Wilde's Coursera Capstone

Car accidents

Introduction: Business Problem

Car accidents occurs everywhere worldwide and is one of the leading causes for people. According to World Health Organization (WHO), roughly 1.35 million people die from traffic collisons. More than half of road traffic deaths involve users such as pedestrians, cyclists, and motorcyclists. Road traffic accidents also attribute to 3% of the domestic product. The goal is to identity relevant factors and derive insight on what events lead to these car accident and what severity is depended on.

The main audience of this project is road services, rescue services and city administration.

Questions:

- What is the most common cause?
- Are there relationships between causes?
- What are the first things to look for to prevent car accidents?

Data

The dataset, containing roughly ~250,000 records as of Oct 2020, was used for analysis. It contains data collected from 2004 to 2020 and is based on accidents taken place in the state of Washington, Seattle. For each car accident, a severity (1 = prop damage and 2 = injury) code is assigned as well as other relevant information such as:

- location
- speeding involved
- road condition
- collision type
- weather condition
- lighting condition
- driver inattention
- number of people involved

These data points will be analyzed to what the major influences in car collisions.

Methodology

According to the data provided, the most important task is to understand the relationship between different factors.

I will be using visualization and correlation tools.

In this case, machine learning tools are not really needed, since there is no task of predicting something in the project. Our main task is to understand the relationships.

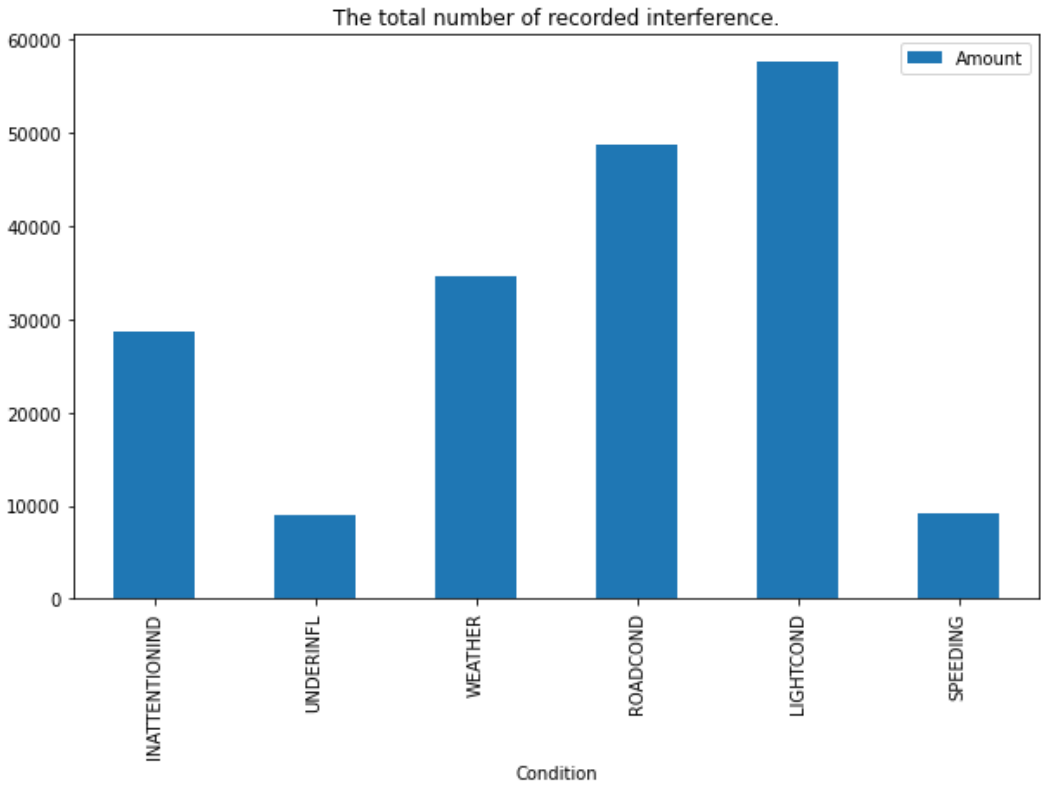
Analysis

Graph of the number of incidents by year, 2004-2020.

Correlations

	INATTENTIONIND	UNDERINFL	WEATHER	ROADCOND	LIGHTCOND	SPEEDING
INATTENTIONIND	1.000000	-0.036292	-0.031633	-0.035152	-0.063941	-0.059372
UNDERINFL	-0.036292	1.000000	0.004086	0.008116	0.225326	0.088081
WEATHER	-0.031633	0.004086	1.000000	0.757512	0.161445	0.108626
ROADCOND	-0.035152	0.008116	0.757512	1.000000	0.178592	0.134975
LIGHTCOND	-0.063941	0.225326	0.161445	0.178592	1.000000	0.087299
SPEEDING	-0.059372	0.088081	0.108626	0.134975	0.087299	1.000000

Bad Weather + Bad road have 75% correlation!
Dark + drugs = 22%
Bad weather + dark = 16% corr
Bad weather + speeding = 10%
Bad road + dark = 17%
Bad road + speeding = 13% corr.



Results and Discussion

A bad road is the main aggravating factor. As we can see, Bad Weather + Bad road have 75% correlation, Bad road + dark = 17% and Bad road + speeding = 13% corr.
Poor lighting is also a big problem - the number one factor.
The driver's inattention, on the contrary, does not correlate with anything, which suggests that this is an independent factor that cannot be controlled.
Drugs and alcohol go very badly with poor lighting.
Speeding does not go well with bad weather and bad roads.

Conclusion

We can’t control weather, but can control the condition of the roadway and artificial lighting.
But at the same time, the total number of accidents has been steadily decreasing since 2016, which indicates the correct trend.

As a recommendation, I recommend that you especially carefully monitor areas with poor road surfaces and poor lighting in bad weather.