

# Project: Forecasting Sales

## Step 1: Plan Your Analysis

### 1. Does the dataset meet the criteria of a time series dataset?

To meet the criteria of a time series dataset, each measurement of data taken across a continuous time interval is sequential and of equal intervals, each time unit having at most one data point, ordering matters in the list of observations and dependency of time.

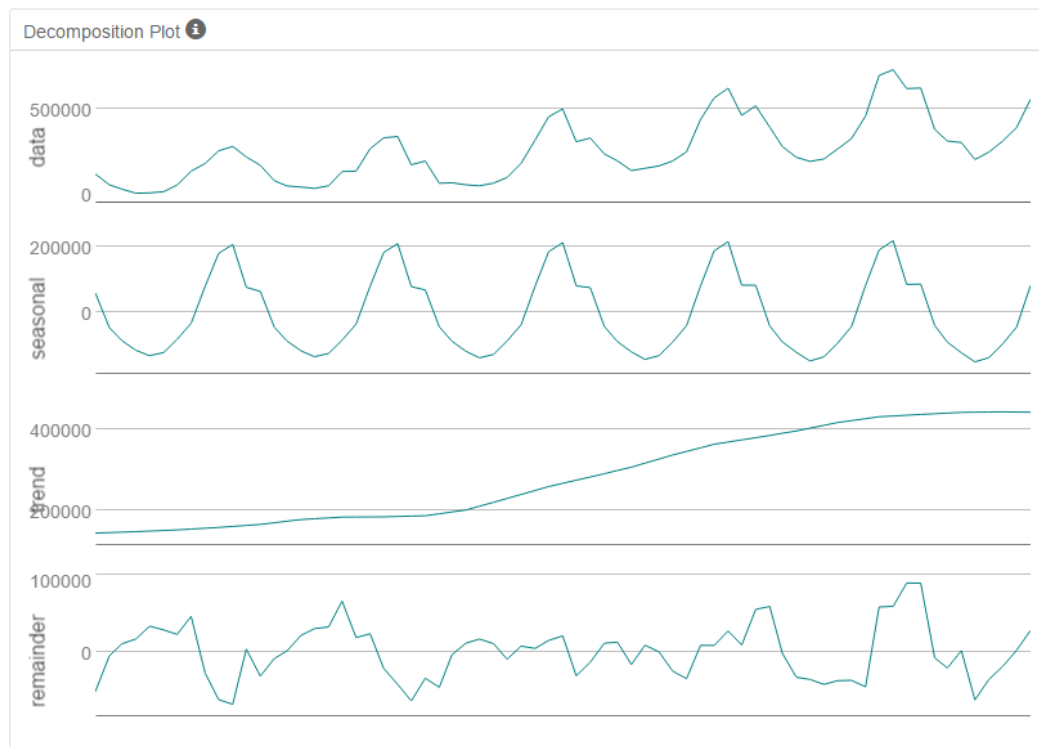
### 2. Which records should be used as the holdout sample?

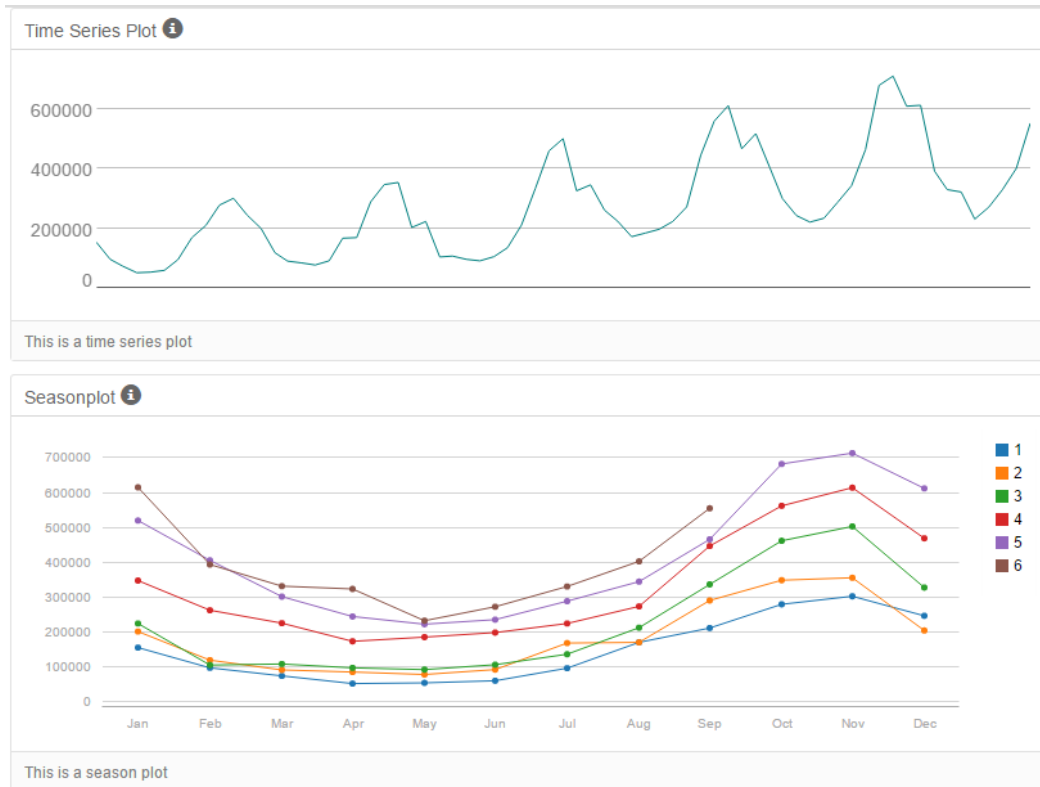
Holdout sample size depends on how far the prediction is. Since we need to predict the sales for the next 4 months, the sample should be 4-long one.

## Step 2: Determine Trend, Seasonal, and Error components

### 1. What are the trend, seasonality, and error of the time series?

The time series and decomposition plots are generated using TS plot function. The seasonality and trend show increasing trends, thus multiplication and addition should be applied respectively. For error plot, there isn't a trend but rather fluctuations and thus should be applied multiplicatively as well.





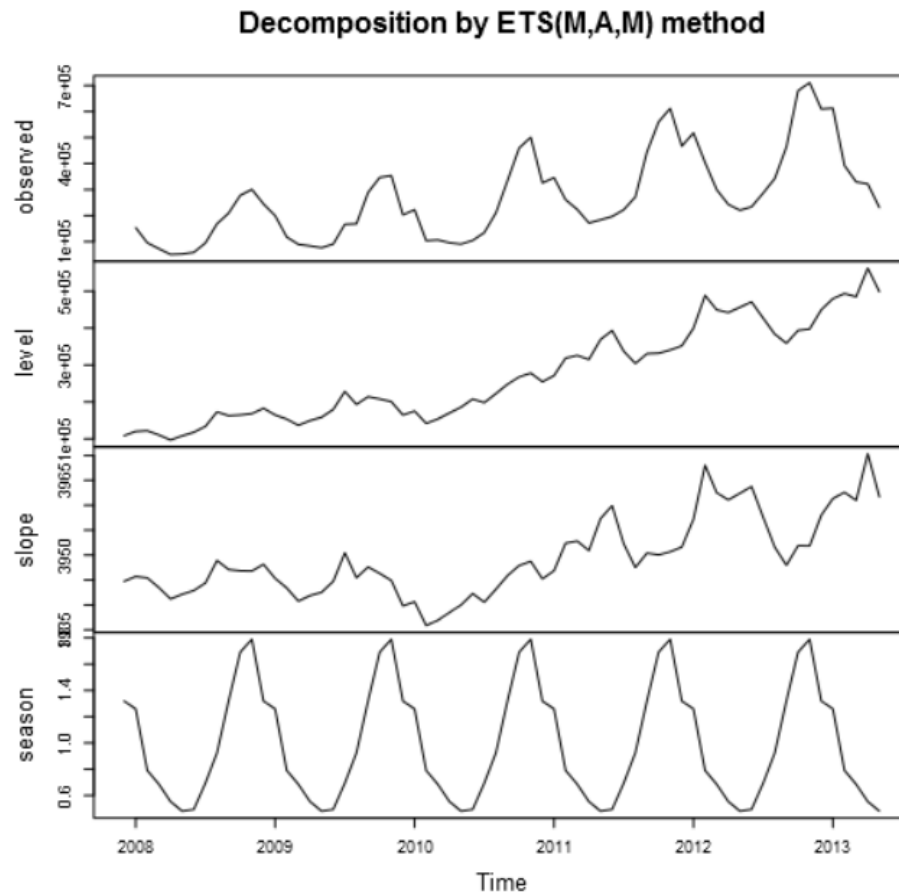
## Step 3: Build your Models

### ETS

#### 1. What are the model terms for ETS?

ETS (M,A,M) is chosen based on the decomposition plot above. A dampened and non-dampened ETS models are run with a holdout sample of 4 months.

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

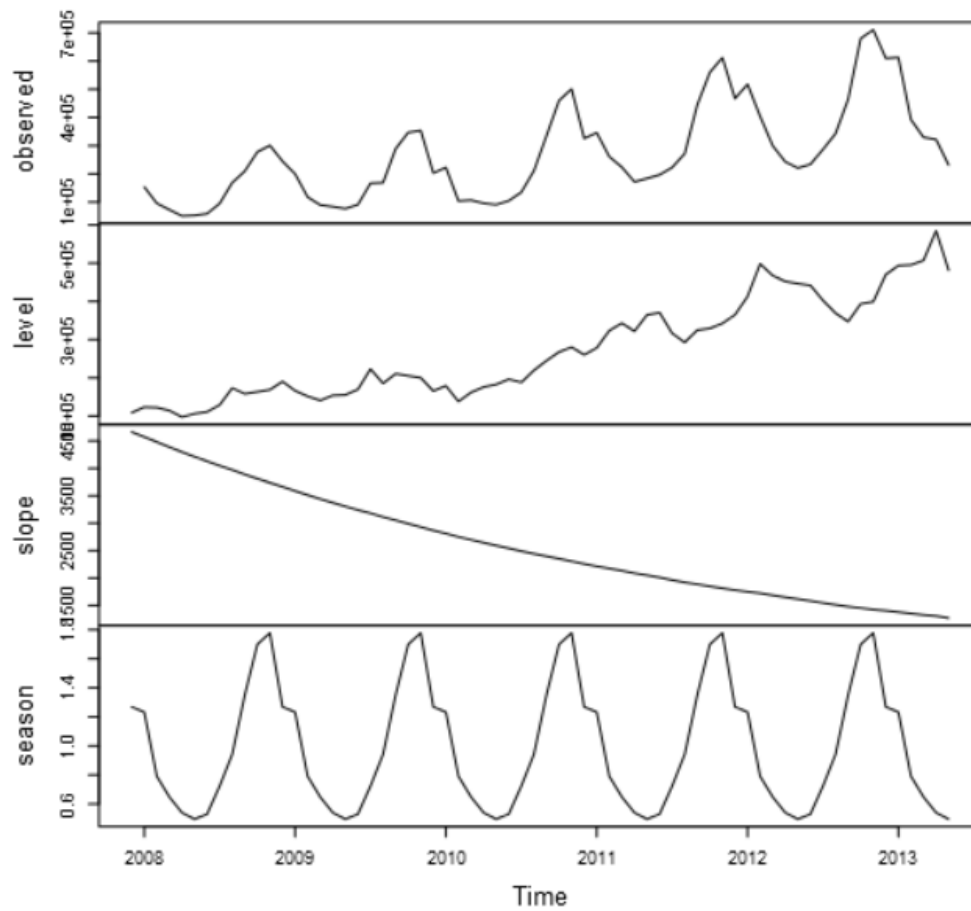
Information criteria:

AIC	AICc	BIC
1639.7367	1652.7579	1676.7012

The AIC value is 1639.74, RMSE (Root Mean Square Error) is 32992.73 and MASE (Mean Absolute Percentage Error) is 0.3727.

Dampened ETS Model:

### Decomposition by ETS(M,Ad,M) method



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

The AIC value is 1639.47, RMSE is 33153.53 and MASE is 0.3675.

Non-Dampened:

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

Dampened:

Actual and Forecast Values:

Actual	ETS
271000	248063.01908
329000	351306.93837
401000	471888.58168
553000	679154.7895

Accuracy Measures:

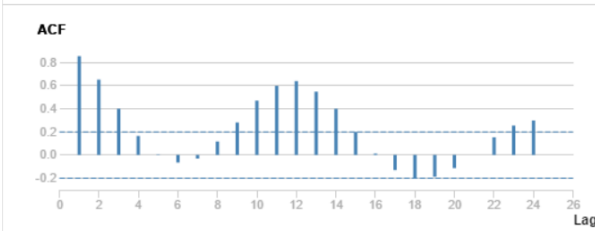
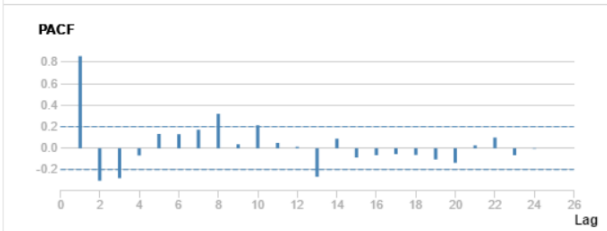
Model	ME	RMSE	MAE	MPE	MAPE	MASE	NA
ETS	-49103.33	74101.16	60571.82	-9.7018	13.9337	1.0066	NA

By comparing the forecast and actual results, dampened model is chosen due to its higher accuracy. The dampened model's RMSE & MASE are lower and could offset its marginally lower AIC.

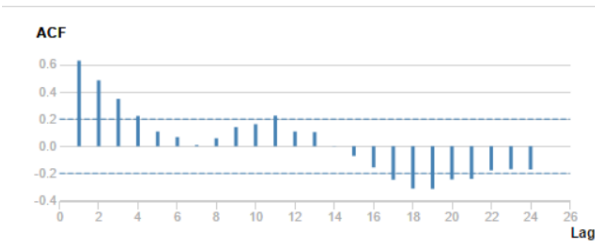
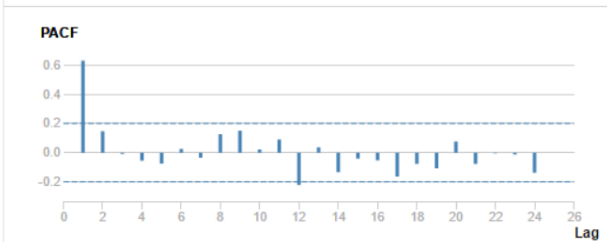
## ARIMA

### 2. What are the model terms for ARIMA?

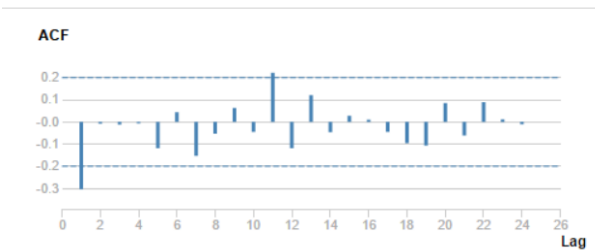
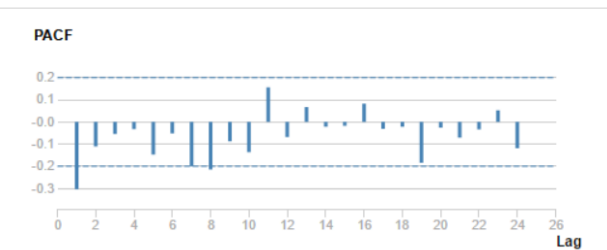
Without differencing, the time series and seasonal component's Auto-Correlation Function (ACF) shows high correlation and the Partial Autocorrelation Function (PACF) shows a significant lag at period 13 which is due to seasonal effect.

Autocorrelation Function Plot Partial Autocorrelation Function Plot 

A seasonal difference is then taken. However, the ACF still shows high correlation while the data doesn't have strong correlation in PACF after a seasonal difference is applied.

Autocorrelation Function Plot Partial Autocorrelation Function Plot 

A seasonal first difference is performed and ACF plot doesn't show strong correlation anymore.

Autocorrelation Function Plot Partial Autocorrelation Function Plot 

ARIMA (0,1,1)(0,1,0)12 is used as lag-1 is negative and the number of period is 12 months.

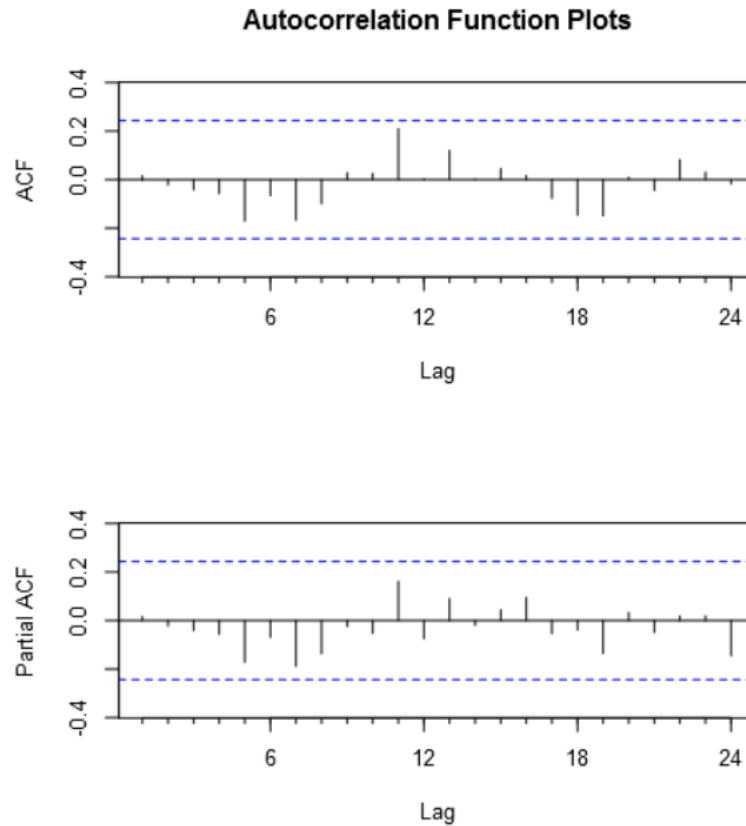
#### Information Criteria:

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

As shown above, the AIC is 1256.60, RMSE is 36761.53 and MASE is 0.3646.



Both ACF and PACF doesn't shows significant correlation and no additional AR or MA terms needed.

## Forecast

### 1. Which model did you choose?

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:

Actual and Forecast Values:

Actual	ARIMA
271000	263228.48013
329000	316228.48013
401000	372228.48013
553000	493228.48013

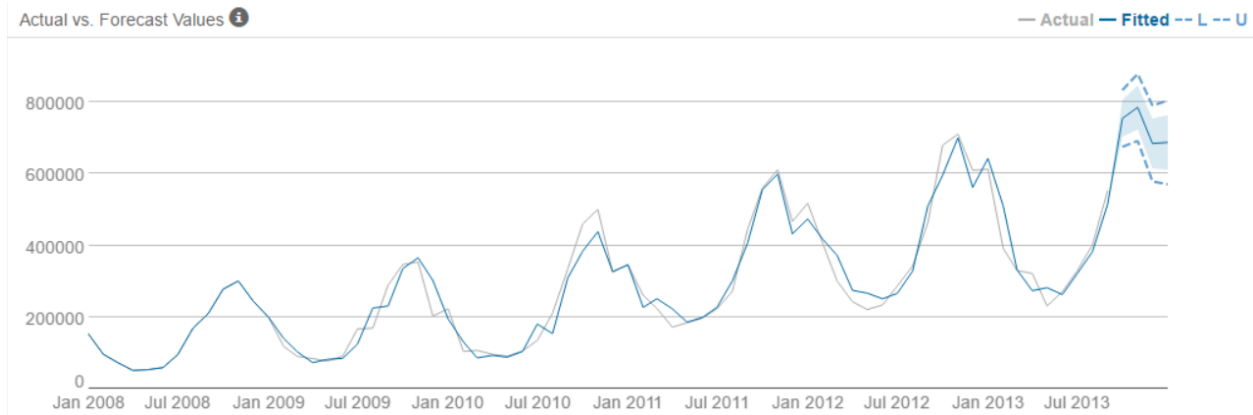
Accuracy Measures:

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

ARIMA model is better at forecasting sales using holdout sample as validation data as the MAPE and ME value are lower than ETS model.

The RMSE for ARIMA is 33999.79 compared to ETS' RMSE at 60176.47. ARIMA's MASE value of 0.4532 is also lower than ETS' MASE value of 0.8116. It is clear that ARIMA model is better since its in-sample error measurements and forecast error measurements are smaller.

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



The forecast for the next 4 periods (Oct-13 till Jan-14) are 754,854, 785,854, 684,654 and 687,854.