Road Traffic Accident Severity Analytics

Accident Risk Insights

Sector: Urban Transportation & Public Safety

Team Members:

- Aaditya Yadav (2401010004)
- Aditya Pal (2401020082)
- Ishan Maheshwari (2401010194)
- Navodit Sharma (2401010294)
- Naitik Pandey (2401010287)
- Samarth Khera (2401010407)

Faculty mentor: Ayushi Vashisht & Satyaki Das

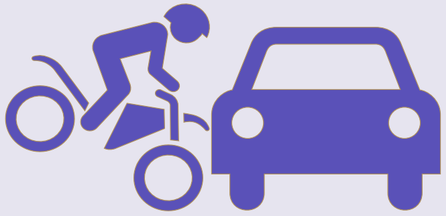
Problem Statement

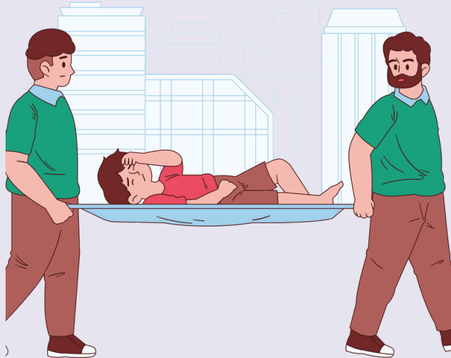
How do environmental conditions, vehicle types, and driver demographics influence the severity of accidents?

The dataset contains 12,316 road traffic accident records collected between 2017 and 2020 in Addis Ababa. The data includes key variables such as accident severity, time of occurrence, weather conditions, road surface type, vehicle category, and driver demographics.

Preliminary exploration suggests that accident severity varies significantly across different environmental and behavioral conditions. A considerable proportion of accidents result in serious or fatal injuries, indicating the need for targeted safety interventions. Patterns may exist between accident severity and factors such as night-time driving, adverse weather conditions, heavy vehicle involvement, and driver age.

The objective of this project is to identify high-risk conditions associated with severe and fatal accidents to support data-driven traffic safety decisions and resource allocation.





Data Engineering

Origins of Sources

01

Platform: Kaggle

Domain: Urban Transportation & Public Safety

Original Size: 12,000+

Cleaned Size: 12317 (final analysis set)

Time Period: Snapshot data (2017-2020)

02

1. Categorical Data
   - Missing or blank values were replaced with the mode (most frequently occurring value) or labeled as “Unknown”.
2. Numerical Data
   - Empty or missing numeric values were replaced with 0 or the median of the range.

03

Calculated key performance indicators (KPIs) from the dataset.

Examples include:

% Serious Accidents

% Fatal Accidents

% Slight Injuries

These KPIs are used for dashboard visualization and analysis.

04

1. Final analysis dataset contains 12,317 records after cleaning.
2. Added derived metrics for dashboard visualization (Scorecards, Bar Charts, etc.).
3. Dataset ready for exploratory analysis and KPI reporting.

KPI & Metrics Framework

KPI	Formula	Business Value
Total Accidents	=Count(Total Accidents)	12316
Total Casualties	=Count(Casualties)	1907
Serious Accidents	= (Number of Serious Accidents / Total Accidents) * 100	14.15%
Fatal Accidents	= (Number of Fatal Accidents / Total Accidents) * 100	1.28%
Serious Injury	= (Number of Serious Injury Accidents / Total Accidents) * 100	14.15%
Slight Injury	= (Number of Slight Injury Accidents / Total Accidents) * 100	84.56%
Most Common Cause	mode(most common cause)	No Distancing

Key Insights

Overall Accident Trends

- Total accidents and casualties have decreased
  - Accidents: –12,316
  - Casualties: –19,067
- Serious accidents % down by 14.15%
- Fatal accidents % down by 1.28%

This suggests overall road safety improvements compared to the previous period.

Injury Severity Distribution

- Slight injuries dominate (84.56%)
- Serious injuries ≈ 14%
- Fatal injuries ≈ 1–1.5%

👉 Most accidents are non-fatal, but the volume of slight injuries is very high — indicating frequent minor crashes.

Most Common Cause of Accidents

- "No Distancing" (Failure to maintain safe distance) is the leading cause.

🔎 This highlights:

1. Poor defensive driving
2. Traffic congestion effects
3. Need for awareness campaigns on safe following distance

Time-Based Patterns

- Slight injuries are consistently high across all days.
- Fatal accidents show noticeable spikes toward weekend days (especially Saturday).
- Midweek has relatively stable accident severity levels.

📌 Interpretation:

Weekend mobility and possibly higher speeds may increase fatal risk.

Vehicle Type Insights:

Certain vehicle categories (e.g., automobiles and motorcycles) show higher involvement in injuries.

Public and commercial transport contribute significantly to serious injury share.

👉 Targeted enforcement and training by vehicle category could reduce impact.

Road Surface Impact

Dry road surfaces account for the majority of accidents.

This suggests:

Accidents are driven more by human behavior than weather conditions.

Infrastructure alone is not the primary issue.

ADVANCED ANALYSIS

Segmentation Analysis

What was done:
Grouped accidents by driving experience, vehicle type, and day of week to identify high-risk segments.

New understanding:
Mid-experience drivers and weekend periods show higher severity risk. Risk is not limited to new drivers — overconfidence may be a key factor.

Forecasting Analysis

What was done:
Trend and predictive modeling of accident severity using factors like driving experience, vehicle type, day, and cause.

New understanding:
Although total accidents are decreasing, slight injuries remain very high (84%+), indicating persistent minor crash frequency. Behavioral factors (especially "No Distancing") strongly influence severity outcomes.

Root Cause Analysis

What was done:
Analyzed accident causes and linked them to injury severity levels.

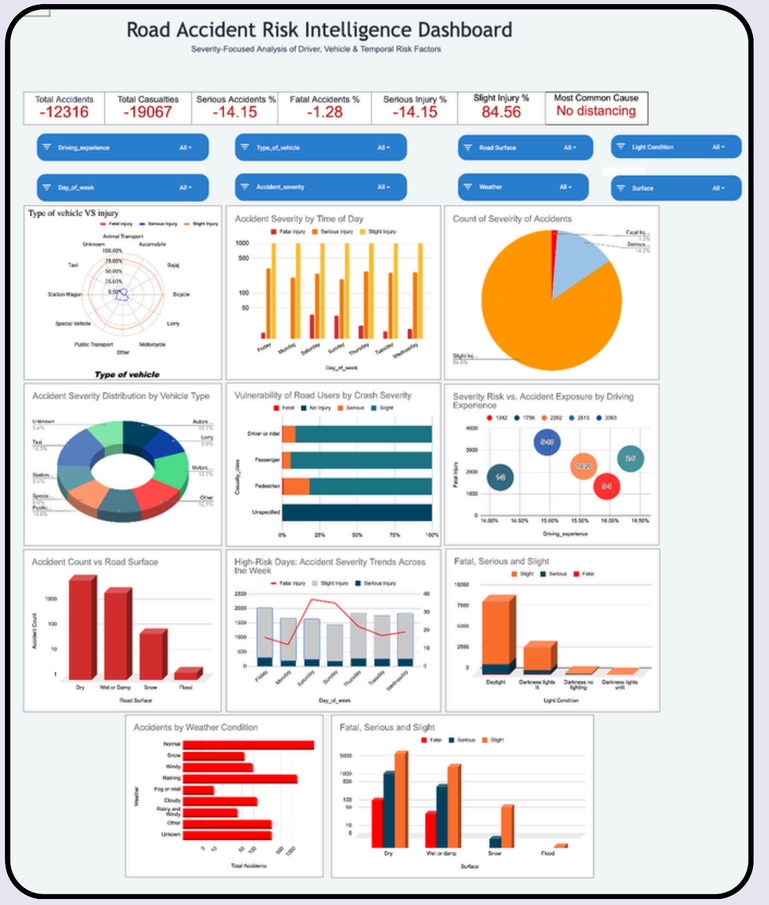
New understanding:
"No Distancing" is the dominant cause, and most accidents occur on dry roads — confirming that human behavior, not road condition, is the primary driver of accidents.

Dashboard Walkthrough

Accident Severity Analysis by Vehicle, Driver & Time

Executive View:

KPIs: Total Accidents(12316), Total Casualties(19067),
%Serious Accidents(14.15%), %Fatal Accidents(1.28%),
%Serious Injuries(14.15%), %Slight Injuries(84.56%),
Most Common Cause(No Distancing)



Operational View:

- Type of vehicle VS injury
- Accident Severity by Time of Day
- Count of Seveirity of Accidents
- Accident Severity Distribution by Vehicle Type
- Vulnerability of Road Users by Crash Severity
- Severity Risk vs. Accident Exposure by Driving Experience
- Accident Count vs Road Surface
- High-Risk Days: Accident Severity Trends Across the Week
- Fatal, Serious and Slight
- Accidents by Weather Condition

RECOMMENDATIONS

Implement strict enforcement and awareness campaigns to reduce “No Distancing” violations.

Increase traffic monitoring and patrol presence during high-risk weekend periods.

Introduce mandatory refresher safety training for mid-experience drivers.

Apply targeted safety regulations and technology adoption for high-risk vehicle categories

IMPACT & VALUE

1. **Cost Reduction:**

Targeted behavioral enforcement can significantly reduce accident-related medical, insurance, and infrastructure repair costs.

2. **Time Efficiency:**

Fewer accidents mean reduced traffic congestion and faster commute times, improving overall productivity.

3. **Operational Efficiency:**

Data-driven enforcement allows authorities to focus on high-risk drivers, vehicles, and time periods instead of broad, costly campaigns.

4. **Revenue & Economic Stability:**

Lower accident severity reduces economic losses from fatalities and workforce injuries, preserving human capital and long-term economic output.

LIMITATIONS & NEXT STEPS

Limitations

1. **Limited Granularity:**

Data lacks detailed location, speed, and traffic density information, restricting deeper spatial and behavioral analysis.

2. **Potential Reporting Bias:**

Minor accidents may be underreported, affecting severity distribution accuracy.

3. **No Exposure Data:**

Absence of vehicle population or driver mileage data limits true risk-rate calculation.

4. **Limited Time Horizon:**

Short-term data reduces accuracy of long-term forecasting and seasonal pattern analysis.

🔎 Next Steps

1. Integrate geospatial and traffic volume data for hotspot analysis.
2. Incorporate driver demographic and vehicle registration data to calculate true risk rates.
3. Apply advanced machine learning models for more accurate severity prediction.
4. Develop a real-time monitoring dashboard for continuous risk tracking.