

```
In [25]: import pandas as pd
import plotly.express as px
```

```
In [26]: #Reading the csv file
df_airbnb = pd.read_csv('AB_NYC_2019.csv')
```

```
In [27]: df_airbnb.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 [28]: #Filling blank spaces with zero
df_airbnb.fillna(value = 0)
```

	id	name	host_id	host_name	neighbourhood_group	neighbourhood	latitude	longitude	room_type	price	minimum_nights	number_of_reviews	last_review	
	1	2595	Skylit Midtown Castle	2845	Jennifer	Manhattan	Midtown	40.75362	-73.98377	Entire home/apt	225	1	45	2019-05-21
	2	3647	THE VILLAGE OF HARLEM....NEW YORK !	4632	Elisabeth	Manhattan	Harlem	40.80902	-73.94190	Private room	150	3	0	0
	3	3831	Cozy Entire Floor of Brownstone	4869	LisaRoxanne	Brooklyn	Clinton Hill	40.68514	-73.95976	Entire home/apt	89	1	270	2019-07-05
	4	5022	Entire Apt- Spacious Studio/Loft by central park	7192	Laura	Manhattan	East Harlem	40.79851	-73.94399	Entire home/apt	80	10	9	2018-11-19

	48890	36484665	Charming one bedroom - newly renovated rowhouse	8232441	Sabrina	Brooklyn	Bedford-Stuyvesant	40.67853	-73.94995	Private room	70	2	0	0
	48891	36485057	Affordable room in Bushwick/East Williamsburg	6570630	Marisol	Brooklyn	Bushwick	40.70184	-73.93317	Private room	40	4	0	0
	48892	36485431	Sunny Studio at Historical Neighborhood	23492952	Ilgar & Aysel	Manhattan	Harlem	40.81475	-73.94867	Entire home/apt	115	10	0	0
	48893	36485609	43rd St. Time Square-cozy single bed	30985759	Taz	Manhattan	Hell's Kitchen	40.75751	-73.99112	Shared room	55	1	0	0
	48894	36487245	Trendy duplex in the very heart of Hell's Kitchen	68119814	Christophe	Manhattan	Hell's Kitchen	40.76404	-73.98933	Private room	90	7	0	0

48895 rows × 16 columns

```
In [29]: df_top10host = df_airbnb['host_id'].value_counts().head(10)
```

```
In [31]: df_top10host = df_top10host.reset_index()
df_top10host
```

	level_0	index	host_id
0	0	219517861	327
1	1	107434423	232
2	2	30283594	121
3	3	137358866	103
4	4	16098958	96
5	5	12243051	96
6	6	61391963	91
7	7	22541573	87
8	8	200380610	65
9	9	7503643	52

```
In [32]: df_top10host.rename(columns = {'index':'host_id','host_id':'number_of_listing'},inplace = True)
```

```
In [35]: #names of top 10 host_id
df_hostname = df_airbnb.loc[df_airbnb['host_id'].isin(df_top10host['host_id'])]
df_hostname
```

	id	name	host_id	host_name	neighbourhood_group	neighbourhood	latitude	longitude	room_type	price	minimum_nights	number_of_reviews	last_review	re
	2631	1390532	Amazing Designer Loft in BK Factory	7503643	Vida	Brooklyn	Greenpoint	40.72456	-73.94343	Entire home/apt	129	30	7	2019-03-27
	2636	1392673	Spacious Factory Converted Loft	7503643	Vida	Brooklyn	Greenpoint	40.72756	-73.94230	Entire home/apt	129	30	6	2018-09-02
	2708	1460017	Lovely Factory Loft in Greenpoint!	7503643	Vida	Brooklyn	Greenpoint	40.72732	-73.94076	Entire home/apt	129	30	6	2018-09-11
	2749	1486099	All New Bohemian Chic 1BR Loft	7503643	Vida	Brooklyn	Greenpoint	40.72267	-73.94362	Entire home/apt	149	30	9	2018-02-18
	2782	1524302	Huge 1 Bedroom Loft at Habitat 101!	7503643	Vida	Brooklyn	Greenpoint	40.72686	-73.94177	Entire home/apt	149	30	7	2018-10-31

	48723	36404684	Smart Nolita 1BR w/ W/D near famous cafes by B...	107434423	Blueground	Manhattan	Nolita	40.72283	-73.99472	Entire home/apt	316	30	0	NaN
	48724	36404784	Dapper Hell's Kitchen 2BR w/ Gym + W/D + Door...	107434423	Blueground	Manhattan	Hell's Kitchen	40.76082	-73.99709	Entire home/apt	385	30	0	NaN
	48725	36404815	Trendy Hell's Kitchen 1BR w/ Gym + W/D + Doom...	107434423	Blueground	Manhattan	Hell's Kitchen	40.76083	-73.99727	Entire home/apt	267	30	0	NaN
	48726	36404936	Beautiful Williamsburg 2BR w/ Skyline views + ...	107434423	Blueground	Brooklyn	Williamsburg	40.71493	-73.96365	Entire home/apt	278	30	0	NaN
	48727	36404972	Hip East Village 1BR w/ Gym, W/D, Doorman, nea...	107434423	Blueground	Manhattan	East Village	40.73206	-73.98776	Entire home/apt	365	30	0	NaN

1270 rows × 16 columns

```
In [36]: #last occurrence of host_name
df_hostnamelast = df_hostname.groupby('host_name').last()
```

```
In [37]: df_hostnamelast.reset_index(inplace = True)
```

```
In [38]: df_top10host.sort_values(by = 'host_id',inplace = True)
```

```
In [39]: df_hostnamelast.sort_values('host_id',inplace=True)
```

```
In [42]: df_top10host['host_name'] = df_hostnamelast['host_name']
df_top10host

C:\Users\User\AppData\Local\Temp\ipykernel_17864\161119726.py:1: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
```

	host_id	host_name	number_of_listing
9	7503643	Vida	52
5	12243051	Ken	96
4	16098958	Kazuya	96
7	22541573	Sonder	87
2	30283594	Jeremy & Laura	121
6	61391963	Pranjal	91
1	107434423	Corporate Housing	232
3	137358866	Kara	103
8	200380610	Sonder (NYC)	65
0	219517861	Blueground	327

```
In [43]: df_top10host = df_top10host[['host_id','host_name','number_of_listing']]
```

```
In [44]: df_hostnameaverage = df_hostname.groupby('host_name')['price'].mean()
```

```
In [45]: px.bar(df_top10host,x = 'host_name',y = 'number_of_listing',title = 'host vs number of listings')
```

