**Introduction**

This project focuses on exploring the **AB\_NYC\_2019** dataset, which contains detailed information about Airbnb listings in New York City for the year 2019. The dataset includes various features such as host information, listing prices, geographical coordinates, neighborhood characteristics, and reviews. The primary aim of this project is to uncover insights regarding trends, pricing, host performance, and neighborhood popularity.

**Dataset Overview**

* **Source**: The dataset was obtained from Inside Airbnb.
* **Key Columns**:
  + **id**: Unique identifier for each listing.
  + **host\_id**: Unique identifier for each host.
  + **neighbourhood\_group**: Borough classification of listings (e.g., Manhattan, Brooklyn).
  + **price**: Price per night of the listing.
  + **latitude** & **longitude**: Geographical coordinates for mapping.
  + **number\_of\_reviews**: Total reviews received per listing.
  + **reviews\_per\_month**: Average number of reviews received per month.

**Key Findings and Insights**

* The top host in the dataset has **327 listings**, indicating a high level of engagement with the Airbnb platform.
* An analysis of boroughs revealed that certain neighborhoods exhibit significantly higher listing densities, indicating varied demand across the city.
* A geographical heatmap was created to visualize the pricing trends across different neighborhoods, highlighting areas with both high and low costs.
* The title analysis showed common naming trends among listings, helping identify popular keywords and themes.

**Methods Used for Analysis**

1. **Data Cleaning**:
   * Removed duplicates and handled missing values to ensure data integrity.
   * Parsed columns for easier analysis, particularly in the title and review sections.
2. **Exploratory Data Analysis (EDA)**:
   * Used descriptive statistics to summarize key features.
   * Created visualizations using libraries such as Matplotlib and Seaborn to explore relationships and distribution of data.
3. **Geographical Analysis**:
   * Mapped listings using latitude and longitude coordinates, generating heatmaps to represent pricing variations visually.
4. **Text Analysis**:
   * Parsed and analyzed the listing titles to determine the most frequently used words, providing insight into naming conventions and popular trends.
5. **Reviews Analysis**:
   * Investigated the relationship between the number of reviews and listing prices to identify top-performing hosts.

**Summary of Visualizations**

* **Geographical Heatmap**: Displays pricing variations across different neighborhoods in NYC.
* **Bar Charts**: Illustrate the number of listings across boroughs and the relationship between reviews and pricing.
* **Word Clouds**: Represent common words used in listing titles to highlight popular themes.