Project Update: Predicting Airbnb Accommodation Prices in Bangkok for 2025

Thus far I gathered and loaded data from Airbnb's public datasets, which include over 22,000 listings in Bangkok. Performed initial cleaning, which involved handling missing values, removing unnecessary columns, and converting data types for better processing efficiency. Conducted an in-depth EDA to understand the variables' distributions, identify outliers, and observe correlations between different features such as neighborhood, room type, and price. Extracted useful features that are likely to influence the price, such as the type of room, number of reviews, and specific amenities offered. I tried different regression models to try to predict the price. Thus far, those did not give me reasonable results and I need to do more work with the data before running the models. Some challenges I encountered was that some neighborhoods had significantly fewer listings, which made it difficult to draw reliable conclusions about these areas. Additionally there was substantial variability in prices within the same neighborhood and room type, likely influenced by unobserved factors.

The next steps will be to try the models with different features, to try to get the MSE to be reasonable for bangkok housing prices. I'd like to try binning the prices. I'll also try to predict which neighborhood the home is in as well based on some factors. As well as answer these questions: My main research question is can we predict accomodation prices in bangkok for 2025? Smaller questions:

- What neighbourhood in Bangkok is cheapest?
- Whats the cheapest season for each neighbourhood?
- How does seasons affect the pricing and availability of listings?
- How does customer satisfaction, as reflected by reviews per month, affect the likelihood of a property being booked?
- What are the common characteristics of highly rated listings?
- Which types of properties (apartments, villas, cottages) are most popular among renters?
- How does the number of reviews correlate with the demand for certain listings?
- How do the lengths of stays vary by season and what impact does this have on pricing?
- How do the availability metrics (availability_30, availability_60, availability_90, availability_365) correlate with pricing? Does higher availability lead to lower prices?
- Can seasonal availability data predict pricing trends for peak and off-peak seasons?
- Which neighborhoods show the highest variability in pricing within the same house type?
- How do various review scores (cleanliness, check-in, communication) affect the price and booking frequency of listings?
- Does the overall review score rating correlate with the price fluctuations throughout the year?
- Can you use the features such as number of reviews, review scores, and host information to predict future pricing trends or the probability of a listing being booked?