

**Affordable Housing Snapshot
(Housing Affordability Index) -
Team 2**

Deliverable 1

1. Data Collection and Preprocessing

- A collection of datasets covering different aspects and questions of the project were provided in the project proposal.
- All the provided data that was publicly accessible was preprocessed for preliminary analysis.
- Essentially, the preprocessing steps were:
 - a. **Dealing with NaN (Missing) Values:** Columns which had over 50-60% missing values were dropped from the dataset. If a column was significantly important, then the missing rows were filled with the mean value (for continuously-valued columns) and with the highest occurring value (for discrete-valued columns). If the number of rows with NaN values were low (3-5%), then these rows were simply dropped.
 - b. **Data Normalization:** All the continuously real-valued columns were normalized using z-standardization (subtracting each value by the mean of the column and then dividing the result with the standard deviation of the column).
 - c. **Data (Attribute) Selection:** For each dataset, a few columns (such as ID) were not relevant to the information the data was representing. Hence, these columns were simply dropped from the database.
 - d. **Outlier Analysis:** A preliminary (basic) outlier analysis was also performed on the data. Given the size of each dataset, we might need to perform some more operations to further analyze the outliers.
- More preprocessing steps will be taken before further analysis of the data.
- Apart from the data provided in the project proposal, we also found the following dataset which was handy in this preliminary analysis:

Buildings Featuring Existing Affordable Rental Units

(Source - <https://www.bostonplans.org/housing/finding-housing/property-listings>)

2. Preliminary Analysis of the Data

For Deliverable 1, our main focus was on distribution of affordable housing inventory in D9.

In the collected data, some useful information was found:

1. The number of buildings with affordable housing in each neighborhood, and the number of affordable units in each building.
2. The different bedroom sizes available in each building.
3. The AMI range of renters who could afford an apartment in each building.

Data Preprocessing:

1. Removal of some redundant symbols that might affect understanding. Example: ‘\n’, ‘\t’.
2. Splitting the “Bedroom Sizes” column into five columns, representing “studio”, “1BR”, “2BR”, “3BR”, and “4BR”.
3. Splitting the “AMI Categories” column into two columns, “AMI_min” and ” AMI_max”. Extracting data in the original column (AMI Categories), recording the maximum and minimum values of AMI, and updating the new columns.
4. Adding two columns: “latitude” and “longitude”. Using neighborhood banes to specify the specific location of the building.

	Neighborhood	Building Address	For Current Vacancies and More Information	Number of Units Tracked	Bedroom Sizes	AMI Categories	1BR	2BR	3BR	4BR	Studio	AMI	AMI_max	AMI_min
0	Allston	8 Griggs Street	The Mount Vernon Company: 617.267.0006	2.0	1BR	80-100%; 100%	1	0	0	0	0	80 100 ; 100	100	80
1	Allston	Continuum, 219 Western Avenue	617.603.5660 info@continuumallston.com	42.0	Studio, 1BR, 2BR, 3BR	70%	1	1	1	0	1	70	70	70
2	Allston	The Edge, 60–66 Brainerd Road	National Development: 857.576.6160	10.0	Studio, 1BR, 2BR	70%	1	1	0	0	1	70	70	70
3	Allston	The Element , 65 Brainerd Road	National Development: 857.576.6160	4.0	Studio, 1BR	70%	1	0	0	0	1	70	70	70
4	Allston	The Eco Allston, 75 Brainerd Road	National Development: 857.576.6160	14.0	Studio, 1BR, 2BR	70%	1	1	0	0	1	70	70	70
...
83	South End	Troy Boston, 275 Albany Street	Equity Residential:617-517-7302	38.0	Studio, 1BR, 2BR	70%	1	1	0	0	1	70	70	70
84	West End	Avalon North Station	Avalon Communities: 617.236.1369	30.0	Studio, 1BR, 2BR, 3BR	80%, 80-120%	1	1	1	0	1	80 80 120	120	80
85	West End	Emerson Place, 72 Stanford Street, (1 & 10 Em...	Equity Residential: 617.973.2200	19.0	Studio, 1BR, 2BR	90%	1	1	0	0	1	90	90	90
86	West Roxbury	Belgrade Place, 446 Belgrade Avenue	Maloney Properties: 617.209.5406 Belgrade@malo...	5.0	Studio, 1BR, 2BR, 3BR	70%	1	1	1	0	1	70	70	70
87	West Roxbury	Residences at Snow Place, 1215 Centre Street	Rogerson Communities, Inc.: 617.325.7900	12.0	1BR	80%	1	0	0	0	0	80	80	80

87 rows × 14 columns

Table 1: Cleaned Data

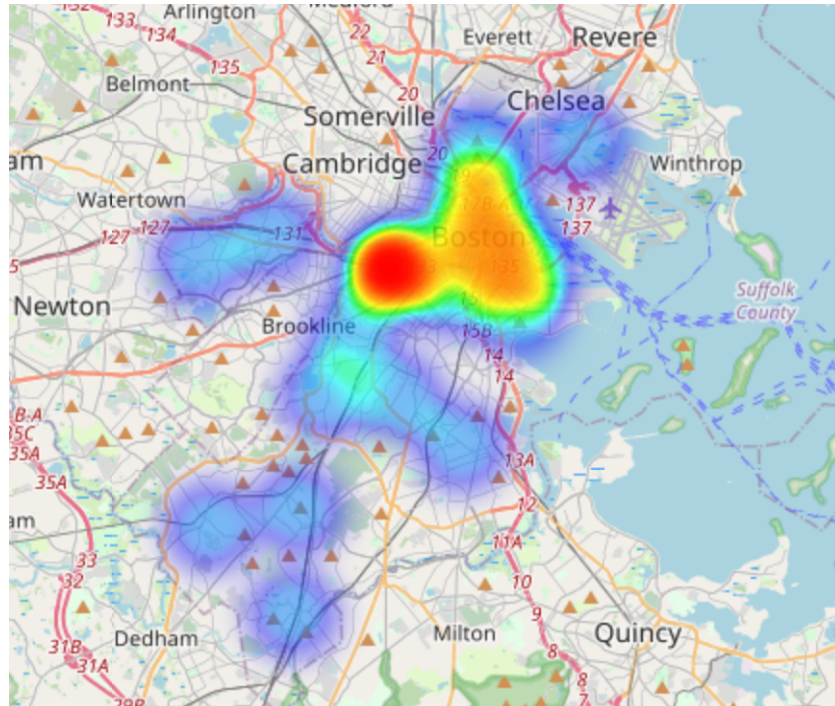


Fig 1. HeatMap of distribution of affordable housing inventory

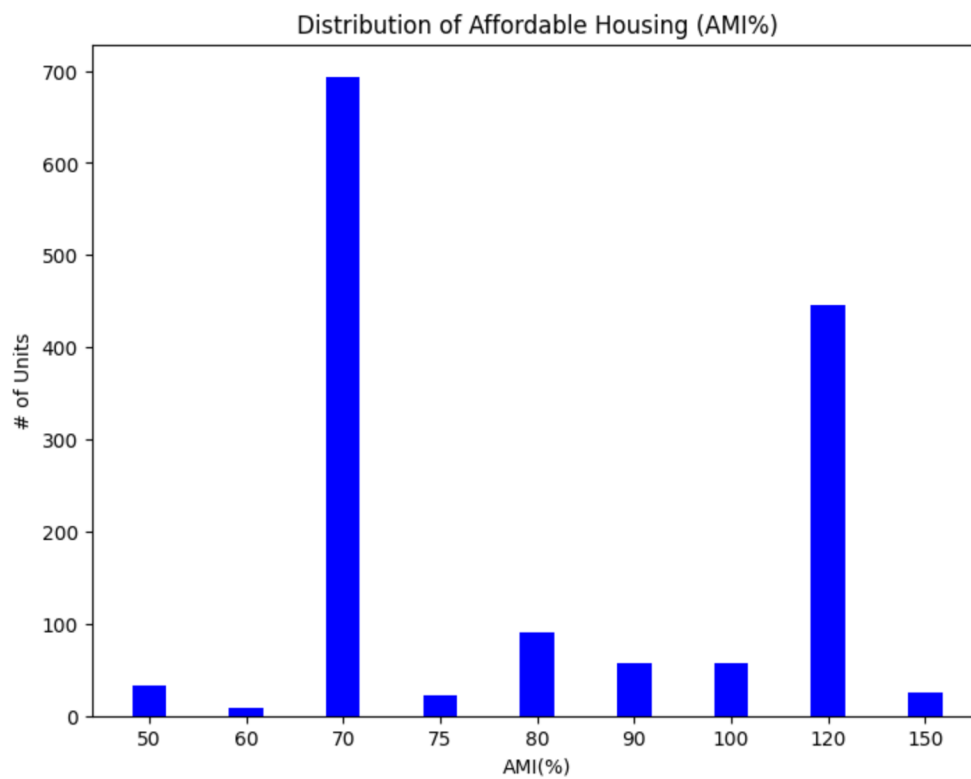


Fig 2. Distribution of affordable housing based on AMI

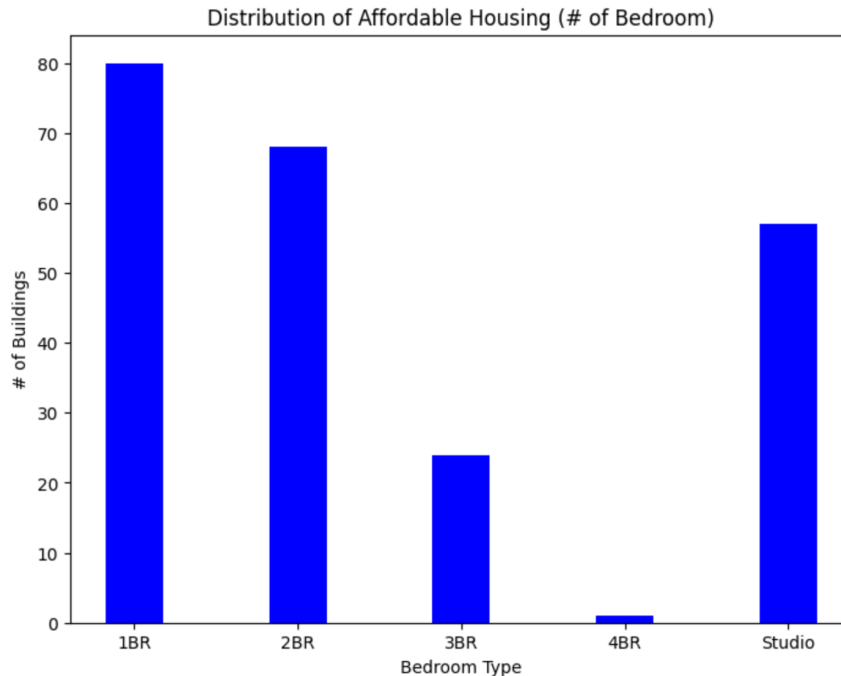


Fig 3. Distribution of affordable housing based on number of Bedroom

3. Answering One Key Question

The question we are concentrating on in this section is “What is the distribution of affordable housing inventory in D9 based on AMI% and # of units (e.g. 1 bedroom etc.)?”

According to the heatmap (Fig 1.), the highest concentration of available affordable housing units are located in the area south of Kenmore on Boylston Street near Charles Gate. Overall, available and affordable houses are spread all over Boston city with units in Jamaican Plane, West Roxbury, Brighton, Allston, Charlestown, Downton Boston, South Boston, and a few other major areas in Boston City.

Fig 2. represents a bar graph visualizing the number of available and affordable houses based on AMI. Majority of the houses (about 46% of the total affordable houses) have an AMI of 70%. Another major chunk of affordable houses (about 32% of the total affordable houses) have an AMI of 120%. AMIs range from 50% to 150% with majority houses (78%) in the above mentioned AMI percentages.

When it comes to the distribution of affordable housing based on the number of bedrooms (Fig 3.), the majority of the affordable houses are 1 Bedroom and 2 Bedroom houses (33% and 29%

respectively). Studio apartments are the third highest in the number of affordable houses (about 25%). The number of affordable 4 Bedroom houses are negligible.

4. Refining Project Scope and Project Limitations

The following are a few updated limitations of this project:

- The information presented on the Dashboard might also be used by realtors and landlords (who rent their house(s)) to increase prices (misuse of information)
- This project has multiple datasets (i.e., multiple sources of data), which might sometimes be a disadvantage and confusing.
- In Data Science projects, there is always the threat of Data Privacy Violation.
- The datasets might contain a bias. The datasets will be tested for that as well.
- The dataset collected concerning building features describes the types of units in each building. However, there is no information about how many of each type of unit is in each building, so analysis involving these units is limited.
- The building features dataset is somewhat limited in scope, containing less than 100 rows.
- Some of the datasets present data that is not up to date. Some of the permit data dates back to 2012, for example, and other datasets contain data only up to 2019 or 2020. Much of this data does not reflect up to date changes since the pandemic.
- Some of the building features data shows AMI ranges as opposed to exact groupings, making the data more difficult to categorize.

After the client meeting on 10/17/2022, the project has pivoted from the original description. This project will now focus more on identifying and extracting information about bad landlords. Affordability still remains a part of this project but it can be explored in the later half. Currently we plan on focussing on four main datasets:

1. [311 data](#) (Residents)
2. <https://data.boston.gov/dataset/public-works-violations> (Public Works)
3. <https://data.boston.gov/dataset/building-and-property-violations1> (ISD)
4. [Rentsmart data](#)