

# Exploring Price Determinants of Airbnb Listings in New York: A Comparative Analysis of Mixed Effects and Linear Regression Models

Yuqing(Sophie) Deng

## Abstract

This study examines how location, property features impact Airbnb prices in New York City. By analyzing listing data and employing various models, including EDA and multilevel mixed-effects models, the research seeks to understand pricing trends and inform strategic decisions for hosts and guests in the Airbnb marketplace.

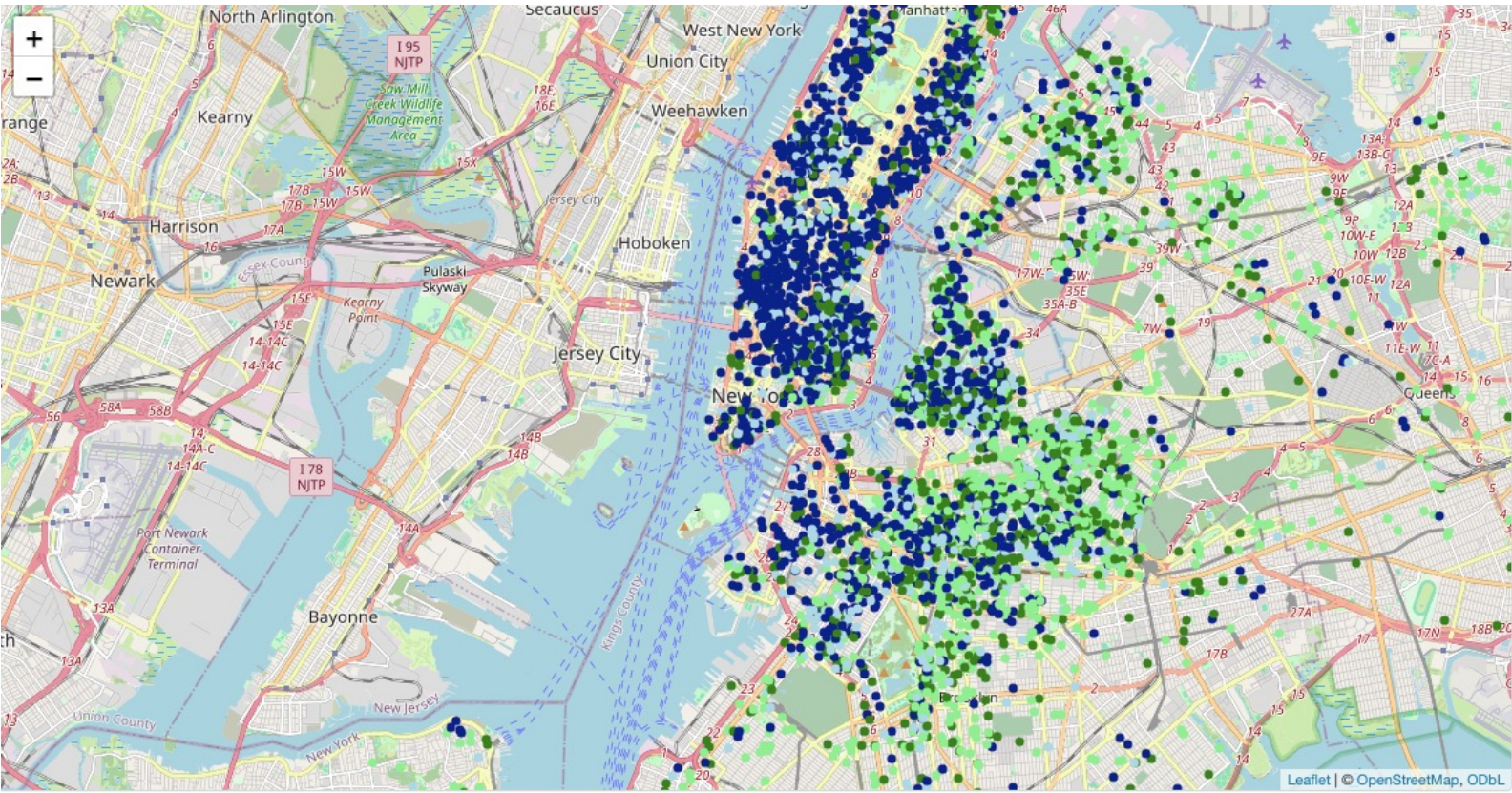
## Introduction

Airbnb has reshaped urban accommodation, especially in New York's diverse landscape. This study examines the factors influencing Airbnb prices, with a focus on location and property details, to aid in understanding of the platform's pricing dynamics.

## Hypothesis

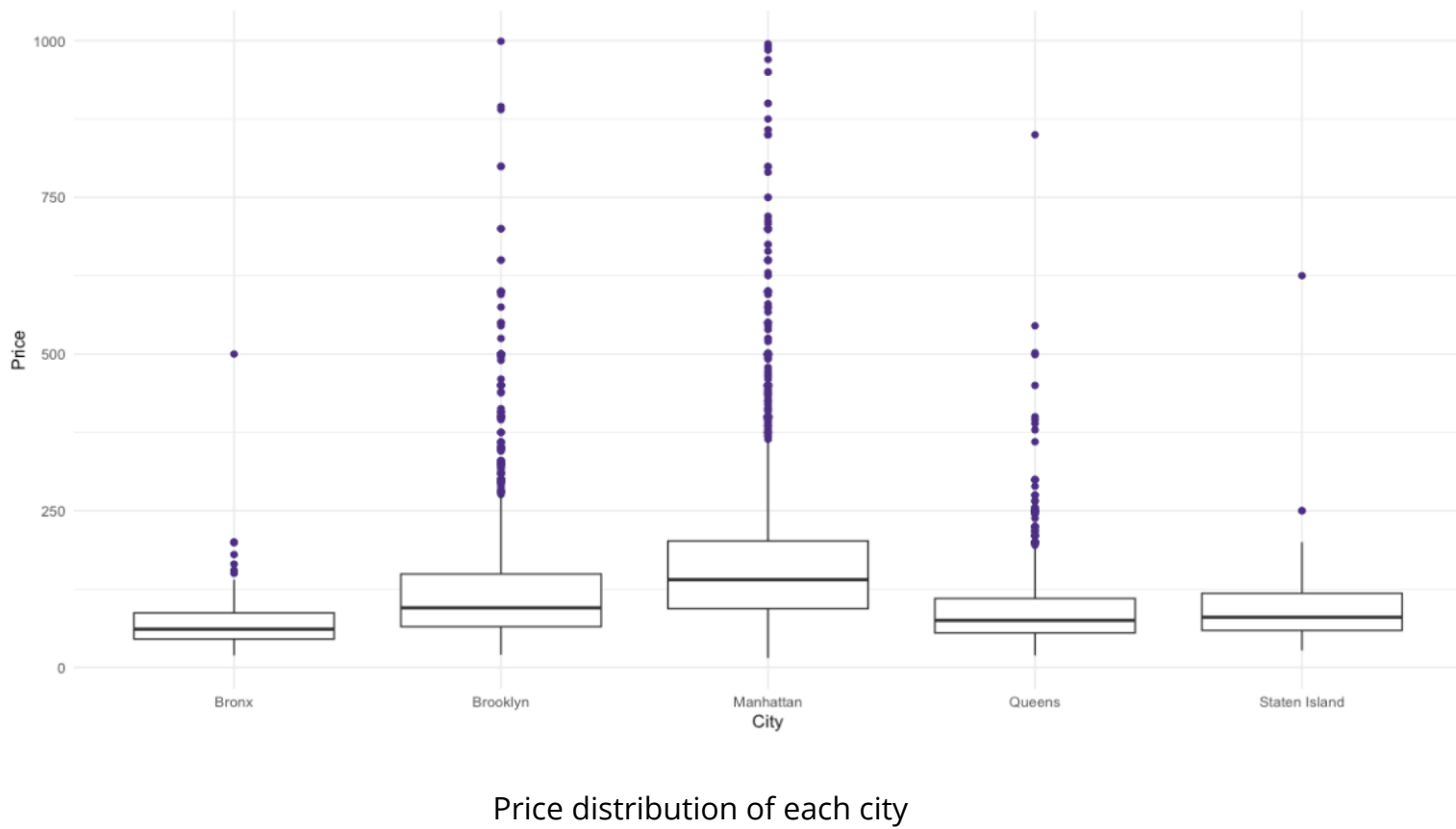
- The primary hypothesis of this study posits that geographical location, specifically at the neighborhood level, has a significant impact on Airbnb listing prices in New York City, and listing prices vary among different locations(neighbourhoods/cities).
- A further hypothesis is that the application of multilevel mixed-effects models will reveal neighborhood-specific price variations that traditional linear regression may not capture due to its limitation in addressing nested data structures.

Listings Categorized by Price



## Methods

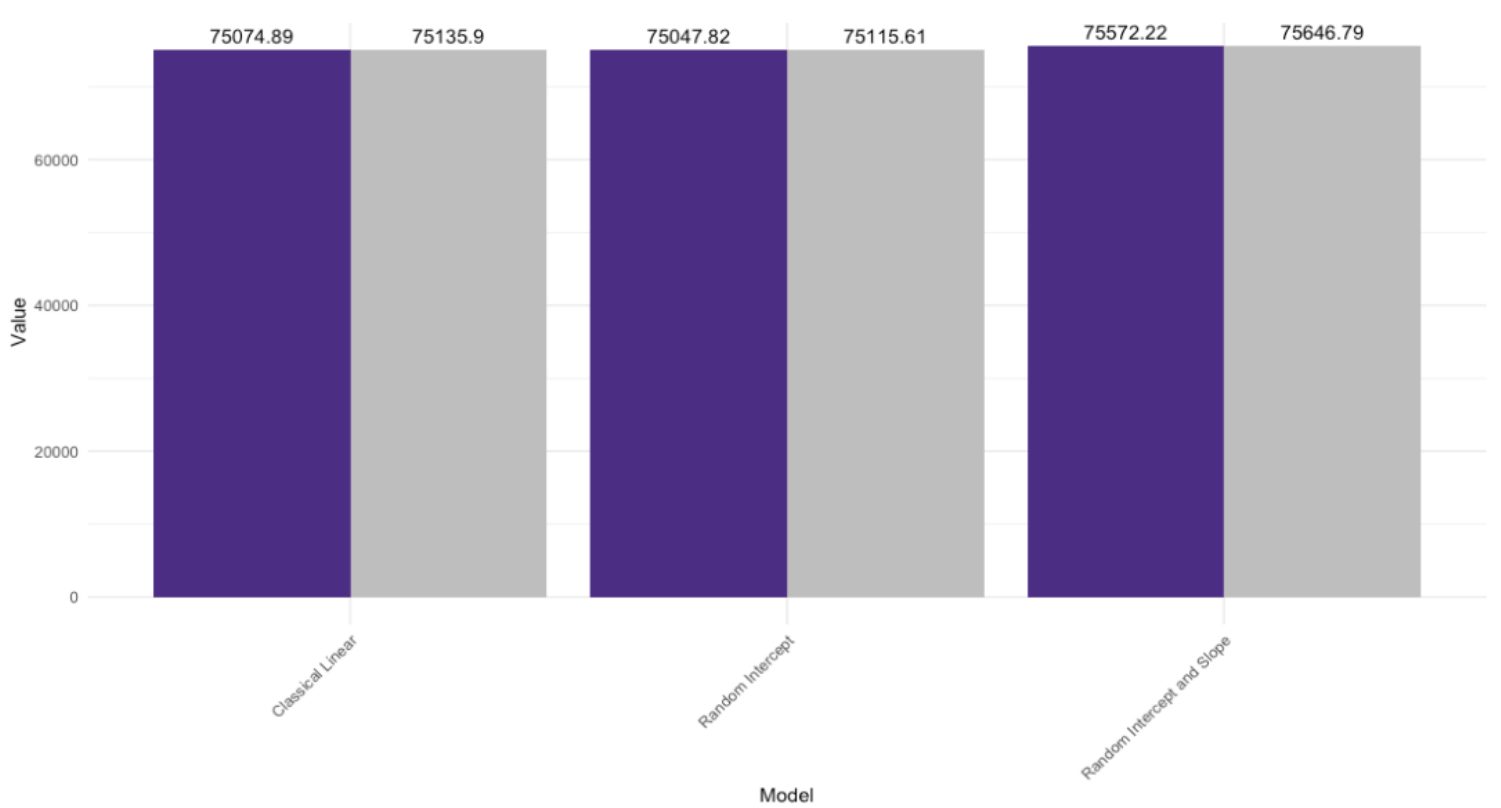
- **Exploratory Data Analysis (EDA):** Analyzed Airbnb data in New York to understand key pricing factors.
- **Linear Regression Model:** Used linear regression to predict Airbnb prices based on property characteristics.
- **Multilevel Mixed-Effects Model - City Level:** Explored pricing variations across different cities with a multilevel model.
- **Multilevel Mixed-Effects Model - Neighborhood Level:** Examined neighborhood-specific price influences using a detailed multilevel model.



## Results Part 1

- Experiment 1: Sample observations using stratified sampling, ensuring that the samples maintain representation across different groups or strata.
- The analysis grouped the data by "City" as the primary variable of interest, allowing for the examination of how pricing varies within each city.
  - Three models were considered to explore pricing variations:  
Random Intercept Model: AIC = 75895.33, BIC = 75956.35  
Classical Linear Model: AIC = 75922.09, BIC = 75976.32  
Random Intercept and Slope Model: AIC = 75572.22, BIC = 75646.79

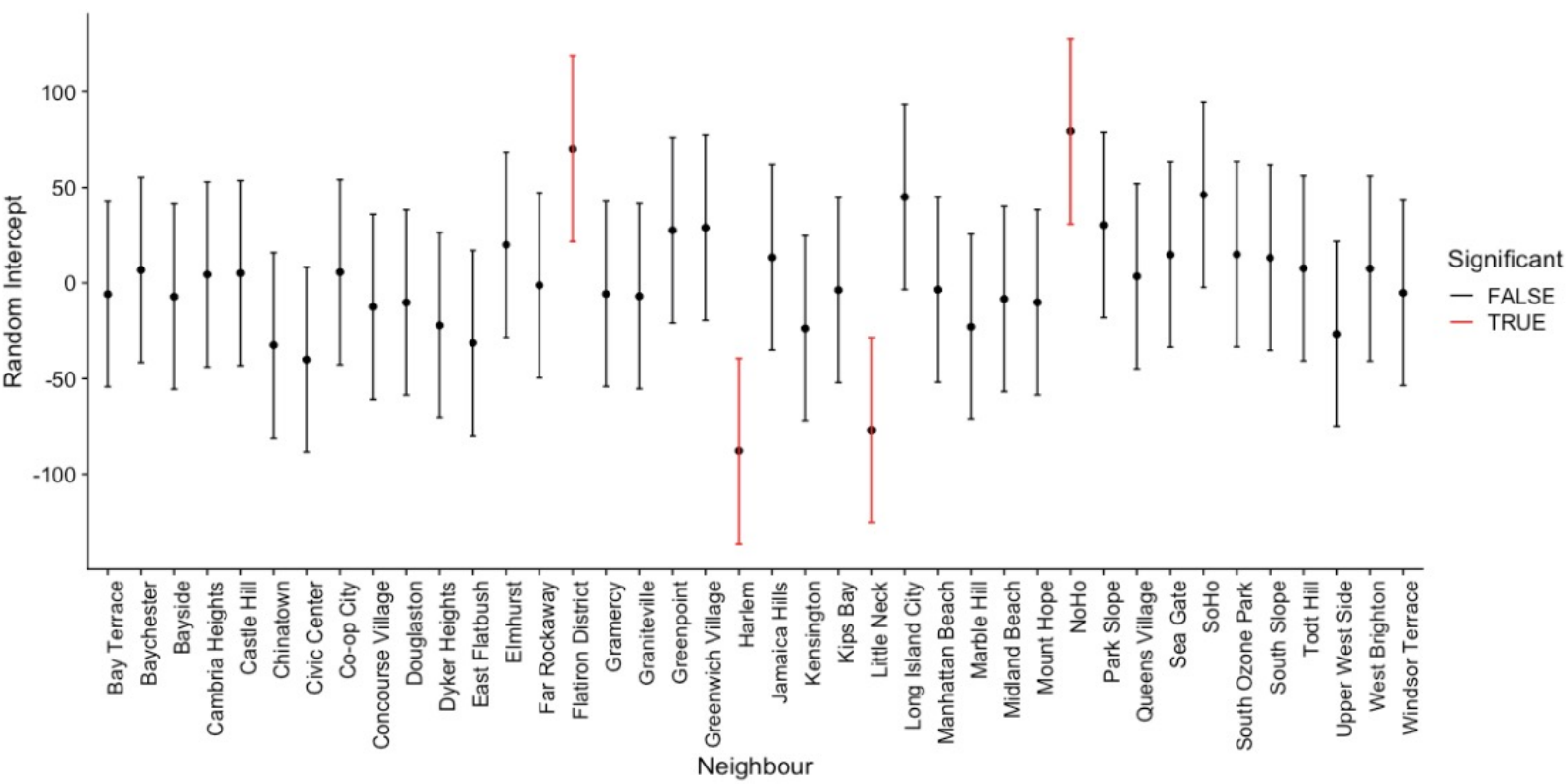
Model Comparison by AIC and BIC



## Results Part 2

- Experiment 2: A total of 40 neighborhoods were randomly selected from the entire dataset, representing a diverse range of locations within New York City.
- The first model is a Classical Linear Model, which provides a traditional approach to understanding pricing dynamics. The second model is a Random Intercept and Slope Model, which incorporates both city-level intercepts and slopes, allowing for more nuanced insights into neighborhood-specific variations.
  - The AIC and BIC values for the three models are as follows:  
The Random Intercept Model has an AIC of 75895.33 and a BIC of 75956.35. The Classical Linear Model has an AIC of 75922.09 and a BIC of 75976.32. Finally, the Random Intercept and Slope Model has an AIC of 75572.22 and a BIC of 75646.79.

95% CI of Random Intercepts per Neighbourhood



## Conclusions

- Prices among different neighbourhoods/cities in this dataset don't significantly differ.
- Slight Improvement in Complex Models  
Complex mixed-effects models, like the Random Intercept Model, performed slightly better in modeling Airbnb pricing in New York City.
- Limited Optimization Gains  
However, the improvement achieved by these complex models was not significantly greater than the simpler Classical Linear Model.

The results could be attributed to several factors. These include limited inter-group variability in pricing across neighborhoods or cities, data quality and feature selection, sample size constraints, and the risk of overfitting associated with complex models.

ICC Comparison of Two Experiments

