If you commute to work in Seattle, accidents pose a potential risk for your health and cause annoying delays in traffic. What, if you had the information to avoid them?

Martin Koschmieder

Getting there safely and in time!

Capstone Project for course “IBM Data Science Professional”

Contents

[Introduction 3](#_Toc50126078)

[Data 3](#_Toc50126079)

[Methodology 3](#_Toc50126080)

[Results 3](#_Toc50126081)

[Discussion 3](#_Toc50126082)

[Conclusion 3](#_Toc50126083)

[Acknowledgments 3](#_Toc50126084)

[Appendix 3](#_Toc50126085)

# Introduction

Most people do not live near their workplace and many need to commute to work by car every working day. As a commuter, you just want to get there safely and in time, so avoiding accidents on your way to work is a top priority. Your GPS may show you the best route based on current traffic, but accidents can suddenly change the traffic situation and before you know it, you are stuck in traffic and sweating to get to your destination in time.

What, if you knew, where and when most accidents happen?

What, if you knew how weather or road conditions affect the severity of accidents?

With this information you might be able to avoid some of the accident hotspots on your route or drive extra carefully in certain situations (e.g. near accident hotspots or in certain weather or road conditions).

Based on traffic from the City of Seattle as an example, this project will show how the following helpful information for commuters can be provided:

1. Top 50 collision hotspots
2. Distribution of collisions throughout the week (day of the week and time of day)
3. Typical situations (characteristics of location, weather and road conditions etc.), in which severe collisions occur

# Data

The dataset used in this project comprises detailed information on almost 200,000 accidents that were reported in Seattle between 1 January, 2004, and 20 May, 2020. For each of these accidents, the dataset provides 38 characteristics describing the location (coordinates, description etc.), the date and time, the type of collision, the weather and road conditions and administrational information (e.g. the police report number etc.)

The data and descriptive metadata is provided by the Seattle Department of Transportation (SDOT) and can be downloaded from the “Transport” section of the “City of Seattle Open Data portal”. The most recent report is dated September 11, 2020 (https://data.seattle.gov/Land-Base/Collisions/9kas-rb8d).

For this project a copy of the data dated provided by Coursera was used:

dataset: <https://s3.us.cloud-object-storage.appdomain.cloud/cf-courses-data/CognitiveClass/DP0701EN/version-2/Data-Collisions.csv>

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

# Methodology

section which represents the main component of the report where you discuss and describe any exploratory data analysis that you did, any inferential statistical testing that you performed, if any, and what machine learnings were used and why.

### Data Preparation and Cleaning

* Deal with missing data
* Straighten out data types
* Convert text data to numbers
* Select features
* Select lables
* Balance lables
* Split data into training and testing sets

### Modelling

….

### Evaluation

….

### (Deployment)

…

# Results

section where you discuss the results.

# Discussion

section where you discuss any observations you noted and any recommendations you can make based on the results.

# Conclusion

section where you conclude the report.

# Acknowledgments

# Appendix

Metadata on Dataset