**Overview & Key points**

|  |  |  |
| --- | --- | --- |
| **Chapter** | **Title** | **Key Points Covered** |
| Summary | Overview of the Study | Overview of accident data analysis (2015-2017) in Kolkata. Focus on accident patterns, and time of occurrence. Includes video analytics, simulations, and an interactive interface for road safety insights. |
| Chapter 1 | Data Collection and Analysis | * Data from official records (2015-2017), categorized by area, victim type, vehicle type, and time. * Manual collection of location coordinates using Google Maps. * Dataset enriched with coordinates, junction types, road widths, and time slots. * Junctions categorized into Link Roads, 4-Way Intersections, T-Junctions, Y-Junctions, and Circular Cross section * Road widths measured manually. * Challenges: Ambiguous locations, similar names, and time-intensive manual work. * QGIS Visualisation |
| Chapter 2 | Computer Vision based Analysis of traffic Video | * YOLOv9 is used for object detection. * The ByteTrack algorithm tracks vehicles and pedestrians. * The system analyzes uploaded traffic videos to generate key statistics. * It detects and tracks potential collision situations. * Further analysis like vehicle trajectories using trajectory clustering techniques are also done. |
| Chapter 3 | Road safety Analysis through Simulations | * Simulation of accident-prone locations using PTV Vissim for realistic traffic modeling. * Variation of key parameters: Vehicle speed, relative flow, and traffic density. * Conflict analysis using SSAM to identify Rear-End, Lane-Change, and Crossing Conflicts. * Conflict data utilized for ML-based accident risk prediction. * Challenges: Manual road network creation, behavioral model assumptions, and high computational load. just in few lines |
| Chapter 4 | Machine Learning based Modelling | * Regression model using RandomForestRegressor for traffic conflict prediction * Data preprocessing with 14 input features (e.g., speed, relative flow, total vehicles) and 4 output features (Crossing, Rear End, Lane Change, Total Conflicts). * Handling missing values using SimpleImputer with mean strategy. * Model evaluation using MSE (10470), MAE (57), and R² (0.75). * Visualization of actual vs. predicted values for each conflict type. * Future work: Hyperparameter tuning, cross-validation, error handling, and incorporating Graph Neural Networks (GNNs) for dynamic traffic flow modeling. * Potential GNN models include STGNN, ST-ResNet, and STDN to enhance prediction accuracy and account for spatial-temporal correlations. |
| Chapter 5 | Interface Development | * Streamlit based application * 3 major sections  1. Home (Video Tracking) 2. Simulation 3. GIS Software |

**Summary & Introduction**

This report analyses road network safety using traffic accident data from 2015 to 2017 in Kolkata. The data, originally in Word files with text-based location descriptions, was manually processed to extract coordinates, junction types, and road widths using Google Maps. This enriched dataset helped identify accident hotspots and understand patterns related to road structure and time of occurrence.

The analysis extends further with video analytics to monitor real-time traffic behaviour, simulations to model accident scenarios, and an interactive interface for visualizing results. The goal is to provide actionable insights for improving road safety, while highlighting the challenges of manual data processing and the potential for automation in future studies.

Road accidents are a leading cause of fatalities and injuries worldwide, with urban areas like Kolkata experiencing a significant share of such incidents. This study focuses on analysing road accident data in Kolkata for the years 2015 to 2017. The dataset, obtained from official records, includes detailed information about accident locations, victim types, vehicle types, and time of occurrence. By examining this data, the study aims to identify high-risk areas and contributing factors, providing a foundation for improving road safety in the city.

**Objectives:**

The primary objectives of this study are:

* **Identify Accident Patterns:** Analyse spatial and temporal trends in accident data to identify recurring patterns and hotspots.
* **Understand Contributing Factors:** Investigate the role of road design, junction types, vehicle types in causing accidents.

To achieve these objectives, the study uses a combination of methods, including manual data collection to get accurate location coordinates and road details, video analytics to observe traffic behaviour and detect patterns, simulations to create accident scenarios based on vehicle type, speed, road width, and junction type - helping find accident hotspots and test safety solutions - and an interactive interface to show accident data, video insights, and simulation results. This interface helps policymakers easily explore accident hotspots and analysis outcomes.

.

**Chapter 1: Analysis of Traffic Accidents in Kolkata in Recent Years**

**1.1 Source of Raw Data**

The accident data used in this analysis was sourced from official Word files provided by relevant authorities. These files contained detailed records of traffic accidents spanning the years 2015, 2016, and 2017. The data was structured on a month-by-month basis and divided into several key categories to provide a comprehensive overview of road incidents.

The primary types of data collected included:

* **Traffic Guard-wise breakdown:** Monthly records of fatal and non-fatal accident cases categorized by different Traffic Guard areas.
* **Division-wise cases:** Data showing fatal and non-fatal cases across various traffic divisions, organized monthly for all three years.
* **Victim type analysis:** A breakdown of accident cases based on victim type (e.g., pedestrians, drivers, passengers).
* **Road user statistics:** Information on people killed or injured, categorized by vehicle type involved (e.g., cars, two-wheelers, trucks).
* **Time-wise accident data:** Details on the time of occurrence (e.g., time of day) and the number of people involved in both fatal and non-fatal accidents.
* **Traffic Post Guard & Time-wise fatal statement:** A detailed, month-by-month breakdown of fatal cases by Traffic Post Guard areas, including the exact location descriptions of each accident.

Since the original Word files provided text-based location descriptions (e.g., "Near XYZ Junction," "Opposite ABC Building," or "Crossing of Two Roads") without geographic coordinates, latitude and longitude coordinates for these locations were manually determined using Google Maps. This manual approach was necessary due to the inaccuracy and ambiguity of automated location detection methods.

**1.2 Location Processing – Determining Latitude and Longitude**

The raw data included detailed accident location descriptions but didn’t have latitude and longitude coordinates.

The image below is an excerpt from the original Word file, which contains accident data categorized by Traffic Guard, location, vehicle type, road user type, and time of occurrence. However, the data only provides location descriptions without geographic coordinates.

To enhance the analysis, each location was manually processed using **Google Maps** to extract precise latitude and longitude coordinates. This allowed for more detailed spatial analysis, including identifying junction types and estimating road widths. The manual approach was necessary due to inconsistencies or incomplete location details that automated tools couldn’t accurately interpret.

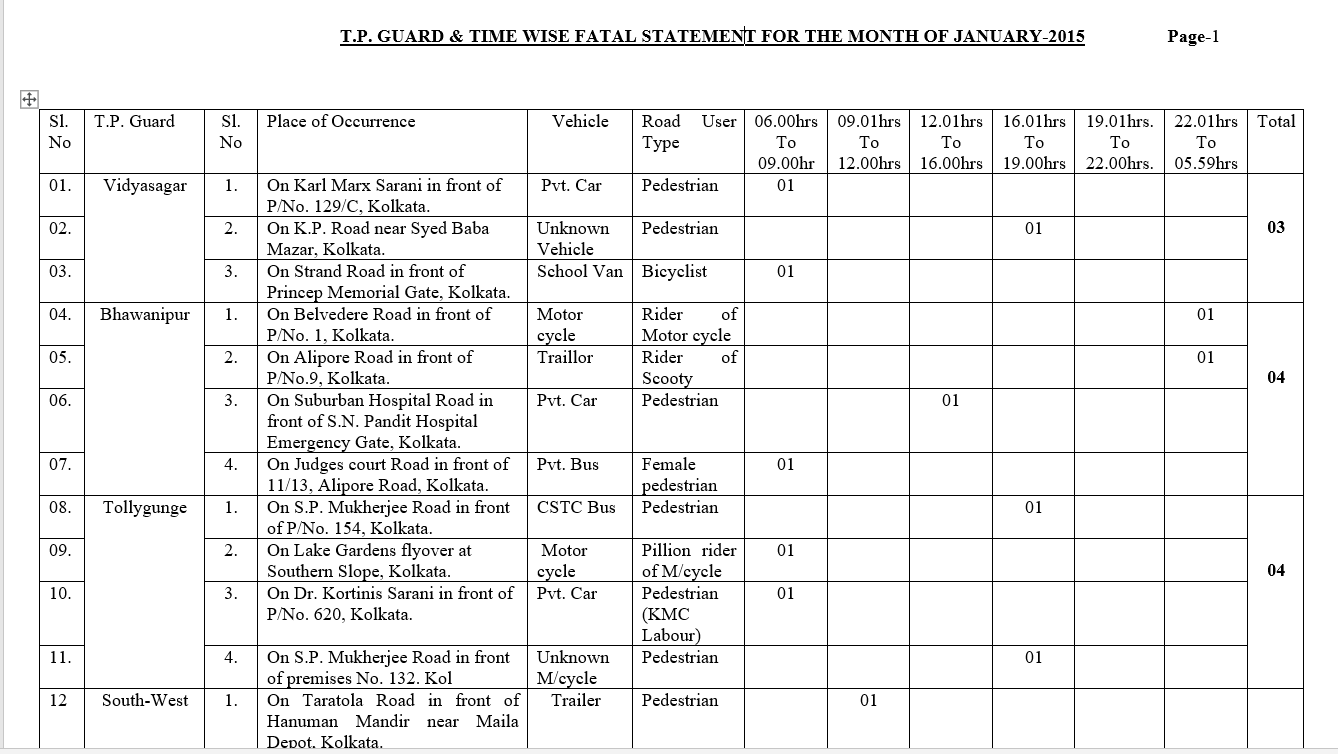


Figure 1: Screenshot of word file with Accident Location Details

The data presented in the above screenshot outlines the occurrences of fatal accidents involving various road users in Kolkata during January 2015. Similar data is available in Word files for all 12 months across 2015, 2016, and 2017, and has been structured systematically to enable a comprehensive analysis over the three-year period.

The steps followed were:

* **Extracted location descriptions** from the Word files (e.g., "Near XYZ Junction," "Opposite ABC Building," or "Crossing of Two Roads").
* **Searched for each location** on Google Maps to determine the precise latitude and longitude.
* **Manually verified the coordinates** to prevent errors, especially for locations with similar names or multiple entries (e.g., different crossings named).
* **Performed this process manually** because automated tools often failed to interpret the location descriptions accurately.

**1.3 Processed Dataset with Coordinates and Road Details**

The image below showcases the finalized dataset after extracting information from the Word files and enriching it with additional location-based details. Each entry now includes:

* **Latitude and Longitude:** Manually gathered using Google Maps for precise geolocation.
* **Junction Type**
* **Road Width**
* **Time of Occurrence:** Grouped into time slots for time-based analysis.

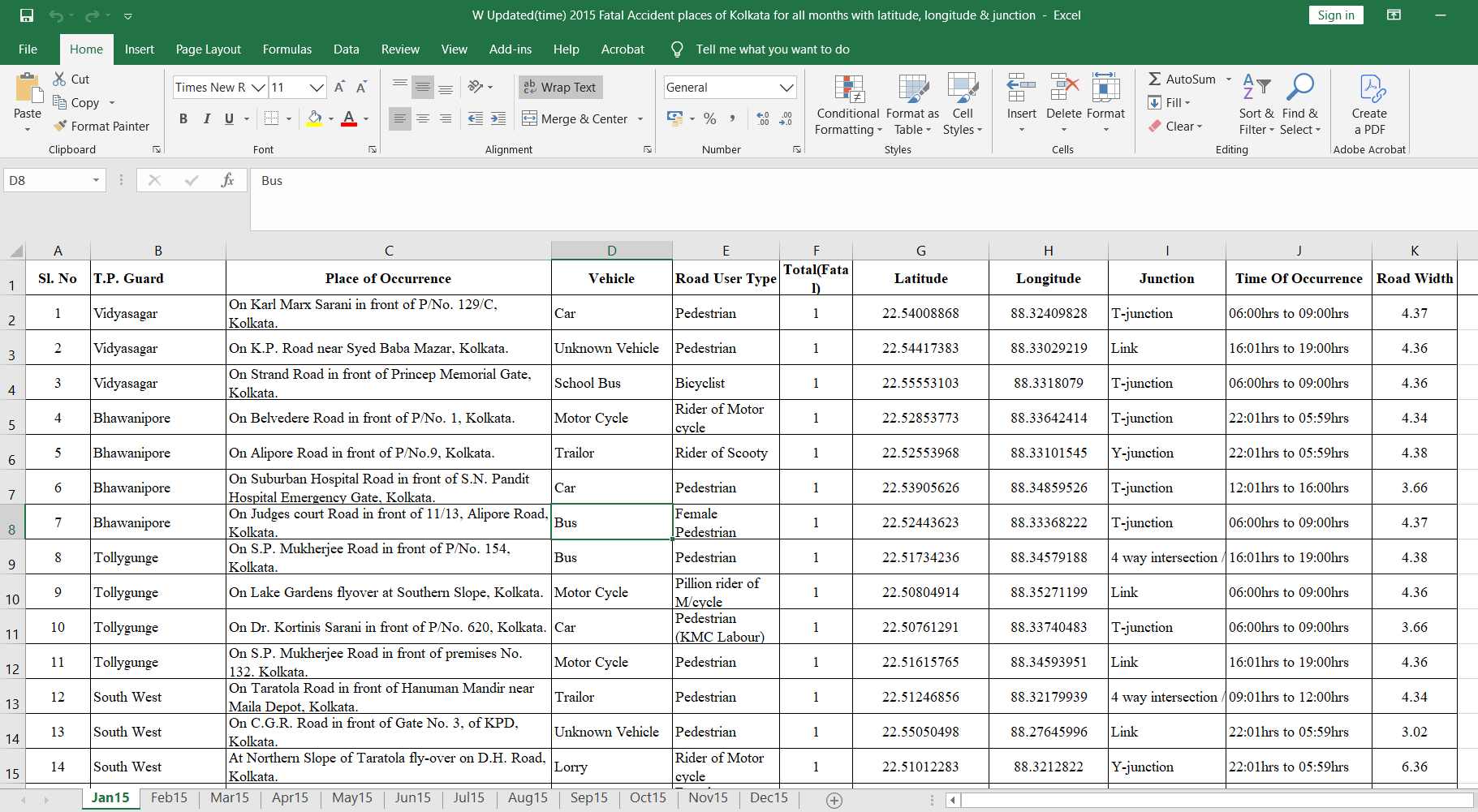


Figure 2: Screenshot of Processed Dataset with Coordinates and Road Details

This improved dataset forms the base for detailed analysis. It helps identify accident-prone areas, create visual maps of hotspots, and find connections between accident frequency and factors like road width, junction type, and time of occurrence. The added coordinates and road details give a clearer picture of accident patterns and the reasons behind them.

* 1. **Junction Type Classification**

After mapping the accident locations, each site was examined to identify the type of road junction present. This was done manually using Google Maps' satellite view and street-level imagery. Each location was categorized into one of the following types:

* **Link Road:** A straight, continuous section of road without intersections.
* **4-Way Intersection:** A crossing where two roads intersect, creating four possible directions.
* **T-Junction:** A road that ends perpendicularly at another road, forming a "T" shape.
* **Y-Junction:** A fork where a road splits into two branches, creating a "Y" shape.
* **Circular Intersection (Roundabout):** A circular intersection where vehicles travel around a central island.

This classification aimed to evaluate how different junction types influence the frequency and severity of accidents.

* 1. **Road Width Measurement**

For a more detailed analysis, road width at each accident site was manually measured using **Google Maps' distance measurement tool**.

* 1. **Challenges and Limitations**
* **Challenges** 
  + Ambiguous location descriptions (e.g., "Near XYZ Junction" could refer to multiple junctions).
  + Similar names for several places (for example, several entries for "Main Road Junction").
* **Limitations**
  + The process was performed manually because automated tools often failed to interpret the location descriptions accurately.
  + Time-consuming nature of manual verification.

**2. Traffic Guard-wise Fatal and Nonfatal Accident Data of Kolkata (2015-2017)**

This gives a summary of traffic accidents in Kolkata in 2015, 2016, and 2017 that are broken down by combine fatal and non-fatal cases for different T.P. Guards.

|  |  |  |
| --- | --- | --- |
| **Traffic Guard** | **Total Cases** | **Total Persons** |
| South West | 793 | 641 |
| Vidyasagar | 722 | 712 |
| South | 705 | 633 |
| Ultadanga | 690 | 640 |
| Bhowanipur | 672 | 548 |
| Shyambazar | 649 | 612 |
| Jorabagan | 635 | 502 |
| East | 619 | 503 |
| Tiljola | 566 | 507 |
| Tollygunge | 543 | 508 |
| Beliaghata | 529 | 475 |
| Sealdah | 523 | 503 |
| Purba Jadavpur | 451 | 372 |
| Thakurpukur | 435 | 406 |
| Head Quarters | 415 | 369 |

The following table outlines the top **15** accident-prone regions, ranked by total cases and total persons affected:

Table 1: Traffic Guard-Wise Distribution of Fatal and Non-Fatal Accident Cases in Kolkata (2015-2017)

* 1. **Traffic Guards with the Highest Number of Injuries:**
* **Vidyasagar**:
  + **712 persons injured** across **722 cases**, making it the most hazardous Traffic Guard.
  + The high number of injuries suggests significant traffic congestion, possibly involving both personal vehicles and public transport.
* **South West**:
  + **641 persons injured** from **793 cases**, the second-highest in terms of casualties.
  + This area likely experiences industrial or commercial traffic, including heavy vehicles like lorries and trailers, contributing to accident severity.
* **Ultadanga**:
  + **640 persons injured** with **690 cases**, indicating a high rate of severe accidents.
  + The mix of public and private vehicles suggests that intersections or road conditions may play a role in the high accident rate.

**2.2 Traffic Guards with Fewer Injuries and Cases:**

Significantly fewer accidents and injuries were reported by certain traffic guards, which could be a sign of improved traffic control or reduced traffic volumes.

* **Metiabruz**:
  + **Only 72 persons injured** across **91 cases**, making it the lowest among all analysed Traffic Guards.
  + This suggests that either the traffic volume is low, or existing safety measures are effective.
* **Garia** and **Jadavpur**:
  + **Garia** reported **189 injuries** from **199 cases**, while **Jadavpur** had **203 injuries** across **225 cases**.
  + These areas could serve as examples of effective traffic management, possibly involving well-maintained roads or lower congestion levels.

**2.3 Correlation between Total Cases and Injuries:**

**The** top three Traffic Guards (Vidyasagar, South West, and Ultadanga) exhibit both high case counts and high injury numbers, suggesting that increased traffic volume correlates with higher accident severity.

* **Total Accident Cases Across All Traffic Guards:**  
  There were **11,582** accident cases reported across all Traffic Guards during the period.
* **Total Persons Involved Across All Traffic Guards:**  
  A total of **10,228** persons were involved in these accidents, including both fatalities and injuries.
  1. **Key Observations:**
* **Vidyasagar** and **South West** are the most critical areas requiring focused interventions to improve road safety and reduce fatalities and injuries.
* **Metiabruz** and other areas with fewer accidents could provide valuable insights into successful traffic management strategies.

**3. Division-Wise and Police Station-Wise Accident Analysis for Kolkata (2015-2017)**

This section highlights critical insights derived from the accident data categorized by divisions, police stations, and monthly trends.

**3.1** **Division-Wise Accident Totals:**

The accident cases and affected persons vary significantly across different divisions:

|  |  |  |
| --- | --- | --- |
| **Division** | **Total Cases** | **Total Persons** |
| South | 2855 | 2533 |
| SED (South East Division) | 1579 | 1378 |
| Central | 1526 | 1283 |
| SSD (South Suburban Division) | 1382 | 1242 |
| Eastern Suburban | 1321 | 1224 |
| SWD (South West Division) | 1134 | 943 |
| North | 1002 | 928 |
| Port | 778 | 697 |

Table 2 **:Division-Wise Accident Totals**

* **South Division** has the highest number of total cases (2855) and affected persons (2533), indicating a major hotspot for traffic accidents.
* **SED (South East Division)** and **Central Division** also report high accident counts, suggesting heavy traffic flow and potential risks.
* **Port Division** reports the lowest number of cases (778) and persons affected (697), indicating relatively safer conditions.

**3.2 Police Station Hotspots:**

According to the analysis, the following police stations have the most accident cases:

|  |  |  |
| --- | --- | --- |
| **Police Station** | **Total Cases** | **Total Persons** |
| Hastings | 664 | 628 |
| P/Maidan | 609 | 540 |
| Behala | 415 | 305 |
| Alipore | 366 | 298 |
| Hare Street | 330 | 284 |

Table 3: Police **Station Hotspots**

* **Hastings** and **P/Maidan** are the top accident-prone areas, with more than 600 cases and significant casualties, suggesting the need for enhanced traffic management.
* **Behala** and **Alipore** also show high accident involvement, indicating heavy vehicular movement or congested areas.
* **Metiabruz** reports the least cases (10) and affected persons (9), highlighting a relatively low accident rate.
  1. **Monthly Accident Trend Analysis:**
     1. **Monthly Accident Cases:**

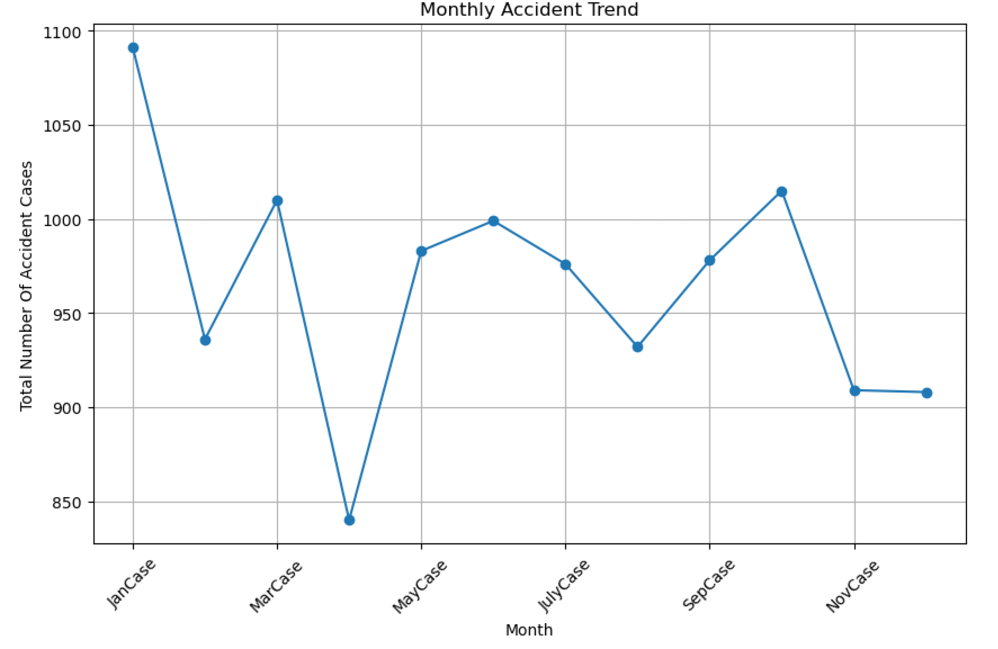
****

Figure :Monthly Distribution of Traffic Accident Cases (2015–2017)

The line graph displays fluctuations in accident cases throughout the year, with notable peaks in **January**, **March**, **June** and **October**.

* + 1. **Monthly Affected Persons:**

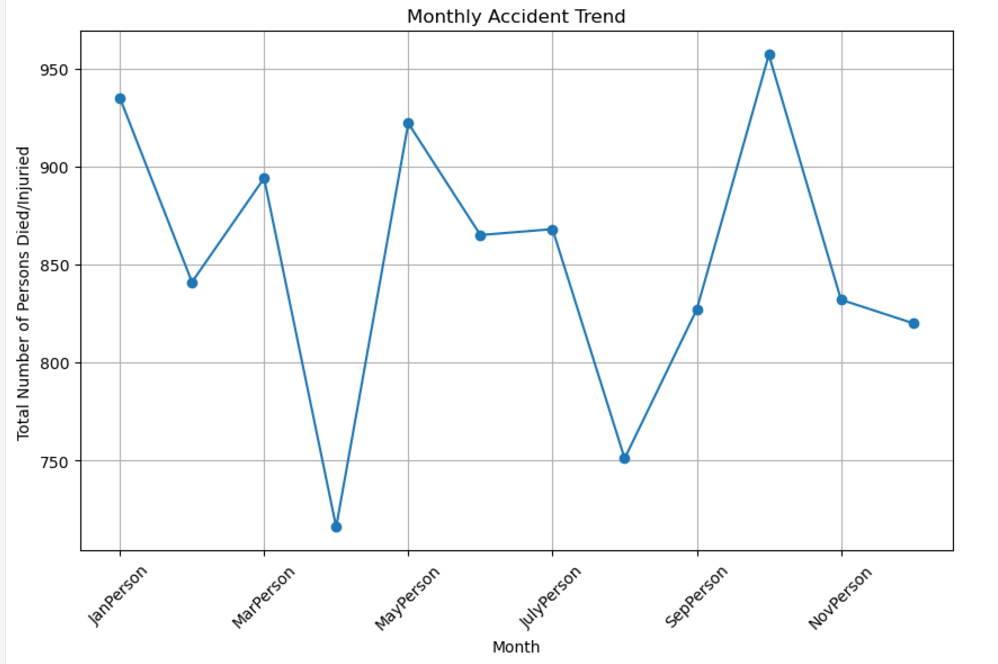
****

Figure :Monthly Distribution of Affected Persons in Traffic Accidents (2015–2017)

The trend for persons injured or killed reveals significant peaks in **January**, **May**, and **October**, suggesting these months experience heightened accident severity.

* + 1. **Victim-Type Injury Analysis (2015-2017)**

The data reveals critical insights into how different victim groups are affected by traffic accidents, emphasizing the need for targeted safety measures:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Victim Type** | **Fatal Male** | **Fatal Female** | **Grievous Injury Male** | **Grievous Injury Female** | **Minor Injury Male** | **Minor Injury Female** |
| Pedestrian | 430 | 149 | 1681 | 513 | 1060 | 390 |
| Passenger | 71 | 16 | 580 | 205 | 643 | 361 |
| Occupant | 23 | 4 | 9 | 2 | 29 | 6 |
| Two-Wheeler Rider | 215 | 6 | 988 | 26 | 867 | 43 |
| Pillion Rider | 65 | 23 | 215 | 65 | 169 | 62 |
| Motor Vehicle Driver | 45 | 0 | 190 | 1 | 354 | 8 |
| |  | | --- | | Non-Motorized |  |  | | --- | |  | | 56 | 2 | 191 | 11 | 161 | 13 |
| Police Persons | 4 | 0 | 59 | 0 | 114 | 2 |
| Other | 35 | 14 | 20 | 2 | 30 | 0 |
| **Total** | **944** | **214** | **3933** | **825** | **3427** | **885** |

Table 4: Victim Type-Wise Distribution of Fatal and Injury Cases

* **Total Victims Overview (2015-2017)**
* **Total Male Victims:** 8,304
* **Total Female Victims:** 1,924
* **Key insights:**
* **Pedestrians** remain the most vulnerable group, accounting for **430 male** and **149** fatalities, with a high number of grievous and minor injuries. This underscores the urgent need for improved pedestrian safety measures such as improved crosswalks, pedestrian signals, and road awareness campaigns.
* **Two-wheeler riders** face significant risks, contributing to **215 male** and **6 female** fatalities, highlighting the necessity of helmet enforcement and better infrastructure.
* **Pillion riders** and **passengers** also show notable injury counts, emphasizing the importance of promoting helmet usage for pillions and enhancing passenger safety measures in vehicles.
* **Motor vehicle drivers** and **non-motorized users** have fewer fatalities but still face many injuries, showing the need for safer roads and better enforcement of traffic rules.
* **Police personnel** experienced **4 fatalities** and **59 grievous injuries**, stressing the importance of protective gear and safer duty practices.
* **The Other category** reflects **49 fatalities** and a range of injuries, indicating the need for comprehensive safety strategies covering all road user types.
  + 1. **Traffic Guard and Vehicle-wise Fatal Accident Analysis (2015-2017)**

This study offers a detailed analysis of traffic accidents by vehicle type and Traffic Police (T.P.) Guard regions. The information highlights the vehicle classifications and most affected areas across a number of places in Kolkata. The following are the main conclusions derived from the data

**5.1 Top T.P. Guards by Accident Volume**

The data indicates that certain T.P. Guards experienced a higher number of accidents involving various vehicles. Here are the top contributors:

**1.Shyambazar:**

* + **Bus**: 30 accidents
  + **Unknown Vehicle**: 17 accidents
  + **Lorry**: 13 accidents
  + **Motor Cycle**: 11 accidents

**2.Sealdah:**

* + **Bus**: 29 accidents
  + **Car**: 6 accidents
  + **Lorry**: 5 accidents

**3.South West:**

* + **Lorry**: 19 accidents
  + **Unknown Vehicle**: 17 accidents
  + **Trailor**: 14 accidents

These regions show high accident rates, with buses, unknown vehicles, and heavy vehicles (lorry, trailors) being major contributors.

* 1. **Vehicle Types with Highest Accident Frequency**

Certain vehicle types are involved in accidents more frequently across multiple T.P. Guards:

* **Buses**: A dominant contributor to accidents in almost every region (e.g., 30 in Shyambazar,29 in Sealdah, 22 in Ultadanga, 18 in Head Quarter).
* **Motorcycles**: Recorded high accident counts in areas like Regent Park (10), Vidyasagar (9), and Thakurpukur (10).
* **Unknown Vehicles**: Significant in areas like Vidyasagar (16), South West (17), and East (11), indicating unidentifiable or unregistered vehicles' involvement.
  1. **Regions with the Highest Heavy Vehicle Accidents**

Heavy vehicles like lorries, trailors, and trucks are frequently involved in accidents:

* **South West**: Lorry (19), Trailor (14), Tanker (5)
* **Vidyasagar**: Lorry (10), Trailor (7)
* **Jorabagan**: Lorry (16), Truck (5)

These regions may experience issues due to industrial traffic.

* 1. **T.P. Guards with Notable Vehicle Trends**

Every T.P. Guard has distinct patterns of vehicle involvement:

* **Beliaghata**: Motorcycles (12 accidents) are the most involved, followed by buses (8) and auto-rickshaws (7), indicating a mix of personal and public transport issues.
* **Head Quarter**: Buses dominate with 18 accidents, followed by Mini Buses (7) and Taxis (5), highlighting heavy public transport use in the area.
* **Ultadanga**: Buses (22) and Cars (15) lead, showing a mix of public and private vehicle involvement.
  1. **Key Observations and recommendations**

1. **High-Risk Areas:**

* **Shyambazar**, **Sealdah**, **South West**, and **Vidyasagar** are the most accident-prone.
* These areas need targeted traffic control measures.

1. **High-Risk Vehicles:**

* Buses and Motorcycles are consistently involved in accidents.
* Better management of public transportation and stricter laws governing motorcycle safety may help lessen this.

1. **Unknown Vehicle Involvement:**

* High counts in multiple regions suggest the need for better monitoring and registration enforcement.

1. **Impact from Heavy Vehicles:**

* Industrial traffic is probably the cause of the high frequency of truck and trailer accidents, which points to problems with heavy vehicle regulation.

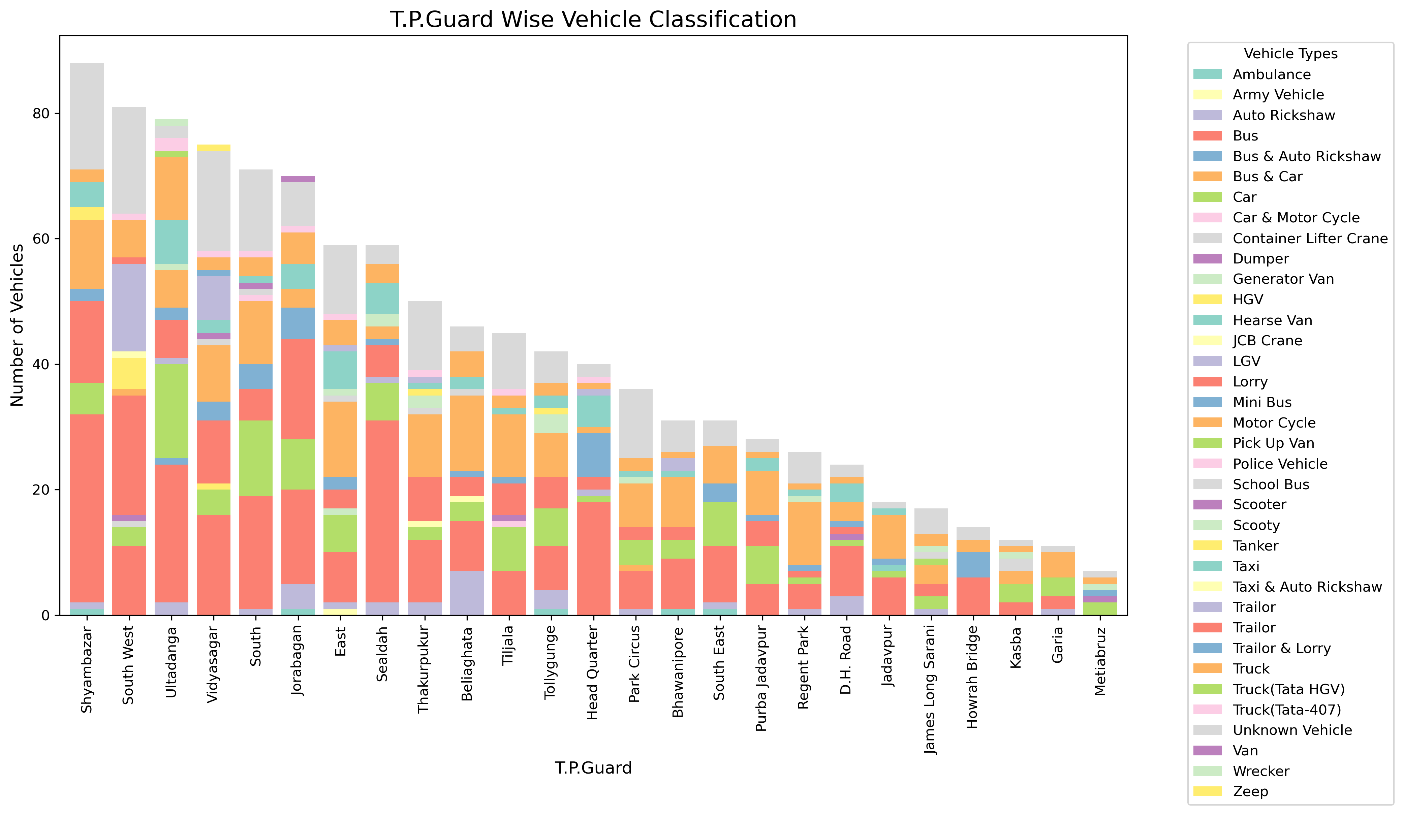


Figure 5: Graphical Representation of T.P. Guard-Wise Vehicle Classification (2015-2017)

The bar chart provides a detailed breakdown of the types of vehicles involved in accidents across different T.P. Guards from 2015 to 2017. Each segment of the stacked bars represents a specific vehicle type, allowing for visual comparison of vehicle distribution.

 **Dominance of Specific Vehicle Types:**

* Certain T.P. Guards have a large concentration of particular vehicle types, suggesting regional trends in vehicle engagement.

 **T.P. Guards with Highest Accidents:**

* The T.P. Guards with the tallest bars indicate the highest total number of accidents, showing key hotspots.

 **Vehicle Diversity:**

* The variation in colour bands within each bar demonstrates the diversity of vehicle types involved across different locations.
  + 1. **Timing Insights for Fatal Traffic Accidents (2015–2017):**

The data reveals significant insights into the timing of traffic accidents over the years 2015 to 2017:

* 1. **Peak Accident Times:**
  + The time slot **21:01 hrs to 23:59 hrs** has the highest number of accidents (152 cases), indicating that late-night hours pose the greatest risk.
  + Evening commutes (**18:01 hrs to 21:00 hrs**) are also risky, with 137 cases, highlighting the dangers during post-work traffic.
  1. **Morning and Afternoon Accident Trends:**
  + Morning peak hours **(09:01 hrs to 12:00 hrs)** reports 134 cases, while **(06:00 hrs to 09:00 hrs)** accounts for 116 cases, reflecting risks during heavy traffic flow.
  + Afternoon hours **(15.00 hrs to 18.00 hrs)** also have a notable count of 116 cases, indicating increased traffic level.
  1. **Night and Early Morning Accidents:**
  + Significant accidents occur even during **00:00 hrs to 03:00 hrs** (114 cases), possibly due to fatigue or low visibility.
  1. **Less Risky Time Periods:**
  + Time slots like **03:01 hrs to 05:59 hrs** (79 cases) and **16:01 hrs to 19:00 hrs** (17 cases) show fewer accidents, suggesting lower traffic volumes.

**Table Format:**

|  |  |
| --- | --- |
| Time of Occurrence | Number of Accidents |
| 21:01 hrs to 23:59 hrs | 152 |
| 18:01 hrs to 21:00 hrs | 137 |
| 09:01 hrs to 12:00 hrs | 134 |
| 06:00 hrs to 09:00 hrs | 116 |
| 15:01 hrs to 18:00 hrs | 116 |
| 00:00 hrs to 03:00 hrs | 114 |
| 12:01 hrs to 15:00 hrs | 110 |
| 03:01 hrs to 05:59 hrs | 79 |
| 12:01 hrs to 16:00 hrs | 28 |
| 22:01 hrs to 05:59 hrs | 27 |
| 19:01 hrs to 22:00 hrs | 19 |
| 16:01 hrs to 19:00 hrs | 17 |
| 06.00 hrs to 09:00 hrs | 5 |
| 03.01 hrs to 05.59 hrs | 3 |

Table 5: Time-Wise Distribution of Accidents

* + 1. **Fatal Traffic Accident Trends by Junction Type (2015-2017)**

This table shows the different types of junctions and the number of accidents associated with each type.

|  |  |
| --- | --- |
| **Junction Type** | **Number of Accidents** |
| Link | 421 |
| 4-way intersection / Crossroads | 284 |
| T-junction | 267 |
| Y-junction | 48 |
| Circular crossroads | 31 |

Table 6: Fatal Traffic Accident Trends by Junction Type (2015-2017)

* **Link Junctions**: The most accident-prone junction type, with 421 cases, requires targeted safety measures.
* **High-Risk Intersections**: 4-way intersections (284 accidents) and T-junctions (267 accidents) also pose significant risks and need safety improvements.
* **Lower Risk Junctions**: Y-junctions (48 accidents) and Circular crossroads (31 accidents) show lower accident rates, indicating relatively safer conditions.
  + 1. **T.P. Guard-Wise Accident Distribution (2015-2017) in QGIS**

This section gives a clear breakdown of accident locations across all Traffic Police (T.P.) Guard regions in Kolkata. The QGIS maps show where accidents are most concentrated, making it easier to spot high-risk areas and regional trends.

* 1. **Accident Density Across T.P. Guards**

QGIS Visualization: The map displays accident points within different T.P. Guard boundaries, highlighting key locations such as **Shyambazar**, **Bhowanipore,** **Ultadanga**, and **Vidyasagar**.

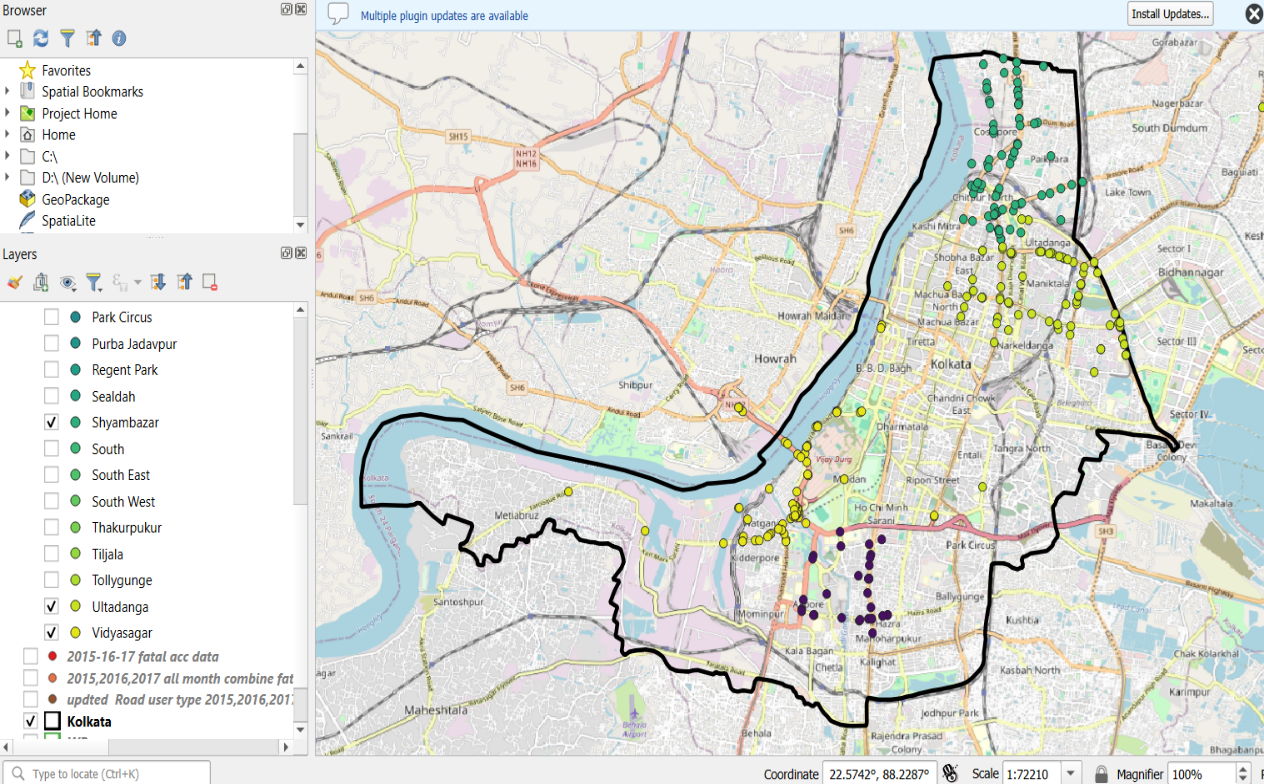


Figure 6: Accident Distribution Map across Different T.P. Guards

* 1. **Vehicle Type Distribution Across T.P. Guards**

QGIS Visualization: The map shows accident points categorized by different vehicle types, such as **Bus, Car** and **Motor Cycle.**



Figure 7: Vehicle Type Distribution Across Different T.P. Guards

* 1. **Victim Wise Category Distribution**

QGIS Visualization: The map displays accident data categorized by victim type — **Pedestrians**, **Bicyclists**, and **Rider of Motor Cycle** — across different regions.

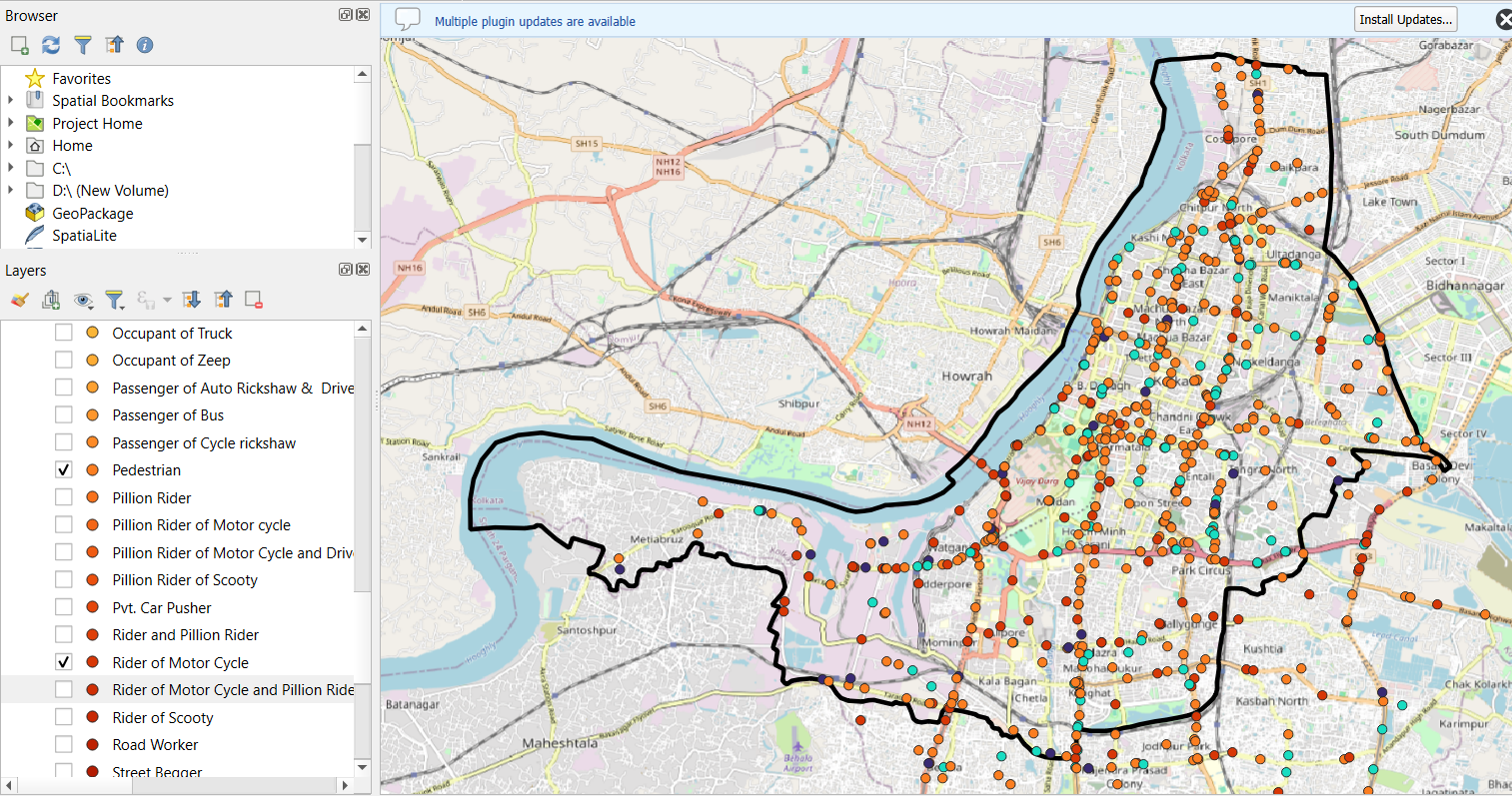


Figure 8: Victim-Wise Category Distribution

* + 1. **Conclusion**

This study provides a comprehensive analysis of traffic accidents in Kolkata from 2015 to 2017, offering critical insights into accident patterns, victim types, high-risk areas, and contributing factors.

**Chapter 2: Computer Vision based Analysis of traffic Video**

This comprehensive report examines a sophisticated traffic analysis system leveraging computer vision techniques and statistical processing to monitor traffic patterns, vehicle behaviors, and potential collision scenarios. The system demonstrates advanced capabilities in real-time object detection, tracking, and statistical analysis of traffic videos.

**2.1 Object Detection and Tracking Implementation**

The traffic analysis system implements a robust computer vision pipeline using state-of-the-art models and algorithms for accurate object detection and tracking. The core components include YOLOv9 for object detection and ByteTrack for maintaining object identity across video frames.

**YOLOv9 Object Detection**

The system employs the YOLOv9 (You Only Look Once) model from the Ultralytics framework to detect various traffic participants in uploaded videos. This implementation provides several key advantages:

* + Real-time object detection with high accuracy
  + Multi-class identification for various traffic participants (cars, buses, trucks, motorcycles, people, bicycles)
  + Bounding box generation for each detected object
  + Classification confidence scores for reliable detection

The implementation utilizes the **draw\_boxes** function to visually represent detected objects by creating bounding boxes with appropriate labels. Each detected object receives a unique identifier along with its category, enabling consistent tracking throughout the video sequence.

**ByteTrack Object Tracking**

Object tracking is managed through ByteTrack, an advanced multi-object tracking algorithm that maintains object identities across frames. The implementation demonstrates several sophisticated features:

* Consistent ID assignment across multiple frames
* Tracking through occlusions and partial visibility
* Association of detections between consecutive frames based on motion and appearance
* Management of object entry and exit from the scene

The tracking system's effectiveness is evidenced by the "ByteTrack complete!" notification displayed in the interface, confirming successful tracking operations throughout the video analysis process.

**Trajectory Analysis and Data Processing**

The system performs sophisticated trajectory analysis by processing tracked objects' positions over time, extracting valuable movement patterns and behaviors.

**Position Calculation and Trajectory Formation**

The **create\_trajectory** function implements a comprehensive approach to trajectory formation:

* Calculation of object center points by averaging bounding box coordinates
* Chronological sorting of position data (by ID, date, and time) for continuous trajectory formation
* Grouping of trajectory data by object ID and category
* Storage of complete movement paths for each detected object

This implementation enables the system to maintain detailed records of each object's path through the monitored area, forming the foundation for subsequent statistical analysis.

**Speed and Distance Computation**

The system calculates critical motion metrics through several specialized functions:

* Euclidean distance calculation between consecutive positions using the **Euclidean** function.
* Time-based velocity determination by dividing distance by elapsed time
* Total distance calculation by summing sequential position changes
* Average speed computation across an object's entire trajectory

These calculations provide essential data for both statistical reporting and safety analysis, creating a comprehensive understanding of traffic flow dynamics.

**Statistical Analysis and Metrics Generation**

The statistical analysis component transforms raw trajectory data into meaningful metrics, enabling deeper insights into traffic patterns and behaviors.

**Category Distribution Analysis (Example)**

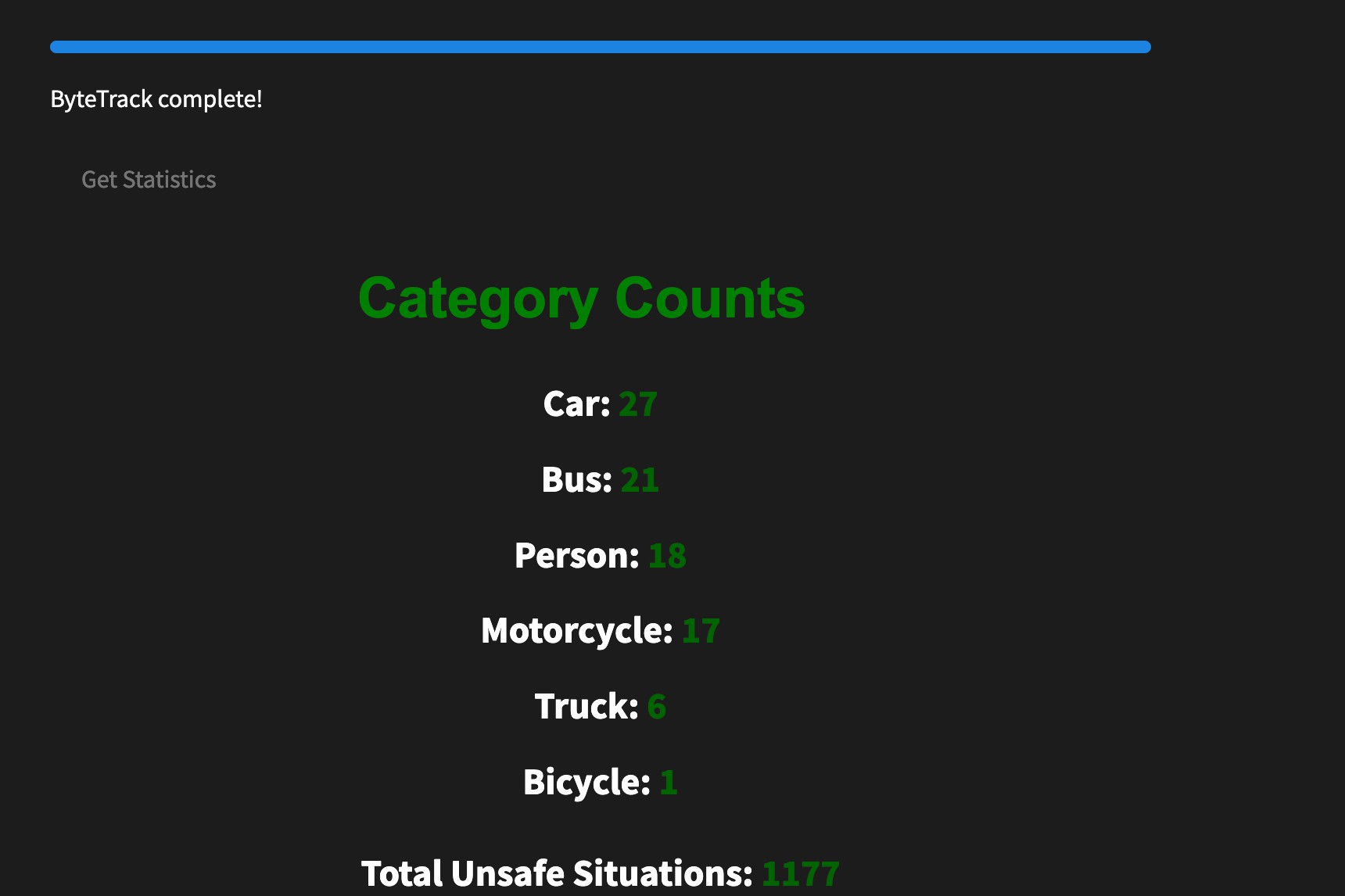
****

Figure 9: Vehicle Counts and Unsafe Situations

As shown in the "Category Counts" screenshot, the system effectively classifies and quantifies traffic participants:

* Cars: 27 instances
* Buses: 21 instances
* People: 18 instances
* Motorcycles: 17 instances
* Trucks: 6 instances

This distribution analysis provides immediate insight into traffic composition, with a total of 1177 unsafe situations identified across all categories.

**Relative Speed and Distance Calculation (Example)**

The system implements sophisticated normalization techniques to generate comparable metrics across different object types:

* The **generate\_statistics** function processes raw speed and distance data.
* **MaxAbsScaler** is applied to normalize values to a consistent scale.
* Values are rescaled to a 0-100 range for intuitive interpretation.
* Relative metrics enable meaningful comparisons between different vehicle type.

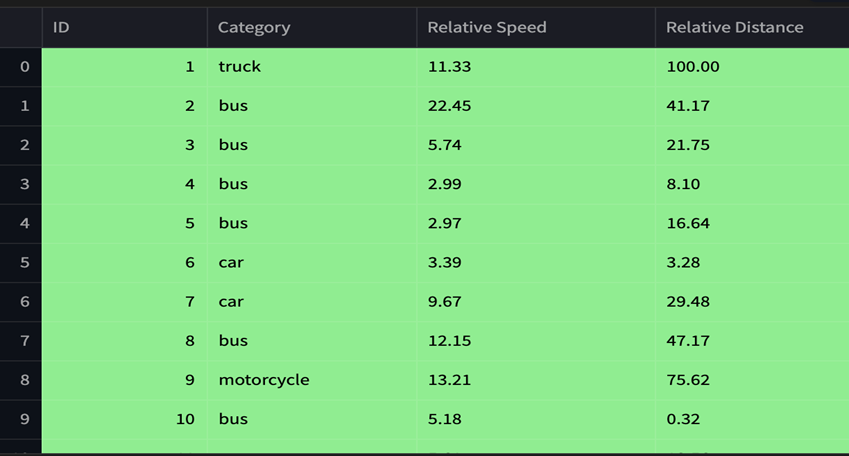


Figure 10: Statistics generated from traffic video

The "Average Relative Speed by Category" data reveals interesting patterns in traffic behavior: (0-100)

* Motorcycles: 49.39
* Trucks: 21.14
* Cars: 21.11
* Buses: 10.59 (lowest relative speed)

These normalized values provide valuable insights into the relative mobility of different traffic participants within the same environment.

**Unsafe Situation Detection**

One of the system's most valuable features is its ability to identify potentially dangerous traffic situations through advanced proximity and speed analysis.

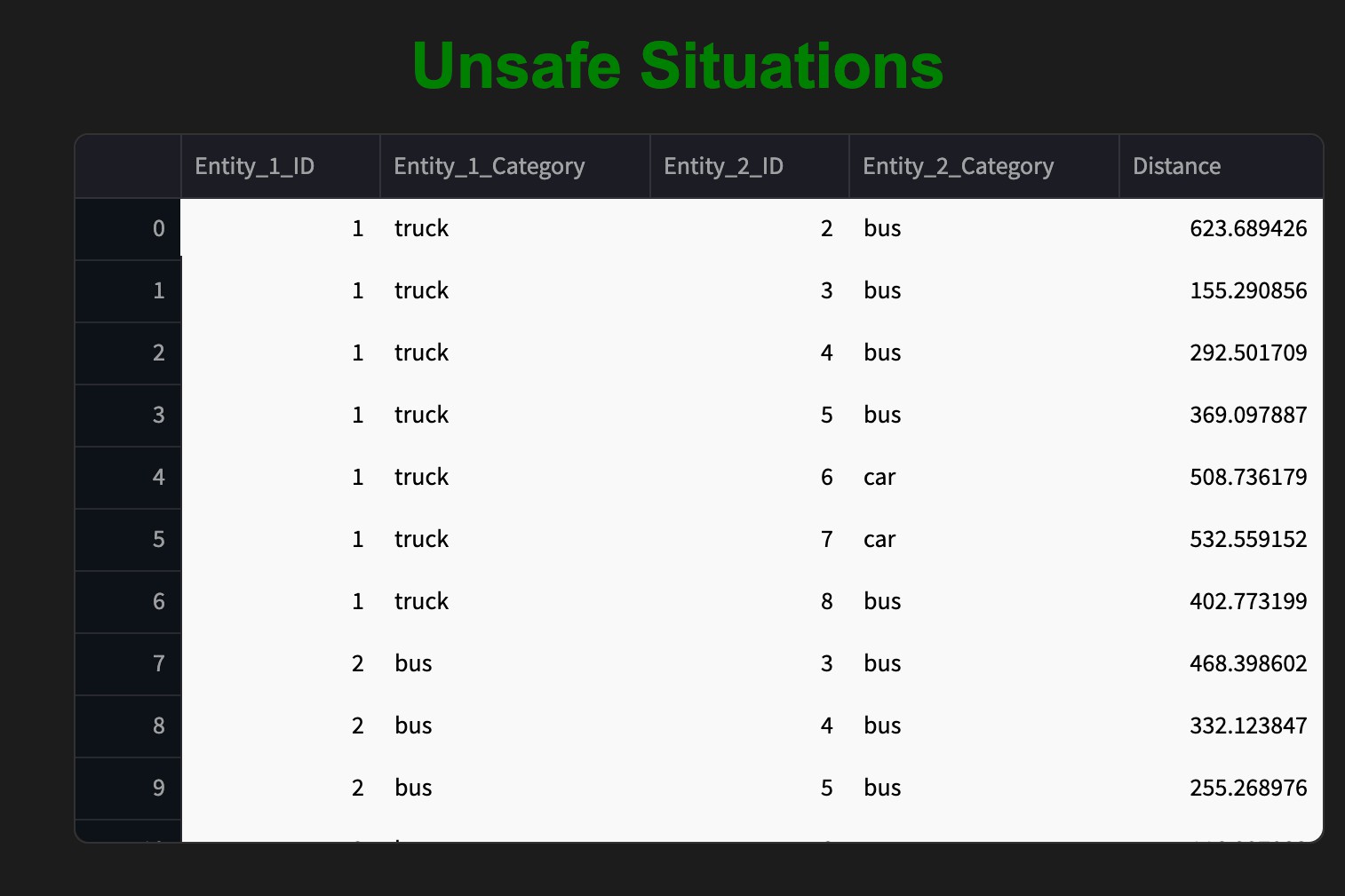


Figure 11: Details about vehicles creating unsafe situations

**Proximity Threshold Implementation**

* The detect\_unsafe\_sutiuations function implements a sophisticated collision risk detection system:
* Calculation of distances between all object pairs at each timestamp
* Comparison against a configurable proximity threshold (default: 500.0 units)
* Filtering to avoid duplicate situation reporting using a **unsafe\_pair** set.
* Special handling to exclude person-to-person interactions.

The system stores detailed information about each unsafe situation, including:

* Timestamp and frame number
* IDs and categories of both entities involved
* Precise distance measurement between entities

**Intelligent Risk Assessment**

The system demonstrates sophisticated risk assessment by considering multiple factors:

* It distinguishes between various vehicle combinations (truck-bus, truck-car, bus-bus, etc.)
* Distance measurements are stored with high precision (e.g., 292.501709, 369.097887)
* The implementation recognizes that proximity alone doesn't constitute danger (e.g., "2 slow vehicles close to each other don't create dangerous situations")
* Multiple vehicle interactions are tracked simultaneously to provide a comprehensive safety assessment

This nuanced approach to risk detection enables more accurate identification of genuinely dangerous situations while minimizing false positives.

**Visualization and User Interface**

The screenshots reveal a well-designed user interface built with **Streamlit** that effectively communicates complex traffic data:

* Tables display unsafe situations with entity IDs, categories, and distances
* Category counts are prominently displayed with color-coding
* Speed metrics are presented in an easily digestible format
* The interface provides immediate access to key statistics upon video upload

The system's visualization approach transforms complex data into accessible information, enabling users to quickly identify and understand traffic patterns and safety concerns.

**Conclusion**

The traffic analysis system represents a sophisticated integration of computer vision and statistical analysis techniques for comprehensive traffic monitoring and safety assessment. The implementation demonstrates several noteworthy achievements:

* Successful integration of state-of-the-art object detection (YOLOv9) and tracking (ByteTrack) algorithms
* Comprehensive trajectory analysis with accurate speed and distance calculations
* Intelligent unsafe situation detection based on multiple risk factors
* Effective statistical processing with meaningful relative metrics
* User-friendly visualization of complex traffic data

The system's ability to process videos in real-time, extract meaningful statistics, and identify potential collision scenarios makes it a valuable tool for traffic monitoring, urban planning, and safety analysis applications.

**Chapter 3: Road Safety Analysis through Simulations in Accident-prone locations in Kolkata**

## **3.1 Introduction**

Urban road safety is a major concern due to high traffic density and complex networks. In Kolkata, several accident-prone locations have been identified through historical data. Understanding accident causes requires a systematic approach combining real-world observations with traffic simulations.

This study uses **PTV Vissim** to model high-risk locations and simulate varied traffic conditions, generating datasets for analysis. The collected trajectory data is processed using the **Surrogate Safety Assessment Model (SSAM)** to identify traffic conflicts and assess safety risks, contributing to data-driven road safety strategies.

## **3.2 Methodology**

## **3.2.1 Data Collection and Road Network Generation**

## The first step in this study involved selecting accident-prone locations in Kolkata using historical accident data. The geographic coordinates of these locations were obtained, which served as the basis for road network modeling. Using PTV Vissim, the road network for each identified location was recreated, ensuring that the simulated environment accurately reflected real-world traffic conditions. Road geometries, lane structures, and signal timings were incorporated to enhance simulation realism.

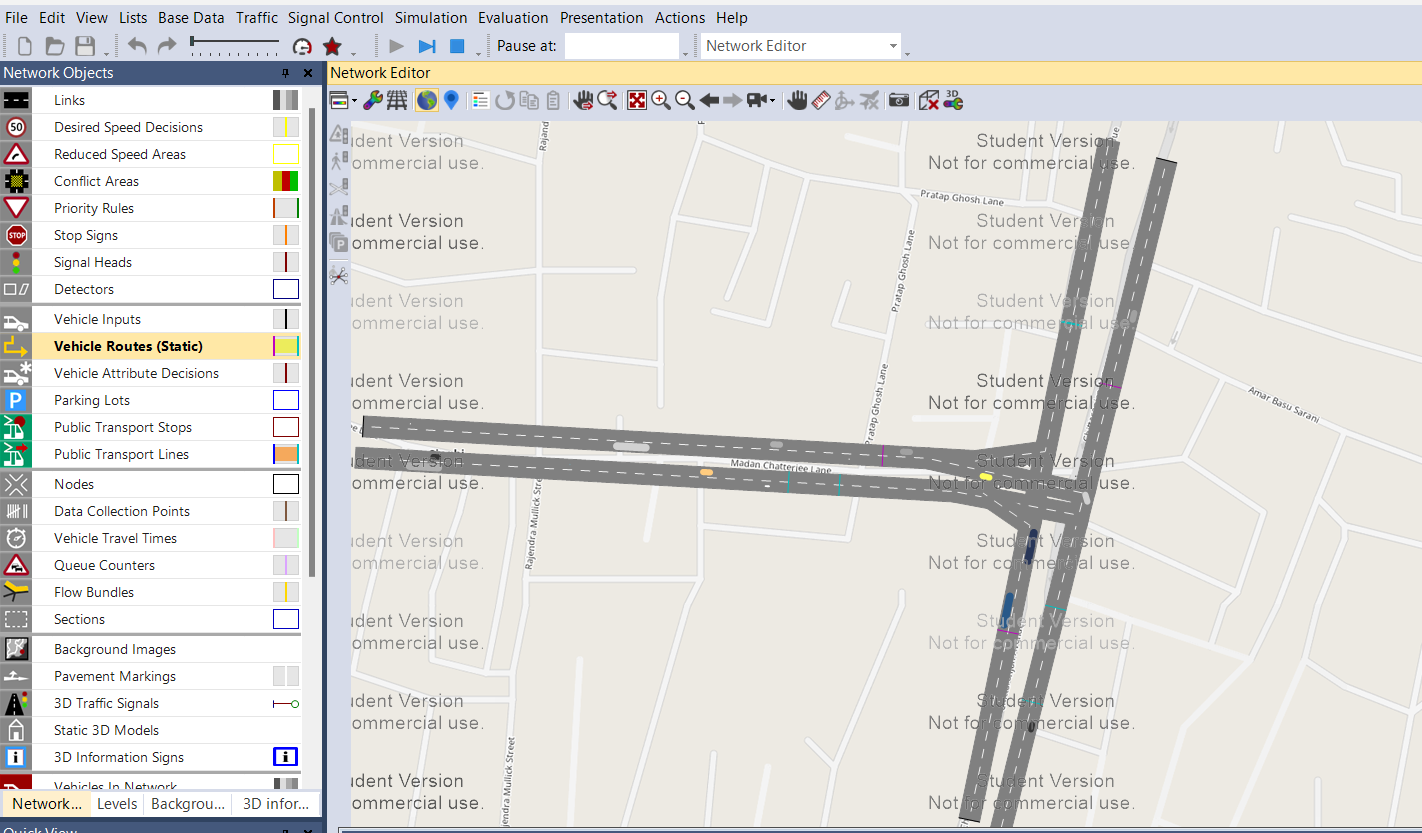


Figure 12:Snapshot of a running simulation on VISSIM

### **3.2.2 Simulation Setup and Parameter Selection**

To develop a comprehensive dataset, simulations were conducted by varying **different traffic parameters** at each selected location. The primary input parameters included:

* **Vehicle Speed:** Speed distribution of different vehicle types like Car, Light Goods Vehicle (LGV), Heavy Goods Vehicle (HGV), Bus, Motorbike, Man
* **Relative Flow (RelFlow):** Traffic demand and flow variations across road sections
* **Number of Vehicles:** Traffic density at different time intervals

These parameters were systematically modified to generate diverse traffic scenarios. The goal was to capture variations in driving behavior and traffic congestion, which directly influence accident risks. Each simulation run generated a trajectory file containing detailed information on vehicle movements, interactions, and speed variations.

### **3.2.3 Conflict Analysis using SSAM**

Once the trajectory data was collected, it was processed using the **Surrogate Safety Assessment Model (SSAM)** to identify and categorize potential conflicts. SSAM analyzes vehicle trajectories to detect high-risk interactions that could lead to collisions. The software classifies conflicts into three primary categories:

* **Rear-End Conflicts:** Sudden braking or tailgating scenarios
* **Lane-Change Conflicts:** Unsafe lane switches and merging issues
* **Crossing Conflicts:** Intersections with high collision risk

By quantifying the number and severity of conflicts, we gained insights into the likelihood of accidents occurring under different traffic conditions. The output from SSAM provides a surrogate measure of road safety, enabling researchers to assess the impact of traffic patterns on accident risks.

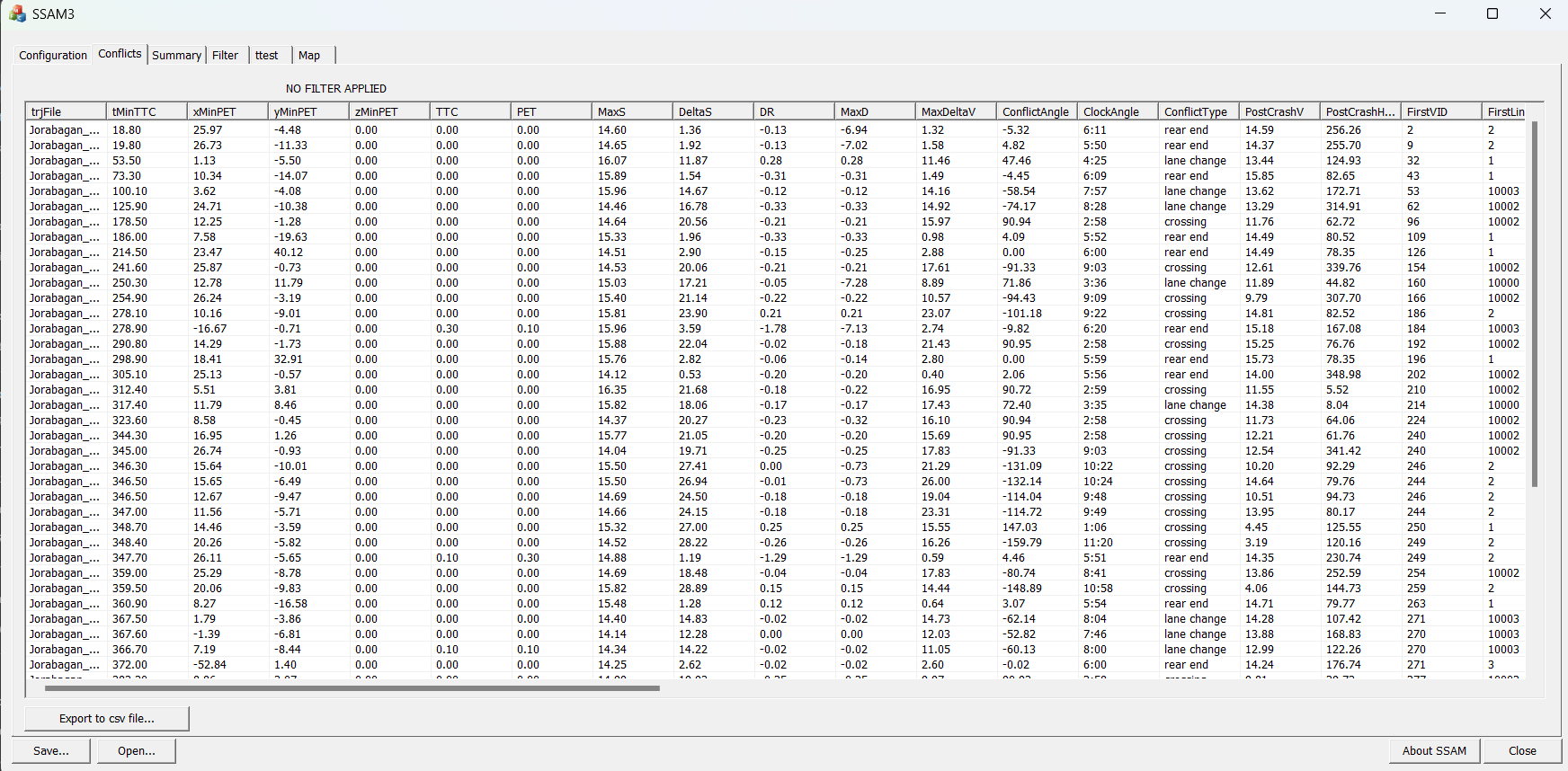


Figure 13:Snapshot collected from the SSAM Software showing the conflict data

## **3.3 Data Output and Utilization**

The extracted conflict data from SSAM served as a training dataset for a machine learning model developed in the next phase of the project. The dataset was shared with the next research team member, who used it to train a RandomForestRegressor for accident risk prediction.

The final dataset included:

* **14 Input Features** (Vehicle Speed, Traffic Flow, Number of vehicles, etc.)
* **4 Output Features** (Types of Traffic Conflicts)

## **3.4 Challenges and Limitations**

* **Manual Road Network Creation:** Some inaccuracies may arise due to limitations in recreating exact real-world road conditions in Vissim.
* **Assumption-Based Traffic Behavior:** Simulations rely on behavioral models that may not fully capture real driver responses in accident-prone situations.
* **Computational Load:** Running 100+ simulation scenarios per location required significant processing time and computational power.

## **3.5 Conclusion**

By simulating accident-prone locations using PTV Vissim and analyzing conflicts with SSAM, this study provided valuable data for data-driven road safety improvements. The generated dataset was later used for machine learning-based accident prediction, forming the basis for further research in traffic safety analytics.

**Chapter 4: Machine Learning based modelling of Road Safety under Interventions**

**4.1 Introduction**

The provided model is designed to handle a regression task using a RandomForestRegressor from the sklearn.ensemble module. It includes data preprocessing, model training, prediction, and evaluation. This report outlines the key components of the model and provide insights into its functionality.

**4.2 Data Preprocessing**

There are 14 input and 4 output features. The input features are as follows: Car\_Speed, LGV\_Speed, HGV\_Speed, Bus\_Speed, Man\_Speed, Bike\_Man\_Speed, Car\_RelativeFlow, LGV\_RelativeFlow, HGV\_RelativeFlow, Bus\_RelativeFlow, Man\_RelativeFlow, Bike\_Man\_Relative Flow, Total Vehicles, Road Width. The output features are Crossing, Rear End, Lane and Total Vehicles

Handling Missing Values: The code uses SimpleImputer from sklearn.impute to fill missing values (NaN) in the target variable (y\_train) with the mean of the existing values. This is a common strategy for handling missing data, especially when the dataset is not too large, and the missing values are not too frequent.

Imputation Strategy: The imputer is fitted on training data and then used to transform both train and test data. This ensures consistency in handling missing values across the training and testing datasets.

* 1. **Model Training**

1. Model Selection: The model chosen for this task is a RandomForestRegressor. This robust ensemble model can handle complex relationships between features and the target variable.
2. Model Fitting: The model is fitted using the imputed y\_train and the original X\_train. This means any missing values in the target variable have been replaced before training the model.

**Prediction and Evaluation**

* 1. Prediction: The trained model is used to predict values for Crossing, Rear End, Lane Change and Total Conflicts.
  2. Evaluation Metrics: The code calculates several metrics to evaluate the model's performance:
  + Mean Squared Error (MSE): Measures the average squared difference between predicted and actual values is 10470
  + Mean Absolute Error (MAE): Measures the average absolute difference between predicted and actual values, which is 57
  + R-Squared (R2): Measures how well the model explains the variance in the target variable, which is 0.75
  + Printing Predictions: The code includes a function to print the first few rows of actual and predicted values side by side for comparison.
  1. A 2x2 subplot matrix, each showing the scatter plot of actual vs. predicted values for one of the four target variables (Y1, Y2, Y3, Y4) where Y1 = Crossing, Y2 = Rear End , Y3 = Lane Change and Y4 = Total Conflicts. The code first creates DataFrames for the true values (y\_test\_df) and predicted values (predict\_df). Subplots Setup: It sets up a 2x2 subplot matrix using plt.subplots(). Scatter Plots: For each target variable, it creates a scatter plot of true vs. predicted values and labels the axes. Finally, it displays the plots using plt.show().

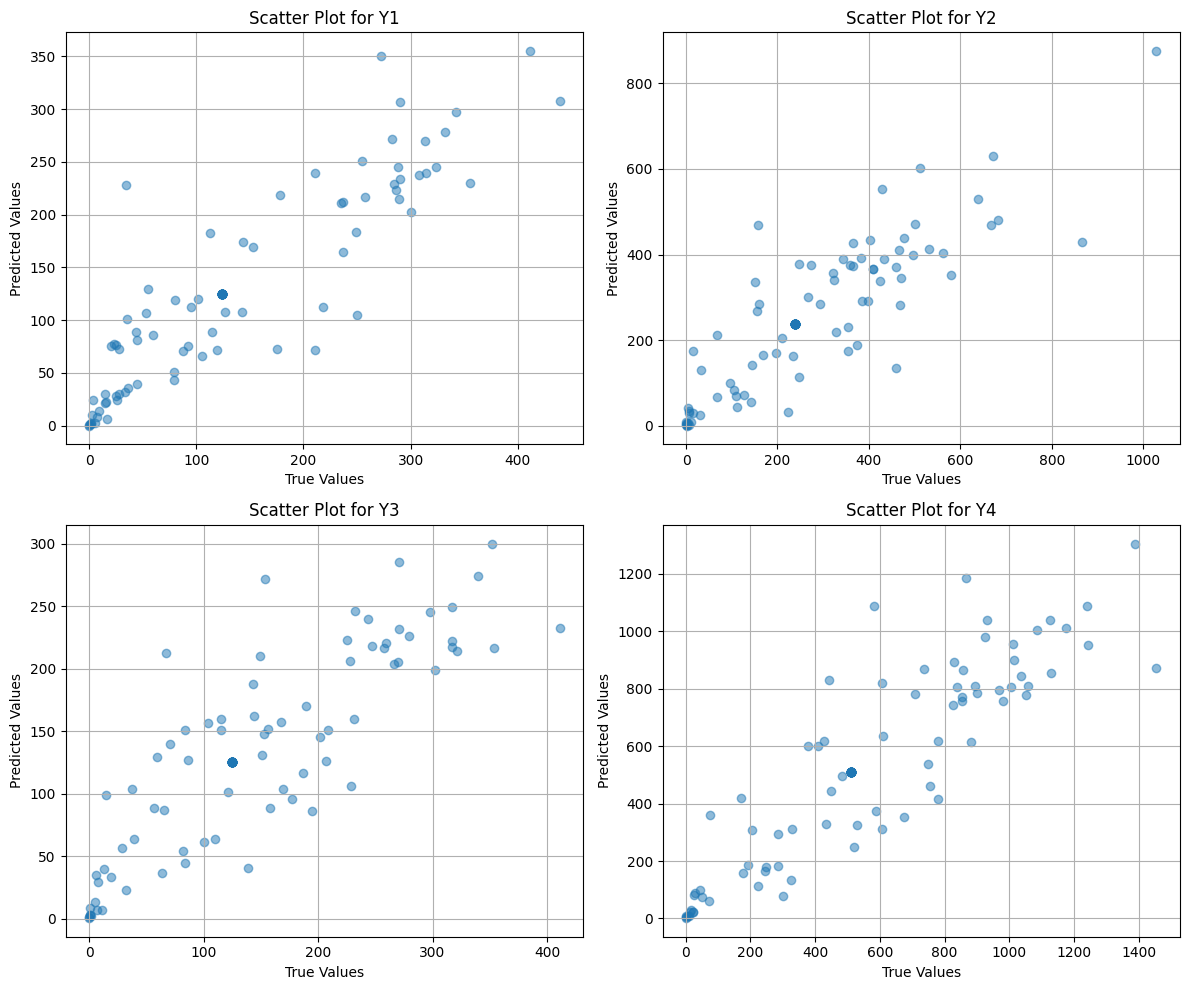


Figure 14: Comparison of Actual vs. Predicted Values for Traffic Conflict

Saving Predictions: The actual and predicted values are saved to a CSV file named “predictions\_conflicts.csv”. This allows for easy comparison and further analysis of the predictions.

**4.5 Future Work**

1. Hyperparameter Tuning using techniques like RandomizedSearchCV or GridSearchCV to optimize the hyperparameters of the RandomForestRegressor.
2. Using cross-validation techniques (e.g., KFold) to more robustly evaluate the model's performance on unseen data.
3. Error Handling by adding checks to handle potential errors, such as ensuring that the imputer is fitted before transforming data.
4. The graph below shows the nodes of a part of Kolkata. We will incorporate these nodes in our Graph Neural Network.
5. Graph Neural Networks (GNNs) are a type of deep learning model that is specifically made to handle and assess data that is shown as graphs. There are nodes (data points) and edges (connections between data points) in a graph. GNNs use this structure to do things like classifying nodes, predicting links, and classifying graphs at the graph level.
6. Some of the graphs that model traffic flow is Spatial Temporal Graph Neural Network (STGNN), ST-ResNet, STDN, etc.
7. Traffic forecasting is essential for optimizing resources and public safety; however, it is a difficult task due to the dynamic spatial correlations and temporal fluctuations in the traffic data. Spatial-temporal networks, including temporal attention networks with full graph attention networks, graph convolution networks with temporal convolution networks, and recurrent neural networks with graph convolution networks, are employed to encapsulate these complex dependencies.



Figure 15: Graph Representation of Kolkata Nodes for GNN Integration

**4.6 Conclusion**

The model provides a basic framework for training a regression model using a RandomForestRegressor. However, there are opportunities for improvement, particularly in model evaluation and optimization. The model's performance and reliability can be enhanced by addressing these areas. This study can be taken further with Graph Neural Networks and implementing it for more accurate and complex traffic prediction tasks.

**Chapter 5: Safar labs Interface design**

This report provides a detailed examination of a sophisticated Streamlit-based traffic analysis and prediction system. The interface integrates computer vision, machine learning, and geospatial visualization to deliver a comprehensive traffic management solution with multiple specialized modules.

**5.1 Interface Overview and Navigation**

The application features a clean, dark-themed interface with three distinct pages accessible via dropdown selectors at the top of the interface:

* Home
* Simulation
* GIS Software

Each page maintains consistent design elements while offering specialized functionality for different aspects of traffic analysis. The interface employs Streamlit's column-based layout system for organized content presentation and uses a wide-mode configuration to maximize screen real estate.

**5.2 Simulation Interface**

**5.2.1 Traffic Conflict Prediction**

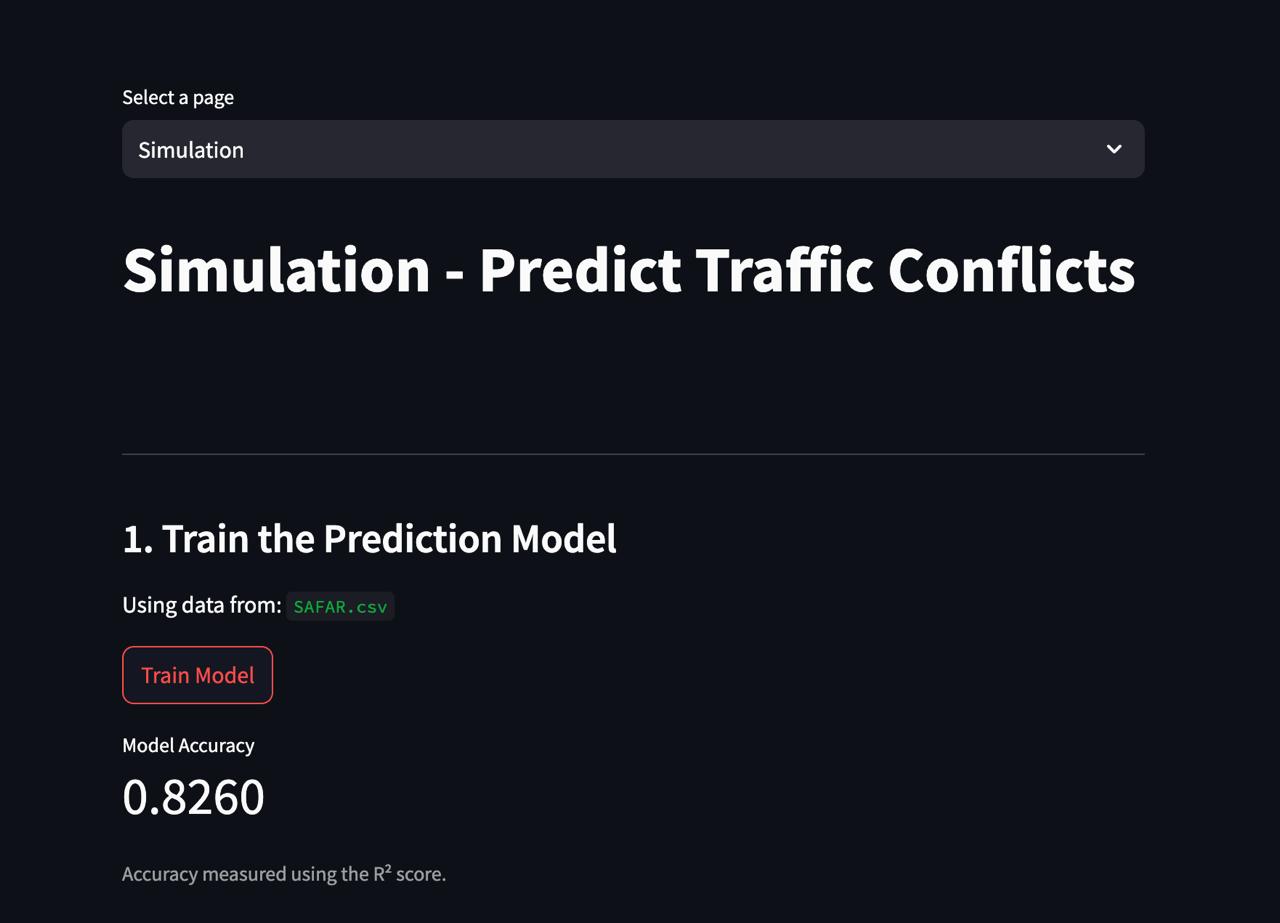
****

Figure 16: Traffic Conflict Prediction

The Simulation page serves as a sophisticated traffic conflict prediction platform with multiple functional sections.

**5.2.2 Model Training and Evaluation**

The top section of the Simulation interface focuses on model preparation:

* Clear header: "Train the Prediction Model"
* Data source indicator showing “Simulation Data.csv”
* Prominent "Train Model" button for initiating the training process
* Model performance metrics display showing an accuracy of 0.8260 (measured using R² score)

This implementation allows users to train a traffic conflict prediction model using the Simulation dataset from Vssim Software and immediately view the model's performance metrics.

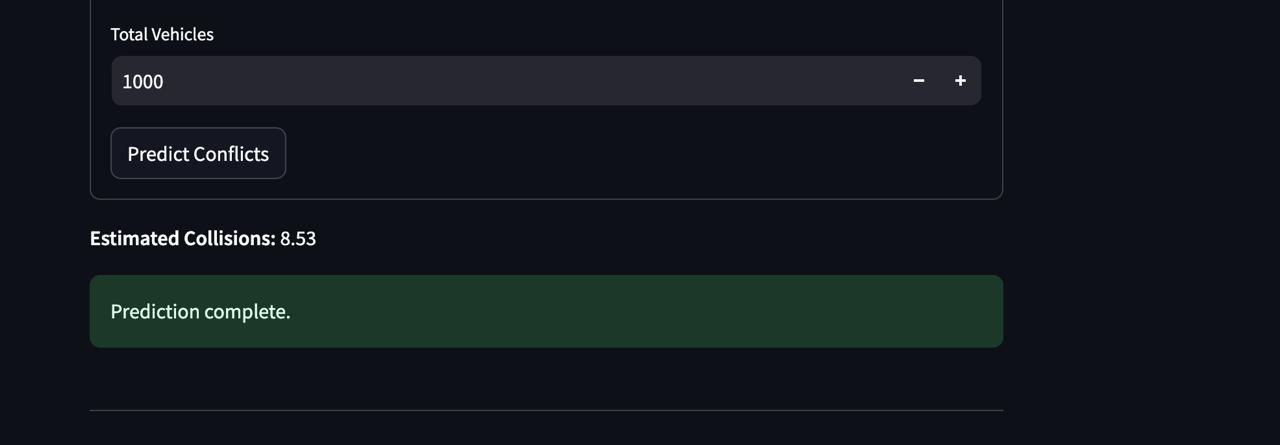


Figure 17: Model Training and Evaluation

**5.3 Traffic Scenario Configuration**

The second major section enables detailed traffic scenario specification:

* Section header: "2. Enter Traffic Scenario Details" with a traffic light icon
* Junction configuration dropdown (set to "4-way intersection")

**5.3.1 Vehicle Speed Parameters (Example Inputs)**

The interface provides granular control over vehicle speeds:

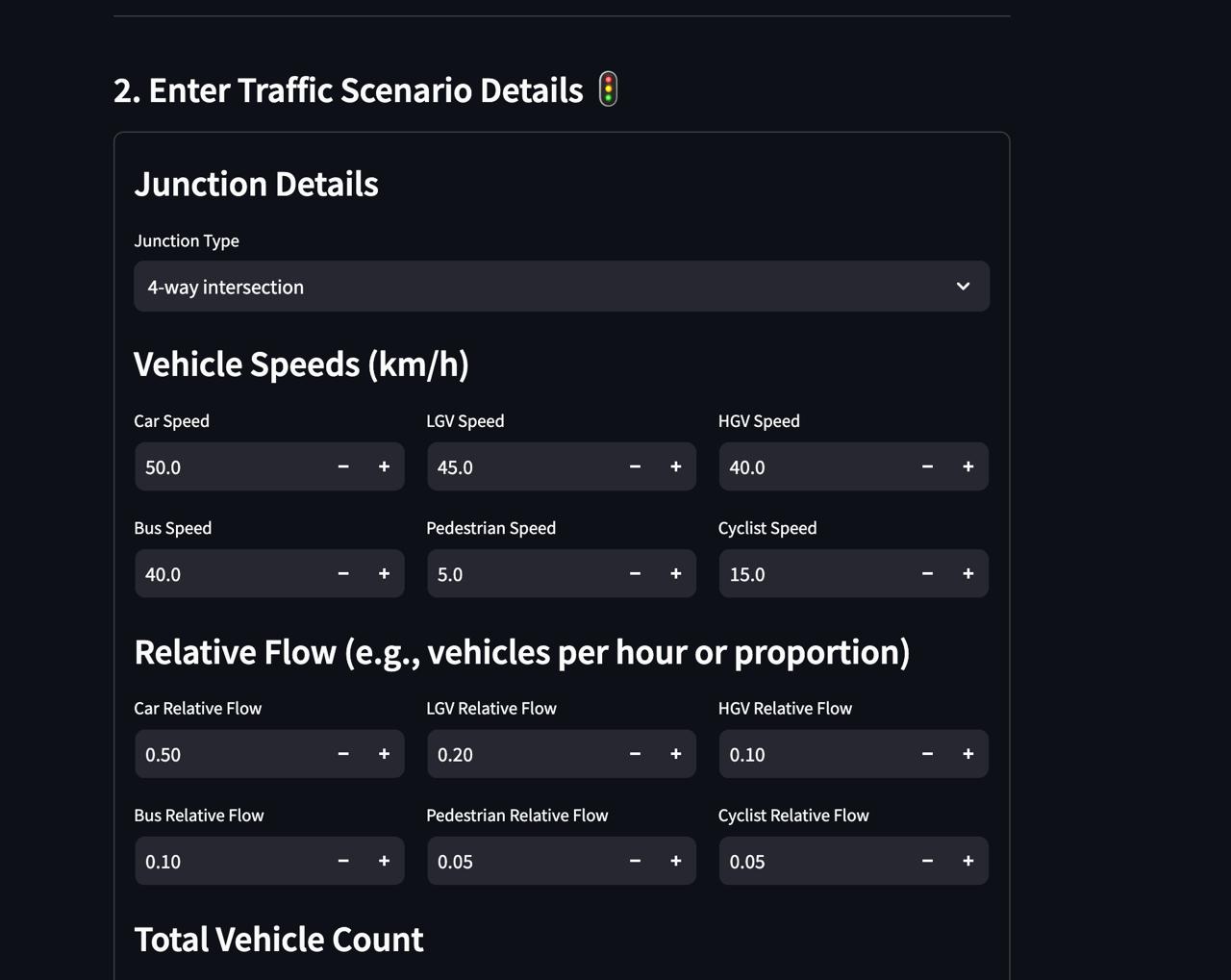


Figure 18: Vehicle Speed Parameters

* Car Speed: 50.0 km/h (with +/- adjustment controls)
* LGV (Light Goods Vehicle) Speed: 45.0 km/h
* HGV (Heavy Goods Vehicle) Speed: 40.0 km/h
* Bus Speed: 40.0 km/h
* Pedestrian Speed: 5.0 km/h
* Cyclist Speed: 15.0 km/h

**5.3.2 Traffic Flow Configuration (Example Inputs)**

* Below the speed settings, users can adjust relative traffic flow parameters:
* Car Relative Flow: 0.50 (50% of total traffic)
* LGV Relative Flow: 0.20
* HGV Relative Flow: 0.10
* Bus Relative Flow: 0.10
* Pedestrian Relative Flow: 0.05
* Cyclist Relative Flow: 0.05

These comprehensive configuration options allow users to simulate various traffic scenarios by adjusting the speed and density of different road users.

**5.4 GIS Software Interface - Geospatial Analysis**

The GIS Software page provides powerful geospatial visualization and analysis capabilities for traffic incident data.

**5.4.1 Data Management**

The top section offers data access functionality:

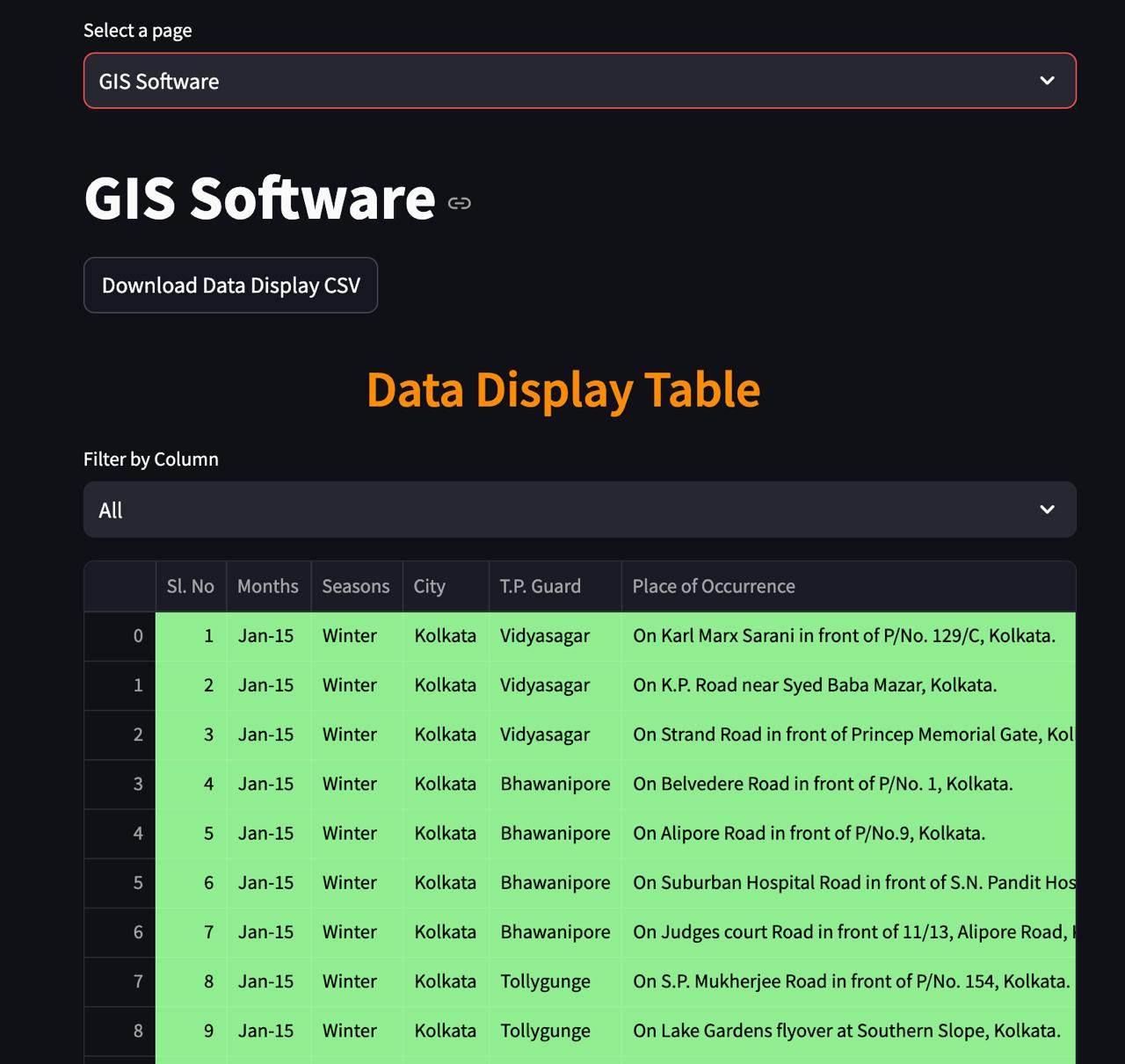


Figure 19: GIS Software Interface

* Page title: "GIS Software" with a link icon
* "Download Data Display CSV" button for exporting data (displays GIS Software data)

**5.4.2 Incident Data Table**

A comprehensive data table displays traffic incident records:

* Filter control allowing selection of specific data columns (set to "All")
* Detailed table with columns: Sl. No, Months, Seasons, City, T.P. Guard, Place of Occurrence
* Data showing January 2015 (Winter) accident records in Kolkata
* Location details with specific addresses (e.g., "On Karl Marx Sarani in front of P/No. 129/C, Kolkata")
* Police jurisdiction information (including Vidyasagar, Bhawanipore, Tollygunge areas)

**5.4.3 Geospatial Visualization**

The "GIS Map" section provides a powerful visual representation of incident data:



Figure 20: Geospatial Visualization

* Interactive map centered on Kolkata
* Orange markers indicating exact accident locations throughout the city
* Clear concentration patterns visible in central Kolkata and along major corridors
* The map uses OpenStreetMap as its base layer, providing detailed street-level context

This visualization allows for immediate identification of accident hotspots and problematic areas within the city.

**5.5 Technical Implementation Features**

Based on the code and interface elements, the system incorporates several sophisticated technical components:

**5.5.1 Computer Vision Pipeline**

* YOLOv9 implementation for real-time object detection in traffic videos
* ByteTrack algorithm for consistent object tracking across video frames
* Classification of multiple traffic participant types (people, cars, buses, trucks, motorcycles, bicycles)

**5.5.2 Statistical Analysis**

* Calculation of relative speeds and distances for different vehicle types
* Normalization of metrics for meaningful comparisons across vehicle categories
* Generation of comprehensive statistics about traffic composition and behavior

**5.5.3 Machine Learning Integration**

* Predictive modeling for traffic conflict situations
* Model training interface with performance evaluation
* Scenario-based prediction capabilities

**5.5.4 Collision Detection System**

* Advanced proximity and speed-based analysis for identifying unsafe situations
* Differentiation between various vehicle combination scenarios
* Intelligent classification that considers both distance and speed factors

**5.6 Conclusion**

The interface represents a comprehensive traffic analysis and prediction system with multiple specialized modules. It combines:

1. Advanced computer vision techniques for traffic monitoring
2. Sophisticated statistical analysis of traffic patterns and behaviors
3. Predictive modeling for traffic conflict scenarios
4. Geospatial visualization of historical accident data

The three-page interface provides a logical workflow from data visualization (GIS Software) to predictive modeling (Simulation), creating a powerful platform for traffic safety analysis, urban planning, and accident prevention. The implementation demonstrates sophisticated use of Streamlit's capabilities to create an intuitive, information-rich application for complex traffic analysis tasks.