Qlik Business Analytics Program

Qlik Analysis of Road Safety And Accident Patterns in India

Github Repository : https://github.com/somethingshawnx/Road-Safety-Project

DEFINE PROBLEM/PROBLEM UNDERSTANDING

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 datadriven 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.

DATA PREPROCESSING:

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

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

Feature Classification and Handling:

Qlik classifies columns in the dataset based on their feature type (categorical, numeric, date, or free text).
It checks for sparsity, constants, and high cardinality (unique values) in columns.
Features with too many null values or low predictive value are excluded.
Feature Scaling and Standardization:
Qlik calculates summary statistics for each column to guide feature scaling. Standardization ensures that features have similar scales, preventing dominance by large values.

PREPROCESSED DATASET:

https://e2r6v1en56xtnwn.sg.qlikcloud.com/dataset/6657f7d9529d271369ea6169
https://e2r6v1en56xtnwn.sg.qlikcloud.com/dataset/6657f9df627669741c961ad0
https://e2r6v1en56xtnwn.sg.qlikcloud.com/dataset/6657f9be96efc118aaede64e
https://e2r6v1en56xtnwn.sg.qlikcloud.com/dataset/6657f95096efc118aaede5eb
https://e2r6v1en56xtnwn.sg.qlikcloud.com/dataset/6657f911529d271369ea6223
https://e2r6v1en56xtnwn.sg.qlikcloud.com/dataset/6657f8f396efc118aaede595
https://e2r6v1en56xtnwn.sg.qlikcloud.com/dataset/6657f8cd96efc118aaede578
https://e2r6v1en56xtnwn.sg.qlikcloud.com/dataset/6657f88a627669741c9619cb
https://e2r6v1en56xtnwn.sg.qlikcloud.com/dataset/6657f84196efc118aaede51a
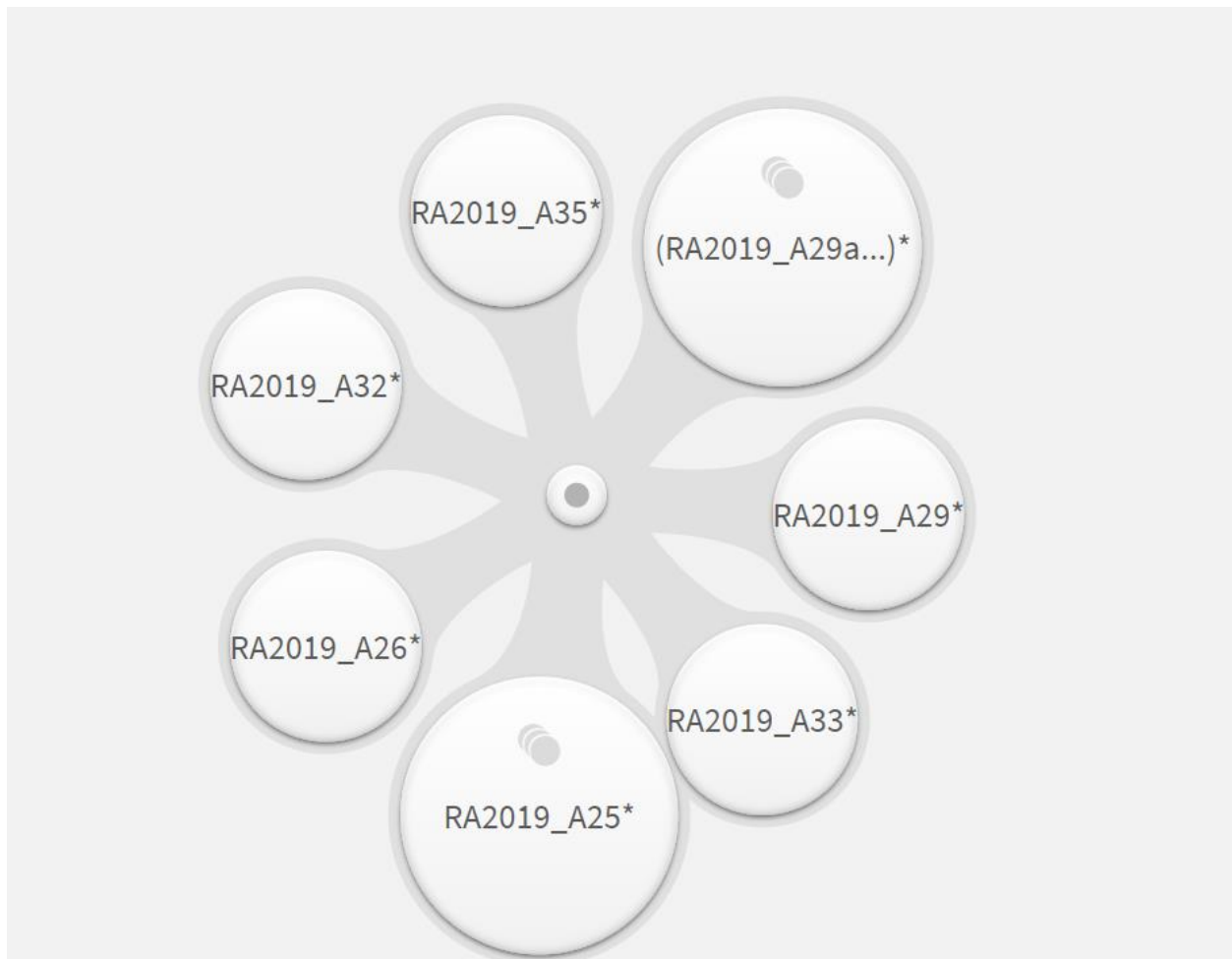
Data Preparation

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.

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.
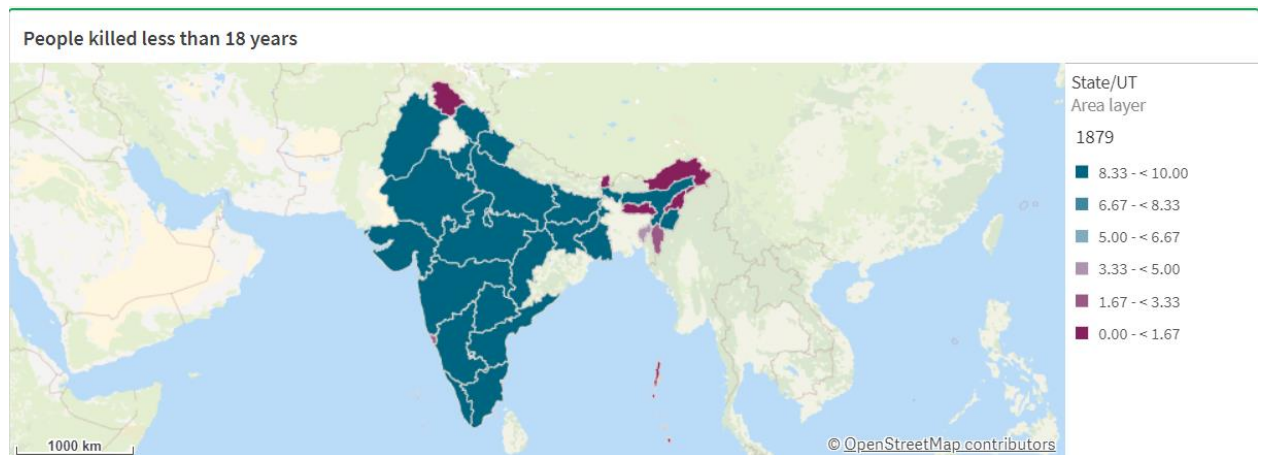
Fig : People killed less than 18 years

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.
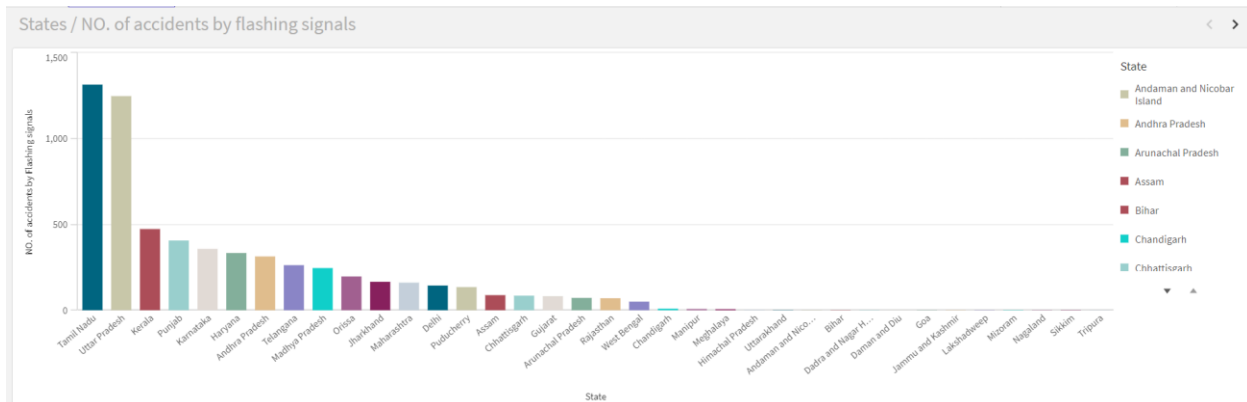

Fig : States vs NO. of accidents by flashing signals

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.
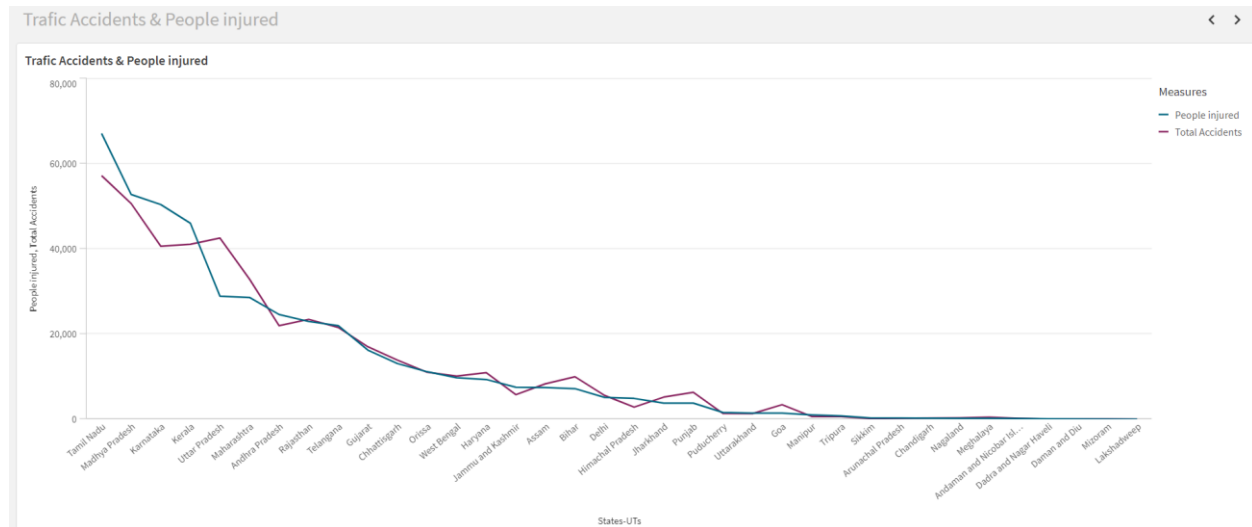
Fig : Traffic Accidents & People injured

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.
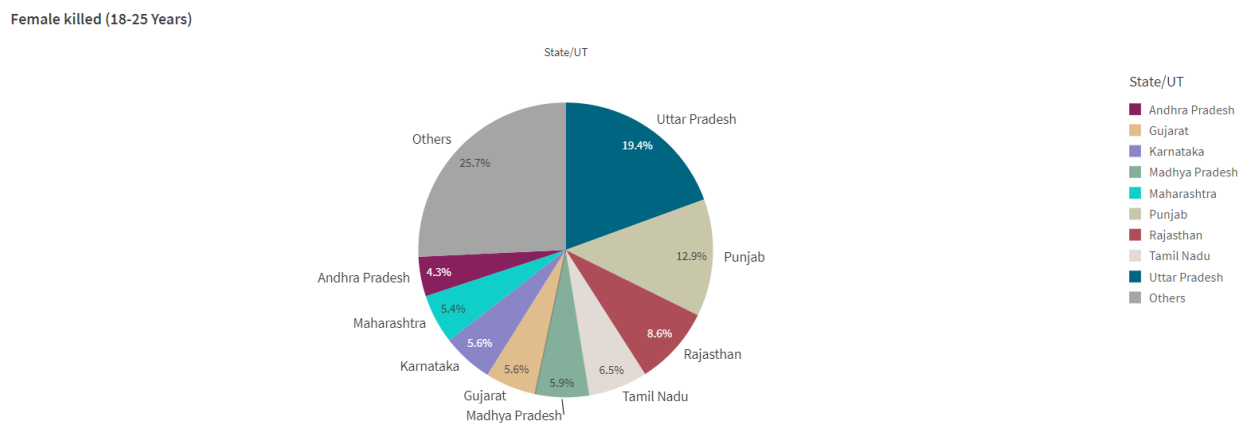
Fig : Female killed (18-25 Years)

Stacked 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.
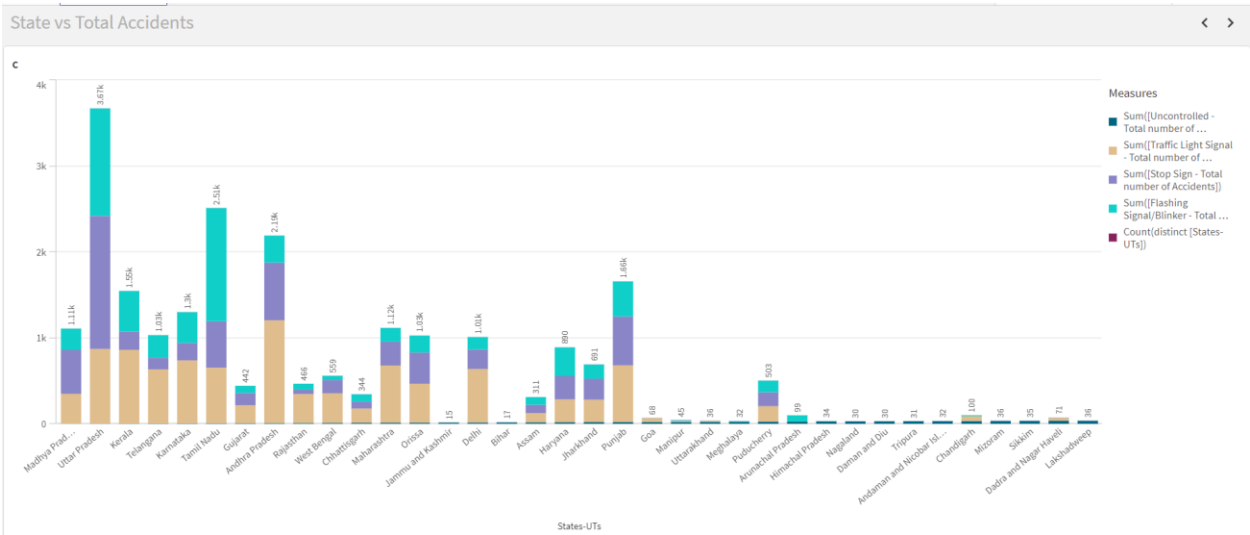
Fig 5 : State vs Total Accidents

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.
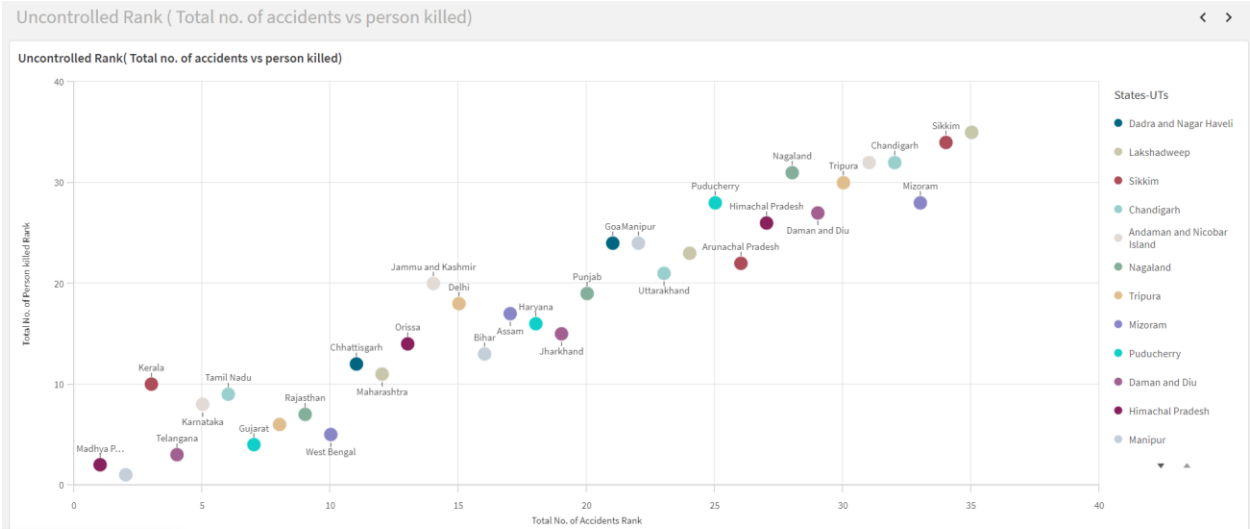
Fig 6 :   Uncontrolled Rank(Total No. of accidents vs person killed)

Multi Line Chart :A line chart using multiple measures, also known as a multiline 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.
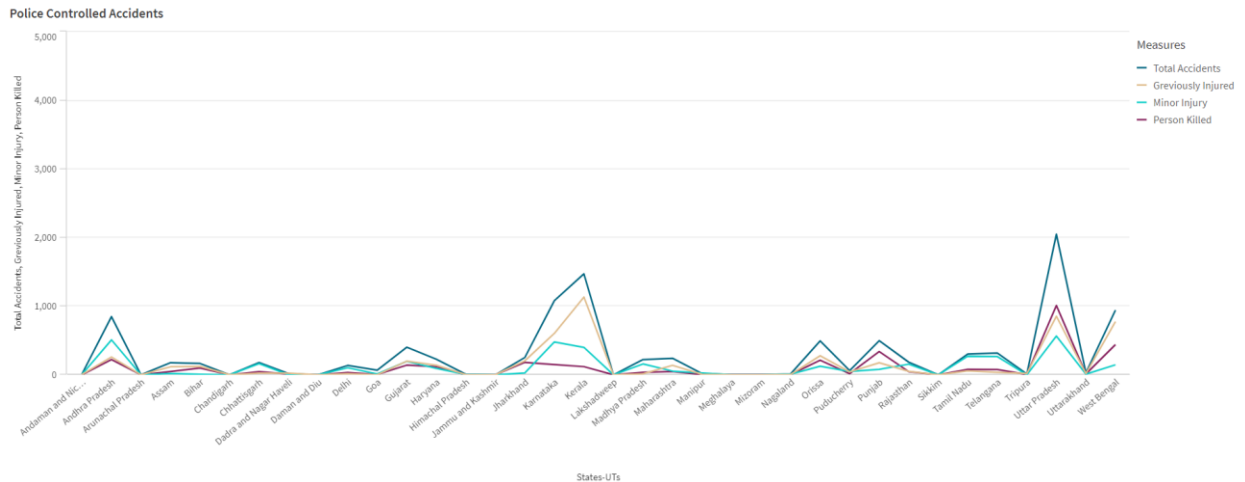

Fig 8 : Police Controlled Accidents

Grouped Bar graph:

A grouped bar graph also known as a clustered bar chart, is a type of bar graph that represents and compares multiple sets of data side by side in distinct groups. Each group contains two or more bars, where each bar represents a different category or variable. These graphs are particularly useful for comparing multiple categories across different groups or time periods.
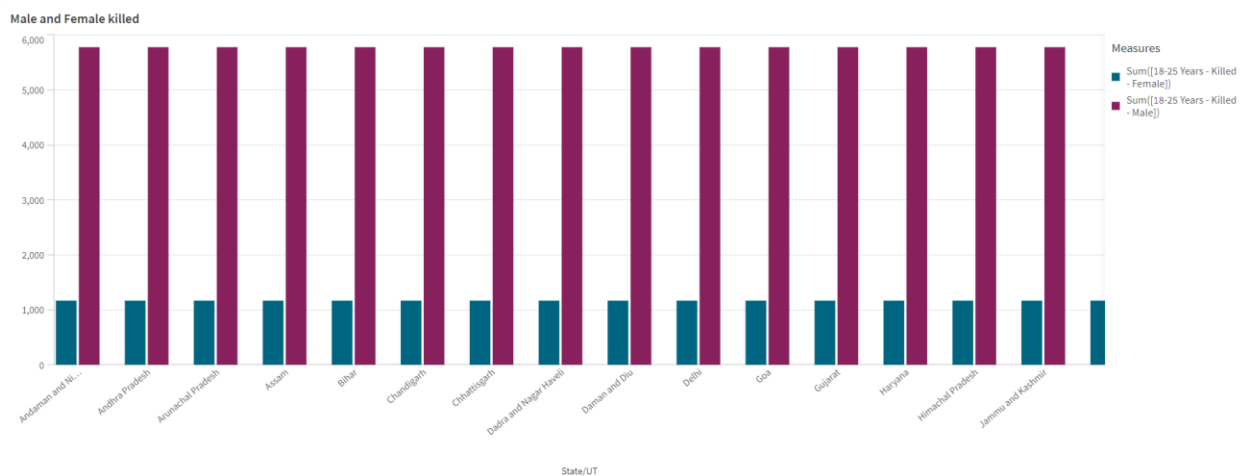
Fig: Male and Female Killed

Dashboard

A dashboard is a graphical user interface (GUI) that displays information and data in an organized and easy-to-read format. Dashboards are often used to provide real-time monitoring and analysis of data. They are typically designed for a specific purpose or use case. Dashboards can be used in a variety of settings, such as business, finance, manufacturing, healthcare, and many other industries. They can be used to track key performance indicators (KPIs), monitor performance metrics, and display data in the form of charts, graphs, and tables.

A dashboard in Qlik refers to a collection of visualizations, data, and analytics designed to provide a comprehensive view of key metrics and insights. Qlik, a leading data analytics and business intelligence platform, allows users to create interactive dashboards that enable real-time data exploration and decision-making.
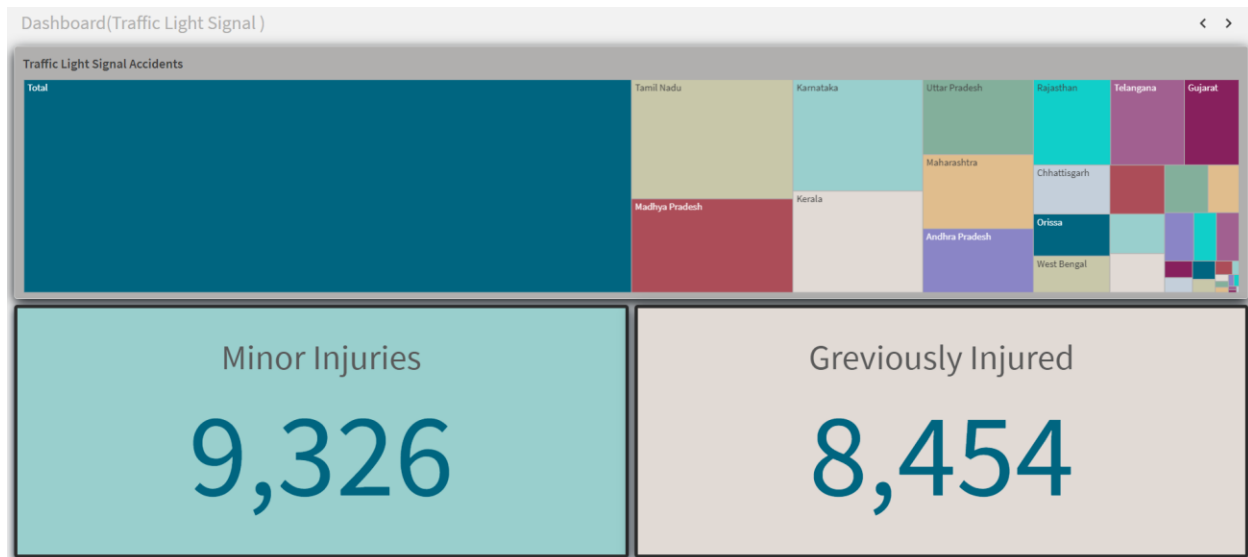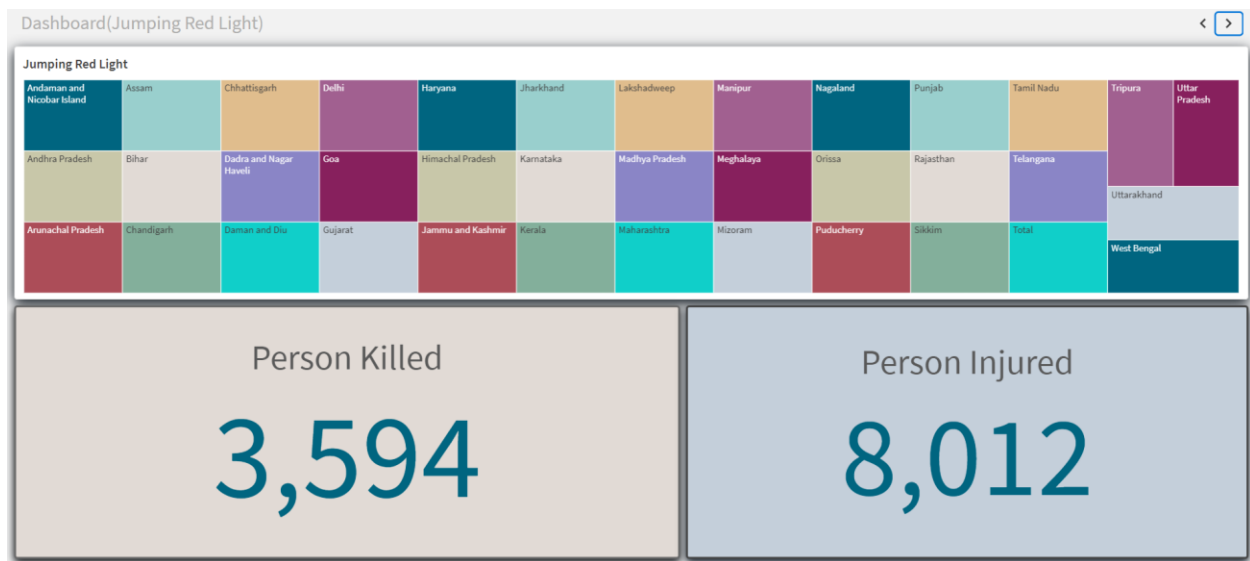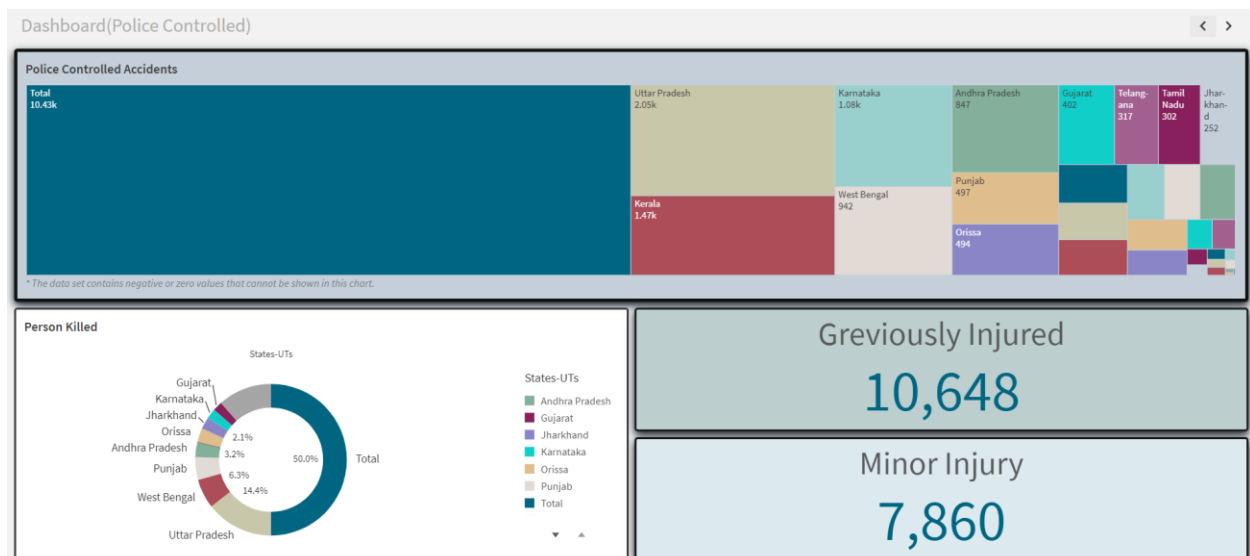

Fig : Dashboard 1

Fig : Dashboard 2

Fig : Dashboard 3

Storytelling

Storytelling in a Qlik project involves crafting a narrative around the insights and findings derived from data analysis using Qlik Sense. It's about presenting data in a compelling and informative way that engages stakeholders and communicates key insights effectively.

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
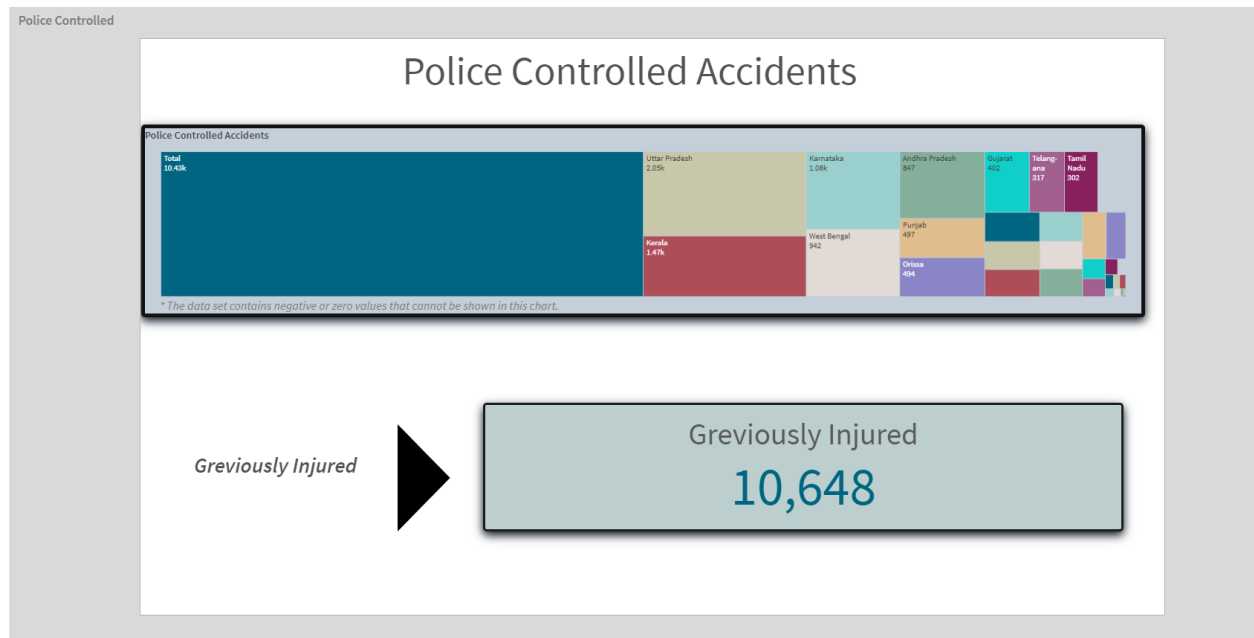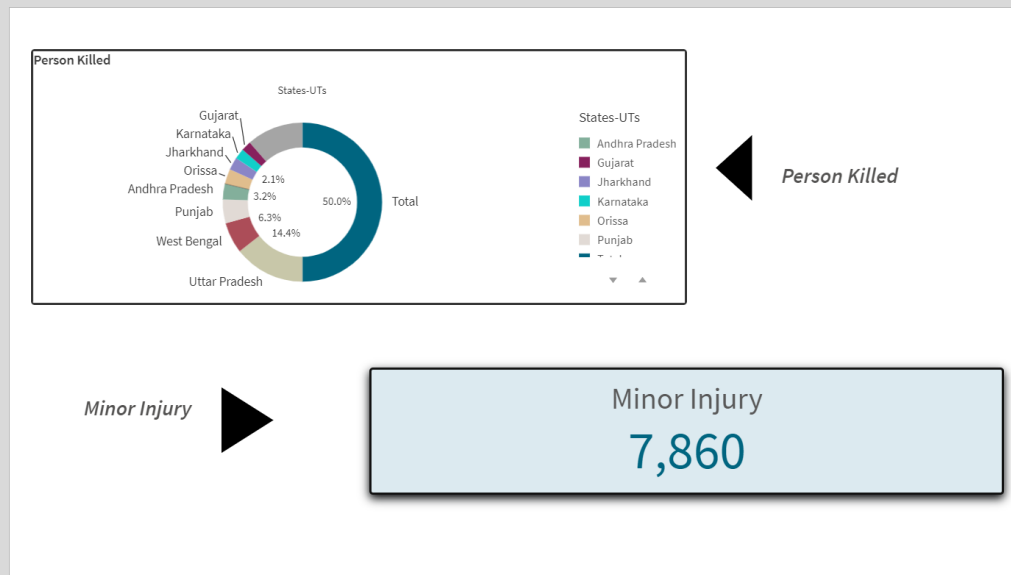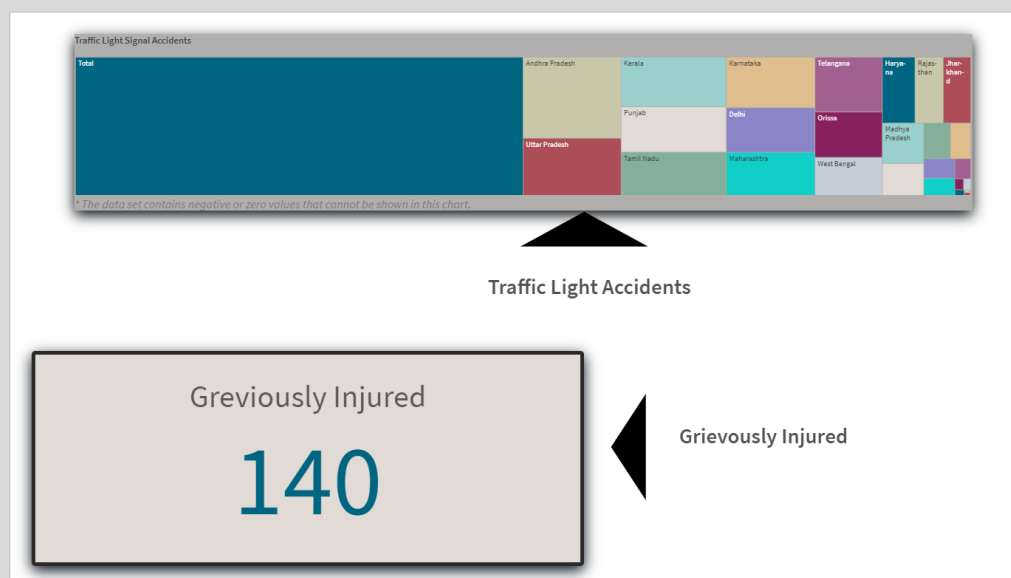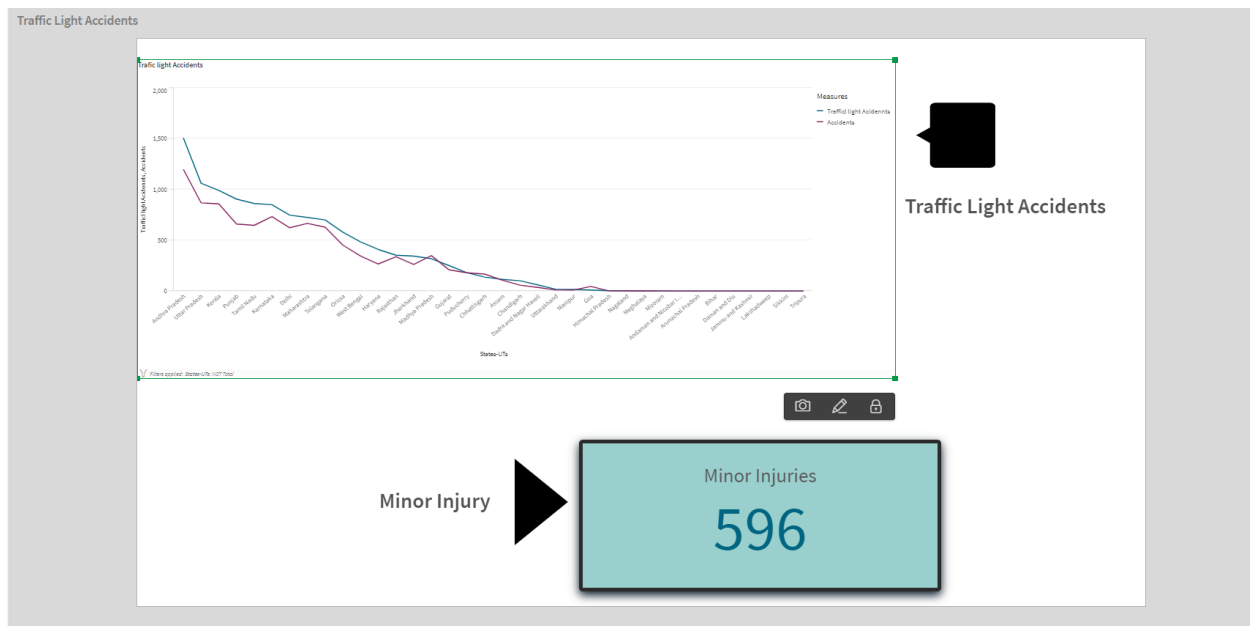
Fig : Storytelling 1

Fig : Storytelling 2

Performance Testing:

Applications of Data Filters

1. 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.
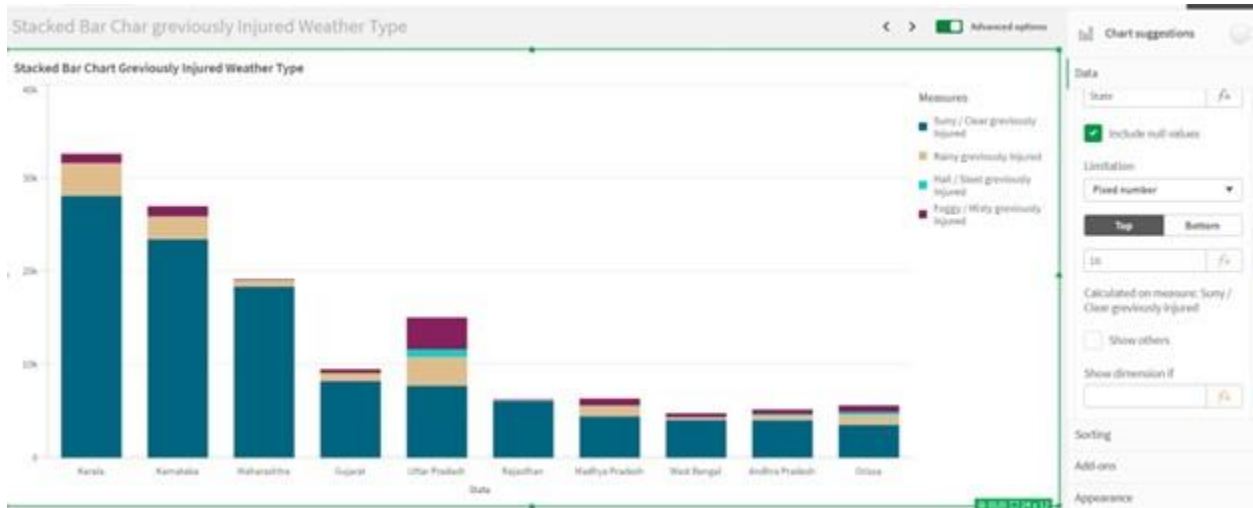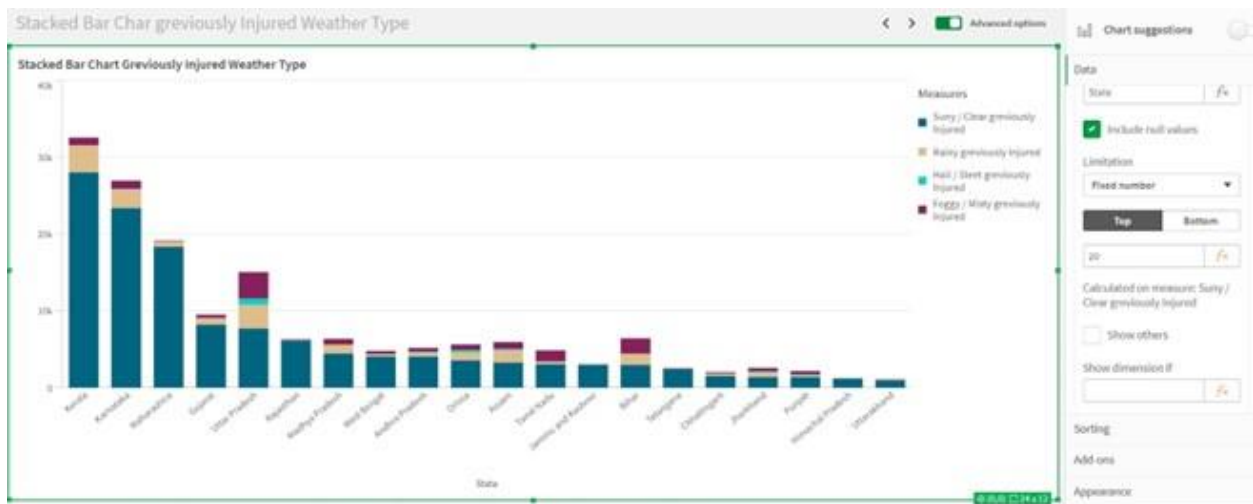

Fig 17 : Data Filter of 10 bars

Fig 18 :Data Filter of 20 bars

Use of Master Items :

Master Items are reusable objects that allow users to centralize and manage dimensions, measures, and visualizations across multiple sheets and apps. They provide a way to maintain consistency and standardization in data representation, while also facilitating efficiency in dashboard development and maintenance.
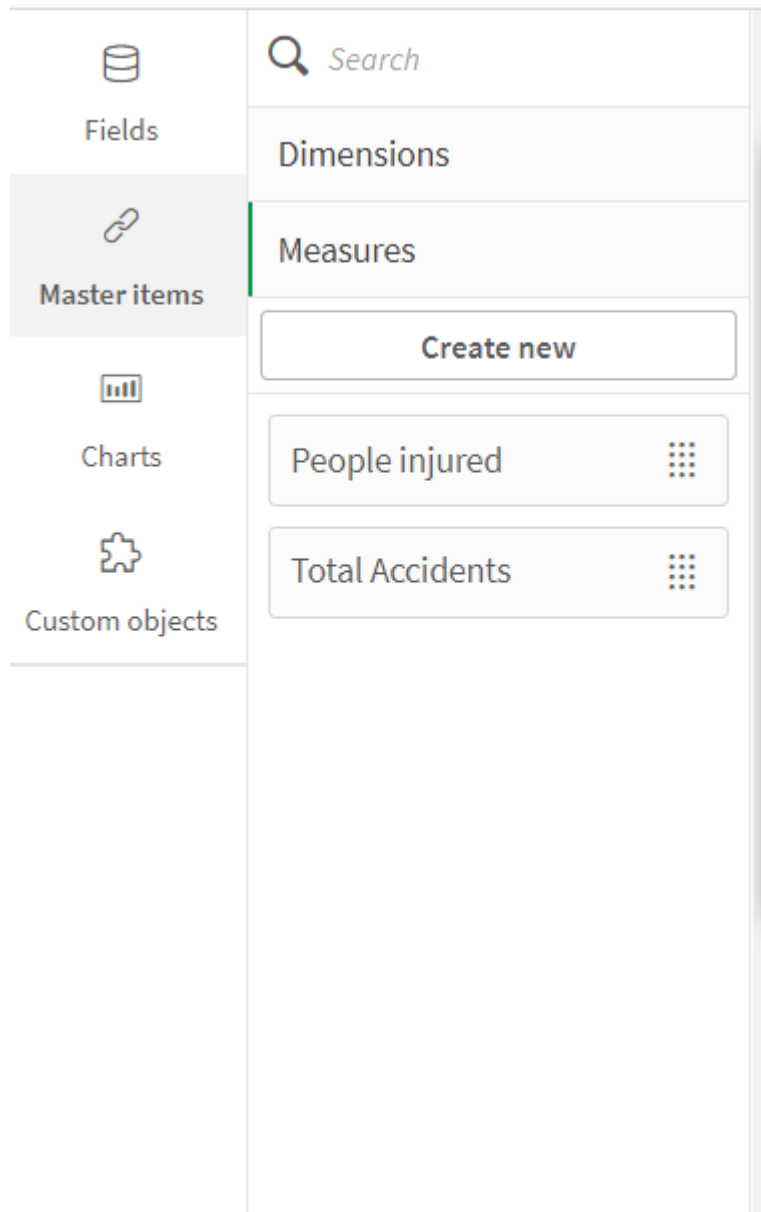

Fig : Master Items

Fields

Master items

Charts

Custom objects

Search

Dimensions

Measures

Create new

People injured

Total Accidents

People injured

people who are injured due to different reasons

Expression:
Sum({<[RA2019_A33.State/UT]=
{'Andaman and Nicobar Islands','Andhra pradesh','Arunachal

*The data set contains negative or zero values that can

Person Killed

States-UTs

More detail into master item

Number of Graphs / Visualizations :

- **Map Chart : People killed less than 18 years**

- **Bar Chart : States vs NO. of accidents by flashing signals**

- **Line Chart : Traffic Accidents & People injured**

- **Pie chart : Female killed (18-25 Years)**

- **Stacked Bar Chart : State vs Total Accidents**

- **Scatter Plot : Uncontrolled Rank(Total No. of accidents vs person killed)**

- **Multi Line chart : Police Controlled Accidents**

- **Grouped Bar graph : Male and Female Killed**

NOTE: there are more visualisation in my qlik cloud I mentioned some here.