Capstone Final Project Data Collision in Seattle

Introduction/Business Problem

I will research and evaluate the dataset "car accident severity" in the city of Seattle in this final capstone project. I will analyze and determine how weather conditions affect the severity of car accidents. It's very common for people to relate bad weather to more road accidents, so we usually tend to drive more carefully when the weather is not favorable. Does bad weather directly translate to more car accidents? That is exactly what we will try to verify – if the bad weather plays a big role in causing more car accidents.

Data

The dataset used for this final project is the Example Dataset, Data-Collisions, city of Seattle, provided by IBM/Cousera. The dataset includes data from 2004 to 2018 that address all types of collisions, it includes many types of attributes that can have some impact in the accident itself and the severity of it. The dataset consists of 194673, each row represents an accident consisting of 38 factors that may affect the severity of the accident.

Methodology

I used the IBM cloud Pak for Data – Watson Studio – to build a notebook and pushed the report on my Github repository to facilitate review. I used Python as the primary language and imported numpy and pandas libraries.

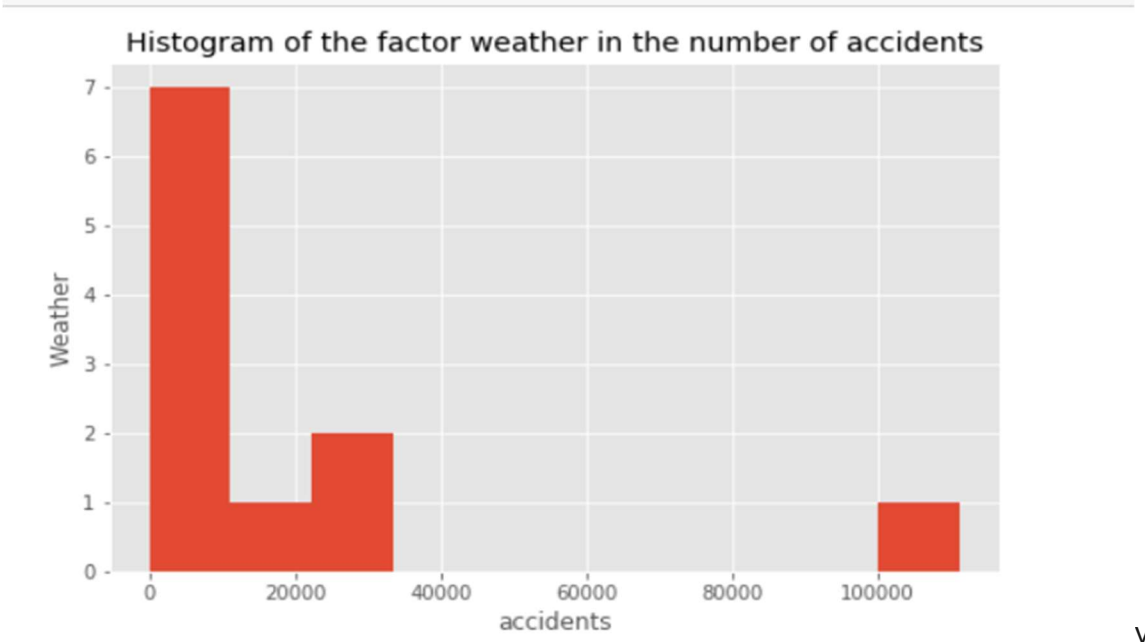
I downloaded the cvs file, printed the data types to review the columns I needed to review and analyze my question. I was also able to retrieve part of the data by printing the "head" of the cvs file. Then, I dropped all the columns I didn't need for this project and renamed the column I would be working with the purpose of making the reading of the results more intuitive. I grouped the data from the two colums (weather, number of accidents) by the weather, so I could get a count of the number of accidents by the

weather condition. I downloaded matplotlib to build some data visualization, histogram and bar graph, to analyze the results**.**

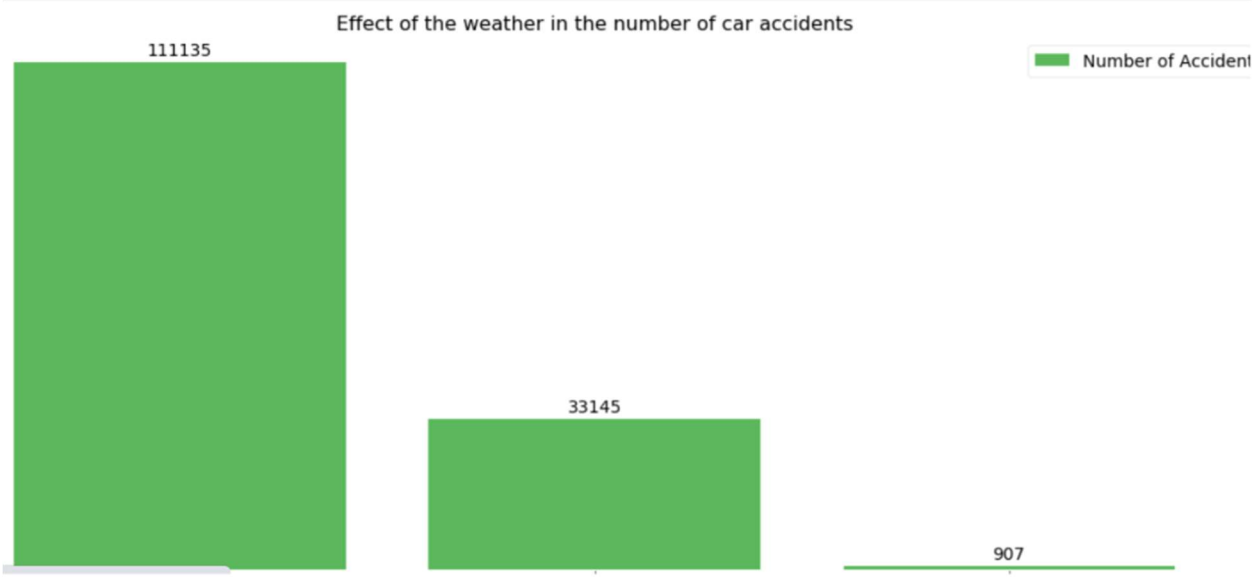
WEATHER	Number of Accidents
Blowing Sand/Dirt	56
Clear	111135
Fog/Smog/Smoke	569
Other	832
Overcast	27714

Results

The results of the data are showing me that car accidents happen mostly in clear weather. Raining weather came in second with the greatest number of accidents and surprisingly weather came in fifth behind overcast and unknown weather.


Histogram of the factor weather in the number of accidents
Weather
accidents
0
1
2
3
4
5
6
7
0
20000
40000
60000
80000
100000

y snow


Effect of the weather in the number of car accidents
Number of Accident
111135
33145
907

Discussion

I consider the results of this data to be completely unexpected as I always thought that bad weather increases the risk and it's associated with more car accidents. Every time it snows, I actually see more accidents in the roads. It's important to mention that this project just take in consideration the weather and there are many other factors that have a great influence in car accidents; for example, road conditions. Also, there are other relevant variables that aren't part of the dataset, such as time of the day, traffic condition, age of the driver etc. I understand that looking at only one variable for this kind of analysis doesn't mimic the reality. So, more variables we use to analyze car accidents, closer we will get from the reality.

Conclusion

If you take in consideration the weather as the only variable influencing car accidents, we will get unexpected results, such as snow weather have less accidents than clear weather. The results of this data tells me that most accidents occur in clear weather, while it may be true to this variable, we have limiting factors that may be more important than the weather in the occurrence of these accidents. Therefore, in order to have results that approximate form the reality other variables would need to be included in this study.

THANK YOU!