



airbnb EDA

Capstone Project

▲ 112.35
▲ 38.70
▼ 20.24
▲ 81.44
32.67

AI



Aditya Bhople



Rupali Dawkore



Rachna Gupta



Piyush Sujeeth



Subhi Yadav

What is AirBnB ?

AI

Airbnb (ABNB) is an online marketplace that connects people who want to rent out their homes with people who are looking for accommodations in specific locales.

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

Data analysis on millions of listings provided through Airbnb is a crucial factor for the company. These millions of listings generate a lot of data - data that can be analyzed and used for security, business decisions, understanding of customers' and providers' (hosts) behavior and performance on the platform, guiding marketing initiatives, implementation of innovative additional services and much more.



Objectives

1. What is the average preferred price on the Airbnb platform by the customers on a particular location?
2. Find top 10 host with most listing on the Airbnb platform?
3. Find total no. of nights spend per location by the customers?
4. What are the total number of nights spend per room types by the customers?
5. Top 10 highest listing neighborhood on Airbnb platform?
6. What can we learn about different hosts and areas?
7. What can we learn from predictions? (ex: locations, prices, reviews, etc)
8. Which hosts are the busiest and why?
9. Is there any noticeable difference of traffic among different areas and what could be the reason for it?

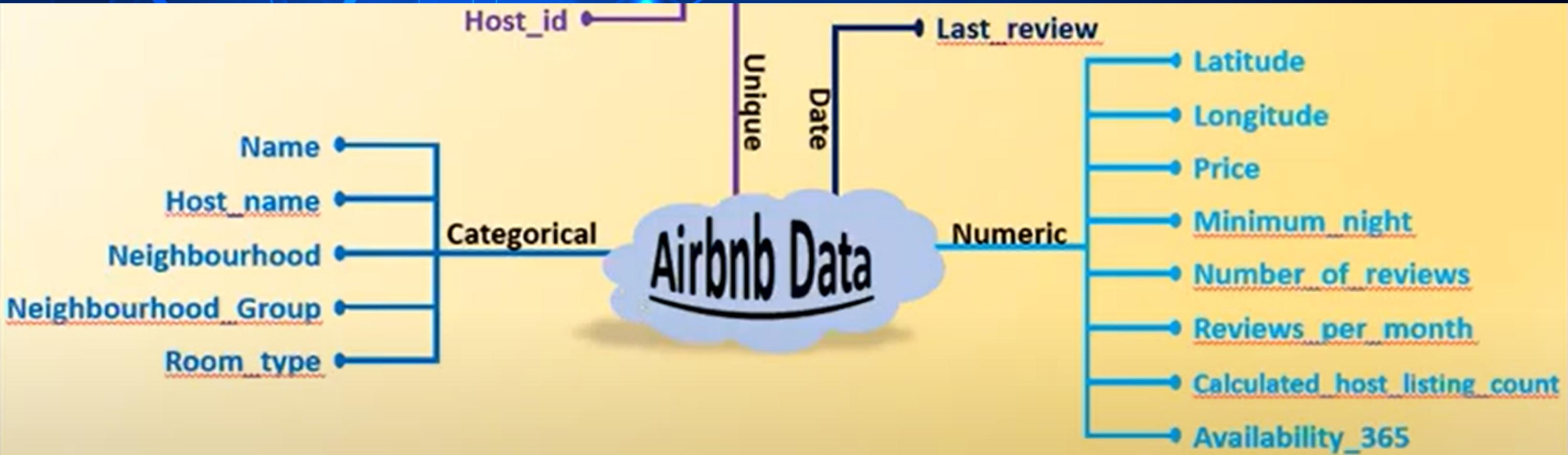
Attributes in the study

The Airbnb data set contains 48895 rows and 16 columns. Here are the columns:

- ID
- Name
- host_id
- host_name
- neighbourhood_group
- neighbourhood
- longitude
- latitude
- room_type
- room_type
- price
- minimum_nights
- number_of_reviews
- last_review
- reviews_per_month
- calculated_host_listings_count
- availability_365

AI

Data Summary



What is EDA ?

Exploratory Data Analysis refers to the critical process of performing initial investigations on data so as to discover patterns,to spot anomalies,to test hypothesis and to check assumptions with the help of summary statistics and graphical representations.

The following steps are involved in the process of EDA

- Acquire and loading the dataset
- Understanding the data and various variables
- Cleaning the dataset
- Exploring and visualizing the dataset
- Analyzing relationship between variables

Libraries used in Project

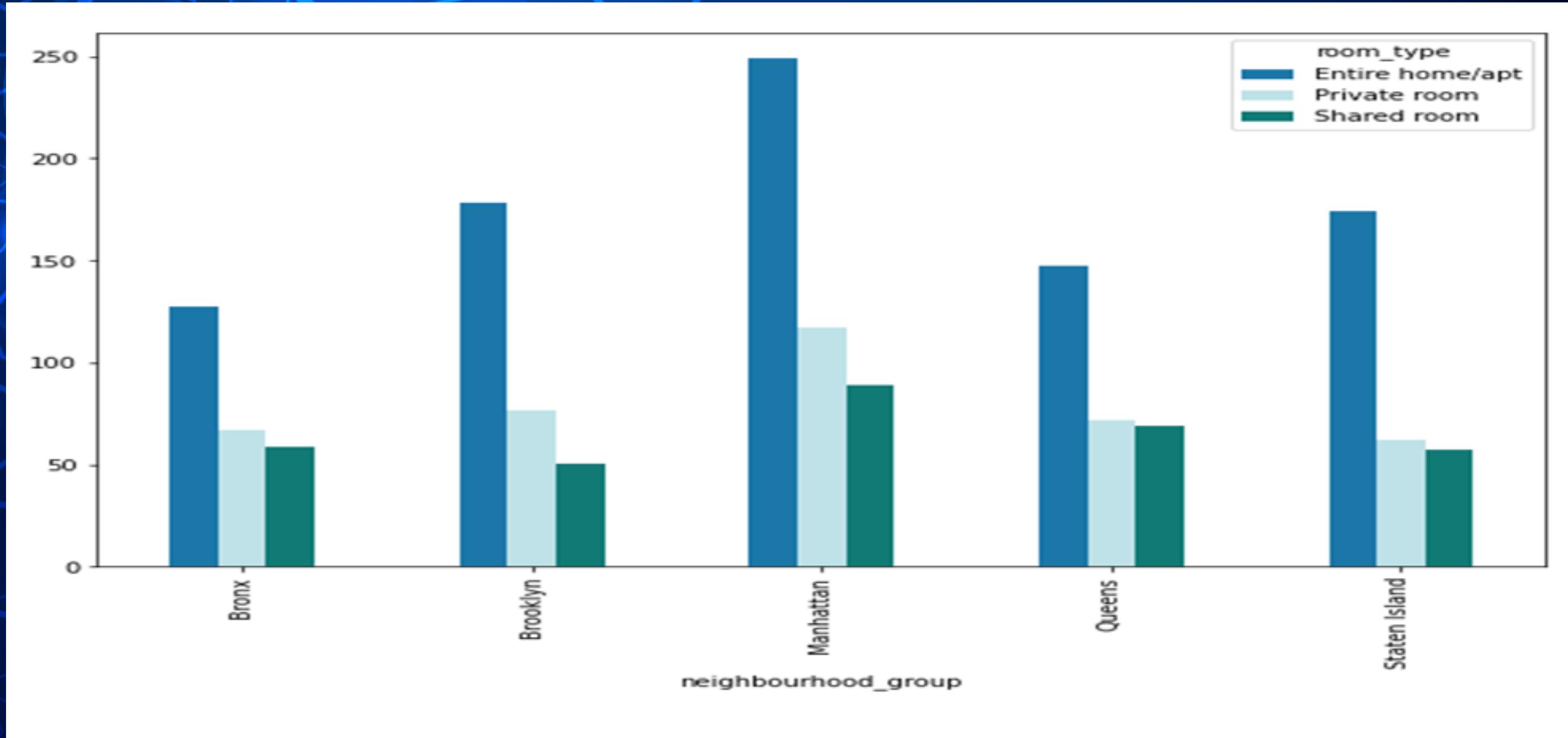
AI

- Matplotlib
- Seaborn
- Pandas
- Numpy



Neighbourhood Group

AI

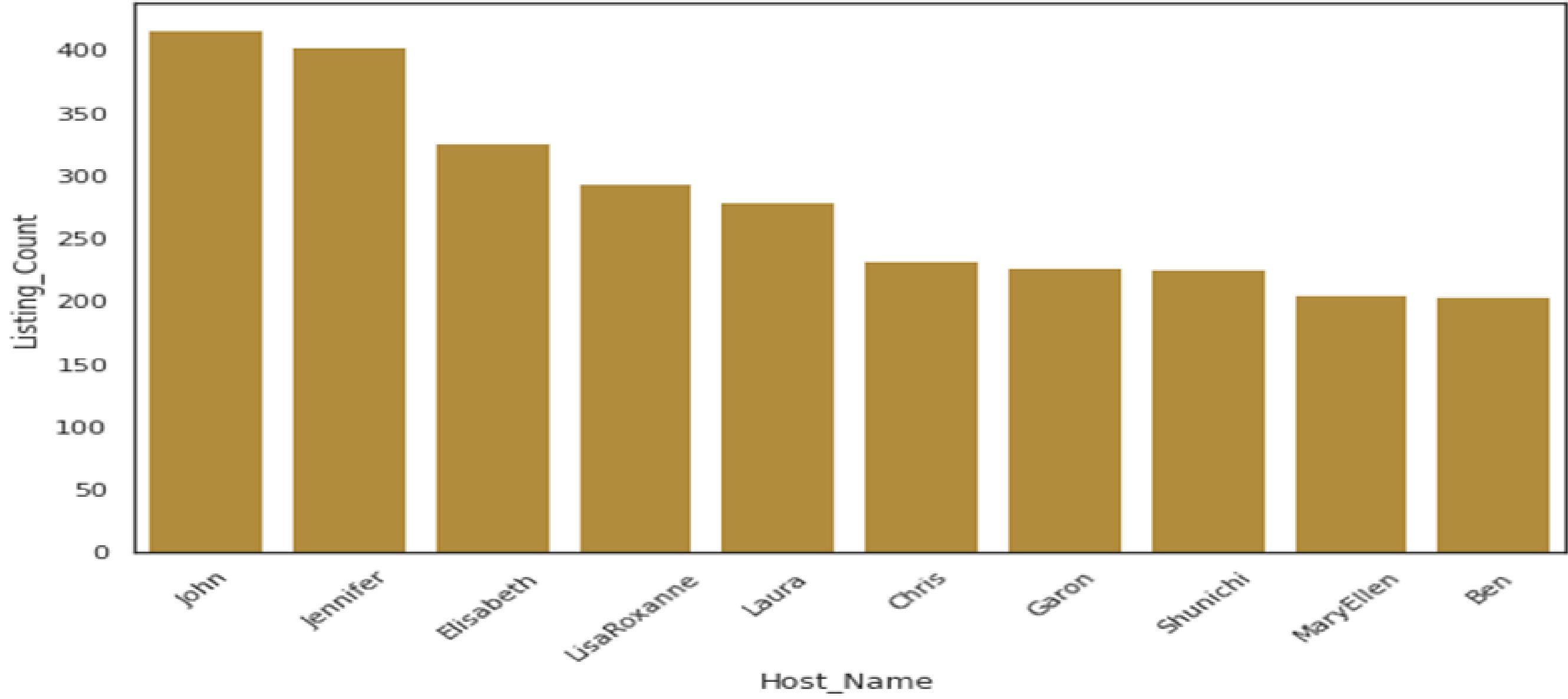


Number of properties available in different neighbourhood groups is more in Manhattan (Home room type)

Host with most listings on the Airbnb Platform

AI

Hosts with most listings on the Airbnb platform



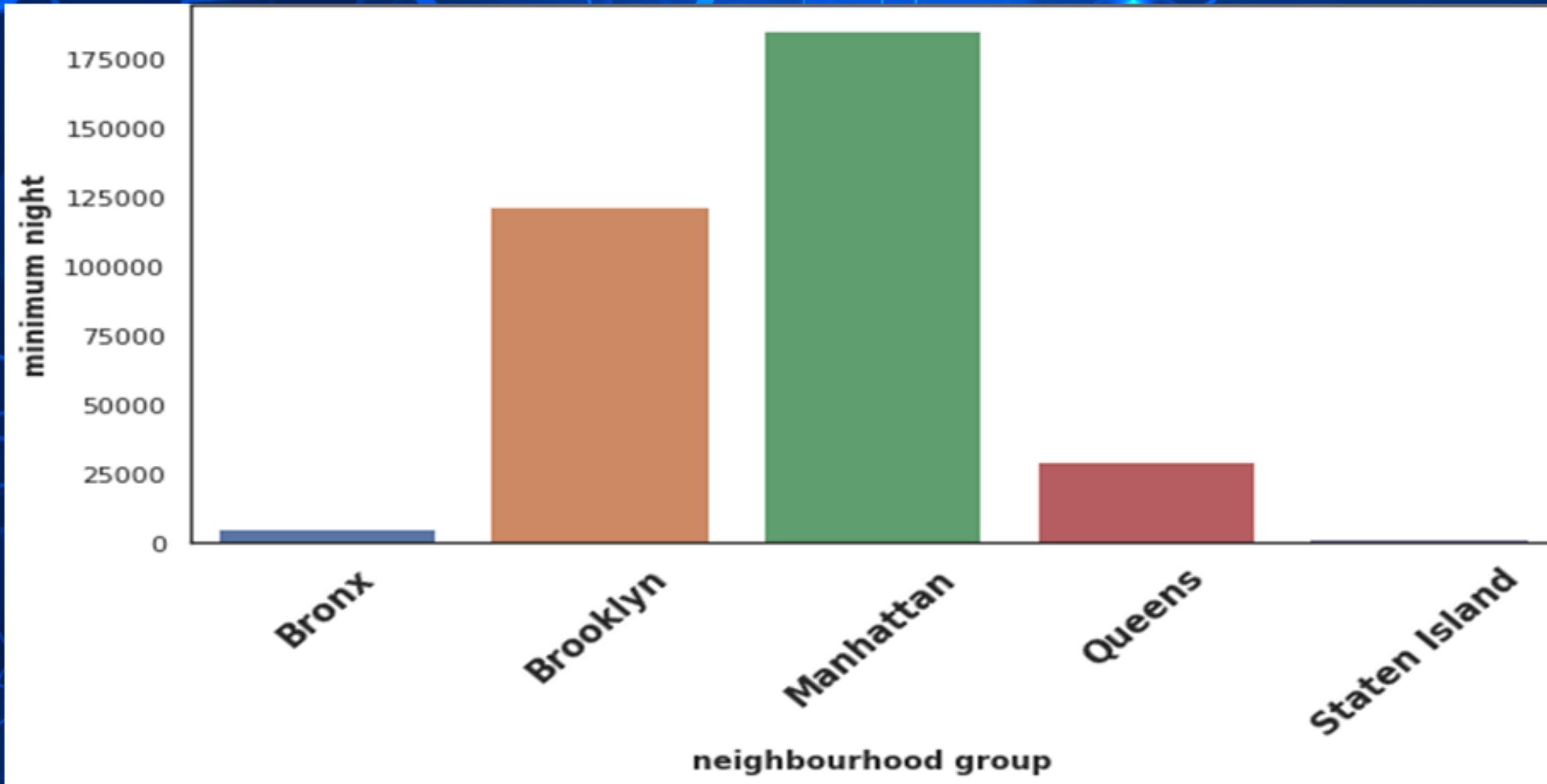
AI

	Host_Name	Counts
0	John	417
1	Jennifer	403
2	Elisabeth	327
3	LisaRoxanne	294
4	Laura	279
5	Chris	232
6	Garon	227
7	Shunichi	226

John has the most listing counts on the airbnb platform

Number of nights spend by the customer in neighbourhood groups

AI



neighbourhood_group	minimum_nights
Staten Island	1802
Bronx	4971
Queens	29356
Brooklyn	121693
Manhattan	184792

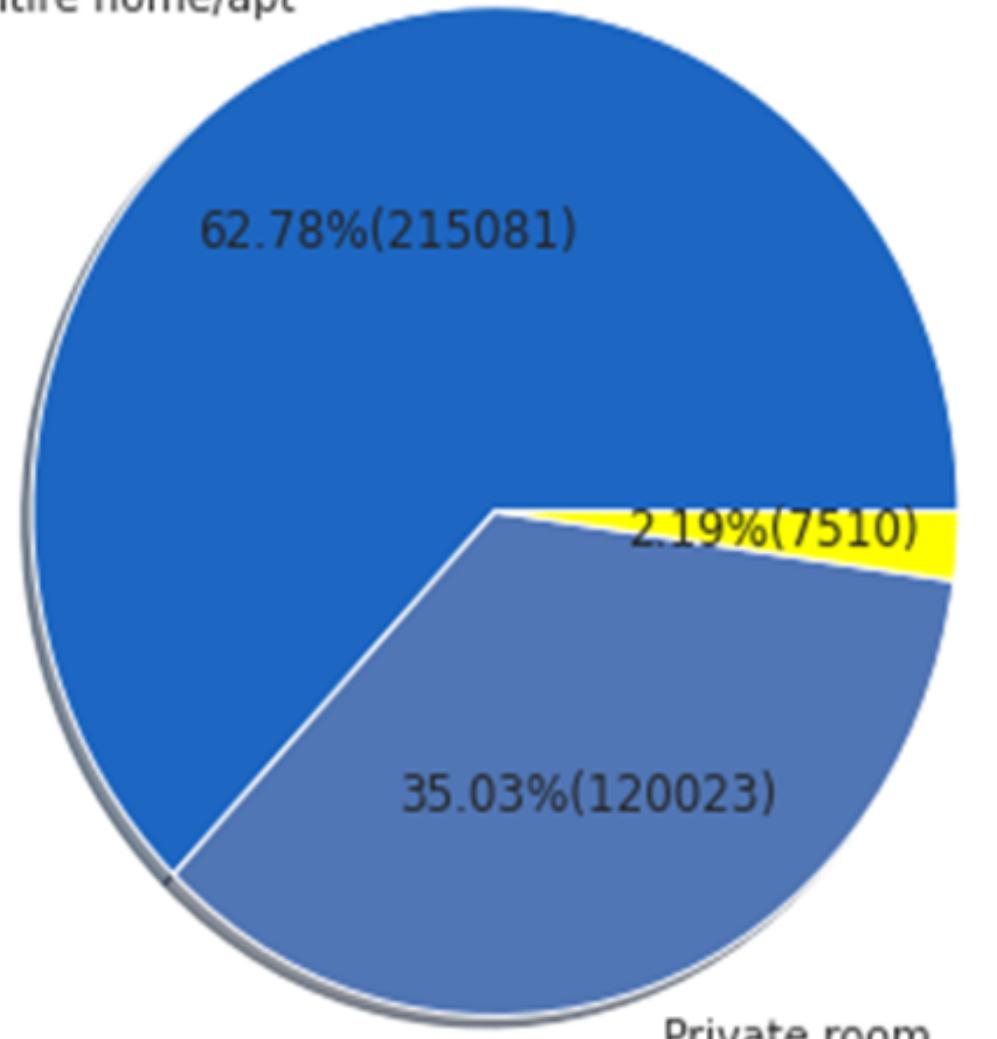
Here we can observe that people love to spend the maximum nights on Manhattan

Total number of nights spends according to room types

AI

Total_No.of nights spend per room types

Entire home/apt



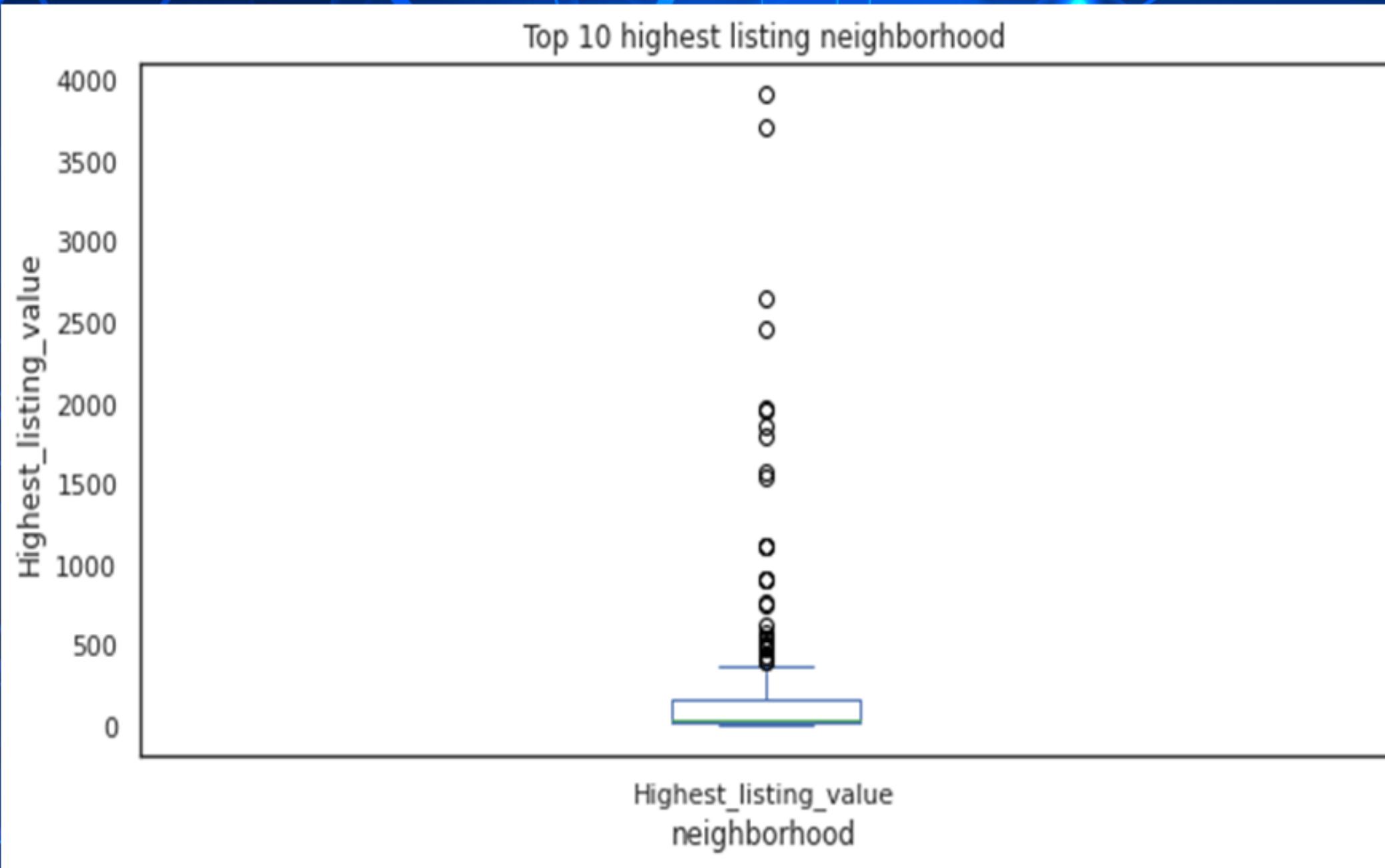
room_type minimum_nights

room_type	minimum_nights
Entire home/apt	215081
Private room	120023
Shared room	7510

Total number of nights spends according to room types

Top 10 Highest Listing on Airbnb Platform

AI

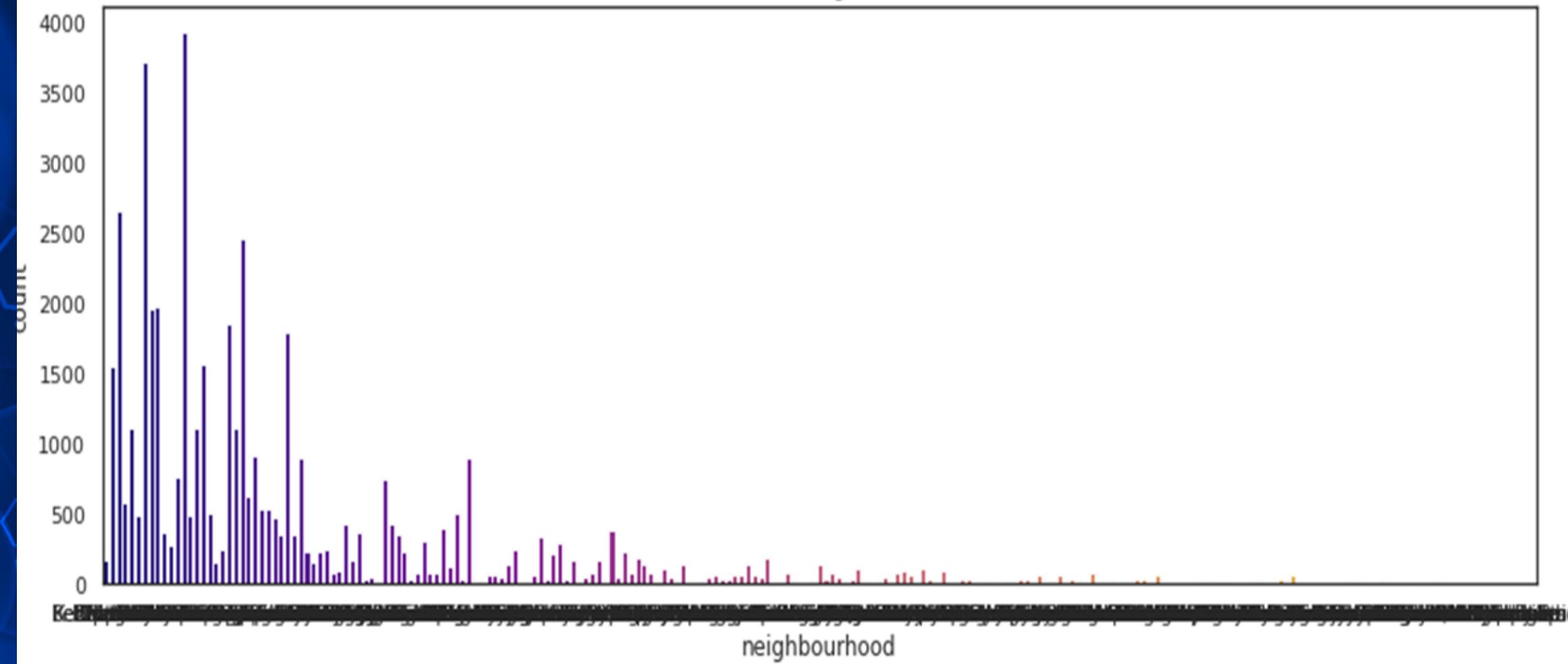


	neighbourhood	Highest_listing_value
0	Williamsburg	3917
1	Bedford-Stuyvesant	3713
2	Harlem	2655
3	Bushwick	2462
4	Upper West Side	1969
5	Hell's Kitchen	1954
6	East Village	1852
7	Upper East Side	1797
8	Crown Heights	1563
9	Midtown	1545

Number of nights spend by the customer in neighbourhood groups

AI

All Different Neighbourhoods



	neighbourhood	Highest_listing_value
0	Williamsburg	3917
1	Bedford-Stuyvesant	3713
2	Harlem	2655
3	Bushwick	2462
4	Upper West Side	1969
5	Hell's Kitchen	1954
6	East Village	1852
7	Upper East Side	1797
8	Crown Heights	1563
9	Midtown	1545

In this airbnb dataset we identified:-

Total_no_of_host_id 48858

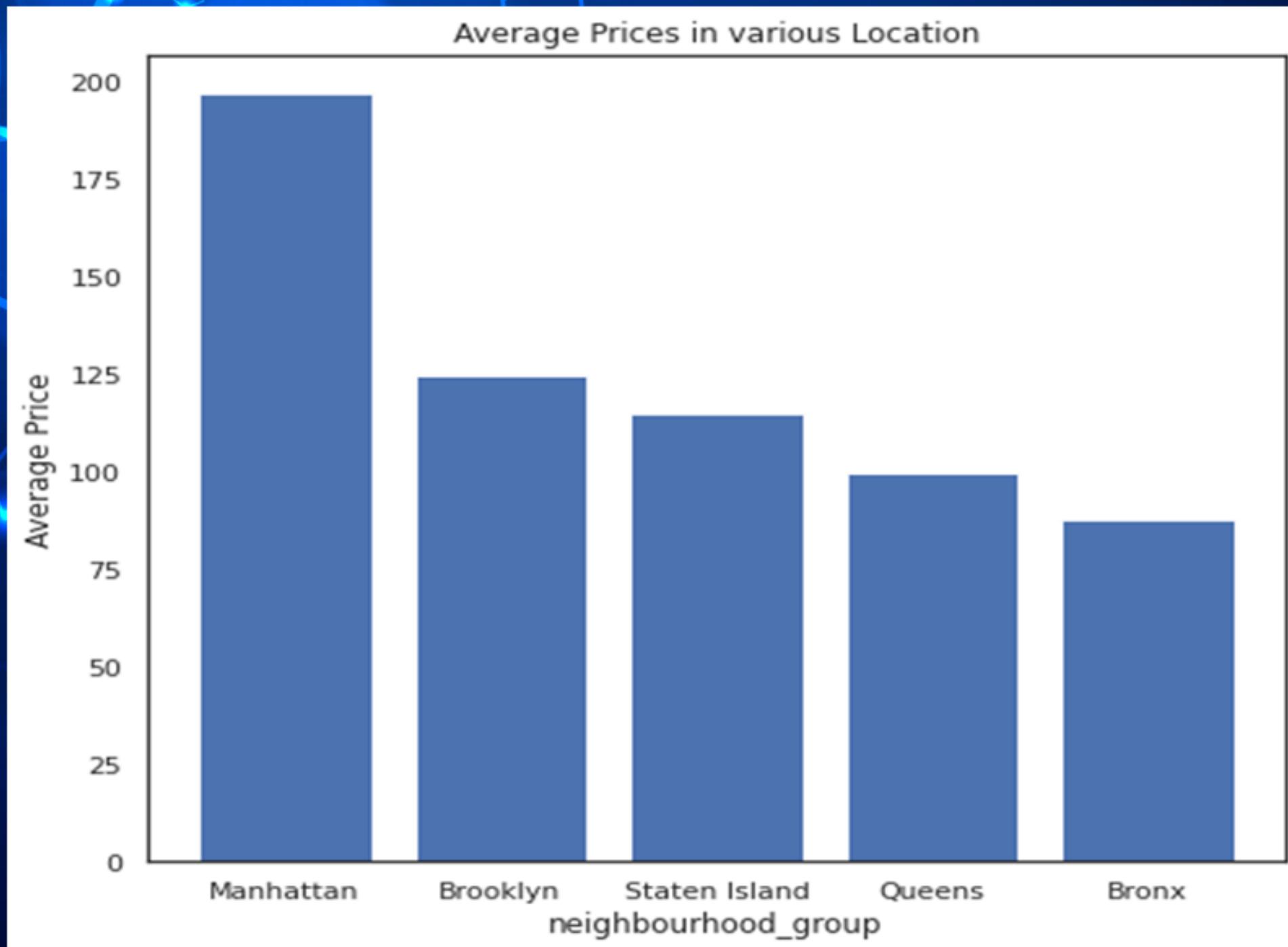
Total_no_of_host_name 48858

Total_unique_host_name 11450

Total_unique_host_id 37425

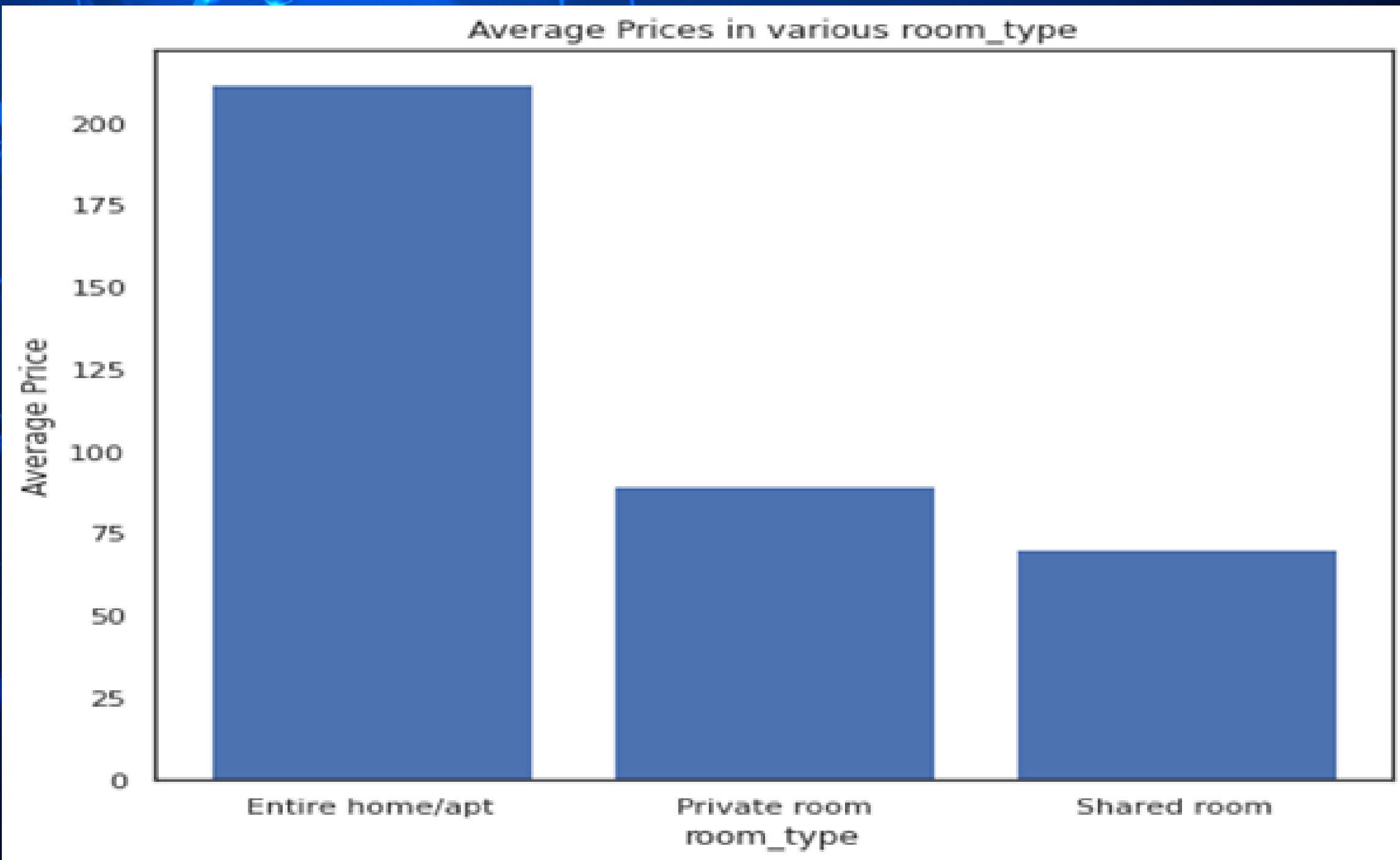
AI

What we can learn from predictions?



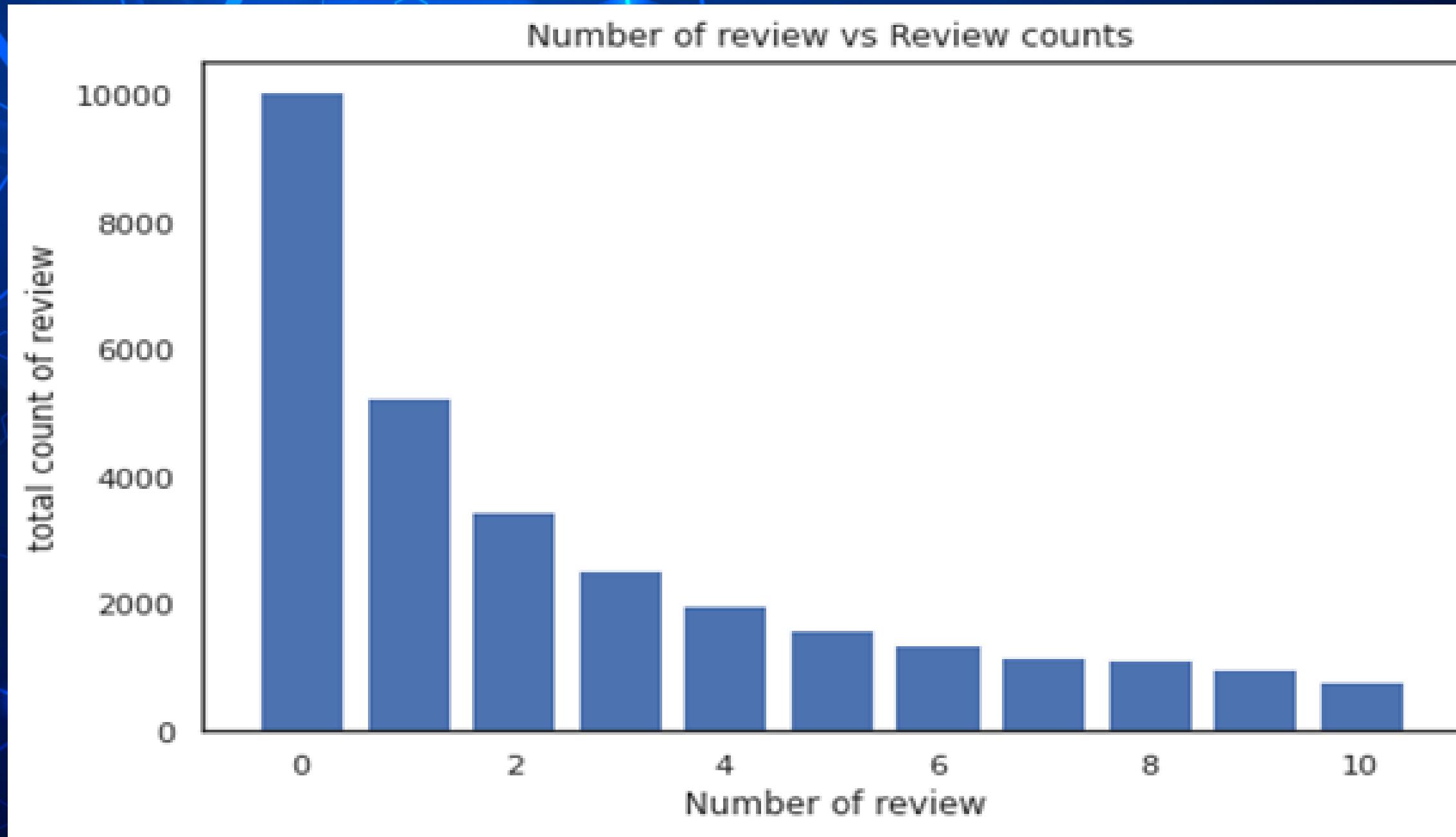
AI

What we can learn from predictions?



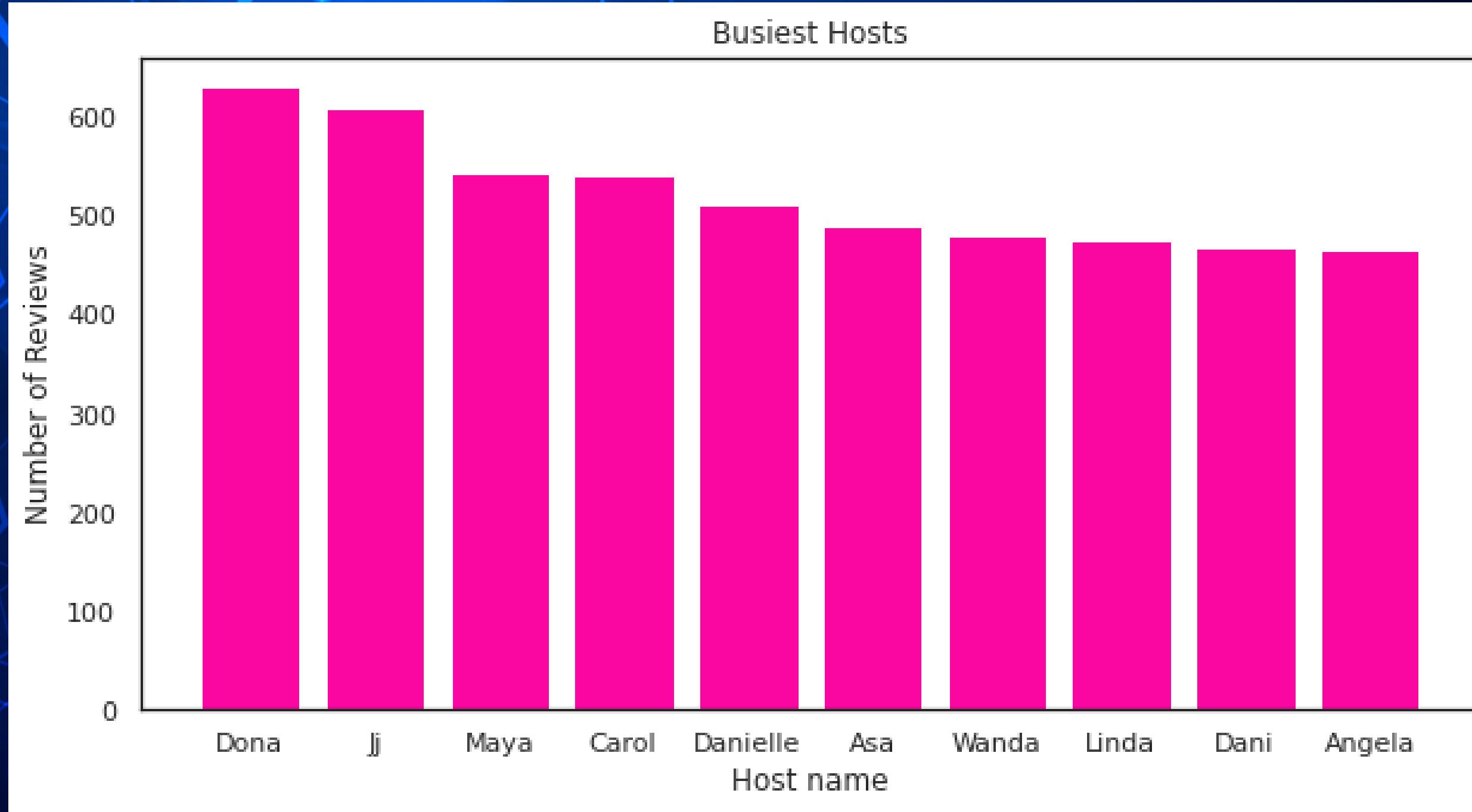
What we can learn from predictions?

AI



Which hosts are busiest and why?

AI



AI

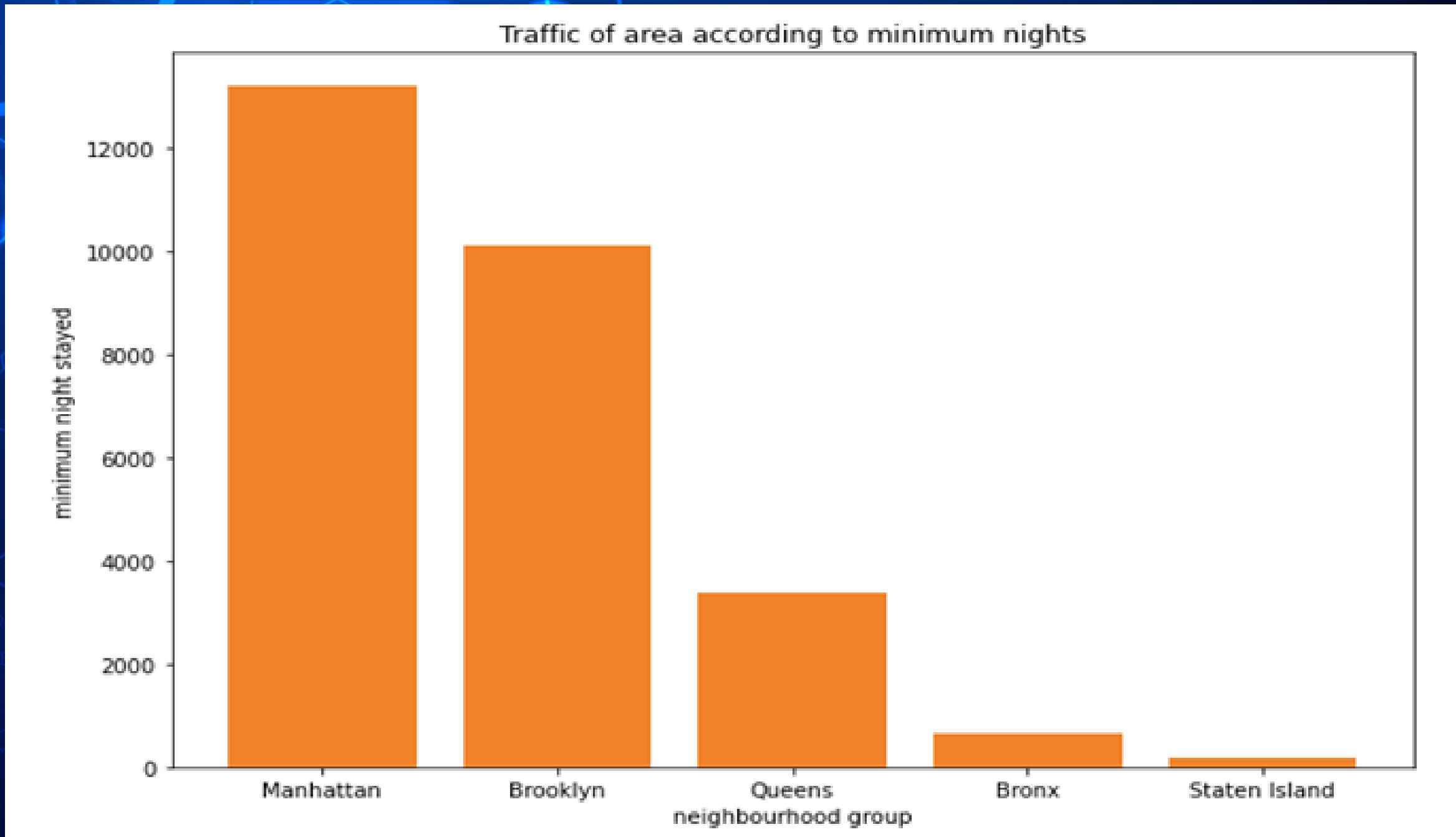
Which hosts are busiest and why?

	host_name	room_type	number_of_reviews
3434	Dona	Private room	629
6332	Jj	Private room	607
8977	Maya	Private room	543
2164	Carol	Private room	540
2975	Danielle	Private room	510
1214	Asa	Entire home/apt	488
13839	Wanda	Private room	480
7902	Linda	Private room	474
2947	Dani	Entire home/apt	467
863	Angela	Private room	466

According to their reviews count. These host are having more reviews because they have a listing of private rooms and entire apartment and people use stay more in private rooms and apartment.

AI

Noticable difference between different areas



Noticable difference between different areas

	neighbourhood_group	room_type	minimum_nights
6	Manhattan	Entire home/apt	13190
4	Brooklyn	Private room	10123
3	Brooklyn	Entire home/apt	9553
7	Manhattan	Private room	7973
10	Queens	Private room	3370
9	Queens	Entire home/apt	2096
1	Bronx	Private room	652
8	Manhattan	Shared room	480
5	Brooklyn	Shared room	413
0	Bronx	Entire home/apt	378
11	Queens	Shared room	198
13	Staten Island	Private room	188
12	Staten Island	Entire home/apt	176
2	Bronx	Shared room	59
14	Staten Island	Shared room	9

From the above analysis we can see that the minimum nights spent by the people are in Manhattan in entire home/apartment.

We can also say that

- Manhattan
- Brooklyn
- Queens

Have the huge traffic because people prefer more to stay in private room and entire home/apartment. We can clearly see that these neighbourhood are proving more private rooms and entire home/apartment.

Limitation

- ✓ Dataset features in terms of modern world, are of very poor quality in deciding the valuation of a property
- ✓ User ratings of hosts aren't available, it would've been better to rank our hosts based on user satisfaction and ratings. Normally a low rated property tends to lower their price.
- ✓ In order to have a better analysis regarding the quality of the properties, it would be interesting if we had an analysis of sentiments with property valuations.
- ✓ The exact number of guests count also missing

Scope of Improvement

- ✓ As dataset has few qualifying attributes to value a property, more features can be added like bedroom, bathroom, property age (it might be one of the most important one), applicable tax rate, distance to nearest airport, hospital or schools.
- ✓ In presence of ratings, hosts can be classified and ranked, special discount or offer can be given to highest rated hosts following marketing strategy
- ✓ Time series analysis can be done to make prediction on occupancy rate based on tourist season.

Conclusion

- ✓ Manhattan and Brooklyn are the two distinguished, expensive & posh areas of NY
- ✓ Most visitors don't prefer shared rooms, they tend to visit private room or entire home.
- ✓ Some properties are having Minimum Nights to stay is more than 365 Days which can be favorable among Students, Low-Income Employees & Immigrants.
- ✓ Though location of property has high relation on deciding its price, but a property in popular location doesn't mean it will stay occupied in most of the time.
- ✓ Top rated hosts as per the number of reviews are Dona and Jj.



AI

Thank You !!