**EDA PROJECT**

**ROAD ACCIDENT DATA ANALYSIS  
  
GROUP MEMBERS:   
 SIJAL FATIMA KAINAT IJAZ**

**L1F21BSDS0030 L1F21BSDS0038**

**1. Problem Statement:**

Road accidents are a significant threat to public safety and result in human and economic losses. This analysis aims to identify the key factors contributing to road accidents in the United Kingdom to understand high-risk areas, times, and conditions. The goal is to provide actionable recommendations to reduce the occurrence and severity of road accidents and improve road safety. This includes understanding trends over time, the impact of different environmental and road conditions, and the types of vehicles most frequently involved.

**2. Data Description:**

**2.1 Data Source:**

The data for this analysis is from a road accident dataset for the United Kingdom.

**2.2 Data Cleaning Process:**

* **Missing Values:** Handling missing values in the variables. The missing data is imputed by mean or mode.
* **Consistency Checks:** Ensuring consistency in categorical data and standardizing them.

**2.3 Features:**

* accident\_severity (categorical): Severity of the accident (Slight, Serious, Fatal).
* day\_of\_week (categorical): Day of the week the accident occurred.
* vehicle\_type (categorical): Type of vehicle involved in the accident.
* road\_type (categorical): Type of road where the accident occurred (e.g., Single carriageway, Dual carriageway).
* weather\_conditions (categorical): Weather conditions at the time of the accident.
* urban\_or\_rural\_area (categorical): Location of the accident (Urban, Rural).
* number\_of\_casualties (numerical): Number of casualties in the accident.
* speed\_limit (numerical): Speed limit at the location of the accident (km/h).
* time (time): Time of the accident.
* junction\_detail (categorical): Details about the junction where the accident occurred.
* road\_surface\_condition (categorical): Condition of the road surface (Wet, Dry, Snow/Ice).

**2.4 Data Volume:**

The dataset contains 307,973 accident records.

**3. Analysis:**

**3.1 Python Analysis:**

Python with libraries like pandas, matplotlib, and seaborn is used for data loading, cleaning, visualization and modeling. Some of the analyses are given below:

* Calculating the number and percentage of accidents by severity, day of the week, vehicle type, road type, weather conditions, and urban/rural area.
* Analyzing the average number of casualties per accident severity.
* Examining accident frequency by hour of the day.
* Creating bar plots, line plots, and other visualizations to explore relationships between variables.

**3.2 SQL Analysis:**

SQL is used for data extraction and aggregation. The queries give the calculation of accident counts by various categories such as severity, day of week, vehicle type etc. and average casualties by severity.

**Table creation and data loading code:**

CREATE TABLE accidents (

accident\_index VARCHAR(50) PRIMARY KEY,

accident\_date DATE,

day\_of\_week VARCHAR(200),

junction\_control VARCHAR(50),

junction\_detail VARCHAR(50),

accident\_severity VARCHAR(200),

light\_conditions VARCHAR(200),

local\_authority VARCHAR(50),

carriageway\_hazards VARCHAR(50),

number\_of\_casualties INT,

number\_of\_vehicles INT,

police\_force VARCHAR(50),

road\_surface\_conditions VARCHAR(200),

road\_type VARCHAR(50),

speed\_limit INT,

time TIME,

urban\_or\_rural\_area VARCHAR(20),

weather\_conditions VARCHAR(50),

vehicle\_type VARCHAR(50)

);

SET datestyle = 'DMY'; -- Set to Day-Month-Year

COPY accidents (accident\_index, accident\_date, day\_of\_week, junction\_control, junction\_detail,

accident\_severity, light\_conditions, local\_authority, carriageway\_hazards,

number\_of\_casualties, number\_of\_vehicles, police\_force, road\_surface\_conditions,

road\_type, speed\_limit, time, urban\_or\_rural\_area, weather\_conditions, vehicle\_type)

FROM 'C:/road\_accident.csv'

DELIMITER ','

CSV HEADER;

**3.3 Excel Analysis:**

Excel is used to create an interactive dashboard summarizing key accident metrics. The dashboard includes:

* KPIs such as total casualties, fatal casualties (with percentage), serious casualties (with percentage), slight casualties (with percentage), and casualties involving cars (with count and percentage).
* Casualties by vehicle type, road type, light condition, and road surface condition.
* Monthly trend comparison of casualties between 2021 and 2022.
* Casualties by urban/rural area.
* Interactive filtering by accident date (year) and urban/rural area.

**3.4 Power BI Analysis:**

Power BI is also used to develop a dashboard with similar information to the Excel dashboard, but with additional features:

* Geographical visualization by map of the UK showing accident locations.
* Direct comparison of current year and previous year metrics.
* Interactive elements such as slicers and filters to explore the data.

**4. Modeling:**

Random Forest is used for prediction. Accuracy achieved by it is: 85.46%

**5. Insights & Conclusions:**

* Slight accidents are the most frequent, followed by serious and then fatal accidents.
* Fridays and Tuesdays show the highest number of accidents.
* Cars are the most common vehicle type involved in accidents.
* Single carriageways are associated with the highest number of accidents.
* "Fine no high winds" is the most frequently reported weather condition during accidents, but adverse weather (rain, snow, fog) also contributes significantly.
* Urban areas have a higher number of accidents overall, but rural roads may have higher severity due to factors like speed.
* Accidents are most frequent during the afternoon and early evening (3 p.m to 6 p.m), likely coinciding with rush hour.
* Fatal accidents have the highest average number of casualties, followed by serious and then slight accidents.

**6. Recommendations:**

* Focus road safety campaigns and interventions on high-risk days (Fridays, Tuesdays), times (afternoon/early evening), and locations (single carriageways, urban areas).
* Promote vehicle safety awareness, particularly for car drivers, given their high involvement in accidents.
* Issue timely warnings and advisories during adverse weather conditions (rain, snow, fog) and encourage drivers to adjust their driving accordingly.
* Implement stricter speed enforcement measures, especially on single carriageways and in areas with a history of high-speed accidents.
* Improve the safety of junctions, particularly T or staggered junctions and crossroads, through better signage, lighting, or redesign.