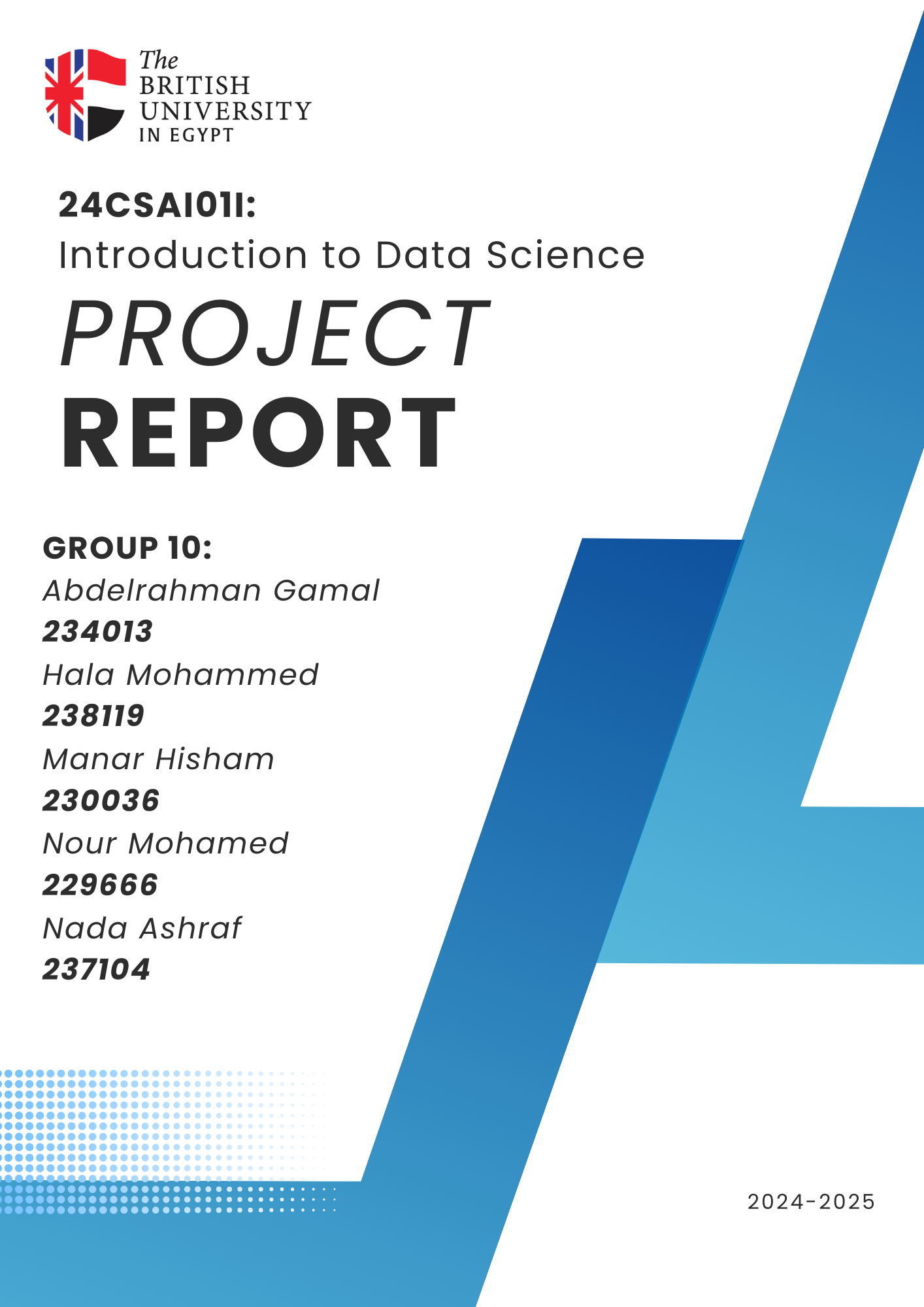
****

**The link to the data set we will be working on:**

[**https://www.kaggle.com/datasets/farshidbahrami021/road-accident-dataset**](https://www.kaggle.com/datasets/farshidbahrami021/road-accident-dataset)

This is an unclean dataset that includes various information regarding road accidents that occurred in different states throughout the USA. It contains a total of **23 features** along with **1,610 records** and it has some null values to reflect real-world scenarios where not all information is available or recorded. It includes information like date and time at which the accident occurred and other information such as environmental conditions, and accident circumstances which provides insights to analyze what patterns and factors contribute to the severity of the accident.

**Description of the data set features:**

**State**: The state in the USA where the accident occurred.

**Date:** The date and time at which the accident occurred.

**Day\_of\_Week:** The day of the week when the accident occurred.

**Time\_of\_Day:** The time of the day when the accident occurred.

**Weather\_Conditions:** The prevailing weather conditions at the time of the accident (e.g., Sunny, Rainy, Snowy, Foggy, Stormy).

**Road\_Conditions:** The condition of the road surface at which the accident occurred for example, Dry, Wet, Icy, Snowy, Muddy

**Light\_Conditions:** The lighting conditions at the time of the accident (e.g., Daylight, Dawn, Dusk, Night).

**Type\_of\_Road:** The type of road where the accident occurred (Highway, Street, Freeway, Rural, Urban).

**Type\_of\_Junction:** The type of junction where the accident occurred ( Intersection, T-Junction, Roundabout, Crossroads, Overpass, Underpass).

**Type\_of\_Accident:** The type of accident that occurred (e.g., Rear-end collision, Head-on collision, Side collision, Rollover, Run-off-road, Pedestrian involved, Cyclist involved).

**Vehicle\_Type:** The type of vehicle involved in the accident (e.g., Car, Truck, Motorcycle, Bicycle, Pedestrian).

**Driver\_Age\_Group:** The age group of the driver involved in the accident (e.g., Teenager, Young Adult, Adult, Senior).

**Num\_Vehicles\_Involved:** The number of vehicles involved in the accident.

**Num\_Casualties:** The number of casualties (injuries or fatalities) resulting from the accident.

**Speed\_Limit:** The speed limit of the road where the accident occurred.

**Distance\_to\_Nearest\_Hospital:** The distance to the nearest hospital from the accident location.

**Distance\_to\_Nearest\_Police\_Station:** The distance to the nearest police station from the accident location.

**Visibility:** The visibility level at the time of the accident.

**Road\_Width:** The width of the road where the accident occurred.

**Road\_Surface\_Friction\_Coefficient:** The coefficient of friction of the road surface.

**Vehicle\_Speed:** The speed of the vehicle involved in the accident.

**Time\_Taken\_for\_Emergency\_Response**: The time taken for emergency response to arrive at the accident scene.

**Research Questions:**

1- **“How do weather conditions impact road conditions and what impact does this combination have on emergency response time?”**

Do bad weather conditions lead to longer delays in response times? Additionally, how does the distance to the nearest hospital affect emergency response times during such conditions?

We first analyze how road conditions behave under different weather conditions, for instance, muddy roads could be more common in rainy weather, and icy roads are more common in winter. Therefore, we can observe how weather conditions affect road conditions influencing the rate of emergency response times even when the distance to the nearest hospital is too short.

**A key question to explore is:** Does a shorter distance to the nearest hospital always result in fast emergency response time? We want to investigate when the distance to the nearest hospital fails to predict faster response times, as this may reveal insights such as if any of these factors contribute to delays, for example, does severe weather and road conditions sometimes lead to major delays even when the hospital is nearby?

**Features Used:** **Weather\_Conditions, Road\_Conditions, Distance\_to\_Nearest\_Hospital, Time\_Taken\_for\_Emergency\_Response**

2- **“How does road visibility along with the light conditions affect the severity of the accident?”**

Would it cause a higher number of casualties at night due to dark surroundings with low visibility compared to daylight where drivers have clearer sight. In addition, bad weather conditions would further limit visibility throughout all times, increasing the rate of accident severity.

To analyze whether such correlation exists or not, we can use data visualizations to uncover a pattern between accident severity regarding road visibility, light conditions, and weather conditions. We would need to observe where a high number of casualties usually occur, as it should be more common with dark surroundings where there is low visibility compared to daylight conditions. Furthermore, we want to investigate how adverse weather conditions like fog or storms play a role in limiting the visibility during both day and night, always making driving conditions at a potential risk.

**Features Used:** **Light\_Conditions, Visibility, Num\_Casualties, Weather\_Conditions**

3-**"Do accidents with vehicle speeds exceeding the speed limit result in more multi-vehicle collisions, and does road width play a role?"**

Are accidents occurring at high speeds destined to have some collateral damage, due to the vehicle being out of control, especially taking the road’s width into account, does the narrower the road result in more vehicles being involved in nearby accidents?

**Features Used:**

* **Vehicle\_Speed** – this column is selected to study the how accident speeds vary and their relation to the road’s speed limit.
* **Num\_Vehicles\_Involved** – to check if higher speeds lead to multi-vehicle accidents.
* **Speed\_Limit** – to analyze the relation between the vehicle’s speed and road width, since speed limits are determined by various factors and width is a factor.
* **Road\_Width** – To investigate whether narrower roads are more prone to multi vehicle crashes.

**How to visualize this question:**

**Scatter Plot**: to create a correlation between the speed difference (Vehicle\_Speed – Speed\_limit) and the number of vehicles involved, color coding the road’s width also to visualize the data accurately.

Plotting the speed difference of speed on the X-axis, and the number of vehicles involved on the Y-axis, monitoring if the dots keep on inclining or declining, then we have our answer.

4- **"How do different states vary in the types of junctions where accidents occur, and which junctions are most accident-prone and result in severe accidents in each state?"**

Certain junction types might lead to more casualties in specific states due to traffic density or accident frequency. States with many intersections may have more casualties per accident compared to those with more roundabouts, or other types of junctions.

**Features Used:**

**State** – To compare accident trends across different states.

**Type\_of\_Junction** – To identify which junction types (e.g., intersections, roundabouts, T-junctions) are most common in each state.

**Num\_Casualties** – To determine the severity of accidents at different junctions in each state.

**How to approach this question visually:**

Bar chart: plotting the “State” attribute on the X-axis to study the most common state that has accidents and the count of these accidents on the Y-axis, color coding the type of junction that these accidents has happened on.

5- **“How does the distance to the nearest police station and hospital along with the road type and time of day influence the emergency response time in each US state?”**

**Potential insights:**

* Longer distances to police stations and hospitals would generally increase emergency response time
* Rural roads might experience longer delays than urban or highways
* Some states might have quicker response times despite longer distances, possibly due to better services and infrastructure
* Accidents that happen during nighttime might face longer delays

**How can we answer the question from the data set:**

* Compare emergency response time across states
* Find correlations between the police station and hospital distances with the response time
* Find patterns between the type of road or time of day
* Make use of visualization techniques such as scatter plots for distance vs. response time and heatmaps to highlight states with long delays.

**Features Used: Distance\_to\_Nearest\_Hospital, Distance\_to\_Nearest\_Police\_Station, Type\_of\_Road, Time\_of\_Day, State**

6- “**How do vehicle speed and road surface friction influence the number of vehicles involved in accidents across different junction types? Do slippery roads with lower friction coefficients lead to more multi-vehicle crashes, especially at higher speeds?”**

**How can we answer the question from the data set:**

* First, we'll analyze the relationship between road surface friction and vehicle speed in a variety of accident situations. We can see how different speed levels affect the severity of accidents under various road friction situations by classifying Vehicle\_Speed into ranges (low, medium, and high).
* Create a heatmap to visualize how different levels of friction and speed contribute to multi-vehicle accidents across various junction types.

**Features Used: Type\_of\_Junction**, **Num\_Vehicles\_Involved**, **Road\_Surface\_Friction\_Coefficient**, **Vehicle\_Speed**

7- **“How do road types and road conditions affect the rate of accidents in different days of the week and different times of the day?”**

**Possible Insight:** predicting the days that have the most amount of accidents on specific roads

**Point to make:** trying to avoid specific roads on a specific days in order to avoid possible accidents in the future

**How can we answer the question from the data set:**

* Gather data: Day\_of\_Week ,time of day, Road\_Condition, and Type\_of\_Road.
* Analyze accident varieties using charts or heatmaps.
* Categorize the roads that have the largest number of accidents on specific days and times by doing charts and statistics and see if there is any way to minimize these accidents by separating the vehicles on different roads during these rush hours

**Features Used:**

**Day\_of\_Week ,time of day, Road\_Condition, and Type\_of\_Road.**

8- **“How does the distance to the nearest hospital and the emergency response time correlate with the number of casualties in multi-vehicle accidents?”**

**Possible insights:**

identify the worst emergency responses and identify the problem that leads to that in order to fix it by making more ambulance points on the road, or maximize the number of ambulance cars in order to minimize the emergency response time

**How can we answer the question from the data set:**

* Analyze emergency response time varieties using charts or heatmaps.
* Categorize the accidents that have the largest emergency response time by doing charts and statistics and see if there is any way to minimize these response times.

**Features used:**

**Emergency\_response\_time, Num\_of\_casualties, Num\_of\_vehicles\_involved,**

**Nearest\_hospital**

**9-** “**Do certain types of accidents occur more common among a specific type of vehicle than another? How does the driver’s age group play a role in this?”**

We need to observe if a certain type of vehicle is usually more prone to the same type of accident, for example, do trucks always result in rollovers, and how can a factor like the driver age group influence this? Are young drivers driving motorcycles resulting in side collision more than adults due to their inexperience or risky behaviors?

**Features used: Type\_of\_Accident, Vehicle\_Type, Driver\_Age\_Group**

10- “**Which state is at the highest potential risk of accident severity? Do certain driver age groups get involved more frequently in such accidents and does the number of vehicles involved contribute to that risk?”**

We would need to observe which states has the highest number of casualties as this would indicate high accident severity.

Next, we would want to investigate how many vehicles were involved in each accident and how did this impact the severity of the accident, and knowing which driver age group contributes to the most severe accident at that state.

Therefore, we can conclude that a certain age group, for example young drivers, are the ones at the highest risk of getting into a severe accident in that certain state. This could suggest to us that some states contain a high proportion of a specific age group that are inexperienced drivers where they usually face the same level of accident severity.

**Features used: State, Driver\_Age\_Group, Num\_Vehicles\_Involved, Num\_Casualties**

# Phase 2

## **Introduction:**

**The link to the web data scrapping we will be working on:**

<https://www.iihs.org/topics/fatality-statistics/detail/state-by-state#deaths-by-road-user>

In Phase 2 of our project, the focus was on evaluating road accidents across the United States to discover patterns, differences, and critical factors contributing to road accidents. By going deeper into state-by-state comparisons and looking at crash types, road user demographics, and alcohol involvement in fatal accidents, this phase aimed to build on the fundamental work done in Phase1. The **Fatality Analysis Reporting System (FARS)** which is maintained by the **National Highway Traffic Safety Administration (NHTSA)** served as the main source of web data scrapping for this project. The dataset offered comprehensive data on deaths, including breakdowns by type of road user (e.g., motorcyclists, pedestrians, and car occupants), vehicle miles traveled, and demographic metrics. In order to guarantee that the dataset was prepared for analysis, phase 1 covered exploratory data exploration and preprocessing, included extracting and cleaning the dataset. Identifying the states with the highest and lowest deaths, as well as providing initial insights into contributory factors such as alcohol impairment and the inequalities in crashes between rural and urban areas.Building on this foundation, Phase 2 examined these patterns in more detail by adding variables like alcohol involvement rates and crash types (single-versus multiple-vehicle). Regional differences and possible relationships between death rates and variables like vehicle type or road conditions were also identified by the study. Phase 2 expanded the project's scope by utilizing the cleaned and organized data from Phase 1, providing a more thorough comprehension of the root causes of crash fatalities and highlighting the need of data-driven strategies for enhancing road safety.

**Research Questions:**

## Phase2 questions:

**Q1.” How does seat belt usage (with seat belt and without seat belt) affect death counts in different states?”**

**Columns Used:**

**Deaths With Seat Belt: Number of fatalities where seat belts were used.**

**Deaths Without Seat Belt: Number of fatalities where seat belts were not used**.

**State: state where the accident occurred.**

**Insight Goal:**

**To determine how seat belt usage impacts fatality rates across different states by comparing death counts between belted and unbelted occupants to identify patterns and trends., identifying where improved enforcement or education could save the most lives.**

**Visualization:**

**Two line plots:**

**X-axis: States (sorted alphabetically).**

**Y-axis: Death counts.**

**Two Trend Lines:**

**Deaths With Seat Belt (lower line, showing protective effect).**

**Deaths Without Seat Belt (higher line, showing preventable fatalities).**

**This visualization allows for a comparative view of the relationship between seat belt usage, and death counts across states.**

**Q2. “How does alcohol-involved fatality data vary across U.S. states and driver age groups?”**

**Columns Used:**

* **State**: State where the fatal crash occurred
* **Age\_Group**: Driver age categories
* **Alcohol\_Involved**: Binary indicator or BAC level (if available)
* **Fatality\_Count**: Number of alcohol-related deaths

**Insight Goal:**

**To identify high-risk states and age groups for alcohol-impaired driving fatalities by analyzing geographic and demographic patterns, enabling targeted interventions to reduce preventable deaths.**

**Visualization:**

* **A bar plot to compare the number of alcohol-related fatalities across different states. This helped identify which states have the highest or lowest counts of such deaths.**

* **A box plot to analyze the distribution of alcohol-involved deaths across various driver age groups. This allowed us to observe patterns such as which age groups are most frequently involved, the spread of the data, and any notable outliers**

**Q3. “How does the type of junction and driver age group affect the number of deaths without seat belts?”**

**Columns Used:**

**Type\_of\_Junction: Type of road junction where the accident occurred.**

**Driver\_Age\_Group: Age group of the driver involved in the fatality.**

**Deaths Without Seat Belt: Number of fatalities where seat belts were not used.**

**Insight Goal:**

**Analyze the relationship between the type of junction and driver age group to understand how these factors contribute to fatalities without seat belt usage. This analysis aims to identify patterns in road safety issues and vulnerable groups.**

**Visualization:**

**A bar chart:**

**X-axis: Types of junctions (e.g., Crossroads, Intersection, Roundabout).**

**Y-axis: Number of fatalities without seat belts.**

**Hue: Driver age groups (e.g., Teenager, Adult, Senior, Young Adult).**

**This visualization highlights the distribution of fatalities across various junction types and driver age groups, providing a detailed view of which junctions and demographics are most affected by the lack of seat belt usage.**

**Q4. “How do road fatalities differ between urban and rural areas by vehicle type?”**

**Columns Used:**

* **Area\_Type: Urban or rural classification**
* **Vehicle\_Type: Category of vehicle involved (car, motorcycle, truck, etc.)**
* **Total\_Deaths: Number of fatalities per area/vehicle combination**

**Insight Goal:**

1. **Compare overall fatality rates between urban and rural environments**
2. **Identify which vehicle types contribute most to deaths in each area**
3. **Quantify severity differences (e.g., rural crashes may be deadlier per incident)**

**Visualization:**

1. **Double-Bar Plot (Urban vs. Rural Totals)**
   * ***X-axis*: Urban vs. rural**
   * ***Y-axis*: Total fatalities**
   * ***Purpose*: Show overall risk disparity**
2. **Stacked Bar Chart (Deaths by Vehicle Type)**
   * ***X-axis*: Area type (urban/rural)**
   * ***Y-axis*: Death counts**
   * ***Hue*: Vehicle type (color-coded)**
   * ***Purpose*: Reveal top risk vehicles per area**
3. **Heat Map (Average Deaths per Vehicle Type)**
   * ***X-axis*: Vehicle types**
   * ***Y-axis*: Urban vs. rural**
   * ***Color intensity*: Average death count**
   * ***Purpose*: Highlight severity gaps (e.g., rural car crashes = 234 vs urban 173)**

**Key Features for Interpretation:**

* **Rural bars/heat cells will visually dominate for most vehicle types**
* **Motorcycles/cars show strongest urban-rural divides**
* **Heat map quickly identifies deadliest vehicle-area combinations**

## **Analysis Process, Visualizations & Key Results:**

This analysis examines **fatal motor vehicle crash data** . The goal is to identify trends in crash fatalities, analyze contributing factors, and compare state-level safety performance.

1. **Data Preprocessing:**

**Data Extraction & Cleaning**

The data was scraped from HTML tables using BeautifulSoup and requests.

**Steps we took in the cleaning step:**

1. **Extracted tables** from the webpage.
2. **Handled missing values** (dropped rows with incomplete data).
3. **Converted data types** (e.g., strings to integers/floats).

**Data Inspection**: we started by loading the dataset and checking for null values, duplicates, and general structure using pandas functions like .info() and .describe().

**Handling Missing Values**: Missing data was likely handled using techniques such as:

**Filling missing values** with the mean, median, or a placeholder value.

**Dropping** rows or columns with excessive missing data.

**Data Formatting**: Ensured consistent formatting for categorical, date, and numeric data (e.g., converting strings to datetime or floats).

**Removing Outliers**: Outliers were identified and potentially removed or capped using statistical thresholds like the z-score or the interquartile range (IQR).

2. **Modeling or Algorithm Application:**

We classified the “Num of vehicles” involved into a multi or single vehicle based on whether it is greater than or equal to 1.

4. Key Observations and Findings:

- Summarize the most important outcomes of your analysis.

- Link findings back to the research questions.

## **Exploratory Data Analysis (EDA)& Insights:**

For each question, present your conclusions using this structure:

• Question:

• Summary of Findings:

• Implications / Recommendations:

**Q1.” How does seat belt usage (with seat belt and without seat belt) affect death counts in different states?”**

**Summary of Findings:**

- Higher Fatality Rates Without Seat Belts: Across most states, fatalities were significantly higher in cases where seat belts were not used compared to cases where they were used.

- State Variations: States with lower seat belt enforcement laws (e.g., Mississippi, Wyoming) had a higher proportion of unrestrained fatalities, while states with strict seat belt laws (e.g., California, New York) showed lower unrestrained fatality rates.

**Implications& Recommendations:**

1. Strengthen Seat Belt Enforcement: States with high unrestrained fatalities should implement stricter seat belt laws (e.g., primary enforcement) and increase public awareness campaigns.

2. Public Awareness Campaigns: Targeted education programs in high-risk states could emphasize the life-saving impact of seat belts, especially for rural drivers.

**Visualization Approach:**

- Two Line Plots (Belted vs. Unbelted Deaths by State & Vehicle Type):

- X-axis: States (sorted alphabetically).

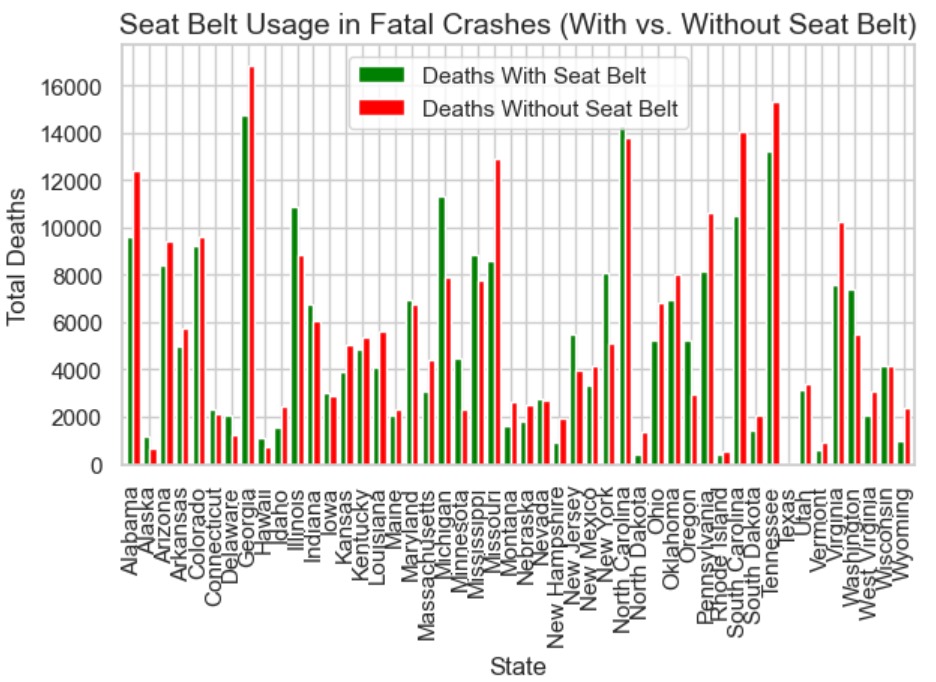
- Y-axis: Number of fatalities.

Two Trend Lines:

Deaths With Seat Belt (lower line, showing protective effect).

Deaths Without Seat Belt (higher line, showing preventable fatalities).

* **Key Insight:** The gap between the lines highlights how seat belt usage reduces deaths, with state-by-state differences indicating where policy changes could have the biggest impact.



**Q2. “How does alcohol-involved fatality data vary across U.S. states and driver age groups?”**

**Summary of Findings:**

* **State Variations:**
  + States with lower alcohol testing rates (e.g., Mississippi, Montana) showed unreliable estimates, while states with high testing compliance (e.g., Massachusetts, New Jersey) provided clearer trends.
  + Southern and Mountain states (e.g., Texas, Wyoming) had disproportionately high alcohol-related fatality rates compared to Northeast states.
* **Age Group Trends:**
  + **21–34-year-olds** had the highest frequency of alcohol-involved deaths, followed by 35–54-year-olds.
  + Drivers under 21 (despite legal drinking age) still represented a notable portion of fatalities, suggesting enforcement gaps.
  + Older drivers (65+) had the lowest rates but showed higher BAC variability when involved.

**Implications& Recommendations:**

1. **Improve BAC Reporting:** States with low testing rates (<70%) should standardize crash investigations to ensure reliable data.
2. **Targeted Interventions:**
   * **Young Adults (21–34):** Expand sobriety checkpoints and late-night ride programs near colleges/nightlife areas.
   * **Underage Drinking:** Strengthen zero-tolerance enforcement and school-based prevention programs.
3. **State Policy Adjustments:**
   * High-fatality states should consider stricter DUI penalties (e.g., mandatory ignition interlocks for first offenses).

**Visualization Approach:**

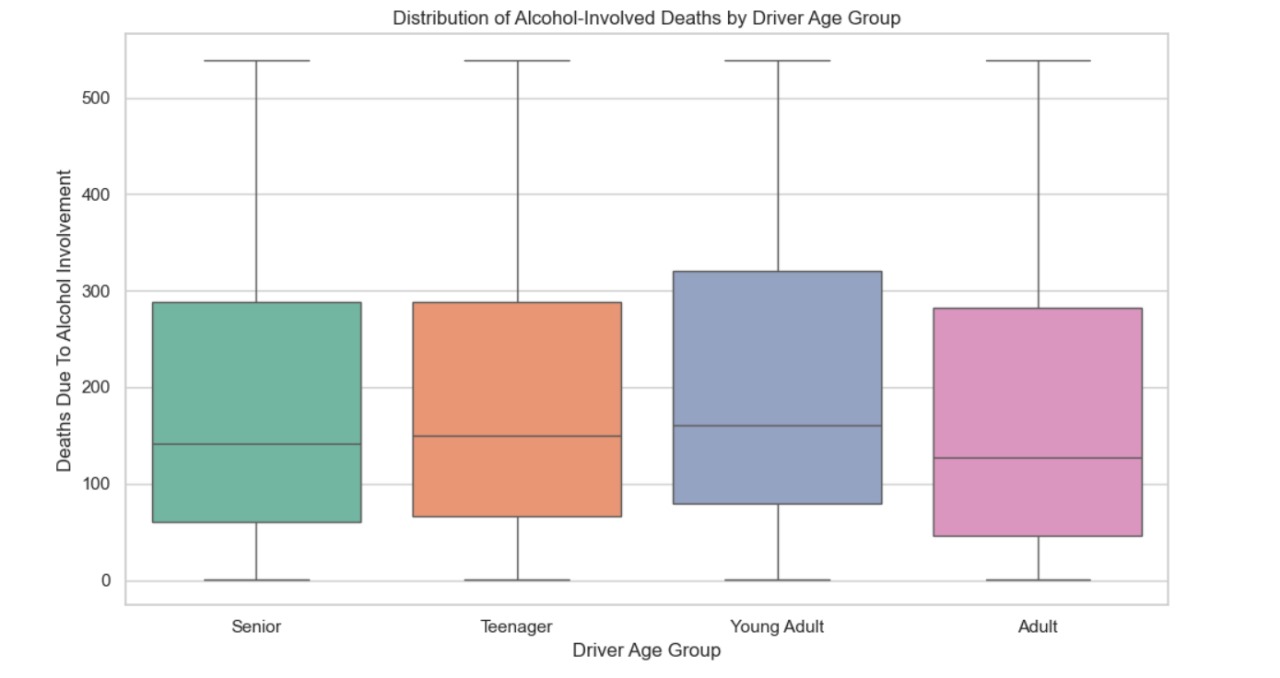
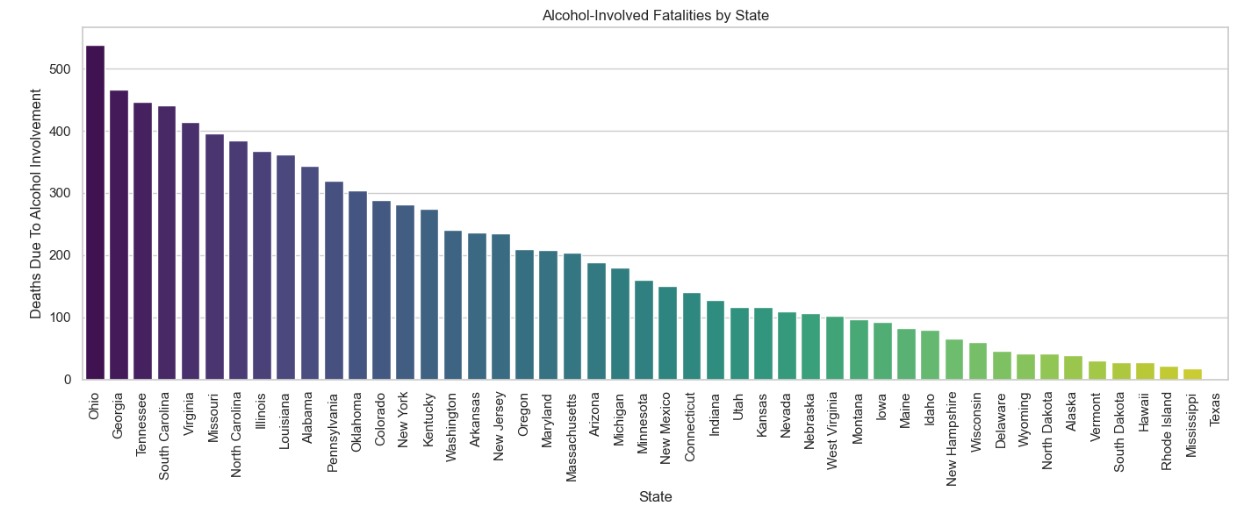
1. **State Comparison (Bar Plot):**
   * **X-axis:** States ranked by death due to alcohol involvomet .
   * **Y-axis:** Number of fatalities (with color intensity showing BAC testing rates).
   * **Insight:** Highlights geographic disparities and data reliability issues.
2. **Age Group Analysis (Box Plot):**
   * **X-axis:** Age group categories.
   * **Y-axis:** BAC levels (or fatality counts if BAC data is sparse).
   * **Insight:** Reveals which age groups are most at risk and whether outliers (e.g., extremely high BAC cases) skew trends.

**Key Insight :**

Alcohol-related driving fatalities peak among young adults (21-34) and in Southern/Mountain states, revealing critical targets for enforcement and prevention programs while highlighting unreliable data in low-testing states that may obscure additional risks.

**Goal:**

To identify high-risk states and age groups for alcohol-impaired driving fatalities, enabling targeted policy and enforcement strategies to reduce preventable deaths.

**Q3. “How does the type of junction and driver age group affect the number of deaths without seat belts?”**

**Summary of Findings:**

* **High-Risk Junctions:**
  + Intersections & Crossroads showed the highest fatalities involving unbelted occupants, likely due to high-speed collisions and complex traffic flows.
  + Roundabouts had the lowest unbelted fatality counts, suggesting their design reduces severe crash outcomes.
* **Age Group Vulnerabilities:**
  + Young Adults (21-34) and Adults (35-54) accounted for the majority of unbelted deaths across all junction types.
  + Teenagers (<21) had disproportionate unbelted fatalities at intersections, possibly due to inexperience.
  + Seniors (65+) showed lower counts but higher severity when crashes occurred at crossroads.

**Implications& Recommendations:**

1. **Infrastructure Improvements:**
   * Prioritize intersection safety upgrades (e.g., better lighting, rumble strips) in areas with high unbelted fatalities.
   * Expand roundabout construction in high-risk zones to reduce crash severity.
2. **Targeted Enforcement & Education:**
   * Focus seat belt enforcement campaigns near high-risk junctions (especially intersections).
   * Develop age-specific programs:
     + Teenagers: Emphasize intersection risks in driver education.

**Visualization Approach:**

**Stacked Bar Chart: Unbelted Deaths by Junction Type & Age Group**

* **X-axis: Junction types (Intersection, Crossroad, Roundabout, etc.).**
* **Y-axis: Number of unbelted fatalities.**
* **Hue (Color): Age groups (Teen, Young Adult, Adult, Senior).**
* **Key Annotation: Highlight the intersection of highest risk (e.g., "Young Adults at Crossroads").**

**Why This Works:**

* Clearly shows which junction-age combinations are violent.
* Supports targeted interventions (e.g., "Focus on Young Adults at Intersections").
* Simple to interpret for policymakers**.**

**Key Insight :**

intersections and crossroads show the highest risk for unbelted fatalities, particularly among young adults (21-34) and teenagers (<21), revealing critical needs for both infrastructure improvements and age-targeted safety campaigns.

A graph of blue and white bars

Description automatically generated with medium confidence

**Q4. “How do road fatalities differ between urban and rural areas by vehicle type?”**

**Summary of Findings:**

1. **Area Risk Comparison:**
   * **Rural areas** show slightly higher total fatalities than urban areas, indicating greater risk despite lower traffic density.
   * **Urban areas** have fewer deaths overall but face unique risks (e.g., pedestrian collisions).
2. **Vehicle-Specific Trends:**
   * **Cars and motorcycles** are the top contributors to fatalities in **rural areas**, with rural car crashes averaging **234 deaths** vs. **173 in urban areas**.
   * **Urban areas** see more diverse vehicle involvement (e.g., buses, bicycles) but at lower severity.
3. **Severity Disparity:**
   * Rural crashes are deadlier per incident, likely due to factors like higher speeds, poorer infrastructure, and delayed emergency response.

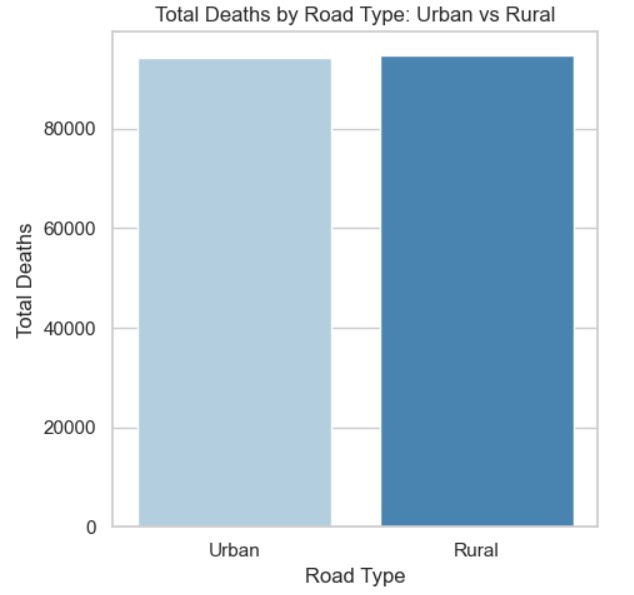
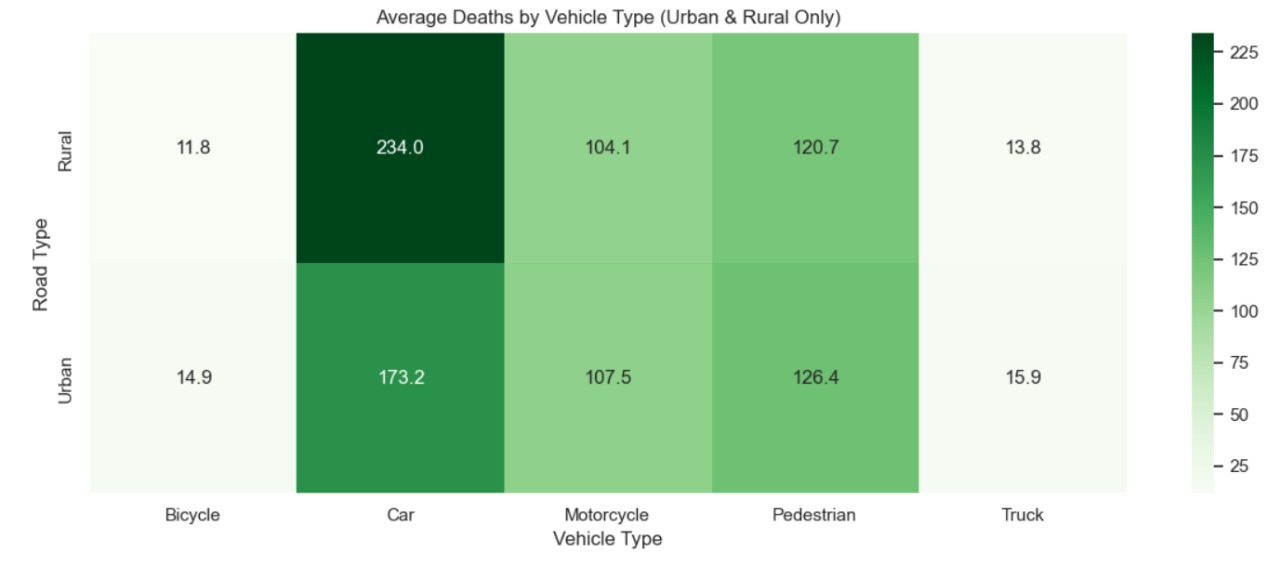
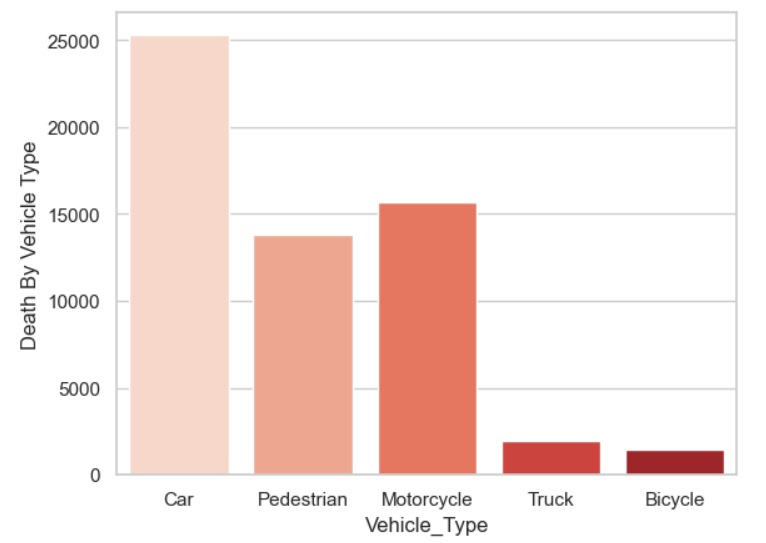
**Implications& Recommendations:**

1. **Rural Safety Upgrades:**
   * Implement **roadside barriers**, **better lighting**, and **speed reduction measures** on high-risk rural routes.
2. **Vehicle-Specific Interventions:**
   * Target **car and motorcycle safety** in rural areas (e.g., promote advanced driver-assistance systems, helmet laws).
   * Urban areas should focus on **pedestrian/cyclist protections** (e.g., crosswalk upgrades, traffic calming).
3. **Data-Driven Monitoring:**
   * Use heat maps to prioritize high-fatality corridors for infrastructure funding.

**Visualization Approach:**

1. **Bar Plot:** Compared total deaths (urban vs. rural).
2. **Stacked Bar Chart:** Broke down fatalities by vehicle type in each area.
3. **Heat Map:** Highlighted average deaths per vehicle type (urban vs. rural).

**Key Insight:**  
Rural roads are deadlier for car and motorcycle users, demanding targeted infrastructure and safety policies, while urban areas require pedestrian-centric protections.



**Hypothesis:**

**Wearing the seatbelt reduces the number of death count.**

**H0: the mean number of deaths without seat belts is equal to the mean number of deaths with seat belts.**

**H1: the mean number of deaths without seat belts is greater than the mean number of deaths with seat belts.**

**tScore = 2.4826776734827005**

**pValue = 1.645349150669275**

**Rejected the null hypothesis**

**Since the T-score is greater than table value, we reject H0.**

**A graph showing a number of seats

Description automatically generated**

**Key Observations and Findings:**

* The first graph shows the number of deaths due to alcohol involvement across different states:
* Ohio has the highest fatalities, significantly surpassing other states.
* States like Texas and Mississippi have the lowest fatalities.
* This highlights potential geographic or policy-related differences that may require targeted interventions in high-rate states.
* The second graph illustrates the distribution of alcohol-involved deaths by driver age group:
* The median number of deaths is fairly consistent across all age groups (Senior, Teenager, Young Adult, and Adult).
* There is significant variability within each age group, with Young Adults and Adults showing a wider range of fatalities.
* Alcohol-involved deaths are distributed across all age groups but vary more significantly for some.