

Importing Libraries

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

Reading DataSet

```
In [2]: airbnb = pd.read_csv("Airbnb NYC 2019.csv")
airbnb
```

Out[2]:

	id	name	host_id	host_name	neighbourhood_group	neighbourhood	latitude	longitude	room_type	pri
0	2539	Clean & quiet apt home by the park	2787	John	Brooklyn	Kensington	40.64749	-73.97237	Private room	1.
1	2595	Skylit Midtown Castle	2845	Jennifer	Manhattan	Midtown	40.75362	-73.98377	Entire home/apt	2
2	3647	THE VILLAGE OF HARLEM....NEW YORK !	4632	Elisabeth	Manhattan	Harlem	40.80902	-73.94190	Private room	1
3	3831	Cozy Entire Floor of Brownstone	4869	LisaRoxanne	Brooklyn	Clinton Hill	40.68514	-73.95976	Entire home/apt	
4	5022	Entire Apt: Spacious Studio/Loft by central park	7192	Laura	Manhattan	East Harlem	40.79851	-73.94399	Entire home/apt	
...	
48890	36484665	Charming one bedroom - newly renovated rowhouse	8232441	Sabrina	Brooklyn	Bedford-Stuyvesant	40.67853	-73.94995	Private room	
48891	36485057	Affordable room in Bushwick/East Williamsburg	6570630	Marisol	Brooklyn	Bushwick	40.70184	-73.93317	Private room	
48892	36485431	Sunny Studio at Historical Neighborhood	23492952	Ilgar & Aysel	Manhattan	Harlem	40.81475	-73.94867	Entire home/apt	1
48893	36485609	43rd St. Time Square-cozy single bed	30985759	Taz	Manhattan	Hell's Kitchen	40.75751	-73.99112	Shared room	
48894	36487245	Trendy duplex in the very heart of Hell's Kitchen	68119814	Christophe	Manhattan	Hell's Kitchen	40.76404	-73.98933	Private room	

48895 rows × 16 columns

Data Description & Information

In [3]: `airbnb.describe()`

Out[3]:

	id	host_id	latitude	longitude	price	minimum_nights	number_of_reviews	reviews_per_mor
count	4.889500e+04	4.889500e+04	48895.000000	48895.000000	48895.000000	48895.000000	48895.000000	38843.0000
mean	1.901714e+07	6.762001e+07	40.728949	-73.952170	152.720687	7.029962	23.274466	1.3732
std	1.098311e+07	7.861097e+07	0.054530	0.046157	240.154170	20.510550	44.550582	1.6804
min	2.539000e+03	2.438000e+03	40.499790	-74.244420	0.000000	1.000000	0.000000	0.0100
25%	9.471945e+06	7.822033e+06	40.690100	-73.983070	69.000000	1.000000	1.000000	0.1900
50%	1.967728e+07	3.079382e+07	40.723070	-73.955680	106.000000	3.000000	5.000000	0.7200
75%	2.915218e+07	1.074344e+08	40.763115	-73.936275	175.000000	5.000000	24.000000	2.0200
max	3.648724e+07	2.743213e+08	40.913060	-73.712990	10000.000000	1250.000000	629.000000	58.5000

In [4]: `airbnb.isnull().sum()` *#Checked Missing values*

Out[4]:

id	0
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
number_of_reviews	0
last_review	10052
reviews_per_month	10052
calculated_host_listings_count	0
availability_365	0

dtype: int64

In [5]: `airbnb.columns`

Out[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')

Dropping the unnecessary columns

In [6]: `airbnb.drop(['latitude', 'longitude', 'last_review', 'reviews_per_month'], axis=1, inplace=True)`

In [7]:

airbnb.head(10)

Out[7]:

	id	name	host_id	host_name	neighbourhood_group	neighbourhood	room_type	price	minimum_nights	number_of_reviews
0	2539	Clean & quiet apt home by the park	2787	John	Brooklyn	Kensington	Private room	149	1	
1	2595	Skylit Midtown Castle	2845	Jennifer	Manhattan	Midtown	Entire home/apt	225	1	
2	3647	THE VILLAGE OF HARLEM....NEW YORK !	4632	Elisabeth	Manhattan	Harlem	Private room	150	3	
3	3831	Cozy Entire Floor of Brownstone	4869	LisaRoxanne	Brooklyn	Clinton Hill	Entire home/apt	89	1	
4	5022	Entire Apt: Spacious Studio/Loft by central park	7192	Laura	Manhattan	East Harlem	Entire home/apt	80	10	
5	5099	Large Cozy 1 BR Apartment In Midtown East	7322	Chris	Manhattan	Murray Hill	Entire home/apt	200	3	
6	5121	BlissArtsSpace!	7356	Garon	Brooklyn	Bedford-Stuyvesant	Private room	60	45	
7	5178	Large Furnished Room Near B'way	8967	Shunichi	Manhattan	Hell's Kitchen	Private room	79	2	
8	5203	Cozy Clean Guest Room - Family Apt	7490	MaryEllen	Manhattan	Upper West Side	Private room	79	2	
9	5238	Cute & Cozy Lower East Side 1 bdrm	7549	Ben	Manhattan	Chinatown	Entire home/apt	150	1	

1. What can we learn about different hosts and areas?

In [8]:

host_areas =airbnb.groupby(['host_name','neighbourhood_group'])['calculated_host_listings_count'].max().reset_index()
host_areas.sort_values(by='calculated_host_listings_count',ascending=False).head(5)

Out[8]:

	host_name	neighbourhood_group	calculated_host_listings_count
13217	Sonder (NYC)	Manhattan	327
1834	Blueground	Manhattan	232
1833	Blueground	Brooklyn	232
7275	Kara	Manhattan	121
7480	Kazuya	Queens	103

2.What we learn from room type and their prices according to area?
We find that Host name Sonder(NYC) has listed highest number of listings in Manhattan followed by Blueground

```
In [10]: room_price_area_wise = airbnb.groupby(['neighbourhood_group', 'room_type'])['price'].max().reset_index()
room_price_area_wise.sort_values(by='price', ascending=False).head(10)
```

Out[10]:

	neighbourhood_group	room_type	price
3	Brooklyn	Entire home/apt	10000
6	Manhattan	Entire home/apt	10000
10	Queens	Private room	10000
7	Manhattan	Private room	9999
4	Brooklyn	Private room	7500
12	Staten Island	Entire home/apt	5000
9	Queens	Entire home/apt	2600
1	Bronx	Private room	2500
11	Queens	Shared room	1800
0	Bronx	Entire home/apt	1000

We found that Entire home/apt is the highest number of room types overall and prices are high in the brooklyn and Manhattan for entire home/apt.
3.What can we learn from Data?

```
In [11]: area_reviews = airbnb.groupby(['neighbourhood_group'])['number_of_reviews'].max().reset_index()
area_reviews
```

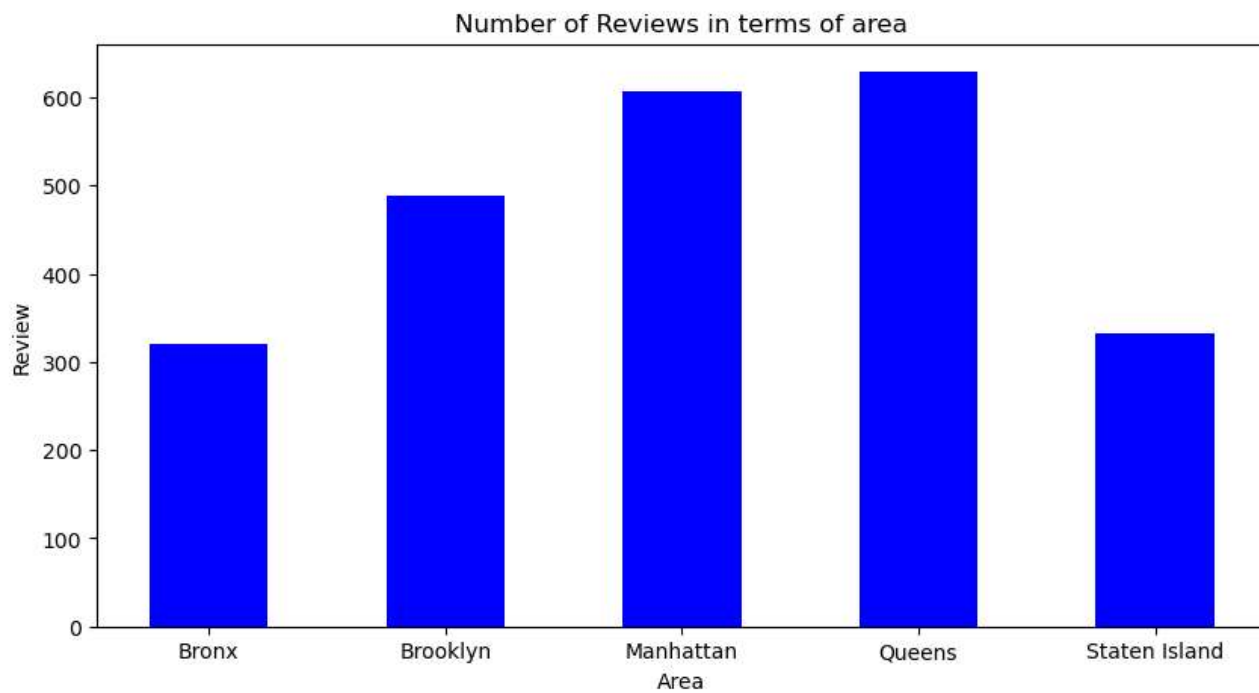
Out[11]:

	neighbourhood_group	number_of_reviews
0	Bronx	321
1	Brooklyn	488
2	Manhattan	607
3	Queens	629
4	Staten Island	333

Visualize the data

```
In [12]: area = area_reviews['neighbourhood_group']
review = area_reviews['number_of_reviews']
fig = plt.figure(figsize =(10,5))

plt.bar(area, review, color ="blue", width =0.5)
plt.xlabel('Area')
plt.ylabel('Review')
plt.title("Number of Reviews in terms of area")
plt.show()
```



```
In [13]: price_area = airbnb.groupby(['price'])['number_of_reviews'].max().reset_index()
price_area.head(10)
```

Out[13]:

	price	number_of_reviews
0	0	95
1	10	93
2	11	113
3	12	8
4	13	9
5	15	19
6	16	43
7	18	1
8	19	76
9	20	116

```
In [14]: price_list = price_area['price']  
review = price_area['number_of_reviews']  
fig = plt.figure(figsize =(10,5))  
  
plt.scatter(price_list, review)  
plt.xlabel('Price')  
plt.ylabel('Number of reviews')  
plt.title('Number of Reviews VS Price')  
plt.show()
```

