

Find 5 places in Tokyo to open Milk Tea shops

Ding Xu

Capstone Project - The Battle of Neighborhoods (Week 2)

1. Introduction

As pearl milk tea is becoming more and more popular in Japan, the stakeholder wants to open a milk tea shop as a new business. He needs me to recommend a location in Tokyo to open that shop.

He told me this business mode is very mature in China. The milk tea shops always open near shopping mall, university, bakery and subway station.

He wants me to find 5 best places in Tokyo to open the milk tea shop by analysing the nearby environment. I will select 5 possible places in Tokyo by comparing with the milk tea locations in Shanghai.

2. Data

2.1 Data sources

First, i use the geographic coordinate of milk tea shops in Shanghai and nearby environment. I use these data to analyze the features of places of milk tea shops in Shanghai.

Secondly, i use the the geographic coordinate of some feature places such as shopping centers in Tokyo and nearby environment. I use these data to find out if these places is similar with milk tea shops in Shanghai.

I get the shanghai's data from 'amap' — a map like 'google maps' in china, it provides the free api, so i can take enough data.

	Name	Address	Longitude	Latitude
0	奶茶会所	光明路与高东新路交叉口西北50米	121.617630	31.332172
1	奶茶咖啡	人民路与梧桐路交叉口西北50米	121.496488	31.227198
2	一点奶茶	共和新路647号	121.463955	31.254808
3	四季奶茶	三牌楼路87号	121.492453	31.222938
4	悸动奶茶	光明镇光明路319号	121.516284	30.906174

This is the data i get from amap's api.

I get Tokyo's data from Internet, i find some famous shopping centers in Tokyo, and then use coordinate position conversion to get those longitude and latitude.

	Name	Longitude	Latitude
0	Tokyo Tower	139.745433	35.658580
1	Shinjuku	139.703356	35.693825
2	Ginza	139.766486	35.671223
3	Akihabara	139.774473	35.702259
4	Ikebukuro	139.707731	35.734831
5	Mastuya Ginza	139.766698	35.672256
6	东京上野	139.774154	35.708730
7	Daikanyama	139.704221	35.650547
8	Nakamise-dori Street	139.796454	35.711841
9	Shibuya Center Street	139.699783	35.660046
10	東京駅一番街	133.918413	34.665540
11	Ameyoko Arcade	139.774383	35.710008

This is the data of Tokyo.

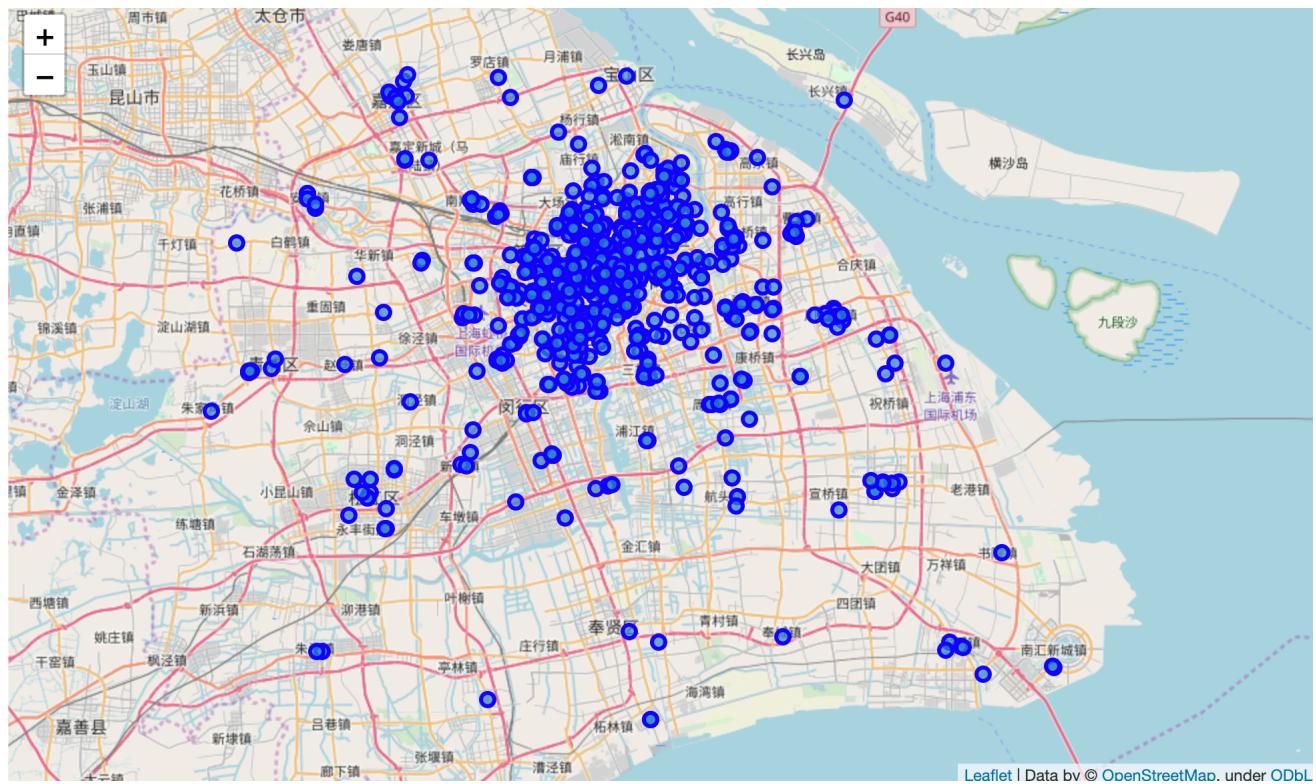
And i get both venues of shanghai and Tokyo from foursquare.

2.2 Data pretreatment

As the longitude and latitude got from 'amap' is together, my first things is split them.

Then conside the number of Tokyo's data, i reduce my shanghai's data from 850 to 100. One hundred is enough to find out the features.

Next, i show all the milk tea shops in shanghai in the map:



Also, i show the shopping centers in Tokyo i collected in the map:



Finally, i consider i would do k-means, so i use one hot encoding to venues of Shanghai and Tokyo.

3. Methodology

My methodology is

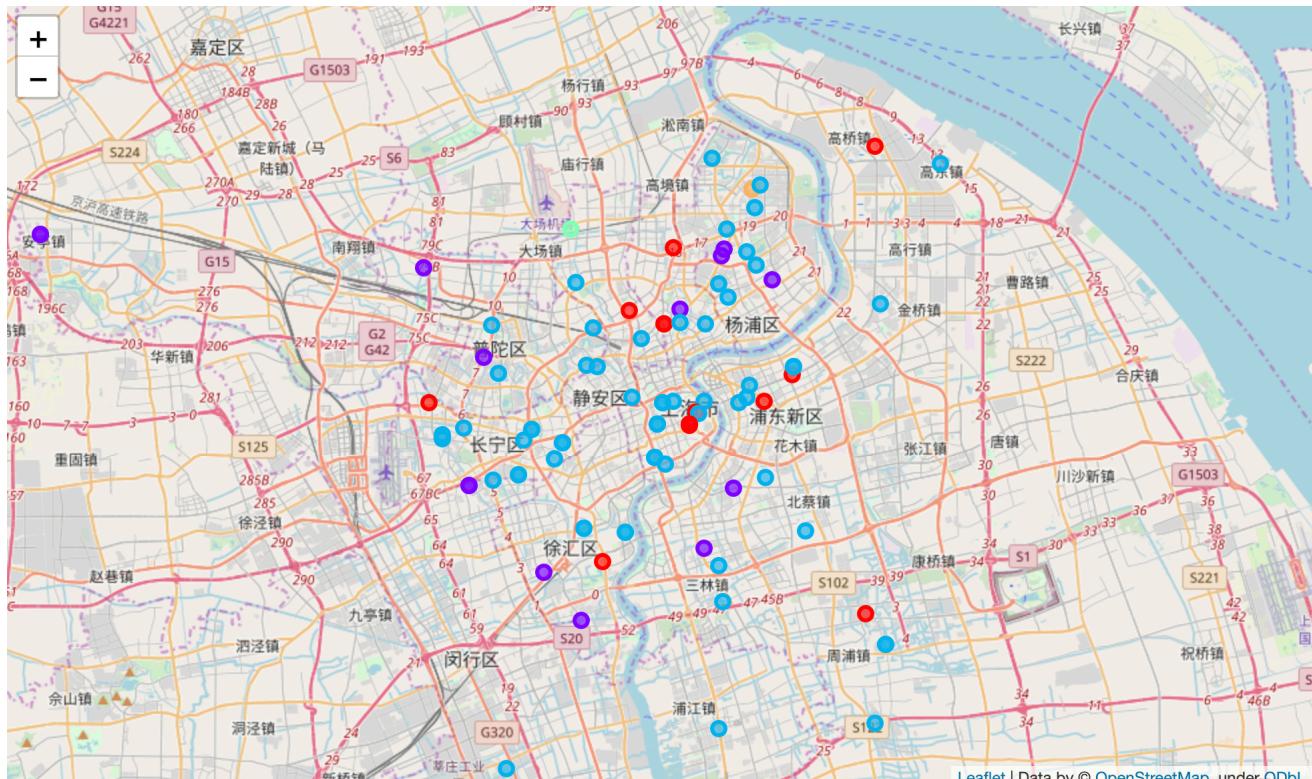
1. Research the places of Shanghai's milk tea shops. Find out the most places would appear nearby. Then choose some possible places in Tokyo.
2. Do the cluster using Shanghai's data and Tokyo's possible places to find out the places which has the same cluster label with the most cluster label of Shanghai's milk tea shop.

3.1 Clustering of Shanghai's data

I use K-means to cluster the Shanghai's data:

	Name	Address	Longitude	Latitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	奶茶会所	光明路与高东新路交叉口西北50米	121.617630	31.332172	2	Airport	Farmers Market	Food	Golf Course	German Restaurant
1	奶茶咖啡	人民路与梧桐路交叉口西北50米	121.496488	31.227198	2	Hotel	Clothing Store	Steakhouse	Art Gallery	Furniture / Home Store
2	一点奶茶	共和新路647号	121.463955	31.254808	2	Hotel	Supermarket	Bubble Tea Shop	Hotpot Restaurant	Metro Station
3	四季奶茶	三牌楼路87号	121.492453	31.222938	0	Hotel	Clothing Store	Convenience Store	Museum	Coffee Shop
5	tina奶茶	淞虹路683号	121.361598	31.212718	2	Fast Food Restaurant	Grocery Store	Ramen Restaurant	Sporting Goods Shop	Coffee Shop

And i show in the map with different colors:



It is obviously that most milk tea shops in the center are similar.

3.2 Find the most places in Shanghai

Because the venues are all different but similar, i split the venues to single words to find out the most words in all venues

```

words=[]
for i in range(0,81):
    for j in range(2,6):
        words.append(name_venues_sorted.loc[i][j].split(' '))

from collections import Counter
words=sum(words,[])
word_counts = Counter(words)
top_five = word_counts.most_common(5)
print(top_five)

```

Then i get the top 5 words in venues are

```
[('Restaurant', 119), ('Shop', 41), ('Chinese', 29), ('Coffee', 25), ('Food', 25)]
```

3.3 Choose possible places in Tokyo

As 3.2 shows, restaurant, shop, coffee and food is the most things near milk tea shops So, i choose some shopping malls in Tokyo, there are always many restaurants and coffee in the shopping malls, i can choose five places from these 12 shopping centres.

3.4 Clustering of both data

I merge shanghai's data and Tokyo's data to do the cluster to find out the most 5 similar places in possibles places.

This is the cluster labels of all data. It is obviously that milk tea shops' nearby environments are similar and also exist shopping centers have similar environments.

4. Result

Compare the cluster labels of milk tea shops in shanghai and possible places in Tokyo, i can choose

1. Tokyo Tower
 2. Ginza
 3. Daikanyama
 4. Shibuya Center Street
 5. Ameyoko Arcade

This result is suitable, because some milk tea shops have opened in these places actually.

5. Discussion

I recommend the stakeholder to open the milk tea shops in these places quickly, as the milk tea is more and more popular in Japan.

Also, when i do the cluster to shanghai's milk tea shops, i found the shops in city centers are all the same cluster label. It is interesting, and i think the reason is the city center always has anything, so the features would be found near those places.

6. Conclusion

I used the places of milk tea shops in shanghai to find out the feature of those places and then use the features to find 5 suitable places to open a milk tea shops in Tokyo.

The result is so great and i believe stakeholder would have a huge success.

