IBM Applied Data Science Capstone
Predicting Road Accidents in Seattle

Author: Ilan Gil

Date: 20/09/2020

---

1. Problem Description

Road accidents are one of the leading causes of death globally. In the USA alone, more than 38'000 people died in motor vehicle collisions last year. This particular project will focus on the American city of Seattle. Located on the West Coast of the USA, it is the largest city in the state of Washington with a population above 700'000. Naturally, this high concentration of people implies a high number of motor vehicles and the corresponding accidents usually associated with them.

These accidents often involve dramatic consequences. On top of direct health damages, people involved in such accidents may expect large medical bills, significant property damage, or missing revenue from their inability to work. Immediately after occurring, these accidents also require a very swift response from the authorities. Preparation and coordination between the paramedic team, the police department, and the nearest hospitals is key in order to ensure an appropriate response.

This project will therefore focus on the ''response'' side of these accidents. In other words, the upcoming analysis will strive to shine light on some factors that may indicate a higher likelihood that accidents will take place, as well as at the potential ways that authorities may prepare themselves in order to respond to these accidents.

Being able to anticipate and better prepare oneself to incoming accidents could prove very useful for a number of federal bodies. In detail:

- Hospitals could use those predictions in order to coordinate their staff strength with respect to the amount of accidents prone to occurring on a given day. For instance, should the data analysis indicate that accidents are more likely to occur on say Fridays, hospitals could respond by increasing their headcount on that specific day.
- Police departments could be interested in those predictions insofar as they would allow them to adapt their level of alertness if the probability of severe accidents turns out to be higher than usual on certain days.
- Governing authorities may also use those predictions, notably in order to warn people of a potentially dangerous day when it comes to road accidents.

2. Data

This analysis will use an extensive dataset from the Seattle Police Department, with over 190'000 observations since 2004.

The dataset includes a number of factors that could help us determine the likelihood of a road accident occurring. These factors pretty much describe the overall context in which an accident takes place. Some of the most important factors include:

- **SEVERITYCODE**: A code that corresponds to the severity of the collision
- **PERSONCOUNT**: The total number of people involved in the collision
- **VEHCOUNT**: The number of vehicles involved in the collision
- **WEATHER:** A description of the weather conditions during the time of the collision
- **LIGHTCOND:** The light conditions during the collision
- **ROADCOND:** The condition of the road during the collision
- **Etc.**