

EXPLORATORY DATA ANALYSIS (EDA) ON ROAD TRAFFIC ACCIDENTS IN GREAT BRITAIN (2020)

Introduction

- **Objective:** Analyze road traffic accident data from 2020 to provide insights into accident patterns and help advise government agencies on improving road safety.
- **Data Source:** The dataset is provided in a SQLite database `accident_data_v1.0.0_2023.db`. **Download dataset at <https://rb.gy/r19j12>**
- **Data Context:** The data includes detailed records of road traffic accidents in Great Britain for the year 2020.
- **Project timeline:** **24th June to 13th July 2024**

Section 1 – 24th to 25th June

2. Data Extraction and Loading

- **Task:** Extract the relevant data for the year 2020 from the SQLite database.
- **Tools:** Use pandas and SQLite libraries in Python.
- **Deliverables:** A pandas DataFrame containing the data for 2020.

3. Data Cleaning and Preprocessing

- **Task:** Ensure the data is clean and ready for analysis.
 - Handle missing values.
 - Convert data types where necessary (e.g., datetime columns).
 - Ensure consistency in categorical variables.
- **Deliverables:** A cleaned DataFrame ready for analysis.

4. Data Overview and Summary Statistics

- **Task:** Provide a summary of the data.

- Number of accidents, vehicles, and casualties.
- Basic statistics for numerical columns (mean, median, standard deviation).
- Distribution of categorical variables (e.g., accident severity, road type).
- **Deliverables:** Summary statistics and initial observations.

Section 2 – 26th to 29th June

5. Temporal Analysis

- **Task:** Analyze the temporal aspects of accidents.
 - Accidents by hour of the day.
 - Accidents by day of the week.
 - Monthly trends.
 - Analysis of seasonal patterns.
 - Analysis of public holidays vs. regular days.
- **Deliverables:** Visualizations (e.g., line plots, bar charts) and insights on temporal patterns.

6. Spatial Analysis

- **Task:** Analyze the geographical distribution of accidents.
 - Distribution of accidents based on latitude and longitude.
 - Accidents by local authority district and highway.
 - Analysis of urban vs. rural areas.
 - Heatmaps of accident hotspots.
 - Geospatial clustering of accidents.
- **Deliverables:** Geographical visualizations (e.g., heat maps, choropleth maps) and insights on spatial patterns.

Section 3 – 30th June to 7th July

7. Analysis of Accident Severity

- **Task:** Investigate factors contributing to accident severity.
 - Severity distribution by road type, speed limit, weather conditions, and light conditions.
 - Analysis of junction details and controls.
 - Impact of road surface conditions and special site conditions.
 - Role of carriageway hazards.
- **Deliverables:** Visualizations and insights on factors affecting accident severity.

8. Vehicle and Casualty Analysis

- **Task:** Analyze vehicle and casualty data.
 - Distribution of vehicle types involved in accidents.
 - Analysis of vehicle maneuvers and directions.
 - Casualty demographics (age, sex) and severity.
 - Analysis of passenger types (car, bus/coach).
 - Impact of vehicle factors (e.g., skidding, overturning, point of impact).
- **Deliverables:** Visualizations and insights on vehicle and casualty patterns.

9. Specific Analysis: Motorbike Accidents

- **Task:** Focus on motorbike accidents.
 - Significant hours of the day and days of the week for different motorbike categories (125cc and under, over 125cc up to 500cc, and over 500cc).
 - Analysis of accident severity and conditions for motorbikes.
 - Impact of light and weather conditions on motorbike accidents.
- **Deliverables:** Detailed analysis and visualizations for motorbike accidents.

10. Specific Analysis: Pedestrian Accidents

- **Task:** Focus on pedestrian accidents.
 - Significant hours of the day and days of the week for pedestrian involvement.
 - Analysis of pedestrian location and movement during accidents.
 - Impact of pedestrian crossing facilities and human control.
 - Analysis of pedestrian road maintenance worker involvement.
- **Deliverables:** Detailed analysis and visualizations for pedestrian accidents.

Section 4 – 8th to 10th July

11. Analysis of External Factors

- **Task:** Investigate external factors affecting accidents.
 - Weather conditions and their impact on accidents.
 - Light conditions during accidents (daylight, darkness, street lights).
 - Impact of police presence at accident scenes.
 - Analysis of trunk road involvement.
- **Deliverables:** Visualizations and insights on external factors affecting accidents.

12. Socioeconomic Analysis

- **Task:** Analyze socioeconomic factors related to accidents.
 - Analysis of accidents by casualty home area type and IMD decile.
 - Analysis of driver home area type and IMD decile.
 - Correlation between socioeconomic factors and accident severity.
- **Deliverables:** Visualizations and insights on socioeconomic factors.

13. Driver and Vehicle Analysis

- **Task:** Analyze driver and vehicle characteristics.
 - Distribution of driver ages and genders.
 - Analysis of engine capacity and vehicle age.
 - Impact of driver journey purpose on accidents.
 - Analysis of vehicle towing and articulation.
- **Deliverables:** Visualizations and insights on driver and vehicle characteristics.

Section 5 – 11th to 13th July

14. Insights and Recommendations

- **Task:** Summarize key findings and provide actionable recommendations.
 - Identify high-risk times and locations.
 - Suggest interventions to improve road safety.
 - Propose policy changes based on findings.
- **Deliverables:** A report summarizing insights and recommendations for government agencies.

Tools and Libraries

- Python
 - pandas
 - numpy
 - matplotlib
 - seaborn
 - folium (for geographical visualizations)
 - SQLite

NOTE: You are not restricted to these tools or libraries. You are at liberty to use whatsoever libraries to perform an extensive EDA as is required!

Submission

- **Code:** Well-documented Jupyter notebooks or Python scripts.
- **Report:** A detailed report in PDF or markdown format summarizing the findings and recommendations.
- **Presentation:** A PowerPoint presentation highlighting key insights and recommendations.