Senator Markey's Earmarks - Team D

Zuizz Saeed (Class of 2025) zuizzms@bu.edu

Jeremy Bui (Class of 2025) jbui02@bu.edu

Rithvik Nakirikanti (Class of 2025) rithvikn@bu.edu

Grace Elias (Class of 2025) gelias@bu.edu

Jiang Cheng Liu (Class of 2025) jiangcl@bu.edu

Looker Studio Dashboard

Trello

GitHub

Introduction

Federal budget earmarks, officially referred to as Congressionally Directed Spending funding (CDS), represent a crucial mechanism through which Congressional representatives allocate discretionary funds directly to their district or state for specific projects. This process plays a significant role in shaping local landscapes and communities—whether it is funding local community projects, infrastructure developments, or any other targeted initiatives. FY22 was the first year of CDS funding after a decade without earmarks.

Given the importance of these funding allocations, it was imperative that we carefully analyzed the process of evaluating CDS requests through an equity lens, ensuring that the distribution of resources is fair and inclusive. This project, under the advisory of Senator Edward J. Markey, seeks to develop a transparent methodology to assess the distribution and effects of CDS funding.

At a big-picture level, our goal has been to analyze the allocation requests funded by Senator Markey and determine which demographics (if any) are being underserved. After doing this, we can further examine the data to best determine if any implicit biases may be favoring certain demographics over others. This is an important task and one that many senators and other legislators should take on, to ensure that they are most fairly representing their constituents. They must work to improve the area that they serve as a whole and that no subsection is forgotten.

To achieve our goal, we worked with data from 2022, 2023, and 2024 to seek out trends in which demographics and subcommittees have been underserved. Ultimately, we came up with a few suggestions for Senator Markey as to how he might better spread out his CDS funding.

Base Analysis

To begin this project, we spent a lot of time working with and organizing our data. We were given spreadsheets with data regarding allocation spending in 2022 and 2023. The datasets that we had access to included columns for the funded cities, project names, and the amount they had received from the years 2022 and 2023. One of the very first objectives was to obtain the census tract data and then analyze the demographic populations of each city to depict whether or not the BIPOC populations are being benefited. We first used Google Maps API to retrieve the latitude and longitude of cities and then fed that into the census tract API to retrieve that city's census tract ID and county number. This was used further with a script that uses another API endpoint from the same API to get the counts of Whites, African Americans, Asians, Alaskan Natives, Native Americans, Pacific Islanders, and Other races.

We were also able to look at PDF files with the requests made for 2024. Using Tabula, we converted the data from the PDFs into data frames and then exported them as spreadsheets. We then similarly used Google Maps API to retrieve the geodata and demographic data for the 2024 requests.

Once we had our data organized for the base analysis, we were able to begin considering certain questions. What demographics have been underrepresented thus far in terms of receiving allocation funding? Which demographics have been largely benefitting from having their requests funded? How are trends changing over time? What other relevant patterns or gaps are visible that may point to inequities for different populations?

In our analysis, we employed various graphs to determine which demographic groups have been over or underserved by allocation funding. Before delving into the demographic breakdown of served cities, we first examined the allocation requests across different

subcommittees. These subcommittees include Energy and Water Development, Agriculture, Commerce, Justice, Science, Financial Services, Homeland Security, Interior, Environment, Transportation, Housing and Urban Development, Labor, Health, Human Services, Education, and others. Aligning these subcommittees with Senator Markey's stated priorities—such as jobs and the economy, energy and climate change, health and medical research, immigration, gun violence, LGBTQ+ rights, homeland security, foreign relations, telecommunications, internet privacy, and education—allowed us to assess the degree of alignment between funded requests and Senator Markey's agenda. This comprehensive approach provided valuable insights into the distribution and alignment of allocation funding with key policy objectives.

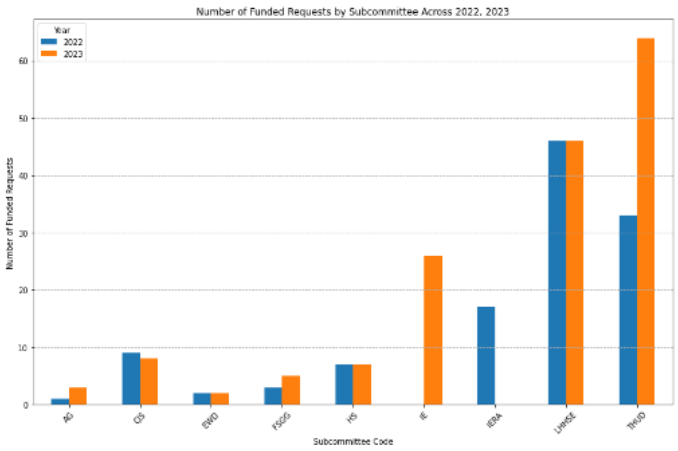
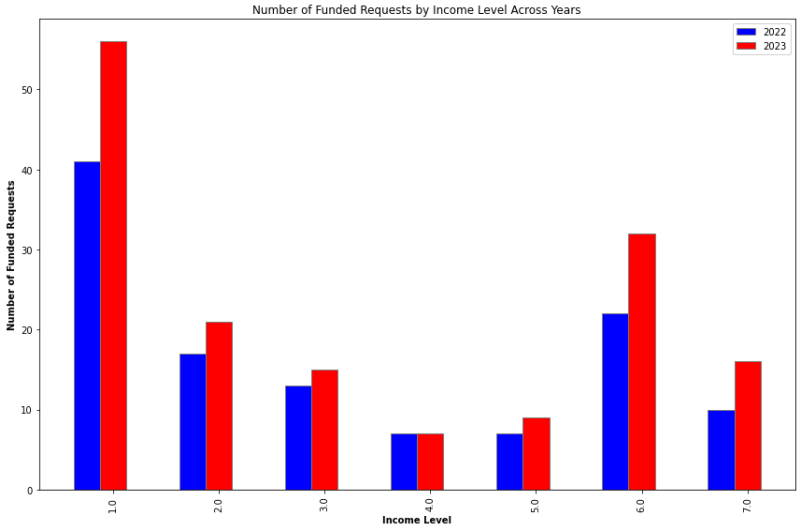


Figure 1: Number of funded requests by Subcommittee Code

Looking at the chart with our data, it was interesting to note that the largest amount of approved funding requests in 2023 went to the Transportation, Housing and Urban Development, and Related Agencies subcommittee. This committee does not strongly align with any of Senator Markey's priorities. High results from Labor, Health and Human Services, Education, and Related Agencies aligned with our expectations, as well as followed up by Interior, Environment, and Related Agencies and Commerce, Justice, Science, and Related Agencies. It appears that all subcommittees not including Transportation, Housing and Urban Development, Related Agencies, Agriculture, and potentially Financial Services and General Government align very closely with Senator Markey's goals. So we would have expected these three to have received the least amount of funded

requests. Since this was not the case for Transportation, Housing and Urban Development, and Related Agencies, it would be interesting to know whether they submitted an abnormally large amount of funding requests to the point that they could not be ignored or if each of their requests were for very little money. If neither of these was the case, this may be an area Senator Markey should look into funding less. Some of the money that had previously gone to it might be better off being put towards Energy and Water Development, and Related Agencies.

One way to categorize the cities of Massachusetts is by income (per capita) level. As explained in the table below, each level represents a certain range of income (per capita). Using our geodata, we were able to determine the income (per capita) of each city that has made a request, assign them to their respective level, and create a chart showing the amount of funded requests per income (per capita) level.



Income (per capita):	
7	$65k +
6	$55-65k
5	$50-55k
4	$45-50k
3	$40-45k
2	$30-40k
1	$30k -

Figure 2: Number of funded requests by income level and income range index

In 2022 and 2023, we noticed a trend where funded requests were predominantly coming from cities with either high or low incomes per capita. Interestingly, cities with incomes falling in the middle range didn't receive as many funded requests. This observation sparks curiosity about the

submission patterns across different income levels. It prompts the question of whether all income brackets are equally active in submitting requests, or if the disparity in funded requests is influenced by varying submission rates. Understanding the submission dynamics across different income levels will shed light on the factors driving resource allocation and whether there are any underlying biases or barriers in the process.

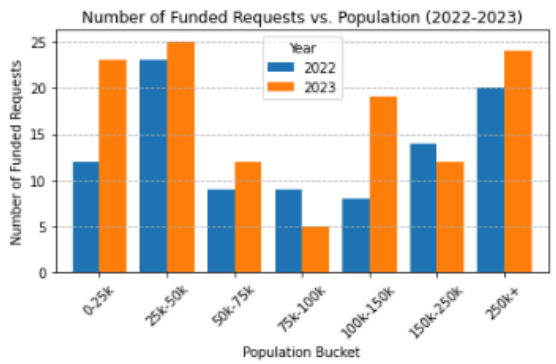


Figure 3: Number of Funded Requests vs. Population

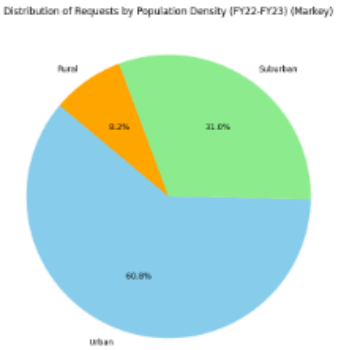


Figure 4: Distribution of Requests by Population Density

Furthermore, we can delve deeper into the analysis by examining population density as a factor. By assessing population density in conjunction with urban, suburban, and rural categorizations, we can gain a more nuanced understanding of the disparities observed in earmark requests. This approach allows us to explore whether population density influences the level of representation within each geographical setting and provides additional insights into the equitable distribution of earmark resources. Additionally, considering the observation from the chart that funded requests were most common among cities with either high or low incomes (per capita), it would be valuable to investigate if all income levels were submitting requests at equal levels, or if those who saw the most funded requests had sent the most requests to begin with.

Furthermore, to visualize how the funds are distributed in regards to cities and their BIPOC percentage, we created a heat map:

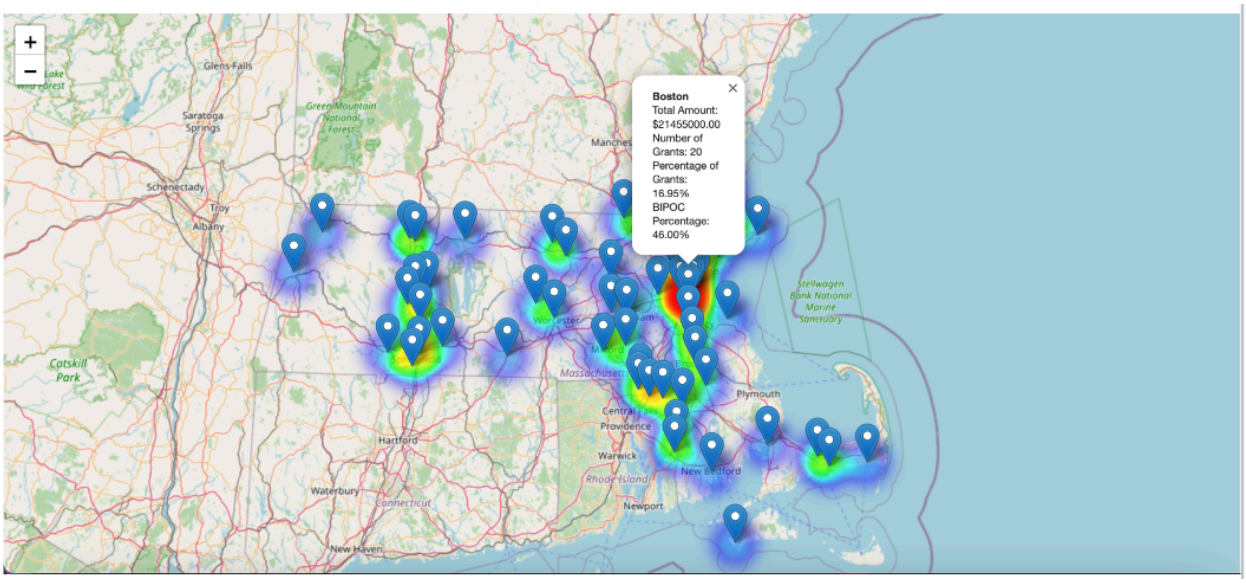


Figure 5: BIPOC heatmap for requests across MA (FY 2022-2024)

We can see from the heatmap that Boston received far more funding than any other city in Massachusetts, which seems to be promising since Boston has the largest percentage of BIPOC in Massachusetts. One of the issues that we thought of was that cities like Boston that have larger populations would lead to more funds being allocated towards those locations, and we tried to mitigate that by giving more weight to the BIPOC percentage. With this, one can see that while the funds are distributed fairly well across Massachusetts, Boston is clearly the epicenter with its large heat value.

Regarding the changing trends over time, the analysis reveals intriguing shifts in earmark funding patterns. Specifically, for the year 2024 (Figure X), there appears to be less pressure on lower-income areas, with funding being more evenly distributed across different income levels compared to previous years. This observation suggests a potential trend towards greater equity in earmark allocation, where resources are allocated based on project merit rather than income level alone. However, it's important to note that our analysis is limited to a three-year timeframe, and thus, further longitudinal studies would be necessary to fully understand how these trends

evolve. Nonetheless, these initial insights provide valuable groundwork for future research and policymaking efforts aimed at promoting fair and inclusive distribution of earmark resources.

In conducting our base analysis, we organized and analyzed data spanning two years (2022-2023) concerning allocation spending. Leveraging spreadsheets and employing APIs, we analyzed demographic populations to assess the representation of BIPOC communities benefiting from earmark requests. Our examination also involved scrutinizing subcommittees handling allocation requests, revealing intriguing disparities between funding distribution and Senator Markey's stated priorities. Notably, we observed a concentration of funded requests among cities with high or low incomes (per capita), prompting further inquiry into submission rates across income levels. Additionally, a comparative analysis of funding trends over time highlighted a potential shift towards more equitable distribution in 2024, particularly in addressing disparities based on income levels. Nonetheless, limitations in our analysis time frame necessitate further longitudinal studies to fully grasp trends. These findings lay a robust foundation for future research and policymaking endeavors aimed at the fair and inclusive distribution of earmark resources.

After carefully studying our initial findings, we were excited to dig deeper into our project extension. We wanted to know if the data we looked at truly pointed to unfair treatment of certain cities and groups, or if some cities just hadn't asked for much in the first place. This next phase was all about finding out what was going on behind the numbers. We wanted to understand why some places seemed to get fewer approvals, whether it was because they weren't asking for much or if there were other reasons at play. We aimed to get to the bottom of how resources were being distributed and to figure out if any issues needed addressing to make things fairer for everyone involved.

Extension Analysis

Our extension project aimed to give Senator Markey an interactive dashboard that looks at the requests and approval funding between 2022, 2023, and 2024 to look at the distribution of the allocated funds. It was set up with different filters for years, demographics, and subcommittees, as well as income levels. And using these filters we, as well as Senator Markey's team, have been able to better analyze the request vs approval data and offer suggestions to Senator Markey about how he may better distribute his CDS funding. However, it is important to note that we only had access to 2024's requests to date, so approvals for 2024 were not taken into consideration by our board.

Before we could create our dashboard, we once again had to organize the data we had received. One of the challenges that we faced was collecting FY22 Requests Data. While we did have FY22-funded data available and provided to us within a spreadsheet, we had to come up with a strategy to retrieve the requested data for that year since it was all on a website. We initially tried to just use BeautifulSoup4 and pandas to parse the data but we ran into a lot of issues with data inconsistencies as the website didn't have the data in a consistent format. Some and honestly most of the data was presented in different ways, so we eventually had to resort to using Selenium web scraping which provided better flexibility and more advanced parsing capabilities so we could rectify the data formatting inconsistency issues. After we were able to get all of the data using Selenium, we utilized Pandas capabilities to sort and organize this data which we later uploaded to our spreadsheets. The FY23 Request Data, however, was available in similar PDFs to the FY22 and FY23 Allocation Data, so we were able to use a similar Python script, incorporating Tabula, to convert the data into a spreadsheet. We then used similar

techniques to what we used when collecting FY24 Request Data to retrieve demographic information for the cities that made requests in 2022 and 2023.

To create our dashboard, we utilized Google Looker Studio. Google Looker Studio allowed us to import an all-inclusive spreadsheet containing the data from all of the individual spreadsheets we had been working with and then create charts using that data. We were then able to create various filters that a user can switch between so they can see different portions of the data depicted in the charts to get a better understanding of how request and allocation data has looked in the past few years.

Our extension project aimed to answer the question of whether or not certain demographics were being over/underserved based on both request and allocation data. In the first phase of our project, we merely examined the allocations being given, but to understand the full situation, it is important to also look at what cities/demographics are submitting requests. In the next phase of our project, we will delve into the demographic composition of request submitters, analyzing factors such as geographic location, socioeconomic status, and representation. By correlating this data with approval rates and allocation patterns—bearing in mind the 57% approval ratio—we aim to identify any discrepancies or biases in the distribution of resources. Additionally, we intend to explore how Senator Markey's sponsored requests fare across different committees, shedding light on whether certain areas or demographics receive preferential treatment. Through

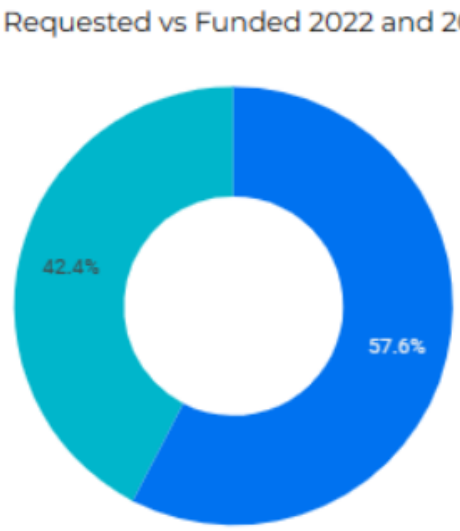


Figure 6: Requested vs Funded percentages (FY22-23)

this comprehensive analysis, we aspire to contribute to a more equitable and informed decision-making process.

As mentioned earlier, our focus shifted to examining the breakdown of subcommittees to determine if they were receiving appropriate funding in line with their requests. Additionally, we aimed to delve into the specific monetary breakdown of each subcommittee's requests and the corresponding allocation they received. This analysis was crucial for evaluating the alignment between requested funds and actual allocations across various subcommittees. By scrutinizing these details, we sought to identify any discrepancies or inefficiencies in the allocation process and propose recommendations for optimizing resource distribution in line with the priorities and needs of each subcommittee.

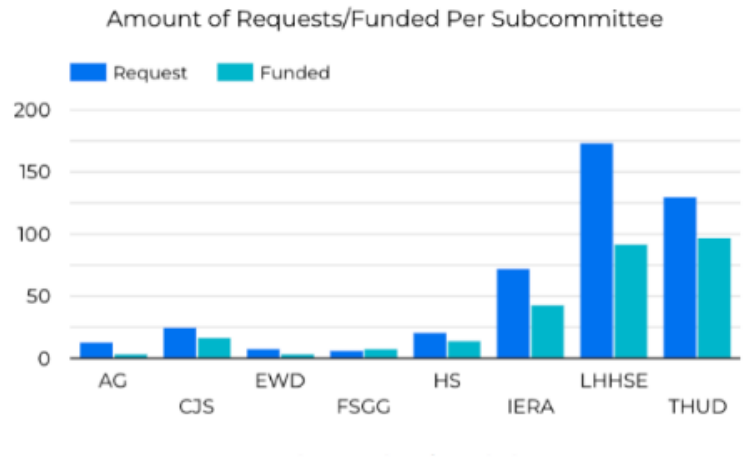


Figure 7: number of requests vs funded per subcommittee

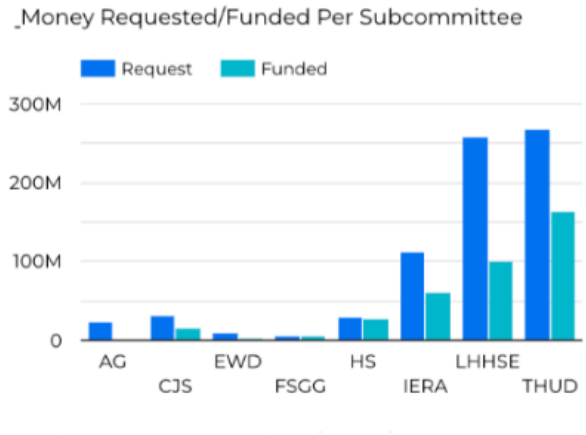


Figure 8: Monetary requests vs funded per subcommittee

Based on what we discussed earlier regarding the subcommittees that Senator Markey claims to prioritize, it was surprising for us to see that Transportation, Housing and Urban Development, and Related Agencies have had more requests approved than Labor, Health and Human Services, Education, and Related Agencies, even though it has submitted less. Labor,

Health and Human Services, Education, and Related Agencies are the subcommittees with the lowest rate of success out of them all.

When we look at the actual money requested and funded, we see that the Labor, Health and Human Services, Education, and Related Agencies subcommittee has a low rate of success at securing dollars as well. Senator Markey and his team may want to more closely examine the exact requests being submitted for the Labor, Health and Human Services, Education, and Related Agencies subcommittee to get a better idea of why they haven't been receiving funding recently and to determine whether or not this should change going forward. They may also want to more closely examine the Transportation, Housing and Urban Development, and Related Agencies subcommittee for the opposite reason.

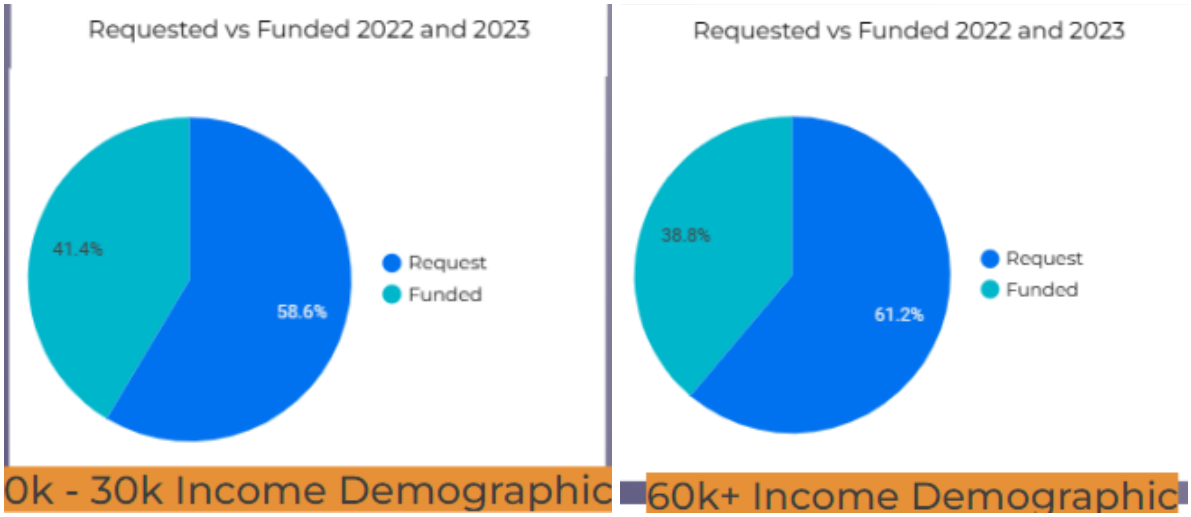


Figure 9: Quantity requests vs funded (FY 22-23) 0k-30k demographic and 60k+ demographic

As we can see from the graph above, the rate of request approvals is approximately 3% higher for the lower-income class when compared to the higher-income class. For the lower income class, while the amount of allocated money to approved requests is double the amount for the higher income class, the amount of money *requested* for the lower income class is almost double that of the higher income class, meaning that while the rates of approvals are similar, the lower

income class is still requesting a lot more money than the higher income class. This discrepancy is noted in the graphs below.

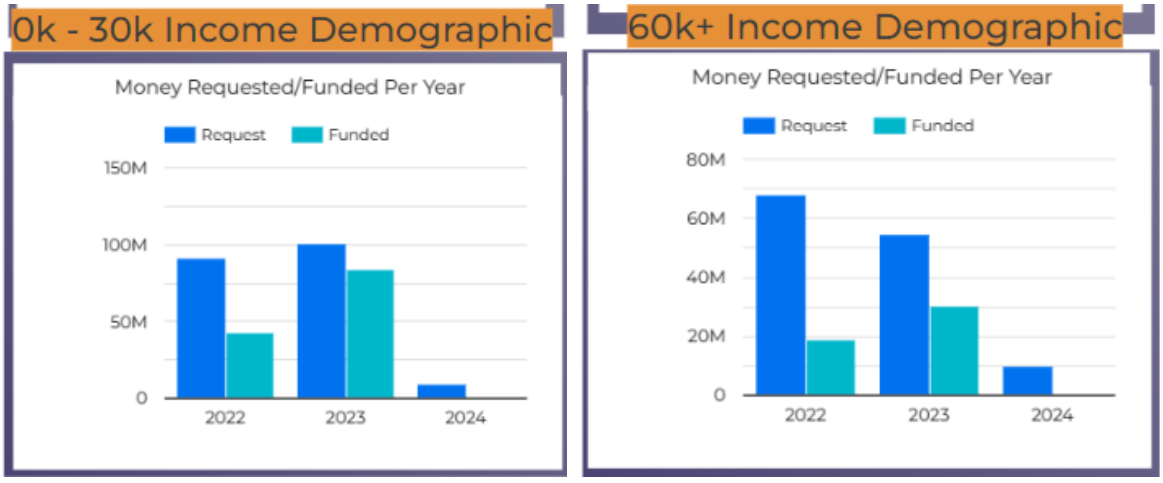


Figure 10: Monetary requests vs funded (FY 22-23) 0k-30k demographic and 60k+ demographic

As we mentioned earlier, a key aspect of our extension project was to analyze the demographics (BIPOC percentages to be precise) of cities that have submitted requests. We aimed to analyze these demographics about whether their requests were approved and to identify if cities with high BIPOC populations are being adequately served as this is one of Senator Ed Markey's highest priorities.

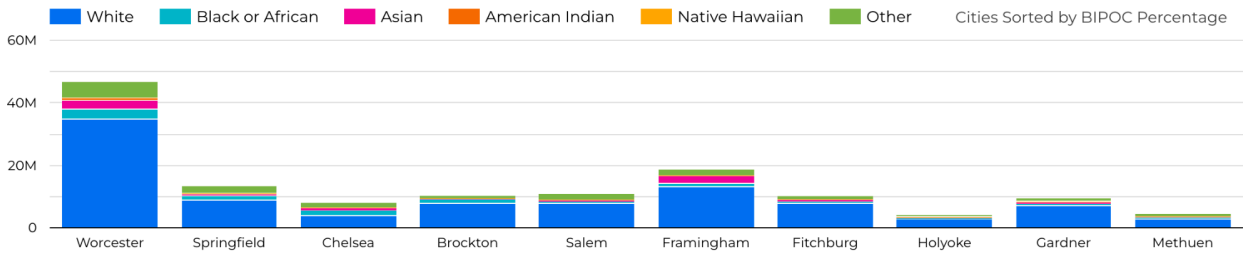


Figure 11: Distribution of demographics across MA cities

In the graph above, we displayed the cities with the highest BIPOC percentages. The only city that we currently excluded from this visualization was Boston so we could get a clearer picture of the demographic spread across Massachusetts as a whole.

Comparing just the requests versus funds for the cities with the highest BIPOC percentages (excluding Boston once again), we can see that just over 50% of the requests were funded. While this is a respectable number by any standard if we also take into account the income levels of these cities, it reveals that many of these cities have very low average income levels.

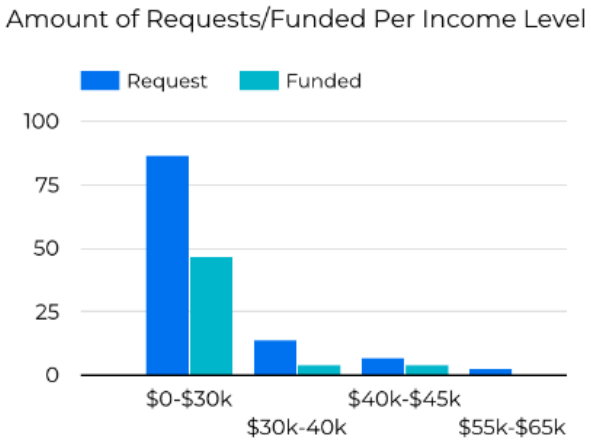


Figure 13: Requests vs Funded across income levels

The graph highlights a notable trend: the majority of requests originate from cities falling within the $0-$30k income range, whereas there are notably fewer requests originating from cities in the $55k-$65k range. This stark contrast underscores a critical observation: higher BIPOC percentage cities tend to have lower income levels. This correlation between demographics and income levels underscores the intersecting challenges faced by marginalized communities. It becomes imperative for Senator Ed Markey's administration to prioritize these cities with higher BIPOC percentages, ensuring they receive the necessary funding and support to address their specific needs. By directing resources towards these communities, Senator Markey can actively work towards addressing systemic inequalities

and fostering inclusive growth and development for all constituents, regardless of socioeconomic status or racial background.

Based on our findings, we have a few key recommendations for Senator Markey and his team. Firstly, we suggest a thorough examination of the allocation distribution among subcommittees, particularly about the observed disparity in approval rates. It may be beneficial to reassess the allocation proportions, considering reallocating resources from areas such as Transportation, Housing and Urban Development, and Related Agencies towards other subcommittees that demonstrate greater need or potential for impact. By reallocating resources strategically, Senator Markey's team can optimize the utilization of funds and better address pressing issues across various sectors.

Additionally, our analysis underscores the importance of targeting cities highlighted as both high in Black, Indigenous, and People of Color (BIPOC) populations and low in income. These cities represent communities that are often disproportionately affected by systemic inequalities and in greater need of assistance. Therefore, Senator Markey's team should prioritize initiatives and allocate resources to these areas to ensure equitable distribution and effective support for marginalized communities. By focusing efforts on these specific demographics, Senator Markey can demonstrate a commitment to addressing systemic disparities and promoting inclusive growth and development.

Future Scope

In considering the future scope of our project analyzing Senator Markey's earmarks, several avenues for further exploration and development emerge. Firstly, we propose an expansion of our analysis to encompass a longer timeframe beyond the three years currently examined. Extending the analysis over multiple decades could unveil more profound trends and patterns in earmark allocation, allowing for a deeper understanding of how resource distribution has evolved. This approach would provide valuable insights into the effectiveness and impact of earmark allocation strategies, facilitating evidence-based policymaking and resource allocation decisions.

Secondly, our analysis primarily focused on the distribution of earmark funding across different demographic groups and subcommittees. A future direction could involve a more granular examination of specific project types and their outcomes. By categorizing earmark-funded projects based on their objectives, such as infrastructure development, community services, or environmental initiatives, we can assess their efficacy in addressing pressing societal needs and achieving desired outcomes. This project-level analysis would enable policymakers to identify successful project models and allocate resources more effectively based on project merit and impact.

Moreover, incorporating additional data sources and analytical techniques could enhance the depth and scope of our analysis. Integration of qualitative data, such as stakeholder interviews or case studies, could provide rich contextual insights into the factors influencing earmark allocation decisions and their on-the-ground impact. Furthermore, employing advanced statistical methods, such as predictive modeling or machine learning algorithms, could enable the

identification of predictive factors influencing earmark funding approval, thereby enhancing the accuracy and efficiency of resource allocation processes.

Another aspect of future exploration involves the development of interactive visualization tools and dashboards to facilitate data-driven decision-making and stakeholder engagement. Building upon the interactive dashboard created in our extension analysis, future iterations could incorporate real-time data updates, customizable filters, and dynamic visualizations to provide policymakers with actionable insights and facilitate transparent communication with constituents. Additionally, expanding the scope of our analysis beyond Massachusetts to encompass a broader geographical context could yield valuable comparative insights into earmark allocation practices across different states and regions. By benchmarking Senator Markey's earmark allocation strategies against those of other lawmakers, policymakers can identify best practices and opportunities for improvement in resource allocation processes.

In summary, the future scope of our project entails extending the analysis over a longer timeframe, conducting project-level evaluations, integrating additional data sources and analytical techniques, developing interactive visualization tools, and expanding the geographical scope of analysis. These future directions aim to enhance the comprehensiveness, accuracy, and utility of our analysis, ultimately contributing to more equitable and effective earmark allocation practices.

INDIVIDUAL CONTRIBUTIONS:

Zuizz Saeed:

Many of my contributions consisted of creating graphs and visuals for the data that was given. For example, I worked on creating graphs that highlighted the various distributions for the different subcommittees and looking into disparities among them (and demographics for each one). I also worked on cleaning up some of the data in cases where there were inconsistent entries for things like the location of the project. For the client presentations, the slides were split up evenly among the group members, and I presented these slides to the client as the team representative and answered any questions.

Jeremy Bui:

My contributions mainly dealt with Jupyter Notebooks in creating the graphs for initial analysis and some in-depth (some examples are income, distribution among rural/suburban/urban, BIPOC distribution, etc). Moreover, I was also able to further expand this once we received our requests/approval for the mid-semester report for BIPOC and income. I also helped to gather the number of requests for Markey's 2023 and convert them to CSV format. Lastly, I helped create some graphs and the template for our LookerStudio dashboard for the extension project. For the client presentations, the slides were all assigned evenly across members.

Rithvik Nakirikanti:

At the beginning of the project, since much of the work was data cleaning, I worked a lot with pandas to help convert data that was in PDFs with tables into tabular form. I was also helping assist with the census tract API and finding the right endpoints with John because there were a

lot of options and we needed to find the demographic percentages that would be most useful for our project. Then I helped assist with creating some of the graphs, and for the extension project, I worked on the Selenium web scraping and data cleaning to get the FY22 requests data into spreadsheets. Overall, we all helped each other with making slides, presenting etc.

Grace Elias:

I worked a lot in the beginning with the 2024 request data to convert it to tabular form. I later used similar techniques on the 2023 request data. As we worked on our early insights, mid-semester, and final reports, I created many graphs and analyzed the data significantly. I then used this understanding of the data to help work on our dashboard to determine which features we should show and let users filter by.

Jiang Cheng Liu:

Majority of contributions were done with the data processing where I analyzed the files and used libraries like Pandas to iterate and edit through the datasets. For instance, one of the earliest stepping stones was being able to get the census tract data for each city, and to achieve that I had to use the Google Maps API to first get the latitude and longitude and then feed that towards the census tract API to get important information such as the census tract ID number and county number. Furthermore, I was tasked to change how we gathered the demographic information by using a different API endpoint and thus changing from percentages to count. For the client presentations, the slides were delegated evenly and accordingly.