**From Listing to Booking: Predicting a 'Perfect' Airbnb Listing**

Hannie Pham, Pham Duy Anh Vu

Data 400: Capstone of Data Analytics

The rise of short-term rental platforms like Airbnb has transformed the hospitality industry, offering travelers diverse accommodation choices and enabling new income opportunities. Yet, with increasing listing volume and variability, it remains challenging for hosts to optimize their offerings and for guests to identify high-quality stays. While past studies have explored pricing and guest preferences, few have combined exploratory data analysis (EDA) with predictive modeling to systematically define and forecast listing success.

In this study, we integrated EDA and supervised machine learning to identify features that distinguish highly rated Airbnb listings, defined as those with guest ratings above 4.5. Using a dataset of over 4,000 listings from major urban centers, we analyzed factors such as host responsiveness, quality and quantity of amenities, location, and pricing. Our EDA revealed consistent behavioral patterns across successful listings, while machine learning models—including logistic regression and ensemble methods - identified amenities, pricing, and host response rate as the strongest predictors of listing success. Here we show that our models, particularly the Random Forest classifier, can predict successful listings with over 95% accuracy and provide interpretable insights by ranking features based on their contribution to guest satisfaction.

These findings offer both descriptive and predictive value - supporting hosts in improving listings, assisting guests in identifying quality stays, and informing platform designers and policymakers aiming to improve transparency and user experience. Future extensions may include incorporating review sentiment, optimizing pricing strategies, and developing real-time feedback tools for dynamic listing improvement.