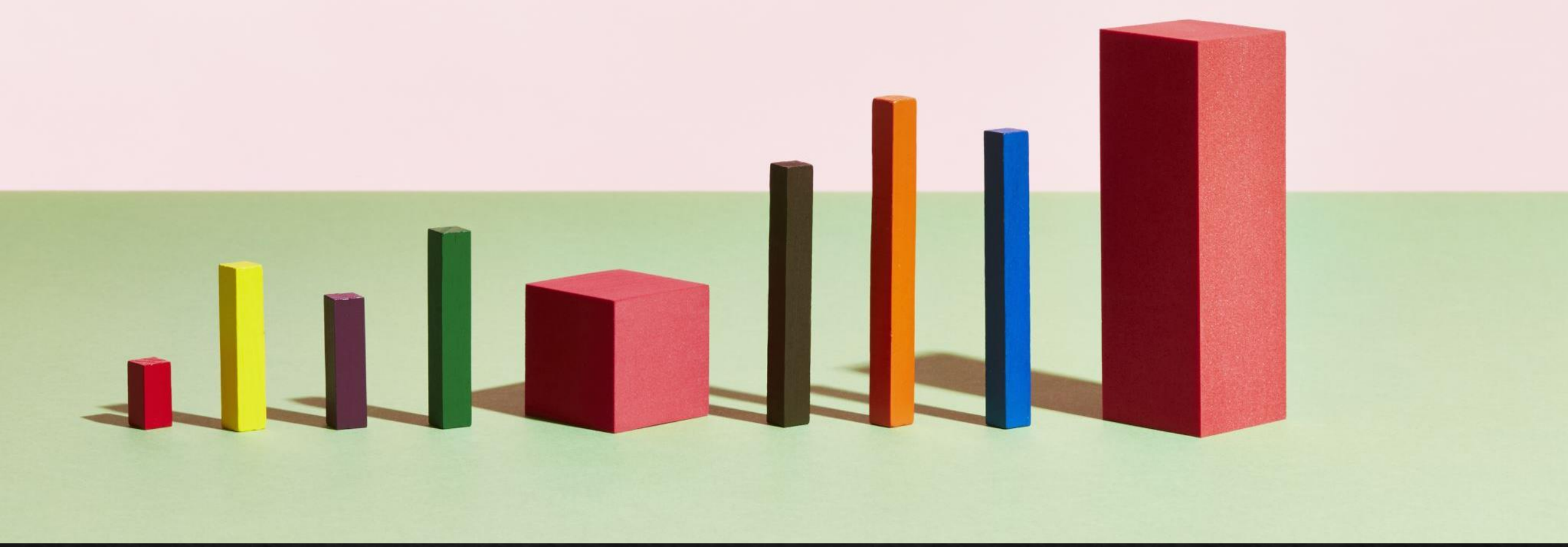


PREDICT SIMILAR BETWEEN NEIGHBOURHOODs

PROBLEMS

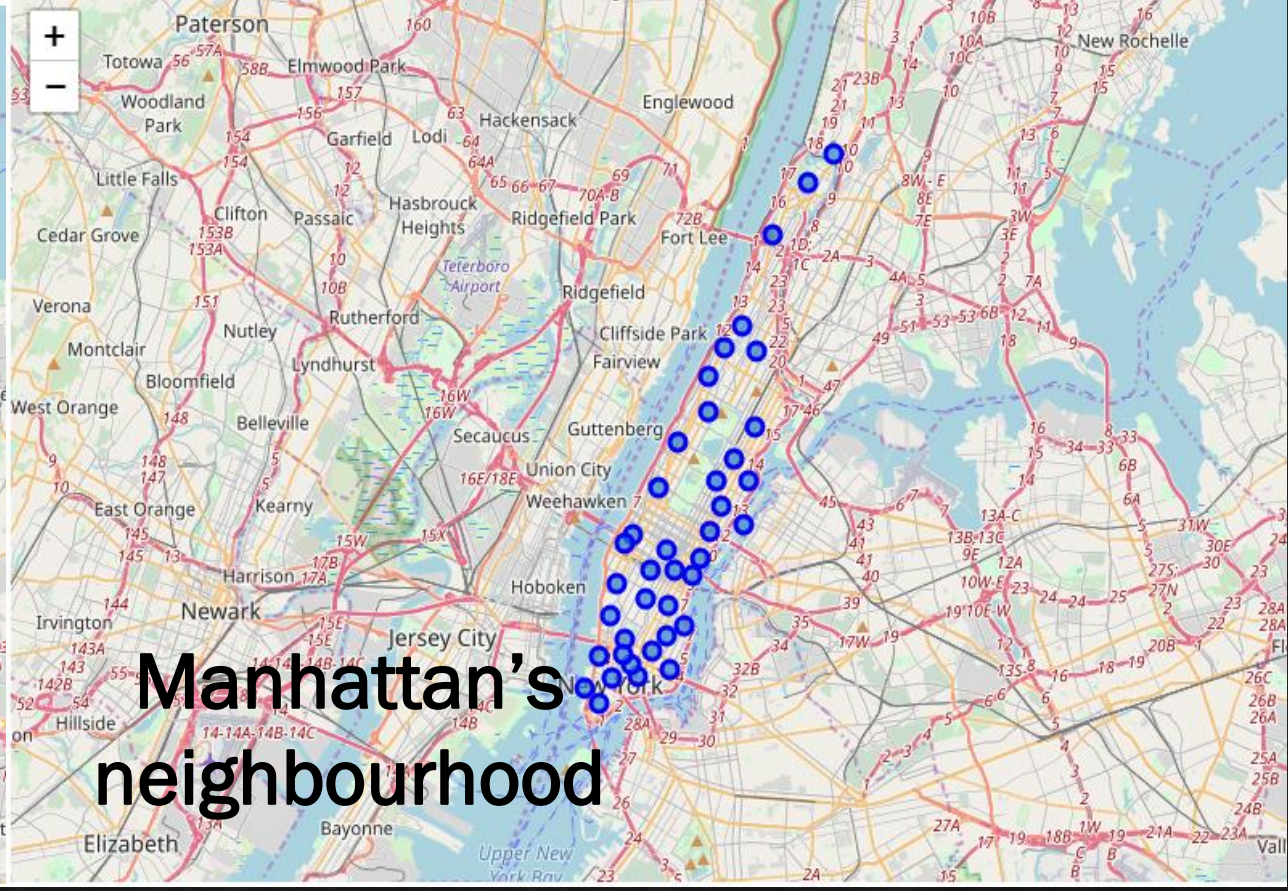
- Selecting where to live is very important.
- Many people need accommodations having many similar with their current places.
- Predicting where is the most suitable place for living is a difficult question now.
- For instance: Where is suitable with you when you go from Canada to USA?

QUESTIONS and INFORMATIONS

- Let's look at a question from my friend:

Now he is living in Manhattan, NY, and he want to get the most relevant neighbourhood for living in Bronx. He know informations about neighbourhoods and every venues in both Borough. Which is the way he should do to get the purpose above?

- The questions is now "WHICH WAY" with dataset about:
  - All neighbourhoods in New York City from this link: https://geo.nyu.edu/catalog/nyu_2451_34572
  - All venues in 2 Borough above from Foursquare API



QUESTIONS and INFORMATIONS

METHOD (THE WAY)

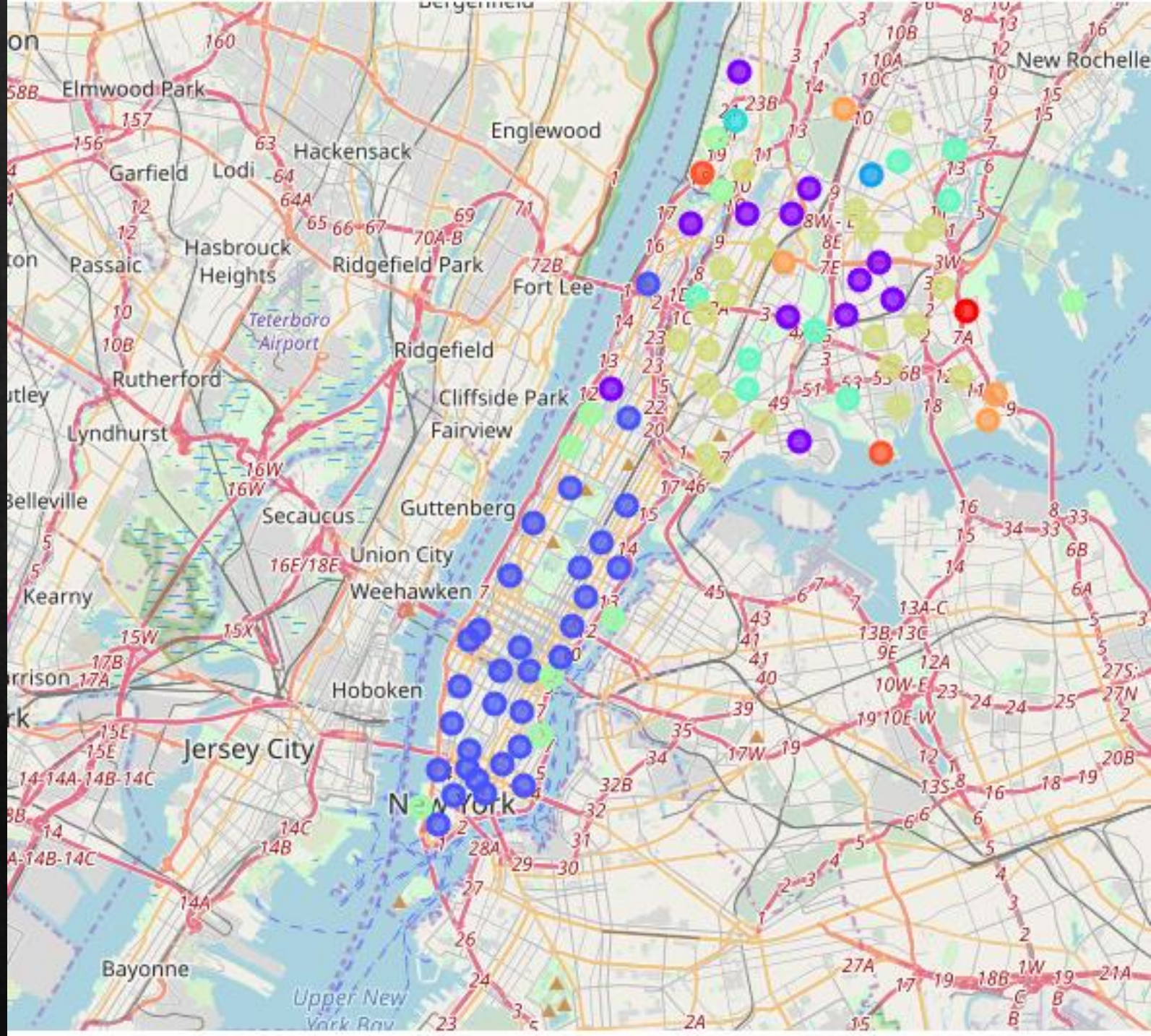
1. Merge dataset from both boroughs into dataframe X.
2. Merge dataset about venues each neigbhourhoods into dataframe X.
3. Use one-hot encoding for venues's categories of dataframe X.

Now, you can see apart of dataframe X on the right.

	Borough	Neighborhood	Accessories Store	Adult Boutique	Afghan Restaurant	African Restaurant	American Restaurant	Antique Shop	Arcade	Arepa Restaurant	Argentinian Restaurant	Art Gallery	Art Museum	
0	Bronx	Allerton	0.000000	0.00	0.00	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.(
1	Bronx	Baychester	0.000000	0.00	0.00	0.000000	0.000000	0.00	0.041667	0.000000	0.000000	0.000000	0.000000	0.(
2	Bronx	Bedford Park	0.000000	0.00	0.00	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.(
3	Bronx	Belmont	0.000000	0.00	0.00	0.000000	0.010204	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.(
4	Bronx	Bronxdale	0.000000	0.00	0.00	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.(
5	Bronx	Castle Hill	0.000000	0.00	0.00	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.(
6	Bronx	City Island	0.000000	0.00	0.00	0.000000	0.034483	0.00	0.000000	0.000000	0.000000	0.034483	0.000000	0.(
7	Bronx	Claremont Village	0.000000	0.00	0.00	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.(
8	Bronx	Clason Point	0.000000	0.00	0.00	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.(
9	Bronx	Co-op City	0.000000	0.00	0.00	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.(
10	Bronx	Concourse	0.000000	0.00	0.00	0.038462	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.038462	0.(
11	Bronx	Concourse Village	0.000000	0.00	0.00	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.(
12	Bronx	Country Club	0.000000	0.00	0.00	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.(
13	Bronx	East Tremont	0.000000	0.00	0.00	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.(
14	Bronx	Eastchester	0.000000	0.00	0.00	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.(
15	Bronx	Edenwald	0.000000	0.00	0.00	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.(

METHOD (THE WAY)

- Next, this method use K-mean Clustering for segmentation.
- Finally, we should look at something more intuitive (on the right again).
- Circles describe neigbourhoods.
- Color describe the similar between them.



CONCLUSION

- This is just a simple way to solve “Finding accommodations problems”.
- As you know, there are many independent variables which are not using in my method, such as price of apartments, quality of atmosphere or characteristics of people living around you. This is the first method I building, and I am going to upgrade this method in the future.

THANK YOU