

DELIVERABLE 2

1. Problem Statement, Data Cleaning/Collection

For this deliverable, we have conducted an analysis on 3 datasets (Operating Budget, Capital Budget, and Revenue Budget), addressing all the key questions outlined in the base project. The datasets provide information on budget allocations for different departments and expense categories spending for the city of Boston from FY2021 to FY2024. Our focus is on understanding the expenditure patterns for these years. This involves dissecting the allocation of funds to various departments and programs, as well as examining the distribution of funds across different areas.

We obtained our data for the Operating Budget, Capital Budget Plan, and Revenue Budget from the City of Boston website, requiring no additional steps beyond downloading the associated CSV files. As for our current extension project, we've sourced datasets from the City of Boston that provides information on census tracts in 2020 and neighborhood demographics within the city.

Data cleaning primarily involved addressing missing values, particularly in numeric fields. In these fields, the string "#MISSING" was used instead of a NaN value, sometimes with surrounding whitespace. To address this, each numeric column was initially treated as a string, whitespace was stripped, and then all rows containing values equal to this "missing" string were removed. Subsequently, columns were safely converted to numeric values.

2. Exploratory Data Analysis (1 page, only add the meaningful plots)

The initial exploration reveals that the City of Boston operates with a comprehensive structure. There are 82 distinct departments, 220 programs, and 7 expenses categories, highlighting the diverse range of initiatives undertaken by the city.

To discern the city's priorities, we investigate the distribution of funds across departments, programs, and expense categories over years based on the operating budget. Firstly, we utilize bar charts to highlight the funding distribution among various entities. This includes identifying the top-funded departments, programs, geographic areas, and expense categories. The graphical representation enables a clear comparison of the relative funding levels, spotlighting entities that have received the highest share of the budget.

Additionally, we utilize pie charts to present a holistic view of the budget distribution by showcasing the proportional contribution of top-funded departments to the overall budget. This succinct visualization method offers stakeholders an intuitive understanding of the concentration of financial resources within specific departments.

Besides, our analysis in `'revenue_2.ipynb'` also introduces a series of scatter plots, each designed to shed light on different dimensions of the dataset. The initial scatter plots focus on the frequency distribution of various revenue categories, accounts, cabinets, and departments within the budget. These visual representations are crucial as they offer a clear picture of the prevalent categories and the most actively involved departments and accounts in the city's financial planning.

3. Visualizations, methodology for finding underlying patterns, and insights for key base questions

Key question #1: **How does the city of Boston spend its annual budget and how has this changed over time?**

First, we calculate the sum of each year's operating budget over time from FY21-FY22.

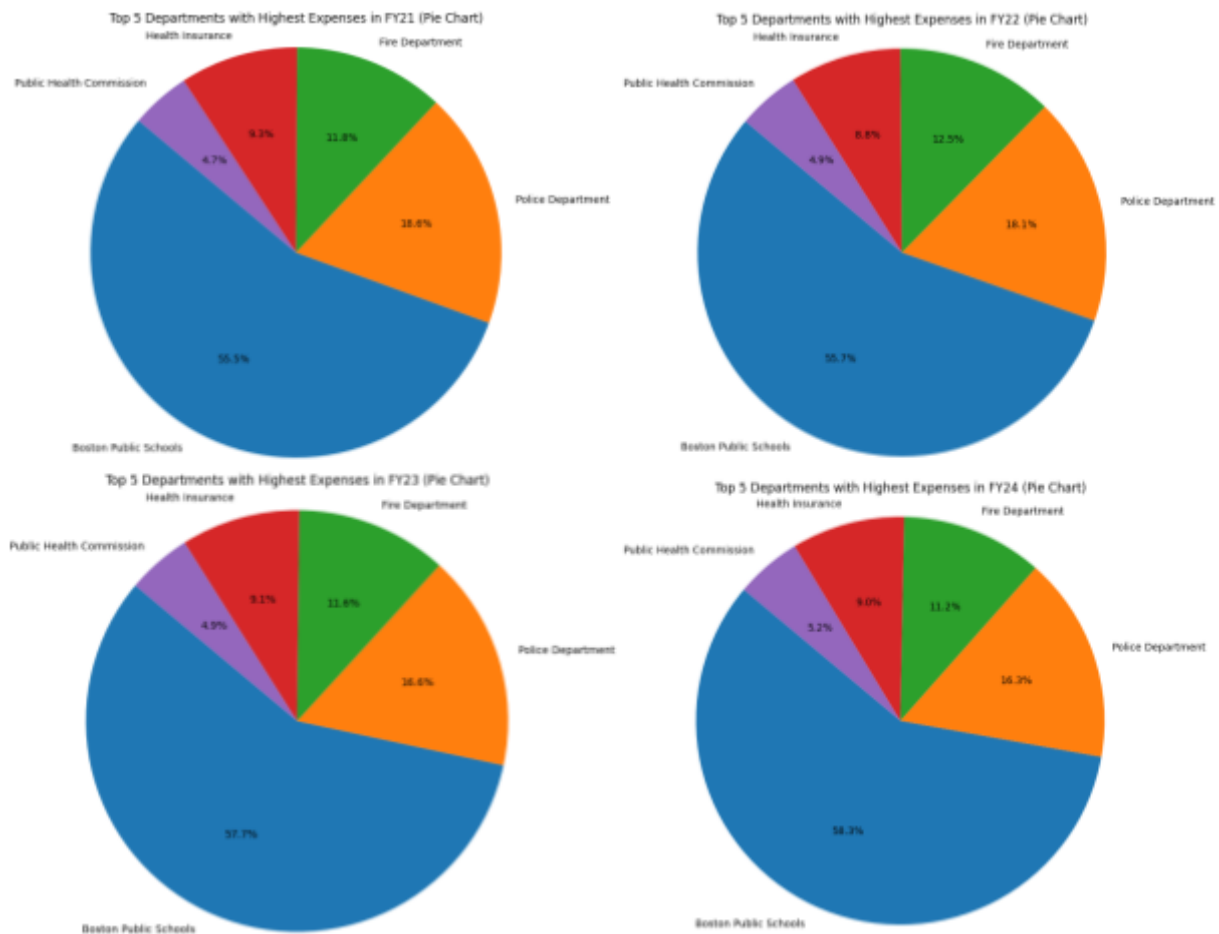
The average annual increase in the annual budget is approximately **5.59%** per year. Specifically, there was a **6.56%** increase from 2021 to 2022, a **3.56%** increase from 2022 to 2023, and a projected **6.65%** increase from 2023 to 2024.

After that, we explore how the city of Boston spends its annual budget. The five departments that experienced the highest increases in funding from 2021 to 2024 were

- Boston Public Schools (representing **55.5%**, **55.7%**, **57.7%**, and **58.3%** of the budget during this period, respectively.).

- Police Department (representing **18.6%**, **18.1%**, **16.6%**, and **16.3%** of the budget).
- Fire Department (representing **11.8%**, **12.5%**, **11.6%**, **11.2%** of the budget).
- Health Insurance (representing **9.3%**, **8.8%**, **9.1%**, **9.0%** of the budget).
- Public Health Commission (representing **4.7%**, **4.9%**, **4.9%**, **5.2%** of the budget).

Visualization:

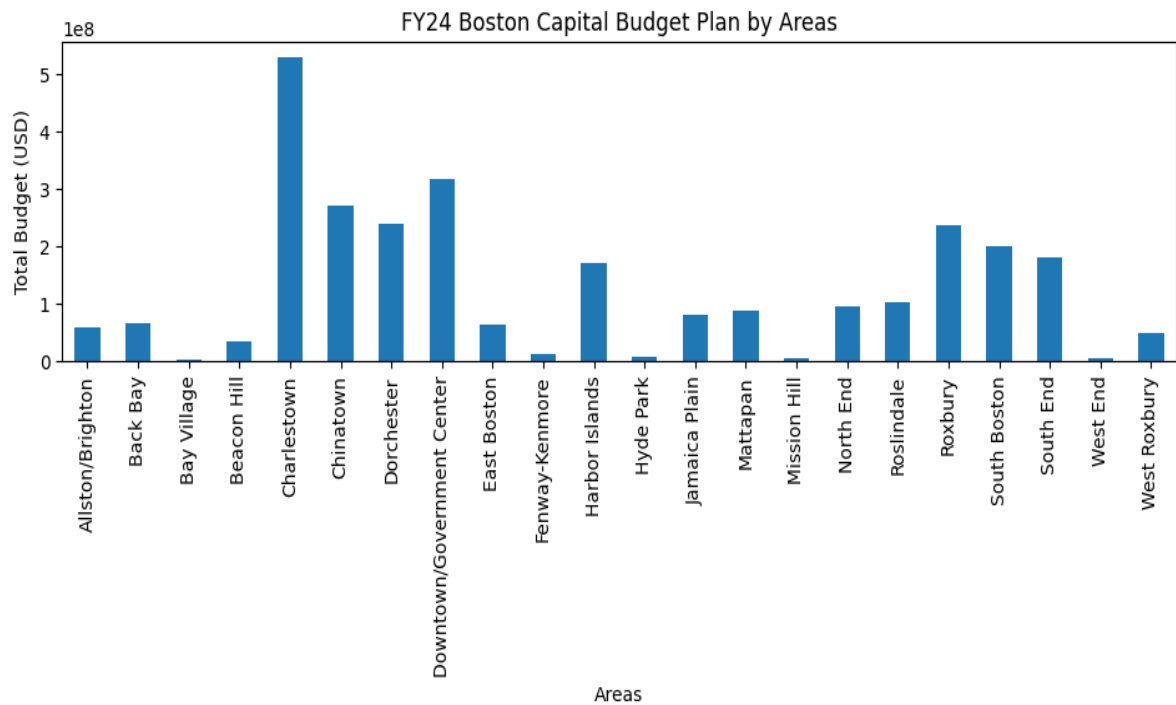


Key question #2: **How is spending allocated to different locations in the city?**

We observe that Charlestown stands out as the neighborhood with the highest investment, receiving roughly \$500 million dollars. Following Charlestown, we also observe

significant investments in Downtown/Government Center, Chinatown, Dorchester, and Roxbury.

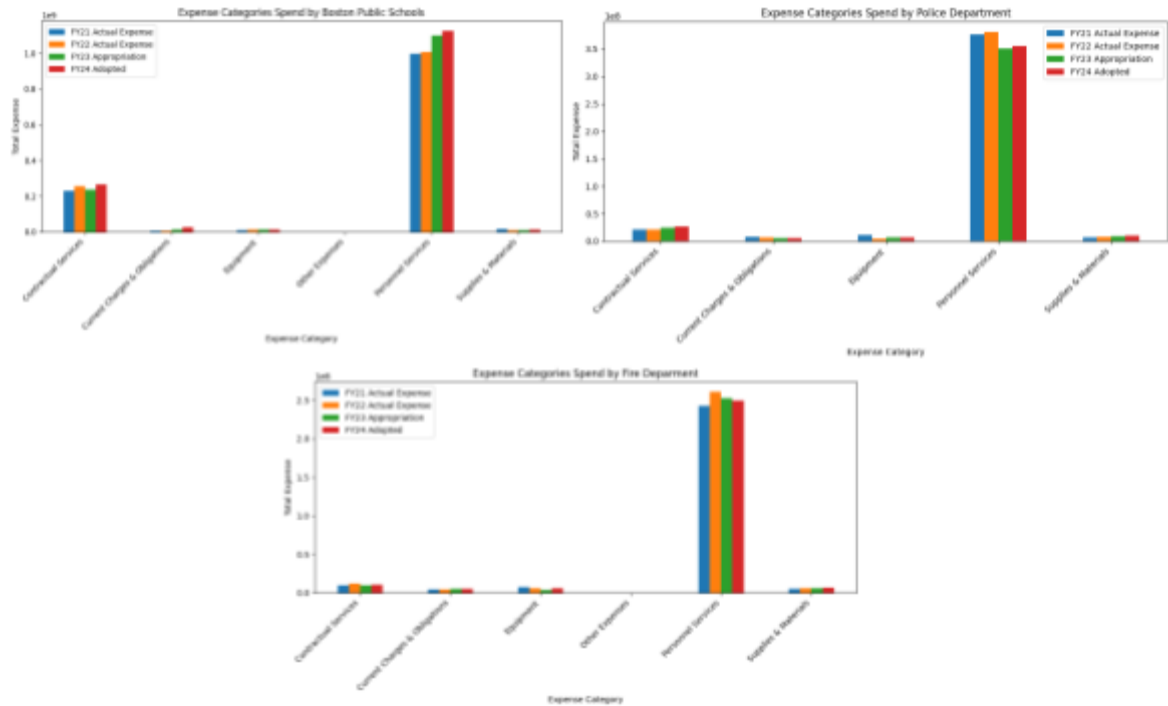
Visualization:



Key Question #3: Budget Allocation for Expense Categories

So far, we have determined that the Boston Public School, Police Department, and Fire Department allocate the majority of their budget to personnel services. This implies that the highest funding is directed towards the salaries of teachers, police officers, and firefighters.

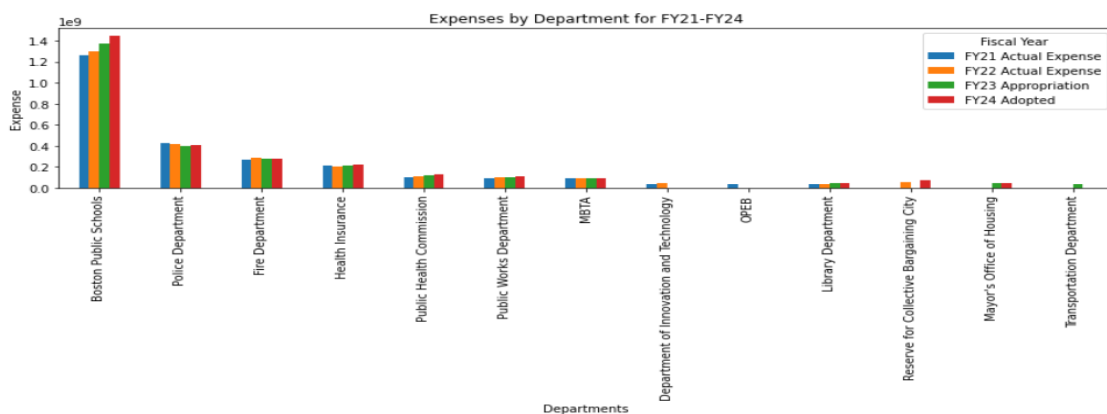
Visualization:



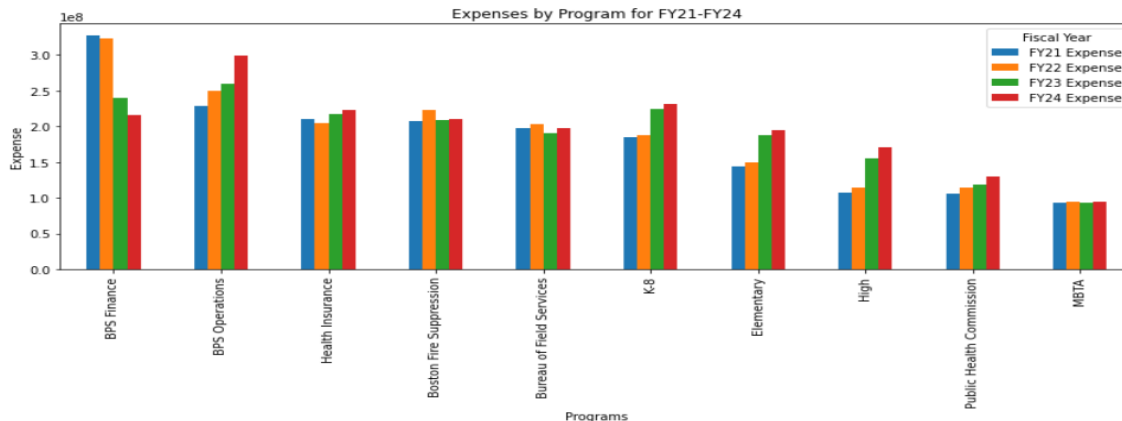
Key Question #4: Expenses by Department and Program.

To provide a comprehensive overview of our financial allocations, we will delve into the breakdown of expenses by department and program. This detailed analysis will illuminate the distribution of funds, shedding light on how resources are allocated across various facets of our organization.

Visualization:



Boston Public Schools dominate in terms of the City's operational budget allocation. We also see a healthy upward trend in the BPS Budget allocation from FY21 to FY24.



BPS Finance and BPS Operations represent the most generously funded programs in the City of Budget. It's noticeable that while BPS Finance experiences a declining trend in its operating budget, BPS Operations consistently receives increased investment every year.

4. Your extension proposal

When analyzing a city's budget, it is important to consider where resources are being committed and what ends those resources are serving. We can see now how spending is allocated by department or by neighborhood, but we don't have a clear picture of who or what is being served by the current budget. Furthermore, digging deeper into demographic information can point towards potential inequities or gaps in resources for certain groups.

Generally, the information we want to look into is how spending is correlated by age. Are older neighborhoods more highly invested in? Is the city catering its spending to younger families? College students?

Currently, we have access to Census data:

<https://data.boston.gov/dataset/2020-census-for-boston> and neighborhood demographics data:

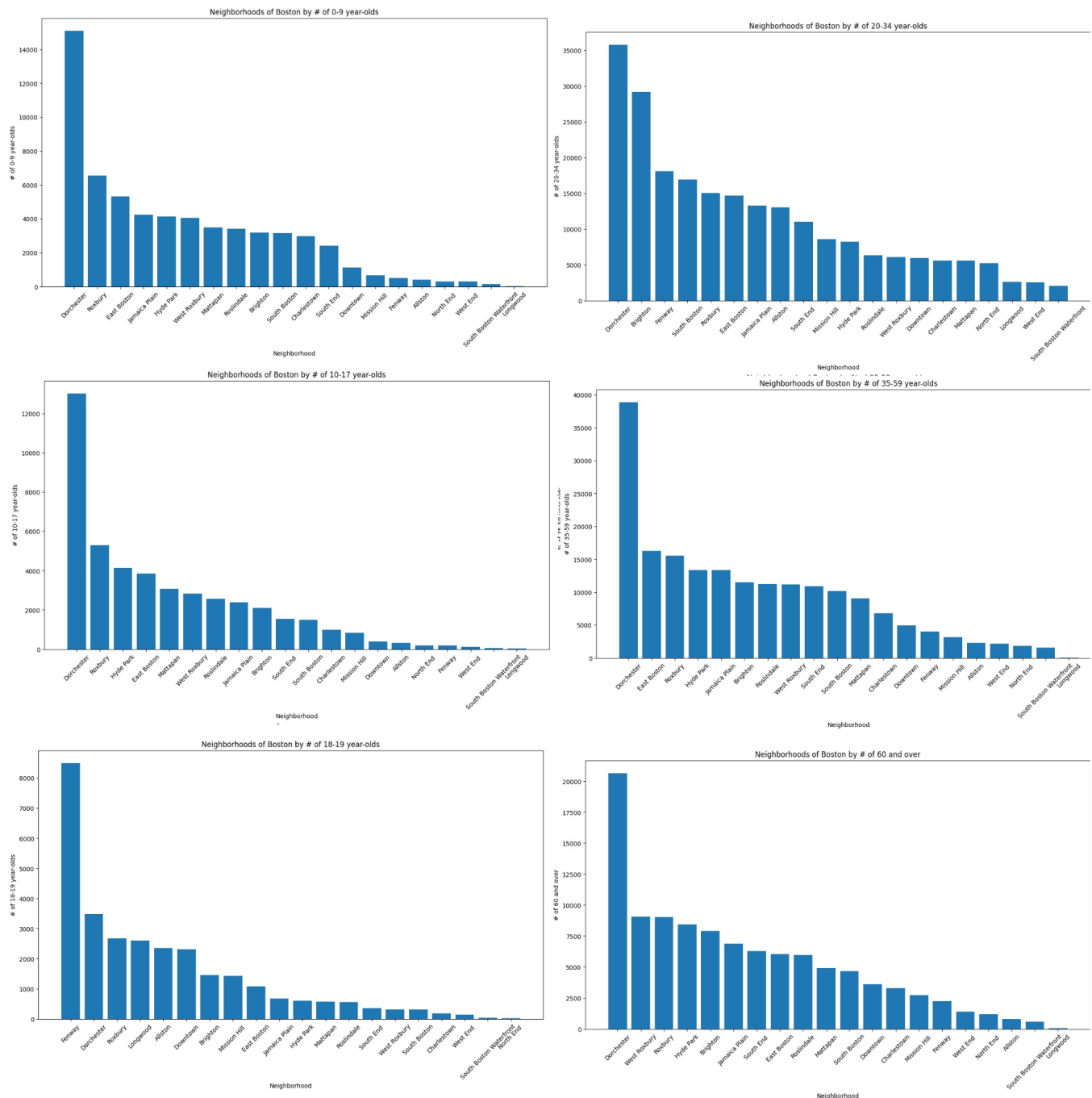
<https://data.boston.gov/dataset/neighborhood-demographics>. These have age demographic breakdowns by neighborhood as well as a plethora of other information we can potentially dig into to expand our analysis.

This extension project was inspired in part by an article in the Boston Globe:

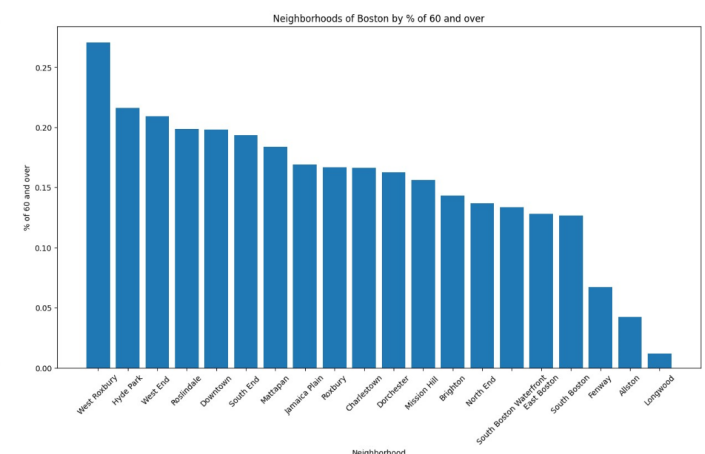
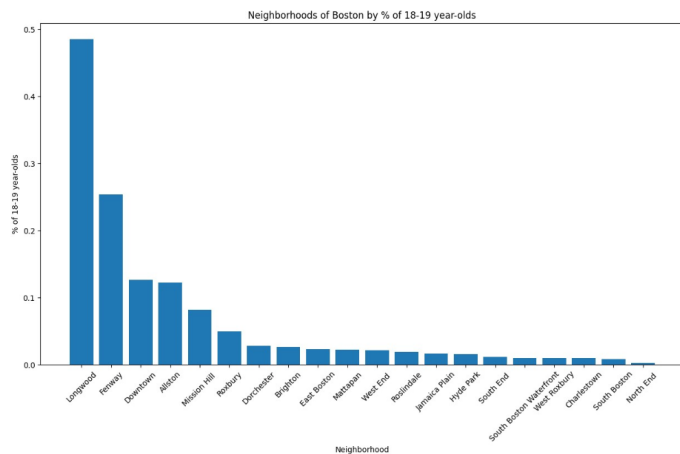
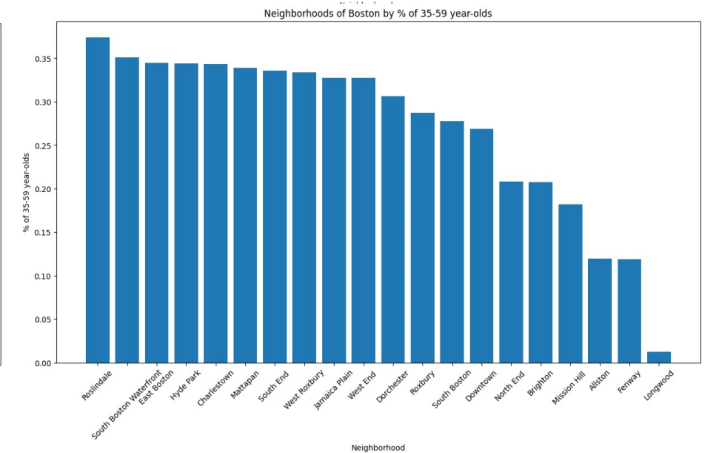
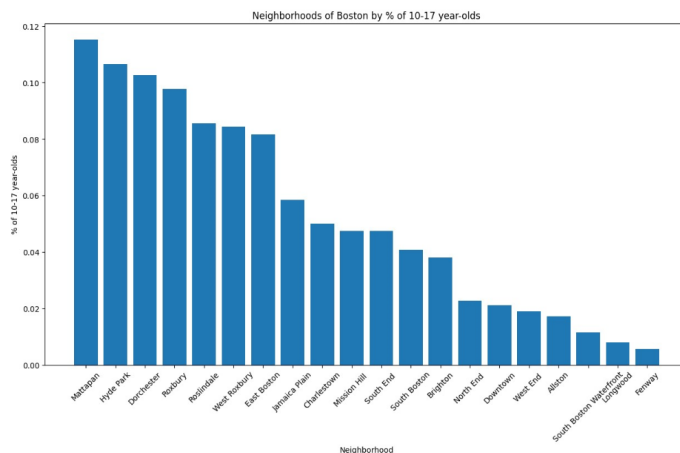
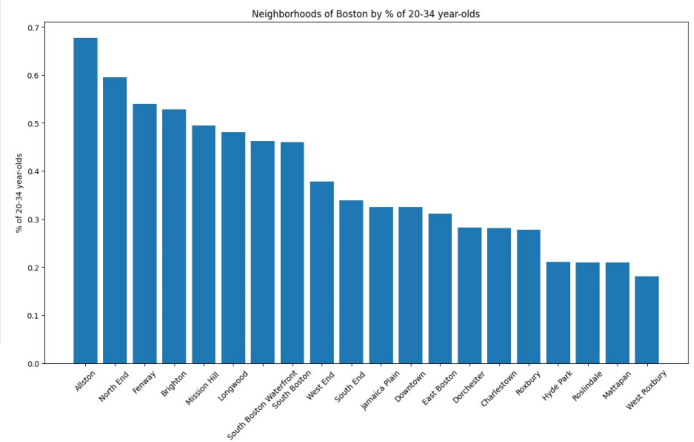
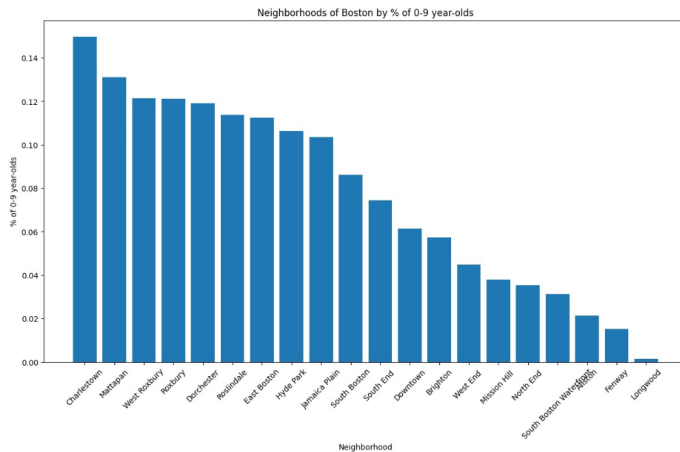
<https://www.bostonglobe.com/2023/10/01/opinion/affordable-housing-suburbs-senior-housing-zoning/> about affordable housing being disproportionately allocated towards seniors rather than younger adults or families. This, while ostensibly increasing the amount of affordable housing, does not allow younger families to thrive. We therefore wanted to see if we could find if

Boston's budget could potentially be distributed in a similar manner, with a disproportionate amount of funds being allocated towards older communities (communities which may be voting at a higher rate as well), or whether Boston is able to balance the needs of children, students, and families.

5. Visualization and insights for extension proposal



For our preliminary analysis of age demographic data, the total number of people by age group tells us very little, as it essentially conforms to the population of each neighborhood, with a slight exception for the 18-19 age group in Fenway, likely due to the population of college students, in part from this very school.



What is significantly more interesting is how Charlestown, the neighborhood with the highest proportion of 0-9 year-olds, is also the neighborhood with the highest capital budget investment plan for 2024. This warrants further investigation, since the capital budget is used for expenses such as the construction of roads and schools, so it could be that the City of Boston is investing in its youth by allocating such funding to that area.

We still have to dig into per-capita data, but right now this is slightly challenging as the capital budget definition of a neighborhood differs from the census or neighborhood demographics data definition of a neighborhood. We hope to investigate further into the specific neighborhood boundaries used to see if we can put together those two datasets and perform a more rigorous analysis.

6. The individual contribution of each team member.

Ruoxi — I helped do the question 2 and 3 in the report, and did the revenue_2.ipynb file for more analysis. I focused on processing and analyzing a fiscal year's adopted revenue budget dataset. More information about what I did specifically is included in Question 2 and 3.

Logan — Worked on all parts of the extension proposal including background research, finding data sources, generating figures, and performing data analysis. This required researching census and neighborhood demographic datasets, and formatting them properly for our analysis.

Shreyas - I cleaned data, conducted exploratory data analysis (EDA), and visualized operational budget allocations for departments and programs from FY21 to FY24.

Analyzing the top 10 departments and programs, I aimed to compare funding trends over the four years, identifying those with increased funding and those without.

Priscilla - In this report, I've worked on part 1 + 2 + 3. In accordance with the work carried out by my team, I have synthesized and addressed the key questions of the base project. In the technical domain, recently, I constructed a bar chart illustrating how each department allocates its investment across various expense categories. Additionally, I have identified challenges, limitations, expectations, and a completion plan that we have.