Deliverable 2 - Boston Permitting D

1. Introduction

Real estate development significantly contributes to the economic growth of many U.S. cities, including Boston. In Boston, numerous new buildings and renovations are constructed and conducted over days, months, and years. Commencing any development project in Boston requires acquiring an official permit from the city, a process encompassing several steps such as approval stages, community involvement, and public hearings. In this project, the city aims to analyze the aforementioned process in depth.

According to the City of Boston, the permitting process begins with submitting an application and paying the underlying fees. This is followed by a review of the plans against the city's zoning laws. At this stage, applications not adhering to the zoning regulations may lead to permit denial. The development project can only start once an official permit card has been issued, valid for six months and subject to extension.

For rejected applicants, there is an option to pursue an appeal. The appeal process involves filing the appeal, engaging in a community process, undergoing a public hearing, and awaiting the final decision by the Zoning Board of Appeal (ZBA). A new one can be filed after a year if the appeal is denied. Furthermore, larger projects exceeding 20,000 square feet must undergo the Article 80 review process, a significant focus of this analysis project.
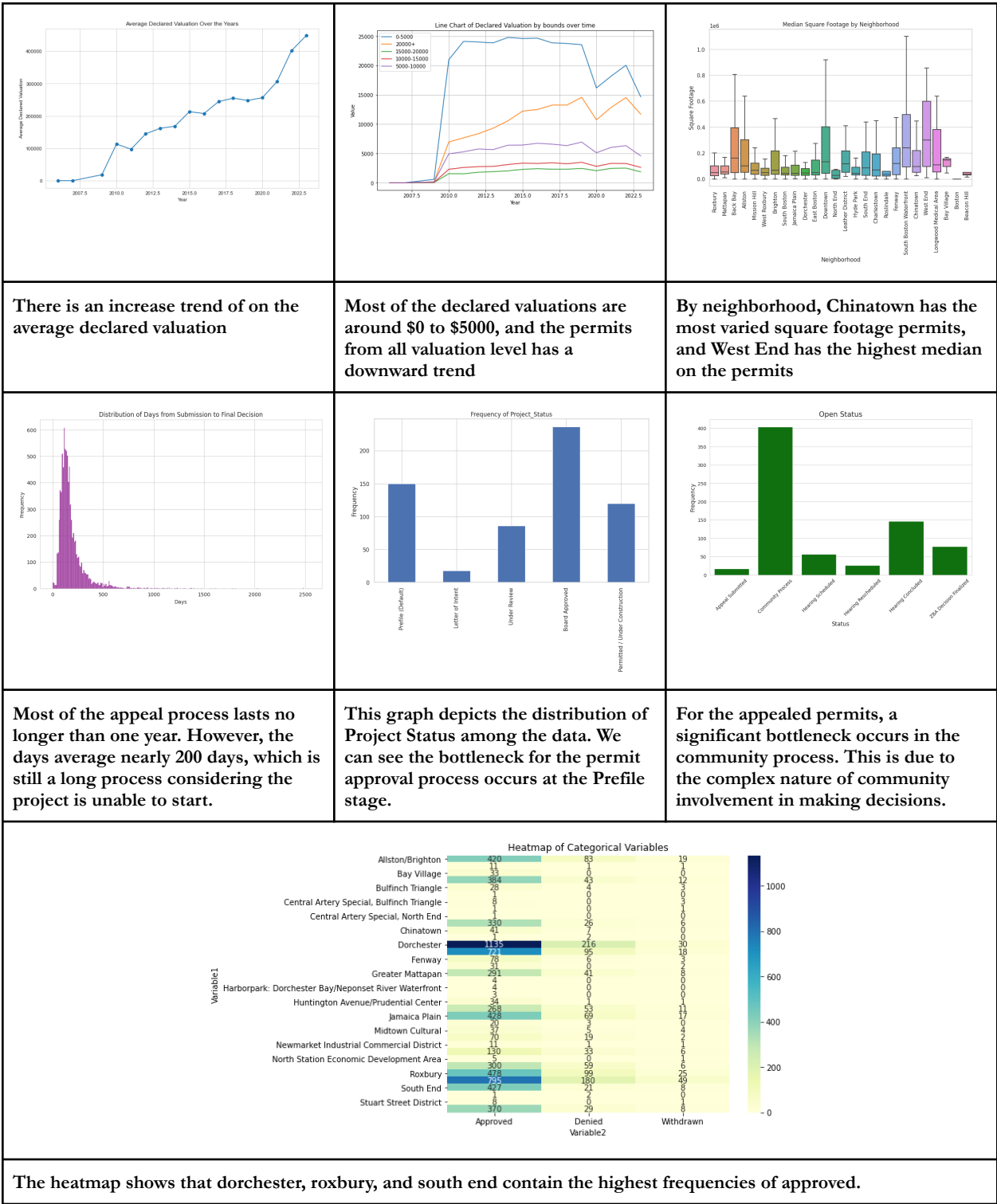
The scope of the analysis of Boston Real Estate Permitting extends to societal, political, and environmental issues, which are pivotal for the future growth of Boston.

The project utilizes several datasets for comprehensive analysis:

No	Name	File	Source	Details
1	Approved Building Permits	abp.csv	Analyze Boston	Data on approved building permits, including location, date, fees, valuation, work type, etc.
2	Article80 Development Projects	a80.csv		Data on the development projects subject to Article 80 review process, including location, date, status, size, etc.
3	Zoning Board of Appeal Tracker	zba.csv		Data on denied permit applications undergoing the appeal process.
4	Census Data for Demographics	census.csv	US Census	Demographic data for the Boston locations analyzed in the project
5	Income Data per ZIP	IRS.csv	IRS	Per-capita earned income by ZIP code extracted from IRS records

For the Approved Building Permits dataset, unnecessary columns were dropped, numeric columns were cleaned, and new columns were derived from dates while retaining certain columns for future analysis and keeping null values. Similarly, the Article 80 Development Projects dataset received simplification and renaming of columns, date-derived columns, and retention of specific columns for future use, with null values maintained. The Zoning Board of Appeal Tracker underwent a more extensive clean-up, including removing unnecessary columns, converting categorical values to numerical values, correcting human errors, and refining values in certain columns. Lastly, the Census Data for Demographics was processed by renaming columns, joining demographic data with shapefiles, removing unrelated columns, and normalizing data to get demographic proportions for census block groups. This initial and continuous cleaning process was crucial for ensuring the datasets were clean, relevant, and amenable to further analysis in the project.

2. Exploratory Data Analysis

There is an increase trend of on the average declared valuation

Most of the declared valuations are around $0 to $5000, and the permits from all valuation level has a downward trend

By neighborhood, Chinatown has the most varied square footage permits, and West End has the highest median on the permits

Most of the appeal process lasts no longer than one year. However, the days average nearly 200 days, which is still a long process considering the project is unable to start.

This graph depicts the distribution of Project Status among the data. We can see the bottleneck for the permit approval process occurs at the Profile stage.

For the appealed permits, a significant bottleneck occurs in the community process. This is due to the complex nature of community involvement in making decisions.

The heatmap shows that dorchester, roxbury, and south end contain the highest frequencies of approved.

3. Analysis Questions

1. What building permits are approved yearly by type (work type), description, valuation (declared valuation), square footage, and occupancy type?

A.   Work Types & Description

As shown in Figure 1, the top work types for approved building permits are electrical, plumbing, and gas, followed by conversion and interior renovation. Due to the frequent need for improvements and maintenance in building facilities, as well as compliance with safety regulations, it's typical for work types like electrical, plumbing, and gas to be common.
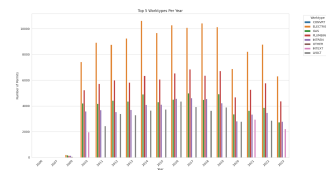
Figure 1. Top 5 Worktypes Per Year

B.   Declared Valuation

As illustrated in Figure 2, the top declared valuations for approved building permits are 0-1,750 USD, 1,750-3,500 USD, and 3,500-5,250 USD. Despite some outliers with significantly high declared valuations, the average declared valuation of the approved permits generally remains lower.
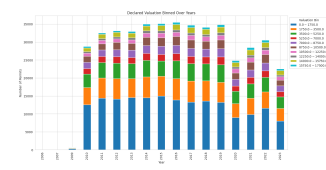

Figure 2. Top 5 Valuations Per Year

C.   Occupancy Types

Figure 3 depicts the top occupancy types for approved building permits are 1-3 family residential buildings, multi-family residential buildings, mixed-use developments, and commercial properties. Considering Boston's common types of buildings, it makes intuitive sense that residential and commercial buildings would have the highest frequency in obtaining approved permits.
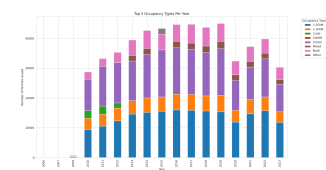

Figure 3. Top 5 Occupancy Types Per Year

2. How have these changed over the past five years i.e. a year-over-year analysis?

A.   Work Types & Description

Reviewing Figure 4, over the last five years, there has generally been a declining trend in the number of approved permits across the work types and descriptions. However, electrical, plumbing, and gas works remain consistently the most commonly approved.
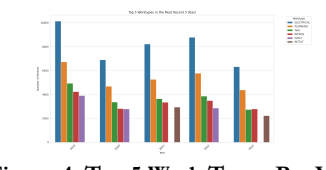

Figure 4. Top 5 Work Types Per Year (Recent Five Years)

B.   Declared Valuation

For declared values for the approved permits, there has been a steady increase each year over the five-year period. While the average declared value had a modest rise, the spike in the values of outliers is substantial.
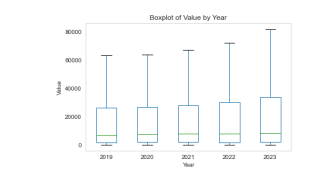

Figure 5. Top 5 Valuations Per Year (Recent Five Years)

C.   Occupancy Types

Looking at Figure 6, which shows the data based on occupancy type, the number of approved permits peaked around 2019. Subsequently, a decline through 2020 and a recovery in 2022 occurred. An assumption can be made that the COVID-19 pandemic was the potential cause of this fluctuation in recent years.
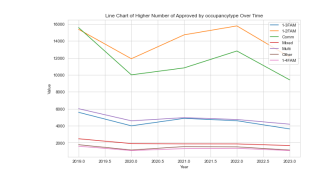

**Figure 6. Top Occupancy types for past 5 years**

3. Who is applying for building permits by geography (neighborhood, zip code, zoning district)?

Historically, most approved permits have been in Boston, Dorchester, and Roxbury. Specifically concerning permits under the Article 80 review process, they predominantly originate from Dorchester and Roxbury. Additionally, permits appealed to the zoning board are primarily from Dorchester, Boston, South Boston, and Roxbury. Over the last five years, South and East Boston have shown a notable increase in permits falling under the Article 80 process, leading to permit approvals during this period. Based on the geography, we were assuming that in urban areas, more permits are applied and approved, whereas the suburban area has less permits applied and approved. The region that is closer to central Boston will have a higher number of approved permits.
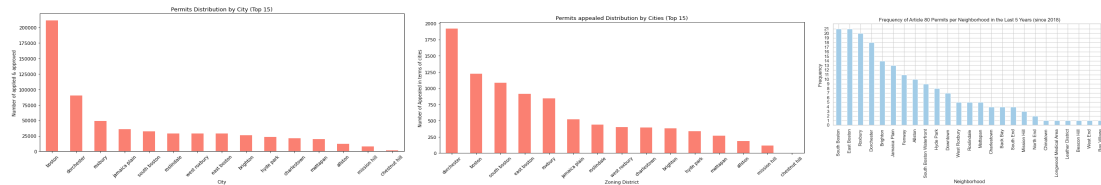
4. What are the year-over-year trends visible in the Zoning Board of Appeal approvals and denials by geography (neighborhood - listed as city, zip code, zoning district)?

The year-over-year trend concerning approved permits versus denials on the zoning board exhibits distinct variations across different regions. For instance, the South End experienced minimal approved permits, apart from an extreme peak in 2019. In contrast, West Roxbury showcased separate peaks in 2016 and 2021. Moreover, Dorchester witnessed a recent increase in the approval ratio, specifically in 2022.
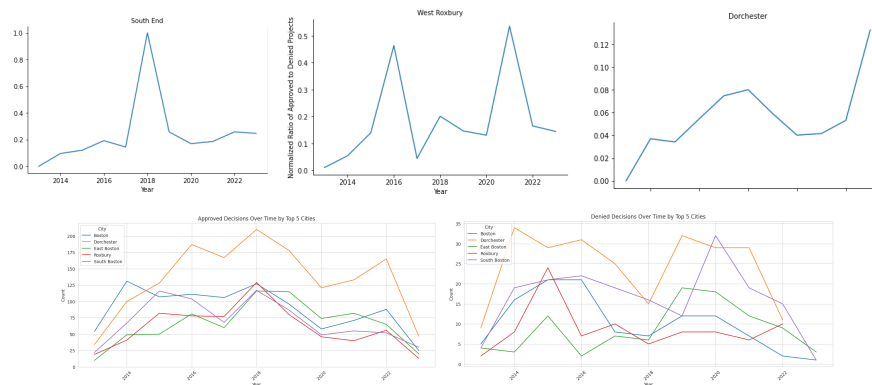
4. What are the geographic profiles of the census tracts of the addresses for the permits submitted and zoning board approvals and denials?

The geographic profiles for the permits are shown in figures 7 and 8. These plots were constructed using census data for block group divisions and the IRS tax return data for per-capita income by ZIP code. Geographic distributions were constructed for all census-defined demographic groups and were used for further analysis.
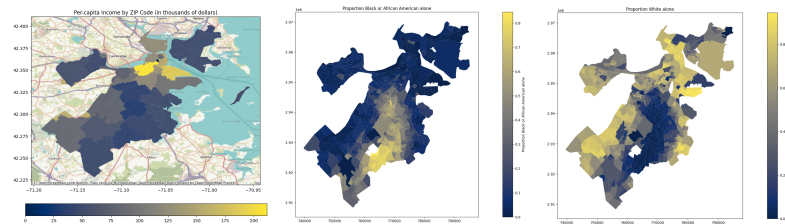

Figure 7, 8 Income and Demographic Distributions

Spatial joins between the development coordinates and the census block groups / ZIP code shapes were used to calculate the number of developments within each area. The Lorenz curves for approved and rejected developments given below show how the CDF of income ranked from lowest to highest varies with the CDF of developments (P-P plot). Additional Lorenz graphs of population by demographic group vs development distribution were also produced. The linear graph in the middle represents perfect equality and deviation of the real curve shows unequal distribution. The x and y-axes show cumulative percent with respect

to each distribution. By finding the Gini coefficient for each of these curves, we measure income and demographic inequality in the developments. The Lorenz curves for approved and rejected requests by increasing income level are below.

Lorenz Curve for Approved Requests by Income

Lorenz Curve for Rejected Requests by Income
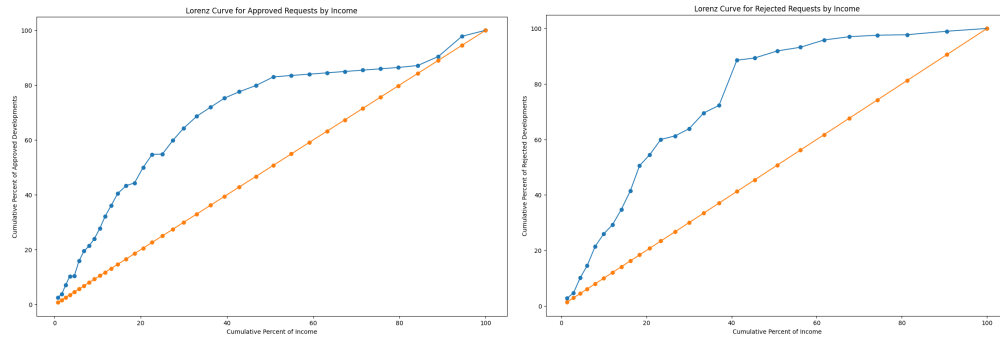
Both distributions are skewed, where more developments occur in ZIP codes with lower income levels. The bottom 50% of incomes account for 80% of approved and 92% of rejected permitting requests. There is clear economic inequality since permit rejections are more likely to occur in lower-income ZIP codes. The higher levels of development in lower-income areas can be explained by city renewal projects, gentrification, or lower development costs. The Gini coefficient for approved is -0.395, and for rejected is -0.512 (negative sign used to show skew toward lower income). We can see that the rejected requests distribution is more unequal than approved with respect to income.
Two Lorenz curves for the distribution of developments vs population by demographic group are below.

Lorenz curve for White alone

Lorenz curve for Black or African American alone
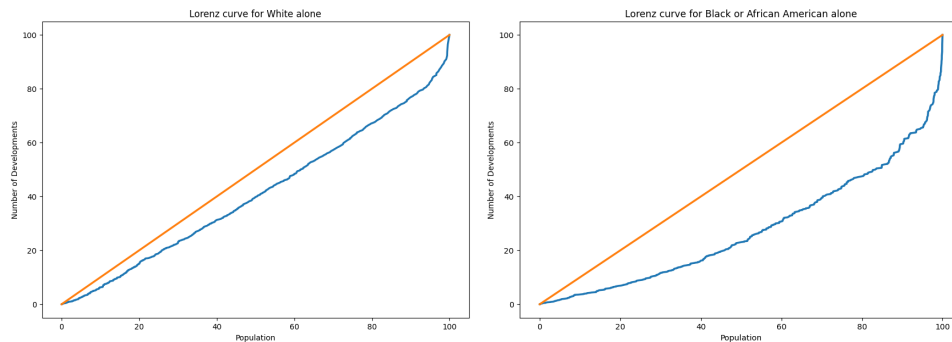
The Lorenz curve relating the population of the "white alone" group is much more favorable than the curve for the "Black or African American alone" group. The Gini coefficients for these two plots are 0.181 and 0.446, showing a much higher level of inequality for the Black or African American group than the white alone group with respect to the development distribution.
The total Gini coefficient with respect to developments for the population was 0.303. The Gini was 0.409 for the Hispanic or Latino group, 0.335 for the Asian alone group, 0.409 for American Indian and Alaska Native alone and 0.455 for Native Hawaiian and Other Pacific Islander alone group. White alone was the only demographic group with a Gini lower than the total. These results show unequal distributions of developments with respect to different demographic groups. Lower Gini for the white alone group shows that developments are more evenly proportional to population than for other demographic groups.

4. Proposal for Extension Projects on Boston Permitting Dataset

Our initial findings in Boston's real estate permitting process have provided insights into the approved, appealed, and Article 80 building permits. As we extend this project, we propose to look into the decision-making factors for permit approvals and rejections from the machine learning perspective. Also, as a second extension, we will explore the correlation between COVID-19 data and the permitting process.

- **Extension Project 1: Decision Tree Analysis for Permit Outcomes**

The first extension project will combine the Approved Building Permits (abp.csv) and the Zoning Board of Appeal Tracker (zba.csv) datasets. We will create a boolean 'approved' column, assigning 'true' for data from abp.csv and 'false' for zba.csv. A decision tree algorithm will be utilized to identify and illustrate the key factors leading to permit outcomes. This analysis aims to discover the factors influencing the permitting decisions, which will be helpful for predicting future approvals and rejections.

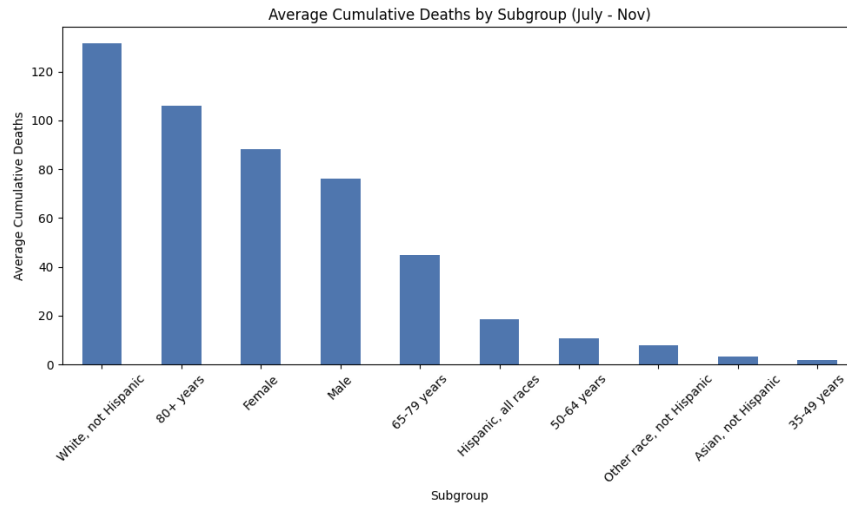- **Extension Project 2: Correlating COVID-19 Impact with Permitting Data**

The second extension will establish a correlation between Massachusetts COVID-19 data and the Boston permitting dataset. By integrating trends in COVID-19 cases with permit application counts, issuance dates, and denials, we aim to understand how the pandemic has affected the Boston permitting process. This analysis will be crucial in assessing the resilience of the city permitting process and the city's responsiveness to unprecedented health crises.

Summary of Extension Projects:

Objectives
- Using a decision tree M.L. model, identify the predictive factors for permit decisions.   - Analyze the impact of the COVID-19 pandemic on the permitting process in 2020-2023   - To provide actionable insights to optimize the permitting process from societal, political, and environmental perspectives.
Methodology
- Data Integration: - Merging abp.csv and zba.csv datasets and derive an 'approved' column.   - Combining the COVID-19 data with the permit data.   - Decision Tree: Utilize machine learning model to predict permit outcomes through features.   - Correlation: Analyze the correlation matrix between COVID-19 data and permitting data.
Expected Outcomes
- A fairly accurate decision tree model illustrating the importance of each feature.   - Insights from the correlations found between COVID-19 and Permit data.   - Recommendations for the city to implement in future permit decisions and permitting processes during future pandemics.

5. Extension Project Preliminary Analysis and Visualizations

Correlating COVID-19 Impact with Permitting Data	
monthly Confirmed Deaths	monthly Confirmed Cases
There is a huge amount of deaths at the beginning of the pandemic; and there are three death peaks.	There are two case peaks, and the one that with the highest is when Omicron emerged
Yearly Confirmed Deaths	Yearly Confirmed Cases
The confirmed deaths on a yearly basis has significantly decreased.	The yearly confirmed cases are peaked at 2022, but then dropped at 2023
Comparison of Tests and Percent Positivity	Cumulative Confirmed Cases over Time
We can see tests vs percent probability in this graph for Covid is Massachusetts. We can see as tests go up the percent positive falls.	This is the cumulative confirmed cases of Covid through the past few months. We see an uptrend of confirmed cases as the winter months start.

Here are the average cumulative Covid-19 deaths by subgroup. We want to examine further the impact of Covid-19 in different areas, and whether its effect has been disproportionate in areas with different concentrations of subgroups. In this data we see the highest cumulative deaths among white, not hispanic people, which is surprising as typically white demographics have better survival rates compared to many ethnic minorities in the United States. Further analysis is needed to explain this discrepancy.
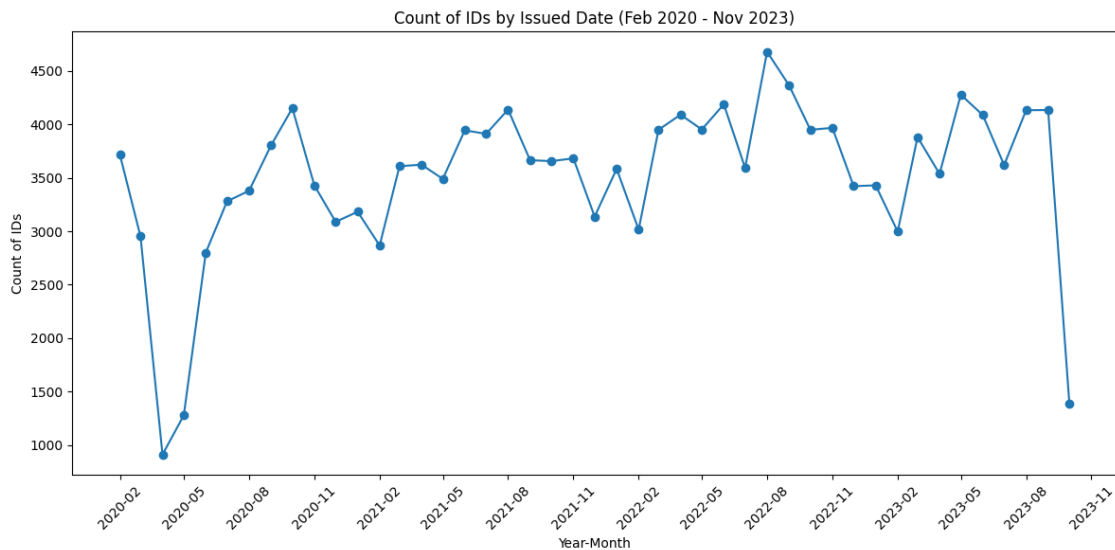
When compared with the COVID-19 graph, there is a large correlation shown when the number of approved permits drop around March and April of 2020, the initial spike in pandemic confirmed deaths.

Decision Tree Analysis for Permit Outcomes

merged.csv = abp.csv and zba.csv merged

- The object_id, city, zip, description, and Approved columns from the abp.csv have been merged with the boa_apno, city, zip, project_description, and Approved features from the zba.csv.

	ID	city	zip	Description	Approved
0	1	boston	02109	Interior/Exterior Work	1
1	2	roxbury	02132	City of Boston	1
2	3	charlestown	02129	Other	1
3	4	roxbury	02118	Interior/Exterior Work	1
4	5	brighton	02135	Renovations - Exterior	1

	ID	city	zip	Description	Approved
631311	BOA264014	boston	02111	Change from residential to office family assoc...	0
631312	BOA262723	hydepark	02131	Change occupancy from single family to two Blo...	0
631313	BOA262758	boston	02111	Remove proviso to this petitioner only per pla...	0
631314	BOA262762	boston	02116	Underground basement addition An indoor pool 4...	0
631315	BOA217901	roxbury	02132	WHEN PLANS COME BACK FROM BOA DW OP HOLD APPLI...	0

- Preview of approved rows in merged.csv

- Preview of rejected rows in merged.csv

Words used in Approved

Words used in Rejected
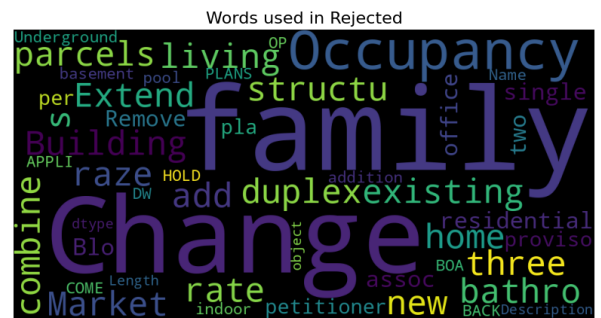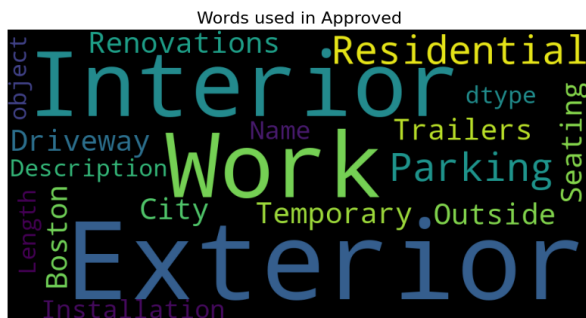
Dominant Words: "Exterior", "Interior", "Work", "Residential", "Renovations"

Other Notable Words: "Installation", "Driveway", "Parking", "Temporary", "City", "Boston"
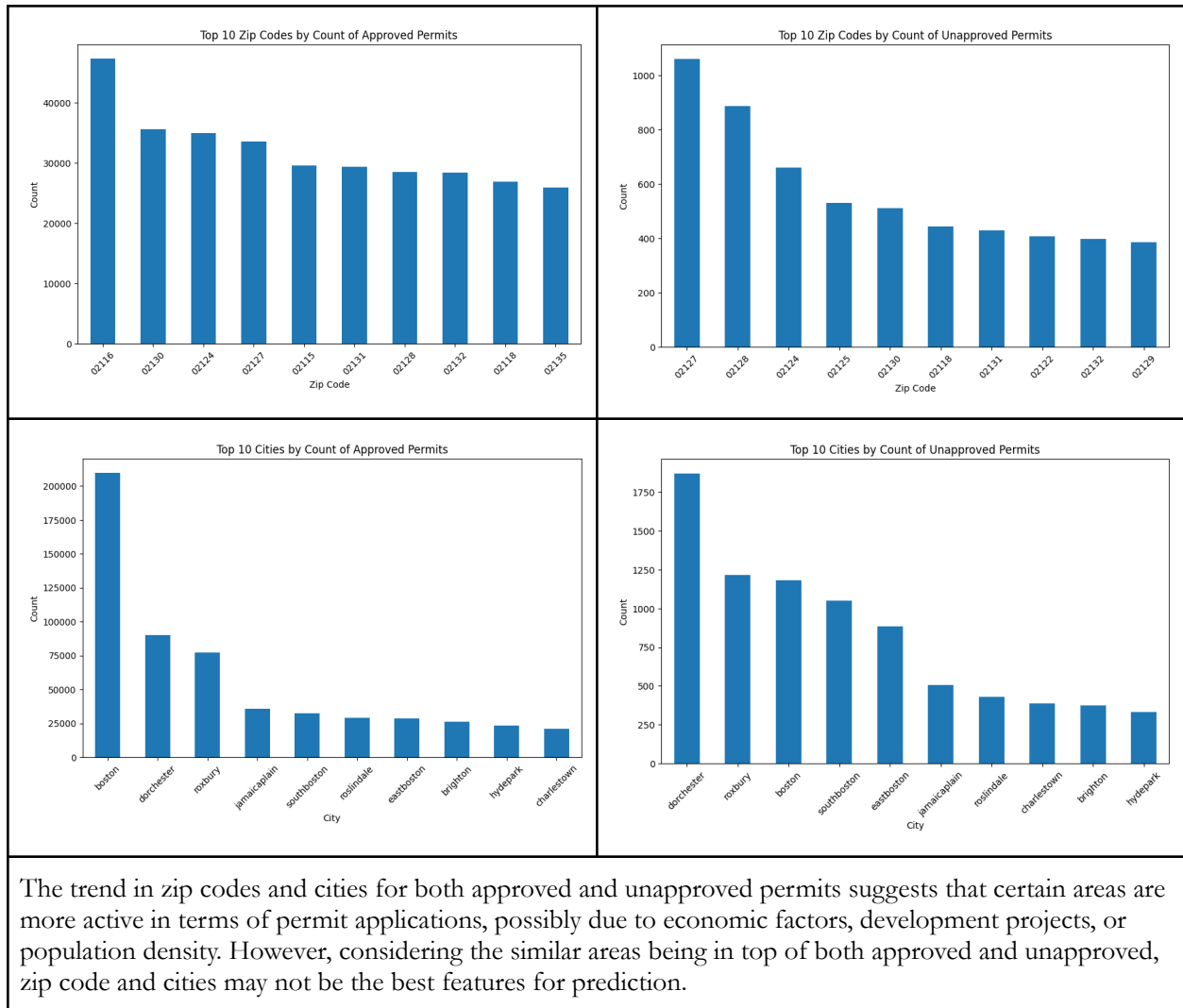
Assumptions
- Focus on construction or renovation work that's likely external or internal, hinting at a common approval for renovation projects.
- The presence of "Residential" suggests that approved projects may be associated with residential properties.
- Words like "Driveway" and "Parking" indicate that modifications to property access points are commonly approved.
- The term "Temporary" could imply that temporary structures or changes are often approved.

Dominant Words: "Change", "Family", "Building", "Occupancy"

Other Notable Words: "Extend", "Remove", "Single", "New", "Duplex", "Residential"

Assumptions
- The word "Change" is prominent, suggesting that applications involving significant changes may be often rejected.
- "Family" alongside "Single" and "Duplex" could indicate issues related to zoning or intended use, possibly with attempts to change a single-family dwelling into a multi-family one facing rejections.
- "Occupancy" being large may point to applications being rejected due to occupancy-related issues, which could be related to regulatory standards.
- "Building" and "New" suggest that new construction or significant building modifications may be often subject to rejection.

The trend in zip codes and cities for both approved and unapproved permits suggests that certain areas are more active in terms of permit applications, possibly due to economic factors, development projects, or population density. However, considering the similar areas being in top of both approved and unapproved, zip code and cities may not be the best features for prediction.

6. Individual Contributions

Name	Contributions
David Euijoon Kim (Team Lead)	Drafted the deliverable 2 file, conducted preliminary analysis and exploratory data analysis for key questions, then wrote the introduction part, exploratory data analysis on extension project; brainstorming the extension proposal and asked specific questions. NLP analysis as part of visualization Merged dataset of approved and zoning board (extension project).
Efim Sokolov	Found and imported datasets for census data, income and geographic boundaries, performed spatial and attribute joins to merge the data, plotted maps and other visualizations relating to demographic, income and development geographic distributions, plotted Lorenz curves for income and demographics, calculated Gini coefficients to analyze equality of distribution.
Zhihuan Hao	Base project partially conducted exploratory data analysis, answered 4 key questions with visualization & insights (1-4) and the corresponding presentation slides on checkpoint A and Deliverable 2; Cleaned the dataset of Covid-19 cases and did some exploratory data analysis around it. Made out the comparison for the pattern of the peak of pandemic and compared it with team members permits graph.
Lukas Werk	Merged data, worked on cleaning data and extracting/choosing/ removing columns and features for the various merges, analyses and extension project. Some interesting work here was looking at cleaning anomalies like NaNs or incorrect values and investigating why they were present. Some were entry issues and typos, that gave interesting or unexpected behavior when improperly cleaned, for example the coordinate fields for the merged geographic analyses we did. I also took the chance to look at some of the individual entries (especially in a80) to really understand what larger projects were present and what was the most relevant and impactful to our analysis.
Jackson Fisk	Converted article 80 temporal data into usable metrics for graphs to accurately represent Boston's current article 80 permitting state. Helped with the consensus on fields to data clean for each dataset. Transformed other categorical data into numeric for potential machine learning applications. Created graphs and designed presentation slides regarding article 80 temporal data and insights in the permitting process. Began exploratory data analysis for the extension project using two Covid-19 datasets and presented findings regarding race data and . Proofread and edited final deliverable.