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.**

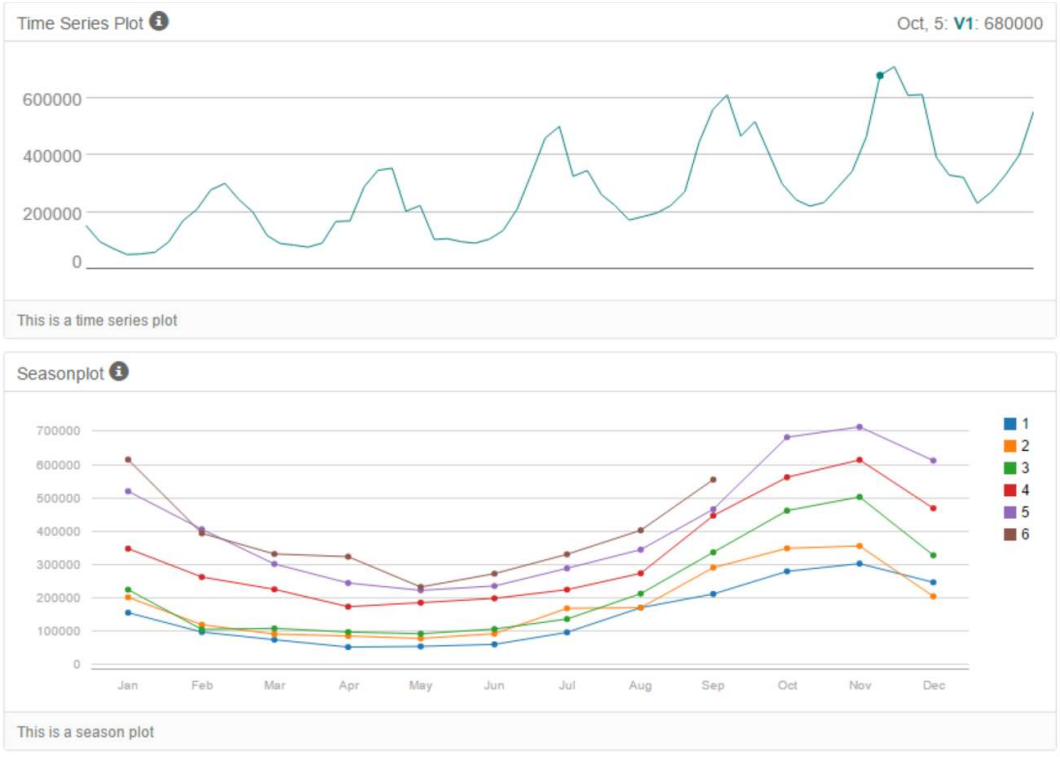
   We have at most one datapoint for each time measurement as well as a sequence to the data, it is ordered by the time, the data covers a continuous time interval and there is an equal spacing between each time measurement.

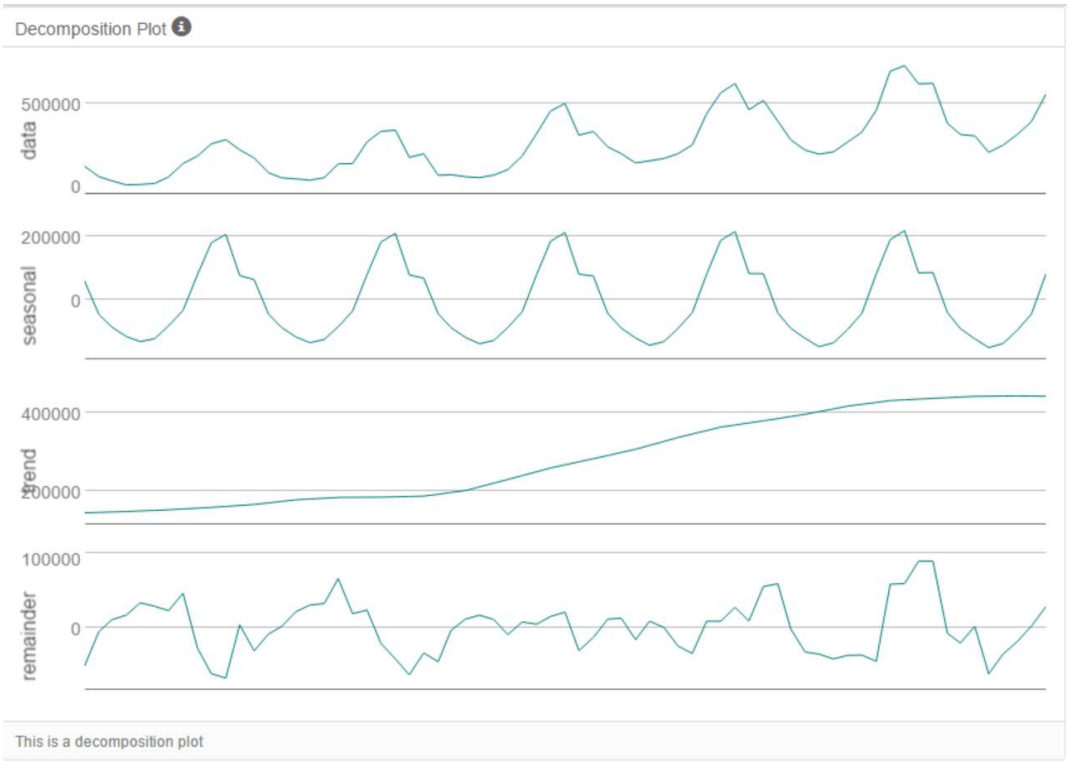
2. **Which records should be used as the holdout sample?**

   The holdout sample depends on the time frame we must predict. Because we will predict the sales for the next 4 months, the holdout sample shall be built by using the data of the least 4 months in the dataset.

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.**





The time series and decomposition plot are created using the TS-Plot function. Seasonality shows a small growth in volatility over time thus we should use multiplication for seasonality. The trend seems like a linear development which corresponds to an additive component. There is no visible trend in the graph for the error, but the volatility is lightly higher in the end, so we should use a multiplicative component.

Step 3: 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 with a holdout sample of 4 months.

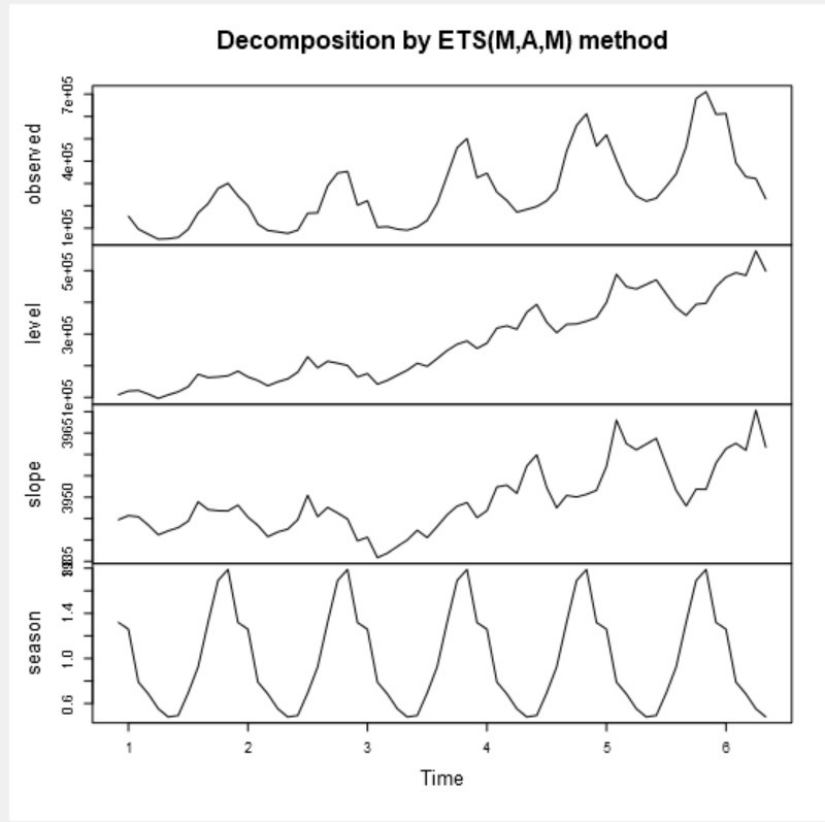
   Non-dampened ETS model:

   We can see that the AIC value is 1639.74, RMSE (Moot Mean Square Error) is 32992.73 and MASE (Mean Absolute Percentage Error) is 0.3727.

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



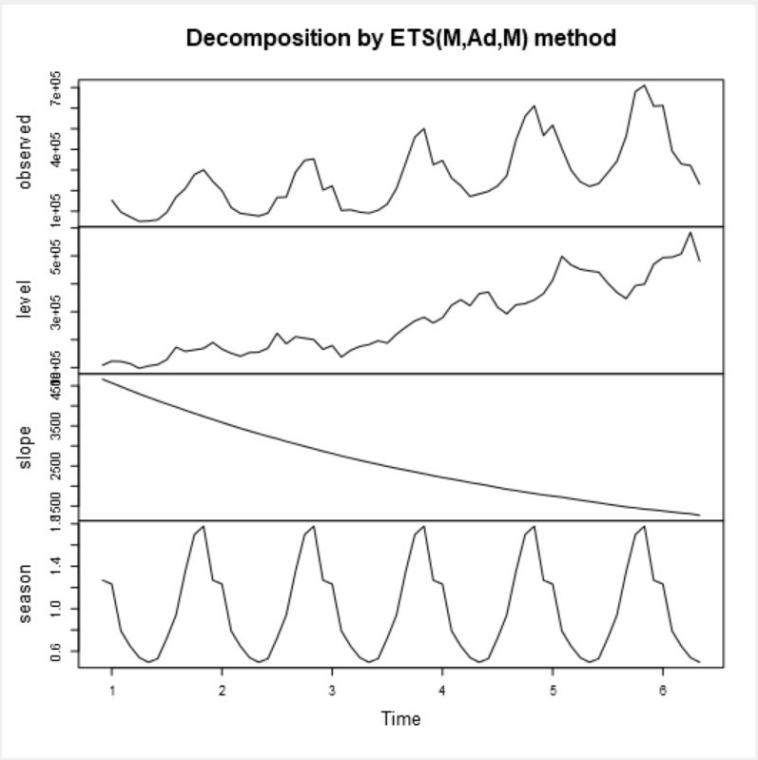
Dampened ETS Model:

We can see that the AIC value is 1639.47, RMSE (Moot Mean Square Error) is 33153.53 and MASE (Mean Absolute Percentage Error) is 0.3675.

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



Let's look at the actual forecasts for dampened and non-dampened ETS Models.

Non-dampened ETS Model:

Actual and Forecast Values:

Actual	ETS_M_A_M_
271000	248063.01908
329000	351306.93837
401000	471888.58168
553000	679154.7895

Accuracy Measures:

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

Dampened ETS Model:

Actual and Forecast Values:

Actual	ETS_M_A_M_
271000	255966.17855
329000	350001.90227
401000	456886.11249
553000	656414.09775

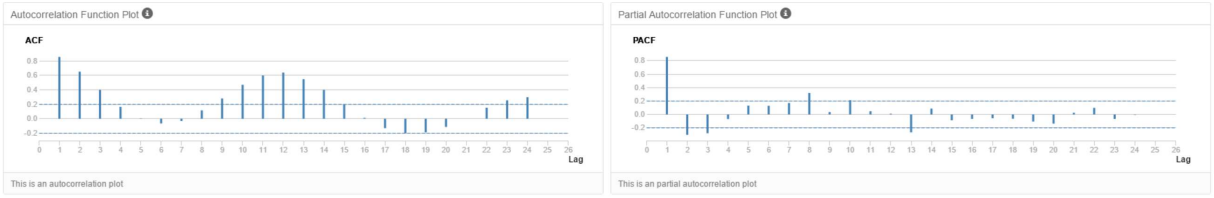
Accuracy Measures:

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

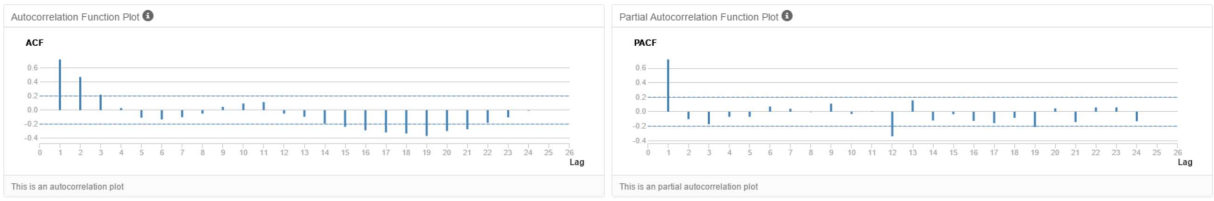
If we compare the forecast and the actual results, it seems that, the dampened model should be chosen due to its higher accuracy. The dampened model's RMSE & MASE are lower and should offset its slightly lower AIC.

ARIMA Model

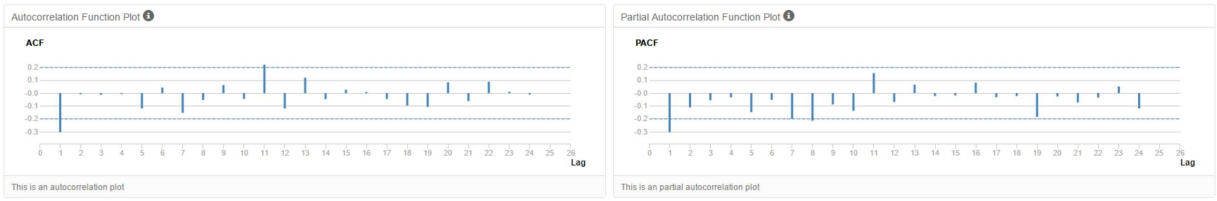
First, we look at the auto-correlation function (ACF) and the partial auto-correlation function (PACF) without differencing the dataset. We see that, the ACF shows a high correlation and the PACF shows significant lags in the periods 2,3 and 13 which could due to seasonal effects.



After producing a first seasonal differencing, the ACF still shows a high correlation. The PACF looks better but we should difference the data one more time.



We can now see that the ACF and PACF shows a stationarized data.



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

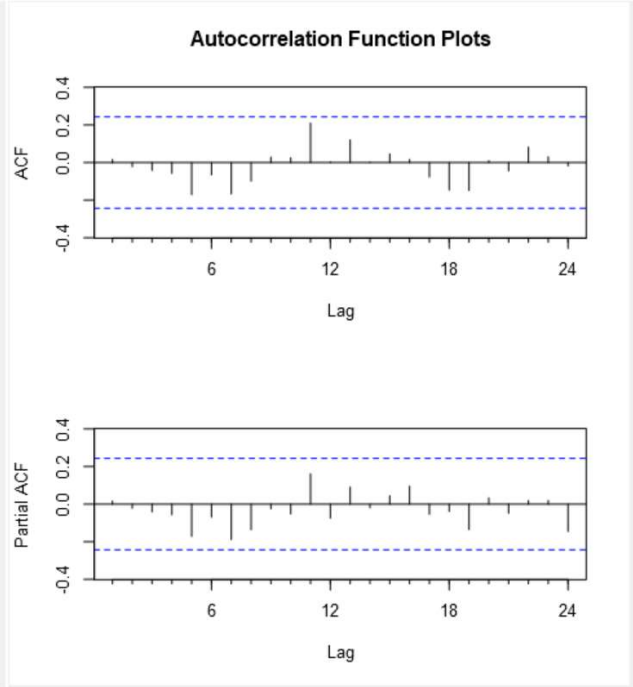
Information Criteria:

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



The ACF and PACF plots from the ARIMA (0,1,1)(0,1,0)12 Model show no significant correlation. This suggests that no other AR or MA terms are needed.

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.**

ARIMIA 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

ETS Model:

Actual and Forecast Values:

Actual	ETS_M_A_M_
271000	255966.17855
329000	350001.90227
401000	456886.11249
553000	656414.09775

Accuracy Measures:

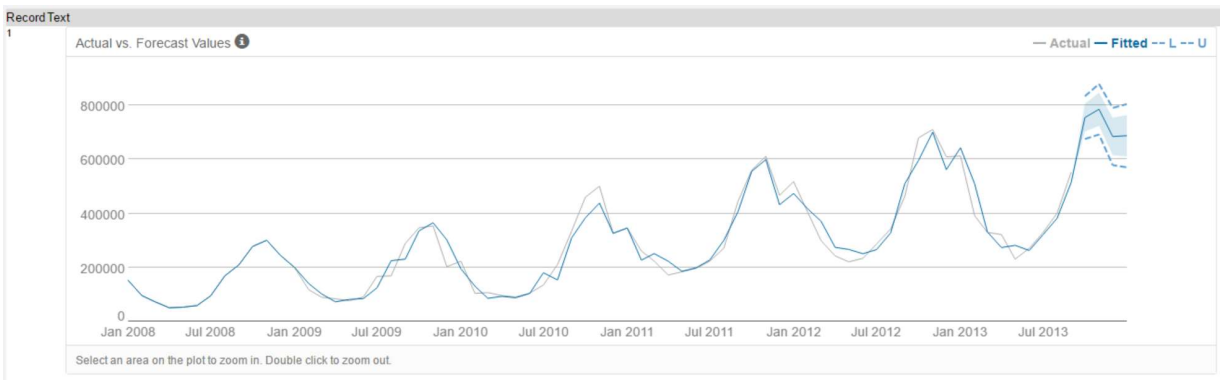
	Model	ME	RMSE	MAE	MPE	MAPE	MASE	NA
ETS_M_A_M_		-41317.07	60176.47	48833.98	-8.3683	11.1421	0.8116	NA

The ARIMA model is better at the forecasting of monthly sales compared to the ETS model using the holdout dataset as a validation set. The MAPE and ME values are lower using the ARIMA model. The RMSE for the ARIMA model is 33999.79 compared to RMSE at 60176.47 for the ETS model. The MASE value of 0.4532 for the ARIMA is also lower than the MASE value of 0.8116 for the ETS. The ARIMA model should be the better choice 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 are 754,854, 785,854, 684,654 and 687,854.

Period	Sub_Period	Prognose	Prognose_high_95	Prognose_high_80	Prognose_low_80	Prognose_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



Before you Submit

Please check your answers against the requirements of the project dictated by the rubric here. Reviewers will use this rubric to grade your project.