Project Report: Exploratory Data Analysis (EDA) on Accident Severity

Author: Deeksha Singh

Abstract:

This report presents an exploratory data analysis (EDA) of a dataset related to road accidents. The dataset contains information about various attributes associated with road accidents, including accident severity, location, weather conditions, vehicle details, and more. The primary objective of this analysis is to gain insights into the factors contributing to accident severity.

Introduction:

Road traffic accidents are a significant public safety concern worldwide. Understanding the factors that influence accident severity is crucial for developing effective preventive measures and improving road safety. This EDA aims to provide a comprehensive overview of the dataset and identify key patterns and trends.

Data Description:

The dataset used for this analysis consists of two main tables: "Accidents" and "Vehicles." The "Accidents" table contains information about individual accidents, including attributes such as accident severity, location, time, and weather conditions. The "Vehicles" table contains details about the vehicles involved in these accidents.
You can download the data from the Kaggle repository
:https://www.kaggle.com/datasets/tsiaras/uk-road-safety-accidents-and-vehicles

Data Preprocessing:

The following data preprocessing steps were performed before conducting the analysis:

1. Handling Missing Data: Columns with excessive missing data were dropped from the dataset, and records with missing values were removed.
2. Datetime Format: The date column was converted from an object data type to a datetime data type for time-related analysis.

3. Feature Engineering: The "Hour" and "Daytime" columns were created to categorize accidents by time of day, allowing for better analysis.

Exploratory Data Analysis

- Descriptive Statistics

1. Numerical Attributes
   We started by conducting a descriptive analysis of the numerical attributes:

   Number of Casualties: The mean number of casualties per accident is approximately 1.24, with a maximum of 27 casualties in a single accident.

   Number of Vehicles: The mean number of vehicles involved in an accident is approximately 1.67, with a maximum of 20 vehicles.

2. Categorical Attributes
   We also analyzed categorical attributes, including "Accident Severity," "Road Type," and "Weather Conditions." Here are some key insights:

   Accident Severity: Most accidents in the dataset are categorized as "Slight" (approximately 87.9%), followed by "Serious" (approximately 11.1%) and "Fatal" (approximately 0.88%).

   Road Type: The majority of accidents occur on "Single carriageways" (approximately 54.2%), followed by "Dual carriageways" (approximately 15.9%) and "Roundabouts" (approximately 3.7%).

   Weather Conditions: "Fine no high winds" is the most common weather condition during accidents (approximately 56.2%), followed by "Unknown" conditions (approximately 26.4%) and "Raining no high winds" (approximately 10.1%).

3. Correlation Analysis
   We conducted correlation analysis to identify relationships between numeric attributes. Key findings from the correlation analysis include:

   A moderate positive correlation between the number of casualties and the number of vehicles involved in accidents (correlation coefficient ≈ 0.31).

Weak correlations between age-related vehicle attributes (e.g., "Age of Vehicle" and "Age Band of Driver") and other attributes.

4. Time Analysis
   We analyzed accidents based on the time of day, categorizing them into five groups:

   "Early Morning," "Office Hours," "Afternoon Rush," "Evening," and "Night." The analysis revealed that most accidents occur during "Office Hours" (between 10 and 15), followed by the "Afternoon Rush" (between 15 and 19).

Conclusion:

This EDA provided valuable insights into road traffic accidents and their severity. Key takeaways include the predominant occurrence of "Slight" accidents, the influence of factors like road type and weather conditions, and the importance of time-related analysis. These findings can serve as a foundation for further in-depth analysis and the development of safety measures to reduce accident severity on the roads.

Future Directions:

Further analysis of the relationship between specific attributes and accident severity.
Advanced predictive modeling to identify factors contributing to severe accidents.
Geospatial analysis to identify accident hotspots for targeted interventions.
Temporal analysis to detect trends and seasonality in accidents.
This EDA lays the groundwork for more comprehensive research on road safety and accident prevention.