

Car Accidents analysis

Correlation between accidents from 2004 to 2019

Problem

We want to analyze a dataset containing the accidents made in Seattle from 2004 to 2019. The dataset is rich in information and will allow us to make some analyzes on road deaths and any correlations with factors related to drunk or drug driving and wheater conditions.

Interest

The analysis will arouse interest in local surveillance, law enforcement but also the health department.

Data cleaning

Before carrying out analyzes it is necessary to proceed with a data cleaning. The table contains a lot of data, some deemed superfluous for our purpose.

The columns we have left are those that indicate the coordinates of the incident, the severity code, the total people, the pedestrians and cars involved, the date of the incident, the weather and road conditions and finally driving under alcohol or drugs. This last column has been normalized bringing all values to 0 or 1 to indicate driving in an unaltered state and driving in a confused state, respectively.

Some columns have been converted to other formats to allow better use of the data within them. An example is the date column of which only the date was taken with precision to the single day. The hours have been deleted.

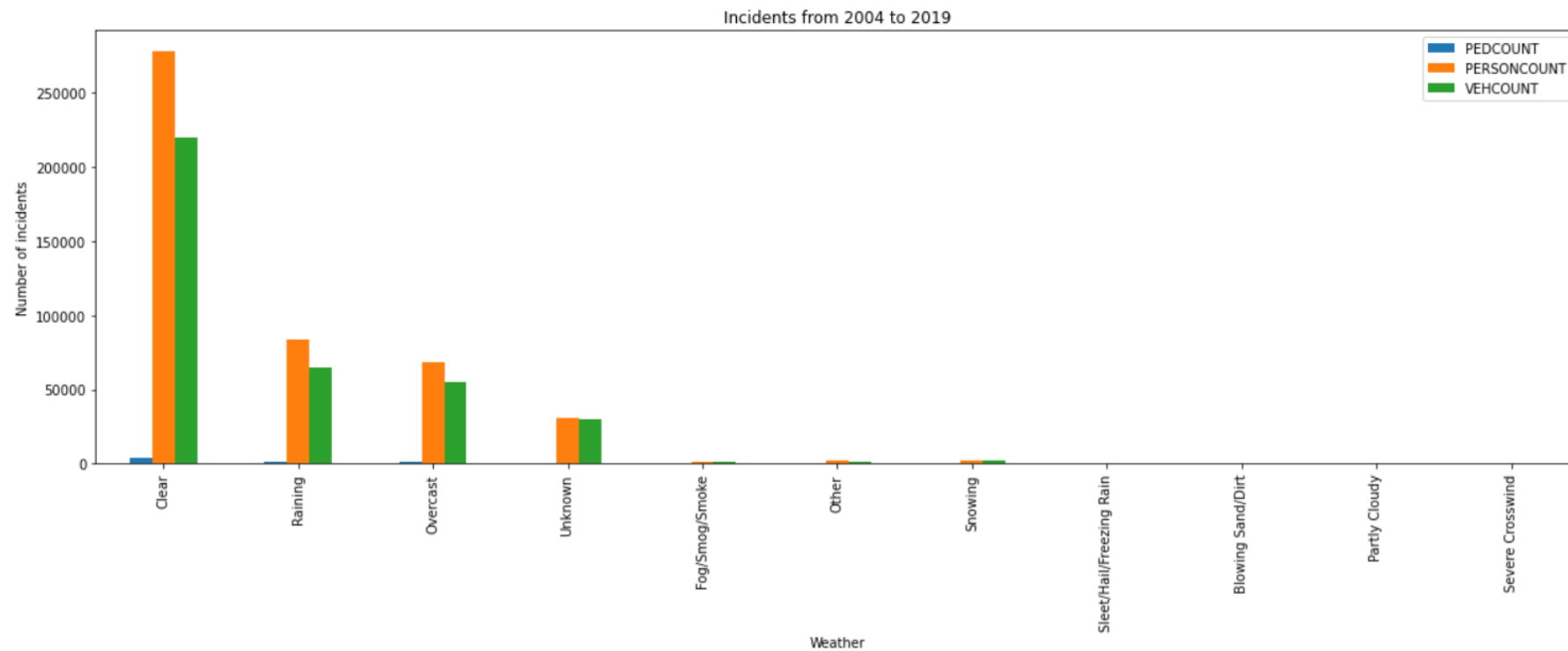
All rows containing NAN values have been deleted.

Data handling

A smaller data frame was created with the selected columns. Then a sorted pivot table was created in order to understand how climatic conditions affected the accidents in numerical terms.

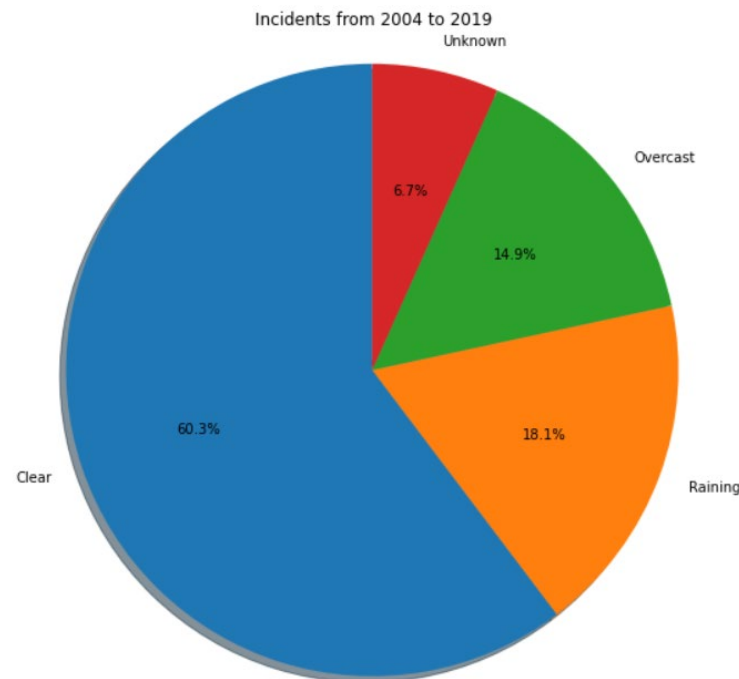
Graphics

This graph displays accidents based on weather conditions.



Relationship between weather and accidents

The previous graph surprisingly shows that most accidents happen in the least dangerous weather conditions! This is likely due to drivers being more cautious in adverse weather conditions and less traffic in rainy weather. Below is a pie chart that highlights the four most dangerous weather situations.



Linear correlation between variables

One wonders if there is a correlation between the number of people involved and drivers in an altered state of consciousness. Numerical evidence is shown below. The analysis was done considering three subsets of data randomly taken from the entire dataset.

```
[0.00046160848415111033, 0.00032322613981727866, 0.000602270230692481]
```

We note how the correlation is almost zero. Therefore the variables involved are not dependent on each other.

Taking the variable degree of severity of the accident and the drivers in an altered state of consciousness we notice how the situation changes very little. However it is greater than before.

```
[0.0014649152292338652, 0.0017240418236708077, 0.002288182844122133]
```

Conclusions

Contrary to what was expected, accidents are not related to the type of driver in an altered state or not, but to the danger of drivers in the best climatic conditions. Inattention and the feeling of lack of danger is more dangerous than the meteorological danger!