City of Boston: Remodeling and Unit Loss Team C Final Report

Team Members	Class Year	Email
Rishab Sudhir (Team Lead)	CAS ‘25	rsudhir@bu.edu
Maha Alali	CAS ‘25	maha9@bu.edu
Wilbert Limson	CAS ‘24	wlimson2@bu.edu
Ashley Harlow	CAS ‘24	aeharlow@bu.edu
Anderson Lawrence	CAS ‘24	andersgl@bu.edu

Table of Contents

Introduction

The City of Boston has been experiencing significant changes in its housing landscape over the past two decades. As the city continues to attract new residents and businesses, the demand for housing has increased, leading to a complex interplay of factors that influence the availability, affordability, and quality of housing units. Understanding these changes is crucial for the City of Boston to develop effective housing policies and urban planning initiatives that meet the needs of its residents and ensure a sustainable future for the city.

This report delves into the impact of remodeling and zoning conversions on the housing markets in Boston, with a specific focus on how these activities may reduce the number of available housing units as buyers convert multi-unit homes into larger, single-unit dwellings. To gain a comprehensive understanding of the situation, our team has thoroughly analyzed various datasets, including property assessment data covering the past 20 years, building permits, and demographic information, which can be found in the appendix. These datasets provide valuable insights into the housing market trends, construction activities, and socio-economic characteristics of affected areas.

Our analysis is divided into four main sections, each focusing on a specific aspect of the housing market:

1. Appreciation of housing prices and its correlation with remodeling and unit loss
2. Loss and gain of residential properties across different neighborhoods
3. Heatmaps and demographic analysis of affected areas
4. Building permits and sentiment analysis of construction activities

The base questions we aimed to answer were:

- What communities are building more housing units?
- Which ones are losing housing units?

- Where are housing remodels and renovations happening?
- How many housing units are lost to remodels on average, each year?

Based on the feedback and our answers to the base questions, we asked the following extension questions:

- What is the correlation between bedrooms and living area?
- If we're not losing bedrooms due to renovations, then what is causing it?
- What are the demographics in areas of loss and gain?

By examining these questions, we aim to uncover trends and patterns that will help inform the City of Boston's decision-making process regarding housing policies and urban planning initiatives. The insights gained from this study will shed light on the communities most impacted by potential shifts in the housing market, the factors contributing to these changes, and the implications for Boston's residents and policymakers. Ultimately, this report seeks to provide actionable recommendations that can help the City of Boston address the challenges faced by its housing market and ensure a more equitable and sustainable future for all its residents.

Appreciation of housing prices and its correlation with remodeling and unit loss

The rationale for this exploration was three-fold: firstly, to determine the magnitude of the problem and whether the City of Boston should take immediate action to avoid putting financial pressure on its citizens; secondly, to identify which areas in Boston are most impacted by this issue; and thirdly, to uncover the greatest correlates of appreciation, such as remodeling and unit loss.

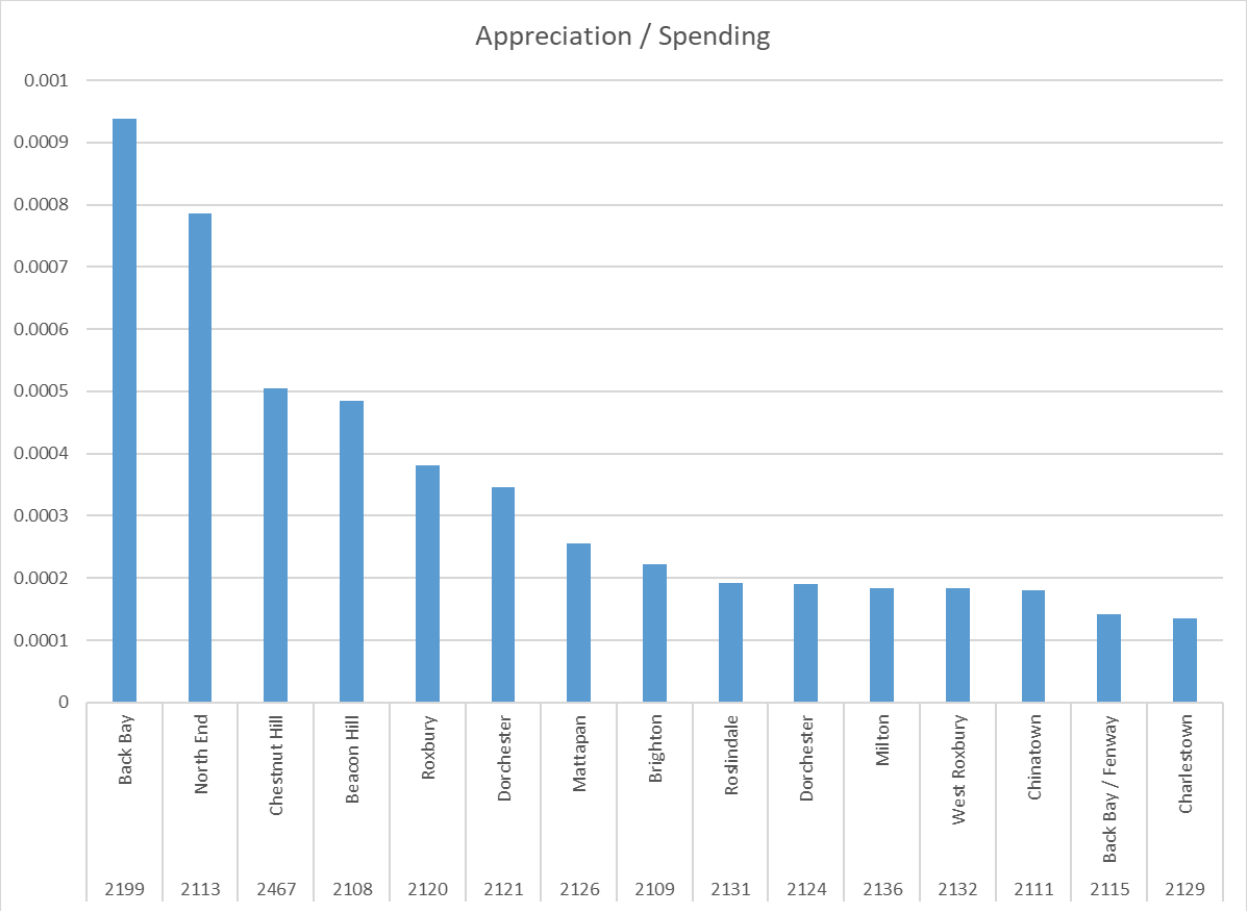
To analyze the correlation between housing appreciation and remodeling in the greater Boston area, we observed the difference in housing prices from 2009 to 2024, as the building permits data starts from 2009. The preprocessing steps involved cleaning the building permits data by removing null, NaN, and 'zero value' rows within the 'declared_valuation' column, converting the object-type 'zip' to a string by accepting the first slice of those zips which contained a hyphen, and resolving the 'declared_valuation' column as a float type. For the property assessment data, we converted the 2024 data to a string and removed comma-separations in 'BLDG_VALUE' and 'TOTAL_VALUE' columns. We then summed up the values for each zip code using each property's 'declared_valuation.' The resulting data captured the total declared valuations for each zip, which was appended to a separate Excel file to record these values for further analysis in future steps.

We initially considered using the total value column from the property assessment data to determine the appreciated value. However, we decided to use the building value instead, as it better captures the potential correlates we were examining, such as spending and remodeling, irrespective of an increase in demand.

To investigate the relationship between spending and appreciation further, we calculated the ratio of spending from the 'total_valuation' data divided by the appreciation of housing prices. The analysis revealed that neighborhoods with low spending on remodeling had exceptionally high ratios, ranging from sixty to one hundred, compared to the majority of neighborhoods with ratios between 10-4 and 10-3. This suggests that the original hypothesis regarding the correlation between spending and inflation was not entirely correct.

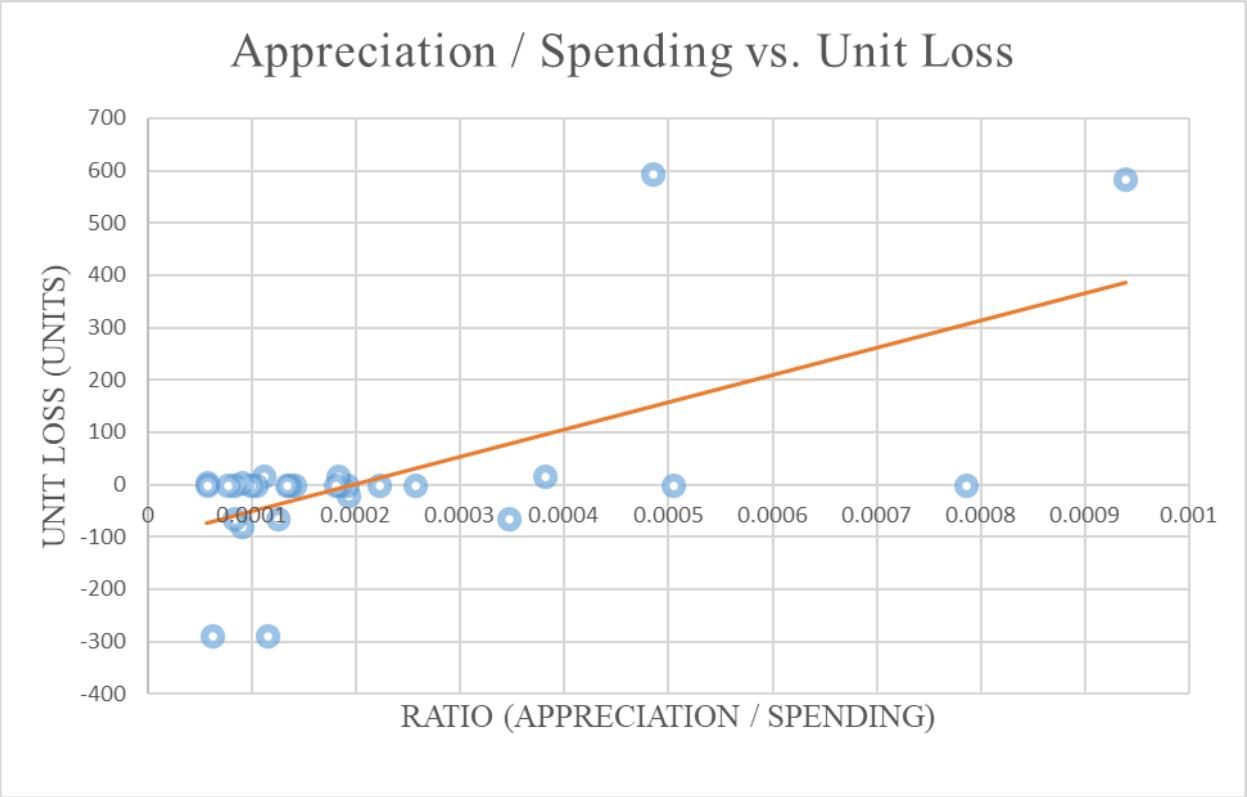
To determine if unit loss was correlated with price appreciation, we compared the total spending data against unit loss in each neighborhood. The data was condensed from 'zip' format to ‘neighborhood’ format, and the fifteen neighborhoods with the highest ratios were plotted (see chart below).

Appreciation / Spending Ratio (Graph)



Finally, we compared the ratios against unit loss for each neighborhood to determine if a positive, negative, or zero correlation existed. The resulting chart (see below) shows a clear positive correlation between unit loss and price appreciation, as the linear forecast line indicates.

Appreciation / Spending Ratio vs Unit Loss (Graph)



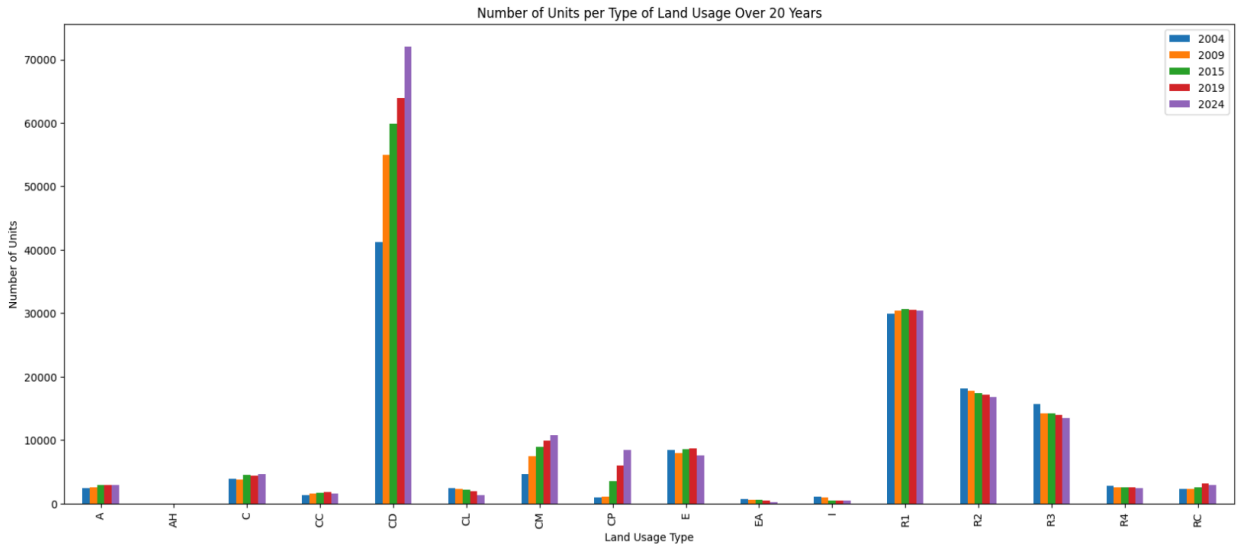
In conclusion, our analysis reveals that unit loss is positively correlated with housing price appreciation in Boston. This finding highlights the need for the City of Boston to address the issue of unit loss, as it directly impacts citizens' ability to afford and maintain their housing status. By identifying the areas most affected by this issue and understanding the key correlates of appreciation, policymakers can develop targeted strategies to mitigate the negative impacts of unit loss on Boston's residents.

Loss and gain of residential properties across different neighborhoods

Residential Property Loss and Gain

When working to determine how many residential properties are lost per year, we started by looking for general trends in the number of properties in Boston. We used the land usage feature of the property assessment data set which tells us if a given property is residential, commercial, industrial, and so forth. To look at general trends, we looked at 20 years' worth of data in 5-year intervals, where we grouped the data by land usage type and counted the number of properties per land usage type for each year. We then plotted the counts for each kind of land usage for each year together to get an idea of which land usage types were gaining properties and which were losing properties.

Number of Units Per Type of Land Usage Over 20 Years (Graph)



Land Usage Codes (Index)

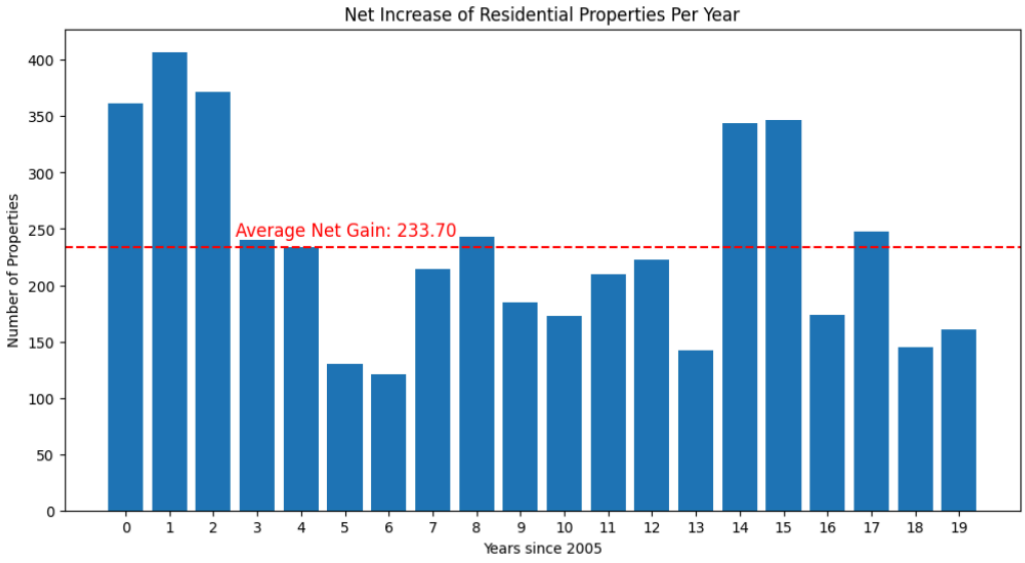
Code	Type
A	Residental (7 or more Units)
AH	Agricultural/Horticultural
C	Commercial
CC	Commercial Condominium
CD	Residental Condominium Unit
CL	Commercial Land
CM	Condominium Building (excluding units)
CP	Condominium Parking
E	Tax Exempt
EA	Tax Exempt (121A)
I	Industrial
R1	Residental One-Family Home
R2	Residental Two-Family Home
R3	Residental Three-Family Home
R4	Residental Four-Family Home
RC	Mixed Use (Residental and Commercial)

Here, we can see that for most types of land usage, we either have around the same number of properties over all 20 years or gain property. In particular, we are gaining lots of residential condominium units, condominium buildings, and condominium parking lots; however, we are ever so slightly losing multi-family properties. However, since condominiums and family homes are residential, this could suggest that we are both losing and gaining residential properties, but we may still be gaining more than we are losing.

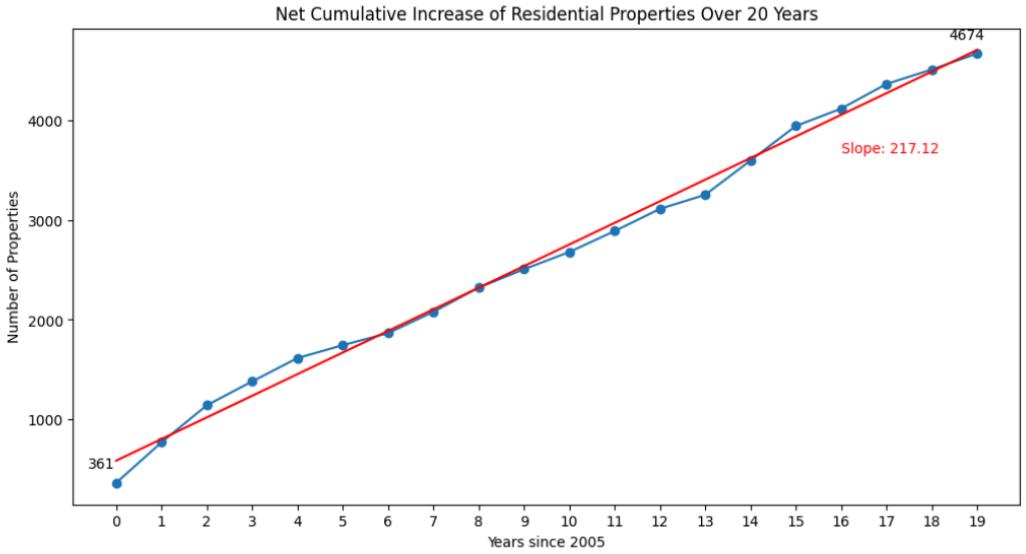
After considering the client feedback from the early insights meeting, we started looking specifically at the gain and loss of only residential properties and how many of the lost residential properties were due to remodeling. Again, by using the land usage feature of the property assessment data set, we were able to find the number of residential properties we had

per year in the city of Boston. We grabbed a subset of properties with the land usage codes R1, R2, R3, and R4, which stand for one, two, three, and four-family homes, respectively, as well as properties with land usage codes RC, A, and CM which stand for mixed-use residential and commercial properties, residential properties with seven or more units, and condominium buildings. Once we had this subset for each year of data, we could use the unique parcel ID values associated with each of the residential properties per year and check all other years for that same parcel ID number. We can then count the number of properties that were not present the year before but are present now and the number of properties that were present that year but not the year after to get the net gain or loss of properties that year. We performed this check on all 20 years of data that we had, giving us the following two plots.

Net Increase of Residential Properties Per Year (Graphs)



Net Cumulative Increase of Residential Properties Over 20 Years (Graph)

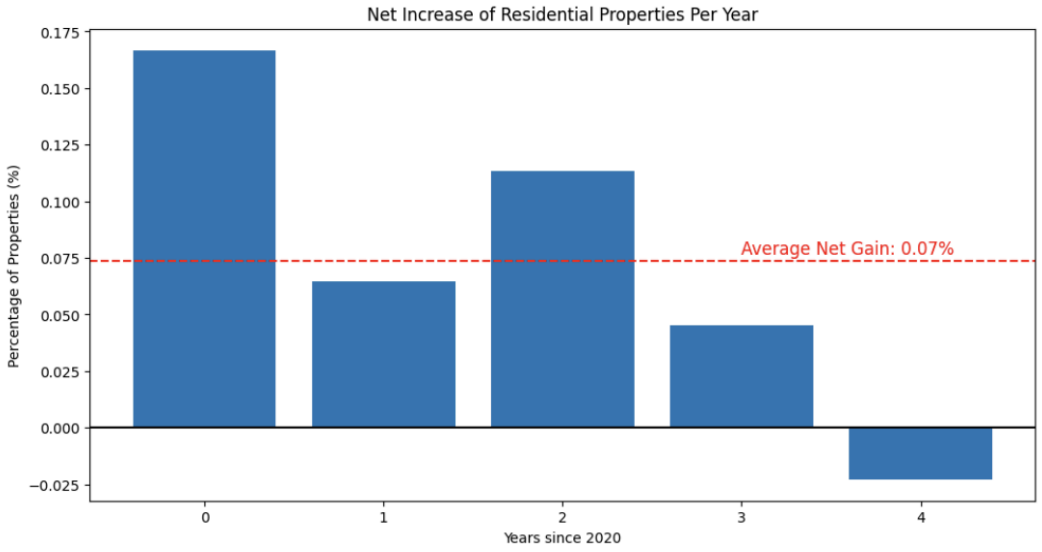


As we can see, over the last 20 years, we have only seen a net increase in the number of residential properties, with an average of around 230 residential properties per year and we have gained 4674 residential properties in total. Even though we do have a net gain, that doesn't mean that we don't lose any properties. It just means that we gain more than we lose, so now let's turn our attention to another feature of the property assessment data set that tells us the year of the most recent remodel of a given property. If a property in our data set had a remodel between the years of 2004 and 2024, we can look at the land usage of that property for the year before and the year after to see if the land usage has changed and if it has, see if the new land usage is residential. After working with the data, we found that there were roughly 200,000 documented remodels between 2004 and 2024, and of those remodels, around 114,000 were related to residential properties. Then, after comparing the land usage before and after a year where a remodel occurred, we found that a mere 0.22% of the remodels were residential properties remodeled into non-residential properties, equating to around 250 properties. In addition to this, we found that 0.94% of the remodeled properties, or around 1,068 properties, were converted from non-residential to residential, and the remaining 98.84% of

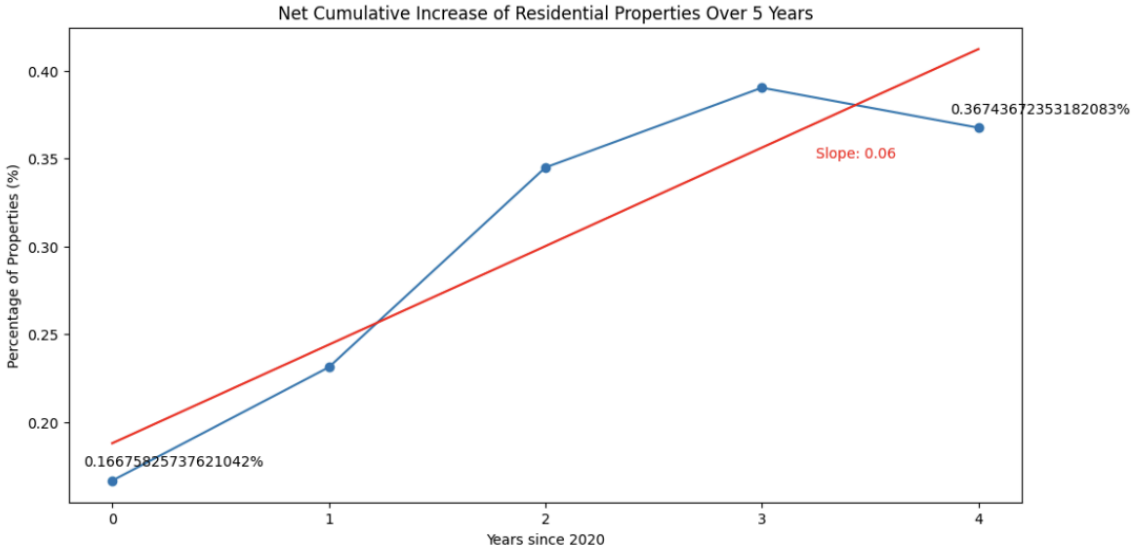
properties started as residential and remained residential. However, we want to note that this does not take into account the difference between remodels that did not change the land usage at all and remodels that started as and stayed residential but changed types of residential, such as a residential property that began as a four family home that became a two family home, or a one-family home that became a three family home, just to name a few examples.

After considering the client feedback from the mid-semester presentation, we decided to look at how the observed net gain of residential properties is spread across the neighborhoods of Boston to determine which neighborhoods, if any, are losing more properties than others. We were able to determine the neighborhoods of residential properties by merging the property assessment data that we have already been using with the SAM data set, which tells us the mailing neighborhood of a given property. However, it is worth noting that we experienced a decent amount of data loss when merging these two data sets. Naturally, when working with two separate data sets, there are some properties that are in one data set but not the other, so if we have any residential properties that exist in the property assessment data and not the SAM data set and vice versa, they are lost in the merge. We experienced around a 33% data loss per year. Since we lost about a third of the data, let's revisit the analysis of Boston as a whole before splitting the data up into individual neighborhoods to check if this data loss significantly impacted the original result we saw.

Net Increase of Residential Properties Per Year (Graph)



Net Cumulative Increase of Residential Properties Over 5 Years (Graph)

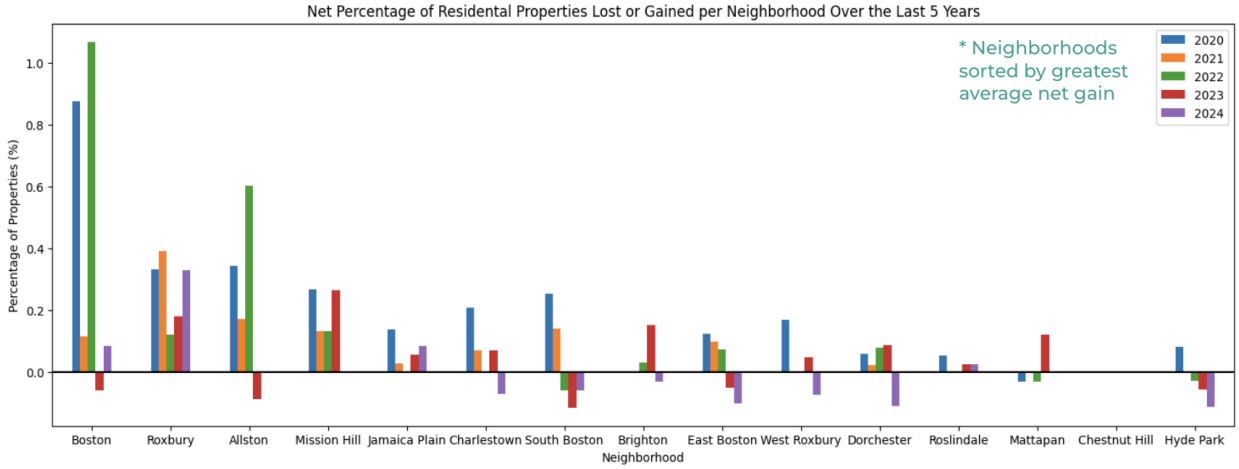


As we can see, the data loss significantly impacted our previous results, as we now observe a net loss in 2024 instead of the steady net gain we saw across all years of data previously. On average, we still have a net gain of properties of 0.07%; however, because of the significant data loss, it is hard to say if this is a genuinely representative result. Also, we would like to note that in accordance with the feedback we received from the client, we began focusing

on the last five years of data rather than all twenty years. Instead of focusing on the count of properties, we are now looking at percentage loss and gain to better understand how large of an impact the loss or gains have on each neighborhood.

We then grouped the merged data set by neighborhood, counted the number of residential properties per neighborhood per year, and plotted them together to see if we could observe any property gain or loss trends for each neighborhood.

Net Percentage of Residential Properties Lost or Gained per Neighborhood Over the Last 5 Years (Graph)



As the highest net gain for a single year however, it is also important to note that the highest net gain for a single year is still just over a 1% net gain of properties, so even at our most significant net gain, we are still talking about a very small number of properties. Looking at the other neighborhoods, we see that most of the neighborhoods are experiencing mostly net gain, with net loss happening primarily in 2023 or 2024, which, as we mentioned before, the city-wide net loss that we observe in 2024 may or may not be representative of the actual number of gained and lost properties because of the data loss. However, it is interesting to note that in South Boston, East Boston, and Hyde Park, we can observe a very steady, smooth decrease in properties, starting with net gains in 2020 and ending with net losses in 2024. This is in contrast to all the other neighborhoods, which seem to have a more sporadic net gain or loss of

properties with little to no trend year to year, which may suggest that there is, in fact, an overall loss of properties in South Boston, East Boston, and Hyde Park which may continue as time goes on.

Insights on Bedroom and Living Area Loss/Gain

Early Insights

To gain a better understanding of the trends in Boston's housing market, we analyzed the property assessment datasets, specifically examining the 'LIVING_AREA' and 'BED_RMS' columns over a 20-year period as we had a feeling that this could give us more granular insight into the loss/gain of residential units. We filtered the data based on residential codes and removed rows with missing information on the bedroom and living area. We then plotted the average values for living areas and bedrooms for residential properties by year.

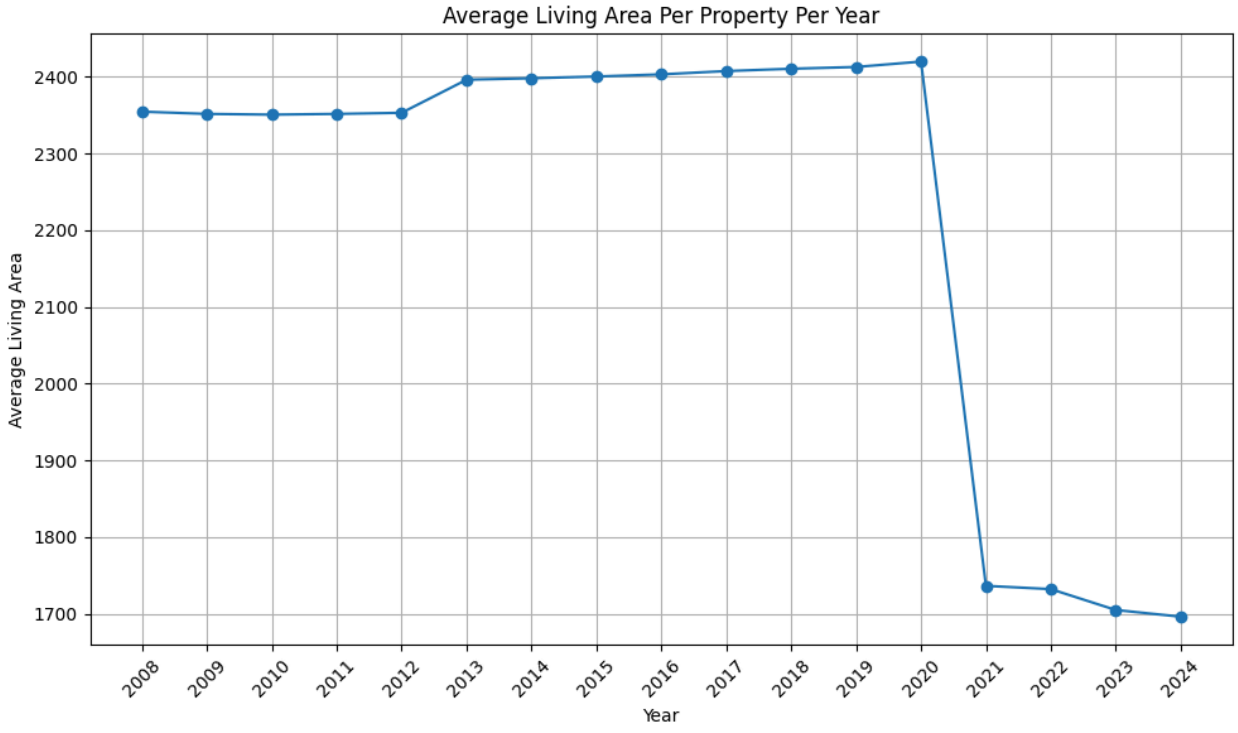
Average Number of Bedrooms in Residential Properties by Year (Graph)

The graph indicates a stable average number of bedrooms in residential properties from 2008 until around 2020, maintaining around 4.5 bedrooms. Post-2021, there's a sharp decline in

the average number of bedrooms, which suggests a significant change in housing trends or data collection methods.

The decline may be related to renovations but also given the timing it may also be related to COVID-19 impact. The pandemic definitely has influenced housing trends, such as increased demand for larger, multifunctional spaces, which could have lead homeowners to combine bedrooms to create larger or more functional spaces. But im not sure if we have a way to actually verify this for now, so we shall continue with the hypothesis that its related to renovations.

Average Living Area in Residential Properties by Year (Graph)

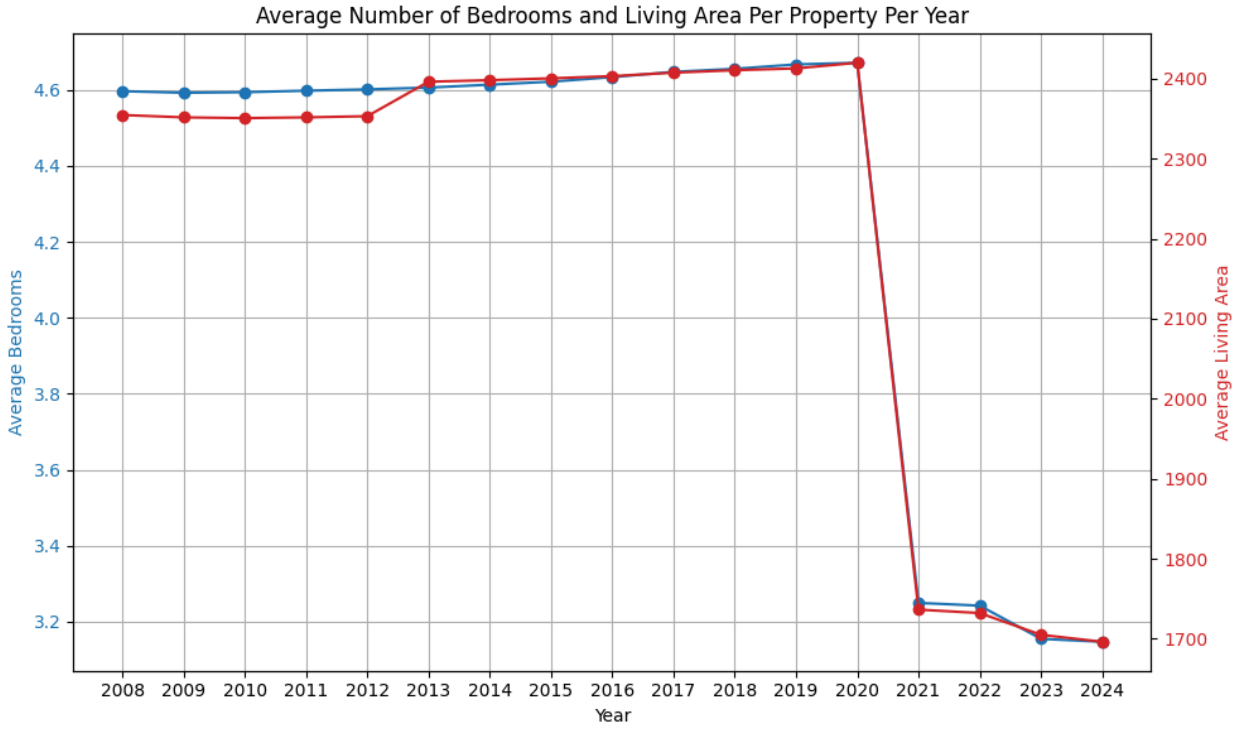


This graph showing the average living area per property from 2008 to 2024 depicts a similar sharp decline post-2020, just as we saw with the average number of bedrooms. The

consistency in the timing of these declines supports the idea that there might be a common underlying factor affecting both the number of bedrooms and the overall living area.

Given how both graphs show similar trends it may be possible that renovations happened due to COVID-19 or other societal changes during the pandemic, it would be plausible that such renovations could lead not only to the combination of bedrooms for more oversized bedrooms or other living spaces but also to a general reduction in the usable living area. This might occur if homeowners are repurposing space for home offices, fitness areas, or for more open living configurations which could reduce the total 'counted' living area if measurements are changed or areas are repurposed.

Average Number of Bedrooms and Living Area Per Property Per Year (Graph)



Overlaying the graphs of the average number of bedrooms and living area per property provides a clear visual correlation between the two metrics over time. Both show remarkable stability from 2008 until 2020, followed by a sharp decline. The simultaneous decrease in both the number of bedrooms and the overall living area from 2021 onwards suggests that the

changes in one are likely connected to changes in the other. So the next step was to investigate whether renovations were causes these decreases.

Is Renovations Causing Loss?

To investigate the impact of renovations on the number of bedrooms and living areas in residential properties, we performed the following analysis:

1. Grouping Properties: We grouped properties by their unique parcel ID (PID) across all available years in the dataset (2008-2024).

2. Identifying Renovated Properties: For each property, we checked if it had been remodeled within the time frame of our dataset. If a property was renovated, we stored the year before the renovation and the year after the renovation (or 2024 if it was the last available year) in a separate dataset.

3. Comparing Before and After Renovation: We created a new dataset called "renovation_comparison" that contains paired rows for each renovated property, with one row representing the year before the renovation and the other row representing the year after the renovation. This dataset allows us to calculate the differences in bedrooms and living areas for each renovated property.

Renovation Comparision Dataset (Dataset)

```
[20]  print(renovation_comparison.head(20))
```

```
...          PID  DATA_YEAR  LIVING_AREA  YR_REMODELLED  BED_ROOMS  REMODEL_YEAR
0   100028000       2013       2376.0         1988.0        6.0        2015.0
1   100028000       2019       3275.0         2015.0        8.0        2015.0
2   100029000       2013       4457.0         2004.0        9.0        2015.0
3   100029000       2018       4475.0         2015.0        9.0        2015.0
4   100035000       2013       2820.0         1985.0        3.0        2018.0
5   100035000       2020       2806.0         2018.0        5.0        2018.0
6   100058000       2013       3840.0         1975.0        8.0        2015.0
7   100058000       2018       3651.0         2015.0       11.0        2015.0
8   100104000       2019       3308.0         2017.0        8.0        2020.0
9   100104000       2022       3308.0         2020.0        8.0        2020.0
10  100105000       2013       3677.0         1991.0        6.0        2016.0
```

The "renovation_comparison" dataset includes the following columns: PID, DATA_YEAR, LIVING_AREA, YR_REMODELLED, BED_ROOMS, and REMODEL_YEAR. For example, rows 0 and 1 represent a property with PID 100028000, with row 0 containing data from 2013 (before renovation) and row 1 containing data from 2019 (after renovation).

Our analysis included 10,203 rows, representing 5,101 renovated properties.

4. Calculating Changes in Bedrooms and Living Areas: To visualize the changes in bedrooms and living areas after renovations, we calculated the differences for each property by subtracting the "before" values from the "after" values.

5. Analyzing the Results: We calculated the number of properties that gained, lost, or had no change in bedrooms and living areas after renovations.

Gained/No Change/Lost in number of Bedrooms/Living Area (Table)

	Gained	No Change	Lost
Bedrooms:	942	4,339	399
Living Area:	1,825	2,870	985

We also calculated the average changes in bedrooms and living areas, both including and excluding properties with no change:

Average Changes in Bedrooms/Living Area

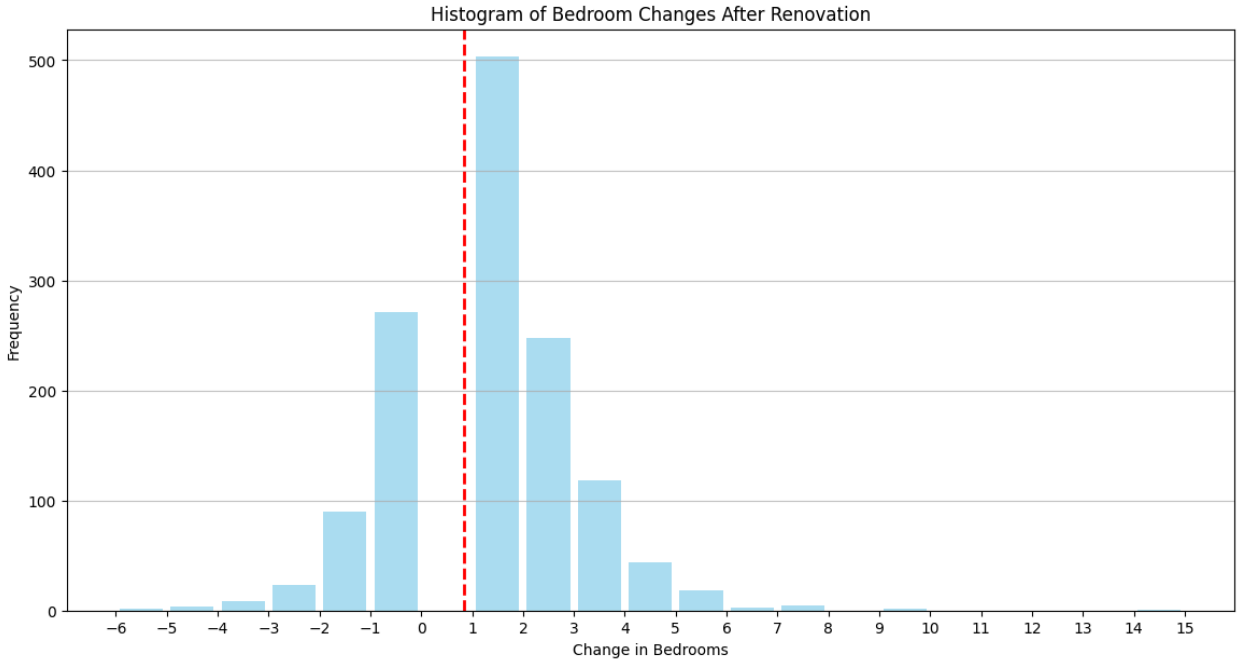
	All Properties	Excluding No Change
Average Change in Bedrooms	+0.196	+0.832
Average Change in Living Area	+81.99 sq ft	+ 165.73 sq ft

For bedrooms, the average change among properties that had any change (either gain or loss) in the number of bedrooms is approximately +0.83 bedrooms. This suggests that, on average, properties tended to gain a small number of bedrooms when there was a change. For living area the average change among properties that had any change (either gain or loss) in living area is approximately +142 square feet. This indicates that properties that saw changes tended to gain more living area on average.

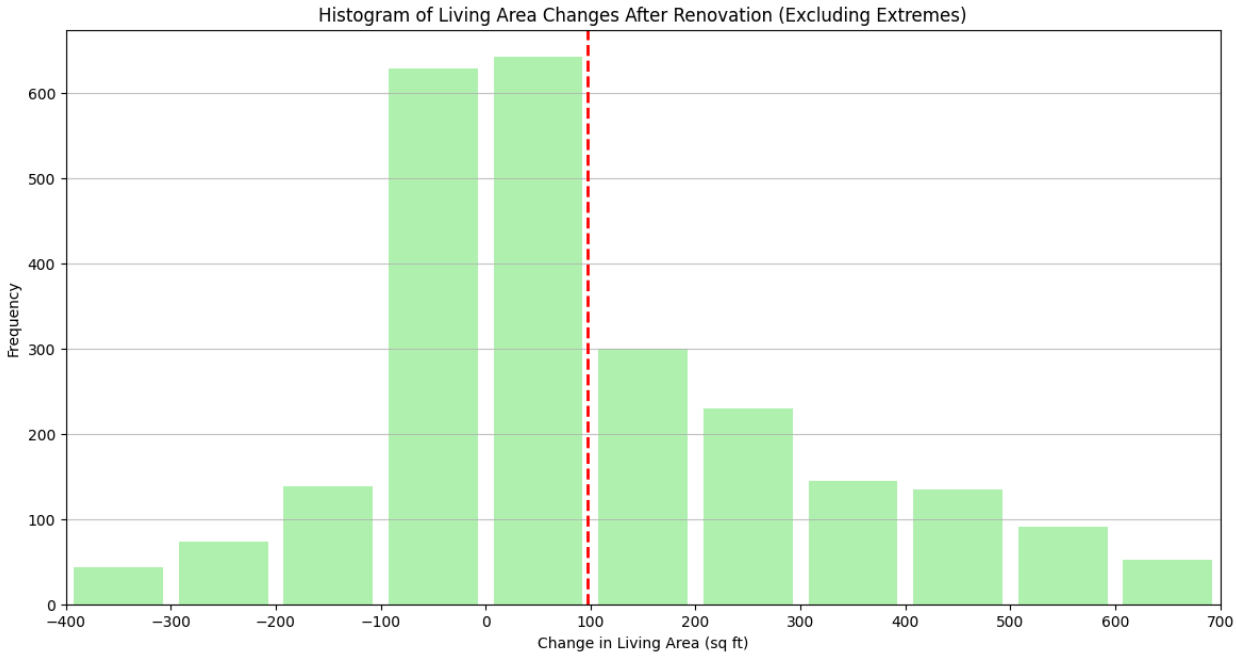
However, it's important to keep in mind that this doesn't reflect the distribution of changes, as it's possible for a few large increases to skew the average upward, even if most changes are small. To get a complete picture of the renovations' impact, let's look at the distribution of changes and consider both the averages and the totals of increased and decreased living areas and bedrooms.

Based on the feedback from this section of analysis I was given two tasks. One was to look into the correlation between living area and bedrooms. The second was to see if renovations aren't causing renovations then what are? An analysis of the other primary factors. On top of that, an extension task we proposed was looking into the demographics that reside in neighborhoods of loss, which is explored in the Heatmaps and Demographic Analysis section.

Distribution of Bedroom Changes (Graph)



Distribution of Living Area Changes (Graph)



The histograms show the distribution of changes in bedrooms and living area after renovation. The bedroom histogram shows that the average change, represented by the red line, lies around the modal class of +1 bedroom, suggesting that renovations tend to result in a small gain in bedrooms rather than a loss. The living area histogram reveals that most renovated properties experience changes of around +/- 100 sq ft, with some extreme cases of larger gains skewing the average upward.

For the living area histogram we had to remove extreme values that make the rest of the data hard to read. These extreme values, also known as outliers, caused the scale of the x-axis to be very wide, which compressed the majority of the data into a small number of bins. We decided a reasonable range using the IQR of the data, and defined the bounds as such **lower_bound = Q1 - 1.5 * IQR**, **upper_bound = Q3 + 1.5 * IQR**.

Our analysis indicates that, on average, renovations do not result in a loss of bedrooms or living areas in residential properties. In fact, net averages show modest gains in both bedrooms and living areas. However, it's important to note that the majority of renovated

properties experience no change in these attributes, and we have also had a significant amount of data loss. Since each property assessment data has varying numbers of rows (e.g., 2008 has 155,564, 2017 has 170,910, and 2024 has 182,242), and we ended up with only around 5,000 properties that have been renovated and that we have data on before and after renovation, we are only analyzing a small chunk of the properties given. Therefore, it is possible that this is not representative of the larger Boston Area. Further research with a more comprehensive dataset would be necessary to draw definitive conclusions about the impact of renovations on bedrooms and living areas in Boston's residential properties.

Correlation between Living Area and Bedrooms

Based on the feedback from this section, we were given two tasks. One was to look into the correlation between the living areas and bedrooms. The second was to see if renovations aren't causing renovations then what are? An analysis of the other primary factors. On top of that, an extension task we proposed was looking into the demographics that reside in neighborhoods of loss, which is explored in the Heatmaps and Demographic Analysis section.

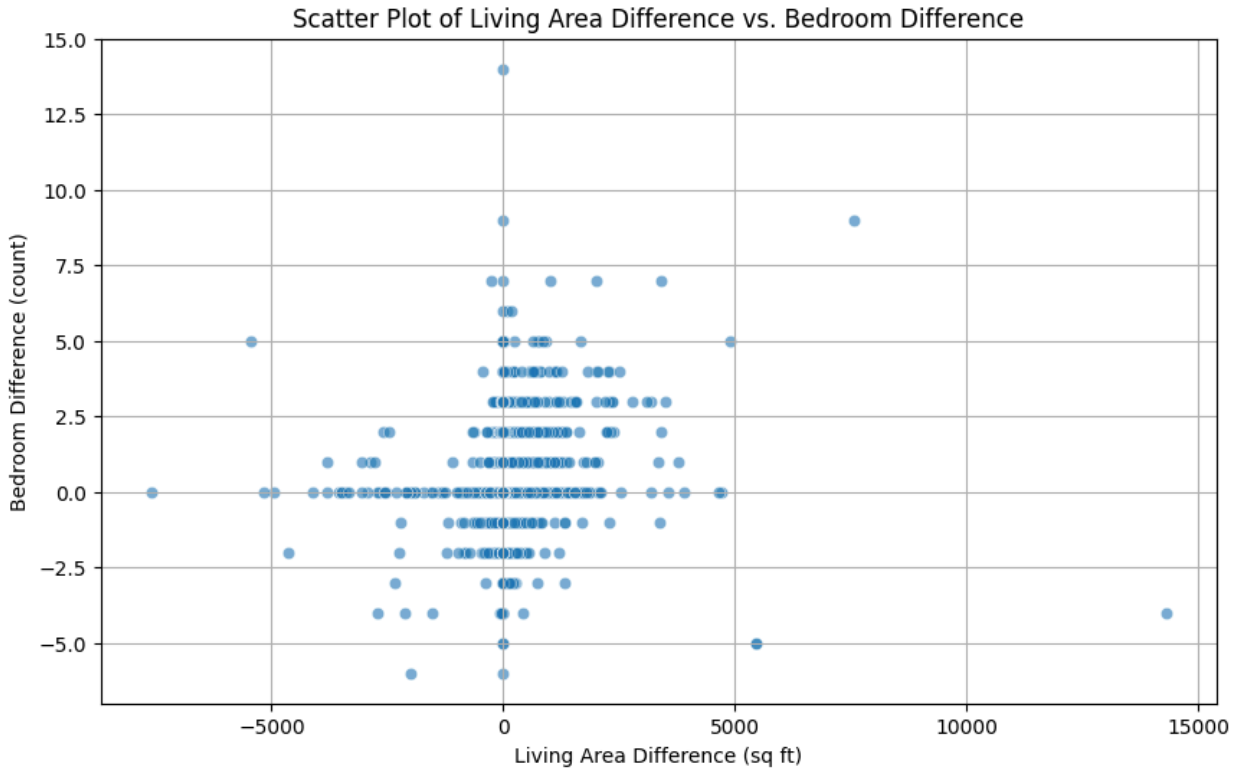
To investigate the relationship between changes in living area and changes in the number of bedrooms due to renovations, we first calculated the Pearson correlation coefficient using the .corr() function.

Living Area/Bedroom Difference Correlation Matrix (Table)

```
Correlation matrix:
                 living_area_diff  bedrooms_diff
living_area_diff          1.000000       0.159131
bedrooms_diff             0.159131       1.000000
```

The Pearson correlation coefficient, r, can take a range of values from +1 to -1. A value of 0 indicates that there is no association between the two variables. The correlation matrix revealed a positive but relatively weak correlation of approximately 0.159 between living_area_diff and bedrooms_diff. This suggests that while larger renovations that increase living space may sometimes also increase the number of bedrooms, this is not a consistent trend. To visualize this relationship, we created a scatter plot with living_area_diff on the x-axis and bedrooms_diff on the y-axis.

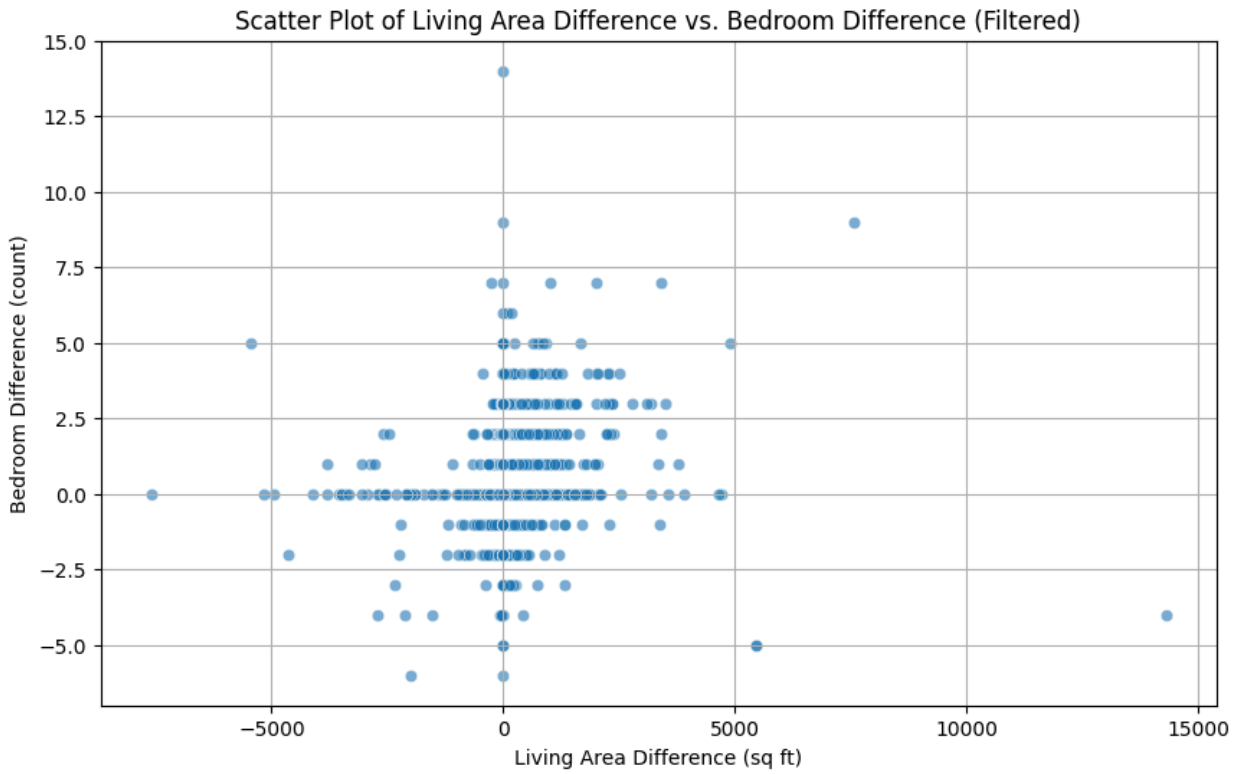
Scatter Plot of Living Area Difference and Bedroom Difference (Graph)



The scatter plot confirms the weak positive relationship between the two variables. The wide distribution of points indicates that changes in living area do not consistently correspond with changes in the number of bedrooms. This could be because some renovations expand the living area without altering the bedroom count, such as by enlarging living rooms or adding new non-bedroom spaces. To further investigate this relationship, we filtered out entries where there

was no change in either the living area or bedrooms, hypothesizing that this might provide a clearer view of the relationship between the variables.

Scatter Plot of Living Area Difference and Bedroom Difference Filtered (Graph)



However, the filtered data showed only a slight decrease in the correlation coefficient, from 0.159 to 0.1419. The scatter plot of the filtered data also showed a wide spread of points, confirming that the changes in living area and bedrooms due to renovations are largely independent of each other. Given the weak linear relationship between changes in living area and changes in the number of bedrooms, we decided to use a linear regression model to explore other variables that might be more strongly related to these changes.

Investigating Predictors of Bedroom Differences Using Linear Regression

To explore potential predictors of bedroom differences, we selected several variables that we hypothesized might be related to changes in the number of bedrooms during renovations. These variables included LIVING_AREA, LAND_SF, TOTAL_VALUE, GROSS_TAX, GROSS_AREA, HEAT_TYPE, AC_TYPE, EXT_FIN, FPLACE, OVERALL_COND, and ROOF_STRUCTURE. We focused on data from 2015-2024 to align with the client's request for more recent data.

A more in-depth analysis of the preprocessing is found in the notebook Linear Regression on Bedrooms.ipynb, but high level, here is what we did Before applying linear regression, we performed several data preprocessing steps:

1. We used regular expressions to standardize the data in the 'ROOF_STRUCTURE', 'EXT_FNISHED', 'HEAT_TYPE', and 'AC_TYPE' columns, extracting the first letter of each entry.
2. We applied one-hot encoding to convert categorical variables into a format suitable for linear regression.
3. We checked for missing values and found that the data loss was not significant enough to warrant imputation. We dropped the rows with missing values, reducing the dataset from 1,476 to 1,420 rows.
4. We scaled the data using Sklearn's StandardScaler to normalize the variables with a mean of 0 and a standard deviation of 1.

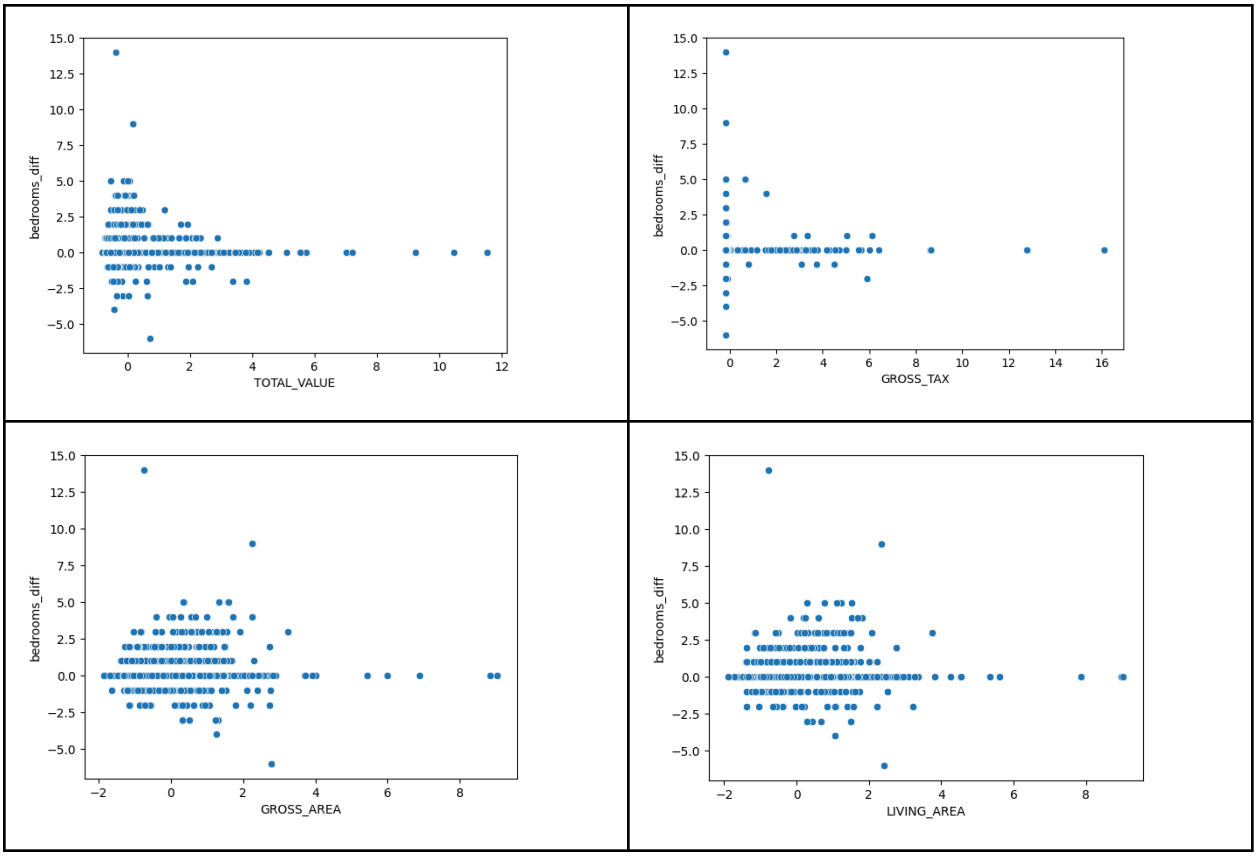
Before proceeding with linear regression, we checked if the assumptions of linearity, independence, homoscedasticity, and normality of residuals were met. We created scatter plots of the numeric predictor variables against the target variable (bedroom difference).

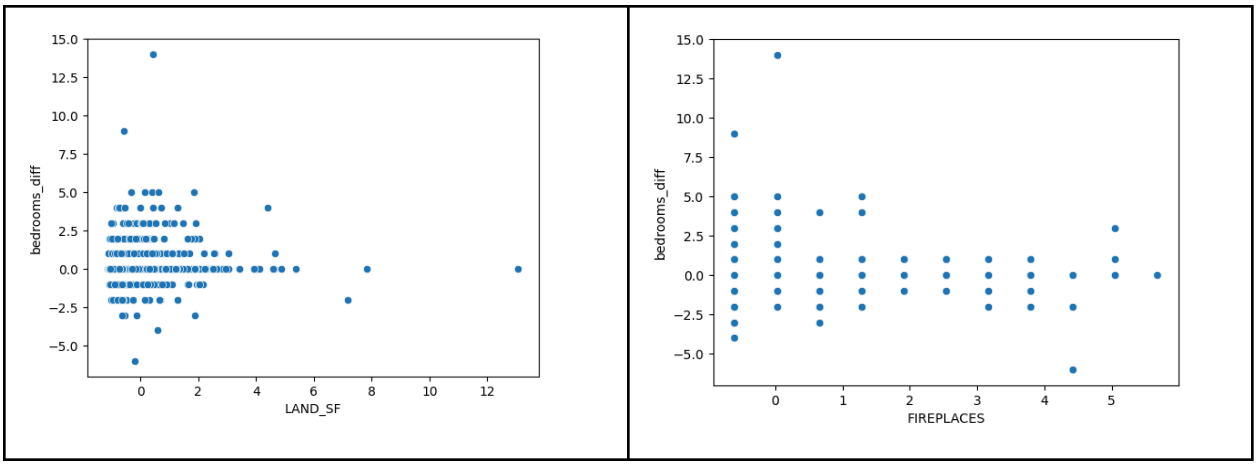
Snippet of Linear Regression Table (Dataset)

	TOTAL_VALUE	GROSS_TAX	GROSS_AREA	LIVING_AREA	LAND_SF	FIREPLACES	BED_ROOMS	living_area_diff	bedrooms_diff	LU_R2	...	EXT_FNISHED_U	EX
1	-0.404587	-0.179563	0.530974	0.403184	-0.483862	-0.604689	1.466583	-0.273993	0.0	0	...	0	
3	-0.729476	-0.186297	-1.291872	-1.292828	-0.483862	-0.604689	-0.876518	-0.274798	0.0	0	...	0	
5	-0.742812	-0.186567	-1.401610	-1.242148	-0.483862	-0.604689	-0.876518	-0.273993	0.0	0	...	0	
7	-0.517219	-0.181992	-0.950793	-1.124181	-0.483862	-0.604689	-0.876518	-0.273993	0.0	0	...	0	
9	-0.255644	-0.176503	1.719710	1.859148	-0.565463	-0.604689	3.341064	-0.273993	0.0	0	...	0	
...	...	...	...	...	...	...	...	...	...	...	...	...	
2848	-0.528390	0.658437	0.337004	0.772608	-0.326331	0.023444	0.529343	2.028384	5.0	1	...	0	
2850	-0.585036	0.545162	-0.781143	-0.269695	-0.714803	1.907845	0.529343	-0.343762	0.0	1	...	0	
2852	-0.682139	0.350985	-0.855884	-0.925500	-0.717323	-0.604689	0.060723	-0.273993	0.0	1	...	0	

Scatter Plot of Dependant Variables(X) Against Independent Variable(Y)

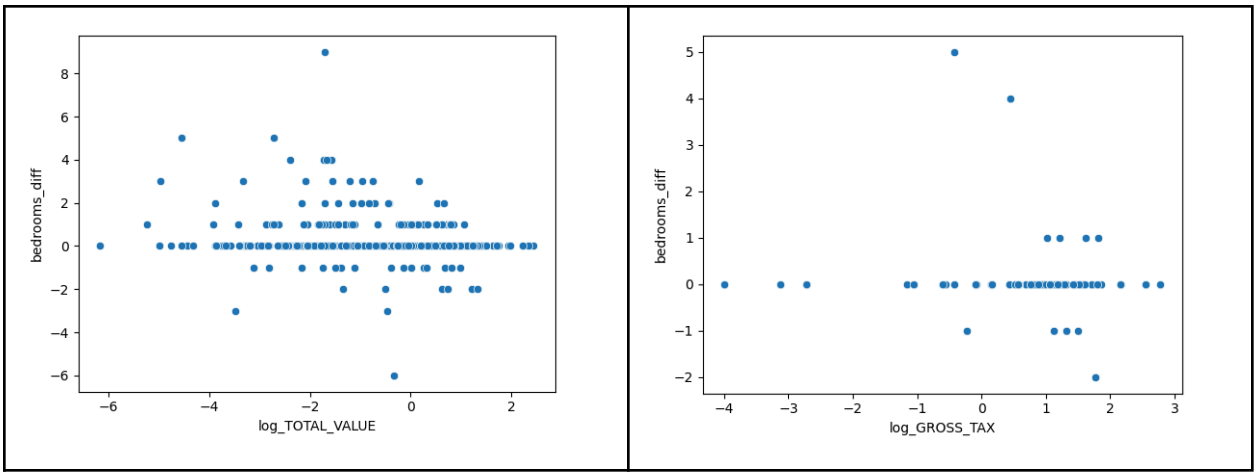
Bedroom_Diff (Graph)

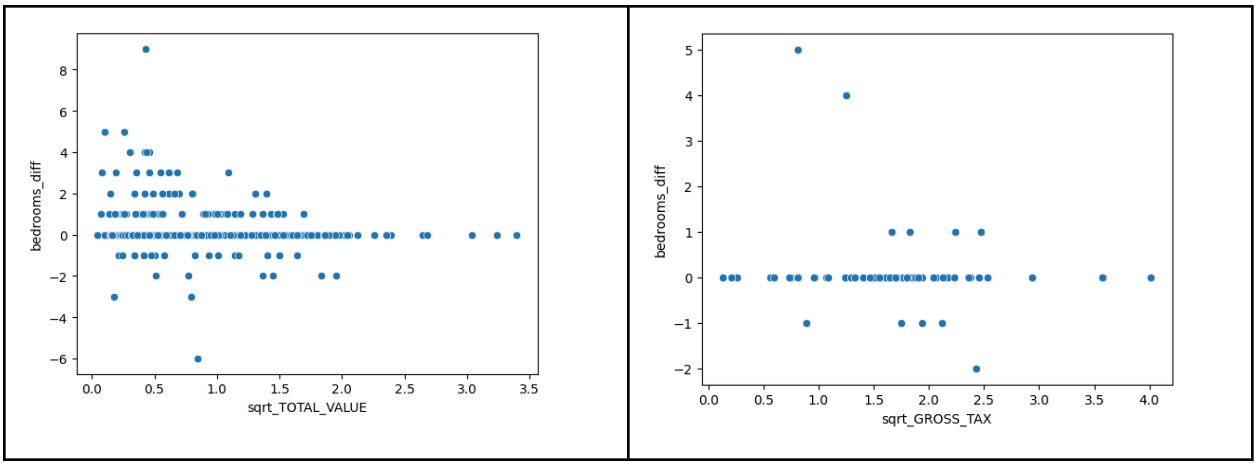




The scatter plots revealed that the relationships between the predictor and target variables were not strictly linear. We observed heteroscedasticity, outliers, and no clear linear patterns, suggesting that linear regression might not be the most suitable model for this data. However, before stopping, I wanted to try some transformations of the dependent variables to try to linearize the data. To attempt to linearize the relationships, we applied log and square root transformations to the TOTAL_VALUE and GROSS_TAX variables.

Scatter Plot of Transformed Dependant Variables(X) Against Independent Variable(Y) Bedroom_Diff (Graph)

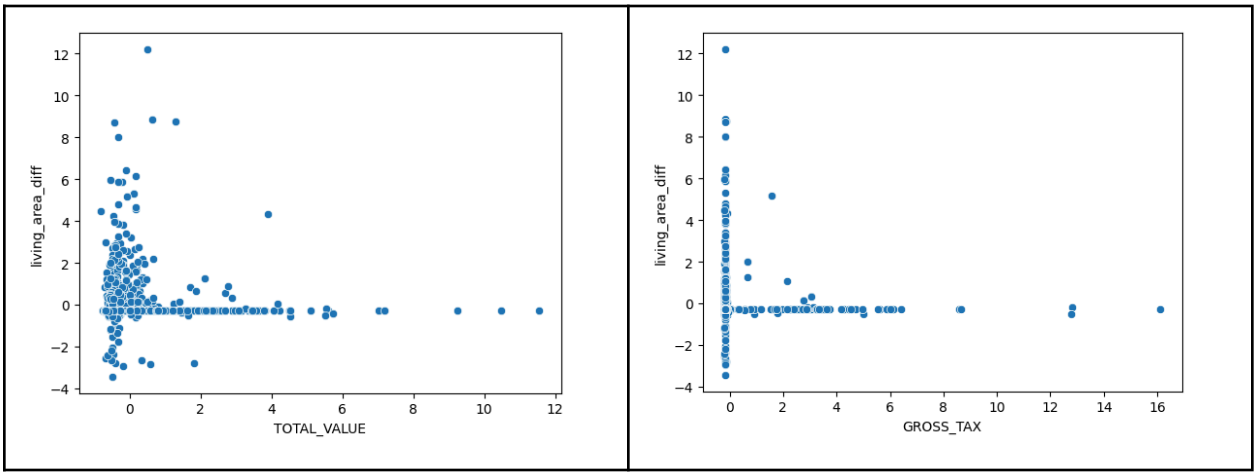


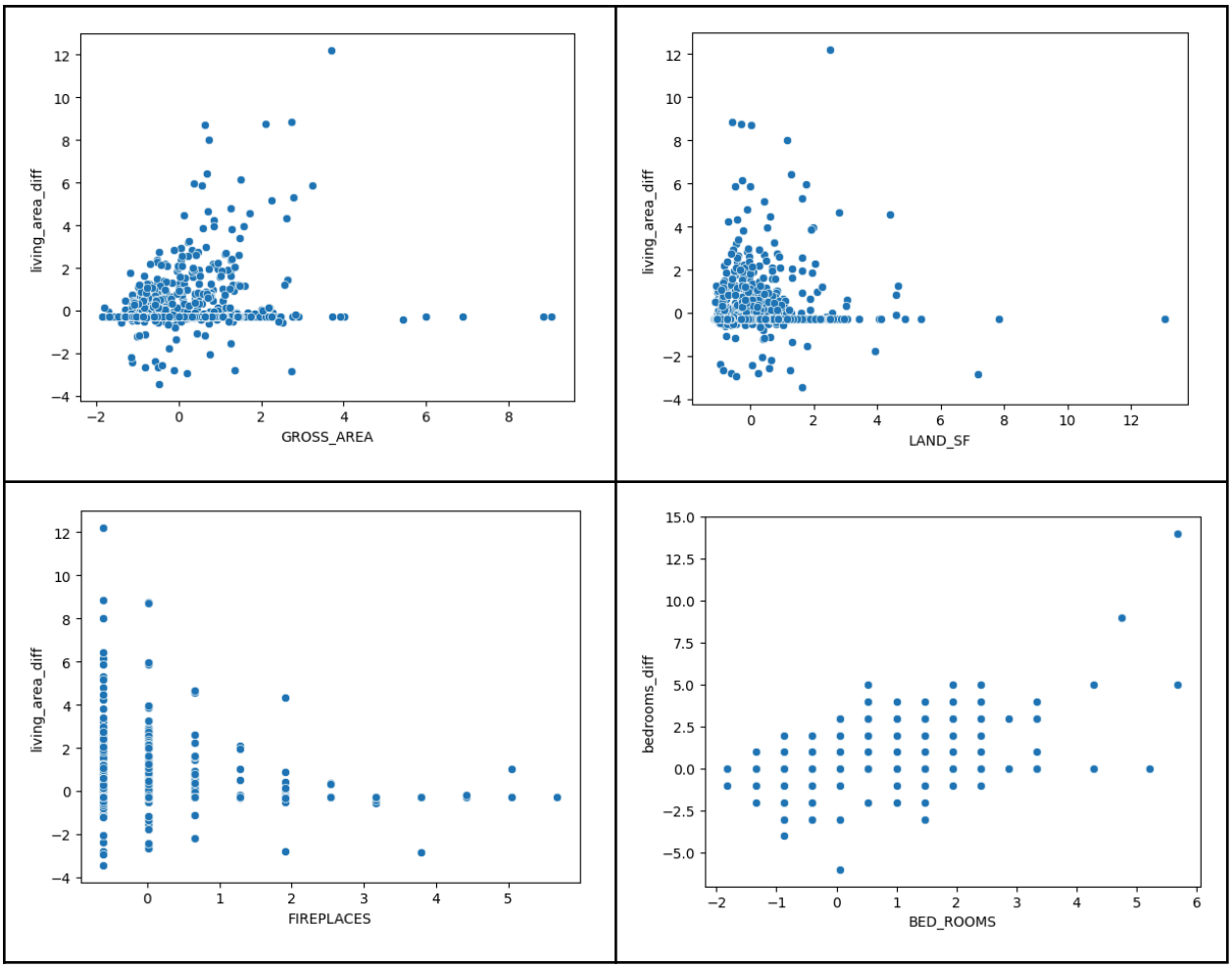


However, even after these transformations, we still did not observe clear linear relationships between the predictor variables and the target variable. After exploring transformations we still didn't see any linear relationships between the predictor variables and the target variable, so linear regression is not the most suitable model for the data... but we checked how it performs for living_area_diff.

Scatter Plot of Dependant Variables(X) Against Independent Variable(Y)

Living_Area_Diff (Graph)





Given the lack of clear linear relationships, we concluded that linear regression was not the most appropriate model for predicting bedroom and living area differences. Due to time constraints, we were unable to explore other models that do not explicitly require linearity, such as classification trees, Naive Bayes, K-Nearest Neighbors, or Support Vector Machines. These models could potentially provide better insights into the factors influencing bedroom differences during renovations.

While we were unable to pursue alternative models within the scope of this project, we believe that investigating these models could yield valuable insights. We recommend further exploration of these alternative models in future research to better understand the predictors of bedroom and living area differences in residential renovations.

Given the limitations of linear regression for this dataset, we decided to shift our focus to investigating the demographics of areas experiencing residential property loss, which is explored in this report's "Heatmaps and Demographic Analysis" section.

Heatmaps and demographic analysis of affected areas

In this section, we explore the demographic characteristics of neighborhoods experiencing changes in the number of bedrooms and residential units due to renovations. By visualizing these changes using heatmaps and analyzing the demographic composition of affected areas, we aim to gain insights into the populations most impacted by these housing market shifts.

Data Preparation and Challenges Bedrooms

To create heatmaps displaying changes in bedrooms by neighborhood, we first needed to merge our existing datasets (the increased/decreased and no change bedrooms data) with the SAM (Street Address Management) dataset containing latitude and longitude information for each property. However, during this process, we encountered significant data loss, with nearly a 50% reduction in the number of rows after merging. Despite this challenge, we proceeded with the analysis using the available data, acknowledging that the results may not be fully representative of all neighborhoods in Boston.

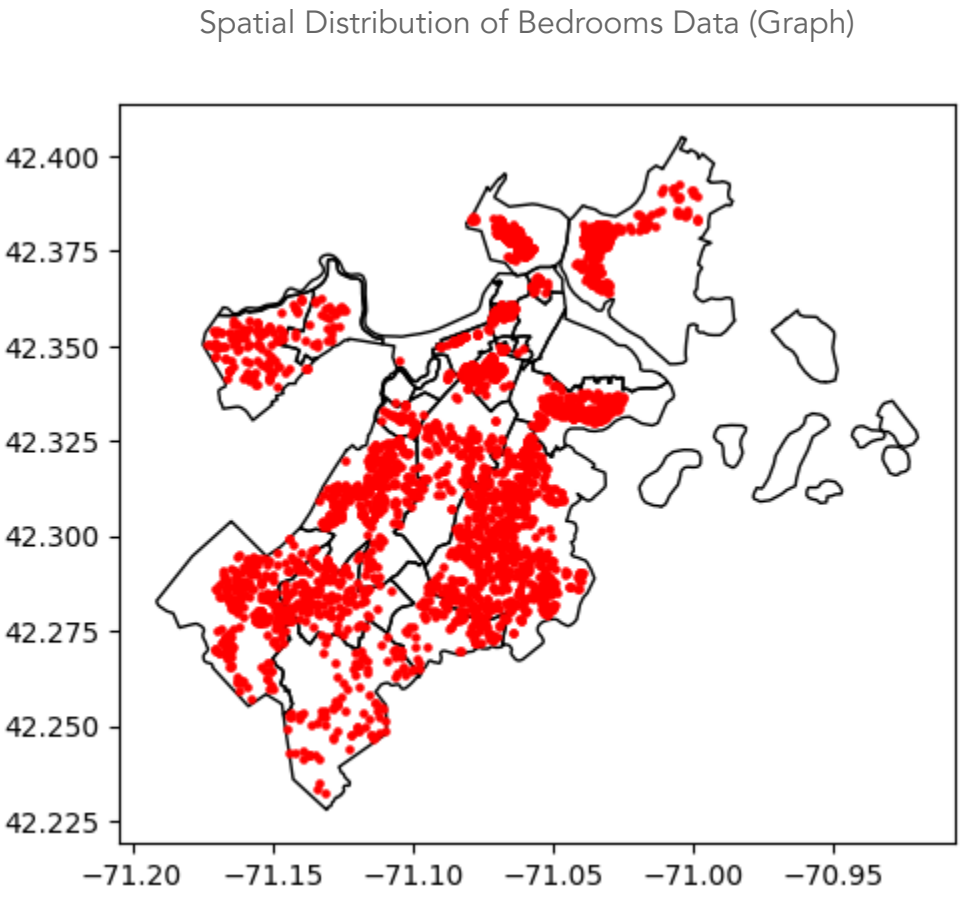
Bedroom Data Before and After Merging with SAM Data (Table)

Bedrooms Data	Before Merging	After Merging
Properties that had increases in bedrooms after renovation	956	617

Properties that had decreases in bedrooms after renovation	409	244
Properties that had no change in bedrooms after renovation	4402	2215

Heatmap Analysis on Bedrooms

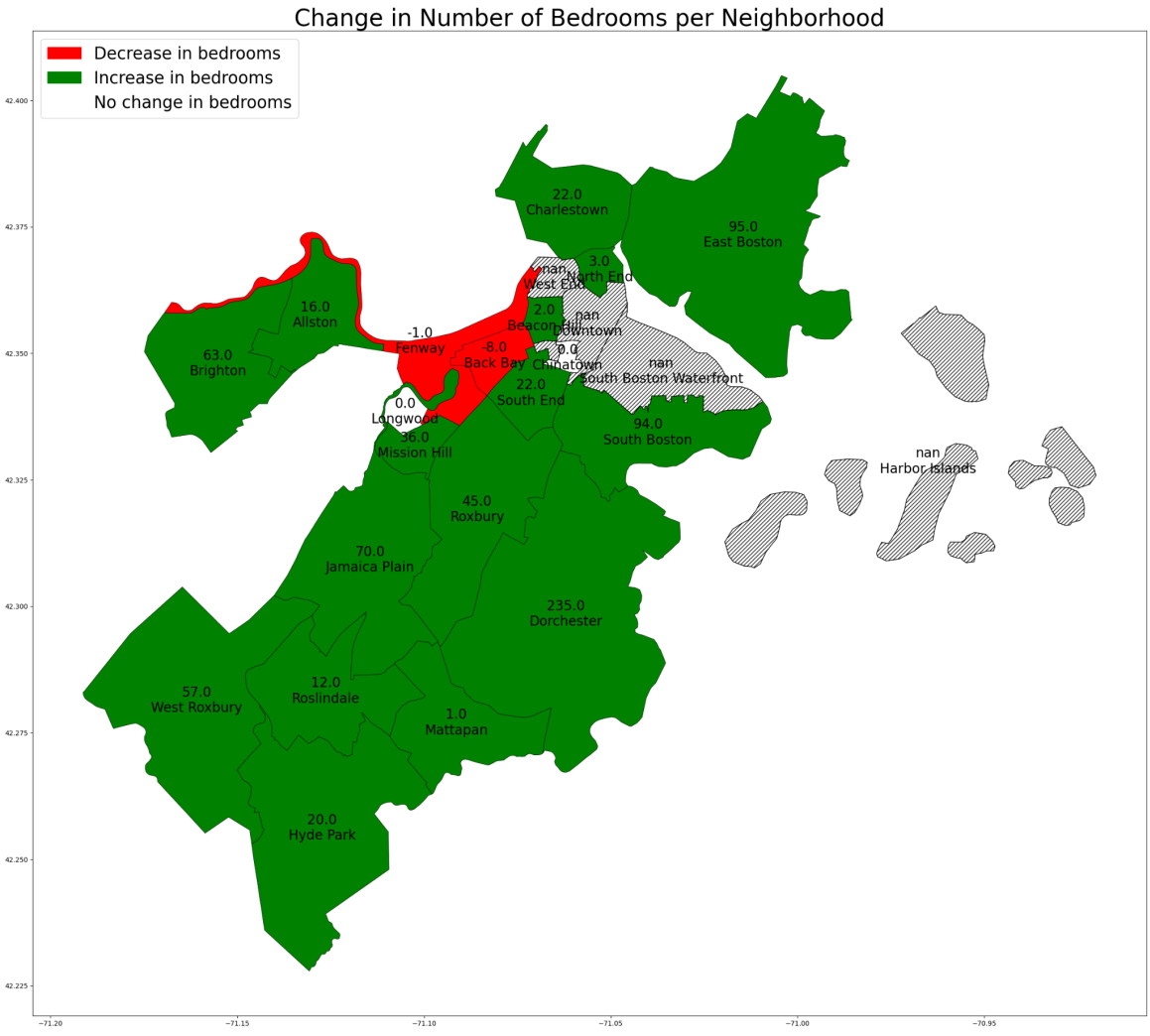
After merging the datasets and aggregating the changes in bedrooms and residential units by neighborhood, we created heatmaps to visualize the spatial distribution of these changes.



Upon plotting the data seems quite well-dispersed, except for East Boston and the harbor islands, for which we don't have data or which were lost during the merge.

The rest of the steps were relatively simple, with just some function calls. They're fully detailed in the notebook associated with this analysis in Updated_Heatmap_Renovations&Bedrooms.ipynb. The original analysis was in Heatmap_Renovations&Bedrooms.ipynb, but the neighborhood boundaries in that file are from 2010; the updated one uses data from 2020, so it's more up-to-date. Also, the 2010 data is inaccessible now for some reason, so only the updated one is runnable. Once we aggregated the differences in bedrooms, we got this map.

Change in the Number of Bedrooms per Neighborhood (Graph)

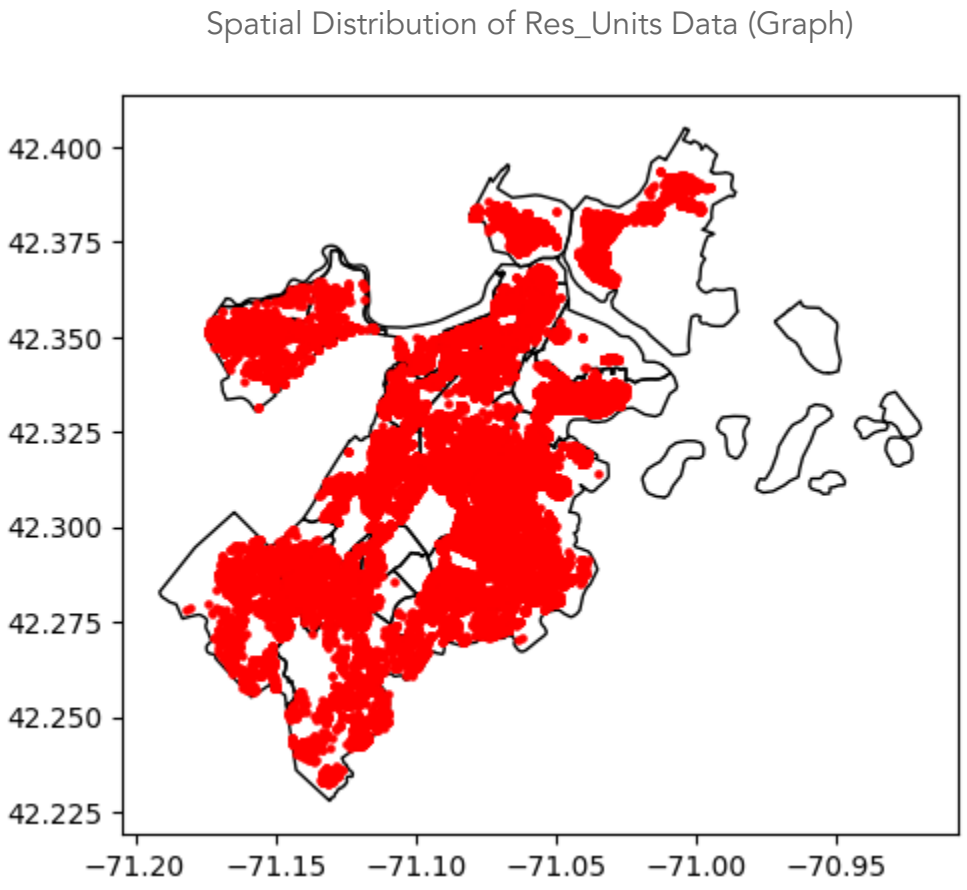


The bedrooms heatmap reveals that most neighborhoods experienced an increase in the number of bedrooms after renovations, with Dorchester showing the most significant increase of 235 bedrooms. Conversely, Fenway and Backbay saw slight decreases, while Chinatown and Bay Village remained unchanged.

Data Preparation and Challenges Res_Units

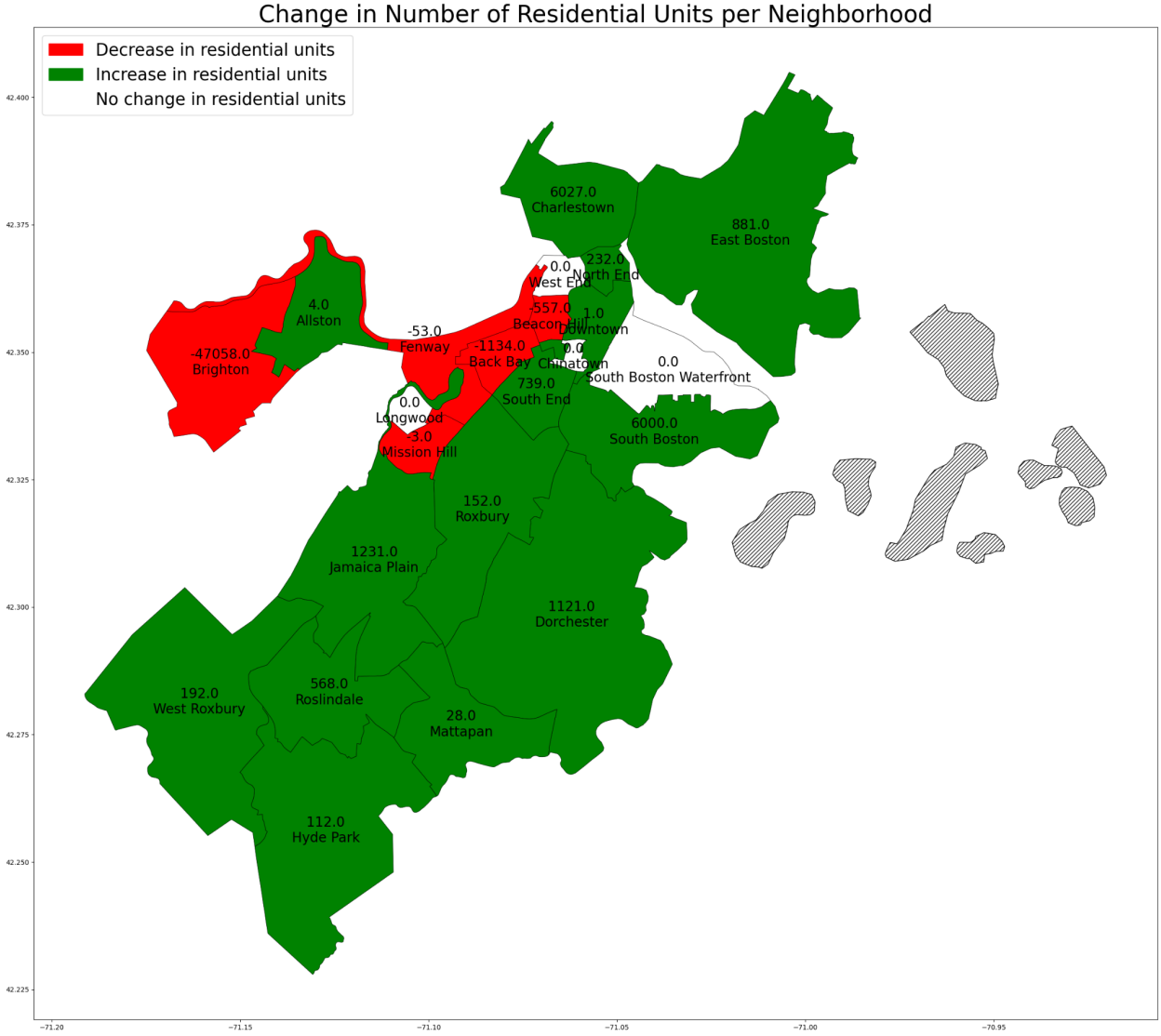
We followed the same process for res_units. However, in the beginning, we had to create res_unit datasets that contained changes in res_units before and after renovations. The process is detailed in the Heatmap_Renovations&ResidentialUnits.ipynb notebook.

Heatmap Analysis on Bedrooms



Similar to the bedroom data, it seems pretty well-dispersed, except for East Boston and the harbor islands, on which we don't have data or that were lost during the merge, but given that we don't have points for either of them, it seems like that we just don't have data on those parts of Boston.

Change in the Number of Res_Units per Neighborhood (Graph)



Using the same method we used on the bedroom data, we visualized the neighborhood changes based on residential units. This map shows a notable decrease in the number of units lost in Brighton compared to other neighborhood losses, with almost 47,000 units lost in the past

20 or so years, which is quite shocking. We can also see that a significant number of neighborhoods, such as Allston, Fenway, North End, and Mission Hill, have remained the same. However, this might be explained due to a lack of data on the residential units in these neighborhoods. The other neighborhoods that increased in number did not increase significantly except for Charlestown, which increased by 5586 units.

Demographic Analysis

To understand the demographics of the neighborhoods affected by these changes, we analyzed the age, race, and household composition data from the census. Using similar mapping techniques that we used earlier, we can map each neighborhood with its majority Demographics.

The code wasn't too complex To start, we had to separate the census data into age, race, and household tables. Then, we added unique column names for simpler data parsing that I could parse with REGEX. Then, after cleaning some data, all the percent columns were stored as objects, so I had to convert them to floats. Then, we were able to get the Modal age/race/household type by neighborhood by using idxmax along the columns of the age/race/household data to grab the column with the highest value and store that value in modal_x.

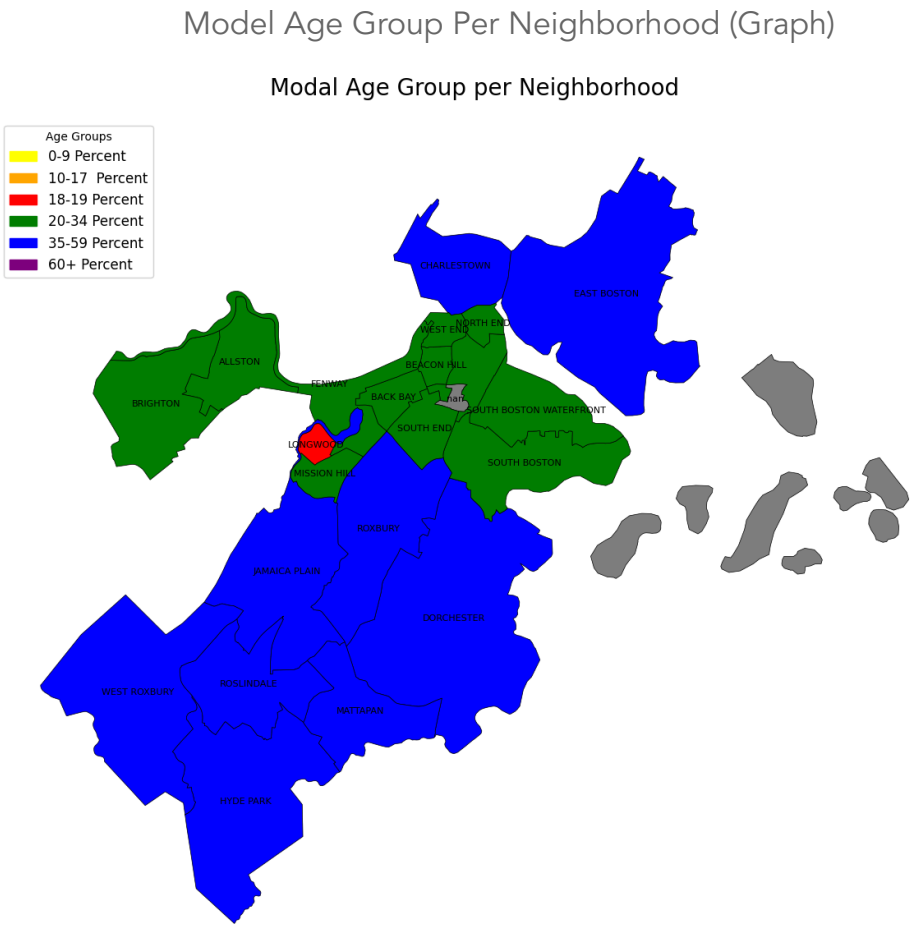
Snippet of Merged Demographic Data (Dataset)

	OBJECTID	BlockGr202	Shape_Leng	Shape_Area	geometry	point_count	bedrooms_diff	color	Neighborhood	Modal_Age_Group
0	1	ALLSTON	35808.619278	4.154760e+07	POLYGON ((-71.12123 42.36775, -71.12069 42.367...	30	16.0	green	ALLSTON	20-34 Percent
1	2	BACK BAY	18815.103609	1.538724e+07	POLYGON ((-71.07315 42.35554, -71.07302 42.355...	49	-8.0	red	BACK BAY	20-34 Percent
2	3	BEACON HILL	11668.951169	7.891524e+06	POLYGON ((-71.06291 42.36123, -71.06286 42.360...	138	2.0	green	BEACON HILL	20-34 Percent
3	4	BRIGHTON	47051.804654	7.658156e+07	POLYGON ((-71.13737 42.35876, -71.13747 42.358...	127	63.0	green	BRIGHTON	20-34 Percent
4	5	CHARLESTOWN	33910.754786	5.127021e+07	POLYGON ((-71.06700 42.39401, -71.06741 42.393...	221	22.0	green	CHARLESTOWN	35-59 Percent
5	6	CHINATOWN	10843.828683	3.436019e+06	POLYGON ((-71.05801 42.35235, -71.05817 42.352...	2	0.0	white	NaN	NaN
6	7	DORCHESTER	80692.139164	2.193038e+08	POLYGON ((-71.05733 42.32804, -71.05720 42.328...	679	235.0	green	DORCHESTER	35-59 Percent

With this data, we were able to plot all the demographic data. First, we plotted the modal class for each demographic on the neighborhood map and then pie charts for the breakup of each demographic in each neighborhood. This allowed us to make inferences about the population in each neighborhood that was being affected by renovations.

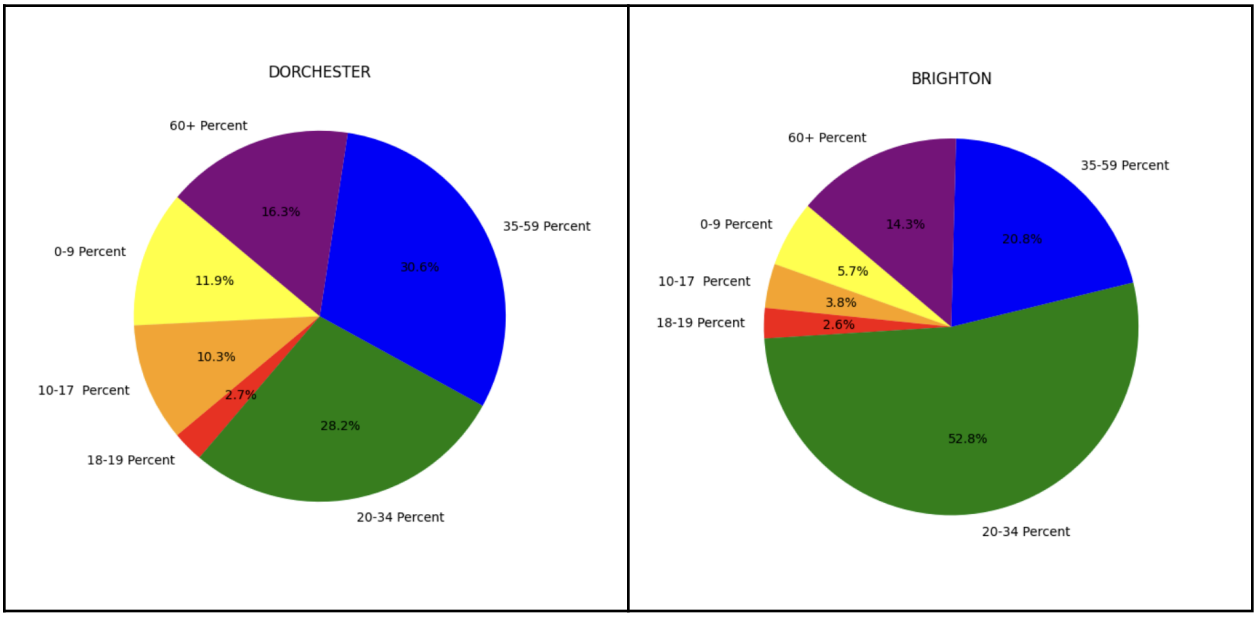
Age Demographics

Model Age Group Per Neighborhood (Graph)



The age demographic analysis shows that the majority of the population in the affected neighborhoods falls within the 20-34 and 35-59 age ranges. Brighton, which experienced the greatest loss in residential units, has a predominantly young population (20-34), while Dorchester, which gained the most units, has a larger middle-aged population (35-59).

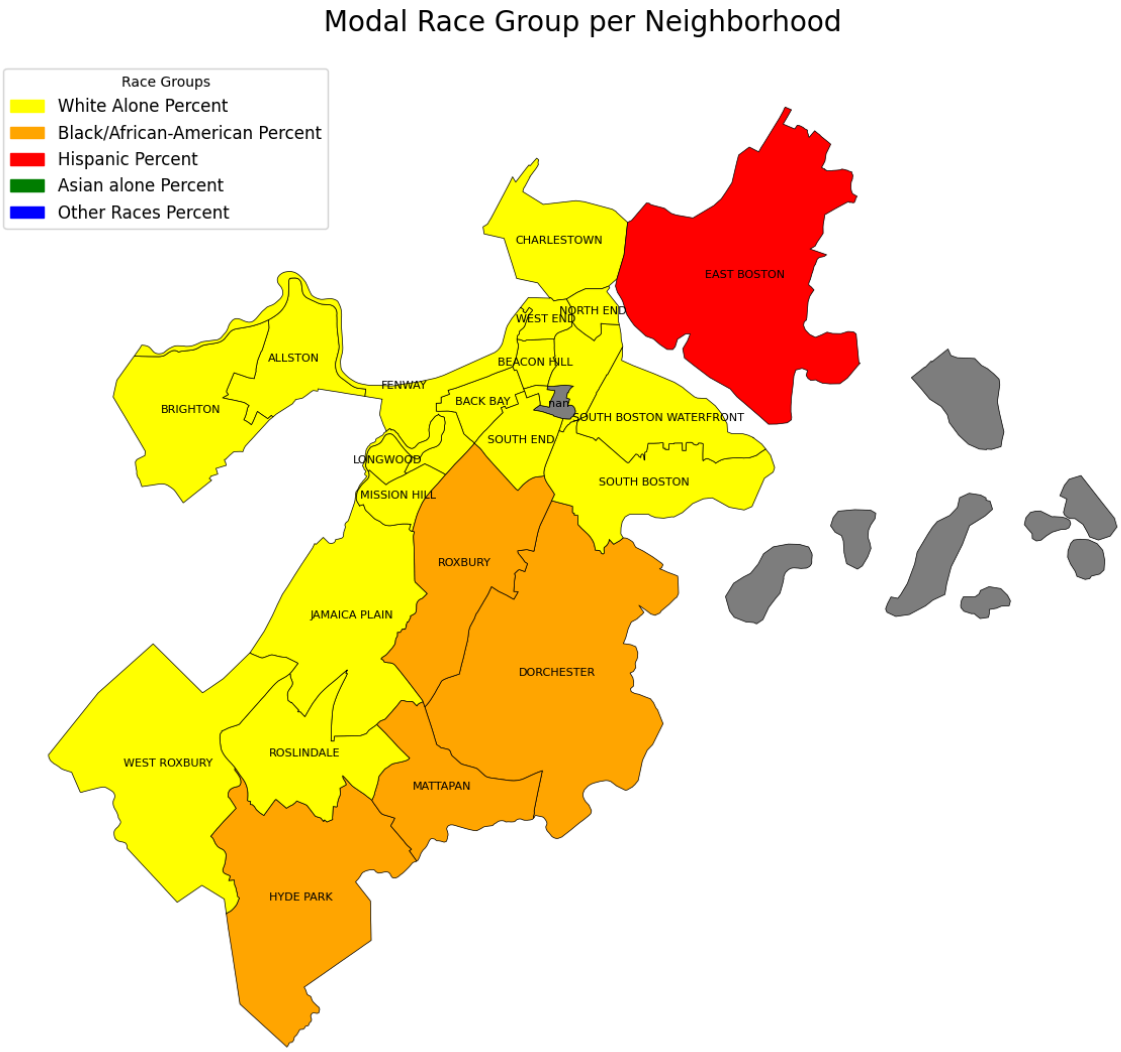
Age Groups in Brighton and Dorchester (Graph)



From the pie charts, we can see that in Brighton, the majority of those affected are in the 20-59 age range. In Dorchester, however, there seems to be a mix of ages affected. All the pie charts for each neighborhood can be found here.

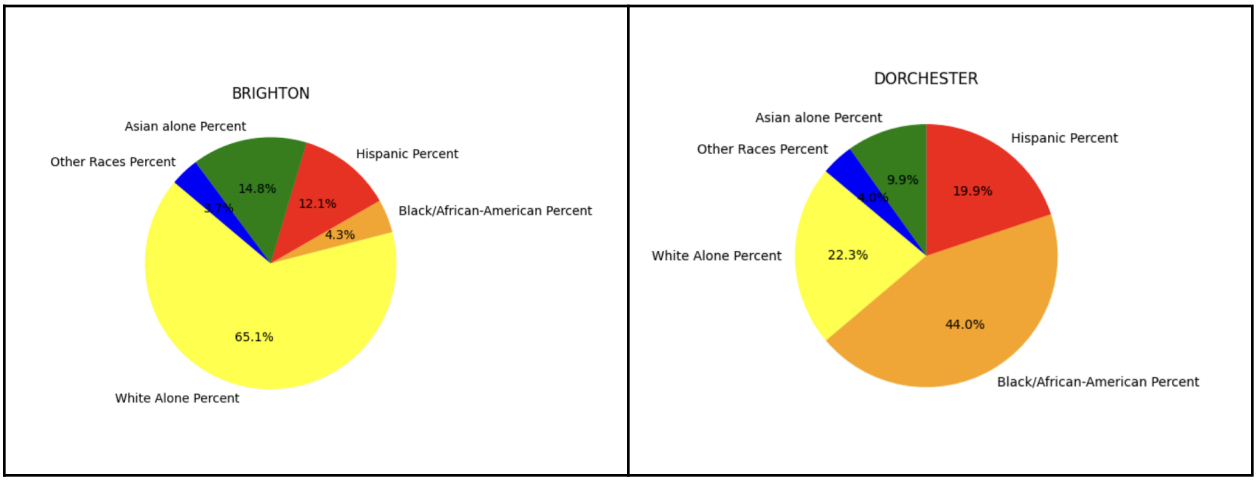
Race Demographics

Model Race Group Per Neighborhood (Graph)



The racial demographic analysis reveals that the majority of the population in the affected neighborhoods is either White Alone or Black/African-American. Brighton, which lost the most units, is predominantly White Alone, while Dorchester, which gained the most units, has a larger Black/African-American population.

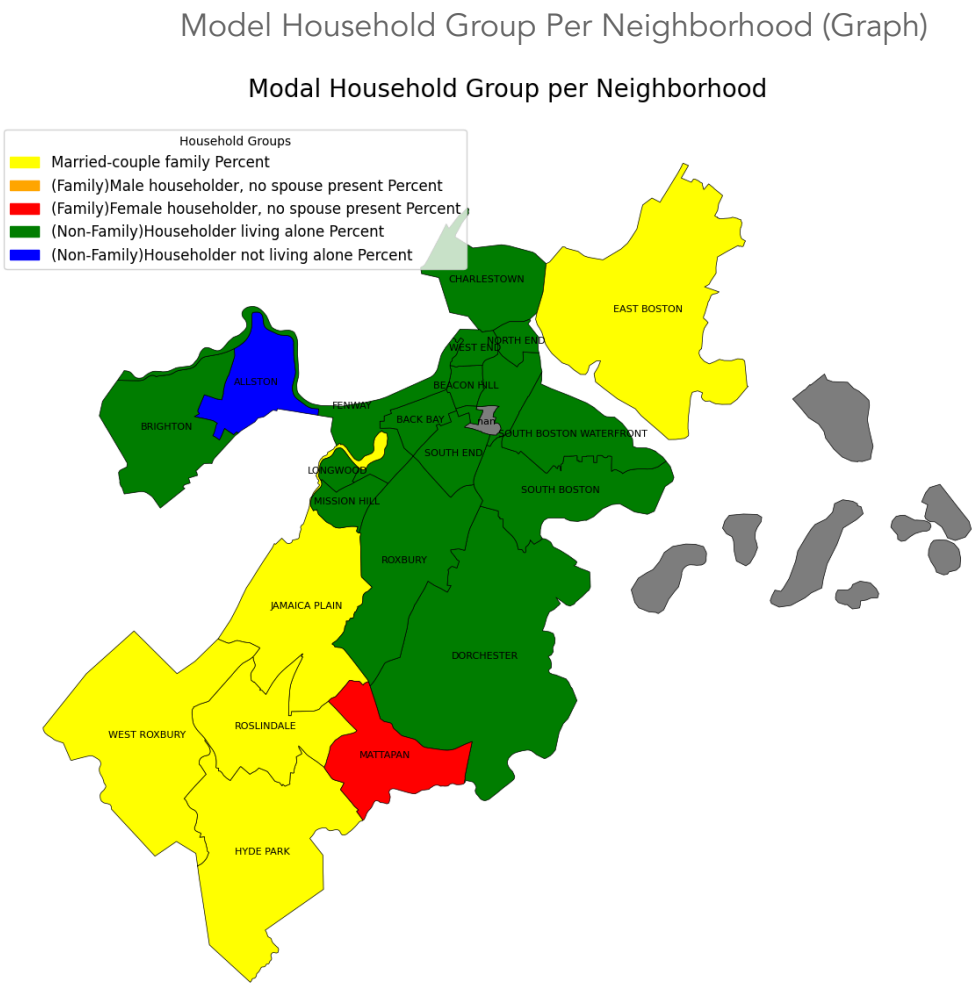
Race Groups in Brighton and Dorchester (Graph)



In Brighton, Whites alone seem to be the majority race, but there's a bit more of a mix in Dorchester, with Whites Alone and Hispanics making up around 20 percent each. All the pie charts for each neighborhood can be found here.

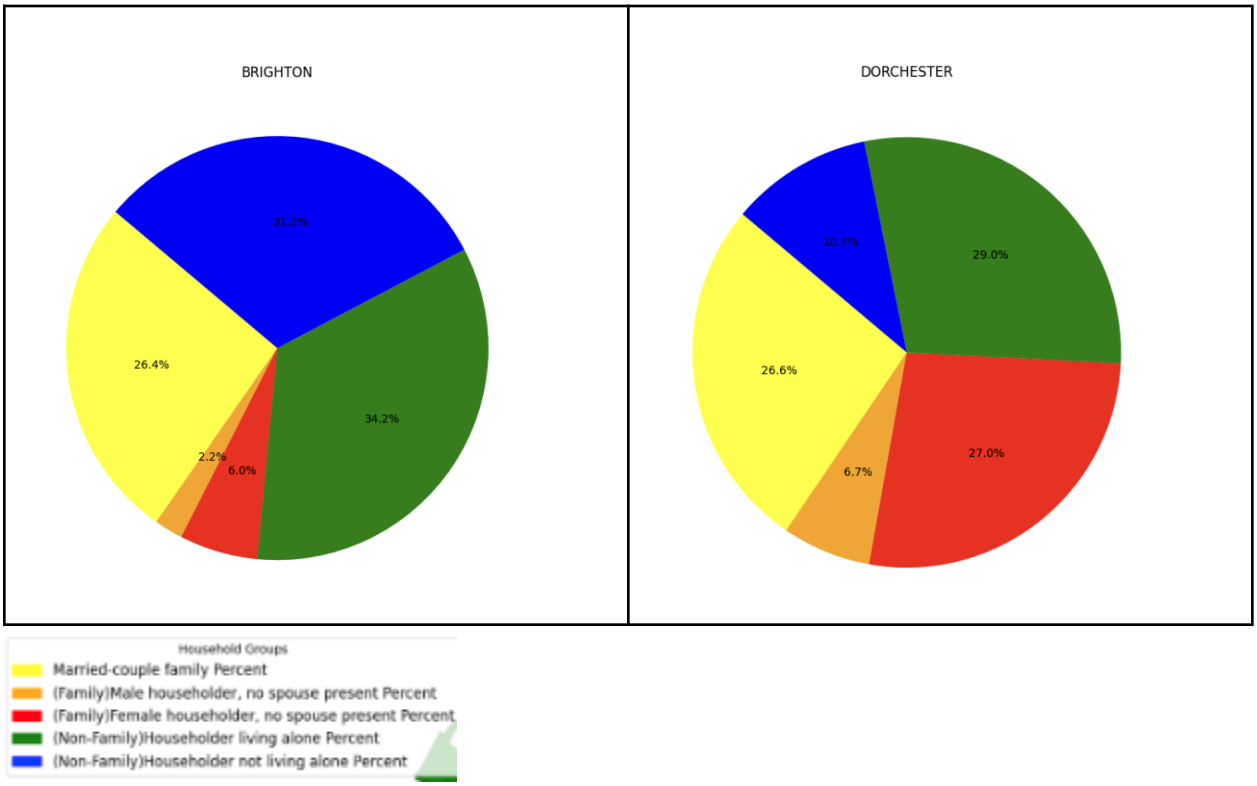
Household Demographics

Model Household Group Per Neighborhood (Graph)



The household composition analysis indicates that the majority of households in the affected neighborhoods are either Married-couple families or Householders living alone. In Brighton, there is a mix of non-family roommate housing and couple homes, while in Dorchester, there are gains in both living-alone units and shared units.

Household Groups in Brighton and Dorchester (Graph)



From these pie charts, we can tell that people living alone just barely exceed other categories. In Brighton, there is a lot of non-family roommate housing and couple homes. And in Dorchester, there are gains in living-alone units and sharing ones. All the pie charts for each neighborhood can be found here.

Conclusion

Our heatmap and demographic analysis provide insights into the neighborhoods and populations that may be affected by changes in bedrooms and residential units due to

renovations. Brighton, with its predominantly young, White population, experienced the greatest loss of units, while Dorchester, with its middle-aged, Black/African-American population, saw the most significant gains. However, it is important to note that due to data limitations, we cannot definitively conclude that these specific demographic groups are directly benefiting from or being disadvantaged by the changes in the housing market. Further research with more comprehensive data is needed to fully understand the impact of these changes on different demographic groups within each neighborhood. Despite these limitations, our analysis highlights the importance of considering demographic factors when assessing the effects of housing market changes and renovations on communities in Boston.

Building Permits and Sentiment Analysis of Construction Activities

This analysis utilizes two key datasets: the APPROVED BUILDING PERMITS data and the LIVE STREET ADDRESS MANAGEMENT (SAM) ADDRESSES data. The primary goal is to investigate the relationship between property locations and their corresponding geographic coordinates (x and y values). By moving away from an analysis based solely on zip codes and instead leveraging precise coordinate information, we can create detailed heatmaps that provide a more granular visualization of development density patterns. This coordinate-based approach allows for a deeper understanding of the spatial distribution of construction and development activities within the study area.

Base Analysis

Total Fees and Declared Valuation Per Property Stats (Table)

Total Fees	
Count:	555,815
Mean:	$ 916
Std:	26,022
Min:	$ 0
25%:	32
50%:	70
75%:	190
Max:	$ 13,080,420

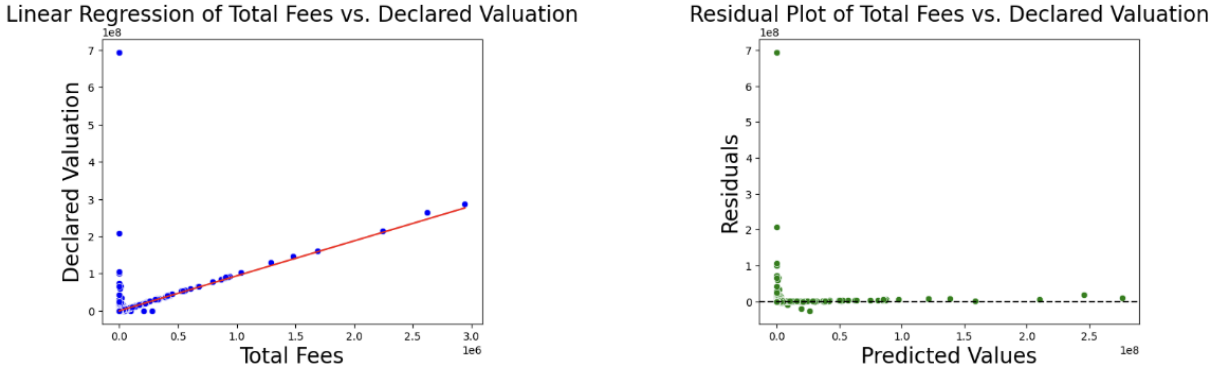
Declared Valuation	
Count:	555,815
Mean:	$ 116,650
Std:	4,040,626
Min:	$ -1,000,000
25%:	1,500
50%:	5,500
75%:	20,000
Max:	$ 2,100,000,000

The initial analysis focused on quantifying the construction activity within the neighborhood based on the recorded data. The dataset contains 555,815 permits, indicating a

high level of development, renovation, and infrastructure projects across Boston. To assess the financial aspect, the wide range of permit fees (from $0 to over $13 million) and declared valuations (from -$1 million to $2.1 billion) was examined. This diversity suggests that the projects range from minor repairs or modifications to major developments. The mean values ($916 for fees and $116,650 for declared valuations) provide an estimate of the average cost and scale, while the substantial standard deviations highlight the significant variability among projects.

Furthermore, the dataset sheds light on Boston's regulatory environment for housing and construction. The city's framework appears to be aimed at ensuring safety, compliance with building codes, and orderly development. The variety of permits issued, including short and long form building permits, as well as electrical, plumbing, and gas permits, reflects the comprehensive nature of this regulatory oversight.

Linear Regression and Residual Plot (Total Fees vs. Declared Valuation) (Graph)



The RMSE value of 2305872.84 indicates that, on average, the model's predictions deviate from the actual values by about $190,658.13.

After understanding the nature of construction activities and permits, the analysis examined the relationship between total fees and declared valuation to determine if higher fees correlate with higher declared valuations. The scatter plot revealed a positive correlation

between these two variables, suggesting that as total fees increase, declared valuations tend to increase as well. The linear regression line captures this trend, indicating a linear relationship.

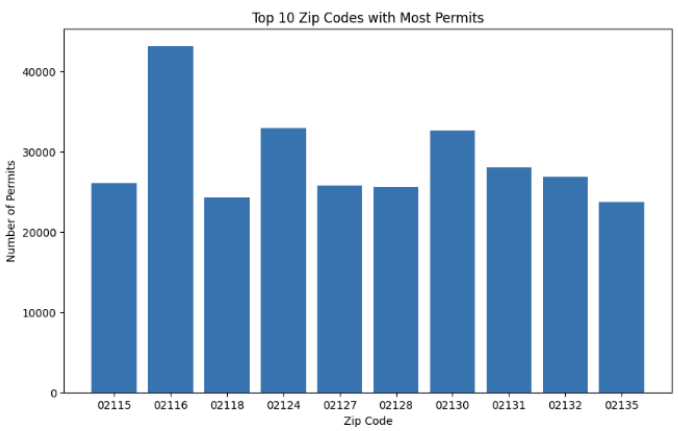
The positive slope of the regression line, with a coefficient of 93.86, confirms the positive relationship between total fees and declared valuation. This means that as the declared valuation of building permits rises, the associated total fees also tend to increase. The y-intercept of 30,216.71 suggests that even when the declared valuation is zero, there is a base amount of fees on average, possibly representing a minimum fee or baseline cost for obtaining a permit, regardless of project valuation.

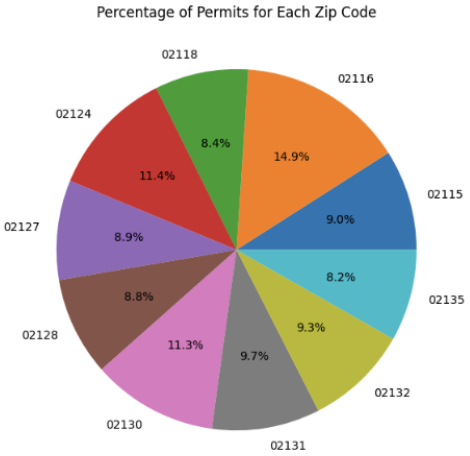
However, it is important to note that the Root Mean Square Error (RMSE) of 2,305,872.84 is quite high, indicating a considerable spread of the actual values around the predicted values. This suggests that while a positive relationship exists, total fees cannot be precisely predicted based solely on declared valuation due to other influencing factors not captured in this simple linear model.

The residual plot identified outliers, representing large residuals or differences between the observed and predicted values. These outliers could be indicative of particularly large projects, errors in data reporting, or instances where the fee structure deviates significantly from the norm due to special circumstances.

Understanding this relationship between fees and declared valuation can assist Boston's policymakers and city planners in setting fair and appropriate fee structures that align with the economic impact of construction projects.

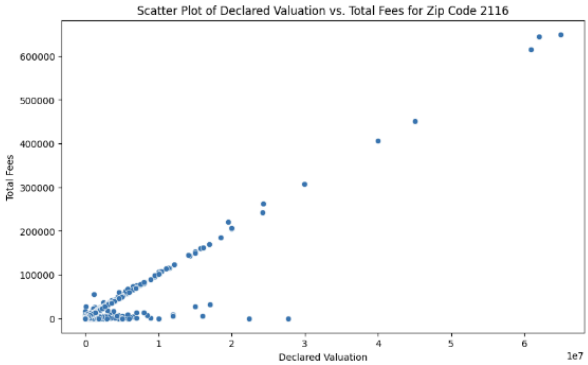
Top 10 Zip Codes with the Most Permits (Graph)

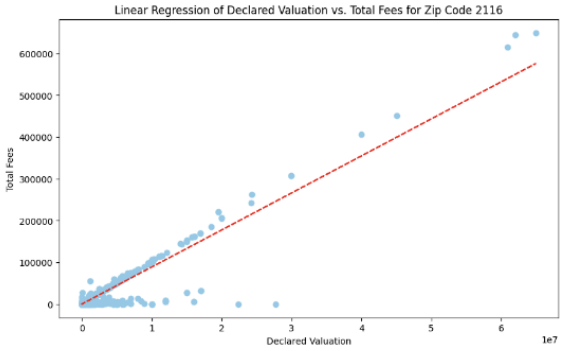




These are the top 10 zip codes from 75 entries

Declared Valuation vs. Total Fees for Zip Code 02216 (Graph)





Correlation coefficient between declared valuation and total fees for zip code 2116: 0.9323180081064679

The analysis then focused on identifying the zip code with the highest number of permits, with the intention of determining the areas with the most development activity. The zip code 02116 was found to have the highest concentration of permits. Specifically, for this zip code, the correlation coefficient of 0.9323 indicates a very strong positive linear relationship between

declared valuation and total fees. As the declared valuation increases, the total fees tend to increase in a closely related manner.

The correlation coefficient close to 1 suggests that the linear model provides a good fit for the data in the 02216 zip code. This implies that the total fees can be predicted with a fair degree of accuracy based on the declared valuation. The trend in the data points and the slope of the regression line suggest that projects with higher valuations in this particular zip code are associated with significantly higher permit fees.

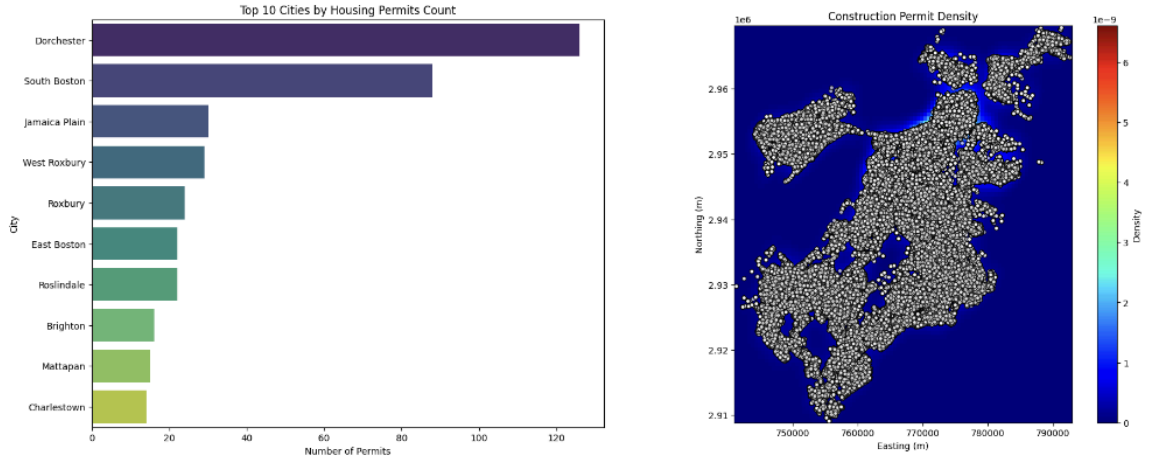
However, it is noteworthy that most of the data points are clustered at the lower end of declared valuations, while as the valuation increases, the total fees also rise, and the spread becomes wider. This pattern might indicate that for the majority of projects, the fees are relatively modest, but for a few high-value projects, the fees can be quite substantial.

Although the overall relationship is strong, there are still some outliers or deviations from the trend line. These outliers could represent exceptional cases or specific types of permits that have different fee structures. It is important to note that the strong correlation found in the 02116 zip code may not be applicable to other zip codes, as the relationship between fees and valuations could vary across different areas.

For stakeholders, such as developers and builders in the 02116 zip code, understanding this relationship can assist in budgeting and forecasting project costs more accurately. For policymakers, this data provides insights into how building permit fees correlate with the scale of investment in construction within a specific area of the city, enabling more informed decision-making.

Top 10 Cities by Housing Permit Coust

Top 10 Cities by Housing Permit Counts (graph)

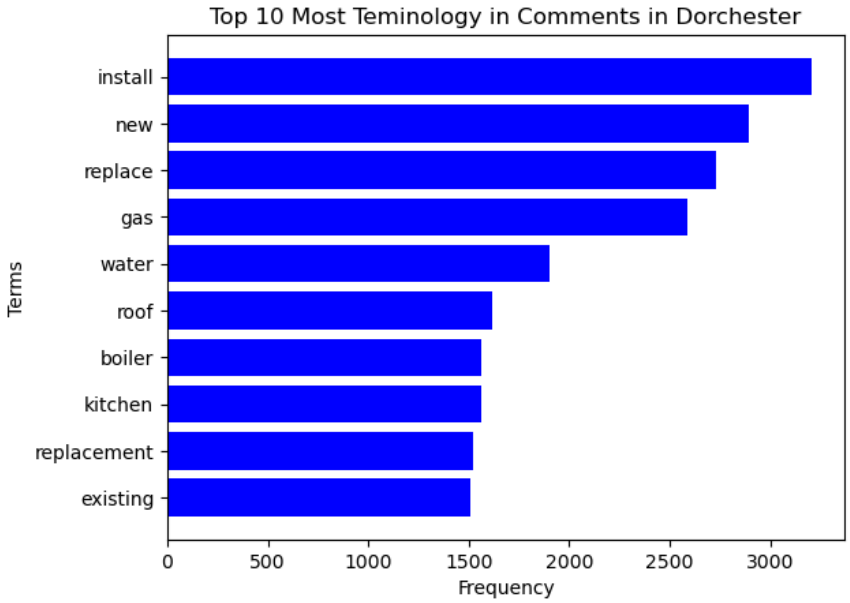


Dorchester being the highest area with highest housing permits followed by South boston. By taking the Y_COORD and X_COORD we can see which area has the highest density.

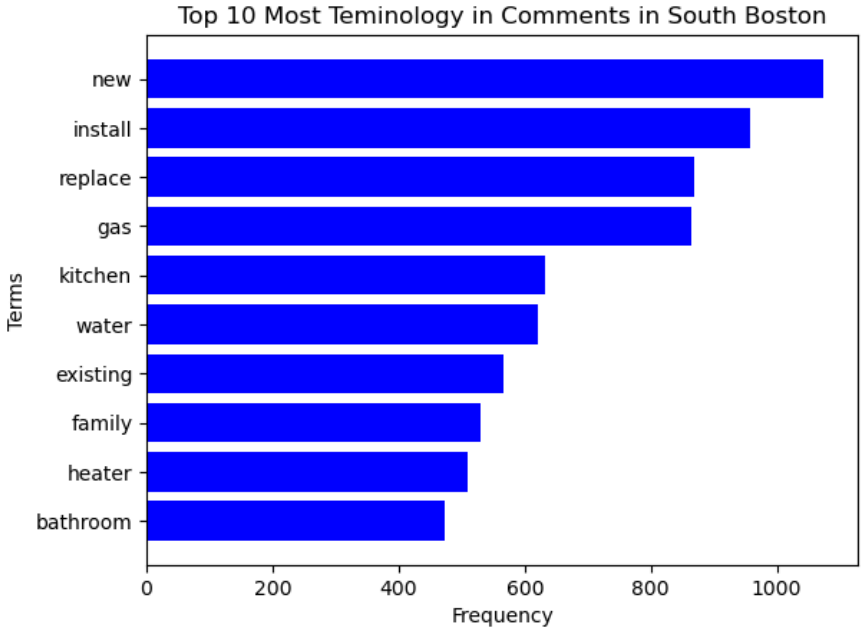
The analysis then shifted its approach to examine specific zip code locations within the area, with the purpose of obtaining more detailed information on the spatial distribution of development. By utilizing the geographic coordinates (gpsx and gpsy parameters) from the Building_permits.csv file and the corresponding X_COORD and Y_COORD values from the Live_Street_Address_Management_(SAM)_Addresses.csv file, the data was plotted on a heatmap. This visualization revealed that the neighborhoods of Dorchester and South Boston had the highest concentration of development activities, far outweighing other areas. This finding aligns with the earlier analysis, which indicated that Dorchester had the highest unit gains.

Following this spatial analysis, a qualitative assessment was performed by analyzing the words used in the comments provided by clients when issuing permits. This word analysis aimed to understand the underlying reasons and motivations behind the issuance of permits in these high-development areas.

Most Common Words Found in Comments in Dorchester (Graph)



Most Common Words Found in Comments in South Boston (Graph)

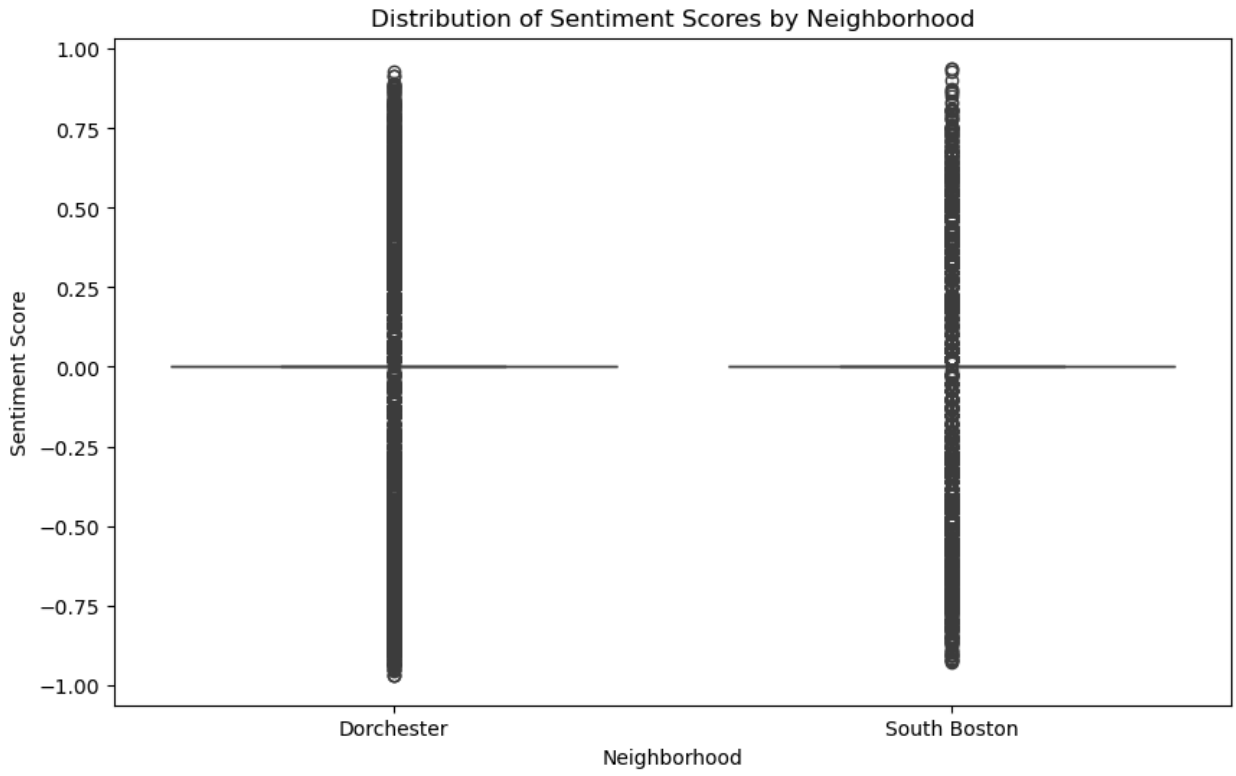


By conducting a deeper analysis of the word frequencies in the comments related to development trends in Dorchester and South Boston, several prominent terms emerge. The prevalence of words such as "install," "remove," "replace," "water," "heater," "basement," "unit,"

and "new" provides a general impression of the types of construction activities occurring in these neighborhoods.

The specific comments from the Boston Housing Department data offer concrete examples of these activities, including amendments for basement finishing, bathroom additions, and the installation of fire alarm systems. The terminology used indicates a mix of both renovation projects (evident from words like "reframing," "repairs," "existing," "renovation") and new construction ("newly constructed," "add," "new roof deck"). This combination of renovation and new development suggests an environment characterized by growth and renewal, with a strong emphasis on modernizing and potentially expanding existing properties.

Distribution of Sentiment Scores by Neighborhood (Graph)



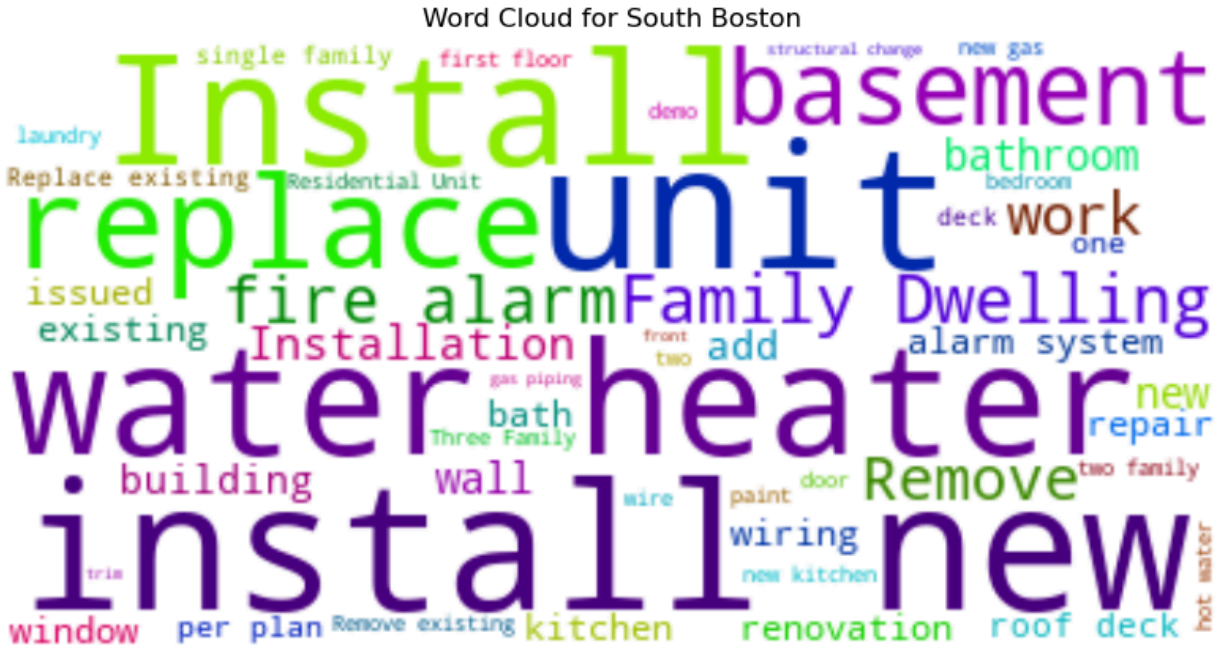
The sentiment scores attached to each comment provide a quantitative measure of the emotional tone conveyed in the text. While sentiment may not always directly correlate with the nature or outcome of the development project, it offers an indirect indicator of how these

projects are perceived or described in the comments – whether they are viewed positively or as neutral descriptions of the work to be undertaken.

To further enrich the analysis, the Term Frequency-Inverse Document Frequency (TF-IDF) technique can be employed. TF-IDF allows for a comparison between the word usage in the comments and the broader discourse within the Boston Housing Department. This approach would enable the identification of words that are distinctive or prevalent in the comments related to construction in Dorchester and South Boston, as compared to the general housing and construction discussions within the department. Words that are common in the comments but less frequent in the wider departmental documents would have higher TF-IDF scores, signaling their importance in the specific context of development in these neighborhoods.

Word Cloud for Dorchester (Graph)

Word Cloud for South Boston (Graph)



The word clouds and comments provide a foundational understanding of the landscape of construction activity within the study area.

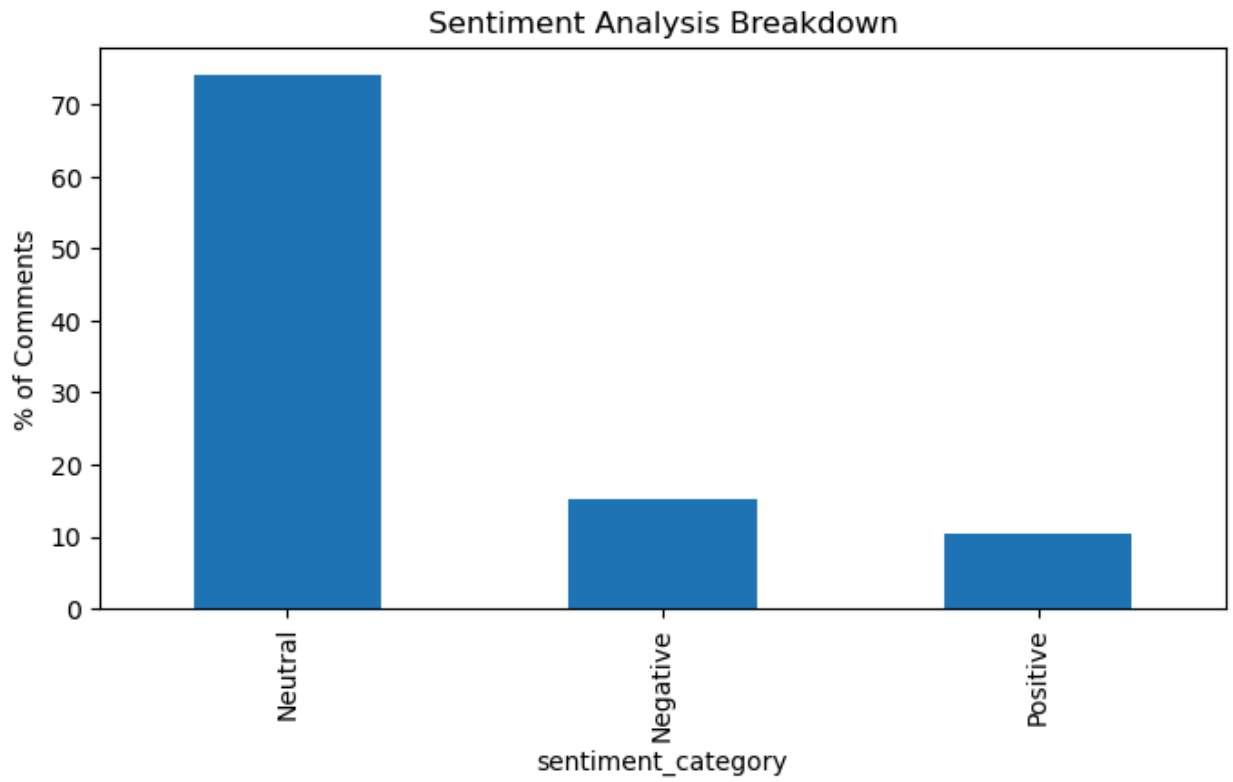
Permit data, such as that from the APPROVED BUILDING PERMITS and LIVE STREET ADDRESS MANAGEMENT (SAM) databases, offers an empirical basis for analyzing regulatory compliance. These datasets are rich in details about the nature of construction activities and their geographic distribution, enabling cross-referencing with municipal codes and standards to assess compliance.

An examination of the comments and permit details suggests that the majority of construction activities involve improvements to existing structures, as evidenced by frequent mentions of terms like "amending," "reframing," "renovation," and "installation." This implies that property owners and developers are not only investing in the physical enhancement of their properties but are also navigating the city's regulatory framework, which likely necessitates seeking amendments for changes to previously approved plans.

Terms such as "replace" and "install," associated with substantial components like "water heater," "fire alarm system," and "windows," suggest that many of these projects are essential upgrades to meet current building codes, which may have evolved since the original construction of the property. By focusing on these key terms, it can be inferred that the work being carried out is not merely cosmetic but aligns with the city's safety, energy, and building standards.

Moreover, the attention given to "basement" projects indicates a trend in utilizing below-grade spaces, which typically face strict regulatory requirements due to concerns such as egress, moisture control, and fire safety. The fact that permits are being issued for such projects strongly indicates that these construction activities are meeting the regulatory standards set forth for basement conversions.

Sentiment Analysis Breakdown (Graph)



Considering the positive and negative sentiments expressed in the comments could provide further insights. A deeper analysis of the language used to describe these activities would be beneficial. For instance, negative comments might highlight challenges or issues faced during construction, such as phrases like "no work to be done" or "no structural changes," which could suggest bureaucratic hurdles or restrictions impeding development. Conversely, positive comments mentioning installations and replacements could reflect satisfaction with improvements or upgrades, potentially indicating investment and development within the neighborhood.

Employing the TF-IDF (Term Frequency-Inverse Document Frequency) analysis would not only enrich our understanding of the types of construction activities but also provide valuable insights into how these activities are integrated into the urban fabric and the regulatory landscape. Such an understanding could prove instrumental for city planners, policymakers, and stakeholders in shaping the future development of these areas. For example, if the term "fire alarm system" has a high TF-IDF score, it may indicate a heightened focus on safety regulations in recent developments. Similarly, if "water heater" is a prominent term, it could signify a wave of upgrades to more energy-efficient systems in response to environmental concerns or policy incentives.

Overall, this approach allows us to move beyond a surface-level interpretation of the data and engage in a nuanced analysis of the socio-economic dynamics and regulatory environment that influence development patterns in Dorchester and South Boston. By combining quantitative data analysis, qualitative text analysis, and spatial visualization techniques, we can gain a deeper understanding of the interplay between construction activities, community sentiments, regulatory frameworks, and broader societal factors shaping the urban landscape.

Conclusion

Our comprehensive analysis of Boston's housing market has provided valuable insights into the complex dynamics at play, revealing trends and patterns that contribute to the changing landscape of housing availability and affordability in the city. By examining property assessment data, building permits, and demographic information, we have uncovered key findings that can inform policymakers and stakeholders in their efforts to develop equitable and sustainable housing solutions.

The appreciation of housing prices and its correlation with remodeling and unit loss have been significant concerns. Our analysis has shown that unit loss is positively correlated with price appreciation, highlighting the need for the City of Boston to monitor and address this issue to prevent further financial strain on its residents.

While the city has experienced an overall net gain in residential properties over the past 20 years, certain neighborhoods have been disproportionately affected. Brighton, with its predominantly young, white population, has experienced the greatest loss of units, while Dorchester, with its middle-aged, Black/African-American population, has seen the most significant gains. However, it is important to note, we do not know if those specific demographics are being affected by gains/losses. These findings emphasize the importance of implementing targeted policies and initiatives that address the unique challenges and opportunities present in each community.

Our investigation into the impact of renovations on bedrooms and living areas has revealed that, on average, renovations do not result in a loss of these features. However, the weak positive correlation between changes in living area and changes in the number of bedrooms suggests that these two variables are largely independent of each other during renovations. Furthermore, the linear regression analysis did not yield a strong predictive model

for bedroom and living area differences, indicating that other factors not included in the model may be influencing these changes.

The building permit analysis showed a high number of permits (555,815) and a wide range of permit fees and declared valuations, reflecting the diversity of projects from minor repairs to major developments. The zip code 02116 was found to have the highest concentration of permits, with a strong positive linear relationship between declared valuation and total fees. The word frequency analysis of comments related to development trends in Dorchester and South Boston revealed a mix of renovation and new construction activities, with a focus on modernizing and expanding existing properties.

In conclusion, this report offers a comprehensive examination of the factors influencing Boston's housing market, with a particular focus on the impact of remodeling and zoning conversions on housing availability and affordability. By understanding the trends, patterns, and demographic characteristics of affected communities, policymakers and stakeholders can make informed decisions that promote equitable and sustainable housing solutions for all of Boston's residents. The insights gained from this study serve as a foundation for future research and policy development, emphasizing the need for a data-driven approach to addressing the complex challenges faced by Boston's housing market.

Potential next steps

Appreciation of housing prices and its correlation with remodeling and unit loss

- Conduct a deeper exploration into the neighborhoods with the highest unit loss to identify specific factors driving this trend, such as demographic changes, development patterns, or regulatory issues.
- Analyze additional datasets, such as demographic data or housing supply data, to gain a more comprehensive understanding of the factors influencing housing price appreciation and unit loss.

Loss and gain of residential properties across different neighborhoods

- Investigate the reasons behind the observed net loss of residential properties in certain neighborhoods, particularly focusing on South Boston, East Boston, and Hyde Park, to identify underlying factors contributing to this trend.
- Examine the relationship between property loss and neighborhood characteristics, such as gentrification, urban development projects, or changes in zoning regulations, to understand the broader context of property dynamics.

Insights on Bedroom and Living Area Loss/Gain

- Explore alternative modeling techniques, such as classification trees, Naive Bayes, K-Nearest Neighbors, or Support Vector Machines, to predict bedroom and living area differences during renovations, considering the lack of clear linear relationships observed in the data.
- Investigate the impact of external factors, such as the COVID-19 pandemic, on housing trends and renovations to better understand the drivers behind the observed changes in bedroom and living area.

Heatmaps and demographic analysis of affected areas

- Conduct a deeper dive into neighborhoods like East Boston and the harbor islands to understand their specific trends and impacts, possibly by acquiring additional or more detailed data for these areas.
- Translate the findings of the analysis into actionable recommendations for policymakers, urban planners, and community leaders to inform decision-making and interventions aimed at addressing disparities and mitigating adverse impacts of housing market changes on vulnerable populations.

Building Permits and Sentiment Analysis of Construction Activities

- Conduct a temporal analysis to examine how construction activity and development density patterns have evolved over time, identifying trends, seasonality effects, or significant fluctuations in permit issuance rates.
- Translate the findings of the analysis into actionable policy recommendations for optimizing fee structures, streamlining permitting processes, and promoting sustainable development practices that align with the economic and social goals of the community.

Contributions

The length of a section is not indicative of a persons contribution. Each persons specific work can be found in the juypter notebooks found in the Github. Each notebook has the authors name at the top. However, each person is responsible for the findings and interpretations found in each section of the report detailed below:

Anderson Lawrence - Appreciation of housing prices and its correlation with remodeling and unit loss

Ashley Harlow - Loss and gain of residential properties across different neighborhoods: Residential Property Loss and Gain

Rishab Sudhir (Team Lead) - Loss and gain of residential properties across different neighborhoods: Insights on Bedroom and Living Area Loss/Gain

Maha Alali - Heatmaps and demographic analysis of affected areas

Wilbert Limson - Building Permits and Sentiment Analysis of Construction Activities