

FINAL PROJECT: INITIAL ANALYSIS



**ALY6070 - Communication and Visualization for Data Analytics**

**Final Project: Initial Analysis**

Prof. Mykhaylo Trubskyy

Report by: Group 3

Aadira Anil Ramakrishnan

Kareena Ramrakhya

Sayali Sudhir Shinde

Sakshi Singh

Isha Zade

## INTRODUCTION

Understanding and communicating complex data information is a vital skill. As part of our project, we'll be using data visualization techniques to tell stories with the data we've collected from Airbnb. We'll create dashboards, scorecards, and spatial representations to make the information easier to understand.

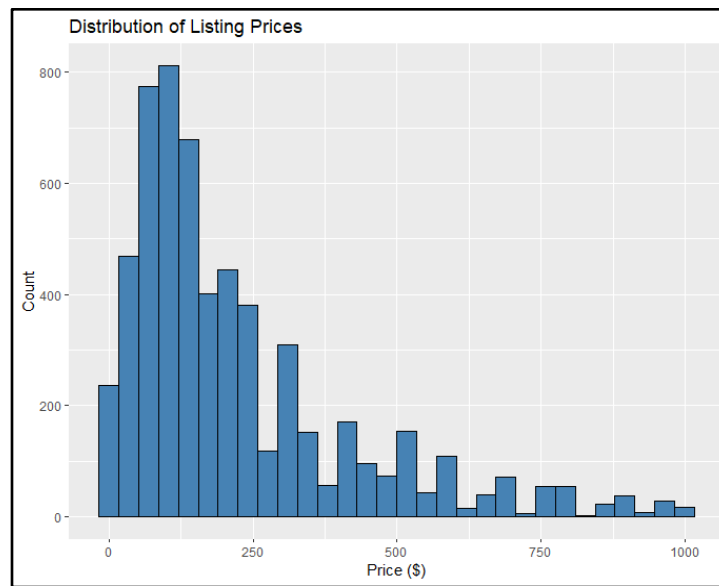
In different roles, being able to evaluate and propose the right visualizations for a specific audience is crucial. That's why we'll focus on using key design concepts to ensure our visualizations are effective for our target audience.

For this assignment, we'll start by analyzing the dataset. We'll use data to help us make informed decisions. Through our analysis and visualization efforts, we aim to provide valuable information that can benefit various stakeholders in the hospitality industry.

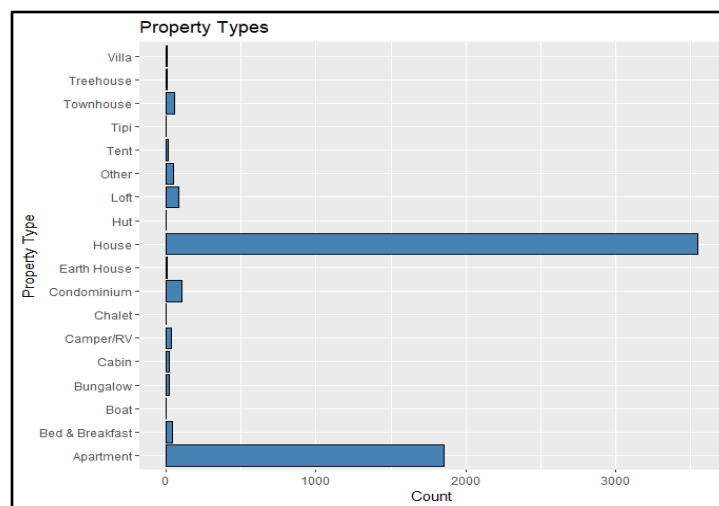
## ANALYSIS

### I. Variables in the Data

After some data exploration and analysis, we came up with a few data visualizations to understand the dataset, variables, and correlations. These are the visualizations:



Graph 1: Histogram showing the distribution of price of listing.



Graph 2: Bar graph showing the count of different property types.

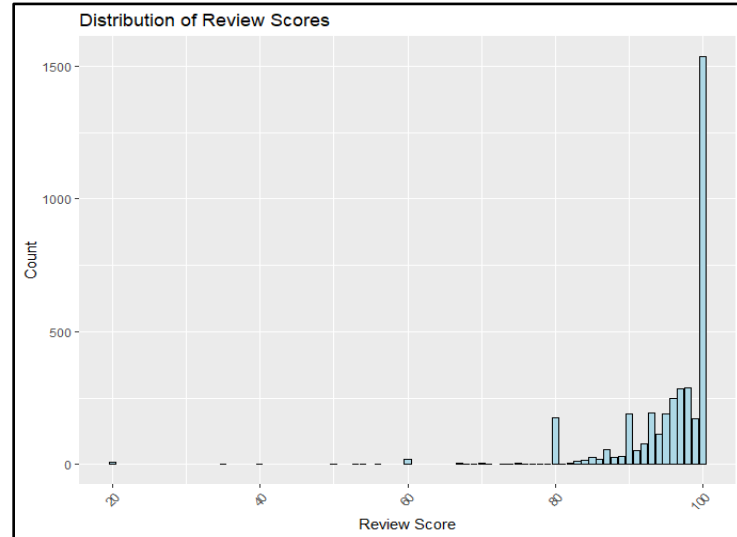


Image Bar Chart showing Count of Review Scores

Variables:

- A. From the first visualization (Histogram of Prices), we can identify "Price (\$)" as one of the variables in the dataset. This variable represents the pricing of listings.
- B. The second visualization (Bar Plot of Property Types) reveals another variable named "Property Type." This categorical variable indicates the different types of properties available in the dataset, such as houses, apartments, condos, etc.
- C. Lastly, the third visualization (Distribution of Review Scores) displays the variable "Review Score," representing the ratings given to listings by customers.

## II. Correlations, Patterns, and Trends:

- A. Correlations: While the provided visualizations don't directly show correlations between variables, we can infer potential correlations from the data. For example, we might expect a positive correlation between price and certain property features (e.g., number of bedrooms, amenities), as higher-priced listings tend to offer more amenities or have larger floor areas.
- B. Patterns: The histogram of prices shows the distribution of listing prices, which can reveal patterns such as price ranges that are more common or the presence of outliers. Similarly, the bar plot of property types illustrates the frequency of different property types, allowing us to identify which types are more prevalent in the dataset.
- C. Trends: The visualizations don't explicitly depict trends over time. However, if the dataset includes temporal information, such as listing dates or review timestamps, one could analyze trends over time, such as changes in listing prices or review scores over different time periods.

## III. Audience and their Areas of Interest

The audience for a dataset like this one will have a lot of stakeholders who can make use of this dashboard such as:

- A. Airbnb Hosts: They might be interested in understanding market trends, pricing strategies, and the features that attract guests.
- B. Travelers: Travelers could use this data to find suitable accommodations based on location, price, amenities, and availability.
- C. Real Estate Investors: Investors might analyze this data to identify profitable locations, property types, and rental trends.

- D. Urban Planners: Planners may use this data to understand the impact of short-term rentals on neighborhoods and housing markets.
- E. Tourism Boards: Tourism boards could use this data to assess the impact of Airbnb on local tourism and plan promotional strategies.
- F. Researchers: Researchers from various fields such as hospitality, economics, sociology, or urban studies might analyze this data to understand the dynamics of the sharing economy, housing markets, and tourism patterns.
- G. Regulatory Bodies: Government agencies or regulatory bodies might use this data to monitor compliance with local regulations and assess the impact of short-term rentals on communities.

Their areas of interest and the questions that can be answered with these questions are as follows:

- A. How much does it cost to stay in different areas and types of places?
- B. How quickly do hosts respond to messages?
- C. What is the 'check-in' and 'check-out' time?
- D. What do guests say in their reviews?
- E. How long do people usually stay?
- F. What makes some hosts "superhosts"?

### III. Proposed Questions for the Dashboard

Below are some key questions that can be answered through the use of data visualization.

#### A. Dynamic Pricing Optimization:

- a. What are the most influential factors in determining optimal pricing, and how do they vary across different market segments?

#### B. Host Performance and Guest Satisfaction:

- a. How do metrics such as host response time, acceptance rate, and listing cleanliness correlate with guest reviews and overall satisfaction?
- b. Are there specific host behaviors or characteristics that consistently result in higher guest ratings and positive reviews?

#### C. Market Segmentation and Targeting:

- a. Are there discernible segments within the Airbnb market based on traveler preferences, demographics, and booking behaviors?
- b. What are the key drivers of demand within each segment, and how can we position our offerings to capture market share and maximize revenue?

#### D. Spatial Analysis and Expansion Opportunities:

- a. Where are the geographic hotspots or underserved areas with high demand potential for Airbnb listings?
- b. Can we identify emerging trends or shifts in traveler preferences that may signal opportunities for expansion into new markets or neighborhoods?

#### E. Customer Lifetime Value and Loyalty:

- a. How do guest behaviors, such as repeat bookings, referral rates, and engagement with host amenities, correlate with long-term customer value?

F. Competitive Benchmarking and Market Positioning:

- a. How do our pricing, availability, and guest satisfaction metrics compare to competitors within our target markets?

*IV. The Use and Benefits of a Dashboard for Data Clarity*

A dashboard can be very useful for a dataset such as this one, which displays AirBnB details in several ways:

- A. Clarity and Organization: The data for this dataset comes from many different sources and is an aggregate of a number of related datasets. The main purpose of a dashboard is to make such an aggregate of datasets clearer and easier to interpret for everyone.
- B. Visualizing Trends: There are many stakeholders for whom this dashboard can be useful, such as the hosts. Investors, tourism companies, etc. This dashboard can display the dynamic trends related to listing prices, types, property value, and rating over time, which can be helpful for making business decisions.
- C. Ease of Implementation: Creating a dashboard like this using Tableau is not overly complicated to set up, which makes it a very efficient solution when it comes to datasets with complex inner connections.
- D. Predictive Analytics: Predictive analysis models can be implemented on the dashboard. This can help users forecast trends, patterns, and dynamics pertaining to the area of interest. This can help them make changes to aspects like policies, pricing, etc. accordingly.
- E. Risk Management: This dashboard can help users identify potential risks that can affect their strategies and business, such as negative reviews, regulatory issues, compliance issues, geographical issues, etc.



- F. Customer Segmentation: It can be useful for segmenting customers according to different areas of interest, such as demographics, past preferences, booking behaviors, price range, etc. This information can be useful for several business decisions, such as determining the most efficient marketing strategy while cutting costs.
- G. Geospatial Analysis: Maps can be incorporated into the dashboard to create dynamic visualizations such as heat maps and graphs that can help highlight hotspots, expansion opportunities, low opportunity area, growing popularity, etc. This information can help identify demand and potential in order to boost business.
- H. Comparative Analysis: The dashboard will have dynamic and interactive features that can help display and compare metrics like popularity, pricing, demand, trends, guest satisfaction, etc., which can help hosts analyze their performance compared to competitors in the area. It can help them understand the areas they need to work on or opportunities for making their listing more attractive.

#### *V. Types of Data Visualizations for Communicating the Information Clearly*

- A. What are the most influential factors in determining optimal pricing, and how do they vary across different market segments?
  - a. Multivariable scatter plot: For every market segment, this kind of graph could display the relationship between price and various factors. Pricing and other significant factors could be represented by the axes, and each segment could be color-coded.
- B. How do metrics such as host response time, acceptance rate, and listing cleanliness correlate with guest reviews and overall satisfaction?

- a. Correlation Matrix: The correlation coefficient between each metric and overall satisfaction or guest reviews can be shown in a matrix.
- C. Are there specific host behaviors or characteristics that consistently result in higher guest ratings and positive reviews?
  - a. Box Plot: The distribution of guest ratings for various host traits or behaviors could be displayed using this kind of graph. A box plot representing each category of behavior or attribute would make it simple to identify which categories typically have higher median ratings and how evenly distributed the ratings are.
- D. Are there discernible segments within the Airbnb market based on traveler preferences, demographics, and booking behaviors?
  - a. Stacked Bar Chart: A stacked bar chart could display the percentage of each variable within segments to represent multiple variables, such as traveler preferences and demographics.
- E. What are the key drivers of demand within each segment, and how can we position our offerings to capture market share and maximize revenue?
  - a. Heat Map: This could display how much each driver of demand contributes to the total demand in each market. Every driver would be a segment of the bar, color-coded for clarity, and each segment would have its own bar.
- F. Where are the geographic hotspots or underserved areas with high demand potential for Airbnb listings?
  - a. Choropleth Map: Areas are shaded on this kind of map according to a metric, like the possible demand for Airbnb listings. You could quickly determine which areas are underserved and which are hotspots thanks to it.

- G. Can we identify emerging trends or shifts in traveler preferences that may signal opportunities for expansion into new markets or neighborhoods?
- a. Line graphs: These can display patterns for different traveler preferences over time. Several lines could be used to represent various neighborhoods or market niches.
- H. How do guest behaviors, such as repeat bookings, referral rates, and engagement with host amenities, correlate with long-term customer value?
- a. Scatter Plot: You can plot each behavior (repeat bookings, referral rates, and engagement with amenities) against the long-term customer value of the data using a scatter plot with a regression line. This would make it easier to see how each behavior and customer value relate to one another, both in strength and direction.
- I. How do our pricing, availability, and guest satisfaction metrics compare to competitors within our target markets?
- a. Bar Chart: Each metric could be displayed as a set of bars for each competitor side by side in a grouped bar chart. This makes it possible to compare each metric directly between rivals.

## CONCLUSION

To sum up, our examination of the Airbnb dataset has provided valuable insight on the workings of the business as well as the potential of the dashboard. We laid the foundation for further analysis by identifying important variables through data visualization methods, such as listing pricing, property kinds, and guest reviews. We were able to deduce probable links that could influence market dynamics and pricing tactics. This lays a good foundation for us to work on for future milestones.

In order to meet to the different needs of the stakeholders, we recommended questions for the dashboard that could help in well-informed decision-making. These inquiries, which range from client segmentation to dynamic pricing optimization, demonstrate the complexity of the various dynamic variables involved. Through the use of this dashboard, stakeholders can gain access to actionable insights and confidently and clearly navigate the complexities of such a dynamic business.

The benefits and usefulness of a dashboard is based on its ability to improve data clarity, simplify predictive analytics, and support comparative analysis. By employing techniques such as trend visualization, hotspot identification, and competitive benchmark evaluation, stakeholders can make well-informed strategic decisions that accelerate business growth and elevate guest satisfaction. Our analysis essentially highlights how data visualization can significantly impact how the industry and its various stakeholders develop in the future.

## REFERENCES

1. Tufte, Edward R. *The Visual Display of Quantitative Information*. Graphics Press, 2001.
2. Knaflic, Cole Nussbaumer. *Storytelling with Data: A Data Visualization Guide for Business Professionals*. Wiley, 2015.
3. Yau, Nathan. "FlowingData." [Blog]. Available at: <https://flowingdata.com/>
4. Datawrapper Blog. [Blog]. Available at: <https://blog.datawrapper.de/>
5. Yau, Nathan. *Data Points: Visualization That Means Something*. Wiley, 2013.
6. What is a dashboard? A complete overview. (n.d.). Tableau.  
<https://www.tableau.com/learn/articles/dashboards/what-is>
7. Pedamkar, P. (2023, March 21). What is Tableau Dashboard? EDUCBA.  
<https://www.educba.com/what-is-tableau-dashboard/>

## APPENDIX

```
1 # necessary libraries
2 library(tidyverse)
3 library(ggplot2)
4
5 # Loading data into data frame
6 df <- read.csv("listings.csv")
7
8 # Checking dimensions
9 print(paste("Shape of the data:", dim(df)))
10
11 # Checking data types
12 print("Data types:")
13 print(str(df))
14
15 # Checking for missing values
16 print("Missing values per column:")
17 print(colSums(is.na(df)))
18
19 # Basic summary statistics
20 print("Summary statistics:")
21 print(summary(df))
22
23 # Load required libraries
24 library(ggplot2)
25 library(dplyr)
26 library(stringr)
27
28
29 #data viz
30
31 # Histogram of prices
32 ggplot(listings, aes(x = price)) +
33   geom_histogram(bins = 30, fill = "steelblue", color = "black") +
34   labs(title = "Distribution of Listing Prices", x = "Price ($)", y = "Count")
35
36 # Bar chart of property types
37 ggplot(property_types, aes(x = count, y = property_type)) +
38   geom_bar(stat = "identity", fill = "steelblue", color = "black") +
39   labs(title = "Property Types", x = "Count", y = "Property Type") +
40   theme(axis.text.y = element_text(angle = 0, hjust = 1))
41
42 # Bar
43 ggplot(df, aes(x = review_scores_rating)) +
44   geom_bar(fill = "lightblue", color = "black") +
45   labs(title = "Distribution of Review Scores", x = "Review Score", y = "Count") +
46   theme(axis.text.x = element_text(angle = 45, hjust = 1))
47
```