

Context:

Since 2008, guests and hosts have used Airbnb to expand on traveling possibilities and present more unique, personalized way of experiencing the world. This project deals with “**Airbnb NYC 2019**” Dataset.

As a data modeling and visualization specialist, you are required to work on the dataset to perform EDA and fulfill the business requirement.

The data file “Airbnb NYC 2019” includes all the necessary information needed to find out more about the hosts, geographical availability, necessary metrics to make predictions and draw conclusions.

Activities To Perform:

Perform the following activities on the dataset using R Language. You may use any R centric libraries as may be needed to complete the operations.

1) Perform Pre-processing on the dataset as required.

- a. Treat missing values
- b. Remove duplicates if any
- c. Identify and treat outliers if any
- d. Transform Data based on judgement and explain the methodology used

2) Perform Exploratory Data Analysis on the ‘Airbnb NYC 2019’ Dataset

- a. What can we learn about different hosts and areas?
- b. Which hosts are the busiest and why?
- c. Is there any noticeable difference in traffic among different areas and what could be the reason for it?

3) Plot Graphs and charts to demonstrate the following:

- a. Cluster of hosts according to neighborhood group.
- b. Plot a graph showing room_type chosen commonly with neighborhood group.
- c. Plot average price with neighborhood.

d. Plot highest, lowest, and median price of stay with neighborhood.