Data Science in Car Accident Severity

1. Introduction/Business Problem

Introduction where you discuss the business problem and who would be interested in this project.

Car accidents are a huge problem in our world. This project aims in analyzing "car accident severity" in terms of human fatality, traffic delay, property damage, or any other chance of fatality. In order to reduce car collisions in a community, a data science model must be trained to predict the severity of an accident. Taking consideration of the current weather, road and visibility conditions. When conditions are bad, this model will alert drivers to remind them to be more careful.

2. Data section

Data where you describe the data that will be used to solve the problem and the source of the data.

The dataset for this project is taken from a shared data link which is collected from Seattle SPOT Traffic Management Division. This is the shared data for Seattle city. The dataset is in the form of .CSV file. This includes all types of collisions. Collisions will display at the intersection or mid-block of a segment. The target label for the dataset is severity, which describes the fatality of an accident. The shared data has unbalanced labels. This dataset is updated weekly and is from 2004 to present. It contains information such as severity code, address type, location, collision type, weather, road condition, speeding, etc.,. There are 37 attributes in this dataset.

This Project for everyone who really care about the traffic records, especially in the transportation department.
This model is to improve the predictability of the accident severity and to reduce accidents in the future.

The result helps SHSP, DMV, stakeholders, insurance company, car manufacturers, and partnerships to allocate budget for education and enforcement to act on the result in order to achieve the goal of minimizing fatal/injury car crash.

The link for the dataset: https://s3.us.cloud-object-storage.appdomain.cloud/cf-courses-data/CognitiveClass/DP0701EN/version-2/Data-Collisions.csv

The link for the Metadata of the dataset: https://s3.us.cloud-object-storage.appdomain.cloud/cf-courses-data/CognitiveClass/DP0701EN/version-2/Metadata.pdf

Some of the Metadata given with the dataset:

The target Data to be predicted under (**SEVERITYCODE** 1-prop damage 2-injury) label.
Other important variables include:

- **ADDRTYPE**: Collision address type: Alley, Block, Intersection
- **LOCATION**: Description of the general location of the collision
- **PERSONCOUNT**: The total number of people involved in the collision helps to identify severity involved
- **PEDCOUNT**: The number of pedestrians involved in the collision helps to identify severity involved
- **PEDCYLCOUNT**: The number of bicycles involved in the collision helps to identify severity involved
- **VEHCOUNT**: The number of vehicles involved in the collision helps to identify severity involved
- **JUNCTIONTYPE**: Category of junction at which collision took place helps to identify where most collisions occur
- **WEATHER**: A description of the weather conditions during the time of the collision
- **ROADCOND**: The condition of the road during the collision
- **LIGHTCOND**: The light conditions during the collision
- **SPEEDING**: Whether speeding was a factor in the collision (Y/N)
- **HITPARKEDCAR**: Whether the collision involved hitting a parked car