Estimate Optimal Store Locations

For a fast-fashion company in Shanghai, China

Di He

APRIL 19, 2020

Content

1.Introduction

1.1 Background

With the fast development of China consumer market, a global fast-fashion brand plans to enter China market. And they have chosen Shanghai, the most fashion city in China, as their first entered city. At first stage, the company plans to open about 8~10 stores, based on their judgments of the market and company's financial ability.

1.2 Business Problem

The business problem is where are the best locations to open those stores. Without considering rent costs, the company should open a store in the most popular place, where bring stores high traffic and target customers. Therefore, we must know where the company's target customer mostly like to visit. It can be breakdown to two questions. First, how to find the place our target customers like to visit. In this case, our customer is those 18~30 girls who like fashion clothes, shoes, and bags. Second, where is the most popular places among all the places they like.

Another problem is to decide the store opening sequence and the possible shopping malls near those optimal locations to open those stores. Maybe we can rank those popular places according to their popularity and open store according to this rank. And we can find nearby shopping malls as a list to give the company's mangers as a reference

1.3 Interest

Those companies with similar target customer market and plans to enter China market may have an interest.

2. Data Acquisition

2.1 Relevant Data

2.1.1 Store Locations of Similar Brands in Shanghai

We want location data stands for places that our target young ladies like to visit. Stores which have similar target market may be a good one. Even those brands may have different products to sell, as long as they target fashion young ladies, their store locations can be our reference. Those young ladies may be interested in clothes, shoes, and bags. Therefore, I selected those brands in Shanghai as similar brands, please see the table. Then store locations of these brands are good reference as places young ladies like to visit.

Products Categories	Brands
Clothes	UNIQLO, ZARA, H&M, GAP, Urban Revivo, ONLY, Vero Moda, Peacebird
Shoes	Clarks, Bella, Hotwind
Bags	Michael Kors

2.1.2 Shopping Mall Locations in Shanghai

Compared with store locations, shopping mall locations can reflect places with high traffic. Therefore, I also used shopping mall locations as relevant data.

2.1.3 Transportation Station Locations in Shanghai

Shanghai is a big city with many metro lines. In some extant, metro station locations stand for popularity. Central business places often have more lines and stations. And those central places are also where the company want to open their store.

2.1.4 Shopping Mall Locations near Estimated Optimal Locations

After deciding optimal locations to open store, we have to find nearby shopping malls to open those stores. Therefore, we need shopping mall locations near those optimal locations we estimate.

2.2 Data Source and Data Cleaning

2.2.1 Gaode Map API

I get store locations, shopping mall locations and metro station locations in Shanghai through Gaode map API. You can find their web here: https://lbs.amap.com/dev.

One row in the metro station data stands only one station. If two metro lines cross and there is just one station. However, I want the station appears twice to better stands the popularity. Therefore, I cleaned this part of data.

2.2.1 Foursquare API

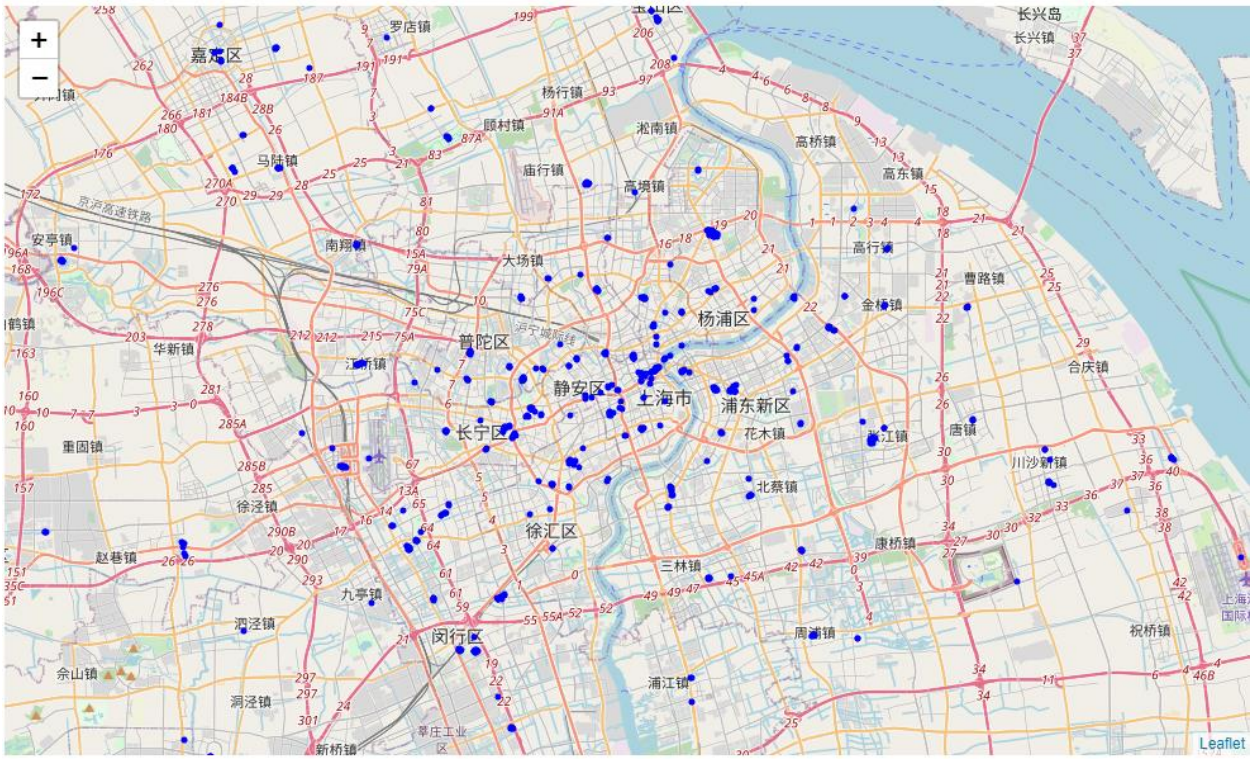
I get shopping mall locations near estimated optimal locations through Foursquare API.

Because I search the venue use the keyword 'Shopping Mall' , and there are several output venues not fit my criteria, such as 'Starbucks'. Therefore, I filtered the output venues to make sure they are really shopping malls.

3. Exploratory Data Analysis

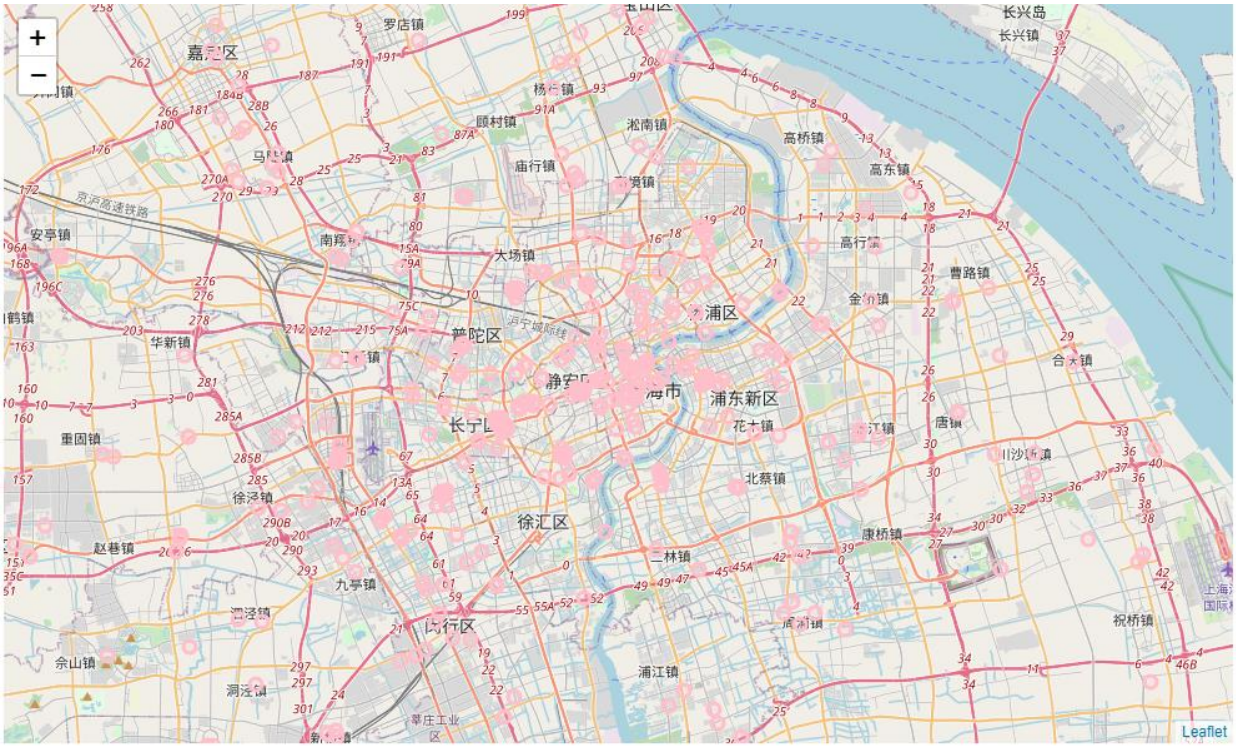
3.1 Store Map

According to the data we get from Gaode API, those similar brand's stores quantity are UNIQLO:76, ZARA:20, H&M:31, GAP:29, Urban REVIVO:21, ONLY:103, Vero Moda:69, Peacebird:33, hotwind:98, Charles & Keith:19, Clarks:33, Belle:76, Michael Kors:13. In this Store map, every blue point stands for a store. From the map, we can see there are more stores at central places.



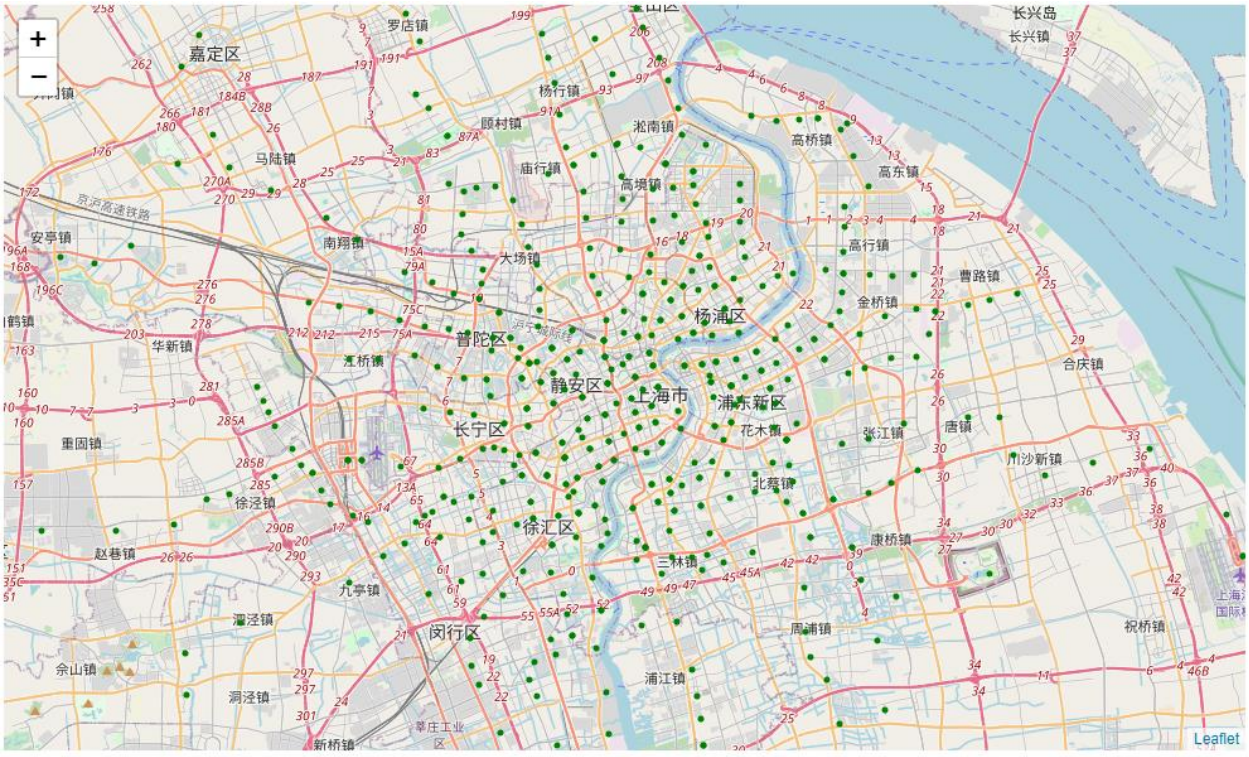
3.2 Shopping Mall Map

According to the data we get from Gaode API, those are 461 shopping malls in Shanghai. In this Shopping Mall map, every pink circle stands for a shopping mall. From the map, we can see there are more shopping mall at central places.



3.3 Metro Station Map

According to the data we get from Gaode API, those are 449 metro stations in Shanghai. In this Metro Station map, every green point stands for a metro station. From all the three map, we assume we should open more stored in the central place.



4. Clustering to Estimate Optimal Store Locations

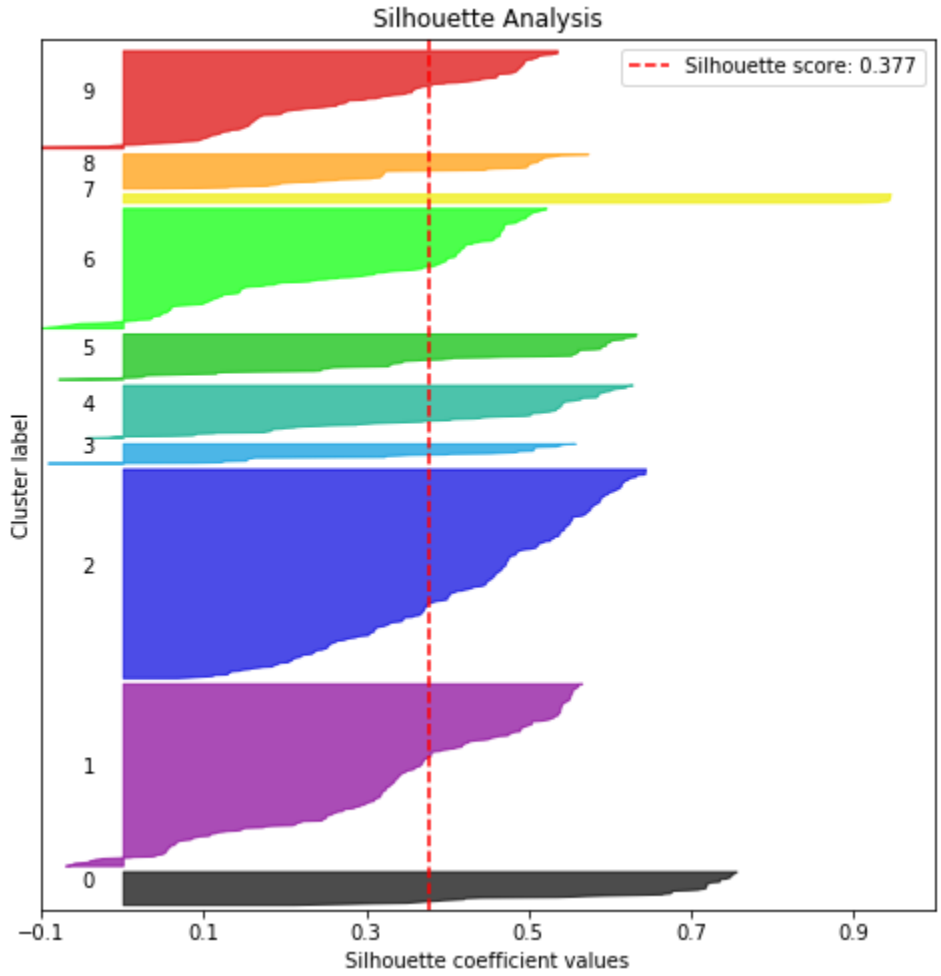
4.1 Explanation

We use latitude and longitude information of those locations in the three maps as input data of clustering. The output centroids are the optimal store locations we want. The centroids should stand for most popular places for young ladies in Shanghai.

4.2 K-Means

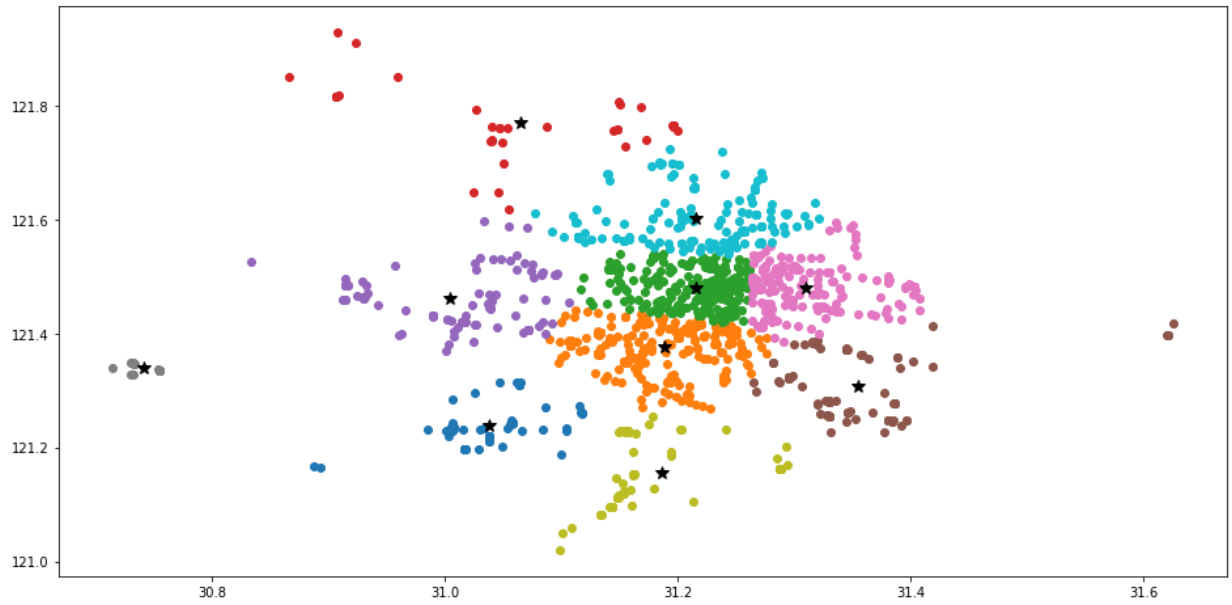
4.2.1 Clustering Outcome

I set Number of Clusters is 10. And the Silhouette Plots shows there is an imbalance among the points numbers in different clusters. For business, it seems that those cluster with less point may cannot attract people.



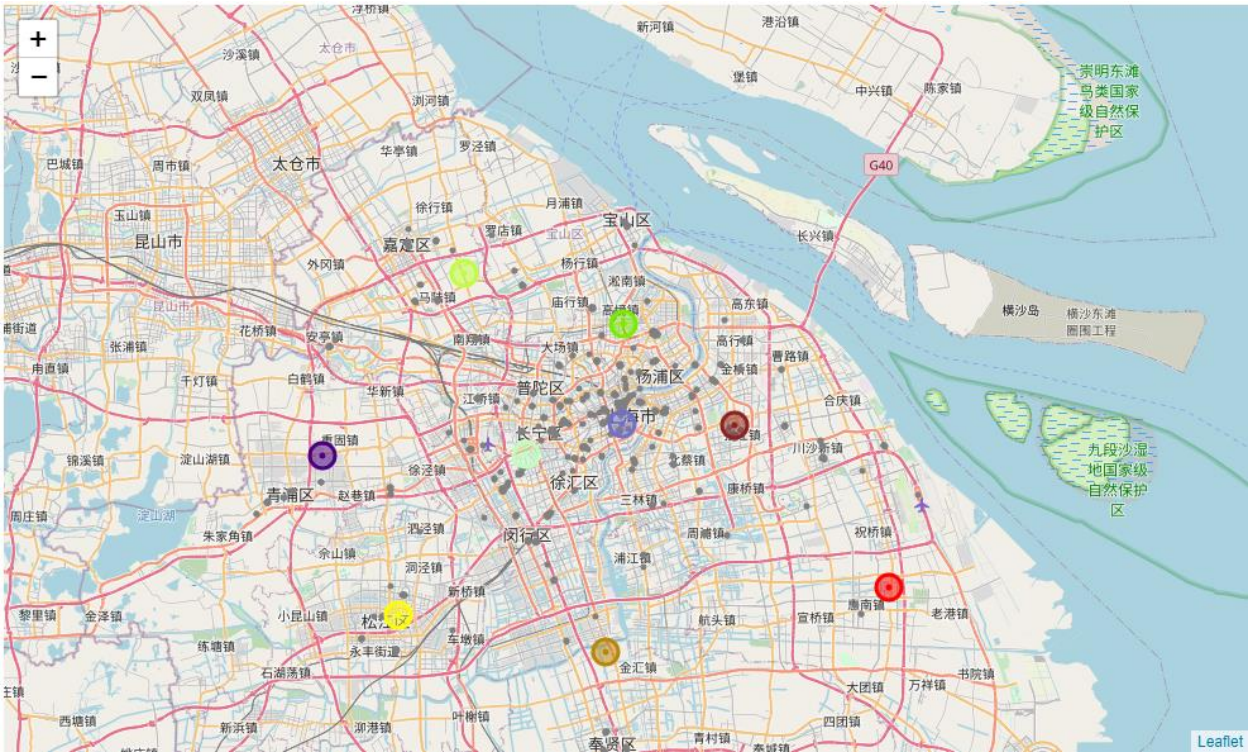
4.2.2 Clustering Plot

In the plot, every color stands a cluster, the black star stands one cluster centroid. From the methodology of K-Means and also the plot, we know it regards the center of data points as the center of the corresponding cluster. However, it is not very proper in our business context, because we want most popular place, rather than most central place.



4.2.3 Clustering Map

In the map, every color stands a cluster centroid. According to my experience, those points are just near popular business places, they are not popular themselves. This outcome also reflects the defect we analysis from the clustering plot and its methodology. Therefore, to synchronies with business, I try another clustering algorithm, DBSCAN, which is based on density.

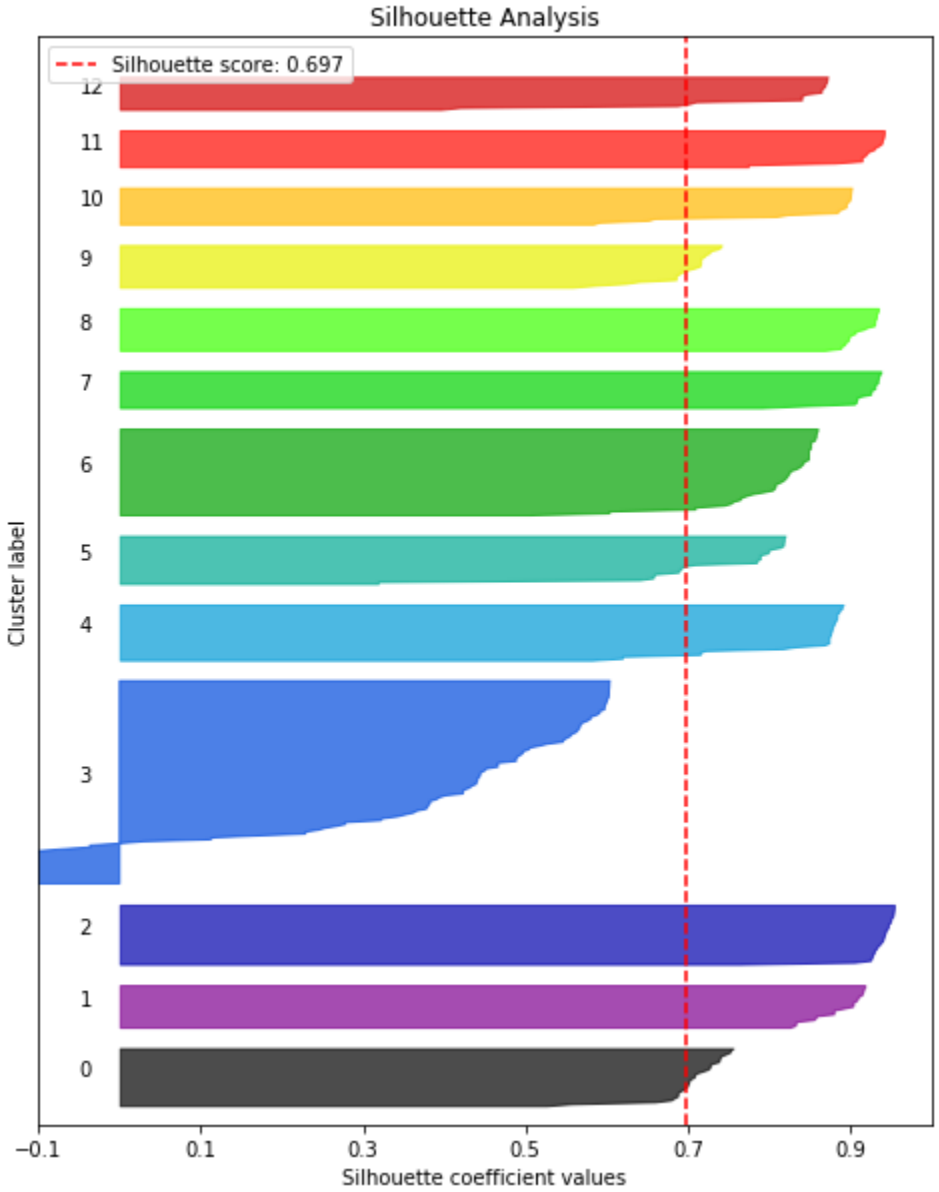


4.3 DBSCAN

4.2.1 Clustering Outcome

I set the maximum distance between two samples is 0.008 and minimum samples is 10. In the DBSCAN outcome, the Number of Clusters is 10. While the noise ratio is 80.33%, which means there are lots of data points are noises. Maybe those points are in low density, which means those places are not popular enough.

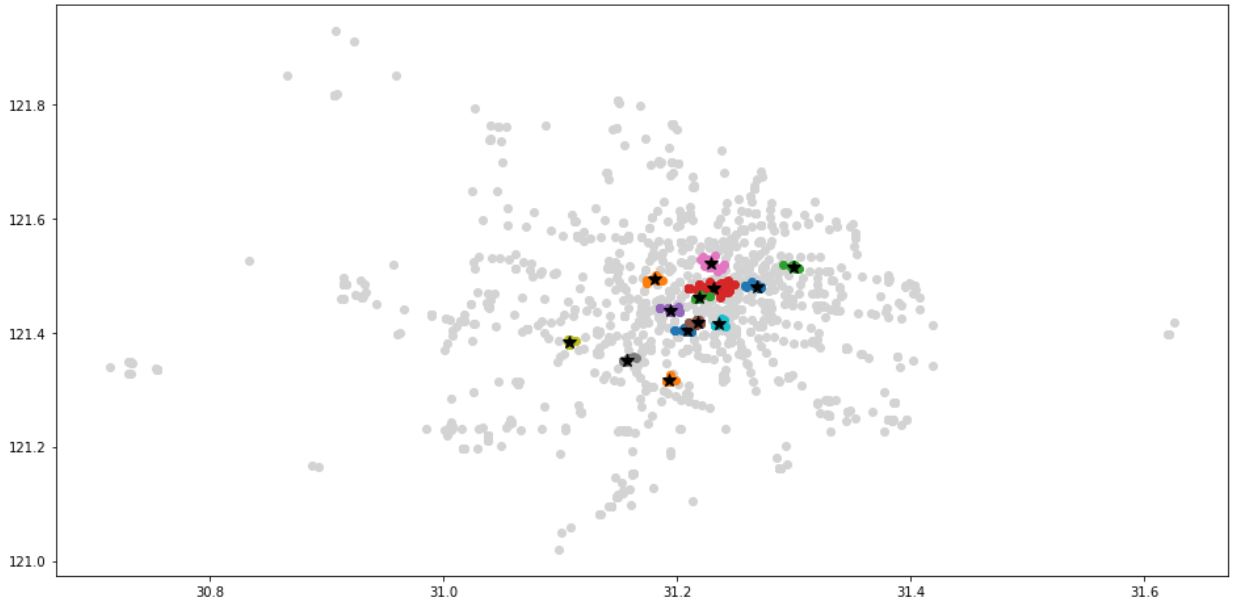
Apart from those noises, the Silhouette Plot seems very nice.



4.3.2 Clustering Plot

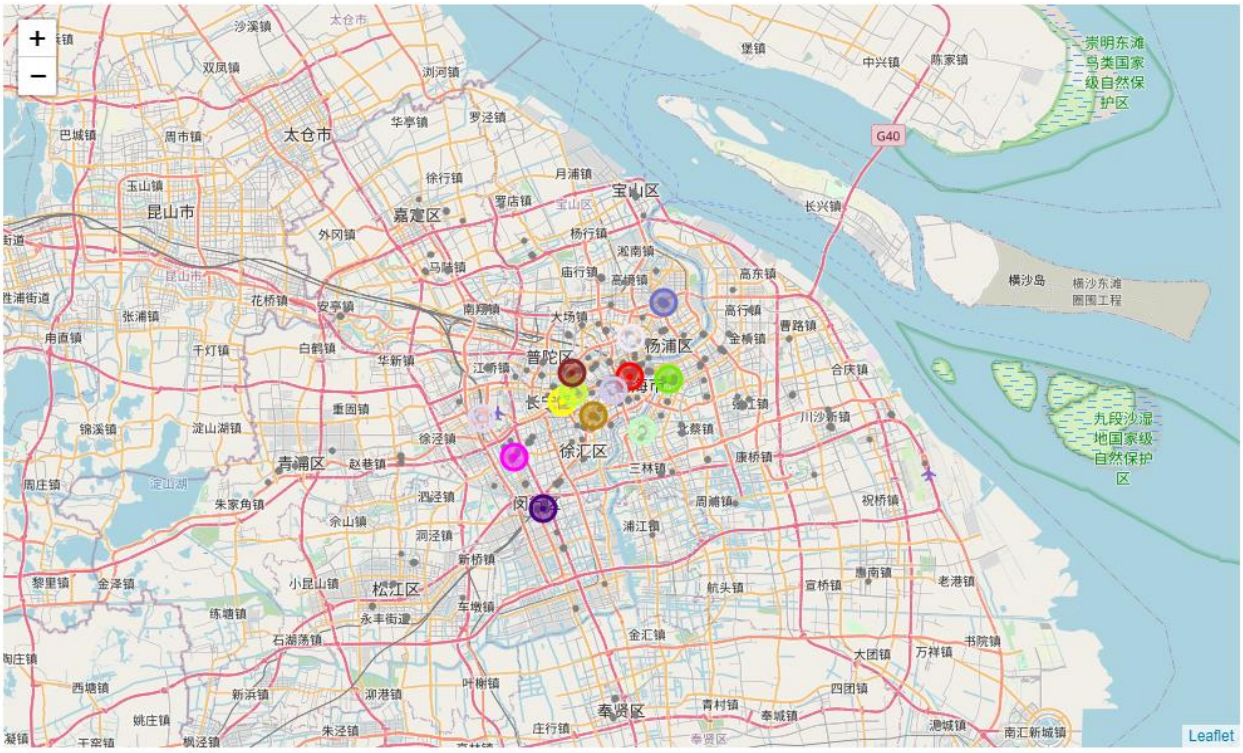
In the plot, every color stands a cluster, except the grey ones, which stands for noises. And The black star stands one cluster centroid. The centroid is not the direct output of DBSCAN, but the average location of points in each cluster calculated by myself.

From the methodology of DBSCAN and also the plot, we know it considers the data which is in the region with high density of the data space to belong to the same cluster. And that is fit our business logic, to find the most popular place.



4.2.3 Clustering Map

In the map, every color stands a cluster centroid. Compared to the K-Means clustering map, those points are more central. And according to my experience, those points are very popular business places. This outcome also reconciles foregoing analysis.



4.4 Clustering Result

Based on those analysis, we finally use the clustering outcome of DBSCAN as top 10 popular places in Shanghai that fit our business.

5. Rank those popular Locations and find Nearby Shopping Malls

5.1 Explanation

We use shopping malls data near the top 10 popular places as a list of possible shopping malls to contact to open store. We also use the number of those malls as a popularity ranking standard. The place with most nearby shopping malls is the place the company should open its first store, because of its popularity.

5.2 Final Map and Usage

Every point stands for recommended store open locations. The radius stands for ranking. The biggest point is the recommended first store to open. And the words popup is the relevant shopping malls nearby to open the store. The company's managers can then visit those shopping malls to evaluate their popularity and quality and contact those shopping malls to negotiate the price.

6. Limitation and Future Improvement

6.1 Limitation

First, population data is not included, which is also a very important consideration. Second, In Shanghai, business centers are changing quickly, this analysis cannot catch this trend. It still relies on business managers' experiences and judgements.

6.2 Future Improvement

Instead of giving numbers of stores (in this case, 8~10), we can use Elbow methods or Silhouette plots to decide the optimal numbers of stores should be open. Further, we can make an interactive platform to let business managers set the numbers of stores they plant to open and see where is recommended by our model.