Airbnb San Francisco

January 2, 2024

[1]:
```
#import libraries
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import matplotlib.image as mpimg
%matplotlib inline
import seaborn as sns
```

[2]:
```
#data extraction
df=pd.read_csv('listings.csv')
df.head()
```

[2]:

	id	host_id	host_name	name	neighbourhood	RATING	RATING_FINAL	BEDROOM_FINAL	BED-FINAL	BEDROOM	…	room_type	price	minimum_nights	number_of_reviews	last_review	reviews_per_month	calculated_host_listings_count	availability_365
0	138592.0	648553	Julie & Brad	Home in San Francisco	Noe Valley	4.91	4.91	4 bedroom	4 bed	4 bedroom	…	Entire home/apt	1250	5	34	18-08-2023	0.23	1	80
1	474107.0	2318089	Inés	Home in San Francisco	Excelsior	4.91	4.91	2 bedroom	2 bed	2 bedroom	…	Private room	79	2	149	14-11-2023	1.92	1	196
2	487019.0	2410550	Cecile	Rental unit in San Francisco	Potrero Hill	4.53	4.53	3 bedroom	4 bed	3 bedroom	…	Entire home/apt	249	5	36	18-07-2023	0.26	1	202
3	487499.0	1682183	Daniel	Rental unit in San Francisco	Mission	4.78	4.78	1 bedroom	1 bed	1 bedroom	…	Entire home/apt	125	30	196	04-11-2023	1.41	1	149
4	1163946.0	1918239	Jen	Condo in San Francisco	Mission	2 bedroom	0.00	2 bedroom	2 bed	2 bed	…	Entire home/apt	70	30	2	07-10-2016			

```
4              0.02                                1                 0

   number_of_reviews_ltm              license
0                      2  Pending Application
1                     20          STR-0007772
2                      7          STR-0001599
3                      4                  NaN
4                      0                  NaN

[5 rows x 25 columns]
```

[3]:
```
#drop duplicate columns
df = df.drop('RATING', axis=1)
df = df.drop('BEDROOM', axis=1)
df = df.drop('bed', axis=1)
df = df.drop('BATH', axis=1)
df.rename(columns={'RATING_FINAL':'rating'},inplace=True)
df.rename(columns={'BEDROOM_FINAL':'bedroom'},inplace=True)
df.rename(columns={'BED-FINAL':'beds'},inplace=True)
df.rename(columns={'BATH FINAL':'bathroom'},inplace=True)
df.head()
```

[3]:
```
           id  host_id     host_name                           name  \
0   138592.0   648553  Julie & Brad         Home in San Francisco
1   474107.0  2318089          Inés         Home in San Francisco
2   487019.0  2410550        Cecile  Rental unit in San Francisco
3   487499.0  1682183        Daniel  Rental unit in San Francisco
4  1163946.0  1918239           Jen        Condo in San Francisco

  neighbourhood  rating     bedroom    beds       bathroom  latitude  …  \
0    Noe Valley    4.91  4 bedroom   4 bed          2 bath  37.75635  …
1     Excelsior    4.91  2 bedroom   2 bed   1 shared bath  37.72369  …
2  Potrero Hill    4.53  3 bedroom   4 bed          2 bath  37.75622  …
3       Mission    4.78  1 bedroom   1 bed          1 bath  37.75884  …
4       Mission    0.00  2 bedroom   2 bed          1 bath  37.76111  …

         room_type price  minimum_nights  number_of_reviews  last_review  \
0  Entire home/apt  1250               5                 34   18-08-2023
1     Private room    79               2                149   14-11-2023
2  Entire home/apt   249               5                 36   18-07-2023
3  Entire home/apt   125              30                196   04-11-2023
4  Entire home/apt    70              30                  2   07-10-2016

  reviews_per_month  calculated_host_listings_count  availability_365  \
0              0.23                               1               80
1              1.92                               1              196
2              0.26                               1              202
```

```
3              1.41                                1                149
4              0.02                                1                  0

   number_of_reviews_ltm             license
0                      2  Pending Application
1                     20         STR-0007772
2                      7         STR-0001599
3                      4                 NaN
4                      0                 NaN

[5 rows x 21 columns]
```

[4]:
```
#data cleaning
df['bedroom']=df['bedroom'].str.replace('bedroom','')
df['bedroom']=df['bedroom'].str.replace('Studio','0')
df['bedroom']=df['bedroom'].str.replace(' ','')
df['beds']=df['beds'].str.replace('bed','')
```

[5]:
```
#adding new cols
df['bathroom'] = df['bathroom'].astype(str)
print(df['bathroom'].isna().sum())

# Extract numerical values from 'bath' column
df['num_bathroom'] = df['bathroom'].str.extract('(\d+\.?\d*)', expand=False).
 ↪astype(float)

df.loc[df['bathroom'].str.contains('shared|half'), 'num_bathroom'] = 0
df['num_bathroom'] = df['num_bathroom'].fillna(0).astype(int)

# Add a new column 'is_private_bath' with 1 for private bath and 0 for shared/
 ↪half bath
df['is_private_bathroom'] = ~df['bathroom'].str.contains('shared|half',␣
 ↪case=False).astype(int)
df['is_private_bathroom'].replace(-2, 0, inplace=True)
df['is_private_bathroom'].replace(-1, 1, inplace=True)

df = df.drop('bathroom', axis=1)
```

```
0
```

[6]:
```
df.dtypes
```

[6]:
```
id                              float64
host_id                           int64
host_name                        object
name                             object
neighbourhood                    object
```

```
rating                              float64
bedroom                              object
beds                                 object
latitude                            float64
longitude                           float64
room_type                            object
price                                 int64
minimum_nights                        int64
number_of_reviews                     int64
last_review                          object
reviews_per_month                   float64
calculated_host_listings_count        int64
availability_365                      int64
number_of_reviews_ltm                 int64
license                              object
num_bathroom                          int32
is_private_bathroom                   int32
dtype: object
```

[7]:
```
#correcting data types
df['bedroom']=df['bedroom'].astype('int')
df['beds']=df['beds'].astype('int')
df.dtypes
```

[7]:
```
id                                  float64
host_id                               int64
host_name                            object
name                                 object
neighbourhood                        object
rating                              float64
bedroom                               int32
beds                                  int32
latitude                            float64
longitude                           float64
room_type                            object
price                                 int64
minimum_nights                        int64
number_of_reviews                     int64
last_review                          object
reviews_per_month                   float64
calculated_host_listings_count        int64
availability_365                      int64
number_of_reviews_ltm                 int64
license                              object
num_bathroom                          int32
is_private_bathroom                   int32
dtype: object
```

[8]:
```
df.columns
```

[8]:
```
Index(['id', 'host_id', 'host_name', 'name', 'neighbourhood', 'rating',
       'bedroom', 'beds', 'latitude', 'longitude', 'room_type', 'price',
       'minimum_nights', 'number_of_reviews', 'last_review',
       'reviews_per_month', 'calculated_host_listings_count',
       'availability_365', 'number_of_reviews_ltm', 'license', 'num_bathroom',
       'is_private_bathroom'],
      dtype='object')
```

[9]:
```
#missing values
df.isna().sum()
```

[9]:
```
id                                 0
host_id                            0
host_name                          1
name                               0
neighbourhood                      0
rating                             0
bedroom                            0
beds                               0
latitude                           0
longitude                          0
room_type                          0
price                              0
minimum_nights                     0
number_of_reviews                  0
last_review                     1884
reviews_per_month               1884
calculated_host_listings_count     0
availability_365                   0
number_of_reviews_ltm              0
license                         2965
num_bathroom                       0
is_private_bathroom                0
dtype: int64
```

[10]:
```
#handling missing values
df.last_review.fillna(method="ffill",inplace=True)
df.fillna({'reviews_per_month':0}, inplace=True)
df.fillna({'license':'Not Updated/No liscence'}, inplace=True)
df.dropna(inplace=True)
#examing changes
df.isna().sum()
```

[10]:
```
id                 0
host_id            0
```

```
host_name                          0
name                               0
neighbourhood                      0
rating                             0
bedroom                            0
beds                               0
latitude                           0
longitude                          0
room_type                          0
price                              0
minimum_nights                     0
number_of_reviews                  0
last_review                        0
reviews_per_month                  0
calculated_host_listings_count     0
availability_365                   0
number_of_reviews_ltm              0
license                            0
num_bathroom                       0
is_private_bathroom                0
dtype: int64
```

[11]:
```
#value counts for each col
index=['id', 'host_id', 'host_name', 'name', 'neighbourhood', 'rating',
       'bedroom', 'beds', 'latitude', 'longitude', 'room_type',
       'price', 'minimum_nights', 'number_of_reviews', 'reviews_per_month',
       'calculated_host_listings_count', 'availability_365',
       'number_of_reviews_ltm', 'license', 'num_bathroom',␣
 ↪'is_private_bathroom']

for i in index:

    print(df[i].value_counts(), "\n")
    print("-----------------------------------------------------------------")
```

```
id
8.086380e+17    11
7.840150e+17     9
8.086210e+17     9
8.086220e+17     7
7.840130e+17     7
                ..
2.810560e+07     1
2.808599e+07     1
2.808173e+07     1
2.807441e+07     1
1.037700e+18     1
Name: count, Length: 7718, dtype: int64
```

```
------------------------------------------------------------
host_id
542041520    249
107434423    164
4430421      155
520931919    148
173206762     61
            …
169948922      1
29535485       1
28190541       1
156365053      1
161769990      1
Name: count, Length: 3967, dtype: int64

------------------------------------------------------------
host_name
Allen         251
Blueground    164
Chris         158
Landmark      155
Michael       120
             …
Christiaan      1
Inés            1
Kameh           1
Gabrielle       1
Nazanina        1
Name: count, Length: 2153, dtype: int64

------------------------------------------------------------
name
Rental unit in San Francisco                  2650
Home in San Francisco                         1996
Condo in San Francisco                        1124
Hotel in San Francisco                         724
Guest suite in San Francisco                   563
Boutique hotel in San Francisco                283
Serviced apartment in San Francisco            174
Townhouse in San Francisco                     110
Guesthouse in San Francisco                     81
Hostel in San Francisco                         74
Loft in San Francisco                           70
Aparthotel in San Francisco                     65
Resort in San Francisco                         29
Bed and breakfast in San Francisco              25
Vacation home in San Francisco                  19
```

```
Cottage in San Francisco                          17
Place to stay in San Francisco                    14
Villa in San Francisco                            12
casa particular in San Francisco                   6
Bungalow in San Francisco                          5
Tiny home in San Francisco                         3
Rental unit in Tiburon                             3
Rental unit in San Francisco, Hayes Valley         1
Rental unit in Noe Valley - San Francisco          1
Floor in San Francisco                             1
Home in san francisco                              1
In-law in San Francisco                            1
Home in San Francisco                              1
cycladic house in San Francisco                    1
Castle in San Francisco                            1
Name: count, dtype: int64

------------------------------------------------------------
neighbourhood
Downtown/Civic Center    1010
Western Addition          690
Mission                   661
South of Market           558
Nob Hill                  488
Outer Sunset              360
Bernal Heights            332
Castro/Upper Market       312
Haight Ashbury            299
Noe Valley                262
Marina                    223
Inner Richmond            206
Outer Richmond            197
Financial District        193
Russian Hill              185
Bayview                   179
Pacific Heights           176
North Beach               174
Excelsior                 171
Potrero Hill              170
Inner Sunset              169
Parkside                  157
Chinatown                 139
Ocean View                128
Outer Mission             125
West of Twin Peaks        104
Visitacion Valley          84
Lakeshore                  67
Glen Park                  61
```

```
Twin Peaks                  60
Crocker Amazon              45
Presidio Heights            30
Diamond Heights             14
Presidio                    14
Seacliff                     9
Golden Gate Park             3
Name: count, dtype: int64

------------------------------------------------------------
rating
0.00    3037
5.00    1118
4.96     145
4.88     145
4.89     144
        …
3.93       1
3.92       1
3.63       1
4.01       1
2.33       1
Name: count, Length: 121, dtype: int64

------------------------------------------------------------
bedroom
1     4527
2     1508
0      976
3      733
4      214
5       52
6       26
7        8
8        5
10       2
9        2
11       1
15       1
Name: count, dtype: int64

------------------------------------------------------------
beds
1     4358
2     2092
3      857
4      396
5      126
```

```
0      103
6       69
7       23
8       17
10       5
9        5
12       3
11       1
Name: count, dtype: int64

------------------------------------------------------------
latitude
37.788990    30
37.792521    26
37.787257    24
37.787865    20
37.769420    19
             ..
37.735280     1
37.760450     1
37.734160     1
37.736310     1
37.775939     1
Name: count, Length: 5814, dtype: int64

------------------------------------------------------------
longitude
-122.412791    30
-122.410035    26
-122.396321    24
-122.419520    21
-122.410416    20
               ..
-122.480420     1
-122.417160     1
-122.418760     1
-122.496480     1
-122.467666     1
Name: count, Length: 5880, dtype: int64

------------------------------------------------------------
room_type
Entire home/apt    4981
Private room       2948
Shared room          68
Hotel room           58
Name: count, dtype: int64
```

```
------------------------------------------------------------
price
99      198
100     153
150     147
90      130
250     116
       …
454       1
2107      1
481       1
1093      1
431       1
Name: count, Length: 675, dtype: int64

------------------------------------------------------------
minimum_nights
30      2976
1       1844
2       1605
3        594
31       264
4        212
5        139
365      106
7         68
60        45
90        36
32        24
45        16
6         15
10        15
180       15
28        14
120       11
14        10
500        3
360        3
364        2
20         2
12         2
55         2
70         2
21         2
183        2
300        2
18         2
150        1
```

```
9           1
200         1
62          1
50          1
61          1
91          1
182         1
192         1
89          1
190         1
59          1
29          1
141         1
1000        1
1125        1
240         1
15          1
88          1
40          1
113         1
13          1
Name: count, dtype: int64

------------------------------------------------------------
number_of_reviews
0      1884
1       684
2       407
3       304
4       254
       …
696       1
626       1
345       1
705       1
590       1
Name: count, Length: 474, dtype: int64

------------------------------------------------------------
reviews_per_month
0.00    1884
0.06     135
0.04     106
0.02      99
0.07      97
        …
8.14       1
3.47       1
```

```
6.59       1
4.13       1
3.39       1
Name: count, Length: 674, dtype: int64

------------------------------------------------------------
calculated_host_listings_count
1      2939
2      1054
3       579
4       396
5       330
249     249
164     164
9       162
155     155
148     148
48      144
6       138
7       133
32       96
12       96
30       90
10       90
25       75
15       75
23       69
11       66
8        64
61       61
20       60
19       57
55       55
53       53
43       43
37       37
18       36
17       34
34       34
33       33
16       32
31       31
29       29
28       28
26       26
24       24
22       22
21       21
```

```
14       14
13       13
Name: count, dtype: int64

------------------------------------------------------------
availability_365
0      1428
365     533
364     327
269     305
179     196
       …
202       3
223       2
294       2
183       2
196       2
Name: count, Length: 364, dtype: int64

------------------------------------------------------------
number_of_reviews_ltm
0      3217
1       932
2       615
3       426
4       297
       …
118       1
106       1
99        1
96        1
133       1
Name: count, Length: 126, dtype: int64

------------------------------------------------------------
license
Not Updated/No liscence        2965
Exempt                         1101
License not needed per OSTR     266
City registration pending       172
pending                          70
                                …
2023-009670STR                     1
STR-0002439                        1
2022-009061STR                     1
STR-0002763                        1
2023-027852STR                     1
Name: count, Length: 2317, dtype: int64
```

```
------------------------------------------------------------
num_bathroom
1     5457
0     1202
2     1125
3      199
4       42
5       27
6        2
10       1
Name: count, dtype: int64

------------------------------------------------------------
is_private_bathroom
1    6928
0    1127
Name: count, dtype: int64

------------------------------------------------------------
```

[12]:
```
#stat summary
df.describe()
```

[12]:

	id	host_id	rating	bedroom	beds
count	8.055000e+03	8.055000e+03	8055.000000	8055.000000	8055.000000
mean	3.093048e+17	1.478846e+08	2.992223	1.387213	1.760397
std	4.133539e+17	1.780713e+08	2.336953	1.041436	1.175001
min	9.580000e+02	1.169000e+03	0.000000	0.000000	0.000000
25%	2.172707e+07	6.998119e+06	0.000000	1.000000	1.000000
50%	4.500046e+07	5.578278e+07	4.680000	1.000000	1.000000
75%	7.630695e+17	2.635022e+08	4.920000	2.000000	2.000000
max	1.037700e+18	5.490278e+08	5.000000	15.000000	12.000000

	latitude	longitude	price	minimum_nights
count	8055.000000	8055.000000	8055.000000	8055.000000
mean	37.769007	-122.430098	392.096710	20.612539
std	0.023130	0.027096	2182.891277	48.646240
min	37.708480	-122.512460	10.000000	1.000000
25%	37.753888	-122.442360	95.000000	2.000000
50%	37.774214	-122.422380	145.000000	3.000000
75%	37.787865	-122.411020	249.000000	30.000000
max	37.809810	-122.358480	50000.000000	1125.000000

	number_of_reviews	reviews_per_month	calculated_host_listings_count
count	8055.000000	8055.000000	8055.000000
mean	47.043824	1.002677	23.934078

```
std            98.824562           1.702522                      54.420559
min             0.000000           0.000000                       1.000000
25%             1.000000           0.025000                       1.000000
50%             7.000000           0.280000                       3.000000
75%            44.000000           1.220000                      12.000000
max          1134.000000          35.880000                     249.000000

       availability_365  number_of_reviews_ltm  num_bathroom  \
count       8055.000000            8055.000000   8055.000000
mean         177.621477               9.262570      1.071260
std          136.039850              19.335322      0.706764
min            0.000000               0.000000      0.000000
25%           33.000000               0.000000      1.000000
50%          176.000000               1.000000      1.000000
75%          317.500000               8.000000      1.000000
max          365.000000             426.000000     10.000000

       is_private_bathroom
count          8055.000000
mean              0.860087
std               0.346918
min               0.000000
25%               1.000000
50%               1.000000
75%               1.000000
max               1.000000
```

[13]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 8055 entries, 0 to 8055
Data columns (total 22 columns):
 #   Column                          Non-Null Count  Dtype
---  ------                          --------------  -----
 0   id                              8055 non-null   float64
 1   host_id                         8055 non-null   int64
 2   host_name                       8055 non-null   object
 3   name                            8055 non-null   object
 4   neighbourhood                   8055 non-null   object
 5   rating                          8055 non-null   float64
 6   bedroom                         8055 non-null   int32
 7   beds                            8055 non-null   int32
 8   latitude                        8055 non-null   float64
 9   longitude                       8055 non-null   float64
 10  room_type                       8055 non-null   object
 11  price                           8055 non-null   int64
 12  minimum_nights                  8055 non-null   int64
 13  number_of_reviews               8055 non-null   int64
```

```
 14  last_review                     8055 non-null   object
 15  reviews_per_month               8055 non-null   float64
 16  calculated_host_listings_count  8055 non-null   int64
 17  availability_365                8055 non-null   int64
 18  number_of_reviews_ltm           8055 non-null   int64
 19  license                         8055 non-null   object
 20  num_bathroom                    8055 non-null   int32
 21  is_private_bathroom             8055 non-null   int32
dtypes: float64(5), int32(4), int64(7), object(6)
memory usage: 1.3+ MB
```

[14]:
```
#analysis - top reviewed
top_reviewed_listings=df.nlargest(10,'number_of_reviews')
top_reviewed_listings
```

[14]:

	id	host_id	host_name	\
3118	35642179.0	265029065	Grant Plaza	
183	545685.0	2676602	Su	
192	585326.0	2676602	Su	
714	4464347.0	22931450	Sarah	
1041	8356380.0	44046204	Cheryl	
1449	14804950.0	24949158	Angela	
16	8739.0	7149	Ivan & Wendy	
1389	13845578.0	9194713	Lance	
1078	9051149.0	47224934	Elmer	
1169	10469182.0	34963239	Landy	

	name	neighbourhood	rating	\
3118	Boutique hotel in San Francisco	Financial District	4.25	
183	Guest suite in San Francisco	Outer Richmond	4.82	
192	Guest suite in San Francisco	Outer Richmond	4.80	
714	Bed and breakfast in San Francisco	North Beach	4.75	
1041	Guesthouse in San Francisco	Crocker Amazon	4.94	
1449	Guest suite in San Francisco	Bernal Heights	4.89	
16	Condo in San Francisco	Mission	4.92	
1389	Guest suite in San Francisco	Visitacion Valley	4.93	
1078	Rental unit in San Francisco	Western Addition	4.93	
1169	Home in San Francisco	Outer Richmond	4.92	

	bedroom	beds	latitude	longitude	…	minimum_nights	\
3118	1	1	37.79035	-122.40619	…	1	
183	1	1	37.77502	-122.48035	…	2	
192	1	2	37.77547	-122.48116	…	1	
714	1	8	37.79821	-122.40521	…	1	
1041	1	1	37.71329	-122.43633	…	1	
1449	0	1	37.74486	-122.41814	…	1	
16	1	1	37.76030	-122.42197	…	1	

```
1389        1     1  37.71907 -122.40276  …              1
1078        1     1  37.77133 -122.43612  …              1
1169        1     1  37.78028 -122.50252  …              2

      number_of_reviews  last_review  reviews_per_month  \
3118               1134   03-12-2023              21.65
183                1071   30-11-2023               7.69
192                 996   29-11-2023               7.18
714                 959   30-11-2023               8.70
1041                891   01-12-2023               9.16
1449                809   26-11-2023               9.32
16                  805   12-11-2023               4.61
1389                791   26-11-2023               8.79
1078                777   26-11-2023               7.88
1169                777   22-11-2023               8.17

      calculated_host_listings_count  availability_365  number_of_reviews_ltm  \
3118                               3                0                    300
183                                2              114                     93
192                                2                0                     99
714                                4              298                    354
1041                               1              164                    119
1449                               1               62                    113
16                                 2                0                     44
1389                               6                0                    134
1078                               1              301                     60
1169                               3              147                    114

                          license  num_bathroom is_private_bathroom
3118  License not needed per OSTR             1                   1
183                   STR-0004160             1                   1
192                   STR-0004160             1                   1
714                        933345             0                   0
1041                  STR-0000771             1                   1
1449                  STR-0002030             1                   1
16                    STR-0000028             1                   1
1389                  STR-0005709             1                   1
1078                  STR-0001444             0                   0
1169                  STR-0002581             1                   1

[10 rows x 22 columns]
```

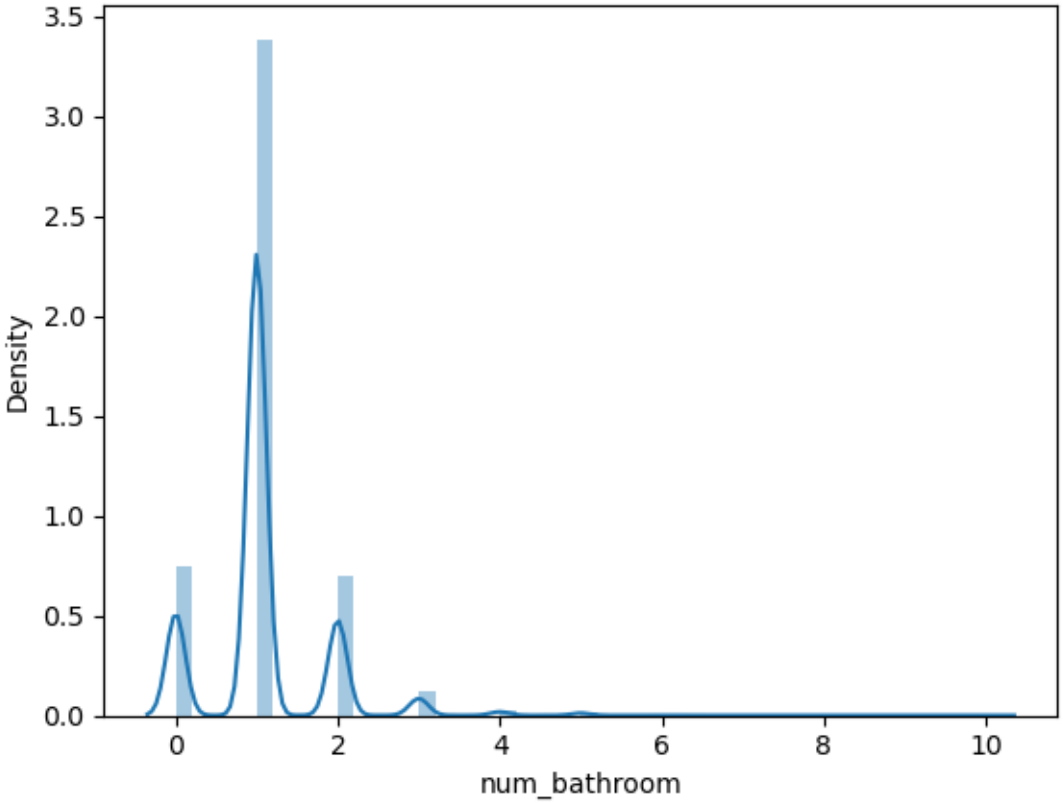
[15]:
```
#avg price
price_avrg=top_reviewed_listings.price.mean()
print('Average price per night: {}'.format(price_avrg))
```

```
Average price per night: 101.2
```

[16]:
```
#data vizualtion
sns.distplot(df["num_bathroom"])
```

```
C:\Users\DHESIKA\AppData\Local\Temp\ipykernel_31108\175651117.py:2: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(df["num_bathroom"])
```

[16]: <Axes: xlabel='num_bathroom', ylabel='Density'>

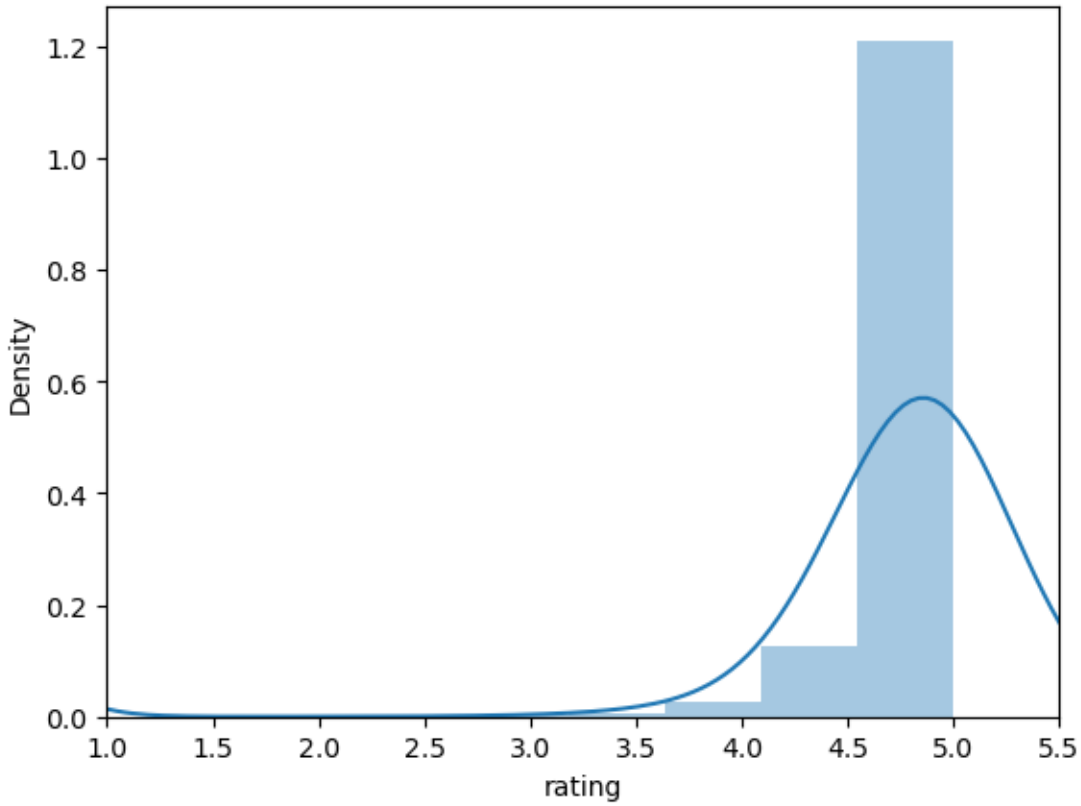


[17]:
```
sns.distplot(df["rating"])
plt.xlim(1,5.5)
```

```
C:\Users\DHESIKA\AppData\Local\Temp\ipykernel_31108\315059631.py:1: UserWarning:
```

```
`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(df["rating"])
```

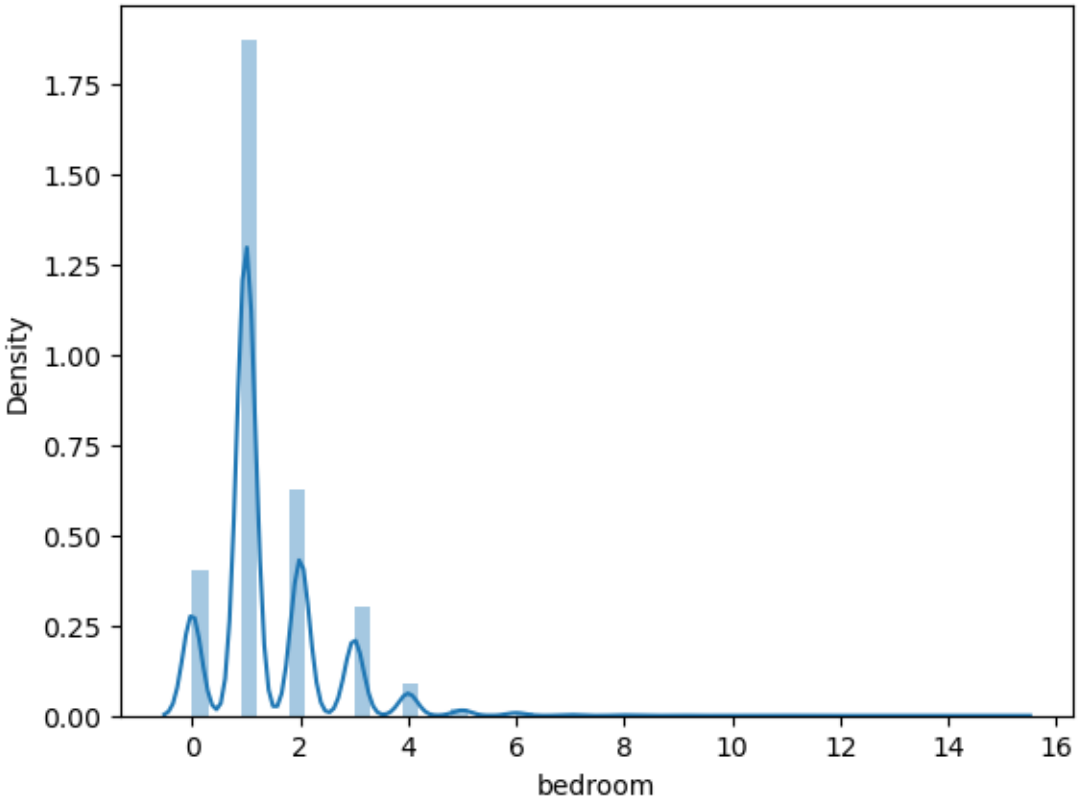
[17]: (1.0, 5.5)



[18]:
```
sns.distplot(df["bedroom"])
plt.show()
```

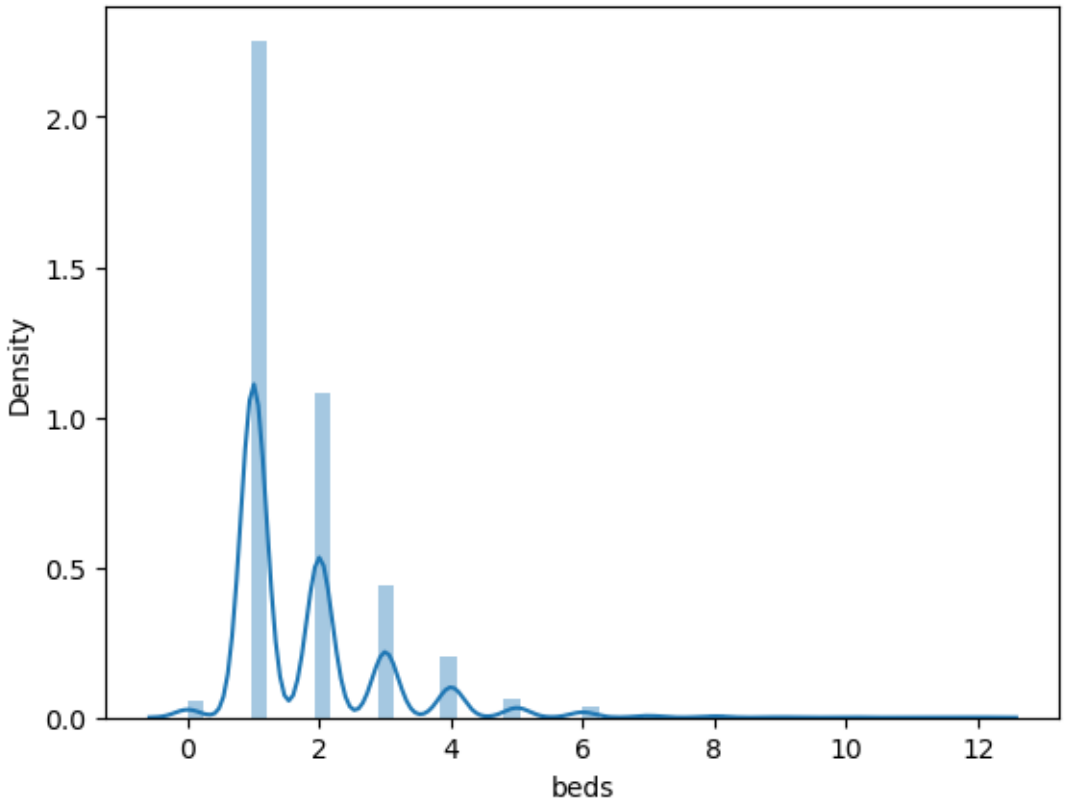
```
C:\Users\DHESIKA\AppData\Local\Temp\ipykernel_31108\4044275101.py:1:
UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).
```

```
For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(df["bedroom"])
```



[19]:
```
sns.distplot(df["beds"])
plt.show()
```

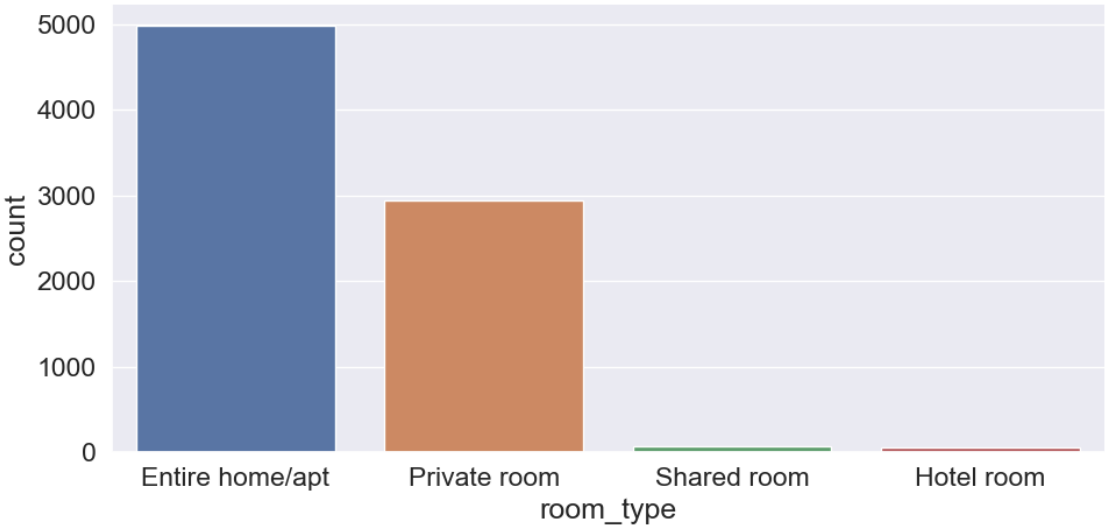
```
C:\Users\DHESIKA\AppData\Local\Temp\ipykernel_31108\3721278667.py:1:
UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(df["beds"])
```



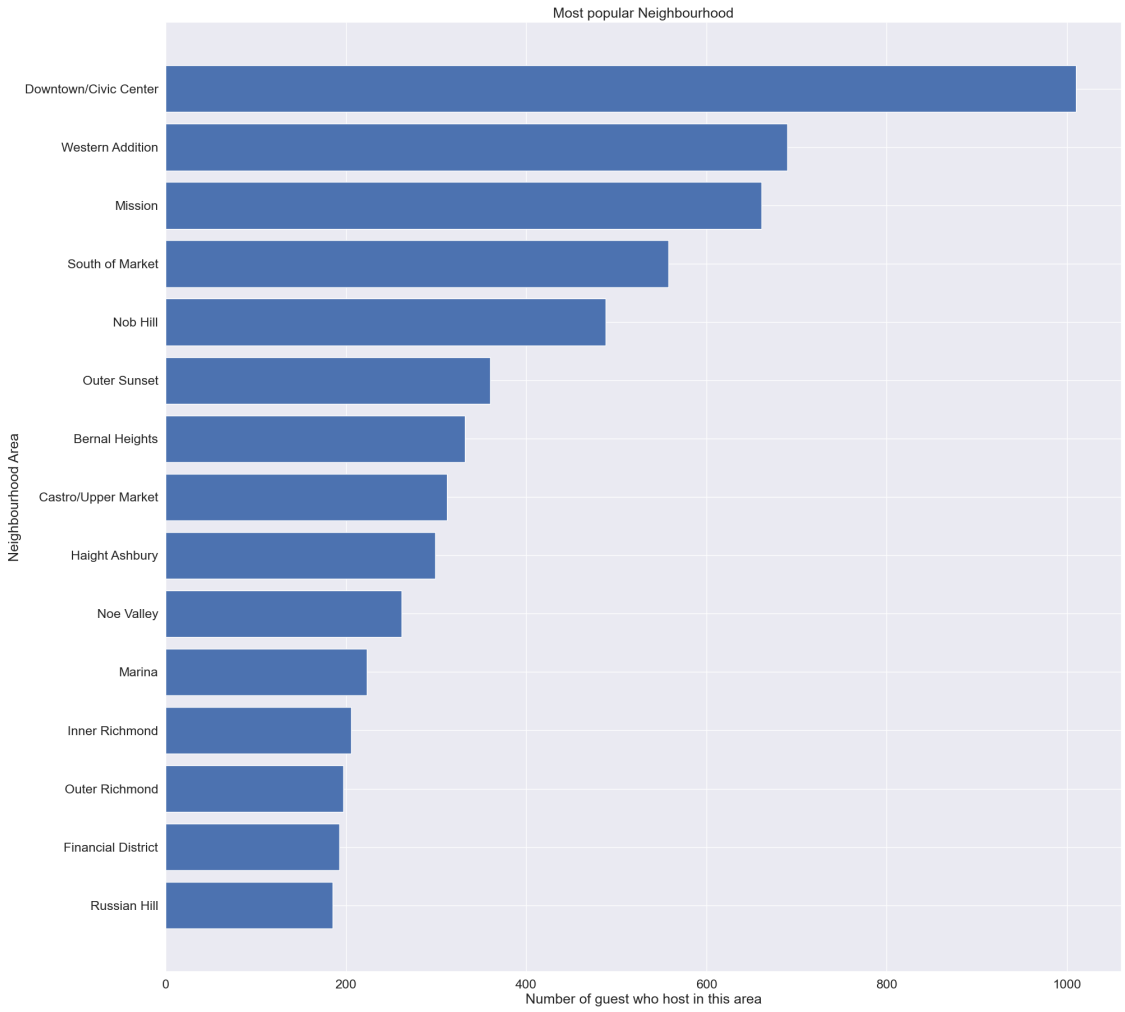
[20]:
```
def plot_catplot(h,v,he,a):
    sns.set(font_scale=1.5)
    sns.catplot(x=h,kind=v,data=df,height=he, aspect = a)

# Function to plot catplot graphs
def plot_piechart(h):
    sns.set(font_scale=1.5)
    fig = plt.figure(figsize=(5,5))
    ax = fig.add_axes([0,0,1,1])
    ax.axis('equal')
    langs = list(df[h].unique())
    students =list(df[h].value_counts())
    ax.pie(students, labels = langs,autopct='%1.2f%%')
    plt.show()
```

[21]:
```
plot_catplot("room_type", "count", 5, 2)
```

```
D:\Users\DHESIKA\anaconda3\Lib\site-packages\seaborn\axisgrid.py:118:
UserWarning: The figure layout has changed to tight
  self._figure.tight_layout(*args, **kwargs)
```



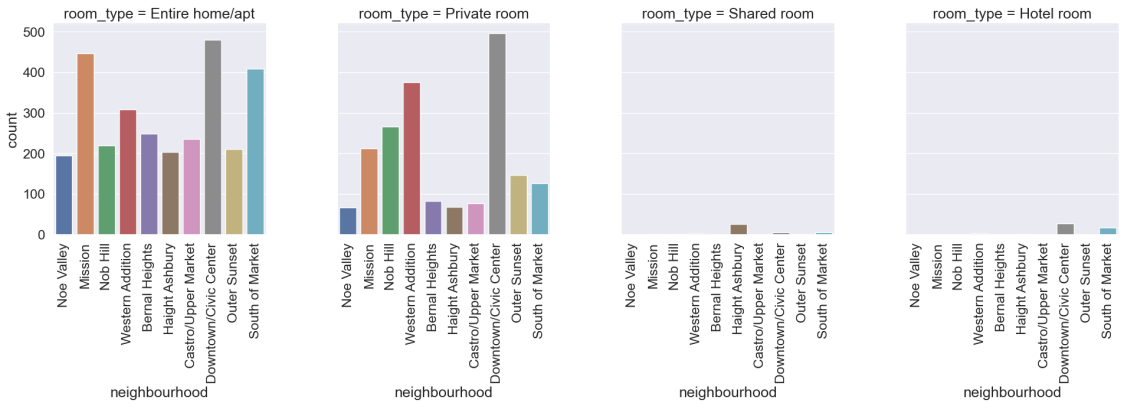
[22]:
```
data = df.neighbourhood.value_counts()[:15]
plt.figure(figsize=(22,22))
x = list(data.index)
y = list(data.values)
x.reverse()
y.reverse()

plt.title("Most popular Neighbourhood")
plt.ylabel("Neighbourhood Area")
plt.xlabel("Number of guest who host in this area")

plt.barh(x,y)
plt.show()
```



[23]:
```
sub_7=df.loc[df['neighbourhood'].isin(['Downtown/Civic Center','Western␣
 ↪Addition','Mission','South of Market', 'Nob Hill', 'Outer Sunset', 'Bernal␣
 ↪Heights', 'Castro/Upper Market','Haight Ashbury','Noe Valley'])]
#using catplot to represent multiple interesting attributes together and a count
viz_3=sns.catplot(x='neighbourhood', col='room_type', data=sub_7, kind='count')
viz_3.set_xticklabels(rotation=90)
```

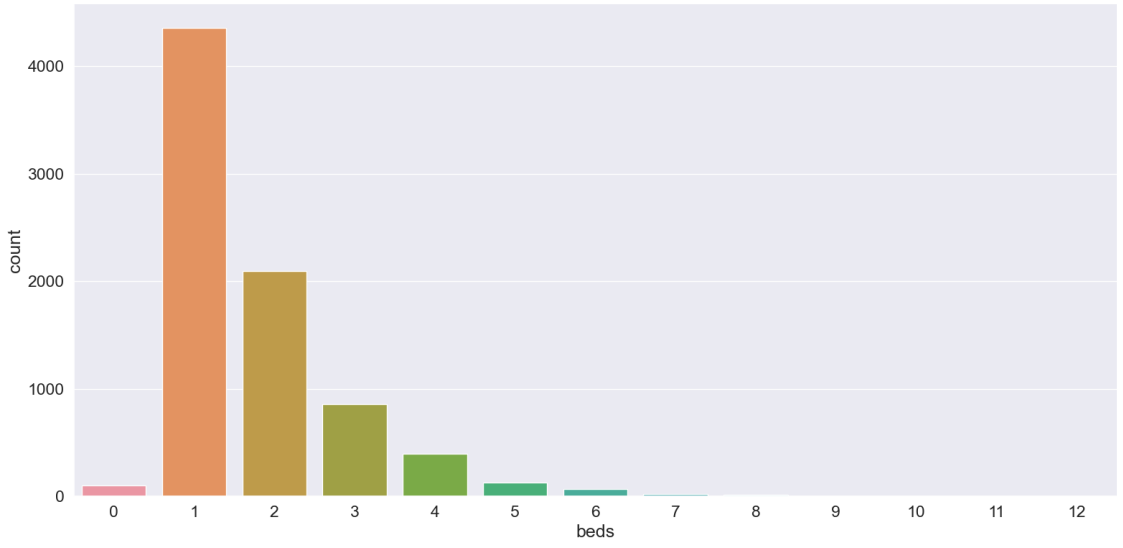
```
D:\Users\DHESIKA\anaconda3\Lib\site-packages\seaborn\axisgrid.py:118:
UserWarning: The figure layout has changed to tight
  self._figure.tight_layout(*args, **kwargs)
```

[23]:
```
<seaborn.axisgrid.FacetGrid at 0x1e272817dd0>
```



[24]:
```
plot_catplot("beds","count",8, 2)
```

```
D:\Users\DHESIKA\anaconda3\Lib\site-packages\seaborn\axisgrid.py:118:
UserWarning: The figure layout has changed to tight
  self._figure.tight_layout(*args, **kwargs)
```



[25]:
```
#top hosts
top_host=df['host_name'].value_counts().head(10)
top_host.head()
```

[25]:
```
host_name
Allen          251
Blueground     164
Chris          158
Landmark       155
```

```
Michael       120
Name: count, dtype: int64
```

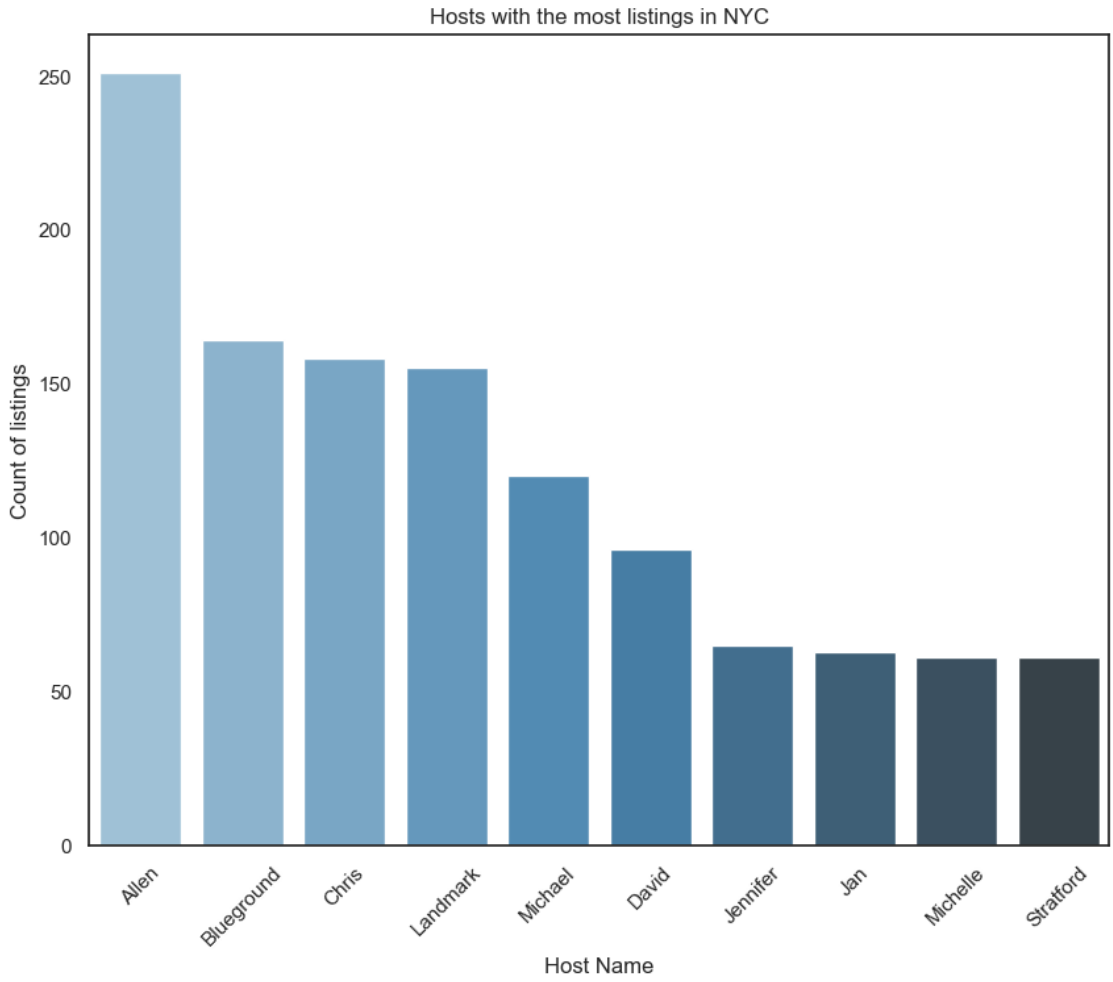
[26]:
```
top_host_df=pd.DataFrame(top_host)
top_host_df.reset_index(inplace=True)
top_host_df
```

[26]:

	host_name	count
0	Allen	251
1	Blueground	164
2	Chris	158
3	Landmark	155
4	Michael	120
5	David	96
6	Jennifer	65
7	Jan	63
8	Michelle	61
9	Stratford	61

[27]:
```
sns.set(rc={'figure.figsize':(10,8)})
sns.set_style('white')
viz_1=sns.barplot(x="host_name", y="count", data=top_host_df,
                  palette='Blues_d')
viz_1.set_title('Hosts with the most listings in NYC')
viz_1.set_ylabel('Count of listings')
viz_1.set_xlabel('Host Name')
viz_1.set_xticklabels(viz_1.get_xticklabels(), rotation=45)
```

[27]:
```
[Text(0, 0, 'Allen'),
 Text(1, 0, 'Blueground'),
 Text(2, 0, 'Chris'),
 Text(3, 0, 'Landmark'),
 Text(4, 0, 'Michael'),
 Text(5, 0, 'David'),
 Text(6, 0, 'Jennifer'),
 Text(7, 0, 'Jan'),
 Text(8, 0, 'Michelle'),
 Text(9, 0, 'Stratford')]
```



[28]:
```
#categorical data handling
categorical_col = []
for column in df.columns:

    if df[column].dtypes != "float64" and df[column].dtypes != "int32" and␣
 ↪df[column].dtypes != "int64":
        categorical_col.append(column)
categorical_col
```

[28]:
```
['host_name', 'name', 'neighbourhood', 'room_type', 'last_review', 'license']
```

[29]:
```
#heatmap for corr
from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
for col in categorical_col:
    df[col] = le.fit_transform(df[col])
```

```
    plt.figure(figsize = (40,40))
sns.heatmap(df.corr(), annot=True, fmt=".2f", cmap="seismic")
plt.show()
```

```
<Figure size 4000x4000 with 0 Axes>
```

```
<Figure size 4000x4000 with 0 Axes>
```

```
<Figure size 4000x4000 with 0 Axes>
```

```
<Figure size 4000x4000 with 0 Axes>
```

```
<Figure size 4000x4000 with 0 Axes>
```