Qlik Business Analytics Virtual Internship Program

DEFINE PROBLEM/ PROBLEM UNDERSTANDING

01. Specify The Business Problem:

   The problem of road accidents is multifaceted, involving various factors that contribute to high incidence rates and severe consequences. Human factors such as drivers behaviour----speeding, reckless driving, and distracted driving are significant contributors, as is pedistrain behavior like jaywalking. Environment factors , including poor road conditions and adverse weather, further exacerbate the risk of accidents.Vehicle-related issues such as poor maintenance and inherent design flaws also play a critical role.Inadequate infrastructure, charaterized by poorly designed roads and inefficient traffic management, adds to the problem.Legal and regularly shortcoming,including inconsistent enforcement of traffic laws and outdated legislation, hinder effective road safety measures. Socioeconomic factors like limited access to road safety education and economic constraints that prevent proper vehicle maintenance contribute to the persistence of road accidents. The consequences are severe, with high rates of fatalities and injuries leading to significant physical, emotional, and economic burdens. Addressing this issue requires comprehensive strategies, including education and awareness campaigns, infrastructure improvements, stricter vehicle safety standards, better enforcement of laws, and the adoption of advanced technological solutions. A concerted effort from government authorities, law enforcement, civil society, and individuals is essential to mitigate the problem of road accidents effectively.

02. Business Requirements:

   This analysis aims to provide key insights into user demographics, accident patterns, and problem areas. The primary focus is on creating interactive, visually compelling dashboards to support strategic planning and operational improvements. These insights will be crucial for making informed decisions, enhancing safety protocols, and ensuring regulatory compliance. By leveraging this data, organizations can better understand their users, identify trends, and address critical issues effectively. The ultimate objective is to foster a safer, more efficient environment through datadriven decision-making and proactive measures.

03. Literature Survey:

A literature survey for analyzing road safety and accident patterns involves researching and reviewing previous studies, articles, reports, and statistics on the topic. This includes examining methods and techniques used for analyzing accident data, as well as the findings and conclusions of these studies. Recommended sources include academic databases like PubMed, IEEE Xplore, Google Scholar, and institutional repositories. Additionally, government reports and publications can offer valuable insights into recent developments in the field.

04. Social Impact:

Social Impact Analysis:

- Create visualizations to display the demographic distribution of accidents across the country.

- Compare the severity of accidents in different areas of traffic control.

- Explore any correlation between speeding, weather, and total accidents.

- Identify the leading causes of accidents.

- Examine the distribution of age groups and gender of the victims.

- Investigate the contribution of diverse types of vehicles to the total number of accidents.

DATA PREPROCESSING:

Data preprocessing is a critical step in preparing raw data for analysis. Qlik employs several techniques to enhance data quality, consistency, and suitability for modeling. Here are some key aspects of Qlik's data preprocessing:

Automatic Data Preparation and Transformation:
Qlik's AutoML (Automatic Machine Learning) feature automatically preprocesses datasets before model training.
The preprocessing steps include data preparation and transformation to improve data quality and model accuracy.
Techniques such as imputing missing values, encoding categorical variables, and generating new features are applied.

Feature Classification and Handling:

Qlik classifies columns in the dataset based on their feature type (categorical, numeric, date, or free text).

It checks for sparsity, constants, and high cardinality (unique values) in columns.

Features with too many null values or low predictive value are excluded.

Feature Scaling and Standardization:

Qlik calculates summary statistics for each column to guide feature scaling. Standardization ensures that features have similar scales, preventing dominance by large values.

DATA VISUALTIZATION:

Data visualization is a powerful technique for transforming raw data into readily understandable graphical representations. This report explores the core principles and benefits of data visualization, highlighting its role in maximizing the value of information.

Data visualization offers a multitude of advantages over traditional, text-based data analysis methods. Here are some key benefits:

Enhanced Clarity:

Visual representations simplify complex data sets, allowing users to grasp trends and patterns at a glance.

Improved Communication:

Data visualizations effectively communicate insights and findings to a broader audience, fostering better understanding and collaboration.

Identification of Trends and Patterns: Graphical representations readily reveal trends and correlations that might be obscured in raw data.

Detection of Outliers:

Data visualizations can highlight outliers, data points that deviate significantly from the norm, which may hold valuable insights.

Informed Decision Making: By providing a clear understanding of the data, data visualization empowers users to make informed and data-driven decisions.

DASHBOARD:

A dashboard is a graphical user interface (GUI) that displays information and data in an organized and easy-to-read format. Dashboards are often used to provide real-time monitoring and analysis of data. They are typically designed for a specific purpose or use case. Dashboards

can be used in a variety of settings, such as business, finance, manufacturing, healthcare, and many other industries. They can be used to track key performance indicators (KPIs), monitor performance metrics, and display data in the form of charts, graphs, and tables.

STORYTELLING:

A data story is a way of presenting data and analysis in a narrative format, with the goal of making information more engaging and easier to understand. A data story typically includes a clear introduction that sets the stage and explains the context for the data, a body that presents the data and analysis in a logical and systematic way and a conclusion that summarizes the key findings and highlights their implications. Data stories can be told using a variety of media, such as reports, presentations, interactive visualizations and videos.

PERFORMANCE TESTING:

01. Applications of data filters:

Selections within data allow users to filter data based on individual fields or dimensions, enabling a focused analysis of relevant subsets. Users can choose specific values to include or exclude, set numeric and date ranges, and create complex filters using logical conditions like AND, OR, and NOT. Advanced features include dynamic, interactive filters such as dropdowns, sliders, and drill-down capabilities, as well as automated filters based on user behavior or predefined templates. These filtering capabilities are crucial for refining data analysis, improving decision-making, and enhancing operational efficiency across various applications like business intelligence, financial analysis, healthcare, and operational management.

02. Use of master items/Calculated fields:

Qlik Sense enhances data analysis by enabling the creation of reusable filter objects such as Master Items and Calculated Fields, which streamline the application of consistent filters across multiple visualizations and dashboards. Master Items allow users to define and save frequently used dimensions and measures, like "Region" or "Total Sales," ensuring uniform definitions and calculations throughout various analyses. These items can be easily applied across different charts and tables, maintaining consistency and saving time. Similarly, Calculated Fields enable users to create custom calculations, such as "Profit Margin," that automatically update across all visualizations using these fields when underlying data changes. This approach simplifies the application of complex filters, ensuring consistent criteria and reducing errors. Furthermore, shared resources like Master Items and Calculated Fields promote standardized data analysis

within teams, enhancing collaboration and ensuring uniform reporting. Overall, Qlik Sense's reusable filter objects improve efficiency, accuracy, and consistency in data analysis and reporting.

03. Number of Graphs/ Visualizations:

There are in total 15 visualizations created in this project.

01. States / NO. of accidents by flashing signals

02. State vs stop sign

03. State vs Total Accidents

04. Trafic Accidents & People injured

05. Uncontrolled Rank (Total no. of accidents vs person killed)

05. Map

06. Male and Female killed

07. Accidents due to Drunk and driving

08. State vs Phone usage

09. Accident due to weather condition

10. Pedestrian Killed: Gender wise

11. Female killed (18-25 Years)

12. Total Accidents

13. Police Controlled Accidents

14. Driving Wrong Side

15. Females Killed: Age distribution