Eric Ngo & Kaylee Trinh

Project Report

An analysis of Airbnb Ratings & Reviews

May 05, 2024

Executive Summary

Since its establishment in 2007, Airbnb has transformed the travelling and accommodation landscape. The company is now one of the most popular lodging websites for travelers across the world. As an accommodation platform, Airbnb relies heavily on customers' feedback to inform other customers' decision on where to stay. Yet, much of the dynamics that shape customer reviews and ratings are unclear. In this project, we investigate patterns in customer reviews and factors influencing customer feedback (both in numeric ratings and written reviews) using sentiment analysis, word clouds, two-mode text networks, and simultaneous regression models. We also use these analytical tools to compare these patterns across Bangkok and San Francisco – two major metropolitans and tourist destinations of the world.

Overall, the findings provide interesting insights: while they confirm some apparent knowledge that listings with good/helpful hosts, good location, and clean and comfortable space invite more positive customer feedback, analyses also show that how hosts advertise about themselves, and their properties also matters. Particularly, hosts with positive introductions/biographies & positive neighborhood descriptions tend to receive higher customer ratings and more positive reviews.

In addition, the findings also reveal some inherent difference of Airbnb reviews in Bangkok and San Francisco. The former's districts can be divided into three explicitly clear clusters based on the level/amount of material features that their Airbnb listings possess (such as good locations or premium amenities). In contrast, San Francisco has notably more clusters that are overlapping and tend to have a more distinct qualitative characteristics or "vibes".

Problem Statement

Our research aims to analyze the factors influencing customer ratings and reviews on Airbnb, alongside identifying common themes in these reviews. Our focus will be on two global metropolitan and major tourist destinations, Bangkok and San Francisco, where we'll analyze patterns and compare similarities and differences between the two cities.

Data Description

We began with two datasets for each city: one containing listing details and the other listing reviews. We merged these datasets utilizing the common identifier 'listing'. The dataset for Bangkok comprises roughly 250,000 reviews, while San Francisco's dataset contains around 350,000 reviews. Each dataset has over 20 variables. Among these variables, we concentrate on several for our analysis, alongside the target columns—reviews and rating score. These include the availability of host information section and its corresponding sentiment, host experience (indicating how long the host has been on Airbnb), super-host status, presence of neighborhood description and its sentiment if available, listing price, and the number of reviews for each listing.

Data cleaning process:

- Language detection applied for reviews & removal of non-English reviews.
- Remove special characters, stop words (for word clouds & text networks), etc.
- Correct variables format: int/float for numeric, datetime for date.

Analysis

Sentiment Analysis

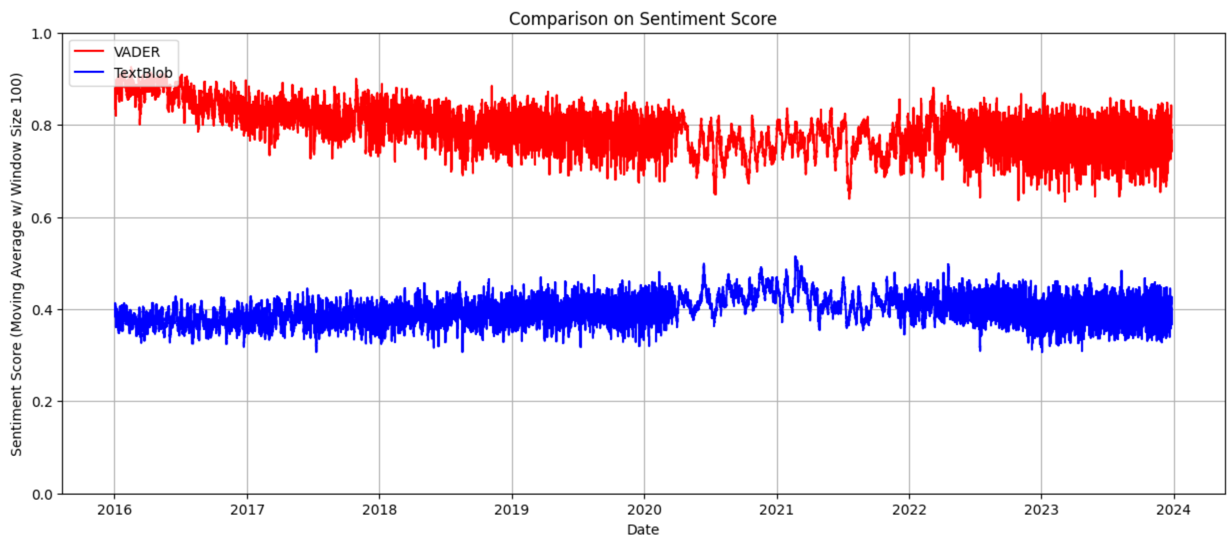

Figure 1A. Bangkok Reviews Sentiment Score

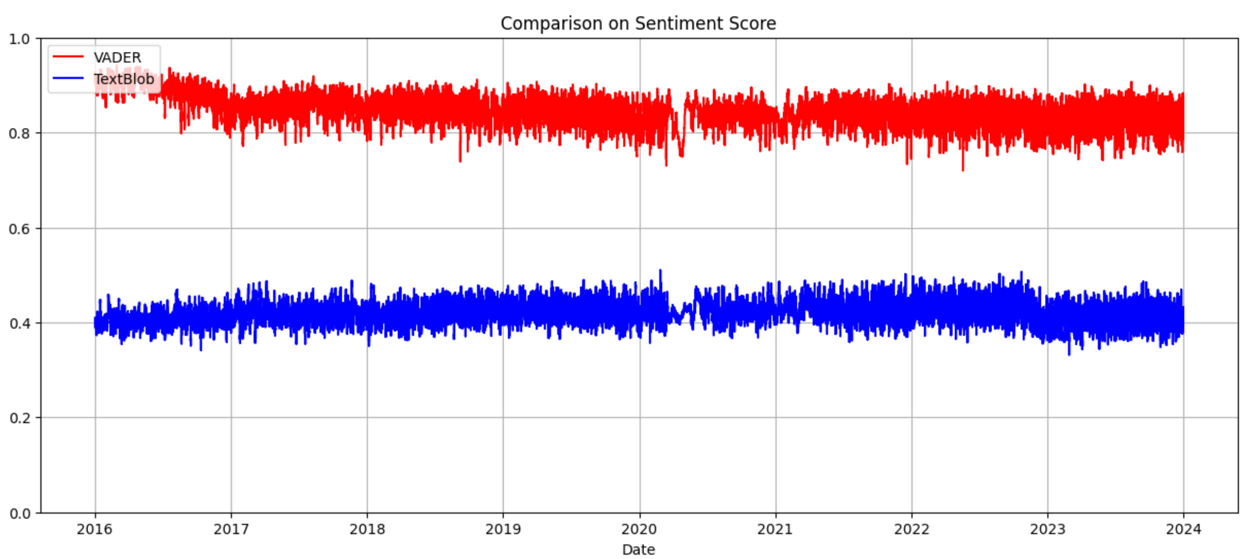

Figure 1B. San Francisco Reviews Sentiment Score

A glance at figures 1A and 1B reveals a clear similarity between the overall sentiment of Airbnb customer reviews in Bangkok and San Francisco: both cities have similar levels of sentiment, either measured by VADER or Text Blob. Most reviews are positive, which is unsurprising as this is perhaps due to a selection effect: people tend to spend time reviewing Airbnb properties that they have good experiences with. However, VADER seems to give considerably higher scores compared to TextBlob. Since VADER is generally considered a more superior metric (especially when it comes to short texts like customer reviews), we will mostly refer to VADER for comparisons between Bangkok and San Francisco (San Fran or SF).

Despite the apparent similarity, there are slight distinctions between the two major cities. First, the VADER score for Bangkok appears to be a bit lower than that of San Fran, with the former being generally under 0.8 whereas the latter being close to 0.9. Second, during the COVID years (from 2020 to early 2022), the trend line for Bangkok became notably thinner, reflecting the decrease in the number of reviews (and thus travelers). In contrast, the thickness of SF's trend line remained relatively the same. This corresponds to the fact that Bangkok, being in Thailand, followed a very restrictive COVID policies compared to San Francisco, which followed a much more relaxed pandemic regulations, allowing for many travelers to visit SF.

Common Themes with Word Clouds

To further investigate the patterns across the cities, we step beyond overall sentiments and explore common themes mentioned in customer reviews using word clouds.

A. Bangkok B. San Francisco

Figure 2. Word Clouds of Customer Reviews in Bangkok & San Francisco

Similar to the sentiment visualizations above, placing Bangkok's and SF's word clouds side by side highlights striking resemblance between guest reviews of Airbnb in the two cities. Particularly, the most mentioned words in both word clouds are “place”, “great”, “stay”, “location”, “clean”, “host”, and “nice”. They also include words such as “quiet” and “close.” This suggests that guests in both tourist destinations express their attention to characteristics of the property such as location, host, cleanness, quietness, etc. In other words, this confirms our (seemingly apparent) intuition that there are essential features of an Airbnb property that guests tend to care about, regardless of where their destination is.

Common Themes by Neighborhood Using Text Networks

The above comparisons so far seem to suggest no distinctions in guest behavior between Bangkok and San Francisco. However, the above sentiment time series plots and word clouds only reveal the most overall patterns in the data without considering a major factor unique in each city: urban planning. Specifically, each city is expected to have a different design, with various districts/neighborhoods and different arrangements. Such features and arrangements of neighborhoods reflect the geographical terrain and history of a city, making each urban area unique.

To incorporate such spatial makeup of the two cities into our analysis, we utilize text networks to visualize common themes mentioned by guests by BK’s and SF’s neighborhoods. Each figure below contains a two-mode text network, which visualizes relationships between 2 types of nodes/vertexes: common words (represented by red circles) and neighborhood (represented by blue squares). This approach allows us to not only see whether there are variations in customer reviews due to the distinct features of each district but also compare variations across two cities.

One necessary note: due to the computationally taxing nature of mapping 2-mode text networks, we randomly sampled 10000 reviews within 2023 *from each city* to make figures 3A and 3B.

A. Bangkok B. San Francisco

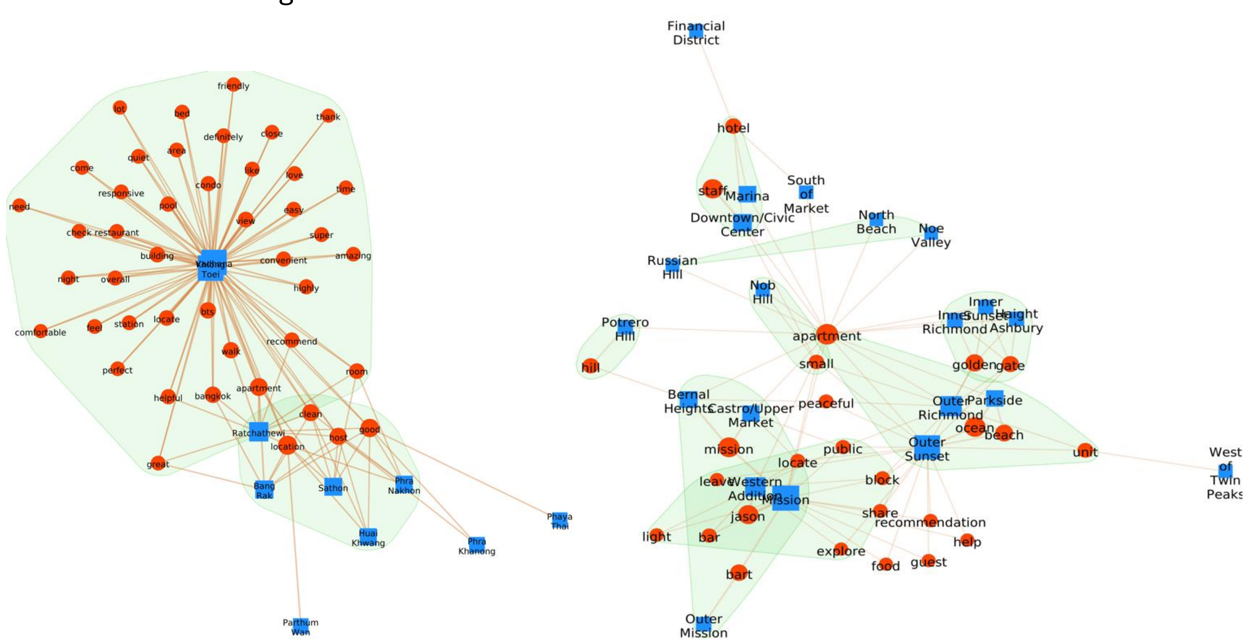

Figure 3. Text Networks of Common Words by Neighborhood in BK and SF

Figure 3A and 3B confirms our expectation that there are indeed variations by neighborhood in each city. Particularly, in Bangkok, there appear to be three distinct clusters of neighborhoods. The cluster receiving the most relevant terms correspond to two central districts in downtown Bangkok – Vadhana and Khlong Toei – suggesting that these two neighborhoods have the most Airbnb listings and/or most guest reviews.

In addition, the two central districts are associated with many positive terms – which can be divided into 3 themes. First, among popular terms are locate, close, bts (BK's subway system), station, and restaurants – all of which suggest that the 2 central neighborhoods have good location, with easy access to the subway system and other public transportation stations as well as many restaurants. Second, Airbnb properties in these two districts also receive terms such as "quiet", "view", "comfortable", and "pool", indicating that these listings have premium amenities such as pools, nice views, and quiet environemnt – all of which give guests comfortable and positive experience. Third, customers also mention words such as "host", "responsive", "helpful", and "friendly". These terms demonstrate that Airbnb hosts in central districts' are helpful, responsive, and friendly. Overall, with central locations, premium amenities, comfortable environments, and good hosts, it is not unsurprising that Airbnb listings in Vadhana and Khlong Toei receive the most positive terms, including "love", "amazing", "perfect", and "super".

The second cluster in Bangkok are composed of what we call "good neighborhoods", which are associated with terms such as "location", "good", "host", "clean", and "recommend". Such mentions suggest that the second cluster still have good and essential features that customers care about (as suggested by the previous word clouds), only that they may not have as many premium amenities or as central locations compared to the first cluster. However, this also means that Airbnb properties in these neighborhoods can be good deals for travellers.

Finally, the third "cluster" actually includes 3 isolates in the Bangkok network – Phaya Thai, Phra Khanong, and Parthum Wan. These neighborhoods are associated with not as much positive terms. This indicates a lower quality and popularity of these three neighborhoods.

While neighborhoods in Bangkok are partitioned into three clear-cut clusters based on their location and amenities, the clustering in San Francisco are based less on material features but instead more on their distinct "vibes" and characteristics. In addition, the clusters in SF appear to be less clearly separated and rather overlapping with each other.

Specifically, there are eight clusters of neighborhoods, each with a distinct set of relevant terms. However, three main clusters can be identified. The first cluster includes the central and financial districts, namely Downtown/Civic Center, Marina, South of Market, and Financial District. Interestingly, these neighborhoods are mostly associated with the terms "hotel" and "staff", suggesting that most Airbnb guests mention their experience being in hotels and interacting with hotel staff. On the other hand, the second cluster consists of neighborhoods such as Outer Sunset, Outer Richmond, Outer

Sunset, and many more – all of which are associated with terms like “ocean”, “beach”, and “Golden Gate”. This shows that the Airbnb listings in these neighborhoods can be well-suited for nature-lovers who enjoy the beaches, ocean, and the Golden Gate. The third main cluster include districts such as Western Addition and Mission, with the most commonly associated words being “bar”, “light”, and “locate” – indicating an area suited for travelers who want to enjoy bars and night life.

Interestingly, both the second and third cluster of SF are also associated with terms such as “explore”, “food”, “recommendation”, which suggests the two areas to have good location, are accessible by public transportation, and have many food options and tourist attractions for people to explore.

While text networks revealed interesting variations by clusters of neighborhoods across the two cities, we are aware that most reviews tend to be positive. As such, we mainly know about positive things mentioned by guests across districts – not what they complain about. To deal with this issue, we picked 5000 reviews with the lowest VADER scores in 2023 and perform the same text network analysis to explore customers’ complaints across different neighborhoods.

Negative Themes by Neighborhood Using Text Networks

While text networks revealed interesting variations by clusters of neighborhoods across the two cities, we are aware that most reviews tend to be positive. As such, we mainly know about positive things mentioned by guests across districts – not what they complain about. To deal with this issue, we picked 5000 reviews with the lowest VADER scores in 2023 and perform the same text network analysis to explore customers’ complaints across different neighborhoods.

A. Bangkok B. San Francisco

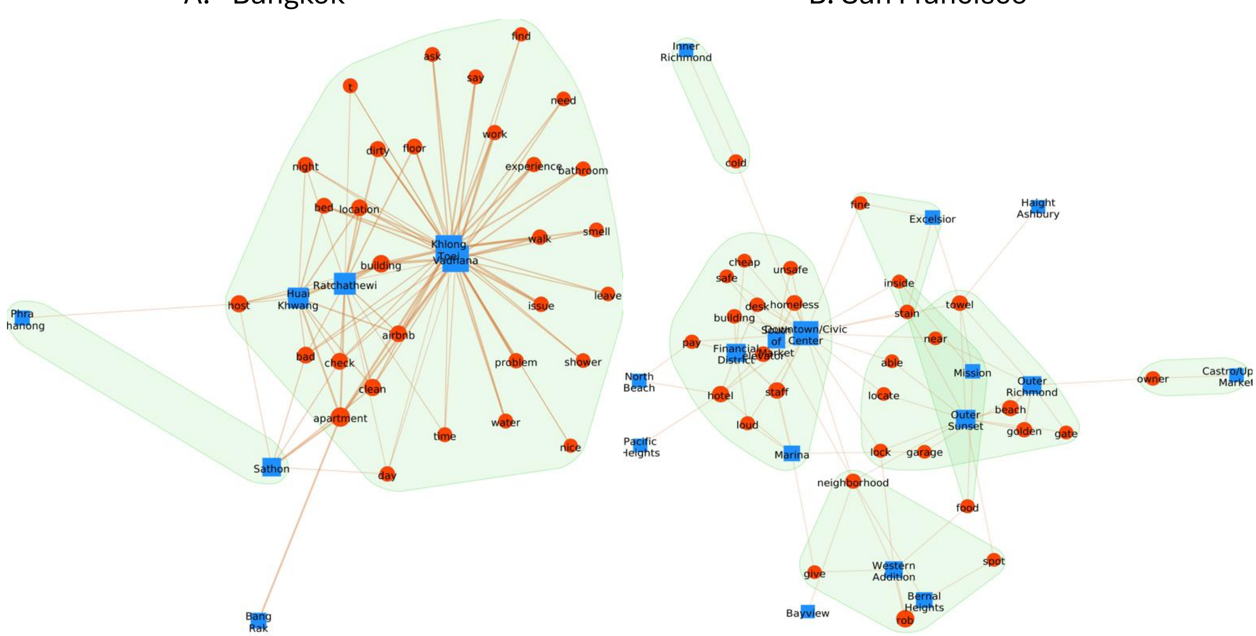

Figure 4. Two-Mode Text Network for Most Negative Customer Reviews in BK and SF

Similar to figure 3A, figure 4A shows a very clear-cut network in Bangkok. In fact, the network for negative comments is even simpler, with only 2 main clusters. While the two central neighborhoods – Vidhana and Khlong Toei – receive the most love and praises, they also receive the most complaints. This is understandable due to their popularity: the more people these neighborhoods attract, the more complaints they have. Of those negative comments, guests mentioned “problem”, “smell”, “dirty”, “water”, and “clean”. This means that most complaints come from guests’ concern for hygiene and cleanness. Similar complaints are recorded for Acathexia and Hua Khwang – 2 good neighborhoods in figure 3A. Some other districts like Pura Khanong, Sathon, and Bang Rack have few complaints about either hosts or the apartments themselves.

For San Francisco, figure 4B again shows a more complex network structure. However, most of the notable concerns relate to the central downtown districts – guests complain that these areas are “loud”, “unsafe”, and have a huge presence of “homeless” people. Some similar safety concerns are spotted in Western Addition and Bernal Heights (the night light districts discussed in figure 3B) as some customers mentioned the term “rob”. Otherwise, some other areas also receive guests’ complaints on hygiene issues such as “stain” and “towels”.

Regression Analysis

While the analyses so far have shown interesting patterns, they are mostly descriptive observations unable to answer one of our guiding questions: which factors drive customers feedback on Airbnb?

To answer this question, we define customer feedback in the form of 2 outcome variables – guest rating (1) and guest review sentiment (2) – and use regression models to analyze the effect on different variables/features on these 2 outcomes.

However, we do not apply a separate linear regression model for each outcome variable. This is because customer ratings and review sentiment are simultaneously determined. In other words, we argue that each Airbnb guest decides the numeric rating and the written review at the same time. As such, predicting one variable using the other as a predictor will create simultaneity bias (a form of endogeneity bias), while completely leaving out the other outcome variable will also create omitted variable bias.

To address this problem, we opt to use a structural equations model – in this case, the seemingly unrelated regression model (SUR) – to analyze the data. Specifically, the SUR model allows two simultaneous equations (with two different dependent variables) to be estimated in parallel. These two equations are connected by having their error terms correlated. Having this model design enables us to account for the two parallel processes that influence customers’ numeric rating and written reviews.

For this analysis, we employ all available observations in the dataset for both cities. For each model, we use around 20 variables as predictors/regressors, which belong to one of the three categories:

explanatory variables relate to the host, neighborhood, or the listing itself. The following are a list of descriptions of the variables:

- Host-related variables:
  - Host Info Available: A binary variable indicating whether the host has an introduction/biography about him/herself.
  - Host Info Sentiment: A VADER score of the sentiment of the host biography (if any). If the host does not have a biography, we substitute the missing values with 0.5 (as we assumed this to be a more neutral value).
  - Host Experience & Host Exp squared: We include a count of months for how long one has been a host, as well as a quadratic term for this count. This is because we suspect more experienced hosts will have better customer feedback, but this effect is not linear but rather an inverted U-shape curve. For example, a host with 48 months (about 4 years) of experience may have a much higher customer ratings and positive reviews compared to a host with only 2 months, but not much higher compared to one with 40 months (about 3 and a half years) of experience.
  - Super host: a binary of whether a host has a super-host title.
  - Host's Total Listings and Total Houses: A host's total number of listings on Airbnb (may include separate rooms/apartments of a same house) and total number of houses on Airbnb (only count separate houses).
  - Host Responds in a few Hours: A binary indicator of whether a host responds quickly (within a few hours).
- Neighborhood-related variables:
  - Neighborhood Description Available: A binary indicator of whether an Airbnb listing has a description of the surrounding neighborhood.
  - Neighborhood Description Sentiment: the sentiment of the written description of the neighborhood (to measure how positive a host writes it), measured by VADER score. Missing values are replaced with 0.5.
- Listing-related variables:
  - Price: Price per night of a listing (in dollars)
  - Occupancy: A count of maximum occupancy to capture the size of the listing.
  - No. of baths: Count of bathrooms.
  - Availability (last 30 days): Captures a listing's (lack of) popularity by using the number of nights (in the last 30 days) that the listing is available for booking.
  - Room Type – Entire House & Room Type – Private Room: Each variable is a binary indicator showing whether a listing is an entire house or a private room (the baseline category includes room types that are hotel rooms or shared rooms).
  - Number of Reviews: Another measure of the popularity of the listing using the total number of reviews (up until that observation).

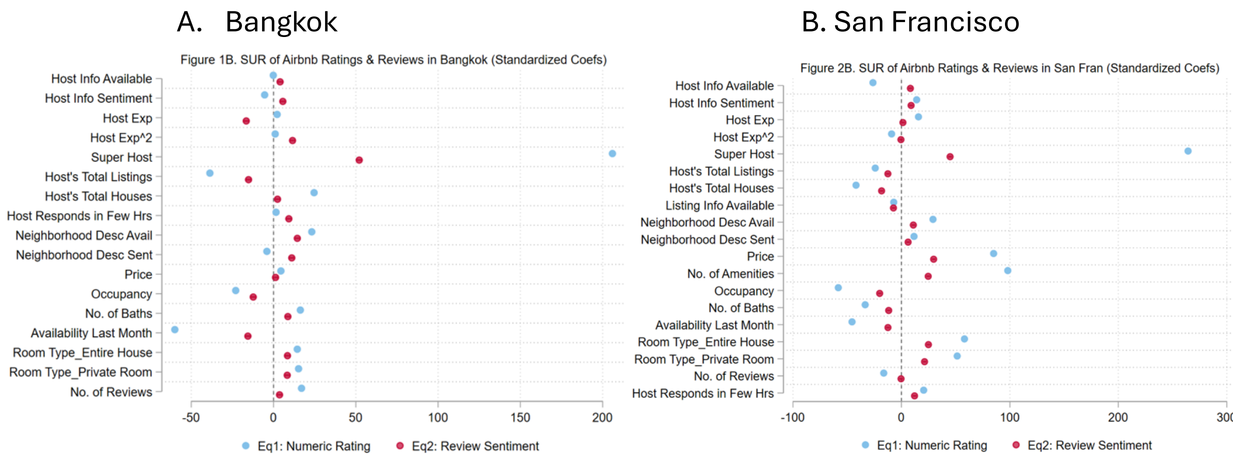

Figure 5. Standardized Coefficients of Seemingly Unrelated Regression in BK and SF

In Figure 5 above, we observe coefficient plots from the SUR model for Bangkok and San Francisco individually. Our focus lies in understanding the patterns shown by certain variables across both equations, and comparing their effects in two cities through the model outputs. The coefficients of variables impacting numeric ratings are depicted in blue, while those for review sentiment are represented in red. The positioning of the blue or red dots relative to the 0 line indicates negative or positive coefficients, while their distance from the line signifies the magnitude of these coefficients.

We begin our analysis by analyzing host-related variables' impact on both numeric ratings and review sentiment. Notably, the coefficients for variables host info available and host info sentiment are positive in the review sentiment equation for both cities. This suggests that a positive host information section contributes favorably to customer reviews. However, the effects of these variables on numeric ratings vary between the two cities, presenting a complexity that needs further investigation. Moving on, we delve into the coefficients related to host experience and its quadratic term count to assess their influence on ratings and reviews. In San Francisco, the findings regarding numeric ratings align with our hypothesis, indicating a positive impact of host experience, but up to a certain threshold (the quadratic term has negative coefficient). However, discrepancies emerge in Bangkok and for the review sentiment equation in San Francisco, suggesting underlying patterns we have yet to fully understand.

The superhost variable stands out with the most significant positive coefficient across both equations and cities, emphasizing the critical role of being a good host in driving customer satisfaction.

Transitioning to listing-related variables, we find that a positive neighborhood description section tends to enhance customer reviews and ratings in both cities, as evidenced by their positive coefficients. Similarly, the variable price shows a positive impact in both locations, particularly in San Francisco with the larger coefficient magnitude. While we initially found this finding counterintuitive, as we expect customers to prefer properties with more affordable price. We suspect this effect stems from a selection bias, wherein higher-priced rooms often offer superior quality and amenities, thus fostering a more positive guest experience.

Lastly, we examine the impact of the number of reviews a listing has. In Bangkok, the positive coefficients for both equations align with our intuition, suggesting that listings with more reviews are more highly valued. However, in San Francisco, this variable either lacks significance or negatively impacts customer ratings and reviews, presenting a notable departure from the expected trend.

We only focus on analyzing a subset of variables deemed significant or displaying intriguing patterns in this SUR model. While similarities exist between Bangkok and San Francisco, such as the positive influence of superhost status and higher-priced rooms, differences also arise, likely stemming from unaccounted underlying factors. Nonetheless, we believe our findings from the regression analysis still reveal some insights about the factors driving customer ratings and review sentiments, which is beneficial for hosts seeking to enhance guest experiences.

Conclusion

Airbnb has developed as an essential option for accommodation for travelers across the world. As customers on this platform, feedback from other customers is essential in how we decide on which listing(s) to stay at. Yet, many of the dynamics that shape customer reviews and ratings are unclear. In this project, we investigate patterns in customer reviews and factors influencing customer feedback (both in numeric ratings and written reviews), as well as compare these patterns across Bangkok and San Francisco – two major metropolitans and tourist destinations of the world.

Overall, the analyses in this project have revealed interesting patterns. First and foremost, reviews are often positive in both places. We believe that this is at least partly due to a selection effect as customers who write reviews and give ratings tend to be those to have positive experiences with Airbnb listing. Second, guests also mention similar aspects of an Airbnb property in their reviews – such as cleanness, good host(s), good location, and/or premium amenities. These two patterns are similar across Bangkok and San Francisco. This further confirms the intuition that guests, regardless of where they are, tend to care about a few fundamental features of an Airbnb accommodation. Third, the geography within a city's urban layout also influences guest feedback – in other words, an Airbnb listing's neighborhood affects guest experience.

Next, we also found that how a host describes & advertises about him/herself and his/her property matters for customer feedback. Particularly, more positive host biographies & neighborhood descriptions tend to positively affect customer feedback in both cities.

In addition, the findings also confirm the apparent assumption that being a good host is important. This is demonstrated by not only the qualitative word clouds and text networks but also the quantitative regression analysis: super host (a title that represents the quality of a host) is the single most powerful predictor of the customer rating & review sentiment.

This project also shows some interesting differences between Airbnb in Bangkok and San Francisco. First, Bangkok neighborhoods are mostly distinguished by their amenities & features – the material factors, whereas San Fran neighborhoods appear to have more distinct characteristics or "vibes" (there are beach/nature districts, nightlife/bar neighborhoods, and downtown areas, etc.) and thus are harder to be fully distinguished.

The analyses also demonstrate subtle differences in the broader context and historical development in which each city is embedded. For example, while homelessness is not mentioned in Bangkok's Airbnb reviews, it is a common concern for guests in San Francisco, which reflects a broader problem faced by not only SF but also many other major cities in the US.