Midterm Report on the Impact of Remodeling and Zoning Conversions on Boston's Housing Market

Yuchen Cao, Jialong Ke, Jingbo Wang, Zihan Li

Introduction

This report extends our initial analysis of the Live Street Address Management (SAM) Addresses database to include a detailed examination of housing unit changes from 2009 to 2024 and an in-depth analysis of the "APPROVED BUILDING PERMITS" and "Property Assessment - Dataset" database. Our goal is to understand the trends in housing development and the impact of remodeling on housing units.

Previous Research

1. What communities are building more housing units?

In Central Boston, Dorchester, South Boston, and East Boston are the neighborhoods where the most new housing units are being built. Of these, Dorchester leads with 1,267 building permits. It shows that these neighborhoods are key areas for meeting the needs of urbanization and population growth.

2. Which ones are losing housing units?

Dorchester, Central Boston, and South Boston are the communities that have lost the most housing units. Most reduced the supply of homes by demolition or similar means.

3. Where are housing remodels and renovations happening?

Central Boston was the community that underwent the most frequent home remodeling and renovations, at 29,714, indicating a strong demand for upgrades or renovations to existing properties. Dorchester and Roxbury, which follow closely behind, also support this trend.

Data Preliminary Analysis

- Analyzed the changes in each type of unit undering 2009 - 2024
- Analyzed the trend of permits in different regions for each year.
- After our group's analysis, we found that the Live Street Address Management (SAM) Addresses table has no corresponding key reference and a large number of units are missing, only 1994 units can be effectively analyzed. units is a key that has many data types that need to be cleaned up before we can use it. The key units has a lot of data types that need to be cleaned up before it can be used for statistical purposes.

Midterm Research

Answering the last question: How many housing units are lost to remodels on average, each year?

In the preliminary analysis presented in our report, we conducted a trend examination through linear regression and assessed the variations in housing unit numbers from 2009 to 2024. This analysis provided insights into the locations experiencing growth in housing units and those witnessing declines. Utilizing this data, we addressed a crucial query: What is the average annual loss of housing units due to remodeling? To calculate this, we analyzed the changes in unit numbers over the specified period and divided by the number of years between 2009 and 2024, yielding the annual average. The initial presentation of this data in Figure 1, though informative, lacked clarity in its visual representation.

Zip Code		Zip Code	
02108	6.800000	02128	135.133333
02109	6.200000	02129	31.066667
02110	48.733333	02130	-15.333333
02111	-1.466667	02131	39.333333
02113	7.266667	02132	-12.000000
02114	-6.133333	02134	24.000000
02115	27.400000	02135	137.200000
02116	-637.733333	02136	24.133333
02118	-89.533333	02137	NaN
02119	-9.400000	02146	NaN
02120	0.866667	02199	0.000000
02121	7.066667	02201	NaN
02122	23.533333	02210	64.466667
02124	5.266667	02215	39.533333
02125	34.066667	02445	NaN
02126	-1.000000	02446	0.000000
02127	64.066667	02467	-13.266667

Figure 1. Average unit change in zip code

To enhance comprehension and visual appeal, we restructured the data visualization. By assigning zip codes to the x-axis and the changes in unit numbers to the y-axis, we created a more intuitive graph Figure 2. This graphical representation allows for immediate identification of regions with increasing or decreasing housing unit trends.
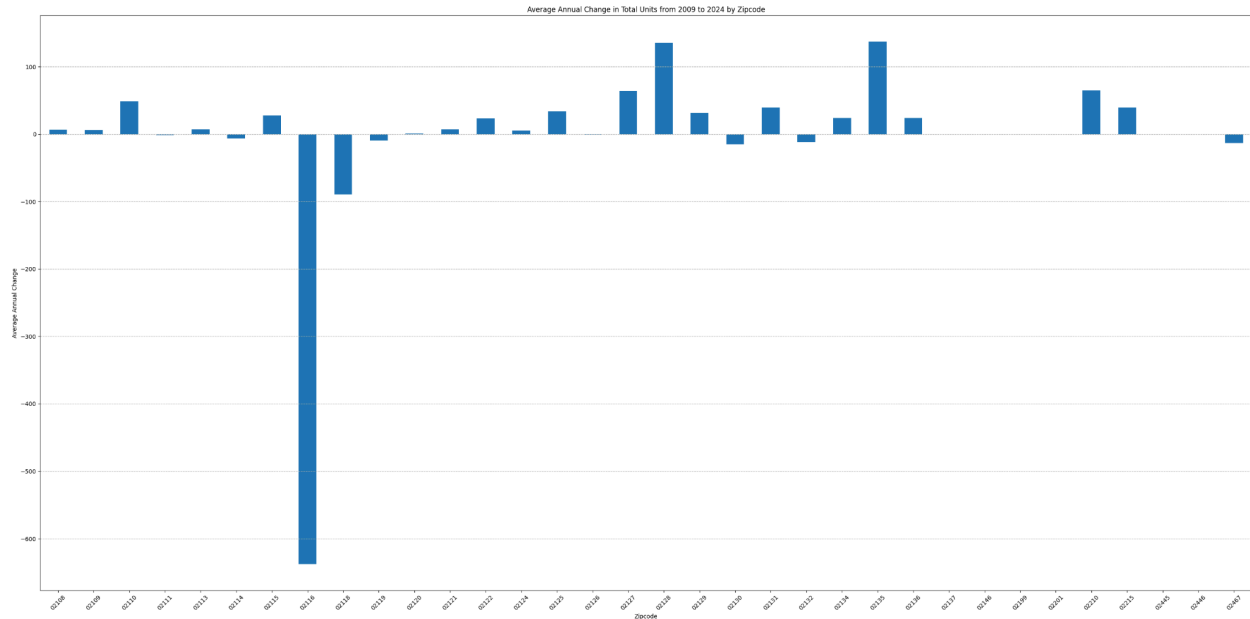
Figure 2. Average unit change in zip code in graph

These findings offer valuable insights into the dynamics of housing availability and the impact of remodeling on housing units. Understanding these trends is crucial for policymakers, developers, and community planners in making informed decisions related to housing development and renovation policies.

Coming up with another question: How has the remodeling taken place over different years?

Besides the key questions, our research team posed an additional question: How has remodeling taken place over different years? To address this, we consulted the "APPROVED BUILDING PERMITS" database, meticulously examining all relevant fields. Our analysis revealed that the 'issued_date' column denotes the date that the permit was granted, while the 'status' column reflects the permit's current condition. These columns enabled us to track the frequency of remodeling activities over time.

Initially, we intended to use the 'city' column to identify project locations. However, upon analysis, we found the data within this column to be inconsistent and poorly formatted, complicating direct usage. As shown in Figure 3, there are a lot of different formats.
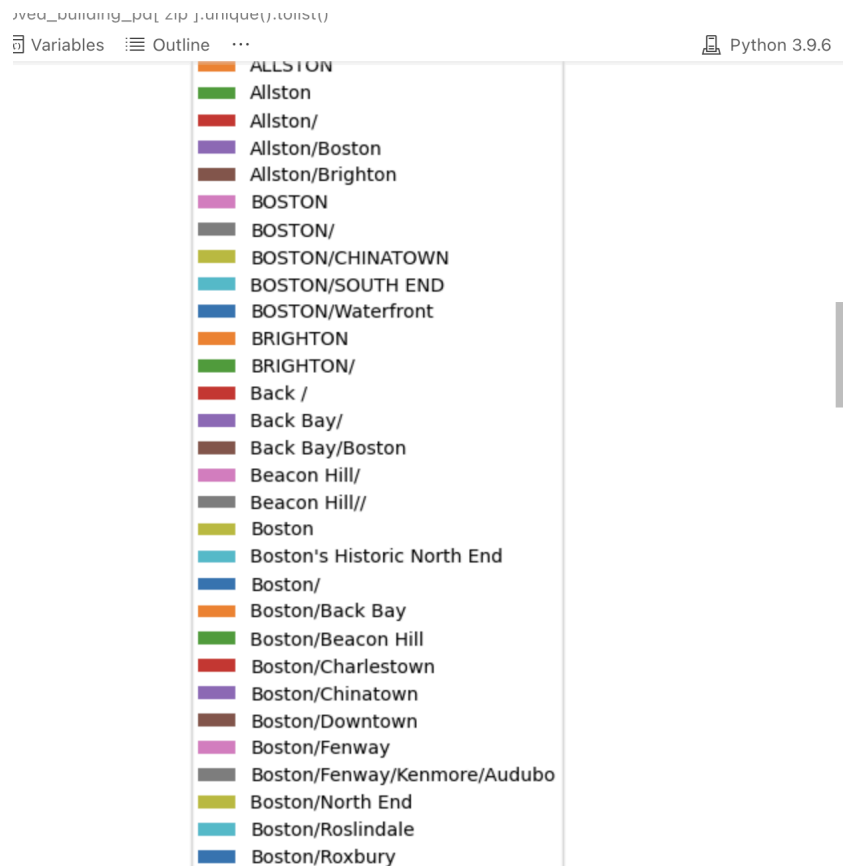
- ALLSTON
- Allston
- Allston/
- Allston/Boston
- Allston/Brighton
- BOSTON
- BOSTON/
- BOSTON/CHINATOWN
- BOSTON/SOUTH END
- BOSTON/Waterfront
- BRIGHTON
- BRIGHTON/
- Back /
- Back Bay/
- Back Bay/Boston
- Beacon Hill/
- Beacon Hill//
- Boston
- Boston's Historic North End
- Boston/
- Boston/Back Bay
- Boston/Beacon Hill
- Boston/Charlestown
- Boston/Chinatown
- Boston/Downtown
- Boston/Fenway
- Boston/Fenway/Kenmore/Audubo
- Boston/North End
- Boston/Roslindale
- Boston/Roxbury

Figure 3. Data in 'city' column of "APPROVED BUILDING PERMITS" database

Consequently, we opted to use the 'zip' code as a more reliable indicator of location for each permit.

We then proceeded to create a series of graphs displaying the volume of remodeling projects across different areas and years. As shown in Figure 4, the horizontal axis represents the year, the vertical axis represents the number of permits, and each line represents a different area (distinguished by zip code).
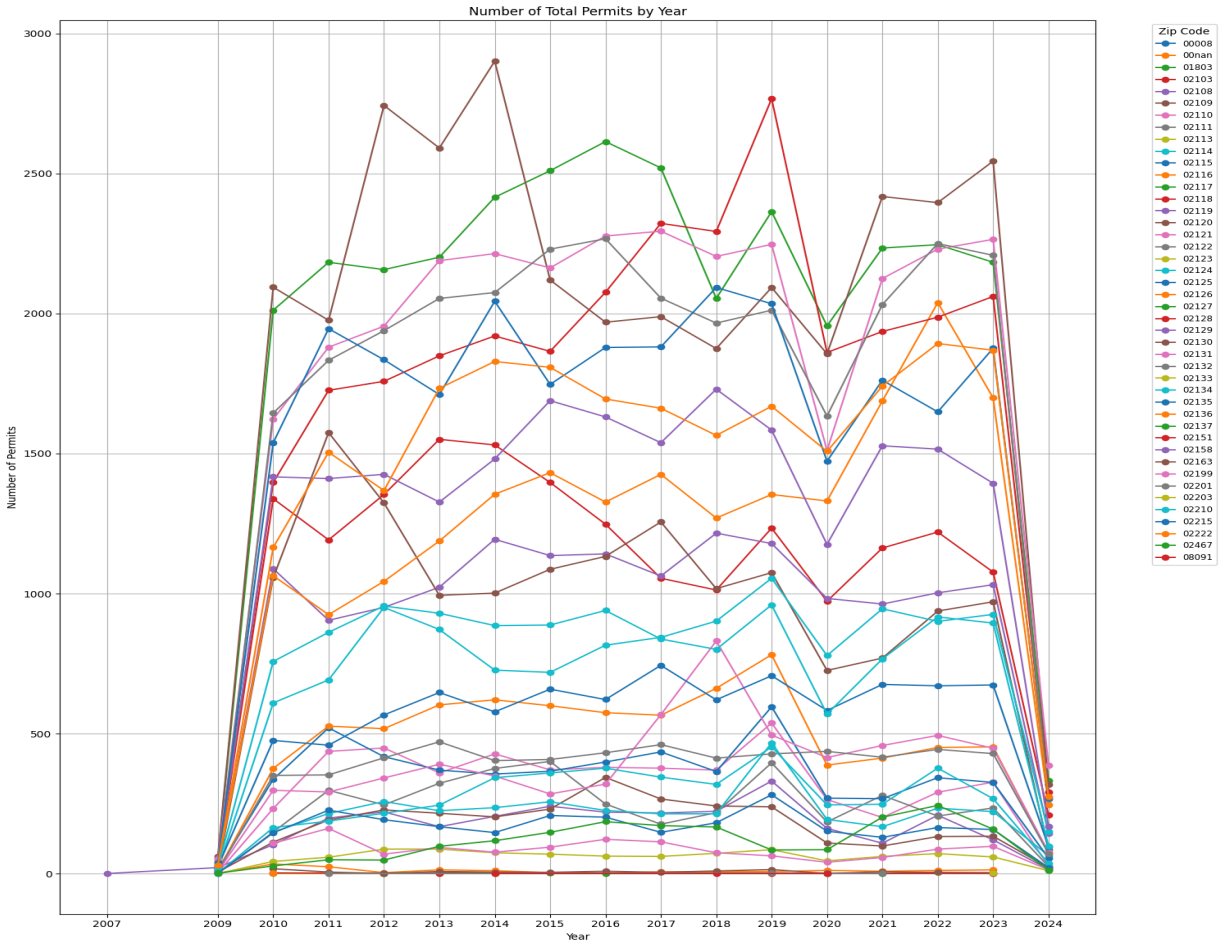
Figure 4. Volume of remodeling projects across different areas and years

Our findings indicated four distinct statuses for permits, including a 'Stop work' status, which signifies permits that were issued but later halted due to non-compliance or other issues. We excluded permits with this status from our analysis. The remaining statuses—'open' and 'closed'—indicate whether a permit is currently valid or not, respectively. We generated separate graphs for each status("Open" and "Close") to provide clearer insights, which are shown in Figure 5.
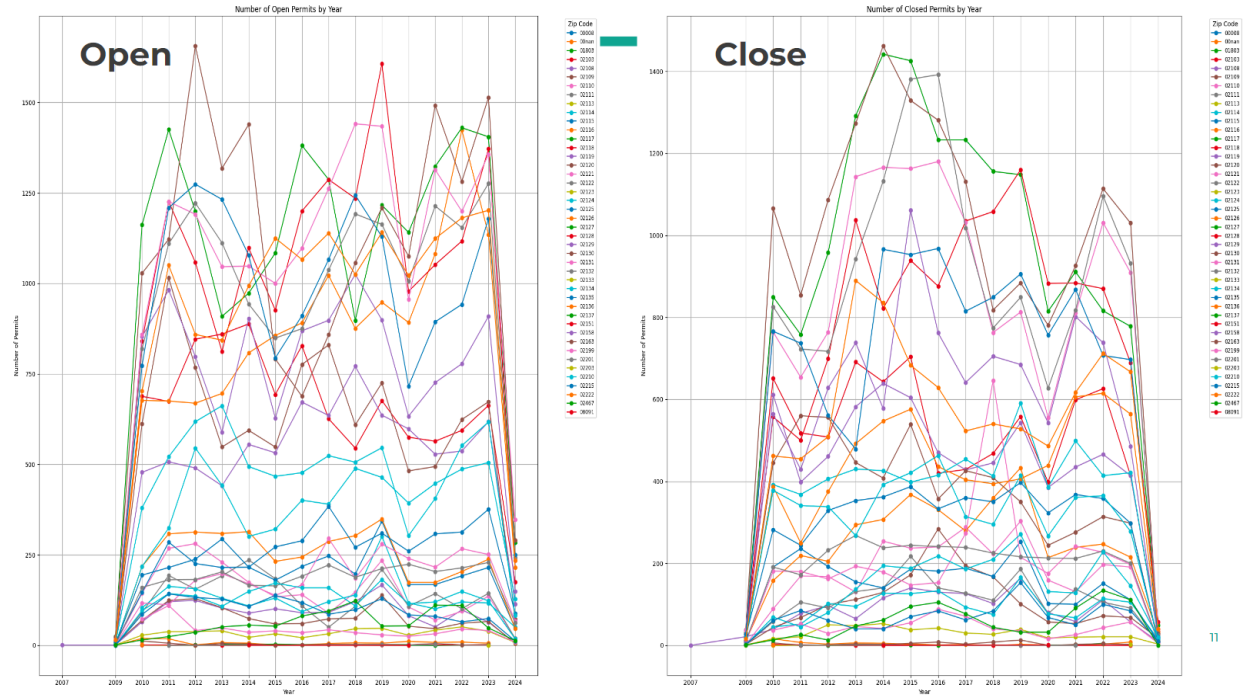
Figure 5. Volume of remodeling projects across different areas and years with status "Open" and "Close"

However, we observed that the initial graphs were overly complex and hindered easy analysis. To remedy this, we correlated zip codes with their corresponding neighborhoods, focusing on the top 10 to simplify the visual presentation. This approach significantly reduced the number of lines in each graph, enhancing clarity. As shown in Figure 6, the horizontal axis represents the year, the vertical axis represents the number of permits, and each line represents a different area (distinguished by neighborhoods) and only top 10 neighborhoods are selected.

Similar to above, we generated separate graphs for each status("Open" and "Close") to provide clearer insights, which are shown in Figure 7.

Figure 6. Volume of remodeling projects in top 10 neighborhoods across different years
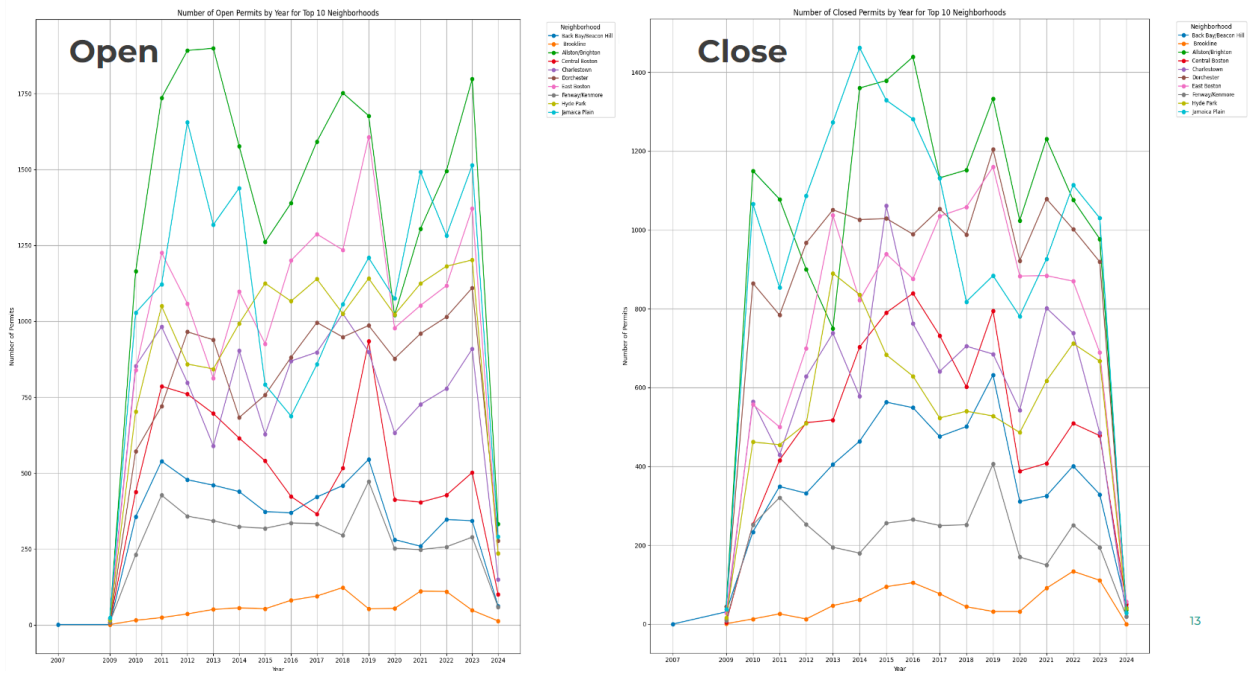

Figure 7. Volume of remodeling projects in top 10 areas across different areas and years

During this process, we identified several incorrect zip codes, such as 01803 and 00008. By cross-referencing the 'address' and 'city' columns, we accurately reassigned these entries to their correct locations, ensuring the integrity of our analysis.

This detailed exploration into remodeling activities sheds light on the construction trends across different neighborhoods, offering insights into the areas experiencing significant renovation activities. Such data is essential for understanding the changing landscape of housing and can guide future development strategies.

Conclusion

Through our comprehensive analyses, we have unearthed significant trends in housing unit changes and remodeling activities across various areas. These findings underscore the importance of robust data collection and analysis methods in understanding the impact of remodeling on housing trends. Furthermore, the identification of data inconsistencies and challenges emphasizes the need for continuous improvement in data management practices. This report has bridged the gap between initial insights and a more detailed understanding of housing unit changes and remodeling activities. Through rigorous data analysis and visualization, we have laid a foundation for ongoing and future research into these critical areas, providing stakeholders with actionable insights and a framework for informed decision-making.

Discussion

Our analyses have provided significant insights into housing unit trends and the scope and impact of remodeling activities. The methodologies adopted, from linear regression to strategic data visualization and geographical mapping, have allowed us to uncover patterns that inform on both the current state and potential future trends in housing and remodeling.

Questions from Stakeholders

1. **How accurate are the insights drawn from the available data, considering the missing units and disorganized city column information?**

The accuracy of insights is contingent upon the data's completeness and cleanliness. We've employed rigorous data cleaning techniques to mitigate these issues. However, stakeholders should note that while trends and patterns identified are indicative, the precision of predictions might be slightly affected by these data limitations.

2. **Can we quantify the impact of remodeling on the availability of housing units?**

Yes, by analyzing the net change in housing units from 2009 to 2024 and examining the data on approved building permits, we can estimate the annual average loss or gain in housing units due to remodeling. It's important to note, however, that these figures are averages and may not capture year-on-year volatility.

3. **How can the analysis be improved with the current limitations of the data?**

Improvement can be achieved through continuous data cleaning efforts, seeking

additional data sources to fill gaps, and employing advanced analytical techniques like machine learning for more nuanced insights. Engaging with local government bodies for more comprehensive data can also be beneficial.

Blockers for Stakeholders

1. Incompleteness of Key Data Fields

We recommend establishing partnerships with local government and other data providers to access more complete datasets. Additionally, implementing a more robust data collection and management system for future projects could prevent such issues.

2. Disorganized and Incorrect Geographical Data

Utilize advanced data cleaning tools and techniques to standardize and correct geographical information. Involving experts in GIS (Geographic Information Systems) could also provide a methodological framework for addressing these issues systematically.

3. Potential Biases in Data Collection and Analysis

Conduct thorough audits of data collection and analysis processes to identify and mitigate any biases. This may involve reviewing the methodologies employed in data cleaning, analysis, and interpretation stages to ensure comprehensive and unbiased insights.

4. Difficulty in Visualizing Complex Data Sets

Leverage more sophisticated data visualization tools and techniques. Collaboration with data visualization experts can help in creating more intuitive and insightful graphical representations of the data.

Challenges and Limitations

- Data completeness and cleanliness were significant challenges, particularly in the 'units' key of the SAM Addresses database and the 'city' column of the permits database.
- The exclusion of 'Stop work' status permits may omit relevant data on halted remodels, potentially affecting the comprehensiveness of our analysis.

Future Research Directions

- Further analysis could investigate the reasons behind halted remodels and the impact of such stops on housing trends.
- A deeper demographic analysis could enhance understanding of the socio-economic factors influencing housing and remodeling trends.

Next Steps and Task Distribution

- **Continued Data Cleaning:** Further cleanse and standardize data across databases for improved analysis.

- **Advanced Trend Analysis:** Apply more sophisticated statistical methods to understand underlying trends in housing unit changes and remodeling activities.
- **Stakeholder Engagement:** Present findings to stakeholders and engage in discussions to refine research questions and approaches based on feedback.
- **Collaborative Research Initiatives:** Explore partnerships with local authorities and real estate developers to enrich our data sets and analyses.

Github link: https://github.com/BU-Spark/ds-boston-remodeling/tree/team-b