### problem description

Our goal is to predict the 'fair price' of an available house posted on Airbnb, depending on its features such as location, facilities nearby, size of the house, type of beds and so on.

First we need to summarize the features of each house in the list. Based on the dataset named 'listing' and 'venues'.

\_\_\_

### Summary

#### Our main work is as follows:

1. We plot the house distribution

2. We analyzed the types of venues

3. We predicted the housing price

4. We analyzed the volatility of housing price in one year.

\_\_\_

### main results:

1. Distribution of Airbnb houses in NYC

2. Tags and distribution of venues

We can get the venues by listing all the tags listed in 'Venue'.

For each entry, we can get its tags and get an unique set of all the tags.

Then we denote this information in a sparse matrix. For each factor, we use 1 to show that 1 certain place has certain tag such as 'transit\_station', 'post\_office', etc. And we use 0 to show that it does not have one.

For example, entry'Times Sq - 42 St' has the tags ['transit\_station', 'point\_of\_interest', 'establishment']

Such is a tiny version of this matrix:

venue\_id | food | bar |point\_of\_interest|establishment

----|------|----

Times Sq - 42 St| 0 | 0|1|1

Checkers | 1 | 0|1|1

Chez Kama| 1 | 1|1|1

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| venue\_id | food | bar | point\_of\_interest | establishment |
| Times Sq - 42 St | 0 | 0 | 1 | 1 |
| Checkers | 1 | 0 | 1 | 1 |
| Chez Kama | 1 | 1 | 1 | 1 |

And the one we generated from our program is just much larger.

Once we get this sparse matrix, combined with its location details (longitude&latitude) with certain houses' location information(longitude&latitude), we can get the distribution of certain tags around each Airbnb house listed in the dataset. And all this is for possible later use