

SETTLING DOWN IN CHICAGO

ARIBETH21

FOR IBM DATA SCIENCE COURSE

MOVING TO CHICAGO ?

Suppose you have a friend who, due to his change of job, will be moving to Chicago but know little of the city. He is of upper-middle class and has a family of four, he and his wife, together with their two young kids.

Which community will you recommend ?



MOVING TO CHICAGO ?

After a discussion with your friend, you both agree that this community should meet 2 requirements,

- **Safe.** Since we all know that Chicago is by no means a safe city, so it is the first thing that we would consider.
- **Relaxed.** As we are looking for a community to live a life, this community should provide a calm and relax environment, and of course with sufficient venues to support daily life, such as dry cleaning, restaurants, playgrounds for kids, etc.

MOVING TO CHICAGO ?

In this project, we are going to use data science knowledge to sort out the ideal community(s) for your friend to settle down in Chicago based on the 2 criteria.
Here is how we are going to solve the problem.

- Data
- Methodology
- Results
- Discussion
- Conclusion

DATA - REGARDING SAFETY

https://en.wikipedia.org/wiki/Community_areas_in_Chicago

North Side		
Number	Community area	Neighborhoods
05	North Center	. Horner Park
06	Lake View	. Boystown . Lake View East . Graceland West . South East Ravenswood . Wrigleyville
07	Lincoln Park	. Old Town Triangle . Park West . Ranch Triangle . Sheffield Neighbors . Wrightwood Neighbors
21	Avondale	. Belmont Gardens . Chicago's Polish Village . Kosciuszko Park
22	Logan Square	. Belmont Gardens . Bucktown . Kosciuszko Park . Palmer Square
Far North side		
Number	Community area	Neighborhoods
01	Rogers Park	. East Rogers Park
02	West Ridge	. Arcadia Terrace . Peterson Park . West Rogers Park
03	Uptown	. Buena Park . Argyle Street . Margate Park . Sheridan Park

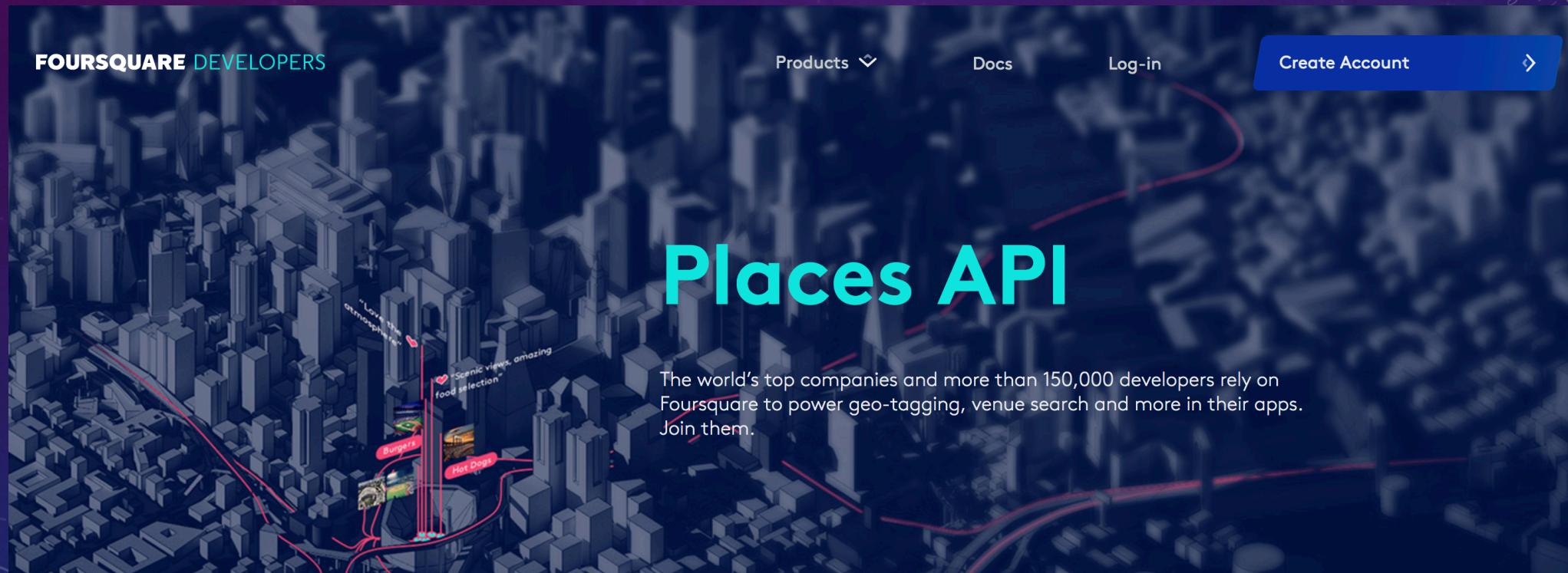
Community Areas	Crime Counts	Community Names
0	1	2706 Rogers Park
1	2	646 West Ridge
2	3	723 Uptown
3	4	3082 Lincoln Square
4	5	1505 North Center
5	6	2017 Lake View
6	7	2718 Lincoln Park
7	8	1971 Near North Side
8	9	3658 Edison Park
9	10	3569 Norwood Park

<https://data.cityofchicago.org/>

CHICAGO DATA PORTAL	
Crimes - 2018	
Based on Crimes - 2001 to present	
Community Area	FBI Code
67	11
74	14
71	14
31	08B
71	08B
25	04B
71	08B
5	08B
23	14
6	06
3	26
69	15
19	18
44	15
1	08B
8	06
6	11
16	14

DATA - REGARDING RELAXATION

Foursquare: is a technology company that built a massive dataset of location data.
We will use Foursquare API to explore the communities in Chicago.



METHODOLOGY

**1. Sort out the
safe
communities**

1.1 Data
acquisition and
cleaning

1.2 Sort out the
safe
communities

**2. Cluster these
safe
communities**

2.1 Explore each
community
candidate

2.2 Cluster
communities

METHODOLOGY – 1.1 DATA ACQUISITION AND CLEANING

Data Acquisition
↓
Data Cleaning
↓
Adding Geocode

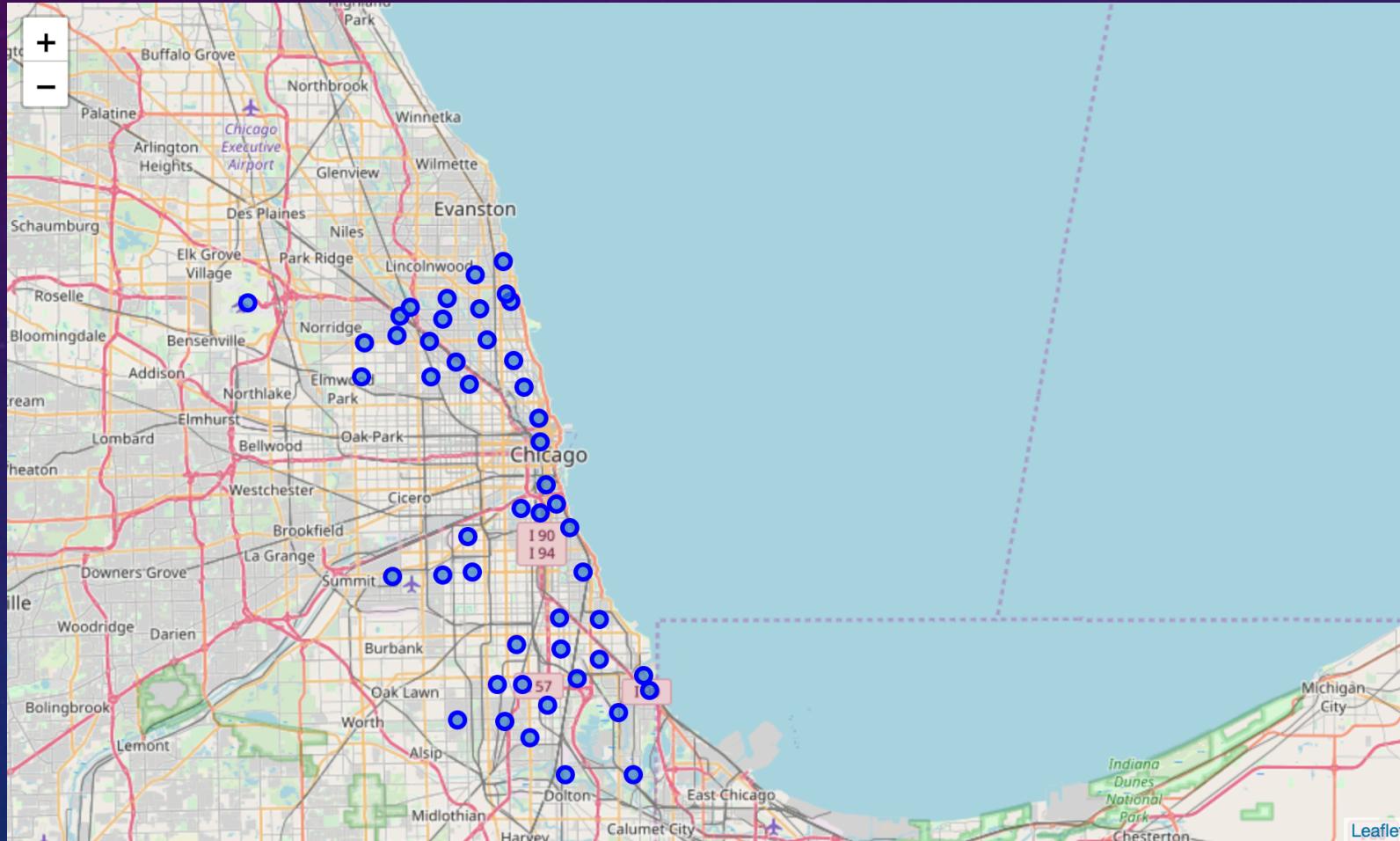


Community Areas	Crime Counts	Community Names	Latitude	Longitude	
0	1	2706	Rogers Park	42.008820	-87.666180
1	2	646	West Ridge	41.999480	-87.692660
2	3	723	Uptown	41.981230	-87.660000
3	4	3082	Lincoln Square	41.975700	-87.689140
4	5	1505	North Center	41.954110	-87.681420
5	6	2017	Lake View	41.939820	-87.656820
6	7	2718	Lincoln Park	41.921840	-87.647440
7	8	1971	Near North Side	41.900150	-87.634330
8	9	3658	Edison Park	42.007890	-87.813990
9	10	3569	Norwood Park	41.985470	-87.806110
10	11	1093	Jefferson Park	41.970460	-87.763460
11	12	474	Forest Glen	41.976400	-87.753610
12	13	989	North Park	41.982940	-87.719150
13	14	2395	Albany Park	41.968290	-87.723380
14	15	3183	Portage Park	41.957770	-87.765040

7

METHODOLOGY – 1.2 SORT OUT THE SAFE COMMUNITIES

Select the communities with less crime counts than the average of Chicago and show them on the map.
There are 49 safe communities in total.



METHODOLOGY – 2.1 EXPLORE EACH COMMUNITY CANDIDATE

Use the *Foursquare API* to explore community candidates in Chicago, to get the Top 10 most common venue categories in each community.

	Community Names	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Albany Park	Liquor Store	Asian Restaurant	Cosmetics Shop	Dance Studio	Park	Paper / Office Supplies Store	Bus Stop	Light Rail Station	Bar	Bank
1	Armour Square	Bus Station	Chinese Restaurant	Food Truck	Park	Diner	Dance Studio	Discount Store	Donut Shop	Rental Car Location	Road
2	Auburn Gresham	Pizza Place	Deli / Bodega	Intersection	Playground	Pharmacy	Park	Record Shop	Optical Shop	Chinese Restaurant	Sandwich Place
3	Avondale	Wings Joint	Mexican Restaurant	Financial or Legal Service	Pizza Place	Donut Shop	Filipino Restaurant	Field	Fast Food Restaurant	Farmers Market	Exhibit
4	Beverly	Park	Boutique	Donut Shop	Lounge	Fast Food Restaurant	Dry Cleaner	Financial or Legal Service	Filipino Restaurant	Field	Farmers Market

To see more details, please refer to https://github.com/Aribeth21/Coursera_Capstone/blob/master/Capstone_Project.ipynb

METHODOLOGY – 2.2 CLUSTER COMMUNITIES

Use *k-means* algorithm to segment the community candidates, with the feature of most common venue categories in each community.

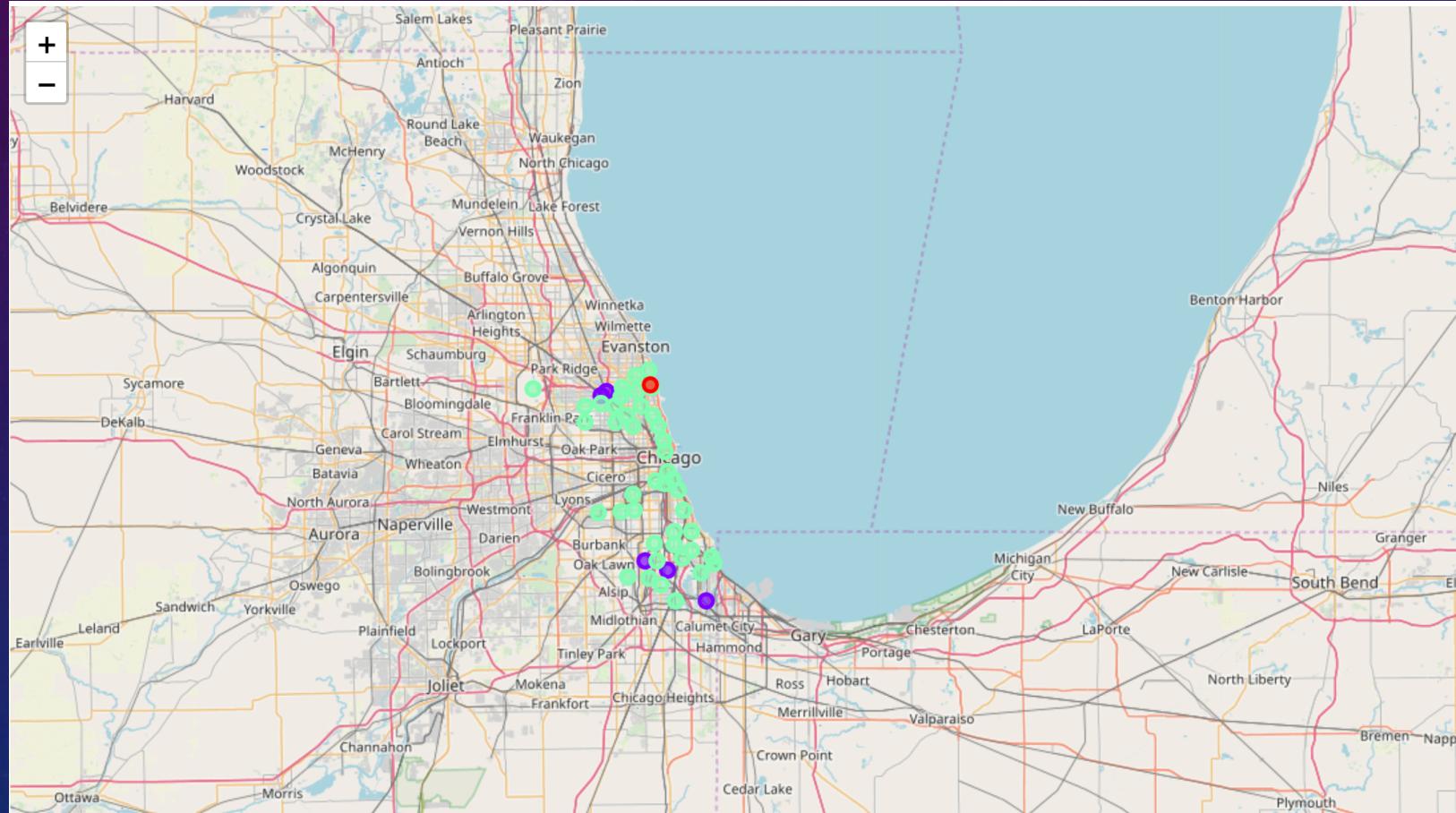
Based on our sample size (49), we decide to set the number of clusters to 3.

Community Areas	Crime Counts	Community Names	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	
0	1	2706	Rogers Park	42.00882	-87.66618	2	Theater	Pizza Place	American Restaurant	Bar	ATM	Gas Station	Mexican Restaurant	Sandwich Place	Café
1	2	646	West Ridge	41.99948	-87.69266	2	Indian Restaurant	Pakistani Restaurant	Grocery Store	Fast Food Restaurant	Football Stadium	Donut Shop	Market	Fruit & Vegetable Store	Clothing Store
2	3	723	Uptown	41.98123	-87.66000	2	Pizza Place	Sandwich Place	Sushi Restaurant	Asian Restaurant	Vietnamese Restaurant	Bus Station	Chinese Restaurant	Coffee Shop	Theater
3	4	3082	Lincoln Square	41.97570	-87.68914	2	Bus Station	Bar	Café	Convenience Store	Pizza Place	Korean Restaurant	Sandwich Place	Liquor Store	Food & Drink Shop
4	5	1505	North Center	41.95411	-87.68142	2	Bar	Coffee Shop	Bank	Mobile Phone Shop	Boutique	American Restaurant	Dive Bar	Pub	Pharmacy

To see more details, please refer to https://github.com/Aribeth21/Coursera_Capstone/blob/master/Capstone_Project.ipynb

METHODOLOGY - CLUSTER THESE SAFE COMMUNITIES

Let's display them on the map. We can see that there dots are consist of 3 colors, each color represents a cluster.



RESULTS - CLUSTER 1

Crime Counts	Community Names	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	
76	233	Edgewater	0	Intersection	Chinese Restaurant	Yoga Studio	Donut Shop	Financial or Legal Service	Filipino Restaurant	Field	Fast Food Restaurant	Farmers Market	Exhibit

As we can see, cluster 1 only contains 1 community. Based on its Top 10 Most Common Venue Categories, it looks like a community with a slow-paced life, as venues like Yoga Studio, Farmers Market, Exhibit are very common, although the 1st common venue is intersection, which may seem a little awkward.

RESULTS - CLUSTER 2

Crime Counts	Community Names	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
10	1093	Jefferson Park	1 Neighborhood	Park	Theater	Yoga Studio	Donut Shop	Filipino Restaurant	Field	Fast Food Restaurant	Farmers Market	Exhibit
11	474	Forest Glen	1 Park	Mexican Restaurant	Bar	Yoga Studio	Dry Cleaner	Financial or Legal Service	Filipino Restaurant	Field	Fast Food Restaurant	Farmers Market
48	1085	Roseland	1 Park	Gas Station	Liquor Store	Clothing Store	Seafood Restaurant	Dog Run	Field	Fast Food Restaurant	Farmers Market	Exhibit
54	2224	Hegewisch	1 Park	Food & Drink Shop	Bus Station	Yoga Studio	Donut Shop	Filipino Restaurant	Field	Fast Food Restaurant	Farmers Market	Exhibit
71	3113	Beverly	1 Park	Boutique	Donut Shop	Lounge	Fast Food Restaurant	Dry Cleaner	Financial or Legal Service	Filipino Restaurant	Field	Farmers Market

Cluster 2 looks exactly the type of community we are looking for. There are venues like parks, dry cleaners, fields, farmers markets, exhibits and of course restaurants, which perfectly meet our requirement of a relaxed, laid-back environment, and the parks and fields are 'must-haves' for family with kids.

RESULTS - CLUSTER 3

	Crime Counts	Community Names	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	2706	Rogers Park	2	Theater	Pizza Place	American Restaurant	Bar	ATM	Gas Station	Mexican Restaurant	Sandwich Place	Café	Donut Shop
1	646	West Ridge	2	Indian Restaurant	Pakistani Restaurant	Grocery Store	Fast Food Restaurant	Football Stadium	Donut Shop	Market	Fruit & Vegetable Store	Clothing Store	Juice Bar
2	723	Uptown	2	Pizza Place	Sandwich Place	Sushi Restaurant	Asian Restaurant	Vietnamese Restaurant	Bus Station	Chinese Restaurant	Coffee Shop	Theater	Mexican Restaurant
3	3082	Lincoln Square	2	Bus Station	Bar	Café	Convenience Store	Pizza Place	Korean Restaurant	Sandwich Place	Liquor Store	Food & Drink Shop	Karaoke Bar
4	1505	North Center	2	Bar	Coffee Shop	Bank	Mobile Phone Shop	Boutique	American Restaurant	Dive Bar	Pub	Pharmacy	Yoga Studio
5	2017	Lake View	2	Café	Japanese Restaurant	Bakery	Gym / Fitness Center	Pizza Place	Bagel Shop	Coffee Shop	Performing Arts Venue	Clothing Store	Sports Bar
6	2718	Lincoln Park	2	Pizza Place	Sandwich Place	Coffee Shop	Bar	Taco Place	Breakfast Spot	Fast Food Restaurant	Art Gallery	Mexican Restaurant	American Restaurant
7	1971	Near North Side	2	Gym / Fitness Center	Gym	Restaurant	American Restaurant	Coffee Shop	Cycle Studio	Breakfast Spot	Sandwich Place	Café	Pub
12	989	North Park	2	Coffee Shop	Pharmacy	Theater	Video Store	Bar	Food Truck	Sushi Restaurant	Fried Chicken Joint	Supermarket	Park
13	2395	Albany Park	2	Liquor Store	Asian Restaurant	Cosmetics Shop	Dance Studio	Park	Paper / Office Supplies Store	Bus Stop	Light Rail Station	Bar	Bank
14	3183	Portage Park	2	Bus Station	Breakfast Spot	Bike Rental / Bike Share	Bakery	Sandwich Place	Asian Restaurant	Coffee Shop	Food Truck	Radio Station	Convenience Store
15	2794	Irving Park	2	Discount Store	Accessories Store	Mexican Restaurant	Taco Place	Fried Chicken	Snack Place	Fast Food Restaurant	Pet Store	Seafood Restaurant	Mobile Phone Shop

The remaining 43 communities all belong to Cluster 3 according to k-means algorithm. Table 14 just takes a snap of these communities. By first look, it is a bit chaotic, but if we look closely, we could still find some common features in these communities. Most of them give us the impression of a fast-paced city life. For example, the most common venues include coffee shops, fast food restaurants, pizza places, bus stations, etc. They all remind me of the hustle bustle of downtown areas and CBDs. So, I don't think we'll recommend the communities in Cluster 3 to our friend.

DISCUSSION

- Based on the observations above, we'll recommend a community from Cluster 2. These communities perfectly meet his requirements of safety and relaxation.
- But exactly which one to choose is still up to the friend's preference. Say, if he thinks safety is the first priority, then Forest Glen with the least crime records should be his choice. Or if he works in south Chicago and would like to settle his family close by, then he should choose from Roseland, Hegewisch or Beverly.
- But as we discussed in the previous part, Cluster 1 which only includes 1 community called Edgewater could also be an option. Especially if the friend want to live near lake.

CONCLUSION

- In this project, we firstly narrow down our choices by select those safe communities based on crime records. Secondly, we use the Foursquare API to explore these communities and then use k-means algorithm to group the communities into clusters based on the feature of most common venues.
- We obtain 3 clusters and according to our observations, we recommend Cluster 2 to our friend, which meet his requirements the most.
- But, as we can see, there are still some issues regarding to this segmentation. For instance, Cluster 1 and 2 seems alike in terms of their most common venues and within Cluster 3 we can see that there are some communities more similar to Cluster 2. I think this is because the venue categories in Foursquare do not always correctly reflect our intension.
- So maybe for future improvement, we should explicitly hand-pick the venue categories that we need to solve the problem and try different clustering algorithms as well to get a better result.