**Introduction**

**Problem Statement and Motivation**

Airbnb has revolutionized the travel and hospitality industry by offering unique and localized accommodations beyond traditional hotels. However, escalating prices for Airbnb listings, particularly in high-demand cities like New York, have introduced challenges for travelers, especially younger generations eager to explore while remaining budget-conscious. Rising accommodation costs lead to hesitation among users, who now carefully evaluate options before booking. This highlights the need for a systematic understanding of pricing determinants.

This problem is particularly interesting because addressing it directly contributes to simplifying the booking process for price-sensitive users. Moreover, it aligns with the broader goal of promoting accessible and equitable travel opportunities. Use cases extend to enabling better decision-making for travelers, strategic pricing for hosts, and informed policy recommendations for Airbnb platforms in competitive urban markets.

**Proposed Approach**

To tackle this problem, we aim to:

1. Identify key features impacting price using statistical and machine learning models.
2. Segment the market into distinct categories based on price ranges and listing characteristics, such as budget, mid-range, and luxury accommodations.

We employed a combination of predictive modeling and clustering algorithms. Random Forest and XGBoost are also used for feature importance analysis, given their capabilities to handle mixed data types and identify complex patterns. For market segmentation, we used K-Nearest Neighbors (KNN) and other distance-based algorithms will help identify similar listings intuitively, emphasizing geographic and characteristic-based similarities.

**Strengths of the Approach**

Our approach builds upon existing methodologies while incorporating flexibility and scalability. Random Forest and XGBoost excel in identifying non-linear relationships and interactions between variables, providing robust performance for pricing predictions. These models are complemented by Logistic Regression for a simple and interpretable baseline.

Compared with previous studies, our approach uniquely integrates clustering for practical segmentation and user-friendly market insights. This two-pronged strategy—combining predictive modeling with segmentation—differs from solely predictive approaches by offering actionable recommendations for both travelers and hosts.

**Key Components**

The project involves several key steps:

1. Data Cleaning: Addressing missing values and preparing data for analysis.
2. Feature Importance Analysis: Using Random Forest and XGBoost to identify determinants such as location, room type, and availability.
3. Clustering for Market Segmentation: Grouping listings into price categories to facilitate user decision-making.
4. Visualization: Employing feature importance plots and clustering visualizations to ensure interpretability.