Coursera/IBM Data Science Program

Project Report: THE BATTLE OF NEIGHBORHOODS

Di Wang

Content

Chapitre 1

Introduction

1.1 Project Description

This is for the final project of the Data Science Specialization. A 9-courses series created by IBM, hosted on Coursera platform. The problem and the analysis approach are left for the learner to decide, with a requirement of leveraging the Foursquare location data to explore or compare neighborhoods or cities of your choice or to come up with a problem that you can use the Foursquare location data to solve.

1.2 Promblem Description

Restaurants are a driving force in Illinois's economy. They provide jobs and build careers for thousands of people, and play a vital role in local communities throughout the state. According to Illinois Restaurant Association[1], Every dollar spent in the table service segment contributes $\$1.93$ to the state economy. Every dollar spent in the limited-service segment contributes $\$1.66$ to the state economy.

FIGURE 1.1 – Illinois Restaurant Association

Besides, Illinois's 474,500 eating-and-drinking-place jobs represent the majority of the state's total restaurant and food service workforce of 588,700 jobs, with the remainder being non-restaurant food service positions.

	U.S. SENATORS	EATING AND DRINKING PLACES: Establishments in the state	EATING AND DRINKING PLACES: Employees in the state*
	Richard Durbin (D) Tammy Duckworth (D)	25,488	474,500
	U.S. REPRESENTATIVES	**Establishments in the state**	**Employees in the state***
1	Bobby L. Rush (D)	1,067	19,857
2	Robin Kelly (D)	927	17,252
3	Daniel Lipinski (D)	1,295	24,109
4	Jesús G. García (D)	1,152	21,450
5	Mike Quigley (D)	2,074	38,603
6	Sean Casten (D)	1,613	30,038
7	Danny K. Davis (D)	2,398	44,634
8	Raja Krishnamoorthi (D)	1,444	26,878
9	Jan Schakowsky (D)	1,655	30,807
10	Bradley Scott Schneider (D)	1,427	26,575
11	Bill Foster (D)	1,132	21,070
12	Mike Bost (R)	1,291	24,041
13	Rodney Davis (R)	1,568	29,197
14	Lauren Underwood (D)	1,226	22,830
15	John Shimkus (R)	1,191	22,163
16	Adam Kinzinger (R)	1,351	25,159
17	Cheri Bustos (D)	1,312	24,418
18	Darin LaHood (R)	1,365	25,420
	TOTAL	**25,488**	**474,500**

FIGURE 1.2 – Illinois Restaurant Association

The main goal will be exploring the neighborhoods of Chicago city in order to analyze the relationships among population, incomes, race, rental fees, and restaurant numbers.

Our goals can be summarized as follows :

1.List and visualize all major parts of Chicago city that has great restaurants.

2.Visualize the relationships among population, incomes, race, rental fees, and restaurants numbers.

3.Which areas show potential for new restaurants market ? (having the similar features with the best locations but fewer restaurants numbers)

1.3 Interest

Our latent audiences include :

1. Chicago Restaurant Inspection Department. This department needs to know the details of restaurants distribution in Chicago City and its relationship with other features(population, rates, etc).
2. New restaurant openers who have problems in finding ideal restaurants positions.
3. Existing restaurants owners who are hesitate about whether to invest more money in restaurants updating and extending delivery services.

Chapitre 2

Data acquisition and cleaning

2.1 Data sources

Chicago data set contains neighborhood name, population, income, latinos, blacks, white, asian, other race can be found at link[2]. And the average house rent fees for each neighborhoods can be scraped at website[3]. However, based on the difference of Chicago area definition, the neighborhoods data in these two data set are not matched. And there are some missing data in the house rent fees data set. Restaurants data is available through the Foursquare API.

2.2 Data preprocessing

It is necessary to take a glance at the data shape and type before later processing.

	neighborhood	population	income	latinos	blacks	white	asian	other
0	Rogers Park	54991	39482	0.244	0.263	0.393	0.064	0.036
1	West Ridge	71942	47323	0.204	0.111	0.427	0.225	0.032

FIGURE 2.1 – Data Shape

```
neighborhood     object
population       object
income           object
latinos          object
blacks           object
white            object
asian            object
other            object
Latitude        float64
Longitude       float64
dtype: object
```

FIGURE 2.2 – Data type

Based on the Fig. 2.1, the geographical data is required to be added with python geo API. Besides, according to the Fig. 2.2, some of the object type are necessary to be changed into float type.

To understand our data deeply, we plot the histogram figures from multiple viewpoints.

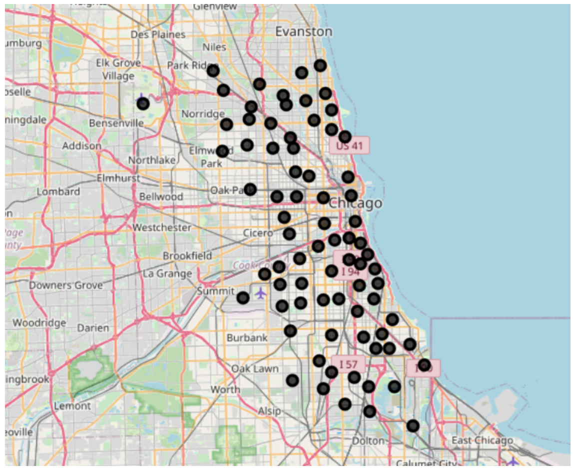

FIGURE 2.3 – Geo Map of Chicago Neighborhoods

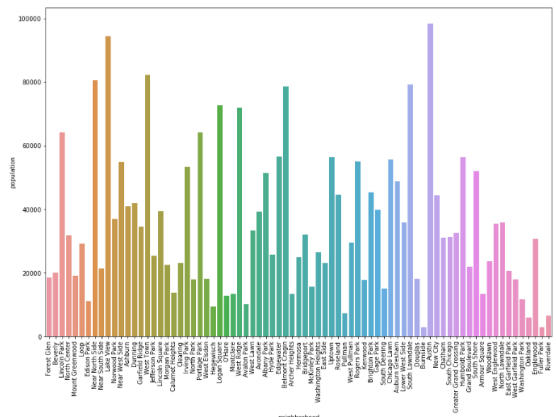

FIGURE 2.4 – Population distribution

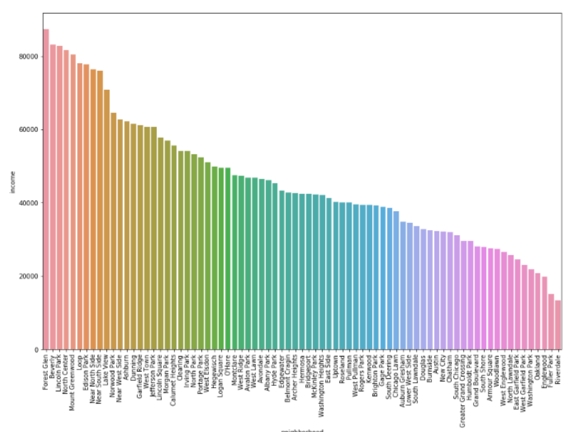

FIGURE 2.5 – Incomes distribution

According to the above figures, we find that the population feature and the incomes feature are not in the linear relationship. For instance, Forest Glen neighborhood has low population but high incomes. Lake View neighborhood has the same high population and incomes. One main influence factor beneath is the geographical location, which should be considered in our later analysis.

```
( 1 / 77 ) Resturants in Rogers Park, :22
( 2 / 77 ) Resturants in West Ridge, :17
( 3 / 77 ) Resturants in Uptown, :34
( 4 / 77 ) Resturants in Lincoln Square, :12
( 5 / 77 ) Resturants in North Center, :23
( 6 / 77 ) Resturants in Lake View, :17
( 7 / 77 ) Resturants in Lincoln Park, :16
( 8 / 77 ) Resturants in Near North Side, :20
( 9 / 77 ) Resturants in Edison Park, :9
( 10 / 77 ) Resturants in Norwood Park, :5
( 11 / 77 ) Resturants in Jefferson Park, :9
( 12 / 77 ) Resturants in Forest Glen, :4
( 13 / 77 ) Resturants in North Park, :8
( 14 / 77 ) Resturants in Albany Park, :14
( 15 / 77 ) Resturants in Portage Park, :7
( 16 / 77 ) Resturants in Irving Park, :7
( 17 / 77 ) Resturants in Dunning, :6
( 18 / 77 ) Resturants in Montclare, :7
( 19 / 77 ) Resturants in Belmont Cragin, :14
( 20 / 77 ) Resturants in Hermosa, :8
( 21 / 77 ) Resturants in Avondale, :16
( 22 / 77 ) Resturants in Logan Square, :26
( 23 / 77 ) Resturants in Humboldt Park, :9
```

FIGURE 2.6 – Restaurants Data

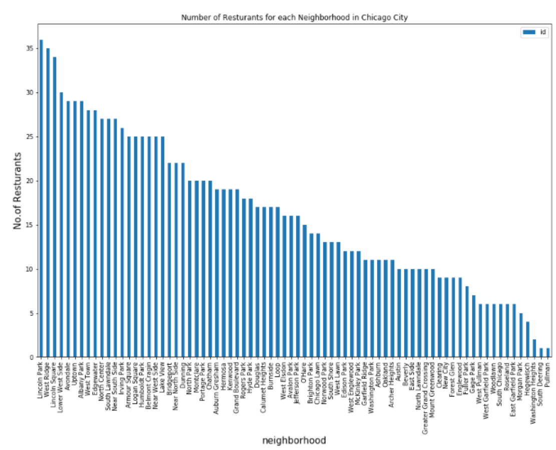

FIGURE 2.7 – Restaurants Data

After connecting with Foursquare API, we obtain the restaurants data in each neighborhoods. The gap of the number of restaurants in different neighborhoods is high. Because the restriction of the free version API, we set the radius at 800 meters. For instance, there are 35 restaurants within the 800 meters of the neighborhood Lincoin Park center. Though comparison with all above figures, we find that the place with most restaurants are different from the place with largest population and that with highest incomes. Thus, we need to consider other features like house rent fees or races.

The top ten neighborhoods with the most restaurants are as follows :

2	Uptown
76	Edgewater
23	West Town
32	Near South Side
21	Logan Square
40	Hyde Park
4	North Center
30	Lower West Side
0	Rogers Park
7	Near North Side

FIGURE 2.8 – Top 10 Neighborhoods

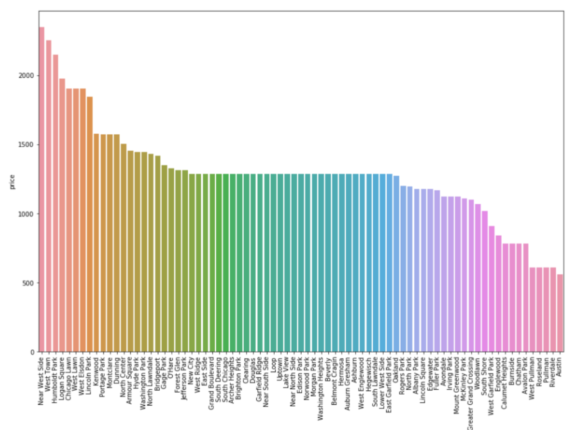

FIGURE 2.9 – House Rent Fees

After obtain the price data from the website, we select the house rent data we need and fill the missing data with mean values. Similarly, the rent fees differences in different neighborhoods are large.

Fig. **??**.

Chapitre 3

Data Analysis

3.1 Data Visualization

Before selecting the reasonable method to cluster our neighborhoods, it is essential to visualize our combined data set as Fig. 3.1.

	neighborhood	population	income	latinos	blacks	white	asian	other	Latitude	Longitude	price	restaurants
0	Rogers Park	54991.0	39482.0	0.244	0.263	0.393	0.064	0.036	42.010531	-87.670748	1199.000000	22.0
1	West Ridge	71942.0	47323.0	0.204	0.111	0.427	0.225	0.032	42.003548	-87.696243	1288.520833	17.0
2	Uptown	56362.0	40324.0	0.142	0.200	0.516	0.114	0.028	41.966630	-87.655546	1288.520833	34.0
3	Lincoln Square	39493.0	57749.0	0.191	0.038	0.631	0.111	0.029	41.975990	-87.689616	1180.000000	12.0
4	North Center	31867.0	81524.0	0.136	0.023	0.773	0.045	0.022	41.956107	-87.679160	1504.000000	23.0

FIGURE 3.1 – Our Combined Data

The mathematics analysis of our data set is as follows :

	neighborhood	population	income	latinos	blacks	white	asian	other	Latitude	Longitude	price	restaurants
count	77	77.000000	77.000000	77.000000	77.000000	77.000000	77.000000	77.000000	77.000000	77.000000	77.000000	77.000000
unique	77	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
top	Rogers Park	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
freq	1	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
mean	NaN	35007.766234	45823.064935	0.255403	0.391805	0.282688	0.054104	0.015948	41.840406	-87.675310	1288.520833	9.246753
std	NaN	22400.350739	17571.879139	0.281042	0.402650	0.281802	0.102284	0.008530	0.099928	0.069661	349.628411	8.384126
min	NaN	2876.000000	13380.000000	0.007000	0.003000	0.003000	0.000000	0.003000	41.653646	-87.906768	562.000000	0.000000
25%	NaN	18109.000000	32553.000000	0.035000	0.033000	0.021000	0.002000	0.010000	41.766886	-87.718388	1178.000000	3.000000
50%	NaN	31028.000000	42418.000000	0.115000	0.143000	0.165000	0.010000	0.013000	41.831700	-87.666762	1288.520833	7.000000
75%	NaN	48743.000000	55669.000000	0.453000	0.909000	0.515000	0.064000	0.021000	41.931698	-87.624774	1330.000000	12.000000
max	NaN	98514.000000	87394.000000	0.892000	0.978000	0.884000	0.726000	0.041000	42.010531	-87.532781	2351.000000	34.000000

FIGURE 3.2 – Mathematics Analysis

One essential studying aspect in data visualization and model selection is to analyze the correlation of different features.

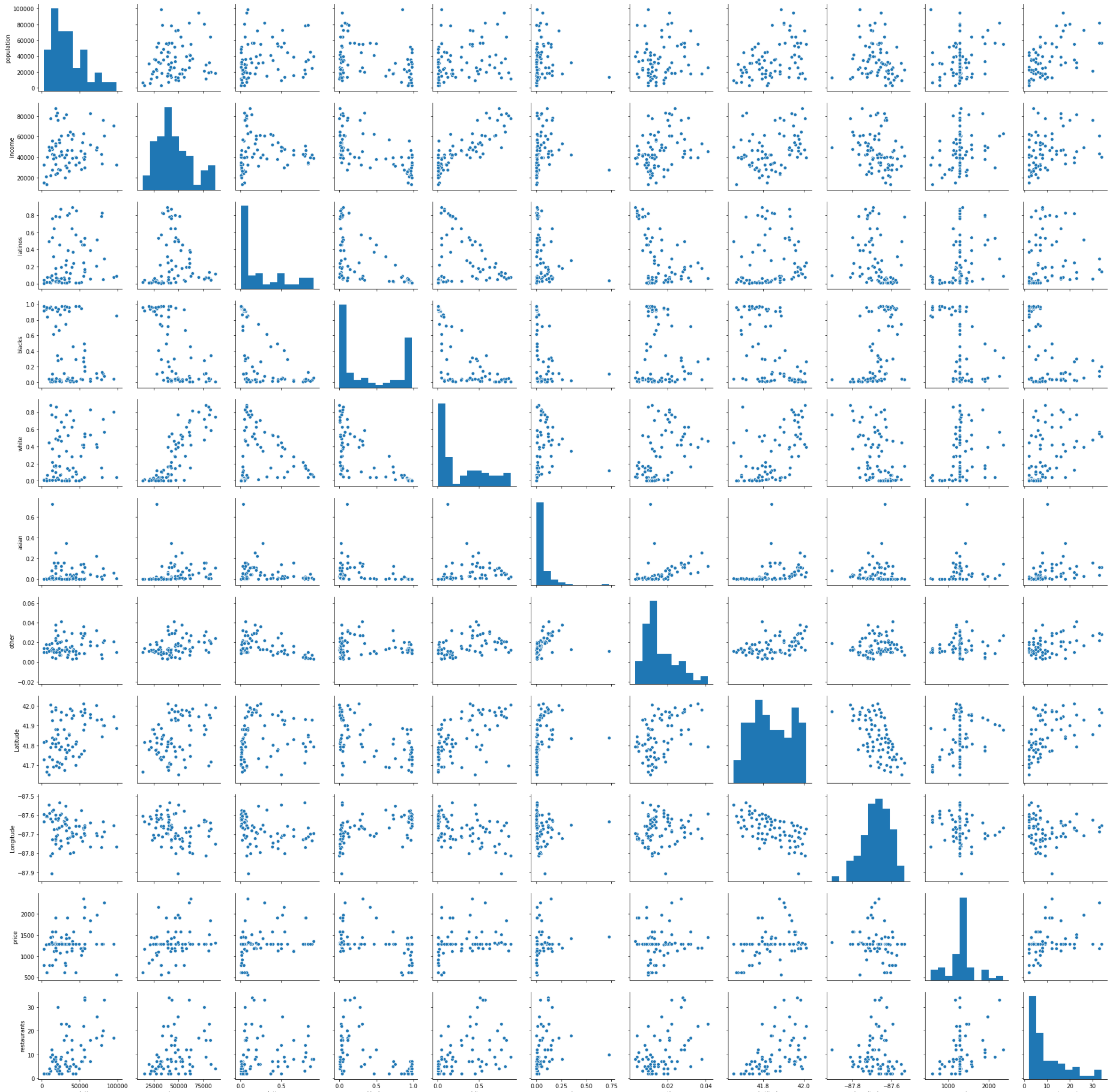

FIGURE 3.3 – Correlation Analysis

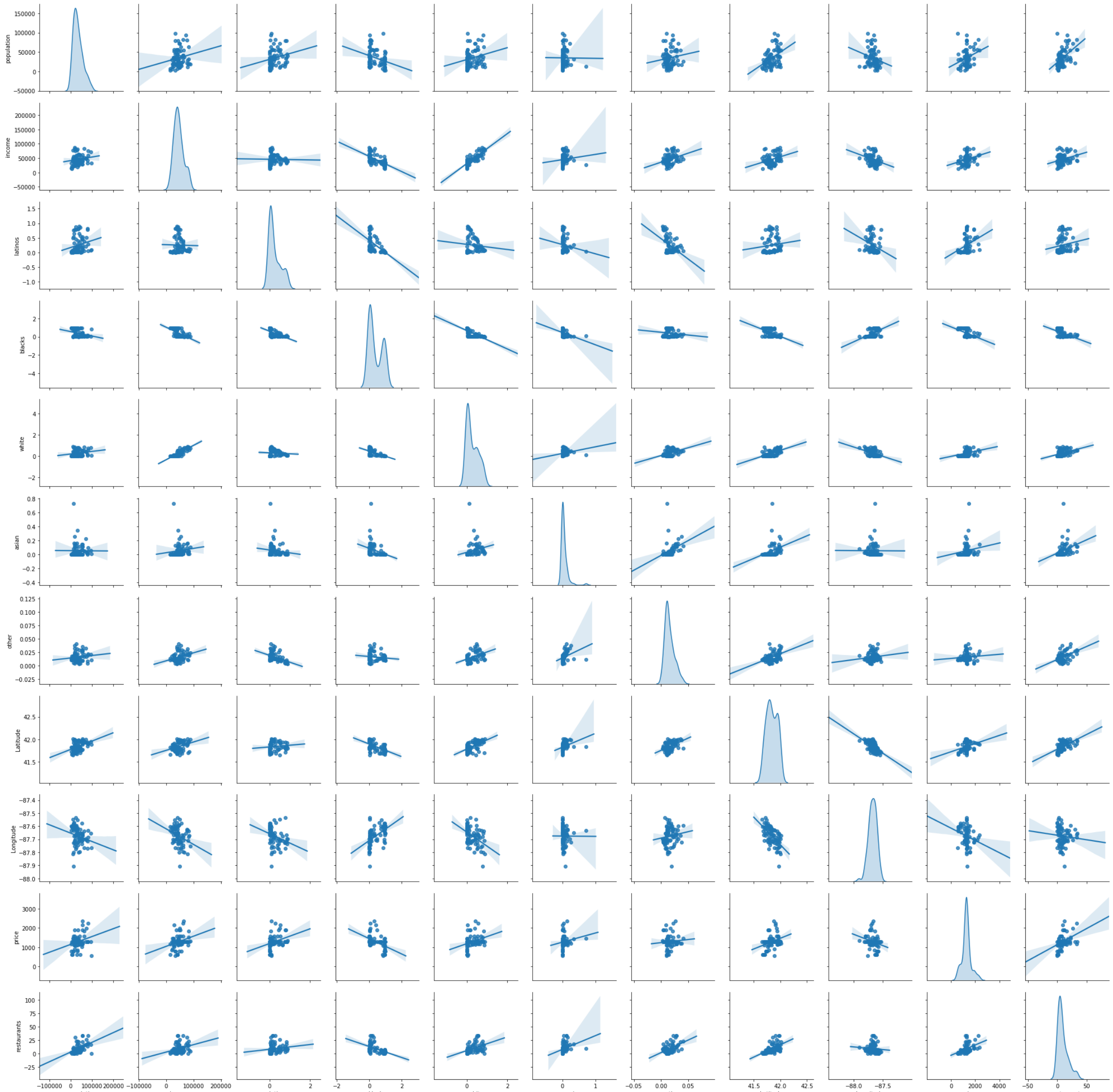

FIGURE 3.4 – Correlation Analysis

The data distribution of the number of restaurants and incomes, population and the rent fees are shown as follows :

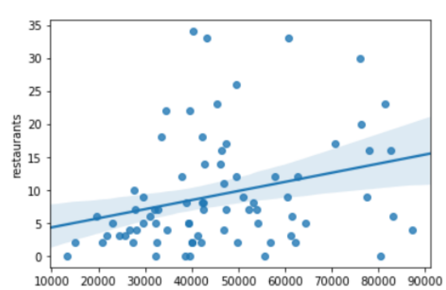

FIGURE 3.5 – Restaurants - Incomes

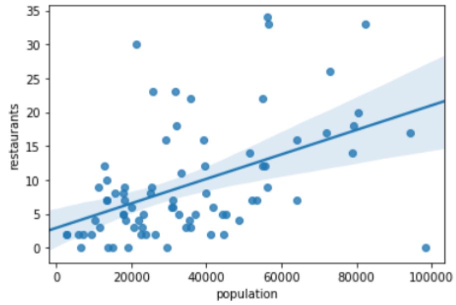

FIGURE 3.6 – Restaurants - Population

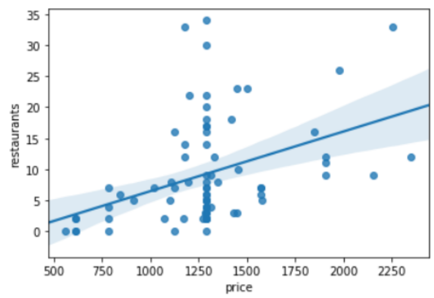

FIGURE 3.7 – Restaurants - Rent Fees

The correlations among features can be calculated as follows :

	population	income	latinos	blacks	white	asian	other	Latitude	Longitude	price	restaurants
population	1.000000	0.158677	0.188791	-0.268153	0.193299	-0.005859	0.120195	0.414476	-0.205366	0.261643	0.483141
income	0.158677	1.000000	-0.017895	-0.608502	0.839576	0.106144	0.306844	0.326617	-0.300464	0.265904	0.288821
latinos	0.188791	-0.017895	1.000000	-0.598503	-0.088472	-0.113886	-0.422515	0.107208	-0.289621	0.308689	0.148088
blacks	-0.268153	-0.608502	-0.598503	1.000000	-0.704699	-0.339418	-0.128790	-0.568606	0.506661	-0.460743	-0.494046
white	0.193299	0.839576	-0.088472	-0.704699	1.000000	0.225200	0.447044	0.577106	-0.437646	0.297632	0.430785
asian	-0.005859	0.106144	-0.113886	-0.339418	0.225200	1.000000	0.352828	0.314864	-0.003653	0.138924	0.311180
other	0.120195	0.306844	-0.422515	-0.128790	0.447044	0.352828	1.000000	0.465925	0.120102	0.079069	0.477621
Latitude	0.414476	0.326617	0.107208	-0.568606	0.577106	0.314864	0.465925	1.000000	-0.606663	0.318327	0.549015
Longitude	-0.205366	-0.300464	-0.289621	0.506661	-0.437646	-0.003653	0.120102	-0.606663	1.000000	-0.236302	-0.083881
price	0.261643	0.265904	0.308689	-0.460743	0.297632	0.138924	0.079069	0.318327	-0.236302	1.000000	0.400682
restaurants	0.483141	0.288821	0.148088	-0.494046	0.430785	0.311180	0.477621	0.549015	-0.083881	0.400682	1.000000

FIGURE 3.8 – Correlation

To check the confidence of the above correlation data, we calculate the Pearson values. The Pearson Correlation Coefficient between the number of restaurants and the population is 0.48314085413317565 with a P-value of P = 8.565985422491125e-06. The Pearson Correlation Coefficient between the number of restaurants and the rent fees is 0.40068247552309183 with a P-value of P = 0.00030500957612873453. Thus we can trust our correlation table confidently.

Based on the above figures, conclusions can be summarized as follows :

1. Neighborhoods with more population demands more restaurants .

2. It is counterintuitive that not neighborhoods with higher incomes have more restaurants ; Actually, neighborhoods whose incomes are among 37500 and 75000.

3. Among all races, the neighborhoods where more white people dwell the more restaurants will be there ; the number of restaurants increases sharply with the increase of asian population.

4. The neighborhoods in the eastern Chicago have more restaurants.

5. The rent fees for most restaurants are among 1000 dollars and 1500 dollars.

6. Most restaurants have less than 10 restaurants(because the query restriction of free version of foursquare API, we set the radius as 800).

3.2 Data Analysis

In this section, we select the K-Means method to cluster the neighborhoods. The is essential to regularize the initial data before train the model.

The regularized data can be described as :

	neighborhood	population	income	latinos	blacks	white	asian	other	Latitude	Longitude	price	restaurants
0	Rogers Park	0.897945	-0.363231	-0.040839	-0.321992	0.394017	0.097386	2.366179	1.713637	0.065916	-0.257725	1.531093
1	West Ridge	1.659636	0.085920	-0.184100	-0.701967	0.515461	1.681760	1.894169	1.643297	-0.302462	0.000000	0.930817
2	Uptown	0.959550	-0.314999	-0.406154	-0.479481	0.833356	0.589428	1.422159	1.271426	0.285581	0.000000	2.971754
3	Lincoln Square	0.201544	0.683145	-0.230659	-0.884455	1.244119	0.559905	1.540162	1.365707	-0.206716	-0.312424	0.330541
4	North Center	-0.141130	2.045031	-0.427643	-0.921952	1.751323	-0.089590	0.714145	1.165434	-0.055623	0.620351	1.651148

FIGURE 3.9 – Regularization

The crucial hyperparameter in K-Means is the number of clusters. To tune the hyperparameter, we use the elbow method.

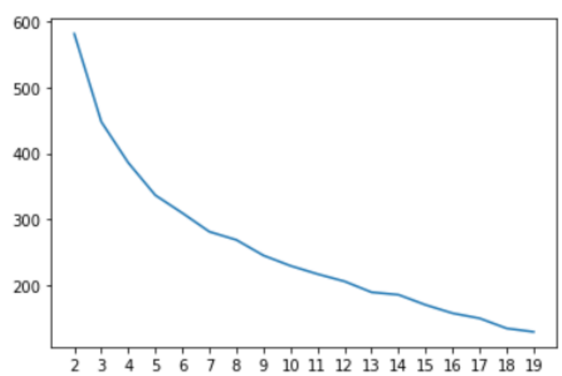

FIGURE 3.10 – Elbow Figure

We can find that the best hyperparameter is 7. And the neighborhood clustering map is depicted as Fig. 3.11.

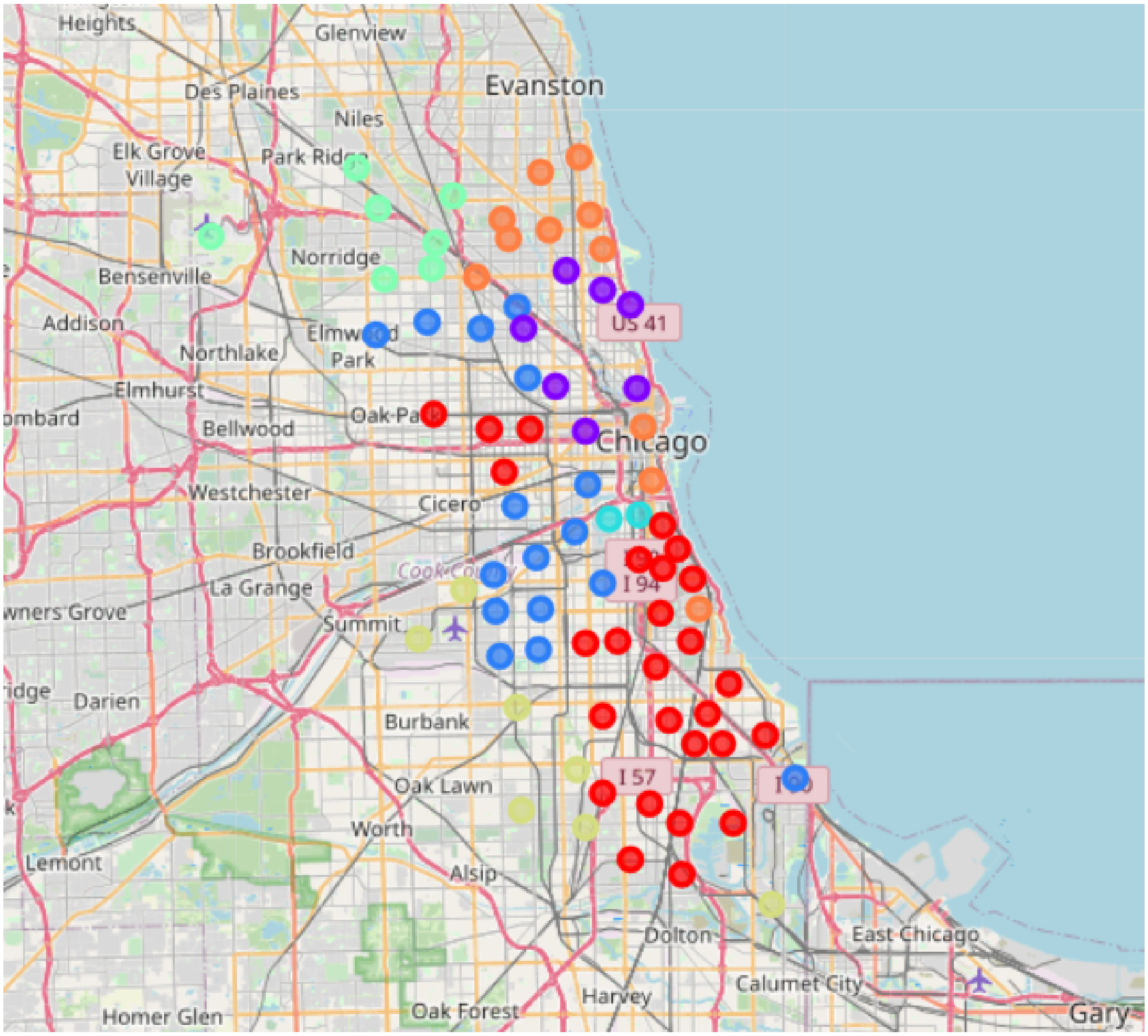

FIGURE 3.11 – Neighborhood Cluster Map

	neighborhood	population	income	latinos	blacks	white	asian	other	Latitude	Longitude	price	restaurants	cluster
0	Rogers Park	54991.0	39482.0	0.244	0.263	0.393	0.064	0.036	42.010531	-87.670748	1199.000000	22.0	6
12	North Park	17931.0	53305.0	0.180	0.032	0.493	0.257	0.038	41.980587	-87.720892	1198.000000	8.0	6
40	Hyde Park	25681.0	45335.0	0.063	0.304	0.467	0.124	0.041	41.794446	-87.593924	1447.000000	23.0	6
1	West Ridge	71942.0	47323.0	0.204	0.111	0.427	0.225	0.032	42.003548	-87.696243	1288.520833	17.0	6
32	Near South Side	21390.0	75995.0	0.056	0.281	0.481	0.155	0.027	41.856700	-87.624774	1288.520833	30.0	6
31	Loop	29283.0	78124.0	0.069	0.115	0.627	0.159	0.031	41.881609	-87.629457	1288.520833	16.0	6
15	Irving Park	53359.0	54048.0	0.456	0.033	0.417	0.070	0.025	41.953365	-87.736447	1124.000000	7.0	6
13	Albany Park	51542.0	46198.0	0.494	0.040	0.292	0.144	0.029	41.971937	-87.716174	1180.000000	14.0	6
76	Edgewater	56521.0	43331.0	0.165	0.143	0.547	0.116	0.029	41.983369	-87.663952	1178.000000	33.0	6
2	Uptown	56362.0	40324.0	0.142	0.200	0.516	0.114	0.028	41.966630	-87.655546	1288.520833	34.0	6
3	Lincoln Square	39493.0	57749.0	0.191	0.038	0.631	0.111	0.029	41.975990	-87.689616	1180.000000	12.0	6
54	Hegewisch	9426.0	49924.0	0.496	0.039	0.449	0.005	0.011	41.653646	-87.546988	1288.520833	2.0	5
63	Clearing	23139.0	54123.0	0.453	0.012	0.515	0.009	0.010	41.780588	-87.773388	1288.520833	5.0	5
69	Ashburn	41081.0	62238.0	0.368	0.462	0.152	0.007	0.011	41.747533	-87.711163	1288.520833	2.0	5
71	Beverly	20034.0	83092.0	0.046	0.341	0.588	0.006	0.019	41.718153	-87.671767	1288.520833	6.0	5
55	Garfield Ridge	34513.0	61206.0	0.392	0.059	0.532	0.010	0.007	41.803617	-87.745489	1288.520833	3.0	5
73	Mount Greenwood	19093.0	80505.0	0.072	0.052	0.860	0.007	0.010	41.698089	-87.708662	1123.000000	0.0	5
74	Morgan Park	22544.0	56886.0	0.027	0.667	0.287	0.004	0.014	41.690312	-87.666716	1288.520833	2.0	5
75	O'Hare	12756.0	49601.0	0.095	0.032	0.772	0.083	0.019	41.973101	-87.906768	1330.000000	12.0	4
10	Jefferson Park	25448.0	60592.0	0.194	0.010	0.687	0.089	0.021	41.969738	-87.763118	1313.000000	9.0	4
9	Norwood Park	37023.0	64477.0	0.120	0.004	0.815	0.046	0.015	41.985590	-87.800582	1288.520833	5.0	4
14	Portage Park	64124.0	52356.0	0.388	0.013	0.535	0.046	0.017	41.957809	-87.765059	1575.000000	7.0	4
8	Edison Park	11187.0	77678.0	0.078	0.003	0.884	0.024	0.012	42.005734	-87.814016	1288.520833	9.0	4
11	Forest Glen	18508.0	87394.0	0.115	0.007	0.746	0.107	0.024	41.991752	-87.751674	1313.000000	4.0	4
16	Dunning	41932.0	61584.0	0.238	0.007	0.704	0.038	0.013	41.952809	-87.796449	1575.000000	6.0	4
59	Bridgeport	31977.0	42382.0	0.270	0.021	0.351	0.345	0.013	41.837938	-87.651028	1419.000000	18.0	3
33	Armour Square	13391.0	27619.0	0.035	0.106	0.123	0.726	0.011	41.840033	-87.633107	1457.000000	10.0	3

FIGURE 3.12 – Neighborhood Cluster

60	New City	44377.0	32222.0	0.573	0.296	0.106	0.016	0.008	41.807533	-87.656440	1288.520833	5.0	2
17	Montclare	13426.0	47460.0	0.540	0.045	0.375	0.028	0.012	41.925309	-87.800893	1575.000000	7.0	2
58	McKinley Park	15612.0	42327.0	0.648	0.015	0.171	0.157	0.010	41.831700	-87.673664	1110.000000	8.0	2
30	Lower West Side	35769.0	34573.0	0.824	0.031	0.124	0.010	0.010	41.854200	-87.665609	1288.520833	22.0	2
29	South Lawndale	79288.0	33593.0	0.826	0.131	0.039	0.001	0.004	41.843644	-87.712554	1288.520833	18.0	2
51	East Side	23042.0	41196.0	0.784	0.034	0.172	0.002	0.007	41.713569	-87.532781	1288.520833	3.0	2
57	Brighton Park	45368.0	39289.0	0.853	0.012	0.081	0.050	0.004	41.818922	-87.698942	1288.520833	5.0	2
18	Belmont Cragin	78743.0	42842.0	0.789	0.032	0.152	0.020	0.008	41.931698	-87.768670	1288.520833	14.0	2
19	Hermosa	25010.0	42418.0	0.874	0.030	0.076	0.012	0.007	41.928643	-87.734502	1288.520833	8.0	2
20	Avondale	39262.0	46519.0	0.644	0.025	0.284	0.030	0.016	41.938921	-87.711168	1124.000000	16.0	2
56	Archer Heights	13393.0	42571.0	0.760	0.010	0.215	0.010	0.005	41.811422	-87.726165	1288.520833	7.0	2
65	Chicago Lawn	55628.0	37779.0	0.452	0.493	0.043	0.003	0.009	41.775033	-87.696441	1906.000000	12.0	2
21	Logan Square	72791.0	49610.0	0.512	0.054	0.392	0.025	0.017	41.928568	-87.706793	1976.000000	26.0	1
7	Near North Side	80484.0	76290.0	0.049	0.108	0.721	0.101	0.020	41.900033	-87.634497	1288.520833	20.0	1
6	Lincoln Park	64116.0	82707.0	0.056	0.043	0.829	0.051	0.021	41.940298	-87.638117	1846.000000	16.0	1
5	Lake View	94368.0	70746.0	0.076	0.039	0.804	0.060	0.021	41.947050	-87.655429	1288.520833	17.0	1
27	Near West Side	54881.0	62770.0	0.092	0.315	0.420	0.146	0.027	41.880066	-87.666762	2351.000000	12.0	1
23	West Town	82236.0	60720.0	0.291	0.078	0.572	0.038	0.022	41.901421	-87.686166	2255.000000	33.0	1
4	North Center	31867.0	81524.0	0.136	0.023	0.773	0.045	0.022	41.956107	-87.679160	1504.000000	23.0	1
66	West Englewood	35505.0	26654.0	0.022	0.963	0.004	0.001	0.011	41.778089	-87.666718	1288.520833	4.0	0
67	Englewood	30654.0	19743.0	0.011	0.974	0.003	0.001	0.011	41.779756	-87.645884	842.000000	6.0	0
68	Greater Grand Crossing	32602.0	29663.0	0.012	0.969	0.006	0.001	0.013	41.766886	-87.620845	1103.000000	5.0	0
70	Auburn Gresham	48743.0	34767.0	0.009	0.978	0.003	0.001	0.009	41.743387	-87.656042	1288.520833	4.0	0
72	Washington Heights	26493.0	42053.0	0.010	0.974	0.005	0.000	0.012	41.706423	-87.656160	1288.520833	2.0	0

FIGURE 3.13 – Neighborhood Cluster

Based on our cluster results, cluster 6 is ideal place for new restaurants. Especially, Rogers Park and Hyde Park belong to the 6th cluster and have less than 30 restaurants now.

According to the above data, we find that cluster 1 is competitive and has high amounts of restaurants. Then we want to know is our cluster accurate ? Or should we combine cluster 1 and 6 together ? We can use ANOVA method to check the difference between the cluster 1 and cluster 6.

ANOVA : Analysis of Variance The Analysis of Variance (ANOVA) is a statistical method used to test whether there are significant differences between the means of two or more groups. ANOVA returns two parameters : F-test score : ANOVA assumes the means of all groups are the same, calculates how much the actual means deviate from the assumption, and reports it as the F-test score. A larger score means there is a larger difference between the means. P-value : P-value tells how statistically significant is our calculated score value. If our price variable is strongly correlated with the variable we are analyzing, expect ANOVA to return a sizeable F-test score and a small p-value.

F-statistics = MSB / MSE where MSB is mean squared between and MSE is mean squared error. MSB measures the variance of each cluster to the whole population. MSE measures the variance of each cluster itself. If F is large(MSB is large, MSE is small), at least there exists one cluster is difference from the others. And all clusters' variances are small. If F is small(MSB is small and MSE is large), there are two possible cases, one of which is the mean values of clusters are similar. Another case is the variances of all cluster are large. In our case, F= 0.1051366776809853 , P = 0.7499507171065818. The F value is small, thus we can not reject our null hypothesis(clster 1 and 6 are the same). Thus, cluster 6 and 1 are ideal place for new restaurants. Rogers Park, Hyde Park and Logan Square are competitive.

Chapitre 4

Conclusion

Restaurants are a driving force in Chicago's economy and employment ratios. In this report, we obtain our data set from the open data source, Foursquare API, GEO API and website. After pre-processing and cleaning the data set, we visualize and analyze the relationships among population, incomes, races, rent fees, the number of restaurants in different neighborhoods. And K-Means method is applied to cluster the neighborhoods. Our conclusion can be summarized as follows : 1. the top ten neighborhoods with most restaurants are Uptown, Edgewater, West Town, Near South Side, Logan Square, Hyde Park, North Center, North Center, Lower West Side, Rogers Park and Near North Side. 2. Neighborhoods with more population demands more restaurants . 3. It is counterintuitive that not neighborhoods with higher incomes have more restaurants ;Actually, neighborhoods whose incomes are among 37500 and 75000. 4. Among all races, the neighborhoods where more white people dwell the more restaurants will be there ; the number of restaurants increases sharply with the increase of asian population. 5. The neighborhoods in the eastern Chicago have more restaurants. 6. The rent fees for most restaurants are among 1000 dollars and 1500 dollars. 7. Rogers Park,Hyde Park and Logan Square are competitive for new restaurant center.

Chapitre 5

Reference

[1]Illinois Restaurants Association, https ://www.illinoisrestaurants.org/page/IndustryStatistics [2]Chicago Data, https ://github.com/dssg/411-on-311 [3]Chicago Average House Rent Fees, https ://www.rentcafe.com/average-rent-market-trends/us/il/chicago/