Battle of the Neighborhood

Introduction

In recent years, more and more people have chances to study abroad, but some students generally have not gone to the destination country before, so that they have some anxiety, such as they are not sure whether the countries, cities or communities are safe. They have no idea, whether there are restaurants, banks, or supermarkets around the neighborhoods, and they do not know how the local people evaluate the community that they will move in. This article or .ipynb file uses the Foursquare API to help the potential students understand the culture and venues around their University. This article takes myself as an example, it is similar to use for other students. For example, I will go to New York University next year. Since I have a scholarship, I do not need to worry too much about my rent. So I mainly consider the safety of the cultures and venues in the neighborhood and the food around the location. Generally speaking, I prefer Chinese food, sushi, Italian Restaurants, pizzerias and ice cream stores, I want to live in a neighborhood that is closer to the university and fits my requirements well.

Who can benefit from the project?

- Students who will study abroad and want to find a good place to live
- Traveler who wants to make a better plan
- Employer who wants to book a hotel for business trip with a specified surroundings
- Other people who want to explore a specified place

Data Requirements and Description

Data that will be required for solving the problem:

1. Download New York City crime data from `https://data.cityofnewyork.us/Public-Safety/NYC-crime/qb7u-rbmr`, in this project we use it as "**NYPD_Complaint_Data_Historic.csv**", It gives all the crime data in New York.

2. Download the New York geojson file from `https://data.cityofnewyork.us/City-Government/Borough-Boundari` `tqmj-j8zm`, we use it as "**Borough_Boundaries.json**"to visualize the choropleth map, which gives a intuitionistic impression about the quantity of crimes in each borough of New York.

3. Download the New York neighborhood data from `https://data.cityofnewyork.us/City-Government/Borough-Boundaries/tqmj-j8zm`, we use it as "**newyork_data.json**"to get the data of all neighborhoods in New York.

4. "**crim_manh_data.csv**ïs the data of total crime numbers in each neighborhoods in Manhatten, this is a dataframe that I created myself by Data cleaning from **NYPD_Complaint_Data_Historic.csv**.

5. "**cny_crime_neigh.csv**ïs the data of crimes with information of neighborhoods, this is a dataframe that I created myself by Data cleaning from **NYPD_Complaint_Data_Historic.csv** and **newyork_data.json**.

Methodology

1 New York Map

First of all, I want to know the specific location of New York University. I only have a general impression of New York City. I know there are 5 Boroughs. Now I want to find the location of New York University. I use **geopy** to get the latitude and longitude of New York University and open it with the **folium** method. Now I get a map which is centered on New York University, and the red dot in Abbildung 1 is New York University, located in borough Manhattan.

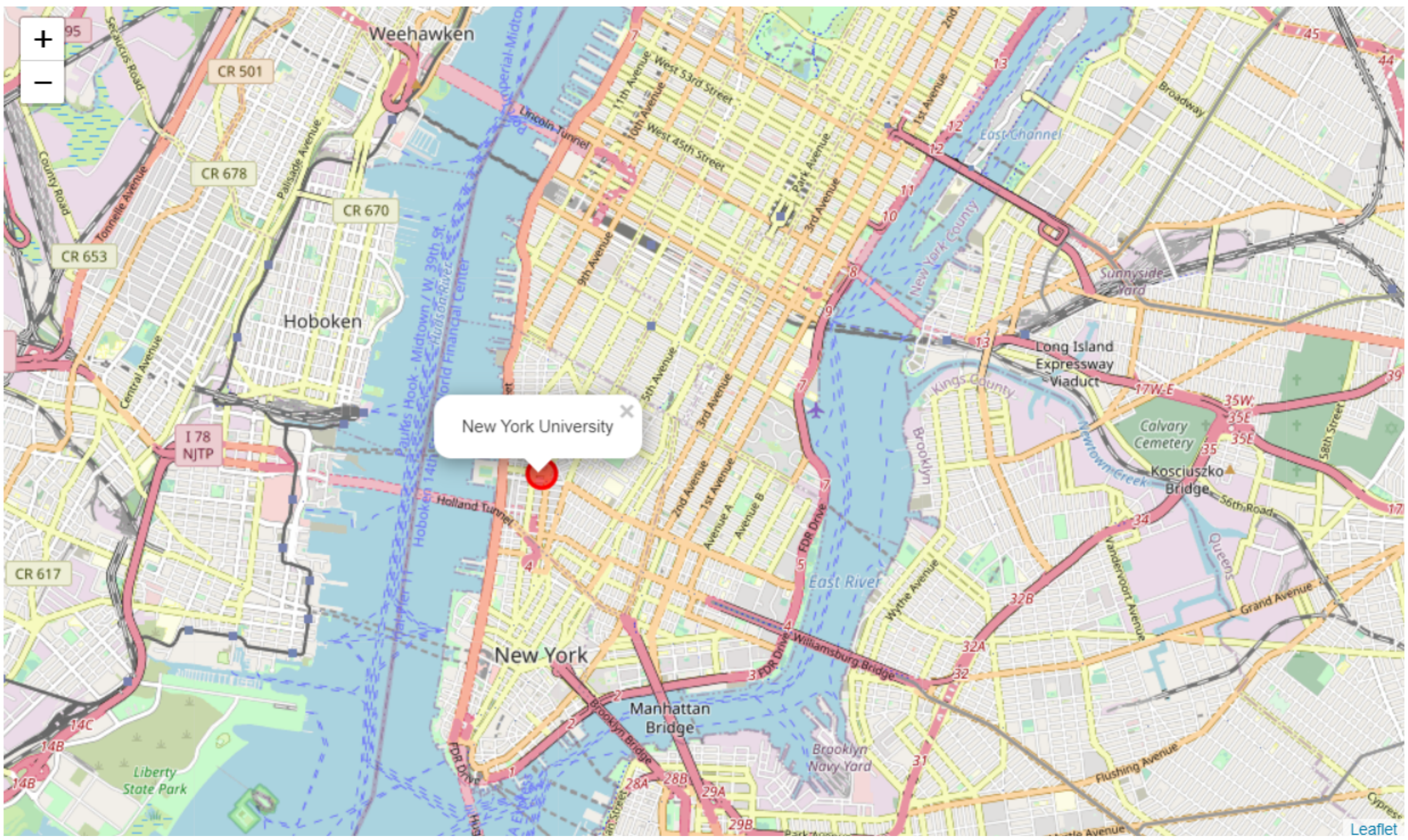

Abbildung 1: Location of New York University

2 NYC Crime Data

We read the crime data of New York from **NYPD_Complaint_Data_Historic.csv**, through Data Cleaning, select the crime data of the whole year of last year, 2019, we want to check how many crimes have occurred in each borough, and see which borough we can live so that we can be more safe. The result is in Abbildung 2:

There are 5 rows in the Dataframe, which looks not so intuitive.Naturally we thought of using choropleth map to represent the results, we use **choropleth** method to give the corresponding color to each borough. The

	Borough	Total	latitude	longitude
0	Bronx	173171	40.846651	-73.878594
1	Brooklyn	226484	40.650104	-73.949582
2	Manhattan	198241	40.789624	-73.959894
3	Queens	157290	40.749824	-73.797634
4	Staten Island	32443	40.583456	-74.149605

Abbildung 2: Crime in each borough of NYC

more dangerous the borough is, the darker is the color. we also use the **popup** method, so that we can get the number of crimes in the corresponding borough by clicking the mouse. Here is the Visualization (Abbildung 3).

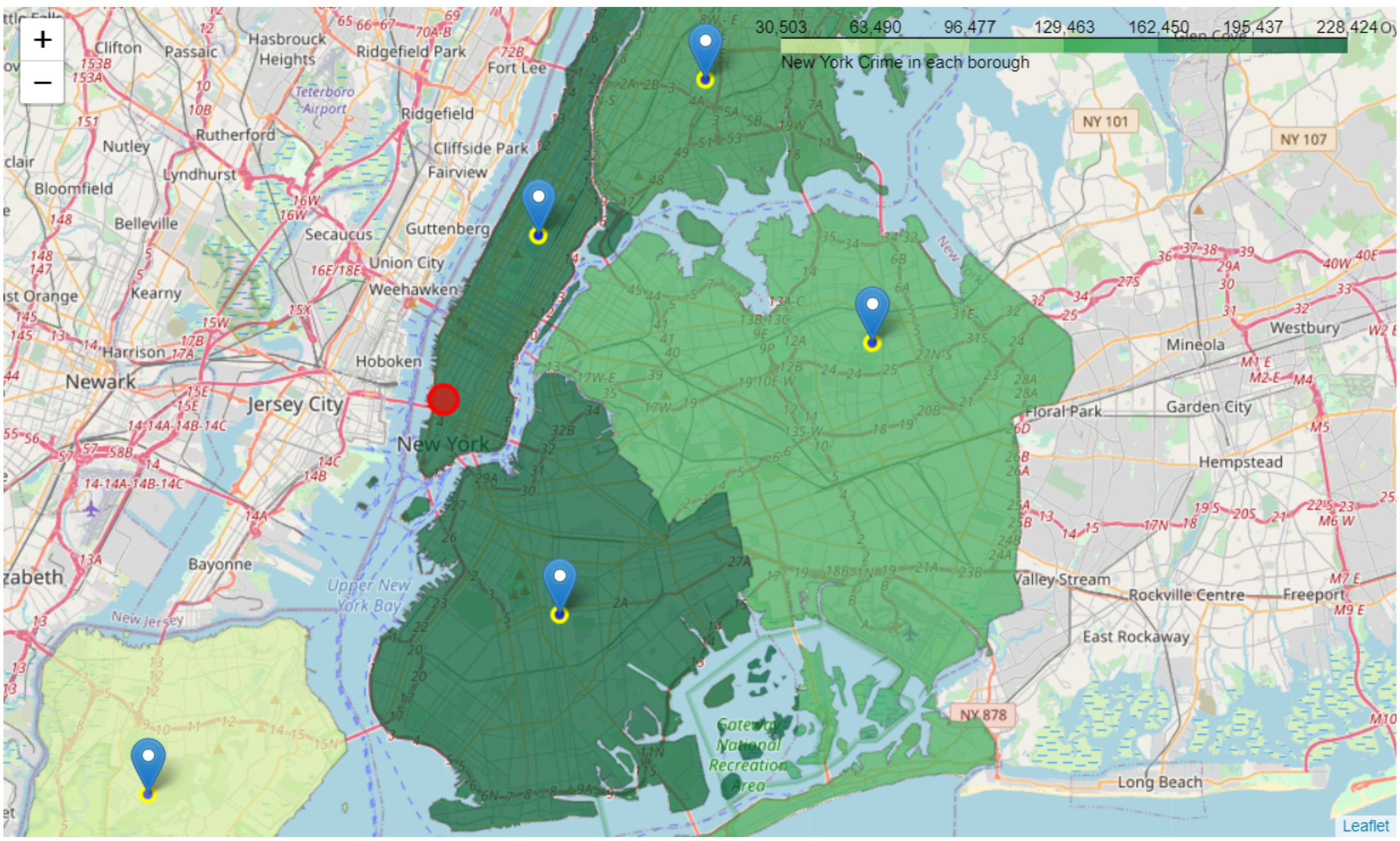

Abbildung 3: choropleth map for crime in each borough

We can clearly see from the figure that Staten Island has the fewest crimes. It can be speculated that this area is relatively remote and has relatively few people. Brooklyn has the most crimes. Manhattan and Bronx are similar, and Queens is less. But Queens is too far from the University. Based on these data, we find out that it is better to live in Manhattan. The number of crimes is similar to Bronx, but it is closer to the University. So, we will explore the neighborhoods in Manhatten.

We read the data from **ny_crime_neigh.csv** and find all the crime data in Manhattan through Data Cleansing, which contains all the neighborhoods, we have to choose a place to live in, of course, this cannot be our only reason to choose, we will continue to explore other characteristics of the neighborhoods, such as food, and then comprehensively, select the neighborhood we want to live in, we store the Dataframe(Abbildung 4) first, and prepare for the final choice.

	Neighborhood	Latitude	Longitude	Total
0	Battery Park City	40.708067	-74.010176	3386
1	Central Harlem	40.812100	-73.945226	11666
2	Chinatown	40.717211	-73.992731	3697
3	Clinton	40.761623	-73.990241	5306
4	East Harlem	40.796745	-73.940868	11789
5	East Village	40.727226	-73.985959	2643
6	Gramercy	40.736526	-73.983767	2365
7	Hamilton Heights	40.826026	-73.947521	3379
8	Hudson Yards	40.743988	-73.996093	8990
9	Lenox Hill	40.766140	-73.957565	1883
10	Lincoln Square	40.774318	-73.983275	3117
11	Lower East Side	40.718234	-73.981860	4572
12	Manhattanville	40.817060	-73.954483	2176
13	Marble Hill	40.867986	-73.917868	3244
14	Midtown-Midtown South	40.754522	-73.984904	11742
15	Morningside Heights	40.807648	-73.960606	2752
16	Murray Hill	40.743632	-73.977666	2759
17	Soho	40.719880	-74.002647	5501
18	Stuyvesant Town	40.732694	-73.976938	181
19	Turtle Bay	40.755019	-73.970609	3093
20	Upper East Side	40.772455	-73.960921	3800

Abbildung 4: Crimes in Manhattan neighborhoods

3 Neighborhood around University

Now let's explore the neighborhoods of Manhattan. We download and read the **newyork_data.json** file to get the names and latitude and longitude of each neighborhood. We can find that the latitude and longitude of each block are quite similar, but obviously, some neighbohoods are still far from the university But according the dataframe, the distance is not easy to read. In order to solve the distance problem, after clearing the table, we define a **distance** function and use the Haversine formula to calculate the distance of each neighborhood coordinate from the university coordinate. By sorting the dataframe, we get the nearest top 5 neighborhoods to the university using **head()** attribute. The following dataframe 5 shows the result.

	Borough	Neighborhood	Latitude	Longitude	Distance to NYU(km)
0	Manhattan	Greenwich Village	40.726933	-73.999914	0.575227
1	Manhattan	West Village	40.734434	-74.006180	0.576233
2	Manhattan	Soho	40.722184	-74.000657	0.906546
3	Manhattan	Tribeca	40.721522	-74.010683	0.945405
4	Manhattan	Little Italy	40.719324	-73.997305	1.325842

Abbildung 5: neighbohoods closest to University

According to the dataframe, we found that 33 neighborhoods are within 10km from the university, and 6 neighborhoods are at least 10 km far away from that. I do n't want to live too far from the University, so my following exploration is limited to these 33 neighborhoods , The Abbildung 6 below are the rest that are no longer considered.

	Borough	Neighborhood	Latitude	Longitude	Distance to NYU(km)
0	Manhattan	Manhattanville	40.816934	-73.957385	10.574777
1	Manhattan	Central Harlem	40.815976	-73.943211	10.998178
2	Manhattan	Hamilton Heights	40.823604	-73.949688	11.513660
3	Manhattan	Washington Heights	40.851903	-73.936900	14.827319
4	Manhattan	Inwood	40.867684	-73.921210	16.967513
5	Manhattan	Marble Hill	40.876551	-73.910660	18.239554

Abbildung 6: neighbohoods not be considered any more due to distance

Now we make further decisions, excluding the neighborhoods that are far away, they are indicated by small red dots on the map 7, and blue dots indicate the neighborhoods within 10km from the university, and the **popup** method is also used to display the names of the neighborhoods for viewing.

Now, to make further decisions in these closed neighborhoods, I like delicious food, so I use restaurants in these neighborhoods as a standard, to decide where to live. First, I want to see what types of venues (in Abbildung 8) are provided using **Foursquare API**. Among all categories, food is used as a further criterion, but other audiences can use more venues types as selection criteria.

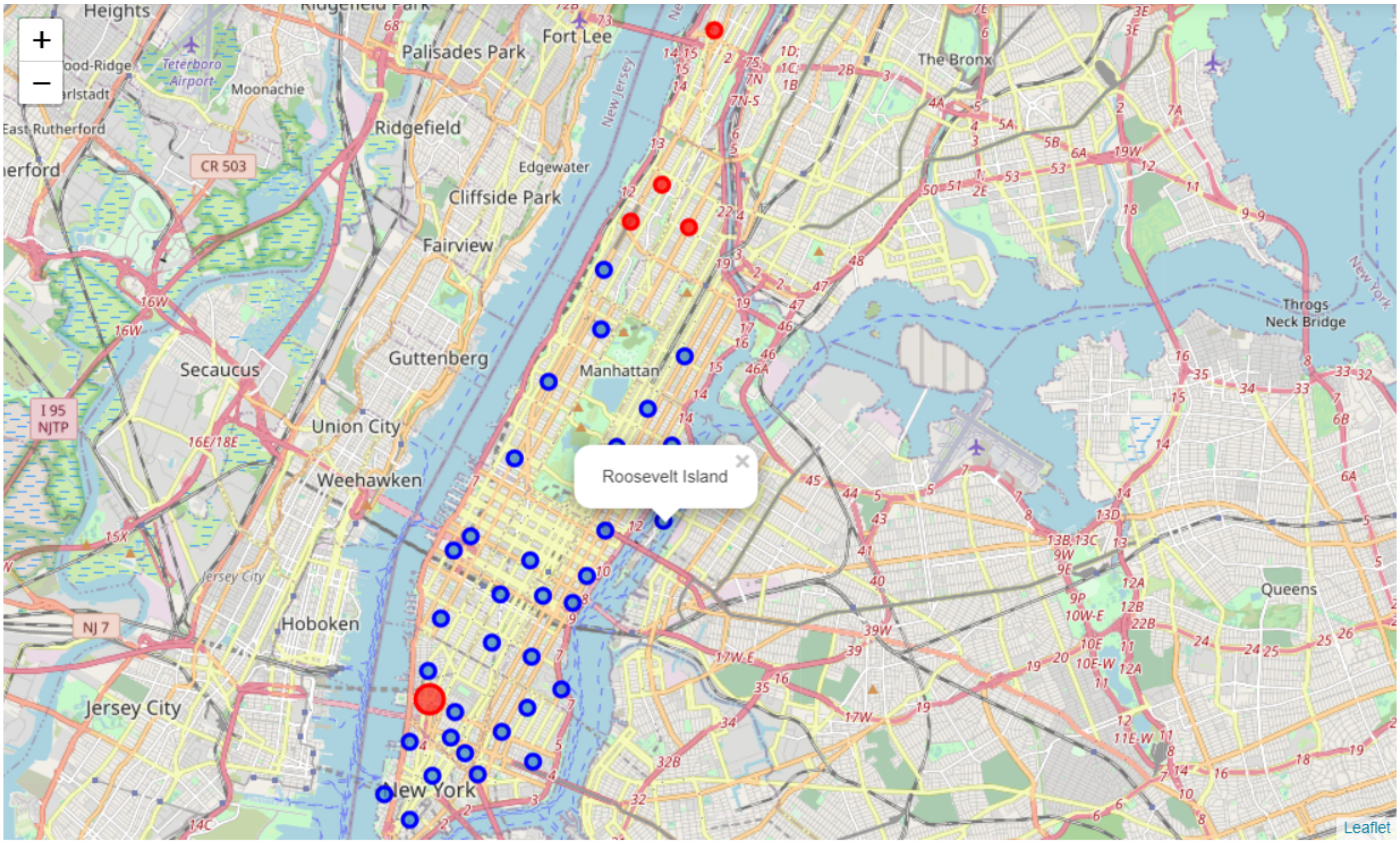

Abbildung 7: neighbohoods visualization, blue dots to explore, red dots to give up

Let 's explore neighborhoods using the data provided by **Foursquare** with the coordinates of each neighborhood as the center of a circle and a radius of 1km. We have defined a **getNearbyFood** function. For the audiences with other requirements, as long as the **categoryId** in the function is changed to the ID Abbildung 8. In the end, we get a dataframe of size (1650,7). Note that this size will change every time we run the program because the data provided by Foursquare may change frequently. We continue to clean the dataframe and assign values to each neighborhood according to the frequency of venues to get the top 10 food vebues with the highest frequency in their neighborhood as follows 9.

Only 13 neighborhoods of data are shown here, actually there are 33 rows of course. Now we use these dummies data as features and use the **K-Means** algorithm to cluster these neighborhoods. But we don't know how many clusters to divide into, here we use **The Silhouette Method**, The silhouette value measures how similar a point is to its own cluster (cohesion) compared to other clusters (separation), we first assume that k can be from 2 to 30 , through a for-loop detection, the following Abbildung 10 is drawn to choose a better k.

We can see from Abbildung 10 that there is a peak, also local maximum when k is equal to 10, we choose $k = 10$, use K-Means for clustering, and get the following Abbildung 11.

If you observe carefully at each cluster, you can find out that each cluster has indeed similar food venues. The clustering with $k = 10$ is pretty reasonable. I personly prefer Chinese food, sushi, Italian restaurants, pizzerias, and coffee shops. So I can easily select neighborhoods in cluster 3, the image below 12 shows all the neighborhoods included in cluster 3.

```
4d4b7104d754a06370d81259 Arts & Entertainment
4d4b7105d754a06372d81259 College & University
4d4b7105d754a06373d81259 Event
4d4b7105d754a06374d81259 Food
4d4b7105d754a06376d81259 Nightlife Spot
4d4b7105d754a06377d81259 Outdoors & Recreation
4d4b7105d754a06375d81259 Professional & Other Places
4e67e38e036454776db1fb3a Residence
4d4b7105d754a06378d81259 Shop & Service
4d4b7105d754a06379d81259 Travel & Transport
```

ABBILDUNG 8: get the categories of venues provided by Foursquare

Fortunately, the neighborhoods in cluster 3 are relatively close to the university. Next, I will choose a relative safe area among these neighborhoods to live in. We read in the clean data processed in part 2, "**crim_manh_data.csv**", which provided the crime data with information of neighborhoods, the four neighborhoods: SoHo, TriBeCa, Civic Center, and Little Italy are considered as a whole area in the dataframe. This is not a problem, because they are very close to each other. Let 's check the total number of crimes in each 3 aera: Chinatown, Lower East Side, SoHo-TriBeCa-Civic Center-Little Italy. The result is shown in the Abbildung 13 below.

Chinatown has the least number of crimes, so for comprehensive consideration, I plan to live in chinatown. Let'go further now, I want to find the categories of crimes in Chinatown, and check some serious crimes, such as drugs or shooting case. Because if such crimes categories take a large proportion, then Chinatown is the mostdangerous place, although the total number of crimes is the least. The result is in Abbildung 14

Results

It can be seen from the figure 14 that the most types of crimes are Petit Larceny, and the proportion of some dangerous crimes such as drug crimes and robbery is not large. Therefore, considering the number of crimes and the categories, the distance from the University and the surrounding food venues, **I will live in Chinatown**.

Discussion

After completing this project, I found that Jupyternotebook runs slower than Python IDE,for example Pycharm, but the run result for dataframe is more intuitive. In addition, when you want to compare the amount, use bar charts, when you want to check the proportion, the pie chart is better.

Conclusion

This project is not only helpful for me to choose a neighorhood to live, but also for the previous types of people meintioned in Introduction. For the selection of surrounding venues, just by changing the ID which is already given in this article, you can select different venues that are important for different kinds of people. The venues can also be not limited in New York, for other places, just change a local .csv file and .json file, then you can explore them in the same way shown in this article.

	Neighborhood	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	Battery Park City	Coffee Shop	Burger Joint	Deli / Bodega	Fast Food Restaurant	Donut Shop	Food Court	Mexican Restaurant	Sandwich Place	Mediterranean Restaurant	American Restaurant
1	Chelsea	Coffee Shop	Bakery	Fast Food Restaurant	Bagel Shop	Mexican Restaurant	BBQ Joint	Wine Bar	Café	Pizza Place	Grocery Store
2	Chinatown	Chinese Restaurant	Bakery	Bubble Tea Shop	Italian Restaurant	Pizza Place	Malay Restaurant	Sandwich Place	Coffee Shop	Food Court	Sushi Restaurant
3	Civic Center	Chinese Restaurant	Coffee Shop	Bakery	Food Court	Grocery Store	Salad Place	Sandwich Place	Mexican Restaurant	Bubble Tea Shop	Taco Place
4	Clinton	Coffee Shop	Donut Shop	Fast Food Restaurant	Mexican Restaurant	Pizza Place	Bubble Tea Shop	Food Court	Burger Joint	Taco Place	Restaurant
5	East Harlem	Deli / Bodega	Fast Food Restaurant	Café	Coffee Shop	American Restaurant	Donut Shop	African Restaurant	Bagel Shop	Italian Restaurant	Sandwich Place
6	East Village	Coffee Shop	Bakery	Pizza Place	Deli / Bodega	Bagel Shop	Fried Chicken Joint	Sushi Restaurant	American Restaurant	Sandwich Place	Café
7	Financial District	Coffee Shop	Food Court	Fast Food Restaurant	Sandwich Place	Donut Shop	Burger Joint	Deli / Bodega	Mexican Restaurant	Taco Place	Restaurant
8	Flatiron	Coffee Shop	Donut Shop	Fast Food Restaurant	Bagel Shop	Fried Chicken Joint	Grocery Store	Deli / Bodega	Ice Cream Shop	Café	Burger Joint
9	Gramercy	Coffee Shop	Bagel Shop	American Restaurant	Deli / Bodega	Café	Bakery	Fast Food Restaurant	Ice Cream Shop	Donut Shop	Japanese Restaurant
10	Greenwich Village	Coffee Shop	Bakery	Pizza Place	Italian Restaurant	Sandwich Place	Donut Shop	Bagel Shop	Sushi Restaurant	Japanese Restaurant	New American Restaurant
11	Hudson Yards	Coffee Shop	Café	Donut Shop	Mexican Restaurant	Pizza Place	Burger Joint	Fast Food Restaurant	Deli / Bodega	Food Court	American Restaurant
12	Lenox Hill	Coffee Shop	Salad Place	Bar	Bakery	Burger Joint	Diner	Deli / Bodega	Bagel Shop	Italian Restaurant	Ice Cream Shop

ABBILDUNG 9: get the top 10 food venues for each neighborhood

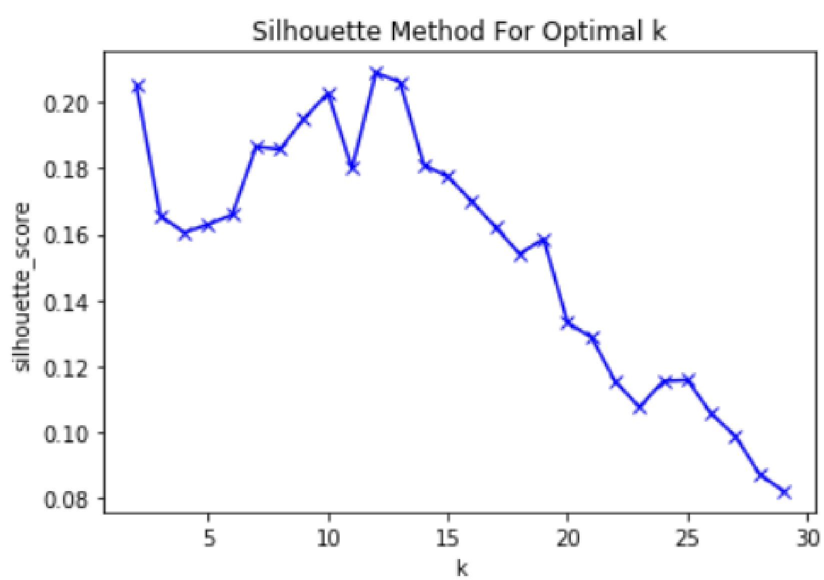

ABBILDUNG 10: result of silhouette value to choose k

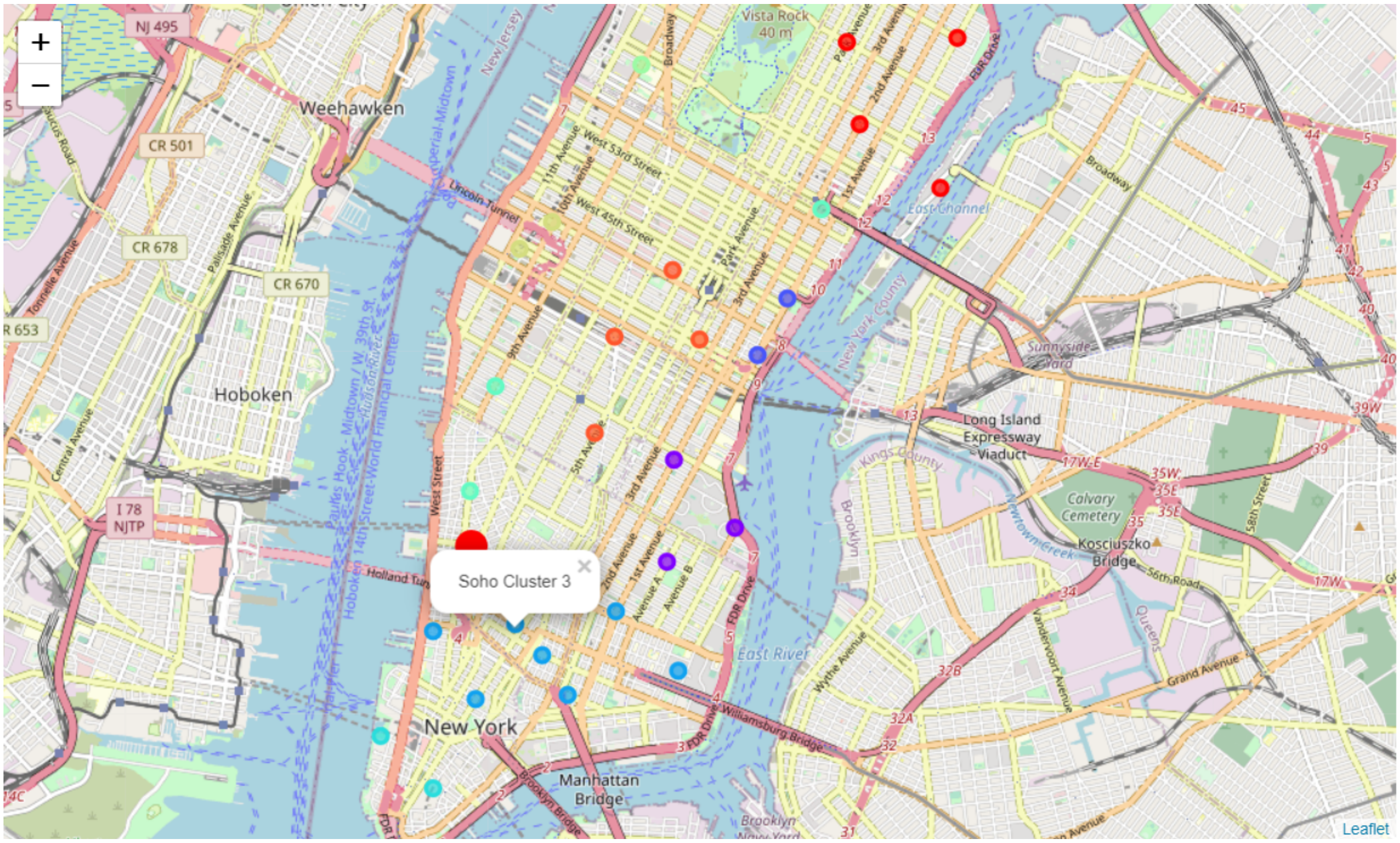

Abbildung 11: neighborhood clustering using K-Means based on food venues

	Neighborhood	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	Distance to NYU(km)
2	Chinatown	Chinese Restaurant	Bakery	Bubble Tea Shop	Italian Restaurant	Pizza Place	Malay Restaurant	Sandwich Place	Coffee Shop	Food Court	Sushi Restaurant	1.810183
3	Civic Center	Chinese Restaurant	Coffee Shop	Bakery	Food Court	Grocery Store	Salad Place	Sandwich Place	Mexican Restaurant	Bubble Tea Shop	Taco Place	1.560254
14	Little Italy	Bakery	Coffee Shop	Chinese Restaurant	Sushi Restaurant	Ice Cream Shop	Bubble Tea Shop	Sandwich Place	Fast Food Restaurant	Pizza Place	Italian Restaurant	1.325842
15	Lower East Side	Coffee Shop	Bakery	Pizza Place	Chinese Restaurant	Sushi Restaurant	Sandwich Place	Food Court	Bubble Tea Shop	Bagel Shop	Cocktail Bar	2.470414
21	Noho	Bakery	Coffee Shop	Pizza Place	Sandwich Place	Fried Chicken Joint	Sushi Restaurant	Italian Restaurant	Fast Food Restaurant	Chinese Restaurant	Hot Dog Joint	1.624577
23	Soho	Italian Restaurant	Coffee Shop	Sandwich Place	Bakery	Chinese Restaurant	Sushi Restaurant	Ice Cream Shop	Pizza Place	Salad Place	Bubble Tea Shop	0.906546
26	Tribeca	Coffee Shop	Sandwich Place	Burger Joint	Pizza Place	Bakery	Salad Place	Mexican Restaurant	Bubble Tea Shop	Italian Restaurant	French Restaurant	0.945405

Abbildung 12: choose cluster 3 according the food venues I prefer

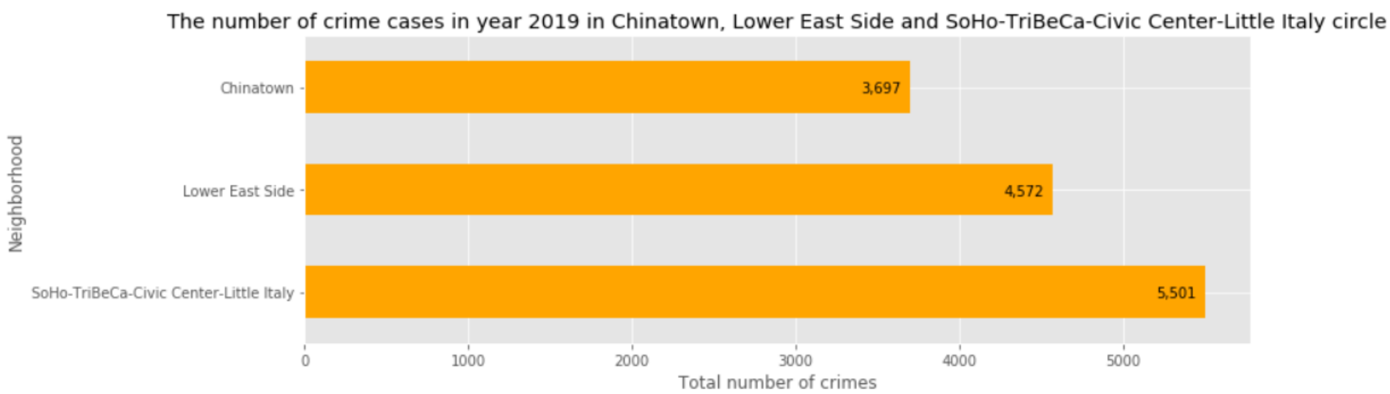

ABBILDUNG 13: the total number of crimes in Chinatown, Lower East Side, SoHo-TriBeCa-Civic Center-Little Italy

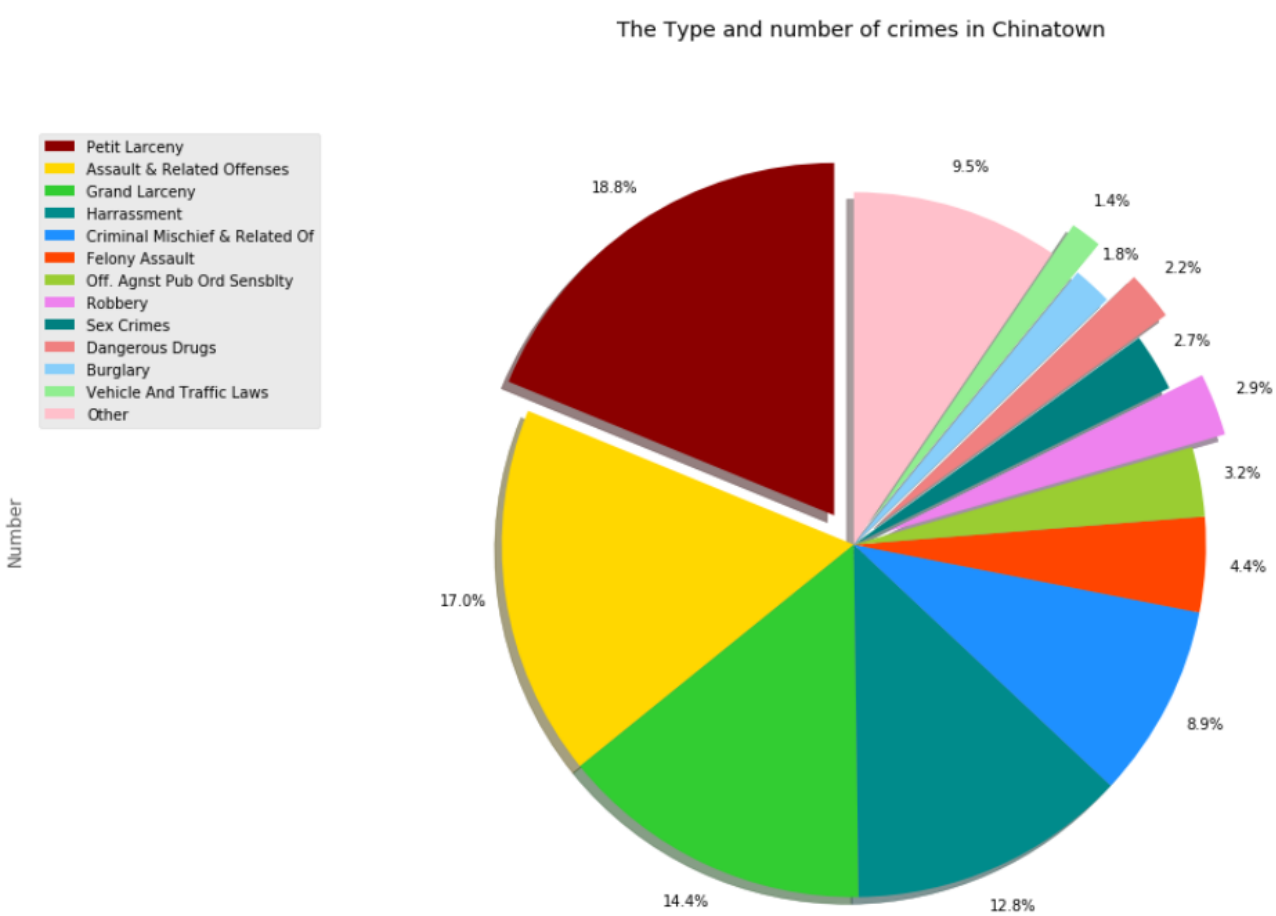

ABBILDUNG 14: categories of crimes in Chinatown