AirBnB NYC Case Study – PPT 1

-BY AVLEEN SINGH & NAND DAVE

Importing the necessary libraries and the dataset provided

In [3]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
```

In [4]:
```
df = pd.read_csv("AB_NYC_2019.csv")
df.head()
```

Out[4]:

	id	name	host_id	host_name	neighbourhood_group	neighbourhood	latitude	longitude	room_type	price	minimum_nights	number_of_revi
0	2539	Clean & quiet apt home by the park	2787	John	Brooklyn	Kensington	40.64749	-73.97237	Private room	149	1	
1	2595	Skylit Midtown Castle	2845	Jennifer	Manhattan	Midtown	40.75362	-73.98377	Entire home/apt	225	1	
2	3647	THE VILLAGE OF HARLEM....NEW YORK !	4632	Elisabeth	Manhattan	Harlem	40.80902	-73.94190	Private room	150	3	
3	3831	Cozy Entire Floor of Brownstone	4869	LisaRoxanne	Brooklyn	Clinton Hill	40.68514	-73.95976	Entire home/apt	89	1	
4	5022	Entire Apt: Spacious Studio/Loft by central park	7192	Laura	Manhattan	East Harlem	40.79851	-73.94399	Entire home/apt	80	10	

Information about the dataset

Cleaning of data

Cleaning of Data

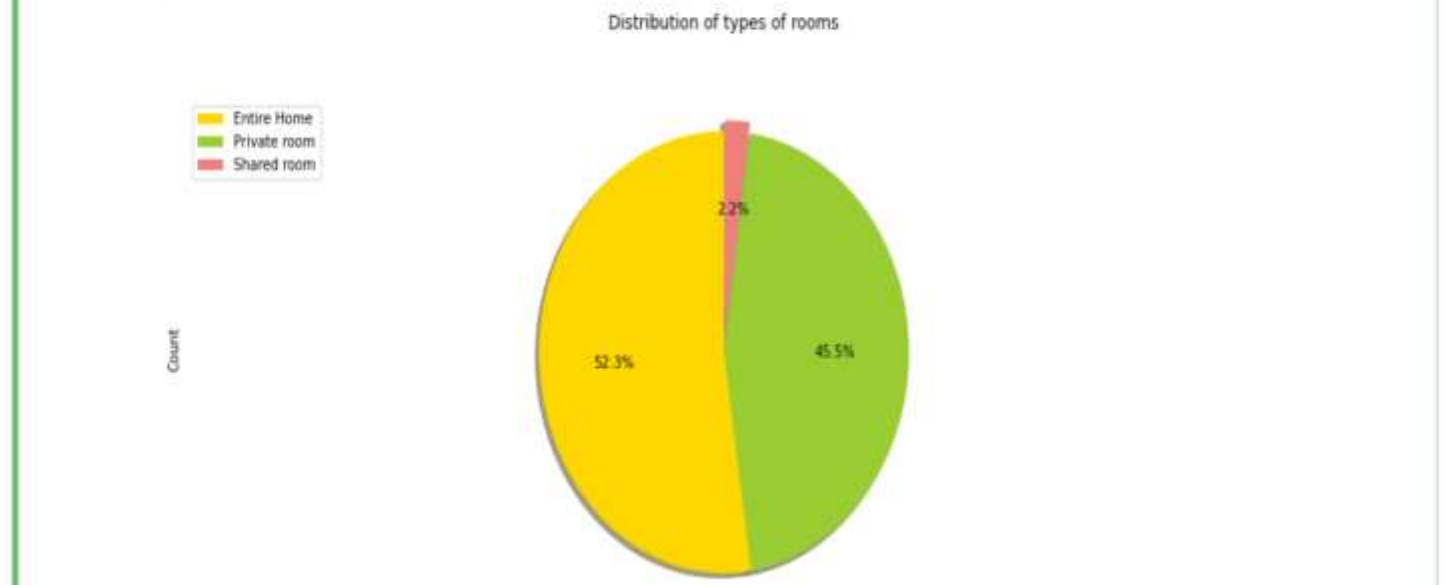
```
In [9]: df.dropna(subset=['last_review','reviews_per_month'],axis=0,inplace=True)
        df.count()
```

```
Out[9]: id                                38843
        name                              38837
        host_id                           38843
        host_name                         38827
        neighbourhood_group               38843
        neighbourhood                     38843
        latitude                          38843
        longitude                         38843
        room_type                         38843
        price                             38843
        minimum_nights                    38843
        number_of_reviews                 38843
        last_review                       38843
        reviews_per_month                 38843
        calculated_host_listings_count    38843
        availability_365                  38843
        dtype: int64
```

- Dropped the missing values from the columns “last_review” and “reviews_per_month” which were not necessary.

Pie-Chart visualization of different types of rooms

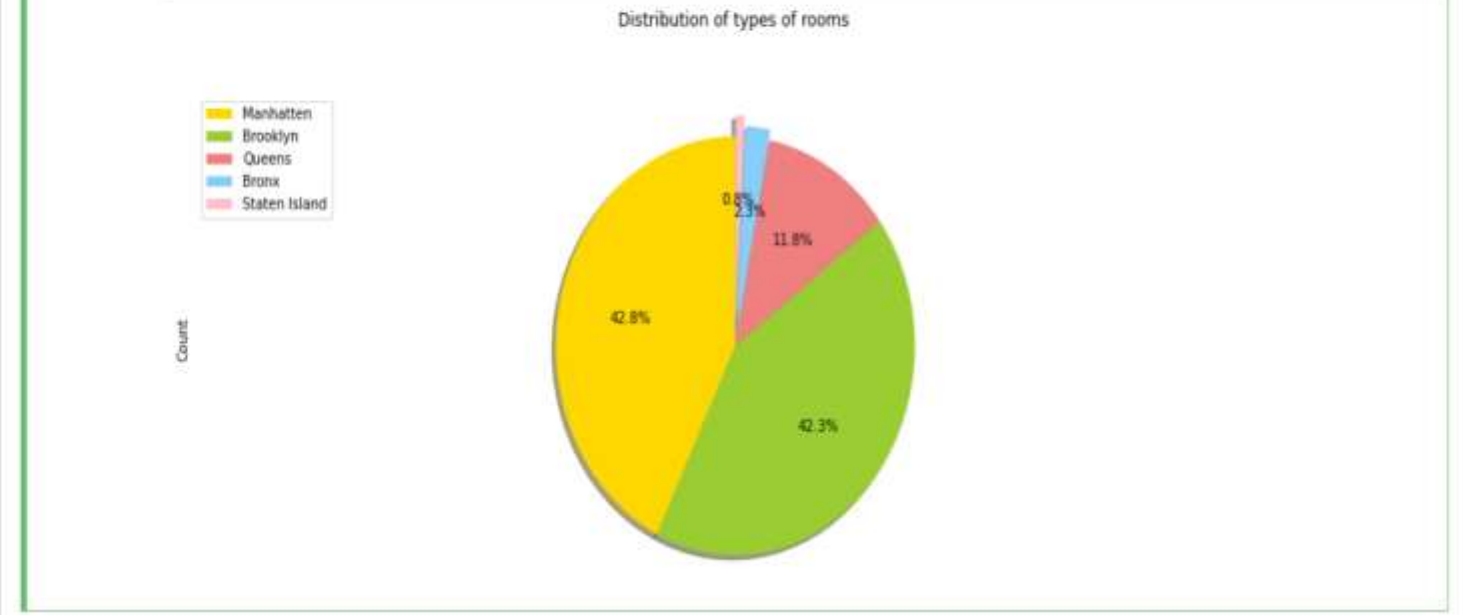
```
In [11]: df_type_of_rooms['Count'].plot(kind='pie', figsize=(15,6), autopct="%1.1f%%", startangle=90, shadow=True, labels=None,
                                        colors=['gold','yellowgreen','lightcoral'], explode=[0,0,0.05])
         plt.title('Distribution of types of rooms',y=1.12)
         plt.axis('equal')
         plt.legend(labels=['Entire Home','Private room','Shared room'],loc='upper left')
         plt.show()
```



- We found out that Most of the people offer either a entire home or a private room

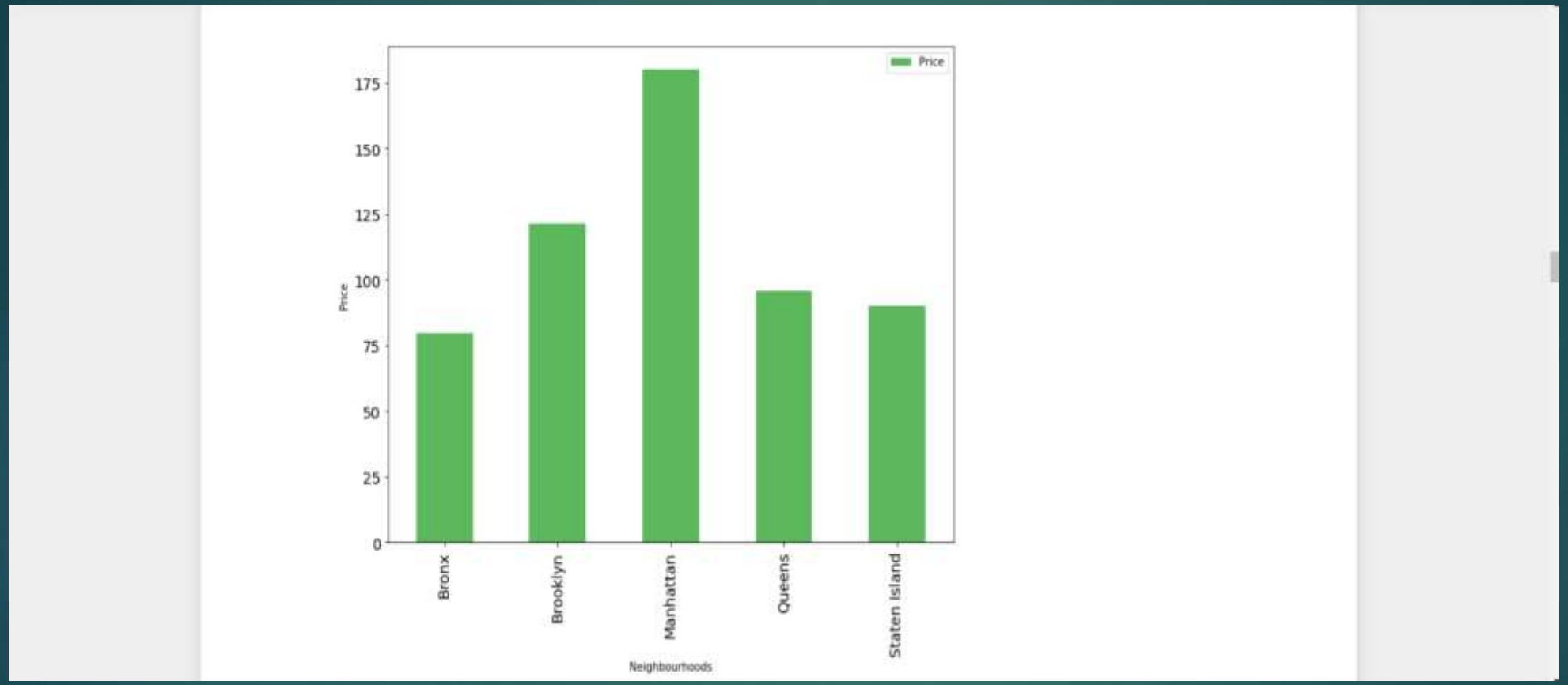
Pie chart visualization on distribution of types of rooms based on neighborhood

```
In [13]: df_neighbourhood_group['Count'].plot(kind='pie', figsize=(15,6), autopct="%1.1f%%", startangle=90, shadow=True, labels=None,
                                              colors=['gold','yellowgreen','lightcoral','lightskyblue','pink'], explode=[0,0,0,0.05,0.1])
         plt.title('Distribution of types of rooms',y=1.12)
         plt.axis('equal')
         plt.legend(labels=['Manhatten','Brooklyn','Queens','Bronx','Staten Island'],loc='upper left')
         plt.show()
```



- We found out that more than half of the houses are based in either Manhatten or Brooklyn

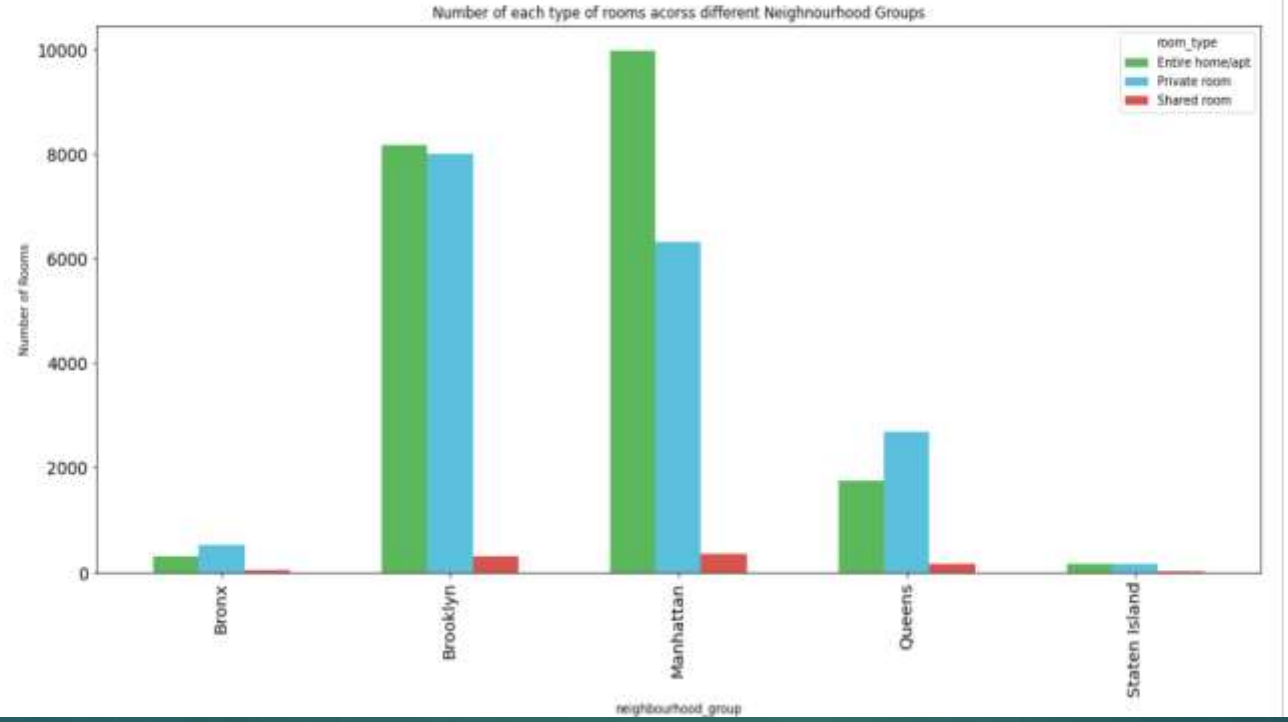
Average Price of Rooms across different neighborhood groups



- We found out that Manhattan has the highest average price of 175$

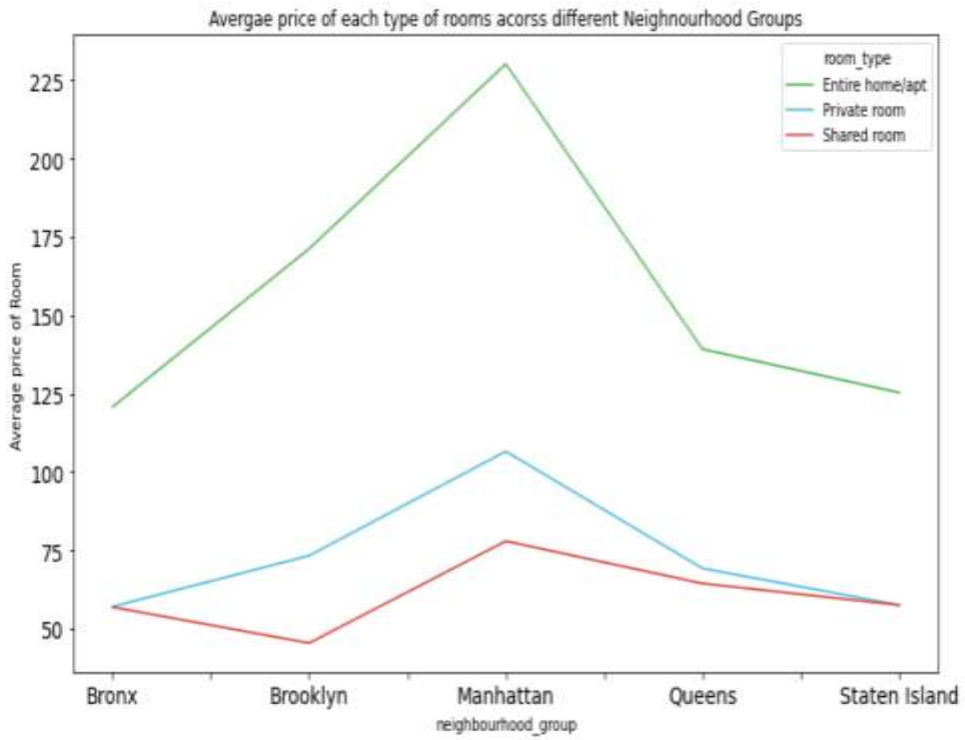
Number of each type of rooms across different Neighborhood Groups

```
In [17]: df_neighbourhood_1.plot(kind='bar', figsize=(18,8), width=0.6, fontsize=14, color = ['#5cb85c','#5bc0de','#d9534f'])
         plt.title('Number of each type of rooms acorss different Neighnourhood Groups')
         plt.ylabel('Number of Rooms')
         plt.show()
```



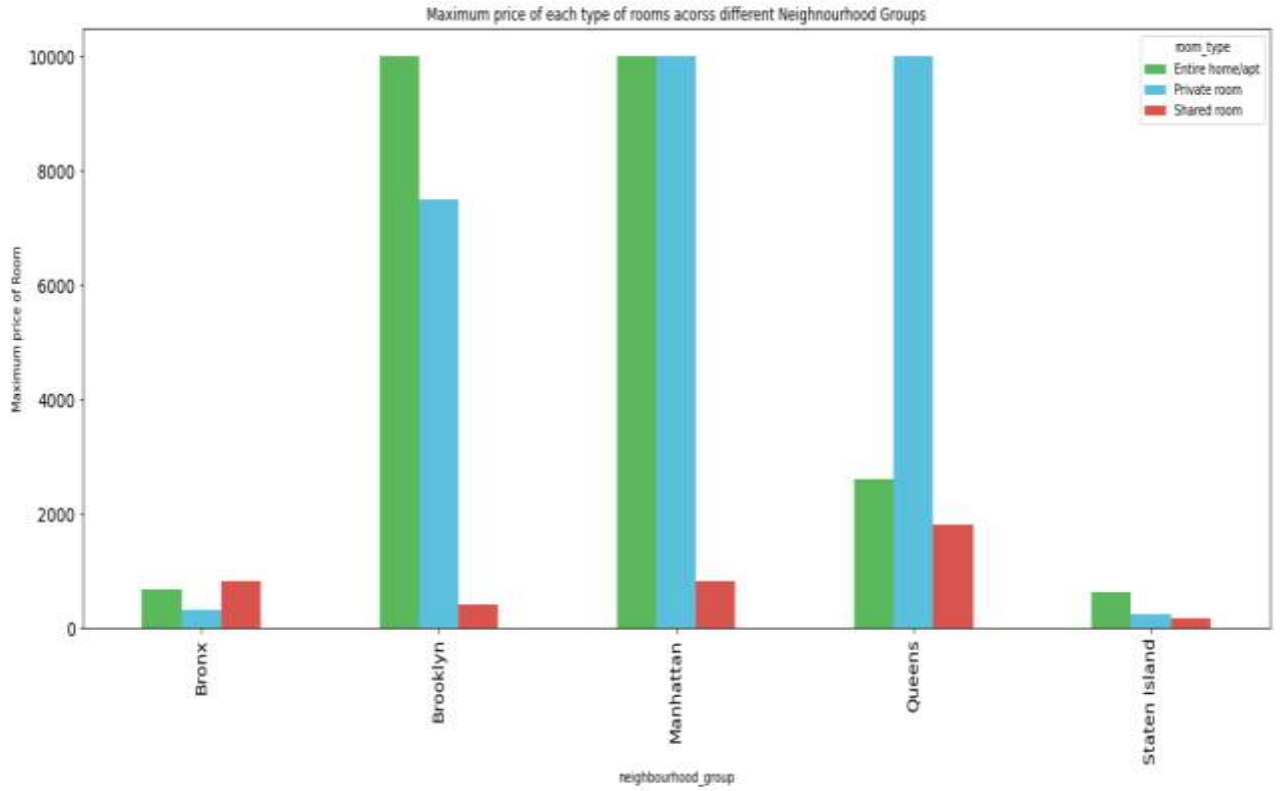
Average price of each type of rooms across different Neighborhood Groups

```
In [19]: df_neighbourhood_2.plot(kind='line', figsize=(12,7), fontsize=14, color = ['#5cb85c','#5bc0de','#d9534f'])
         plt.title('Avergae price of each type of rooms acorss different Neighnourhood Groups')
         plt.ylabel('Average price of Room')
         plt.show()
```



Maximum price of each type of rooms across different Neighborhood Groups

```
In [21]: df_neighbourhood_3.plot(kind='bar', figsize=(20,8), fontsize=14, color = ['#5cb85c','#5bc0de','#d9534f'])
         plt.title('Maximum price of each type of rooms acorss different Neighnourhood Groups')
         plt.ylabel('Maximum price of Room')
         plt.show()
```



Conclusion

- There are 5 different neighbourhood groups present in our database: 1.Manhattan 2.Bronx 3.Brooklyn 4.Queens 5.Staten Island
- The one that shows up the most is Manhattan with over 16 thousand entries.
- There are three room types listed: 1.Private room 2.Shared room 3. Entire Home/apt.
- We plotted a pie chart for visualising different types of rooms and we found out that most of the people offered either an Entire home/apt or a private room. We plotted a pie chart for distribution of different types of rooms and we found out that more than half of the houses are based in either Manhatten or Brooklyn.
- Then checked the average price of Rooms across different neighbourhood groups and found out that Manhattan has the highest average price of 175$. Then we found out the maximum price of each type of rooms acorss different Neighnourhood Groups and Manhattan leads with room_type Entire home/apt and private room.
- Then we checked for the neighborhood groups that are available and which shows up the most and found out that , there were 5 neighborhood groups available and Manhattan shows up the most, with 16632. If we analyze the same data based on the average number of reviews per month between all listings of a certain host, Sonder has a total of 397.56 reviews per month followed by Row NYC with 111.72.
- Looking at the price by neighbourhood group: The highest average price for a room in NYC is in Manhattan with a little over 196 $/night. The highest total price for listings in NYC are in Manhattan almost double the total price in Brooklyn
- Based on the total price of all the listings we have a TOP 5 hosts in NYC: Sonder is leading with a total of $55920 followed by Blueground, Sally, Red Awning, and Kara.
- We analyzed the entries with zero available days in 2019 and more than 100 reviews for the year and found out that there are 162 entries that satisfy these conditions.
- By summing the total prices per host we found that "Sonder NYC" has the highest total of prices (55920) and is located in Manhattan. Also, Sonder is the leader in the total calculated listings count with 207 entries.
- We looked into a specific case of Danielle from Queens and her listings sorted by the date of the last review. We found that there are 4 different Danielles in Queens. However, the latest review for a Danielle in our database was for Danielle with the host_id 26432133 and it was posted on 2019-07-08. Danielle from Queend/Long Island City got her latest review on 2019-06-20, Danielle from Queens/Astoria got her latest review on 2018-01-02 and the last Danielle in our database did not get any reviews.
- Finally, we looked into how many entries had a year round availability and found that 841 properties were available for 365 days in 2019, also we found that by the total number of reviews left on the platform, the "Entire home/apt" option is leading by far in this category.

Appendix

- Importing the necessary libraries and the dataset provided
- Cleaning of data
- Pie-Chart visualization of different types of rooms
- Pie chart visualization on distribution of types of rooms based on neighborhood
- Average Price of Rooms across different neighborhood groups
- Number of each type of rooms across different Neighborhood Groups
- Average price of each type of rooms across different Neighborhood Groups
- Maximum price of each type of rooms across different Neighborhood Groups
- Conclusion