# Data Visualization Final Project

**Analysing Traffic Collision Data: Enhancing Road Safety in Montgomery County**

# Contents

Abstract

Introduction

Description of Dataset

Preprocessing the dataset

Outlier Detection and Analysis

Exploratory Data Analysis

Business Question Plots

Conclusion

**1** **Abstract**

The Montgomery County Traffic Collision Dataset is a valuable resource containing detailed information on 172,105 traffic accidents that occurred on county and local roads within Montgomery County. This dataset, sourced from the Maryland State Police's Automated Crash Reporting System, spans 43 columns and offers insights into various aspects of collision occurrences and involved drivers. Analyzing this dataset can contribute significantly to identifying accident hotspots, enhancing road safety measures, and influencing policy choices aimed at preventing traffic collisions and improving road safety in the area. The dataset's accessibility promotes community collaboration, facilitates efficient resource allocation, and encourages research into innovative traffic collision prevention solutions.

**2** **Introduction**

The Montgomery County Traffic Collision Dataset is a comprehensive compilation of 172,105 records spanning 43 columns. It offers a detailed overview of traffic accidents occurring on county and local roads within Montgomery County. Sourced from the Maryland State Police's Automated Crash Reporting System, this dataset provides insights into collision occurrences and involved drivers. Analyzing this dataset can lead to crucial insights that can help save lives and enhance road safety measures.

**3 Data Description:**

- Report Number: Unique identifier for each collision report.

- Local Case Number: Local case number assigned to the collision by the reporting agency.

- Agency Name: Name of the law enforcement agency reporting the collision.

- ACRS Report Type: Type of crash report according to the Automated Crash Reporting System.

- Crash Date/Time: Date and time when the collision occurred.

- Route Type: Type of route where the collision occurred (e.g., Maryland State, County).

- Road Name: Name of the road where the collision occurred.

- Cross-Street Type: Type of street intersecting with the road where the collision occurred.

- Cross-Street Name: Name of the intersecting street.

- Off-Road Description: Description of the location off the road where the collision occurred.

- Municipality: Municipality where the collision occurred.

- Related Non-Motorist: Involvement of non-motorists (e.g., pedestrians, cyclists) in the collision.

- Collision Type: Type of collision (e.g., rear-end, head-on).

- Weather: Weather conditions at the time of the collision.

- Surface Condition: Condition of the road surface at the time of the collision.

- Light: Lighting conditions at the time of the collision.

- Traffic Control: Type of traffic control at the collision site (e.g., traffic signal).

- Driver Substance Abuse: Whether the driver was under the influence of substances.

- Non-Motorist Substance Abuse: Whether the non-motorist involved was under the influence of substances.

- Person ID: Unique identifier for each person involved in the collision.

- Driver At Fault: Whether the driver was at fault for the collision.

- Injury Severity: Severity of injuries sustained in the collision.

- Circumstance: Circumstances surrounding the collision.

- Driver Distracted By: Factors distracting the driver at the time of the collision.

- Drivers License State: State issuing the driver's license.

- Vehicle ID: Unique identifier for each vehicle involved in the collision.

- Vehicle Damage Extent: Extent of damage to the vehicle.

- Vehicle First Impact Location: Location on the vehicle where the first impact occurred.

- Vehicle Second Impact Location: Location on the vehicle where the second impact occurred.

- Vehicle Body Type: Body type of the vehicle involved.

- Vehicle Movement: Movement of the vehicle at the time of the collision.

- Vehicle Continuing Dir: Direction in which the vehicle continued after the collision.

- Vehicle Going Dir: Direction in which the vehicle was traveling at the time of the collision.

- Speed Limit: Speed limit on the road where the collision occurred.

- Driverless Vehicle: Whether the vehicle involved was driverless.

- Parked Vehicle: Whether the vehicle involved was parked.

- Vehicle Year: Year of manufacture of the vehicle.

- Vehicle Make: Make or manufacturer of the vehicle.

- Vehicle Model: Model of the vehicle.

- Equipment Problems: Any equipment problems reported with the vehicle.

- Latitude: Latitude coordinate of the collision location.

- Longitude: Longitude coordinate of the collision location.

- Location: Combined latitude and longitude coordinates of the collision location.

The data was collected from the Crash Reporting - Drivers Data dataset on Data.gov. It's sourced from the Maryland State Police's Automated Crash Reporting System and reported by local law enforcement in Montgomery County. You can access it on the Data.gov website.

**Source:** <https://catalog.data.gov/dataset/crash-reporting-drivers-data>

**4 Dataset Preprocessing**

The dataset has undergone multiple preprocessing stages in order to address missing values for improved analysis.

**4.1 Identifying Missing Values:**

We have used isnull() to determine the percentage of missing values in each dataset column.sum() method, and we discovered that the dataset contains a large number of missing values.

**4.2 Handling Missing Values:**

* Replaced missing values in 'Collision Type' with the most common value.
* Replaced missing values in 'Weather' and 'Surface Condition' with their respective modes.
* Replaced missing values in 'Traffic Control' with 'Unknown' or 'Not Specified'.
* Replaced missing values in 'Vehicle Model' with the mode of the column.
* Converted 'Crash Date/Time' to a normal datetime format.
* Converted object-type columns to strings.

**4.3 Feature Engineering:**

* Using the replace() function and a predefined dictionary (corrections), inconsistent or incorrect values are corrected to ensure uniformity and accuracy in the dataset for the vehicle make column.
* We used the random state ML algorithm to handle the data after resampling it because it was difficult to read and visualize due to the large amount of data.
* We have exported the CSV file to work on Tableau.

**4.4 Reverification:**

Finally, we have rechecked the data making sure there are no null values in the dataset.

A screenshot of a computer screen

Description automatically generated A screenshot of a computer

Description automatically generated

**5 Exploratory Data Analysis**  
A screenshot of a computer screen

Description automatically generated

The waffle chart illustrating the distribution of accidents by route type reveals significant insights into the occurrence of accidents across different road categories. Notably, Maryland state roads stand out as the most frequent location for accidents, constituting a substantial proportion of the total incidents at 44.93%. Conversely, service roads experience the lowest incidence of accidents, representing a mere 0.02% of the total. This visualization effectively highlights the disparity in accident rates across various types of routes, emphasizing the prominence of Maryland state roads in accident occurrence while underscoring the relatively low frequency of accidents on service roads.

A close up of words

Description automatically generated

In the word cloud chart depicting vehicle makes, Toyota emerges as the most prominent company, indicating its prevalence in the dataset. This observation underscores Toyota's significant presence in reported accidents compared to other car manufacturers. Similarly, Honda, Chevy, and Ford also stand out prominently in the word cloud, suggesting their frequent occurrence in recorded accidents. This insight offers valuable information about the distribution of accidents across different vehicle brands, with Toyota notably leading in frequency within the dataset.

A map of a city

Description automatically generated

The chloropleth chart depicting the distribution of accidents across Montgomery County offers valuable insights into accident patterns within the region. Specifically, the chart highlights that state highways in Maryland serve as the primary location for accidents, with a notable majority occurring in this category. This observation underscores the significance of state highways as high-traffic areas prone to accidents within Montgomery County. By visually representing the geographical distribution of accidents, the chloropleth chart effectively communicates the concentration of incidents on state highways, thereby facilitating a better understanding of accident hotspots within the county.A graph of a graph with a pie chart and text

Description automatically generated with medium confidence

The bar graph illustrates the number of accidents categorized by the year the vehicles were manufactured, offering insights into accident trends relative to vehicle production years. It is evident that the highest number of accidents occurred among vehicles manufactured between 2013 and 2019. This observation aligns with the likelihood that more vehicles were produced during this period, reflecting increased vehicle ownership and usage. The visual representation effectively highlights the correlation between vehicle manufacturing years and accident frequency, emphasizing the significance of understanding temporal patterns in accident data analysis.

**6 Smart Questions**

6.1 How many accidents does Montgomery County account for every year and how severe are they? A graph of a green and blue line

Description automatically generated with medium confidence

The bar chart we've constructed serves to analyze the annual occurrence of accidents within the county, presenting a breakdown of accident numbers by year alongside their severity, denoted by distinct colors. By visually representing this data, we can discern trends in accident frequency over time and gain insights into the severity distribution across different years. This visualization aids in identifying patterns and fluctuations in accident rates, thereby facilitating informed decision-making and targeted interventions aimed at improving road safety within the county.

4.2 Are drivers responsible for accidents?A graph with blue and orange squares

Description automatically generated

In response to our analysis into the role of drivers in collisions, our analysis reveals that a significant portion of collisions can be attributed to driver error. Furthermore, the majority of these collisions appear to result in no obvious injuries. This finding underscores the importance of understanding and addressing driver behavior as a key factor in preventing accidents and promoting road safety. By identifying the prevalence of driver-related errors in collisions and their potential consequences, we can prioritize interventions and initiatives aimed at enhancing driver education, awareness, and adherence to traffic regulations, ultimately contributing to a reduction in accidents and injuries on the roads.

4.3 Is there a third party involvement in the accidents?A screenshot of a white and orange graph

Description automatically generated Despite driver error being the primary contributing factor to the majority of incidents, our analysis reveals that pedestrian interference is a notable factor in accidents involving third parties. This observation suggests that third parties, such as pedestrians, play a significant role in contributing to accidents. By recognizing the involvement of third parties in accidents, particularly in cases of pedestrian interference, we can identify areas for targeted interventions and safety measures aimed at reducing pedestrian-related accidents. Additionally, this insight emphasizes the importance of promoting pedestrian safety awareness and implementing measures to enhance pedestrian-vehicle interaction, ultimately fostering safer road environments for all road users.

4.4 Is there any reason behind the accident’s occurrence? A screenshot of a computer

Description automatically generated In our analysis depicted in the graph, we aimed to identify contributing factors to driver accidents. Our findings reveal that, while some incidents are associated with drivers ingesting alcohol and other substances, the majority of accidents occur when drivers haven't consumed any substances. This insight highlights the significance of understanding the various factors contributing to driver accidents, including substance use, and underscores the importance of promoting responsible driving behavior and adherence to traffic regulations. By recognizing the prevalence of accidents unrelated to substance consumption, we can focus on comprehensive strategies aimed at addressing all potential causes of driver accidents, ultimately striving towards safer road environments for all.

4.5 Are accidents based on weather conditions/age of the vehicle?A screenshot of a computer

Description automatically generated

In our graph, we aimed to assess whether weather conditions and the age of vehicles are factors influencing accidents. Based on our analysis of nearly 8,000 accidents occurring over a nearly three-year period from 2010 to 2015, we have observed that both weather conditions and the age of vehicles indeed have an impact on the frequency of accidents. This finding underscores the importance of considering multiple variables when analyzing accident data and highlights the need for comprehensive approaches to road safety that address factors such as weather conditions and the condition of vehicles. By recognizing the influence of these factors, we can develop targeted interventions and strategies aimed at reducing the occurrence of accidents and promoting safer road environments.

4.6 Does the speed limit on the road affect the severity of an accident?A graph with green lines and orange and green lines

Description automatically generated with medium confidence The graph presented illustrates the frequency of accidents across different types of roads with varying speed limits. Upon analysis of the data, we can infer that a significant portion of accidents occur on roads with a speed limit of 35 mph. This observation leads us to conclude that maintaining the speed limits on these roads could be instrumental in reducing the number of accidents that occur. By recognizing the correlation between speed limits and accident frequency, we can advocate for measures aimed at promoting safer driving behaviors and enforcing speed limits on roads where accidents are prevalent. This proactive approach aligns with efforts to enhance road safety and mitigate the risk of accidents on high-traffic routes.

4.7 Does the speed limit on the road affect the rate of the injury?A graph of different colored bars

Description automatically generated As depicted in the graph, the majority of accidents are associated with no obvious injuries, and a significant proportion of these accidents happen on roads with a speed limit of 35 mph. This observation underscores the importance of considering both the severity of accidents and the road conditions where they occur when assessing road safety measures. By recognizing the prevalence of accidents with no obvious injuries and their concentration on roads with a 35 mph speed limit, we can prioritize targeted interventions aimed at reducing accident rates and promoting safer driving practices on these specific road types. This holistic approach aligns with efforts to enhance road safety and minimize the impact of accidents on individuals and communities.

**5 Conclusion**

In conclusion, our analysis sheds light on several key findings regarding accidents within the dataset. Firstly, a significant portion of accidents can be attributed to third parties, with pedestrians playing a prominent role. Secondly, these accidents predominantly occur during daylight hours and often transpire on roads with a speed limit of 35 mph. Thirdly, the year 2017 stands out as having the highest frequency of accidents among the years considered in the dataset. Finally, it is noteworthy that the majority of accidents result in no apparent injuries. These insights provide valuable information for understanding accident patterns and underscore the importance of implementing targeted interventions to enhance road safety and mitigate the risk of accidents, particularly those involving pedestrians and occurring within specific road conditions and timeframes.

**5.1 Challenges**

- Limited utilization of available data rows.

- Presence of numerous null values hindering analysis depth.

**5.2 Future Directions:**

Integration of Machine Learning:

- Opportunity to leverage predictive capabilities for forecasting and anomaly detection.

- Potential for deeper insights and more flexible analysis.

**5.3 Lessons Learned:**

- Understood the importance of preprocessing. Spent more time on cleaning the data and handling the missing values.

- Some charts are complex, so we utilized the tableau to make the charts.

- We got to know how impactful the data can be to make data driven decision if we make insightful visualizations.