AIRBNB IN NYC

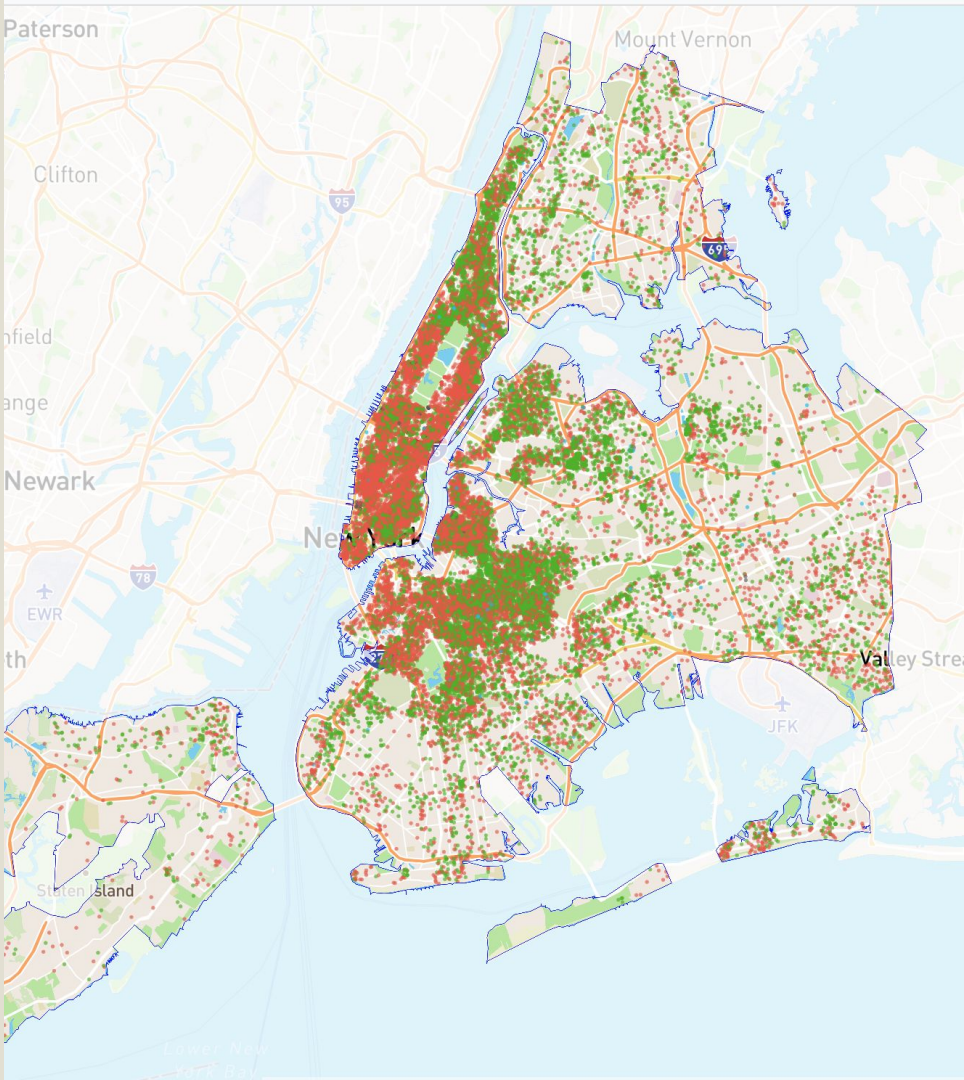
TEAM LOL: TAMSIN, JASMINE, HANNAH

INTRO

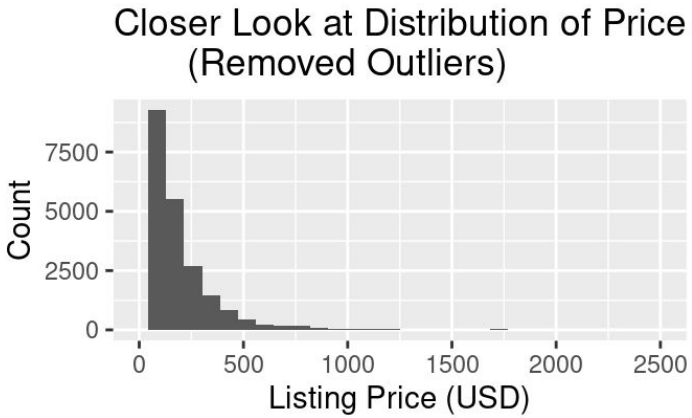
- ➔ Subject: Airbnb prices in NYC
- ➔ Research Question: How do various factors, such as bedroom number, room type, review scores, and neighborhood, influence the price of Airbnb listings in New York City?
- ➔ Motivation: Airbnb has transformed the travel industry

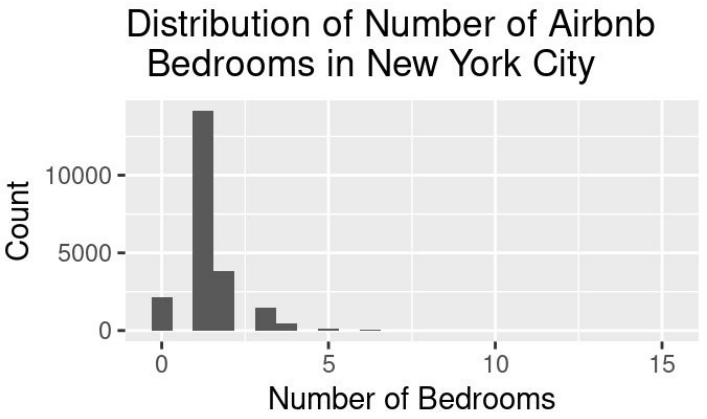
Dataset

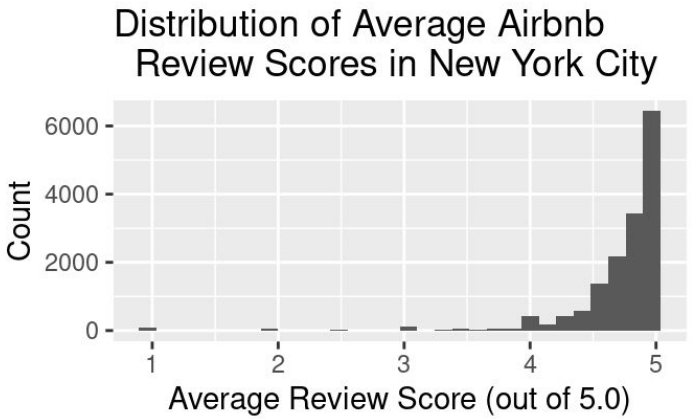
- From InsideAirbnb on March 1, 2025
- 58 variables
- 22,308 observations
- Focusing on price, bedroom number, room type, review score, and neighborhood



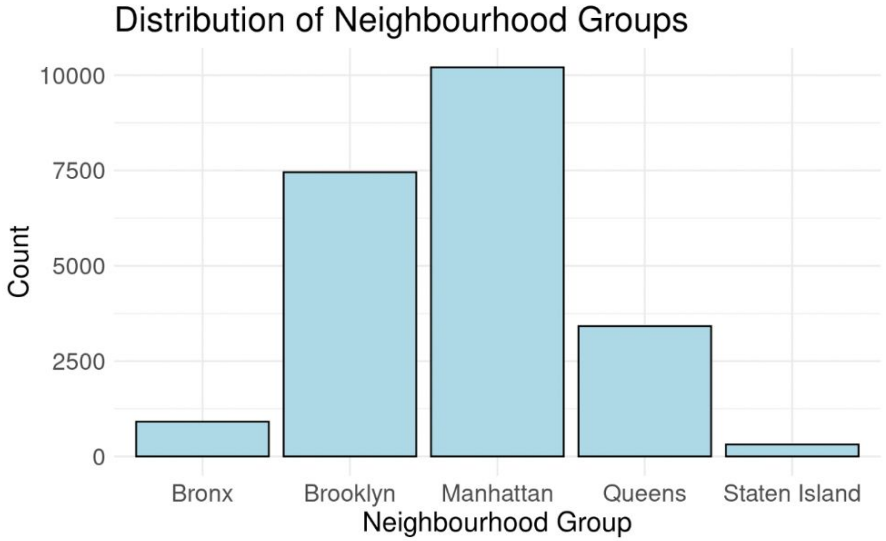
Univariate EDA

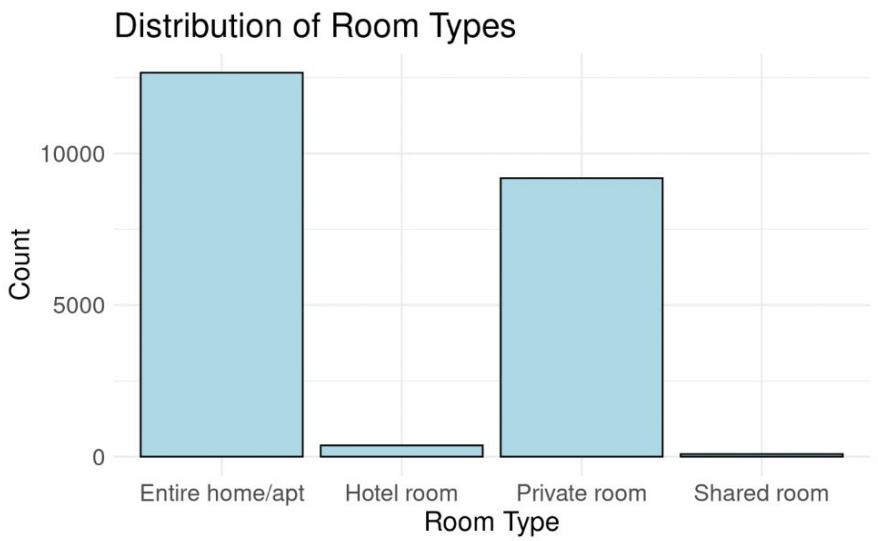




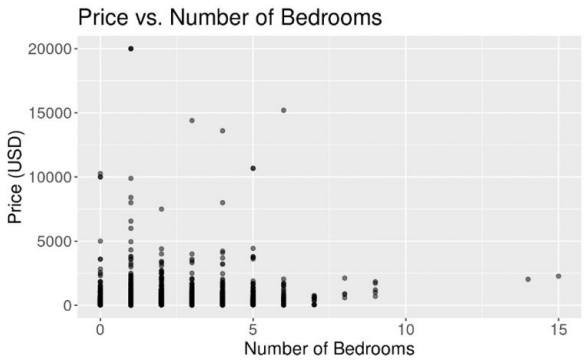


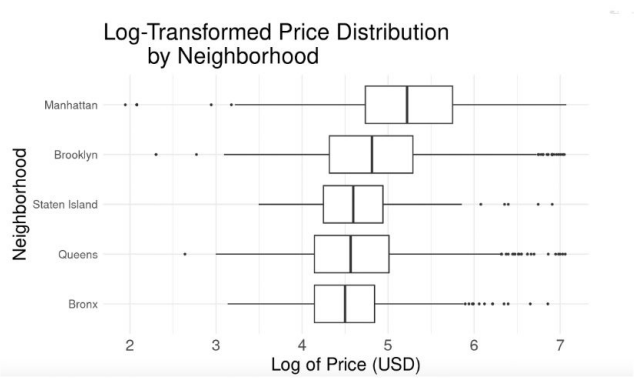
Univariate EDA

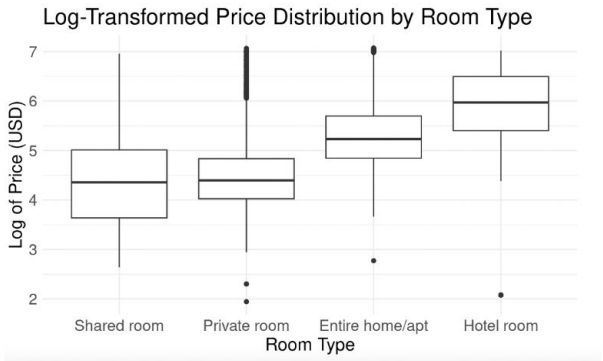




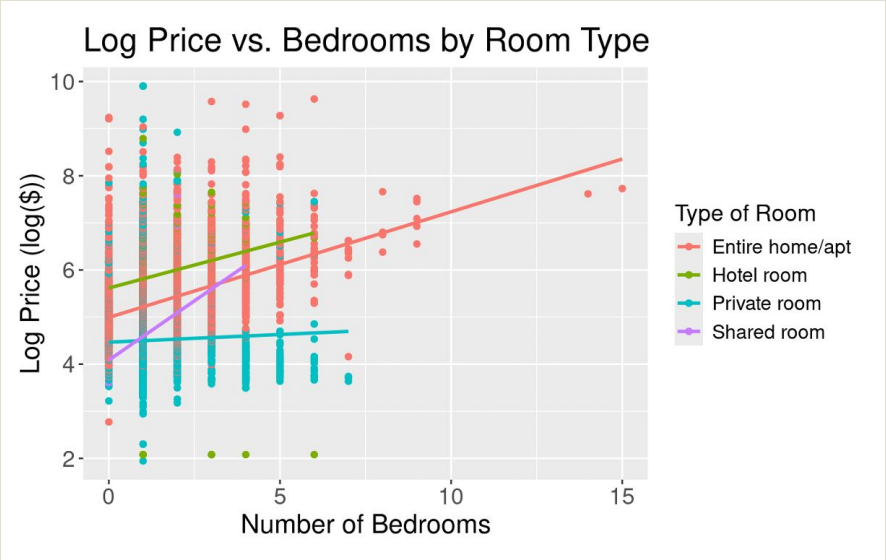
Bivariate EDA

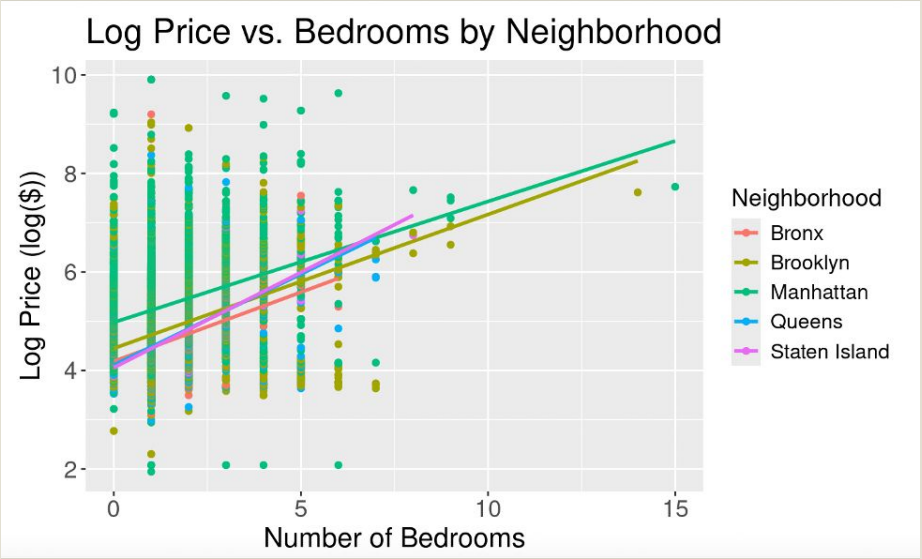






Interaction Effects





Initial Modeling Strategies

- **Model Choice:** Multiple Linear Regression (MLR)

  Response variable: log-transformed price
  Predictors: Bedrooms, Bathrooms, Accommodation Capacity, Room Type, Review Score, Neighborhood, Distance from City Center

- **Strategies:**

  Apply log(price) as the response and fit multiple linear regression models with and without interaction terms

  Evaluate model fit and compare models using Adjusted R^2, residual plots, and p-values

  Check assumptions: linearity, normality, homoscedasticity, independence

- **Mathematical Expression of the Model:**

$$\log(\text{Price}) = \beta_0 + \beta_1 \cdot \text{Bedrooms} + \beta_2 \cdot \text{Bathrooms} + \beta_3 \cdot \text{Capacity} + \beta_4 \cdot \text{Review Score} + \beta_5 \cdot \text{RoomType\_Private} + \beta_6 \cdot \text{RoomType\_Shared} + \beta_7 \cdot \text{Neighborhood\_Brooklyn} + \cdots + \varepsilon_i$$

Next Steps

- Finalize feature selection and handle remaining missing data
- Fit and compare multiple linear regression models:
  - ❖ With vs without interaction terms
  - ❖ Evaluate using Adjusted R^2, residual plots, p-values
- Conduct full assumption checks (linearity, homoscedasticity, normality, independence)
- Investigate multicollinearity and consider dropping or combining correlated predictors if necessary