Boston Permitting Process Analysis

E. Kim, L. Werk, E. Sokolov, Z. Hao, J. Fisk

Boston University

CS506: Data Science Tools and Applications

TEAM D

Euijoon (David) Kim	dk98@bu.edu	CDS '24	*Team Rep.*
Lukas Werk	lmwerk@bu.edu	CAS '24	
Efim Sokolov	efim@bu.edu	CAS '25	
Zhihuan (Richard) Hao	haozh@bu.edu	CAS '24	
Jackson Fisk	jfisk@bu.edu	CAS '24	

Code and Data for this Project can be found on our GitHub Repository:

https://github.com/BU-Spark/ds-boston-permitting

Introduction

The real estate and development sector significantly drives economic growth in many U.S. cities, particularly in Boston. The city witnesses a continuous flow of construction activities, ranging from new buildings to extensive renovations. The initiation of any development project in Boston begins with securing an official permit, a comprehensive process that involves multiple stages of approval, active community involvement, and public hearings. This permitting analysis project aims to thoroughly examine and analyze this complex process.

The permitting process in Boston commences with the submission of an application accompanied by the required fees. The next stage in the process is the review of the project plans against the zoning laws of the city. Projects failing to comply with these regulations are at risk of permit denial. Only upon the issuance of an official permit card, which holds validity for a period of six months with the possibility of extension, can a development project officially kick off.

In instances where applications are rejected, applicants are provided with the option to file an appeal. This appeal process is complex, encompassing the filing of the appeal itself, a period of community engagement, a public hearing, and then the wait for a final decision from the Zoning Board of Appeal (ZBA). Should the appeal be denied, the opportunity to file a new appeal arises after a year.

Furthermore, this analysis project also looks into larger-scale projects, specifically those exceeding 20,000 square feet, which are subject to the Article 80 review process. This represents a phase for large development projects, requiring an extensive and thorough scrutiny by various city departments and stakeholders.

The scope of this analytical project is wide-ranging, extending beyond the technicalities of the permitting process to encompass the societal, political, and environmental considerations that are crucial for the sustainable and integrated growth of Boston.

This extensive analysis is further enriched by two additional extension projects. The first of these employs a decision tree algorithm to analyze building permits and zoning board data, unveiling key insights such as the determining factors in permit outcomes, the prevalence of specific word usage in permit descriptions, and the relationship between permit applications and a variety of variables including city, zip code, and project type. This analysis offers valuable predictions about permit outcomes.

The second extension project takes a deeper look into the impact of the COVID-19 pandemic on the permitting process. This part of the study brings to light significant shifts in the patterns of permit issuance and the types of projects that received approval during the pandemic, underscoring the effects of the health crisis on the urban development landscape of Boston.

Data Overview

The data used for this project consists of five datasets from the City of Boston, State of Massachusetts and US Census. The datasets and our cleaning methods are described below.

▤ *Approved Building Permits*

The Approved Building Permits dataset includes approved permits for a large variety of constructions. The cleaning process for this dataset included basic text preprocessing and NaN removal, and the extraction of features such as dates into year, month and day columns.

▤ *Article 80 Permits*

The Article 80 Permits dataset contains a subset of permits subject to article 80, or larger projects. This dataset received simplification and renaming of columns, date-derived columns, and retention of specific columns for future use, with null values maintained.

▤ *Zoning Board of Appeal*

The Zoning Board of Appeal Tracker contains information on rejected permits submitted to the Zoning Board of Appeal. This dataset 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.

▤ *Census Data*

We pulled the census data from the Analyze Boston website and processed it by renaming columns, joining demographic data with shapefiles, removing unrelated columns, and normalizing data to get demographic proportions for census block groups.

▤ *COVID19 Data*

In addition to our base datasets, we pulled COVID-19 data from the CDC's COVID Data Tracker. The cleaning of this dataset requires dropping na values, selecting only the confirmed death, confirmed cases and the dates columns. Then the whole day-by-day tracking will be converted to a month-to-month format where the date column is in year-month and with the corresponding cases and deaths being calculated for that month. Finally, the cleaned dataset is merged with all the permitting datasets with respect to the project dates in year month form.

No	Original	Data Cleaning Explained	Cleaned Data
1	abp.csv 27 columns	1. Drop unnecessary columns A. applicant, owner, address, state, property_id, parcel_id, some geo columns 2. Numeric Columns A. Process monetary columns to floats B. Strip permit number prefixes C. Format zipcodes D. Extract (year, month, day) from dates 3. Text Columns A. Levenshtein match cities to correct typos B. Process text with standard NLP methods 4. Null values A. Keep the null values, as some are not mistakes.	cleaned_abp.csv 20 columns
2	a80.csv 25 columns	1. Drop unnecessary columns A. X, Y, Project_Street_Name, Project_Street_Suffix B. contact - Personal information not required	Cleaned_a80.csv 33 columns
		2. Change column name A. Lower case letters and simpler names	
		3. Derive columns from dates (year, month, day) A. Filed_Date, BPDA_Board_Approval, First_Building_Permit, COO_Permit_Date, Last_Project_Update_Date	
		4. Keep for future use A. objectid, projectid, lat, lon B. name, description - wordcloud	
		5. Null values A. Keep the null values, as some are not mistakes.	

3	zba.csv 18 columns	1. Drop unnecessary columns C. address - Use city, zip, and zoning_district variables instead D. contact - Personal information not required	cleaned_zba.csv 35 columns
		2. Convert categorical variable values to numerical form - Remains categorical A. status - 1 to 7 in order of the appeal process - for simplicity B. appeal_type - Zoning : 0, Building: 1	
		3. Derive columns from dates (year, month, day) B. submitted_date, hearing_date, final_decision_date C. Calculate the differences among the three variables (duration)	
		4. Correct human error A. decision 'AppProv' == 'Approved'; 'DeniedPrej' == 'Denied'; ' ' == 'nan';	
		5. Clean up values A. zoning_district Eliminate the term 'Neighborhood' from the values	
		6. Keep for future use C. parent_apno, boa_apno, ever_deferred, num_deferrals, city, zip, ward D. project_description - wordcloud	
		7. Null values B. Keep the null values, as some are not mistakes.	
4	census.csv	1. Imported Census block group data and shapefiles from City of Boston datasets	cleaned_census.csv
		2. Renamed columns from census identifiers to readable descriptions	
		3. Performed an attribute join between demographic data and shapefile on	

		GeoID	
		4. Dropped columns unrelated to demographics (mostly on federal institutions such as prisons, juvenile facilities and military).	
		5. Dropped columns with unnecessary geographic identifiers (same values for all rows, repeated identifiers due to join)	
		6. Plotted shapefile geometries for block groups	
		7. Plotted a heatmap of demographic data for each census group, mapping colors to demographic distributions across block groups	
		8. Normalized data by total to retrieve proportions of each demographic group for a given census block group and plotted heatmap.	
5	covid.csv	1. Extracted Confirmed Cases, Confirmed Deaths, and Dates	cleaned_covid.csv
		2. Transformed Year-Month-Day into Year-Month	
		3. Dropped Null values	
		4. Merged zbp and zba with respect to the Year-Month column	

Base Project

To gain general insights on the Boston permitting process, we prepared a Base Project report to answer key questions related to the permitting process over the last five years.

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.

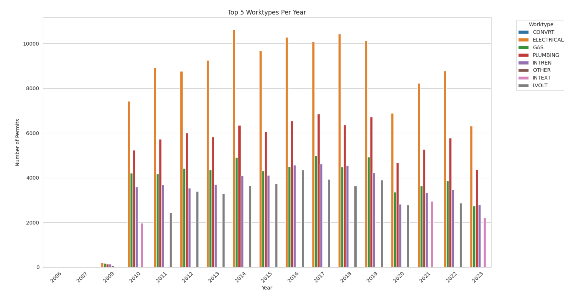


Figure 1. Top 5 Worktypes Per Year

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.

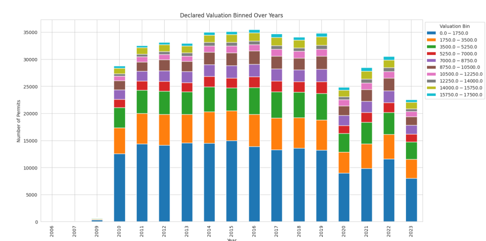


Figure 2. Top 5 Valuations Per Year

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.

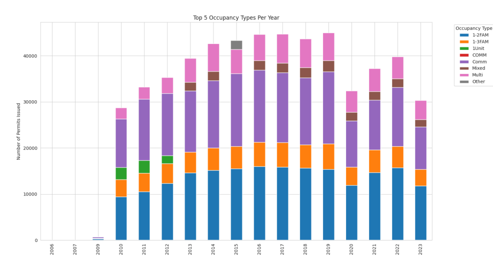


Figure 3. Top 5 Occupancy Types Per Year

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.

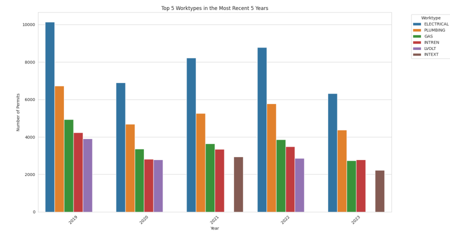


Figure 4. Top 5 Work Types Per Year (5 Years)

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.

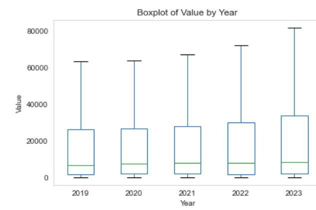


Figure 5. Top 5 Valuations Per Year (5 Years)

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. The Covid-19 pandemic is a likely cause of this fluctuation in recent years.

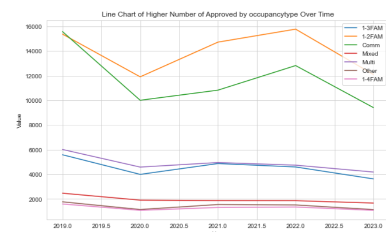


Figure 6. Top Occupancy types (5 Years)

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. Based on the geography, we assumed 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.

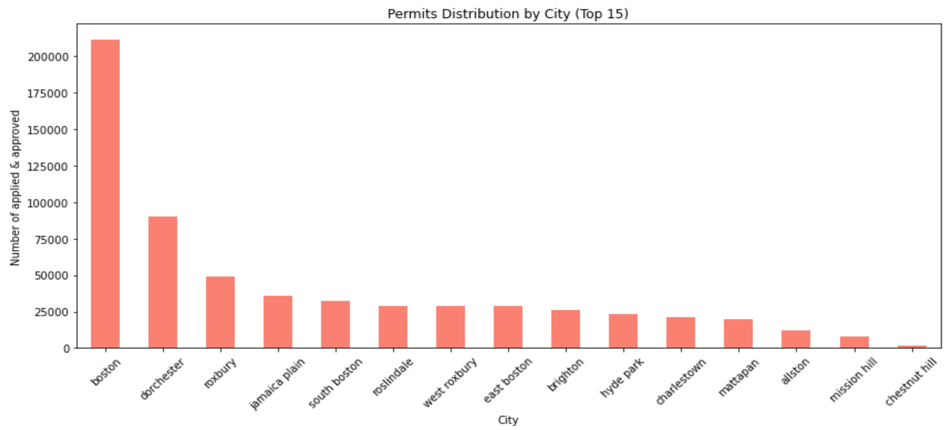


Figure 7. Permits Distribution by City

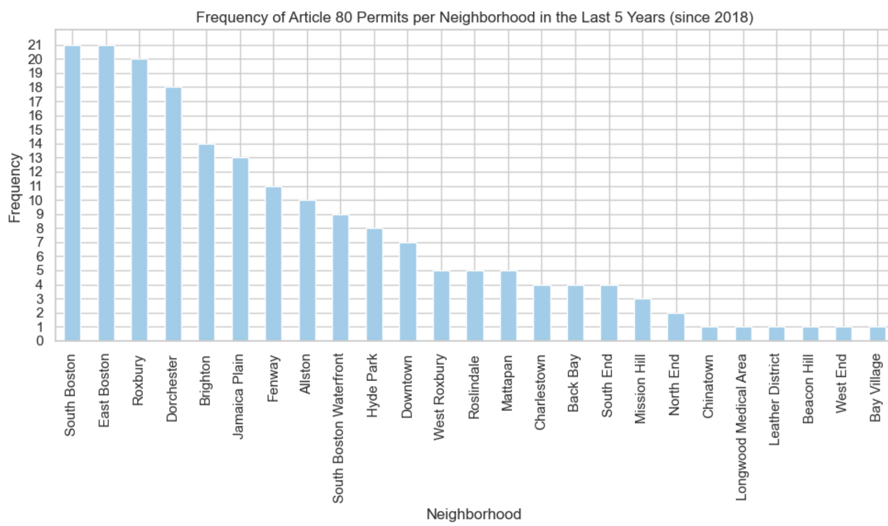


Figure 8. Frequency of Article 80 permits per Neighborhood in last 5 years

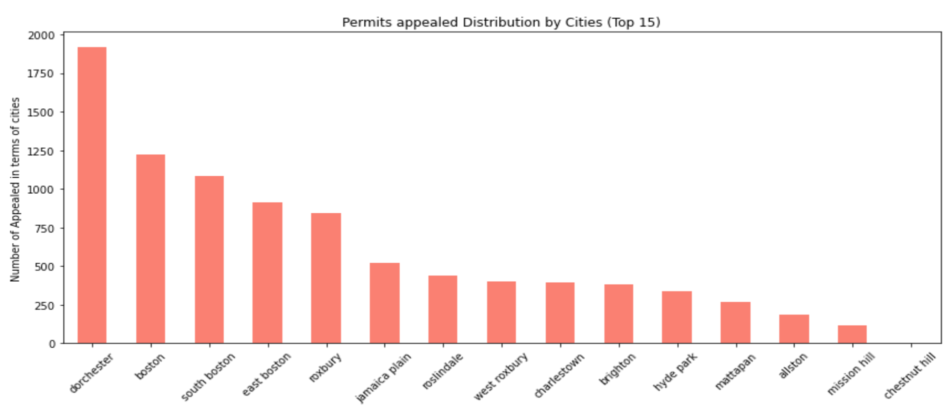
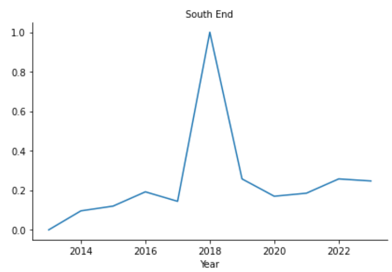


Figure 9. Permits Appealed Distribution by Top 15 Cities

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.



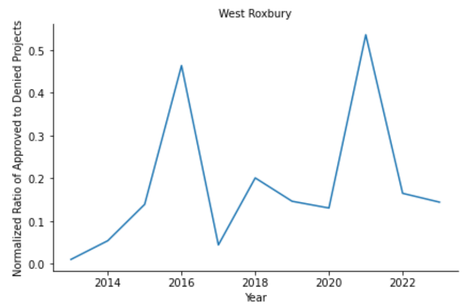
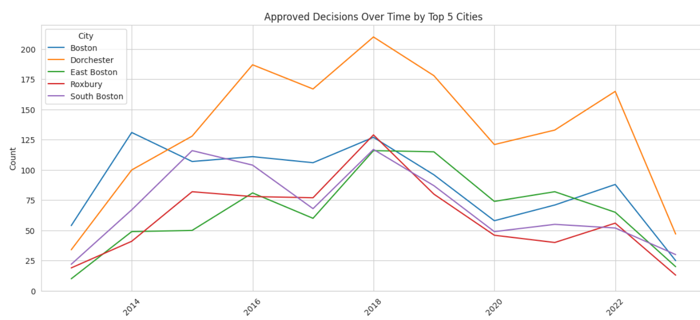


Figure 10.a (South End) and Figure 10.b (West Roxbury).

The normalized ratio of number of Approved over Denied projects in South End and West Roxbury



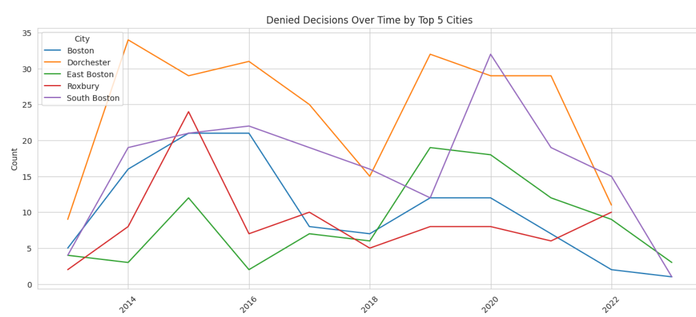


Figure 11.a (Above) and Figure 11.b (Below), Number of Approved & Denied Projects by Top 5 Cities

5. 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 Figure 12. 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.

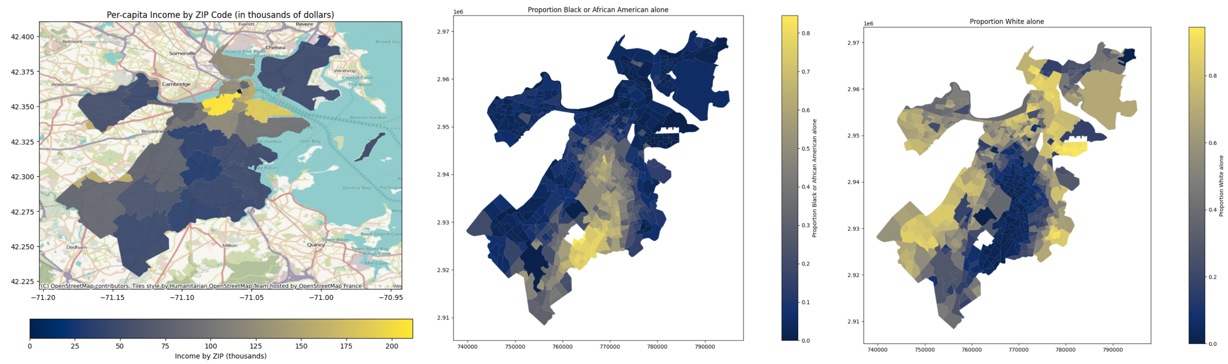


Figure 12 a b c. Income and Demographic Distributions

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.

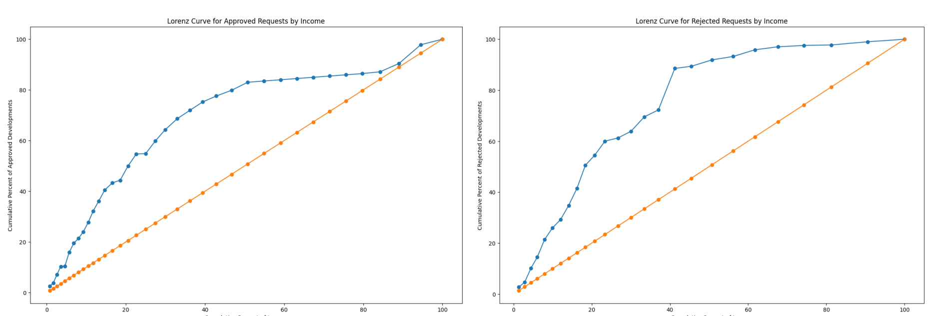


Figure 13 a b. Lorenz Curve for Approved Requests by Income

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.

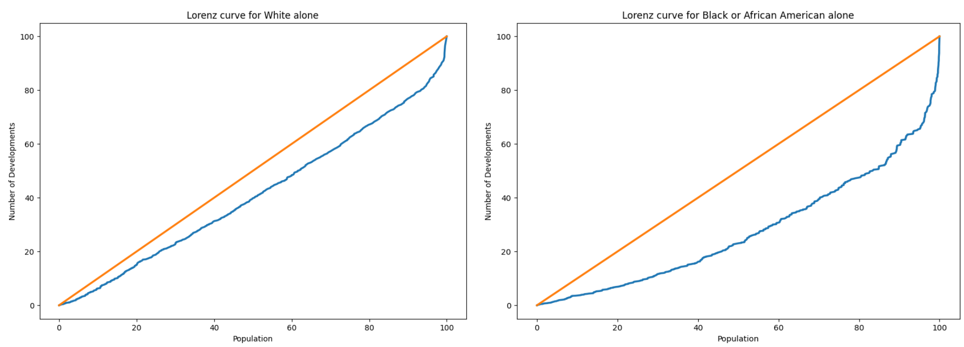


Figure 14 a b. Lorenz Curve for Different Races

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.

Extension Projects

Extension Project I - Decision Tree Analysis for Permit Outcomes

This extension project combines the Approved Building Permits (abp.csv) and the Zoning Board of Appeal Tracker (zba.csv) datasets. A decision tree algorithm was utilized to identify and illustrate the key factors leading to permit outcomes and to uncover novel features. Our decision tree was trained on a vectorized word matrix containing every word used in the text fields of the dataset, which allowed us to analyze trends in word usage in descriptions.

Before proceeding with the decision tree analysis, 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.

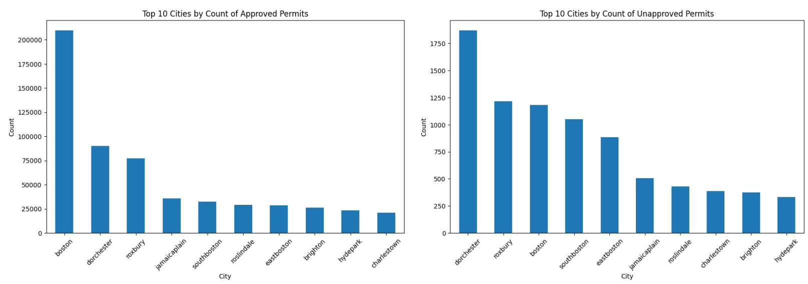


Figure 15 a b. Top 10 Cities by Count of Permits

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. This is possibly due to economic factors, development projects, or population density. However, considering similar areas topped both approved and unapproved as shown in Figure 15, we came to the conclusion that zip code and cities shouldn't be used for prediction.

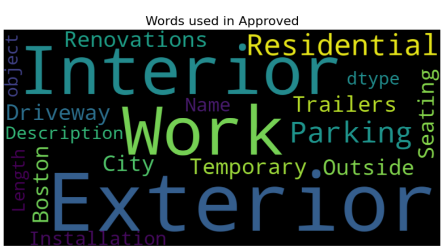


Figure 16a. Words used in Approved Permits

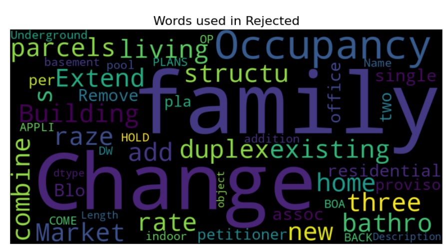


Figure 16b. Words used in Rejected Permits

Dominant Words: "Exterior", "Interior", "Work", "Residential", "Renovations"

Other Notable Words: "Installation", "Driveway",

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

Other Notable Words: "Extend", "Remove", "Single", "Parking", "Temporary", "City", "Boston"

Conclusions:

- Focus on construction or renovation work that's likely external or internal hints 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.

"New", "Duplex", "Residential"

Conclusions:

- 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 Figure 17 shows a decision tree plot and a performance metric table for a classification model. The decision tree diagram illustrates the rules derived from the data to predict whether a permit will be approved or not, with nodes representing the conditions based on features like 'parking', 'family', 'change', 'issued', 'service', and 'new'. The leaves represent the final decision of 'Approved' or 'Not Approved'.

The performance metrics indicate an accuracy of 0.79. Precision, recall, and F1-score are provided for both classes (False for not approved, True for approved). The model shows better recall for the approved class (0.95) than the not approved (0.61), indicating it's better at identifying true approvals than true rejections.

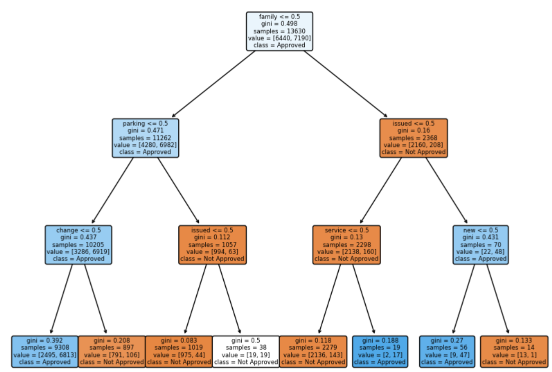


Figure 17a. Decision Tree Plot

Accuracy: 0.79

	precision	recall	f1-score	support
False	0.92	0.61	0.73	1598
True	0.73	0.95	0.83	1810
accuracy			0.79	3408
macro avg	0.83	0.78	0.78	3408
weighted avg	0.82	0.79	0.78	3408

Figure 17b. Decision Tree Statistics

Overall, the words 'parking', 'family', 'change', 'issued', 'service', and 'new' are important in identifying approved and rejected permits in Boston.

Extension Project II - Correlating COVID-19 Impact with Permitting Data

This extension project establishes 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. Such an analysis is crucial in assessing the resilience of the city permitting process and the city’s responsiveness to unprecedented health crises.

1. What is the general trend of the Covid cases, deaths and how may or may not have an effect on the permitting approvals?

As Figure 18 shows, there is a huge burst of Covid deaths in 2020.05, 2021.02, and 2022.01. Whereas in Figure 19, for the confirmed cases, there are several small and large peaks in 2020.04, 2021.01, 2022.01 and 2022.05. By merging the columns of Covid dataset and Approved permitting projects, normalized columns are also made so that it can be shown in Figure 20 for a better comparison. It is easy to see that on 2020.04, the peak of covid deaths corresponded to an extreme minimum point of number of projects that were approved on that month. Same trend goes to 2021.01, the increase and burst of both confirmed cases and deaths are followed with significant decreases in the number of projects; And so as the 2022.01. Based on the visualization, there is a high probability that the Covid has an effect on the number of approved permitting projects.

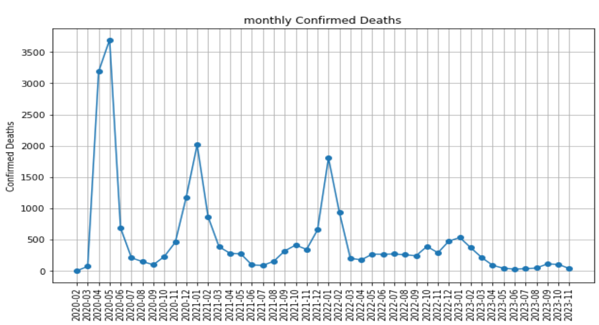


Figure 18. Monthly Confirmed Covid Deaths

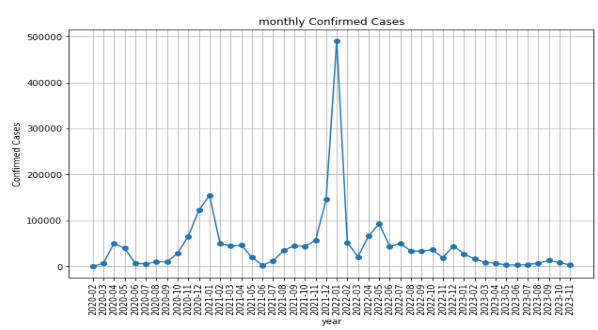


Figure 19. Monthly Confirmed Covid Cases

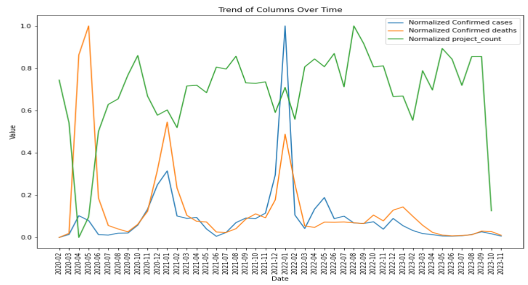


Figure 20. The Combination of Normalized Approved Projects, Confirmed Cases and Confirmed Deaths Each Month

2. How strong is the relationship between confirmed cases, deaths and approved projects?

Are both the cases and deaths strongly correlating with the trend of approved permitting projects?

Based on the correlation matrix, there is a positive correlation between confirmed cases and confirmed deaths, which makes sense. And there is only a very little negative correlation between confirmed cases and number of approved projects; But, there is a strong negative correlation between confirmed deaths and number of approved projects.

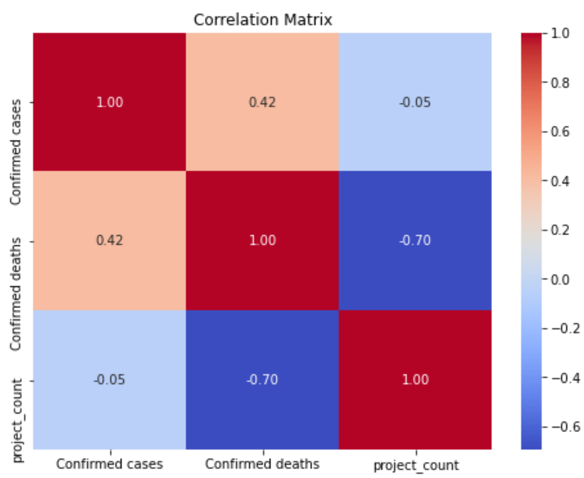


Figure 21. Correlation Matrix Between Confirmed Cases, Deaths and Number of Approved Projects

Based on the Q-Q plots between the distributions of Confirmed Cases and Approved/Denied Projects by the Zoning Board of Appeals, it's clear that the distributions are similar with stronger similarity between denied counts and confirmed cases since the plot trends closer to the identity line. This indicates a stronger relationship between COVID cases and denied permits. For both plots, most project counts are clustered in the smaller quantiles of the confirmed cases.

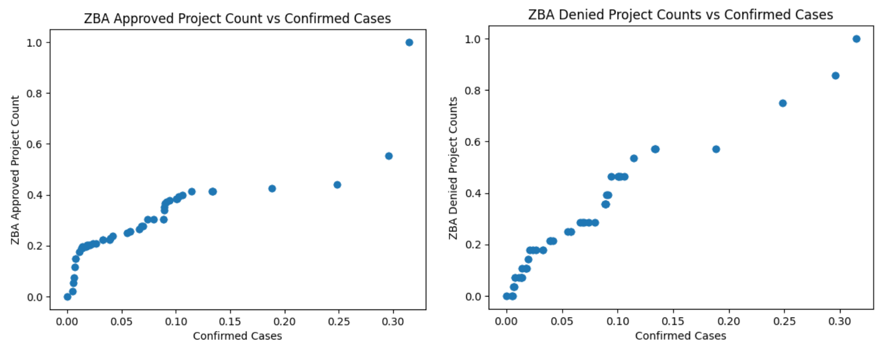


Figure 22.Q-Q plots for Confirmed Cases, Deaths and Number of Approved Projects

3. What kind of relationship could it be between the confirmed deaths and number of approved projects?

Based on the regression, the R-squared means that approximately 48.4% of the data can be explained by the linear regression. And since the p-value of Confirmed Deaths equals 0, we reject the null hypothesis that the coefficient of the variable Confirmed Deaths is 0. And -0.6902 means that approximately every one person is confirmed dead by the Covid, the number of approved projects decreases by 0.69.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:          project_count   R-squared:                       0.484
Model:                            OLS   Adj. R-squared:                  0.472
Method:                 Least Squares   F-statistic:                     40.27
Date:                Mon, 11 Dec 2023   Prob (F-statistic):           1.15e-07
Time:                        19:47:21   Log-Likelihood:                -346.94
No. Observations:                  45   AIC:                             697.9
Df Residuals:                      43   BIC:                             701.5
Df Model:                           1
Covariance Type:            nonrobust
====================================================================================
                       coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------------
const             3861.1528     99.282     38.891      0.000    3660.933    4061.373
Confirmed deaths    -0.6902      0.109     -6.346      0.000      -0.910      -0.471
==============================================================================
Omnibus:                       33.031   Durbin-Watson:                   1.458
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              108.580
Skew:                          -1.767   Prob(JB):                     2.64e-24
Kurtosis:                       9.739   Cond. No.                     1.10e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 1.1e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

Figure 23. Simple Linear Regression of Number of Projects Approved and Number of Confirmed Deaths

Conclusion

A comprehensive examination of the Boston real estate and development sector's permitting process was conducted in this report, detailing its impact on economic growth and the influence of socio-political factors on the process. The analysis utilized datasets from the city of Boston, the State of Massachusetts, and the US Census, which were meticulously cleaned and preprocessed. These datasets included information on approved building permits, Article 80 permits, the Zoning Board of Appeal Tracker, and demographic data.

Key insights from the base project revealed that the most common work types for approved permits were electrical, plumbing, and gas works, followed by conversions and interior renovations. This reflected the ongoing needs for maintenance and compliance with safety regulations. The reported valuations for approved permits generally remained low, with the most common valuations being under $5,250. The predominant occupancy types for approved permits were residential and commercial buildings. A year-over-year analysis showed a declining trend in the number of approved permits, with a slight increase in the average declared valuation and a peak in permit approvals around 2019, which was then followed by a decline in 2020 due to the pandemic. Geographical analysis indicated that most approved permits came from urban areas like Boston and Dorchester, with South and East Boston experiencing a rise in permits under Article 80.

The Lorenz curves indicated a skewed distribution of developments, particularly in lower-income ZIP codes, where a disproportionate number of permit approvals and rejections occurred. Notably, 80% of approved and 92% of rejected permits were concentrated in the lower 50% of income levels, suggesting an economic disparity where lower-income areas saw more development, possibly due to urban renewal, gentrification, or lower costs. Demographically, development distribution was also unequal. The Gini coefficient for the white population (0.181) indicated a relatively equitable development distribution, whereas for the Black or African American group, the coefficient was significantly higher (0.446), signaling a stark disparity.

The Decision Tree Analysis for Permit Outcomes was a part of the study that used a decision tree algorithm to predict permit outcomes based on features extracted from the data, including word usage trends in permit descriptions. The analysis suggested that words like 'parking', 'family', 'change', 'issued', 'service', and 'new' were critical in identifying approved and rejected permits. The decision tree model achieved an accuracy of 0.79, with better performance in identifying approved rather than rejected permits.

The section correlating the COVID-19 Impact with Permitting Data established a correlation between COVID-19 data and permitting activity, showing that peaks in COVID-19 deaths were correlated with decreases in approved permits. This suggested that the pandemic had a tangible effect on the permitting process. The report uncovered a strong negative correlation between COVID-19 deaths and the number of approved projects, reinforcing the need to consider the effects of systemic natural disasters or health crises on urban processes like permitting.

In summary, the report provided valuable insights into the complexities of the permitting process in Boston, also highlighting the impact of external factors such as the COVID-19 pandemic. It suggested that economic and social variables had to be considered when forecasting development patterns and that there was a need to understand and mitigate the disparities in development distribution among different demographic and income groups.

Challenges & Limitations

Initially, grasping the project goals and delving into the complexities of the Boston Permitting Process demanded extensive reading and data exploration. This early phase involved a deep dive into the background information and data, leading to a foundational understanding. Subsequently, the task of cleaning, merging, and analyzing the data presented its own set of challenges, such as incomplete information and ambiguous data.

In the Decision Tree extension project, one of the hurdles was merging the abp and zba datasets due to a lack of compatible columns. Consequently, the merged dataset primarily featured dates and descriptions for prediction analysis. However, the dates showed no significant correlation with the permitting outcomes, necessitating the use of descriptions and natural language processing techniques for prediction. The process included cleaning the descriptions by removing stopwords and tokenizing the text. Yet, aiming for interpretability with a simpler decision tree compromised the accuracy of the predictions.

For the Covid extension project, while the data visualization and linear regression yielded noteworthy insights, the study could benefit from a larger sample size spanning more than just several months. Alternatively, comparing the current data with similar analyses from another historical disease outbreak could strengthen the argument about the impact of diseases on permitting project approvals. A broader analysis encompassing multiple disease periods could make the results more compelling. Another limitation is the linear regression's focus on a single variable – the number of confirmed deaths. Although significant, incorporating additional factors could help address issues of multicollinearity and omitted variable bias, making the analysis more robust.

Contributions

Contributions by team members are listed below. The team hosted weekly meetings and operated on a scrum collaboration framework. Additional time for collaboration was organized on an as-needed basis. The team representative for this project was David Euijoon Kim.

David Euijoon Kim

- Worked on data cleaning, eda, in-depth analysis, and visualization for the base project with the team
- Focus on organizing overall deliverables, data cleaning & analysis, and extension project 1.
- As team rep: organizing, reporting in-person, scrum reports, submissions, and video edits.

Lukas Werk

- Data Cleaning, Processing and Merging
- Feature extraction and text field processing
- Visualizations on base questions and extension project 1
- Initial processing and Decision Trees for EP1
- Reporting and Presentation format and creation

Efim Sokolov

- Initial Data Cleaning
- Locating, cleaning and merging (spacial, attribute) of geospatial and income datasets and shapefiles
- Geographical and inequality analysis and visualizations (maps, Lorenz curves, Gini)
- Q-Q plots and correlation analysis for COVID extension

Zhihuan Hao

- Data Cleaning and EDA for the base project and answering some of the key questions.
- Data cleaning, EDA, extension question creation and analysis for the Covid-19 extension project.
- Build line charts, correlation matrix and linear regression—referring to the extension-2 ipynb

Jackson Fisk

- Preprocessing / Data Cleaning on Article 80 dataset, created graphs, did basic and complex analysis
- For extension cleaned Covid-19 dataset, did feature analysis and created graphs showing different covid death trends in Boston. Reviewed and provided suggestions to further improve Decision Trees.
- Creation and editing of reports on findings