Searching the best place for starting a new coffee shop in Seoul, South Korea

Jaecheol Hwang

1. Introduction

1.1 Background

As of July 2019, in South Korea, there were 71,000 Coffee shops. And the number of Coffee shops are increasing even now. However, many of them were closed before 3 years.

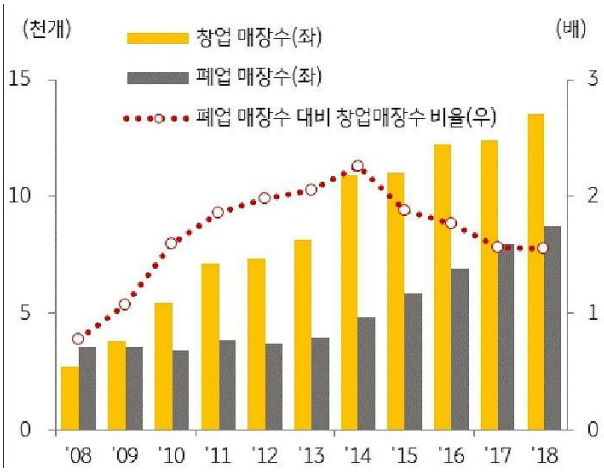

Fig. 1. Number of start-up and shut-down coffee shop in South Korea. [link]

It means that many people tried to open coffee shop but a few of them succeed to keep their shop. Then does we just give up to open our own coffee shop? My answer is NO There are report, column, or wikipedia page about coffee culture of South Korea. We don't need to give up now, however we should know that points of the success.

I want to find the points using geographic data and Foursquare location data, and figure out which area is the most potential for starting coffee shop.

1.2 Problem Statement

- In Seoul, South Korea, where is the best place for starting a new coffee shop?
- Seoul has 25 Districts/Counties("Gu", in korean) and Districts are separated to several Cities/Towns("Dong", in korean).

Fig. 2. Map of Seoul with districts. [link]

- So, our goal is found best "Dong" for starting a new coffee shop using Foursquare location data. We may use location data for Seoul additionally.

2. Methodology

2.1 Exploratory data analysis

I will use 2 datasets. First is the geographic data of Seoul. It contains District and Dong names and its latitude, longitude information. Second one is Foursquare location data, with Foursquare API. We can get information of venues in Seoul.

Let's check geographic data first. It has names of District, Dong and their lat, lng information. Also I have boundary json file of Seoul, districts, Dongs. At Fig. 3., blue circles mean the locations of each Districts of Seoul, and Fig. 4., blue dots mean the locations of each Dongs.

	district	district_eng	lat	lng
0	종로구	Jongno-gu	37.592128	126.979420
1	중구	Jung-gu	37.557335	126.997985
2	용산구	Yongsan-gu	37.528582	126.981987
3	성동구	Seongdong-gu	37.548240	127.043114
4	광진구	Gwangjin-gu	37.543059	127.088351
	dong	**dong_eng**	**lat**	**lng**
0	가락동	Garak-dong	37.494280	127.117631
1	가리봉동	Garibong-dong	37.483017	126.887363
2	가산동	Gasan-dong	37.475469	126.884239
3	가양동	Gayang-dong	37.567984	126.844436
4	가회동	Gahoe-dong	37.582806	126.984865

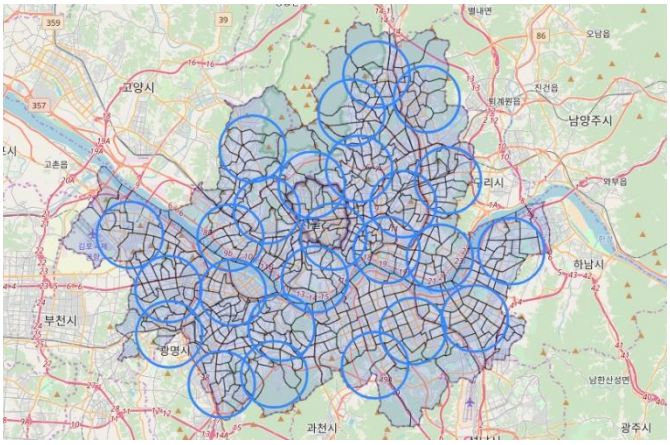

Fig. 3. Map of Seoul using boundary data(25 Districts).

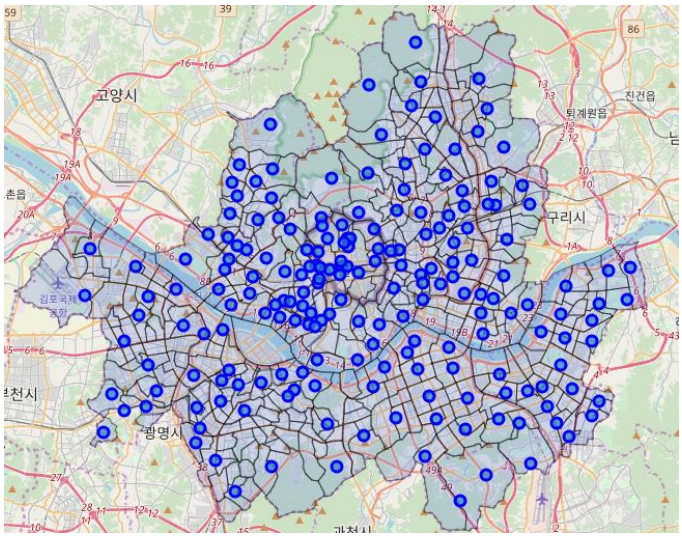

Fig. 4. Map of Seoul using boundary data(198 Dongs).

District Name	# of Dongs
Jongno-gu	18
Mapo-gu	13
…	…
Dobong-gu	3
Geumcheon-gu	3

Of all Districts, Jongno-gu has 18 Dongs, but Geumcheon-gu has only 3 Dongs.

Using FourSquare API, I can get the information of 4,707 venues in Seoul.

Category Name	# of venues
Korean Restaurant	643
Cafe	285
Coffee Shop	270
Chinese Restaurant	217
Japanese Restaurant	180
…	…

There are Coffee shop and Coffee-shop-like categories, like Café, Bakery, Dessert shop, Donut shop, Bubble Tea shop. So, I can separate Coffee shop and Not Coffee shop.

As mentioned above, Jongno-gu has 18 Dongs, so Jongno-gu can be a good example for showing my methodology. At Jongno-gu, I can get 471 venues from Foursquare API, and there are 56 Korean Restaurants, 36 Cafés, 29 Coffee shops, etc.

```
1 df_jongno.categories.value_counts().to_frame().head(10)
```

	categories
Korean Restaurant	56
Café	36
Coffee Shop	29
Historic Site	23
Italian Restaurant	14
Art Museum	13
Chinese Restaurant	12
Bakery	12
BBQ Joint	12
Art Gallery	12

Let’s plot Coffee shop and Not Coffee shop venues.

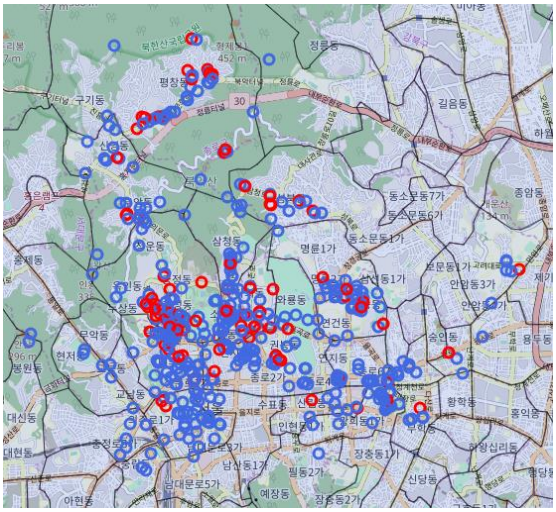

Fig. 5. Map of Jongno-gu with venues(471 venues).

Red circles show Coffee shop venues and blue circles show Not Coffee shop venues. We can get insight that a place with a lot of red circles are fiercely competitive place and a place with a lot of blue circles are less competitive place for coffee shop. None of circles means that there are none of venues, and not suitable for coffee shop.

We also have Dong information of Jongno-gu.

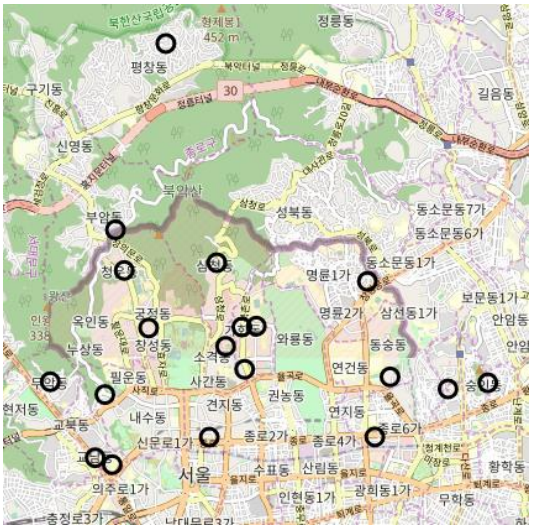

Fig. 6. Map of Jongno-gu with Dong(18 Dongs).

Let's make score for each Dong about starting Coffee shop. There are two standards for scoring. First is that the Dong, far from existing Coffee shop, get high score. And second, the Dong, close from existing Not Coffee shop, get high score.

So I calculate distance of every Dongs between center of Dongs and each venues using latitude and longitude data. After that I counted the number of Coffee shop and Not Coffee shop with in 1KM from center of each Dongs. Scoring is simply calculated by number of Not Coffee shop subtract number of Coffee shop.

	dong	dong_eng	lat	lng	district	cafe_count	non_cafe_count	score
158	종로1가	Jongno 1(il)-ga	37.570698	126.980334	종로구	19	108	89
189	화동	Hwa-dong	37.580686	126.982610	종로구	24	94	70
153	재동	Jae-dong	37.578159	126.985253	종로구	24	82	58
193	효자동	Hyoja-dong	37.582704	126.971898	종로구	28	86	58
145	이화동	Ihwa-dong	37.577325	127.005471	종로구	7	64	57
159	종로5가	Jongno 5(o)-ga	37.570730	127.003350	종로구	6	61	55
176	평동	Pyeong-dong	37.567673	126.966835	종로구	8	63	55
10	계동	Gye-dong	37.582904	126.986877	종로구	19	73	54
75	사직동	Sajik-dong	37.575384	126.965826	종로구	24	77	53
4	가회동	Gahoe-dong	37.582806	126.984865	종로구	19	72	53
17	교남동	Gyonam-dong	37.568435	126.964474	종로구	8	54	46
169	창신동	Changsin-dong	37.576003	127.013525	종로구	5	34	29
79	삼청동	Samcheong-dong	37.589878	126.981380	종로구	8	33	25
185	혜화동	Hyehwa-dong	37.587822	127.002323	종로구	6	31	25
174	청운동	Cheongun-dong	37.589029	126.968516	종로구	12	28	16
69	부암동	Buam-dong	37.593496	126.967285	종로구	4	19	15
57	무악동	Muak-dong	37.576791	126.958227	종로구	4	16	12
177	평창동	Pyeongchang-dong	37.613968	126.974302	종로구	8	18	10
101	숭인동	Sungin-dong	37.576675	127.019065	종로구	2	10	8

Fig. 7. Score table of Jongno-gu.

As the result, "Jongno 1(il)-ga" get the highest score, and "Sungin-dong" get the lowest score. Fig. 7. shows venues of Jongno-gu. Blue circle represents "Jongno 1(il)-ga", the best place for Coffee shop in Jongno-gu, and grey circle represents "Hwa-dong", the 2nd best place, and red circle represents "Sungin-dong", the worst place.

“Jongno 1(il)-ga” and “Sungin-dong” has almost same Coffee shop / Not Coffee shop rate. However “Jongno 1(il)-ga” has lots of venues(127) and “Sungin-dong” has only 12 venues. So I can make difference between these to Dongs.

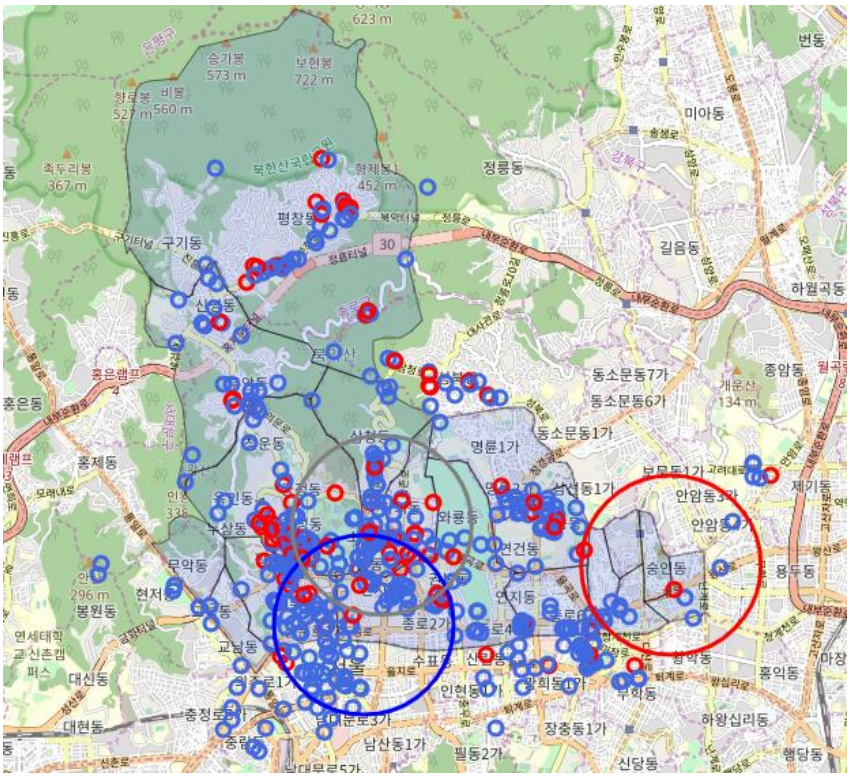

Fig. 7. Map of Jongno-gu with scoring results.

3. Results

We checked the result of Jongno-gu, and it is quite acceptable. So let’s move on to the whole of Seoul.

```
df_result.sort_values('score', ascending=False)
```

	dong	dong_eng	lat	lng	district	cafe_count	non_cafe_count	score
158	종로1가	Jongno 1(il)-ga	37.570698	126.980334	종로구	19	108	89
144	이태원동	Itaewon-dong	37.538862	126.992278	용산구	23	107	84
86	서교동	Seogyo-dong	37.553650	126.920080	마포구	42	120	78
95	소공동	Sogong-dong	37.563898	126.979989	중구	13	91	78
33	다동	Da-dong	37.567883	126.981194	중구	15	90	75
...	...	...	...	...	...	...	...	...
96	송정동	Songjeong-dong	37.554104	127.062161	성동구	2	2	0
103	신길동	Singil-dong	37.505826	126.912651	영등포구	1	1	0
102	시흥동	Siheung-dong	37.449703	126.910697	금천구	1	1	0
139	월계동	Wolgye-dong	37.629556	127.056253	노원구	1	1	0
167	진관동	Jingwan-dong	37.642252	126.934085	은평구	6	5	-1

198 rows × 8 columns

Fig. 8. Score table of Seoul.

“Jongno 1(il)-ga”, “Itaewon-dong”, “Seogyo-dong”, “Sogong-dong”, “Da-dong” got best 5 and “Songjeong-dong”, “Singil-dong”, “Siheung-dong”, “Wolgye-dong”, “Jingwan-dong” got worst 5. Let’s visualize the result.

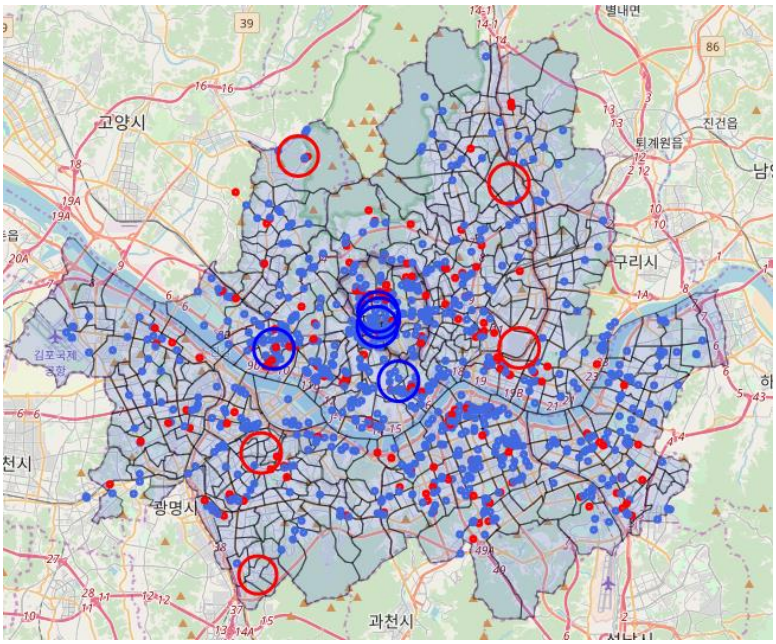

Fig. 9. Map of Seoul with scoring result.

Fig. 9. shows the result. 5 blue circles show 5 best places for Coffee shop, and 5 red circles show 5 worst places for Coffee shop in Seoul. As you can see, blue circles locate center of Seoul and contain a lot of venues, on the other hand, red circles contain a few of venues.

4. Discussion

Although the results have been obtained, it should be noted that there is a problem with the dataset. In particular, there is a lack of venue datasets. Although I brought venue data using Foursquare API, many Korean venues are not registered because this service is not used much in Korea. And because famous places have a lot of venue data, and few places don't, I think this imbalance has had a huge impact on the results. According to recent Korean statistics, the number of coffee shops in Seoul is close to 20,000. Therefore, it is necessary to interpret the results by referring to these points.

5. Conclusion

In this project, we used the venue data provided by Foursquare to find the best place to succeed as a coffee shop in South Korea. As a result, the place that people generally visited was a lot of venue data, and these places received the highest score. On the contrary, places where people visit less often get lower score. However, since Korea is a country that does not use much Foursquare, a lack of venue data can be problem. Thus, data gathering is needed to reinforce the result.