Jinyan Xiang

Sep 27 · 6 min read · Listen

Save

Predicting the Airbnb Listing Price

A data-based approach of over 200 thousand Airbnb listings in 31cities and 19 states in the United States

Introduction to Airbnb

Airbnb was founded in 2008 and shortly grew into a unicorn in the accommodation sharing market and, more broadly, the sharing economy. After a decade of development, Airbnb operates in over 220 countries and regions and over 100,000 cities with 5.6 million active listings and 150 million active users. More importantly, it has become a significant participant in local economies; for example, the host-guest community generated over $117 billion in economic impact across 30 cities in 2019 (Meyer, 2022).

Why Predicting Listing Price

So, it is not surprising to see burgeoning research on Airbnb, and **listing price** is always the center of such research because it is so critical to the platform, hosts, and guests

the platform's revenue, and the health of the host-guest sharing community and local economies.

As a researcher who studies the engagement in sharing economy, I am also curious about antecedents of listing price, which could have a profound influence on hosts' and guests' long-term engagement in accommodation sharing (not limited to Airbnb) and the development of the sharing economy.

What Could Predict Listing Price

In other words, I ask one question here: what predicts the listing price on Airbnb? This question can be boiled down to three questions: how do 1) environment-related factors (e.g., regional difference), 2) lodge-related factors (e.g., lodge capacity, lodge type), and 3) host-related factors (e.g., host tenure) affect listing price?

What Predicts Listing Price

To investigate these questions, I constructed a dataset (*source: Kaggle*) that contains information about 236,823 unique Airbnb listings (*active in September 2022*) in 31 cities and 19 states in the United States.

I. Environment-related: substantial regional difference

There is a substantial difference in listing price across cities. Seaside cities (i.e., Rhode Island, Santa Cruz, Hawaii, Pacific Grove, and San Diego) have a significant price premium compared to urban cities (e.g., Los Angeles, New York City). Specifically, the average listing price in seaside cities is $351, which is about 50% and 65% more than the average listing price across all cities ($237) and in urban cities ($213). Such price premium in seaside cities is presumably due to the views premium and the high tourism-driven demand.



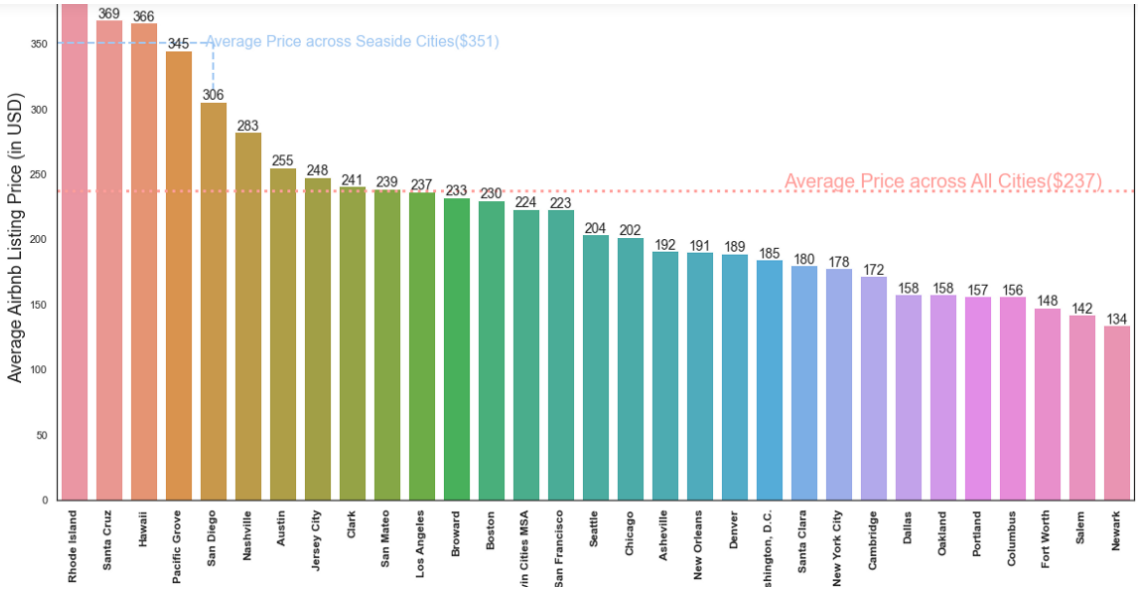
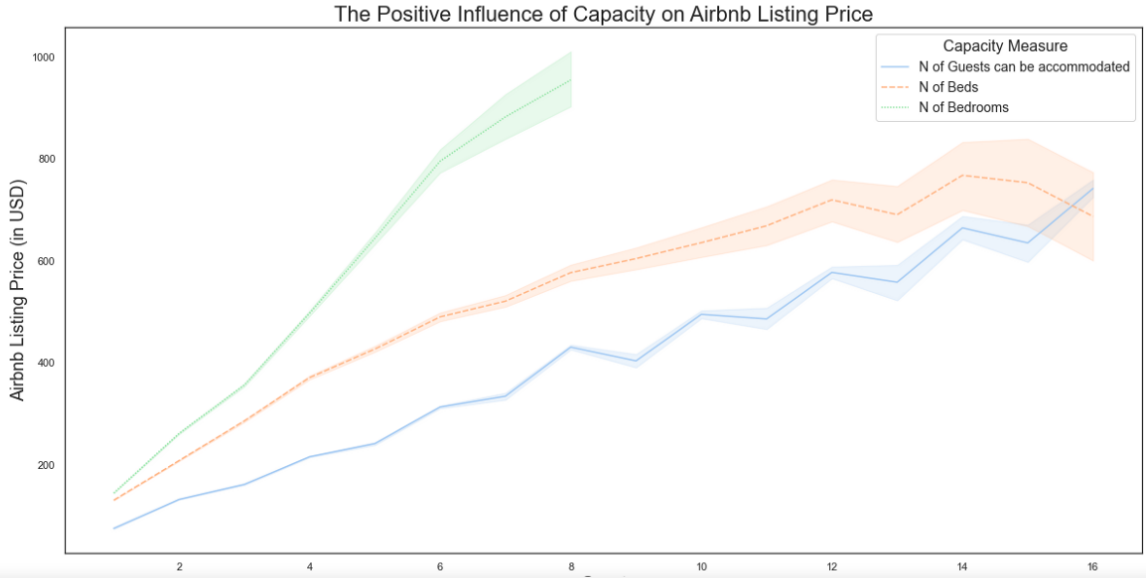



Figure 1: Influence of Cities on Listing Price

II. Lodging-related: strong positive influence of capacity

Capacity strongly influences listing price: the larger the capacity of a lodge in terms of the maximum number of guests that can be accommodated, the number of beds, and the number of bedrooms, the higher the listing price. In addition, the maximum number of guests has the most decisive influence, followed by the number of beds and bedrooms.



III. Lodging-related: significant difference in property and room types

Property type and room type are also critical categorical predictors. The average listing price for villas ($526) is more than double that across all property types ($237). On the other hand, the average listing price for guest suites ($136) is about half of that across all property types. As for room type, privacy matters. Shared rooms ($71) value only one-fifth or about one-fourth of hotel rooms ($363) or entire homes/apartments ($275).

Figure 3(a): Influence of Property and Room Types on Listing Price

Such property types or/and room types could be partly attributed to capacity differences. For example, villas — the most expensive type of property — can accommodate seven guests on average, while this number for all property types combined is only four.

To explore this speculation, I calculate the average listing price per guest, similar to most property and room types. The diminishing difference confirms the hypothesis. Interestingly, hotels and hotel rooms have prominent price-per-guest premiums compared to other property types and room types, which cannot be attributed to capacity.

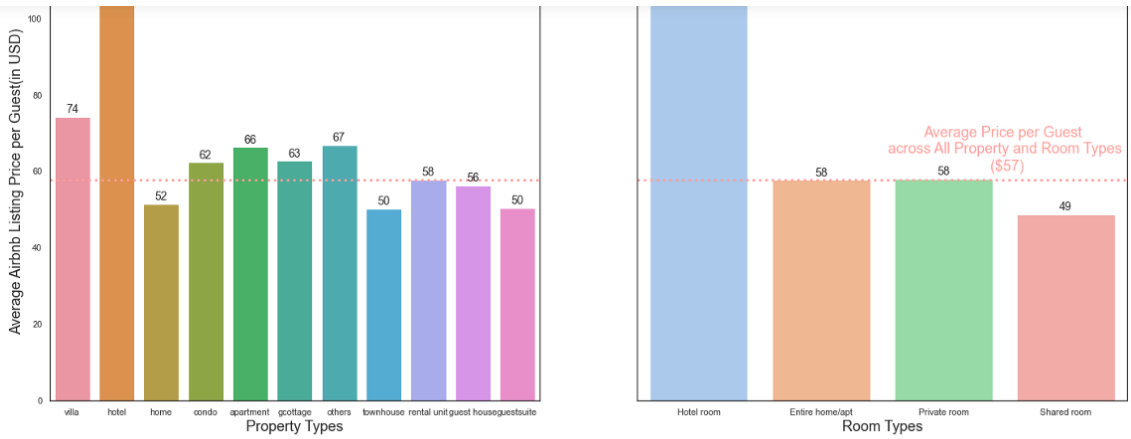


Figure 3(b): Influence of Property and Room Types on Listing Price per Guest

IV. Lodging/Host-related: varied influence of ratings

Over 97% of listings have over 4 out of 5 ratings in each of the seven rating dimensions (i.e., overall, accuracy (of the description), check-in, cleanliness, communication, location, value). Due to the low variance, the influence of ratings on listing price is challenging to detect in general. The overall rating has a weak positive impact on price.

However, some aspects of the listing or host manifested in specific rating dimensions have stronger influences than others. For example, lodge location is the most positive predictor of price, followed by cleanliness. In contrast, the perceived value of the lodge is the most negative predictor of price, followed by check-in. Communication and accuracy — the host-related aspects have negligible influence on price. The fact that location is the strongest predictor also implies that views or location premium of seaside cities could explain listings' high price there.

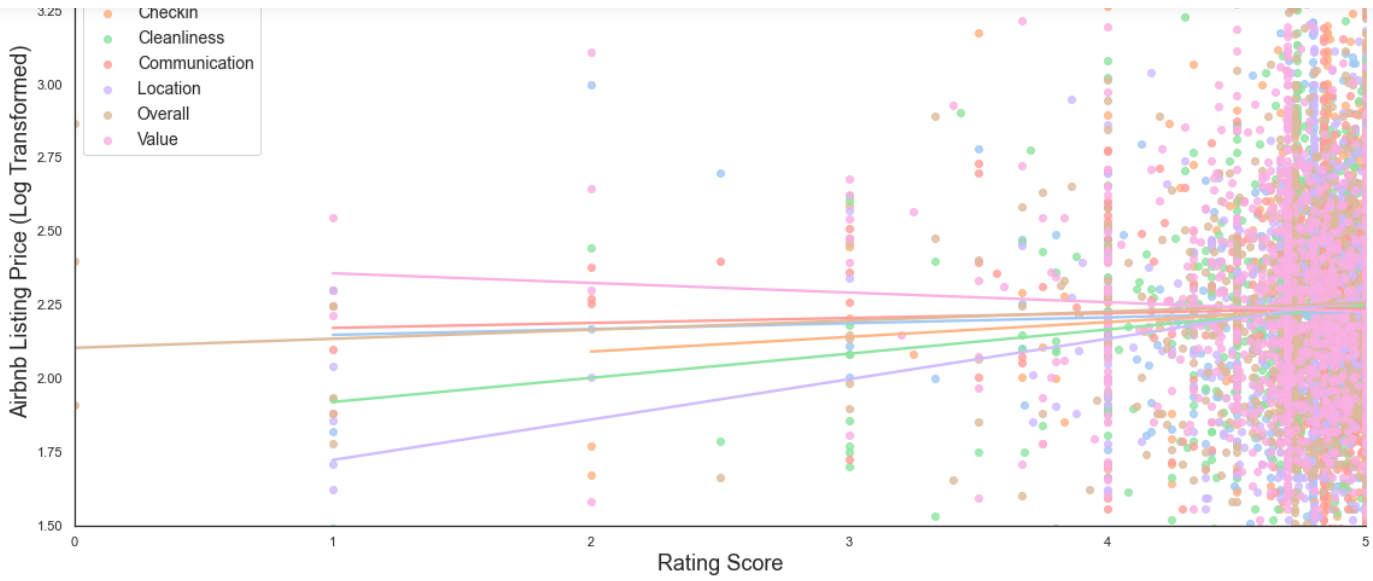


Figure 4: Influence of Ratings (by dimension) on Listing Price

V. Host-related: weak positive influence on host tenure and host badges

Surprisingly, host-related factors are the weakest predictors of listing price, compared to environment-related and lodge-related factors. Hosts' tenures (i.e., the experience of being an Airbnb host in terms of years) and their badges (i.e., whether they are super hosts, whether their identity is verified) can only slightly increase the listing price (price difference is less than $10), partly because the majority of hosts are NOT super host (66.5%) but have their identity identified (84%).

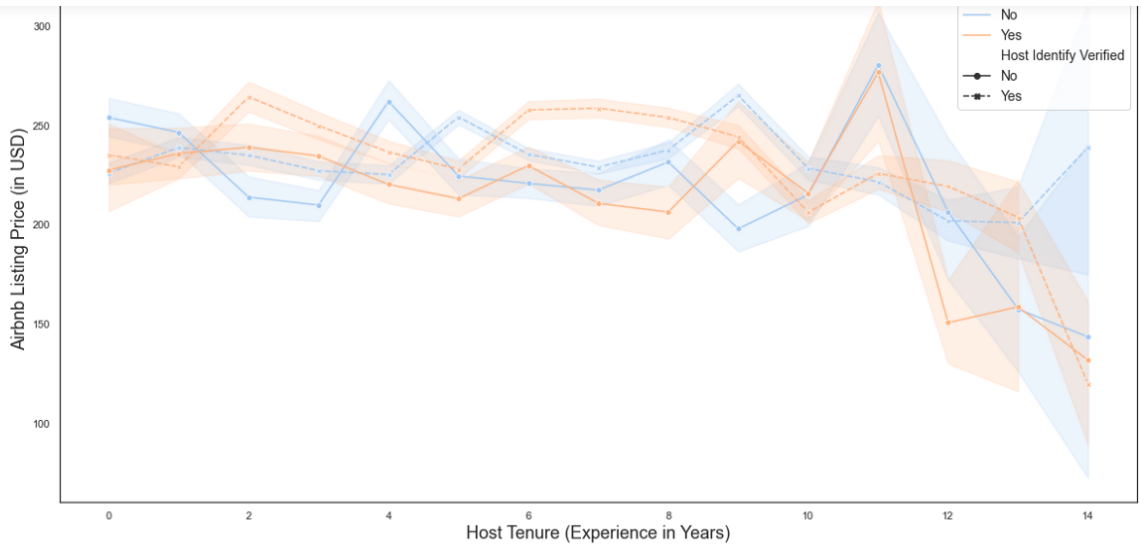


Figure 5: Influence of Host Tenure and Badges (super host, identity verification) on Listing Price

Conclusion

Using a rich Airbnb listing dataset (N = 236,823 unique in 31 American cities and 19 states), I explored the determinants/predictors of listing price from three facets: environmental, lodge, and host. **Key takeaways are**:

1. There is a substantial regional difference: listings in regions that have views premium or/and are tourism hotspots (e.g., seaside cities) are more expensive than in other areas.

2. Lodge capacity is the strongest predictor of listing price: the more guests a lodge can accommodate, the more expensive the listing is.

3. Although listing price differs across property and room types, such difference can be partly attributed to the capacity difference. Interestingly, hotels (hotel rooms) have the most significant price premium after controlling for the capacity difference: they are more expensive in terms of listing price per guest.

4. Ratings matter for predicting price, but the weight varies by dimension. Location and perceived value of the lodge matter most, while communication of the host and accuracy of the description have almost no impact.

Caveat: all the conclusions are essentially observational and correlational, and any causal inference must be drawn cautiously.