Description

Since 2008, guests and hosts have used Airbnb to expand on traveling possibilities and present more unique, personalized way of experiencing the world.

Data

This dataset describes the listing activity and metrics in NYC, NY for 2019. This data file includes all needed information to find out more about hosts, geographical availability, necessary metrics to make predictions and draw conclusions.

The dataset has these columns.

Column	Description
id	listing ID
name	name of the listing
host_id	host ID
host_name	name of the host
neighbourhood_group	location
neighbourhood	area
latitude	latitude coordinates
longitude	longitude coordinates
room_type	listing space type
price	price in dollars
minimum_nights	amount of nights minimum
number_of_reviews	number of reviews
last_review	latest review
reviews_per_month	number of reviews per month
calculated_host_listings_count	amount of listing per host
availability_365	number of days when listing is available for booking

PART 1- Coding (Please email your code in an hour, even if you have not finish all questions)

Q1: How many neighborhoods have more than 1000 listings on Airbnb? How many neighborhoods have more than 1000 hosts on Airbnb?
Q2: Find the average price of each neighborhood listing with both listing id and host_id greater than 1500?
Q3: Find the host name who has the greatest number of listings in New York? For that host, how many listings of “private room” type does this host have?
Q4: Let’s define popularity of a listing as the total *number_of_reviews* for that listing. What are the three most popular listing (id and host_id) in each neighborhood group?

PART 2- Insight

Explore the dataset and show what insight you can find in 1h to 1.5h. Some suggested directions to take when exploring the data to find insight are listed below, but feel free to take other directions.

- What can we learn about different hosts and areas?
- What can we learn from various columns? (ex: locations, prices, reviews, etc.)
- Which hosts are the busiest and why?
- Is there any noticeable difference of traffic among different areas and what could be the reason for it?

A visualization of this dataset would be a great way to explore more such questions.