Nocturnal vs. Diurnal: A Comparison of Land Use by Type of Commercial Districts

Dawon Oh

shfam6@snu.ac.kr

Contents

1. Intro

- Jacobs (1961) argued that the diversity of physical environments, such as mixed use, is necessary to create a vibrant space
- The heterogenous visiting pattern is common in commercial areas: peak hours in daytime and nighttime.

1. Classify commercial districts whether the population is concentrated during the daytime or nightime
2. Analyzing which land use affects the difference
3. Find out which factors drive a larger number of visitors

2. Literature Review

Ⅰ Relationship between land use and travel behavior

Cervero·Duncan(2003): Land-use mix with commercial use has a positive effect on individual travel.

Frank et al.(2008): Mixed land use increases both commuting and non-commuting traffic.

Ⅰ Research on travel patterns using location-based big data

Sofiane et al.(2018) , Tao et al.(2014), Juan et al.(2018), Enwei et al.(2018):
When extraordinary events such as rallies or heavy rain are excluded, a certain type of time-series pattern appears in individuals' travel behavior.

Shan et al.(2012): Distinguish time-series patterns of individual travel behavior in Chicago using K-means cluster analysis.

Differentiation from previous research

- Previous studies have not been able to define time-series travel pattern due to data limitations.
- Construct a model that reflects the actual land use situation of each building.

The geographical scope of this research covers 251 commercial districts located in the city of Seoul; commercial districts are places where several geographically adjacent stores are concentrated.

The data were collected for 730 days from January 1, 2017 to December 31, 2018.

Data

1) "Seoul Living Population" (SLP)

Data regarding the number of people in each of the 19,153 Census Tally Districts(CTD)s in Seoul by sex and age (5-year interval, aged between 10 to 79, under 9, and over 80) collected on a daily basis.

It is based on mobile signal data: The data are recorded at the point of time when a mobile phone moves from a base station to another, or when it uses data, regardless of whether its user is using the mobile phone or not.

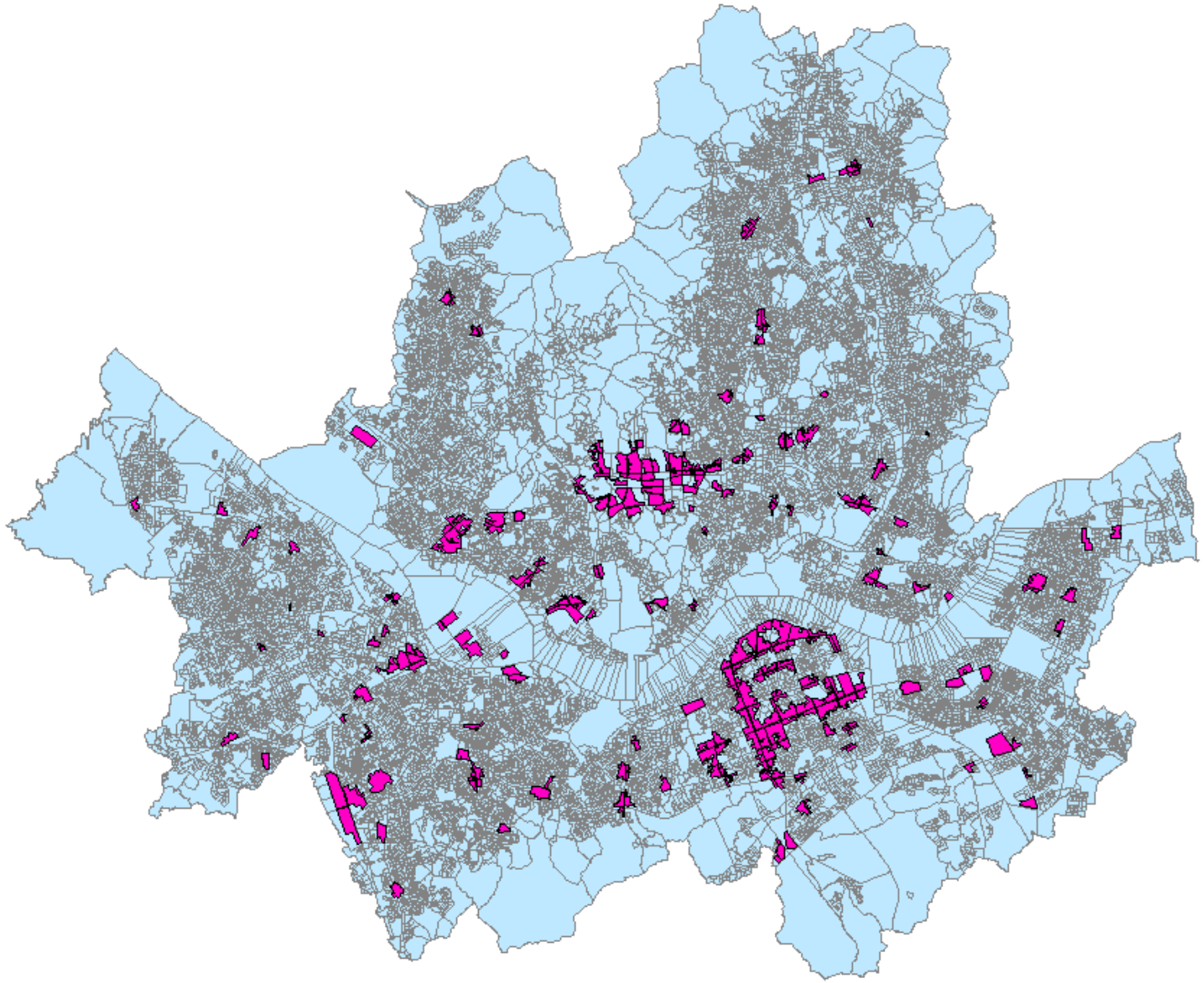
2) Building DB (a summary of the current status of buildings throughout Seoul – total 654,595 buildings)

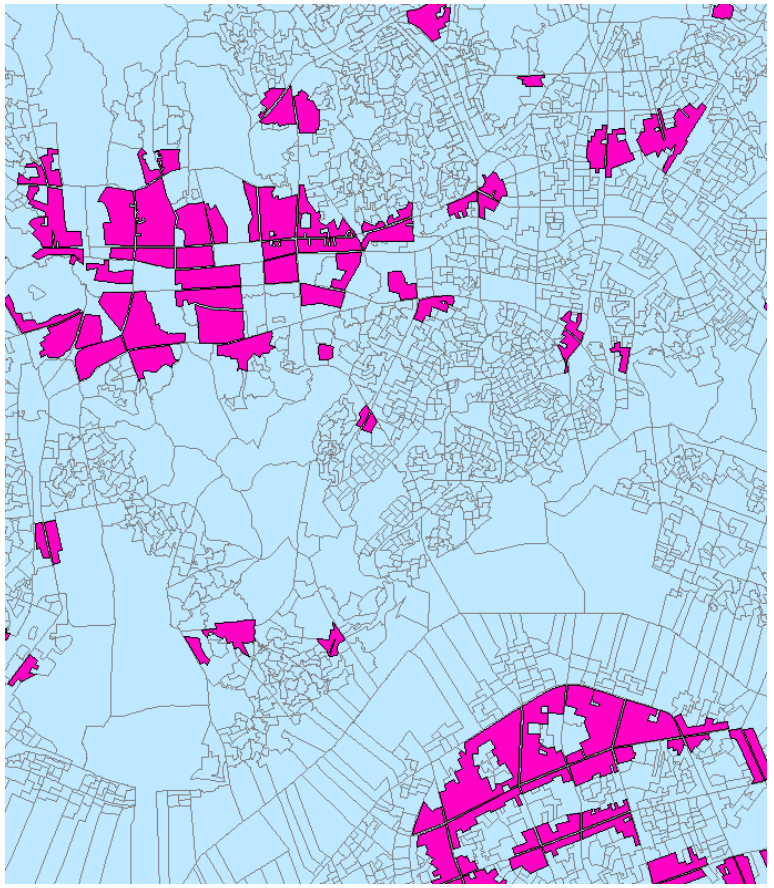
It contains information such as each building's use, year of construction, total area, floor area.

* SLP is a different measure than floating population as it represents the de facto population at the time of measurement. Therefore, it is not identical to census data, and has a limitation in that it cannot reveal which area the individual has moved to.

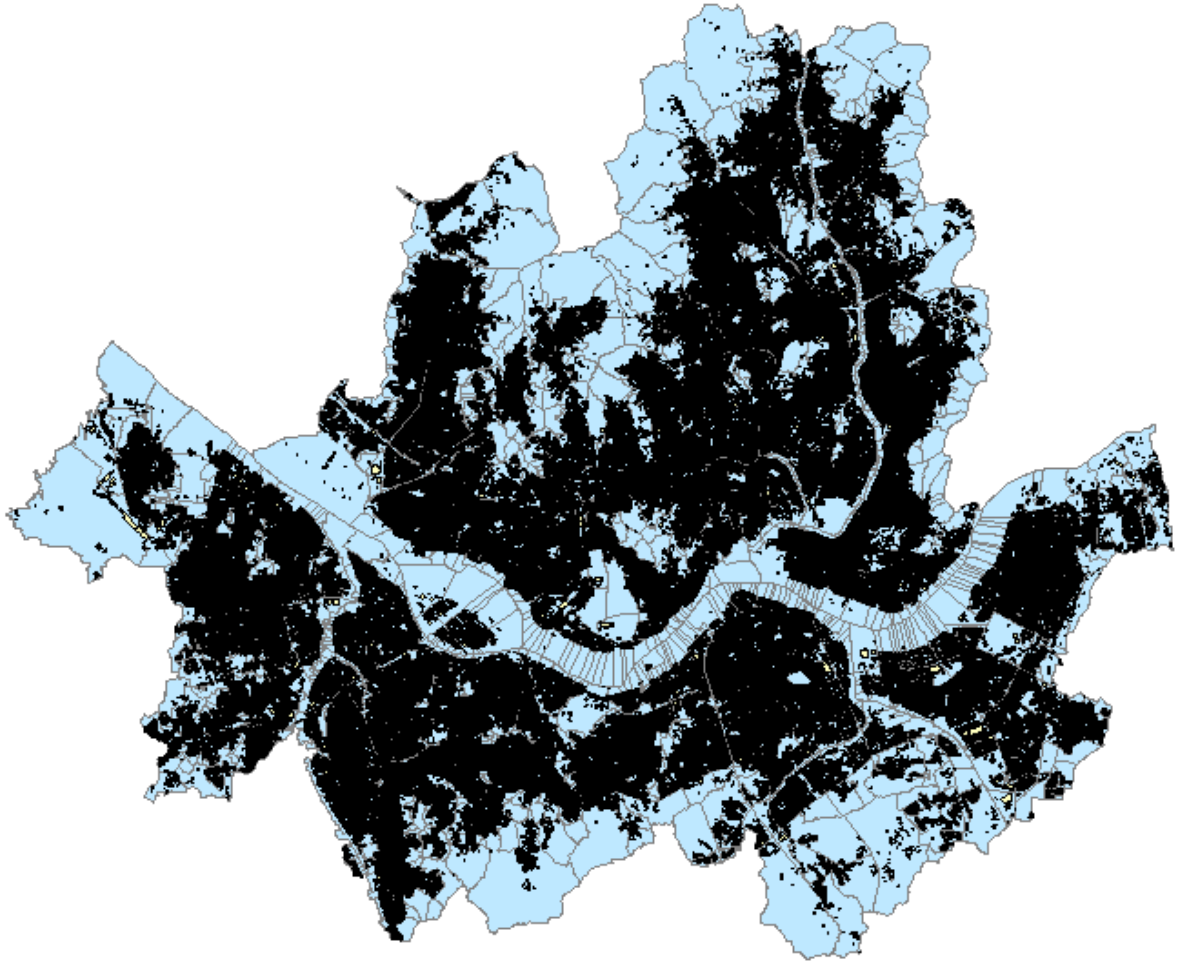
Example of “Seoul Living Population” (SLP)

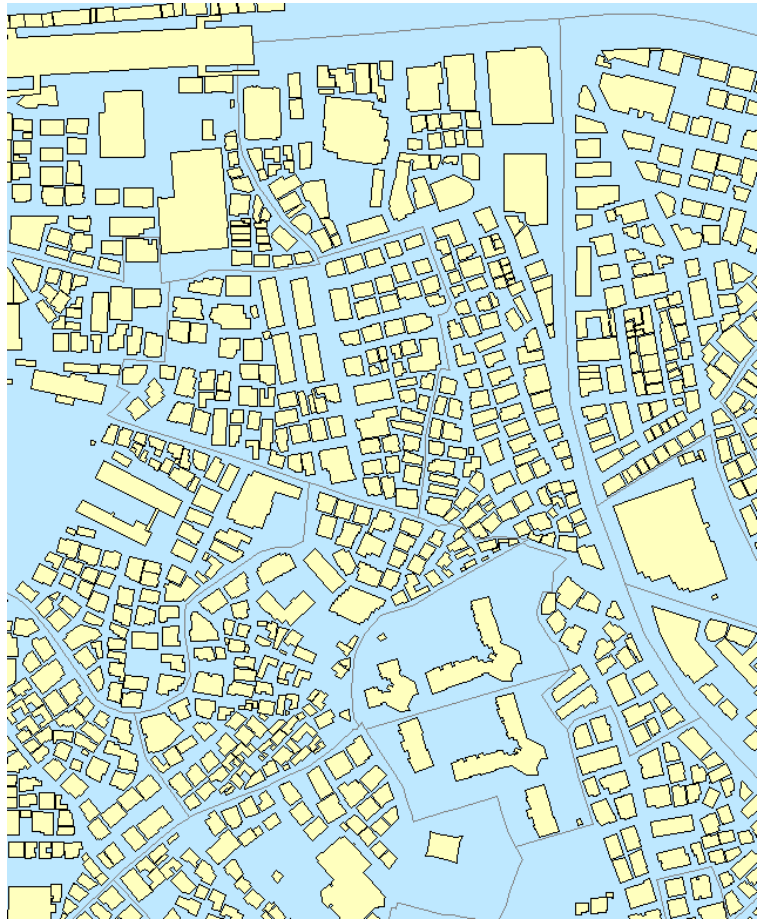
	A	B	C	D	E	F	G	H	I	J	K	L	M
1	DATE_ID	TIME	ADMINISTRATIVE_UNIT	CENSUS_TALLY_CODE	TOTAL_POPULATION	MALE_AGE_0~9	FEMALE_AGE_0~9	MALE_AGE_10~14	FEMALE_AGE_10~14	MALE_AGE_15~19	FEMALE_AGE_15~19	MALE_AGE_20~24	FEMALE_AGE_20~24
2	20180101	0	11260610	1107062080001	272	*	*	*	*	4	5	6	9
3	20180101	1	11215710	1105053030006	19	*	*	*	*	*	*	*	*
4	20180101	2	11215770	1105058010003	557	8	10	4	5	16	16	20	21
5	20180101	3	11500641	1116072010205	361	10	13	5	6	5	6	9	9
6	20180101	4	11305645	1109074010012	1304	22	33	11	16	25	21	44	32
7	20180101	5	11440710	1114071010001	407	6	10	*	4	7	4	14	20
8	20180101	6	11560560	1119056041202	312	19	8	7	*	6	6	8	7
9	20180101	7	11710646	1124081020102	731	52	38	27	19	19	20	18	19
10	20180101	8	11110700	1101070010006	1105	25	26	12	13	25	21	38	41
11	20180101	9	11380570	1112058020002	3975	83	84	38	43	88	85	137	139
12	20180101	10	11200720	1104069010014	344	10	*	5	*	6	6	15	18
13	20180101	11	11290715	1108071030004	360	25	15	13	8	9	10	10	7
14	20180101	12	11170570	1103057010005	455	6	8	*	7	18	14	10	14
15	20180101	13	11620525	1121052010001	211	12	21	5	9	4	*	*	8
16	20180101	14	11680565	1123078051701	66	*	*	*	*	*	*	*	*
17	20180101	15	11500611	1116067010005	319	19	19	10	11	16	10	6	13
18	20180101	16	11350695	1111077070203	93	*	*	*	*	*	*	*	*
19	20180101	17	11410710	1113071020116	568	16	23	9	13	16	18	18	13
20	20180101	18	11320515	1110063010009	17	*	*	*	*	*	*	*	*
21	20180101	19	11620525	1121052020003	481	10	16	4	7	12	7	15	20
22	20180101	20	11530550	1117055011401	111	*	*	*	*	*	*	*	5
23	20180101	21	11500641	1116072010502	308	11	10	5	4	6	*	10	5
24	20180101	22	11380625	1112074030033	252	7	6	*	*	4	6	5	9
25	20180101	23	11215750	1105056010004	361	5	8	*	4	10	13	8	11
26	20180102	0	11170640	1103064010001	646	6	31	*	14	14	14	7	17
27	20180102	1	11545620	1118053020017	401	9	10	5	6	8	7	10	9
28	20180102	2	11320522	1110065020007	631	13	8	5	*	15	14	15	18
29	20180102	3	11380690	1112071040313	535	25	16	16	10	13	15	9	12
30	20180102	4	11500510	1116051020004	2103	74	48	38	28	49	40	84	107
31	20180102	5	11440710	1114071010002	402	5	11	*	5	11	7	16	13
32	20180102	6	11470611	1115063010004	1784	83	108	43	52	43	43	48	48
33	20180102	7	11200670	1104067040201	198	5	*	*	*	4	5	5	4
34	20180102	8	11320521	1110064010001	4120	65	43	35	20	110	35	113	87
35	20180102	9	11500641	1116072020208	139	*	8	*	4	*	*	*	*
36	20180102	10	11260540	1107054020007	664	18	14	9	6	20	20	31	23
37	20180102	11	11350695	1111077060102	449	18	21	9	10	15	11	14	10
38	20180102	12	11500611	1116067012403	124	6	9	*	5	*	*	*	*
39	20180102	13	11620525	1121052010002	2948	94	132	41	58	48	62	71	92
40	20180102	14	11320514	1110062020109	183	5	6	*	4	5	*	6	6
41	20180102	15	11500510	1116051021001	68	*	*	*	*	*	*	*	*



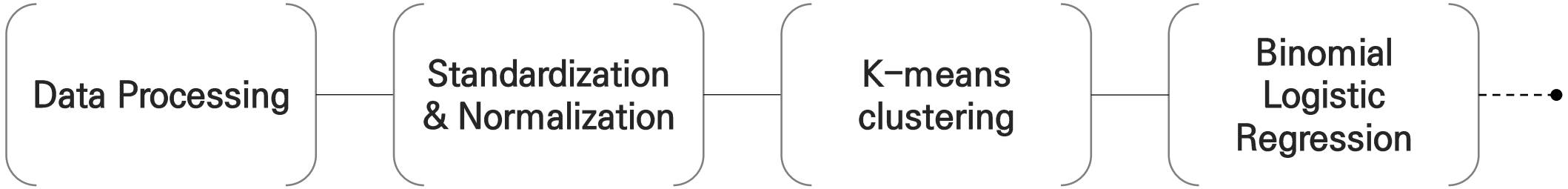


Commercial districts in Seoul





Buildings in Seoul


Data Processing
Standardization & Normalization
K-means clustering
Binomial Logistic Regression

Data Aggregation

The total number of people present in each of the 19,153 CTDs were divided hourly from 0 to 23 o'clock, and the number of people present in each time zone were counted.

(The population aged 0 to 19 years and above 70 was excluded.*)

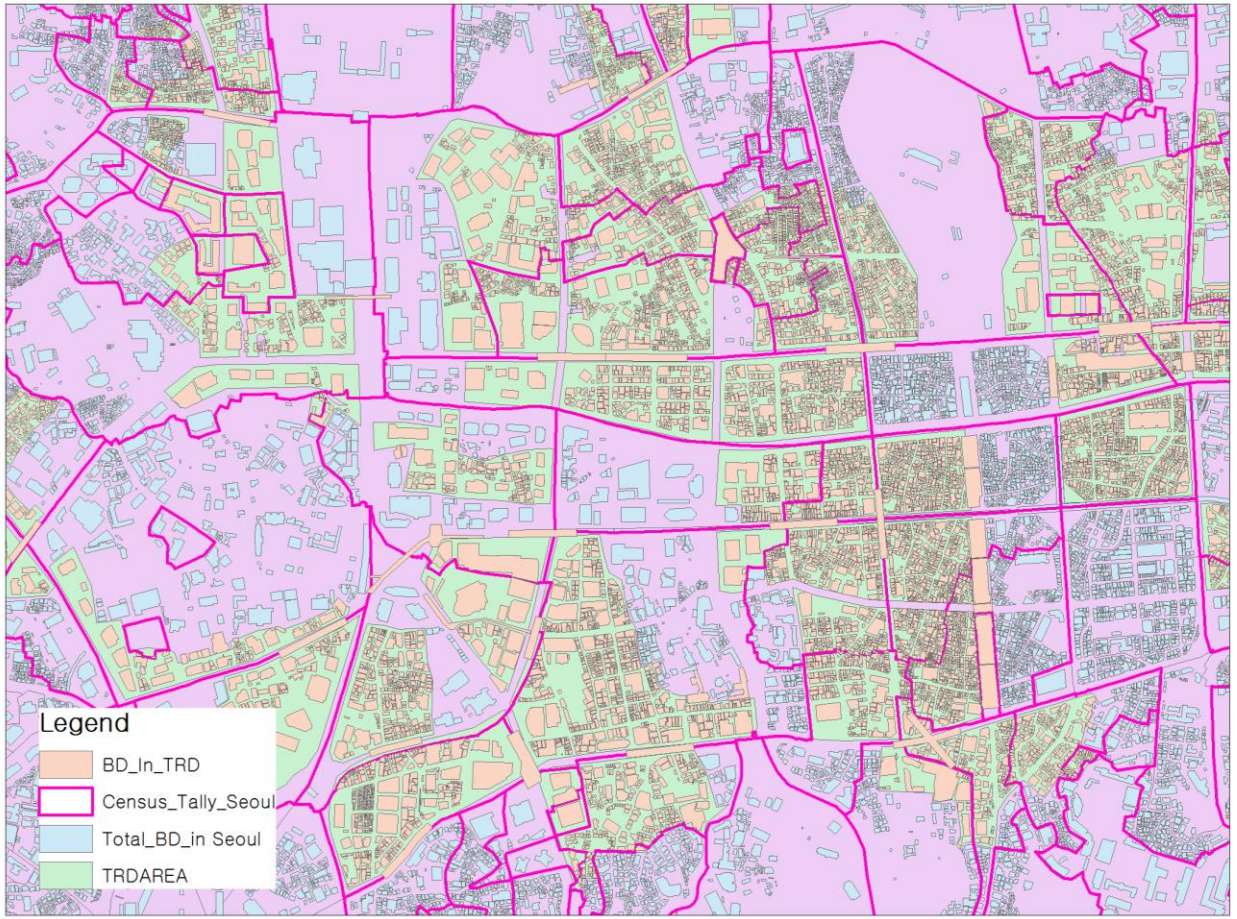
GIS – Spatial Join

1. join feature: 'have their center in'
Only buildings in which the center of the building is located in the relevant commercial area are extracted as a separate layer.

2. join feature: 'Intersect'
How the names of commercial districts in buildings closest to the center of the CTDs are combined

The inconsistency between the boundaries of CTDs and commercial districts was solved by determining which CTDs should be included or not. To avoid double counting when several commercial districts are included in a large CTD, we excluded 8 commercial districts, including only 243 commercial districts in the analysis.

* Ther is a strong tendency that age group over 70 are not fully equipped with mobile phones and age group under 19 are mostly in school during the daytime .



In order to obtain the annual average value by commercial district, each value is divided by 730.
(Using average value to exclude the influence of seasonal changes or one-time events such as parades, festivals, or rallies.)

$$X_{at} = \frac{\sum_{n=1}^{k} X_n}{k}$$

X^{at}: Average value of SLP for CTD 'a' and time zone 't'
k : 730 (sum of total days, from 01/01/2017-12/31/2018)

Standardization & Normalization

$$Z_{at} = \frac{X_{at} - A_a}{S_a}$$

$$Z'_{at} = \left(\frac{X_{at} - Min}{Max - Min}\right) * 2 + Z_{at}$$

Z_{at}: Standardized value of SLP for CTD 'a' and time zone 't'
X^{at}: Average value of SLP for CTD 'a' and time zone 't'
A_a : Total average value of SLP for CTD 'a'
S_a : Standard deviation of SLP for CTD 'a'
Z'_{at}: Standardized normalized value of SLP for CTD 'a' and time zone 't'
Max: Maximum value of the averages for all CTDs and time zones
Min : Minimum value of the averages for all CTDs and time zones

* Z-SCORE standardization
* Normalization

(In the second equation, the min-max scaling multiplied by 2 gives a deviation of the maximum and minimum values of each commercial district's volume by ± 1 standard deviation. This roughly corresponds to the deviation between the upper and lower 15%.)

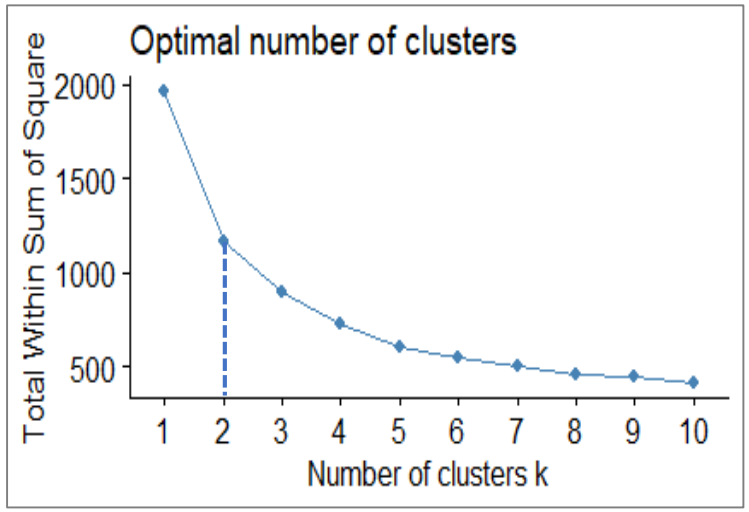
To partially reflect the difference in the volume of the absolute value in the analysis,
excluding the effect of the difference in size between individuals on the clustering process.

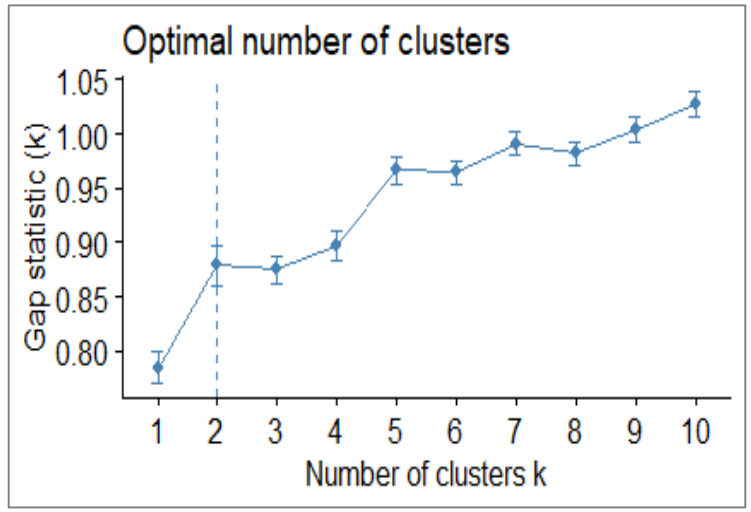
Nocturnal vs. Diurnal

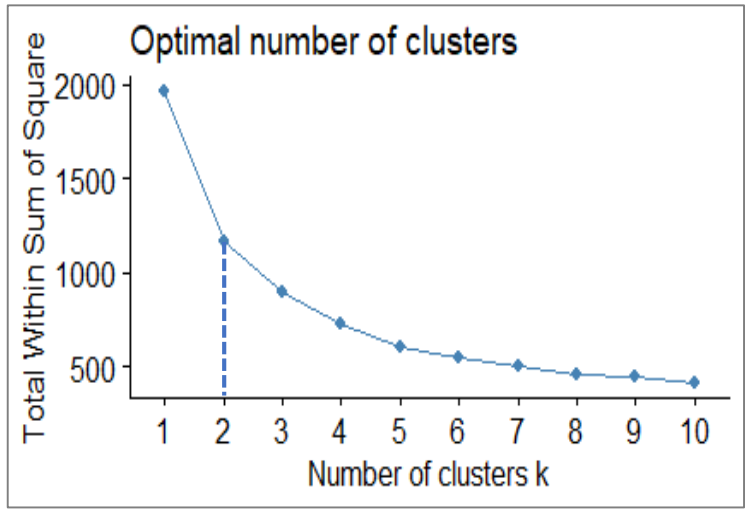
K-means Clustering

1. Weekdays
2. Weekend

K-means Clustering – Weekdays

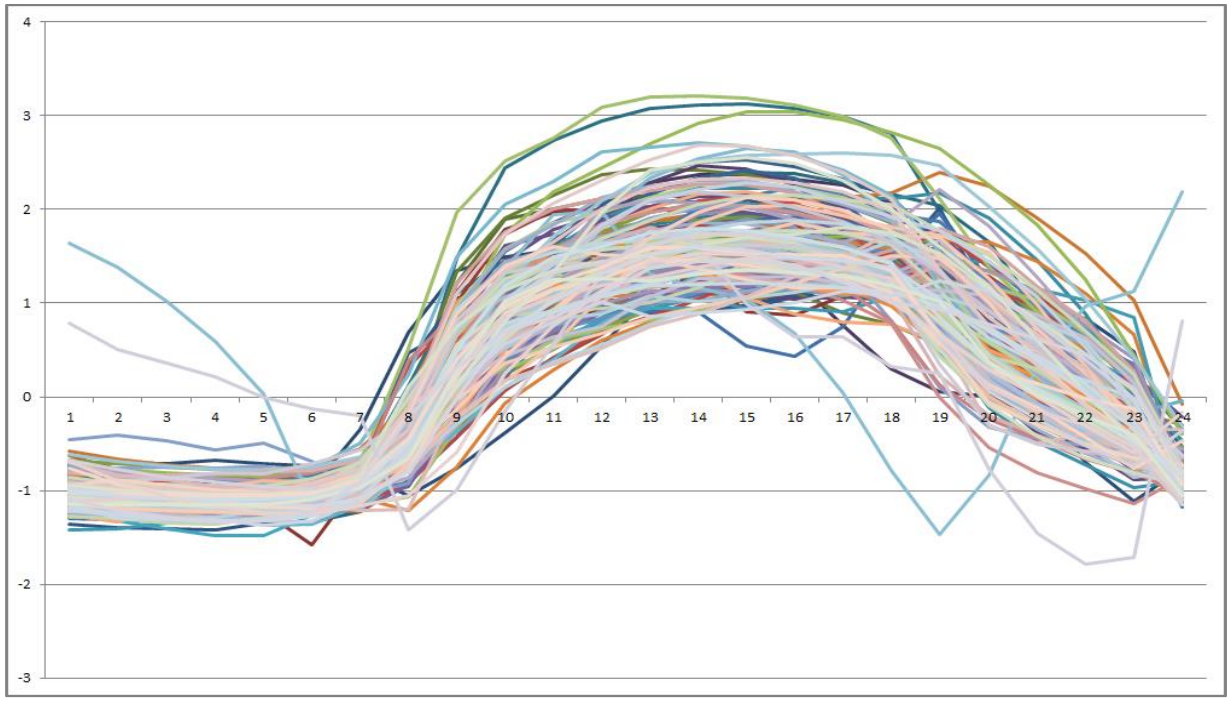






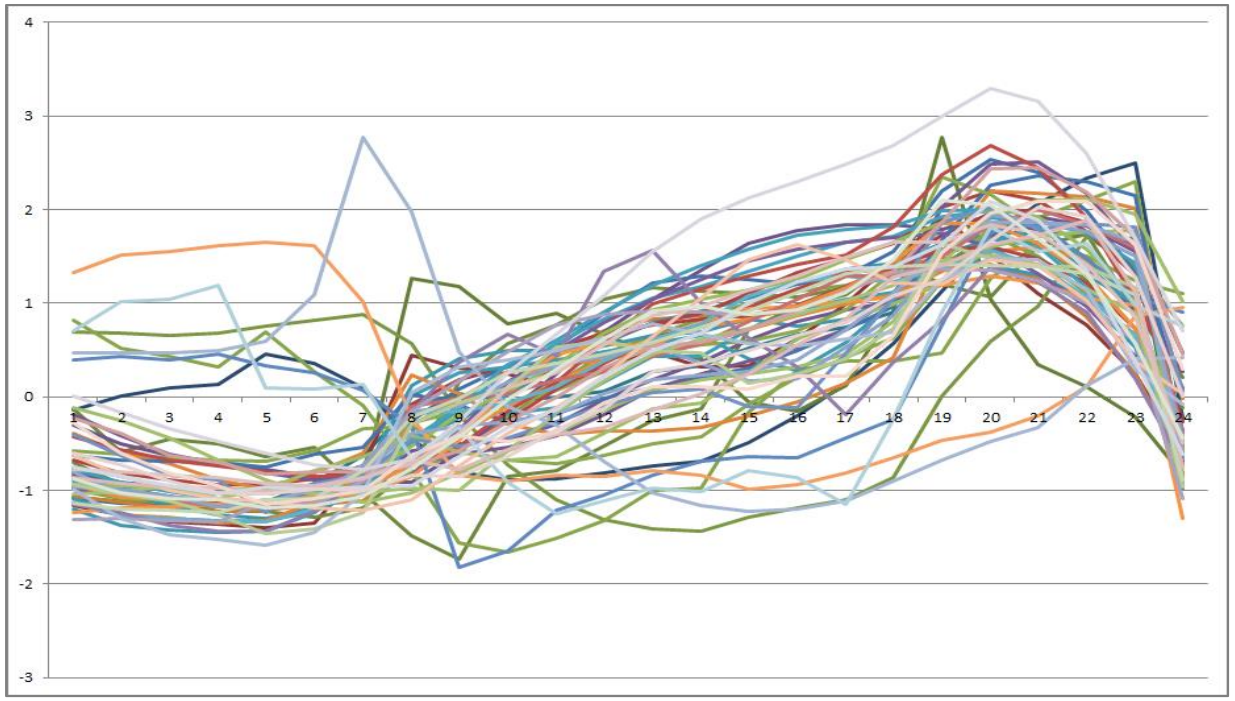
Number of optimal clusters: 2

K-means Clustering – Weekdays



Diurnal Commercial Districts (184)

Peak hours between 12 p.m. ~ 6 p.m.

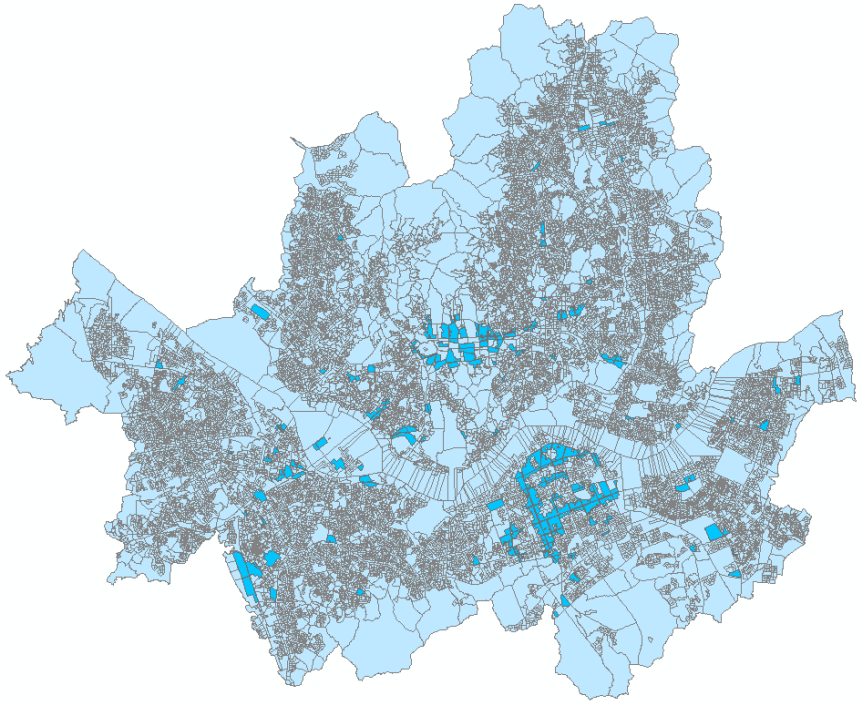


Nocturnal Commercial Districts (59)

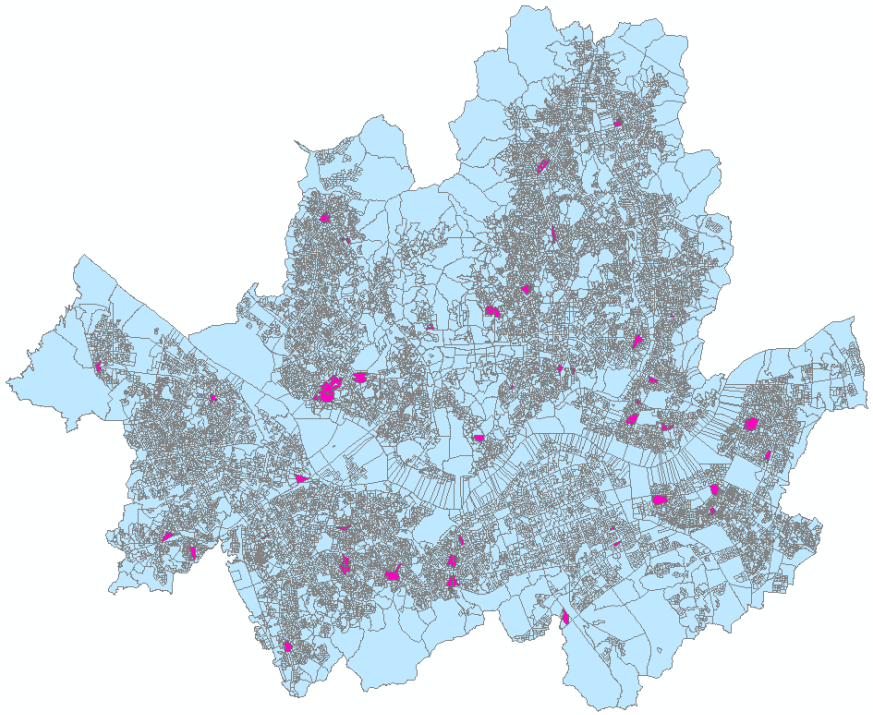
Peak hours between 6 p.m. ~ 12 a.m.

Weekdays

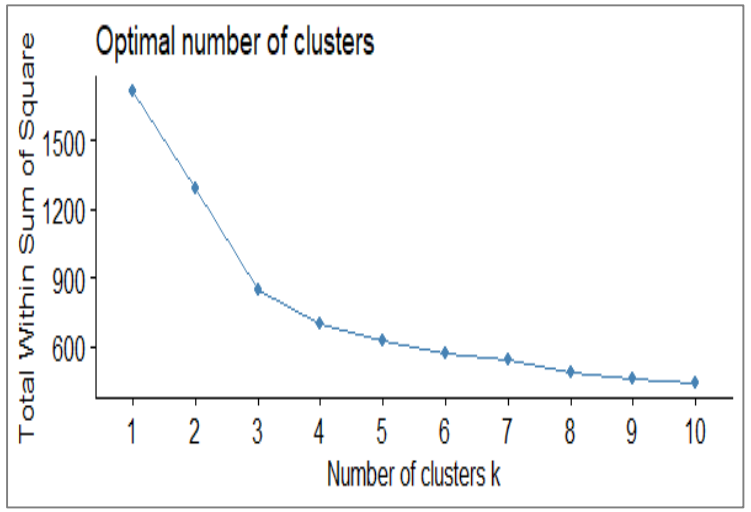
Diurnal Commercial Districts



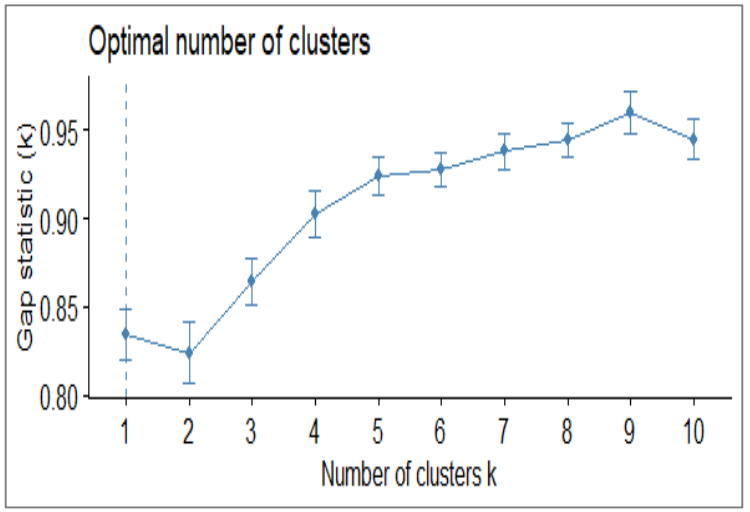
Nocturnal Commercial Districts



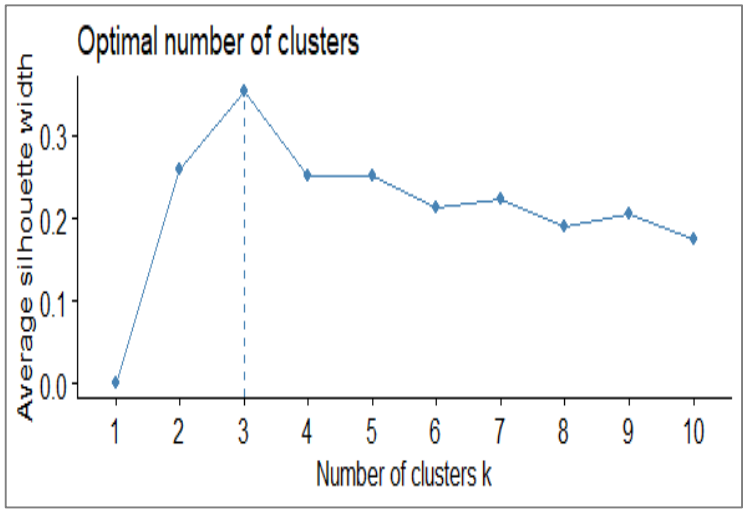
K-means Clustering – Weekends



WSS(The Elbow Method)



Gap Statistic

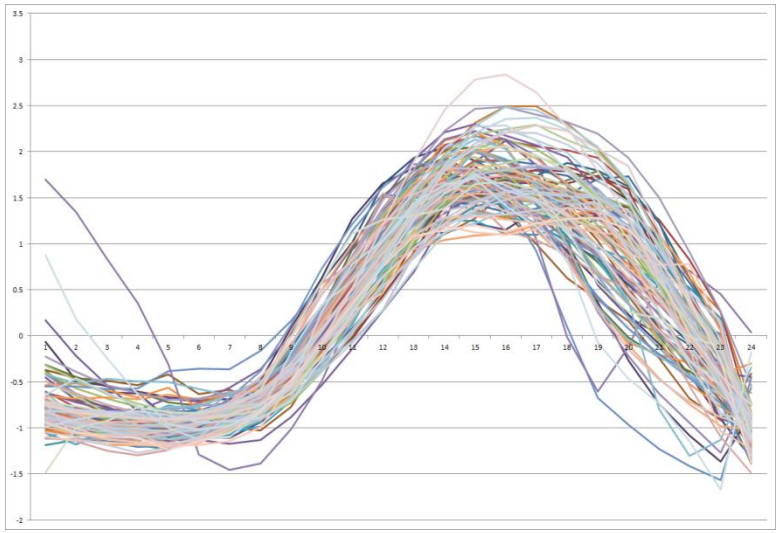


Silhouette

Number of optimal clusters: 2 or 3

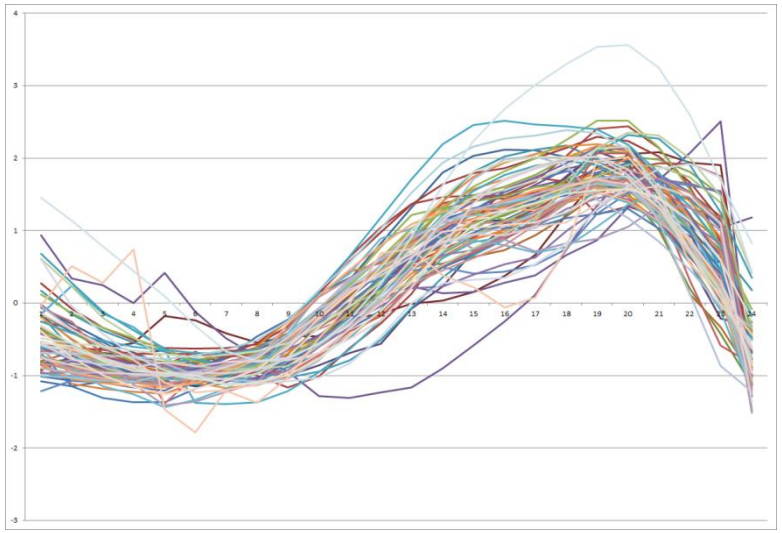
(need to be verified with graphs)

K-means Clustering – Weekends



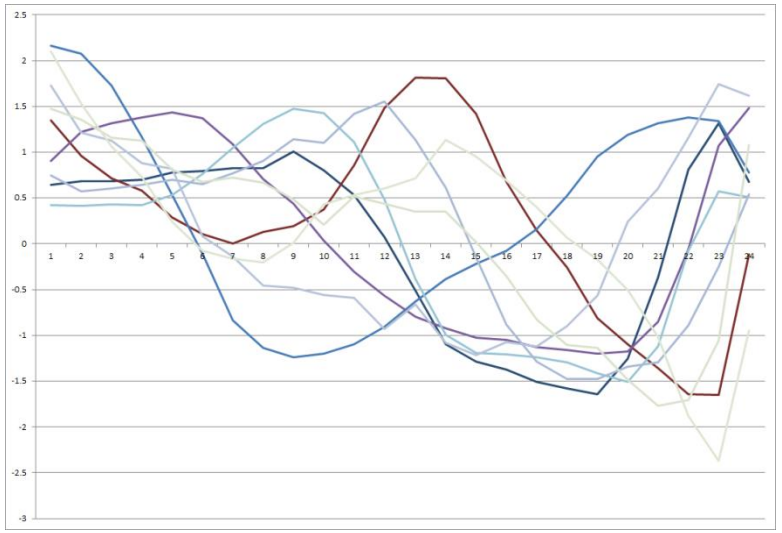
Diurnal Commercial Districts (160)

Peak hours between 12 p.m. ~ 6 p.m.



Nocturnal Commercial Districts (74)

Peak hours between 6 p.m. ~ 12 a.m.

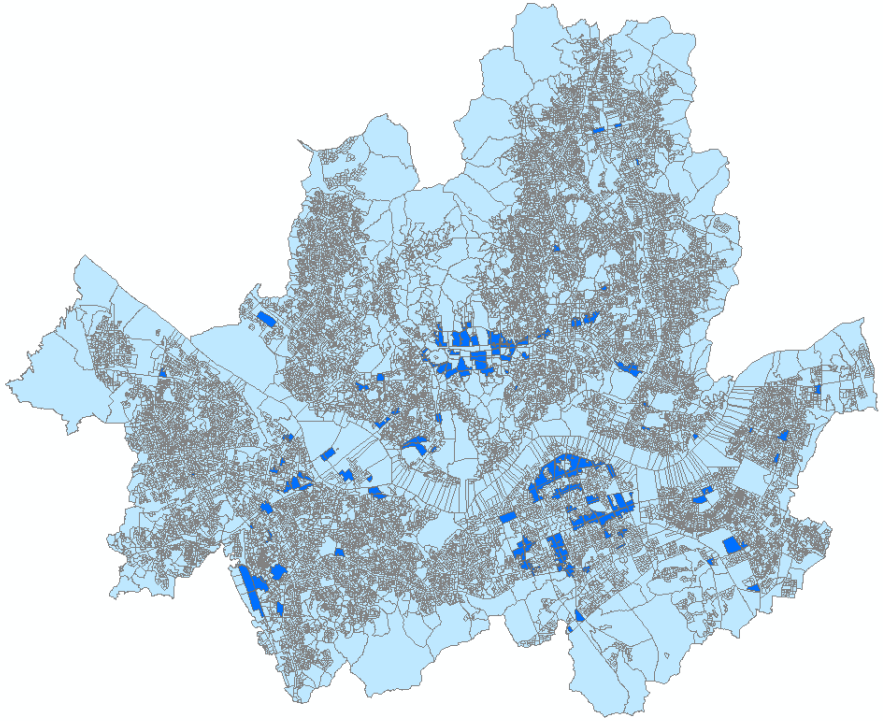


Outliers (9)

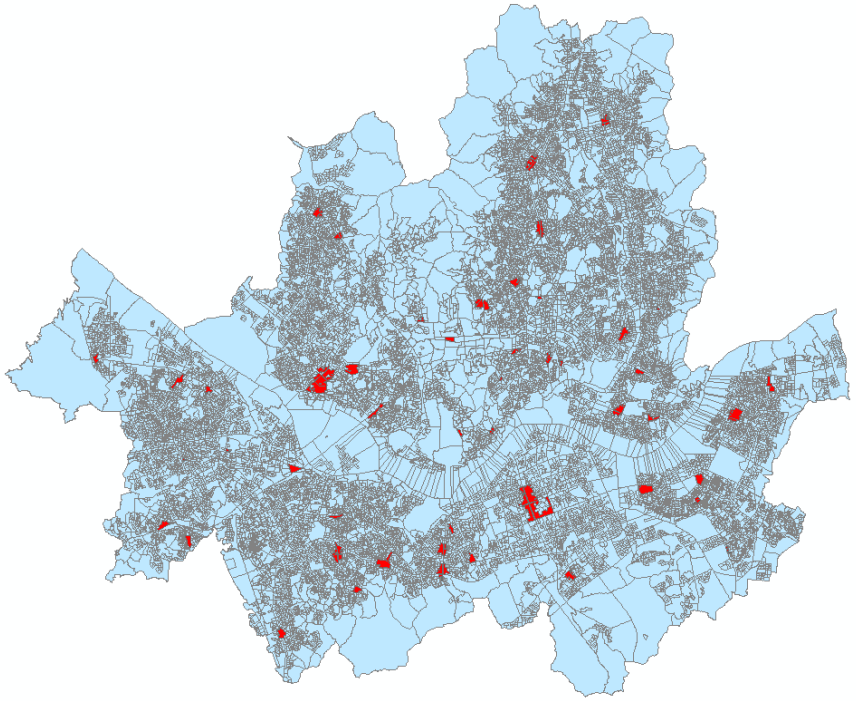
Peak hours between 12 a.m. ~ 12 p.m.

Weekends

Diurnal Commercial Districts



Nocturnal Commercial Districts



Model Description & Result

Binomial Logistic Regression

1. Weekdays
2. Weekends

Land use The ratio by each use to the total floor area

Category	Land use
1	Residential
2	Retail/ service
3	Restauran·pub
4	Culture/ recreation
5	Accommodation/ tourism
6	Offices/ small complex
7	Medical care
8	Sports
9	Traditional market
10	Education/ social service
11	Terminal (Transportation)
12	Etc.

Land Use Mix (LUM) Entropy Index

(0≤ LUM ≤1, The closer the value of LUM to 1, the higher the complexity)

$$LUM = -\sum P_U \ln(P_U)/\ln(n)$$

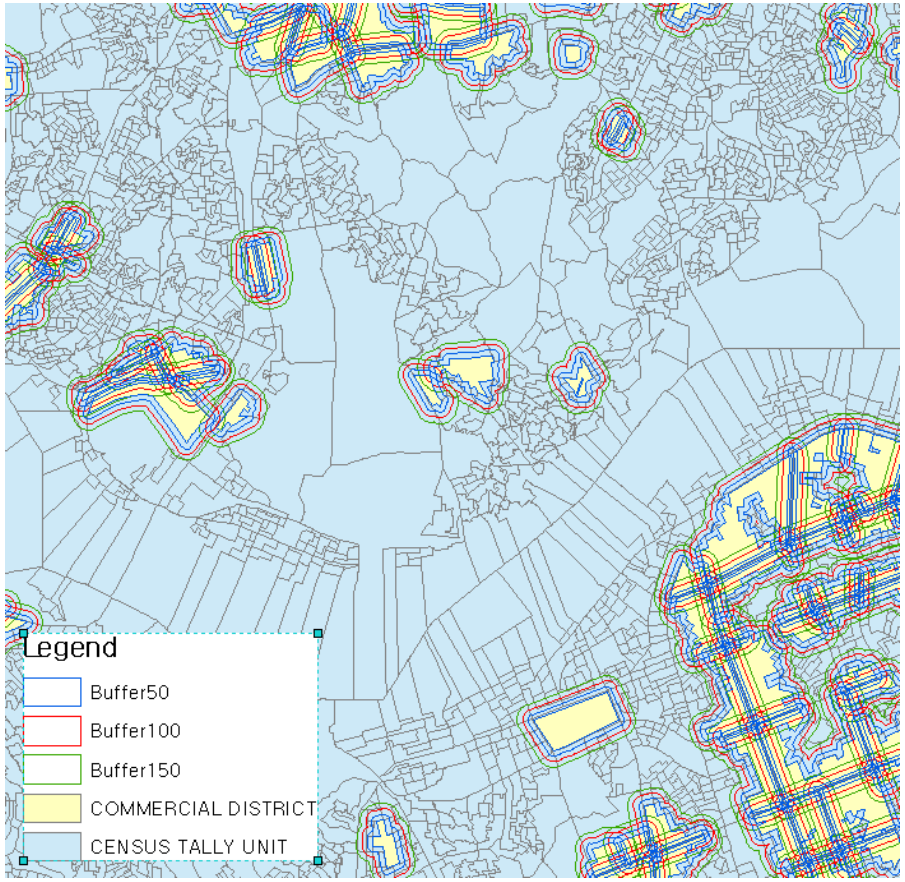
P_U: Ratio of total floor area of land use type 'U' (calculated based on the use classification on left)

n: Number of land uses in the unit (total: 12)

* Lee·Moundon (2006) :

Define the mixed types based on the number of facilities and the total floor area

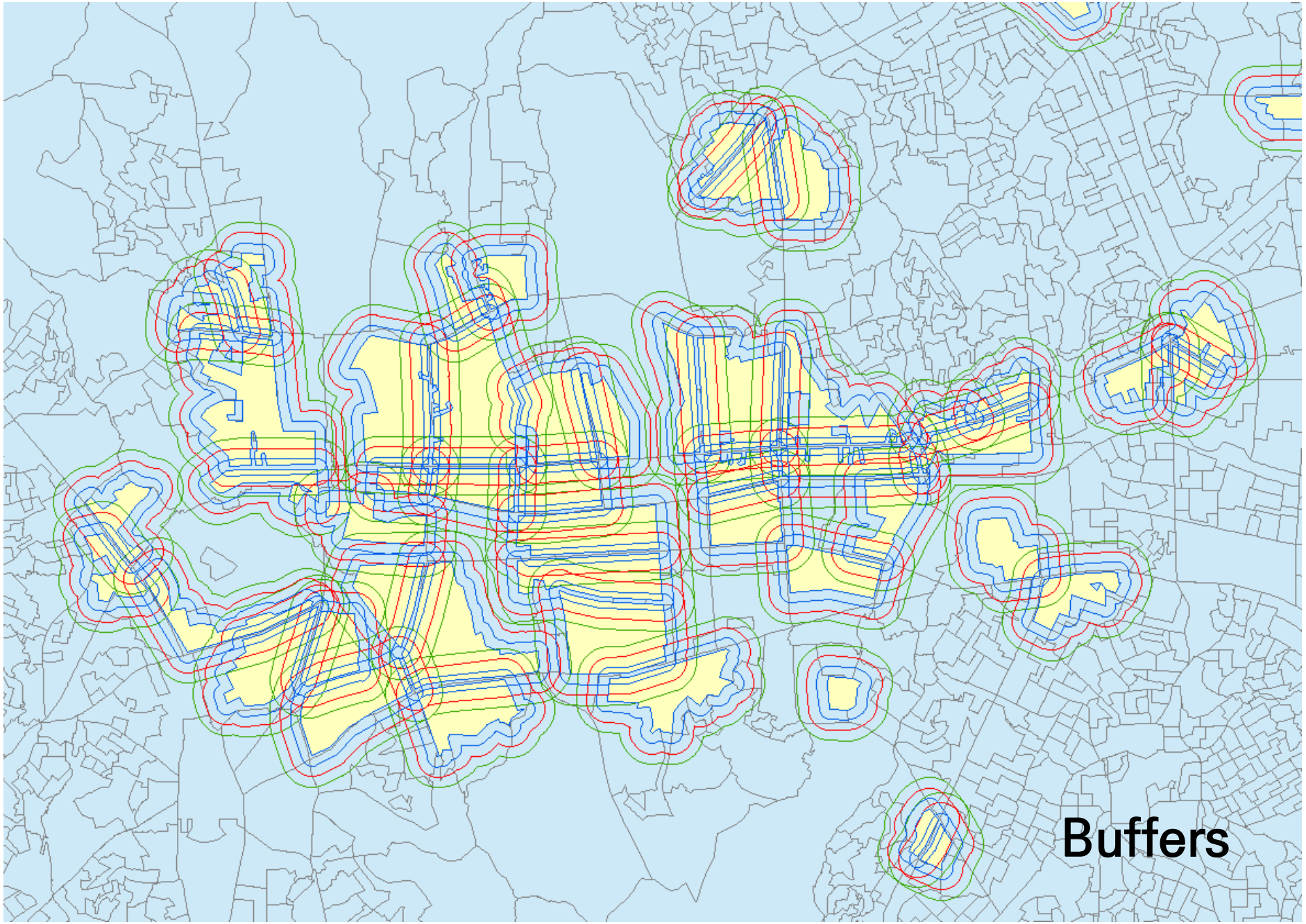
3. Result



Binomial Logistic Regression

Use of dummy-type dependent variable (1 night, 0 day)

Model				Weekday		Weekend	
Dependent variable				Nocturnal=1	Diurnal=0	Nocturnal=1	Diurnal=0
Independent Variable	Inner commercial District			Size of commercial district (m²)			
				LUM entropy index			
				Land use ratio (ratio of each of the 12 land use categories)			
				Ratio of people in their 20s and 30s (by total population)			
	50m Radius	Adjacent area (Excluding the inner area)		LUM entropy index			
				Land use ratio			
	100m Radius			Number of residents per unit area (m²)			
				Number of employees per unit area (m²)			
	150m Radius			Number of businesses per unit area (m²)			
				Average number of employees per business			


Buffers

주중		Commercial area	Commercial area Inc. 50m buffer	Commercial area Inc. 100m buffer	Commercial area Inc. 150m buffer
		Coef.	Coef.	Coef.	Coef.
상권 내부	LUM	6.016***	7.745**	4.581*	4.816**
	Residential	20.723***	13.910***	8.805***	12.267***
	Retail & service	18.686***	10.876***	5.720**	10.367**
	Restaurant·pub	28.015***	22.911**	23.755***	20.241**
	Accommodation & tourism	20.353***	18.180***	8.634**	11.392**
	Culture & recreation	19.428**	23.829**	19.816**	
	Medical care	26.256**	20.614**		20.988**
	Traditional market	18.560**			
상권 외부	LUM		7.414**		
	Medical care		−36.438***		
	Education & social service			−10.282**	
	Restaurant·pub			37.571*	
	Office & small complex			−8.223**	0.000*
	Average number of employees per business				0.229*
	Number of residents per unit area (m^2) (business density)		276.182***	209.162***	196.598***
_cons		−22.999***	−28.480***	−14.158***	−19.467
Log likelihood		−87.108	−52.075	−59.010	−62.351
Hosmer−Lemeshow chi2		11.73	12.62	7.73	5.17
Correctly classified		82.99%	90.00%	87.55%	87.97%
area under ROC curve		0.890	0.959	0.948	0.946
Pseudo R2		0.351	0.611	0.560	0.535

주말		Commercial area	Commercial area Inc. 50m buffer	Commercial area Inc. 100m buffer	Commercial area Inc. 150m buffer
		Coef.	Coef.	Coef.	Coef.
상권 내부	LUM	3.503**	3.280*	3.537**	3.989**
	Residential	10.628***	10.813***	9.915***	10.046***
	Retail & service	9.930***	9.001***	9.072***	9.006***
	Restaurant·pub	21.440***	21.559**	18.454**	16.142**
	Accommodation & tourism	11.599***	12.550***	10.643***	10.324***
	Office & small complexs	5.352*	8.701***	7.888**	7.858**
	Medical care	11.371*			
	Sports		−51.292*	−58.449**	−55.026*
	20~30s ratio (population)		4.919**	6.411**	6.259**
상권 외부	LUM				−5.695***
	Retail & service			9.816***	
	Number of residents per unit area (m^2) (business density)		−410.859**	−433.392**	
	Average number of employees per business		−0.427***	−0.414***	−0.384**
_cons		−12.735***	−11.783**	−15.817***	−7.181**
Log likelihood		−113.303	−113.303	−88.956	−92.656
Hosmer−Lemeshow chi2		11.01	11.01	8.60	27.95
Correctly classified		79.49%	79.49%	85.90%	83.76%
area under ROC curve		0.831	0.831	0.894	0.894
Pseudo R2		0.224	0.224	0.391	0.365
Number of obs		234			

***p<0.01 **p<0.05 *p<0.1

4. Conclusion

The higher LUM and the ratio of residential use, the more concentrated visitors are at night on both weekdays and weekends.

Medical facilities inside the commercial area also attracts visitors at nighttimes.

The job center and commercial center are separated on weekend nights.

Thank you.