Road Accident Analysis Report (India, 2020)

A. Objective

The purpose of this analysis is to explore and visualize road accident patterns, major causes, and their outcomes in million-plus cities across India for the year 2020. This report aims to identify accident hotspots, highlight critical causes, and support data-driven decisions for road safety improvement.

B. Dataset Overview

- **Source**: Government database (2020 Road Accident Records)
- **Records**: 9,550
- **Key Columns**:
    - **Million Plus Cities**
    - **Cause category**
    - **Cause Subcategory**
    - **Outcome of Incident**
    - **Count** (number of incidents)

Data Cleaning Summary

- Missing values: 3 rows had missing 'Count' values
- Action taken: Removed missing rows
- Count column: Converted to integer
- Final record count: **9,547**

C. Exploratory Data Analysis

1. Total Accidents by Cause Category

A bar plot was generated to represent the total number of accidents under each major cause category. Prominent categories included: - **Driver Fault** - **Traffic Control Failures** - **Vehicle Condition** - **Road Environment**

These categories accounted for the bulk of accident counts across cities.

2. Top 10 Cities by Total Accidents

The top 10 cities with the highest accident counts were identified using a grouped sum by "Million Plus Cities." Cities such as Delhi, Mumbai, and Bengaluru stood out as high-frequency accident zones. These cities also tend to have high vehicular density and complex traffic management needs.

3. Heatmap: Cities vs. Cause Categories

A pivot table was created with cities as rows and cause categories as columns. This was visualized as a heatmap to understand the variation of causes by city. Some notable trends included: - **Delhi**: High counts from *Driver Fault* and *Traffic Control* - **Mumbai**: Significant share of *Road Infrastructure* causes - **Bengaluru**: High in *Pedestrian Negligence* and *Overspeeding*

4. Outcome Distribution (All Cities)

The dataset's outcome column included various consequences of road accidents: - **Minor Injury** - **Grievously Injured** - **Persons Killed** - **Total Accidents**

A countplot visualized the frequency of each outcome. While minor injuries were the most frequent, the number of fatalities was also considerable, signaling the need for urgent preventive measures.

5. Cause Subcategories (Top 10)

Beyond broad categories, subcategories helped identify specific triggers. Top subcategories included: - **Overspeeding** - **Driving Under Influence** - **Mechanical Failure** - **Obstacles on Road** - **Lack of Signage or Signals**

These accounted for a large share of total accident counts and were visualized using a horizontal bar chart.

6. Heatmap: Outcome by Cause Category

A second heatmap was created to map outcomes (e.g., persons killed, minor injuries) to their respective cause categories. This analysis was crucial in understanding which causes are most fatal.

7. Distribution of Accidents Across Cities

Finally, a horizontal bar plot showed accident counts across all million-plus cities in descending order. This offered an at-a-glance view of urban risk profiles.

D. Key Insights

- **Urban Risk Zones**: Metropolitan cities with high population density experienced the most accidents. Delhi, Mumbai, Chennai, and Hyderabad ranked consistently high.
- **Dominant Causes**: *Driver Fault*, *Overspeeding*, and *Failure to Observe Traffic Rules* were among the top contributors to accidents.
- **Injury vs Fatality**: While most accidents resulted in minor injuries, grievous injuries and deaths were notably high for categories like *Drunken Driving* and *Speeding*.

- **City-Specific Patterns**: Some cities had distinct accident profiles, e.g., Kolkata showed more pedestrian-related issues, while Pune had higher incidents from poor vehicle conditions.

E. Conclusion and Recommendations

This data-driven investigation reveals patterns that can guide policymaking and preventive strategies. It confirms that high-density cities and reckless driving behaviors are leading contributors to India's urban accident burden.

Recommendations

1. **Enhanced Enforcement**: Target overspeeding and intoxicated driving through stricter surveillance and penalties.
2. **Infrastructure Overhaul**: Upgrade poor road conditions, particularly in cities where these causes dominate.
3. **City-Specific Safety Plans**: Design customized action plans for cities with unique causal patterns.
4. **Awareness Campaigns**: Focused education for drivers and pedestrians on high-risk behaviors.
5. **Outcome Monitoring**: Use outcomes-to-cause mapping for smarter emergency response and medical preparedness.

F. Future Work

- **Time Series Analysis**: Monthly trends in accidents can reveal seasonal risks.
- **Machine Learning Prediction**: Build models to predict accident likelihood based on location, time, and driver behavior.
- **Integration with Weather and Traffic Data**: Enhance causal analysis using external variables.