Project: Forecasting Sales

Plan Your Analysis

Q.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, the dataset met the criteria of the time series dataset. The criteria of the time series depend on 4 things:

1. It's over a continuous time interval
2. There are sequential measurements across that interval
3. There is equal spacing between every two consecutive measurements
4. Each time unit within the time interval has at most one data point

Our given dataset meets all of the above criteria.

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

Hold out sample is based on for how long we need to do the prediction for in the future. In the give business problem, the prediction has to be made for next 4 months hence the holdout sample will be of 4 months starting from Jun'13 and will go up to September'13.

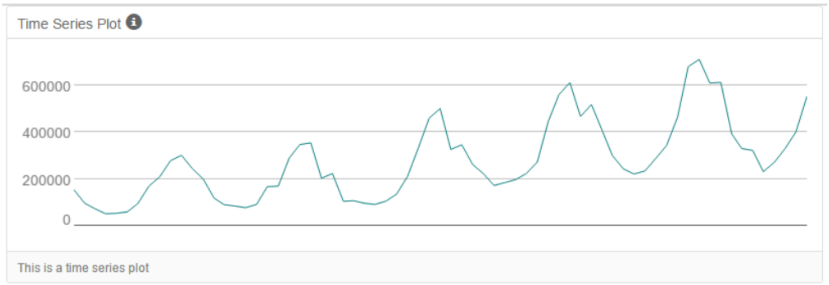
Determine Trend, Seasonal, and Error components

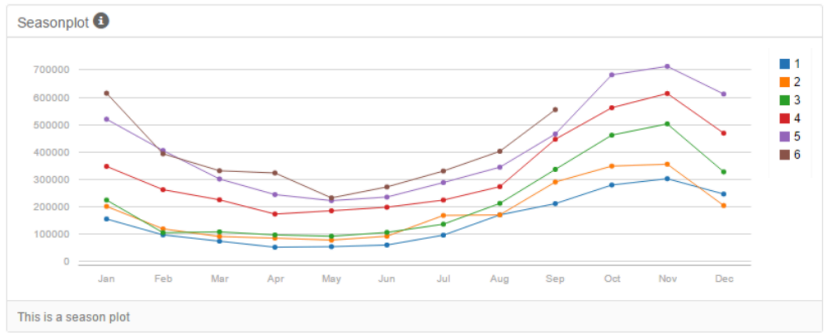
Q.1: What are the trend, seasonality, and error of the time series? Show how you were able to determine the components using time series plots. Include the graphs.

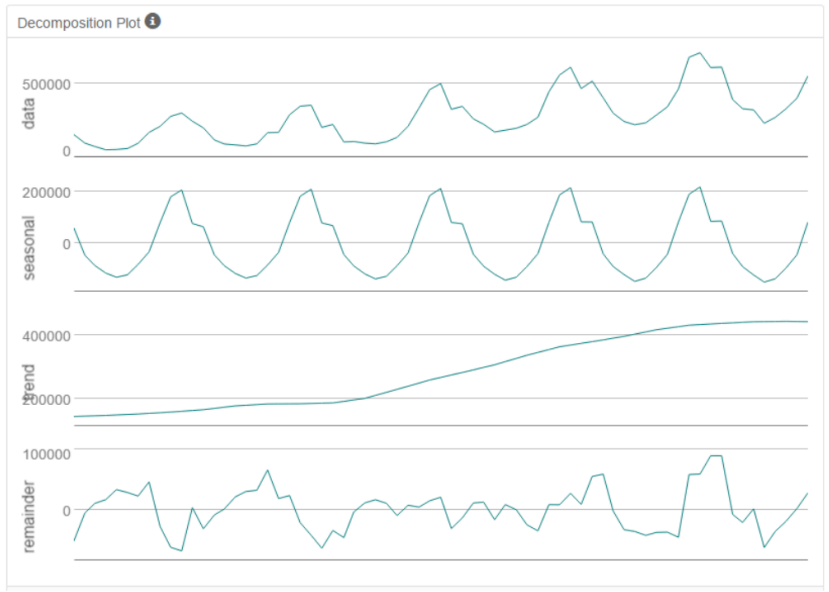
The trend, seasonality and error of the given time series is as below:

From the below graph we can deduce the following things about trend, error and seasonality.

Error: From the graph we can deduce that the fluctuations in the error are not constant in magnitude over the time hence we will apply it multiplicatively.
Trend: The trend in here shows the linear behavior hence it will be applied additively.
Seasonality: The seasonality is also increasing over the time hence it will be applied multiplicatively.







Build your Models

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

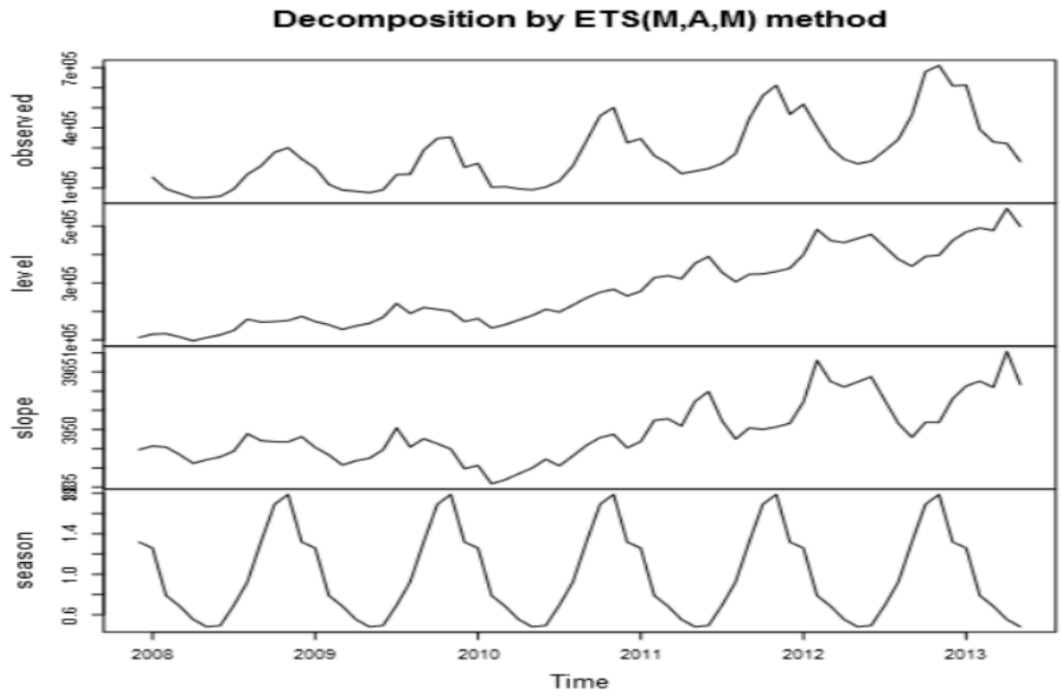
As deduced from the above graph the ETS model was applied in the M, A, M manner i.e. Error is applied multiplicatively, trend is applied additively, and seasonality is applied multiplicatively.

Q.1 a: Describe the in-sample errors. Use at least RMSE and MASE when examining results

We have applied our ETS model with and without the dampening factor and the results are shown below:

ETS without the dampening factor

The graph of the ETS without the dampening factor is as follows:



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

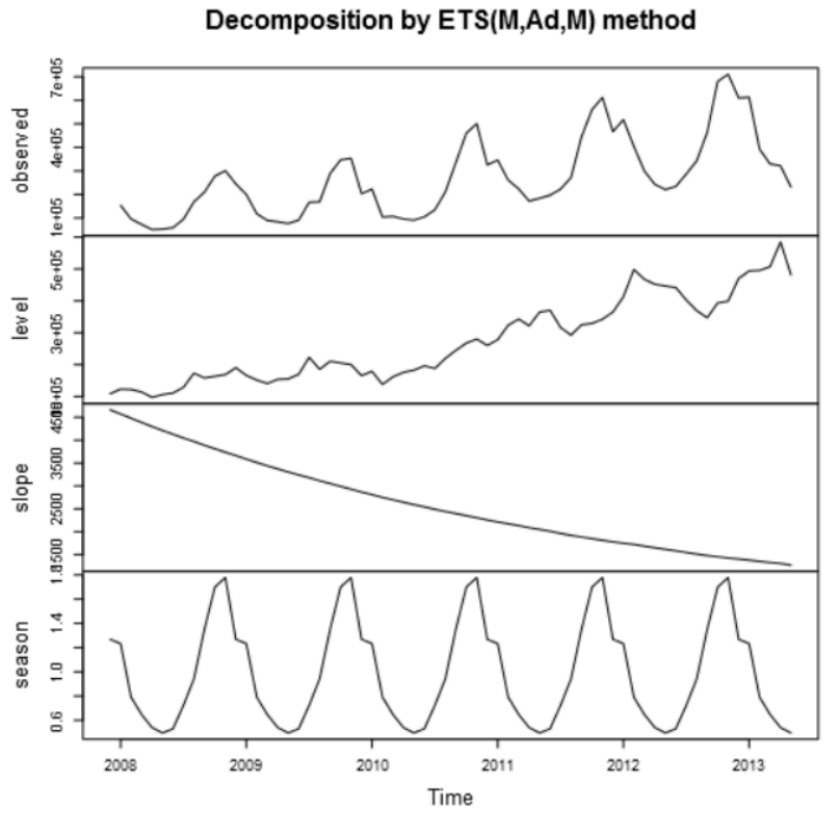
Information criteria:

AIC	AICc	BIC
1639.7367	1652.7579	1676.7012

RMSE value: 32992.726 MASE value: 0.372685, AIC value is 1639.7367

ETS with the dampening factor

The graph of the ETS without the dampening factor is as follows:



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

Information criteria:

AIC	AICc	BIC
1639.465	1654.3346	1678.604

RMSE value: 33153.526 MASE value: 0.3675478, AIC value is 1639.465

Now to decide between non-dampening model and dampening model which is to be chosen we analyze the actual and forecast value of both of the models, which are shown below:

ETS without dampening:

Actual and Forecast Values:

Actual	ETS
271000	248063.01908
329000	351306.93837
401000	471888.58168
553000	679154.7895

Accuracy Measures:

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

ETS with dampening factor:

Actual and Forecast Values:

Actual	ETS_damped
271000	255966.17855
329000	350001.90227
401000	456886.11249
553000	656414.09775

Accuracy Measures:

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

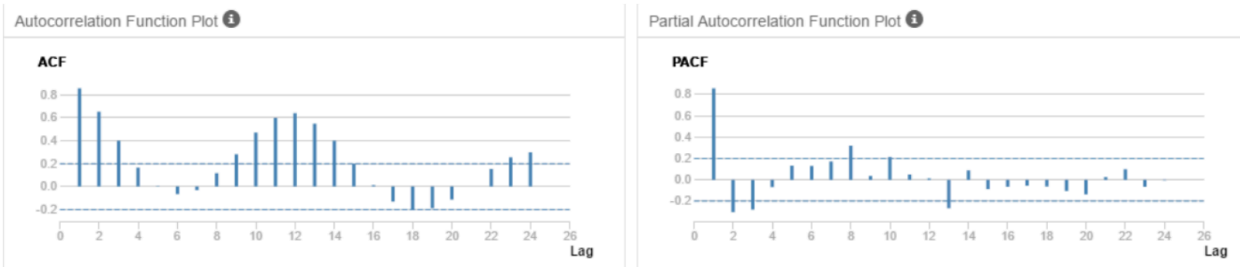
From above tables we can deduce that ETS model with dampening model is more accurate than the ETS model without dampening factor.
Also, the AIC value of dampening and non-dampening ETS model is almost the same but the RMSE and MASE of ETS dampening model is lower hence, **ETS model with dampening factor** will be chosen for our analysis.

Q.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. Re-graph ACF and PACF for both the Time Series and Seasonal Difference and include these graphs in your answer.

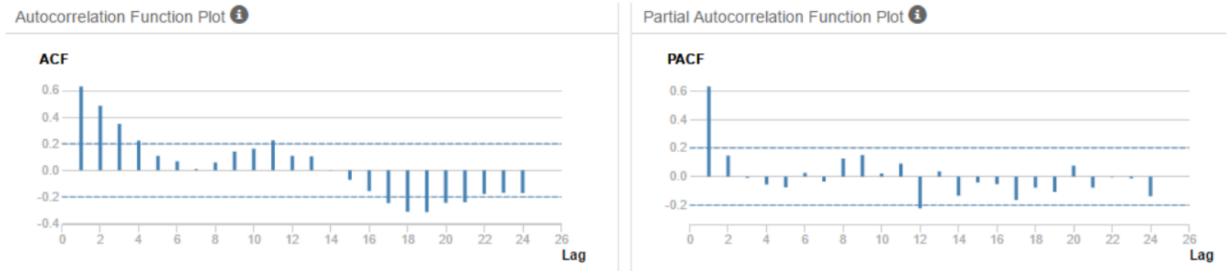
ARIMA stands for **A**uto **R**egressive **I**ntegrated **M**oving **A**verage model. ARIMA model is of two types i.e. season and non-seasonal model. We need to get the ACF and PACF graph to decide which ARIMA model needs to be applied for our analysis.

The initial ACF and PACF graphs are shown below (without differencing):



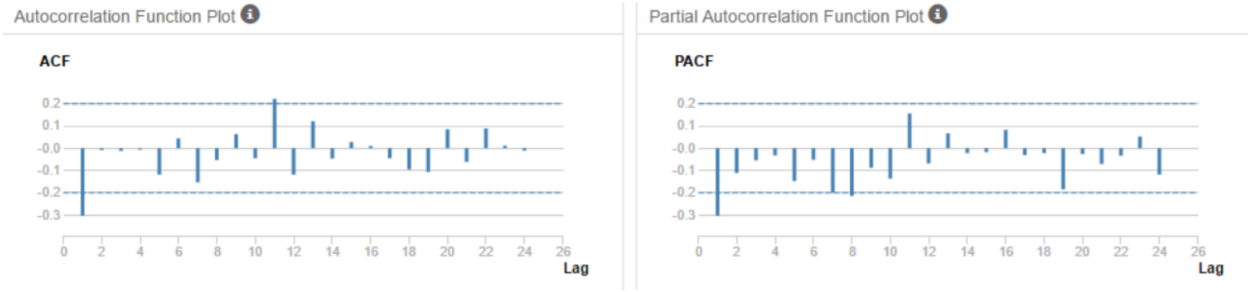
We can observe from the above plots that ACF has strong correlation whereas in PACF there is a significant lag at period 13 we can deduct that this is because of the seasonal effect. Hence the ARIMA seasonal model will be used for our analysis.

Now we take the seasonal difference and produce the ACF and PACF graphs:



We can see that still there is a strong correlation in the ACF graph whereas for the PACF graph there is no strong correlation.

We will again take the normal difference and produce ACF and PACF graph:



Now there is no correlation in the ACF graph

From the above graphs we can deduce that the ARIMA model of the following form will be used: ARIMA(0,1,1)(0,1,0)12 since lag-1 is negative and also the period is of 12 months.

Information Criteria:

AIC	AICc	BIC
1256.5967	1256.8416	1260.4992

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 value: 36761.528 MASE value: 0.3646109, AIC value is 1256.5967

Forecast

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

To decide which model is to be used we should compare the actual and forecast values of both the models:

ETS Model:

ETS Model:

Actual and Forecast Values:

Actual	ETS_damped
271000	255966.17855
329000	350001.90227
401000	456886.11249
553000	656414.09775

Accuracy Measures:

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

ARIMA Model:

ARIMA Model:

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

From above table we can observe that ARIMA model is more accurate in forecasting the values hence we will use ARIMA model.

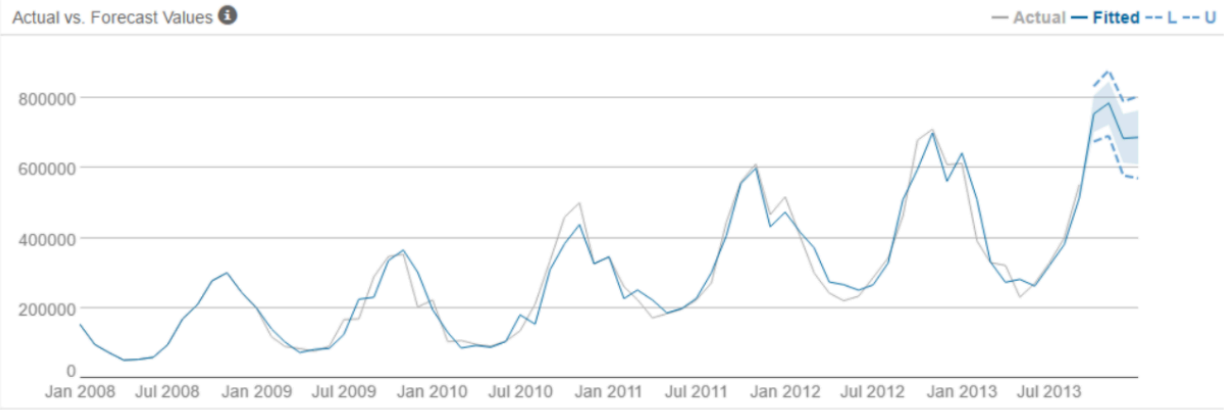
The MAPE value of ARIMA is 6.1833 and MAE value is 27271.52
MAPE value of ETS model is 11.1421 and MAE value is 48833.98

These above errors value also suggest that we should use ARIMA instead of ETS as for ARIMA we have small values.

Also, based on the Accuracy measures we can observe that ARIMA has RMSE of 33999.79 whereas ETS has RMSE of 60176.47, MASE for ARIMA is 0.4532 and for ETS is 0.8116.

Since, the RMSE and MASE value for ARIMA is less than ETS hence it is the better choice.

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



The forecasted value for next 4 months:

October'13 : 754,854
November'13 : 785,854
December'13 : 684,654
January'13. : 687,854