IS 445 Final Project Data Story Message

This data story targets mainly on non-technical audience, especially drivers in Montgomery County in the United States. Traffic violations relate to both personal and social attributes, including drivers' habits, road conditions, surrounding buildings, etc. Drivers need to get familiar with possible traffic violation types due to certain surroundings to avoid such violations. By revealing internal relationships between common types of traffic violations and geographical information via intuitive data visualization, non-technical audiences like drivers in this area are more likely aware of potential challenges and cautions when driving in Montgomery County. Apart from giving reminders and warnings for non-technical audiences like drivers, this data story message might also be delivered to technical audiences to improve research on predicting traffic violations with geographical information.

This work and data story derives from a social-aware issue. Vehicle usage is increasing in daily life, with an increase in traffic violations and car accidents. Regulations and laws have explicit penalties for different types of traffic violations as global reminders and warnings for drivers. However, not every traffic violation is deliberate, and some types of traffic violations are greatly affected by environmental factors like geographic information or road conditions. Investigation of such traffic violations with clustering effect and certain geographic relationships is essential and meaningful for drivers to avoid unintentional traffic violations. Since different areas have unique geographic features, this work focuses on the impact of geographic factors on traffic violations in Montgomery County in the United States.

Existing works have already reached some success in figuring out the geographic impact of traffic violations. Sukhaia et al. (2013) correlate road conditions to traffic violations using road traffic fatalities in South Africa. Li et al. (2020) estimate the space and time patterns of traffic violation behaviors to investigate the relationship between traffic violations and urban surroundings. Elfahim et al. (2023) perform a comparison between different clustering optimizations to detect common types of traffic violations within certain districts. However, these existing methods are mostly generalized to simple impacts of common geographic features or concentrating on fatal traffic violations, indicating their weaknesses in analyzing area-specific and geographically sensitive data acquired from places like Montgomery County. Those disadvantages of existing work motivate us to develop new and targeted research on traffic violations in Montgomery County.

Back to our focus, investigating geographic effects on traffic violations in Montgomery County is an unprecedented research problem. I would like to figure out the potential impacts of geographical locations or surroundings on the number and types of traffic violations. Specifically, it makes sense to verify if some categories of traffic violations are dominant in areas with certain geographical locations. There might also be some hidden patterns of geographical locations influencing the distribution of traffic violations. Finally, I hope with

the relationship between geographical locations and certain types of traffic violations, we can further think of how certain traffic violations can be avoided, from the drivers' side and society's side.

When it comes to resolving the above research question, several separate steps can be considered in sequence. First of all, it is necessary to obtain a visualization plot showing the positions of all traffic violations. Moreover, it makes sense to investigate the individual effect of geographic locations on the number and types of traffic violations via model fitting and visualization. Furthermore, it is advised to figure out potential patterns hidden between different types of traffic violations with multivariate visualization tools. While visualizing the relationships between abundant variables, it is worth considering if those relationships make sense according to our life experience. If the relationships violate our intuition, investigation into additional geographic information might be necessary. Last but not least, time-series analysis can be applied to traffic violations in Montgomery County to visualize the trends of traffic violations.

Though it is a rarely studied topic, several reasonable hypotheses can be concluded from previous preparations. Based on previous article analyses (Li et al. (2020)) and life experience, some types of traffic violations might dominate in certain areas. Places with large populations or traffic currents are more likely linked with some minor traffic violations like running a red light. Traffic violations, especially unintentional ones, might be closely related to certain geographic surroundings like forests, hills, or high buildings. Common traffic violations might not only relate to geographic locations but also certain periods like morning work hours or evening off-work time. As time goes by, the surroundings and road conditions might have changed a lot, indicating potential fluctuations of different traffic violations in those areas.

Those steps and hypotheses are obtained via a third perspective. Data points are collected from drivers in Montgomery County from daily driving logs between 2012 and 2016. To make sure the results are independent of any personal and emotional effects on drivers, only information regarding driving habits is included in the data. Visualization results will be demonstrated in presentation-like manners to all drivers and interested audiences. For further explanations of the results, conclusions will also be obtained from objective analyses of data but not drivers themselves.

After obtaining the results of visualizations and modeling, it is important to think of improvements and adjustments to avoid future traffic violations. For traffic violations triggered by misleading surroundings, new settings along roads can be tested via similar data collection and visualization methods. For traffic violations related to certain times, further attributes like drivers' ages, jobs, or work hours can be included in visualization and modeling. Simple but explicit instructive signs can be pulled up along the road to indicate

potential traffic violations as well. Finally, I hope that diverse relationships between geographic factors and traffic violations are useful reminders and warnings for all drivers in Montgomery County.

References

[1] Elfahim, O., El Midaoui, M., Youssfi, M., & Bouattane, O. (2023). Traffic violations analysis: Identifying risky areas and common violations. *Heliyon*, *9*(9).

[2] Sukhaia, A., & Jones, A. P. (2013). Understanding geographical variations in road traffic fatalities in South Africa. *South African Geographical Journal= Suid-Afrikaanse Geografiese Tydskrif*, *95*(2), 187-204.

[3] Li, Y., Abdel-Aty, M., Yuan, J., Cheng, Z., & Lu, J. (2020). Analyzing traffic violation behavior at urban intersections: A spatio-temporal kernel density estimation approach using automated enforcement system data. *Accident Analysis & Prevention*, *141*, 105509.