Name: Anupama Mohan Rao

SUID: 809474520

**IST 652: Scripting for Data Analysis**

**Mini Project-1 Report**

**Data**: New York City Airbnb Open Data (Sourced from Kaggle.com). Dataset describes the listing activity and metrics in New York City for 2019.

**Data Exploration & Data Cleaning***:*

In the data cleaning process, only the columns that are necessary for to answer the business questions have been retained. The other columns that do not pertain to the business questions that I’m trying to answer are dropped.

**Removing NULL values**: It is difficult to analyze data with NULL values in it. The next steps I have taken is to replace NULL values with zero and verifying those columns again to check if there are any more NULL values that are existing. To check for null values again, I summed the column values, if there contains even a single NULL value, the result will be NULL.

**Business Question:**

1. Who are the top 10 hosts with most listings?

Unit of analysis: host\_id

Comparison: Host ID with No. of listings

Output: Finding the hosts with most number of listings and using head() to subset top 10 values of hosts with most listings.

1. Which are top 10 properties that are most reviewed?

Unit of analysis: Number of reviews

Comparison: All the other factors in the table.

Using nlargest(), finding all the properties which have the most reviews, and filtering it out to top 10 values.

1. Reviews by neighborhood listing and room type.

Unit of Analysis: Five boroughs- Bronx, Brooklyn, Manhattan, Queen, Staten Island.

Comparison: room-type- Entire home/apt, Private room, Shared room

Output: Summary table of the five boroughs against room-type with number of reviews.

**Program Description:**

The program is broken down into following parts

* Reading the file and understanding the data.
* Data wrangling & Data cleaning
* Exploratory Data analysis

**Output Description**:

Top 10 hosts with the most listings:

Table

Description automatically generated

Top 10 properties that are most reviewed:

Table

Description automatically generated

Reviews by neighborhood listing and room type:

Table

Description automatically generated