AKINOLA MARY TOSIN

VEPH/20B/DA103

Traffic Accident Analysis

Outlines

Introduction
Story of data
Data splitting and Preprocessing
Pre Analysis
Post Analysis
Data Visualization and Charts
Recommendation and Observation
Conclusion

INTRODUCTION

This project focuses on analyzing traffic accident data to uncover patterns, causes, and consequences of crashes across various trafficway types and time periods. The dataset includes detailed attributes such as crash dates, weather and lighting conditions, crash types, road defects, contributing causes, injuries, and damages. The primary objective of this analysis is to identify high-risk areas and trends in crash occurrences to support data-driven recommendations for improving road safety.

Story of the Data

The dataset used in this project captures detailed records of traffic accidents, encompassing thousands of crash incidents reported over several years. Each row represents a single crash event, with attributes ranging from **temporal details** (such as crash date, hour, day of the week, and month) to **environmental factors** (like weather, lighting, and road conditions), as well as **crash characteristics** (first crash type,

damage extent, trafficway type, and traffic control devices). It also includes **contributory causes**, **intersection involvement**, and **injury details**, offering a holistic view of each incident.

This rich and diverse dataset provides a unique opportunity to analyze road safety issues from multiple angles. By leveraging the structured information, we can trace how various conditions influence crash patterns and outcomes. It helps identify where and when crashes most frequently occur, the types of roads and intersections involved, and the primary causes leading to accidents and injuries.

Ultimately, the data tells a compelling story about the realities of road safety—highlighting not just raw crash numbers but also human consequences in the form of injuries and fatalities.

Data Splitting and Preprocessing

Before analysis, the dataset was imported into Power BI and loaded into Power Query Editor for cleaning and transformation. This step was crucial to ensure consistency, accuracy, and usability of the data for effective analysis and visualization.

1. Text Formatting

Some columns contained inconsistent capitalization (e.g., values in all uppercase). To maintain a clean and readable presentation, categorical fields such as `Weather_Condition`, `Lighting_Condition`, and `Crash_Type` were reformatted using the “Capitalize Each Word” transformation.

2. Time and Date Parsing

The `Crash_Date` column contained full timestamps (e.g., “8/13/2023 12:11:00 AM”). From this, several useful time-based columns were extracted:

- Crash Month: Originally numeric (1, 2, 3…), it was converted into short month names (e.g., Jan, Feb) using a custom column with a DAX switch formula.

- Crash Day of Week and Crash Hour: These were extracted to enable trend analysis across weekdays and times of day.

3. Derived Columns

Additional columns were created to enhance analysis:

- Year: Extracted from the `Crash_Date` to support year-over-year trend comparisons.

- Month Name and Day Name: Added as human-readable time dimensions.

4. Handling Missing or Ambiguous Data

Some fields, such as `Prim_Contributory_Cause` and `Road_Defect`, contained ambiguous or non-specific values like "Unable to Determine" or "Unknown". These were retained but flagged during analysis to account for possible limitations in interpretation.

Through this preprocessing phase, the dataset was cleaned, enriched, and structured to enable efficient analysis in Power BI using both visual tools and DAX measures.

Pre-Analysis

The pre-analysis stage involved setting clear objectives and identifying the key metrics and dimensions to be explored throughout the project. This step served as a strategic foundation for guiding both the visual and analytical components of the dashboard.

Objectives

- To identify trends in traffic crashes over time.

- To evaluate crash severity across different trafficway types, intersection types, and crash types.
- To analyze the distribution and causes of injuries and fatalities.
- To determine the most frequent contributory causes of crashes.
- To assess the impact of road conditions, lighting, and weather on accident frequency and severity.

Key Metrics

- Total Crashes
- Injuries Total
- Injuries Fatal
- Injuries Incapacitating
- Crash Counts by Year, Month, and Day
- Crash Distribution by Time of Day
- Crash Type and Damage Range Analysis

Key Dimensions

- Crash Date and Time (Year, Month, Day, Hour)
- Trafficway Type

- First Crash Type
- Intersection Related
- Primary Contributory Cause
- Lighting and Weather Conditions
- Road Surface and Defects

Challenges and Limitations

- A notable number of records listed causes as "Unable to Determine" or "Unknown," limiting precise root-cause identification.
- Some data fields (like `Damage`) are recorded as text ranges (e.g., "Over $1,500") rather than numeric values, restricting detailed monetary analysis.
- Missing values in certain categories may reduce the accuracy of segment-specific insights.

Despite these limitations, the dataset provides enough structured information to extract meaningful patterns and support practical road safety recommendations.

Post-Analysis

After conducting the data exploration and visualization in Power BI, several patterns and insights emerged regarding traffic accident trends, injury distributions, and key risk factors.

Crash Trends Over Time

The data revealed fluctuations in crash volumes across the years, with **2024 recording the highest number of crashes**. Monthly and daily breakdowns showed increased crash occurrences during certain months and peak hours, highlighting periods of elevated traffic risk.

Injury and Fatality Insights

Analysis of injuries showed that **"Not Divided" trafficway types** consistently recorded high injury totals, with **angle and turning crashes** being the most prevalent crash types. Fatal injuries, though relatively low in frequency, were most often associated with **Four-Way and T Intersection** roads.

Top Contributing Causes

"**Unable to Determine**" was frequently listed as the top primary cause, suggesting either incomplete data reporting or a lack of clear evidence in many crash investigations. Other common causes included **failure to yield right of way**, **disregarding traffic signals**, and **following too closely**.

Intersection and Trafficway Analysis

The majority of crashes occurred at intersections, particularly **Y and Four-Way** intersections. These locations showed a higher density of crashes and injuries, emphasizing the need for focused safety improvements in intersection design and signaling.

Environmental Conditions

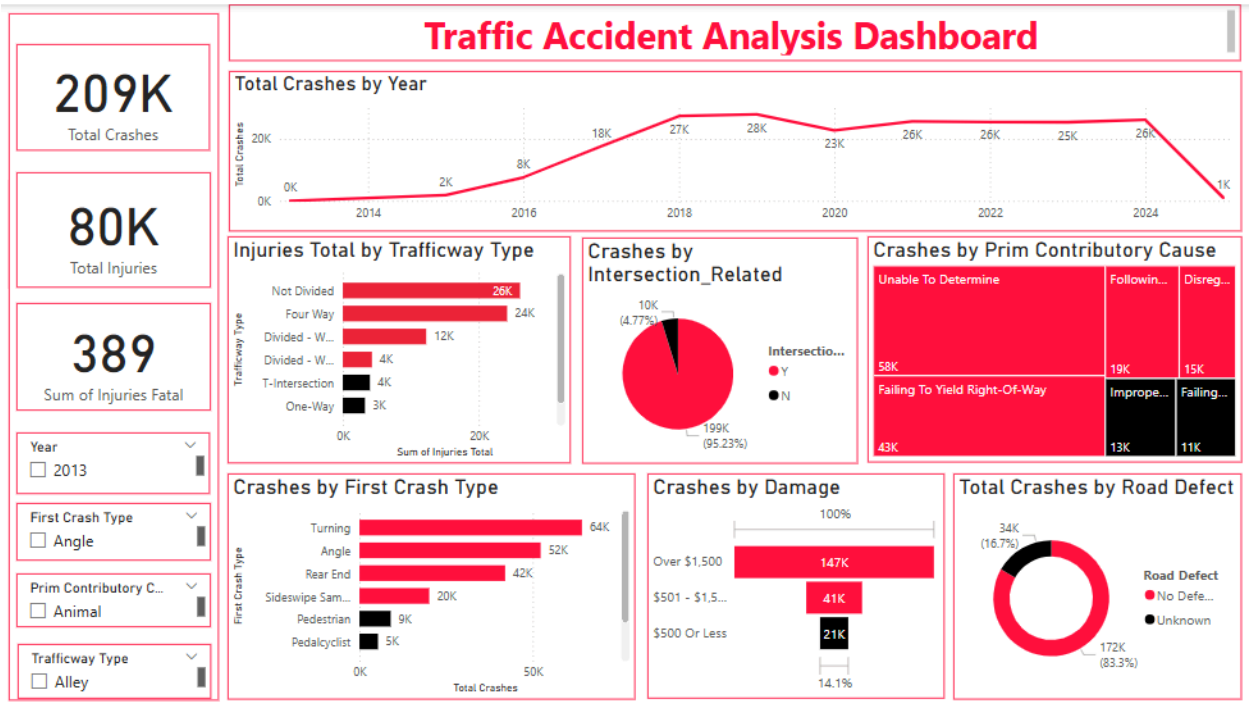
The impact of **weather and lighting conditions** was also examined. While most crashes occurred under clear and dry conditions—likely due to higher traffic volume—there were still significant incidents recorded during rain, snow, or poor lighting, indicating areas where visibility or road grip may be a factor.

Crash Severity and Damage

Most crashes resulted in **damage over $1,500**, highlighting the financial burden of traffic incidents even when injuries are minor or absent. Crash types such as rear-end and sideswipe (same direction) also appeared prominently in cases with high damage categories.

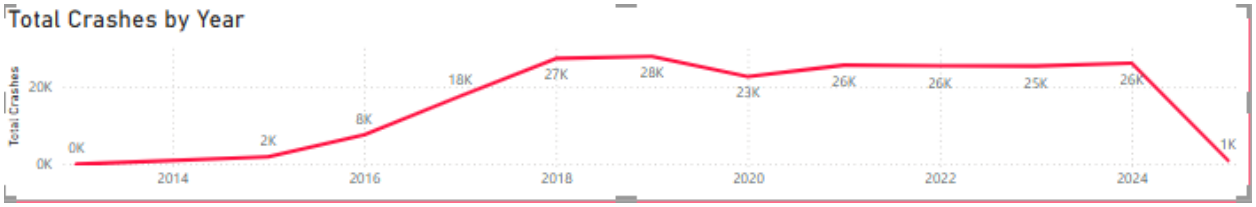
Data Visualization and Dashboard

Power BI was used to design an interactive and visually engaging dashboard that communicates key insights from the traffic accident dataset. A variety of chart types were employed to represent crash patterns, injury distributions, and other critical metrics in an intuitive and accessible way.



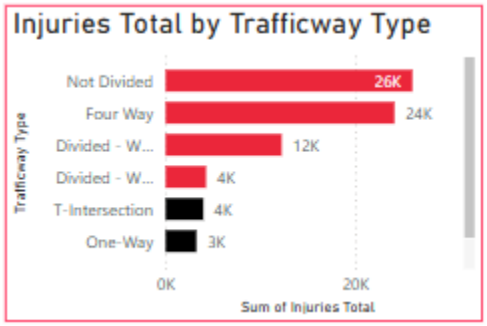
Key Visualizations

- **Total Injuries by Year**



A column chart displays the sum of injuries per year, helping to quickly identify peak years in terms of accident severity.

Injuries Total by Trafficway Type

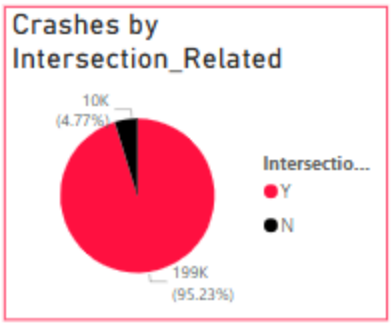


A bar chart compares the total number of injuries across different trafficway types such as Not Divided, Four Way, and T Intersection, revealing which road configurations are most accident-prone.

Injury Breakdown by Trafficway Type

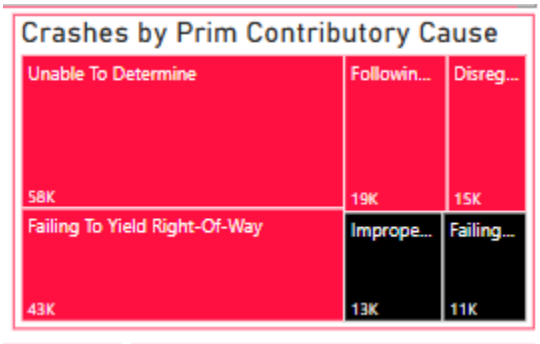
A stacked bar chart visualizes the distribution of injury types (fatal, incapacitating, non-incapacitating, not evident) across each trafficway category, highlighting where the most severe injuries occur.

Crashes by Intersection Related



A donut or pie chart was used to show the proportion of crashes that occurred at intersections, emphasizing the role of road layout in accident frequency.

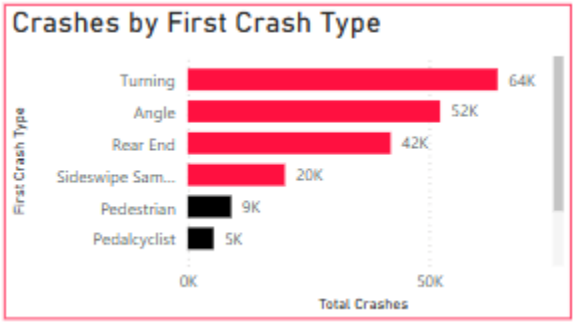
Primary Contributory Causes



A treemap shows the top causes of accidents. "Unable to Determine" leads in frequency, followed by specific driver errors like failure to yield and ignoring signals.

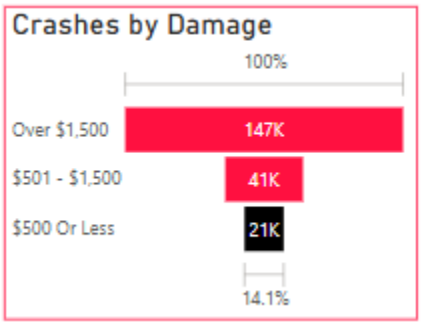
Crashes by First Crash Type

Crashes by First Crash Type

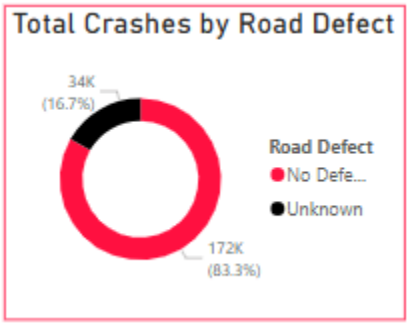


This chart outlines the initial impact type (e.g., angle, turning, rear-end), helping identify which maneuver types are most risky.

Crashes by Damage Range



A bar chart was used to show the volume of crashes by damage estimates—most commonly "Over $1,500," giving insight into economic impact.



Recommendations and Observations

Based on the analysis of the traffic accident dataset from 2018 to 2025, several critical observations and actionable recommendations emerged. These are intended to support traffic authorities, urban planners, and public safety officials in reducing crash frequency and severity.

Observations

- **High Frequency at Not Divided Trafficways**: This road type recorded the highest number of injuries, particularly in 2018. Combined with frequent intersection

involvement, this suggests a need for improved control measures in these areas.

- **"Unable to Determine" as Leading Cause**: This repeated value limits the depth of analysis into root causes. It indicates potential data collection or reporting issues at the crash scene.

- **Angle and Turning Crashes Dominate**: These crash types are the most prevalent across all years and trafficway types, indicating maneuvering and visibility challenges.

- **Most Crashes Occur in Clear Conditions**: The majority of incidents happened under clear weather and dry surface conditions, suggesting that human error plays a larger role than environmental factors.

- **Damage Costs Skewed High**: Most crashes involved damage exceeding $1,500, indicating a high financial toll even when injuries are not severe.

Recommendations

- **Install Better Traffic Control Devices at High-Risk Intersections**
  Enhanced signage, traffic lights, and roundabouts should be considered for high-crash intersections, particularly Y and Four-Way types.

- **Improve Data Collection at Crash Scenes**
  Training for field officers or automated reporting systems could help reduce the number of cases labeled “Unable to Determine,” allowing for clearer root cause analysis.

- **Target Driver Education on Turning and Angle Maneuvers**
  Educational campaigns or mandatory re-training may be beneficial, especially in

areas where such crash types are most common.

- **Implement Speed Control and Monitoring**
  Given the high impact crashes (in terms of damage), speed monitoring technologies like speed cameras should be installed, especially in straight roads and undivided trafficways.

- **Use Time-Based Interventions**
  Since crash volumes spike during certain hours and days, targeted enforcement during peak times can reduce traffic violations and collisions.

- **Prioritize Road Safety Improvements in “Not Divided” Trafficways**
  These roads need attention for lane separation, speed calming, and better visibility at junctions.

- **Leverage Power BI Dashboards for Ongoing Monitoring**
  Stakeholders can use the interactive dashboard to continuously track traffic safety metrics and quickly identify new risk trends.

Conclusion

This Power BI project provided a comprehensive analysis of traffic accident data spanning multiple years, highlighting key trends, risk factors, and areas for safety improvement. Through effective data preprocessing, visual storytelling, and the use of dynamic dashboards, the project revealed that intersection types, trafficway configurations, and certain crash types—particularly angle and turning crashes—contribute significantly to injury totals and accident severity.

The use of DAX measures and interactive filters enabled deep exploration of crash patterns by time, location, and contributing factors. The findings emphasize the critical role of human behavior, infrastructure design, and data completeness in road safety outcomes.

By translating raw crash data into actionable insights, this report supports informed decision-making for policymakers, traffic authorities, and public safety planners. Continued use of analytics tools like Power BI will be vital for monitoring traffic trends and proactively addressing emerging road safety challenges.