D.C. Car Crash Analysis

CSE 6242 Project Proposal

Introduction and Problem Definition (Q1, Q4)

Every year, car accidents are always among the top causes of death in Washington, D.C. In response, the city has launched its Vision Zero Initiative, an effort to reduce vehicle-related crashes to zero by 2024. The team will determine the root causes of car accidents in Washington, D.C. by analyzing patterns, seasonality, and trends in the data collected from the Vision Zero website and other sources. **Our goal is to create an interactive and robust visualization dashboard to communicate the true impact of traffic-related crashes in Washington, D.C. to stakeholders who would be most interested in D.C.'s Vision Zero plan, such as policymakers, police, and local residents (Q1 & Q4).**

Plan of Activities (Q8)

Task[1]	Status	Assigned To	Start Date – End Date
Project Ideation	Completed	All Team Members	10/02 – 10/08
Set-up/Training of GitHub	Completed	Justin Schulberg	10/05 – 10/09
Project Proposal Writeup	In progress	All Team Members	10/09 – 10/15
Project Proposal Slides	Not started	All Team Members	10/09 – 10/15
Project Video	Not started	Ryan Doogan, Maynard Miranda, Adam Peir	10/09 – 10/15
Identification of Datasets	In progress	Ryan Doogan	10/16 – 10/29
Data Pre-processing	Not started	All Team Members	10/23 – 11/06
Progress Report	Not started	All Team Members	10/23 – 11/04
Data Integration	Not started	All Team Members	10/30 – 11/13
Data Analysis/Modeling	Not started	All Team Members	11/06 – 11/29
Data Visualization/Dashboard	Not started	All Team Members	11/06 – 12/03
Finalizing Final Report & Project Poster (Q9)	Not started	All Team Members	11/27 – 12/10

Unless otherwise stated, tasks will be distributed equally amongst the team members. The team will use agile methodology by having biweekly meetings (Sunday and Thursday) and tracking all relevant changes to analytic products via GitHub. Every Sunday, the team will meet to determine what needs to be done during the week and distribute the upcoming tasks equally to each team member. Every Thursday, the team will regroup to discuss the work done and compile each member's output. The cycle will repeat until the project is completed.

Expected Innovation (Q3 , Q7)

Although D.C.'s Vision Zero website includes numerous data visualizations and analytical products, they lack the following features, that we aim to improve upon (Q2):

- **Disjointed Visualizations** | Even though all of the visualizations are built in Tableau and portray insightful information, they are not connected in one seamless dashboard, allowing users to see relationships across multiple features at once.
- **Disparate Tools Used** | Currently, the Vision Zero Team creates its visualization products in both Tableau and ArcGIS. This allows for more powerful mapping capabilities in ArcGIS; however, the two products are kept separately, making it difficult to understand the relevance of the geospatial data in the context of other visualizations.
- **Lack of Diversity in Datasets Used** | Currently, only data on direct crashes are being used.
- **Insufficient Incorporation of Geographic Analysis** | While current data is mapped on the D.C. Vision Zero site, there is substantial room for additional spatial analysis of existing crash data.

To build upon this, we will not only integrate multiple data sources, but we will centralize all the data and corresponding visualizations in Tableau. Because we will only be using Tableau, and it offers free licenses to those with University emails, the cost of this project is $0 (Q7). Additionally, we will use time-series modeling to project traffic-related crashes into the future. Lastly, the team plans to integrate not only crash/fatality data, but also demographic, socio-economic, and land use data such as the location of new road infrastructure.

Intended Impact (Q5, Q6, Q9)

We will integrate multiple data sources to see how different road calming measures (raised surfaces, slowed speed limits, etc.) affect crash frequency before/after implementation. If this effort, or other efforts like it, are unable to find effective methods of crash mitigation, more lives will continue to be at stake. Potential risks exist in the methods for pulling the data and also in joining data. For example, joining crash data, which occurs at a specific location, to road calming efforts, which usually occur in a broader area or segment of road can be tricky. (Q6)

For this effort to be successful, we will adhere to the aforementioned Plan of Activities and drive towards effective data integration of our various data sources, modeling of the data that provides meaningful insight into future trends, and development of our interactive dashboard (Q9). Success of this project will contribute to the success of D.C.'s goal of zero crash-related fatalities by 2024, which the D.C. government will measure by the end of 2024 (Q5 & Q6).

Literature Review (Q2/Q3)

On top of D.C.'s Vision Zero plan, many traffic safety researchers have approached the issue of reducing vehicle-related crashes using analytical approaches. Researchers have identified common Vision Zero measures and analyzed their effectiveness across countries and cultures (Kim, et al. 2017). Crashes in Washington D.C. were predicted using two separate models: ARIMA and Heston. ARIMA assumes that any volatility in the data is constant, while Heston assumes that the volatility is arbitrary. The study shows that Heston has better accuracy than ARIMA (Shannon & Fountas, 2022). Crashes were evaluated based on different collision types, such as rear-end crash, sideswipe crash, and angle crash. Bayesian analytics was used to perform the evaluation. The result shows that each collision type has different rates and risk factors (Guo et. al, 2019). Different factors affect pedestrian crashes in Texas' county-level areas using OLS Regression. The result suggests that homelessness, median household income, and poverty positively correlate with pedestrian crashes (Bernhardt & Kockelman, 2021). The results and assumptions used in the said studies will be leveraged in our analysis and model. The major shortcoming of the studies is that they do not have an interactive dashboard that can better communicate their results to non-data stakeholders.

One component of our crash-related analysis is on bicycle safety. Some of the most common forms of vehicle-related crashes are between cars and bicycles (Daraei et. al, 2021). Many studies often cite the presence of cycling lanes as a causal link to reducing crashes for cyclists; however, some of these studies only look at the absolute presence of a cycling lane (i.e., does one exist or not) as a factor for reducing crashes (Marshall et. al, 2019). In actuality, different types of cycling lanes exist and have different effects on road safety. Our research will look at these different types of road-calming and cycling infrastructure measures, trying to enhance some of the work done in papers that only use observational methods to analyze the impact on vehicle-bicycle crashes between different cycling infrastructures (Jensen, 2007).

Geographic analysis is another crucial component of analyses of pedestrian safety. Numerous studies have incorporated spatial analysis into crash analyses in order to understand how factors specific to certain geographic units influence crash frequency. Researchers in North Carolina (Pulugurtha et. al, 2010) and China (Wang et. al, 2016) analyzed the factors specific to signalized intersections and traffic analysis zones, respectively, which predicted crashes in the relevant zones. Another set of researchers in Florida (Lee et. al, 2017) went even farther, comparing the performance of crash prediction models at different levels of geographic aggregation.

Finally, an interactive and robust visualization communicates our findings to stakeholders in D.C. In the IEEE paper, researchers superimposed photo enforcement citation data on crash data to illustrate the effectiveness of automated enforcement (Rogers et. al, 2016). Additionally, researchers have also used transportation, mobile, and demographic data to determine safety risks in the state of Maryland (Xiong et. al, 2021). Potential shortcomings of these visualizations include their limited scopes.

Bibliography (APA)

Bernhardt, M., & Kockelman, K. (2021, June 3). An analysis of pedestrian crash trends and contributing factors in Texas. Journal of Transport & Health. Retrieved October 12, 2022, from https://www.sciencedirect.com/science/article/pii/S2214140521001201

Daraei, S., Pelechrinis, K. & Quercia, D. A data-driven approach for assessing biking safety in cities. EPJ Data Sci. 10, 11 (2021). https://doi.org/10.1140/epjds/s13688-021-00265-y

Guo, Y., Li, Z., Liu, P., & Wu, Y. (2019, April 29). Modeling correlation and heterogeneity in crash rates by collision types using full Bayesian random parameters multivariate Tobit model. Accident Analysis & Prevention. Retrieved October 12, 2022, from https://www.sciencedirect.com/science/article/pii/S0001457518311576

Jensen, S. U. (2007, November 7). Bicycle tracks and lanes: A before-after study - researchgate. Retrieved October 14, 2022, from https://www.researchgate.net/profile/Soren-Jensen-16/publication/237524182_Bicycle_Tracks_and_Lanes_a_Before-After_Study/links/5a548377458515e7b732688e/Bicycle-Tracks-and-Lanes-a-Before-After-Study.pdf?origin=publication_detail

Kim, E., Muennig, P., & Rosen, Z. (2017, January 9). Vision Zero: A toolkit for road safety in the modern era - injury epidemiology. SpringerLink. Retrieved October 12, 2022, from https://link.springer.com/article/10.1186/s40621-016-0098-z

Kondo MC, Morrison C, Guerra E, Kaufman EJ, Wiebe DJ. Where do bike lanes work best? A Bayesian spatial model of bicycle lanes and bicycle crashes. Saf Sci. 2018 Mar;103:225-233. doi: 10.1016/j.ssci.2017.12.002. PMID: 32713993; PMCID: PMC7380879.

Lee, J., Abdel-Aty, M., & Cai, Q. (2017, March 21). *Intersection crash prediction modeling with macro-level data from various geographic units*. Accident Analysis & Prevention. Retrieved October 12, 2022, from https://www.sciencedirect.com/science/article/abs/pii/S0001457517301070

Marshall, W. E., &; Ferenchak, N. N. (2019, May 29). Why cities with high bicycling rates are safer for all road users. Journal of Transport & Health. Retrieved October 13, 2022, from https://www.sciencedirect.com/science/article/abs/pii/S2214140518301488

Rogers, J. M., Dey, S. S., Retting, R., Jain, R., Liang, X., & Askarzadeh, N. (2016, December 5). Using automated enforcement data to achieve vision zero goals: A case study. Retrieved October 12, 2022, from https://ieeexplore.ieee.org/abstract/document/7841099/authors#authors

Shannon, D., & Fountas, G. (2022, March 3). Amending the heston stochastic volatility model to forecast local motor vehicle crash rates: A case study of washington, d.c. Transportation Research Interdisciplinary Perspectives. Retrieved October 12, 2022, from https://www.sciencedirect.com/science/article/pii/S2590198222000392

Wang, X., Yang, J., Lee, C., Ji, Z., & You, S. (2016, July 29). *Macro-level safety analysis of pedestrian crashes in Shanghai, China*. Accident Analysis & Prevention. Retrieved October 12, 2022, from https://www.sciencedirect.com/science/article/abs/pii/S0001457516302573

Xiong, C., Mahmoudi, J., Luo, W., Yang, M., Zheng, J., & Delion, C. (2021, August 31). A data-driven safety dashboard assessing Maryland statewide density exposure of pedestrians, bicycles, and e-scooters. A Data-Driven Safety Dashboard Assessing Maryland Statewide Density Exposure of Pedestrians, Bicycles, and E-Scooters. Retrieved October 12, 2022, from https://rosap.ntl.bts.gov/view/dot/61218