**NEW YORK CITY**

**AIRBNB GUEST GUIDE**

**Shiny App Development**

Rachel Fagan

[fagan3@purdue.edu](mailto:fagan3@purdue.edu)

Mu-Hua Hsu

[hsu269@purdue.edu](mailto:hsu269@purdue.edu)

Yi-Hsuan Hsu

[hsu274@purdue.edu](mailto:hsu274@purdue.edu)

Yen-Tsz Huang

[huan1627@purdue.edu](mailto:huan1627@purdue.edu)

Gadhiya, Harsh Manoj

[hgadhiya@purdue.edu](mailto:hgadhiya@purdue.edu)

**Abstract:** Our project aims to simplify and aid the searching process for Airbnb guests through our website. Guests can obtain housing and location information by putting simple inputs into our app and clicking one button. It may be a daunting task for travelers to target the most ideal place to stay via Airbnb, as it provides so many listing options. It is time-consuming for the guests to scan through miscellaneous listings, and collect fragmented information from the Airbnb website. We determined the vital elements that guests will consider—for example, listing locations, price, number of people that a location can accommodate, and amenities— and then leveraged appropriate visualization approaches to display the relationship between these elements to provide comprehensive information to guests. Furthermore, we conducted a price prediction to provide an additional decision-making reference.

**Link to App:** <https://rachelfagan1.shinyapps.io/final/>

**Business Problem Definition:** The fact that there are many listings on Airbnb makes it even harder for the potential guests to find the most ideal place to stay, as they have no easy way to view comprehensive information about listings. Based on the current design of the Airbnb website, it may be challenging for guests to acquire the right information to make optimal staying decisions. This lack may also cause negative impacts on the hosts and the enterprise Airbnb itself. For the hosts, if guests are not able to find the right place to stay, some listings will have fewer chances to be booked by guests. The loss of guests for hosts is also a loss for Airbnb’s profit.

The problem is amendable with an R Shiny app because the current data available on Airbnb is vast. We decided to develop an app that can generate comprehensive insights on one site. However, we are working under one major constraint—it is difficult to find detailed data about Airbnb listings from past years. We use data from 2020 and 2021 in our analysis. This may be indirectly beneficial; we suspect Airbnb data is very much affected by the pandemic. With help from our tool, guests may be more likely to use Airbnb, and Airbnb’s revenue and hosts’ earnings would each consequently increase, causing a win-win situation for all stakeholders. There are recommendations outside of Airbnb online on how to find the right listing (such as on one blog, *simplysaratravel*), which demonstrates that our main stakeholders (the guests) need tips and information in order to make better staying decisions.

**Analytics Problem:** We would like to identify the relationships between different elements that are vital for the guests to decide where to stay in New York City by *descriptive analysis*, so guests can make more appropriate selection criteria combinations. We also would like to conduct a *predictive analysis on the price* to enable the guests to make their decisions with insightful price information that can serve as a future estimate.

We enable the users of our App to input several elements that reflect their traveling needs, including neighborhoods, number of guests, room types, and amenities, and generate visuals with different aspects of listing information and pricing being driven based on these inputs. In the descriptive analysis, we assumed that price and rating scores are closely related and the relationship could be shown via scatter plots. In the predictive analysis, we assumed price is influenced by the number of guests, room type, neighborhood, and amenities. A successful descriptive analysis generates clear and insightful visuals that help guests formulate better selection criteria and decisions, and a successful predictive analysis results in an accurate-enough number to help the guests judge whether the price of listings on Airbnb are reasonable. Price prediction helps guests avoid the Airbnb pricing mechanism which shows higher “smart price” instead of “base price” to the guests whose travel is approaching sooner, as they desire (Airbnb Community Center).

**Data:** We acquired Airbnb datasets with (1) general information of the listings, (2) influential decision factors such as ratings, and (3) listings across time from 2020 to 2021. The first two items give us the ability to filter the proper listings for guests, and the third gives us the historical data for prediction. The dataset is sourced from Inside Airbnb. Ideally, we would have had access to more years of detailed Airbnb data, but we suspect the data trends pre-pandemic and post-pandemic would vary widely, so it may be wise that we are only mostly pandemic-time data. For data cleansing, we removed NA values where appropriate, converted columns into the correct data types for our analysis, and deleted erroneous rows of data. We also tokenized the neighborhood description data for our word cloud development. After cleaning the data, we conducted EDA and checked for multicollinearity so we could select drivers for our initial linear regression model. Then, we used the min-max method for standardizing. As for the relationships between features, all of our user inputs (neighborhood, kinds of amenities, room type, and number of guests (the “accommodates” variable) influence the listing prices.

**Methodology Selection (Descriptive and Predictive Analytics):** Descriptive: First, we used data re-shaping to merge two datasets. Then, we produced multiple kinds of graphs to visualize the relationship between factors, including maps, scatterplots, and bar charts. Second, to express the character of a particular area, we use text mining to analyze the neighborhood descriptions with a word cloud. Predictive: The price prediction is filtered by neighborhood, room type, and amenities. We used linear regression to predict the relationship between price and number of people a location can accommodate based on the user filtered data. These techniques (both descriptive and predictive) are successful because they visualize information for guests and provide guests additional insight.

R is a smart tool to use because it includes many different kinds of libraries for us to apply. Not only can we preprocess data with R, but we can also use clustering, data mining, data visualization, and various functions that we can use to build a useful model. Additionally, R Shiny apps provide users a designed, easy-to-use way to access and filter our efforts.

**Model Building:** We created an initial linear regression model, and the R-squared was low. We adjusted the model by creating dummy variables, changing binary dummies into factor variables, and standardizing our predictors. When we integrated the linear regression model into our app, we had to take a different approach. We filtered the data by user input, and then used that filter data in our model, which evaluated “price” against “accommodates” using train() in the caret library. We found that the output of our model predicted reasonable prices despite a low R-squared (for example, the predicted price increased as more accommodation was needed and more expensive areas of NY had higher predicted prices). Our model is constrained by our limited years of data, based on limitations of user constraints, and we are assuming that choosing factors that the guest has control over (the inputs) would predict price reasonably well. We could have the guest input too many items, or the usability of our app would become poor.

**Functionality:** Our Shiny App shows listings mapped, number of listings, predicted price, average price by neighborhood mapped, scores vs. price scatter plots, bar charts, a detailed data table, and word plots, all based on user input. Guests can make informed decisions with our easy-to-understand visuals and clear predictions. We used 20+ packages, but found dplyr to be the most useful. We were able to filter based on user input with dplyr, which was critical to the design and success of our app. If we had more time, we would add a prescriptive analysis that would benefit guests. For example, guests might want to find different housing for different days of traveling so that they can minimize their costs. We also would like to improve our prediction model performance.

**Conclusions:** Our app successfully visualizes data through multiple techniques and predicts prices for guests. Our app works well and is cleanly designed. Also, we were able to learn a lot through this project about the R language, R Shiny Apps, usability, and different R packages. We feel that we have a solid project that we can display to technical and non-technical audiences.

**References:**

Airbnb Data Source: Inside Airbnb <http://insideairbnb.com/get-the-data.html>

Shiny app Code Reference: Github (Author:gl2668) <https://github.com/gl2668/airbnb_priceR>

Map by Price Code Reference: RPubs-NYC Maps (Author: Jake Hofman) <https://rpubs.com/jhofman/nycmaps>

Dynamic Plot Code Reference: Github (Author:wch) <https://gist.github.com/wch/5436415/>

Icon Source for PowerPoint Presentation: United State Icon Pack <https://www.flaticon.com/packs/united-states-4>

Free Vector <https://www.freepik.com/popular-vectors>

[My Method on How to Select the Perfect Airbnb Accommodations](http://www.simplysaratravel.com/home/airbnb):<http://www.simplysaratravel.com/home/airbnb>

Airbnb Community Center: [Guests seeing different prices - Airbnb Community (withairbnb.com)](https://community.withairbnb.com/t5/Help/Guests-seeing-different-prices/td-p/125396)