

Capstone project -1

Airbnb Bookings Analysis

By- Ganesh Walimbe

What is Airbnb Bookings

- **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.
- The company has come a long way since 2007, when its co-founders first came up with the idea to invite paying guests to sleep on an air mattress in their living room. According to Airbnb's latest data, it has in excess of six million listings, covering more than 100,000 cities and towns and 220-plus countries worldwide.

Methodology

AI

Introduction to data

Data preprocessing

Defining Objectives

Exploratory data analysis (EDA)

Conclusions

The data consists of 16 variables as follows

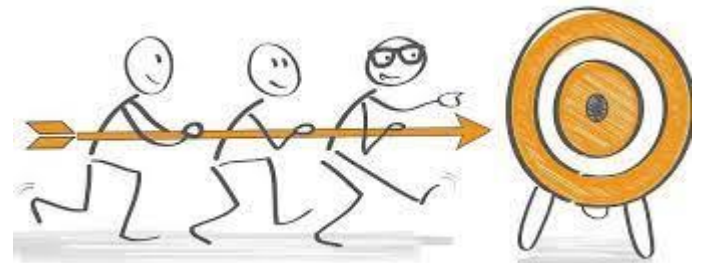
- Id –unique id
- Name – name of the vacation rentals ,cabins,etc
- Host _name –name of the host
- Host_id – id of the host
- Neighbourhood_groups –neighbourhood groups in that location
- Neighbourhood – neighbourhood areas in that place
- Lattitude –lattitude of the place
- Longitutude –longitude of the place
- Room_type –room type of cabin ,rental place

Introduction to data

- Number_of_reviews –number of reviews to the rental rooms
- Last _review – date of the latest review given to the rental place
- Price –price of the room
- Minimum_nights-minimum nights spend in room by guest
- calculated_host_listings_count – count of host listings in rental place
- Availability_365 –availability of the rooms

EDA objectives

- **Understanding the relationship between different hosts and areas.**
- **Finding how factors like prices, reviews affects the renting of the houses .**
- **Finding the top busiest hosts and reasons behind it .**
- **Checking if there any noticeable difference of traffic among different areas**



Data preprocessing

Missing value treatment

The variables containing missing values are name and host_name
There are 16 missing values in name column And 21 missing values in host_name column i dropped the rows which containing missing values

id	0
name	16
host_id	0
host_name	21
neighbourhood_group	0
neighbourhood	0
latitude	0
longitude	0
room_type	0
price	0
minimum_nights	0
number_of_reviews	0



Data preprocessing

Variable selection :

I selected only necessary variables for analysis .

Those variables are as follows :

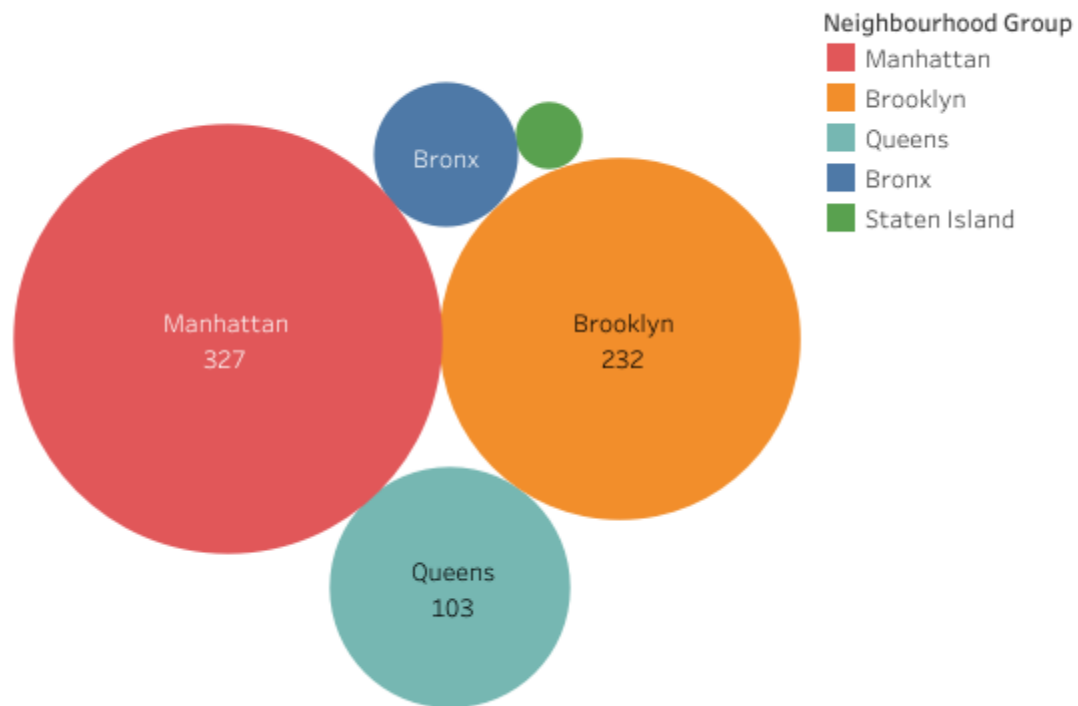
id,host_id,host_name,neighbourhood_group,
Neighbourhood,room_type,price,minimum_nights,
number_of_reviews,calculated_host_listings,availability_365

Understanding the relationship between different hosts and areas.

	host_name	neighbourhood_group	calculated_host_listings_count
13217	Sonder (NYC)	Manhattan	327
1834	Blueground	Manhattan	232
1833	Blueground	Brooklyn	232
7275	Kara	Manhattan	121
7480	Kazuya	Queens	103

From above dataframe we can say that the highest number of hostings are in Manhattan created by host Sonder(NYC) total host listings are 327 and host Blueground 232 total host listings and followed by brooklyn created by host Blueground with 232 host listings

Host and areas relation

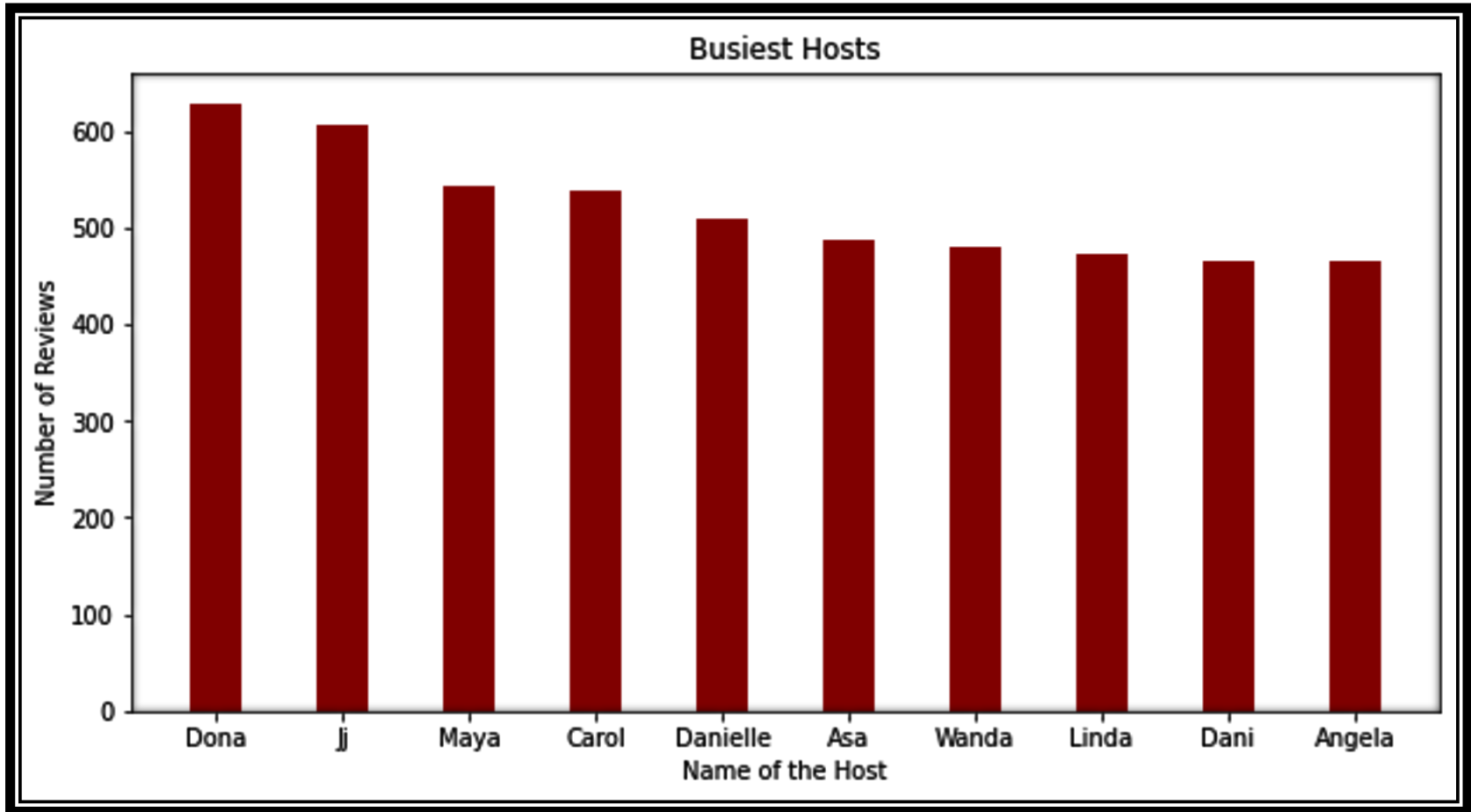


Finding how factors like prices, reviews affects the renting of the houses .



- From the Price v/s Number of reviews scatterplot we can say that most people prefer to stay in place where price is less.

Finding the top busiest hosts and reasons behind it



	host_id	host_name	room_type	number_of_reviews
24474	47621202	Dona	Private room	629
7706	4734398	Jj	Private room	607
22203	37312959	Maya	Private room	543
4589	2369681	Carol	Private room	540
19436	26432133	Danielle	Private room	510
13704	12949460	Asa	Entire home/apt	488
1974	792159	Wanda	Private room	480
5055	2680820	Linda	Private room	474
161	42273	Dani	Entire home/apt	467

From above analysis can see that the busiest hosts are:

- Dona having 629 total number of reviews for private room apartment .
- Ji having having 607 total number of reviews for private room apartment.
- Maya having having 543 total number of reviews for private room apartment.
- Followed by Carol,Dinielle,Asa ,etc are the busiest host .

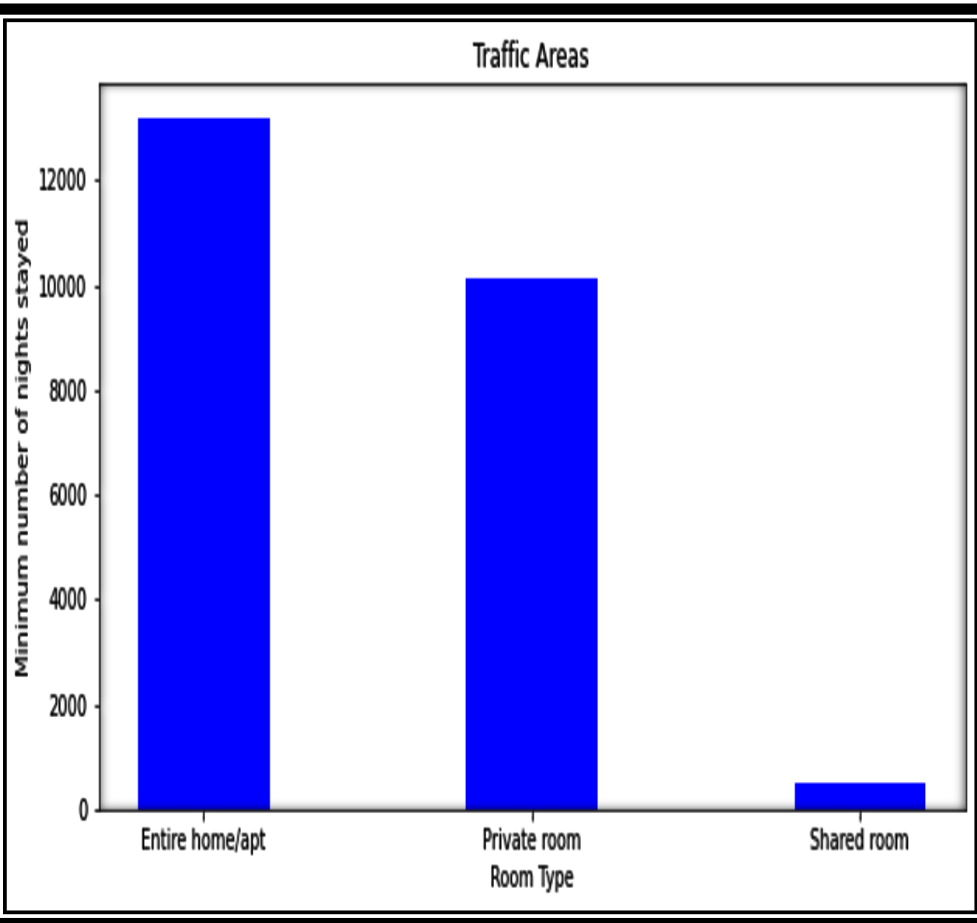
From the above analysis we can see that these host are busiest hosts because people prefer private room and Entire home/apt most rather than other room types.

Checking if there any noticeable difference of traffic among 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

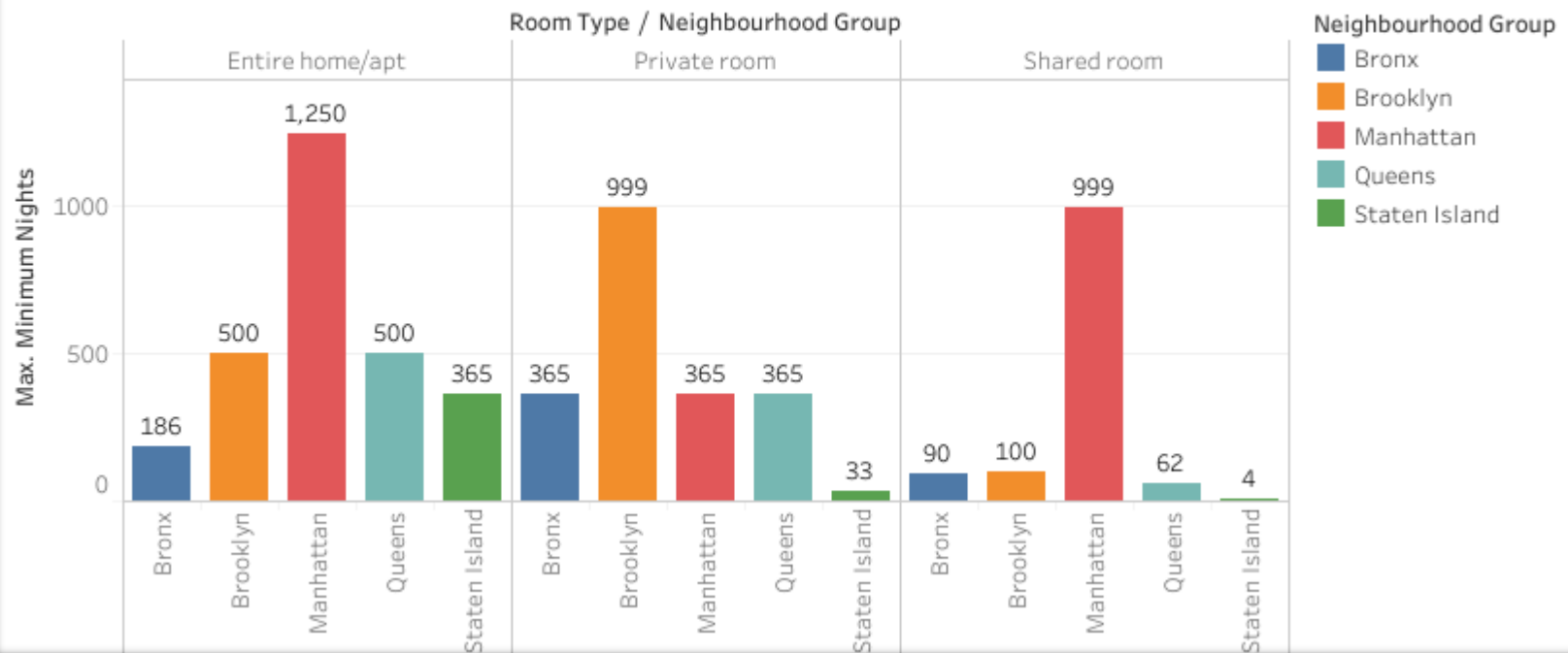
From the data frame we can see that maximum number nights are stayed by people in Manhattan neighbourhood group in entire home room type .followed by Brooklyn neighbourhood group in private room apartment .

Checking if there any noticeable difference of traffic among different areas



From the barplot we can see that people are preferring entire home or private room apartment rather than shared room

Minimum no of nights stayed in room types



Conclusions

- The people are preferring those areas for stay which are having low prices.
- If there are more number of Reviews for particular Neighbourhood group that means that place is a tourist place.
- The highest number of host listings are from Manhattan and are created by Sonder(NYC) and followed by Blueground .
- Busiest hosts are :Dona, Ji, Maya, Carol, Danielle, etc and these host are busiest host because they provides Entire room and private room these room types which are most preferred by people
- People are preferring Entire home/apt or Private room which are present in Manhattan, Brooklyn, Queens
- If people are not staying more then one night means they are travellers.

THANK YOU