COMPREHENSIVE ANALYSIS OF ROAD TRAFFIC ACCIDENTS

CORETECH LABS

Chituru Alerechi
Excel Capstone Project

AGENDA

INTRODUCTION

PROBLEM STATEMENT

DATA CLEANING &PROCESSING

DESCRIPTIVE STATISTICS & CENTRAL TENDENCIES

DASHBOARD OVERVIEW

KEY INSIGHTS

RECOMMENDATION

LIMITATIONS & CONSIDERATIONS

CONCLUSION

INTRODUCTION

- CoreTech Labs is a technology company that provides smart digital solutions through software development, AI integration, and data analytics.
- As a Data Scientist intern, I was responsible for analyzing road accident data, performing statistical analysis, and developing an interactive dashboard to uncover trends and support road safety decisions.

PROBLEM STATEMENT

Road traffic accidents remain a major public safety concern, influenced by various factors such as road conditions, weather, and traffic control systems. However, the available accident data often exists in unstructured formats, making it difficult for stakeholders to extract meaningful insights or take timely action.

To address this, there is a growing need to apply data science techniques to organize, analyze, and visualize accident data. This project aims to uncover patterns, identify high-risk factors, and support data-driven decisions that can improve road safety and optimize resource allocation.

PROJECT OBJECTIVES

1. Clean and standardize accident dataset
2. Analyze accident frequency and severity
3. Identify high-risk factors
4. Create interactive dashboard
5. Provide actionable safety recommendations

DATA CLEANING & PROCESSING

1. Removed duplicate records from the dataset
2. Standardized date and time formats
3. Formatted categorical variables for consistency
4. Rounded geographic coordinates to 2 decimal places
5. Standardized the Vehicle Type column using a nested SUBSTITUTE formula
    - This grouped similar entries and simplified analysis by merging long or inconsistent labels into unified categories like 'Motorcycle', 'Van', 'Bus', and 'Car'

DESCRIPTIVE STATISTICS & CENTRAL TENDENCIES

As part of the analysis, I examined the central tendencies and distribution of key numerical variables to summarize the data and identify outliers. I calculated the mean, median, standard deviation, min, max, and skewness for variables such as:

- Number of Casualties
- Number of Vehicles
- Speed Limit

The results revealed that:

- Number of Casualties had a high positive skew (5.69), indicating that while most accidents resulted in one casualty, a few extreme cases involved significantly more.
- Number of Vehicles was moderately skewed, with most accidents involving 1–2 vehicles.
- Speed Limit showed a moderate right skew (1.14), with limits ranging from 10 to 70 km/h, and a mean of 39 km/h. This suggest that higher speed limits are less common but present in the data.

DESCRIPTIVE STATISTICS & CENTRAL TENDENCIES
CONT.

VARIABLES	MEAN	MEDIAN	STANDARD DEVIATION	MIN	MAX	SKEWNESS
Number of Casualties	1	1	0.82	1	48	5.69
Number of Vehicles	2	2	0.71	1	32	1.85
Speed Limit	39	30	14.03	10	70	1.14

DASHBOARD OVERVIEW

Accident Analysis Dashboard

Acciden...
- fatal
- Serious
- Slight

Urban_...
- Rural
- Urban

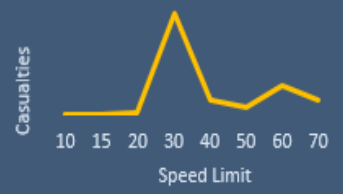
Road_T...
- Dual carriag...
- One way street
- Other
- Roundabout
- Single carria...
- Slip road

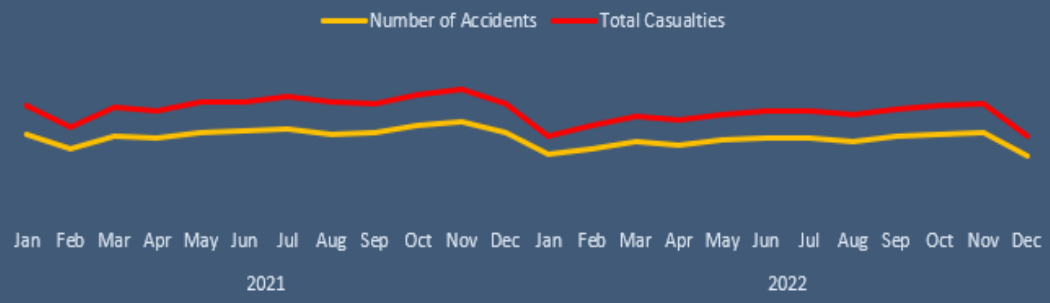
Total Accidents
308K

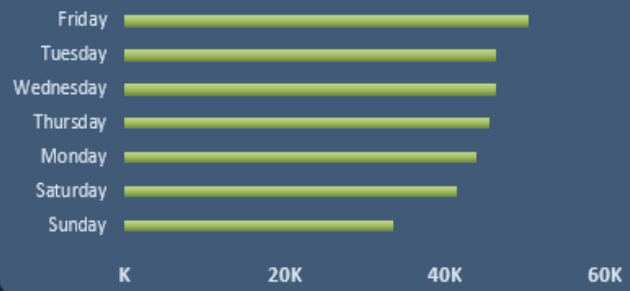
Total Casualties
418K

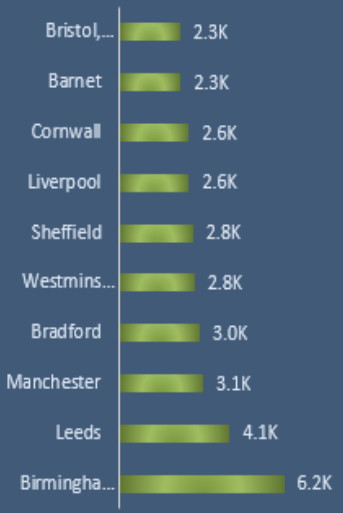
Risky Day
Friday

Accident Peak Time
5 PM

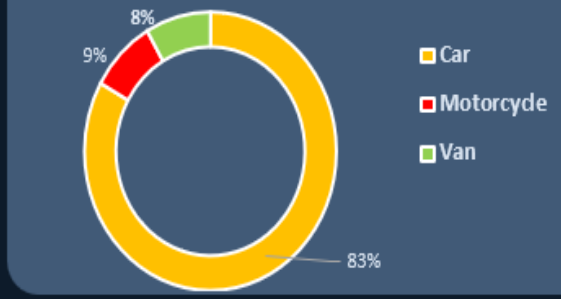
Casualties by Speed Limit


Monthly Trend of Reported


Accident Frequency by Day of the Week


Accident Hotspots


Most Accident-Prone Weather Conditions


Top 3 Vehicle Types Involved in Accidents


Most Accident-Prone Lighting Conditions


KEY INSIGHTS

Overall Impact

A total of 308K accidents were recorded, resulting in 418K casualties, highlighting the significant scale of traffic incidents over the two-year period.

Time-Based Patterns

Friday is confirmed as the most accident-prone day, with the highest frequency.

5 PM is the peak accident time, likely due to evening traffic and workday commutes.

The monthly trend shows a steady number of accidents across both 2021 and 2022, with slight dips in early and late months, possibly linked to weather or travel restrictions.

Vehicle Types Involved

Cars dominate accident involvement at 83%, while motorcycles (9%) and vans (8%) make up the rest.

This reflects typical traffic compositions and may also highlight vulnerability for smaller vehicles like motorcycles.

Environmental Conditions

More accidents (208K) occurred in “Fine, no high winds” conditions, showing that most accidents happen during good weather — likely due to higher traffic volume or driver overconfidence.

Daylight conditions account for the highest number of accidents (277K), followed by lit darkness, suggesting visibility isn’t the only contributing factor.

Road Types & Hotspots

The top accident hotspots are concentrated in urban areas, with Birmingham (5.4K) and Leeds (3.6K) leading the list.

Other high-incident districts include Manchester, Bradford, Sheffield, and Cornwall.

Speed & Severity

Most roads have lower speed limits (mean: 39 km/h; max: 70 km/h)

Casualties peak in the 30–40 km/h range, not at the highest speeds.

Suggests urban traffic, pedestrian zones, and road complexity are key risk factors

RECOMMENDATIONS

Focus Safety Campaigns Around Peak Times

Fridays at 5 PM show the highest accident frequency — suggesting a need for targeted enforcement or awareness campaigns during evening rush hours.

Prioritize Urban Zones for Intervention

Most accidents occur in urban areas with moderate speed limits (30–40 km/h).

Consider traffic calming, pedestrian zones, and infrastructure improvements in high-density districts like Birmingham and Leeds.

Don't Rely Solely on Weather as a Risk Indicator

Most accidents happen in clear daylight — suggesting that driver behavior and traffic volume, not just poor conditions, contribute to risk.

Promote defensive driving education, even in good weather.

Address Vehicle-Specific Risk

With cars making up over 80% of accidents, encourage tech adoption like automatic braking or alert systems.

Motorcycle involvement is also notable — suggesting the need for protective gear enforcement and lane-sharing policies.

Use Data to Guide Resource Allocation

Emergency response teams can be better deployed in districts with consistently high accident rates.

Invest in lighting and signage upgrades in high-casualty zones.

LIMITATIONS & CONSIDERATION

- Casualty counts may not reflect accident severity accurately (e.g., multiple injuries in one record vs. fatality in another).
- The overwhelming number of accidents in good weather raises questions about reporting bias or missing context like traffic volume.
- No fields indicate driver behavior, road signage, or traffic enforcement presence, which may be key contributors to outcomes.
- The presence of extreme outliers (e.g., 48 casualties in one accident) might skew averages or highlight data entry issues.

CONCLUSION

This project successfully applied data science techniques to transform raw, unstructured road accident data into meaningful insights through cleaning, statistical analysis, and visualization.

The interactive dashboard revealed key patterns in accident severity, timing, location, and environmental factors — highlighting critical risk zones and high-impact conditions.

These findings support data-driven decision-making for traffic safety initiatives, resource allocation, and policy planning.

Ultimately, the project demonstrates how structured analysis can drive safer roads and smarter strategies in real-world applications.

THANK YOU

Chituru Alerechi

Data Scientist Intern

@10alytics