Abhijith Varma
2211CS010007

Road Accidents Data Analysis Report

1. Introduction

This report presents a comprehensive analysis of road accidents in India (2001–2014) using a dataset containing state-wise accident records. The analysis was carried out using PySpark, Pandas, and visualization libraries such as Matplotlib and Seaborn.

The primary objective of this analysis is to uncover patterns and trends in road accidents, identify high-risk states and time periods, and provide insights that can assist in policy-making, traffic management, and public safety measures.

2. Dataset Overview

Total Rows: 868

Total Columns: 24

Years Covered: 2001–2014

Missing Values: 0

Attributes include:

STATE/UT – State or Union Territory where accidents were recorded

YEAR – Year of record

JAN–DEC – Number of accidents in each month

TOTAL14 – Annual total accidents for that state/UT

0–3 hrs to 21–24 hrs – Distribution of accidents by time of day

Abhijith Varma
2211CS010007

3. Key Findings

3.1 Accident Distribution

Accident data spans 29 states/UTs over 14 years.States such as Tamil Nadu, Maharashtra, and Uttar Pradesh consistently report the highest accident counts.Smaller UTs and northeastern states report significantly fewer accidents.Certain months show spikes in accident counts, suggesting seasonal or festival-related surges in traffic.

3.2 Time-of-Day Patterns

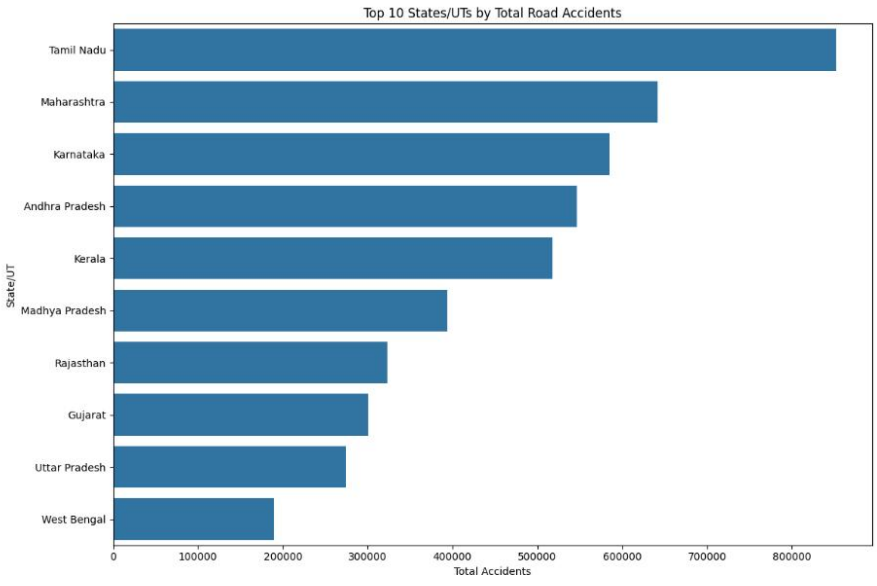
Afternoon and evening hours (15–21 hrs) recorded the highest accident frequencies, coinciding with peak traffic.Night-time (0–6 hrs) accidents are also significant, likely due to fatigue, poor lighting, and visibility issues

3.3 Correlations

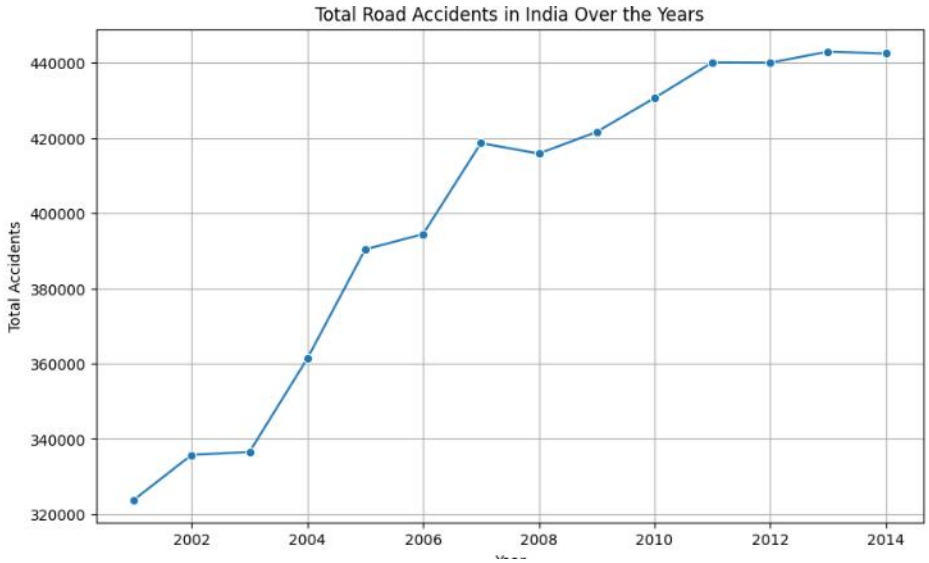
Strong correlation between monthly totals and annual totals, ensuring data consistency. Accident counts show patterns strongly linked with population density, urbanization, and vehicle growth

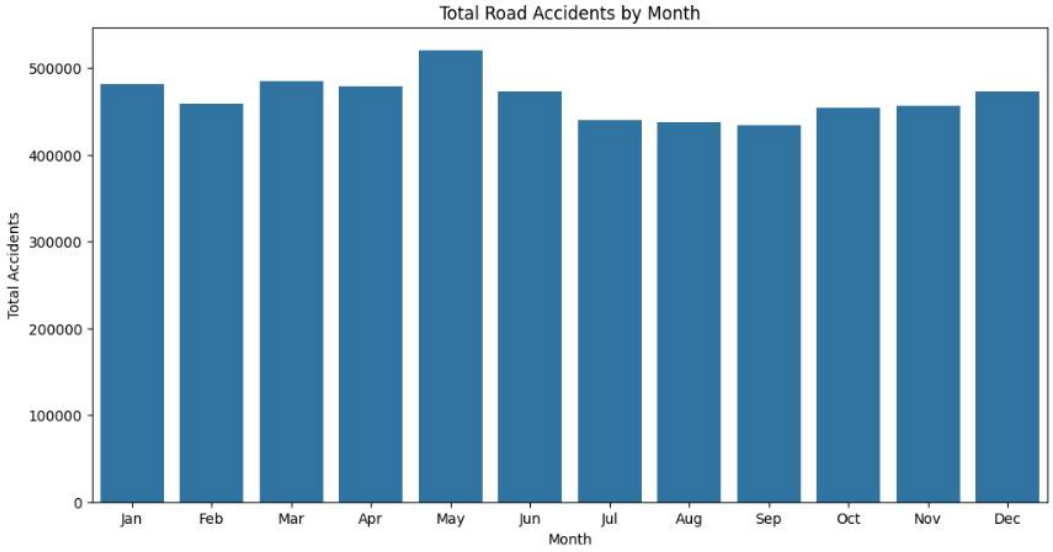
4. Data Visualizations

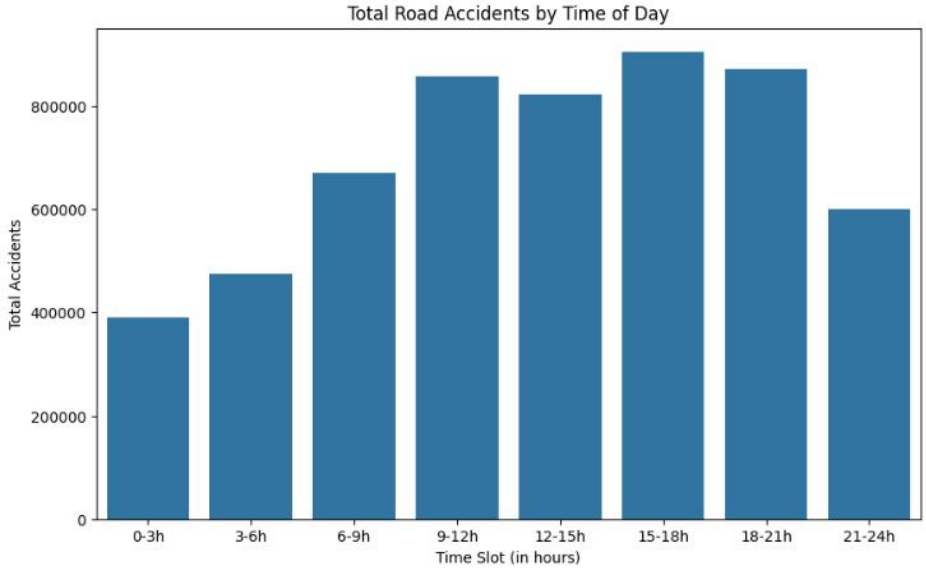
The following visualizations illustrate the insights obtained from the analysis:



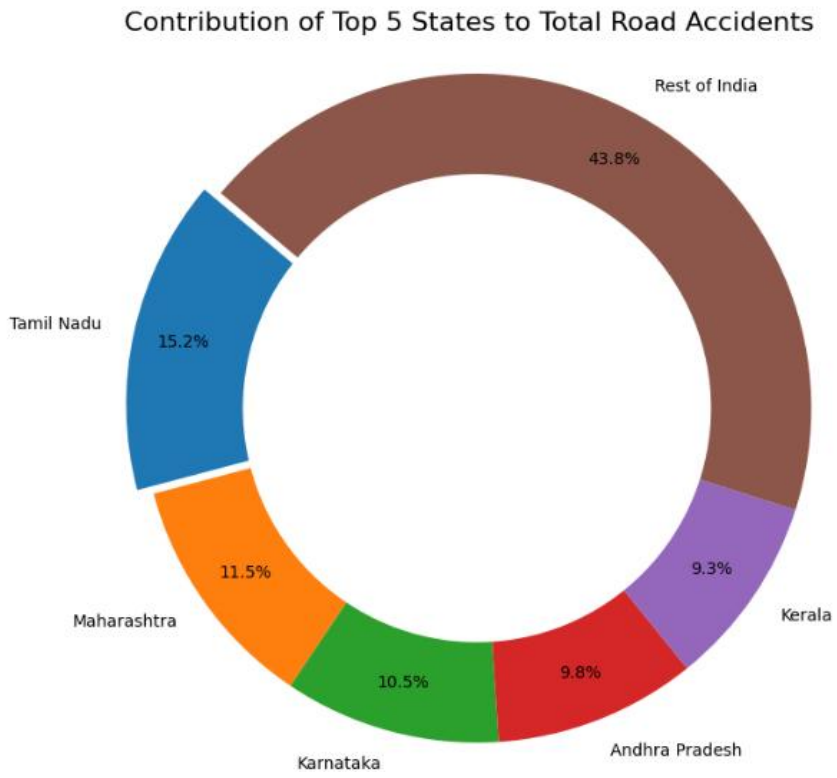
Abhijith Varma
2211CS010007







Abhijith Varma
2211CS010007



5. **Conclusion**

The analysis of road accidents in India from 2001–2014 provides valuable insights into regional disparities, temporal patterns, and risk factors:

Accidents are not evenly distributed — some states consistently emerge as hotspots.

Evening and night slots are high-risk, requiring targeted interventions.

Southern and highly populated states report higher accidents, correlating with urbanization and vehicle density.

Implications:

Policymakers should **focus interventions** on high-accident states and peak hours.

Road safety campaigns** and **better enforcement** during evening/night periods are essential.

Investments in **traffic management, infrastructure upgrades, and awareness programs** will play a key role in reducing accident rates.