Exploring the Dynamics of Fatal and Serious Injury Vehicle Accidents in Minnesota

Luke Zaruba, MGIS Program, University of Minnesota

Advised by Dr. Bryan Runck, University of Minnesota

September 20, 2023

Abstract

This paper details the development and implementation of a near-real-time geospatial data pipeline of fatal and serious traffic accidents across the State of Minnesota. To accomplish this, data is scraped and preprocessed via a serverless, automated pipeline on the cloud, before various spatial and temporal analytics are run on the resulting data to provide insights into crash hotspots and clusters across both space and time. The novelty of this pipeline lies in the creation of an open, near-real-time, geospatial web application to view serious traffic crashes in Minnesota, as well as investigations made into the integration of generative artificial intelligence (AI) into the geocoding process and the use of simple spatial analyses to study the stability of point clustering footprints over time. This project shows several insights into the historical patterns of car crashes across the State, specifically concerning: 1) the need for advances in transportation engineering and increased enforcement and emergency services in areas of high concern; and 2) the effects that COVID-19 has had on traffic safety.

Keywords: spatial data science, spatiotemporal analysis, transportation, traffic safety

1. Introduction

Traffic safety is at the forefront of thought amongst the public, policymakers, and engineers alike when it comes to thinking about how transportation can change to meet the expectations and needs of society. Traffic accidents are increasingly recognized as a major issue, due to the extensive economic and social losses that they cause [1]. Minnesota is no exception in recognizing the threats and challenges that traffic accidents pose to the State, with the State's Towards Zero Deaths program operating with an interdisciplinary approach towards mitigating traffic crashes, ultimately seeking to change the culture around traffic accidents [2]. In 2021 alone, there were 488 fatalities from traffic accidents, with contributing factors including speed, a lack of seatbelt usage, drunk driving, distracted driving, and more [3]. The use of Geographic Information Science (GIScience) and data science to gain insights into patterns of traffic crashes across both space and time is not a novel idea by any means [1]. However, it can be difficult to find detailed information and analyses that are open for all to see and easily understandable from a spatiotemporal perspective. In Minnesota, for example, most publicly available information has been aggregated to the County level at the yearly or monthly temporal aggregation level.

The goal of this project is to introduce a more fine spatial resolution set of data to the public, by utilizing fatal and serious injury crash data from across Minnesota, allowing anyone to interact with the data and analyze it both spatially and temporally. In addition to public engagement, other key efforts associated with this project are to explore and push the boundaries of different spatiotemporal analysis approaches and to engineer a high-performance, near-real-time pipeline to ingest, process, and analyze the data, while exploring the potential for integrating new technologies like generative AI into the process. These include a simple, yet novel approach to exploring the stability and robustness of point clustering over time, as well as an investigation into the capabilities of generative AI to enhance geocoding processes. As a result, several key insights can be gleaned from the analyses that have been performed, including data that can help identify areas of key concern, as well as perceptions of how traffic accidents have changed across space and time in relation to COVID-19.

The structure of this paper is as follows: first, a thorough review of the various facets of spatial and spatiotemporal analysis, as well as geocoding and traffic safety will be examined to provide context for the analyses performed in this project. Next, an exploration of the data and methodologies employed in this project is presented, followed by a discussion of the results and their interpretations. Finally, the paper concludes with a discussion of the implications, challenges, and future opportunities for research in the field.

2. Related Work

2.1. Traffic Safety

From how roads can be engineered more safely [4] to the psychological effects of traffic accidents on children [5], research into traffic safety often encompasses a variety of disciplines, given the far-reaching and broad implications that they have on society. Given the intrinsically spatial dimension of traffic accidents, it is not surprising that GIScience is often incorporated into efforts to better understand and study traffic safety. Typically, GIScience is used to examine where accidents are occurring, while data science and statistics are used to answer non-spatial questions like how, why, and when they are occurring [6]. However, they are often used in conjunction with one another, especially to examine the relationship between when and where through spatiotemporal analyses. In addition to exploring the patterns of accidents spatially, it is also common to explore the patterns of factors and causes, which can then be used to assist in modeling and predicting traffic accidents. Together, spatial and non-spatial attributes of accidents can be used to aid in the development of crash prediction models [7].

2.2. Geocoding and Generative AI

Geocoding is oftentimes an essential part of any GIScience project, as it allows for the seamless translation of textual information to spatial data, providing data that can then be utilized for spatial analysis [8]. In essence, geocoding works by comparing some input textual description of a location to a reference dataset composed of explicit and relative descriptions of reference features, tying the spatial information of the most probable match in the reference dataset to the input description. It is a complicated process, which borrows heavily from fields like "information theory, decision theory, probability theory, and phonetics" [9]. There have also been many studies that seek to integrate deep learning and natural language processing (NLP) into the geocoding process, mainly for standardization processes and toponym recognition tasks [10]. However, for NLP to be used effectively, a domain-specific model is often required to produce reasonably accurate results. This can be a time-consuming process and can be quite difficult to develop an effective model, which needs large amounts of training data to learn. With the rise of generative AI, specifically Large Language Models (LLMs), like OpenAI's GPT models, there are new opportunities to integrate these generalized models into the geocoding process without the need for developing a new model, since LLMs are foundational models, capable of performing many different tasks. There has been a dramatic boost in this area of research, with studies attempting to detail and analyze the capabilities of LLMs for geospatial tasks, typically with poor to moderate accuracy and success, given the lack of explicitly geospatial training information that is incorporated into the training process of the models [11, 12, 13]. Hence, this creates a need for the development of fine-tuned, domain-specific LLMs to truly enable generative AI to make a significant impact on GIScience and geospatial-related tasks, like geocoding [14].

2.3. Spatial Analysis & Statistics

Spatial analysis is essential for understanding how accident patterns occur across space, especially for answering questions about concentrations or hotspots of accidents. LISA, or the Local Indicators of Spatial Autocorrelation, is one such method that utilizes the Local Moran's *I* statistic to find statistically-significant, spatial clusters and outliers of high and low values, of polygons [15]. By using a measure like LISA, polygons can be divided into five groups, polygons with non-significant values, high-high clusters, low-low clusters, high-low outliers, and low-high outliers, which in the case of accidents can show how accident concentrations compare from one spatial unit (e.g., county or city) to its neighbors. This can especially be useful to identify clusters of high accident counts where mitigation efforts can be deployed, or to look at outliers where there are high accident counts surrounded by low counts, to further explore why this might be the case. There are also other methods and statistical measures like the Getis-Ord Gi* statistic which measures the spatial association of values, essentially showing spatial clusters [16].

Point-process analyses are also a key area of the literature in spatial analytics and statistics. In the context of this paper, point clustering and data mining will be the primary focus. Being able to identify point clusters across space has many applications, but for traffic accidents, it allows for the identification and classification of points into clusters based on the similarity that points share with each other. However, inherently the data is subject to large amounts of noise since traffic accidents could occur anywhere, hence the need for an algorithm that can recognize noise. Density-Based Spatial Clustering of Applications with Noise, or DBSCAN, is one of the most frequently used data mining algorithms to identify clusters while also recognizing that noise exists within the data [17]. Removing noise, or points that don't fit well into clusters, improves the usability and ease of interpretation of results, since clusters are resistant to the effects of outliers and the number of clusters is based solely on the density distribution, rather than a user-defined parameter like what is required in other algorithms, such as K-Means Clustering. Various extensions of DBSCAN provide additional capabilities, including ST-DBSCAN for adding a third, temporal dimension, which enables spatiotemporal clustering [18], HDBSCAN for a hierarchical-based approach to clustering [19], as well as A-DBSCAN for creating robust, cluster-based areal delineations [20]. A-DBSCAN is a recent development, that examines cluster stability by using an ensemble approach to iteratively run DBSCAN on a random subsample of the data points and use a nearest-neighbor algorithm to expand the values to include the rest of the non-sampled points, therefore allowing for some measure of stability to be calculated, given the stochastic nature of the approach [20]. Although the original usage of A-DBSCAN was for delineating urban areas through building footprints, the algorithm holds great promise for use in studying traffic accidents, to delineate areas of high accident concentration, where mitigation and traffic safety efforts could be directed more precisely.

Spatial analytical and statistical methods are more commonly used on polygonal or point data, given the inherent simplicities in comparison to network, or line-based data. In the case of traffic accidents, this poses a challenge. However, if a network is topologically sound and connected correctly, this does provide the ability to perform methods like LISA on the network to identify spatial clusters and outliers, with spatial weights constructed through connectivity. This is comparable to constructing spatial weights via a Queen's adjacency matrix for a polygonal dataset. Additional methodologies exist for network-based point pattern analyses, such as using Kernel Density Estimation in the context of a one-dimensional network space [21, 22], and modeling spatial interactions [23], amongst many others. One additional concern of using network-based data is the limitation that is associated with geocoding traffic accidents. It is far easier to successfully and accurately match a point to the correct polygon that it resides in than it is to match the polygon to the correct nearest segment of a network. These implications have a fair amount of literature regarding how to best deal with geocoding quality, which includes options such as validation through volunteered geographic information [24].

2.4. Spatiotemporal Analysis & Time Series Analysis

When a temporal dimension is added to spatial data, this changes the approaches that can be taken if time is accounted for in any analysis. The main difference in approaches is whether or not space is explicitly taken into account or not. For example, if a forecasting algorithm is run on multiple time series for different spatial units, space is not taken into account, whereas if ST-DBSCAN is used to identify spatiotemporal clusters, the spatial dimensions of the data are inherently taken into account. Spatial analyses can also be applied to subsets of data across time, providing results that can be compared to one another, which offers itself as a simple, yet intuitive and useful solution to working with spatiotemporal data.

When working with time series data, forecasting is often thought of as the primary form of analysis, however, several other options exist outside of forecasting, including change point detection or discontinuity analysis. The goal of change point detection is to find points in a signal (i.e., a time series) where abrupt, sudden changes in the underlying signal occur [25]. From this change detection, the signal can effectively be segmented into multiple components, and breakpoints can be identified. In the case of a traffic accident time series, breakpoints would be where major shifts in the trends and patterns of accident occurrences exist, therefore providing potentially useful information when looking at how policy changes and other events affect the time series.

3. Methods

3.1. Data

The data used in the subsequent analyses of this paper were obtained from the Minnesota State Patrol's Crash Updates site, which contains records starting from January 1, 2017, to the present time [26]. In this analysis, the most recent records are from early September 2023. The site is scraped using a custom, Python-based web scraper, which obtains all accident information that can be found within the site. The scraper extracts the data from the website(s) and then uploads all the data to several tables in a relational database, hosted on the cloud.

3.2. Preprocessing

After the raw, extracted data is uploaded to the relational database, several key preprocessing and data engineering tasks must be performed before analysis, with the first being geocoding. Based on a small experiment into the performance of various geocoding processes [8], the raw location descriptions associated with each accident are fed directly into the Esri geocoding API. The resulting accident data, featuring geospatial data, is uploaded to a spatial table within the database. Given the need for quality assurance and control, some basic spatial filtering is applied to the table to remove errant points that were geocoded incorrectly, resulting in a spatial table with all of the accidents that were geocoded within Minnesota. With the geocoding complete, the next step in the pipeline is to run various SQL queries to aggregate the data, spatially and temporally. These operations resulted in a new spatial table featuring the total number of accidents within each City, Township, or Unorganized Territory (CTU) across the State, as well as a time series table showing the total accident count across the State for each week across the entire length of the time series, as well as updates made to existing tables based on the information created in these steps. After this, the data is ready for any analysis to be performed.

3.3. Spatial Analysis

Several spatial analyses are performed on the data when the pipeline is executed. First, LISA is run on the CTU table to identify spatial clusters and outliers of the CTUs, based on the number of traffic accidents within each spatial unit. Rather than passing through the raw count of accidents into the model, the values are standardized by the total length of all roads within the CTU. This standardization allows for easier comparison between CTUs with differing amounts of roads, populations, and concentrations of traffic. Additionally, the LISA model uses spatial weights constructed through the use of a Queen's contiguity-based neighborhood. After the analysis is run, the output metrics are updated in the spatial database.

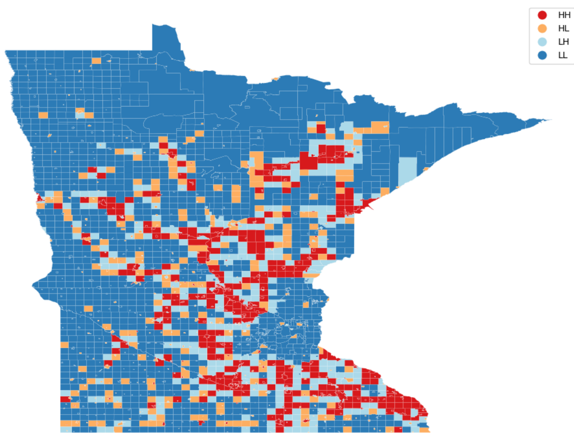
Next, the A-DBSCAN algorithm is run in a few different settings, each producing polygonal outputs that represent the cluster footprints. First, the entire set of traffic accidents is sent through the algorithm to produce an up-to-date, current set of cluster footprints. Second, the traffic accidents are split into subsets based on the year of the incidents, before the algorithm is run on each subset (i.e., year) of data, before aggregating the footprints created from each run into a single table stored in the database. Finally, to examine the temporal stability of these clusters across time, a basic overlay analysis is performed on the yearly footprints, which yields a new set of polygons that have counts associated with how many times an overlap in footprints occurred for an area. In other words, if an area was contained within a cluster footprint for three different years then the polygon over that area would have a value of three. This means that based on this dataset, areas that have been consistently included in accident clusters will have high values, which reveals where the most stable and robust clustering areas are. These areas are the areas that may be of the highest concern since they frequently contain high accident counts.

3.4. Time Series Analysis

Time series analysis is not a part of the near-real-time pipeline, but rather is a focus solely within this paper. The main focus of the time series analysis was the discontinuity or change point detection analysis. After a thorough exploratory data analysis (EDA), change point detection was conducted on the time series through Python, specifically with the ruptures package. Four different change point algorithms were utilized, each of which used squared Euclidean distance as the measure of similarity, and were tested using minimum segment sizes of 10, 20, and 30. These minimum segment sizes indicate the number of observations, or the minimum distance required, between two change points in the signal. The four algorithms used include the Pelt, Binary Segmentation, Window, and Bottom Up models, which all take different approaches to determining the most ideal way to detect changes within the underlying signal. After running the models and detecting the change points through each, the results can then be compared to one another to come to a better understanding of where significant changes occurred throughout the time series.

3.5. API & Web Application Development

To make the data open and accessible, the first step in developing this capability was to create an Application Programming Interface, or API, to serve out the data from the relational database to users. After experimentation with various configurations, including through the use of different languages and frameworks like Flask for Python as well as several implementations using the programming language Go, the decision was made to use Flask and Python for the production version of the API, due to the simplicity and speed of development, as well as the performance. The API makes use of Google Cloud Run to more easily scale up to meet user needs and simplify the deployment process through containerization with Docker. In addition, to best follow modern software development practices, a Continuous Integration/Continuous Development (CI/CD) pipeline was created through GitHub and Google Cloud Build to automatically deploy changes made to the API to Cloud Run. Through the use of an API, the data can then be easily integrated into a variety of geospatial web applications, allowing non-technical users to view and interact with the data.



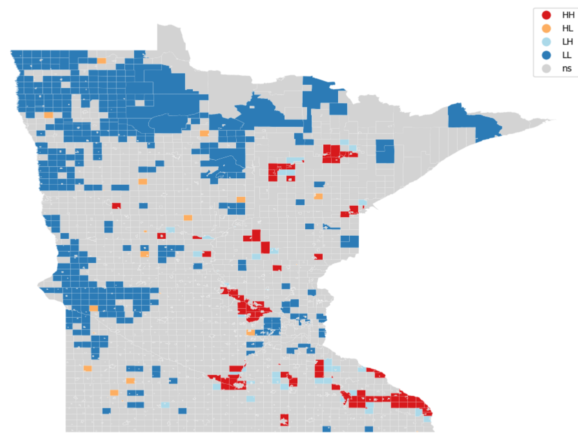


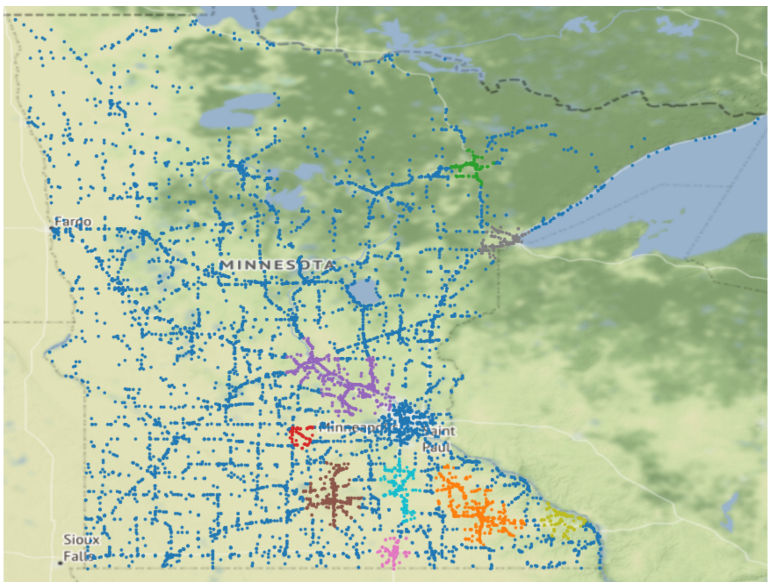
Figure 1. LISA results (left) & *Figure 2.* Statistically significant LISA results (right).

4. Results

4.1. Polygon-based Analysis Results

After running the polygon-based, LISA analysis on the CTUs, several observations can be made quite quickly. Without taking the statistical significance into account (Fig. 1), a “horseshoe-like” pattern emerges of high accident clusters, which surrounds the Twin Cities Metropolitan Area (TCMA). The high clusters that exist mostly correspond with a few large roads like Interstates 90 and 94, as well as other more-populated areas like Rochester, Mankato, Duluth, Saint Cloud, and Hibbing, to name a few. The TCMA itself appears as a low cluster, which is likely due to the standardization, meaning that the high volume of roads within the area minimizes the effect of having likely higher raw accident counts. Much of the outer portions of the State are also low clusters, which corresponds with both the low accident counts and the smaller volumes of roads in these rural areas.

If these results are filtered to only show statistically significant Local Moran's *I* values from the LISA analysis, these patterns can also be seen at much smaller scales (Fig. 2). The low clusters that existed across much of the State are filtered down to only including the far Northeast and Western parts of the State. Meanwhile, the largest groupings of high clusters are centered around Mankato, as well as stretches of Interstate 94 from Saint Cloud to the Northwestern suburbs of the TCMA and from Rochester to Winona through Interstate 90 and U.S. Highway 14.



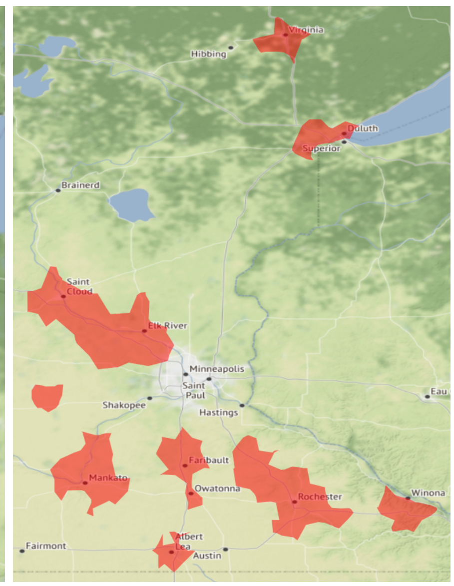
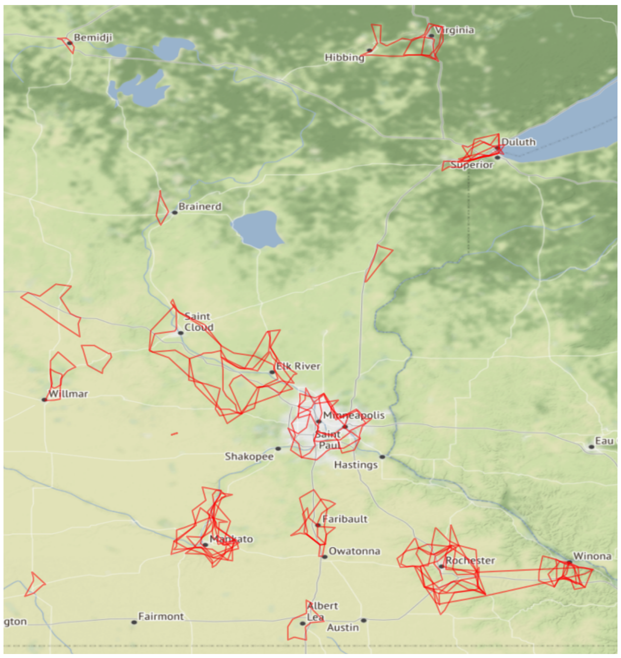


Figure 3. A-DBSCAN results (left) & *Figure 4.* A-DBSCAN footprint results (right).

4.2. Point Pattern/Clustering Results

Unsurprisingly, the point clustering results share many similarities with the LISA analysis performed on the CTU polygons. However, there are some clear patterns in the point patterns that are not picked up. In Figures 3 and 4, the A-DBSCAN clusters are displayed through points and as footprints, respectively. The point patterns reveal the roadways quite well, with several prominent roads like Interstate 94, U.S. Highway 10, and State Highway 23 being shown very distinctly. The clusters themselves are similarly concentrated around Saint Cloud, Duluth, Hibbing, and several prominent cities in Southern Minnesota. The final results of the clustering analysis are for the yearly cluster footprint overlay analysis. The footprints share a lot of similarities with one another and are more or less generally placed in similar areas, with the occasional footprint occurring in different areas throughout the State, including in the TCMA. Figure 5 shows the outlines of these footprints to visualize how the details of the footprints change at a fine resolution, while Figure 6 shows a smoother visualization of the number of times that each area has been classified as being within a cluster. In Figure 6, we can see that the most stable clusters of accidents tend to be in the areas surrounding Mankato and Rochester.



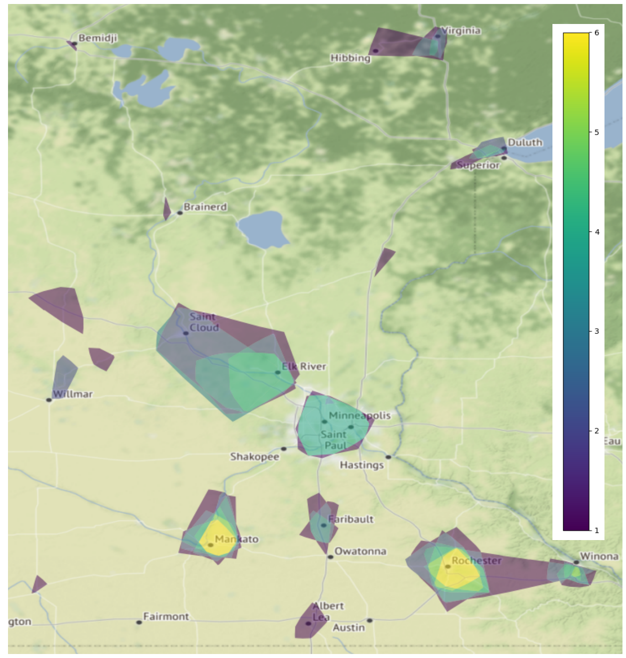


Figure 5. Yearly A-DBSCAN outlines (left) & *Figure 6.* Yearly A-DBSCAN smoothed overlay (right).

4.3. Time Series Analysis Results

The change point detection analysis revealed some key patterns after the different model results were aggregated together. In Figure 7, the change points are overlaid on the time series of traffic accidents, for the four models with the different minimum segment parameter values. In Figure 8, all of the change points are aggregated to show the most commonly detected change points across all models and configurations. Arguably the most interesting parts of the analyses are the trends and change points from 2020 to 2022, during the height of the COVID-19 pandemic, especially when they are lined up with a timeline of policies and events related to the pandemic in Minnesota. The first observation is a strong decline in accidents from January to about May 2020, which corresponds with the initial stage of the pandemic. In May, as retailers were allowed to reopen at 50% capacity [27], a resurgence in accidents was also seen. Throughout the next year, the pattern of accident trends varied up and down, which likely were influenced by COVID-19 to some extent, but potentially to a lesser degree. In early 2022, as the Omicron variant of COVID-19 surged [28], there was also a drop in accident counts. After that, the time series has become more similar to the pre-COVID patterns, although it is difficult to determine if the volume of accident occurrences has returned to what it was prior to the start of the pandemic.

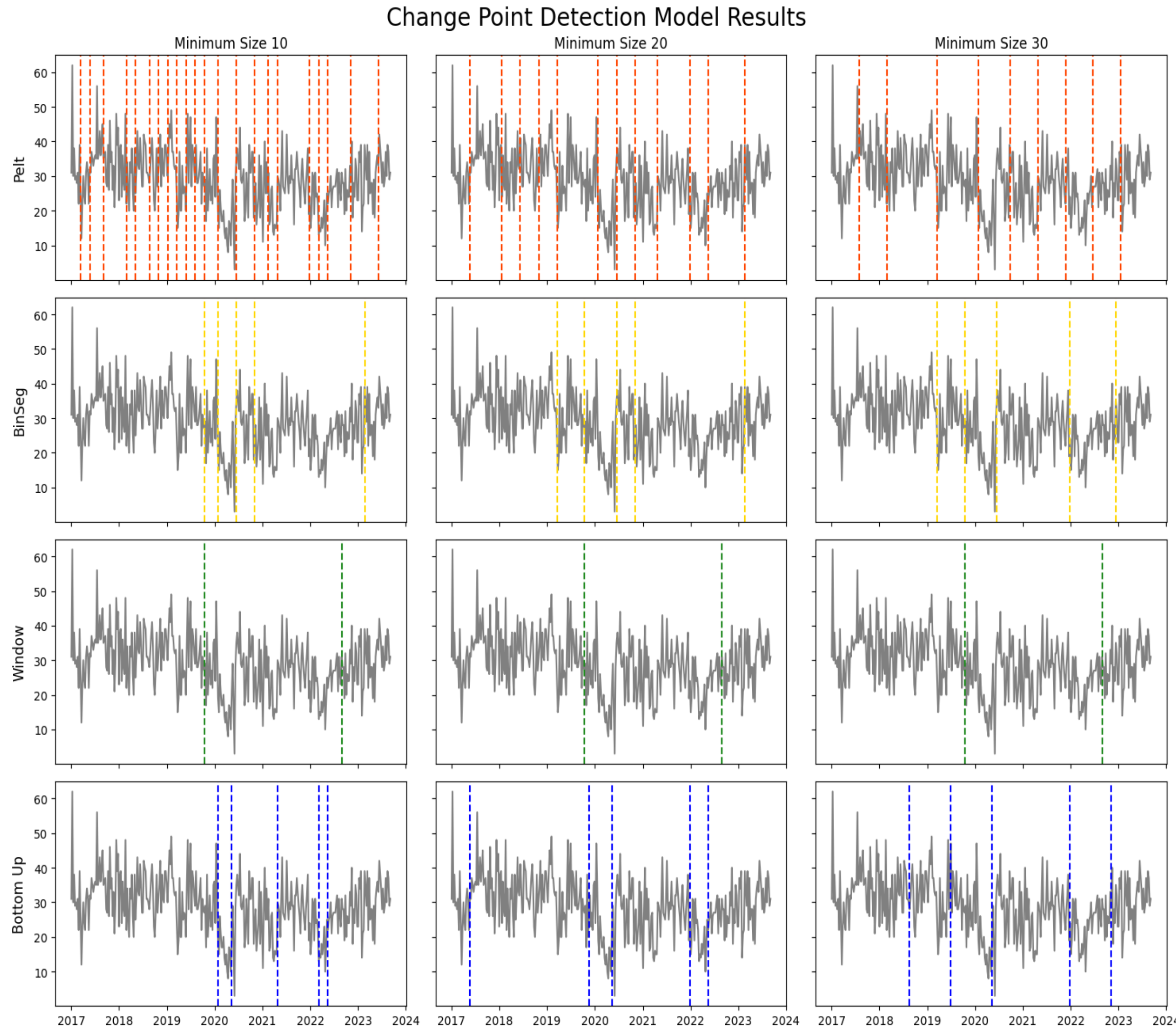


Figure 7. Change point detection results by model and minimum segment size.

Further trends in the time series through exploratory data analysis show that accidents have not quite reached pre-pandemic levels, but until 2023 is over, it is unknown how close the year will be to those levels. The analysis also showed that accidents have more frequently occurred in the Summer and Winter months, which can be seen in Figure 10. Figure 11 shows accidents by time of day, which shows that accidents are most frequent during the rush hour periods of the day, specifically in the evenings. Figure 12 shows accidents by day of the week, with which Fridays had the most accidents, while Sundays had the least. Lastly, figure 13 shows the overall daily pattern of traffic accidents, along with a smoothed average of these values.

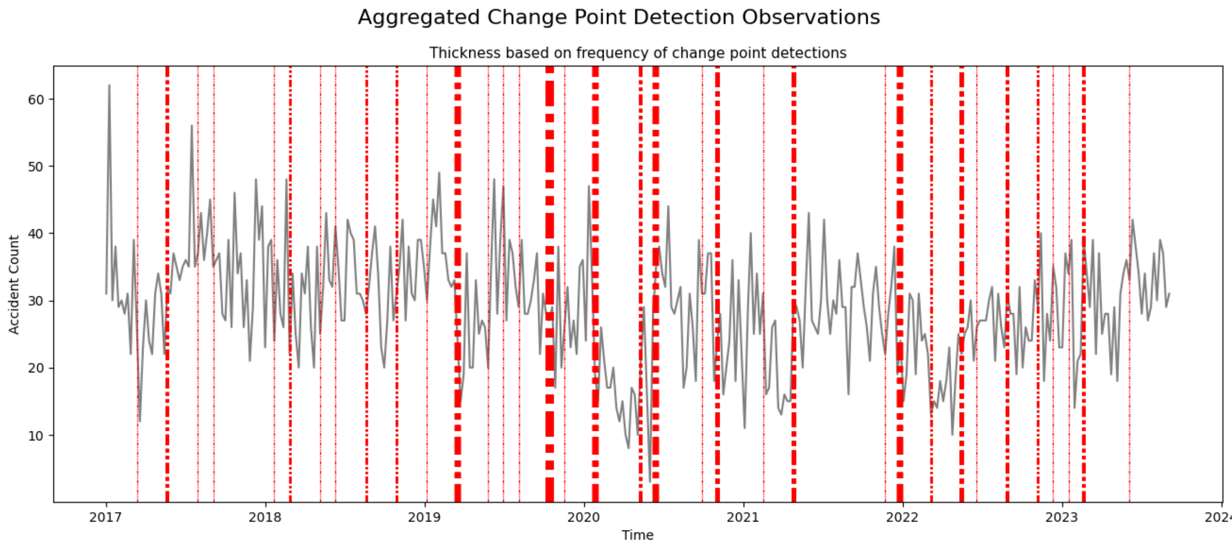
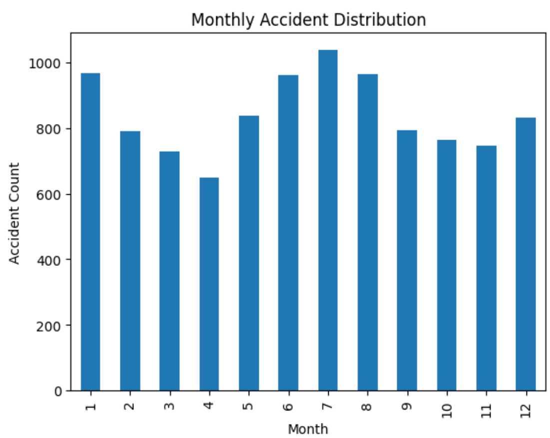
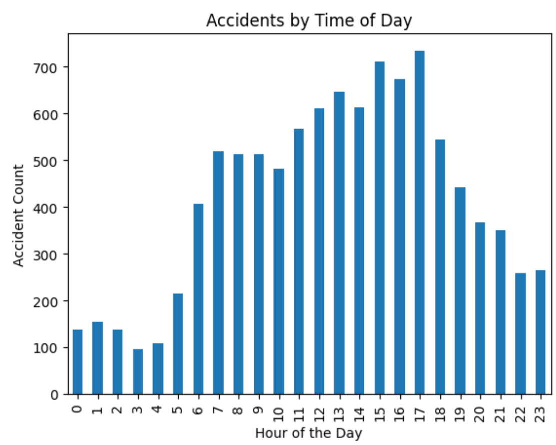
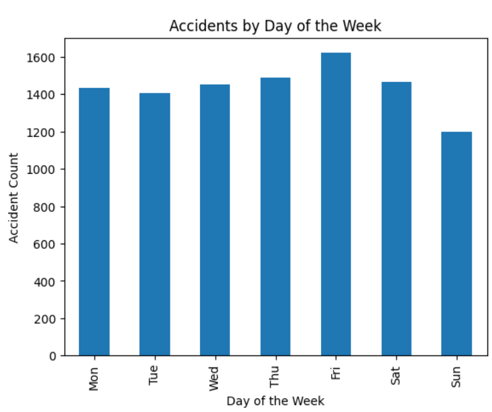


Figure 9. Aggregated change point detection results.







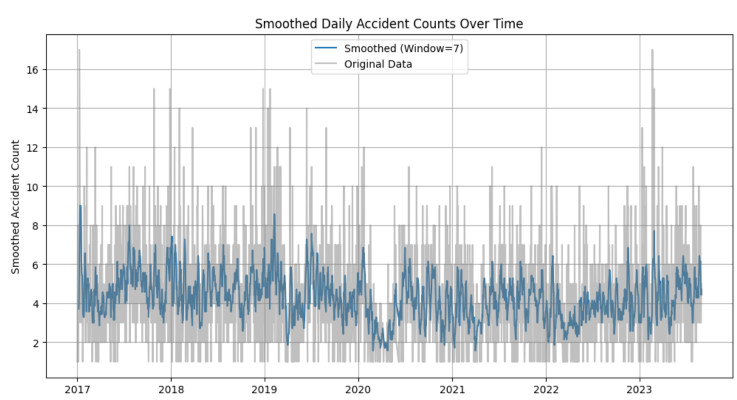


Figure 10. Monthly accident distribution (top-left), *Figure 11.* Hour-of-day accident distribution (top-right), *Figure 12.* Day of week accident distribution (bottom-left), & *Figure 13.* Smoothed accident time series (bottom-right).

5. Discussion & Conclusion

5.1. Discussion

In this work, a near-real-time data pipeline was constructed to analyze the spatial and temporal dynamics of traffic accidents across Minnesota. The work done here opens up many more opportunities to utilize the data for the development of more sophisticated and novel spatiotemporal analytical processes. Specifically, this work could continue in directions related to network-based point pattern analysis, spatiotemporal prediction through methods like Agent-Based Modeling (ABM) and deep learning, and interweaving NLP into spatial analysis, to name a few of the many potential applications. However, this work itself does have its limitations, especially related to the process of geocoding unstructured location descriptions into accident locations. As briefly discussed in this paper, there are several avenues of potential ways to remedy these issues, but one in particular that could be interesting is the use of Volunteered Geographic Information, or VGI, to perform quality assurance and control through a web application, by simply making edits to accident points to correct the imperfect predictions made in the geocoding process. Additionally, the development of a custom geocoder that could make use of other datasets like the State's Linear Referencing System (LRS) and the most up-to-date road centerline data could significantly improve the accuracy of the process.

5.2. Conclusion

Traffic accidents are a continued problem in communities across the State, but through the efforts of driver education, improving emergency services, increased enforcement of traffic laws, and engineering roads with safety as a top priority [2] there are opportunities to change the culture around traffic accidents as we know it. This work seeks to provide resources to educate the public and policymakers on where and when accidents are occurring so that they can make more informed and thoughtful decisions about how to best continue with the mission of eliminating the regularity of traffic accidents in Minnesota.

References

1. Satria, R., & Castro, M. (2016). GIS Tools for Analyzing Accidents and Road Design: A Review. *Transportation Research Procedia*, *18*, 242–247. https://doi.org/10.1016/j.trpro.2016.12.033
2. *About Minnesota TZD*. (2023). MN TZD. https://www.minnesotatzd.org/about
3. Minnesota DPS: Office of Traffic Safety. (2022). *Minnesota Motor Vehicle Crash Facts, 2021*. https://dps.mn.gov/divisions/ots/reports-statistics/Documents/CFmod_2021_Doc.pdf
4. Shah, S. A. R., Ahmad, N., Shen, Y., Kamal, M. A., Basheer, M. A., & Brijs, T. (2019). Relationship between road traffic features and accidents: An application of two-stage decision-making approach for transportation engineers. *Journal of Safety Research*, *69*, 201–215. https://doi.org/10.1016/j.jsr.2019.01.001
5. Bryant, B., Mayou, R., Wiggs, L., Ehlers, A., & Stores, G. (2004). Psychological consequences of road traffic accidents for children and their mothers. *Psychological Medicine*, *34*(2), 335–346. https://doi.org/10.1017/S0033291703001053
6. Chand, A., Jayesh, S., & Bhasi, A. B. (2021). Road traffic accidents: An overview of data sources, analysis techniques and contributing factors. *Materials Today: Proceedings*, *47*, 5135–5141. https://doi.org/10.1016/j.matpr.2021.05.415
7. Mohammed, A. A., Ambak, K., Mosa, A. M., & Syamsunur, D. (2019). A Review of the Traffic Accidents and Related Practices Worldwide. *The Open Transportation Journal*, *13*(1), 65–83. https://doi.org/10.2174/1874447801913010065
8. Zaruba, L. (2023, September 13). The Language of Locations: Evaluating Generative AI's Geocoding Proficiency. *Towards Data Science*. https://towardsdatascience.com/the-language-of-locations-evaluating-generative-ais-geocoding-proficiency-2878b61c575
9. Zandbergen, P. A. (2008). A comparison of address point, parcel and street geocoding techniques. *Computers, Environment and Urban Systems*, *32*(3), 214–232. https://doi.org/10.1016/j.compenvurbsys.2007.11.006
10. Küçük Matci, D., & Avdan, U. (2018). Address standardization using the natural language process for improving geocoding results. *Computers, Environment and Urban Systems*, *70*, 1–8. https://doi.org/10.1016/j.compenvurbsys.2018.01.009
11. Mai, G., Huang, W., Sun, J., Song, S., Mishra, D., Liu, N., Gao, S., Liu, T., Cong, G., Hu, Y., Cundy, C., Li, Z., Zhu, R., & Lao, N. (2023). *On the Opportunities and Challenges of Foundation Models for Geospatial Artificial Intelligence* (arXiv:2304.06798). arXiv. http://arxiv.org/abs/2304.06798
12. Mai, G., Cundy, C., Choi, K., Hu, Y., Lao, N., & Ermon, S. (2022). Towards a foundation model for geospatial artificial intelligence (vision paper). *Proceedings of the 30th International Conference on Advances in Geographic Information Systems*, 1–4. https://doi.org/10.1145/3557915.3561043
13. Roberts, J., Lüddecke, T., Das, S., Han, K., & Albanie, S. (2023). *GPT4GEO: How a Language Model Sees the World's Geography* (arXiv:2306.00020). arXiv. http://arxiv.org/abs/2306.00020
14. Ling, C., Zhao, X., Lu, J., Deng, C., Zheng, C., Wang, J., Chowdhury, T., Li, Y., Cui, H., Zhang, X., Zhao, T., Panalkar, A., Cheng, W., Wang, H., Liu, Y., Chen, Z., Chen, H., White, C., Gu, Q., … Zhao, L. (2023). *Domain Specialization as the Key to Make Large Language Models Disruptive: A Comprehensive Survey* (arXiv:2305.18703). arXiv. http://arxiv.org/abs/2305.18703
15. Anselin, L. (2010). Local Indicators of Spatial Association-LISA. *Geographical Analysis*, *27*(2), 93–115. https://doi.org/10.1111/j.1538-4632.1995.tb00338.x
16. Getis, A., & Ord, J. K. (2010). The Analysis of Spatial Association by Use of Distance Statistics. *Geographical Analysis*, *24*(3), 189–206. https://doi.org/10.1111/j.1538-4632.1992.tb00261.x
17. Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. *Kdd*, *96*(34), 226–231.
18. Birant, D., & Kut, A. (2007). ST-DBSCAN: An algorithm for clustering spatial–temporal data. *Data & Knowledge Engineering*, *60*(1), 208–221. https://doi.org/10.1016/j.datak.2006.01.013
19. Campello, R. J. G. B., Moulavi, D., & Sander, J. (2013). Density-Based Clustering Based on Hierarchical Density Estimates. In J. Pei, V. S. Tseng, L. Cao, H. Motoda, & G. Xu (Eds.), *Advances in Knowledge Discovery and Data Mining* (Vol. 7819, pp. 160–172). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-37456-2_14
20. Arribas-Bel, D., Garcia-López, M.-À., & Viladecans-Marsal, E. (2021). Building(s and) cities: Delineating urban areas with a machine learning algorithm. *Journal of Urban Economics*, *125*, 103217. https://doi.org/10.1016/j.jue.2019.103217
21. Xie, Z., & Yan, J. (2008). Kernel Density Estimation of traffic accidents in a network space. *Computers, Environment and Urban Systems*, *32*(5), 396–406. https://doi.org/10.1016/j.compenvurbsys.2008.05.001
22. Borruso, G. (2008). Network Density Estimation: A GIS Approach for Analysing Point Patterns in a Network Space. *Transactions in GIS*, *12*(3), 377–402. https://doi.org/10.1111/j.1467-9671.2008.01107.x
23. Batty, M. (2005). Network geography: Relations, interactions, scaling and spatial processes in GIS. In *Re-presenting GIS* (pp. 149–170).
24. Zandbergen, P. A. (2009). Geocoding Quality and Implications for Spatial Analysis: Geocoding Quality and Implications for Spatial Analysis. *Geography Compass*, *3*(2), 647–680. https://doi.org/10.1111/j.1749-8198.2008.00205.x
25. Truong, C., Oudre, L., & Vayatis, N. (2020). Selective review of offline change point detection methods. *Signal Processing*, *167*, 107299. https://doi.org/10.1016/j.sigpro.2019.107299
26. Minnesota State Patrol. (2023). *Minnesota State Patrol Crash Updates* [dataset]. https://app.dps.mn.gov/MSPMedia2/
27. Melo, F. (2020, July 13). A COVID-19 timeline for Minnesota. *Pioneer Press*. https://www.twincities.com/2020/07/11/a-coronavirus-covid-19-timeline-for-minnesota/
28. *Governor Walz Announces Actions to Support Hospitals During Omicron Surge*. (2022). Office of Governor Tim Walz. https://mn.gov/governor/covid-19/news/#/detail/appId/1/id/515219