**ANALYSIS OF AIRBNB LISTINGS IN SAN FRANCISCO**

MSMI 603 - Applied Statistics in Marketing Intelligence

Group 10: Sean McDevitt, Taiyi Huang, Jayce Guan, Quan Nguyen

December 12th, 2024

**Table of Contents**

**1. Identifying Optimal Listings for Airbnb Recruitment**

1.1 Research Hypothesis ……..............................................................................

1.2 Defending the Hypothesis …………..............................................................

1.3 Approach and Methodology ………...............................................................

1.4 Result ………..................................................................................................

1.5 Recommendation ……....................................................................................

**2. Host Profiles – Should Airbnb Partner with Individuals or Companies?**

2.1 Research Hypothesis …..….............................................................................

2.2 Defending the Hypothesis …..….....................................................................

2.3 Approach and Methodology ...…….................................................................

2.4 Result ......….....................................................................................................

2.5 Recommendation …….....................................................................................

**3. Amenities That Matter – What Should Airbnb Incentivize?**

3.1 Research Hypothesis …………………...................................................................

3.2 Defending the Hypothesis…………………………………………………………

3.3 Approach and Methodology……………………………………………………….

3.4 Result……………………………………………………………………………….

3.5 Recommendation………………………………………………………………….

**4. Appendix**

**1. Identifying Optimal Listings for Airbnb Recruitment**

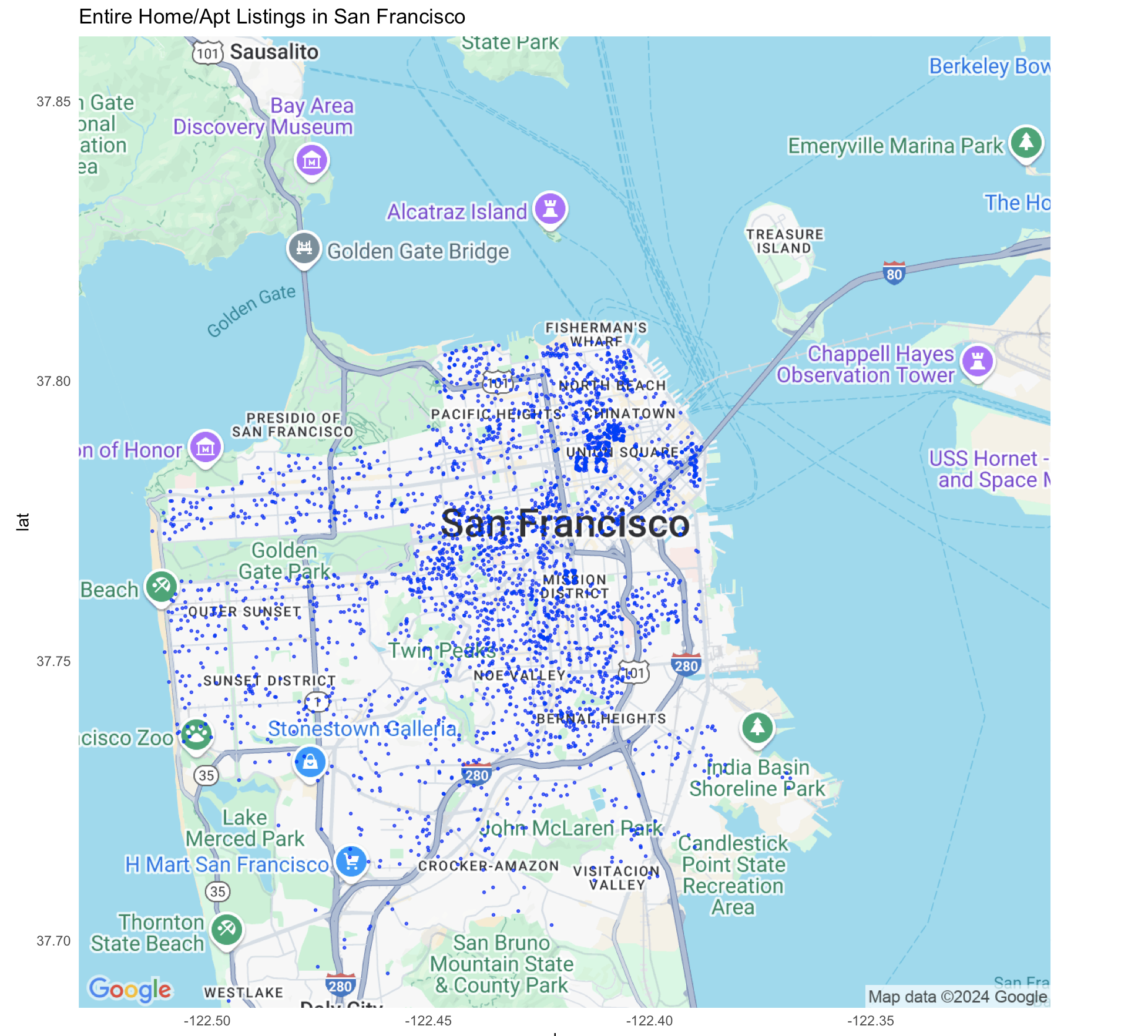
*Question 1: What types of listings should Airbnb recruit?*

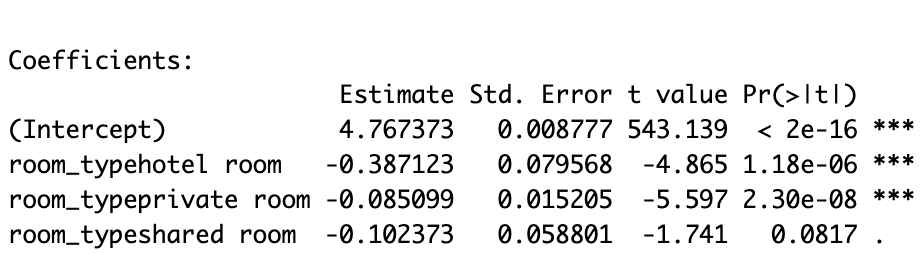
1. **Research Hypothesis**

Entire homes tend to receive higher ratings compared to sharehouses.

1. **Defending the Hypothesis:**

To evaluate the relationship between room type and ratings, we systematically analyze the data with a clear structure of variables and hypotheses. The independent variable in this analysis is Room Type, categorized into "Entire home/apt" and "Shared room," while the dependent variable is Average Ratings. This hypothesis is grounded in the assumption that privacy and comfort offered by entire homes are critical factors that influence guest satisfaction. By contrast, shared rooms often involve compromises on these aspects, which may lead to lower ratings.

1. **Approach and Methodology:** 987
2. **Results:** 



Our analysis confirms that entire homes outperform shared and private rooms in average ratings, making them the most reliable choice for host recruitment specialists. The linear model, using room type as the independent variable and average rating as the dependent variable, revealed a statistically significant positive relationship between entire homes and higher guest ratings. This indicates that guests consistently prefer the privacy, comfort, and projection provided by entire homes. As a basic data visualization, there is a basic conclusion that more people are using Entire Home in San Francisco. Apartment, as a basic insight for the data.

The data highlights that entire homes not only deliver superior guest experiences but also align with Airbnb’s goal of offering higher-quality, satisfaction-driven stays. By recruiting entire home listings, host specialists can ensure a strong track record of guest satisfaction, increasing repeat bookings and long-term customer loyalty. These findings emphasize the strategic value of prioritizing Entire Homes in recruitment efforts to maximize Airbnb’s reputation and market competitiveness.

1. **Recommendation:**

Based on the hypothesis that entire homes receive higher average ratings than shared or private rooms, we recommend prioritizing the recruitment of entire home listings. Our analysis identifies room type as a critical, independent variable influencing ratings (dependent variable), a key metric of guest satisfaction. Entire homes typically offer greater privacy, comfort, and autonomy, which align with the preferences of high-value guests and contribute to consistently higher ratings. To ensure a high standard of service, focus recruitment on entire homes with a proven track record of excellence, such as those with previous high ratings or exceptional reviews. Additionally, consider niche but popular property types, like unique cabins or modern apartments, that blend exclusivity with high performance. By recruiting hosts with entire homes and thoughtfully selected, best-quality properties, Airbnb can continue to enhance its reputation for delivering exceptional guest experiences.

**2. Host Profiles – Should Airbnb Partner with Individuals or Companies?**

*Question 2: Should Airbnb focus on individuals or companies as hosts?*

1. **Research Hypothesis:**

Listings managed by individual hosts have higher average ratings than those managed by companies, when controlling for the number of amenities provided.

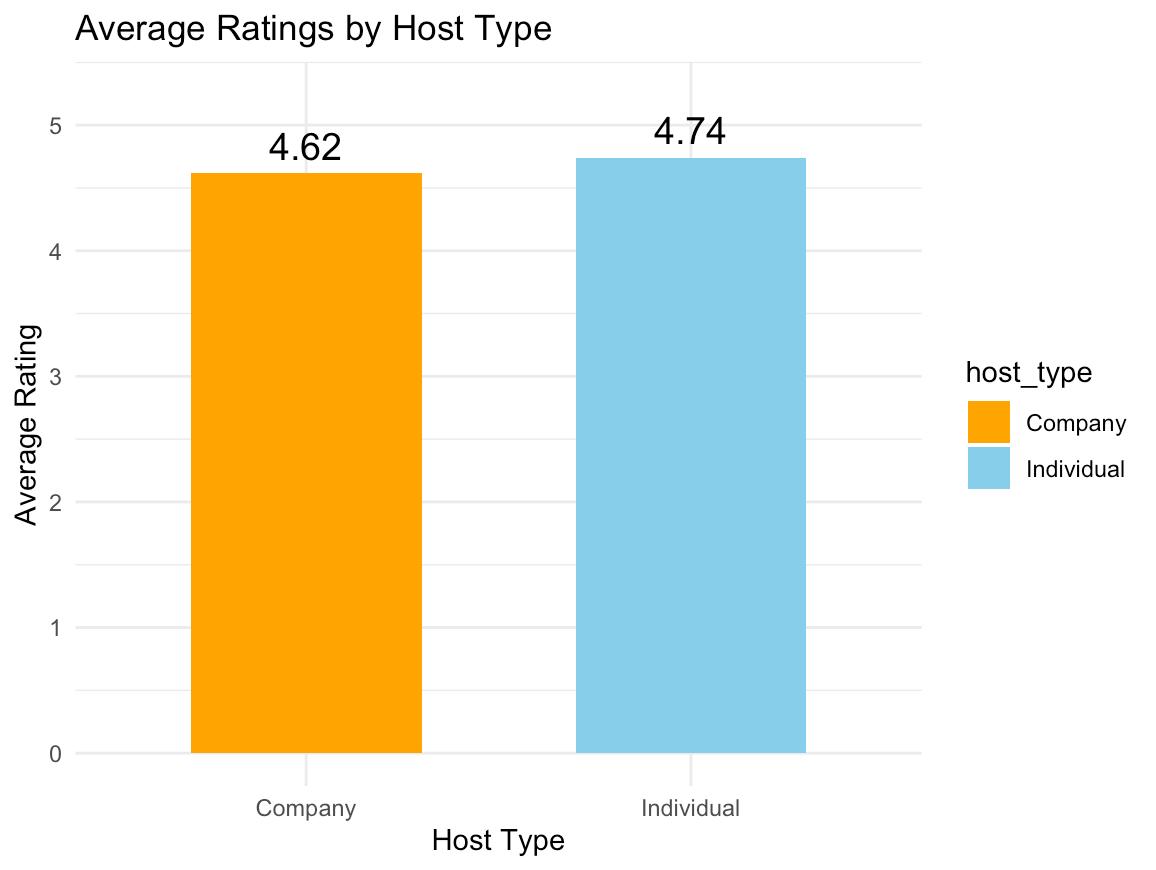
1. **Defending the Hypothesis**:

To consider that individuals have higher average ratings than companies, even controlling for the number of amenities, we must address the relationship between the average ratings and amenities. As we see the positive correlation between average ratings and the number of amenities, we can assume that the hosts who provide more amenities tend to have higher ratings than those who do not provide as many amenities. This supports our hypothesis of finding out which host type would affect the ratings most when we control for amenities.

1. **Approach and Methodology:**

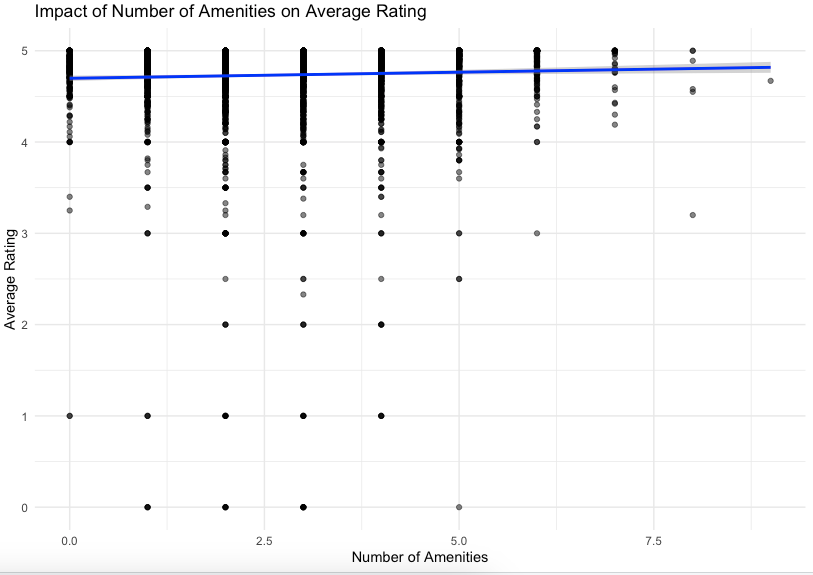
One of the first things we do is to classify our host\_type data to consider how we define company and individual within that data frame. At first, we would put the matter where hosts with more than one listing are considered the company’s listings; otherwise, it is an individual listing. But then we want to test different assumptions for classifying the different host types, so we look into the whole data frame to see how the data is distributed among tables. We recognized that there are lots of hosts that have more than five listings, so we decided to put it to the test to see how it differs from our first assumption. Following that step, we now compare the mean of the average ratings and number of amenities for both individual and company hosts; we expect to see the different distribution levels of each host type for each variable we chose. Next, we need to see the relationship between the average ratings and the number of amenities to confirm the controlling variable. Once we have all the data ready, we will run a multiple regression where the dependent variable is the average ratings, and the dependent variables are the host type and the number of amenities (which is *summary(lm(data=listings, avg\_rating ~ host\_type + number\_of\_amenities.*) With this regression, we hope to discover the host type that attracts the most customers as well as which amenities the host added to increase the average rating overall.

1. **Results:**

Under our first assumption, we categorized hosts with exactly one listing as individuals and hosts with more than one listing as companies. When we compared the mean average rating for both individuals and companies, the individual hosts received ratings 0.10 point higher than company hosts with really high confidentiality, as we can see on the graph below.

In addition to the first assumption, we felt that putting people with one listing as an individual would not be effective in determining the host type. Using the more complex assumption allows for a more precise classification of host types. In this method, we define hosts with more than five listings as companies, whereas those with five or fewer listings are classified as individuals. This guarantees that we concentrate on larger-scale activities more typical of professional hosting companies, distinguishing them from smaller, semi-professional, or individual hosts. This improved classification lowers overlap between host categories, which helps in identifying the distinct traits and actions of professional businesses. This allows us to more accurately examine differences in average ratings and number of amenities.

As I run the analysis test for the number of amenities on average rating to determine if it is a positive correlation or not. The coefficient came out 0.004, which means for every additional amenities added there will be a 0.004 increase in rating point. This result confirms that amenities significantly influence ratings, regardless of host type.



The results of the analysis highlight significant differences in the performance of Airbnb listings managed by individual hosts versus those managed by companies. Individual hosts consistently achieved higher average ratings, as reflected in the graphical comparisons, which demonstrate a higher distribution of top ratings for individual hosts. This reinforces the hypothesis that individual hosts provide a more personalized and satisfactory guest experience. Hence, the findings indicate that individual hosts do well in fulfilling visitor demands, which leads to higher ratings and more guest satisfaction. These findings highlight the strategic importance of emphasizing individual hosts in recruitment efforts to improve Airbnb's reputation and customer loyalty.

1. **Recommendation:**

Based on the findings of this study, it is evident that individual hosts on Airbnb consistently achieve higher average ratings compared to company-managed listings, even when controlling for the number of amenities provided. The personalized and tailored experiences offered by individual hosts play a significant role in enhancing guest satisfaction and driving higher ratings. These insights offer valuable guidance for Airbnb to refine its strategy and capitalize on the strengths of individual hosts while addressing opportunities for improvement among company-managed listings. Airbnb should emphasize the recruitment of individual hosts, as they are shown to excel in delivering personalized guest experiences. Providing targeted support programs, such as hosting workshops, mentorship, and resources for new individual hosts, can help further enhance their performance and maintain high customer satisfaction levels.

1. **Amenities That Matter – What Should Airbnb Incentivize?**

*Question 3: What amenities should Airbnb incentivize?*

1. **Research Hypothesis:**

Airbnb should incentivize hosts to offer high-speed Wi-Fi, fully equipped kitchens, dedicated workspaces, and unique amenities like outdoor spaces or hot tubs to cater to modern travelers' preferences for comfort, functionality, and memorable experiences.

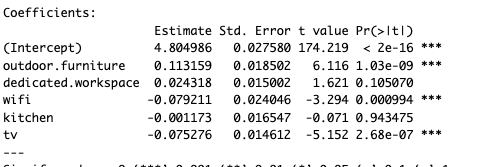
1. **Defending the Hypothesis**:

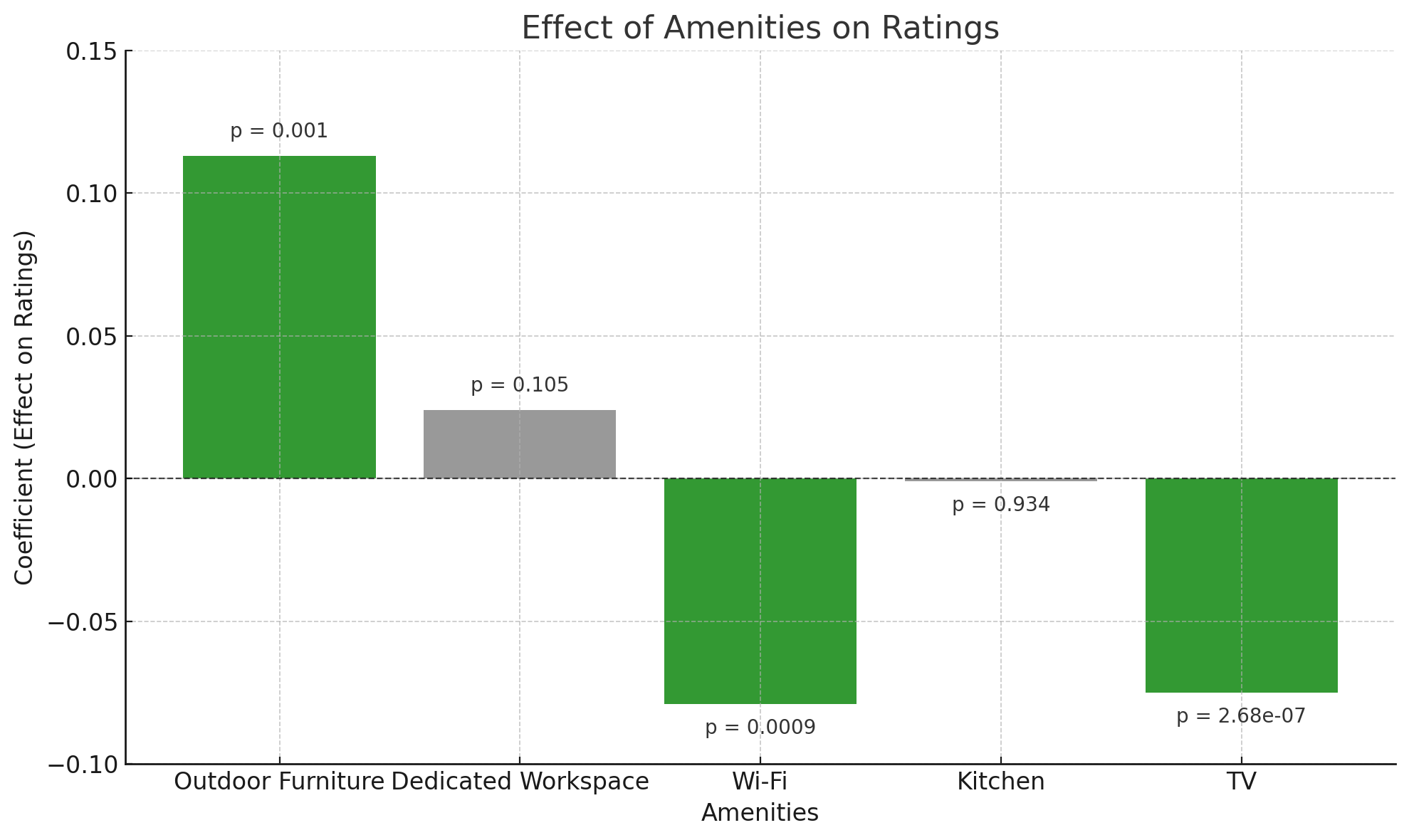
The hypothesis that Airbnb should incentivize hosts to offer high-speed Wi-Fi, fully equipped kitchens, dedicated workspaces, and unique amenities like outdoor furnitures or hot tubs is strongly supported by both data and market trends. High-speed Wi-Fi stands out as an essential amenity for both leisure and business travelers. Text analysis of guest reviews reveals over 5,654 mentions of Wi-Fi, emphasizing its critical role in meeting guest expectations. Fully equipped kitchens cater to families, long-term travelers, and budget-conscious guests who prefer to prepare meals instead of dining out. This feature is particularly appealing for guests who value convenience and cost savings during their stay. Additionally, dedicated workspaces address the needs of the increasing number of travelers who combine work and leisure. The pandemic has accelerated the adoption of remote work, creating a segment of guests who prioritize ergonomic and functional spaces for productivity. Hot tubs and outdoor furnitures provide unique and memorable experiences, which are highly valued by travelers looking for distinctive accommodations. While these features involve higher costs, their potential to command premium nightly rates and attract high-value bookings makes them a worthwhile investment.

1. **Approach and Methodology**:

One of the first things we do is compare the mean of average ratings for properties with and without these amenities. This comparison helps us identify patterns, such as whether listings with high-speed Wi-Fi or outdoor spaces consistently receive higher ratings than those without. Additionally, we analyze the total number of amenities provided by each listing to see how the overall quantity of features impacts guest satisfaction. This helps us determine whether the total number of amenities could serve as a controlling variable in subsequent analyses. Following these comparisons, we examine the relationship between average ratings and the presence of specific amenities using regression analysis. Our model includes average ratings as the dependent variable, with binary indicators for the selected amenities as the independent variables. By doing so, we aim to quantify the individual impact of each amenity on guest satisfaction while controlling for the total number of amenities.

Once the regression model is run, we analyze the results to determine which amenities contribute most significantly to higher ratings. For example, if Wi-Fi or dedicated workspaces show strong positive effects on ratings, it would suggest that these features are critical for guest satisfaction. Ultimately, this approach helps us understand the value of each amenity and provides actionable insights for Airbnb to incentivize hosts to offer features that cater to modern travelers’ preferences for comfort, functionality, and memorable experiences.



1. **Results**: 

Outdoor furniture significantly improved ratings by 0.113 points with a high confidentiality of p value is < 0.001. Dedicated workspaces had a lesser positive effect, increasing evaluations by 0.024 points, but this result was not statistically significant p value equals 0.105, implying that their appeal may be limited, catering mostly to remote workers and business travelers. In contrast, amenities such as Wi-Fi and TVs had unexpectedly negative effects on ratings (-0.079 and -0.075 points, respectively), both of which were statistically significant. Certain findings may reflect concerns with guest expectations or dissatisfaction with the quality of certain amenities. Kitchens had a minor and non-significant influence (-0.001 points), indicating that they are more of an expected baseline element than a guest-attracting factor.

1. **Recommendation**:

To increase guest satisfaction and ratings, Airbnb should emphasize paying hosts to offer outdoor furniture, which has the greatest positive influence on ratings and is highly correlated with data. Dedicated workspaces, while not meaningful for all guests, do provide value for remote workers and business travelers, indicating a potential for targeted marketing to these specialist groups. Furthermore, the unfavorable connections between Wi-Fi and TV ratings need additional examination, as these amenities may fail to exceed guest expectations in terms of quality or functionality. Though showing a negative impact on ratings, kitchen remains a baseline expectation and should continue to be included in listings. To better understand other drivers of satisfaction, further research into factors such as location, price and host interaction is recommended. By addressing these areas and enhancing transparency in listings, Airbnb can better align its offerings with guest preferences and ultimately elevate the overall guest experience.

**4. Appendix**

**Appendix A: Question 1:**

register\_google(key = "AIzaSyBi74uxncSt1uCUXsHXErWFAG1-VcLKNso")

map\_data <- SF\_Listings[!is.na(SF\_Listings$latitude) &

!is.na(SF\_Listings$longitude) &

SF\_Listings$room\_type == "shared room", ]

sf\_map <- get\_map(location = c(lon = -122.4194, lat = 37.7749),

zoom = 12, maptype = "roadmap")

ggmap(sf\_map) +

geom\_point(data = map\_data,

aes(x = longitude, y = latitude),

color = "red", size = 0.5, alpha = 0.7) +

labs(title = "Shared Room Listings in San Francisco") +

theme\_minimal()

######

map\_data <- SF\_Listings[!is.na(SF\_Listings$latitude) &

!is.na(SF\_Listings$longitude) &

SF\_Listings$room\_type == "shared room", ]

sf\_map <- get\_map(location = c(lon = -122.4194, lat = 37.7749),

zoom = 12, maptype = "roadmap")

geom\_point(data = map\_data,

aes(x = longitude, y = latitude),

color = "red", size = 0.5, alpha = 0.7) +

labs(title = "Entire Home/Apt Listings in San Francisco",

color = "Room Type") +

theme\_minimal()

######

summary(lm(data = SF\_listings, avg\_rating ~ room\_type))

**Appendix B: Question 2:**

listings <- listings %>%

mutate( host\_type = ifelse(host\_listings\_count > 5, "Company", "Individual"),

num\_amenities = sapply(strsplit(as.character(amenities), ", "), length)

##Visualization: Average Ratings by Host Type

ggplot(listings, aes(x = host\_type, y = avg\_rating, fill = host\_type)) +

geom\_boxplot() +

labs( title = "Average Ratings by Host Type ", x = "Host Type", y = "Average Rating" ) +

theme\_minimal()

ggplot(listings, aes(x = avg\_rating, fill = host\_type) +

geom\_histogram(binwidth = 0.1, alpha = 0.7, position = "identity") +

labs( title = "Distribution of Average Ratings by Host Type (Stricter Assumption)",

x = "Average Rating",

y = "Count",

fill = "Host Type" ) +

theme\_minimal()

**Appendix C: Question 3:**

summary(lm(data = SF\_listings, avg\_rating ~ outdoor.furniture + dedicated.workspace + wifi + kitchen + tv))

amenities <- data.frame(

Amenity = c("Outdoor Furniture", "Dedicated Workspace", "Wi-Fi", "Kitchen", "TV"),

Coefficient = c(0.113, 0.024, -0.079, -0.001, -0.075),

p\_value = c(0.001, 0.105, 0.0009, 0.934, 0.000000268))

# Add a significance column based on p-value

data$Significance <- ifelse(data$p\_value < 0.05, "Significant", "Not Significant")

# Create the bar chart

ggplot(data, aes(x = Amenity, y = Coefficient, fill = Significance)) +

geom\_bar(stat = "identity", position = "dodge", alpha = 0.8) +

geom\_text(aes(label = paste0("p = ", format(p\_value, digits = 3))),

vjust = ifelse(data$Coefficient > 0, -0.5, 1.5), size = 4) +

scale\_fill\_manual(values = c("Significant" = "green", "Not Significant" = "gray")) +

labs( title = "Effect of Amenities on Ratings", x = "Amenity", y = "Coefficient (Effect on Ratings)" ) +

theme\_minimal() +

theme ( plot.title = element\_text(size = 16, face = "bold"), axis.title = element\_text(size = 12))+

geom\_hline(yintercept = 0, color = "black", linetype = "dashed", size = 0.8)