

CSE 564 Final Project Proposal - Group 21

Visualization of Airbnb Listings in New York City

Arjun Shrivatsa Omampuliyur Balakrishnan (114999692)

Ansh Nirav Mehta (115135282)

Background

New York City is one of the most dynamic and diverse housing rental markets in the world, with a wide range of neighborhoods, property types, and price ranges. The NYC housing rental market has experienced significant fluctuations over the past few years, with changes in demand, supply, and regulations affecting house prices and affordability.

The COVID-19 pandemic has also had a profound impact on the NYC rental market, with many residents leaving the city for suburbs or other cities, while others have taken advantage of lower prices and interest rates to buy properties.

Given these complex dynamics, there is a need for data-driven insights and visualizations that can help renters navigate this. Our project aims to fill this gap by collecting and analyzing data on NYC Airbnb house rental prices and related variables and visualizing the results in an accessible and interactive format.

Airbnb is a popular online marketplace that allows individuals to rent out their homes or apartments to travelers. With over 40,000 listings in New York City in the past year (2022) alone, it can be challenging to navigate through the vast amounts of data available to find the ideal listing. Data visualization can help simplify this process and make it easier for users to find the most appropriate and desired Airbnb listings.

Problem Statement

As outlined above, the increased number of choices for accommodation has made choosing an appropriate housing space overwhelming. Thus there is a requirement to better aid the user in making informed choices. When choosing a particular housing space, there is a need to equip the user with basic information in a top down manner such as

- The locality to be considered
- Average Price of Rentals in the locality
- Availability of rentals throughout the year to give a better perspective on how much in advance the booking needs to be made
- Top hosts of the area

- Distribution of apartment types like Studio, Private room etc for targeted decision making.

Our dashboard aims to visualize these trends for New York City and simplify the problem of house rental booking for the end user.

Dataset

We work with a dataset of NYC Airbnb Listings data for the year 2022 as provided [in this link](#). The primary dataset is the listings.csv file and following is the dataset description. It contains data for 42,000+ listings.

Sno	Variable Name	Description
1	id	Unique ID for the listing
2	name	Name of the Listing
3	host_id	Unique ID for the Host
4	host_name	Name of the host
5	neighbourhood_group	Borough of the listing: Eg: Brooklyn, Bronx
6	neighborhood	Neighborhood of the listing
7	latitude	Geographic Latitude of the listing
8	longitude	Geographical Longitude of the listing
9	room_type	Categorical variable describing room type: Private/ Shared room/ Apartment etc
10	price	Price per night of the listing
11	minimum_nights	Minimum no of nights required per booking
12	number_of_reviews	Total no of reviews on airbnb platform for the particular listing
13	last_review	Date of the most recent review
14	calculated_host_listings_count	No of listings on airbnb platform for the particular host. To identify multi-listings hosts.
15	availability_365	No of nights the property was available (not booked) in 2022
16	no_of_nights_booked	No of nights the property was booked in 2022. Calculated as 365 - availability_365

In addition to the listings.csv file, we plan to use neighborhoods.csv and neighborhoods.geojson file which has the geographical location of every neighborhood in the borough to plot the choropleth map.

Dataset Preprocessing

- The original listings.csv file consists of 17 columns out of which reviews_per_month and number_of_reviews_ltm were removed for lesser significance.
- No of nights booked is a derived column from availability_365. Calculated as $365 - \text{availability_365}$.
- Additional preprocessing if required is to be determined through the course of the project work.

Proposed Dashboard and Expected Behaviour (Approach)

Visualizations

We plan to implement a dashboard with the following visualizations:

1. Choropleth Map of NYC showing all the boroughs. This will be color gradient scaled to reflect the number of listings in each borough.
2. Room type Donut Chart: This chart would show the distribution of the different types of rooms available ie. Private Rooms/ Shared Rooms/ Entire Apartments/ Hotel Listings
3. Boxplot with Jitters: This plot would show the distribution of price per night of the listings as a boxplot with jitters to better understand median and quartile pricing of the listings.
4. Pie Chart: This is a pie chart to understand the split up of short term and long term rental listings.
5. Stacked bar Chart: This chart shows the top 10 hosts along the X axis and every bar represents the number of listings of the hosts segregated by the room types.
6. PCP Plot: This enables the user to visualize the relationship between the different variables in the dataset and observe correlation patterns.
7. PCA Biplot: The two highest components with the highest square root of sum of loading values are obtained and the data points are plotted as a PCA Biplot.

Filters and Interactivity

By default, upon loading all the graphs displayed contain data for entire New York City. A universal filter for Borough is implemented in all the visualizations ie. all the graphs adjust to the choice of Borough the user wishes to visualize.

For example, the user can select a particular borough or a set of boroughs from the choropleth map and all the remaining graphs change to reflect data for that particular selection.

Similarly, all data points in the PCA Biplot and Boxplot will be color coded to reflect their belonging to a particular borough and the user can click and drag a selection of points which filters the other graphs only for the borough selections made. This is planned to be realized using the brush functionality in D3.

Implementation

- We plan to divide the frontend into a card-style layout with each card designated for a particular visualization. There would be 7 cards for each of the seven planned plots. In addition, there would be a reset button to undo all the applied filters.
- Tech Stack:
 - HTML, CSS - Skeleton of the webpage and styling
 - Bootstrap - Template for card layout of frontend.
 - Javascript - For DOM manipulation.
 - D3.js - Using SVG for plots
 - Flask (python) - Backend server.

Dashboard Layout


