

BUSINESS ANALYTIC PROJECT

PROJECT TITLE

Qlik Analysis Of Road Safety And Accident Patterns In India

BY – TAMMU MANIVARDHAN

PROJECT DESCRIPTION -

The Qlik analysis of road safety and accident patterns in India provides an in-depth, data-driven examination of the factors contributing to road accidents across various states and union territories. Utilizing advanced analytics, this study identifies key trends, highlights critical areas for intervention, and suggests actionable insights for enhancing road safety.

.Key Findings:

1. **High-Risk States and UTs:** The Qlik analysis highlights Tamil Nadu and Madhya Pradesh as having the highest road accident rates, followed by Uttar Pradesh and Maharashtra, with key factors being over-speeding, negligent driving, and poor road conditions.
2. **Contributing Factors:** Analyze key contributing factors that lead to road accidents.
3. **Demographic Insights:** The Qlik analysis reveals that young males (18-35) are most vulnerable to road accidents, with higher rates in urban areas due to dense traffic and significant safety challenges in rural areas due to underdeveloped infrastructure.
4. **Temporal Patterns:** The Qlik analysis reveals that most road accidents in India occur during peak traffic hours and late at night, with a noticeable increase during the monsoon season due to poor visibility and slippery roads.

BUISNESS PROBLEM

The business problem addressed by the Qlik analysis of road safety and accident patterns in India is the high incidence of road accidents, which result in significant loss of life, injuries, and economic costs. By identifying patterns and key factors contributing to these accidents, the analysis aims to provide actionable insights to policymakers, law enforcement, and transportation authorities to implement effective safety measures, reduce accidents, and improve overall road safety.

BUISNESS REQUIRMENT

The business requirements for the Qlik analysis of road safety and accident patterns in India are:

- 1. Data Integration and Collection** - Collect comprehensive, up-to-date data on accidents, fatalities, and injuries across all states and union territories from sources like police records, hospital reports, and traffic surveys.
- 2. Data Quality and Accuracy** - Ensure data is clean, accurate, and consistent with validation checks.
- 3. Data Segmentation** - Segment data by location, time, weather, road type, and accident cause.
- 4. Analytical Tools and Techniques** - Utilize advanced Qlik tools to identify patterns, correlations, and trends, and predict accident hotspots.
- 5. Visualization and Reporting**
 - Develop interactive dashboards and customizable reports for stakeholders.
- 6. User Access and Security**
 - Define user roles and implement security measures to protect data privacy.
- 7. Actionable Insights**
 - Provide insights and recommendations for road safety improvements and targeted interventions.
- 8. Performance Monitoring**
 - Monitor and evaluate safety measures, updating the analysis with new data regularly.
- 9. Stakeholder Collaboration**
 - Facilitate collaboration among stakeholders and engage with local authorities and
- 10. Compliance and Standards**
 - Ensure compliance with national and international road safety standards and align with government policies.

Literature Survey for Qlik Analysis of Road Safety and Accident Patterns in India

Introduction

India faces significant challenges in road safety, necessitating data-driven solutions. Qlik's analytical tools can provide valuable insights for improving road safety measures.

Road Accident Analysis in India

1. Accident Statistics and Trends--

- NCRB Reports : Provide comprehensive data on accident frequency, causes, and impacts across states.
- MoRTH Studies : Identify accident-prone zones and assess road safety measures.

2. Contributing Factors---

- Human Factors: Over-speeding, drunken driving, and mobile phone use are major causes.
- Environmental Factors : Weather conditions, infrastructure, and traffic density affect accident rates.
- Vehicle Factors : Vehicle condition and type influence accident severity.

Data Analytics in Road Safety

1. Big Data and Predictive Analytics-

- Real-Time Monitoring**: Big data predicts accident hotspots and high-risk scenarios.
- Machine Learning**. Classifies causes and predicts accident severity using historical data.

2. Visualization Tools--

- Qlik's Capabilities: Effective for handling large datasets and creating interactive visualizations.
- Comparative Studies: Qlik offers superior user interactivity and data integration.

Case Studies and Applications

1. Regional Studies-

- State-Specific Analyses: Provide insights into local accident patterns and interventions.
- Urban vs. Rural Trends: Highlight the need for tailored safety measures.

2. Policy Implications

- Impact Assessment:-- Analyzes the effect of policy changes like stricter traffic laws on accident rates.

Conclusion

Leveraging Qlik for road safety analysis can provide actionable insights, helping to develop targeted interventions and reduce accidents in India. Data-driven approaches are essential for enhancing road safety and reducing fatalities.

Data Collection & Extraction From Database

EXTRACTION OF DATASET -

Dataset is extracted from kaggle Platform for furthure Data analyticses Kaggle is the world's largest data science community with powerful tools and resources to help you achieve your data science goals.

Data set link - <https://www.kaggle.com/datasets/aryakittukrishnasai/road-accidents-in-india>

About Dataset

State/UT-wise Pedestrians killed according to classification of age and sex 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 There are 9 csv format file which as name given below:

RA2019_A24.csv

- RA2019_A25.csv
- RA2019_A26.csv
- RA2019_A29.csv
- RA2019_A29a.csv
- RA2019_A29c.csv
- RA2019_A32.csv
- RA2019_A33.csv AND RA2019_A35.csv

UNDERSTANDING OF DATSET

Here's a simplified table format for the dataset description:

Data set	Key Columns
1. Pedestrians Involved in Accidents	State/UT, Less than 18 years – Male/Female, 18- 25 Years – Male/Female, 25-35 Years – Male/Female, 35- 45 Years – Male/Female, 45-60 Years – Male/Female, 60 and Above – Male/Female, Age not known – Male/Female
2. Pedestrians Killed	State/UT, Less than 18 years - Male/Female, 18-25 Years - Male/Female, 25-35 Years - Male/Female, 35-45 Years - Male/Female, 45-60 Years - Male/Female, 60 and Above - Male/Female, Age not known - Male/Female
3. Pedestrians Killed – Impacting Vehicles	State/UT, Bicycles, Two Wheelers, Auto Rickshaws, Cars, Taxis, Vans, and LMV, Trucks/Lorries, Buses, Other Non-Motorized Vehicles (E-rickshaw etc.), Others, Total
4. Traffic Control Type	State/UT, Traffic Light Signal: Total Accidents, Persons Killed, Persons Injured (Grievously Injured, Minor Injury, Total Injured), Police Controlled: Total Accidents, Persons Killed, Persons Injured (Grievously Injured, Minor Injury, Total Injured), Stop Sign: Total Accidents, Persons Killed, Persons Injured (Grievously Injured, Minor Injury, Total Injured), Flashing Signal/Blinker: Total Accidents, Persons Killed, Persons Injured (Grievously Injured, Minor Injury, Total Injured), Uncontrolled: Total Accidents (Number, Rank), Persons Killed (Number, Rank), Persons Injured (Grievously Injured, Minor Injury, Total Injured), Others: Total Accidents, Persons Killed, Persons Injured (Grievously Injured, Minor Injury, Total Injured)
5. Weather	State/UT, Sunny/Clear: Total Accidents (Number, Rank), Persons Killed (Number, Rank), Persons

	Injured (Grievously Injured, Minor Injury, Total Injured), Rainy, Foggy and Misty, Hail/Sleet, Others:Total Accidents, Persons Killed, Persons Injured (Grievously Injured, Minor Injury, Total Injured)
6. Killed on Two Wheelers – Impacting Vehicles	State/UT, Bicycles, Two Wheelers, Auto Rickshaws, Cars, Taxis, Vans, and LMV, Trucks/Lorries, Buses, Other Non-Motorized Vehicles (E-rickshaw etc.), Others, Total
7. Road Users Killed – Gender	State/UT, Pedestrian, Bicycles, Two Wheelers, Auto Rickshaws, Cars, Taxis, Vans and LMV, Trucks/Lorries, Buses, Other Non-Motor vehicles (E-Rickshaw), Others, Each Category Split into: Male, Female, Total, Two Wheelers: Ran
8. Causes of Accidents	State/UT, Over-Speeding: Number of Accidents (Number, Rank), Persons Killed (Number, Rank), Persons Injured (Grievously Injured, Minor Injury, Total Injured), Drunken Driving/Consumption of Alcohol and Drugs, Driving on Wrong Side, Jumping Red Light, Use of Mobile Phone, Others: Number of Accidents, Persons Killed, Persons Injured (Grievously Injured, Minor Injury, Total Injured)
9. Accidents – Severity and Vehicles	State/UT, Pedestrian, Bicycles, Two Wheelers, Auto Rickshaws, Cars, Taxis, Vans and LMV, Trucks/Lorries, Buses, Other Non-Motorized Vehicles (E-rickshaw etc.), Others, Each Category Split into: Number of Road Accidents, Number of Persons Killed, Number of Persons Grievously Injured, Number of Persons Minor Injured, Total: Number of Road Accidents, Number of Persons Killed, Number of Persons Grievously Injured, Number of Persons Minor Injured

Each Data set contains more detailed columns as per the specific categories. For instance, the “Pedestrians Involved in Accidents” worksheet has columns for different age groups and genders, while the “Traffic Control Type” worksheet includes details about accidents and injuries by different types of traffic control.

DATA PREAPARATION

Data Preparation for Visualization

FOR preparing the given datasets for visualization, we have perform these steps:

1. **Consolidate the Data:** Combine the data from different sheets into a single, comprehensive dataset

FORMATION OF ASSOCIATION WITH ALL DATASET WHICH WAS THERE :-

Associating the various datasets enables comprehensive analysis and visualization of pedestrian and vehicular accidents in India. By linking data on accident causes, traffic control types, weather conditions, and demographic details, we can identify patterns, trends, and correlations. This holistic approach helps in understanding the factors contributing to accidents, thereby aiding in the formulation of effective safety measures and policies.

The screenshot shows the Qlik Sense Data Manager interface. At the top, there are tabs for 'Prepare Data manager' and 'Analyze Sheet'. The current sheet is titled 'Road-Safety-And-Accident-Patterns-In-India'. Below the tabs, there are buttons for '+ Add data' and 'Concatenate or join'. The main area displays a network diagram where several datasets are represented as nodes connected by lines, indicating associations. One prominent node is labeled '(RA2019_A29a...)'. Below the diagram, a table is shown with data from the file 'RA2019_A29a.csv'. The table has columns for State/Uts, Total, Public Transport, Private Transport, STATE/UTS, and Means of road transport. The data includes entries for Andaman and Nicobar Island across various categories like Two Wheelers, Auto Rickshaws, and Buses.

State/Uts	Total	Public Transport	Private Transport	STATE/UTS	Means of road transport
Andaman and Nicobar Island	6	-	-	ANDAMAN AND NICOBAR ISLAND	Cars, Taxis, Vans and LMV
Andaman and Nicobar Island	6	-	-	ANDAMAN AND NICOBAR ISLAND	Other Non-Motorized Vehicles(E-rickshaw)
Andaman and Nicobar Island	6	-	-	ANDAMAN AND NICOBAR ISLAND	Trucks/Lorries
Andaman and Nicobar Island	6	-	-	ANDAMAN AND NICOBAR ISLAND	Two Wheelers
Andaman and Nicobar Island	13	-	-	ANDAMAN AND NICOBAR ISLAND	Buses
Andaman and Nicobar Island	13	-	-	ANDAMAN AND NICOBAR ISLAND	Cars, Taxis, Vans and LMV

2. **Data Cleaning:** Ensure the data is clean, with no missing values or incorrect data entries.

- cleaning null value by removing the value

The screenshot shows the Qlik Sense Data Manager interface with a focus on the 'Set nulls' dialog box. The dialog box is overlaid on a data grid. The grid contains data for various states and vehicles types, such as Andaman and Nicobar Island, Andhra Pradesh, and Arunachal Pradesh, with counts for Bicycles, Two Wheelers, Auto Rickshaws, Cars, Trucks/Lorries, and Buses. The 'Set nulls' dialog box has tabs for 'Bucket', 'Replace', and 'Set nulls'. It also includes a section to 'Select the distinct values you want treated as null values' and a 'Manual null values' input field where the value '0' is listed. Buttons at the bottom include 'Cancel' and 'Set nulls'.

The first screenshot shows a table with 74 rows and 10 columns. A modal dialog is open to 'Replace' or 'Set nulls' for the 'Buses' column, which contains many null values. The second screenshot shows the same table after null values have been replaced with 0. The third screenshot shows the final table where all null values have been removed.

4.

Table after removing null values looks like this

In the pre-processing stage of data preparation, splitting involves dividing the dataset into different parts or subsets for various purposes.

This screenshot shows the 'Split' feature being used to create new fields from a single column. The original column 'Age distribution of total number killed' is being split into nine new fields based on age ranges: 'Age distribution of total number killed 18-25 killed', 'Age distribution of total number killed age<18 Killed', 'Age distribution of total number killed 25-35 Years killed', and 'Age distribution of total number killed 35-45 Years Killed'. The preview shows the count of age distribution for each category.

This screenshot shows the 'Split' feature being used to create new fields from a single column. The original column 'Total | 25-35 Years killed' is being split into four new fields: 'Age distribution of total number killed 18-25 killed', 'Age distribution of total number killed age<18 Killed', 'Age distribution of total number killed 25-35 Years killed', and 'Age distribution of total number killed 35-45 Years Killed'. The preview shows the count of age distribution for each category.

These splits are essential for building robust models, validating their performance, and ensuring that the data is used effectively in the analysis and modeling processes.

Unpivoting

In Qlik Sense is a data transformation process that converts columns into rows, making the dataset more

suitable for analysis:-

The screenshot shows the Qlik Sense Data Manager interface with the title "Road-Safety-And-Accident-Patterns-in-India". A table titled "RA2019_A33" is open, showing data from "RA2019_A33.csv" with 37 rows and 22 columns. An "Unpivot" dialog box is displayed over the table, with the "Cancel" button visible at the bottom right. The dialog lists several columns as attributes and one as a data field, indicating the transformation of wide-format data into long-format data.

on going to "Data Manager" or "Data Model Viewer" to see the transformed table.

Analyze Unpivoted Data:

You can now create visualizations using the unpivoted data

The screenshot shows the Qlik Sense Data Model Viewer interface with the same title "Road-Safety-And-Accident-Patterns-in-India". The table "RA2019_A39a..." is shown with 74 rows and 13 columns. An "Unpivot" dialog box is open, listing various transport modes and state/UTS categories as attributes and their corresponding counts as data fields. This view allows for easier analysis and visualization of the data.

This process helps in converting wide-format data to a long-format, making it easier to analyze and visualize different categories in Qlik Sense.

3. Data Structuring: Structure the data for different types of visualizations, like bar charts, line charts, heatmaps, and others.

CALCULATED FIELD -

calcualted is created in order to get new coloumn by using expression

figure :-1



figure 1.2

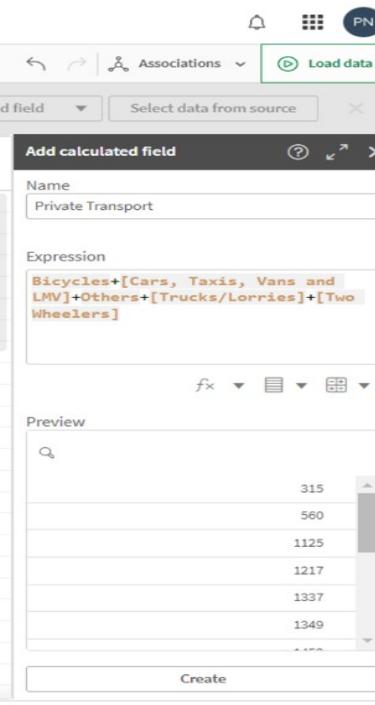


figure 1.3

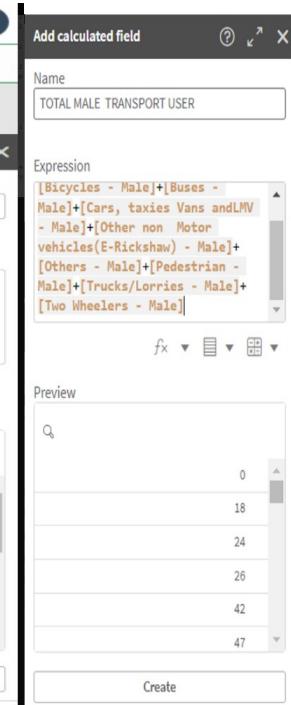


figure 1.4

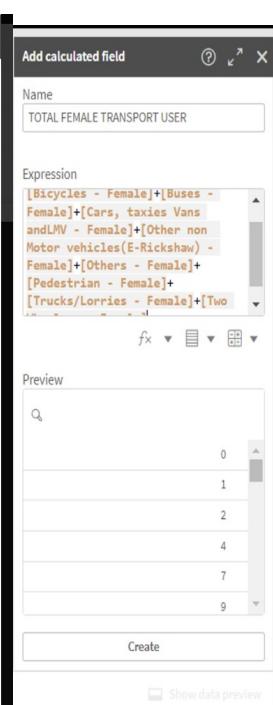
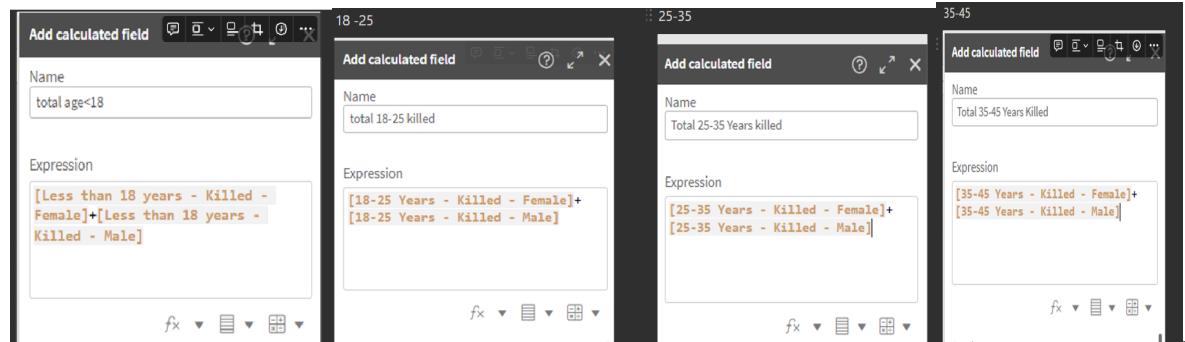


figure 1.1 - This is creating calculating field of Public Transport by using addition expression figure

1.2 - This is creating calculating field of Private Transport by using addition expression

figure 1.3 - This is creating calculating field of Total Male Transport user by using addition expression

figure 1.4- This is creating calculating field of Total Female Transport user by using addition expression



Above all four image :-calculated field is created by using addition Expression

for Less than 18 years, 18-25 Years, 25-35 Years, 35-45 Years, 45-60 Years, 60 and Above, Age not known also such as above ceated

Data Visualization

Data visualization involves the creation of graphical representations to facilitate the comprehension of information. Its primary objective is to render complex data sets more accessible, intuitive, and interpretable. Through the use of visual elements such as charts,

graphs, and maps, data visualization enables the rapid identification of patterns, trends, and outliers within the data. This approach enhances analytical capabilities and supports informed decision-making by presenting data in a clear and concise manner.

Here are the individual visualizations:-

FIRST SHEET -Road Safety And Accident Patterns In India

KPI (SUM) :-

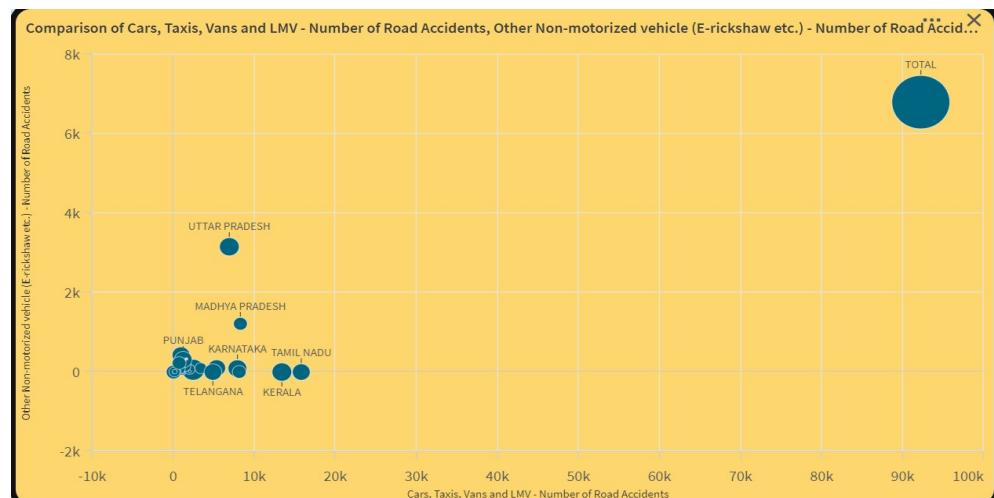
TOTAL TRANSPORT USER



TOTAL PUBLIC TRANSPORT USER



TOTAL PRIVATE USER

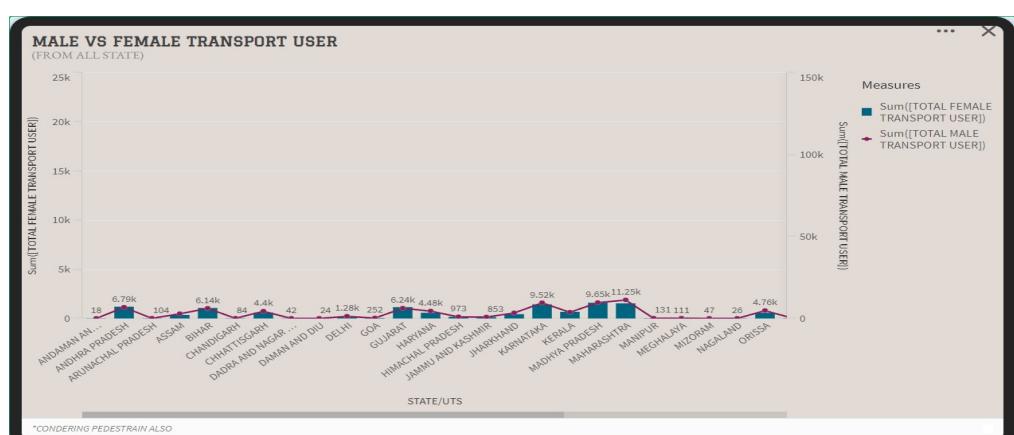


scatter plot

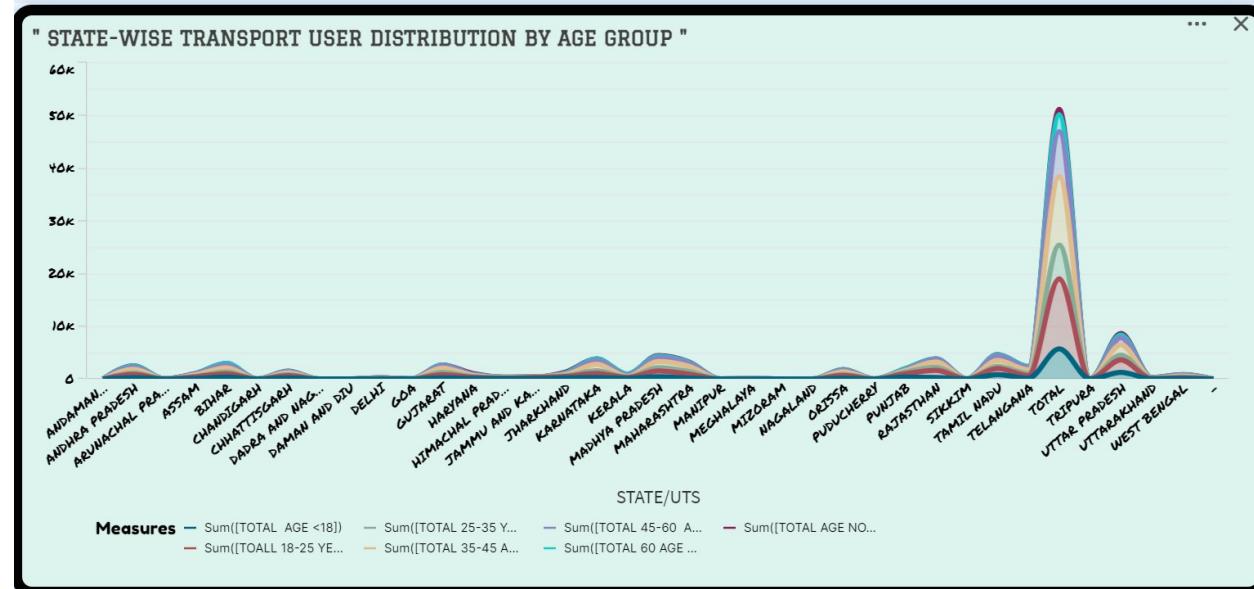


TEXT & IMAGE

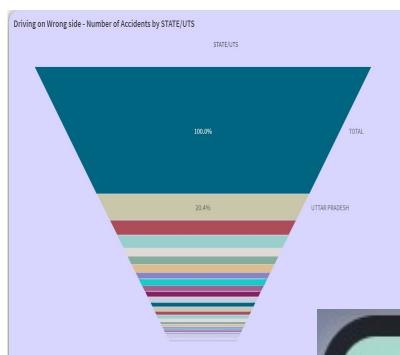
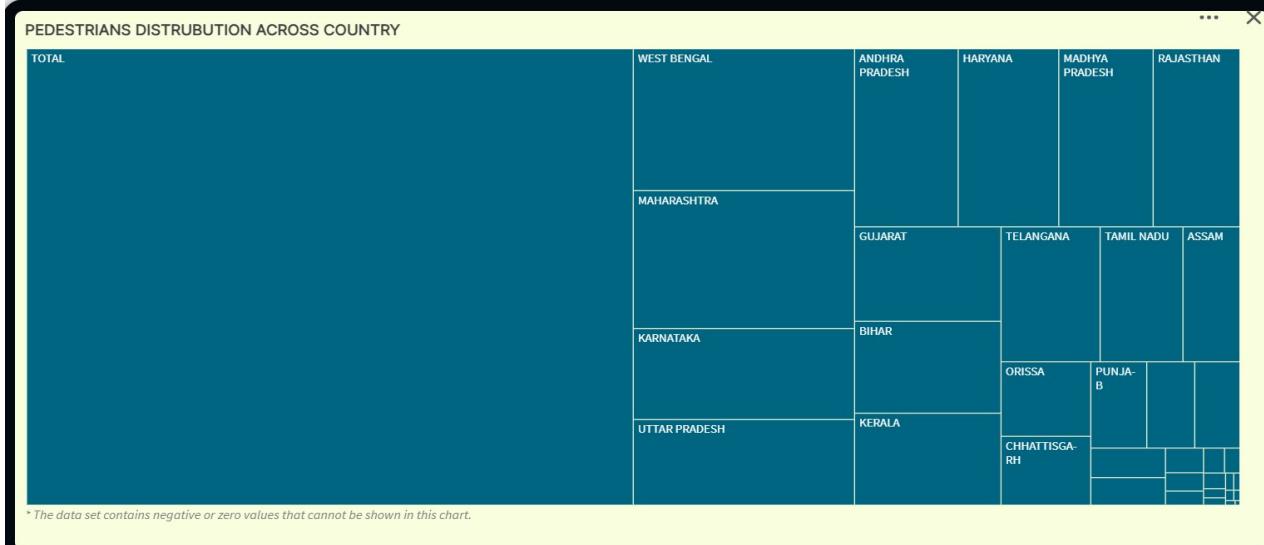
COMBO CHART



LINE CHART



NL
insights



BUTTO



FILTE

PREVIOUS SHEET

NEXT SHEET

Age- distribution of total num...

STATE/UTS

MALE TRANSPORT USER ANALYSIS ACROSS STATE /UTS

CALCULATED MEASURE (KPI)

- The total Sum([TOTAL MALE TRANSPORT USER]) is 258.6k.

RANKING

- The total Sum([TOTAL MALE TRANSPORT USER]) is 258.63k.
- The top STATE/UTS is TOTAL with Sum([TOTAL MALE TRANSPORT USER]) that is 50% of the total.
- 78.9% of Sum([TOTAL MALE TRANSPORT USER]) is represented by top 8 STATE/UTS.

MUTUAL INFORMATION

- The mutual dependence between TOTAL MALE TRANSPORT USER and Total is 99.95% while TOTAL MALE TRANSPORT USER and...

CORRELATION

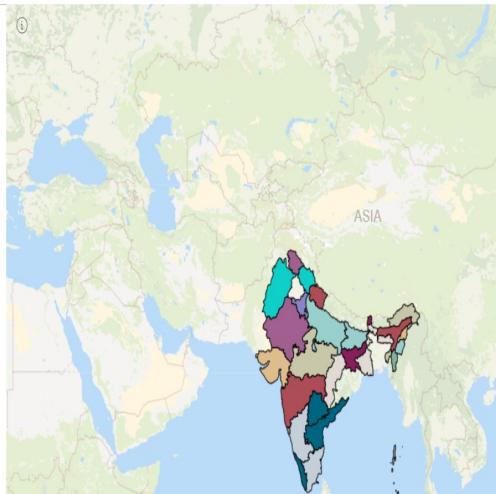


GAUGE

MAP

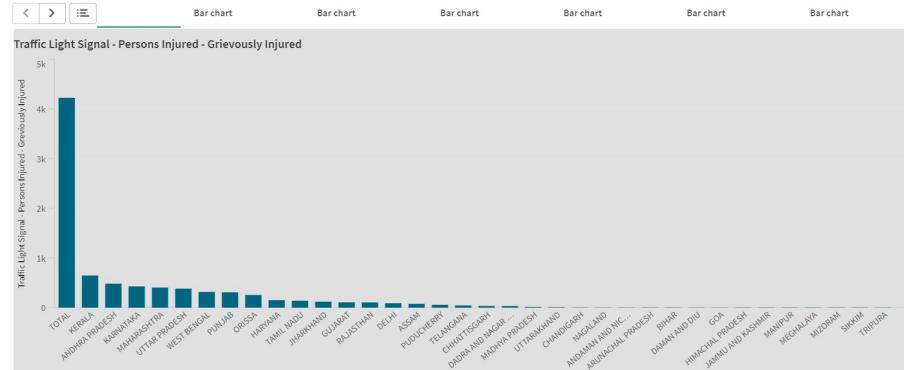
"TRAGIC TOLL: ROAD ACCIDENT FATALITIES AMONG MINORS ACROSS STATES"

STATE/UTS
Area layer
Number of Persons - Minor...

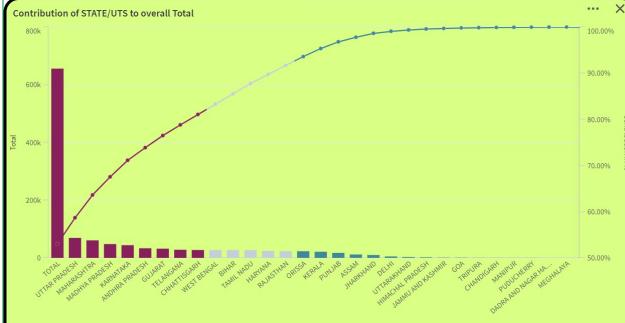


CONTAINER

DETAILED ANALYSIS OF Traffic Light Signal ,Police Controlled ,Stop Signal Causalities Across State/UTs



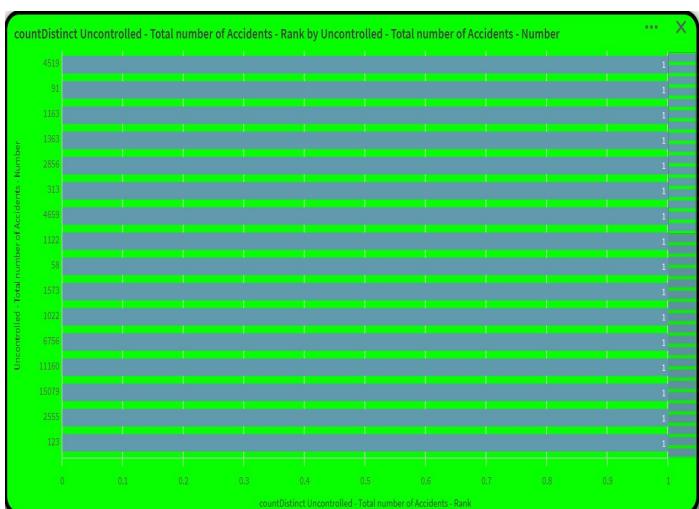
STATE/UTS NAGALAND



Top STATE/UTS by max Traffic Light Signal - Total number of Accidents for Age distribution of total numbers of killed and Age not known - Male

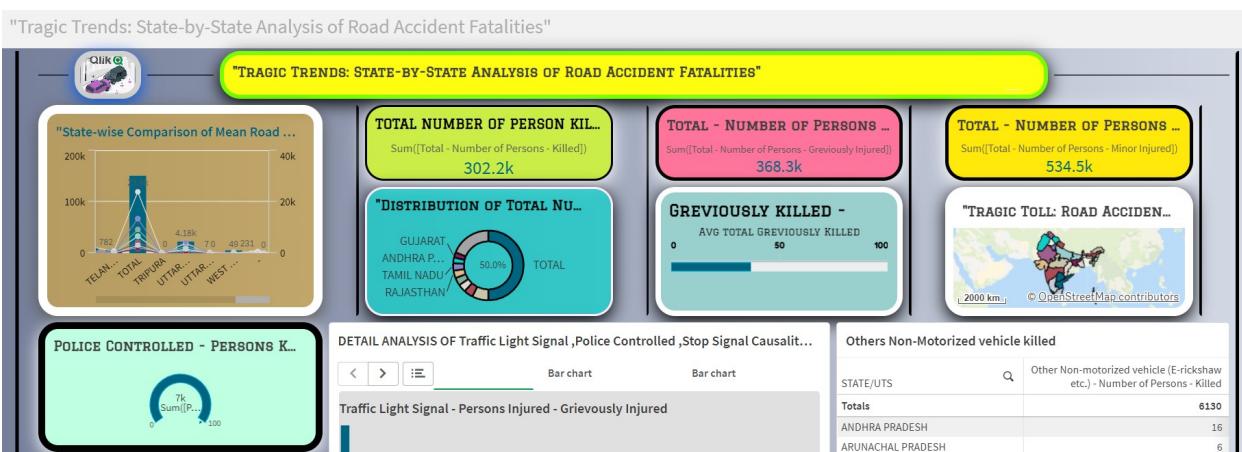
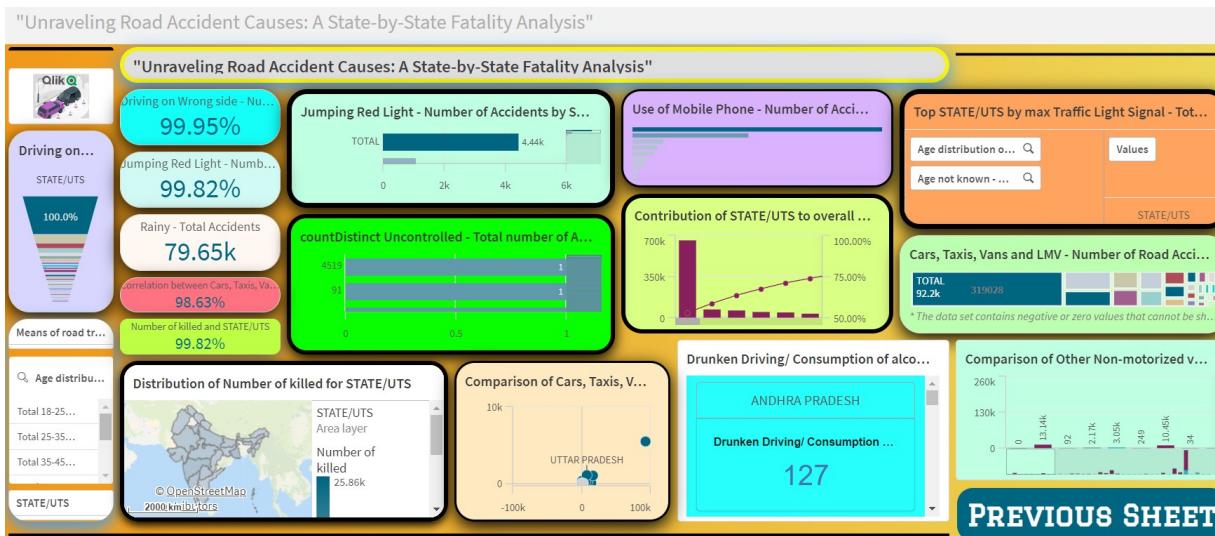
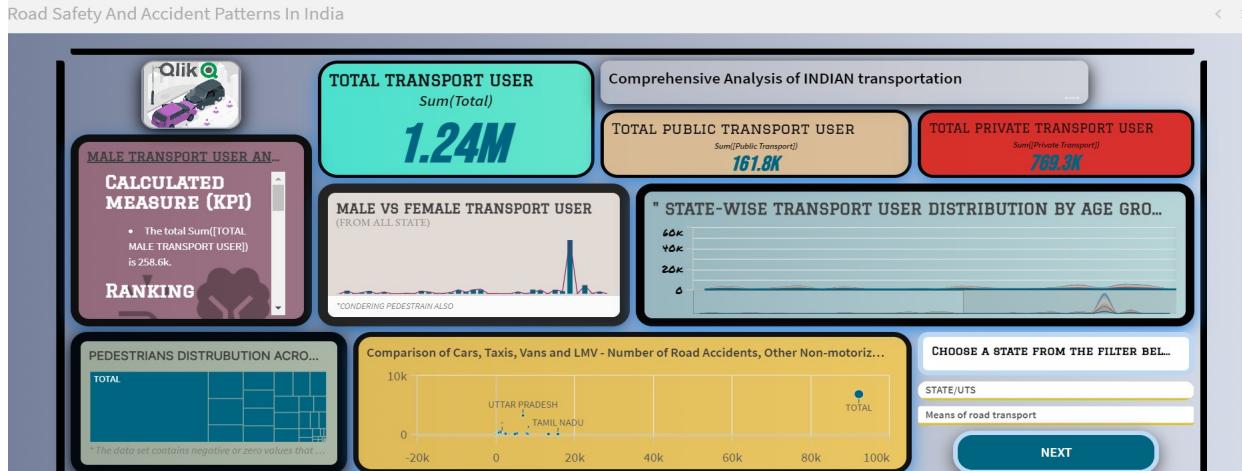
Age distribution o... Q	Values	STATE/UTS	Traffic Light Signal - Total number of Accidents
Age not known ... Q			
ANDHRA PRADESH	TOTAL	ANDHRA PRADESH	1187
Total 18-25 killed	TOTAL		9719
Total 25-35 Years killed	TOTAL		9719
Total 35-45 Years Killed	TOTAL		9719
Total 45-60 Years Killed	TOTAL		9719
Total 60 Years above Killed	TOTAL		9719
Total Age not Known Killed	TOTAL		9719
Total age<18 Killed	TOTAL		9719

TABLE



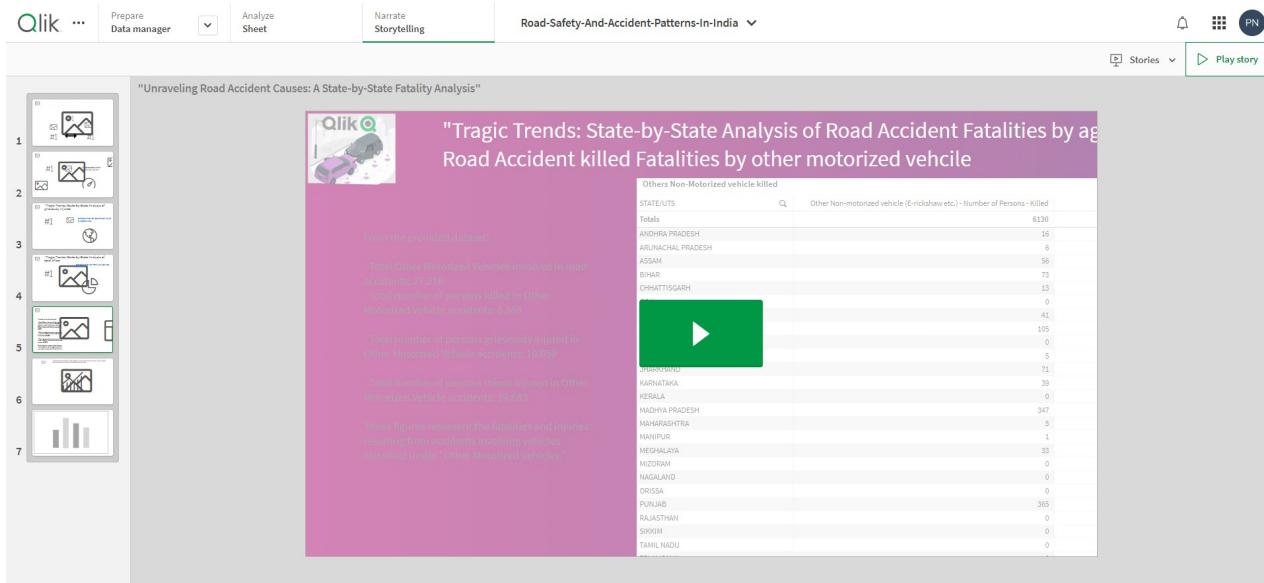
DASHBOARD

A dashboard is a sophisticated graphical user interface (GUI) designed for structured and easily interpretable presentation of data, commonly used for real-time monitoring and analysis. Employed across various sectors, dashboards track key performance indicators (KPIs) and display data through charts, graphs, and tables, thereby facilitating informed decision-making and enhancing operational efficiency.

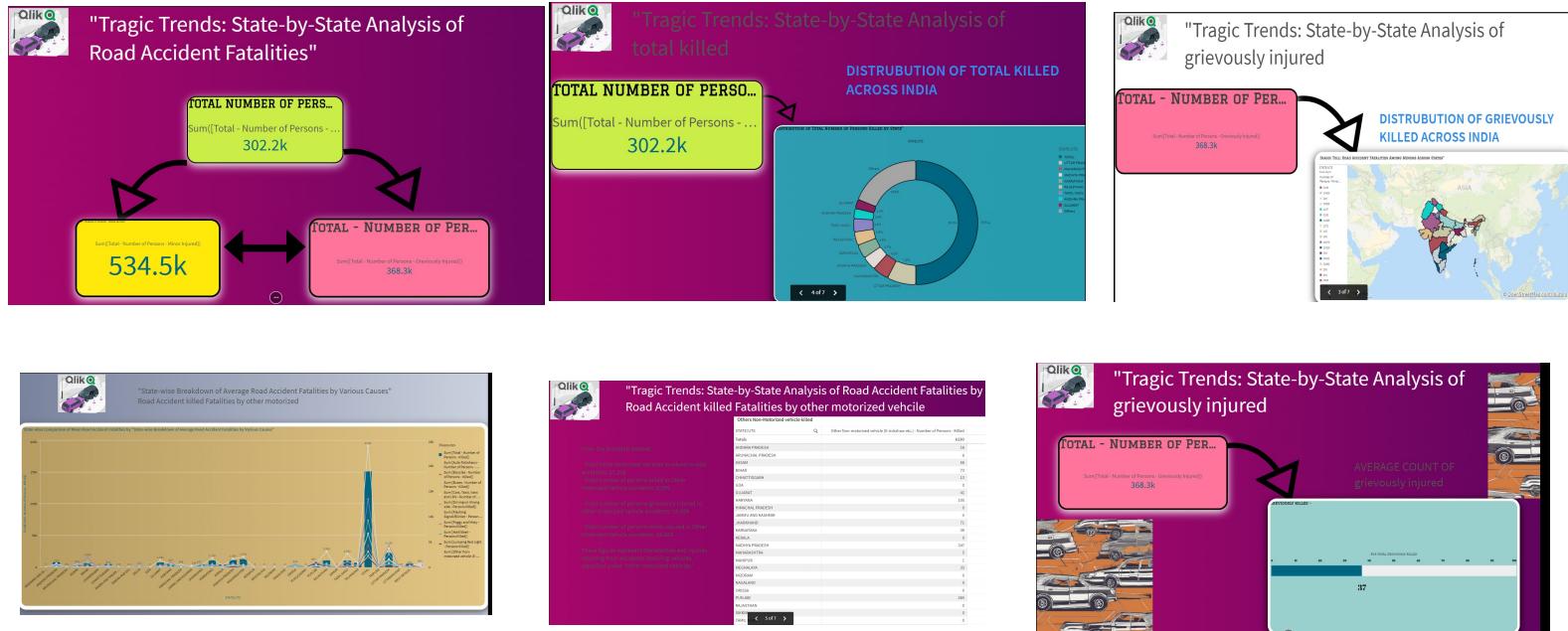


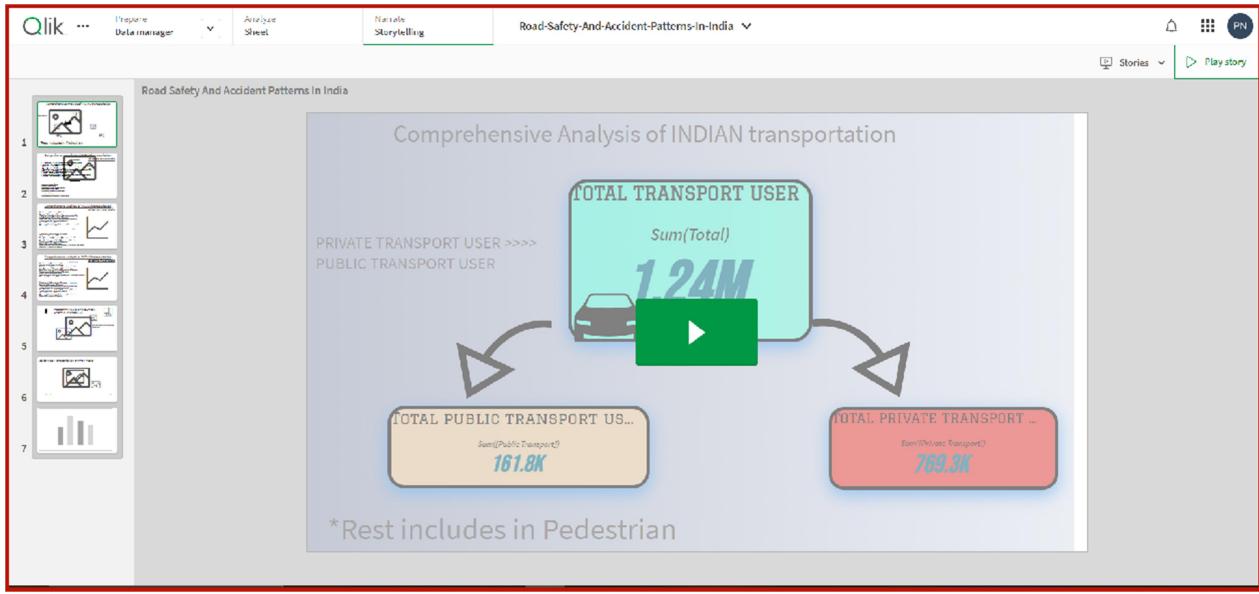
STROYTELLING

storytelling is the art of conveying events, experiences, and messages through narrative. It involves the use of words, images, and sounds to engage an audience, evoke emotions, and communicate ideas in a compelling manner. Effective storytelling captures attention, builds connections, and leaves a lasting impression, making it a powerful tool in various fields such as literature, marketing, education, and entertainment.



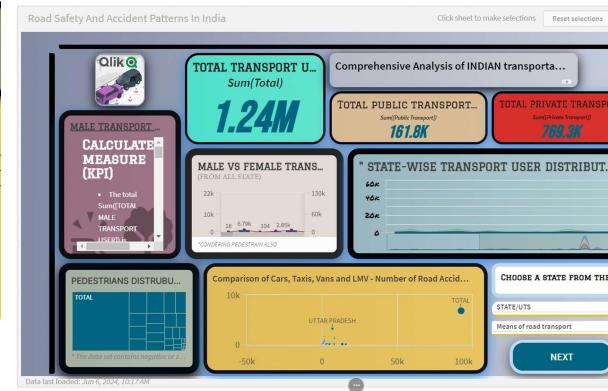
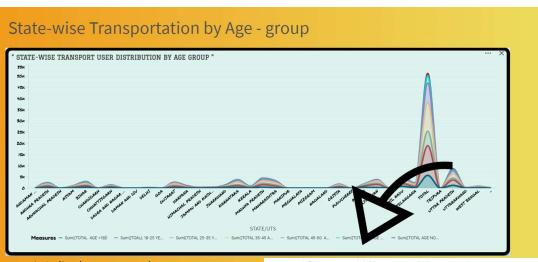
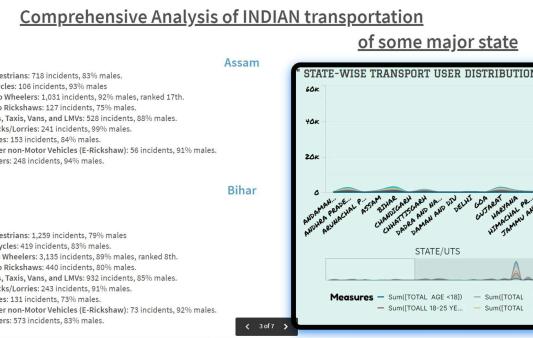
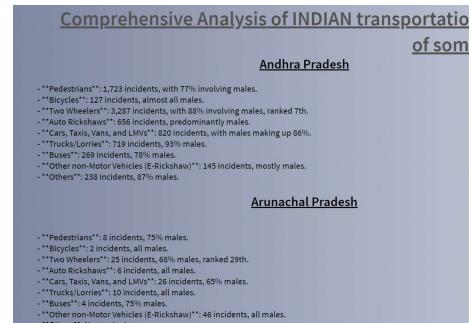
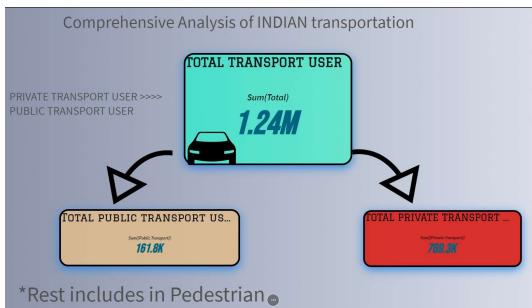
Glimpses OF STROY





STORY 2 :-

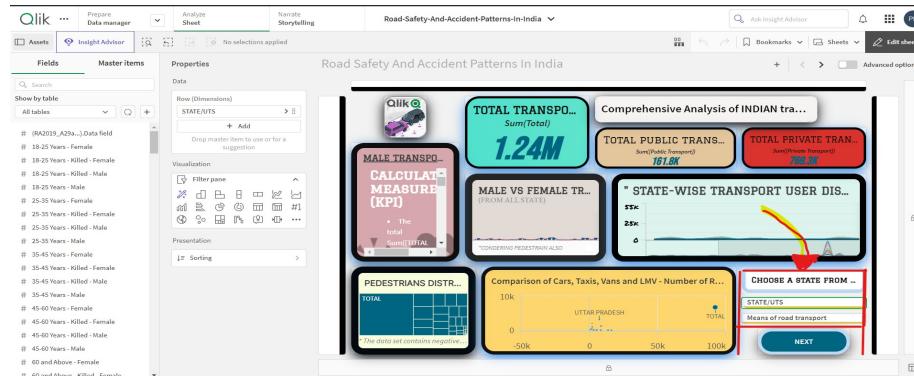
GLIMES OF STORY 2



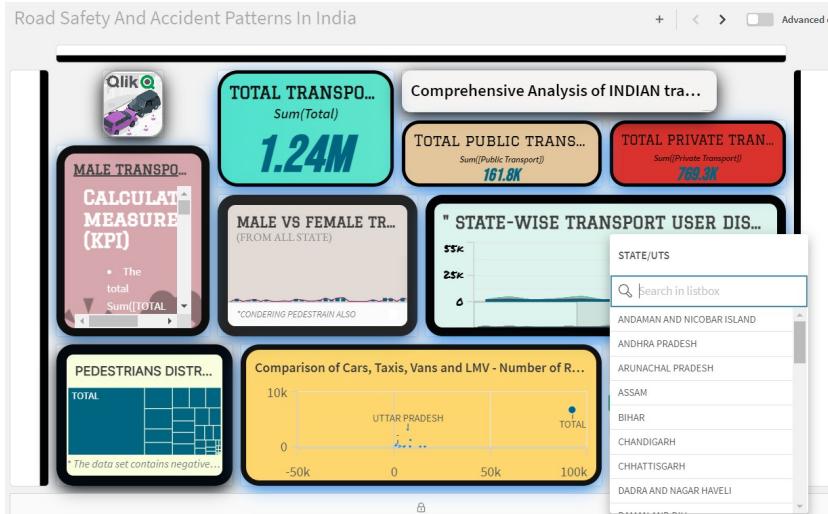
PERFORMANCE TESTING

Application of Filters in Qlik Sense for Road Safety and Accident Pattern Analysis

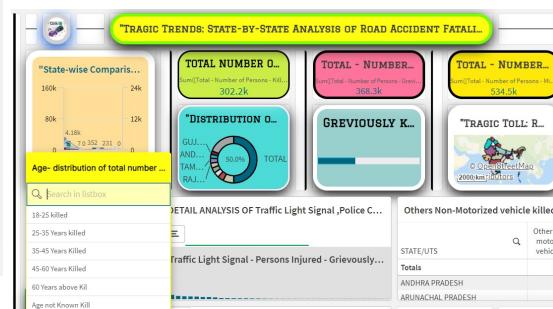
Using Qlik Sense, filters can be effectively applied to analyze road safety and accident patterns. This allows for a comprehensive understanding of the factors contributing to accidents and helps in developing strategies to improve road safety.



Location Filters:



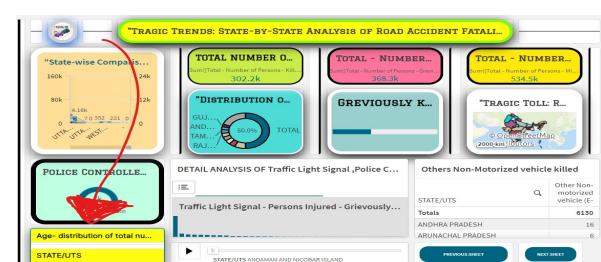
Geographical Areas: Focus on specific regions, cities, or intersections.
Road Type: Compare accident rates on highways, urban streets, and rural
Vehicle Filters:



- **Vehicle Type:** Analyze accidents involving cars, trucks, motorcycles, bicycles, etc.
- **Vehicle Condition:** Assess the impact of vehicle age, maintenance, and safety .

Driver Filters :

- **Driver Age:** Examine accident patterns across different age groups.



Use of Master items / calculated items :-

PUBLIC TOTAL TRANSPORT

Add calculated field

Name: Public Transport

Expression: `[Auto Rickshaws]+[Other Non-Motorized Vehicles(E-rickshaw, etc.)]+Buses`

Preview:

	Total
1	6
2	13
3	1723
4	3352
5	8
6	26
7	1031
8	70
9	1259
10	399
11	3135
12	46
13	36
14	578
15	226
16	665
17	2894
18	25
19	15
20	37
21	10
22	65
23	11
24	92
25	94
26	50
27	186
28	4
29	1291
30	2755

Create

PRIVATE TRANSPORT

Add calculated field

Name: Private Transport

Expression: `Bicycles+[Cars, Taxis, Vans and LMV]+Others+[Trucks/Lorries]+[Two wheelers]`

Preview:

		Total
1		315
2		560
3		1125
4		1217
5		1337
6		1349

Create

CAPITAL :-

USING UPPER EXPRESSION TITLE OF FIELD IS CONVERTED INTO CAPITAL

Add calculated field

Name: STATE/UTS

Expression: `Upper([State/Uts])`

35-45

Add calculated field

Name: TOTAL 35-45 AGE USER

Expression: `[35-45 Years - Female]+[35-45 Years - Male]`

45-60

Add calculated field

Name: TOTAL 45-60 AGE USER

Expression: `[45-60 Years - Female]+[45-60 Years - Male]`

60 ABOVE

Add calculated field

Name: TOTAL 60 AGE ABOVE USER

Expression: `[60 and Above - Female]+[60 and Above - Male]`

DIFFERENT AGE GROUP USER

Add calculated field

Name: TOTAL AGE <18

Expression: `[Less than 18 years - Female]+[Less than 18 years - Male]`

18-25

Add calculated field

Name: TOALL 18-25 YEARS USER

Expression: `[18-25 Years - Female]+[18-25 Years - Male]`

25-35

Add calculated field

Name: TOTAL 25-35 YEARS USER

Expression:

NO Of Graphs/ Visualizations

- 1. Total transport user**
- 2. Total private transport and public transport user analysis**
- 3. Total number of public killed across country**
- 4. Male vs female Public transport user killed**
- 5. Accidents due to Drunken Driving**
- 6. State-wise Mobile Phone Usage**
- 7. Vehicle Contribution towards Total Accidents**
- 8. Correlation - Speeding and Number of accidents**
- 9. Accidents by Weather Type**
- 10. Minors Injured across the country**
- 11. Pedestrians Killed: Gender**
- 12. Pedestrians Killed: Age groups**
- 13. Road Users Killed: Vehicle Distribution**
- 14. Pedestrian killed analysis**
- 15. Rainy season killed visualization**