**1. PROJECT INTRODUCTION**

**THE PROBLEM**

Driving bears a risk of collisions, with accidents sadly resulting in 148 road fatalities in Ireland in 2019 (reference: www.rsa.ie). The technology for self-driving cars is fast advancing but, for the vast majority of drivers, it will be a while yet before self-driving cars are accessible. This project seeks to understand the risk factors which increase the severity of a road collision, in a bid to better understand how to make driving a safer activity.

**WHO WILL BE INTERESTED?**

The results of this project will be of interest to city councils who may be able to educate members of the public with road safety advice based on research. Through developing a better understanding of what increases the severity of car accidents, we can make our roads safer. By understanding what factors contribute to car accidents, city councils can inform the public on what to be aware of, to hopefully prevent a number of collisions or to reduce their severity.

The results will also be of interest to car manufacturers who can use this information to develop new features to make safer cars. This includes self-driving cars – for example, if environmental conditions increase the risk of a more severe collision taking place, this needs to be understood in the development of the self-driving technology.

**2. DATA**

**DESCRIPTION OF THE DATA**

This dataset shares information on all vehicle collisions as provided by the Seattle Police Department and recorded by Traffic Records. The dataset has 38 attributes and 194,673 rows. There are a mix of numerical and categorical attributes, and some have missing data.

The attribute ‘SEVERITYCODE’ indicates the assigned severity of the accidents, with the numbers representing the following accident severity levels:

* 3—fatality
* 2b—serious injury
* 2—injury
* 1—prop damage
* 0—unknown

The dataset includes a number of other attributes containing further detail and information about recorded vehicle collisions in Seattle City. This project seeks to determine if any of these other attributes, or combination of attributes, can operate as a predictor for Accident Severity.

Some potential hypotheses could suggest that environmental conditions on the day of the collision (for example weather, road conditions, light conditions and so on) or driver conditions (for example speeding on the part of the driver, or whether the driver was under the influence) might be predictors of the severity of an accident.

The data dictionary can be found here:

<https://s3.us.cloud-object-storage.appdomain.cloud/cf-courses-data/CognitiveClass/DP0701EN/version-2/Metadata.pdf>

**HOW THE DATA WILL SOLVE THE PROBLEM**

The dataset in question only includes accidents which have a Severity Code (attribute SEVERITYCODE) of 1 or 2, indicating collisions resulting in ‘Property damage’ or ‘Injury’ respectively. As such, the model will attempt to determine what factors increase the likelihood of a more serious accident taking place, where an injury occurs rather than just property damage.

One avenue that might be explored is whether road conditions impact the severity of a collision. The following are the potential values for the attribute ‘ROADCOND’, indicating ‘Road Condition’:

* Dry
* Ice
* Oil
* Other
* Sand/Mud/Dirt
* Snow/Slush
* Standing Water
* Unknown
* Wet
* (blank)

This project will comprise some initial exploratory data analysis to better understand the dataset. Once this data exploration is complete, a model will be developed and then evaluated against the problem of: can we predict accident severity level?