

Capstone Project

Airbnb Bookings Analysis

by

Team X

Darpan Agrawal
Aadarsh Pandey



Introduction

Airbnb, Inc. is an American company that operates an online marketplace for lodging, primarily homestays for vacation rentals, and tourism activities. Based in San Francisco, California, the platform is accessible via website and mobile app

Dataset Information (48895, 16)

Null count:

name - 16

host_name - 21

last_review - 10052

reviews_per_month - 10052

RangeIndex: 48895 entries, 0 to 48894

Data columns (total 16 columns):

#	Column	Non-Null Count	Dtype
0	id	48895 non-null	int64
1	name	48879 non-null	object
2	host_id	48895 non-null	int64
3	host_name	48874 non-null	object
4	neighbourhood_group	48895 non-null	object
5	neighbourhood	48895 non-null	object
6	latitude	48895 non-null	float64
7	longitude	48895 non-null	float64
8	room_type	48895 non-null	object
9	price	48895 non-null	int64
10	minimum_nights	48895 non-null	int64
11	number_of_reviews	48895 non-null	int64
12	last_review	38843 non-null	object
13	reviews_per_month	38843 non-null	float64
14	calculated_host_listings_count	48895 non-null	int64
15	availability_365	48895 non-null	int64

dtypes: float64(3), int64(7), object(6)

Dataset Information (48895, 16)

Null count:

name - 16

host_name - 21

last_review - 10052

reviews_per_month - 10052

RangeIndex: 48895 entries, 0 to 48894

Data columns (total 16 columns):

#	Column	Non-Null Count	Dtype
0	id	48895 non-null	int64
1	name	48879 non-null	object
2	host_id	48895 non-null	int64
3	host_name	48874 non-null	object
4	neighbourhood_group	48895 non-null	object
5	neighbourhood	48895 non-null	object
6	latitude	48895 non-null	float64
7	longitude	48895 non-null	float64
8	room_type	48895 non-null	object
9	price	48895 non-null	int64
10	minimum_nights	48895 non-null	int64
11	number_of_reviews	48895 non-null	int64
12	last_review	38843 non-null	object
13	reviews_per_month	38843 non-null	float64
14	calculated_host_listings_count	48895 non-null	int64
15	availability_365	48895 non-null	int64

dtypes: float64(3), int64(7), object(6)

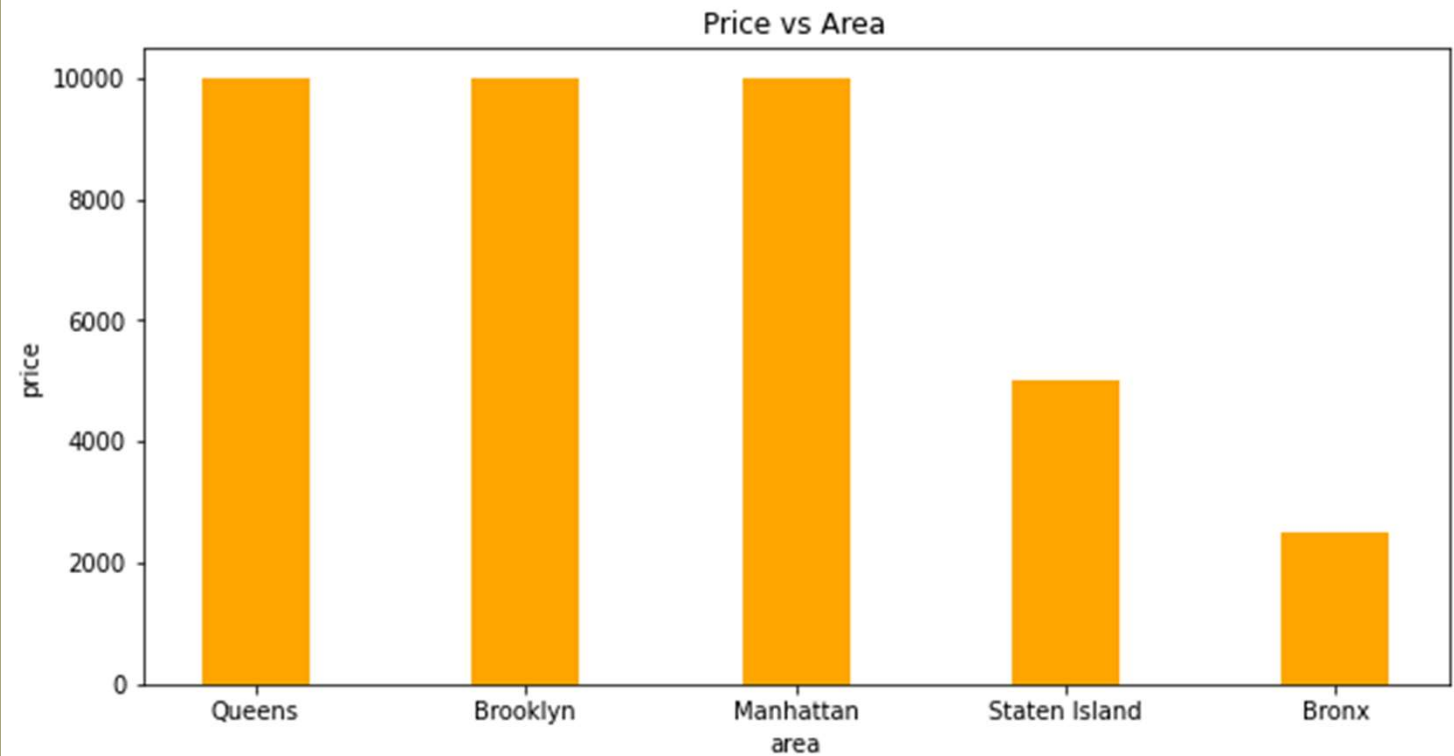
There are 211 unique neighbourhoods
falls under 5 groups

5 unique groups :

**Manhattan
Brooklyn
Queens
Staten Island
Bronx**

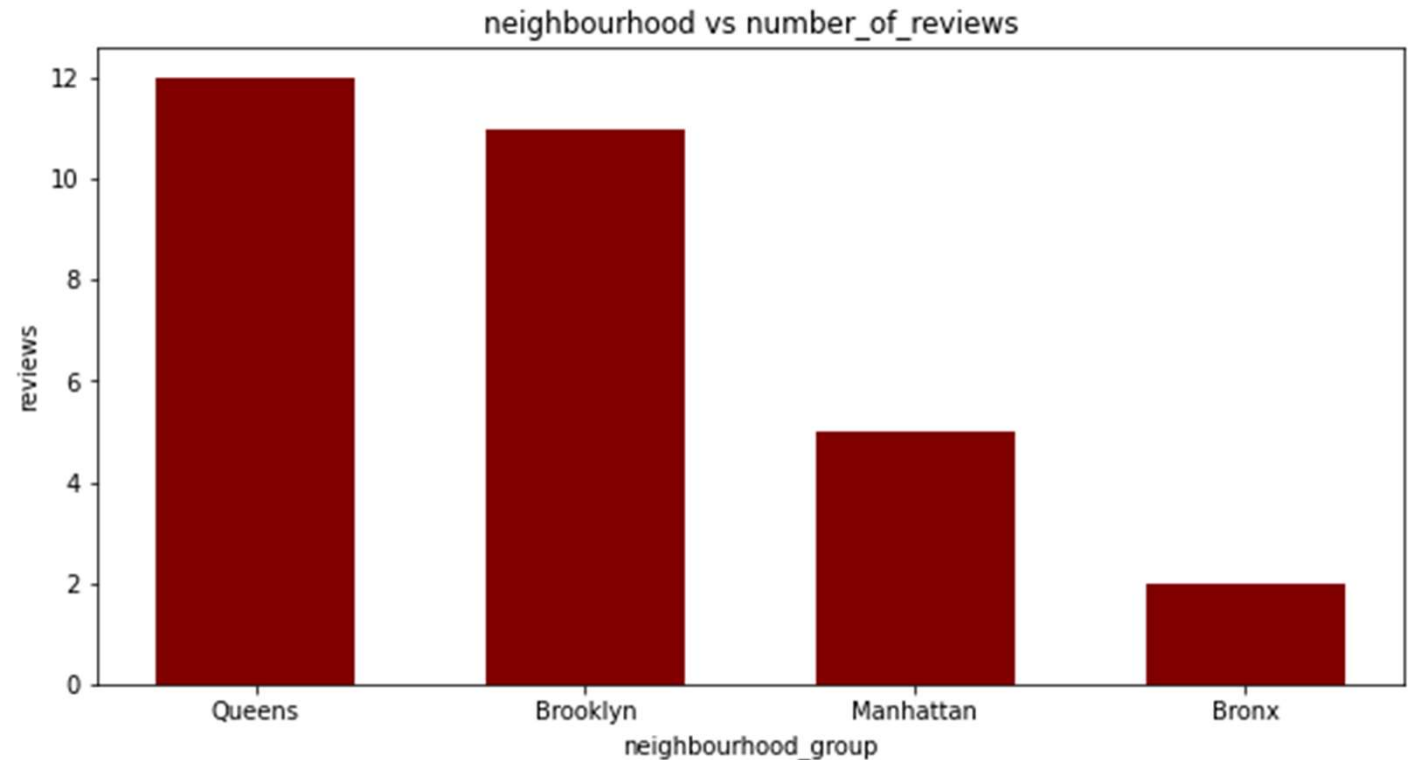
**Location of
Neighbourhood
Groups**

Hosts and Area



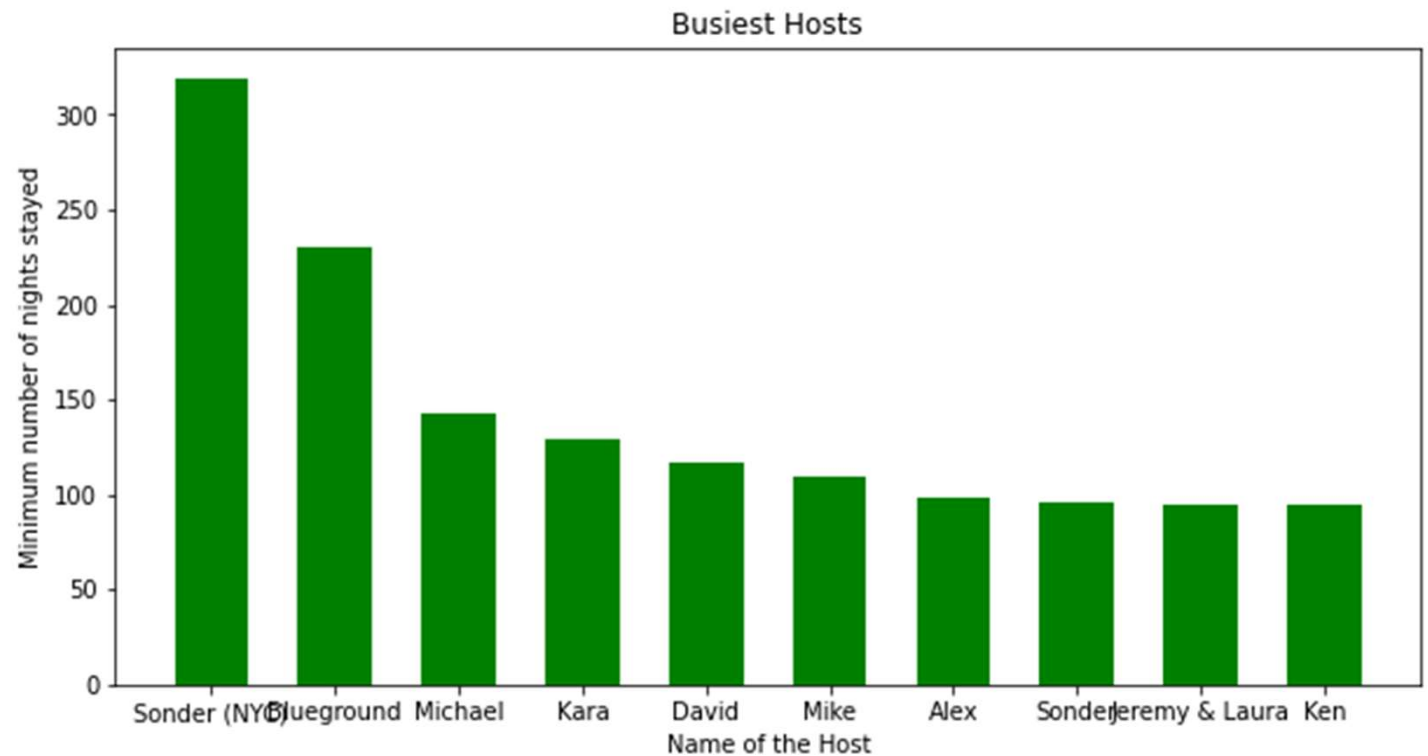
"Queens", "Brooklyn" and "Manhattan" had almost similar prices which is around 10000 and followed by them we had "Staten Island" which is second highest and "Bronx" as 3rd highest

Reviews vs Neighbourhood groups



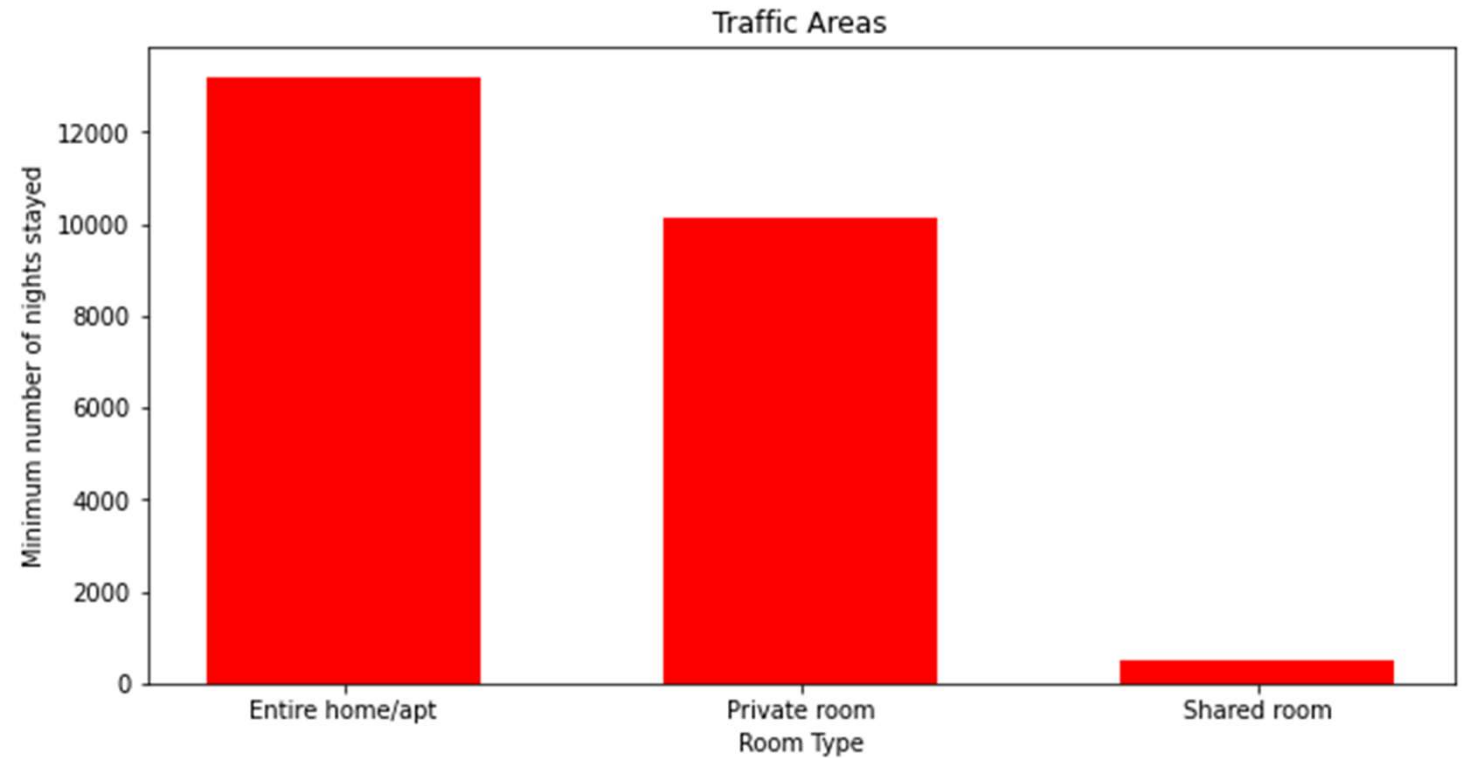
"Queens" has the highest number of reviews followed by that we have "Brooklyn" on the second and then we have "Manhattan" as 3rd highest reviews and then we have "Bronx" in the last that has lowest reviews compare to 1st 3 neighbourhood.

Busiest Hotel



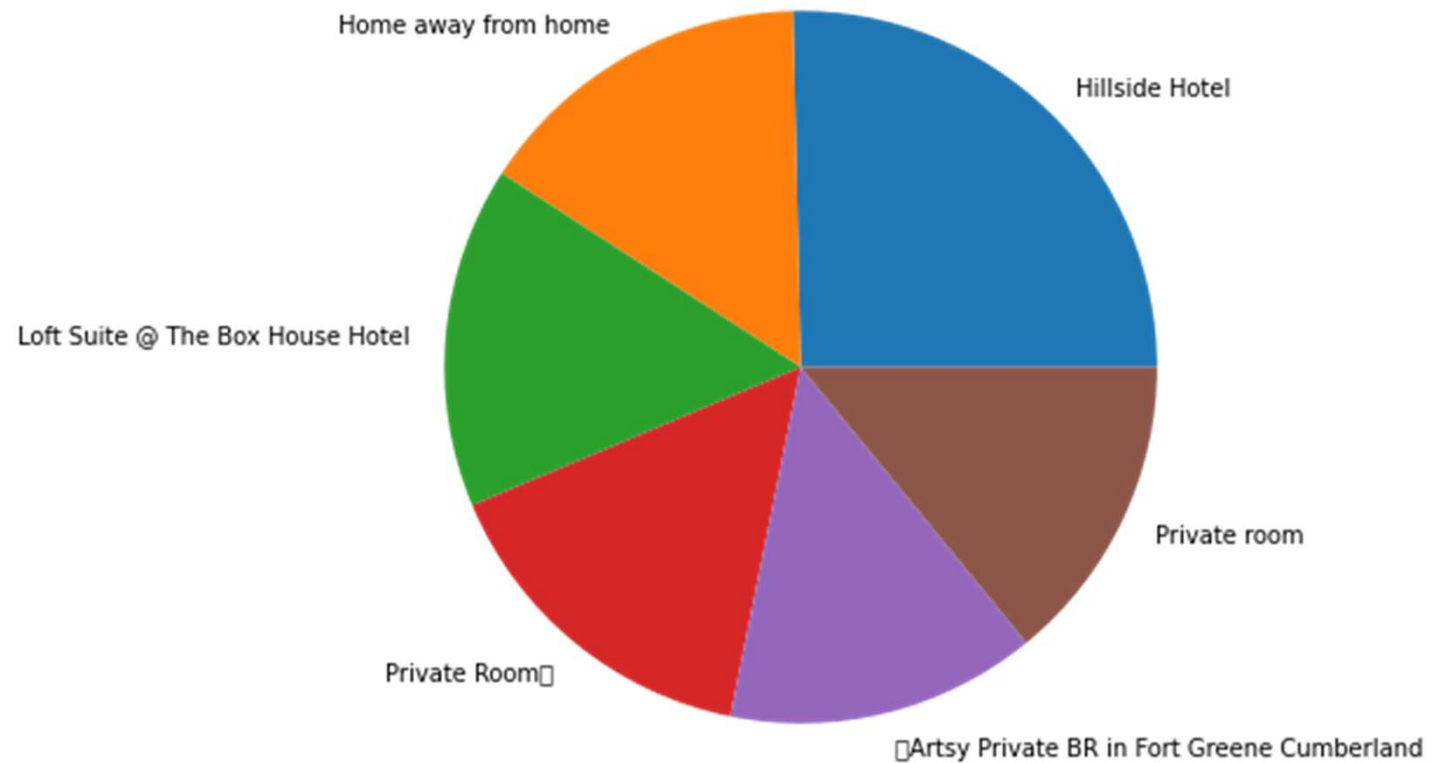
Sonder(NYC) attracts more tenants or tourists across hence making itself the busiest amongst others

Traffic areas



Entire home/apt attracts people to stay for longer durations followed by Private room type

Room Availability



Hillside Hotels shows the highest availability of rooms for visitors.

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

Most visitors don't prefer shared rooms, they tend to visit private room or entire home.

Manhattan, Brooklyn and Queens are the costliest neighbourhood groups.

Queens received highest number of reviews thus making it a trustworthy stay.

Sonders(NYC) is the busiest place among all.

Hillside hotels had the highest number of availability of rooms thus making it favourable place to attract more people to boost the business.



Thank You

As part of EDA Capstone Project by

