



Price, Rating, and Host Analytics

Team: 49 Seahawks

Members: Aditya Narayanan, Disha Srivastava, Meghana Jaladanki, Nikhil Rajkumar

OIM 350: Business Intelligence and Analytics | Spring 2021

Professor Yi Zhou

Agenda

1. Business Context
2. Project Goal
3. Research Questions
4. Dataset Description
5. Methodology
6. Findings and Visualizations
7. Limitations
8. Takeaways and Recommendations



Business Context

Airbnb is one of the world's most valuable vacation listing services and with that comes a very large user base. To select the best listing, customers of Airbnb primarily look at reviews of listings, hosts and prices to determine a good fit for a new vacation stay.

Analyzing these metrics help Airbnb decide how to prioritize their listings and showcase top hosts and top priced listings. The potential business implications that we can apply with our insights would involve helping Airbnb and their customers better understand hosts, listings, and how ratings are finally decided based on these different criteria/attributes.

[Why reviews matter | Reviews build trust in the community and help drive your business.](#)

- *“Reviews are essential to the entire Airbnb community, helping guests choose their travel plans wisely and enabling hosts like you to open their homes with confidence and attract guests who will love their stay.”*



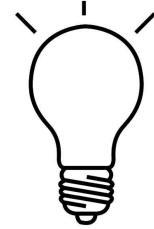
Project Goal

In this BIA project we are hoping to utilize the [LA Airbnb Listings](#) dataset in order to test the impact of certain metrics (such as price, location, host profile, amenities) on an Airbnb's overall ratings. The reviews left by Airbnb guests help power their recommendations algorithm, increase sales for hosts on the platform, and fuel their overall ebookings business. Through looking at past reviews and how certain metrics may have influenced them, we hope to discover which aspects of the guest experience Airbnb's strategy team should focus on improving to provide their customers with the best experience possible and boost their reviews/ratings on the platform.

A majority of Airbnb customers as well as hosts rely on reviews. *Reviews are crucial; a strong or poor review can make or break a host's prospect for future profits on this platform.* We want to analyze what other factors go into affecting a customer's review, in order to increase the credibility of a review as we can validate it based on the factors we have taken into account. This would also motivate hosts to increase their quality of stay for customers by focusing on the most important metrics.



Research Questions



1. Is there a correlation between the neighborhood and ratings?
2. Is there a correlation between price and ratings?
3. What are the most common types of rooms seen in LA?
4. Discover neighborhood clusters on a map to find out popular neighborhoods.
5. Does the number of total listings a host has affect their communication scores?
6. How does a host's identity verification status impact review scores?

We aim to analyze Airbnb's data to discover how different factors influence ratings - whether it is a host's rating or a rating of their listing. Our research questions will be used to find correlations of ratings with price and the location of a neighborhood. Based on this, we also want to investigate the most popular areas where Airbnb's are located around the city of Los Angeles. Our questions analyze how the price of a listing would change with location of the listing around Los Angeles as well as observe how ratings of a listing change with the price of a listing. Next, we wish to find out the most common types of rooms in LA and possibly find trends in room prices and room type. We are also interested in learning more about the hosts of these Airbnb's, and therefore wish to investigate how a host's identity being verified relates to their score. Finally, we wish to investigate whether the total number of listings a host has affects their communication scores with their customers.



Data Description

- We use the LA Airbnb Listings dataset to observe the impact of metrics like (such as price, location, host profile, amenities) on an Airbnb's overall ratings.
- Airbnb is an online platform connecting users who want to rent out their properties to those who are looking for short-term leases or rentals. After the guests leave, they may leave any reviews and ratings about the listing and host, which depict their experience staying there. The dataset used in this study is from Kaggle.com, which is originally sourced from Inside Airbnb, an investigatory website on how Airbnb is being used in cities around the world, for Los Angeles locations.
- The original dataset contains 31,253 rows and 95 columns and we have taken a sample size of 5000 rows and 21 columns for our analysis.
- **Time-Frame:** The data is scraped from 1-May-2017 to 3-May-2017.
- **Key Variables:** TopRatingID, Id, Host_id, Host_is_superhost, Host_total_listings_count, Host_identity_verified, Neighbourhood_cleansed, Latitude, Longitude, Property_type, Room_type, Amenities, Price, Availability_365, Number_of_reviews, Review_scores_communication, Review_scores_location, Review_scores_value, Cancellation_policy, Reviews_per_month, bedrooms



Methodology

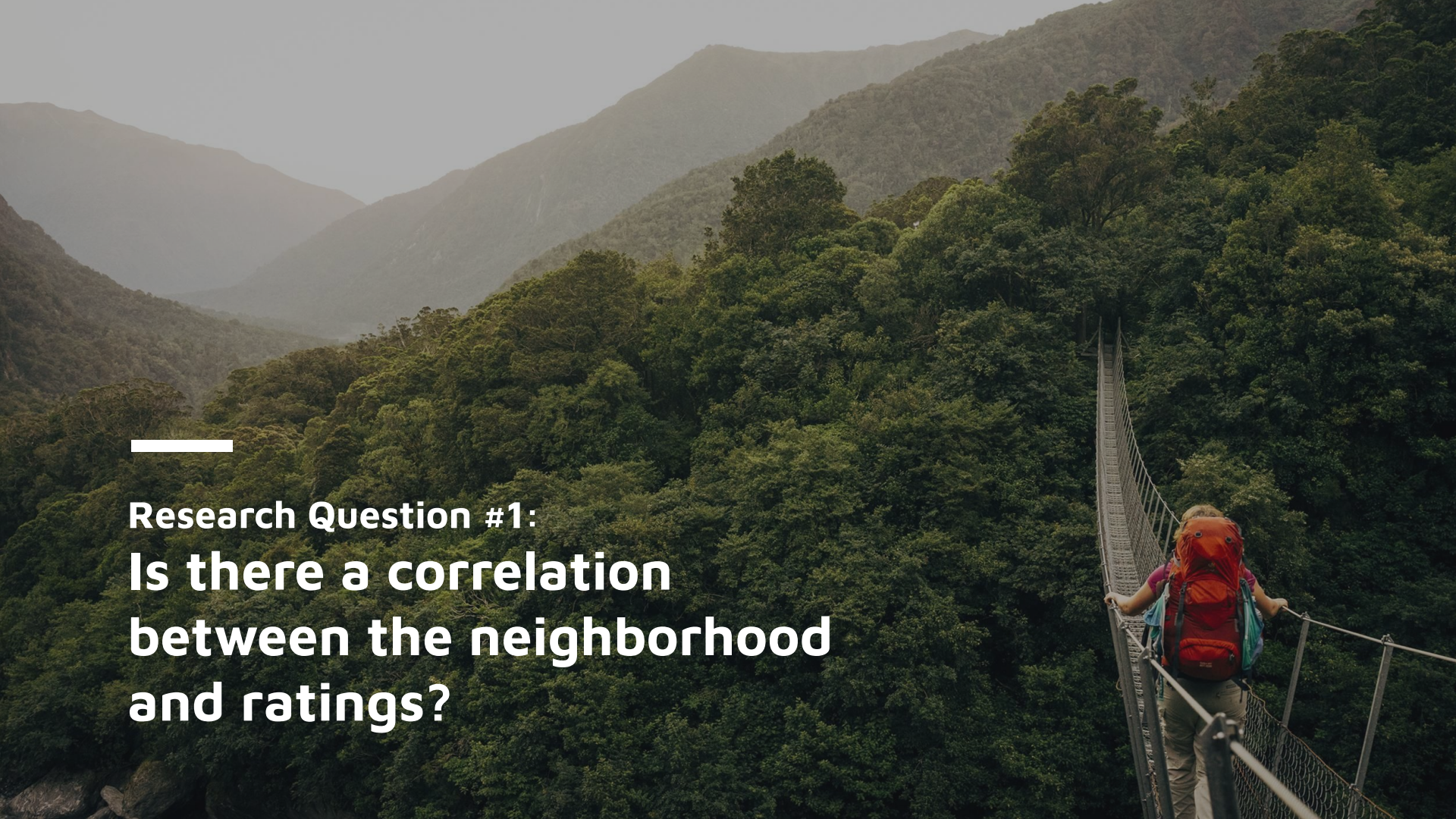


Data Cleaning

- First we deleted columns irrelevant to our research questions and essentially created a subset of our dataset
- We then deleted all null values and replaced them with ' '

Applications/Tools

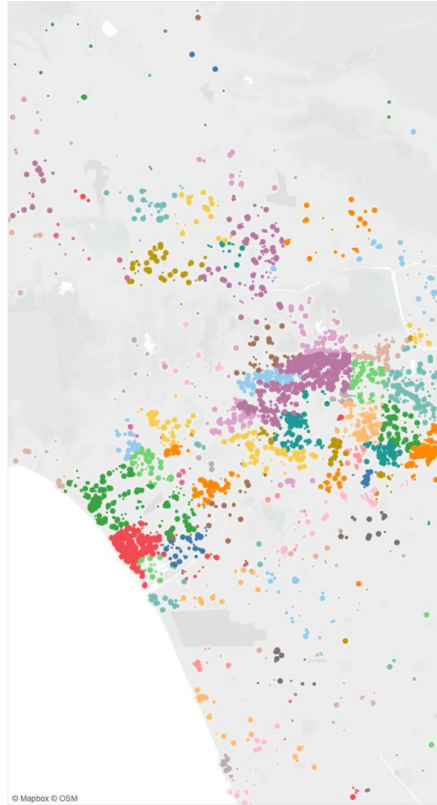
- We performed the data cleansing through Excel
- A github repository was created where we could easily share the data and make changes to it if needed
- We created visualizations through Tableau, such as airbnb locations plotted on maps, bar charts showing different distributions, and bubble charts

A person with a red backpack is walking across a suspension bridge that spans a deep, dense forest. The bridge is made of metal cables and a wooden plank. The person is seen from behind, looking towards the end of the bridge. The forest is lush with green trees, and in the background, there are misty, rolling mountains under a soft, hazy sky. The overall scene is serene and adventurous.

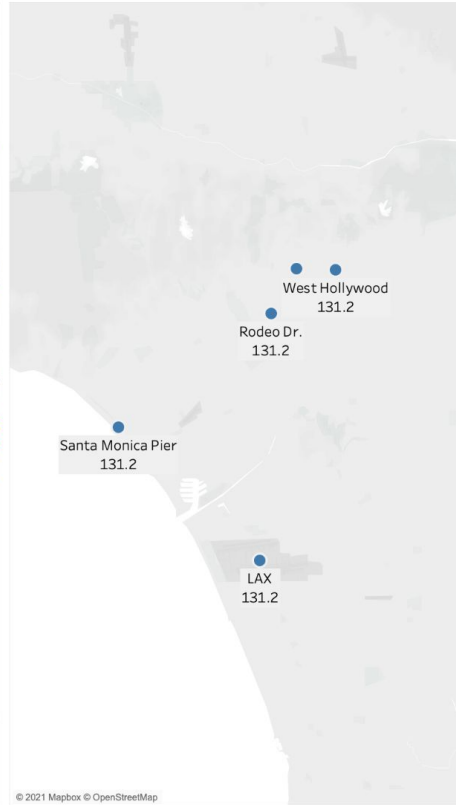
Research Question #1:
Is there a correlation
between the neighborhood
and ratings?

Analyzing Effect of Neighborhoods on Ratings

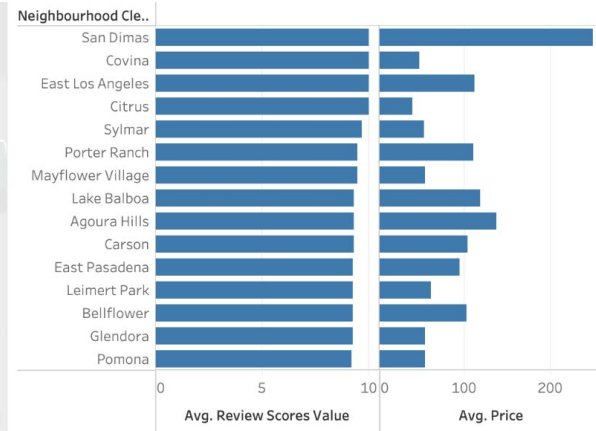
Map of Listings by Neighborhood

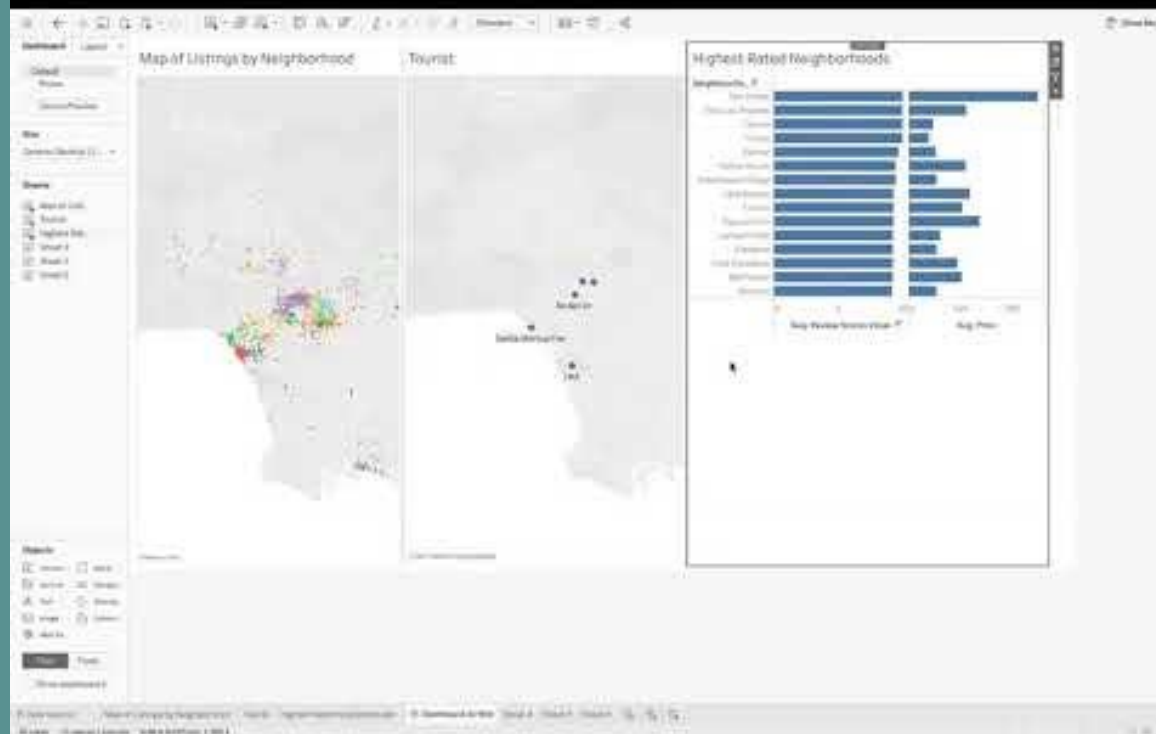


Tourist



Highest Rated Neighborhoods








Impact of Neighborhood on Ratings

- Is there a correlation between the neighborhood and ratings?
 - The top rated neighborhoods seem to have higher ratings due to their lack of airbnb bookings.
 - The top rated neighborhoods do not necessarily have cheaper prices. Meaning their prices do not have an effect on the ratings
 - With areas such as San Dimas and Covina, there are not as many ratings compared to popular tourist spots, so if an airbnb in the neighborhood is highly rated even once, it really affects the overall rating of the neighborhood
- Is there a correlation between neighborhoods that are closer to tourist spots and ratings?
 - The neighborhoods with tourist spots have a high amount of airbnb's as shown in the map, meaning they also have more ratings
 - It does not seem as though tourist spots have a negative impact on ratings. If neighborhoods with tourist spots are looked at, such as Venice, Hollywood, and Downtown LA, they are not rated considerably lower than neighborhoods that do not have top tourist spots

There is no negative effect observed of tourist neighborhood on ratings and the average price for neighborhoods does not have an effect on their ratings.

A person with a red backpack is walking across a suspension bridge that spans a deep, dense forest. The bridge is made of metal cables and a wooden plank. The person is seen from behind, looking towards the end of the bridge. The forest is lush with green trees, and in the background, there are misty, rolling mountains under a soft, hazy sky. The overall scene is serene and adventurous.

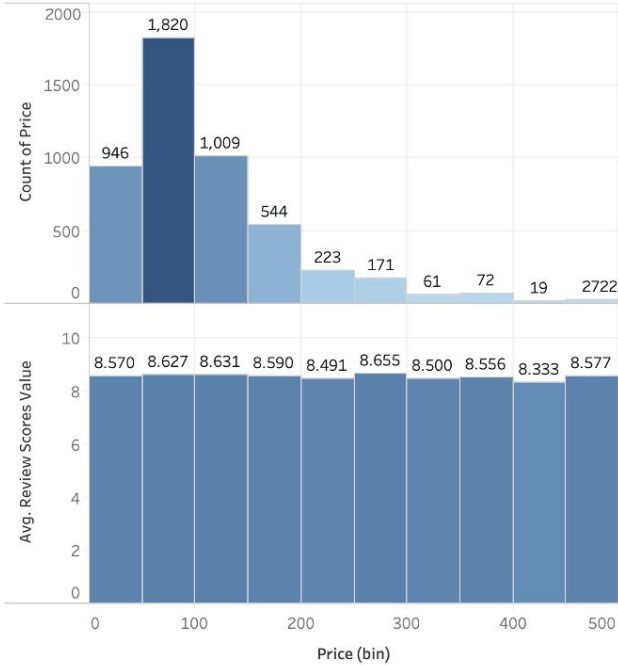
Research Question #2:
Is there a correlation
between price and ratings?

Analyzing Impact of Price (less than \$500) on Ratings

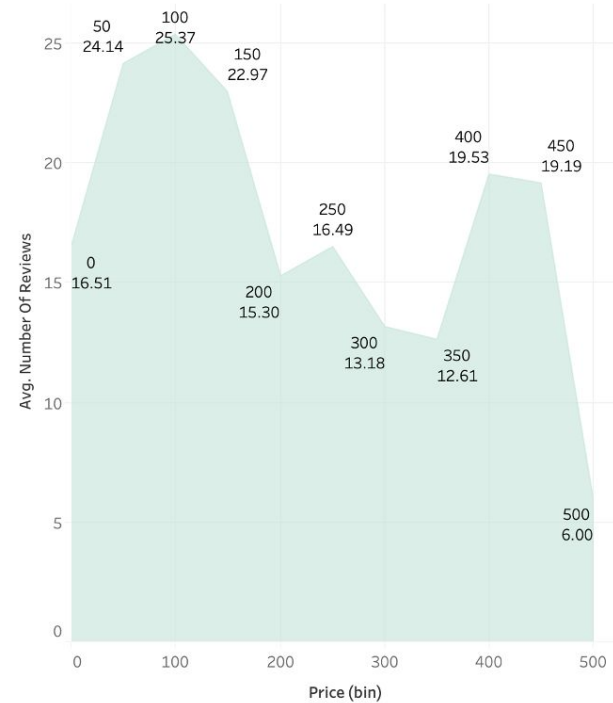
Distribution of Prices

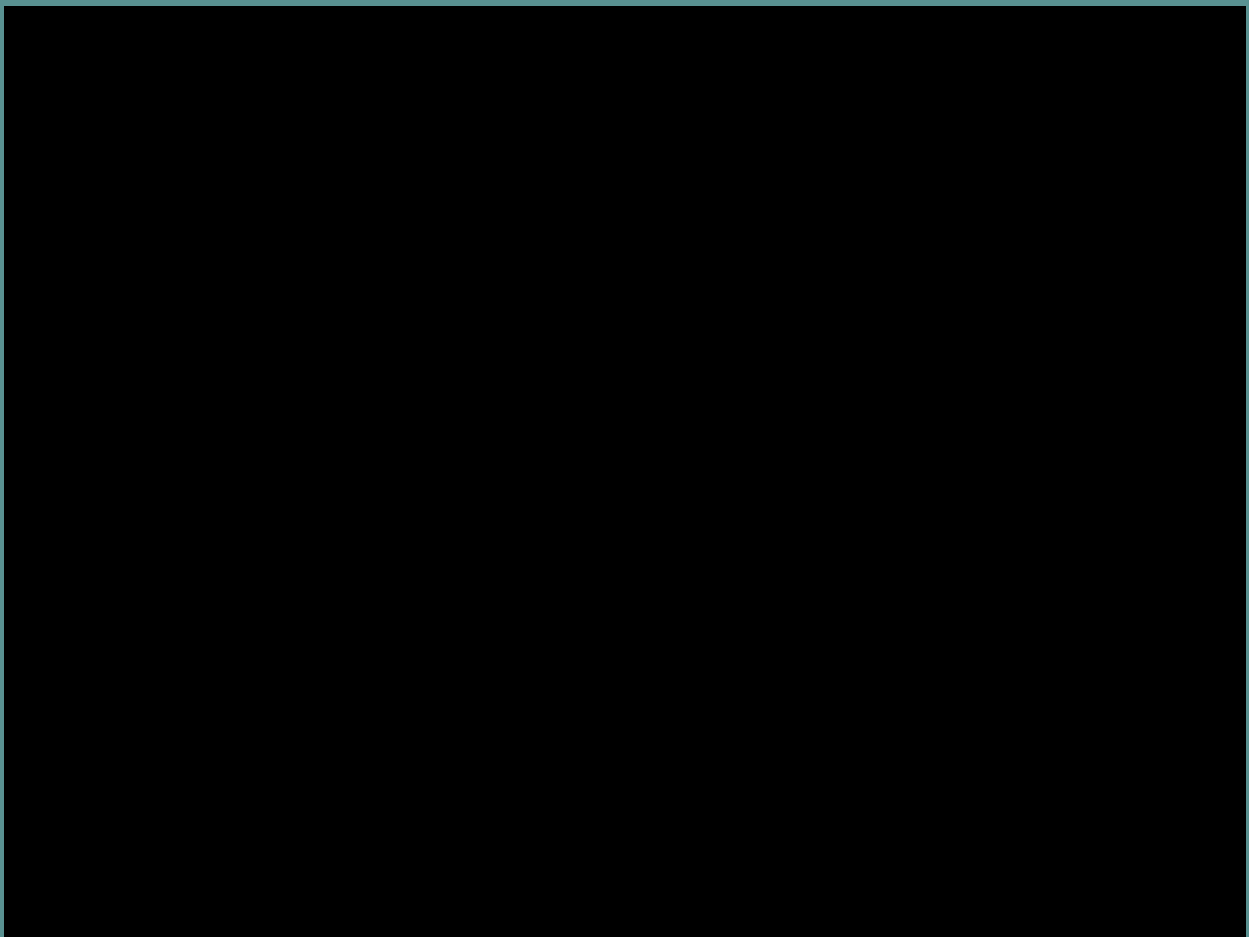


Distribution of Reviews and Number of Listings per Price Bin



Average Number of Reviews per Price Bin







No Correlation between Price and Ratings

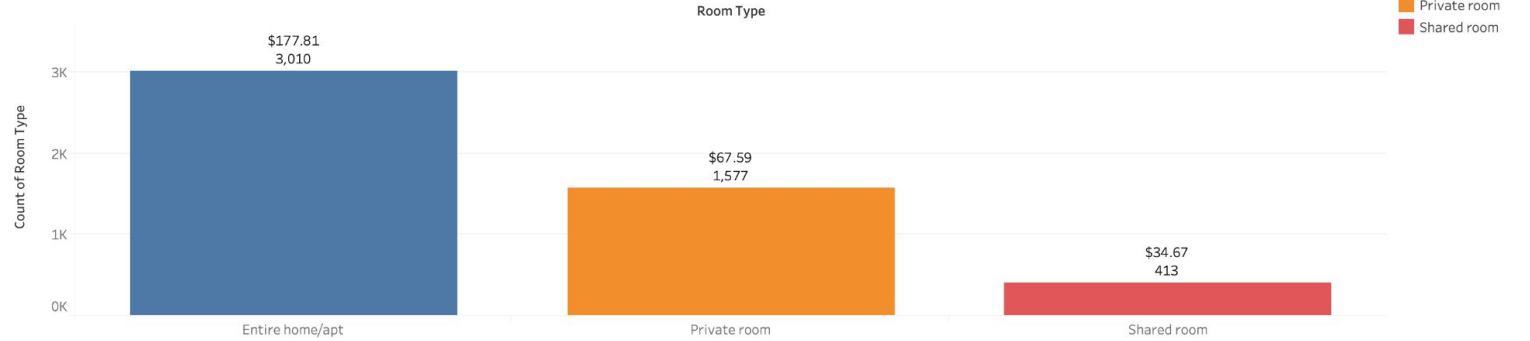
- The distribution of prices is from \$10.00/night (minimum) to \$4,000/night (maximum).
- The distribution of ratings is from 2.00 (lowest) to 10.00 (highest).
- Price does not seem to have a strong correlation with Airbnb ratings. As price increases, there is no linear relationship between ratings and price. This goes for decrease in price as well. We can see from our histogram that each price bin fluctuates around an average price rating of 8.00.
- Apart from listings with prices above \$1,000/night (which can be considered outliers as they cater to a small subset of the general population), most price ranges have well distributed amounts of guests leaving reviews; the lower the price, the higher amount of reviews.
 - There is also a sharp uptake in number of reviews at \$50/night, with the \$50.00 - \$100.00 price bin experiencing the highest amount of review traffic.
- The data has been filtered to remove null values and focus our analysis on Airbnb listings under \$500.00 a night.
- If hosts are looking to improve their average rating on the platform, they should look at other metrics outside of price ranges.

A person with a red backpack is walking across a suspension bridge that spans a deep, dense forest. The bridge is made of metal cables and a wooden plank. The person is seen from behind, looking towards the end of the bridge. The forest is lush with green trees, and in the background, there are misty, rolling mountains under a soft, hazy sky. The overall scene is serene and adventurous.

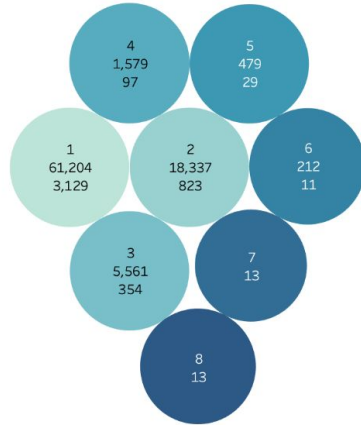
Research Question #3:
**What are the most common
types of rooms listed on
Airbnb in LA?**

Analysis of Airbnb Rooms

Rooms in Listings

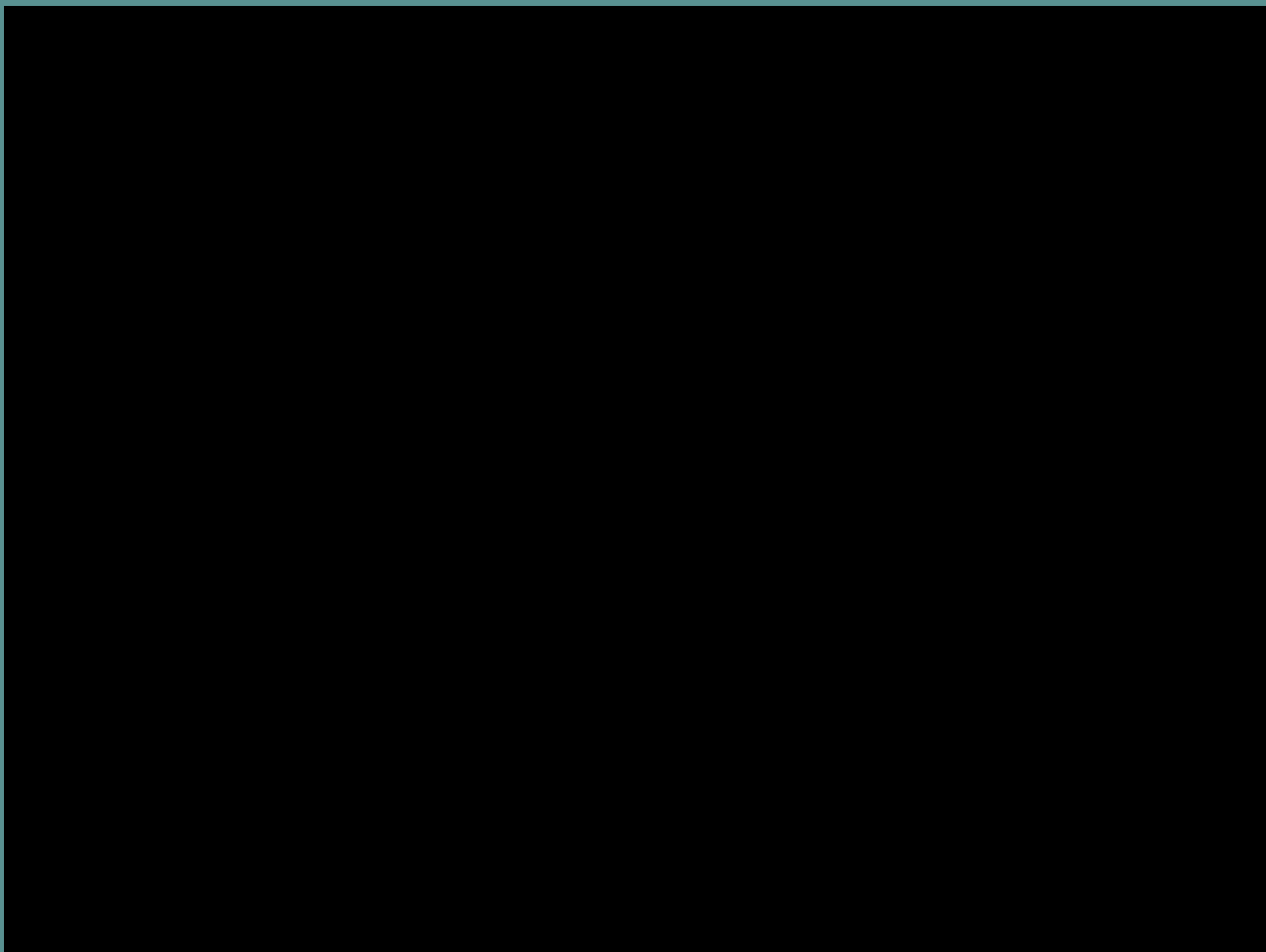


Bedrooms vs Rating



Avg Price vs Bedrooms







Analysis of Airbnb Rooms and Room Types

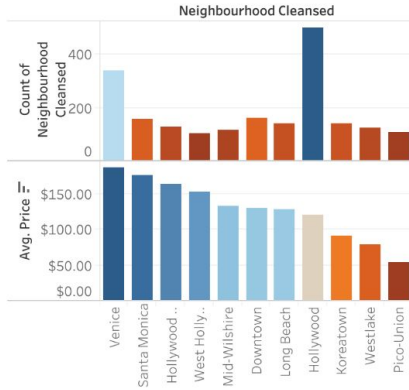
- Entire homes/apartments are the most popular room types in LA followed by private rooms and shared rooms, from our sample.
 - The average price of entire apartments is \$177.81, private rooms is \$67.59, and shared rooms is \$34.67. There are 3010 “entire apartment” listings, 1577 private room listings, and 413 shared room listings in this sample.
 - Groups like families or larger groups of friends are likely to prefer getting entire homes/apartments for their stays as compared to single travelers who are in LA for a short period of time, whether it is business or transit.
- Looking at the relationship between price and number of bedrooms, we can see that listings that are entire apartments have the highest average price when the apartment has 1-2 bedrooms. Second, looking at private rooms, again are the highest at 1-2 bedrooms.
- 4-bedroom listings have the highest rating among listings with bedroom numbers ranging between 1-8.

A person with a red backpack is walking across a suspension bridge that spans a deep, dense forest. The bridge is made of metal cables and a wooden deck. The forest is lush with green trees, and in the background, there are misty, rolling mountains. The scene is captured from a low angle, looking up at the person and the bridge.

Research Question #4:
**Highlight neighborhood clusters
on a map to discover popular
neighborhoods.**

Location Analytics on Neighborhood Clusters

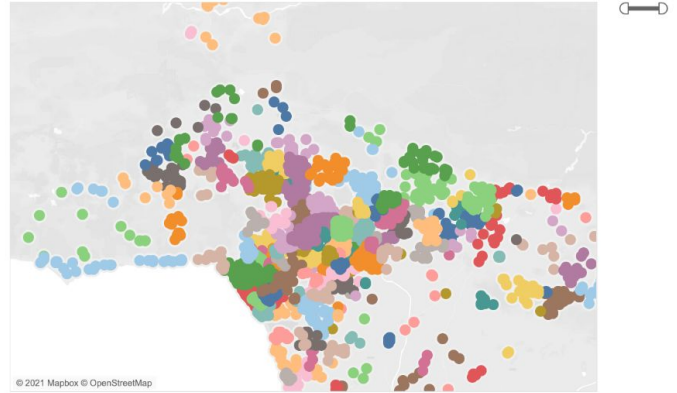
Most neighborhood listings



Expensive Neighborhoods



Map of LA Listings

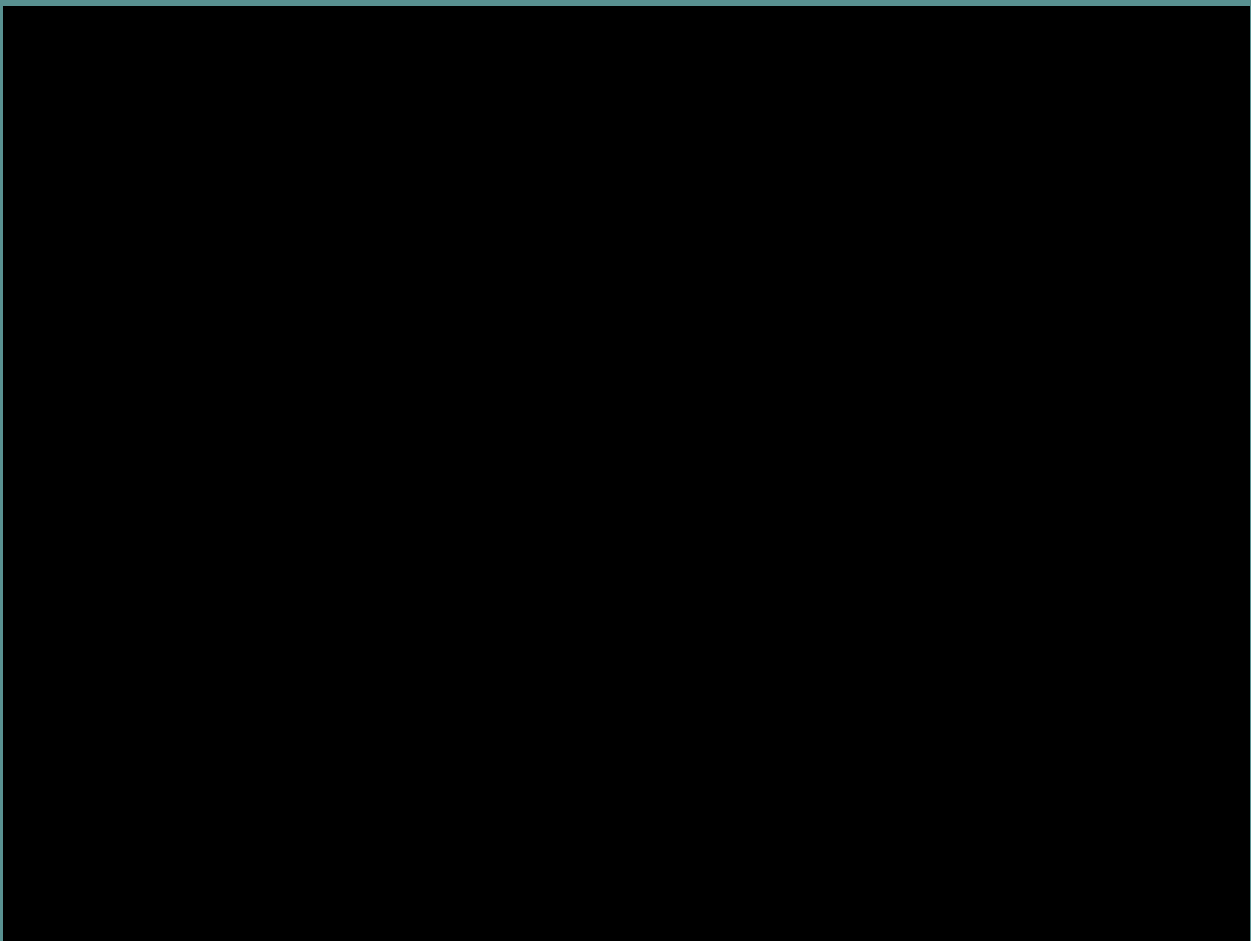


Neighborhood Cluster Ratings - Hollywood



Neighborhood Cluster Ratings - Venice







Airbnb Location Analytics

- The different neighborhoods are marked in “Map of LA Listings”. The most popular neighborhood is Hollywood followed by Venice, in this sample.
- Venice was also the most expensive area to stay in due to the highest average price in the sample. Hollywood was not very expensive as compared to Venice which shows why it was so popular as well!
- The two visualizations of Neighborhood Cluster Ratings showcase that neighborhoods which are popular and have a greater number of listings do tend to have higher average ratings than listings located in less popular areas.
- Lastly, there are neighborhoods that form clusters because of their prices. Malibu, and Bel-Air were the most expensive neighborhoods on average, but Agua Dulce had only one listing which was the most expensive listing in the sample.

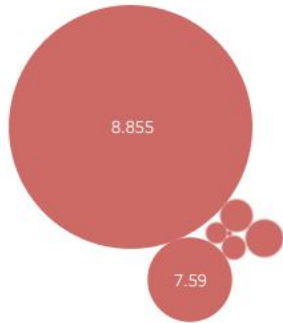
A person with a red backpack is walking across a suspension bridge that spans a deep, dense forest. The bridge is made of metal cables and a wooden deck. The forest is lush with green trees, and in the background, there are misty, rolling mountains. The scene is captured from a low angle, looking up at the person and the bridge.

Research Question #5:

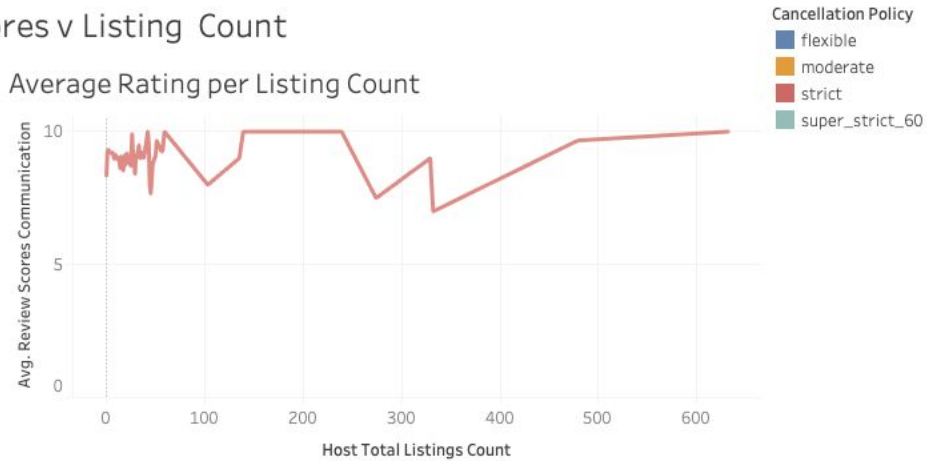
Does the number of total listings a host has affect their communication scores?

Communication Scores v Listing Count

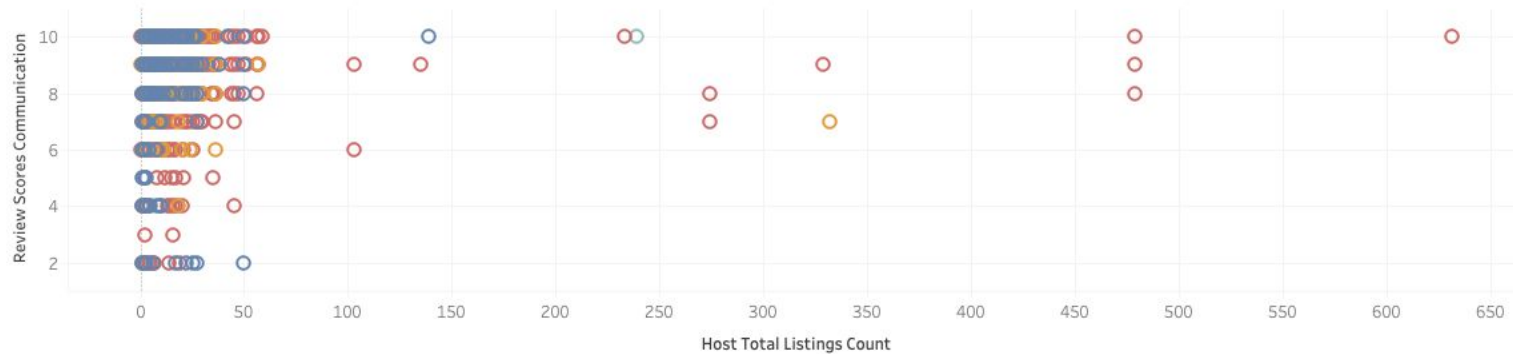
Distribution of Communication Scores

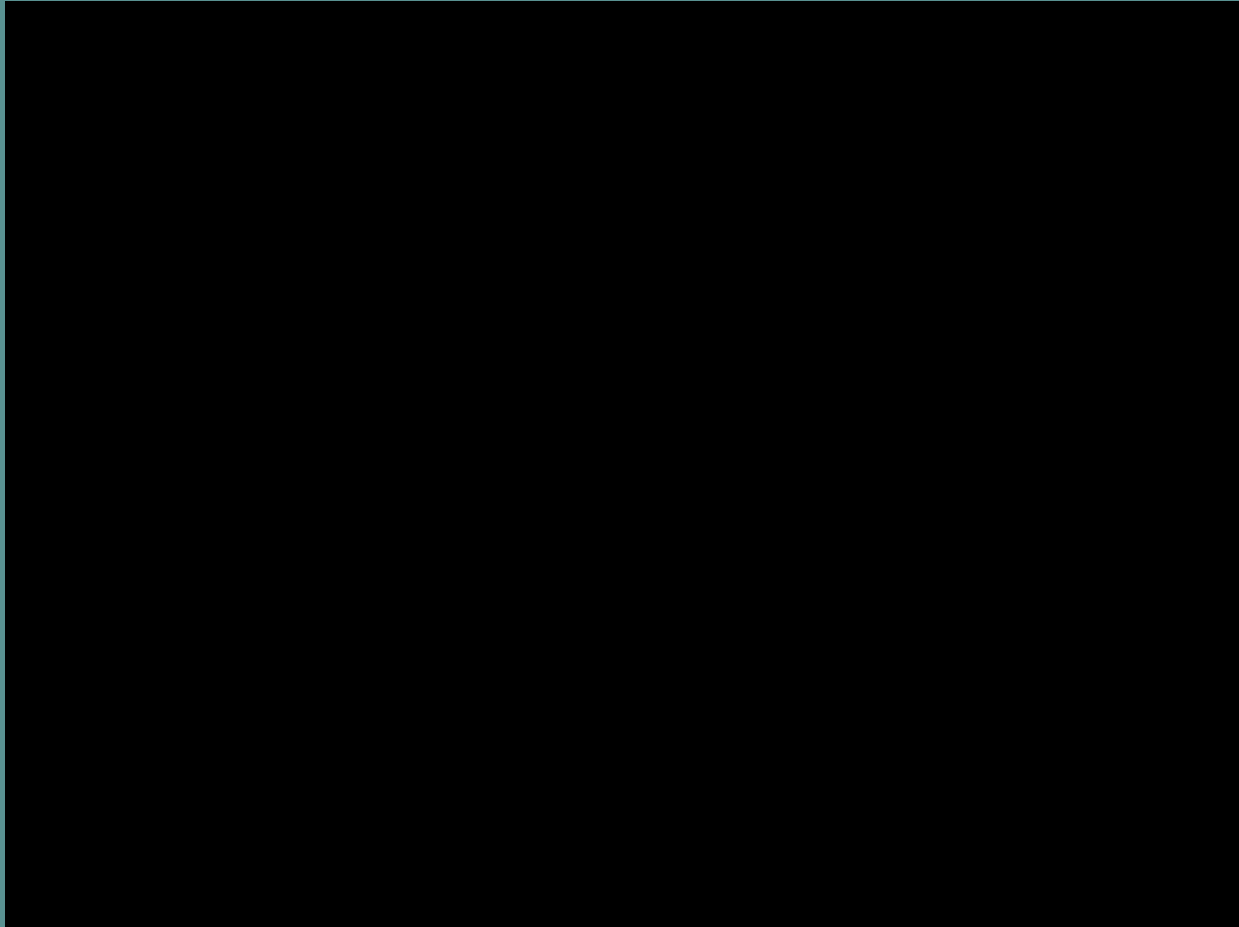


Average Rating per Listing Count



Communication Scores v Listing Counts








Communication Scores and Listing Count

- Firstly, we notice how the communication scores are distributed over the data set. We note that most of the hosts have received an average of 8.85 as a communication score.
- Next, based on the number of listings, we see how the average communication rating varies. This shows us a more general trend of how a listing count could affect communication.
- Finally, we analyze our scatter plot displaying correlations between communication scores and listing counts. This shows us if a correlation exists or not, and according to the plot we notice that most of the listings have reviews varying from 0-10 and only the outliers show trends. I.E. The hosts who have listings >100 have reviews > 6.0 and that is a trend that holds true for the whole dataset. But when listings < 100 , the reviews are distributed over the entire range and therefore does not show any correlation.

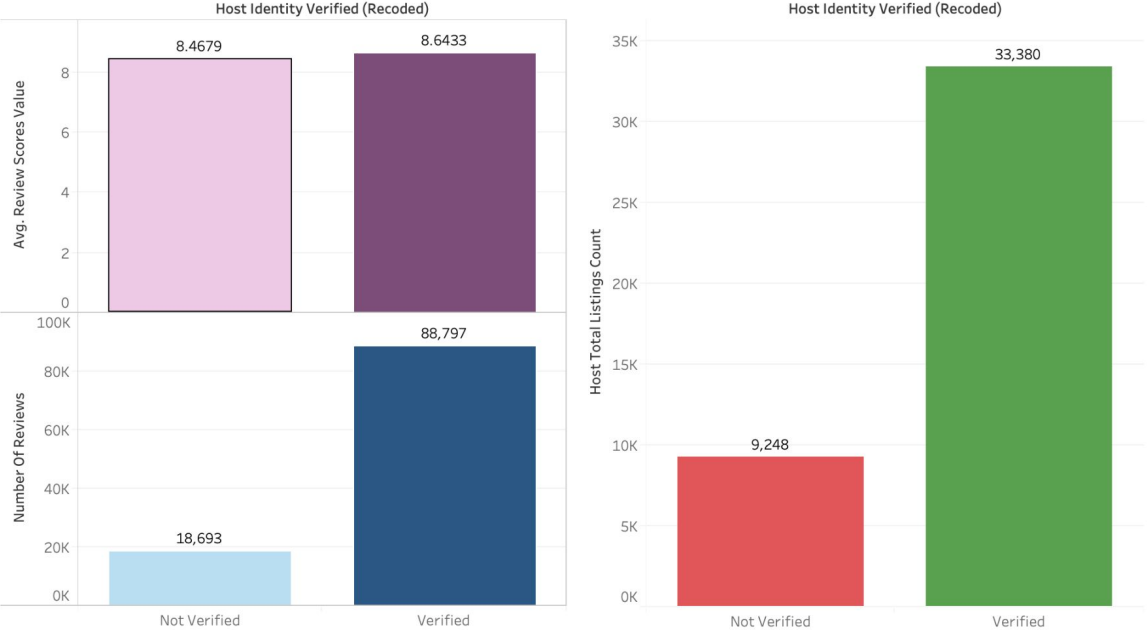
A person with a red backpack is walking across a suspension bridge that spans a deep, dense forest. The bridge is made of metal cables and a wooden plank. The person is seen from behind, looking towards the end of the bridge. The forest is lush and green, with many trees. In the background, there are more mountains and hills, some of which are shrouded in mist or fog. The overall scene is a beautiful, natural landscape.

Research Question #6:
**How does a host's identity
verification status impact
review scores?**

Analyzing Impact of Host Verification Status on Review Scores



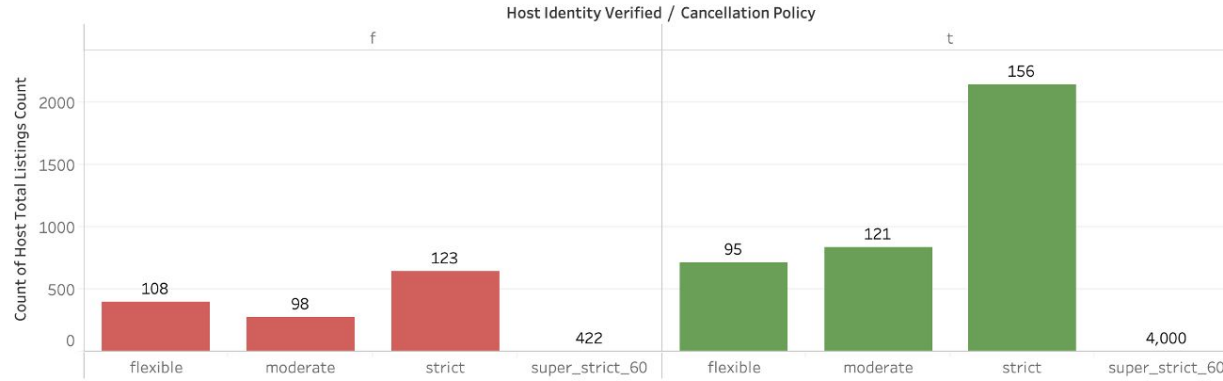
Average Review Score per Host Verification Status Number of Listings per Host Verification Type



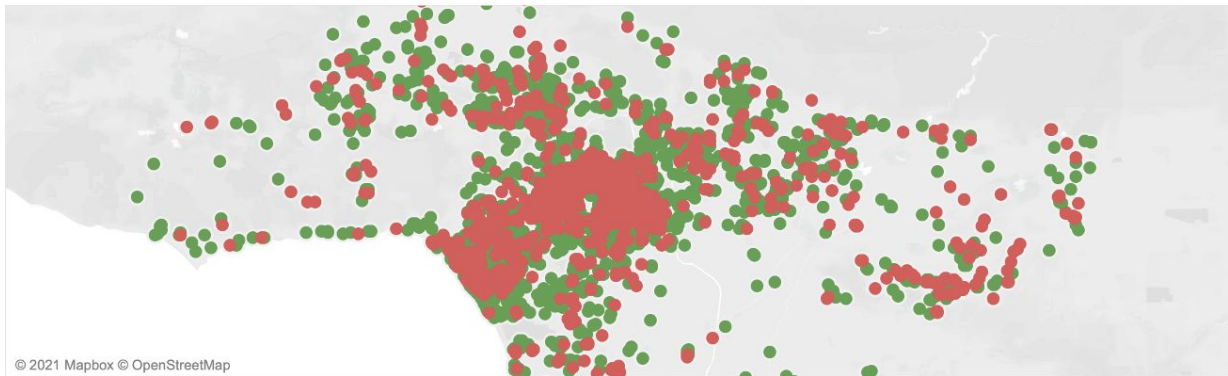
Analysis of Host Verification Status Benefits

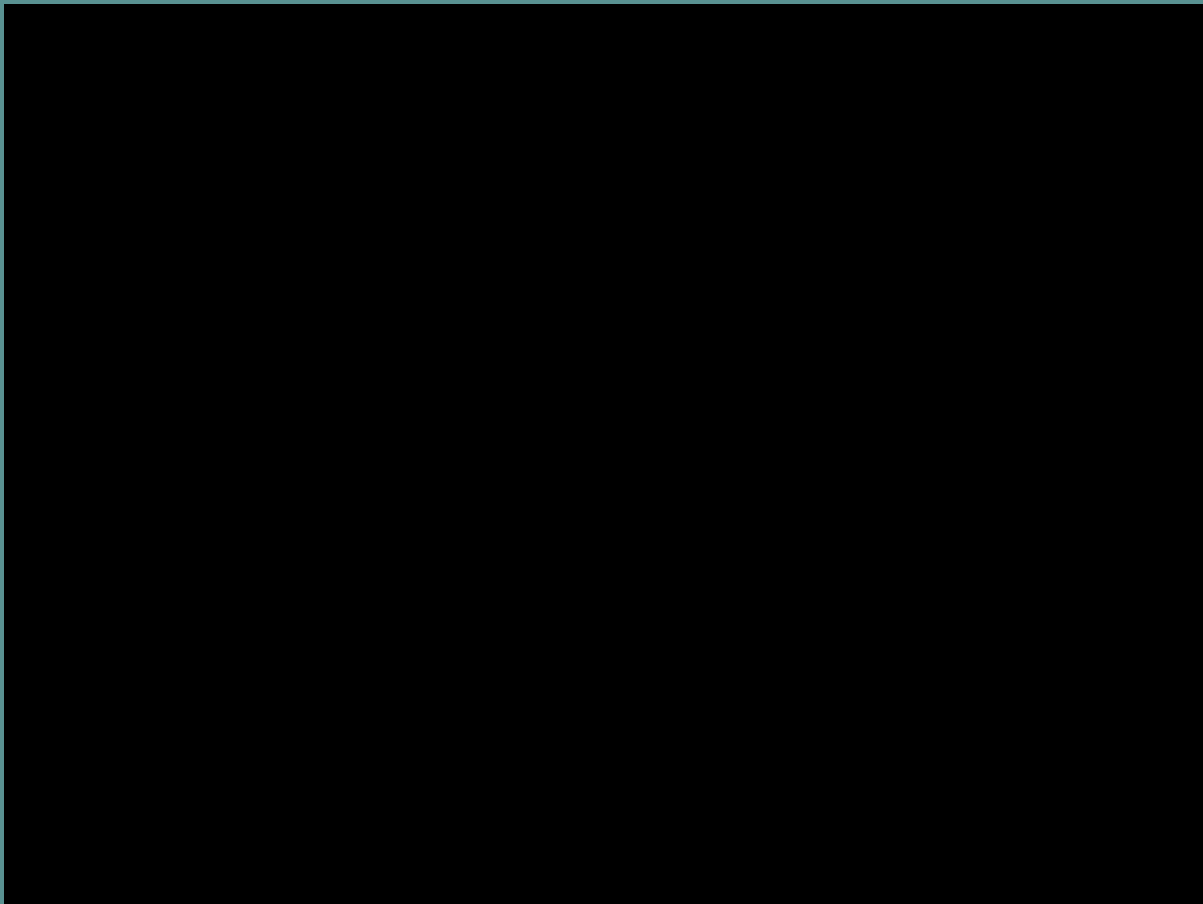
Host Identity Verified
f t

Number of Listings per Host Verification Type



Map of Verified Hosts





p.s. I apologize for the two cameras! I think Zoom was trying to make up for not showing my camera in my first recording!



Host Verification (T/F) Impact on Reviews

- The average review score for non-verified hosts is 8.4679 and the average review score for verified hosts is 8.6433.
 - Hence, the average rating review for non-verified hosts deviates from the average rating review for verified hosts by 0.1754. This does not create a large enough difference for hosts to be motivated to complete the extra steps to become verified.
- Taking the extra steps to get verified will most likely not help an Airbnb host in Los Angeles increase their overall rating; however, it might help them perform better in terms of being able to propose stricter cancellation policies and possibly make their guests feel safer if their location deviates from downtown or highly populated areas.
- A larger majority of hosts have chosen to get verified (there are 2,672 more verified hosts than non-verified), possibly to help guests feel more comfortable and safe (a metric we don't have quantitative measure for).
- Verified hosts have more reviews than unverified hosts so there is no sampling error present.



Limitations

- Overall, we were able to achieve our data analysis and exploration objectives using this dataset. To further analyze the impact of different variables on Airbnb prices and ratings, we could have used categorical variables to determine the trust and safety of areas and hosts. We could have asked questions such as: “Does host verification make guests feel safer and more trustworthy?”
- Second, we believe that the ratings could be biased as some guests could have poorer experiences at the airbnb as compared to others, and these ratings are not uniform across all guests.
- Third, we found some valuable data like income levels of neighborhoods, crime ratings, and tourist attractions nearby were missing. We could have used this data to pinpoint areas that are more popular than others and done a correlation analysis around this.



Takeaway & Recommendation #1

Takeaway #1: An Airbnb in LA's average review score is more impacted by location than price.

- Among the highest rated neighborhoods, there is no correlation between their price and rating. Furthermore, tourist spots improve the ratings of airbnb's in those neighborhoods.

 *Recommendation:*

Amazing luxury penthouse with Ocean view!

[1 review](#) · [Marina del Rey, California, United States](#)

- *For guests:* It is recommended that guests book airbnb's in tourist areas if they can afford the price, since bookings in neighborhoods with tourist spots seem to be ranked higher on average compared many neighborhoods that are not.
- *For hosts:* Certain neighborhoods require greater upkeep and higher rent, so the price of airbnb's can be increased since it does not affect ratings negatively. This is shown by the top-ranked airbnb's, due to the fact that there is no correlation between some of their high prices and ratings.



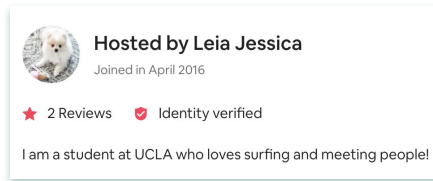
Takeaway & Recommendation #2

Takeaway #2: Obtaining a verified status as a host may not help increase their listing's average review score, but it will help with adjacent hosting perks.

- The average review score for non-verified hosts only deviates slightly (by .01754) from the average review score for verified hosts. However, verified hosts are able to place stricter cancellation policies than non-verified hosts which is a huge perk for hosts. It also appears that hosts in remote locations (away from downtown areas, houses are further spread apart) do complete the extra step to get verified, possibly to make their guests feel safer. LA geographically has three variants of location: costal, downtown, and valleys. Price, ratings, and verification status are impacted by the variance in geographical location types.



Recommendation:



- *For guests:* If a guest is not verified, it might be that they haven't completed the process yet. If you are staying in downtown or near tourist attractions, you are most likely staying with an experienced host regardless of verification.
- *For hosts:* If you Airbnb listing is not in downtown or near a heavily populated tourist area, it is beneficial to complete the verification process to attract more guests. If you are looking to increase your review scores, verification might not help as much. If you would like to have a stronger cancellation policy, it is recommended to become a verified host.



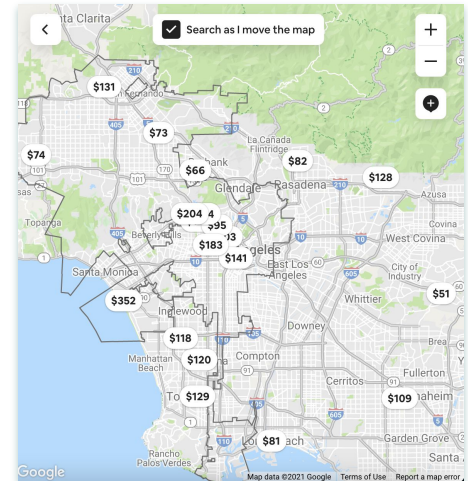
Takeaway & Recommendation #3

Takeaway# 3: LA's Airbnbs are scattered amongst coastal lines, downtowns, and exclusive neighborhoods. The number of listings in each area differs due to host upkeep.

- LA's Airbnb listings have prices and ratings that vary by neighborhood. We found clusters like Hollywood which have high ratings, many verified hosts and average prices, and Venice which had high costs, not many verified hosts due to it being a high tourist area.

Recommendation:

- *For guests:* If a guest wishes to stay in a more tourist-friendly or coastal areas, they should be ready to spend more money as the average price of these areas are above average. Hollywood is a very popular area to stay in due to its price being about average (\$120/night) and having a high number of verified listings.
- *For hosts:* If a host wishes to have more number of guests, they should keep listings in tourist-friendly areas, provided that they can afford to maintain them due to high costs. Areas which are not coastal or tourist-friendly tend to have lower costs, but may not attract many tourists. Listings in downtown areas may be popular among business travelers.





Takeaways & Recommendations #4

Takeaway #4: For guests, communications scores of a host do not increase because they are a more experienced host (with a higher number of total listings).

Recommendation:

- *For guests:* If the host has ≥ 100 listings, a guest can ensure solid communication with their hosts. However guests face the risk of dealing with a strict cancellation policy and must keep an eye out for that when booking with such hosts.
- *For hosts:* More listings (≥ 100) owned could lead to that host being more responsive and having better communication with their guests. This could be as the hosts are purely focussing all their efforts on ensuring their guest has the best experience and hence maintains top-notch communication with their guests.



Hosted by Hotel Pepper Tree

Joined in February 2018



532 Reviews



Identity verified

During your stay

Our front desk is open 24 hours a day to help with you any questions or needs.

Response rate: 99%

Response time: within an hour

Contact host