

```
In [1]: import pandas as pd
import numpy as np
import re
import plotly.express as px
import matplotlib.pyplot as plt
import seaborn as sns
```

```
In [2]: import warnings
warnings.filterwarnings("ignore")
```

```
In [3]: airbnb_df = pd.read_csv(r'C:\Users\Arindham Krishna\OneDrive\Desktop\Projects for Portfolio\Data
```

```
In [4]: airbnb_df.head()
```

Out[4]:

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

```
In [5]: 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	id	48895 non-null	int64
1	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	number_of_reviews	48895 non-null	int64
12	last_review	38843 non-null	object
13	reviews_per_month	38843 non-null	float64
14	calculated_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
```

```
In [6]: airbnb_df.isnull().sum()
```

```
Out[6]: 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
```

We can see that we have got empty values in Name, Host name, last review and reviews per month. Lets get in deeper

```
In [7]: airbnb_df[airbnb_df["name"].isnull()]
```

Out[7]:

	id	name	host_id	host_name	neighbourhood_group	neighbourhood	latitude	longitude	room	
	2854	1615764	NaN	6676776	Peter	Manhattan	Battery Park City	40.71239	-74.01620	horr
	3703	2232600	NaN	11395220	Anna	Manhattan	East Village	40.73215	-73.98821	horr
	5775	4209595	NaN	20700823	Jesse	Manhattan	Greenwich Village	40.73473	-73.99244	horr
	5975	4370230	NaN	22686810	Michaël	Manhattan	Nolita	40.72046	-73.99550	horr
	6269	4581788	NaN	21600904	Lucie	Brooklyn	Williamsburg	40.71370	-73.94378	F
	6567	4756856	NaN	1832442	Carolina	Brooklyn	Bushwick	40.70046	-73.92825	F
	6605	4774658	NaN	24625694	Josh	Manhattan	Washington Heights	40.85198	-73.93108	F
	8841	6782407	NaN	31147528	Huei-Yin	Brooklyn	Williamsburg	40.71354	-73.93882	F
	11963	9325951	NaN	33377685	Jonathan	Manhattan	Hell's Kitchen	40.76436	-73.98573	horr
	12824	9787590	NaN	50448556	Miguel	Manhattan	Harlem	40.80316	-73.95189	horr
	13059	9885866	NaN	37306329	Juliette	Manhattan	Chinatown	40.71632	-73.99328	F
	13401	10052289	NaN	49522403	Vanessa	Brooklyn	Brownsville	40.66409	-73.92314	F
	15819	12797684	NaN	69715276	Yan	Manhattan	Upper West Side	40.79843	-73.96404	F
	16071	12988898	NaN	71552588	Andrea	Bronx	Fordham	40.86032	-73.88493	S
	18047	14135050	NaN	85288337	Jeff	Brooklyn	Bedford-Stuyvesant	40.69421	-73.93234	F
	28889	22275821	NaN	49662398	Kathleen	Brooklyn	Bushwick	40.69546	-73.92741	horr

We can see that the missing values in the bnb names column also seems to have majority of 0 number of reviews and also last review date is missing. Clearly states the possibility that the specific airbnbs have not got any reviews till date.

Moving forward, we can go ahead and drop the airbnbs with missing names with 0 reviews recieved and we will only keep which have got 1 or more reviews and then we will replace the missing airbnb names with unknown.

```
In [8]: list_columns = ['name', 'last_review']
airbnb_df.dropna(subset=list_columns, inplace=True, how='all')
```

```
In [9]: airbnb_df[airbnb_df["name"].isnull()]
```

Out[9]:		id	name	host_id	host_name	neighbourhood_group	neighbourhood	latitude	longitude	room
	3703	2232600	NaN	11395220	Anna	Manhattan	East Village	40.73215	-73.98821	hor
	5775	4209595	NaN	20700823	Jesse	Manhattan	Greenwich Village	40.73473	-73.99244	hor
	5975	4370230	NaN	22686810	Michaël	Manhattan	Nolita	40.72046	-73.99550	hor
	11963	9325951	NaN	33377685	Jonathan	Manhattan	Hell's Kitchen	40.76436	-73.98573	hor
	13401	10052289	NaN	49522403	Vanessa	Brooklyn	Brownsville	40.66409	-73.92314	F
	28889	22275821	NaN	49662398	Kathleen	Brooklyn	Bushwick	40.69546	-73.92741	hor

Now, we can see that in the above frame that out of 16 missing values we are left with 6 missing values. Reason for keeping these 6 missing values is that they do have received reviews from the guests. Let's replace these missing value with value 'Unknown'

```
In [10]: airbnb_df["name"].fillna('Unknown', inplace = True)
```

```
In [11]: airbnb_df[airbnb_df["name"] == 'Unknown']
```

Out[11]:		id	name	host_id	host_name	neighbourhood_group	neighbourhood	latitude	longitude	rc
	3703	2232600	Unknown	11395220	Anna	Manhattan	East Village	40.73215	-73.98821	
	5775	4209595	Unknown	20700823	Jesse	Manhattan	Greenwich Village	40.73473	-73.99244	
	5975	4370230	Unknown	22686810	Michaël	Manhattan	Nolita	40.72046	-73.99550	
	11963	9325951	Unknown	33377685	Jonathan	Manhattan	Hell's Kitchen	40.76436	-73.98573	
	13401	10052289	Unknown	49522403	Vanessa	Brooklyn	Brownsville	40.66409	-73.92314	
	28889	22275821	Unknown	49662398	Kathleen	Brooklyn	Bushwick	40.69546	-73.92741	

```
In [12]: airbnb_df.isnull().sum()
```

```
Out[12]: id          0
         name        0
         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   10042
         reviews_per_month  10042
         calculated_host_listings_count  0
         availability_365  0
         dtype: int64
```

1. We can see now that missing names of airbnb have been replaced from NaN to 'Unkown'
2. Also, we can see that there are 0 missing values in the name columns.

```
In [13]: airbnb_df[airbnb_df["host_name"].isnull()]
```

Out[13]:

	id	name	host_id	host_name	neighbourhood_group	neighbourhood	latitude	longitude
360	100184	Bienvenue	526653	NaN	Queens	Queens Village	40.72413	-73.76133
2700	1449546	Cozy Studio in Flatbush	7779204	NaN	Brooklyn	Flatbush	40.64965	-73.96154
5745	4183989	SPRING in the City!! Zen-Style Tranquil Bedroom	919218	NaN	Manhattan	Harlem	40.80606	-73.95067
6075	4446862	Charming Room in Prospect Heights!	23077718	NaN	Brooklyn	Crown Heights	40.67512	-73.96146
6582	4763327	Luxurious, best location, spa inc'l	24576978	NaN	Brooklyn	Greenpoint	40.72035	-73.95359
8163	6292866	Modern Quiet Gem Near All	32722063	NaN	Brooklyn	East Flatbush	40.65263	-73.93219
8257	6360224	Sunny, Private room in Bushwick	33134899	NaN	Brooklyn	Bushwick	40.70146	-73.92792
8852	6786181	R&S Modern Spacious Hideaway	32722063	NaN	Brooklyn	East Flatbush	40.64345	-73.93643
9138	6992973	1 Bedroom in Prime Williamsburg	5162530	NaN	Brooklyn	Williamsburg	40.71838	-73.95630
9817	7556587	Sunny Room in Harlem	39608626	NaN	Manhattan	Harlem	40.82929	-73.94182
14040	10709846	Sunny, spacious room in Greenpoint	7822683	NaN	Brooklyn	Greenpoint	40.73539	-73.95838
14631	11553543	Cozy Room Astoria	26138712	NaN	Queens	Ditmars Steinway	40.77587	-73.91779
15174	12113879	Sunny, Large West Village 1 BR Near Everything	5300585	NaN	Manhattan	Chelsea	40.73949	-73.99807
19565	15648096	Spacious 2 bedroom close to Manhattan	100971588	NaN	Bronx	Highbridge	40.83844	-73.92489
27777	21787094	Spacious and Bright Midtown East Apartment	415290	NaN	Manhattan	Upper East Side	40.76436	-73.96836

	id	name	host_id	host_name	neighbourhood_group	neighbourhood	latitude	longitude
27962	21867710	Amazing room in the centre on Manhattan! Welcome!	159156636	NaN	Manhattan	Hell's Kitchen	40.75668	-73.99097
28042	21901156	Best Deal! Lovely place in Manhattan! Time Squ...	159156636	NaN	Manhattan	Hell's Kitchen	40.75656	-73.99063
28274	21990808	***SLEEPING IN THE GOLDEN PLACE***	159156636	NaN	Manhattan	Hell's Kitchen	40.75835	-73.99063
30570	23660605	Private Bedroom w/ Balcony view of Central Park	177146433	NaN	Manhattan	East Harlem	40.79766	-73.94824
32193	25180572	Beautiful Cozy Studio Apt. UWS	119609345	NaN	Manhattan	Upper West Side	40.77701	-73.97663
38992	30468335	LARGE Prospect Lefferts Gardens Bedroom	228750026	NaN	Brooklyn	Flatbush	40.65152	-73.95277

Again even in the host name column we have missing values, let's drop the missing host names with 0 number of reviews and missing last review.

```
In [14]: list_columns_names = ['host_name', 'last_review']
airbnb_df.dropna(subset=list_columns_names, inplace=True, how='all')
```

```
In [15]: airbnb_df[airbnb_df["host_name"].isnull()]
```

Out[15]:

	id	name	host_id	host_name	neighbourhood_group	neighbourhood	latitude	longitude
360	100184	Bienvenue	526653	NaN	Queens	Queens Village	40.72413	-73.76133
2700	1449546	Cozy Studio in Flatbush	7779204	NaN	Brooklyn	Flatbush	40.64965	-73.96154
5745	4183989	SPRING in the City!! Zen-Style Tranquil Bedroom	919218	NaN	Manhattan	Harlem	40.80606	-73.95061
6582	4763327	Luxurious, best location, spa inc'l	24576978	NaN	Brooklyn	Greenpoint	40.72035	-73.95355
8163	6292866	Modern Quiet Gem Near All	32722063	NaN	Brooklyn	East Flatbush	40.65263	-73.93215
8257	6360224	Sunny, Private room in Bushwick	33134899	NaN	Brooklyn	Bushwick	40.70146	-73.92792
8852	6786181	R&S Modern Spacious Hideaway	32722063	NaN	Brooklyn	East Flatbush	40.64345	-73.93643
9817	7556587	Sunny Room in Harlem	39608626	NaN	Manhattan	Harlem	40.82929	-73.94182
14040	10709846	Sunny, spacious room in Greenpoint	7822683	NaN	Brooklyn	Greenpoint	40.73539	-73.95838
14631	11553543	Cozy Room Astoria	26138712	NaN	Queens	Ditmars Steinway	40.77587	-73.91775
15174	12113879	Sunny, Large West Village 1 BR Near Everything	5300585	NaN	Manhattan	Chelsea	40.73949	-73.99801
19565	15648096	Spacious 2 bedroom close to Manhattan	100971588	NaN	Bronx	Highbridge	40.83844	-73.92489
27962	21867710	Amazing room in the centre on Manhatten! Welcome!	159156636	NaN	Manhattan	Hell's Kitchen	40.75668	-73.99097
28042	21901156	Best Deal! Lovely place in Manhattan! Time Squ...	159156636	NaN	Manhattan	Hell's Kitchen	40.75656	-73.99063

	id	name	host_id	host_name	neighbourhood_group	neighbourhood	latitude	longitude
28274	21990808	***SLEEPING IN THE GOLDEN PLACE***	159156636	NaN	Manhattan	Hell's Kitchen	40.75835	-73.99065
30570	23660605	Private Bedroom w/ Balcony view of Central Park	177146433	NaN	Manhattan	East Harlem	40.79766	-73.94824

```
In [16]: airbnb_df["host_name"].fillna('Unknown', inplace = True)

In [17]: airbnb_df[airbnb_df["host_name"] == 'Unknown']
```

Out[17]:

	id	name	host_id	host_name	neighbourhood_group	neighbourhood	latitude	longitude
360	100184	Bienvenue	526653	Unknown	Queens	Queens Village	40.72413	-73.76133
2700	1449546	Cozy Studio in Flatbush	7779204	Unknown	Brooklyn	Flatbush	40.64965	-73.96154
5745	4183989	SPRING in the City!! Zen-Style Tranquil Bedroom	919218	Unknown	Manhattan	Harlem	40.80606	-73.95061
6582	4763327	Luxurious, best location, spa inc'l	24576978	Unknown	Brooklyn	Greenpoint	40.72035	-73.95355
8163	6292866	Modern Quiet Gem Near All	32722063	Unknown	Brooklyn	East Flatbush	40.65263	-73.93215
8257	6360224	Sunny, Private room in Bushwick	33134899	Unknown	Brooklyn	Bushwick	40.70146	-73.92792
8852	6786181	R&S Modern Spacious Hideaway	32722063	Unknown	Brooklyn	East Flatbush	40.64345	-73.93643
9817	7556587	Sunny Room in Harlem	39608626	Unknown	Manhattan	Harlem	40.82929	-73.94182
14040	10709846	Sunny, spacious room in Greenpoint	7822683	Unknown	Brooklyn	Greenpoint	40.73539	-73.95838
14631	11553543	Cozy Room Astoria	26138712	Unknown	Queens	Ditmars Steinway	40.77587	-73.91775
15174	12113879	Sunny, Large West Village 1 BR Near Everything	5300585	Unknown	Manhattan	Chelsea	40.73949	-73.99801
19565	15648096	Spacious 2 bedroom close to Manhattan	100971588	Unknown	Bronx	Highbridge	40.83844	-73.92489
27962	21867710	Amazing room in the centre on Manhatten! Welcome!	159156636	Unknown	Manhattan	Hell's Kitchen	40.75668	-73.99097
28042	21901156	Best Deal! Lovely place in Manhattan! Time Squ...	159156636	Unknown	Manhattan	Hell's Kitchen	40.75656	-73.99063

	id	name	host_id	host_name	neighbourhood_group	neighbourhood	latitude	longitude
28274	21990808	***SLEEPING IN THE GOLDEN PLACE***	159156636	Unknown	Manhattan	Hell's Kitchen	40.75835	-73.99065
30570	23660605	Private Bedroom w/ Balcony view of Central Park	177146433	Unknown	Manhattan	East Harlem	40.79766	-73.94824

All the missing host names have been updated with Unknown, however these were not removed as these hotels have recieved reviews.

In [18]: `airbnb_df.info()`

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 48880 entries, 0 to 48894
Data columns (total 16 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   id                                     48880 non-null  int64
1   name                                  48880 non-null  object
2   host_id                               48880 non-null  int64
3   host_name                             48880 non-null  object
4   neighbourhood_group                    48880 non-null  object
5   neighbourhood                          48880 non-null  object
6   latitude                               48880 non-null  float64
7   longitude                              48880 non-null  float64
8   room_type                             48880 non-null  object
9   price                                  48880 non-null  int64
10  minimum_nights                         48880 non-null  int64
11  number_of_reviews                      48880 non-null  int64
12  last_review                            38843 non-null  object
13  reviews_per_month                     38843 non-null  float64
14  calculated_host_listings_count         48880 non-null  int64
15  availability_365                       48880 non-null  int64
dtypes: float64(3), int64(7), object(6)
memory usage: 6.3+ MB
```

In [19]: `airbnb_df.isnull().sum()`

```
Out[19]: id                0
name                    0
host_id                 0
host_name               0
neighbourhood_group     0
neighbourhood           0
latitude                0
longitude               0
room_type               0
price                   0
minimum_nights          0
number_of_reviews       0
last_review             10037
reviews_per_month       10037
calculated_host_listings_count  0
availability_365        0
dtype: int64
```

```
In [20]: airbnb_df[airbnb_df["last_review"].isnull()]
```

Out[20]:

	id	name	host_id	host_name	neighbourhood_group	neighbourhood	latitude	longitud
2	3647	THE VILLAGE OF HARLEM....NEW YORK !	4632	Elisabeth	Manhattan	Harlem	40.80902	-73.9415
19	7750	Huge 2 BR Upper East Cental Park	17985	Sing	Manhattan	East Harlem	40.79685	-73.9487
26	8700	Magnifique Suite au N de Manhattan - vue Cloitres	26394	Claude & Sophie	Manhattan	Inwood	40.86754	-73.9265
36	11452	Clean and Quiet in Brooklyn	7355	Vt	Brooklyn	Bedford-Stuyvesant	40.68876	-73.9437
38	11943	Country space in the city	45445	Harriet	Brooklyn	Flatbush	40.63702	-73.9632
...
48890	36484665	Charming one bedroom - newly renovated rowhouse	8232441	Sabrina	Brooklyn	Bedford-Stuyvesant	40.67853	-73.9495
48891	36485057	Affordable room in Bushwick/East Williamsburg	6570630	Marisol	Brooklyn	Bushwick	40.70184	-73.9337
48892	36485431	Sunny Studio at Historical Neighborhood	23492952	Ilgar & Aysel	Manhattan	Harlem	40.81475	-73.9486
48893	36485609	43rd St. Time Square-cozy single bed	30985759	Taz	Manhattan	Hell's Kitchen	40.75751	-73.9917
48894	36487245	Trendy duplex in the very heart of Hell's Kitchen	68119814	Christophe	Manhattan	Hell's Kitchen	40.76404	-73.9893

10037 rows × 16 columns

last_review and reviews per month both are empty and basically when the number of reviews are 0 then automatically this will be null, so lets just drop all the null values

```
In [21]: airbnb_df = airbnb_df.dropna()
```

```
In [22]: airbnb_df.isnull().sum()
```

```
Out[22]: id          0
         name        0
         host_id     0
         host_name    0
         neighbourhood_group  0
         neighbourhood  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
         dtype: int64
```

Now that we have cleaned the data, lets do the visualization.

```
In [23]: len(airbnb_df.host_name.unique())
```

```
Out[23]: 9887
```

```
In [24]: top_hostnames = airbnb_df.host_name.value_counts().head(10)
         print(top_hostnames)
```

```
Michael      335
David        309
John         250
Alex         229
Sonder (NYC) 207
Sarah        179
Maria        174
Jessica      170
Daniel       170
Anna         160
Name: host_name, dtype: int64
```

```
In [25]: top_hosts = airbnb_df.host_id.value_counts().head(10)
         print(top_hosts)
```

```
219517861    207
61391963      79
16098958      61
137358866     51
7503643       49
190921808     46
30283594      43
1475015       42
120762452     40
2119276       39
Name: host_id, dtype: int64
```

```
In [26]: top_places = airbnb_df.neighbourhood.value_counts().head(10)
         print(top_places)
```

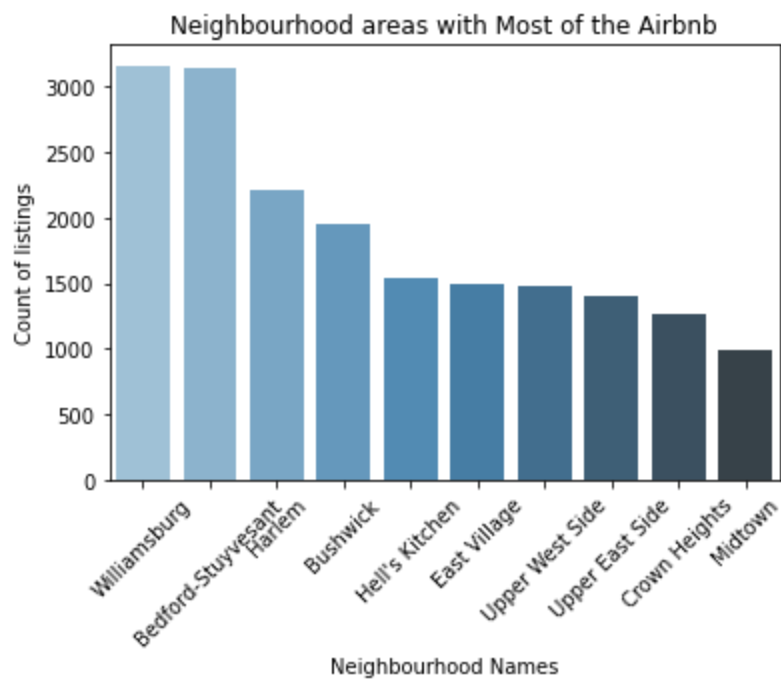
```
Williamsburg      3163
Bedford-Stuyvesant 3141
Harlem            2206
Bushwick          1944
Hell's Kitchen    1532
East Village      1490
Upper West Side   1482
Upper East Side   1405
Crown Heights     1265
Midtown           986
Name: neighbourhood, dtype: int64
```

```
In [27]: top_places_df = pd.DataFrame(top_places)
top_places_df.reset_index(inplace=True)
top_places_df.rename(columns={'index': 'Neighborhood Name', 'neighbourhood': 'count'}, inplace=True)
top_places_df
```

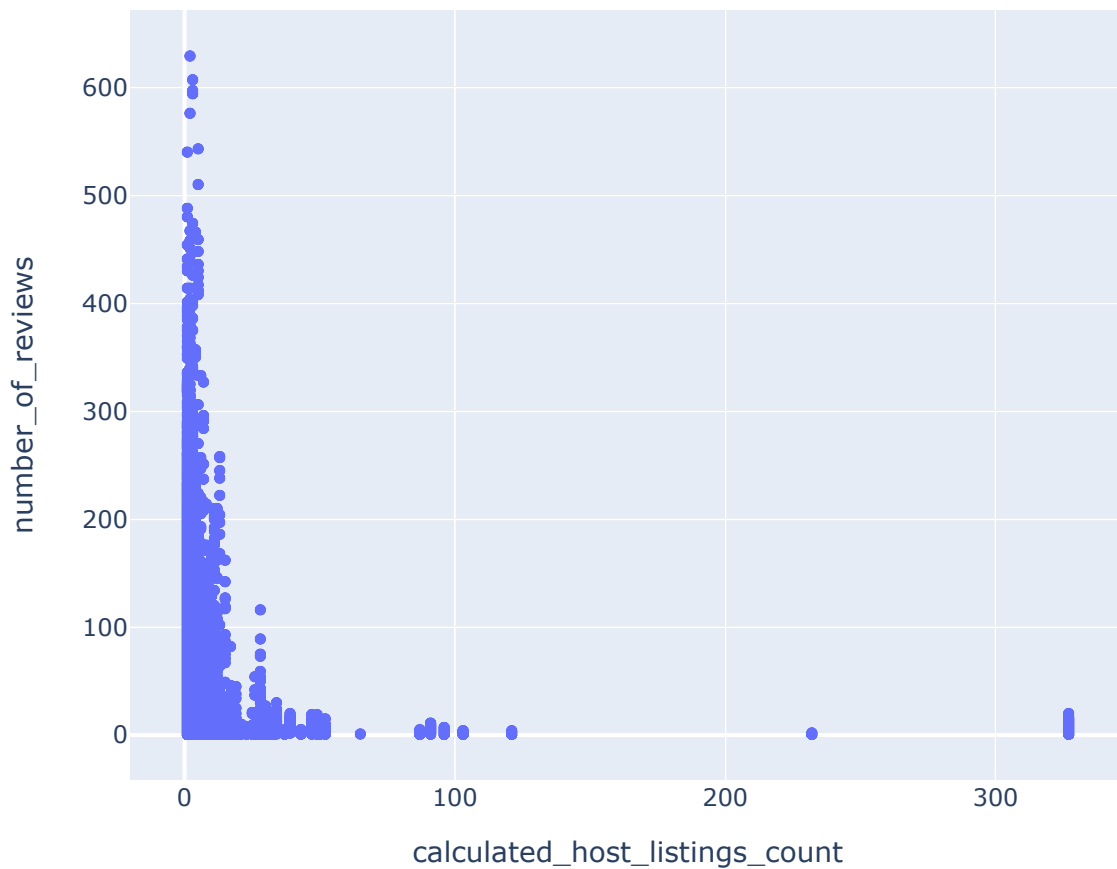
```
Out[27]:
```

	Neighborhood Name	count
0	Williamsburg	3163
1	Bedford-Stuyvesant	3141
2	Harlem	2206
3	Bushwick	1944
4	Hell's Kitchen	1532
5	East Village	1490
6	Upper West Side	1482
7	Upper East Side	1405
8	Crown Heights	1265
9	Midtown	986

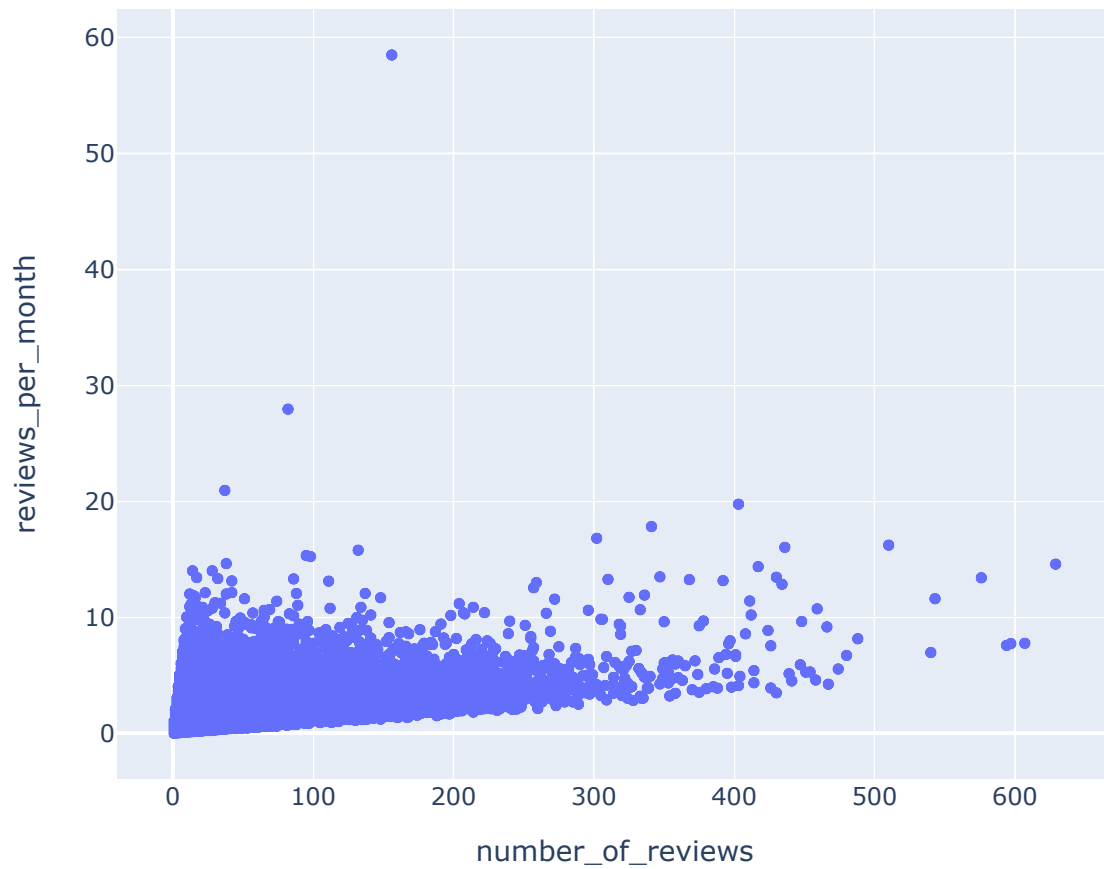
```
In [28]: viz_1=sns.barplot(x="Neighborhood Name", y="count", data=top_places_df,
                           palette='Blues_d')
viz_1.set_title('Neighbourhood areas with Most of the Airbnb')
viz_1.set_ylabel('Count of listings')
viz_1.set_xlabel('Neighbourhood Names')
viz_1.set_xticklabels(viz_1.get_xticklabels(), rotation=45)
plt.show()
```



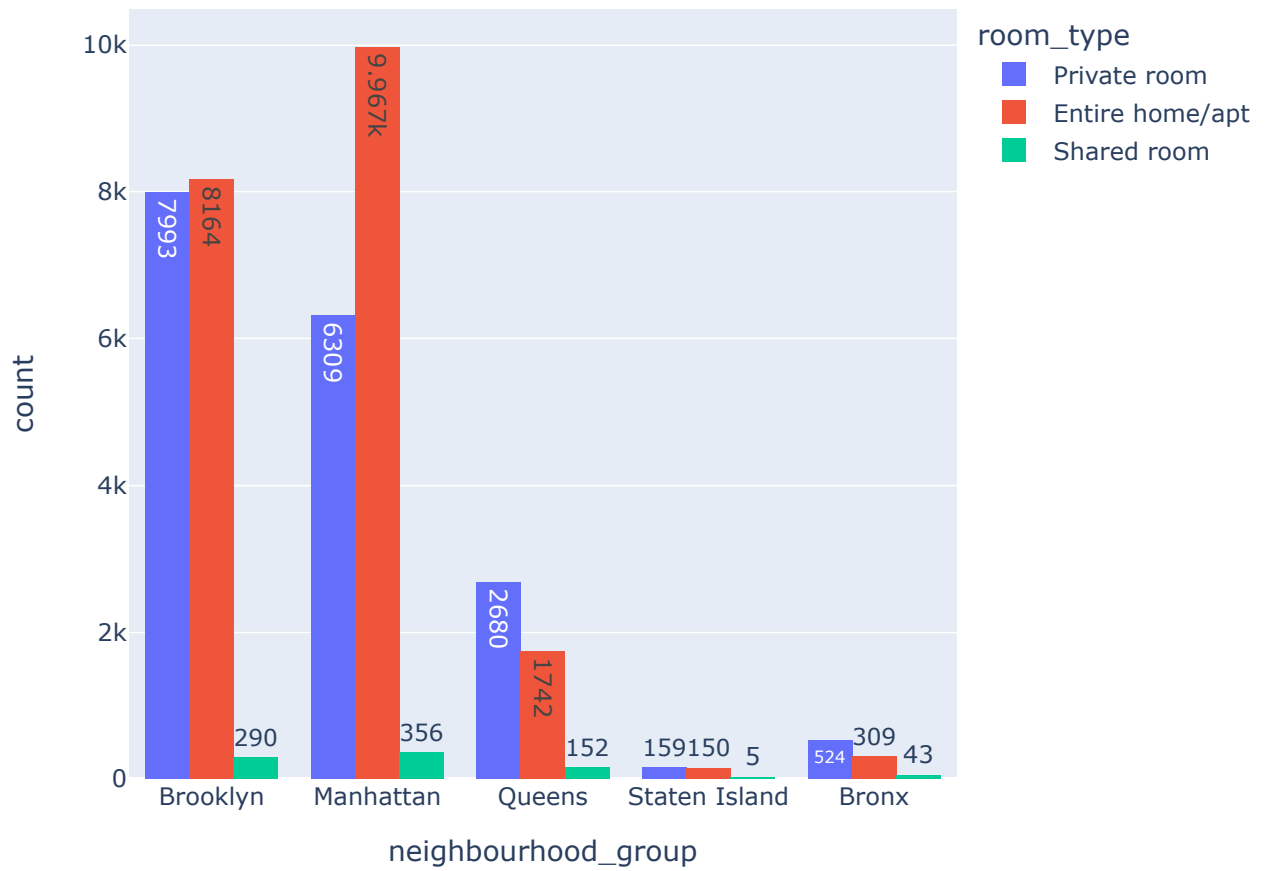
```
In [29]: fig = px.scatter(airbnb_df, x="calculated_host_listings_count", y="number_of_reviews")
fig.show()
```



```
In [30]: fig2 = px.scatter(airbnb_df, x="number_of_reviews", y="reviews_per_month")
fig2.show()
```



```
In [31]: # data_canada = px.data.gapminder().query("country == 'Canada'")
fig3 = px.histogram(airbnb_df, x='neighbourhood_group', color = 'room_type', barmode='group', text_
fig3.show()
```

```
In [32]: airbnb_df.nlargest(10, 'number_of_reviews')
```

Out[32]:

	id	name	host_id	host_name	neighbourhood_group	neighbourhood	latitude	lon
	11759	9145202	Room near JFK Queen Bed	47621202	Dona	Queens	Jamaica	40.66730
	2031	903972	Great Bedroom in Manhattan	4734398	Jj	Manhattan	Harlem	40.82085
	2030	903947	Beautiful Bedroom in Manhattan	4734398	Jj	Manhattan	Harlem	40.82124
	2015	891117	Private Bedroom in Manhattan	4734398	Jj	Manhattan	Harlem	40.82264
	13495	10101135	Room Near JFK Twin Beds	47621202	Dona	Queens	Jamaica	40.66939
	10623	8168619	Steps away from Laguardia airport	37312959	Maya	Queens	East Elmhurst	40.77006
	1879	834190	Manhattan Lux Loft.Like.Love.Lots.Look !	2369681	Carol	Manhattan	Lower East Side	40.71921
	20403	16276632	Cozy Room Family Home LGA Airport NO CLEANING FEE	26432133	Danielle	Queens	East Elmhurst	40.76335
	4870	3474320	Private brownstone studio Brooklyn	12949460	Asa	Brooklyn	Park Slope	40.67926
	471	166172	LG Private Room/Family Friendly	792159	Wanda	Brooklyn	Bushwick	40.70283

In [33]:

airbnb_df.nlargest(10, 'price')

Out[33]:

	id	name	host_id	host_name	neighbourhood_group	neighbourhood	latitude	long	
	9151	7003697	Furnished room in Astoria apartment	20582832	Kathrine	Queens	Astoria	40.76810	-73.
	17692	13894339	Luxury 1 bedroom apt. -stunning Manhattan views	5143901	Erin	Brooklyn	Greenpoint	40.73260	-73.
	6530	4737930	Spanish Harlem Apt	1235070	Olson	Manhattan	East Harlem	40.79264	-73.
	12342	9528920	Quiet, Clean, Lit @ LES & Chinatown	3906464	Amy	Manhattan	Lower East Side	40.71355	-73.
	30268	23377410	Beautiful/Spacious 1 bed luxury flat-TriBeCa/Soho	18128455	Rum	Manhattan	Tribeca	40.72197	-74.
	4377	2953058	Film Location	1177497	Jessica	Brooklyn	Clinton Hill	40.69137	-73.
	45666	34895693	Gem of east Flatbush	262534951	Sandra	Brooklyn	East Flatbush	40.65724	-73.
	3537	2110145	UWS 1BR w/backyard + block from CP	2151325	Jay And Liz	Manhattan	Upper West Side	40.77782	-73.
	43009	33397385	Midtown Manhattan great location (Gramacy park)	16105313	Debra	Manhattan	Midtown	40.74482	-73.
	2698	1448703	Beautiful 1 Bedroom in Nolita/Soho	213266	Jessica	Manhattan	Nolita	40.72193	-73.

In [34]:

airbnb_df.nlargest(10, 'minimum_nights')

Out[34]:

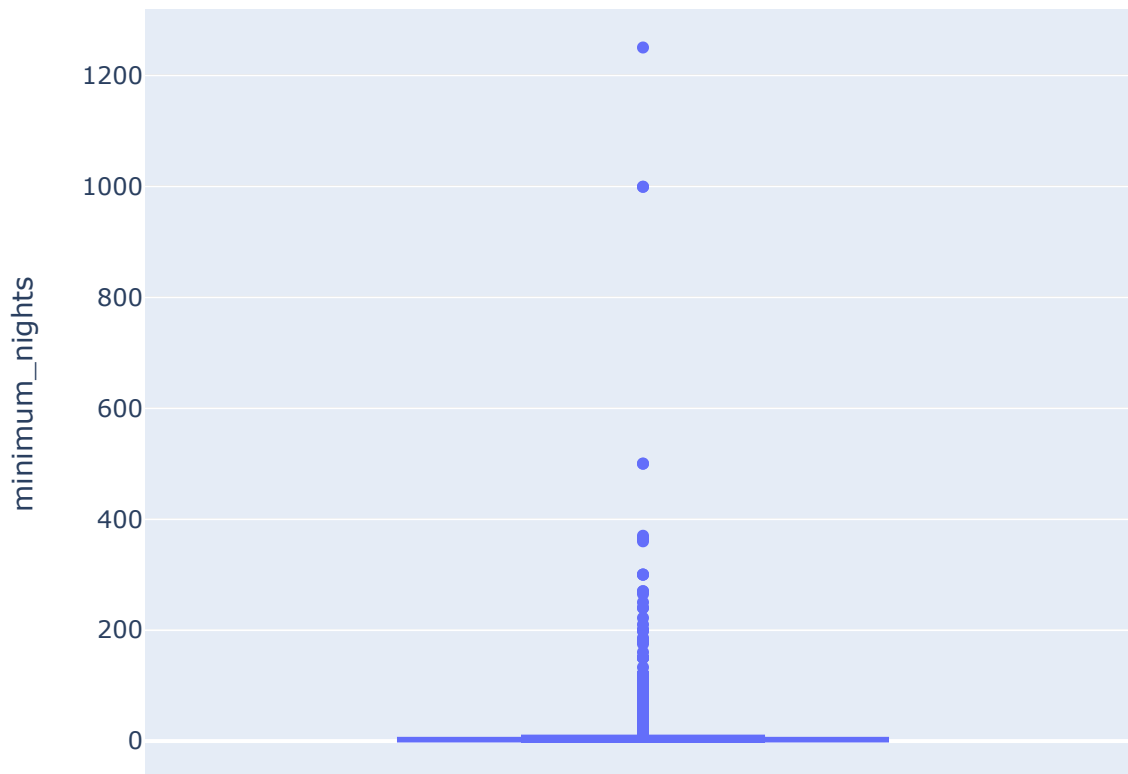
	id	name	host_id	host_name	neighbourhood_group	neighbourhood	latitude	longitude
5767	4204302	Prime W. Village location 1 bdrm	17550546	Genevieve	Manhattan	Greenwich Village	40.73293	-73.99782
13404	10053943	Historic Designer 2 Bed. Apartment	2697686	Glenn H.	Manhattan	Harlem	40.82915	-73.94034
26341	20990053	Beautiful place in Brooklyn! #2	151084261	Angie	Brooklyn	Williamsburg	40.71772	-73.95059
7355	5431845	Beautiful Fully Furnished 1 bed/bth	3680008	Aliya	Queens	Long Island City	40.75104	-73.93863
11193	8668115	Zen Room in Crown Heights Brooklyn	8996336	Laura	Brooklyn	Crown Heights	40.67255	-73.94914
1305	568684	800sqft apartment with huge terrace	2798644	Alessandra	Brooklyn	Bushwick	40.70202	-73.92402
700	258690	CHELSEA 1 Bdrm Plus Sleeping Loft!!	1359611	Andrea	Manhattan	Chelsea	40.74618	-74.00392
754	271694	Easy, comfortable studio in Midtown	1387370	James	Manhattan	Midtown	40.75282	-73.97315
2150	992977	Park Slope Pre-War Apartment	4000059	Shahdiya	Brooklyn	Park Slope	40.67359	-73.97434
3945	2454507	Close to everything - Jr 1 bedroom	9991763	Andrew	Manhattan	Financial District	40.70781	-74.00701

Now, lets look at this, here we can see that airbnb homes have set up requirement of minimum nights of upto 1250 days which is not possible in practical terms as well as legally.

Minimum nights restriction can mostly go upto 30 days, legally if a guest is staying for around 30 days then they are no more called guests and are referred as tennants and hotels have to generate a differenrent invoice and go ahead with differnet paper procedures. So, in this case lets remove the the minimum nights which are more than 150 days.

In [35]:

```
fig_nights = px.box(airbnb_df, y="minimum_nights")
fig_nights.show()
```



There are several ways to remove the outliers but I will go ahead with IQR filtering. Where we will first find the Q1 and Q3 which is 25 and 75 percentile of your data. Post that we will find the IQR which $IQR = Q3 - Q1$.

After we get our IQR, we can find out the upper limit and lower limits of our data. $upper\ limit = Q3 + (1.5 \times IQR)$ and $lower\ limit = Q1 - (1.5 \times IQR)$

```
In [36]: Q1 = airbnb_df['minimum_nights'].quantile(0.25)
Q3 = airbnb_df['minimum_nights'].quantile(0.75)
print("value of Q1 and Q3 are",Q1 , 'and', Q3)
```

value of Q1 and Q3 are 1.0 and 4.0

```
In [37]: IQR = Q3 - Q1
print("value of IQR is",IQR)
```

value of IQR is 3.0

```
In [38]: lower_limit = Q1 - (1.5*IQR)
upper_limit = Q3 +(1.5*IQR)
print("upper and lower limits of minimum nights feature is", upper_limit, 'and', lower_limit)
```

upper and lower limits of minimum nights feature is 8.5 and -3.5

```
In [39]: #You can find the list of records that are outliers in the dataset.
airbnb_df[airbnb_df['minimum_nights'] > upper_limit]
```

Out[39]:

	id	name	host_id	host_name	neighbourhood_group	neighbourhood	latitude	longitu	
	4	5022	Entire Apt: Spacious Studio/Loft by central park	7192	Laura	Manhattan	East Harlem	40.79851	-73.943
	6	5121	BlissArtsSpace!	7356	Garon	Brooklyn	Bedford- Stuyvesant	40.68688	-73.955
	14	6090	West Village Nest - Superhost	11975	Alina	Manhattan	West Village	40.73530	-74.005
	27	9357	Midtown Pied- a-terre	30193	Tommi	Manhattan	Hell's Kitchen	40.76715	-73.985
	29	9657	Modern 1 BR / NYC / EAST VILLAGE	21904	Dana	Manhattan	East Village	40.72920	-73.985

	47906	35974123	Downtown NYC - Across From the Freedom Tower!	73986440	Daniel	Manhattan	Financial District	40.71047	-74.007
	48024	36038287	Huge, Clean & Quiet room in a 2 BR Apt	167769983	Sara	Manhattan	Stuyvesant Town	40.73221	-73.979
	48203	36137740	Cozy Kips Bay Apartment	30933227	Dipti	Manhattan	Kips Bay	40.74133	-73.977
	48486	36280646	Cable and wfi, L/G included.	272872092	Chris	Queens	Forest Hills	40.73657	-73.850
	48636	36351543	2-MONTH SUBLEASE (WITH EARLY MOVE-IN)	52984497	Friday	Brooklyn	Bedford- Stuyvesant	40.68914	-73.924

4464 rows × 16 columns

In [40]:

```
#No outliers in the lower limit area.  
airbnb_df[airbnb_df['minimum_nights'] < lower_limit]
```

Out[40]:

id	name	host_id	host_name	neighbourhood_group	neighbourhood	latitude	longitude	room_type	price	n
----	------	---------	-----------	---------------------	---------------	----------	-----------	-----------	-------	---

In [41]:

```
#Lets trim and remove the outliers present in upper limit, please note that the angle bracket is  
#In the above line it was > upper_limit and here for trimming we use < upper_limit  
newairbnb_df = airbnb_df[airbnb_df['minimum_nights'] < upper_limit]  
newairbnb_df.shape
```

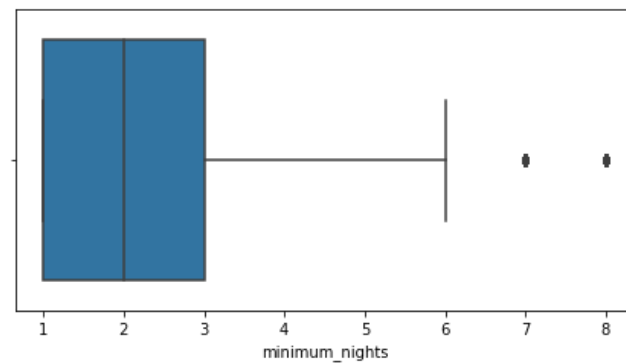
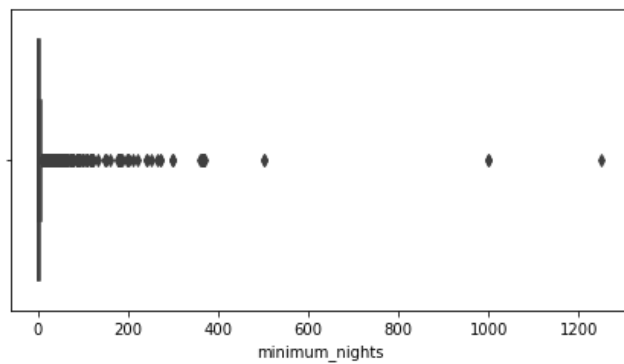
Out[41]: (34379, 16)

Lets compare the boxplots of previous data and new data.

In [42]:

```
plt.figure(figsize=(16,8))  
plt.subplot(2,2,1)  
sns.boxplot(airbnb_df['minimum_nights'])  
plt.subplot(2,2,2)
```

```
sns.boxplot(newairbnb_df['minimum_nights'])  
plt.show()
```



```
In [43]: newairbnb_df.nlargest(10, 'minimum_nights')
```

Out[43]:

	id	name	host_id	host_name	neighbourhood_group	neighbourhood	latitude	longitude	room_type
149	34760	Sunny Room in Old Historical Brooklyn Townhouse	149929	Obed	Brooklyn	Fort Greene	40.69101	-73.97312	
181	45910	Beautiful Queens Brownstone! - 5BR	204539	Mark	Queens	Ridgewood	40.70382	-73.89797	hco
422	144087	LUXURY OF THE HORIZON	616825	Rinaldo	Manhattan	Harlem	40.80473	-73.95320	hco
810	289020	Sunny 1BR Center of East Village!	347036	Jasmine	Manhattan	East Village	40.72948	-73.98694	hco
1094	460999	Beautiful duplex apt in Harlem	128890	Sara	Manhattan	Harlem	40.81620	-73.94747	hco
1407	629770	Slick Studio with High Ceilings	3129020	Ian	Manhattan	Gramercy	40.73326	-73.98245	hco
1685	764753	Beautiful 3 bedroom apartment!!	3371859	Jenny & Jose	Brooklyn	Crown Heights	40.67679	-73.93776	hco
1780	799945	Very Small Room, Old Historic Brooklyn Townhouse	149929	Obed	Brooklyn	Fort Greene	40.68537	-73.97713	
1817	813225	Quirky, Exposed-Brick Cozy Room Brooklyn Townh...	149929	Obed	Brooklyn	Fort Greene	40.68871	-73.97181	
1857	825550	Luxury 1BD in Historic Carrol Gardens Brownstone	4331864	Jeff	Brooklyn	Gowanus	40.67869	-73.99197	hco

In [44]:

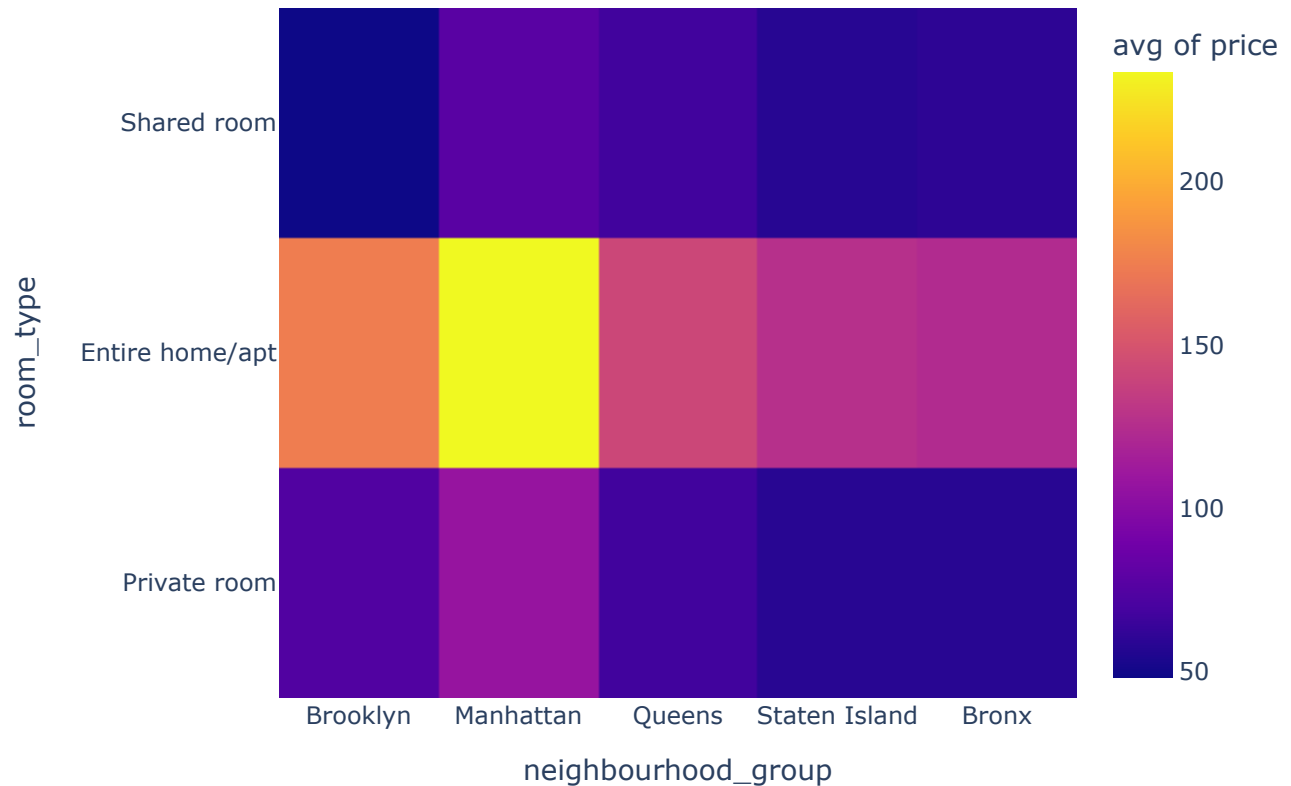
```
checkfig = px.density_heatmap(airbnb_df, x="neighbourhood_group", y="room_type", z="price", hist=
checkfig.show()
```


Average Price of differnt Rooms in Different areas



```
In [45]: checkfig1 = px.density_heatmap(newairbnb_df, x="neighbourhood_group", y="room_type", z="price", l
checkfig1.show())
```

Average Price of differnt Rooms in Different areas



We can see that there is not much of difference in avergae pricing in areas and room type before and after we removed the outlier.

```
In [46]: newtop_places =newairbnb_df.neighbourhood.value_counts().head(10)
print(newtop_places)
```

```
Williamsburg      2857
Bedford-Stuyvesant 2827
Harlem            1984
Bushwick          1763
East Village      1346
Hell's Kitchen    1306
Upper West Side   1192
Crown Heights     1148
Upper East Side   1099
East Harlem        871
Name: neighbourhood, dtype: int64
```

No difference in the top places as well.

```
In [47]: top_roomtype =airbnb_df.room_type.value_counts().head(10)
print(top_roomtype)
```

```
Entire home/apt    20332
Private room       17665
Shared room         846
Name: room_type, dtype: int64
```

```
In [48]: newtop_roomtype =newairbnb_df.room_type.value_counts().head(10)
```

```
print(newtop_roomtype)
```

```
Entire home/apt      17453
Private room         16167
Shared room           759
Name: room_type, dtype: int64
```

```
In [49]: airbnb_df.nlargest(10, 'number_of_reviews')
```

Out[49]:	id	name	host_id	host_name	neighbourhood_group	neighbourhood	latitude	l
11759	9145202	Room near JFK Queen Bed	47621202	Dona	Queens	Jamaica	40.66730	.
2031	903972	Great Bedroom in Manhattan	4734398	Jj	Manhattan	Harlem	40.82085	.
2030	903947	Beautiful Bedroom in Manhattan	4734398	Jj	Manhattan	Harlem	40.82124	.
2015	891117	Private Bedroom in Manhattan	4734398	Jj	Manhattan	Harlem	40.82264	.
13495	10101135	Room Near JFK Twin Beds	47621202	Dona	Queens	Jamaica	40.66939	.
10623	8168619	Steps away from Laguardia airport	37312959	Maya	Queens	East Elmhurst	40.77006	.
1879	834190	Manhattan Lux Loft.Like.Love.Lots.Look !	2369681	Carol	Manhattan	Lower East Side	40.71921	.
20403	16276632	Cozy Room Family Home LGA Airport NO CLEANING FEE	26432133	Danielle	Queens	East Elmhurst	40.76335	.
4870	3474320	Private brownstone studio Brooklyn	12949460	Asa	Brooklyn	Park Slope	40.67926	.
471	166172	LG Private Room/Family Friendly	792159	Wanda	Brooklyn	Bushwick	40.70283	.

```
In [50]: newairbnb_df.nlargest(10, 'number_of_reviews')
```

Out[50]:

	id	name	host_id	host_name	neighbourhood_group	neighbourhood	latitude	lon
	11759	9145202	Room near JFK Queen Bed	47621202	Dona	Queens	Jamaica	40.66730
	2031	903972	Great Bedroom in Manhattan	4734398	Jj	Manhattan	Harlem	40.82085
	2030	903947	Beautiful Bedroom in Manhattan	4734398	Jj	Manhattan	Harlem	40.82124
	2015	891117	Private Bedroom in Manhattan	4734398	Jj	Manhattan	Harlem	40.82264
	13495	10101135	Room Near JFK Twin Beds	47621202	Dona	Queens	Jamaica	40.66939
	10623	8168619	Steps away from Laguardia airport	37312959	Maya	Queens	East Elmhurst	40.77006
	1879	834190	Manhattan Lux Loft.Like.Love.Lots.Look !	2369681	Carol	Manhattan	Lower East Side	40.71921
	20403	16276632	Cozy Room Family Home LGA Airport NO CLEANING FEE	26432133	Danielle	Queens	East Elmhurst	40.76335
	4870	3474320	Private brownstone studio Brooklyn	12949460	Asa	Brooklyn	Park Slope	40.67926
	471	166172	LG Private Room/Family Friendly	792159	Wanda	Brooklyn	Bushwick	40.70283

In [51]:

```
pd.pivot_table(airbnb_df,index= 'room_type',values = ["number_of_reviews"],aggfunc = "sum")
```

Out[51]:

	number_of_reviews
room_type	
Entire home/apt	580403
Private room	538346
Shared room	19256

In [52]:

```
pd.pivot_table(airbnb_df,index= 'neighbourhood_group',values = ["number_of_reviews"],aggfunc = "sum")
```

Out[52]:

	number_of_reviews
neighbourhood_group	
Bronx	28371
Brooklyn	486574
Manhattan	454569
Queens	156950
Staten Island	11541

In [53]:

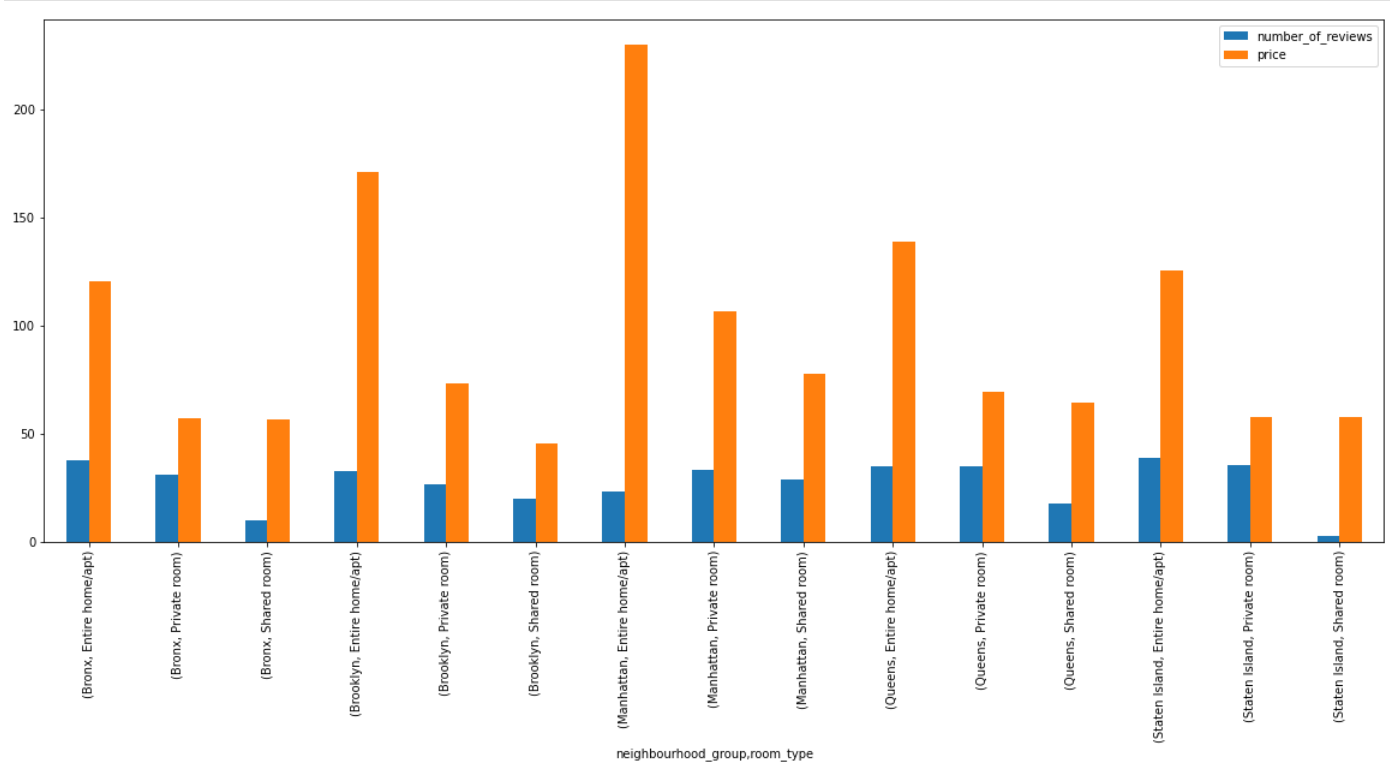
```
avg_review_price = pd.pivot_table(airbnb_df,index= ['neighbourhood_group','room_type'],values = ["number_of_reviews"],aggfunc = "sum")
```

Out[53]:

neighbourhood_group		room_type	number_of_reviews	price
Bronx		Entire home/apt	37.627832	120.789644
		Private room	31.129771	57.093511
		Shared room	10.046512	56.930233
Brooklyn		Entire home/apt	32.720235	171.249265
		Private room	26.730014	73.339797
		Shared room	19.975862	45.458621
Manhattan		Entire home/apt	23.592555	230.190830
		Private room	33.151054	106.603265
		Shared room	28.853933	77.980337
Queens		Entire home/apt	34.812859	139.225603
		Private room	34.910821	69.286567
		Shared room	18.059211	64.467105
Staten Island		Entire home/apt	39.046667	125.366667
		Private room	35.660377	57.584906
		Shared room	2.800000	57.600000

In [54]:

```
# plt.figure(figsize=(20,20))
avg_review_price.plot(kind='bar',figsize=(20,8))
plt.show()
```



From the plot above, you can see that on an average staten island and Queens cities have got more number of reviews for the airbnb. Whereas the pricing is high for Manhattan City and then you have got Brooklyn City

