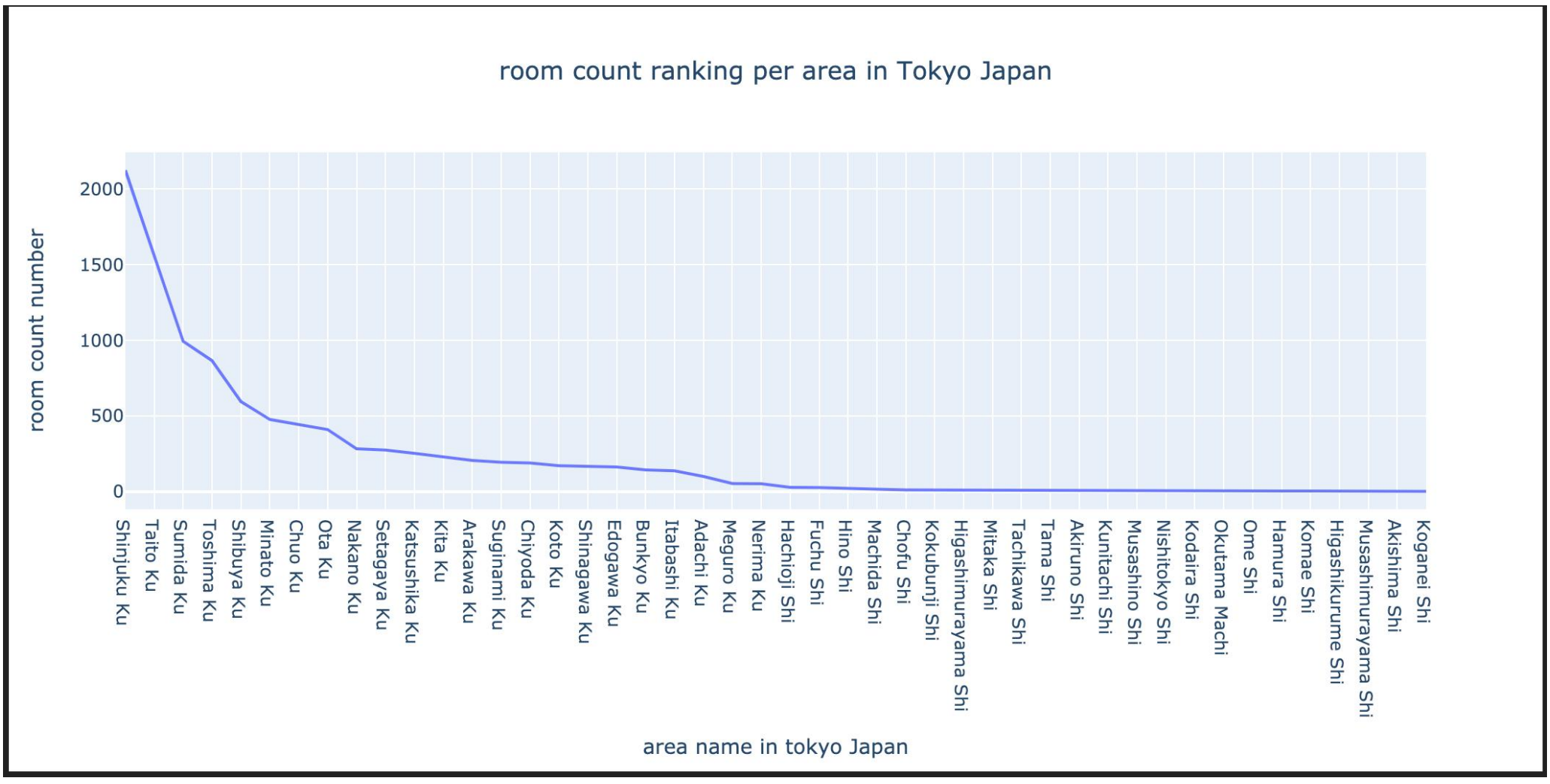
# Do you know the details about Tokyo AirBNB market?




# 01 choose a dataset for this project

in this project, Tokyo AirBNB data is used for data analysis and it's also available in the following link.
http://insideairbnb.com/get-the-data.html
the screen capture is also attached here. including some listings, calendar and reviews data.

**Tokyo, Kantō, Japan**
See Tokyo data visually here.

| Date Compiled | Country/City | File Name | Description |
|---|---|---|---|
| 28 December, 2021 | Tokyo | listings.csv.gz | Detailed Listings data for Tokyo |
| 28 December, 2021 | Tokyo | calendar.csv.gz | Detailed Calendar Data for listings in Tokyo |
| 28 December, 2021 | Tokyo | reviews.csv.gz | Detailed Review Data for listings in Tokyo |
| 28 December, 2021 | Tokyo | listings.csv | Summary information and metrics for listings in Tokyo (good for visualisations). |
| 28 December, 2021 | Tokyo | reviews.csv | Summary Review data and Listing ID (to facilitate time based analytics and visualisations linked to a listing). |
| N/A | Tokyo | neighbourhoods.csv | Neighbourhood list for geo filter. Sourced from city or open source GIS files. |
| N/A | Tokyo | neighbourhoods.geojson | GeoJSON file of neighbourhoods of the city. |

# 02 Key Steps for Project

## 02.01 data and github url

downloaded the csv files for listings, calendar and reviews from above dataset. and included in the following github project:
https://github.com/allhanz/AirBNB_data_analysis

## 02.02 several issues for data analysis

1) how many host or listings(houses) are avaliable in the tokyo airBNB market japan?
2) is there any relationship between price and the datetime?
3) which area have the most review count number and may be the most popular for customers?
4) which area have the most listings(rooms) ?
5) how the price for different room type?
6) how many hosts are available in the AirBNB tokyo market and how many listings(rooms) for each host?

## 02.03 data analysis details by creating jupyter notebook

### 02.03.01 some insight for data analysis

imaging i am a host who will decide to put some houses into the tokyo AirBNB market. i'd better to understand how the market situation. for example, which

location, which area, what kind of rom etc will be the popular to the local customers and for Foreign customers. with these kinds of questions, to check is there any answer infotmation included in this tokyo AirBNB dataset via data analysis.

## 02.03.02 data analysis details

## 1)how many host or listings(houses) are avaliable in the tokyo airBNB market japan?

as the following figure, we can check the following information:

1) at the end and beginning of the year, the room count per day decrease hardly. because this period is during the Japanese happy new year holidays which around from the last weekend in Dec to the first weekend in Jan.

2) after the first weekend in Jan, the room count number increases hardly.

3) at the day of 27th,March, the room count number also decrease hardly. about the reason,as my guess, it may have some relationship with the account end period of the year.

4) at the day of 26th,june, the room count number also decrease hardly. about the reason, as my guess, it may have some relationship with the summuer holiday in Japan.01. the host count numer for tokyo AirBNB market.

```
#plot the room count number per day
plot_data(avaliable_counts,"date","count")
```

✓ 0.1s Python


## 2)is there any relationship between price and the datetime?

as the following figure, we can check the points listed here.

1) the average price per room is increasing from jan to Dec.

2) compared with the above figure which shows the available room count number, we can see that, the price is becoming higher and higher as the available room count number is becoming smaller and smaller.

3) and also we can see that the price is highest in the Friday, and second high in Saturday in almost every week.

```
# plot the average price of all the rooms per day
plot_data(mean_price_per_date,"date","adjusted_price")
```

✓ 0.5s Python


zoom in the price data in the period from 13th,March to 1st,May


3)which area have the most review count number and may be the most popular for customers?

as the following figure, we can see the points listed here.

1) the shinjyuku-ku and shibuya-ku are also very popular for customers.

2) but also some not so famous area is also very popular for customers, such as ome-shi, komae-shi and okutama-machi.

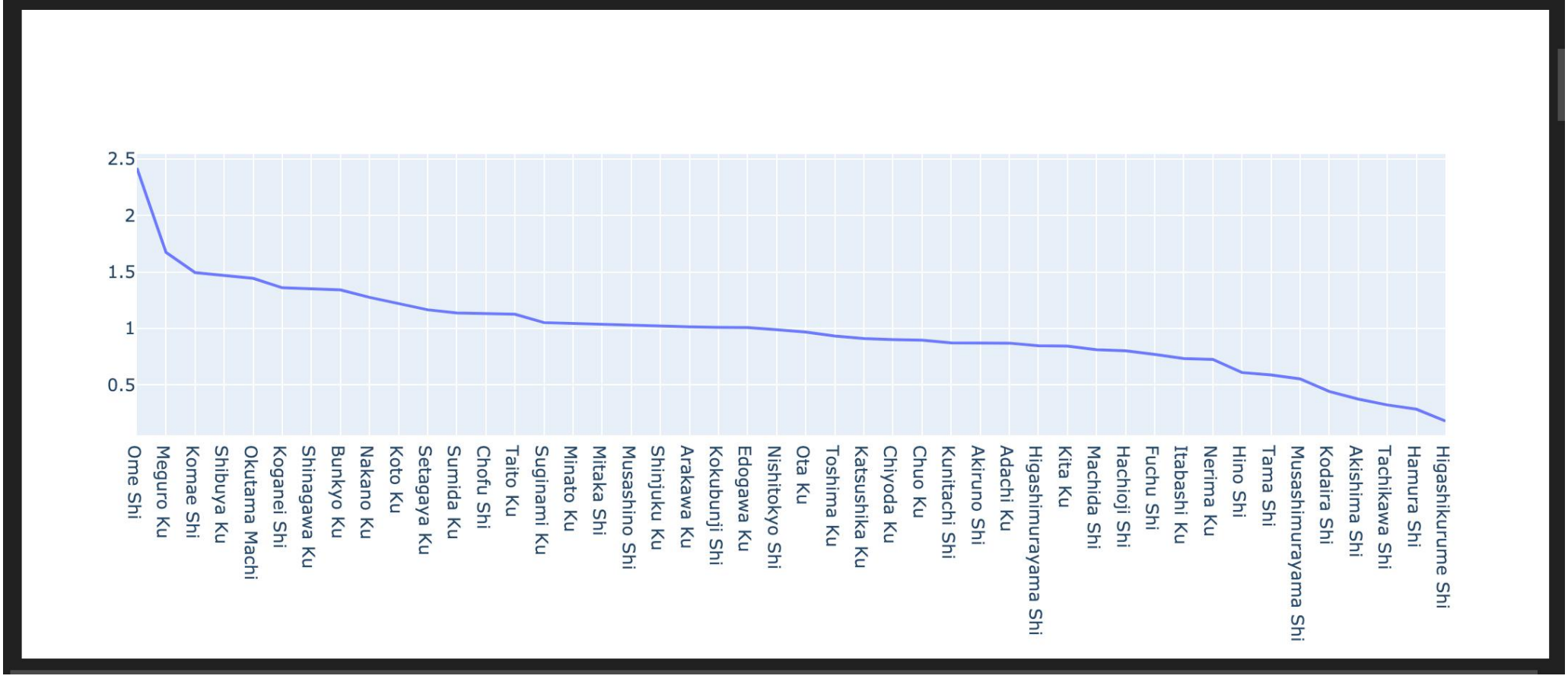
3) for also famous area like chiyoda-ku and chuo-ku, is not so highly reviewed. about the reason, need to analyze further. maybe about the too high price.

we also can check the top-5 ranking information here:

ume-shi> megoro-ku> komae-shi> shibuya-ku> okutama-machi

```
# plot the average review count per month for each area
plot_data(mean_reviws_sorted,"neighbourhood","reviews_per_month")
```

✓ 0.6s Python

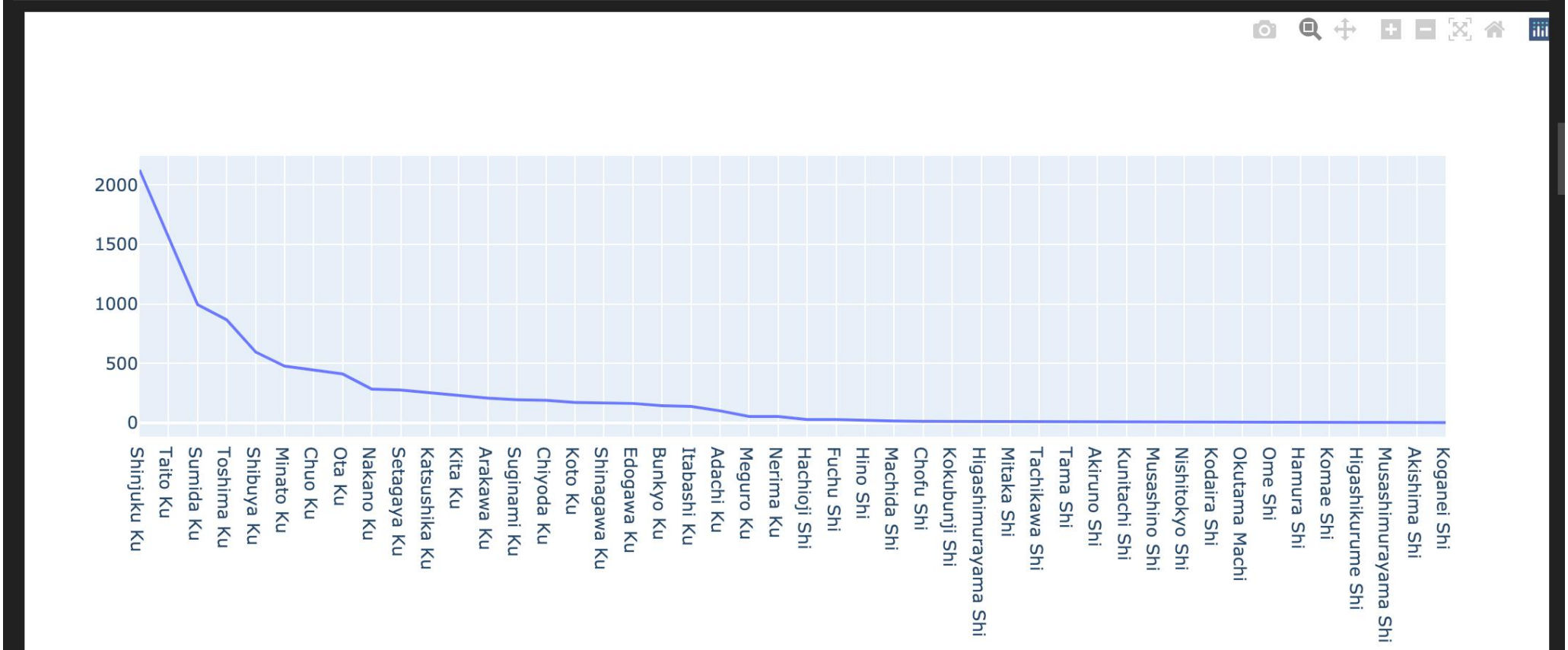



## 4)which area have the most listings(rooms) ?

here is the listing(room) count ranking for each area. we can see shinjyuku-ku area has the most listings(rooms). shibuya-ku area is only in 4th place. but taito-ku area is in the second place, maybe it is more closer than other area to the narita international Airpot.

```
plot_data(neigh_count_sorted,"neighbourhood","count")
```

✓ 0.1s Python




## 5)how the price for different room type?

the following are the average price for each type room. we can see that the entire room is the most expensive, private room is in the second place, hotel room is the 3rd. and of course the shared room is the cheapest. this is the same as my imagination.

| | room_type | price |
|---|---|---|
| 0 | Entire home/apt | 18426.9 |
| 1 | Private room | 16125 |
| 2 | Hotel room | 15045 |
| 3 | Shared room | 4915.67 |

4 rows × 2 columns

the following figure shows that Entire home have the largest count number and the private room is the second. also the same ranking with the price ranking.

| | room_type | count |
|---|---|---|
| 0 | Entire home/apt | 6922 |
| 1 | Private room | 2619 |
| 2 | Hotel room | 397 |
| 3 | Shared room | 376 |

4 rows × 2 columns

## 6)how many hosts are available in the AirBNB tokyo market and how many listings(rooms) for each host?

about 2554 hosts are available and hold around 10000 rooms in the AirBnB tokyo market. and also we can see that around 70% hosts have less than 3 lists(houses) in the tokyo AirBNB market.

```
    # calculate the percentage values for each type host which have 1 room,2 rooms, 3 rooms etc.
    listing_count_list=list(set(tokyo_list_count_sorted["count"].to_list()[0]))
    for item in listing_count_list:
        print("{} hosts have {} listings: about {}% :".format(tokyo_list_count_sorted[f["count"]==item,:].shape[0],item,tokyo_list_count_sorted[f[
```

✓ 0.4s Python

```
1110 hosts have 1 listings: about 43.46123727486296% :
403 hosts have 2 listings: about 15.779169929522318% :
232 hosts have 3 listings: about 9.08379013312451% :
178 hosts have 4 listings: about 6.9694596711041505% :
105 hosts have 5 listings: about 4.111198120595145% :
106 hosts have 6 listings: about 4.150352388410337% :
64 hosts have 7 listings: about 2.505873140172279% :
61 hosts have 8 listings: about 2.3884103367267033% :
41 hosts have 9 listings: about 1.6053249804228662% :
31 hosts have 10 listings: about 1.2137823022709475% :
32 hosts have 11 listings: about 1.2529365700861395% :
27 hosts have 12 listings: about 1.05716523101018% :
19 hosts have 13 listings: about 0.7439310884886452% :
23 hosts have 14 listings: about 0.9005481597494126% :
8 hosts have 15 listings: about 0.31323414252153486% :
14 hosts have 16 listings: about 0.548159749412686% :
13 hosts have 17 listings: about 0.5090054815974941% :
8 hosts have 18 listings: about 0.31323414252153486% :
9 hosts have 19 listings: about 0.3523884103367267% :
9 hosts have 20 listings: about 0.3523884103367267% :
7 hosts have 21 listings: about 0.274079874706343% :
3 hosts have 22 listings: about 0.11746280344557558% :
5 hosts have 23 listings: about 0.19577133907595928% :
1 hosts have 24 listings: about 0.03915426781519186% :
10 hosts have 25 listings: about 0.39154267815191857% :
4 hosts have 26 listings: about 0.15661707126076743% :
2 hosts have 27 listings: about 0.07830853563038372% :
2 hosts have 28 listings: about 0.07830853563038372% :
1 hosts have 29 listings: about 0.03915426781519186% :
3 hosts have 30 listings: about 0.11746280344557558% :
2 hosts have 31 listings: about 0.07830853563038372% :
3 hosts have 32 listings: about 0.11746280344557558% :
1 hosts have 33 listings: about 0.03915426781519186% :
2 hosts have 35 listings: about 0.07830853563038372% :
1 hosts have 36 listings: about 0.03915426781519186% :
3 hosts have 37 listings: about 0.11746280344557558% :
2 hosts have 39 listings: about 0.07830853563038372% :
1 hosts have 40 listings: about 0.03915426781519186% :
1 hosts have 41 listings: about 0.03915426781519186% :
1 hosts have 42 listings: about 0.03915426781519186% :
```

for more details, you can check in the notebook file which is “data_analysis.ipynb”