***DSCI 5360.006 – DATA VISUALIZATION FOR ANALYTICS***

***GROUP 4***

***Understanding Road Safety Trends: An Analysis of Road Accidents From January 2021 To December 2022***

**Team Members:**

**Puja Madhav-11714654**

**Taruni Madugula-11711823**

**Saadana Shyam Sunder-11724719**

**Deepthi Sree Boonapalli-11640627**

**Hari Charan Reddy Karra-11714445**

**Kanaka Sri Brahmini Bandla-11595731**

**Introduction and Background:**

Road accidents remain a persistent challenge worldwide, causing significant human and economic losses. Understanding the trends and patterns of road accidents is crucial for developing effective strategies to improve road safety. This proposal aims to analyze road accident data spanning from January 2021 to December 2022, providing insights into the factors influencing accident occurrence and severity in an urban area.

**Brief History and Context:**

Road safety has been a longstanding concern globally, with efforts to mitigate accidents dating back several decades. Despite advancements in vehicle technology and infrastructure, road accidents continue to claim numerous lives and impose substantial societal costs. The period from January 2021 to December 2022 witnessed various socio-economic changes, technological advancements, and shifts in transportation patterns, all of which may have influenced road safety outcomes.

**Motivation for Topic Selection:**

The motivation for investigating road safety trends over the period from January 2021 to December 2022 arises from the need to understand the evolving dynamics of road accidents. By analyzing comprehensive data covering multiple years, we aim to identify long-term trends, seasonal variations, and emerging patterns in accident occurrence and severity. This knowledge is essential for formulating targeted interventions and policy measures to address specific challenges and improve overall road safety.

**Sources of Inspiration:**

Our motivation is inspired by a combination of academic research, policy reports, and real-world observations. Studies conducted by organizations such as the International Road Assessment Programme (iRAP) and the Institute of Transportation Engineers (ITE) offer valuable insights into the factors contributing to road accidents and effective countermeasures. Additionally, media coverage of road safety issues and public discourse surrounding transportation policies inform our understanding of the broader societal context in which road accidents occur.

**Conclusion:**

In conclusion, the proposed analysis of road accidents from January 2021 to December 2022 represents a comprehensive effort to understand road safety trends and inform evidence-based decision-making. By leveraging detailed data on accident characteristics and contextual factors over a multi-year period, we seek to identify underlying patterns, assess the effectiveness of existing interventions, and develop targeted strategies to enhance road safety outcomes. Through collaboration with stakeholders across sectors, we aim to create safer road environments and reduce the incidence and severity of accidents, ultimately saving lives and improving quality of life for all road users.

**Objectives and Goals:**

**Learning and Accomplishment:** The primary objective of this project is to gain insights into the trends, patterns, and contributing factors of road accidents occurring from January 2021 to December 2022. Specifically, we aim to:

* Identify temporal trends in accident occurrences, including seasonal variations and long-term changes.
* Analyze the spatial distribution of accidents to identify hotspots and assess geographic patterns.
* Investigate the factors associated with different levels of accident severity, including road conditions, weather, and junction characteristics.
* Assess the impact of socio-economic factors, such as urbanization and population density, on road safety outcomes.
* Explore the relationship between vehicle types, speed limits, and accident severity.

**Questions to Answer:** Throughout the analysis, we seek to answer several key questions, including:

* What are the temporal trends in road accidents over the study period?
* Are there specific geographic areas or road segments with disproportionately high accident rates?
* What factors contribute most significantly to the occurrence and severity of road accidents?
* How do weather conditions, road surface conditions, and lighting affect accident likelihood and severity?
* Are certain vehicle types or road users more prone to accidents, and if so, why?

**Project Benefits:** This project has the potential to benefit various stakeholders, including policymakers, urban planners, transportation agencies, and the general public. The insights gained from the analysis can:

* Inform evidence-based decision-making and policy formulation aimed at improving road safety.
* Enable targeted interventions and resource allocation to high-risk areas and vulnerable road users.
* Enhance public awareness and education initiatives regarding road safety and accident prevention.
* Facilitate the development of innovative technologies and infrastructure improvements to mitigate accident risks.
* Ultimately, contribute to the reduction of injuries, fatalities, and economic losses associated with road accidents, thereby improving the overall quality of life for communities, and enhancing societal well-being.

Top of Form

**Dataset Description:**

**Source:** The dataset utilized in this analysis is sourced from Kaggle, a platform renowned for sharing datasets relevant to various domains, including transportation and road safety. Specifically, this dataset provides a simulated collection of records detailing road accidents that purportedly occurred between January 2021 and December 2022.

**Format:** Structured in the form of a CSV (Comma-Separated Values) file, the dataset is well-suited for seamless importation into statistical software, facilitating comprehensive analysis.

**Participants/Cases:** Comprising a total of 300,000 records, each entry within the dataset represents a unique road accident case, providing ample data for robust analysis and insights.

**Variables of Interest:** Notable variables included in the dataset encompass a wide array of factors pertinent to road accidents, such as:

* **Accident\_Index:** A unique identifier for each accident record.
* **Accident Date:** The date on which the accident occurred (format: DD/MM/YYYY).
* **Day\_of\_Week:** The day of the week when the accident took place.
* **Junction\_Control** : Describes the type of junction control at the accident location (e.g., "Give way or uncontrolled").
* **Junction\_Detail:** Provides additional details about the junction where the accident occurred (e.g., "T or staggered junction").
* **Accident\_Severity:** Indicates the severity of the accident (e.g., "Serious").
* **Latitude:** The geographic latitude of the accident location.
* **Light\_Conditions:** Describes the lighting conditions at the time of the accident (e.g., "Daylight").
* **Local\_Authority\_(District)**: The local authority district where the accident occurred.
* **Carriageway\_Hazards:** Describes any hazards present on the carriageway at the time of the accident (e.g., "None").
* **Longitude:** The geographic longitude of the accident location.
* **Number\_of\_Casualties:** The total number of casualties involved in the accident.
* **Number\_of\_Vehicles:** The total number of vehicles involved in the accident.
* **Police\_Force:** The police force that handled the accident.
* **Road\_Surface\_Conditions:** Describes the surface conditions of the road at the time of the accident (e.g., "Dry").
* **Road\_Type:** Specifies the type of road where the accident occurred (e.g., "One way street").
* **Speed\_limit:** The speed limit applicable to the road where the accident occurred.
* **Time:** The time of day when the accident happened (format: HH:MM).
* **Urban\_or\_Rural\_Area:** Indicates whether the accident occurred in an urban or rural area.
* **Weather\_Conditions:** Describes the weather conditions at the time of the accident (e.g., "Fine no high winds").
* **Vehicle\_Type:** Specifies the type of vehicle involved in the accident (e.g., "Car," "Taxi/Private hire car").

**Data Collection and Funding:** As a demo dataset, it is important to note that the data may not have been collected as part of a funded project. Instead, it serves educational and analytical purposes, mimicking real-world scenarios to facilitate learning and experimentation in the field of data analysis and road safety.

**Publications and Ideas:** While there may not be specific publications directly associated with this dataset, it serves as an invaluable resource for researchers and analysts interested in studying road safety trends and patterns. The dataset can be utilized to develop innovative analytical approaches, identify key factors contributing to accidents, and devise strategies for mitigating road safety risks.

**Scope and Dimension: Locations Included:** While the precise locations covered in the dataset are not explicitly specified, the inclusion of geographic coordinates (latitude and longitude) enables mapping of accidents at various spatial scales, from local to regional and potentially national levels.

**Scope and Dimension: Timeframe Included:** The dataset spans the period from January 2021 to December 2022, allowing for temporal analyses to uncover trends, seasonal patterns, and temporal variations in road accident occurrences. Additionally, the inclusion of date and time variables facilitates the exploration of accident timelines and temporal dynamics.

**Utility and Constraints of Datasets in Underpinning Project Objectives**

Our project relies on datasets to elucidate road accident patterns and associated factors, aligning with our core objectives. However, we must recognize the inherent utility and constraints within these datasets:

**Utility:**

1. **Informed Decision-Making:** The datasets serve as foundational pillars for informed decision-making processes. They offer comprehensive insights into accident severity, temporal trends, junction control types, weather conditions, road surface conditions, road types, vehicle types, location, and casualties, aiding in the formulation of effective road safety strategies and interventions.
2. **Identification of Hotspots:** Through meticulous analysis of accident locations, the datasets enable the identification of geographical hotspots with heightened accident rates. This geographical insight facilitates targeted interventions and resource allocation to enhance safety measures in these areas.
3. **Temporal Trend Analysis:** Leveraging the datasets for time series analysis allows us to discern temporal trends such as seasonal variations, peak accident periods, and long-term changes. This temporal understanding serves as a compass for devising strategies that address dynamic road accident patterns effectively.
4. **Comparative Examination:** By juxtaposing various variables within the datasets, such as accident severity, weather conditions, road types, and vehicle types, we uncover correlations and patterns that inform tailored interventions. This comparative approach enhances the efficacy of our strategies by addressing specific risk factors associated with different contexts.

**Constraints:**

1. **Data Quality Concerns:** The utility of the datasets is contingent upon the quality and reliability of the data. Issues such as inaccuracies, inconsistencies, and missing values may undermine the validity of analyses and subsequent decision-making processes.
2. **Limited Scope:** While comprehensive, the datasets may have limitations in terms of depth and breadth. Critical factors influencing road accidents, such as driver behavior, road maintenance practices, and vehicle safety features, may not be adequately captured, restricting the breadth of analysis.
3. **Temporal and Spatial Granularity:** The granularity of the data, both temporally and spatially, may vary, impacting the depth of analysis. Limited temporal resolution hampers the identification of short-term trends, while coarse spatial resolution obscures localized variations in accident patterns.
4. **Missing parameters:** Missing parameters such as age groups are critical considerations when utilizing accident data. These omissions can hinder the analysis by obscuring age-specific patterns and risk factors. Addressing these gaps is essential for developing comprehensive road safety strategies and ensuring equitable outcomes.

Top of Form

**Visualizations:**Top of Form

**1) Accident Severity Distribution by Day of the Week (Bar Chart):**

Variables: Accident severity (from dataset), Day of the week (from dataset)

Description: This chart will show the distribution of accident severity (e.g., serious, slight) across different days of the week.

Question: Do certain days of the week exhibit a higher proportion of slight accidents compared to others?

A screenshot of a computer

Description automatically generated

This bar chart breaks down the severity of accidents across different days of the week, making it easy to compare. Let's take a closer look:

**What are your findings from the graph?**

* + Slight Injuries: If you look at Fridays, you'll notice that there were more minor accidents compared to other days. Fridays seem to have the highest number of slight injuries throughout the week.
  + Serious Injuries: There were around 6412 serious accidents on Fridays, which is quite a lot. This means Fridays are days when serious accidents happen more frequently than other days.
  + Fatal Injuries: Saturdays stand out here. While fatal injuries overall are around 610, Saturdays have the highest number of fatal accidents, totaling 704. This suggests that Saturdays might be riskier days on the road.
  + Fatal Injuries Severity: There were only 7 accidents classified as having fatal injuries on Fridays, and Wednesdays, which is noticeably lower than Tuesday.

**What can be done to prevent them?**

This information is crucial for the traffic department to take appropriate action. For instance, they might want to increase patrols or awareness campaigns on Fridays to prevent slight and serious accidents. Similarly, they could focus on Saturdays to address the higher rate of fatal accidents. By understanding these patterns, they can better allocate resources and improve road safety for everyone.

**Design Principles:**

* Clarity: Ensure the visualization clearly communicates the distribution of accident severity across different days of the week.
* Simplicity: Keep the design clean and straightforward, avoiding unnecessary clutter or complexity.
* Color Contrast: Use distinct colors for each severity category to make them easily distinguishable.
* Labeling: Provide clear labels for the axes and data points to enhance understanding.
* Consistency: Maintain consistency in design elements throughout the visualization to avoid confusion.

**2) Speed Limit in Rural and Urban vs Accident Severity**

* Variables: Accident severity, speed limit, Urban or Rural Area, Road Accident(count)
* Description: Through this area graph we will be exploring the relationship between the speed limit in urban and rural areas and the accidents that have occurred and their severity.
* Question: What are the predominant speed limits associated with fatal accidents in rural and urban areas, and how do they differ?

A screenshot of a graph

Description automatically generated

**Explain the graph:**

The area graph is divided into 2 parts, one for Rural and another for Urban. X – Axis represents the speed limit in the respective regions, and Y – Axis represents the count of road accidents in rural and urban regions. The accident severity is divided into 4 types Fatal, Fetal, Serious, and Slight.

**What are your findings from the graph?**

* Regarding the Fatal accidents most of the fatal accidents in the rural area have occurred where the speed limit is 60m/hr and 70m/hr, on the contrary in the urban area most of the fatal accidents have occurred where the speed limit is 30m/hr.
* There were 0 Fetal accidents in the rural area and 49 Fetal accidents in the urban region which have occurred are caused where the speed limit is 30m/hr.
* Serious and slight injuries in the rural and urban areas are kind of similar in the urban areas most of the accidents have occurred at speed limits 30m/hr. In the Rural areas a large no of accidents can be seen at 30, 60 and 70m/hr speed limit.

**What can be done to prevent them?**

* Putting Accident prone area boards on the sides of the road in areas where a major portion of the accidents have been occurring will increase awareness of the drivers and make them thread slowly in those areas.
* Reducing the Speed limits in the areas where the large no of accidents have been occurring, can contribute towards the decrease in the no of accidents and the increase the safety of the people.

**Design Principles:**

* Clarity: Ensure the visualization effectively communicates the relationship between speed limits, accident severity, and urban/rural areas.
* Differentiation: Use distinct colors or patterns to differentiate between different accident severities and urban/rural areas for easy understanding.
* Consistency: Maintain consistency in labeling and design elements to avoid confusion and aid interpretation.
* Focus: Highlight the key insights and findings to guide the viewer's attention towards important trends and patterns.
* Simplicity: Keep the design clean and uncluttered to enhance readability and comprehension.

**3) Time Series Analysis of Accidents**

Variables: Accident date, Road Accident(count)

Description: Through this line graph we will be exploring the no of accidents that have occurred during the years 2021 and 2022

Question: What are the trends in the number of road accidents over the years 2021 and 2022, and are there any notable fluctuations or patterns observed within this time frame?

A graph on a computer screen

Description automatically generated

**Explain the graph:**

The line graph shows the number of accidents that have occurred during the years 2021 and 2022. The x – axis represents the years, and y – axis represents the count of road accidents.

**What are your findings from graph?**

From the line graph we can observe that the number of accidents followed a upward trend until the month of November during both the years followed by a seep decrease for the month of December.

At the start of the year 2021 the number of accidents started at 13,417, it has reached a peak of 15,473 and it had ended with 13,709, for the year 2022 it has started at 9,967, reaching a peak of 13,622 and it had ended at 9,625.

**Have the accidents reduced from 2021 to 2022 or they have increased?**

By looking at both the numbers we can conclude that the number of accidents had decreased compared to 2021. Which indicates that necessary steps have been put in place to reduce the frequency of accidents.

**Design Principles:**

* Clarity: Ensure the visualization effectively communicates the trend in the number of road accidents over the specified time frame.
* Simplicity: Keep the design clean and uncluttered to enhance readability and comprehension.
* Focus: Highlight the key insights and trends in the data to guide the viewer's attention towards important patterns.
* Accuracy: Present the data accurately without distortion or misrepresentation.
* Context: Provide relevant context or annotations to aid interpretation and understanding of the graph.

**4) Number of accidents frequency per quarter and Year-over-Year Percentage Variations**

**A screenshot of a computer

Description automatically generated**

The variables used in the time series analysis of accident visualization are as follows:

* **YEAR (Accident Date)**: Represents the year in which the accidents occurred. It likely corresponds to the x-axis, indicating the time dimension.
* **QUARTER (Accident Date)**: Represents the quarter of the year (Q1, Q2, Q3, Q4) in which the accidents occurred. It also contributes to the x-axis, providing finer granularity within each year.
* **CNT (Number of Casualties)**: This variable likely represents the count of casualties (injuries or fatalities) resulting from each accident. It is used for the y-axis, indicating the magnitude of accidents.
* **CNT (Number of Casualties per Accident):** Represents the average number of casualties (injuries or fatalities) per accident. It helps assess accident severity. The quick table calculation computes year-over-year growth, comparing this average between consecutive years.

**Questions that can be answered from the Time Series Analysis:**

**1Q) How did the quarterly accident count in 2022 compare to the corresponding quarters in 2021, and what were the year-over-year percent changes?**

The visualization effectively showcases the quarterly trends in accident counts for 2021 and 2022, along with the year-over-year percent changes.

In the first quarter (Q1) of 2022, there was a notable 11.515% decrease in accident counts compared to Q1 of 2021. This significant reduction in casualties suggests that safety measures or other factors were effective in improving road safety during this period.

While the accident count increased from Q1 to Q2 in 2021, there was a smaller 9.312% decrease in Q2 of 2022 compared to the same quarter in the previous year. However, it's important to note that the accident count in Q2 2022 was still higher than Q1 2022, indicating a potential seasonal or periodic increase in casualties.

The downward trend continued in the third quarter (Q3) of 2022, with a 9.170% decrease in accident counts compared to Q3 2021. This sustained reduction in casualties reinforces the effectiveness of the safety measures or contributing factors during this period.

Interestingly, the fourth quarter (Q4) of 2022 exhibited the most substantial decrease in accident counts compared to the previous year, with a 16.4379% reduction. This significant improvement in road safety during the final quarter of 2022 is particularly noteworthy and may warrant further investigation into the specific factors or interventions that contributed to such a substantial decrease in casualties.

Overall, the visualization highlights the positive impact of safety measures or other factors in reducing accident counts throughout 2022, with the most substantial improvements observed in Q1 and Q4. However, it also reveals potential periodic or seasonal variations in accident counts, as evident from the quarter-to-quarter fluctuations within each year.

**Design Principles**

The design principle for the visualization provided focuses on comparing the number of accidents between two consecutive years, quarter by quarter. Here’s a breakdown of its key elements:

**Dual Graphs:** It uses a line graph to show the actual numbers and a bar graph to display the percentage change.

**Data Representation:** The line graph details accident counts for each quarter in 2021 and 2022, while the bar graph shows the percentage decrease from Q1 to Q3.

**Interactivity:** Filters and column selectors are available for customizing the view

**5) By Accident Severity and Location (District)**

A screenshot of a computer screen

Description automatically generated

**Variables used in the chart:**

* **Columns:** (Longitude)
* **Rows:** (Latitude)

**Filter used in the chart: (**Accident Severity Filter: True) to toggle between different accident severity types Fatal, Serious, slight & Fetal.

**Variables used in the marks shelf:**

* **Color: (**Accident Severity parameter)
* **Size:** Sum (Number of casualties)
* **Detail:** Local Authority (District), Sum(Number of casualties)**,** & Sum(Number of vehicles)

**Text in tooltip is achieved using the following statements.**

Local Authority (District): **<Local Authority (District)>**

Latitude: **<AVG(Latitude)>**

Longitude: **<AVG(Longitude)>**

Number of Casualties: **<SUM(Number of Casualties)>**

Number of Vehicles: **<SUM(Number of Vehicles)>**

**Process of creating the visualization:**

**STEP 1)** By placing Longitude in Columns and Latitude in Rows we will be able to get the map which is the initial step for the geospatial map. Next, we place Local Authority (District) in the Detail card to map the districts on the map.

**STEP 2)** Then we place accident severity in the Colour card to differentiate between the types of accident severity, and in the filter to interact with the accident severity variable. Then we place Sum (Number of casualties)**,** & Sum(Number of vehicles) in the detail card to get the information necessary in the map.

**STEP 3)** Finally, we place the Sum(Number of casualties) on the Size card, which allows the circles representing casualties to adjust based on the severity of accidents: Fatal, Serious, Slight, and Fetal

**Applying Design Principles:**

**Bubble Size:** The size of each circle corresponds to the sum of number of casualties in the local authority district. Larger bubbles indicate a higher number of casualties in that district. The bubbles provide a visual representation of the distribution of accidents across local authority districts, with each bubble indicating both the accident severity and number of casualties in a particular district.

**Bubble Color:** The color of each bubble represents the severity of the accidents. Red bubbles represent fatal accidents, orange bubbles represent fetal accidents, blue bubbles represent serious accidents, and green bubbles represent slight accidents.\

**Questions that can be asked from the visualization.**

**1Q)** Create a symbol map which displays the number of casualties along with the number of vehicles that were involved in the accident and it’s severity type?

**2Q)** Which regions of theGreat Britan can we observe a major no of casualties across all accident severity types?

**3Q)** By changing the accident severity types, what differences are taking place in the symbols map and the measure scale of the Sum (Number of casualties)?

**Insights that can be gained from the visualization.**

Across all the accident severity types that had occurred in Great Britan, Fetal accidents have occurred only in England. For fatal, Serious, and slight accident severity, majority of them have occurred in England. Whereas Wales and Scotland have their fair share of accidents compared to England they are very low. Coming to Northen Ireland they have no accidents across the 4 accident severity types.

**6) Number of Accidents occured in Morning and Evening with Number of Casuality**

A screenshot of a computer

Description automatically generated

**Variables used in the chart:**

**Columns:**

* **Accident Date:** The date on which the accident occurred (format: DD/MM/YYYY).

**Rows:**

* **Morning or Evening:** It defies whether the accident occurred during morning, or Evening, this variable is created using the following formulae.
* IF DATEPART('hour', [Time]) < 12 THEN 'Morning' ELSE 'Evening' END
* **(CY casualties)** Number of casualties that were involved in the accident this variable is created using the following formulae.
* SUM(IF YEAR([Accident Date])=[Current Year] THEN [Number of Casualties] END).

**Filter used in the chart :** CY casualtiesThis filter can be used to adjust the total no of casualties so that we can set a range for the casualties that will be displayed in the graph .

**Variables used in the marks shelf:**

* **Color: (**Morning or Evening)
* **Detail:** Local Authority (District)

**Text in tooltip is achieved using the following statements.**

Month of Accident Date: **<MONTH(Accident Date)>**

Morning or Evening: **<Morning or Evening>**

Local Authority (District): **<Local Authority (District)>**

CY casualties: **<AGG(CY casualties)>**

**Process of creating the visualization:**

**STEP 1)** By placing Accident Date in the Columns and Morning or Night and CY casualties in Rows we will be able to get necessary graph we will further enhance it by adding filters and variables in marks shelf.

**STEP 2)** Then we place CY casualties in the filters to adjust the number of casualties. In the marks shelf we will add Morning or Evening to the color card to assign different color to morning and evening, to segregate the area chart we will be adding Local Authority (District) to the Detail card so that each line in the area graph represents each districts casualty.

**Applying Design Principles:**

**Month’s Format:**  The month’s display format has been changed from entire month name to First letter of the month name. This is done by going to X-axis format, once the format opens in header go to Dates and change it to First letter so that just the months first letter displays in the graph.

**Line Color:** The color of each line represents its respective feature; the yellow line represents the casualties that have occurred during the mornings and the blue line represents the casualties that have occurred during the evenings.

**Questions that can be asked from the visualization.**

**1Q) How** does the number of accidents in the mornings compare to the evening’s over the years?

**2Q)** Are there any observable patterns for the number of casualties during mornings and evening’s over the time period?

**3Q)** How do accident trends vary among different local authorities?

**Insights that can be gained from the visualization.**

The number of accidents that have occurred in the mornings are consistently lower than in the evenings over the years across all the districts. The evening-time casualties are almost double in number when compared to the casualties that have occurred in the mornings.

In 2021 for the months January, February, and March we can see that they are moving in same directions, from March to September there is no similar patterns when mornings and nights are compared. From September 2021 to January 2022 both the graphs are moving in same direction.

For the Year 2022, from January to march and September to December we can observe that the number of casualties is moving in a similar pattern. And for the rest of the months there is no particular pattern that can be observed.

**7) Visualizing Road Accident Severity and Frequency Across Local Authority Districts**

A screenshot of a computer

Description automatically generated

**Variables used in the chart:**

* **Accident Severity:** This categorical variable represents the severity of accidents and is used to categorize accidents into different levels such as fatal, fetal, serious, and slight.
* **Count of Accident Index:** This quantitative variable represents the count of accident index, which indicates the number of accidents that occurred in each local authority district. It serves as the measure for the size of the bubbles in the jitter chart.
* **Local Authority:** This categorical variable represents the local authority district where the accidents occurred. It provides geographic context for the accidents and is used to categorize and differentiate the data points in the chart.
* **Random Rows and Columns:** These variables are introduced to create jitter in the chart, ensuring that data points are not overlapped and providing a clearer representation of the distribution of accidents.

**Questions that can be answered from the Jitter Chart:**

What is the distribution of accident severity levels across different local authority districts?

**Data Preparation:**

* The data is prepared by creating a summary of accident counts, indicating the count of accident index. This summary is represented as the "sum (bar height)" in the visualization.
* **Setting Up the Chart:**

The accident severity, represented by categories such as fatal, fetal, serious, and slight, is dragged into the Columns shelf. This sets up the x-axis of the chart to display severity levels.

* The "sum (bar height)" representing the count of accident index is placed in the “Columns shelf” to establish the y-axis, creating the bars that indicate the total number of accidents for each severity level.
* **Randomization:** Random rows are placed in the “Rows shelf” to introduce jitter, ensuring that data points are not overlapped and providing a clearer representation of the distribution of accidents.

**Applying Design Principles:**

* **Bubble Size:**

The size of each bubble corresponds to the count of accidents in the local authority district. Larger bubbles indicate a higher number of accidents in that district. The bubbles provide a visual representation of the distribution of accidents across local authority districts, with each bubble indicating both the severity and frequency of accidents in a particular district.

* **Bubble Color:**

The color of each bubble represents the severity of the accidents. Blue bubbles represent fatal accidents, orange bubbles represent fetal accidents, green bubbles represent serious accidents, and red bubbles represent slight accidents.

* **Bar Height:**

The height of the bars in the background represents the total number of accidents for each severity level. These bars serve as a reference point for understanding the overall distribution of accidents across severity levels.

**Here are some potential findings that could be derived from the jitter chart described:**

* **Spatial Distribution of Accidents:** The jitter chart reveals the spatial distribution of accidents across different local authority districts, highlighting areas with higher concentrations of accidents compared to others.
* **Severity Variation:** By observing the size and color of the bubbles, it becomes evident which severity levels are more prevalent in specific local authority districts. For instance, districts with larger blue bubbles may indicate a higher frequency of fatal accidents.
* **Local Authority Hotspots:** Certain local authority districts may emerge as hotspots for specific severity levels of accidents. These districts may require targeted interventions to improve road safety and reduce the incidence of accidents.
* **Geographic Patterns:** The jitter chart can help identify any geographic patterns or clusters of accidents with specific severity levels. Clusters of larger bubbles of the same color may suggest localized areas with a higher prevalence of accidents of a particular severity level.
* **Comparative Analysis:** By comparing the distribution of accidents across different severity levels within each local authority district, insights can be gained into the relative impact and severity of accidents in different areas.
* **Data Dispersal:** The jitter effect introduced in the chart ensures that data points are spread out, reducing overlap, and providing a clearer representation of the distribution of accidents across severity levels and local authority districts.

These findings can serve as a basis for further analysis and decision-making aimed at improving road safety and reducing the incidence and severity of accidents in specific areas. By comparing the distribution of bubble colors across different local authority districts, stakeholders can gain insights into the geographic patterns of accident severity and identify areas that may require targeted interventions or safety measures.

**Based on the insights derived from the jitter chart, here are some suggestions that could be made:**

* **Targeted Interventions:** Identify local authority districts with high concentrations of accidents, particularly those with a high prevalence of fatal or serious accidents. Implement targeted interventions such as improved signage, road infrastructure upgrades, or increased enforcement in these areas to reduce accident rates.
* **Safety Campaigns:** Develop and implement safety campaigns tailored to address specific severity levels of accidents prevalent in different local authority districts. For example, campaigns targeting speeding and reckless driving in districts with a high prevalence of fatal accidents.
* **Infrastructure Improvements:** Invest in infrastructure improvements in areas with clusters of accidents, such as road redesigns, installation of traffic calming measures, or enhancements to pedestrian and cyclist safety infrastructure.
* **Data-Driven Policy:** Use the insights from the jitter chart to inform data-driven policy decisions aimed at improving road safety. For example, allocate resources and funding based on the severity and frequency of accidents in different areas.

**8) Jitter Chart: Top 10 Local Authority Districts with Most Accidents"**

A screenshot of a computer

Description automatically generated

**Findings from the Chart:**

**Variation in Accident Severity:** The data reveals variations in accident severity across the ten listed local authority districts. For example, some districts have higher numbers of fatal accidents compared to others, while others have more fetal or slight accidents.

**Magnitude of Fatal Accidents:** The number of fatal accidents ranges from 4 in Lambeth to 184 in Westminster. This indicates significant differences in the risk of fatal accidents between different districts.

**Distribution of Fetal Accidents:** Fetal accidents also vary across districts, with count 5 in Westminster. While fetal accidents are less severe than fatal accidents, they still represent a serious risk to road safety.

**Slight Accidents:** Slight accidents, which are typically less severe, show a wide range across districts, with counts ranging from 1,690 in Southwark to 5,402 in Birmingham. This suggests differing levels of traffic safety measures or enforcement across districts.

**Serious Accidents:** Serious accidents, which are more severe than any other accidents, also shows the distribution across the local districts ranging from 230 in Barnet to 706 in Birmingham. Analyzing the number of serious accidents alongside fatal, fetal, and slight accidents provides a comprehensive understanding of road safety issues within each district and help identify areas for targeted interventions or improvements.

**Further analysis –**

**Overall, Road Safety:** While the provided data gives insights into the severity of accidents, it's important to note that it doesn't provide a comprehensive picture of overall road safety. Factors such as population density, road infrastructure, traffic patterns, and enforcement efforts may also influence accident rates.

**9) Exploring the Correlation: Speed Limits, Casualties, and Vehicles in Road Accidents**A screenshot of a computer

Description automatically generated

**Variables Involved:**

1. Distinct count of Speed limit: The number of different speed limits within a district, indicating the diversity in speed regulation.
2. Number of Casualties: The total casualties resulting from road accidents in each district.
3. Number of Vehicles: The total number of vehicles involved in road accidents within each district.
4. Local Authority (District): The specific districts or administrative regions where the accidents have occurred.
5. Correlation Coefficient: A statistical measure that describes the size and direction of a relationship between two variables (not visible in the chart but mentioned in the tooltip).

The visualization you've provided appears to be a set of scatter plots showing the correlation between various factors related to road accidents. Here's a breakdown of each plot and the story they tell together:

Plot Descriptions:

1.Upper Left Plot - District Count vs. Speed Limit:

- This plot displays the number of unique speed limits against the count of districts. Each point represents a district, and the line indicates a trend or relationship between the diversity of speed limits and the district count.

2. Upper Right Plot - Casualties vs. Speed Limit:

- Here, we see a comparison between the number of casualties and the number of distinct speed limits in districts. The trend line suggests that as the variety of speed limits in a district increases, so does the number of casualties.

3. Bottom Left Plot - Casualties vs. Vehicles:

- This scatter plot correlates the number of casualties with the number of vehicles involved in accidents. The positive slope of the trend line indicates that districts with more vehicles involved in accidents tend to also have a higher number of casualties.

4. Bottom Right Plot - Vehicles vs. Speed Limit:

- Similar to the upper right plot, this shows the relationship between the number of vehicles and the variety of speed limits. Again, there seems to be a positive correlation, with more vehicles involved in accidents in districts with a greater diversity of speed limits.

**Insights and Questions Addressed:**

- Is there a relationship between the complexity of speed limits and accident severity?

- The plots suggest that a higher diversity of speed limits within a district is associated with both an increase in the number of casualties and the number of vehicles involved in accidents.

- Are districts with more diverse speed regulations experiencing more accidents?

- While the plots show a correlation, causation cannot be determined from correlation alone. However, it does raise questions about whether more complex speed limit landscapes contribute to higher accident rates.

- Does the number of vehicles involved in accidents correlate with the number of casualties?

- Yes, the direct relationship shown in the scatter plot supports the notion that more vehicles involved in accidents usually mean more casualties.

**Professional Explanation for Stakeholders:**

These scatter plots employ statistical methods to examine the relationships between the diversity of speed limits in a district and accident severity metrics. The trend lines indicate that increased complexity in speed regulations might be linked to the severity and frequency of accidents, suggesting that policy simplification could be beneficial. Further investigation is warranted to explore whether standardizing speed limits across larger areas could enhance road safety. Stakeholders should consider these findings when discussing traffic regulation policies and district-level road safety measures.

**10) "Accident Clustering by District: Exploring Severity, Frequency, and Variance"**

A map of the united kingdom

Description automatically generated

A screenshot of a report

Description automatically generated

A table with numbers and text

Description automatically generated

Variables Involved in the K-Means Clustering:

1. **Sum of Number of Casualties**: The total number of casualties in accidents within each district.
2. **Sum of Number of Vehicles**: The total number of vehicles involved in accidents within each district.
3. **Accident Severity**: The most common level of severity of accidents within each district.
4. **Local Authority (District)**: The specific district or area where the data is collected.
5. **Number of Clusters**: The number of distinct groups formed through the clustering process.

* The map showcases the clustering of road traffic accidents based on severity and location within a specific region that includes parts of the UK, Ireland, the Netherlands, and Belgium.
* Each cluster is color-coded, and the size of the dots appears to correspond to the number of casualties in each accident.

**K-Means Clustering Analysis:**

* **Inputs for Clustering**:
  + Variables included are the sum of the number of casualties, the sum of the number of vehicles, and the accident severity.
  + Data has been normalized and is detailed by the local authority (district).
* **Summary Diagnostics**:
  + There are 4 clusters created from 244 data points (locations).
  + The within-group sum of squares is low (1.5388), suggesting that the points within each cluster are close to their respective centroids.
  + The between-group sum of squares is substantially higher (13.204) relative to the within-group sum, indicating that the clusters are distinct from one another.
* **Clusters Information**:
  + Cluster 1 has the most items (163 locations) but lower average casualties and vehicles involved.
  + Cluster 2 has fewer items (54 locations) but a higher average of casualties and vehicles, indicating more severe accidents.
  + Clusters 3 and 4 have very few items, with cluster 3 showing particularly high averages, likely indicating specific areas where severe accidents are less frequent but more impactful when they occur.
* **Analysis of Variance (ANOVA)**:
  + The F-statistic and associated p-value for both the number of casualties and the number of vehicles are significant (p-value = 0.0), suggesting that there is a statistically significant difference in the means of these variables across the four clusters.

**Questions Answered by the Visualization:**

* **Which districts have the highest severity of accidents?**
  + By looking at the color-coding and size of the dots on the map, we can identify areas with a higher severity of accidents.
* **Are accidents evenly distributed across the region?**
  + The visualization indicates that accidents are not evenly distributed, with certain clusters showing a higher frequency or severity of accidents.
* **Do some districts have a higher number of vehicles involved in accidents?**
  + Yes, the clusters with larger dot sizes suggest areas with a higher number of vehicles involved in accidents.
* **How distinct are the accident severities between clusters?**
  + The significant ANOVA results indicate clear distinctions in the accident severities across clusters.

**Professional Explanation for Stakeholders:**

* The k-means clustering analysis has successfully segmented the road traffic accident data into four distinct groups based on the number of casualties, number of vehicles involved, and accident severity.
* This geographical segmentation helps us identify patterns and hotspots, which is essential for targeting interventions and improving road safety measures.
* The significant variation among clusters shows that our approach is on the right track in identifying areas that require different levels of attention and resources.
* Our findings will be crucial in guiding infrastructure development, policy-making, and emergency services deployment to reduce accident rates and improve safety on our roads.

These visualizations provide actionable insights for stakeholders, emphasizing areas that might benefit from increased road safety measures and potential changes in traffic regulations.

**Dashboard**

**Dashboard 1: Road Accident Analysis Dashboard**

A screenshot of a computer screen

Description automatically generated

**Introdution:** The dashboard provides a comprehensive analysis of road accident data, offering insights into the frequency, severity, and distribution of accidents across various parameters. By examining these patterns, we aim to identify areas of concern and inform targeted interventions to enhance road safety.

**Charts included:**

* Casualties by Accident Severity and Location (District)
* Accidents by Road Type
* Accident Severity Distribution by Day of the Week
* Speed Limit in Rural and Urban vs Accident Severity

**Questions accident by dashboard**

"What are the key factors and patterns associated with road accidents, including the severity of casualties, the distribution of accidents across different road types, the temporal trends in accident severity, and the relationship between speed limits in rural and urban areas and accident severity?

**Dashboard Design:Bottom of Form**

To design the dashboard with the filter for accident severity:

1. **Select Relevant Charts:** we have chosen charts such as "Casualties by Accident Severity and Location (District)," "Accidents by Road Type," "Accident Severity Distribution by Day of the Week," and "Speed Limit in Rural and Urban vs. Accident Severity."
2. **Create Filters:** For each chart, create a filter based on accident severity. This filter will allow users to select the severity level they want to analyze. you have created filters by right-clicking on the relevant field (accident severity) and selecting "Show Filter."
3. **Apply Filters to Charts:** Once the filters are created, we apply them to each chart. This will ensure that the data displayed in the charts changes dynamically based on the selected severity level. we dragged the severity filter to each chart or use the "Filters" shelf to apply the filter.
4. **Arrange the Dashboard:** Arrange the charts on the dashboard layout as desired, positioned them in a grid format or arrange them vertically .Ensuring that there is enough space for the filter(s) to be placed.
5. **Add Filters to the Dashboard:** Add the severity filter(s) to the dashboard. Position them at the top or side of the dashboard for easy access. drag the severity filter(s) from the filter shelf onto the dashboard canvas.
6. **Test Interactivity:** Testing the dashboard to ensure that the charts update dynamically based on the severity level selected in the filter. Verify that all charts respond appropriately to changes in the filter selection.
7. **Formatting and Final Touches:** Format the dashboard to improve readability and aesthetics. This includes adding titles, axis labels, legends, and other annotations as needed. Ensure that the layout is visually appealing and intuitive for users to navigate.

**Design principles used:**

* Consistency: Consistent use of colors, fonts, and layout elements across charts and filters helps create a cohesive visual experience and makes it easier for users to navigate the dashboard.
* Hierarchy: The layout of the dashboard likely follows a logical hierarchy, with the severity filter prominently placed at the top or side for easy access. This ensures that users understand the primary interactive element of the dashboard.
* Simplicity: The dashboard design is likely kept simple and uncluttered to avoid overwhelming users with unnecessary information. Each chart and filter serves a specific purpose, and extraneous elements are minimized to maintain focus.
* Interactivity: The use of interactive elements, such as filters, allows users to customize their analysis and explore the data based on their preferences. This enhances user engagement and facilitates deeper insights into the dataset.

**Conclusion:**

In essence, the dashboard provides insights into the various dimensions of road accidents, allowing stakeholders to understand the underlying factors, identify areas of concern, and inform targeted interventions to improve road safety.

The dashboard concludes by emphasizing the importance of data-driven approaches in understanding road accident patterns and informing road safety strategies. By leveraging these insights, stakeholders can prioritize interventions, allocate resources effectively, and work towards creating safer road environments for all.

**Dashboard 2 - Time Analysis and Comparison of Road Accidents:**

A screenshot of a computer

Description automatically generated

**Introduction:**

The dashboard provides a detailed exploration of temporal patterns in road accidents, aiming to uncover trends over time and compare current data with previous years. By analyzing the distribution of accidents across different days of the week and times of the day, we seek to identify key insights that can inform targeted interventions and enhance road safety measures.

**Charts included:**

* Current year Casualties by Day of Week
* No: of Accidents compared to previous year
* Time Series Comparison for Road accidents
* Number of Accidents occurred in Morning and Evening with Number of Casuality

**Questions accident by dashboard:**

"What are the temporal trends in road accidents, and how do they compare to previous years? Additionally, how do the number of casualties vary by day of the week and time of day?"

**Dashboard Design:Bottom of Form**

1. Data Preparation: Gather and prepare the relevant data on road accidents, including variables such as accident dates, casualties, days of the week, times of day, and year-over-year comparisons.
2. Select Charts: Choose the appropriate charts that will provide insights into temporal trends and comparisons. For this dashboard, consider using charts such as bar charts, line charts, and time series plots.

* Current Year Casualties by Day of the Week:
* No. of Accidents compared to previous year:
* Time Series Comparison for Road Accidents:
* Number of Accidents occurred in Morning and Evening:

1. Dashboard Layout: Arrange the charts on the dashboard layout in a logical order. grouped related charts together and leaving space for filters or additional annotatio
2. Filtering and Interactivity: Add filters if necessary to allow users to interact with the dashboard. For example, added filters to select specific time periods or accident severity levels.
3. Formatting and Final Touches: Format the dashboard to improve readability and aesthetics. Add titles, axis labels, legends, and annotations as needed. Ensuring that the layout is visually appealing and easy to understand.
4. Testing: Test the dashboard to ensure that all charts and filters work as expected. Verify that the dashboard provides meaningful insights into temporal trends and comparisons of road accidents.

**Design principles used:**

1. Responsive Design: Ensure that the dashboard is responsive and adaptable to different devices and screen sizes. Design the layout and visualizations to scale appropriately and maintain usability across desktops, tablets, and mobile devices.

2. Aesthetic Appeal: Pay attention to the aesthetic aspects of the dashboard, including color choices, typography, and graphic elements. Create a visually appealing design that enhances the overall user experience and encourages engagement with the data.

3. Iterative Design Process: Adopt an iterative approach to dashboard design, soliciting feedback from users and stakeholders at various stages of development. Incorporate user feedback to refine the design, improve usability, and address any issues or concerns.

**Conclusion:**

In essence, the dashboard provides insights into the temporal patterns of road accidents over time, comparing the current year's data with previous years. It also examines the distribution of accidents across different days of the week and times of the day, shedding light on when accidents are most likely to occur and their associated casualties.

The dashboard concludes by emphasizing the importance of understanding temporal trends in road accidents and their implications for road safety initiatives. By leveraging insights from the data, stakeholders can prioritize interventions, allocate resources effectively, and work towards reducing the incidence and severity of road accidents.

Bottom of Form

**Dashboard 3: Year over Year Accident Comparison**

**Introduction:**

The dashboard offers a comprehensive analysis of road accident data, focusing on year-over-year comparisons to uncover trends and patterns in accident occurrences and severity levels. By examining changes in total accidents, casualties, and severity categories over time, we aim to identify areas of progress and areas needing attention in road safety initiatives.

**Charts included :**

* Total accidents Kpi and sparkline
* Total Causalities Kpi and sparkline
* Total fatal Causalties Kpi and sparkline
* Total Serious causalties Kpi and sparkline
* Total Slight causalties Kpi and Sparkline

A screenshot of a computer

Description automatically generated

**Questions accident by dashboard:**

"How do road accidents and their associated casualties compare year-over-year, and what are the trends in total accidents, casualties, and severity levels over time?"

**Dashboard Design:Bottom of Form**

To create Dashboard 3 - Year over Year Accident Comparison with current year and previous year filters

1. Data Preparation: Gather and prepare the relevant data on road accidents, including variables such as accident dates, casualties, and severity levels.
2. Select Charts: Choosing the appropriate charts that will provide insights into year-over-year comparisons of total accidents, casualties, and severity levels. For this dashboard, consider using KPIs, bar charts, and sparklines.
3. Total Accidents Year-over-Year:
   * Create a KPI for total accidents in the current year and add a sparkline to show the trend over time.
4. Total Casualties Year-over-Year:
   * Create KPIs for total casualties (including fatalities, serious injuries, and slight injuries) in the current year and add sparklines to show the trends over time.
5. Total Fatal Casualties Year-over-Year:
   * Repeat the same process for total fatal casualties, creating KPIs, sparklines, and bar charts comparing the current year with the previous year.
6. Total Serious and Slight Casualties Year-over-Year:
   * Create separate KPIs, sparklines for total serious and slight casualties, comparing the current year with the previous year.
7. Filters for Current Year and Previous Year:
   * Create filters for selecting the current year and previous year. These filters will allow users to switch between different years and dynamically update the data displayed in the charts.
8. Dashboard Layout: Arrange the charts on the dashboard layout in a logical order, with KPIs at the top and bar charts and sparklines below. Leaving space for the filters to be placed at the top or side of the dashboard.
9. Testing: Test the dashboard to ensure that all charts and filters work as expected. Verify that users can select different years using the filters and that the data updates accordingly.

Formatting and Final Touches: Format the dashboard to improve readability and aesthetics. Add titles, axis labels, legends, and annotations as needed. Ensuring that the layout is visually appealing and easy to understand.

**Design principles used:**

1. User-Centric Design: Design the dashboard with the end user in mind, understanding their needs, preferences, and level of expertise. Ensure that the layout, navigation, and visualizations cater to the target audience's expectations and requirements.

2. Contextualization: Provide context and background information to help users understand the significance of the data and the insights derived from it. Include annotations, descriptions, and contextual notes to guide users in interpreting the visualizations effectively.

3. Progressive Disclosure: Present information gradually and progressively, starting with high-level summaries and allowing users to drill down into more detailed data as needed. This approach prevents information overload and facilitates a more focused analysis.

4. Storytelling: Use narrative techniques to tell a compelling story with the data, guiding users through a logical sequence of insights and conclusions. Incorporate storytelling elements such as introductions, plot development, and conclusions to engage users and communicate key messages effectively.

**Conclusion:**

In essence, the dashboard provides insights into the trends and changes in road accident occurrences and severity levels by comparing data from different years. It allows stakeholders to understand the overall trajectory of road safety efforts and identify areas of improvement or concern based on historical data.

The dashboard concludes by emphasizing the importance of continuous monitoring and evaluation of road safety efforts through year-over-year comparisons. By leveraging insights from historical data, stakeholders can identify trends, assess the effectiveness of interventions, and implement evidence-based strategies to improve road safety outcomes for all road users.

**StoryBoard**

**Comprehensive Road Safety Analysis: Patterns, Risks, and Temporal DynamicsYearly**

* **Road Safety Overview - A Tale of Decline and Opportunity**

A screenshot of a computer

Description automatically generated

This dashboard encapsulates a detailed narrative of road traffic accidents and casualties over the year. It tells a positive story of overall decline, with total accidents and casualties down by approximately 11.7% and 11.9%, respectively. However, beneath this overarching trend, the data also point to specific challenges, such as the significant increase in accidents involving buses and motorcycles. The sparklines provide a monthly rhythm to the data, revealing the ebb and flow of accident rates through the seasons. By dissecting these trends, this story point emphasizes the complexity of road safety and the continuous need for targeted, data-driven interventions

* **Navigating the Elements - Casualties in Different Driving Conditions:**

A graph of different colored lines

Description automatically generated with medium confidence

The interconnected Sankey diagrams reveal the multifaceted nature of road traffic accidents and their casualties. The flow of casualties is highest during daylight and in fine weather, suggesting a higher volume of traffic during these perceived ‘safer’ conditions. Conversely, fewer casualties are associated with traditionally hazardous conditions like darkness and inclement weather, possibly due to more cautious driving behavior. Road type analysis indicates that more complex or high-speed routes, such as dual carriageways and roundabouts, are associated with a higher number of casualties. This visualization challenges assumptions, guiding us to rethink safety measures and focus on times and places where accidents are unexpectedly more frequent.

* **Junctions and Casualties – Where Roads Cross, Risks Rise**

A black background with lines

Description automatically generatedA black background with red and yellow lines

Description automatically generated

Our data streams diverge into patterns of risk at the crossroads of travel. The Sankey diagrams trace the flow of casualties through the arteries of junction control and detail, revealing the critical junctions where risk culminates. "Give way or uncontrolled" junctions and "T or staggered junctions" emerge as focal points of concern, channeling the highest numbers of casualties. These flows of data illuminate the cross-sections of decision and danger, guiding us to reconsider the design and control of these junctions. As we trace the paths of these diagrams, we commit to a course of action that could transform risky intersections into safe passages.

* **Dissecting Accident Dynamics - Location, Time, and Type**

A screenshot of a computer screen

Description automatically generated

dashboard takes us on a visual journey through the critical factors contributing to road accidents. The clustered map spots tell a story of geographic accident distribution, potentially identifying regions needing urgent road safety improvements. The bar chart converses about the dangers of single carriageways, which dwarf other road types in accident counts. A weekly rhythm of accidents emerges, peaking on Saturdays, hinting at behavioral risks or increased weekend travels. Finally, the urban-rural divide reveals an urban concentration of accidents at lower speed limits, contrasting with rural high-speed mishaps. This single story point emphasizes the multifactorial nature of road safety and the importance of tailored, data-driven responses.

* **Temporal Trends in Road Safety**

A screenshot of a computer

Description automatically generated

This dashboard distills the ebb and flow of road accidents into a temporal story of risk. Casualties peak on weekends, urging a review of leisure-time driving habits and enforcement measures. A heartening downtrend in quarterly accidents year over year offers a silent nod to effective safety interventions. The daily cadence of accidents is revealed through hourly data, with urban life dictating a rhythm of risk around typical work hours, while rural roads tell a tale of steadier, yet persistent, danger. As daylight fades, so does safety, with evenings proving more perilous. This singular story point draws our attention to the critical hours and days where our vigilance against road mishaps must be most keen.

* **Unraveling the Ties Between Speed Diversity and Accident Severity**

A screenshot of a computer screen

Description automatically generated

This set of scatter plots illuminates the intricate relationship between the diversity of speed limits across districts and the resulting road safety outcomes. We observe that districts with a wider range of speed limits tend to report a higher number of accidents and casualties. The plots collectively suggest a potential link between the complexity of speed regulations and the increase in both the volume of traffic incidents and the severity of their consequences. As we dissect these correlations, we prompt a vital discussion on the impact of speed limit diversity on road safety, leading to considerations for policy refinement and uniformity that may simplify driving conditions and enhance overall traffic safety

* Diverse Impact - A Closer Look at Accident Severity Across Districts

A screenshot of a computer

Description automatically generated

Our jitter chart showcases the distribution of road traffic accidents categorized by severity across the top local authority districts. At a glance, we can see a clear disparity: while 'Slight' accidents dominate the landscape, often in densely populated areas such as Leeds, 'Serious' and 'Fatal' incidents, although less frequent, punctuate the chart with a pressing reminder of their grim presence. The size variation in our data points is not just a visual cue but a call to action, as each enlarged circle represents a higher number of casualties and a district in need of focused attention and tailored road safety interventions.

* **Unveiling Hotspots and Prioritizing Action**

A map of the united kingdom

Description automatically generated

By employing k-means clustering, we've identified four distinct groups of traffic accidents across regions, differentiated by the number of casualties and vehicles involved. The visualization on the map highlights areas with varying accident severities, which are crucial for directing our road safety efforts. The data-driven clusters range from high-frequency, lower-severity incidents to less common but more severe accidents. Statistical analysis validates the significance of these clusters, ensuring that our targeted safety measures are grounded in robust evidence. This insight is pivotal for allocating our resources efficiently and formulating tailored interventions that could potentially save lives.

**Top of Form**

**Summary and Conclusions**

Throughout our research, we have unearthed significant findings that shed light on road accident patterns and their underlying factors. Our analysis has provided valuable insights into various aspects of road safety, aiding in the formulation of informed strategies and interventions. Here are the key findings and implications of our research:

1. Temporal Trends: Time series analysis revealed notable temporal trends in road accidents, including seasonal variations and long-term changes. Understanding these patterns allows for targeted interventions during peak accident periods and facilitates the allocation of resources effectively.
2. Geographical Hotspots: By analyzing accident locations, we identified geographical hotspots with heightened accident rates. Targeting these areas with tailored interventions and infrastructure improvements can mitigate accident risks and enhance overall road safety.
3. Risk Factors: Factors such as weather conditions, road surface conditions, junction control types, and vehicle types significantly influence accident occurrence and severity. Addressing these risk factors through appropriate measures and interventions can help reduce accident rates and mitigate their impact.
4. Demographic Considerations: While demographic information such as age groups was not fully available, its inclusion is crucial for understanding age-specific risk factors and tailoring interventions accordingly. Future research should focus on comprehensive data collection to address this gap.
5. Success in Addressing Objectives: Our project has successfully addressed the initial questions and objectives by leveraging data-driven analyses to uncover actionable insights. By examining various aspects of road accidents and their contributing factors, we have laid the groundwork for informed decision-making and the development of targeted road safety initiatives.

In conclusion, our research underscores the importance of data-driven approaches in understanding road accident patterns and informing road safety strategies. While our findings provide valuable insights, there remains a need for ongoing research and collaboration to further refine our understanding and enhance road safety outcomes. By continuing to leverage data and evidence-based practices, we can work towards creating safer road environments for all.

**References:**

1. Elvik, R., Hoye, A., Vaa, T., & Sorensen, M. (2009). The handbook of road safety measures. Emerald Group Publishing.
2. Antoniou, C., Yannis, G., & Papadimitriou, E. (2017). Road safety in European countries: Trends, statistics, and national policies. Journal of Transportation Safety & Security, 9(3), 266-286.
3. Xie, Y., Liu, C., Wang, J., & Li, J. (2020). A data-driven approach for visualizing road traffic accidents using geographic information systems: A case study in China. Safety Science, 131, 104903.
4. Shi, H., Wang, J., Zhang, D., & Gong, P. (2021). Visualization of road traffic accidents using spatial clustering and time series analysis. International Journal of Geographical Information Science, 35(5), 947-971.
5. Wang, Y., Li, J., & Wang, J. (2021). Visualizing road traffic accidents using GIS: A case study in Shenzhen, China. Sustainability, 13(15), 833
6. World Health Organization (WHO). (2021). Road traffic injuries. Retrieved from <https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries>

**Team Members Contributions:**

**1) Saadana Shyam Sunder:**

* Spearheaded the development of chart visualization strategies.
* Drafted the project proposal outlining the visualization plan.
* Created initial draft charts as prototypes for further development.

**2) Puja Madhav:**

* Conducted the collection and organization of the dataset.
* Enhanced the design of the charts based on project requirements and feedback.
* Provided valuable input and insights during the review process of the project proposal.

**3) Hari Charan Reddy Karra:**

* Analyzed and interpreted the data to identify key insights and trends.
* Added critical elements to the charts to improve clarity and understanding.
* Collaborated closely with the team to ensure the charts effectively communicated the project's objectives.

**4) Taruni Madugula**

* Contributed to the analysis of the charts, providing valuable perspectives and recommendations.
* Played a crucial role in the selection of the dataset, ensuring it aligned with the project's goals and objectives.
* Participated actively in discussions to refine the visualization approach based on the data characteristics.

**5) Kanaka Sri Brahmini Bandla**

* Contributed valuable ideas and suggestions regarding the types of charts to be created.
* Engaged in brainstorming sessions to explore various visualization options and concepts.
* Provided creative input that contributed to the development of the overall charting strategy.

**6) Deepthi Sree Boonapalli**

* Provided essential information and context for the project proposal.
* Conducted thorough research and collected relevant references to support the project's findings and recommendations.
* Collaborated with the team to ensure the proposal was comprehensive and well-supported by evidence.