Project Proposal

Team lol - Tamsin Connerly, Hannah Lee, Jasmine Xiang

```
library(tidyverse)
library(tidymodels)

airbnb <- read.csv("data/NYC-Airbnb-2023.csv")
```

Introduction

The rise of short-term rental platforms, particularly Airbnb, has significantly disrupted the traditional hospitality industry and transformed urban housing markets worldwide. In New York City, one of the world's most popular tourist destinations, the impact of Airbnb has been particularly pronounced, raising questions about its effects on local communities, housing affordability, and the broader urban economy.

Previous research has identified several factors that impact Airbnb pricing. Wang and Nicolau (2017) found that host attributes, site and property attributes, amenities and services, rental rules, and online review ratings all play significant roles in determining listing prices (Wang and Nicolau 2017). Furthermore, recent studies have provided evidence of Airbnb's influence on housing markets. Barron et al. (2020) found that a 1% increase in Airbnb listings leads to a 0.018% increase in rents and a 0.026% increase in house prices(Barron, Kung, and Proserpio 2018). This effect is more pronounced in areas with a lower share of owner-occupiers, suggesting that non-owner-occupiers are more likely to reallocate their properties from long-term to short-term rentals.

Our research question is: "How do various factors, such as bedroom/bathroom number, accommodation capacity, room type, review scores, and distance from city center, influence the price of Airbnb listings in New York City?"

Understanding the determinants of Airbnb pricing in New York City is crucial for several reasons. Firstly, it can provide valuable insights for policymakers grappling with the challenges posed by the growth of short-term rentals, including potential impacts on housing affordability and neighborhood character(Toader et al. 2021). Secondly, it can help hosts make more informed pricing decisions, potentially leading to more efficient market outcomes.

Based on existing literature and our understanding of the New York City housing market, we hypothesize that:

- Listings with more bedrooms and bathrooms will command higher prices, reflecting the premium placed on space in urban environments.
- The number of guests a property can accommodate will positively correlate with price, as larger groups often have higher budgets.
- The type of room (entire home/apartment vs. private room) will significantly impact pricing, with entire homes/apartments having a higher price.
- Higher review scores will be associated with higher prices, as positive feedback may justify premium pricing.
- Properties closer to the city center will be priced higher due to their convenient location and proximity to attractions.
- The impact of these factors on price may vary across different boroughs or neighborhoods.

Data description

...

Exploratory data analysis

...

Analysis approach

Potential Predictor Variables of Interest:

Our potential predictors capture many key aspects of Airbnb listings — space, amenities, perceived quality, and location—that are likely to have significant influence on nightly rates.

- Number of Bedrooms (quantitative): Reflects the amount of sleeping space in a listing. We anticipate more bedrooms generally commanding higher prices.
- Number of Bathrooms (quantitative): Indicates the level of comfort/convenience. Listings with more bathrooms are typically priced higher.
- Accommodation Capacity (quantitative): Represents how many guests can stay in the listing. We expect larger capacity to have higher nightly rates.

- Room Type (categorical): Classified primarily as "Entire home/apt," "Private room," or "Shared room." We plan to use one-hot encoding, as room type strongly affects price (entire units usually cost more).
- Review Scores (quantitative): Represents overall rating (on a typical 1–5 scale, or 1–100 in some datasets). Higher-rated listings may set premium prices.
- Distance from City Center (quantitative): Measures proximity to major landmarks or downtown areas. We expect closer listings to charge higher prices.

Regression model technique:

Our responsible variable is price, which is a continuous variable, and we will employ a Multiple Linear Regression model to estimate the effect of each predictor on price while controlling for the others.

Model specification:

$$Price = \beta_0 + \beta_1(Bedrooms) + \beta_2(Bathrooms) + ... + \epsilon$$

Assumption checking:

- Linearity: The relationship between predictors and our response variable price should be approximately linear.
- **I**ndependence of Residuals: We assume each listing is an independent observation.
- Homoscedasticity: Residuals should have constant variance across all fitted values.
- Normality of Residuals: Residuals should be roughly normally distributed.

Model evaluation:

In evaluating our MLR model, we will begin by examining the R^2 and Adjusted R^2 value. They can help us understand how well the model explains the variability in price while adjusting for the number of predictors used. We will also look at p-values for each predictor, which will indicate which variables have a significant impact on the listing's price.

Feature engineering:

Depending on the distribution of our predictors and response variable, we may consider examining the potential for interaction terms, particularly if we hypothesize that the effect of one variable depends on another. For example, the impact of accommodating additional guests may differ significantly if the listing is an entire home versus a private room.

In addition, we plan to carefully encode categorical variables, such as Room Type, using dummy variables. This will allow the model to compare and contrast distinct categories effectively. We plan to capture the detailed relationships among our features, thus producing more robust, interpretable insights into Airbnb pricing.

Data dictionary

The data dictionary can be found here.

Barron, Kyle, Edward Kung, and Davide Proserpio. 2018. "The Sharing Economy and Housing Affordability: Evidence from Airbnb." *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3006832.

Toader, Valentin, Adina Letiţia Negruşa, Oana Ruxandra Bode, and Rozalia Veronica Rus. 2021. "Analysis of Price Determinants in the Case of Airbnb Listings." *Economic Research-Ekonomska Istraživanja* 35 (1): 2493–2509. https://doi.org/10.1080/1331677x.2021.1962380.

Wang, Dan, and Juan L. Nicolau. 2017. "Price Determinants of Sharing Economy Based Accommodation Rental: A Study of Listings from 33 Cities on Airbnb.com." *International Journal of Hospitality Management* 62 (April): 120–31. https://doi.org/10.1016/j.ijhm.2016.12.007.