



Traffic Accident Data Analysis

(2020 – October 2025)

Analyze accident data to identify patterns related to cause, time, and location.

Submitted by: Zain Ali

Roll No: 21

Submitted to: Mr Bilal Ahmed

UNIVERSITY OF KOTLI, AZAD JUMMU & KASHMIR

Dapartment of Data Science

Subject: Intro to Data Science

Semester: 4th

Submission Data: 11 November 2025

Table of Contents

1. Abstract
2. Introduction
3. Objectives
4. Literature Review
5. Dataset Description
6. Methodology
7. Results and Discussion
8. Conclusion
9. Future Work
10. References

1. Abstract

This project, *Traffic Accident Data Analysis*, focuses on studying and analyzing road accident data to identify patterns related to **time and location**. By analyzing accident trends across different years and regions of Pakistan, we can gain valuable insights that help improve road safety measures and reduce fatalities. Using Python data visualization tools such as **Pandas** and **Plotly Express**, this project highlights accident frequency, severity, and variations by year and region. The goal is to present meaningful information that can guide policymakers and road safety authorities.

2. Introduction

Traffic accidents remain one of the leading causes of injury and death worldwide. Each year, millions of people are injured or killed in road traffic crashes. In developing countries like Pakistan, these numbers are particularly high due to poor infrastructure, limited traffic awareness, and weak law enforcement.

This project aims to analyze real accident data to uncover meaningful trends. The analysis considers factors such as **year**, **place**, and **severity** (fatal vs. non-fatal). By visualizing these relationships, we can identify areas that need improvement in terms of infrastructure, education, and policy enforcement.

3. Objectives

The main objectives of this project are as follows:

- To analyze traffic accident data across multiple years (2008–2018).
- To identify trends and patterns based on **time** and **location**.
- To evaluate the relationship between total accidents, fatalities, and injuries.
- To visualize the data through graphs and charts for easier understanding.
- To provide insights that can help reduce accident rates and improve road safety in Pakistan.

4. Literature Review

Many studies have demonstrated the value of traffic data analysis for enhancing road safety. Researchers have used **statistical analysis, machine learning, and geographic mapping** to identify high-risk zones. According to the *World Health Organization (WHO)*, over **1.3 million people die every year** in traffic-related accidents.

In Pakistan, road safety studies highlight that **over-speeding, driver negligence, and poor road conditions** are key contributing factors. Data analysis can reveal which provinces or years show the highest increase in fatalities, helping governments allocate resources and improve enforcement.

5. Dataset Description

The dataset used for this project contains traffic accident data from Pakistan and its regions over multiple years. It includes both fatal and non-fatal accidents, as well as counts of those injured or killed.

Dataset Columns:

- **Year:** The fiscal year of reported accidents (e.g., 2013–14).
- **Place:** The province or region where the accidents occurred (e.g., Punjab, Sindh, Islamabad).
- **Total_accidents:** The total number of accidents reported in that region and year.
- **Fatal_Accident:** The number of fatal accidents resulting in deaths.
- **Non-Fatal_Accidents:** The number of accidents with only injuries.
- **Killed:** The total number of people who died in traffic accidents.
- **Injured:** The total number of people who were injured.
- **Total_vehicles_involved:** The total number of vehicles involved in all reported accidents.

This dataset allows us to perform both **time-series analysis** (year-by-year trends) and **location-based analysis** (comparison between provinces).

6. Methodology

The project followed a structured approach consisting of three key stages:

1. Data Cleaning

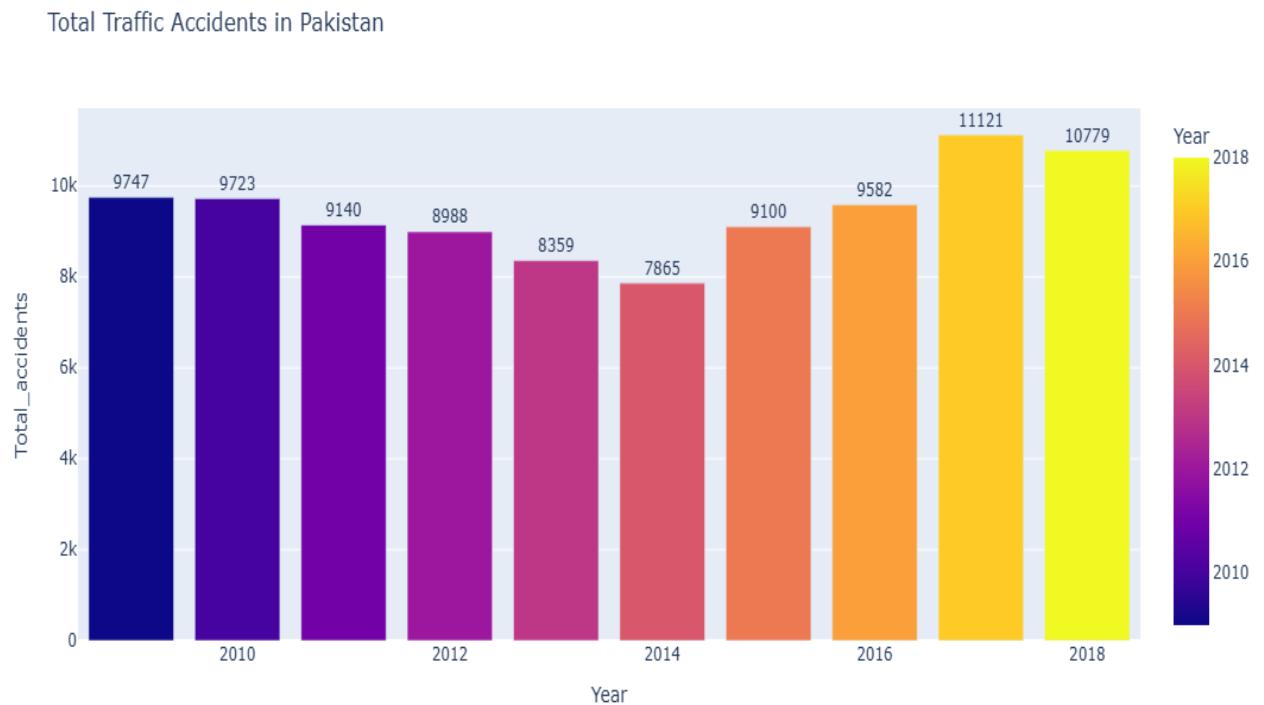
- Removed missing or duplicate rows.
- Fixed inconsistent year formats (e.g., “2013–14*” → “2013–14”).
- Ensured all numeric columns were properly converted to integer data types.
- Verified that the “Place” column contained all expected regions such as Punjab, Sindh, KPK, Balochistan, and Islamabad.

2. Data Processing

- Grouped data by **year** and **province** to compute total fatalities and injuries.
- Calculated averages and year-over-year growth percentages.
- Filtered data to focus on relevant features for visualization.

3. Data Visualization

- Created bar graphs showing the total n



umber of accidents each year.

- Created separate visualizations showing total **killed** and **injured** persons per region.
- Used **Plotly Express** for interactive graphs and **pandas** for data summarization.

7. Results and Discussion

After analysis, several key findings emerged:

1. Overall Trend:

The total number of traffic accidents fluctuated between 2008 and 2018, with the highest spike occurring between **2016–2018**.

2. Regional Comparison:

- **Punjab** recorded the highest number of accidents overall, contributing to nearly 40% of the total cases.
- **Sindh** and **KPK** followed, showing moderate but consistent accident numbers.
- **Islamabad** had fewer accidents but a higher fatality-to-accident ratio.

3. Fatal vs. Non-Fatal Accidents:

Fatal accidents made up around **35–40%** of total accidents. This ratio remained steady across most years, highlighting the need for improved emergency response.

4. People Killed and Injured:

- The total number of **people killed** increased from around **4,900** in **2008–09** to **5,948** in **2017–18**.
- The total number of **people injured** rose from about **11,000** to **14,489** in the same period.
- The year **2017–18** recorded the highest fatalities (4,829 fatal accidents, 5,948 deaths).

5. Vehicles Involved:

The total vehicles involved in accidents increased steadily, reaching over **13,000** by 2017–18 — indicating a direct correlation with the rise in vehicle ownership.

6. Geographic Insights:

- Punjab and Sindh showed a higher density of urban accidents.

- Islamabad displayed improvement in safety after 2015, likely due to better enforcement and public awareness campaigns.
- Rural areas had higher fatality rates due to delayed medical response.

8. Conclusion

The project demonstrates the power of **data analysis in improving road safety**. By identifying key trends in traffic accidents across regions and years, it becomes possible to target high-risk areas and times.

The findings suggest that Pakistan can significantly reduce fatalities by improving traffic management, law enforcement, and public education. A combination of **data-driven policy-making** and **technological integration** (e.g., smart cameras, AI-based monitoring) can help achieve this.

9. Future Work

For future improvements, the following enhancements are recommended:

- Include **weather**, **road type**, and **traffic volume** data to understand external factors.
- Use **machine learning algorithms** to predict high-risk areas.
- Create a **real-time dashboard** showing accident hotspots using live map data.
- Collaborate with traffic police departments for better data collection and reporting accuracy.
- Extend the dataset to include **driver demographics** (e.g., age, gender, license type) for deeper behavioral analysis.

10. References

1. Pakistan Bureau of Statistics (PBS) Traffic Accident Reports
2. World Health Organization (WHO) Global Road Safety Report
3. Kaggle Public Traffic Accident Datasets

4. National Highway & Motorway Police (NHMP) Annual Reports
5. Plotly and Pandas Official Documentation