Forecasting Sales

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.

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, a 4-month long holdout sample from record 66 to record 69 should be used.

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.**

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.

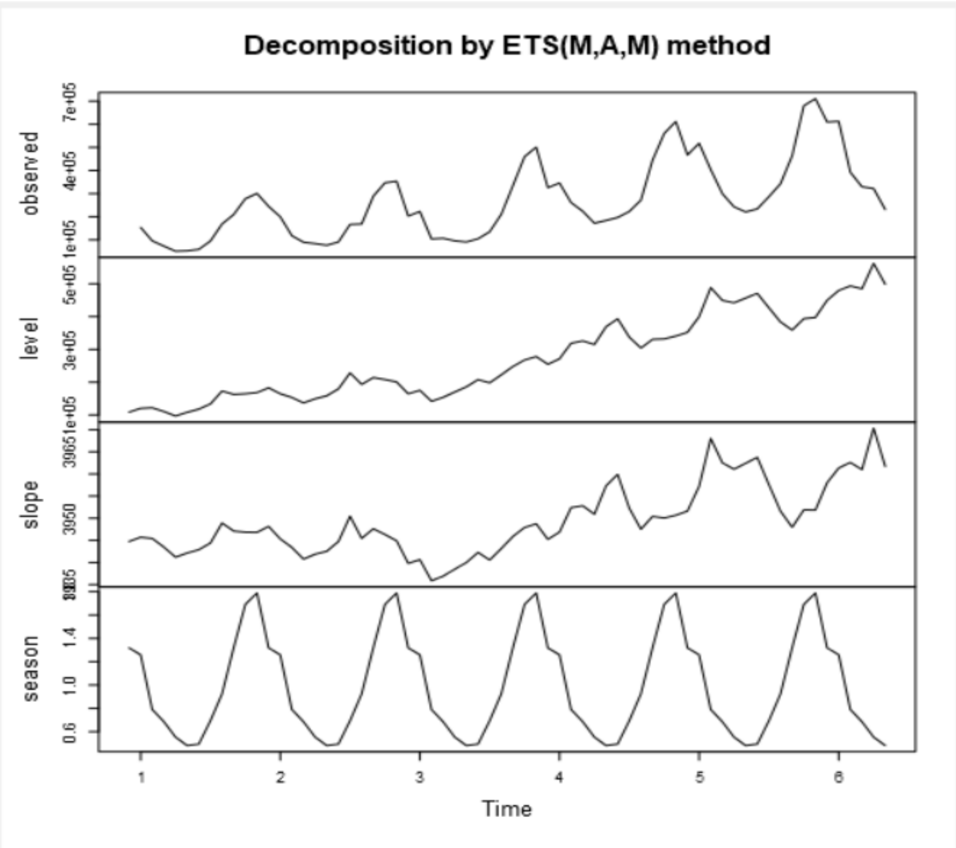
Build your Models

ETS Model

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

ETS(M,A,M) is chosen based on the decomposition plot above. A dampened and non-dampened ETS models are run.

Non-dampened ETS model:



5 Method:
ETS(M,A,M)

6 In-sample error measures:

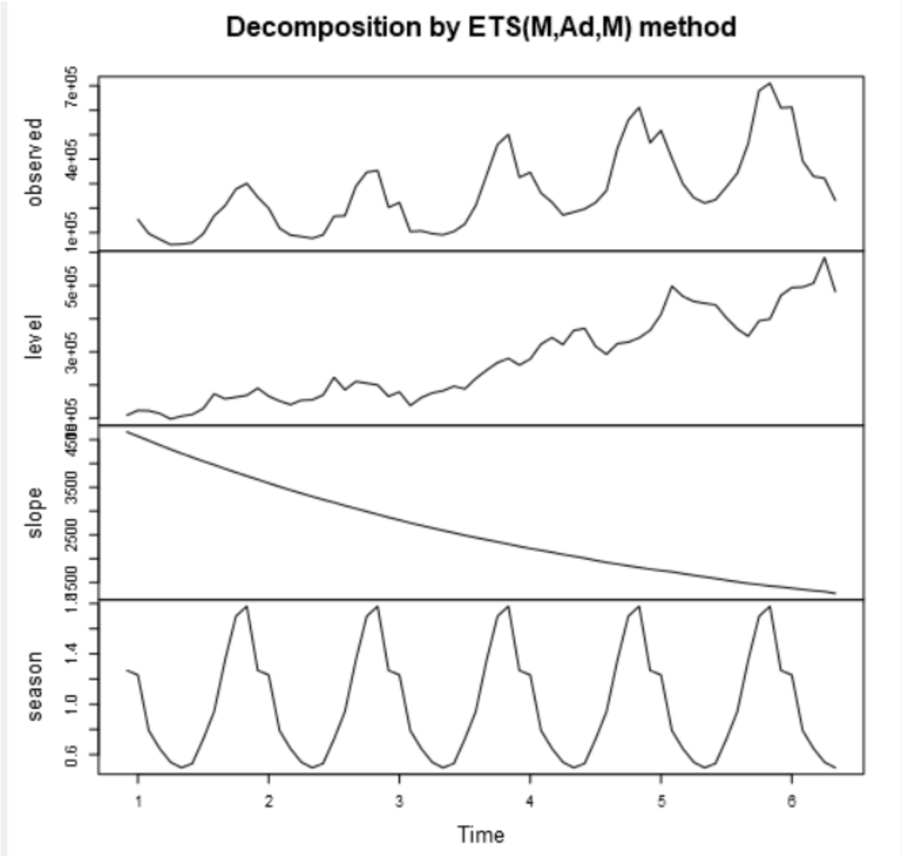
ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
2818.2731122	32992.7261011	25546.503798	-0.3778444	10.9094683	0.372685	0.0661496

7 Information criteria:

AIC	AICc	BIC
1639.7367	1652.7579	1676.7012

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

Non-dampened ETS model:



5 Method:
ETS(M,Ad,M)

6

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

7

Information criteria:

AIC	AICc	BIC
1639.465	1654.3346	1678.604

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

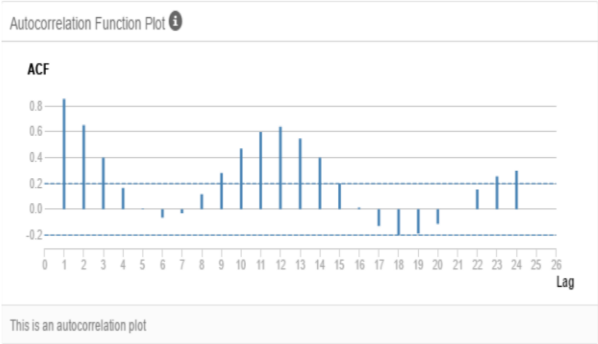
RMSE (Root mean square error) and MASE (Mean absolute percentage error) are two most important error statistics.
The lower the RMSE, the better fit is the model. The fit is determined by looking at what is the predicted value using the model and what was actually observed.
MASE is a relative measure of error applicable only to time series data. It measures relative reduction in error compared to a naïve model. MASE has a threshold of goodness which is one so models close to one or above one are not considered very reliable. Both the dampened and the undampened model have MASE values significantly lower than one, which means both the models are good.

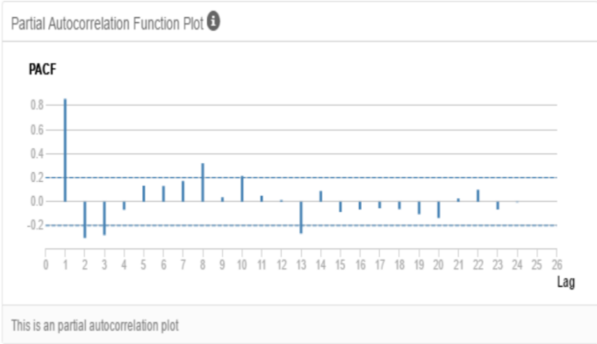
By comparing the forecast and actual results, **dampened model** is chosen due to its higher accuracy. The dampened model's RMSE & MASE are lower and could offset its marginally lower AIC.

ARIMA Model

2. What are the model terms for ARIMA? Explain why you chose those terms.

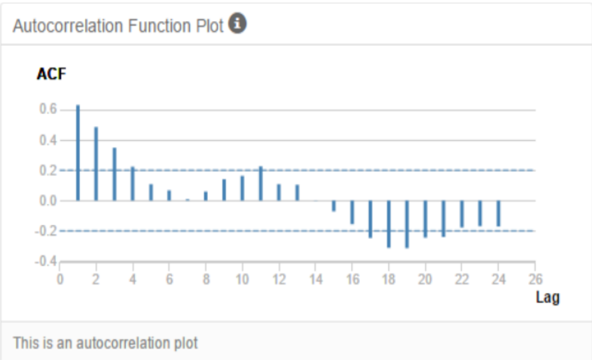
To determine the Model terms, we need to determine ACF and PACF.

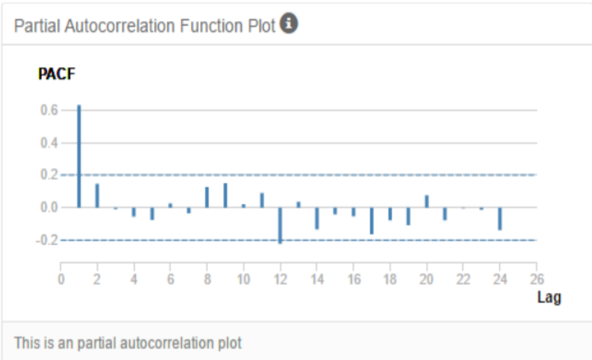




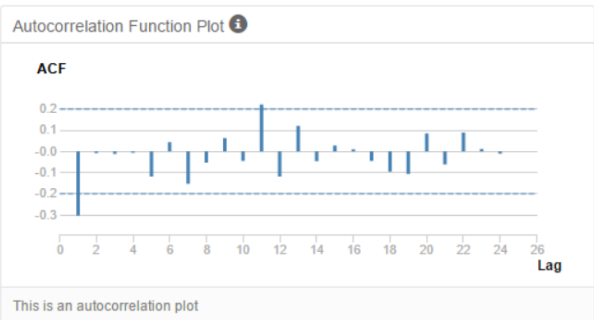
From the above graph, it can be noted that ACF is slowly decreasing towards 0 with seasonal increases at the Lags. This indicates a serial correlation, so we need to difference the series.

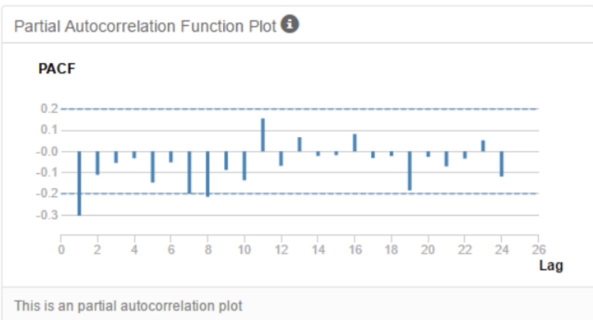
The below graph, with Seasonal difference, is still similar to the previous one. But, the correlation is lot lesser, so we will do the Seasonal first differencing.





After the Seasonal first differencing, we see that the significant correlation is removed.





With the Seasonal first differencing we also get a stationary time series as shown below.

The ACF lag-1 term is negative and then a sharp cut-off is seen so p=0 and q=1, therefore add an MA term. For seasonal components, the ACF lag-1 is negative so we can skip s-AR term or P=0. All seasonal lags (12,24) do not show a spike so we can skip adding an s-MA term also or Q=0. We used seasonal and non-seasonal differencing or d=0 and D=1.

ARIMA)(0,1,1)(0,1,0)12 will be the chosen model.

Information Criteria:

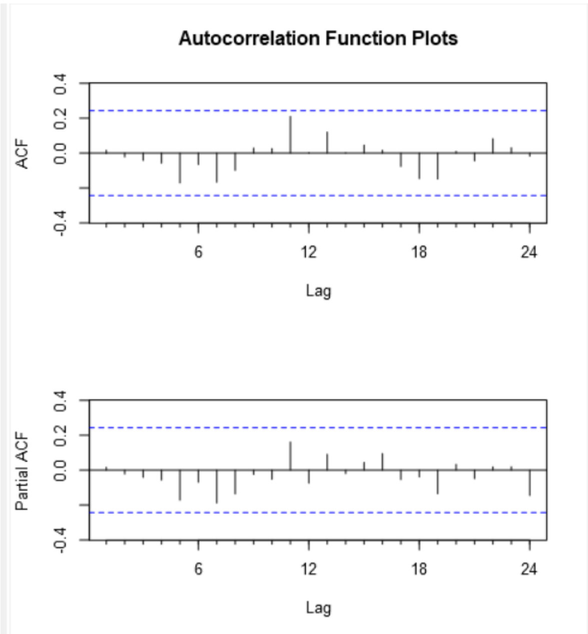
AIC	AICc	BIC
1256.5967	1256.8416	1260.4992

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

The **RMSE** and **MASE** values obtained from the in-sample error measures for the ARIMA (0,1,1)(0,1,0)12 are **36761.5** and **0.365** respectively.

The lower the RMSE, the better fit is the model. The fit is determined by looking at what is the predicted value using the model and what was actually observed.
MASE is a relative measure of error applicable only to time series data. It measures relative reduction in error compared to a naïve model. MASE has a threshold of goodness which is one so models close to one or above one are not considered very reliable. Both the dampened and the undampened model have MASE values significantly lower than one, which means both the models are good.



Forecast

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

ETS

<u>Non-Dampened ETS</u>

Method:
ETS(M,Ad,M)

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

<u>Dampened ETS</u>

Method:
ETS(M,A,M)

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

ARIMA

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

The ARIMA model has smaller absolute values in most of the metrics. The ARIMA model appears to produce smaller predictive errors in more metrics than the ETS model. RMSE is one metric where ETS model gives a lower value than ARIMA model, but it is not significantly large. From the above tables, we can gather that overall the ARIMA model misses its forecast by a lesser amount than the ETS model.

Actual and Forecast Values:

Actual	ETS_Model	ARIMA_Model
271000	255966.17855	263228.48013
329000	350001.90227	316228.48013
401000	456886.11249	372228.48013
553000	656414.09775	493228.48013

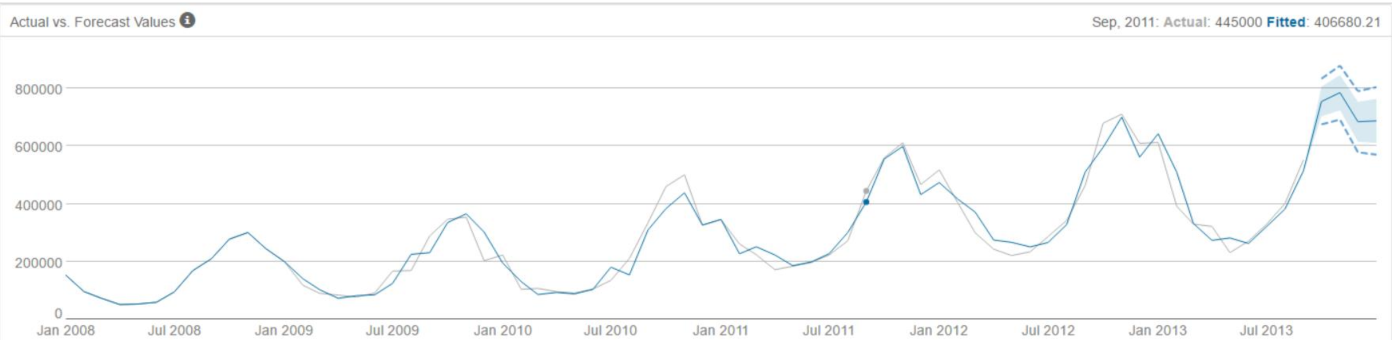
Accuracy Measures:

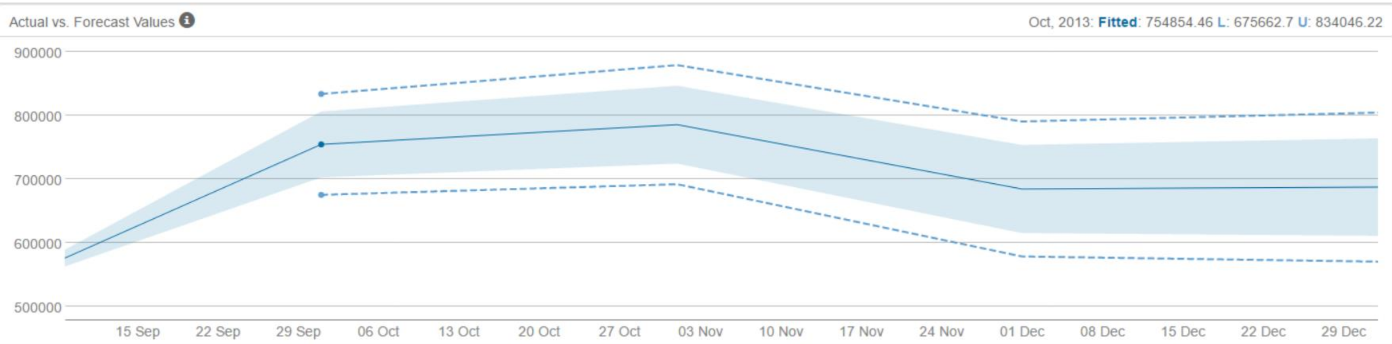
Model	ME	RMSE	MAE	MPE	MAPE	MASE	NA
ETS_Model	-41317.07	60176.47	48833.98	-8.3683	11.1421	0.8116	NA
ARIMA_Model	27271.52	33999.79	27271.52	6.1833	6.1833	0.4532	NA

ARIMA model has lower forecast accuracy measure MASE value than ETS model. AIC value of ARIMA is also less than ETS model. Also, predicted values of ARIMA model is closer to the actuals of the holdout sample than the ETS model. Therefore, ARIMA model is better to forecast the result.

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

Period	Sub_Period	Final_Forecast	Final_Forecast_high_95	Final_Forecast_high_80	Final_Forecast_low_80	Final_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 actual values are shown in gray and the forecasted values in blue. The shaded light blue region in the plot shows the 95% confidence interval, and the dotted dark blue lines show the 80% confidence interval.