Qlik Analysis Of Road Safety And Accident Patterns In India

Define Problem / Problem Understanding :

Specify the business problem:

Technological advancement in transportation has minimised the distances but has increased the risk to life. Every year, accidents result in loss of lakhs of lives and serious injuries to crores of people. A study to analyse road safety and accident trends in India is to be conducted using Qlik Sense, a data analytics platform. This study involves examining data related to road incidents, such as types of accidents, locations, causes, and potentially factors contributing to road safety or risks. The use of Qlik Sense is a data-driven approach, utilizing visualizations and insights generated from the analysis to understand patterns and potentially inform strategies for improving road safety in India.

Business requirements:

The analysis aims to provide valuable insights into user demographics, accident patterns, and problem areas. The primary focus is on creating interactive and visually compelling dashboards to support strategic planning and operational improvements. The insights derived from this analysis will be instrumental in making informed decisions, implementing better safety protocols, and ensuring compliance with regulations.

Literature Survey:

For my literature survey on Road Safety and Accident Patterns analysis, I

researched and reviewed numerous previous studies, articles, reports, and figures on the topic. My exploration included delving into the various methods and techniques used for analyzing accident data, as well as examining the results and conclusions of these studies. I utilized several academic databases, including PubMed, IEEE Xplore, Google Scholar, and various institutional repositories, to gather comprehensive and relevant information. Additionally, I examined government reports and publications to gain insights into the latest developments in road safety and accident analysis. This thorough approach allowed me to compile a robust and detailed overview of the current state of research in this field

Social Impact:

As part of my analysis on road safety and accident patterns, I carried out several tasks to provide a comprehensive understanding of the data:

1. Created visualizations to display the demographic distribution of accidents across the country: I developed various charts and maps to illustrate how accidents are distributed among different demographic groups across various regions. These visualizations included heat maps and bar graphs to highlight areas with higher accident rates and the demographics most affected.

2. Compared the severity of accidents in different areas of traffic control: I analyzed accident data to compare how the severity of accidents varies in areas with different traffic control measures, such as intersections with traffic lights versus stop signs. I used statistical methods to quantify these differences and visualized them using comparative bar charts and scatter plots.

3. Explored any correlation between speeding, weather, and total accidents: I conducted a correlation analysis to determine how speeding and weather conditions influence the total number of accidents. This involved collecting data on speed limits, recorded speeds, weather conditions at the time of accidents, and total accident counts. I then used scatter plots and correlation matrices to present the relationships.

4. Identified the leading causes of accidents: I performed a thorough analysis of the accident data to identify the most common causes of accidents. This involved categorizing accidents by cause and using pie charts and bar graphs to represent the frequency and percentage of each cause.
5. Examined the distribution of age groups and gender of the victims: I analyzed the demographic information of accident victims to understand how different age groups and genders are affected. I created age distribution histograms and gender ratio charts to visualize this data.
6. Investigated the contribution of diverse types of vehicles to the total number of accidents: I looked into how various types of vehicles contribute to the overall number of accidents. This analysis included categorizing accidents by vehicle type (e.g., cars, trucks, motorcycles) and using stacked bar charts and pie charts to illustrate the proportions and frequencies of each vehicle type involved in accidents. These steps provided a detailed and multifaceted view of road safety and accident patterns, allowing for more targeted and effective recommendations for improving road safety.

Data Collection & Extraction From Database:

Data collection is the process of gathering and measuring information on variables of interest, in an established systematic fashion that enables one to answer stated research questions, test hypotheses, evaluate outcomes and generate insights from the data.

Data Preparation:

To prepare the data for visualization in my analysis of road safety and accident patterns, I followed a meticulous process to ensure the data was clean, well-organized, and ready for creating insightful visualizations. Here are the steps I undertook:

1. Data Cleaning:

- Removed Irrelevant Data: I filtered out any data points that were not pertinent to the analysis, such as unrelated fields or records outside the scope of the study.
- Handled Missing Data: I addressed missing values by either imputing them with appropriate estimates or removing incomplete records if they were not critical.

2. Data Transformation:

- Formatted Data: I transformed the data into a consistent format, ensuring that all variables were standardized, such as converting dates to a common format and normalizing numerical values.
- Categorized Data: I grouped the data into meaningful categories, such as accident severity levels, vehicle types, age groups, and weather conditions, to facilitate easier analysis.

3. Data Exploration:

- Identified Patterns and Trends: I conducted exploratory data analysis (EDA) to uncover initial patterns, trends, and anomalies within the dataset. This involved generating summary statistics and preliminary visualizations like histograms and scatter plots.
- Explored Relationships: I examined relationships between different variables to identify potential correlations and causations, which would be critical for my visualizations.

4. Data Filtering:

- Focused on Specific Subsets: I filtered the data to concentrate on specific subsets that were most relevant to the research questions. For instance, I isolated data related to high-risk demographics or particular weather conditions.
- Refined Data Scope: I ensured that the data subsets chosen for visualization were comprehensive enough to provide meaningful insights

while avoiding information overload.

6. Ensured Data Accuracy and Completeness:

- Cross-Verification: I cross-verified the data against original sources to confirm its accuracy.
- Completeness Check: I ensured that all necessary data points were present and adequately represented in the dataset.
By carefully preparing the data through these steps, I was able to create clear, accurate, and insightful visualizations that effectively conveyed the key findings and trends in road safety and accident patterns. This preparation made the data easily understandable and ready for in-depth analysis and decision-making.

Data Preprocessing - Qlik Sense Script:

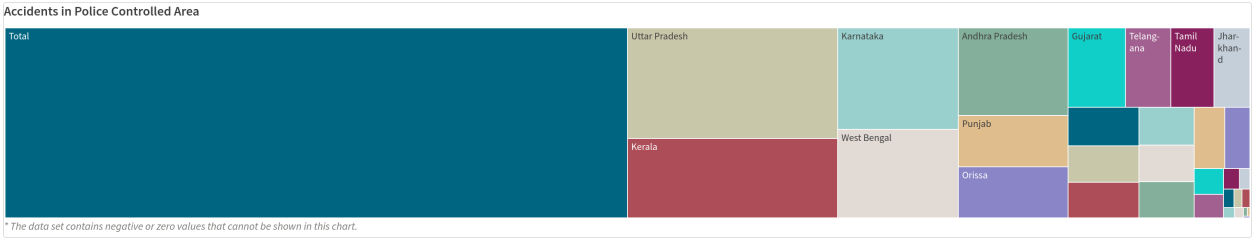
```
SET ThousandSep=',';
SET DecimalSep='.';
SET MoneyThousandSep=',';
SET MoneyDecimalSep='.';
SET MoneyFormat='$ ###0.00;-$ ###0.00';
SET TimeFormat='h:mm:ss TT';
SET DateFormat='M/D/YYYY';
SET TimestampFormat='M/D/YYYY h:mm:ss[.fff] TT';
SET FirstWeekDay=6;
SET BrokenWeeks=1;
SET ReferenceDay=0;
SET FirstMonthOfYear=1;
SET CollationLocale='en-US';
SET CreateSearchIndexOnReload=1;
SET MonthNames='Jan;Feb;Mar;Apr;May;Jun;Jul;Aug;Sep;Oct;Nov;Dec';
SET LongMonthNames='January;February;March;April;May;June;July;August;September;October;November;December';
SET DayNames='Mon;Tue;Wed;Thu;Fri;Sat;Sun';
SET LongDayNames='Monday;Tuesday;Wednesday;Thursday;Friday;Saturday;Sunday';
SET NumericalAbbreviation='3:k;6:M;9:G;12:T;15:P;18:E;21:Z;24:Y;-3:m;-6:µ;-9:n;-12:p;-15:f;-18:a;-21:z;-24:y';

```

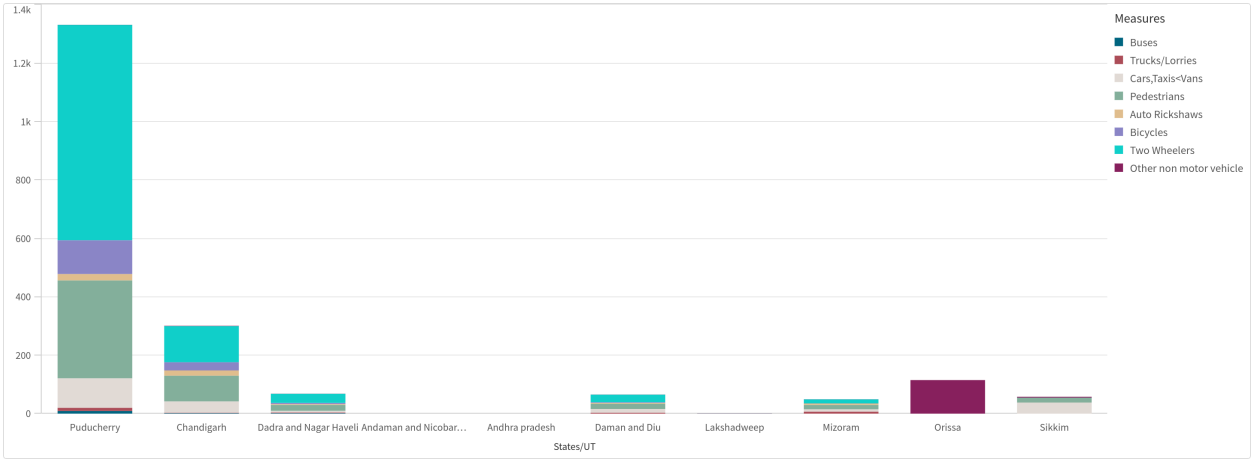
Data Visualisation:

Total Number of Road Accidents -

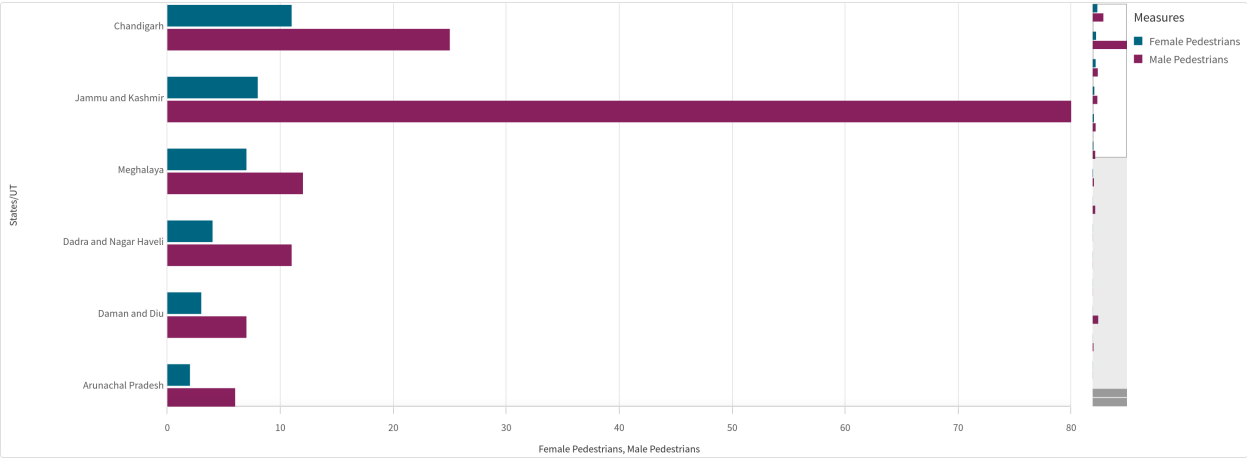
Accidents in Police Controlled Area



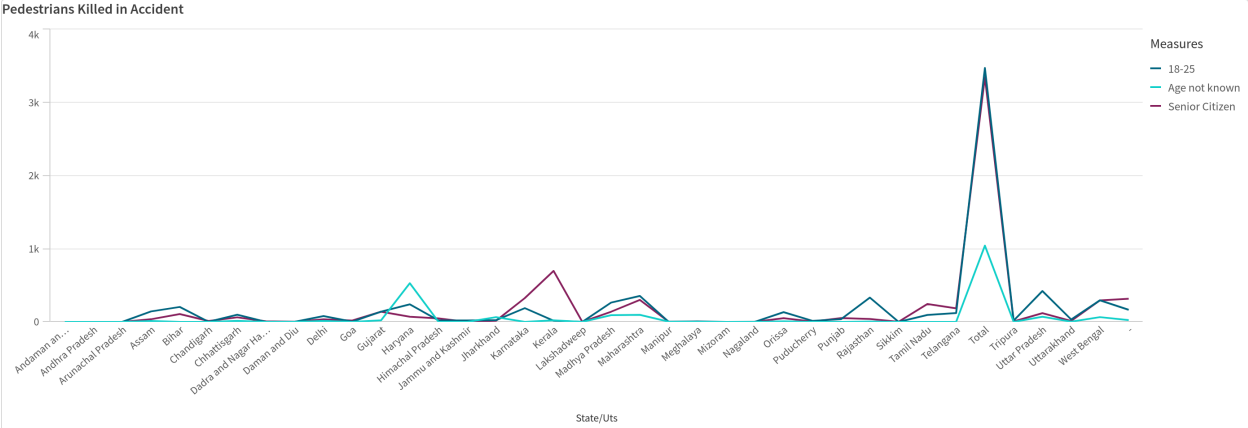
Road users killed Vehicle Distribution



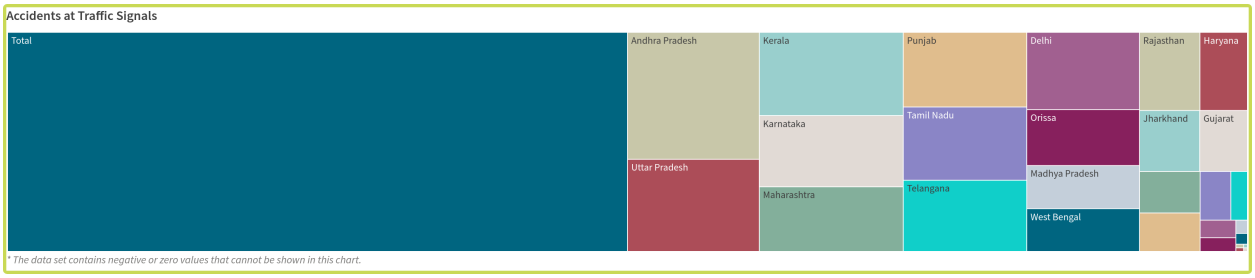
Pedestrians killed



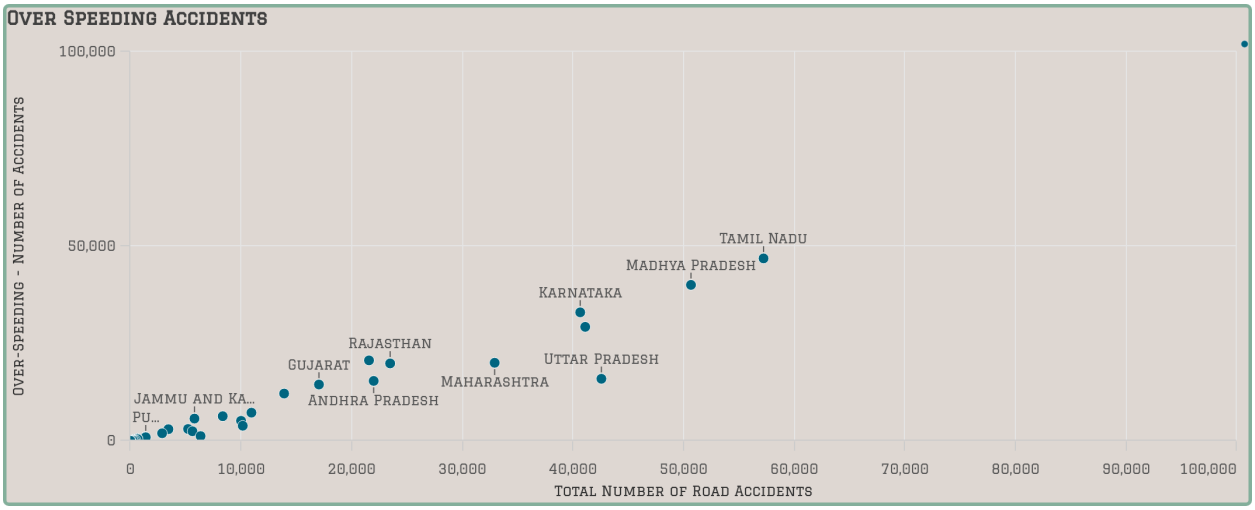
Pedestrians killed in Accident



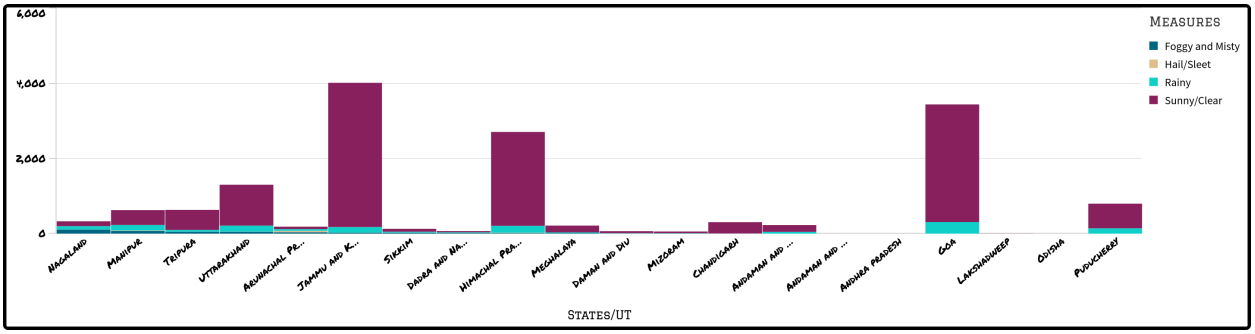
Accidents at traffic signals



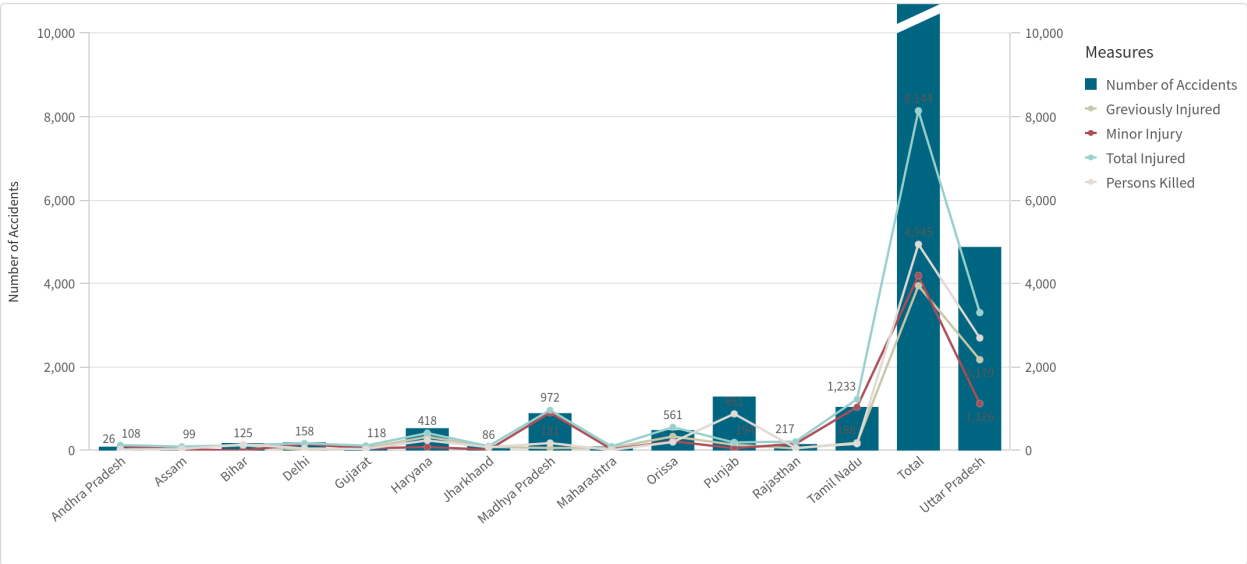
Over Speeding Accidents



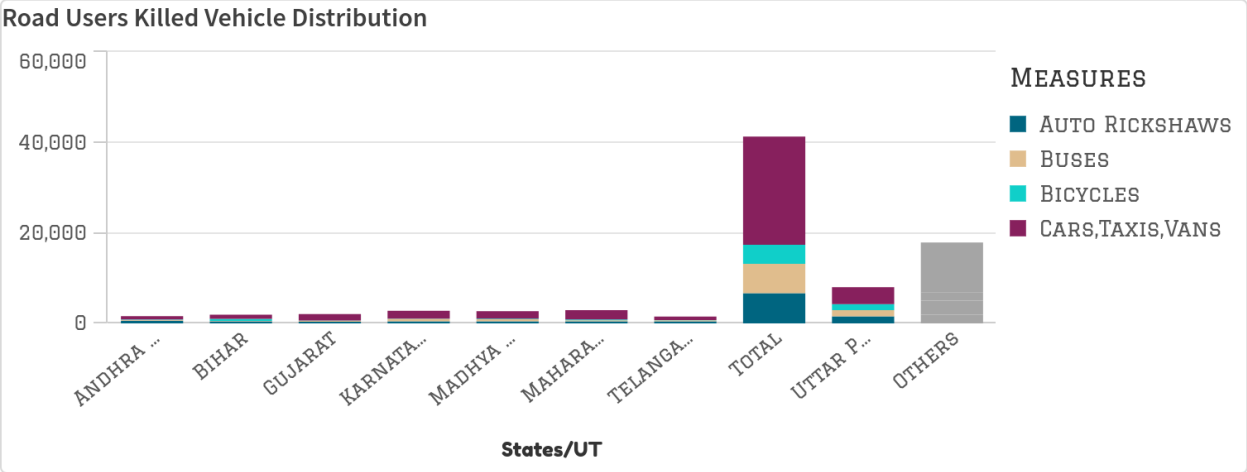
Road users killed



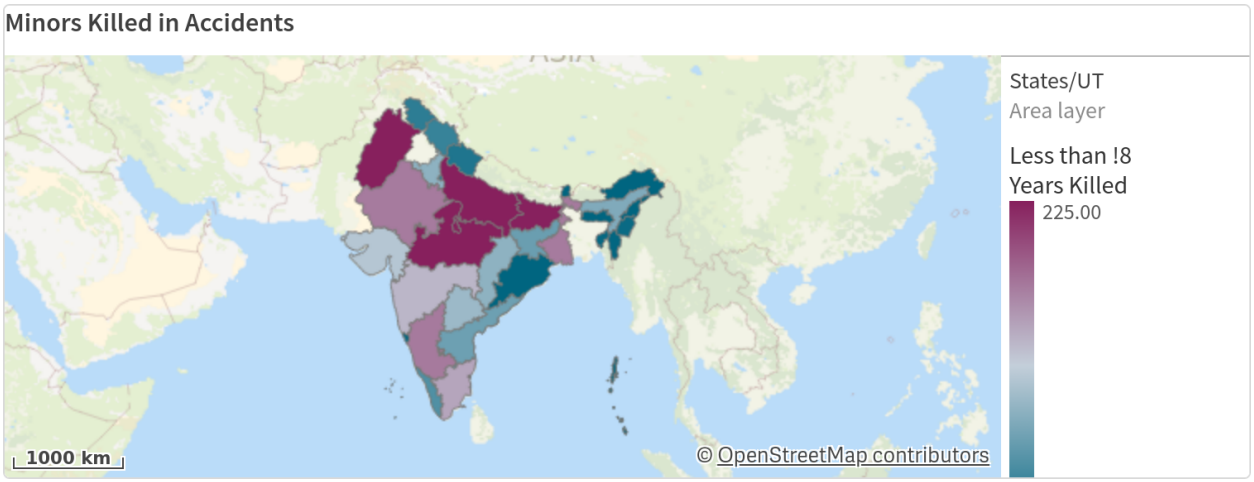
Mobile Usage



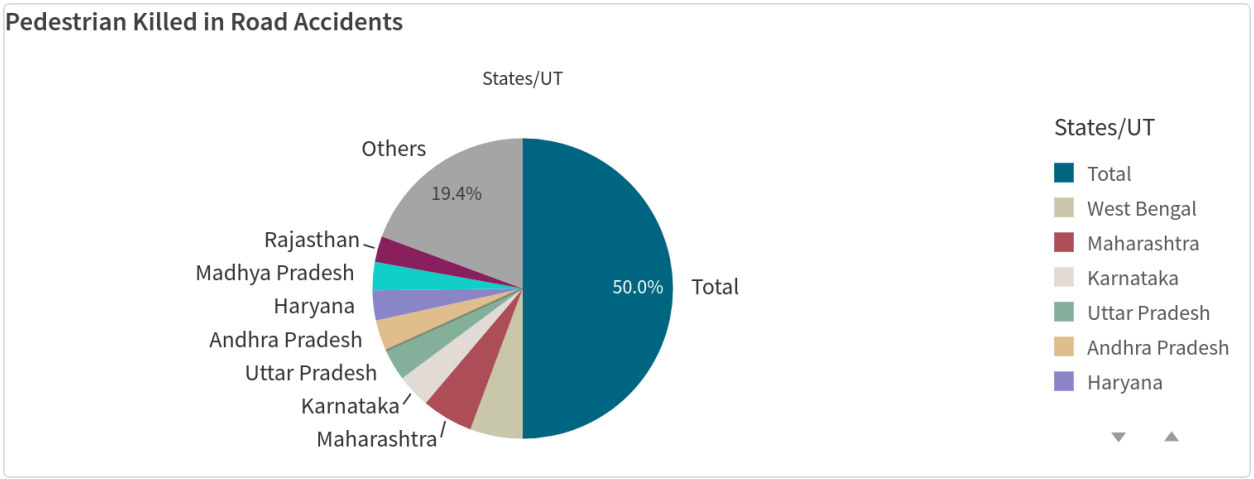
Road users killed vehicle distribution



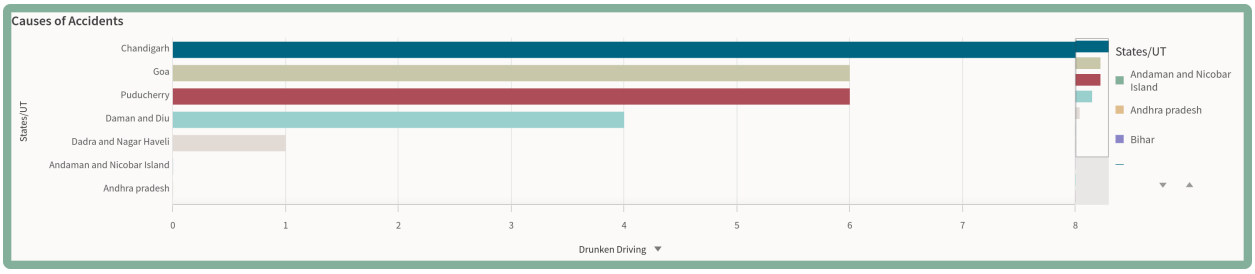
Minors Killed in accidents



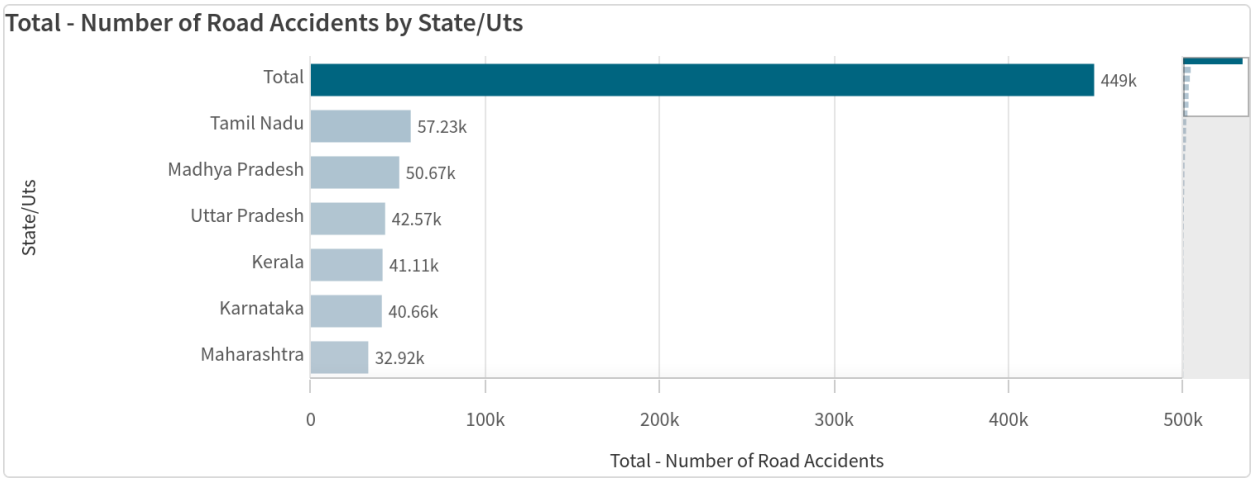
Pedestrians killed in road accidents



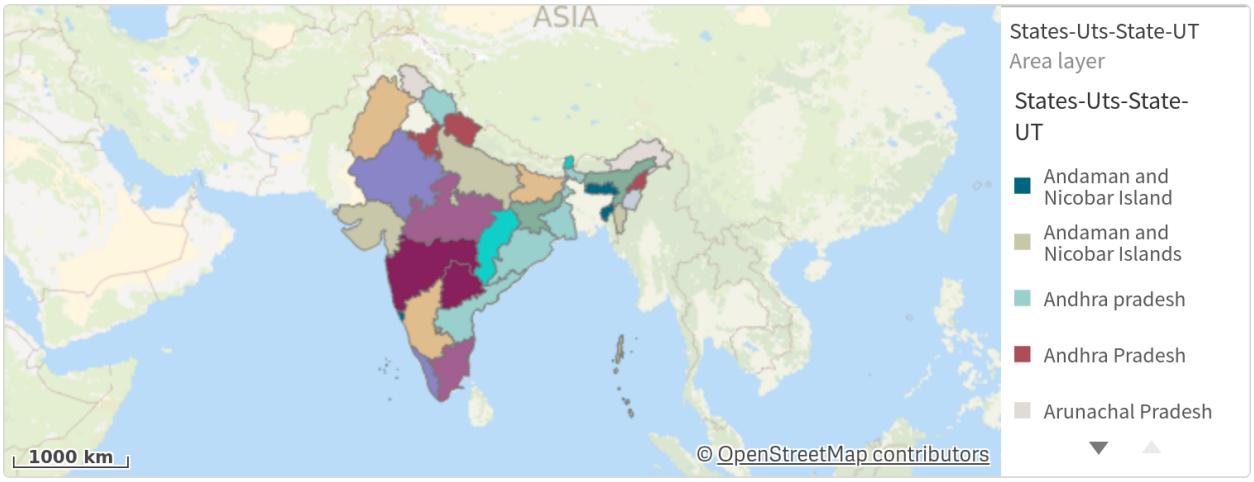
Causes of Accidents



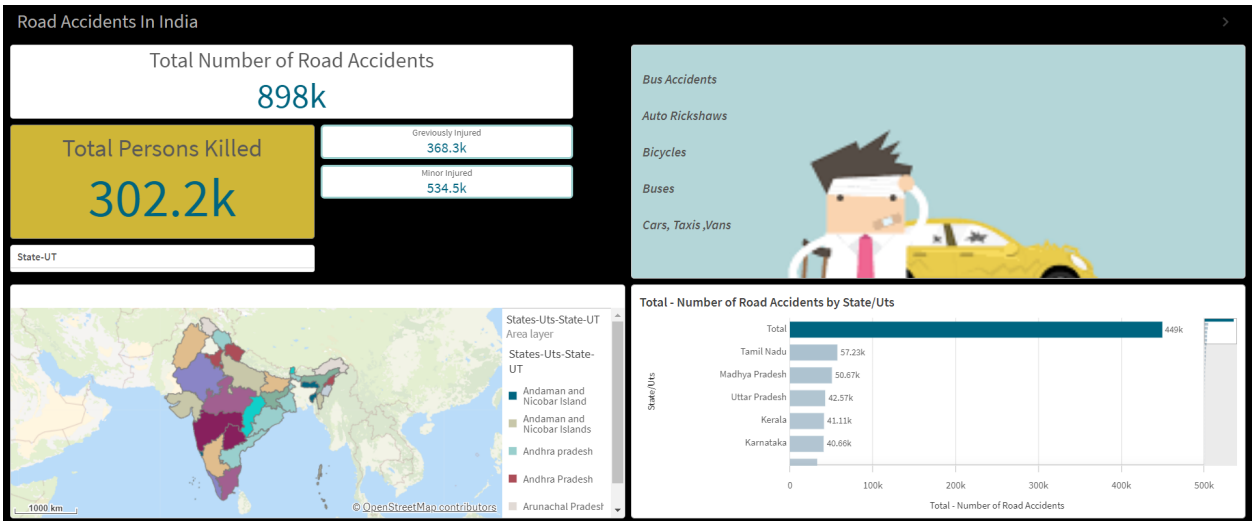
Total number of road accidents by State/UT

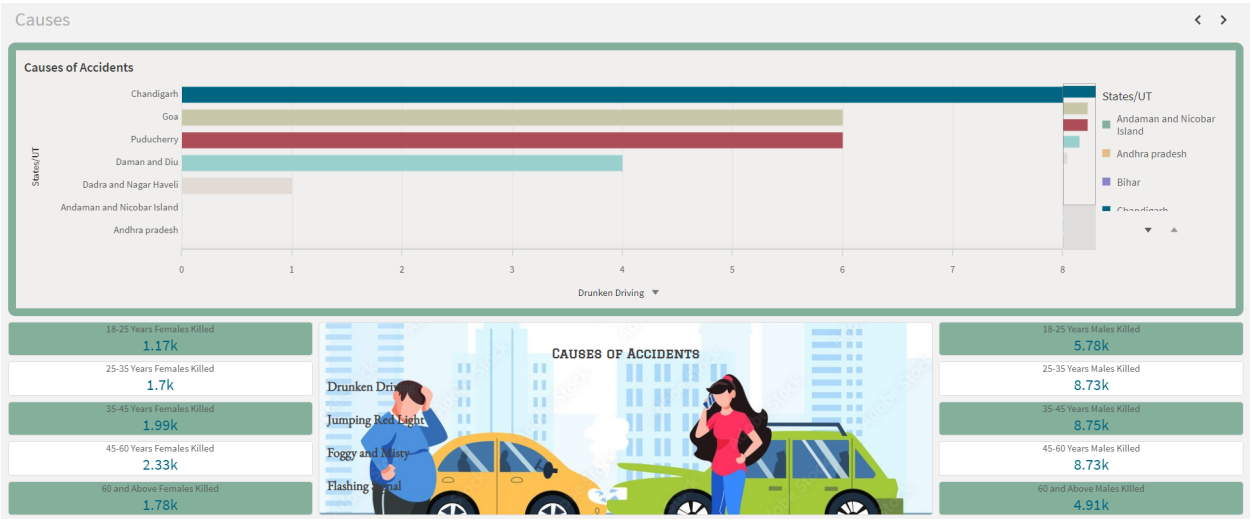


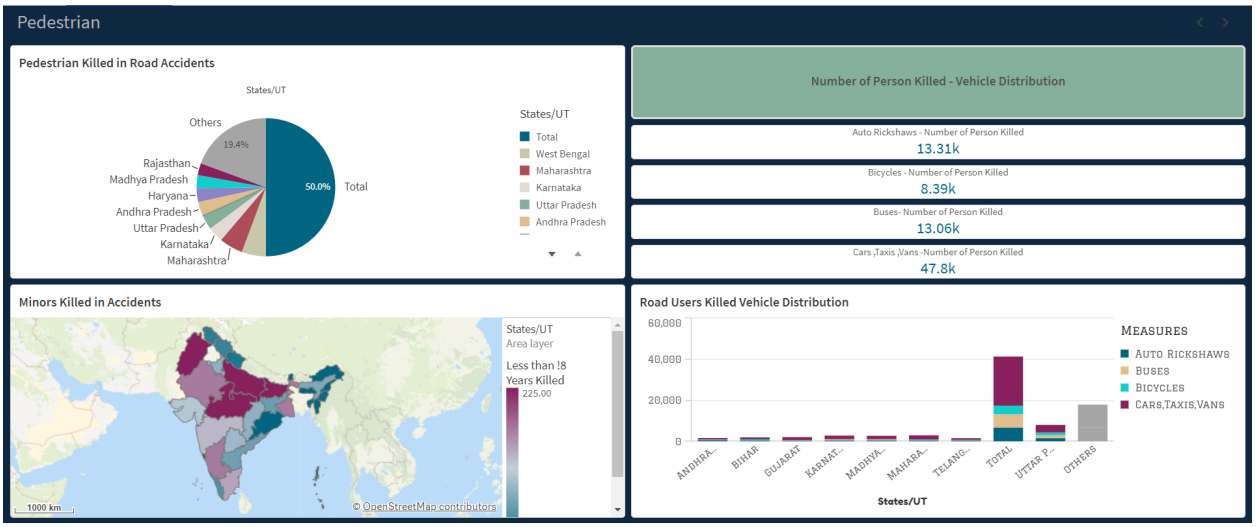
Road accidents distribution

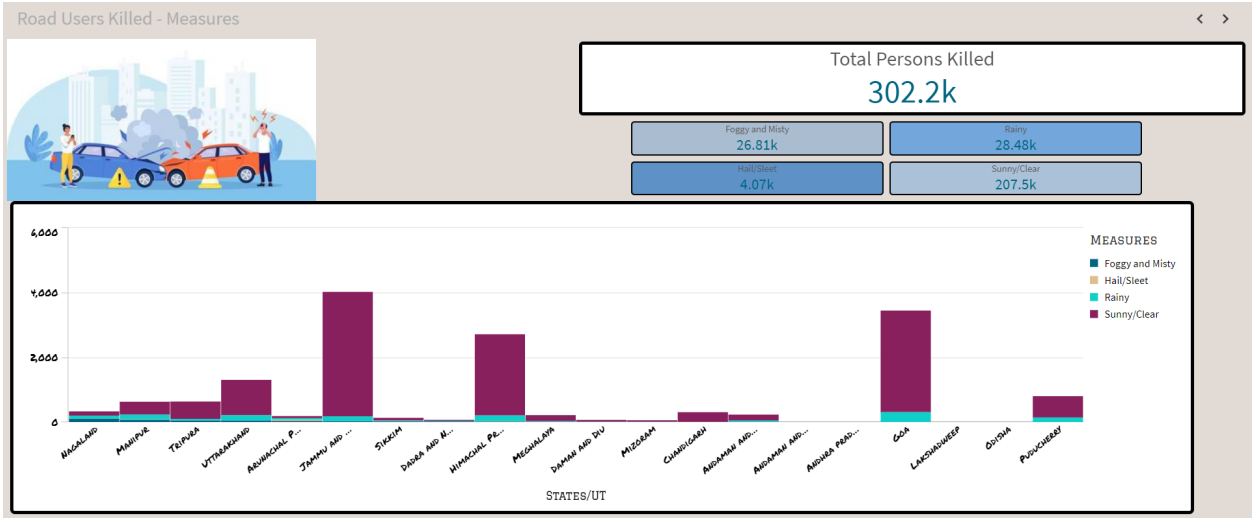


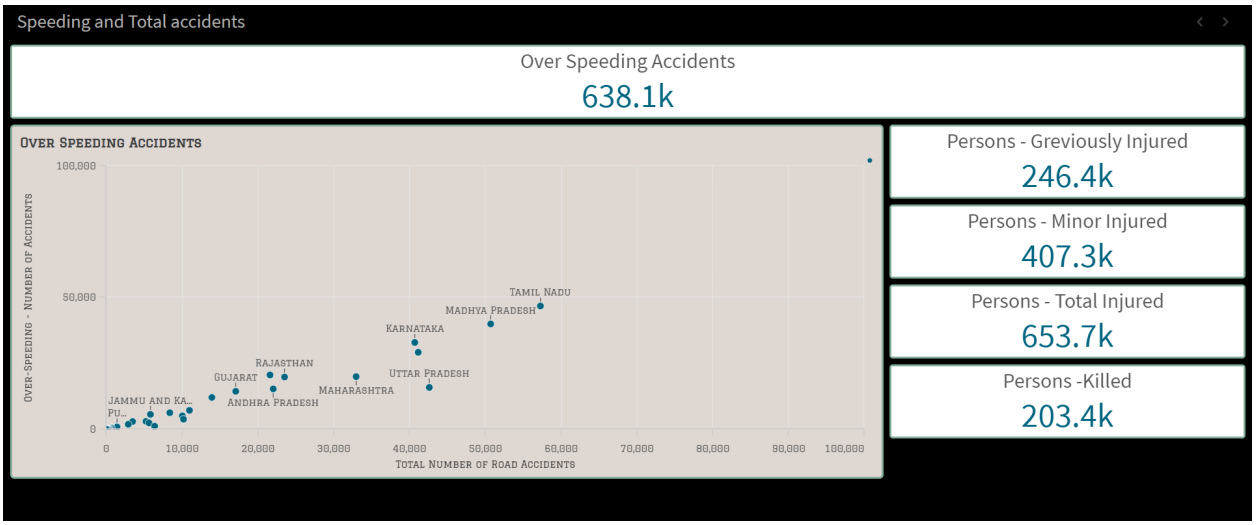
Dashboard:


Road Accidents In India
Total Number of Road Accidents
898k
Total Persons Killed
302.2k
Greviously Injured
368.3k
Minor Injured
534.5k
State-UT
Bus Accidents
Auto Rickshaws
Bicycles
Buses
Cars, Taxis ,Vans
States-Uts-State-UT
Area layer
States-Uts-State-UT
Andaman and Nicobar Island
Andaman and Nicobar Islands
Andhra pradesh
Andhra Pradesh
Arunachal Pradesh
1000 km
© OpenStreetMap contributors
Total - Number of Road Accidents by State/Uts
Total
Tamil Nadu
Madhya Pradesh
Uttar Pradesh
Kerala
Karnataka
449k
57.23k
50.67k
42.57k
41.11k
40.66k
State/Uts
0
100k
200k
300k
400k
500k
Total - Number of Road Accidents


Causes
Causes of Accidents
Chandigarh
Goa
Puducherry
Daman and Diu
Dadra and Nagar Haveli
Andaman and Nicobar Island
Andhra pradesh
States/UT
0
1
2
3
4
5
6
7
8
Drunken Driving
States/UT
Andaman and Nicobar Island
Andhra pradesh
Bihar
18-25 Years Females Killed
1.17k
25-35 Years Females Killed
1.7k
35-45 Years Females Killed
1.99k
45-60 Years Females Killed
2.33k
60 and Above Females Killed
1.78k
CAUSES OF ACCIDENTS
Drunken Driving
Jumping Red Light
Foggy and Misty
Flashing Signal
18-25 Years Males Killed
5.78k
25-35 Years Males Killed
8.73k
35-45 Years Males Killed
8.75k
45-60 Years Males Killed
8.73k
60 and Above Males Killed
4.91k


Pedestrian
Pedestrian Killed in Road Accidents
States/UT
Others
19.4%
Rajasthan
Madhya Pradesh
Haryana
Andhra Pradesh
Uttar Pradesh
Karnataka
Maharashtra
50.0%
Total
States/UT
Total
West Bengal
Maharashtra
Karnataka
Uttar Pradesh
Andhra Pradesh
Number of Person Killed - Vehicle Distribution
Auto Rickshaws - Number of Person Killed
13.31k
Bicycles - Number of Person Killed
8.39k
Buses- Number of Person Killed
13.06k
Cars ,Taxis ,Vans -Number of Person Killed
47.8k
Minors Killed in Accidents
States/UT
Area layer
Less than !8 Years Killed
225.00
1000 km
© OpenStreetMap contributors
Road Users Killed Vehicle Distribution
60,000
40,000
20,000
0
ANDHRA...
BIHAR
GUJARAT
KARNAT...
MADHYA...
MAHARA...
TELANG...
TOTAL
UTTAR P...
OTHERS
States/UT
MEASURES
AUTO RICKSHAWS
BUSES
BICYCLES
CARS,TAXIS,VANS


Road Users Killed - Measures
Total Persons Killed
302.2k
Foggy and Misty
26.81k
Rainy
28.48k
Hail/Sleet
4.07k
Sunny/Clear
207.5k
6,000
4,000
2,000
0
NAGALAND
MANIPUR
TRIPURA
UTTARAKHAND
ARUNACHAL P...
JAMMU AND ...
SIKKIM
DADRA AND N...
HIMACHAL PR...
MEGHALAYA
DAMAN AND DIU
MIZORAM
CHANDIGARH
ANDAMAN AND...
ANDAMAN AND...
ANDHRA PRAD...
GOA
LAKSHADWEEP
ODISHA
PUDUCHERRY
STATES/UT
MEASURES
Foggy and Misty
Hail/Sleet
Rainy
Sunny/Clear



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.To effectively communicate my analysis of road safety and accident patterns, I created a compelling data story. This approach enabled me to present the data and analysis in a narrative format, making the information more engaging and easier to understand.

Performance Testing:

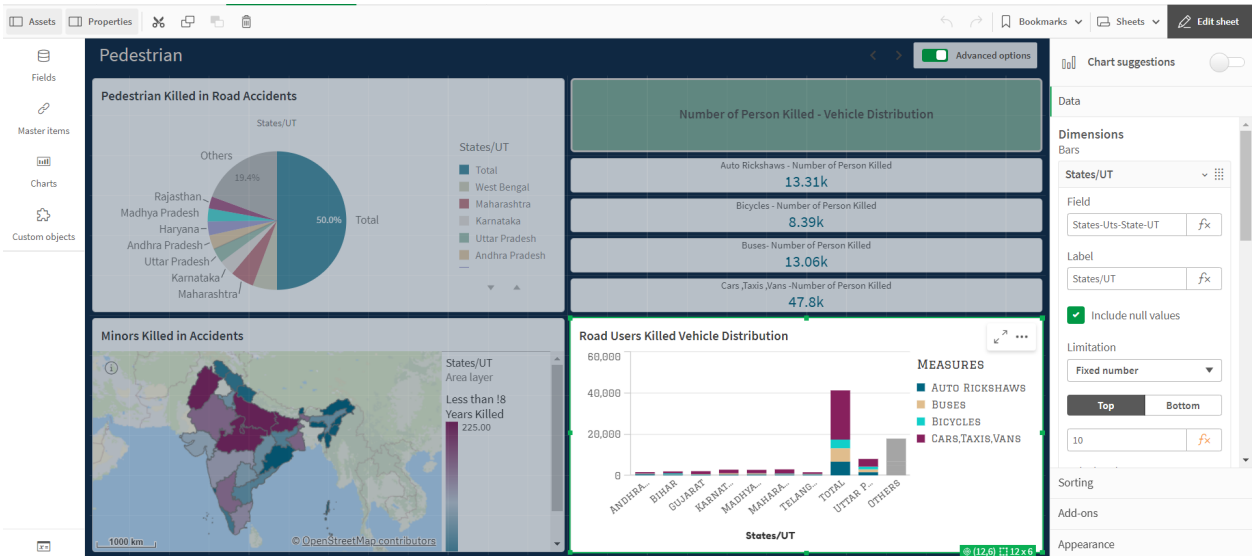
To enhance the interactivity and depth of my road safety and accident patterns analysis, I implemented various selections and filters within the data. This allowed users to customize their views and focus on specific subsets of data based on individual fields or dimensions.

Application Of Data Filters:

To enhance the interactivity and utility of my road safety and accident patterns analysis, I

implemented selections within the data. This feature allowed users to filter data based on individual fields or dimensions, choose specific

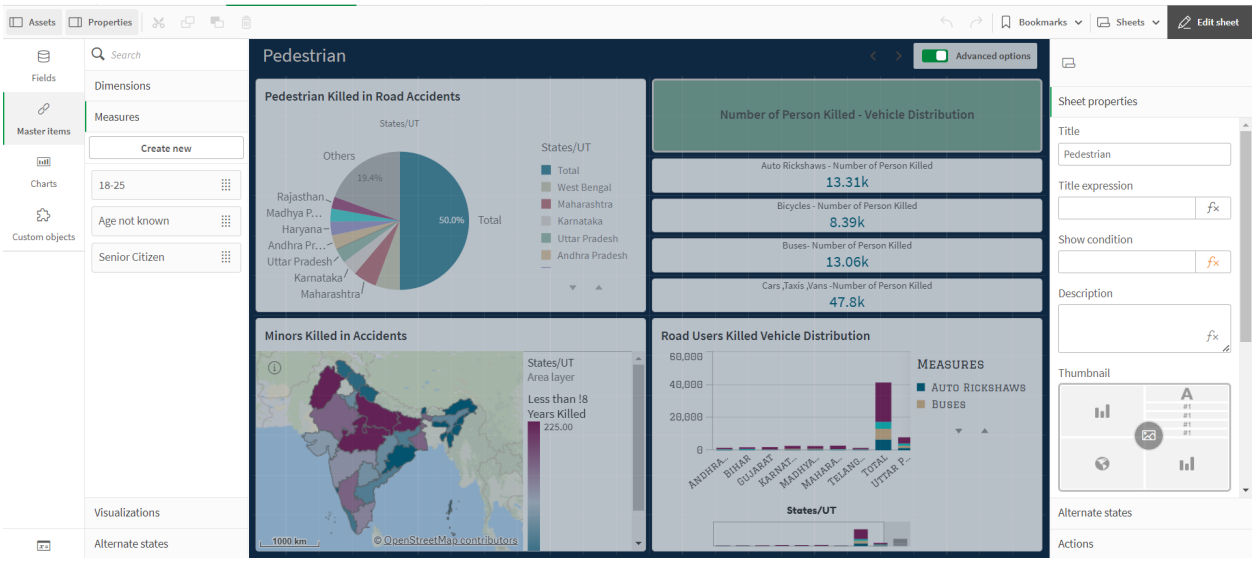
values within a field to include or exclude from analysis, and create complex filters based on predefined conditions and logic.



Use Of Master Items/Calculated Fields:

To streamline and enhance my road safety and accident patterns analysis, I leveraged Qlik

Sense's powerful features, such as reusable filter objects, Master Items, and Calculated Fields. These tools allowed for consistent and efficient application of filters across multiple visualizations and dashboards. Here's how I utilized these features:



Implementing Reusable Filter Objects in Qlik Sense

Master Items:

Creation of Master Items: I created Master Items for frequently used fields such as age group, gender, geographic region, accident severity, and weather conditions. By defining these as Master Items, I ensured consistency and ease of use across all visualizations and dashboards.

Consistent Filtering: Using Master Items allowed me to apply the same filters across different charts and graphs without having to redefine them each time. This not only saved time but also maintained consistency in the analysis.

Calculated Fields:

Defining Calculated Fields: I created Calculated Fields for complex filters that combined multiple conditions. For example, a Calculated Field for “High-Risk Accidents” might combine conditions such as speeding in adverse weather conditions and nighttime driving.

Simplified Analysis: These Calculated Fields enabled me to apply complex logic in a straightforward manner, making it easier to analyze specific scenarios without repeatedly setting up complex conditions.