RETHINK THE ROADS:
Unexpected Statistics on Vehicle Collisions in New York State (2018-2022)

Team Members

Lekhith Reddy Kambham
Rithvik Rangaraj
Sampada Regmi
Sandeep Kanaram Parmar
Ayushi Sharma

Data Source: Motor Vehicle Crashes at NY - Motor Vehicle Crashes - Case Information: Three Year Window | State of New York (ny.gov)

Project Description:

Motor vehicle crashes pose a significant threat to public safety, with far-reaching consequences for individuals, communities, and governments. Understanding the underlying patterns and factors contributing to crashes is crucial for implementing effective preventive measures and improving overall road safety. This project aims to transform raw data into visually compelling and actionable information for a diverse range of stakeholders, from transportation officials to law enforcement and the public. The Motor Vehicle Crashes - Case Information dataset used, spanning a three-year window, serves as a comprehensive repository of information that, when harnessed effectively, provided valuable insights. Tableau's ability to integrate various data types, including geospatial data, makes it an ideal platform for uncovering hidden trends and correlations within the motor vehicle crashes dataset. We have uncovered various significant trends in the history of vehicle crashes in New York from 2018 to 2022, examining both the timing and severity of the accidents.

Dashboard Description:

1. ***Home:*** The homepage dashboard offers a concise overview of the entire project, aiding users in comprehending the subsequent dashboards. It includes well-coordinated navigation buttons for each dashboard, enhancing user-friendliness. Additionally, an interactive filled map of NY state counties displays the number of accidents over the years. This feature dynamically updates the "Total Vehicle Crashes" block as users hover over each county, making the homepage both informative and visually appealing.

2. ***Chronological Vehicle Collision Summary:*** This dashboard presents a comprehensive summary of vehicle accidents in New York over the years 2018-2022. The dashboard, structured around time divisions, offers users a clear overview of accident statistics over this period, with further breakdowns by month, day, and hour. This detailed visualization demonstrates the relationship between the timing of

accidents and their occurrence in New York. To enhance clarity and improve visualization, we have consolidated the years into two groups: 2018-2019, 2020 & 2021-2022, using calculated fields and parameters. Similarly, for hours of the day, we have organized them into four day-part segments, also utilizing calculated fields and parameters.

3. ***Unveiling Accident Severity:*** Building on the vehicle collision summary provided in the previous dashboard, this dashboard focuses on the severity of accidents over the past four years. It is highly interactive where users can filter data by specific years and levels of accident severity. Leveraging the capabilities of calculated fields, parameters, and dashboard actions in Tableau, we have crafted an engaging and dynamic experience. Users can easily explore the following information from this dashboard, filtered by year and accident severity:

    - The number of accidents in each county is visualized on a filled map.
    - The causes of collisions are organized by the frequency of accidents using a bubble chart.
    - Counts of accident incidents on each weekday using a line chart.
    - The influence of weather conditions on the number of accidents using a bar graph
    - The effect of road surface conditions on the number of accidents using a tree map.

4. ***Most & Least crash-prone Counties:*** This dashboard delivers insights into the counties with the highest and lowest incidence of crashes for each year group, featuring an interactive filter to facilitate analysis.

Supporting Graphs:

We also have several graphs that were not a part of the main dashboard but helped us find trends in the data that enabled us to design our dashboard and be able to tell a story to convey the message in the most efficient way possible.

Predictive Analytics with Our Data

Our data is such that it has the potential to provide useful insights, provided we dig deep into it. We decided to use this flexibility and implement a few machine learning models to predict our most valuable column: the number of vehicles involved i.e, predict the maximum likelihood of weather/road surface/lighting conditions that would result in the maximum number of accidents

Models Used

On initial analysis, preprocessing, and feature analysis, to predict the weather condition that would cause the most number of accidents, we used Poisson regression.

Results:

- ***The Coefficient for Weather Conditions:*** Fog/Smoke is the strongest negative value whereas snow/ice is the strongest positive value which indicates that Fewer accidents are to occur when it is foggy and most accidents occur when it is snowy. A chart in Tableau shows the number of vehicles involved if the weather conditions are foggy/snowy and visually, there are more accidents.

- ***Association Rule Mining:*** We used it to detect the combination of weather, road surface, and lighting conditions that would result in more vehicles being involved in the accidents.

- ***Results:*** We found that bright light, Clear Weather, and slightly slippery roads lead to the most accidents. We also analyzed the relationship between these factors to determine if they actually played a role in determining the likelihood of accidents. The result was that most factors out of all possible road, lighting, and weather conditions showed a high confidence of greater than 75%, showing that they do indeed play a role.