

CS 506 - Data Science Fundamentals

City of Boston: City Budget Team A

Group Members: Toby Ueno, Deep Patel, Lucia Kisova, Thian Amarasekera

Problem Statement:

For this deliverable, we have been working on analyzing both datasets given and answering all questions outlined in the base project. We have been given data regarding the department and category spending for the city of Boston, as well as the proposed operating budget for FY24. We are working on finding out how money has been spent for the years 2021, 2022, 2023, 2024, by breaking down how spending has been allocated to different departments and programs, and how much different locations receive. Additionally, we are looking at how much money is spent per capita for all the different programs.

Data Collection/Cleaning:

Our data was pre-collected from the City of Boston website for the Operating Budget Data and Planned Capital Budget Data; no extra steps for collection were necessary beyond downloading the associated CSV files. For the extension project we have decided on for now, we have found a dataset from the city of Boston regarding how many people live in a particular neighborhood in the neighborhoods of Boston.

To clean the data, the main step was to get rid of missing values, particularly in numeric fields. Numeric fields had the string “#MISSING” instead of a NaN value, sometimes with whitespace surrounding the string, so each numeric column was first treated as a string, stripped of whitespace, and then all rows with any values equal to this “missing” string were dropped. Columns were then safely cast to numeric values.

Hypothesis: One interesting thing to note is that the Execution of Courts department and Annual Audit costs reported a significant percent increase in funding. We hypothesize that these two are correlated. We hypothesize that the city is increasing the funding in these departments to potentially increase revenue in future years.

Key Question #1: How has spending per department changed over time?

On average, the city budget increased by **5.59%** per year. Budget had a **6.56%** increase from 2021-2022, a **3.56%** increase from 2022-2023, and a projected **6.65%** increase from 2023-2024.

The three departments which gained the largest percent-increase in funding per year over this four-year period were:

- Execution of Courts (under the Finance Cabinet), with a 136.81% increase
- Annual Audit Costs (under “Other”), with a 133.88% increase
- Reserve for Collective Bargaining City (also under “Other”), with a 117.21% increase

Notably, this excludes the OPAT cabinet, which received no funding in 2021 (it was conceived in 2022).

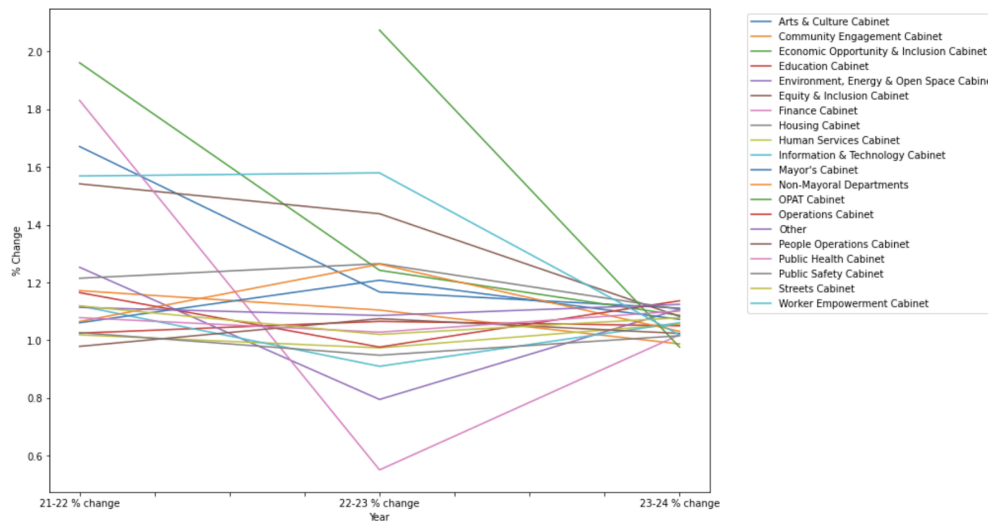
The three departments which lost the largest percent-decrease in funding per year over this four-year period were:

- Office of Finance (under the Finance Cabinet), with a 20.12% decrease

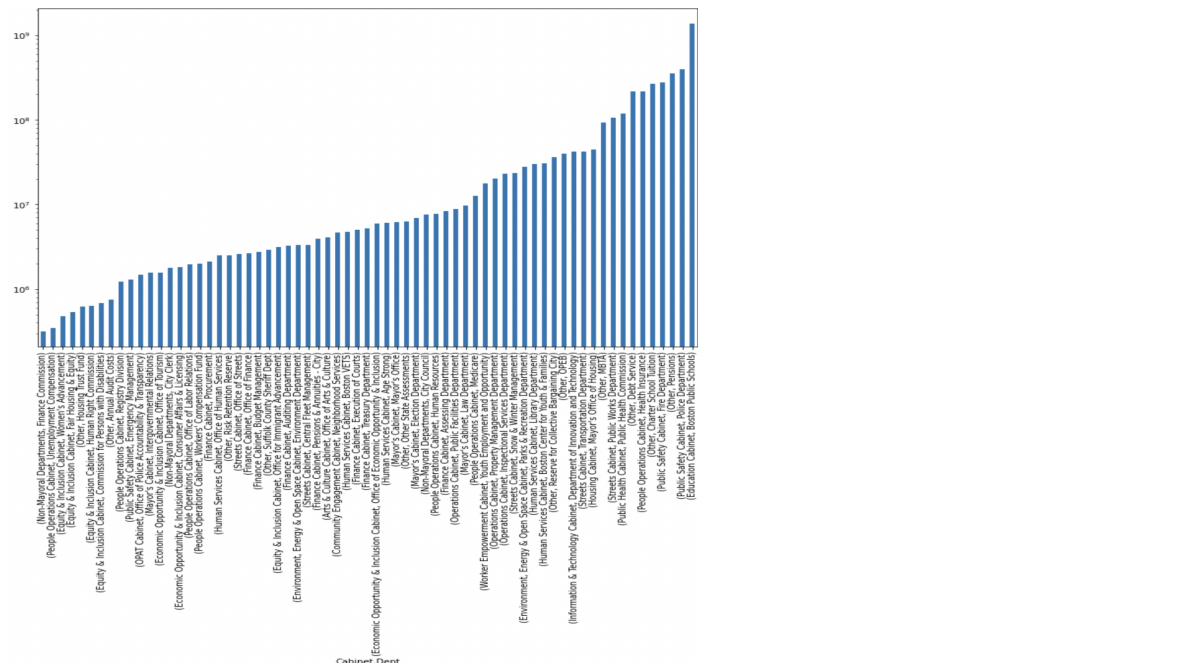
- Suffolk County Sheriff Dpt. (under “Other”), with an 8.32% decrease
- Library Department (under the Human Services Cabinet), with a 4.89% decrease

The Police Department and the Central Fleet Management were the only other two departments to report an average yearly decrease in funding over this period. All other 61 departments reported an increase.

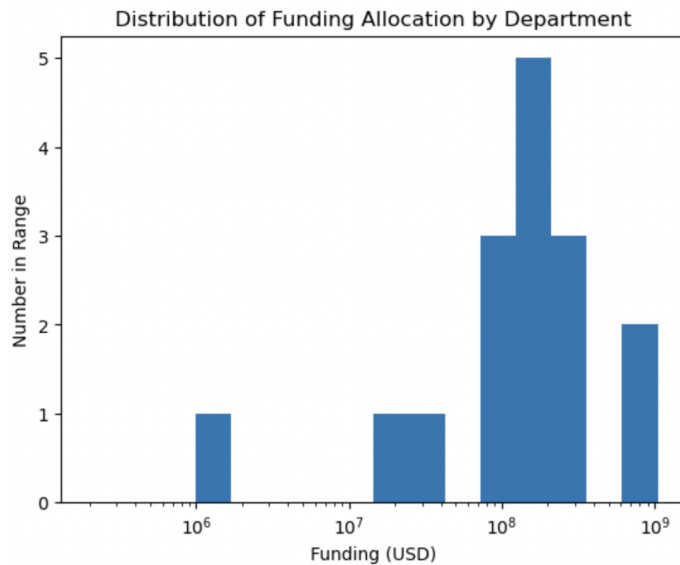
Visual: The data given shows this increase as well. A general increase year over year can be seen by looking at the percentage increase from each year. This data shows that while there have been some outliers, most departments show an increase in spending.



Visual: This bar chart shows the expenditure of the various cabinet departments in the city of Boston for the 2023 fiscal year (actual + project).



Visual: A similar bar chart shows the distribution of funding for the year 2024 in terms of how many departments receive that much funding. As seen, many departments are receiving a large amount of funding.



Key Question #2: **How has spending per program changed over time?**

The three programs which gained the largest percent-increase in funding over this four-year period were:

- Trust (under the Treasury Department), with a 4854.92% increase
- Small & Local Business (under the Office of Economic Opportunity & Inclusion), with a 656.51% increase
- Reserve for Collective Bargaining City (under department of the same name), with a 444.38% increase

The Trust project is somewhat of an outlier, since its budget in 2021 was relatively low (\$631.62).

Again, this excludes the OPAT cabinet, which received no funding in 2021 (it was conceived in 2022).

The three programs which lost the largest percent-decrease in funding over this four-year period were:

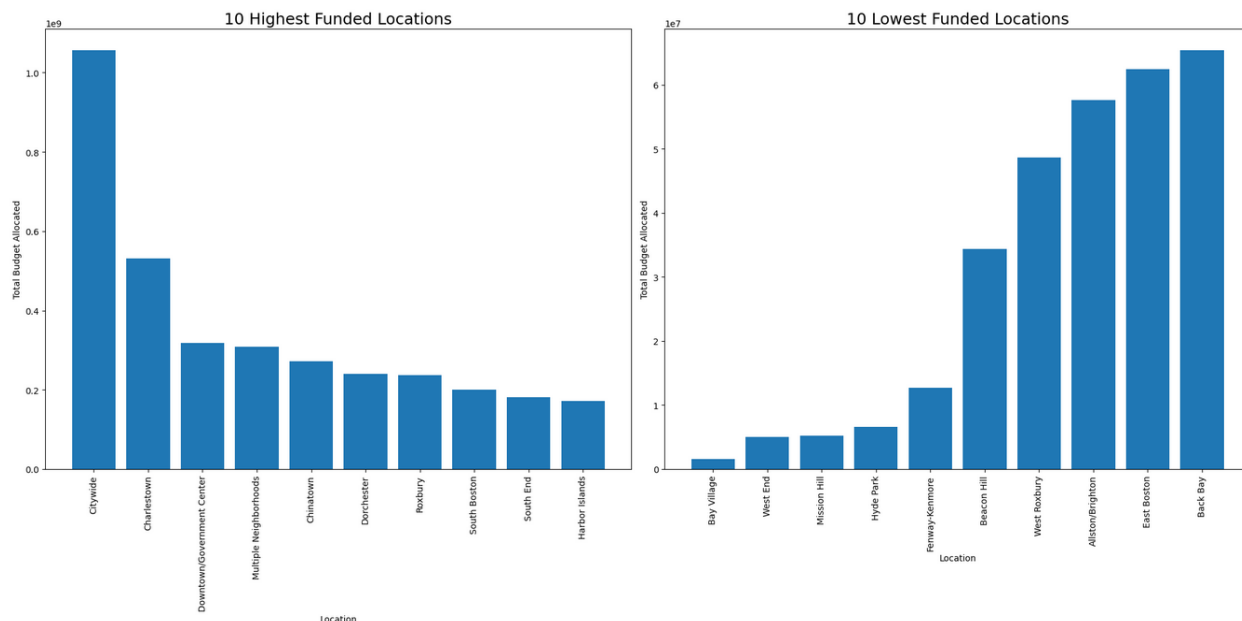
- Affirmative Action (under Human Resources), with a 97.31% decrease
- Consumer Affairs (under Consumer Affairs & Licensing), with an 56.63% decrease
- Alterations & Repair (under the Property Management Department) with a 54.68% decrease

151 projects (82.07%) received an increase in funding, whereas 33 (17.93%) received a decrease.

Key Question 3: How is spending allocated to different locations in the city?

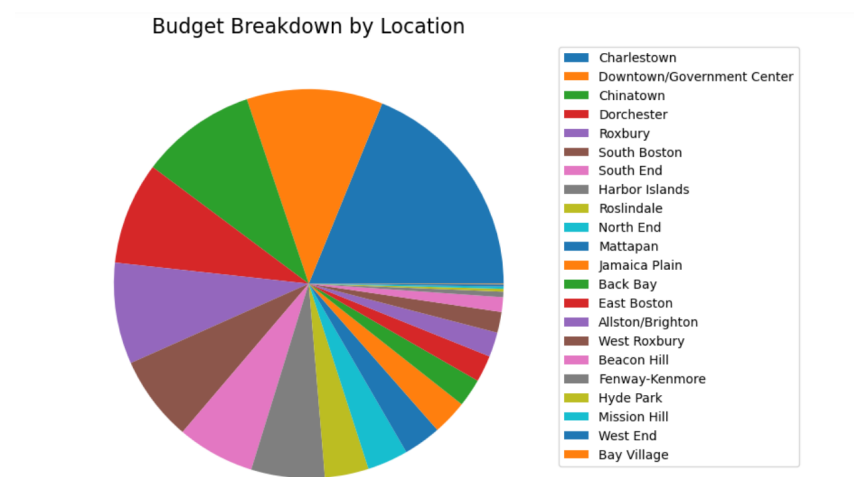
Excluding citywide projects, Charlestown was by far the most funded neighborhood, with a total budget of \$531.2 million. Bay Village was the lowest funded neighborhood with a total budget of only \$1.53 million. Charlestown's budget is 18.84% of the total neighborhood-specific project budget, while Bay Village's budget is only 0.05% of the total neighborhood-specific budget.

Visual: This bar chart shows the allocation of funding among the 10 highest funded and 10 lowest funded neighborhoods in the city.



A total of 21 neighborhoods had projects that were included in the Capital Budget Plan. We aim to analyze characteristics of neighborhoods that receive more/less funding in the extension project.

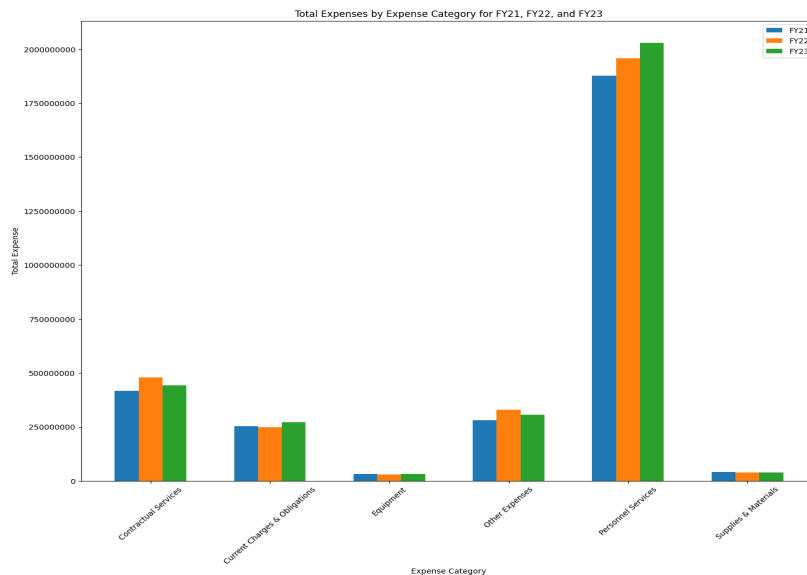
Visual: Pie chart showing full breakdown of funding allocated to specific neighborhoods. Excludes citywide and multiple neighborhood projects.



Key question #4: How is spending broken down by budget category?

There are 7 main expense categories that the City of Boston has. These are: Personnel Services, Contractual Services, 'Supplies & Materials, Current Charges & Obligations, 'Equipment, Other Expenses, and Fixed Expenses. Of these, the lowest spending category are the Fixed Expenses. For the fiscal years 2021, 2022, and 2023 these expenses on average have been about \$191,092,897. The Personnel Expenses are by far the largest expense category each year. Personnel expenses have been consistently above 1.8 billion dollars per year for the past 3 years. One interesting thing to note is that it has also been the one of the only expense categories where the expense has risen for the past 3 years. All other categories, besides fixed expenses, have had some fluctuation year to year. This can be attributed to an expanding workforce and increased wages to compensate for inflation and cost of living adjustments.

Visual: Bar graph depicting total expense by category for the years 2021, 2022, 2023. This excludes the fixed expenses which are far lower than the rest of the expenses shown. As seen, personnel services cost the city of Boston upwards of \$1,800,000,000 per year.



Extension Project Proposal:

Extension Project Proposal	
Extension Pitch	In this extension project, our primary focus is to uncover and analyze trends in neighborhood spending per capita, shedding light on the variations and disparities that exist across different neighborhoods in Boston. By delving into this exploration, we hope to provide valuable insights that can inform local policy-markers, business strategies, and community initiatives to foster a more equitable and informed approach to urban planning and development regarding city budget.

Rationale	<p>The extension project is important as it can unravel the intricate relationship between neighborhood funding and population density. This could offer a better understanding of how expenditures vary across diverse communities in Boston. The significance lies in the visibility of these expenditures to local residents, making the findings essential for shaping policies and initiatives that are designed to address the needs and preferences of each respective neighborhood. By bridging the gap between spending patterns and population density, the project aims to empower communities, local authorities, and businesses with actionable insights. This can be important when looking to the future in order to create a more responsive and community-centric approach to urban development in Boston.</p>
Questions for Analysis	<p>Our analysis will delve into pivotal questions surrounding neighborhoods with higher spending per capita, aiming to hone in on trends and similarities among those areas. By investigating whether proximity and smaller population density contribute to heightened spending patterns, the project seeks to validate a preliminary hypothesis. This will provide valuable insights into the spatial and demographic dynamics that shape neighborhood expenditure. These questions serve as the foundation for uncovering patterns that can inform targeted interventions and policies, fostering a more comprehensive understanding of the economic dynamics at the community level in Boston.</p>
Data Sets & Sources	<p>The primary dataset, sourced from the City of Boston, provides essential information on population per neighborhood with the corresponding neighborhood names and population figures. By leveraging this comprehensive dataset, we aim to conduct an analysis of funding across various neighborhoods with different population densities and spending per capita. This can allow us to draw correlations between population density and expenditure patterns.</p>
Data Visualizations	<p>Our data visualization centers around the creation of an informative and visually compelling map of Boston, detailing spending per capita for each neighborhood. By employing geospatial visualization, this map can provide a clear and accessible representation of spending patterns across the city. This visual tool not only enhances the communication of our findings but also facilitates a more intuitive understanding of the economic landscape in Boston which can aid policy-makers in making informed decisions related to neighborhood development and resource allocation.</p>
Additional Information	<p>Ensuring the accuracy and reliability of our analysis is essential and to address this concern we are focused on accessing a synchronized dataset for population per neighborhood and spending per neighborhood for the same</p>

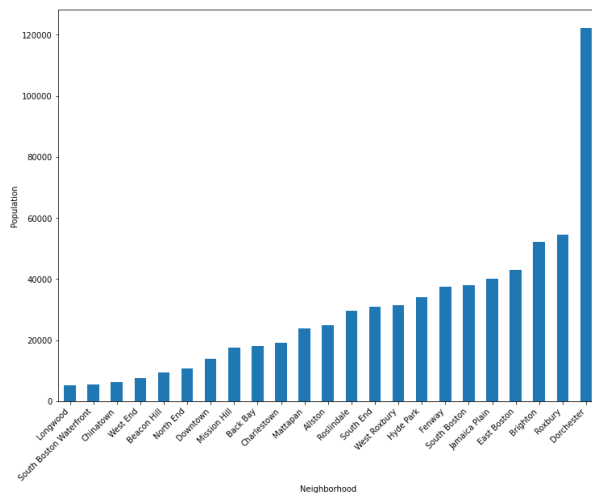
year(s). Mismatched data years have the potential to introduce inaccuracies into our findings and make it more difficult to draw meaningful conclusions for our analysis. We are working on finding a more comprehensive dataset in order to enhance the precision of our conclusions to provide a more robust foundation for informed decision-making in understanding the dynamics of Boston's neighborhoods.

Visualization and Insights for Extension Proposal:

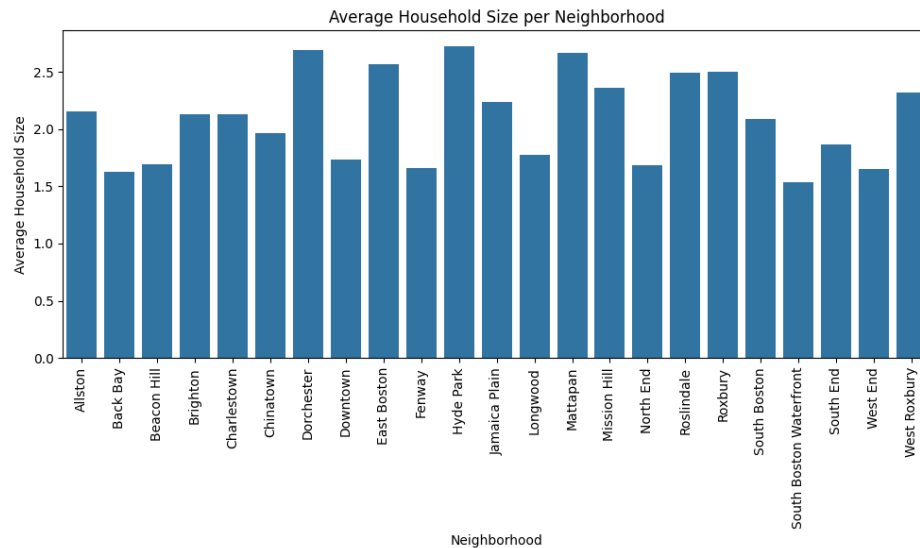
Work on our extension is very much still in its early stages, as there has been insufficient time in between our deciding on the direction of our extension and the submission of this report to draw any large or significant insights from the extension data. We have focused so far on obtaining, cleaning, and performing preliminary exploratory analysis on a new dataset that will give more information on neighborhood demographics and other statistics.

The dataset selected was a subset of the 2020 Census data, available on the City of Boston website. The dataset contained population statistics for neighborhoods, including breakdowns by race/ethnicity and age, and some other miscellaneous information averaged by neighborhood, including household size, university housing amount, number of correctional facilities, etc. Each column was identified via a code, such as P0020001, which was manually cross-referenced with a Census summary file to extract the meaning of the code. Cleaning and preprocessing this dataset involved renaming the columns to carry semantic value instead of these arbitrary codes, as well as typical data type conversion and removing or rearranging out-of-place rows and columns (such as that of the Neighborhood, which was moved to the index of the dataframe).

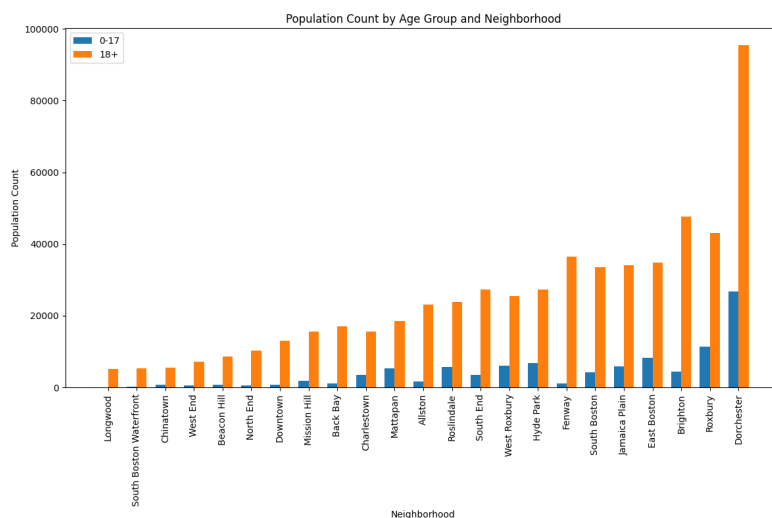
Visual: Bar graph showcasing the overall population by neighborhood. Dorchester notably had the highest population by a large margin (almost double that of second-place Roxbury), whereas Longwood had the smallest.



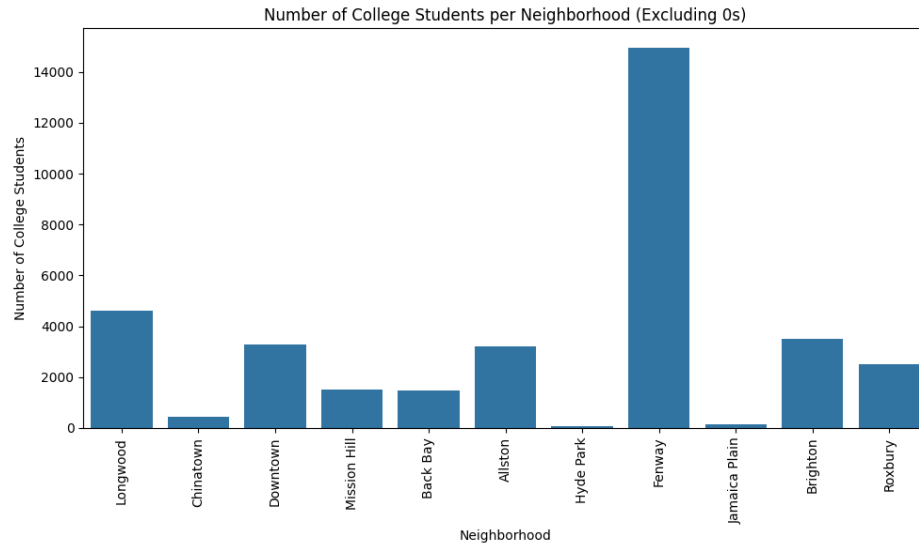
Visual: Bar graph displaying the average household size by neighborhood. There does not seem to be any particular neighborhood that is an outlier for this metric. The typical range of household size ranges from 1.53 (South Boston Waterfront) to 2.72 (Hyde Park).



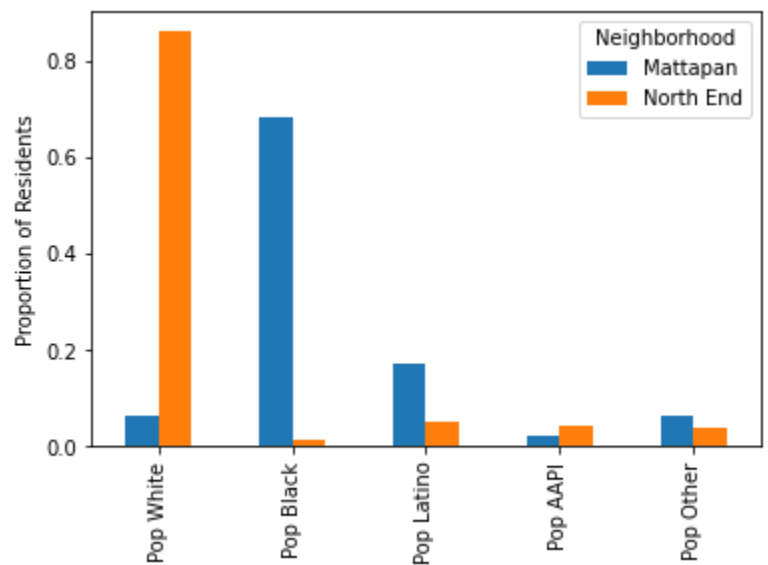
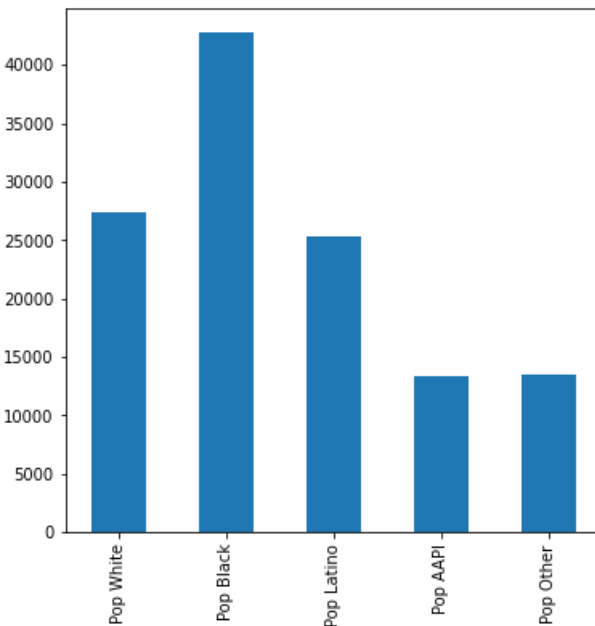
Visual: Bar graph displaying age groups by neighborhood. The two age groups created per neighborhood are 0-17 and 18+. In every neighborhood, the number of individuals who are older than 18 is far greater than the number of individuals who are 0-17. This is to be expected. One thing to note is that some neighborhoods have a disproportionately larger number of individuals who are 18+ and have a very small percentage of the population who are 0-17 years of age.



Visual: Bar chart depicting the number of college students in each neighborhood of Boston, excluding those that do not have college students. Fenway appears to be a very popular location for students to live.



Visual: Bar chart showing the race demographics of Dorchester, the neighborhood with the aforementioned largest population. Black or African American residents comprise the largest single racial group of Dorchester residents, at around 35% of the total population. White and Hispanic/Latino residents are about even at around 21% each, followed by Asian/Pacific Islanders and Other (which includes two or more races) at around 11% each.



Visual: Comparative bar chart between the demographics of Mattapan and the North End, the neighborhoods with the largest and smallest proportions of Black residents, respectively. The y-axis is normalized to proportion of population instead of absolute numbers to account for the

neighborhoods being different sizes. Here, we observe an anecdotal inverse relationship between Black and white populations; when one is high, the other is not. Furthermore, the percentage of residents of other races is relatively low for races other than the majority.

Contributions of Team Members:

Toby: Toby primarily analyzed the Operating Budget dataset. He cleaned and preprocessed the data, performed exploratory insight-gathering. He grouped the spending data by cabinet, department, and program, and he analyzed those groups with the goal of answering the related base questions for this project. He also performed some exploratory data analysis for the initial extension dataset, cleaning and preprocessing that before creating some initial visualizations. He additionally contributed to the project scrum reports, write-ups, and presentations.

Lucy: Lucy focused on analysis for the Capital Budget Plan. She cleaned and preprocessed the data using the procedure outlined above, analyzed the breakdown of funding by department and location and created several visuals to answer these questions. She also began doing research for the extension project and finding additional datasets, while communicating with the TPM to ensure all requirements for the base project were met.

Deep: Deep worked on the analysis for the adopted Operating Budget dataset. He cleaned and preprocessed the data in order to get it into an analysis ready format. He worked on using this data to answer a few of the key questions required by the base project. In addition, Deep contributed to working on developing an extension project proposal and found some data that may be useful in the extension project. Deep also helped contribute to the presentation.

Thian: Thian focused on writing up the extension project proposal. He took the preliminary information for the Early Insights Presentation and researched more about the topic and how it could be researched. From his research, he had a better understanding about the project goals, rationale and analysis. He was able to outline the extension pitch in a detailed manner using the recommended template.