<div align="center">**Dataset Documentation**</div>

Dataset Overview

This dataset, which was obtained from the official website of the UK government, contains road safety statistics for 2019. The Department for Transport (DfT) collects and maintains the information, which is divided into three separate datasets: accidents, casualties, and vehicles. These datasets are connected using the variable "Accident Index." The records provide comprehensive information about auto collisions that resulted in injuries and were reported to the police within 30 days of the collision.

Dataset: Accidents, Casualties, and Vehicles in 2019 for UK Road Safety. This dataset includes thorough data on UK traffic collisions, with a focus on collisions that caused injuries to people and were reported to the police within 30 days. The dataset, which spans the year 2019, is made up of three distinct datasets: accidents, casualties, and vehicles.

Data Sources and Data Collection Procedure:

The Road Safety Data dataset was obtained from the UK Government's official website. The Department for Transport (DfT) oversees gathering and maintaining the data. The following link contains additional details about the data:

https://data.gov.uk/dataset/cb7ae6f0-4be6-4935-9277-47e5ce24a11f/road-safety-data

The following three datasets were used in this study:

a. Road Safety Data - Accidents 2019
b. Road Safety Data - Casualties 2019
c. Road Safety Data - Vehicles 2019
The variable "Accident_Index" is used to link these datasets together.

Dataset Descriptions:
a. Road Safety Data - Accidents 2019
- The Accidents dataset contains 31 variables and 117,536 data points.
- Every data point corresponds to a distinct instance of an accident that was reported in the UK in 2019.

- The police gather information either at the scene of the accident or when it is reported to them at a police station by a member of the public.

b. Road Safety Data - Casualties 2019

- The Casualties dataset contains 16 variables and 153,158 data points.
- Data points are linked to the Accidents dataset using the 'Accident_Index' variable.
- The number of data points is higher than in the accident dataset because more than one casualty may be involved in an accident.

c. Road Safety Data - Vehicles 2019

- The Vehicles dataset contains 23 variables and 216,381 data points.
- Data points are linked to the Accidents dataset using the 'Accident_Index' variable.
- The number of data points represents the total number of vehicles involved in accidents in 2019.

Variable Descriptions:

A variable lookup table is provided to assist with understanding the variables and fields collected in each dataset. This table helps define and explain the variables used across the Accidents, Casualties, and Vehicles datasets.

Age of the driver: Younger and less experienced drivers, such as teenagers, may be more likely to be in collisions due to their inexperience behind the wheel and higher risk-taking tendencies. A decrease in cognitive and physical capacities may also provide difficulties for elderly drivers, raising the danger of collisions.

Light condition: Poor visibility in low-light situations, including at night or in foggy weather, can raise the risk of accidents. Drivers find it more difficult to identify risks and respond quickly when vision is poor, which increases the severity of accidents.

Weather conditions: Unfavorable weather conditions, such as rain, snow, ice, or fog, can make roads slick and cut down on visibility, which can lead to an increase in the frequency and severity of accidents. It is more difficult for drivers in these situations to keep control of their cars and react to unexpected changes in the road.

Week Day Distribution: There can be a wide range in how accidents are distributed throughout the week. For instance, weekends or particular days of the week may have greater

accident rates because of more traffic, more drinking, or tired drivers. On certain days, there may be greater traffic and other circumstances, which can lead to more serious accidents.

Number of Vehicles: As compared to accidents involving only one or two vehicles, collisions involving more than two vehicles are more likely to result in significant damage, serious injuries, or fatalities. Chain reactions, larger impact forces, and more complicated dynamics can occur in multi-vehicle collisions, increasing the severity of the crash.

Speed Limit: Accident severity typically rises at higher speeds. Drivers who drive faster than the speed limit need more space to stop their cars and have less time to react to sudden situations. Accidents that happen at faster speeds are therefore more likely to result in serious injuries or fatalities.

Age Band: This variable most likely alludes to various age brackets or ranges of drivers. Age-related differences in risk-taking behaviour, driving ability, and total experience may have an impact on how serious accidents are. Drivers who are younger or older, for example, may experience accidents with greater severity than drivers who are in their middle years.

Road Surface: The state of the road, such as icy, wet, or uneven surfaces, can impact a vehicle's stability and traction. Unsafe driving practices such as skidding, losing control, or taking longer to stop might result in worse accidents.

Data Cleaning and Pre-processing:

The dataset provided by the UK government is a raw dataset. Users are advised to perform data cleaning and pre-processing steps based on their specific research questions and requirements, such as handling missing values, data transformation, or outlier detection and treatment.

Data Quality and Limitations:

The dataset is regarded as trustworthy because it came from the official website of the UK Government and is kept up to date by the Department for Transportation. (DfT). The dataset only includes accidents reported to the police within 30 days of the incident, thus users should be aware that there may be underreporting of road traffic accidents.

Data Access and Distribution:

The dataset can be accessed and downloaded from the UK Government's official website: https://data.gov.uk/dataset/cb7ae6f0-4be6-4935-9277-47e5ce24a11f/road-safety-data

Data Citation:

Users should cite the dataset as follows:

UK Department for Transport. (2019). Road Safety Data - Accidents, Casualties, and Vehicles 2019 [Data set]. Retrieved from https://data.gov.uk/dataset/cb7ae6f0-4be6-4935-9277-47e5ce24a11f/road-safety-data

Versioning and Updates:

The dataset provided covers the year 2019. Users should check the UK Government's official website for updates or newer versions of the dataset, if available.

Contact Information:

For any questions or assistance with the dataset, users can contact the Department for Transport (DfT) at the following email address: roadacc.stats@dft.gov.uk

The dataset can be used to answer the following research questions:

Research question 1: How does driver age relate to the severity of road traffic accidents in the UK, with a focus on Casualty Status 19? Are certain age groups more likely to be involved in severe accidents?

Answer: The dataset contains information on the age of the driver involved in each accident, as well as the severity of the accident. This information can be used to investigate the relationship between driver age and accident severity. For example, researchers could use the dataset to determine whether younger or older drivers are more likely to be involved in severe accidents.

Research question 2: Can seasonal trends be observed in the rate of road traffic accidents in the UK, and how do these trends Answer: differ for accidents of varying severity? When are more severe accidents more likely to occur?

Answer: The dataset includes information about each accident's date. Investigating seasonal patterns in the frequency of traffic accidents may be done using this data. The dataset, for instance, might be used by academics to find out whether accidents are more common in the

winter than in summer. They might also look at if there are seasons when accidents are more likely to be more serious.

*Research questions 3:*What is the impact of various conditions, such as lighting, weather, road quality, and road size, on the number of casualties in road traffic accidents in the UK, with a focus on Casualty Status 19?

Answer: The collection includes data on a range of factors, including illumination, weather, road quality, and road size, that may have an impact on how serious an accident is. Investigating the effect of these circumstances on the number of fatalities in traffic accidents may be done using this information. The dataset, for instance, might be used by academics to establish if poor weather or low light conditions increase the likelihood of serious accidents.

These are only some of the research problems that the dataset may be utilised to resolve. For any researchers who want to learn about the cause and effects of traffic accidents, the dataset is a great resource.

Stats-19:

The STATS19 database contains information on all traffic collisions that resulted in injuries and were reported to the police within 30 days of the collision. The information is obtained by the police at the accident site or when it is reported to them by a member of the public in a police station. The Standing Committee for Road Accident Statistics (SCRAS) and the Association of Chief Police Officers have agreed to use the Department of Transportation's stated variables and fields. (ACPO). Direct data delivery occurs to the DfT or the relevant local authorities (or groups of local authorities). Before providing data to DfT, local authorities verify it.

The dataset can be used to develop data-supported policies and actions that will increase road safety. Using this dataset and rigorous statistical analysis, researchers can gain a better understanding of the multivariate factors that affect the severity of traffic accidents. Additionally, they can aid in the creation of practical plans to lessen the frequency and severity of traffic accidents, ultimately enhancing public safety and well-being.