ROAD SAFETY AND SEVERITY ANALYSIS

IN GREAT BRITAIN

INTRODUCTION

Backgrounds

According to the Department for Transport of Great Britain, there were 1.671 reported road deaths in 2018, 23.165 serious injuries in road traffic accidents reported to the police and 97.799 slight injuries. All this adds up to 122.635 reported road accidents, this is 6% lower than in 2017 and is the lowest level on record. There is no single underlying factor that drives road accidents. Instead, there are a number of influences. These include:

• The distance people travel.

• The mix of transport modes used.

• Behavior of drivers, riders and pedestrians.

• The mix of groups of people using the road (e.g. changes in the number of newly qualified or older drivers)

• External effects such as the weather, which can influence behavior (for instance, encouraging / discouraging travel, or closing roads) or change in the risk on roads (by making the road surface more slippery).

The patterns involved in dangerous crashes and road accidents can be helpful in developing road safety policies or greater awareness of the driver himself. The study of these patterns, causes and severity of injuries can predict the severity or probability of an accident, which would be very useful for the well-being of any driver or pedestrian

Business Understanding

The analysis of these accidents, causes and severity of injuries can be used to predict the severity or probability of an accident. Therefore, it would be very useful for the well-being of any driver or pedestrian, to offer the possibility that, knowing the risks, the probability of an accident, the most dangerous routes and the less favorable weather conditions, among many other factors, people drive carefully or even change their trip.

<u>Interest</u>

The main objective is to identify the severity or probability of a possible accident, so this work can be useful for anyone interested in assessing risks while driving, the traffic control departments or simply for anyone who drives a vehicle, offering the possibility that, knowing the risks and probabilities of an accident, people drive with more care. or even change their trip. Undoubtedly, this would help all drivers to have a better perspective of possible accidents, the most dangerous routes, the less favorable weather conditions and with all these data, to make a more informed decision.

DATA ACQUISITION

<u>Data understanding</u>

This project is developed over the data provided by the UK open data website, section road safety data [(https://data.gov.uk/dataset/cb7ae6f0-4be6-4935-9277- 47e5ce24a11f / road-safety-data / datafile / 36f1658e-b709-47e7-9f56-cca7aefeb8fe / preview)](https://data.gov.uk/dataset/cb7ae6f0-4be6-4935-9277-47e5ce24a11f) for year 2018. This repository has the benefit of having a well-formatted and structured data set, however it presents labels with missing data encoded as -1, which are not easily identifiable when searching for null or blank records. The set presents a good number of variables in order to build a model, some more predictive than others and although it presents useful geospatial data to produce visualizations, they are not useful for the present project.

<u>Data Preparation</u>

In order to generate more effective predictions, it is necessary to review the quality, balance and integrity of the dataset, avoiding skewed metrics. In terms of balance, particularly in the **Accident severity** metric, which will be our target variable, the dataset encodes three categories of severity:

Accident Severity	Code
Fatal	1
Serious	2
Slight	3

For the analysis and creation of a predictive model and because the dataset is labeled, I am going to use supervised machine learning. Once identified the data will be cleaned up, after the information will be analyzed to determine which of the attributes have the highest correlation with the target variable, among which are probably: Light Conditions, Weather Conditions, Road Surface Conditions, Junction Detail, and some others.

The variables of the dataset will be group into three main sets:

Accident details

1. Accident Index
2. Police Force
3. Accident Severity
4. Number of Vehicles
5. Number of Casualties
6. Did Police Officer Attend Scene of Accident
7. 1st Road Class
8. 1st Road Number
9. Road Type
10. Speed limit
11. Junction Detail
12. Junction Control
13. 2nd Road Class
14. 2nd Road Number
15. Pedestrian Crossing-Human Control
16. Pedestrian Crossing-Physical Facilities

Location and time

1. Location Easting OSGR (Null if not known)
2. Location Northing OSGR (Null if not known)
3. Longitude (Null if not known)
4. Latitude (Null if not known)
5. Date (DD/MM/YYYY)
6. Day of Week
7. Time (HH:MM)
8. Local Authority (District)
9. Local Authority (Highway Authority - ONS code)
10. Urban or Rural Area

Environmental issues

1. Carriageway Hazards
2. Light Conditions

3. Weather Conditions
4. Road Surface Conditions
5. Special Conditions at Site

With all the data above, it is possible to determine interesting situations and make questions about them, like:

1. When do most accidents happen? On weekdays or the weekend?

2. Is weather a key factor in road accidents?

3. Is the number of people affected related to the severity of the accident?

4. Which causes more accidents?