

A nighttime photograph of the Seoul skyline, featuring the N Seoul Tower on a hill in the foreground, with the city lights and other buildings visible in the background under a dark blue sky.

Prediction of Hot Places in Seoul based on Machine Learning

Hyun Seung Hong
Hye Rim Lee
Ju Hee Kim

Hongik University

Index

I. Introduction

Background
Purpose

II. Data Collection

Data for selecting
Data for predicting

III. Data Analysis

Hot Place Selection
Variable Extraction
Hot Place Prediction Algorithm

IV. Result

V. Conclusion

Prediction of Hot Places
in Seoul
based on Machine Learning

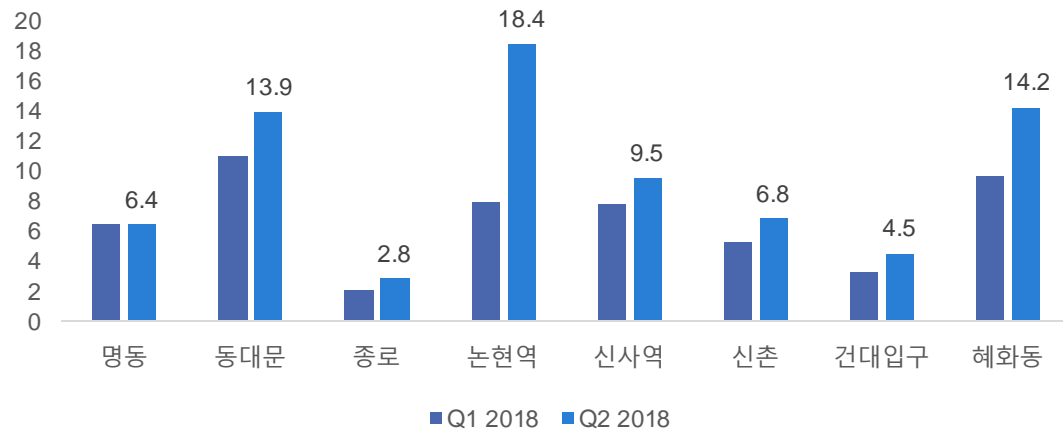
1. Introduction

Purpose & Outline

Background

Seoul Marketing Area Status

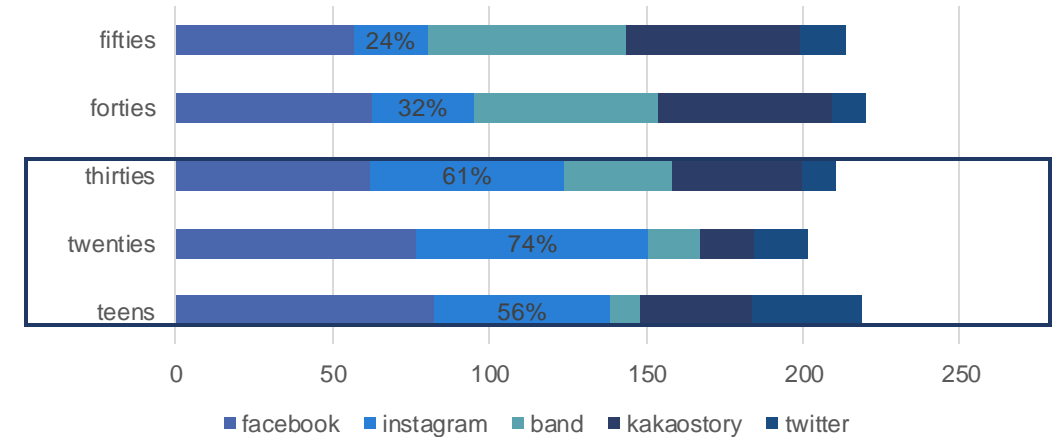
Figure 1. Seoul Major Marketing Area Vacancy Rate



- "Gentrification", the phenomenon that indigenous people are pushed out due to the inflow of outsiders and money and the increase of rents as the underdeveloped area near the city is activated.
- Because of gentrification phenomenon, **Seoul marketing areas' vacancy rate steadily increases**, there areas' businesses are going downturn.
- **Entrepreneurs worry about choosing a business location.**




Increasing SNS Utilization

Figure 2. SNS Service Utilization by Age

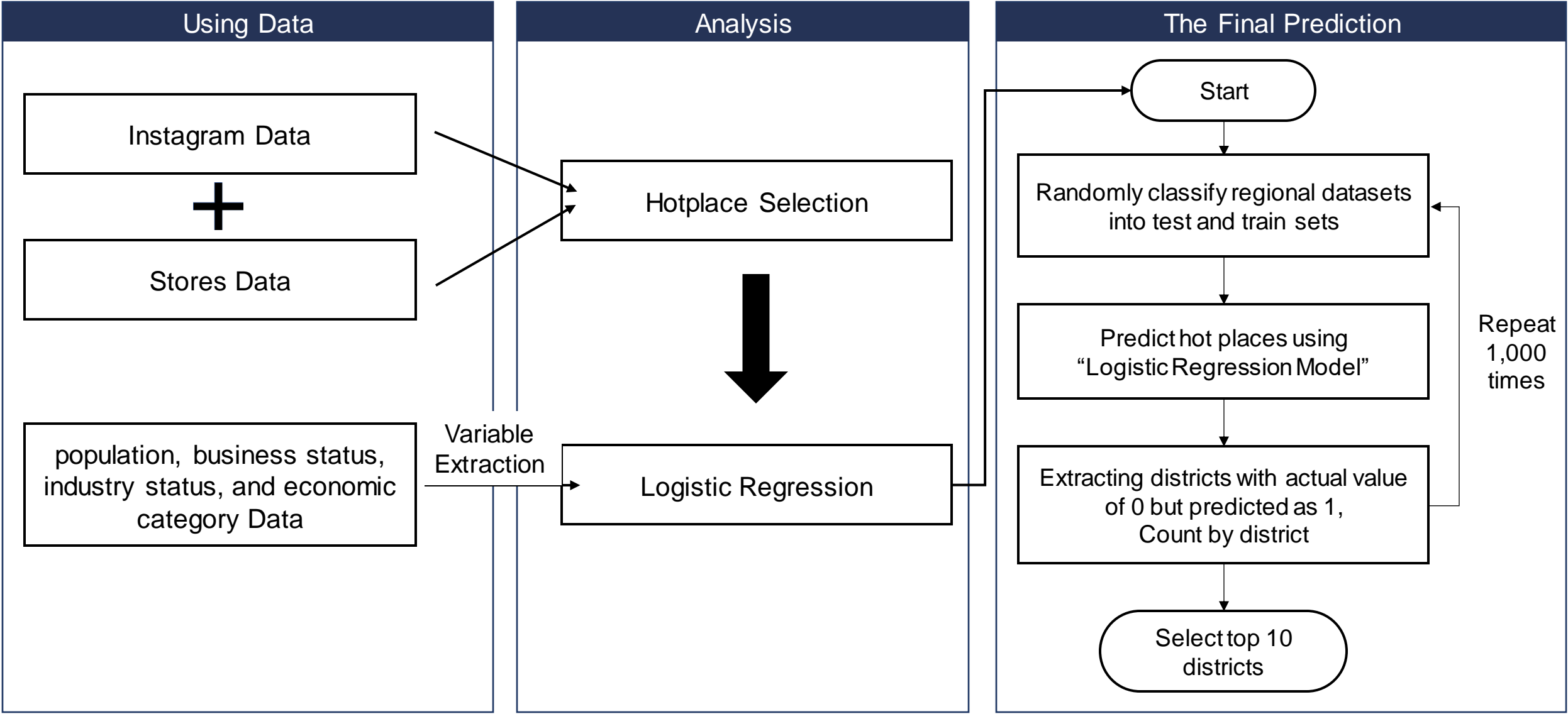


- Recently, people in their 20s and 30s actively share consumption tendencies **using SNS**, especially Instagram.
- Based on this, we will select a hot place and analyze the variety characteristics of the district to **predict the district that will emerge as a hot place in the future.**

Purpose

Purpose		
		
Establish countermeasures against gentrification problems	Providing location insight to future entrepreneurs in their 20s&30s	Help local revitalization
<p>As a result of this study, it is expected that the prediction of the ‘hot place’ can be made in advance to establish the policy and the countermeasures against it.</p>	<p>The prediction of hot places would provide location insights to future entrepreneurs preparing to start a business.</p>	<p>By selecting highly correlated factors among the variables, provide insights to selected hot place candidates to help revitalize the region.</p>

Outline

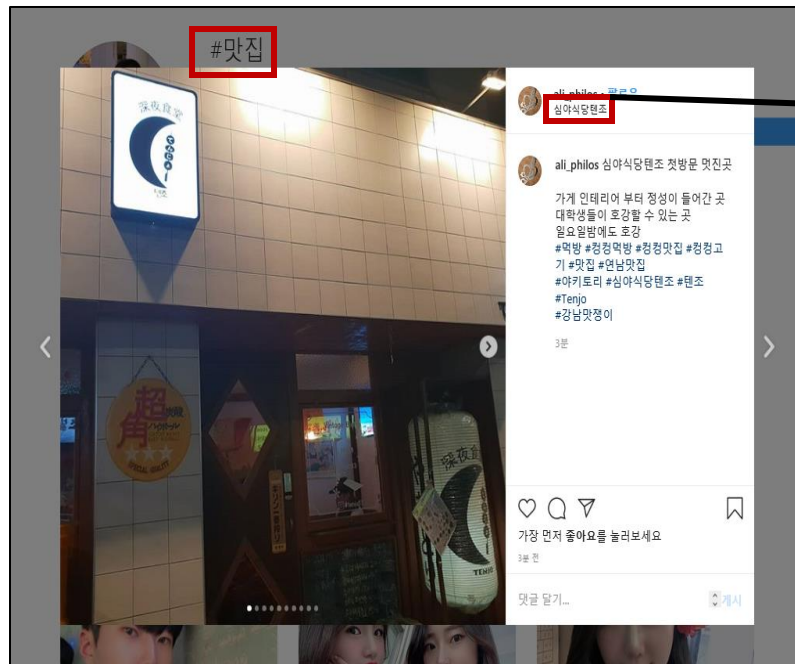


II. Data Collection

Data for selecting
Data for predicting

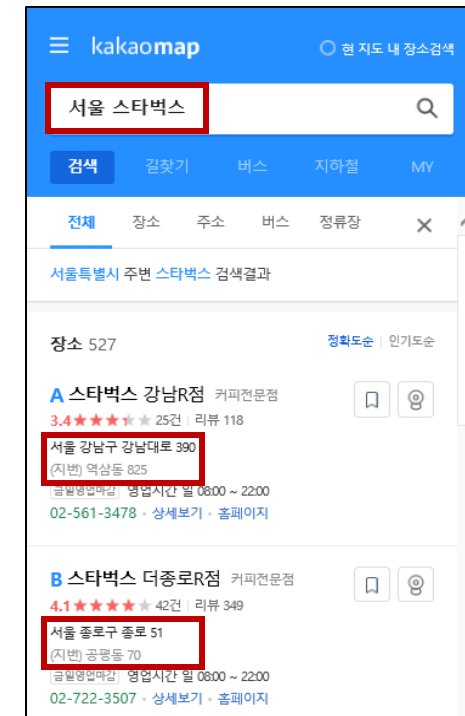
Data for Selecting

Instagram Crawling



- By searching for tags related to hot places (jmt, waiting, dating, restaurants, cages), we **crawled location about 300~400 posts per tag**
- the collected location was searched by the crawler on Kakao map, and only the location in Seoul was **extracted and classified**

Stores Crawling



- The number of stores in each region in Seoul was extracted through a crawler that **collects locations after searching for keywords in Starbucks, McDonald's, and Mara soup** on the 'Kakao map'

II. Data Collection

Data for Predicting

- Collection for hot place analysis, prediction



Category	Data
Population	Moving in, Moving out, single household, 2 households, 3 or more household, weekly population index, resident registration population
Industry Status	Food, Living Service, Leisure, Sports, Academic Education, Accommodation
Business	Operation Store / Closed Store
Economy	GDP, Level Index

Table 1. Extracted data by category

III. Data Analysis

Hot Place Selection
Variable Extraction
Hot Place Prediction Algorithm

Hot Place Selection

District Scoring by Instagram Hash Tag

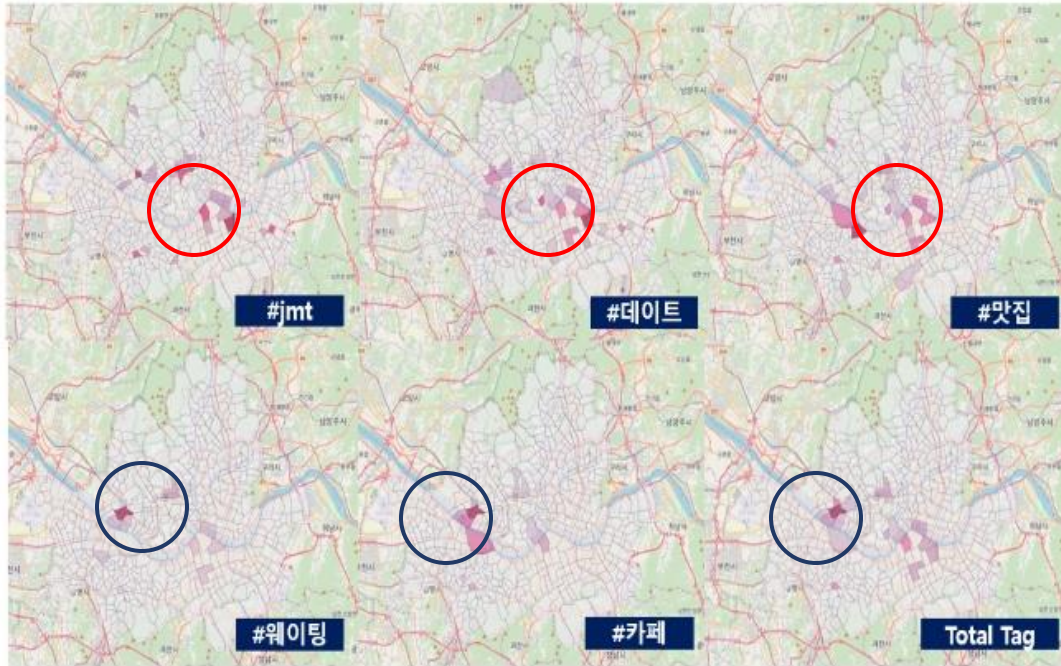


Figure 3. The number of each tag for each district

- The Instagram location hashtag is determined to be sufficient to select a representative hot place in Seoul through hash tag visualization.

District Scoring by the Number of Stores

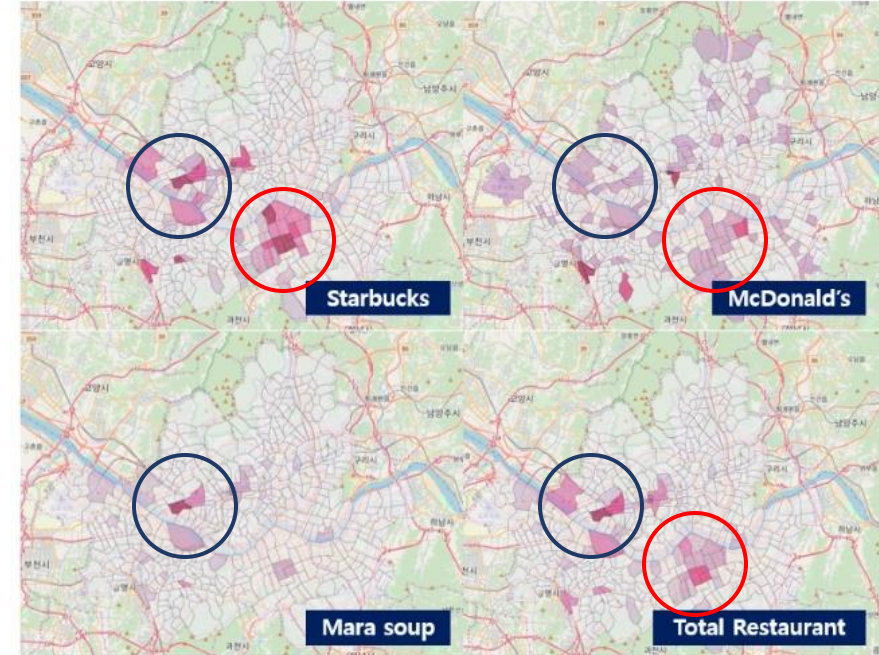


Figure 4. The number of each store for each district

- The number of Starbucks and McDonald's stores is **similar to** the Instagram hashtag, and is a good example of a representative hot place.
- In the case of Mara soup, Instagram hashtags appear in the **most crowded** places, and the importance of store weight should be weighted.

Hot Place Selection

District scoring

Because Total Mara soup stores are significantly smaller than Starbucks / McDonald's

Score of Store by region =
Number of Marasoup stores per region * 2 + Number of Starbucks per region + McDonald's per region

+

SNS post count by district = SNS score by district

Total score by distric =
SNS score by district * 0.7 +store score by district *0.3

Becuase it's about selecting districts that 2030 population "actually goes to"

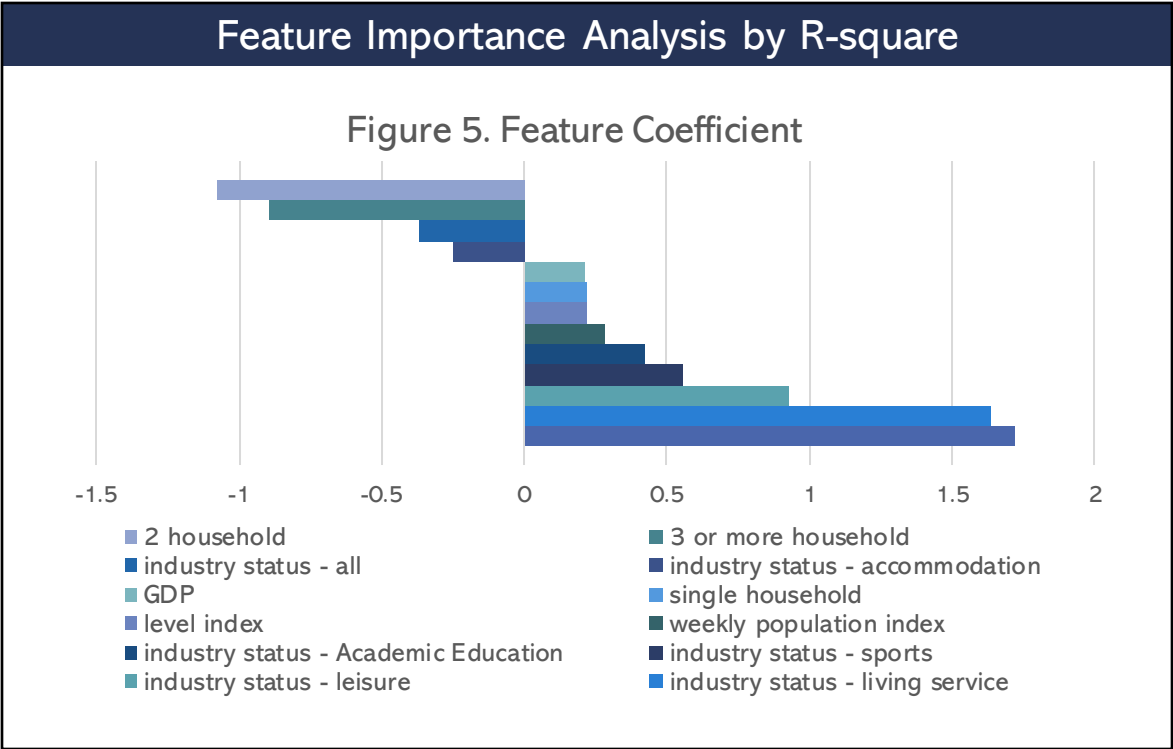
Hot place selection result

rank	name	Total Score
1	서교동	37.9
2	신사동(강남)	31.7
3	역삼1동	17.8
4	종로1.2.3.4가동	16.7
5	여의동	13.6
6	신촌동	13.2
7	청담동	13.1
8	연남동	13.1
9	삼성1동	12.4
10	이태원1동	9.7

rank	name	Total Score
275	상도4동	0
276	부암동	0
277	무악동	0
278	교남동	0
279	창신2동	0
280	창신3동	0
281	송인2동	0
282	신당5동	0
283	황학동	0
284	다산동	0

- The score difference between the top 40 districts and the bottom 384 districts was large.
- Out of a total of 424 districts, 40 districts(Seogyo-dong, Sinsa-dong, Yeoksam 1-dong, Jongro 1,2,3,4 Ga ...) are hot places.
- 384 districts(Burn 3dong, Suyu 3dong, ..)were classified as non-hot places.

Variable Extraction



Exp 1. Variable Coefficient

$$z = D(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_d x_d$$

Exp 2. Logistic expression

$$p(y = 1|x) = g(z) = \frac{e^z}{1 + e^z} = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_d x_d}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_d x_d}}$$

Feature Selection

Dependent variable	Hot Place O/X
Independent variable	
Population	1-person households, 2-person households, 3-person households or more, Weekly Population Index, Registration Population
Industry Status	Food, Living Service, Leisure, Sports, Academic Education, Accommodation
Business	Operation Store/ Closed Store
Economy	GDP, Level Index

Table 2. Final variables

- Exp 1's $\beta_0, \beta_1, \beta_2, \dots, \beta_n$ value is the slope that the variable represents in logistic regression, which means the influence to dependent variable of each independent variable.
- Variables with β_n between 0.2 and 1.7 absolute values are selected as final independent variables.**

Hot Place Prediction Algorithm

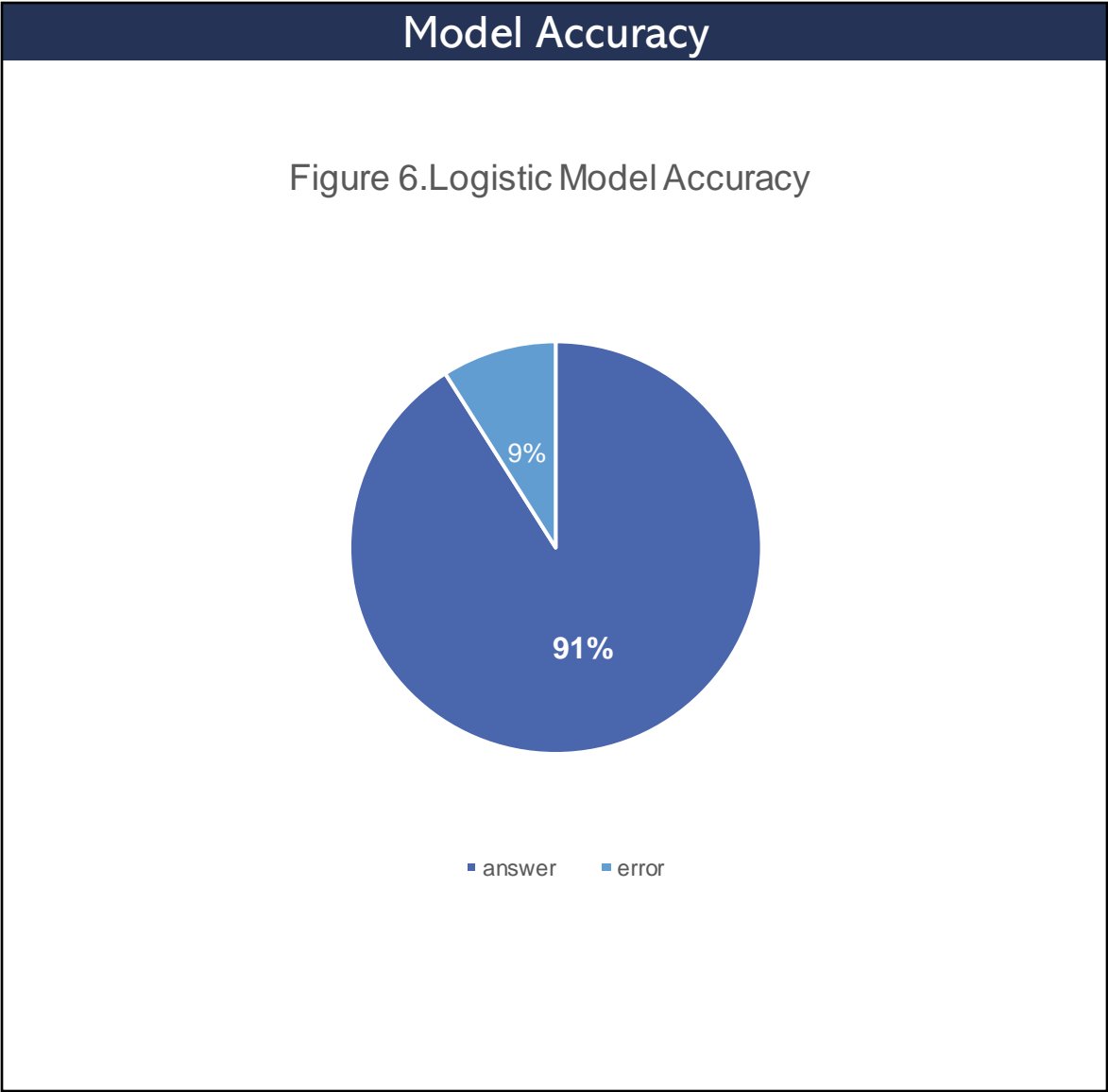
Confusion Matrix

	Predict Non-Hot Place	Predict Hot Place
Actually Non-Hot Place	Actually Non-Hot Place	Non-Hot Place, But Predicted Hot Place
Actually Hot Place	Hot Place, But Predicted Non-Hot Place	Actually Hot Place

Table 3. Confusion Matrix

- β error:** Regions that Non-hot places but logistic model predicted to hot place
- The logistic model is suitable for binary classification since the above graph is derived by matching a kind of sigmoid function when calculating the probability.
- By training the logistic model 1,000 times, we can **extract the regions predicted by the model as hot places from the non-hot places.**

IV. Result



Predicted Hot Place

Figure 4. Top 10 districts

rank	name	count
1	중림동	340
2	삼청동	333
3	필동	308
4	남영동	294
5	종로 5,6가동	286
6	장충동	260
7	신당5동	255
8	대치4동	243
9	한강로동	222
10	광희동	213

V. Conclusion

Conclusion

Junglim-dong

전통과 핫플레이스가 공존하는 곳... '반전 매력' 중림동 뜬다

원문 2020.01.10 18:05 | 수정 2020.01.11 15:10 | 제1면 A1

오랫동안 과거에 머물러 있던 중림동이 크게 달라지고 있다. 2017년 봄 서울역 고가가 '서울로7017'로 탈바꿈한 이후 해마다 동네의 모습이 바뀐다. '남대문시장과 회현동에서 시작되는 고가 보행로가 사람들의 발길을 중림동과 만리동으로 쉽게 이어준 덕분'이라고 상인들은 말한다.



"Young people have been crowded since the opening of Seoulro 2017 and 'hot places' including restaurants appears in every alley"

동네로 유입되는 사람들이 늘면서 맛집 등 '핫플레이스'들이 골목마다 등장했다. 독특한 콘셉트의 식당과 카페엔 주중에도 젊은이들이 붐빈다. 양식당 서울부띠끄, 퓨전레스토랑 동네로 유입되는 사람들이 늘면서 맛집 등 '핫플레이스'들이 골목마다 등장했다. 독특한 콘셉트의 식당과 카페엔 주중에도 젊은이들이 붐빈다. 양식당 서울부띠끄, 퓨전레스토랑 베리스트릿키친 등이 대표적이다. 설렁탕집 중림장, 닭고치 가게인 호수집 등 노포(老鋪)들은 전국적인 맛집으로 떠올랐다. 근대 서울의 모습과 젊은이들이 즐겨 찾는 '핫플레이스'가 공존하는 곳. 중림동이 서울 도심 명소의 하나로 거듭나고 있다.

Recent developments in the region and the formation of commercial zones are likely to be hot places.

Pil-dong

서울시, 도심 싹틔줄 골목길 재생사업 본격 추진

조성신 기자 | 입력 : 2018.02.08 10:24:35 | 댓글 0



"Alleyway revitalization with 'Seoul-type alleyway recycling business' "

서울시는 도시의 역사와 함께해온 역사문화유산이자 삶의 공간인 골목길을 일·살·놀이가 어우러진 곳으로 재생하는 '서울형 골목길 재생사업'을 올해부터 본격 추진한다고 8일 밝혔다.

'서울형 골목길 재생사업'은 도시재생활성화지역 등 일정 구역을 정해서 '먼' 단위로 재생하는 기존 도시재생사업과 달리 골목길을 따라 1km 이내의 현장 밀착형 소규모 방식의 '선' 단위 재생 개념을 도입해 추진한다. 담장 낮추기, 골목 마당 공유, 내 집 수선하기 같은 사업도 주민 주도로 함께 진행한다.

시는 지난해 8월에 착수한 '서울형 골목길 재생 기본계획' 용역을 3월경에 마무리하고 오는 5월까지 골목길 재생사업 가이드라인을 마련해 기본계획을 수립할 계획이다. 6월에는 자치구 공모를 통해 사업대상지를 추가 선정하고 재생사업을 본격 확대 추진한다. 기본계획 수립과 병행해 시는 용산구와 성북구 2곳 골목길을 시범사업지로 선정하고 5월 중 실행계획을 수립, 연내 사업에 착수할 예정이다.

Namyong-dong

남영동 골목여행에서 만난 모듬스테이크, 숙대입구맛집 '서지스테이크'

최태인 기자 | 승인 2018.12.13 15:45 | 댓글 0

"With the addition of restaurants and cafes, it is attracting new attention as a hot place."

서울에서 만날 수 있는 대표적인 골목길로는 용산구 이태원동의 경리단길, 마포구 망원동의 망리단길, 관악구 서울대 입구의 사로수길 등이 있다. 용산구 남영동의 열정도로도 새롭게 주목받고 있는 핫플레이스다.

남영동 열정도로는 숙대입구역과 남영역 인근 지역을 일컫는다. 이 지역이 눈길을 끌기 시작한 것은 청년사업가들이 지역 상권 살리기에 발 벗고 나서면서다.

중소규모의 인쇄업체가 모여 있던 골목길에 감각과 위트가 가득한 식당, 카페 등이 하나둘 들어서면서 사람들의 발걸음을 모으고 있다.

Thank you.