

For this task, use **New York City Airbnb Open Data** from Kaggle.

It is available at the link after registration.

[New York City Airbnb Open Data](#)

## Visualization Task: Analyzing Airbnb Listings in New York City

### Objective:

Create a visual analysis of Airbnb listings in New York City using the dataset. This task will involve multiple visualizations that explore various aspects of the data, including neighborhood distribution, pricing trends, availability, room types, and review patterns.

### Task Breakdown:

#### 1. Neighborhood Distribution of Listings

- **Plot:** Create a bar plot to show the distribution of listings across different **neighbourhood\_group**.
- **Details:** Label each bar with the count of listings, use distinct colors for each neighborhood group, and add titles and axis labels.

#### 2. Price Distribution Across Neighborhoods

- **Plot:** Generate a box plot to display the distribution of **price** within each **neighbourhood\_group**.
- **Details:** Use different colors for the box plots, highlight outliers, and add appropriate titles and axis labels.

#### 3. Room Type vs. Availability

- **Plot:** Create a grouped bar plot to show the **average availability\_365** for each **room\_type** across the neighborhoods.
- **Details:** Include error bars to indicate the standard deviation, use different colors for room types, and add titles and axis labels.

#### 4. Correlation Between Price and Number of Reviews

- **Plot:** Develop a scatter plot with **price** on the x-axis and **number\_of\_reviews** on the y-axis.
- **Details:** Differentiate points by **room\_type** using color or marker style, add a regression line to identify trends, and include a legend, titles, and axis labels.

#### 5. Time Series Analysis of Reviews

- **Plot:** Create a line plot to show the trend of **number\_of\_reviews** over time (**last\_review**) for each **neighbourhood\_group**.

- **Details:** Use different colors for each **neighborhood group**, smooth the data with a rolling average, and add titles, axis labels, and a legend.

#### 6. Price and Availability Heatmap

- **Plot:** Generate a heatmap to visualize the relationship between **price** and **availability\_365** across different **neighborhoods**.
- **Details:** Use a color gradient to represent the intensity of the relationship, label the axes, and include a color bar for reference.

#### 7. Room Type and Review Count Analysis

- **Plot:** Create a stacked bar plot to display the **number\_of\_reviews** for each **room\_type** across the **neighbourhood\_group**.
- **Details:** Stack the bars by **room type**, use different colors for each room type, and add titles, axis labels, and a legend.

#### Execution and Submission:

- **Python Script:** Develop a single Python script using Matplotlib to generate all the visualizations. The script should be modular, with functions dedicated to each plot.
- **Data Preparation:** Ensure the dataset is cleaned and pre-processed as necessary (e.g., handling missing values, categorizing data) before visualization.
- **Output:** Save each plot as a separate image file (e.g., PNG) and also display them sequentially within the script.

#### Deliverables:

- A Python script (.py) that generates all the specified visualizations.
- Individual image files for each plot (e.g., neighborhood\_distribution.png, price\_distribution.png).