Car Collision Severity Analysis

1. Introduction

1.1 Purpose of the Research

The purpose of the project is to understand the possible serverity of a car accident given the address type, location, road and weather condition, etc. This allow drivers to thoroughly evaluate condition and more careful in situations with high serverity level, reducing collisions, and give government an understanding on what type of road conditions need to be fix or add sign to reduce accident.

1.2 Interest

The main audience of the project is government, taking the result to further estimate solutions for preventing car accidents. But also for general car drivers to understand how street condition could affect the severity.

2. Data acquisition and cleaning

2.1 Data Source

The collision data is provided by SPD and recorded by Traffic Records.

The data will be used are:

a. Collision address type: including Alley, Block, and Intersection

b. SEVERITYCODE: 3—fatality

2b—serious injury

2—injury

1—prop damage

0—unknown

c. COLLISIONTYPE: ex. Angles, Sideswipe, Parked Car

d. WEATHER: weather conditions. ex. Raining, Snowing, cloudy

e. ROADCOND: condition of the road during the collision. ex. wet, dry

f. LIGHTCOND: The light conditions during the collision. ex. Dark - Street Lights On, Daylight

2.2 Data Cleaning

There are several problems with the datasets. First, there were a lot of missing values from the data downloaded. Some columns even missing more than 50% of the data. Since data missing would affect the result of the research, I decide to remove rows with missing data.

Second, there were redundancy in the columns. For example, there are two columns of severity. Therefore, one of the columns is also deleted.

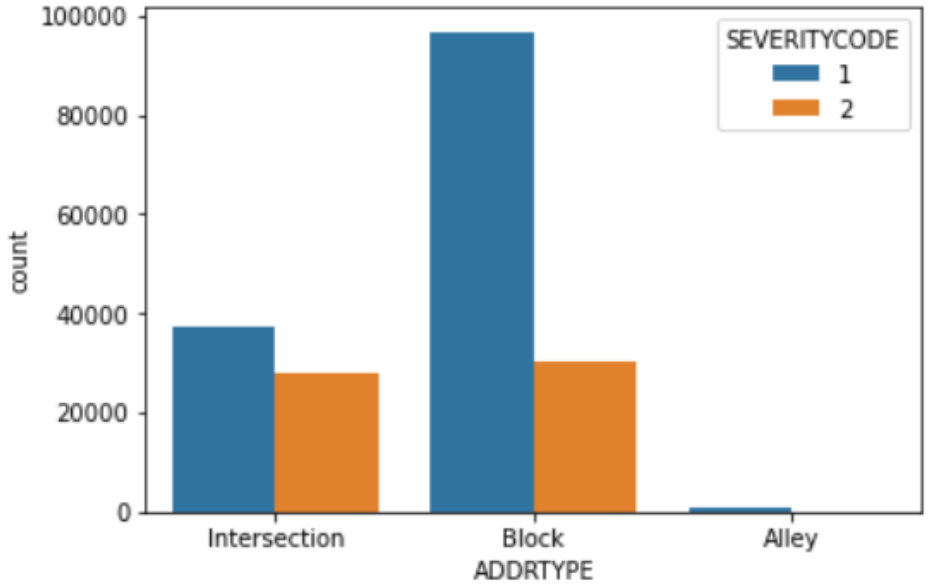
2.3 Feature Selection

As the main purpose is the evaluate the how road condition would affect the severity of car collision,

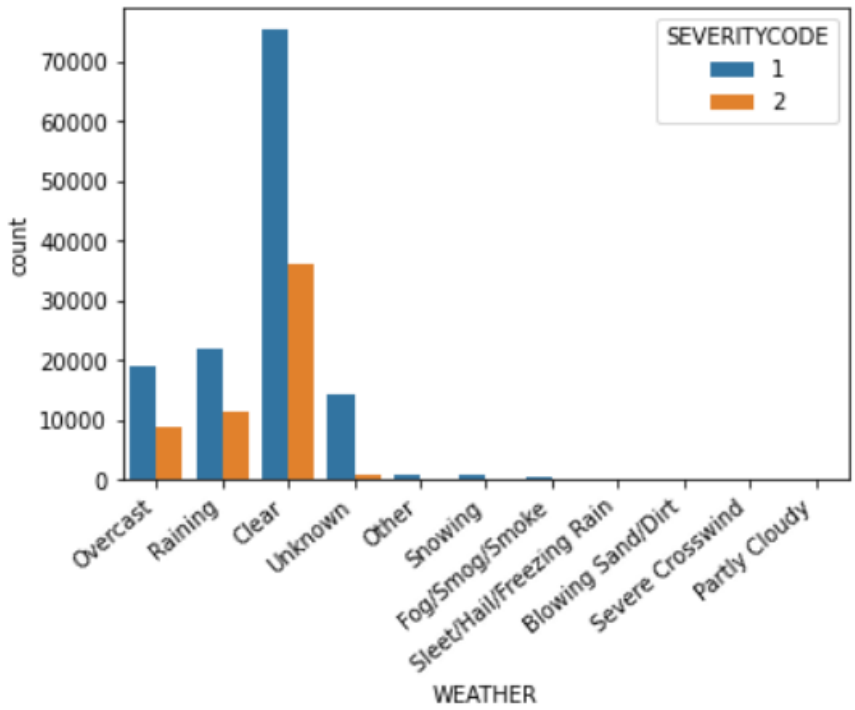
upon examining the meaning of each feature, I select the feature that might be most related, which are ADDRTYPE','COLLISIONTYPE','WEATHER','ROADCOND','LIGHTCOND'.

3. Exploratory Data Analysis

3.1 Relationship between ADRTYPE and SERVERITYCODE
Most of the collisions were happening in the block address. As the injury case amount is relative same in intersection and block, block has more prop damage cases.

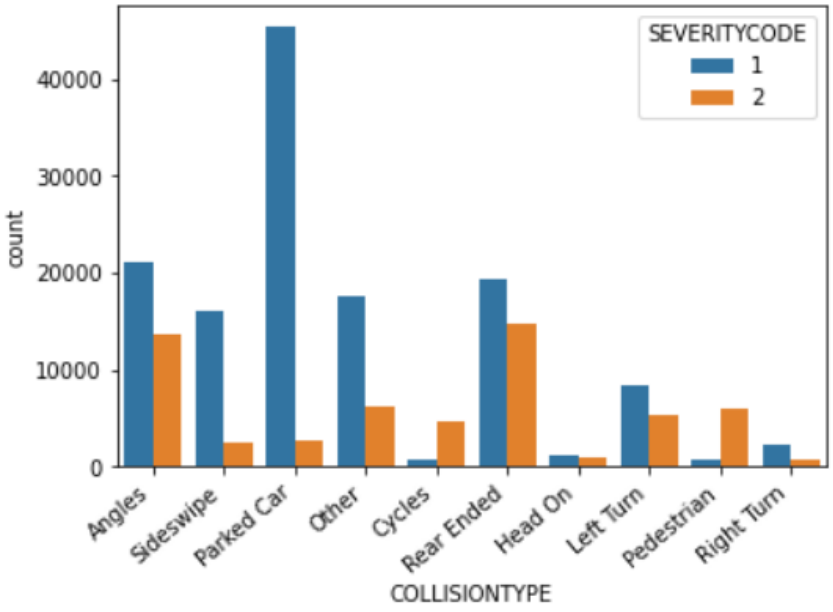


3.2 Relationship between WEATHER and SERVERITYCODE
While it is widely thought that collisions happened when it was raining, the data set shows that there are more case when the weather is clear.



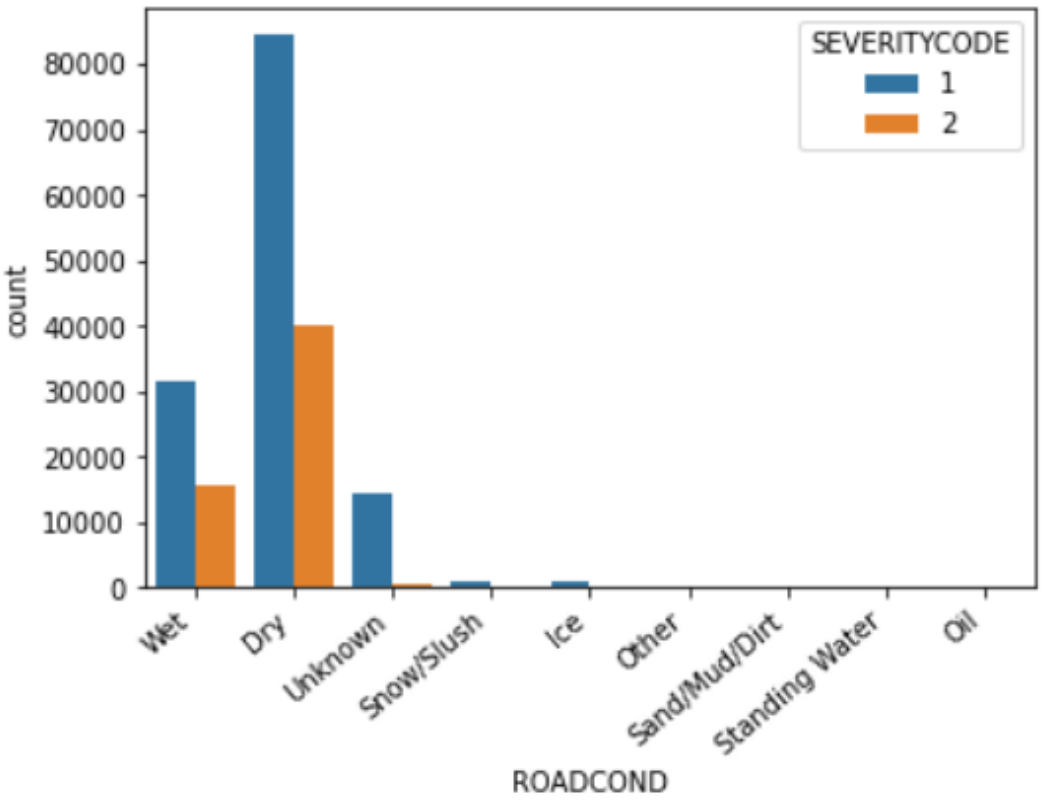
3.3 Relationship between COLLISIONTYPE and SERVERITYCODE

Most of the collisions were on parked car, causing prop damage. However, rear ended collisions were having more injury case. Besides, collisions on cycles and pedestrians were most likely caused injury.



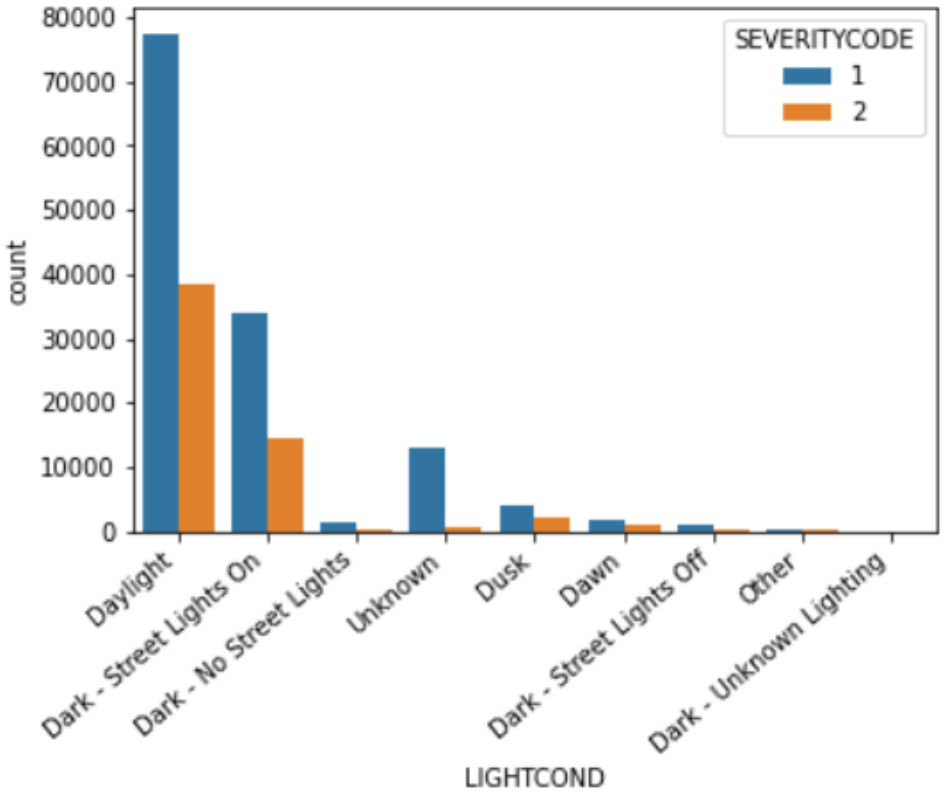
3.4 Relationship between ROADCOND and SERVERITYCODE

While it is widely thought that collisions happened on wet or ice road condition, the data show that most collision cases were happened dry condition.



3.5 Relationship between LIGHTCOND and SERVERITYCODE

There were more collisions during daylight. This could be that less car on the street in the evening and at night.



4. Predictive Modeling

In this research, I apply two model to predict the severity given ach conditions.

4.1 Applying Decision Tree Model and K Nearest Neighbor(KNN) Model
The accuracy for two model applied is approximately the same, whereas decision tree model results in slightly higher accuracy. The model could give government a more comprehensive view on what type of conditions would causing more serious injury, and could further planning on improving road condition, creating sign, or designing propaganda.

4.2 Problems for further Analysis

The model creates in this research is to predict the collision severity given road conditions. However, it is also worth to investigate the relationship between different conditions and having deeper analysis on collisions.