Coursera Capstone

IBM Applied Data Science Capstone project

Car accidents in South Moravia (CZ)

By: Ing. Josef Sevcik

Brno 2020

Introduction/Business Problem

Brno and South Moravia are the second biggest city and region in the Czech Republic and currently have a big issue with transport, especially by car. In 2008 South Moravia has more than 7 thousand registered car accident. In this time Brno has around 400 thousand inhabitants. Main proposes of this project is to find and highlight the main traffic area map with the highest possibility for car accident based on data source from Car accident police records datasheet. I choose for the now smaller area as the city. In the future, I can extend this are to NUTS (Nomenclature of Territorial Units for Statistics) 3 (Regions) or NUTS 1 (Country).

Hypotheses:

Which area has a higher possibility of a car accident?

What time and date have those area higher possibility of a car accident?

Prerequisites:

Set rules for accidents' severity 0, 1, 2 (2 highest).

Include data which could have higher k-value for car accident and injuries, death coll and higher value of the damage.

Exclude from data set those data, which does not have k-value, such could be the technical malfunction of a vehicle.

Future of project:

An extended project could be assign with: Including weather as the possible key of the current statuses on the roads. Including roads obstacles and repair via GIS. Including live car accidents on the road (real data).

The project could help drivers, to inform them or avoid to travel via those place, where is a higher possibility of the car accident. This source could be possibly applied to Mapy.cz (local service provider similar to Google Maps or Waze) to inform drivers before they drive or to calculate another road.

Clearly define a problem or an idea of your choice. Remember that data science problems always target an audience and are meant to help a group of stakeholders solve a problem, so make sure that you explicitly describe your audience and why they would care about your problem.

This submission will eventually become your Introduction/Business Problem section in your final report. So I recommend that you push the report (having your Introduction/Business Problem section only for now) to your Github repository and submit a link to it.

Data section

Currently, there are more sources for this project. However, I consider choosing this data: https://kod.brno.cz/dataset/8afbbf81-6c34-4182-b8a7-bba959f0dfe1/resource/7cbaeb26-2bb0-4f20-99a6-cdbbc5802d35/download/nehody-brno.xlsx

Also, there are available statistics which are not a good source as data. But also there is a source available here: https://www.policie.cz/clanek/statistika-nehodovosti-900835.aspx and it is for the whole Czech Republic but in the local language. In this case, you can download the RAR object, which contains all NUTS 3 data (Regions), but not only for Brno city. Those data could be possibly translated to English and could be collected for many years in the back for timeline which could cause better interpretation for future and in case of extending for NUTS 1 (Country).

By adding API to this could be add also are which are currently blocked or there are any other obstacles. https://www.doprava-brno.cz/

First data set to contains a lot of information, including the weather, light, status of the road, victims of accident, time and date and the most important GPS location. I would like to evaluate the car accident severity number from those data. The possibility will be the death toll as the highest rating.

Also weather condition could by possibly ad via API throw various data sources.

Methodology

The report is represents by main components as mentioned above from Police record of car accidents in South Moravia in Czech republic. Unfortunately, as it can be visible in python file, there is very low correlation. That why was created column called score, which assume by correlation, that few columns are influencing the car accidents.

GPS data (can be find in dataset as X and Y) have to be transferred to readable format, which is support by folium.

Data was dived into 3 groups. Group 0 was excluded as there was low possibility for car accident or it was coincidence. Also all data where reports said that was connected with alcohol and drugs were excluded.

Results

Statistically we can said that during year 2018 in South Moravia were 3187 car accidents. From that 42 persons died. Heavy Injuries have 291 persons. Light Injuries has 2514 people. The damage was calculated as 181.095.100 CZK (around 8 146 795 USD) meaning that each car accident has damage around 3.000 CZK (135 USD).

The most car accidents were during the week high except Saturday and Sunday, as it is visualized in bar graph below. As hypothesis was told to predict time and date, this hypothesis cannot be answer, as on the market there are no data source collecting time.
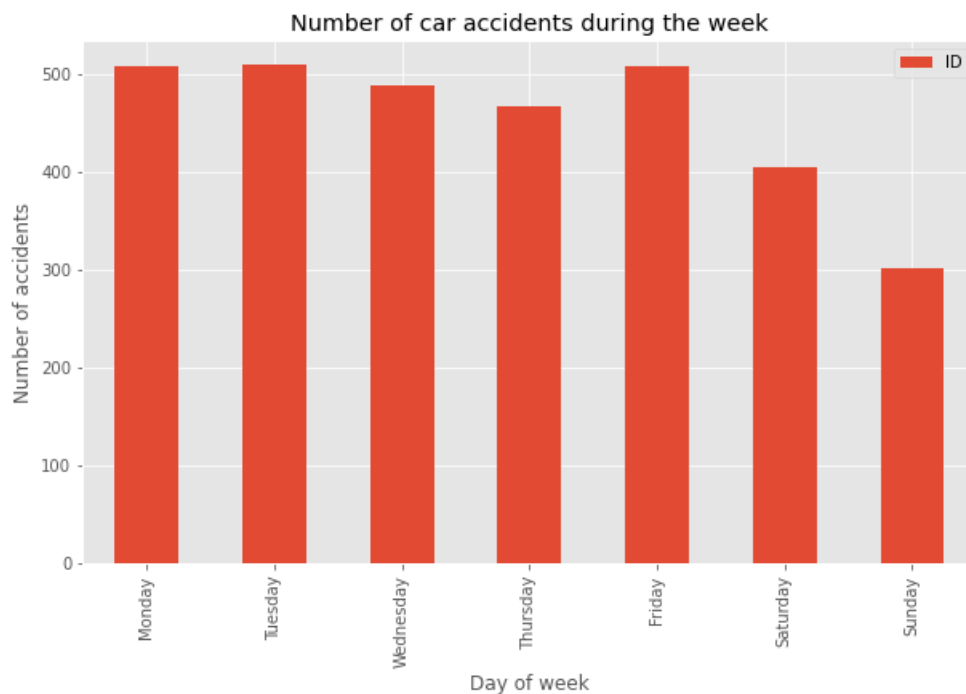
As date we can say that during February and March is lower change to have car accident that in other Month of year.
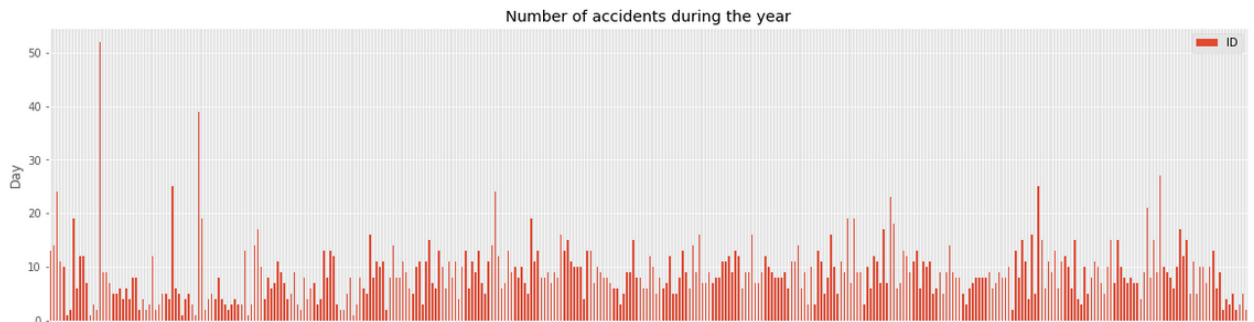

The main reason for car accident were inappropriate driving or inappropriate speed.
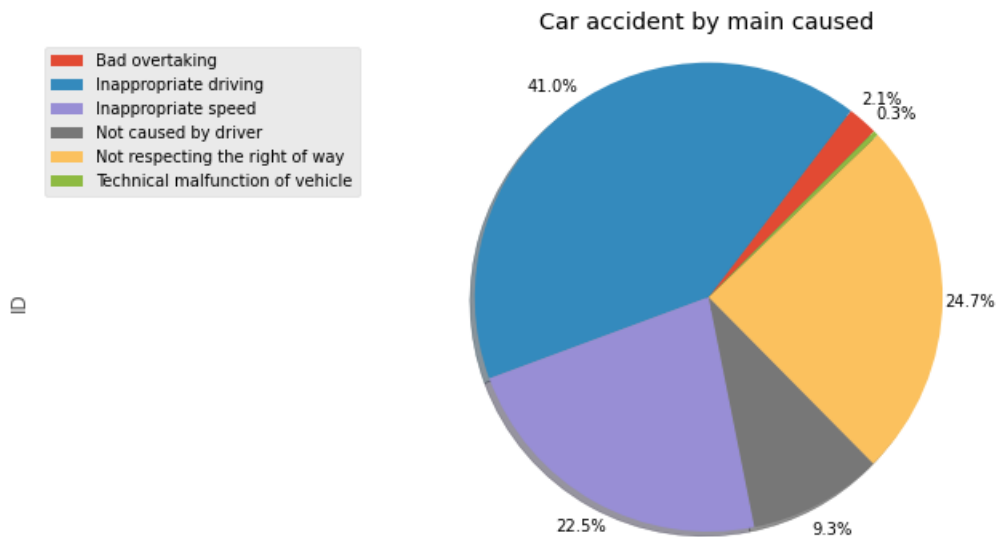
Caused place of accident was usually not reported. The main contributor otherwise is pedestrian crossing or nearby. Others are transport stops, bridges, overpass or underpass.

Number of car accidents by caused place
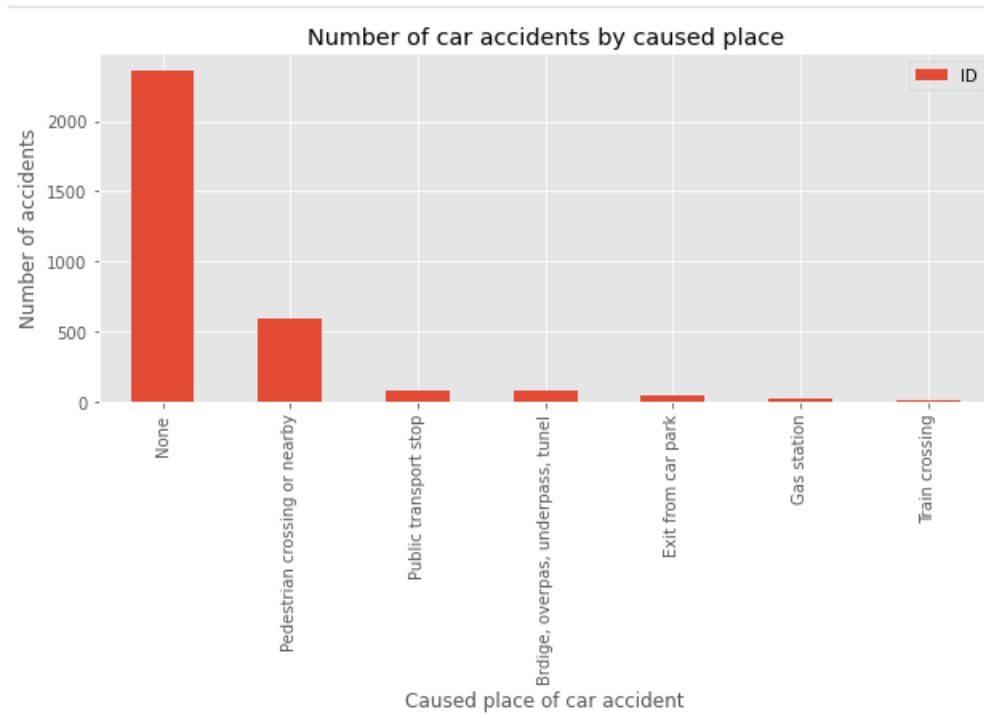
In the map is easy demonstrated the value of higher risk accidents place.
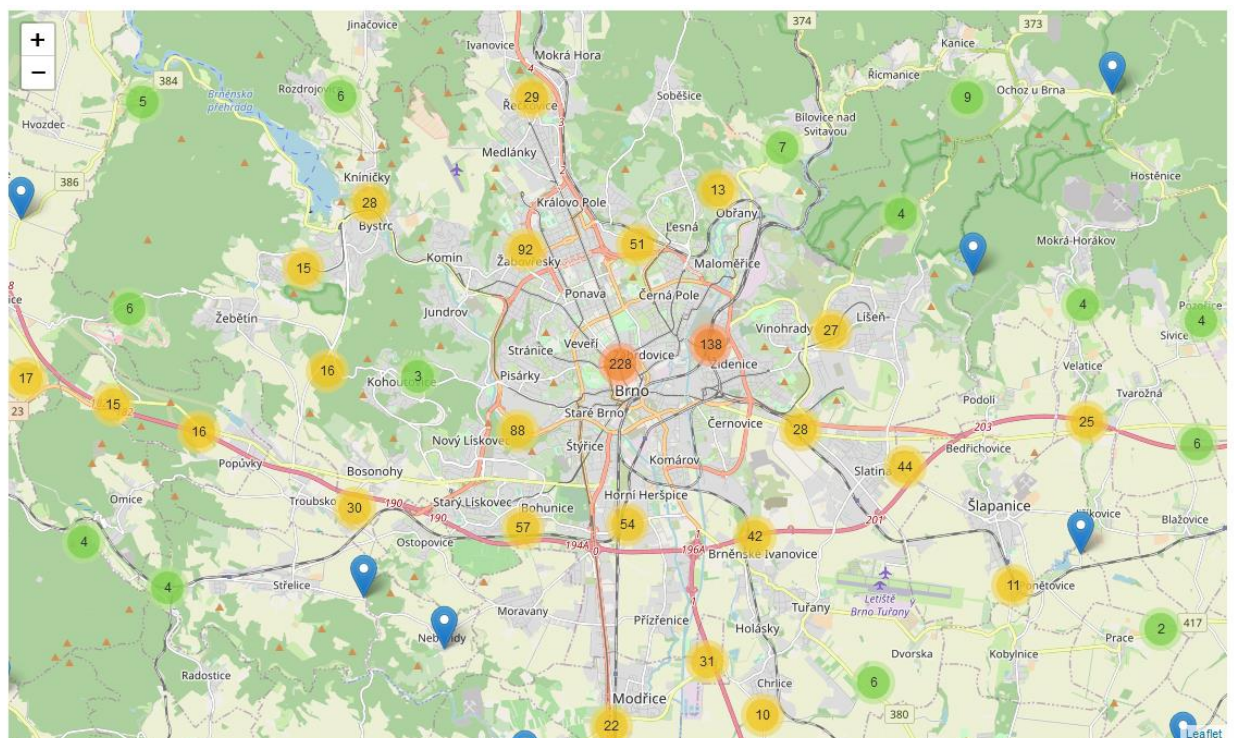
Discussion

Based on my skills I think I did maximum for this project. However I see the opportunity to make much bigger conclusion based on those data. If those data will used much in more details, it could possibly help to save lives, health or damage.

Conclusion

The risk of the car accident is mainly at the center of Brno city and highways. The lower risk/chance to have car accident is on Saturday and Sunday and in February and March.

In the next contribution on this project could possibly be the deep analyzing and prediction data for better understating and exact highlighted place of car accident (generally we can say be aware of pedestrian crossing and around, but would be much helpful to target on those one, which could possibly lead to crash or death.