

Airbnb NYC 2019

Analysis

Team member

Priya Debrani

According to Airbnb's latest data it has in excess of 5.6 million listings covering more than 100000 cities and towns and 220+ countries worldwide.

Abstract:

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. Hosts and travelers use Airbnb platform to list their properties and book their stay respectively.

The dataset consists of 49000 records of Airbnb listings in New York City and my task was to explore and find insights about hosts, neighbourhoods, and room types which can help the company, hosts, and travellers to make better decisions. The main purpose of EDA is to detect any errors, outliers as well as understand patterns in the data. EDA on the Airbnb listings data will help us understand many insights about the hosts, neighbourhood groups, room types, prices etc.

1.Introduction

Airbnb 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 2008, when its co-founders first came up with the idea to invite paying guests to sleep on an air-mattress in their living room.

□ The idea behind Airbnb is simple:

- Hosts list out their property details on Airbnb along with other factors like pricing, amenities provided etc.
- Airbnb sends a professional photographer (if available) to the property location in order to take high quality photographs.
- Travelers search for a property in the city where they wish to stay and browse available options according to the price, amenities etc.
- Booking is made through Airbnb where travellers pay the amount mentioned by the host and some additional money as transaction charges.
- Host approves the booking, the travellers stays there and finally Airbnb pays the amount to the host after deducting their commissions.
- Airbnb's business model is quite profitable. The company like Uber, Lyft and others has capitalized on the sharing economy, essentially making money renting out property that it doesn't own. Every time a reservation is made, Airbnb takes a cut.

- Exploratory Data Analysis is a set of techniques that were developed by Tukey, John Wilder in 1970. The philosophy behind this approach was to examine the data before building a model. Exploratory Data Analysis or EDA is used to take insights from the data. Data Scientists and Analysts try to find different patterns, relations, and anomalies in the data using some statistical graphs and other visualization techniques.
- The main purpose of EDA is to detect any errors, outliers as well as to understand different patterns in the data. It allows analysts to understand the data better before making any assumptions. The outcomes of EDA helps businesses to know their customers, expand their business and take decisions accordingly.

2.Problem Definition:

The dataset has around 49000 observations and 16 columns in it which is a mix of categorical and numerical variables. The main objective here is to explore and analyse the data to find out insights such as:

- 1) What can we learn about different hosts and areas?
- 2) How are rentals distributed among the five boroughs?
- 3) What's the price distribution and what's the range of fair prices available?
- 4) Which hosts are the busiest?
- 5) Which are the top 10 hosts on the basis of reviews?
- 6) Which are the top 10 hosts on the basis of count of listings?

3.EDA on given Data set

EDA will help us generate such insights not limited to these which will help the company, hosts and travelers to make better decisions in future.

The different features the dataset contains:

id: It is an unique id given to the property listed in airbnb NYC which is a numerical variable.

name: It represents the name of the airbnb listed property which is a categorical variable.

host_id: This is an unique id given to the host of the property which is a numerical variable.

host_name: The name of the host of the property listed which is a categorical variable.

neighbourhood_group: This represents a big neighborhood inside which there are many mini neighborhoods which is a categorical variable. There are 5 neighborhood groups in the data:

- i) Manhattan
- ii) Brooklyn
- iii) Staten Island
- iv) Queens
- v) Bronx

4.Count for neighbour groups (EDA)

From this distribution plotting, it implies that Manhattan has the maximum number of airbnbs in the whole of NYC.

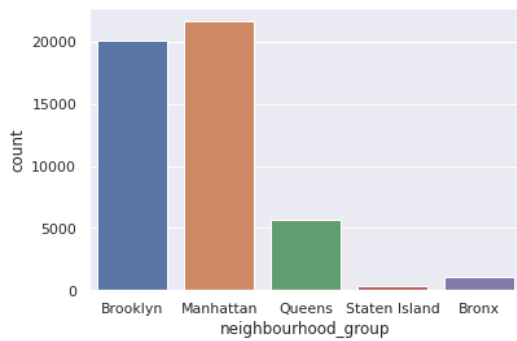


Fig 1: Rating Distribution

5. Rental Distribution vs Borough

As it's reflected in the visualization, Brooklyn and Manhattan boroughs concentrate the majority of the listed rentals on Airbnb, adding up more than 40,000 rentals between the two of them. This means that the bulk of visitors of New York stay in properties, rooms or residencies located in these areas.

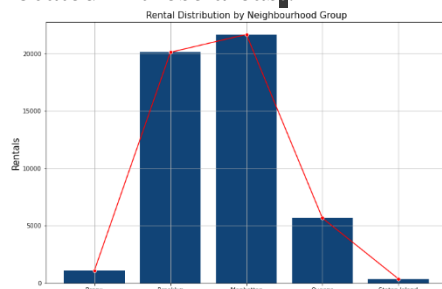


Fig 2: Rental Distribution

6. Type Rental

From the whole rentals available in the dataset, 52% of them correspond to entire-home apartments, 46% to private-room rentals and the minority remaining corresponds to shared rooms with a 2% of the sample.

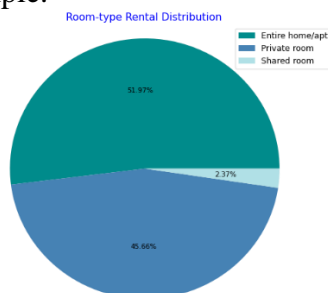
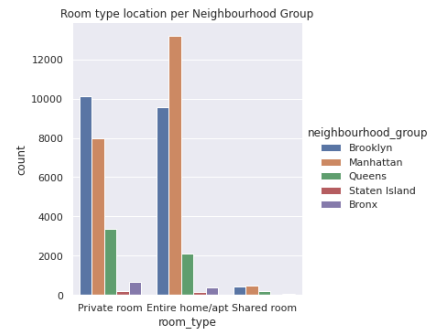


Fig 3: Type Rental

7. Room type location per Neighbourhood Group

Here we conduct the graph which show type

Room type location per Neighbourhood Group



For Manhattan the count of entire home/apt is maximum while for Brooklyn private room count is maximum.

8. Conclusion

We have reached the end of our analysis of Airbnb listings in NYC. We started from looking out for duplicate values, then missing value treatment and finally used EDA to discover many insights from the dataset. To summarize few of the important insights we gathered:

- Host Maya is the busiest host in NYC and there are multiple reasons in favor of it like price, minimum nights, availability and number of reviews. She has a total of 5 properties listed in the same neighborhood.
- Manhattan and Brooklyn are the most expensive neighborhoods and they receive the most traffic as well. Due to many tourist attractions and the number of properties available, people tend to visit these two areas comparatively more than other ones.
- Entire Home/Apt is the costliest room type available but still the most

preferred ones for the customers.
Entire Home/Apt and Private
Rooms receive way more traffic
than Shared Rooms and as a
result Shared Rooms stay
available for most of the time out
of the 365 days.

These insights generated can definitely
help everyone make better decisions in
future to enhance their experience of
staying in an Airbnb in NYC.

9.References

1. Tutorialspoint
2. Analytics Vidhya
3. Stack Overflow