Capstone Project
Project Title: Airbnb Bookings Analysis

By an individual

- Ashutosh Kumar Dubey

LET’S START THE ANALYSIS OF THE BOOKINGS

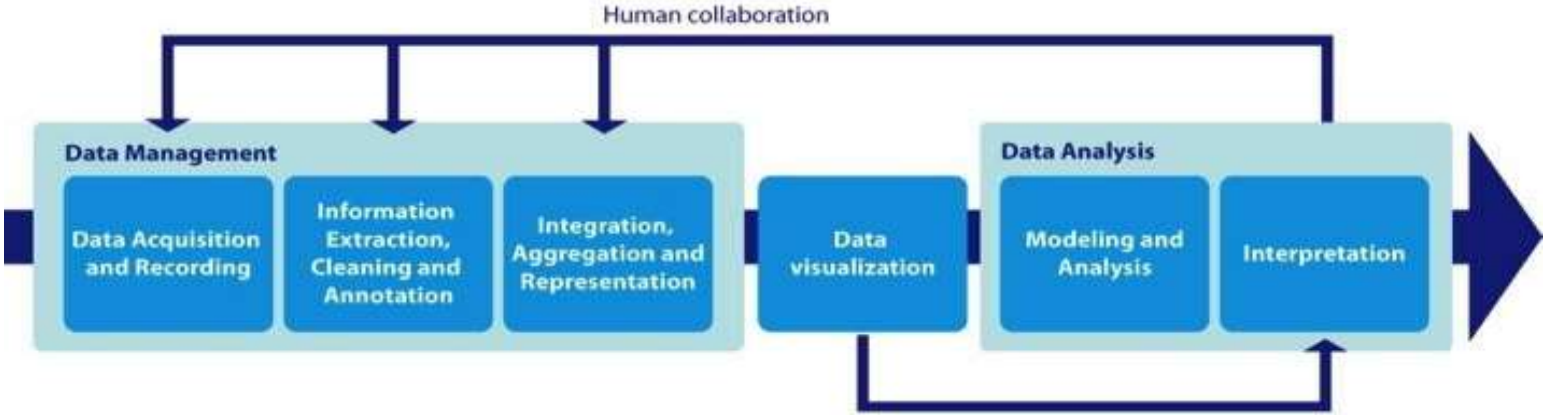
1. Introduction to Airbnb
2. Defining problem statement
3. EDA with Python Programming
4. Data analysis
5. Data wrangling and Visualization
6. Observation and conclusion

Introduction

- Airbnb(AirBed&Breakfast) is an online marketplace that connects people who want to rent out their homes with people who are looking for accommodations in specific locales.
- Airbnb offers people an easy, relatively stress-free way to earn some income from their property.
- Guests often find Airbnb is cheaper, has more character, and is homier than hotels.

Data Pipeline

- Data processing-1: In this first part we've removed unnecessary features. Since there were nearly many columns with all null values.
- Data processing-2: In this part, we manually go through each column selected from part 1, And encoded the categorical and filled the null values by 0
- EDA: In in this part, we do some exploratory data analysis (EDA) on the features selected in part-1 and 2 to see the trend.



How Airbnb works

1. Hosts sign up on the platform and list their spaces
2. Travelers browse the website and book spaces
3. Admins inform hosts of bookings
4. Travelers pay the rental fee
5. Travelers check in and check out
6. Travelers can rate and review their stays

PROBLEM STATEMENT

- Since 2008, guests and hosts have used Airbnb to expand on traveling possibilities and present a more unique, personalized way of experiencing the world. Today, Airbnb became one of a kind service that is used and recognized by the whole 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.

Let’s go through the ‘Dependent Columns’

- id : a unique id identifying an Airbnb listing
- name : name representing the accommodation
- host_id : a unique id identifying an Airbnb host
- host_name : name under whom host is registered
- neighborhood_group : a group of area
- neighborhood : area falls under neighborhood group
- latitude : coordinate of listing
- longitude : coordinate of listing
- room_type : type to categorize listing rooms
- price : price of listing
- minimum_nights : the minimum nights required to stay in a single visit
- number_of_reviews : total count of reviews given by visitors

Correlation Heatmap

- *Correlation between different variables*

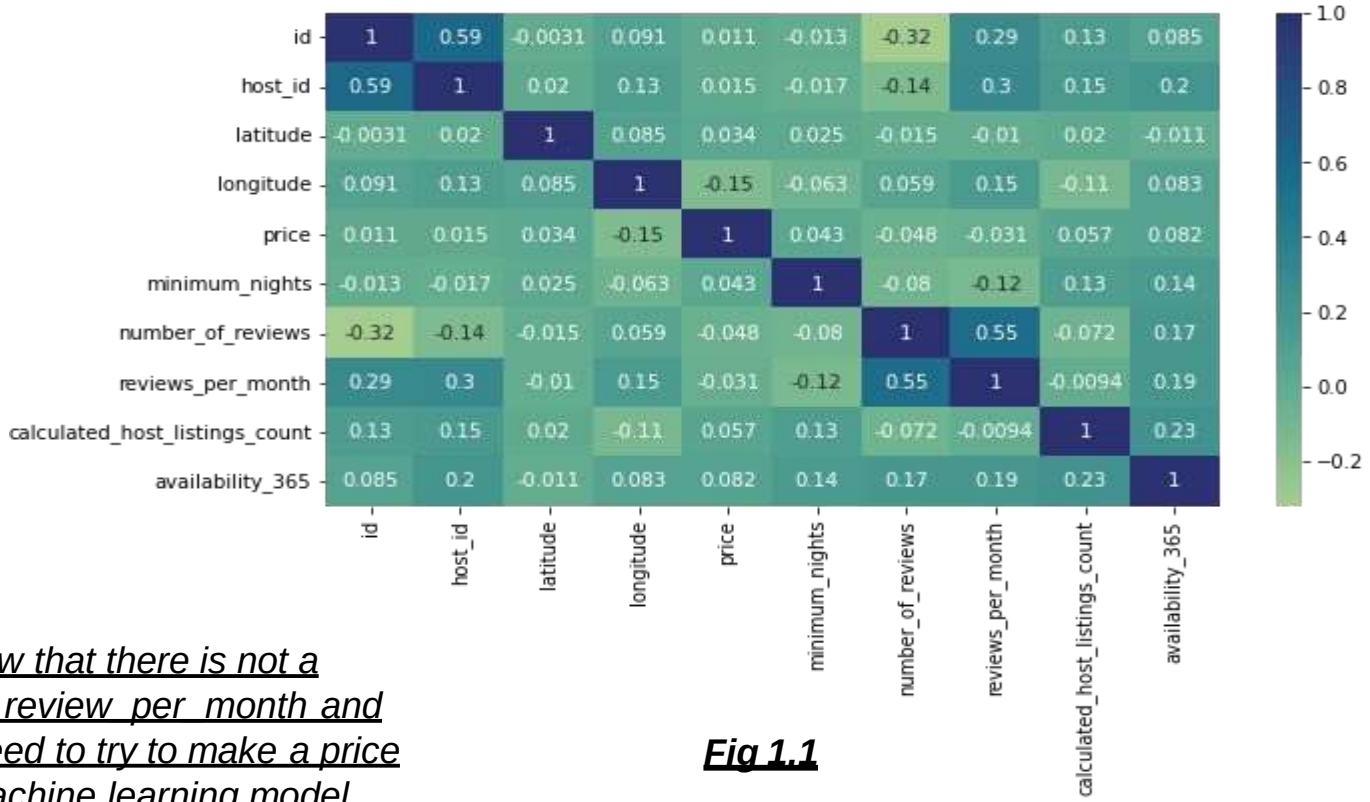


Fig 1.1

From the Fig 1.1, we know that there is not a strong correlation except review_per_month and number_of_review.We need to try to make a price prediction using some machine learning model.

Dependent Variable

The pie chart represents the Neighborhood group as a function of list of Counting's Neighborhood group here is a dependent variable.

From Fig 1.2 we got to know the percentage of every neighborhood group as

- Brooklyn : “**41.1%”**
- Manhattan : **44.3%**
- Staten Island : **0.8%**
- Bronx : **2.2%**
- Queens : **11.6%**

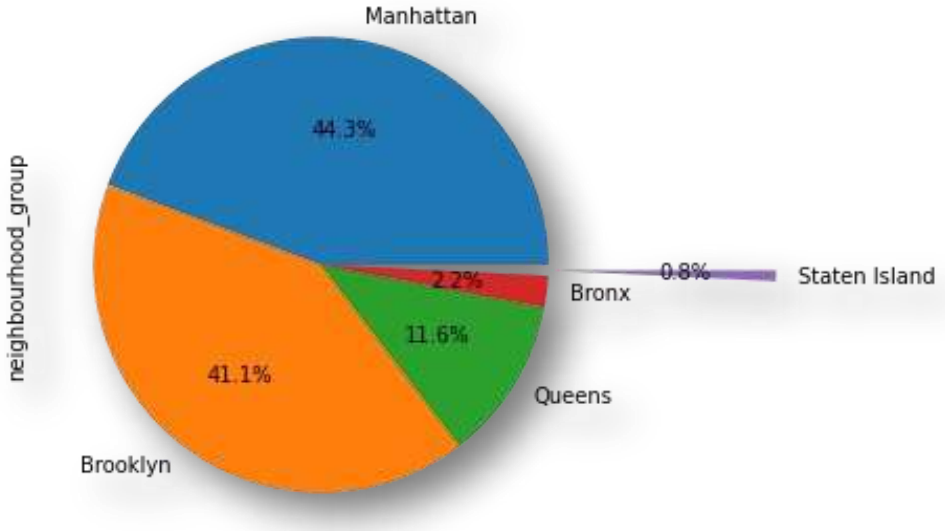


Fig 1.2

- Conclusion: The maximum number of listings are from MANHATTAN group followed by Brooklyn..

Map of Neighborhood Group with respect to Longitude and Latitude

In Fig 1.3 we have represented the neighborhood groups as

- Brooklyn : “**Blue”**
- Manhattan : **“Orange”**
- Staten Island : **“Green”**
- Bronx : **“Red”**
- Queens : **“Purple”**

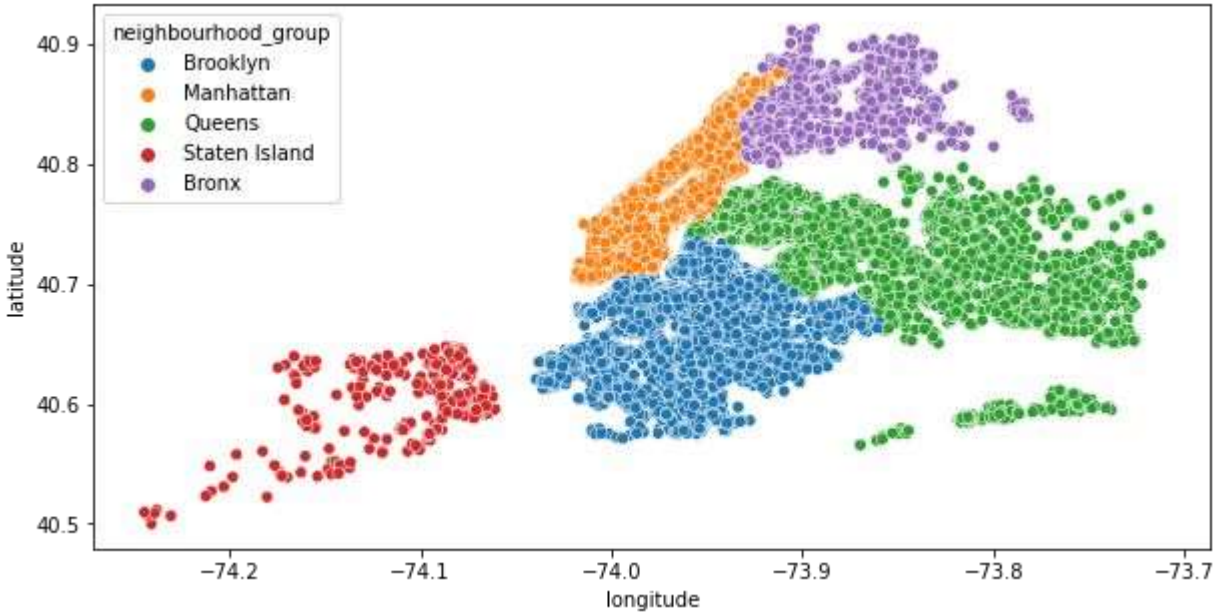


Fig 1.3

Predictions for location

- From the Analysis of Fig 1.4 we can say that most people prefer to stay at a place where price is less and the number of reviews are high and in this case ***Queens*** is that neighborhood area.

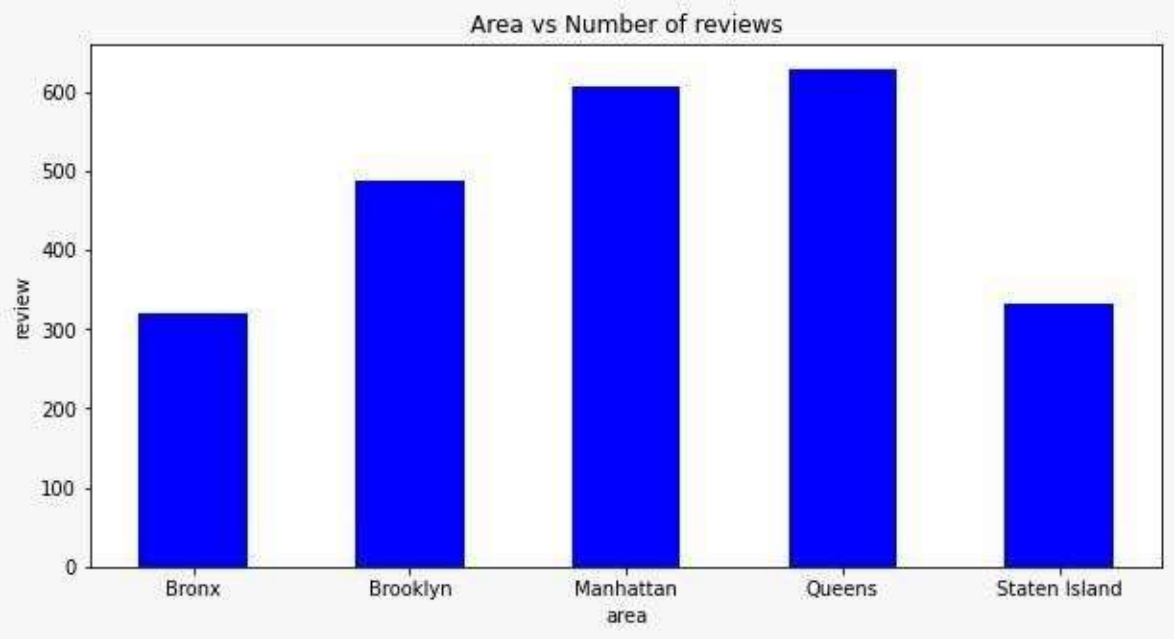


Fig 1.4

Predictions for Max Price

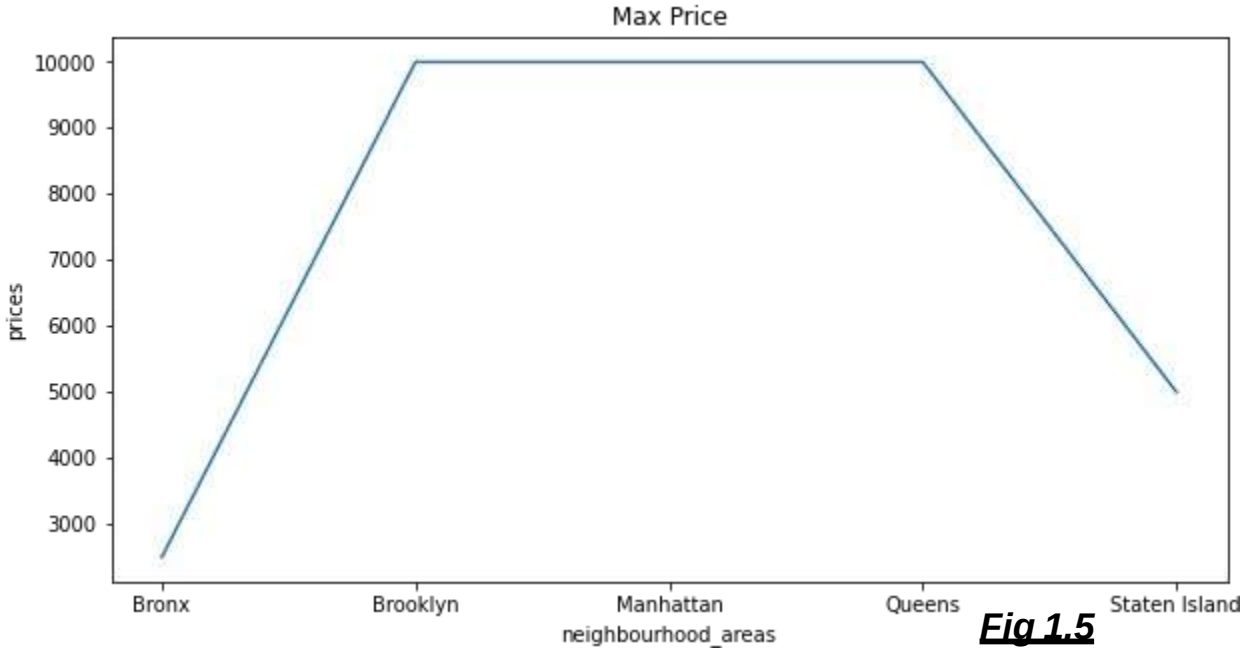


Fig 1.5

From fig. 1.5 we can conclude that the neighborhood groups **Brooklyn, Manhattan and Queens** have Maximum price of **'$10,000' each** for lodging.

Predictions for Min Price

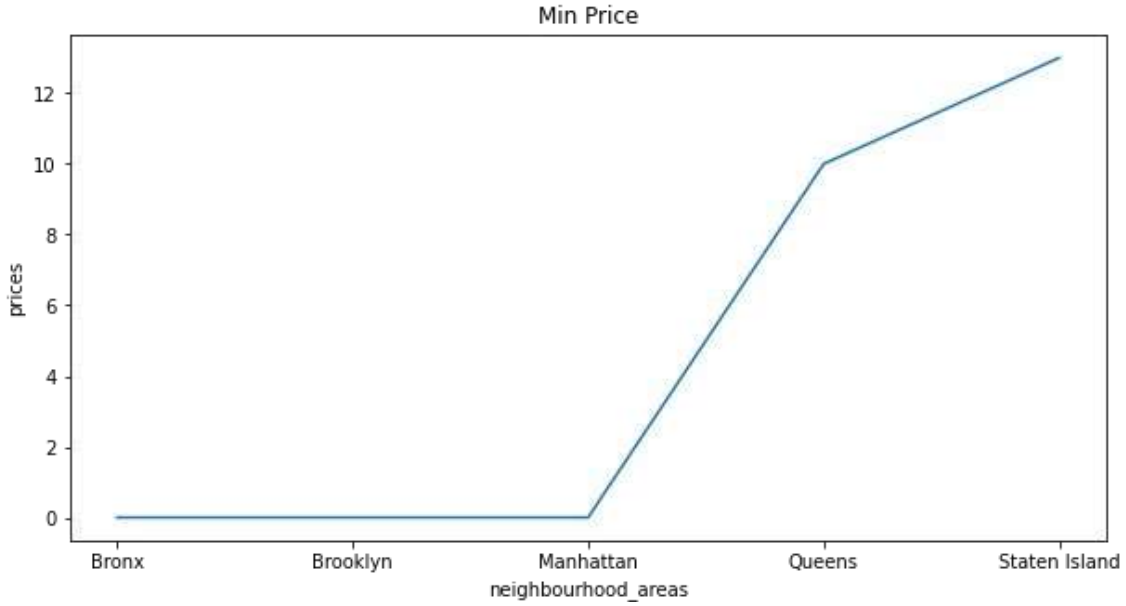
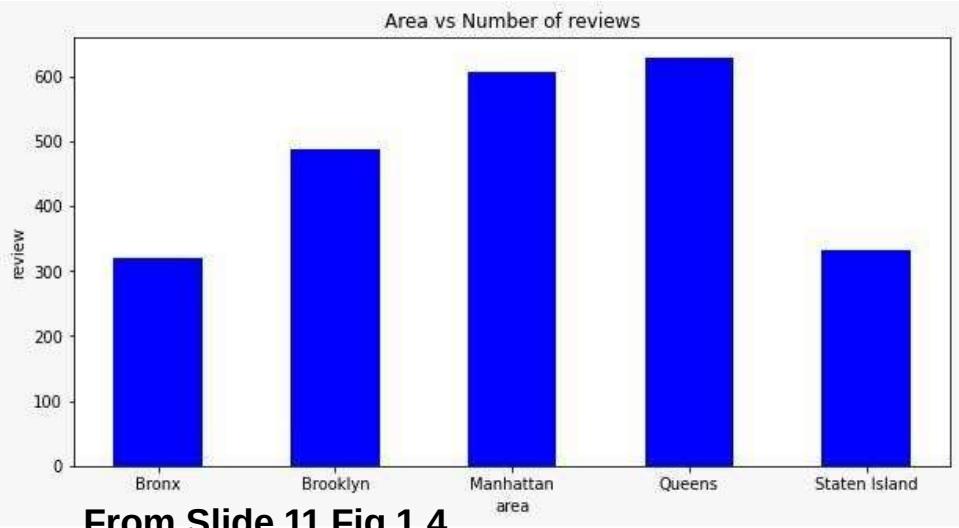


Fig 1.6

From fig. 1.5 we can conclude that the minimum price of lodging for neighborhood groups **Bronx, Brooklyn and Manhattan are '$0' each**

Neighborhood & Price plot



From Slide 11 Fig 1.4

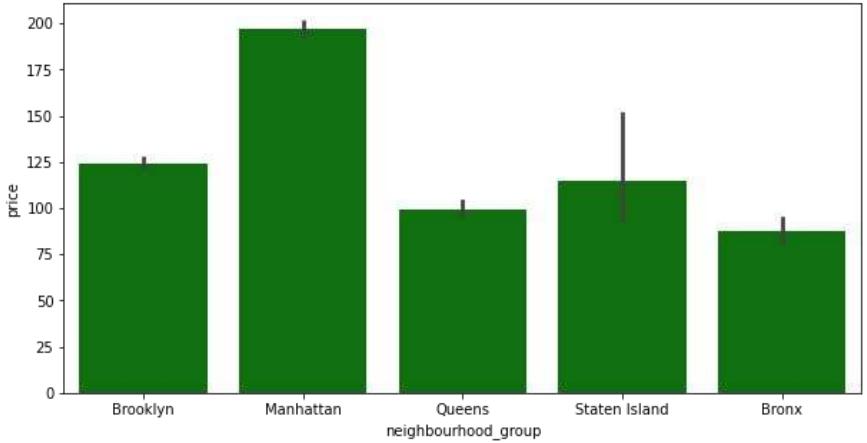


Fig 1.7

In fig 1.7 the Neighborhood Group **Manhattan** has a price range between **‘175-200$’ (‘Maximum’**). **Queens** has a price range between **’75-100$’ (‘Second Lowest’)**

From both the figures, we can conclude that the customer will choose Queens (less price and great reviews)

Which of the host are busiest ?

host_name	Neighborhood_group	minimum_nights	
13217	Sonder (NYC)	Manhattan	327
1834	Blueground	Manhattan	230
9742	Michael	Manhattan	212
3250	David	Manhattan	202
9741	Michael	Brooklyn	159
6808	John	Manhattan	151
3249	David	Brooklyn	142
7275	Kara	Manhattan	135
432	Alex	Manhattan	134
9856	Mike	Manhattan	134

From Fig 1.8
The Busiest Host are

1. Sonder (NYC)
2. Blueground
3. Michael
4. David
5. Michael

6. John
7. David
8. Kara
9. Alex
10. Mike

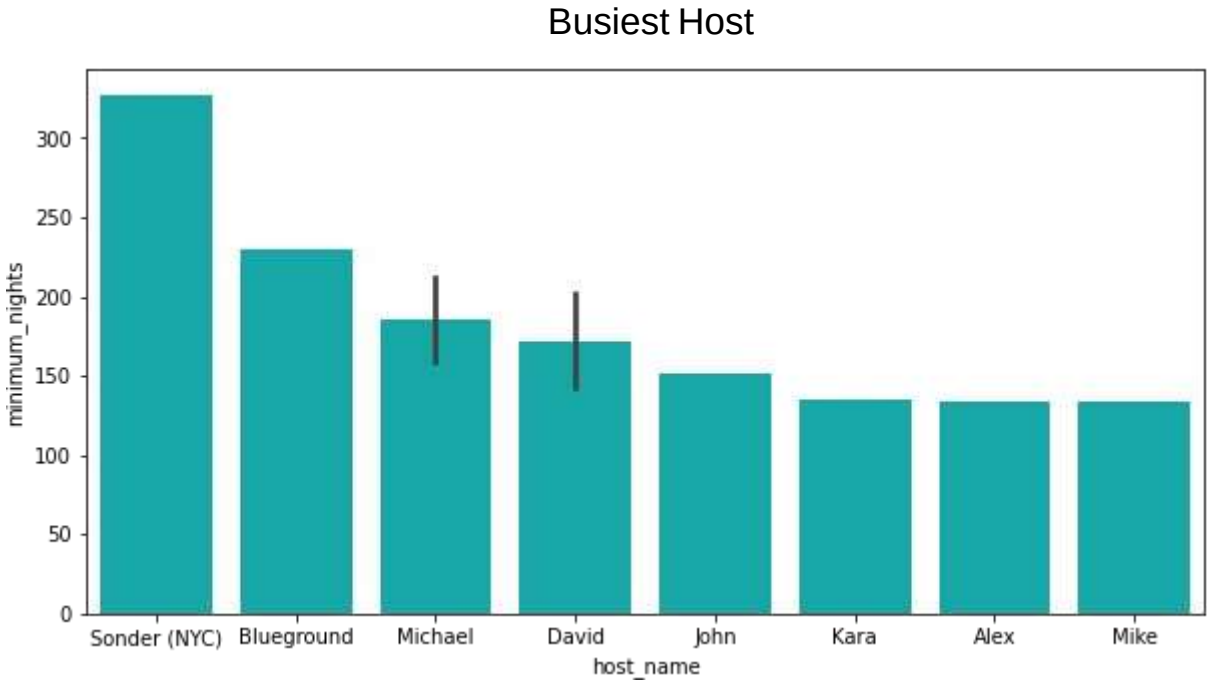


Fig 1.8

TRAFFIC AMONG DIFFERENT AREA WITH NEIGHBOURHOOD GROUP

From the above fig 1.9
We can conclude that the density of price distribution of the neighborhood group **Manhattan** is the most Uniform as compared to any other. While that of **Bronx** is the most Non-Uniform we can state that Manhattan has the highest range of prices for the listings with $150 price as average observation, followed by Brooklyn with $90 per night. Queens and Staten Island appear to have very similar distributions, Bronx is the cheapest of them all.

Fig 1.9

Traffic Among Different area with Neighborhood Group

From Fig 1.10 We can clearly observe that Manhattan and Brooklyn have more private and entire home/apt room type So this could be the reason because of which Brooklyn and Manhattan have more traffic than other neighborhood groups.(we have already seen above the private and entire home/apt room type are mostly preferred by people)

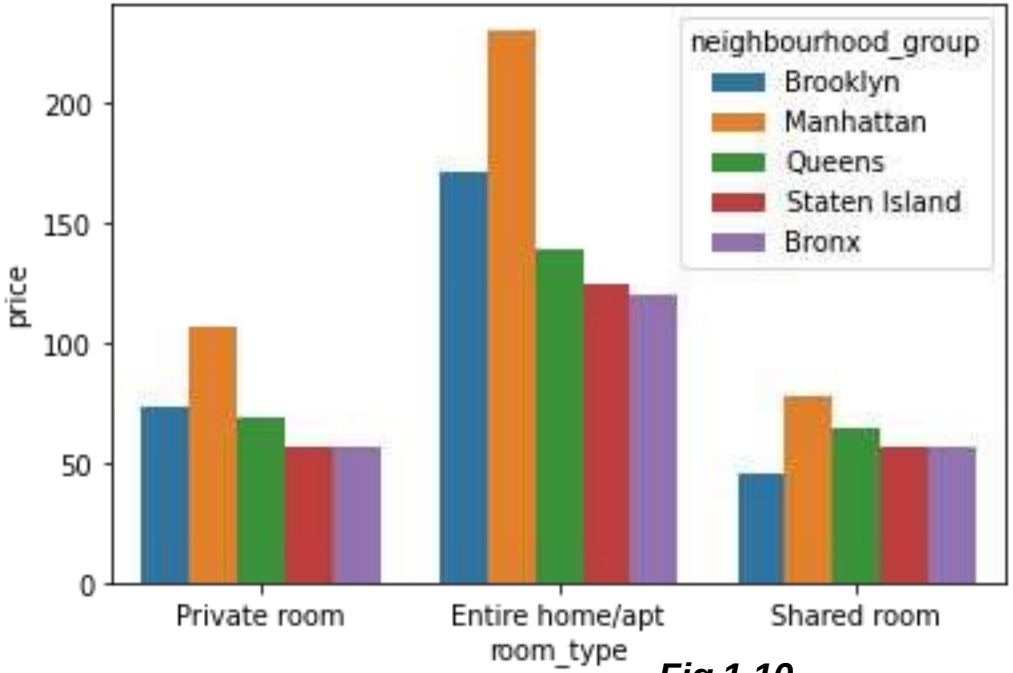


Fig 1.10

Conclusion

- The people who prefer to stay in an Entire home/Apartment stay for a long period of time in that particular Neighborhood only hence, they can be thought of being staying with family or friends. More number of people are there, hence the entire home/ apartment has been booked.
- The people who prefer to stay in Private room (probably businessmen) don't stay for longer as compared to those who prefer to stay in entire Home or Apartment.
- Most people prefer to pay less price for the accommodation they're in.
- The more the number of Reviews for a particular Neighborhood group, the more the chances of the group being a tourist place are there.
- People not preferring to stay for more than a day can most probably be termed as travelers (probably on a business trip or for some sight-seeing purposes).


THANK YOU