

Capstone project- 1

EDA on Airbnb booking

By-

Incredible Data Scientist:

Suvendu Dey

Abishek Kumar

Ranjit Biswal

❖ Problem Statement



- ❖ For this project we are analyzing Airbnb's New York City (NYC) data of 2008. NYC is not only the most famous city in the world but also top global destination for visitors drawn to its museums, entertainment, restaurants, and commerce.
- ❖ Our main objective is to find out the key metrics that influence the listing of properties on the platform. For this, we will explore and visualize the dataset from Airbnb in NYC using basic exploratory data analysis (EDA) techniques.
- ❖ Data analysis on thousands of listings provided through Airbnb is a crucial factor for the company.
- ❖ We will be finding out the distribution of every Airbnb listing based on their location, including their price range, room type, listing name, and other related factors.

❖ Understanding the Data

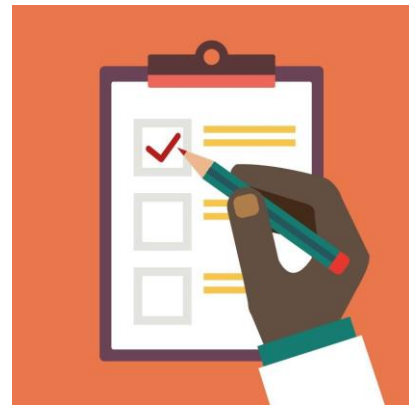
- There are 49,000 observations with various types of fields in our dataset.
- List of fields:

<ul style="list-style-type: none">■ Id■ Name■ Host_id■ Host_name■ Neighbourhood_group■ Neighbourhood■ Latitude■ Longitude■ Room_type	<ul style="list-style-type: none">■ Price■ Minimum_nights■ Number_of_reviews■ Last_review■ Reviews_per_month■ Calculated_host_listing_count■ availabilty_365
--	--

❖ Agenda

→ **We try to answer following questions for Airbnb:**

- Relationship analysis between different variables
- What Can we learn About Different hosts and areas?
- What can we learn from predictions?
- Which hosts are the busiest and why?
- The highest host are 1. sonder (NYC) 2.Blueground 3.Michal
- Finding Total count of each room types?
- Room available per Neighborhood Group
- Price on different neighbourhood_group
- Rooms availability in different areas



Room types and their relation with availability in different neighborhood groups?

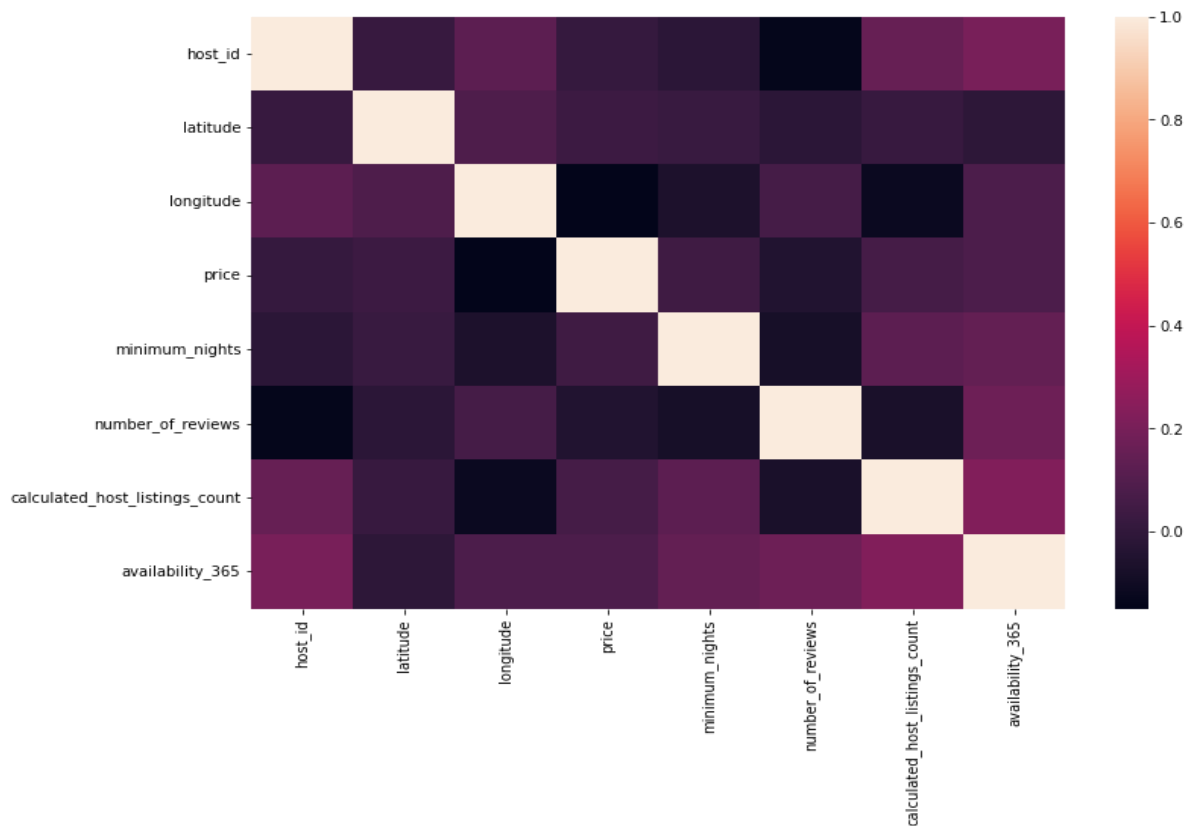
- Total no. of nights spends per room types
- Top 10 highest listing neighborhood
- Which are the top 25 most used words in listing names?
- Find top 10 hosts with most listings
- Find the top three hosts based on their turnover
- Find total no. of nights spend per location

Map of New York City (NYC)



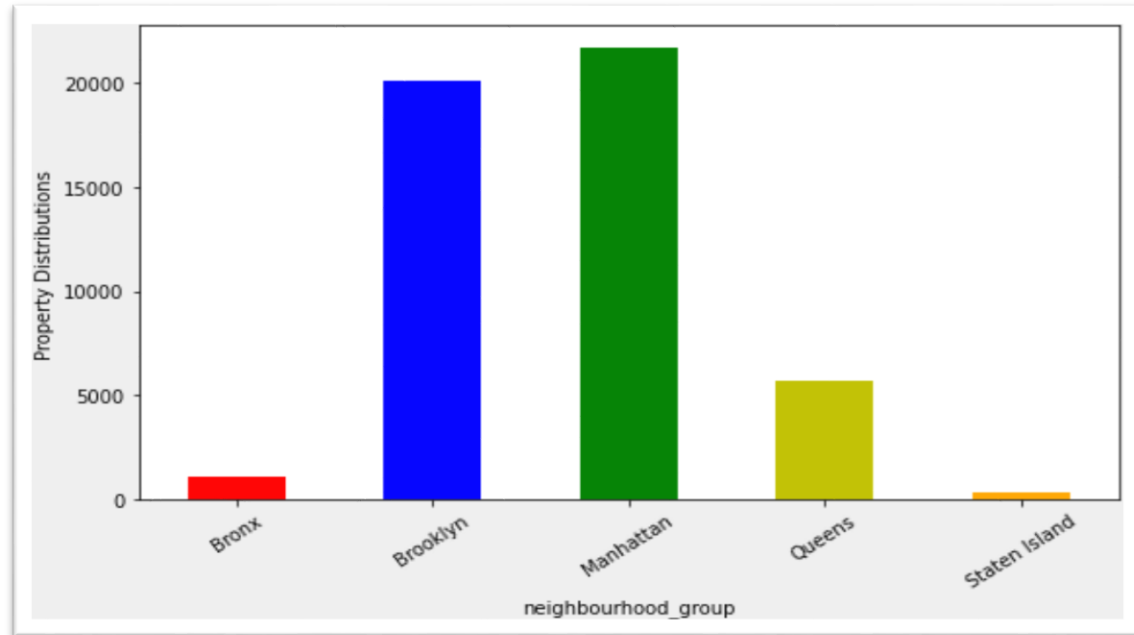
Relationship analysis between different variables:

1. **host_id**
2. **latitude**
3. **longitude**
4. **price**
5. **minimum_nights**
6. **number_of_reviews**
7. **calculated_host_listings_count**
8. **availability_365**



❖ Number of active hosts per location (Where most of the Hosts interested to own property?)

- Manhattan and Brooklyn are the most preferred place by hosts.
- Bronx and Staten Island is low traffic for host as compared to other location
- Queens has a average number of host.

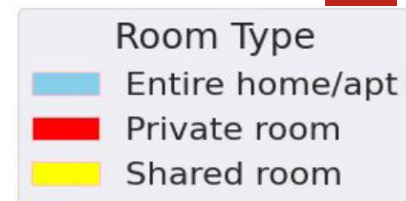
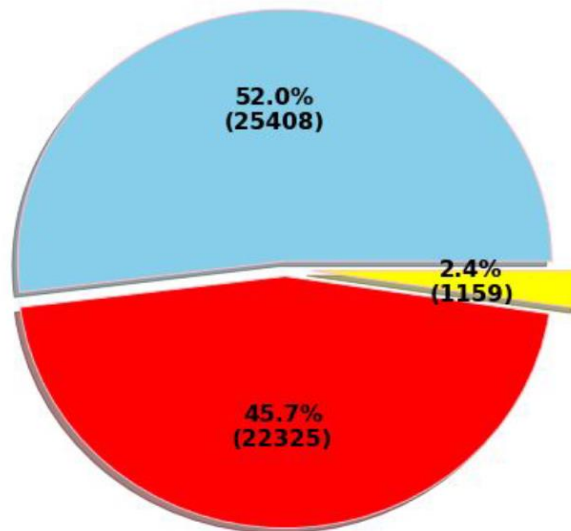


In this analysis we will find out 3 types of rooms available are there:

1. Private room
2. Entire home/apt
3. Shared room

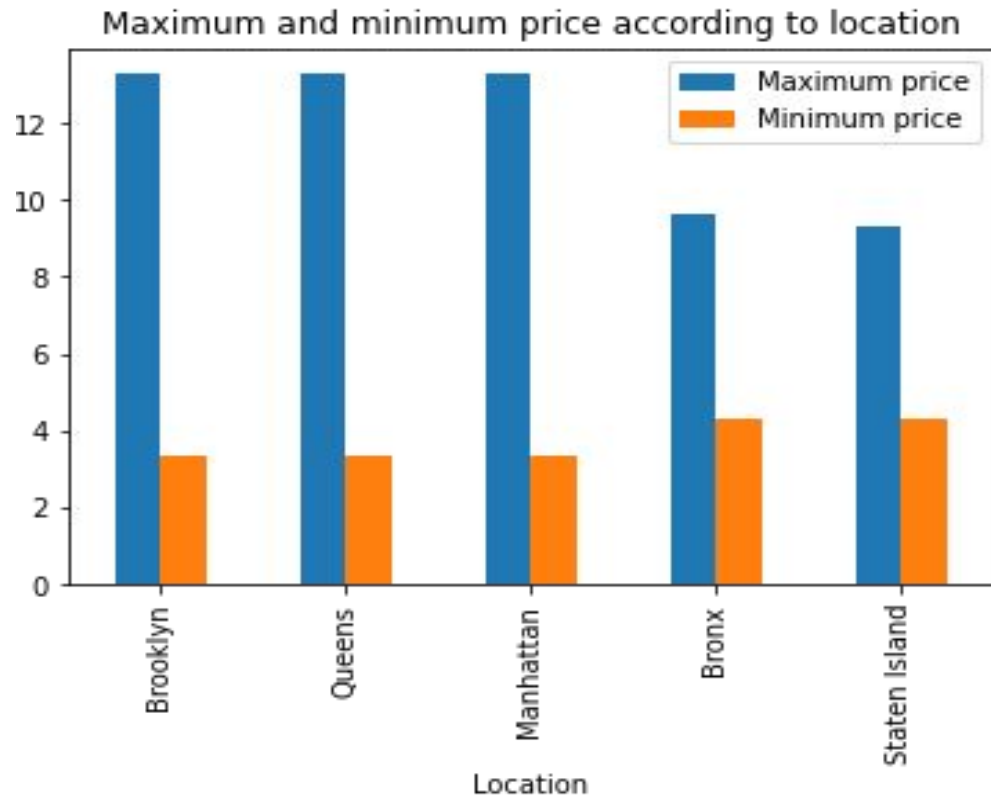
In this chart entire home/apt is the highest rooms availability room are 25409 are in there, and second private room availability is 22326 shared room are available room 1160

Count of Listed Rooms



❖ Where does the customers pay the highest and lowest rent according to location?

- Maximum price of rent in Brooklyn, Queens, Manhattan is almost same.
- Here we used log transformation to display minimum price with maximum price.
- Minimum price in Brooklyn, Queens, Manhattan is also same.



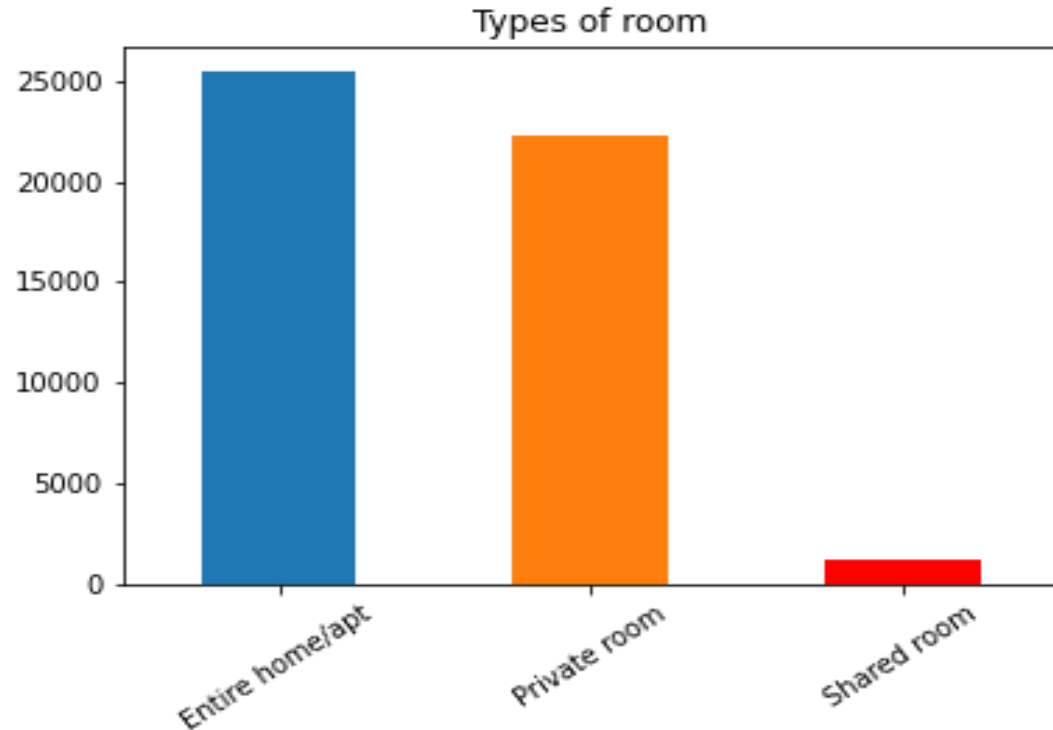
❖ Total count of each room types as per listing.

➤ Based on the Analysis we found that:

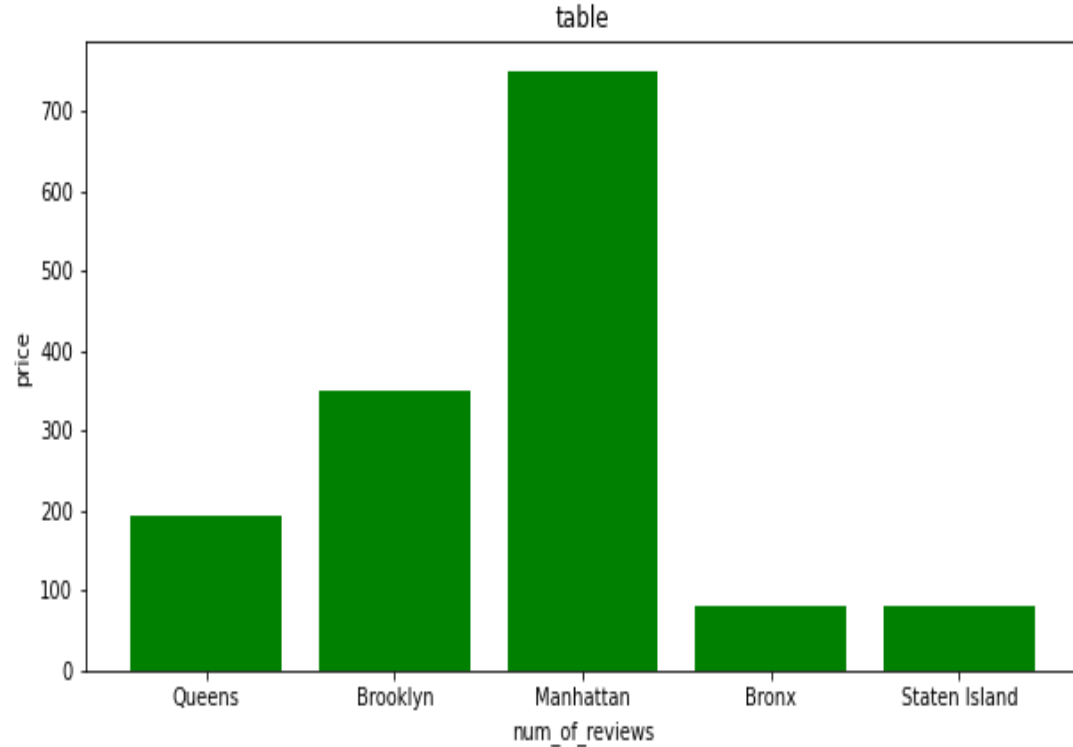
→ **Entire home/apt** has the highest number of listings of among other room types.

→ **Private room** has the second position of highest demanding then other room types.

→ **Shared Room** is the least listed room types among other room types.

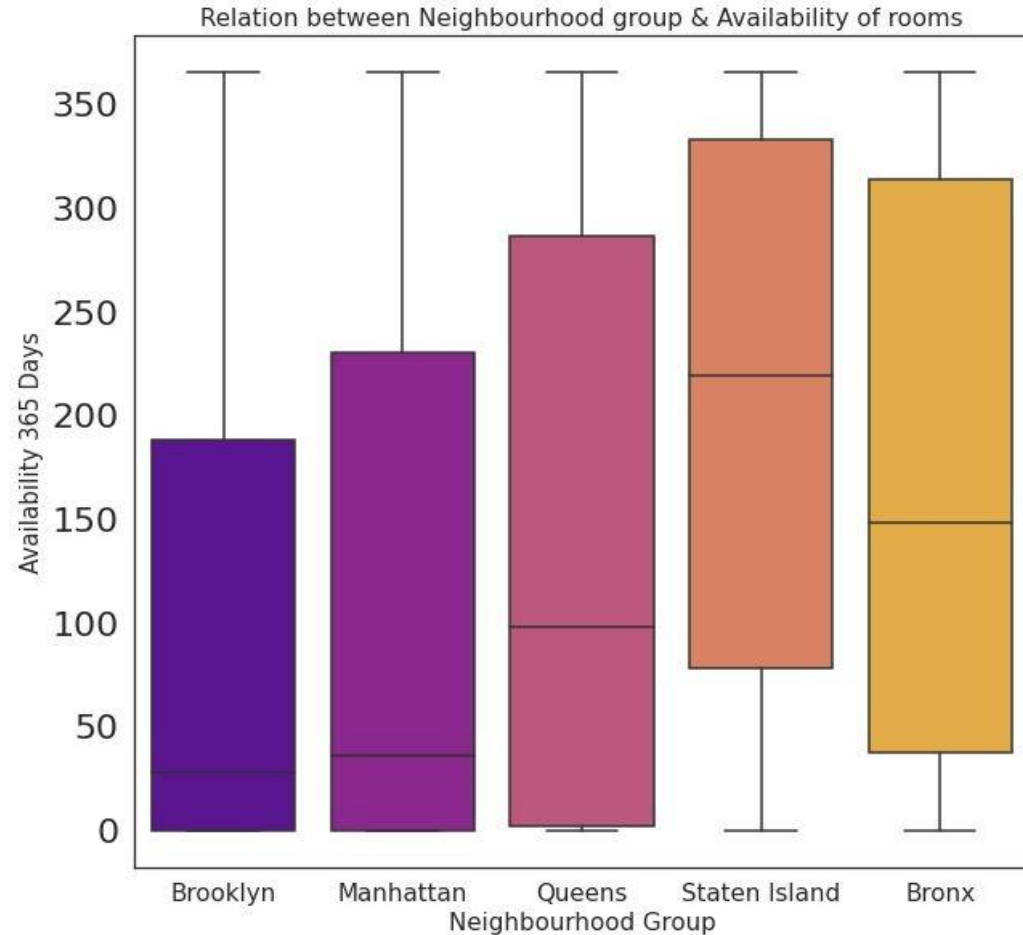


- I did analysis that this based on reviews that Manhattan is the most famous location preferred by peoples.
- Brooklyn is the second highest reviewed location preferred by peoples
- And Queens is the 3rd highest reviewed location preferred by people



Room types and their relation with availability in different neighborhood groups?

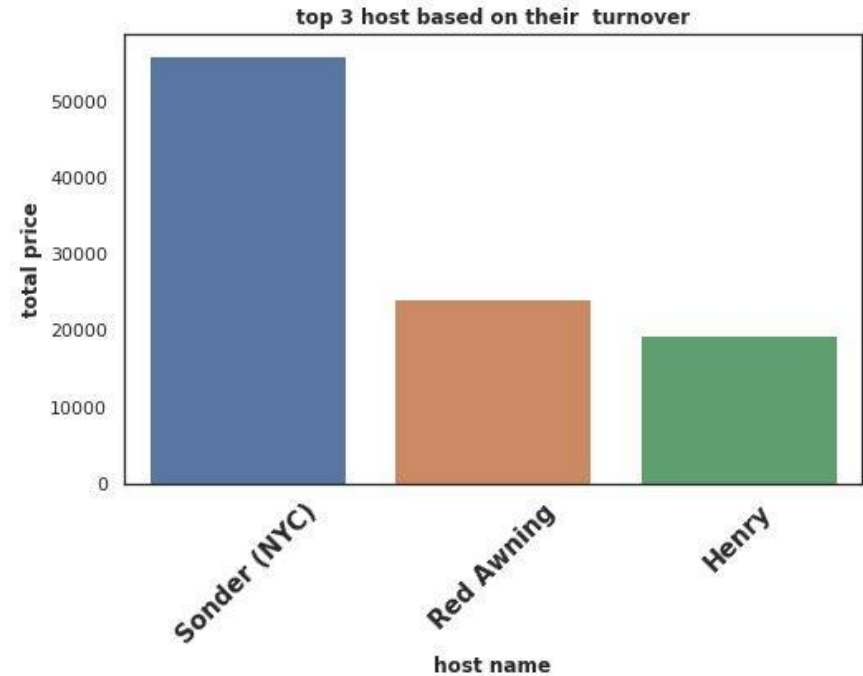
- ❑ Brooklyn and Manhattan have the least availability of rooms overall, as low as 0 days.
- ❑ Staten Island and Bronx has the highest availability rate overall at around 300 days.
- ❑ Form this analysis we can say that people stay for longer duration of time in Private rooms in Brooklyn and Manhattan

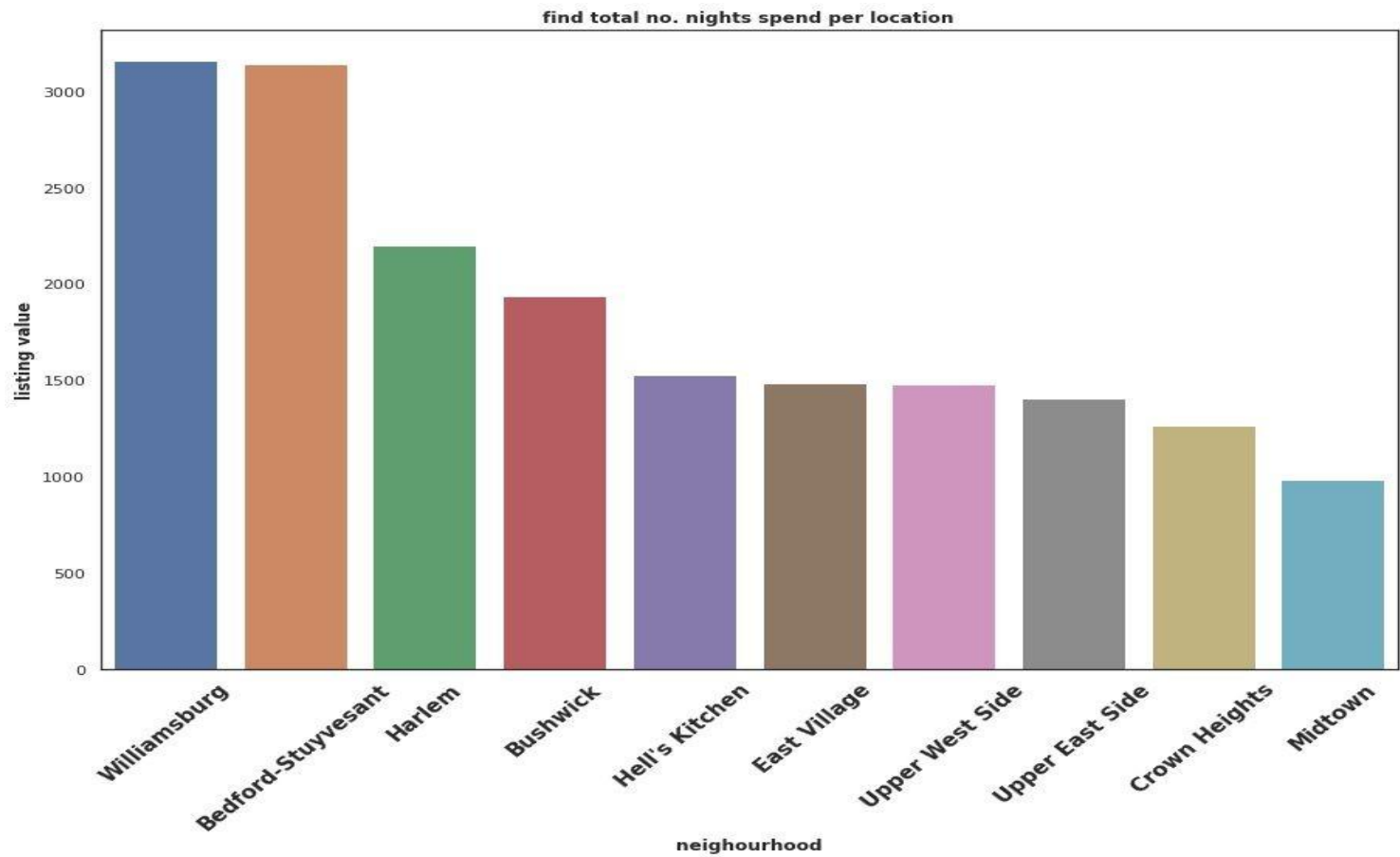


Top three hosts based on their turnover



- Here we are trying to find top three host based on their turnover and compare between those three hosts.
- We can state that Sonder (NYC) is the top host based on his high turnover. Red Awning and Henry secured 2nd and 3rd position respectively.
- Total turnover of Sounder (NYC) is more than 50000\$. Red Awning and Henry is quite behind from the 1st place.





Challenges Faced

- Reading the dataset and understanding the meaning of some columns.
- For answering some of the questions we had to understand the business model of Airbnb that how they work.
- Handling Nan values, null values and duplicates.
- Designing multiple visualizations to summarize the information in the dataset and successfully communicate the results and trends to the reader.





Analysis Summary:



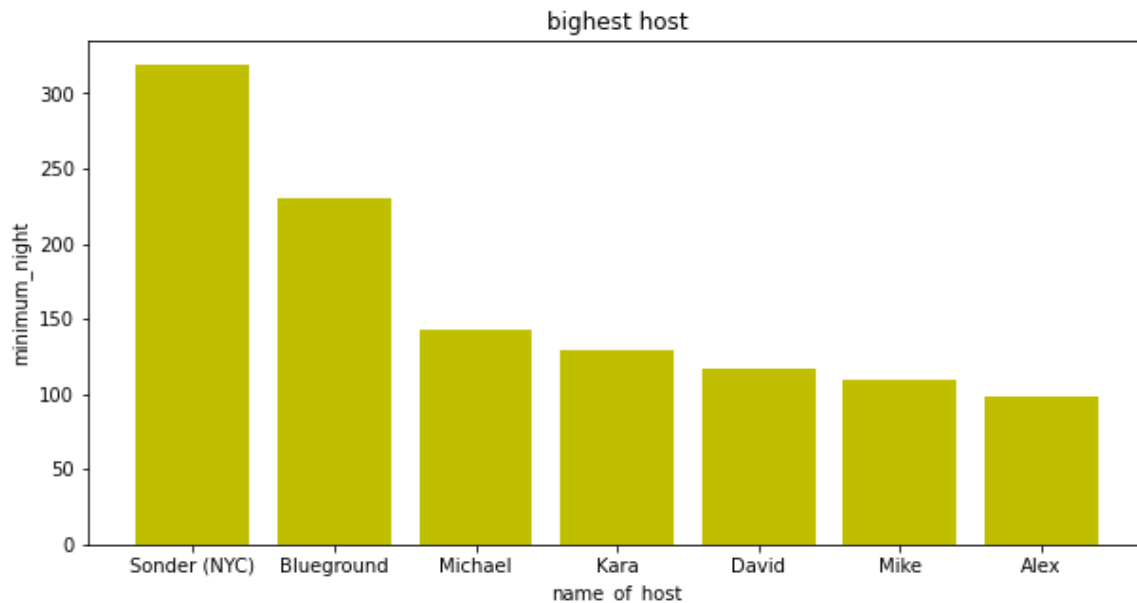
→ We defined some points which can help Airbnb in their business:

- Manhattan is the most focused place in New York for hosts to do their business
- Customers pay highest amount in Brooklyn, Queens and Manhattan that is \$10,000 and lowest amount is \$10.
- For the three types of room type (i.e. Entire home, Shared room, & Private room) average price of entire home is around \$157, for Shared room is around \$60, and for private room is around \$75.
- Top three host base on their turnover are Redawning, Henry and best host is Sonder (NYC)

Top 7 Busiest Host Name are:

- In this Bar Graph We Saw that sonder (NYC) minimum Nights is 319 and it is the topest host of this Airbnb project, In Blueground Minimum night 230

And Michael minimum night 143, kara minimum night 129, David minimum night 117, mike minimum night 109 and Alex minimum night 198



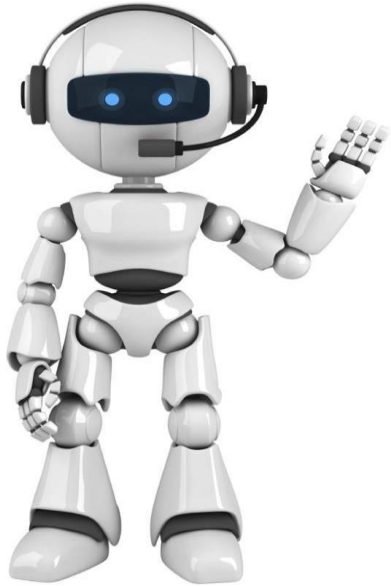
Analysis Summary:



‘Entire home/apt’ room type has the highest number of listing of 52% and ‘Shared Room’ is the least listed room type at only 2.4% in total.



- **People stay for longer duration of time in Private rooms in Brooklyn and Manhattan.**
- **Words such as ‘bedroom’, ‘cozy’, ‘private’, ‘apartment’ and ‘spacious’ are used more frequently than words such as ‘park’, ‘near’, ‘village’ and ‘heart’.**
- **Count of listing by top 10 hosts is almost 2.5% (1270 listings) of the whole dataset.**
- **More customer preferred Manhattan location for night stay than Brooklyn**
- **63.2% customer spend night in Entire home and 1.6% spend night in Shared room**



Thank you

AlmaBetter