ORIE 5741 Final Project Proposal

Jiongjiang Duan(jd2253), Sirawich Tippawanich(st895), Jiaqi Zhang (jz2232)

September 29, 2021

1   Dataset Introduction

The data set we shall use for this project was published on Kaggle by a user named Ahmad Bhat. The data set contains Airbnb (a popular home-sharing platform) listings in 10 major cities including information about hosts, pricing, location, room types and over 5 million historical reviews. Among 30 features, there are 11 nominal, 2 ordinal, 5 discrete and 12 continuous variables. There are a total of 180,000 unique hosts who were responsible for 280,000 unique listings in the data set.

2   Questions to explore

Naturally, the following questions can be explored with the data set:

- Do track records of hosts influence their behavior in determining the listing prices? For review metrics to be good incentives for hosts to maintain their standards, a relationship should exist between higher scores and ability to ask for higher prices. Platform managers should assess the effectiveness of the incumbent review system and find a replacement if it is found to be ineffective.

- Does the effectiveness of the review system vary across major cities or neighborhoods within the city? How do they vary across different property tiers? Do owners of high-value properties wield more power in determining listing prices? Do higher number of claimed amenities allow hosts to price higher relative to their review performances?

2.1   Why are these questions important?

Home-sharing platforms should act as moderators between hosts and guests in ensuring that there is a system in place to protect both parties in all transactions. The ability of guests to provide reviews is a powerful tool that can assure fairness between pricing and underlying service the guests receive. Rigorous analysis of the dynamic between reviews and market pricing of properties can help a platform identify anomalies and vulnerability in their policies and governance as a moderator.

2.2   How does the data set help to answer the question

Our data set contains information of the property, customer service feedback, host information and review scores. Given these information, we will be able to evaluate how the review system and host behavior influence the listing price by constructing a regression model. From the regression result, we can find the features that influence the listing prices most and use correlation map to analyze if the review score has an impact on the listing prices. We can then segment the properties by the features we identified in the first question and build a regression model on each of the segment to see how the review system change over different segments.

2.3   How will the insights developed benefit the stakeholder (Airbnb)

To ensure business longevity, platforms such as Airbnb must satisfy both hosts and guests. Guest dissatisfaction can arise from sources such as unfair pricing and inaccurate listing details, both of which the platform is ultimately responsible for regulating. Similarly, to ensure meritocratic-based competition between hosts, the review process should be effective in establishing a market equilibrium between quality and pricing. Insights developed from review data analysis can bring to light areas of improvement in the platforms' operations.

References

https://www.kaggle.com/mysarahmadbhat/airbnb-listings-reviews.