Capstone Collisions Report

1. Introduction

Roughly 38,000 Americans are killed in car crashes every year and many of these crashes could have been prevented. The issue that we are seeking to understand is why car accidents happen. If through the data set we can see what attributes the crashes have in common and what could be contributing to the problem, then changes can be made to reduce the severity and number of accidents. This will save time, resources and more importantly, lives.

2. Data

The source of the data is the traffic records from the Seattle Police Department. This data contains information about the conditions under which the crash occurred, the location, the severity of the collision as well as other aspects. When looking through the data I became more interested in whether or not weather has an effect on the severity of the accidents. While most accidents happened when it was clear, enough were under different weather conditions, such as raining or overcast, that I thought this could be significant and something I would like to explore. I also wanted to know how weather affects the severity of the accident. While weather is not something that can be changed or controlled, the city could make decisions such as spending more money on snow plowing equipment or providing driving lessons about how to drive in the rain, that could bring the accidents into check if weather was a factor. To get a closer look at the data I wanted to work with I decided to make a separate data frame just looking at weather and the severity code of the accident. This data frame was called df_weather.

3. Methodology

With the new df_weather data frame, I used value_counts() to see the breakdown of accidents under each possible weather condition.

```
In [36]: df['WEATHER'] .value_counts().to_frame()
```

Out[36]:

	WEATHER
Clear	111135
Raining	33145
Overcast	27714
Unknown	15091
Snowing	907
Other	832
Fog/Smog/Smoke	569
Sleet/Hail/Freezing Rain	113
Blowing Sand/Dirt	56
Severe Crosswind	25
Partly Cloudy	5

I couldn't tell too much just from looking at this and had to do further analysis. I calculated the percentage of accidents that happen for each type of weather for the whole data set.

```
In [58]: (df_weather.groupby('WEATHER').size()/df['WEATHER'].count())*100
```

```
Out[58]: WEATHER
         Blowing Sand/Dirt            0.029537
         Clear                       58.617980
         Fog/Smog/Smoke               0.300118
         Other                        0.438837
         Overcast                    14.617705
         Partly Cloudy                0.002637
         Raining                     17.482278
         Severe Crosswind             0.013186
         Sleet/Hail/Freezing Rain     0.059602
         Snowing                      0.478396
         Unknown                      7.959724
         dtype: float64
```

About 58.62% of accidents happened when the weather was clear, so this alone did not give me too much information. To understand how different weather conditions impact the severity of these accidents, I made a separate data frame called df_severity that only contained the

accidents where the crashes were more severe, meaning the severity was 2. With this data frame, I again looked at the percentages of the weather for each accident.

```
In [59]: (df_severity.groupby('WEATHER').size()/df_severity['WEATHER'].count())*100

Out[59]: WEATHER
         Blowing Sand/Dirt            0.026268
         Clear                       62.762679
         Fog/Smog/Smoke               0.327473
         Other                        0.203138
         Overcast                    15.314164
         Partly Cloudy                0.005254
         Raining                     19.571308
         Severe Crosswind             0.012258
         Sleet/Hail/Freezing Rain     0.049033
         Snowing                      0.299454
         Unknown                      1.428972
         dtype: float64
```

Although not by an extremely significant amount, more of the severe accidents happened during clear weather than compared to looking at both the severe and less severe crashes.

4. Results

I saw that 58.6% of crashes happened when the day was clear and 17.48% when it was raining. In the case of only looking at severity=2, 63.76% of crashes happened with clear weather and 19.57% were when it was raining. Therefore, it does not appear like bad weather made accidents more severe like I had thought would be a possibility. While there was a slight increase in percentage of severe accidents with bad weather, such as rain, a higher percentage of the crashes also happened during clear days. This is not the result I had expected.

5. Discussion and Conclusion

I did not see the result I was hoping to in this test, but since it was what I decided to look for in the earlier sections of the course, it was what I had to investigate. The difference in this could be that people are might be more careful when the weather is not good for driving because they know that they have to stay alert to stay safe. This extra attention and care could

potentially outweigh the additional hazards that rain, snow or other extreme weather conditions cause on the road. It appears that investing in strategies to deal with extreme weather or teaching drivers to better deal with un-ideal weather conditions would be a waste of money for the city as it would likely not have much of an impact on lessening the severity of the accidents in Seattle.