

INDIAN INSTITUTE OF SCIENCE, BANGLORE

SUMMER INTERNSHIP
ROAD ACCIDENT ANALYSIS PROJECT
(POWER-BI)
REPORT- FILE

SUBMITTED BY :
ARUNDHATI ARUN TAK
B.TECH (CSE) (DATA SCIENCE)

INTRODUCTION :

The dataset comprises a comprehensive collection of accident records, offering insights into various factors contributing to road incidents. Here is a breakdown of the dataset attributes:

- **Accident_Index:** This attribute serves as a unique identifier for each accident entry, facilitating easy referencing and data management.
- **Accident Date and Time:** These fields provide information regarding the date and time of each accident, enabling temporal analysis to identify patterns such as peak accident hours or seasonal variations.
- **Day_of_Week:** This categorical attribute denotes the day of the week on which the accident occurred, aiding in understanding the influence of weekdays on accident frequency and severity.
- **Junction_Control and Junction_Detail:** These attributes shed light on the nature of junctions involved in accidents and the corresponding control mechanisms (if any), which is crucial for assessing the role of intersections in road safety.
- **Accident_Severity:** This categorical variable categorizes accidents based on their severity, allowing for the prioritization of interventions and resources towards mitigating the most severe accidents.
- **Latitude and Longitude:** These geographical coordinates pinpoint the precise location of each accident, facilitating spatial analysis to identify accident hotspots and plan targeted interventions.
- **Light_Conditions and Weather_Conditions:** These attributes describe the prevailing lighting and weather conditions at the time of the accident, respectively. Understanding their influence on accident occurrence is vital for implementing measures like improved street lighting or weather-specific safety protocols.
- **Local_Authority_(District) and Police_Force:** These fields provide administrative details regarding the jurisdiction and reporting authority associated with each accident, which can be useful for regulatory compliance and inter-agency collaboration.
- **Carriageway_Hazards and Road_Surface_Conditions:** These attributes capture additional contextual information regarding hazards present on the road surface and prevailing road conditions, offering insights into factors contributing to accident causation and severity.
- **Number_of_Casualties and Number_of_Vehicles:** These numerical variables quantify the human and vehicular impact of each accident, respectively, enabling statistical analysis to understand casualty distribution and vehicle involvement patterns.

- **Road_Type and Speed_limit:** These attributes characterize the physical attributes of the road where the accident occurred and the corresponding speed limit, providing essential contextual information for understanding accident dynamics and risk factors associated with different road types and speed limits.
- **Urban_or_Rural_Area and Vehicle_Type:** These categorical variables classify accidents based on the urban or rural setting and the type of vehicles involved, offering insights into the differential risk profiles associated with different environments and vehicle categories.

This dataset serves as a valuable resource for researchers, policymakers, and stakeholders in the field of road safety, offering rich insights into the multifaceted nature of road accidents and informing evidence-based interventions aimed at reducing accident occurrence and mitigating their impact on society.

Excel View Of Dataset:

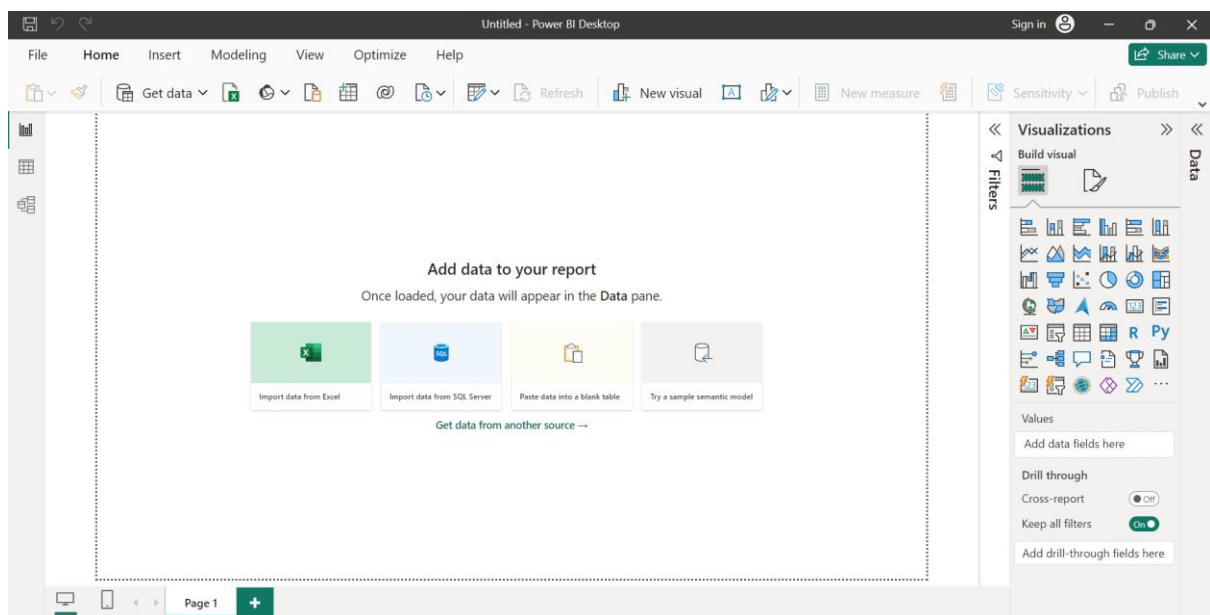
Road Accident Data - Excel										
Arundhati Tak										
File Home Insert Draw Page Layout Formulas Data Review View Help Power Pivot Tell me what you want to do										
Clipboard Font Alignment Number Styles Cells Editing Add-ins										
X88										
	A	B	C	D	E	F	G	H	I	J
1	Accident_Index	Accident Date	Day_of_Week	Junction_Control	Junction_Detail	Accident_Severity	Latitude	Light_Conditions	Local_Authority_(District)	Carriageway_Hazards
2	BS0000001	01-01-2021	Thursday	Give way or uncontrolled	T or staggered junction	Serious	51.51227	Daylight	Kensington and Chelsea	None
3	BS0000002	05-01-2021	Monday	Give way or uncontrolled	Crossroads	Serious	51.5144	Daylight	Kensington and Chelsea	None
4	BS0000003	04-01-2021	Sunday	Give way or uncontrolled	T or staggered junction	Slight	51.48667	Daylight	Kensington and Chelsea	None
5	BS0000004	05-01-2021	Monday	Auto traffic signal	T or staggered junction	Serious	51.5078	Daylight	Kensington and Chelsea	None
6	BS0000005	06-01-2021	Tuesday	Auto traffic signal	Crossroads	Serious	51.48208	Darkness - lights lit	Kensington and Chelsea	None
7	BS0000006	01-01-2021	Thursday	Give way or uncontrolled	T or staggered junction	Slight	51.49342	Daylight	Kensington and Chelsea	None
8	BS0000007	08-01-2021	Thursday	Give way or uncontrolled	T or staggered junction	Serious	51.48018	Daylight	Kensington and Chelsea	None
9	BS0000008	02-01-2021	Friday	Auto traffic signal	Crossroads	Slight	51.49196	Daylight	Kensington and Chelsea	None
10	BS0000009	07-01-2021	Wednesday	Give way or uncontrolled	T or staggered junction	Slight	51.49646	Daylight	Kensington and Chelsea	None
11	BS0000010	10-01-2021	Saturday	Auto traffic signal	Crossroads	Slight	51.48115	Daylight	Kensington and Chelsea	None
12	BS0000011	07-01-2021	Wednesday	Auto traffic signal	Crossroads	Slight	51.48208	Darkness - lights lit	Kensington and Chelsea	None
13	BS0000012	16-01-2021	Friday	Auto traffic signal	Crossroads	Slight	51.495	Darkness - lights lit	Kensington and Chelsea	None
14	BS0000013	12-01-2021	Monday	Data missing or out of	Not at junction or withi	Slight	51.49878	Daylight	Kensington and Chelsea	None
15	BS0000014	09-01-2021	Friday	Give way or uncontrolled	T or staggered junction	Slight	51.50619	Daylight	Kensington and Chelsea	None
16	BS0000015	17-01-2021	Saturday	Give way or uncontrolled	T or staggered junction	Slight	51.49308	Daylight	Kensington and Chelsea	None
17	BS0000016	25-01-2021	Sunday	Auto traffic signal	Crossroads	Serious	51.48208	Darkness - lights lit	Kensington and Chelsea	None
18	BS0000017	26-01-2021	Monday	Give way or uncontrolled	Crossroads	Slight	51.48867	Darkness - lights lit	Kensington and Chelsea	None
19	BS0000018	26-01-2021	Monday	Data missing or out of	Not at junction or withi	Slight	51.48236	Darkness - lights lit	Kensington and Chelsea	None
20	BS0000019	19-01-2021	Monday	Give way or uncontrolled	T or staggered junction	Slight	51.49391	Daylight	Kensington and Chelsea	None
21	BS0000020	27-01-2021	Tuesday	Data missing or out of	Not at junction or withi	Slight	51.5093	Darkness - lights lit	Kensington and Chelsea	None

Data Preparation:

1. Importing Data to PowerBI:

- Open Power BI Desktop: Launch Power BI Desktop on your computer.
- Get Data: Click on the "Home" tab in the ribbon at the top of the window. Then, click on "Get Data" and choose "Excel" from the dropdown menu.
- Select Excel File: Navigate to the location of your Excel file and select it. Click "Open" to continue.
- Choose Data: In the Navigator window, you'll see a list of sheets and tables available in the Excel file. Select the sheets or tables you want to import data from by checking the box next to them. You can also preview the data by clicking on the sheet or table.
- Load Data: After selecting the desired sheets or tables and optionally transforming the data, click on the "Load" button. This will import the data into Power BI.

IMPORTING VIEW IN POWERBI :



Cleaning Data:

1. Error Correction:

- **Open Power Query Editor:** Click on the "Transform Data" button in the Home tab of the Power BI Desktop ribbon. This will open the Power Query Editor window.
- **Select Column:** In the Power Query Editor, navigate to the column containing the values you want to correct i.e. Accident Severity
- **Replace Values:**
 - i. Right-click on the column header and select "Replace Values" from the context menu.
 - ii. In the "Replace Values" dialog box, you'll see two text boxes labelled "Value to Find" and "Replace With".
 - iii. In the "Value to Find" text box, enter the incorrect value (e.g., "fetal").
 - iv. In the "Replace With" text box, enter the correct value (e.g., "fatal").
 - v. Click "OK" to apply the replacement.
 - vi. Review Changes: Power Query will apply the replacement to all occurrences of the incorrect value in the selected column. You can review the changes in the preview window to ensure they are correct.
- **Close and Apply:** Once you are satisfied with the changes, click on the "Close & Apply" button in the Home tab of the Power Query Editor to close the editor and apply the changes to your data model.

CLEANING VIEW IN POWERBI :

Accident_Index	Accident Date	Day of Week	Junction Control	Junction_Detail	Accident_Severity	Latitude	Light Conditions	Local Authority
BS0000139	08 April 2021	Wednesday	Give way or uncontrolled	T or staggered junction	Slight	51.534472	Daylight	Westmir
BS0000158	22 April 2021	Wednesday	Give way or uncontrolled	T or staggered junction	Slight	51.532526	Daylight	Brent
BS0000188	29 April 2021	Wednesday	Give way or uncontrolled	T or staggered junction	Slight	51.522967	Daylight	Westmir
BS0000331	15 July 2021	Wednesday	Give way or uncontrolled	T or staggered junction	Slight	51.532512	Daylight	Camden
BS0000367	29 July 2021	Wednesday	Give way or uncontrolled	T or staggered junction	Slight	51.554499	Daylight	Camden
BS0000524	28 October 2021	Wednesday	Give way or uncontrolled	T or staggered junction	Slight	51.528382	Daylight	Camden
BS0000531	28 October 2021	Wednesday	Give way or uncontrolled	T or staggered junction	Slight	51.526805	Daylight	Camden
BS0000551	28 October 2021	Wednesday	Give way or uncontrolled	T or staggered junction	Slight	51.524501	Daylight	Westmir
BS0001040	07 January 2021	Wednesday	Give way or uncontrolled	T or staggered junction	Slight	51.534472	Daylight	Westmir
BS0001071	11 February 2021	Wednesday	Give way or uncontrolled	T or staggered junction	Slight	51.532526	Daylight	Brent
BS0001356	25 March 2021	Wednesday	Give way or uncontrolled	T or staggered junction	Slight	51.522967	Daylight	Westmir
BS0001370	08 April 2021	Wednesday	Give way or uncontrolled	T or staggered junction	Slight	51.532512	Daylight	Camden
BS0001374	08 April 2021	Wednesday	Give way or uncontrolled	T or staggered junction	Slight	51.554499	Daylight	Camden
BS0001386	08 April 2021	Wednesday	Give way or uncontrolled	T or staggered junction	Slight	51.528382	Daylight	Camden
BS0001444	22 April 2021	Wednesday	Give way or uncontrolled	T or staggered junction	Slight	51.526805	Daylight	Camden
BS0001621	20 May 2021	Wednesday	Give way or uncontrolled	T or staggered junction	Slight	51.524501	Daylight	Westmir
BS0001891	01 July 2021	Wednesday	Give way or uncontrolled	T or staggered junction	Slight	51.534472	Daylight	Westmir
BS0001921	24 June 2021	Wednesday	Give way or uncontrolled	T or staggered junction	Slight	51.532526	Daylight	Brent
BS0001930	08 July 2021	Wednesday	Give way or uncontrolled	T or staggered junction	Slight	51.522967	Daylight	Westmir
BS0002086	05 August 2021	Wednesday	Give way or uncontrolled	T or staggered junction	Slight	51.532512	Daylight	Camden
BS0002092	05 August 2021	Wednesday	Give way or uncontrolled	T or staggered junction	Slight	51.554499	Daylight	Camden
BS0002299	16 September 2021	Wednesday	Give way or uncontrolled	T or staggered junction	Slight	51.528382	Daylight	Camden
BS0003307	30 September 2021	Wednesday	Give way or uncontrolled	T or staggered junction	Slight	51.526805	Daylight	Camden

Data Transformation:

- **Rename Columns:**
 - "Accident_Index" can be renamed to "Index".
 - "Accident Date" can be renamed to "Date".
 - "Day_of_Week" can be kept as is.
 - Similarly, rename other columns to more concise names if necessary.
- **Change Data Types:**
 - Change "Accident Date" column to the date data type.
 - Ensure that numeric columns like "Latitude", "Longitude", "Speed_limit", etc., are in the number format.
- **Handle Missing Values:**
 - Identify any missing values in columns and decide whether to replace them or remove them based on the analysis requirements.
- **Standardize Categorical Values:**
 - Standardize categorical values in columns like "Accident_Severity", "Light_Conditions", "Road_Surface_Conditions", etc., for consistency.
- **Split Columns:**
 - If "Accident Date" contains both date and time, split it into separate columns for date and time.
- **Remove Unnecessary Columns:**
 - Remove columns like "Local_Authority_(District)" if they are not relevant to the analysis.
- **Add Calculated Columns:**
 - If needed, add calculated columns using DAX expressions for additional analysis, such as calculating the time difference between accidents.
 - Like create a new column for Month write the DAX query as "Month = `format('Calender'[Date], "mmm")`" gives the months in names.
 - For year write DAX query "Year = `year('Calender'[Date])`"

- **Close and Apply:** Once satisfied with the transformations, click on the "Close & Apply" button in the Home tab to close the Power Query Editor and apply the changes to your Power BI data model.

POWER QUERY VIEW :

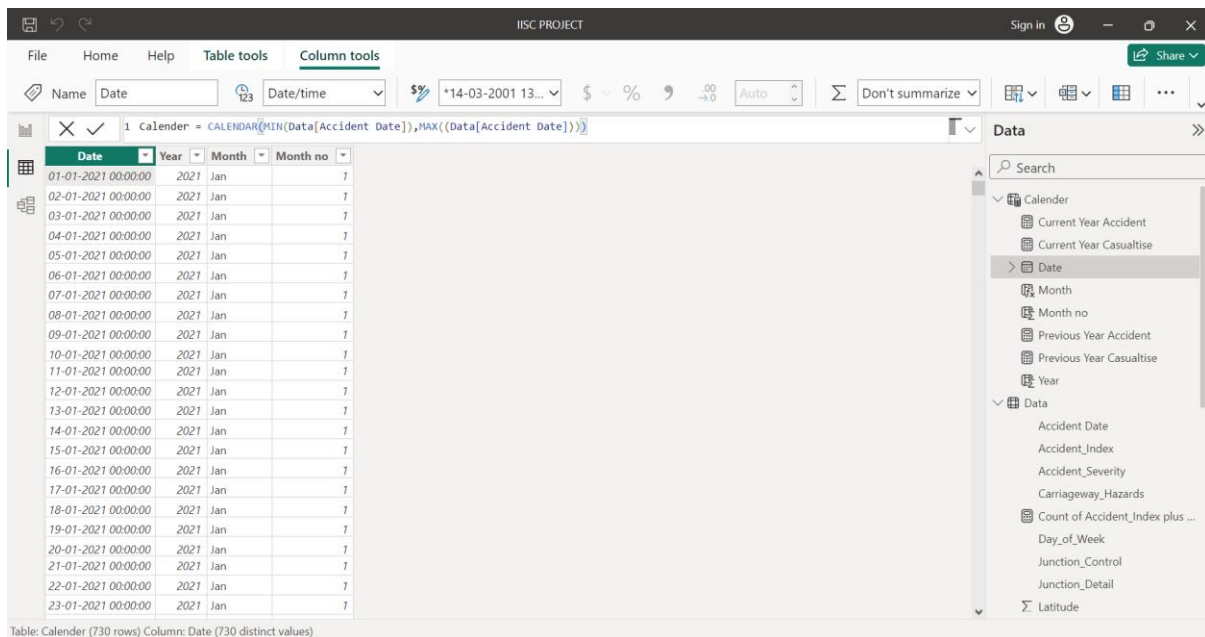
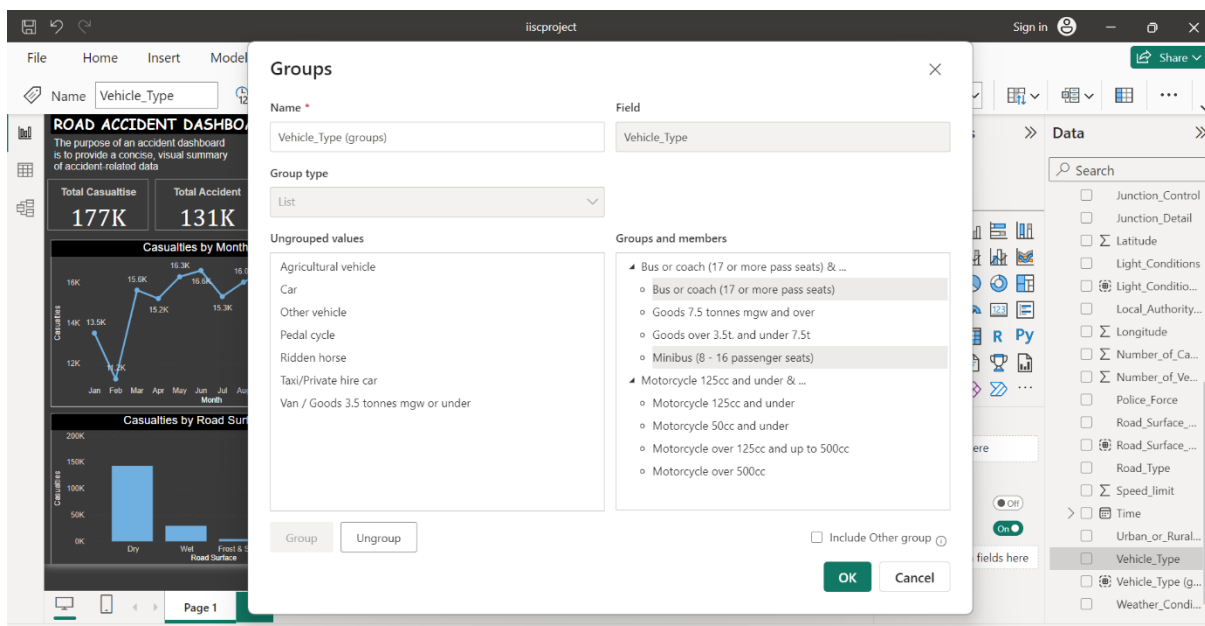


Table: Calender (730 rows) Column: Date (730 distinct values)

Date	Year	Month	Month no
01-01-2021 00:00:00	2021	Jan	1
02-01-2021 00:00:00	2021	Jan	1
03-01-2021 00:00:00	2021	Jan	1
04-01-2021 00:00:00	2021	Jan	1
05-01-2021 00:00:00	2021	Jan	1
06-01-2021 00:00:00	2021	Jan	1
07-01-2021 00:00:00	2021	Jan	1
08-01-2021 00:00:00	2021	Jan	1
09-01-2021 00:00:00	2021	Jan	1
10-01-2021 00:00:00	2021	Jan	1
11-01-2021 00:00:00	2021	Jan	1
12-01-2021 00:00:00	2021	Jan	1
13-01-2021 00:00:00	2021	Jan	1
14-01-2021 00:00:00	2021	Jan	1
15-01-2021 00:00:00	2021	Jan	1
16-01-2021 00:00:00	2021	Jan	1
17-01-2021 00:00:00	2021	Jan	1
18-01-2021 00:00:00	2021	Jan	1
19-01-2021 00:00:00	2021	Jan	1
20-01-2021 00:00:00	2021	Jan	1
21-01-2021 00:00:00	2021	Jan	1
22-01-2021 00:00:00	2021	Jan	1
23-01-2021 00:00:00	2021	Jan	1

Grouping Variables:



ROAD ACCIDENT DASHBOARD

The purpose of an accident dashboard is to provide a concise, visual summary of accident-related data.

Total Casualties: 177K
Total Accident: 131K

Casualties by Month

Month	Casualties
Jan	12K
Feb	13.5K
Mar	15.8K
Apr	15.3K
May	16.3K
Jun	16.0K
Jul	15.3K
Aug	16.0K

Casualties by Road Surface

Road Surface	Casualties
Dry	150K
Wet	50K
Frost & Ice	50K

Groups

Name: Vehicle_Type (groups)
Field: Vehicle_Type
Group type: List

Ungrouped values

- Agricultural vehicle
- Car
- Other vehicle
- Pedal cycle
- Ridden horse
- Taxi/Private hire car
- Van / Goods 3.5 tonnes mgw or under

Groups and members

- Bus or coach (17 or more pass seats) & ...
 - Bus or coach (17 or more pass seats)
 - Goods 7.5 tonnes mgw and over
 - Goods over 3.5t. and under 7.5t
 - Minibus (8 - 16 passenger seats)
- Motorcycle 125cc and under & ...
 - Motorcycle 125cc and under
 - Motorcycle 50cc and under
 - Motorcycle over 125cc and up to 500cc
 - Motorcycle over 500cc

Include Other group ☐

OK Cancel

Exploratory Data Analysis:

Create slicers :

- Create slicers for weather condition and year.
 - **Open Power BI Desktop:**
Launch Power BI Desktop application on your computer.
 - **Import Data:**
Import the dataset you want to work with by clicking on the "Get Data" button in the Home tab and selecting your data source (e.g., Excel file).
 - **Design Report Page:**
Once the data is imported, Power BI will open a new report page. Here, you'll design your report layout.
 - **Add Visualizations:**
Drag and drop fields from your dataset into the canvas to create visualizations. For example, you can add a bar chart, line chart, or table.
 - **Create Slicer:**
To create a slicer, follow these steps:
Click on the "Slicer" visualization icon in the Visualizations pane (it looks like a funnel).
Drag the field you want to use as a slicer from the Fields pane into the "Fields" section of the slicer visualization.
Power BI will automatically create a slicer based on the selected field.
 - **Customize Slicer:**
You can customize the slicer appearance and behaviour by selecting the slicer visualization on the canvas and using the options available in the Visualizations pane. For example, you can change the slicer orientation, adjust the size, or enable/disable multi-select.
 - **Interact with Slicer:**
Once the slicer is created, you can interact with it to filter other visuals on the report page. Click on different values in the slicer to filter the data displayed in other visualizations accordingly.

Slicer View:

Weather Condition				Year	
Fine	Other	Raining	Snowing/Fog	2021	2022

Create cards :

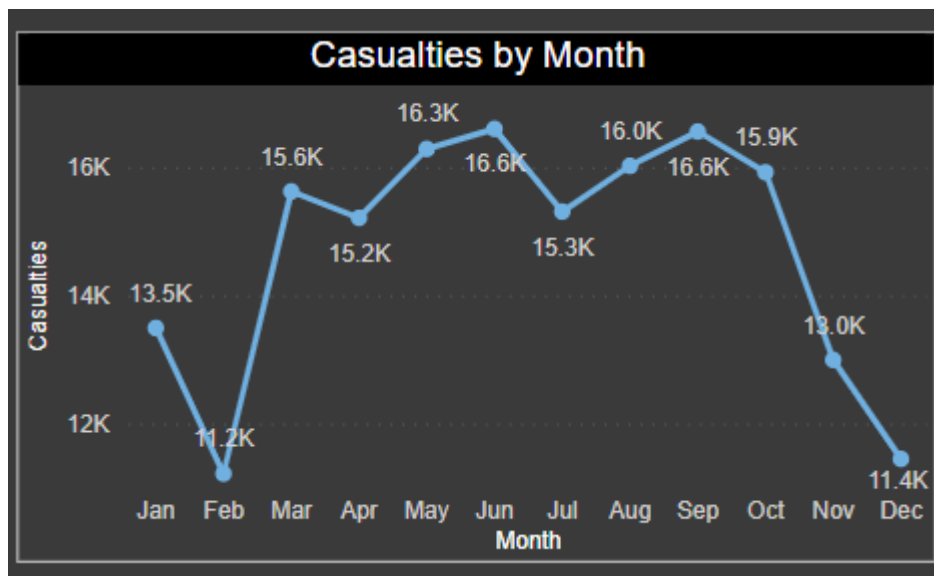
- Create cards for Total Accidents, Total Casualties, and based on vehicle type.
- **Add Cards:**
 - To create cards for Total Accidents and Total Casualties, follow these steps:
 - Click on the "Card" visualization icon in the Visualizations pane (it looks like a card).
 - Drag the "Accident_Index" field into the "Fields" section of the card visualization.
 - Power BI will automatically calculate the total number of accidents and display it in the card.
 - Repeat the same process to create a card for Total Casualties by dragging the "Number_of_Casualties" field into another card visualization.
- **Create Card for Vehicle Type:**
 - To create a card based on vehicle type, you can use a slicer combined with a card visualization:
 - Drag the "Vehicle_Type" field into the canvas to create a slicer.
 - Next, create a card visualization by clicking on the "Card" visualization icon.
 - Drag the "Vehicle_Type" field into the "Fields" section of the card visualization.
 - Power BI will automatically count the occurrences of each vehicle type and display it in the card based on the slicer selection.
 - **Customize Cards:**
 - You can customize the appearance of the cards by selecting each card visualization on the canvas and using the formatting options available in the Visualizations pane. You can adjust the font size, color, alignment, and other properties to match your report style.

Cards View :

Total Casualties	Total Accident	MOTORCYCLE	CYCLE/HORSE	GOODS CARRIER	CAR	BUS
177K	131K	15K	50	14K	141K	4886

Creating Graphs:

1. Casualties VS Month :



Interpretation:

Seasonal Variations: there are many noticeable patterns or fluctuations in casualty numbers across different months. For example, higher casualties during certain months may indicate factors like weather conditions or increased traffic.

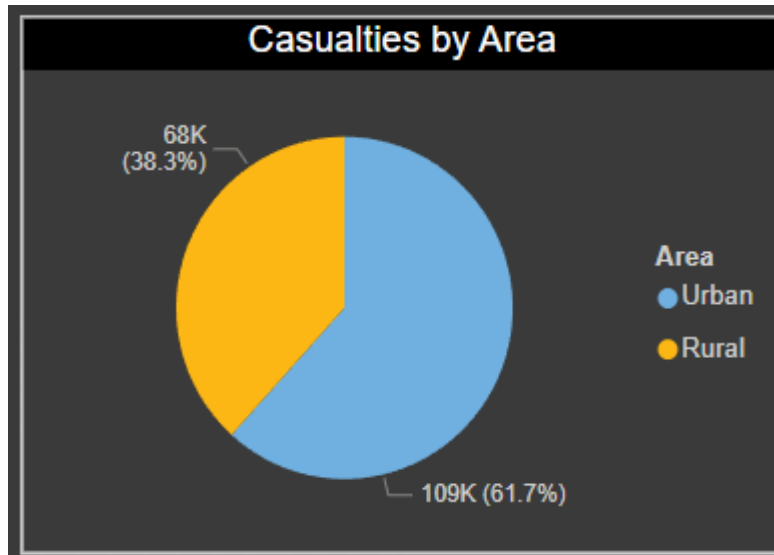
Peak Months: Identify any months with significantly higher or lower casualty numbers compared to others and provide possible explanations for these variations. Peak months are March, May, June, Aug, Sep.

Month-to-Month Comparison: Compare the casualty numbers between different months and highlight any significant differences or trends over time.

Conclusion:

Mostly accidents are caused in month of June, September, August so we are concluded that wet road type causes more accidents as compared to winter and summer road conditions.

2. Casualties VS Area:



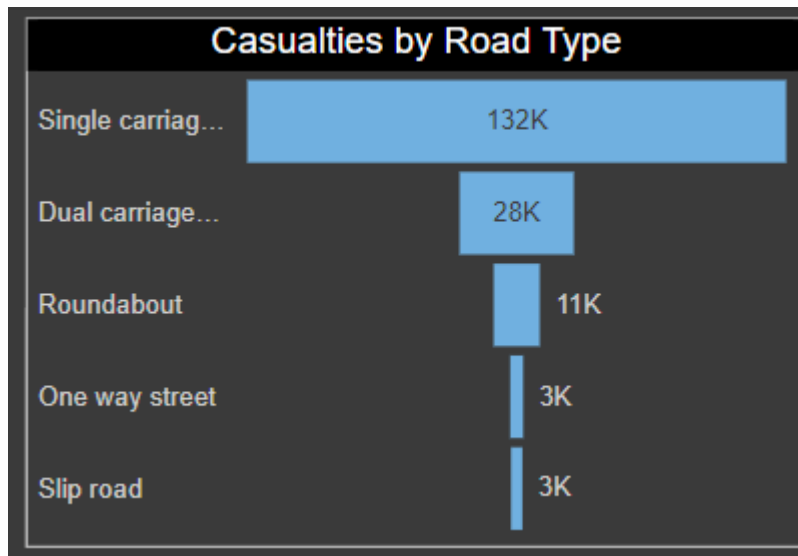
Urban vs. Rural Comparison:

Difference between urban and rural areas and discuss their significance in terms of road safety and casualty rates. This explains any differences in infrastructure, traffic volume, and driving behaviour between these areas.

Casualty Counts:

Casualties are more in urban area as compared to rural area.

3. Casualties Vs Road Type:



Description of the Analysis:

- Explain the methodology used to examine casualties by road type, such as grouping the data based on the classification of roads (e.g., motorways, dual carriageways, single carriageways).

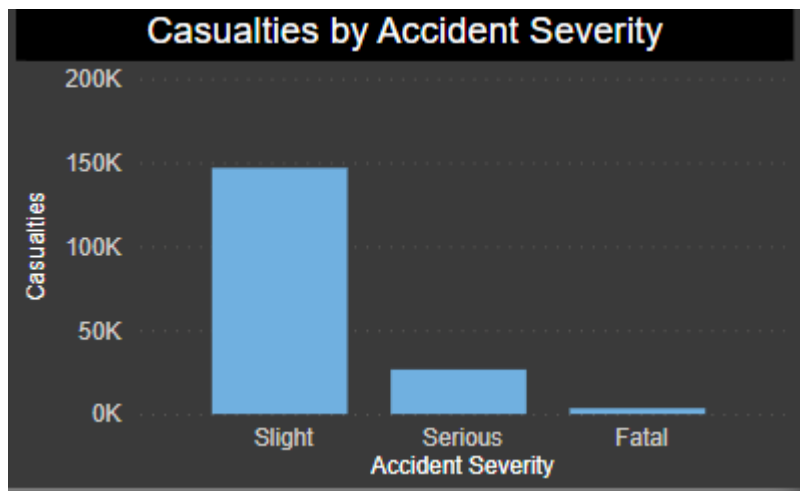
Classification of Road Types:

- Motorways: High-speed roads designed for long-distance travel, usually separated by barriers.
- Dual Carriageways: Roads with two or more lanes of traffic moving in opposite directions, often separated by a central reservation.
- Single Carriageways: Roads with a single lane of traffic in each direction, typically with lower speed limits and more frequent intersections.

Casualty Counts:

- More in Single carriage as compared to rest road types

4. Casualties VS Severity:



Description of the Analysis:

- The approach taken to categorize casualties based on the severity of their injuries, typically classified into several categories such as:
- Fatalities: Cases where individuals involved in accidents succumbed to their injuries and died.
- Serious Injuries: Cases where individuals sustained severe injuries requiring hospitalization or medical treatment.
- Slight Injuries: Cases where individuals sustained minor injuries that did not require hospitalization or extensive medical treatment.

Severity Levels:

- Fatalities: Injuries resulting in death at the scene or shortly after the accident.
- Serious Injuries: Injuries causing significant harm or impairment to individuals, often requiring extended medical care and rehabilitation.
- Slight Injuries: Minor injuries such as bruises, cuts, or whiplash that do not pose a significant threat to life or require extensive medical intervention.

Casualty Distribution:

- Present the distribution of casualties across different severity levels, either as absolute counts or percentages.
- Utilize visual aids like pie charts or stacked bar charts to illustrate the proportion of casualties in each severity category.

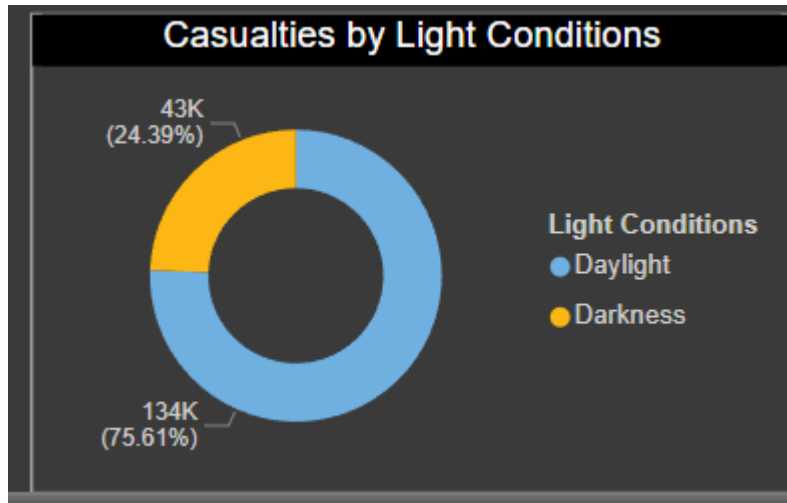
Interpretation:

Factors Influencing Severity:

- factors that may influence the severity of injuries sustained in road accidents, including:
- Vehicle speed and impact force
- Use of safety equipment (e.g., seat belts, helmets)

- Vehicle safety features (e.g., airbags, crumple zones)
- Emergency response times and quality of medical care

5. Casualties VS light Conditions:



Description of the Analysis:

- Daylight: Accidents occurring during daylight hours when natural sunlight provides adequate visibility.
- Darkness- no lighting: Accidents occurring at night or in areas with no artificial lighting, resulting in reduced visibility.
- Darkness- lighting: Accidents occurring at night but in areas with artificial lighting, such as streetlights or vehicle headlights.

Light Condition Categories:

- Daylight: Visibility is generally good, allowing drivers to see road signs, hazards, and other vehicles clearly.
- Darkness- no lighting: Reduced visibility increases the risk of accidents due to decreased perception of hazards and obstacles.
- Darkness- lighting: Although visibility may be improved compared to areas with no lighting, drivers may still encounter challenges, such as glare from oncoming headlights or shadows.

Interpretation:

- Casualties are more happened in daylight as compared to darkness

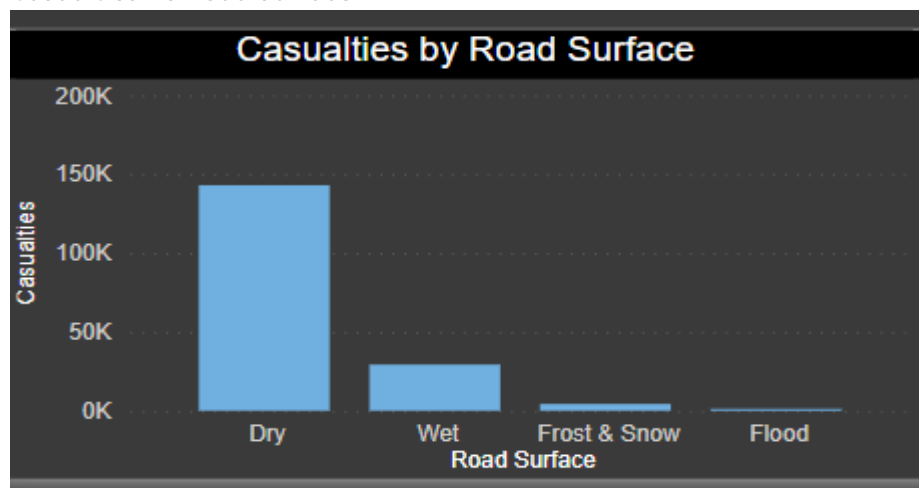
Factors Influencing Light Condition Impact:

- Time of day: Differences in traffic volume and driver behaviour during daylight and nighttime hours.
- Weather conditions: Adverse weather conditions such as fog, rain, or snow can further reduce visibility and exacerbate accident risks.
- Infrastructure: Quality and placement of streetlights, road signage, and reflective markings can affect visibility and safety.

Preventive Measures:

- Installing additional lighting in poorly illuminated areas to enhance visibility for drivers and pedestrians.
- Educating road users about the importance of adjusting driving behavior and using headlights appropriately in low-light conditions.
- Implementing technological solutions such as adaptive lighting systems or road surface treatments to improve visibility and reduce glare.

6. Casualties VS Road Surface:



Description of the Analysis:

- Dry: Road surfaces free from moisture or precipitation, providing good traction for vehicles.
- Wet or damp: Road surfaces affected by light rain, drizzle, or dew, potentially reducing traction and increasing braking distances.
- Snow-covered: Road surfaces covered by snow, which can significantly reduce traction and visibility, increasing the risk of accidents.
- Ice: Road surfaces covered by ice, posing severe hazards due to reduced traction and slippery conditions.

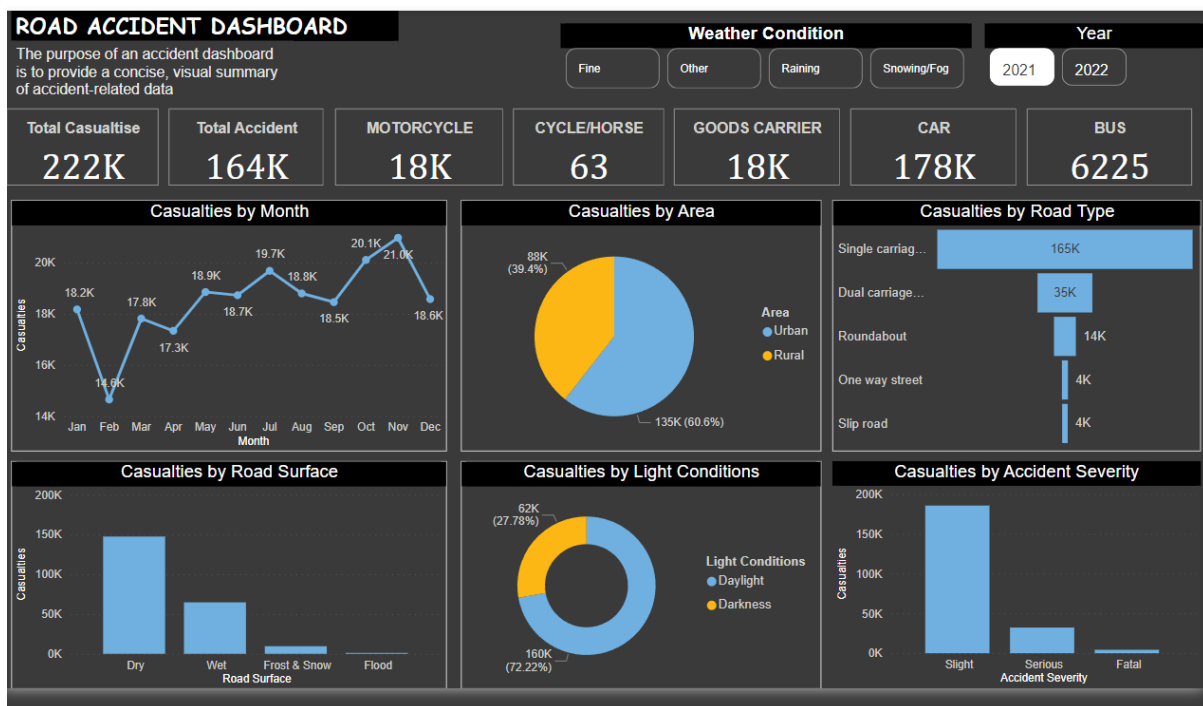
Road Surface Categories:

- Dry: Generally considered the safest condition for driving, with optimal traction and reduced risk of skidding or hydroplaning.
- Wet or damp: Increased risk of accidents due to reduced tire grip and longer braking distances, particularly during heavy rainfall or wet conditions.
- Snow-covered: Heightened risk of accidents due to decreased traction and impaired vehicle control, requiring cautious driving and slower speeds.
- Ice: Extremely hazardous conditions with minimal traction, significantly increasing the likelihood of accidents, especially if drivers fail to adapt their driving behavior.

Preventive Measures:

- Implementation of road surface treatments (e.g., anti-icing agents, sand/salt application) to improve traction and minimize the impact of adverse weather conditions.
- Public education campaigns to raise awareness of safe driving practices in various road surface conditions.

VIEW OF DASHBOARD FOR YEAR(2021):



VIEW OF DASHBORAD FOR YEAR(2022):

