<u>Project: Forecasting Sales</u>

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 given dataset meets the required characteristics of a time series dataset, namely, the data is continuous, sequential, there is equal spacing between consecutive data points - each point is a month, so 12 intervals in a year, and lastly there is only a single data point for corresponding time interval.

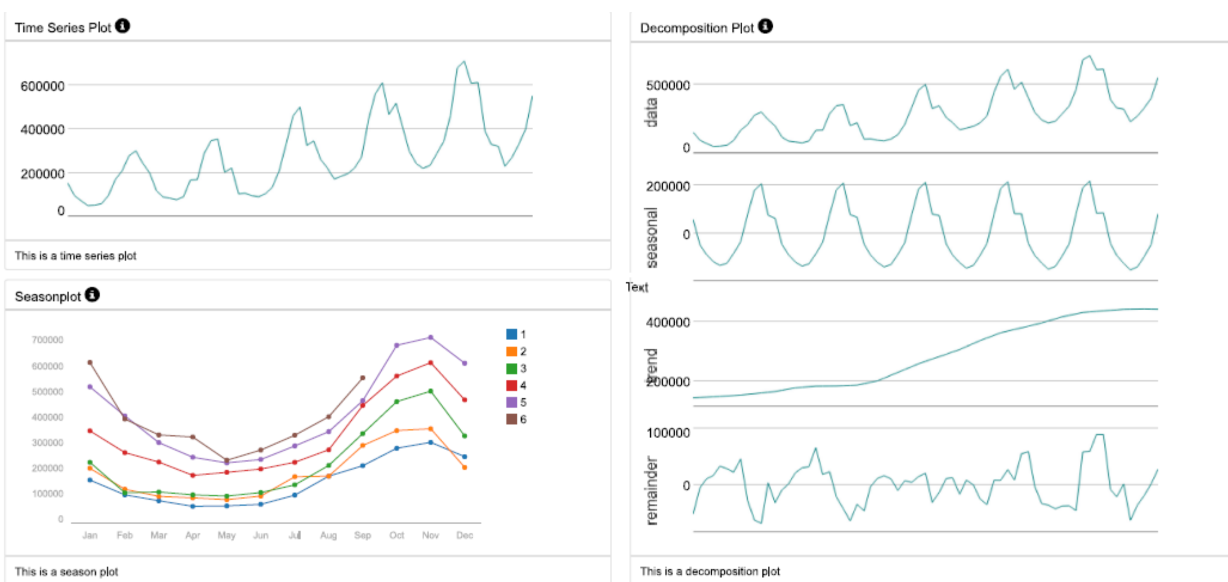
2. Which records should be used as the holdout sample?

   Since we need to do a forecast of the next four months, we should hold the last four months as our holdout sample i.e. the period from 2013-06 to 2013-09.

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.

   From the decomposition we can easily make out that,
   Trend – increasing linearly,
   Error – fluctuating irregularly
   Seasonality – while a bit difficult to notice, the peaks seem to be increasing but very slightly



Step 3: Build your Models

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

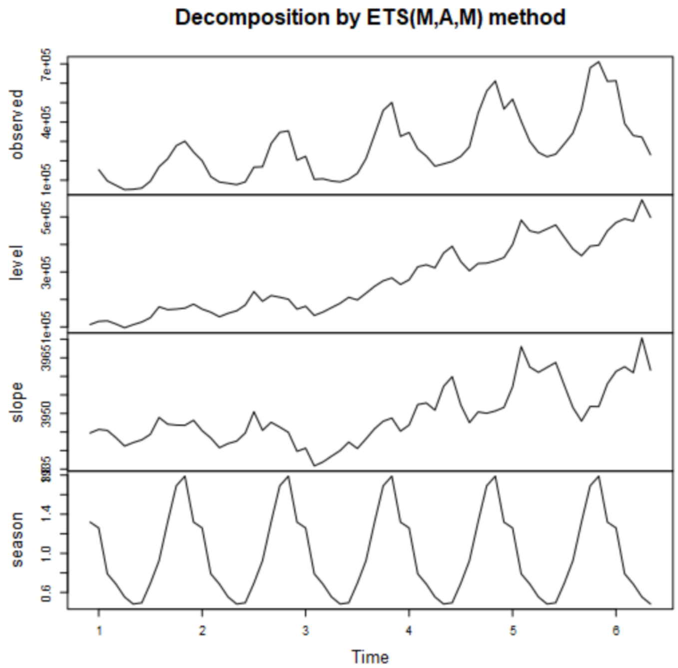
    a. Describe the in-sample errors. Use at least RMSE and MASE when examining results

The pattern observed for ETS or Error, Trend and Seasonality was:

Component	Pattern	Model terms
Error	irregular	Multiplicative (M)
Trend	linearly increasing	Additive (A)
Seasonality	slight peaks	Multiplicative (M)

ETS (M,A,M) model is chosen and implemented in Dampenedd and Non-Dampened mode.

Non-Dampened ETS model



In-sample error measures:

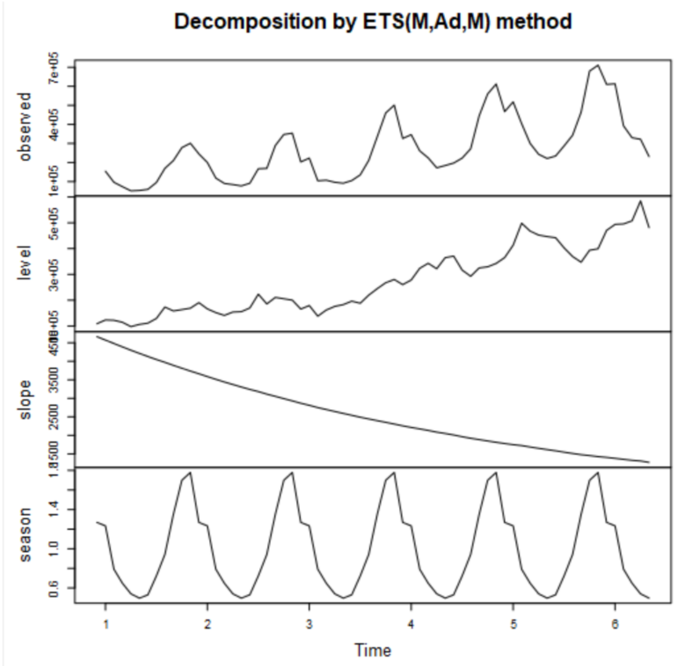
ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
2818.2731122	32992.7261011	25546.503798	-0.3778444	10.9094683	0.372685	0.0661496

Information criteria:

AIC	AICc	BIC
1639.7367	1652.7579	1676.7012

AIC for Non-Dampened ETS model is 1639.7

Dampened ETS 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

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

Information criteria:

AIC	AICc	BIC
1639.465	1654.3346	1678.604

AIC for Dampened ETS model is 1639.5

AIC is comparable for both Dampened and Non-Dampened ETS models. While Non-Dampened has lower RMSE, it has more MAE, MPE and MAPE.

Non-Dampened

Accuracy Measures:

Model	ME	RMSE	MAE	MPE	MAPE	MASE	NA
ETS_MAM	-49103.33	74101.16	60571.82	-9.7018	13.9337	1.0066	NA

Dampened

Accuracy Measures:

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

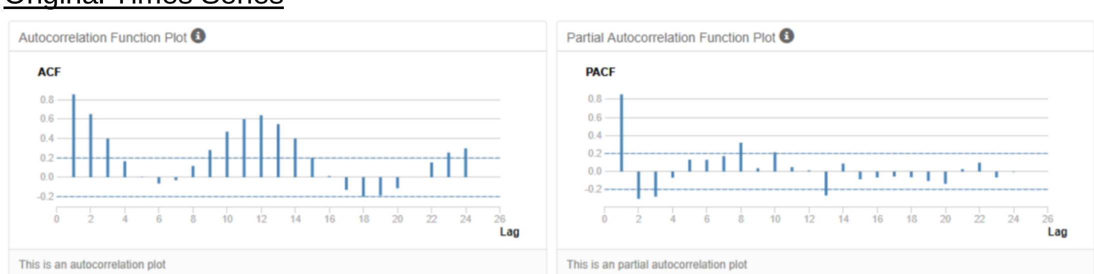
If we look at the Forecast accuracy measures, errors in all parameters for Dampened are lower than Non-Dampened. Additionally, for an ETS model, MASE < 1, which is true for Dampened ETS model. **So out of these two, Dampened ETS model is better choice**.

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.

Component	Pattern	Model terms
AR (p)	No positive correlation	0
I (d)	first difference used to remove trend	1
MA (q)	Negative correlation after first difference at lag 1	1
AR (P)	no seasonal correlation at next cycle, i.e. lag 12, lag 24 and so on	0
I (D)	first seasonal difference needed to make series stationary	1
MA (Q)	no correlation	0
m	monthly	12

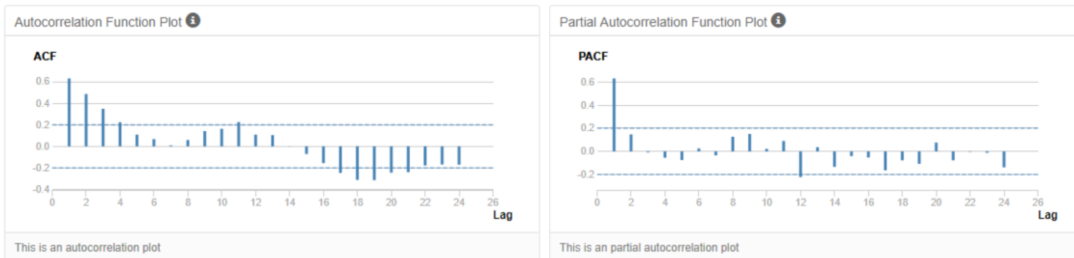
The original time series has linear trend and seasonality so differencing will be done to remove those and make the series stationary.

Original Times Series



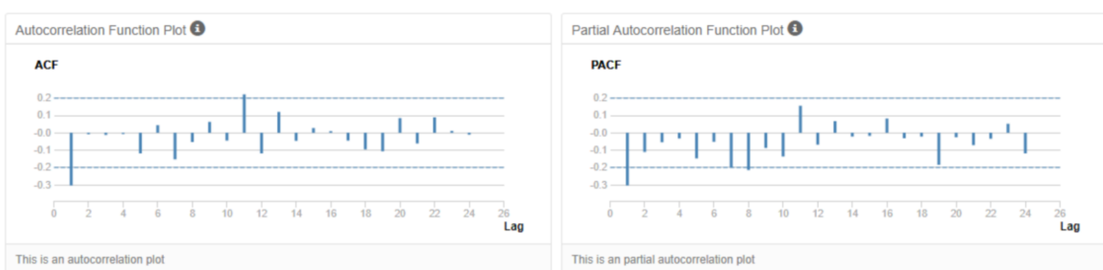
ACF – positive correlation at lag 1 and then eventual drop to zero and repeat. The pattern clearly speaks of seasonality. We need to do differencing to remove it.
PACF – positive correlation at lag 1 and sudden drop.

<u>Time Series with seasonal difference</u>



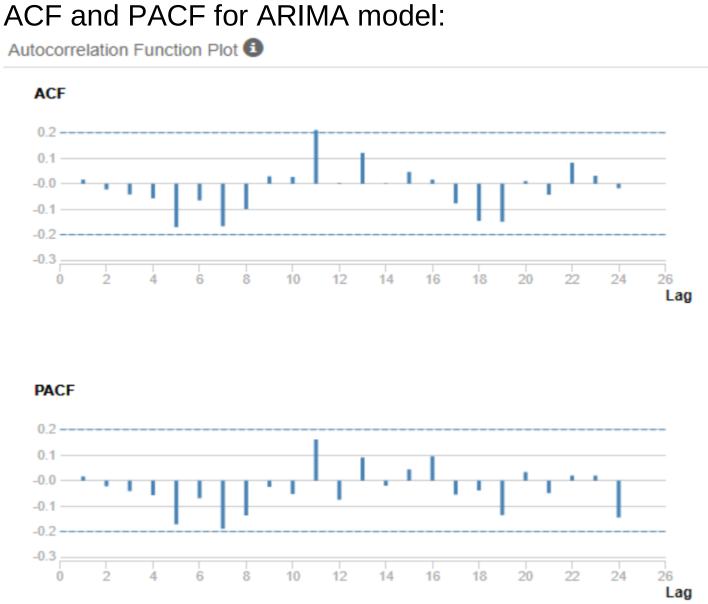
ACF – positive correlation at lag 1 and gradual drop towards 0, will need more differencing to remove seasonal pattern
PACF – positive correlation at lag 1 and then sudden drop.

Time Series with first seasonal difference



ACF – negative correlation at lag 1 and drops off, no further differencing needed
PACF – negative correlation at lag 1

ACF and PACF for ARIMA model:



ARIMA model in-sample error measures:

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

Information Criteria:

AIC	AICc	BIC
1256.5967	1256.8416	1260.4992

AIC for the model is 1256.6

Step 4: Forecast

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.

   ARIMA model is the better choice, lets see why:

In-sample error measures: **ARIMA Model**

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

In-sample error measures: **Dampened ETS Model**

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

Actual and Forecast Values: **ARIMA Model**

Actual	ARIMA
271000	263228.48013
329000	316228.48013
401000	372228.48013
553000	493228.48013

Actual and Forecast Values: **Dampened ETS Model**

Actual	Damped_ETS_MAM
271000	255966.17855
329000	350001.90227
401000	456886.11249
553000	656414.09775

Accuracy Measures:

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

Accuracy Measures:

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

Looking at in-sample error measures and forecast error measures, it's clear that ARIMA Model is more reliable. RMSE for ARIMA is very much lower at 33999.9 vs 60176.47 for ETS. MAE is smaller
at 27271.52 for ARIMA vs 48883 for ETS. MASE should be lower than 1. While that is true for both of them MASE for ARIMA at 0.4532 is 45% lower than that of ETS at .8116.

ARIMA also does a better job of forecasting values closer to the actual values for the holdout sample. We will go ahead with ARIMA model.

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

Forecast for the next four periods using ARIMA Model

Period	Sub_Period	forecast
6	10	760617.152585
6	11	786812.700678
6	12	683059.130563
7	1	684480.980021