IBM Data Science Professional Certificate Capstone Project – Car Accident Severity

Moshe E.

August 2020

Introduction

- The Capstone Project of the IBM Data Science Specialization covers all the topics taught in the courses
- The project is based on a provided dataset of car accidents occurred in Seattle since 2004
- The data was recorded by Traffic Records and collected by the Seattle Police Department
- The dataset can be found here (link to metadata)
- The dataset includes attributes such as:
  - Severity
  - Location
  - Collision type
  - Number of injuries
  - Weather, road and light conditions, etc.

The Business Problem

- Not all accidents can be predicted
- Many uncontrollable factors exist in every accident:
  - Weather
  - Location
  - Time, etc.

The Business Problem

- However, manageable recorded factors may include:
  - Lighting – if many accidents occur in dark areas ➲ proper lighting should be installed
  - Pedestrians – if many pedestrians involved in some areas ➲ noticeable crosses are needed
  - Cyclists – if many cyclists involved in some areas ➲ bicycle lanes should be paved
  - Parked cars – if many accidents involve parked cars ➲ proper parking is needed
- The impact of one such factor can be huge and save lives
- The local authority may benefit a lot from the analysis
- A safer living space for the citizens can be provided

The Data

- The dataset gathers all collision events in Seattle since 2004
- The attributes in the dataset include, among others:
  - Severity of collision (damage level)
  - Collision type (head on, involved pedestrians or cyclists)
  - Time of accident date and time
  - Affected persons (also if cyclists, pedestrians or passengers were involved)
  - Involved parked cars
  - Address (alleys, blocks or intersections)
  - Weather, road and light conditions
- Overall there are 194673 accidents recorded
- As mentioned before, this report will focus on the manageable factors