**Predicting the occurrence of severe car accidents in Seattle, Washington, USA**

# Introduction

## 1.1 Background

Road traffic accidents and subsequent injuries have a remarkable social impact for individuals and families and are one of the leading causes of death across the globe. Although Seattle in one of the safest cities in the United States of America (USA), according to Seattle government more than 10000 collisions a year still occur, resulting in an average of 20 people losing their lives and 150 people being seriously injured. During the first half of the year 2019, 101 people were seriously injured or killed in 98 collisions on Seattle streets. That is the highest number of crashes in the first six months of a year since 2010, according to preliminary police reports analysed by the Seattle Department of Transportation (SDOT). The city of Seattle is aiming at ending traffic deaths and serious injuries on city streets by 2030, however, some reports argue that if better measures don’t take place Seattle will be far from meeting its goal.

Seattle has a warm temperate oceanic climate characterized by cool, wet winters and mild, relatively dry summers. Summertime is mildly warm and moderately dry with temperatures ranging between 20-23 Celsius degrees, with daylight hours up to sixteen hours. In contrast, winters are cold and wet, with low temperatures between 2-10 Celsius degrees, and due to the rainfall most days roads are wet and slippery. Rainfall is experienced for an average of eighteen days per month, and the rain gauge collects approximately 147.3 mm. Snow falls are high and regular, accumulating up to 43.2 mm, mostly in December and February. The average sunshine hours per day are two. Adverse conditions and road status could play a pivotal road in the occurrence of fatal car accidents.

**1.2 Problem and Interest**

In order to establish new measures to reduce the number of mortalities and injuries caused by car collisions, this study aims to find meteorological features as well as road conditions that can foster car collisions, with the ultimate goal of preventing or diminishing car crashes, making Seattle’s roads safer for all inhabitants, extendable to other cities across USA and the rest of the world.

**2. Data**

**2.1 Data acquisition and cleaning**

The data used for this study comprises a csv file and metadata from all collisions in Seattle provided by the Seattle Police Department (SPD) and recorded by Traffic Records from 2004 to present. I will use the data to structure a classification model and determine whether certain weather conditions (snow, fog, rain, crosswind), as well as road status (ice, snow, wet) can promote car collisions leading to injured people.

**2.2 Data cleaning and processing**

Data contained 194673 rows and 37 columns. There were no missing values for the dependent variable “Severity of accident” but 5081 and 5012 missing values were found in “weather” and “road conditions” independent variables. Since the number of total observations is high, and the independent variables are not continuous but categorical with several subsets, I decided to remove the rows comprising missing values in those two independent variables. I cleaned the data in the table and make a new table with my independent variable “severity of accident” and my two dependent variables “weather” and “road condition”. After cleaning the data, I had 189543 rows and 3 columns.

This is how my cleaned table looks like:

A screenshot of a cell phone

Description automatically generated

I will continue processing the data for classification models. Details will be updated in the next report.