ARIMA Models

Summary: ARIMA which stands for Autoregressive Integrated Moving Average helps you forecast data for seasonal and nonseasonal data

STEP 1: TIME SERIES DECOMPOSITION PLOT

A time series decomposition plot allows you to observe the seasonality, trend, and error/remainder terms of a time series.

Useful Alteryx tool: TS Plot

STEP 2: DETERMINE THE ARIMA TERMS

Nonseasonal ARIMA models are displayed in the terms (p,d,q) which stand for p - periods to lag for, d - number of transformations used to make the data stationary, q - lags of the error component

Stationary - mean and variance are constant over time vs Non-Stationary - mean and variance change over time

Differencing - take the value in the current period and subtract it by the value from the previous period. You might have to do this several times to make the data stationary. This is the Integrated component which is d in the model terms.

Autocorrelation - How correlated a time series is with its past values, if positive at Lag-1 then AR if negative then MA

Partial Autocorrelation - The correlation between 2 variables controlling for the values of another set of variables. If the partial autocorrelation drops of quickly then AR terms, if it slowly decays then MA

Seasonal ARIMA models are denoted (p,d,q)(P,D,Q)m

These models may require seasonal differencing in addition to non-seasonal differencing. Seasonal differencing is when you subtract the value from a year previous of the current value.

Useful Alteryx tool: TS Plot

STEP 3: BUILD AND VALIDATE THE ARIMA MODEL

Build the ARIMA model using the terms determined in step 2. You can use internal and external validation to validate the quality of the model.

Internal validation: Look at in-sample error measures, particularly RMSE (Root-Mean-Square Error) and MASE (Mean Absolute Scaled Error).

External validation: Determine the accuracy measures by comparing the forecasted values with the holdout sample. This is especially important for comparing ARIMA models to other types of models, such as ETS.

Pick the ARIMA model with lowest AIC value. If the AIC values are comparable, use calculated errors to pick one that minimizes error the most. Many software tools will automate the selection of the model by minimizing AIC.

Useful Alteryx tools: ARIMA, TS Compare

STEP 4: FORECAST!

Use the best ARIMA model to forecast for the desired time period. Make sure to add the holdout sample back into the model. Plot the results along with 80% and 95% confidence intervals.

Useful Alteryx tool: TS Forecast