

Personal coffee-cafe launching location recommendation in Mega-city, SEOUL, Korea

Seungwon Seo

June 7, 2019

1. Introduction

1.1 Background

Nowadays younger people who live and work in Seoul city, South Korea, have been ever thinking of opening own business like coffee-cafe at least once. That's the one of preferred business types after retiring from the company as salaryman. It is truly obvious item for their 2nd life, and its fact is on several trend report. There is huge market for the coffee item and also many of official skill courses for bringing up baristas in society either. It becomes easy for whoever thinking of own coffee cafe (like small deli, dessert café, etc) and lower some burden to start to open own cafe. In natural it brings hot competition of having own coffee cafe around hot-place like where many people gathered and spend most of times around like workplace in downtown, etc. When considering own coffee cafe, selecting good location considering to avoid hot competition with other cafes, or strategically focusing the spot where people easily access to the cafe is very important. In this report, given that here is true downtown of Seoul city, which is amazing mega-city living 10 millions people in South Korea, recommended place to have own coffee cafe thru the consideration of segmenting venues around downtown. Coffee is sort of the medium to required in communication for Korean people when they spend time with other people each other. In that point of view, the famous down, Seoul City-hall, is selected for the area having coffeecafe and segmentation after exploring nearby area characteristics of venues has been executed. With this segmentation, strategically feasible spot of opening coffee cafe will be recommended for better successful business opening.

1.2 Problem

When opening coffee cafe in the downtown, selecting coffee cafe location is very important and should be proceed carefully. Without thorough investigation of downtown venues including other competitive coffee cafe, easily own cafe will be falling into the failure, which no one likes to visit. Considering the famous downtown, Seoul city-hall, there are so many venues tempting people, picking good positioning for the location is definitely key factor for the success. If ignoring specific characteristics, types, kinds, categories of nearby venue, the aimed coffee cafe will be one of super normal cafe having destiny failure. Even if it sounds very hassle, venue info nearby location of downtown should be obtained and also segmented to look for strategically doable, reasonable place thru area clustering

1.3 Interest

Who may dream about having own coffee cafe business around the downtown in Seoul city will get interest of this report. Referring the recommendation about feasible, profitable spots after clustering results, dreamers can have the sight to see good spot to get more people and do sustainable coffee cafe by themselves.

2. Data acquisition and cleaning

2.1 Data sources

To have venues information around the aimed down town, Seoul city-hall, data from ‘Foursquare.com’ has been used. Interface to Foursquare is thru unique API for request and response interface. Current time line condition is applied and radius distance from City-hall to somewhere place to see venues is set as 15km. The number of nearby venues is assumed as fifty.

2.2 Data cleaning

In selecting wisely of the coffee-cafe location, other venues information is critically important. With search queary ‘food’, relevant venue related data has been obtained. Hence obtained data thru Foursquare API (‘Foursquare Agent’) should have venues, categories, address with priority, underscored data such as distance, formattedaddress, cc, country, labeledLatLngs, State was dropped from the table. All the venues has been listed up as much as fifty and its latitude, longitude info was leveraged to get exact location. Non particular NaN data in all the columns is removed for better data set sources. Also its dependency for the analyzed result is not that big as assumed.

2.3 Feature selection

After data cleansing with derived table along with 2.2 above, 4,988 number of venues has been came up with regarding the down town spot info of Seoul city as 37.56 for latitude and 126.97 for longitude. This venues data has 184 unique categories for the venue, that is, with that categories information all the characteristics of venue is easily figured. It will be much of being helpful decide which nearby venue address from downtown is better selection for the cafe.

To group these data with actual address information, ‘groupby’ action has been done for dataframe and finally 34 rows and 182 columns of shape was derived. For the information, here below figure.1 shows about first 5 rows of address groupby data.

In addition, even if it has thousands of venues data, only top 10 limit of venues were considered for the feature selection too.

Against final data set for 10 limit venues with 50 counts of row, clustering will be executed for better segmentation in order to hit high profit proven population selection and to avoid meaningless venue area together. Machine learning method for clustering will be ‘k-means’ algorithm, that is already prepared in python library.

address	Cheonggyecheon 40	Lotte Mart	N서울타워 Plaza 1F	강남구 도산대로 442	강남구 도산대로89길 7
1st Most Common Venue	Hotel	Hotel	Korean Restaurant	Japanese Restaurant	Coffee Shop
2nd Most Common Venue	Korean Restaurant	Korean Restaurant	Café	Italian Restaurant	BBQ Joint
3rd Most Common Venue	Café	Café	Hotel	Korean Restaurant	Japanese Restaurant
4th Most Common Venue	Coffee Shop	Chinese Restaurant	Theater	Coffee Shop	Cocktail Bar
5th Most Common Venue	Chinese Restaurant	Coffee Shop	Gym / Fitness Center	Cocktail Bar	Italian Restaurant
6th Most Common Venue	BBQ Joint	Dessert Shop	Park	BBQ Joint	Korean Restaurant
7th Most Common Venue	Historic Site	Bookstore	Bakery	Sushi Restaurant	Whisky Bar
8th Most Common Venue	Theater	Cosmetics Shop	Trail	Café	French Restaurant
9th Most Common Venue	Plaza	BBQ Joint	Italian Restaurant	Chinese Restaurant	Modern European Restaurant
10th Most Common Venue	Dessert Shop	Historic Site	Scenic Lookout	Whisky Bar	Salon / Barbershop

Fig 1. Top 10 limit of venue value by top 5 location nearby Seoul-city

3. Methodology

3.1 Exploratory Data Analysis

To have proper point to locate café in the down town, venues nearby the center point of Seoul city (37.56, 126.97) was derived thanks to Foursquare location API. Since key information was about the ‘address’ around the downtown, venues & venue category was included in derived table of data, shown in Fig 2. and Fig 3. below.

	address	address Latitude	address Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	중구 남대문로 81	37.564695	126.98116	Soo Fitness & Spa	37.565317	126.981374	Gym / Fitness Center
1	중구 남대문로 81	37.564695	126.98116	청담스포피아	37.565317	126.981374	Golf Course
2	중구 남대문로 81	37.564695	126.98116	The Westin Chosun Seoul (웨스틴조선호텔)	37.564378	126.980058	Hotel
3	중구 남대문로 81	37.564695	126.98116	Nanta Theater (난타극장)	37.563688	126.983687	Theater
4	중구 남대문로 81	37.564695	126.98116	Arc N Book (아크엔북)	37.565930	126.981370	Bookstore

Fig 2. A head information about dataframe having ‘address’, ‘Venue’, and ‘Venue Category’

With the address information, detailed data about Venue Category had to be obtained to see more detail provision in aimed address where one can possibly pick for the café. First of all, number of venues around the down town was investigated thru the Foursquare API and received result as below Fig 3. Most of nearby area has hundred of venues in there, which reflects there are sort of hot place people want to gather often. That’s the analysis for 34 nearby places where potentially own café will be located.

address	address Latitude	address Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Cheonggyecheon 40	100	100	100	100	100	100
Lotte Mart	100	100	100	100	100	100
NA서울타워 Plaza 1F	100	100	100	100	100	100
강남구 도산대로 442	100	100	100	100	100	100
강남구 도산대로 89길 7	100	100	100	100	100	100
강남구 압구정로 165	100	100	100	100	100	100
강남구 인주로 30길 21	100	100	100	100	100	100
마포구 성산동 515 난지도길 66	97	97	97	97	97	97
서대문구 신촌동 연세세브란스 본관 3층	100	100	100	100	100	100
서대문구 신촌로 83	100	100	100	100	100	100
서대문구 신촌역로 19-2	100	100	100	100	100	100
서대문구 이화여대길 33	100	100	100	100	100	100
서울특별시 종구 명동	100	100	100	100	100	100
성동구 행당동 한양대병원 동관 1층	100	100	100	100	100	100
양천구 목동동로 257	91	91	91	91	91	91
연주로 30길 57	100	100	100	100	100	100
용산구 신흥로 20길 45	100	100	100	100	100	100
이태원로 192	100	100	100	100	100	100
자하문로 10	100	100	100	100	100	100
장충단로 7길 9-1	100	100	100	100	100	100
종로구 관수동	100	100	100	100	100	100
종로구 동승동	100	100	100	100	100	100
종로구 신문로 1가 163 광희문오피시아 지하1층	100	100	100	100	100	100
종로구 신문로 2가 12-5 3층	100	100	100	100	100	100
종로구 안국동 21-1	100	100	100	100	100	100
중구 남대문로 81	100	100	100	100	100	100
중구 남대문시장 4길 12-2	100	100	100	100	100	100
중구 동호로 330	100	100	100	100	100	100
중구 소공로 63 신세계백화점 본점 B1	100	100	100	100	100	100
중구 신당2동 종하빌딩 A B	100	100	100	100	100	100
중구 을지로 5길 26	100	100	100	100	100	100
중구 장충단로 275	100	100	100	100	100	100
중구 충무로 1가 52-5	100	100	100	100	100	100
중구 태평로	100	100	100	100	100	100

Fig 3. Number of venues around the down town (limit is on 100 for the cap)

Then more specific information is need to see categorical types about specific venues against one single address area. Hence, mean of the frequency for occurrence of each category was derived too like Fig 4. Its total number of addressss groupby venue category is 40, but here is briefly shown as 10. Also, the number of category reached to 334, so below Fig 4. can't present them all. Simply it present 12 categories to see features little bit.

	address	African Restaurant	American Restaurant	Aquarium	Arcade	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	BBQ Joint	Bagel Shop	Bakery	Bar	Baseball Stadium
0	Cheonggyecheon 40	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.000000	0.040000	0.01	0.020000	0.00	0.000000
1	Lotte Mart	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.000000	0.030000	0.00	0.030000	0.02	0.000000
2	N서울타워 Plaza 1F	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.000000	0.020000	0.00	0.040000	0.02	0.000000
3	강남구 도산대로 442	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.000000	0.060000	0.00	0.000000	0.02	0.000000
4	강남구 도산대로89길 7	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.000000	0.100000	0.00	0.000000	0.01	0.000000
5	강남구 압구정로 165	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.000000	0.080000	0.00	0.050000	0.02	0.000000
6	강남구 언주로30길 21	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.010000	0.080000	0.00	0.100000	0.01	0.000000
7	마포구 성산동 515 난지도길 66	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.010309	0.072165	0.00	0.041237	0.00	0.000000
8	서대문구 신촌동 연 세세브란스 본관 3층	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.000000	0.040000	0.02	0.060000	0.00	0.000000
9	서대문구 신촌로 83	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.000000	0.070000	0.02	0.060000	0.00	0.000000

Fig 4. Number of venues around the down town (limit is on 100 for the cap)

And, tens of venues along each address should be marked on the map in order to have more clear vision of it. Fig 5 shows that key venues location has been distributed nearby the center point of Seoul city, the most largest down town in South Korea. It assumed that 15km radius distance from the downtown was considered for the analysis.

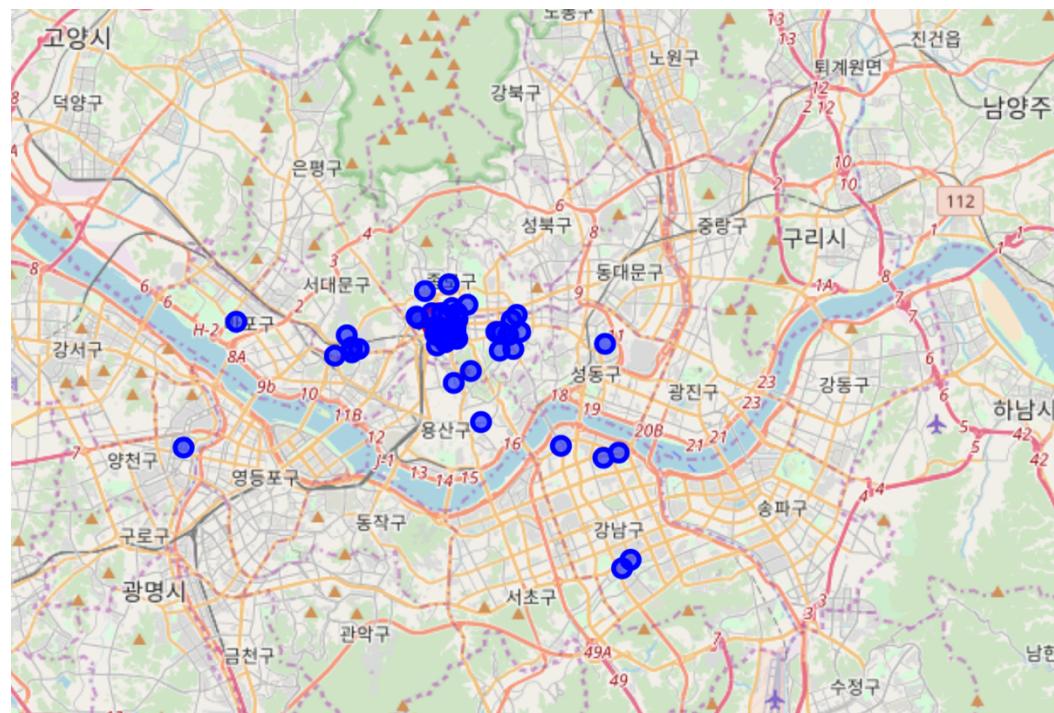


Fig 5. Derived venues location map nearby the center point of Seoul, which is the core down town

To see better vision around the down town inside, zoom-in feature was obtained as can see below Fig 6..

Red color means that is Seoul city-hall, the center of Seoul. Blues circle spots represent the key venues around.

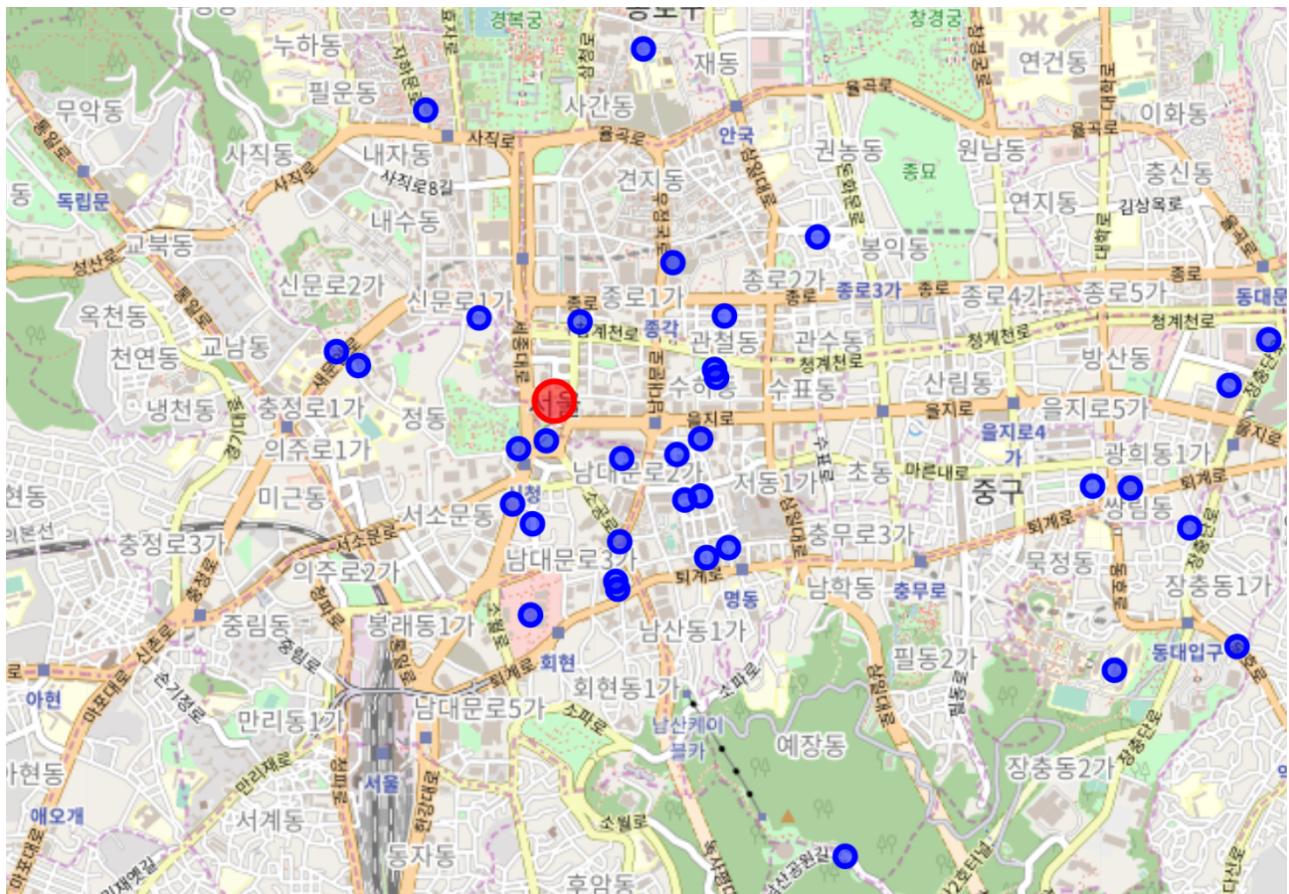


Fig 6. Venues locations nearby the Seoul city-hall with zoom-in sight view

After checking key venue distribution around the down town, clustering work was required to see the segmentation like Fig 7.

5 clustering was executed in this experiment and specific characteristic information by each cluster was obtained as Fig 8.

5 clusters was represent by 5 different colors to figure out the segmentation easily. In this report, there was some of insufficient dataset on the cluster being violet color by Foursquare dataset to South Korea region. Sort of far distributed dataset for a cluster has been seen. However, rest of 4 different color based clusters shows hot-place around the down town actually, and it is quite matched with public preference to hang out to the down town.

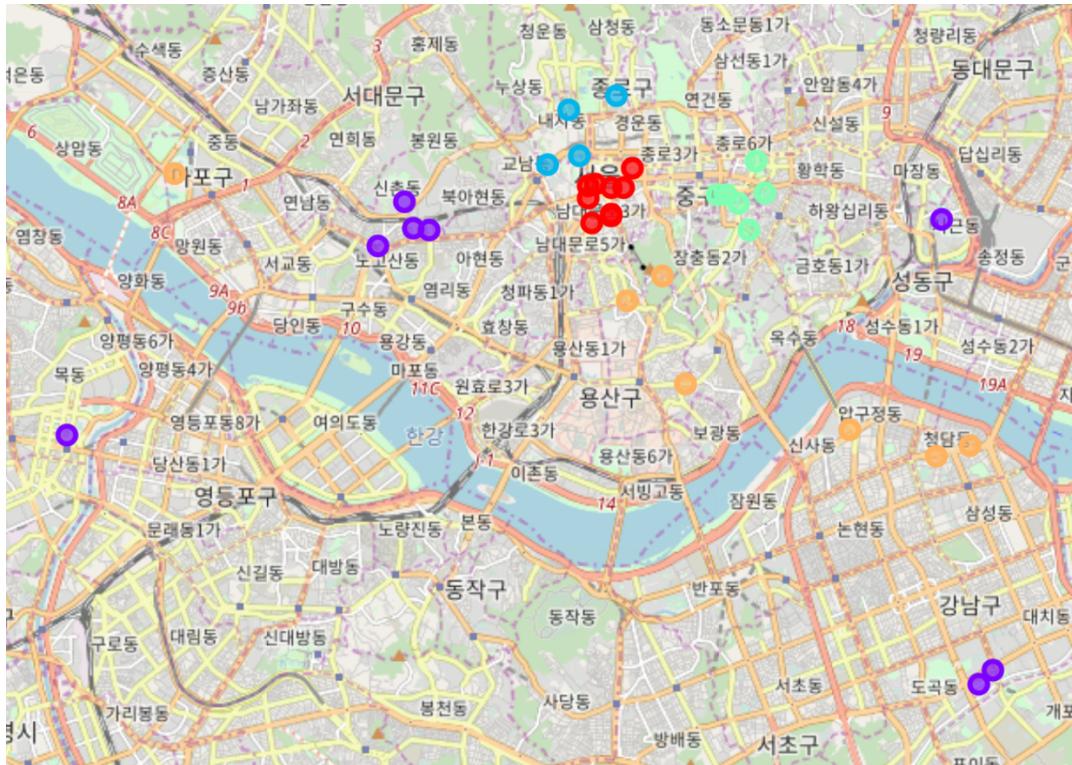


Fig 7. 5 clusters representation with different colors on the map of Seoul city.

Finally, in order to catch reasonable coffee café location in the Seoul, each venue category information was checked thru Fig 8.

1st Cluster	categories	Cluster Label	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
Food Court		0	Hotel	Korean Restaurant	Café	Chinese Restaurant	Coffee Shop
Food Truck		0	Hotel	Korean Restaurant	Café	BBQ Joint	Coffee Shop
Gourmet Shop		0	Hotel	Korean Restaurant	Café	Bakery	Chinese Restaurant
Food Court		0	Hotel	Korean Restaurant	Café	Bakery	BBQ Joint
Grocery Store		0	Korean Restaurant	Hotel	Coffee Shop	Historic Site	Café
Korean Restaurant		0	Hotel	Korean Restaurant	Café	Coffee Shop	Chinese Restaurant
Korean Restaurant		0	Hotel	Korean Restaurant	Café	Chinese Restaurant	Bakery
Greek Restaurant		0	Hotel	Korean Restaurant	Café	Coffee Shop	Chinese Restaurant
Seafood Restaur		0	Hotel	Korean Restaurant	Café	Chinese Restaurant	Coffee Shop

2nd Cluster	categories	Cluster Label	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
Burger Joint		1	Café	Coffee Shop	Korean Restaurant	Cocktail Bar	Hotel
Bunsik Restaurant		1	Coffee Shop	Korean Restaurant	Japanese Restaurant	Bakery	Dessert Shop
Turkish Restaurant		1	Coffee Shop	Korean Restaurant	Bakery	Lounge	Cocktail Bar
Food Court		1	Coffee Shop	Café	Bakery	Chinese Restaurant	Dessert Shop
Food Court		1	Park	Korean Restaurant	BBQ Joint	Coffee Shop	Café
Bunsik Restaurant		1	Coffee Shop	BBQ Joint	Italian Restaurant	Cocktail Bar	Japanese Restaurant
Grocery Store		1	Japanese Restaurant	Italian Restaurant	Korean Restaurant	BBQ Joint	Coffee Shop
Grocery Store		1	Korean Restaurant	Japanese Restaurant	Coffee Shop	BBQ Joint	Bakery
Grocery Store		1	Coffee Shop	BBQ Joint	Dessert Shop	Korean Restaurant	Bakery
Snack Place		1	Coffee Shop	Bakery	Café	Korean Restaurant	BBQ Joint

3rd Cluster	categories	Cluster Label	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
Food Court		2	Korean Restaurant	Hotel	Noodle House	Coffee Shop	Bakery
Food Court		2	Korean Restaurant	Coffee Shop	Noodle House	Hotel	Market
Restaurant		2	Korean Restaurant	Hotel	Noodle House	Coffee Shop	Market
Cosmetics Shop		2	Korean Restaurant	Hotel	Noodle House	Coffee Shop	Bakery
Office		2	Korean Restaurant	Coffee Shop	Hotel	Noodle House	Bakery
Food Truck		2	Korean Restaurant	Coffee Shop	Hotel	Noodle House	Chinese Restaurant

4th Cluster	categories	Cluster Label	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
Supermarket		3	Korean Restaurant	Coffee Shop	Bakery	BBQ Joint	Chinese Restaurant
Supermarket		3	Korean Restaurant	Coffee Shop	Bakery	BBQ Joint	Café
Food Court		3	Coffee Shop	Korean Restaurant	BBQ Joint	Bakery	Café
Bunsik Restaurant		3	Korean Restaurant	Coffee Shop	Seafood Restaurant	Café	BBQ Joint

5th Cluster	categories	Cluster Label	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
Japanese Curry Res		4	Coffee Shop	Korean Restaurant	Café	Italian Restaurant	Art Gallery
Bunsik Restaurant		4	Café	Tea Room	Coffee Shop	Italian Restaurant	Historic Site
Restaurant		4	Korean Restaurant	Coffee Shop	Café	Hotel	Historic Site

Fig 8. Most common venue information by each cluster. (For better capturing of data, only 5 priority on common

venue was set for the summary)

When taking looking at the top3 common venues, it can be determined which cluster will be proper to get in avoiding hot competition of café crowded in nearby downtown. Considering the size of cluster and venue kinds, 1st and 3rd cluster seems to be proper to get in softly.

4. Result

In summary, about 50 venues nearby the down town was investigated and its categorical info was obtained to see thru the characteristics. Within 15 km radius from the center point of downtown, which is Seoul city-hall, all the possible venues were derived as long as Foursquare API support possible dataset, even for South Korea area. Finally 5 clusters were gained on the map, it shows practical and proven distribution of venues along with physical voice of public people hangin around. 5 Clusters has own different configuration of venue categories and it is helpful to figure out there business kind of situation based on data. Looking up cluster 1 to 5, business decision can be made to get in the down town place wisely.

5. Discussion

Based on 5 clusters characteristics, the area where one new café does not need to get in is easily figure. For example, 2nd cluster has 1st common venue of café types in the priority. There are full of café already. Hot competition is foreseen thru this data analysis ahead. It should be avoided. In that point of view, 1st cluster and 3rd cluster seems to be proper and reasonable considering size of how much people gather and already taken venue categories. If more easy access to the downtown, comfortable locating of café with nearby venues 1st cluster will be recommended in this report for one's soft landing business of own coffee café.

6. Conclusion

In report, geographical venue data (categories, address, latitude, longitude, etc) were investigated and sanitized to see better understanding of aimed region, when one start coffee-café business around the down town, megacity Seoul, in South Korea. K-mean clustering method was used to have feasible segmentation around the down town given condition 15 km radius distance from the city hall. Looking at the differentiated characteristics of each cluster one can eaily figure out its venue style of candidate area for making better and safe business decision. If more sufficient data like workplace kinds, companies, average salaries of people working or living in the cluster, more feasible and accurate decision could be made of either.