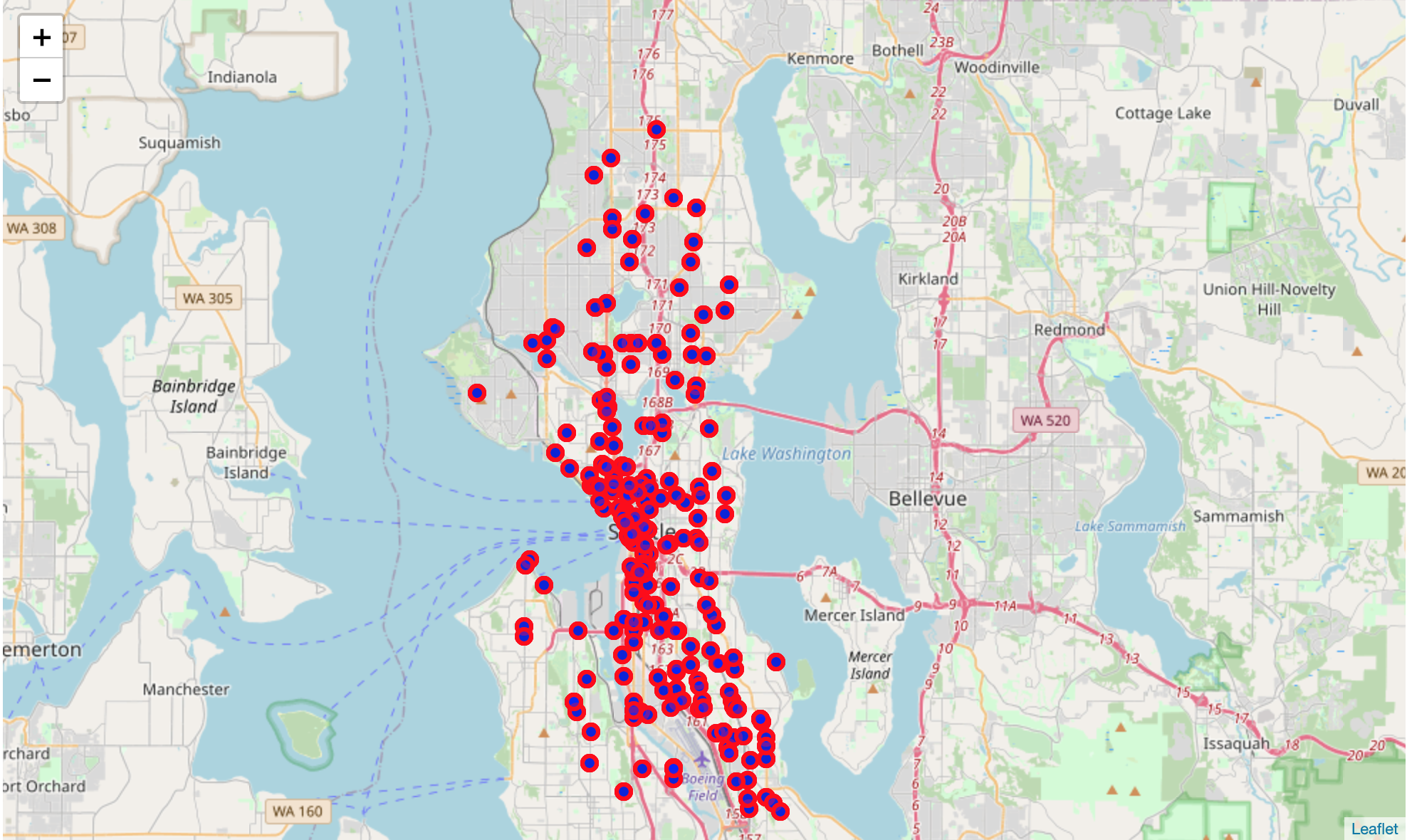
# Capstone Project - Car accident severity

## 

## **Introduction**

Road safety is one of the main concerns for every government in the world. Approximately 1.35 million people die in road crashes each year, on average 3,700 people lose their lives every day on the roads, while 20-50 million people suffer non-fatal injuries (resulting in long-term disabilities). For United States, the road crash is also a security risk for drivers and pedestrian. More than 38,000 people die every year in crashes on U.S. roadways. The traffic fatality rate is 12.4 deaths per 100,000 inhabitants and the terrible traffic security condition also cause 4.4 million people suffering from serious injure. The economic and societal damages of this single problem cost every citizen 871 billion dollars. To ease the harsh traffic security issue and offer a solution for the relevant department in the U.S. government, a prediction for the severity of car accidents and a detailed data analysis is crucial for the policy maker.

Seattle is also troubled with public traffic and road safety issue. Due to the assistance of the modern techniques, the Seattle Department of Transportation (SDOT) has been collecting and maintaining the public traffic data (including volumes, speeds and collisions) since 2004. There are over 229 heavy incidents in which more than 10 people are involved in last 16 years. Many of them are concentrated in the city area. It is crucial for the government to analyze the historical *Figure 1 The incidents (**over 10 persons involved) in last 16 years*



data. With the relevant data, I could use data science techniques to explore the incident severity in Python. The core aim of this report is to locate the key factor that lead to an incident in Seattle, as well as to offer some basic descriptions about the Seattle traffic incidents. To achieve this target, this report will present the answers to several issues. I shall start with the location of the incident on the map and the hot areas for the incident; then I will turn to the exploratory analysis for the collision type, location type and severity (the incident scale); in the end, I would build up a model to analyze the factors which may affect the incident severity.

**Data Section**

To explore the problems, I utilize some database to acquire data.

* The example dataset comes from the web page of Coursera Course *Applied Data Science Capstone.* The .csv file contains 19,4673 pieces of collision records of Seattle. The data for this capstone project is offered by SDOT Traffic Management Division and recorded by Traffic Records Group. It covers the annual collisions data from 2004 to present. The time frequency of this dataset is weekly and it shows the traffic collision records in Seattle.The example datasets contain 194,673 pieces of records starting from 2004. The attributes in the datasets covers the weahter condition, road condition, collision type and fatality.
* The other relevant data comes from the database of Seattle government. The main data source is from Seattle Department of Transportation. The basic background information is offered by this department.
* The geographic data, the map of Seattle is offered by Google Map via the Folium package.

Methodology