## **NYC Airbnb Market Analysis**

# **📌 Project Summary**

This project explores **New York City’s Airbnb market** using **data importing, cleaning, and exploratory data analysis (EDA)** techniques. The primary goal is to extract valuable insights related to **pricing trends, host behavior, customer preferences, and business opportunities** for potential Airbnb hosts, travelers, and business investors.

By analyzing the dataset, we can uncover patterns in **listing prices, availability, demand, host activity, customer reviews, and location preferences**. These insights can help **Airbnb hosts optimize their pricing, investors identify profitable neighborhoods, and customers make informed decisions** when booking accommodations.

**Project Title: Exploring the NYC Airbnb Market**

**Overview:** This project analyzes the NYC Airbnb market using a dataset containing information about listings, pricing, host details, customer reviews, and availability. The objective is to derive meaningful insights into market trends, revenue potential, customer preferences, and optimal pricing strategies.

**📚 Technologies & Libraries Used:**

* **Python:** Primary language for data analysis.
* **Google Colab:** Environment for running the analysis.
* **Pandas:** Data manipulation and preprocessing.
* **NumPy:** Numerical computations.
* **Matplotlib & Seaborn:** Data visualization.
* **Folium:** Geographic data visualization (heatmaps).
* **WordCloud:** Text analysis of customer reviews.

**🔍 Key Insights & Analyses:**

1. **Revenue Trends Over Time:**
   * Monthly average price analysis using bar charts.
   * Seasonal pricing trends to identify high and low-demand periods.
2. **Neighborhood Price Analysis:**
   * Identified the most expensive and cheapest neighborhood groups based on average price.
   * Highlighted areas with high customer interest but lower supply.
3. **Demand for Listings:**
   * Heatmap visualization of listing availability across NYC.
   * Analysis of top hosts with the highest number of listings.
4. **Instant Booking & Pricing:**
   * Percentage of instant bookable listings.
   * Average price comparison across different room types.
5. **Customer Sentiment Analysis:**
   * WordCloud visualization of frequently mentioned words in house rules.
6. **Impact of Ratings on Bookings:**
   * Scatter plot showing correlation between number of reviews and price.
7. **Optimal Pricing for Maximum Bookings:**
   * Identified the price ranges that maximize total booking revenue.
8. **Superhosts vs. Regular Hosts:**
   * Comparison of revenue generated by verified hosts vs. non-verified hosts.
9. **Cleaning Fee vs. Guest Satisfaction:**
   * Analyzed correlation between service fees and customer ratings.
10. **Best Time of Year for Promotions:**
    * Seasonal pricing analysis to determine when hosts should adjust prices.
11. **Room Type Performance:**
    * Comparison of price and availability for different room types.
12. **Expansion Opportunities:**
    * Identified neighborhoods with high reviews but fewer listings, indicating potential growth areas.

## **🛠 Errors And Solutions**

🔹 Dtype Warning: Mixed Data Types in Columns

* Cause: Columns like price, service fee contain numbers stored as strings due to symbols ($, ,).

Fix: Convert them to numeric after removing non-numeric characters.  
  
 python  
CopyEdit  
df["price"] = pd.to\_numeric(df["price"].replace({'\$': '', ',': ''}, regex=True), errors='coerce')

🔹 NaT (Not a Time) Error

* Cause: last review column has missing values or incorrect formats.

Fix: Convert using errors='coerce' to avoid breaking the code.  
  
 python  
CopyEdit  
df["last review"] = pd.to\_datetime(df["last review"], errors="coerce")

🔹 KeyError: Column Not Found

* Cause: Typo in column names (e.g., host id instead of host\_id).

Fix: Print column names and correct spelling.  
  
 python  
CopyEdit  
print(df.columns)

🔹 AttributeError: 'float' object has no attribute 'str'

* Cause: Applying .str methods on a column with NaN or non-string values.

Fix: Convert column to string before using .str.  
  
 python  
CopyEdit  
df["neighbourhood"] = df["neighbourhood"].astype(str)

🔹 ZeroDivisionError

* Cause: Performing division when denominator is 0.

Fix: Replace 0 with NaN or add a check.  
  
 python  
CopyEdit  
df["reviews per month"].replace(0, np.nan, inplace=True)

🔹 MemoryError (Out of Memory Issue)

* Cause: Large dataset causing excessive RAM usage.

Fix: Use chunking while reading large files.  
  
 python  
CopyEdit  
chunk\_size = 10000 # Read 10,000 rows at a time

df = pd.read\_csv("your\_file.csv", chunksize=chunk\_size).

🔹 Matplotlib: No Numeric Data to Plot

* Cause: Trying to plot non-numeric data.

Fix: Convert to numeric before plotting.  
  
 python  
CopyEdit  
df["price"] = pd.to\_numeric(df["price"], errors="coerce")

### **📋 Uses of the Project:**

1. **Market Insights for Hosts & Airbnb:**
   * Helps Airbnb hosts understand pricing trends, demand, and optimal pricing strategies.
   * Identifies areas where Airbnb can expand its presence based on demand and competition.
2. **Revenue Optimization:**
   * Analyzes seasonal trends to suggest the best times to adjust pricing and offer promotions.
   * Identifies factors that impact revenue, such as host verification, superhost status, and service fees.
3. **Customer Behavior & Preferences:**
   * Analyzes guest booking patterns, instant bookability preferences, and satisfaction levels.
   * Identifies top-performing listing types and amenities that attract customers.
4. **Operational Efficiency & Risk Management:**
   * Identifies hosts with multiple listings, aiding in fraud detection and operational planning.
   * Analyzes availability trends to assess market saturation and gaps.
5. **Improving Guest Experience:**
   * Uses word cloud analysis of house rules to understand common restrictions or preferences.
   * Analyzes the correlation between service fees and guest satisfaction.

### **📝 Analyses Performed:**

1. **Revenue Trends Analysis:**
   * Monthly revenue trends using average price per listing over time.
2. **Geospatial Demand Analysis:**
   * Heatmap visualization to show high-demand areas based on availability and bookings.
   * Identification of neighborhoods with high reviews but fewer listings (potential for Airbnb expansion).
3. **Host Performance Analysis:**
   * Top hosts by the number of listings.
   * Comparison of revenue between superhosts and regular hosts.
4. **Customer Sentiment Analysis:**
   * Word cloud analysis of house rules to understand guest concerns and common restrictions.
5. **Impact of Ratings on Bookings:**
   * Scatter plot analysis of price vs. number of reviews.
   * Correlation between service fees and guest satisfaction.
6. **Optimal Pricing Strategy:**
   * Price segmentation and revenue impact analysis by price range.
   * Seasonal pricing trends based on guest reviews.
7. **Room Type Performance:**
   * Average price and availability of different room types to identify the most profitable options.

**Conclusion:** This project provides valuable insights into Airbnb market dynamics in NYC. It helps hosts optimize pricing, understand customer preferences, and identify expansion opportunities. The use of visualization techniques makes the data easily interpretable for stakeholders looking to make data-driven decisions.