

Airbnb Exploratory Data Analysis (EDA): Factors Influencing Customer Ratings and Demand

Jonathan Ho, Sarah Ki, Victor Ndayambaje and Ravi Siddappa
08/05/2024

Presentation: https://docs.google.com/presentation/d/1t4-oyCv7GZ_9qMkxWDCO_jD4JtnZtfL24uO2OYlnhc/edit?usp=sharing

Notebook: https://github.com/UC-Berkeley-I-School/Project2_Ki_Ho_Siddappa_Ndayambaje

Introduction

Shelter, a fundamental element of human existence, has evolved significantly over thousands of years. Traditionally, travelers had limited options, mainly hotels, which were often owned by large corporations. This landscape was disrupted by Airbnb, which launched in 2008. Airbnb is an online platform that connects hosts offering accommodations with guests seeking short-term stays. Users can list, discover, and book spaces ranging from apartments and houses to more unconventional lodgings. By 2024, Airbnb had surpassed 150 million users, with over 1.5 billion stays and 7.7+ million listings. The platform has also introduced new services like Airbnb Experiences, Airbnb Plus, Airbnb Luxe, and Airbnb Adventures to cater to different traveler preferences.

This report aims to explore the characteristics and features of Airbnb listings that influence customer ratings and demand. In particular, we will address the following key questions:

1. What are the top 10 countries with the highest number of listings?
2. Are there any correlations between the listings' features and rating?
3. How do popular and highly rated listings compare with less popular and more lowly rated listings in their features?
4. Which are the countries with both a high number of popular and highly rated listings, and high percentage of such listings?
5. Which locations in these countries in question four should Airbnb put more resources in?

We seek to provide actionable insights that can help leadership understand the landscape of Airbnb listings better and to make informed decisions about where to allocate resources.

Data Processing

Data Description and Cleaning

The Airbnb dataset provides detailed information on various rental properties, including both numerical and categorical features. The dataset includes 12,805 entries, each representing a

unique property listing. Below is an overview of the columns, their respective descriptions, and a brief description of how they were processed:

1. rating: A rating for the property, generally a float value, with a minimum of 3.0. Unrated properties are marked as "new." All "New" ratings were replaced as -1 to convert the column to float.
2. reviews: The number of reviews the property has received, to be converted from object to integer.
3. address: The address of the property.
4. features: A textual summary of the property's features, to be dropped as detailed features are available in other columns.
5. amenities: A list of amenities available at the property, provided as a comma-separated string. These were converted into boolean columns for amenities of interest. Total number of amenities was also calculated and input in a new column.
6. safety_rules: Safety rules for the property, also provided as a comma-separated string. This was dropped as none of the rules were of interest.
7. house_rules: Rules set by the host for the property. These were converted into boolean columns for rules of interest. Total number of rules was also calculated and input in a new column.
8. img_links: URLs to images of the property, potentially useful for analyzing the impact of images on reviews. Total number of links was calculated and input into a new column.
9. country: The country where the property is located, with values containing whitespaces that were removed.
10. checkin: The check-in time for the property.
11. checkout: The check-out time for the property, to be converted to datetime format and filled null values with 'Flexible.'

This dataset provides a comprehensive view of various properties listed on Airbnb, including their amenities, rules, and basic attributes. The data is essential for analyzing factors that influence customer ratings and demand.

Assumptions

1. All prices are in the country's currency. We used a third-party API to convert all foreign currencies to USD. Initial exploration showed the price range from 1,907,963 (listing in Turkey) to 393 (listing in India). This large range of prices with mixed currency settings caused inconsistencies in the conversion process, so price was not considered in our analysis.
2. When there is no check-in or check-out time provided, we assume it is flexible.
3. High rating and high number of reviews indicate a stellar listing, which means it is in higher demand than other listings (we considered listings in the 75th percentile as "stellar").

Analysis

Goal

The goal of this EDA is to understand the characteristics of Airbnb listings, and to derive insights that could guide leadership in what area(s) to invest more resources in.

What are the top 10 countries with the highest number of listings?

We first examine the top 10 countries with the highest number of Airbnb listings. Understanding these countries can provide several strategic benefits, including marketing opportunities, revenue insights, customer preferences, and more.

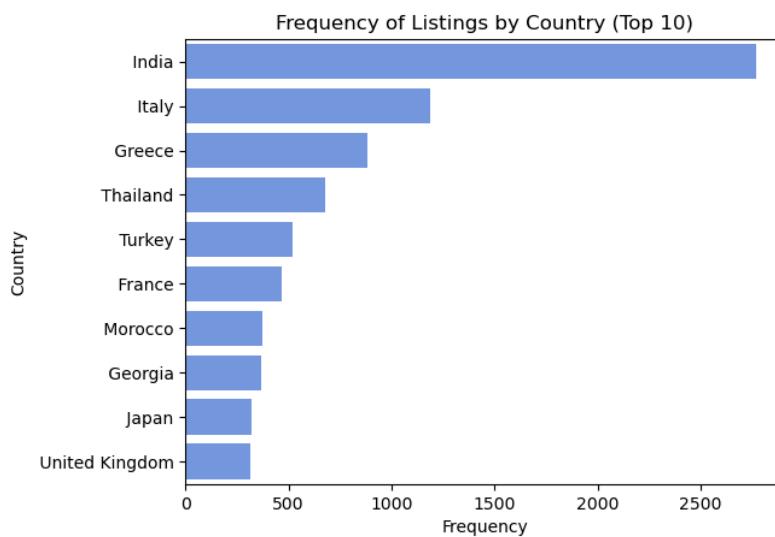


Figure 1: Distribution of Airbnb Listings by Country

As shown in Figure 1, India has the highest number of Airbnb listings. This presents a significant opportunity for Airbnb, given India's rapidly growing middle class and increasing (Bhutia, 2024) domestic travel. There is substantial potential for expanding Airbnb's presence in both urban and rural areas. Targeting emerging cities and regions with promotional campaigns can help capture a larger share of this growing market.

Many of the countries in the top 10 are well-known tourist destinations. Countries like France, Italy, and Japan are likely major contributors to Airbnb's revenue due to their high tourism volumes. Strategies to maximize revenue in these markets should focus on enhancing listings, and leveraging high tourist traffic to increase bookings.

Are there any correlations between the listings' features and rating?

Analyzing correlations between listing features and ratings helps identify opportunities to enhance customer experience and optimize host performance.

Correlation can be defined as the relationship between two variables. The closer the number is to either 1 or -1, the stronger the positive or negative correlation, respectively. The heatmap below reflects the correlation between ratings and different features of the listings.

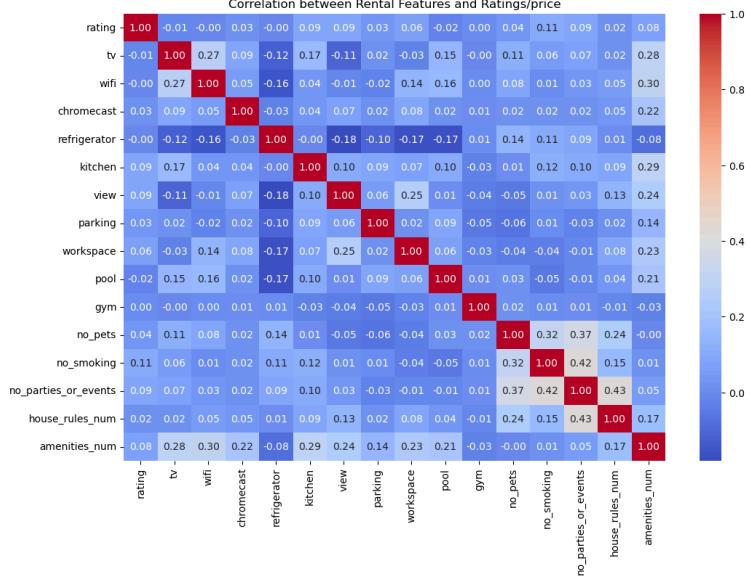


Figure 2: Correlation between Listing Features and Ratings

In Figure 2, we observe that listings often have more than one rule (e.g., no parties, no pets, no smoking), meaning that if a listing has one rule, it is likely to have others as well. The presence of multiple rules indicates that hosts are setting clear expectations for their guests. This helps in managing guest behavior, leading to higher satisfaction and better reviews.

Allowing hosts to set multiple rules can attract guests who are comfortable and willing to comply with these conditions, potentially leading to higher satisfaction and better reviews. While Airbnb is primarily for vacation guests, it also needs to manage the behavior of renters on their platform. Encouraging hosts to implement multiple rules can provide positive feedback to the hosts and eventually allow more renters to list their properties on Airbnb.

We were unable to determine which listing features are most strongly correlated with higher ratings. Consequently, we cannot identify the features that guests value the most. Given this outcome, we will categorize listings into "stellar" and other listings. By comparing the features of highly-rated listings, we aim to identify the characteristics that are more favorable and contribute to higher ratings.

What is a stellar listing?

We will identify highly regarded listings, termed "stellar," by looking at those with high ratings and numerous reviews. By comparing these stellar listings to "non-stellar" ones, we aim to inform Airbnb policies and identify locations with a higher potential for stellar listings, guiding resource allocation. Below, we examine the distribution of ratings and review counts.

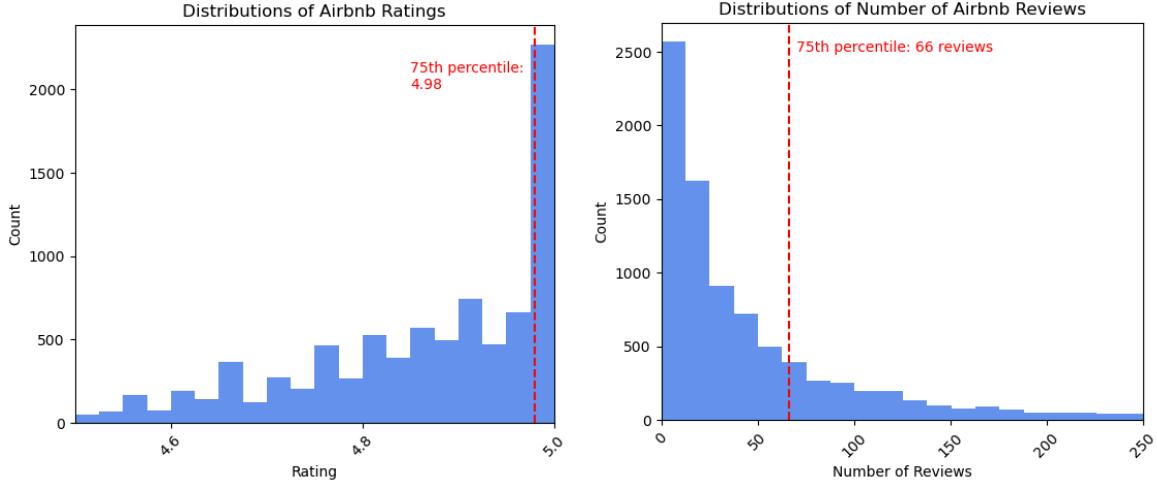


Figure 3: Distribution of Ratings and Number of Reviews

Figure 3 shows both the distribution of Ratings on the left and the Number of Reviews on the right. The ratings distribution indicates a left-skewed pattern, where a large number of listings have high ratings close to 5.0. The Number of Reviews distribution is right-skewed, with many listings receiving a low number of reviews, and the count decreasing as the number of reviews increases.

We define a stellar listing as one that is at or above the 75th percentile for both ratings and number of reviews. The 75th percentile threshold effectively captures the top quartile of listings, highlighting those that perform significantly better than the majority. Listings with ratings and review counts above this threshold are likely to offer consistently positive guest experiences, as reflected in both high ratings and a substantial number of reviews.

By considering both the number of reviews and the rating, this approach balances quantity with quality. A high number of reviews indicates popularity and broad appeal, while a high rating reflects guest satisfaction. Using both metrics together ensures that only listings with both broad appeal and high satisfaction are considered "stellar."

The graph shows that while it is common for users to give high ratings, it is not as common for listings to have more than 50 total reviews. Due to the tendency of Airbnb users to give high ratings, it is essential to incorporate the 75th percentile of the number of reviews into consideration.

How do stellar and non-stellar listings compare in the number of house rules?

As seen in the correlation plot (Figure 2), listings tend to have more than one rule. It is interesting to examine if stellar listings tend to have more or fewer rules. Below, we observe the distribution of the number of house rules.

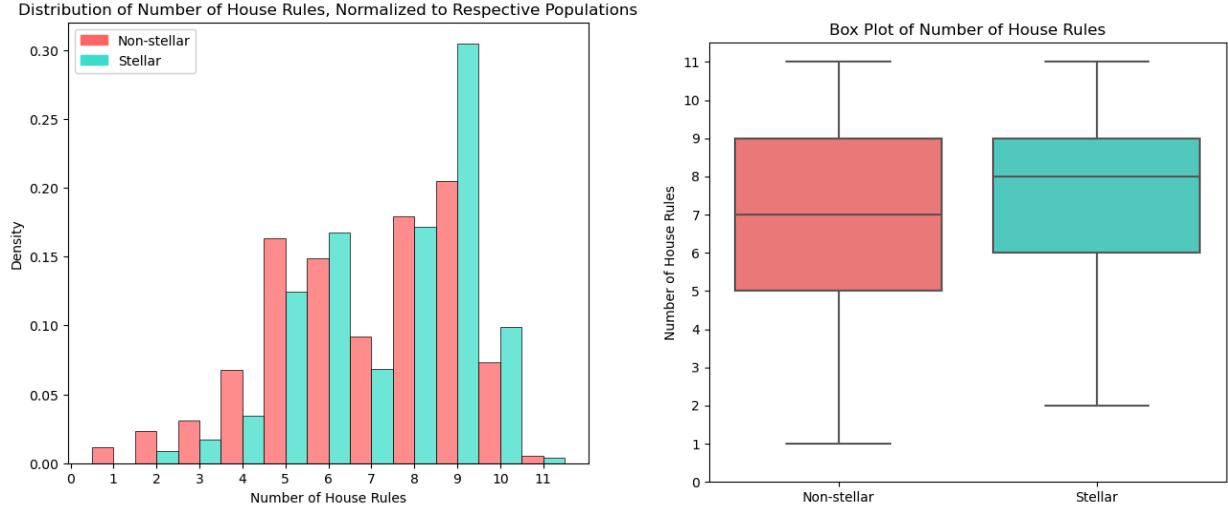


Figure 4: Distribution and Box Plot of Number of House Rules for Stellar and Non-Stellar Listings

We observe that stellar listings do tend to have more house rules in general. This could be because stellar listings might have clear and detailed house rules, setting expectations for guests and leading to better adherence and higher satisfaction. Additionally, it might be possible that more house rules reflect a host's commitment to maintaining a certain standard of quality and order, creating a more predictable and enjoyable experience for guests, thus contributing to higher ratings and positive reviews.

How do stellar and non-stellar listings compare in number of image links and number of amenities?

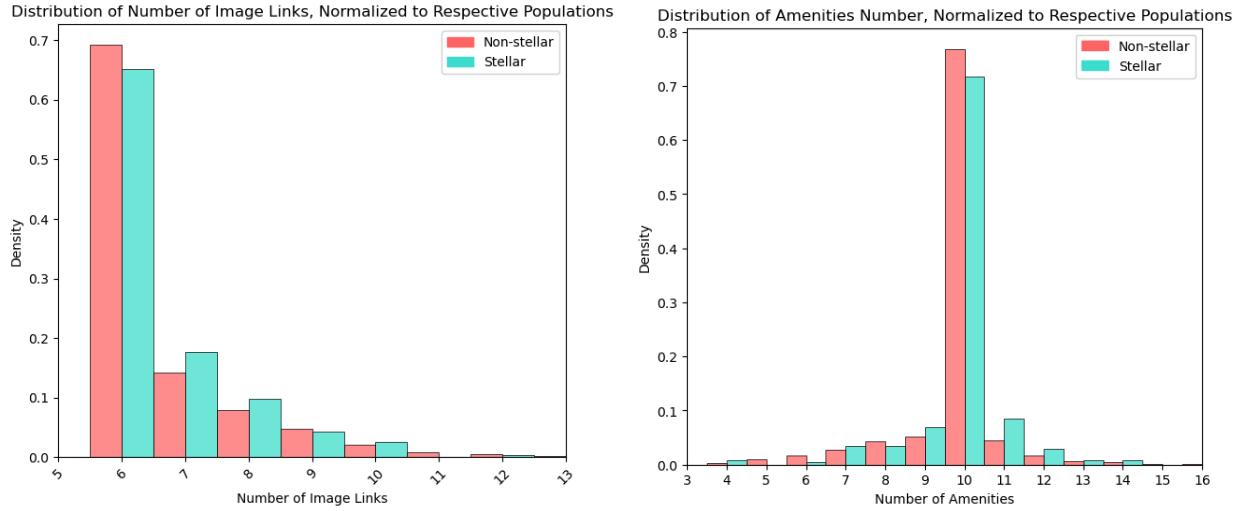


Figure 5: Distribution of Number of Image Links and Number of Amenities

We observe that there is not much difference between stellar and non-stellar listings in terms of the number of image links and the number of amenities. Both types of listings tend to have a

similar distribution for these features, indicating that the number of images and amenities might not be significant differentiators in achieving stellar status.

How do stellar and non-stellar listings compare in type of amenity?

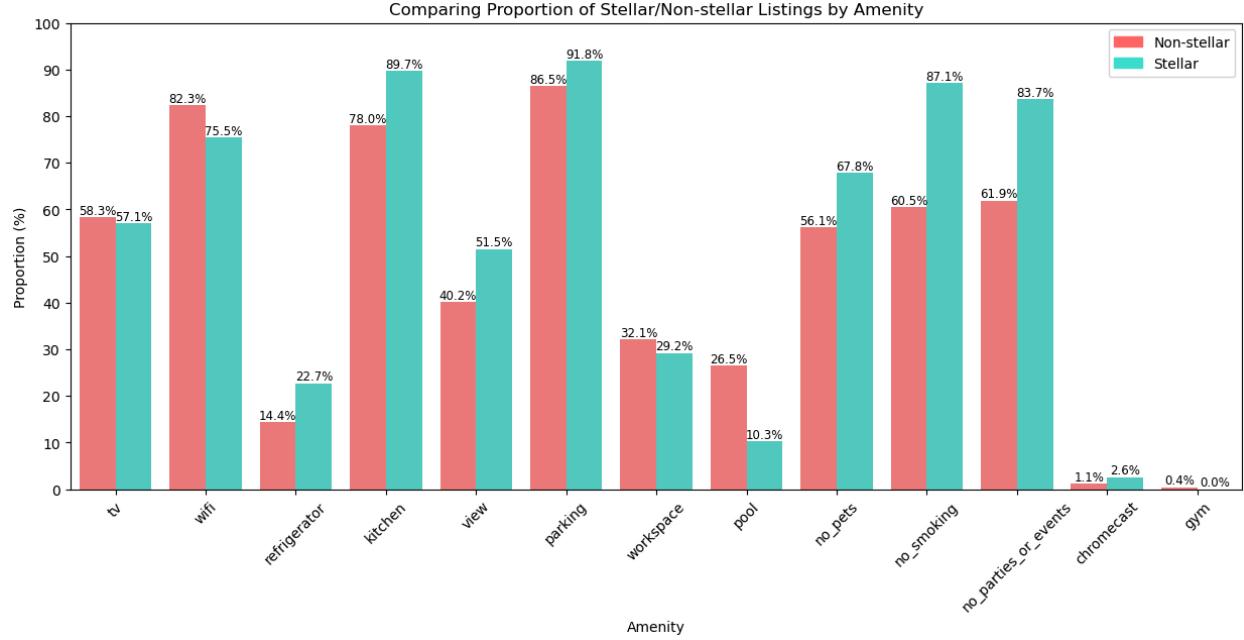


Figure 6: Proportion of Stellar and Non-Stellar Listings by Amenity

We observe that stellar listings tend to have specific amenities such as house rules, a view, parking, a kitchen, and a refrigerator. Interestingly, WiFi and TV do not seem as essential. This could be because these listings could be in rural areas with scenic views, leading visitors to place less importance on TV and WiFi and more on the experience and natural surroundings. Additionally, stellar listings often have no-pet and no-smoking policies, suggesting that these rules contribute to a higher standard of cleanliness and appeal to guests seeking a more controlled environment.

Which are the countries with both a relatively high number of stellar listings, and high percentage of stellar listings?

We study the tables below, of the countries with the highest number of stellar listings, and highest percentage of stellar listings.

	Number of Stellar Listings (Percentage Stellar Listings)	Number of Non-stellar Listings
United Kingdom	60(20%)	236
Greece	18 (2%)	730
United States	18 (10 %)	158
France	16 (4%)	397
Italy	15 (2%)	713

Poland	8 (8%)	90
Norway	8 (4%)	176
Romania	7 (9%)	70
Iceland	7 (23%)	23
Japan	7 (2%)	302

Table 1: Top 10 Highest Number and Percentage of Stellar Listings

We note that the United Kingdom, United States, and Iceland have both a relatively high number and percentage of stellar listings. Even though Iceland has few listings in general, a high percentage of stellar listings might indicate that returns on investments could be higher there. Focusing on quality over quantity in certain markets could yield substantial benefits. We now look at where these listings are on the map to see if we can spot any clustering and trends.

Which locations in the United States, United Kingdom, and Iceland should Airbnb put more resources in?

We look at the three maps below with the stellar listings plotted out. Notably, we dropped four listings from the UK and one from Iceland due to insufficient coordinates.

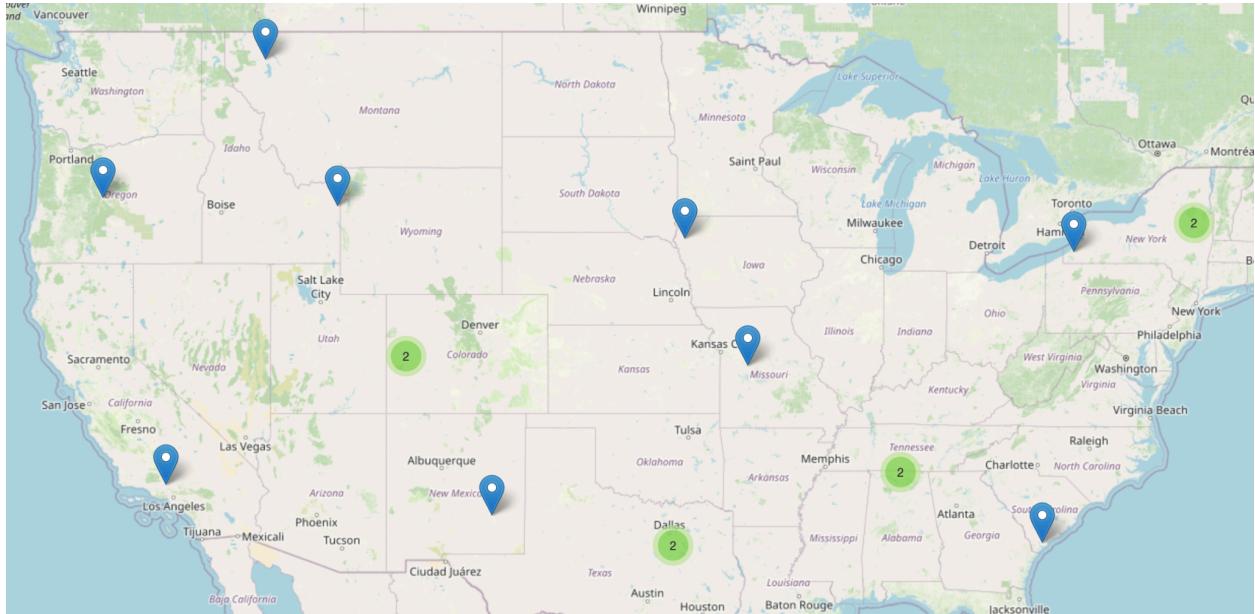


Figure 7: Distribution of Stellar Listings in the United States

We notice for the United States stellar listings, there are no clustering around major cities. This might indicate that the quality of the listing, including that these listings were mostly quieter places with a view and further away from the city, was more important than the convenience of the location.

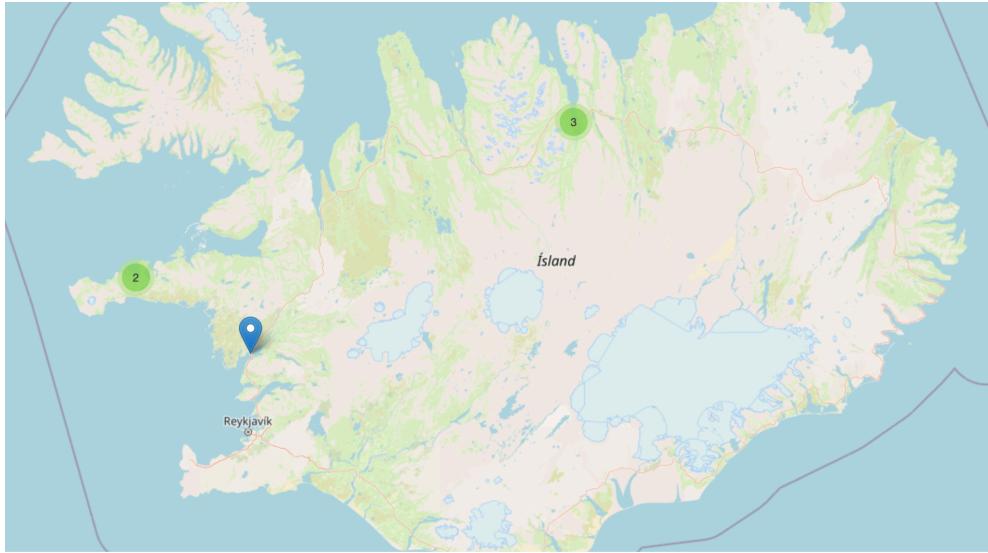


Figure 8: Distribution of Stellar Listings in Iceland

We notice for Iceland that stellar listings are concentrated around popular tourist destinations, including Reykjavik and the northern regions. Perhaps resources could be focused on enhancing the quality and offerings in these high-traffic areas to maximize guest satisfaction and encourage more bookings.



Figure 9: Distribution of Stellar Listings in the United Kingdom

We notice for the United Kingdom that these listings are spread across various locations, including Maidenwells, Wales; Cheshire East, England; Pembrokeshire, Haverfordwest; and Great Ayton, England. Many listings are shepherd huts or unique lodges located in scenic rural areas, providing a serene escape from city life.

Conclusion

This analysis of Airbnb listings provides insights into factors affecting customer ratings and demand. We identified “stellar” listings—those with high ratings and numerous reviews. A key finding is that stellar listings often feature clear and detailed house rules, which enhance guest satisfaction by setting clear expectations. Airbnb could encourage hosts to adopt comprehensive house rules by showcasing their benefits through success stories and data, providing templates, or offering incentives like increased search visibility.

Our analysis shows that the United Kingdom, United States, and Iceland have a high number and percentage of stellar listings. Investing in these specific regions by enhancing listing quality and offerings could be highly beneficial. Overall, it seems that the stellar listings in the United States, United Kingdom, and Iceland are mostly in quiet escapes from large cities. Perhaps more resources could be put into such types of listings.

Countries like India and Italy, with many listings but fewer stellar ones, offer significant growth opportunities. By adopting strategies used in the UK, US, and Iceland—such as implementing clear house rules and improving overall listing quality—Airbnb can enhance listings in these high-growth markets. This approach can help capture a larger market share and boost guest satisfaction.

References

- Reddy, Sathya Narayana (2023) How Airbnb Became a Global Phenomenon: The Growth Story You Need to Know. -
https://medium.com/@satya_/how-airbnb-became-a-global-phenomenon-the-growth-story-you-need-to-know-f5436cf5ed8#:~:text=By%202011%2C%20Airbnb%20had%20reached,guests%20and%204%20million%20listings.
- Bhutia, Peden Doma. "Is the World Ready to Meet the Indian Middle Class?" Skift, 9 January 2024,
<https://skift.com/2024/01/09/is-the-world-ready-to-meet-the-indian-middle-class-megatrends-2024/>. Accessed 4 August 2024.
- Airbnb Statistics [2024]: User & Market Growth Data