

Locations of Accidents in Seattle

Author: Ryan Kennon

Business Problem

How can accident response teams be best deployed? Every City has a limited amount of accident response resources including police, paramedics, and fire departments. This analysis will help inform them about how to better manage staffing and general preparedness for accidents during different weather conditions. Additionally it may help guide infrastructure decisions, what about these high level of accidents locations is causing the spike in accidents and if there is a way to fix them.

Data

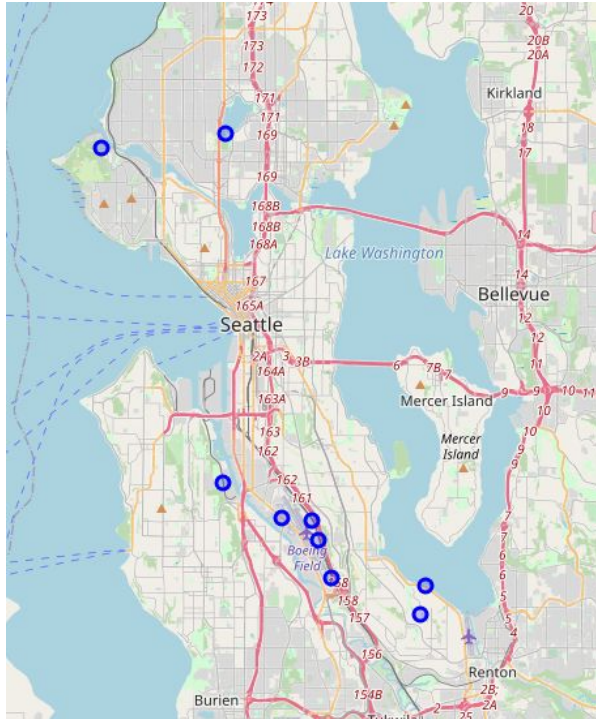
The data set used is a set of all accidents that have occurred in the Seattle metro area since 2004. The data was collected by the SPD and recorded by Traffic Records. The data contains many useful data points regarding our business problem, specifically accident severity, weather conditions, and latitude and longitude of the specific accident. With this data the most common points of severe vs non-severe accidents can be found to recommend where the severe accident response teams and equipment can be better located. Additionally the weather data can be split up and processed to recommend where response resources can be relocated to in the event of a specific weather condition.

Methodology

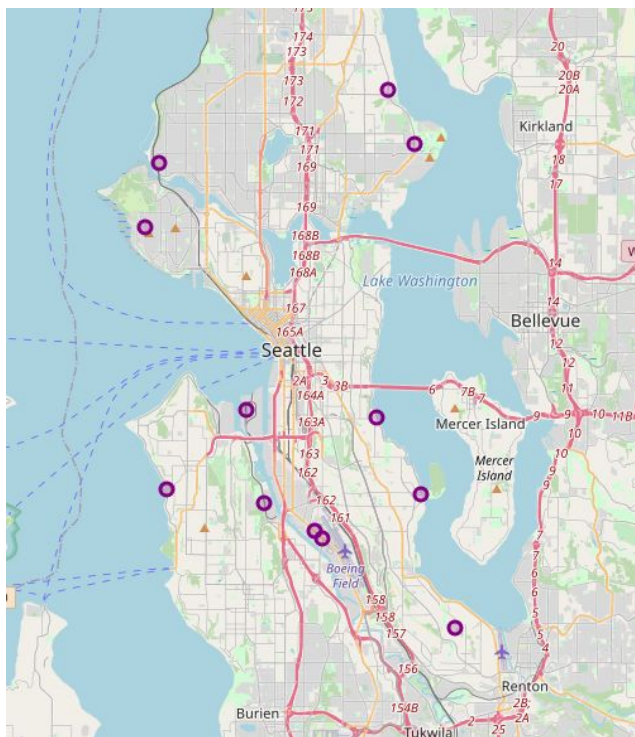
The data was first preprocessed to remove any records from our data set where the main criteria for processing was unfit for analysis. To do this all unknown or unrecorded weather data was removed; followed by the categories of weather where the number of samples would not give a useful amount of clusters of data. Furthermore the excess data that was not useful to this analysis was trimmed off only leaving the columns of severity, weather, latitude and longitude. The data was then split into data frames ready to be processed by a clustering algorithm; based off of severity and into specific weather conditions. Each data frame was then processed with the DBSCAN clustering algorithm and put into clusters the centroid point of the clusters was then calculated. Finally creating a list of geospatial data representing the center of hot spots of where these accidents were occurring to be plotted on a map and analyzed by the SDOT/SPD accident response teams.

Results

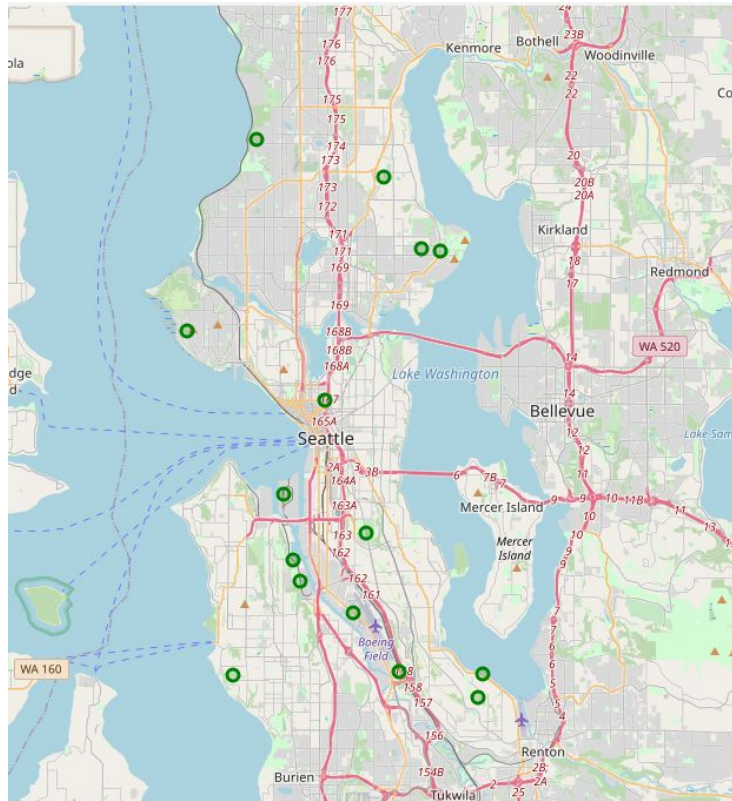
Once the data had been processed and plotted on a map several conclusions could be drawn about accident location from the clustered data, below the blue circles represent accident clusters during clear weather conditions.



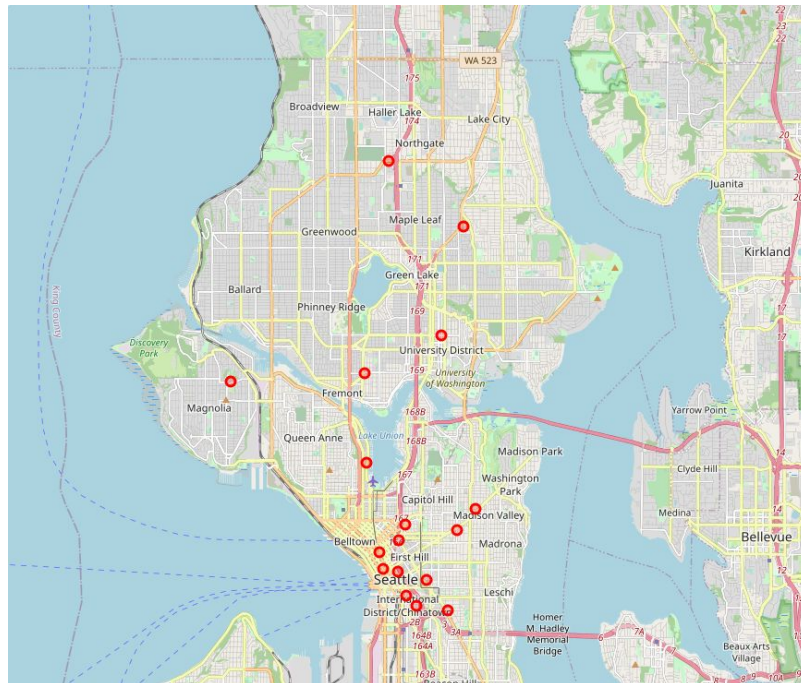
Below represented by the purple circles are the clusters of accidents during overcast weather conditions.



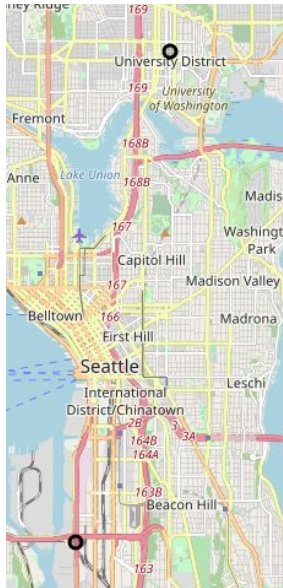
Below represented by the green circles are the clusters of accidents during rainy weather conditions.



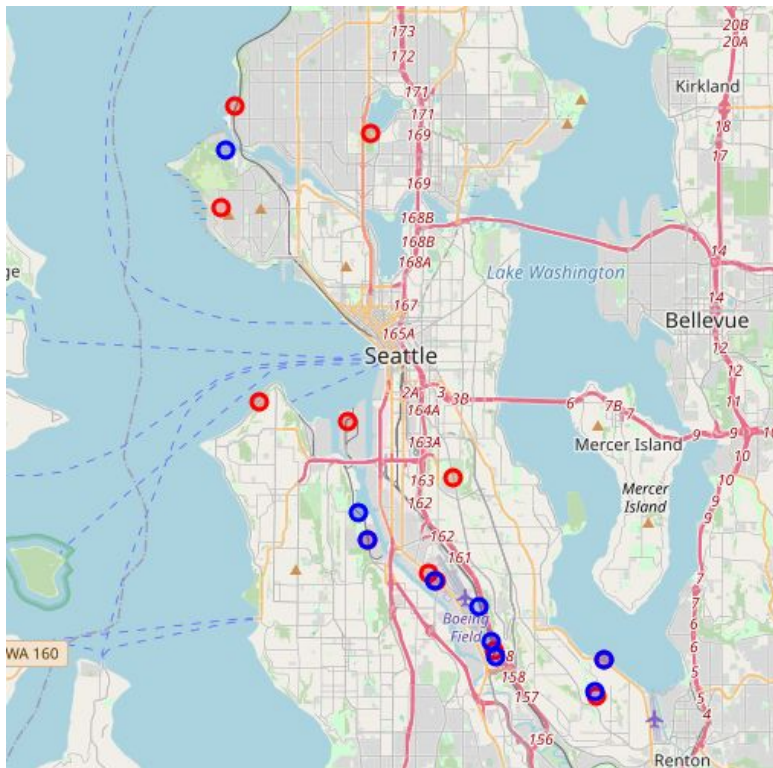
Below represented by the red circles are the clusters of accidents during snowy weather conditions.



Seen below is the Fog/smog/smokey conditions hotspot clusters shown in the 2 black circles below. Clearly demonstrated by the Fog/smog/smokey conditions there was clearly not enough data to gather any real insights into accident patterns which is why there was no more processing of the less commonly seen weather conditions.



And finally the analysis of the severity of the accidents is seen below with the red circles indicating Injury accidents and the blue representing the property damaged in the accidents.



Discussion

Based on the maps shown in the results section a few significant recommendations can be made to the stakeholders in the SPD and SDOT. When the Snowy conditions map is compared with any other conditions there is a significantly higher amount of clusters located in the downtown area indicating that the bulk of resources can be committed to that area when it snows.

Secondly after evaluating the all conditions maps it is clear that a large number of clusters are shared located on coastal roads that are long with many turns, this could be remedied by an analysis of the infrastructure and speed limits in these similar clusters causing high concentrations of accidents.

And finally in the severity of accidents comparison map it is clear that there are high numbers of similarly located clusters to the south of the downtown area that lead to property damage which may help response times and to locate more response teams geared towards property damage south of the city center.

Conclusion

Of the few suggestions I made by this clustering data, more useful insights can be gleaned from the information given the knowledge of the stakeholders, the SPD and SDOT. Additionally more insight may be gained from the inclusion of the road condition data however there were too few data samples left blank to learn much; if this was collected more frequently this metric could provide more insight into infrastructure problems causing accidents.