1. Data Visualization and Analytics

Road Traffic Accident Severity Analytics

Team Members:

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

Institute:

Newton School of Technology

Faculty: Aayushi Vashishth

2. Problem Statement(Executive Summary)

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.

Approach

We analyzed a real-world Road Traffic Accident dataset using Google Sheets. The project focused on:

- Data cleaning and preprocessing
- Handling missing values and inconsistent categorical entries
- Standardizing categorical variables (e.g., time of day, vehicle type, weather conditions)
- Creating calculated KPIs such as Fatality Rate and Severe Injury Rate
- Building pivot tables and visual dashboards
- Identifying relationships between accident severity and contributing factors

Key Insights

- **Accident Severity Increases During Night Hours**

  Accidents occurring during evening and night-time periods show a higher proportion of serious and fatal injuries compared to daytime incidents.

- **Adverse Weather Elevates Risk**

  Rainy and poor visibility conditions are associated with an increased percentage of severe accidents due to reduced control and braking efficiency.
- **Heavy Vehicles Contribute to Higher Fatality Rates**

  Accidents involving trucks and buses demonstrate a greater likelihood of fatal outcomes compared to smaller vehicles.

- **Younger Drivers Show Higher Involvement Rates**

  Younger age groups appear more frequently in accident records, suggesting potential behavioral and experience-related risk factors.

- **Poor Road Conditions Amplify Severity**

  Accidents occurring on damaged or slippery road surfaces show a higher probability of resulting in serious injuries.

Key Recommendations

- **Increase traffic enforcement** and monitoring during high-risk time periods (especially evening and night hours).
- **Improve road infrastructure** maintenance, particularly in areas prone to slippery or uneven surfaces.
- **Introduce stricter monitoring** and compliance checks for heavy vehicles.
- **Launch awareness campaigns** targeting high-risk driver demographics.
- **Deploy emergency response** resources strategically in high-severity zones and time windows.

3. Sector & Business Context

a. Sector Overview

The road transportation and public safety sector plays a critical role in urban development and economic activity. In rapidly growing cities like Addis Ababa, increasing vehicle density, mixed traffic systems, infrastructure limitations, and varying enforcement levels contribute to frequent road traffic accidents.

Government agencies and traffic authorities rely on accident data to design policies, improve infrastructure, allocate enforcement personnel, and reduce fatalities. Data-driven decision-making is essential to enhance road safety and minimize public health and economic losses caused by severe accidents.

b. Current Challenges

- High number of serious and fatal road accidents
- Limited identification of high-risk time periods and conditions
- Resource constraints in traffic enforcement and emergency response
- Infrastructure limitations (road surface quality, lighting, signage)

- Difficulty in targeting specific high-risk driver or vehicle categories

c. Why This Problem Was Chosen

This project was selected to analyze real-world traffic accident data and identify the key factors contributing to severe and fatal crashes. Understanding these relationships helps simulate how traffic authorities can make evidence-based decisions regarding enforcement deployment, infrastructure improvement, and public safety interventions.

The analysis supports practical decision-making aimed at reducing accident severity, improving urban road safety, and optimizing resource allocation.

---

4. Problem Statement & Objectives

a. Problem Statement: To identify the key environmental, vehicle-related, and driver-related factors that influence the severity of road traffic accidents.

b. Objectives
   - Evaluate the impact of time of day and weather conditions on accident severity.
   - Analyze the relationship between vehicle type and fatal accident rates.
   - Examine how driver demographics (such as age) affect accident outcomes.

- Develop an interactive Google Sheets dashboard to support data-driven traffic safety decisions.

c. Project Scope
   - Accident-level analysis using the Road Traffic Accident dataset (12,316 records).
   - Data cleaning, transformation, and categorization performed in Google Sheets.
   - Severity measured using accident severity categories (Slight Injury, Serious Injury, Fatal Injury)
   - KPI creation including Fatality Rate and Severe Injury Rate.
   - Visualization using pivot tables and charts for decision support.
   -

d. Success Criteria
   - Cleaned and structured dataset with standardized variables.
   - Interactive dashboard highlighting high-risk conditions.
   - Clearly defined KPIs (e.g., Fatality Rate, High-Risk Time Period).
   - Actionable insights for traffic authorities and safety planners.

---

5. Data Description

a. Dataset Source
   - **Platform:** Kaggle
   - **Domain:** Road Traffic Accidents (Urban Transportation & Public Safety)
   - **Geographic Coverage:** Addis Ababa
   - **Time Period:** 2017–2020

- **Approximate Size:** 12,316 accident records with 30+ variables

b. Data Structure

Each row represents **one recorded road traffic accident**.

The dataset captures environmental conditions, vehicle characteristics, and driver-related factors associated with each accident.

c. Key Columns (Cleaned Dataset)

- **Accident_Severity** – Slight Injury / Serious Injury / Fatal Injury
- **Time_of_Day** – Morning / Afternoon / Evening / Night
- **Weather_Conditions** – Clear / Rainy / Foggy / etc.
- **Road_Surface_Type** – Dry / Wet / Damaged / etc.
- **Vehicle_Type** – Car / Bus / Truck / Motorcycle / etc.
- **Driver_Age_Band** – Age group category of the driver
- **Light_Conditions** – Daylight / Darkness / No lighting

d. Data Limitations

- Data is limited to a single city (Addis Ababa), which may restrict generalizability.
- Dataset represents recorded accidents only (unreported incidents are excluded).
- No exact geographic coordinates for hotspot mapping (if not included).
- No detailed behavioral variables (e.g., exact speed, intoxication level).
- Snapshot dataset — does not include long-term historical trend comparisons beyond 2017–2020.

6. Data Cleaning & Preparation

a. Missing Values Handling

- Blank or missing categorical fields (e.g., weather conditions, road surface type) were standardized and replaced with **"Unknown"** where appropriate.
- Missing numerical values (e.g., driver age where applicable) were reviewed and categorized into grouped age bands to reduce distortion.
- Inconsistent or incomplete severity entries were validated and corrected based on available labels.

b. Data Type Standardization

- Accident severity and categorical variables were formatted consistently for pivot analysis.
- Driver age fields were converted into standardized **age bands**.
- Ensured uniform text formatting across categories (e.g., vehicle types, weather conditions).

c. Transformations

- Standardized time-of-day categories (e.g., grouping hours into Morning, Afternoon, Evening, Night).
- Cleaned and normalized categorical labels to remove spelling inconsistencies and extra spaces.
- Structured selected variables to support KPI calculations (Fatality Rate, Severe Injury Rate).

d. Feature Engineering

- Created **Severity Indicator fields** (e.g., binary flag for Fatal Injury).

- Created calculated KPIs such as:
    - Fatality Rate
    - Severe Injury Rate
- Grouped similar vehicle categories for clearer analysis (e.g., heavy vehicles vs light vehicles).
- Reordered columns for efficient pivot table creation.
- Removed irrelevant or redundant columns not required for severity analysis.

e. Outlier Treatment

- Reviewed severity distribution for anomalies.
- Verified that all severity values fall within valid predefined categories.
- Checked driver age entries for unrealistic values and grouped them appropriately.
- Examined unusual frequency spikes in certain categories to ensure they were not data-entry errors.

d. Assumptions

- Missing categorical values labeled as “Unknown” do not significantly distort severity patterns.
- Reported accident records accurately reflect actual road incidents.
- Severity classification (Slight, Serious, Fatal) is consistently recorded across all records.

- External factors not included in the dataset (e.g., exact speed, intoxication level) may influence severity but are outside the project scope.

---

7. KPI & Metric Framework

KPI Definition: Key Performance Indicators (KPIs) are quantifiable, outcome-based metrics used to evaluate how effectively traffic authorities can identify and reduce high-risk accident conditions.

1. **KPI 1**: Fatal Accident Rate (%)
    a. Formula: ***= (Number of Fatal Accidents / Total Accidents) * 100***
    b. Why it matters: Measures the proportion of accidents resulting in death. It reflects the overall severity level of road safety conditions.
    c. Objective Mapping: Directly supports the objective of identifying high-risk conditions contributing to fatal crashes.
2. **KPI 2**: Serious Injury Rate (%)
    a. Formula: ***= (Number of Serious Injury Accidents / Total Accidents) * 100***
    b. Why it matters: Indicates the burden of severe injuries on healthcare systems and emergency response.
    c. Objective Mapping: Helps analyze environmental and behavioral factors influencing serious accident outcomes.
3. **KPI 3**: Slight Injury Percentage (%)
    a. Formula: ***= (Number of Slight Injury Accidents / Total Accidents) * 100***
    b. Why it matters: Provides baseline comparison to understand severity distribution.

c. Objective Mapping: Supports overall severity pattern analysis.

4. **KPI 4**: High-Risk Time Period
    a. Formula (Conceptual): Time of day with highest Fatal + Serious %
    b. Why it matters: Identifies when enforcement and monitoring should be increased.
    c. Objective Mapping: Supports data-driven allocation of traffic personnel.
5. **KPI 5**: High-Risk Vehicle Category
    a. Formula (Conceptual): Vehicle type with highest Fatal %
    b. Why it matters: Helps identify vehicle categories contributing disproportionately to fatal outcomes.
    c. Objective Mapping:Supports regulatory decisions and safety monitoring for heavy vehicles.

---

8. Exploratory Data Analysis (EDA)

a. Trend Analysis

- Initial exploration of the dataset reveals that accident severity is not evenly distributed. Slight injuries account for the majority of recorded accidents, while serious and fatal injuries represent a smaller but critical proportion.
- Severity distribution indicates that although fatal accidents are relatively low in percentage, their impact is significant from a public safety perspective.
- Time-based and environmental trends suggest variation in severity levels across different driving conditions.

b. Observations

- Slight injuries constitute the largest share of accidents (approximately 80%+ of total cases).
- Serious injuries represent a moderate percentage, indicating substantial healthcare burden.
- Fatal accidents account for a smaller percentage but are strategically important for policy intervention.
- Night-time and evening periods show relatively higher proportions of serious and fatal accidents compared to daytime.
- Heavy vehicles appear more frequently in high-severity accident categories.
- Poor weather and unfavorable road surface conditions correlate with increased accident severity.

c. Comparison Analysis

Comparison analysis was conducted across key accident risk indicators.

a. Time of Day vs Severity
    - Accidents during evening and night hours show a higher proportion of serious and fatal injuries compared to morning and afternoon periods.

b. Vehicle Type vs Severity
    - Heavy vehicles (e.g., trucks, buses) show a greater association with fatal outcomes compared to smaller vehicles such as bicycles or motorcycles.

c. Driving Experience vs Severity
    - Drivers with lower driving experience show relatively higher involvement in severe accident categories, suggesting behavioral or skill-related risk factors.

d. Distribution Analysis

Distribution analysis helps understand how accident severity and risk factors are spread across categories.

a. Severity Distribution
    - The dataset is highly concentrated in Slight Injury cases, with smaller proportions of Serious and Fatal injuries.

b. Vehicle Type Distribution
    - Certain vehicle categories dominate accident counts, but severity levels vary significantly across type.

e. Correlation Analysis

Correlation relationships were analyzed between key risk variables.

a. **Time of Day vs Severity**

Night-time conditions show a positive association with higher accident severity.

b. **Vehicle Type vs Fatality Rate** Heavy vehicles demonstrate a stronger relationship with fatal outcomes compared to light vehicles.

f. Written Insights with Charts

Based on dashboard visualizations:

- Slight injuries dominate overall accident counts, but serious and fatal injuries are concentrated under specific high-risk conditions.
- Night-time driving significantly increases the probability of severe outcomes.
- Heavy vehicles contribute disproportionately to fatal accidents.
- Poor road conditions and adverse weather amplify accident severity.
- Certain driver experience groups exhibit elevated risk patterns, suggesting need for targeted training or regulation.

9. **Advanced Analysis**

a. Forecasting (Applicability)

Since the dataset represents a multi-year snapshot (2017–2020) rather than structured time-series data with continuous intervals, long-term predictive forecasting was not performed.

However, comparative severity analysis across time-of-day and environmental conditions was used to estimate short-term risk exposure patterns. This helps simulate how accident severity may fluctuate under specific conditions (e.g., night-time or rainy weather).

b. Segmentation

Accidents were segmented based on key risk drivers:

Segment 1: High Severity + Night-Time

- Fatal and serious accidents occurring during evening/night.
- High-priority enforcement window.
- Requires increased monitoring and patrol deployment.

Segment 2: High Severity + Heavy Vehicles

- Accidents involving trucks and buses with fatal outcomes.
- Indicates structural and impact-related risk.
- Candidate for stricter regulatory checks.

Segment 3: Low Driving Experience + Serious/Fatal Severity

- Drivers with limited experience involved in high-severity crashes.
- Suggests training or licensing intervention opportunities.

Segment 4: Adverse Weather + Serious/Fatal Severity

- Severe accidents occurring under rainy or poor visibility conditions.
- Indicates infrastructure and precautionary gaps.

c. Root Cause

Low-severity vs high-severity accident patterns were examined to identify contributing factors.

Primary Contributors to Severe Accidents:

- Night-time driving conditions
- Heavy vehicle involvement
- Adverse weather and poor road surface
- Limited driver experience
- Poor lighting conditions

These factors appear repeatedly across fatal and serious accident categories.

d. Risk / Anomaly Analysis

Identified Risks

- Concentration of fatal accidents in specific time windows → Enforcement risk.
- Heavy vehicle involvement in high-severity crashes → Regulatory risk.
- Higher vulnerability of pedestrians and passengers → Public safety risk.
- Road surface and lighting issues → Infrastructure risk.

Anomalies

- Certain vehicle categories show high accident frequency but relatively low severity, suggesting exposure-based rather than risk-based occurrence.
- Some mid-experience driver groups show unexpectedly high fatal involvement, indicating possible behavioral or contextual factors beyond inexperience.

e. Scenario Analysis

Scenario 1: Increased Night-Time Enforcement

If traffic monitoring is intensified during high-risk evening/night hours:

→ Potential reduction in serious and fatal accident proportion.

Scenario 2: Heavy Vehicle Regulation & Monitoring

Stricter inspection and compliance enforcement for heavy vehicles:
→ Potential decrease in fatal crash contribution from this segment.

Scenario 3: Infrastructure & Visibility Improvement

Improved road lighting and surface maintenance in high-risk areas:

→ Reduction in severity during adverse conditions.

---

10. Dashboard Design

The dashboard is designed to help traffic authorities and safety planners analyze key risk factors influencing accident severity, including driver behavior, vehicle type, time-based exposure, and environmental conditions.

It focuses on identifying high-risk segments to support data-driven enforcement, infrastructure, and policy decisions.

a. Filters

Interactive filters are provided for:

- **Driving Experience**

- **Type of Vehicle**
- **Day of Week**
- **Accident Severity**

These filters allow dynamic segmentation of accident patterns across risk dimensions.

b. KPI Summary

c. **Total Accidents:** 12,316
d. **Total Casualties:** 19,067
e. **Serious Accident %:** ~14%
f. **Fatal Accident %:** ~1%
g. **Serious Injury %:** ~14%
h. **Slight Injury %:** ~84%
i. **Most Common Cause:** No Distancing

The distribution highlights that while slight injuries dominate numerically, serious and fatal accidents represent critical safety concerns.

Visual Components

1. Radar Chart – Vehicle Type vs Injury Severity

Displays how different vehicle categories contribute to slight, serious, and fatal injuries.
→ Helps identify high-risk vehicle segments (e.g., heavy vehicles).

2. Column Chart – Accident Severity by Time of Day

Compares severity levels across days/time periods.

→ Identifies high-risk temporal windows.

3. Pie Chart – Severity Distribution

Shows overall composition of Slight, Serious, and Fatal accidents.

→ Provides high-level severity overview.

4. Donut Chart – Severity Distribution by Vehicle Type
Breaks down accident severity within each vehicle category.

→ Supports regulatory focus on specific vehicle groups.

---

11. Insights Summary

1. **Accident Severity is Highly Concentrated in Slight Injuries** Slight injuries account for the majority of recorded accidents (~80%+), while serious and fatal injuries represent a smaller but critical portion requiring targeted intervention.
2. 🌙 **Night-Time Driving Increases Severity Risk**
   Accidents occurring during evening and night hours show a higher proportion of serious and fatal outcomes compared to daytime incidents.
3. 🚛 **Heavy Vehicles Contribute Disproportionately to Fatal Accidents** Trucks and buses are more frequently associated with fatal and serious injury cases compared to smaller vehicles, indicating higher impact severity.

4. 🛣️ **Road Surface Conditions Influence Accident Occurrence**

   Accident counts are higher on certain road surface types, suggesting infrastructure quality plays a role in crash frequency and severity.

5. 👥 **Vulnerable Road Users Face Greater Severity Exposure**

   Pedestrians and passengers demonstrate higher vulnerability in severe crashes compared to drivers.

---

12. Recommendations

1. Increase Night-Time Traffic Enforcement

- **Insight:** Higher proportion of serious and fatal accidents occur during evening and night hours.
- **Action:** Deploy additional traffic patrols, speed monitoring, and sobriety checks during high-risk time windows.
- **Impact:** Reduction in severe and fatal accidents during peak-risk hours.
- **Feasibility:** High (requires operational scheduling adjustments

2. Strengthen Heavy Vehicle Monitoring

- **Insight:** Heavy vehicles are disproportionately involved in fatal accidents.
  Action: Conduct stricter inspections, enforce speed limits, and implement compliance checks for trucks and buses.

Impact: Lower fatality contribution from high-impact vehicle categories.

Feasibility: Medium–High (requires regulatory coordination).

3. Improve Road Infrastructure & Lighting

- **Insight:** Poor road surface and low visibility conditions increase severity risk.

  Action: Prioritize maintenance of damaged roads and improve street lighting in high-risk areas.

  Impact: Reduced accident severity under adverse conditions.

  Feasibility: Medium (infrastructure investment required).

4.Implement Targeted Driver Training Programs

- Insight: Certain driving experience groups show higher involvement in severe crashes.
- Action: Introduce awareness campaigns, defensive driving programs, and stricter licensing evaluations for high-risk groups.
- Impact: Improved driving behavior and reduced severity probability.
- Feasibility: Medium–High.

5. Focus on Vulnerable Road User Protection

- **Insight:** Pedestrians and passengers are highly vulnerable in severe crashes.
- **Action:** Improve pedestrian crossings, enforce right-of-way rules, and install traffic calming measures.
- **Impact:** Enhanced protection for vulnerable road users.
- **Feasibility:** Medium.

---

13. Impact Estimation

Safety Impact

- Increasing enforcement during high-risk night-time hours could potentially reduce serious and fatal accident rates.

- Strengthening heavy vehicle monitoring may significantly decrease fatal crash contribution from high-impact vehicle categories.

- Infrastructure improvements (road surface and lighting) can lower accident severity under adverse conditions.

Cost Reduction Impact

- Targeted, data-driven enforcement reduces unnecessary resource deployment.
- Preventive infrastructure maintenance reduces long-term public healthcare and emergency response costs.
- Early intervention in high-risk segments lowers economic loss from severe accidents.

Operational Efficiency Improvement

- Risk-based resource allocation improves traffic personnel deployment efficiency.
- Focused monitoring of high-risk vehicle and driver categories enhances enforcement effectiveness.
- Data-driven decision-making reduces guesswork in safety planning.

Public Service Improvement

- Improved road safety increases public confidence in traffic management systems.
- Reduction in severe accidents improves overall community well-being.

- Better infrastructure and awareness campaigns enhance long-term road user safety.

Risk Reduction

- Identification of high-risk time windows helps prevent fatal crashes proactively.
- Monitoring behavioral risk factors (e.g., no distancing) reduces preventable accidents.
- Concentrating interventions on vulnerable road users lowers fatality exposure.

---

14. Limitations

Data Issues

- Dataset represents a multi-year snapshot (2017–2020), not a continuous real-time dataset.
- Some categorical fields required cleaning and standardization before analysis.
- Certain variables lack granular detail (e.g., exact speed, intoxication level).

- Data limited to a single city (Addis Ababa), reducing generalizability.

Assumption Risks

- Missing categorical values were treated as “Unknown” without external validation.

- Severity classifications were assumed to be consistently recorded across all records.

- Driving experience and environmental factors were assumed to directly influence severity without controlling for exposure rates.

What Cannot Be Concluded

- Exact causal relationships between variables (analysis is observational, not experimental).

- Real-time accident prediction accuracy without additional time-series data.

- Impact of unrecorded factors such as driver behavior specifics or vehicle speed.

- National-level traffic safety conclusions beyond the dataset’s geographic scope.

15. Future Scope

- Time-Series Severity Trend Analysis
- Geospatial Accident Hotspot Mapping
- Predictive Modeling for Accident Severity.
- Behavioral Risk Factor Integration
- Emergency Response Impact Analysis
- Policy Effectiveness Evaluation Framework

16. Conclusion

This project analyzed real-world road traffic accident data using Google Sheets to identify the key factors influencing accident severity. The analysis demonstrates that temporal conditions (especially night-time driving), vehicle type (particularly heavy vehicles), road surface conditions, and driver experience significantly impact the likelihood of serious and fatal accidents.

The dashboard provides a structured, data-driven view of accident patterns, enabling traffic authorities to identify high-risk segments and prioritize targeted interventions. By leveraging these insights, decision-makers can optimize enforcement deployment, improve infrastructure planning, and implement behavioral safety programs.

Applying the recommended strategies has the potential to reduce severe and fatal accidents, improve public safety outcomes, and enhance the overall effectiveness of traffic management systems.

---

17. Appendix

Data Dictionary (Summary)

Column Name	Description
Accident_Severity	Classification of accident outcome (Slight Injury, Serious Injury, Fatal Injury).
Time_of_Day	Time period when the accident occurred

	(Morning, Afternoon, Evening, Night).
Day_of_Week	Day on which the accident occurred.
Weather_Conditions	Weather condition at the time of the accident (Clear, Rainy, Foggy, etc.).
Road_Surface_Type	Condition of the road surface (Dry, Wet, Damaged, etc.).
Light_Conditions	Lighting situation during the accident (Daylight, Darkness with lighting, Darkness without lighting).
Type_of_Vehicle	Category of vehicle involved (Car, Bus, Truck, Motorcycle, etc.).

Driving_Experience	Experience level of the driver involved (categorized into bands).

18. Contribution Matrix

Team Member	Dataset & Sourcing	Cleaning	KPI & Analysis	Dashboard	Report	PPT	Overall Role
Samarth Khera	✔️	✔️	✔️	✔️		✔️	Strategy Lead
Ishan Maheshwari	✔️	✔️	✔️	✔️	✔️		Project Lead
Navodit Sharma	✔️	✔️	✔️	✔️			Dashboard Lead

Aditya Pal	✔️	✔️	✔️	✔️		✔️	Analysis Lead
Aaditya Yadav	✔️	✔️	✔️	✔️		✔️	Data Lead
Naitik Pandey	✔️		✔️			✔️	PPT & Quality Lead

Declaration: We confirm that the above contribution details are accurate and verifiable through version history and submitted artifacts.