# INTRODUCTION

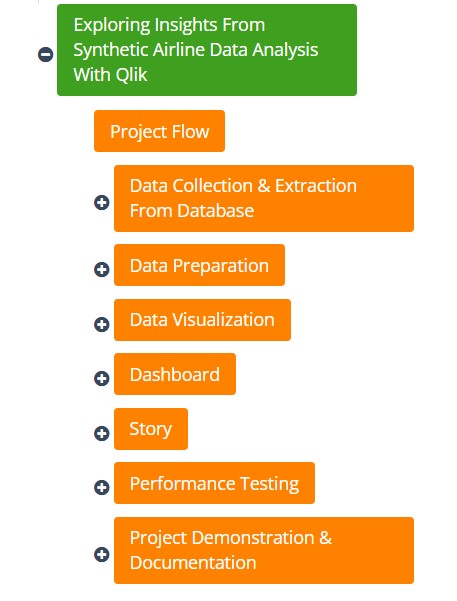
* 1. **Overview:**

Road safety is a pressing concern in India, with a significant number of accidents occurring annually, leading to substantial loss of life and property. The complexity of factors contributing to road accidents—ranging from traffic volume and road conditions to weather patterns and driver behavior—necessitates a comprehensive and data-driven approach to understanding and mitigating these incidents. This project leverages Qlik's advanced data analytics platform to analyze diverse datasets, including traffic data, accident reports, weather conditions, road infrastructure details, and demographic information. By integrating and analyzing these data sources, the project aims to uncover critical insights that can inform strategic decisions and interventions to enhance road safety across India.

* 1. **Purpose:**

The primary purpose of this project is to harness the power of data analytics to address the ongoing challenge of road safety in India. Specifically, the project aims to:

* **Identify Accident Hotspots:** Pinpoint areas and road segments with a high frequency of accidents by analyzing spatial and temporal patterns in the data.
* **Analyze Trends:** Understand the recurring factors and trends that lead to accidents, such as specific types of collisions, seasonal effects, and risky driver behaviors.
* **Develop Predictive Models:** Utilize predictive analytics to forecast potential accident scenarios, enabling proactive measures to prevent accidents before they occur.
  1. **Technical Architecture:**



# Define Problem / Problem Understanding

* 1. **Specify the Business Problem:**

The project addresses the critical issue of road safety in India by analyzing accident patterns and identifying contributing factors. With a high frequency of road accidents leading to significant loss of life and property, there is an urgent need for data-driven solutions to enhance road safety measures.

* 1. **Business Requirements:**

The project will utilize Qlik's data analytics platform to achieve the following objectives:

* **Hotspot Identification:** Pinpoint regions or roads with a high incidence of accidents by correlating traffic data, road conditions, and time of day.
* **Trend Analysis:** Analyze historical accident data to uncover patterns related to accident types, seasonal variations, and driver behavior.
* **Predictive Modeling:** Forecast potential accident scenarios using real-time data inputs like weather conditions and traffic flow patterns, enabling proactive safety interventions.
  1. **Literature Survey:**

The literature review will encompass:

* **Previous Year Accident Data:** India's road safety has been a critical issue for years, with thousands of accidents occurring annually, resulting in significant loss of life and property. According to the National Crime Records Bureau (NCRB), there were over 150,000 fatalities in road accidents in 2021, making India one of the countries with the highest number of road accident deaths globally. The complexity of factors contributing to these accidents—ranging from vehicle types, weather conditions, and traffic controls to driver behavior—highlights the need for comprehensive analysis and intervention strategies.
* **Key Studies:** Several studies have utilized similar datasets to analyze road safety trends and develop predictive models. For example, research has shown that over-speeding and distracted driving are leading causes of accidents, while improved road infrastructure and effective traffic management significantly reduce accident rates. Existing predictive models, such as those incorporating real-time traffic and weather data, have demonstrated the potential to forecast accident-prone areas and times, enabling preemptive actions to prevent accidents.
* **Existing Models:** A review of existing predictive models and analytical tools used in road safety, identifying their strengths and limitations to inform the development of the project’s methodology.

1. **Data Collection** 
   1. **Collect the Dataset:**

The To facilitate a detailed analysis of road safety, the project leverages the "ROAD ACCIDENTS IN INDIA" dataset from Kaggle. This dataset comprises nine Excel worksheets, each focusing on different aspects of road accidents in India during 2019. The data includes:

* **Pedestrians Involved and Killed:** This data is classified by age, gender, and type of impacting vehicles, providing insights into the most vulnerable groups and the types of vehicles most often involved in fatal accidents.
* **Traffic Control Types:** Analyzes accidents based on the presence and type of traffic controls, such as traffic lights, police controls, stop signs, and uncontrolled intersections, helping to assess the effectiveness of different traffic management systems.
* **Weather Conditions:** Examines the impact of weather conditions like sunny, rainy, foggy, and others on the severity and frequency of accidents, revealing how environmental factors contribute to road safety.
* **Causes of Accidents:** Classifies accidents based on causes such as over-speeding, drunken driving, and the use of mobile phones while driving, offering a comprehensive view of the behavioral factors leading to accidents.
* **Severity by Vehicle Type:** Provides detailed information on the number of accidents, fatalities, and injuries by vehicle type, including pedestrians, bicycles, two-wheelers, auto rickshaws, and more.

## Connect Data with Qlik Sense

To analyze the dataset using Qlik Sense, follow these steps:

1. **Extract the Dataset:**

■ After downloading the dataset, extract the files to a specific location on your device.

1. **Create a New Qlik Sense App:**

■ Open Qlik Sense and create a new app named “Road Safety”.

■ Open the newly created app.

1. **Add Data to Qlik Sense:**

■ Click on "Data Manager."

■ Click on "Add data" and select the dataset file from the location where it was extracted.

1. **Data Integration:**

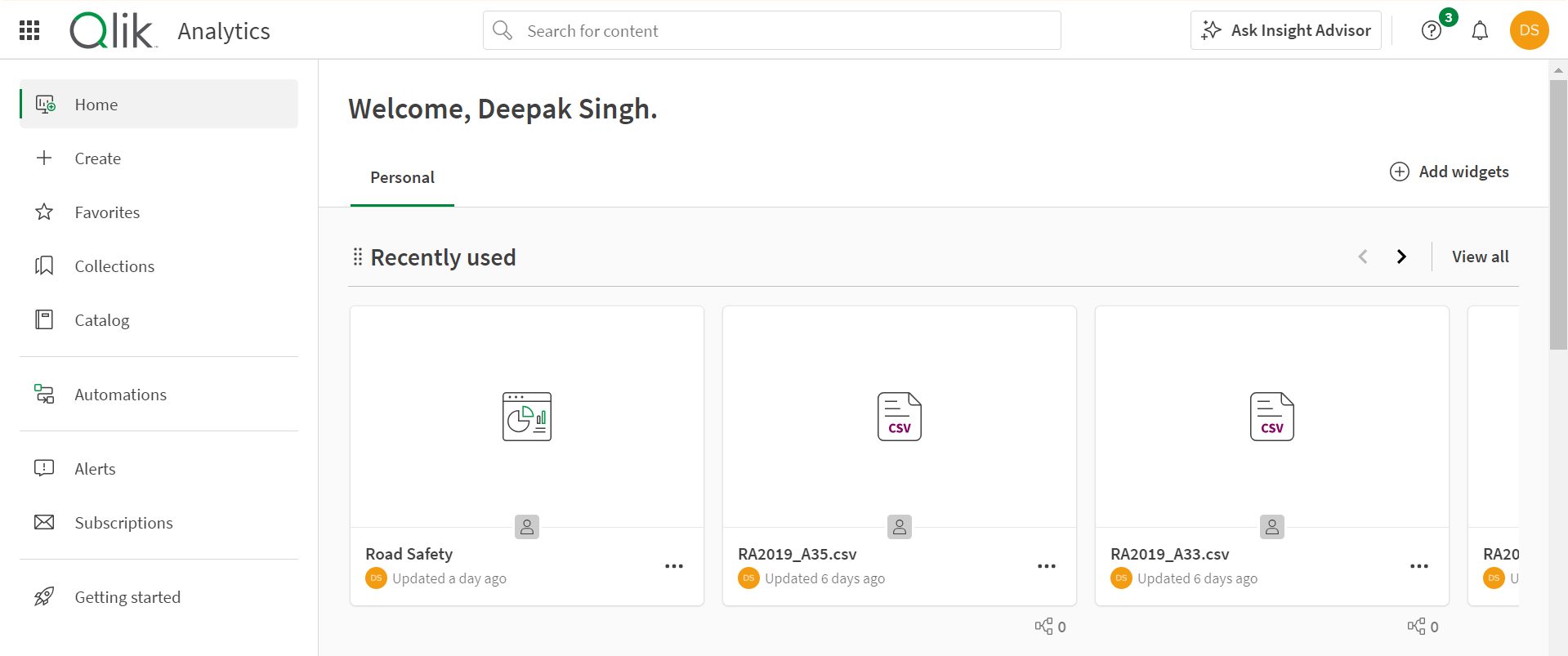
■ Ensure that all relevant fields from the dataset are correctly mapped in Qlik Sense.

■ Check for any inconsistencies or missing values in the dataset and clean the data if necessary.

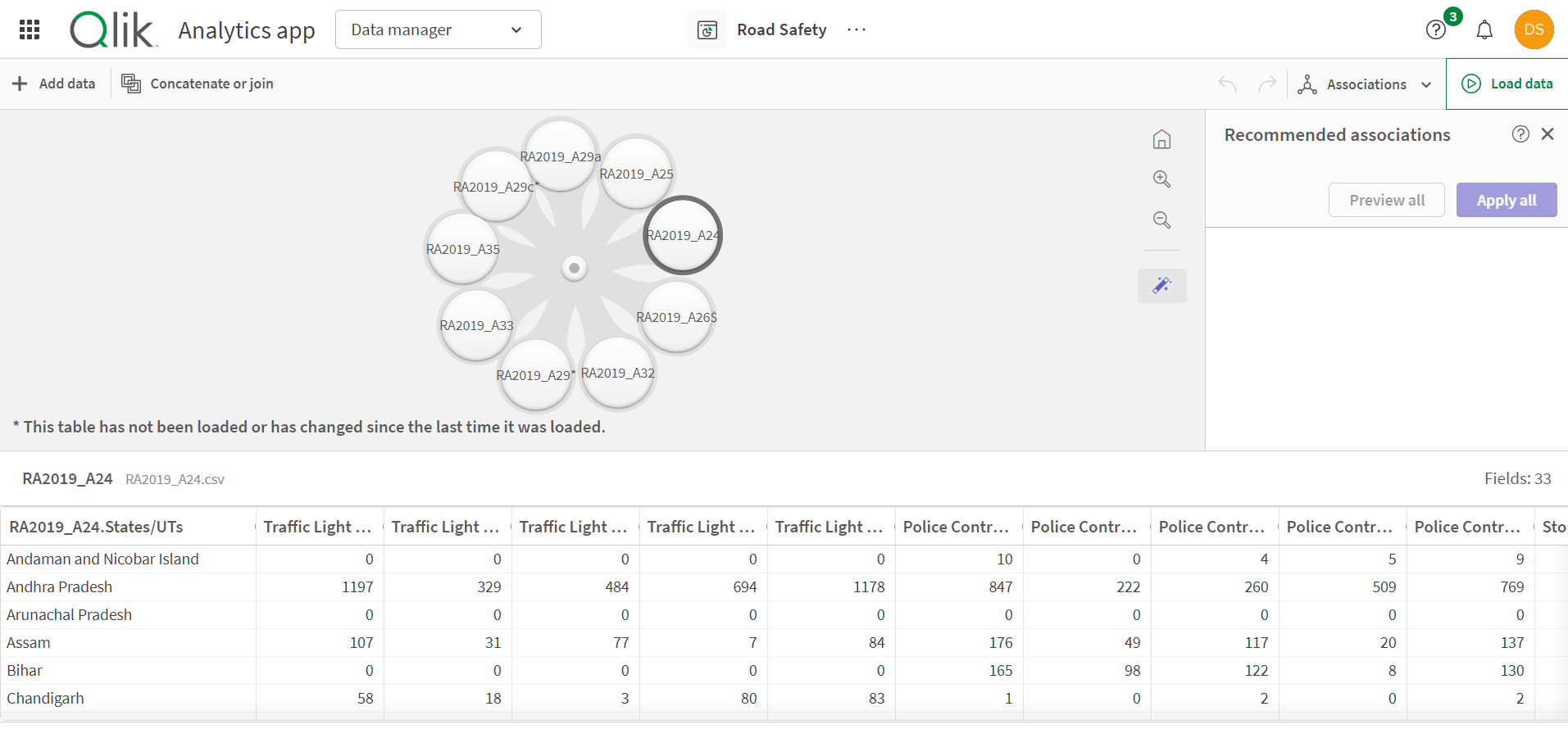
1. **Data Mapping:**

■ Map fields such as State/UT to identify the geographical region or state in India where the accident data is recorded, ensuring it is correctly recognized by Qlik Sense for spatial analysis. Additionally, map age and gender categories, including fields like "Less than 18 years – Male," "Less than 18 years – Female," and "18-25 Years – Male," to categorize pedestrian involvement and fatalities accurately. For vehicle-related analysis, map fields such as "Bicycles," "Two Wheelers," "Auto Rickshaws," "Cars, Taxis, Vans and LMV," "Trucks/Lorries," and "Buses" to examine the types of vehicles involved in accidents, particularly those resulting in fatalities. Traffic control types, including "Traffic Light Signal," "Police Controlled," "Stop Sign," "Flashing Signal/Blinker," and "Uncontrolled," should also be mapped to classify accidents based on traffic control measures, aiding in the evaluation of their effectiveness. Weather condition fields like "Sunny/Clear," "Rainy," "Foggy and Misty," and "Hail/Sleet" should be mapped to understand the influence of different weather conditions on accident frequency and severity. Furthermore, map causes of accidents fields, such as "Over-Speeding," "Drunken Driving/Consumption of Alcohol and Drugs," "Driving on Wrong Side," "Jumping Red Light," and "Use of Mobile Phone," to identify and analyze the root causes of road accidents. Severity by vehicle type should also be mapped, with fields like "Pedestrian - Number of Road Accidents," "Bicycles - Number of Road Accidents," and "Two Wheelers - Number of Road Accidents" to assess the severity of accidents based on the vehicle type involved. Lastly, map accident outcomes fields, including "Persons Killed," "Persons Injured - Grievously Injured," and "Persons Injured - Minor Injury," to facilitate the analysis of the human impact of road accidents.

By following these steps, you can successfully integrate and prepare your synthetic airline dataset in Qlik Sense for comprehensive analysis and visualization.



* + Open Data Manager and we need to click on add data on left side top corner to add the data.



1. **Data Preparation** 
   1. **Prepare the Data for Visualization**:

**Clean the Data**

* + - **Remove Inconsistencies:**

■ After downloading the dataset and converting it from CSV to Excel format, inspect the data for inconsistencies and anomalies.

■ Rectify any discrepancies in the data entries to ensure uniformity.

* + - **Handle Missing Values:**

■ Identify and address any missing values in the dataset.

■ Fill in missing data points with appropriate values or remove records with substantial missing information.

**Transform the Data**

* + - **Format for Analysis:**

■ Ensure the data is in a suitable format for analysis and visualization.

■ Check that dates, times, and numerical values are correctly formatted.

■ Ensure all fields are appropriately labeled.

* 1. **Aggregate and Categorize Data** 
     + **Remove Extra Columns:**

■ Identify and eliminate unnecessary columns such as 'others' and

'average' that are not relevant to the project's analysis.

■ During the data addition process in Qlik Sense, select only the columns

required for analysis and discard extraneous ones.

* + - **Remove 'Total' Rows:**

■ Identify rows that contain "total" values, which are direct additions of each column.

■ Remove these rows from the dataset to ensure that aggregate data does not skew the analysis. In my case I have applied filter on it add excluded the total column from each visualization.

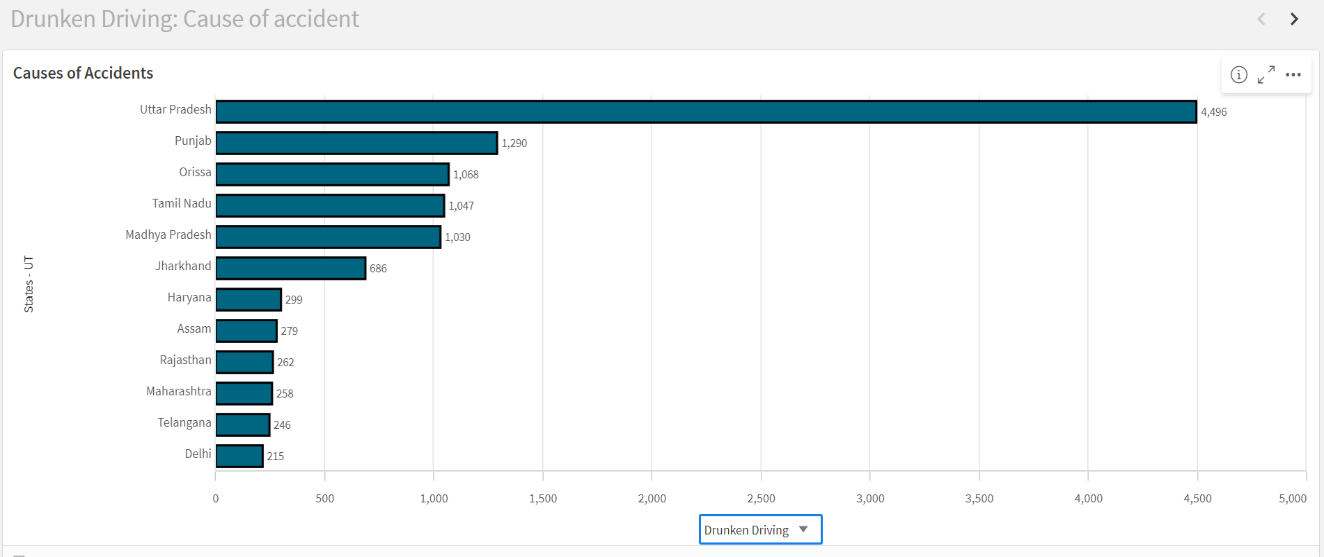
* + - **Re-upload Cleaned Data:**

■ After cleaning the dataset by removing unwanted columns and rows, re-upload the cleaned files to Qlik Sense.

* 1. **Data Association** 
     + **Qlik Sense Recommendations:**

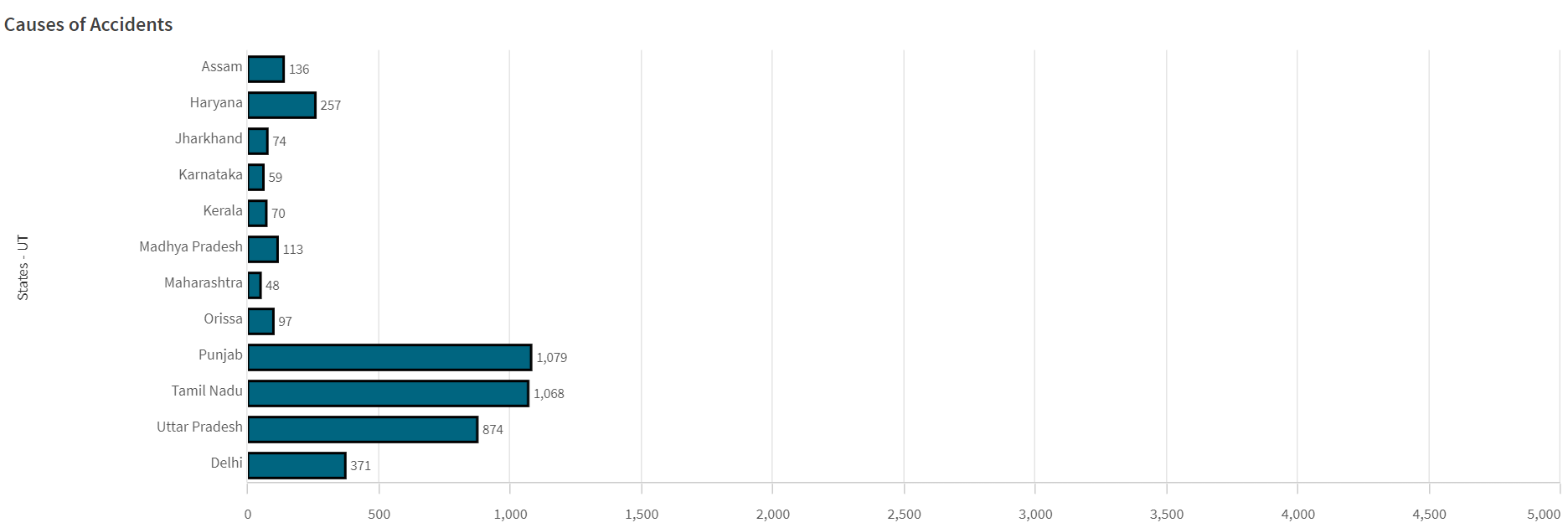
■ Utilize Qlik Sense's recommendations for data associations to link related data fields across different tables.

■ Ensure that the data is properly connected and ready for comprehensive analysis.

1. **Data Visualizations:** 
   1.  **Bar Chart:**

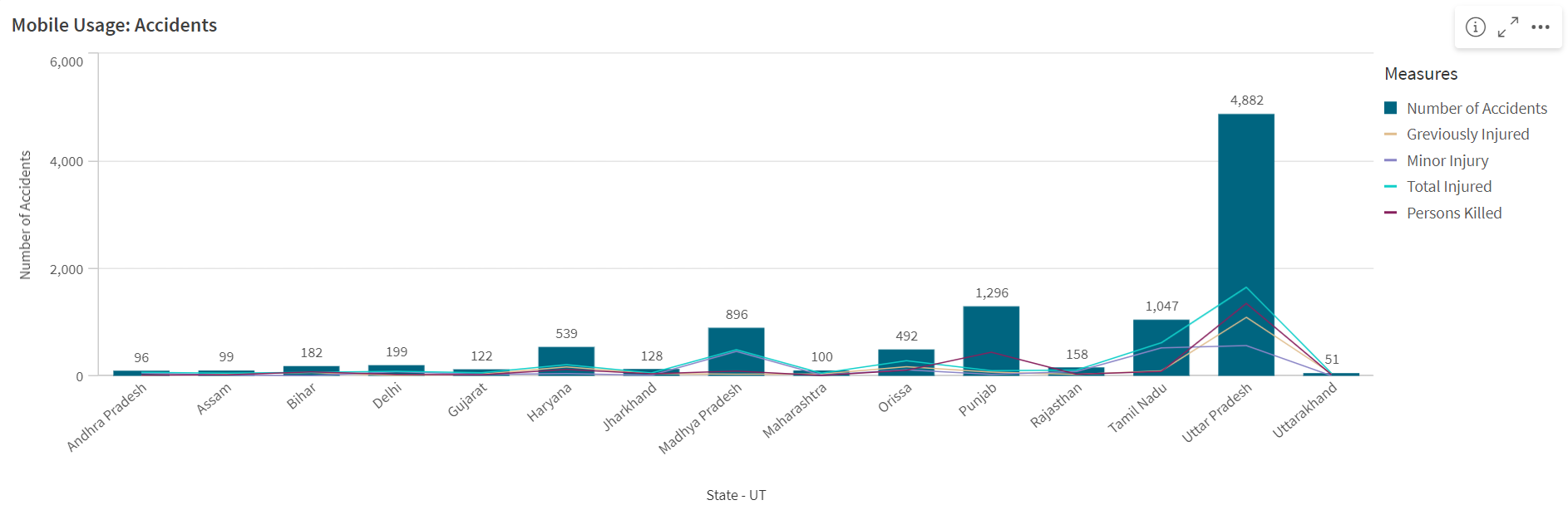
# Accidents due to Drunken Driving

* + The bar chart visualizes the frequency of accidents caused by drunken driving across various regions or states. The height of each bar represents the number of accidents, providing a clear comparison between different areas.
  + **Insight**: The insight drawn from this chart indicates that certain regions have significantly higher incidents of drunken driving, suggesting a need for targeted awareness campaigns and stricter enforcement of driving regulations in these high-risk areas.



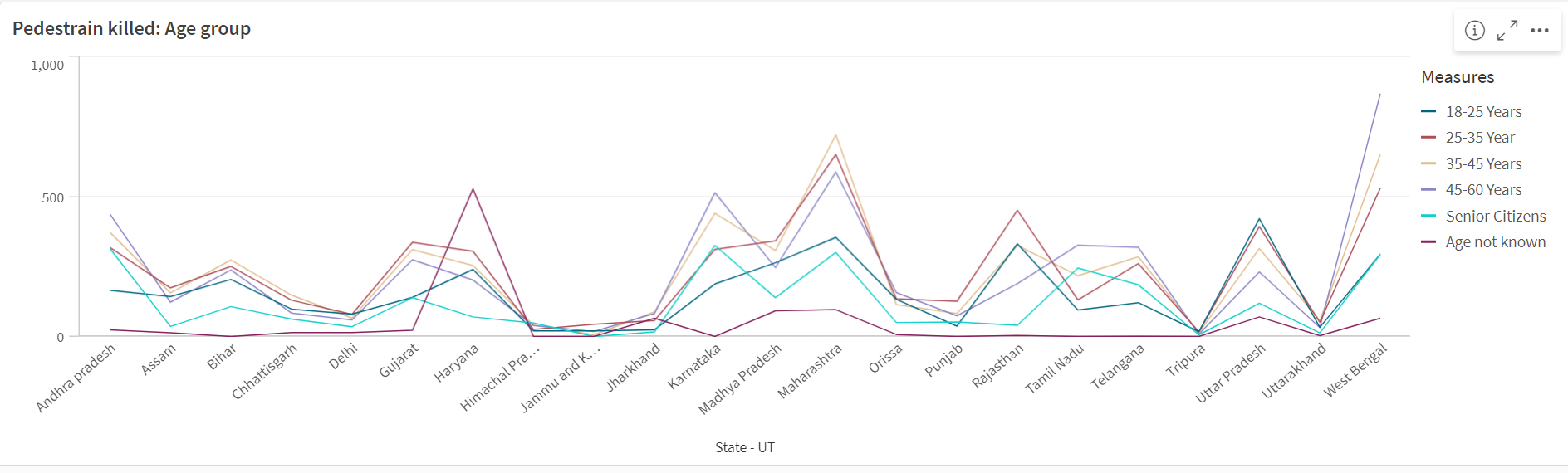
# Accidents due to Jumping Red light

* + The horizontal bar chart visualizes the count of accidents caused due to Jumping Reg light across various regions or states. The height of each bar represents the number of accidents, providing a clear comparison between different areas.
  + **Insight**: The insight drawn from this chart indicates that certain regions have significantly higher incidents of red light, suggesting a need for targeted awareness campaigns and stricter enforcement of driving regulations in these high-risk areas.



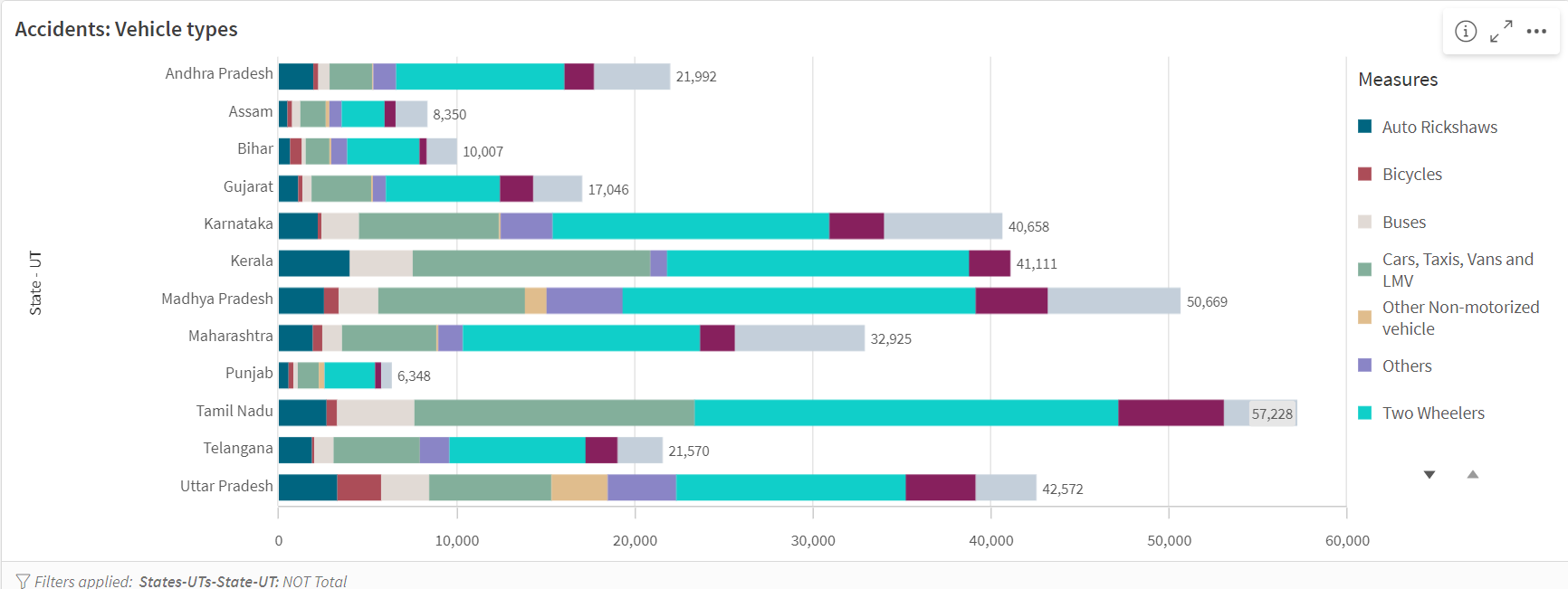
# Accidents during Mobile usage

* + This combo bar chart highlights the number of accidents associated with mobile phone usage while driving. The distribution across different states or regions shows how prevalent this risky behavior is and helps identify areas where mobile usage while driving is particularly problematic.
  + The bar shows the number of Accidents and the line shows that what type of injuries they had.
  + **Insight**: The Insights suggest that regions with higher accidents due to mobile usage might benefit from campaigns focusing on the dangers of distracted driving and increased penalties for mobile usage behind the wheel.
  1. **Line Chart**



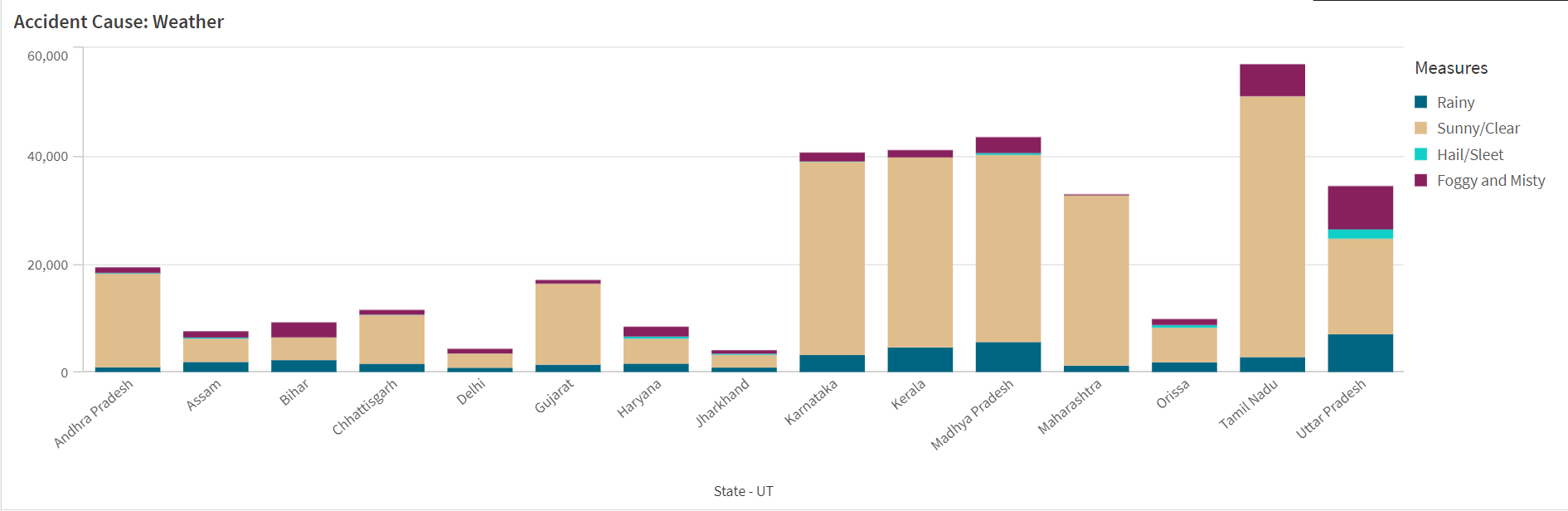
# Pedestrians Killed by Age Group

* + This line chart tracks the number of pedestrian fatalities across different age groups. The slope of the line indicates which age groups are more vulnerable to road accidents.
  + **Insight**: The Insights from this chart reveal that certain age groups, such as the elderly or young adults, might be more susceptible to fatal accidents, highlighting the need for age-specific safety measures, such as better pedestrian crossings or targeted public safety messages.
  1. **Stacked Bar** **Chart**



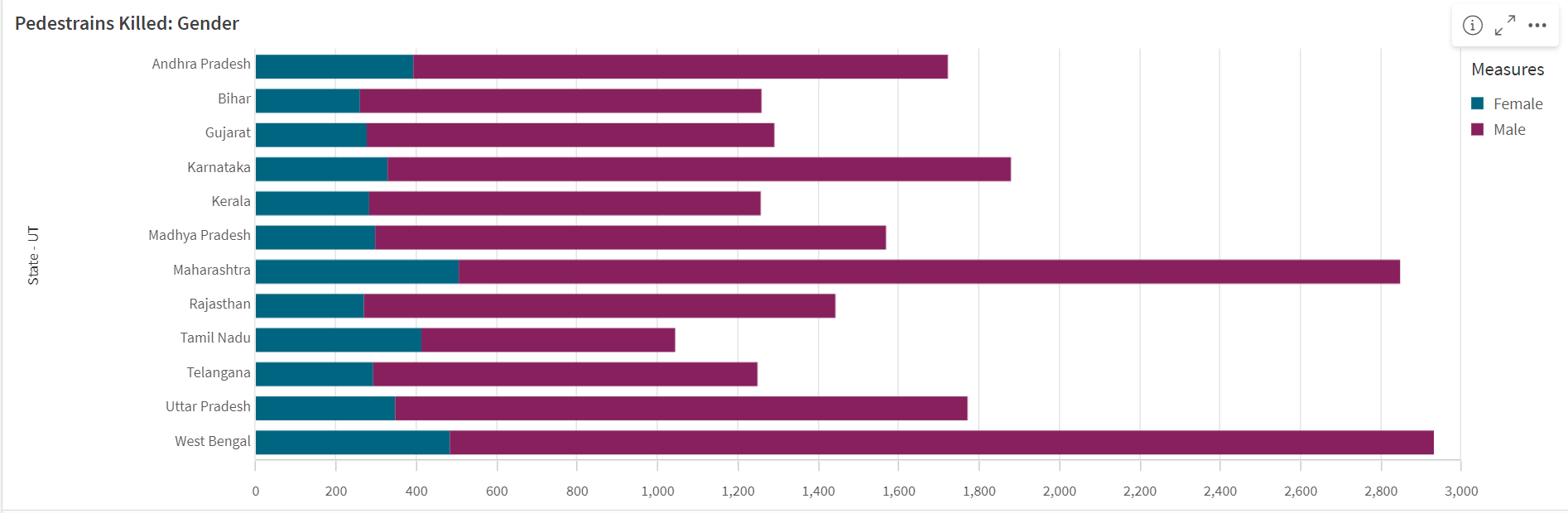
# Accidents by Vehicle Types

* + This stacked bar chart shows the proportion of accidents involving different types of vehicles, such as two-wheelers, cars, and trucks. Each segment within the bar represents a different vehicle type, providing a visual representation of the distribution of accidents across vehicle categories.
  + **Insight**: The Insights suggest that certain vehicle types, like two-wheelers, may be involved in a disproportionately high number of accidents, indicating a need for enhanced safety measures for these vehicles, such as better protective gear for riders or more stringent traffic regulations.



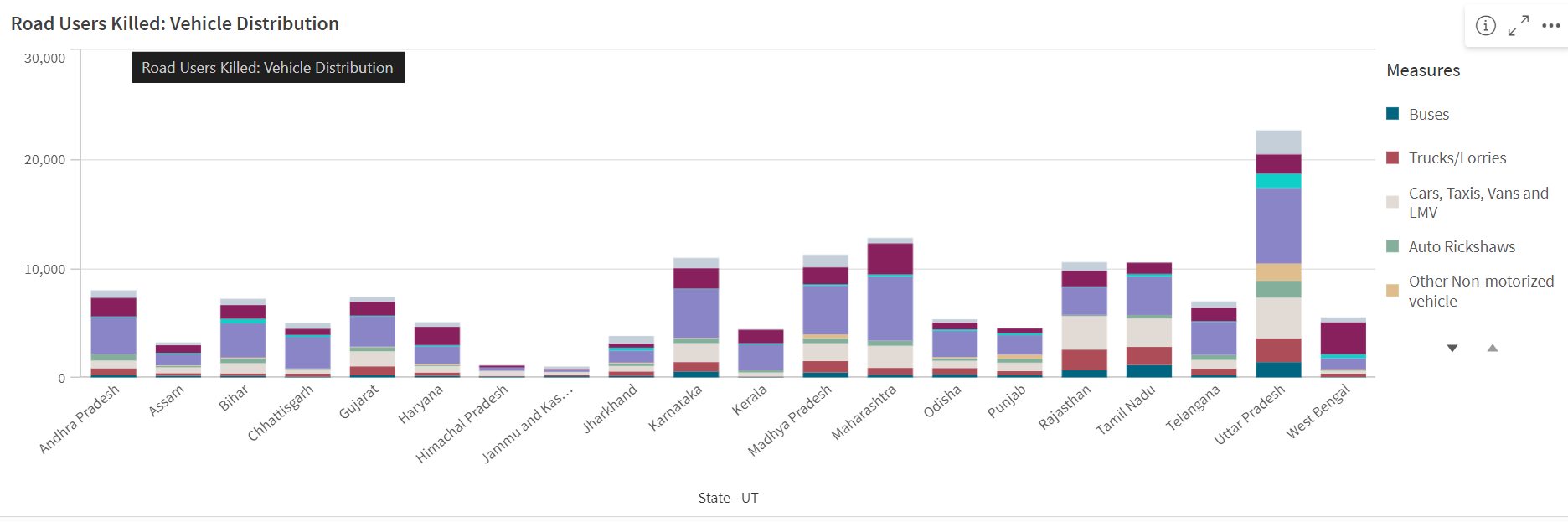
# Accidents Based on Weather Conditions

* + The stacked bar chart compares the number of accidents under various weather conditions, such as clear, rainy, and foggy days. Each segment of the bar corresponds to a specific weather condition, allowing for a comparison of how weather impacts road safety.
  + **Insight**: The Insights drawn from this chart indicate that adverse weather conditions like rain or fog significantly contribute to the occurrence of accidents, suggesting the importance of improving road infrastructure, such as better drainage systems and clearer road markings, to mitigate these risks during bad weather.



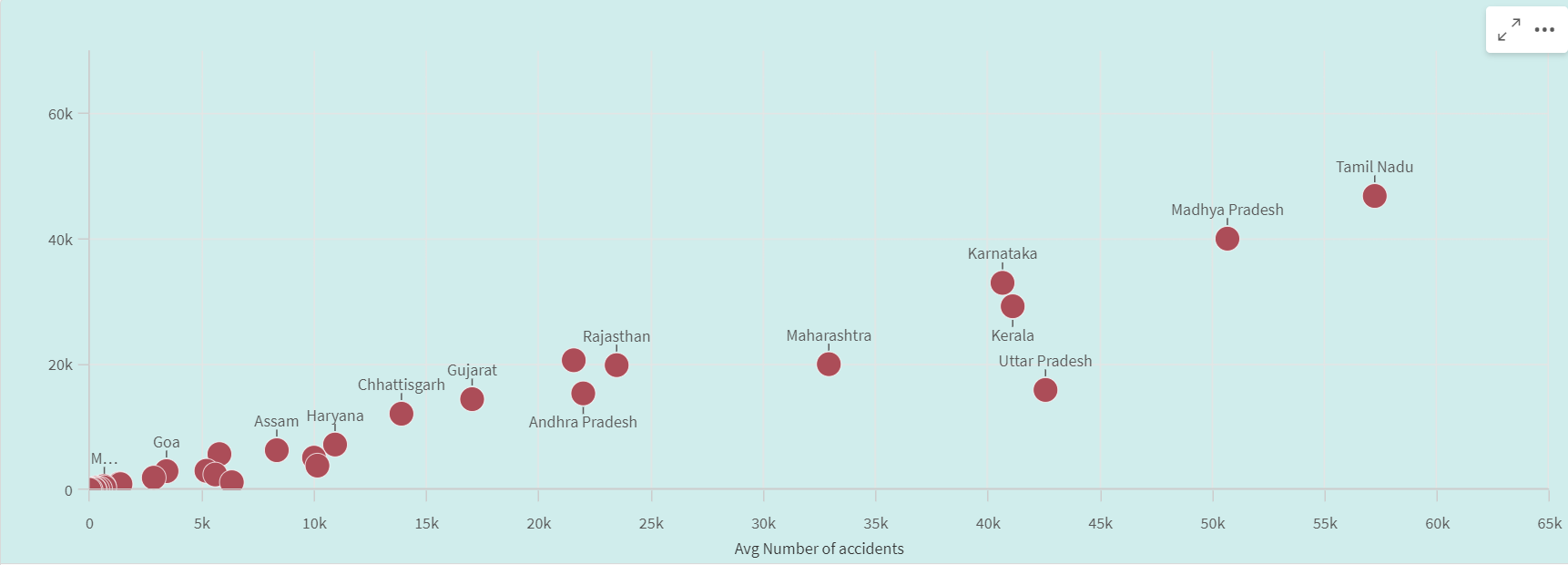
# Overall Pedestrian Fatalities by Gender

* + This visualization provides an overall view of pedestrian fatalities categorized by gender. It reveals the disparity between male and female pedestrian deaths, offering insights into which gender is more vulnerable in road accidents.
  + **Insight**: The insights suggest that targeted interventions might be required to address the specific safety needs of the more vulnerable gender group, such as enhancing street lighting and improving pedestrian facilities in areas where these fatalities are high.



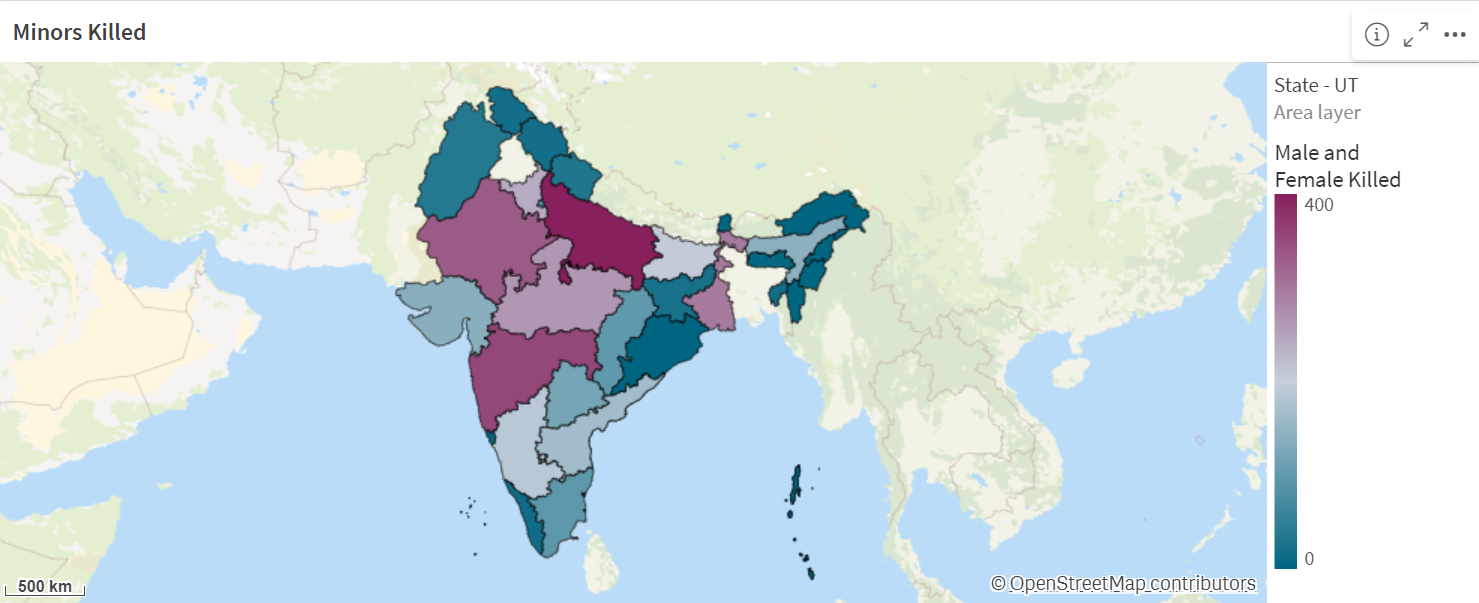
# Accidents Due to Vehicle Distribution

* + This stacked bar chart visualizes the distribution of accidents based on different types of vehicles involved. Each segment within the bar represents a specific vehicle category, such as two-wheelers, cars, or trucks.
  + **Insight**: The insights from this chart reveal which types of vehicles are most frequently involved in accidents, helping to identify high-risk vehicle categories. This information can guide safety interventions, such as targeted driver education programs or enhanced safety regulations for specific vehicle types.
  1. **Scatter Plot**



# Relationship Between Average Number of Accidents in All States

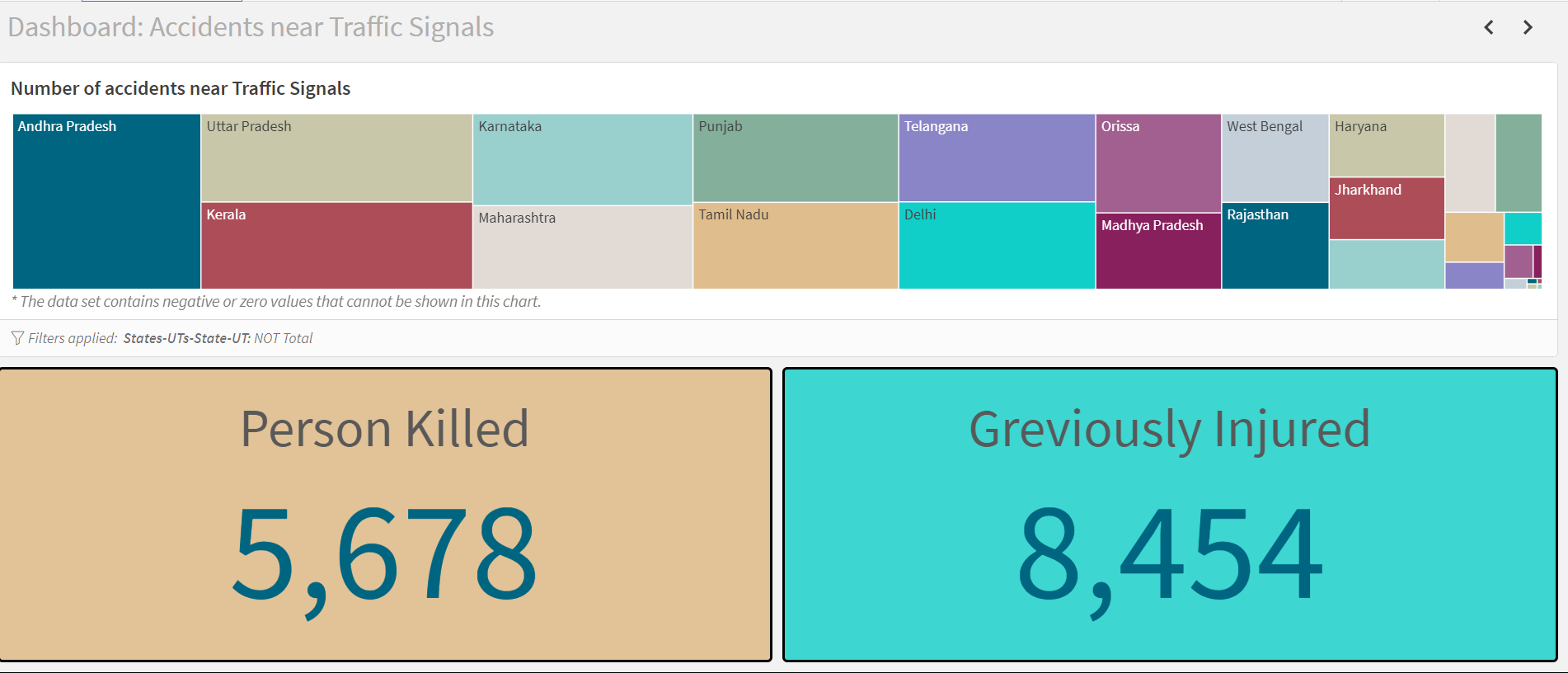
* + The scatter plot demonstrates a positive relationship between the average number of accidents across all states. Each point on the plot represents a state, with its position indicating the frequency of accidents.
  + **Insight**: The Insights from this visualization suggest that states with higher average accident rates may share common factors contributing to road accidents. This understanding can prompt further investigation into shared characteristics, such as infrastructure quality or traffic enforcement, to devise strategies that can be applied across multiple states.
  1. **Map Visualization**



# Accidents Due to Vehicle Distribution

* + This map visualization highlights the states with the highest number of male and female fatalities in road accidents. The color intensity or markers on the map indicate the severity and location of these fatalities.
  + **Insight**: The insights drawn from this map reveal regional patterns in gender-based fatalities, suggesting where gender-specific safety interventions might be necessary. For example, if certain states show a higher number of male fatalities, it could indicate risky driving behaviors prevalent among male drivers that need to be addressed.

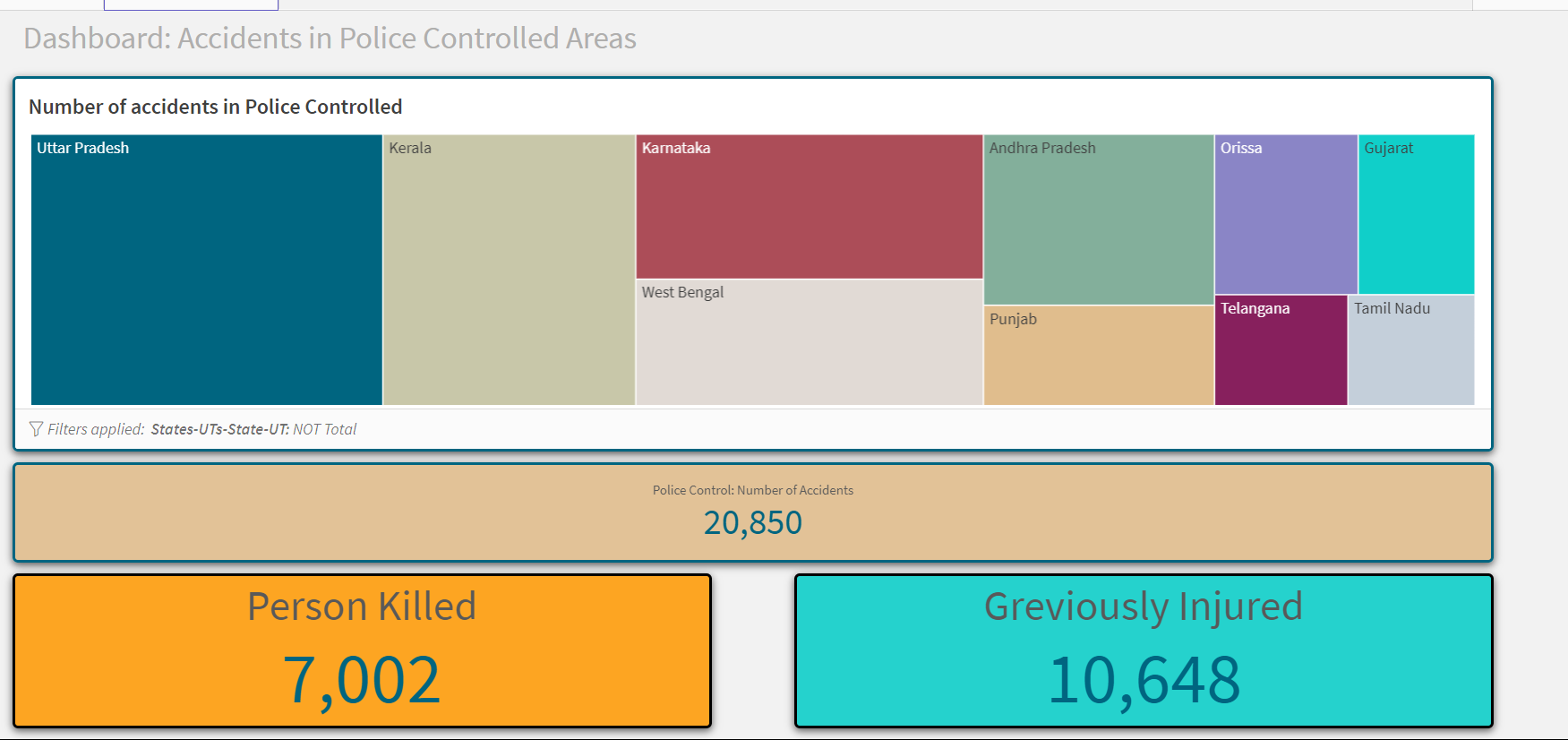
1. **Dashboard:**



**Tree Map and KPIs for Accident and Fatality Analysis**

This dashboard utilizes a tree map to display the number of accidents across all states, with KPIs showing the total number of persons killed and grievously injured. The hierarchical structure of the tree chart allows for a detailed breakdown of accident data by state, while the Key Performance Indicator (KPIs) provide a quick overview of the most critical metrics. KPI show the number of persons killed that is 5.6k approximately and another KPI show the Injured person which are 8.4k approximately.

Insights from this dashboard suggest which states have the highest accident rates and the most severe outcomes, guiding resource allocation for emergency response and road safety improvements.



**Tree Map and KPIs for Police-Controlled Area Accidents.**   
This dashboard focuses on accidents occurring in police-controlled areas, using a tree chart to show the number of accidents by state. The KPIs highlight the total number of accidents, persons killed, and grievously injured in these areas.

Insights from this dashboard can help assess the effectiveness of police-controlled traffic management and identify states where additional police resources or improved traffic control measures might be needed to reduce accidents and fatalities.

1. **Report**
   1. **Key Findings**
      1. **Number of Accidents**

The analysis reveals a significant number of road accidents occurring across various states in India, with notable variations in the frequency and severity of accidents based on vehicle types, weather conditions, and traffic management practices. The data indicates that certain states have a disproportionately high number of accidents, which could be attributed to factors such as infrastructure quality, traffic volume, and enforcement of road safety regulations.

* + 1. **Number of Accidents by Age and Gender**

The data also highlights distinct patterns in the distribution of accidents by age and gender. Younger individuals, particularly those aged 18-35, are more frequently involved in accidents, reflecting the higher risk-taking behavior and greater mobility of this demographic group. The analysis shows that male drivers are significantly more involved in accidents compared to female drivers, which aligns with global trends of higher accident rates among males due to factors like driving frequency, risk exposure, and driving behavior.

* 1. **Recommendations and Actionable Insights**
  2. **Enhance Road Safety Measures in High-Risk States.**

Based on the analysis, certain states exhibit a significantly higher number of road accidents. It is recommended that these states receive prioritized attention for road safety improvements.

Actionable steps include:

* + - * **Deploying additional traffic monitoring systems**: Implement more speed cameras and red-light cameras in high-risk areas to deter dangerous driving behaviors.
      * **Improving road infrastructure**: Invest in better road design, including wider lanes, better lighting, and clearer road markings, particularly in accident-prone areas.
  1. **Targeted Safety Campaigns for High-Risk Age Groups**  
     The data shows that younger individuals, particularly those aged 18-35, are more frequently involved in accidents. To address this, it is recommended to launch targeted safety campaigns focused on this demographic.

Actionable steps include:

* + - * **Awareness campaigns**: Create educational campaigns highlighting the dangers of speeding, distracted driving, and driving under the influence, tailored specifically to younger drivers.
      * **Driver training programs**: Implement advanced driver training programs that emphasize defensive driving techniques and safe driving practices for younger drivers.
  1. **Gender-Specific Interventions**  
     Given the higher involvement of male drivers in accidents, gender-specific interventions should be considered.

Actionable steps include:

* **Promoting safe driving among male drivers**: Launch initiatives that encourage safe driving behaviors among men, such as driving within speed limits and avoiding aggressive driving.
* **Incorporating gender perspectives in road safety policies**: Ensure that road safety policies consider the different risk factors for male and female drivers and pedestrians.
  1. **Strengthen Traffic Control Measures**  
     Accidents occurring in police-controlled areas suggest the need for enhanced traffic control measures.

Actionable steps include:

* **Increase police presence during peak hours**: Deploy more traffic police in areas and at times with high accident rates to manage traffic and enforce rules.
* **Evaluate and optimize traffic control devices**: Regularly assess the effectiveness of traffic lights, stop signs, and other control devices, making necessary adjustments to improve safety.
  1. **Prepare for Adverse Weather Conditions**  
     The impact of weather on accident rates indicates a need for preparedness in adverse conditions.

Actionable steps include:

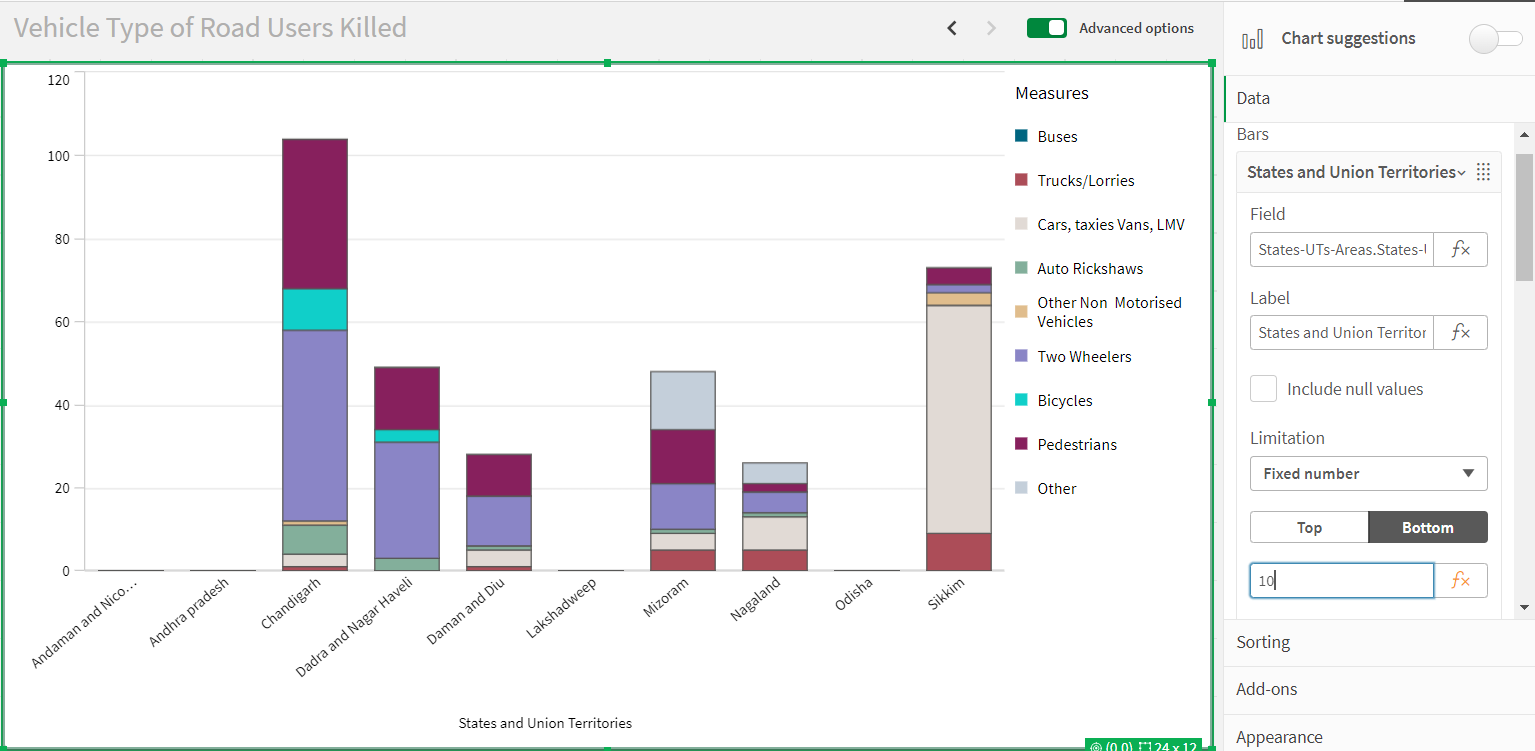
* **Improve road conditions**: Ensure roads are well-maintained to handle adverse weather, with proper drainage systems and skid-resistant surfaces.
* **Weather-related public alerts**: Implement systems to notify drivers of adverse weather conditions, advising them to exercise caution or avoid travel if possible.
  1. **Implement Predictive Analytics for Proactive Safety Measures**  
     Leverage predictive models to forecast accident-prone periods and locations.

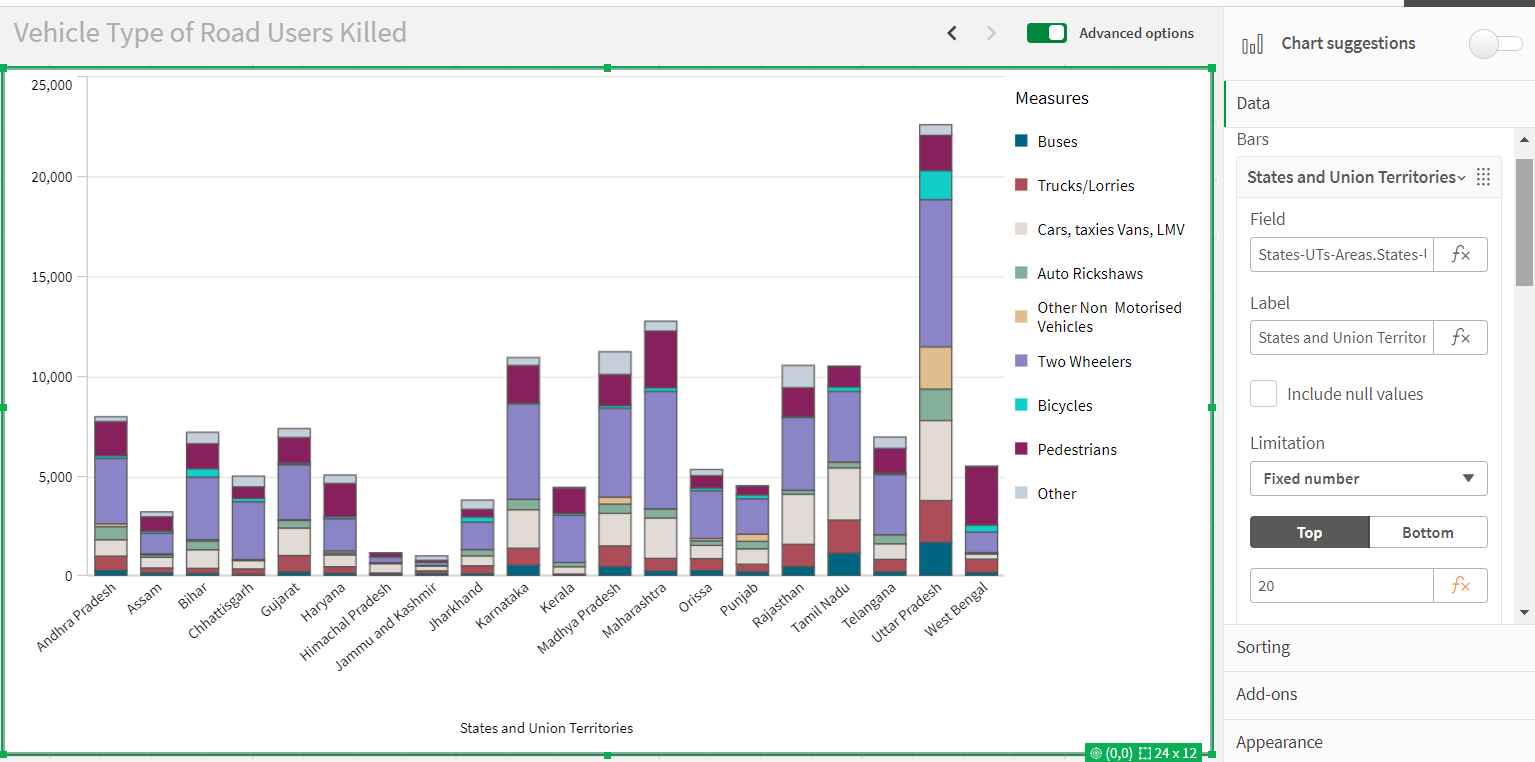
Actionable steps include:

* **Real-time data integration**: Integrate real-time traffic and weather data into predictive models to anticipate accidents and deploy preventive measures accordingly.
* **Strategic resource allocation**: Use predictions to allocate emergency services and traffic enforcement resources more effectively, reducing response times and accident severity.

These recommendations and actionable insights provide a clear path forward for improving road safety based on the data analysis, ensuring that interventions are both targeted and effective.

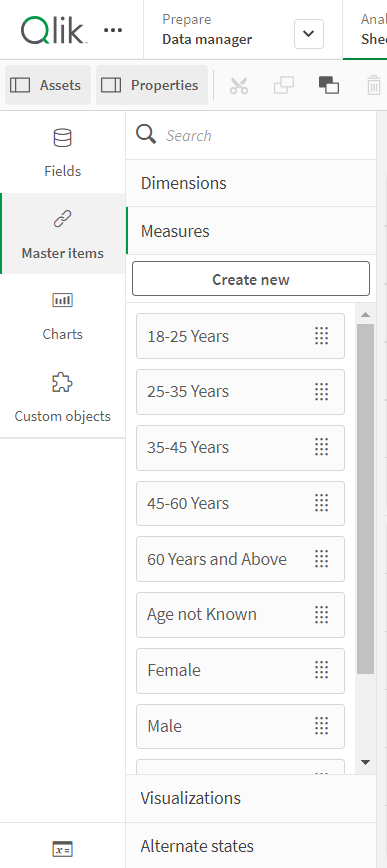
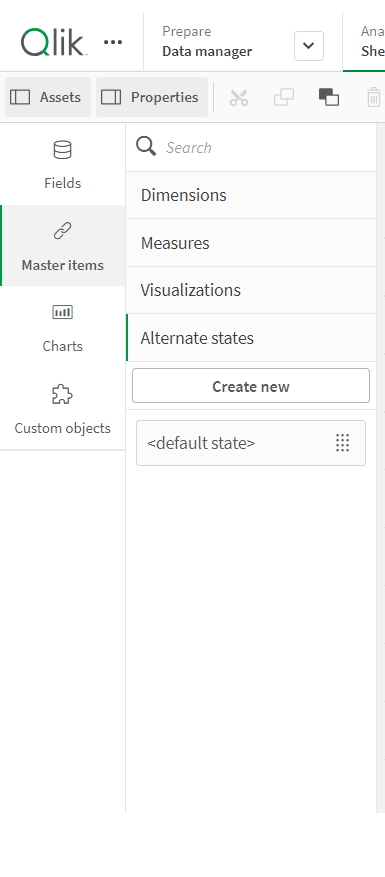
1. **Performance Testing:** 
   1. **Application Of Data Filters:**





Selections within the data allows users to filter data based on individual fields or dimensions. Users can choose specific values within a field to include or exclude from analysis. Complex filters based on predefined conditions and logic can also be created

* 1. **Calculated Fields:**



Qlik Sense allows the creation of reusable filter objects like Master Items, Calculated Fields which can simplify the process of applying consistent filters across multiple visualizations and dashboards.

* 1. **Number of Visualizations** 
     1. Accidents due to Drunken Driving
     2. State-wise Mobile Phone Usage
     3. Vehicle Contribution towards Total Accidents
     4. Correlation - Speeding and Number of accidents
     5. Accidents by Weather Type
     6. Minors Injured across the country
     7. Pedestrians Killed: Gender
     8. Pedestrians Killed: Age groups
     9. Road Users Killed: Vehicle Distribution

BY

DEEPAK SINGH