Data

The data used in this analysis are traffic accident records collected by the Seattle Police Department for the years 2014-2018.(N= 194,673).   Cases labeled as property damage (N= 136,485) and personal injury (N= 58,188) were extracted. Here, we aim to gain insights into the mechanisms of serious road accidents by analyzing the characteristics linked to personal injury. The five features used for prediction are the address type of the accident location (intersection or not), day of the week (weekend or not), weather (clear or not), road conditions (dry or not), and daylight conditions (daylight or not).   These variables were examined for their recorded content, and binary-coded described as above (resulting in missing values being coded as 0). The relationships between the dependent variable, accident severity, and these features are graphed as follows.
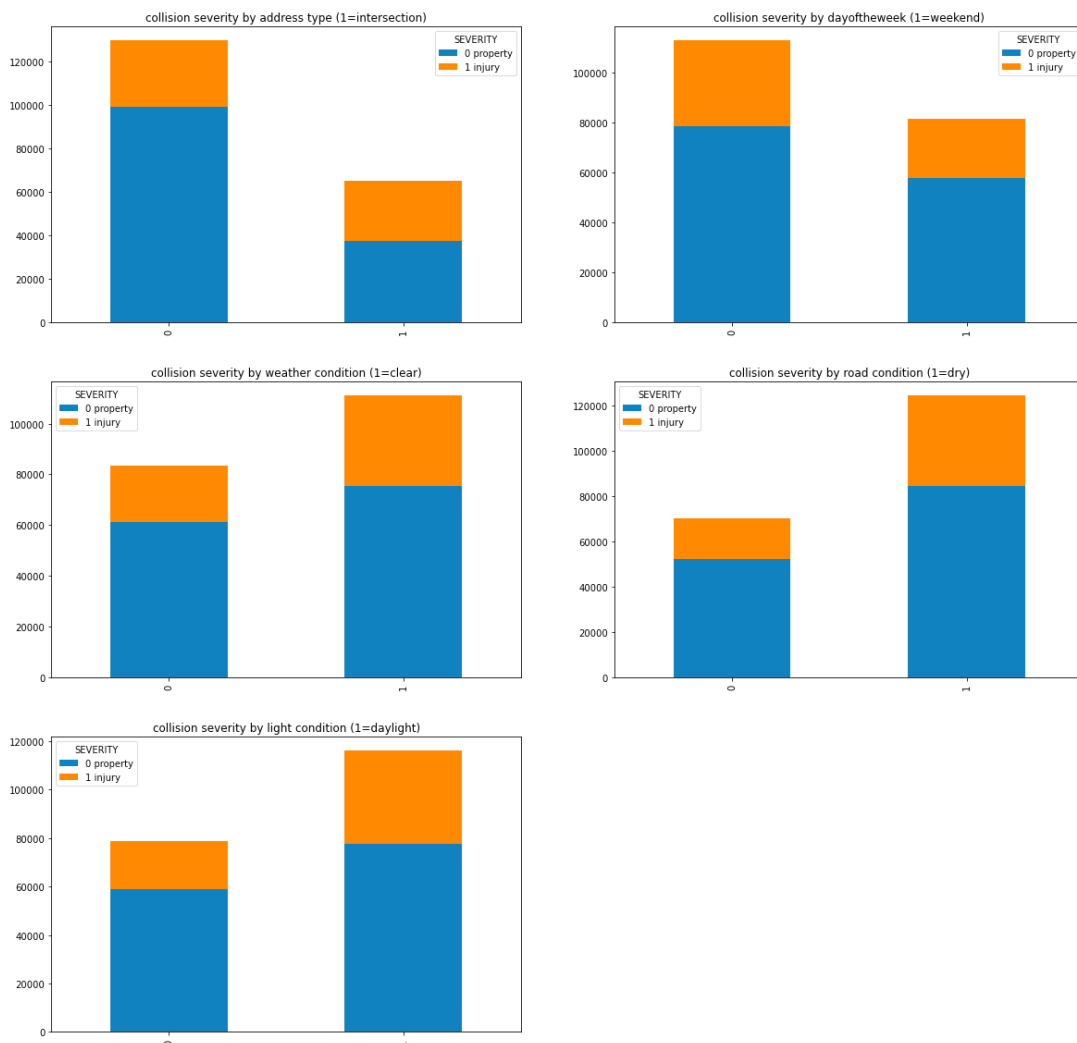
As can be read from the graphs, accidents are more likely to occur at non-intersections, during the weekdays, on sunny days, on dry roads and under the daylight. It is likely that this result is due to the high volume of traffic and the number of people out in such conditions, but injury ratios vary from situation to situation, and it is unclear what combinations of these features cause them. In this analysis, we apply machine learning models such as decision trees to examine how these features explain the outcome variables.