Jeneen Gnilka, Tim Mandel, Austin Vasquez

Project 1: AirBnb Analysis

University of Richmond Data Bootcamp

Written Analysis

Introduction:

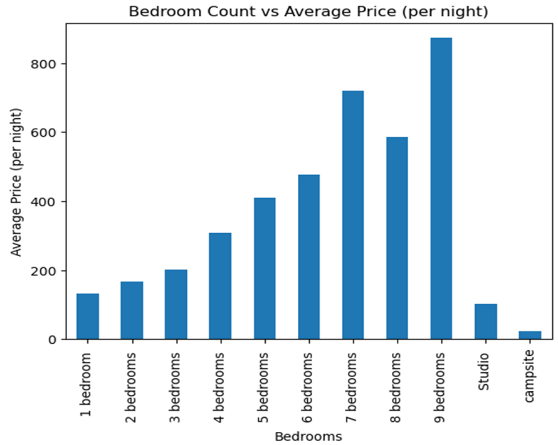
In this project we chose to analyze the Airbnb data from three different areas of the United States: Asheville, NC, Columbus, OH and Rhode Island. We gathered the data from an online source called insideairbnb.com. From there, we cleaned the data using Pandas in Jupyter Notebook. We then utilized Matplotlib in order to create visualizations for the data. We then compiled our findings into a PowerPoint which we then presented to the class. The following are our findings.

Price:

An initial analysis of AirBnbs was conducted based on the three categories: number of bedrooms, number of beds, and number of bathrooms. As expected, the average price per night became more expensive as numbers increased in each of the three categories. Interestingly, there was only a marginal increase between number of half bathrooms in relation to number of bedrooms (i.e., a two-bedroom two and a half bath property only had a slight increase in average price per night compared to a two-bedroom two-bathroom property).

A graph of bathroom and average price

Description automatically generatedA graph of blue and white lines

Description automatically generated

One other factor that was analyzed related to price was the type of rooms available. Excluding one-bedroom properties, other types of properties generally were divided into two groups: Entire home/apt or private room. This followed a trend counter to what people would initially hypothesize (Private room would be thought to be cheaper than an entire property). The table above breaks down the different pricing for the different room types grouped by number of bedrooms.

A screenshot of a room

Description automatically generated

Reviews by Room Type:

After analyzing pricing of AirBnb properties next came the analysis related to reviews. There are 6 different weighted categories for ratings that reviews that produce a final review score presented on a property’s listing. The categories are: review scores accuracy, review scores cleanliness, review scores value, review scores communication, review scores check-in, and review scores location that are combined to create the overall review scores rating. Reviews were grouped by number of bedrooms and type of room. Average review scores of the different categories followed a similar pattern between all data sets (number of rooms and type of rooms). One interesting point to note is how the “value” category for private rooms vs entire home was scored lower for private rooms than entire homes/apartments. This could be related to only a minimal price difference between the two types of rooms. Customers might believe they receive a “better bang for your buck” when renting an entire property instead of a private A graph of different colored lines

Description automatically generatedroom.

A screenshot of a graph

Description automatically generated

Reviews By Host Size:

The last point of analysis was comparing hosts based on the total number of properties listed on AirBnb. Hosts were split into three groups: Hosts with one listing, hosts that have between 2-10 properties, and hosts that have greater than 10 properties. At initial glance, there appears to be differences in the average score ratings between the three groups. Scores follow the trend of the larger the more listings a host has, the lower overall rating is scored. One could hypothesize that this is due to the extra care single listing hosts can give than bigger hosts whom would have to take care of multiple properties.

A screenshot of a test

Description automatically generated

A screenshot of a computer

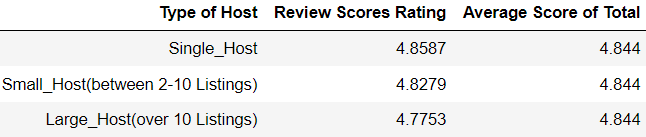
Description automatically generatedA screenshot of a computer

Description automatically generated

Which of the three areas has the best market for Airbnbs:

This is the second question that was asked in our original project proposal. A few different factors go into answering this question. The first is the amount of Airbnbs available. Rhode Island had the most with ~5,300. Asheville had the second most with ~3,300 and Columbus the least with ~2,500. The next factor, average price of an Airbnb followed a similar pattern. Rhode Island was the most expensive at $345 per night on average. Then came Asheville at $177 per night and finally Columbus with an average price of $154. We found that the third factor, average rating, was not statistically significant. Therefore we looked outside the data to find another factor: the population of each area. Rhode Island and Columbus both have a population of around 1 million people. However, Asheville’s population is only about 94,000. Given the small population, the lower end of average price, and having more than 3,000 listings, we determined that Asheville, North Carolina has the best market for Airbnbs.

Statistical Analysis:



A screenshot of a computer

Description automatically generatedA computer code with numbers

Description automatically generated



The table above shows the mean scores for all three groups of hosts for all three areas that were researched. Three individual T-tests were conducted for each group compared to the total as well as ANOVA test used to compare all four groups together. However, after conducting a Shapiro test and plotting to check for normality, it was determined that the data is in fact not normally distributed. Therefore, a Kruskal-Wallis test was used to check for statistical significance since the data is not normally distributed. The data shows an extremely low pvalue across each of the three tests, therefore the null hypothesis can be rejected and accept that these means have a statistically significant difference. The more listing a host has on Airbnb is correlated with a lower rating scores receive and should be considered when booking your next trip!