Project: Forecasting Sales

Step 1: Plan Your Analysis

1. Does the dataset meet the criteria of a time series dataset?
   In order for a dataset to meet the criteria of a time series dataset the dataset have the following characteristics:
   a. The dataset series is spread over a continuous time interval – this dataset meets this characteristic given the data are spread over the period from Jan 2008 through September 2013
   b. The dataset has sequential measurement across the interval – this dataset shows monthly sales across the entire interval
   c. The dataset has equal spacing between every two consecutive measurements – this dataset has a one-month time interval between every measurement
   d. Every time unit within the interval has at most one datapoint – this dataset has monthly sales data present for every month
2. Which records should be used as the holdout sample?
   Since we were asked to forecast the next four months of sales we will withhold the following four months' data as the sample:
   a. June 2013
   b. July 2013
   c. August 2013
   d. September 2013

Step 2: Determine Trend, Seasonal, and Error components

1. What are the trend, seasonality, and error of the time series?
   a. Seasonal plots show peaks and valleys for seasonality with a gradual increase over time indicating use of a multiplicative component.
   b. Trend plot shows a linear relationship and therefore this will indicate an additive component being utilized.
   c. Given that the remainder (error) plot shows a significant amount of variation over time a multiplicative component will be used.

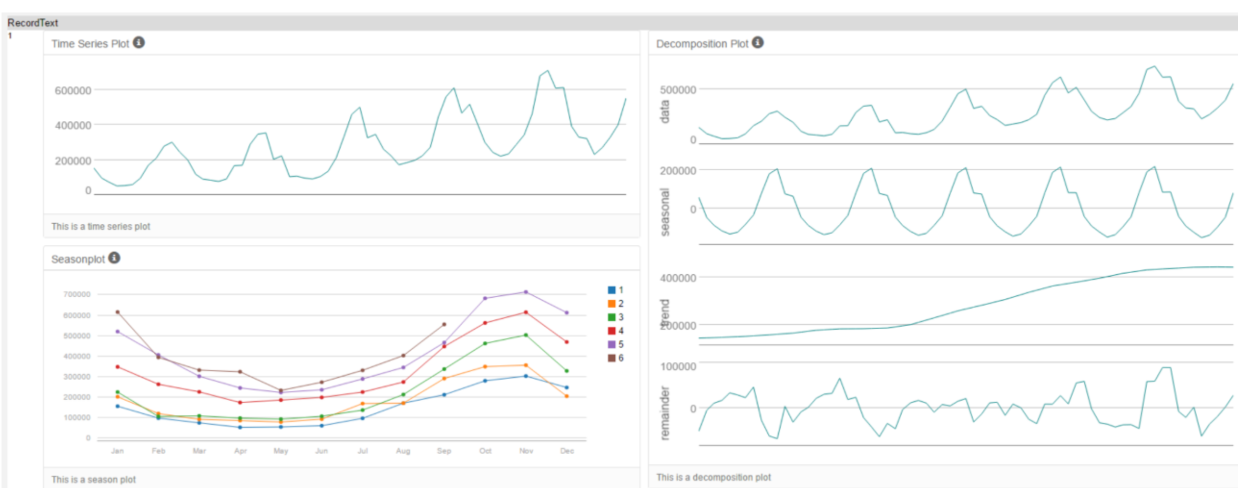

Figure 1: Trend, Seasonal and Error Plots

Step 3: Build your Models

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 terms for the ETS model are M,A,M as describe above given the associated decomposition plot. Additionally, a dampened and non-dampened ETS model with a 4-period holdout sample was created.

ETS (M,A,M) non-Dampened

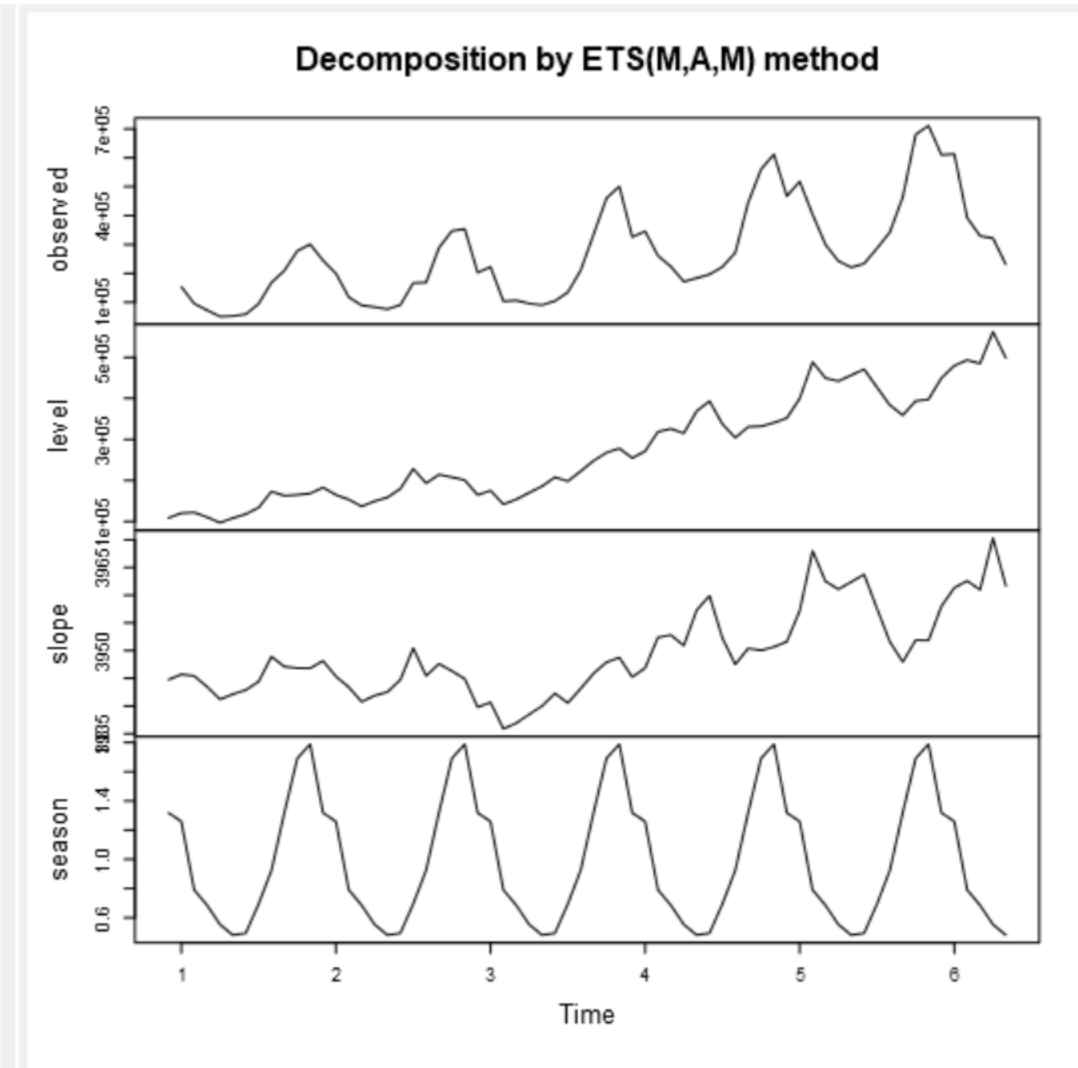

Figure 2: ETS (M,A,M) Non-Dampened

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 3: Sample Errors Non-Dampened ETS (M,A,M)

Actual and Forecast Values:

Actual	ETS_VideoDemand_MAM
271000	248063.01908
329000	351306.93837
401000	471888.58168
553000	679154.7895

Accuracy Measures:

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

Figure 4: Actual vs. Forecasted (non-dampened)

The non-dampened ETS (M,A,M) model gives us an AIC of 1639.7, RMSE (root mean squared error) of 74101.16, and a MASE (mean average scaled error) of 1.01.

ETS (M,A,M) Dampened

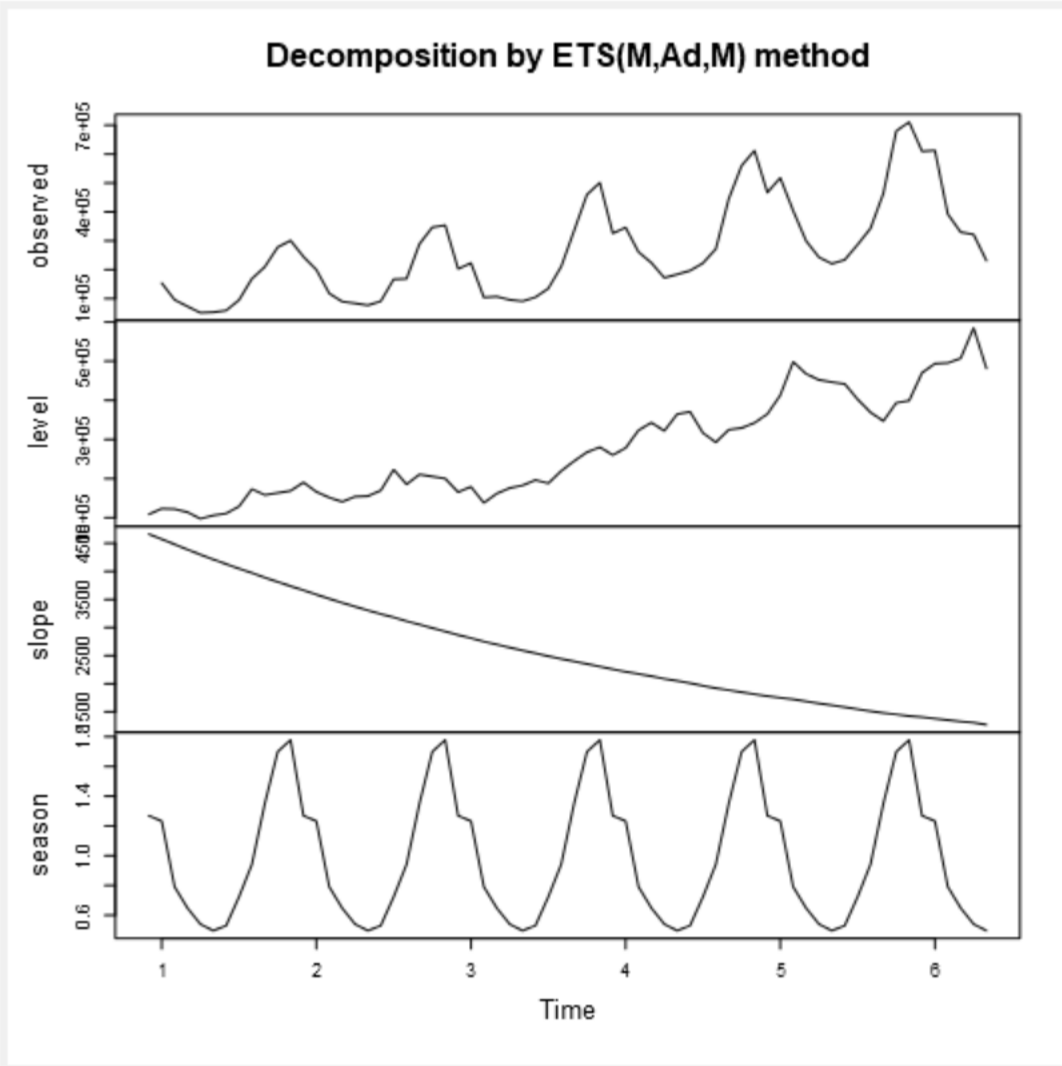

Figure 5:ETS (M,A,M) Dampened

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 6: Sample Errors Dampened ETS (M,A,M)

Actual and Forecast Values:

Actual	ETS_VideoDemand_MAM_damp
271000	255966.17855
329000	350001.90227
401000	456886.11249
553000	656414.09775

Accuracy Measures:

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

Figure 7: Actual vs. Forecasted (dampened)

The dampened ETS (M,A,M) model gives us an AIC of 1639.5, RMSE (root mean squared error) of 60176.5, and a MASE (mean absolute scaled error) of .817

Comparing the forecast values and accuracy measures from both the dampened ETS model and the non-dampened ETS model we see that the dampened model gives better accuracy against the holdout sample. In particular, given a lower mean absolute scaled error when we measure performance against the holdout sample for the dampened ETS of .8116 versus 1.0066 of the non-dampened model we should choose the ETS (M,A,M) Dampened model.

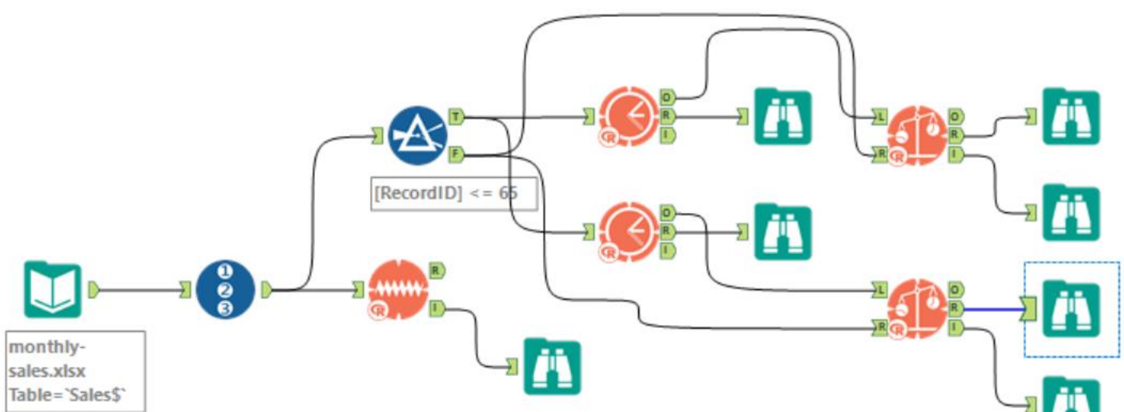

Figure 8: Alteryx ETS Model Workflow

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.

We know form our initial time series plot in figure 1 that the series is not stationary. The ACF and PACF plots are shown below are without differencing. We can see the time series and seasonal component's ACF shows high correlation and the PACF shows a significant lag at period 13. This is due to a seasonal effect

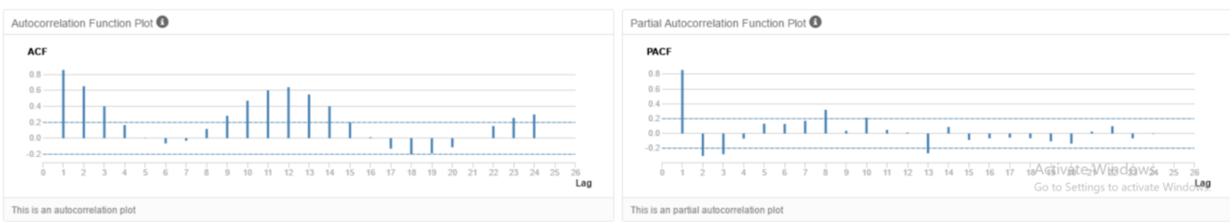

Figure 9: ACF/PACF no differencing plots

Next a seasonal difference is taken and the results shown below. We see in the time series plot that the series is still not stationary. And again, there is high correlation in the ACF plot, but this time we see that the high correlation in the PACF plot is gone.

Figure 10: Seasonal difference time series plot

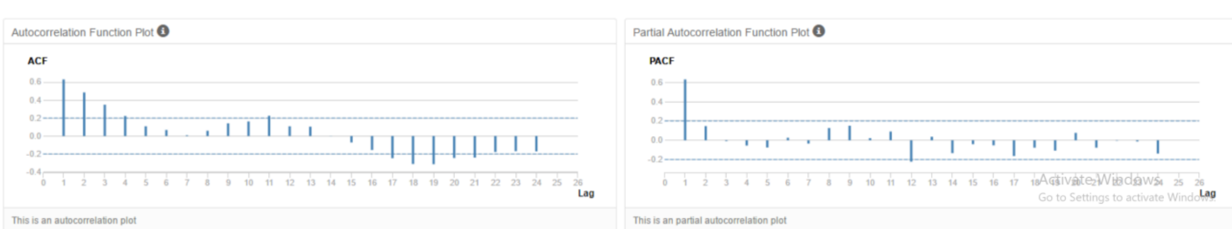

Figure 11: ACF/PACF seasonal differencing plots

Continuing on, we complete a seasonal first difference. The ACF plot no longer shows the strong correlation and all the significant lags are smoothened and we see a stationary time series plot. There will be no further need for differencing.

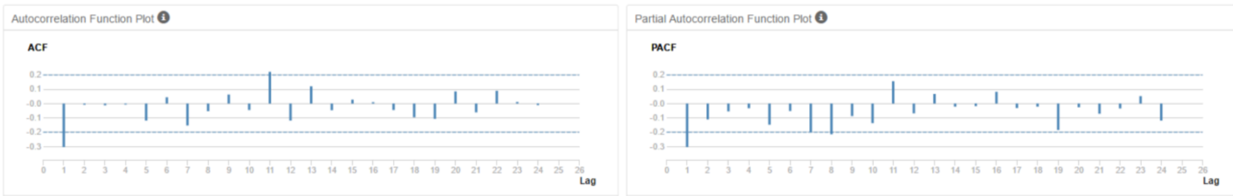

Figure 12: ACF/PACF seasonal first difference plots

Figure 13: Seasonal first difference time series plot

Given the ARIMA model ARIMA(p,d,q)(P,D,Q)m, we should follow 0, 1, 1 for the non-seasonal components and 0, 1, 0 for the seasonal components with 12 periods giving us the following notation ARIMA (0,1,1)(0,1,0)12 due to lag -1 being negative.

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

   The Root Mean Squared Error of the model is 36761.53 and the Mean Absolute Squared Error is .3646 and is at an acceptable level being less than one.

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

Figure 14: Measure of in-sample errors

b. Regraph ACF and PACF for both the Time Series and Seasonal Difference and include these graphs in your answer.

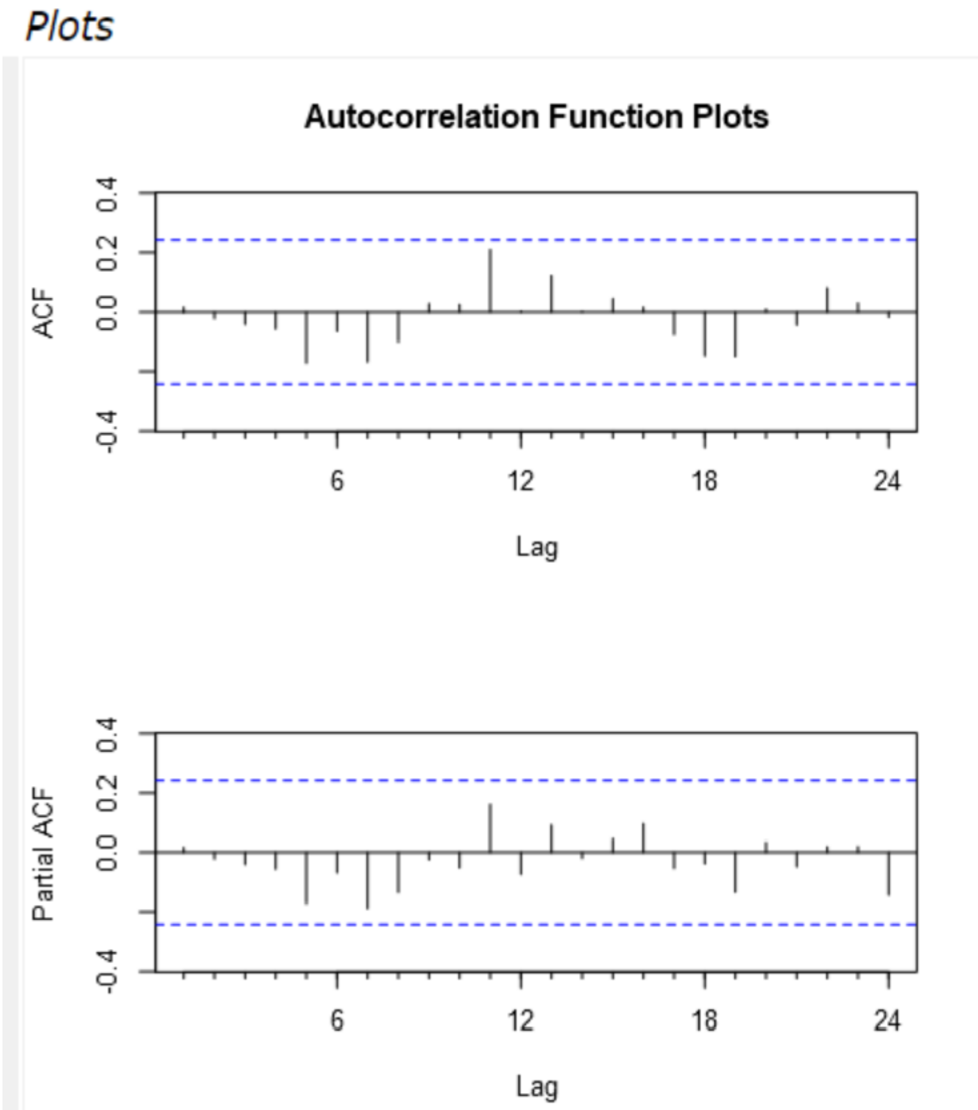

Figure 15: ACF/PACF of ARIMA model

Step 4: Forecast

Actual and Forecast Values:

Actual	Video_ARIMA	Video_ETS_damp
271000	263228.48013	255966.17855
329000	316228.48013	350001.90227
401000	372228.48013	456886.11249
553000	493228.48013	656414.09775

Accuracy Measures:

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

Figure 16: Model Comparison (In-sample errors and accuracy measures)

1. Which model did you choose? Justify your answer by showing: in-sample error measurements and forecast error measurements against the holdout sample.
   Given the model comparison above, I chose to the ARIMA (0,1,1)(0,1,0)12 model over the dampened ETS model. The ARIMA model gave higher accuracy predicting monthly sales for the holdout samples. It also had a lower RMSE and a MASE that was much lower than the ETS model.

2. What is the forecast for the next four periods? Graph the results using 95% and 80% confidence intervals.
   The table below represents the forecast for the next four periods including the 95% and 80% confidence intervals. Also included is a zoomed in version of the forecast including confidence intervals.

Period	Sub_Period	forecast	forecast_high_95	forecast_high_80	forecast_low_80	forecast_low_95
6	10	754854.5	834046.2159	806635.166	703073.7541	675662.7041
6	11	785854.5	879377.7531	847006.0545	724702.8656	692331.167
6	12	684854.5	790787.8282	754120.5664	615588.3537	578921.0919
7	1	687854.5	804889.2866	764379.4199	611329.5002	570819.6335

Figure 17: Forecast Table next four periods

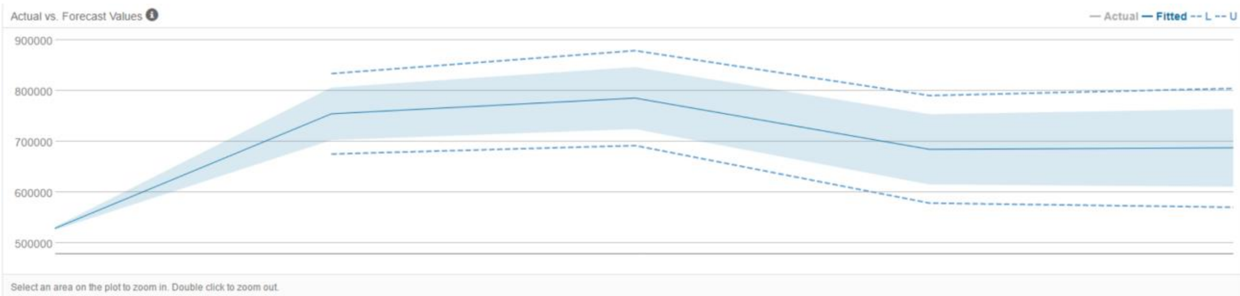

Figure 18: Forecast Graphic next four periods

Workflows:

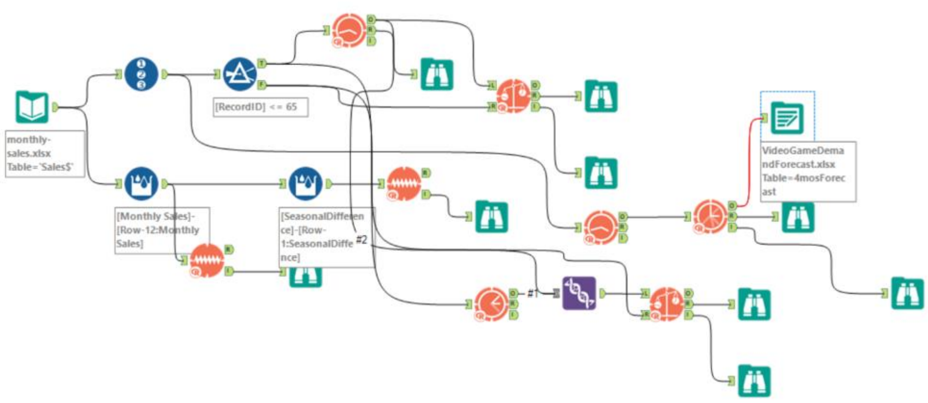

Figure 19: Complete Workflow

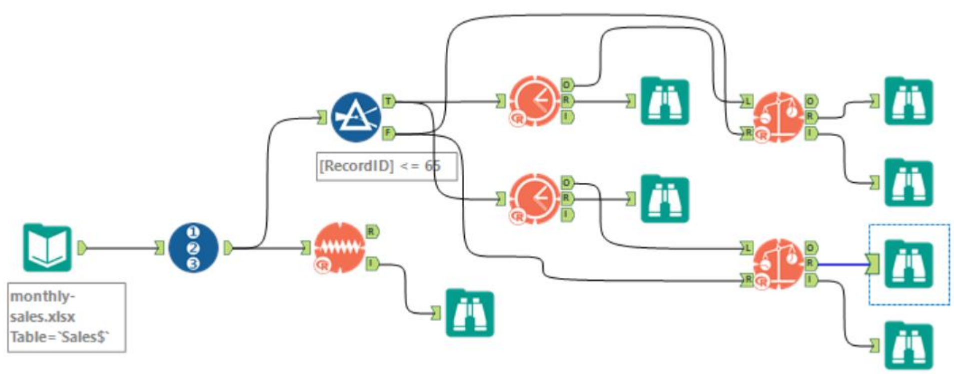

Figure 20: ETS Workflow