Documenters – Leadership Dossiers

By Annie Wilcox (wilcox63@msu.edu), Minh Anh Dang (danganh1@msu.edu), Ryan Flanery (flaneryr@msu.edu), Ryan Hanks (hanksrya@msu.edu), Samuel Corder (cordersa@msu.edu)

Abstract

The goal of this project is twofold: Create an automated process to collect a plethora of information for a set of City of Detroit appointed officials, as well as to create a user-friendly interface where this data can be displayed.

To collect our data, we utilized web-scraping technology, as well as a variety of API's. Between the five of us, we split up finding social media handles, phone numbers, emails, campaign finance data, and nonprofit organization affiliations for all the provided appointed officials. Regarding the collection of our data, the results were mixed. Some officials had plenty of information available, while some weren't as public.

In order to create a user-friendly interface, we compiled all our data into a single spreadsheet using Google Sheets[1]. Then, we created a website using GitHub Pages[2] that pulls data from our Google Sheet. We have made several adjustments to the interface of the website to make it more user-friendly.

The result of these processes is an intuitive website where anyone can search for a City of Detroit appointed official and find out more information about them. We have also created automated processes, meaning much of this project can be adapted for use on different officials from a different city.

Background and Motivation

In today's political climate, there is a great lack of transparency. Many people don't understand who is making the decisions that affect their lives. Despite this, we believe that most people are

interested in the truth. This is where Documenters[3] comes in. Documenters is a nonprofit organization that focuses on trying to make local government more transparent. They send members of the community to public meetings to record and share everything that happened. This information is then given to journalists, who can report more accurately. Our goal as data scientists is to create a new tool supporting initial research for Documenters in the form of a search engine interface that allows journalists and the interested public alike to access information on their locally appointed officials, some of which includes names, positions, contact information, various social media handles, and campaign donations.

Methodology

From the start, we were provided with a list of 368 individuals, along with their department/district, and board/commission/department. When creating the back end of the project, i.e. the collection and storage of the data, we had to adapt our methods for to the wide variety of data types and availability. For the most part, we focused on utilizing API's and building relationships with their publishers, but in other cases, web-scraping and manual data collection were a better fit.

The first type of information we tried to find was contact information, namely phone number and email, we found a handy resource in the City of Detroit Directory Smartsheet[4]. Using Pandas, a Python library specializing in data structure storage within Python, we converted the data into a CSV file that could easily be merged with our own list of names, populating the dataset with contact information for each official. One roadblock we ran into is that some of this contact information is for the department the individual belongs to, not their personal contact information. We decided to leave this as is after agreeing that users would most likely still be able to reach the officials of interest with the given contact information, as well as possible ethical concerns with seeking out individual contact information. We also found a City of Detroit website[5] that contained the phone numbers and emails for every board. Since there were only a handful, we manually collected this information.

Collecting social media handles presented a unique challenge, as each platform has varying regulations on how much of their data is available. We started off by trying to automate data collection via web-scraping but found that scraping tends to contradict with most social media platforms' terms of use. Instead, we decided to focus on trying to obtain API access to X, LinkedIn, and Meta via email and our sponsor, David Palmer's connections. For the LinkedIn API, we started with a premium subscription provided by David but found that the website restricts most profile information (usernames) behind higher tiers of the subscription service. For this, and for a lot of social media information, we were forced to defer to the potential future team who might have greater access to these resources. To see a financial report that outlines the cost and use case for these sections of the data collection, see Appendix C. For demonstration purposes, we collected a selection of LinkedIn handles manually. For Twitter and Instagram profile information, we experimented with Grok, X's AI implementation. There was some success with this, but the data needed to be manually added to the dataset, and the information was only provided for a small fraction of the officials. There were roadblocks in regard to the verification of this information, as many people share the same name, or have both personal and business accounts. Grok also has access to Meta's data, so it proved beneficial as a resource to locate some Instagram handles, but the information still needed to be added to the dataset manually.

To access campaign finance data for these officials, we turned to the Wayne County Finance System[6]. Web-scraping was available as an option, and so we turned to using Selenium WebDriver - a web framework that permits you to execute cross-browser tests - to automate the process of inputting the name of the committee into the system, searching, then extracting the available tabular data. However, the loading time of the website is relatively long, and the program can expectedly cost up to 5 hours of computation time to search for all 368 names in the original dataset. It is possible to optimize this process via the MSU HPCC (High Performance Computing Center); however, we were limited by the project timeline. The contribution records dataset was used to generate pie charts representing donations made by certain individuals and is separated by ZIP code into multiple ones. Since people can share the same name, we want to make sure that no chart has the risk of representing multiple people's donations at once. As such, we separated them by ZIP codes, and we assumed that people which live in different ZIP codes are different individuals. The pie charts are made using Matplotlib and Plotly – which are open-source graphing Python libraries – to enable interactive hovering. The code is available via GitHub for any future teams to optimize for the best results. While we weren't able to fully implement the campaign finance information into the front end, for proof of concept for how this information could be implemented onto the website, see Appendix F.

In terms of nonprofit affiliation, we wanted to search for what positions these officials hold within nonprofit organizations as well as what type of compensation they're receiving. We turned to Candid[7], an organization that stores data about nonprofits. We attempted to use their flagship product GuideStar, an API that allows users to pull this data by searching the name of an official. They store I-990 tax forms, which contain information on charitable contributions made by people. Luckily, Michigan State University provides its students with a subscription to Candid, however, we only had access to the lowest tier of the subscription, which does not provide API access. For this reason, we collected information for a handful of individuals manually. Due to time constraints, we were not able to implement this information onto our website, but it can be found in Appendix E. Additionally, we reached out to Candid to see if they would provide us with API access, but unfortunately, the communication was still in progress at the time of project "completion". The outline of costs can be found in Appendix C.

For the front end, we started with a search engine template in a format that combined HTML, CSS, and JavaScript. We focused on learning the syntax of each of these languages in order to connect our data to the search engine via Google Sheets API. Google Sheets API is a free service that allows a user to generate an access key which they can embed in their code, giving free access to any public Google Sheet[8] via HTML. This allowed us to update our website with ease due to Google Sheets automatic save feature. Every time a user searches for an official on our website, a new call is made with the most up to date data. Later on, we decided to transfer our website from local files to GitHub pages, in order to improve accessibility. After connecting both ends of the project, we focused on simplifying and improving the UI. These changes included basic formatting - such as font size and type, background images, and text readability – as well as simplicity of filtering to streamline searching for users who might not be looking for a specific official, or want information on a whole department.

Results

From the provided datasheet, we created/used methods to collect emails, phone numbers, and X/Instagram profiles. We also began work on automatically collecting LinkedIn handles, campaign finance data, and nonprofit affiliation/compensation. Our final data set is saved in a Google Sheet that is the basis for our search engine. Using our website, anyone interested can learn more about their Detroit locally appointed officials.

Concluding Discussion and Future Work

There were a few issues we ran into during this project. One issue is regarding the amount of data we could find, as these are appointed officials, not elected ones. Many of these individuals don't have a massive digital footprint, and many of their social media accounts are personal. Their campaign information is not nearly as publicized as their elected peers, and it was often layered deeply into various navigation pages as bureaucratic websites are not often known for their agreeable user interface. In addition to some data not being readily available, there were also possible ethical concerns for the collection process. For example, there were a few Instagram accounts that were private, and we weren't sure if something like that should be published. The team worked under the guidance of our community partner, David Palmer, to be sure not to inappropriately invade anyone's privacy. Another issue we had is that we currently have no way of automatically updating the data to be current. If someone changes their X handle, then our website would be inaccurate.

Overall, we have created a website that successfully returns information on Detroit's elected officials in a simpler and more accessible format than was originally available. By nature, this project will never be truly completed. There will always be room to expand on the individuals in the system, as well as the information collected on them. Future teams picking this up could add to the areas covered, and they could search for more information about the individuals already researched. We were tasked with supporting initial research for Documenters to inform them on the ways their mission might be supplemented by the collection of other information, and ultimately, we believe we have created a product that is easy to expand upon, as well as adaptable for each specific area of interest.

In the future, teams will have a variety of choices for starting points on the expansion of this project. With the provided code, they could choose to focus on the implementation of fully automated data collection, or they could choose to focus on client outreach by remaining in contact with the organizations we have provided as references.

Acknowledgements

- A special thanks to David Palmer, our community partner, for guiding us in ethical data collection, as well as providing us with many connections to further aid our research.
- Thanks to Documenters.org for supporting our initial research via David and giving us the opportunity to present our final product.
- Thank you to Dr. Dirk Colbry for helping us navigate the professional world as well as providing us with these initial connections.

References

1. Google. (2025, March 21). *Google Sheets API Overview | Google for Developers.* https://developers.google.com/workspace/sheets/api/guides/concepts
2. GitHub Pages. (2025). https://pages.github.com/
3. City Bureau. (n.d.). *Making local government more accountable*. Documenters.org. https://www.documenters.org/
4. *City of Detroit Directory of Depts.* (n.d.). https://publish.smartsheet.com/9def816c9e6a4a4395d2903039bf714d
5. City of Detroit. (2025). *Boards*. City of Detroit. https://detroitmi.gov/government/boards
6. Wayne County Finance System (n.d.). https://wccampaignfinance.com
7. *Candid | foundation center and GuideStar are now candid*. Candid. (2025). https://candid.org/
8. Corder, S. (2025, April 16). *DetroitDF*. Google Sheets. https://docs.google.com/spreadsheets/d/1x6aJpjxxsb_-TWj-42QgpOGXleo8rbemAy95CrEmyI4/edit?gid=0#gid=0
9. Wilcox, A., Dang, M. A., Hanks, R., Flanery, R., & Corder, S. (2025, March). Find a Detroiter. https://minh-msu.github.io/Detroit_Leadership_Dossiers/

Appendices

Appendix A: Deliverables

Here are all our deliverables, as well as a short description of each:

- **2025_02_02_Documenters_Plan.pdf:** a copy of our original plan for the project
- **Project_Plan_Video:** a folder containing a storyboard and video for our project plan

- **Outreach_Assignment:** a folder containing the blog post and poster for our Outreach Assignment
- **MVP_Video:** a folder containing the storyboard and video for our MVP video
- **Final_Presentation_and_Final_Video:** a folder containing the storyboard and video for our final video, as well as a script for our in-person presentation
- **FinalReport.docx:** our final report containing all our findings
- **WidgetDemo.url:** a link to our website where our data can be displayed

Appendix B: Installation / Reproducibility Instructions

The full instructions can be found here. The process involves creating your own copy of the project, using the web scraper, creating a Google Sheets API account (for free), and updating the website code to reflect your data.

Appendix C: Financial Report

There are many components of this project that require a paid subscription. Additionally, we are going to meet with members of Outlier Media and the City Bureau and present this project to them and see if it's something they would like to use. For these reasons, we decided to compile all of the paid services into one easy to use table.

Below is a table showing the prices for all the services to reproduce our results.

Service Name	Description	Annual Cost	Link	Purpose
Candid - Nonprofit search API	Access to data from 4.9 million nonprofit organizations	$4800	https://candid.org/use-our-data/apis	Allows user to collect nonprofit positions and compensation for anyone. This can be used to increase transparency by showing the public who is supporting an organization, and if there is money involved.
LinkedIn People Profile API	Provides access to LinkedIn profile data, including name, work	Custom pricing, only available via consult ~$708	https://developer.linkedin.com/content/developer/global/en_us/index/partner-programs/apply	Allows the user to collect LinkedIn handles. This can be used to provide more contact information about a particular person. However, the LinkedIn APIs are notoriously hard to access.

	history, education, etc.			
X API Pro	Pull live profile and post data	$60,000	Twitter API Documentation | Docs | Twitter Developer Platform	Allows the user to collect profile and post data about anyone. This could be used to display Twitter handle, as well as popular Tweets, potentially showing what the individual stands for.
Total Annual Cost	-	$65,508	-	-

Appendix D: Manually Collected Data

For a variety of reasons, there were instances where we manually collected data. To ensure that our results are reproducible and to avoid confusion, we have created a table showing what data was manually collected.

Below is a table showing the data we collected manually.

Data Type	Link to Source	Why We Collected It Manually
Board/Department Email	https://detroitmi.gov/government/boards	We came across the website by chance. Since there was only a handful of emails, none of which were for individuals, we decided to grab it by hand.
Board/Department Phone	https://detroitmi.gov/government/boards	We came across the website by chance. Since there was only a handful of phone numbers, none of which were for individuals, we decided to grab it by hand.

LinkedIn	https://www.linkedin.com/feed/	We tried a variety of methods for collecting LinkedIn handles. Along the way, we found a selection of handles manually.
Nonprofit	https://www.guidestar.org/search	We never got access to the Candid API, which would allow us to automate the process of finding this information. For demo purposes, we decided to find this information for a selection of people manually.

Appendix E: Nonprofit Data

Although we didn't get access to the Candid API, we were able to manually search for this data on Candid's website. As MSU students, while we have access to GuideStar Pro, we do not have access to the Candid API. We manually collected this information for a handful of individuals to show what we could do in the future with a Candid API subscription.

Below is a table showing some individuals and their nonprofit affiliations.

Name	**Organization**	**Position**	**Role**	**Location**	**Compensation**	**Source & Fiscal Year**
Mike Duggan	Detroit Regional Partnership	Director	Independent Trustee	Detroit, MI 48226	N/A	Form 990 2023
Mike Duggan	Detroit Regional Partnership Foundation	Director	Independent Trustee	Detroit, MI 48826	N/A	Form 990 2023
Mike Duggan	Detroit Economic Growth Association	Director	Independent Trustee	Detroit, MI 48826	N/A	Form 990 2023
Mike Duggan	Detroit Economic Growth Corp	Director	Independent Trustee	Detroit, MI 48826	N/A	Form 990 2023
Mike Duggan	Metropolitan Affairs Coalition	Director	Independent Trustee	Detroit, MI 48826	N/A	Form 990 2023
Mike Duggan	Voices of Detroit Initiative	Board Member	Independent Trustee	Detroit, MI 48201	N/A	Form 990 2013

Mike Duggan	Detroit Economic Club	Director	Independent Trustee	Detroit, MI 48826	N/A	Form 990 2023
Mary Sheffield	Detroit Economic Growth Association	Director	Independent Trustee	Detroit, MI 48226	N/A	Form 990 2023
Mary Sheffield	Detroit Economic Growth Corp	Director	Independent Trustee	Detroit, MI 48226	N/A	Form 990 2023
Mary Sheffield	Detroit Institute for Law and Organizing	Director	Independent Trustee, Officer	Detroit, MI 48207	N/A	Form 990 2023
Mary Sheffield	Detroit Riverfront Conservancy Inc	Director	Independent Trustee	Detroit, MI 48243	N/A	Form 990 2023
Scott Benson	Tour De Troit	Member	Independent Trustee	Detroit, MI 48201	N/A	Form 990 2023
Scott Benson	International Brotherhood of Electrical Workers (aka 0131 Local Union)	Treasurer	Officer	Kalamazoo, MI 49001	$9,428 + $4,204 (Other Compensation) = $13,632	Form 990 2024
Scott Benson	Detroit Greenways Coalition	N/A	N/A	Detroit, MI 48232	N/A	Updated by the nonprofit in 2024
Anthony Bria	Detroit Regional Partnership	Director	Independent Trustee	Detroit, MI 48226	N/A	Form 990 2023

Appendix F: Campaign Finance Information

Due to time constraints, we couldn't complete and implement the campaign finance information on our website. However, we wanted to provide proof of concept. Below are a handful of pie charts representing donations made by Mary Sheffield and Scott Benson. The data shows what campaigns this individual has donated to, and it is separated by ZIP code into multiple pie charts. The final figures have hover texts that will display additional information such as the transaction date, contribution type, contributor type, and contribution amount. All of these cannot be displayed using static image, as such, there's one example to what it looks like. We separated them by ZIP code because there are many people who may have the same name. Separating them by ZIP code is a first step in verifying that this information is true. Although some people may have moved, or have multiple addresses, there is a possibility that they are someone else entirely.

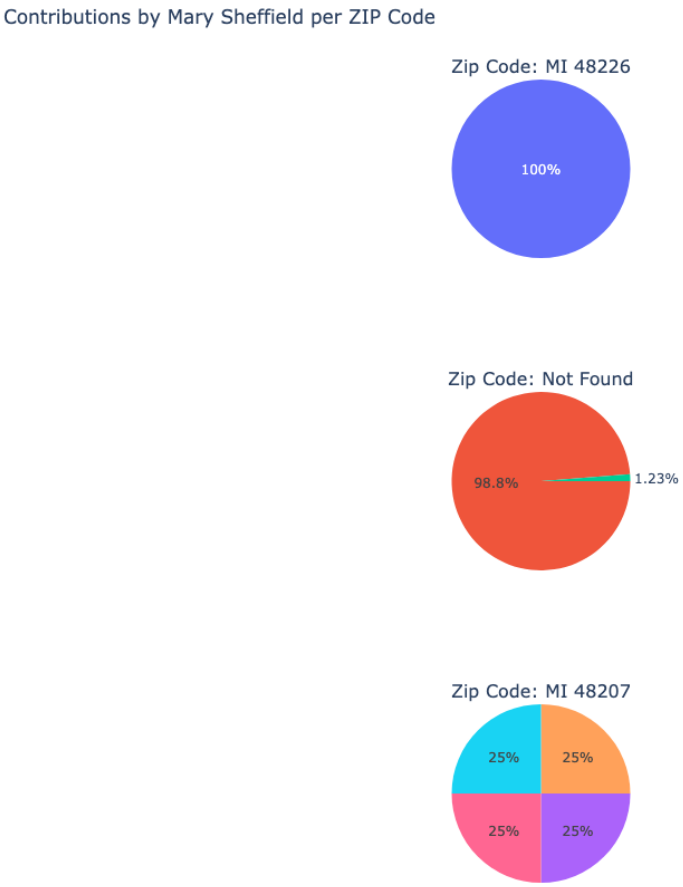

Figure 1. Contributions Record of Mary Sheffield per ZIP Code

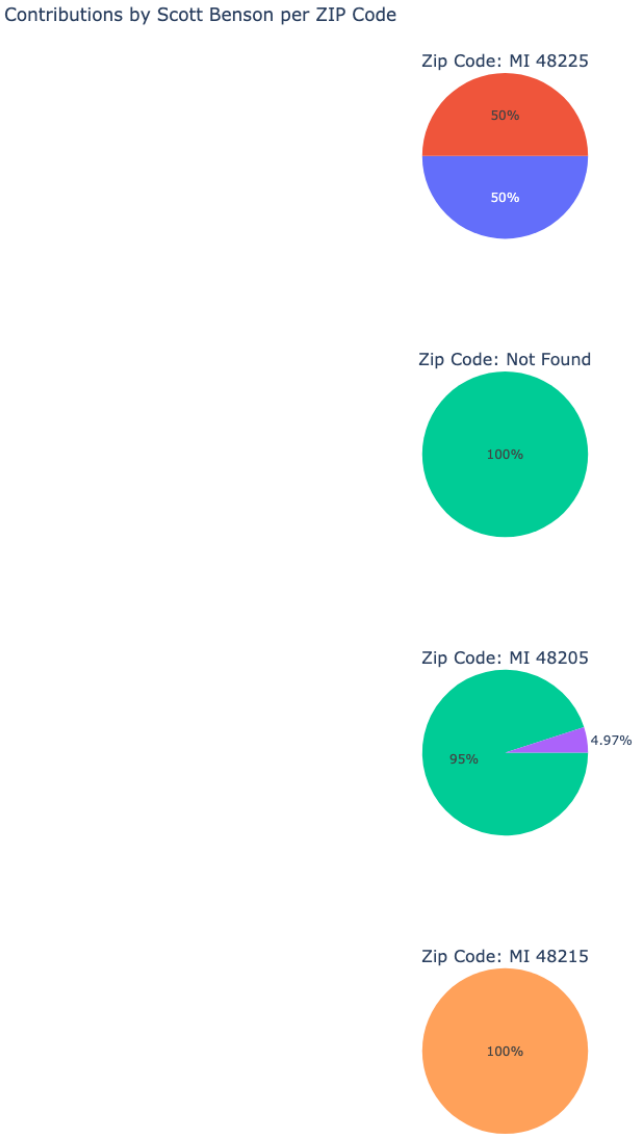

Figure 2. Contributions Record of Scott Benson per ZIP Code

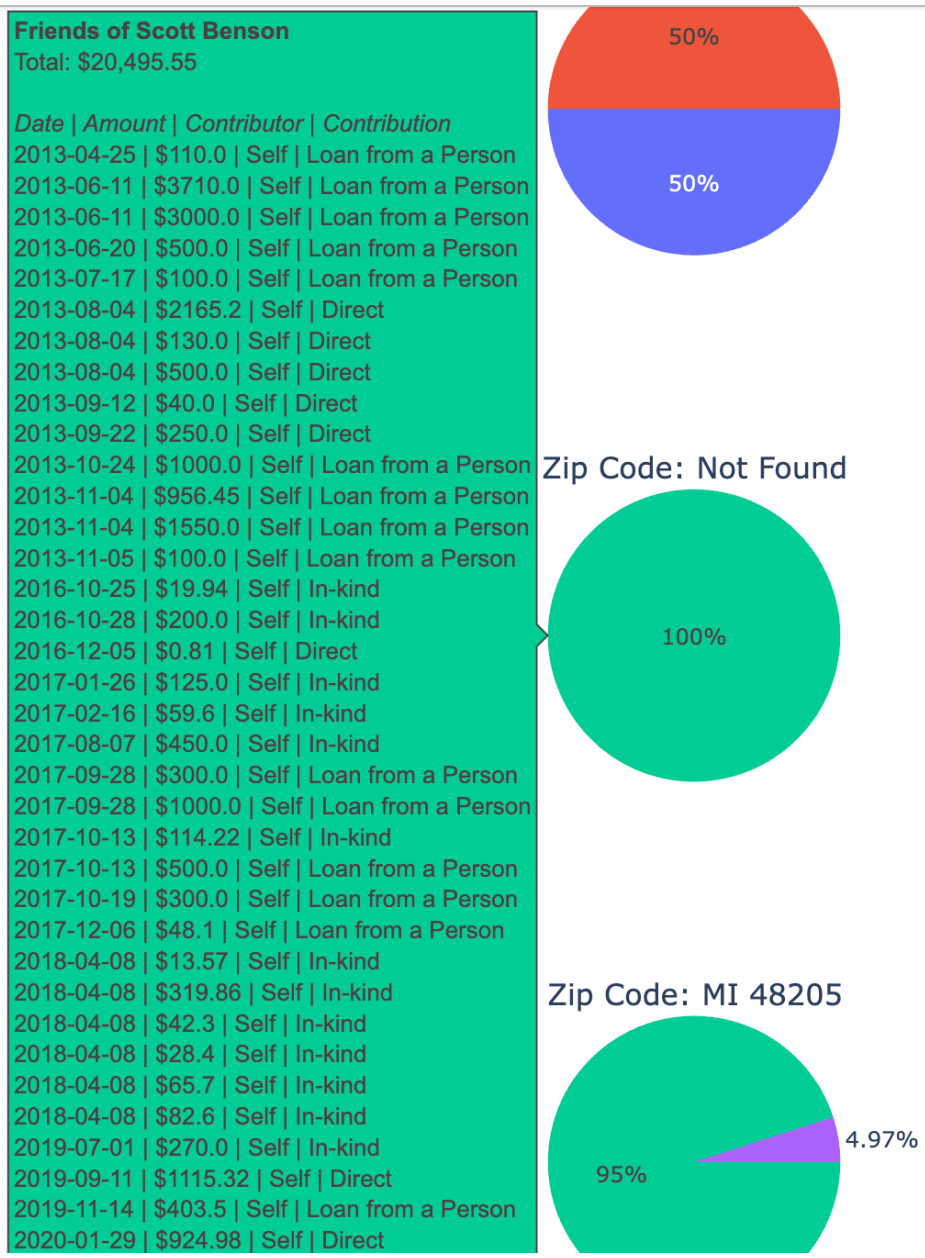

Figure 3. Example of Hover Text in the Final Figure (taken from **Figure 2**)