City of Boston: Permitting Team C
Dima Kazlouski, Aryan Agarwal, Brianna Huang, John Dohyun Kim, Jasper Hoong

Deliverable 2

Base questions:

1. What type of building permits are approved each year by type (worktype), description, valuation (declared valuation), square footage, occupancy type?
2. How have these changed over the past 5 years i.e. a year over year analysis?
3. Who is applying for building permits by geography (neighborhood, zip code, zoning district)?
4. What are the year-over-year trends visible in the zoning board of appeal approvals and denials by geography (neighborhood - listed as city, zip code, zoning district)? You'll want to normalize the data, perhaps the ratio of permits to approvals or denials, etc.
5. What are the geographic profiles of the census tracts of the addresses for the permits submitted and zoning board approvals and denials (use project address and match to census data)? Specifically, look at the race/ethnicity of the census tract income level (average income in census tracts of approved permits)

Introduction to our problem statement:

It is an undeniable truth that real estate and development play a critical role in the city of Boston's economic well-being. When discussing this topic of real estate and development, it is impossible to exclude building permits, which are required for any building to take place. However, before obtaining a building permit, there is a process every property owner must go through where they must get the approval of the Zoning Board of Appeals to confirm that they are in line with the city regulations. With this process, there are crucial questions we must consider: Who is applying for these permits, and what are the visible year-over-year trends in the Zoning Board of Appeal approvals and denials? What are the geographic profiles of the census tracts for the permits submitted and zoning board approvals and denials, focusing on the race/ethnicity of the census tract income level? As we analyze the data and answer these questions by dividing the data by geography, we delve into and investigate the patterns behind the approval or denial of these appeals, looking into things such as race/ethnicity of the census tract income level to see how they potentially are correlated with the likelihood of the Zoning Board approving or denying their requests.

Details of the data collection or cleaning steps you've undertaken:

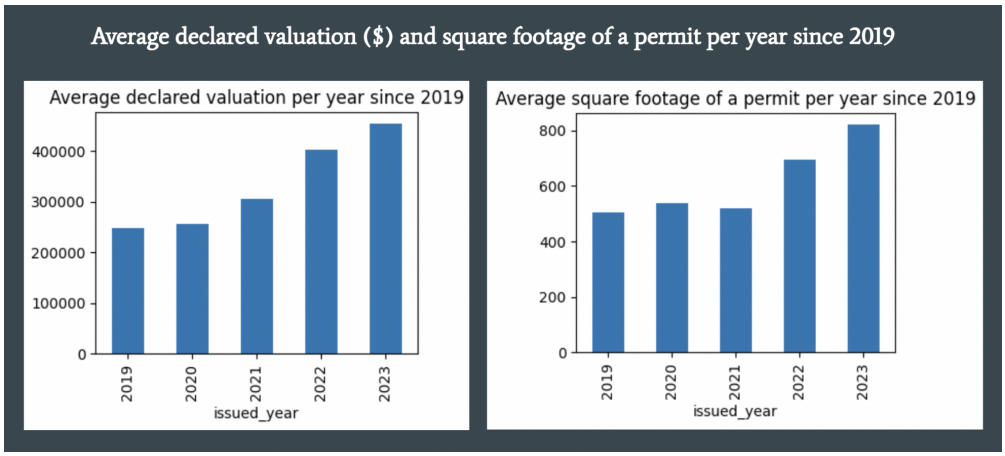
Considering that our project has multiple data sets, we had to go through various collection and cleaning steps for all data sets. Specifically, we had to locate all of our data, create separate notebooks for each dataset for the sake of simplicity, and perform cleaning steps involving:

- Handling missing values
- Converting data types (ie string of dates to date data types) or breaking them down to obtain only desired parts (ie using .split())
- Removing outliers
- Splitting the data set into different parts to only check and analyze necessary data

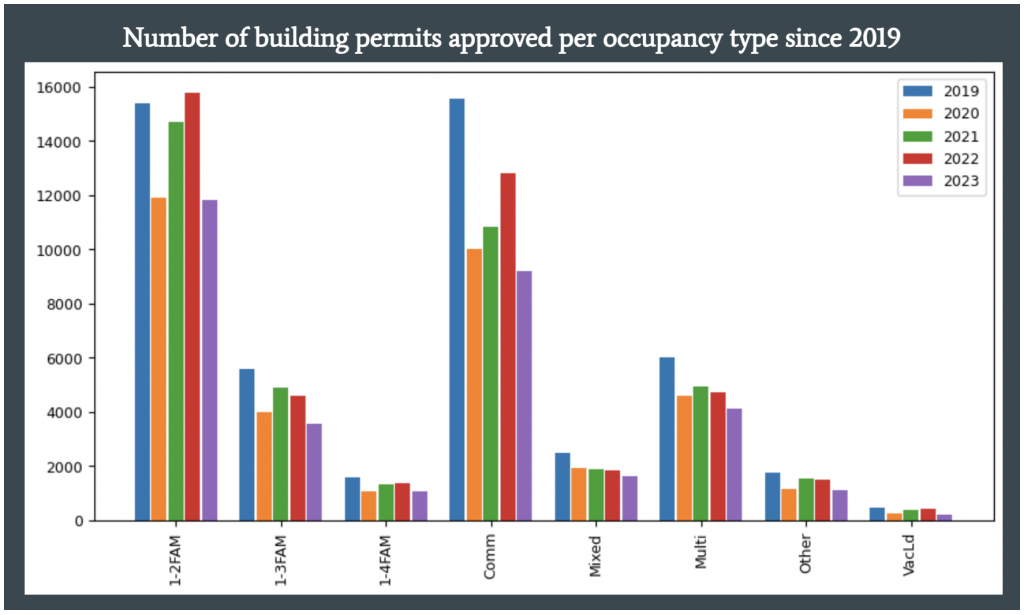
Exploratory Data Analysis:

The graphs and charts below were products of us performing EDA on our data. Once we were able to understand the basic structure of our data and generate key statistics, we were able to create visualizations relevant to our project, specifically relevant to gaining insights into the last three base questions listed at the top of this report.

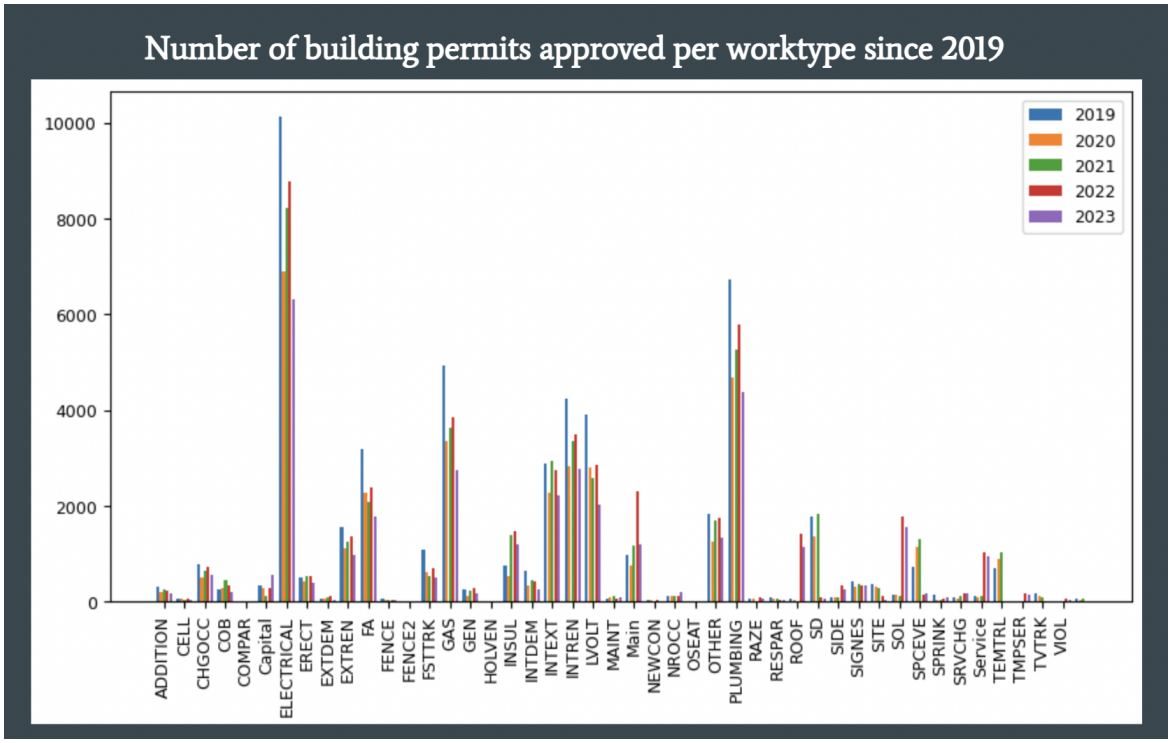
The following results are obtained from approved permits data. We aimed to create graphs to emphasize the relationship between the types of permits approved such as valuation, work type, square footage, and occupancy type, and the year it was approved, starting from 5 years ago.I obtained the dataset directly from the city of Boston's website, and was able to download it as a CSV file. From there, I converted it into a Dataframe and was able to preprocess the data and perform preliminary data analysis on the fields of interest. Here are the early results:



As expected, we see a steady increase in average declared valuation in the past 5 years, likely due to increase in housing costs and thus services for building and renovation. However, the average square footage of an approved permit has also increased, which warrants further investigation.

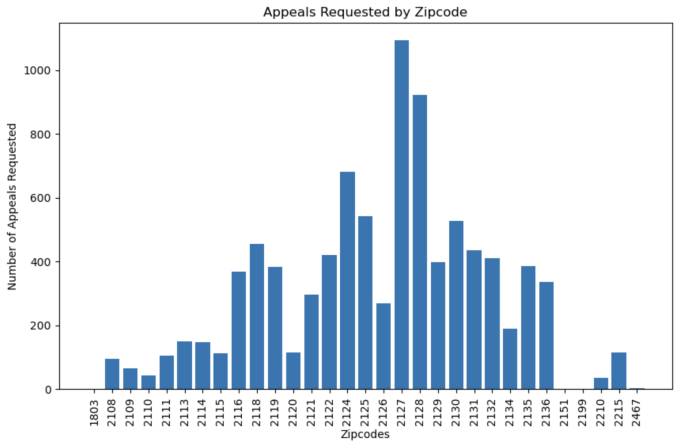


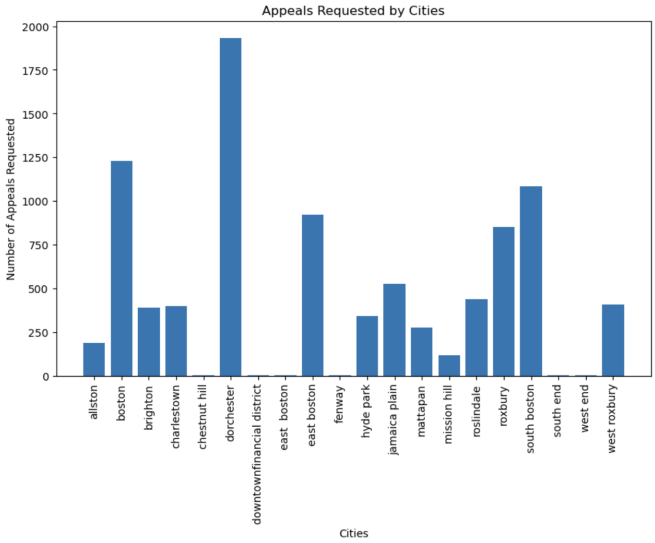
We see a pretty significant decrease in approved building permits in 2020, compared to 2019 and the following years. We would expect this to potentially be due to coronavirus and the community being quarantined, which would certainly decrease the amount of work and construction being done on buildings. Nevertheless, there remains to be a large amount of building permits for 1-2 family and community buildings compared to other occupancy types. This could simply be because there are more of those types of buildings, but could also potentially be related to accessibility to resources to getting permits approved.

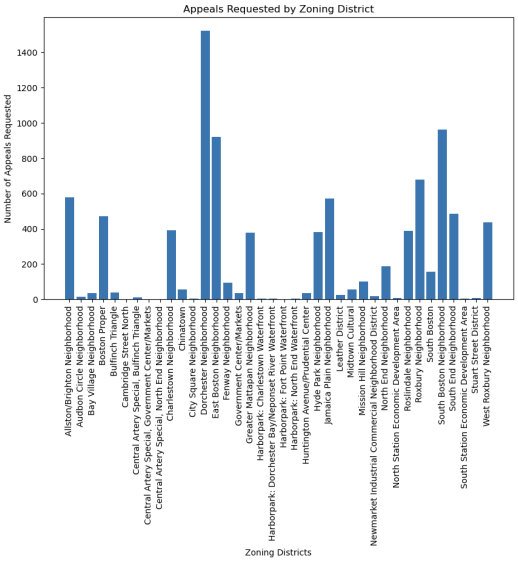


Similarly to the previous graph, we see a significant decrease in approved building permits in 2020 for potentially covid-related reasons. Otherwise, the approved permits seem mostly electrical or plumbing-related, as opposed to larger construction projects like new construction additions. It is possible that this is due to the long process of getting permits approved by the city, especially for larger projects that are expensive and affect other aspects of the city.

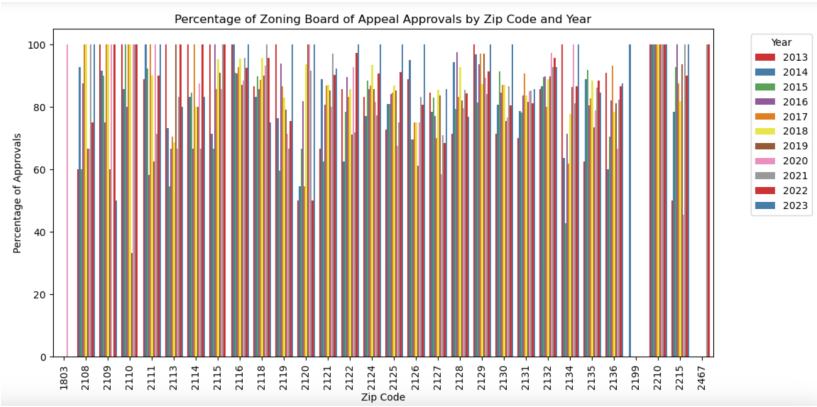
When considering who is applying for building permits by geography, we can check the number of appeals submitted to the Zoning Board of Appeals by neighborhood as it is necessary for the appeals to be approved before a building permit can be obtained. As shown below, certain zip codes, cities, and locations have a much larger number of appeals requested than others. Areas such as Dorchester, Boston (East and South included), and Roxbury, have a significantly higher number of appeals compared to areas such as Fenway, West End, South End, or Chestnut Hill. This is important when further analyzing percentage of approvals, as a smaller sample size tends to result in more extreme data as outliers tend to affect the results a lot more.

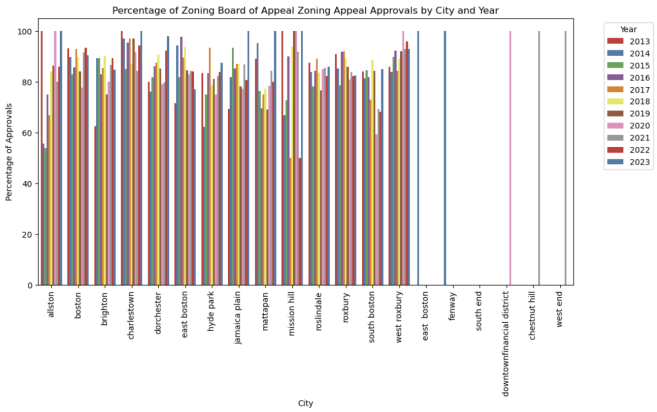




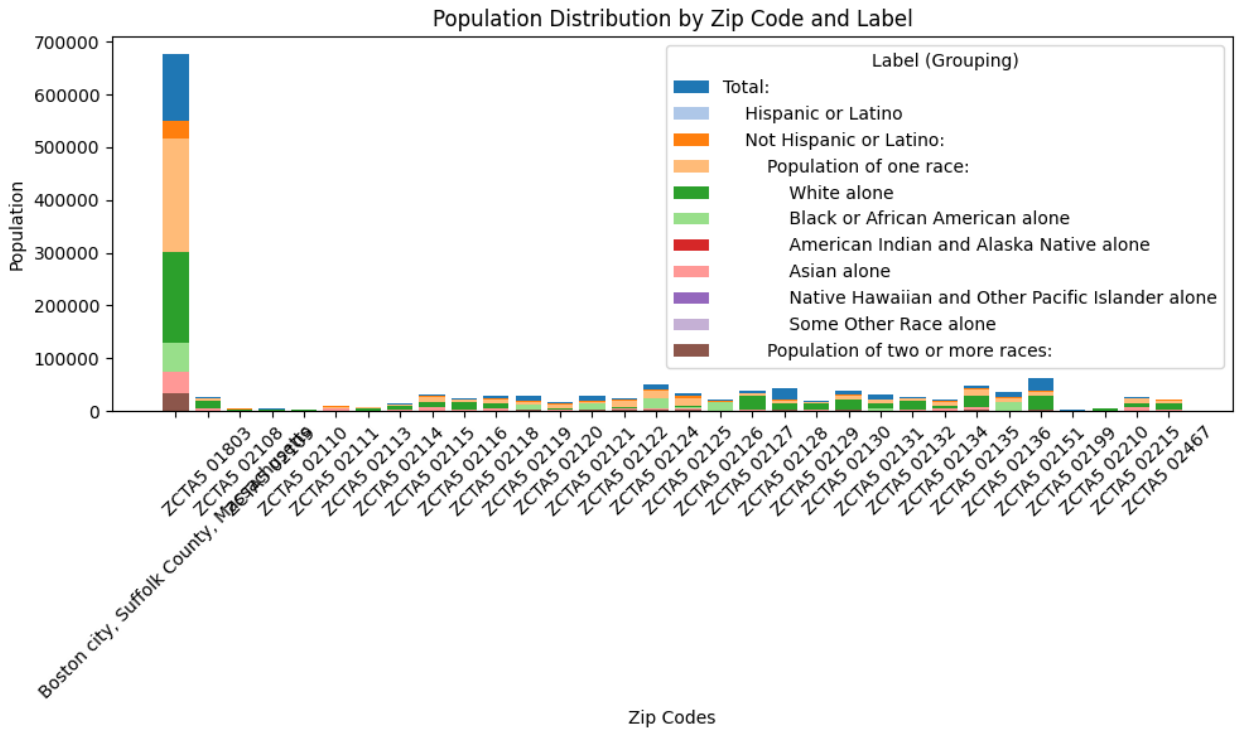


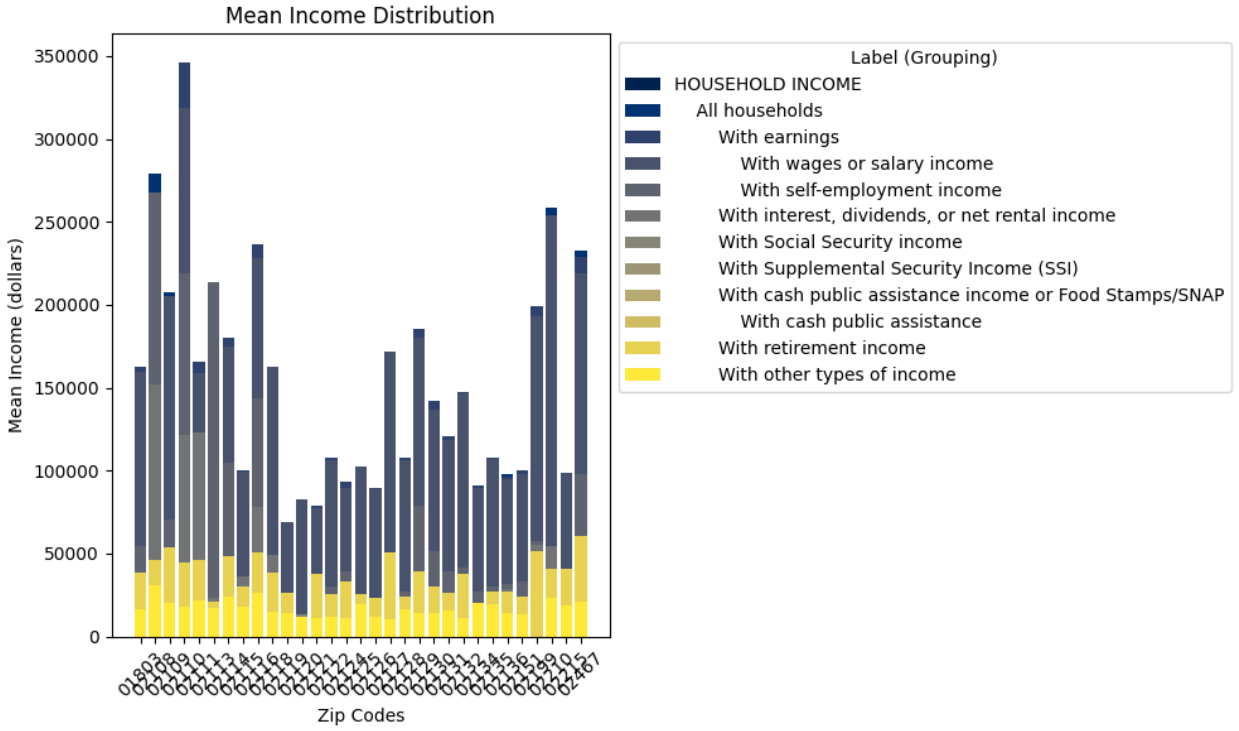
We can see the year-over-year trends visible in the Zoning Board of Appeal approvals and denials by geography by finding the percentage of appeals approved by year for each location. Shown below are graphs organized by zip code and cities, which show that overall, the approvals seem to range between 60% to 100% every year, especially when looking at them by zip code. It is worth noting that some locations such as South End and Fenway seem to be either 100% or 0% yearly which is the result of them having a small sample size as noted before when finding who is applying for these appeals by geography. As a result, if they only have one appeal submitted for the year and it is rejected, the percentage automatically becomes 0 despite the fact that if sample sizes were bigger, they may be closer to the percentages for the other locations. However, for the neighborhoods with large sample sizes, the approval rate appears to be relatively similar every year, especially in places like Boston and Dorchester with huge sample sizes.

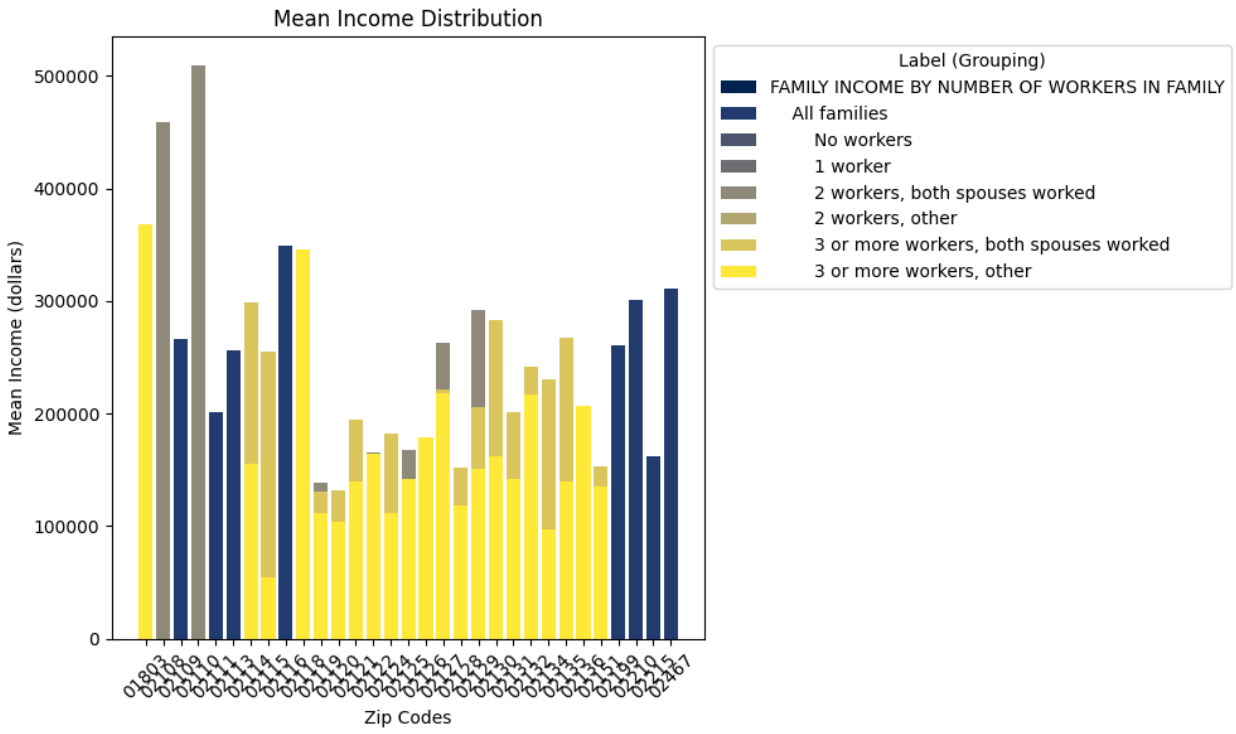


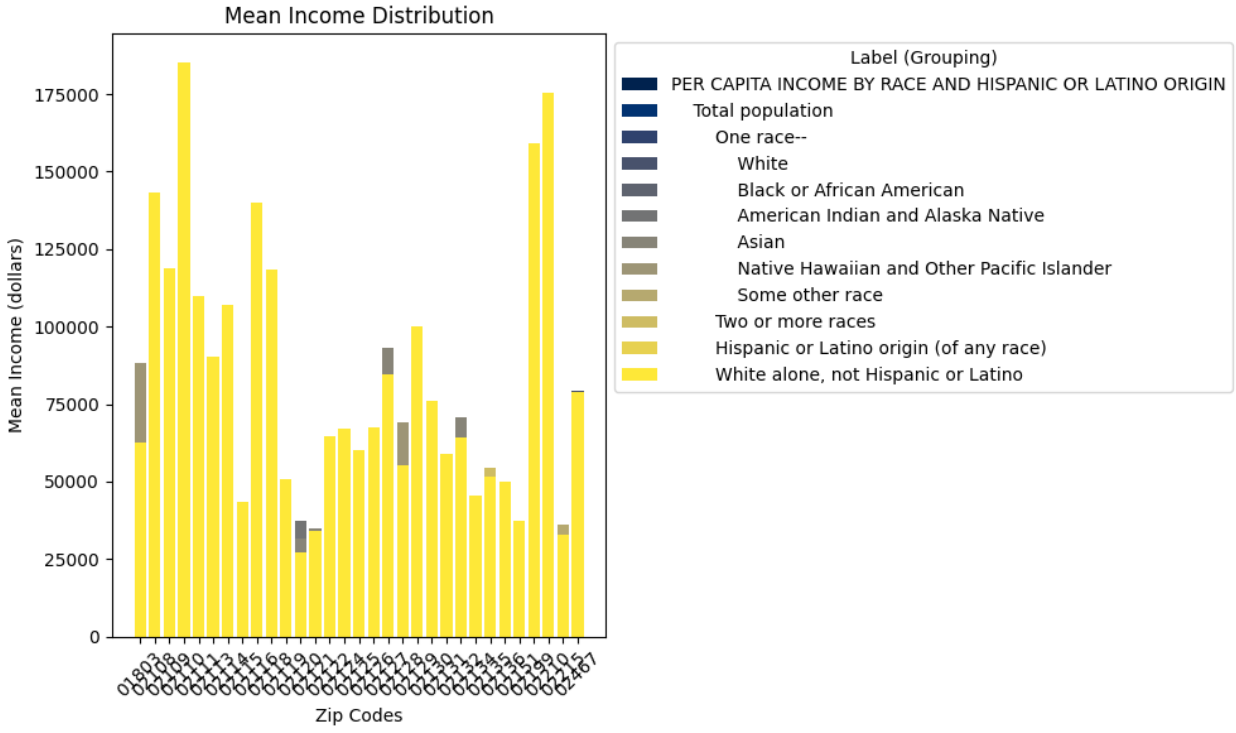


What are the geographic profiles of the census tracts of the addresses for the permits submitted and zoning board approvals and denials (use project address and match to census data)? Specifically, look at the race/ethnicity of the census tract income level (average income in census tracts of approved permits)


Population Distribution by Zip Code and Label
Population
Zip Codes
Label (Grouping)
Total:
Hispanic or Latino
Not Hispanic or Latino:
Population of one race:
White alone
Black or African American alone
American Indian and Alaska Native alone
Asian alone
Native Hawaiian and Other Pacific Islander alone
Some Other Race alone
Population of two or more races:
Boston city, Suffolk County, Massachusetts
ZCTA5 01803
ZCTA5 02108
ZCTA5 02109
ZCTA5 02110
ZCTA5 02111
ZCTA5 02113
ZCTA5 02114
ZCTA5 02115
ZCTA5 02116
ZCTA5 02118
ZCTA5 02119
ZCTA5 02120
ZCTA5 02121
ZCTA5 02122
ZCTA5 02124
ZCTA5 02125
ZCTA5 02126
ZCTA5 02127
ZCTA5 02128
ZCTA5 02129
ZCTA5 02130
ZCTA5 02131
ZCTA5 02132
ZCTA5 02134
ZCTA5 02135
ZCTA5 02136
ZCTA5 02151
ZCTA5 02199
ZCTA5 02210
ZCTA5 02215
ZCTA5 02467


Mean Income Distribution
Mean Income (dollars)
Zip Codes
Label (Grouping)
HOUSEHOLD INCOME
All households
With earnings
With wages or salary income
With self-employment income
With interest, dividends, or net rental income
With Social Security income
With Supplemental Security Income (SSI)
With cash public assistance income or Food Stamps/SNAP
With cash public assistance
With retirement income
With other types of income



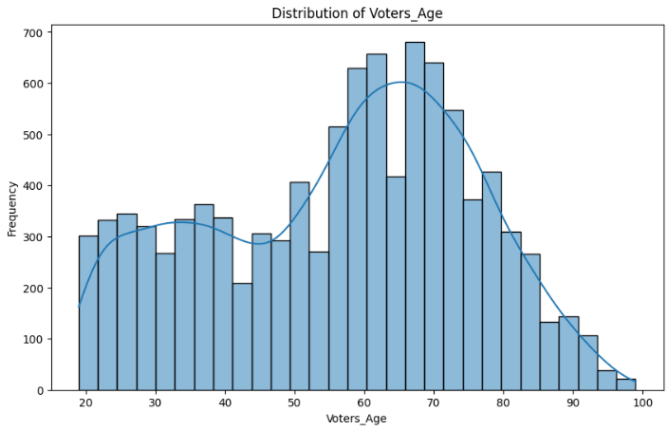


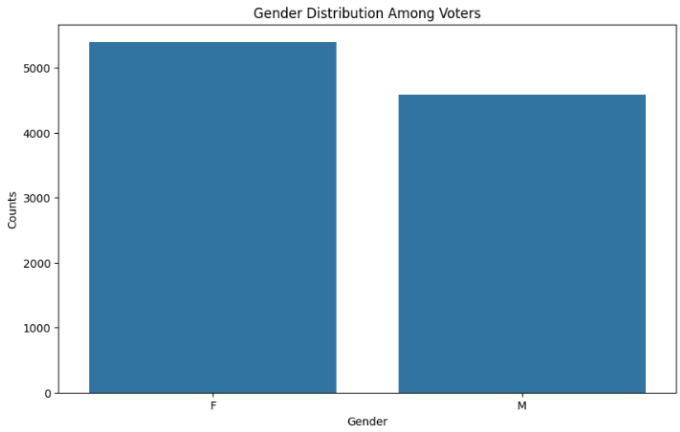
Extension Proposal:

For our extension proposal, we will look into Voter File for Boston data and conduct analysis in regards to what the voting profile for census tracts where there is a high ratio of approved permits and zoning board of appeal decisions is. Voting is important when concerning approved permits and the zoning

board of appeal as zoning regulations are largely determined by the local government and will vary from community to community. Analyzing this data would show us the correlation between the government and the approval of appeals and whether there are systemic inequalities engrained in municipal politics in permitting. By analyzing the Voter File for Boston data, we hope to find out how the local government affects approvals

Visualization and insights for extension proposal:





Very basic EDA work for voting data, we will ramp up the work heading into Checkpoint B.

Individual Contribution:

Jasper: Selectively performed EDA for census data for relevant census tracts. Added graphs of relevant census data to checkpoint A slides to answer base questions.

John Dohyun Kim: Performed EDA for zoning board of appeals data and answered third and fourth base questions. Created and added relevant graphs go the checkpoint A slides and worked on Deliverable 2, including looking into the new data set on Voting Files.

Dima: Organizing notebooks, creating visualizations for Article80 data, graphs relating to the number of permits approved for each year by project type, graphs relating to distribution of project types approved each year, and others. Loaded voting data, created notebook.

Aryan: Data cleaning and EDA work on Article80 data. Visualizations relating to the number of projects filed each year, distributions of projects by neighborhood, and others.

Brianna: Finished extracting and visualizing important data from approved permits data for the first two base project questions for Checkpoint A and Deliverable 2, and made informational slides. Started looking into the extension proposal to determine which data sets to focus on.