Airbnb_Analysis

July 20, 2024

1 Airbnb Bookings Analysis

2 Project Summary -

- **The purpose of the analysis:** understanding the factors that influence Airbnb prices in New York City, or identifying patterns of all variables and Our analysis provides useful information for travelers and hosts in the city and also provides some best insights for Airbnb business.
- This project involved exploring and cleaning a dataset to prepare it for analysis. The data exploration process involved identifying and understanding the characteristics of the data, such as the data types, missing values, and distributions of values. The data cleaning process involved identifying and addressing any issues or inconsistencies in the data, such as errors, missing values, or duplicate records and remove outliers.
- Through this process, we were able to identify and fix any issues with the data, and ensure that it was ready for further analysis. This is an important step in any data analysis project, as it allows us to work with high-quality data and avoid any potential biases or errors that could affect the results. The clean and prepared data can now be used to answer specific research.
- Once the data has been cleaned and prepared, now begin exploring and summarizing it with describe the data and creating visualizations, and identifying patterns and trends in the data. in explore the data, may develop the relationships between different variables or the underlying causes of certain patterns or trends and other methods.
- using data visualization to explore and understand patterns in Airbnb data. We created various graphs and charts to visualize the data, and wrote observations and insights below each one to help us better understand the data and identify useful insights and patterns.
- Through this process, we were able to uncover trends and relationships in the data that would have been difficult to identify through raw data alone, for example factors affecting prices and availability. We found that minimum nights, number of reviews, and host listing count are important for determining prices, and that availability varies significantly across neighborhoods. Our analysis provides useful information for travelers and hosts in the city.
- The observations and insights we identified through this process will be useful for future analysis and decision-making related to Airbnb. and also Our analysis provides useful information for travelers and hosts in the city.

3 Problem Statements -

1. What are the most popular neighborhoods for Airbnb rentals in New York City? How do prices and availability vary by neighborhood?
2. How has the Airbnb market in New York City changed over time? Have there been any significant trends in terms of the number of listings, prices, or occupancy rates?
3. Are there any patterns or trends in terms of the types of properties that are being rented out on Airbnb in New York City? Are certain types of properties more popular or more expensive than others?
4. Are there any factors that seem to be correlated with the prices of Airbnb rentals in New York City?
5. the best area in New York City for a host to buy property at a good price rate and in an area with high traffic ?
6. How do the lengths of stay for Airbnb rentals in New York City vary by neighborhood? Do certain neighborhoods tend to attract longer or shorter stays?
7. How do the ratings of Airbnb rentals in New York City compare to their prices? Are higher-priced rentals more likely to have higher ratings?
8. Find the total numbers of Reviews and Maximum Reviews by Each Neighborhood Group.
9. Find Most reviewed room type in Neighborhood groups per month.
10. Find Best location listing/property location for travelers.
11. Find also best location listing/property location for Hosts.
12. Find Price variations in NYC Neighborhood groups.

there is a lot of problem statements and we have to finds information and insights through different different problem statements so now lets start…

3.0.1 Importing the necessary libraries

[1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt      #for visualization
%matplotlib inline
import seaborn as sns                #for visualization
import warnings
warnings.filterwarnings('ignore')
```

3.0.2 Load Airbnb Dataset

[3]:
```
Airbnb_df = pd.read_csv('C:/Users/KIIT/Downloads/Airbnb NYC 2019.csv')
Airbnb_df
```

[3]:
```
             id                                              name   host_id  \
0          2539                Clean & quiet apt home by the park      2787
1          2595                             Skylit Midtown Castle      2845
2          3647               THE VILLAGE OF HARLEM…NEW YORK !      4632
3          3831                   Cozy Entire Floor of Brownstone      4869
4          5022  Entire Apt: Spacious Studio/Loft by central park      7192
…           …                                                …        …
48890  36484665   Charming one bedroom - newly renovated rowhouse   8232441
48891  36485057     Affordable room in Bushwick/East Williamsburg   6570630
48892  36485431           Sunny Studio at Historical Neighborhood  23492952
48893  36485609              43rd St. Time Square-cozy single bed  30985759
48894  36487245  Trendy duplex in the very heart of Hell's Kitchen  68119814

           host_name neighbourhood_group       neighbourhood  latitude  \
0               John            Brooklyn          Kensington  40.64749
1           Jennifer           Manhattan             Midtown  40.75362
2          Elisabeth           Manhattan              Harlem  40.80902
3        LisaRoxanne            Brooklyn        Clinton Hill  40.68514
4              Laura           Manhattan         East Harlem  40.79851
…                …                  …                  …        …
48890        Sabrina            Brooklyn  Bedford-Stuyvesant  40.67853
48891        Marisol            Brooklyn            Bushwick  40.70184
48892  Ilgar & Aysel           Manhattan              Harlem  40.81475
48893            Taz           Manhattan      Hell's Kitchen  40.75751
48894     Christophe           Manhattan      Hell's Kitchen  40.76404

       longitude        room_type  price  minimum_nights  number_of_reviews  \
0      -73.97237     Private room    149               1                  9
1      -73.98377  Entire home/apt    225               1                 45
2      -73.94190     Private room    150               3                  0
3      -73.95976  Entire home/apt     89               1                270
4      -73.94399  Entire home/apt     80              10                  9
…            …               …    …             …                 …
48890  -73.94995     Private room     70               2                  0
48891  -73.93317     Private room     40               4                  0
48892  -73.94867  Entire home/apt    115              10                  0
48893  -73.99112      Shared room     55               1                  0
48894  -73.98933     Private room     90               7                  0

      last_review  reviews_per_month  calculated_host_listings_count  \
0      2018-10-19               0.21                               6
1      2019-05-21               0.38                               2
2             NaN                NaN                               1
3      2019-07-05               4.64                               1
4      2018-11-19               0.10                               1
…            …                 …                             …
48890         NaN                NaN                               2
```

```
48891         NaN                NaN                                  2
48892         NaN                NaN                                  1
48893         NaN                NaN                                  6
48894         NaN                NaN                                  1

       availability_365
0                   365
1                   355
2                   365
3                   194
4                     0
…                  …
48890                 9
48891                36
48892                27
48893                 2
48894                23

[48895 rows x 16 columns]
```

4 About the Dataset – Airbnb Bookings

- This Airbnb dataset contains nearly 49,000 observations from New York , with 16 columns of data.
- The Data includes both categorical and numeric values, providing a diverse range of information about the listings.
- This Dataset may be useful for analyzing trends and patterns in the Airbnb market in New York and also gain insights into the preferences and behavior of Airbnb users in the area.
- This dataset contains information about Airbnb bookings in New York City in 2019. By analyzing this data, you may be able to understand the trends and patterns of Airbnb use in the NYC.

5 UNDERSTAND THE GIVEN VARIABLES

Listing_id :- This is a unique identifier for each listing in the dataset.

Listing_name :- This is the name or title of the listing, as it appears on the Airbnb website.

Host_id :- This is a unique identifier for each host in the dataset.

Host_name :- This is the name of the host as it appears on the Airbnb website.

Neighbourhood_group :- This is a grouping of neighborhoods in New York City, such as Manhattan or Brooklyn.

Neighbourhood :- This is the specific neighborhood in which the listing is located.

Latitude :- This is the geographic latitude of the listing.

Longitude :- This is the geographic longitude of the listing.

Room_type :- This is the type of room or property being offered, such as an entire home, private room, shared room.

Price :- This is the nightly price for the listing, in US dollars.

Minimum_nights :- This is the minimum number of nights that a guest must stay at the listing.

Total_reviews :- This is the total number of reviews that the listing has received.

Reviews_per_month :- This is the average number of reviews that the listing receives per month.

Host_listings_count :- This is the total number of listings that the host has on Airbnb.

Availability_365 :- This is the number of days in the next 365 days that the listing is available for booking.

6 Data Exploration and Data Cleaning

[4]:
```
Airbnb_df.head().T
```

[4]:
```
                                                              0  \
id                                                         2539
name                               Clean & quiet apt home by the park
host_id                                                    2787
host_name                                                  John
neighbourhood_group                                    Brooklyn
neighbourhood                                        Kensington
latitude                                               40.64749
longitude                                             -73.97237
room_type                                          Private room
price                                                       149
minimum_nights                                                1
number_of_reviews                                             9
last_review                                          2018-10-19
reviews_per_month                                          0.21
calculated_host_listings_count                                6
availability_365                                            365

                                                 1  \
id                                            2595
name                          Skylit Midtown Castle
host_id                                       2845
host_name                                 Jennifer
neighbourhood_group                      Manhattan
neighbourhood                              Midtown
latitude                                  40.75362
longitude                                -73.98377
```

```
room_type                                 Entire home/apt
price                                                 225
minimum_nights                                          1
number_of_reviews                                      45
last_review                                    2019-05-21
reviews_per_month                                    0.38
calculated_host_listings_count                          2
availability_365                                      355

                                                             2  \
id                                                        3647
name                            THE VILLAGE OF HARLEM….NEW YORK !
host_id                                                   4632
host_name                                            Elisabeth
neighbourhood_group                                  Manhattan
neighbourhood                                           Harlem
latitude                                              40.80902
longitude                                             -73.9419
room_type                                         Private room
price                                                      150
minimum_nights                                               3
number_of_reviews                                            0
last_review                                                NaN
reviews_per_month                                          NaN
calculated_host_listings_count                               1
availability_365                                           365

                                                         3  \
id                                                    3831
name                            Cozy Entire Floor of Brownstone
host_id                                               4869
host_name                                      LisaRoxanne
neighbourhood_group                               Brooklyn
neighbourhood                                  Clinton Hill
latitude                                          40.68514
longitude                                        -73.95976
room_type                                  Entire home/apt
price                                                   89
minimum_nights                                           1
number_of_reviews                                      270
last_review                                     2019-07-05
reviews_per_month                                     4.64
calculated_host_listings_count                           1
availability_365                                       194

                                                                              4
id                                                                         5022
```

```
name                                Entire Apt: Spacious Studio/Loft by central park
host_id                                                                          7192
host_name                                                                       Laura
neighbourhood_group                                                         Manhattan
neighbourhood                                                             East Harlem
latitude                                                                     40.79851
longitude                                                                   -73.94399
room_type                                                             Entire home/apt
price                                                                              80
minimum_nights                                                                     10
number_of_reviews                                                                   9
last_review                                                                2018-11-19
reviews_per_month                                                                 0.1
calculated_host_listings_count                                                      1
availability_365                                                                    0
```

[5]:
```python
#checking what are the variables here:
Airbnb_df.columns
```

[5]:
```
Index(['id', 'name', 'host_id', 'host_name', 'neighbourhood_group',
       'neighbourhood', 'latitude', 'longitude', 'room_type', 'price',
       'minimum_nights', 'number_of_reviews', 'last_review',
       'reviews_per_month', 'calculated_host_listings_count',
       'availability_365'],
      dtype='object')
```

- **so now first rename few columns for better understanding of variables -**

[6]:
```python
rename_col = {'id':'listing_id','name':'listing_name','number_of_reviews':
 ↪'total_reviews','calculated_host_listings_count':'host_listings_count'}
```

[7]:
```python
# use a pandas function to rename the current function
Airbnb_df = Airbnb_df.rename(columns = rename_col)
Airbnb_df.head(2)
```

[7]:
```
   listing_id                      listing_name  host_id host_name  \
0        2539  Clean & quiet apt home by the park     2787      John
1        2595             Skylit Midtown Castle     2845  Jennifer

  neighbourhood_group neighbourhood  latitude  longitude        room_type  \
0            Brooklyn    Kensington  40.64749  -73.97237     Private room
1           Manhattan       Midtown  40.75362  -73.98377  Entire home/apt

   price  minimum_nights  total_reviews last_review  reviews_per_month  \
0    149               1              9  2018-10-19               0.21
1    225               1             45  2019-05-21               0.38
```

```
   host_listings_count  availability_365
0                    6               365
1                    2               355
```

[8]:
```
#checking shape of Airbnb dataset
Airbnb_df.shape
```

[8]: (48895, 16)

[9]:
```
#basic information about the dataset
Airbnb_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 48895 entries, 0 to 48894
Data columns (total 16 columns):
 #   Column               Non-Null Count  Dtype
---  ------               --------------  -----
 0   listing_id           48895 non-null  int64
 1   listing_name         48879 non-null  object
 2   host_id              48895 non-null  int64
 3   host_name            48874 non-null  object
 4   neighbourhood_group  48895 non-null  object
 5   neighbourhood        48895 non-null  object
 6   latitude             48895 non-null  float64
 7   longitude            48895 non-null  float64
 8   room_type            48895 non-null  object
 9   price                48895 non-null  int64
 10  minimum_nights       48895 non-null  int64
 11  total_reviews        48895 non-null  int64
 12  last_review          38843 non-null  object
 13  reviews_per_month    38843 non-null  float64
 14  host_listings_count  48895 non-null  int64
 15  availability_365     48895 non-null  int64
dtypes: float64(3), int64(7), object(6)
memory usage: 6.0+ MB
```

So, host_name, neighbourhood_group, neighbourhood and room_type fall into categorical variable category.

While host_id, latitude, longitude, price, minimum_nights, number_of_reviews, last_review, reviews_per_month, host_listings_count, availability_365 are numerical variables

---

[10]:
```
# check duplicate rows in dataset
Airbnb_df = Airbnb_df.drop_duplicates()
Airbnb_df.count()
```

[10]:
```
listing_id              48895
listing_name            48879
host_id                 48895
host_name               48874
neighbourhood_group     48895
neighbourhood           48895
latitude                48895
longitude               48895
room_type               48895
price                   48895
minimum_nights          48895
total_reviews           48895
last_review             38843
reviews_per_month       38843
host_listings_count     48895
availability_365        48895
dtype: int64
```

so, there is no any duplicate rows in Dataset

---

[11]:
```python
# checking null values of each columns
Airbnb_df.isnull().sum()
```

[11]:
```
listing_id                 0
listing_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
total_reviews              0
last_review            10052
reviews_per_month      10052
host_listings_count        0
availability_365           0
dtype: int64
```

host_name and **listing_name** are not that much of null values, so first we are good to fill those with some substitutes in both the columns first.

[12]:
```python
Airbnb_df['listing_name'].fillna('unknown',inplace=True)
Airbnb_df['host_name'].fillna('no_name',inplace=True)
```

[13]:
```
#so the null values are removed
Airbnb_df[['host_name','listing_name']].isnull().sum()
```

[13]:
```
host_name       0
listing_name    0
dtype: int64
```

now, the columns **last_review** and **reviews_per_month** have total 10052 null values each.

last_review column is not required for our analysis as compared to **number_of_reviews** & **reviews_per_month**. We're good to drop this column.

listing_id also not that much of important for our analysis but i dont remove because of **listing_id** and **listing_name** is pair and removing listing_id it still wont make much difference. make sense right ?

[14]:
```
Airbnb_df = Airbnb_df.drop(['last_review'], axis=1)     #removing last_review␣
↪column beacause of not that much important
```

[15]:
```
Airbnb_df.info()      # the last_review column is deleted
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 48895 entries, 0 to 48894
Data columns (total 15 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   listing_id            48895 non-null  int64
 1   listing_name          48895 non-null  object
 2   host_id               48895 non-null  int64
 3   host_name             48895 non-null  object
 4   neighbourhood_group   48895 non-null  object
 5   neighbourhood         48895 non-null  object
 6   latitude              48895 non-null  float64
 7   longitude             48895 non-null  float64
 8   room_type             48895 non-null  object
 9   price                 48895 non-null  int64
 10  minimum_nights        48895 non-null  int64
 11  total_reviews         48895 non-null  int64
 12  reviews_per_month     38843 non-null  float64
 13  host_listings_count   48895 non-null  int64
 14  availability_365      48895 non-null  int64
dtypes: float64(3), int64(7), object(5)
memory usage: 5.6+ MB
```

The **reviews_per_month** column also containing null values and we can simple put 0 reviews by replacing NAN's i think this is make sense -

[16]:
```
Airbnb_df['reviews_per_month'] = Airbnb_df['reviews_per_month'].
↪replace(to_replace=np.nan,value=0).astype('int64')
```

[17]:
```
# the null values are replaced by 0 value
Airbnb_df['reviews_per_month'].isnull().sum()
```

[17]: 0

so there is no null value now in 'reviews_per_month' column because we replaced null value by 0 value. this will make sense because there is no any such data to find those null value

[18]:
```
#so now check Dataset columns changed and null values, last_review column␣
↪removed.
Airbnb_df.sample(5)
```

[18]:
```
       listing_id                                        listing_name  \
41590    32333972                                     Pacheco's House.
18492    14556634                                         Brooklyn Nook
29816    22941772  COZY Bedroom in S. Williamsburg /BedStuy Brooklyn
42433    32933643   Spacious Studio in the UES 94th st (30 days MIN)
37333    29646041  Comfortable place, 15 min from JFK & 30 min to…

         host_id  host_name neighbourhood_group            neighbourhood  \
41590  242758145     Adrian            Brooklyn  Prospect-Lefferts Gardens
18492    9974520    Michael            Brooklyn                 Greenpoint
29816    1466154  Stephanie            Brooklyn         Bedford-Stuyvesant
42433  159598333        Sol           Manhattan            Upper East Side
37333  210339363    Dilenia              Queens                  Woodhaven

       latitude  longitude        room_type  price  minimum_nights  \
41590  40.65447  -73.96160     Private room     60               2
18492  40.72804  -73.94599     Private room     45               2
29816  40.69640  -73.94696     Private room     36              28
42433  40.78331  -73.94646  Entire home/apt     99              30
37333  40.68639  -73.86088     Private room     60               5

       total_reviews  reviews_per_month  host_listings_count  availability_365
41590              6                  1                    1               144
18492              1                  0                    1                 0
29816              1                  0                    1               297
42433              0                  0                    5               332
37333             11                  1                    2                17
```

6.0.1 Check Unique Value for variables and doing some experiments -

[19]:
```
# check unique values for listing/property Ids
# all the listing ids are different and each listings are different here.
Airbnb_df['listing_id'].nunique()
```

[19]: 48895

[20]:
```
# so there are 221 unique neighborhood in Dataset
Airbnb_df['neighbourhood'].nunique()
```

[20]: 221

[21]:
```
#and total 5 unique neighborhood_group in Dataset
Airbnb_df['neighbourhood_group'].nunique()
```

[21]: 5

[22]:
```
#so total 11453 different hosts in Airbnb-NYC
Airbnb_df['host_name'].nunique()
```

[22]: 11453

[23]:
```
# most of the listing/property are different in Dataset
Airbnb_df['listing_name'].nunique()
```

[23]: 47906

Note - so i think few listings/property with same names has different hosts in different areas/neighbourhoods of a neighbourhood_group

[24]:
```
Airbnb_df[Airbnb_df['host_name']=='David']['listing_name'].nunique()

# so here same host David operates different 402 listing/property
```

[24]: 402

[ ]:
```
Airbnb_df[Airbnb_df['listing_name']==Airbnb_df['host_name']].head()

# there are few listings where the listing/property name and the host have same␣
↪names
```

[ ]:
```
       listing_id    listing_name    host_id       host_name  \
9473      7264659         Olivier    6994503         Olivier
10682     8212051           Monty   43302952           Monty
16422    13186374            Sean   35143476            Sean
23996    19348168             Cyn   74033595             Cyn
24152    19456810  Hillside Hotel  134184451  Hillside Hotel

      neighbourhood_group       neighbourhood  latitude  longitude  \
9473            Manhattan     Upper West Side  40.78931  -73.97520
10682            Brooklyn       East Flatbush  40.66383  -73.92706
16422            Brooklyn     Windsor Terrace  40.65182  -73.98043
23996            Brooklyn  Bedford-Stuyvesant  40.67850  -73.91478
24152              Queens           Briarwood  40.70454  -73.81549
```

```
            room_type  price  minimum_nights  total_reviews  \
9473   Entire home/apt    200               5             12
10682      Shared room     95               2              7
16422  Entire home/apt    400               7              0
23996     Private room     75               2              1
24152     Private room     93               1              2

       reviews_per_month  host_listings_count  availability_365
9473                   0                    1                25
10682                  0                    1               238
16422                  0                    1                 0
23996                  0                    1                 0
24152                  0                   18                90
```

[ ]:
```
Airbnb_df.loc[(Airbnb_df['neighbourhood_group']=='Queens') &␣
 ↪(Airbnb_df['host_name']=='Alex')].head(4)

# Same host have hosted different listing/property in different or same␣
 ↪neighbourhood in same neighbourhood groups
# like Alex hosted different listings in most of different neighbourhood and␣
 ↪there are same also in queens neighbourhood_group!
```

[ ]:
```
       listing_id                       listing_name   host_id host_name  \
3523      2104910  SPACIOUS APT BK/QUEENS w/BACKYARD!  10643810      Alex
4512      3116519   Large 900 sqft Artist's Apartment   3008690      Alex
6178      4518242             Zen MiniPalace Astoria  23424461      Alex
10543     8090529        Modern studio in Queens, NY  17377835      Alex

      neighbourhood_group neighbourhood  latitude  longitude        room_type  \
3523               Queens     Ridgewood  40.70988  -73.90845  Entire home/apt
4512               Queens     Ridgewood  40.70124  -73.90941  Entire home/apt
6178               Queens       Astoria  40.76369  -73.91601  Entire home/apt
10543              Queens     Sunnyside  40.74674  -73.91881  Entire home/apt

       price  minimum_nights  total_reviews  reviews_per_month  \
3523      99               2             57                  0
4512      70              10              0                  0
6178      80               1              3                  0
10543    250               3              0                  0

       host_listings_count  availability_365
3523                     1                42
4512                     1                 0
6178                     1                 0
10543                    1               364
```

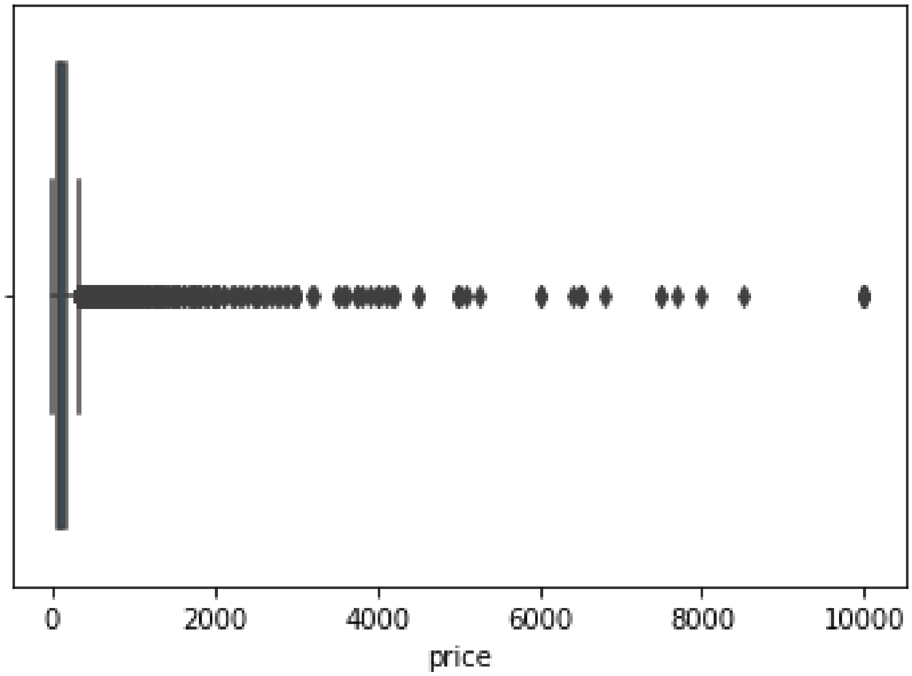
7 Describe the Dataset and removing outliers

[ ]:
```
# describe the DataFrame
Airbnb_df.describe()
```

[ ]:
```
         listing_id       host_id      latitude     longitude         price  \
count  4.889500e+04  4.889500e+04  48895.000000  48895.000000  48895.000000
mean   1.901714e+07  6.762001e+07     40.728949    -73.952170    152.720687
std    1.098311e+07  7.861097e+07      0.054530      0.046157    240.154170
min    2.539000e+03  2.438000e+03     40.499790    -74.244420      0.000000
25%    9.471945e+06  7.822033e+06     40.690100    -73.983070     69.000000
50%    1.967728e+07  3.079382e+07     40.723070    -73.955680    106.000000
75%    2.915218e+07  1.074344e+08     40.763115    -73.936275    175.000000
max    3.648724e+07  2.743213e+08     40.913060    -73.712990  10000.000000

       minimum_nights  total_reviews  reviews_per_month  host_listings_count  \
count    48895.000000   48895.000000       48895.000000         48895.000000
mean         7.029962      23.274466           0.806258             7.143982
std         20.510550      44.550582           1.502767            32.952519
min          1.000000       0.000000           0.000000             1.000000
25%          1.000000       1.000000           0.000000             1.000000
50%          3.000000       5.000000           0.000000             1.000000
75%          5.000000      24.000000           1.000000             2.000000
max       1250.000000     629.000000          58.000000           327.000000

       availability_365
count      48895.000000
mean         112.781327
std          131.622289
min            0.000000
25%            0.000000
50%           45.000000
75%          227.000000
max          365.000000
```

Note - price column is very important so we have to find big outliers in important columns first.

[ ]:
```
sns.boxplot(x = Airbnb_df['price'])

plt.show()
```



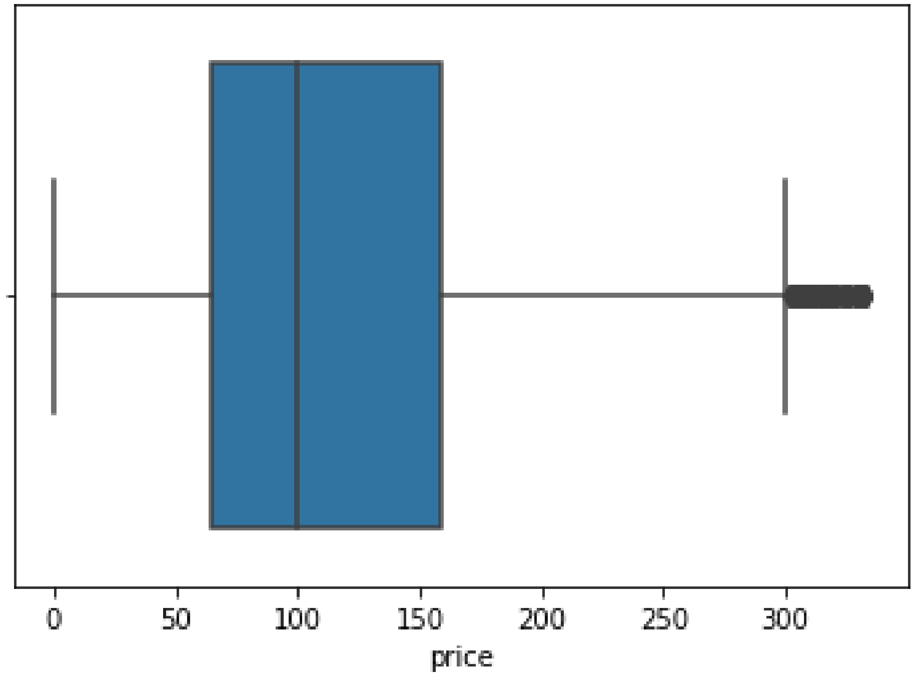
7.0.1 using IQR technique

[ ]:
```
# writing a outlier function for removing outliers in important columns.
def iqr_technique(DFcolumn):
  Q1 = np.percentile(DFcolumn, 25)
  Q3 = np.percentile(DFcolumn, 75)
  IQR = Q3 - Q1
  lower_range = Q1 - (1.5 * IQR)
  upper_range = Q3 + (1.5 * IQR)                        # interquantile range

  return lower_range,upper_range
```

[ ]:
```
lower_bound,upper_bound = iqr_technique(Airbnb_df['price'])

Airbnb_df = Airbnb_df[(Airbnb_df.price>lower_bound) & (Airbnb_df.
 ↪price<upper_bound)]
```

[ ]:
```
# so the outliers are removed from price column now check with boxplot and also␣
 ↪check shape of new Dataframe!

sns.boxplot(x = Airbnb_df['price'])
print(Airbnb_df.shape)
```

(45918, 15)




```
[ ]: # so here outliers are removed, see the new max price
print(Airbnb_df['price'].max())
```

333

---

8 Data Visualization

(1) Distribution Of Airbnb Bookings Price Range Using Histogram

```
[ ]: # Create a figure with a custom size
plt.figure(figsize=(12, 5))

# Set the seaborn theme to darkgrid
sns.set_theme(style='darkgrid')

# Create a histogram of the 'price' column of the Airbnb_df dataframe
# using sns distplot function and specifying the color as red
sns.distplot(Airbnb_df['price'],color=('r'))

# Add labels to the x-axis and y-axis
plt.xlabel('Price', fontsize=14)
```

```
plt.ylabel('Density', fontsize=14)

# Add a title to the plot
plt.title('Distribution of Airbnb Prices',fontsize=15)
```

[ ]: Text(0.5, 1.0, 'Distribution of Airbnb Prices')

observations –>

- The range of prices being charged on Airbnb appears to be from **20 to 330 dollars** , with the majority of listings falling in the price range of **50 to 150 dollars.**
- The distribution of prices appears to have a peak in the **50 to 150 dollars range**, with a relatively lower density of listings in higher and lower price ranges.
- There may be fewer listings available at prices above **250 dollars**, as the density of listings drops significantly in this range.

---

(2) Total Listing/Property count in Each Neighborhood Group using Count plot

```
[ ]: # Count the number of listings in each neighborhood group and store the result␣
  ↪in a Pandas series
counts = Airbnb_df['neighbourhood_group'].value_counts()

# Reset the index of the series so that the neighborhood groups become columns␣
  ↪in the resulting dataframe
Top_Neighborhood_group = counts.reset_index()

# Rename the columns of the dataframe to be more descriptive
```

```
Top_Neighborhood_group.columns = ['Neighborhood_Groups', 'Listing_Counts']

# display the resulting DataFrame
Top_Neighborhood_group
```

[ ]:

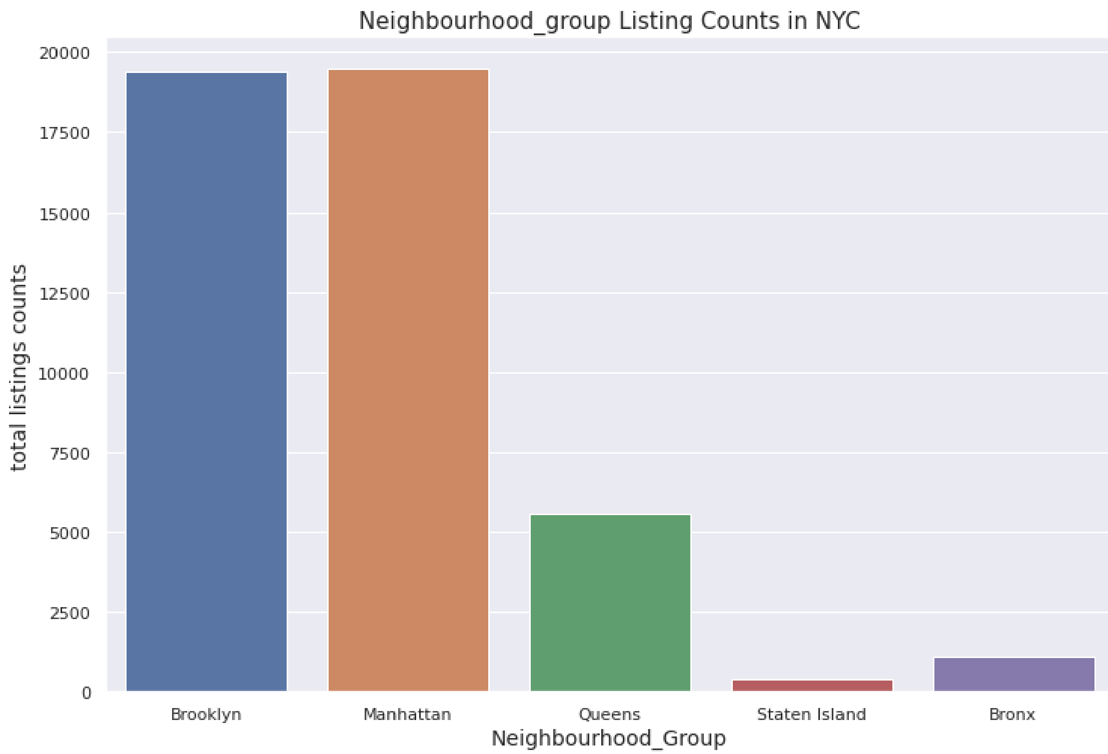
```
  Neighborhood_Groups  Listing_Counts
0           Manhattan           19501
1            Brooklyn           19415
2              Queens            5567
3               Bronx            1070
4       Staten Island             365
```

[ ]:

```
# Set the figure size
plt.figure(figsize=(12, 8))

# Create a countplot of the neighbourhood group data
sns.countplot(Airbnb_df['neighbourhood_group'])

# Set the title of the plot
plt.title('Neighbourhood_group Listing Counts in NYC', fontsize=15)

# Set the x-axis label
plt.xlabel('Neighbourhood_Group', fontsize=14)

# Set the y-axis label
plt.ylabel('total listings counts', fontsize=14)
```

[ ]:

```
Text(0, 0.5, 'total listings counts')
```



Observations –>

- Manhattan and Brooklyn have the highest number of listings on Airbnb, with over 19,000 listings each.
- Queens and the Bronx have significantly fewer listings compared to Manhattan and Brooklyn, with 5,567 and 1,070 listings, respectively
- Staten Island has the fewest number of listings, with only 365.
- The distribution of listings across the different neighborhood groups is skewed, with a concentration of listings in Manhattan and Brooklyn.
- Despite being larger in size, the neighborhoods in Queens, the Bronx, and Staten Island have fewer listings on Airbnb compared to Manhattan, which has a smaller geographical area.
- This could suggest that the demand for Airbnb rentals is higher in Manhattan compared to the other neighborhoods, leading to a higher concentration of listings in this area.
- Alternatively, it could be that the supply of listings is higher in Manhattan due to a higher number of homeowners or property owners in this neighborhood who are willing to list their properties on Airbnb.

---

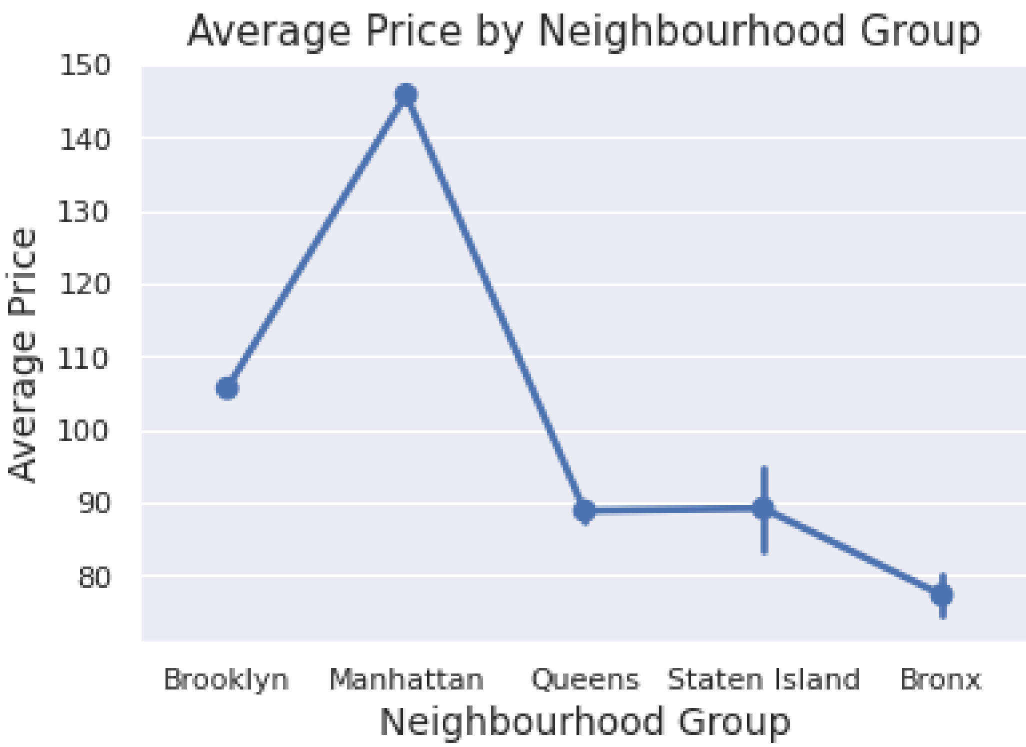
(3) Average Price Of Each Neighborhood Group using Point Plot

```
[ ]: # Group the Airbnb dataset by neighborhood group and calculate the mean of each␣
      ↪group
     grouped = Airbnb_df.groupby("neighbourhood_group").mean()

     # Reset the index of the grouped dataframe so that the neighborhood group␣
      ↪becomes a column
     neighbourhood_group_avg_price = grouped.reset_index()

     # Rename the "price" column to "avg_price"
     neighbourhood_group_avg_price = round(neighbourhood_group_avg_price.
      ↪rename(columns={"price": "avg_price"}),2)

     # Select only the "neighbourhood_group" and "avg_price" columns
     neighbourhood_group_avg_price[['neighbourhood_group', 'avg_price']].head()
```

```
[ ]:   neighbourhood_group  avg_price
     0               Bronx      77.37
     1            Brooklyn     105.70
     2           Manhattan     145.90
     3              Queens      88.90
     4       Staten Island      89.24
```

```
[ ]: #import mean function from the statistics module
     from statistics import mean

     # Create the point plot
     sns.pointplot(x = 'neighbourhood_group', y='price', data=Airbnb_df, estimator =␣
      ↪np.mean)

     # Add axis labels and a title
     plt.xlabel('Neighbourhood Group',fontsize=14)
     plt.ylabel('Average Price',fontsize=14)
     plt.title('Average Price by Neighbourhood Group',fontsize=15)
```

```
[ ]: Text(0.5, 1.0, 'Average Price by Neighbourhood Group')
```



Observations –>

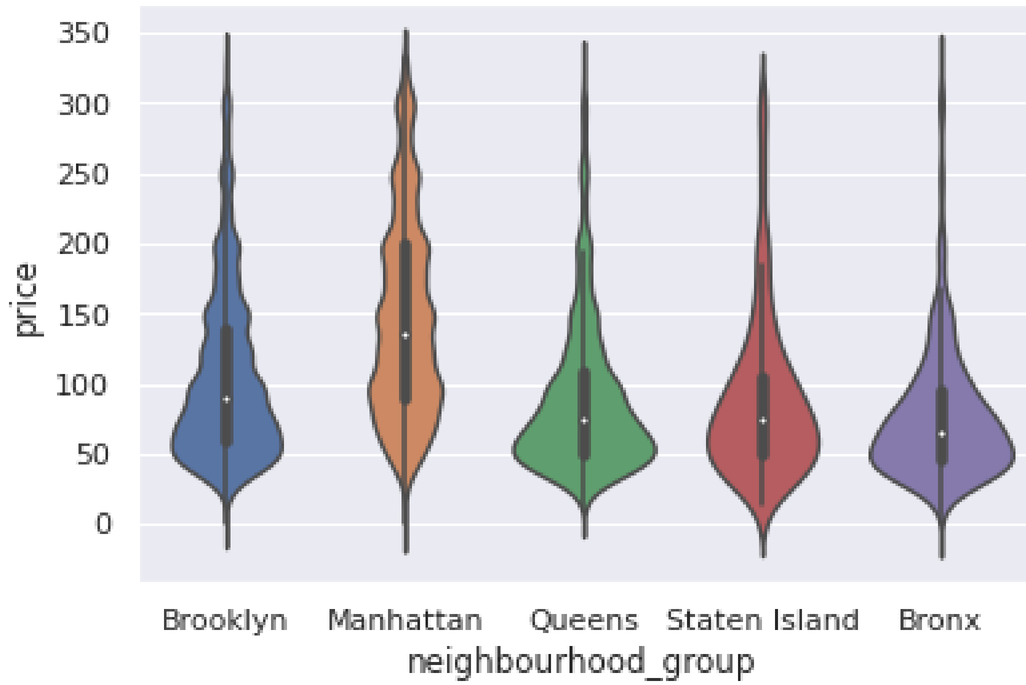
- The average price of a listing in New York City varies significantly across different neighborhoods, with **Manhattan having the highest 146 dollars/day average price** and **the Bronx having the lowest near 77 dollars/day.**
- In second graph price distribution is very high in Manhattan and Brooklyn. but Manhattan have more varity in price range, you can see in second violinplot.
- The average price increases as you move from the outer boroughs (Bronx, Brooklyn, Queens, and Staten Island) towards the center of the city (Manhattan).
- The average price in queens and Staten Island is relatively similar, despite being in different parts of the city.
- The data suggests that the overall cost of living in New York City is higher in the center of the city (Manhattan) compared to the outer boroughs. This is likely due to the fact that Manhattan is the most densely populated and commercially important borough, and therefore has higher demand for housing in the centrally located neighborhoods

---

(4) Price Distribution Of Each Neighborhood Group using Violin Plot

[ ]:
```
# Create the violin plot for price distribution in each Neighbourhood_groups
```

```
ax= sns.violinplot(x='neighbourhood_group',y='price',data= Airbnb_df)
```



Observations –>

- price distribution is very high in Manhattan and Brooklyn. but Manhattan have more Diversity in price range, you can see in violin plot.
- Queens and Bronx have same price distribution but in Queens area more distribution in 50$ to 100$ but diversity in price is not like Manhattan and Brooklyn.

---

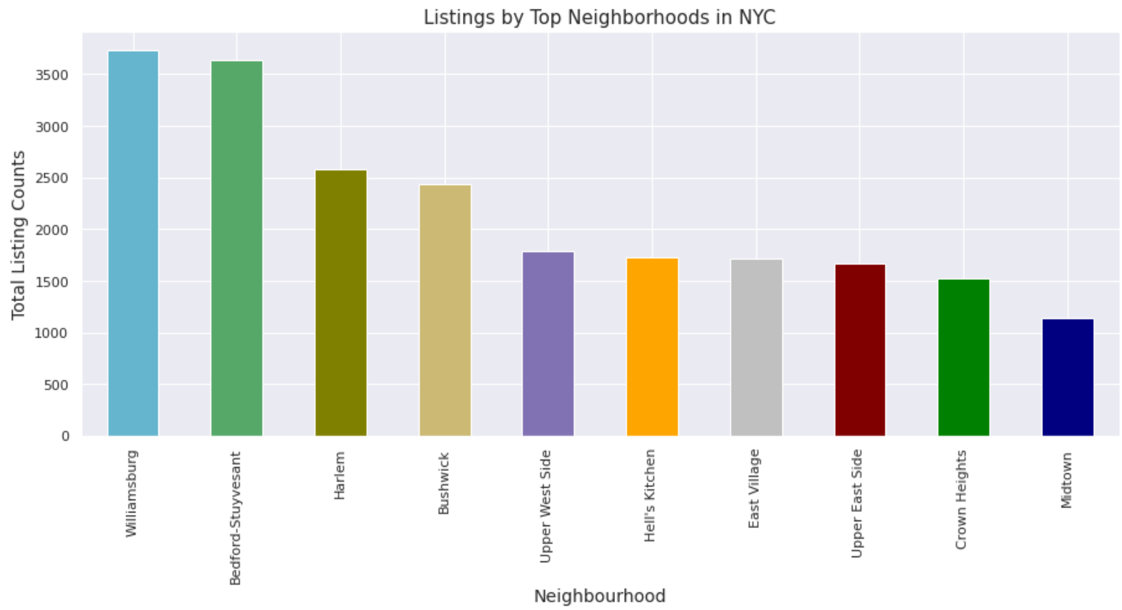
(4) Top Neighborhoods by Listing/property using Bar plot

```
[ ]: # create a new DataFrame that displays the top 10 neighborhoods in the Airbnb␣
     ↪NYC dataset based on the number of listings in each neighborhood
     Top_Neighborhoods = Airbnb_df['neighbourhood'].value_counts()[:10].reset_index()

     # rename the columns of the resulting DataFrame to 'Top_Neighborhoods' and␣
     ↪'Listing_Counts'
     Top_Neighborhoods.columns = ['Top_Neighborhoods', 'Listing_Counts']

     # display the resulting DataFrame
     Top_Neighborhoods
```

```
[ ]:    Top_Neighborhoods  Listing_Counts
     0        Williamsburg            3732
     1  Bedford-Stuyvesant            3638
     2              Harlem            2585
     3            Bushwick            2438
     4     Upper West Side            1788
     5      Hell's Kitchen            1731
     6        East Village            1714
     7     Upper East Side            1670
     8       Crown Heights            1519
     9             Midtown            1143
```

```
[ ]: # Get the top 10 neighborhoods by listing count
     top_10_neigbourhoods = Airbnb_df['neighbourhood'].value_counts().nlargest(10)

     # Create a list of colors to use for the bars
     colors = ['c', 'g', 'olive', 'y', 'm', 'orange', '#C0C0C0', '#800000',␣
      ↪'#008000', '#000080']

     # Create a bar plot of the top 10 neighborhoods using the specified colors
     top_10_neigbourhoods.plot(kind='bar', figsize=(15, 6), color = colors)

     # Set the x-axis label
     plt.xlabel('Neighbourhood', fontsize=14)

     # Set the y-axis label
     plt.ylabel('Total Listing Counts', fontsize=14)

     # Set the title of the plot
     plt.title('Listings by Top Neighborhoods in NYC', fontsize=15)
```

```
[ ]: Text(0.5, 1.0, 'Listings by Top Neighborhoods in NYC')
```



Observations –>

- The top neighborhoods in New York City in terms of listing counts are Williamsburg, Bedford-Stuyvesant, Harlem, Bushwick, and the Upper West Side.
- The top neighborhoods are primarily located in Brooklyn and Manhattan. This may be due to the fact that these boroughs have a higher overall population and a higher demand for housing.
- The number of listings alone may not be indicative of the overall demand for housing in a particular neighborhood, as other factors such as the cost of living and the availability of housing may also play a role.

---

(5) Top Hosts With More Listing/Property using Bar chart

[ ]:
```
# create a new DataFrame that displays the top 10 hosts in the Airbnb NYC␣
↪dataset based on the number of listings each host has
top_10_hosts = Airbnb_df['host_name'].value_counts()[:10].reset_index()

# rename the columns of the resulting DataFrame to 'host_name' and␣
↪'Total_listings'
top_10_hosts.columns = ['host_name', 'Total_listings']

# display the resulting DataFrame
top_10_hosts
```

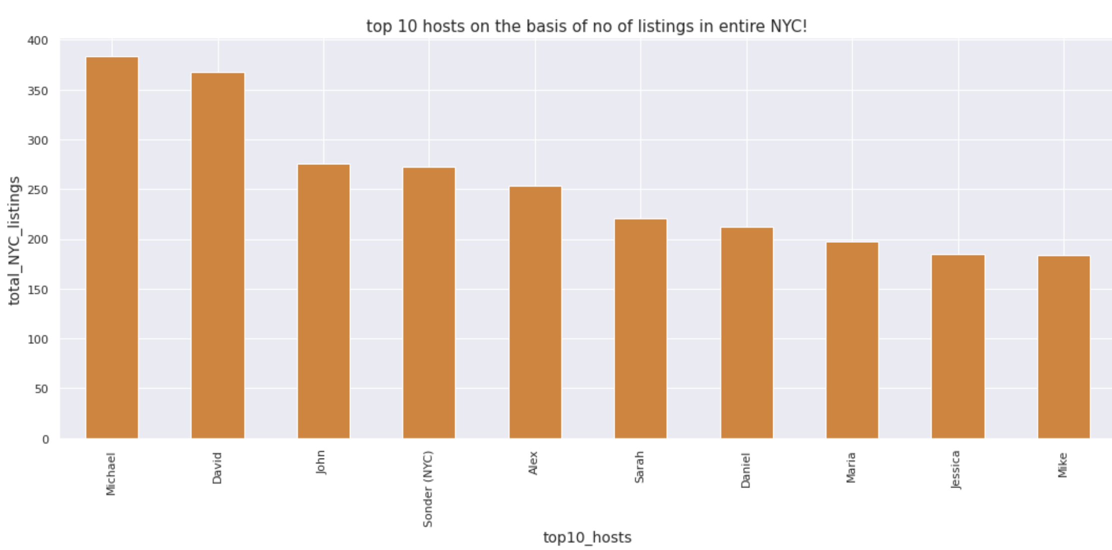
[ ]:

	host_name	Total_listings
0	Michael	383
1	David	368
2	John	276
3	Sonder (NYC)	272
4	Alex	253
5	Sarah	221
6	Daniel	212
7	Maria	197
8	Jessica	185
9	Mike	184

[ ]:

```python
# Get the top 10 hosts by listing count
top_hosts = Airbnb_df['host_name'].value_counts()[:10]

# Create a bar plot of the top 10 hosts
top_hosts.plot(kind='bar', color='peru', figsize=(18, 7))

# Set the x-axis label
plt.xlabel('top10_hosts', fontsize=14)

# Set the y-axis label
plt.ylabel('total_NYC_listings', fontsize=14)

# Set the title of the plot
plt.title('top 10 hosts on the basis of no of listings in entire NYC!',
 ↪fontsize=15)
```

[ ]: Text(0.5, 1.0, 'top 10 hosts on the basis of no of listings in entire NYC!')



Observations –>

- The top three hosts in terms of total listings are Michael, David, and John, who have 383, 368, and 276 listings, respectively.
- There is a relatively large gap between the top two hosts and the rest of the hosts. For example, john has 276 listings, which is significantly fewer than Michael's 383 listings.
- In this top10 list Mike has 184 listings, which is significantly fewer than Michael's 383 listings. This could indicate that there is a lot of variation in the success of different hosts on Airbnb.
- There are relatively few hosts with a large number of listings. This could indicate that the Airbnb market is relatively competitive, with a small number of hosts dominating a large portion of the market.

---

(6) Number Of Active Hosts Per Location Using Line Chart

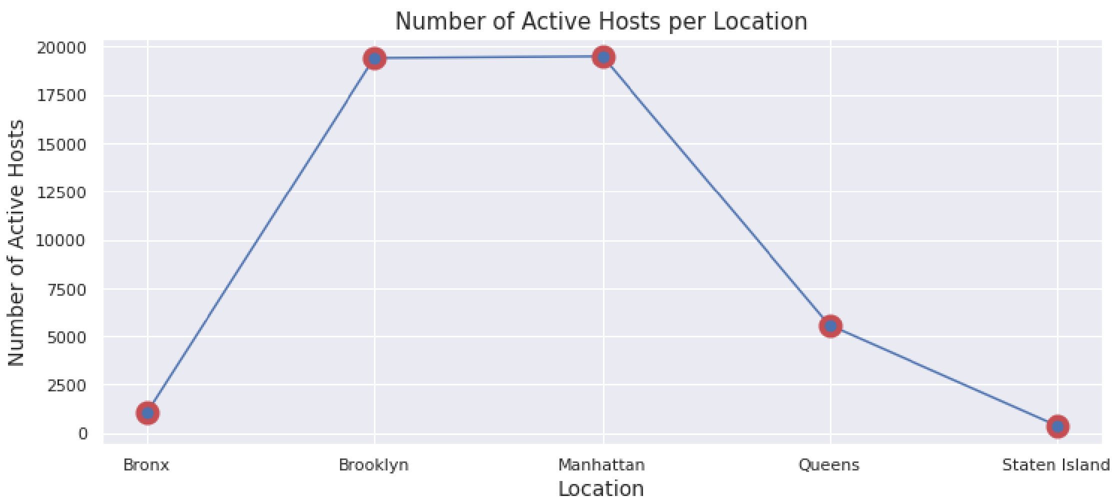
[ ]:
```
# create a new DataFrame that displays the number of hosts in each neighborhood␣
↪group in the Airbnb NYC dataset
hosts_per_location = Airbnb_df.groupby('neighbourhood_group')['listing_id'].
↪count().reset_index()

# rename the columns of the resulting DataFrame to 'Neighbourhood_Groups' and␣
↪'Host_counts'
hosts_per_location.columns = ['Neighbourhood_Groups', 'Host_counts']

# display the resulting DataFrame
hosts_per_location
```

[ ]:

	Neighbourhood_Groups	Host_counts
0	Bronx	1070
1	Brooklyn	19415
2	Manhattan	19501
3	Queens	5567
4	Staten Island	365

[ ]:
```
# Group the data by neighbourhood_group and count the number of listings for␣
↪each group
hosts_per_location = Airbnb_df.groupby('neighbourhood_group')['listing_id'].
↪count()

# Get the list of neighbourhood_group names
locations = hosts_per_location.index

# Get the list of host counts for each neighbourhood_group
host_counts = hosts_per_location.values
```

```python
# Set the figure size
plt.figure(figsize=(12, 5))

# Create the line chart with some experiments using marker function
plt.plot(locations, host_counts, marker='o', ms=12, mew=4, mec='r')

# Add a title and labels to the x-axis and y-axis
plt.title('Number of Active Hosts per Location', fontsize='15')
plt.xlabel('Location', fontsize='14')
plt.ylabel('Number of Active Hosts', fontsize='14')

# Show the plot
plt.show()
```



Observations –>

- Manhattan has the largest number of hosts with 19501,Brooklyn has the second largest number of hosts with 19415.
- After that Queens with 5567 and the Bronx with 1070. while Staten Island has the fewest with 365.
- Brooklyn and Manhattan have the largest number of hosts, with more than double the number of hosts in Queens and more than 18 times the number of hosts in the Bronx.

---

(7) Average Minimum Price In Neighborhoods using Scatter and Bar chart

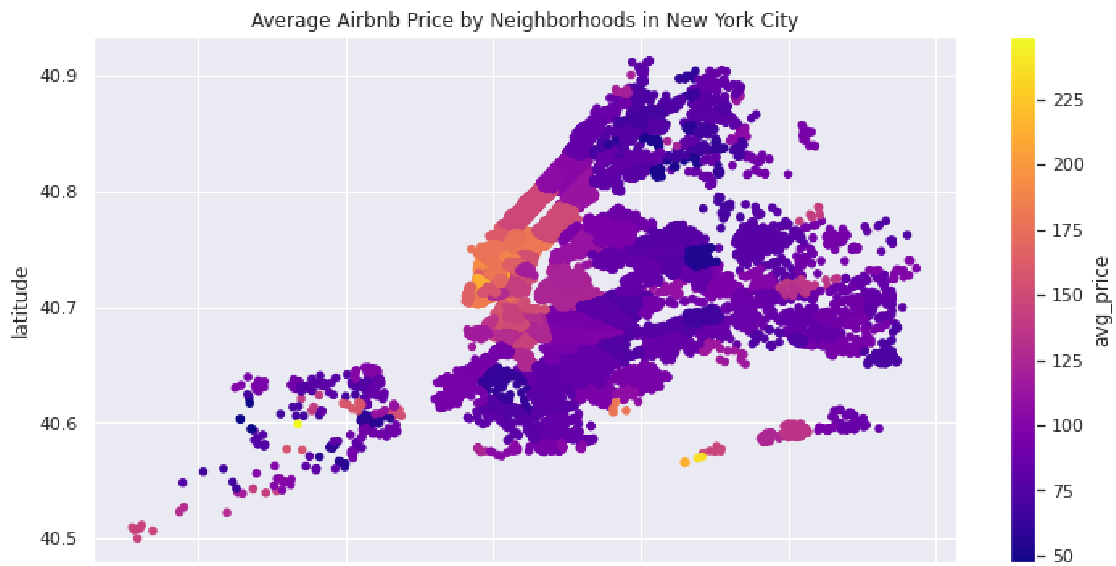
[ ]:
```python
# create a new DataFrame that displays the average price of Airbnb rentals in␣
↪each neighborhood
```

```
neighbourhood_avg_price = Airbnb_df.groupby("neighbourhood").mean().
 ↪reset_index().rename(columns={"price": "avg_price"})[['neighbourhood',␣
 ↪'avg_price']]

# select the top 10 neighborhoods with the lowest average prices
neighbourhood_avg_price = neighbourhood_avg_price.sort_values("avg_price").
 ↪head(10)

# join the resulting DataFrame with the 'neighbourhood_group' column from the␣
 ↪Airbnb NYC dataset, dropping any duplicate entries
neighbourhood_avg_price_sorted_with_group = neighbourhood_avg_price.
 ↪join(Airbnb_df[['neighbourhood', 'neighbourhood_group']].drop_duplicates().
 ↪set_index('neighbourhood'),
                                                                                ␣
 ↪on='neighbourhood')

# Display the resulting data
display(neighbourhood_avg_price_sorted_with_group.style.hide_index())
```

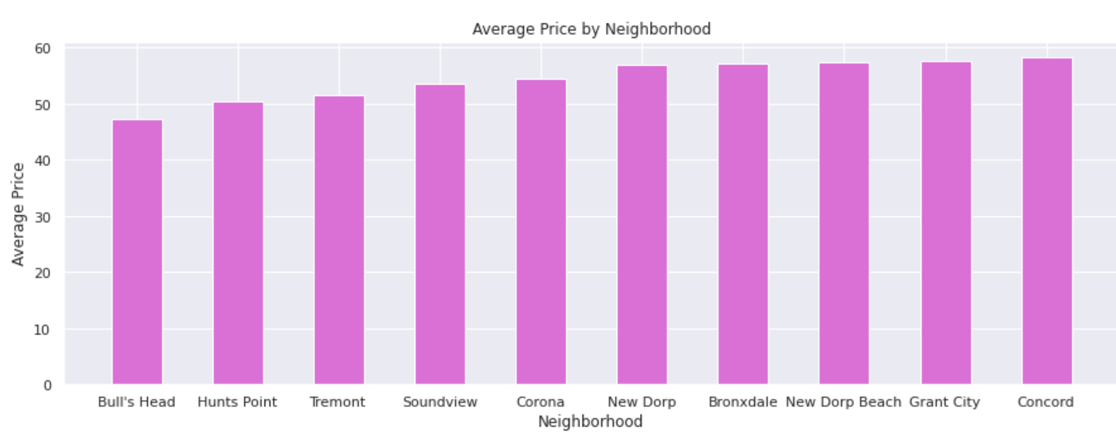
```
<pandas.io.formats.style.Styler at 0x7fb129c8d220>
```

[ ]:
```
neighbourhood_avg_price = (Airbnb_df.groupby("neighbourhood").mean().
 ↪reset_index().rename(columns={"price": "avg_price"}))[['neighbourhood',␣
 ↪'avg_price']]
neighbourhood_avg_price = (neighbourhood_avg_price.sort_values("avg_price"))

# Group the data by neighborhood and calculate the average price
neighbourhood_avg_price = Airbnb_df.groupby("neighbourhood")["price"].mean()

# Create a new DataFrame with the average price for each neighborhood
neighbourhood_prices = pd.DataFrame({"neighbourhood": neighbourhood_avg_price.
 ↪index, "avg_price": neighbourhood_avg_price.values})

# Merge the average price data with the original DataFrame#trying to find where␣
 ↪the coordinates belong from the latitude and longitude
df = Airbnb_df.merge(neighbourhood_prices, on="neighbourhood")

# Create the scattermapbox plot
fig = df.plot.scatter(x="longitude", y="latitude", c="avg_price",␣
 ↪title="Average Airbnb Price by Neighborhoods in New York City",␣
 ↪figsize=(12,6), cmap="plasma")
fig
```

[ ]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x7fb1299374f0>
```



[ ]:
```python
# Extract the values from the dataset
neighborhoods = neighbourhood_avg_price_sorted_with_group['neighbourhood']
prices = neighbourhood_avg_price_sorted_with_group['avg_price']

# Create the bar plot
plt.figure(figsize=(15,5))
plt.bar(neighborhoods, prices,width=0.5, color = 'orchid')
plt.xlabel('Neighborhood')
plt.ylabel('Average Price')
plt.title('Average Price by Neighborhood')

# Show the plot
plt.show()
```



Observations –>

- All of the neighborhoods listed are located in the outer boroughs of New York City (Bronx, Queens, and Staten Island). This suggests that these neighborhoods may have a lower overall cost of living compared to neighborhoods in Manhattan and Brooklyn.
- Most of these neighborhoods are located in the Bronx and Staten Island. These boroughs tend to have a lower overall cost of living compared to Manhattan and Brooklyn.
- These neighborhoods may be attractive to renters or buyers looking for more affordable housing options in the New York City area.

---

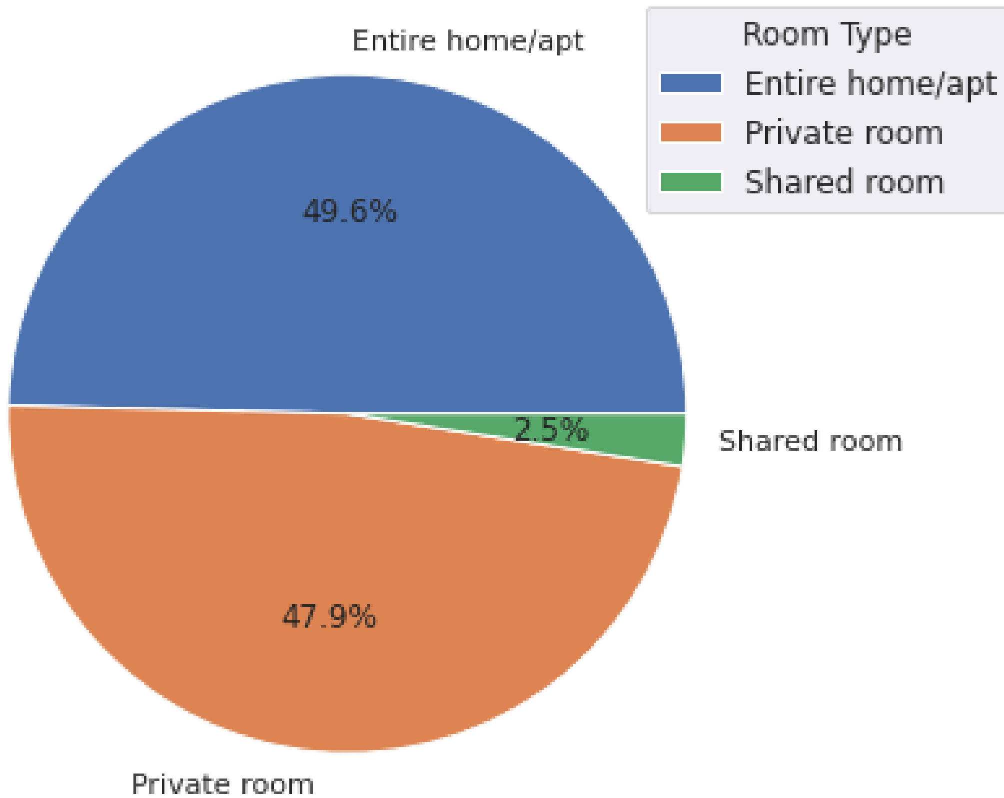
(8) Total Counts Of Each Room Type

[ ]:
```
# create a new DataFrame that displays the number of listings of each room type␣
↪in the Airbnb NYC dataset
top_room_type = Airbnb_df['room_type'].value_counts().reset_index()

# rename the columns of the resulting DataFrame to 'Room_Type' and␣
↪'Total_counts'
top_room_type.columns = ['Room_Type', 'Total_counts']

# display the resulting DataFrame
top_room_type
```

[ ]:
```
         Room_Type  Total_counts
0  Entire home/apt         22784
1     Private room         21996
2      Shared room          1138
```

[ ]:
```
# Set the figure size
plt.figure(figsize=(10, 6))

# Get the room type counts
room_type_counts = Airbnb_df['room_type'].value_counts()

# Set the labels and sizes for the pie chart
labels = room_type_counts.index
sizes = room_type_counts.values

# Create the pie chart
plt.pie(sizes, labels=labels, autopct='%1.1f%%')

# Add a legend to the chart
plt.legend(title='Room Type', bbox_to_anchor=(0.8, 0, 0.5, 1), fontsize='12')
```

```
# Show the plot
plt.show()
```



Observations –>

- The majority of listings on Airbnb are for entire homes or apartments, with 22784 listings, followed by private rooms with 21996 listings, and shared rooms with 1138 listings.
- There is a significant difference in the number of listings for each room type. For example, there are almost 20 times as many listings for entire homes or apartments as there are for shared rooms.
- The data suggests that travelers using Airbnb have a wide range of accommodation options to choose from, including private rooms and entire homes or apartments

---

---

(9) Stay Requirement counts by Minimum Nights using Bar chart

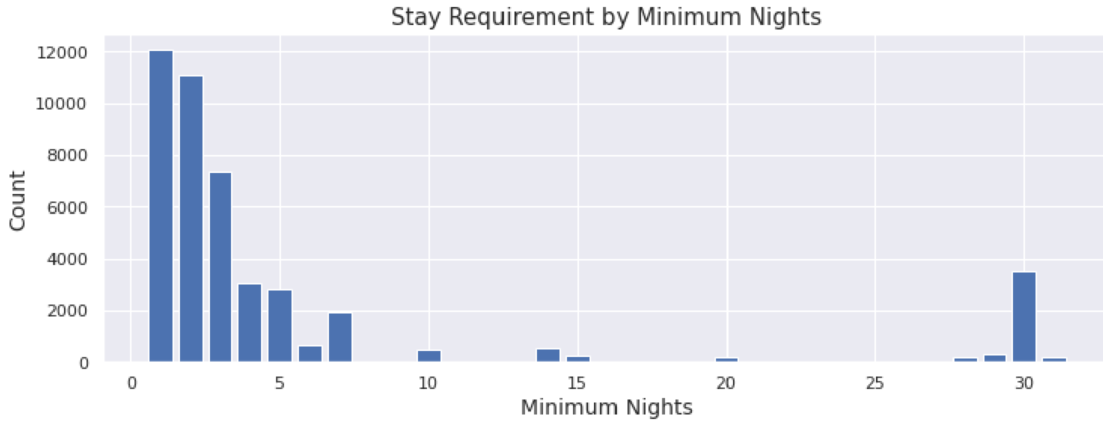
[ ]:
```python
# Group the DataFrame by the minimum_nights column and count the number of rows␣
↪in each group
min_nights_count = Airbnb_df.groupby('minimum_nights').size().reset_index(name␣
↪= 'count')

# Sort the resulting DataFrame in descending order by the count column
min_nights_count = min_nights_count.sort_values('count', ascending=False)

# Select the top 10 rows
min_nights_count = min_nights_count.head(15)

# Reset the index
min_nights_count = min_nights_count.reset_index(drop=True)

# Display the resulting DataFrame
min_nights_count
```

[ ]:

	minimum_nights	count
0	1	12067
1	2	11080
2	3	7375
3	30	3489
4	4	3066
5	5	2821
6	7	1951
7	6	679
8	14	539
9	10	462
10	29	327
11	15	272
12	20	215
13	31	189
14	28	173

[ ]:
```python
# Extract the minimum_nights and count columns from the DataFrame
minimum_nights = min_nights_count['minimum_nights']
count = min_nights_count['count']

# Set the figure size
plt.figure(figsize=(12, 4))

# Create the bar plot
plt.bar(minimum_nights, count)

# Add axis labels and a title
plt.xlabel('Minimum Nights', fontsize='14')
plt.ylabel('Count', fontsize='14')
```

```
plt.title('Stay Requirement by Minimum Nights', fontsize='15')

# Show the plot
plt.show()
```



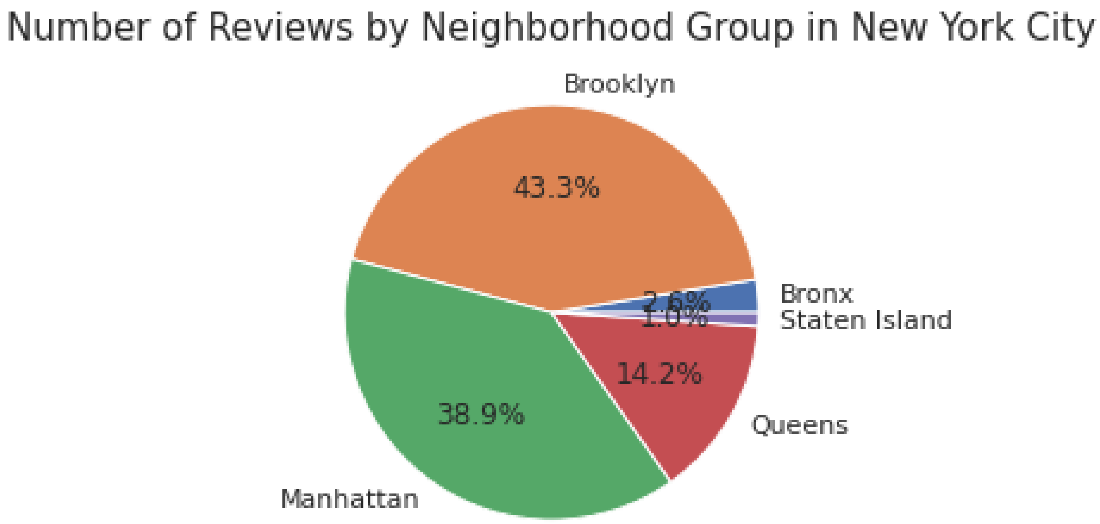
Observations –>

- The majority of listings on Airbnb have a minimum stay requirement of 1 or 2 nights, with 12067 and 11080 listings, respectively.
- The number of listings with a minimum stay requirement decreases as the length of stay increases, with 7375 listings requiring a minimum stay of 3 nights, and so on.
- There are relatively few listings with a minimum stay requirement of 30 nights or more, with 3489 and 189 listings, respectively.

---

(10) Total Reviews by Each Neighborhood Group using Pie Chart

[ ]:
```
# Group the data by neighborhood group and calculate the total number of reviews
reviews_by_neighbourhood_group = Airbnb_df.
 ↪groupby("neighbourhood_group")["total_reviews"].sum()

# Create a pie chart
plt.pie(reviews_by_neighbourhood_group, labels=reviews_by_neighbourhood_group.
 ↪index, autopct='%1.1f%%')
plt.title("Number of Reviews by Neighborhood Group in New York City",␣
 ↪fontsize='15')

# Display the chart
plt.show()
```



Observations –>

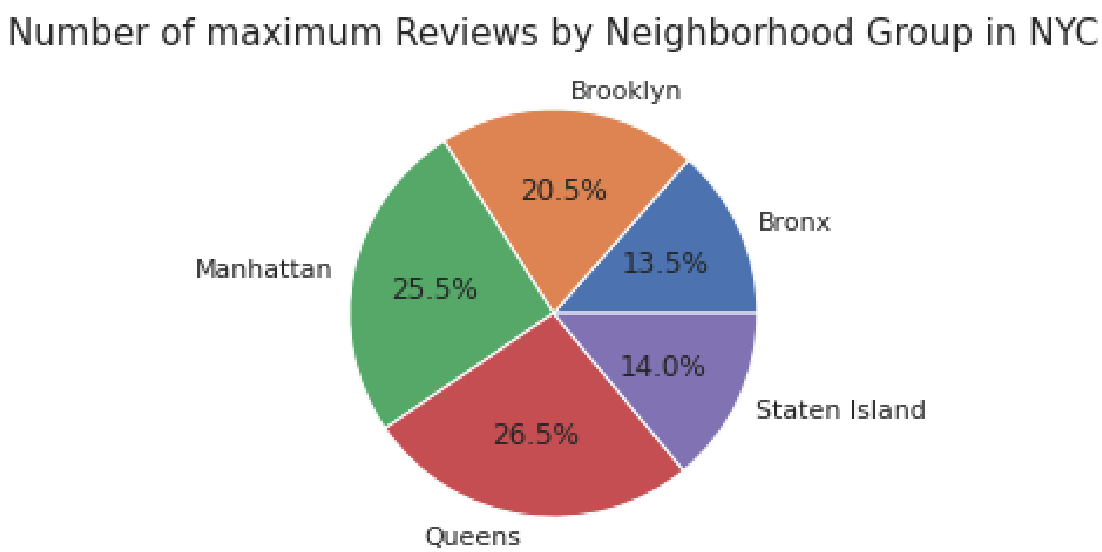
- Brooklyn has the largest share of total reviews on Airbnb, with 43.3%, followed by Manhattan with 38.9%.
- Queens has the third largest share of total reviews, with 14.2%, followed by the Bronx with 2.6% and Staten Island with 1.0%.
- The data suggests that Airbnb is more popular in Brooklyn and Manhattan compared to the other neighborhood groups.
- Despite having fewer listings, Brooklyn has more reviews on Airbnb compared to Manhattan. This could indicate that Airbnb users in Brooklyn are more likely to leave reviews, or that the listings in Brooklyn are more popular or successful in generating positive reviews. It is worth noting that there could be a number of other factors that could contribute to this difference in reviews, such as the quality of the listings or the characteristics of the travelers who use Airbnb in these areas.

---

(11) Number of Max. Reviews by Each Neighborhood Group using Pie Chart

[ ]:
```
# Group the Airbnb data by neighbourhood group
reviews_by_neighbourhood_group = Airbnb_df.
  ↪groupby("neighbourhood_group")["total_reviews"].max()

# Create a pie chart to visualize the distribution of maximum number of reviews␣
  ↪among different neighbourhood groups
```

```
plt.pie(reviews_by_neighbourhood_group, labels=reviews_by_neighbourhood_group.
    ↪index, autopct='%1.1f%%')

# Add a title to the chart
plt.title("Number of maximum Reviews by Neighborhood Group in NYC", 
    ↪fontsize='15')

# Display the chart
plt.show()
```



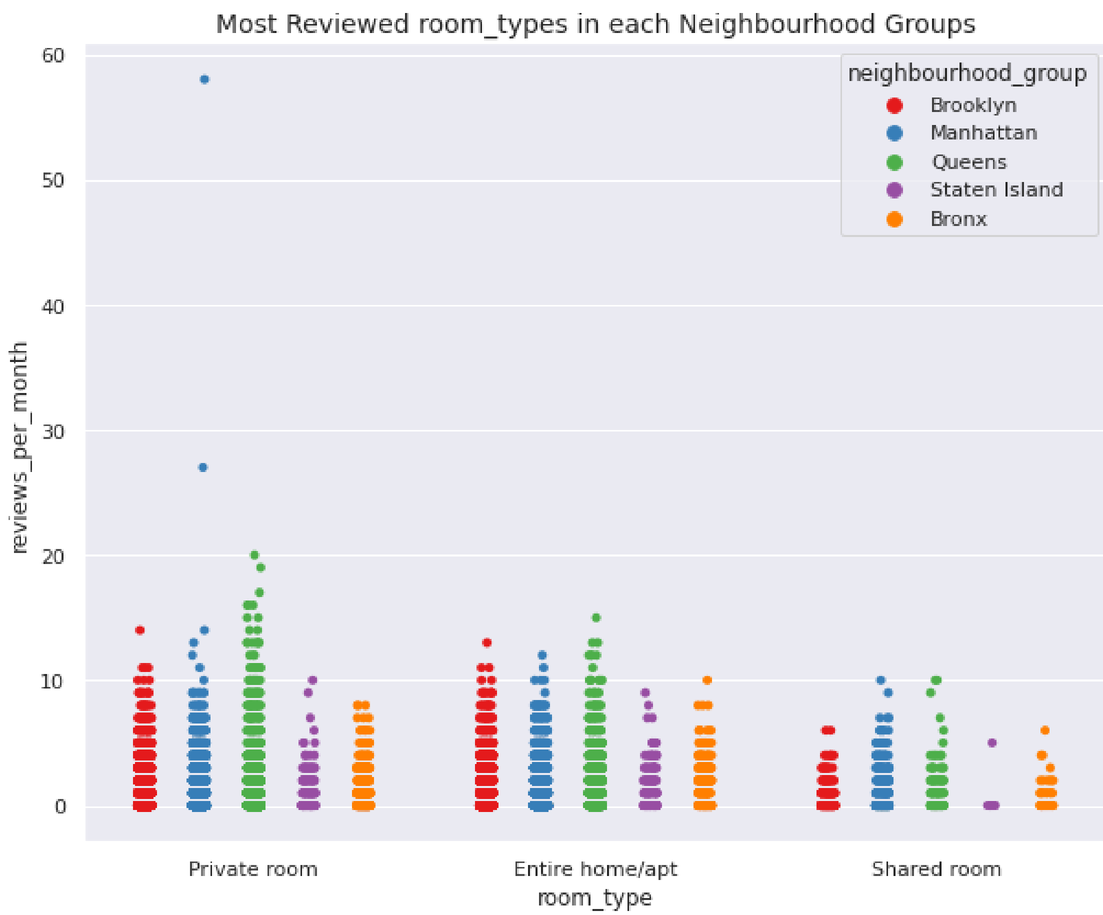
Observations –>

- Queens and Manhattan seem to be the most popular neighborhoods for reviewing, as they have both high number of maximum reviews.
- Queens has the highest percentage of reviews at 26.5%, but it has the third highest number of listings, behind Manhattan and Brooklyn. This suggests that Queens may be a particularly popular destination for tourists or visitors, even though it has fewer listings compared to Manhattan and Brooklyn.
- Manhattan and Brooklyn also have a high percentage of reviews, at 25.5% & 20.5%. This indicates that it is a popular destination for tourists or visitors as well. (number of listings higher than queens)
- Overall, this data suggests that Queens, Manhattan, and Brooklyn are the most popular neighborhoods for tourists or visitors, based on the high number of reviews they receive.

(12) most reviewed room type per month in neighbourhood groups

[ ]:
```
# create a figure with a default size of (10, 8)
f, ax = plt.subplots(figsize=(10, 8))

# create a stripplot that displays the number of reviews per month for each␣
↪room type in the Airbnb NYC dataset
ax = sns.stripplot(x='room_type', y='reviews_per_month',␣
↪hue='neighbourhood_group', dodge=True, data=Airbnb_df, palette='Set1')

# set the title of the plot
ax.set_title('Most Reviewed room_types in each Neighbourhood Groups',␣
↪fontsize='14')
```

[ ]:
```
Text(0.5, 1.0, 'Most Reviewed room_types in each Neighbourhood Groups')
```



Observations –>

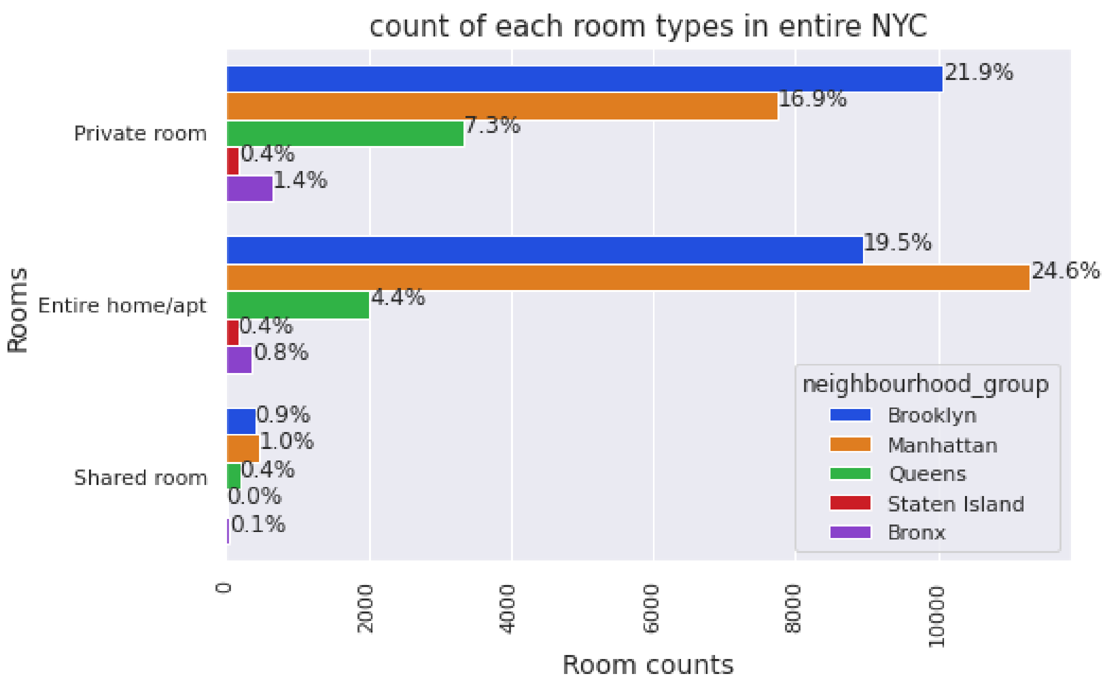
- We can see that Private room recieved the most no of reviews/month where Manhattan had the highest reviews received for Private rooms with more than 50 reviews/month, followed by Manhattan in the chase.

- Manhattan & Queens got the most no of reviews for Entire home/apt room type.
- There were less reviews recieved from shared rooms as compared to other room types and it was from Staten Island followed by Bronx.

---

(13)Count Of Each Room Types In Entire NYC Using Multiple Bar Plot

[ ]:
```python
# Now analysis Room types count in Neighbourhood groups in NYC

# Set the size of the plot
plt.rcParams['figure.figsize'] = (8, 5)

# Create a countplot using seaborn
ax = sns.countplot(y='room_type', hue='neighbourhood_group', data=Airbnb_df,␣
    ↪palette='bright')

# Calculate the total number of room_type values
total = len(Airbnb_df['room_type'])

# Add percentage labels to each bar in the plot
for p in ax.patches:
        percentage = '{:.1f}%'.format(100 * p.get_width()/total)
        x = p.get_x() + p.get_width() + 0.02
        y = p.get_y() + p.get_height()/2
        ax.annotate(percentage, (x, y))

# Add a title to the plot
plt.title('count of each room types in entire NYC', fontsize='15')

# Add a label to the x-axis
plt.xlabel('Room counts', fontsize='14')

# Rotate the x-tick labels
plt.xticks(rotation=90)

# Add a label to the y-axis
plt.ylabel('Rooms', fontsize='14')

# Display the plot
plt.show()
```



Observations –>

- Manhattan has more listed properties with Entire home/apt around 24.6% of total listed properties followed by Brooklyn with around 19.5%.
- Private rooms are more in Brooklyn as in 21.9% of the total listed properties followed by Manhattan with 16.9% of them. While 7.3% of private rooms are from Queens.
- Very few of the total listed have shared rooms listed on Airbnb where there's negligible or almost very rare shared rooms in Staten Island and Bronx.
- We can infer that Brooklyn,Queens,Bronx has more private room types while Manhattan which has the highest no of listings in entire NYC has more Entire home/apt room types.

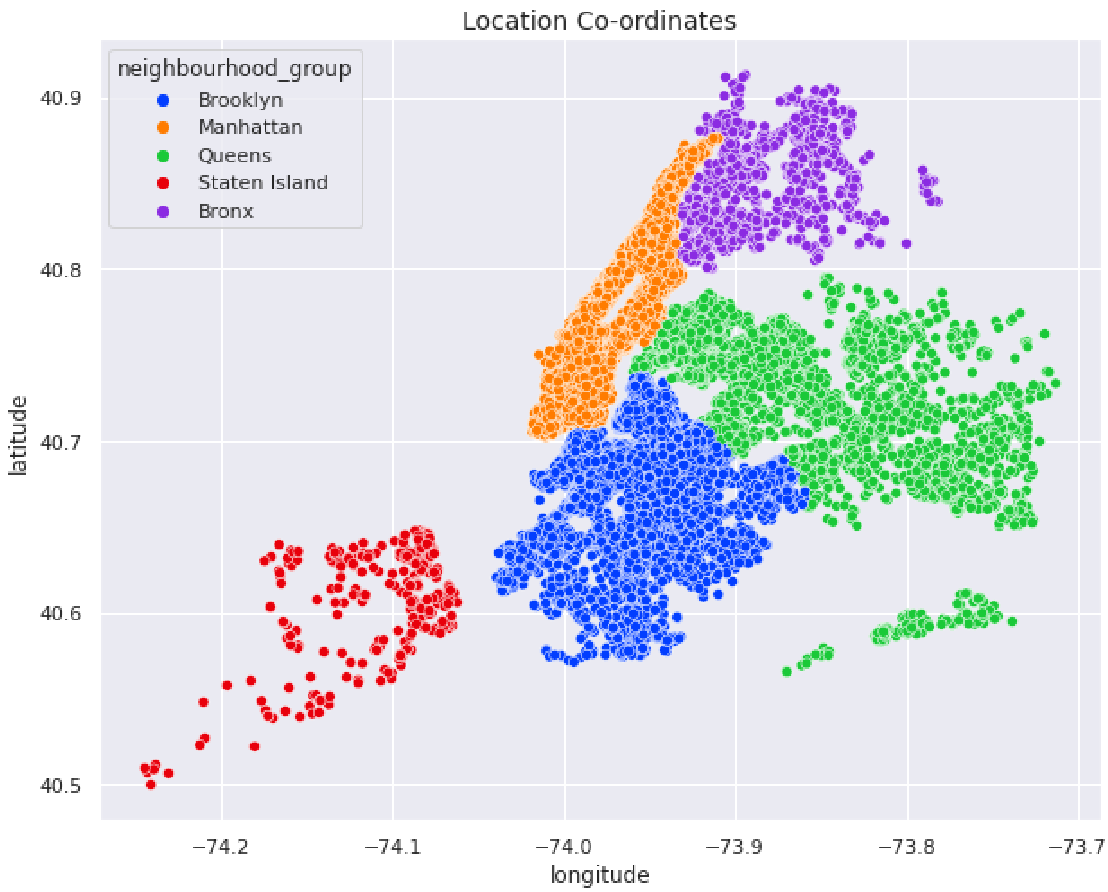
---

(14) use latitude and longitude in scatterplot map and find neighbourhood_groups and Room types in map

```
[ ]: #trying to find where the coordinates belong from the latitude and longitude

# set the default figure size for the seaborn library
sns.set(rc={"figure.figsize": (10, 8)})

# create a scatter plot that displays the longitude and latitude of the␣
 ↪listings in the Airbnb NYC dataset
```

```
ax = sns.scatterplot(data=Airbnb_df, x="longitude", y="latitude",␣
 ↪hue='neighbourhood_group', palette='bright')

# set the title of the plot
ax.set_title('Location Co-ordinates', fontsize='14')
```

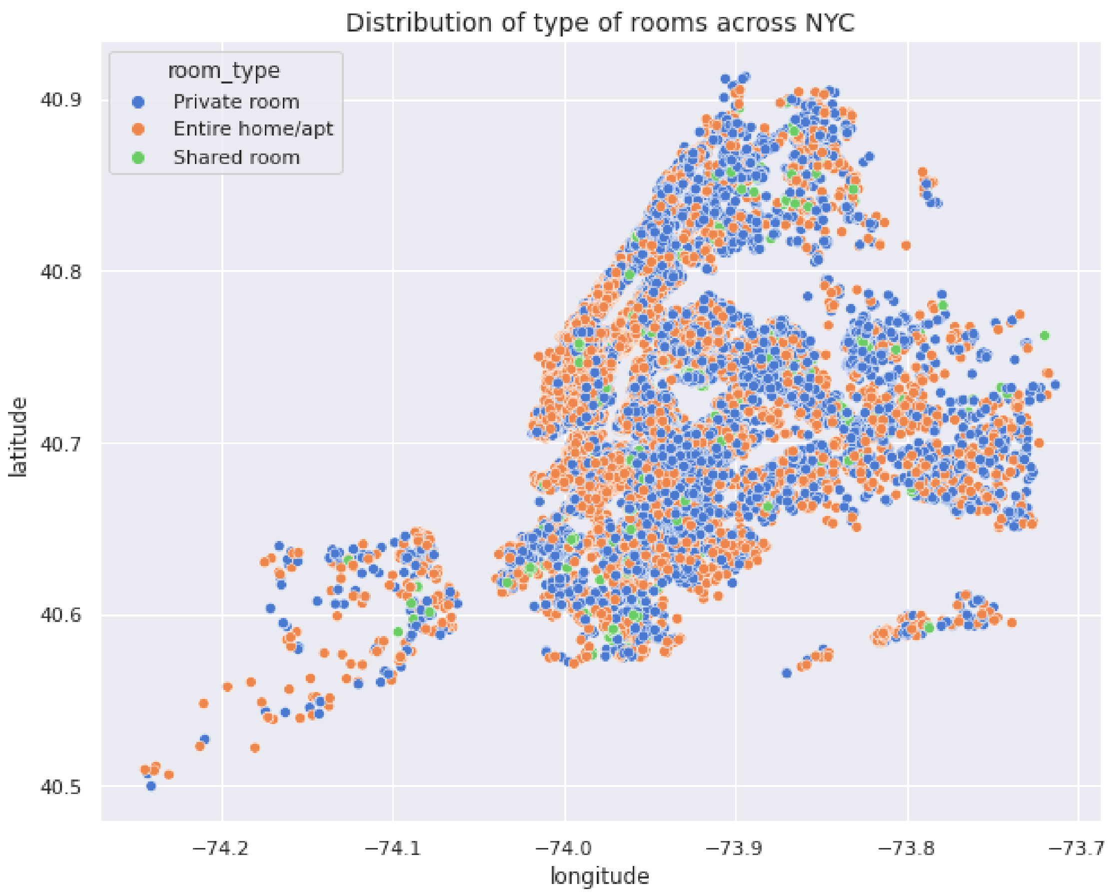
[ ]: Text(0.5, 1.0, 'Location Co-ordinates')



[ ]:
```
# Let's observe the type of room_types

# set the default figure size for the seaborn library
sns.set(rc={"figure.figsize": (10, 8)})

# create a scatter plot that displays the longitude and latitude of the␣
 ↪listings in the Airbnb NYC dataset with room_types.
ax = sns.scatterplot(x=Airbnb_df.longitude, y=Airbnb_df.latitude, hue=Airbnb_df.
 ↪room_type, palette='muted')

# set the title of the plot
```

```
ax.set_title('Distribution of type of rooms across NYC', fontsize='14')
```

[ ]: Text(0.5, 1.0, 'Distribution of type of rooms across NYC')

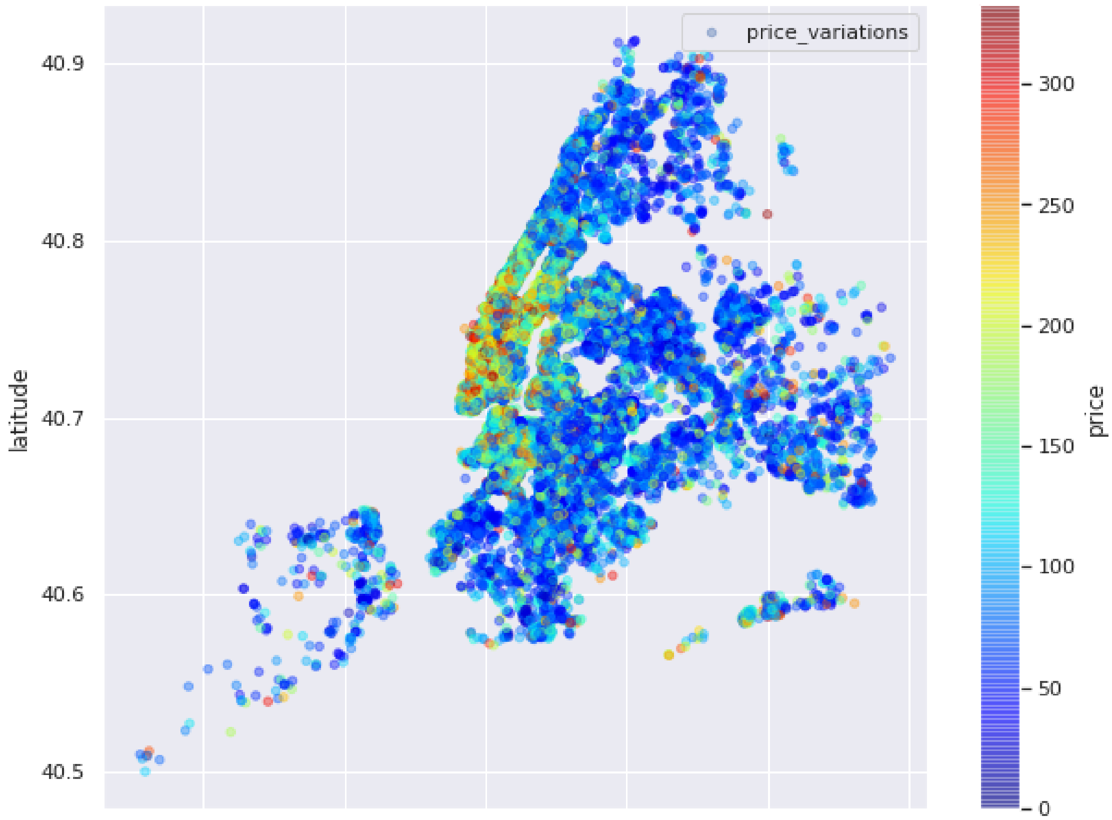


---

(15) Price variations in NYC Neighbourhood groups using scatter plot

[ ]:
```
# Let's have an idea of the price variations in neighborhood_groups

# create a scatter plot that displays the longitude and latitude of the␣
↪listings in the Airbnb NYC dataset, with the color of each point indicating␣
↪the price of the listing
lat_long = Airbnb_df.plot(kind='scatter', x='longitude', y='latitude',␣
↪label='price_variations', c='price',
                    cmap=plt.get_cmap('jet'), colorbar=True, alpha=0.4,␣
↪figsize=(10, 8))

# add a legend to the plot
lat_long.legend()
```

```
[ ]: <matplotlib.legend.Legend at 0x7fb127d67700>
```



Observations –>

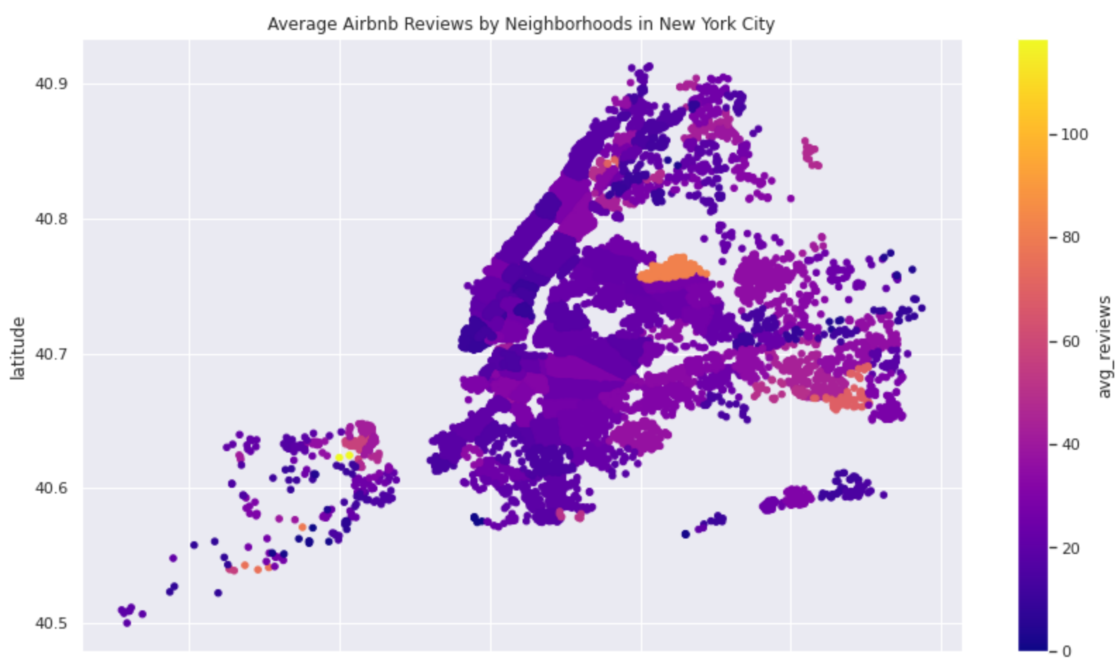
- The range of prices for accommodations in Manhattan is particularly high, indicating that it is the most expensive place to stay in NYC due to its various attractive amenities, as shown in the attached image.
- they are likely to attract a lot of tourists or visitors because of more valuable things to visit so price is higher than other neighbourhood groups.
- Travelers are likely to spent more days in this area because of popular amenities, high concentration of tourist attractions and public transports.

---

(16) Find Best Location Listing/Property Location For Travelers and Hosts

[ ]:
```python
# Group the data by neighborhood and calculate the average number of reviews
neighbourhood_avg_reviews = Airbnb_df.groupby("neighbourhood")["total_reviews"].
↪mean()

# Create a new DataFrame with the average number of reviews for each␣
↪neighborhood
neighbourhood_reviews = pd.DataFrame({"neighbourhood":␣
↪neighbourhood_avg_reviews.index, "avg_reviews": neighbourhood_avg_reviews.
↪values})

# Merge the average number of reviews data with the original DataFrame
df = Airbnb_df.merge(neighbourhood_reviews, on="neighbourhood")

# Create the scattermapbox plot
fig = df.plot.scatter(x="longitude", y="latitude", c="avg_reviews",␣
↪title="Average Airbnb Reviews by Neighborhoods in New York City",␣
↪figsize=(14,8), cmap="plasma")

# Display the scatter map
fig
```

[ ]: <matplotlib.axes._subplots.AxesSubplot at 0x7fb127ecf730>



[ ]:
```python
#from IPython.display import Image
```

```
# Replace "path/to/photo.jpg" with the actual path to your photo on Google Drive
#Image("path/to/photo.jpg", resize=(600, 400))  # Set the width to 600 pixels␣
↪and the height to 400 pixels
```

Observations –>

- I have attached a photo of this map because of some valuable insight. The neighborhoods near the airport in Queens would have a higher average number of reviews, as they are likely to attract a lot of tourists or visitors who are passing through the area. The proximity to the airport could make these neighborhoods a convenient and appealing place to stay for travelers.
- There could also be other factors contributing to the high average number of reviews in these neighborhoods. For example, they may have a higher concentration of high-quality listings or attractions that attract more visitors and result in more reviews and Airport is key factor i think this is make sense.

---

(17) Correlation Heatmap Visualization

[ ]:
```
# Calculate pairwise correlations between columns
corr = Airbnb_df.corr()

# Display the correlation between columns
corr
```

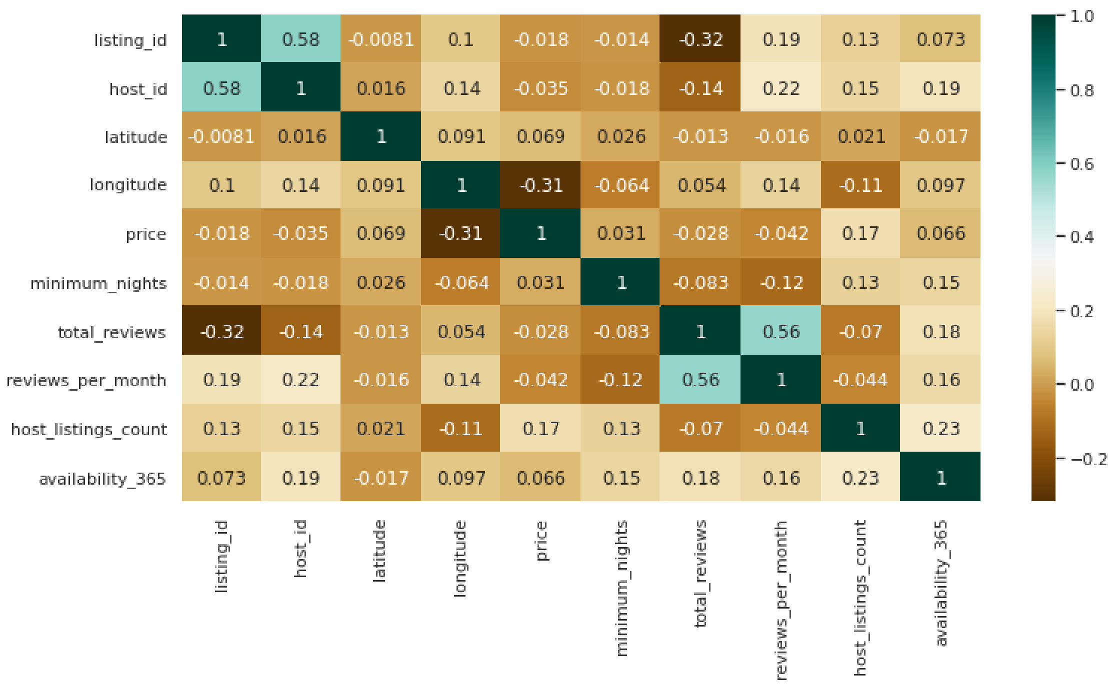
[ ]:

	listing_id	host_id	latitude	longitude	price
listing_id	1.000000	0.581439	-0.008072	0.101403	-0.018180
host_id	0.581439	1.000000	0.015965	0.144330	-0.034812
latitude	-0.008072	0.015965	1.000000	0.091354	0.068789
longitude	0.101403	0.144330	0.091354	1.000000	-0.306922
price	-0.018180	-0.034812	0.068789	-0.306922	1.000000
minimum_nights	-0.013841	-0.017972	0.025853	-0.064128	0.031141
total_reviews	-0.320428	-0.136529	-0.012515	0.053831	-0.027547
reviews_per_month	0.189768	0.216020	-0.015752	0.135783	-0.041992
host_listings_count	0.125179	0.147276	0.021285	-0.107333	0.172891
availability_365	0.073188	0.193673	-0.017492	0.097181	0.066179

	minimum_nights	total_reviews	reviews_per_month
listing_id	-0.013841	-0.320428	0.189768
host_id	-0.017972	-0.136529	0.216020
latitude	0.025853	-0.012515	-0.015752
longitude	-0.064128	0.053831	0.135783
price	0.031141	-0.027547	-0.041992
minimum_nights	1.000000	-0.082851	-0.117291

```
total_reviews            -0.082851      1.000000          0.562593
reviews_per_month        -0.117291      0.562593          1.000000
host_listings_count       0.133237     -0.070357         -0.043678
availability_365          0.146329      0.183707          0.156463

                     host_listings_count  availability_365
listing_id                      0.125179          0.073188
host_id                         0.147276          0.193673
latitude                        0.021285         -0.017492
longitude                      -0.107333          0.097181
price                           0.172891          0.066179
minimum_nights                  0.133237          0.146329
total_reviews                  -0.070357          0.183707
reviews_per_month              -0.043678          0.156463
host_listings_count             1.000000          0.225251
availability_365                0.225251          1.000000
```

[ ]:
```python
# Set the figure size
plt.figure(figsize=(12,6))

# Visualize correlations as a heatmap
sns.heatmap(corr, cmap='BrBG',annot=True)

# Display heatmap
plt.show()
```



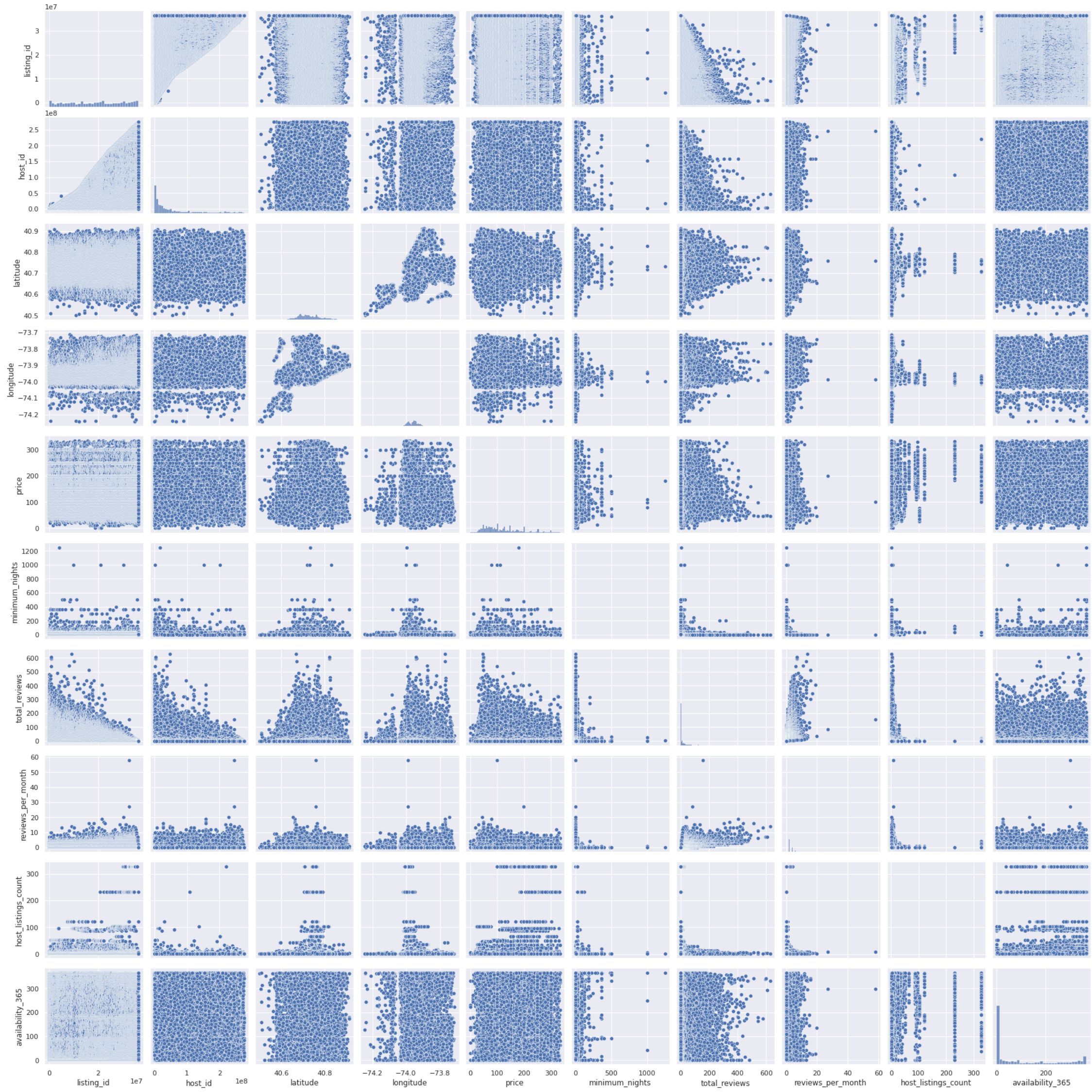
Observations –>

- There is a moderate positive correlation (0.58) between the host_id and id columns, which suggests that hosts with more listings are more likely to have unique host IDs.
- There is a weak positive correlation (0.17) between the price column and the calculated_host_listings_count column, which suggests that hosts with more listings tend to charge higher prices for their listings.
- There is a moderate positive correlation (0.23) between the calculated_host_listings_count column and the availability_365 column, which suggests that hosts with more listings tend to have more days of availability in the next 365 days.
- There is a strong positive correlation (0.58) between the number_of_reviews column and the reviews_per_month column, which suggests that listings with more total reviews tend to have more reviews per month.

---

(18) Pair Plot Visualization

[ ]:
```
# create a pairplot using the seaborn library to visualize the relationships␣
↪between different variables in the Airbnb NYC dataset
sns.pairplot(Airbnb_df)

# show the plot
plt.show()
```



- A pair plot consists of multiple scatterplots arranged in a grid, with each scatterplot showing the relationship between two variables
- It can be used to visualize relationships between multiple variables and to identify patterns in the data.

8.1 BUSINESS CONCLUSION :-

- Manhattan and Brooklyn have the highest demand for Airbnb rentals, as evidenced by the large number of listings in these neighborhoods. This could make them attractive areas for hosts to invest in property.

- Manhattan is world-famous for its parks, museums, buildings, town, liberty, gardens, markets, island and also its substantial number of tourists throughout the year ,it makes sense that demand and price both high.
- Brooklyn comes in second with significant number of listings and cheaper prices as compared to the Manhattan: With most listings located in Williamsburg and Bedford Stuyvesant two neighborhoods strategically close to Manhattan tourists get the chance to enjoy both boroughs equally while spending less.
- Williamsburg, Bedford-Stuyvesant, Harlem, Bushwick, and the Upper West Side are the top neighborhoods in terms of listing counts, indicating strong demand for Airbnb rentals in these areas.
- The average price of a listing in New York City is higher in the center of the city (Manhattan) compared to the outer boroughs. This could indicate that investing in property in Manhattan may be more lucrative for Airbnb rentals. But Manhattan and Brooklyn have the largest number of hosts, indicating a high level of competition in these boroughs.
- The data suggests that Airbnb rentals are primarily used for short-term stays, with relatively few listings requiring a minimum stay of 30 nights or more. Hosts may want to consider investing in property that can accommodate shorter stays in order to maximize their occupancy rate.
- The majority of listings on Airbnb are for entire homes or apartments and also Private Rooms with relatively fewer listings for shared rooms. This suggests that travelers using Airbnb have a wide range of accommodation options to choose from, and hosts may want to consider investing in property that can accommodate multiple guests.
- The data indicates that the availability of Airbnb rentals varies significantly across neighborhoods, with some neighborhoods having a high concentration of listings and others having relatively few.
- The data indicates that there is a high level of competition among Airbnb hosts, with a small number of hosts dominating a large portion of the market. Hosts may want to consider investing in property in areas with relatively fewer listings in order to differentiate themselves from the competition.
- The neighborhoods near the airport in Queens would have a higher average number of reviews, as they are likely to attract a lot of tourists or visitors who are passing through the area. The proximity to the airport could make these neighborhoods a convenient and appealing place to stay for travelers for short-term stay with spending less money because The price distribution is high in Manhattan and Brooklyn.

---

9 Thank You