Paola Torres

Wicaksa Munajat

Cheuk On Yim

Matthew Stoney

CST-383, Prof. Memo

June 17, 2022

**Final Project Report:
Can weather conditions predict the severity of Accidents in Santa Clara County?**

Introduction

Millions of accidents occur each year across the United States. They range by type, consequence, and severity. With this project we hoped to study if a specific factor, namely weather, had any effect on the severity of accidents. Some insight into this might help motorists rethink their driving habits during these conditions.

In this project we posed the question of do weather conditions, specifically precipitation and visibility, have an effect on the severity of the accident?

Selection of Data

The dataset that we used was titled *US Accidents (2016 - 2021).* It was obtained from Kaggle. This was a large dataset containing data points from across the United States. The dataset contained nearly 3 million rows and 47 columns, so we decided to narrow our focus to data specifically from the County of Santa Clara. Along with this we dropped columns that had missing information and created a subdataset that contained columns related to different weather conditions (i.e. visibility, windspeed, precipitation. etc.)

Methods

We primarily used numpy, pandas, matplotlib, and sklearn to process and begin to answer the research question through a Google Colab notebook. We used the following from sklearn for the machine learning portion of the project: KNeighborsClassifier, StandardScaler, train_test_split, classification_report, confusion_matrix, accuracy_score, and mean_squared_error.

Our dataset was scaled and trained. The features used were precipitation and visibility. The target was severity. The machine learning model used was KNeighbor Classification.

Results

Based on the data, San Jose has the most accidents in the County of Santa Clara (Refer to *Figure 1* at the end of the report). This makes sense as it is the biggest city in the county. Plotting the data by weather condition reflects the most number of accidents when the weather is fair (Refer to *Figure 2*). However, the combination of accidents by all other conditions is greater. There was also a positive correlation between humidity and number of accidents (Refer to *Figure 3*). When it came to the number of accidents by severity, the most accidents occurred in the level above minimal severity (Refer to *Figure 4*). In regards to the two weather conditions of focus, precipitation and visibility, precipitation carried the most accidents between the two.

The correlation was not strong between individual weather conditions and the severity of accidents. However, based on our graphs, there is a definite effect caused by weather on the severity of the accidents.

One caveat to consider is that the weather in Santa Clara County is mostly fair, and this can explain the fact that most accidents occur during fair weather. This may go against common sense that bad weather causes accidents, but if there is never snow or ice, and very little rain or fog, most accidents will occur during fair weather. This can skew the results. This made it difficult to predict with the machine learning training model we created whether an accident would be severe or not based on the weather conditions. Further exploration of the question could be done using a more expansive dataset that includes information from all over the United States.

Discussion

The finding that individual weather conditions do not have a strong correlation with the severity of accidents is important because this shows that all types of weather can cause accidents in general. However, together, weather as opposed to a fair day, does have an effect on accidents. This latter part has been replicated by other researchers. Future research on the subject might look into what factors do affect the severity of accidents.

Summary

The most important take on this is that millions of accidents occur across the United States every year. Even though specific weather does not necessarily predict how bad an accident will be. Weather conditions can increase your chances of being involved in one.
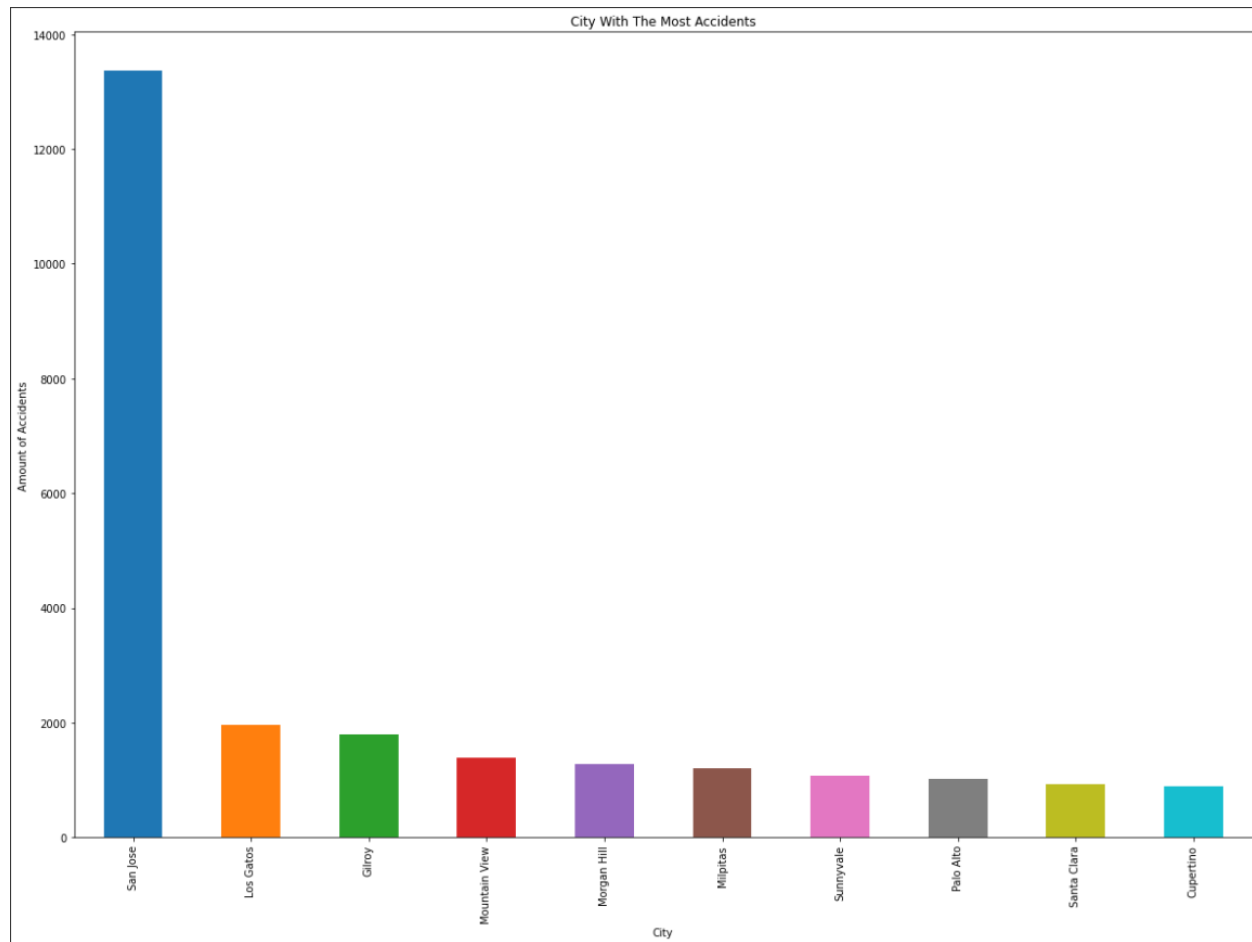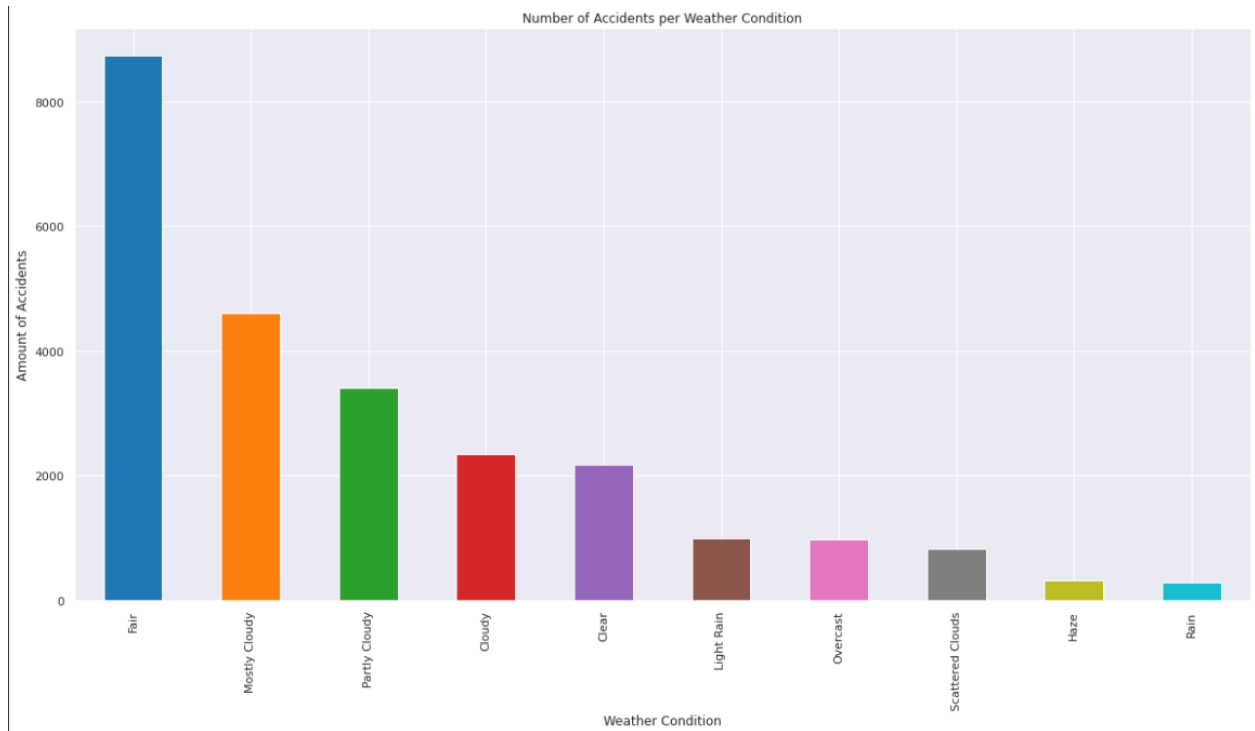
Figure 1: City with the Most Accidents

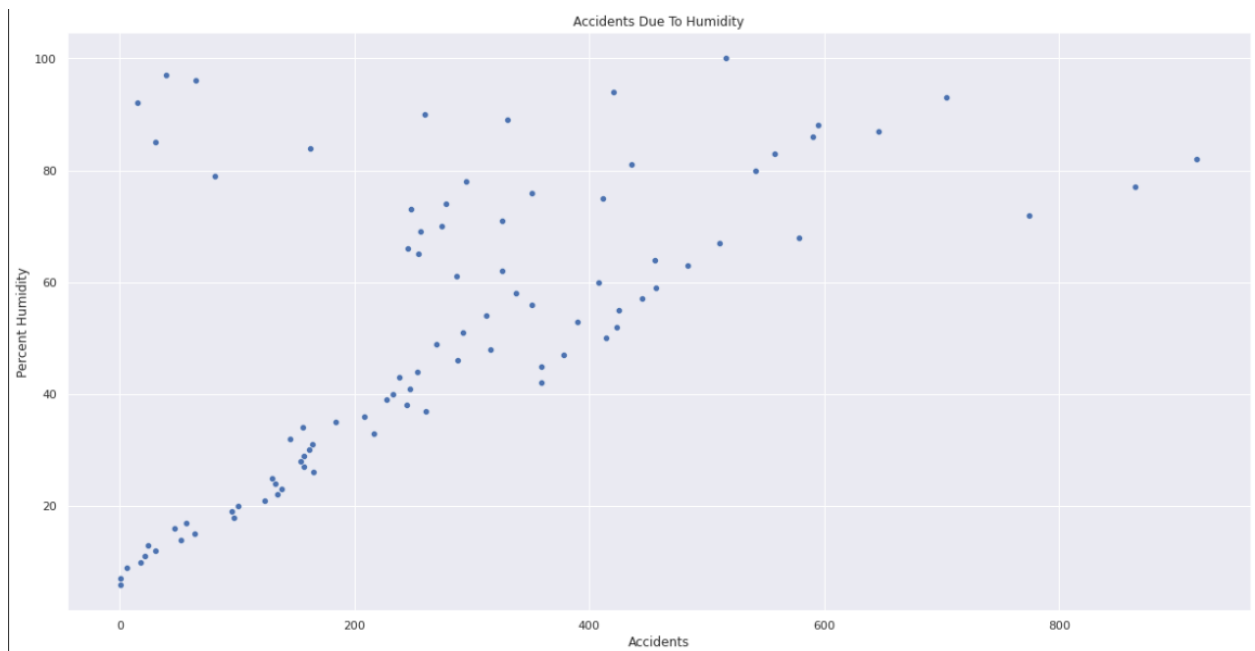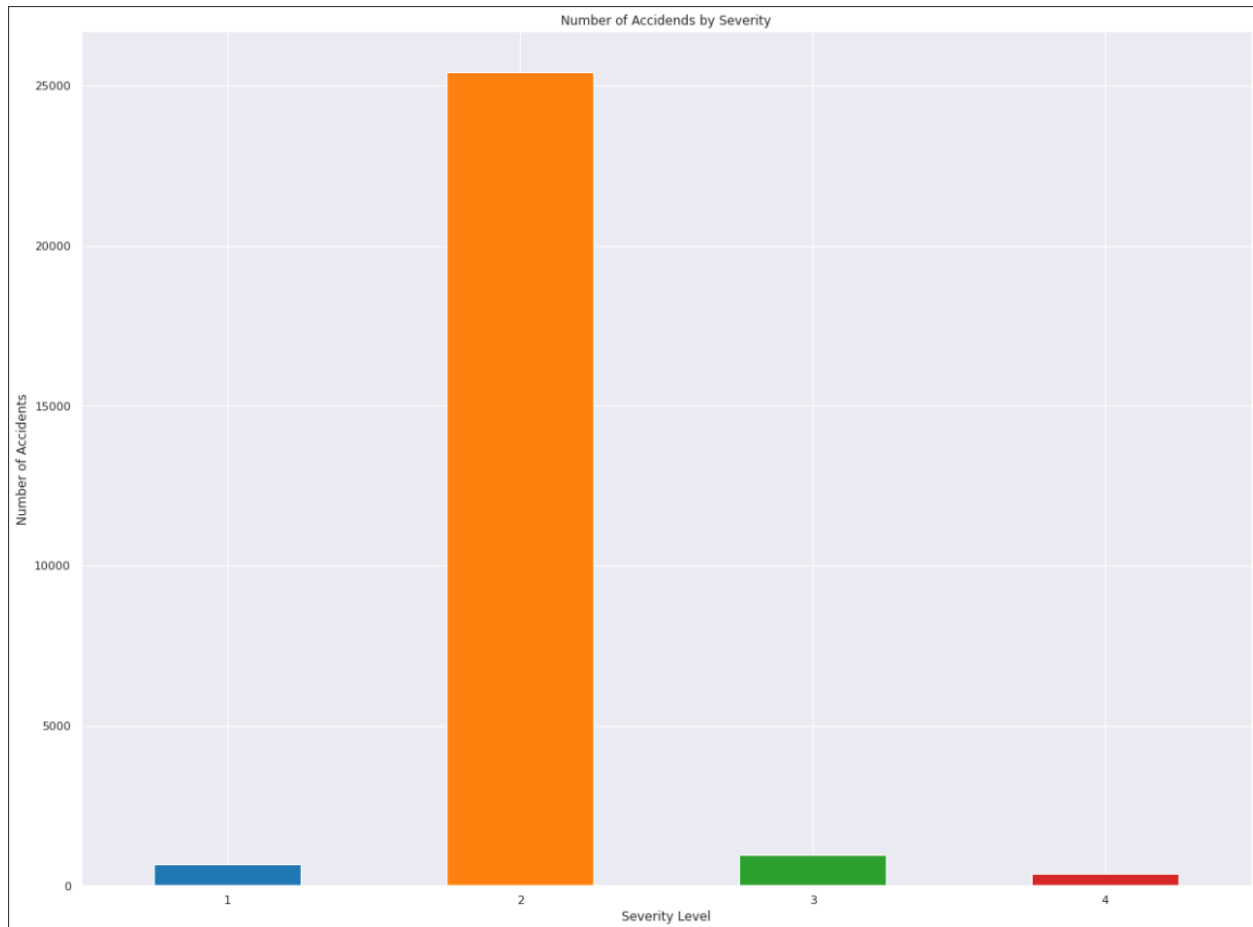Figure 2: Number of Accidents per Weather Condition

Figure 3: Accidents Due to Humidity

Figure 4: Number of Accidents by Severity

References

Moosavi, Sobhan, Mohammad Hossein Samavatian, Srinivasan Parthasarathy, and Rajiv Ramnath. "A Countrywide Traffic Accident Dataset.", arXiv preprint arXiv:1906.05409 (2019).

Moosavi, Sobhan, Mohammad Hossein Samavatian, Srinivasan Parthasarathy, Radu Teodorescu, and Rajiv Ramnath. "Accident Risk Prediction based on Heterogeneous Sparse Data: New Dataset and Insights." In proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, ACM, 2019.