Data Analysis Report

Road Accidents in India

Data Analysis using MySQL & Power BI

By Mayur Rajput

Project Overview

The "Road Accidents in India" data analysis project aims to explore and derive insights from a dataset obtained from Kaggle comprising three CSV files: Accidents, Casualties and Vehicles. The dataset contains detailed information about road accidents, casualties and vehicles involved. MySQL was utilized for data analysis to uncover patterns, trends and factors contributing to road accidents in India.

Project Goal:

The primary goal of this project is to leverage data analytics techniques, specifically using MySQL, to gain valuable insights into the patterns, trends, contributing factors of road accidents and also to provide recommendations based on the analysis to reduce accidents in India.
The insights derived from this analysis can inform policy decisions, guide targeted interventions and contribute to efforts aimed at improving road safety in India. Understanding the dynamics of road accidents is crucial for developing effective strategies to reduce their frequency and severity.

Objective:

1. **Total Accidents vs Total Casualties Analysis:** Examine the distribution of accidents and casualties over different months and days.
2. **Vehicle Distribution Analysis:** Analyze the distribution of vehicles involved in accidents over different months.
3. **Temporal Trends Analysis:** Investigate variations in accident severity and casualties severity over months.
4. **Severity-Based Distribution:** Explore the distribution of accidents and casualties based on accident severity.
5. **Top 10 Districts Analysis:** Identify and rank the top 10 Local Authority Districts by total casualties and total accidents.
6. **Driver's Age Impact Analysis:** Analyze the impact of driver's age on accidents, casualties, and their severity.
7. **Driver's Gender Impact Analysis:** Examine the impact of driver's gender on accidents, casualties, and their severity.
8. **Road and Weather Impact Analysis:** Analyze the impact of road types, weather conditions and road surface conditions on accidents.
9. **Casualty Gender Distribution Analysis:** Investigate the distribution of casualties based on their gender.
10. **Urban and Rural Analysis:** Explore the distribution of accidents and casualties in urban and rural areas.
11. **Top Districts by Severity Analysis:** Identify and rank the top 10 Local Authority Districts by accident and casualty severity.

Project Overview

Methodology

The project employs MySQL for Database creation, Data Collection, Data Cleaning, data querying, aggregation, data analysis and Data Visualizations using Power BI. Each task corresponds to specific SQL queries tailored to extract meaningful insights from the dataset.

Tools and Technologies:

1. **Database Management System:** MySQL
2. **Data Visualization:** Power BI for graphical representation.

Expected Outcomes:

1. **Comprehensive Distribution Insights:** Gain a thorough understanding of how accidents, casualties, and severity are distributed over time and across different variables.
2. **Identification of High-Risk Districts:** Rank districts based on total casualties, accidents, and severity, aiding in targeted interventions.
3. **Demographic and Environmental Impact:** Uncover the impact of driver age, gender, road types, and weather conditions on accidents and casualties.

Database Setup and Data Import

Database Creation:

The first step in our project involved creating a MySQL database to organize and store the road accident data. A dedicated database named "road_accidents_india" was created to facilitate efficient data management.

```
CREATE DATABASE IF NOT EXISTS road_accidents_india;
USE road_accidents_india;
```

Table Creation:

To store the data, relevant tables were created based on the CSV files—Accidents, Casualties and Vehicles. The table structures were designed to accommodate the specific attributes of each dataset.

Accidents Table

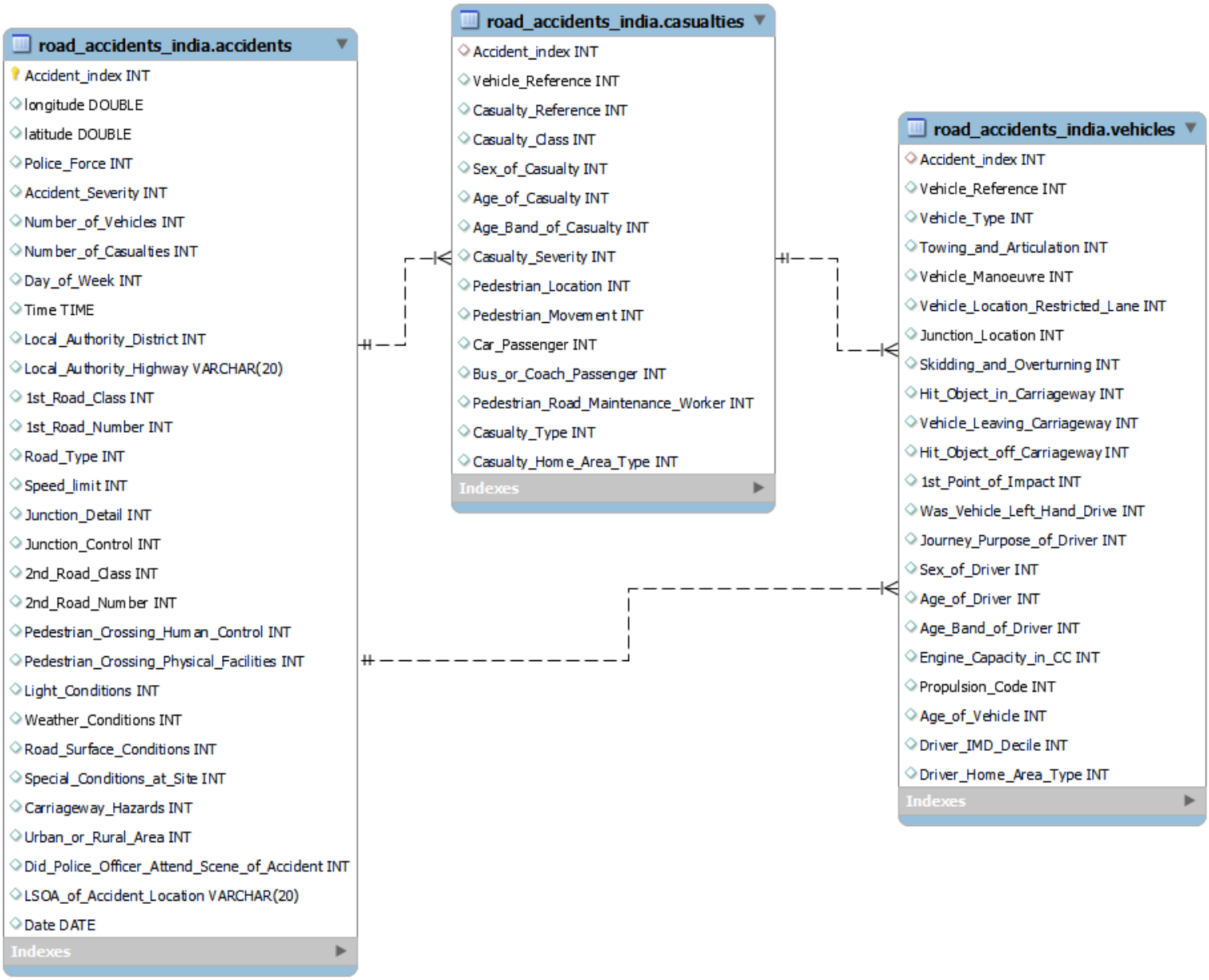
```sql
-- Creating Accident Table --
CREATE TABLE accidents(
    Accident_index INT PRIMARY KEY,
    longitude DOUBLE,
    latitude DOUBLE,
    Police_Force INT,
    Accident_Severity INT,
    Number_of_Vehicles INT,
    Number_of_Casualties INT,
    Day_of_Week INT,
    Time TIME,
    Local_Authority_District INT,
    Local_Authority_Highway VARCHAR(20),
    1st_Road_Class INT,
    1st_Road_Number INT,
    Road_Type INT,
    Speed_limit INT,
    Junction_Detail INT,
    Junction_Control INT,
    2nd_Road_Class INT,
    2nd_Road_Number INT,
    Pedestrian_Crossing_Human_Control INT,
    Pedestrian_Crossing_Physical_Facilities INT,
    Light_Conditions INT,
    Weather_Conditions INT,
    Road_Surface_Conditions INT,
    Special_Conditions_at_Site INT,
    Carriageway_Hazards INT,
    Urban_or_Rural_Area INT,
    Did_Police_Officer_Attend_Scene_of_Accident INT,
    LSOA_of_Accident_Location VARCHAR(20),
    Date DATE);
    SHOW TABLES;
    SELECT * FROM accidents;
```

Similar import statements were executed for creating Casualties and Vehicles tables.

Data Import:

The next step involved importing the records from the provided CSV files into the corresponding tables. MySQL's Table Data Import Wizard method was used for this purpose. Similar import statements were executed for the Casualties and Vehicles tables.
This database setup ensures a structured and organized environment for our analysis, allowing for seamless retrieval and manipulation of the data.

Entity Relationship Diagram:



Data Cleaning

- Columns that are not required for analysis or lack clear interpretation, such as lighting conditions, Junction_Detail, etc were removed to streamline the dataset.
- Eliminating irrelevant or unclear columns ensures a more focused and manageable dataset for analysis.

Tables overview:

```
SELECT * FROM accidents
LIMIT 10;
```

Accident_index	Accident_Severity	Number_of_Vehicles	Number_of_Casualties	Day_of_Week	Local_Authority_District	Road_Type	Weather_Conditions	Road_Surface_Conditions	Urban_or_Rural_Area	Date
1	2	1	1	3	12	6	2	2	1	2018-01-04
2	3	1	1	4	12	3	1	1	1	2018-01-05
3	3	2	1	5	12	6	1	1	1	2018-01-06
4	3	1	1	6	12	6	1	1	1	2018-01-07
5	3	1	1	2	12	6	1	2	1	2018-01-10
6	3	2	1	3	12	6	2	2	1	2018-01-11
7	3	2	1	5	12	6	1	1	1	2018-01-13
8	3	1	2	6	12	3	1	1	1	2018-01-14
9	3	2	2	7	12	6	1	1	1	2018-01-15
10	3	2	5	7	12	6	1	1	1	2018-01-15

```
SELECT * FROM casualties
LIMIT 10;
```

Accident_index	Sex_of_Casualty	Age_of_Casualty	Casualty_Severity
1	1	37	2
2	1	37	3
3	1	62	3
4	1	30	3
5	1	49	3
6	2	30	3
7	1	31	3
8	2	13	3
9	2	13	3
10	1	35	3

```
SELECT * FROM vehicles
LIMIT 10;
```

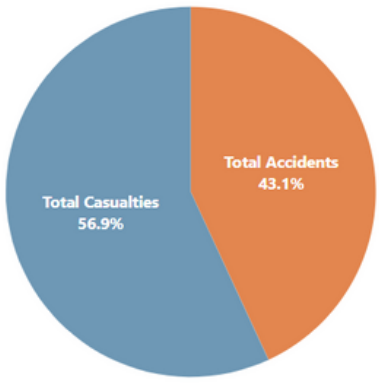
Accident_index	Sex_of_Driver	Age_of_Driver
1	2	74
2	1	42
3	1	35
4	1	62
5	2	49
6	1	49
7	1	51
8	2	30
9	1	31
10	1	41

Data Analysis Tasks

- Analyze the distribution of Accidents and Casualties

```
-- Problem Statement: Analyze the distribution of Accidents and Casualties
SELECT COUNT(accident_index) AS Total_Accidents, SUM(Number_of_Casualties) AS Total_Casualties
FROM Accidents;
```

	Total_Accidents	Total_Casualties
▶	59974	79208



Data Analysis Tasks

Task 1: Analyze the distributions of accidents, casualties caused over months and day of Week.

```
SELECT DATE_FORMAT(Date, '%m') AS Month, COUNT(accident_index) AS Accident_Count,
SUM(Number_of_Casualties) AS Total_Casualties FROM accidents
GROUP BY Month
ORDER BY Month;
```

Month	Accident_Count	Total_Casualties
01	4988	6543
02	4570	6005
03	4763	6230
04	5076	6715
05	5387	7079
06	5303	7064
07	4975	6637
08	4660	6199
09	4847	6325
10	5186	6850
11	5431	7189
12	4788	6372

```
SELECT Day_of_Week, COUNT(accident_index) AS Accident_Count,
SUM(Number_of_Casualties) AS Total_Casualties FROM accidents
GROUP BY Day_of_Week
ORDER BY Day_of_Week;
```

Day_of_Week	Accident_Count	Total_Casualties
1	6542	9240
2	8405	10863
3	9011	11564
4	9173	11787
5	8912	11599
6	9788	12769
7	8143	11386

Task 2: Analyze the distribution of vehicles involved in accidents over months.

```
SELECT DATE_FORMAT(Date, '%m') AS Month,
SUM(number_of_vehicles) AS Number_of_vehicles FROM accidents
GROUP BY Month
ORDER BY Month;
```

Month	Number_of_vehicles
01	9038
02	8393
03	8590
04	9337
05	9804
06	9826
07	9072
08	8577
09	8919
10	9524
11	10026
12	8764

Task 3: Analyze the variations in accident severity and casualties severity over months.

```
SELECT DATE_FORMAT(a.Date, '%m') AS Month, ROUND(AVG(a.accident_severity)) AS Average_Accident_Severity,
ROUND(AVG(c.Casualty_Severity)) AS Average_Casualty_Severity FROM accidents a
INNER JOIN casualties c
ON a.Accident_index = c.Accident_index
GROUP BY month
ORDER BY month;
```

Month	Average_Accident_Severity	Average_Casualty_Severity
01	3	3
02	3	3
03	3	3
04	3	3
05	3	3
06	3	3
07	3	3
08	3	3
09	3	3
10	3	3
11	3	3
12	3	3

Data Analysis Tasks

Task 4: Analyze the Distribution of Accidents and Casualties based on Accident Severity.

```
SELECT Accident_Severity, COUNT(Accident_index) AS Accident_count,
SUM(Number_of_Casualties) AS Total_Casualties FROM accidents
GROUP BY Accident_Severity
ORDER BY Accident_Severity;
```

Accident_Severity	Accident_count	Total_Casualties
1	646	1202
2	7372	10473
3	51956	67533

Task 5: Identify Top 10 Local Authority District by Total Casualties and Total Accidents.

```
SELECT ROW_NUMBER() OVER (ORDER BY Total_accidents DESC) AS Rank_,
Local_Authority_District, Total_Accidents, Total_Casualties FROM (
    SELECT Local_Authority_District, count(Accident_Index) AS Total_Accidents,
    sum(Number_of_Casualties) AS Total_Casualties
    FROM accidents
    GROUP BY Local_Authority_District
    ORDER BY Total_accidents DESC
    LIMIT 10) AS Ranked_Districts;
```

Rank_	Local_Authority_District	Total_Accidents	Total_Casualties
1	102	2336	3173
2	91	2021	3037
3	1	1564	1762
4	9	1176	1335
5	20	1141	1412
6	30	1137	1356
7	27	1108	1318
8	100	1061	1471
9	147	1033	1351
10	8	1018	1148

Task 6: Analyze the impact of Driver's Age on Accidents, Casualties and their Severity.

```
SELECT
    CASE
        WHEN v.Age_of_Driver BETWEEN 0 AND 17 THEN '0-17'
        WHEN v.Age_of_Driver BETWEEN 18 AND 24 THEN '18-24'
        WHEN v.Age_of_Driver BETWEEN 25 AND 34 THEN '25-34'
        WHEN v.Age_of_Driver BETWEEN 35 AND 44 THEN '35-44'
        WHEN v.Age_of_Driver BETWEEN 45 AND 64 THEN '45-64'
        WHEN v.Age_of_Driver >= 65 THEN '65+'
        ELSE 'Unknown'
    END AS Age_Group, COUNT(a.Accident_index) AS Accident_count,
    SUM(a.Number_of_Casualties) AS Total_Casualties,
    ROUND(AVG(a.Accident_Severity)) AS Average_Accident_Severity,
    ROUND(AVG(Casualty_Severity)) AS Average_Casualty_Severity
    FROM accidents a
INNER JOIN vehicles v
ON a.Accident_index = v.Accident_index
INNER JOIN casualties c
ON a.Accident_index = c.Accident_index
GROUP BY Age_Group
ORDER BY Age_Group;
```

Age_Group	Accident_count	Total_Casualties	Average_Accident_Severity	Average_Casualty_Severity
0-17	1488	2040	3	3
18-24	7568	10123	3	3
25-34	12435	16470	3	3
35-44	11807	15593	3	3
45-64	10817	14395	3	3
65+	2210	2984	3	3

Data Analysis Tasks

Task 7: Analyze the impact of Driver's Gender on Accidents, Casualties and their Severity.

```
SELECT v.Sex_of_Driver AS Driver_Gender, COUNT(a.accident_index) AS Accident_Count,
SUM(a.Number_of_Casualties) AS Total_Casualties,
ROUND(AVG(a.Accident_Severity)) AS Average_Accident_Severity,
ROUND(AVG(Casualty_Severity)) AS Average_Casualty_Severity
FROM accidents AS a
INNER JOIN vehicles AS v
ON a.Accident_index = v.Accident_index
INNER JOIN casualties c
ON a.Accident_index = c.Accident_index
GROUP BY Driver_Gender;
```

Driver_Gender	Accident_Count	Total_Casualties	Average_Accident_Severity	Average_Casualty_Severity
1	33984	45049	3	3
2	12341	16556	3	3

Task 8: Analyze the impact of road types, Wheather Conditions and Road Surface Condition on Accidents.

```
SELECT * FROM accidents;
SELECT Road_Type, COUNT(Accident_index) AS Accident_Count,
ROUND(AVG(Accident_Severity)) AS Average_Accident_Severity FROM accidents
GROUP BY Road_Type
ORDER BY Road_Type;
```

Road_Type	Accident_Count	Average_Accident_Severity
1	2993	3
2	1630	3
3	8780	3
6	45632	3
7	470	3
9	469	3

```
SELECT Weather_Conditions, COUNT(Accident_index) AS Accident_Count,
ROUND(AVG(Accident_Severity)) AS Average_Accident_Severity FROM accidents
GROUP BY Weather_Conditions
ORDER BY Weather_Conditions;
```

Weather_Conditions	Accident_Count	Average_Accident_Severity
1	49789	3
2	7073	3
3	406	3
4	581	3
5	522	3
6	49	3
7	227	3
8	756	3
9	571	3

Data Analysis Tasks

```sql
SELECT Road_Surface_Conditions, COUNT(Accident_index) AS Accident_Count,
ROUND(AVG(Accident_Severity)) AS Average_Accident_Severity FROM accidents
GROUP BY Road_Surface_Conditions
ORDER BY Road_Surface_Conditions;
```

Road_Surface_Conditions	Accident_Count	Average_Accident_Severity
-1	50	3
1	43545	3
2	15481	3
3	231	3
4	617	3
5	50	3

Task 9: Analyze the distribution of casualties based on their Gender.

```sql
SELECT c.sex_of_Casualty, SUM(a.number_of_casualties) AS Casulaties_Count FROM accidents a
INNER JOIN casualties c
ON a.Accident_index = c.Accident_index
GROUP BY Sex_of_Casualty
ORDER BY Sex_of_Casualty;
```

sex_of_Casualty	Casulaties_Count
1	45095
2	30778

Task 10: Analyze the distribution of Accidents and Casualties in urban and rural areas.

```sql
SELECT Urban_or_Rural_Area, COUNT(accident_index) AS Accident_Count FROM accidents
GROUP BY Urban_or_Rural_Area
ORDER BY Urban_or_Rural_Area;
```

Urban_or_Rural_Area	Accident_Count
1	48901
2	11073

Task 11: Identify Top 10 Local Authority District by Accident and Casualty Severity.

```sql
SELECT a.local_authority_district, ROUND(AVG(a.accident_severity)) AS Average_Accident_Severity,
ROUND(AVG(c.casualty_severity)) AS Average_Casualty_Severity FROM accidents a
INNER JOIN casualties c
ON a.accident_index = c.accident_index
GROUP BY a.local_authority_district
ORDER BY Average_Accident_Severity, Average_Casualty_Severity
LIMIT 10;
```

local_authority_district	Average_Accident_Severity	Average_Casualty_Severity
12	3	3
1	3	3
11	3	3
570	3	3
5	3	3
8	3	3
4	3	3
3	3	3
2	3	3
10	3	2

Key Findings

1. **Total Accidents and Casualties:**
   - Total Accidents: 59,974
   - Total Casualties: 79,208
   - The substantial number of casualties highlights the urgent need for comprehensive road safety initiatives.

2. Monthly and Weekly Distributions:
   - Accidents peak in November, with consistently high casualties throughout the year.
   - Weekdays, especially Friday, witness a higher occurrence of accidents and casualties.
   - Implementing targeted safety measures during peak months and days is crucial for mitigating specific challenges.

3. Vehicle Involvement Over Months:
   - The number of vehicles involved remains relatively consistent throughout the year.
   - Continuous monitoring and safety interventions are required to address consistent challenges.

4. Severity Analysis:
   - Average accident severity and casualty severity remain constant at level 3 throughout the year.
   - Identifying specific contributing factors to severity level 3 accidents requires further investigation.

5. Accident Severity and Casualty Distribution:
   - Most accidents fall under severity level 3, contributing to the majority of casualties.
   - Focusing on reducing severity level 3 accidents could substantially reduce casualties.

6. Top 10 Local Authorities by Casualties and Accidents:
   - Local Authority Districts with higher accidents may benefit from enhanced safety measures.
   - Implementing targeted safety programs in these districts could lead to a substantial reduction in accidents and casualties.

7. Impact of Driver's Age and Gender:
   - Accidents and casualties are spread across age groups, indicating the need for age-specific safety initiatives.
   - Male drivers are involved in a significantly higher number of accidents and casualties.
   - Tailored awareness campaigns for different age groups and gender-specific safety measures are crucial.

Key Findings

8. **Impact of Road Types, Weather and Surface Conditions:**
    - Accidents are frequent on single carriageways, during clear weather and on dry surfaces.
    - Focusing on improving road infrastructure, weather-aware driving campaigns and road maintenance can contribute to accident reduction.

9. **Casualties Based on Gender and Urban/Rural Areas:**
    - Male casualties outnumber females, emphasizing gender-specific safety concerns.
    - Accidents are more prevalent in urban areas.
    - Urban areas require targeted safety interventions and gender-specific initiatives can enhance overall road safety.

10. **Top 10 Local Authorities by Severity:**
    - Local Authorities exhibit uniformity in accident and casualty severity.
    - Despite uniform severity, local interventions may address specific challenges in these areas.

Recommendations

1. **Targeted Safety Campaigns:**
    - Implement targeted safety campaigns during peak months and days to address season-specific challenges.

2. **Continuous Monitoring:**
    - Implement continuous monitoring of vehicle involvement trends to identify emerging patterns and risks.

3. **Severity-Focused Interventions:**
    - Develop interventions specifically targeting severity level 3 accidents to reduce casualties.

4. **Localized Safety Measures:**
    - Implement localized safety measures in districts with high accident and casualty rates.

5. **Age and Gender-Specific Initiatives:**
    - Develop age-specific and gender-specific awareness campaigns to address the unique challenges faced by different demographic groups.

6. **Infrastructure Improvement:**
    - Focus on improving road infrastructure, especially on single carriageways, to reduce accidents.

Recommendations

7. **Weather-Aware Driving Programs:**
   - Launch campaigns to raise awareness about safe driving practices during adverse weather conditions.
8. **Urban Safety Initiatives:**
   - Implement urban-specific safety initiatives to address the unique challenges posed by urban road environments.
9. **Community Engagement:**
   - Ingage local communities in identifying and addressing safety concerns to ensure effective interventions.
10. **Continuous Evaluation:**
    - Regularly evaluate the impact of implemented interventions and adjust strategies based on evolving road safety needs.

Conclusion

This comprehensive analysis provides a holistic understanding of road accidents in India, laying the foundation for informed and targeted interventions. The recommended strategies encompass a range of initiatives, from focused awareness campaigns to infrastructure improvements, with a continuous evaluation framework ensuring adaptive and effective road safety measures. The implementation of these recommendations is critical for creating a safer and more secure road environment across the country.