

# At the corner of speeding and distracted

## An analysis of intersection collisions in Seattle caused by speeding and distracted drivers

Leslie McHugh  
Sept. 2020

### Introduction

#### 1.1 Background:

Although distracted driving is a more recent phenomenon, speeding has been a problem for as long as people have been driving.

Ranked as the 5<sup>th</sup> worst in the nation<sup>1</sup>, Seattle traffic is notoriously bad. The lack of a good mass transit system means more people are forced to drive everywhere which, combined with Seattle's rainy weather, means collisions are an all too common occurrence.

#### 1.2 Problem:

While SDOT (Seattle Department of Transportation) reports that overall accidents are trending downward, the city has a goal to eliminate all traffic fatalities.<sup>2</sup> To that end, the agency collects volumes of traffic accident related data. The focus of this analysis is on creating a very targeted initiative to identify the most intersections in the city where the most severe collisions occur where the driver is both speeding and distracted.

#### 1.3 Interest:

SDOT and city officials are interested in these findings, as they'll be able to more effectively target efforts and resources in reducing severe collisions. They will also be more equipped to set up possible check points and / or speed traps to watch for speeding and distracted drivers.

For residents and others traveling in around the city, they would have the knowledge to know which areas are more dangerous and prone to serious collisions and so could be more alert or possible avoid those areas if they so desired.

### Data acquisition and cleaning

#### 2.1 Data set:

The dataset used can be found here: <https://s3.us.cloud-object-storage.appdomain.cloud/cf-courses-data/CognitiveClass/DP0701EN/version-2/Data-Collisions.csv> . Metadata can be found here: <https://s3.us.cloud-object-storage.appdomain.cloud/cf-courses-data/CognitiveClass/DP0701EN/version-2/Metadata.pdf> The data was collected by SPD and SDOT from 2004 to present and contains detailed information about traffic collisions in Seattle including severity of the collisions, weather, light and road

1 <https://accidentdatacenter.com/us/washington/seattle-tacoma-wa>

2 <https://sdotblog.seattle.gov/2018/12/04/2018-traffic-report-still-growing-moving-and-getting-safer/>

conditions, location, latitude and longitude specifications, whether the driver was under the influence, distracted or speeding. The initial data set contained a total of 38 features.

## **2.2 Data cleaning:**

The availability of latitude and longitude and other specifics about each crash, made mapping the collisions seem the visualization that would be most useful. However, the data needed to be narrowed down to make map plotting readable. The first step was to eliminate the features that were irrelevant to the target. The dataset was trimmed down to 11 features.

## **2.3 Feature selection:**

Since the focus is on severe collisions, the severity code was the key feature for the study. Other features necessary were those that contain data related to possible causes of a collision. Weather condition, road condition, light condition and possible driver impairment.

## **Methodology**

### **3.1 Removing missing data from latitude and longitude**

Since the final visualization of this analysis will be a map, latitude and longitude data were crucial. An examination of that data specifically revealed that were rows in the dataset that were missing latitude and / or longitude. There were relatively few rows with missing data and the data can't be replaced by other means, those rows were eliminated.

### **3.2 Removing other missing data**

Examining other critical features related to collision cause – road condition, light conditions and weather conditions – revealed missing data as well. As with the latitude and longitude data, these did not lend themselves well to replacement methods and there were relatively few rows with missing data, so they were removed as well.

### **3.3 Narrowing down the selections to intersection related collisions only**

The Junction Type feature showed 4 unique labels. Two of which were intersection specific, so the dataset was further narrowed to include only those collisions.

### **3.4 Examining values in columns with possible other contributing factors**

- Adverse weather conditions such as rain and snow were removed.
- Adverse road conditions were eliminated so only those collisions on dry roads were included.
- Adverse light conditions such as dark – streetlights off were removed.
- All collisions where driver was under the influence were removed.

1 <https://accidentdatacenter.com/us/washington/seattle-tacoma-wa>

2 <https://sdotblog.seattle.gov/2018/12/04/2018-traffic-report-still-growing-moving-and-getting-safer/>

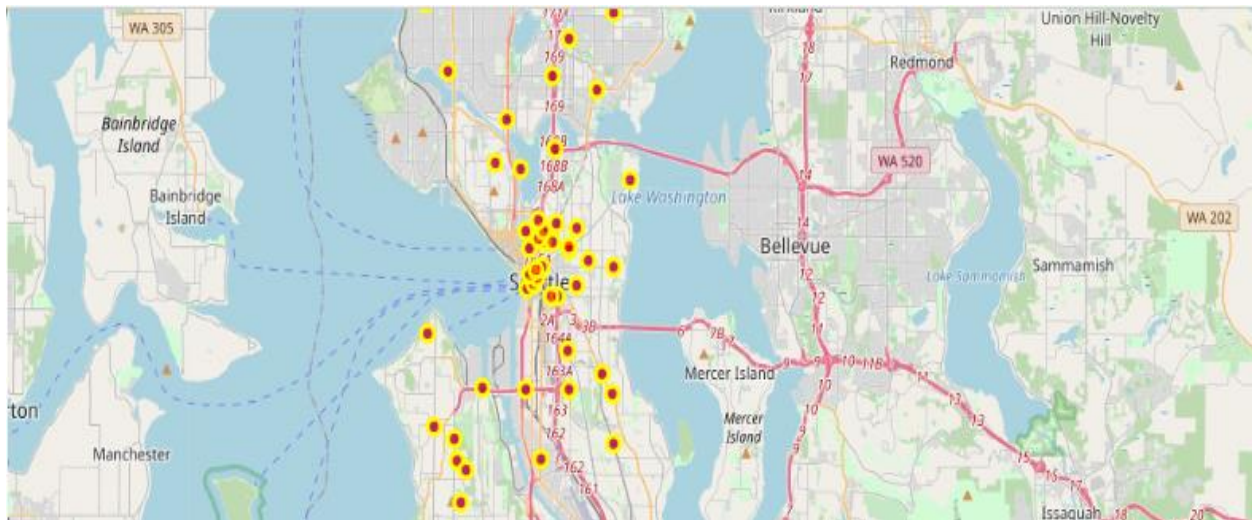
### 3.5 Speeding and distracted driver features

Most rows in the speeding and distracted features were blank. I chose to focus on only those that were specifically noted as Y for speeding and distracted.

## Mapping the resulting dataset

### 4.1 Plotting the individual collisions

The initial map of the data showed an obvious concentration area of these collisions. However, the individual points were not as actionable as clusters would be as to allow police to focus resources in general areas or neighborhoods.

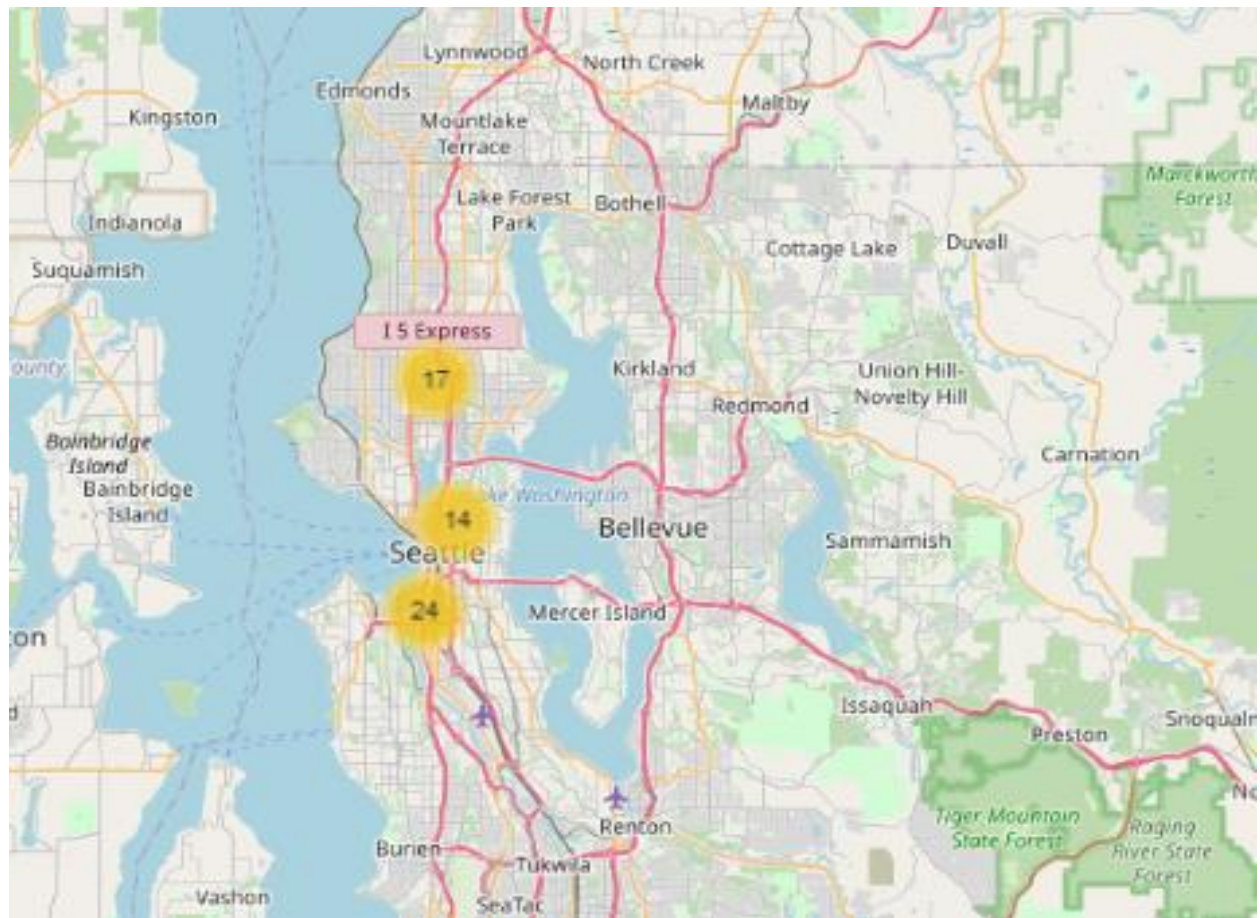


### 4.2 Clustering the plotted collisions

Clustering the data revealed more usable information, but still didn't provide detailed enough information as to which areas or neighborhoods needed the most attention.

1 <https://accidentdatacenter.com/us/washington/seattle-tacoma-wa>

2 <https://sdotblog.seattle.gov/2018/12/04/2018-traffic-report-still-growing-moving-and-getting-safer/>



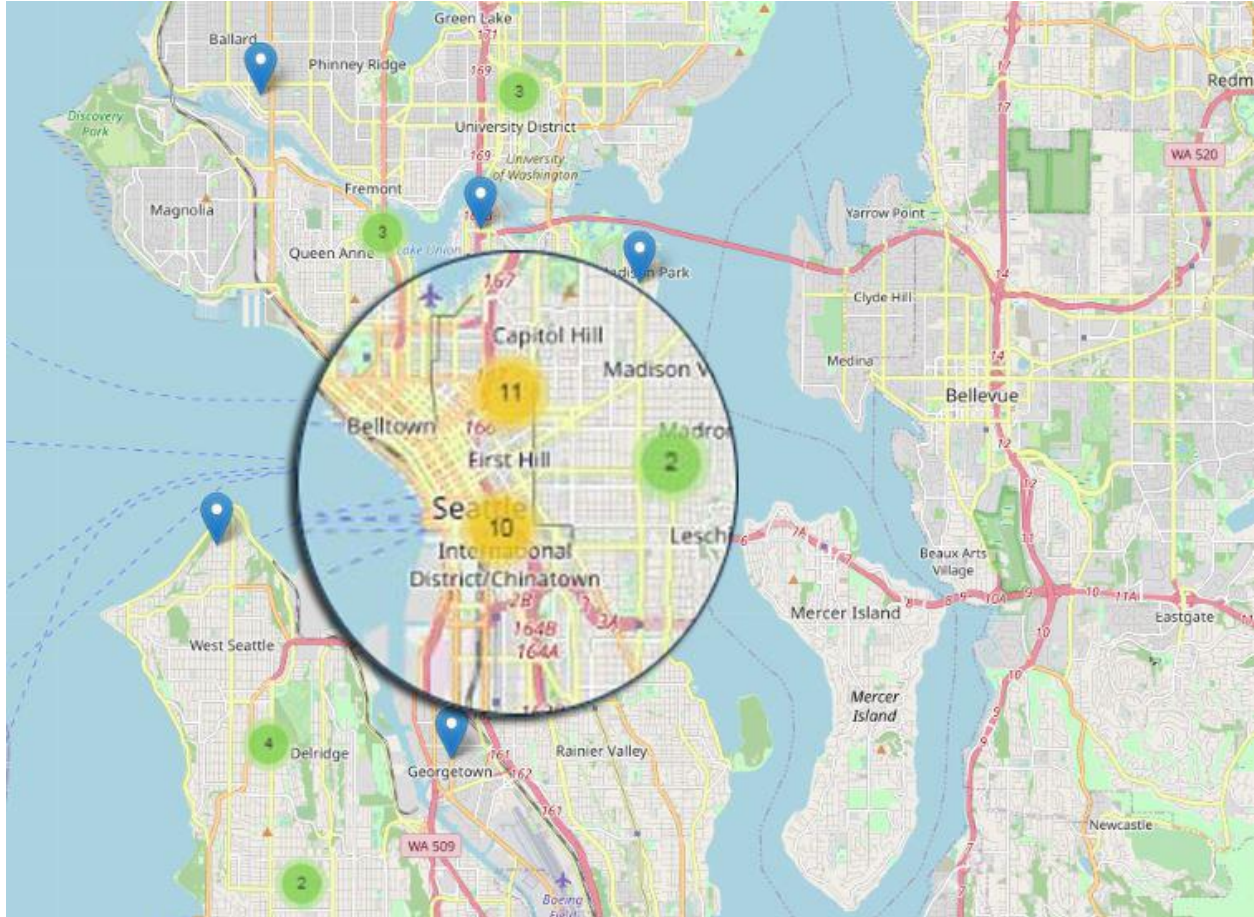
1 <https://accidentdatacenter.com/us/washington/seattle-tacoma-wa>

2 <https://sdblog.seattle.gov/2018/12/04/2018-traffic-report-still-growing-moving-and-getting-safer/>



### 4.3 Expanding the clusters to the neighborhood level

Further analysis of the collision clusters reveals that outside the downtown area, they are quite scattered. The exceptions are the First Hill and International District/Chinatown areas.



## Conclusion

### 5.1 Conclusion:

Approximately 34% of severe collisions involving a speeding and distracted driver with no other causes happen in the First Hill and International District/Chinatown areas of Seattle. A focus on reducing the number of these types of car accidents should focus most heavily on those 2 sections of the city.

Drivers traveling to or through those areas should be especially alert in these neighborhoods.

1 <https://accidentdatacenter.com/us/washington/seattle-tacoma-wa>

2 <https://sdotblog.seattle.gov/2018/12/04/2018-traffic-report-still-growing-moving-and-getting-safer/>

## Recommendations for further analyses

### 6.1 Suggestions for further studies

- A repeat of this study at 6 months and 1 year after official measures are in place to try and mitigate these types of crashes.
- A separate analysis studying only distracted drivers regardless of speeding
- A separate analysis studying only speeding drivers regardless of distraction
- A study comparing the relationship between various conditions (road, weather, light) and accident severity

<sup>1</sup> <https://accidentdatacenter.com/us/washington/seattle-tacoma-wa>

<sup>2</sup> <https://sdotblog.seattle.gov/2018/12/04/2018-traffic-report-still-growing-moving-and-getting-safer/>