



THE BATTLE OF NEIGHBORHOOD

Where to Start Business in Guangzhou?

Author: Ziqing Xu

Introduction

- In our lives, people are diverse. Some people prefer a cozy lifestyle and choose to work for someone else. However, some people are passionate about challenges, and they start their own businesses to follow their dreams and fulfill their passion.
- Types of entrepreneurship are diverse, such as opening restaurants, coffee shops, IT companies and so on. Then, a significant problem arises, namely, in which specific location to recommend entrepreneurs to start their businesses.
- The goal of the project is to segment and cluster the neighbourhoods by exploring and comparing the neighbourhoods. By analyzing the clusters, we can figure out the best-recommended location to start a specific type of business in Guangzhou.

Data

- Resource: Neighbourhood data of Guangzhou extracted from:
https://en.wikipedia.org/wiki/List_of_township-level_divisions_of_Guangdong
- Geospatial Coordinates of Guangzhou Neighbourhoods will be obtained by using the **geocode** package. Geospatial Coordinates data is important for getting venues data.
- Venues data of each Neighbourhood will be retrieved by using **Foursquare API**

Methodology – Data Requirement

- The basic requirement of a neighbourhood should contain its borough, city, latitude, longitude.

Methodology – Data Collection

	City	Borough	Neighbourhood
0	Guangzhou	Baiyun	Jingtai
1	Guangzhou	Baiyun	Songzhou
2	Guangzhou	Baiyun	Tongde
3	Guangzhou	Baiyun	Huangshi
4	Guangzhou	Baiyun	Tangjing
5	Guangzhou	Baiyun	Xinshi
6	Guangzhou	Baiyun	Sanyuanli
7	Guangzhou	Baiyun	Tonghe
8	Guangzhou	Baiyun	Jingxi
9	Guangzhou	Baiyun	Yongping

- ❑ Web Scraping

- ❑ Tools:

- ❑ Requests library

- ❑ BeautifulSoup library

Methodology – Data Understanding and Analysis (1)

	City	Borough	Neighbourhood	Latitude	Longitude
0	Guangzhou	Baiyun	Jingtai	23.171167	113.260877
1	Guangzhou	Baiyun	Tongde	23.166263	113.229654
2	Guangzhou	Baiyun	Huangshi	23.205192	113.260667
3	Guangzhou	Baiyun	Tangjing	23.175695	113.248646
4	Guangzhou	Baiyun	Xinshi	23.187983	113.255349

Geospatial data is added.



Map of Neighborhoods in Guangzhou

Methodology – Data Understanding and Analysis (2)

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Jingtai	23.171167	113.260877	Wanda International Cinemas (万达国际电影城)	23.173979	113.261186	Multiplex
1	Jingtai	23.171167	113.260877	Wanda Plaza (万达广场)	23.175312	113.261407	Shopping Mall
2	Jingtai	23.171167	113.260877	SUBWAY (赛百味)	23.174236	113.260859	Sandwich Place
3	Jingtai	23.171167	113.260877	Hannashan Korean BBQ	23.174181	113.260906	Korean Restaurant
4	Jingtai	23.171167	113.260877	Boya Holiday Hotel	23.172244	113.263937	Hotel

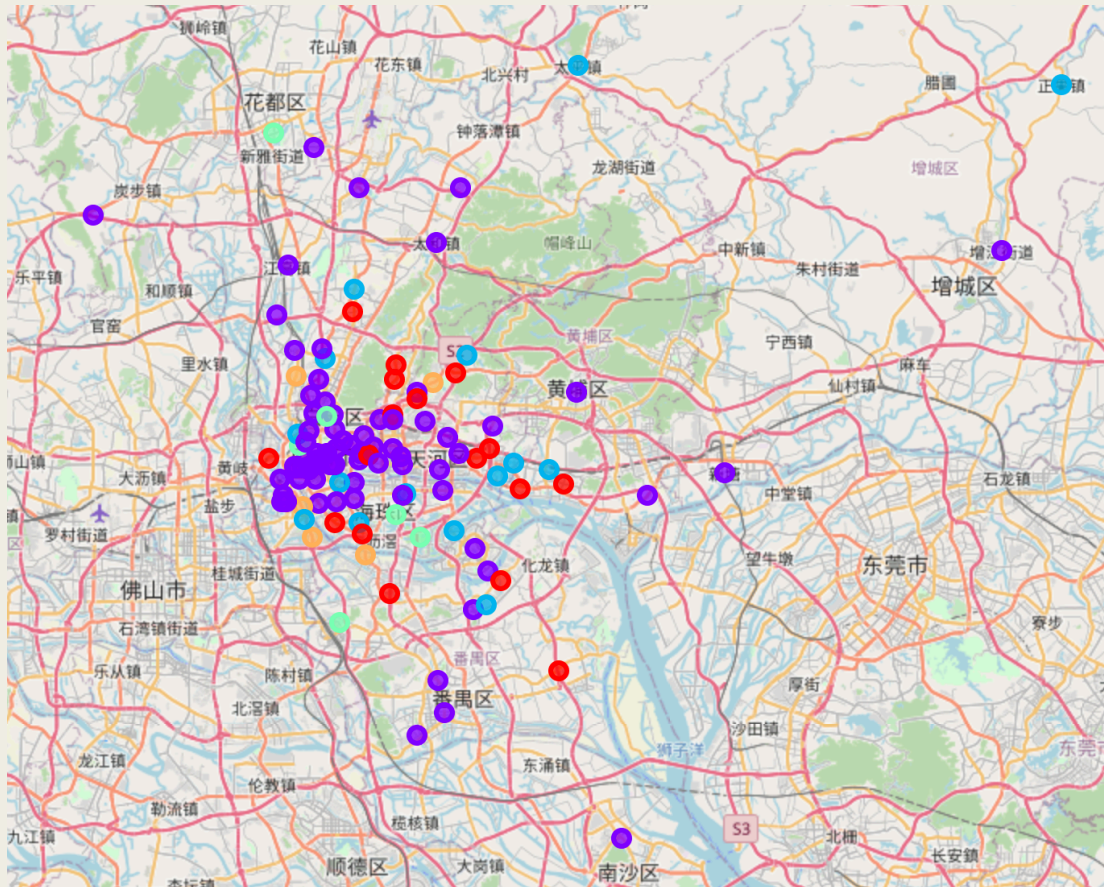
Getting Nearby Venues for each
Neighborhood by Foursquare API

Methodology – Data Understanding and Analysis (3)

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Baihedong	Hotpot Restaurant	Chinese Restaurant	Department Store	Dim Sum Restaurant	Diner	Discount Store	Dog Run	Dongbei Restaurant	Dumpling Restaurant	Farmer's Market
1	Baiyun	Vietnamese Restaurant	Shopping Mall	Food	Asian Restaurant	Clothing Store	Chinese Restaurant	Fast Food Restaurant	Women's Store	Farmers Market	Food
2	Beijing	Nightclub	Pizza Place	Chinese Restaurant	Hotel	Fast Food Restaurant	Restaurant	Convenience Store	Jewelry Store	Noodle House	De
3	Binjiang	Convenience Store	Pharmacy	Bus Station	Noodle House	Sandwich Place	Coffee Shop	Pizza Place	Discount Store	Dog Run	De
4	Caihong	Convenience Store	Hotel	Fast Food Restaurant	Noodle House	French Restaurant	Dim Sum Restaurant	Diner	Discount Store	Dog Run	Don

Top 10 Venues for Each Neighborhood

Methodology – Modeling and Evaluation



- ❑ K means Clustering
- ❑ By using K means clustering, we can analyze the data of each cluster to get insight of the pattern of venues.

Results

- Cluster 1 (Red): restaurants (Chinese, Dim Sum, Cantonese, Fast Food), metro station, and stores(shopping mall, convenience store) are most recommended.
- Cluster 2 (Purple): the types of business are diverse. Based on the result, it is likely to be an area for entertainment. Cafe, Restaurant, Shopping Mall, Resort, Park, and Sport Court are recommended.
- Cluster 3 (Blue): Opening a Chinese Restaurant is most recommended as the first most common venues are Chinese Restaurants for most of the neighbourhood in this cluster. Department Store and Women's Store are also recommended.
- Cluster 4 (Mint Green): Opening a hotel is most recommended as the first most common venues are hotels for most of the neighbourhood in this cluster. From the result, the neighbourhoods in this cluster are likely to be areas for tourism. Starting up a tourism-relative business is recommended.
- Cluster 5: Opening a Restaurant is most recommended as most of the common venues are about dining. However, Cantonese Food is most popular.

Discussion

- As we see in the result section, the number of neighbourhoods in cluster 2 is much more than in the other clusters. The type of venues are diverse in cluster 2, so it is not easy to make a recommendation in the area covered by cluster 2. Using a bigger K may get a better result.

Conclusion

- Overall, as demonstrated in the result section, we can effectively make a recommendation for entrepreneurs to start up different types of business. I hope this can be applied to different cities.

Thank you