

Predicting Car Accident Severity in Seattle, USA

IBM Data Science Capstone

13th September 2020

Ryan W

Table of Contents

1	Business Problem	1
1.1	Introduction	1
1.2	Intended Audience.....	2
2	Data	2
2.1	Data Sources	2

1 Business Problem

1.1 Introduction

In the United States, road traffic crashes are a leading cause of death for people aged between 1 - 54 years. Globally, every day, almost 3,700 people are killed in road traffic crashes and are estimated to be the eighth leading cause of death.¹

The impact of road traffic crashes is felt economically as well for countries, costing countries 3% of their gross domestic product.² A study conducted in the United States showed that in 2010 the total economic cost of motor vehicle crashes was \$242 billion. These costs were derived from the 32,999 fatalities, 3.9 million non-fatal injuries, and 24 million damaged vehicles. However, this number increases significantly when quality-of-life valuations are considered for those involved in road traffic crashes, increasing the total value of societal harm from road traffic crashes to \$836 billion.³

With such a large impact on both the lives of those living in the United States, as well as the country's economy, there is a keen interest to develop machine learning models to better predict causes of road traffic crashes to develop preventative measures.

¹ <https://www.cdc.gov/injury/features/global-road-safety/index.html>

² <https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries>

³ <https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812013>

1.2 Intended Audience

The following report will provide insight into various key indicators for road traffic crashes in the Seattle district of the United States. This information will prove a range of insights that can have direct applications for a range of industries such as the Seattle council, insurance and health care.

For the Seattle council, insights will provide information related to areas prone to areas to accidents and their cause effect, which can lead towards the development of interventions in speed management, infrastructure design and enforcement of traffic laws.

For insurance companies, insights will provide information to district areas that have high accident counts for both fatalities and minor crashes such as sideswiping of parked cars. This can lead towards requiring owners pay an increased premium on their car insurance to reduce the risk for the insurance company.

For the health care department, insights will provide information on areas that have high accident counts that require medical assistance and the type of treatment required. This can lead towards improved resource allocation in areas of high road traffic crashes, development of post-crash survival strategies and training for relevant staff.

2 Data

2.1 Data Sources

Data that outlines the collisions in the Seattle area from the year 2004 to current can be found [here](#) and the metadata can be found [here](#). The data is collected by the SDOT Traffic Management Division and is updated weekly. It includes data of all types of collisions across all types of transportation methods. The dataset, at the time of writing this report, contains 194,673 reported accidents. There are 37 different attributes that describe the accident:

Location Based Attributes (7):

Attribute	Data type, length	Description
SHAPE	Geometry	ESRI geometry field
ADDRTYPE	Text, 12	Collision address type: <ul style="list-style-type: none">• Alley• Block• Intersection
INTKEY	Double	Key that corresponds to the intersection associated with a collision
LOCATION	Text, 255	Description of the general location of the collision
JUNCTIONTYPE	Text, 300	Category of junction at which collision took place
SEGLANEKEY	Long	A key for the lane segment in which the collision occurred.

CROSSWALKKEY	Long	A key for the crosswalk at which the collision occurred.
--------------	------	--

Unique Accident Identifier Attributes (5):

Attribute	Data type, length	Description
OBJECTID	ObjectID	ESRI unique identifier
INCKEY	Long	A unique key for the incident
COLDETKEY	Long	Secondary key for the incident
EXCEPTRSNCODE	Text, 10	If the accident has sufficient data associated
EXCEPTRSNDESC	Text, 300	If the accident has sufficient data associated

Accident Details Attributes (22):

Attribute	Data type, length	Description
SEVERITYCODE	Text, 100	A code that corresponds to the severity of the collision: 3–fatality 2b–serious injury 2–injury 1–prop damage 0–unknown
SEVERITYDESC	Text	A detailed description of the severity of the collision
COLLISIONTYPE	Text, 300	Collision type
PERSONCOUNT	Double	The total number of people involved in the collision
PEDCOUNT	Double	The number of pedestrians involved in the collision. This is entered by the state.
PEDCYLCOUNT	Double	The number of bicycles involved in the collision. This is entered by the state.
VEHCOUNT	Double	The number of vehicles involved in the collision. This is entered by the state.
INJURIES	Double	The number of total injuries in the collision. This is entered by the state.
SERIOUSINJURIES	Double	The number of serious injuries in the collision. This is entered by the state.
FATALITIES	Double	The number of fatalities in the collision. This is entered by the state.
INCDATE	Date	The date of the incident.
INCDTTM	Text, 30	The date and time of the incident.
SDOT_COLCODE	Text, 10	A code given to the collision by SDOT.
SDOT_COLDESC	Text, 300	A description of the collision corresponding to the collision code.

INATTENTIONIND	Text, 1	Whether or not collision was due to inattention. (Y/N)
UNDERINFL	Text, 10	Whether or not a driver involved was under the influence of drugs or alcohol.
PEDROWNOTGRNT	Text, 1	Whether or not the pedestrian right of way was not granted. (Y/N)
SDOTCOLNUM	Text, 10	A number given to the collision by SDOT.
SPEEDING	Text, 1	Whether or not speeding was a factor in the collision. (Y/N)
ST_COLCODE	Text, 10	A code provided by the state that describes the collision. For more information about these codes, please see the State Collision Code Dictionary.
ST_COLDESC	Text, 300	A description that corresponds to the state's coding designation.
HITPARKEDCAR	Text, 1	Whether or not the collision involved hitting a parked car. (Y/N)

Road Condition Attributes (3):

Attribute	Data type, length	Description
WEATHER	Text, 300	A description of the weather conditions during the time of the collision.
ROADCOND	Text, 300	The condition of the road during the collision.
LIGHTCOND	Text, 300	The light conditions during the collision.