Analysis of Car Accident Severity

Veronica Zhao

Introduction/Business Problem

- One of the leading causes of death in the United States is accidents, including those caused by motor vehicles
- Knowing which factors and how much they affect car accidents will provide important information that can be used to caution drivers during certain road conditions through road signs or warnings
- Project Focus: Seeing which environmental and road factors affect car accidents the most and building a model that best helps predict the severity of a car accident under certain conditions.

Data

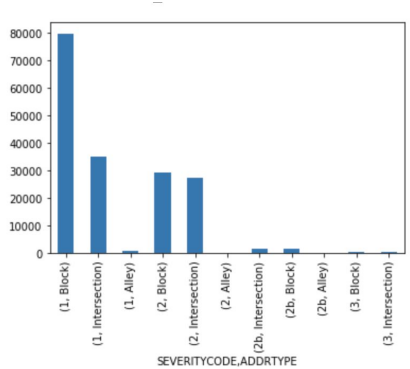
- Seattle city data provided by SPD and recorded by Traffic Records
- Includes all collisions that occurred at the intersection or mid-block of a segment from years 2004 to 2020
- Primarily looking at variables related to weather, road conditions, and location
    - ADDRTYPE
    - WEATHER
    - ROADCOND
    - LIGHTCOND
- Predicting severity of car accident
    - SEVERITYCODE

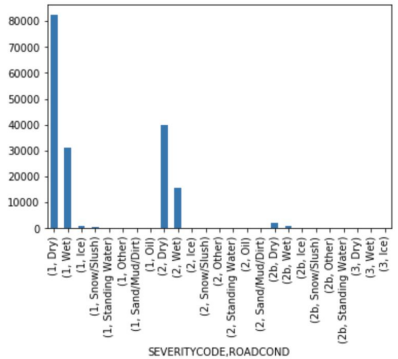
Data Collection and Preparation

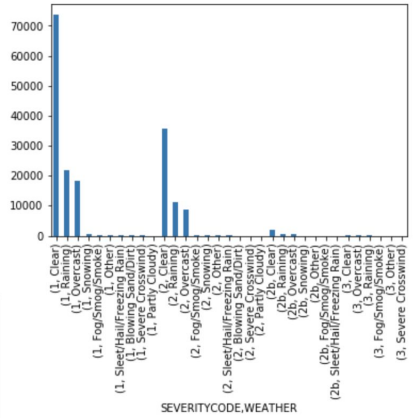
- Imported CSV file of car collisions from the Seattle city data
- Dropped all variables except:
  - ADDRTYPE
  - WEATHER
  - ROADCOND
  - LIGHTCOND
  - SEVERITYCODE
- Dropped all NaN values
- Dropped all unknown entries

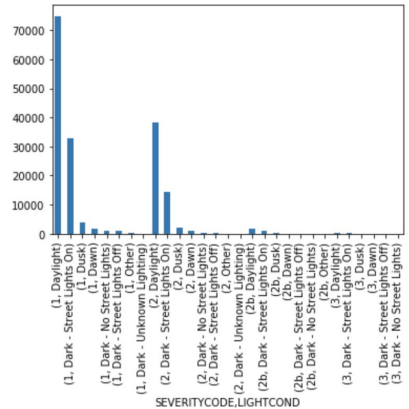
Exploratory Data Analysis

Grouped each predictor variable by SEVERITYCODE and plotted its frequency









Methodology

1. Loaded the data and kept only the variables we need for the analysis which are the variable we are trying to predict
    a. Cleaned the data by removing unknown entries
    b. Plotted each predictor variable value and grouped them by SEVERITYCODE
2. Prepare our data for model building
    a. Label encode categorical variables to produce numerical labels
    b. Split data into training and test sets for our model
3. Fit data into different models and evaluate them to see which produces the highest accuracy
    a. Models: SVM, K-Nearest Neighbors, Logistic Regression, Decision Tree

Modeling and Evaluation

- SVM
  - Accuracy
    - Train set: 0.6597271663434234
    - Test set: 0.6572912489736686
  - F1-score: 0.5213709856308176
  - Jaccard Similarity Score: 0.6572912489736686
- K-Nearest Neighbors (k = 7)
  - Accuracy
    - Train set: 0.6358235337441387
    - Test set: 0.6317236638597697
  - F1-score: 0.5395379811233175
  - Jaccard Similarity Score: 0.6317236638597697
- Logistic Regression
  - Accuracy
    - Train set: 0.6597271663434234
    - Test set: 0.6572912489736686
  - F1-score: 0.5213709856308176
  - Jaccard Similarity Score: 0.6572912489736686
  - Log Loss: 0.7087700067484126
- Decision Tree
  - Accuracy
    - Train set: 0.6597598998355142
    - Test set: 0.6572912489736686
  - F1-score: 0.5213709856308176
  - Jaccard Similarity Score: 0.6572912489736686

Results and Discussion

- For all accident groups, it appears that the accident occurred the most during daylight, when the weather was clear, and road condition was dry
- Second most likely conditions that cause accidents are rainy weather, wet road conditions, and being in the dark with street lights
- Most of the accidents that resulted in prop damage occurred at a block
- Almost all accidents that resulted in injury occurred at a block or intersection
- Accidents that resulted in fatality only occurred at blocks and intersections
- All models' accuracy were around 63-65%
- Decision Tree train set accuracy is slightly higher so we can say that this model is more accurate

Recommendation

- Pay more attention on the road when faced with conditions that do not seem to bring much risk
- More caution and speed limits should be enforced during rainy and wet conditions
- Installing signs and lights that light up in the dark to warn drivers at night
- Use decision tree as model for future predictions

Conclusion

- Focused on
  - finding the major environmental factors and road conditions that affect car accidents
  - building a model that can help predict the severity of car accidents based on these conditions
- Made recommendations to improve the safety of drivers and those on the road during certain road and weather conditions
- Suggested a model that produces the best results for further analysis of car accident severity based on the same predictors