* The data was prepared and analyzed to evaluate its potential for forecasting. Three response variables were tested: raw crash counts and two different denoised count variables.
* Multiple model types incorporating various combinations of variables were fitted for each response variable. These included ARIMA, TSLM, and neural network models, along with baseline models such as naïve, seasonal naïve, and drift models.
* Each model for every response variable was also fitted using varying train-test split proportions. The test set size was incrementally increased from 5% to 30% in 5% intervals. The root mean square error (RMSE) and residuals for each model were captured and analyzed.