

Analysis of Venus from Districts of Shanghai and Beijing and A Tentative Clustering Scheme of Those Districts

Yuhui Cao
May 28, 2020

1. Data Collection: districts vs. Locations

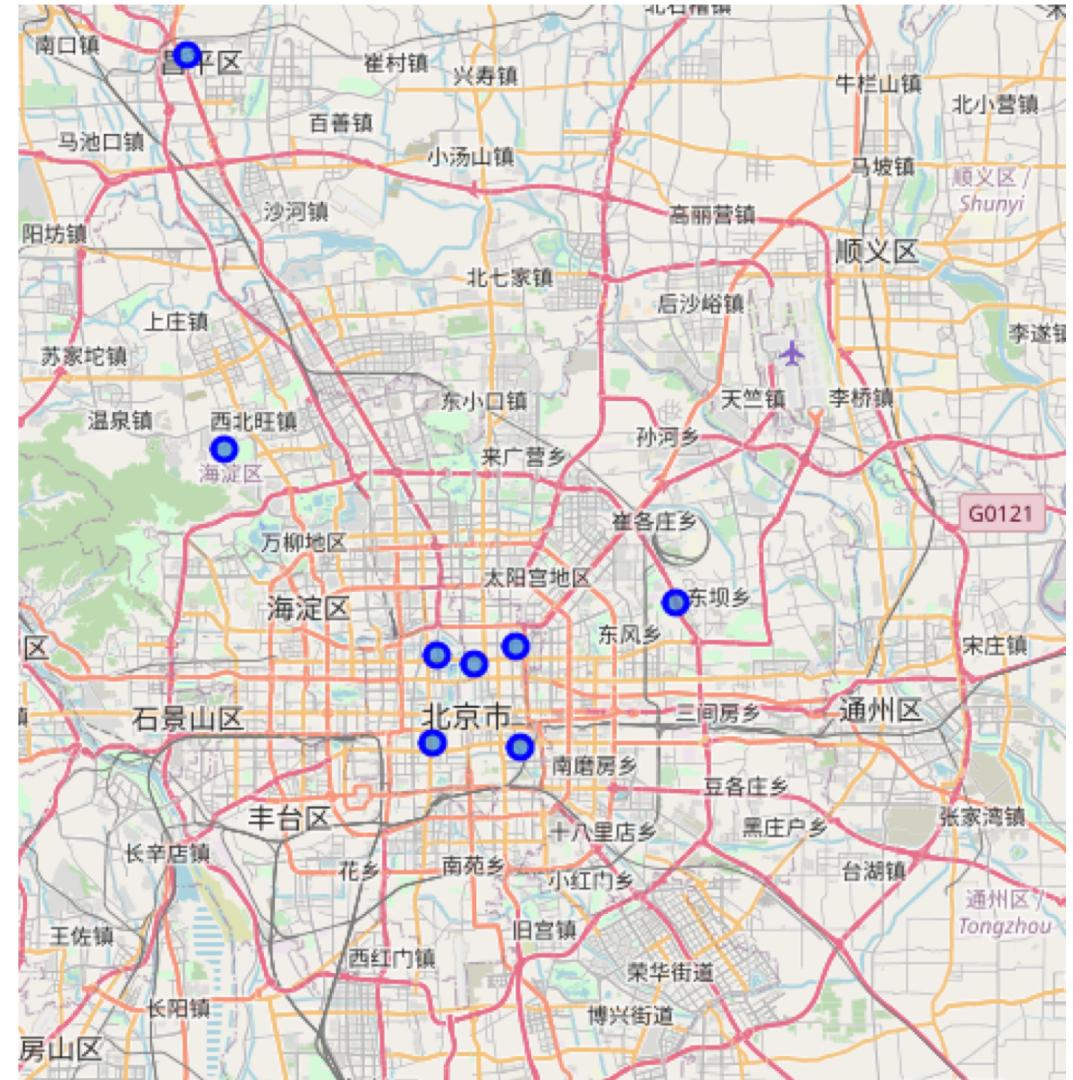
Original Data shows Postal Code, name of the city, Longitude and Latitude of 21 districts for the 2 cities.

```
1 frames = [Beijing_df, Shanghai_df]
2 two_city = pd.concat(frames)
3 print(two_city.shape)
4 two_city
```

(21, 5)

	Postal Code	City	District	Longitude	Latitude
0	110100	Beijing	Beijing	116.395645	39.929986
1	110101	Beijing	Dongcheng	116.421885	39.938574
2	110102	Beijing	Xicheng	116.373190	39.934280
3	110103	Beijing	Chongwen	116.424636	39.889292
4	110104	Beijing	Xuanwu	116.369352	39.891531
5	110105	Beijing	Chaoyang	116.521695	39.958953
6	110108	Beijing	Haidian	116.239678	40.033162
7	110114	Beijing	Changping	116.216456	40.221724
0	310100	Shanghai	Shanghai	121.487899	31.249162
1	310101	Shanghai	Huangpu	121.496072	31.227203
2	310104	Shanghai	Xuhui	121.446235	31.169152
3	310105	Shanghai	Changning	121.387616	31.213301
4	310106	Shanghai	Jingan	121.454756	31.235381
5	310107	Shanghai	Putuo	121.398443	31.263743
6	310108	Shanghai	Zhabei	121.457769	31.288044
7	310109	Shanghai	Hongkou	121.491919	31.282497
8	310110	Shanghai	Yangpu	121.535717	31.304510
9	310112	Shanghai	Minhang	121.425024	31.093538
10	310114	Shanghai	Jiading	121.251014	31.364338
11	310117	Shanghai	Songjiang	121.226791	31.021245
12	310119	Shanghai	Nanhui	121.769956	31.052602

2. Data Visualization: Location of Districts



Visualization of each district on map by using Folium library.

3. Venues Information from Foursquare

185 venues were obtained for the 2 cities together (from Foursquare).

```
1 print(two_city_venues.shape)
2 two_city_venues.head()
```

(185, 7)

	District	District Latitude	District Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Beijing	39.929986	116.395645	TRB Hutong	39.926528	116.397204	French Restaurant
1	Beijing	39.929986	116.395645	The Georg by Georg Jensen	39.933440	116.392740	Scandinavian Restaurant
2	Beijing	39.929986	116.395645	City Walls Courtyard House Beijing	39.928131	116.392442	Hostel
3	Beijing	39.929986	116.395645	Peking Hostel (北平国际青年旅舍)	39.934350	116.396886	Hostel
4	Beijing	39.929986	116.395645	Nanluogu Alley (南锣鼓巷)	39.932498	116.396925	Pedestrian Plaza

4. Unique Categories

- 80 unique venues categories (“Venue Category”) were obtained for the 2 cities, together.

District	District Latitude	District Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Beijing	17	17	17	17	17	17
Changning	20	20	20	20	20	20
Changping	5	5	5	5	5	5
Chaoyang	2	2	2	2	2	2
Chongwen	5	5	5	5	5	5
Dongcheng	24	24	24	24	24	24
Haidian	1	1	1	1	1	1
Hongkou	6	6	6	6	6	6
Huangpu	14	14	14	14	14	14
Jiading	4	4	4	4	4	4
Jingan	26	26	26	26	26	26
Minhang	1	1	1	1	1	1
Nanhui	1	1	1	1	1	1
Putuo	4	4	4	4	4	4
Shanghai	23	23	23	23	23	23
Songjiang	2	2	2	2	2	2
Xicheng	9	9	9	9	9	9
Xuanwu	7	7	7	7	7	7
Xuhui	6	6	6	6	6	6
Yangpu	4	4	4	4	4	4
Zhabei	4	4	4	4	4	4

Let's find out how many unique categories can be curated from all the returned venues

```
1 print('There are {} uniques categories.'.format(len(two_city_venues['Venue Category'].unique())))
```

There are 80 uniques categories.

5. Onehot: Venue Category for Each District

```
6 | two_city_onehot.head()
```

The dimensionality of two_city_onehot dataset is (185, 81)

	District	Art Gallery	Art Museum	Asian Restaurant	Athletics & Sports	Auto Garage	Bakery	Bar	Beer Bar	Beijing Restaurant	...	Sporting Goods Shop	Steakhouse	Supermarket	Sushi Restaurant	Tea Room
0	Beijing	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
1	Beijing	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
2	Beijing	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
3	Beijing	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
4	Beijing	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0

5 rows x 81 columns

Onehot encoding for “Venue Category”.

6. Averaged “Venue Category” for each district.

```
1 two_city_grouped = two_city_onehot.groupby('District').mean().reset_index()
2 two_city_grouped
```

District	Art Gallery	Art Museum	Asian Restaurant	Athletics & Sports	Auto Garage	Bakery	Bar	Beer Bar	Beijing Restaurant	... Sporting Goods Shop	Steakhouse	Supermarket	S Restau	
0 Beijing	0.000000	0.00	0.00	0.00	0.0	0.000000	0.058824	0.000000	0.058824	...	0.000000	0.000000	0.0	0.000
1 Changning	0.000000	0.05	0.00	0.05	0.0	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.0	0.000
2 Changping	0.000000	0.00	0.20	0.00	0.0	0.200000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.0	0.000
3 Chaoyang	0.000000	0.00	0.00	0.00	0.5	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.0	0.000
4 Chongwen	0.000000	0.00	0.00	0.00	0.0	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.2	0.000
5 Dongcheng	0.000000	0.00	0.00	0.00	0.0	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.0	0.000
6 Haidian	0.000000	0.00	0.00	0.00	0.0	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.0	0.000
7 Hongkou	0.000000	0.00	0.00	0.00	0.0	0.166667	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.0	0.000
8 Huangpu	0.071429	0.00	0.00	0.00	0.0	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.071429	0.0	0.000
9 Jiading	0.000000	0.00	0.00	0.00	0.0	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.0	0.000
10 Jingan	0.000000	0.00	0.00	0.00	0.0	0.000000	0.000000	0.000000	0.000000	...	0.038462	0.000000	0.0	0.000
11 Minhang	0.000000	0.00	0.00	0.00	0.0	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.0	0.000
12 Nanhui	0.000000	0.00	0.00	0.00	0.0	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.0	0.000
13 Putuo	0.000000	0.00	0.00	0.00	0.0	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.0	0.000
14 Shanghai	0.000000	0.00	0.00	0.00	0.0	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.0	0.000
15 Songjiang	0.000000	0.00	0.00	0.00	0.0	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.0	0.000
16 Xicheng	0.000000	0.00	0.00	0.00	0.0	0.000000	0.000000	0.111111	0.111111	...	0.000000	0.000000	0.0	0.000
17 Xuanwu	0.000000	0.00	0.00	0.00	0.0	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.0	0.000
18 Xuhui	0.000000	0.00	0.00	0.00	0.0	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.0	0.168
19 Yangpu	0.000000	0.00	0.00	0.00	0.0	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.0	0.000
20 Zhabei	0.000000	0.00	0.25	0.00	0.0	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.0	0.000

21 rows x 81 columns

Let's confirm the new size

```
1 two_city_grouped.shape
```

(21, 81)

7. Top 10 most common venues for each district.

```
22 print(District_venues_sorted.shape)
23 District_venues_sorted.head()
```

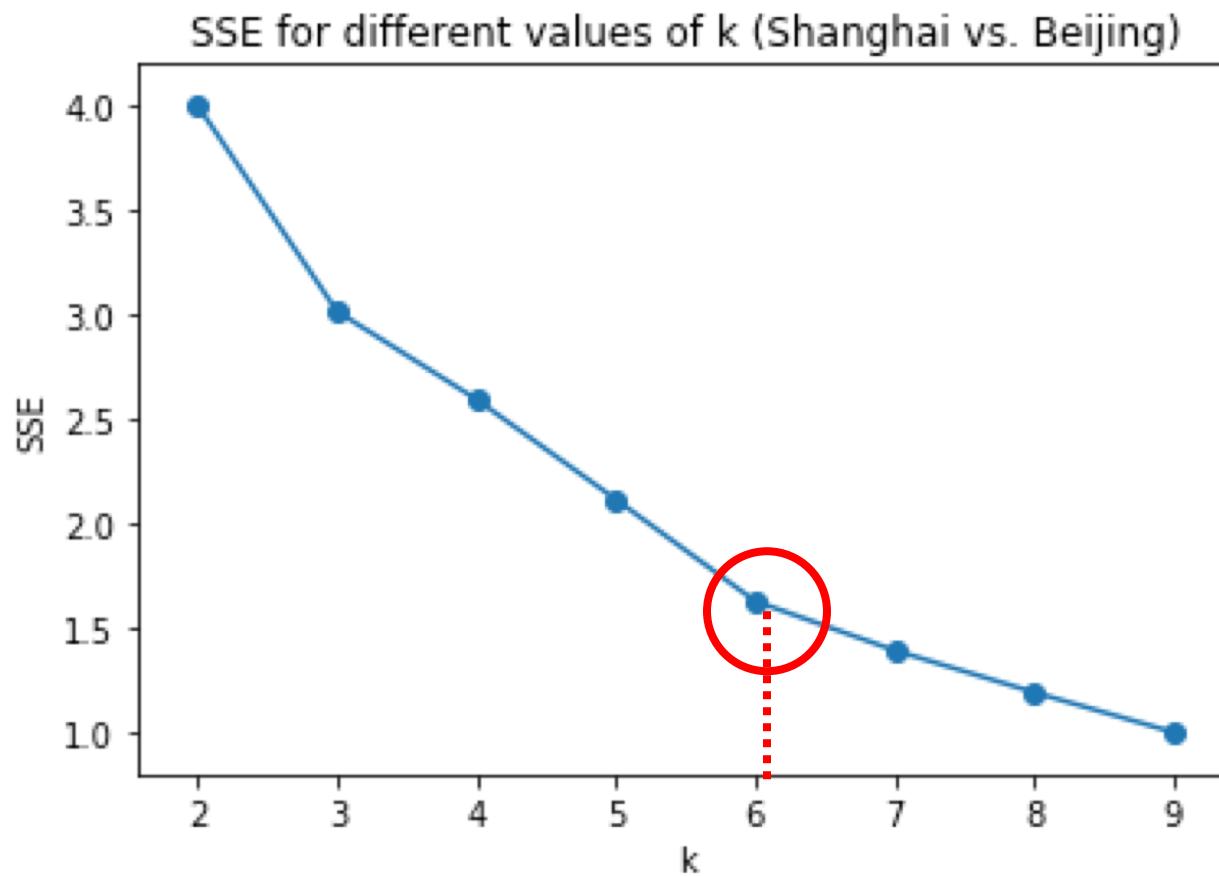
(21, 11)

This data is going to be used for final clustering analysis ----
districts from 2 cities combined.

	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	Beijing	Hostel	Hotel	Hotpot Restaurant	Metro Station	French Restaurant	Chinese Restaurant	Restaurant	Café	Scandinavian Restaurant	Bus Station
1	Changning	Chinese Restaurant	Japanese Restaurant	Hotel	Gym Pool	Athletics & Sports	Tea Room	Movie Theater	Art Museum	Ramen Restaurant	Dongbei Restaurant
2	Changping	Shopping Mall	Asian Restaurant	Chinese Restaurant	Bakery	Korean Restaurant	Fast Food Restaurant	Dessert Shop	Dim Sum Restaurant	Dongbei Restaurant	Dumpling Restaurant
3	Chaoyang	Park	Auto Garage	Wedding Hall	Farmers Market	Department Store	Dessert Shop	Dim Sum Restaurant	Dongbei Restaurant	Dumpling Restaurant	Electronics Store
4	Chongwen	Metro Station	Supermarket	Chinese Restaurant	Fast Food Restaurant	Convenience Store	Gym	Grocery Store	Gym / Fitness Center	Furniture / Home Store	French Restaurant

Above is each district along with the top 10 most common venues for, both Shanghai and Beijing.

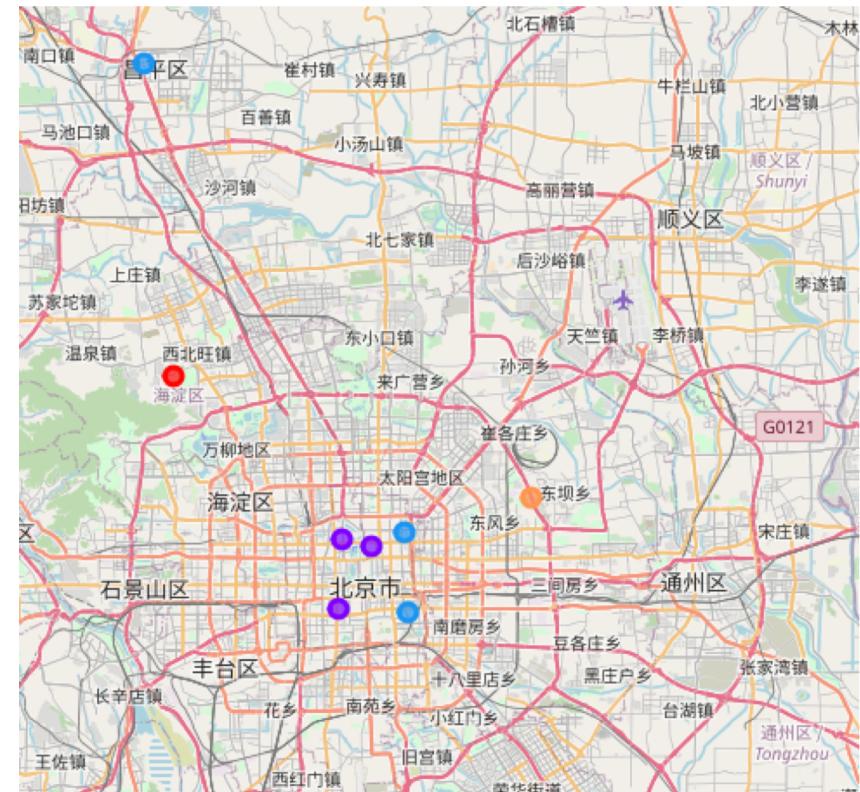
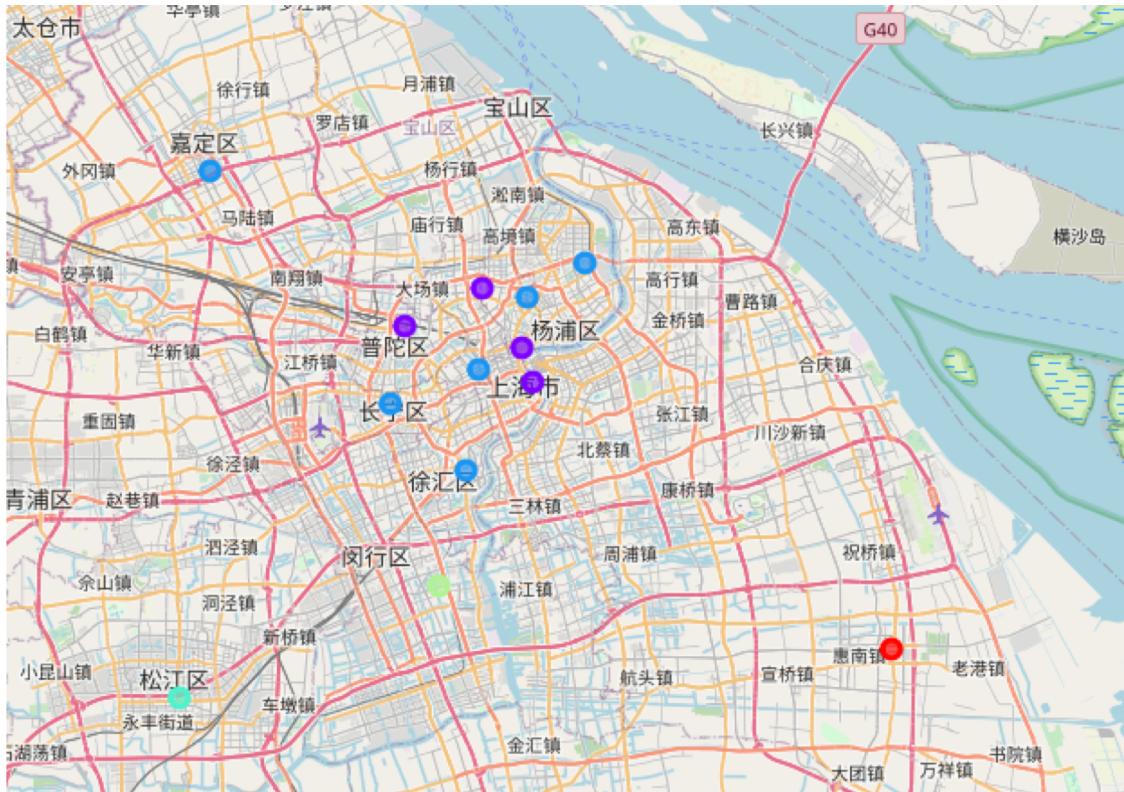
8. “Elbow” Analysis



Finally I chose K = 6 for the final clustering of the 2 cities together.

9. Clustering Visualization

- Clustering Visualization of each district for the 2 cities, together (color-coded). There are 6 clusters (colors).



9. Examination of Clusters

Name were given to each cluster, according to the characteristics of them.

Table 3. Final clustering of Shanghai and Beijing together. There are 6 clusters and corresponding names.

Cluster 1 “College Town”		Cluster 2 “Traditional Cosmopolitan Area”		Cluster 3 “Newly Developed Cosmopolitan Area”	
City	District	City	District	City	District
6	Beijing	Haidian	0	Beijing	Beijing
12	Shanghai	Nanhui	2	Beijing	Xicheng
Cluster 4 “Minhang High-tech Development Zone”		Cluster 5 “Densely Populated Area”		Cluster 6 “Old Shanghai”	
City	District	City	District	City	District
9	Shanghai	Minhang	5	Shanghai	Putuo
1	Shanghai	Huangpu	6	Shanghai	Zhabei
0	Shanghai	Shanghai	11	Shanghai	Songjiang
7	Beijing	Changping	10	Shanghai	Jiading
3	Beijing	Chongwen	8	Shanghai	Yangpu
2	Shanghai	Xuhui	7	Shanghai	Hongkou
3	Shanghai	Changning	4	Shanghai	Jingan