CHANDIGARH UNIVERSITY
Discover. Learn. Empower.

NAAC GRADE A+ ACCREDITATED UNIVERSITY

University Institute of Computing

In partial fulfilment for the requirement of the award of the degree

Master in Computer Application

Report

on

Road Accident Dashboard

Business Analytics Report

Submitted to

Miss Amanjot Kaur.

Submitted by

Ashutosh Ranjan

23MCA20049

MCA 5B

Session: 2023-2025

ACKNOWLEDGMENT

I would like to express my sincere gratitude to all those who contributed to the successful completion of my road accident dashboard project. First and foremost, I would like to thank my academic advisor for their invaluable guidance and support throughout this project. Their expertise and encouragement have been instrumental in shaping my understanding of data visualization and road safety analysis.

I am also grateful to my peers and friends who provided insightful feedback and motivation during the development of this dashboard. Their collaborative spirit and constructive criticism were crucial in refining my work.

A special thanks to the Kaggle community for making the road accident dataset accessible. The data provided was essential for this analysis, allowing me to create a comprehensive visualization that highlights important trends in road safety.

I would also like to acknowledge the resources and tools offered by Tableau, which facilitated the creation of interactive and user-friendly visualizations. The capabilities of this software significantly enhanced the quality and effectiveness of my analysis.

Lastly, I would like to extend my heartfelt appreciation to my family for their unwavering support and encouragement. Their belief in my abilities and their constant motivation have been my driving force throughout this journey.

This project stands as a testament to the combined efforts of all those involved, and I am truly grateful for each contribution.

ABSTRACT

This report presents the development of a comprehensive road accident dashboard utilizing Tableau, aimed at enhancing the understanding of road safety trends through effective data visualization. With road accidents being a significant public safety concern, analyzing and interpreting data related to these incidents is crucial for developing preventive measures. The project leverages a dataset obtained from Kaggle, encompassing a wide range of variables such as accident type, location, time, weather conditions, and contributing factors.

The dashboard serves as an interactive platform that allows users to explore key insights and patterns in road accident occurrences. It includes various visualizations, such as bar charts, scatter plot, pie chart, and line graphs, providing a multifaceted view of the data. The dashboard enables users to filter information by different parameters, such as time periods and accident types, facilitating a deeper analysis of trends and correlations.

Key findings from the dashboard reveal significant patterns in accident frequency, peak times for incidents, and correlations between weather conditions and accident rates. These insights underscore the importance of targeted interventions and policy-making to enhance road safety. The report concludes with recommendations for further research and practical applications of the dashboard in informing stakeholders about critical safety measures.

Overall, this project demonstrates the power of data visualization in transforming complex datasets into actionable insights, highlighting the potential for improved road safety through informed decision-making.

INDEX

iii. Limitations of the Project
iv. Future Directions for Research

VII. Chapter 7 : References

i. Data Source References
ii. Literature on Road Safety and Data Visualization
iii. Tools and Technologies Used

I. Introduction

i. Background of Road Accidents

Road accidents are a leading cause of injury and death globally, posing a significant public health and safety challenge. According to the World Health Organization (WHO), approximately 1.3 million people die each year as a result of road traffic accidents, with millions more sustaining serious injuries. Factors contributing to road accidents include human behavior (such as reckless driving, speeding, and driving under the influence of alcohol), environmental conditions (such as poor weather and road infrastructure), and vehicle-related issues (like mechanical failures).

The increasing number of vehicles on the road, coupled with urbanization and population growth, has exacerbated the situation, leading to heightened traffic congestion and a greater likelihood of accidents. Understanding the patterns and trends of road accidents is crucial for developing effective strategies and interventions aimed at reducing fatalities and improving road safety.

ii. Importance of Data Visualization in Road Safety

Data visualization plays a critical role in transforming complex datasets into understandable and actionable insights. By presenting data in visual formats such as charts, graphs, and maps, stakeholders can easily identify trends, patterns, and anomalies that may not be immediately apparent in raw data.

In the context of road safety, effective data visualization allows policymakers, researchers, and the general public to grasp the scope of the problem quickly. It facilitates informed decision-making by highlighting areas of concern, such as high-accident zones, peak times for incidents, and the influence of various factors on road safety. Additionally, visualizations can be used to communicate findings to a broader audience, fostering awareness and encouraging community engagement in road safety initiatives.

iii. Objectives of the Dashboard Project

The primary objective of this dashboard project is to create an interactive and user-friendly tool that provides insights into road accident data, enabling stakeholders to analyze trends and make informed decisions regarding road safety. The specific objectives include:

1. **Data Exploration**: To explore and analyze the dataset obtained from Kaggle, identifying key features and trends related to road accidents.

2. **Visual Representation**: To design and develop visualizations that effectively communicate the data, highlighting important aspects such as accident frequency, geographical distribution, and temporal patterns.
3. **User Interaction**: To implement interactive elements within the dashboard that allow users to filter and customize their analysis based on various parameters, enhancing the overall user experience.
4. **Insights Generation**: To derive actionable insights from the visualizations that can inform policymakers, traffic authorities, and the general public about critical road safety issues and potential interventions.
5. **Awareness Promotion**: To raise awareness about road safety and the factors contributing to accidents, encouraging proactive measures to enhance public safety on the roads.

II. Data Source

i. Overview of the Kaggle Dataset

The dataset used for this project is sourced from Kaggle, a popular platform for data science competitions and sharing datasets. This particular dataset contains comprehensive records of road accidents across various regions and time frames, offering a rich resource for analyzing trends and patterns in road safety. The dataset encompasses a diverse range of information, including details about the accidents, the conditions under which they occurred, and various contributing factors.

Kaggle datasets typically undergo validation by the community, ensuring that users have access to high-quality data. This dataset is no exception, as it includes extensive documentation regarding its structure, making it suitable for in-depth analysis. The availability of such a dataset is crucial for researchers and analysts aiming to understand and mitigate road safety issues.

ii. Features Included in the Dataset

The dataset comprises several key features that are essential for analyzing road accidents. These features include:

1. **Accident ID**: A unique identifier for each accident record, facilitating easy reference and analysis.

2. **Date and Time**: Timestamp information indicating when the accident occurred, allowing for temporal analysis of accident frequency.
3. **Location**: Geographic coordinates or descriptions of the accident locations, which can be mapped to visualize spatial trends.
4. **Accident Type**: Categorizations of the type of accident (e.g., collision, rollover, pedestrian accident), providing insights into the nature of incidents.
5. **Weather Conditions**: Information on the weather at the time of the accident (e.g., clear, rainy, foggy), which is crucial for understanding environmental factors influencing road safety.
6. **Vehicle Type**: Data on the types of vehicles involved in the accidents (e.g., cars, trucks, motorcycles), allowing for analysis of vehicle-related risk factors.
7. **Casualties**: The number of injuries and fatalities resulting from the accidents, providing a direct measure of the impact of road incidents.
8. **Road Conditions**: Information regarding the state of the road at the time of the accident (e.g., dry, wet, icy), which is vital for assessing the relationship between road conditions and accident frequency.

iii. Data Collection Methods

The data collection process for the Kaggle dataset typically involves aggregating information from various sources, including governmental traffic safety agencies, insurance companies, and law enforcement reports. The dataset is curated to ensure a comprehensive representation of road accidents, encompassing different regions, times, and conditions.

Data collection methods may include:

1. **Surveys and Reports**: Gathering information from police reports, accident investigations, and victim surveys to compile accurate records of accidents.
2. **Public Databases**: Utilizing data from publicly accessible databases maintained by government agencies focused on transportation and public safety.
3. **Automated Systems**: Employing technology such as traffic cameras and sensors that record accident data in real-time, which can be integrated into the dataset.
4. **Community Contributions**: Encouraging users and researchers to contribute additional data or corrections, enhancing the dataset's accuracy and comprehensiveness.

iv. Data Preprocessing Steps

Before analysis, the dataset undergoes several preprocessing steps to ensure its quality and usability. These steps include:

1. **Data Cleaning**: Identifying and addressing missing values, duplicate records, and outliers that could skew the analysis. This process ensures that the dataset is accurate and reliable.
2. **Data Transformation**: Converting data types where necessary (e.g., converting date strings into datetime objects) and normalizing data formats to maintain consistency across records.
3. **Feature Selection**: Evaluating and selecting relevant features that contribute to the analysis while removing unnecessary or redundant variables. This step helps in simplifying the dataset for more focused analysis.
4. **Encoding Categorical Variables**: Transforming categorical data (e.g., weather conditions, accident type) into numerical formats using techniques such as one-hot encoding to facilitate quantitative analysis.
5. **Data Aggregation**: Summarizing the data to derive insights, such as calculating the total number of accidents per month or the average number of casualties per accident type, which aids in visualization and reporting.

III. Methodology

i. Overview of Tableau as a Visualization Tool

Tableau is a powerful and versatile data visualization tool widely used for transforming raw data into interactive and shareable dashboards. It enables users to connect to various data sources, including spreadsheets, databases, and cloud services, and facilitates the creation of visually appealing representations of complex data sets.

One of Tableau's key strengths is its user-friendly interface, which allows individuals with limited technical expertise to create sophisticated visualizations. The software supports a variety of visualization types, including bar charts, line graphs, pie charts, scatter plots, maps, and more, catering to diverse analytical needs.

Tableau also emphasizes interactivity, allowing users to filter data, drill down into specifics, and view data from different perspectives. This interactivity is crucial for enabling stakeholders to gain insights and make informed decisions based on the data presented. Moreover, Tableau's capability for real-time data analysis and its compatibility with web publishing make it an ideal tool for creating dynamic dashboards that can be accessed by a wider audience.

ii. Data Import and Preparation in Tableau

The first step in utilizing Tableau for the road accident dashboard project involves importing the cleaned and preprocessed dataset. This process typically includes the following steps:

1. **Connecting to the Data Source**: Tableau allows users to connect to a variety of data sources. In this project, the Kaggle dataset is imported directly into Tableau, ensuring seamless integration.
2. **Data Exploration**: Once the data is imported, Tableau provides tools to explore the dataset's structure and contents. Users can examine variable types, detect any inconsistencies, and ensure that all relevant fields are available for analysis.
3. **Data Preparation**: During this phase, any additional data preparation tasks are conducted within Tableau. This may include renaming fields for clarity, creating calculated fields for specific metrics (e.g., calculating the accident rate), and ensuring data types are correctly assigned.
4. **Data Blending**: If multiple data sources are used, Tableau allows for blending data from different sources to provide a comprehensive view. However, in this project, the focus remains on a single dataset for clarity.

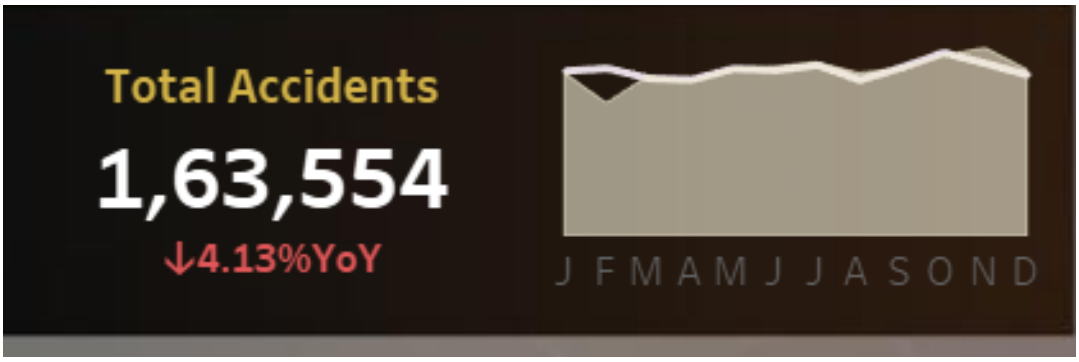
iii. Creation of Visualizations

After the data preparation, the next step is to create various visualizations that highlight key insights from the road accident data. The following types of visualizations are typically included:

1. **Bar Charts**: Used to display the frequency of accidents by type, allowing for a clear comparison of the most common accident types.

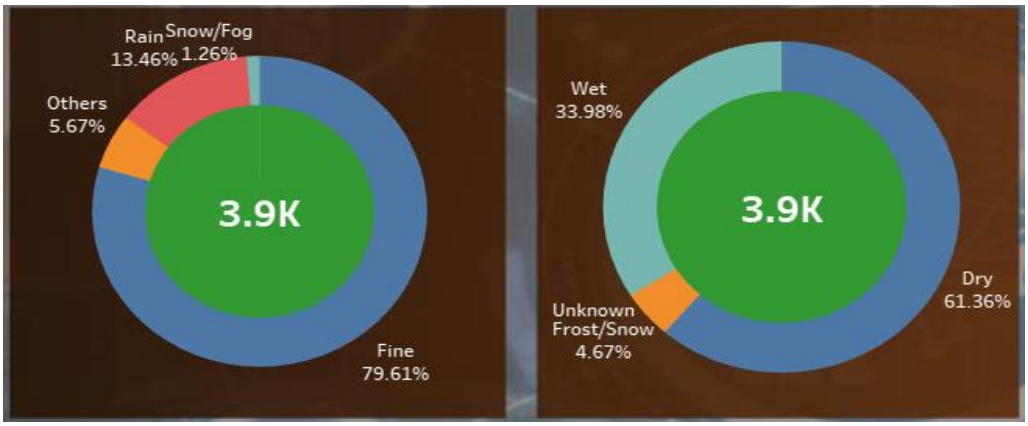
Picture showing bar graph of fatal by road surfaces

2. **Area Chart**: Employed to illustrate trends over time, such as the number of accidents per month or year, highlighting peaks and troughs in accident rates.



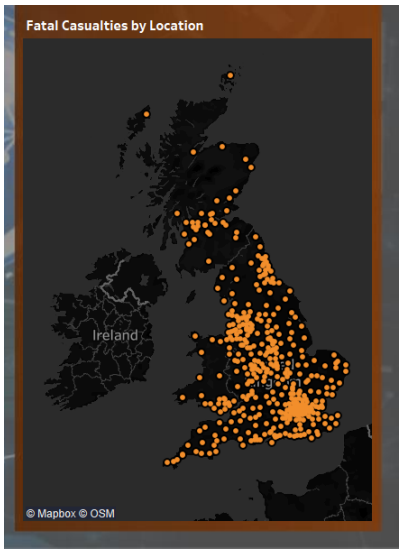
Picture of area chart showing total accident per month and showing data year on year

3. **Pie Charts**: Utilized to represent the proportion of accidents based on different categorical variables, such as vehicle types or weather conditions.



Picture of the donut chart

4. **Scatter Plots**: Analyzed to explore correlations between various factors, such as the relationship between weather conditions and accident frequency.



Scatter plot showing fatal casualties by loaction

iv. Interactive Dashboard Features

The culmination of the project is the creation of an interactive dashboard that allows users to explore the road accident data dynamically. Key interactive features include:

1. **Filter Options**: Users can apply filters to the dashboard to narrow down data based on specific criteria, such as date ranges, accident types, or locations. This interactivity enables a tailored analysis of the data.
2. **Drill-Down Functionality**: Users can click on visual elements (e.g., a bar in a bar chart) to drill down into more detailed views, accessing underlying data points for a more granular analysis.
3. **Tooltip Information**: Hovering over data points reveals additional information through tooltips, providing context and enhancing the understanding of the data without cluttering the visualizations.
4. **Dynamic Parameters**: Users can adjust parameters, such as time frames or categories, to see how the visualizations update in real-time, fostering an exploratory analysis of road safety trends.
5. **Export Options**: The dashboard includes options for users to export visualizations and data, allowing stakeholders to share insights with others or incorporate them into reports.

IV. Dashboard Design

Image showing the final dashboard

i. Overview of Dashboard Layout

The dashboard is designed to provide a user-friendly interface that facilitates the exploration of road accident data. The layout is organized logically to ensure that users can easily navigate through various visualizations and interact with the data.

1. **Grid Structure**: The dashboard is arranged in a grid format, allowing for a clear separation of different visualizations while maintaining an organized look. Key visualizations are placed in prominent sections, ensuring that critical insights are immediately visible to users.
2. **Navigation Bar**: A navigation bar at the top provides quick access to different sections of the dashboard, including links to detailed reports and additional resources related to road safety.
3. **Color Scheme**: A consistent color palette is utilized throughout the dashboard to enhance visual appeal and improve readability. Colors are selected strategically to represent different categories (e.g., accident types, weather conditions), aiding users in interpreting the data more effectively.

4. **Responsive Design**: The dashboard is optimized for various screen sizes, ensuring that users on different devices—such as desktops, tablets, and smartphones—can access the information without losing functionality or clarity.

ii. Key Visualizations

The dashboard features several key visualizations, each designed to highlight specific aspects of the road accident data:

a. Bar Charts: Accident Frequency by Type

Bar charts are employed to illustrate the frequency of accidents categorized by type (e.g., collisions, rollovers, pedestrian accidents). Each bar represents the total number of incidents for a specific type, allowing users to quickly identify the most prevalent types of accidents.

- **Design Features**: The bar chart uses distinct colors for each accident type, enhancing clarity. Data labels on each bar indicate the exact number of accidents, making it easy for users to assess the information at a glance.
- **Insight**: This visualization helps stakeholders understand which accident types are most common, guiding targeted interventions for road safety improvements.

b. Heat Maps: Geographic Distribution of Accidents

Heat maps visualize the geographic distribution of road accidents, using color intensity to represent areas with higher concentrations of incidents.

- **Design Features**: The map is interactive, allowing users to zoom in and out and hover over specific regions to reveal additional details about the number of accidents in those areas.
- **Insight**: This visualization identifies high-risk zones, enabling local authorities and safety organizations to focus resources and implement preventive measures in those regions.

c. Line Graphs: Trends Over Time

Line graphs are utilized to showcase trends in road accidents over time, highlighting changes in accident frequency on a monthly or yearly basis.

- **Design Features**: Different lines represent various accident types or contributing factors, with markers indicating significant events (e.g., policy changes, public awareness campaigns).

- **Insight**: This visualization allows stakeholders to assess the impact of interventions over time and identify seasonal trends, aiding in strategic planning and resource allocation.

iii. User Interaction Elements

User interaction is a critical aspect of the dashboard design, enhancing the overall experience and enabling tailored analysis:

a. Filters for Custom Analysis

The dashboard includes filter options that allow users to customize their analysis based on specific criteria, such as:

- **Date Range**: Users can select specific time periods to analyze trends in road accidents, allowing for seasonal comparisons or evaluations of particular initiatives.
- **Accident Type**: Users can filter the data to focus on specific types of accidents, enabling a more targeted analysis of relevant issues.
- **Location Filters**: Users can narrow down the analysis to particular geographic areas, facilitating localized assessments of road safety.

b. Tooltips and Data Labels

Tooltips and data labels enhance user interaction by providing additional context and information:

- **Tooltips**: When users hover over data points on visualizations, tooltips appear, offering more detailed insights (e.g., exact accident counts, contributing factors, or weather conditions at the time of the accident).
- **Data Labels**: Important metrics, such as total accident counts or percentages, are displayed directly on visualizations, ensuring that key information is easily accessible without needing to interact further.

V. Analysis and Insights

i. Key Findings from the Visualizations

The visualizations presented in the dashboard yield several key findings that contribute to a deeper understanding of road accidents:

1. **Prevalence of Specific Accident Types**: The bar chart indicates that collisions are the most common type of road accident, accounting for a significant proportion of incidents. This highlights the need for targeted campaigns focused on reducing collisions, especially at busy intersections.
2. **Geographic Hotspots**: The heat map reveals distinct geographic patterns, with certain areas experiencing a disproportionately high number of accidents. These hotspots are critical for policymakers and law enforcement to focus their road safety initiatives.
3. **Trends Over Time**: The line graphs show that while accident rates have fluctuated over the years, a noticeable increase in accidents is observed during specific months, suggesting possible seasonal influences. This can guide future traffic management strategies during peak months.
4. **Impact of Road Safety Initiatives**: By comparing data before and after the implementation of specific road safety initiatives, the visualizations suggest a potential decline in accident rates, indicating the effectiveness of such measures.

ii. Correlation Analysis: Weather Conditions and Accident Rates

To gain further insights into the factors influencing road accidents, a correlation analysis was conducted focusing on weather conditions and accident rates. The analysis revealed:

1. **Inclement Weather Effects**: The data showed a positive correlation between adverse weather conditions (e.g., rain, snow, fog) and increased accident rates. During rainy days, the frequency of accidents rose significantly compared to clear weather days.
2. **Specific Weather Patterns**: Analysis of the data indicated that accidents involving pedestrians and cyclists increased during periods of low visibility, such as foggy conditions. This suggests the need for enhanced visibility measures, such as improved street lighting and signage, during such weather.
3. **Recommendations for Weather-Related Awareness**: The findings advocate for the implementation of public awareness campaigns about safe driving practices in adverse weather conditions, including the importance of reducing speed and maintaining safe distances.

iii. Temporal Analysis: Peak Accident Times

The temporal analysis of accident data uncovered important trends regarding peak accident times, which can help in designing effective interventions:

1. **Time of Day**: Analysis of the data indicated that the majority of accidents occur during evening hours, particularly between 5 PM and 9 PM. This trend suggests an increased volume of traffic and potential distractions as drivers navigate home after work.
2. **Weekday vs. Weekend Patterns**: The data also revealed that weekends saw a higher incidence of accidents compared to weekdays, possibly due to increased recreational

3. travel and nightlife activities. This insight can inform targeted law enforcement efforts during weekends.
4. **Holiday Periods**: Certain holidays and events were associated with spikes in accident rates, highlighting the need for heightened road safety measures during these times, including increased police presence and public awareness campaigns.

iv. Recommendations for Road Safety Measures

Based on the analysis and insights derived from the visualizations, several recommendations can be made to improve road safety:

1. **Targeted Traffic Management**: Focus efforts on high-accident areas identified in the heat map by implementing traffic management strategies such as improved signage, traffic signals, and speed bumps.
2. **Public Awareness Campaigns**: Develop campaigns that educate drivers about the risks associated with specific accident types and promote safe driving practices, particularly during peak hours and adverse weather conditions.
3. **Enhanced Infrastructure**: Invest in infrastructure improvements, such as better street lighting in high-risk areas, pedestrian crossings, and bike lanes, to enhance safety for vulnerable road users.
4. **Monitoring and Evaluation**: Establish a continuous monitoring system for road safety initiatives, utilizing dashboard insights to assess effectiveness and make necessary adjustments over time.
5. **Collaboration with Stakeholders**: Engage with local authorities, community organizations, and the public to foster collaboration in addressing road safety issues and implementing effective solutions.

VI. Conclusion

i. Summary of Findings

This project successfully developed a comprehensive road accident dashboard utilizing data sourced from Kaggle, providing valuable insights into the factors contributing to road accidents. Key findings include:

1. **Dominance of Collisions**: The analysis revealed that collisions are the most prevalent type of road accident, necessitating focused intervention strategies.
2. **Geographic Hotspots**: The geographic distribution of accidents highlighted specific areas that require immediate attention and resource allocation for road safety improvements.

3. **Weather Impact**: A correlation analysis indicated that adverse weather conditions significantly increase accident rates, emphasizing the need for public awareness regarding safe driving practices during such conditions.
4. **Temporal Patterns**: The analysis of temporal data uncovered peak accident times, particularly during evenings and weekends, guiding targeted law enforcement and public safety campaigns

ii. Significance of the Dashboard in Promoting Road Safety

The dashboard serves as a vital tool in promoting road safety by:

1. **Data Accessibility**: By visualizing complex data in an easily interpretable format, the dashboard enhances understanding among stakeholders, including policymakers, traffic authorities, and the general public.
2. **Informed Decision-Making**: The insights derived from the dashboard empower stakeholders to make data-driven decisions regarding traffic management, resource allocation, and the implementation of road safety initiatives.
3. **Community Engagement**: The interactive nature of the dashboard fosters community engagement, allowing users to explore data relevant to their local areas and contribute to discussions about improving road safety.

iii. Limitations of the Project

Despite the comprehensive nature of the dashboard, several limitations should be acknowledged:

1. **Data Quality and Completeness**: The accuracy of the findings is dependent on the quality and completeness of the data from Kaggle. Any gaps or inconsistencies in the dataset may affect the reliability of the insights.
2. **Generalization of Findings**: While the dashboard provides valuable insights, the findings may not be generalizable to all regions or contexts. Local factors and conditions may influence accident rates differently.
3. **Static Nature of the Analysis**: The analysis conducted is based on historical data, which may not account for recent changes in traffic patterns, policies, or road conditions. Ongoing data collection and analysis are necessary to keep insights relevant.

iv. Future Directions for Research

To build upon the findings of this project and further enhance road safety initiatives, future research directions may include:

1. **Longitudinal Studies**: Conducting longitudinal studies to assess the long-term impact of implemented road safety measures and the effectiveness of public awareness campaigns.
2. **Expansion of Data Sources**: Incorporating additional datasets, such as traffic volume, demographic information, and socioeconomic factors, to gain a more comprehensive understanding of the factors contributing to road accidents.
3. **Real-Time Data Integration**: Exploring the integration of real-time data sources, such as traffic cameras and sensors, to provide up-to-date insights on road conditions and accident occurrences.
4. **Behavioral Analysis**: Investigating driver behavior and attitudes towards road safety to inform more effective interventions and educational campaigns tailored to specific demographics.

VII. References

i. Data Source References

1. Kaggle. (n.d.). Road Accidents Dataset. Retrieved https://www.kaggle.com/datasets
    - This dataset provides comprehensive information on road accidents, including various features such as accident type, location, weather conditions, and time of occurrence.
2. National Highway Traffic Safety Administration (NHTSA). (2021). Fatality Analysis Reporting System (FARS). Retrieved from https://www.nhtsa.gov/research-data/fatality-analysis-reporting-system-fars
    - This resource offers a wealth of data on fatal motor vehicle crashes in the United States, serving as a benchmark for comparative analysis.
3. World Health Organization (WHO). (2020). Global Status Report on Road Safety. Retrieved from https://www.who.int/publications/i/item/9789241565684
    - This report provides insights into road safety trends worldwide and offers valuable statistics and recommendations for improving road safety.

ii. Literature on Road Safety and Data Visualization

1. Elvik, R., & Vaa, T. (2004). The Handbook of Road Safety Measures. Elsevier.
    - This book compiles various road safety measures and evaluates their effectiveness, providing a foundation for understanding the factors influencing road accidents.

2. Kwan, M.-P., & Ding, G. (2008). The Role of Data Visualization in the Understanding of Road Safety. Transportation Research Part A: Policy and Practice, 42(10), 1396-1409.
    - This paper discusses the significance of data visualization in analyzing road safety data, highlighting its role in communicating insights to stakeholders.
3. Few, S. (2012). Show Me the Numbers: Designing Tables and Graphs to Enlighten. Analytics Press.
    - This book presents principles and techniques for effective data visualization, emphasizing clarity and understanding in presenting complex data.

iii. Tools and Technologies Used

1. Tableau Software. (n.d.). Tableau: Business Intelligence and Analytics Software. Retrieved from https://www.tableau.com/
    - Tableau is a powerful data visualization tool that allows users to create interactive and shareable dashboards. It is widely used for analyzing and visualizing complex datasets.
2. Microsoft Excel. (n.d.). Excel: Spreadsheets for Data Analysis and Visualization. Retrieved from https://www.microsoft.com/en-us/microsoft-365/excel
    - Excel is a spreadsheet application that enables data organization, analysis, and visualization, often used for preliminary data manipulation before importing into visualization tools.
3. Google Maps API. (n.d.). Google Maps Platform Documentation. Retrieved from https://developers.google.com/maps/documentation
    - The Google Maps API provides geolocation services and mapping capabilities, which can enhance the geographic visualizations within the dashboard.

Blog link - > https://road-accident-dashboard.blogspot.com/2024/10/exploring-road-safety-trends-through.html