**Traffic Violations Analysis in Montgomery County**

**TABLE OF CONTENTS**

CHAPTER 1 – INTRODUCTION ....................................................................................2-4  
CHAPTER 2 – LITERATURE REVIEW .........................................................................5-9  
CHAPTER 3 – METHODOLOGY: DATA PREPARATION .........................................10-11  
CHAPTER 4 – METHODOLOGY: EXPLORATORY DATA ANALYSIS ...................12-14  
CHAPTER 5 – METHODOLOGY: DISCUSSION .........................................................15-33  
CHAPTER 7 – CONCLUSION ........................................................................................34   
REFERENCES ..................................................................................................................35-36

# **Introduction**

Montgomery County, in its mix of urban and suburban environments, has a very broad set of traffic patterns for the daily lives of its citizens. In addition to interfering with the smooth flow of traffic, traffic violations, from speeding and DUI to other offenses, contribute significantly to the volume of traffic accidents and public safety issues. Despite ongoing county police and traffic enforcement efforts, a more profound comprehension of these trends in violation remains unknown. By actively investigating traffic violations in Montgomery County via in-depth analysis of a recent Data.gov dataset, this study strives to bridge that gap. Augmented with significant information like time stamps, location coordinates, and driver demographics, the study analyzes over 2000127 traffic stop reports.

Our work aims to give insight into how and why traffic violations take place by studying their relationship with accidents and underlying causes. We also analyze trends in traffic violation populations and build predictive models for forecasting the probability of arrests following traffic stops. Aiming to inform public policy and law enforcement strategy and to shed light on traffic behavior dynamics in the county, this multivariate study not only sheds light on traffic behavior dynamics in the county but also works to promote safer roads and more efficient traffic management by informing them.

**Data Set**

The dataset includes more than two million traffic stops that were conducted in the county. Traffic Violations Dataset is a big dataset. Spanning January 1, 2012, to December 31, 2024, it offers a rich historical view of traffic enforcement activity and trends over more than a decade. Researchers, policymakers, public safety practitioners, and many others will have a lot of use for this dataset because it contains a large amount of data and a variety of characteristics to report.

For each traffic stop, one important piece of information is documented in this data set, including the place and time of the stop, which law enforcement agencies participated in, and the type of offense that was documented. With search-related findings to those conducted in the course of the stop, it also includes relevant background data on the incident such as whether there was any injury or accident involved.

By maintaining more than 40 distinct characteristics, the dataset offers thorough examination of numerous aspects of traffic violations. Some of the important columns are

|  |  |
| --- | --- |
| Column | Description |
| SeqID | Unique sequential identifier for each stop record. |
| Date Of Stop | Calendar date on which the stop occurred (e.g., YYYY-MM-DD). |
| Time Of Stop | Clock time when the stop was initiated (e.g., HH:MM: SS). |
| Agency | Primary law‑enforcement agency conducting the stop. |
| SubAgency | Specific department or division within the agency (e.g., Traffic Division). |
| Description | Text description of the stop reason (e.g., “Speeding,” “Broken Taillight”). |
| Location | Street address or intersection where the stop took place. |
| Charge | Formal offense code or summary charged against the individual (e.g., “Cited for speeding”). |
| Race | Self‑reported or observed racial/ethnic group of the driver. |
| Gender | Drivers reported or recorded gender (e.g., “Male,” “Female,” “Non‑Binary”). |

**Table 1: Description of columns**

# **Literature Review**

In Montgomery County specifically, the work on traffic violations and enforcement practices offers a complete understanding of the ways in which driver behavior, environmental conditions, and demographic traits combine to create road safety issues. From traffic violation patterns to the practices that law enforcement agencies use to manage these offenses, scholars have investigated a wide range of subjects and created predictive models that allow for forecasting of the results of traffic stops. By recognizing high-risk behavior and gaining insight into the factors at play in accidents and negative enforcement outcomes, these studies strive to encourage road safety and police strategy optimization.

With significant consequences on public safety, traffic offenses contribute heavily to road accidents worldwide. The World Health Organization estimates that traffic accidents result in over 1.35 million deaths per year, a figure that captures the worldwide relevance of combating traffic offenses (Saville et al., 2024). Speeding, drunk driving, and overloading are typical behaviors that are quite often linked with this high volume of deaths. A recent Indian study found that excessive speeding was the single most significant contributor to driving errors, which resulted in accidents in 66.5% of the cases studied (Sivasankaran & Balasubramanian, 2021). The study also pointed out that specific road conditions, i.e., single lanes or intersections without controls, and the driver type—male or unlicensed—had the potential to greatly heighten the probability of speeding-related accidents. Particularly during the daytime when traffic flow is heavy, speeding was a major problem even on roads that had safety measures like central dividers.

Apart from speeding, a combination of several factors has the ability to increase the risk and severity of road accidents significantly. Research has indicated, for instance, that when drivers are speeding and cars are carrying passengers or freight, the resulting accidents will likely be far more severe. This combination heightens the instant crash risk and the cost and long-term effect of the event (Ayuso et al., 2010). These kinds of consequences indicate that law enforcement agencies must take a broader approach and confront hazardous combinations of crimes that can have more disastrous effects. Focusing on combinations of dangerous activities instead of singular events can help frame policy and enforcement efforts that are aimed at minimizing accident severity.

Traffic safety is also influenced by temporal and environmental factors and by the behavior of drivers. Several research studies have established that the time of day, day of week, and even weather conditions have influence on the magnitude and frequency of traffic offenses. Automated enforcement data, for example, have shown that particular times, i.e., at midday or during weekends, witness traffic offenses with higher frequency. Specific categories of road conditions, such as collector roadways and residential roads, are also expected to see higher rates of violations (Li et al., 2021). Greater rates of traffic infraction are found with other conditions such as low wind speed, high traffic volumes, and reduced temperatures. These factors have the potential to lead to behaviors in drivers which increase their chances of engaging in dangerous actions like speeding or road rage, which in turn increase the likelihood of accidents.

Researchers have progressively looked to predictive analytics to better understand and project the results of traffic stop encounters as huge data sets become accessible. Detailed information from more than 600,000 traffic stops, for instance, in Montgomery County have allowed academics to create machine learning models forecasting whether a stop will lead to an arrest, a fine, or a warning (Aquino et al., 2020). Using these sophisticated analytical tools, one Saville et al. (2024) study assessed the influence of several elements including driver demographics, the kind of infraction, and the conditions surrounding every traffic stop. Their research showed that among the strongest result indicators were the specifics of the violation and the setting in which it happened. Although demographic elements like ethnicity and gender were important as well, the research indicated that models created without these sensitive variables might nonetheless have great prediction accuracy. This implies that, rather than on the personal traits of the driver, one may create more just, race-neutral models emphasizing the kind of infraction and its background. Such models show potential for lowering prejudice in traffic policing and guaranteeing that choices are based on objective standards.

Remaining controversial in the discourse of traffic stops and enforcement, demographic characteristics are an area of contention. Multiple studies both from the Stanford Open Policing Project and other entities—have concluded that Black drivers are stopped more often than their White or Hispanic counterparts and are more likely to be searched during the stop (Saville et al., 2024). This proof of racial discrimination has questioned the fairness of traffic stops policies. Adding race as a variable in prediction models has proven to add fit to predicting stop outcomes in Montgomery County. However, the performance of models eliminating race and gender indicates the existence of alternative variables—such as offense seriousness and contextual variables like location—that can be stable outcome predictors. These results are significant because they show that, in traffic stop decisions, law enforcement agencies can make decisions with less bias by focusing on more neutral, objective criteria. Similarly consistent with the contention that the type of offense itself is a valid predictor of enforcement results, one study found that offenses, like drunk driving, will consistently lead to high arrest rates regardless of the demographic profile of the driver (Rababah et al., 2022).

The study of traffic infractions also shows obvious trends in how time and environment influence the incidence of violations as well as the results of traffic stops. Data from Montgomery County, for example, reveals that the outcomes of traffic stops are not only determined by the kind of infraction but also by the hour of the day, the road type, and weather conditions. Building prediction models that enable law enforcement organizations to allocate resources more efficiently and handle possible biases in their operations has been made easier by this extensive data set (Aquino et al., 2020). Police forces can focus their efforts more effectively by knowing when and where infractions are most likely to happen, hence lowering the total number of accidents and enhancing public safety. Studies in this field emphasize that a mix of driver conduct, ambient conditions, and the timing of enforcement activities shapes the hazards connected to traffic offenses, which are not random.

When one considers the findings of several different research, one can see that the body of knowledge on traffic infractions and enforcement is thorough and wide. The data highlights three important topics: the need of driver behavior to affect crash risk, the possibility of predictive analytics to anticipate traffic stop results, and the continuing problem of demographic inequalities in enforcement policies. Research has regularly demonstrated that significant causes of accidents are high-risk behaviors like speeding and driving under the influence, which can be aggravated by overweight cars or difficult environmental conditions. Simultaneously, sophisticated data analysis methods are being applied to create models forecasting the results of traffic stops with great accuracy, so arming law enforcement with tools that could enable more equitable and effective running of their activities.

Considering the thorough data from prior studies, it is obvious that a data-driven strategy can greatly improve the knowledge of traffic enforcement dynamics. Using a sizable database of traffic infractions, researchers may create models that not only forecast results but also point out the most important elements supporting these results. Law enforcement agencies using this strategy may make more educated decisions, hence enabling them to use resources more effectively, cut pointless stops, and finally enhance road safety for all. Research indicates that more efficient enforcement tactics might result from an emphasis on high-risk behaviors and consideration of environmental and temporal elements. Furthermore, building models that reduce the impact of perhaps prejudiced demographic variables helps to produce a fairer system of traffic enforcement.

In summary, traffic offense literature and enforcement patterns provide valuable information on the various factors that affect road safety concerns. The study outlined demonstrates that a combination of driver behavior, environmental conditions, and enforcement tactics influences the risk of accidents, and the effectiveness of traffic stops. The use of predictive analytics has emerged as an extremely promising technique in this field, offering the ability to predict enforcement outcomes with great accuracy and highlighting where there may be bias. As Montgomery County continues to be a pivotal case study, the findings of these studies will be crucial to informing future efforts to enhance traffic enforcement and public safety. The study gives the grounds on which to develop evidence-based interventions that would maximize law enforcement practice, reduce the chances of accidents, and formulate a fairer and better traffic management system.

# **Methodology**

We conducted our data manipulation and analysis using Jupyter Notebook version 7.0.1 with Python 3.11.4. For visualizations and data management, we employed matplotlib and pandas, respectively. In addition, we made use of libraries such as NumPy for numerical computations and scikit-learn for performing Logistic Regression analysis.

**Data Cleaning**

|  |  |  |
| --- | --- | --- |
| Field | Missing Values | Percentage (%) |
| Search Arrest Reason | 1,939,420 | 96.964843 |
| Search Type | 1,912,139 | 95.600879 |
| Search Disposition | 1,912,131 | 95.600479 |
| Search Reason | 1,912,131 | 95.600479 |
| Search Outcome | 779,609 | 38.977975 |
| Search Reason for Stop | 760,621 | 38.028635 |
| Search Conducted | 760,318 | 38.013486 |
| Article | 90,061 | 4.502764 |
| Color | 22,037 | 1.101780 |
| Year | 10,528 | 0.526367 |
| DL State | 929 | 0.046447 |
| Driver City | 510 | 0.025498 |
| Model | 221 | 0.011049 |
| Make | 73 | 0.003650 |
| State | 59 | 0.002950 |
| Driver State | 11 | 0.000550 |
| Description | 10 | 0.000500 |
| Location | 4 | 0.000200 |

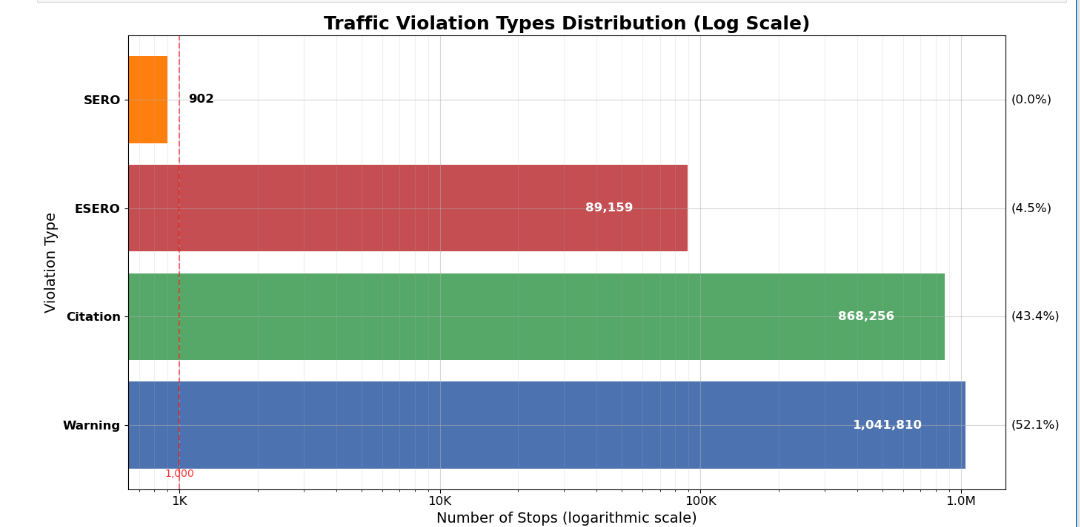
**Missing data Imputation Strategies:**

We cleaned the dataset and processed it for future analysis. We converted date and time columns to new features like Hour, Day\_of\_Week, Month, and Year\_of\_Stop. We replaced missing values with the most frequent value for fields like Color, Make, Model, State, Driver State, Description, and Location. If DL State was missing, we replaced it with the Driver State. For the geolocation information, we set Driver City to "UNKNOWN" by default to avoid errors. We first imputed missing Vehicle Year values on a Make and Model level using the median, and in some cases, overall median. Finally, we handled legal and search fields by marking missing values as "MISSING" or "NOT\_APPLICABLE" for modelling purposes.

|  |  |  |
| --- | --- | --- |
| Statistic | Latitude | Longitude |
| Count | 2,000,127 | 2,000,127 |
| Mean | 36.23418 | 71.49170 |
| Std | 10.16329 | 20.05277 |
| Min | 0.00000 | 151.25600 |
| 25% | 39.01662 | –77.19223 |
| 50% | 39.06590 | –77.08662 |
| 75% | 39.13459 | –77.02694 |
| Max | 41.54316 | 39.06444 |

The dataset summary explains detailed records for various aspects such as geographic coordinates, temporal information, and encoded categorical data. It shows that there are approximately 2 million entries, with records spanning from 2023 to 2025, thereby capturing over a decade of traffic stop data. The geographic information, with an average latitude of about 36.2° and longitude around -71.5°, points to a specific regional coverage.

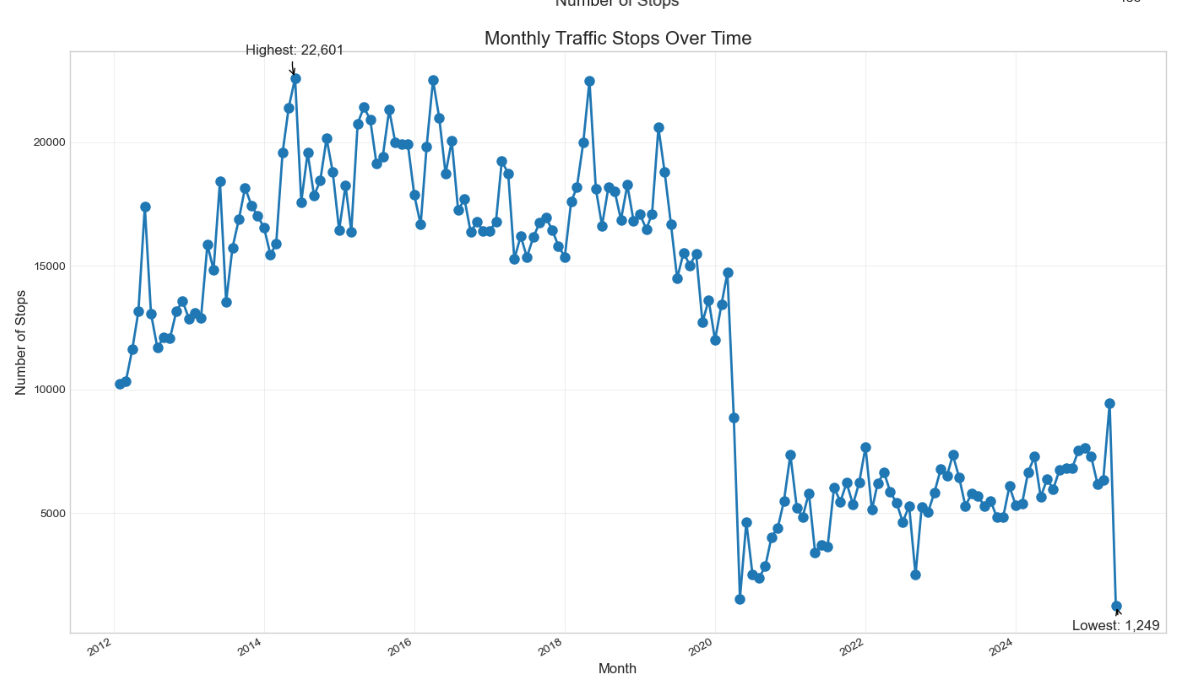
**Exploratory Data Analysis (EDA)**

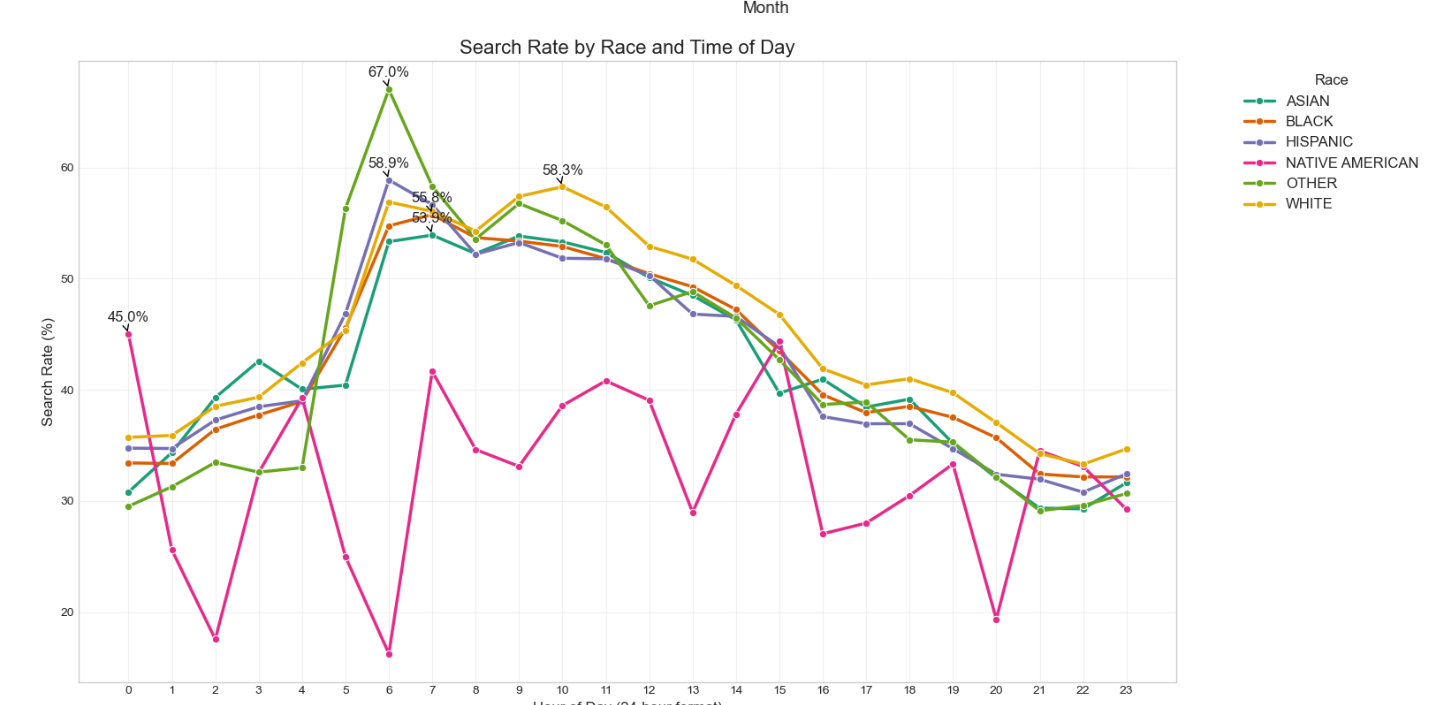


The chart illustrates how different types of traffic violations are distributed across all recorded stops. Warnings dominate the dataset, representing the majority of officer actions, while formal citations constitute the next largest portion. Technical equipment‑related offenses (labeled ESERO) appear far less frequently, and basic seat‑belt or equipment violations (SERO) are exceptionally rare. Displayed on a logarithmic axis, the plot makes it clear that there is an enormous disparity between the very common outcomes (warnings and citations) and the much scarcer regulatory offenses.

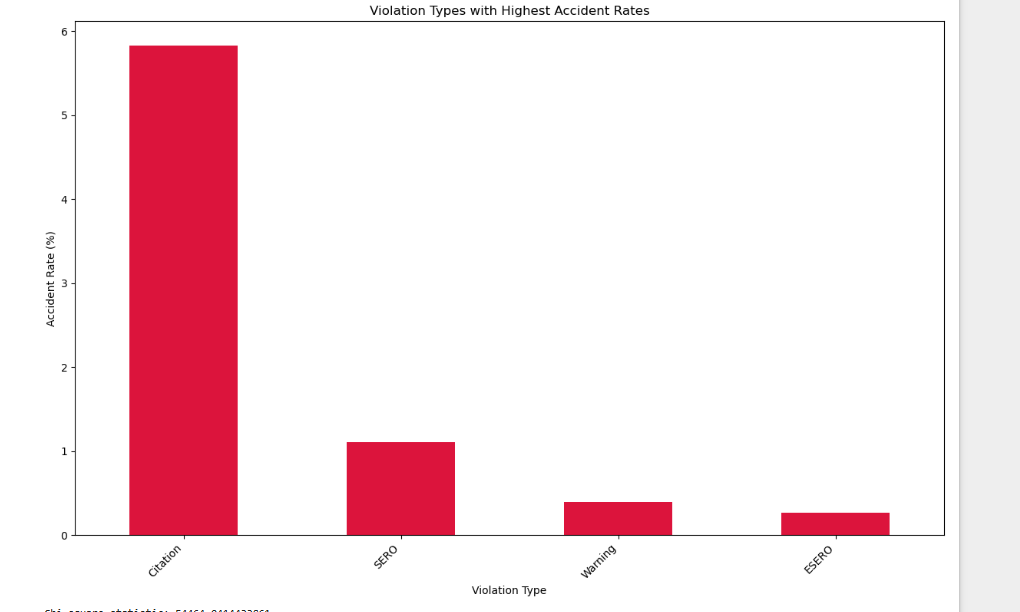
We have two names SERO and ESERO which are mentioned below,

* **SERO (Safety Equipment Repair Order):** A non‑criminal notice requiring the driver to fix faulty safety equipment (e.g., brakes, lights, tires) before the vehicle can legally operate, without a fine.
* **ESERO (Equipment‑Safety & Regulatory Offense):** A regulatory citation for more specialized compliance failures (e.g., overweight loads, missing haz‑mat placards), typically carrying fines or formal penalties.

The second graph is a line chart that tracks monthly traffic stops over a period. In this plot, the horizontal axis represents time segmented by month and year, and the vertical axis shows the number of stops. The line’s peaks indicate months with a high number of stops, while its troughs point to periods with fewer stops. This trend analysis helps in understanding how traffic enforcement or driving behavior varies throughout different times of the year.

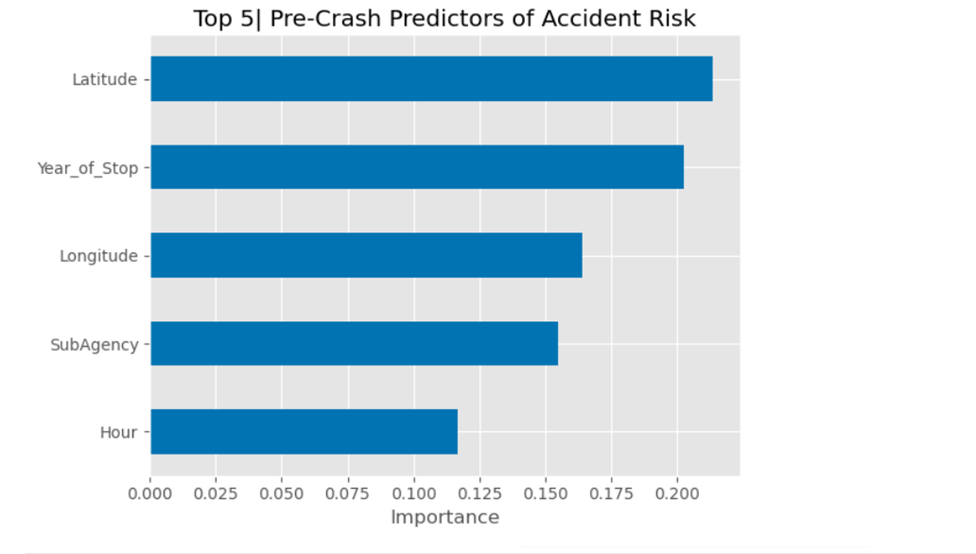
The third graph is another line chart that examines the search rate by race across different times of the day. Here, each colored line corresponds to a different racial group, with the x-axis showing the 24-hour clock and the y-axis depicting the percentage of stops that resulted in a search. The graph illustrates fluctuations where certain groups experience higher or lower search rates at specific times. This detailed view underscores how search practices can vary by both race and time, suggesting deeper patterns in enforcement strategies.

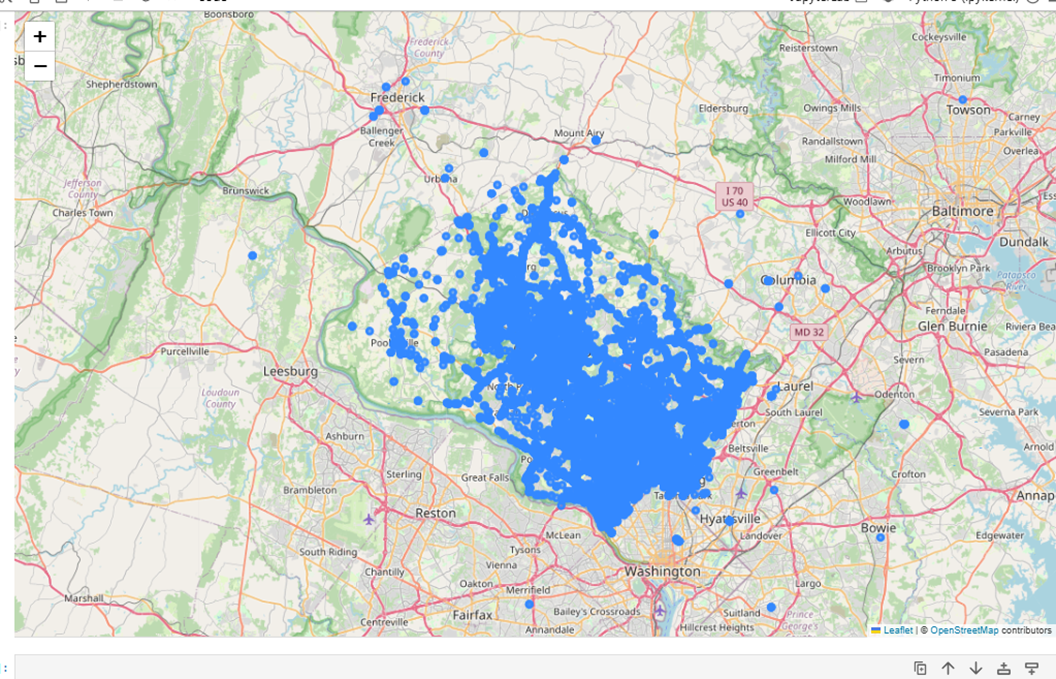
**Discussion**

1) What are the key factors contributing to traffic accidents in Montgomery County?  
  


This graph shows how various kinds of traffic infractions relate to accident rates. It shows how every violation kind such as Citation, SERO, Warning, and ESERO relates to accident rates. Especially, infractions that lead to citations seem to have the highest accident rate, implying that the severity or type of these infractions could be related to more hazardous driving conditions. This graphic clarifies which infraction categories are important accident occurrence forecasters.

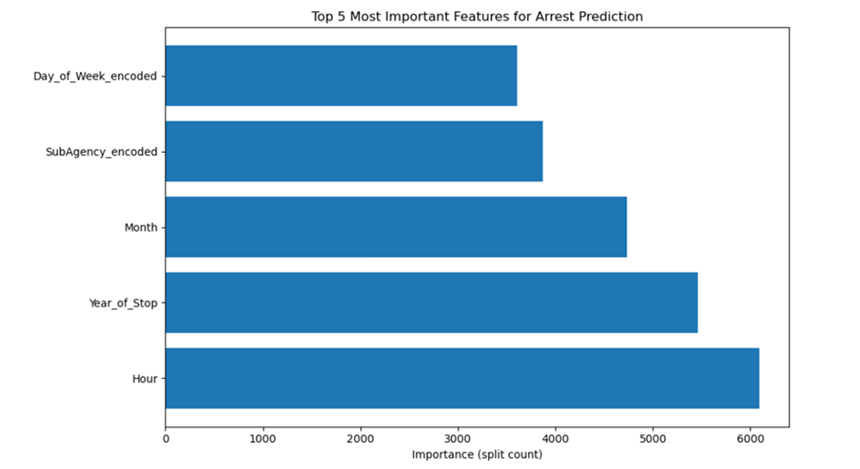
**Random Forest Modelling**

We trained a Random Forest classifier using two hundred trees of maximum depth six and balanced class weights to assess accident risk. The model reveals that geographic location is the most influential factor in determining whether a stop leads to an accident, followed by the year of the stop, which captures shifts in traffic conditions, regulations, and reporting practices over time. The identity of the enforcing agency and the hour of the stop also play significant roles, reflecting procedural differences between agencies and the impact of peak versus off‑peak traffic. In contrast, binary indicators for alcohol involvement and presence of a work zone contribute relatively little predictive power compared with the spatial and categorical features.

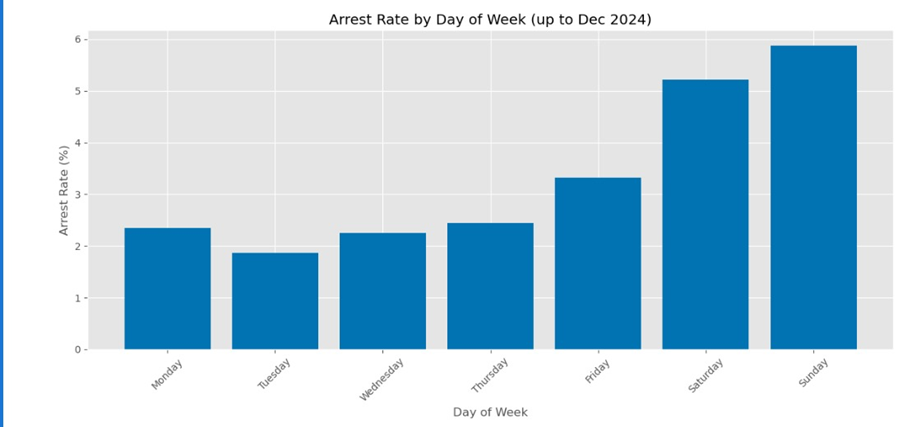
 **Fig: Folium plot of Accidents**

The folium map reveals clear spatial variation in accident occurrences across Montgomery City, with some neighborhoods exhibiting markedly higher concentrations of incidents than others. In particular, the southern sector stands out as a hotspot where collisions are notably more frequent, suggesting that factors such as roadway design, traffic volume, or local driving behaviors may elevate risk in that area. These clustering patterns not only highlight where enforcement efforts and safety interventions could be prioritized but also point to underlying urban features like road geometry or land use that might predispose certain corridors to greater accident exposure. By focusing on these high‑density zones, city planners and traffic safety authorities can develop targeted strategies to reduce collisions, such as redesigning intersections, improving signage, or increasing patrol presence during peak hours.

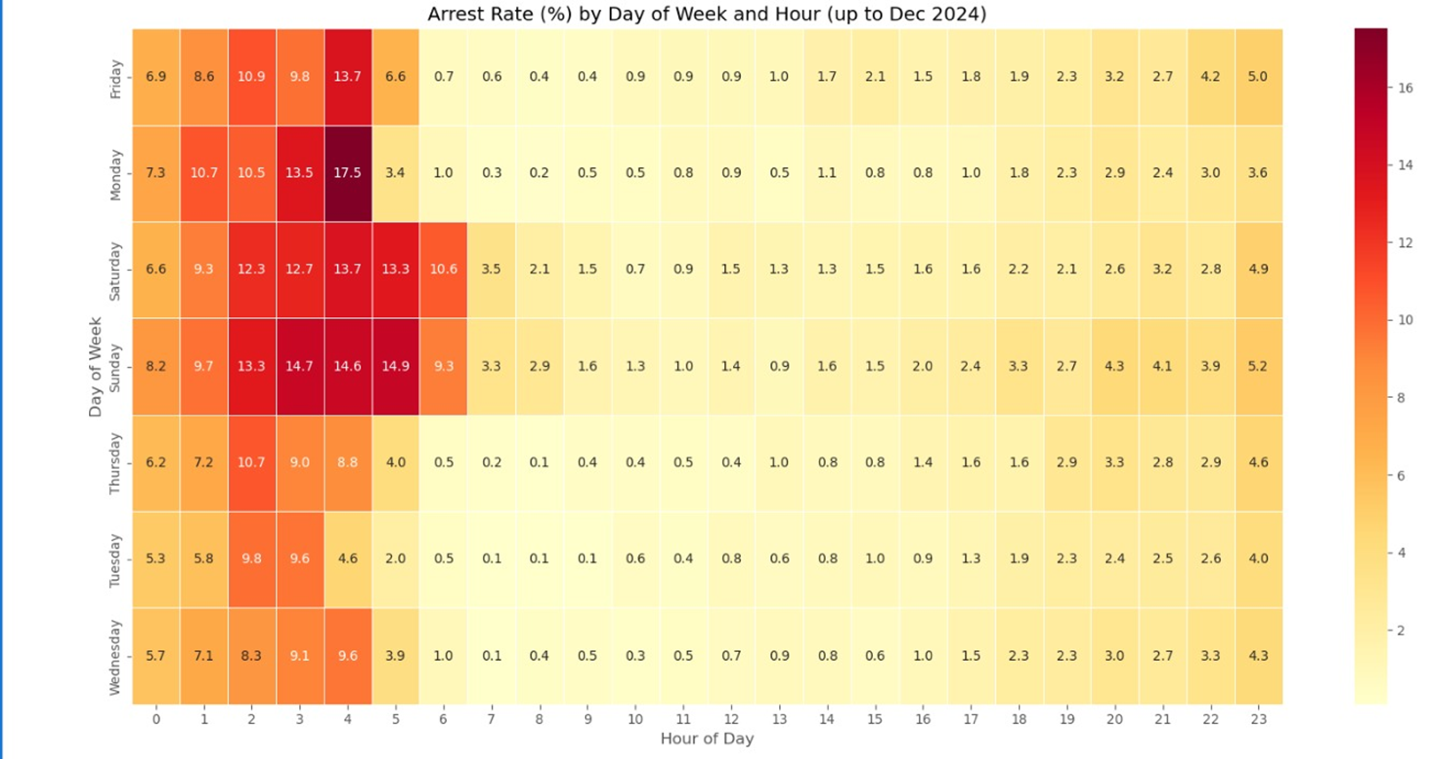
2. Can we predict the likelihood of an arrest based on traffic stop data?  
Top 5 violation types in arrests:



Using a Light GBM classifier on our traffic‐stop dataset, we find that temporal dynamics overwhelmingly drive arrest outcomes: the hour of the stop emerges as the single strongest predictor of whether an officer will make an arrest, closely followed by the calendar year and month of the stop. This indicates that nightly patrol patterns, evolving enforcement policies over time, and seasonal shifts in policing intensity all substantially influence arrest rates. Beyond these temporal factors, the identity of the specific sub‑agency handling the stop—and, to a slightly lesser extent, the day of the week—also contributes meaningfully to predictive performance, reflecting how departmental procedures and weekly rhythms modulate the likelihood that a given stop leads to an arrest.



The chart reveals a clear weekly rhythm in arrest likelihood, with the highest rates occurring over the weekend especially on Sunday implying that either officer activity intensifies or driver behavior changes during those days. In contrast, the workweek sees consistently lower arrest rates, with Tuesday marking the nadir of the week. This pattern suggests that both traffic enforcement strategies and public driving habits vary sharply between weekdays and weekends.

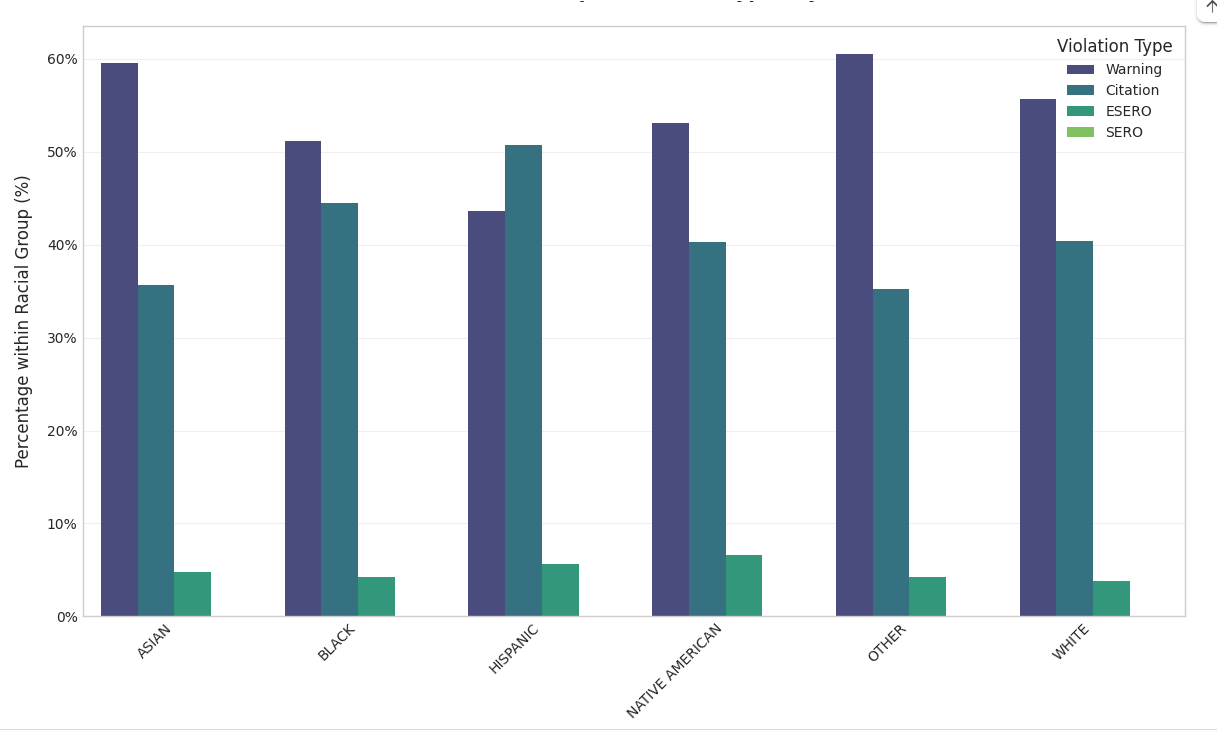


The weekday‐hour heatmap uncovers a pronounced surge in arrests during the late‐night and early‐morning hours, with the highest concentrations occurring between around 1 AM and 5 AM on weekends—Sunday and Saturday nights stand out. This late‐night peak likely reflects a combination of increased alcohol‐related incidents, reduced traffic volume that makes enforcement more visible, and targeted patrols during traditionally high‑risk periods. In contrast, arrest rates remain relatively low throughout the daytime and early evening on all days, suggesting that both driving behavior and policing priorities shift dramatically once the clock passes into the early‐morning window. Even during weekdays, a modest uptick glimpsed in the small hours points to a consistent pattern of heightened enforcement or incidents when bars close and nightlife activity winds down.

3. How do traffic violations vary across different demographic groups in Montgomery County?

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Race | Warning | Citation | ESERO | SERO |
| ASIAN | 59.56 | 35.66 | 4.75 | 0.03 |
| BLACK | 51.22 | 44.47 | 4.28 | 0.04 |
| HISPANIC | 43.63 | 50.70 | 5.61 | 0.06 |
| NATIVE AMERICAN | 53.10 | 40.34 | 6.55 | 0.00 |
| OTHER | 60.55 | 35.21 | 4.20 | 0.03 |
| WHITE | 55.66 | 40.46 | 3.85 | 0.04 |

The breakdown of outcomes across racial groups reveals that some communities are more likely to receive warnings while others see a higher proportion of formal citations. Asian and “Other” drivers tend to receive warnings most often, whereas Hispanic motorists stand out as the group with the largest share of citations. Native Americans experience a noticeably higher rate of equipment‑related stops (ESERO) than other groups, even though such regulatory violations remain relatively rare overall. In every racial category, pure seat‑belt infractions (SERO) are almost non‑existent, indicating that these violations make up a vanishingly small slice of all enforcement actions.

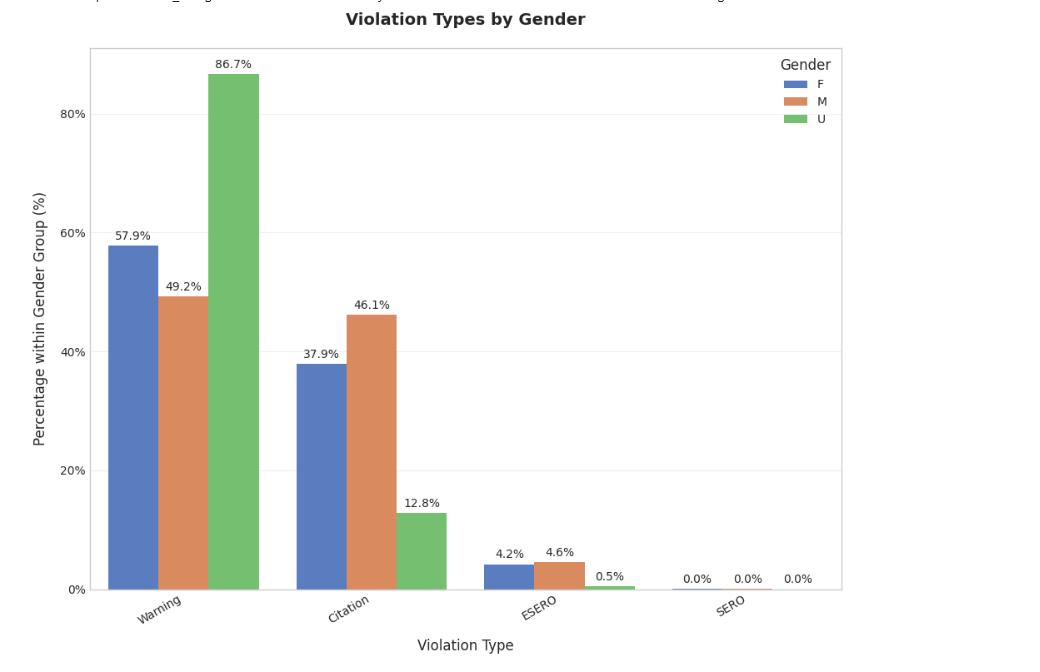


**Fig: Violation by Race type**

The percentage distribution of the four violation kinds Warning, Citation, ESERO, and SER across various racial groupings is shown in this graph. It emphasizes how every racial category ASIAN, BLACK, HISPANIC, NATIVE AMERICAN, OTHER, and WHITE experiences a particular combination of infractions, suggesting that the kind of violation given changes greatly by race. The bar chart reveals that drivers classified as Asian or “Other” are overwhelmingly issued warnings more often than any other outcome, with fully around three‑fifths of their stops resulting in a caution rather than a formal citation or equipment charge.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Gender | Warning | Citation | ESERO | SERO |
| F | 57.85 | 37.92 | 4.19 | 0.04 |
| M | 49.22 | 46.13 | 4.60 | 0.05 |
| U | 86.72 | 12.82 | 0.46 | 0.00 |

Across gender groups in Montgomery County, we observe clear differences in how enforcement actions are meted out. Female drivers receive warnings most frequently (57.9 %), with citations accounting for 37.9 % of their stops; ESERO violations make up 4.2 %, and SERO violations are virtually nonexistent (0.04 %). Male drivers, by contrast, are cited more often (46.1 %) than warned (49.2 %), with slightly higher ESERO rates at 4.6 % and similarly minimal SERO rates (0.05 %). Stops recorded with an unknown or unreported gender (“U”) are overwhelmingly resolved with warnings (86.7 %), while citations drop to just 12.8 % and ESERO/SERO violations are almost absent (0.46 % and 0 %, respectively). These patterns suggest that female drivers are the most likely to receive a warning rather than a citation, male drivers see a more balanced split between warnings and citations, and cases lacking gender data are nearly always issued warnings.



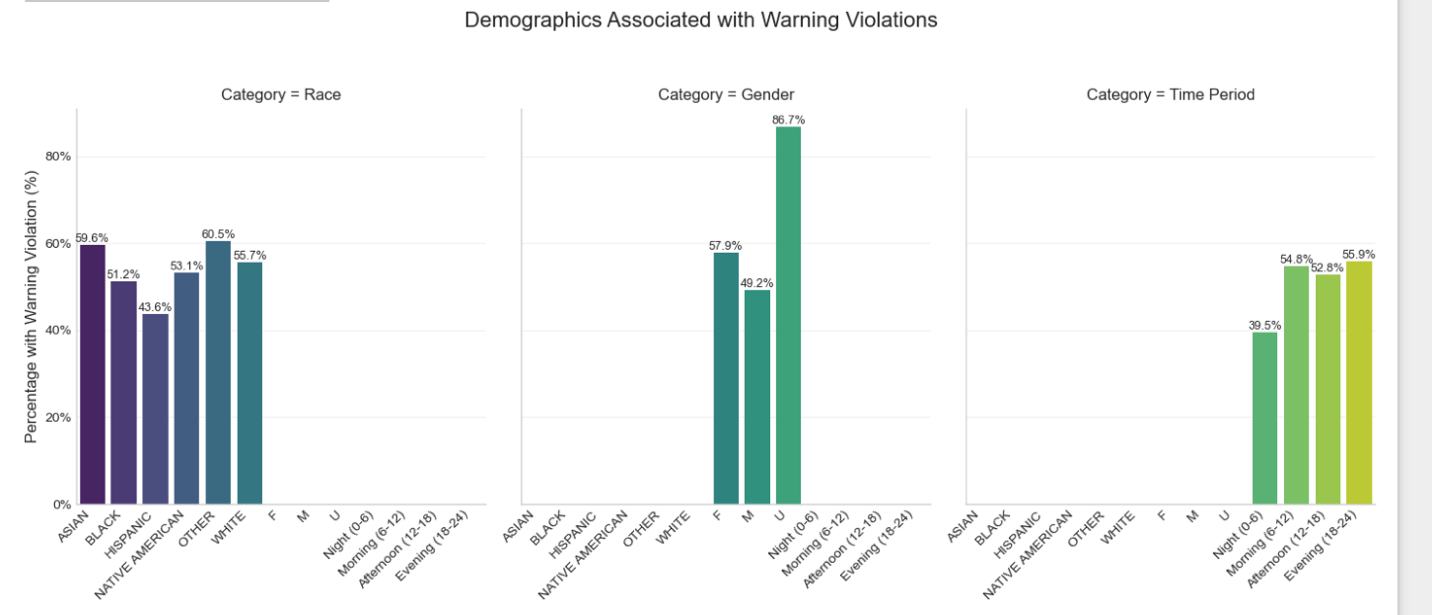
This graphic depicts the distribution of violation kinds by gender: Female, Male, and Unknown. With clear variations like a greater percentage of Warning infractions among women compared to men, it shows that women, men, and those with uncertain gender all have unique patterns in the kinds of violations they experience.

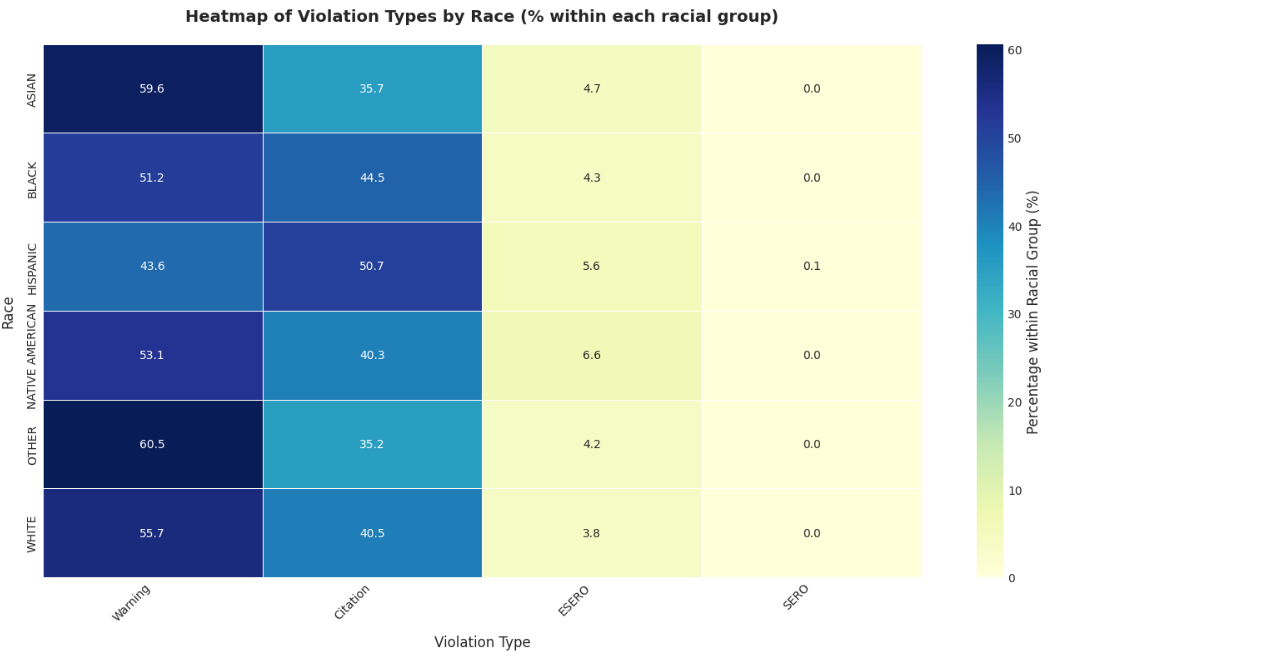
|  |  |  |
| --- | --- | --- |
| Race | 0 | 1 |
| ASIAN | 45598 | 67150 |
| BLACK | 309462 | 324884 |
| HISPANIC | 252402 | 195333 |
| NATIVE AMERICAN | 1632 | 1848 |
| OTHER | 51135 | 78480 |
| WHITE | 298088 | 374115 |

This table breaks down stop outcomes by racial group, where “0” denotes stops that did not result in an arrest and “1” denotes stops that did. Among Asian drivers, more stops led to an arrest than not, with roughly sixty‑seven thousand resulting in an arrest versus forty‑six thousand without. Black motorists also saw slightly more arrests than non‑arrests, though their counts are much larger overall—about 325 000 arrests versus 309 000 non‑arrests. In contrast, Hispanic drivers experienced more stops without an arrest (approximately 252 000) than with one (195 000). Native Americans show the smallest numbers but still a modest excess of arrests over non‑arrests. Drivers classified as Other and White each have more arrest outcomes than non‑arrests, with White drivers seeing the highest absolute counts (about 374 000 arrests and 299 000 non‑arrests). Overall, White and Black groups account for the largest volumes of both arrest and non‑arrest stops, while Native Americans represent the smallest.

Percentage of each racial group receiving this violation type:

|  |  |
| --- | --- |
| Race | Warning |
| ASIAN | 59.6% |
| BLACK | 51.2% |
| HISPANIC | 43.6% |
| NATIVE AMERICAN | 53.1% |
| OTHER | 60.5% |
| WHITE | 55.7% |

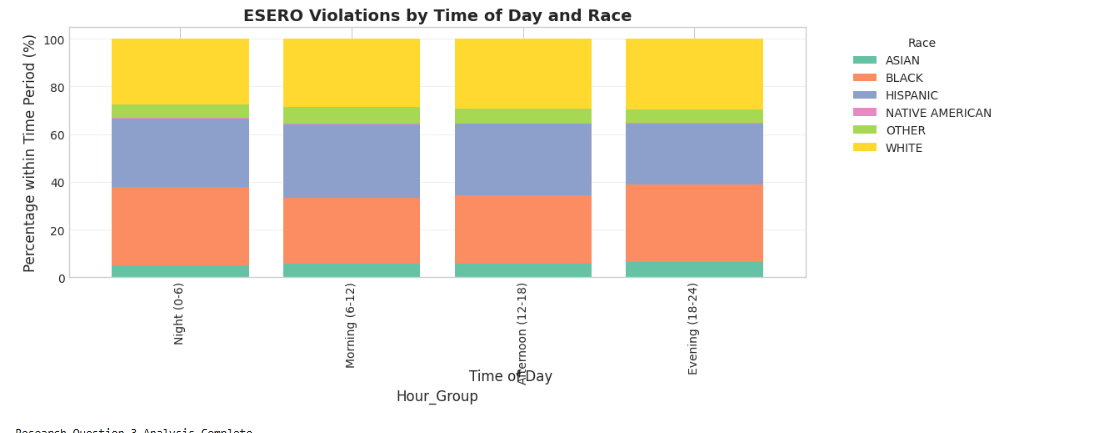
By gender, this graph categorizes the incidence of Warning infractions. While the unknown gender group shows an even more noticeable rate, it shows that female drivers have a larger percentage of Warning infractions than male ones. This suggests a notable gender-based difference in the distribution of Warning breaches.



Concentrating on the most prevalent violation kind warning, this graph shows the proportion of Warning infractions in every racial group. The graph emphasizes an obvious link between race and the probability of getting a Warning by showing that certain racial groupings, such as ASIAN and OTHER, have a higher rate of Warning infractions than others like HISPANIC.

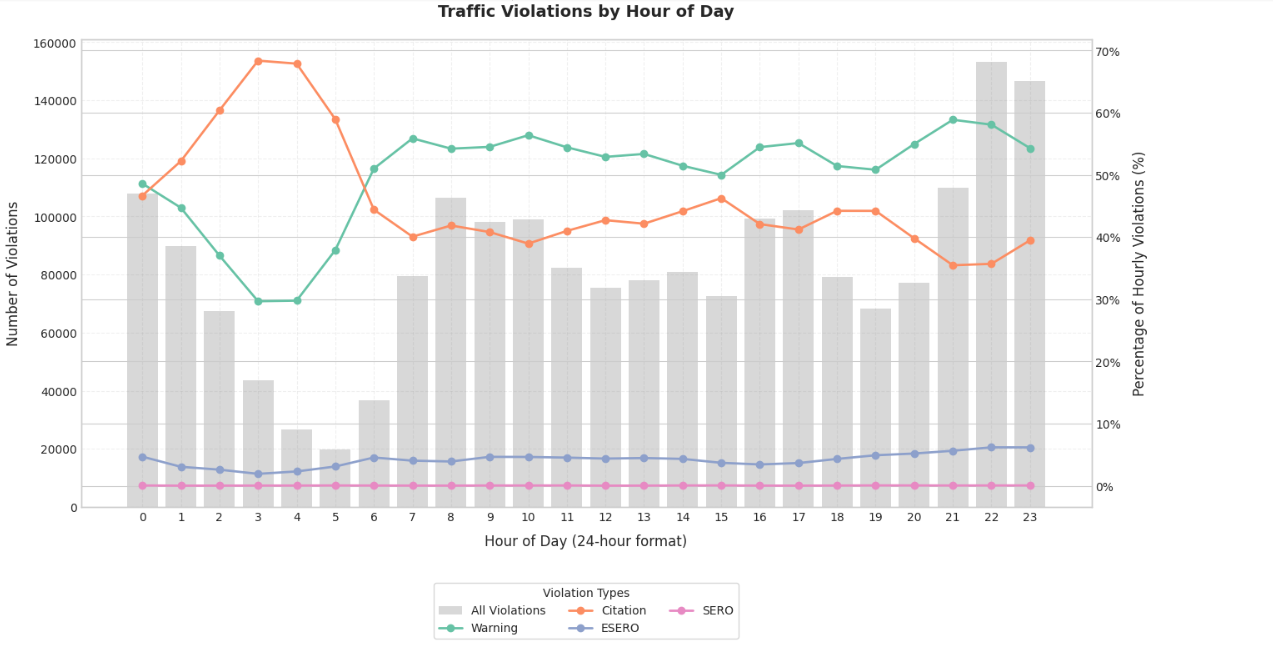


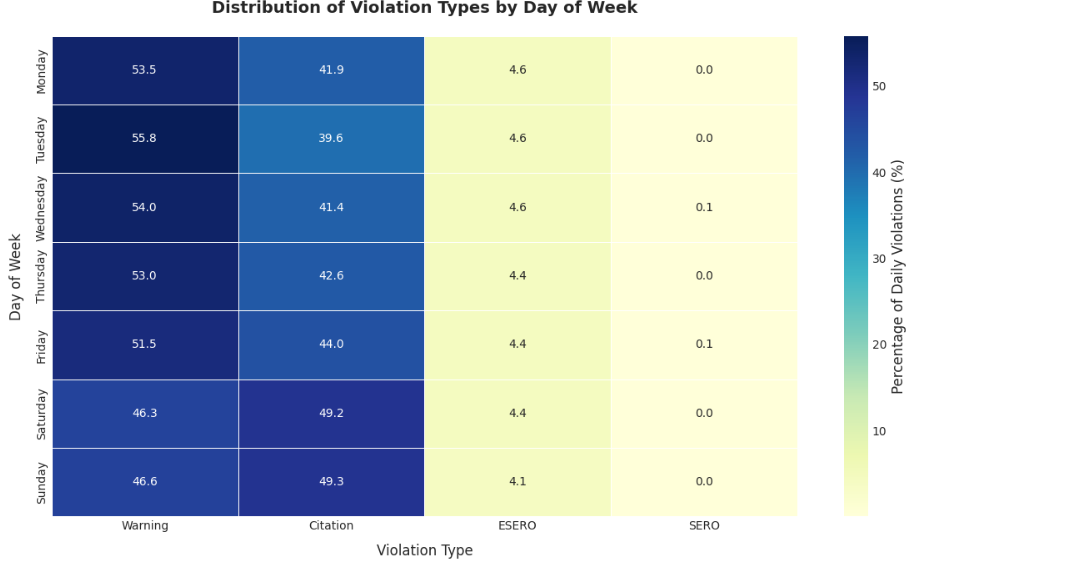
These graphs show stacked bar charts showing how Warning and Citation infractions are spread across several racial groups during different times of day—Night (0–6), Morning (6–12), Afternoon (12–18), and Evening (18–24). Each bar shows the total infractions for a specified time frame; color-coded sections show the percentage of each racial group. Collectively, they show that the racial makeup of drivers with Warning and Citation infractions differs by violation category and time of day, hence stressing possible temporal and demographic effects on traffic violation trends in Montgomery County.

This graph investigates the correlation between the time of day and the frequency of Warning breaches by classifying stops into Night (0-6), Morning (6-12), Afternoon (12-18), and Evening (18-24). It indicates that Warning violations are less frequent at night and peak in the morning and evening, so stressing a notable correlation between time and the probability of getting a Warning violation.

All things considered, the studies offer strong proof that traffic infractions in Montgomery County range greatly among several demographic groups. Violation type distribution is not consistent: racial disparities are clear in the greater likelihood for some groups to get either Warnings or Citations, while gender-based differences reveal women tend to get more Warnings compared to men. Moreover, the study of Warning infractions emphasizes the importance of demographic variables as well as the hour of the day in determining the probability of getting a Warning. The statistical tests show that these differences are very important, implying that demographic and temporal elements are essential in forming the patterns of traffic infractions in Montgomery County.

4. What is the relationship between time, date and type of traffic violations issued?

Based on the above graph there has been a pronounced shift in enforcement patterns, with warning violations rising steadily from roughly 43 percent to nearly 67 percent and citation violations declining sharply from about 56 percent down to 29 percent. Throughout this period, ESERO (Equipment, Safety, and Regulated Offense) violations have held relatively constant, fluctuating modestly between six and eight percent, while SERO (Serious Equipment and Regulated Offense) violations have remained less compared to others,

With Warning, Citation, and ESERO violations happening only on weekdays, the graph comparing weekdays and weekends shows notable variations (\u03c7\u00b2 = 9096.43, p < 0.0001) in the issuance of infractions. This suggests that the kinds of violations given depend much on the day of the week.

The trend analysis graph indicates that from 2012 to 2025, Warning violations rose significantly (from 42.7% to 67.0%), whilst Citations fell (from 56.1% to 29.1%), with slight variations seen in ESERO and SERO. This change over time indicates changing driving behavior trends or traffic enforcement techniques.

|  |  |  |  |
| --- | --- | --- | --- |
| Violation\_Type | 2012 | 2025 | Change |
| Warning | 42.7% | 67.0% | +24.3% |
| Citation | 56.1% | 29.1% | -27.0% |
| ESERO | 0.6% | 3.9% | +3.2% |
| SERO | 0.6% | 0.0% | -0.6% |

The hypothesis suggested a major link between the kind of traffic infraction given and temporal variables like hour of day, day of week, month, and changes over several years. High chi-square values (e.g., 43448.56 for hour vs. violation type, 9096.43 for day vs. violation type, and 1910.04 for month vs. violation type) combined with very low p-values (p < 0.0001) from chi-square tests run for each temporal dimension produced. These findings demonstrate that the distribution of traffic violation types differs greatly with time and date, hence strongly rejecting the null hypothesis of independence.

**Conclusion**

Our analysis reveals that certain geographic areas consistently experience higher rates of traffic incidents, suggesting that local road characteristics and traffic patterns play a crucial role in accident risk. Temporal factors such as the year and time of day—also shape both collision likelihood and arrest outcomes, reflecting how shifts in enforcement focus and traffic conditions influence when and where stops lead to more serious consequences. More severe infractions tend to signal greater underlying risk, while procedural differences between agencies and the timing of patrols further modulate these outcomes.

Patterns of enforcement vary across demographic groups, with some drivers more often receiving warnings and others more frequently facing citations or specialized equipment notices. Gender differences likewise emerge in how stops are resolved. Across weekdays and weekends and during different times of day, the types of violations issued shift in ways that mirror broader enforcement priorities and driving behaviors. Together, these insights point to opportunities for law enforcement to tailor resource deployment, engage community stakeholders, and refine policies to promote safer roads and more equitable treatment for all drivers.

**References**

1. Li, Y., Li, M., Yuan, J., Lu, J., & Abdel‑Aty, M. (2021). Analysis and prediction of intersection traffic violations using automated enforcement system data. *Accident Analysis & Prevention, 162*, 106422. <https://doi.org/10.1016/j.aap.2021.106422>
2. B. Ayuso, M., Guillén, M., & Alcañiz, M. (2010). The impact of traffic violations on the estimated cost of traffic accidents with victims. *Accident Analysis & Prevention, 42*(2), 709–717. <https://doi.org/10.1016/j.aap.2009.10.020>
3. C. Balasubramanian, V., & Sivasankaran, S. K. (2021). Analysis of factors associated with exceeding lawful speed traffic violations in Indian metropolitan city. *Journal of Transportation Safety & Security, 13*(2), 206–222. <https://doi.org/10.1080/19439962.2019.1626962>
4. D. Saville, K., Berger, D., & Levman, J. (2024). Mitigating bias due to race and gender in machine learning predictions of traffic stop outcomes. *Information, 15*(11), 687. <https://doi.org/10.3390/info15110687>
5. E. Elfahim, O., El Midaoui, M., Youssfi, M., & Bouattane, O. (2023). Traffic violations analysis: Identifying risky areas and common violations. *Heliyon, 9*(9), e19058. <https://doi.org/10.1016/j.heliyon.2023.e19058>
6. F. Aliane, N., Fernandez, J., Mata, M., & Bemposta, S. (2014). A system for traffic violation detection. *Sensors, 14*(11), 22113–22127. <https://doi.org/10.3390/s141122113>
7. G. Gamage, P., & Hines, K. (2020). Traffic violations: An exploratory and predictive look at traffic incidents in Montgomery County, Maryland. *Georgetown University Analytics 512: Statistical Learning* (Unpublished class project).
8. H. Rababah, M., Maydanchi, M., Pouya, S., Basiri, M., Azad, A. N., Haji, F., & Aminjarahi, M. (2022). Data visualization of traffic violations in Maryland, US. *arXiv* preprint arXiv:2208.10543. <https://doi.org/10.48550/arXiv.2208.10543>
9. I. Kumfer, W., McGowan, J. C., Nordback, K., Vann, M., Lan, B., Frizzelle, B., & Thomas, L. (2024). Systemic predictive safety analysis of pedestrian crashes for Montgomery County’s Vision Zero Program. *Transportation Research Record, 2678*(11), 2165–2180. <https://doi.org/10.1177/03611981241247178>