Capstone Project

Recommendation for new house location

Nov. 5th, 2019

Hyejeong Kang

Agenda

1. Executive Summary
2. Business Problem
3. Data
4. Methodology
5. Results and Discussion
6. Conclusion

1. Executive Summary

This report will provide the methodology and result in finding the optimal residential locations for employees who move and work for Samsung Digital City in Suwon, South Korea. Since the result is based on the limited data and skills of the author, I would like to say that you regard this report as reference only.

2. Business Problem

My stakeholders plan to move to a new neighborhood near Samsung Digital City in Suwon, South Korea. They work for Samsung Electronics and want to find the best residential area around their office. Below are elements that optimal house location should have

- Near Samsung Digital City at Suwon
- Close to Park or Green place
- Easy to access basic amenities

3. Data

I mainly collected two types of datasets using Google geocode API and Foursqure API.

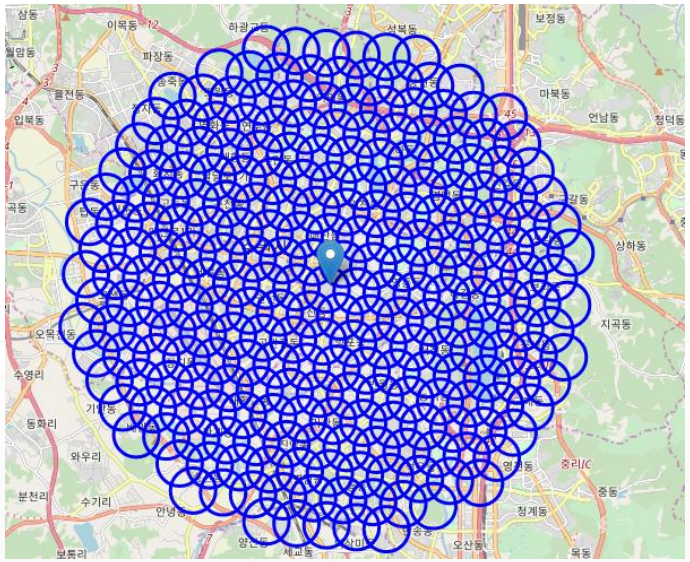
1) Samsung Digital City (SDC) Neighborhood data that includes address, latitude, longitude using **Google geocode API**.

2) Venues category data of all neighborhoods within 10km from SDC using **Foursquare API**.

4. Methodology

1) Create neighborhoods around Samsung Digital City
(code in reference to https://cocl.us/coursera_capstone_notebook).

- Get the latitude and longitude of SDC center point using Google geocode API.
- Create neighborhoods that are equally spaced, centered around SDC and within ~10km from SDC, and each neighborhood is defined as circular area with a radius 600m.
- Calculate the latitude and longitude of all neighborhoods



4. Methodology

2) Get the address of each neighborhood (of the central point of each circle)

- Get the address of all neighborhoods and save them into data frame.

```
df_locations.head(10)
```

	Address	Latitude	Longitude	X	Y	Distance from center
0	414 Songjuk-dong, Jangan-gu, Suwon, Gyeonggi-d...	37.299649	127.000938	6.530165e+06	1.290957e+07	9763.196198
1	906 Jeongja 2(i)-dong, Jangan-gu, Suwon, Gyeon...	37.293556	126.995879	6.531365e+06	1.290957e+07	9488.940931
2	111-11 Jeongja 2(i)-dong, Jangan-gu, Suwon, Gy...	37.287464	126.990822	6.532565e+06	1.290957e+07	9361.623791
3	256-6 Seodun-dong, Gwonseon-gu, Suwon, Gyeongg...	37.281373	126.985766	6.533765e+06	1.290957e+07	9387.225362
4	247-2 Seodun-dong, Gwonseon-gu, Suwon, Gyeongg...	37.275282	126.980712	6.534965e+06	1.290957e+07	9564.517761
5	67-2 Tap-dong, Gwonseon-gu, Suwon, Gyeonggi-do...	37.269192	126.975659	6.536165e+06	1.290957e+07	9885.342685
6	843 Jowon-dong, Jangan-gu, Suwon, Gyeonggi-do,...	37.305286	127.015136	6.528365e+06	1.291061e+07	9501.578816
7	893 Gyeongsu-daero, Jowon 1(il)-dong, Jangan-g...	37.299194	127.010073	6.529565e+06	1.291061e+07	8982.204629
8	69-6 Jeongja 2(i)-dong, Jangan-gu, Suwon, Gyeo...	37.293102	127.005013	6.530765e+06	1.291061e+07	8600.000000
9	344-65 Hwaseo-dong, Paldal-gu, Suwon, Gyeonggi...	37.287010	126.999953	6.531965e+06	1.291061e+07	8373.768566

4. Methodology

3) Get venue categories of all neighborhoods using Foursquare API

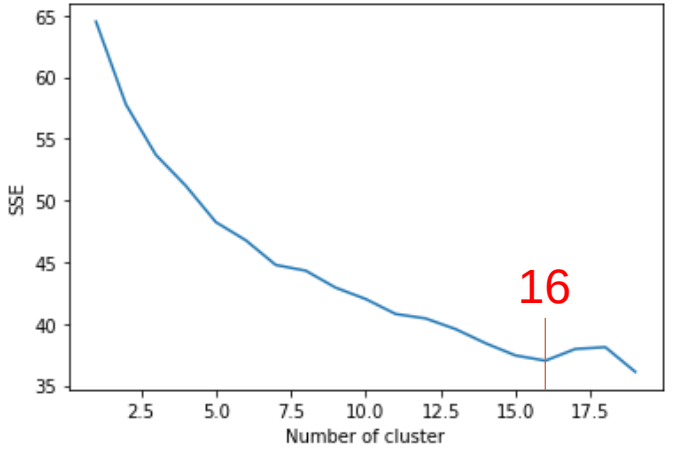
- Get the address of all neighborhoods and save them into data frame.

	Neighbourhood	Accessories Store	Airport Service	American Restaurant	Arcade	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	...	Train Station	Tunnel	Udon Restaurant
0	1 Maesanno 3(sam)-ga, Paldal-gu, Suwon, Gyeong...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	...	0.0	0.0	0.000000
1	1 Seocheon-dong, Giheung-gu, Yongin-si, Gyeong...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	...	0.0	0.0	0.000000
2	1-6 Seodun-dong, Gwonseon-gu, Suwon, Gyeonggi-...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	...	0.0	0.0	0.000000
3	10-78 Hwasan-dong, Hwaseong-si, Gyeonggi-do, S...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.166667	0.0	...	0.0	0.0	0.000000
4	1009-6 Yeongtong-dong, Yeongtong-gu, Suwon, Gy...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.022222	0.0	...	0.0	0.0	0.022222

4. Methodology

4) Make clusters based on the similarity of venue categories using K-Mean.

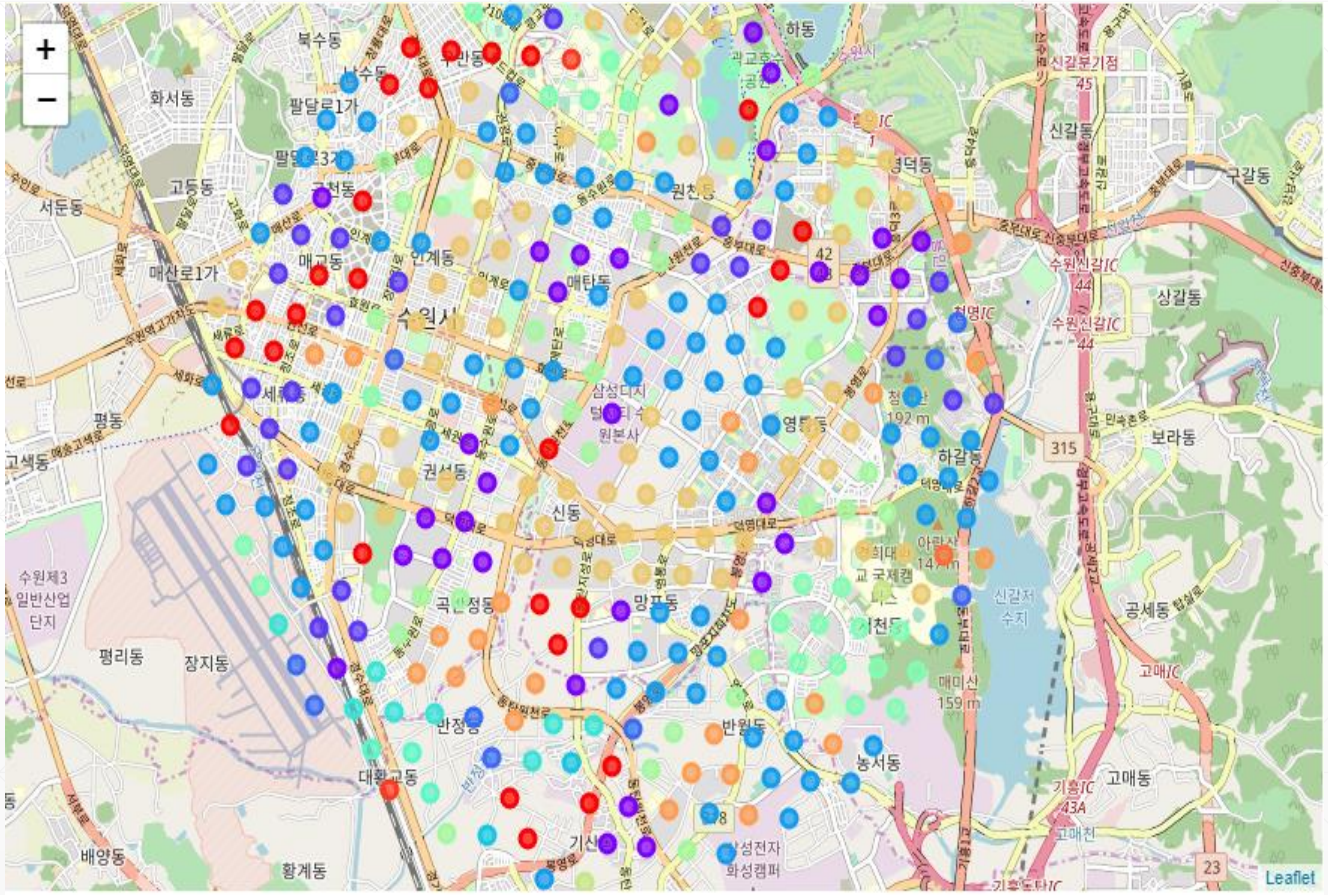
- Create the data frame that each neighborhood has top 10 most common venues
- Make the number of clusters from 1 to 20 and evaluate SSE.



- Create 16 clusters that has the lowest SSE and analyze venues of each cluster

5. Results and Discussion

- There are 241 neighborhoods and 169 venue categories (241X169) to be clustered.
- 241 neighborhoods are clustered into 16 groups based on the similarity of venue categories.



5. Results and Discussion

- It won't be smart to discuss all 16 clusters in detail. Instead, I made the summary table of the 1st common venue of 16 clusters to sense their characteristics (Highlighting the key venues of each cluster)

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Park				1				2				6				
Restaurant	24	6		11	1	3	11		6		20	2		2		
Desert(Bakery/Donut,,		2		1				3	2							
Bus Station		1		1				1								
Bus Stop				1								2				2
Café	3	1						18	1		1			6		
Concert/Museum				1							1					
Convenience Store									1					4		
Fried Chicken Joint	2	1						2	2				2			
Golf Course		1		1	8				1							
Gym/Yoga		1							2							
Historical Site		2							1							
Intersection	3	20							5					3		
Activity		2						3	4							
Store	1	4	2					1	6	1			1		1	

5. Results and Discussion

- Since it is the most important to have a park near the optimal residential area, I will select the cluster 11, which has “park” as the largest 1st common venue, for further analysis.

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Park				**1**				**2**				**6**				
Restaurant	24	6		11	1	3	11		6		20	2		2		
Desert(Bakery/Donut,,		2		1				3	2							
Bus Station		1		1				1								
Bus Stop				1								2				2
Café	3	1						18	1		1			6		
Concert/Museum				1							1					
Convenience Store									1					4		
Fried Chicken Joint	2	1						2	2				2			
Golf Course		1		1	8				1							
Gym/Yoga		1							2							
Historical Site		2							1							
Intersection	3	20							5					3		
Activity		2						3	4							
Store	1	4	2					1	6	1			1		1	

5. Results and Discussion

- Cluster 11 has 10 neighborhoods with parks in common. Now let's take a look at each neighborhood to see if there are any positive or negative venues for the residential option (List in near from SDC center order).

positive negative

	Address	Distance from center	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
125	441-4 Sin-dong, Yeongtong-gu, Suwon, Gyeonggi-...	200.000000	11.0	Park	Food	Korean Restaurant	Yoga Studio	Electronics Store	Food & Drink Shop	Flower Shop	Fishing Store	Fast Food Restaurant	Farmers Market
189	164 Seonong-dong, Giheung-gu, Yongin-si, Gyeon...	4161.730409	11.0	Park	Café	Dance Studio	Hotel	Juice Bar	Asian Restaurant	Chinese Restaurant	Korean Restaurant	Supermarket	Electronics Store
112	445-9 Gokbanjeong-dong, Gwonseon-gu, Suwon, Gy...	4521.061822	11.0	Bus Stop	Park	Outdoor Supply Store	Korean Restaurant	Food & Drink Shop	Food	Flower Shop	Fishing Store	Fast Food Restaurant	Farmers Market
119	57-7 Ha-dong, Yeongtong-gu, Suwon, Gyeonggi-do...	7000.000000	11.0	Park	Reservoir	Pet Café	Scenic Lookout	Gym / Fitness Center	Donut Shop	Flower Shop	Fishing Store	Fast Food Restaurant	Farmers Market

5. Results and Discussion

- Bus Stop or Metro Station is also very Important for better residential area. Also Farmer markets, Café, Yoga studio/Gym are good places to have.

21	199-155 Godeung-dong, Paldal-gu, Suwon, Gyeong...	7318.469785	11.0	Park	Bridal Shop	Korean Restaurant	Yoga Studio	Food & Drink Shop	Food	Flower Shop	Fishing Store	Fast Food Restaurant	Farmers Market
195	486-3 Byeongjeom 1(il)-dong, Hwaseong-si, Gyeo...	8487.638070	11.0	Bus Stop	Dessert Shop	Korean Restaurant	Park	Electronics Store	Food & Drink Shop	Food	Flower Shop	Fishing Store	Fast Food Restaurant
40	648 Iui-dong, Yeongtong-gu, Suwon, Gyeonggi-do...	9206.519429	11.0	Park	BBQ Joint	Steakhouse	Korean Restaurant	Food & Drink Shop	Food	Flower Shop	Fishing Store	Fast Food Restaurant	Farmers Market
3	256-6 Seodun-dong, Gwonseon-gu, Suwon, Gyeongg...	9387.225362	11.0	Park	Café	Metro Station	Yoga Studio	Food & Drink Shop	Food	Flower Shop	Fishing Store	Fast Food Restaurant	Farmers Market
0	414 Songjuk-dong, Jangan-gu, Suwon, Gyeonggi-d...	9763.196198	11.0	Korean Restaurant	Fried Chicken Joint	Park	Electronics Store	Food & Drink Shop	Food	Flower Shop	Fishing Store	Fast Food Restaurant	Farmers Market
26	912-10 Gosaek-dong, Gwonseon-gu, Suwon, Gyeong...	9957.911428	11.0	Korean Restaurant	Gym / Fitness Center	Café	Park	Yoga Studio	Food	Flower Shop	Fishing Store	Fast Food Restaurant	Farmers Market

5. Results and Discussion

- I think the cluster 11 has a lot of good house locations and I like to recommend the different locations considering the possible different life style of target stakeholders.

A) Life Convenience : 112 (445-9 Gokbanjeong-dong), 195 (486-3,Byeongjeom), 3 (256-6,sodun)

- All have a public transportation (bus stop or metro station) nearby
- All have restaurants, shops, market nearby

※ If you find the nearest location, 112 would be the best option.

B) Nature & Pet friendly : 119 (57-7, Ha-dong)

- It has a scenic lookout, reservoir, and pet café nearby

6. Conclusion

I like to summarize the report as below

- Target : Recommend the optimal residential area to the employees who plan to move near the Office of Samsung Digital City, Suwon, South Korea
- Process
  1) Data gathering using Google geocode and Foursquare API
  2) Create Dataset that has neighborhood(address), venue categories
  3) Cluster using K-Mean clustering, one of machine learning algorithm
  4) Find the appropriate cluster for the residential area
  5) Analyze each neighborhood in the cluster and find the recommendable neighborhoods
- Results : Three neighborhoods are good for the convenient life and one is good for the nature friendly life.