



**CAL STATE
EAST BAY**

Airbnb Listings Analysis

by

Niharika Patil (md6339)
Pratiksha Gadhe (ln2977)
Yashi Agarwal (sy5386)

A project paper submitted in partial satisfaction of requirement for the degree of

Master of Science
in
Statistics
With
Data Science concentration
in the

Graduate Division
of the

California State University, East Bay

Professor Incharge
Prof. Eric Fox
Fall 2023

1. Introduction

Our dataset consists of Airbnb Listings in Texas, and our objectives are as follows :

- a. Explore the key factors influencing Airbnb pricing, including the impact of room type, type of houses,location, and property characteristics.
- b. Investigate the variation in pricing across different cities to identify the factors that lead to price differences in diverse Airbnb markets, Top hosts based on ratings, Popular house type in dataset.

2. Data Exploration

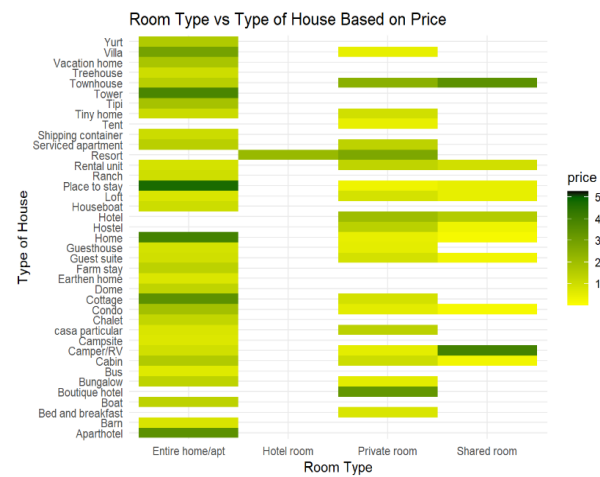
- a. Data Source: The analysis uses Kaggle dataset with 14,861 observations and 18 variables. The key variables of interest are Price, Location, House Type, Hosts, Bedrooms, Baths, Ratings, and Review Date.
- b. Data Pre-processing: As a preliminary step, missing values were addressed, and outliers were removed through data cleaning. New variables were introduced to capture location, house type, ratings, bedrooms, and baths for specific neighborhoods from string data. The dataset now comprises 13,246 observations and 21 variables after cleaning.

c. Exploratory Data Analysis: We calculated summary statistics and utilized data visualization techniques like histograms, scatter plots, heat maps, and bar plots to identify patterns, outliers, and relationships in the dataset. Comparative analysis, such as neighborhood vs. price and room type vs. price, helped assess the impact of categorical variables on pricing. We calculated top hosts based on ratings and popular week day for airbnb to get a review.

3. Key Results

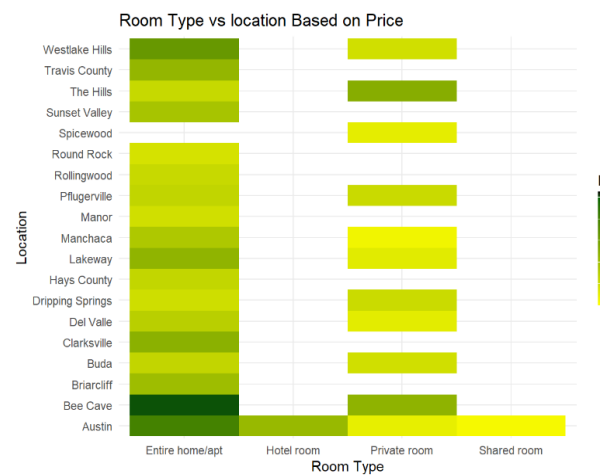
a. Does the room type (e.g., private room, entire home) based on house type influence the pricing of Airbnb listings?

There are 39 types of houses, 4 room types. Following diagram shows prices based on types of houses and room types.



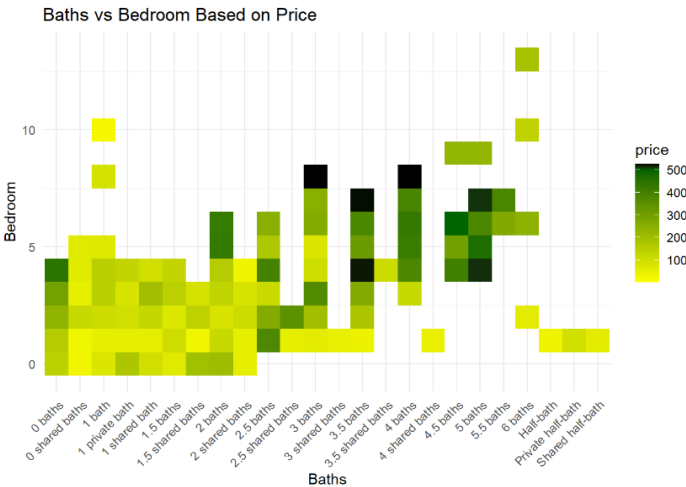
Different room types have varying price trends. "Entire home/apt" listings are generally pricier, while "Shared room" options are budget-friendly and consistent across house types. Within "Entire home/apt," "Place to stay" is the costliest, followed by "Home," "Aparthotel," and "Bus," with "Barn" and "Casa Particular" being the inexpensive. In "Hotel rooms," "resort" prices are moderate. In "Private room," "bungalows" are pricier, while "home" and "guesthouse" are more affordable. Among "Shared room" listings, "Camper/RV" and "Townhouses" are costlier, while the others remain budget-friendly.

b. Is the location a significant factor in determining the price of an Airbnb listing?



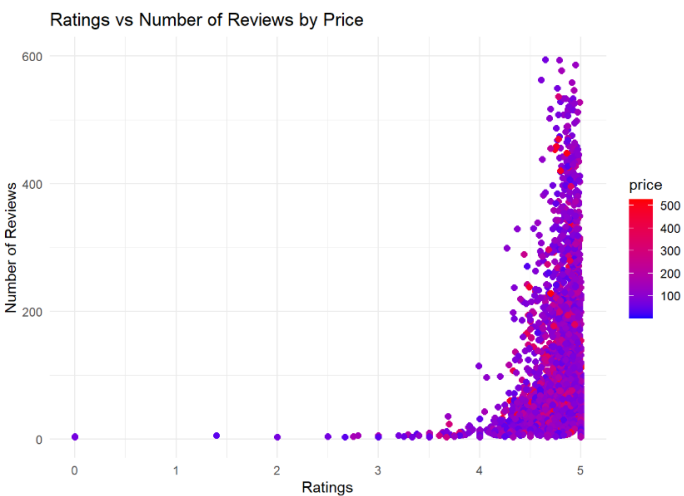
There are 21 locations in Texas, particularly the Austin area. Room Type and Location Impact: "Entire home/apartment" listings consistently command higher prices in all locations, while "Shared room" options remain budget-friendly. Among locations, Bee Cave boasts the highest prices, followed by Austin and WestLake Hills. However, Austin stands out as budget-friendly for shared room bookings, with Manchaca location also offering cost-effective shared room options.

c. Do the number of bedrooms and beds affect the price of an Airbnb listing?
There are 36 types of bathrooms and bedrooms ranging from 0 to 13.



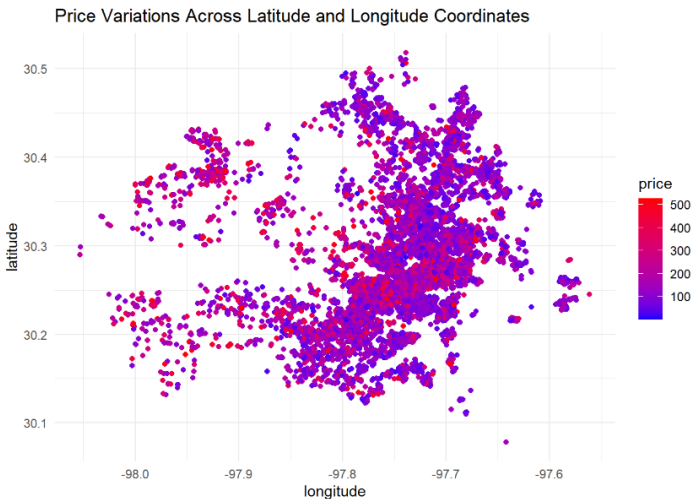
The price is higher for bathrooms in range of 3 to 5.5 and bedrooms in range 4 to 8. And Rest of the types of bedrooms and baths have lesser prices.

d. How do ratings and reviews impact the pricing of Airbnb listings?
Following graph shows the price based on number reviews and ratings. Ratings are given out of 5.



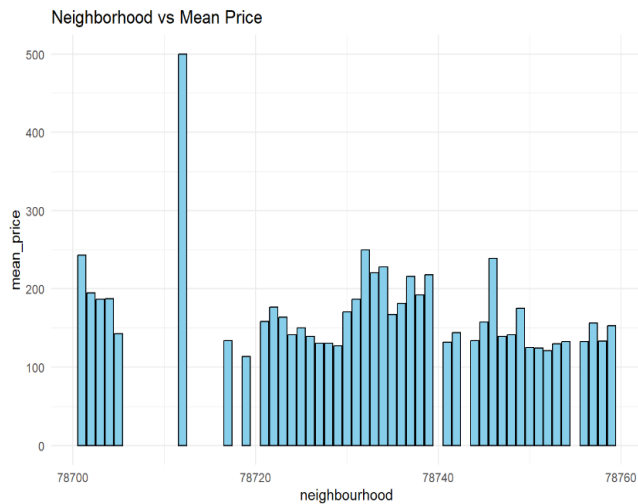
The scatter plot reveals a scattered distribution of data points, suggesting a lack of strong, linear correlation between ratings, the number of reviews, and prices. Listings with different ratings and review counts are found across a range of price points. Outliers may be present, characterized by high prices but varying ratings and review counts, potentially indicating unique, premium listings. While not displaying a clear linear trend, some loose clusters or patterns are noticeable. For instance, highly-rated listings tend to concentrate in a specific price range, and lower-rated listings are often more budget-friendly, although these clusters are not very distinct.

e. Does Latitude and Longitude also have an effect on Price?



From the above plot we can see that there is no strong correlation of latitude, Longitude with prices. Although we can see that at longitudes greater than -97.8 there are some lesser price Airbnb's along with some higher price ones.

f. Is the neighborhood a significant factor in determining the price of an Airbnb listing?



Analyzing Airbnb prices across neighborhoods, 78712 stands out with the highest mean price of \$500, while 78719 offers the lowest mean price at \$113.84. Each bar on the x-axis corresponds to a different neighborhood, representing the average price visitors can expect when booking accommodations. This analysis highlights the varying price ranges, with some neighborhoods offering expensive listings and others providing more affordable options.

g. Who are the top hosts in the dataset? Which region are they located in? How many ratings do they have?

host_name <chr>	Ratings <dbl>	City <chr>	Type of House <chr>
Molly	5	Austin	Home
Billy	5	Austin	Bungalow
Belinda	5	Austin	Home
Rachel	5	Austin	Home
Tracy	5	Austin	Loft
Eric	5	Austin	Home
Janet	5	Austin	Home
Julie	5	Austin	Home
Teresa	5	Austin	Home
Mike	5	Austin	Home

The 10 hosts are displayed in the above figure. They are based in the Austin region and have 5/5 ratings.

h. What is the most popular house type in this dataset?

There are total 39 types of houses out of which “Home” is the most popular with 4228 ratings, followed by “Rental Units” with 1791 ratings and “Condos” with 918 ratings.

4. Conclusion

Room type, house type, location, neighborhood, number of bedrooms and bathroom types are significant determinants of Airbnb listing prices in Texas.

Our analysis suggests that ratings and reviews have some influence on pricing, although the correlation is not strong.

Future Aspects- Research can be done by expanding the dataset to include more cities and longer time frames.

5. Reference

- 1.) R for Data Science by Hadley Wickham and Garret Grolemund <https://r4ds.had.co.nz>
- 2.) Dataset from Kaggle <https://www.kaggle.com/datasets/deeplearner09/airbnb-listing>

6. Code Link

Github Link : https://github.com/Niharika-Patil9/Airbnb_EDA