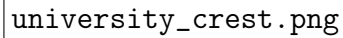


# **An investigation into the car accident likelihood and severity in Seattle**

A square box containing the text "university\_crest.png".

university\_crest.png

**Rosany Antonyvincent**

Data Science Professional Specialisation  
IBM

August 2020

## **Acknowledgements**

## **Abstract**

# Table of contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Background . . . . .	1
1.2	Problem . . . . .	1
<b>2</b>	<b>Data acquisition and cleaning</b>	<b>2</b>
2.1	Data sources . . . . .	2
<b>3</b>	<b>Exploratory data analysis</b>	<b>3</b>
<b>4</b>	<b>Predictive modelling</b>	<b>4</b>
<b>5</b>	<b>Conclusion</b>	<b>5</b>
	<b>References</b>	<b>6</b>
	<b>Appendix A</b>	<b>7</b>
	<b>Appendix B</b>	<b>8</b>

# Chapter 1

## Introduction

### 1.1 Background

Seattle is the largest city in the state of Washington and consists of roughly 710000 residents[1]. The best way to travel within the city is by car, and therefore evaluating the likelihood and severity of road accidents is of great importance. Not only to the safety and fatality of the civilians, but also for the economy. Reducing road accidents, reduces potential time off work for busy commuters which allows for less demand for paid sick leave and for employees to continue working[2]. The most common job type in the area, as of 2018, is software developer which receives an average income of \$91000[3]. This overtook retail salesperson in 2016, which had an average salary of \$28000[4]. This jump in average wage, shows the necessity for preserving and furthering the economy.

### 1.2 Problem

Although across the years the number of car accidents have been decreasing across Seattle, and across Washington, the number of vehicles involved in fatal accidents are slowly increasing as are the fatalities[5]. Equally the number of pedestrians involved in fatal accidents are increasing. The most likely time of accident takes place on Saturday afternoons.

The aim of this project is to predict the likelihood and the severity of a road accident in order to establish routes and times that could be high risk and therefore avoided in order to reduce fatalities.

# Chapter 2

## Data acquisition and cleaning

### 2.1 Data sources

The labelled data used will be the Data Collision records in Seattle across all years provided by SeattleGeoData[6]. Severity will be the predominant dependent variable, with varying attributes. Attributes include location, collision type, incident time and date, weather and junction type. To this end, all other attributes will be discarded. The factors this report will look for is whether the weather has a significant impact on accidents and to what extent. Further to this, it will explore the location of incidents. It will also explore if there is any significance to the time and date of accidents and fatality. This will be done using histograms and clustering techniques.

If a correlation is established, the focus will then be placed on fatalities. Fatalities will then be converted to binary form, which allows for test/train predictions. The accuracy of the test/train prediction will be qualified using kNN, f1-score and Jaccard index.

## **Chapter 3**

### **Exploratory data analysis**

# **Chapter 4**

## **Predictive modelling**



## **Chapter 5**

## **Conclusion**

# References

- [1] 2019 washington state car accident statistics reports. Feb 2018.
- [2] Most common job in seattle isn't in retail anymore, and 4 out of 5 of these workers are men. June 2018.
- [3] Average software developer salary in seattle, washington.
- [4] Average retail sales associate hourly pay in seattle, washington.
- [5] Fatal car crashes and road traffic accidents in seattle, washington.
- [6] Collisions. May 2018.

# Appendix A

## **Appendix B**