# housingXpert: Navigating NYC and Boston Airbnbs

INFO 3300: P2 Final Report

Henry Chen (hc659), Carnell Zhou (cz375), Nicole Jeon (nhj7), Christina Zeng (cwz6)

## Introduction

Annually, millions of travelers reliably choose Airbnb for their accommodation needs, drawn by its unique offerings and diverse range of housing options. In this project, we delve into a comprehensive analysis of Airbnb housing data in two vibrant cities: New York City and Boston. Our focus is to compare and contrast these cities in terms of pricing, guest ratings, and other key attributes, offering a deeper understanding of the Airbnb landscape. Additionally, our study incorporates geographical insights, as each data point is meticulously plotted on a map, allowing us to explore spatial patterns and trends. Through our detailed analysis and interactive visualizations, we aim to provide valuable insights that could guide travelers in making informed choices for their next Airbnb stay in these iconic cities.

## Data and Information

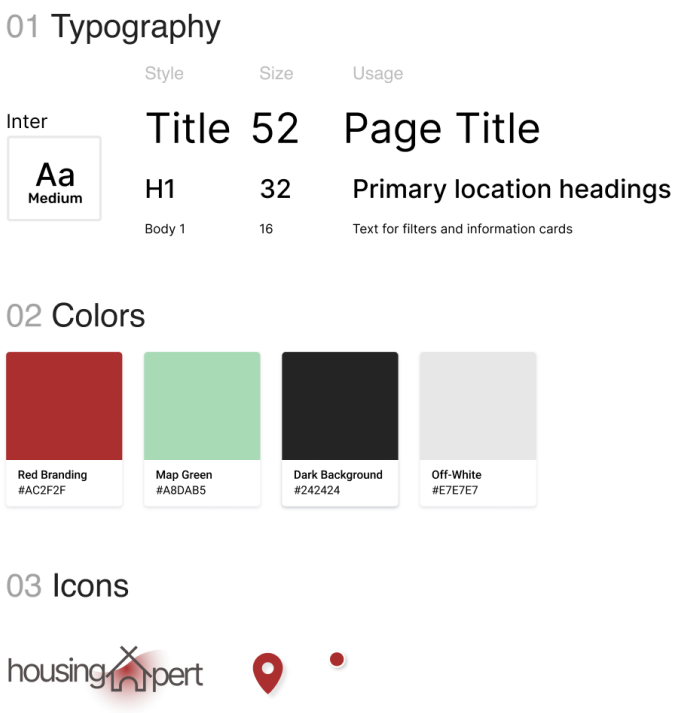
Within the dataset, the analysis comprises the housing listings from Boston and New York City. The goal within this visualization was to gain insight into the housing market by examining the various factors such as the price, location, and the amenities. The data was sourced from merging 2 CSV files into one JSON file. The data for nyc was found by randomly sampling a portion of the nyc csv file as it contained an immense amount of information while the entirety of the boston dataset was used. The data sources that were used for this visualization was the CSV file, “listings.-csv” and “listing-2.cvs” and two JSON files which were “boston.topojson” and “NTA.topojson”. The CSV files contained the information regarding the housing listings in Boston and NYC while the JSON includes the topological data required for mapping the neighborhoods within Boston and NYC. The dataset includes the amenities in the listing name for the housing listing and from string parsing over the name field we were able to obtain information crucial for ratings, the number of beds, and the number of baths.The name field was the variable that contained various information regarding the amenities of the listing. Latitude and longitude are provided for each of the listings which allows for the precise mapping and spatial analysis. The creation of the map was provided by the shapefiles “boston.topojson” and “NTA.topojson” files. Topojson was used to create the outlines for the map and the mesh for the neighborhoods while D3.js was utilized to create the interactivity of the visualization. The code included the functionality for dynamically filtering the data based on the user input which allowed for the exploration of specific neighborhoods, price ranges, and other criteria.

## Interactivity and Design

When exploring the dataset it requires the clear extraction of relevant information and envisioning what data will be necessary for relevant visualization. We focused on the housing datasets of Boston and NYC, and employed clear, interactive and visually appealing tools. Our goal was to strike a balance, providing users with valuable insights without overwhelming them with excessive information.

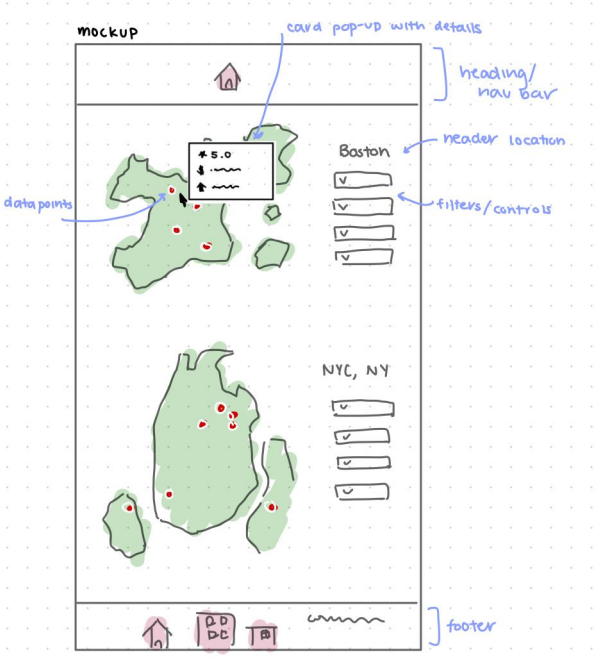
To achieve our objective, we initiated the process by developing a UI Kit, as illustrated in Figure 1, with the aim of ensuring consistency across both cities. The use of red serves to emphasize crucial information in the infographic, guiding the users' attention to elements such as interactive map pins and city-indicating headers. Incorporating green contributes to user trust and familiarity, tapping into the color's association with common navigation platforms. Furthermore, we established an information hierarchy by adopting a unified style for Titles, Headers, and Body text. Given our overarching goal of assisting users in making decisions in unfamiliar locations, building trust is crucial. By utilizing conventional design patterns we are cultivating a professional atmosphere, instilling confidence in our users to use our platform.

**Figure 1**
UI Kit for typography, color scheme, and icons



Our next steps were to create a few sketches of potential layouts of our page. Through numerous iterations, we settled on the design depicted in Figure 2. This design aligns with our primary objective of delivering valuable insights to users without providing excessive information. The design facilitates individual city viewing, ensuring that users can focus on one city at a time. Additionally, we decided to reveal housing details only when the user hovers over the drop pin, as opposed to presenting all data points simultaneously to further reduce clutter and excessive information all at once.

**Figure 2**
Low-fidelity sketch of the layout

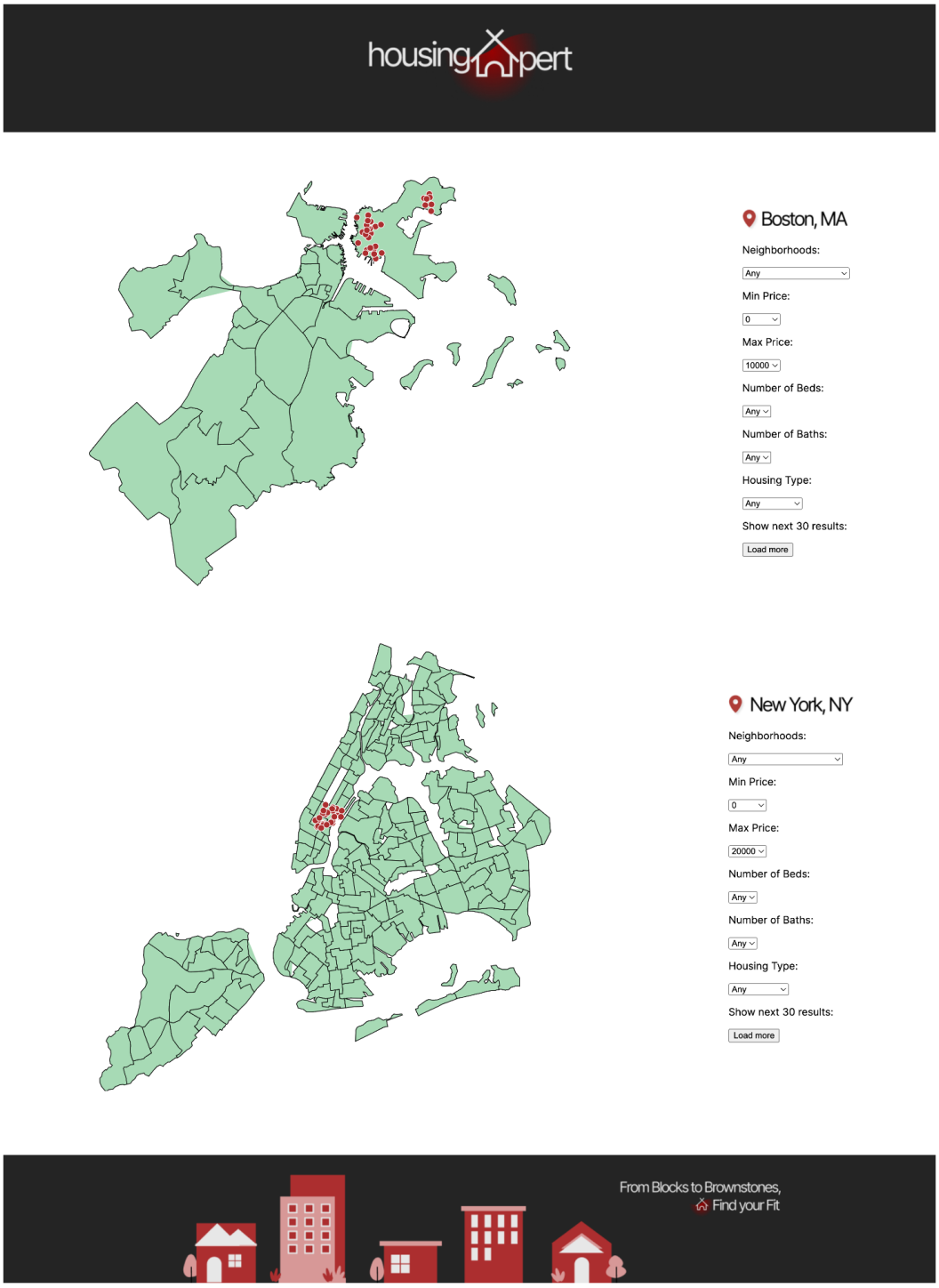



One central element of the project was to create an interactive map utilizing D3.js. The map serves as a visual representation of housing listings with each property marked as a hoverable point. To enhance the user experience, we incorporated smooth animations and relevant information for each of the listings. In addition, the project includes dynamic filters that are used to refine the display listings based on criteria such as price, bedrooms, bathroom, and housing type. The users would then be able to interact with the dropdowns fields to set their preference, and the map will update in real time. Furthermore, we balanced information density and clarity by displaying information for a given dynamic query in array groups to avoid overwhelming the user. When users hover over the listings on the map, the information card pop-up will display information for each of the listings and give details such as the price, rating, and housing type. For more granular results the user can use the dropdown menu and filter all of the listings perhaps for a destination or neighborhood they have in mind or if they are just exploring the area or potential areas for housing. We originally played around with different ideas for representing

the data points, such as using the d3 ordinal scale to assign each datapoint a color, but realized this failed to convey consistency among the plots and even could be confusing at times (do different colors signify a quality or feature about a specific rental unit?). As a result, we stuck to red points with white outlines that not only helped distinguish overlapping points but also helped the points pop out from the green fill of the maps.

**Figure 3**

Our design for a housingXpert site featuring Boston and NYC Airbnb listings



## The Story

The visualization of the housing data collected from Boston and New York City was able to serve as a lens through which one can unravel the intricacies of urban living. We wanted to offer the multi-faceted exploration of the housing landscape into two major urban centers. Through this, one can navigate through a wealth of information from the distribution property type to the spatial arrangement of high-rated accommodation. One of the key insights was the distinct housing profiles offered from both locations. They both manifested in the prevalence of certain property types as well as their geographic concentrations. For instance, the visualization revealed a higher density of serviced apartments in one city compared to the other which hints at the unique urban living preferences.

An intriguing revelation that lied within the spatial distribution of housing prices allowed us to shed light on the affordability of a specific area. The visualization allows one to discern patterns of affordability and identify pockets of relative costliness or affordability within each city. The spatial dimension leads to insights into urban development, socioeconomic disparities, and the influence of neighborhood characteristics and particularly location on housing costs. The insights would be crucial for not only policymakers, but urban planners and individuals seeking housing options that align with their financial means.

When doing the visualization, we believed that the visualization would be relevant as it may uncover the trends in the neighborhood and the housing demand for a given area where demand for a given area can be reflected from the relative housing prices in that area. For property owners this can give them an idea of what price they may want to list their housing for given their specific areas and the amenities for that housing. Likewise as travelers, typically it is cheaper to rent an airbnb out instead of living in a hotel when visiting a city and having a visualization that finds lower cost airbnb and the number of people that it houses is extremely useful and then the pricing scheme will also give us a general idea of how safe a given area may be where higher prices typically are associated with wealthier areas and lower crime.

The dynamic filters empower users to customize their exploration based on their specific needs by taking into account the most relevant features such as amenities and the pricing associated with those features. Users are able  to filter by price range, property type, and other parameters to facilitate a more targeted and meaningful exploration of the dataset as well as learning more about housing and the cost of housing in these cities.

In conclusion, the visualization provides a narrative that unfolds the story of urban housing in Boston and NYC. The insights derived from spatial and categorical analysis empowers users with a comprehensive understanding of housing dynamics and how city and location may affect those dynamics. Surprising discoveries may challenge preconceptions,

encouraging a deeper exploration of the factors influencing housing choices and costs. Ultimately, the visualizations invite users to step into an intricate analysis of urban living and armed with this knowledge they are given a better perspective on the housing landscape of two modern cities.

## Team Contributions

Henry – Filtered and merged datasets, and implemented user inputs in dropdown menus and performed additions to the rest of the code.
Carnell – Plotted map outlines for both cities and implemented data filtering according to user inputs. Helped with datapoint plotting and information card popup.
Nicole – Implementation of datapoint plotting and appropriate design choices (jiggle function, color scheme) and led write up.
Christina – Designed page visualizations such as logos and led the design process for the project. Helped with design choices (plot colors, layouts) and made additions to code base.

Time spent in filtering dataset and merging data: 2 hours
Time spent in creating map outlines using topo: 1.5 hours
Time spent in implementing logic for user inputs: 3 hours
Time spent in plotting datapoints and filter logic: 3 hours
Time spent in designing legends: 2 hour
Time spent on writeup: 3 hours