The Battle of Neighborhoods—Capstone Project
Opening a Shopping Mall in Shanghai

1.Introduction

1.1 Problem Background

Nowadays, choosing optimal locations while opening commercial sites lays a critical role of whether the business would success. Stakeholders are trying to find the best location in the city to locate their businesses in order to attract more targeted consumers and to gain more revenue. Shanghai is one of the four municipalities in China, and it's the second most populous urban area in the world. Because Shanghai is a global center for finance, more and more stakeholders are attracted to expanding their businesses in Shanghai. However, in order to make sure the location of a commercial site won't have negative impact on growth of revenue, it's critical to optimize a location that attracts the most customers and has the best opportunity of growth.

1.2 Business Understanding

This project will try to find an optimal location (district) to open a shopping mall in Shanghai, China. Since there are so many shopping malls, detecting locations that are not already crowded but still attractive is the main task here. By analyzing data and visualizing maps, I will come up with two to three promising district areas based on the criteria. After filtering down to the promising district areas, I will perform further analysis in details of these areas, and select the best neighborhood to open a shopping mall. The opportunities would be compared with existing neighborhood. The results would benefit the stakeholders who seek opportunity to open a shopping mall in Shanghai.

1.3 Problem Description

A shopping mall is essential in a city where people spend their leisure time purchasing needs, having meals, and enjoying entertainment facilities with friends. Especially in Shanghai, a populous city and a prominent financial center, it becomes a culture for young people to spend time in shopping malls during weekends to relax. Popular shopping malls in Shanghai are:

- Grand Gateway 66
- Shanghai K11
- IAPM Mall
- Shanghai IFC Mall
- Plaza 66
- Metro City Plaza

It is clear that to survive in such competitive market, it's very critical to make strategic plans. Various factors need to be studied in order to decide the optimal location, including but not limited to:

1. Shanghai Population by District
2. Shanghai Demographics
3. Population density
4. Shopping mall density across the districts

5. The potential competitors around the optimal location
6. Segmentation of the borough

1.4 Target Audience

Since the objective is to find the optimal location to open a shopping mall in Shanghai, this report would interest any business leaders/stakeholders who want to expand their businesses or start business for the first time in Shanghai. Once this strategic plan is achieved, it can be applied to multiple other areas.

2. Data

The city will be explored is Shanghai, China. Below are the datasets that are going to be analyzed:

2.1 List of districts in Shanghai

There are sixteen districts in total, the names of districts are obtained from Wikipedia. In order to load them into dataset, I used BeautifulSoup library to scrap from Wikipedia. And the result is as follow:

	Neighborhood		Neighborhood
0	Huangpu District	8	Baoshan District
1	Xuhui District	9	Minhang District
2	Changning District	10	Jiading District
3	Jing'an District	11	Jinshan District
4	Putuo District	12	Songjiang District
5	Hongkou District	13	Qingpu District
6	Yangpu District	14	Fengxian District
7	Pudong	15	Chongming District

The link to the website is:
https://en.wikipedia.org/wiki/List_of_administrative_divisions_of_Shanghai

2.2 Geographical coordinates of districts in Shanghai

To obtain the geographical coordinates—latitude and longitude for each district in Shanghai, I applied geopy library and implement the geocoder, and then combined the data frame with corresponding order. The result is as follow:

	Neighborhood	Latitude	Longitude		Neighborhood	Latitude	Longitude
0	Huangpu District	31.23780	121.47810	9	Minhang District	31.10880	121.37472
1	Xuhui District	31.19594	121.44709	10	Jiading District	31.36637	121.22153
2	Changning District	31.21739	121.42105	11	Jinshan District	30.92025	121.25199
3	Jing'an District	31.22000	121.41583	12	Songjiang District	31.03595	121.21460
4	Putuo District	31.25100	121.38970	13	Qingpu District	31.15394	121.11408
5	Hongkou District	31.25000	121.48917	14	Fengxian District	30.83381	121.52128
6	Yangpu District	31.26193	121.51904	15	Chongming District	31.21739	121.42105
7	Pudong	31.23513	121.52759				
8	Baoshan District	31.41639	121.48000				

2.3 Population Density in each district

In order to optimize the location, we want the shopping mall located in area with high population density, so that the range of our target customers would be broad. I used the same data from the Wikipedia page, and got the following result:

	Neighborhood	Area(km^2)	Population	Density(/km^2)
0	Huangpu District	20.46	658,600	32,190
1	Xuhui District	54.76	1,089,100	19,889
2	Changning District	38.3	691,100	18,044
3	Jing'an District	37.37	1,000,000	27,000
4	Putuo District	54.83	1,288,000	23,491
5	Hongkou District	23.46	809,400	34,501
6	Yangpu District	60.73	1,315,200	21,657
7	Minhang District	370.75	2,537,900	6,845
8	Baoshan District	270.99	2,022,900	7,465
9	Jiading District	464.2	1,568,231	3,378
10	Pudong New Area	1,210.41	5,474,900	4,523
11	Jinshan District	586.05	798,000	1,362
12	Songjiang District	605.64	1,760,200	2,906
13	Qingpu District	670.14	1,209,100	1,804
14	Fengxian District	687.39	1,159,900	1,687
15	Chongming District	1,185.49	696,400	587

2.4 Venue Data from Foursquare API

Foursquare API will provide various categories of venue data, and for our interest, we will be only using the shopping mall category to solve the problem. Following is the screen shot of part of the data:

	Neighborhood	DistrictLatitude	DistrictLongitude	VenueName	VenueLatitude	VenueLongitude	VenueCategory
19	Huangpu District	31.2378	121.4781	Three on the Bund	31.236266	121.486486	Shopping Mall
38	Huangpu District	31.2378	121.4781	Bund18 (外滩18号)	31.240480	121.485575	Shopping Mall
39	Huangpu District	31.2378	121.4781	K11 Art Mall (上海K11购物艺术中心)	31.225486	121.469001	Shopping Mall
54	Huangpu District	31.2378	121.4781	Joy City (大悦城)	31.244856	121.467333	Shopping Mall
68	Huangpu District	31.2378	121.4781	IFC Mall (国际金融中心商场)	31.238492	121.497902	Shopping Mall

3. Methodology and Results

The optimal location for a shopping mall should include three main factors: high population density, low competition (low shopping mall density), and better visibility.

3.1 Population Density Analysis

While opening a shopping mall, we don't want the area have little population, because it's important to maintain a large customer base. Therefore, a k-means cluster approach is utilized to cluster the districts by population density.

Following is the basic k-means algorithm:

1: Select K points as initial centroids.
2: repeat
3: Form K clusters by assigning each point to its closest centroid.
4: Recompute the centroid of each cluster.
5: until Centroids do not change.

After performing k-means algorithm by importing scikit-learn library, I got the following result:

	Neighborhood	Latitude	Longitude	Density(/km^2)	Cluster Labels
7	Pudong	31.23513	121.52759	6845	0
8	Baoshan District	31.41639	121.48000	7465	0
9	Minhang District	31.10880	121.37472	3378	0
10	Jiading District	31.36637	121.22153	4523	0
11	Jinshan District	30.92025	121.25199	1362	0
12	Songjiang District	31.03595	121.21460	2906	0
13	Qingpu District	31.15394	121.11408	1804	0
14	Fengxian District	30.83381	121.52128	1687	0
15	Chongming District	31.21739	121.42105	587	0
0	Huangpu District	31.23780	121.47810	32190	1
3	Jing'an District	31.22000	121.41583	27000	1
5	Hongkou District	31.25000	121.48917	34501	1
1	Xuhui District	31.19594	121.44709	19889	2
2	Changning District	31.21739	121.42105	18044	2
4	Putuo District	31.25100	121.38970	23491	2
6	Yangpu District	31.26193	121.51904	21657	2

It's hard to visualize through the table, so here is the map:
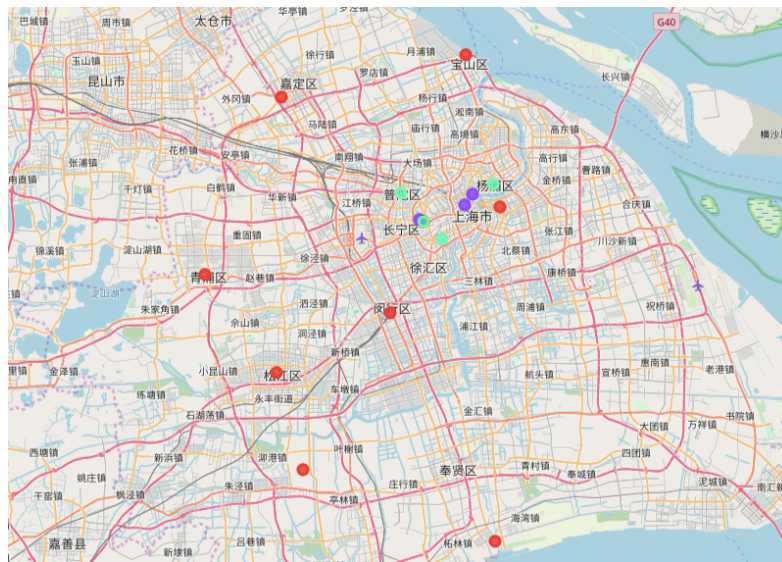
(Red: Cluster 0, Purple: Cluster 1, Green: Cluster 2)

We can see that cluster 1 and 2 have districts with larger population density, and they are all located near the city center.
Based on the bar graph below, it's clearer to determine our candidates based on population density. And the list is:
- Changning District
- Putuo District

- Yangpu District
- Huangpu District
- Xuhui District
- Jing'an District
- Hongkou District

Population Density

Sum of Density(/km^2) for each Neighborhood. Color shows sum of Cluster Labels.
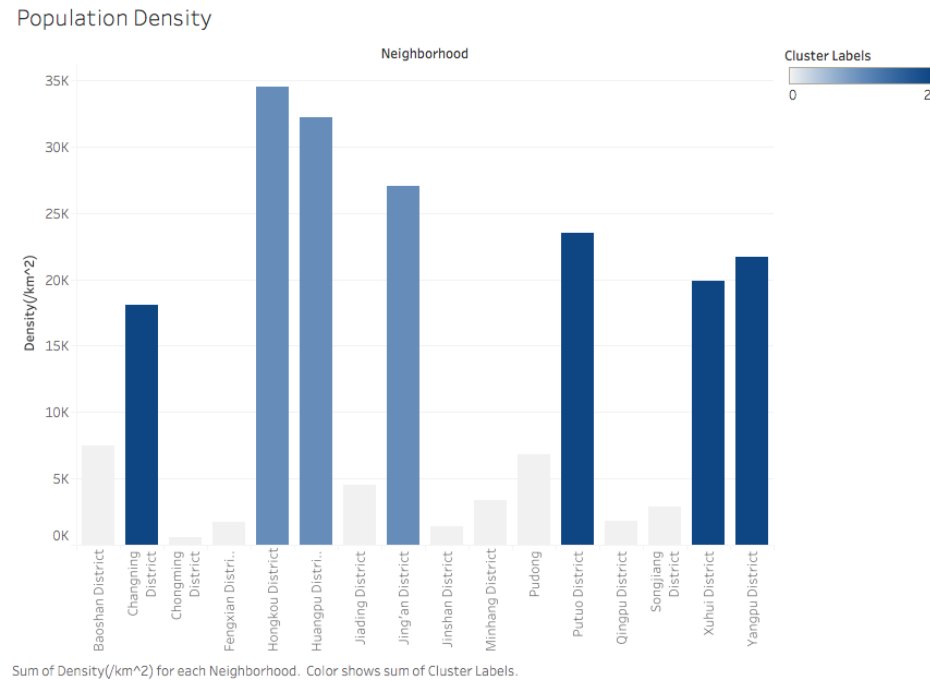
3.2 Competition Analysis

To get familiar with the current market, it's important to analyze the competitive environment. If we choose a shopping mall location that is surrounded by high density shopping malls, we won't have too much competitive advantage. Therefore, other than considering population density, we also need to take competitor density into account. In this method, I use data gathered from Foursquare API to divide the districts into three clusters based on their shopping mall density. Again, the k-means clustering approach is used.

Here is the result represented in the data frame:

|	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude	Density(/km^2)
8	Minhang District	0.076923	0	31.10880	121.37472	3378
10	Putuo District	0.077922	0	31.25100	121.38970	23491
11	Qingpu District	0.111111	0	31.15394	121.11408	1804
12	Songjiang District	0.095238	0	31.03595	121.21460	2906
1	Changning District	0.020000	1	31.21739	121.42105	18044
2	Chongming District	0.020000	1	31.21739	121.42105	587
3	Fengxian District	0.000000	1	30.83381	121.52128	1687
4	Hongkou District	0.030000	1	31.25000	121.48917	34501
5	Huangpu District	0.050000	1	31.23780	121.47810	32190
6	Jiading District	0.000000	1	31.36637	121.22153	4523
7	Jing'an District	0.020000	1	31.22000	121.41583	27000
9	Pudong	0.050000	1	31.23513	121.52759	6845
13	Xuhui District	0.010000	1	31.19594	121.44709	19889
14	Yangpu District	0.020833	1	31.26193	121.51904	21657
0	Baoshan District	0.166667	2	31.41639	121.48000	7465

And here is the map visualization:
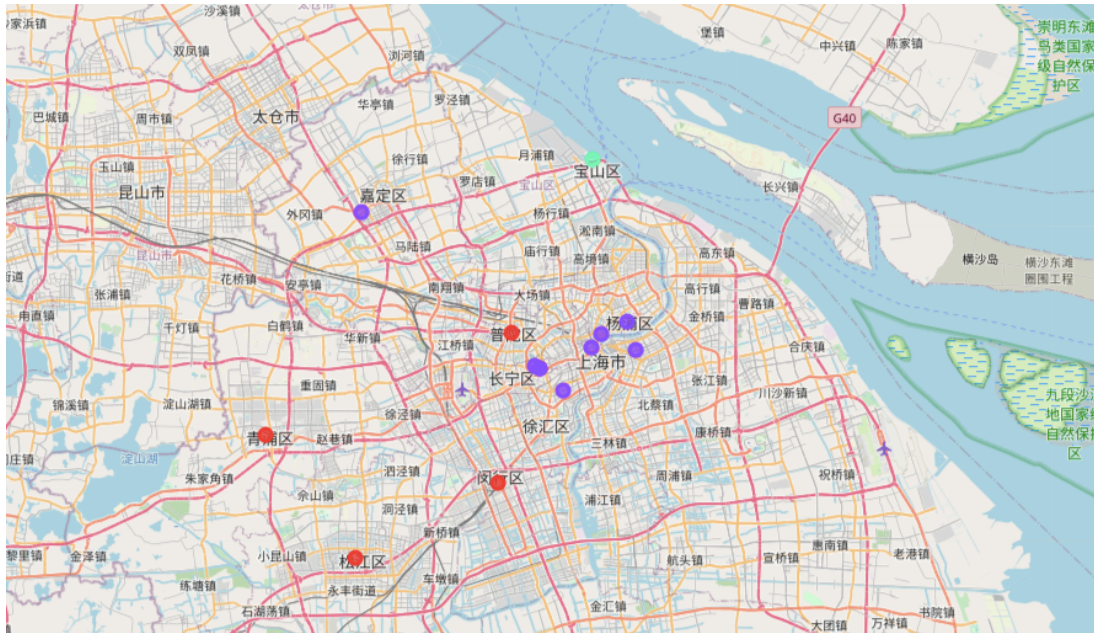
(Red: Cluster 0, Purple: Cluster 1, Green: Cluster 2)

In this step, we came up with the candidate list by selecting districts in cluster 1 that have relatively lower shopping mall density:

- Changning District
- Chongming District
- Fengxian District
- Hongkou District
- Huangpu District
- Jiading District
- Jing'an District
- Pudong District
- Xuhui District
- Yangpu District

3.3 Visibility Analysis

The next step is to check whether our candidates are "visible". We need to make sure the shopping mall to be opened is near the city center, so that not only local residents who can easily reach the location, but also tourists who come to the city and would preferably choose city center to visit. In this method, I used a heatmap to visualize the shopping mall distributions within radius of 1000, 3000, 5000 from the city center.
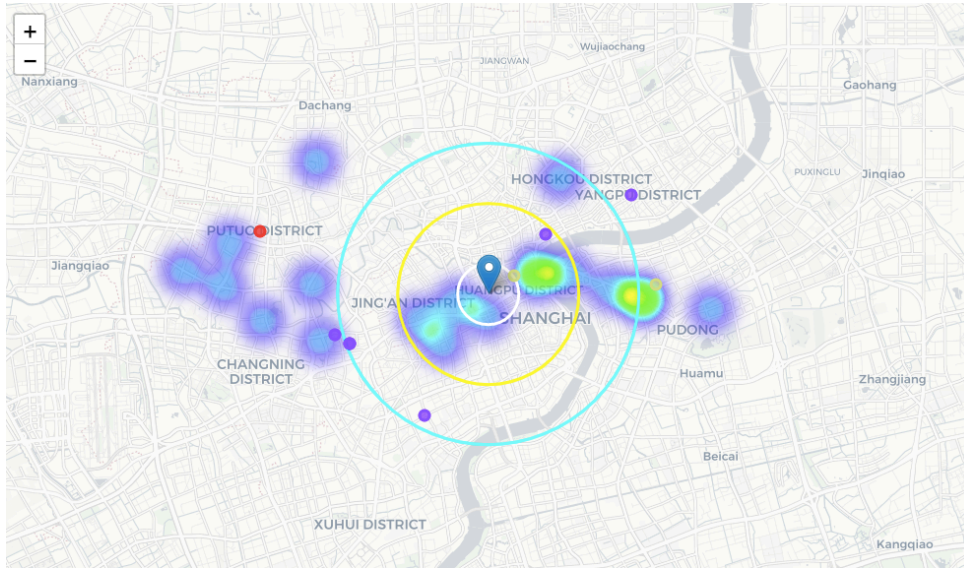
Through the heatmap, we can see that the districts that are within the 5000 km range are:

- Huangpu District
- Hongkou District
- Chongming District
- Xuhui District
- Jing'an District

4. Discussion

As observations and results get from the last step, denser population in Shanghai is concentrated in the city center, with the highest in cluster 2, and moderate in cluster 1. Nevertheless, most of the shopping malls are concentrated in the city center, which makes the competition very intense. In order to avoid high competition, cluster 1 has the strongest candidates, where most have larger opportunities to gain competitive advantage. Meanwhile, districts in cluster 0 and cluster 2 are likely to suffer from intense competition due to oversupply and high concentration. Finally, Shanghai is very famous for tourism. Visitors who come to Shanghai tend to spend most time in the city center to enjoy city sight. Thus, higher visibility of the shopping mall plays a critical role. According to the visibility analysis, five districts that are within 5000 km range from the center are identified.

To combine the results, potential districts are ranked and summarized into top 5. Here is the list:

Result/Rank	Population	Competition	Visibility
1	Hongkou	Jiading & Fengxian	Huangpu
2	Huangpu	Xuhui	Hongkou
3	Jing'an	Changning, Chongming & Jing'an	Xuhui
4	Putuo	Yangpu	Chongming
5	Yangpu	Hongkou	Jing'an

According to this list, Hongkou District and Jing'an District appear in each three result, and Hongkou has a relatively better performance.

5. Conclusion

The purpose of this project is to find an optimal location/district to open a shopping mall in Shanghai. Data including population density, geographical coordinates, shopping mall data are obtained from both Wikipedia and Foursquare API. In the methodology part, three analysis are performed: population density analysis, competition analysis, and visibility analysis. K-means algorithm is mainly used to cluster the districts, and folium map is used for visualization. Tableau is also used for a visualization tool. The final result after analysis is that Hongkou District and Jing'an District are the optimal locations to open the shopping mall. And Hongkou might be preferred, since it has a relatively better performance—high population density, low shopping mall density, and better visibility (close to the city center). This result may benefit the stakeholders who are seeking market opportunity to open a shopping mall in Shanghai successfully.

6. Limitation and Further Improvement

In this project, only three factors are considered for identifying the potential location to open the shopping mall in Shanghai. Other factors that can also be included in the analysis are traffic and household income. Most residents in Shanghai still take subway as transportation tool, so places with more metro stations might be preferred. In addition, areas with higher household income may represent higher purchase ability, which indicates that people in these areas might spend more money on luxuries. Future research can use more accurate clustering and classification approach to determine the candidates and perform deeper analysis. What's more, this project made use of the free account of Foursquare API that has limitations with the number of API calls and results returned. Future research can use paid account to bypass these limitations.