Human Trafficking Data Collection and Aggregation

Authors:

Aakila Mohamed Rafih (moham119@msu.edu)
Jack Nugent (nugentj3@msu.edu)
Tanmay Patel (patelta2@msu.edu)
Sneha Sundar (sundarsn@msu.edu)

Abstract

This paper proposes a methodology for collecting and aggregating public data sources related to human trafficking in Michigan. The current lack of reliable and consistent measurements on the extent of human trafficking in the United States is a significant challenge, with most measures being based on WAGs (Wildly Approximated Guesses) and not well documented in publications. Our project aims to address this gap by exploring potential public data sources that could provide a more comprehensive understanding of the problem, including risks such as home insecurity, child welfare, poverty, substance abuse, domestic abuse, sex/gender identity, immigration status, encounters with law enforcement, and travel infrastructure.

Our methodology involved a preliminary search for publicly available data, which uncovered a considerable amount of obscured information. To enhance accessibility, we recorded the raw datasets in our GitHub repository for publications that referenced publicly available data.

Additionally, for datasets that were not available in tabular format, we leveraged the Python library Tabula to extract tables and store them in our repository.

To facilitate collaboration and data exploration, we opted for an Excel spreadsheet format for storage, which includes a tagging system for efficient identification of relevant data sources. Through consultation with our sponsors at QSIDE, we expanded the scope of information recorded for each resource pertaining to human trafficking in Michigan, providing a brief description of each dataset and rationale for its inclusion.

This paper emphasizes the importance of obtaining a comprehensive understanding of the human trafficking problem through a horizontal view, which involves understanding how various risks interact with each other, either by location or other variables. Our project lays the groundwork for future research to have access to meaningful data to explore the interplay of these risks and their impact on human trafficking victims within Michigan.

Background and Motivation

Most American organizations acknowledge that human trafficking is a significant problem. However, measurements on the extent of the problem are inconsistent and/or not well documented in publications. Previous investigation into these methods resulted in the conclusion that most measures are just WAGs (Wildly Approximated Guesses). The purpose of this project is to explore potential sources of public data to get a true handle on the extent of the human trafficking problem.

The goal when working in this space is to minimize the number of people who are victimized by human trafficking. This is a difficult problem to solve, and often victims are identified on a case-by-case basis. Ideally, we would want there to be a dataset that has information on every instance of human trafficking in Michigan, however this does not exist. Because this is the first step of QSIDE addressing this problem in Michigan, we need to understand what data is out there and how it can all be related with human trafficking. QSIDE gave us a list of risks that all potentially relate to human trafficking. These risks include home insecurity, child welfare, poverty, substance abuse, domestic abuse, sex/gender Identity, immigration status, encounters with law enforcement, and travel infrastructure (Appendix section B discusses these in more detail). However, without data, there is not a really good place to start solving this problem. For our project we are laying the foundation for future projects to have access to meaningful data to begin to see how these risks play together and impact human trafficking victims within Michigan.

There have been many vertical studies and efforts to highlight individual risk factors. But there have not been many studies within Michigan that try to get a horizontal view of the problem. Having a deep understanding on individual risks like substance abuse or immigration in Michigan does not provide the whole picture on how these risks interact and how they impact human trafficking. This projects aim is to explore and document how these risks are all connected, either by location or another variable, to help generate a horizontal view and understanding of the human trafficking problem.

Methodology

During our preliminary search for data related to human trafficking, we encountered a considerable amount of obscured information. The findings reported in publications and journals were typically based on privately collected data or a handful of public datasets, such as the United States Census. For publications that referenced publicly available data, our team took the initiative to dive deeper and locate and record the raw datasets in our GitHub repository for accessibility. For datasets that were not obtainable in tabular format, we leveraged the Python library Tabula to extract tables from the publications and store them in our repository.

To streamline data storage and facilitate collaboration, our team opted to use an Excel spreadsheet. This decision was made with consideration to our project's mixed audience, consisting of data and social scientists. This format enables us to share our data without requiring users to have Python installed, while simultaneously supporting data exploration and analysis.

In consultation with our sponsors at QSIDE, we expanded the scope of information recorded for each resource pertaining to human trafficking in Michigan. We implemented a tagging system that facilitates quick and efficient identification of relevant data sources (e.g., "show me Scientific Journals related to Immigration"). Additionally, we provided a brief description of each dataset and rationale for its inclusion.
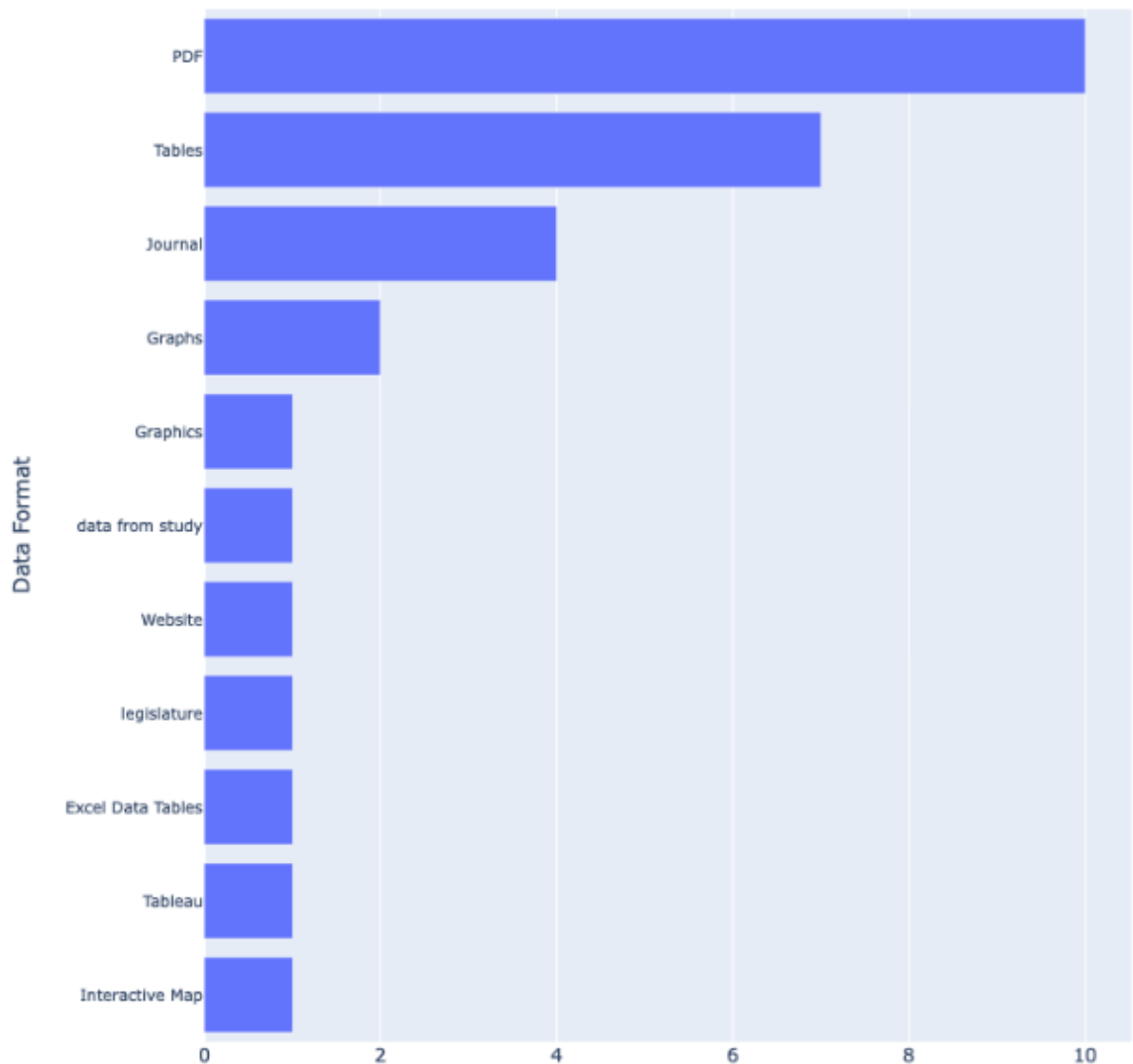
Results

The tangible product we produced was a data bibliography that contains a repository of data we found from the scraping different web links from the internet. We created an Excel sheet that contained headings and tags to categorize data we found. The headings included Title, Data Format, Github Reference, Original Data Source Link, Original Data Source, Article Link, Tags, Summary, and Relevance. Figure 1 below shows an example entry of the bibliography. More detailed explanations of the columns can be found in Appendix section A.

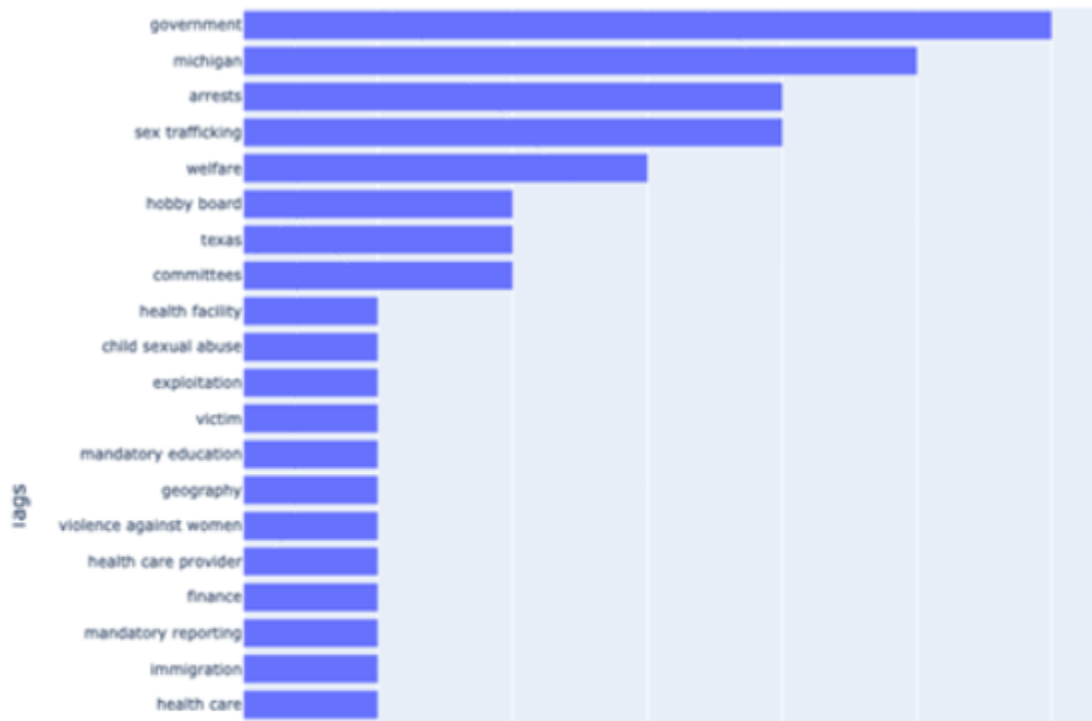| Title	Data Format	GitHub Reference	Original Data Source Link	Original Data Source	Article Link	Tags	Summary	Relevance
2017 Report to the Governor – Michigan Human Trafficking Commission	PDF, Tables	https://github.com/jacknuge28/QSIDE_Human_Trafficking/blob/main/Data/Raw_Data/001.001_ReportToGovernor_arrest_data.csv		Michigan Data Collection and Research Subcommittee	https://www.michigan.gov/ag/-/media/Project/Websites/AG/publications/2017/2017_Michigan_Human_Trafficking_Commission_Report.pdf?rev=fd8380535b6048fda23baf9495b54b5e&hash=08A0736C82785B6BAE49B6919C1B4F31	Government, Committees, Arrests	2017 Report from the Michigan Human Trafficking Commission on committees that are working with human trafficking. Has arrests made between Sept 1, 2016 through Nov 30, 2016 and if those arrests were suspected of being involved with human trafficking.	This resource contains information on various governmental committees around Michigan dealing with Human Trafficking. Contact names and department names for these prevention/prosecution resources can be found. Also includes arrest data for a small window in time relating to human trafficking.

Additionally, we did some initial analysis on the data bibliography we created. The figures below are the counts for the type of data formats present in the research and the counts for our self-implemented tagging system type (outside of the major factor tags).

A majority of the sources we have collected are in PDF format, and many contain tables within the document. This is largely due to the fact many of our sources are in the form of visualizations or publications for a wider audience, so the data has been presented in a more

visually appealing format compared to the raw data. Figure 2 shows a graph detailing the frequency of the most common data format types we saw in the data we collected.

With our self-implemented tagging system, users of our bibliography can add tags representing (at a glance) what the data source may reference. While currently the system allows for any tags, we hope to have a robust bank of common tags that can be used to help an end user be more efficient in their search for information that suits their needs. Figure 3 shows a horizontal bar plot detailing the occurrence of tags, helping the end user visualize the different sources included in the bibliography.

Concluding Discussion and Future Work

The main outcome of our project was to create a foundation that would be improved on in future.

For the next steps, our group came up with a couple ideas for what future groups could possibly work on. One of our ideas included analyzing eight risk factors for human trafficking, specifically in Michigan. These eight risk factors include: These risks include Home-Insecurity, Interactions with Child Welfare, Poverty, Substance Abuse (Exposure or Personal), Abuse/Domestic Violence, Gender/Sexual Identification, Immigration Status, Interaction with Law Enforcement and Travel. This could help us understand how these factors may contribute to the problem and identify ways to improve the current conditions of Human Trafficking in Michigan. Our main question would be to research how these risks can increase/decrease the likelihood of being trafficked. Are there risks that outrank others? How can public organizations use this information to help lower the number of at-risk people?

The use of visualizations and data models, such as modeling the likelihood of human trafficking, can be difficult to do as there is always the possible threat of putting at-risk individuals at harm. Models created would have to be kept private within the organizations. Working with QSIDE, and seeing if they themselves have data, or connections with other organizations that may have data, would be a great start to producing informative models. We also acknowledged that creating an accurate model would take perfect data, which we are aware may be difficult to find. If the obstacle of not having enough data exists, the next step would be to compare the relationships between the risk factors using the instances of trafficking, to further explore how human trafficking can be combated. For example, if we were to collect data on substance abuse and poverty, we would then try to identify any overlaps and see if there are connections in previously written articles or papers related to human trafficking.

In conclusion, we began this project very aware that we were not going to solve the issue of Human Trafficking. Instead, we wanted to build a project that would serve as the starting point that we did not have. Our work is laying the groundwork on the data analytic angle of human trafficking. We hope that our work will be insightful and contribute for future groups who will continue to combat the ongoing issue of Human Trafficking in Michigan, and eventually the rest of the country.

Acknowledgements

Resources

A copy of our updated data bibliography can be found at the following link: (https://github.com/jacknuge28/QSIDE_HumanTrafficking/blob/main/Data_Bibliography.xlsx (https://github.com/jacknuge28/QSIDE_HumanTrafficking/blob/main/Data_Bibliography.xlsx))

Our full repository of work can be found at the following link: (https://github.com/jacknuge28/QSIDE_HumanTrafficking (https://github.com/jacknuge28/QSIDE_HumanTrafficking))

References

The following references were used to provide a comprehensive background of the task at hand. These resources helped to provide insight into the research question, and to identify existing gaps in current publications.

[1] Aghazarm, C., & Laczko, F. (2008). Human trafficking: New directions for research. Geneva International Organization for Migration.

[2] Albright, E., & D'Adamo, K. (2017). The media and human trafficking: A discussion and critique of the dominant narrative. Human trafficking is a public health issue: A paradigm expansion in the United States, 363-378.

[3] Anderson, V. R., Kulig, T. C., Sullivan, C. J., & Farrell, A. (2019). Estimating the prevalence of human trafficking in Ohio.

[4] Cunningham, S., DeAngelo, G., & Tripp, J. (2019). Craigslist reduced violence against women. Baylor University.

[5] Denton, E. (2016). Anatomy of offending: Human trafficking in the United States, 2006–2011. Journal of Human Trafficking, 2(1), 32-62.

[6] Dimas, G. L., Konrad, R. A., Lee Maass, K., & Trapp, A. C. (2022). Operations research and analytics to combat human trafficking: A systematic review of academic literature. PLoS one, 17(8), e0273708.

[7] Dubrawski, A., Miller, K., Barnes, M., Boecking, B., & Kennedy, E. (2015). Leveraging publicly available data to discern patterns of human-trafficking activity. Journal of Human Trafficking, 1(1), 65-85.

[8] Freeman, N. K., Keskin, B. B., & Bott, G. J. (2022). Collaborating with Local and Federal Law Enforcement for Disrupting Sex Trafficking Networks. INFORMS Journal on Applied Analytics, 52(5), 446-459.

[9] Konrad, R. A., Maass, K. L., Dimas, G. L., & Trapp, A. C. (2022). Perspectives on How To Conduct Responsible Anti-Human Trafficking Research in Operations and Analytics. European Journal of Operational Research.

[10] National Academies of Sciences, Engineering, and Medicine (2019). Estimating the Prevalence of Human Trafficking in the United States: Considerations and Complexities: Proceedings of a Workshop in Brief. Washington, DC: The National Academies Press. https://doi.org/10.17226/25550 (https://doi.org/10.17226/25550).

[11] Owens, C., Dank, M., Breaux, J., Bañuelos, I., Farrell, A., Pfeffer, R., ... & McDevitt, J. (2014). Understanding the organization, operation, and victimization process of labor trafficking in the United States. Washington, DC: Urban Institute.

[12] Sweileh, W. M. (2018). Research trends on human trafficking: A bibliometric analysis using

Appendix

A. Data Bibliography Column Descriptions

Column Name	Description
Title	Identifying name of the data source/paper.
Data Format	What format is the data source or publication stored in? Options include PDF, tables, graphs, graphics, etc.
GitHub Reference	Direct link to the data source (if available) in the QSIDE Human Trafficking project repository.
Original Data Source Link	Link to the original source of the raw data.
Original Data Source	Title of the original data source/paper.
Article Link	Link to the original data source/paper.
Tags	User-generated tags about the data denoting keywords about the source. Helpful for searching and for ease-of-reference.
Summary	A summary of the data source/publication and the contents inside.
Relevance	How does this source relate to the broader human trafficking problem? Why was it included?

B. Factor Tag Descriptions

Factor Tag	Description
Poverty	State or condition in which one lacks the financial resources and essentials for a certain standard of living.
Child Welfare	Involvement in the system of programs and services meant to protect the well-being of children who may be at risk of abuse, neglect, or exploitation.
Domestic Violence	Physical abuse of a household member, especially one's spouse or domestic partner.
Gender / Sexual Identity	Sex : person's biological status and is typically assigned at birth, usually on the basis of external anatomy. Sex is typically categorized as male, female or intersex. Gender : social construct of norms, behaviors and roles that varies between societies and over time. Gender is often categorized as male, female or nonbinary.
Substance Abuse	The use of an illegal or unprescribed drug, or the inappropriate habitual use of another drug or alcohol, especially when resulting in addiction.
Immigration	Travel into a country for the purpose of permanent residence.
Encounters with Law enforcement	Encounters with police or other law enforcement officials. Severity may range from traffic stops to home visits, and different intensity/frequencies may be related to a household's susceptibility to trafficking.
Home Insecuity	Housing insecurity is the lack of security in an individual shelter that is the result of high housing costs relative to income, poor housing quality, unstable neighborhoods, overcrowding, and, but may not include, homelessness.
Travel Infrastruture	Points of travel (highways, airports, seaports, etc.) can be related to human trafficking as it provides both potential victims as well as routes for transporting them across state/national borders.