

Capstone Project - The Battle of Neighborhoods

Where should I open my restaurant in downtown Shanghai?

Issac Ma
2021.03.02

Shanghai is fancy and cruel.

This project, through data analysis, tries to provide insights for decision makers by answering the following questions:

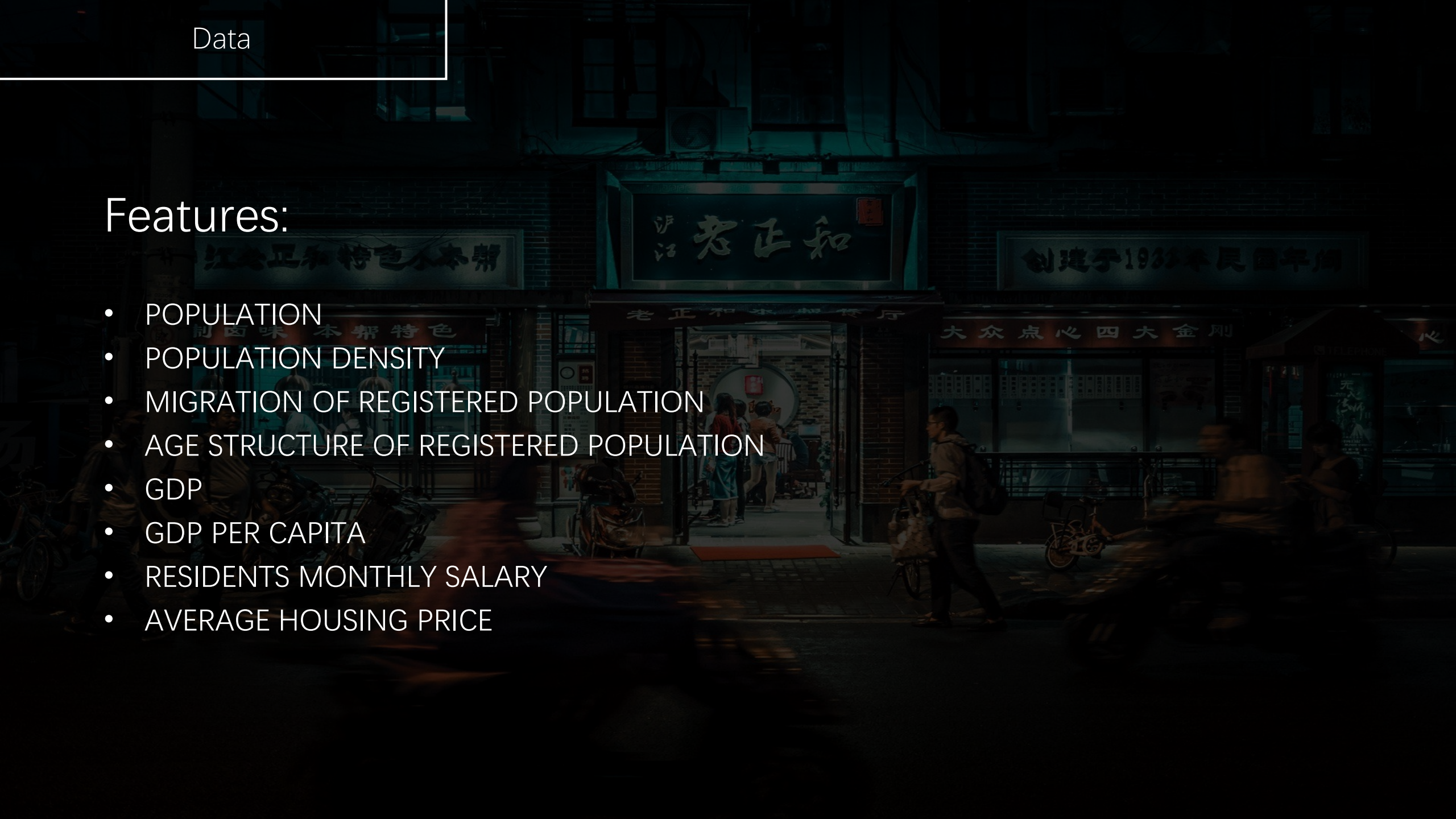
- What kind of restaurant should I open?
- Where should my restaurant be located to ensure better profitability?

Shanghai Statistical Yearbook 2019

- RESIDENT FOREIGNERS IN SHANGHAI IN MAIN YEARS
- GDP Per Papita By District (2019)
- Residents Monthly Salary and Average Housing Prices By Districts (2019)
- LAND AREA, POPULATION AND DENSITY OF POPULATION IN DISTRICTS (2018)
- MIGRATION OF REGISTERED POPULATION IN DISTRICTS (2017 ~ 2018)
- AGE STRUCTURE OF REGISTERED POPULATION IN DISTRICTS (2018)

Features:

- POPULATION
- POPULATION DENSITY
- MIGRATION OF REGISTERED POPULATION
- AGE STRUCTURE OF REGISTERED POPULATION
- GDP
- GDP PER CAPITA
- RESIDENTS MONTHLY SALARY
- AVERAGE HOUSING PRICE



Also, because we are focusing our analysis on downtown Shanghai, so we will only consider districts that are considered as the urban core of Shanghai historically, which include:

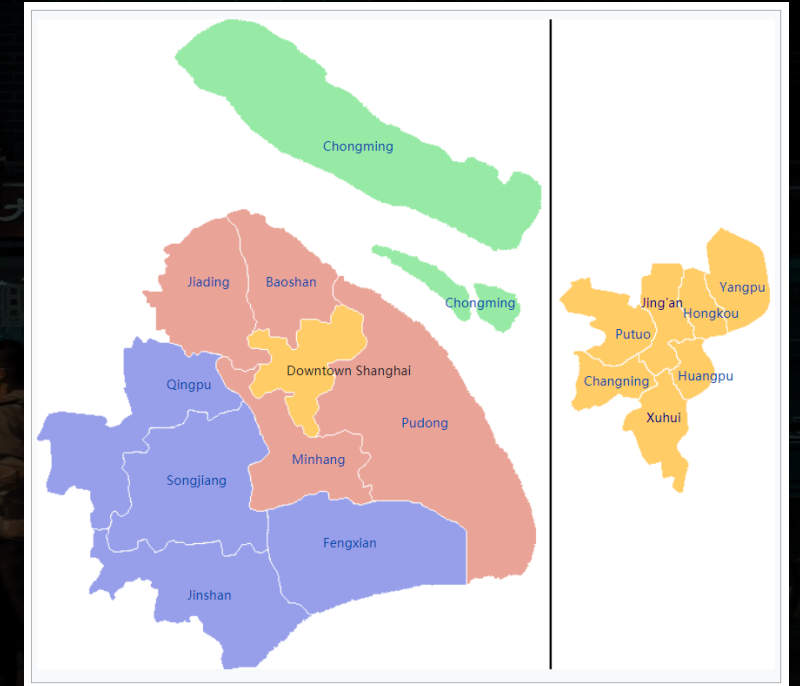
- Huangpu District
- Xuhui District
- Jingan District
- Changning District
- Hongkou District
- Yangpu District
- Putuo District

We will try to obtain all relevant data and manipulate them to eventually form 2 datasets. They are:

dist_info_center, containing all demographic and relevant information regarding the 7 central administrative districts in Shanghai mentioned above.

dist_neigh, containing neighbourhoods in central districts that are arbitrarily defined and their latitudes and longitudes.

These 2 datasets will be our primary data for later analysis.



source: wikipedia

Data

	district	neighbourhood_eng	neighbourhood_cn	latitude	longitude
0	Huangpu	Dongjiadu, Shanghai	董家渡	31.220094	121.509060
1	Huangpu	Huangpu Riverside, Shanghai	黄浦滨江	31.202310	121.488060
2	Huangpu	Laoximen, Shanghai	老西门	31.224795	121.490066
3	Huangpu	East Nanjing Road, Shanghai	南京西路	31.240739	121.465932
4	Huangpu	Penglai Park, Shanghai	蓬莱公园	31.208241	121.487516
...
72	Putuo	Zhenguang, Shanghai	真光	31.259050	121.393681
73	Putuo	Zhenru, Shanghai	真如	31.253539	121.409050
74	Putuo	COSCO two Bay City, Shanghai	中远两湾城	31.258663	121.451892
75	Putuo	Ganquan, Shanghai	甘泉	31.271251	121.437584
76	Putuo	Yichuan, Shanghai	宜川	31.266578	121.447124

dist_neigh

Data

	district	population	population_density	migration_in_17	migration_out_17	migration_in_18	migration_out_18	age_17_and_below	age_18_to_34
1	Huangpu	653,800	31,955	2,890	692	3,108	800	84,400	144,300
2	Xuhui	1,084,400	19,803	9,832	6,225	10,657	5,885	125,500	167,000
3	Changning	694,000	18,120	4,280	1,661	5,086	1,585	64,400	100,500
4	Jing'an	1,062,800	28,818	4,246	904	5,295	903	104,500	155,600
5	Putuo	1,281,900	23,380	4,868	1,065	5,585	945	102,000	142,700
6	Hongkou	797,000	33,944	3,913	1,377	4,460	1,396	71,700	124,700
7	Yangpu	1,312,700	21,615	16,485	8,679	17,043	8,006	116,000	210,800

dist_info_center

First, we will do some exploratory analyses by exploring the demographic information of the 7 central districts to understand them better. One of the important questions is:

How are most people in this district different from the others?

We will use **FourSquare API** to fetch information of all restaurants that can be found in central districts, try to cluster them into different groups using **K Means Clustering** and further break them down using the same clustering method.

We will eventually find the best neighbourhood to open a new restaurant based on the following principles:

- Try to avoid locations where there are too many restaurants of the same kind, which means fierce competition;
- Prefer places where most residents have a relatively higher consuming capability;
- Prefer places where the rent is relatively affordable.

1. What's the most populous district?

	district	population
0	Yangpu	1312700
1	Putuo	1281900
2	Xuhui	1084400
3	Jing'an	1062800
4	Hongkou	797000
5	Changning	694000
6	Huangpu	653800

Looks like a lot of people are living in **Yangpu** and **Putuo** since these 2 districts are historically considered to be '老城区', meaning old urban areas with large population.

2. Which district has the highest population density?

	district	population_density
0	Hongkou	33944
1	Huangpu	31955
2	Jing'an	28818
3	Putuo	23380
4	Yangpu	21615
5	Xuhui	19803
6	Changning	18120

Hongkou tops the list as it is another old urban area with a large number of ordinary Shanghai residents. Although near city center, many of whom are living in spaces no larger than 20 square meters, eagerly waiting to be relocated by the government, which means getting one or multiple apartments for free. Each apartment, could be worth of at least 2 million RMB now in Shanghai.

Following Hongkou are **Huangpu** and **Jing'an**. These 2 districts are conventionally considered the places for the rich people and tourists with limited land areas, which explains relatively higher density of population.

4. Which district has relatively higher GDP per capita? What about monthly salary?

	district	monthly_salary	gdp_avg
0	Jing'an	8380	217300
1	Changning	8030	237800
2	Hongkou	7970	145300
3	Putuo	7720	87100
4	Xuhui	7640	257400
5	Yangpu	7220	159600
6	Huangpu	7160	397100

Unsurprisingly, **Huangpu** tops the list in terms of GDP per capita, yet the monthly income of Huangpu residents is no match for Jing'an and Changning. In fact, Huangpu residents earn the least among all 7 central districts.

People working in **Jing'an** and **Changning** earn more than 8,000 RMB per month on average. After all, they are basically RICH districts with countless office buildings home to numerous knowledge-intensive businesses and a lot of foreign workers.

It is interesting to see that **Hongkou** and **Putuo** residents also enjoy a decent monthly salary close to 8,000 RMB with apparently lower GDP per capita.

5. Which district is more 'YOUNG', with higher ratio of laborforce (age 18-59) to total population?

	district	laborforce	population	ratio
0	Huangpu	422100	653800	0.6456
1	Hongkou	369200	797000	0.4632
2	Jing'an	470700	1062800	0.4429
3	Xuhui	476400	1084400	0.4393
4	Changning	301400	694000	0.4343
5	Yangpu	566300	1312700	0.4314
6	Putuo	447500	1281900	0.3491

Yangpu is indeed a populous urban area, followed by Xuhui and Jing'an, also 2 areas favored by major workforce.

Huangpu, **Hongkou** and **Jing'an** are apparently 'younger' than the other districts.

6. Which district is more friendly to people who want to settle in downtown Shanghai?

	district	housing_price
0	Putuo	55738
1	Hongkou	58927
2	Yangpu	59443
3	Jing'an	66228
4	Changning	68491
5	Xuhui	71064
6	Huangpu	81375

	housing_price_avg
0	65895.1429

The average housing price in Shanghai is **65895 RMB/m²**.

The biggest challenge to settle in a first-tier city in China is whether one can afford to buy an apartment or not.

Currently, with the down payment of more than 300,000 RMB in Shanghai on average, It seems like **Putuo**, **Hongkou**, and **Yangpu** might be a good place to start a new life in downtown Shanghai, where housing prices are lower than the Shanghai average by as much as 5,000 RMB/m².

Fetch restaurant information using FourSquare API

```
#### Send GET request and examine results

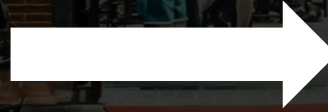
import requests

results_huangpu = requests.get('https://api.foursquare.com/v2/venues/by/latlng?lat=31.234410019618384&lng=121.47994779956817&v=20180801&requestId=603c597a&type=notificationTray', headers={'X- foursquare-client-id': '4cbd9032035d236a08'})
results_huangpu

{'meta': {'code': 200, 'type': 'notificationTray', 'notifications': [], 'response': {'version': '20180801', 'name': 'Fukui Japanese Restaurant (福井日本料理)', 'location': {'lat': 31.234410019618384, 'lng': 121.47994779956817, 'crossStreet': 'Shandong M Rd', 'latLngs': [{'label': 'display', 'lat': 31.234410019618384, 'lng': 121.47994779956817}], 'labeledLatLngs': [{'lat': 31.234410019618384, 'lng': 121.47994779956817}], 'distance': 91, 'cc': 'CN', 'city': 'Huangpu', 'state': '上海市', 'country': '中国', 'formattedAddress': ['339 Guangdong Rd', 'Huangpu', '上海市', '中国']}]}
```

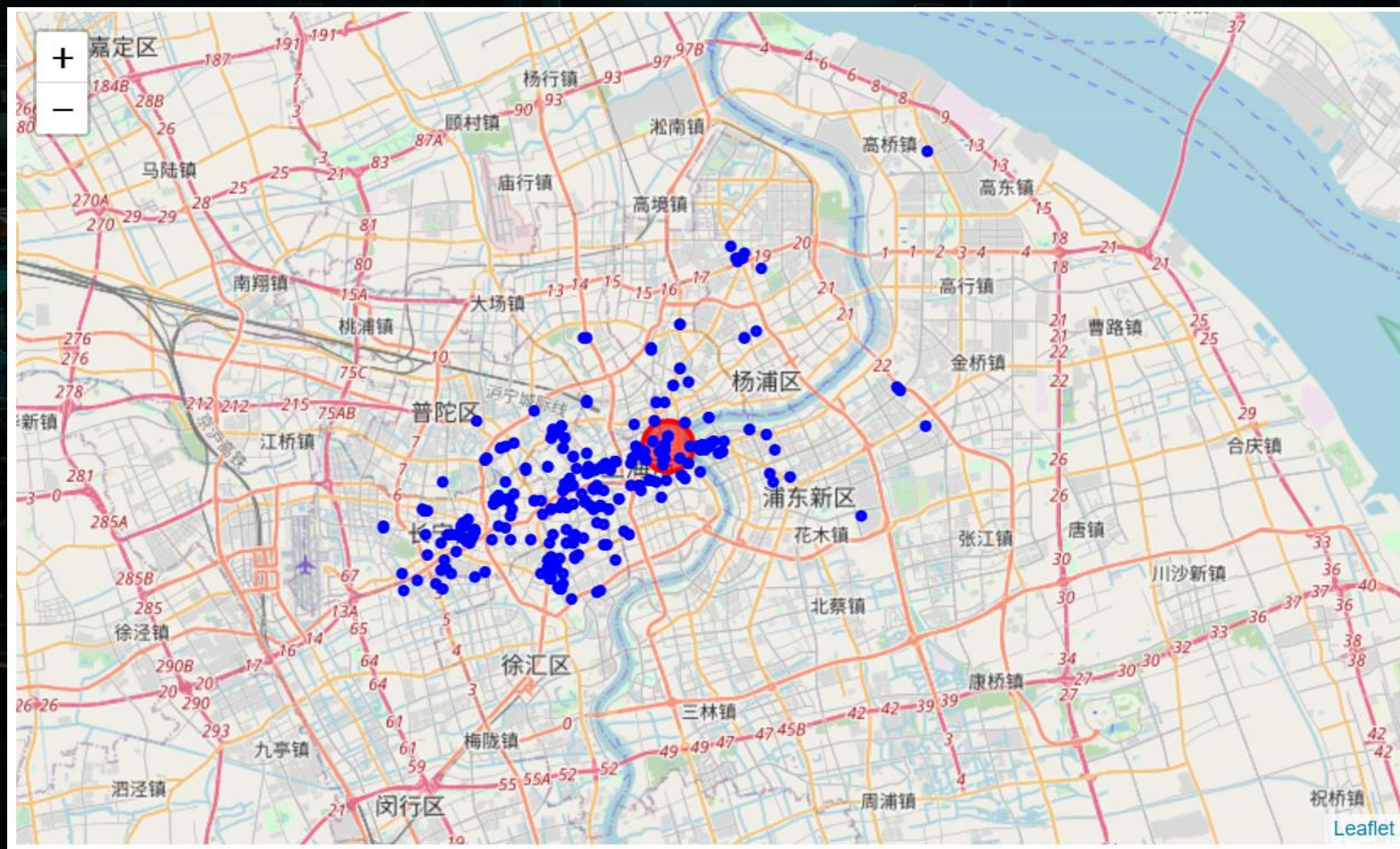


FOURSQUARE

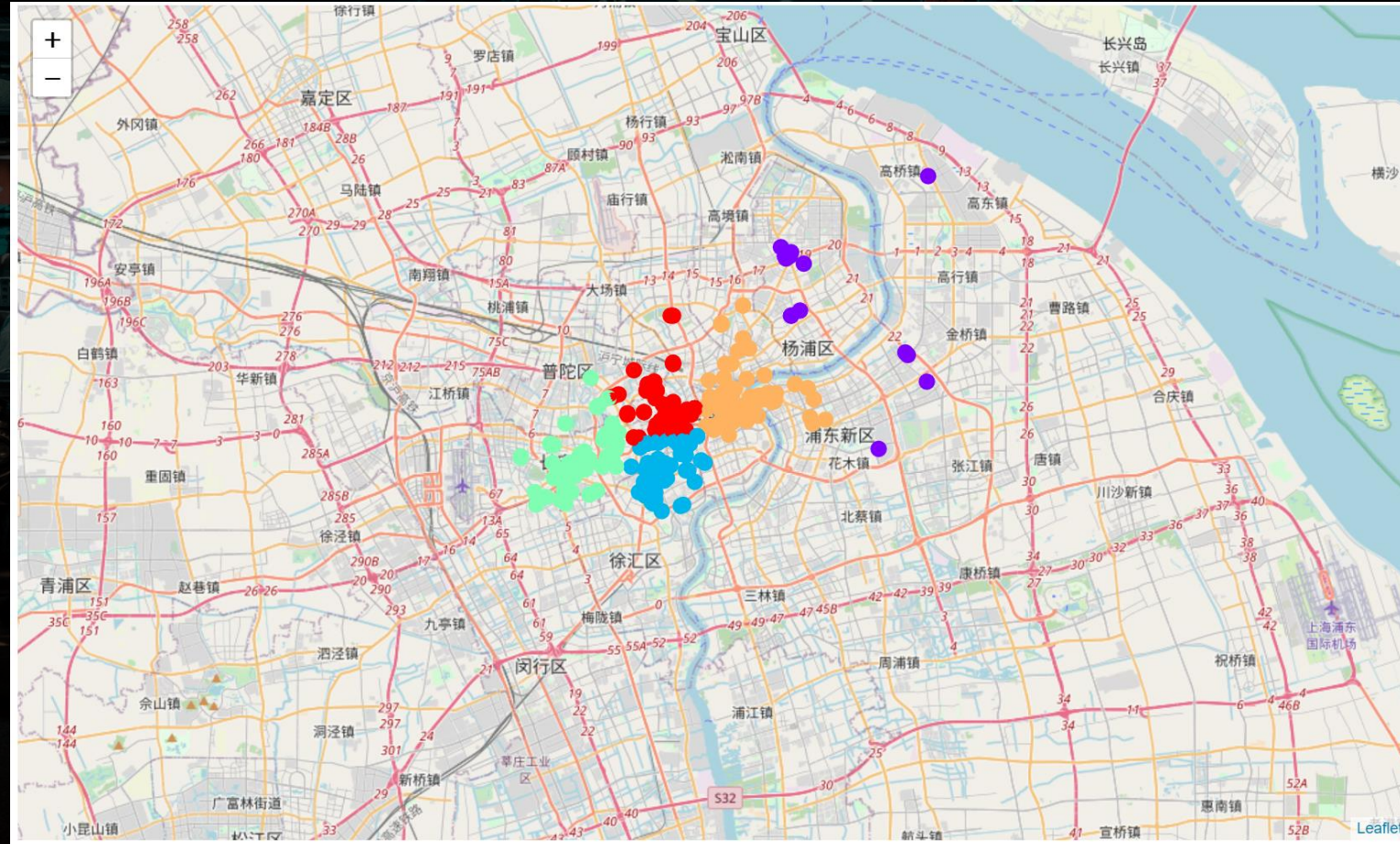


	name	categories	address	crossStreet	lat	lng	lat	crossStreet	lat	distance	cc	city	state	country
0	Fukui Japanese Restaurant (福井日本料理)	Sushi Restaurant	339 Guangdong Rd	Shandong M Rd	339 Guangdong Rd	Shandong M Rd	31.234410	121.479948	121.479948	91	CN	上海市	中国	
1	Three Street Restaurant (三玛路)	Chinese Restaurant	黄浦区汉口路413号	Chinese Restaurant	黄浦区汉口路413号	NaN	31.237718	121.477995	121.477995	31.2				
2	Xinwang Restaurant (新旺茶餐厅)	Chinese Restaurant	汉口路309号 309 Hankou Rd.	Shandong Rd. (M)	31.237748	121.479868	31.23							
3	Nanxiang Steamed Bun Restaurant (南翔馒头店)	Dumpling Restaurant	85 Yuyuan Rd 豫园老街85号	NaN	31.228740	121.486932								
4	Waya Restaurant (蓝方餐厅)	Chinese Restaurant	410 Sichuan M Rd	opp Xinfaji	410 Sichuan M	opp Xinfaji	31.238							
5	Tai Kang Tang Bao Guan	Dumpling	2/F 766 Nanjing E Rd	Xizang M	24.000000	101.430000								

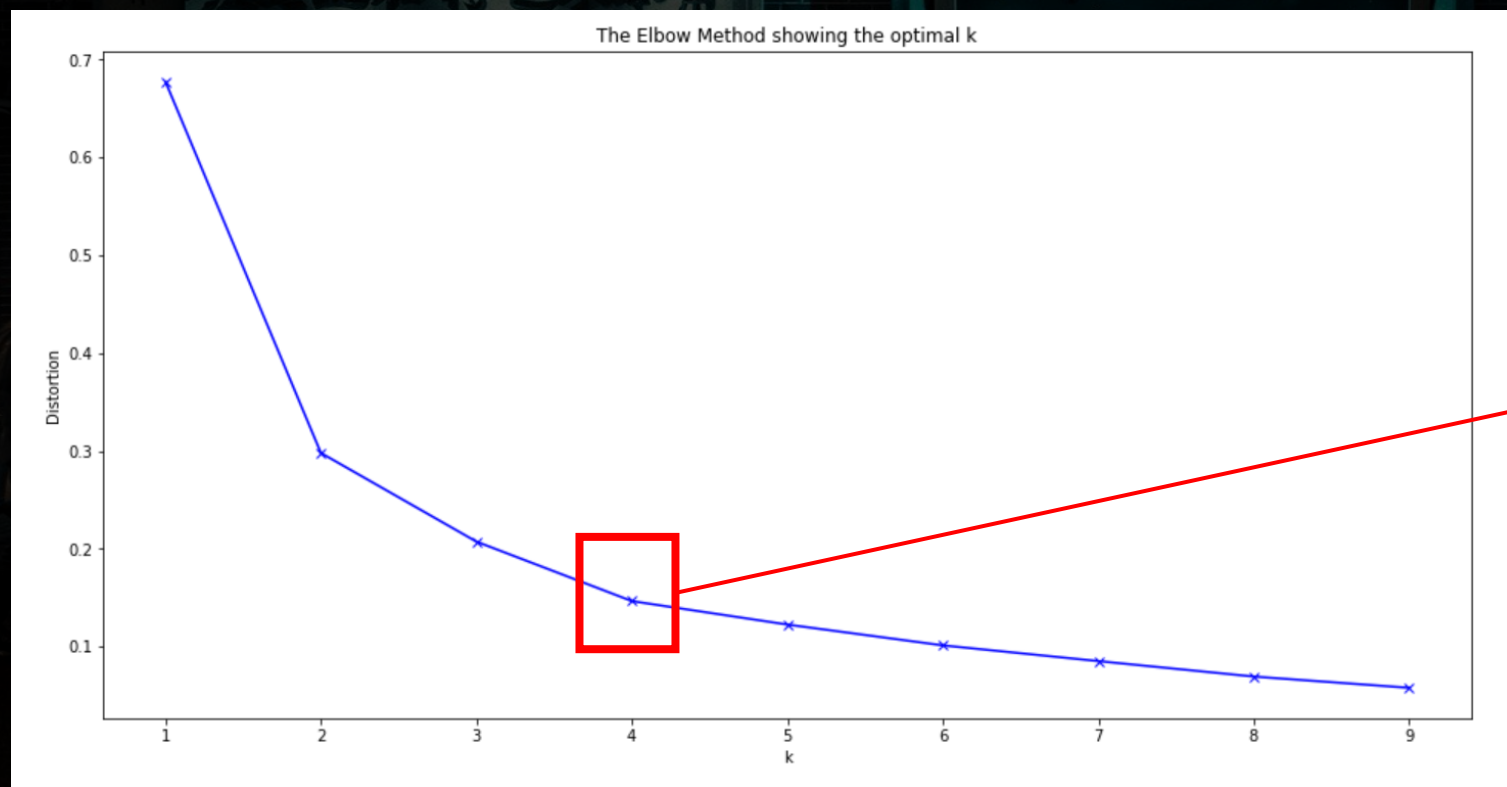
Plotting



K Means Clustering—— All 7 Central Districts (k = 5)

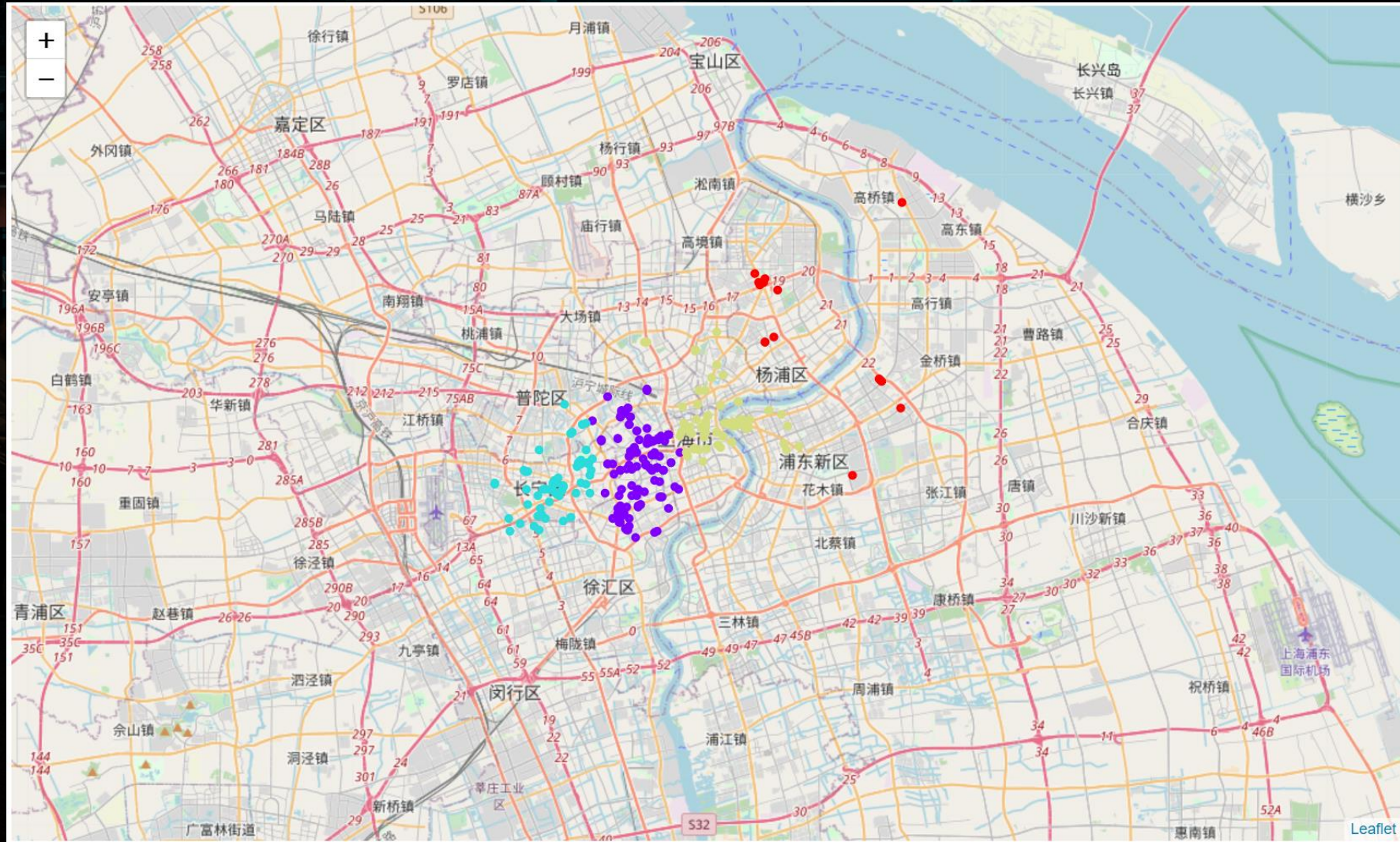


Find the optimal K using the Elbow Method



optimal $k = 4$

K Means Clustering—— All 7 Central Districts with the optimal k value

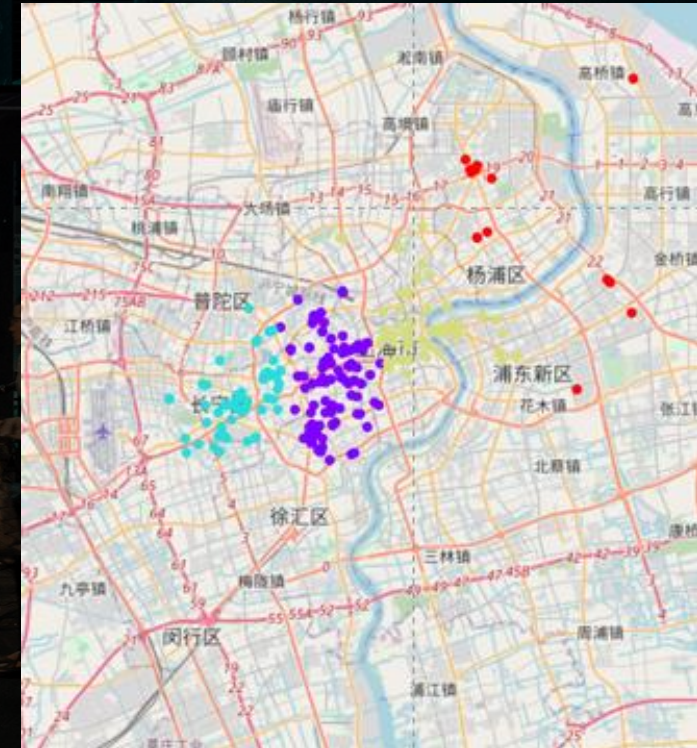


Business Understanding

We finally have our clustering map, in which the restaurants are divided into 4 groups.

Now let's decide which cluster we need to explore further based on the business principles we mentioned in the Methodology section:

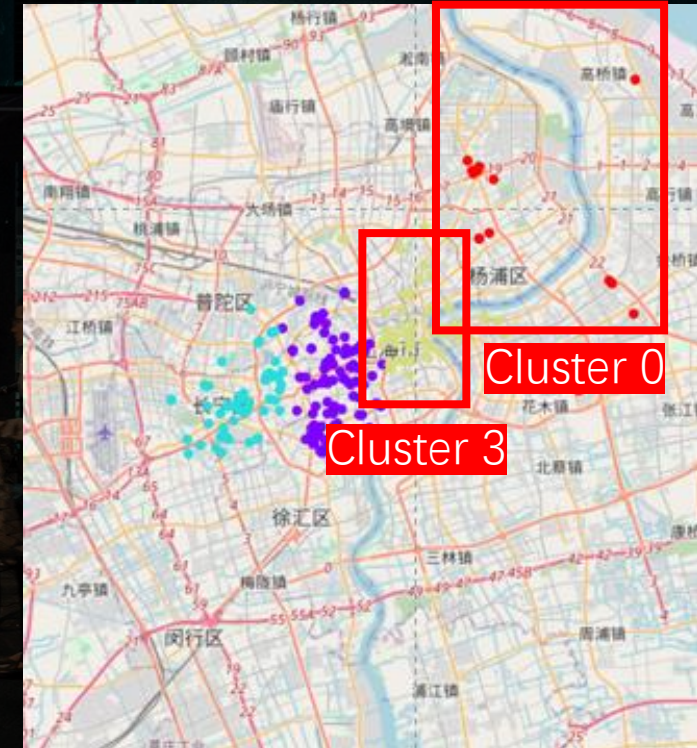
- Try to avoid locations where there are too many restaurants of the same kind, which means fierce competition;
- Prefer places where most residents have a relatively higher consuming capability;
- Prefer places where the rent is relatively affordable.



Business Understanding

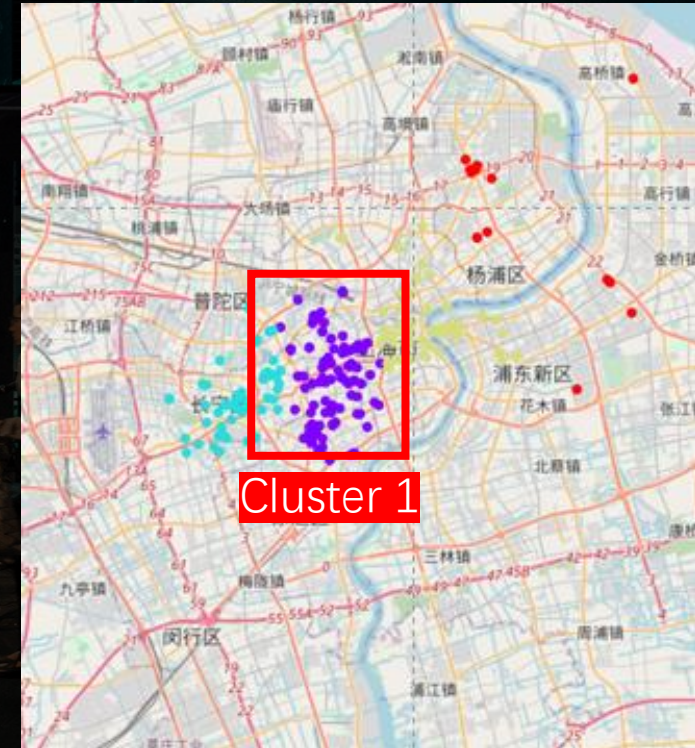
Cluster 0, which is red, seems to be a group of outliers far from downtown Shanghai and therefore deserves no further consideration.

Cluster 3, which is yellow, covers the entire area of Huangpu district as well as Lujiazui, which is the central part of Pudong, which is not part of the central districts we define in this project, where famous skyscrapers like the Oriental Pearl TV Tower, Shanghai Tower and Internatinal Financial Center are located. This cluster is basically equivalent to center of downtown Shanghai. Even though the purchasing power of Huangpu residents is not as high as housing prices within the area, all dots are densely concentrated within the cluster, which means fierce competition and the average rents might be one of the highest among Shanghai. Since we do not have many expericences of running a restaurant, for risk-control consideration, we will also skip this cluster.



Business Understanding

Cluster 1, which is purple, covers almost the entire area of Xuhui, Jing'an, and part of Putuo, Changning and Huangpu district. Limilar to Cluster 3, this cluster also have densely concentrated dots. One of the biggest differences is that it covers a bigger area compared to Cluster 3. Other features like purchasing power, rents, GDP percapita are not very different to Cluser 3. This cluster is OK because the average rents might be relatively lower than Cluster 3 and it covers a wide range of areas, which means diverse customer groups plus decent residents' purchasing power based the results from previous Exploratory Data Analysis. But is the our best choice?



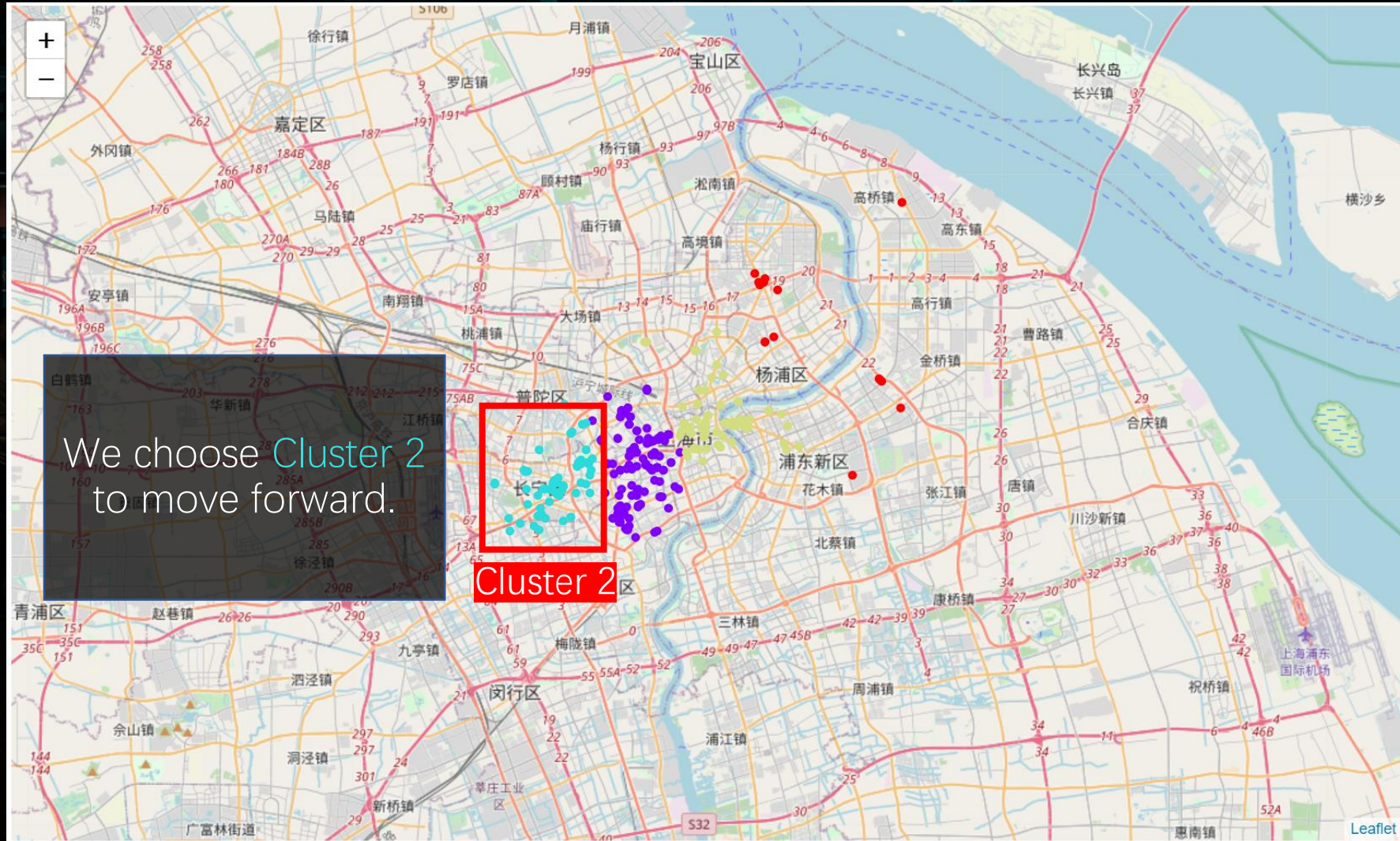
Business Understanding

Cluster 2, which is blue, covers most areas of Changning and Putuo district. And we decide to move forward with this cluster for the following reasons:

- Most restaurant are not densely concentrated within the area compared to Cluster 2;
- The entire area is relative further than that of Cluster 2, which means average rents might be more affordable;
- People in this cluster also have a decent monthly salary.



K Means Clustering—— All 7 Central Districts with the optimal k value



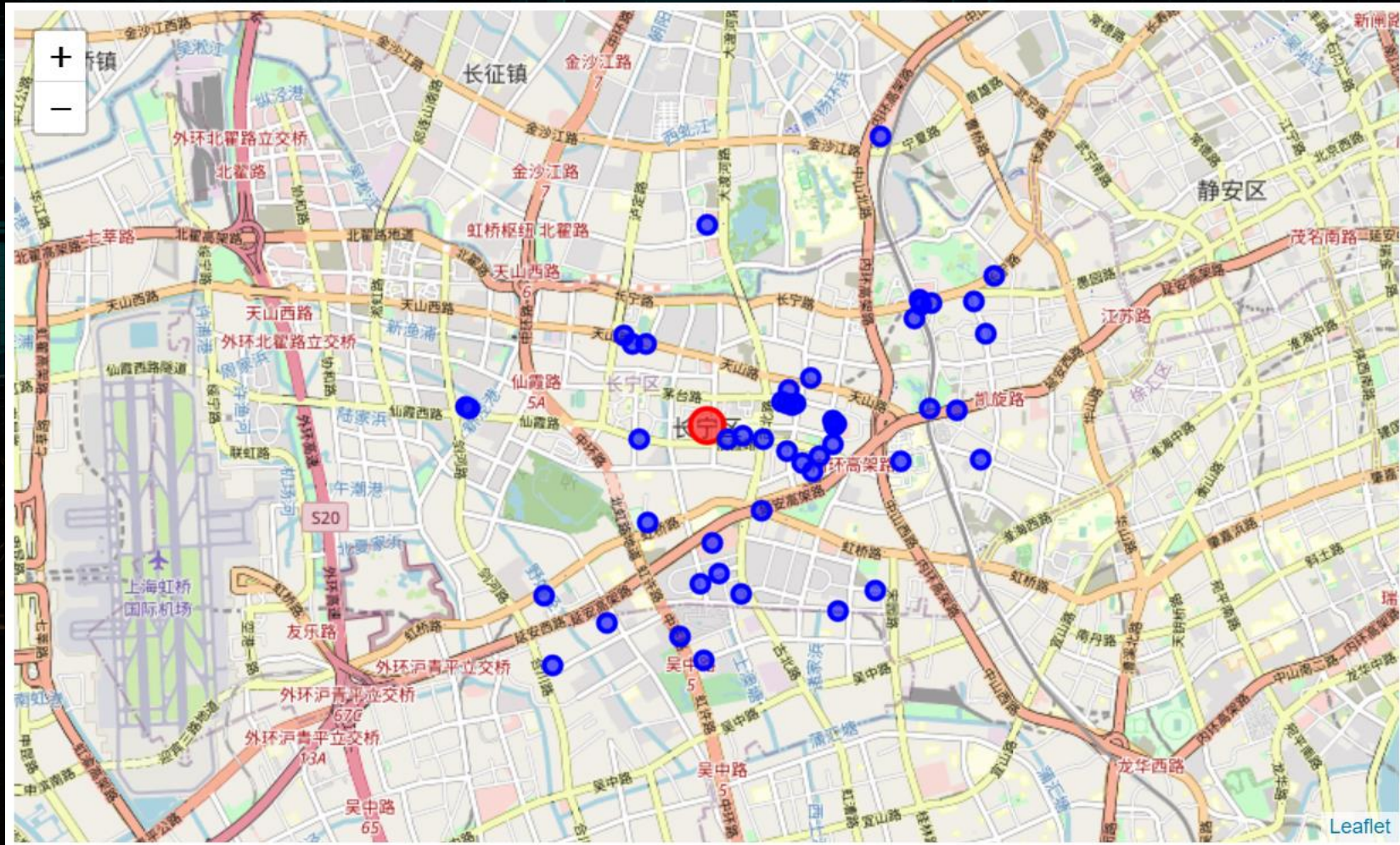
Business Understanding

Cluster 2, which is blue, covers most areas of Changning and Putuo district. And we decide to move forward with this cluster for the following reasons:

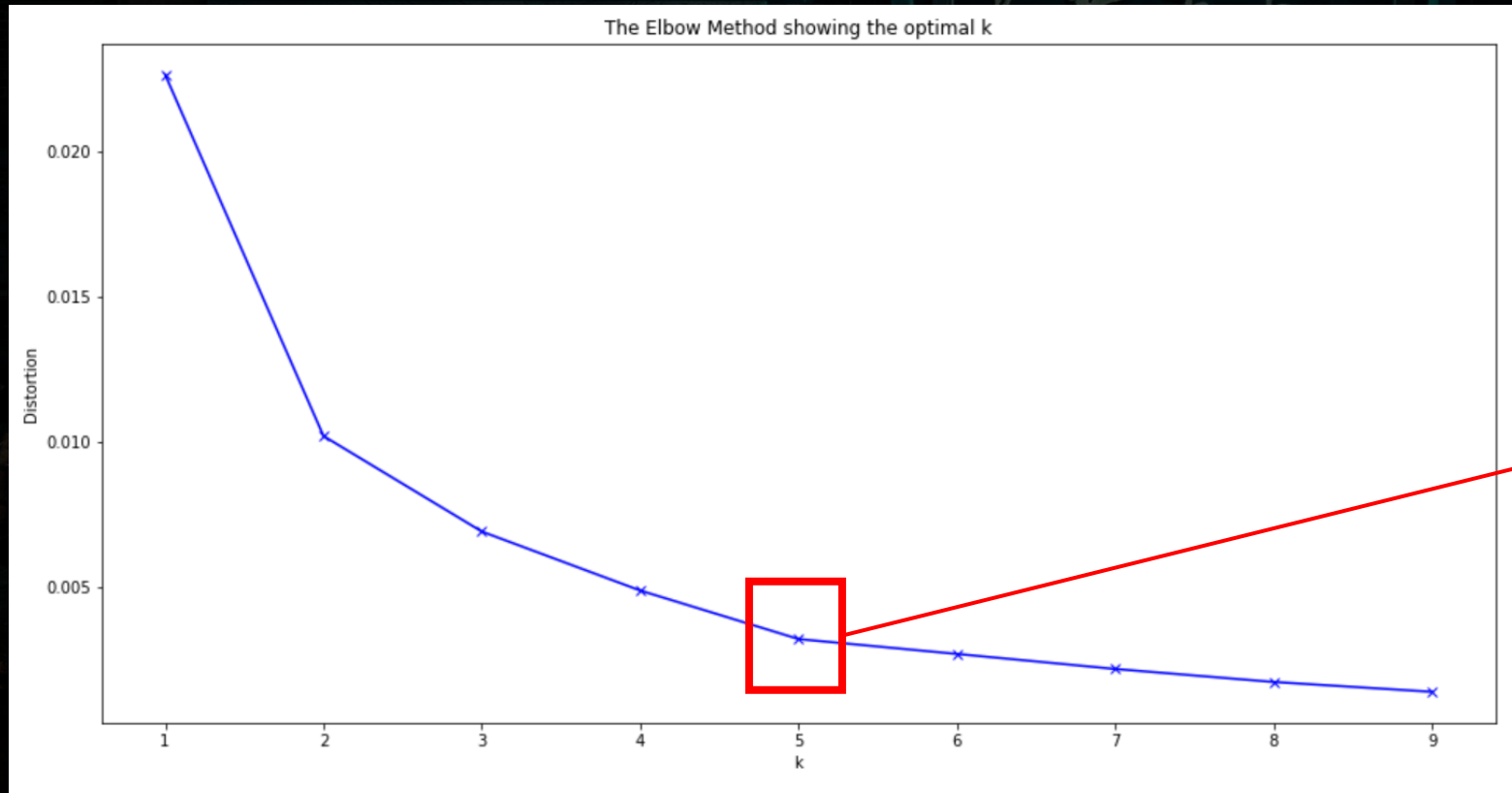
- Most restaurant are not densely concentrated within the area compared to Cluster 2;
- The entire area is relative further than that of Cluster 2, which means average rents might be more affordable;
- People in this cluster also have a decent monthly salary.



Plotting——Cluster 2

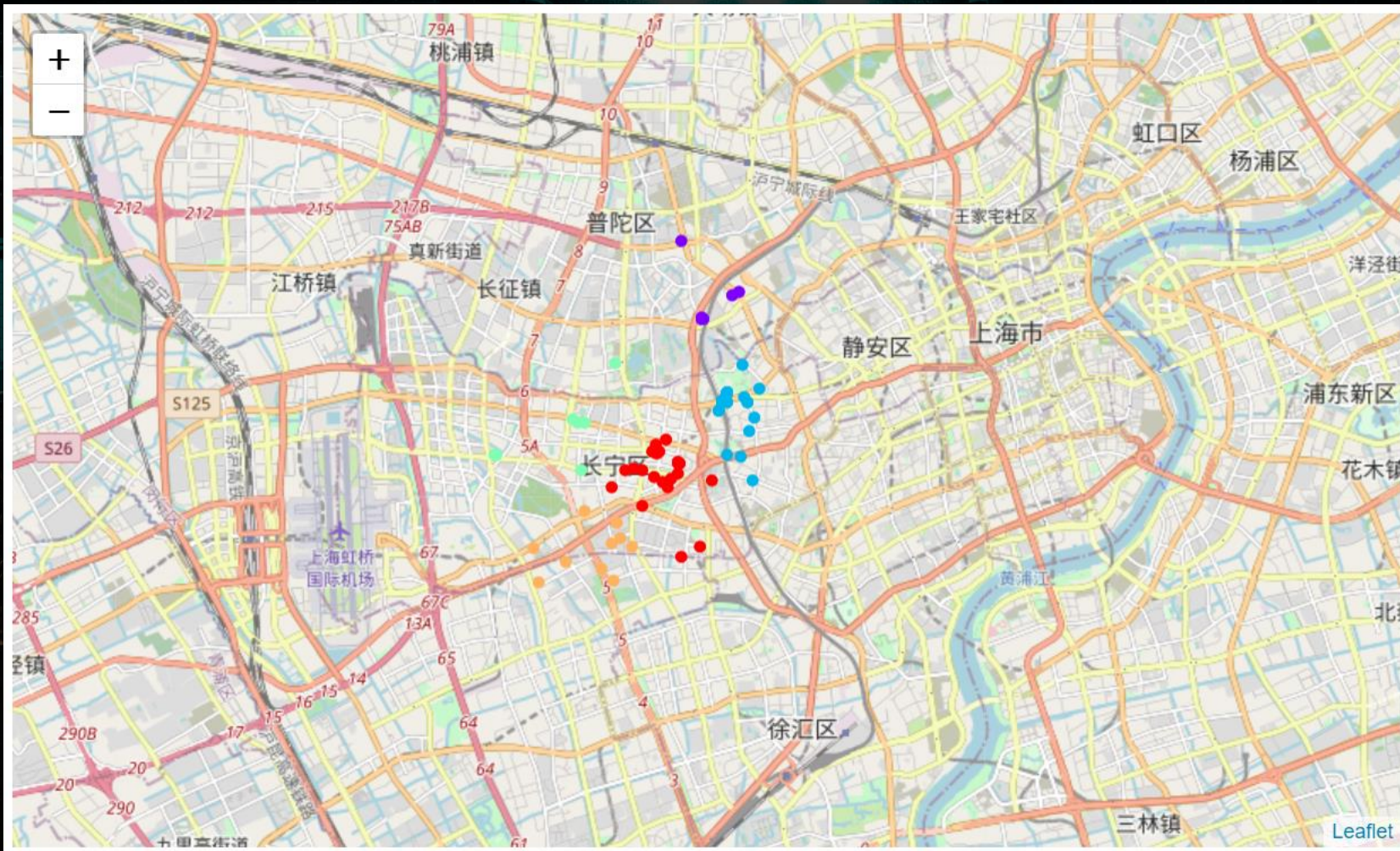


Find the optimal K using the Elbow Method



optimal k = 5

K Means Clustering—— Cluster 2 (k = 5)



Business Understanding

	categories	number
0	Chinese Restaurant	21
1	Cantonese Restaurant	8
2	Japanese Restaurant	6
3	Asian Restaurant	5
4	American Restaurant	3
5	Xinjiang Restaurant	3
6	Fast Food Restaurant	3
7	Seafood Restaurant	3
8	Hong Kong Restaurant	3
9	Taiwanese Restaurant	2
10	Shanghai Restaurant	2
11	Hotpot Restaurant	2

How about the restaurants categories in this cluster?

We can see that the top 5 categories are Chinese, Cantonese, Japanese, Asian and American.

Business Understanding

2.11 RESIDENT FOREIGNERS IN SHANGHAI IN MAIN YEARS

Types	2005	2018
Total	1,171,076	1,171,076
Residence Permitted Foreigners	1,171,076	1,171,076
By Country and Region		
# Japan	87,525	87,525
Republic of Korea	21,561	21,561
Singapore	25,739	25,739
Germany	42,769	42,769
United Kingdom	29,845	29,845
Canada	98	102
United States	98	95
Australia	948	2,404
France	2,404	5,439
By Types		
Work	948	2,404
Study	2,404	5,439
Reunit	5,439	7,311
Private Transaction	7,311	
Journalists		
Long-term Visa Foreigners Above Half Year		

By exploring **2.11 RESIDENT FOREIGNERS IN SHANGHAI IN MAIN YEARS**

(<http://tjj.sh.gov.cn/tjnj/nj19.htm?d1=2019tjnje/E0211.htm>), we can see that there're **an overwhelming number of Japanese** in Shanghai and a lot of them are actually working or living in **Changning** district.

However, there are only 6 Japanese restaurants in this cluster, which is kind of under-supply and also means low competition.

Business Understanding



Considering that Japanese cuisine, compared to Asian and American food, is also loved by many Chinese and foreigners from western countries, we decide to open a new

Japanese restaurant



Data Analysis

Business Understanding



Now that the category is decided, where should we start our new business?

Business Understanding

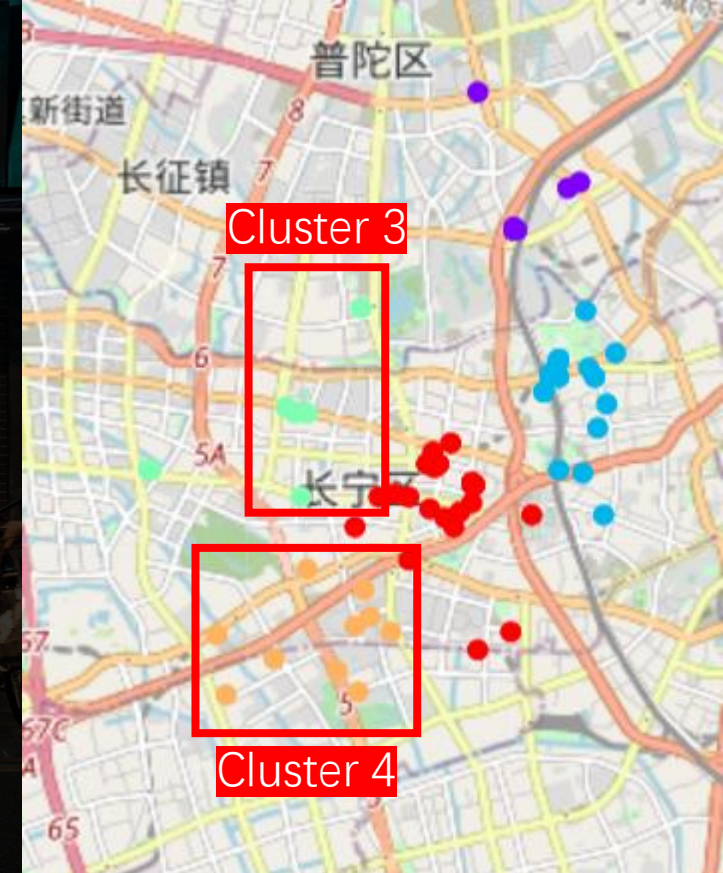
Considering profitability and the insights we gained so far, we will choose our final neighbourhood based on the following principles:

- the neighbourhood is preferable if it is near to residential areas or business areas;
- the neighbourhood is preferable to subway stations;
- there are not too many restaurants of the same kind within the neighbourhood;
- rents are generally affordable.

Business Understanding

Cluster 3 (Green): An outlier without distinctive features.

Cluster 4 (Orange): Covers most of the areas usually called by Shanghainese as 'Gubei', which is historically a wealthy community. Not densely-populated. Filled with residential areas. No big shopping malls or business areas nearby and easy access to subway stations.

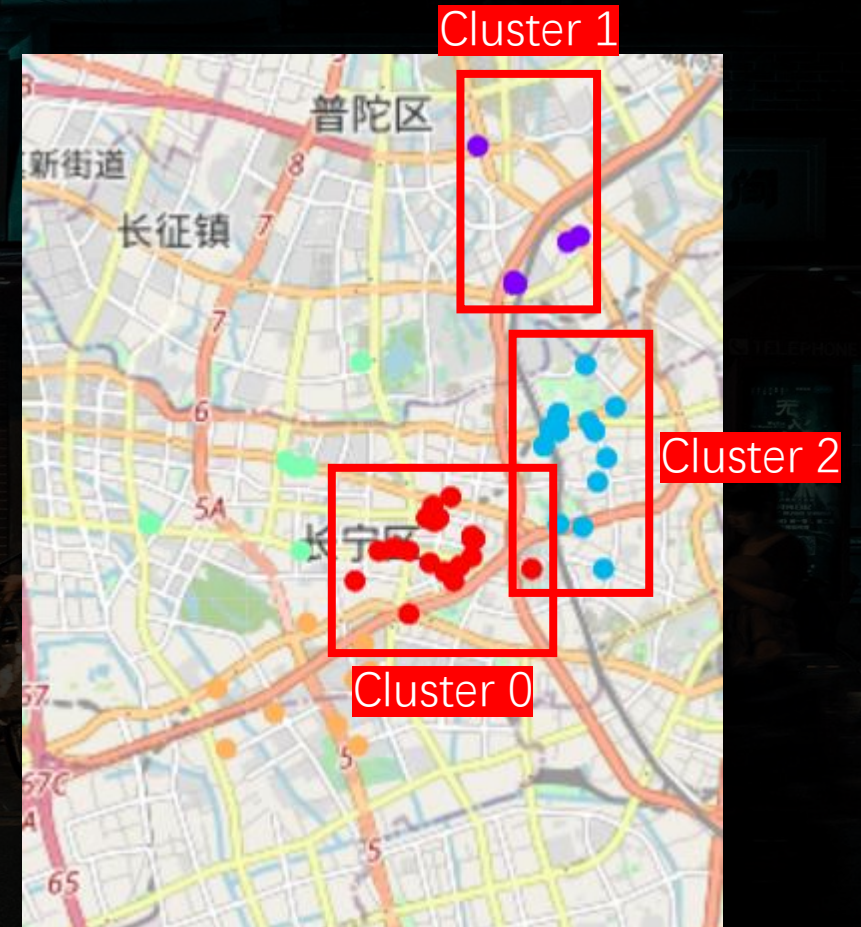


Business Understanding

Cluster 0 (Red): The center of all 5 clusters. Also a sub-regional commercial center surrounded by a lot of shopping malls, business buildings and residential areas. Easy access to subway. Potentially higher rents.

Cluster 1 (Purple) : A sub-regional commercial center with the magnificent Global Harbor Mall and East China Normal University nearby. Easy access to subway. Potentially lower rents.

Cluster 2 (Blue): A famous sub-regional commercial center near Zhongshan Park. Big shopping malls and business areas nearby. Easy access to subway. Potentially higher rents.



Business Understanding

Let's quantify the features of each cluster to decide which one is the final winner.

Here I select 7 most distinctive features arbitrarily bases on business understanding. They are:

- **Rents (lower rent gets higher mark)**
- **Proximity to business areas(the closer the higher mark)**
- **Proximity to residential areas(the closer the higher mark)**
- **Proximity to shopping malls(the further the higher mark)**
- **Proximity to subway stations(the closer the higher mark)**
- **Number of competitors within the cluster(less number gets higher mark)**
- **Purchasing power of residents nearby(greater purchasing power gets higher mark)**

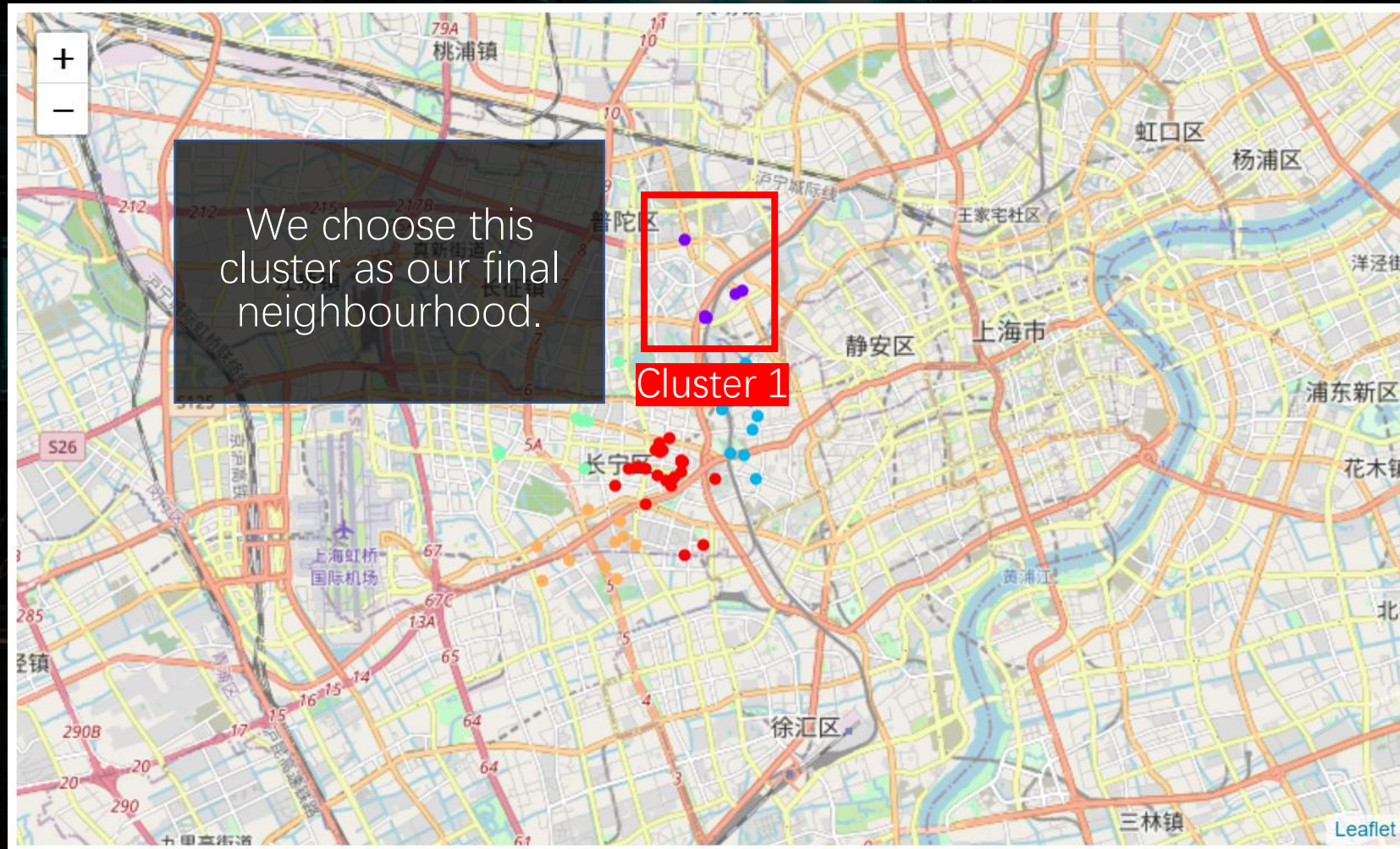
The rating mark will be ranging from 0 to 5, and the total mark will be calculated.

Business Understanding

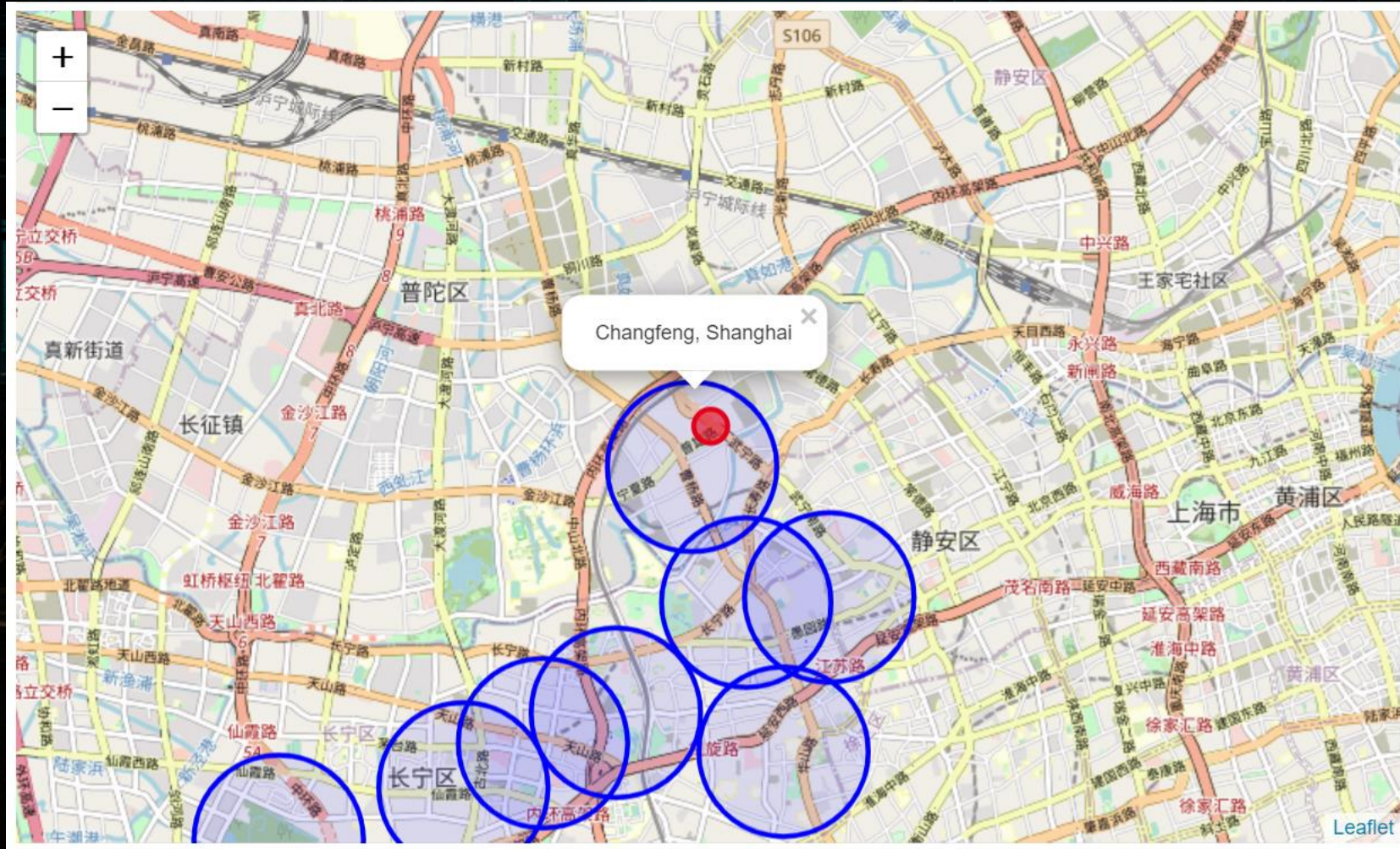
	Cluster	Rents	Proximity to business areas	Proximity to residential areas	Proximity to shopping malls	Proximity to subway stations	Number of competitors within the cluster	Purchasing power of residents nearby	Total score
1	Cluster 1	4	3	5	5	4	5	3	29
3	Cluster 3	5	2	5	5	3	5	3	28
4	Cluster 4	2	3	5	3	5	5	5	28
2	Cluster 2	2	4	4	2	5	4	4	25
0	Cluster 0	2	5	4	2	5	1	4	23

Our winner is Cluster 1, with a total score of 29!

Business Understanding



Final Neighbourhood

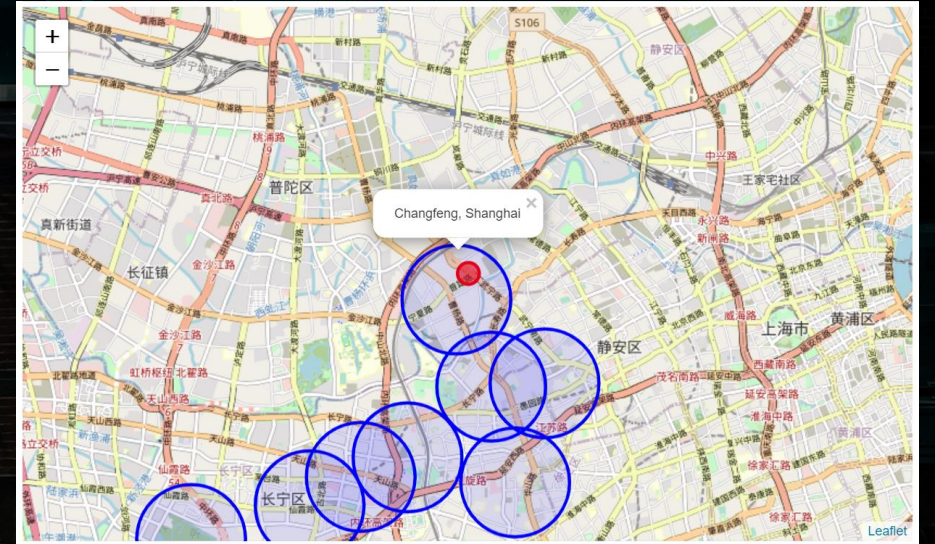


Conclusions

As seen on the map, **Changfeng** is the nearest neighbourhood to Cluster 1.

As such, we can now conclude that:

- The kind of restaurant we will open is **Japanese restaurant**;
- Our new restaurant will be in the neighbourhood of **Changfeng** located in Changning district, Shanghai.



In this project, we utilized multiple data science tools like Python, MySQL and FourSquare API to find out the best neighbourhood in downtown Shanghai to open a new restaurant. Along the way, we also generated useful insights by analyzing the demographic information of 7 central districts in Shanghai. By going through the full cycle of data science using multiple tools, we formed a deeper understanding of what data, programming and business are all about.