

The Battle Of Neighborhood - Final Week

1. Introduction & Business Problem

Introduction

- Pre-analysis of market before Business Opening

Seoul, officially the Seoul Special City, is the capital and largest metropolis of South Korea. Most of the largest business or market of South Korea is located in Seoul. It is not limited in Financial business, but also including other like Manufacturing, wholesale, retail, and more.

Due to the above reason, it is the big challenge to open business in Seoul. Business market is highly competitive but owner might grab the change to get high return from their business.

The report will provide business insight in advance to start any kinds of business in Seoul by segmented, visualized information.

Business Problem

- what kind of business you want to start? It will provide simple insight to start the business at specific location.

There are several business venues can expend to real business industry. It can be a restaurant, a Cafe, or even a company in special industry. Based on the kind of venue, weighted segmentation standards are different. But most of small business owners do not know what kind of standards are important to set the point of business start.

And also, it is highly required to understand compattors to start business. For example, if owner would like to open a korean food restaurant, they need to get the information of the location nearby.

Various information & factors need to be utilized for the above insight.

- Geolocation Data of Seoul District, Neighborhood
- Living Population Data of Seoul by District, Neighborhood
- Geolocation Data of Seoul Venues
 - Any convenient venues located nearby the place?
 - Any competitors are located nearby the place?
- Segmentation of the Seoul District

Target

This information insight will locate and recommend the neighborhood to business owner which is best 5 choices to start their business.

2. Data

In this project, the City called Seoul will be the main place to analyse.

Data 1: Seoul Geolocation data with district, Neighborhood

Seoul has 25 Districts and 425 Neighborhoods. South Korea government proceeding a big data project to document these data as structured one, and it is sharing through various sources.

In this project, structured Seoul districts, neighborhoods geolocation data will be used.

- Data is gathered from (sgis.kostat.go.kr)
- Data is shared from (https://github.com/busk/admdongkor)
- Data source link is (https://github.com/busk/admdongkor/blob/master/ver20200101/HangJaongDang_ver20200101_gesjson)

Data 2: Seoul Living population by Datetime, age, district, neighborhood

To analyse market in Seoul, living population listed by time, age and places is highly required. As part of the South Korea big data project, Government has signed contract with Korea telecommunication companies like KT, SKT, LGT, and gathered living population data with them.

- Data is gathered from (data.seoul.go.kr)
- Data source is available from (http://data.seoul.go.kr/dataList/GA-15438/SC1/datasetView.do)

Data 3: Seoul geographical coordinates data from Foursquare

In this project, Foursquare API will be utilized to gather data of geolocation, various venue data like category, name geolocation and etc.

3. Methodology

Business understanding

In this project, Top 5 recommended District or Neighborhood and simple analysis report will be provided to business user based on their requirements.

- Restaurant Owner - Based on population status and venue, neighborhoods will be segmented and sorted
- Office Owner - Based on population status, simple analysis will be performed to understand best place to locate office.

Analytic Approach

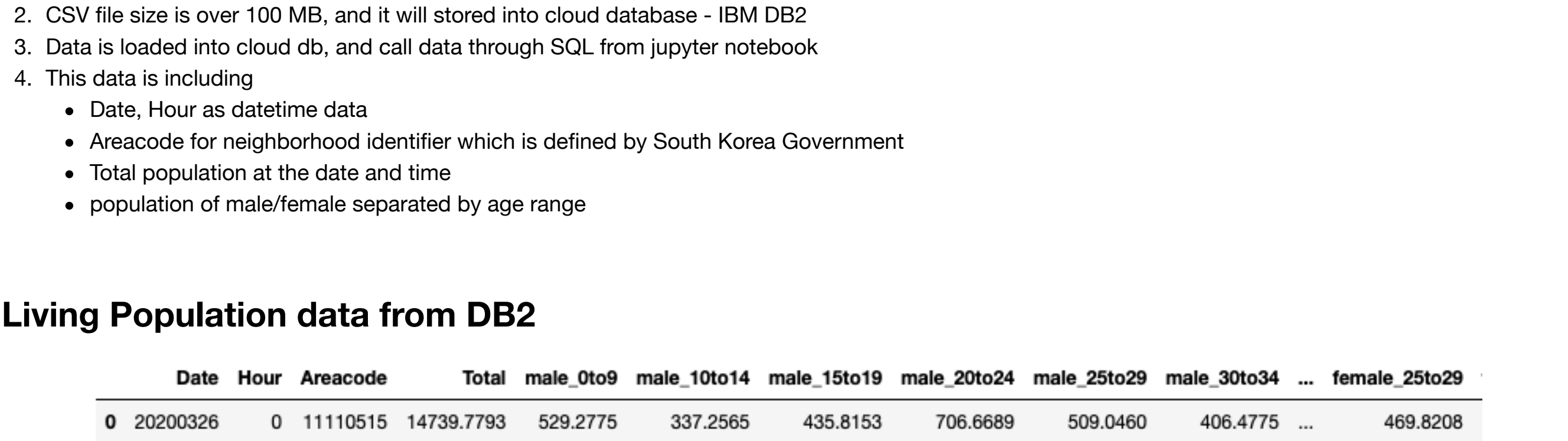
As mentioned above, Seoul has 25 Districts and 425 Neighborhoods. Based on living population analysis, we are going to proceed segmentation by district area. Details will be followed at the EDA below.

Explortory Data Analysis

Data 1: Seoul Geolocation data with district, Neighborhood

- Data is provided with GeoJSON format from github.
- Data is loaded by GeoPandas - read file function as Data Frame
- This Data Frame is including
 - District, Neighborhood name and government management code
 - Geometry data by multipolygon lat, long
- Using centroid function to get the center point of geolocation multipolygon data
- Using Folium to visualize Seoul district, neighborhood data into map
- Foursquare API will be utilized to get venue data

Seoul Map with District Area, Neighborhood point



Data 2: Living Population Data of Seoul

- Data is provided by CSV format from data.seoul.go.kr
- CSV file size is over 100 MB, and it will stored into cloud database - IBM DB2
- Data is loaded into cloud db, and call data through SQL from jupyter notebook
- This data is including
 - Date, Hour as datetime data
 - Areacode for neighborhood identifier which is defined by South Korea Government
 - Total population at the date and time
 - population of male/female separated by age range

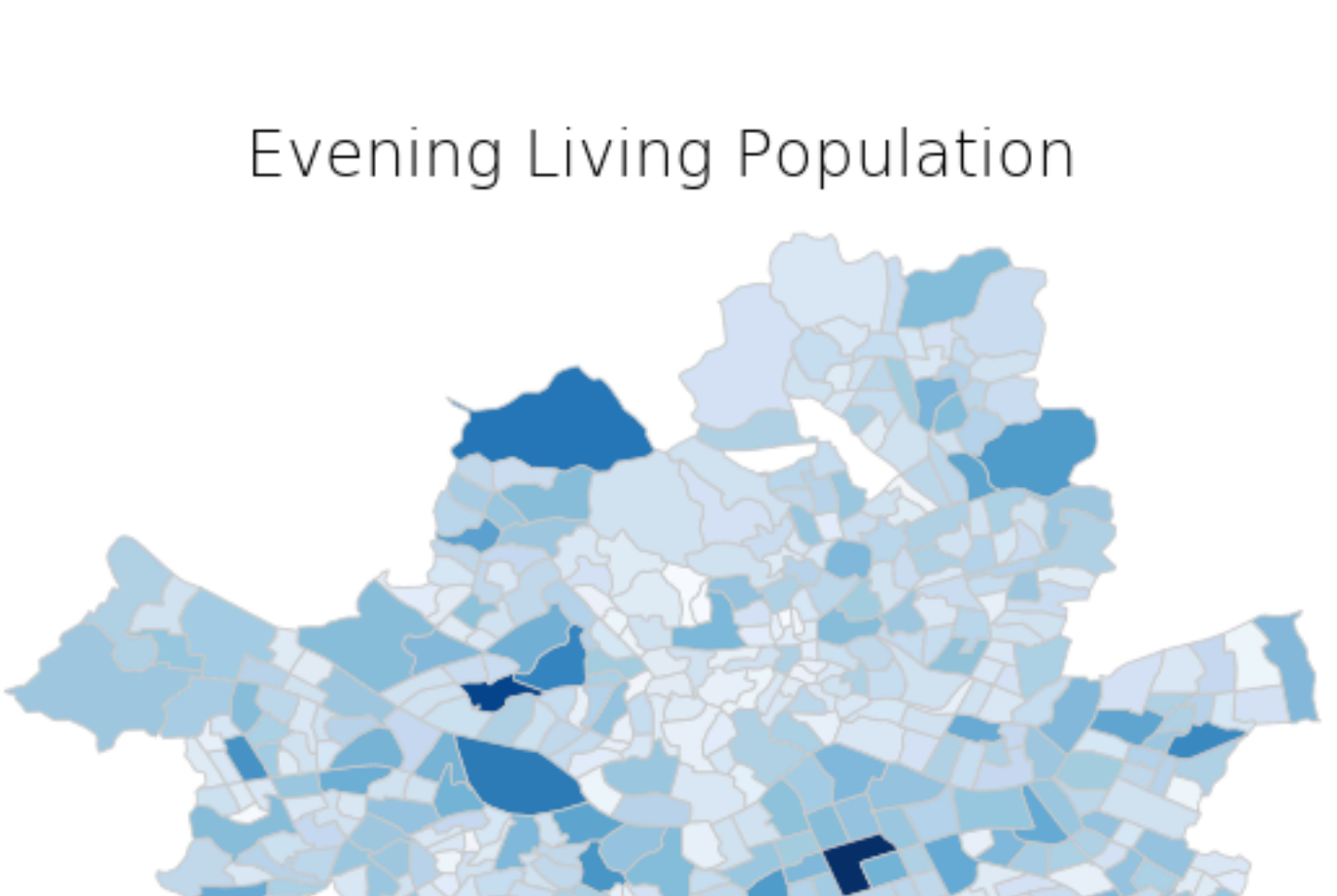
Living Population data from DB2

	Date	Hour	Areacode	Total	male_0to9	male_10to14	male_15to19	male_20to24	male_25to29	male_30to34	...	female_25to29
0	20200326	0	11110515	14739.7793	529.2775	337.2565	435.8153	706.6689	509.0460	406.4775	...	469.8208
1	20200326	0	11110530	14253.1012	326.4137	145.0726	220.0557	430.0698	513.6943	541.8583	...	532.2772
2	20200326	0	11110540	4542.9356	78.4293	60.0854	139.9614	393.9535	153.0175	180.0634	...	123.0728
3	20200326	0	11110550	16053.9286	467.9358	362.8310	426.2707	510.7543	458.0618	383.8269	...	442.4556
4	20200326	0	11110560	19814.0498	526.4606	380.8060	634.5152	650.6258	539.8961	396.1611	...	486.8294

Seoul Population - insight from Data

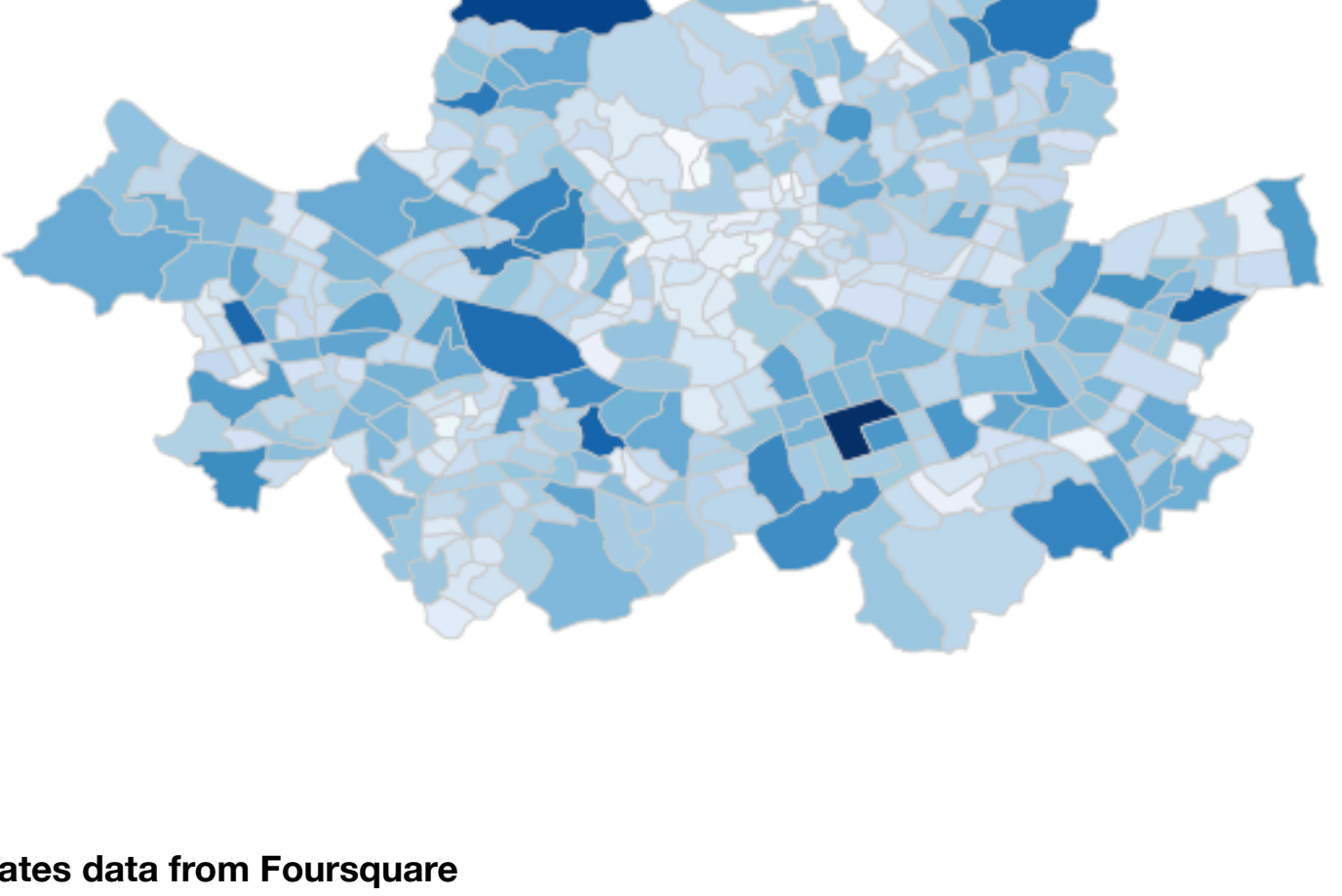
- Based on living population data within Working Hour, Top 100 Neighborhoods contains 45 percent of total population.
 - 38 % of above Top 100 population are from only 4 Districts.
 - Gang-nam, Seo-cho, Song-pa and Gang-seo
 - 20 % of Top 100 population are also from only 4 Districts.
 - No-won, Dong-jak, Ma-po, Gu-ro
 - It shows that Major business neighborhoods of Seoul are located at the above districts.
 - 58 % of Top100 population
- However, it also shows that some specific Neighborhoods which are not in above listed districts.
 - Top 5 Neighborhood
 - 역삼1동 - Yeok-sam 1dong (Gang-nam)
 - 여의동 - Yeo-ui dong
 - 종로1·2·3·4가동 - Jong-ro 1,2,3,4 dong
 - 서교동 - Se-kyo dong (Hong Dae)
 - 가산동 - Ga-san dong

Working Hour Living Population



- Based on living population data after working hour, Top 100 Neighborhoods contains 38 percent of total population.
 - Mainly, listed districts are similar as major business districts, but highest district is different.
 - Song-pa is the highest district of this data value.
 - Based on data, we can say that if employees need to spend time after working hour, they are spending near their companies.
- Some specific neighborhoods which are not listed in above districts list, are hotplaces of Seoul to spend evening time.
 - like hong-dae, sin-chon
- Top 5 Neighborhood
 - 역삼1동 - Yeok-sam 1dong (Gang-nam)
 - 서교동 - Se-kyo dong (Hong Dae)
 - 진관동 - Jin-Gwan dong
 - 여의동 - Yeo-ui dong
 - 신촌동 - Sin-Chon dong

Evening Living Population



- Based on living population data in Dawn time, Top 100 Neighborhoods contains 38 percent of total population.
 - Listed districts are almost same as above - after working hour data.
 - Hotplaces of Seoul still shows high population value in dawn time.
- Top 5 Neighborhood
 - 역삼1동 - Yeok-sam 1dong (Gang-nam)
 - 진관동 - Jin-Gwan dong
 - 길동 - Gil-dong
 - 상도1동 - Sang-do 1dong
 - 화곡1동 - Hwa-gok 1dong

Dawn time Living Population



Data 3: Seoul geographical coordinates data from Foursquare

- Geolocation data is utilized as input for the Foursquare API
- Set python fuction to get venue data
- one hot encoding of all venue categories from Neighborhood, and calculate mean of category
- based on above result, get highest 10 venues

Top 10 Venues from Neighborhoods

Neighborhood	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
가산동	Korean Restaurant	BBQ Joint	Coffee Shop	Japanese Restaurant	Hotel	Chinese Restaurant	Beer Bar	Japanese Restaurant	Sauna	Chinese Restaurant
가산동	Korean Restaurant	BBQ Joint	Coffee Shop	Japanese Restaurant	Hotel	Chinese Restaurant	Beer Bar	Japanese Restaurant	Sauna	Chinese Restaurant
가산동	Korean Restaurant	BBQ Joint	Coffee Shop	Japanese Restaurant	Hotel	Chinese Restaurant	Beer Bar	Japanese Restaurant	Sauna	Chinese Restaurant
가산동	Korean Restaurant	BBQ Joint	Coffee Shop	Japanese Restaurant	Hotel	Chinese Restaurant	Beer Bar	Japanese Restaurant	Sauna	Chinese Restaurant
가산동	Korean Restaurant	BBQ Joint	Coffee Shop	Japanese Restaurant	Hotel	Chinese Restaurant	Beer Bar	Japanese Restaurant	Sauna	Chinese Restaurant
가산동	Korean Restaurant	BBQ Joint	Coffee Shop	Japanese Restaurant	Hotel	Chinese Restaurant	Beer Bar	Japanese Restaurant	Sauna	Chinese Restaurant
가산동	Korean Restaurant	BBQ Joint	Coffee Shop	Japanese Restaurant	Hotel	Chinese Restaurant	Beer Bar	Japanese Restaurant	Sauna	Chinese Restaurant
가산동	Korean Restaurant	BBQ Joint	Coffee Shop	Japanese Restaurant	Hotel	Chinese Restaurant	Beer Bar	Japanese Restaurant	Sauna	Chinese Restaurant
가산동	Korean Restaurant	BBQ Joint	Coffee Shop	Japanese Restaurant	Hotel	Chinese Restaurant	Beer Bar	Japanese Restaurant	Sauna	Chinese Restaurant
가산동	Korean Restaurant	BBQ Joint	Coffee Shop	Japanese Restaurant	Hotel	Chinese Restaurant	Beer Bar	Japanese Restaurant	Sauna	Chinese Restaurant

4. Result

Clustering Neighborhood, and seperately analyze by Area

Business Area

- Clustered by nearby common venue
 - Coffee shop, Restaurant, Stores or Activity, Natural environment
- Top 5 Neighborhood Analyze
 - 역삼1동 - Yeok-sam 1dong (Gang-nam)
 - High Risk, High Return Area to start business. Venues are placed commensurately.
 - You can consider to open activity business in this area.
 - 여의동 - Yeo-ui dong
 - You can consider to open Korean Restaurant in this area.
 - Coffee shop is not recommended.
 - 종로1·2·3·4가동 - Jong-ro 1,2,3,4 dong
 - You can consider to open Japanese Restaurant in the area.
 - 서교동 - Se-kyo dong (Hong Dae)
 - Similar with Gang-nam. You can consider to open activity business in this area.
 - 가산동 - Ga-san dong
 - Venues are placed commensurately.

District	Neighborhood	Areacode	Total	Per	long	lat	Cluster Labels	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
관악구	역삼1동	11680640	93324868	1.207123	127.035491	37.500889	2	Korean Restaurant	BBQ Joint	Coffee Shop	Chinese Restaurant	Hotel	Japanese Restaurant	Sauna	Sushi Restaurant	Italian Restaurant	Sauna
영등포구	여의동	11680540	49989377	1.219887	126.927571	37.527489	0	Coffee Shop	Park	Cafe	Plaza	BBQ Joint	Chinese Restaurant	Sauna	Sushi Restaurant	Hotel	Japanese Restaurant
종로구	종로1·2·3·4가동	11110515	41123842	1.003334	126.989733	37.575077	2	Korean Restaurant	Coffee Shop	Cafe	Coffee Shop	Dumping Restaurant	Beer Bar	Japanese Restaurant	Italian Restaurant	Museum	Hotel
마포구	서교동	11440960	36200535	0.880721	126.920945	37.554772	0	Korean Restaurant	Tea Room	Coffee Shop	BBQ Joint	Dessert Shop	Chinese Restaurant	Bakery	Hotel	Bar	Japanese Restaurant
금천구	가산동	11544510	31985889	0.779247	126.884341	37.478635	0	Coffee Shop	Korean Restaurant	Cafe	BBQ Joint	Sporting Goods Shop	Chinese Restaurant	Steakhouse	Bakery	BBQ Joint	Japanese Restaurant

Hot Place Area

- Clustered by nearby common venue
 - Coffee shop, Restaurant, Stores or Activity, Natural environment
- Top 5 Neighborhood Analyze
 - 역삼1동 - Yeok-sam 1dong (Gang-nam)
 - Same as Business Area clustering information.
 - 서교동 - Se-kyo dong (Hong Dae)
 - Same as Business Area clustering information.
 - 진관동 - Jin-Gwan dong
 - You can consider to open Restaurant business in this area.
 - Any Cuisine can be considered.
 - 여의동 - Yeo-ui dong
 - Same as Business Area clustering information.
 - 신촌동 - Sin-Chon dong
 - You can consider to open activity business in this area.

District	Neighborhood	Areacode	Total	Per	long	lat	Cluster Labels	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
관악구	역삼1동	11680640	12317512	0.679090	127.035491	37.500889	3	Korean Restaurant	BBQ Joint	Coffee Shop	Chinese Restaurant	Hotel	Japanese Restaurant	Cafe	Sauna	Sushi Restaurant	Sauna
영등포구	여의동	11680540	12262365	0.622523	126.940518	37.539279	2	Bakery	Coffee Shop	Cafe	Museum	Planetarium	Bistro	Trail	Food	Plex Market	Fish Market
금천구	가산동	11744965	10838712	0.592049	127.149007	37.539725	3	Korean Restaurant	Plex Market	Noodle House	Cafe	Bakery	Yoga Studio	Food Court	Food	Fish Market	Fast Food Restaurant
동작구	상도1동	11809530	10614079	0.538845	126.953470	37.488329	3	Coffee Shop	Ice Cream Shop	Sandwich Place	Ice Cream Shop	Fast Food Restaurant	Pizza Place	Gym/Fitness Center	Dumping Restaurant	Donut Shop	Fish Market
강서구	화곡1동	11500540	10628261	0.534387	126.840936	37.533780	0	Bakery	Ice Cream Shop	Pizza Place	Bus Station	Grocery Store	Market	Yoga Studio	Food	Plex Market	Multiplex

5. Discussion

- 역삼1동 - Yeok-sam 1dong (Gang-nam) is the place with highest population value not only in business hour, but also in other times.
- Korean Restaurant and Coffee shop are highest common venues from most of places.
- Recommended Business can be listed as above.

6. Conclusion

The report is generated on limited data only. There are many options can be added to get more information, like estate data, employee satisfaction result and so on. It can be improved if better quality of data is available.

Gang-Nam, Yeo-ui, Hong-dae are places where having compattive market, but at the same time there have high potential to success with many options.