NYC Airbnb Project □

Project Introduction□

Since 2008, guests and hosts have used Airbnb to expand on traveling possibilities and present more unique, personalized way of experiencing the world. This project aims to analyze the New York City Airbnb dataset to uncover insights into listing characteristics, pricing patterns, and host popularity. By exploring various aspects of the data, we seek to answer questions related to the influence of neighborhoods, the popularity of room types, and the correlation between host popularity and their number of listings.

Data □

- The dataset is sourced from Kaggle, with data describing Airbnb listing activity and metrics in New York City, 2019.
- This is a public dataset and can be found on this website.
- The file includes data about hosts, geographical availability, room types and customer review dates
- This dataset has about 48895 entries with 16 columns.
- The data is a mix between categorical and numeric values.

Main challenges: Handling missing data, cleaning and renaming columns, and ensuring data integrity. **Strengths:** Rich dataset with a huge sample size. **Weaknesses:** Potential geoethical threat due to the display of sensitive information such as host names. Also, for further analysis, it would have been nice to have additional columns such as a customer review column scored by a likert scale system (for eg: 0-5) for each listing. This would have for example facilitated bivariate analysis on the satisfaction of customers for each listing.

Questions (Hypotheses) □

There are three questions (hypotheses) asked/tested. Each hypothesis was analysed and further used to propose a business opportunity **Hypothesis 1:** The neighborhood group significantly influences the price and number of Airbnb listings Question: Does the neighborhood group significantly influence Airbnb listing prices? Conclusion: Neighborhood groups indeed have a notable impact on listing prices. Certain areas command higher prices than others, due to location or demand. **Hypothesis 2:** Particular room types are more popular and demand higher prices. Question: Are particular room types more popular and demand higher prices? Conclusion: Certain room types are more popular, and they often attract higher prices. Entire homes/apartments tend to be both popular and pricier. Whereas shared rooms are cheaper and less popular. **Hypothesis 3:** Having higher listings has a positive correlation with a higher host popularity index (number of reviews)
Question: Does having higher listings correlate with a higher host popularity index (number of reviews)? Conclusion: There is a weak correlation between the number of listings and host's popularity index. Hosts with higher number of listings do not necessarily have more reviews. | Hypotheses Conclusions | Business Opportunities □ | | ------ | ------ | | H1: Neighborhoods indeed have a notable impact on listing numbers and prices. Certain areas command higher prices than others, potentially due to location or demand. | (Opportunity for planning authority/ minicipality of NYC): Identify and promote listings in low-performing geographic clusters, focusing marketing efforts and optimizing pricing strategies in these areas to decentralize population growth. | | H2: Certain room types are more popular, and they often attract

higher prices. Entire homes/apartments tend to be both popular and pricier. | (Opportunity for hosts) Identify and promote(invest in building) popular room types to attract more guests and increase overall occupancy + revenue. | | H3: There's a weak correlation between the number of listings and a host's popularity index. Hosts with higher number of listings do not necessarily have more reviews. | There's an opportunity to optimize listing promotion strategies to enhance the visibility and popularity of diverse hosts, irrespective of the number of listings. **Recommendations** 1. Quality Emphasis: Encourage hosts to prioritize the quality of their listings over quantity. Implement initiatives that reward hosts for maintaining high-quality listings. 2. Targeted marketing: Implement targeted marketing campaigns that cater to hosts with a diverse number of listings to attract a broader customer range.|

Methodology

> *Steps for data cleaning, wrangling and analysis*

Data loading and exploration:

- On first view of the data,
- The 'name' column was dropped because it was not relevant for the analysis. Additionally, there was a corresponding 'ID' column which served the same purpose. it would have been redundant to keep both.
- The 'host_name' column was similarly dropped because it was not relevant for the analysis. There was a corresponding 'host_ID' column which served the same purpose and contained more unique values (37457), as against (11452) for 'host_name'. In geo-ethics, it is considered best practice to anonymize/avoid names of persons during data analysis for privacy and safety reasons.
- The 'last_review' column was similarly dropped because it was not relevant for the analysis.
- The null values of reviews per month were filled with the mean of values in this column. Considering the nature of values in this column,this aggregation method provides a representative and unbiased estimate, minimizing the impact of missing values on the distribution of this columns data which are used in the analysis.
- The 'availability_365' column was renamed to 'yearly_availability_365'. This is simply because, "availability_365" is not very descriptive. However, "yearly_availability_365" is more intuitive for all readers regardless of their affinity with the dataset.

Analysis

The Jupyter notebook containing all the codes and visualizations.

Conclusions after analysis

This Airbnb dataset of NYC, 2019 proved to be comprehensive, offering diverse column data for in-depth exploration.

The analysis justified two out of three hypotheses we proposed. First, it was proven that, neighborhood groups indeed have a notable impact on listing numbers and prices. Certain areas command higher prices than others, potentially due to location or demand. Subsequently, we analysed the relationship between the different room types and prices. The results show that certain room types are more popular, and they often attract higher prices. Entire homes/apartments tend to be both popular and pricier. Finally, we analysed if there was a positive correlation between having higher listings and higher host popularity index (number of reviews). Surprisingly, the data showed there was a weak correlation between the number of host listings and host's popularity index. Hosts with higher number of listings do not necessarily have more reviews.

Drawing from the conclusions of the above hypothesis, One business opportunity for hosts can be to identify and promote popular room types to attract more guests and increase overall occupancy which translates into revenue. Also, for the planning authority or Minicipality of NYC, they can identify and promote listings in low performing geographic clusters (for example the Bronx), focusing marketing efforts and optimizing pricing strategies to decentralize population settlement across the city. Further, There's an opportunity to optimize listing promotion strategies to enhance the visibility and popularity of diverse hosts, irrespective of the number of listings. Here, some recommendations are 1. Quality Emphasis: Encourage hosts to prioritize the quality of their listings over quantity. 2. Targeted marketing: Implement targeted marketing campaigns that cater to hosts with a diverse number of listings. Highlight unique and appealing features of these listings to attract a broader audience(with different income ranges/ preferences).

For further analysis, it would have been nice to have additional columns such as a customer review column scored by a likert scale system (for eg: 0-5) for each listing. This would have for example facilitated bivariate analysis on the satisfaction of customers for each listing. Nonetheless, the exploration and analysis revealed interesting relationships between various columns (variables) and proved sufficient to test the proposed hypotheses.

Further questions or points for research

Further exploration could be done on additional aspects such as seasonal pricing trends, the impact of amenities on pricing, and the effectiveness of promotional strategies on host success. More specifically:

- Are there seasonal patterns affecting the pricing and availability of Airbnb listings in New York City?
- Is the yearly availability of listings influenced by factors like neighborhood and room type?
- Do hosts with a higher number of listings have different pricing strategies?

Links

Jupyter notebook Dataset