Deliverable 1

Question we focused on for this deliverable:

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?

What we have done as a group:

- Equally split various data sets between each group member
- Performed data cleaning on our respective data sets
- Performed EDA on our data for our respective data sets

Introduction to our problem statement:
Boston's urban landscape is molded by the myriad of building permits granted each year. Yet, essential questions linger: What types of permits, characterized by worktype, valuation, and occupancy, dominate the approvals? How have these trends evolved over the past five years? Beyond the numbers, who are the key players seeking these permits, and what geographical patterns emerge from both the permit requests and the decisions of the zoning board of appeal? As we venture into this analysis, we also strive to uncover the deeper socio-economic layers, probing into the racial, ethnic, and income profiles of neighborhoods where these permits originate.

Details of the data collection or cleaning steps you've undertaken:
Considering that our project has a multitude of 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:

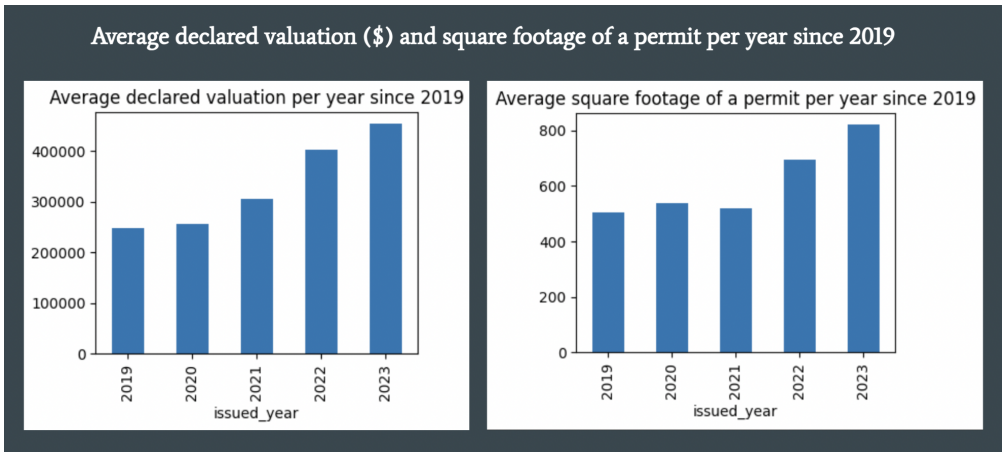
1. Handling missing values
2. Converting data types, ie string of dates to date data types
3. Removing outliers

Exploratory Data Analysis:

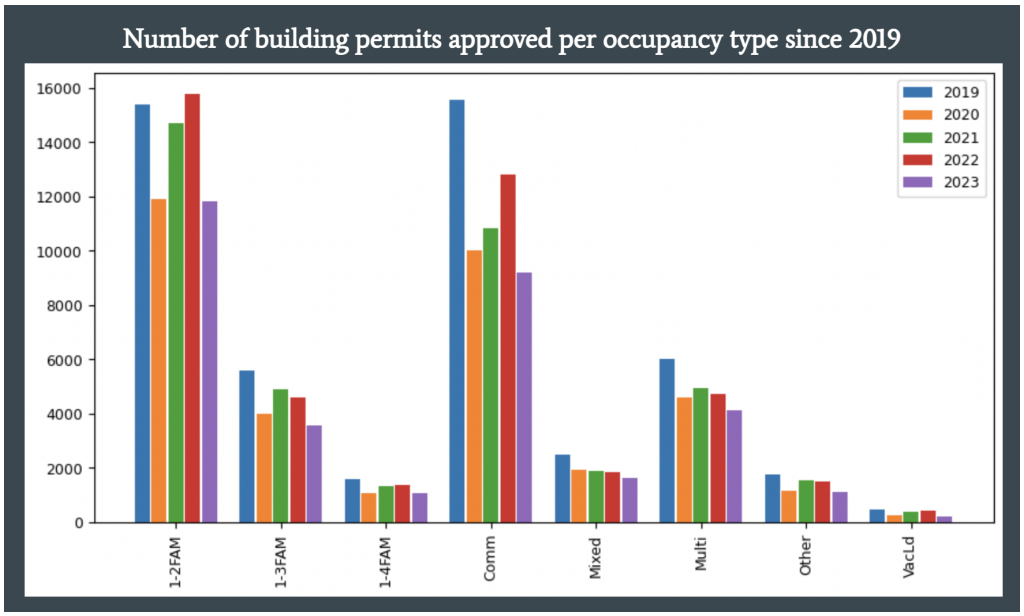
The graphs and charts below were products of us performing EDA on our data. Once we were able to understand the basic structure of our data, generate key statistics, we were able to create visualizations relevant to our project, specifically relevant to gaining insights into the first two base questions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

The following results are obtained from approved permits data. I 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:

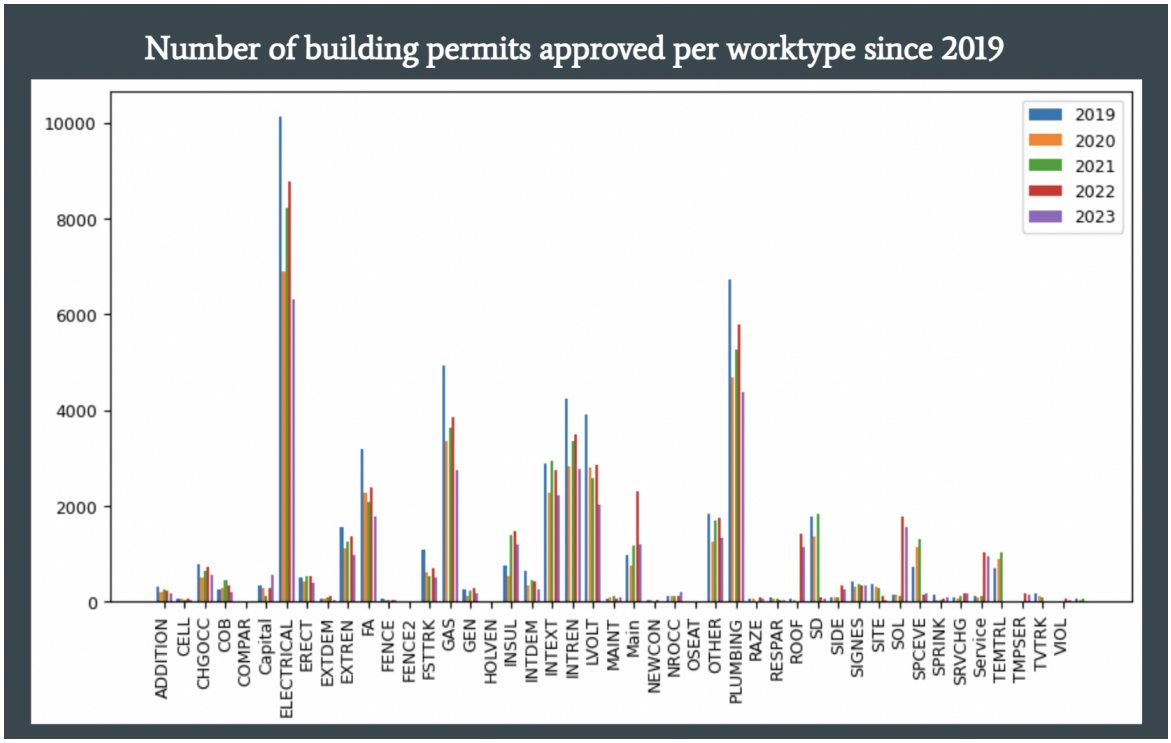


As expected, we see a steady increase in average declared valuation in the past 5 years, likely due to increase in housing costs and thus services for building and renovation. However, the average square footage of an approved permit has also increased, which warrants further investigation.



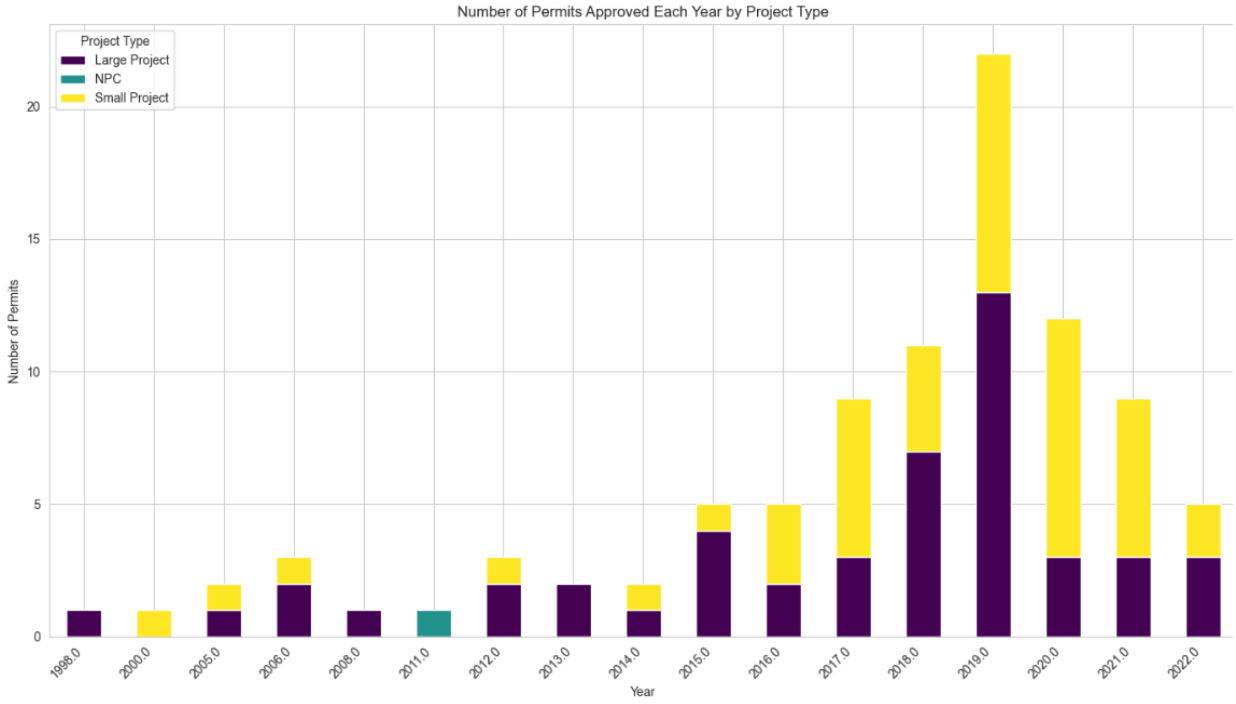
We see a pretty significant decrease in approved building permits in 2020, compared to 2019 and the following years. We would expect this to potentially be due to coronavirus and the community being quarantined, which would certainly decrease the amount of work and construction being done on

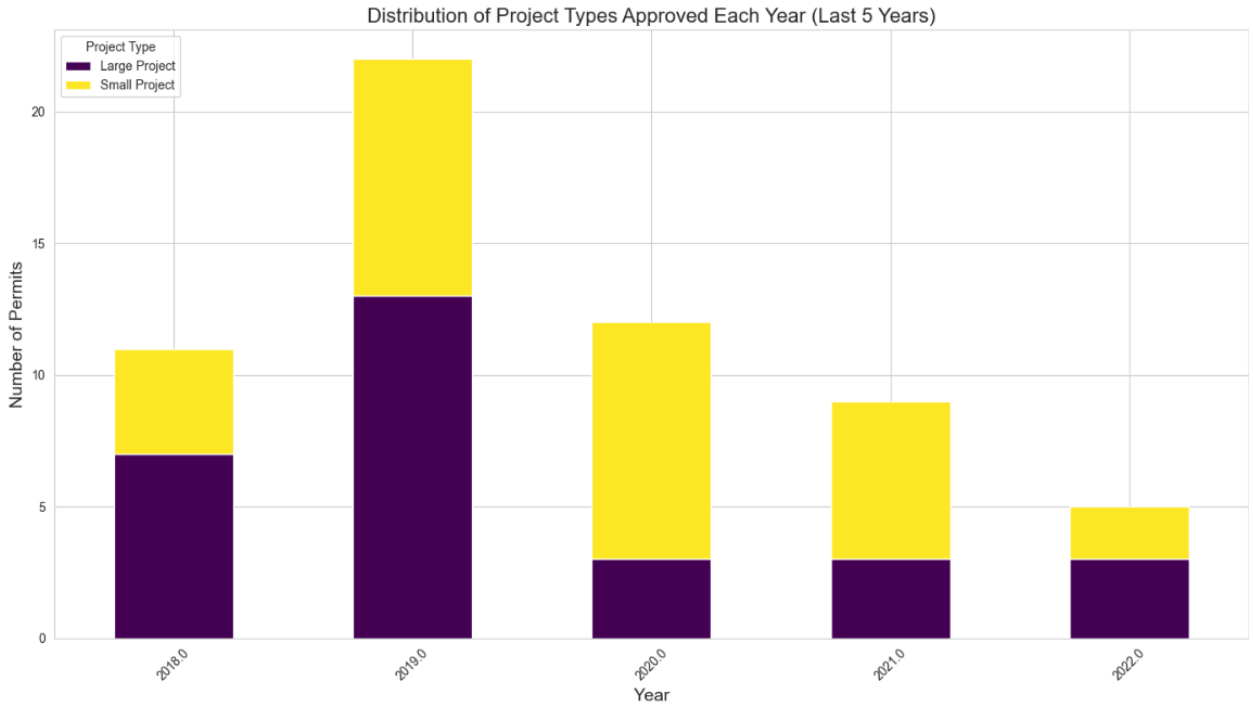
buildings. Nevertheless, there remains to be a large amount of building permits for 1-2 family and community buildings compared to other occupancy types. This could simply be because there are more of those types of buildings, but could also potentially be related to accessibility to resources to getting 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.

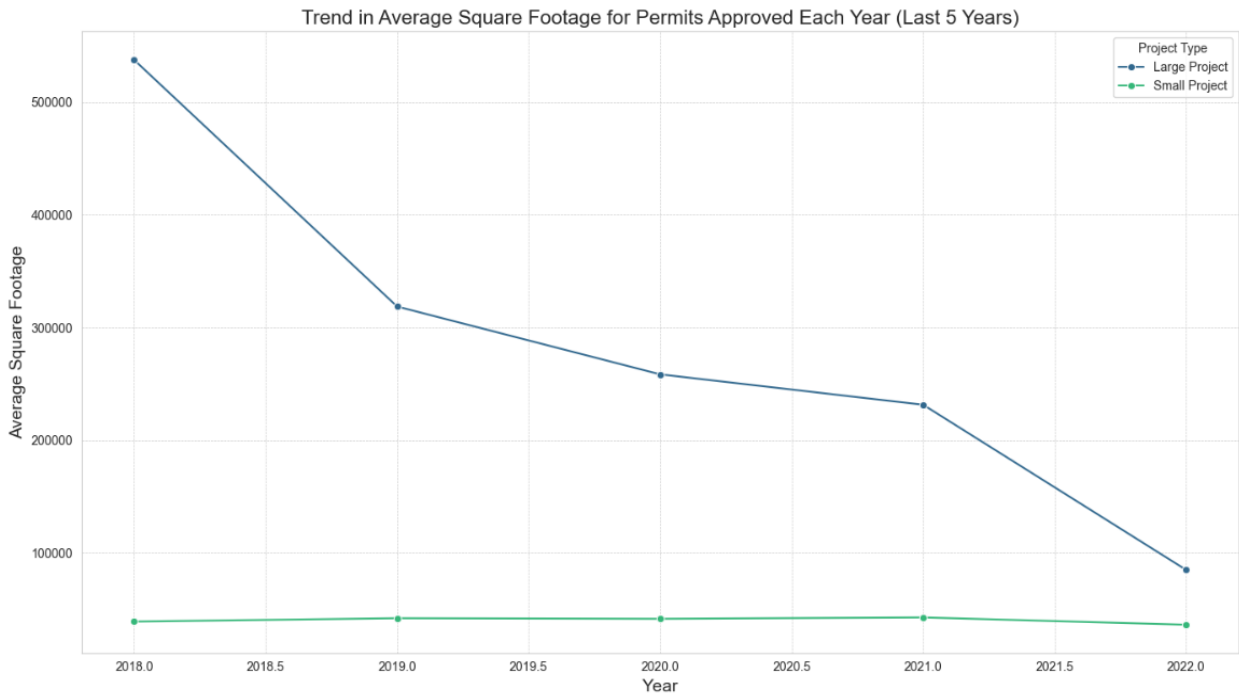
In regards to the Article80 Data:



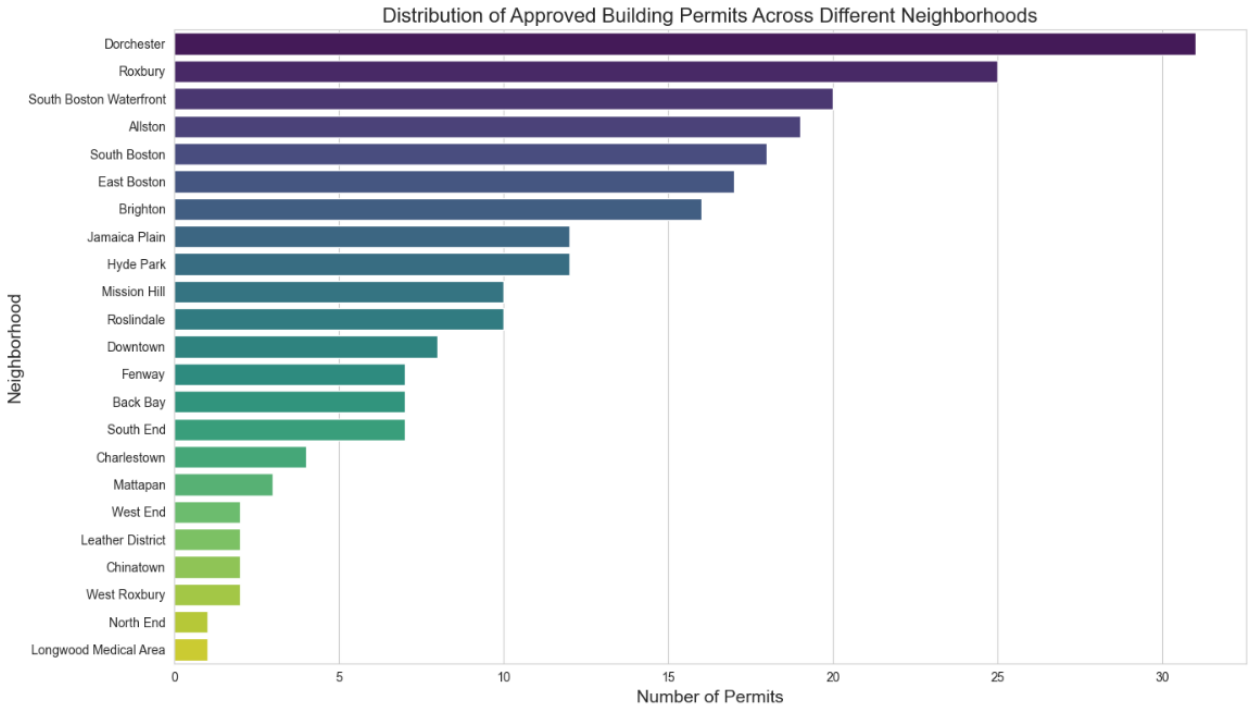


The charts above shows the distribution of building permits by project type over the years. We can observe which project types are more prevalent in certain years, indicating trends or shifts in development focus. Noticeably, over the last 4 years there has been a year to year decrease in small

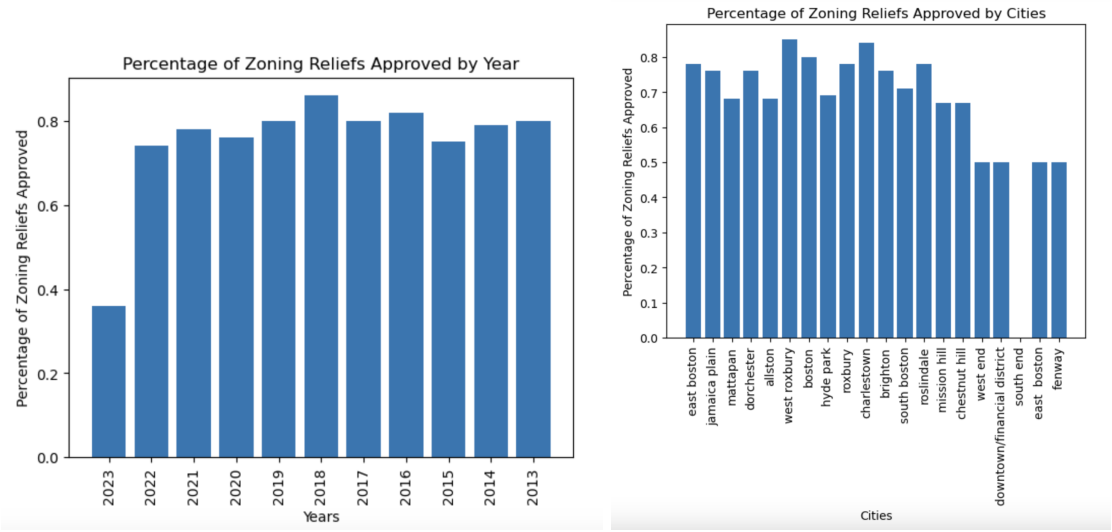
projects. There has been a large drop in the number of large projects between 2019 and 2020 but that number has not changed between 2020 and 2022.



This chart showcases how the average size (in square footage) of approved projects has changed over the last five years. Notably, the average square footage for large projects has been decreasing for the last 5 years, however there has been almost no change for small projects.



This visualization highlights which neighborhoods have the highest number of approved building permits. It helps identify areas of high development activity. The article80 data set does not specify which projects get denied, further analysis on the other data sets will be conducted to answer this question.



From the zoning board of appeal tracker data I created two graphs that show the percentage of approved applications. An approval in zoning relief allows the applicants to continue with obtaining a permit hence its importance. The first graph shows the percentage of zoning reliefs granted by year. The second shows the reliefs granted by the city. As displayed by the high percentage of zoning relief approvals yearly, it is likely for the zoning reliefs to be approved (2023 is low because the year has not

yet completed and many requests are still pending). Similarly, zoning reliefs from most cities are approved as well, although some cities have had less luck in getting approved (although it is worth noting that the sample size for some cities were greater than others; for example, South End had very few applications)

Individual Contributions:

Brianna: EDA and and other work on permits data including downloading, cleaning, performing all EDA on the permits data. Filtered and plotted data for analyzing relationships between building permits approved valuation and their geography. This included zip code, city, and worktype.

John: EDA and other work on zoning data, analyzing by zoning districts and city correlation to approvals.

Jasper: EDA and other work on census data including locating the relevant census data for the city of Boston. Unlike the other datasets, census data required going in deeper to find the relevant data.

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.

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.