Time Series 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.
   Yes, the dataset meets all criteria of time series dataset, because the dataset covers a continuous time interval, have equal spacing between every two consecutive measurements, and each time unit within the time interval has at most one data point.

2. Which records should be used as the holdout sample?
   The length of the holdout sample should be at least same with the periods we are forecasting, in this project, which is 4 months, so the records from 2013-06 to 2013-09 should be used as holdout sample.

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.

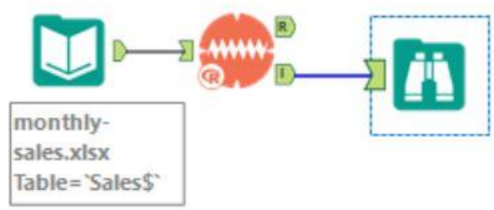
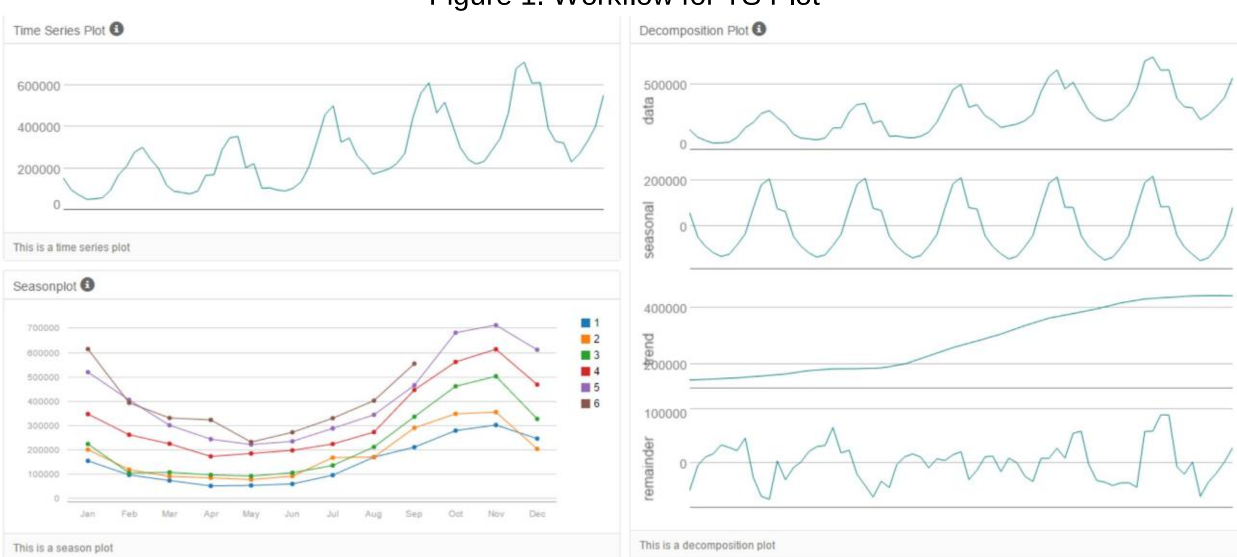


Figure 1. Workflow for TS Plot



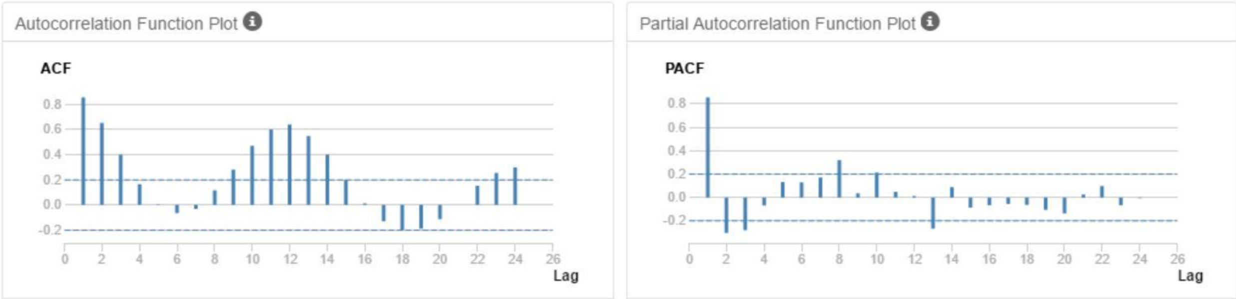


Figure 2. TS Plot

TS Plot tool were used for viewing the trend, seasonality, and error of the dataset.
The figure 2 indicate that the seasonality of the data is increasing slightly, so the multiplication should apply. The trend is obviously increasing linearly, so the addition should apply. The error shows no obvious trend but fluctuate over time, so the multiplication should apply.

Step 3: Build your Models

Analyze your graphs and determine the appropriate measurements to apply to your ARIMA and ETS models and describe the errors for both models. (500 word limit)

Answer these questions:

ETS Model

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 model term for the ETS model are multiplication, addition, multiplication (M,A,M) determined by the above figure 2. TS Plot result.

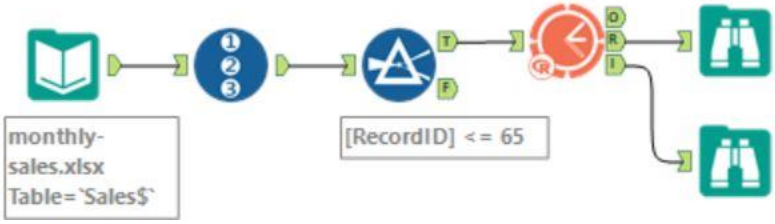


Figure 3. Workflow for building an ETS Model

With Trend Dampening:

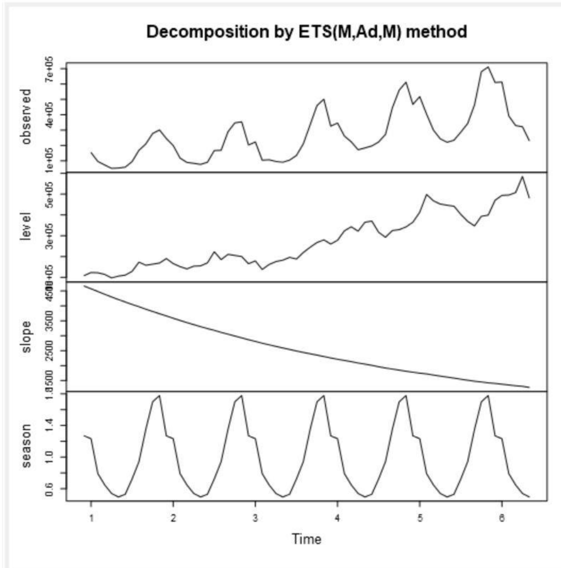


Figure 4. Decomposition by ETS(M,A,M) with Dampening

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

Figure 5. Errors and AIC (Dampening)

For the ETS(M,A,M) with trend dampened, the RMSE(root mean square error) is **33153.52** and the MASE(mean absolute squared error) is **0.3675**. AIC is **1639.465**

With Trend Non-Dampening:

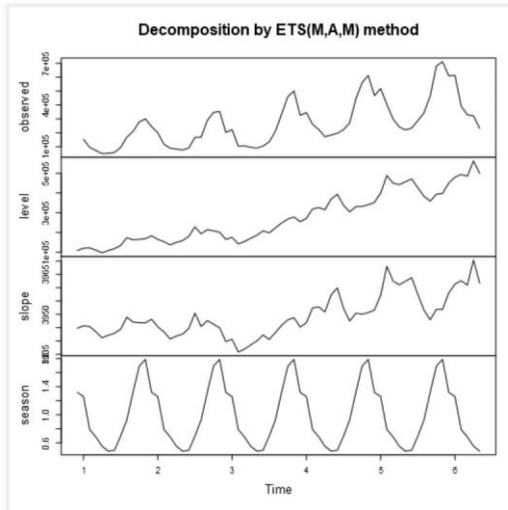


Figure 6. Decomposition by ETS (M,A,M) with Non-Dampening

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

Figure 7. Errors and AIC (Non-Dampening)

ETS(M,A,M) without dampening, the RMSE is **32992.72**, and MASE is **0.3726**, while AIC is **1639.73**

Validate the forecast:

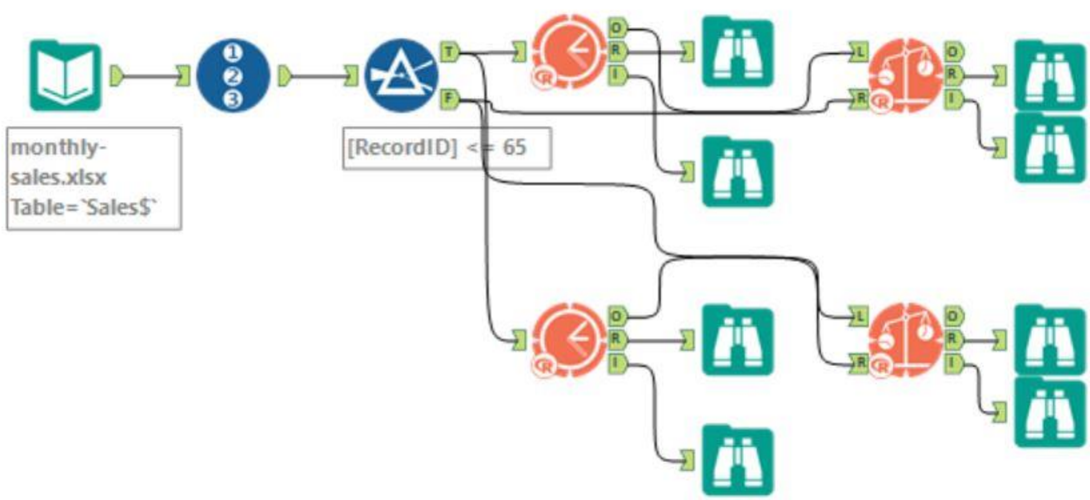


Figure 8. Workflow for ETS Model Validation

Actual and Forecast Values:

Actual	ETS_Model_Dampen
271000	255966.17855
329000	350001.90227
401000	456886.11249
553000	656414.09775

Accuracy Measures:

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

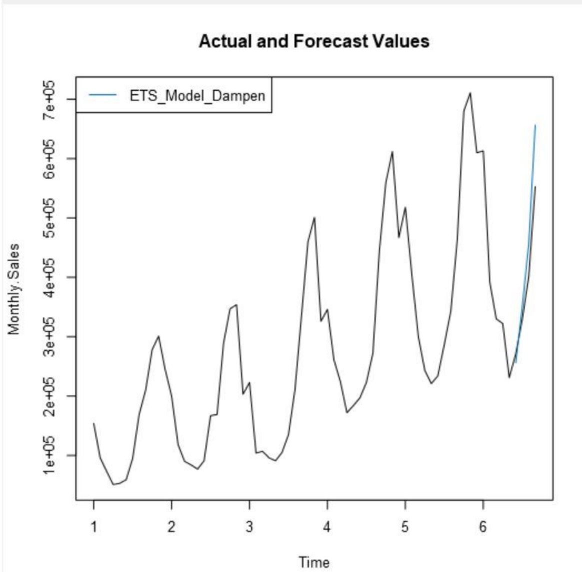


Figure 9. Validation Report (Dampen)

Actual and Forecast Values:

Actual	ETS_Model_NonDampen
271000	248063.01908
329000	351306.93837
401000	471888.58168
553000	679154.7895

Accuracy Measures:

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

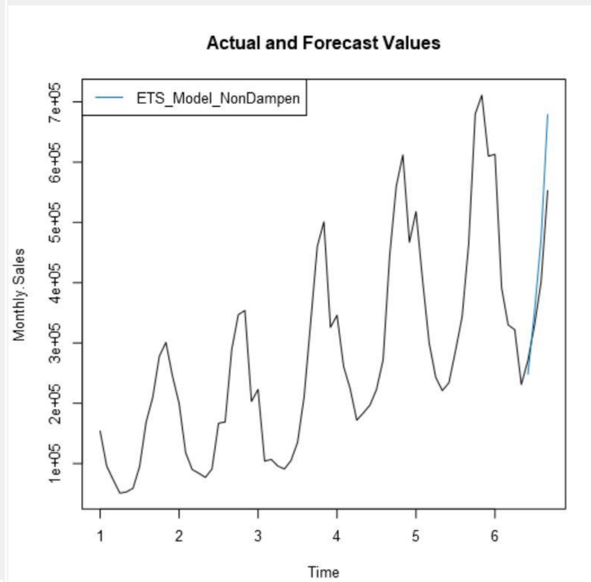


Figure 10. Validation Report (Non-Dampen)

From figure 9 and figure 10, the ETS model with trend dampening has lower errors compares to the one without trend dampening. Because lower errors could lead to lower AIC.

ARIMA Model:

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.

From figure 2 TS Plot, trend and seasonality are observed, so the data need to be stationary at first.

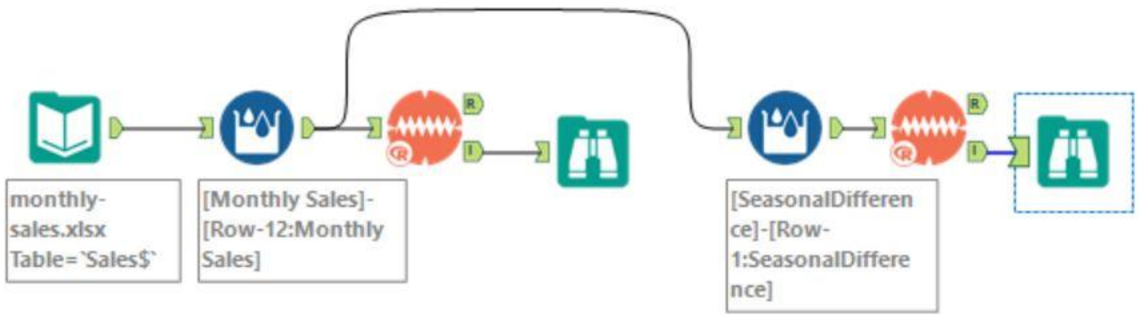


Figure 11. Workflow for Seasonal Differencing

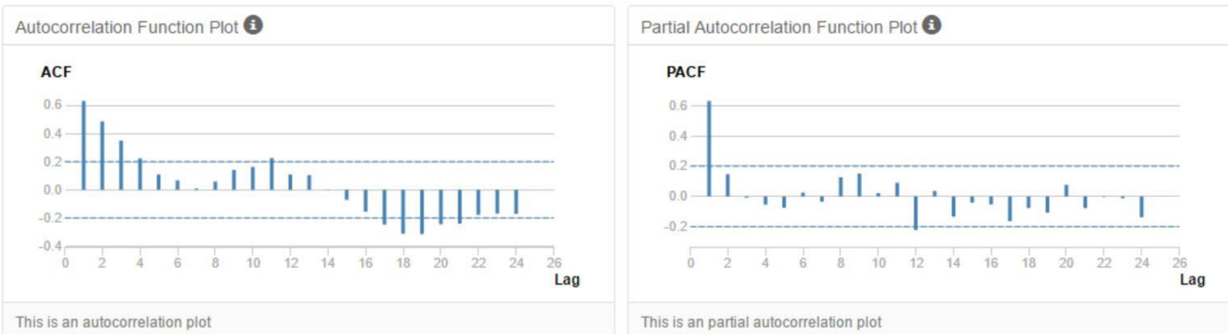


Figure 12. Differencing 1

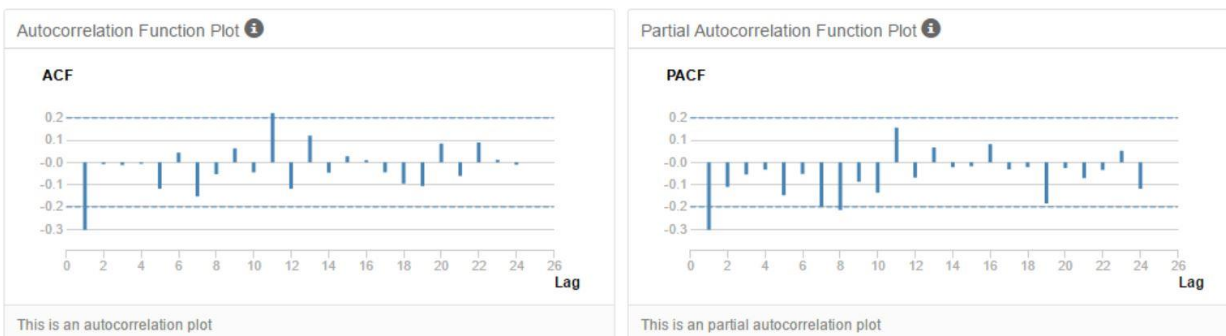


Figure 13. Differencing 2

From the figure 12, the seasonal differencing hasn't adjusted for the effect of seasonality, so the first seasonal differencing need to be taken to correct the first seasonality to make the dataset stationary. Figure 13 indicate that the series is stationary. The serial correlation has now disappeared.

From the figure 13, the ACF and PACF plot, since at lag-1, the autocorrelation is negative, and there is only 1 seasonal differencing, and no seasonal correlations anymore, so the ARIMA(0,1,1)(0,1,0)12 is chosen.

ARIMA (3)	ARIMA: Error in plot.window(...) : need finite 'ylim' values
ARIMA (3)	ARIMA: Execution halted
ARIMA (3)	ARIMA: The R.exe exit code (1) indicated an error.

System Error

Project Ongoing