Project: Forecasting Sale

(This assignment was a part of my Predictive Analytics for Business Nanodegree at Udacity. Acting as a supply chain analyst, and given five years of data, I was tasked with forecasting product sales for the next four months.)

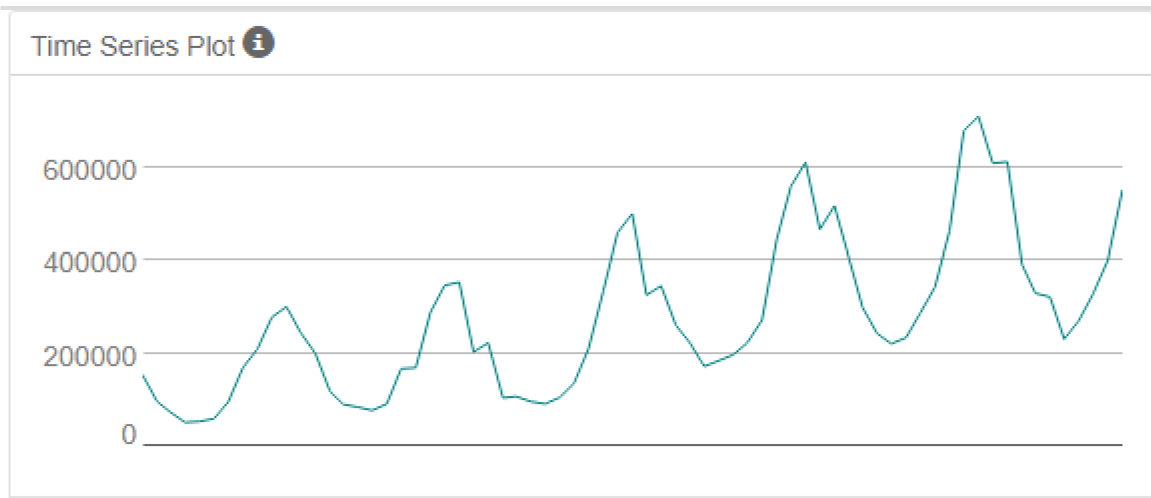
Analysis Planning

The monthly_sales spreadsheet comprises a time series dataset. There is continuous coverage of a period of nearly five years, and the reported intervals are consistent. There is a single value for sales for each month. When these values are plotted, and the plots are decomposed, we find trend, seasonality, and error.

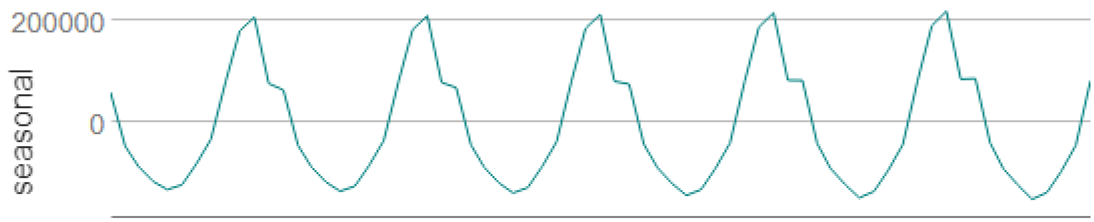
The holdout sample will be the four most recent data points.

Determining Trend, Seasonal, and Error components

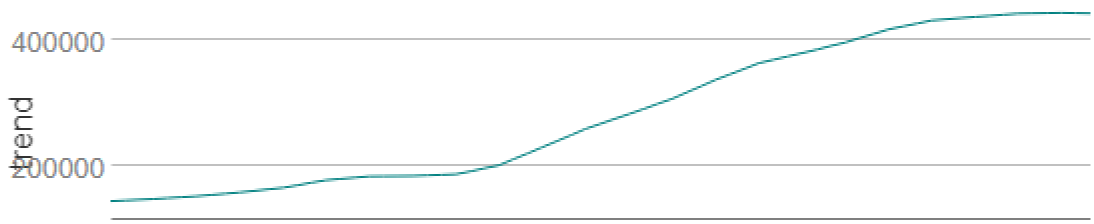
The time series plot of the data (please see below) can be decomposed to investigate its seasonality, trend, and error aspects.



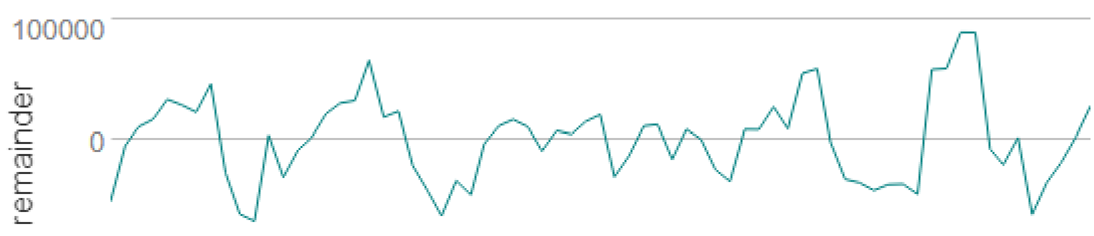
The data has a strong seasonality (please see graph below). Monthly sales hit their lowest values every May and peak every November. The annual highs are consistently increasing by $3,000.



As seen in the graph below, monthly sales have a linear upward trend.



The remainder graph (below) shows that the error measure has a variance that changes over time.

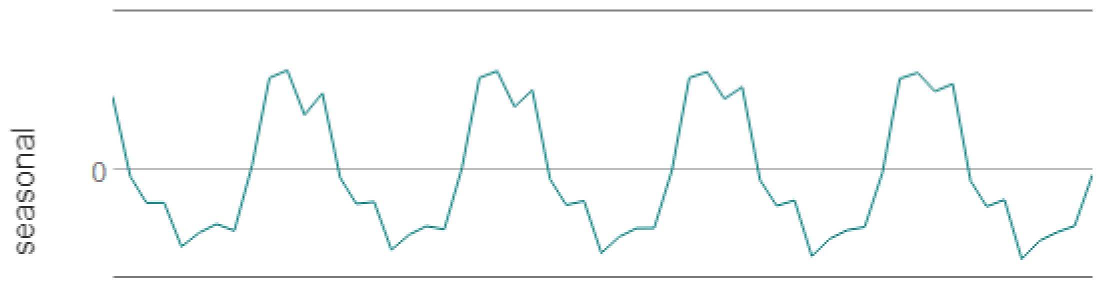


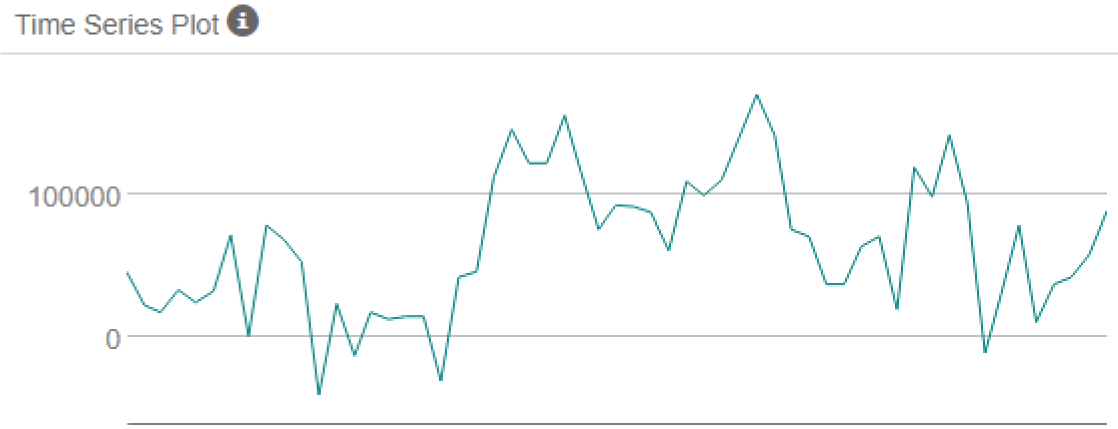
Building the Predictive Models

The ETS model used is MAA. The dataset's Error plot shows changing variance, so Error is applied multiplicatively. Trend is applied additively, because the upward trend is linear. Seasonality does increase over time, but it does so at a consistent rate of $3,000 per year, so Seasonality is applied additively.

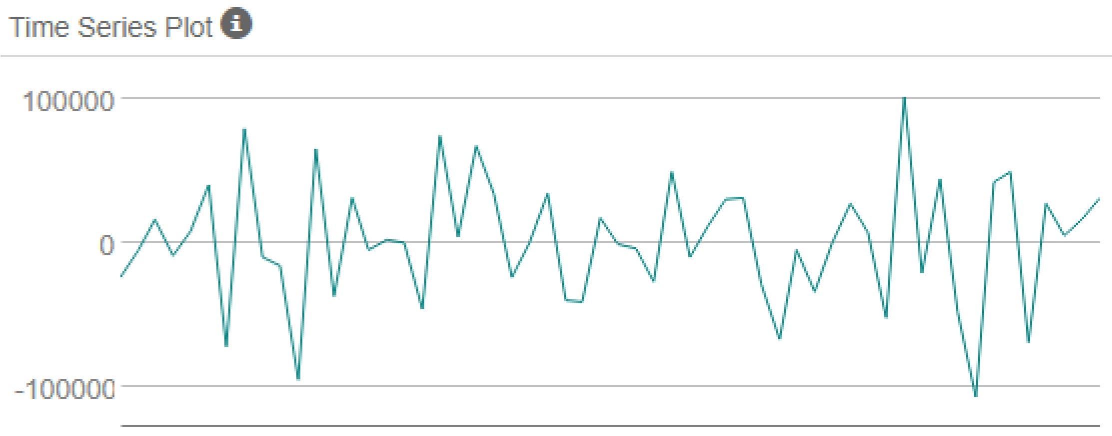
In-sample RMSE is 48,206. This means the standard deviation of the differences between predicted and actual values is equal to 17% of the average monthly sales. This model's MASE is 0.528.

The ARIMA model used is (0,1,1)(0,1,1). Seasonal differencing is done because the data has a strong seasonal component. While this makes seasonality stationary, our time series plot still is not. (Please see the two graphs that follow.)





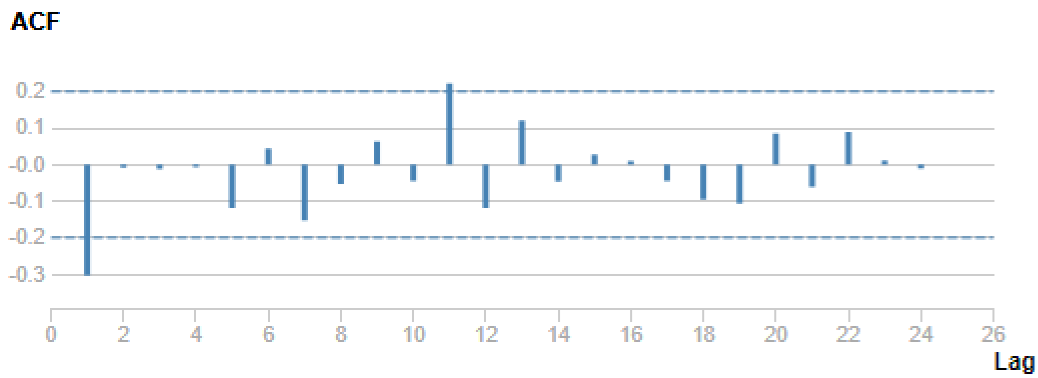
After applying non-seasonal differencing, the time series plot is stationary (graphic follows).



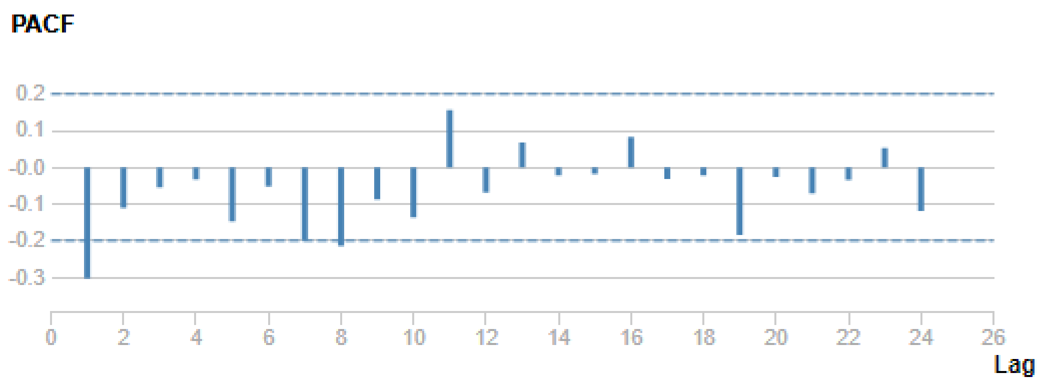
With differencing completed, the ACF and PACF are examined. The resulting graphs follow this paragraph. Both plots show negative autocorrelation at lag 1, indicating using non-seasonal MA.

Similarly, both plots show negative autocorrelation at lag 12, indicating the use of seasonal MA. This gives the ARIMA model the terms (0,1,1)(0,1,1).

Autocorrelation Function Plot



Partial Autocorrelation Function Plot



The standard deviation of the difference between predicted and actual values, RMSE, for the ARIMA model is 36,758, roughly 13% of the average monthly sales. MASE for the ARIMA model is 0.365.

Forecast

The in-sample error measurements for the ETS model are below.

Method:
ETS(M,A,A)

In-sample error measures:

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
-3165.523166	48206.6800824	36215.8613811	-4.3453583	14.7897347	0.5283349	0.6440806

Information criteria:

AIC	AICc	BIC
1673.431	1686.4523	1710.3956

The in-sample error measurements for the ARIMA model are below.

Method: ARIMA(0,1,1)(0,1,1)[12]

In-sample error measures:

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
-358.1274828	36758.4027043	24996.5435416	-1.800917	9.8272386	0.3646619	0.0166958

Information Criteria:

AIC	AICc	BIC
1258.5932	1259.0932	1264.447

The ARIMA model shows superior RMSE, MASE, and AIC in internal validations.

External validations against the holdout sample follow. First for the ETS model:

Accuracy Measures:

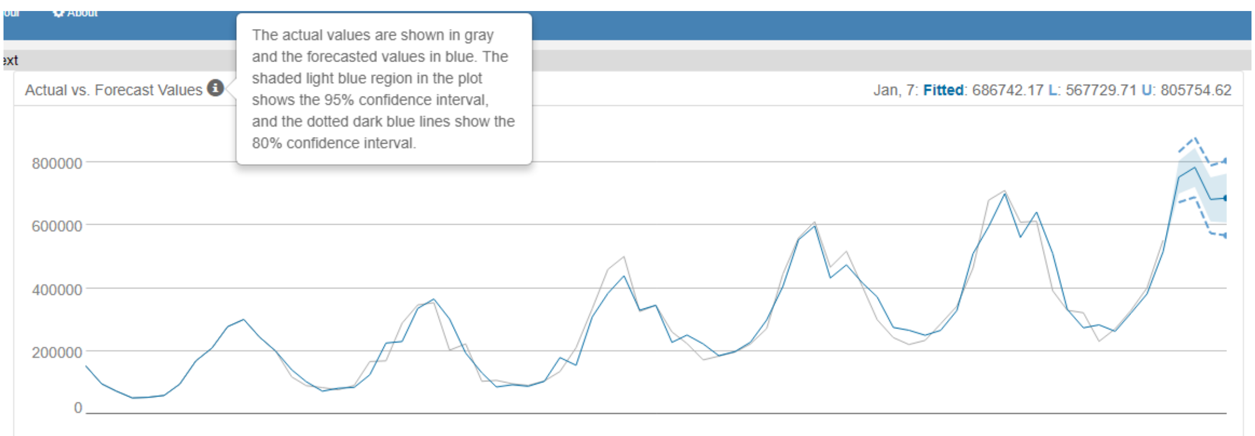
Model	ME	RMSE	MAE	MPE	MAPE	MASE
ETSmaa	-38267.74	46267.66	39285.6	-12.3056	12.4897	0.6529

Next, for the ARIMA model:

Accuracy Measures:

Model	ME	RMSE	MAE	MPE	MAPE	MASE
ARIMA_011011	27184.45	34010.92	27184.45	6.1547	6.1547	0.4518

Again, the ARIMA model outperforms the ETS model in RMSE and MASE, so the ARIMA model is used to forecast the four periods that follow the dataset. The results are plotted below.



Period	Sub_Period	forecast	forecast_high_95	forecast_high_80	forecast_low_80	forecast_low_95
6	10	753163.3696	833072.445371	805413.105801	700913.6334	673254.29383
6	11	784679.954091	879433.81916	846636.176286	722723.731896	689926.089022
6	12	682563.879889	790133.064606	752899.589201	612228.170577	574994.695172
7	1	686742.165916	805754.61928	764560.226417	608924.105416	567729.712552