Influence of outer conditions and driver behavior on severity and size of car accidents

Abstract:

In this text we will analyze the influence of outer conditions like weather, light and road conditions and driver behavior like driving under the influence of alcohol, speeding and being inattentive, disrespecting pedestrians right of way, on the severity and size (number of vehicles and persons involved) of accidents.

Introduction:

In this project we want to find out how the circumstances under which a collision occurs influence its severity. As almost everyone drives, walks, or cycles through cities frequently, this can be interesting for everyone. Certain behavior, for example speeding, driving under the influence of alcohol or drugs, can be avoided by any driver. Knowing the risks can encourage changing the behavior. While we cannot change the weather, we can decide to postpone a trip to the city, take the train or adjust our speed.

As well for politicians and city employees making traffic laws, planning cities (e.g. deciding on speed limits) and the maintenance of roads (e.g. clearing snow), the results of the data can be interesting.

Data:

We will be using a data set of 194673 collisions which occurred in Seattle between the first of January 2004 and the 20th of May 2020.

We are mainly Interested in the following columns contained in the data:

Data describing the severity and size of the accident:

SEVERITYCODE:

A code corresponding to the severity of the collision. The code distinguishes between incidents with only property damage, incidents where at least one person was injured, incidents where at least one person was seriously injured and incidents with fatalities. The data at hand only contains the severity codes for incidents with only property damage and incidents with at least one person injured.

PERSONCOUNT, PEDCOUNT, PEDCYLCOUNT and VEHCOUNT:

Number of people in total, pedestrians, bicycles, and vehicles involved in the accident. For bigger numbers we will not look at the exact numbers, but rather group large accident.

Data describing the general conditions:

WEATHER:

One of the following: 'Overcast', 'Raining', 'Clear', 'Snowing', 'Fog/Smog/Smoke', 'Sleet/Hail/Freezing Rain', 'Blowing Sand/Dirt', 'Severe Crosswind', 'Partly Cloudy' or ‘other’/’unknown’

ROADCOND:

One of the following ‚Wet', 'Dry', 'Snow/Slush', 'Ice', 'Sand/Mud/Dirt', 'Standing Water', 'Oil' or 'Other'.

LIGHTCOND:

One of the following: 'Daylight', 'Dark – Street Lights On', 'Dark – No Street Lights', 'Dark – Street Lights Off', 'Dusk', 'Dawn' or 'Unknown'. We will treat 'Dark – Lights Unknown' as 'Unknown', as we do not have enough data in this category to treat it separately and we have no way of knowing if they should be in the category 'Dark - Lights On' or 'Dark - Lights off'.

Condition/behavior of the driver:

SPEEDING, INATTENTIONIND:

Was speeding, inattention a factor in the incident?

PEDROWNOTGRNT:

Was a pedestrian's right of way not granted?

UNDERINFL:

Was a driver under the influence of alcohol/drugs involved?

Some of the categorial columns contain values like 'Unknown' and 'Other' as well as NaN. We will replace all those values with 'Unknown'. We will convert all data in boolean data, for the columns containing several categories (e.g. WEATHER) we will add a new column for every category, but drop the values for the 'Unknown' columns as they do not add any information.

The dataset contains further data e.g. on the location of the incident and a description of the collision. In this project we will focus on the factors mentioned above. Analyzing all possible factors would be out of scope of the project.

Although only severity code 1 (property damage) and 2 (injury) are contained in the data, we decide to use the data provided. While more data would be interesting for further investigation, we believe, that the data set at hand will give us enough inside for a first investigation and we will get a good idea what to further investigate with more data in later projects.

The data as well tells us very little about who was injured. We neither have the number of injured people nor do we know, if the injured people where in a car, on a bicycle or pedestrians.

Methodology:

Rather then getting a prediction, we want to find out which factors lead to more severe accidents. A decision tree not only gives us a prediction, but we can see, how the predictions is derived from the data. Therefor we will train and visualize a decision tree.

We will first look at all the rows in the table. We will as well have a look at only the accidents without pedestrians or cyclists involved.

Results:

When we look at the decision tree including all the accidents, it becomes evident that the most impo rtant factor if someone gets injured in an accident, is if there where cyclists or pedestrians involved. In total there are 194673 accidents in the table.

There are injuries in 58188 of the accidents, i.e. in 29.9 %.
There are 6939 accidents with pedestrians involved. Among the accidents involving pedestrians there 6241 with injuries, i.e. 89.9%.
There are 5391 accidents with cyclists, but no pedestrians involved, including 4715 with injuries, henc e there are injuries in 87.5 %.
The percentage of accidents with injuries is nearly three times as high when there are pedestrians or cyclists involved.
If we only regard accidents without pedestrians or cyclists, the accuracy score of the prediction by th e decision tree is about the same as when we predict that there are no injuries in all the accidents. In the given data set the influence of the factors that we considered (e.g. weather, light, road condition s, speeding) on the severity of the accident seams to be not very high.

Discussion:

To get more conclusive results, we should look at other models and more data. As well scaling of the data might lead to better results. This first investigation shows little influence of weather, light, speeding, and the other factors on the severity of accidents in Seattle. It would be interesting to look at data from other cities as well as for accidents outside of cities.

Here we only compared the severity of accidents defined by whether there are injuries or not. We could find other definitions for the severity of accidents, like costs of the damage, number of injured people etc.

It would although be interesting to compare the number of accidents per day depending on the weather. While the data gives no good evidence, that accidents are worse when there is ice on the street, there might be more accidents.

In most of the accidents in the table not involving cars there where injuries. That does not necessarily mean, that there are no minor accidents between pedestrians and cyclists or between several cyclists. The reason might also be, that those accidents are often so minor, that the police is not notified, and those accidents are never included in any statistic.

We do not have any data on how many cars, pedestrians and cyclists are participating in traffic, let alone how this proportion changes e.g. due to the weather.

Although it seams likely, that in most accidents where there are cyclists or pedestrians and cars involved, if someone is injured, it is the pedestrian or cyclists, we do not get this from the data. Further investigation with more data would be necessary to get good evidence that walking and riding a bike is dangerous in Seattle.

Conclusion:

This little investigation gives rise to the assumption, that accidents in Seattle have worse consequences for cyclists and pedestrians then for people in cars.

For further investigations on the influence of weather, road conditions, light, speeding, use of substances etc. more or different data would be needed.