

# Final report Capstone – The battle of districts in Shanghai

## Introduction

Shanghai is one of (if not the) biggest cities in the world. With over 20 million inhabitants there has to be a lot of space to house all these people. It will be interesting to see how this translated into different things that can be done in each district. Shanghai is built up of 10 districts, with one district (the urban district) is built up of several smaller districts in the center. The problem to be solved is to provide an overview for tourists to see what each district has to offer. Although a lot of things are in the center and most tourists will travel mostly here, there is still a big area that is quite unexplored, but may be of great interest. This is something we will try to map in this project. With the results we may be able to show some districts that are different from the others and which tourists may want to add to their itinerary.

## Data

The data that will be used is a composed csv file on district names, postal codes and latitudinal and longitudinal values of the in total 18 districts in Shanghai. These come from the websites that can be found in the references. Furthermore the Foursquare API will be used. This API provides data on different venues in different districts. A lot of information can be found on these venues using the Foursquare API and this information is essential for this research, so it will be used to retrieve this information. We will obtain information on things like venue type, category, location and district. Below two maps can be found where the districts are marked. The second map is zoomed in on the center, as most of the districts are located here.



## Methodology

### Clustering Approach

To compare the similarities of districts, we decided to explore districts, segment them, and group them into clusters to find similar districts. To be able to do that, we need to cluster data which is a form of unsupervised machine learning: k-means clustering algorithm.

## Results

### K-Means Clustering

	District	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	Baoshan	Shopping Mall	Coffee Shop	Metro Station	Department Store	Chinese Restaurant	Office	Fast Food Restaurant	Stadium	Gym / Fitness Center	Electronics Store
1	Changning	Chinese Restaurant	Japanese Restaurant	Coffee Shop	Fast Food Restaurant	Noodle House	Beer Bar	Café	Clothing Store	Lounge	Pizza Place
2	Chongming	Chinese Restaurant	Park	Beach	Yunnan Restaurant	Frozen Yogurt Shop	Department Store	Dessert Shop	Dim Sum Restaurant	Diner	Dumpling Restaurant
3	Fengxian	Pizza Place	Asian Restaurant	Auto Garage	Bus Station	Frozen Yogurt Shop	Deli / Bodega	Department Store	Dessert Shop	Dim Sum Restaurant	Diner
4	Hongkou	Plaza	Park	Chinese Restaurant	Shopping Mall	New American Restaurant	Music Venue	Multiplex	Fast Food Restaurant	Deli / Bodega	Department Store
5	Huangpu	Hotel	Chinese Restaurant	Bookstore	Coffee Shop	Noodle House	Concert Hall	Seafood Restaurant	Sandwich Place	Restaurant	Peking Duck Restaurant
6	Jiading	Stadium	Convenience Store	Chinese Restaurant	Bus Stop	Deli / Bodega	Department Store	Dessert Shop	Dim Sum Restaurant	Diner	Dumpling Restaurant
7	Jing'an	Coffee Shop	Hotel	Japanese Restaurant	Bakery	Pizza Place	Chinese Restaurant	Vegetarian / Vegan Restaurant	Theater	Hotel Bar	Shanghai Restaurant
8	Jinshan	Italian Restaurant	Shopping Mall	Convenience Store	Chinese Restaurant	Coffee Shop	Athletics & Sports	Auto Garage	Department Store	Dessert Shop	Dim Sum Restaurant
9	Luwan	Café	Chinese Restaurant	Cocktail Bar	Coffee Shop	Park	Hotel	Shopping Mall	Restaurant	Vegetarian / Vegan Restaurant	Nightclub
10	Minhang	Bakery	Café	Fast Food Restaurant	Plaza	Yunnan Restaurant	French Restaurant	Deli / Bodega	Department Store	Dessert Shop	Dim Sum Restaurant
11	Pudong	Coffee Shop	Bakery	Convenience Store	Sandwich Place	Hotpot Restaurant	New American Restaurant	Metro Station	Pizza Place	Athletics & Sports	Health Food Store
12	Putuo	Seafood Restaurant	Fish Market	Metro Station	Yunnan Restaurant	Frozen Yogurt Shop	Deli / Bodega	Department Store	Dessert Shop	Dim Sum Restaurant	Diner
13	Qingpu	Fast Food Restaurant	Restaurant	Chinese Restaurant	Seafood Restaurant	Shanghai Restaurant	Yunnan Restaurant	Fish Market	Convention Center	Deli / Bodega	Department Store
14	Songjiang	Indian Restaurant	Supermarket	Park	Chinese Restaurant	Metro Station	Convenience Store	Historic Site	Fish Market	Deli / Bodega	Department Store
15	Xuhui	Fast Food Restaurant	Coffee Shop	Metro Station	Hotel	Chinese Restaurant	Bus Station	Brewery	Mobile Phone Shop	Scandinavian Restaurant	Squash Court
16	Yangpu	Hotel	Park	Chinese Restaurant	Metro Station	Flower Shop	Yunnan Restaurant	French Restaurant	Department Store	Dessert Shop	Dim Sum Restaurant
17	Zhabei	Coffee Shop	Fast Food Restaurant	Hotel	Art Gallery	Chinese Restaurant	Café	American Restaurant	Lounge	New American Restaurant	Convenience Store

Most common venues in each district

	District	Postal code	Latitude	Longitude	Cluster Labels	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
0	Huangpu	200001	31.23	121.48	0	Hotel	Chinese Restaurant	Bookstore	Coffee Shop	Noodle House	Concert Hall	Seafood Restaurant	Sandwich Place
1	Luvian	200020	31.22	121.47	0	Café	Chinese Restaurant	Cocktail Bar	Coffee Shop	Park	Hotel	Shopping Mall	Restaurant
2	Zhabei	200070	31.25	121.45	0	Coffee Shop	Fast Food Restaurant	Hotel	Art Gallery	Chinese Restaurant	Café	American Restaurant	Lounge
3	Xuhui	200030	31.18	121.43	0	Fast Food Restaurant	Coffee Shop	Metro Station	Hotel	Chinese Restaurant	Bus Station	Brewery	Mobile Phone Shop
4	Changning	200050	31.22	121.42	0	Chinese Restaurant	Japanese Restaurant	Coffee Shop	Fast Food Restaurant	Noodle House	Beer Bar	Café	Clothing Store
5	Jing'an	200040	31.23	121.45	0	Coffee Shop	Hotel	Japanese Restaurant	Bakery	Pizza Place	Chinese Restaurant	Vegetarian / Vegan Restaurant	Theater
6	Putuo	200333	31.25	121.40	1	Seafood Restaurant	Fish Market	Metro Station	Yunnan Restaurant	Frozen Yogurt Shop	Deli / Bodega	Department Store	Dessert Shop
7	Hongkou	200080	31.27	121.50	0	Plaza	Park	Chinese Restaurant	Shopping Mall	New American Restaurant	Music Venue	Multiplex	Fast Food Restaurant
8	Yangpu	200082	31.27	121.52	0	Hotel	Park	Chinese Restaurant	Metro Station	Flower Shop	Yunnan Restaurant	French Restaurant	Department Store
9	Minhang	201100	31.12	121.38	0	Bakery	Café	Fast Food Restaurant	Plaza	Yunnan Restaurant	French Restaurant	Deli / Bodega	Department Store
10	Baoshan	201900	31.40	121.48	0	Shopping Mall	Coffee Shop	Metro Station	Department Store	Chinese Restaurant	Office	Fast Food Restaurant	Stadium
11	Jiading	201800	31.38	121.27	0	Stadium	Convenience Store	Chinese Restaurant	Bus Stop	Deli / Bodega	Department Store	Dessert Shop	Dim Sum Restaurant
12	Pudong	200135	31.22	121.53	0	Coffee Shop	Bakery	Convenience Store	Sandwich Place	Hotpot Restaurant	New American Restaurant	Metro Station	Pizza Place
13	Jinshan	201500	30.75	121.33	0	Italian Restaurant	Shopping Mall	Convenience Store	Chinese Restaurant	Coffee Shop	Athletics & Sports	Auto Garage	Department Store
14	Songjiang	201600	31.03	121.22	0	Indian Restaurant	Supermarket	Park	Chinese Restaurant	Metro Station	Convenience Store	Historic Site	Fish Market
15	Qingpu	201700	31.15	121.12	0	Fast Food Restaurant	Restaurant	Chinese Restaurant	Seafood Restaurant	Shanghai Restaurant	Yunnan Restaurant	Fish Market	Convention Center
16	Fengxian	201400	30.92	121.47	0	Pizza Place	Asian Restaurant	Auto Garage	Bus Station	Frozen Yogurt Shop	Deli / Bodega	Department Store	Dessert Shop
17	Chongming	202100	31.62	121.40	2	Chinese Restaurant	Park	Beach	Yunnan Restaurant	Frozen Yogurt Shop	Department Store	Dessert Shop	Dim Sum Restaurant



## Cluster result

Below is the cluster result. The blue dot on the island above the city is a cluster, the purple dot in the city is a cluster and all the red dots are a cluster.



## Conclusion & discussion

In this project we have made a model using k-means clustering to cluster districts in Shanghai based on their similarity. As can be seen in the figures above, the result is that 16 of the 18 districts belong to the same cluster. Most tourists going to Shanghai will visit the center, which is the red cluster in the middle of the map. The various red dots around it are in the same cluster, meaning that they are quite similar to the center. The purple dot in the mainland part is the district Putuo. This district is a little outside the center and quite small, but has a population similar to central districts. This makes it logical that it is its own cluster, as there may be things here that are not in other places. In the case of the district Chongming, this is an island quite far from the center. It has less of a big city vibe and has much more nature than Shanghai. These results may imply that tourists going to Shanghai may want to consider going to Putuo and Chongming to experience something different. One thing to note is that the Foursquare API is based on user input. Most of the people that contribute to this API will be from outside China. Since China is not yet a tourist hotspot for these people, it is possible that a lot of places are not mapped. This may have influenced the results quite a bit. However, this is a good starting point for more research on this topic.

## References

<https://jingwei.supfree.net/mengzi.asp?id=811>

<https://www.ip138.com/20/>

[https://en.wikipedia.org/wiki/Shanghai#Administrative\\_divisions](https://en.wikipedia.org/wiki/Shanghai#Administrative_divisions)

[https://raw.githubusercontent.com/AxelIID/Coursera\\_Capstone/main/Shanghai%20Postal%20Code%20data.csv](https://raw.githubusercontent.com/AxelIID/Coursera_Capstone/main/Shanghai%20Postal%20Code%20data.csv)