DAB 201 DATA VISUALIZATION AND REPORTING FALL 2024 002

GROUP NO 10

TABLEAU PROJECT

VISUALIZATION AND REPORT

ON

CAR ACCIDENT DATA

Submitted By

STUDENT NAME

Nivedita	W0826550
Manish	W0836754
Nandeep	W0836669
Neha	W0836775
Sahil	W0837719

Chapter 1 Introduction

In today's modern world, road traffic accidents remain a significant concern, impacting communities, economies, and public safety worldwide. Understanding the underlying factors contributing to these accidents is crucial for implementing effective prevention measures and improving road safety standards. In pursuit of this objective, our project endeavors to analyze a comprehensive dataset of car accidents, aiming to uncover insights and trends that can inform evidence-based decision-making and interventions.

Project Objective and Questions to address:

The primary objective of our project is **to conduct a comprehensive analysis of road traffic accidents using the provided dataset.** By exploring the relationships and patterns within the data, we aim to

1. Identify factors associated with the frequency and severity of accidents, as well as their spatial and temporal distribution.
2. What are the most common contributing factors to road accidents, such as weather conditions, road types, vehicle types, and driver behavior?
3. Are there any patterns or trends in the occurrence of road accidents, such as variations by time of day, day of the week, or month of the year?
4. How do the characteristics of road accidents, such as severity and number of casualties, differ across different regions or jurisdictions?

Dataset Overview:

Dataset No 1 (The reason we selected it)

Website Link: Search | Kaggle

Data collected in Respect to Public Domain, Funded by Government of Royal Borough of Kensington, and Chelsea, which is in England, United Kingdom.

Our dataset encompasses a wide array of parameters related to road traffic accidents, providing a detailed perspective on various contributing factors and circumstances surrounding each incident. Key variables include,

- Accident Index, Accident Date
- Months, Year, Day of the Week
- Junction Control, Junction Detail
- Accident Severity
- Latitude, Longitude
- Light Conditions
- Local Authority (District)
- Carriageway Hazards
- Number of Casualties, Number of Vehicles
- Police Force
- Road Surface Conditions, Road Type
- Speed Limit
- Time
- Weather Conditions
- Vehicle Type

Values and Benefits to Society:

1. With a focus on understanding and addressing road safety concerns, this dataset serves as ***a valuable resource for researchers, policymakers, and stakeholders interested in mitigating the risks associated with road traffic accidents***.
2. By analyzing the data within this dataset, stakeholders can gain valuable insights into the factors contributing to accidents, ***identify trends and patterns, and formulate targeted interventions to improve road safety standards and reduce accident rates***.

Dataset No 2 (The reason we rejected it)

Website Link: [Motor Vehicle Collisions - Crashes (kaggle.com)]

After careful evaluation and analysis, we have decided to reject the second dataset on car accidents due to several reasons.

1. The dataset lacked comprehensive information, containing incomplete and missing data across various parameters essential for meaningful analysis.
2. Important variables such as accident severity, location coordinates, and vehicle types were absent or inadequately represented, hindering our ability to derive meaningful insights.
3. Additionally, the dataset exhibited inconsistencies and data quality issues, including duplicate entries, erroneous values, and formatting inconsistencies, which raised concerns about its reliability and accuracy.
4. Moreover, the lack of contextual information and metadata made it challenging to interpret the dataset effectively and assess its suitability for our analytical objectives.

Considering these limitations and the dataset's inability to provide meaningful insights or support robust analysis, we deemed it unsuitable for further consideration in our study.

Contribution of Team Members:

All Team Members have equally Contributed into this project but that's the final work done after selecting the best visualization chart or work done by team members after discussion.

Data Search	Nivedita, Manish
Data-Cleaning	Nivedita
Visualization Charts, Dashboard	Nivedita, Manish, Nandeep, Neha, Sahil
Presentation	Nandeep, Nivedita
Reporting	Chapter 1 - Introduction - Manish Chapter 2 - Methodology – Nivedita, Manish Chapter 3 - Conclusion – Manish and Nandeep

Chapter No 2 Methodology

This Project analysis will utilize advanced data visualization and statistical techniques, leveraging the capabilities of Tableau software to create insightful visualizations. By visually representing the dataset's parameters, we aim to uncover hidden patterns, correlations, and trends that may not be immediately apparent from raw data alone. Through interactive dashboards and visual representations, we intend to provide stakeholders with actionable insights and recommendations for improving road safety initiatives.

1. **Bar Chart: Number of Casualties by Local Authority District and Road Type**

Parameters Selected from Dataset: For this visualization, we have selected the following parameters from our dataset:

- **Number of Casualties:** This parameter represents the total number of casualties (injuries or fatalities) resulting from road accidents.
- **Local Authority District:** This parameter categorizes the data based on the administrative district responsible for managing road infrastructure and safety regulations.
- **Road Type:** This parameter classifies road segments based on their characteristics, such as motorways, dual carriageways, single carriageways, etc.

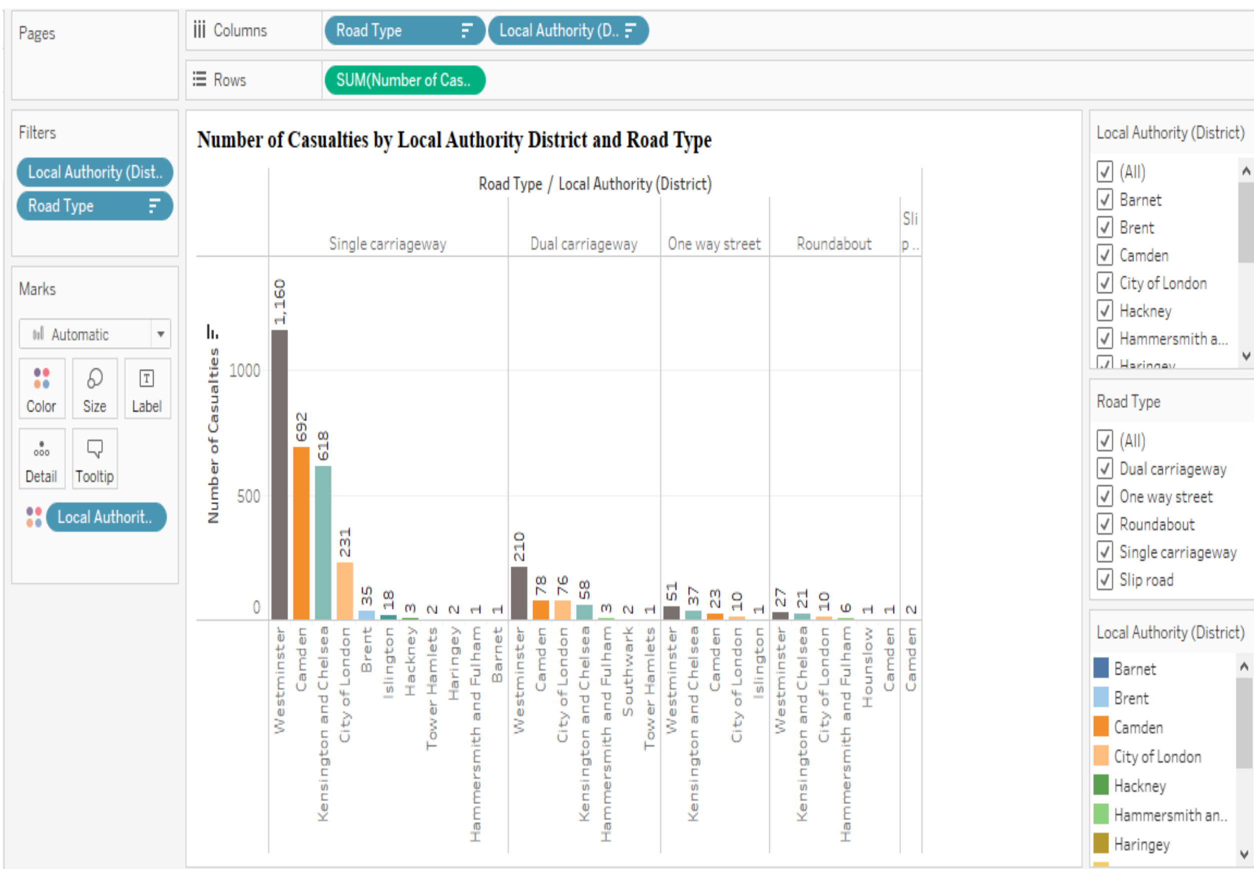
Description:

1. The bar chart provides a visual representation of the ***number of casualties across different local authority districts, segmented by road type***.
2. Each bar in the chart represents a local authority district, with multiple segments within each bar corresponding to different road types.
3. The height of each segment represents the total number of casualties recorded for that specific road type within the respective local authority district.

Suitability of Bar Chart:

1. The bar chart is well-suited for visualizing categorical data, making it an ideal choice for comparing the number of casualties across different local authority districts and road types.
2. By using ***color encoding to differentiate between road types and arranging the bars by local authority district,*** the chart effectively communicates variations in casualty counts within and between districts.
3. Additionally, the use of bars allows for easy comparison of casualty counts across categories, enabling stakeholders to identify areas with higher casualty rates and prioritize interventions accordingly.

***Screenshot of Bar Chart*:**



2. **Geographical Map: Distribution of Casualties by Location and Local Authority**

Parameters Selected from Dataset: For this visualization, we have selected the following parameters from our dataset:

- **Longitude and Latitude:** These parameters represent the geographic coordinates of each road accident location, allowing us to plot the incidents on a geographical map.
- **Local Authority:** This parameter categorizes the data based on the administrative district responsible for managing road infrastructure and safety regulations.
- **Number of Casualties (Shape):** This calculated field represents the total number of casualties (injuries or fatalities) resulting from road accidents, using different shapes to indicate the severity of each incident.

Description:

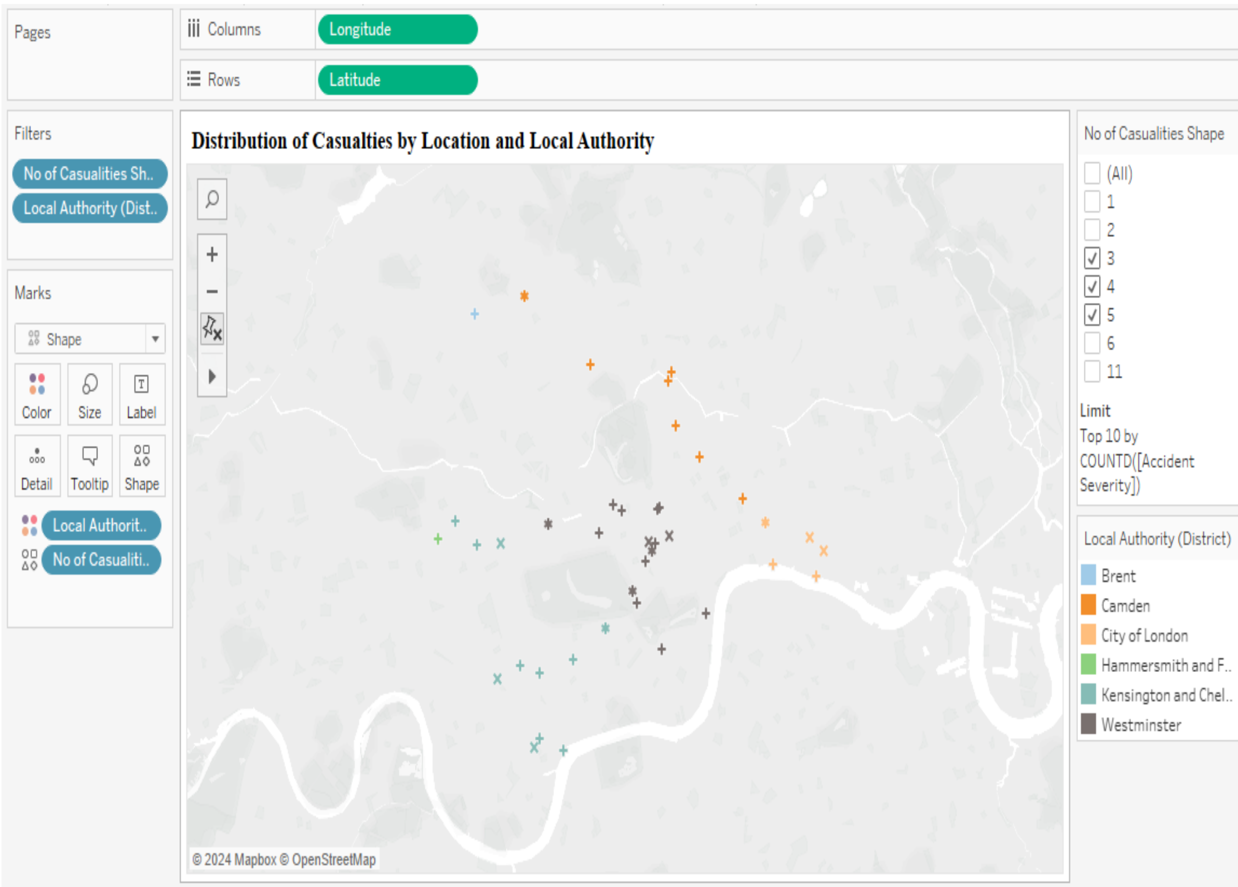
1. The ***geographical map provides a visual representation of the distribution of road accident casualties across different locations, overlaid with local authority boundaries***.
2. Each point on the map represents a specific accident location, with the color of the point indicating the local authority responsible for that area.
3. Additionally, the shape of each point reflects the severity of the incident, with different shapes used to represent varying numbers of casualties.
4. By plotting the incidents on a geographical map, stakeholders can visualize the spatial distribution of accidents and identify hotspots or areas with higher casualty rates.

Suitability of Geographical Map:

1. The ***geographical map is an ideal choice for visualizing spatial data such as accident locations, as it allows stakeholders to understand the geographic distribution of road accidents and their impact on different regions***.

2. By utilizing color encoding for local authority boundaries and shape encoding for casualty counts, the map effectively communicates complex information in a visually intuitive manner.

Screenshot of Geographical Map:



3. **Dual-Axis Chart: Relationship between Speed Limit, Road Type, and Average Number of Accidents**

Parameters Selected from Dataset: For this visualization, we have selected the following parameters from our dataset:

- **Road Type:** This parameter categorizes road segments based on their characteristics, such as motorways, dual carriageways, single carriageways, etc.
- **Local Authority District:** This parameter categorizes the data based on the administrative district responsible for managing road infrastructure and safety regulations.
- **Speed Limit:** This parameter represents the legal speed limit enforced on each road segment.
- **Average Number of Accidents:** This calculated field represents the average number of accidents occurring on each road segment, aggregated over a specific time.

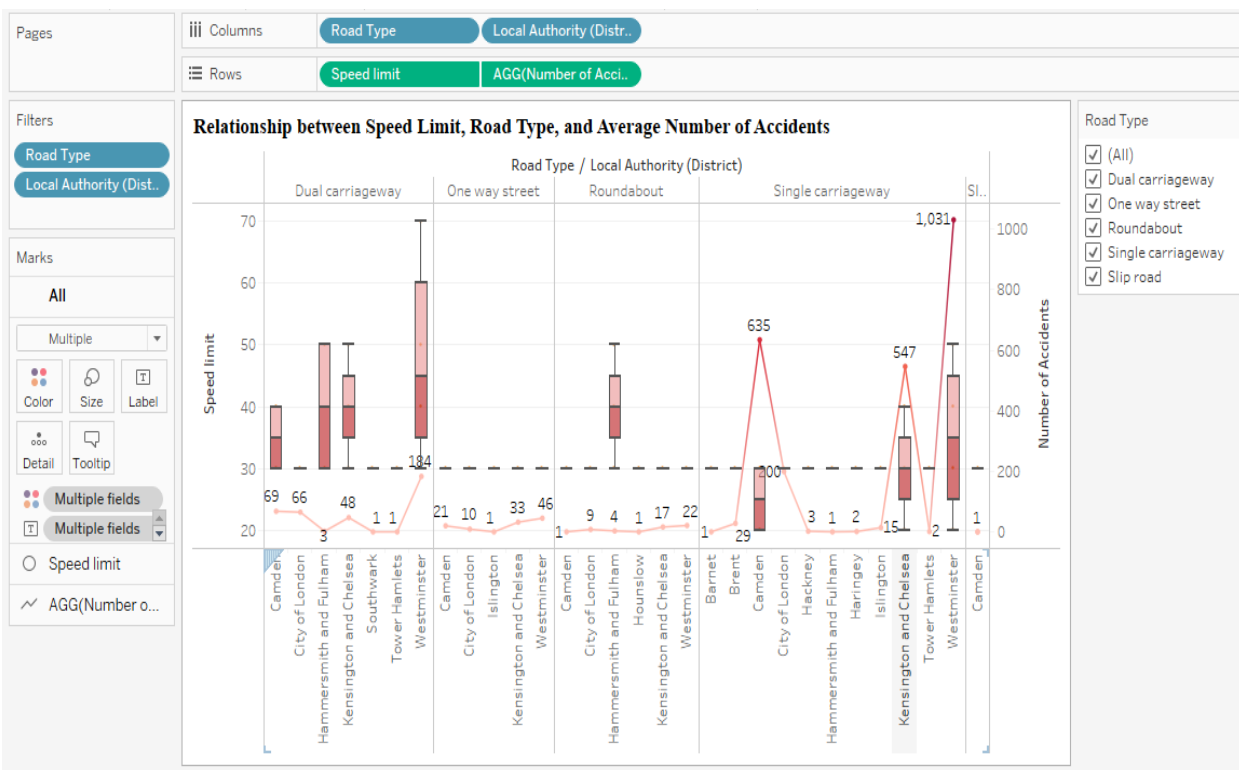
Description:

1. ***The dual-axis chart provides a visual representation of the relationship between speed limit, road type, and the average number of accidents*.**
2. The chart consists of two y-axes, with the left axis displaying the average number of accidents using a line plot, and the right axis displaying the speed limit using a box plot.
3. Each point on the line plot represents the average number of accidents for a specific combination of road type and local authority district, while each box in the box plot represents the distribution of speed limits for the same combinations.

Suitability of Dual-Axis Chart:

1. The ***dual-axis chart is well-suited for visualizing relationships between two related variables (speed limit and average number of accidents)*** across different categories (road type and local authority district).
2. By ***utilizing separate axes for the line plot and box plot***, the chart effectively communicates both the central tendency (average number of accidents) and variability (speed limit distribution) of the data.
3. Additionally, the use of color shading for the line plot and color for the box plot enhances the visual distinction between the two plots, aiding in interpretation.

Screenshot for Dual-Axis Chart:

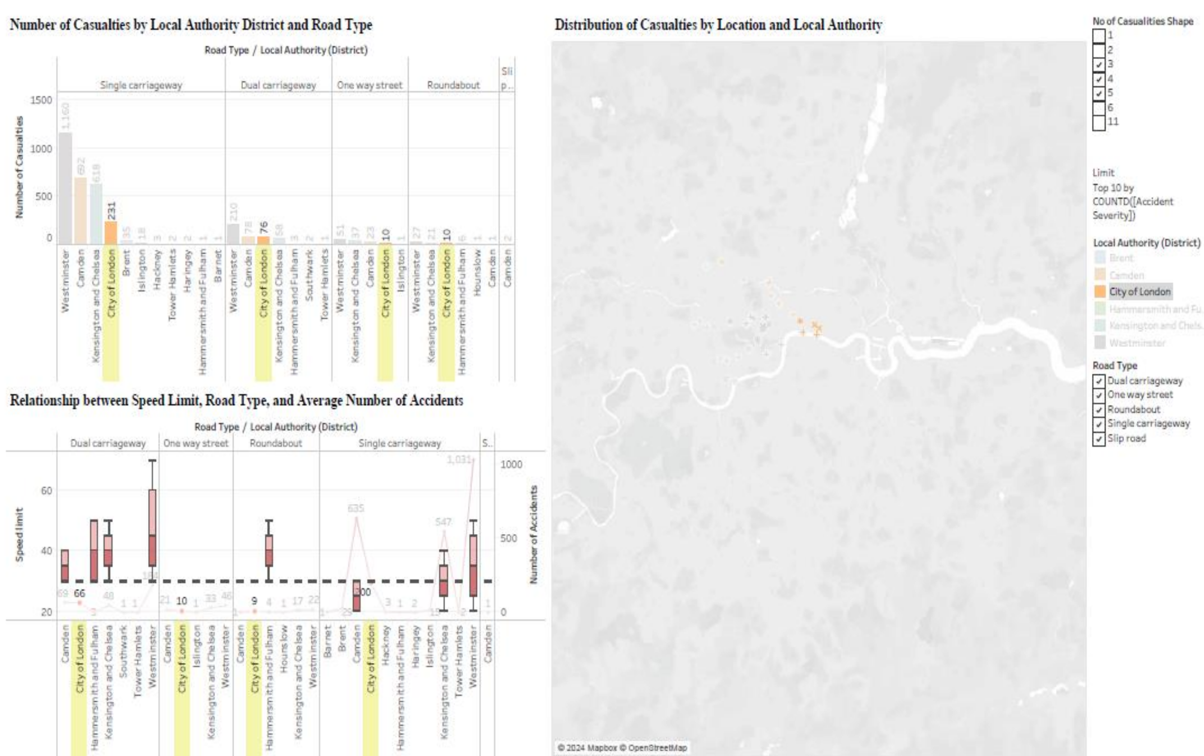


Dashboard 1

Interactivity: The dashboard provides interactive features to enhance user experience and facilitate data exploration. Users can filter the data by **Local Authority District** in the bar chart and dual-axis chart, allowing for focused analysis on specific road types. In the geographical map, users can move over data points to view detailed information about accident locations and casualty counts. Additionally, users can interact with the charts dynamically, selecting different parameters and exploring trends in real-time.

Insights: Through this dashboard, stakeholders can gain valuable insights into road traffic accidents, including spatial distribution, casualty rates, and factors influencing accident frequencies. By analyzing the visualizations, stakeholders can identify high-risk areas, assess the effectiveness of existing safety measures, and prioritize interventions to improve road safety standards and reduce accident rates.

Screenshot of Dashboard 1:



Limitations of Dashboard 1:

1. While Dashboard One provides valuable insights into road traffic accidents, it has several limitations to consider.
2. Firstly, the analysis relies heavily on reported accident data, which may not capture all incidents due to underreporting or data inaccuracies. This could lead to a skewed representation of accident frequencies and patterns, potentially overlooking less severe but still significant incidents.
3. Secondly, the dashboard primarily focuses on visualizing aggregate data at a macro level, which may obscure nuanced variations in accident characteristics at a more granular level.

4. **Line and Dot Chart: Trends in Accident Frequency and Casualty Counts by Month and Day of the Week**

Parameters Selected from Dataset: For this visualization, we have selected the following parameters from our dataset:

- **Month and Day of Week:** These parameters categorize the data based on the month and day of the week when accidents occurred.
- **Average Number of Accidents:** This calculated field represents the average number of accidents occurring on each day of the week, aggregated over a specific time period.
- **Sum of Casualties:** This parameter represents the total number of casualties (injuries or fatalities) resulting from road accidents.

Description:

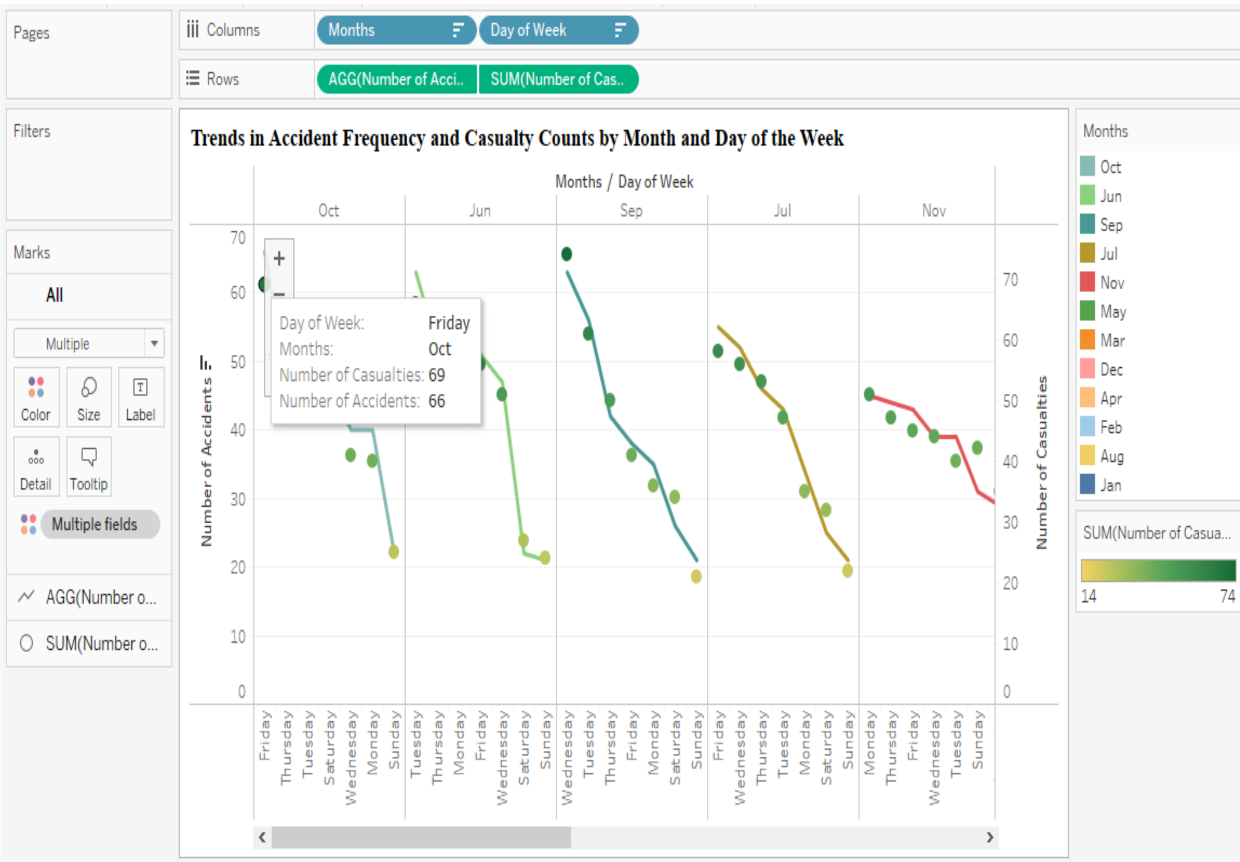
1. The line and dot chart provides a visual representation of trends in accident frequency and casualty counts across different months and days of the week.
2. The chart consists of two y-axes, with the left axis displaying the average number of accidents using a line plot, and the right axis displaying the total number of casualties using a dot plot.
3. Each point on the line plot represents the average number of accidents for a specific month, while each dot represents the total number of casualties for a specific day of the week.

Suitability of Line and Dot Chart:

1. The line and dot chart is well-suited for visualizing temporal trends and patterns in accident frequency and casualty counts, as it allows stakeholders to assess changes over time and across different categories (months and days of the week).
2. By using separate axes for the line plot and dot plot, the chart effectively communicates both the average accident frequency and the total casualty counts, providing a comprehensive overview of road safety dynamics.

3. Additionally, the use of color encoding for months and weekdays enhances the visual distinction between different time periods, aiding in interpretation and analysis.

Screenshot of Line and Dot chart:



5. **Line Chart: Hourly Trends in Accident Frequency by Day of the Week**

Parameters Selected from Dataset: For this visualization, we have selected the following parameters from our dataset:

- **Day of the Week:** This parameter categorizes the data based on the day of the week when accidents occurred.
- **Hour of the Day:** This parameter represents the hour of the day when accidents occurred, allowing for analysis of temporal patterns.
- **Average Number of Accidents:** This calculated field represents the average number of accidents occurring during each hour of the day, aggregated over a specific time period.

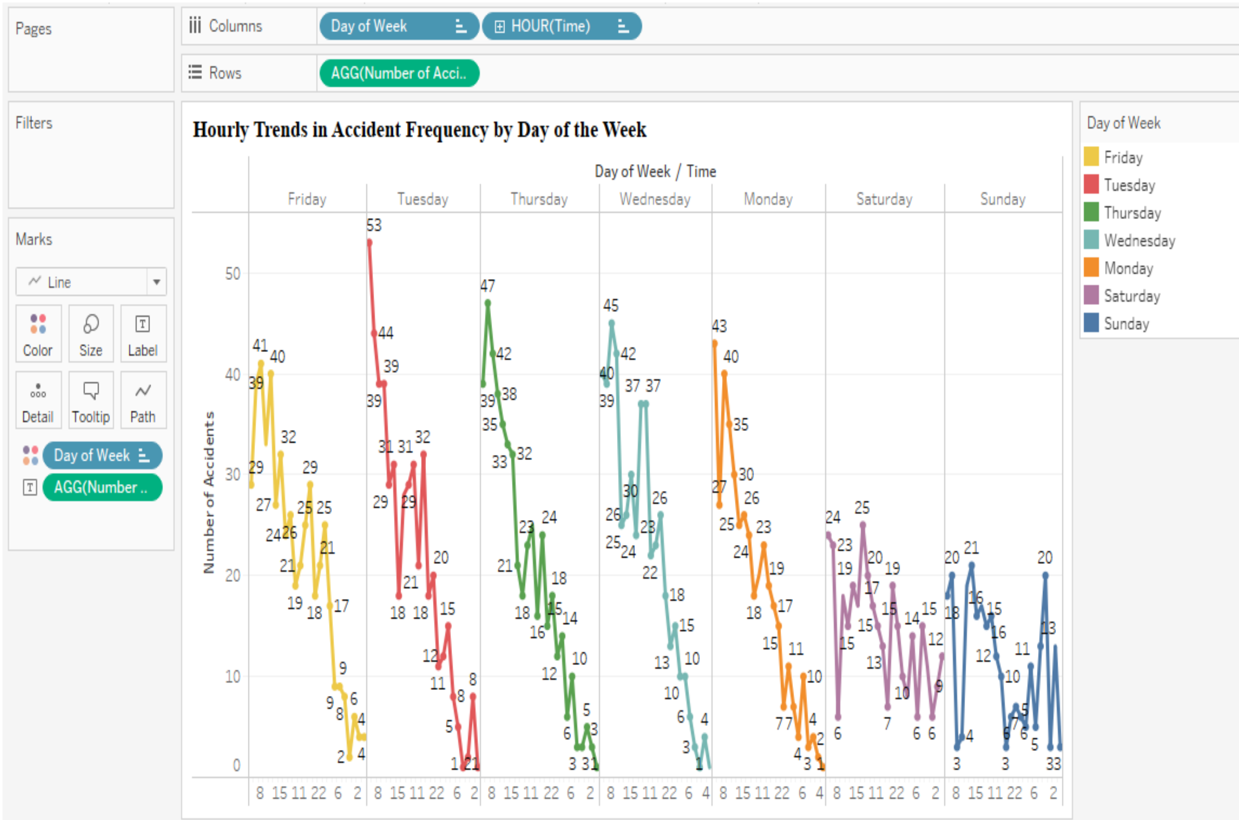
Description:

1. The ***line chart provides a visual representation of hourly trends in accident frequency across different days of the week***.
2. Each line on the chart represents a specific day of the week, with the x-axis indicating the hour of the day and the y-axis representing the average number of accidents.
3. By plotting the average number of accidents for each hour of the day, stakeholders can assess variations in accident occurrence over a 24-hour period and identify any patterns or spikes in activity.

Suitability of Line Chart:

1. The line chart is well-suited for visualizing temporal trends in accident frequency, as it effectively communicates changes over time (in this case, hourly intervals) across different categories (days of the week).
2. By using separate lines for each day of the week, the chart enables stakeholders to compare accident patterns between weekdays and weekends, identify peak hours of activity, and assess any variations in accident occurrence throughout the day.

Screenshot of Line chart:



6. **Highlighted Table: Detailed Analysis of Accident Characteristics**

Parameters Selected from Dataset: For this visualization, we have selected the following parameters from our dataset:

- **Junction Detail:** This parameter describes the specific type of junction or intersection where the accident occurred.
- **Day of the Week:** This parameter categorizes the data based on the day of the week when accidents occurred.
- **Lighting Condition:** This parameter categorizes the lighting conditions at the time of the accident, such as daylight, dusk, or darkness.
- **Vehicle Type:** This parameter classifies the vehicles involved in the accident, such as cars, motorcycles, trucks, etc.
- **Number of Accidents:** This parameter represents the total number of accidents occurring under specific conditions.

Description:

1. The ***highlighted table provides a detailed analysis of accident characteristics, focusing on specific combinations of junction detail, day of the week, lighting condition, and vehicle type***.
2. Each cell in the table represents a unique combination of these parameters, with the color indicating the number of accidents recorded under that specific condition.
3. By filtering the table to show the highest number of accidents, stakeholders can identify the most prevalent accident scenarios and explore additional factors contributing to these incidents.

Suitability of Highlighted Table:

1. The ***highlighted table is well-suited for conducting detailed analysis and exploring the interactions between multiple parameters in accident data***.
2. By organizing the data in a tabular format, the table facilitates easy comparison and interpretation of accident characteristics across different conditions.

3. The use of color encoding for the number of accidents enhances the visual distinction between different scenarios, allowing stakeholders to quickly identify high-risk conditions and prioritize interventions accordingly.

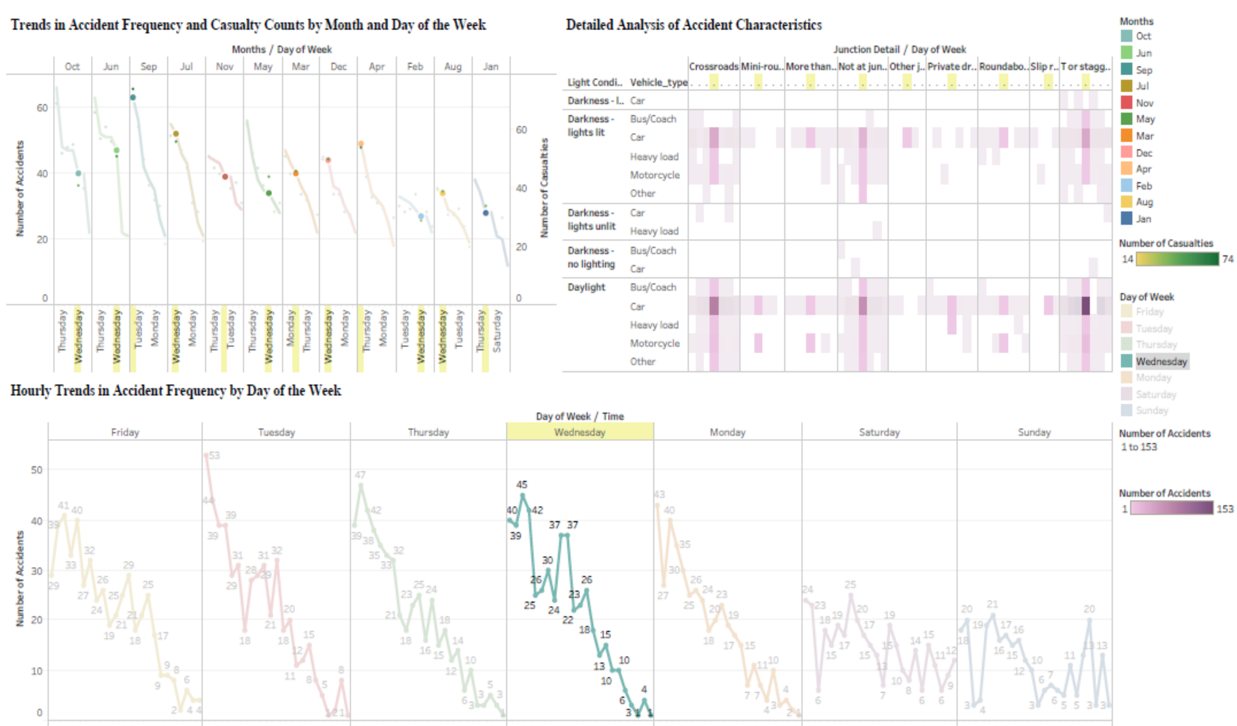
Screenshot of Table:

Dashboard 2

Dashboard Overview: This interactive dashboard offers insights into road traffic accidents through three key visualizations. The highlighted table provides detailed analysis of accident characteristics, allowing users to explore specific combinations of junction detail, day of the week, lighting condition, and vehicle type. The table facilitates granular exploration of accident scenarios, aiding in the identification of common patterns and risk factors.

Insights from Highlighted Table: The highlighted table enables stakeholders to identify prevalent accident scenarios, such as junction types associated with high accident rates on specific days of the week or under lighting conditions. This detailed analysis can inform targeted interventions and safety measures to address specific risk factors and improve road safety standards. Additionally, stakeholders can prioritize resources and allocate funding more effectively by focusing on areas with the highest accident frequencies and severity.

Screenshot of Dashboard 2:



Limitations of Dashboard 2:

1. Dashboard Two offers valuable insights into temporal patterns and spatial distributions of road traffic accidents, yet it has its limitations.
2. Firstly, the analysis is constrained by the availability and quality of data, particularly regarding accident location accuracy and completeness.
3. Inaccurate or incomplete location data could compromise the reliability of spatial analyses and hotspot identification.
4. Secondly, the dashboard primarily focuses on visualizing temporal trends and spatial distributions, overlooking other potential factors contributing to accidents, such as road conditions, driver behavior, or vehicle characteristics. Incorporating additional contextual variables could provide a more comprehensive understanding of accident dynamics and inform targeted interventions.

Chapter 3 Conclusion

Conclusion:

In conclusion, our project endeavors to contribute to the ongoing efforts aimed at reducing road traffic accidents and promoting safer road environments. By leveraging the power of data analysis and visualization, we aspire to provide valuable insights that can inform evidence-based decision-making and support initiatives aimed at enhancing road safety standards and saving lives.

From the dashboard 1,

It is evident that ***Westminster local authority experiences the highest number of accidents, totaling 1031, with casualties numbering at 1160***.

These incidents predominantly occur on single carriageways with a speed limit of 20 mph. Additionally, the data highlights the need for targeted interventions in Westminster to address road safety concerns and reduce accident rates.

Moreover, insights from this analysis underscore the importance of enforcing speed regulations and implementing safety measures on single carriageways to mitigate casualties and improve overall road safety.

From the dashboard 2,

The highest number of accidents occurs on ***Tuesday, Wednesday, and Thursday evenings, particularly between 5 to 6 pm***. These incidents ***predominantly happen in daylight*** conditions involving cars, and the ***road junctions involved are primarily observed as T or staggered junctions***. This insight underscores the importance of targeted safety measures during peak traffic hours and emphasizes the need for enhanced visibility and caution at T or staggered junctions during daylight hours.

Chapter 4 References

1. How To Create Data Reports That Will Skyrocket Your Business Performance, By Bernardita Calzon in Reporting, Mar 22nd 2024

   (25 Dashboard Design Principles, Best Practices & How To's (datapine.com))

2. Dashboard examples: The good, the bad and the ugly

   (Dashboard examples: The good, the bad and the ugly (matillion.com))

3. Best Practices for Designing Dashboards

   (Best Practices for Designing Dashboards | DataCamp)

4. Batt, S., Grealis, T., Harmon, O., & Tomolonis, P. (2020). Learning Tableau: A data visualization tool. *The Journal of Economic Education*, *51*(3–4), 317–328. (https://doi.org/10.1080/00220485.2020.1804503)

5. Analyzing Traffic Accident and Casualty Trend Using Data Visualization, By Abdullah Sakib, Saiful Adli Ismail, Haslina Sarkan, Azri Azmi & Othman Mohd Yusop (Analyzing Traffic Accident and Casualty Trend Using Data Visualization | SpringerLink)
6. Chai, C. P. (2020). The Importance of Data Cleaning: Three Visualization Examples. *CHANCE*, *33*(1), 4–9. (https://doi.org/10.1080/09332480.2020.1726112)