

# **VIRTUAL INTERNSHIP PROGRAM 2024**

## **Qlik Business Analytics Program**

### **Qlik Analysis of Road Safety And Accident Patterns in India**

[01Bharti/Road-Safety-Project](https://github.com/01Bharti/Road-Safety-Project): It contains the data of accidents in different states. ([github.com](https://github.com))

### **Project Flow/Table of Content**

#### **1. Define Problem / Problem Understanding**

- Specify the business problem
- Business requirements
- Literature Survey
- Social or Business Impact

#### **2. Data Collection**

- Collect the dataset
- Connect Data with Qlik Sense

#### **3. Data Preparation**

- Prepare the Data for Visualization

#### **4. Data Visualizations**

- Number of Unique Visualizations

#### **5. Dashboard**

- Responsive and Design of Dashboard

#### **6. Storytelling**

- Story Creation

#### **7. Performance Testing**

- Amount of Data Rendered to DB
- Utilization of Data Filters
- Number of Calculation Fields/Master Items
- Number of Visualizations/Graphs

## **01. DEFINE PROBLEM/PROBLEM UNDERSTANDING**

### **1.1 SPECIFY THE BUSINESS PROBLEM:**

Technological advancements in transportation have shortened distances but increased life-threatening risks. Each year, accidents claim the lives of hundreds of thousands and result in serious injuries to millions in India. To analyze road safety and accident trends, a study will be conducted using Qlik Sense, a data analytics platform. This study will examine data on road incidents, including types of accidents, locations, causes, and other factors contributing to road safety or risks. By leveraging Qlik Sense for data-driven analysis, the study aims to generate visualizations and insights that reveal patterns and inform strategies to enhance road safety in India.

### **1.2 BUSINESS REQUIREMENTS:**

The analysis of road safety and accident trends in India, using the advanced capabilities of Qlik Sense, aims to provide comprehensive insights into user demographics, accident patterns, and critical problem areas. This data-driven approach supports strategic planning and operational improvements to enhance road safety and reduce accidents and injuries.

Technological advancements in transportation have shortened distances but increased risks, leading to hundreds of thousands of fatalities and millions of injuries annually. To address this, the study will examine data related to types of accidents, locations, causes, and other contributing factors. Qlik Sense will transform raw data into meaningful insights through interactive and visually compelling dashboards, revealing patterns and trends that inform better decision-making.

Understanding user demographics is crucial for identifying vulnerable populations and developing targeted interventions. The study will analyze age, gender, and other factors to pinpoint those most at risk. Accident patterns will also be scrutinized to distinguish between different types of collisions and identify conditions under which they occur most frequently.

Identifying high-risk locations through spatial analysis will prioritize areas needing immediate attention, whether through infrastructure improvements, increased law enforcement, or public awareness campaigns. The insights derived will be instrumental in implementing improved safety protocols and ensuring regulatory compliance.

This comprehensive analysis will support continuous monitoring and evaluation of safety measures, allowing for real-time adjustments and long-term improvements. By providing valuable insights, the study aims to reduce the number of accidents and injuries on India's roads, making travel safer for everyone.

### **1.3 LITERATURE SURVEY:**

Conducting a literature survey for the analysis of Road Safety and Accident Patterns entails researching and reviewing previous studies, articles, reports, and figures related to the topic. This comprehensive review involves exploring academic databases such as PubMed, IEEE Xplore, Google Scholar, and institutional repositories to gather a wide range of relevant information.

The literature survey encompasses an examination of various aspects, including the methods and techniques utilized for analyzing accident data, as well as the findings and conclusions of previous studies. By delving into existing research, valuable insights can be gained into the complexities of road safety and accident trends, aiding in the development of effective analytical approaches.

Government reports and publications are also valuable sources of information, providing updates on the latest developments and initiatives in the field of road safety. These documents offer insights into policy measures, regulatory frameworks, and ongoing efforts to address road accidents and improve safety standards.

Overall, a thorough literature survey forms the foundation for the Road Safety and Accident Patterns analysis, providing a wealth of knowledge and insights from previous research endeavours. By synthesizing findings from diverse sources, this approach ensures a comprehensive understanding of the subject matter and facilitates informed decision-making in the analysis process.

### **1.4 SOCIAL OR BUSINESS IMPACT:**

In conducting a Social Impact Analysis, several key objectives emerge:

- Visualize the demographic distribution of accidents nationwide.
- Assess the severity of accidents across various traffic control zones.
- Investigate potential correlations between speeding, weather conditions, and overall accident rates.
- Identify primary causes of accidents.
- Examine the age and gender distribution of accident victims.

- Analyse the contribution of different vehicle types to the overall accident count.

By pursuing these objectives, the analysis aims to provide a comprehensive understanding of the social implications of road accidents, facilitating informed decision-making and the development of targeted interventions to enhance road safety and mitigate the impact of accidents on society.

## 02. DATA COLLECTION

### 2.1 COLLECT THE DATASET:

- Identify Data Sources: Determine where your data resides. This could be in files (such as CSV, Excel), databases (like SQL Server, Oracle), cloud services (such as Google Sheets, Salesforce), or other sources.
- Access Data Load Editor: Open your Qlik Sense application and navigate to the Data Load Editor.
- Load Data from Files: If your data is in files, you can load it directly into Qlik Sense using the "Add data" button. Choose the appropriate file format (CSV, Excel, etc.), navigate to the file location, and load it.
- Dataset Link :  
<https://www.kaggle.com/datasets/aryakittukrishnasai/road-accidents-in-india>
- About the Dataset :
  - State/UT-wise Pedestrians killed according to classification of age during 2019.
  - State/UT-wise Pedestrians killed in Accidents Classified by the type of impacting vehicles during 2019 .
  - State/UT-wise Accidents Classified according to Type of Traffic Control during 2019 .
  - State/UT-wise Accidents classified according to Load Condition of Involved Vehicle during 2019 .
  - State/UT-wise Two Wheelers killed in Accidents Classified by the type of impacting vehicles during 2019 .
  - State/UT-wise Male and Female Persons Killed in Road Accidents in terms of Road User categories during 2019 .
  - State/UT-wise Accidents Victims Classified according to Non-Use of Safety Device (Non Wearing of Helmet) during 2019 etc .

### 2.2 CONNECT DATA WITH QLIK SENSE:

- Prepare Your Data: Ensure that your dataset is in a format that Qlik Sense can work with. This could be a CSV file, Excel spreadsheet, database table, or any other supported data source. Make sure the data is clean and structured properly.
- Launch Qlik Sense: Open Qlik Sense Desktop or access your Qlik Sense environment through a web browser.

- **Create a New App:** In Qlik Sense, create a new app where you'll be loading your data.
- **Load Data:** In the app, go to the "Data manager" or "Data load editor" depending on the version of Qlik Sense you're using. Then, follow these steps to load your data:
  - Click on "Add data" or "Data load editor" to start loading data.
  - Choose the appropriate data source (e.g., file, database, web service).
  - Select the file or database table containing your dataset.
  - Configure the data load options such as file format, delimiter, and field names.
  - Preview and validate the data before loading it into Qlik Sense.
- **Transform and Model Data (Optional):** If needed, you can perform data transformations and modeling during the data loading process. This may include cleaning data, creating calculated fields, joining tables, and applying data model optimizations.
- **Save and Reload Data:** Once you've configured the data load, save your changes and reload the data into the app. Qlik Sense will load the data from the specified source and create an associative data model.
- **Explore Data:** After the data is loaded, you can explore it using Qlik Sense's intuitive user interface. Create visualizations, build dashboards, and analyze the data to uncover insights.

### 03. DATA PREPARATION

Data preparation for visualization is a multi-step process crucial for making data comprehensible and suitable for creating insightful visualizations. The initial step involves cleaning the data, which includes removing irrelevant information and addressing missing values to ensure the dataset's integrity. Once cleaned, the data is transformed into a format conducive to visualization. This transformation might involve restructuring the data, aggregating information, or converting data types to align with the visualization tools being used.

Exploring the data is the next step, where patterns, trends, and anomalies are identified. This exploratory analysis is vital as it guides the direction of the visualization process, highlighting key areas of interest and potential insights. Following this, the data is filtered to focus on specific subsets that are most relevant to the analysis. This step ensures that the visualizations are targeted and meaningful, avoiding the clutter and confusion of extraneous information. Preparing the data for visualization software involves ensuring compatibility with the chosen tools. This preparation might include formatting the data into specific file types, arranging data columns appropriately, and verifying that all necessary variables are included. Accuracy and completeness of the data are paramount throughout this process. Inaccurate or incomplete data can lead to misleading visualizations and erroneous conclusions.



## 04. DATA VISUALIZATION

Preparing the data for visualization involves cleaning the data to remove irrelevant or missing data, transforming the data into a format that can be easily visualized, exploring the data to identify patterns and trends, filtering the data to focus on specific subsets of data, preparing the data for visualization software, and ensuring that the data is accurate and complete. This process helps to make data easily understandable and ready for creating visualizations to gain insights.

### NUMBER OF UNIQUE VISUALIZATIONS:

The number of unique visualizations that can be created with a given dataset. Some common types of visualizations that can be used to analyse include bar charts, line charts, heat maps, scatter plots, pie charts, maps etc. These visualizations can be used to compare, track changes over time, show distribution, relationships between variables, breakdown of one category and much more.

**Map Chart:** A map chart, also known as a geographic or choropleth map, is a type of data visualization tool used to represent data points on a geographical map. It typically displays statistical data related to specific geographic areas such as countries, states, provinces, or regions. The data is often represented visually using colors or shading to indicate different values or categories.

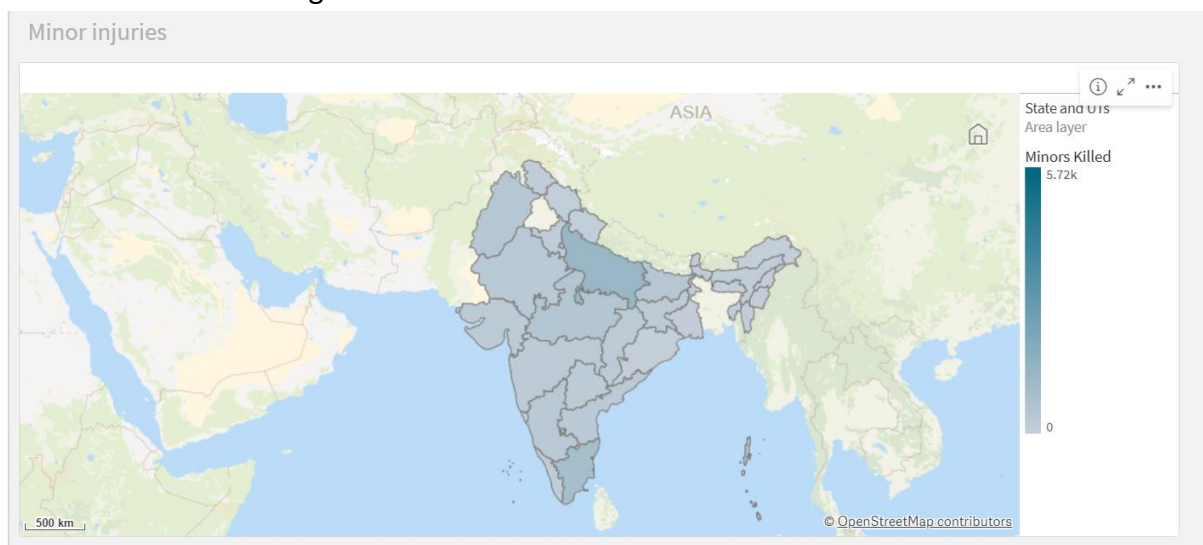


Figure 1 Minors Killed



**Bar Chart:** A horizontal bar chart is a graphical representation of data where individual bars represent different categories. Unlike vertical bar charts, where bars extend vertically from the x-axis, horizontal bar charts have bars that extend horizontally from the y-axis. This type of chart is useful for comparing values across different categories or for visualizing data where category names are too long to fit comfortably below vertical bars.

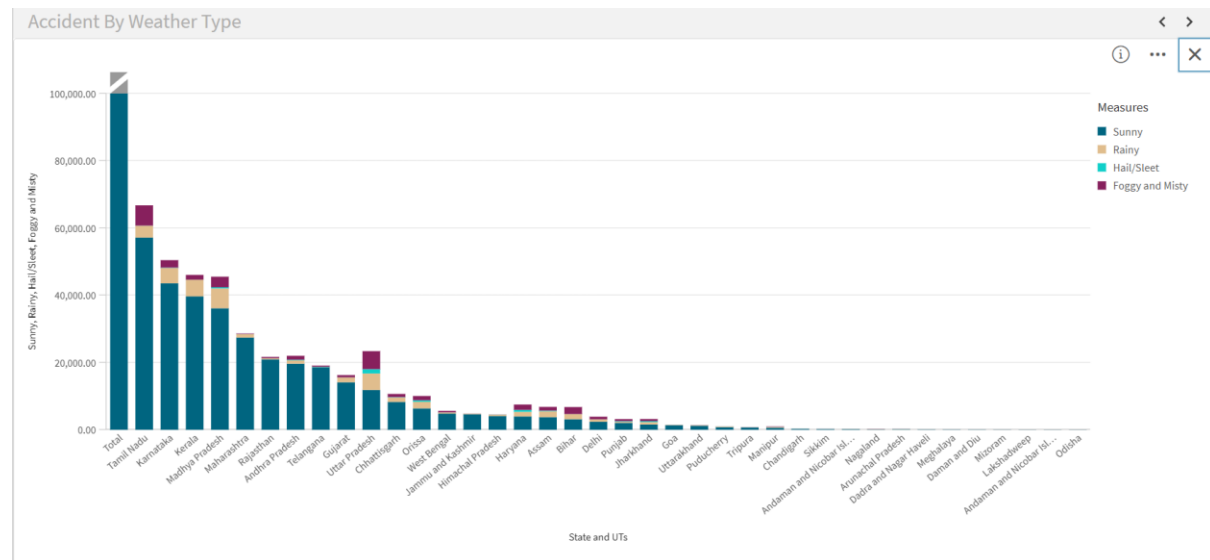


Figure 2 Accident By Weather Type

**Line Chart:** A line chart is a type of graph used to display data points that are connected by straight lines. It is commonly used to show trends over time, making it a popular choice for visualizing changes and patterns in data.

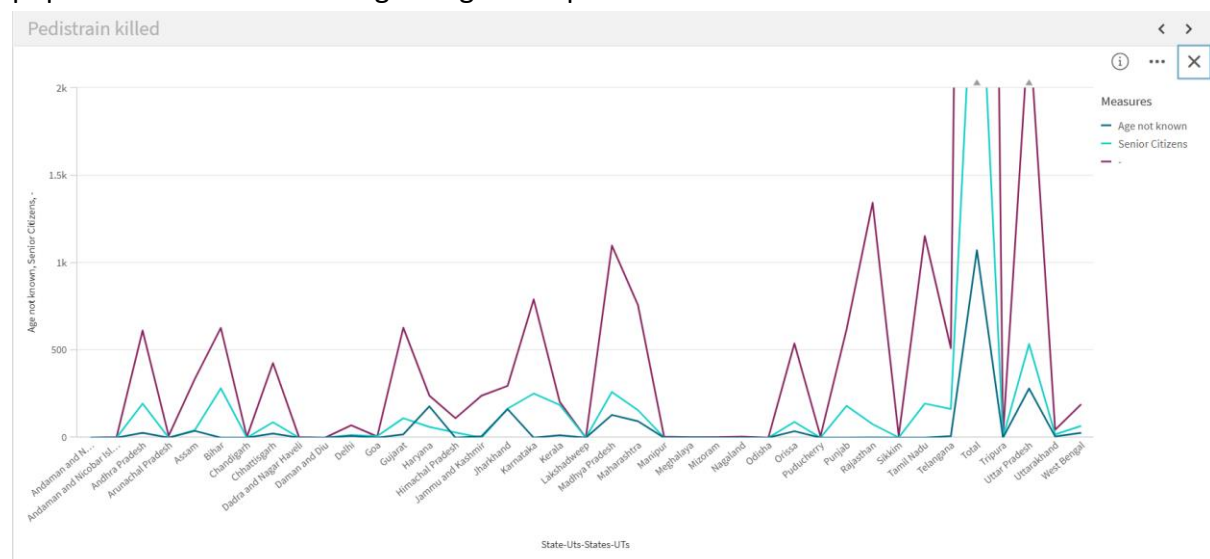


Figure 3 Pedestrian Killed

**Pie chart:** A pie chart is a circular statistical graphic divided into slices to illustrate numerical proportions. Each slice of the pie represents a category's

contribution to the whole, with the size of each slice proportional to the quantity it represents. Pie charts are useful for showing the relative sizes of parts to a whole in a simple and easy-to-understand format.

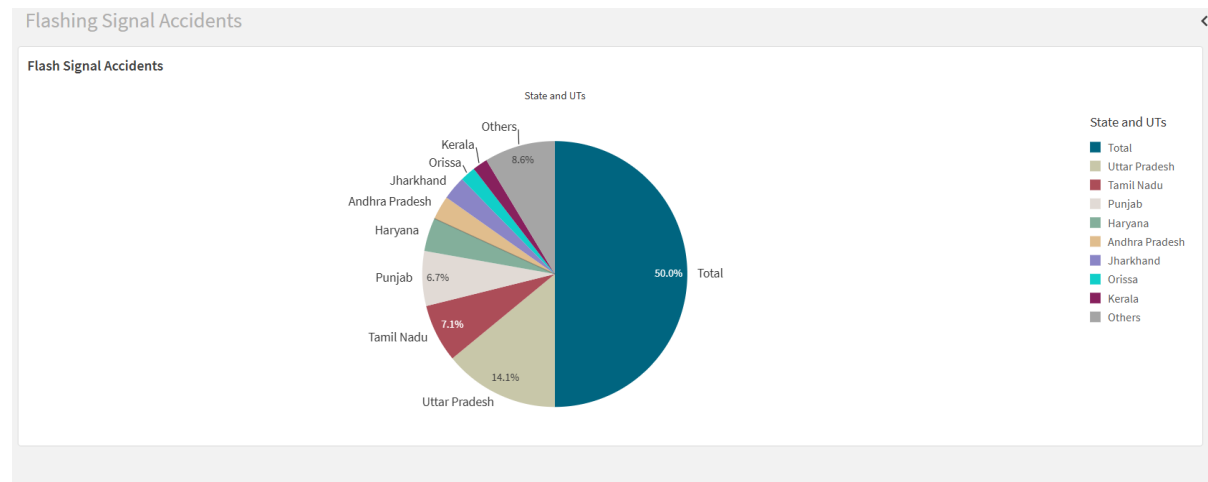


Figure 4 Flash Signal Accidents

**Stacked Horizontal Bar chart:** A stacked horizontal bar chart is a type of bar chart where bars are segmented into sub-bars that represent different categories or components. Each bar extends horizontally and is divided into segments that stack on top of each other, making it easy to see the contribution of each component to the total for that category.

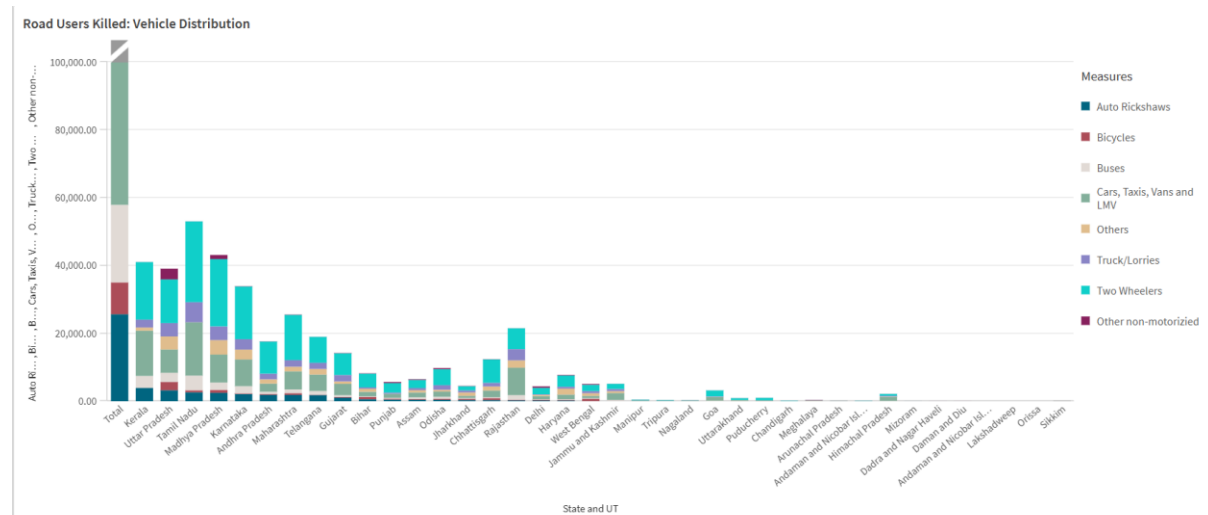


Figure 5 Road Users Killed : Vehicle Distribution

**Scatter Plot:** A scatter plot (also known as a scatter chart or scatter graph) is a type of data visualization that displays values for typically two variables for a set of data. The data is displayed as a collection of points, each representing an observation, where the position of each point on the horizontal and vertical axis indicates values for an individual data point.

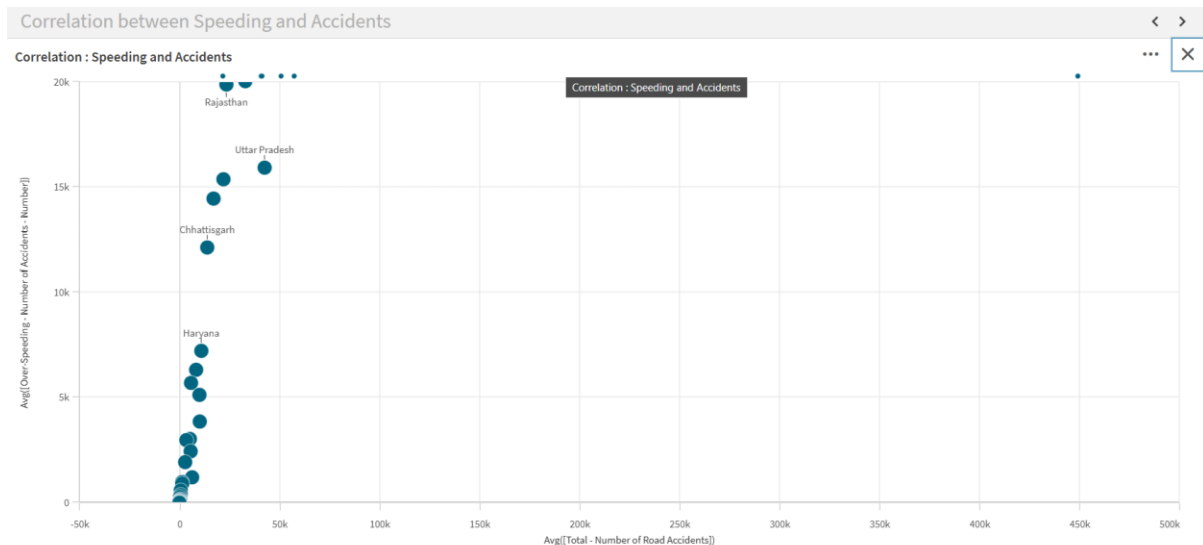


Figure 6 Correlation between Speeding an Accidents

**Multi Line Chart:** A line chart using multiple measures, also known as a multi line chart, is a type of line chart that displays multiple data sets on the same graph. Each data set is represented by its own line, allowing for comparison between different variables or measures over the same period or categories.

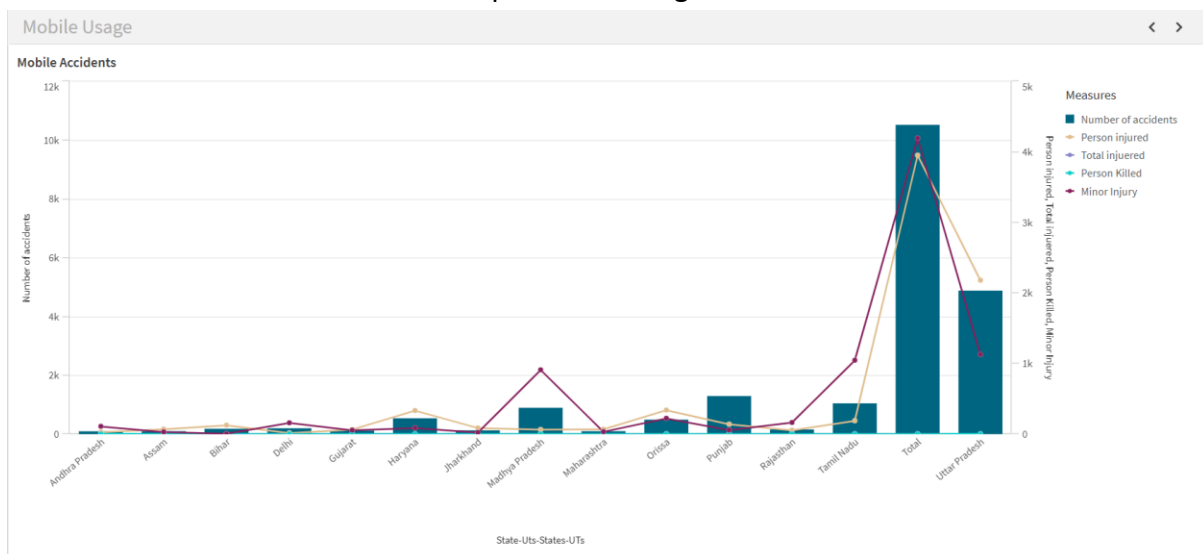


Figure 7 Mobile Usage

**Stacked Horizontal Bar chart:** A stacked horizontal bar chart is a type of bar chart where bars are segmented into sub-bars that represent different categories or components. Each bar extends horizontally and is divided into segments that stack on top of each other, making it easy to see the contribution of each component to the total for that category.

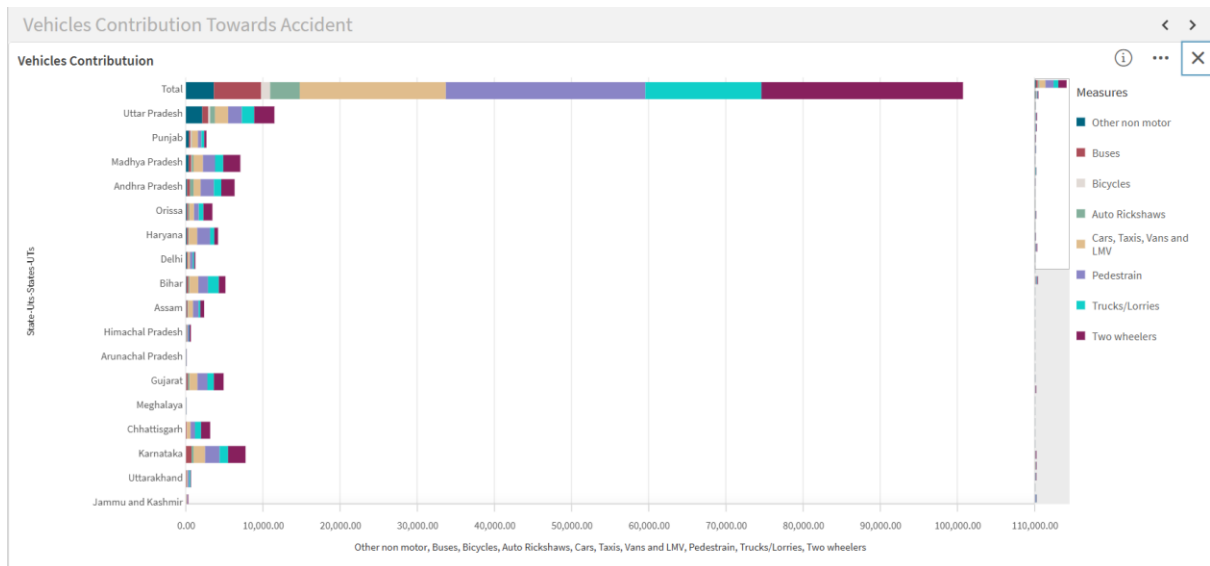
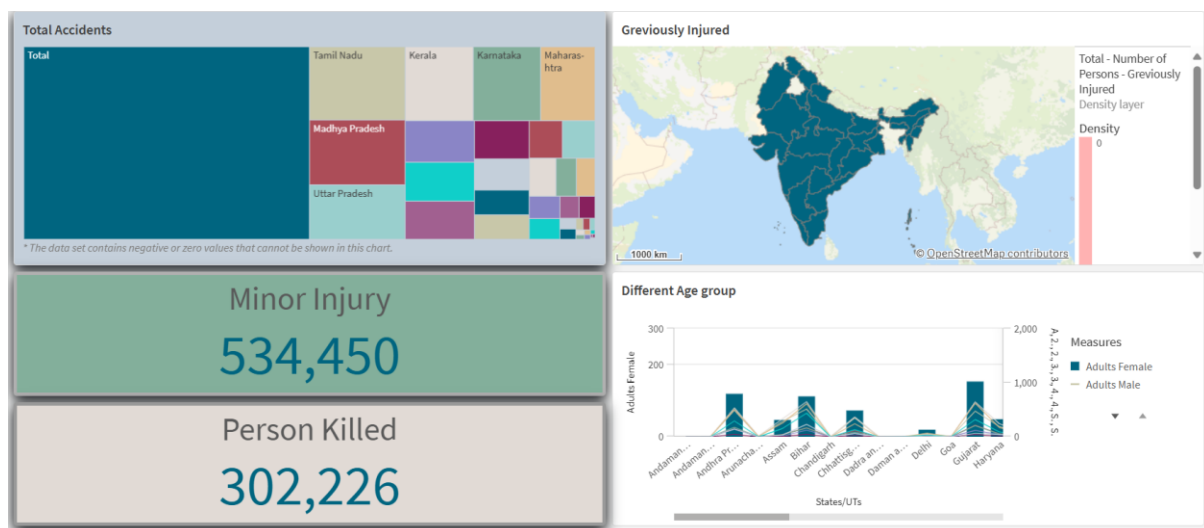


Figure 8 Vehicle Contribution Towards Accidents

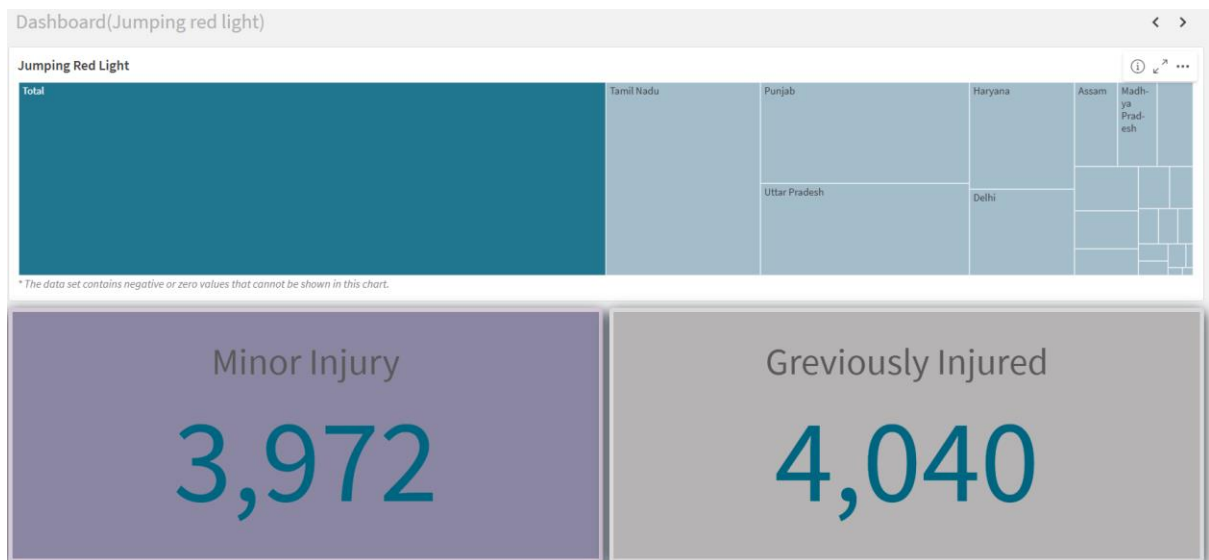
## 05. DASHBORAD

A dashboard is a graphical user interface (GUI) designed to display information and data in an organized, easy-to-read format. These tools are commonly used for real-time monitoring and data analysis, tailored to specific purposes or use cases. Dashboards are utilized across various industries, including business, finance, manufacturing, and healthcare. They help track key performance indicators (KPIs), monitor performance metrics, and present data through charts, graphs, and tables.

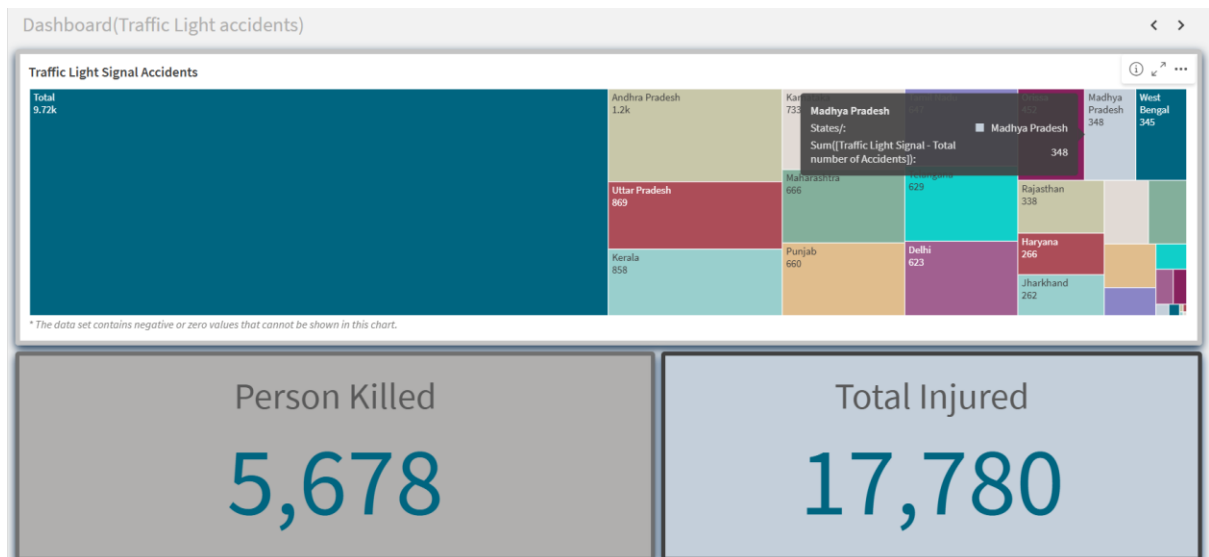
In Qlik, a dashboard refers to a collection of visualizations, data, and analytics that provides a comprehensive view of key metrics and insights. Qlik, a leading platform in data analytics and business intelligence, enables users to create interactive dashboards for real-time data exploration and informed decision-making.



Dashboard 1. Persons Killed



Dashboard 2. Jumping Red Light



Dashboard 3. Traffic Light Accidents

## 06. STORYTELLING

Storytelling in a Qlik project involves crafting a narrative around the insights and findings derived from data analysis using Qlik Sense. This approach focuses on presenting data compellingly and informatively, engaging stakeholders and effectively communicating key insights.

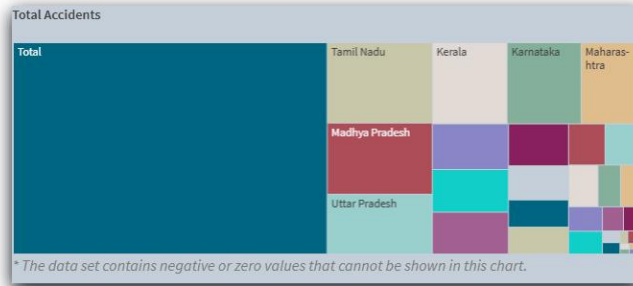
A data story is a method of presenting data and analysis in a narrative format, aiming to make information more engaging and easier to understand. Typically, a data story includes a clear introduction that sets the stage and provides context for the data. The body of the story presents the data and analysis in a logical and systematic manner. Finally, the conclusion summarizes the key findings and highlights their implications. Data stories can be conveyed through various media, such as reports, presentations, interactive visualizations, and videos.

In essence, storytelling in a Qlik project transforms raw data into a narrative that resonates with its audience. By organizing data into a structured narrative, it becomes easier for stakeholders to grasp complex insights and see the practical applications of the data. This method not only enhances understanding but also fosters a deeper connection to the data, prompting more informed decision-making.

Qlik's tools support this storytelling approach by allowing users to create interactive and dynamic visualizations. These visualizations can be seamlessly integrated into data stories, enhancing their impact and accessibility. By leveraging the capabilities of Qlik Sense, users can build narratives that not only inform but also inspire action, making data-driven insights an integral part of strategic discussions and decisions.

## Total Number of Accidents Occur:

Highest In Tamil Nadu



Minor Injury

Minor Injury  
534,450

Previously Injured

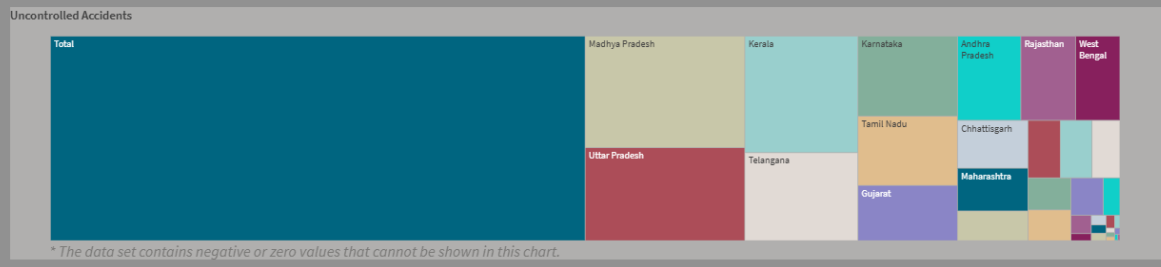
Total Person Killed

Person Killed  
302,226



Total Number of Accidents

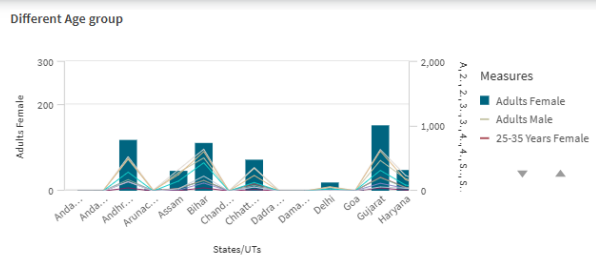
## Total Uncontrolled Accidents



Person Killed

Person killed  
662

Age  
Distribution



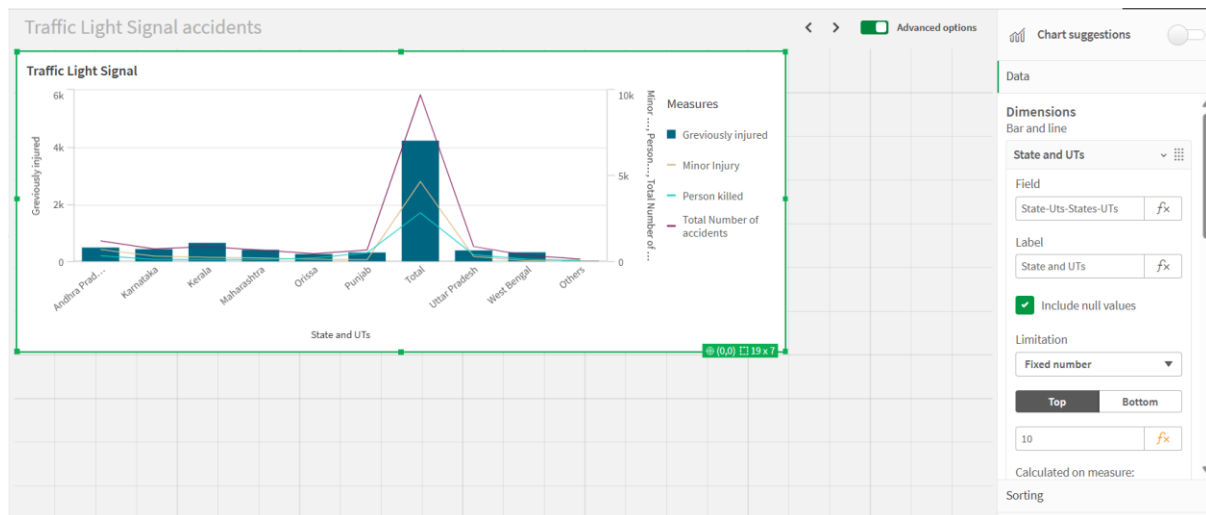
Uncontrolled Accidents



## 07. PERFORMANCE TESTING

### 7.1 APPLICATIONS OF DATA FILTERS:













Selections within the data enable users to filter based on individual fields or dimensions, allowing them to include or exclude specific values from the analysis. This functionality provides a tailored view of the data, facilitating more precise insights. Users can also create complex filters using predefined conditions and logic to meet specific analytical needs. These capabilities enhance the flexibility and depth of data exploration, allowing for a more detailed and customized analysis. By leveraging these selection tools, users can efficiently navigate through large datasets and focus on the most relevant information for their decision-making process.



Limitation Filter

### 7.2 USE OF MASTER FIELD:

Master Items are reusable objects that enable users to centralize and manage dimensions, measures, and visualizations across multiple sheets and apps. This feature ensures consistency and standardization in data representation, streamlining the development and maintenance of dashboards. By using Master Items, users can efficiently apply the same metrics and visual elements throughout their projects, enhancing coherence and reducing the time required for updates. This centralized approach simplifies the management of data components, contributing to a more efficient and reliable dashboard creation process.

	 Search
Fields	Dimensions
	Measures
Master items	<div>Create new</div>
	18 to 25 persons 
Charts	25 to 35 persons 
	35 to 45 Persons 
Custom objects	45 to 60 Persons 
	Age Not Known Tot... 
	Minor Persons 
	Senior Citizen 60 & ... 
	Visualizations

## **08. NUMBER OF UNIQUE VISUALIZATIONS**

There are in total 15 visualizations created in this project:

01. Accidents due to weather conditions
02. Mobile Injuries
03. Accidents By Wrong Side Driving
04. Minors Killed
05. Correlation between Speeding and Accidents
06. Road Users Killed: Vehicles Distribution
07. Pedestrian Killed: Gender
08. Vehicle Contribution Towards Accident
09. Pedestrian Killed
10. Flash Signal Accidents
11. Mobile Usage
12. Traffic Light Signal Accidents
13. Police Controlled Accidents
14. Stop Signed Killed