Business Problem : Open Cafe Business in Seoul

Mo Sung Hoon

August 15, 2020

1. Introduction

Coffee is important for life. Cafe becomes a third place for everyone who wants to stay withouth any housework.

Many people comes to cafe for chatting with acquintances, reading books and studying, or just enjoying taste of coffee.

How about opening cafe in Seoul?

Source : Neilson Korea, 2019. 7. 14.

There are so many Cafes are in Seoul city and Koreans drinks 1.3 cup of coffee each day, men drink 9.56 cup of coffee a week, women drink 9.04 cup of coffee a week, according to the survey from Neilson Korea.

So, opening cafe business in Seoul seems like jumping into the red ocean.

Nevertheless, There are so many opportunities for opening cafe in Seoul.

U.S.A. is the best coffee market all over the world, which records $26.1 billion at 2018. In case of Korea, Sales at coffee shops in Korea have grown from $600 million in 2007 to $4.3 billion (4.87 trillion won) in 2018, this record, Korea ranked 3rd coffee shop market all over the world.

Source : Jungang-Ilbo(Korea Press), 2019. 3. 18.

In this analysis, we'll gather information for opening cafe business.

We will search some places of Seoul, and find proper place for opening cafe business.

1.1 Data for solving this business problem

In this analysis, I'll find proper place for opening cafe business in Seoul.

So, I have to know information of neighborhoods in Seoul.

I will use location data of Seoul from Public Seoul Data hub(https://data.seoul.go.kr/)

and foot traffic data of Seoul from SK Data Hub(https://www.bigdatahub.co.kr/) with SK Data Hub API.

With location data and foursquare api, I will find some cafes in Seoul and make dataframe for it.

with these data, I'll find proper neighborhood.

1.2 Methodology for Data Analysis

First, I'll do Exploratory Data Analysis and check status of the Seoul.

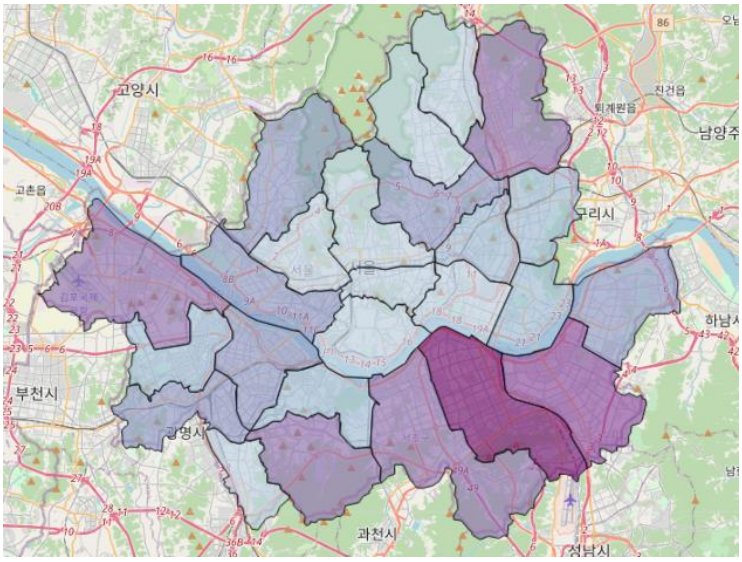
Second, With data, I'll make scores for neighborhoods.

With scores, I'll select a proper neighborhood for opening cafe business and explain result of analysis.

2. Data Analysis

In this part, Exploratory Data Analysis will be processed. First, we do Exploratory Data Analysis with foot traffic data.

2.1 Exploratory Data Analysis with choropleth maps



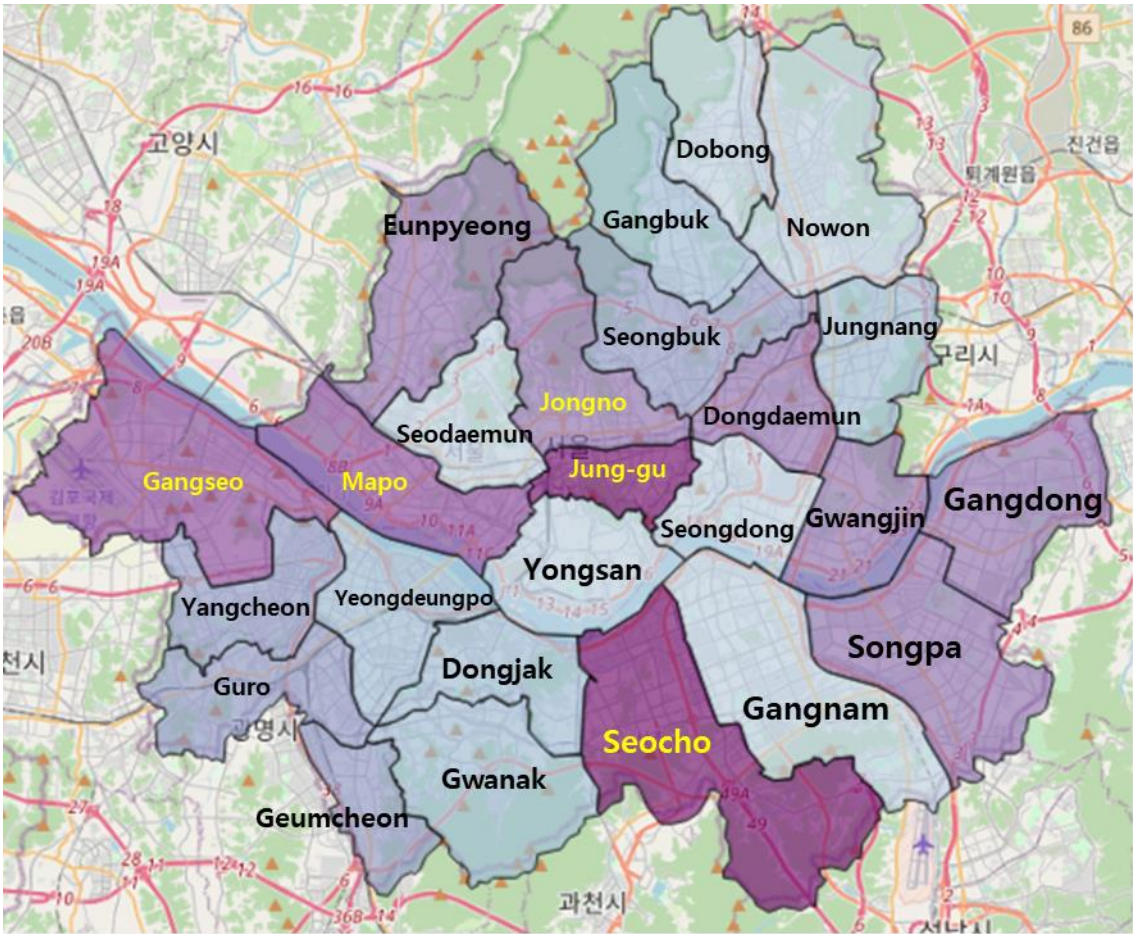
	군구	유동인구수
0	강남구	47965.807292
17	송파구	43007.181839
3	강서구	36449.266304
14	서초구	33899.488225
8	노원구	33040.048687
4	관악구	32827.153533
19	영등포구	32255.355525
6	구로구	29195.157382
16	성북구	28128.134058
12	마포구	28067.927989

- Gangnam, Songpa, Gangseo, Seocho, Nowon are top 5 foot trapic neighborhoods.

- The best foot traffic place is Gangnam.

- the fact that more people walks around means there are many chances of business.

	index	numb_cafes		index	numb_cafes
0	서초구	20	15	관악구	7
1	중구	19	16	강북구	7
2	강서구	17	17	중랑구	7
3	마포구	16	18	동작구	6
4	등대문구	14	19	서대문구	4
5	종로구	13	20	노원구	3
6	광진구	12	21	도봉구	3
7	강동구	12	22	용산구	3
8	은평구	11	23	성동구	3
9	송파구	11	24	강남구	2



* colors on letter just help to recognize district name

- Secho, Jung-gu, Gangseo, Mapo, Dongdaemun districts are top 5 districts that have many cafes.

- Gangnam, Seongdong, Yongsan, Dobong, Nowon districts are last 5 districts that have many cafes.

※ If you want to check names of district map, look up above map

	연령대(10세단위)	유동인구수
1	30	33351.856793
2	40	33152.507880
3	50	31098.679076
0	20	30194.238587
4	60	19925.842120
5	70	15076.064946

- Age group is the critical index for consumption, business analysis.

- 30~40 age groups are major groups for cafe business.

- In fact, 20~49 age groups are critical for consumption according to many articles, and it is a common sense of Korea.

- So, we have to check how many 30~40 age group people

	군구	성별	연령대(10세단위)	유동인구수
7	강남구	여성	30	71501.888587
1	강남구	남성	30	70600.176630
2	강남구	남성	40	65379.320652
6	강남구	여성	20	62473.980978
8	강남구	여성	40	61943.505435
211	송파구	여성	30	56599.334239
212	송파구	여성	40	56364.402174
205	송파구	남성	30	54407.445652
206	송파구	남성	40	53224.836957
213	송파구	여성	50	53135.883152

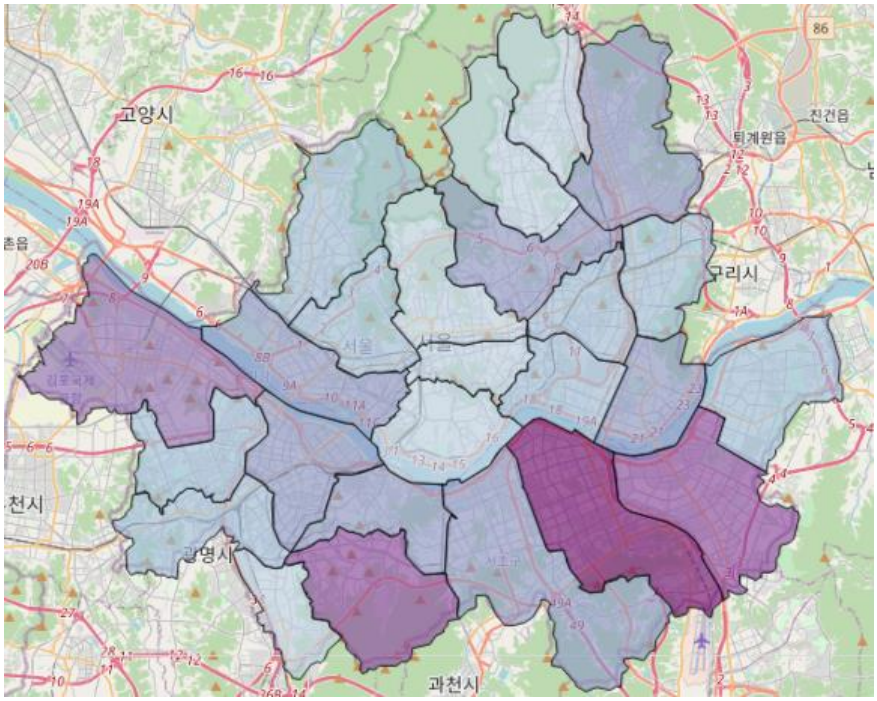
- When it comes to sex, female who's age is in 30 age group are major in Gangnam, where the number of foot traffic is best in Seoul.

- So, in Gangnam, 30-Female group is major target for cafe business.

- After that, we check choropleth footprint map of Seoul by age group.

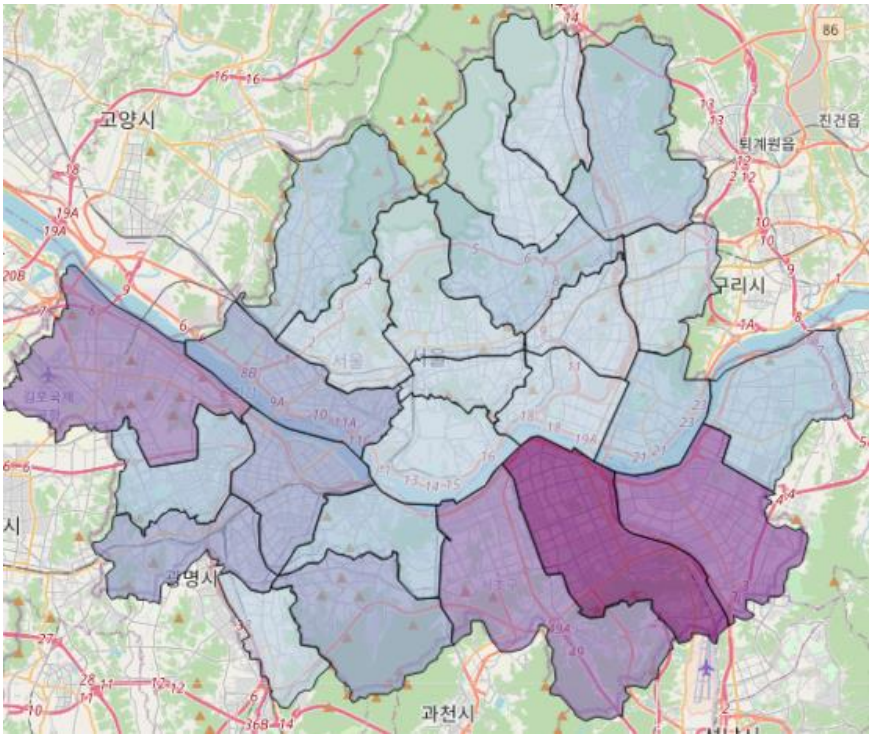
- Footprint map of Seoul (Age group-20)

	군구	유동인구수
0	강남구	112556.562500
17	송파구	90702.445652
4	관악구	89182.961957
3	강서구	75278.885870
12	마포구	73150.692935
14	서초구	72531.467391
8	노원구	72145.013587
16	성북구	68669.782609
11	동작구	66576.467391
19	영등포구	66189.918478



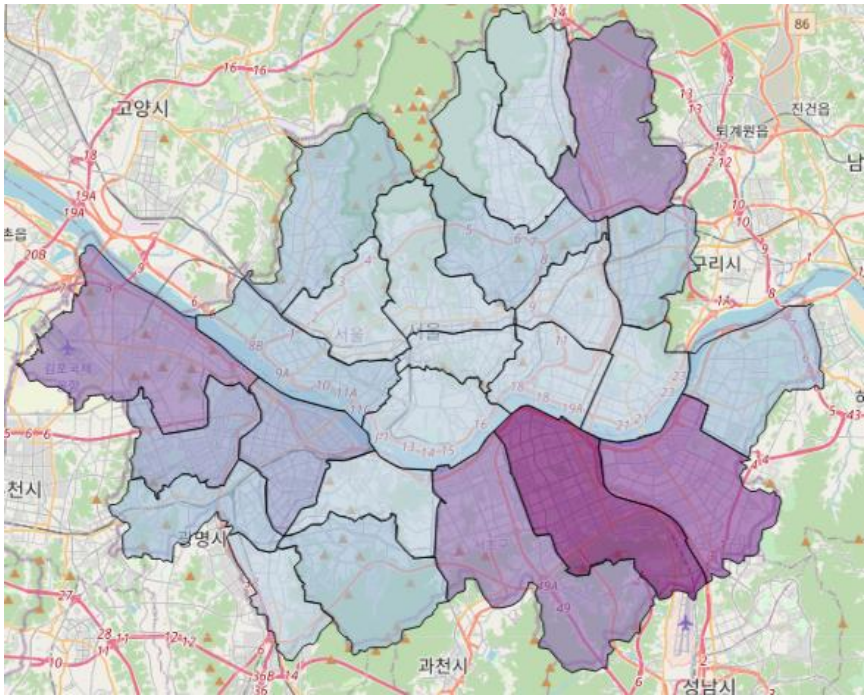
- Footprint map of Seoul (Age group-30)

	군구	유동인구수
0	강남구	142102.065217
17	송파구	111006.779891
3	강서구	95654.524457
14	서초구	93437.024457
19	영등포구	85831.195652
4	관악구	83651.236413
12	마포구	77332.500000
6	구로구	74300.000000
8	노원구	68700.067935
1	강동구	64760.095109



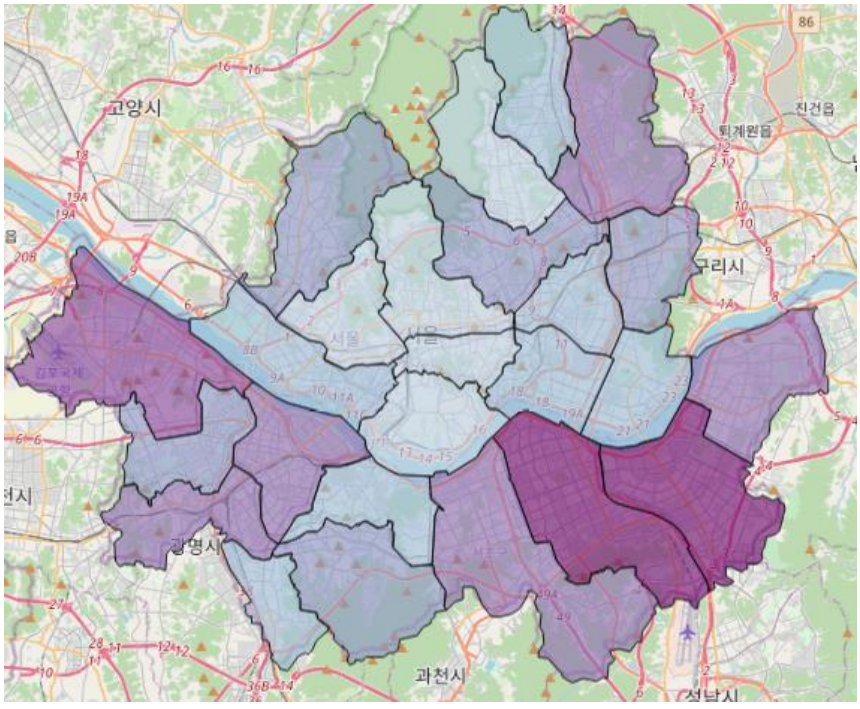
- Footprint map of Seoul (Age group-40)

	군구	유동인구수
0	강남구	127322.826087
17	송파구	109589.239130
14	서초구	91633.940217
3	강서구	87489.782609
8	노원구	87274.809783
19	영등포구	80060.502717
18	양천구	75296.698370
6	구로구	70501.032609
21	은평구	67973.260870
4	관악구	67948.057065



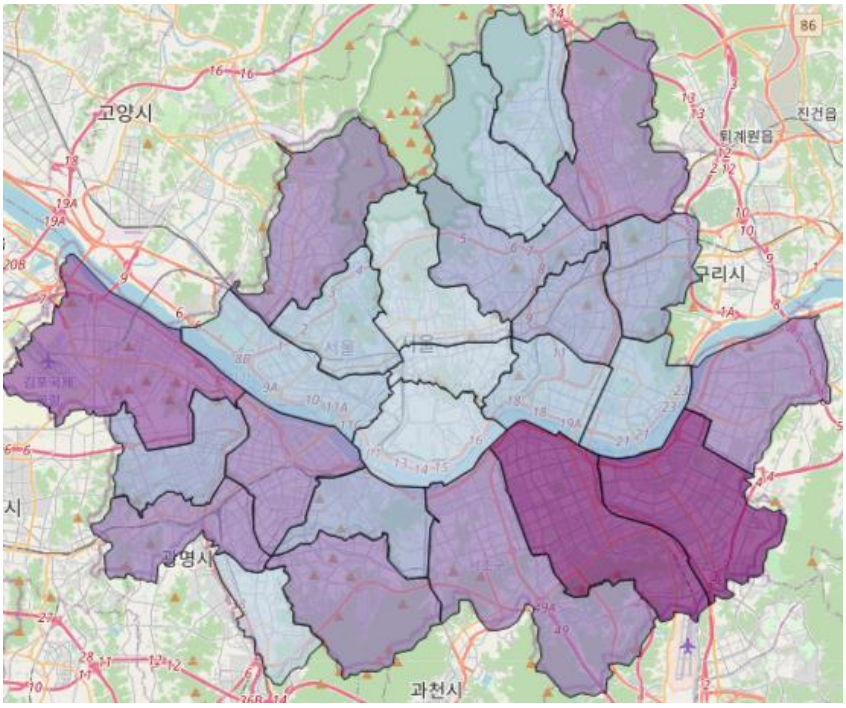
- Footprint map of Seoul (Age group-50)

	군구	유동인구수
17	송파구	100366.915761
0	강남구	97456.902174
3	강서구	85380.475543
8	노원구	78779.633152
19	영등포구	76745.951087
14	서초구	73382.866848
6	구로구	70518.355978
1	강동구	69502.894022
18	양천구	68923.179348
4	관악구	67463.586957



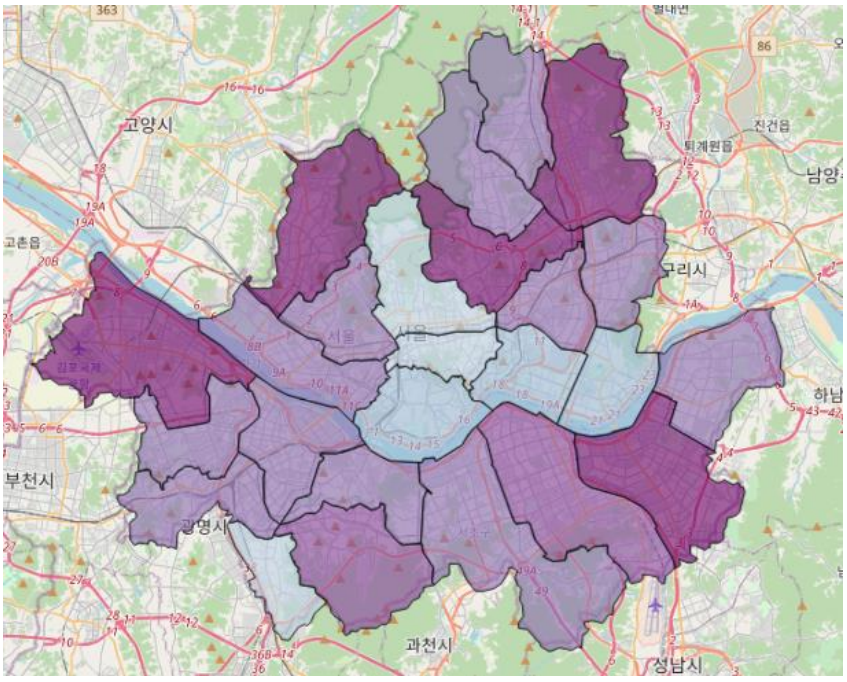
- Footprint map of Seoul (Age group-60)

	군구	유동인구수
17	송파구	64418.505435
0	강남구	58996.942935
3	강서구	54969.144022
4	관악구	48865.638587
8	노원구	48056.182065
6	구로구	46799.864130
19	영등포구	46497.309783
14	서초구	45871.494565
21	은평구	45404.524457
1	강동구	45091.453804



- Footprint map of Seoul (Age group-70)

	군구	유동인구수
8	노원구	41524.877717
17	송파구	40002.296196
21	은평구	39906.467391
3	강서구	38618.383152
16	성북구	37340.230978
0	강남구	37154.388587
4	관악구	36814.361413
10	동대문구	32088.301630
19	영등포구	31739.388587
6	구로구	31403.614130



- Many people pass by Gangnam and Songpa districts.

- Many younger people (from 20 to 40) pass by Gangnam absolutely than elder people.

- 40, 50 age groups are more distributed on Yangcheon district than other age groups.

- all age groups except 30 contains Nowon district on top 10 district of footprints, especially, 70 age group people pass by Nowon district most.

2.2 Scoring districts of Seoul

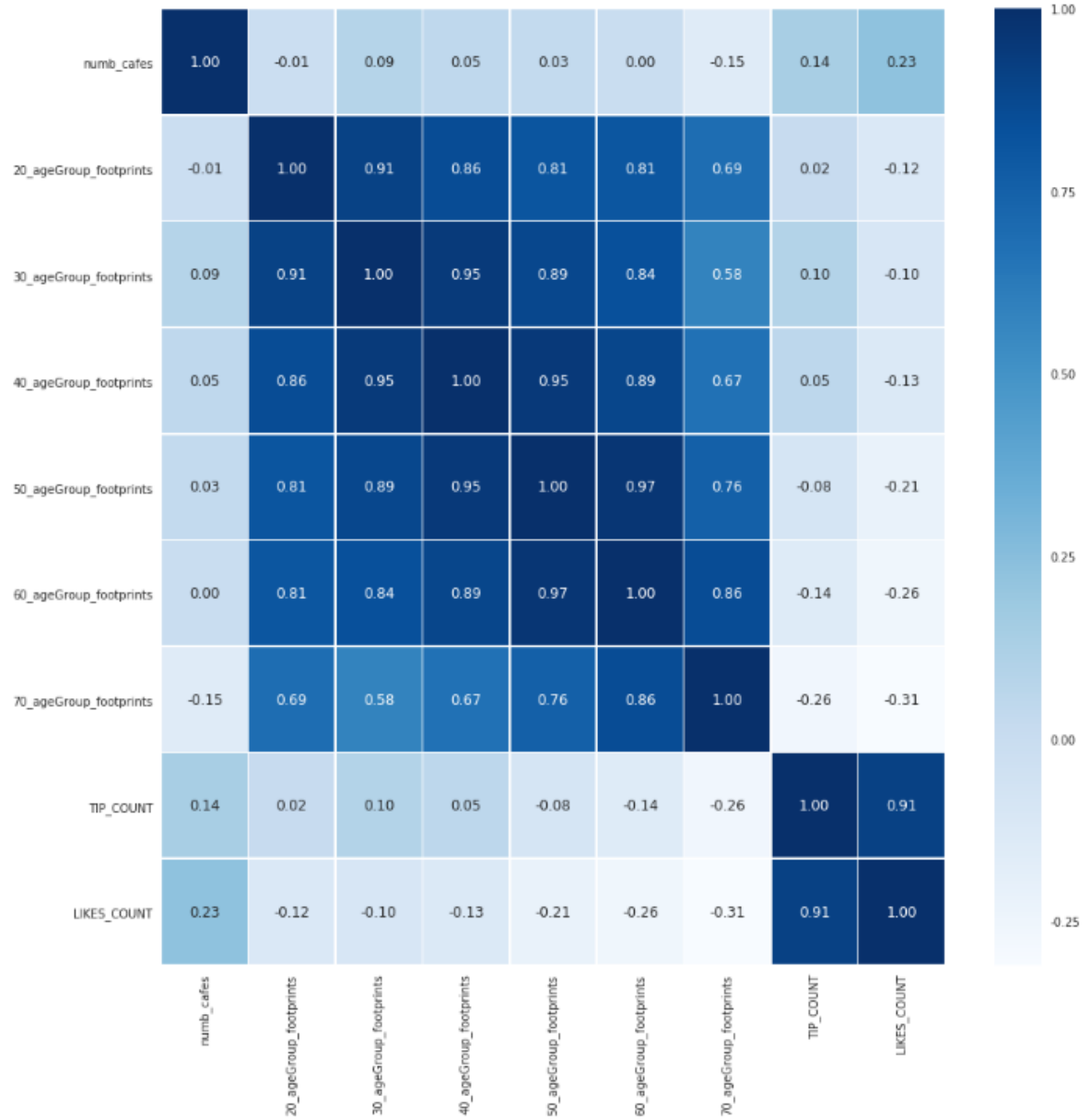
- In this part, with footprints data and cafe data, districts of Seoul are scored for decision making of opening cafe business. Firstable, all data is merged and dataframe data is prepared for scoring districts.

	군구	시군구명_영문	numb_cafes	20_ageGroup_footprints	30_ageGroup_footprints	40_ageGroup_footprints	50_ageGroup_footprints	60_ageGroup_footprints	70_ageGroup_footprints	TIP_COUNT	LIKES_COUNT
0	도봉구	Dobong-gu	3	38750.815217	39985.774457	47534.483696	50481.875000	33278.899457	28992.146739	0.000000	0.000000
1	은평구	Eunpyeong-gu	11	52611.576087	63266.508152	67973.260870	64607.459239	45404.524457	39906.467391	0.272727	0.181818
2	동대문구	Dongdaemun-gu	14	59387.486413	52022.581522	52908.355978	54369.470109	38051.494565	32088.301630	0.214286	0.142857
3	동작구	Dongjak-gu	6	66576.467391	63297.377717	55780.883152	55413.505435	38864.646739	31337.690217	0.166667	0.166667

Example of dataframe data

- Before scoring districts, correlation analysis is processed for recognizing relations among features. In this correlation analysis, we have to scale footprints data and counts of cafes, because of Covariance problem, which means range difference among data can make distortion of covariance. So, we have to match range of data, I will match it similar to range level of TIP_COUNT and LIKES_COUNT.

- After scaling data with MinMaxScaler from scikit-learn library, correlation calculation is processed and heatmap is made for visualization.



- 50, 60, 70 age groups are moving homogeneously and 30, 40 age groups are also like that.

- 50, 60, 70 age groups are negative to people's ratings of cafes.

- 20, 30, 40 age groups are positive slightly to tip_count(which means the number of check-ins).

- When more cafes are in province, their ratings are slightly high.

- In this analysis, we can find some information about age groups.

- 20, 30, 40 age groups are different with 50, 60, 70 age groups.

- So, if we decide to select best place for opening cafe business, we have to divide conditions and select proper place for cafe business for each condition.

- In generally, more cafes means more competitions exist, so we have to give penalty to numb_cafes. we have to divide customer segments into young age(20, 30, 40) groups and elder age(50, 60, 70) groups based on the correlation analysis. So, we can make score for selecting proper place of opening cafe business.

- When it comes to make score, I have to make numbers of cafes into the reciporal, because much the number of cafes means more competitions. so it is rejected for activating cafe business. I also have to plus 1 to features, TIP_COUNT and LIKES_COUNT, because I will make score for multiplying of features, so the value '0' makes distortions. In case of elder people, based on correlation analysis, their footprints trends are negative to likes_count, which means people's positive ratings of cafe, so I divide it.

	군구	score_for_young
9	강남구	318317.878170
6	종로구	120707.263858
24	용산구	55548.810386
17	노원구	45060.719270
8	중랑구	42188.662378
15	서초구	33810.319209
22	관악구	30887.519648
19	서대문구	24546.702899
5	구로구	19359.950349
23	성동구	16921.333031

	군구	score_for_elder
9	강남구	40335.048687
17	노원구	18706.743659
0	도봉구	12528.102355
23	성동구	11296.016002
24	용산구	9160.412138
21	영등포구	7380.126165
3	동작구	6978.657911
18	송파구	6769.841897
22	관악구	6684.839113
5	구로구	5901.660089

3. Results

From the trade area analysis for cafe business, Gangnam is the best place for opening cafe business.

The number of footprints from people in Gangnam is clearly more high than other districts of Seoul, importance of this affects the results.

Elder people moves homogeneously, because their footprints' correlations are one. But, They are heterogeneous from younger people, correlation between footprints of elder people and preferences of cafe at on-line are negative. So, we have to analyze and make score differently.

In case of opening cafe business that targets younger people, Jongno, Yongsan, Nowon, Jungang districts are also charming places to start business. Especially, Gangnam and Jongno districts' scores are distinguished from other districts.

In case of opening cafe business that targets elder people, Gangnam, Nowon, Dobong, Seongdong, Yongsan districts are charming places to start business. Gangnam district is the best place for opening cafe business same as results of targeting elder people. But, unlike results of targeting younger people, Dobong and Seongdong districts are charming places for cafe business targeting elder people.

4. Conclusion

So, after the trade area analysis for cafe business, based on analysis results, Gangnam is the best place for opening cafe business regardless of differences of preferences from age.

Footprints of people in Gangnam is especially distinguished. Footprints means

business activation degree of district, so cafes in Gangnam has high probability to gather people. Also, footprints of Gangnam is the best and the number of cafes in Gangnam is lower than most of other districts. So, Gangnam district is good place to open cafe.

If you want to target mainly young people, Jongno and Yongsan, Nowon districts are also recommended. And for elder people, Dobong and Seongdong districts are recommended.

All these above recommendations are totally based on just analysis.

But, in fact, It is hard to open cafe business at Gangnam district because building rent is too high to run cafe. Because there are so many offices in Gangnam district, it is not a unique findings that footprints in Gangnam is distinguished.

So, actually, I can recommend Nowon district is the alternative for opening cafe. (in my opinion, average rent in Yongsan district is also more high than Nowon district.)

Basically, Nowon district's average footprints are high level, and the number of cafes is lower than other districts. Many people pass by and stay in this district, but the number of cafes can't cover these people. In this point, we can find business opportunity. Maybe we can gather information about what is trending among people about cafe, dessert, beverages, interior of café, etc.

With this information, you can set your strategy and concept, interior of cafe and additional services. But, it is very important to find place that make money, also a standard for business. So, finally, this analysis suggests how to analyze spatial data and make insight from it.

5. Limitations

There are so many limitations and delay to operate this analysis. First, it is hard to

gather superior data for making good results. I have to gather data from so many different sources like SK data hub, Seoul public data center, 4squared. They require each standard to use their data.

So, you have to learn and know how to use each data source's API and code for gathering data. It spends much time for analysis.

What is fatal to this issue is the '4squared service'. 4squared service experienced many failures for vitalizing service. This service seeks people's culture that they feel free to share information about their experience of space.

But, as you can see in my data, count level is so low, maybe the reason is that this service is not activated among people.

Location data of cafes in Seoul can't cover reality. As a matter of fact, the number of cafes in Seoul is over 13,000 according to the press in Korea. 4sqaured location system can get location data from participation of users. So, I could get just 235 cafe information.

From the point of view in accuracy of this report, data gathered isn't enough to see all kinds of situations.

The fact that we can get just limited data from it make our insights narrower. If we get various kinds of data about cafe and spaces, it can makes us possible to analyze thorough many different views,even we can make regression model that can predict the revenue. On that point, this analysis offers a lesson that various and rich data can increase quality of analysis.

Spatial data has infinite chance to catch various kinds of social phenomenon. So, 4squared service is also a good attempt to catch spatial experiences. Experiences of people becomes important than ever, because it drives peoples' decisions in an instant. Time to decide for purchase decreases as days go by because of drastic technology progress. So, in the near future, we have to catch all kinds of glimpses of realities. Analysis on these will offers us more insights of reality.