

Project Summary

Studying the distribution of crash categories between resident and non-resident drivers, we find that the largest difference of case proportion appears in the crash types Lateral Move (Left/Right), Decelerating (Slowing), and Slower, Going Straight.

Analyzing the distribution of driver behavior factors and distractor factors, we conclude that compared to resident drivers, non-resident drivers are more likely to encounter fatal accidents that result from leaving the proper lane. Compared to resident drivers, non-resident drivers in accidents are more likely to be inattentive, distracted by people outside the vehicle, talking or listening on the phone.

Drivers unfamiliar with local traffic and environment tend to do lateral moving, deceleration and lane shifting. This makes non-resident drivers prone to crash, thereby contributing to the high fatal accident rate. Distraction is another reason for them being more likely to encounter crashes.

We find that the distribution of high-risk drivers across zip codes is similar to Pareto distribution, where the curve reaches 80% at 0.4, showing that few zip code areas contribute to most of the high-risk drivers.

Analyzing the population demographic of high-risk drivers, we find that most of the high-risk drivers are aged from 16 to 35. The possible reason might be that young drivers are less skilled and are more likely to engage in dangerous driving behavior. Also, males are predominant among all the high-risk drivers.

After analyzing the time series of the count of crashes over months, strong seasonality can be found. From February through August, the monthly total of crashes increase and peak in either July or August. We assume that the number of fatal accidents is related to temperature and precipitation. The higher the temperature, the more precipitation, the more fatal accidents. The possible reason is that fine weather leads to more outdoor activities, which cause accidents.

A random forest model with an accuracy of 60% is made to predict the crash category based on a set of features that describe environmental conditions. It can be applied to notify drivers when adverse road conditions appear.