Ava Haghighi
Data 205
CRN :22017

Title: Accident Rate and contributing factors

According to the Federal Highway Administration (FHWA) , in the state of Maryland, 23% of accidents are related to county roads. Therefore, analyzing accidents at the individual county level can help identify which areas require more attention for improvements in road infrastructure and other safety factors. This report investigates the relationship between various traffic and environmental factors and accident rates in Montgomery County. Using crash report data, this project explores patterns and aims to guide transportation policy to enhance road safety.

The main dataset used in this project is the Crash Report by Montgomery data, which includes 36 variables related to accidents. Vehicle factors include vehicle year, make, and model. Environmental factors consist of surface conditions, visibility levels, weather conditions, and the time of day and week, as well as the year. Accident characteristics cover factors such as intersection or non-intersection locations and the exact site of the accident.
Another dataset used for traffic volume information in the county is the MDOT SHA Annual Average Daily Traffic (AADT) data. Variables such as annual average daily traffic (AADT) are used for calculations of Vehicle Miles Traveled (VMT) . The source of this data is Maryland.gov Open Data.

As the Federal Highway Administration (FHWA) of the U.S. Department of Transportation highlights, Analyzing both crash average trend and crash rate over the years is important because they provide complementary insights into the behavior of crash patterns.These analyses are crucial for supporting roadway safety initiatives and identifying high-risk locations or patterns. In addition to these analyses, further investigations are conducted to explore the relationship between crash counts and other variables within the crash dataset.By examining these relationships, we aim to identify patterns that can provide a better understanding of the factors influencing crash frequency in Montgomery County.

The method used in this project to calculate accident frequency is provided by the Federal Highway Administration of the U.S. Department of Transportation. This approach suggests analyzing accident data by first calculating a rolling average for accidents. The rolling average of accident counts, calculated over three years, helped smooth fluctuations and highlight longer-term trends in accident patterns. Similarly, the inclusion of AADT data helps contextualize accidents in relation to traffic volumes, ensuring the results are more accurately adjusted for road usage. Additionally, the accident rate was calculated using the formula: crash count multiplied by 100,000, normalized over the Average Daily Traffic (ADT) for each year, and then divided by the length of the road segment. The methods used to understand the relationships between variables in this project include Spearman's method, which was employed to determine whether there is a correlation between two variables, namely accident count and speed limit, which is the only continuous variable analyzed. Linear regression analysis was also applied to explore the relationship between accident count and variables such as surface condition, vehicle make, vehicle model, weather, light conditions, day of the week, and injury severity. All calculations were conducted using the programming tool R.

To determine if accidents occur predominantly around infrastructure areas, choropleth maps were created using ArcGIS Pro. These maps visualize the locations of accidents and infrastructure areas. The count of accidents in each census block was mapped alongside the count of ongoing infrastructure projects, and a correlation was calculated. Another map created for this project is a kernel density map for bike trails and accident locations. This map helps analyze and visualize the relationship between the distribution of bike trails and accident hotspots, providing insights into traffic safety. It also allows for assessing whether Montgomery County's bikeways are effective in reducing the frequency of accidents involving bikes and vehicles near the trails.

Data cleaning for this project primarily focused on the crash data, which contained variables with numerous levels. Key variables were grouped for clarity and analysis. For example, driver distractions were categorized into groups such as "Looked but Did Not See," "Physical Activities Distraction," "Distracted by Other Occupants," "Distracted by Device," and "Lost in Thoughts." Weather conditions were consolidated into categories like "Rainy," "Snowy," "Clear," "Wind," and "Fog." Surface conditions were grouped as "Wet," "Ice," and "Clear." Injuries were classified into "No Injury," "Serious Injury," "Minor Injury," and "Fatal Injury." Additionally, vehicle mode was one-hot encoded into "EV" and "Non-EV" categories for future analysis. This process simplified the dataset, making it more structured and suitable for analytical purposes.The grouping of vehicle types, weather conditions, and injuries into manageable categories allowed for better identification of significant patterns.

Summary statistics for a subset of variables in the dataset, focusing on both numeric and categorical variables, are provided below. The first variable, 'Day of the Week,' shows a clear distribution of accidents across different days. Friday and Saturday have notably higher accident counts compared to other days. The 'Crash Date' variable ranges from January 1st, 2015, to November 11th, 2024. The 'Speed Limit' variable ranges from a minimum of 0 (indicating parked vehicles) to a maximum of 75 mph, with a mean of 31.38 mph and a median of 35 mph. This suggests that the majority of accidents occurred on roads with speed limits around 30-40 mph.

In terms of injury severity, most accidents did not result in injuries, with 83,565 out of 97,472 accidents being injury-free. Among the accidents that involved injuries, the majority were classified as minor (6,015 cases) or serious (6,946 cases), while fatalities were rare, accounting for just 99 cases. The 'Other Serious Injury' category further suggests that although most accidents are less severe, a significant number of serious injuries still occur.

Surface conditions reveal that the majority of accidents occurred on dry surfaces (65,603), followed by wet surfaces (15,911), snow and ice (1,246), and other surface types (141). Lighting conditions also vary, with 23,267 incidents occurring under 'Dark with Lights,' 3,247 incidents under 'Dark, No Lights,' 4,068 incidents during 'Dawn and Dusk,' and 64,332 incidents during

'Daylight.' Notably, substance abuse was involved in only 47 cases out of the 97,472 total accidents.

Based on the bar plot visualizing crash dates by month, October emerges as the month with the highest number of accidents. This summary of key statistics helps in understanding patterns in accident occurrence, providing a foundation for further analysis and the development of safety improvements

Base on the result of From 2015 to 2019, the number of accidents remained relatively stable, with minor fluctuations each year. However, in 2020, there was a significant decline in accident counts, likely due to the impact of the COVID-19 pandemic, which led to reduced traffic and travel. The accident count dropped to 7,591 in 2020, much lower than previous years. In 2021 and 2022, accident numbers began to rise again, showing a recovery trend, although they did not reach the levels observed in 2019. Following the count of the accidents by each year, the rolling average shows a clear reduction in 2020 due to the pandemic but can also reveal the recovery process by showing how quickly the accident numbers bounce back after that drop. But the overall all trend line shows that accident counts in Montgomery County are high but slowly falling.

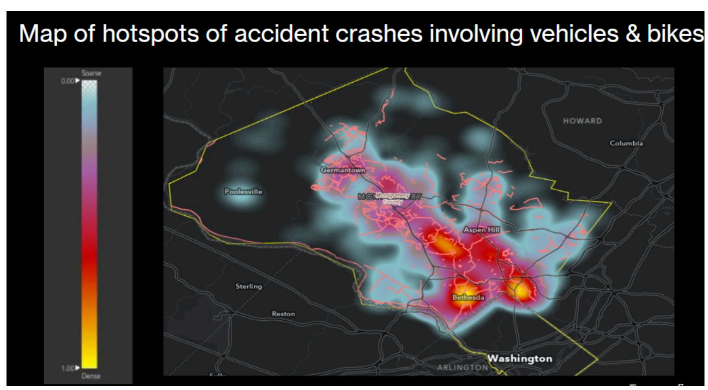
In this project, Spearman's rank correlation test, which is appropriate for non-parametric data, was used to determine whether there was a monotonic relationship between accident counts and speed limits. The test yielded a weak negative correlation with a coefficient of -0.115. However, the p-value was not statistically significant, indicating that the observed correlation may be due to chance. Therefore, it failed to reject the null hypothesis, which suggests there is no meaningful association between accident count and speed limit based on this test. After calculating the number of accidents for each vehicle and dividing it by the total number of accidents, a bar plot was created to visualize the top ten vehicle makes, with Toyota appearing at the top. However, this cannot indicate that the Toyota make is a contributing factor to accidents, as the calculation lacks precision. For a more accurate analysis, additional factors such as the market share of each vehicle make or the number of registered vehicles for each make in Montgomery County should be considered.

To understand if the type of vehicle (EV or none EV) has any relationship between injury severity variable with the null hypothesis indicating no association and alternative indicating strong association for vehicle type with injury severity, a multinomial logistic regression model following manually calculated z-values and p-values to assess the significance of each coefficient was made that indicted the result of the vehicle type, specifically Non-EVs, significantly influences the likelihood of different injury categories. For Minor Injury, the estimate is 3.3677, and the p-value is 0.000931, meaning Non-EVs are more likely to result in minor injuries compared to EVs. Similarly, for No Injury, the estimate is 5.3522 with a highly significant p-value of 9.36e-08, indicating Non-EVs are more likely to result in no injury. The Other injury category also shows significance, with an estimate of 2.5654 and a p-value of 0.0134, indicating a higher likelihood of other injuries in non-EVs. However, for Serious Injury, the p-value is 0.572, indicating no significant effect of vehicle type on serious injuries. Overall, the results suggest that while vehicle type significantly affects minor injuries, no injury, and other injuries, it does not significantly influence serious injuries. Two key factors should be considered when interpreting these results. The first thing is that the sample that was used for the model is based on the estimated number of vehicles per year and that is only 217 vehicles in a total of all accidents happening in Montgomery County between the years 2015 and 2024.

Furthermore, to understand the relationship between accident count and several predictors, factors such as vehicle type (EV, non-EV), surface condition, lighting, weather, injury severity, and days of the week were considered, with Total Registrations of vehicles used as an offset. This number, provided by MVA Maryland, represents the total count of registered EVs and non-EVs, ensuring that the accident count is adjusted relative to the number of vehicles registered. The offset normalizes the data per vehicle, allowing for fair comparisons of accident counts across different vehicle types by accounting for exposure differences due to varying vehicle registration numbers. findings include significant variables with p-values < 0.001. Non-EV vehicles (compared to EVs) have a high positive coefficient (5.34), indicating that non-EVs are associated with a significantly higher number of accidents.

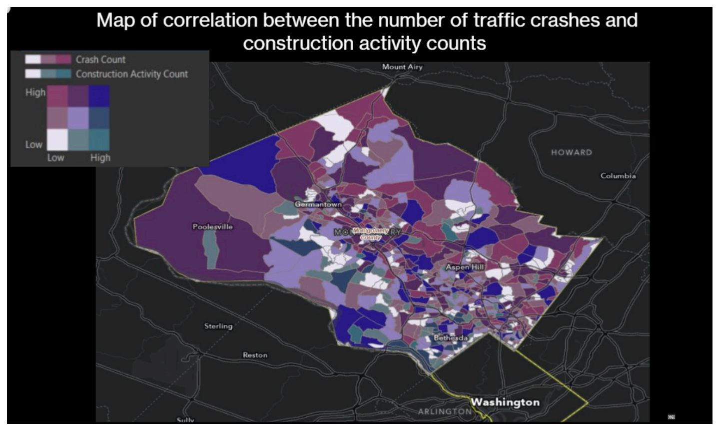
Regarding days of the week, Tue through Sun all show positive coefficients compared to Monday, with Wed and Thu having the highest coefficients, implying a greater likelihood of accidents on those days. Surface conditions like Snow and Ice and Wet conditions have negative coefficients, meaning accidents are less frequent in such conditions. The model also suggests Fatal Injuries and Serious Injuries are associated with lower accident counts, as indicated by the negative coefficients. Regarding Lighting conditions, accidents in daylight have a positive association, while accidents in darkness/no lights, dawn and dusk, and unspecified/other lighting conditions have negative associations. suggesting that accidents in these lighting conditions are less frequent, potentially due to lower traffic volumes.

One fact that should be considered is that while most variables, such as surface condition and injury severity, have been re-leveled (e.g., Dry for surface condition and No Injury as the first level for injury severity), some variables, such as day of the week and lighting conditions, have not been re-leveled. This inconsistency in re-leveling may impact the interpretation of coefficients and comparisons between categories, as the reference level for these variables could influence the results.



The next phase of this project involved analyzing the relationship between accident locations and bike trails to determine whether accidents occur more frequently around trail areas. Another key question explored through this series of maps is whether the bikeways in Montgomery County are effective in reducing accidents involving bicycles and vehicles. The following map shows a density analysis of accidents along bike trails in ArcGIS Pro, identifying "hot spots." This map was created using population density data from the census and crash report geo-locations along bike trails, which are represented as points. The map reveals that higher accident occurrences are found in more populated areas (depicted with darker colors for more densely populated areas and lighter colors for less populated areas). This result suggests that if any changes or improvements are considered for bikeway infrastructure and location, these should primarily focus on areas with higher population densities. The map illustrates that these areas are generally located closer to Washington, D.C.

The second map created for this project is a bivariate map, where the colors represent the correlation between accident counts in census blocks and the number of construction areas. The analysis shows that dark blue areas indicate a high number of accidents associated with significant construction activity, while white and light blue areas represent locations with low correlation. This map helps identify areas that may require more attention to improve driver safety.



References and acknowledgments

MDOT SHA Annual Average Daily Traffic (AADT). (n.d.). https://data.imap.maryland.gov/maps/77010abe7558425997b4fcdab02e2b64/about

MDOT/MVA Electric and Plug-in Hybrid Vehicle Registrations by County as of each month end from July 2020 to September 2024 | Open Data | opendata.maryland.gov. (2024, October 18). https://opendata.maryland.gov/Transportation/MDOT-MVA-Electric-and-Plug-in-Hybrid-Vehicle-Regis/qtcv-n3tc/about_data

MVA Vehicle Registrations by County as of Each Month End from January 2023 to September 2024 | Open Data | opendata.maryland.gov. (2024, October 24). https://opendata.maryland.gov/Transportation/MVA-Vehicle-Registrations-by-County-as-of-Each-Mon/db8v-9ewn/about_data

Roadway safety information Analysis - Safety | Federal Highway Administration. (n.d.). https://safety.fhwa.dot.gov/local_rural/training/fhwasa1210/s3.cfm