Qlik Analysis of Road Safety and Accident Patterns in India

1. INTRODUCTION

1.1 Overview

India faces a significant challenge with road safety, where countless individuals lose their lives or sustain serious injuries in road accidents annually. This project aims to address this critical issue by utilizing Qlik, a powerful business intelligence and data visualization platform, to analyze road safety and accident data in India. By examining extensive data from multiple reliable sources, the project seeks to identify patterns and trends that contribute to road accidents. The ultimate goal is to create a safer road environment, reduce the number of accidents, and enhance overall safety measures.

1.2 Purpose

The primary aim of this project is to comprehensively understand the road environment and analyze accidents in India using the Qlik platform. This endeavor is guided by the following objectives:

Identify Patterns and Trends:

- **Temporal and Spatial Analysis**: Gain insights into the patterns of road accidents over time, identifying specific periods of heightened accident rates. Additionally, pinpoint geographic areas most susceptible to accidents.

Inform Policy Decisions:

- **Actionable Insights:** Extract meaningful insights to aid policymakers in formulating effective road safety measures and policies. By comprehending the underlying causes of accidents, such as driver behavior, road conditions, and environmental factors, targeted policies can be devised to address these issues.
- **Resource Allocation:** Facilitate the efficient allocation of resources by identifying high-risk zones necessitating safety enhancements like operational lights, signage, and traffic regulations.

Enhance Public Awareness:

- **Educational Campaigns**: Utilize findings to develop outreach programs focusing on addressing significant causes of road accidents. Such campaigns could tackle issues like speeding, drunk driving, and the importance of seat belt and helmet usage.

- **Community Engagement:** Foster community participation in preventing severe vehicular accidents by making statistical results accessible to the public through interactive dashboards.

Support Research:

- **Data-Driven Research:** Provide researchers with a comprehensive dataset and a range of analytical tools to study road safety and accident prevention. This information can facilitate academic research aimed at identifying the factors contributing to road accidents.
- **Collaborative Studies**: Establish networks and partnerships with stakeholders such as government departments, NGOs, and educational institutions to enhance the sharing and analysis of road safety information.

Improve Infrastructure Planning:

- **Infrastructure Development**: Assist urban planners and engineers in designing safer roads by identifying areas requiring improvement. Strategies for enhanced road designs, safer pedestrian facilities, and improved traffic control can be developed based on the study's findings.
- **Long-Term Planning:** Support long-term urban and regional planning initiatives by providing a historical perspective on road safety and predicting future accident hotspots.

Through the accomplishment of these objectives, this project will make a substantial contribution to enhancing road safety, reducing accident rates, and ultimately saving lives. The comprehensive analysis facilitated by Qlik will serve as a robust foundation for informed decision-making, effective policy formulation, and community engagement in road safety initiatives.

1.3 Technical Architecture

Data Collection,Ingestion&Modeling:

- Utilize various reliable methods to gather data from diverse sources including databases, traffic sensors, and public records. Ensuring the credibility and reliability of the accumulated data is paramount.
- **ETL Processes:** Implement Extract, Transform, Load (ETL) techniques to facilitate the smooth flow of data into the system. This reduces data inconsistency and enhances compatibility.

- **Schema Design:** Develop a comprehensive dataset definition for the data warehouse, outlining tables, fields, and entities along with their interconnectivity.
- **Normalization and Optimization:** Eliminate redundant data access and organize data to enable faster querying and storage.

Data Transformation:

- **Data Cleaning:** Correct missing data, outliers, and erroneous values to ensure the highest data quality standards.
- **Data Enrichment:** Augment raw data with additional contextual information or calculated fields, such as Severity Indices for accidents involving multiple vehicles or geographic location data.

Data Visualization:

- **Visual Representation:** Present data in easily understandable formats such as charts, graphs, and maps.
- **Interactive Elements:** Incorporate filters and interactive features to empower users to analyze data effectively.

Dashboard Designing:

- **User-Friendly Dashboards:** Design simple and intuitive dashboards for easy information relay, catering to users with varying levels of observation.
- **Customization:** Allow users to customize views and choose relevant data for meaningful analysis.

Storytelling Creation:

- **Storytelling:** Develop meaningful narratives that highlight key findings and data conclusions, bringing business insights to life.
- **Contextual Solution:** Provide contextual solutions to interpret trends or patterns within the data effectively.

Reporting and Analytics:

- **In-Depth Reporting:** Generate comprehensive reports showcasing road safety

and accident patterns, tailored to different audience groups.

- **Advanced Analytics:** Apply advanced analytics techniques including predictive modeling and machine learning to reveal accurate indications and anticipate future risks and trends.

By following this technical architecture, the project aims to effectively analyze road safety and accidents in India, providing actionable insights for informed decision-making and policy formulation.

2. DEFINE PROBLEM/PROBLEM UNDERSTANDING

2.1 Business Problem

The key business problem that this project seeks to address include:

- **High Accident Rates**: The persistent high rate of road accidents leading to loss of life and property.
- **Lack of Insights**: Insufficient detailed analysis and understanding of the factors contributing to road accidents.
- **Policy Ineffectiveness**: Existing road safety policies and measures may not be fully effective due to a lack of data-driven insights.
- **Resource Allocation**: Inefficient allocation of resources for road safety improvements due to inadequate information on accident hotspots and trends.

2.2 Business Requirements

To address these business problems, the project has outlined the following business requirements:

- **Data Integration**: Integrate diverse datasets from multiple sources to create a comprehensive database of road accidents.
- **Interactive Dashboards**: Develop interactive dashboards using Qlik to visualize road accident data in an accessible manner.
- **Trend Analysis**: Perform trend analysis to identify temporal patterns and high-risk periods for road accidents.
- **Geographical Analysis**: Conduct geographical analysis to pinpoint accident hotspots and areas requiring targeted interventions.
- **Cause Analysis**: Analyze the causes of accidents to identify key factors contributing to high accident rates.

2.3 Literature Survey

A literature survey was conducted to understand the current state of research and methodologies used in the analysis of road safety and accident patterns. Key findings from the survey include:

2.3.1 Existing Research

- **Patterns and Trends**: Studies have identified various temporal and spatial patterns in road accidents, emphasizing the importance of detailed data analysis for understanding these patterns.
- **Risk Factors**: Research highlights several risk factors, including driver behavior, road conditions, vehicle types, and environmental conditions, which significantly impact accident rates.
- **Data Visualization**: The use of data visualization tools, like Qlik, has been shown to effectively communicate complex data insights, making it easier for stakeholders to understand and act upon the findings.

2.3.2 Methodologies

- **Data Mining Techniques**: Techniques such as clustering, regression analysis, and association rule mining have been employed to uncover hidden patterns and relationships in road accident data.
- **Geospatial Analysis**: Geographic Information Systems (GIS) and spatial analysis methods are widely used to identify accident hotspots and analyze spatial distribution.
- **Predictive Modeling**: Predictive models have been developed to forecast accident occurrences and identify potential high-risk areas, aiding in proactive safety measures.

2.3.3 Gaps in Existing Research

- **Real-Time Data Analysis**: Limited research on the integration and analysis of real-time data for immediate insights and interventions.
- **Comprehensive Datasets**: Many studies rely on limited datasets, highlighting the need for more comprehensive and integrated data sources.
- **User-Centric Visualizations**: There is a gap in research focused on developing user-centric visualizations that cater to diverse stakeholders, including policymakers, road safety authorities, and the general public.

By addressing these gaps and leveraging the capabilities of Qlik for data visualization and analysis, this project aims to provide a more detailed and actionable understanding of road safety and accident patterns in India.

3. DATA COLLECTION

3.1 Collect the Dataset

Data collection is a crucial step in any data analysis project. For this project on road accidents in India, the dataset was obtained from Kaggle, a renowned platform for data science enthusiasts. The dataset provides comprehensive information on road accidents across various states and territories in India for the year 2019.

Dataset Link: RoadAccidentsinIndiaDataset

3.2 Understand the Dataset

The dataset contains detailed information on various aspects of road accidents, including the number of accidents, fatalities, injuries, types of vehicles involved, and accident locations. It consists of nine Excel worksheets, each focusing on different aspects of road accidents in India during the year 2019.

3.3 Description of the Dataset

The dataset comprises the following Excel worksheets:

1. **Pedestrians**: State/UT-wise pedestrians involved in accidents classified by age and gender during 2019.
2. **Pedestrians Killed**: State/UT-wise pedestrians killed classified by age and gender during 2019.
3. **Pedestrians Killed – Impacting Vehicles**: State/UT-wise pedestrians killed classified by the type of impacting vehicles during 2019.
4. **Traffic Control Type**: State/UT-wise accidents classified by the type of traffic control during 2019.
5. **Weather**: State/UT-wise accidents classified by the type of weather during 2019.
6. **Killed on Two Wheelers – Impacting Vehicles**: State/UT-wise two-wheelers killed classified by the type of impacting vehicles during 2019.
7. **Road Users Killed – Gender**: State/UT-wise male and female persons killed in road accidents by road user categories during 2019.
8. **Causes**: State/UT-wise accident victims classified by the causes of accidents during 2019.
9. **Accidents – Severity and Vehicles**: State/UT-wise vehicle types of victims and

severity of accidents during 2019.

Each worksheet contains specific columns related to the respective category of road accidents, providing detailed insights into various aspects of road safety and accident patterns in India.

3.4 Connect Data with Qlik Sense

To connect the dataset with Qlik Sense for analysis and visualization, follow these steps:

1. **Create a New Data Connection**: Use the data manager in Qlik Sense to create a new data connection.
2. **Select Data Source**: Choose the appropriate data source option based on where the dataset is stored (e.g., local file, web file).
3. **Configure Connection Settings**: Provide the necessary details for the data source, such as file path or URL.
4. **Preview and Load Data**: Preview the data to ensure it is loaded correctly, make any necessary transformations or associations, and then load the data into Qlik Sense.

By connecting the dataset with Qlik Sense, you can leverage its powerful features for data analysis and visualization, gaining valuable insights into road safety and accident patterns in India.

4. DATA PREPARATION

4.1 Prepare the Data for Visualization

The process of preparing the dataset for visualization in Qlik Sense involves several systematic steps, ensuring the seamless integration and optimization of the data for analysis. Below is a detailed walkthrough of the data preparation process:

- **Dataset Link**: The dataset utilized for this analysis is sourced from Kaggle and focuses on road accidents in India. Road Accidents in India Dataset
- **Access Qlik Sense Cloud**: Begin by logging into Qlik Sense Cloud, the platform chosen for conducting data analysis and visualization.
- **Home Page Navigation**: Upon logging in, navigate to the home page of Qlik Sense data analytics, where various options and functionalities are available.
- **Creation of New Analytics App**:
    - Click on the "+ add new" button.
    - Select "New Analytics App" from the dropdown list.
    - Fill in the necessary details in the form provided, assigning a name to the app, such as "Accident Data Analysis App".
    - Confirm the creation of the app.
- **Files and Other Resources**: Once inside the newly created app, locate and click on "Files and Other Resources" to initiate the data loading process.
- **Addition of Datasets**:
    - Utilizing the drag-and-drop functionality, add each of the nine datasets related to road accidents in India into Qlik Sense one by one.

- **Data Manager Interaction**:
    - As each dataset is added, Qlik Sense redirects to the Data Manager interface for further processing.
    - Repeat the data loading process for all datasets to ensure comprehensive inclusion.
- **Combination of Datasets**:
    - Upon successful loading of all datasets, click on "Apply All" to combine them into a unified dataset within Qlik Sense.
- **Optional Dataset Name Updates**:
    - If necessary, modify dataset names or make other adjustments by accessing the associate table within Qlik Sense.

    - Ensure all changes are saved and reflected in the dataset.
- **Data Loading Confirmation**:
    - After any modifications or updates, click on "Load Data" to confirm the changes and finalize the dataset.
- **Data Readiness for Visualization**:
    - With the completion of the data preparation process, the dataset is now ready for visualization within Qlik Sense, facilitating insightful analysis of road accidents in India.

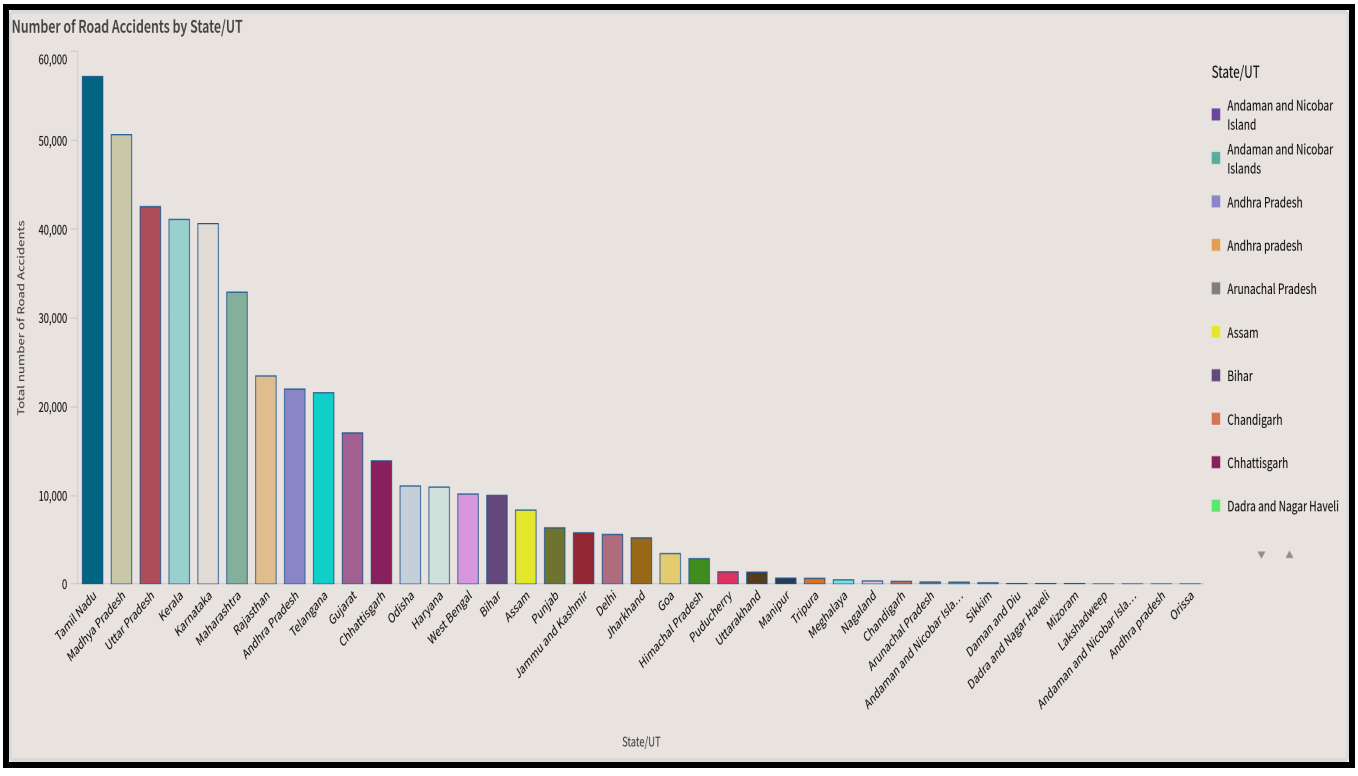
By meticulously following these steps, the dataset is seamlessly integrated into Qlik Sense, laying the groundwork for comprehensive analysis and visualization of road accident patterns.

5. Data Visualization

5.1 Visualizations

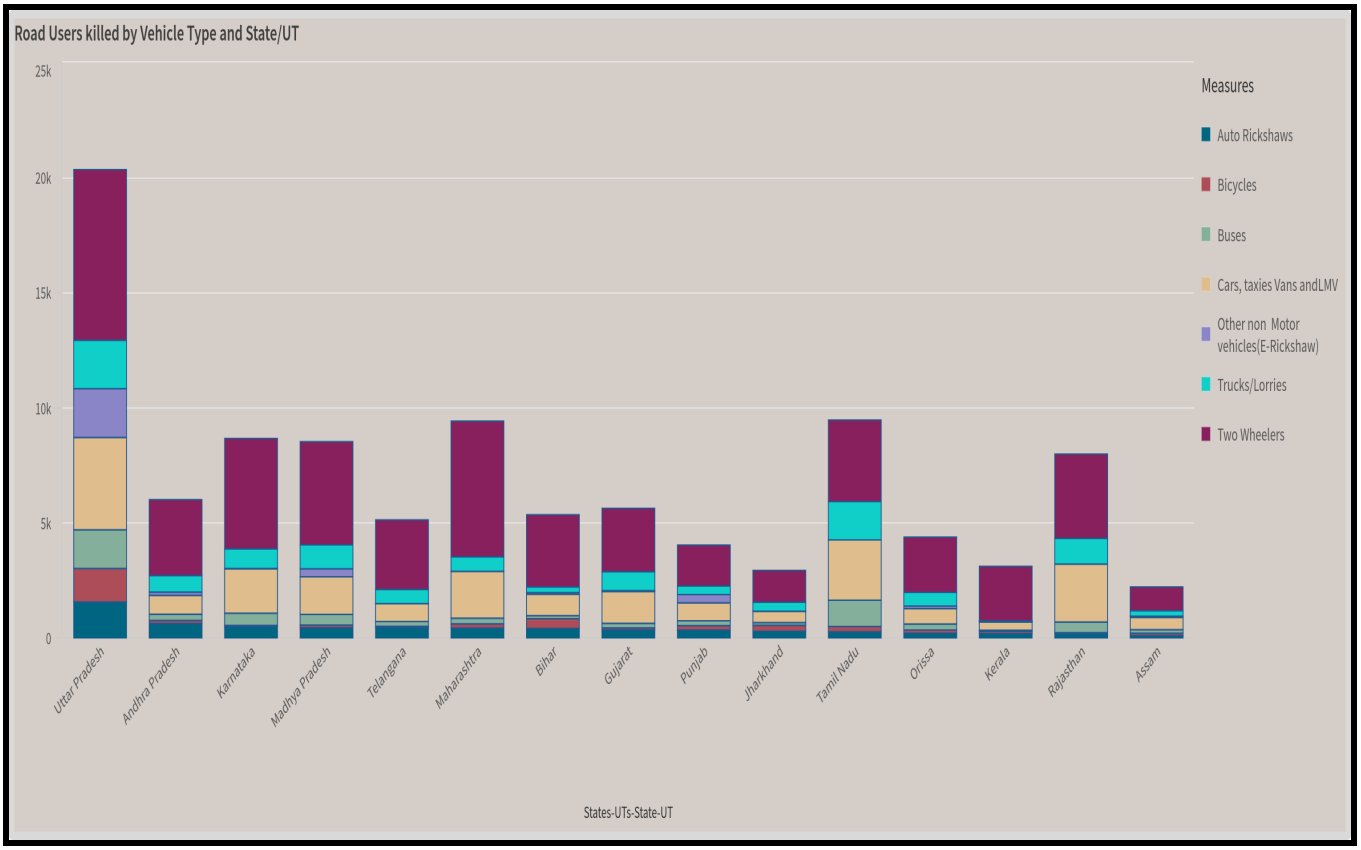
Visualisation 1: Number of Road Accidents by State/UT

The bar chart ranks Indian states and Union Territories by the total number of road accidents. Tamil Nadu tops the list with over 60,000 accidents, followed by Madhya Pradesh and Uttar Pradesh, each with around 50,000. Kerala and Karnataka report approximately 40,000 accidents each, while Maharashtra and Rajasthan have around 30,000. States like Gujarat, Chhattisgarh, and Odisha fall in the 20,000 to 30,000 range. The lowest numbers are seen in Dadra and Nagar Haveli, Lakshadweep, and Arunachal Pradesh, each with fewer than 5,000 accidents. This chart identifies key areas for road safety improvements and resource allocation.

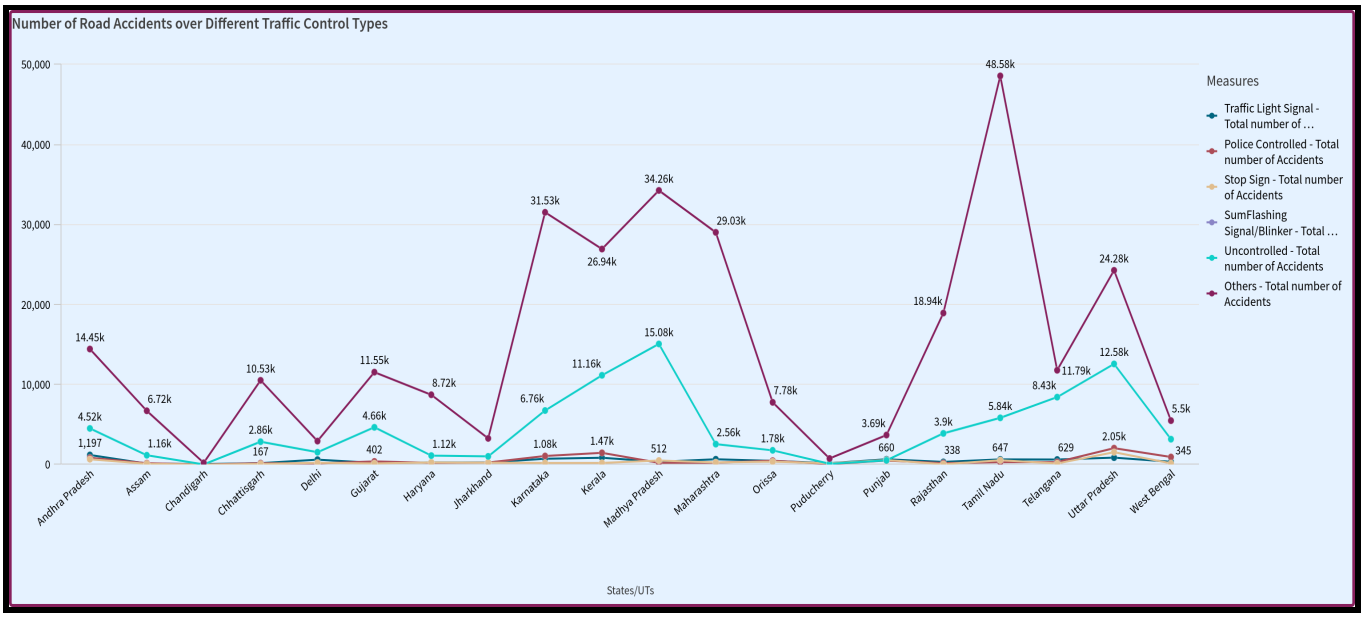


Visualisation 2 : Road Users killed by Vehicle Type and State/UT

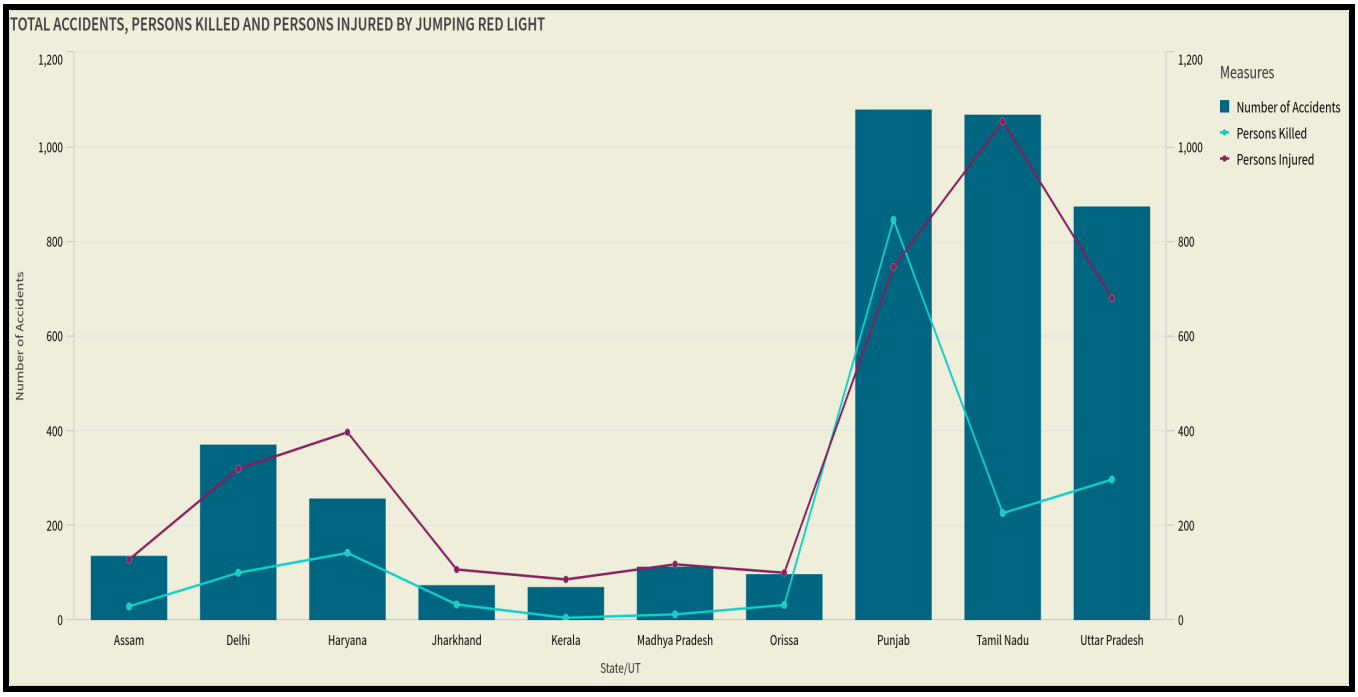
The chart shows the number of road users killed by different types of vehicles in Indian states and union territories. Uttar Pradesh has the highest number of road users killed, followed by Karnataka and Andhra Pradesh. Two-wheelers are the most common type of vehicle involved in road accidents. The chart highlights the need for road safety measures to reduce accidents and fatalities. It can be used to identify areas where measures are most needed and to track their effectiveness over time.



Visualisation 3 : Number of Road Accidents over Different Traffic Control Types

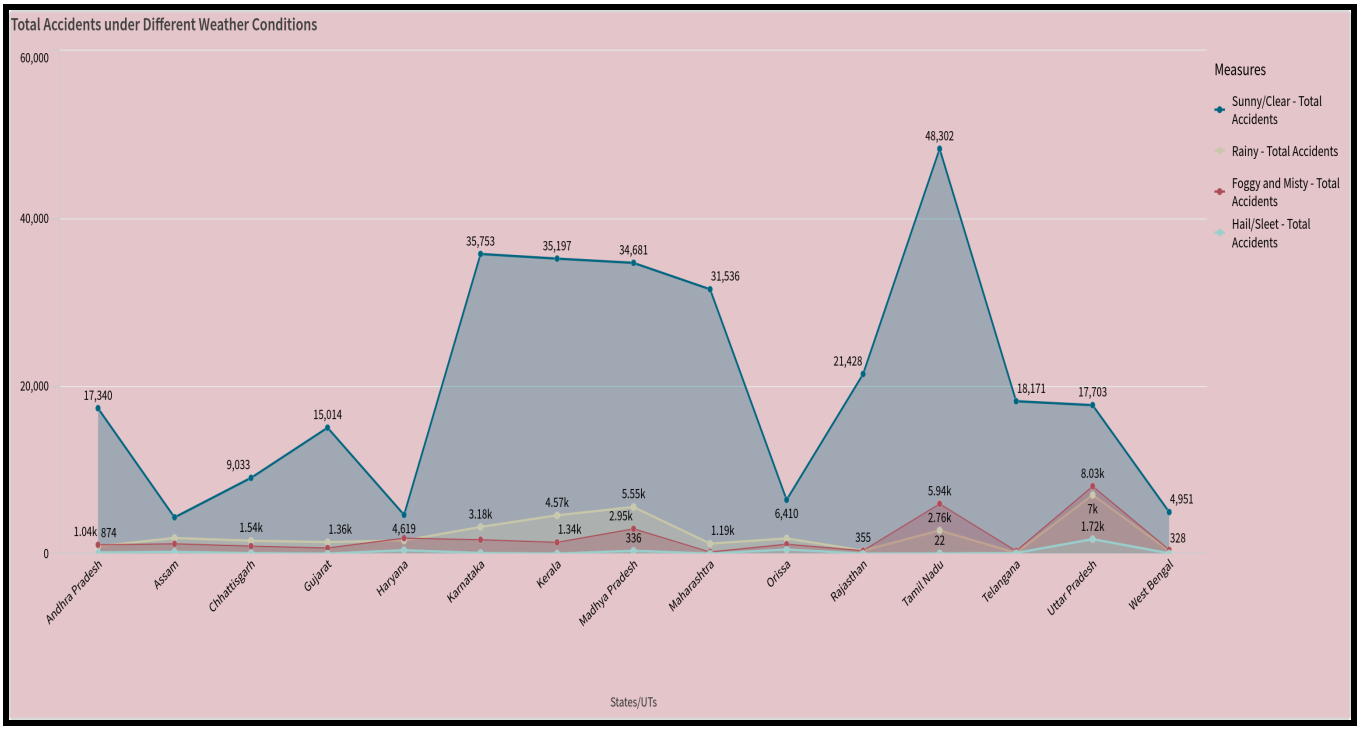


Visualisation 4:Total Accidents, Persons and Persons Injured by Jumping Red Light



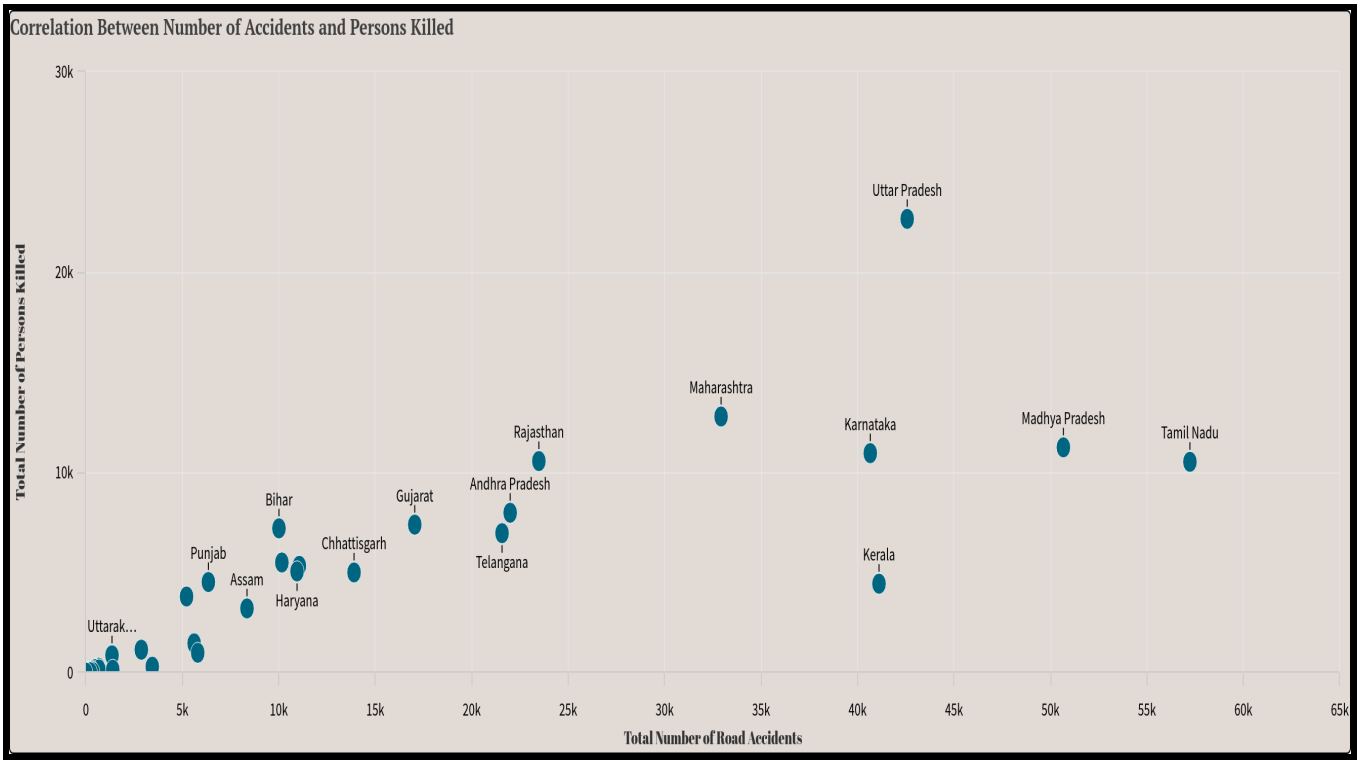
Visualisation 5: Toatal Accidents under different Weather Conditions

This bar chart shows the total number of accidents that occurred under different weather conditions in various states and union territories in India. The state of Andhra Pradesh had the highest number of accidents at 17,340, followed by Karnataka at 35,753, and Tamil Nadu at 48,302. The least number of accidents occurred in West Bengal at 4,951.



Visualisation 6: Correlation between Number of Accidents and Persons killed

This scatter plot shows the correlation between the number of road accidents and the number of people killed in each state of India. It appears there is a strong positive correlation between these two variables, indicating that a higher number of accidents generally results in a higher number of people killed. The states with the highest number of road accidents are also the states with the highest number of fatalities.



Visualisation 7: State/UT wise persond killed and Injured due to Drunken Driving

The table provides insights into the prevalence of drunk driving accidents across various states and union territories in India. Jharkhand records the highest number of accidents at 686, followed by Assam with 279 and Karnataka with 132. Conversely, Dadra and Nagar Haveli report only 1 accident. This data underscores the severity of the drunk driving issue, guiding efforts to target prevention strategies where they're most urgently needed.

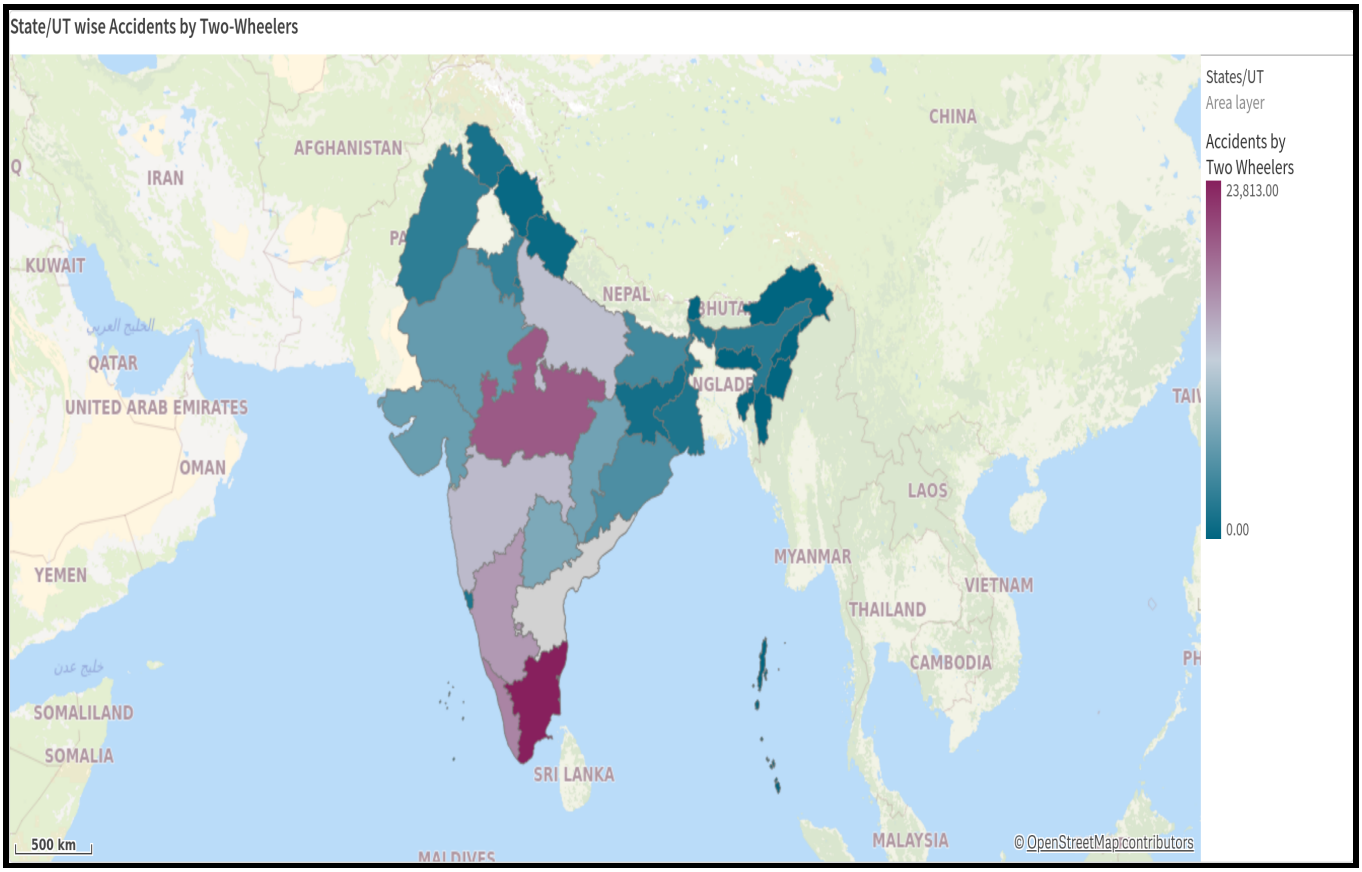
3.5

State-wise Persons killed and Injured due to drunken driving

States/UT	Values			
	Total number of accidents due to drunken driving	Persons Killed	Greviously Injured	Minor Injured
Andaman and Nicobar Island	0	0	0	0
Andaman and Nicobar Islands	41	1	12	24
Andhra Pradesh	127	43	48	96
Andhra pradesh	0	0	0	0
Arunachal Pradesh	46	27	34	36
Assam	279	74	273	33
Bihar	0	0	0	0
Chandigarh	8	0	0	8
Chhattisgarh	134	34	51	61
Dadra and Nagar Haveli	1	0	0	0
Daman and Diu	4	0	2	2
Delhi	215	45	29	173
Goa	6	1	0	0
Gujarat	47	17	18	16
Haryana	299	132	115	165
Himachal Pradesh	75	24	31	75
Jammu and Kashmir	25	5	44	0
Jharkhand	686	578	305	77
Karnataka	132	37	62	104

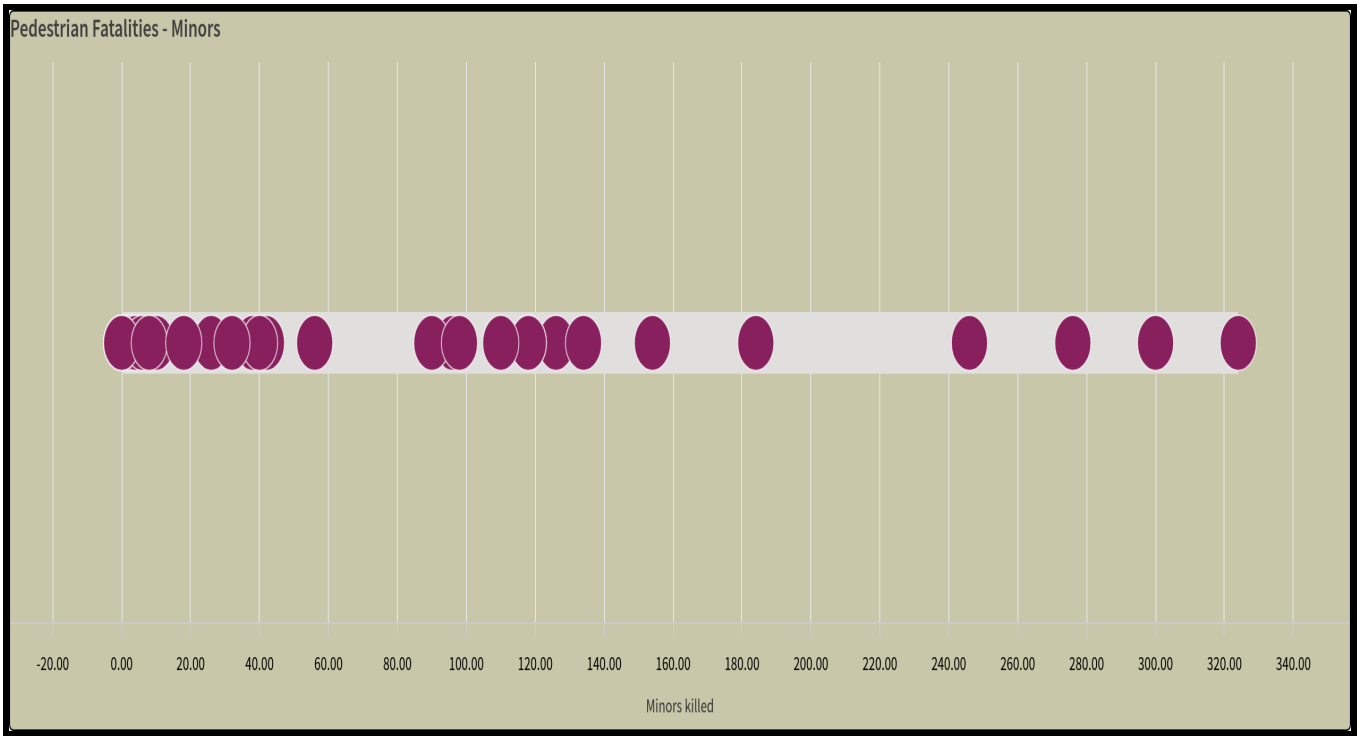
Visualisation 8: State/UT wise Accidents by Two-Wheelers

The map shows the number of accidents by two-wheelers in each state or union territory of India. The data shows a wide range of accidents across different states. The darker the color, the higher the number of accidents. For example, Tamil Nadu and Kerala appear to have the highest number of accidents by two-wheelers. The number of accidents in the North-eastern states appear to be the lowest.



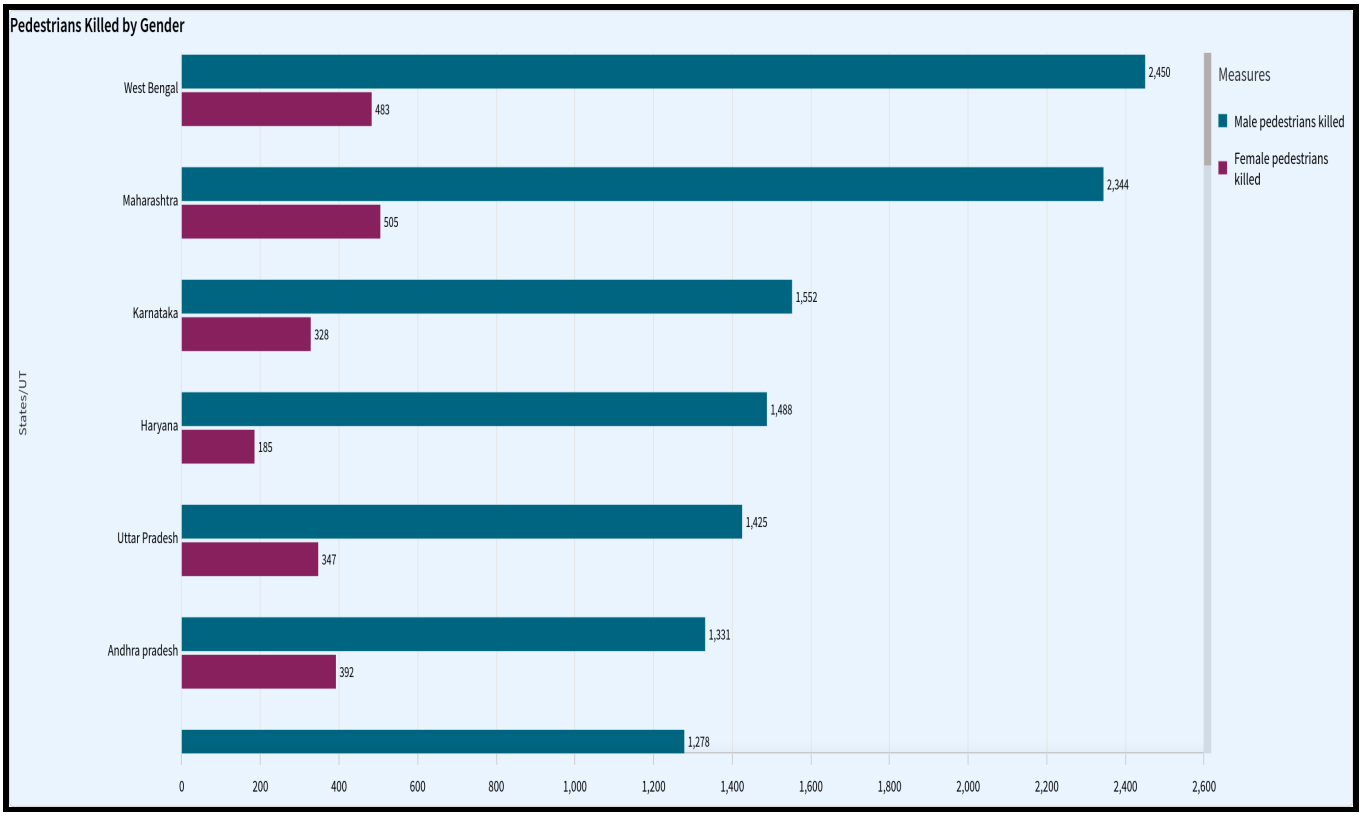
Visualisation 9: Pedestrian Fatalities - Minors

This chart shows the number of minors killed in pedestrian accidents. There were 318 minors killed in pedestrian accidents. This is a significant number, and it highlights the importance of pedestrian safety. Drivers should be aware of their surroundings and be extra cautious when driving near children. Pedestrians should also be aware of their surroundings and take steps to stay safe.



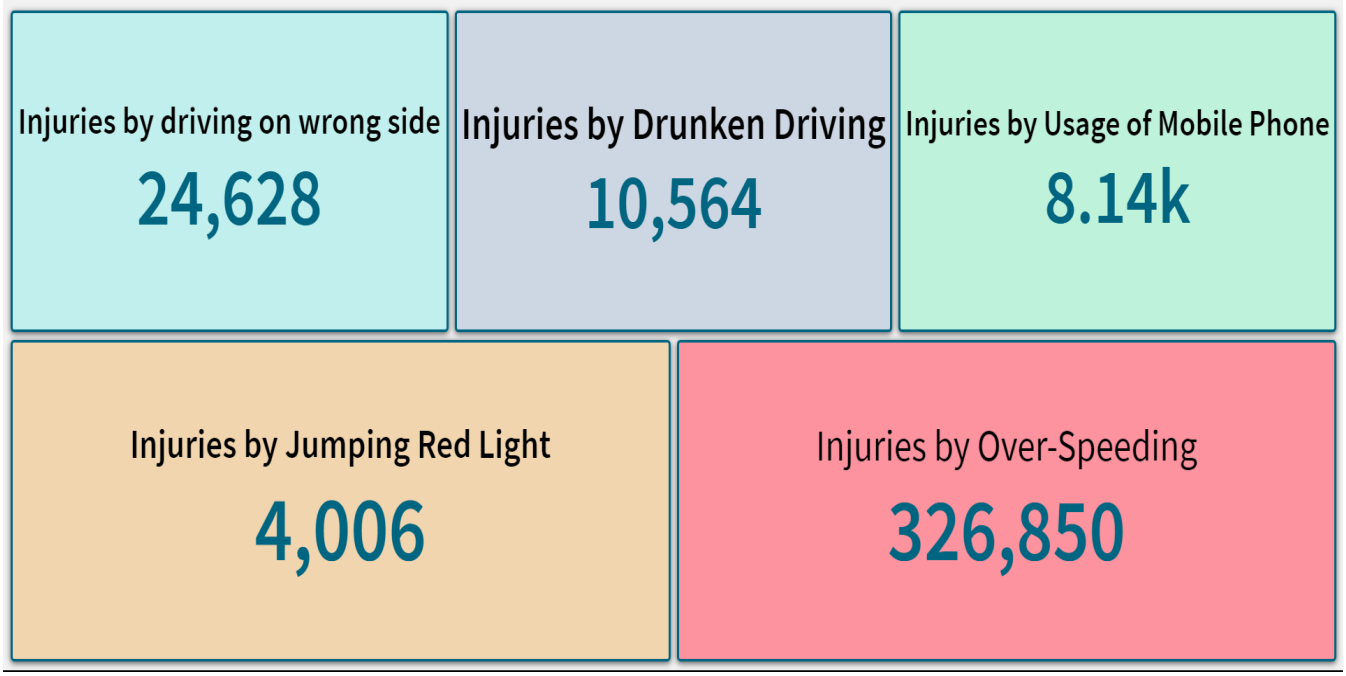
Visualisation 10 : Pedestrians killed by Gender

This bar chart shows the number of male and female pedestrians killed in different states/UTs of India. The states/UTs are sorted by the total number of pedestrians killed. The highest number of pedestrian deaths were recorded in West Bengal, followed by Maharashtra and Karnataka.



Visualisation 11: KPI CHART showing Total Injuries due to various causes

This Visualization shows the number of injuries caused by different driving violations. The most common cause is over speeding, followed by driving on the wrong side of the road. Drunken driving is responsible for a significantly lower number of injuries. Overall, these statistics highlight the importance of safe driving practices to prevent accidents and injuries.



6. DASHBOARD

6.1 Responsive Design of Dashboard [DASHBOARD LINK]

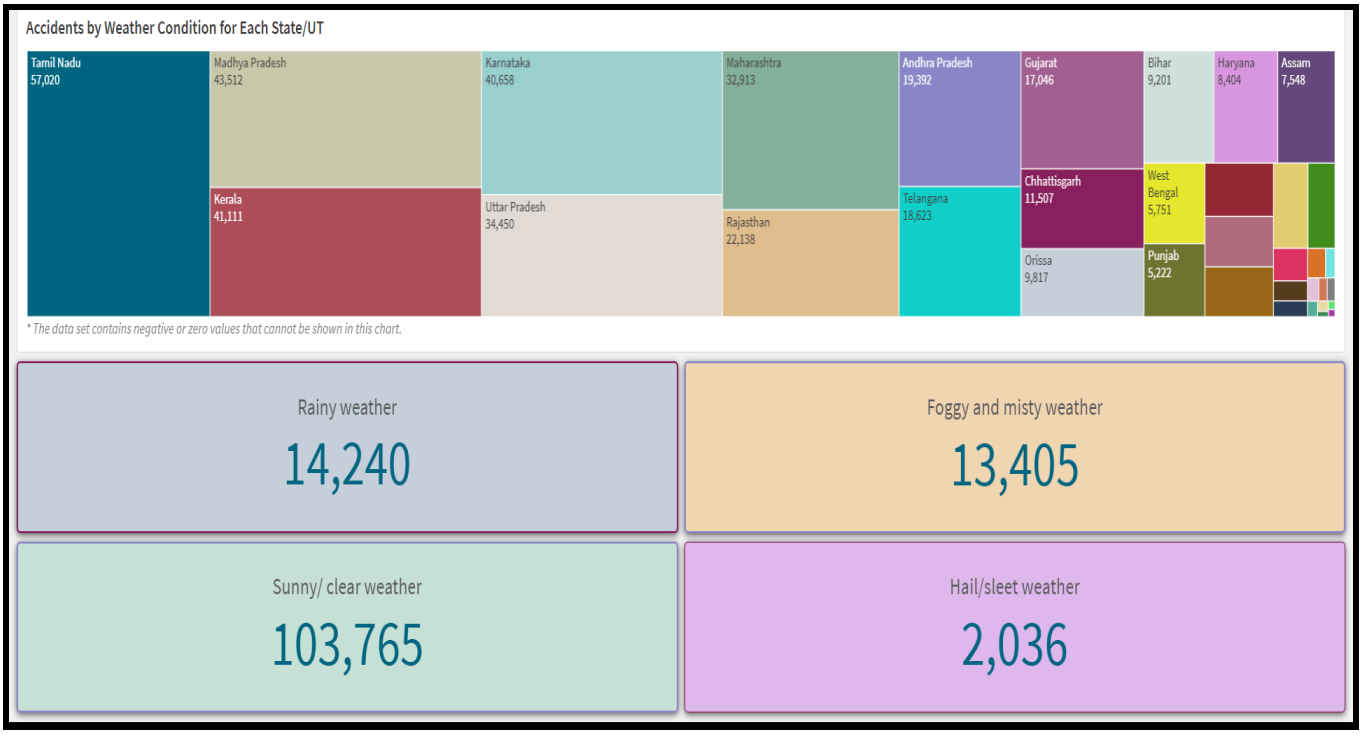
Dashboard 1 :

This Dashboard shows the number of accidents by weather condition for each state/UT in India. The Dashboard is divided into two sections:

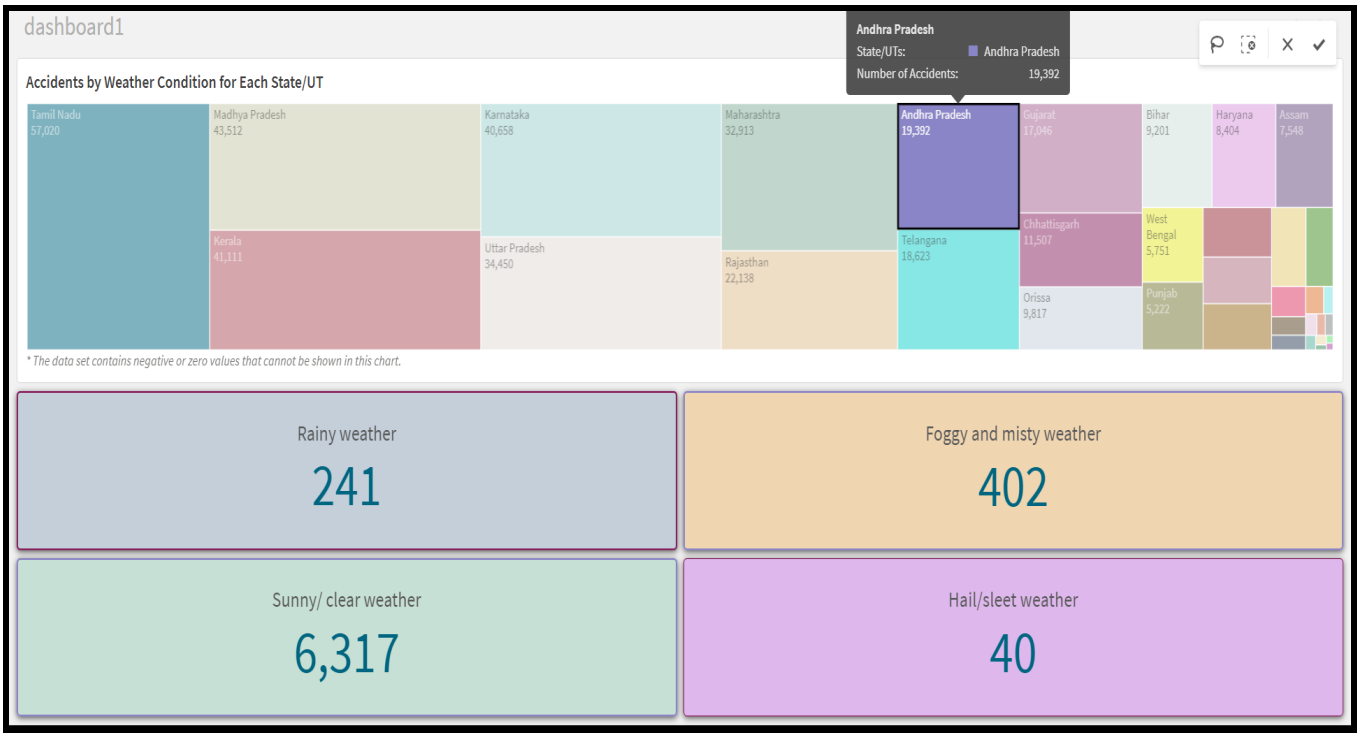
- The top section shows the number of accidents for each state/UT, broken down by weather condition.
- The bottom section shows the total number of accidents for each weather condition, regardless of state/UT.

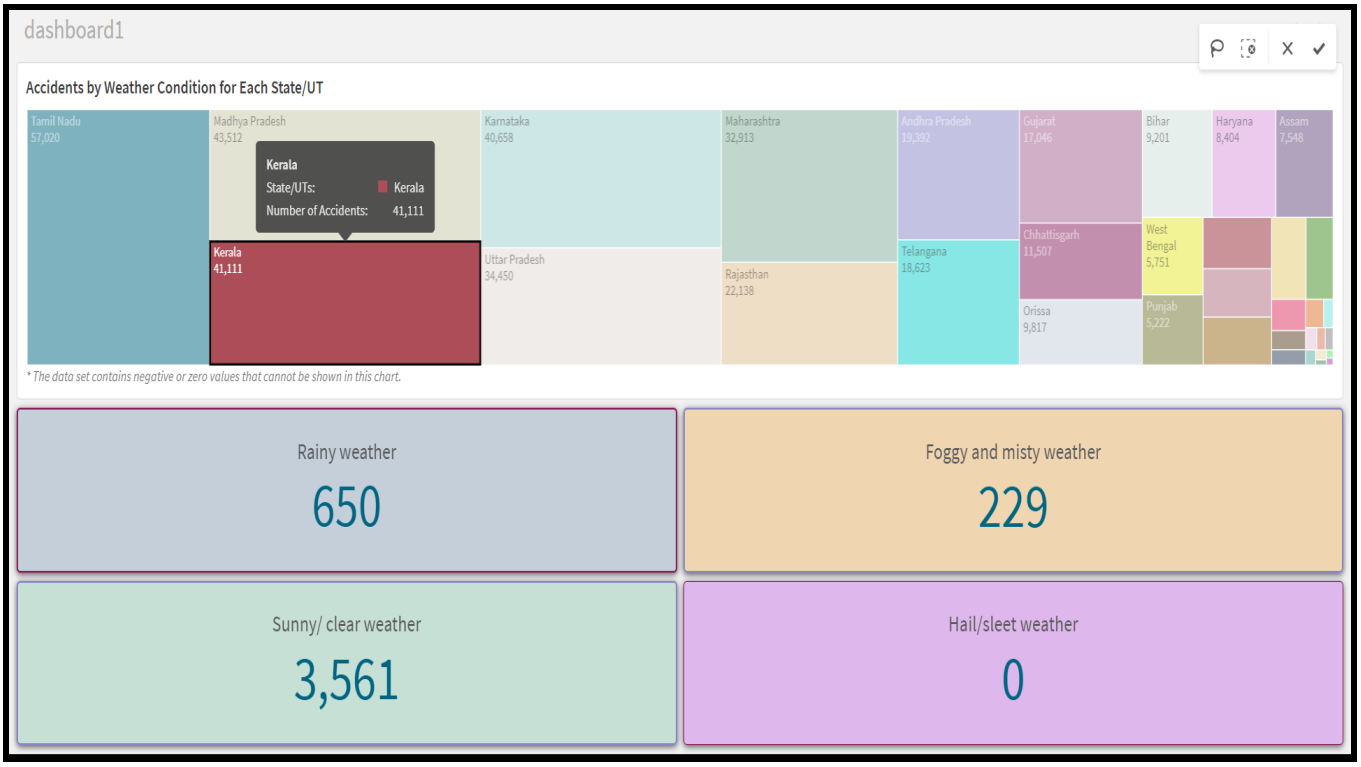
The dashboard shows that:

- The majority of accidents occur during sunny/clear weather.
- The next most common weather condition for accidents is foggy and misty weather.
- Rainy weather accounts for a relatively small proportion of accidents.
- Hail/sleet weather is the least common weather condition for accidents.



The Responsiveness of Dashboard 1 :

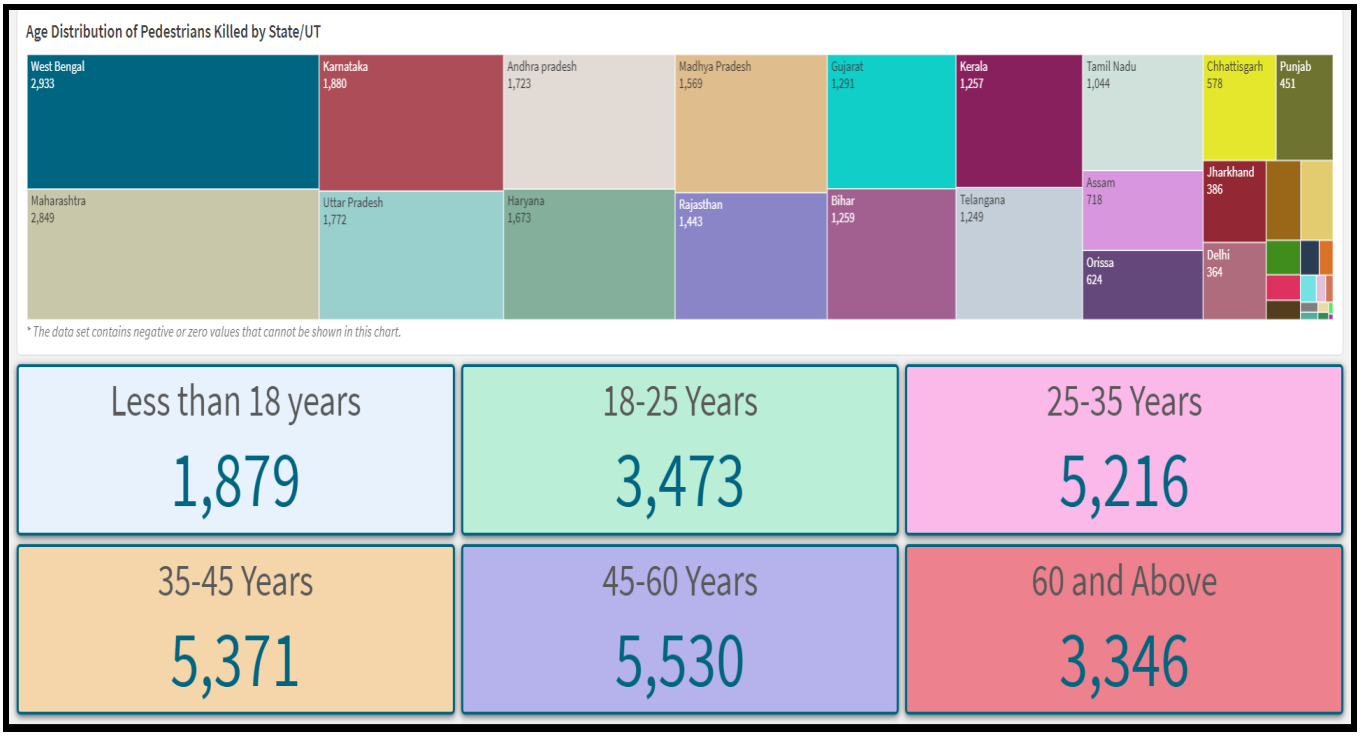




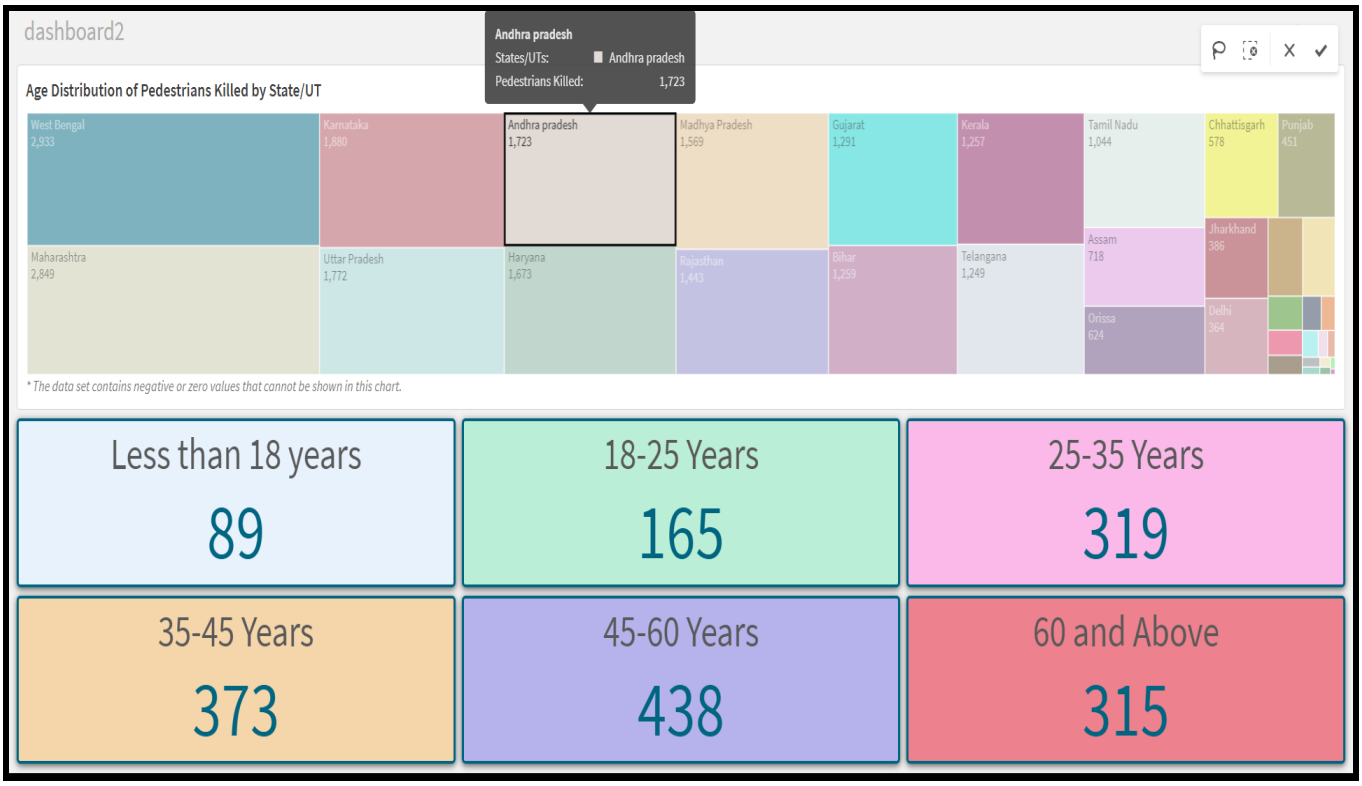
Dashboard 2 :

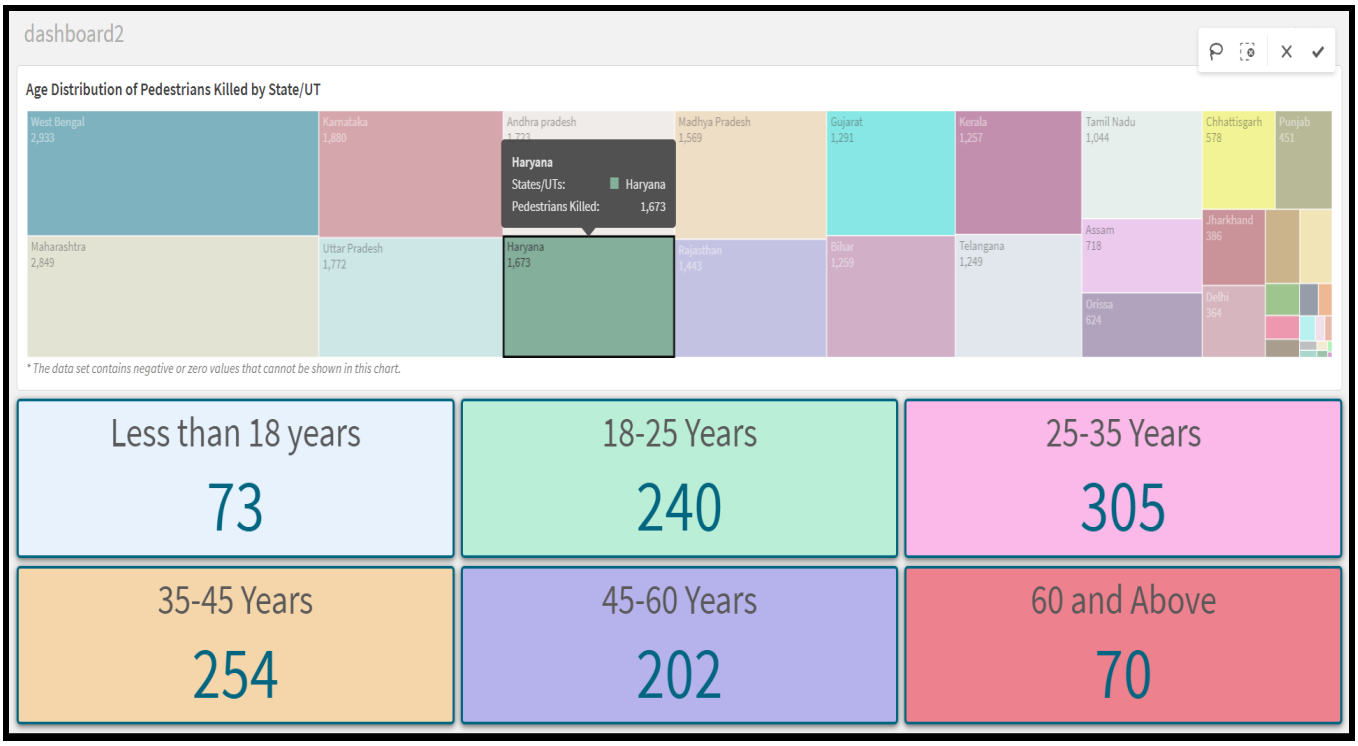
This dashboard shows the age distribution of pedestrians killed by state/UT.

- The top part of the dashboard shows the number of pedestrians killed by state/UT.
- The bottom part of the dashboard shows the number of pedestrians killed in different age groups.
- The age groups are less than 18 years, 18-25 years, 25-35 years, 35-45 years, 45-60 years, and 60 and above.
- This dashboard is useful for understanding the age distribution of pedestrians killed in different states/UTs.



The Responsiveness of Dashboard 2 :

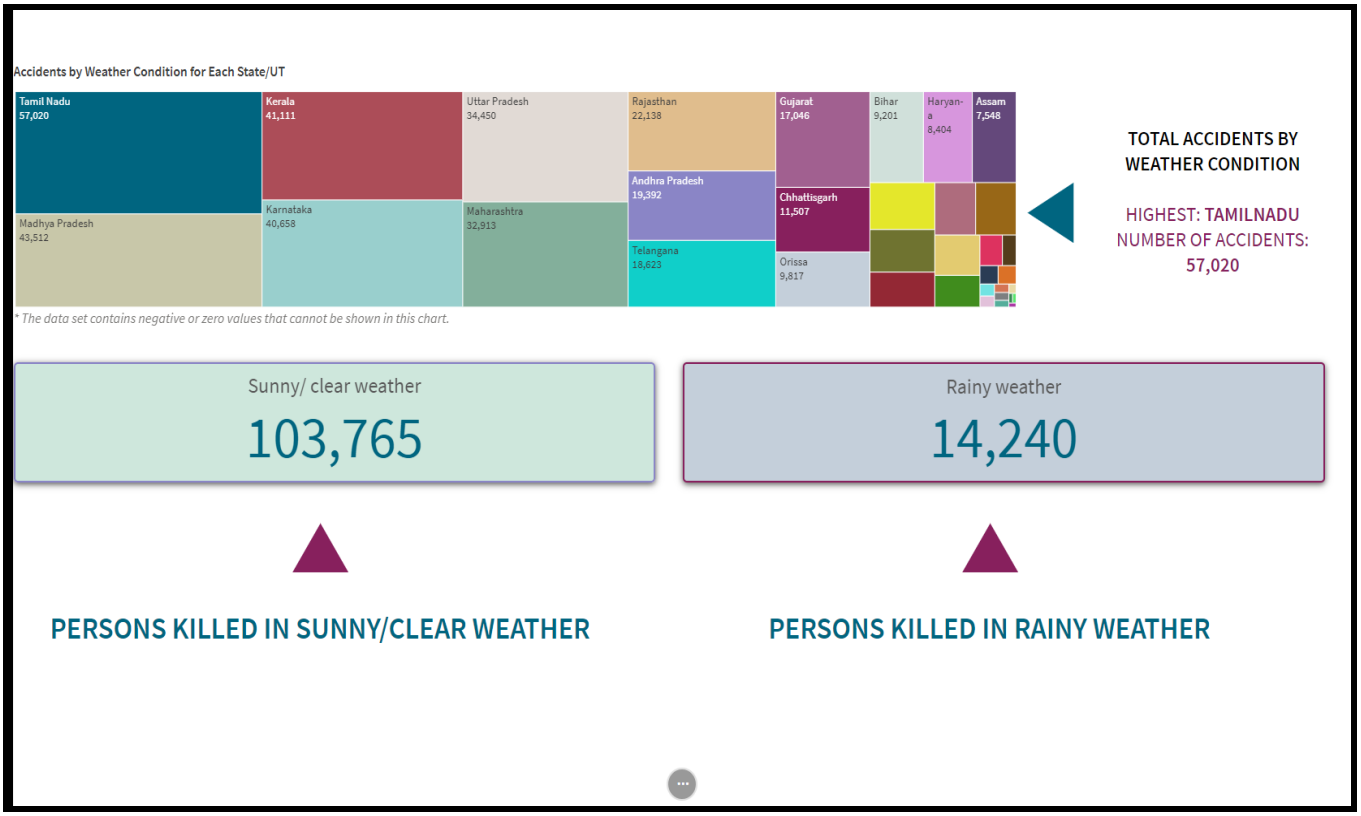




STORY TELLING : [STORYTELLING LINK]

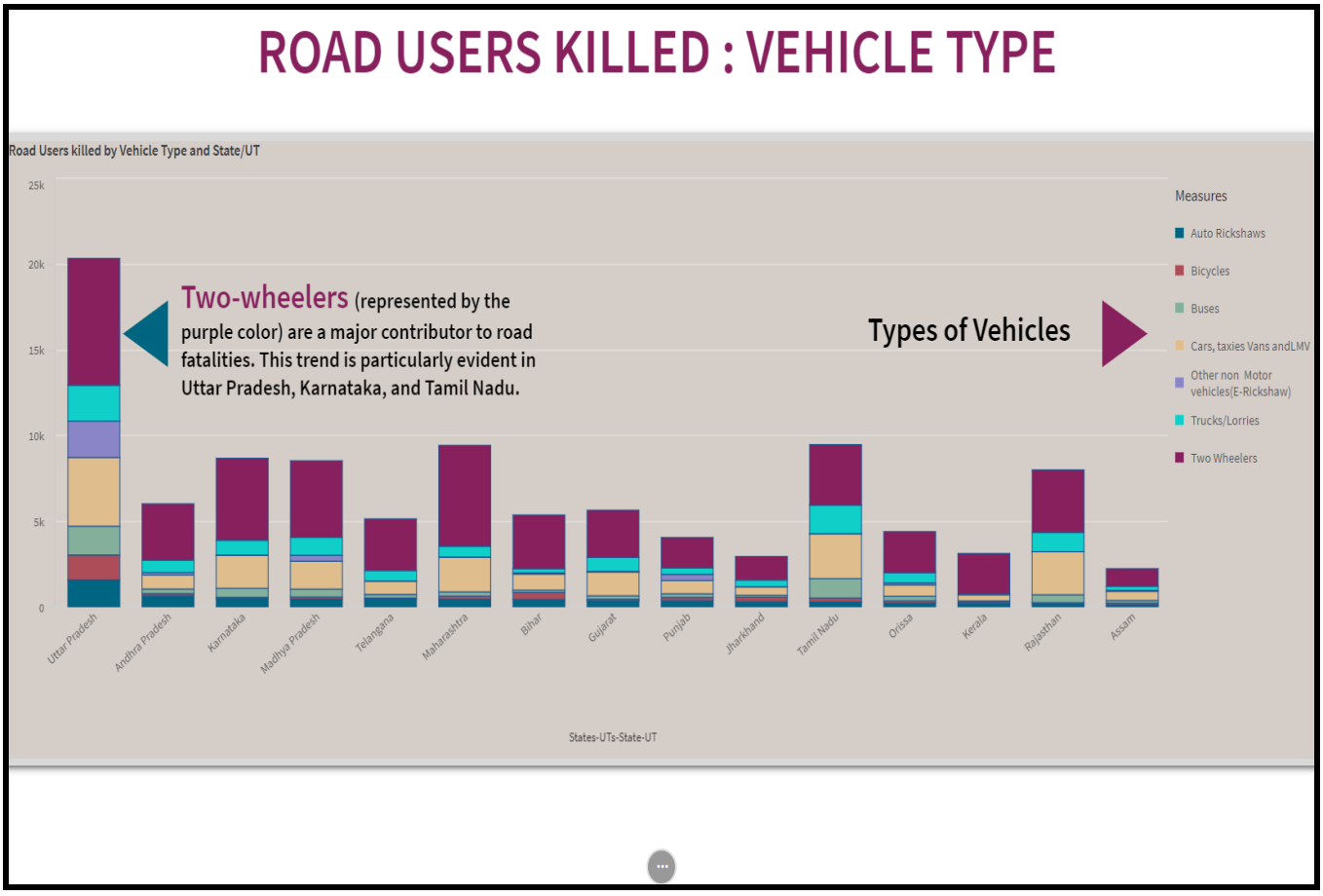
Story Telling 1 :

- This Story shows the number of accidents in different states of India, broken down by weather condition.
- Tamil Nadu has the highest number of accidents, with 57,020 reported.
- The number of accidents are significantly higher during sunny/clear weather compared to rainy weather.
- This suggests that drivers may need to take extra precautions when driving in sunny conditions, as these conditions often lead to increased speeding and accidents.
- The number of accidents are significantly higher during sunny/clear weather compared to rainy weather , highlights the need for better infrastructure and road safety measures to reduce the number of accidents in India.



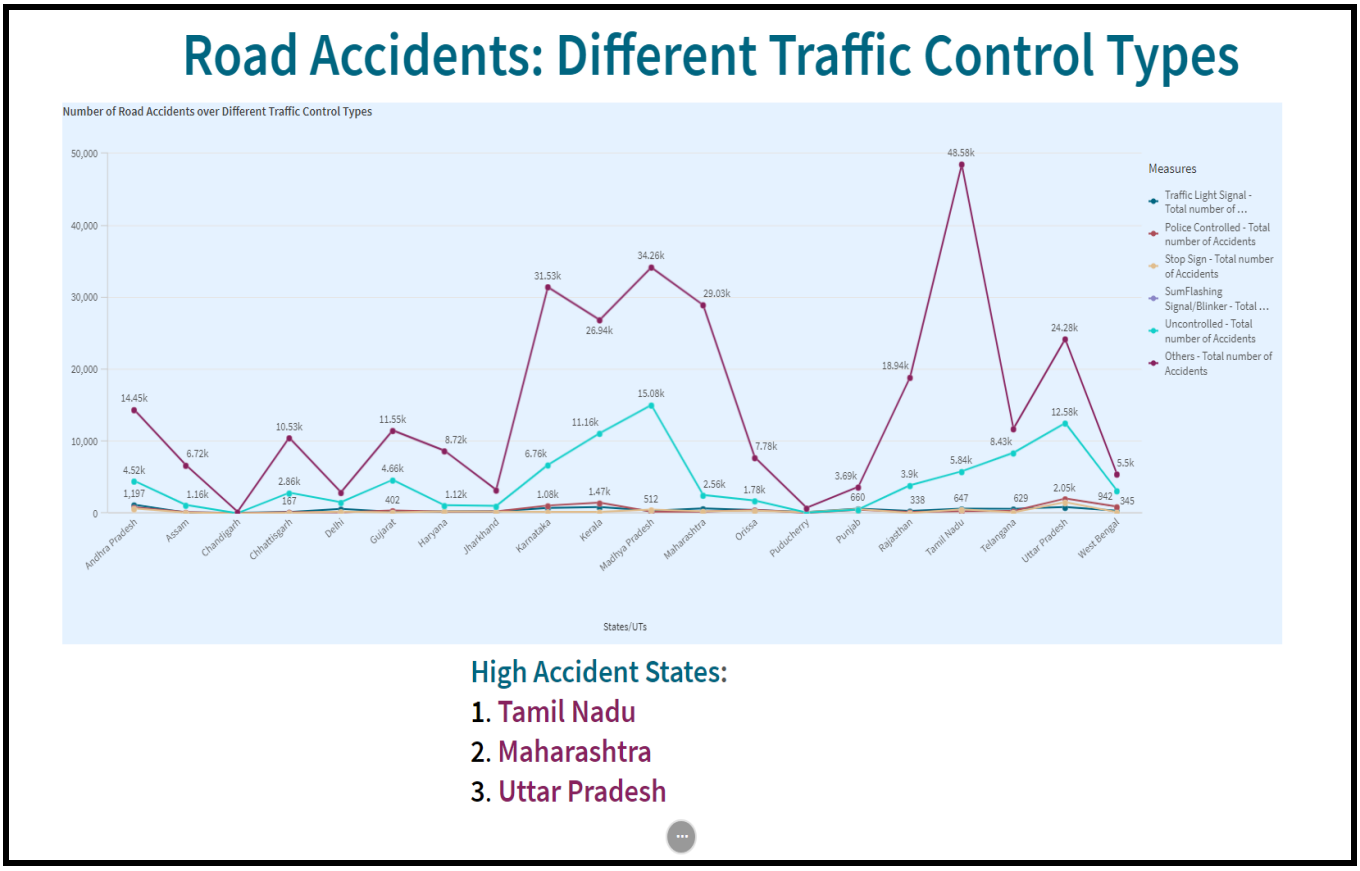
Story Telling 2 :

- This Story illustrates the number of road users killed by different vehicle types across various Indian states and union territories.
- Two-wheelers, depicted in purple, are the most significant contributors to road fatalities, especially in states like Uttar Pradesh, Karnataka, and Tamil Nadu.
- The Story highlights the distribution of fatalities involving auto rickshaws, bicycles, buses, cars/taxis/vans, non-motor vehicles (e-rickshaws), trucks/lorries, and two-wheelers.
- This visual emphasizes the critical need for targeted road safety measures for two-wheeler users to reduce the high fatality rates.



Story Telling 3 :

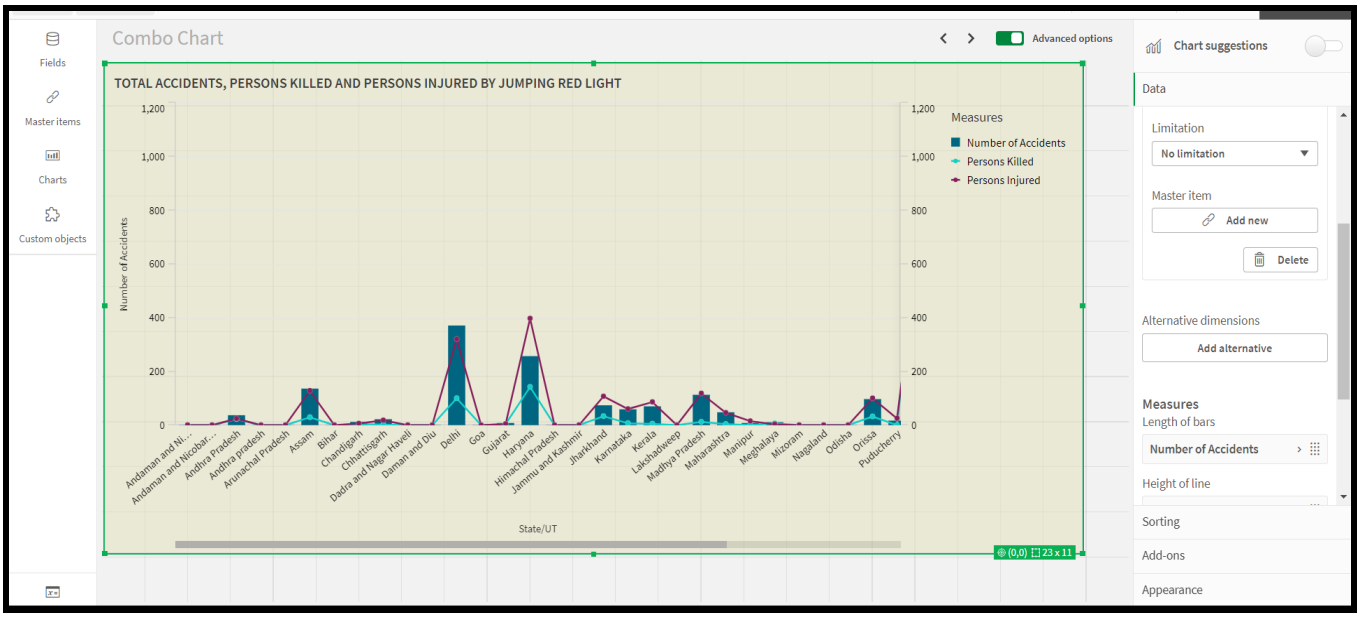
- This story showcases the number of road accidents across various Indian states and union territories, categorized by different traffic control types.
- The data highlights significant accident rates under different control measures such as traffic lights, police control, stop signs, flashing signals, uncontrolled intersections, and others.
- Tamil Nadu, Maharashtra, and Uttar Pradesh emerge as the states with the highest number of road accidents.



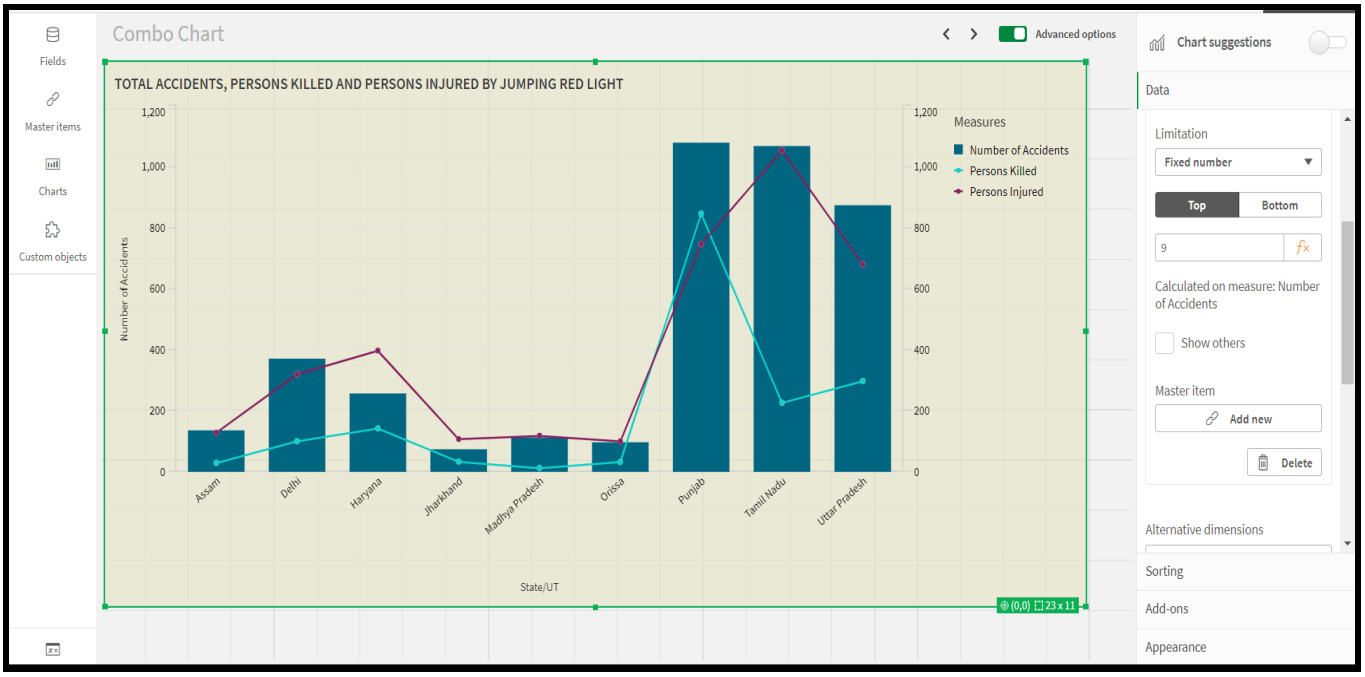
PERFORMANCE TESTING :

Utilization of Data Filters:

Data filters are criteria or conditions applied to datasets to selectively display data that matches those criteria. Here, the criteria is **1. No Limitations**

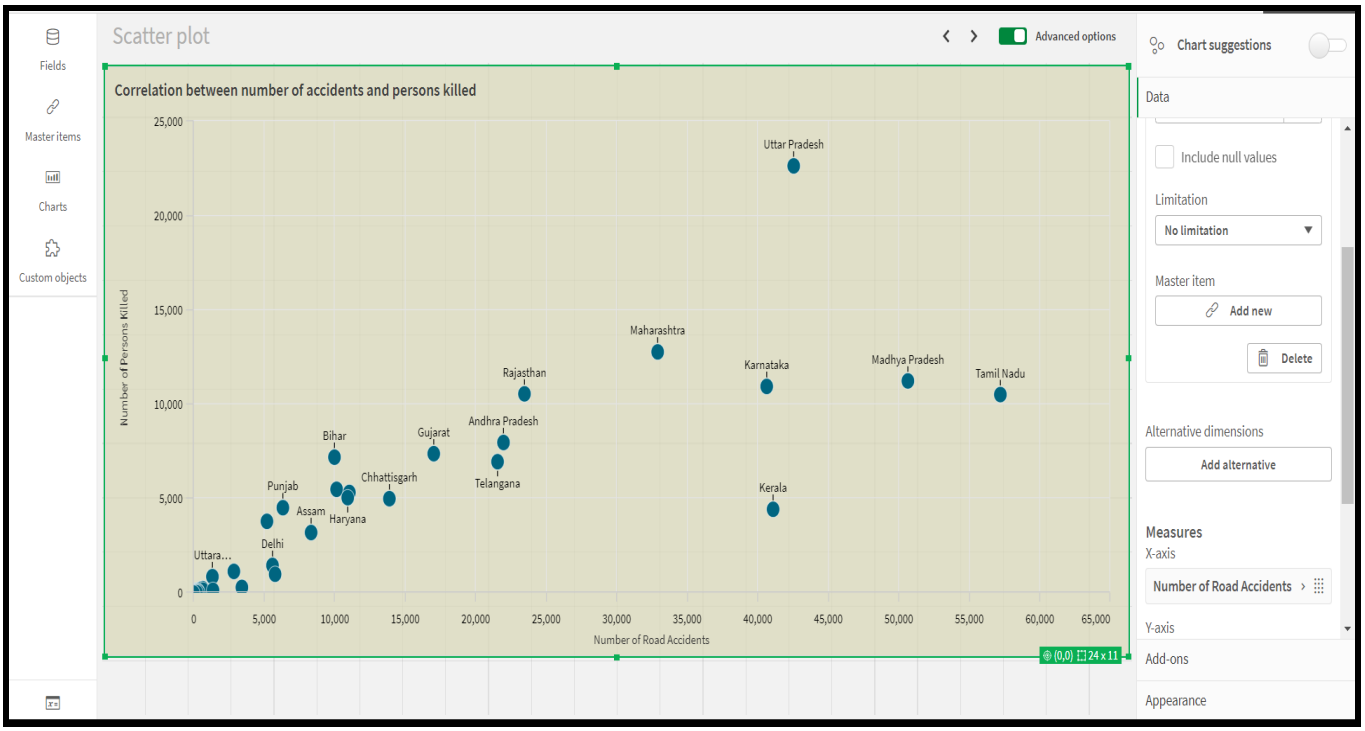


2. Limit is 9

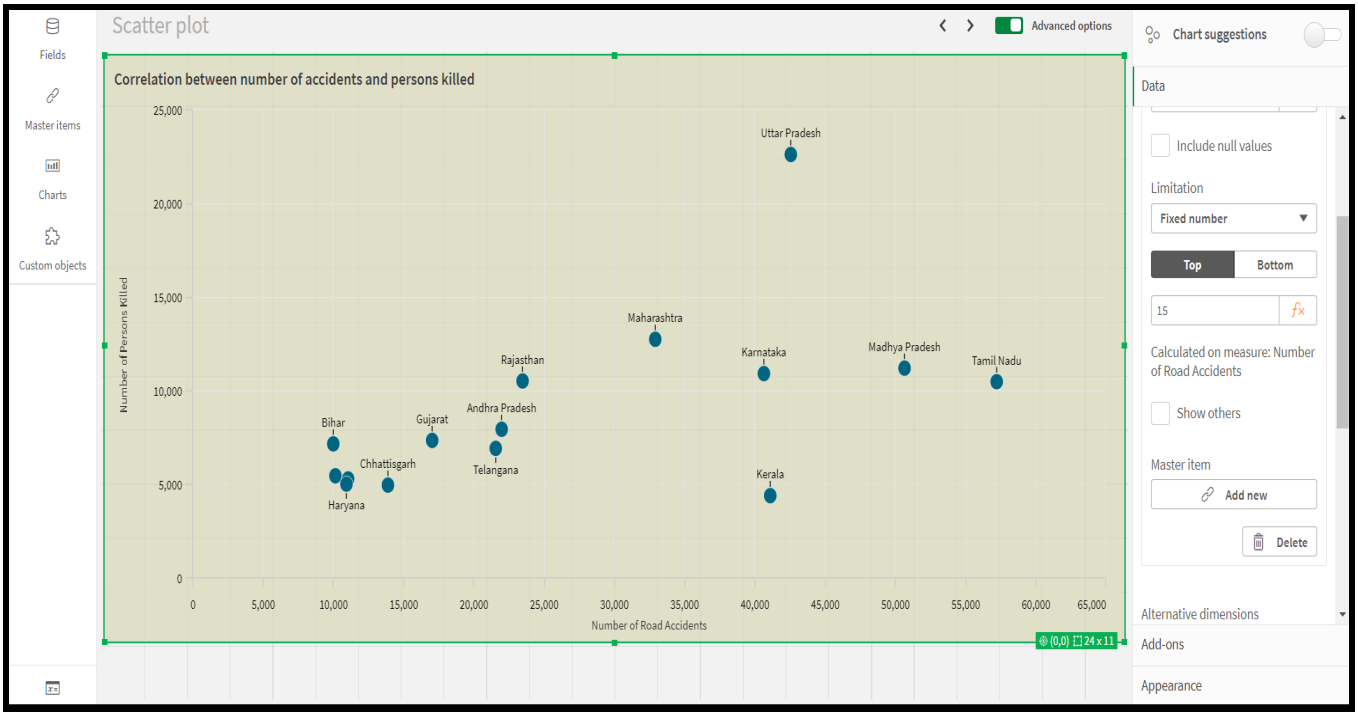


Lets have a look at another visualisation:

1.No limitations

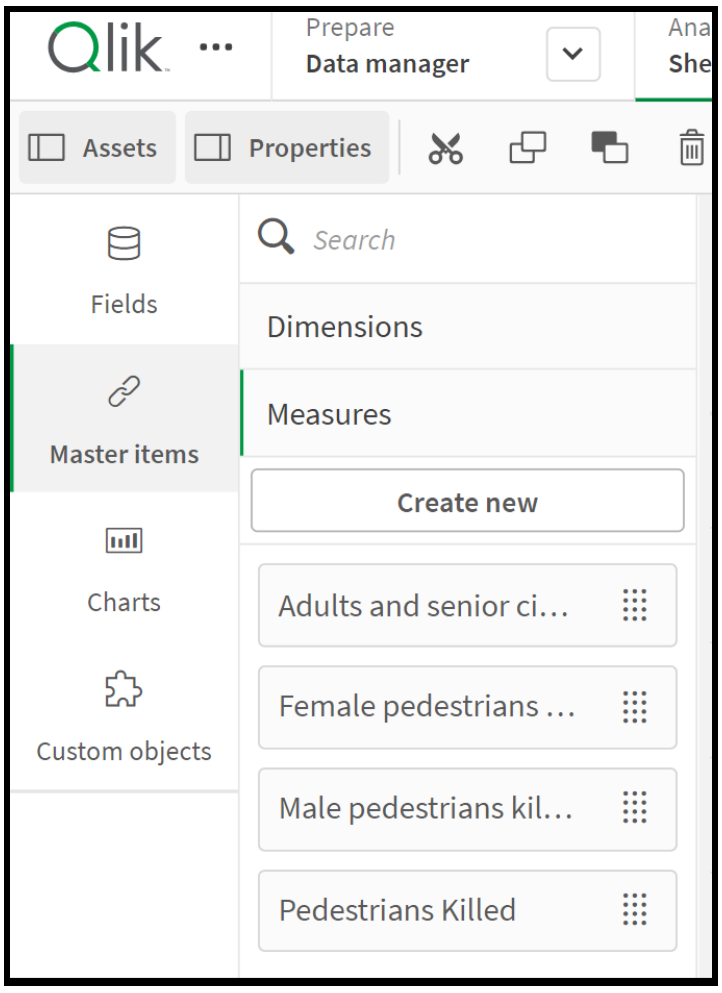


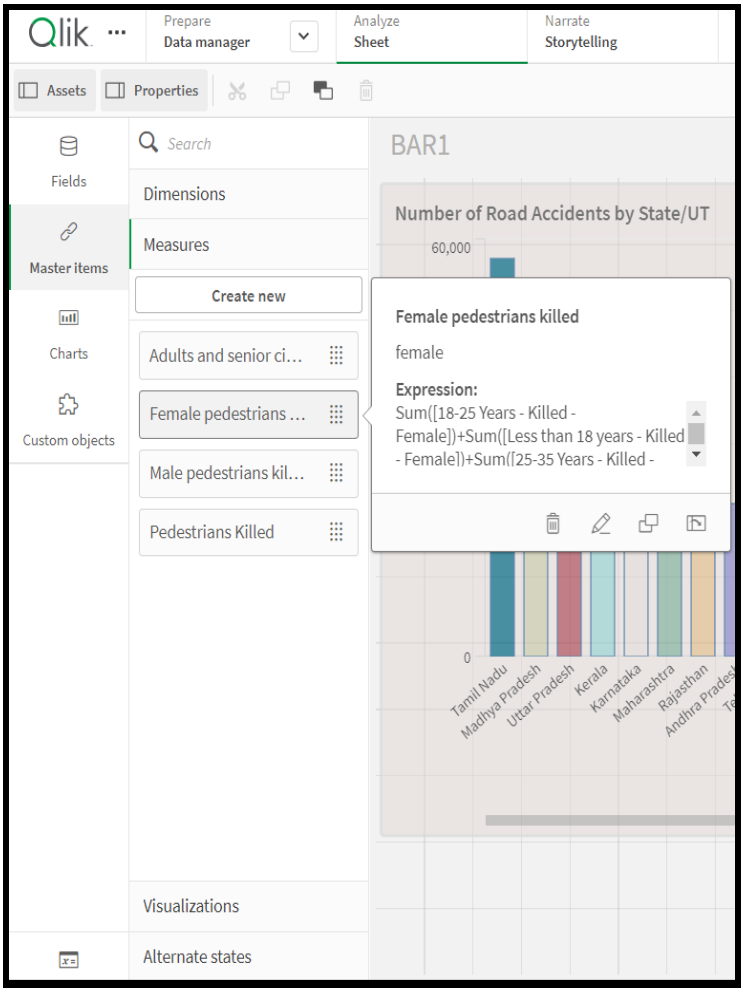
2. Limit is set to 15



UTILIZATION OF MASTER ITEMS:

Master items in Qlik Sense are a way we can create objects once and use them elsewhere in an app (or other apps). These include dimensions, measures and visualizations that keep things in an orderly fashion, increase efficiency, and maintain accuracy in analytics. Goal: They simplify the process of development, bring uniformity in calculations & visualization, and simplify the management and update of the application.





TOTAL VISUALIZATIONS :

- Number of Road Accidents by State/UT
- Road Users killed by Vehicle Type and State/UT
- Number of Road Accidents over Different Traffic Control Types
- Total accidents, persons killed and persons injured by jumping red light
- Total Accidents under Different Weather Conditions
- Correlation between number of accidents and persons killed
- State/UT wise Persons killed and Injured due to drunken driving
- State/UT wise Accidents by Two-Wheelers
- Pedestrian Fatalities - Minors
- Pedestrians Killed by Gender
- Total number of Injuries due to various causes