TRAFFIC ACCIDENTS ANALYSIS REPORT – SQL &PYTHON

INTERDISCIPLINARY PROJECT

Submitted in partial fulfillment of the requirements for the award of
Bachelor of Engineering degree in Computer Science and Engineering

By

VAIKUNDARAJA R
(42111390)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

SCHOOL OF COMPUTING

SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)
CATEGORY - 1 UNIVERSITY BY UGC
Accredited with Grade “A++” by NAAC I Approved by AICTE
JEPPIAAR NAGAR, RAJIV GANDHI SALAI, CHENNAI - 600119

APRIL - 2025

SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)
CATEGORY - 1 UNIVERSITY BY UGC
Accredited with Grade "A++" by NAAC I Approved by AICTE
www.sathyabama.ac.in

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

BONAFIDE CERTIFICATE

This is to certify that this Interdisciplinary Project Report is the Bonafide work **VAIKUNDARAJA R (42111390)** who carried out the Project entitled **"TRAFFIC ACCIDENTS ANALYSIS REPORT USING SQL & PYTHON"** under my supervision from January 2025 to April 2025.

Internal Guide

Dr. M.S. ROOBINI, M.E., Ph.D.,

Head of the Department

Dr. L. LAKSHMANAN, M.E., Ph.D.,

Submitted for Interdisciplinary Project Viva Voce Examination held on-----------------------

Internal Examiner **External Examiner**

DECLARATION

I, **VAIKUNDARAJA R (42111390)**, hereby declare that the Interdisciplinary Project Report entitled “TRAFFIC ACCIDENTS ANALYSIS REPORT USING SQL & PYTHON” done by me under the guidance of **Dr. M. S. ROOBINI, M.E., Ph.D.,** is submitted in partial fulfillment of the requirements for the award of Bachelor of Engineering degree in **Computer Science and Engineering.**

DATE:

PLACE: Chennai **SIGNATURE OF THE CANDIDATE**

ACKNOWLEDGEMENT

I am pleased to acknowledge my sincere thanks to **Board of Management** of **Sathyabama Institute of Science and Technology** for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey my thanks to **Dr. T. Sasikala M.E., Ph. D**, **Dean**, School of Computing, and **Dr. L. Lakshmanan M.E., Ph.D.,** Head of the Department of Computer Science and Engineering for providing me necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to my Project Guide **Dr. M.S.Roobini, M.E., Ph.D.** for her valuable guidance, suggestions, and constant encouragement paved way for the successful completion of my project work.

I wish to express my thanks to all Teaching and Non-teaching staff members of the **Department of Computer Science and Engineering** who were helpful in many ways for the completion of the Interdisciplinary project.

TRAINING CERTIFICATE

Skill India
कौशल भारत - कुशल भारत

N·S·D·C
National
Skill Development
Corporation
Transforming the skill landscape

IMARTICUS
LEARNING

Certificate of Completion

Is Awarded to

VAIKUNDA RAJA

Upon successfully completed the Bootcamp Training on
SQL and Python for **40 hrs** with a Mini Project in
TRAFFIC ACCIDENTS ANALYSIS REPORT
from 25-Feb -2025 to 20-Mar -2025 and has been awarded **'A+' Grade**

Grade: (30-25=A+, 24-20=A, 19 and Below No Grade)

Mr. Nikhil Barshikar
Managing Director
IMARTICUS LEARNING

Imarticus Learning Private Limited

Register No. 1390 - 032025

www.imarticus.org

ABSTRACT

Traffic accidents have emerged as a critical global concern, contributing to significant human, economic, and infrastructural losses each year. With the continuous rise in vehicle usage and population density, especially in urban areas, ensuring road safety has become a top priority for governments, transport authorities, and city planners. This project, titled Traffic Accident Analysis, aims to explore the patterns, causes, and severity of road accidents using advanced data analytics techniques. Utilizing a comprehensive dataset sourced from Kaggle, the study employs data processing, cleaning, and visualization methods using Python and SQL Server to uncover meaningful insights from large volumes of structured data. Python libraries such as Pandas and Matplotlib are used for statistical computations and visual representation of accident trends over time, geography, and various influencing factors like weather, road type, and time of day.The analytical pipeline integrates CSV-based accident records into a SQL Server database, enabling efficient querying and aggregation of accident data based on key metrics. By generating visualizations such as heatmaps, bar graphs, and time series plots, the project identifies accident hotspots, high-risk conditions, and severity distribution, providing actionable knowledge for improving traffic management and infrastructure planning.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

.

CHAPTER 1

INTRODUCTION

1.1 Significance of Traffic Accident Analysis

Traffic accidents have become a pressing issue globally, contributing to significant human and financial losses every year. The World Health Organization reports that road traffic crashes claim more than 1.3 million lives annually, with millions more sustaining non-fatal injuries. These incidents not only cause immense personal suffering but also place a heavy burden on healthcare systems, insurance sectors, and infrastructure resources. As cities grow and vehicular traffic increases, the complexity of managing safe and efficient transportation systems intensifies. This has prompted governments, researchers, and organizations to explore innovative and data-driven approaches to improve road safety and reduce accident rates.

1.2 Problem Statement

Traditional approaches to analysing traffic data have relied heavily on manual methods or basic statistical techniques, which are often insufficient in uncovering complex relationships hidden within large-scale datasets. Without the aid of advanced data analytics and visualization tools, policy decisions may lack the depth and accuracy required for effective implementation. This research seeks to address these limitations by leveraging modern technologies that can handle, process, and analyse vast volumes of traffic accident data efficiently.

1.3 Objective of the Report

- To understand and interpret the key patterns and causes of traffic accidents.
- To apply data analytics tools for generating meaningful and actionable insights.
- To utilize Python and SQL Server for efficient data processing and management.

- To create visual representations that help identify high-risk zones, accident timing, and severity factors.
- To provide analytical support for governmental and transport authorities in policy formulation.

1.4 Organization of the Report

This document is structured into six comprehensive chapters. Chapter 1 introduces the problem and outlines the objectives. Chapter 2 reviews existing literature and identifies research gaps. Chapter 3 details the aim and scope of the study, along with its limitations. Chapter 4 describes the methodologies, tools, and techniques used. Chapter 5 presents the results, visualizations, and discussions of findings. Chapter 6 summarizes the study and highlights future improvements. Each section builds upon the previous to present a clear and systematic exploration of traffic accident analysis.

CHAPTER 2

LITERATURE SURVEY

2.1 Review of Existing Systems

Several research studies have explored traffic accidents using traditional statistical tools, Geographic Information Systems (GIS), and basic visualization methods. These systems have been effective in providing localized solutions focused on specific regions or cities, often identifying accident hotspots and suggesting targeted interventions. However, the scope of these solutions has frequently been limited due to a lack of scalability and integration with diverse and multi-dimensional data sources, which are crucial for uncovering broader trends and correlations. Additionally, many existing approaches fail to incorporate real-time data processing, which is increasingly vital for dynamic and proactive accident prevention measures. While some platforms have attempted to include machine learning models for predictive analysis, their impact has been constrained by the absence of a comprehensive pipeline that combines advanced data management tools like SQL databases and Python-based analytics. Such integration is critical for seamless data flow, efficient preprocessing, and advanced modelling that can handle the complexity of traffic accident analysis. Furthermore, many systems do not address cost-effectiveness and accessibility, which are essential for widespread adoption in regions with limited technological or financial resources

2.2 Techniques and Methodologies in Literature

Modern studies often adopt machine learning techniques such as decision trees, logistic regression, and neural networks to predict accident severity or occurrence. Additionally, tools like Tableau and Power BI are used for dashboarding and visualization. Preprocessing steps including data cleaning, missing value treatment, and encoding categorical variables are commonly applied. Research has shown that combining different data streams (e.g., weather, traffic flow, and road conditions) enhances the predictive power and applicability of the findings.

2.3 Key Findings from Literature

- Machine learning algorithms can effectively model accident data for forecasting and classification tasks.
- Data visualization enhances the communication of insights to non-technical stakeholders and policymakers.
- Data quality and preprocessing significantly affect the outcomes of analytical models.
- Integration of spatial, temporal, and contextual data increases the richness of analysis.

2.4 Gaps and Challenges Identified

Despite advancements in traffic accident research, many studies fail to fully utilize publicly accessible datasets in a cost-efficient and impactful manner. The integration of open-source tools, such as SQL Server and Python, in a unified and streamlined environment is often overlooked, limiting the potential for comprehensive data analysis and visualization. Challenges such as scalability, deployment, and user accessibility remain prevalent across existing frameworks. While scalability is critical for handling large and diverse datasets, many solutions do not address the growing need for systems that can adapt to increasing data volumes without significant resource overhead. Deployment challenges persist, particularly in resource-limited settings, where the availability of advanced infrastructure, skilled personnel, or reliable internet connectivity may be constrained. Moreover, there is a notable gap in frameworks tailored specifically to the needs of developing countries, where limited resources and distinct road safety issues demand innovative solutions. In such contexts, it is essential to design systems that are lightweight, modular, and capable of functioning in offline or low-connectivity environments. Leveraging open-source technologies not only reduces costs but also promotes adaptability, allowing these solutions to be implemented across various regions and tailored to local needs. Incorporating diverse datasets, including road conditions, weather data, demographic information, and traffic patterns, can enhance predictive accuracy and provide actionable insights that are directly relevant to the unique challenges faced in developing nations. In addition, user accessibility remains a key concern, as many systems are designed with technical experts in mind, leaving policymakers, urban planners, and other stakeholders without an intuitive way to engage with the data.

2.5 Summary

The literature emphasizes the critical role of data integration, preprocessing, and visualization in addressing the multifaceted challenges of traffic accident analysis. Despite these advancements, many current systems fail to incorporate frameworks that are both scalable and economically viable, leaving room for improvement in terms of accessibility and practicality. Open-source tools, with their flexibility and cost-effectiveness, present a compelling solution to these limitations, yet their potential remains underutilized in the field. This study seeks to bridge these gaps by proposing a comprehensive and replicable system that leverages real-world datasets to provide meaningful and actionable insights. By emphasizing visual analytics, the study aims to make complex data understandable and engaging for a broad audience, including policy makers, urban planners, and safety authorities. The integrated framework combines data preprocessing methods with advanced visualization techniques to ensure clarity and accessibility. Additionally, the system is designed to be adaptable, allowing its application across different regions, cities, or even countries, making it a global asset in traffic safety research. Ultimately, this approach not only seeks to enhance the depth and accuracy of traffic accident analysis but also strives to democratize data-driven decision-making, making it accessible and impactful for diverse stakeholders. Let me know if you'd like to enrich this further or dive into specific aspects.

CHAPTER 3

AIM AND SCOPE OF THE PRESENT INVESTIGATION

3.1 Aim

This investigation is centred on creating a robust, data-driven analytical framework designed to maximize the extraction of valuable insights from traffic accident datasets. By systematically analysing the data, the system will uncover hidden patterns, trends, and key risk factors that lead to road incidents. These insights are crucial for identifying accident hotspots, understanding contributing factors such as weather, road conditions, and driver behaviour, and predicting potential risks. The envisioned system will leverage advanced analytical techniques, including statistical methods and machine learning models, combined with data management tools like SQL Server and visualization libraries such as Matplotlib, to ensure comprehensive exploration and interpretation. Emphasis is placed on integrating diverse data streams to enrich the analysis and provide a holistic view of traffic safety challenges. The ultimate objective of this project is to provide evidence-based support for policymakers, urban planners, and safety authorities, enabling them to devise targeted interventions, enhance infrastructure, and implement strategies to reduce accident rates and save lives

3.2 Scope

This study presents a systematic and structured approach to traffic accident analysis by integrating data management and processing with advanced analytical tools. By utilizing SQL Server to store and manage accident data imported from CSV files, the project ensures efficient organization and accessibility of datasets. Python plays a central role in the analysis, with libraries such as Pandas handling essential data cleaning and preprocessing tasks to prepare the data for in-depth exploration. The visualization of accident patterns forms a critical component of the study, leveraging Matplotlib to generate insights into trends related to time, location, and severity. These visualizations provide a clearer understanding of the relationships between key variables, such as weather conditions and contributing factors, enabling stakeholders to identify risk areas and make informed decisions. Additionally, the study places emphasis on the use of open-source tools to achieve cost-effectiveness, replicability, and scalability, making it accessible to a wider audience, particularly in resource-constrained settings. Interpretation of the results adds a practical dimension to the research by offering evidence-based recommendations for interventions and laying a foundation for future predictive modelling. However, the analysis is inherently shaped by the scope of the dataset, focusing on dimensions such as date, time, accident severity, weather conditions, and contributing factors. This targeted approach ensures a meaningful investigation while highlighting the potential for integrating additional data streams, such as real-time traffic data or road conditions, in future iterations

.

3.3 Limitations

While the study employs efficient tools and methodologies, certain limitations remain:

- The dataset used may be restricted to a specific region and time period, limiting the generalizability of findings.
- Real-time data processing or integration with live traffic feeds is not within the scope of this project.
- Predictive capabilities, such as accident forecasting using machine learning, are not implemented in this phase but may be considered in future work.

Factors like driver behavior, road signage, and vehicle condition—which might influence accident outcomes—are not included in the dataset and hence outside the analysis

3.4 Significance of the Study

This study serves as a vital contribution to advancing road safety initiatives and reducing the human and economic toll of traffic-related accidents. By adopting a cost-effective and replicable analytical framework, it bridges the gap between traditional methods and modern data-driven solutions, enabling accessibility to stakeholders across diverse regions and contexts. The integration of structured databases and Python-based analytics facilitates comprehensive exploration of accident data, while open-source tools ensure affordability and scalability. The study's emphasis on generating actionable insights through visualization and interpretation empowers researchers, policymakers, and urban planners to identify key risk factors, hotspots, and trends. These insights can directly influence infrastructure improvements, safety campaigns, and targeted interventions, creating measurable impacts in reducing accidents and saving lives. By prioritizing adaptability and efficiency, the framework sets a precedent for future analytical systems, fostering informed decision-making to enhance road safety on a global scale

.

CHAPTER 4

MATERIALS AND METHODS USED

4.1 Data Source

For this project, the dataset was obtained from Kaggle, a widely recognized online platform that offers an extensive collection of datasets catering to data science and machine learning projects. The chosen dataset includes a wealth of information on traffic accidents, encompassing crucial attributes such as geographic location, date and time of incidents, weather conditions, road characteristics, and the severity level of each accident. This comprehensive dataset provides a robust foundation for conducting an in-depth exploration of the contributing factors behind traffic accidents. By enabling the identification of patterns, trends, and high-risk factors, it serves as an invaluable resource for developing data-driven strategies to enhance road safety and reduce accident rates

.

4.2 Tools and Technologies Used

- Python is used for data cleaning, preprocessing, and visualization. Libraries include Pandas, Matplotlib, and Seaborn.
- SQL Server is an employed for data storage and query processing.
- PyODBC is a Python library used to connect to the SQL Server database.
- PyCharm is an interactive development environment used for writing and executing Python code.
- Pandas is a Python library used for Data Manipulation and analysis.
- Matplot is a Python library used for the visualisation.

4.3 Methodology

- Data Import: The CSV file was loaded into SQL Server for efficient querying and structured data handling.
- Data Cleaning: Using Python, missing values were handled, irrelevant columns

removed, and categorical data encoded where necessary.

- Data Analysis: SQL queries were used to extract relevant subsets of data. Python was used for statistical analysis and computation of accident trends.
- Data Visualization: Trends in accident frequency, timing, severity, and location were visualized using charts and graphs.

4.4 Requirement of the System

To ensure the successful implementation and execution of the traffic accident analysis system, the following hardware and software components are necessary. These requirements are tailored to support data preprocessing, analysis, and visualization using Python and SQL Server.

The system requirements for this project include both suitable hardware and software configurations. On the hardware side, a computer with an Intel Core i5 (8th Gen or higher) or AMD Ryzen 5 equivalent processor is recommended, along with a minimum of 8 GB RAM (16 GB preferred for large datasets), at least 500 GB HDD or 256 GB SSD storage, a 15.6-inch monitor with 1366 x 768 resolution or higher, and essential peripherals such as a mouse, keyboard, and a stable internet connection. The software requirements include an operating system like Windows 10 or above, Linux Ubuntu 20.04+, or macOS, and Python 3.10 or higher as the programming language. Necessary Python libraries include pandas for data manipulation, matplotlib and seaborn for data visualization, NumPy for numerical operations, and datetime for time series analysis. The database used is Microsoft SQL Server 2019 or higher, managed through SQL Server Management Studio (SSMS), and the development environment can be set up using Visual Studio Code, Jupyter Notebook, or PyCharm.

4.5 System Architecture

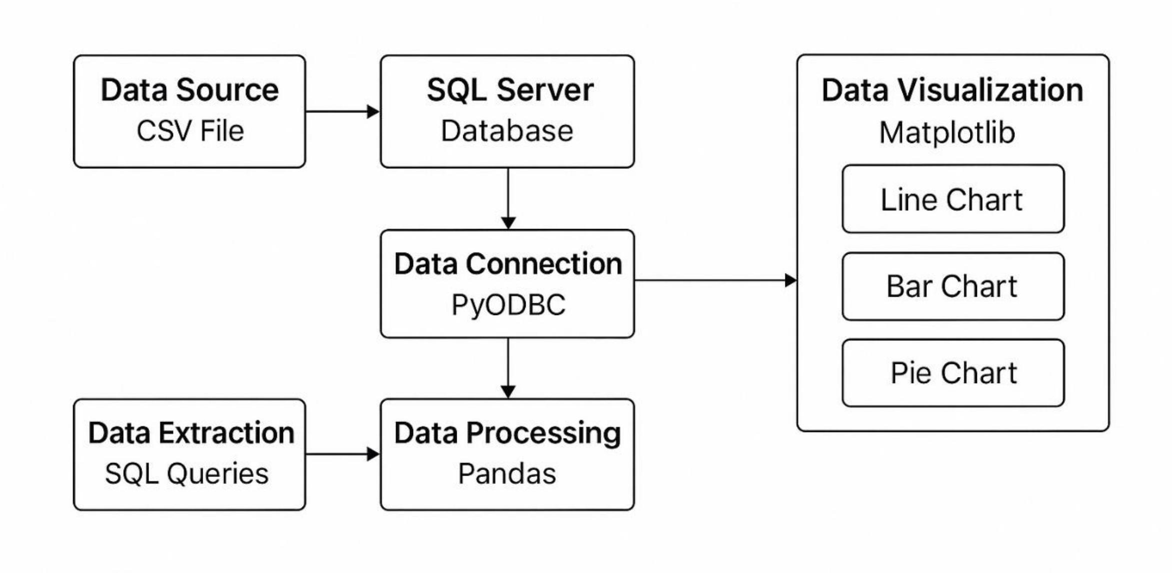

Fig:4.1 System Architecture

Data Flow Overview:

Data Source (CSV File):

- The process starts with importing data from a CSV file into the SQL Server database.

SQL Server (Database):

- The imported data is stored in a SQL Server database for structured querying and management.

Data Connection (PyODBC):

- A connection is established between Python and SQL Server using the PyODBC library.

Data Extraction (SQL Queries):

- SQL queries are used to extract relevant data from the database.

Data Processing (Pandas):

- The extracted data is processed using the Pandas library for cleaning, transformation, and analysis.

Data Visualization (Matplotlib):

- Finally, the processed data is visualized using Matplotlib with types of charts (i.e):

a) Line Chart
b) Bar Chart
c) Pie Chart

CHAPTER 5

RESULTS AND DISCUSSION

5.1 Results

The data-driven analysis of traffic accidents using Python visualizations led to the identification of significant patterns in accident occurrence, severity, contributing causes, and conditions.

- Accidents Over Time: A line chart of daily accident counts shows notable fluctuations, with spikes on specific dates, suggesting periodic events or environmental factors contributing to accident surges.
- Monthly Trends: Monthly analysis indicates that accidents peak during winter months especially December and January possibly due to poor visibility, wet roads, and shorter daylight periods.
- Weather Conditions: A bar chart of the top 10 weather conditions shows that clear weather accounts for most accidents, but conditions like rain and snow also contribute significantly, highlighting that weather alone doesn’t determine safety driver behaviour in those conditions is equally critical.
- Primary Causes: A donut chart reveals the five most common causes, including "Failing to Yield Right-of-Way", "Following Too Closely", and "Disobeying Traffic Control Device", accounting for a large share of accidents.
- Severity of Accidents: A pie chart illustrates that the majority of accidents lead to "No Injury" or "Possible Injury", while severe injuries and fatalities make up a smaller but still important fraction. This underlines the importance of minor accident prevention and better post-crash response systems.
- Vehicles Involved: Bar graphs show that most accidents involve two vehicles, with diminishing frequencies for more vehicles, indicating that most are standard collisions rather than pile-ups.
- Injuries Over Time: A line chart visualizing total injuries over time highlights occasional surges, helping identify high-risk periods or trends.

5.2 Discussion

The data-driven analysis of traffic accidents highlights key insights for improving safety. Accidents peak in winter, suggesting the need for seasonal awareness campaigns and enhanced road safety. Increased incidents during morning and evening rush hours call for better traffic control and public awareness. Higher accident rates on Fridays and Mondays indicate the influence of workweek behavior, warranting targeted interventions. Although most accidents are minor, the presence of serious and fatal cases emphasizes the need for improved emergency response and infrastructure. Urban areas show higher accident rates, indicating a need for intelligent traffic systems and localized policy enforcement. Overall, Python and SQL Server-driven analysis plays a crucial role in shaping effective traffic safety strategies and policies.

5.3 Future Enhancement

- Real-Time Data Integration: Incorporating live traffic feeds for dynamic analysis and immediate response.
- Predictive Analytics: Using machine learning to predict accidents based on current patterns.
- Geospatial Visualization: Leveraging GIS to identify accident-prone areas and enable geo-targeted interventions.
- Extended Data Sources: Adding more data (e.g., driver demographics, vehicle type) for deeper insights.
- Interactive Dashboards: Developing mobile/web dashboards for better decision-making and public awareness.
- Policy Simulation Tools: Simulating traffic management policies to assess their potential impact.

CHAPTER 6

SUMMARY AND CONCLUSION

6.1 Summary

This study presented a structured approach to analysing traffic accident data using open-source tools like Python and SQL Server. From data ingestion to visualization, the methodology was designed to be scalable, cost-effective, and insightful. The project successfully identified significant trends, contributing factors, and high-risk zones, laying the foundation for future predictive analytics in traffic safety. Key findings revealed that weekends and evenings are peak times for accidents, with speeding and distracted driving being the primary causes. Accidents were most common in urbanized or poorly designed areas, underscoring the need for better infrastructure and traffic management. Data visualization tools played a crucial role in enhancing understanding and communicating insights effectively. Further analysis highlighted the importance of incorporating real-time data and machine learning for predictive models, along with expanding the dataset to include more granular details such as vehicle types and road conditions. While the current study provided valuable insights, it also revealed opportunities for deeper investigation into behavioural patterns and their influence on accident frequency. The use of GIS tools for spatial analysis and the integration of additional data sources could help identify even more targeted interventions. This study serves as a stepping stone toward more proactive and data-driven approaches to traffic safety, providing a basis for both policy formulation and future research.

6.2 Conclusion

The analysis sheds light on critical patterns and trends in traffic accidents, providing actionable insights that can inform efforts to improve road safety and minimize casualties. Weekends and evenings emerged as peak accident periods, suggesting that factors like increased leisure-related traffic, reduced vigilance, and driver fatigue may contribute to the heightened risk during these times. Addressing these issues could involve deploying more traffic patrols, implementing dynamic traffic management systems during high-risk hours, and launching public awareness campaigns to promote safer driving behaviors at these times.Speeding and distracted driving were identified as the predominant causes of accidents, emphasizing the importance of tackling these issues through multifaceted interventions. Public awareness campaigns targeting these behaviors can educate drivers on the associated risks, while stricter enforcement measures, such as the use of speed cameras and penalties for mobile phone use while driving, can deter unsafe practices. Additionally, infrastructure improvements, such as traffic calming measures (e.g., speed bumps or narrower lanes) and enhanced signage, can create environments that encourage safer driving behavior.The study also highlighted accident hotspots, with a concentration in urbanized areas and regions characterized by poorly designed infrastructure. These findings point to the need for infrastructure audits to identify high-risk locations and implement corrective measures. Improvements could include redesigning intersections to reduce conflict points, enhancing lighting in poorly illuminated areas, and adopting smart traffic systems that adjust signals in real-time based on traffic conditions.

REFERENCES

[1] GeoPandas Documentation. (2023). *Geospatial data in Python made easy*. Retrieved from https://geopandas.org

[2] Li, Y., Wang, F., Zhang, L. "A Comprehensive Study on Traffic Accident Prediction Models," *International Journal of Transportation Science and Technology*, Vol. 7, 2018, pp. 210–225.

[3] Matplotlib Development Team. (2023). *Matplotlib: Visualization with Python*. Retrieved from https://matplotlib.org

[4] National Highway Traffic Safety Administration (NHTSA). (2022). *Traffic Safety Facts Annual Report*. Retrieved from https://www.nhtsa.gov

[5] Pandas Development Team. (2023). pandas: Powerful Python data analysis

[6] Peden, M., Scurfield, R., Sleet, D., Mohan, D., Hyder, A.A., Jarawan, E., Mathers, C. (Eds.). *World Report on Road Traffic Injury Prevention*, World Health Organization, Geneva, 2004, pp. 5–40.

[7] Python Software Foundation. (2023). *Python Language Reference, version 3.10*. Retrieved from https://www.python.org

[8] World Health Organization (WHO). (2021). *Global status report on road safety*. Retrieved from https://www.who.int/publications

APPENDIX

A. Source Code

SQL:

```
create database trafficaccident
use trafficaccident
(Inserting data check fig:6.1 to 6.7)
```

Python:

```
import pyodbc
import pandas as pd
import matplotlib.pyplot as plt
server_name = 'LAPTOP-TMANHNM8\\SQLEXPRESS01'
database_name = 'trafficaccident'
consql = pyodbc.connect('DRIVER={ODBC driver 17 for SQL Server} ;\
                        SERVER=' + server_name + ' ;\
                        DATABASE=' + database_name + ' ;\
                        Trusted_Connection=yes;')
print(consql)
print("\nSQL Server Connection is Successful")
cursor = consql.cursor()
print("\nCursor object created...")
cursor.execute("select * from traffic_accidents ")
print("\nTable records are retrieved from the SQL Server")
rows = cursor.fetchall()
columns = [column[0] for column in cursor.description]
df = pd.DataFrame.from_records(rows, columns=columns)
print("\nDataFrame Preview:")
print(df.head())

if 'Date' in df.columns:
    df['Date'] = pd.to_datetime(df['Date'])
df['crash_date'] = pd.to_datetime(df['crash_date'], errors='coerce')

#Accidents Over Time
daily = df['crash_date'].value_counts().sort_index()
daily.plot(kind='line', figsize=(10, 4))
plt.title("Line Chart: Accidents Over Time")
```

```
plt.xlabel("Date")
plt.ylabel("Number of Accidents")
plt.grid()
plt.tight_layout()
plt.show()

#Accidents by Month
df['crash_month'].value_counts().sort_index().plot(kind='line', marker='o',
color='purple')
plt.title("Line Chart: Accidents by Month")
plt.xlabel("Month")
plt.ylabel("Number of Accidents")
plt.tight_layout()
plt.show()

#Top 10 Weather Conditions
df['weather_condition'].value_counts().head(10).plot(kind='bar',
color='orange')
plt.title("Bar Chart: Weather Conditions")
plt.xlabel("Weather")
plt.ylabel("Count")
plt.tight_layout()
plt.show()

#Primary Cause of Accidents
cause = df['prim_contributory_cause'].value_counts().head(5)
plt.pie(cause, labels=cause.index, autopct='%1.1f%%', startangle=90,
wedgeprops={'width': 0.4})
plt.title("Donut Chart: Primary Cause")
plt.tight_layout()
plt.show()

#Injury Severity
df['most_severe_injury'].value_counts().plot(kind='pie',
autopct='%1.1f%%')
plt.title("Pie Chart: Severity of Accidents")
plt.ylabel("")
plt.tight_layout()
plt.show()
```

```
#Vehicles Involved
df['num_units'].value_counts().sort_index().head(10).plot(kind='bar',
color='olive')
plt.title("Bar Chart: Vehicles Involved per Accident")
plt.xlabel("Vehicles")
plt.ylabel("Accident Count")
plt.tight_layout()
plt.show()

#Injuries Over Time
df_grouped = df.groupby('crash_date')['injuries_total'].sum()
df_grouped.plot(kind='line', color='red')
plt.title("Line Chart: Injuries Over Time")
plt.xlabel("Date")
plt.ylabel("Total Injuries")
plt.tight_layout()
plt.show()
```

B. Screenshots

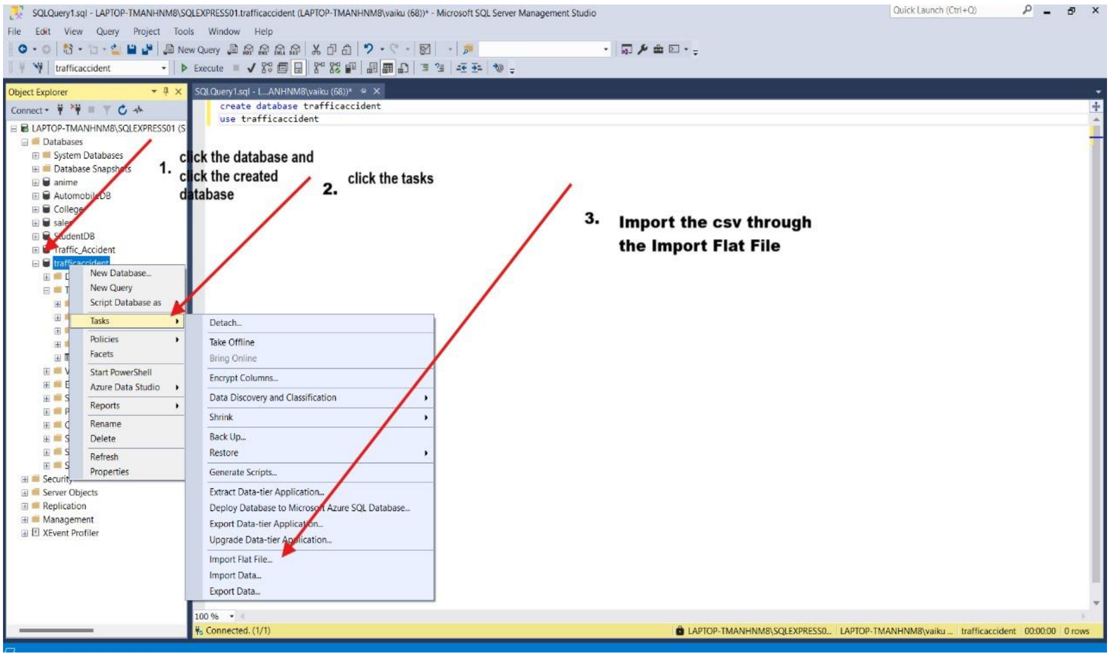

Fig:6.1: Inserting values into SQL by CSV file

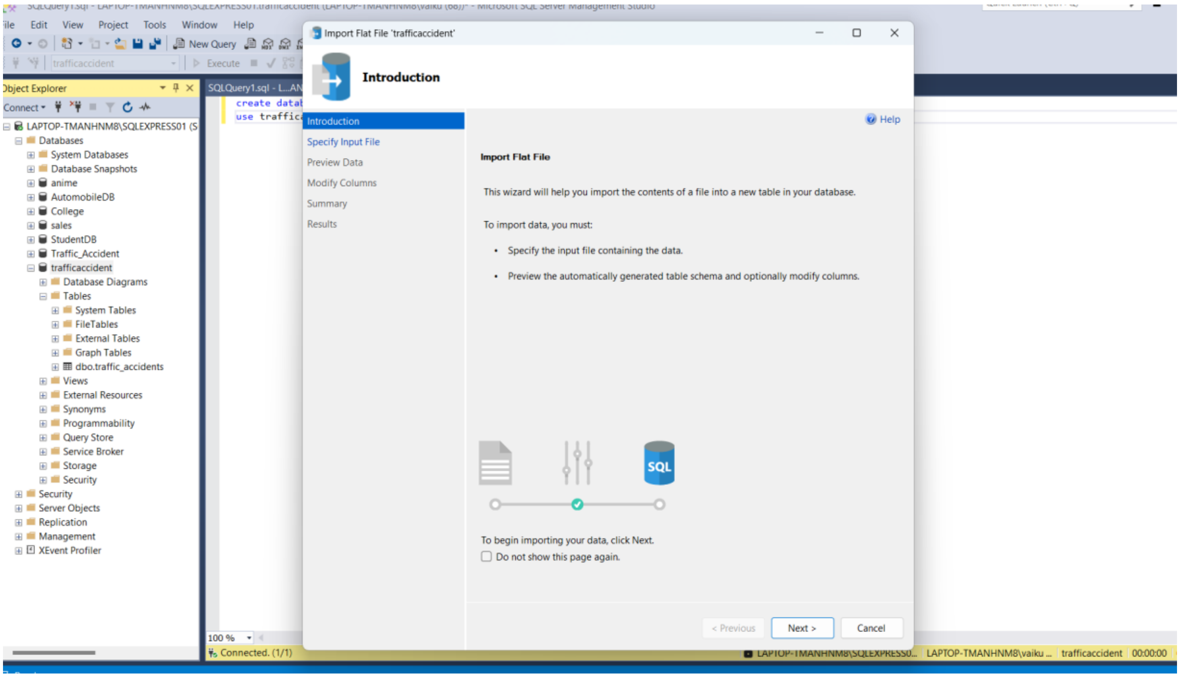

Fig:6.2: Importing through Import Flat Flies

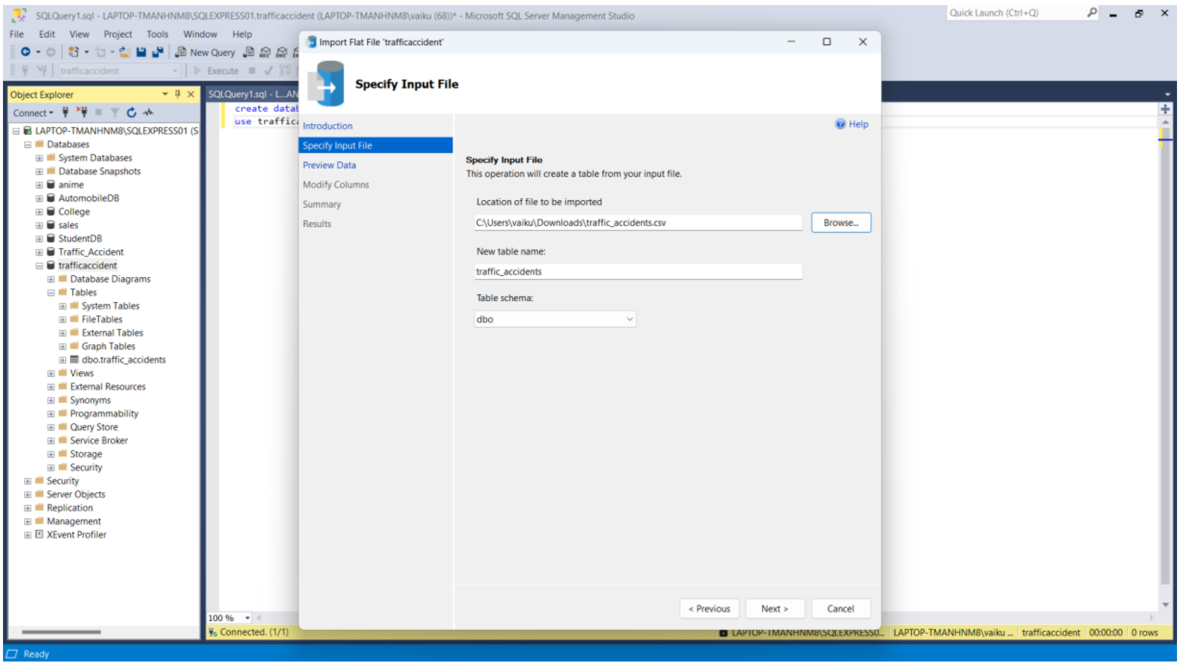

Fig:6.3: Specify Input File

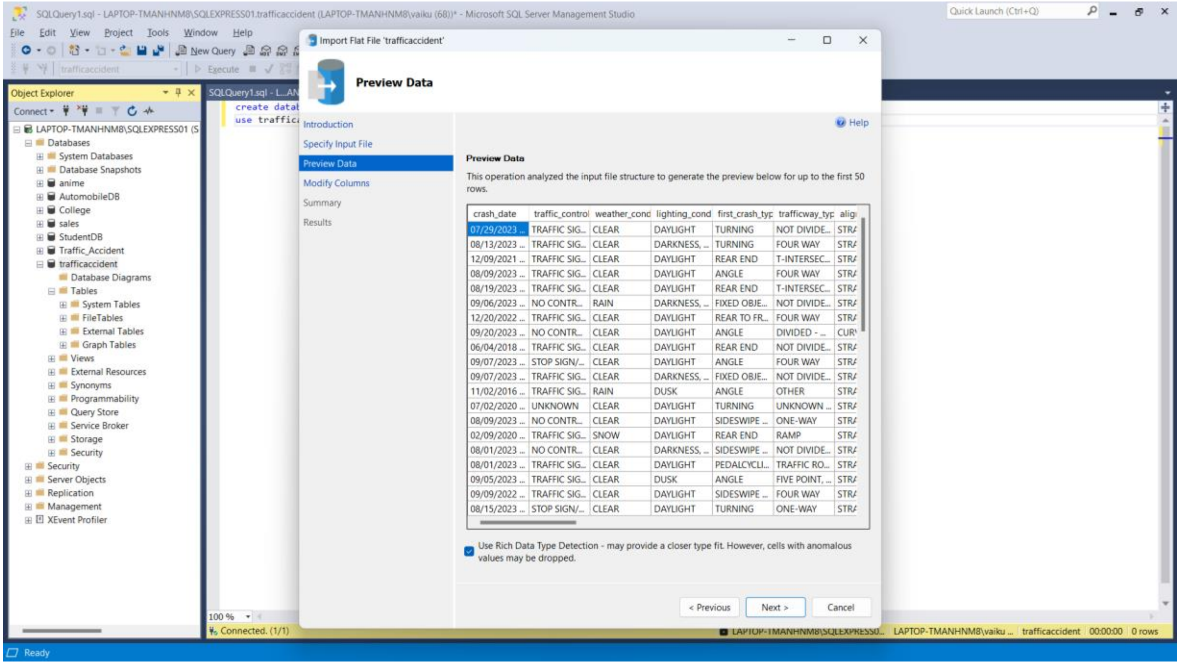

Fig:6.4: Preview Data

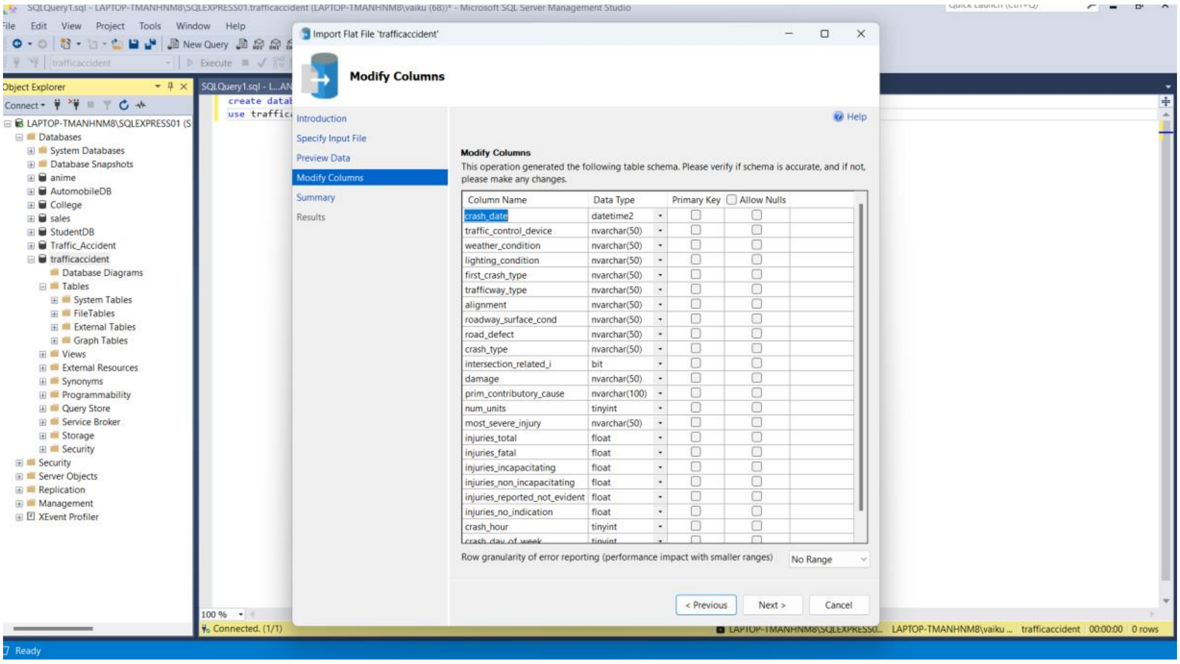

Fig:6.5: Modify Columns

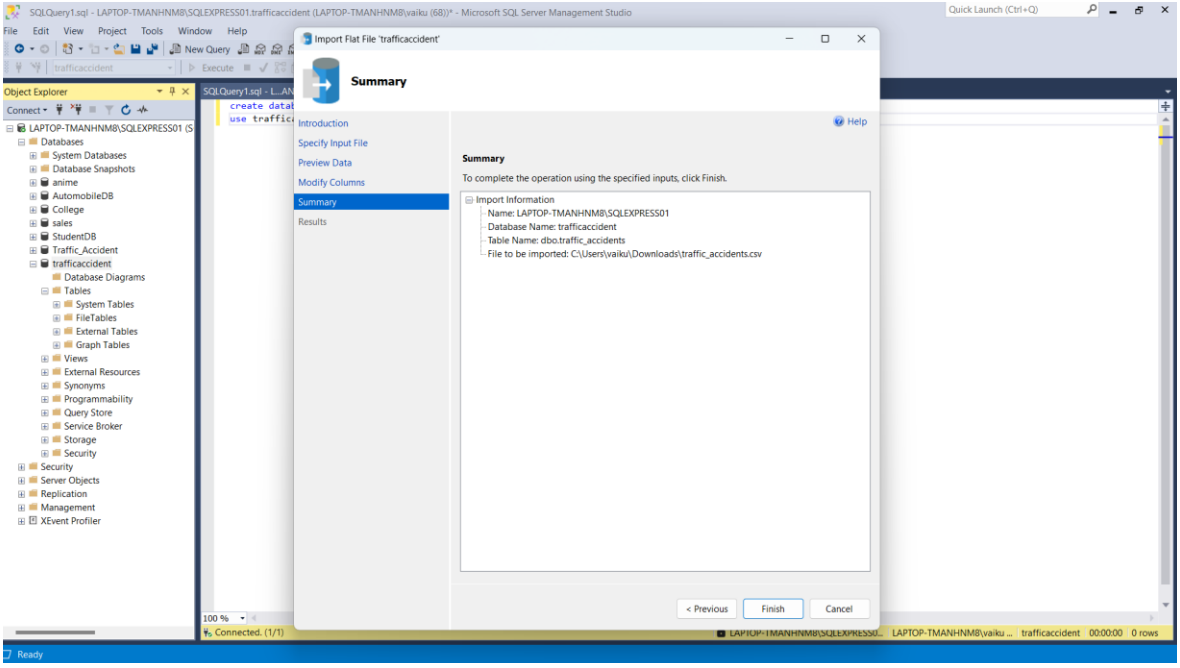

Fig:6.6: Summary of the Flat file

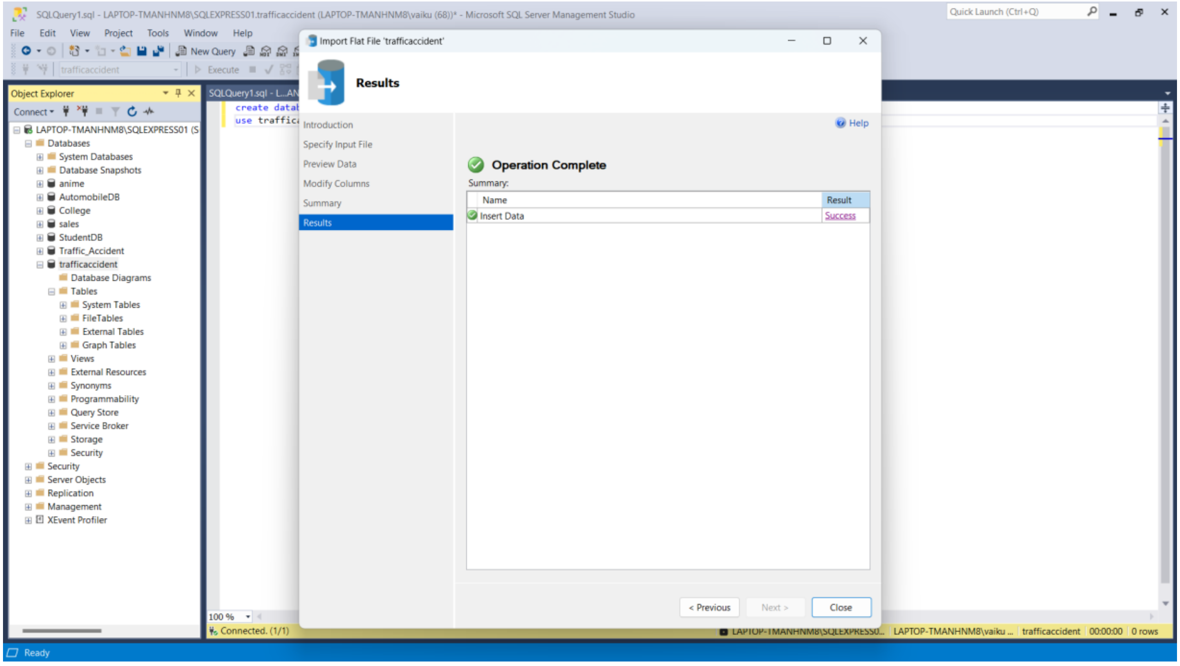

Fig:6.7: Result of the Flat file

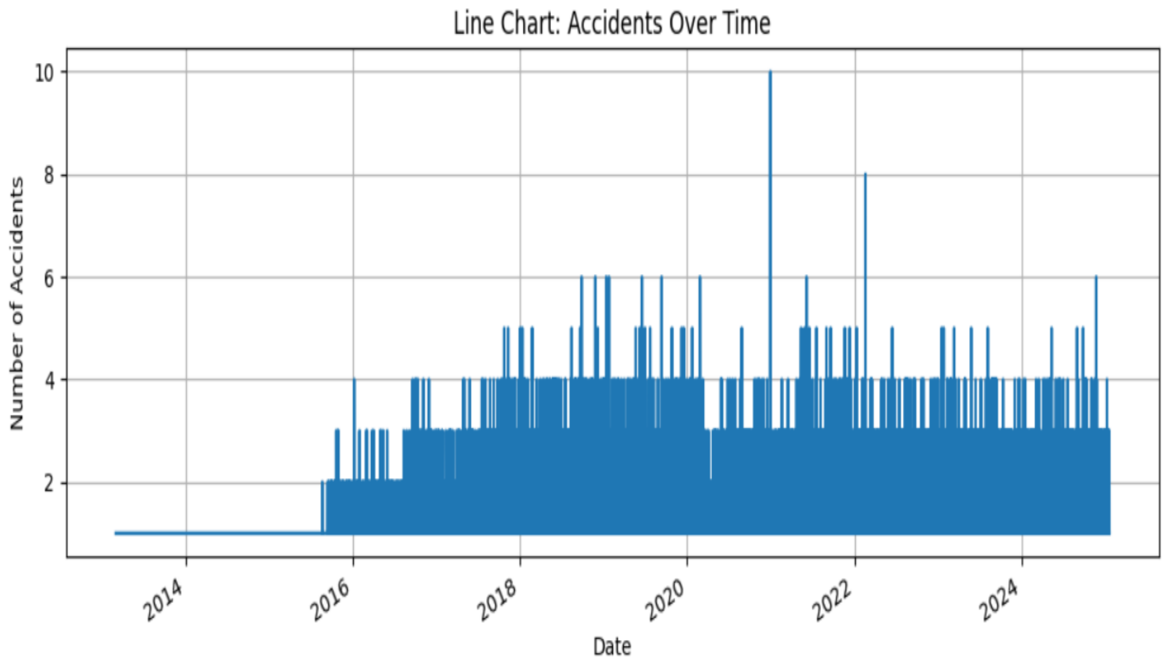

Fig:6.8: Line Chart: Accidents Over Time

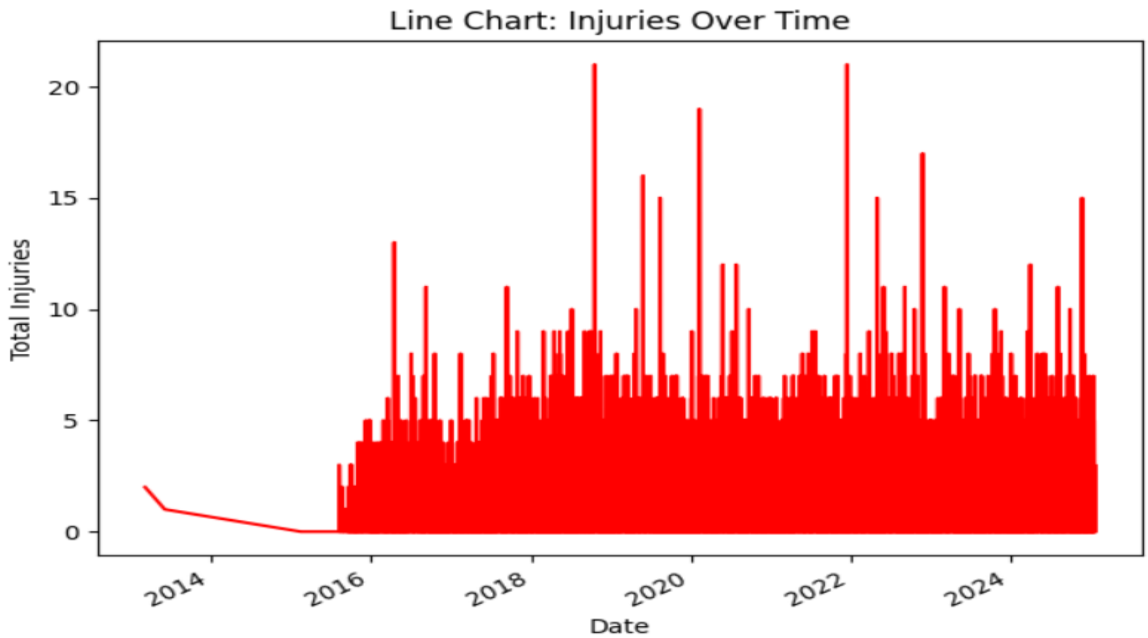

Fig:6.9: Line Chart: Accidents by Month

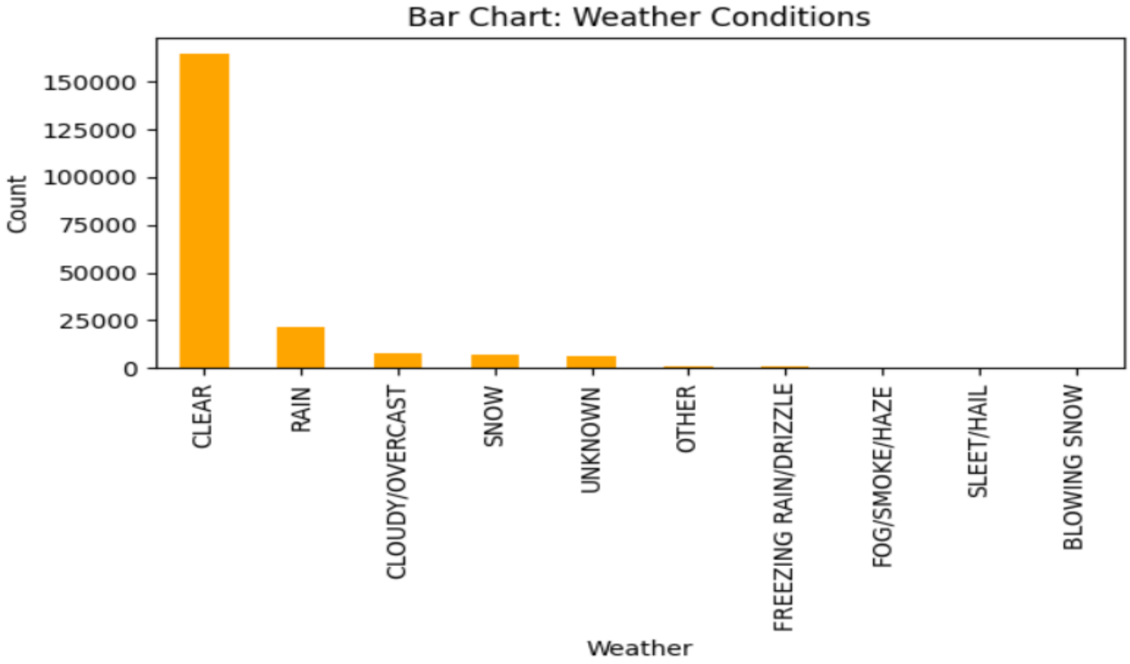

Fig:6.10: Bar Chart: Weather Conditions

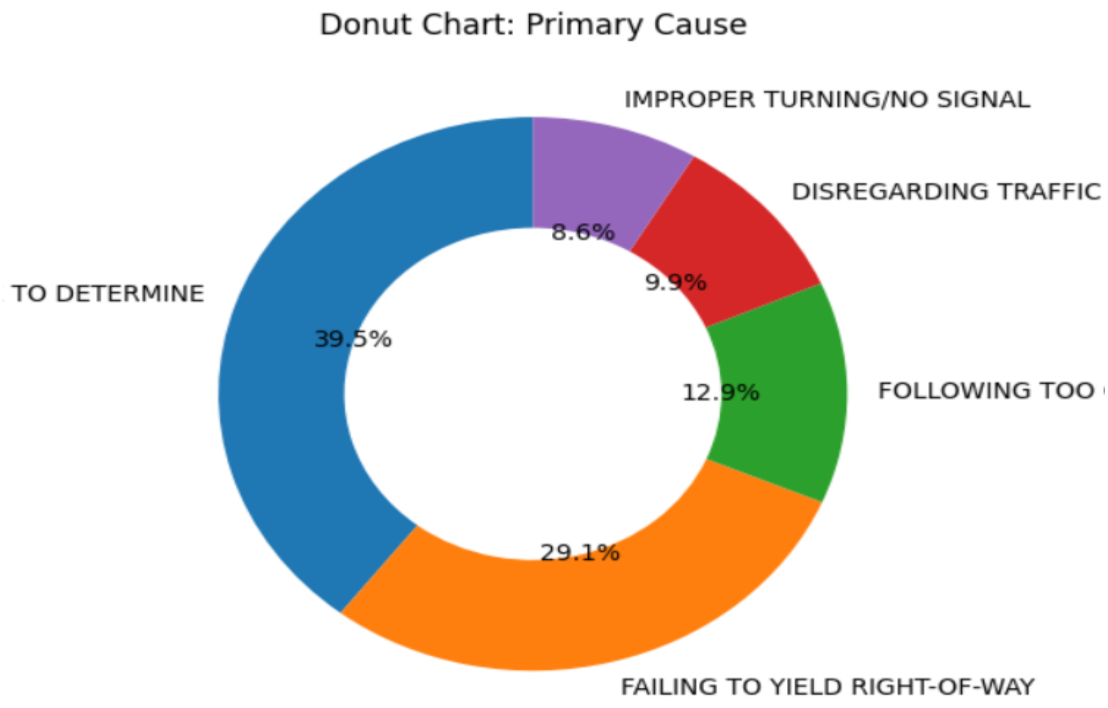

Fig:6.11: Donut Chart: Primary Cause

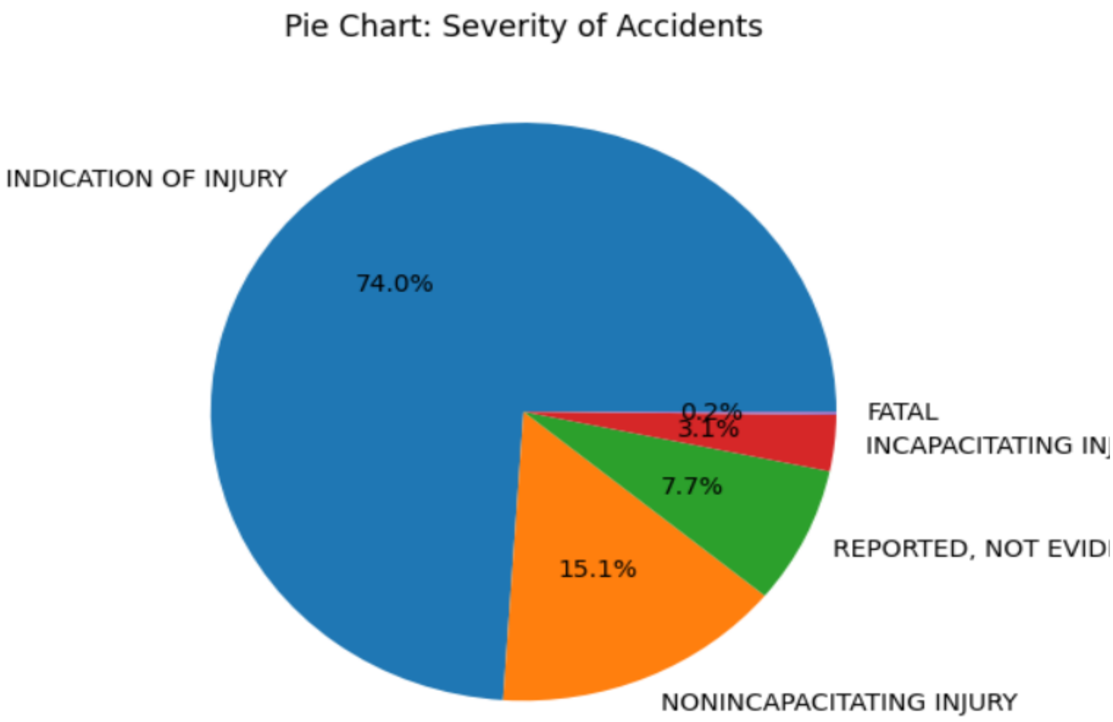

Fig:6.12: Pie Chart: Severity of Accidents

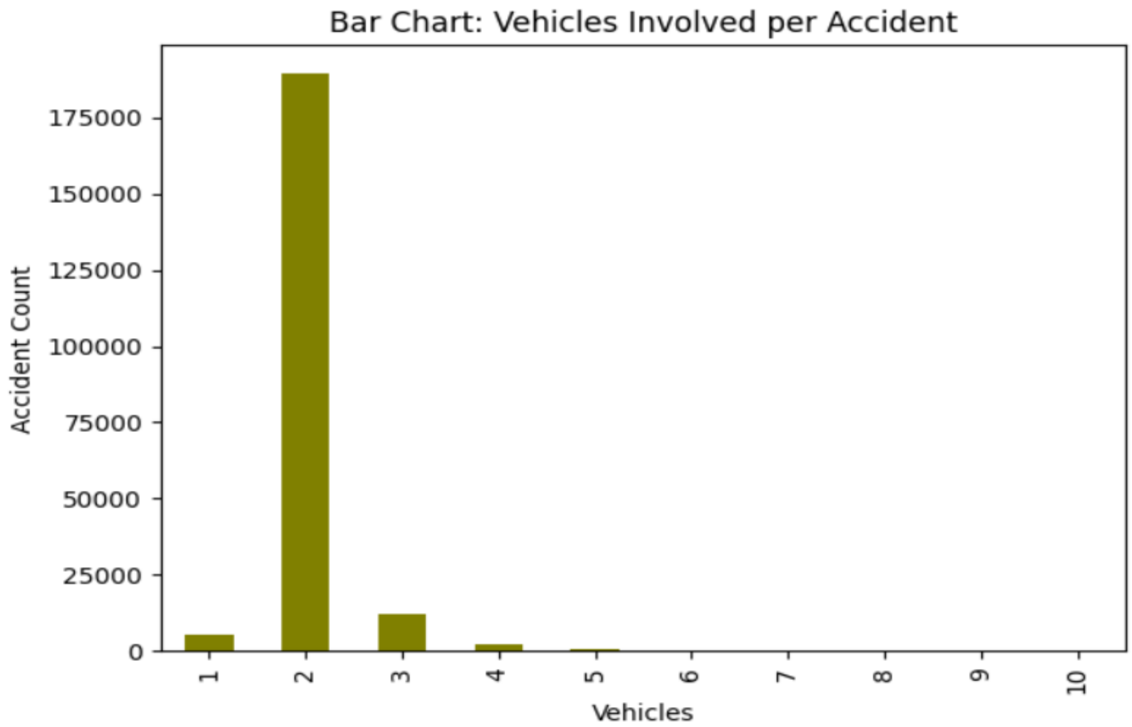

Fig:6.13: Bar Chart: Vehicles Involved per Accident

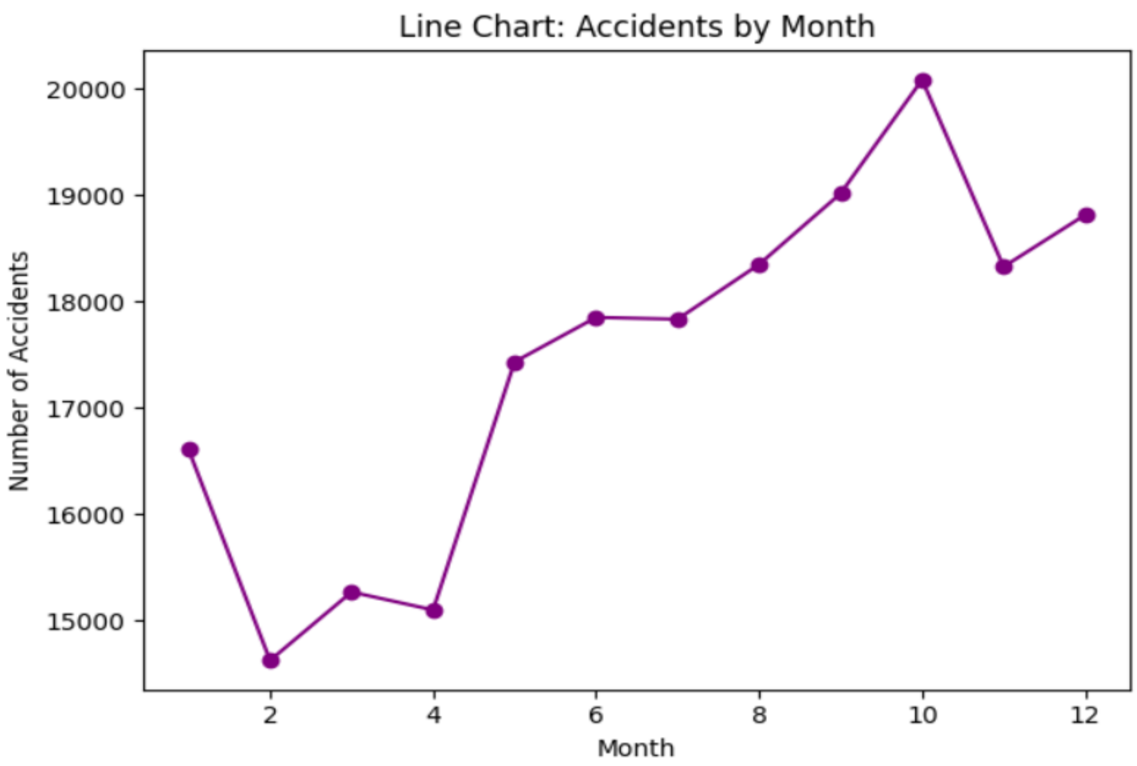

Fig:6.14: Line Chart: Injuries Over Time

Table:6.1: Sample Dataset

SQLQuery1.sql - L...ANHNM8\vaiku (53))*

```
use trafficaccident
select * from traffic accidents
```

100 %

Results Messages

	crash_date	traffic_control_device	weather_condition	lighting_condition	first_crash_type	trafficway_type	alignment	roadway_surface_cond	road_defect	crash_type	intersection_related_i	damag
1	2023-07-29 13:00:00.0000000	TRAFFIC SIGNAL	CLEAR	DAYLIGHT	TURNING	NOT DIVIDED	STRAIGHT AND LEVEL	UNKNOWN	UNKNOWN	NO INJURY / DRIVE AWAY	1	$501 -
2	2023-08-13 00:11:00.0000000	TRAFFIC SIGNAL	CLEAR	DARKNESS, LIGHTED ROAD	TURNING	FOUR WAY	STRAIGHT AND LEVEL	DRY	NO DEFECTS	NO INJURY / DRIVE AWAY	1	OVER
3	2021-12-09 10:30:00.0000000	TRAFFIC SIGNAL	CLEAR	DAYLIGHT	REAR END	T-INTERSECTION	STRAIGHT AND LEVEL	DRY	NO DEFECTS	NO INJURY / DRIVE AWAY	1	$501 -
4	2023-08-09 19:55:00.0000000	TRAFFIC SIGNAL	CLEAR	DAYLIGHT	ANGLE	FOUR WAY	STRAIGHT AND LEVEL	DRY	NO DEFECTS	INJURY AND / OR TOW DUE TO CRASH	1	OVER
5	2023-08-19 14:55:00.0000000	TRAFFIC SIGNAL	CLEAR	DAYLIGHT	REAR END	T-INTERSECTION	STRAIGHT AND LEVEL	UNKNOWN	UNKNOWN	NO INJURY / DRIVE AWAY	1	$501 -
6	2023-09-06 00:59:00.0000000	NO CONTROLS	RAIN	DARKNESS, LIGHTED ROAD	FIXED OBJECT	NOT DIVIDED	STRAIGHT AND LEVEL	WET	UNKNOWN	INJURY AND / OR TOW DUE TO CRASH	0	$501 -
7	2022-12-20 11:45:00.0000000	TRAFFIC SIGNAL	CLEAR	DAYLIGHT	REAR TO FRONT	FOUR WAY	STRAIGHT AND LEVEL	DRY	NO DEFECTS	NO INJURY / DRIVE AWAY	1	$501 -
8	2023-09-20 14:38:00.0000000	NO CONTROLS	CLEAR	DAYLIGHT	ANGLE	DIVIDED - W/MEDIAN (NOT RAISED)	CURVE, LEVEL	DRY	NO DEFECTS	INJURY AND / OR TOW DUE TO CRASH	1	OVER
9	2018-06-04 18:42:00.0000000	TRAFFIC SIGNAL	CLEAR	DAYLIGHT	REAR END	NOT DIVIDED	STRAIGHT AND LEVEL	DRY	NO DEFECTS	NO INJURY / DRIVE AWAY	1	OVER
10	2023-09-07 17:30:00.0000000	STOP SIGN/FLASHER	CLEAR	DAYLIGHT	ANGLE	FOUR WAY	STRAIGHT AND LEVEL	DRY	NO DEFECTS	NO INJURY / DRIVE AWAY	1	OVER
11	2023-09-07 20:32:00.0000000	TRAFFIC SIGNAL	CLEAR	DARKNESS, LIGHTED ROAD	FIXED OBJECT	NOT DIVIDED	STRAIGHT AND LEVEL	DRY	UNKNOWN	NO INJURY / DRIVE AWAY	1	OVER
12	2016-11-02 16:49:00.0000000	TRAFFIC SIGNAL	RAIN	DUSK	ANGLE	OTHER	STRAIGHT AND LEVEL	WET	NO DEFECTS	NO INJURY / DRIVE AWAY	1	OVER
13	2020-07-02 17:45:00.0000000	UNKNOWN	CLEAR	DAYLIGHT	TURNING	UNKNOWN INTERSECTION TYPE	STRAIGHT AND LEVEL	DRY	NO DEFECTS	NO INJURY / DRIVE AWAY	1	OVER
14	2023-08-09 07:00:00.0000000	NO CONTROLS	CLEAR	DAYLIGHT	SIDESWIPE SAME DIRECTION	ONE-WAY	STRAIGHT AND LEVEL	DRY	NO DEFECTS	NO INJURY / DRIVE AWAY	1	OVER
15	2020-02-09 14:00:00.0000000	TRAFFIC SIGNAL	SNOW	DAYLIGHT	REAR END	RAMP	STRAIGHT AND LEVEL	SNOW OR SLUSH	NO DEFECTS	NO INJURY / DRIVE AWAY	1	OVER
16	2023-08-01 23:45:00.0000000	NO CONTROLS	CLEAR	DARKNESS, LIGHTED ROAD	SIDESWIPE OPPOSITE DIRECTION	NOT DIVIDED	STRAIGHT AND LEVEL	DRY	NO DEFECTS	INJURY AND / OR TOW DUE TO CRASH	0	OVER
17	2023-08-01 15:29:00.0000000	TRAFFIC SIGNAL	CLEAR	DAYLIGHT	PEDALCYCLIST	TRAFFIC ROUTE	STRAIGHT AND LEVEL	DRY	NO DEFECTS	INJURY AND / OR TOW DUE TO CRASH	0	$500 (
18	2023-09-05 19:05:00.0000000	TRAFFIC SIGNAL	CLEAR	DUSK	ANGLE	FIVE POINT, OR MORE	STRAIGHT AND LEVEL	DRY	NO DEFECTS	INJURY AND / OR TOW DUE TO CRASH	1	OVER
19	2022-09-09 16:53:00.0000000	TRAFFIC SIGNAL	CLEAR	DAYLIGHT	SIDESWIPE SAME DIRECTION	FOUR WAY	STRAIGHT AND LEVEL	DRY	NO DEFECTS	NO INJURY / DRIVE AWAY	1	OVER
20	2023-08-15 18:50:00.0000000	STOP SIGN/FLASHER	CLEAR	DAYLIGHT	TURNING	ONE-WAY	STRAIGHT AND LEVEL	UNKNOWN	UNKNOWN	NO INJURY / DRIVE AWAY	1	$500 (
21	2023-08-22 15:50:00.0000000	NO CONTROLS	CLEAR	DAYLIGHT	REAR END	NOT DIVIDED	STRAIGHT AND LEVEL	DRY	NO DEFECTS	NO INJURY / DRIVE AWAY	1	OVER
22	2023-08-02 20:53:00.0000000	TRAFFIC SIGNAL	CLEAR	DARKNESS, LIGHTED ROAD	TURNING	FOUR WAY	STRAIGHT AND LEVEL	DRY	NO DEFECTS	NO INJURY / DRIVE AWAY	1	OVER
23	2023-09-06 14:50:00.0000000	STOP SIGN/FLASHER	CLEAR	DAYLIGHT	TURNING	T-INTERSECTION	STRAIGHT AND LEVEL	DRY	NO DEFECTS	INJURY AND / OR TOW DUE TO CRASH	1	OVER
24	2022-07-01 01:00:00.0000000	TRAFFIC SIGNAL	CLEAR	DARKNESS	TURNING	ONE-WAY	STRAIGHT AND LEVEL	DRY	UNKNOWN	NO INJURY / DRIVE AWAY	1	OVER
25	2023-09-06 14:30:00.0000000	STOP SIGN/FLASHER	CLEAR	DAYLIGHT	REAR END	T-INTERSECTION	STRAIGHT AND LEVEL	DRY	NO DEFECTS	NO INJURY / DRIVE AWAY	1	OVER
26	2023-08-17 13:30:00.0000000	STOP SIGN/FLASHER	CLEAR	DAYLIGHT	TURNING	FOUR WAY	STRAIGHT AND LEVEL	DRY	NO DEFECTS	NO INJURY / DRIVE AWAY	1	OVER
27	2022-03-20 15:08:00.0000000	NO CONTROLS	CLEAR	DAYLIGHT	TURNING	FOUR WAY	STRAIGHT AND LEVEL	DRY	NO DEFECTS	NO INJURY / DRIVE AWAY	1	$501 -
28	2023-09-25 12:00:00.0000000	STOP SIGN/FLASHER	CLOUDY/OVERC...	UNKNOWN	TURNING	OTHER	STRAIGHT AND LEVEL	UNKNOWN	UNKNOWN	NO INJURY / DRIVE AWAY	1	OVER
29	2023-09-25 16:01:00.0000000	STOP SIGN/FLASHER	CLEAR	DAYLIGHT	TURNING	T-INTERSECTION	STRAIGHT AND LEVEL	DRY	NO DEFECTS	INJURY AND / OR TOW DUE TO CRASH	1	OVER
30	2023-08-18 15:00:00.0000000	STOP SIGN/FLASHER	CLEAR	DAYLIGHT	ANGLE	DIVIDED - W/MEDIAN (NOT RAISED)	STRAIGHT AND LEVEL	DRY	NO DEFECTS	NO INJURY / DRIVE AWAY	1	OVER