Data Requirements Document (DRD)

Project Title:	
Airbnb NYC: Insights & Engagement	

1. Purpose

The purpose of this DRD is to outline the data requirements for the Airbnb NYC case study analysis project. The objective is to explore insights related to pricing, availability, room types, host activity, and review patterns using cleaned Airbnb dataset and deliver meaningful business recommendations via dashboards and Python visualizations.

2. Source of Data

- Original Dataset: NYC Airbnb Open Data (sourced from Kaggle)
- Cleaned Dataset: cleaned_airbnb.csv
- Cleaning Tool: Python (Pandas, NumPy)
- Visualization Tools:
 - o Power BI (2 dashboards)
 - Jupyter Notebook (6 Python plots)

3. Data Fields (Key Columns Used)

Column Name	Description	Туре
id	Unique identifier for each listing	Integer
name	Title/name of the listing	Text
host_id	Unique identifier for the host	Integer
host_identity_verified	Indicates whether the host's identity is verified	Boolean/Text
host_name	Name of the host	Text
neighbourhood_group	Main city region (e.g., Manhattan, Brooklyn)	Categorical
neighbourhood	Specific area within the neighborhood group	Categorical
lat	Latitude coordinate of the listing	Float
long	Longitude coordinate of the listing	Float
instant_bookable	Indicates if the listing can be booked instantly	Boolean/Text
cancellation_policy	Type of cancellation policy (e.g., flexible, strict)	Categorical
room_type	Type of room offered (Entire home/apt, Private room, Shared room)	Categorical
Construction year	Year the property was constructed	Integer
price	Price per night in USD	Integer
service fee	Additional service fee charged per stay	Integer
minimum nights	Minimum number of nights required for booking	Integer
number of reviews	Total number of reviews received	Integer
last review	Date of the most recent review	Date
reviews per month	Average number of reviews per month	Float
review rate number	Average rating score given by guests	Float
calculated host listings count	Total number of listings by the same host	Integer
availability 365	Number of days the listing is available per year	Integer
total_cost	Total cost = price + service fee	Integer

4. ii Derived/Calculated Measures (Power BI)

Measure Name	Formula/Description
Revenue Potential	price * minimum_nights (summed)
Occupancy Rate	(availability_365 - minimum_nights) / availability_365
Review Category	Categorized as "High" or "Low" based on review score >= 4.5
Review Activity	Categorized as "Active" if reviews per month > 1
Active Hosts Count	Count of listings with frequent reviews

5. **Python Visualizations Used**

up

6. Usage in Dashboards

Dashboard Name Purpose

Room Type Insights Pricing, distribution, and revenue by room type/location

Airbnb Room Trends NYC Host activity, review volume, and occupancy rates

7. Data Requirements Checklist

- All null or NA values removed or handled in Python
- Data types properly assigned
- No duplicate listings (cleaned)
- Derived metrics calculated and validated
- Relationship checks between price, review, availability