EXPLORING INDIAN ROAD ACCIDENT DATA TO IMPROVE SAFETY

A PROJECT REPORT

Submitted by

PRADEEPA (22ALR065)

ROSHINI G (22ALR082)

YOGANATH M (22ALR116)

For

22ADF01 DATA ANALYSIS

DEPARTMENT OF ARTIFICIAL INTELLIGENCE

KONGU ENGINEERING COLLEGE
(Autonomous)

PERUNDURAI, ERODE – 638 060

NOVEMBER 2024

DEPARTMENT OF ARTIFICIAL INTELLIGENCE

KONGU ENGINEERING COLLEGE

(Autonomous)

PERUNDURAI ,ERODE – 638 060

NOVEMBER 2024

DEPARTMENT OF ARTIFICIAL INTELLIGENCE

22ADF01 – Data Analysis Project Report

Signature of course in-charge

Signature of the HOD

Submitted for the continuous Assessment viva voice examination held on____

EXAMINER I

EXAMINER II

ABSTRACT

This Power BI project makes an analysis of a comprehensive database of road accidents in India from 2006 to 2016 as factors like location, driver education, age, climate conditions, and accident severity were recorded across all states and cities. The main objective is to dig deeper in meaningful and probable insights into the causes and trends of road accidents with their implications, which can help in making data-informed decisions for improving road safety. There are six structured CSV files that are used in the dataset. This makes available several rows of data to be loaded, cleaned, and organized in Power BI. Using basic Power BI functionality, we analyze and calculate the number of accidents per state, including dangerous locations such as schools, residential areas, or religious locations for both 2014 and 2016. We apply time intelligence functions to analyze the trends observed for the period from 2009 to 2016 and also use DAX in the model to calculate some more accurate metrics, like the average number of injuries per accident or year-on-year percentage change in the number of accidents. The research also looks at the relationship between driver education level and accident rate, which utilizes DAX measures to highlight states with the highest fatal accident rates. For even further granularity, customized measures assess accident severity and filter data to include report on fatalities among drivers aged below 25 years old, which can be put toward identifying demographic segments with increased risks. Reports and visualizations such as pivot tables, line charts, and geographic maps provide clear actionable insights. The project's filters and slicers enable end-users to personalize reports based on the state, age, and accident location. The main purpose of this analysis is to help identify prevailing issues with road safety and also prepare policymakers, planners, and road safety agencies with targeted interventions, better infrastructure, and safer drivers. The project utilizes the powerful feature of data modeling enabled through interactive capabilities of Power BI so as to transform raw data into valuable insights in its pursuit of promoting safety on roads.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.1 INTRODUCTION

- The project leverages Power BI to analyze road accident data from India across various dimensions including location, driver demographics, weather, and accident severity.
- The dataset spans from 2006 to 2016, covering states, Union Territories, and select cities, with data collected across factors like accident type, location, and driver education level.
- Six CSV files serve as the data source, offering detailed records on accident counts, fatalities, and injuries.
- The goal is to extract insights to support safer infrastructure, targeted interventions, and policy changes.
- Power BI's data modeling and visualization tools enable trend analysis over time, helping identify high-risk zones and demographic patterns.
- DAX functions are employed for specific calculations such as average injuries per accident, year-over-year accident change, and accident severity.
- The project provides dynamic reports, including filters for location, age, and accident types, allowing for customized, user-focused exploration.
- Geographic visualizations in Power BI offer insights into regional accident patterns and hotspots, helping direct resources to the most impacted areas.
- Correlation analysis between driver education levels and accident frequency assists in identifying demographic segments with higher risks.
- Analyzing age-based accident data aids in understanding which age groups are most vulnerable, helping target awareness and training efforts.
- The project ultimately informs data-driven decisions for enhancing road safety measures, supporting policymakers, urban planners, and safety agencies in India.risk areas and inform data-driven decisions to improve road safety measures across India.

1.2 DATA COLLECTION

The process of gathering, measuring, and analyzing precise insights for research using accepted,established methods is known as data collection. A researcher can assess their hypothesis using the data that they have gathered. No of the subject of study, gathering data is typically the first and most crucial phase in the research process. Depending on the type of data needed, different disciplines of research require differentapproaches to data gathering.

DATASET

The dataset is collected from various sources related to road accidents in India and includes six tables, each offering detailed information on different aspects of road accidents. The dataset covers multiple years (2006-2016) and provides insights into factors like the type of location, driver education, age, weather conditions, and accident severity. This rich dataset is structured as follows:

1. **Acc_classified_according_to_Type_of_Location_2014_and_2016.csv**

- Number of Columns: 45
- Number of Rows: 38 (data for 36 states/UTs, including 2 additional rows) This file classifies road accidents according to their proximity to various locations such as schools, hospitals, residential areas, and factories. It provides accident data for 2014 and 2016, including the number of accidents, fatalities, and injuries for each location type.

2. **Accidents_Acc_EduQual_Drivers_above_10_Standard_09-16.csv**

- Number of Columns: 9
- Number of Rows: 38 (data for 36 states/UTs, including 2 additional rows) This file contains data on the number of accidents by state, categorized by the education level of the drivers (those above 10th standard). It spans from 2009 to 2016, giving insights into how education correlates with accident rates.

3. **Accidents_Classified_Acc_To_EduQual_Of_Drivers_9-10_standard-09-16.csv**

- Number of Columns: 9
- Number of Rows: 38 (data for 36 states/UTs, including 2 additional rows) Similar to the previous file, this dataset focuses on accidents based on the education qualification of drivers, specifically those with 9th and 10th-grade education levels. It also covers the period from 2009 to 2016.

4. **Due_to_Weather_Condition-2006-15.csv**
   - Number of Columns: 31
   - Number of Rows: 38 (data for 36 states/UTs, including 2 additional rows)
     This file captures data on road accidents occurring due to various weather conditions from 2006 to 2015. It provides accident statistics based on the weather conditions in different states, including details on fatalities and injuries.

5. **No_of_Road_Acc_acco_to_clf_of_age_of_Driver_2014_2016.csv**
   - Number of Columns: 23
   - Number of Rows: 38 (data for 36 states/UTs, including 2 additional rows)
     This dataset categorizes road accidents based on the age of the driver, providing information on accident severity and the number of fatalities and injuries for various age groups (e.g., 15-24, 25-64, 64+).

6. **Road_Accident_Profile_of_Select_Cities-2011-15.csv**
   - Number of Columns: 29
   - Number of Rows: Varies by city (data for select cities, including multiple rows for each city)
     This file provides detailed information on road accidents in select cities across India between 2011 and 2015. It includes data on fatalities, injuries, and accident severity for each year.

This dataset allows for comprehensive analysis of road accidents in India, helping to identify patterns and correlations with various factors such as location, weather conditions, driver education, and age demographics. By analyzing this data, one can gain valuable insights into the primary causes and impacts of road accidents, which can inform targeted interventions, safety policies, and road infrastructure improvements aimed at reducing accidents and fatalities.

DATASET:https://www.kaggle.com/datasets/arindambaruah/indian-road-accidents-data/data

1.3 PROBLEM STATEMENT

Underlying causes of accidents, fatalities, and injuries are derived to understand the crucial features of road accident analysis. A high-risk zone area with determined high-risk vulnerable groups can be identified considering factors such as location, weather, age demographics of drivers, and educational levels, making an amendment in interventions, road safety policies, and infrastructure optimization lead to reducing accidents and saving lives. Data-driven insights have thereby been used in decision-making to promote safer roads.

1.4 BUSINESS OBJECTIVE

1. To analyze road accident trends across different states/UTs and visualize them using Power BI.
2. To compare the number of accidents, fatalities, and injuries in various locations such as schools, hospitals, and residential areas in 2014 and 2016.
3. To identify the impact of weather conditions on road accidents and analyze yearly trends from 2006 to 2015.
4. To assess the relationship between driver education level and accident rates, providing insights for targeted road safety initiatives.
5. To analyze the accident data by age groups, identifying high-risk drivers and proposing safety measures for young and elderly drivers.

ABOUT POWER BI:

1. Power BI is a business analytics tool from Microsoft that helps users create interactive visualizations and share insights across an organization.
2. It allows business users to track key metrics and goals in one place, making it easy to collaborate and stay on top of performance.
3. With busy schedules, people often lack time for entertainment or detailed cricket analysis, so summarizing events in a simple and attractive way is necessary.
4. Power BI helps create dynamic and interactive dashboards to visualize cricket match data in an easy-to-understand way.
5. Power BI simplifies decision-making by offering a range of interactive visualizations and business intelligence features.

CHAPTER 2

DATA PREPARATION AND MODELING

2.1 DATA CLEANING

Data cleaning is the process of removing errors from the data by filling in missing values, smearing noisy data, analyzing and removing outliers, and smoothing noisy data. Data at various degrees of detail mayoccasionally diverge from what is needed Missing Values – Appropriate values are substituted for missing values. The strategies listed below

- When a tuple contains many attributes with empty values, it is
- Disregarded.
- For the missing value, the values are manually filled in.
- The values may be filled with the same global constant.
- The attribute mean can replace the values that are absent.
- The most likely value can be used to fill in the blanks.

2.2 DATA TRANSFORMATION

1. The process of changing data from one format or structure to another is known as data transformation. It is a crucial component of the majority of data management and integration jobs, including

2. Application **integration**, Data **wrangling**, Data **warehousing**, and Data **integration**. Depending on the required modifications to the data between the source (initial data) and the destination (final data), data transformation can be straightforward or difficult. The process of data transformation often involves both manual and automated procedures.

3. Depending on the format, structure, complexity, and amount of the data being changed, a broad range of tools and technologies may be employed. For decades, corporations have benefited greatlyfrom using conventional data transformation techniques.

4. Since the development of the various tools and technologies (data profiling, data visualization, datapurification, data integration, etc.), most (if not all) businesses now transform massive volumes of data that feed internal and external applications, data warehouses, and other data repositories.

5. So, Data Transformation is a required process inorder to preprocess the loaded data set as per our requirement and apply those changes for future use. It is while Data Analysis and creating DAX functions of those relations respectively.

PROCEDURE

STEP 1

1. Go to HOME tab in ribbon.
2. Click on GET DATA and select data from the system or from any platform where it resides.
3. Here select 6 different tables of CSV format from system and load it to POWER BI.

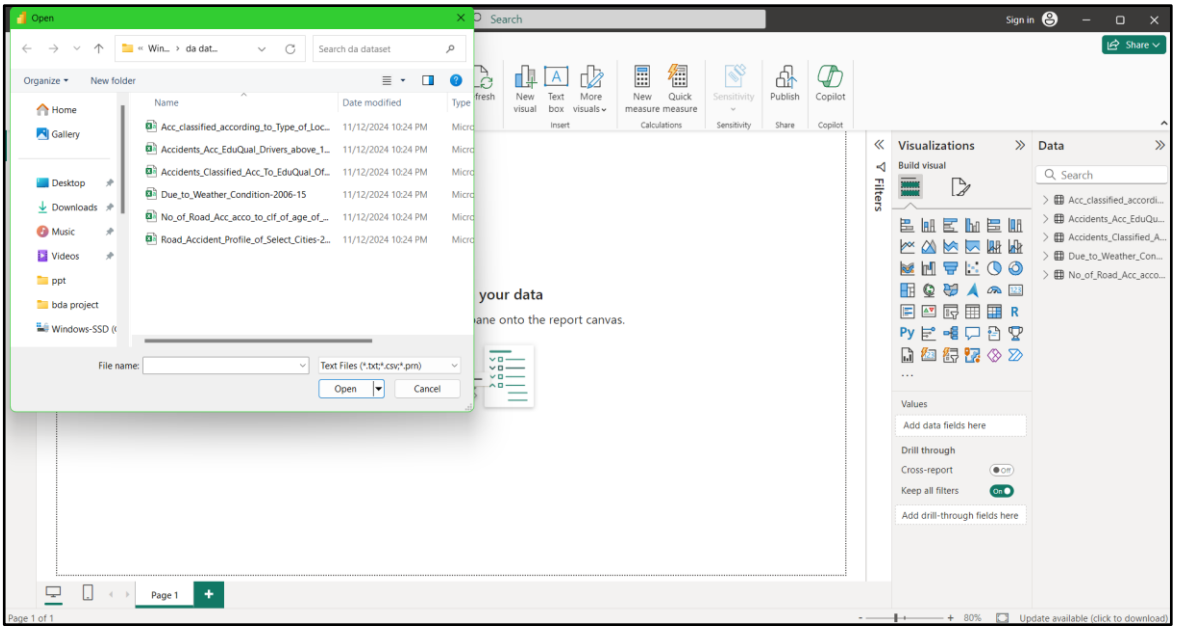


Figure 2.2.1 Select Dataset

STEP 2

1. From the ribbon of HOME tan select TRANSFORM DATA inorder to clean and transform data.

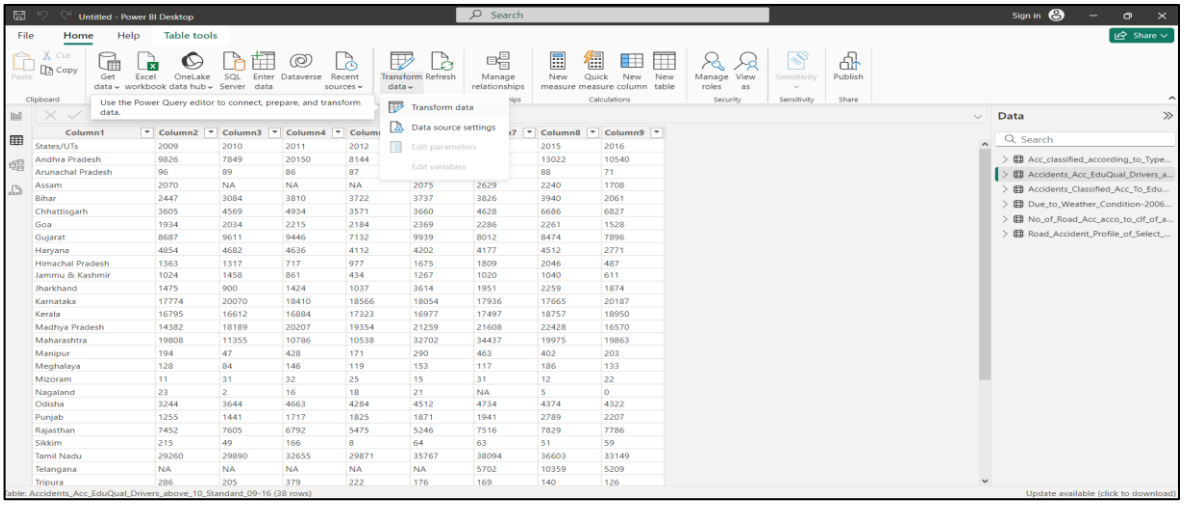


Figure 2.2.2Transform Data

STEP 3

1. After choosing transforming data all the loaded tables and opened in POWER QUERY EDITOR, sothat we can make any changes as per our wish.
2. Then open the SUMMARY table and replace the values which are blank.
3. Then try to add NULL values to the rows in which the matches are cancelled due to some reasons

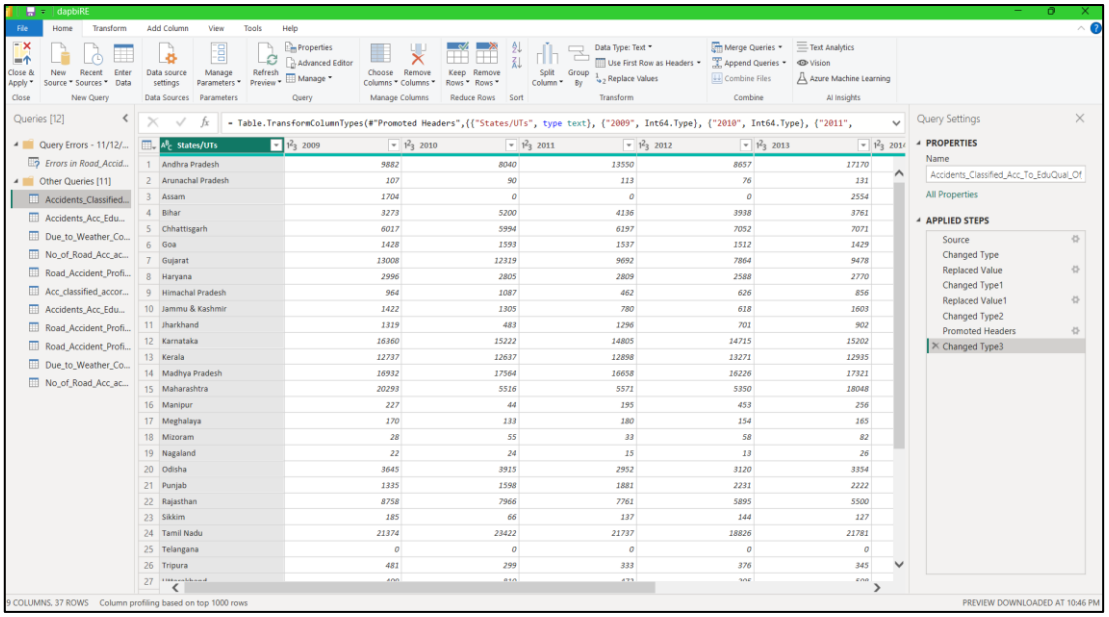


Figure 2.2.3 Power Query Editor

STEP 4

1. Then on same SUMMARY table apply REPLACE VALUES.
2. In this select any column that need new values to be replaced for further processing

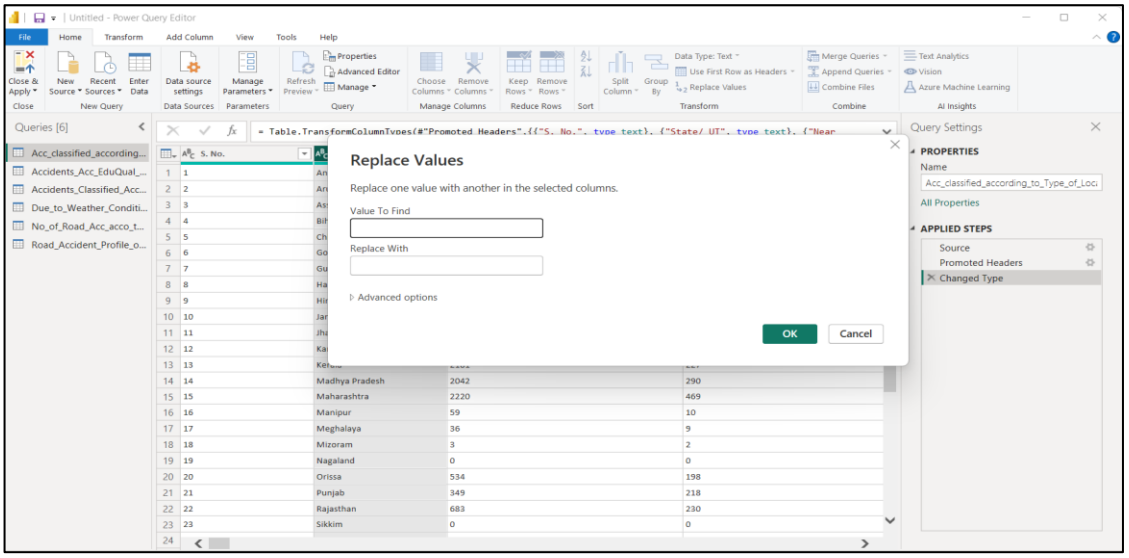


Figure 2.2.4 Replace Value

STEP 5

1. Now select acc_classified_accordind_to_type_of_loc table to clean it.
2. Here applies change datatype so click the column that needed to change the datatype.
3. Select “Near School or College-Total Acc-2024” column then change its datatype to minutes.

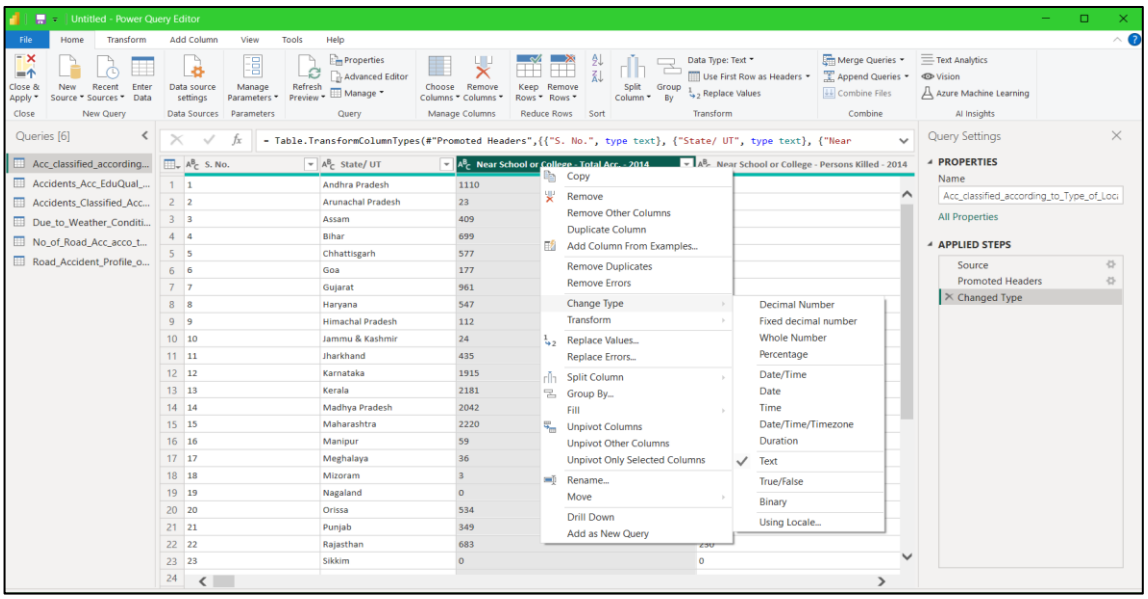


Figure 2.2.5 Change Column Type

STEP 6

1. Select Accidents_Acc_EduQual_Drivers_above_10_Standard_09-16 table to unpivot column
2. Select all year column and apply unpivot for better visualization

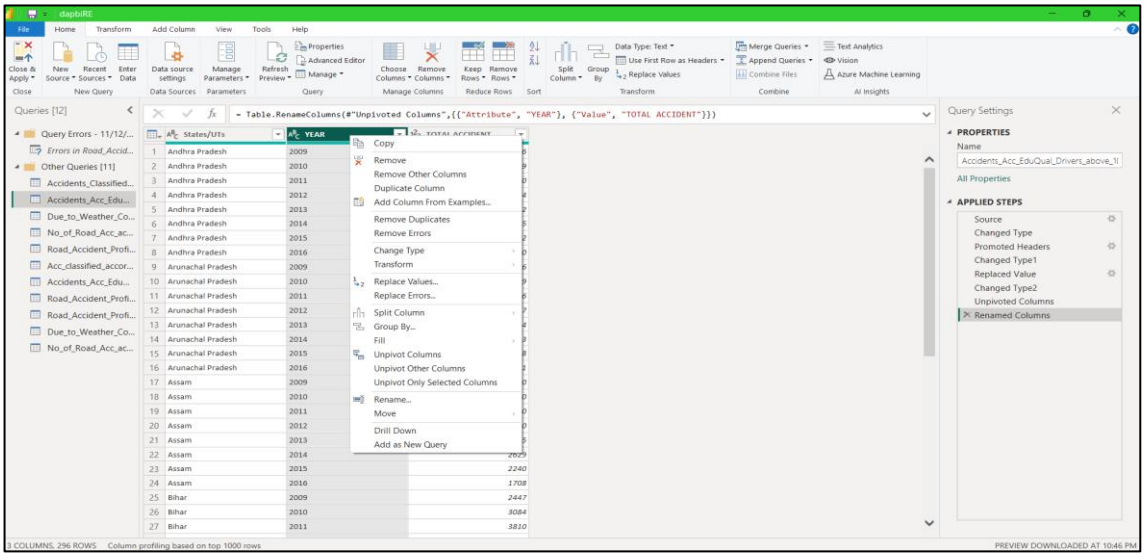


Figure 2.2.6 Unpivot Column

STEP 7

1. Change first row as Column name in Accidents_Acc_EduQual_Drivers _above_ 10_Standard_09-16
2. Select Use first row as Headers to change it

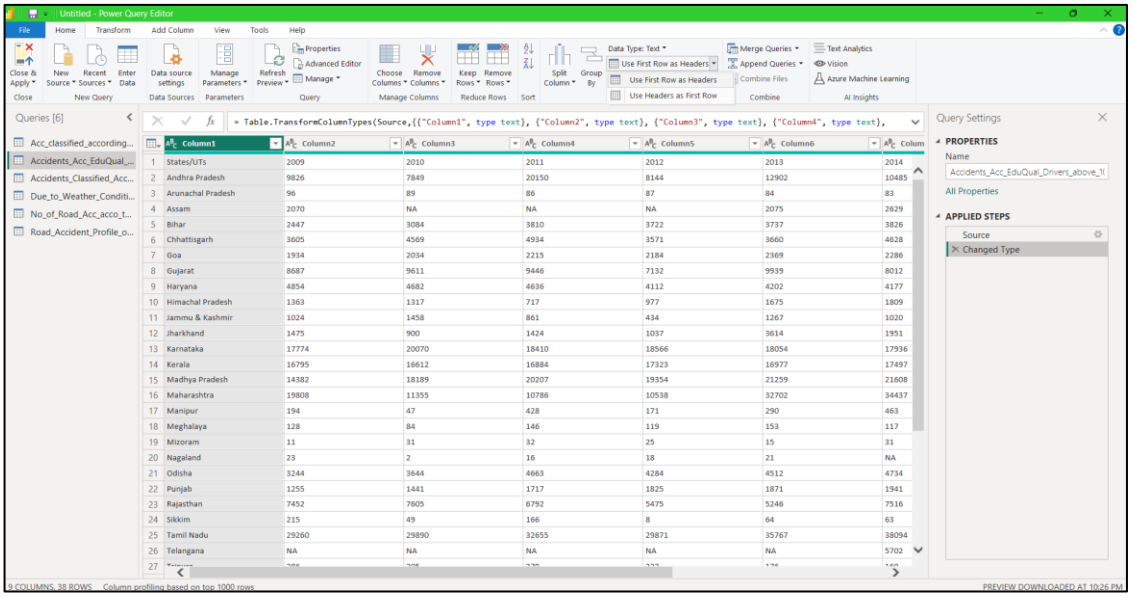


Figure 2.2.7 Change column name

STEP 8

1. Select each table and replace NA values.

2. Replace NA to 0 values.

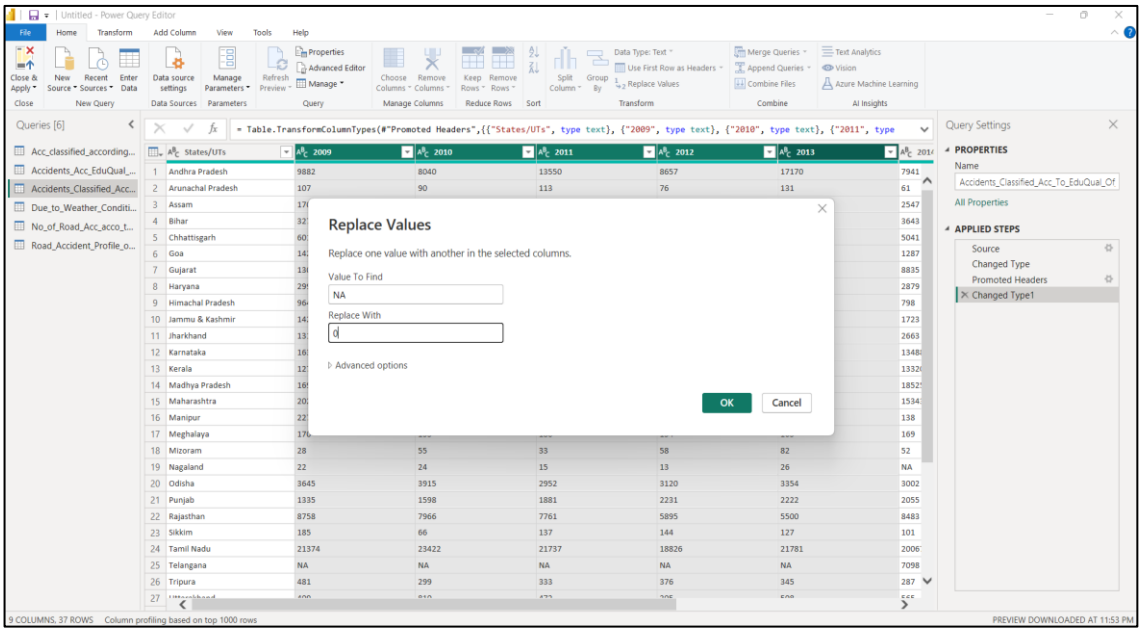


Figure 2.2.8 Replace NA values

STEP 9

1. Create new column to calculate total accident
2. Select add new column and select custom column to create it.

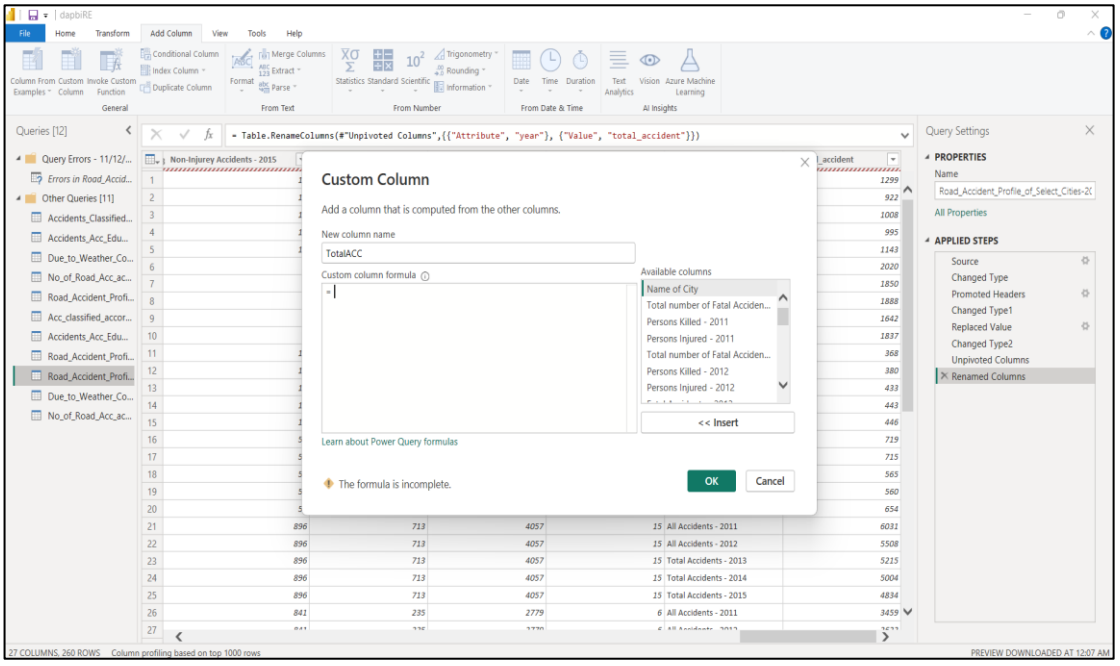


Figure 2.2.9 Create Custom Column

2.3.DATA MODELLING

Data modelling is one of the aspects used in BI tools to establish relationships between various data sources. When using several data sources, you can construct engaging data visualizations by defining the relationships between them.

It can create unique calculations on the already-existing tables using the modelling capability, and thesecolumns can then be easily displayed in Power BI visualizations. This enables companies to create new measures and perform unique calculations for them.

Data Modeling is used to create relationship among the different tables inorder to access the data of different tables to visualize them. There are four types of relations that we can create as,

- ❖ One to One relationship
- ❖ One to Many relationship
- ❖ Many to One relationship
- ❖ Many to Many relationship

PROCEDURE
STEP 1

1. Create One to One relationship.
2. Select Acc_classified_according_to_Type_of_Location_2014_and_2016 and Accidents_Acc_EduQual_Drivers_above_10_Standard_09-16 to create one to one relationship.

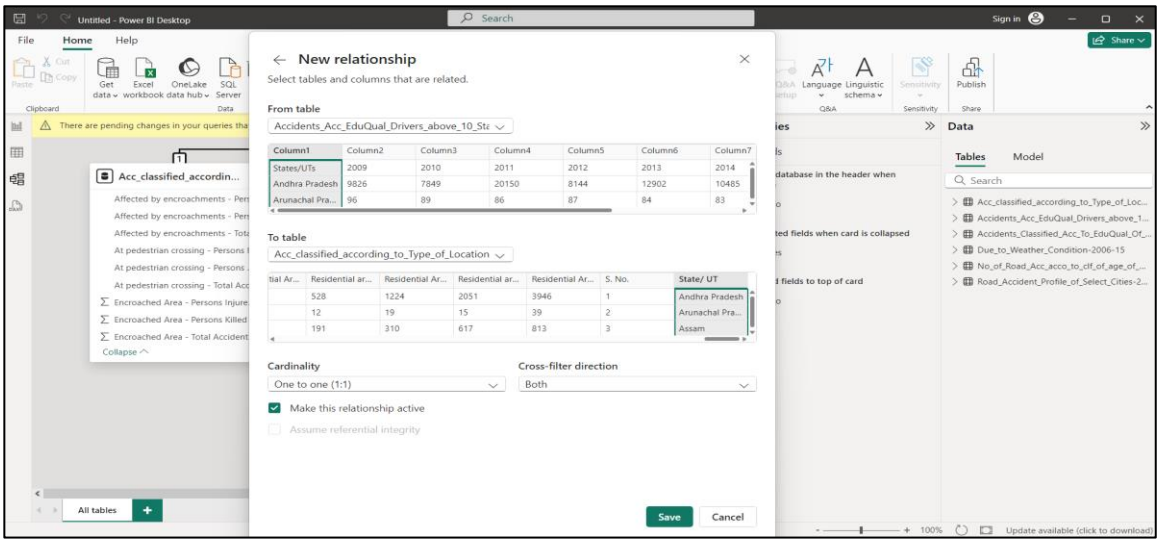


Figure 2.3.1 One to One relationship

STEP 2

1. Create Many to One relationship.
2. Select No_of_Road_Acc_acco_to_clf_of_age_of_Driver_2014_2016and Accidents_Classified_Acc_To_EduQual_Of_Drivers_9-10_standard-09-16 to create Many to one relationship.

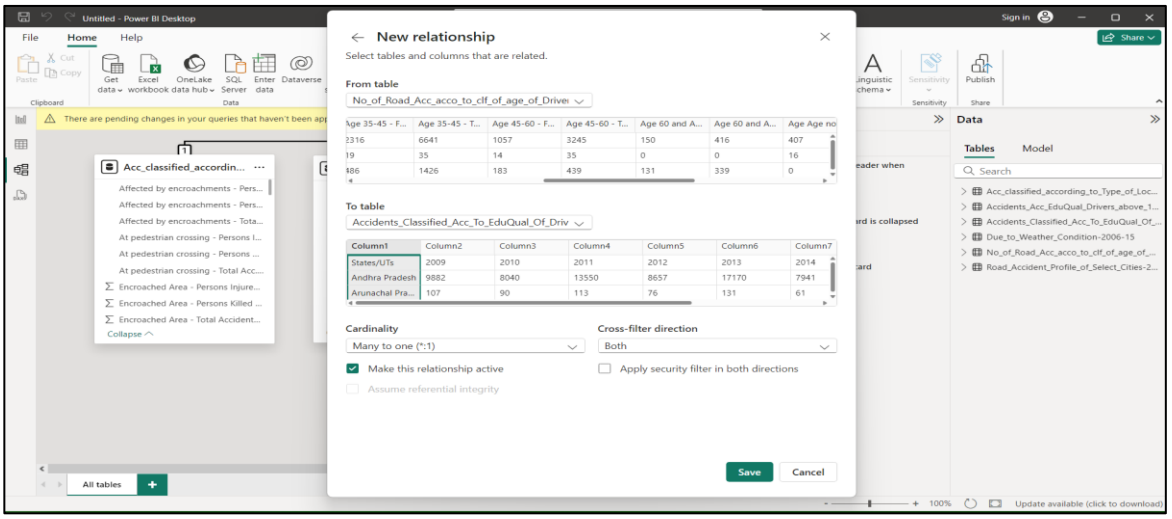


Figure 2.3.2 Many to One relationship

STEP 3

1. Create One to Many relationship.
2. Select Road_Accident_Profile_of_Select_Cities-2011-15 and Acc_classified_according_to_Type_of_Location_2014_and_2016to create one to one relationship.

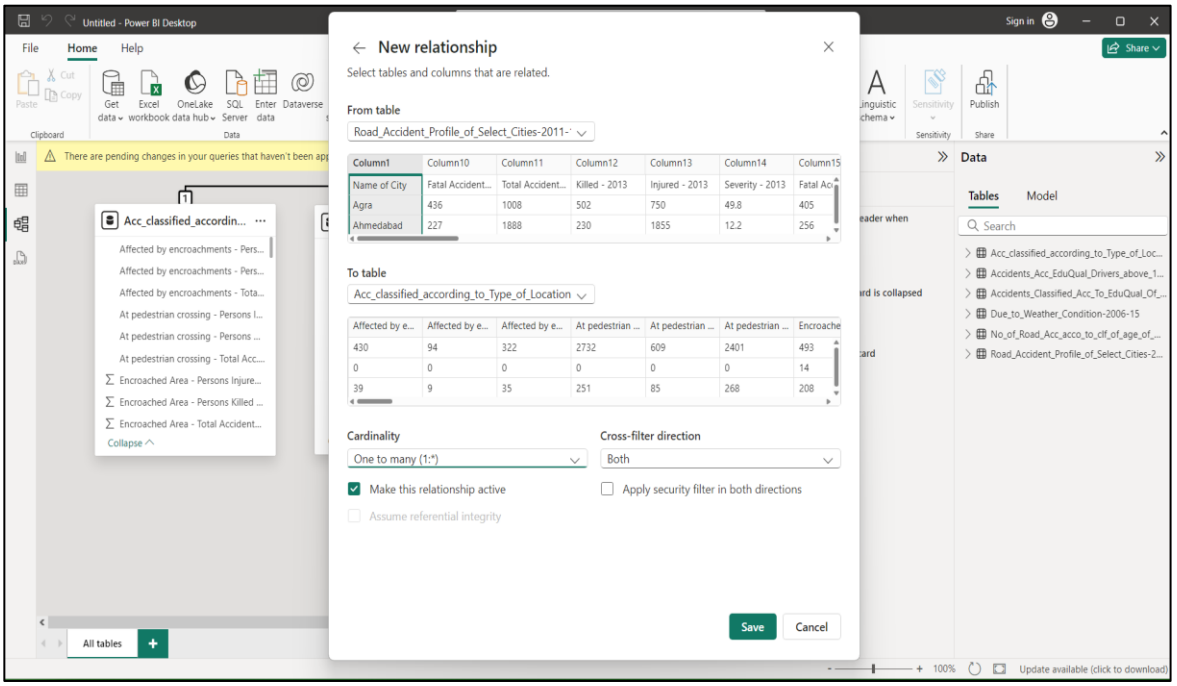


Figure 2.3.3 One to Many relationship.

STEP 4

1. Create all the necessary relationship between the tables

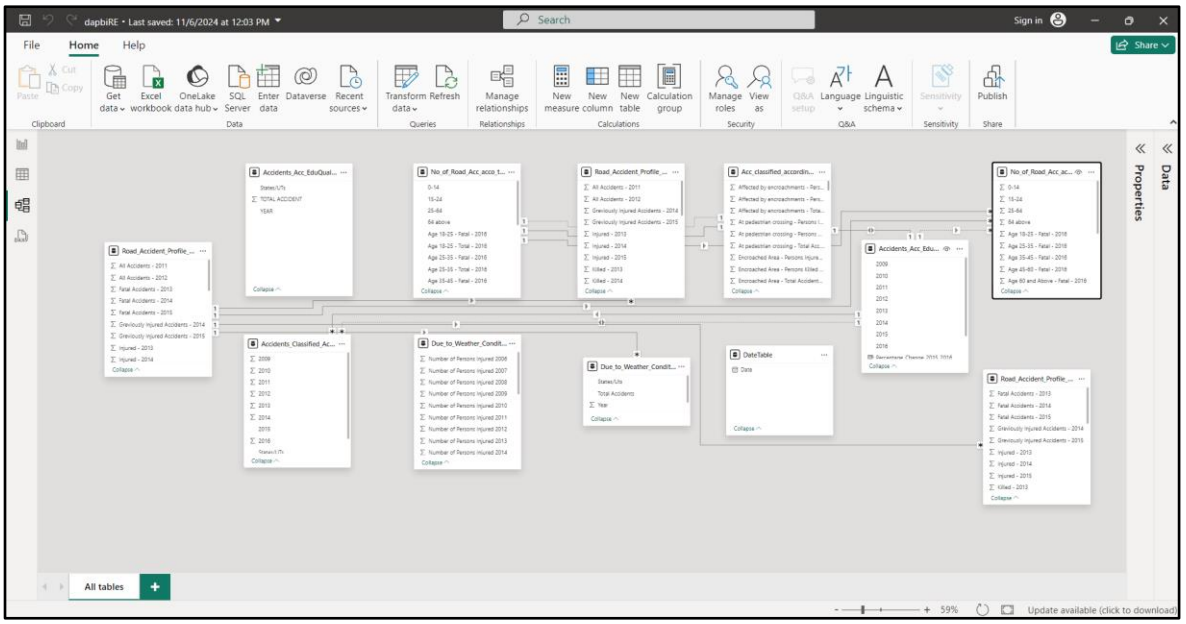


Figure 2.3.4 Model View

2.4 DAX (Data Analysis Expressions)

DAX is a special function that contains collection of operators, formulae, functions, expressions to calculate, process and execute the values from existing table and return one or more values as the result of respective functions. So, it is used to create new information from the data that already exist in the table whilecreating model and analyzing it.

DAX measured of Power Bi are special functions or Programming Language that are used to create the following such as

- Calculated columns
- New measures
- Customized tables
- Quick measures
- Implement Time Intelligence

There exist many formulae for creating the new columns, measures. The time intelligence are specialfunctions the are applicable only for the Time-based columns only.

So, from these formulae and expression we can find results like maximum, minimum, average, count,sum, filters, difference, total, variance, percentage, addition, subtraction, division, etc.….

STEP 1

1. Creating Quick measure for table Summary.
2. Click Quick measure at ribbon and a menu pop up
3. Measure named “Total_Accidents_2016” and “Total_Accidents_2014” for “Acc_classified_according_to_Type_of_Location_2014_and_2016” and click on Enter.

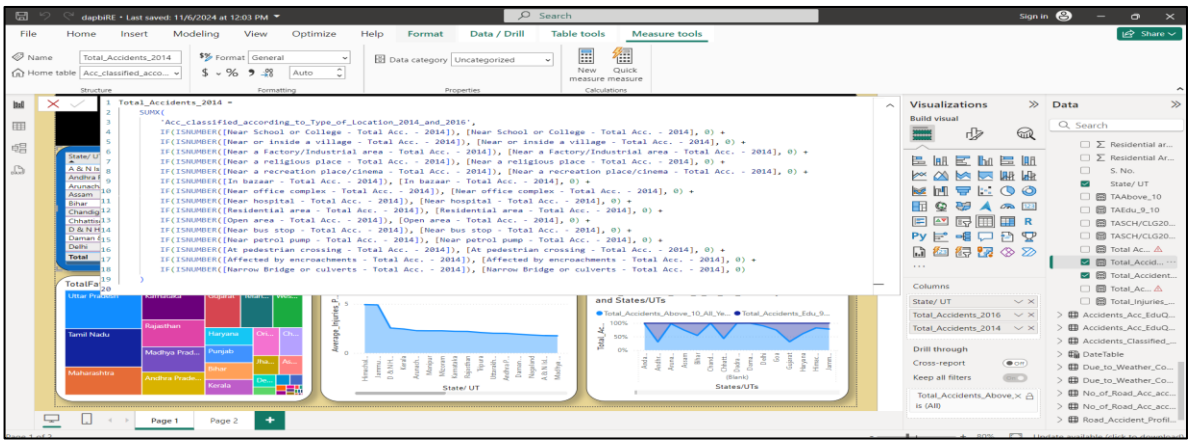


Figure 2.4.1 Quick measure “Total_Accidents”

STEP 2

1. Create New measure to calculate total accidents near schools or college.
2. Select table then select “New measure”.
3. Apply DAX Measure to calculate “TASCH/CLG2014” and “TASCH/CLG2016”.

DAX MEASURE:

TASCH/CLG2014 =
SUM('Acc_classified_according_to_Type_of_Location_2014_and_2016'[Near School or College - Total Acc. - 2014])

TASCH/CLG2016 =
SUM('Acc_classified_according_to_Type_of_Location_2014_and_2016'[NearSchool/College/ any other educational Institutes - Total Accidents - 2016])

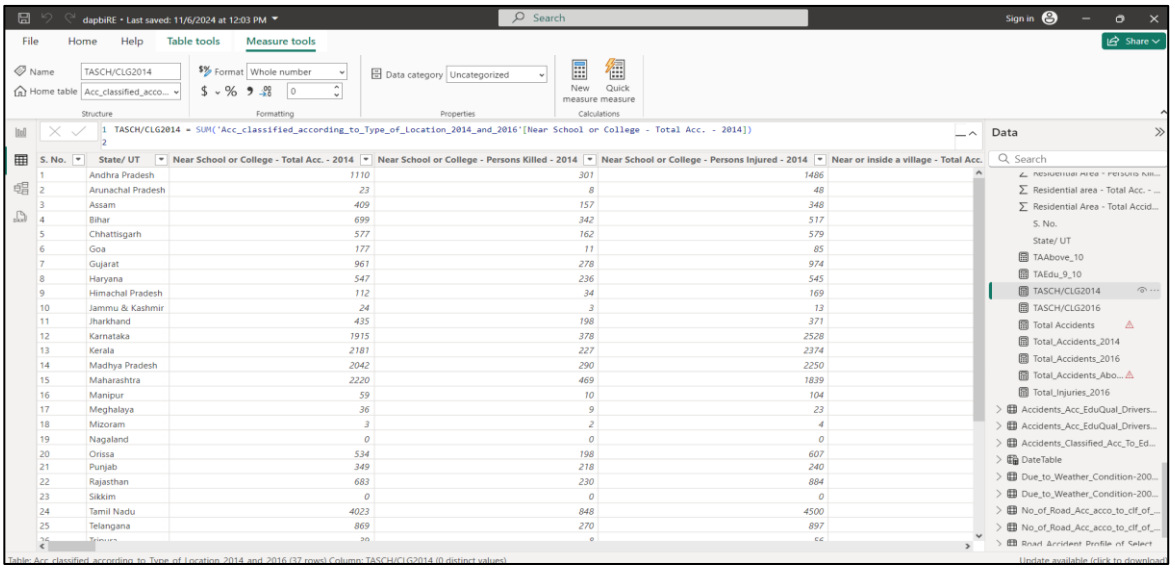


Figure 2.4.2 New Measure for Accident near sch/clg

STEP 3

1. Create New measure to calculate Total Fatalities 2016 .
2. Select table and then select “New measure”.
3. Apply DAX Measure to calculate “TotalFatalities2016”

DAX MEASURE:

TotalFatalities2016 = SUM('No_of_Road_Acc_acco_to_clf_of_age_of_Driver_2014_2016'[Age Less than 18 years - Fatal - 2016]) +

SUM('No_of_Road_Acc_acco_to_clf_of_age_of_Driver_2014_2016'[Age 18-25 - Fatal - 2016]) +
SUM('No_of_Road_Acc_acco_to_clf_of_age_of_Driver_2014_2016'[Age 25-35 - Fatal - 2016]) +
SUM('No_of_Road_Acc_acco_to_clf_of_age_of_Driver_2014_2016'[Age 35-45 - Fatal - 2016]) +
SUM('No_of_Road_Acc_acco_to_clf_of_age_of_Driver_2014_2016'[Age 45-60 - Fatal - 2016]) +
SUM('No_of_Road_Acc_acco_to_clf_of_age_of_Driver_2014_2016'[Age 60 and Above - Fatal - 2016]) +
SUM('No_of_Road_Acc_acco_to_clf_of_age_of_Driver_2014_2016'[Age Age not known - Fatal - 2016])

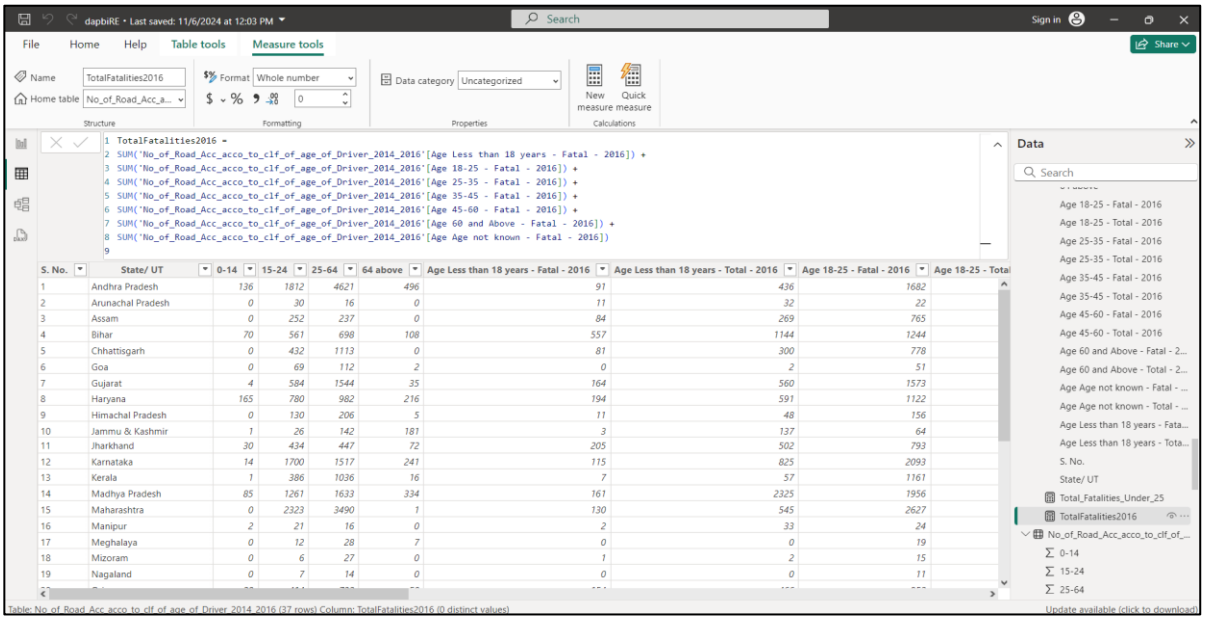


Figure 2.4.3 New Measure created

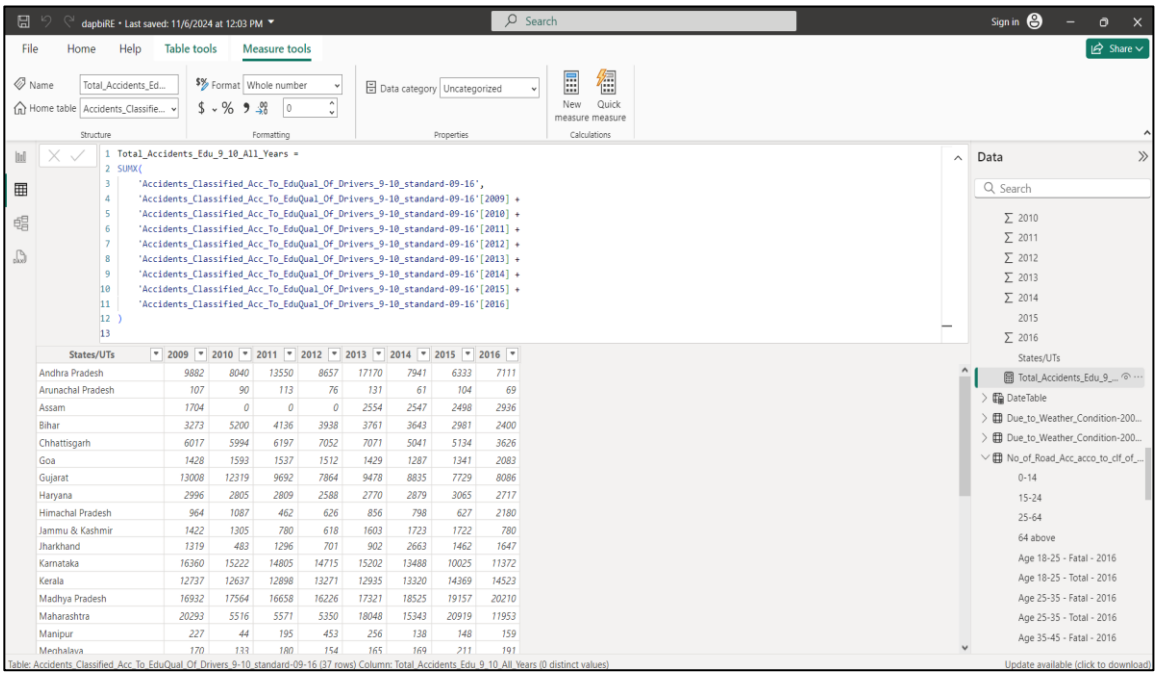
STEP 4

1. Create New measure to calculate Total Accidents of education qual of drivers 9-10.
2. Select table and then select “New measure”.
3. Apply DAX Measure to calculate “Total_Accidents_Edu_9_10_All_Years” and “Total_Accidents_Above_10_All_Years”

DAX MEASURE:

Total_Accidents_Edu_9_10_All_Years =
SUMX(
'Accidents_Classified_Acc_To_EduQual_Of_Drivers_9-10_standard-09-16',
'Accidents_Classified_Acc_To_EduQual_Of_Drivers_9-10_standard-09-16'[2009] +
'Accidents_Classified_Acc_To_EduQual_Of_Drivers_9-10_standard-09-16'[2010] +

'Accidents_Classified_Acc_To_EduQual_Of_Drivers_9-10_standard-09-16'[2011] +
'Accidents_Classified_Acc_To_EduQual_Of_Drivers_9-10_standard-09-16'[2012] +
'Accidents_Classified_Acc_To_EduQual_Of_Drivers_9-10_standard-09-16'[2013] +
'Accidents_Classified_Acc_To_EduQual_Of_Drivers_9-10_standard-09-16'[2014] +
'Accidents_Classified_Acc_To_EduQual_Of_Drivers_9-10_standard-09-16'[2015] +
'Accidents_Classified_Acc_To_EduQual_Of_Drivers_9-10_standard-09-16'[2016]
)



States/UTs	2009	2010	2011	2012	2013	2014	2015	2016
Andhra Pradesh	9882	8040	13550	8657	17170	7941	6333	7111
Arunachal Pradesh	107	90	113	76	131	61	104	69
Assam	1704	0	0	0	2554	2547	2498	2936
Bihar	3273	5200	4136	3938	3761	3643	2981	2400
Chhattisgarh	6017	5994	6197	7052	7071	5041	5134	3626
Goa	1428	1593	1537	1512	1429	1287	1341	2083
Gujarat	13008	12319	9692	7864	9478	8835	7729	8086
Haryana	2996	2805	2809	2588	2770	2879	3065	2717
Himachal Pradesh	964	1087	462	626	856	798	627	2180
Jammu & Kashmir	1422	1305	780	618	1603	1723	1722	780
Jharkhand	1319	483	1296	701	902	2663	1462	1647
Karnataka	16360	15222	14805	14715	15202	13488	10025	11372
Kerala	12737	12637	12898	13271	12935	13320	14369	14523
Madhya Pradesh	16932	17564	16658	16226	17321	18525	19157	20210
Maharashtra	20293	5516	5571	5350	18048	15343	20919	11953
Manipur	227	44	195	453	256	138	148	159
Meghalaya	170	133	180	154	165	169	211	191

Figure 2.4.4 New Measure

STEP 5

1. Create New Column to calculate percentage change in the number of accidents from 2015 to 2016.
2. Select table and then select “New Column”.
3. Apply DAX Measure to calculate “Percentage_Change_2015_2016”.

DAX MEASURE:

Percentage_Change_2015_2016 =

DIVIDE('Accidents_Acc_EduQual_Drivers_above_10_Standard_09-16 (2)'[2016]
'Accidents_Acc_EduQual_Drivers_above_10_Standard_09-16 (2)'[2015],
'Accidents_Acc_EduQual_Drivers_above_10_Standard_09-16 (2)'[2015], 0) * 100

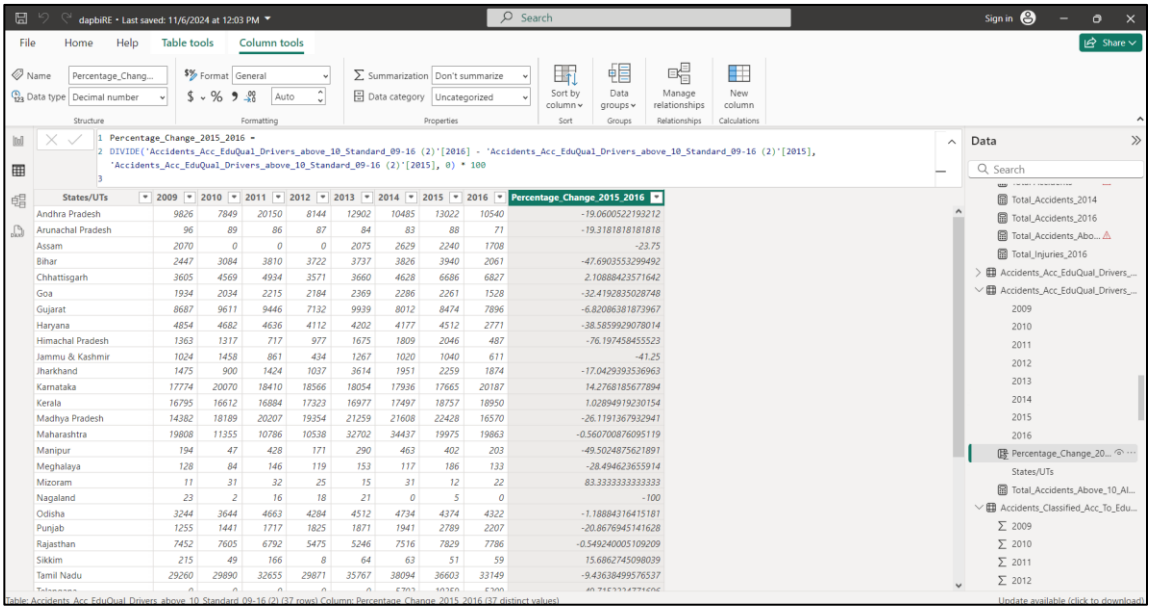


FIG 2.4.5 New Column for percentage change

STEP 5

1. Create New Measure to calculate the average severity of accidents.
2. Select table and then select “New Measure”.
3. Apply DAX Measure to calculate “Average_Severity”.

DAX MEASURE:

```
Average_Severity =
VAR Total_Accidents =
    SUM('Road_Accident_Profile_of_Select_Cities-2011-15'[All Accidents - 2011]) +
    SUM('Road_Accident_Profile_of_Select_Cities-2011-15'[All Accidents - 2012]) +
    SUM('Road_Accident_Profile_of_Select_Cities-2011-15'[Total Accidents - 2013]) +
    SUM('Road_Accident_Profile_of_Select_Cities-2011-15'[Total Accidents - 2014]) +
    SUM('Road_Accident_Profile_of_Select_Cities-2011-15'[Total Accidents - 2015])

VAR Total_Fatalities =
    SUM('Road_Accident_Profile_of_Select_Cities-2011-15'[Persons Killed - 2011]) +
    SUM('Road_Accident_Profile_of_Select_Cities-2011-15'[Persons Killed - 2012]) +
    SUM('Road_Accident_Profile_of_Select_Cities-2011-15'[Killed - 2013]) +
    SUM('Road_Accident_Profile_of_Select_Cities-2011-15'[Killed - 2014]) +
    SUM('Road_Accident_Profile_of_Select_Cities-2011-15'[Killed - 2015])
```

```
VAR Total_Injuries =
    SUM('Road_Accident_Profile_of_Select_Cities-2011-15'[Persons Injured - 2011]) +
    SUM('Road_Accident_Profile_of_Select_Cities-2011-15'[Persons Injured - 2012]) +
    SUM('Road_Accident_Profile_of_Select_Cities-2011-15'[Injured - 2013]) +
    SUM('Road_Accident_Profile_of_Select_Cities-2011-15'[Injured - 2014]) +
    SUM('Road_Accident_Profile_of_Select_Cities-2011-15'[Injured - 2015])
VAR Total_Severity = Total_Fatalities + Total_Injuries
RETURN
IF(Total_Accidents > 0, DIVIDE(Total_Severity, Total_Accidents), 0)
```

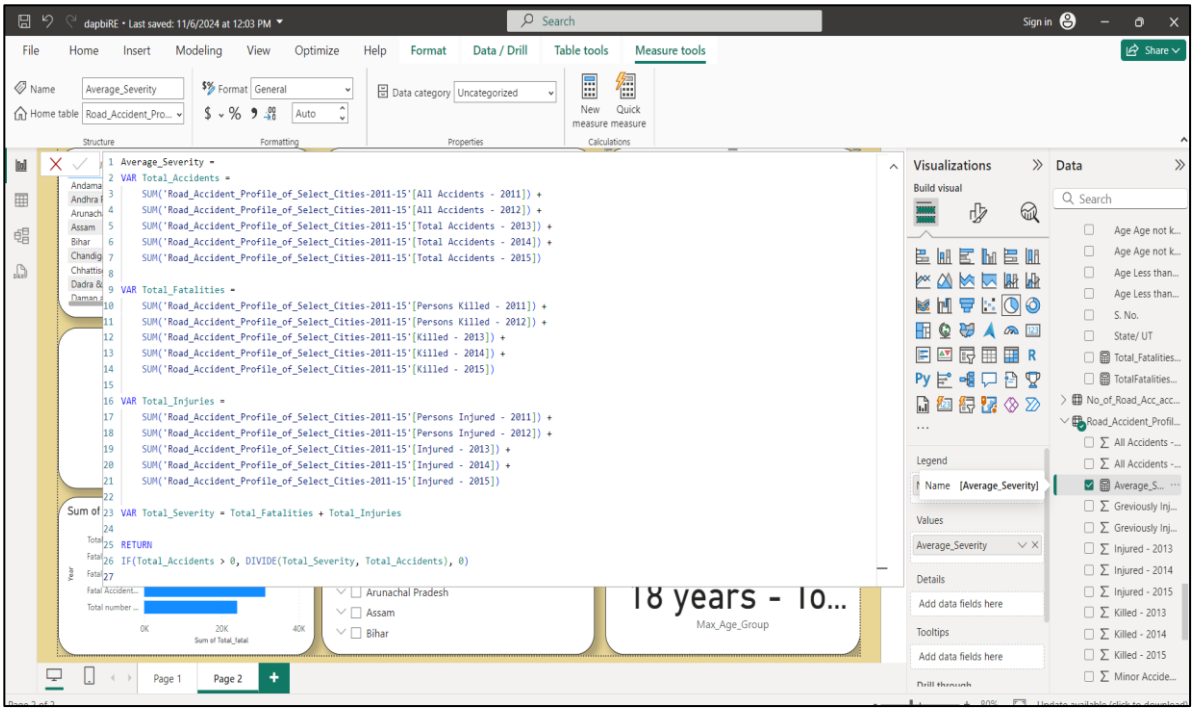


FIG 2.4.6 New Measure Average Severity

STEP 6

1. Create New Measure to calculate which age group has experienced the highest number of accidents.
2. Select table and then select “New Measure”.
3. Apply DAX Measure to calculate “Max_Age_Group”.

DAX MEASURE:

```
Max_Age_Group =
CALCULATE(
    MAX('No_of_Road_Acc_acco_to_clf_of_age_of_Driver_2014_2016 (2)'[Age Group]),
    FILTER('No_of_Road_Acc_acco_to_clf_of_age_of_Driver_2014_2016 (2)',
     'No_of_Road_Acc_acco_to_clf_of_age_of_Driver_2014_2016 (2)'[Total Accidents] =
      [Max_Accidents1])
)
```

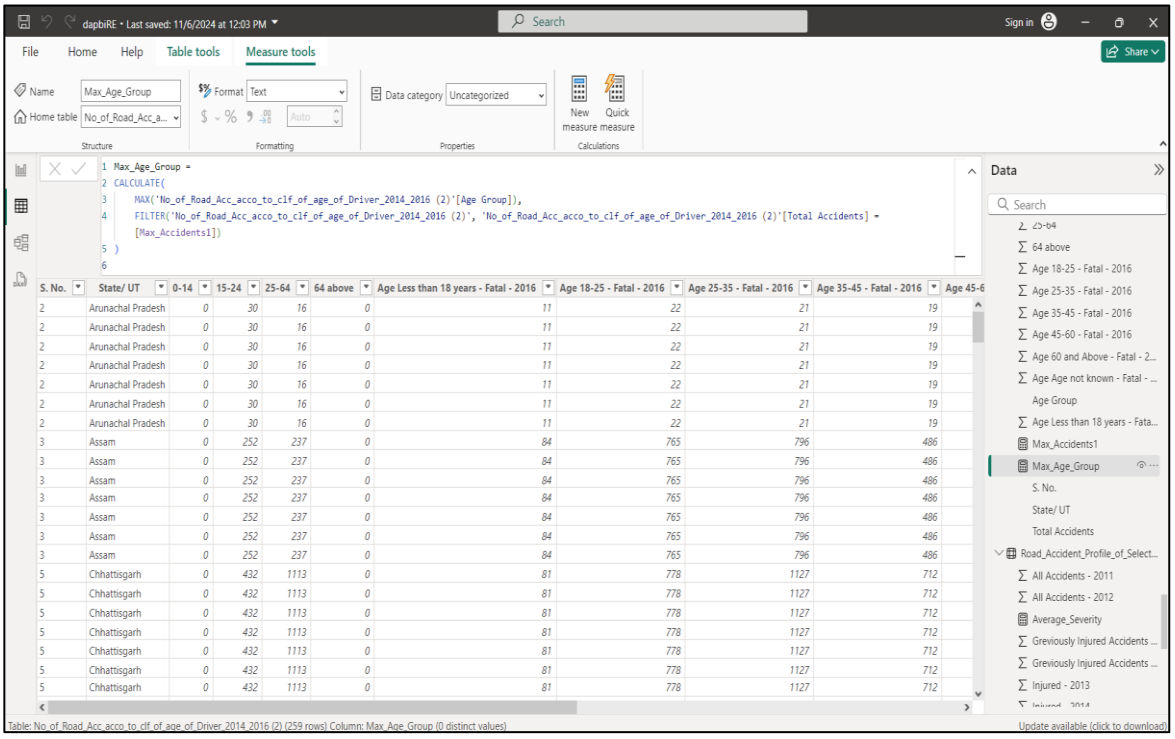


FIG 2.4.7 New Measure for max age group accident

CHAPTER 3

DATA ANALYSIS AND INTERPRETATION

3..1 DATA ANALYSIS

To turn raw data into insightful information, data analysis is the process of analyzing, manipulating, and monitoring. Making the necessary decisions for a business or company's growth is made easier with the use of data insights. Deep data analysis is crucial if need want to manage a firm that is data-driven. Then it is needed to find learning different Power BI data analysis approaches fascinating and useful.

Data analysis includes the following results

- Used to create various charts from Power Bi visuals
- Select datas from various tables, analyse it and convert it into visuals.
- From the analysed result infer the result or final solution.

CHARTS

1. Using the road accident data from 2014 and 2016, analyze the trends in accident frequency across different states and Union Territories (UTs)?

i. Select table Acc_classified_according_to_Type_of_Location_2014_and_2016.
ii. Include calculated measure “Total_Accidents_2016” and “Total_Accidents_2014”
iii. Then select **table chart** for visualization.

State/ UT	Total_Accidents_2016	Total_Accidents_2014
West Bengal	7,049.00	12875
Uttarakhand	860.00	1410
Uttar Pradesh	25,153.00	31034
Tripura	407.00	716
Total	292,446.00	480777
Telangana	15,657.00	20078
Tamil Nadu	50,267.00	67250
Sikkim	161.00	203
Rajasthan	10,953.00	24628
Punjab	4,988.00	6391
Puducherry	1,532.00	1111
Total	**584,892.00**	**961554**

Figure 3.1.1 Number of total accidents in 2014 and 2016

2. Analyze the road accident trends from 2009 to 2016?

i. Select table Accidents_Acc_EduQual_Drivers_above_10_Standard_09-16.

ii. Include "Total Accident" in X-axis and "Year" in Y-axis.

iii. Then select **stacked bar chart** for visualization.

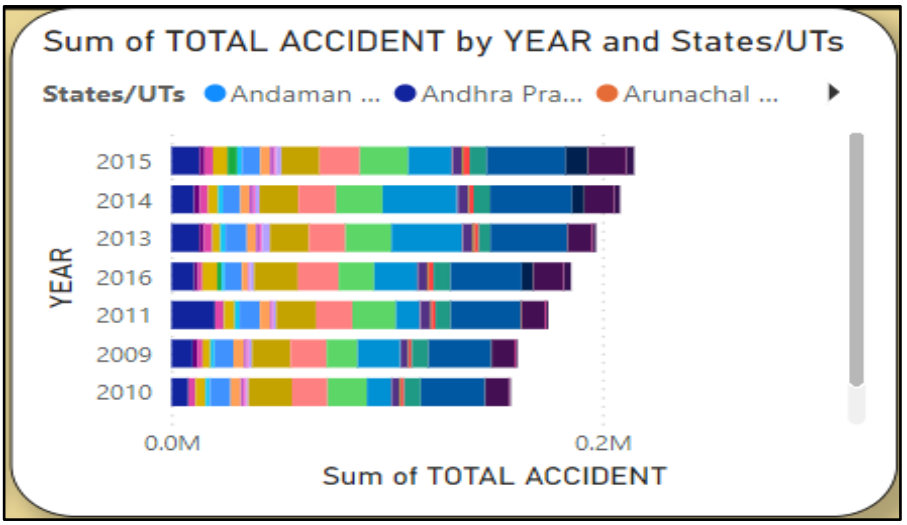


Figure 3.1.2 Total Accident by year

3. Compare the number of road accidents near schools and colleges in 2014 and 2016?

i. Select table Acc_classified_according_to_Type_of_Location_2014_and_2016.

ii. Include calculated measure "TASCH/CLG2016" and "TASCH/CLG2014"

iii. Then select **Clustered Column chart** for visualization.

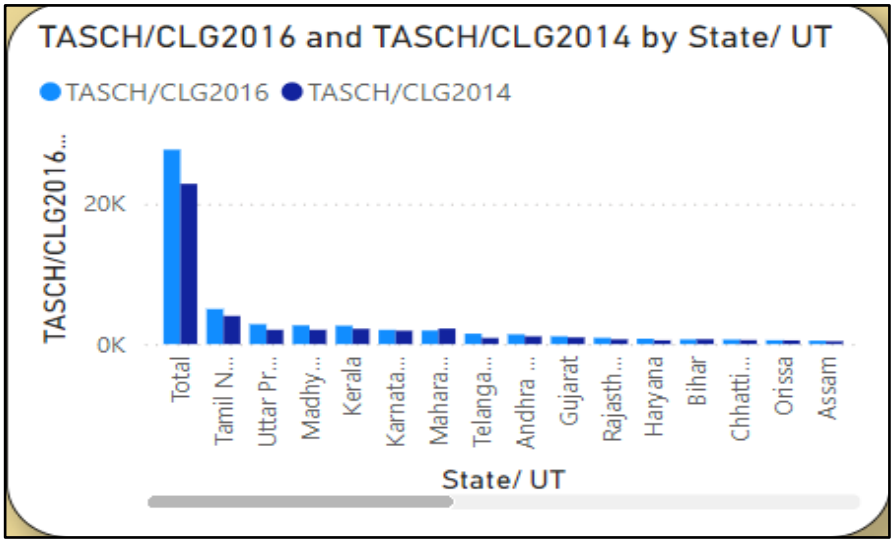


Figure 3.1.3 Accident near school or college

4. Using the data on road accident fatalities from 2016, identify the state with the highest number of fatalities?

i. Select table No_of_Road_Acc_acco_to_clf_of_age_of_Driver_2014_2016.

ii. Include calculated measure "TotalFatalities2016".

iii. Then select **Treemap chart** for visualization.

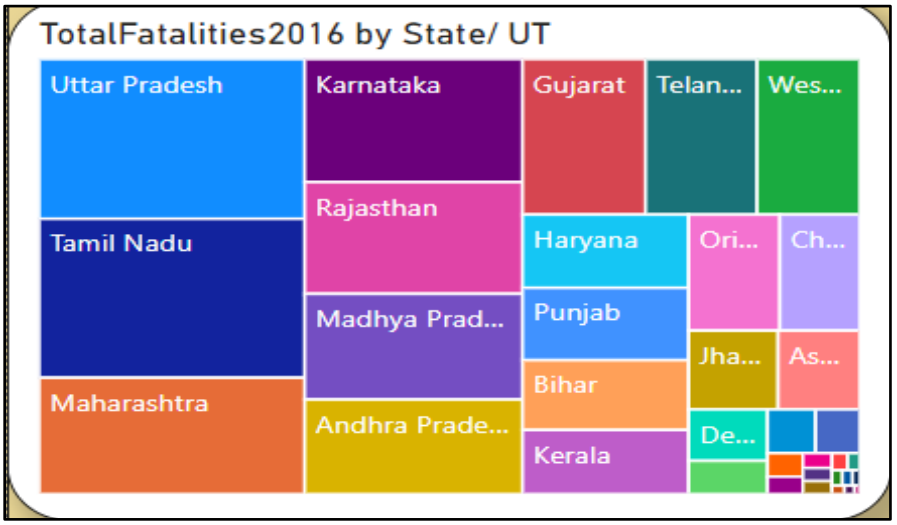


Figure 3.1.4 Highest number of fatalities

5. Calculate the average number of injuries per accident for each state in 2016?

i. Select table Acc_classified_according_to_Type_of_Location_2014_and_2016.

ii. Include calculated measure "Average_Injuries_Per_Accident_2016".

iii. Then select **Area chart** for visualization

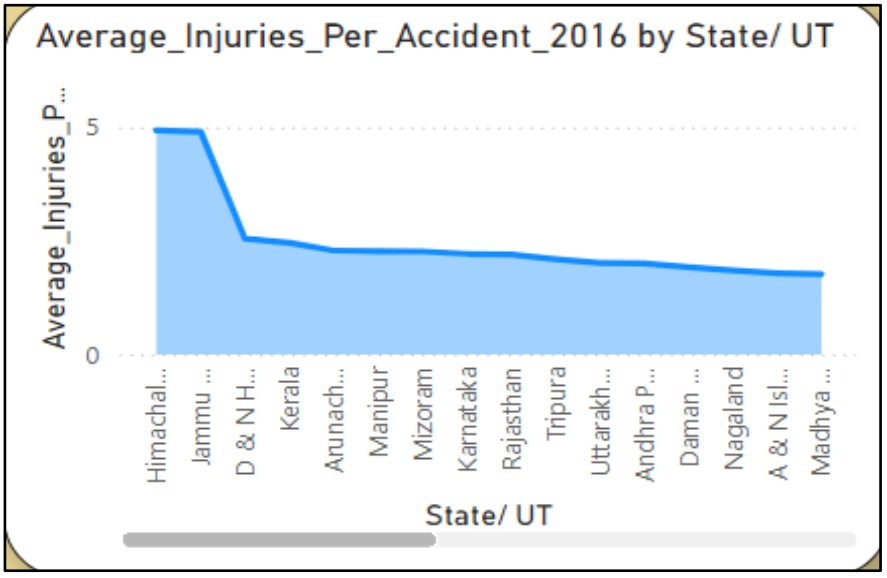


Figure 3.1.5 Average number of injuries

6. Examine the correlation between the education qualification of drivers and the number of accidents?

i. Select table Acc_classified_according_to_Type_of_Location_2014_and_2016.

ii. Include calculated measure “Total_Accidents_Above_10_All_Years” and “Total_Accidents_Edu_9_10_All_Years”.

iii. Then select **100% Stacked area chart** for visualization.

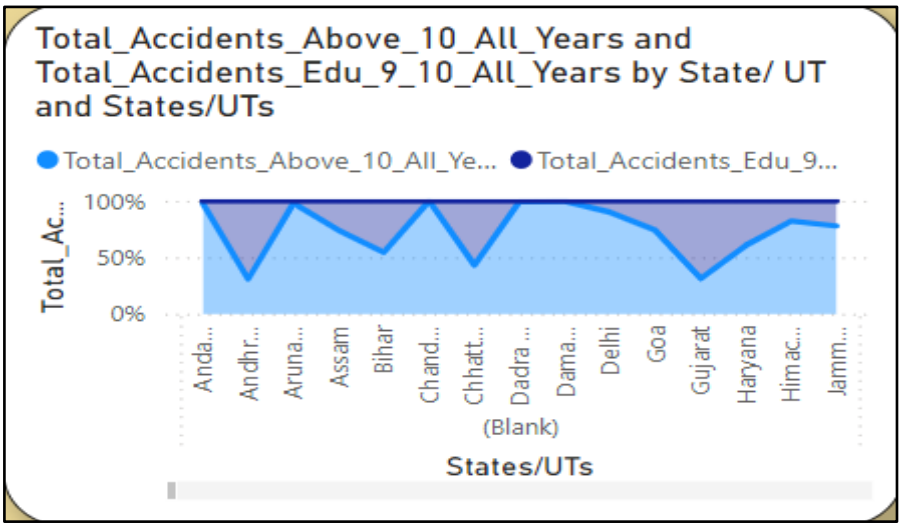


Figure 3.1.6 Education qualification with accident

7. Evaluate the percentage change in the number of road accidents from 2015 to 2016 across different states?

i. Select table Accidents_Acc_EduQual_Drivers_above_10_Standard_09-16.

ii. Include calculated measure “Percentage_Change_2015_2016” and “States/UT”.

iii. Then select **Table chart** for visualization.

States/UTs	Percentage_Change_2015_2016
Andaman & Nicobar Islands	10.00
Andhra Pradesh	-19.06
Arunachal Pradesh	-19.32
Assam	-23.75
Bihar	-47.69
Chandigarh	0.00
Chhattisgarh	2.11
Dadra & Nagar Haveli	-31.25
Daman and Diu	-45.90

Figure 3.1.7 Percentage Change

8. Design a DAX measure to analyze the average severity of accidents by city?

i. Select table Road_Accident_Profile_of_Select_Cities-2011-15.

ii. Include calculated measure “Average_Severity”.

iii. Then select **pie chart** for visualization.

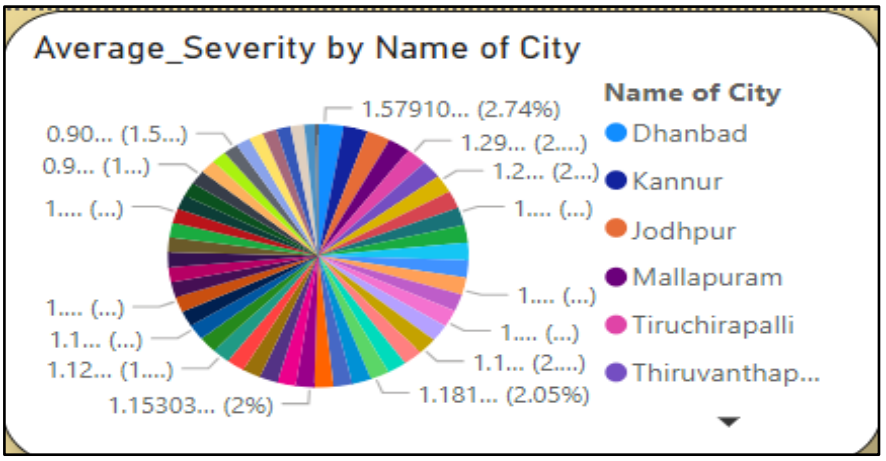


Figure 3.1.8 Average Severity by City

9. Write a DAX measure to calculate the total number of fatalities per state, considering only accidents involving drivers under the age of 25?

i. Select table No_of_Road_Acc_acco_to_clf_of_age_of_Driver_2014_2016.

ii. Include calculated measure “Total_Fatalities_Under_25”.

iii. Then select **Area chart** for visualization.

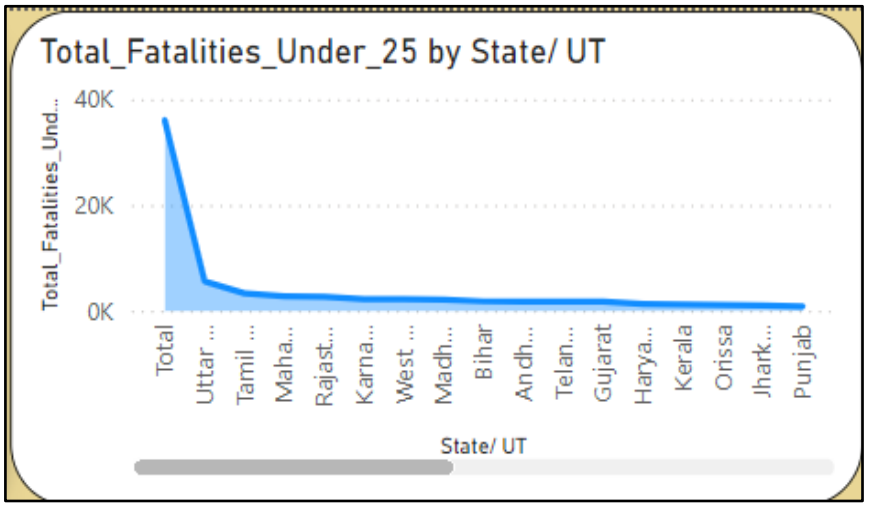


Figure 3.1.9 Total Fatalities under 25

10. Design a DAX measure to calculate the average number of fatalities per accident?

i. Select table Road_Accident_Profile_of_Select_Cities-2011-15.

ii. Include calculated measure “Average Fatalities per Accident”.

iii. Then select **Card** for visualization

Figure 3.1.10 Average number of fatalities

‘

11. Analyze the yearly distribution of accidents caused by weather conditions from 2006 to 2015?

i. Select table Due_to_Weather_Condition-2006-15.

ii. Include calculated measure “Sum of total_accident”.

iii. Then select **Donut chart** for visualization.

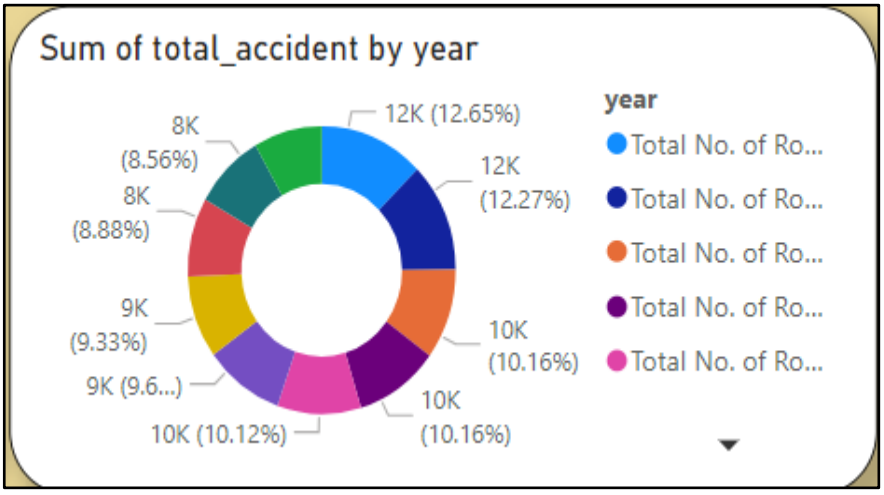


Figure 3.1.11 Sum of total accident due to weather

12. Create a Pivot Table showing the total number of accidents by state and year?

i. Select table Due_to_Weather_Condition-2006-15.

ii. Include calculated measure “Sum of total_accident”.

iii. Then select **Table chart** for visualization.

States/Uts	Total No. of Road Accidents 2006	Total No. of Road Accidents 2007	Total No. of Road Accidents 2008	Total No. of Road Accidents 2009	Total No. of Road Accidents 2010
West Bengal	393	583	778	362	575
Uttarakhand	12	24	18	6	11
Uttar Pradesh	577	557	738	780	831
Tripura	0	1	0	0	0
Telangana	0	0	0	0	0
Tamil Nadu	395	14	120	88	80
Sikkim	16	13	22	39	19
Rajasthan	489	466	13	5	11
Punjab	56	69	71	133	71
Puducherry	0	0	0	0	0
Odisha	198	306	323	378	0
Nagaland	0	0	0	0	0
Total	**8370**	**8792**	**8060**	**7758**	**9534**

Figure 3.1.12 Total accidents by state and year

13. Calculate the total number of fatalities per accident by year?

i. Select table Road_Accident_Profile_of_Select_Cities-2011-15.

ii. Include calculated measure “Sum of Total_fatal”.

iii. Then select Stacked bar chart for visualization.

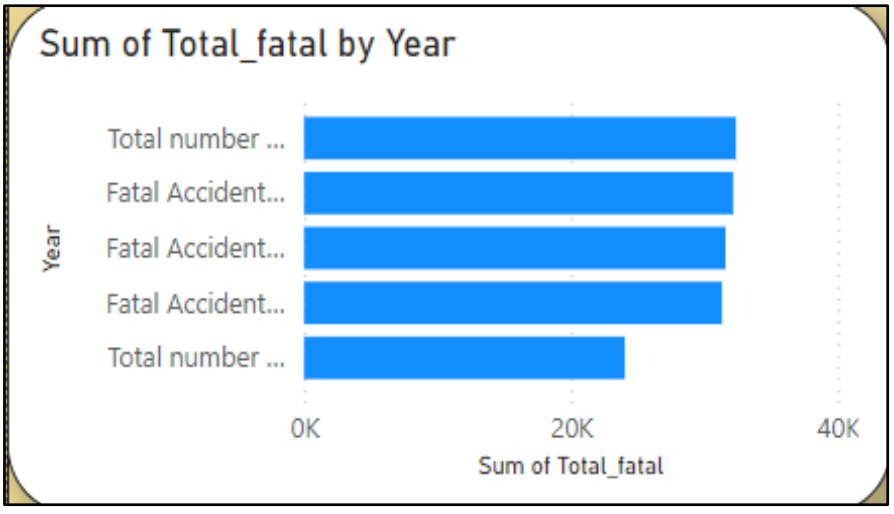


Figure 3.1.13 Total number of fatal

14. Design a process to filter data by specific states for customized reports?

i. Select slicer to filter the data by specific states.

ii. Include “State/UT” in fields

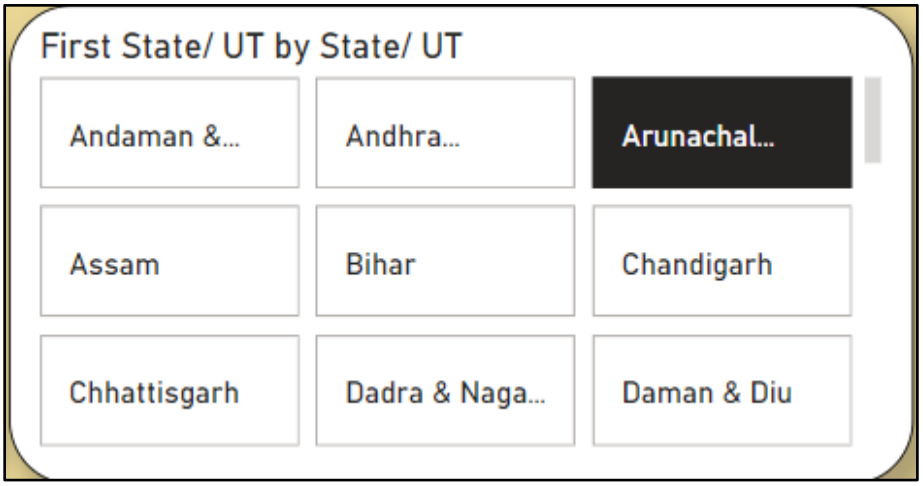


Figure 3.1.14 Slicer

15. Identify the age group that has experienced the highest number of accidents?

i. Select table No_of_Road_Acc_acco_to_clf_of_age_of_Driver_2014_2016.

ii. Include calculated measure “Max_Age_Group”.

iii. Select Card for better visualization.

Figure 3.1.15 Highest number of accident by age

3.2 PUBLISHING DASHBOARD

- Often referred to as a canvas, a Power BI dashboard is a single page that employs visuals to convey a story. A well-designed dashboard only includes the key components of the tale because it is only onepage long. The dashboard's tiles—the visuals you see there—are placed there by report creators.
- The report page where the visualisation was made is often the page you land on after picking a tile. A dashboard's visuals are derived from reports, and each report is built using a single dataset. A dashboard may really be thought of as a portal to the underlying reports and statistics.
- Then it may get the report that was used to produce a visualisation by selecting Dashboards are an excellent method to keep an eye on your company, search for solutions, and quickly view all of your most crucial indicators.
- A dashboard's visualisations might be drawn from a single underlying dataset or several, as well as a single underlying report or many.
- Regardless of where the data is stored, a dashboard may mix on-premises and cloud data to provide a consolidated picture. A dashboard is interactive, and the tiles refresh as the underlying data changes. It is more than simply a lovely picture.

Link for dashboard

https://app.powerbi.com/groups/me/reports/ac4c4570-3e2d-4e75-924c-d7ca45e69311/b58b031030321b3a66 ac?experience=power-bi

Process of creating Dashboard

STEP 1

1. First, create your road accident analysis dashboard in Power BI.
2. In Power BI Desktop, go to File > Publish > Publish to Power BI and select a workspace in Power BI Service to publish your report.

Figure 3.2.1 Publish dash board

STEP 2

1. Open Power Bi serviced in web browser.
2. Login in Power Bi from your account.
3. Click My Workspace and check “Road_Accident_Analysis” dashboard
4. Open the Published Dashboard

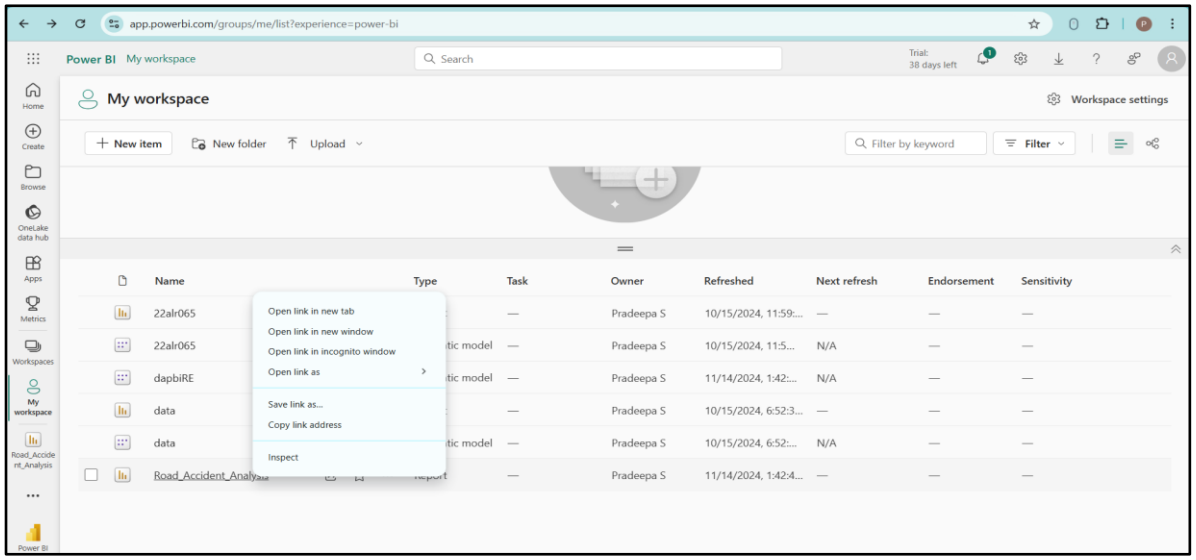


Figure 3.2.2 Open published dashboard

DASHBOARD VIEW OF INDIAN ROAD ACCIDENT ANALYSIS

PAGE 1:

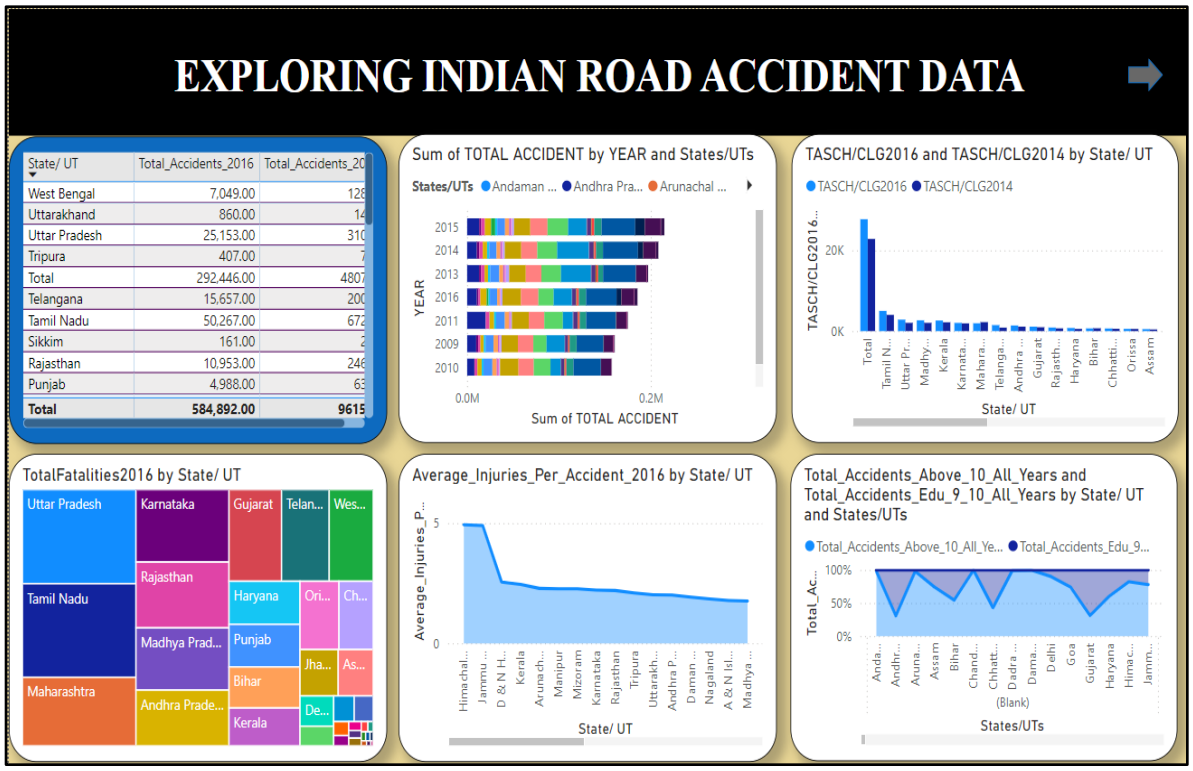


Figure 3.2.3 Road accident dashboard

PAGE 2

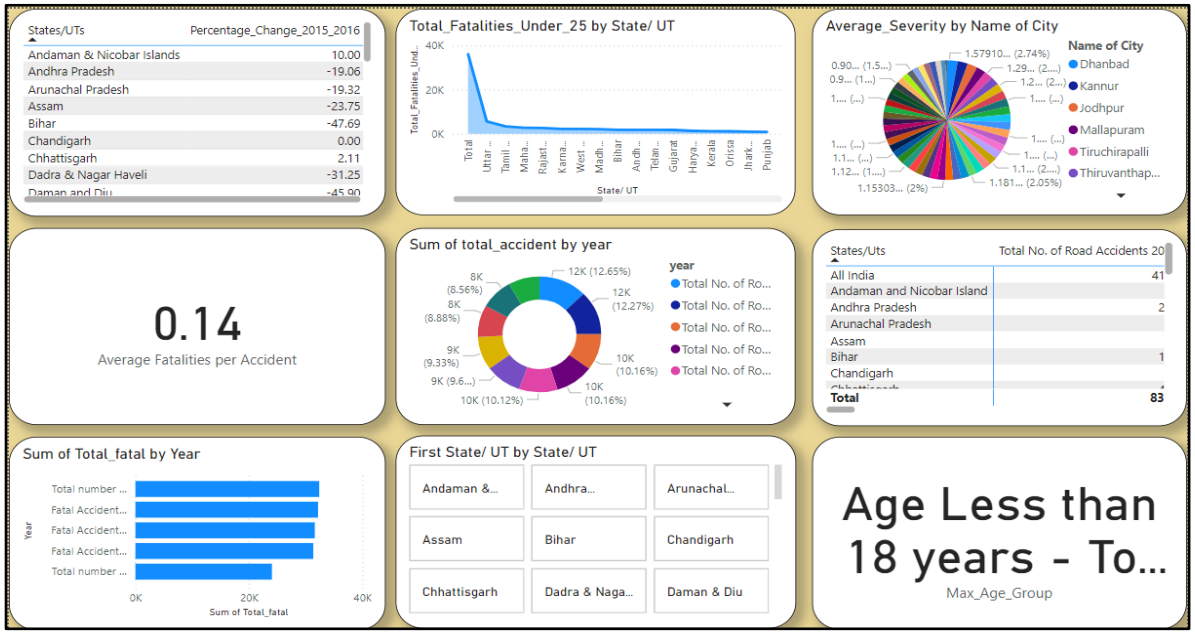


Figure 3.2.4 Road accident dashboard

3.3 INFERENCES

1. **Using the road accident data from 2014 and 2016, analyze the trends in accident frequency across different states and Union Territories (UTs)?**
   - The visualization of the table chart shows the data for accidents in 2016 and 2014.
   - From the visualization, West Bengal had the highest number of accidents in 2016, with a total of 7,049 accidents, and also had the highest accident count in 2014, with 12,875 accidents.
   - The total number of accidents across all states in 2016 was 584,892, while in 2014, the total was 961,554.
2. **Analyze the road accident trends from 2009 to 2016?**
   - The visualization of the road accident trends from 2009 to 2016 highlights key insights.
   - The highest number of accidents occurred in the year 2015.
   - The lowest number of accidents occurred in the year 2010.
   - Overall, Andhra Pradesh recorded the highest number of road accidents among all states during this period.
3. **Compare the number of road accidents near schools and colleges in 2014 and 2016?**
   - Tamil Nadu recorded the highest TASCH/CLG values in both 2016 and 2014.
   - Uttar Pradesh, Madhya Pradesh, and Kerala also show relatively high TASCH/CLG figures.
   - States such as Assam, Orissa, and Chhattisgarh have comparatively low TASCH/CLG values.
   - The "Total" bar represents the combined figures across all states, showing the highest value overall.
4. **Using the data on road accident fatalities from 2016, identify the state with the highest number of fatalities?**
   - The TreeMap chart visualizes road accident fatalities data from 2016.
   - Uttar Pradesh has the highest rate of road accident fatalities in 2016.
   - The second-highest rate of road accident fatalities in 2016 follows closely after Uttar Pradesh.
   - This visualization helps in comparing the accident fatality rates across different states for that year.

5. **Calculate the average number of injuries per accident for each state in 2016?**
   - This visualization uses an Area chart to represent the average number of injuries per accident for each state in 2016.
   - Himachal Pradesh has the highest average number of injuries per accident among the states in 2016.
   - Other states show a gradual decrease in the average number of injuries per accident.

6. **Examine the correlation between the education qualification of drivers and the number of accidents?**
   - This visualization uses a Stacked Area chart to show the correlation between drivers' education levels and the number of accidents.
   - Andhra Pradesh shows the highest number of accidents related to drivers' education levels.
   - The chart highlights variations across different states, indicating how education qualifications may impact accident rates.
   - This visualization effectively illustrates the comparative accident trends by education level for each state.

7. **Evaluate the percentage change in the number of road accidents from 2015 to 2016 across different states?**
   - This visualization uses a Table chart to display the percentage change in road accidents from 2015 to 2016 for each state.
   - The data shows that Andaman & Nicobar Islands had a 10.00% increase in road accidents during this period.
   - In contrast, Bihar experienced a significant reduction in road accidents, with a decrease of 47.69%.
   - The table provides a clear comparison of changes in accident rates across different states from 2015 to 2016.

8. **Design a DAX measure to analyze the average severity of accidents by city?**
   - For the visualization, use a pie chart to show the average severity of accidents by city.
   - From the visualization, the highest number is for Dhanbad city.
   - From this point, the numbers gradually decrease.

9. Write a DAX measure to calculate the total number of fatalities per state, considering only accidents involving drivers under the age of 25?

- For the visualization, use an area chart to show the total number of fatalities per state, considering only accidents.
- Uttar Pradesh has the highest total number of fatalities.
- Punjab has the lowest total number of fatalities**.**

10. Design a DAX measure to calculate the average number of fatalities per accident?

- For the visualization, use a card to display the calculated average number of fatalities per accident.
- From the visualization, the average number of fatalities per accident is 0.14.

11. Analyze the yearly distribution of accidents caused by weather conditions from 2006 to 2015?

- Use an Area Chart to display the total number of fatalities per state, with the X-axis representing the states and the Y-axis representing the total fatalities.
- Highlight Total No_of_Road_Accidents 2015 with the highest total fatalities and Total No_of_Road_Accidents 2016 as the state with the lowest fatalities for easy identification.

12. Create a Pivot Table showing the total number of accidents by state and year

- Uttar Pradesh consistently leads with the highest accidents, peaking at 831 in 2010.
- Punjab shows a decline, from 56 in 2006 to 11 in 2010.
- Odisha fluctuates, rising to 378 in 2009 but dropping to 0 in 2010.
- Tripura, Telangana, and Puducherry report 0 accidents throughout, indicating no incidents or missing data.

13. Calculate the total number of fatalities per accident by year?

- The chart shows total fatalities by year, with the highest nearing 40,000.
- Fatalities are consistent across years with minimal variation.
- Adding data labels and clearer year names would enhance readability.

14. Design a process to filter data by specific states for customized reports?

- Use the slicer visualization to filter data by specific states for customized reports.
- In the visualization, select Arunachal Pradesh.

15. Identify the age group that has experienced the highest number of accidents?

- Use a card visualization to display the group with the highest number of accidents.
- The result shows the group above and below 18 years.

INFERENCE OF DASH BOARD:

- Finally, the dashboard is created using the POWER BI services and is visualized under the name "Indian Road Accident Data Dashboard."
- All visualizations are pinned to a comprehensive dashboard.
- The dashboard contains multiple insights, such as "State-Wise Accidents," "Fatalities by Age Group," and "Yearly Trends in Accidents," which are successfully consolidated into a single interactive dashboard.
- The "State-Wise Accidents" section highlights total accidents, fatalities, and injuries for each state/UT.
- The "Fatalities by Age Group" section provides an in-depth analysis of fatalities involving different age groups, such as drivers under 25 or over 60.
- The "Yearly Trends in Accidents" visualizes fluctuations in road accidents across the years and identifies high-risk periods for specific states.

CHAPTER 4

CONCLUSION AND FUTURE WORK

4.1 RECOMMENDATIONS

- Develop interactive and user-friendly dashboards to summarize cricket matches, helping the audience easily analyze key aspects such as team performance, player statistics, and match highlights.
- Use Power BI to visualize match data dynamically, making it accessible and understandable for both educated and uneducated cricket enthusiasts.
- Provide Key Performance Indicators (KPIs) like strike rates, economy rates, and match-winning contributions to offer quick insights into team and player performance.
- Centralize the match statistics and significant goals to help users track the progress of teams and individual players throughout the tournament.
- Create dashboards that are visually appealing and easy to navigate, catering to busy users who may not have time to analyze detailed match reports.
- Leverage Power BI's collaborative capabilities to share insights with stakeholders, enabling real-time discussions and updates on cricket events.
- Include historical comparisons and predictive analysis using Power BI to enhance decision-making for team strategies and fan engagement.

REFERENCES

1. Sonal, Shristi, and Saumya Suman. "A framework for analysis of road accidents." 2018 International Conference on Emerging Trends and Innovations In Engineering And Technological Research (ICETIETR). IEEE, 2018.
2. Ruikar, Manisha. "National statistics of road traffic accidents in India." Journal of Orthopaedics, Traumatology and Rehabilitation 6.1 (2013): 1-6.
3. Krishnan, Vijay. "Research data analysis with power bi." (2017).
4. Becker, Louis T., and Elyssa M. Gould. "Microsoft power BI: extending excel to manipulate, analyze, and visualize diverse data." Serials Review 45.3 (2019): 184-188.
5. Ferrari, Alberto, and Marco Russo. Analyzing data with power BI and power pivot for excel. microsoft press, 2017.
6. Powell, Brett. Mastering Microsoft Power BI: expert techniques for effective data analytics and business intelligence. Packt Publishing Ltd, 2018.
7. Awad, Mohammed, Aneesa Al Redhaei, and Salam Fraihat. "Using business intelligence to analyze road traffic accidents." Proceedings of the Central and Eastern European eDem and eGov Days. 2022. 83-92.
8. Setiawan, Johan. "Visualization of Road Accident Data In Jabodetabek Area Year 2014-2017." Journal of Physics: Conference Series. Vol. 1196. No. 1. IOP Publishing, 2019.
9. Abualigah, Sayel M., et al. "Visualization of Accidents Data in United States." 2022 13th International Conference on Information and Communication Systems (ICICS). IEEE, 2022.
10. Ruqti, Dina, Mohammad Alkhammash, and Dalia Al-Eisawi. "Visual Data Analytics Dashboard for Insightful Traffic Monitoring: The Case of Jordan’s Transportation Sector." International Conference on Business and Technology. Cham: Springer Nature Switzerland, 2024.