INT374 – DATA ANALYTICS WITH POWER BI PROJECT REPORT

(Project Semester January-April 2025)

Motor Vehicle Accident Analysis

Submitted by

Piyush Saini

12310960

B.Tech CSE K23DG

Course Code – INT374

Under the Guidance of

Aashima (UID 28968)

Discipline of CSE/IT

Lovely School of Computer Science and Engineering Lovely Professional University, Phagwara

CERTIFICATE

This is to certify that Piyush Saini bearing Registration no.12310960 has completed INT374 project titled, "Motor Vehicle Accident Analysis" under my guidance and supervision. To the best of my knowledge, the present work is the result of his/her original development, effort and study.

Signature and Name of the Supervisor Designation of the Supervisor

School of Computer Science and Engineering

Lovely Professional University Phagwara, Punjab.

Date: 11/12/25

DECLARATION

I, Piyush Saini student of Computer Science and Engineering under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date: 11/12/25 Signature

Registration No.12310960 Piyush Saini

ACKNOWLEDGEMENT

I would like to take this opportunity to express my sincere gratitude and appreciation to all those individuals and institutions whose guidance, support, and encouragement contributed to the successful completion of this Minor Project. This project has been a significant learning experience, and its completion would not have been possible without the cooperation and assistance of many people who supported me throughout this academic journey.

First and foremost, I express my deepest sense of gratitude to my project supervisor, Ms. Aashima, for her constant guidance, invaluable suggestions, and continuous encouragement during every phase of this project. Her in-depth knowledge, technical expertise, and insightful feedback helped me develop a clear understanding of the subject and guided me in the right direction whenever challenges arose. She patiently reviewed my work, provided constructive criticism, and motivated me to improve both the technical and analytical aspects of this project. Her mentorship has been a source of inspiration and played a pivotal role in shaping the quality and outcome of this work.

I am also thankful to the respected faculty members of the Department of Computer Science and Engineering for imparting essential theoretical knowledge and practical insights throughout the course of my study. The concepts learned during lectures and laboratory sessions laid a strong foundation that proved extremely useful while working on this project. Their dedication to teaching and commitment to academic excellence greatly influenced my learning process.

I would like to extend my sincere gratitude to Lovely Professional University for providing a well-structured academic environment and the necessary infrastructure required for successful project execution. The availability of modern laboratories, computing resources, library facilities, and online learning platforms enabled me to explore, analyze, and implement various aspects of the project effectively. The university’s emphasis on project-based learning helped me enhance my technical skills, problem- solving abilities, and overall professional competence.

I am grateful to my classmates and friends for their cooperation, valuable discussions, and moral support throughout the project duration. Collaborative interactions, knowledge sharing, and brainstorming sessions with peers helped me gain different perspectives and overcome various technical and conceptual challenges. Their encouragement and friendly support created a positive learning environment that kept me motivated during demanding phases of the project.

I would also like to acknowledge the contributions of researchers, authors, and educators whose books, research papers, journals, and online resources served as important references for this project. Their work provided essential insights and theoretical understanding that guided the design, analysis, and implementation stages of this project.

I express my heartfelt gratitude to my family members for their unwavering support, patience, and motivation throughout my academic journey. Their constant encouragement, emotional backing, and belief in my abilities gave me the confidence to pursue my goals with dedication and determination. They have always been a source of strength, especially during challenging times.

Finally, I would like to thank everyone who has directly or indirectly contributed to the completion of this Minor Project. This project has not only enhanced my technical knowledge but has also helped me develop critical thinking, analytical skills, and a disciplined approach towards problem-solving. I consider this project a valuable milestone in my academic career and hope that the knowledge and experience gained will be beneficial in my future professional endeavors.

Piyush Saini

B.Tech – Computer Science and Engineering Lovely Professional University

CHAPTER 1: INTRODUCTION

Road traffic accidents have emerged as a major public safety and socio-economic concern worldwide. Every year, millions of accidents result in loss of life, serious injuries, and long-term disabilities. Rapid urbanization, increasing vehicle density, and inadequate traffic management contribute significantly to accident occurrences.

With the availability of large-scale accident datasets and modern analytics tools, it has become possible to analyze accident patterns in depth. **Microsoft Power BI** is a powerful business intelligence tool that enables transformation of raw accident data into interactive dashboards and actionable insights.

This project titled **"Motor Vehicle Accidents Analysis Dashboard"** focuses on analyzing motor vehicle crash data to identify **trends, high-risk areas, time patterns, vehicle involvement, and the impact on vulnerable road users such as pedestrians, cyclists, and bikers**.

Dataset: https://catalog.data.gov/dataset/motor-vehicle-collisions-crashes

LinkedIn: https://www.linkedin.com/posts/piyush-saini-666720299_powerbi-dataanalytics-dashboarddesign-activity-7406375144614453249-wUKu?utm_source=social_share_send&utm_medium=android_app&rcm=ACoAAEfi8pMBHctra_B27qI58QA0jdeXly3uuk0&utm_campaign=share_via

GitHub: https://github.com/PIYUSHSAINI24/Motor-Vehicle-Accident-Analysis

CHAPTER 2: OBJECTIVES OF THE PROJECT (EXPANDED)

The primary objectives of this project are:

- To analyze motor vehicle accident data using Power BI
- To study accident trends over different years and months
- To identify high-risk time bands and days
- To analyze area-wise accident and fatality distribution
- To examine the impact of accidents on pedestrians, cyclists, and bikers
- To identify major contributing factors and vehicle types involved
- To design an interactive dashboard for effective data-driven insights
- To demonstrate practical application of data analytics concepts

CHAPTER 3: OVERVIEW OF THE DATASET

The dataset used in this project comprises detailed **motor vehicle collision records**, where each row represents a **single road accident event**. The dataset captures comprehensive information related to the **time, location, vehicles involved, contributing factors, and the resulting human impact** of each accident. This structured format enables systematic analysis of accident patterns and severity across multiple dimensions.

The dataset is sourced from a reliable public data repository and reflects **real-world traffic accident scenarios**. Due to its large size and diversity of attributes, the dataset is highly suitable for **descriptive, comparative, and exploratory data analysis** using business intelligence tools such as Power BI.

One of the key strengths of this dataset is the inclusion of **geographical coordinates (latitude and longitude)**, which allows advanced **geospatial analysis** through map-based visualizations. Additionally, the presence of injury and fatality-related fields enables detailed assessment of **accident severity and risk levels**, making the dataset ideal for road safety analytics.

Overall, the dataset provides a strong foundation for understanding **accident trends, high-risk locations, vulnerable road users, and contributing factors**, thereby supporting meaningful insights and data-driven decision-making.

3.3 Description of Dataset Columns

The dataset consists of several important columns that can be broadly categorized into **temporal, geographical, vehicle-related, and impact-related attributes**. Each category plays a vital role in supporting multidimensional analysis.

---

1. Temporal Attributes

Temporal attributes provide information related to the **timing of accidents** and are essential for trend and pattern analysis over different periods.

- **CRASH DATE**
  Represents the exact calendar date on which the accident occurred.
  This attribute is used for **year-wise, month-wise, and date-level trend analysis**, helping identify long-term changes and seasonal patterns in accident occurrences.
- **CRASH TIME**
  Indicates the specific time at which the accident occurred.
  This column enables **hourly analysis** and supports the creation of **time bands** (such as morning, afternoon, evening, and night) to identify high-risk time periods.
- **Year, Month, and Day Name (Derived Columns)**
  These columns are extracted from the crash date to enhance analytical capabilities.
  - **Year** supports longitudinal trend analysis.
  - **Month** helps identify seasonal variations in accident frequency.
  - **Day Name** (Monday, Tuesday, etc.) assists in analyzing weekday and weekend accident patterns.

---

2. Geographical Attributes

Geographical attributes describe the **location of accidents** and enable spatial and area-based analysis.

- **BOROUGH**
  Specifies the administrative area where the accident occurred.
  This column is used for **area-wise comparison**, allowing identification of regions with higher accident frequency or fatality rates.
- **ZIP CODE**
  Provides a more granular location detail within each borough.
  This attribute supports localized analysis and helps identify accident-prone zones at a finer level.
- **LATITUDE and LONGITUDE**
  These numerical coordinates represent the exact geographical location of each accident.
  They are used in **map and density visualizations** to identify accident hotspots and visualize spatial distribution patterns across regions.

---

3. Vehicle & Cause Attributes

These attributes describe the **vehicles involved in accidents** and the **primary contributing causes**, enabling cause-and-effect analysis.

- **VEHICLE TYPE CODE 1**
  Indicates the type of vehicle involved in the accident, such as sedan, SUV, motorcycle, or bicycle.
  This column is crucial for analyzing **vehicle-wise accident involvement** and understanding which vehicle categories are more frequently associated with accidents.
- **CONTRIBUTING FACTOR VEHICLE 1**
  Represents the primary factor contributing to the accident, such as distracted driving, speeding, or failure to yield.
  This attribute helps identify **major causes of accidents**, supporting risk assessment and preventive insights.

---

4. Impact Attributes

Impact attributes capture the **human consequences of accidents**, making them essential for severity and safety analysis.

- **Persons Injured and Persons Killed**
  These columns record the total number of individuals injured or killed in each accident.
  They are used to measure overall accident severity and calculate key performance indicators such as **total injuries, fatalities, and fatality rate**.
- **Pedestrians, Cyclists, and Motorists Injured & Killed**
  These attributes provide a detailed breakdown of casualties based on road user categories.

This classification enables focused analysis of **vulnerable road users**, helping assess the relative risk faced by pedestrians and cyclists compared to motor vehicle occupants.

3.4 Dataset Structure Summary

- Each row represents a **single motor vehicle accident event**
- Columns include a mix of **categorical and numerical attributes**
- The dataset supports **aggregation, filtering, and drill-down analysis**
- The structure is suitable for **star schema data modeling** in Power BI

3.5 Suitability of the Dataset

The dataset is highly suitable for:

- Time-based trend analysis
- Area-wise and location-based comparisons
- Vehicle and cause-related accident analysis
- Severity assessment using injury and fatality metrics
- Geospatial visualization and hotspot identification

CHAPTER 4: DATA PREPROCESSING

4.2 Data Cleaning Process

Data cleaning involved identifying and correcting inaccuracies, inconsistencies, and redundancies present in the raw dataset. Since the dataset contained a large number of accident records collected over different time periods and locations, careful cleaning was necessary to ensure accurate analysis.

4.2.1 Removal of Irrelevant and Duplicate Columns

The original dataset included several columns that were not directly relevant to the objectives of the project or did not contribute meaningful analytical value.

The following steps were taken:

- Columns unrelated to accident analysis were removed
- Redundant identifier and metadata columns were eliminated
- Duplicate or repeated columns containing similar information were reviewed and removed

This step simplified the dataset, reduced memory usage, and improved overall dashboard performance.

4.2.2 Handling Missing Values

Missing values were present in both numerical and categorical columns. If left untreated, these could lead to incorrect aggregations and misleading visualizations.

The following strategies were applied:

- Missing values in **injury and fatality columns** were replaced with zero (0), as the absence of reported casualties indicates no recorded injuries or deaths
- Missing values in categorical columns such as **vehicle type** and **contributing factor** were replaced with the label **"Unknown"** to maintain consistency
- Records with missing or invalid **latitude and longitude** values were removed to prevent errors in map-based visualizations

These steps ensured that calculations and visuals were based on complete and reliable data.

4.3 Data Transformation

Data transformation was performed to convert the dataset into a standardized format suitable for efficient analysis and visualization.

4.3.1 Data Type Conversion

Several columns were initially stored in incorrect data formats. To address this issue, the following transformations were applied:

- **CRASH DATE** was converted to a Date data type
- **CRASH TIME** was converted to a Time data type
- Injury and fatality columns were converted to numeric (whole number) format
- **Latitude and Longitude** columns were converted to decimal numbers for accurate geospatial mapping
- Text-based fields such as borough names and contributing factors were converted to text format

Correct data typing was essential for applying DAX calculations and creating accurate visualizations.

4.3.2 Standardization of Categorical Fields

Inconsistencies were observed in categorical columns due to variations in spelling, capitalization, and naming conventions.

To standardize these fields:

- Leading and trailing spaces were removed
- Inconsistent naming patterns were corrected using replace and transform operations
- Similar categories were grouped under consistent labels

Standardization ensured accurate grouping and prevented duplication in charts and slicers.

4.4 Feature Engineering

Feature engineering was performed to derive additional attributes that enhance analytical depth and dashboard usability.

4.4.1 Creation of Derived Date Attributes

From the crash date, the following derived columns were created:

- **Year** – Used for long-term trend analysis
- **Month** and **Month Name** – Used to identify seasonal accident patterns
- **Day Name** – Used to compare weekday and weekend accident trends

These derived attributes support detailed time-based analysis.

4.4.2 Time Band Classification

To analyze accident patterns based on time of occurrence, a **Time Band** column was created using crash time:

Time Band	Time Range
Night	00:00 – 05:59
Morning	06:00 – 11:59
Afternoon	12:00 – 17:59
Evening	18:00 – 23:59

This classification helps identify **high-risk time periods** and supports comparative analysis across different times of the day.

4.5 Handling Geographical Data

Geographical attributes required special attention as they directly influenced map-based visualizations.

The following steps were taken:

- Records with missing, zero, or invalid latitude and longitude values were removed
- Geographical fields such as borough were assigned appropriate geographic data categories in Power BI
- Coordinates were validated to ensure correct plotting on maps

These steps ensured accurate representation of accident locations and reliable hotspot analysis.

4.6 Final Dataset Preparation and Validation

After completing data cleaning, transformation, and feature engineering, the final dataset was validated to ensure readiness for analysis.

Validation steps included:

- Verifying sample aggregations for injuries and fatalities
- Checking consistency across categorical fields
- Previewing visualizations to ensure accurate filtering and interaction

The final cleaned and structured dataset was then used for **data modeling and dashboard development**.

CHAPTER 5: DATA MODELING

Data modeling is a crucial step in transforming a cleaned dataset into a structured framework that supports **efficient analysis, accurate aggregations, and smooth interactivity** within Power BI. An effective data model improves report performance, simplifies relationships, and enables meaningful insights across multiple dimensions.

In this project, a **Star Schema data model** was implemented to analyze motor vehicle accident data. The star schema was chosen because of its **simplicity, scalability, and suitability for business intelligence applications**. This modeling approach reduces data redundancy, enhances query performance, and makes the dataset easier to understand and maintain.

5.1 Star Schema Design

The star schema consists of a **central fact table** connected to multiple **dimension tables**. The fact table stores quantitative, event-level data, while dimension tables provide descriptive context for analysis. This design allows users to analyze accident metrics from different perspectives such as **time, location, vehicle type, and contributing factors**.

The star schema structure enables efficient filtering, slicing, and drill-down operations across all visuals in the dashboard.

5.2 Fact Table

Fact_Crashes

The **Fact_Crashes** table serves as the central table in the data model. Each record in this table represents a **single motor vehicle accident event**. The fact table stores numerical measures that are used for calculations and key performance indicators throughout the dashboard.

Key Attributes of Fact_Crashes:

- Accident identifier (Collision ID)
- Date and time references
- Foreign keys linking to dimension tables
- Injury and fatality counts

Measures Stored in Fact_Crashes:

- Total persons injured
- Total persons killed
- Pedestrians injured and killed
- Cyclists injured and killed
- Motorists injured and killed

These measures form the foundation for all analytical insights, including severity analysis, trend analysis, and comparisons across categories.

5.3 Dimension Tables

Dimension tables provide descriptive information that adds context to the numerical data stored in the fact table. In this project, the following dimension tables were created:

5.3.1 Dim_Date

The **Dim_Date** table contains time-related attributes derived from the crash date.

Attributes include:

- Date
- Year
- Month
- Month Name
- Day
- Day Name
- Weekend indicator

This dimension enables **time-based analysis**, allowing users to study accident trends across years, months, and weekdays.

5.3.2 Dim_Time

The **Dim_Time** table focuses on the time of occurrence of accidents.

Attributes include:

- Crash Time
- Hour
- Time Band (Night, Morning, Afternoon, Evening)

This dimension supports **time-of-day analysis**, helping identify high-risk periods and peak accident hours.

5.3.3 Dim_Location

The **Dim_Location** table stores geographical information related to accident locations.

Attributes include:

- Borough
- ZIP Code
- Latitude
- Longitude
- Street Name

This dimension enables **area-wise and spatial analysis**, allowing identification of accident-prone regions and hotspots using map visualizations.

5.3.4 Dim_Vehicle

The **Dim_Vehicle** table captures information about the types of vehicles involved in accidents.

Attributes include:

- Vehicle Type

This dimension allows analysis of **vehicle-wise accident involvement**, helping identify which vehicle categories are more frequently associated with injuries or fatalities.

5.3.5 Dim_Cause

The **Dim_Cause** table represents the contributing factors responsible for accidents.

Attributes include:

- Contributing Factor

This dimension supports **cause-based analysis**, enabling identification of major risk factors such as distracted driving, speeding, or failure to yield.

5.4 Relationships and Model Efficiency

All dimension tables are connected to the **Fact_Crashes** table using **one-to-many relationships**, with dimension tables on the "one" side and the fact table on the "many" side. Single-directional filtering was applied to maintain model clarity and prevent ambiguity.

This relationship structure ensures:

- Accurate aggregations
- Efficient filtering and slicing
- Improved performance for large datasets
- Clear and maintainable data model design

5.5 Advantages of the Star Schema Model

The star schema model implemented in this project offers several benefits:

- Faster query performance
- Simplified DAX calculations
- Improved readability and maintainability
- Easy scalability for future enhancements
- Industry-standard modeling approach

CHAPTER 6 : DASHBOARD

This chapter describes the design, structure, and analytical components of the Motor Vehicle Accidents Analysis Dashboard developed using Microsoft Power BI. The dashboard transforms processed accident data into interactive visual representations that allow users to explore accident trends, severity, and risk factors efficiently.

The dashboard is designed with a dark theme to improve visual focus, reduce clutter, and highlight critical insights. It integrates multiple visual elements such as KPI cards, charts, maps, and slicers to provide a comprehensive and user-friendly analytical experience.

6.1 Dashboard Overview

The Motor Vehicle Accidents Analysis Dashboard provides a holistic view of road accident data, focusing on both overall accident severity and detailed breakdowns across time, location, vehicle type, and contributing factors.

The primary objectives of the dashboard are:

- **To present key accident statistics at a glance**
- **To identify high-risk locations and time periods**
- **To analyze the impact of accidents on different road users**
- **To support interactive exploration through filters and slicers**

The dashboard is structured in a logical layout to ensure ease of navigation and clarity of insights.

6.2 Key Performance Indicator (KPI) Cards

KPI cards are placed at the top section of the dashboard to provide an instant summary of accident severity and impact. These indicators enable users to quickly understand the overall scale of the problem.

The KPI cards display:

- **Total Accidents – Total number of recorded accident events**
- **Total Injured – Aggregate number of people injured in accidents**
- **Total Killed – Total number of fatalities resulting from accidents**
- **Pedestrian Injuries and Fatalities**
- **Cyclist Injuries and Fatalities**
- **Biker (Motorist) Injuries and Fatalities**

These KPIs help in comparing the relative impact of accidents on different categories of road users and highlight the vulnerability of pedestrians and cyclists.

6.3 Time-Based Analysis Visuals

Time-based analysis is a key component of the dashboard, as accident patterns often vary across different time periods.

Accidents by Year and Month

Line charts are used to visualize the trend of accidents and fatalities over time. These visuals help identify:

- **Long-term increases or decreases in accidents**
- **Seasonal patterns**
- **Periods of unusually high accident occurrences**

Time Band Analysis

Accidents are grouped into four time bands:

- **Night**
- **Morning**
- **Afternoon**
- **Evening**

Bar and slicer visuals allow users to compare accident frequency across these time bands and identify high-risk periods during the day.

6.4 Area and Location-Based Analysis

Geographical analysis plays a crucial role in understanding accident distribution.

Accidents by Area

Donut and bar charts are used to represent accident distribution across different boroughs. These visuals highlight:

- **Areas with consistently high accident counts**
- **Regions that require focused safety interventions**

Accident Density Map

A map visualization plots accident locations using latitude and longitude coordinates. The map helps identify:

- **Accident hotspots**
- **Densely affected regions**
- **Spatial clustering of accidents**

This geospatial analysis enhances situational awareness and supports location-specific insights.

6.5 Vehicle Type and Contributing Factor Analysis

Vehicle Type Distribution

A donut chart displays the proportion of accidents involving different vehicle types such as sedans, SUVs, motorcycles, and bicycles. This visual helps identify:

- **Vehicle categories with higher accident involvement**
- **Risk patterns associated with specific vehicle types**

Contributing Factors Analysis

A horizontal bar chart is used to display the most common contributing factors leading to accidents. These factors include:

- **Driver distraction**
- **Speeding**
- **Failure to yield**
- **Unsafe lane usage**

This analysis supports cause-based insights and highlights areas where preventive measures can be implemented.

6.6 Vulnerable Road User Analysis

Special focus is given to vulnerable road users, including pedestrians, cyclists, and bikers.

Dedicated visuals present:

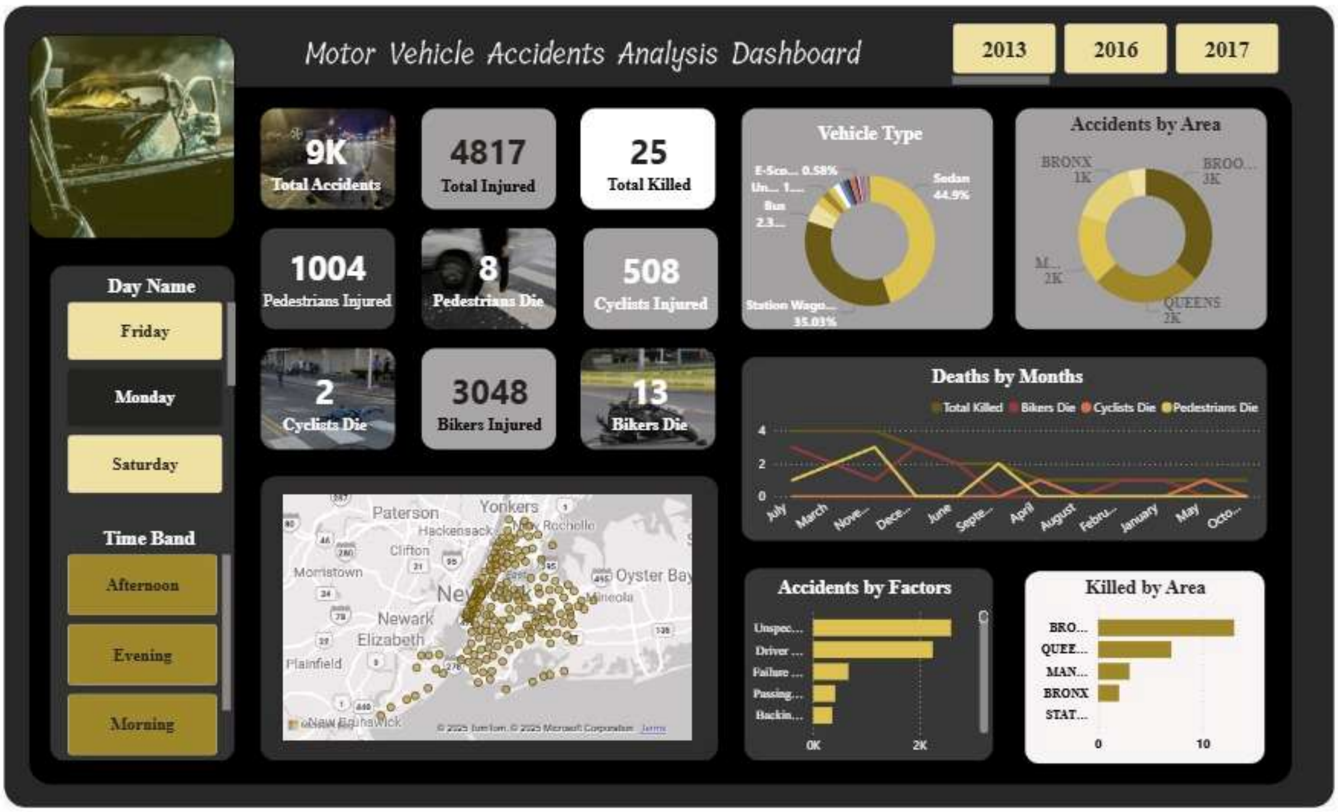
- **Injuries and fatalities for each road user category**
- **Comparative analysis of risk levels**
- **Area-wise and time-based impact on vulnerable users**

This section emphasizes road safety concerns and highlights the need for targeted interventions to protect vulnerable groups.

6.7 Interactive Features and Slicers

To enhance user interaction and analytical flexibility, the dashboard includes multiple slicers:

- **Year**
- **Day Name**
- **Time Band**



Motor Vehicle Accidents Analysis Dashboard
2013
2016
2017
9K
Total Accidents
4817
Total Injured
25
Total Killed
1004
Pedestrians Injured
8
Pedestrians Die
508
Cyclists Injured
2
Cyclists Die
3048
Bikers Injured
13
Bikers Die
Day Name
Friday
Monday
Saturday
Time Band
Afternoon
Evening
Morning
Vehicle Type
Sedan 44.9%
Station Wago... 35.03%
Accidents by Area
BRONX 1K
BROO... 3K
M... 2K
QUEENS 2K
Deaths by Months
Total Killed
Bikers Die
Cyclists Die
Pedestrians Die
Accidents by Factors
Unspec...
Driver ...
Failure ...
Passing...
Backin...
Killed by Area
BRO...
QUEE...
MAN...
BRONX
STAT...

CHAPTER 7: ANALYSIS AND INSIGHTS

This chapter presents the **key findings and analytical insights** derived from the **Motor Vehicle Accidents Analysis Dashboard**. The insights are based on interactive visualizations such as **KPI cards, line charts, bar charts, donut charts, maps, and slicers**, which collectively help in identifying accident patterns, high-risk areas, and vulnerable road user groups.

The analysis focuses on understanding accident trends over time, geographical distribution, vehicle involvement, contributing factors, and the severity of accidents in terms of injuries and fatalities.

7.1 Overall Accident Severity Analysis

The KPI indicators on the dashboard provide an immediate overview of the **scale and severity of motor vehicle accidents**. The total accident count highlights the frequency of crash events, while the injury and fatality metrics reflect the human impact associated with these incidents.

The analysis indicates that although a large number of accidents result in injuries, a smaller proportion lead to fatalities. However, even a small fatality rate represents a significant loss of life, emphasizing the importance of continuous monitoring and safety interventions.

7.2 Time-Based Trend Analysis

Year-Wise and Monthly Trends

The line charts depicting accident and fatality trends over time reveal noticeable variations across different years and months. Certain periods show an increase in accident frequency, which may be attributed to factors such as increased traffic volume, seasonal conditions, or changes in travel behavior.

Monthly trend analysis helps identify seasonal patterns, where specific months record relatively higher accident counts. This insight is valuable for planning **seasonal road safety campaigns and preventive measures**.

7.3 Time Band Analysis

The time band analysis categorizes accidents into **Night, Morning, Afternoon, and Evening** periods. The dashboard reveals that accident occurrences are generally higher during **evening and afternoon hours**, which coincide with peak traffic congestion and commuter movement.

Night-time accidents, although fewer in number, tend to have a higher severity due to reduced visibility and driver fatigue. This observation highlights the need for enhanced night-time road safety measures such as better lighting and stricter enforcement.

7.4 Area and Location-Based Analysis

Geographical analysis shows that accidents are **not uniformly distributed across all areas**. Certain boroughs consistently record higher accident counts and fatalities, indicating localized risk factors such as traffic density, road design, or urban congestion.

The accident density map further reinforces this observation by highlighting **hotspot regions** where accidents are spatially concentrated. These hotspots represent priority areas for infrastructure improvements, traffic regulation, and public awareness initiatives.

7.5 Vehicle Type Analysis

Vehicle-wise analysis reveals that **motor vehicles account for the majority of accidents**, followed by motorcycles and bicycles. This pattern reflects the higher exposure of motor vehicles on roads compared to other modes of transport.

Motorcycle and bicycle-related accidents, although fewer in number, often result in higher injury severity due to limited physical protection. This insight emphasizes the importance of safety measures such as helmet usage, dedicated lanes, and stricter traffic enforcement for two-wheelers.

7.6 Contributing Factor Analysis

The contributing factor analysis identifies **driver-related factors** as the most common causes of accidents. Factors such as distracted driving, speeding, failure to yield, and unsafe lane changes appear frequently in the dataset.

These findings suggest that **human behavior** plays a critical role in accident occurrence. Addressing these issues through awareness programs, stricter penalties, and improved driver education can significantly reduce accident frequency and severity.

7.7 Vulnerable Road User Analysis

A focused analysis of **pedestrians, cyclists, and bikers** highlights the disproportionate impact of accidents on vulnerable road users. Pedestrians and cyclists, in particular, face higher risks due to the lack of physical protection.

The dashboard shows that injuries among these groups are significantly influenced by location and time of day. High pedestrian injury counts in densely populated areas indicate the need for improved crosswalk design, traffic calming measures, and pedestrian safety infrastructure.

7.8 Key Observations

The major observations derived from the analysis include:

- Accident frequency varies significantly across different time periods
- Evening and afternoon hours record the highest number of accidents
- Certain areas consistently act as accident hotspots
- Motor vehicles dominate accident involvement, but two-wheelers face higher injury severity
- Driver-related factors are the leading causes of accidents

CHAPTER 8: CONCLUSION

This project successfully demonstrates the application of **data analytics and visualization techniques** using Microsoft Power BI to analyze motor vehicle accident data. By transforming raw accident records into a structured data model and interactive dashboard, the project provides meaningful insights into **accident trends, severity, high-risk locations, time patterns, and vulnerable road user impact**.

The implementation of a **star schema data model**, combined with effective data preprocessing and DAX-based calculations, ensured accurate aggregations and high performance. The dashboard design integrates key performance indicators, time-based analysis, geographical mapping, and cause-based visuals, enabling users to explore accident data from multiple perspectives.

The insights derived from the dashboard highlight that accidents are influenced by factors such as **time of day, traffic density, vehicle type, and driver behavior**. The focused analysis on pedestrians, cyclists, and bikers emphasizes the vulnerability of these groups and the need for targeted safety interventions.

Overall, the project meets all its defined objectives and demonstrates the practical use of Power BI as a powerful tool for **data-driven road safety analysis**. It also reflects the effective application of data analytics concepts learned during the course, including data cleaning, modeling, visualization, and storytelling with data.

CHAPTER 9: FUTURE SCOPE

Although the current project successfully analyzes historical terrorism data and presents meaningful insights Although the **Motor Vehicle Accidents Analysis Dashboard** successfully delivers meaningful insights through descriptive and exploratory analysis, there remains significant scope for further enhancement. With advancements in data analytics, machine learning, and real-time data systems, this project can be extended to provide **deeper intelligence, predictive capabilities, and greater real-world applicability**.

The following subsections outline potential future enhancements that can substantially improve the analytical depth, usability, and impact of the project.

9.1 Integration of Additional Data Sources

One of the major limitations of the current dashboard is that it relies solely on historical accident records. In future versions, the dataset can be enriched by integrating **additional contextual data sources**, such as:

- Weather conditions (rain, fog, temperature, visibility)
- Road infrastructure data (road type, lane count, speed limits)
- Traffic volume and congestion data
- Lighting conditions and road surface quality

Combining these datasets with accident data would allow a more comprehensive understanding of **environmental and infrastructural factors** contributing to accidents. This integration would enable multi-factor analysis and help identify correlations between external conditions and accident severity.

9.2 Predictive Analytics and Risk Forecasting

The current project focuses on historical analysis, which answers the question *"What has happened?"*. In the future, predictive analytics can be incorporated to answer *"What is likely to happen?"*.

Machine learning techniques such as:

- Time-series forecasting
- Classification models
- Risk scoring algorithms

can be used to predict:

- High-risk time periods
- Accident-prone locations
- Probability of severe injuries or fatalities

This enhancement would transform the dashboard from a **descriptive analytics tool** into a **decision-support system**, capable of assisting traffic authorities and policymakers in proactive safety planning.

9.3 Real-Time and Automated Data Refresh

Another important future enhancement is the integration of **real-time or near real-time accident data** using Power BI Service and cloud-based data pipelines.

By connecting the dashboard to live data sources:

- Accident data can be updated automatically
- Emergency response teams can monitor current accident trends
- Authorities can identify emerging hotspots instantly

Real-time analytics would significantly improve the practical relevance of the dashboard, especially for **traffic management and emergency response planning**.

9.4 Advanced Geospatial and Spatial Analytics

While the current dashboard uses basic map visualizations, future versions can incorporate **advanced geospatial analytics**, such as:

- Heat intensity mapping
- Spatial clustering of accidents
- Hotspot evolution over time
- Distance-based risk analysis near schools, hospitals, or highways

Advanced spatial analysis would enable deeper insights into **how accident patterns evolve geographically** and support more precise location-based interventions.

9.5 Comparative and Multi-Region Analysis

The scope of the project can be expanded to include **comparative analysis across multiple regions**, cities, or states. This would allow:

- Comparison of accident trends between urban and rural areas
- Identification of regions with better safety performance
- Cross-region benchmarking for policy evaluation

Such comparative analysis would be particularly useful for understanding **regional disparities in road safety** and identifying best practices.

9.6 Dashboard and User Experience Enhancements

Future improvements can also focus on enhancing **user experience and interactivity**, making the dashboard more intuitive and user-friendly. Potential enhancements include:

- Drill-through pages for detailed accident analysis
- Custom tooltips with contextual explanations
- Bookmark-based navigation for guided storytelling
- Role-based dashboards for different stakeholders

These features would allow users to explore the data at multiple levels of detail while maintaining simplicity and clarity.

9.7 Integration with Advanced Analytics Tools

The project can be further strengthened by integrating Power BI with external analytics tools such as **Python or R**. This integration would allow:

- Advanced statistical analysis
- Regression and correlation studies
- Machine learning model deployment
- Scenario-based simulations

Such integration would enhance the analytical depth of the project and support more complex research objectives.

9.8 Summary of Future Scope

In summary, the future enhancements proposed above have the potential to significantly elevate the **Motor Vehicle Accidents Analysis Dashboard** from a descriptive analytics project to a **comprehensive, intelligent road safety analytics platform**. By incorporating real-time data, predictive models, advanced geospatial techniques, and enriched datasets, the dashboard can play a vital role in supporting **data-driven traffic safety policies and awareness initiatives**.