Case Study 2: Airbnb New York City Open Data

Background

Since its establishment in 2008, Airbnb has served as a platform for "sharing economy", where hosts rent out their homes to short term guests, as alternatives to hotels and more personalized traveling options. Data were collected on Airbnb listings in New York City, NY from the year of 2019. A total of 48,895 listings are included in the dataset, with 16 variables such as listing name, host name, listing neighborhood, exact location, etc.

Goal

The primary goal of this analysis is to identify any discernible and interesting patterns among the listings in NYC. Scientific questions include: (i) Which are the most influential factors on the popularity/price of a listing? (ii) Is there heterogeneity among boroughs and neighborhoods? And if there is, which neighborhood has the heaviest traffic or highest price? (iii) Does the type of the listing (i.e., private room, entire home, etc.) vary across neighborhoods? (iv) If you were to post a listing in NYC and would like to make it the most popular and expensive one, where should your listing be located and how would you name it?

Variable key

id = listing ID
name = name of the listing
host_id = host ID
host_name = name of the host
neighbourhood_group = name of the borough (Manhattan, Brooklyn, etc.)
neighbourhood = name of the area/neighborhood
latitude = latitude coordinates
longitude = longitude coordinates
room_type = listing space type
price = price in dollars
minimum_nights = amount of nights minimum (that a guest can stay there)
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