Business Analysis with Qlik Virtual Internship Program

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

Introduction:

1.1 Overview: A brief description about the project:

The goal of this project is to increase road safety in India by identifying high-risk regions and analyzing accident patterns using Qlik's data analytics platform. Through the integration of many data sources, including traffic data, accident reports, meteorological data, road infrastructure, and demographic information, the project aims to detect patterns, pinpoint locations where accidents occur frequently, and forecast future incidents. In order to improve safety precautions and lower accident rates, these insights will support data-driven decision-making by government bodies, transportation agencies, and road safety groups.

Situation 1: Locating Hotspots With the use of Qlik's analytics, areas or particular roads in India with a high accident frequency can be found. The platform will identify areas that are more likely to have accidents by analyzing accident data in conjunction with traffic volume, road conditions, and time of day. This data is essential for putting targeted interventions into place, such as enhanced traffic monitoring, better signage, and lowered speed limits.

Situation 2: Examination of Trends In order to find trends and reoccurring causes of accidents, the platform will examine past accident data. This entails looking at various accident kinds (such as collisions and pedestrian accidents), seasonal fluctuations, and driving habits (such as speeding and inattentive driving). In order to address the underlying causes of accidents, awareness campaigns, driver training programs, and legislative reforms will be guided by the insights obtained.

Situation 3 : Predictive Modeling Scenario Qlik will estimate possible accident situations using real-time data and predictive analytics. The platform is designed to provide early warnings and suggest proactive ways to reduce accidents by taking into account several elements, including weather forecasts, traffic flow patterns, and historical accident trends. Authorities will be able to better allocate resources and put preventative safety measures in place because of this predictive capabilities. The overall goal of this initiative is to

employ data analytics to save lives, increase road safety in India, and lower the number of accidents.

1.2. Purpose: The use of the project. What can be achieved using this:

The project's objective is to use the Qlik platform to increase road safety in India. Through the examination of weather patterns, accident reports, road infrastructure, traffic data, and demography, our objectives are:

Determine Hotspots for Accidents.

Identify locations with a high accident rate.

Benefit: To prevent accidents, implement targeted safety measures such improved signs and modified speed limits.

Examine Accident Patterns Finding trends and typical causes of mishaps is the goal.

Benefit: Direct legislative changes, driver education programs, and awareness efforts to address the underlying problems. Estimate and Avoid Mishaps

Goal: Use current data to predict probable accidents.

Benefit: Increase overall road safety by preventing accidents with early warning systems and preventative initiatives.

By building safer roads in India, we hope to save lives, lower accident rates, and make well-informed decisions using Qlik's analytics

Define Problem / Problem Understanding

2.1 Specify the Business Problem

The problem of road accidents is multifaceted, involving various factors that contribute to high incidence rates and severe consequences. Human factors such as drivers behaviour speeding, reckless driving, and distracted driving are significant contributors, as is pedistrain behavior like jaywalking. Environment factors, including poor road conditions and adverse weather, further exacerbate the risk of accidents Vehicle-related issues such as poor maintenance and inherent design flaws also play a critical role. Inadequate infrastructure, charaterized by poorly designed roads and inefficient traffic management, adds to the problem. Legal and regularly shortcoming, including inconsistent enforcement of traffic laws and outdated legislation, hinder effective road safety measures. Socioeconomic factors like limited access to road safety education and economic constraints that prevent proper vehicle maintenance contribute to the persistence of road accidents. The consequences are severe, with high rates of fatalities and injuries leading to significant physical, emotional, and economic burdens. Addressing this issue requires comprehensive strategies, including education and awareness campaigns, infrastructure improvements, stricter vehicle safety standards, better enforcement of laws, and the adoption of advanced technological solutions. A concerted effort from government authorities, law enforcement, civil society, and individuals is essential to mitigate the problem of road accidents effectively

2.2 Business Requirements

This analysis aims to provide key insights into user demographics, accident patterns, and problem areas. The primary focus is on creating interactive, visually compelling dashboards to support strategic planning and operational improvements. These insights will be crucial for making informed decisions, enhancing safety protocols, and ensuring regulatory compliance. By leveraging this data, organizations can better understand their users, identify trends, and address critical issues effectively. The ultimate objective is to foster a safer, more efficient environment through data- driven decision-making and proactive measures.

2.3 . Literature Survey

A literature survey for analyzing road safety and accident patterns involves researching and reviewing previous studies, articles, reports, and statistics on the topic. This includes examining methods and techniques used for analyzing accident data, as well as the findings and conclusions of these studies. Recommended sources include academic databases like PubMed, IEEE Xplore, Google Scholar, and institutional repositories. Additionally, government reports and publications can offer valuable insights into recent developments in the field.

Data Collection

Collect the Dataset

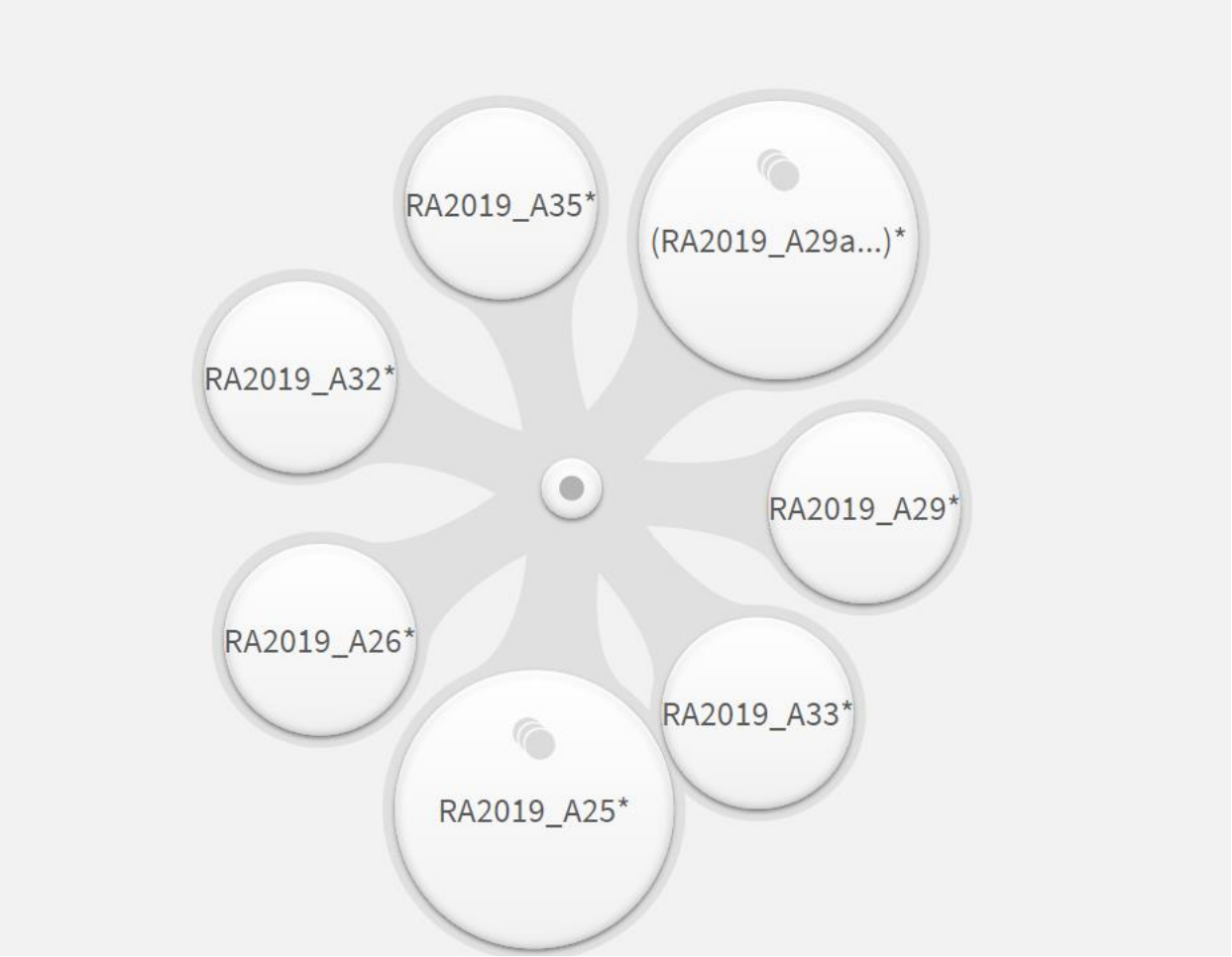
- **Identify Data Sources:** Determine where your data resides. This could be in files (such as CSV, Excel), databases (like SQL Server, Oracle), cloud services (such as Google Sheets, Salesforce), or other sources.
- **Access Data Load Editor:** Open your Qlik Sense application and navigate to the Data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

➤ Connect Data With Qlik Sense

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

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.

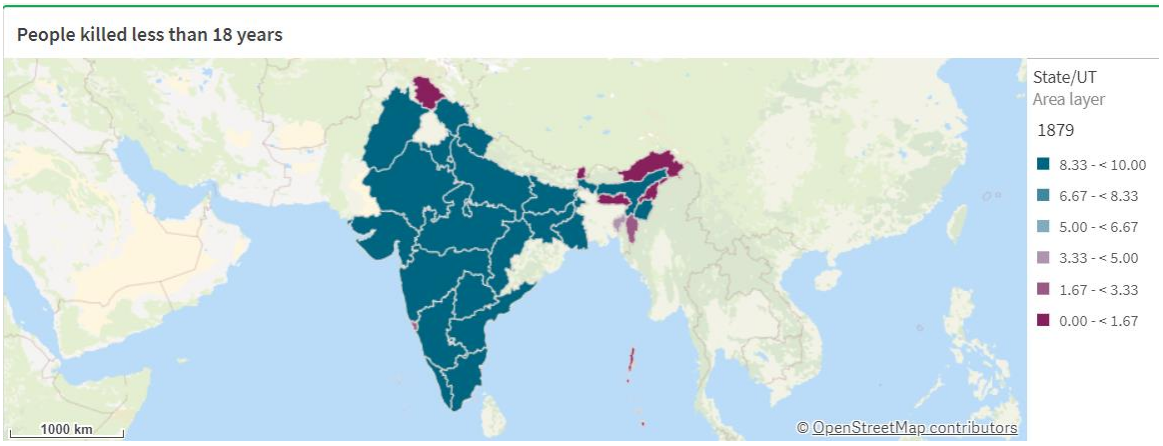


Fig 1 : People killed less than 18 years

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.

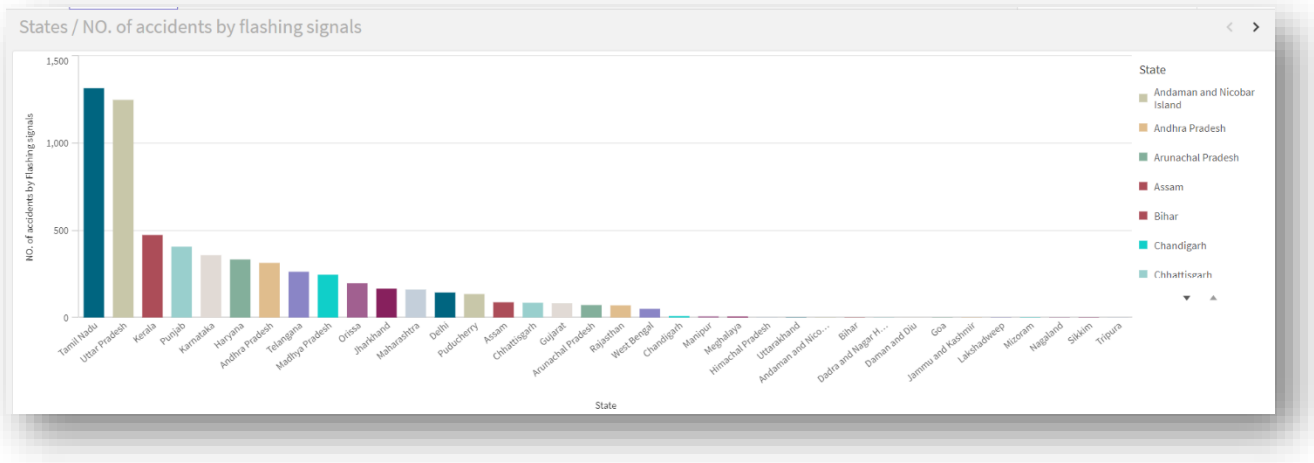
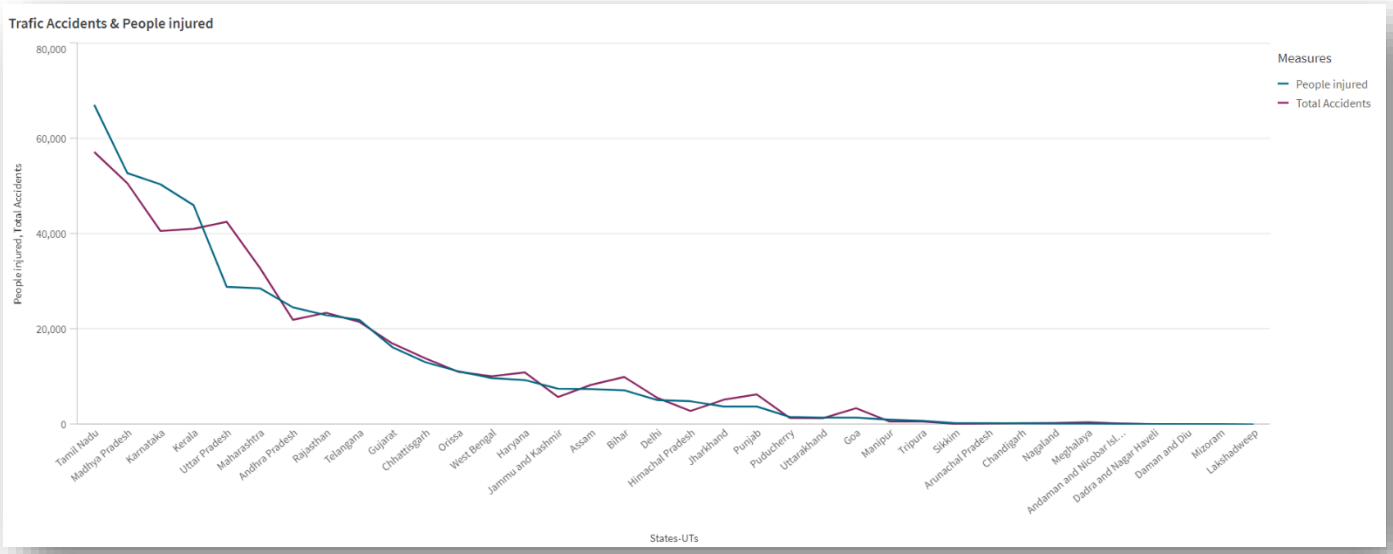


Fig 2 : States vs NO. of accidents by flashing signals

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 3 : 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.

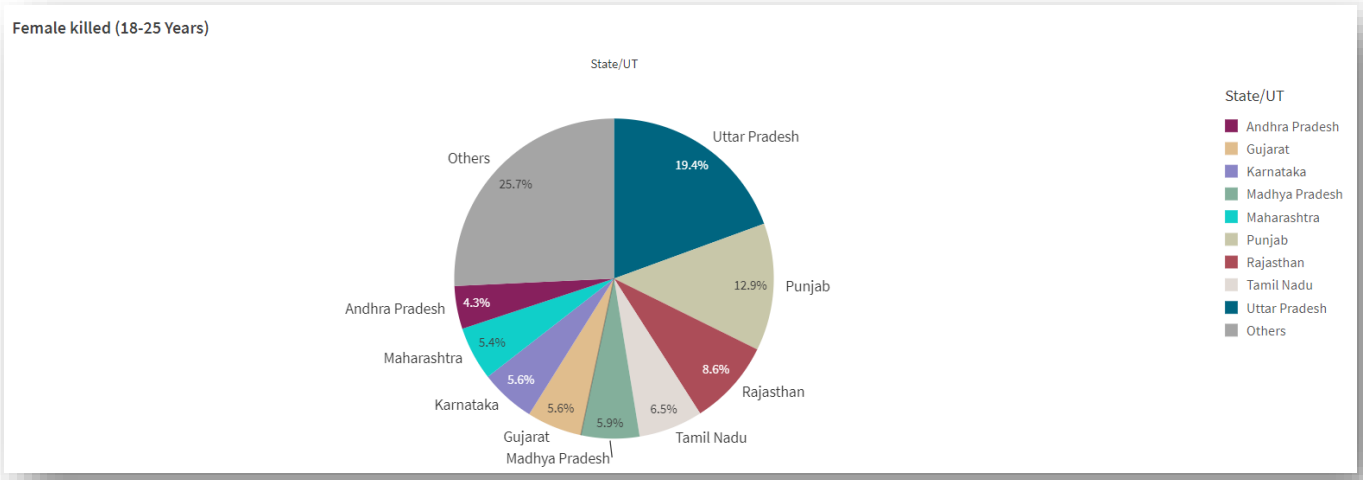


Fig 4 : Female killed (18-25 Years)

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.

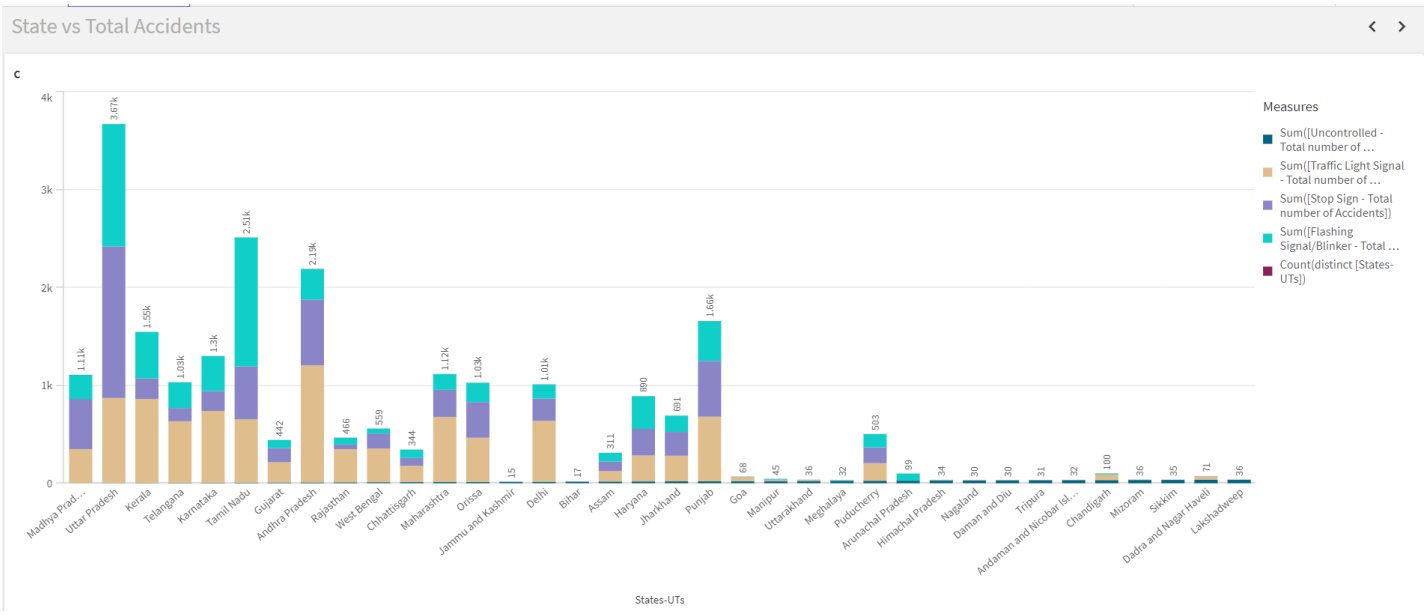


Fig 5 : State vs Total Accidents

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.

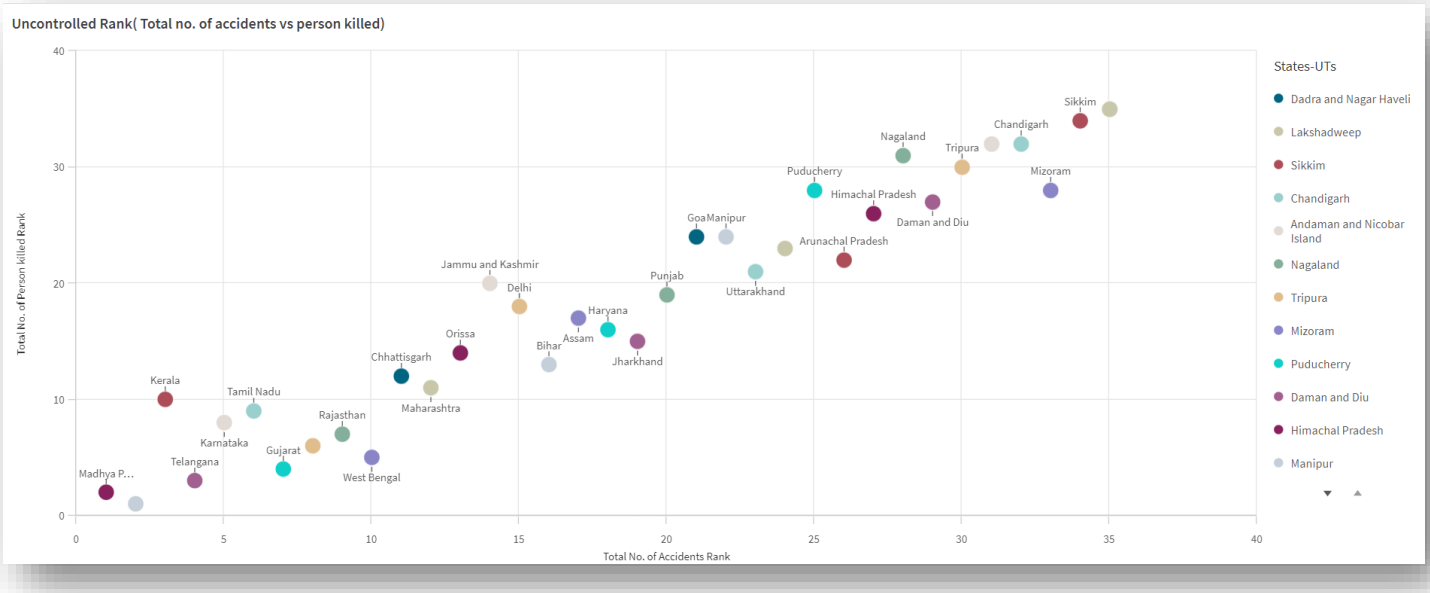


Fig 6 : Uncontrolled Rank(Total No. of accidents vs person killed)

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.

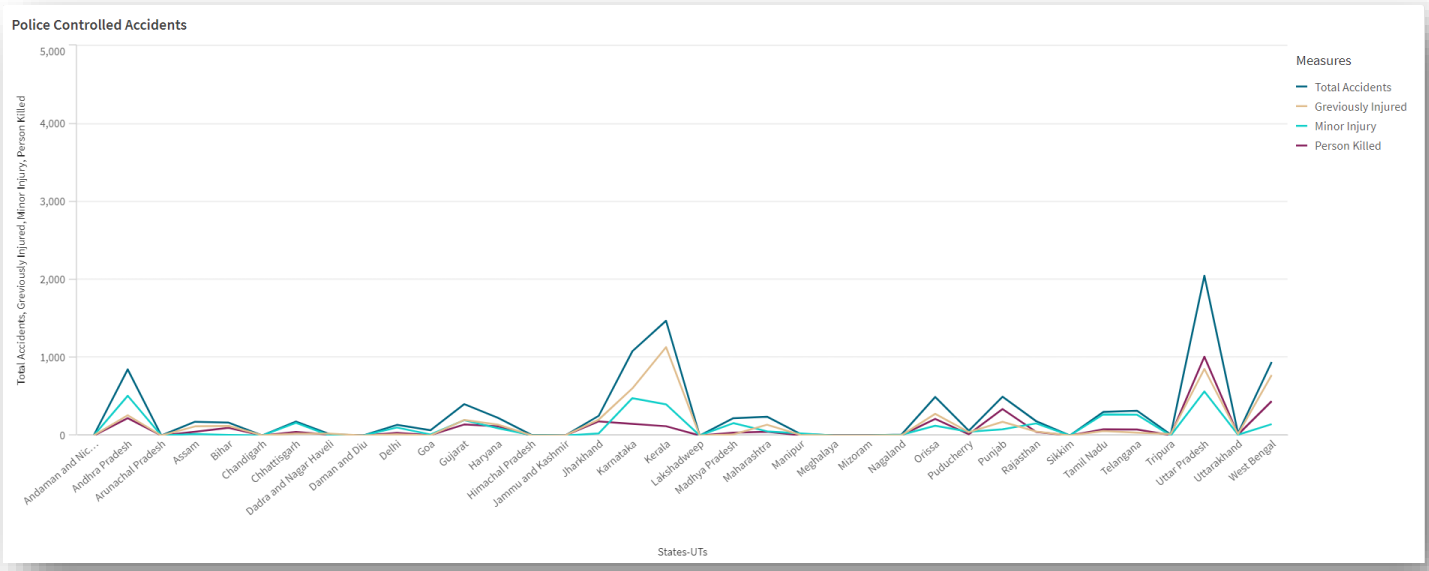


Fig 8 : Police Controlled Accidents

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.

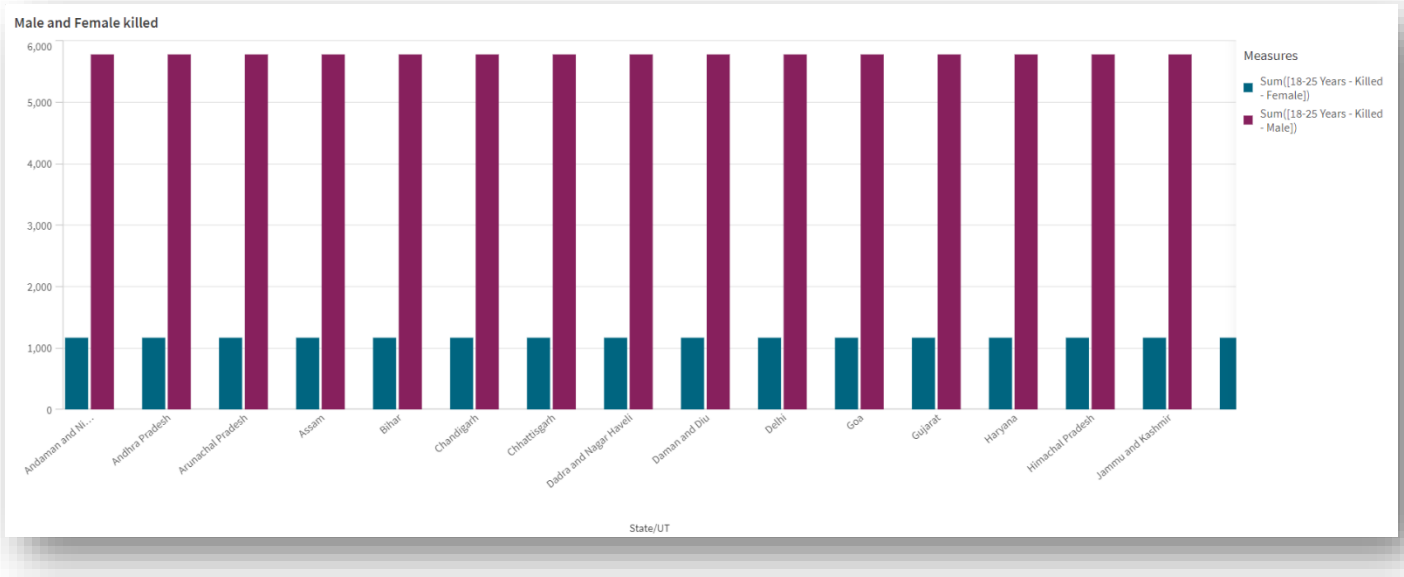


Fig: Male and Female Killed

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.

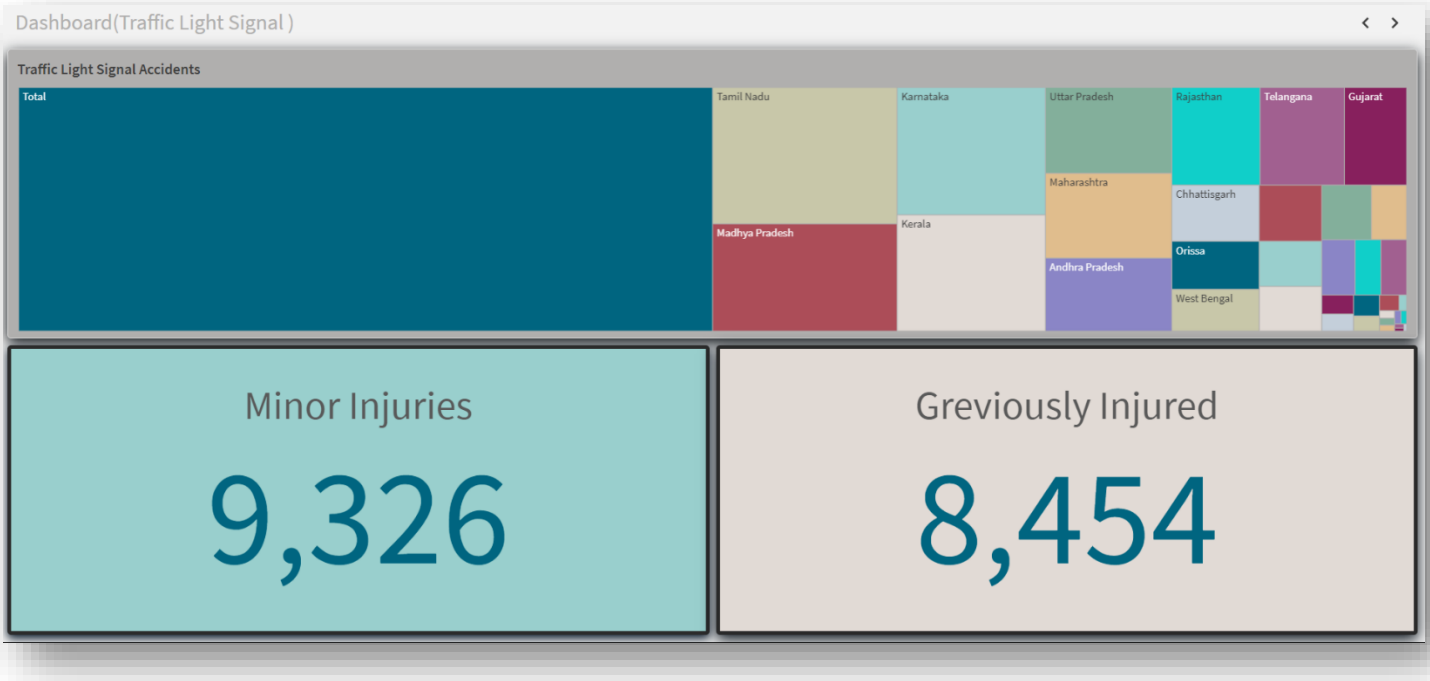


Fig : Dashboard 1

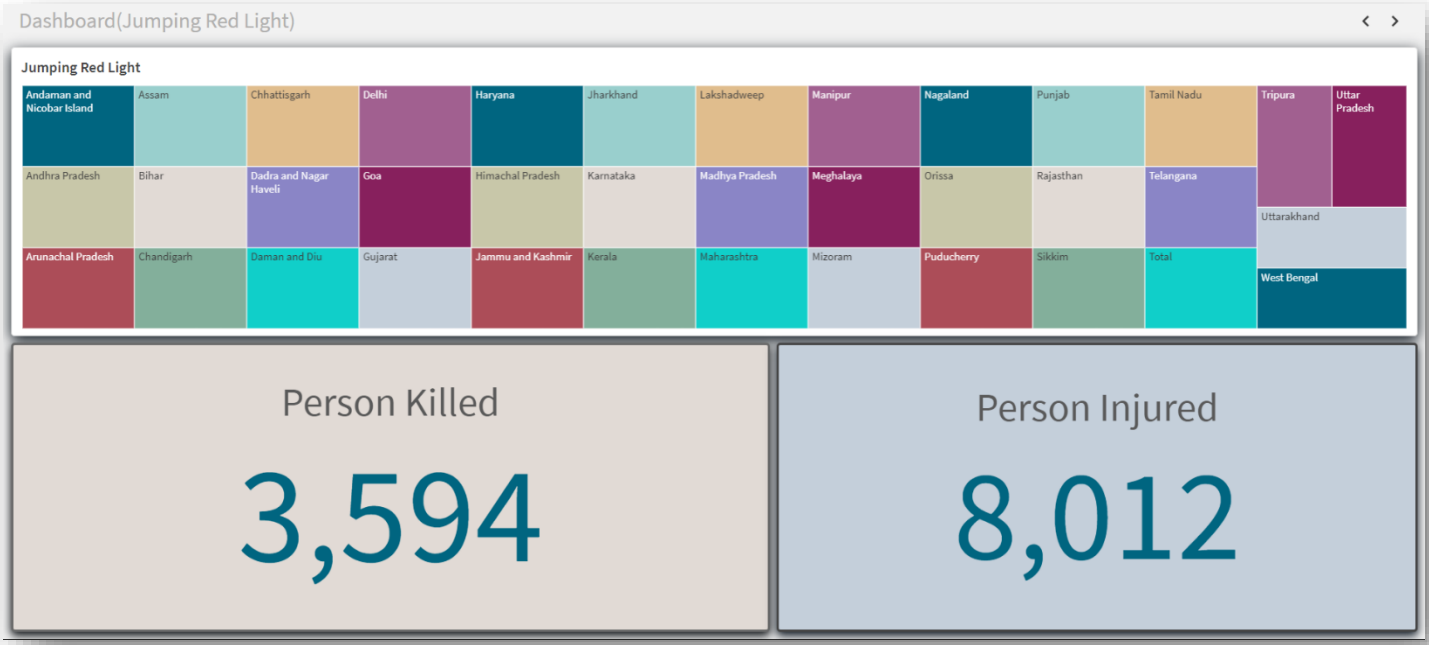


Fig : Dashboard 2

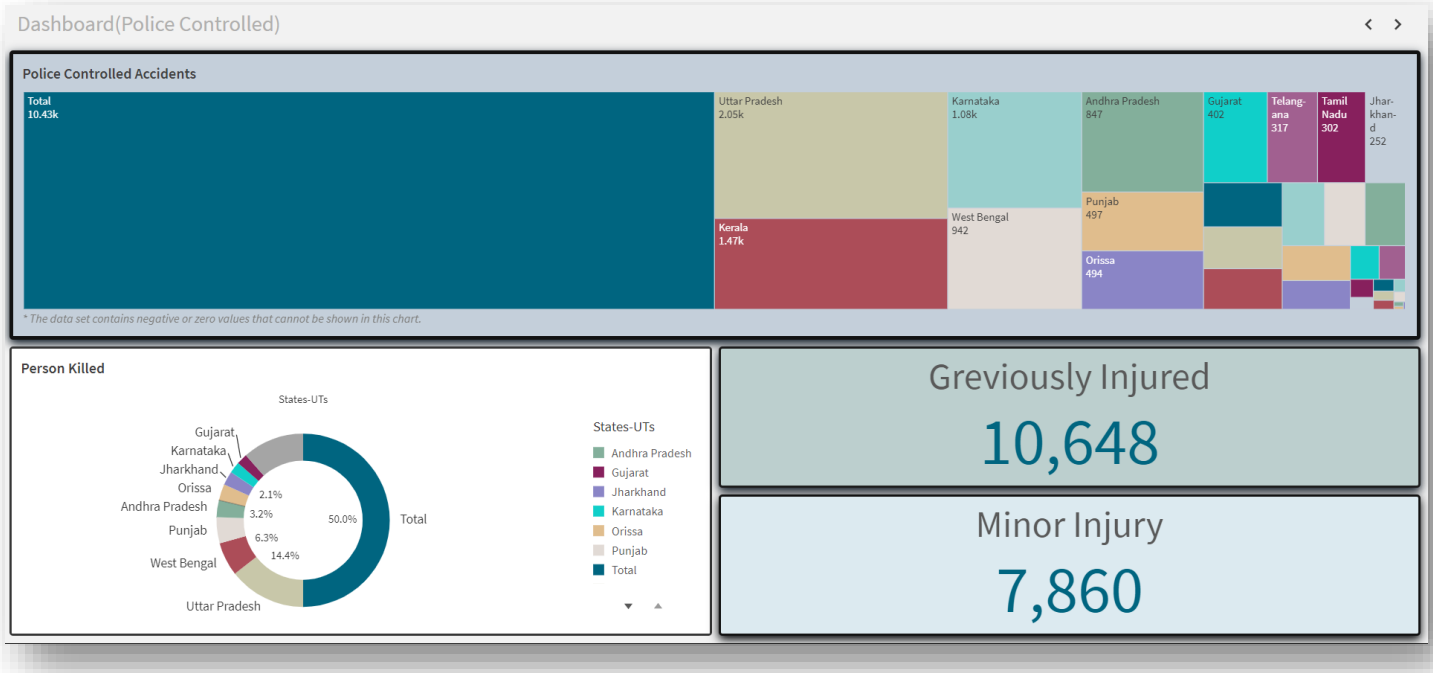
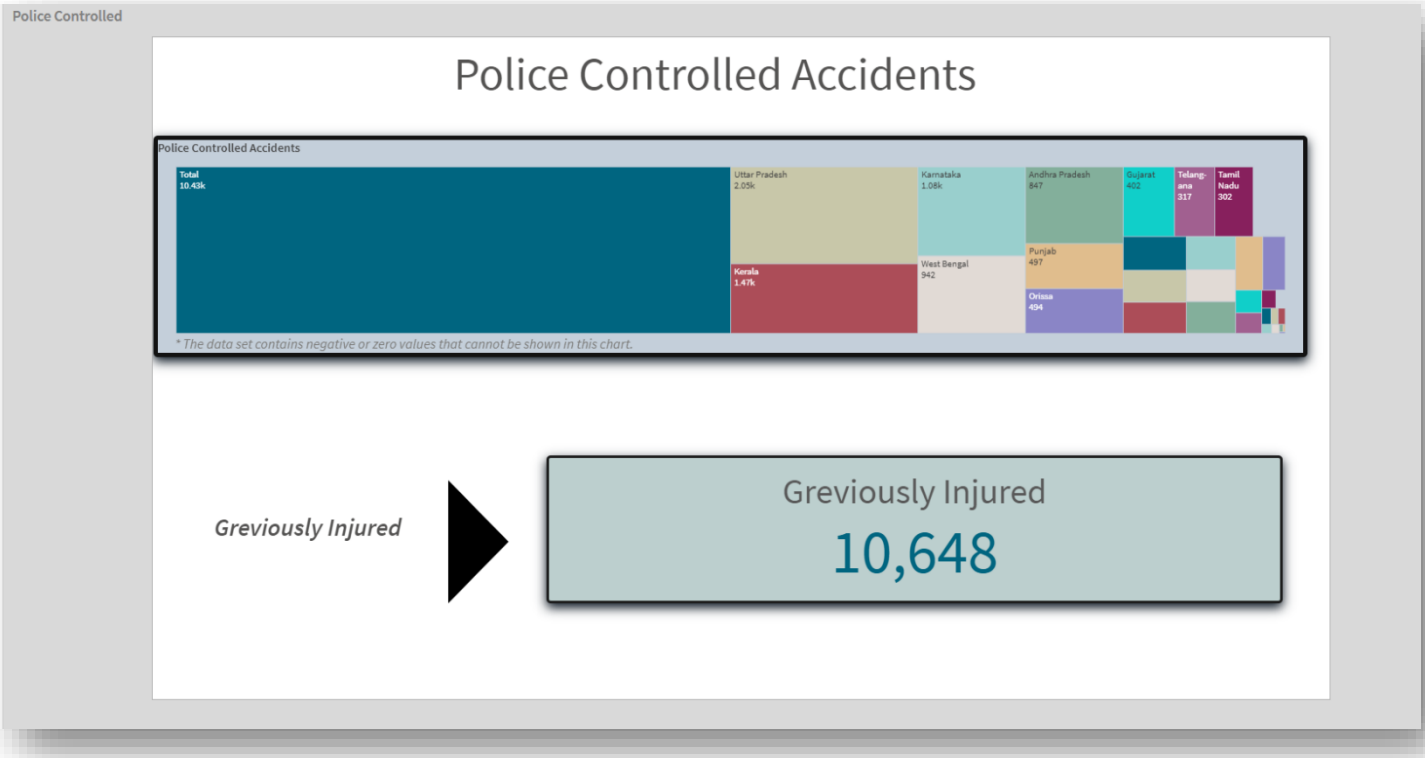


Fig : Dashboard 3

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



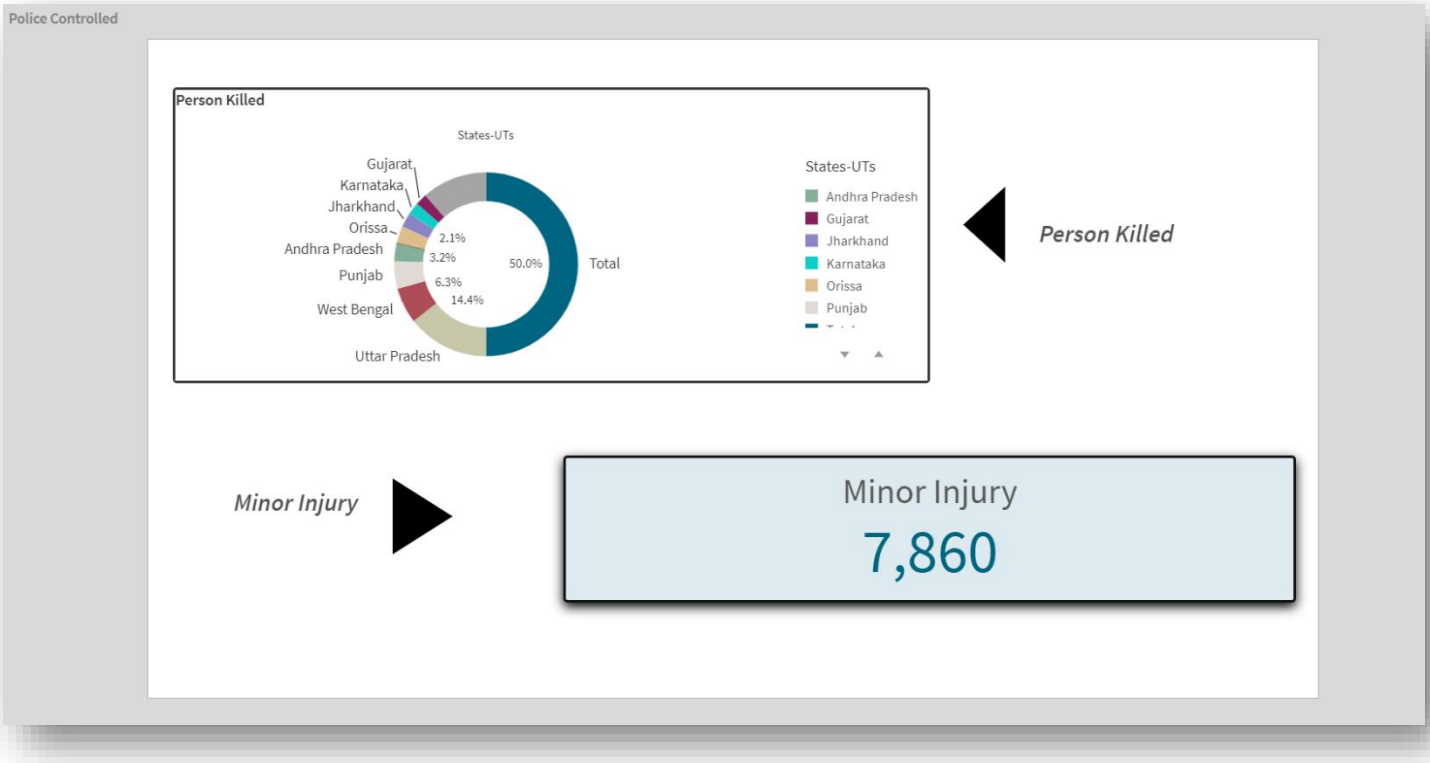


Fig : Storytelling

Performance Testing:

Applications 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.

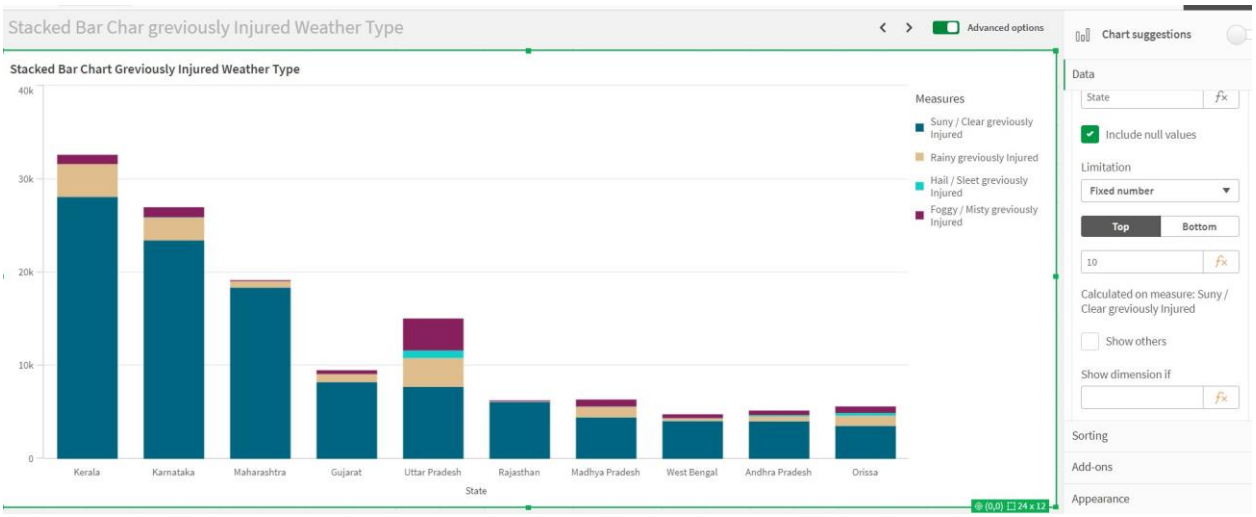


Fig 17 : Data Filter of 10 bars

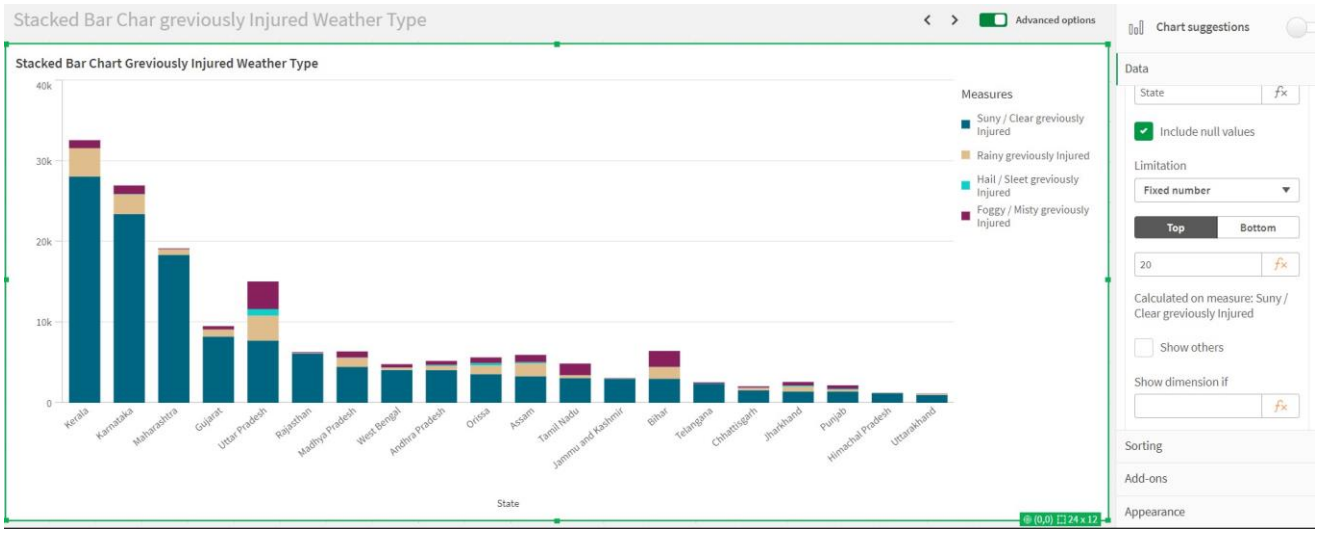


Fig 18 :Data Filter of 20 bars

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.

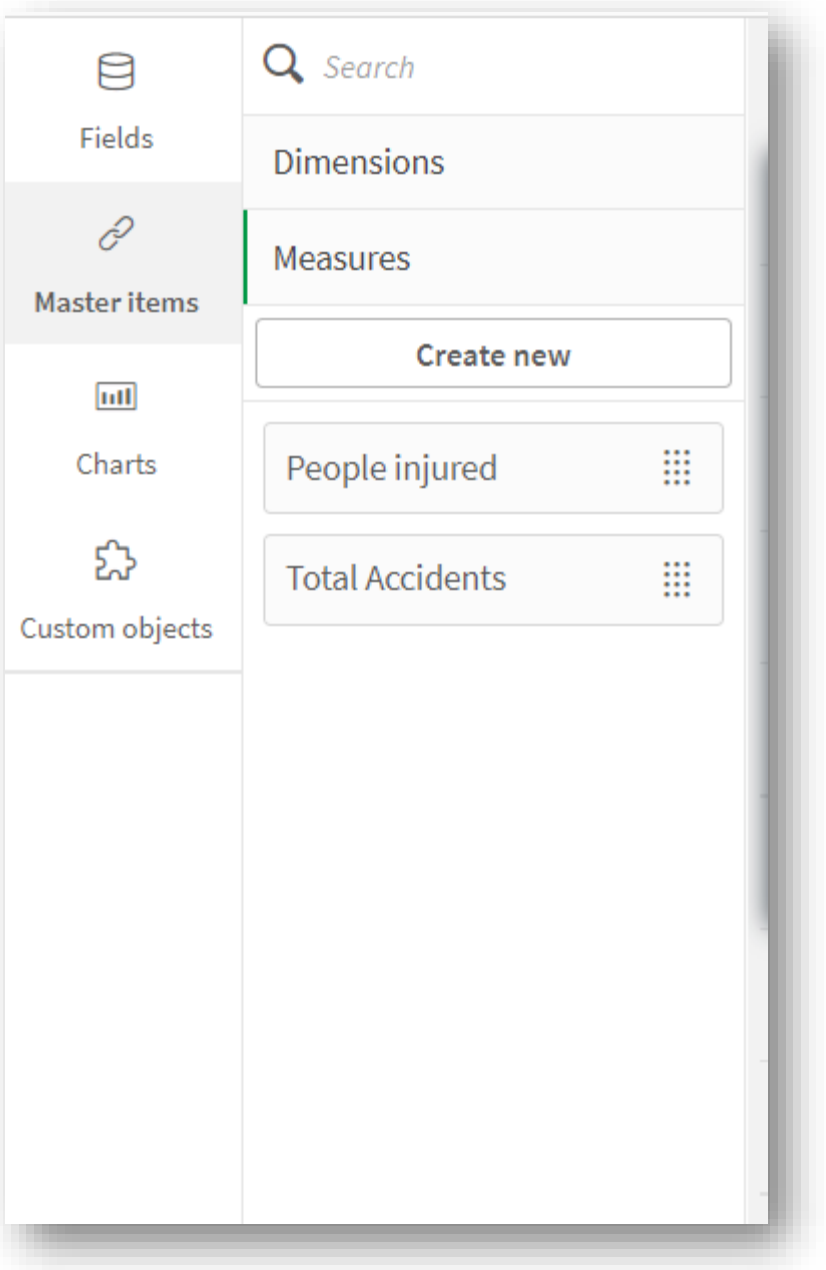
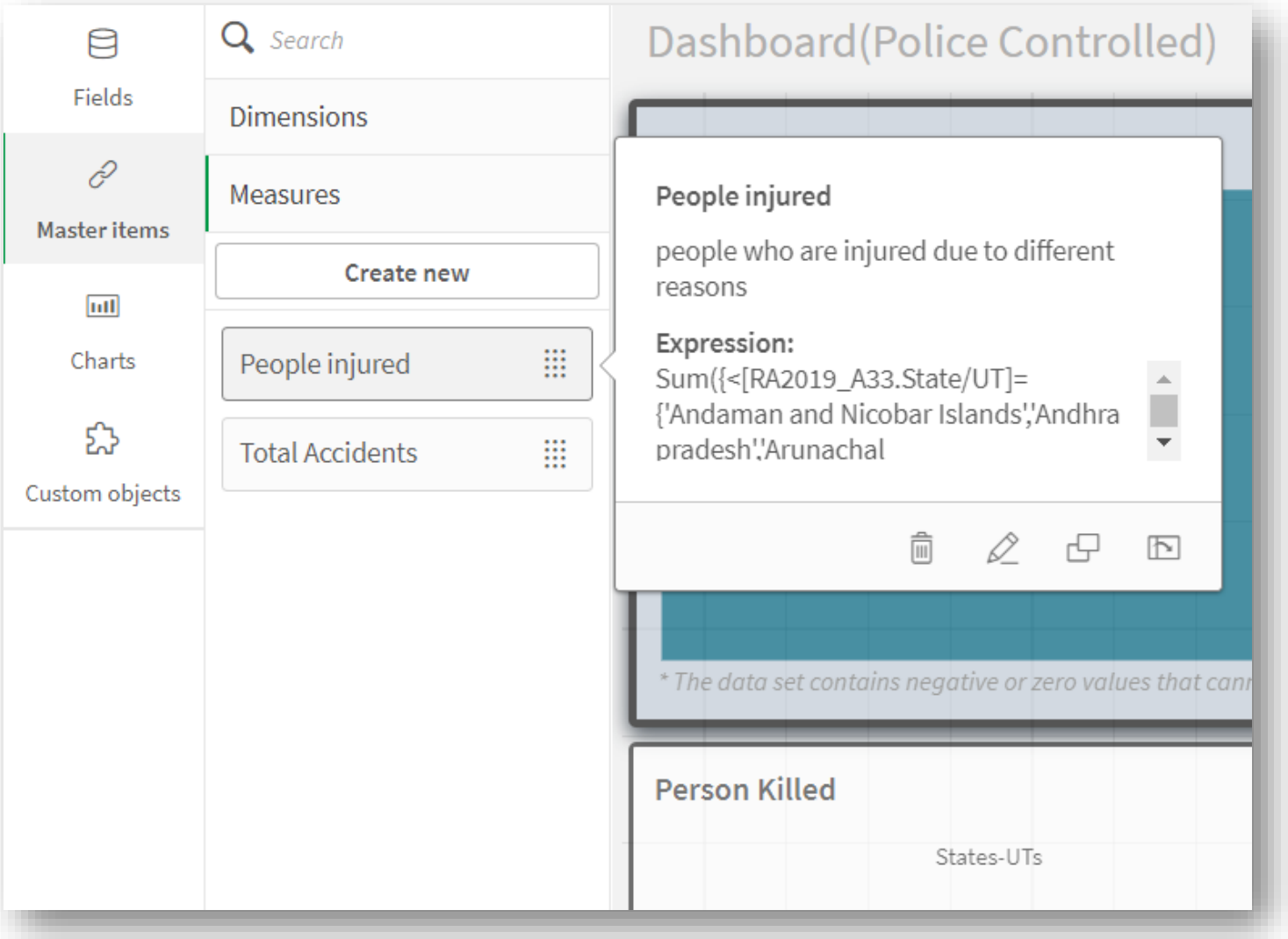


Fig : Master Items



More detail into master item

Number of Graphs / Visualizations :

- Map Chart : Minors killed
- Bar Chart : Accidents on wrong side
- Line Chart : Pedstraints of All Age Group
- Pie chart : Drunken Drive Pedstraints Killed
- Stacked Bar Chart : Persons killed By Weather Type
- Scatter Plot :Correlation Speeding & Number of Accidents
- Multi Line chart : Accidents Due to Driving on Wrong Side (including minors , greviously ,etc)
- Pie chart : Total Accidents Sunny / clear Weather