Milestone

Julien DE VOS – Mathieu COWAN – Enki MILLET

Introduction :

Nowadays, the population is increasingly trying to make their travels easier, quicker, and more efficient. It is a daily occurrence for individuals in active life to move around, and this often involves the use of cars or other vehicles. People have various reasons for choosing to travel by car, but this also comes with the risk of accidents. Every year, there are numerous accidents, and they vary in terms of their causes, severity, the type of journey, the region, and information about the driver. As it is a concerning issue, we have decided to conduct a study on accidents in France. Specifically, through this project, we aim to address the following question: what environmental factors influence French road accidents and how do they impact them, and what can we do to diminish as much as we can the impact of those environmental factors, whether it is weather, road quality or even visibility. To answer that question, we will discuss what our motivation was to work on this project, what objectives we have with this project, what is the best approach to use on the project and what we already have done.

Motivation (WHY):

The impetus for this project arises from a critical need to elevate road safety standards. Our endeavor is to meticulously explore the root causes of severe accidents, providing invaluable insights that can steer the development of targeted preventive measures. Recognizing high-risk locations for severe accidents becomes imperative for optimizing resource allocation and strategically implementing interventions. Moreover, the pursuit of predicting accident severity holds transformative potential, offering a paradigm shift in accident response strategies and bolstering the evolution of advanced driver assistance systems.

In the realm of vehicular accidents, a myriad of studies has been conducted. However, our focus extends beyond the general exploration of accidents to a specific inquiry into the influence of environmental factors on French road accidents. We seek to dissect the impact of various elements, such as weather conditions, road conditions, and visibility, on accident severity.

To underline the significance of our inquiry, we refer to two articles (Link1, Link 2) providing statistics on the percentage of severe accidents attributed to adverse weather conditions, and another article who give statistics about specific roads impact on accident gravity. These articles not only reinforce the relevance of our investigation but also underscore the real-world consequences of environmental factors on road safety.

An illustrative anecdote further emphasizes the gravity of the issue. A friend of mine had who attended a car meeting, had this anecdote about people who tragically lost their lives on the highway due to aquaplaning caused by rain. Despite the efforts of well-intentioned individuals who stopped to assist, a subsequent accident occurred, claiming additional lives. This poignant story underscores the urgency of understanding and mitigating the impact of environmental factors on road safety.

To substantiate our research, we will leverage the dataset available at Kaggle - Accidents in France (2005-2016). This dataset encompasses a comprehensive record of accidents and associated variables, providing a robust foundation for our exploration.

References:

[Link 1 - SOUS LA PLUIE, SUR LES ROUTES: DEUX FOIS PLUS DE RISQUE D'ACCIDENTS]

https://www.awsr.be/wp-content/uploads/2021/03/communique-de-presse-pluie-et-aquaplaning.pdf

[Link 2 - L'état des routes, cet ennemi de la sécurité routière qui coûte de plus en plus cher]

https://www.caradisiac.com/l-etat-des-routes-cet-ennemi-de-la-securite-routiere-qui-coute-de-plus-en-plus-cher-200183.htm

Dataset Reference: Kaggle - Accidents in France (2005-2016)

This comprehensive approach, addressing the WHY, WHAT, and HOW, positions our project as a significant endeavor toward unraveling the complexities of road safety and contributing meaningfully to the discourse on accident prevention and severity mitigation.

Objectives (WHAT):

Concerning the causes of severe accidents, we have as objective to identify key factors which contribute to severe accidents. In a general standpoint, we can affirm that we want to principally analyze factors which are in range of institutional action and prevention. Indeed, it is easier to put measures in place against this type of factors whether than factors which are unitarily linked to human behavior which can be difficult to control in some cases.

We want to analyze historical accident to uncover patterns and correlations. It is important to us to investigate the potential role of variables/elements such as weather conditions, also the luminosity which is linked to the time of the day as well as the type of road in severe accidents, etc. Also concerning accident hotspots, we would want to determine geographic areas with a high incidence of severe accidents. We can then provide actionable insights for targeted safety interventions in identified hotspots.

We want to establish a comparison between the potential different correlations between the factors and the accidents so that we can analyze which elements are the most needed to be treated (for example structures to prevent accidents linked to rain, prevention for safety mainly linked to seatbelts).

Hypothesis we may issue is than in addition to what seems trivial factors such as the non-respect of security rules linked to assets such as seat belt, harsh weather, especially rain should be a prominent factor concerning severity of accidents.

We will separate our work in three parts, first the causes of the severity of the accidents, second the study of the locations where the most severe accidents happen, and finally the prediction of the severity of the accident.

Finally, in the aspect of predicting Accident Gravity, a good objective would be to assess the feasibility of predicting accident severity using machine learning models.

Investigate the potential of developing a predictive model to classify accidents into severity levels.

Approach (HOW):

Introduction: Data plot

- Recuperation of the data
- Data sorting
- Data plot
- Study of the number of accidents per year

1st part: What causes the severity of the accidents?

Through regressions we can study:

- Correlation between the weather and the severity of the accident
- Correlation between the luminosity and the severity of the accident
- Correlation between the age of the driver and the severity of the accident
- Correlation between the sex of the conductor and the gravity of the accident

2nd part: Where do the accidents happen the more often (with high severity) ?

- Correlation between the road surface and the severity of the accident
- Cauterization concerning the geography (the ideal would be to do multiple maps to understand properly)

3rd part: Can we predict the severity of an accident?

- Utilization of model whose goal is to determine the severity of the accident with the data we've kept. The training will be done on a part of our data. Then we will test it on the other part to determine its efficiency (Classification and possibly Decision Tree)
- We could even compare multiple models (as done in exercise 3) to compare the efficiency on our whole data.

In conclusion, this methodological framework is strategically crafted to foster a comprehensive comprehension of severe road accidents, intricately navigating through their causative factors, spatial distribution, and delving into the prospects of predictive modeling for accident severity. The interdisciplinary synergy of statistical analysis, geospatial tools, and machine learning techniques is harnessed to yield actionable insights, constituting a significant stride toward enhancing road safety.

First results

Before using the data and doing our first analysis, we have imported it in the code and filtered it. As we have different csv files containing the data, the first thing to do was to merge them together. After that, we filtered it by selecting only the columns we are interested in. In this way, here is the table we obtained with the 12 features that we kept:

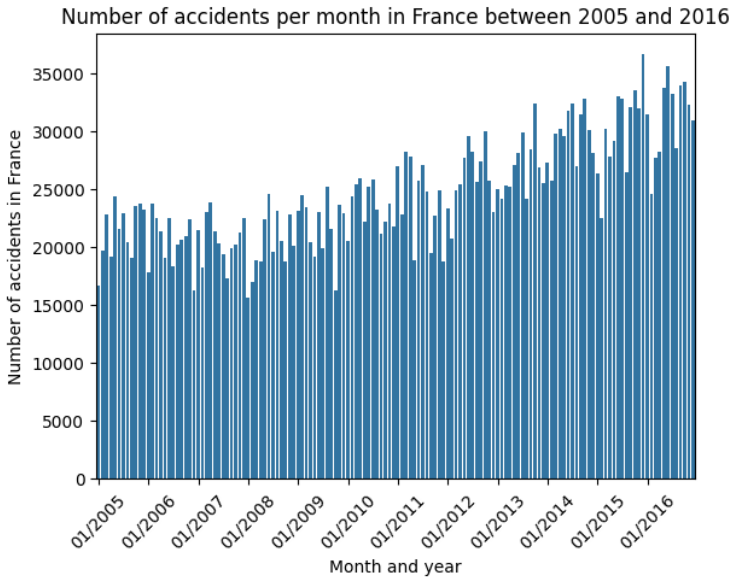
	Year	Month	User category	Severity	Sex	Year of birth	Trip purpose	Securiy	Luminosity	Weather	Type of road	Road surface
0	16	2	1	1	2	1983.0	0.0	11.0	1	8.0	3.0	1.0
1	16	2	1	1	2	1983.0	0.0	11.0	1	8.0	3.0	1.0
2	16	2	1	3	1	2001.0	9.0	21.0	1	8.0	3.0	1.0
3	16	2	1	3	1	2001.0	9.0	21.0	1	8.0	3.0	1.0
4	16	3	1	3	1	1960.0	5.0	11.0	1	1.0	3.0	1.0
...	...	...	...	...	...	...	...	...	...	...	...	...
3553971	5	12	1	4	1	1990.0	5.0	23.0	1	2.0	4.0	1.0
3553972	5	12	1	4	1	1990.0	5.0	23.0	1	2.0	4.0	1.0
3553973	5	12	1	4	1	1990.0	5.0	23.0	1	2.0	4.0	1.0
3553974	5	12	1	4	1	1951.0	0.0	13.0	5	2.0	2.0	2.0
3553975	5	12	2	4	2	1946.0	0.0	13.0	5	2.0	2.0	2.0

Now that the data is ready to be used, we can start to work with it.

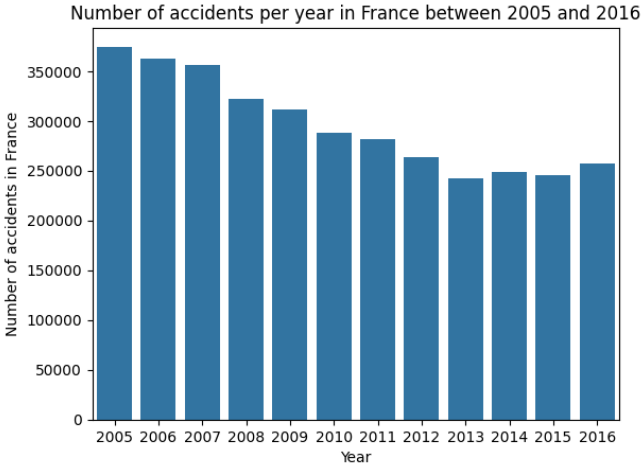
As an introductory work, we decided to have a look at the data. Therefore, we plotted the number of accidents that occurred in France along time.

First, we plot the number of accidents that occurred each month from January 2005 to December 2016:

We can observe that the number of accidents in France tend to increase along time.

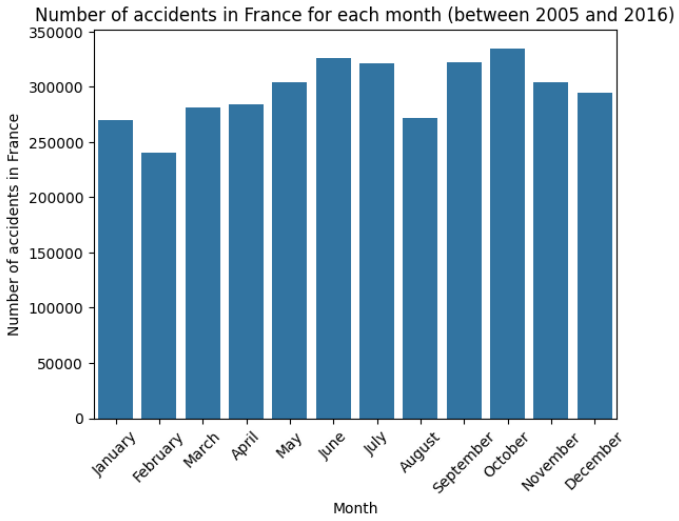


To verify this hypothesis, let's plot the number of accidents through years:



Paradoxically**,** we observe that the number of accidents per year decreases: from above 350000 in 2005 to approximately 250000 in 2016. If we look again at the first plot, we can observe that during the first years, the number of accidents through months is more regular than in the last years (where there are a lot of peaks). Then, maybe this contradiction is because in 2016 there are months during which the number of accidents is way bigger than during the others, so we have the impression that in the first plot the numbers of accidents increases whereas in reality it decreases through years.

Let's now look at the number of accidents for each month:



We could have thought about several hypothesis such as the fact that there will be a lot of accidents during the summer holidays (July and August) because more people are using the car to go in vacation, or the fact that in winter there are much more cars on the roads so it leads to more accidents, but we can see that these hypotheses are false. We observe that the peaks of accidents are reached in June, July, and October while the lowest numbers of accidents are reached in February and August.

In the next parts of this project, we will try to explain more precisely all these values and try to see if natural condition (like weather, road surface, luminosity, etc.) have an influence on them.

GitHub repository

https://github.com/JulienML/AccidentsInFrance/tree/main