Estimate Optimal Store Locations

DI HE

APRIL 19, 2020

Content

1. Introduction

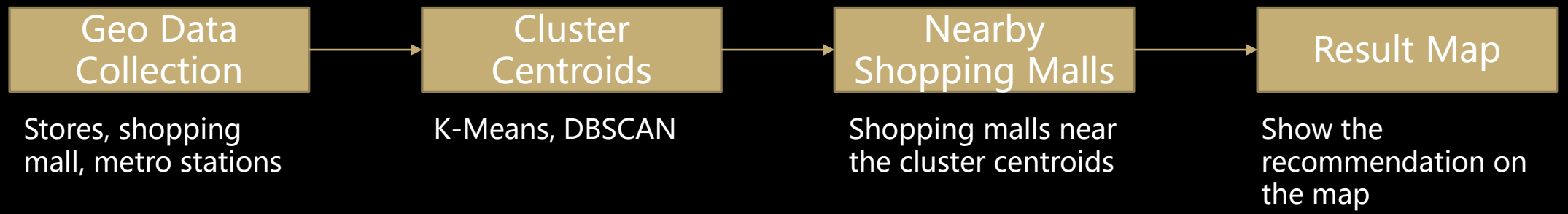
- Introduction
- Framework

Introduction

- Backgrounds: A fast-fashion want to open 8~10 stores in Shanghai, China.
- Business Problem: Where is the best locations to those stores? What are those stores' opening sequence? And the possible shopping malls near those optimal locations to open those stores.
- Delivery: A map with Ranked Recommended Store Locations and Nearby Possible Shopping Malls
- Those companies with similar target market and plans to enter China market may have an interest.

Introduction

- Framework



2. Data Acquisition

- Store Locations of Similar Brands in Shanghai
- Shopping Mall Locations in Shanghai
- Transportation Station Locations in Shanghai
- Shopping Mall Locations near Estimated Optimal Location

Data Acquisition

- **Store Locations of Similar Brands in Shanghai (Source: Gaode Map API)**

They stands for places young ladies like to visits.

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

Data Acquisition

- **Shopping Mall Locations in Shanghai (Source: Gaode Map API)**

Compared with store locations, shopping mall locations can reflect places with high traffic.

- **Transportation Station Locations in Shanghai (Source: Gaode Map API)**

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.

Data Acquisition

- **Shopping Malls near Estimated Optimal Locations (Source: Foursquare API)**

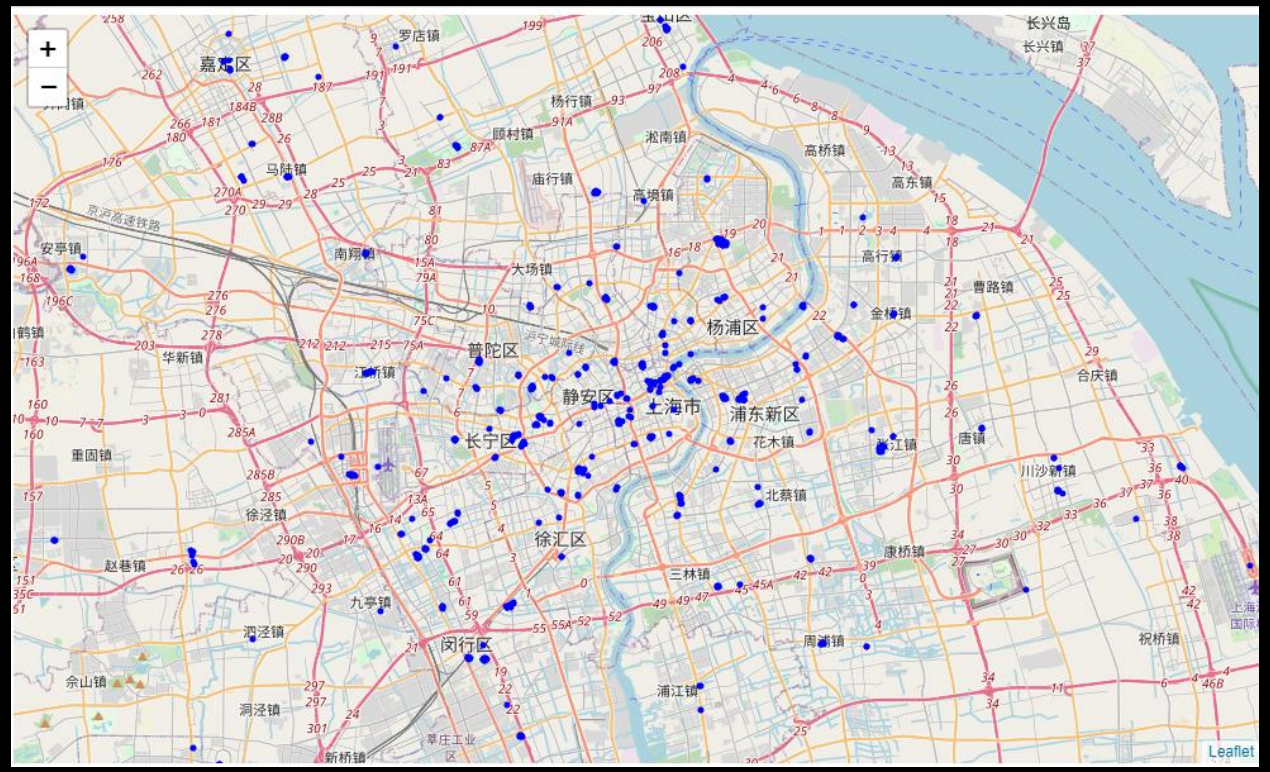
After deciding optimal locations to open store, we have to find nearby shopping malls to open those stores. Therefore, we need shopping mall locations near those optimal locations we estimate.

3. Exploratory Data Analysis

- Store Locations of Similar Brands in Shanghai
- Shopping Mall Locations in Shanghai
- Transportation Station Locations in Shanghai

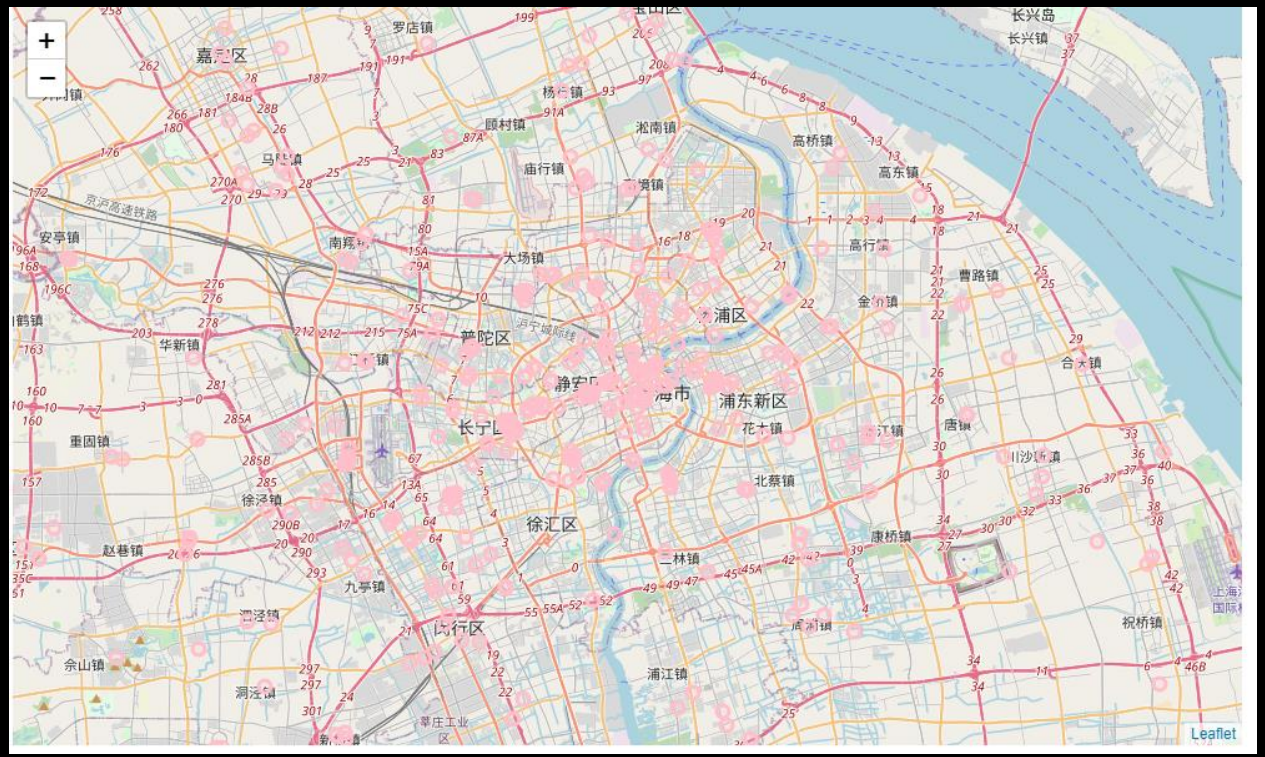
Exploratory Data Analysis

- **Store Map**



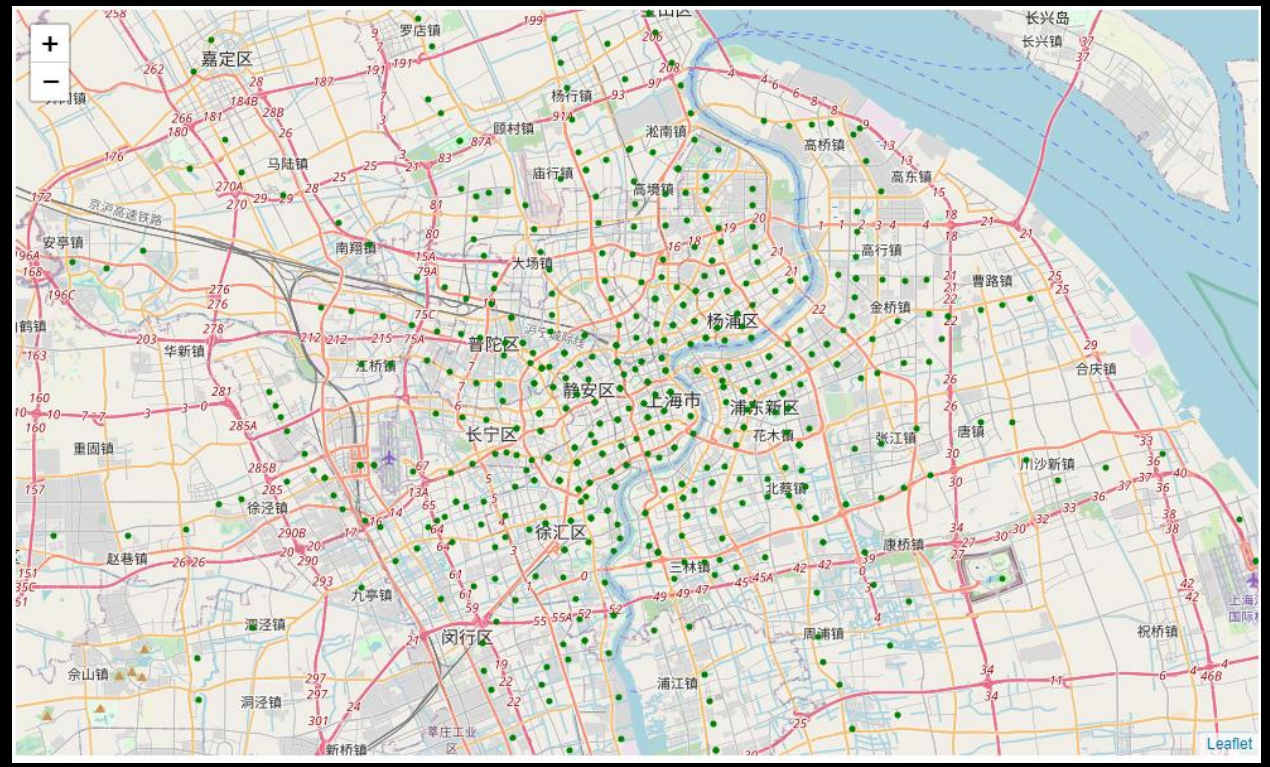
Exploratory Data Analysis

- **Shopping Mall Map**



Exploratory Data Analysis

- Metro Station Map



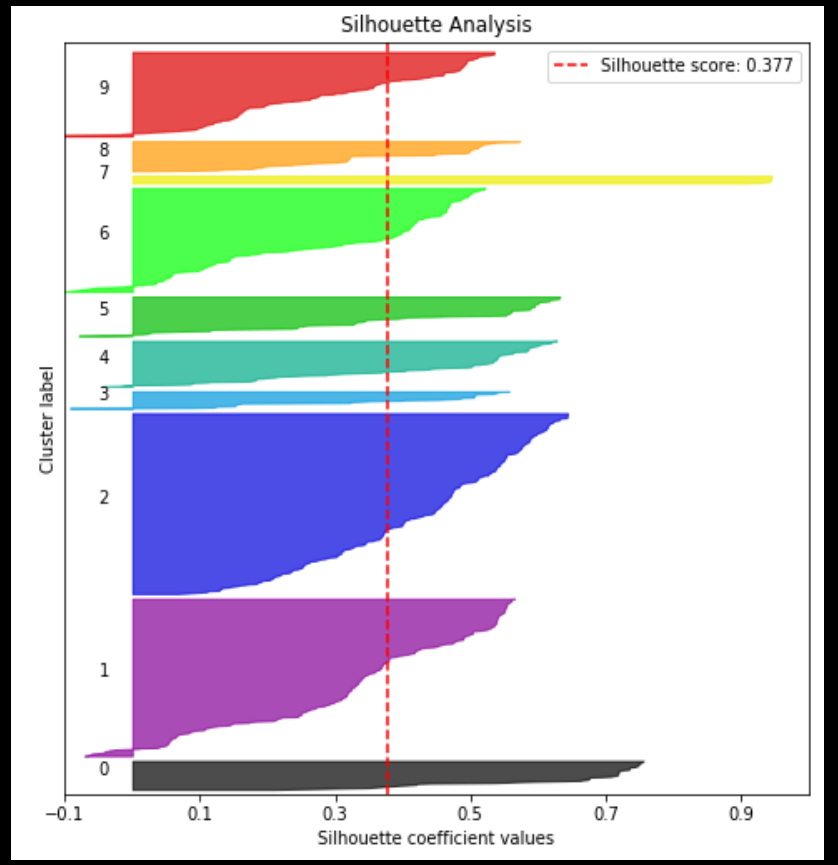
4. Clustering to Estimate Optimal Store Locations

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.

Clustering to Estimate Optimal Store Locations

- **K-Means**

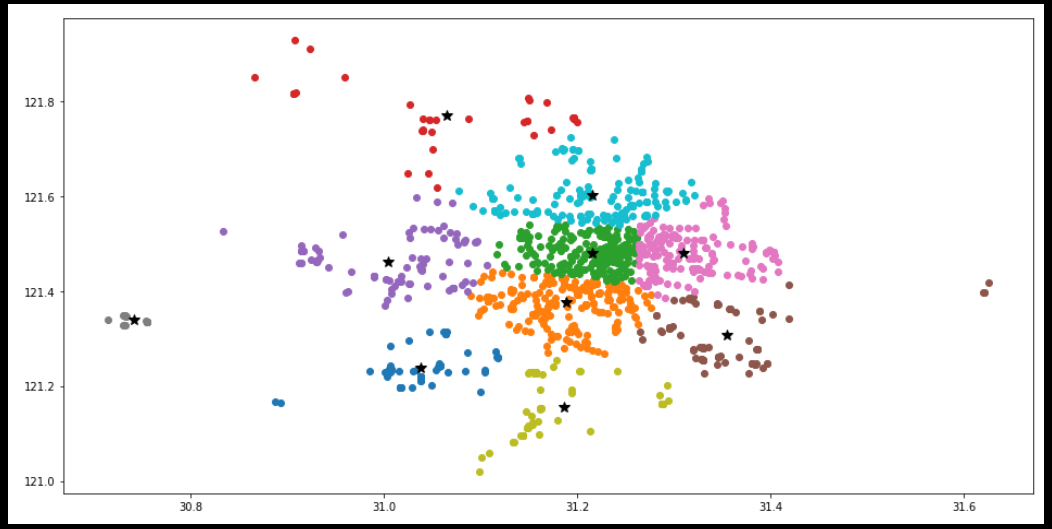
I set Number of Clusters is 10. And the Silhouette Plots shows there is an imbalance among the points numbers in different clusters. For business, it seems that those cluster with less point may cannot attract people.



Clustering to Estimate Optimal Store Locations

- **K-Means**

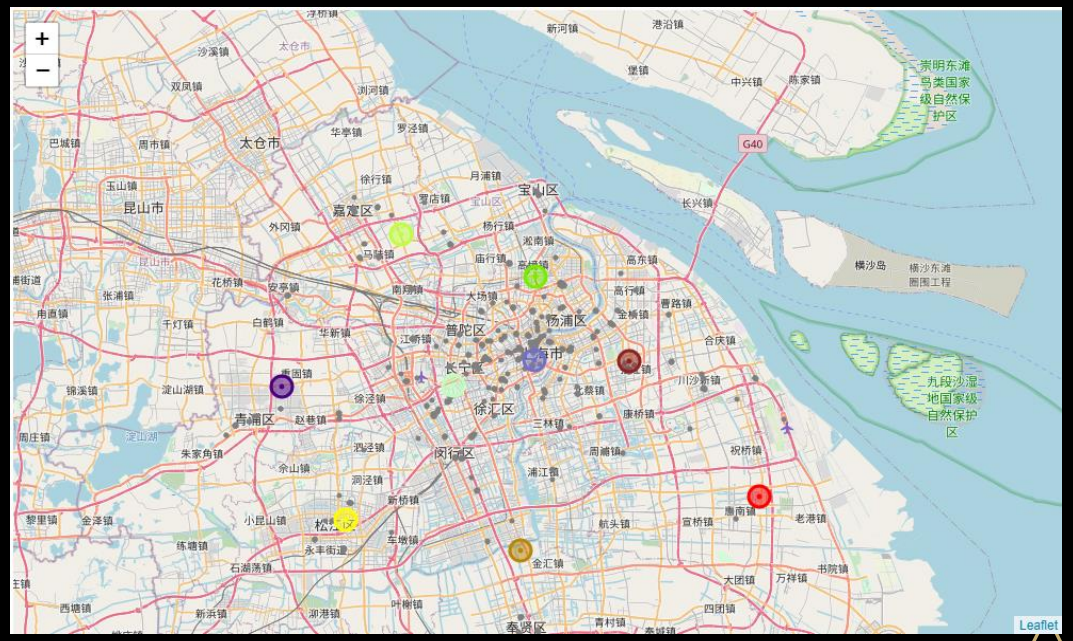
In the plot, every color stands a cluster, the black star stands one cluster centroid. From the methodology of K-Means and also the plot, we know it regards the center of data points as the center of the corresponding cluster. However, it is not very proper in our business context, because we want most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



Clustering to Estimate Optimal Store Locations

- **K-Means**

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.

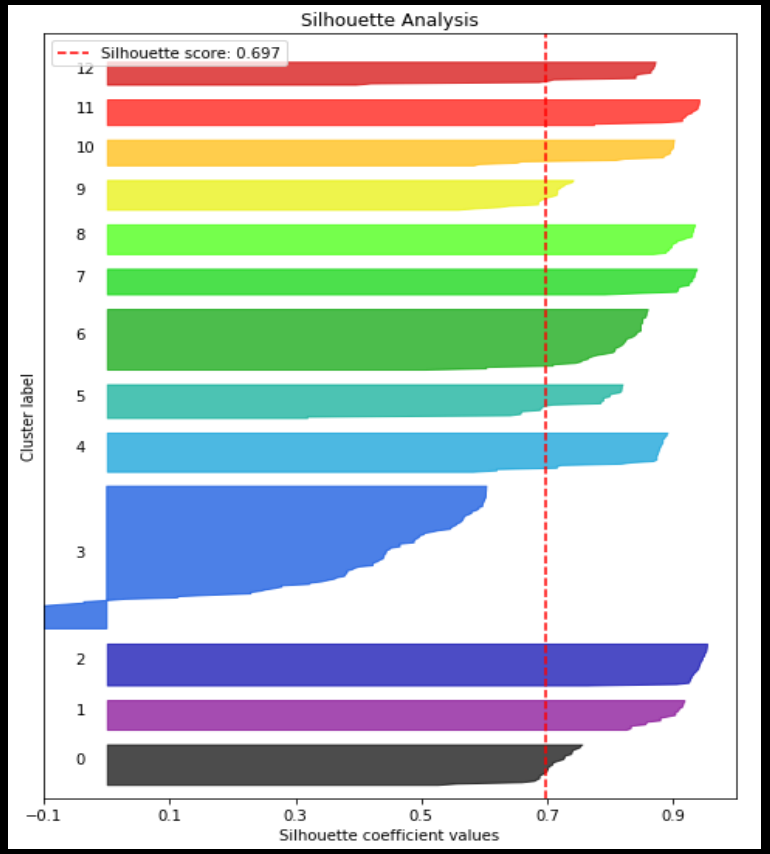


Clustering to Estimate Optimal Store Locations

- **DBSCAN**

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.

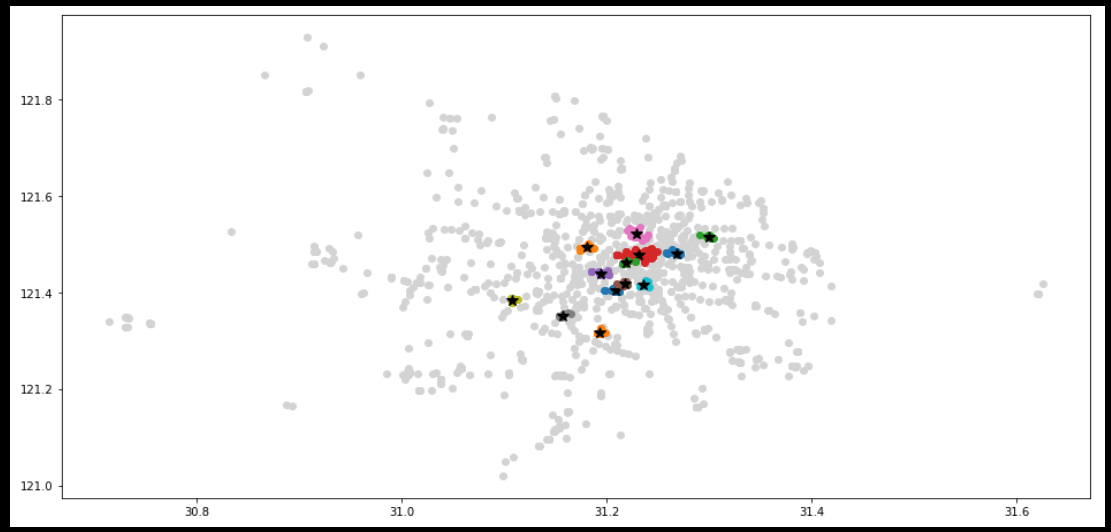


Clustering to Estimate Optimal Store Locations

- **DBSCAN**

In the plot, every color stands a cluster, except the grey ones, which stands for noises. And The black star stands one cluster centroid.

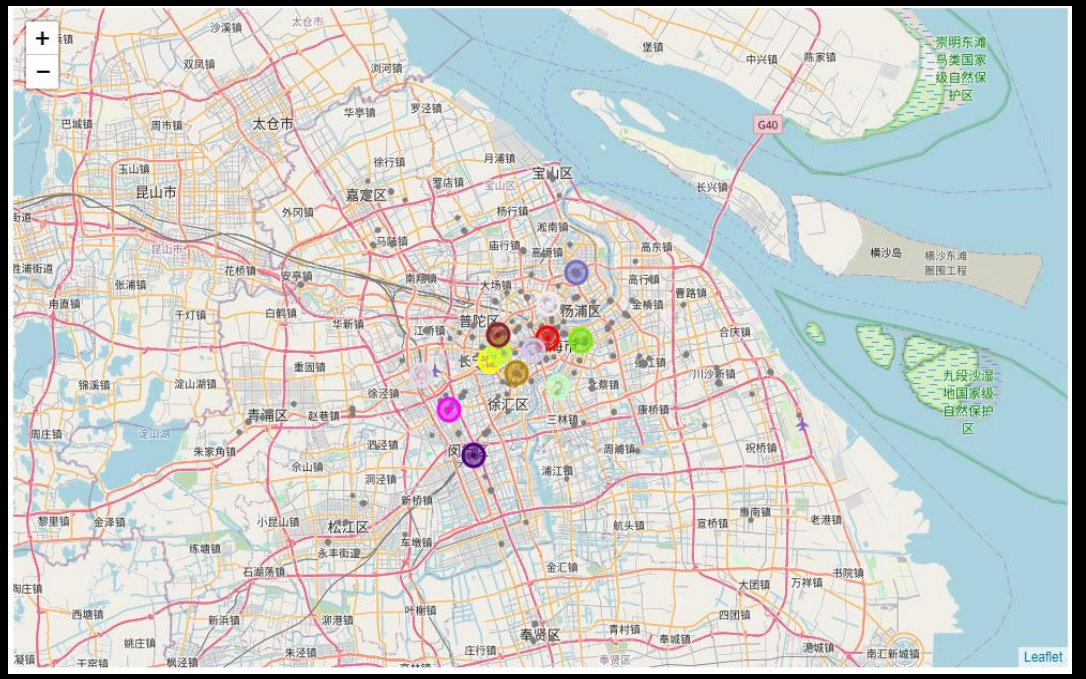
From the methodology of DBSCAN and also the plot, we know it considers the data which is in the region with high density of the data space to belong to the same cluster. And that is fit our business logic, to find the most popular place.



Clustering to Estimate Optimal Store Locations

- **DBSCAN**

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.



Clustering to Estimate Optimal Store Locations

- **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 Malls

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.

Rank those popular Locations and find Nearby Malls

- **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

- Limitation
- Future Improvement

Limitation and Future Improvement

- **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.

Limitation and Future Improvement

- **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.