AirBnB Customer Satisfaction Analysis

Professor Zoe's Favorites: Grace Hobin, Stephen Cassata, Bruce Le, Yejin Kang

Business Context & Our Goal

Business Context:

- Dataset is from Inside Airbnb from Los Angeles location
- Research will focus on on data that describes each listing's characteristics and the ratings received by customers

Project Goal:

- Understand what factors influence the rating of an Airbnb location
- Finding the differences & similarities of different listings and report the satisfaction of customers for each listing

Research Questions

1. Does the number of listings a host has, influence the rating the client will give the AirBnB?
2. Does the cost of the AirBnB correlate to a good or bad experience/review?
    - What is the top 25% (of cost) of Airbnb's average experience/review?
    - What is the bottom 25% (of cost) of Airbnb's average experience/review?
3. Are different neighborhoods in California tied to a higher or lower rating?
    - What is the average rating for different cities in California?
    - What CA neighborhood have the top 25% average rating?
4. What are the room types with the highest to lowest ratings?
    - What are the room types with the highest to lowest ratings?
    - What is the amenity with the highest rating?
5. How does, how many hosts in an area (neighborhood) influence the number of superhosts?
    - How many hosts are in each neighborhood?
    - How many superhosts are in each neighborhood?

Data Description

The Dataset is from Inside Airbnb for the Los Angeles location. "Inside AirBnB" is an independent, non-commercial set of tools and data that allows exploration of how Airbnb is really being used in cities around the world.

Source: Kaggle.com – Kaggle.com is a community of data scientists posting various collections of data for the use of researchers.

Content: The dataset contains information on the description of home/apartment, customer response, customer ratings, owner response, owner ratings, etc.

Size: This set of data is very large, containing 31,254 records across 95 different categories, such as location, response rate, description, ID, URL, etc.

Key Variables: The key variables we will be focusing on are host id, host response rate, city, price, review score rating, amenities, host is super host, etc.

Data Cleansing: Pre-processing

- Program used: Excel
- Duplicates were removed from the data test
  - Ran duplicate testing - 1 duplicate, 31,253 unique values
- Specific columns removed from data set
  - URL columns, host_thumbnaill, country_code, country, calendar_last_viewed, host_acceptance_rate, states not from CA
- Specific data columns of which N/A other invalid values are recorded/replaced
- Specific outliers/extreme data points that are removed from dataset

Data Cleansing: Transformation

Normalization:

We created two new sheets in order to aid or research. These sheets contain queries regarding AirBnB hosts and their properties. The primary key of the host sheet is host id and the primary key for the property sheet is amenities. These sheets include information on:

- Host - number of properties, average response rate, average price, average review score rating, and average review score accuracy
- Property - number of properties, average review score rating, average price, and average review value

Sample selection criteria:

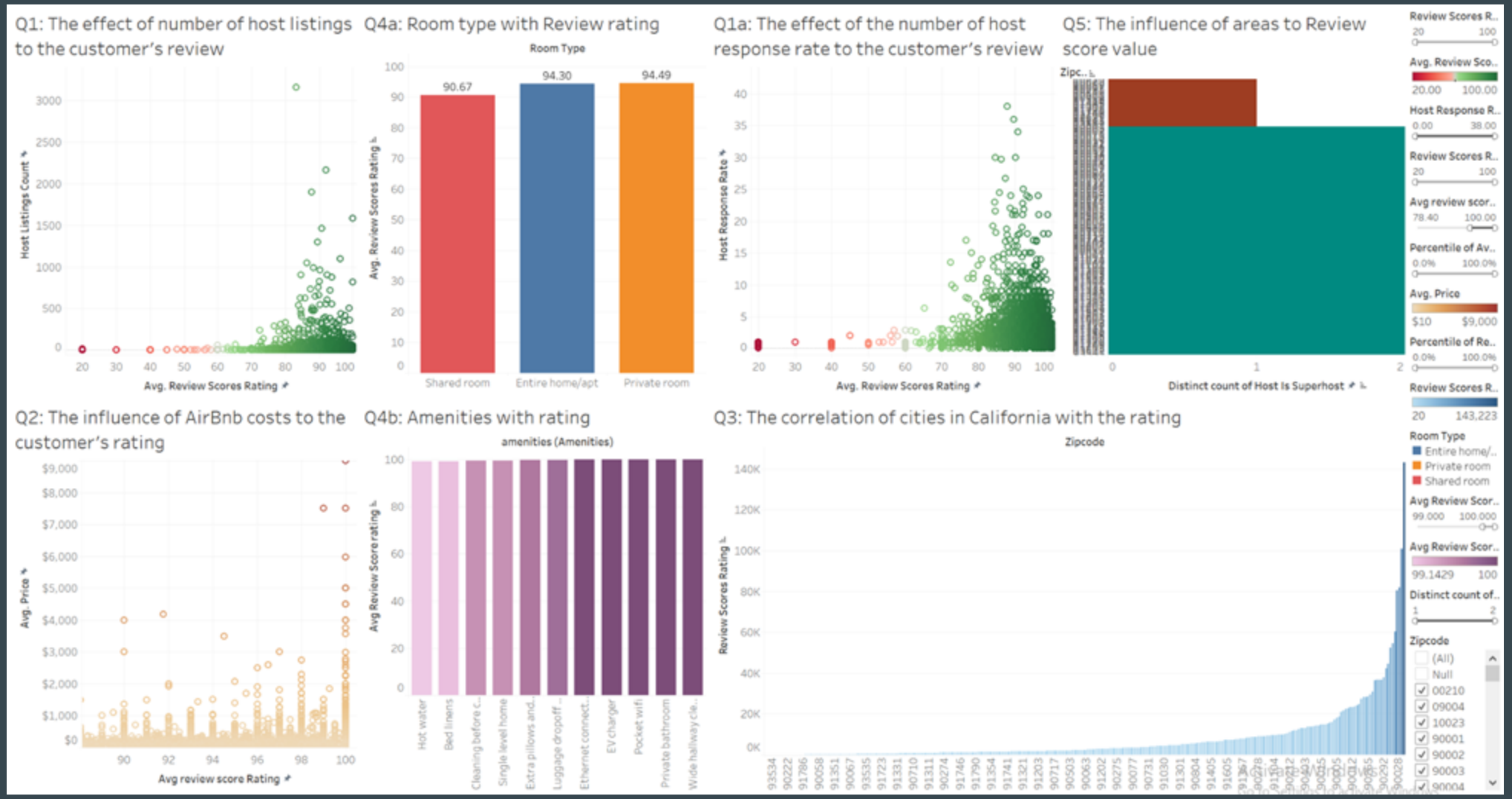
- Must be a host located in CA
- Original Size: 20,810
- Size of Cleansed: 20,807
- Sample size for queries: 93

Transformation...Continued

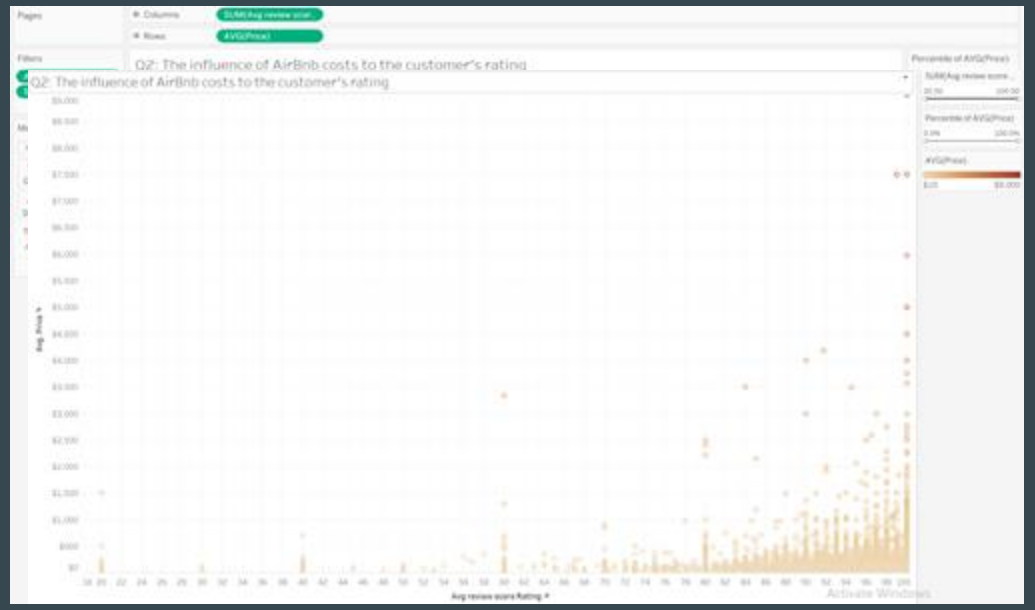
Column Changes/Creation:

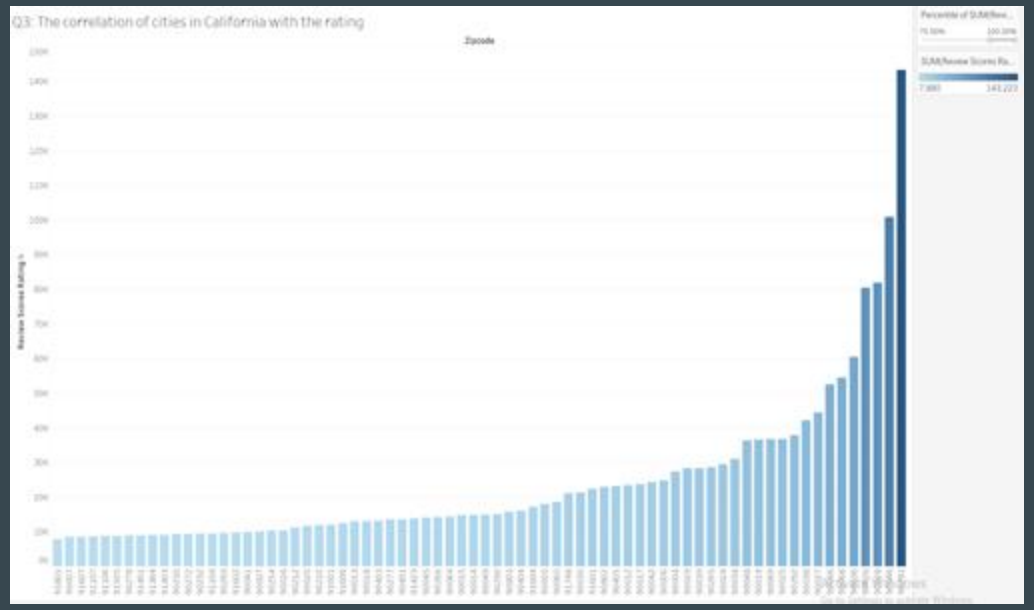
- The column host since was changed from a calendar code to an actual date
- We created a column to calculate the percentage of days out of the year a property was available for rent
  - We did this using the following equation: (365-days available)/365
- This process was then repeated for 30, 60, and 90 days.
  - We used a similar equation and replaced 365 with either 30, 60, or 90.

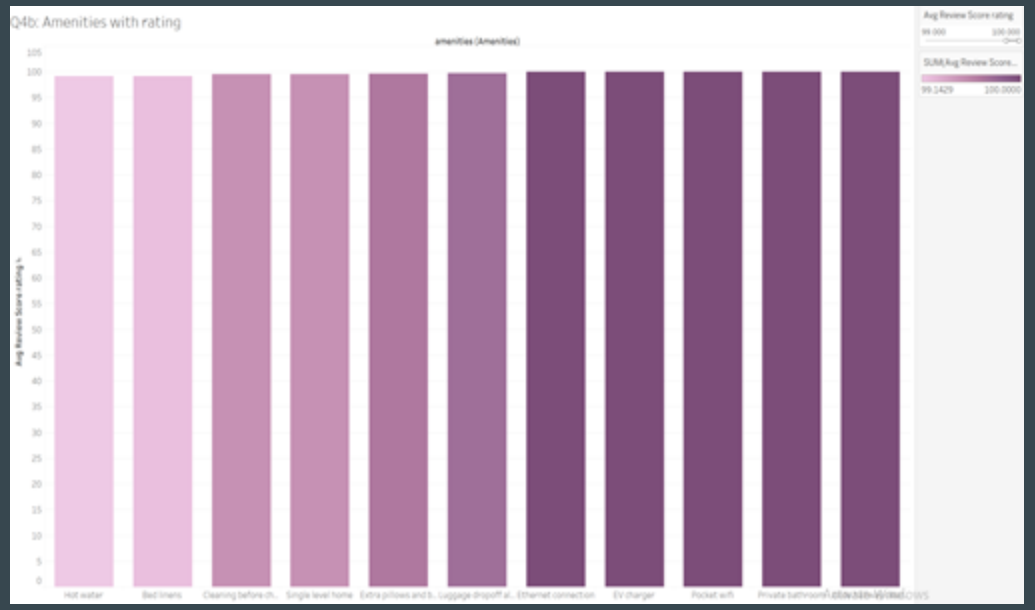
Visual Results

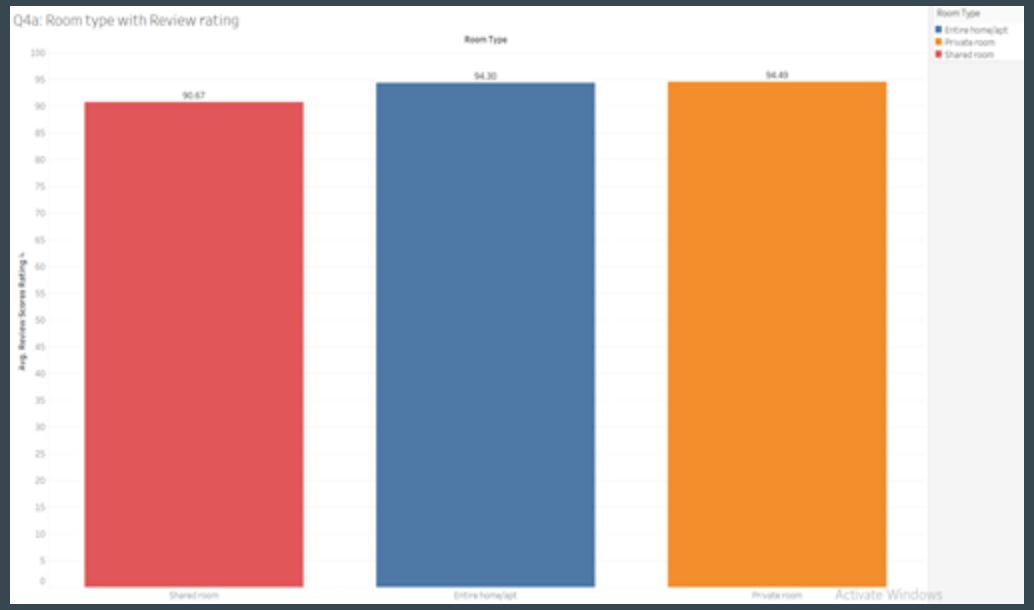


Closer look on the dataview

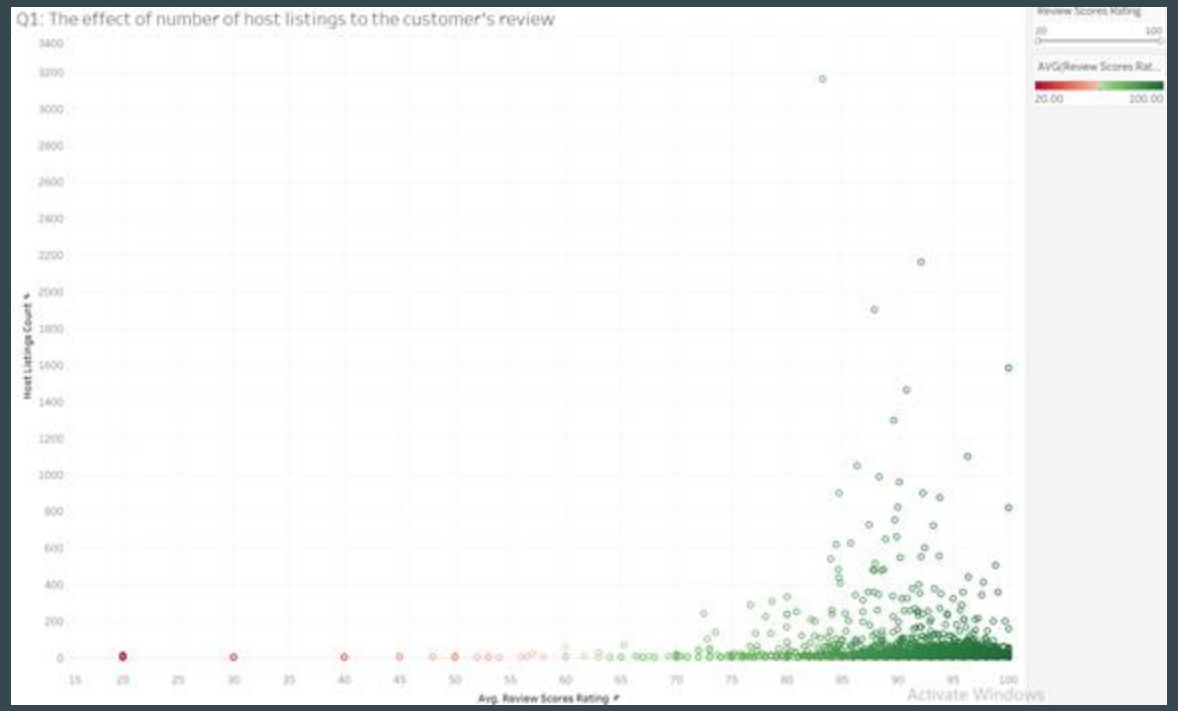


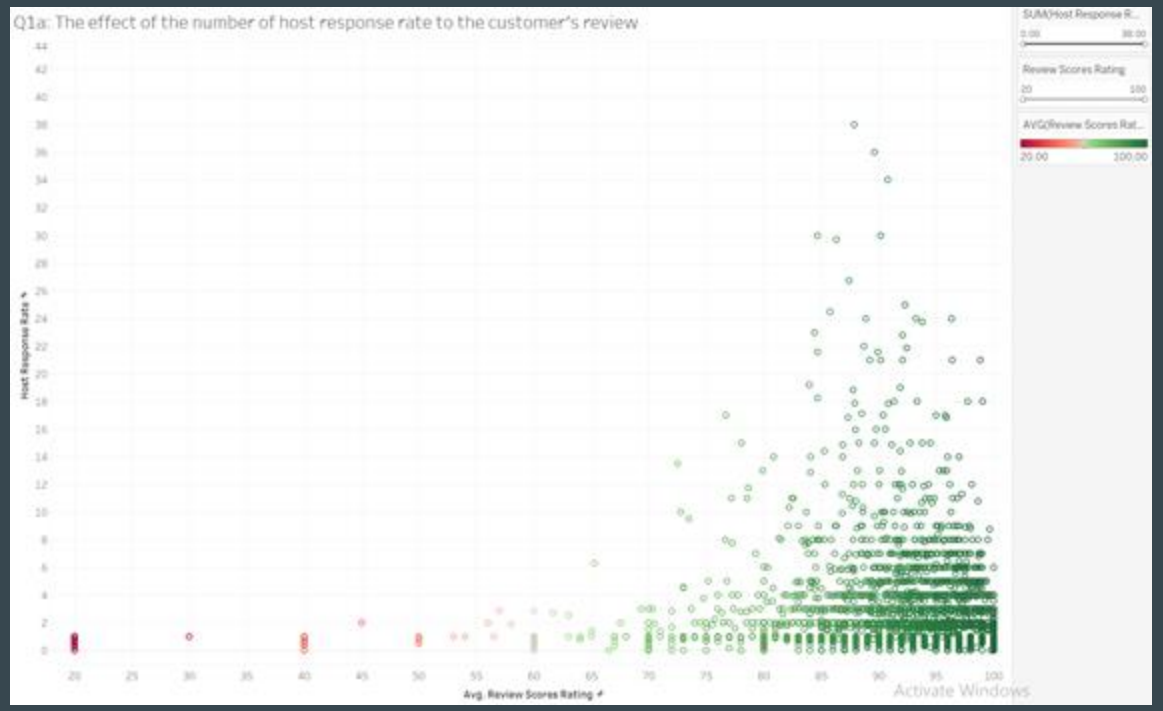






Q1:





- As can be seen in the data visualization, when the host Listing Count is below 20, it demonstrates that more rating scores are below 50 while there are more positive ratings if the host listing count is above 20, ranging from 70 to 100 of rating score.

- Mostly the rating that appears on the data visualization is ranging from 60 to 100, so we can assume the average rating of the total is about 80.

Q2:

2. After creating visuals to investigate our second research question we were able to determine that the cost of an AirBnB does not determine a customer's experience when renting the property. On Tableau we were able to make a scatterplot comparing the average price of a property and its rating. There were properties costing $1,000 with reviews nearing 85 and also 98. We also observed properties costing below $100 with ratings ranging from 80-100. A more expensive sight for rent on the website does not mean that the experience is guaranteed to be amazing.

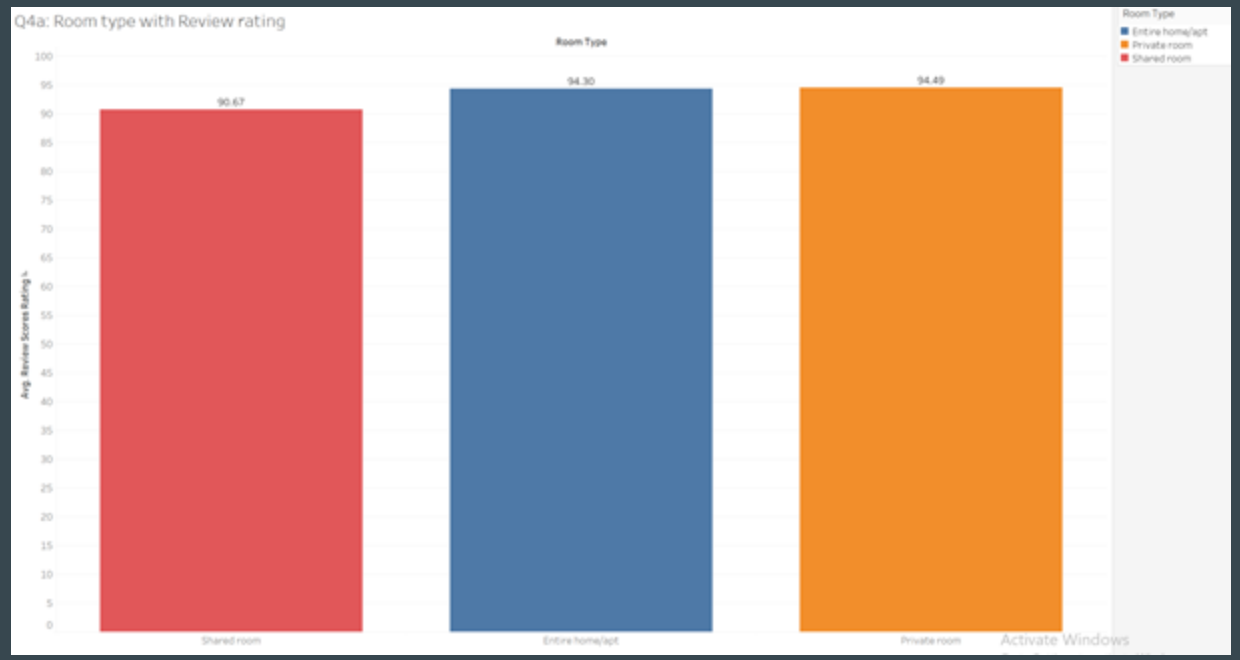
A. The properties in the top 25% of price do not have entirely high ratings like we had assumed. The observations are similar to those in our main question, meaning that AirBnBs with higher costs compared to other properties have both good and bad reviews.
B. Similarly, the bottom 25% of properties when comparing price have both good and bad reviews.

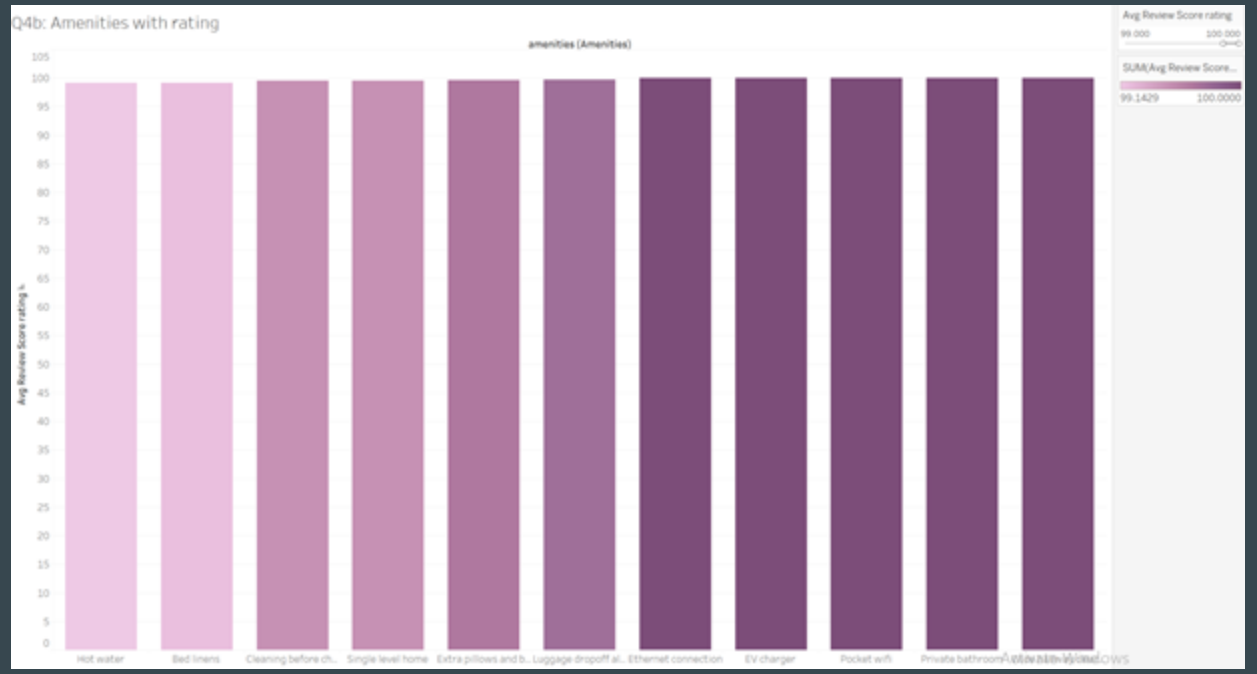
Q3:

3. The visual created for question 3 on Tableau helps us conclude that some neighborhoods in California are tied to higher reviews on AirBnB. We used a bar graph comparing different zip codes and the total/average ratings of properties in those areas. We found that the majority of neighborhoods in California have an average rating above 90. Some neighborhoods have a larger sum of total ratings, meaning there are more properties for rent on the AirBnB site in those neighborhoods.

A. In California, the majority of ratings lie between 90-100. There is one outlier in the group with an average rating of 20, which is very poor. This neighborhood has a zip code of 92821. There were a very small number of neighborhoods with ratings ranging from 80-90.
B. The neighborhood with the zip code 90291 has the highest amount of ratings in California but not the highest average rating. There are multiple neighborhoods with the perfect average rating of 100, and those would be: 90058, 90071, 90221, 90630, 90670, 91708, 91786, 93510, and 93552.

Q4:

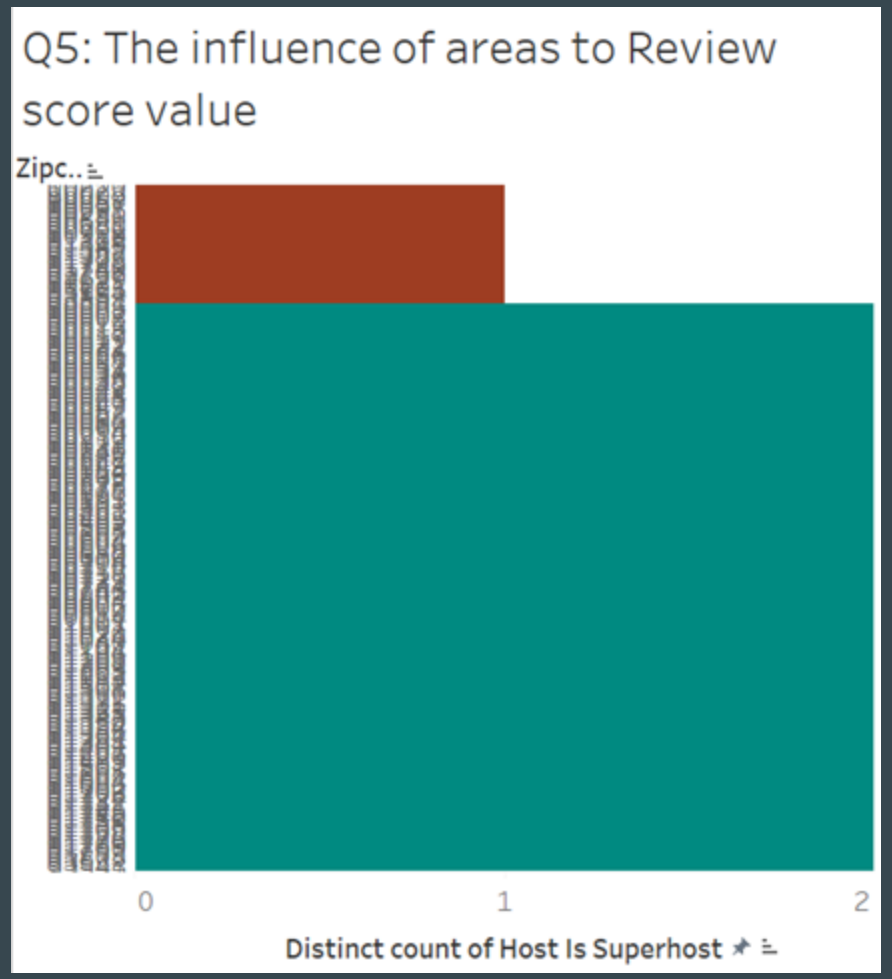




- Private room has the highest review score with a 94.49 average score rating, and the shared room has the lowest rating with a scoring average of 90.67.

- Let the standard of rating from 99 to 100 count as the highest rating. If using that metric to consider, hence Bed linens, Cleaning before checkout, Eternal connection, EV charger, Extra pillows and blankets, Hot water, Luggage dropoff allowed, Pocket wifi, Private bathroom, Single level home, Wide hallway clearance have the highest total rating.

Q5: Visualization Interpretation



- The zip code represents the neighborhood
- The red bar indicates (if the value is 1) the host **is** not a superhost
- The green bar indicates (if the value is 2) indicates the host **is** a superhost

- According to AirBnB, superhosts are “highly rated, experienced, reliable, and responsive.”
  - From this chart we can see that there are a lot more superhosts than normal hosts in these analyzed neighborhoods of California

Question Revisions

1. Will the host-response rate influence rating the client will give AirBnB?
   - → Does the number of listings a host has influence the rating the client will give the AirBnB?
   - (host_response_rate, review_scores, review_scores_rating, review_scores_value)
   - → (host_listings_count, rating)
1. Does the cost of the AirBnB correlate to a good or bad experience/review?
   - (price, review_scores, review_scores_rating, review_scores_value)
   - →(cost, rating)
3. Are different cities in California tied to higher or lower rating rates?
   - (city, host_response_time, host_response_rate)
   - → (location, rating)

Question Revisions... Continued

4. Are the property, room type, amenities, and description correlated with the customer decision to book / review?

- (property_type, room_type, accommodates, bathrooms, bedrooms, beds, amenities, review_score_rating)
- → (property, room type, amenities, review)

5. How does how many hosts in an area (neighborhood) influence the number of superhosts (superhosts are highly rated)?

- → How does how many hosts in an area (neighborhood) influence the number of superhosts? (host_location, host_is_superhost)
- Changed criterias: (city/county) → (neighborhood)

Limitations

- When interpreting the findings we needed to ensure we were using the correct measurements to conclude answers to our research questions.
  - For example, when creating visuals, we discussed whether to use average or sum when adding columns/rows measures

- A technological constraint we encountered was when the original dataset was difficult to download and work with due to its very large size
  - We combatted this problem by cleaning the data and making the file smaller

Takeaways & Conclusions

Our Customer Satisfaction Analysis aims to help AirBnB hosts find better success when hosting their property. In addition, our research is meant to help AirBnB customers get the best possible experience they can by determining what influences a positive or negative experience with hosts and the website as a whole.

Through our research questions and visualizations we landed on certain conclusions:

- Price of a property does not determine if a customer will give a good or bad rating of the rental
- There are some neighborhoods in California with impeccable average ratings on AirBnB, but there aren't many with awful ratings. Overall, most areas have ratings ranging from 88-97.
- There are more superhosts than non-superhosts in the California area.

Actionable solutions:

- Given that there are more superhosts than not, a new AirBnB host should strive to become a superhost by being highly rated, experienced, reliable, and responsive.
- Hosts should include the top amenities related to high ratings when including amenities in their AirBnB listing.
- Customers shouldn't aim to spend the most on an AirBnB to ensure they have a good experience, instead look in areas that interest you and book from there.

Thank You!