**Data Cleaning & EDA Project — Actionable Summary**

**✅ Objective:**

Clean and explore the NYC Airbnb dataset to ensure data quality and extract meaningful insights for decision-making or modeling.

**1. Data Cleaning Process**

**🔹 Data Integrity**

* Converted last\_review to proper datetime format.
* Verified valid ranges for price, minimum\_nights, and location coordinates.

**🔹 Missing Data Handling**

* Dropped rows with missing name and host\_name (non-essential text fields).
* Imputed missing reviews\_per\_month with 0 (no reviews implies 0).
* Preserved last\_review nulls since they represent unreviewed listings.

**🔹 Duplicate Removal**

* Identified and removed **100% of duplicate rows**, ensuring dataset uniqueness.

**🔹 Standardization**

* Standardized categorical values (room\_type, neighbourhood\_group, etc.) using title case formatting.

**🔹 Outlier Detection**

* Removed listings with:
  + price = 0 or price > $1000
  + minimum\_nights > 365  
    These entries were likely errors or not practical.

**2. Exploratory Data Analysis (EDA)**

**🔹 Price Insights**

* Most listings are priced under $200, with a steep drop-off after that.
* High-end outliers were removed to stabilize the distribution.

**🔹 Room Type Trends**

* **Entire home/apt** dominates NYC listings, followed by private rooms.
* Shared rooms are rare.

**🔹 Geographic Insights**

* Manhattan and Brooklyn host the majority of listings.
* Geospatial plot shows dense clusters in downtown Manhattan.

**🔹 Review Activity**

* Many listings have never been reviewed (~20%).
* Listings with reviews often receive between 0–2 reviews/month.

**🔹 Availability Patterns**

* Some listings are available 365 days a year, suggesting full-time rentals.
* Others show short-term availability (seasonal).