Title: Regulatory Affairs of Road Accident Data – India (2020)

Submitted By:

Name: Gagan Ruthwik Chowdary Kolluri

Email: gaganbublu2005@gmail.com

Phone: 9441490852

Acknowledgement

I would like to express my sincere gratitude to Unified Mentor for providing me with the opportunity to undertake this virtual internship in Data Science. This project, titled *"Regulatory Affairs of Road Accident Data – India (2020)"*, has been a significant learning experience and has helped me strengthen my analytical and technical skills.

I extend my heartfelt thanks to the mentors and coordinators at Unified Mentor for their constant support, valuable guidance, and constructive feedback throughout the duration of this project.

I would also like to thank my family and friends for their encouragement and moral support, which kept me motivated during the course of this project.

Last but not least, I am grateful for the availability of open data sources and the tools that made this analysis possible, including Python, Jupyter Notebook, and various data visualization libraries.

Gagan Ruthwik Chowdary Kolluri,

9441490852,

4th May.

Table Of Content

Abstract

Every year, thousands of lives are lost on Indian roads due to preventable accidents. This project takes a deep dive into the road accident data from 2020, aiming to uncover the main factors behind these incidents. By exploring city-wise accident counts, identifying the most frequent and fatal causes, and visualizing the outcomes, this study helps paint a clearer picture of where things go wrong—and more importantly, how we can make things right. The analysis leverages Python for data cleaning, visualization, and even includes a small predictive modeling segment to forecast accident outcomes. The results serve as a foundation for smarter policies and more targeted road safety interventions.

Introduction

Road accidents in India are not just statistics—they're real stories of loss, pain, and missed potential. In 2020 alone, even amid a global pandemic, road accidents claimed thousands of lives and left many more injured. While some causes are well-known, others go under the radar due to lack of awareness or enforcement.

This project aims to step in with data and clarity. By analyzing the official road accident data across Indian cities for 2020, we try to answer key questions:

- Which cities see the most accidents?
- What are the most common causes?
- Which of these causes are most deadly?
- Can this data help predict accident outcomes?

By digging into these areas, we hope to provide insights that are not just informative but actionable—something that can actually guide policymakers, urban planners, and safety officials.

Dataset Description

The dataset used in this project is titled **"Regulatory Affairs of Road Accident Data – India (2020)"**, and it presents a structured record of road accidents that occurred across various Indian cities during the year 2020.

What's in the dataset?

Here's a breakdown of the key columns:

- **City**: The location where the accident occurred.
- **Cause_Category**: The broader classification of what caused the accident (e.g., driver error, vehicle defect).
- **Cause_Subcategory**: A more specific reason under each category (e.g., overspeeding, drunken driving).
- **Outcome**: The result of the accident—either people were **injured**, **killed**, or both.
- **Count**: The number of people affected by the accident outcome.

Dataset Size:

- **Rows**: ~2000 entries
- **Columns**: 5

This dataset provides a powerful combination of both **quantitative** (count of incidents) and **qualitative** (causes, outcomes) data, making it ideal for exploration, visualization, and drawing useful insights.

Data Cleaning & Preprocessing

Before diving into analysis, some cleaning was necessary to ensure the dataset was reliable and usable. Here's what was done:

Step 1: Handling Missing Values

We found a few rows with missing values in the Count column. These rows were removed because incomplete count data could distort our results.

Step 2: Removing Duplicates

A quick check for duplicate rows showed a few repeats, which were dropped to maintain accuracy.

Step 3: Renaming Columns

For clarity and simplicity during analysis, we renamed the columns to:

- City
- Cause_Category
- Cause_Subcategory
- Outcome
- Count

Step 4: Standardizing Data Types

All relevant columns were ensured to be of type string (for categorical values) or integer (for counts). This helped avoid issues during group-by operations and visualizations.

Exploratory Data Analysis (EDA)

EDA helps us understand the structure, patterns, and outliers in the dataset. We used a mix of summary statistics and visualizations to answer key questions like: *Where do the most accidents happen? What causes the most fatalities? Which cities are most at risk?*

---

1. Accidents by City

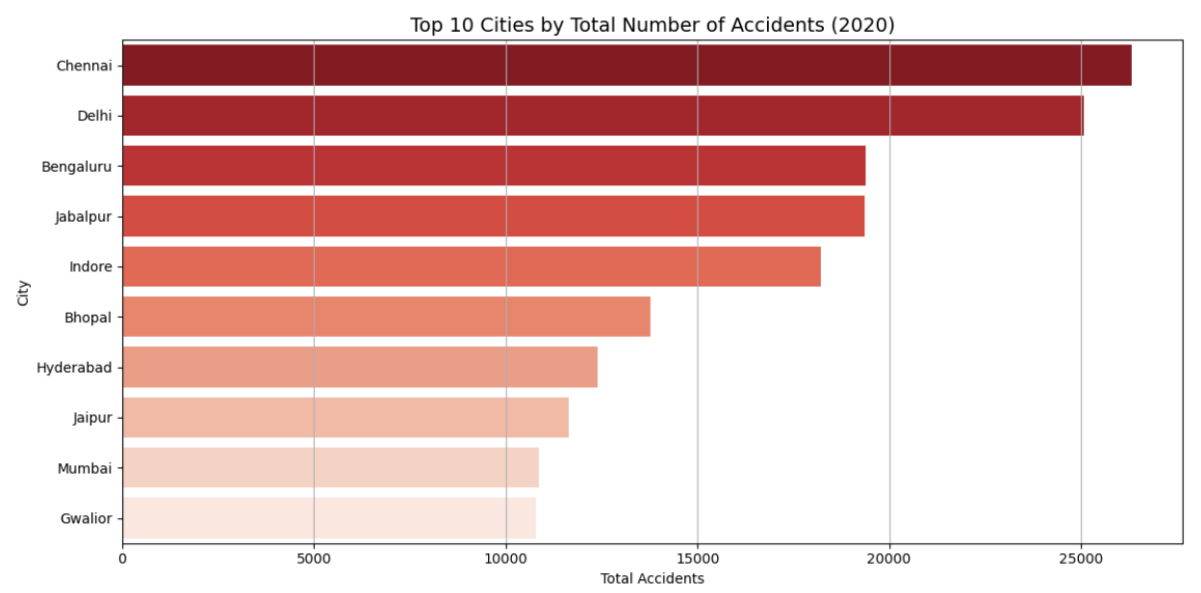
We started by grouping accidents based on city names to see which cities reported the highest number of incidents.

Insight:

- Major cities like **Delhi**, **Mumbai**, **Bangalore**, and **Hyderabad** topped the list.
- These cities, being densely populated with high vehicle density, naturally have a higher accident frequency.

Visualization:

A horizontal bar chart clearly showed the top 10 cities with the most accidents. It helped visualize urban concentration of road risk.



---

2. Distribution of Outcomes

Next, we looked at the types of outcomes reported:

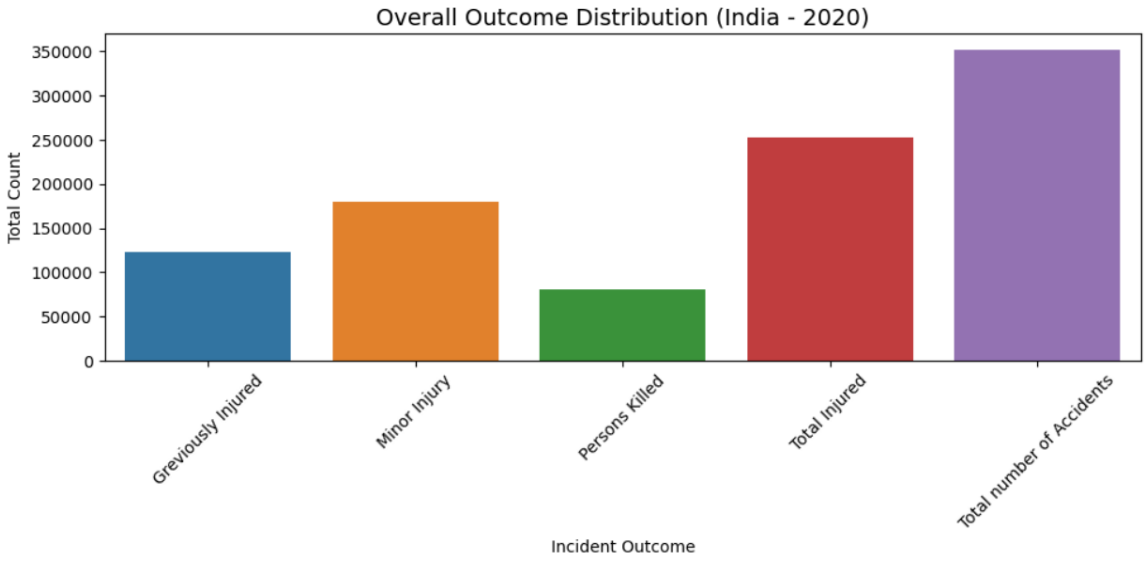
- **Persons Killed**
- **Persons Injured**
- **Fatal Accidents**
- **Total Accidents**

Insight:

- A large number of accidents resulted in **injuries**, but the number of **fatal cases** was still alarmingly high.
- Some outcomes, like "fatal accidents," may include multiple victims, so counts can overlap.

Visualization:

A pie chart and count plot were used to show the share of each outcome type across all reported incidents.



---

3. What's Causing These Accidents?

To understand the *why*, we explored both **Cause Categories** and **Cause Subcategories**.

Top Categories:

- **Driver's Fault** was the most frequent cause category by far.
- Other notable categories included **Vehicle Defects** and **Poor Road Infrastructure**.

Top Subcategories:

- **Overspeeding**

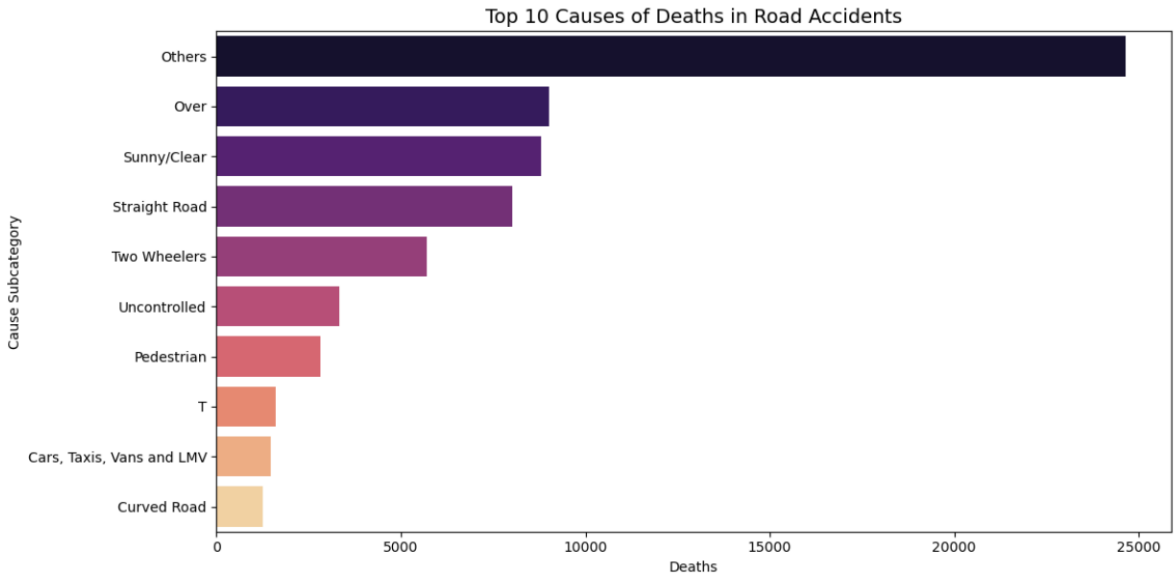
- **Drunken Driving**
- **Red Light Jumping**
- **Distraction (e.g., using mobile phones)**

Insight:

- Overspeeding alone accounted for a major chunk of deaths.
- Drunken driving, although less frequent than overspeeding, had a high fatality ratio.

Visualization:

Separate bar charts were created for both cause categories and subcategories, ranked by total number of fatalities.



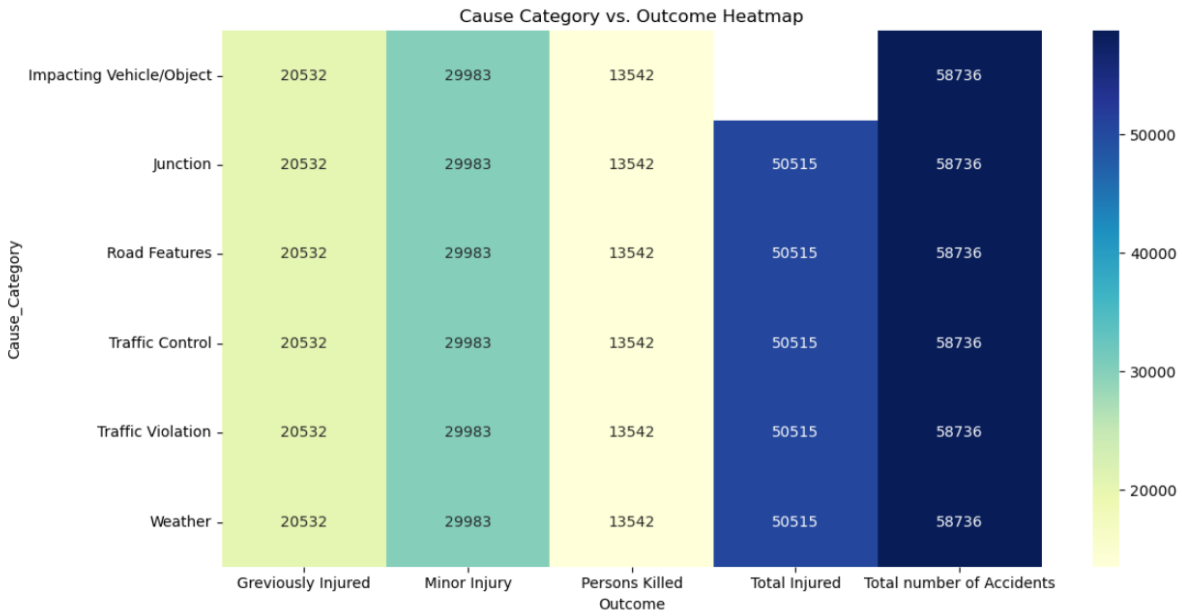
---

4. Heatmap: Causes vs Outcomes

We created a heatmap to visualize how different causes are linked to different outcomes.

Insight:

- **Overspeeding** and **drunken driving** lit up the "Persons Killed" column.
- Some subcategories, while frequent, led more often to **injuries** than deaths (e.g., minor signal violations).



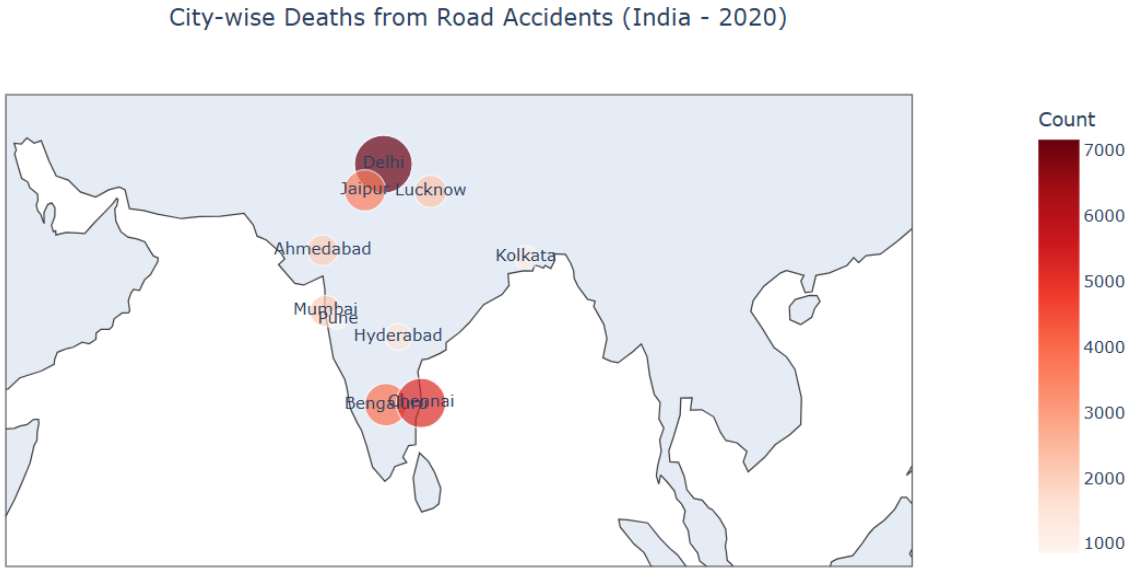
---

5. Bubble Map: Deaths by City

To add a spatial layer to the analysis, we created a bubble map using city coordinates. Each bubble represented a city, and its size was proportional to the number of fatalities.

Insight:

- High fatality clusters were visible in **Delhi NCR**, **Mumbai**, **Chennai**, and **Bangalore**.
- This can guide region-specific enforcement policies.



Findings & Insights

After thoroughly exploring the data, several patterns became clear. These findings don't just reflect numbers—they highlight real-world problems that need urgent attention.

Key Takeaways:

1. **Driver's Fault is the #1 Cause**
    - Categories like **overspeeding**, **negligent driving**, and **drunken driving** dominate the charts.
    - Driver behavior continues to be the single biggest contributor to road accidents.
2. **Overspeeding is Deadlier Than Frequent**
    - While overspeeding happens a lot, it also has a disproportionately high fatality rate.
    - It's not just about frequency—it's about severity.
3. **Urban Centers are High-Risk Zones**
    - Cities like **Delhi**, **Mumbai**, and **Bangalore** see the highest number of accidents and fatalities.
    - These cities combine high vehicle density, aggressive driving behavior, and stressed infrastructure.
4. **Certain Causes Lead to More Deaths than Injuries**
    - Some violations (e.g., **drunken driving**, **head-on collisions**) are more likely to be fatal.
    - Others, like **rear-end collisions** or **signal jumping**, tend to result in injuries more often than deaths.
5. **Inadequate Infrastructure Also Plays a Role**
    - Poor road conditions, lack of proper signage, and mechanical failures (like brake issues) were noted as contributing factors, though less common.

Recommendations

Based on the analysis, here are a few practical steps that could help reduce road accidents and save lives:

---

1. Stricter Law Enforcement for Overspeeding & Drunk Driving

- Use automated speed detection and breathalyzer checkpoints more aggressively in high-risk zones.
- Introduce tiered penalties and mandatory awareness sessions for repeat offenders.

---

2. Awareness Campaigns on Driver Responsibility

- Educate the public about how personal decisions—like phone use while driving or not wearing seat belts—can mean the difference between life and death.
- Targeted social media campaigns can be especially effective among younger drivers.

---

3. Improve Road Infrastructure in High-Incidence Cities

- Expand traffic signal visibility, repair potholes, and clearly mark pedestrian crossings.
- Introduce more roundabouts or smart traffic management systems in congestion-prone areas.

---

4. Mandatory Vehicle Fitness Testing

- Regular check-ups for brakes, steering, and lights should be enforced more strictly, especially for commercial and older vehicles.

---

5. Data-Driven Policing

- Use accident data to predict hotspots and deploy traffic police accordingly.
- Invest in GIS and traffic analytics for smarter urban planning.

Conclusion

India's road accident crisis isn't just a matter of poor driving—it's a complex blend of human error, weak enforcement, and infrastructure gaps. But through the lens of data, we can identify where the cracks lie and how to fix them.

This project made it clear that **overspeeding, careless driving, and urban congestion** are at the heart of the issue. By addressing these with **policy, technology, and education**, we can take a solid step toward safer roads.

While this analysis focused on 2020, similar efforts on a year-over-year basis could reveal deeper trends and help measure the impact of policy interventions. With the right blend of data and action, change is not only possible—it's within reach.

Github Link: https://github.com/Gaganruthwik013/Regulatory-Affairs-of-Road-Accident-Data-2020-India