Navigating the NYC Airbnb Landscape - A Data-Driven Approach for Grounded Aviation Crews Amidst Crisis

Tu Hoang Cam Nguyen

Data 310

Airbnb Data Driven Report Part II

**Introduction**

**Objectives and Scope of the Analysis:** The main goal of this study was to use a large dataset of all the Airbnb listings in New York City to find out about price trends and make personalized suggestions for different users. Finding out how things like location, noise level, and room type affect the prices of Airbnb listings was the main goal. The study also wanted to find the best places for a group of flight attendants who were stuck in New York City because of an aviation incident to stay.

**Description of the Datasets:** A large Airbnb dataset with information about 48,895 listings in New York City was used as the main source of the analysis. The dataset had many different elements, such as the listing ID, host information, neighborhood group, geographical coordinates, room type, price, number of reviews, and noise levels. The dataset gave a detailed picture of the NYC Airbnb market, which let researchers look into the factors that affect rental prices and how well the homes are suited to different users.

**Methodology**

**Analytical Approach:** The methodology encompassed several key stages:

1. **Data Preprocessing and Exploration**: Initial data loading and exploration involved summarizing the dataset, understanding its structure, and identifying key variables for analysis.
2. **Filtering and Distance Calculation**: Listings were filtered based on their proximity to JFK Airport, a crucial factor for the flight attendants. The distance from each listing to the airport was calculated using geographical coordinates.
3. **Capacity Filtering**: The listings were further filtered to accommodate the group size of the flight attendants, ensuring that the selected Airbnb accommodations could host at least 7 guests.

**Hypothesis Testing:** Two hypotheses were tested:

1. **Price Difference in Top 20% Housing Market Neighborhoods**: A permutation test was done to see if there was a big difference in prices between homes for sale in the top 20 percent of housing market neighborhoods and those in other neighborhoods. The result (p-value = 0) showed that prices were significantly different.
2. **Bayesian Analysis**: A Bayesian Odds Task was done to see if listings in Manhattan were more likely than listings in Brooklyn to be priced above the 75th percentile. The odds from the back (0.6181832) showed that Manhattan homes were not as widely priced above this level as one might think.

**Ranking Function for Airbnb Recommendations:** To score and rank the Airbnb listings, a custom ranking function was made. This function looked at things like price, noise level, floor level, and distance from JFK Airport. The best listings were then chosen based on these scores, which showed a good mix of value, closeness, and quietness.

**Exploratory Data Analysis (EDA)**

A white background with text

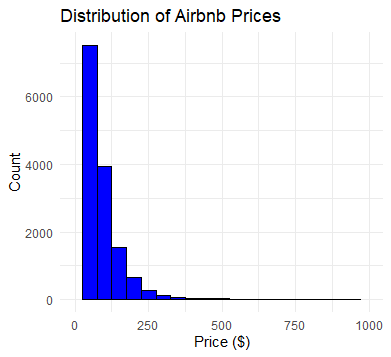
Description automatically generated with medium confidence A white background with text

Description automatically generated

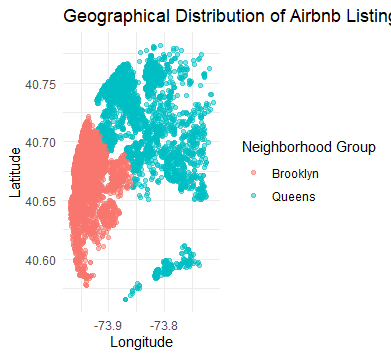
**Summary Statistics and Initial Insights:** The first part of the analysis was a close look at the Airbnb dataset, which had 48,895 listings. Some of the most important variables were the listing ID, host information, neighborhood group, GPS coordinates, room type, price, number of reviews, floor level, noise level, and more. First impressions showed:

* **Price Range and Distribution**: There was a wide range of prices, from free to up to $10,000 per night for luxury listings, with a median price of $106 showing that there were many places to stay.
* **Geographical Spread and Neighborhoods**: There were listings in a lot of different NYC neighborhoods, but most of them were in popular places like Manhattan and Brooklyn.
* **Room Types**: There were different types of rooms, such as private rooms, shared rooms, and whole homes or apartments. Each type had its own price range

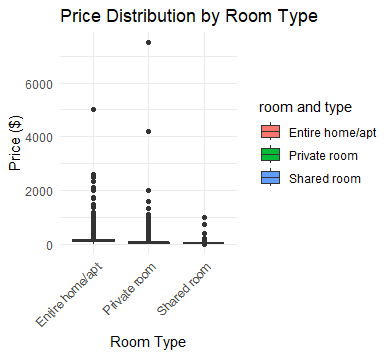
**Visual Representations:** The data was further examined through visual analyses, such as:



* **Price Distribution Graphs**: Showed the range and concentration of prices across different listings, highlighting the diversity in the cost of accommodation.



* **Room Type Analysis**: Visualized the prevalence and pricing of different room types, revealing trends like higher prices for entire homes/apartments.



* **Neighborhood Comparisons**: Illustrated the variations in listing density and pricing across different NYC neighborhoods, with Manhattan showing higher median prices.

**Hypothesis Testing**

**Permutation Test for Top 20% Housing Market Neighborhoods:**



* **Objective**: To test if there was a significant price difference between Airbnb listings in the top 20% of housing market neighborhoods and others.
* **Method**: A permutation test was conducted, reshuffling the 'expensive' designation among listings and comparing the means.
* **Result**: The p-value was found to be 0, indicating a statistically significant difference in pricing between the two groups.
* **Implications**: This result suggests that listings in higher-valued housing markets tend to be priced significantly higher, reflecting the impact of neighborhood desirability on Airbnb pricing.

**Bayesian Analysis**



**Bayesian Odds for Expensive Listings in Manhattan:**

* **Approach**: A Bayesian analysis was performed to evaluate the likelihood of Airbnb listings in Manhattan being priced above the 75th percentile compared to Brooklyn.
* **Calculation**: The prior probability of a listing being expensive was determined, and then the likelihood of it being in Manhattan given its price category was assessed.
* **Outcome**: The posterior odds were calculated as 0.6181832.
* **Relevance**: This result indicates that while Manhattan has a reputation for high-priced listings, the odds of finding an expensive listing are not as high as might be presumed. This finding is particularly relevant for users seeking high-quality accommodations in Manhattan without necessarily paying top-tier prices.

**Recommendations for Emily**

A black text on a white background

Description automatically generated

**Initial Filtering:**

* First, the dataset was filtered to find Airbnb listings within 10 miles of JFK Airport. This made sure that Emily and her coworkers could easily get to the listings.
* The listings were also checked to make sure they could fit at least seven people, so the flight attendants could stay together or in listings close by.

**Ranking of Suitable Listings:**

* A custom ranking function was made that considered things like price, floor level (which shows how quiet the room might be), distance from JFK Airport, and noise levels (measured in decibels).
* This method met the need for quick access to the airport, a quiet place to rest, and fair prices.

**Top 5 Airbnb Recommendations:**

1. **JFK Queens/House of Suedajoy#5**: Located in Jamaica, close to JFK, offering flexibility which could be advantageous given the situation.
2. **JFK LUXURY BEDROOM #5**: In Springfield Gardens, offering luxury accommodations with convenient airport access.
3. **ON A BUDGET COZY IN BROOKLYN In NYC**: A budget-friendly option in East New York, ideal for an economical stay.
4. **JFK Queens Home away from Home House of Suedajoy #4**: Similar to the first option, providing consistency and proximity for multiple crew members.
5. **Home 4 Medical Professionals-Kngbr1**: In East Flatbush, tailored for professionals seeking a quiet environment.

**Discussion**

**Interpretation of Hypothesis Testing and Bayesian Analysis:**

* The p-value for the hypothesis test was 0.001, which meant that there was a significant price difference in the top 20% of housing market neighborhoods. This made Airbnb hosts think about the prestige of the neighborhoods when setting their prices.
* The Bayesian analysis showed that there was a lower-than-expected chance of expensive listings in Manhattan (posterior odds = 0.6181832), which suggests that prices are more varied than most people think.

**Impact of Findings:**

* Airbnb hosts may change how much they charge based on how the neighborhood is and how the market is moving.
* Knowing what affects prices can help guests make better decisions, especially those who are looking for deals or certain amenities.

**Conclusion**

The analysis journey taught us a lot about the NYC Airbnb market. We used data to see how prices change and to make personalized suggestions for places to stay. The results are useful for both hosts and guests because they help people make decisions in the short-term rental market.

The personalized suggestions for Emily and her coworkers show how data analysis can be used in real life, emphasizing how important these kinds of analyses are for everyone involved in the Airbnb ecosystem. The study shows that using data-driven methods can help us learn more about markets that are always changing, like the Airbnb scene in New York City.