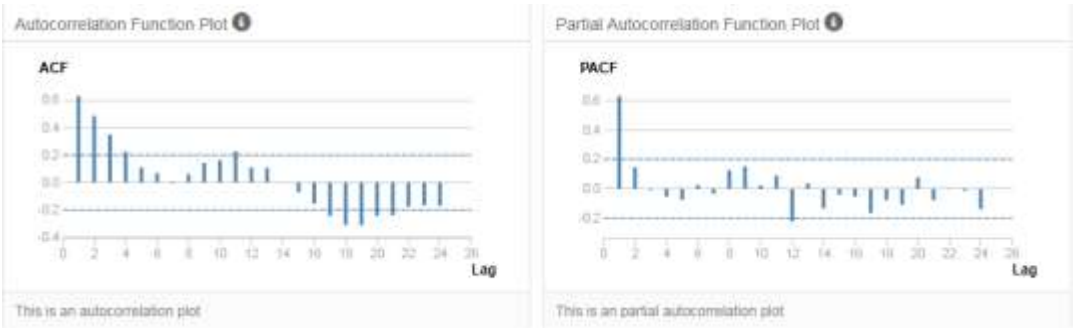
The RMSE is 32992.73 & MASE is 0.3726

After forecasting the values using the holdout sample for ETS dampened & ETS non-dampened model we find that the RMSE & MASE of ETS dampened is less, so we chose the ETS dampened model among these two as the perfect fit.

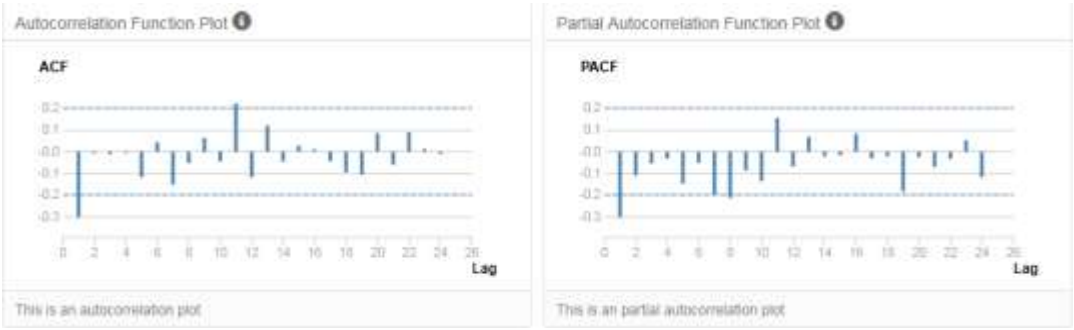
2. What are the model terms for ARIMA? Explain why you chose those terms. Graph the Auto-Correlation Function (ACF) and Partial Autocorrelation Function Plots (PACF) for the time series and seasonal component and use these graphs to justify choosing your model terms.



As we can see from the plots of ACF & PACF above before seasonal differencing that the ACF shows correlations & PACF shows lag at 13.



After seasonal differencing we can see that in ACF it still shows correlation but lesser and there is no correlation displayed in PACF so we perform first seasonal difference.



After performing first seasonal difference we can conclude from the graphs of ACF & PACF that there is no correlation.

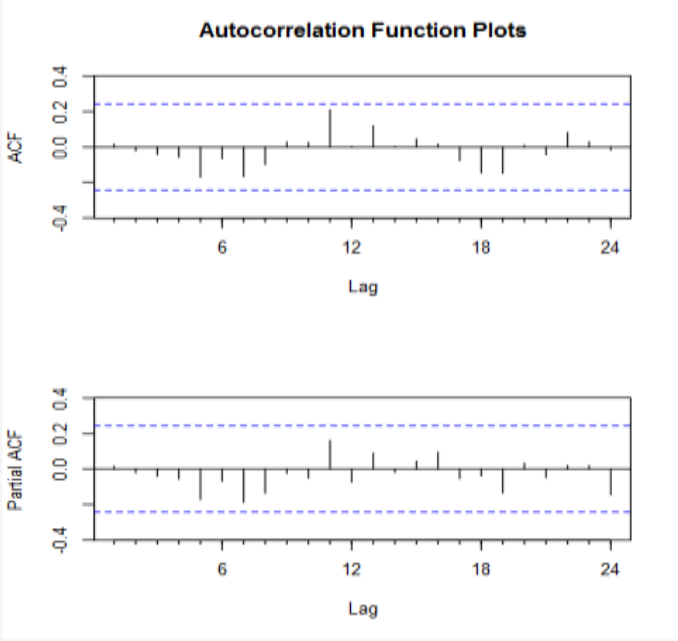
So we terms for ARIMA are

p=0,d=1,q=1 P=0,D=1,Q=0 and period of m=12.
So the model is ARIMA(0,1,1)(0,1,0)12

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

The RMSE is 36761.53 & MASE is 0.3646



There is no correlation as we can see from the plots above.

Step 4: Forecast

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

Actual and Forecast Values:

Actual	ETS_damp
271000	255966.17855
329000	350001.90227
401000	456886.11249
553000	656414.09775

Accuracy Measures:

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

ETS

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	NA
Arima	27271.52	33999.79	27271.52	6.1833	6.1833	0.4532	NA

ARIMA

Comparing the ETS & ARIMA model, we can see that
RMSE & MASE of ARIMA model is less than that of ETS model.
Also the AIC of ARIMA is 1256.60 is less than that of ETS which is 1639.47.
So we choose ARIMA as the perfect fit model for forecasting the values.

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	834046.21595	806635.165997	703073.754099	675662.704146
2013	11	785854.460048	879377.753117	847006.054462	724702.865635	692331.166979
2013	12	684854.460048	790787.828211	754120.566407	615588.35369	578921.091886
2014	1	687854.460048	804889.286634	764379.419903	611329.500193	570819.633462

The image shows the forecast of the next 4 periods from October 2013 to January 2014.
Forecast graph using 95% & 80% confidence interval is shown below