Road Accident Database Analysis

By: Mohammad Ali Asadi

1. Introduction:

Road accidents represent a serious global concern, resulting in loss of life, injuries, and economic repercussions. Analyzing road accident data is vital for understanding the underlying factors and devising effective prevention strategies. In this report, we delve into a road accident database to uncover patterns, trends, and potential interventions aimed at reducing accidents and enhancing road safety.

2. Literature on Traffic Accidents:

Prior research on traffic accidents has identified various factors contributing to their occurrence, including human factors (e.g., driver behavior, age, gender), vehicle-related factors (e.g., vehicle type, maintenance), environmental factors (e.g., weather conditions, road infrastructure), and socio-economic factors (e.g., area deprivation). Studies have also examined the severity of casualties resulting from accidents and the effectiveness of interventions such as road safety campaigns, infrastructure improvements, and policy measures.

3. Description of Analysis Methods:

We conducted exploratory data analysis (EDA) on the road accident database to gain insights into the characteristics of accidents and casualties. This involved data cleaning, visualization, and statistical analysis to understand the distribution of accidents over age, sex, geographical locations, and severity levels. We also employed machine learning algorithms to identify factors associated with accident severity and predict the likelihood of different outcomes.

The dataset analyzed in this report is related to traffic accidents in 2022, which is a famous dataset. This dataset includes the following items:

1. **Status**: The status of the accident (e.g., reported, under investigation).

2. **Accident_Index**: A unique identifier for each reported accident.
   Unique value for each accident. The accident_index combines the accident_year and accident_ref_no to form a unique ID. It can be used to join to Vehicle and Casualty

3. **Accident_Year**: The year in which the accident occurred.

4. **Accident_Reference**: A reference number associated with the accident.

5. **Vehicle_Reference**: A reference number for the involved vehicle in the accident.

6. **Casualty_Reference**: A reference number for the casualty involved in the accident.

7. **Casualty_Class**: Indicates the class of the casualty:

1	Driver or rider
2	Passenger
3	Pedestrian

8. **Sex_of_Casualty**: The gender of the casualty (male or female).

1	Male
2	Female
9	unknown (self reported)
-1	Data missing or out of range

9. **Age_of_Casualty**: The age of the casualty. (-1 Data missing or out of range)
10. **Age_Band_of_Casualty**: Age group to which the casualty belongs (e.g., 0-5, 6-10, 11-15).

1	0 - 5
2	6 - 10
3	11 - 15
4	16 - 20
5	21 - 25
6	26 - 35
7	36 - 45
8	46 - 55
9	56 - 65
10	66 - 75
11	Over 75
-1	Data missing or out of range

11. **Casualty_Severity**: The severity of the casualty's injuries:

1	Fatal
2	Serious
3	Slight

12. **Pedestrian_Location**: The location of the pedestrian at the time of the accident.

0	Not a Pedestrian
1	Crossing on pedestrian crossing facility
2	Crossing in zig-zag approach lines
3	Crossing in zig-zag exit lines
4	Crossing elsewhere within 50m. of pedestrian crossing
5	In carriageway, crossing elsewhere
6	On footway or verge
7	On refuge, central island or central reservation
8	In centre of carriageway - not on refuge, island or central reservation
9	In carriageway, not crossing
10	Unknown or other
-1	Data missing or out of range

13. **Pedestrian_Movement**: The movement of the pedestrian during the accident.

0	Not a Pedestrian
1	Crossing from driver's nearside
2	Crossing from nearside - masked by parked or stationary vehicle
3	Crossing from driver's offside
4	Crossing from offside - masked by parked or stationary vehicle
5	In carriageway, stationary - not crossing (standing or playing)
6	In carriageway, stationary - not crossing (standing or playing) - masked by parked or stationary vehicle
7	Walking along in carriageway, facing traffic
8	Walking along in carriageway, back to traffic
9	Unknown or other
-1	Data missing or out of range

14. **Car_Passenger**: Indicates whether the casualty was a car passenger at the time of the accident (yes/no).

0	Not car passenger
1	Front seat passenger
2	Rear seat passenger
9	unknown (self reported)
-1	Data missing or out of range

15. **Bus_or_Coach_Passenger**: Indicates whether the casualty was a bus or coach passenger (yes/no).

0	Not a bus or coach passenger
1	Boarding
2	Alighting
3	Standing passenger
4	Seated passenger
9	unknown (self reported)
-1	Data missing or out of range

16. **Pedestrian_Road_Maintenance_Worker**: Indicates whether the casualty was a road maintenance worker (yes/no).

0	No / Not applicable
1	Yes
2	Not Known
3	Probable
-1	Data missing or out of range

17. **Casualty_Type**: The type of casualty (e.g., driver/rider, passenger, pedestrian).

0	Pedestrian
1	Cyclist
2	Motorcycle 50cc and under rider or passenger
3	Motorcycle 125cc and under rider or passenger
4	Motorcycle over 125cc and up to 500cc rider or passenger
5	Motorcycle over 500cc rider or passenger
8	Taxi/Private hire car occupant
9	Car occupant
10	Minibus (8 - 16 passenger seats) occupant
11	Bus or coach occupant (17 or more pass seats)
16	Horse rider
17	Agricultural vehicle occupant
18	Tram occupant
19	Van / Goods vehicle (3.5 tonnes mgw or under) occupant
20	Goods vehicle (over 3.5t. and under 7.5t.) occupant
21	Goods vehicle (7.5 tonnes mgw and over) occupant
22	Mobility scooter rider
23	Electric motorcycle rider or passenger
90	Other vehicle occupant
97	Motorcycle - unknown cc rider or passenger
98	Goods vehicle (unknown weight) occupant
99	Unknown vehicle type (self rep only)
-1	Data missing or out of range

18. **Casualty_Home_Area_Type**: The type of area in which the casualty resides (e.g., urban, rural).

1	Urban area
2	Small town
3	Rural
-1	Data missing or out of range

19. **Casualty_IMD_Decile**: The IMD (Index of Multiple Deprivation) decile of the area where the casualty resides (a measure of deprivation).

1	Most deprived 10%

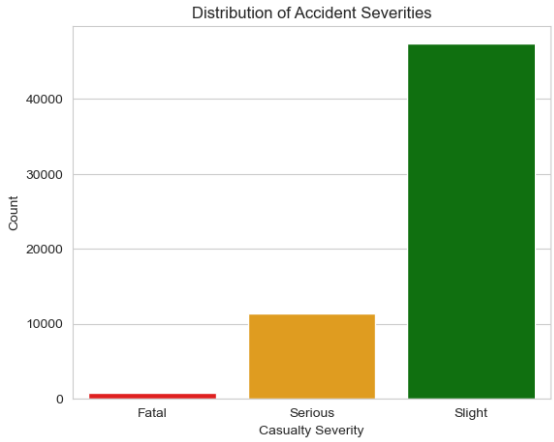
2	More deprived 10-20%
3	More deprived 20-30%
4	More deprived 30-40%
5	More deprived 40-50%
6	Less deprived 40-50%
7	Less deprived 30-40%
8	Less deprived 20-30%
9	Less deprived 10-20%
10	Least deprived 10%
-1	Data missing or out of range

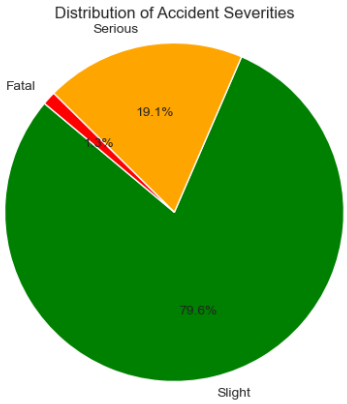
20. **LSOA_of_Casualty**: The Lower Layer Super Output Area (LSOA) associated with the casualty's location. The main measurement of deprivation for England is the Index of Multiple Deprivation (IMD), which is a measure of relative deprivation of lower layer super output areas (LSOAs). LSOAs are small areas roughly equivalent to postcode areas, each with a similar population size. IMD ranks the LSOAs in England from most deprived to least deprived,

4. Analysis Results:

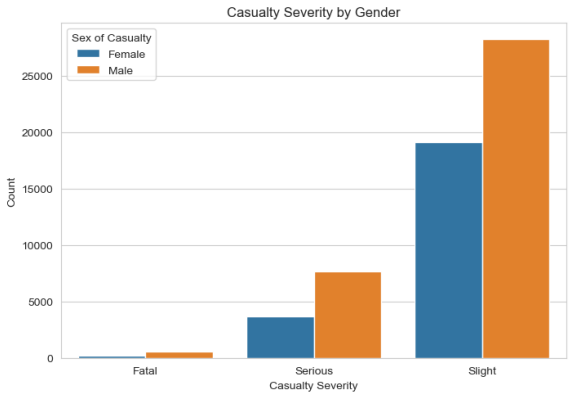
Our analysis yields several key findings:

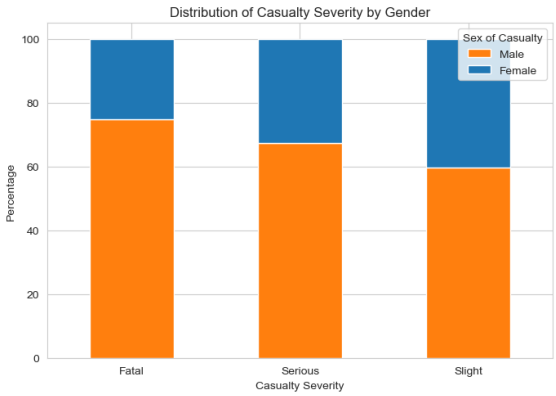
4.1 The majority of accidents are reported as slight, followed by serious and fatal.



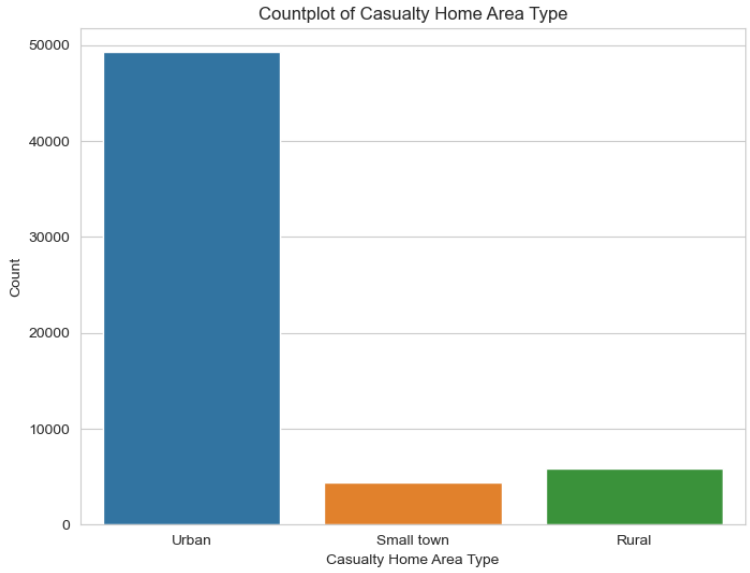


4.2 Male individuals are more frequently implicated in traffic accidents compared to their female counterparts.

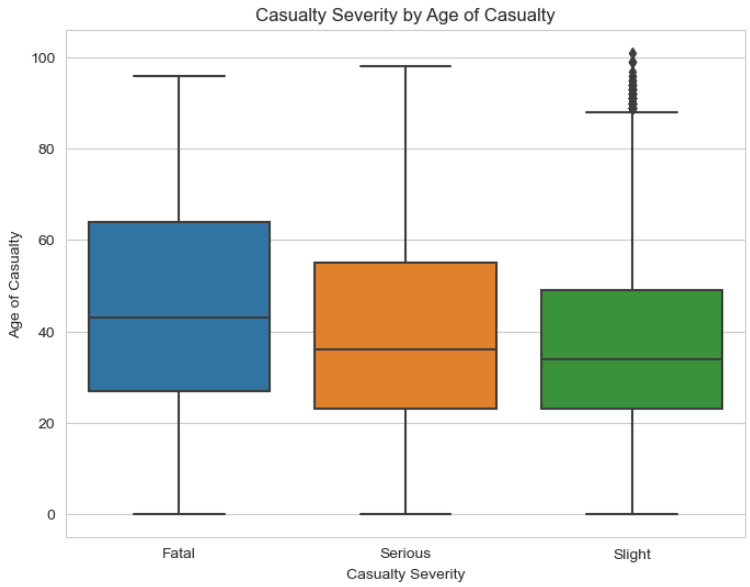




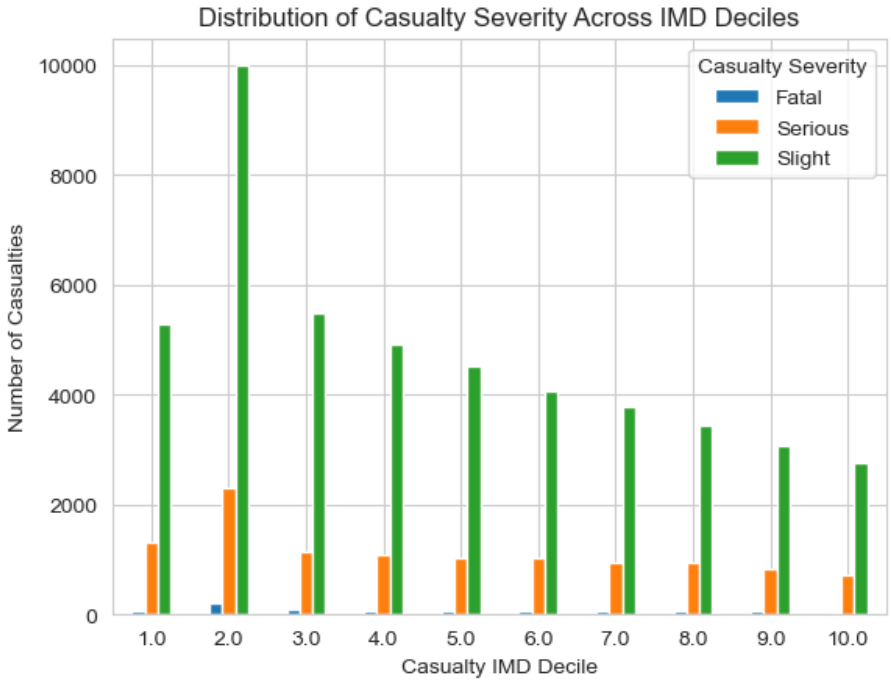
4.3 Most of the traffic accidents happened in urban areas



4.4 Individuals who experienced more severe accidents tend to be older compared to those involved in less severe incidents.



4.5 Certain factors such as IMD decile exhibit significant correlations with accident frequency and severity.



4.6 Machine learning models, including Decision Trees, Random Forests, Gradient Boosting, and XGBoost are utilized to predict accident severity with varying degrees of accuracy.

5. Discussion:

The analysis highlights the complexity of road accidents and the necessity of a multifaceted approach to road safety. Collaboration among government agencies, law enforcement, transportation authorities, and community organizations is essential to address factors like driver behavior, infrastructure deficiencies, and socio-economic disparities. Targeted interventions such as education campaigns, improved road design, and stricter enforcement of traffic laws can effectively mitigate accident risks. While machine learning models offer promise in predicting accident severity, ongoing research is required to refine their accuracy. The findings stress the importance of tailored interventions targeting specific risk factors identified in the analysis. Overall, analyzing road accident data offers valuable insights for developing effective prevention strategies and fostering safer road environments for all. This report lays the groundwork for further research and policy initiatives in road safety and accident prevention.

6. Suggestions to Reduce Traffic Accidents:

Based on our analysis, we recommend the following strategies to reduce traffic accidents:

1. Implement comprehensive road safety education programs targeting drivers, pedestrians, and vulnerable road users.
2. Invest in infrastructure improvements, including better signage, lighting, and pedestrian crossings, to enhance safety for all road users.
3. Enforce stricter regulations on speeding, drunk driving, and distracted driving to deter risky behavior and promote responsible driving.
4. Increase funding for road maintenance and repair to address infrastructure deficiencies and reduce accident risks.
5. Collaborate with local communities and stakeholders to address socio-economic disparities and improve access to safe transportation options.

By implementing these recommendations, policymakers and stakeholders can work towards creating safer roads and reducing the human and economic toll of traffic accidents.

Overall, the analysis of the road accident database provides valuable insights into the factors contributing to accidents and casualties, informing evidence-based strategies for improving road safety and saving lives.