



# 카페 최적 입지 선정을 위한 조건

서울시 카페 폐업률 예측을 통한 상권분석

팀명 : 임 앤 김스

이화여자대학교 통계학과

김미성, 김연정, 임지연

# CONTEXT



## I . 프로젝트 소개

분석배경, 분석목적, 데이터 설명

## II . 분석과정

데이터 시각화, 분석개요

## III . 분석결과

분류모형, 생존분석

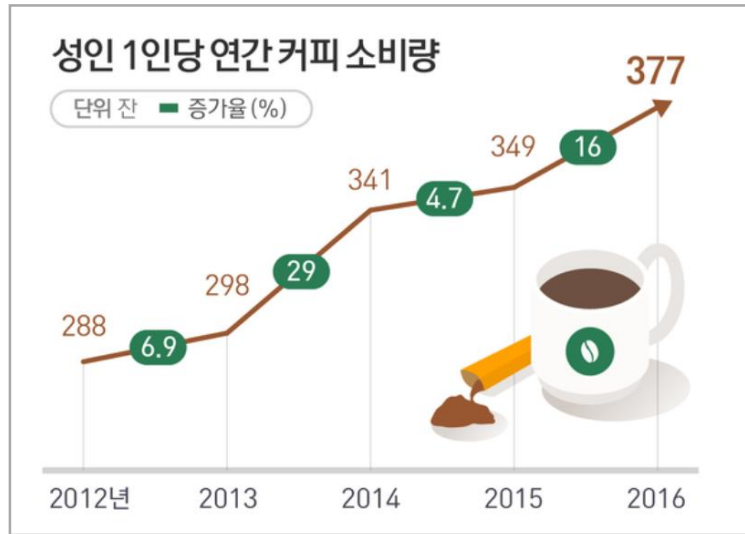
## IV . 결론

입지추천, 위험률 제시, 분석 차별점 및 한계점, 정책 활용 방안

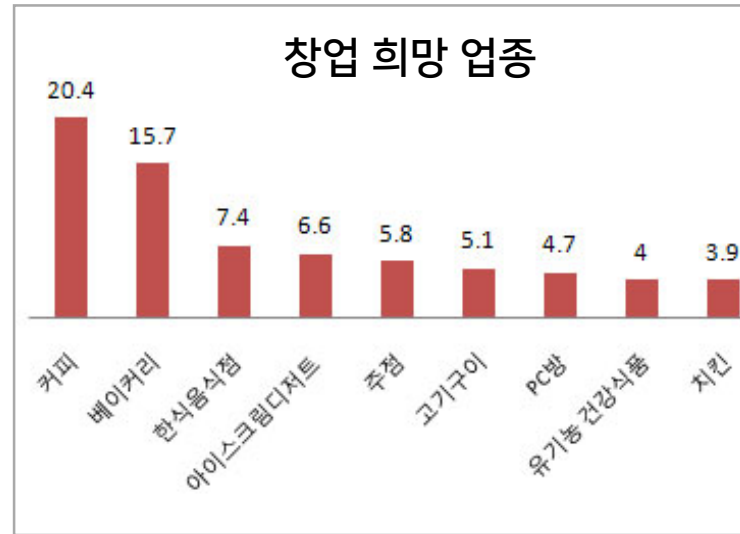
1인당 커피 소비량 ↑, 커가는 커피시장

가장 창업하고 싶은 '카페'

카페는 이미 포화상태



출처 : news1뉴스

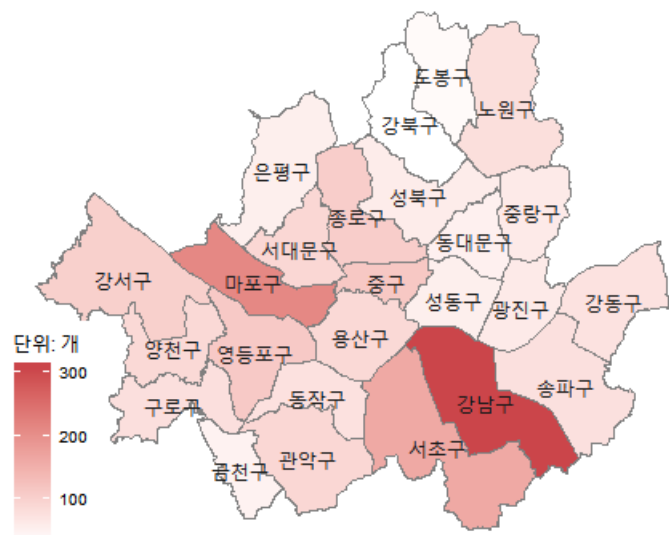


출처 : 이데일리

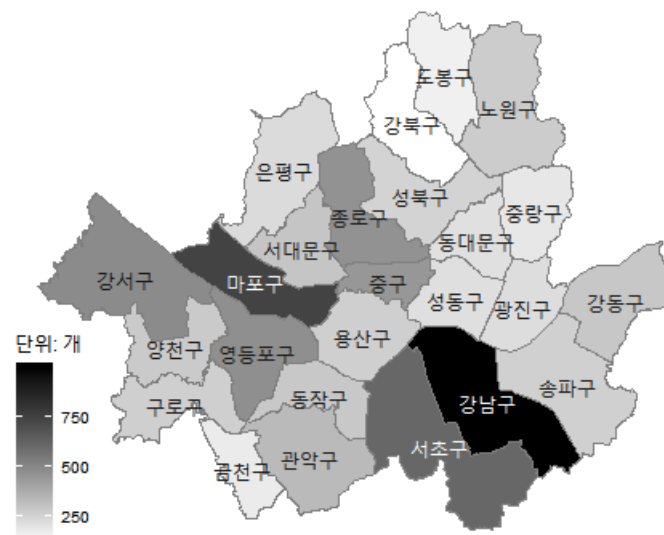


한국인의 커피 소비량이 증가하며 커져가는 커피시장,  
창업자들의 희망 아이템 1위 "커피"

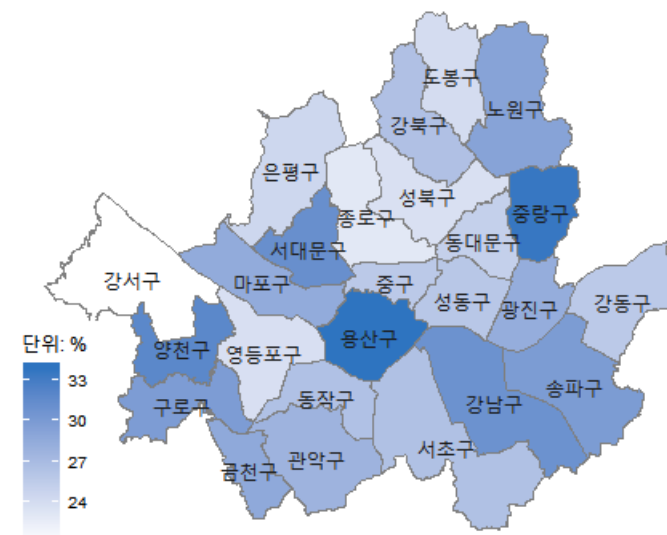
폐업한 카페 수



개업한 카페 수



폐업률 (개업 수 대비 폐업 수)



예비 카페 창업자에게는 적절한 위치를 추천해서 최적의 **입지 선정**을 도와주고,  
기존 카페 창업자에게는 **폐업 위험률**을 제시하자 !!

## 기존의 상권분석 서비스

서울 상권 선택

행정구역

서대문구

신촌동

선택조건 (선택 항목 기준 상위 10%)

기준분기

2019년

2분기

☒ 매출  
☒ 유동인구  
☒ 주거인구  
☐ 직장인구  
☐ 3년 생존율  
☐ 점포수  
☐ 평균영업기간

\*최대 3개

검색

우리 마을 가게

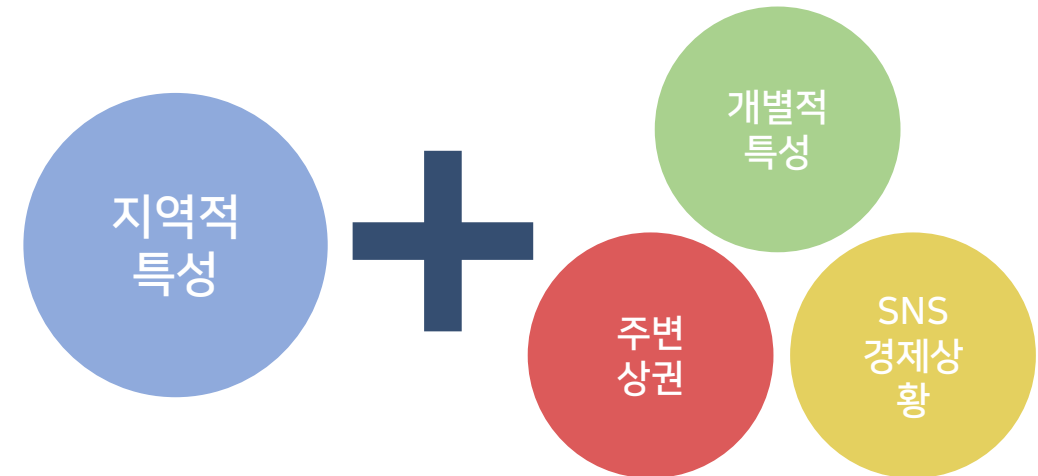
상권분석서비스

지역적인

특성만을 반영

우리 마을 가게 상권분석 서비스에서 창업 위험률 제공 중이나, 지역적인 특성만을 반영하고 있음

## 차별화된 상권분석 서비스



기존 영업 중 카페에게 폐업 위험률 제공  
새로운 창업자에게 **최적 입지&조건** 추천

## 1. 개별 특성 데이터



층 구분	지하, 1층, 2층, ...
총 층수	
총 면적	단위: $m^2$
프랜차이즈 여부	
월, 계절	개업 시점

(출처 : 서울시 열린데이터 광장)

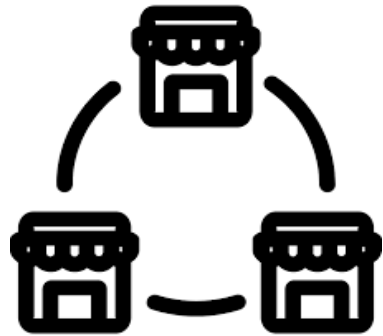
## 2. 지역 특성 데이터



유동/직장/주거 인구	행정동
여성인구 비율	행정동
가구 수, 소득분위	행정동
임대시세, 매출액	행정동
주차장 수	50m 이내
지하철역 수	1000m 이내
집객시설(약국, 학교, ...)	행정동

(출처 : 서울시 열린데이터 광장)

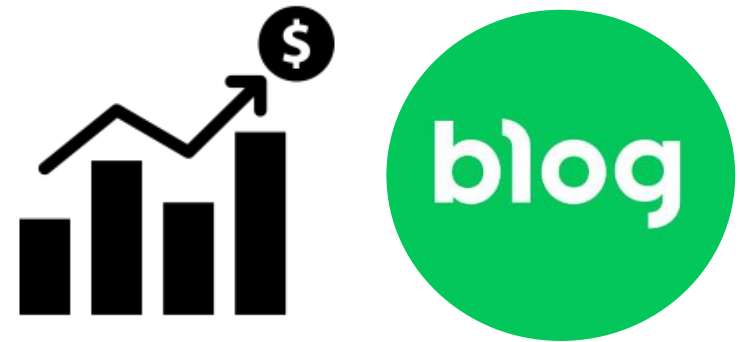
## 3. 주변 상권 데이터



주변 음식점 수	도로명 별
주변 카페 수	도로명 별
주변 프랜차이즈 카페 수	도로명 별
주변 스타벅스 카페 수	도로명 별
이전 달 새로 생긴 카페 수	500m 이내

(출처 : 서울시 열린데이터 광장)

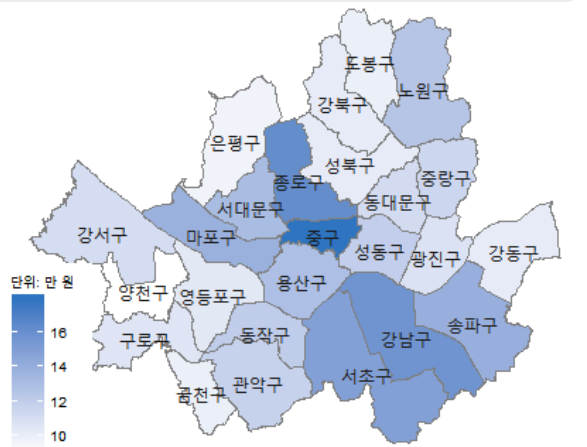
## 4. 기타 데이터



KOSPI 주간/월간 변화율	개업시점
KOSPI 연속 감소 주	개업시점
(출처 : Yahoo Finance)	
NAVER 블로그 건수	지하철 역

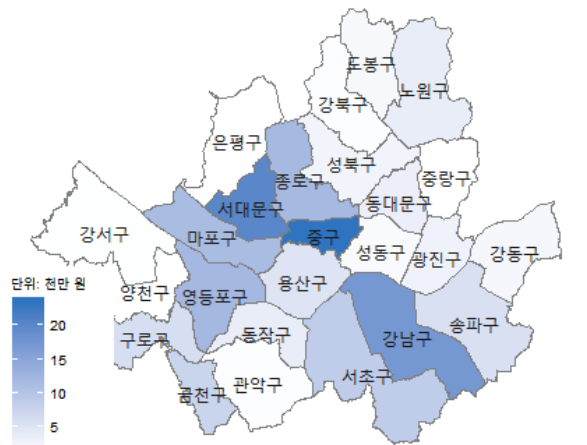
(출처 : 네이버 크롤링)

## 임대료



※ 3.3m<sup>2</sup>당 월 평균 임대료

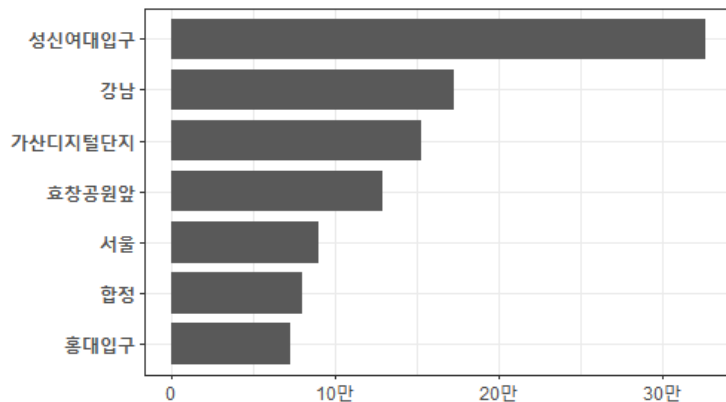
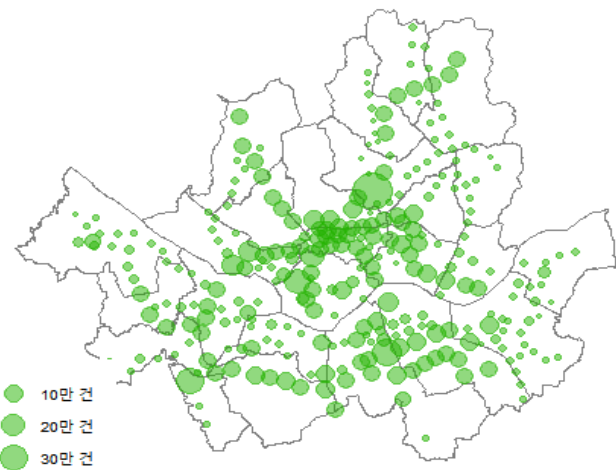
## 매출액



※ 월 평균 매출액

구별로 상이한 임대료/매출액  
임대료와 매출액은 대체로 비례

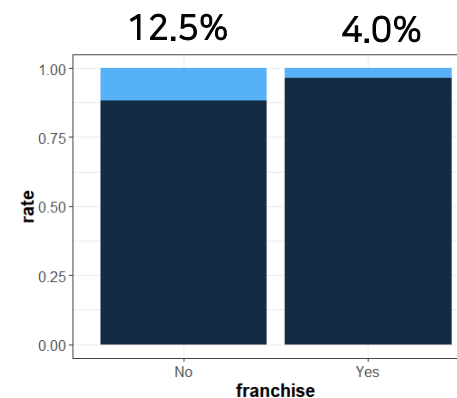
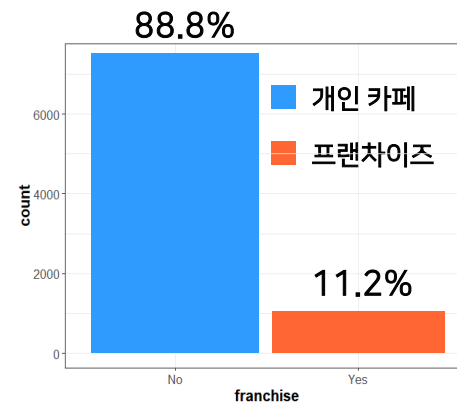
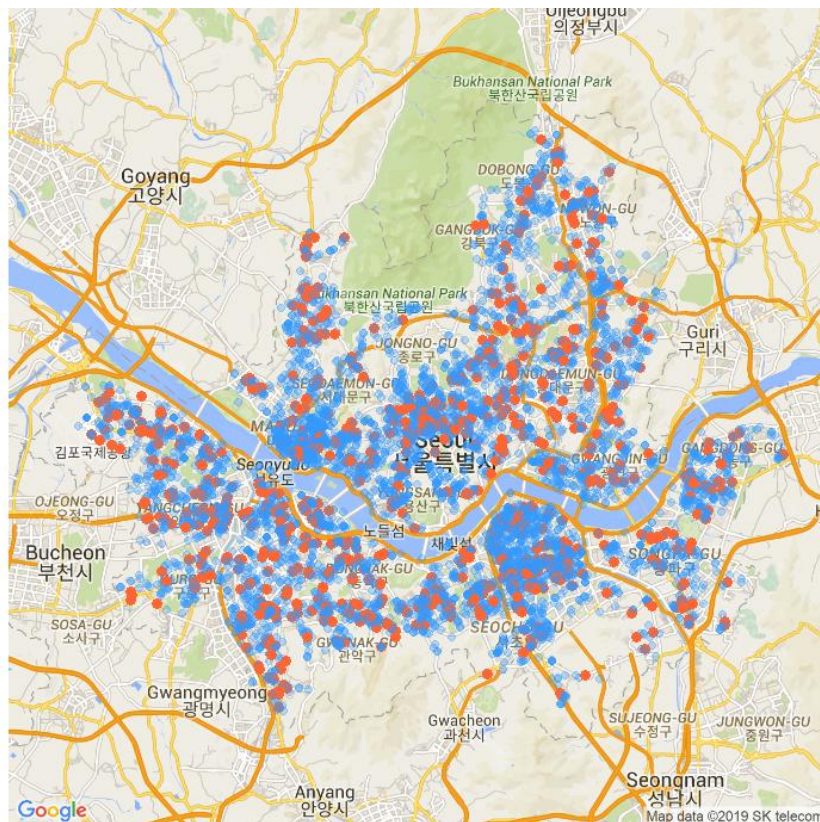
## 지하철역별 블로그 게시물 건수



성신여대 입구역, 강남역, ... 순  
유명 카페 거리 ↑ → 블로그 건수 ↑

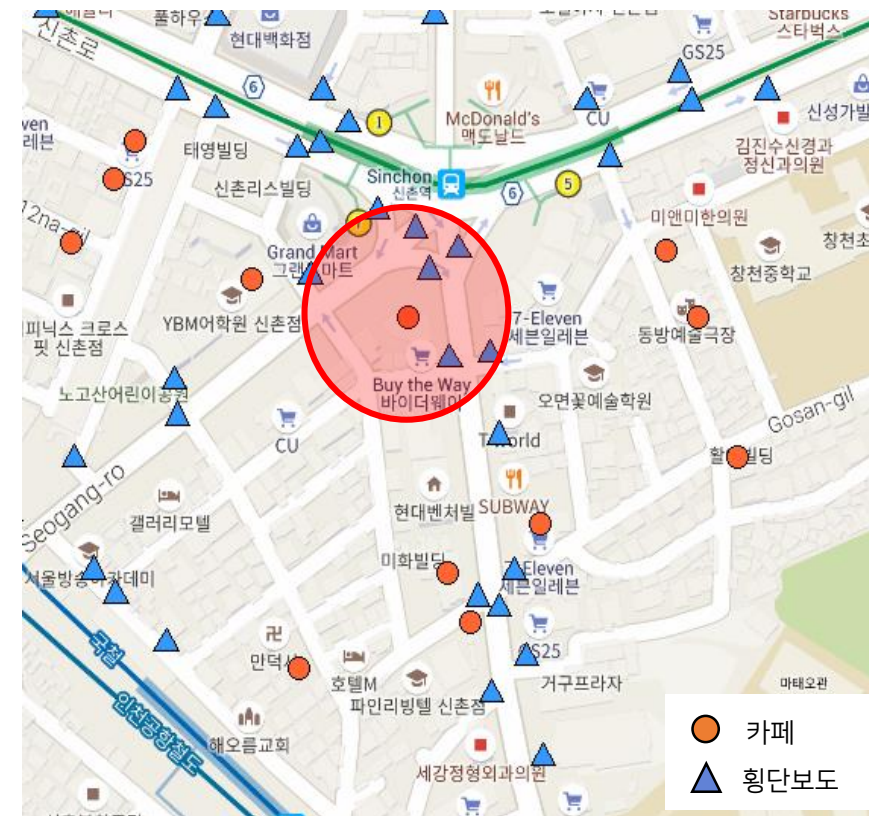


## 프랜차이즈 카페 vs 개인 카페



전체의 약 11%인 프랜차이즈 카페의 폐업률은  
일반 카페 폐업률의 약 1/3

## 50m 이내 횡단보도 수



50m 이내 횡단보도 수 6개

## 01 1년 내 폐업할 카페 예측 및 폐업 원인 분석

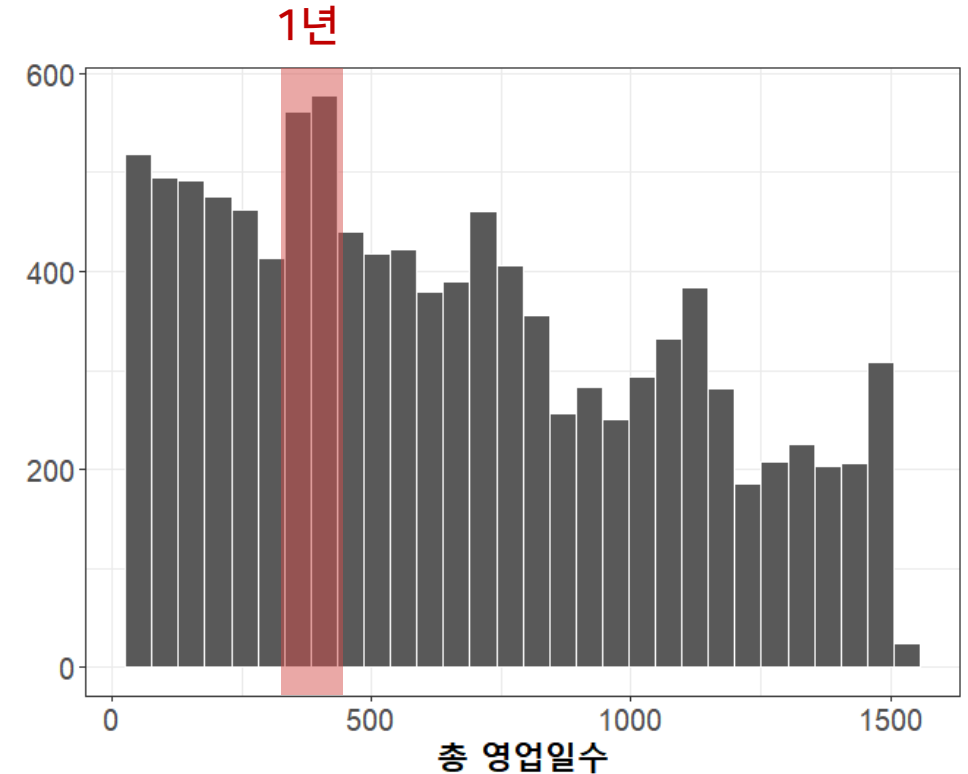
- 2015년 4월 ~ 2018년 5월에 개업한 8312개 카페 분석



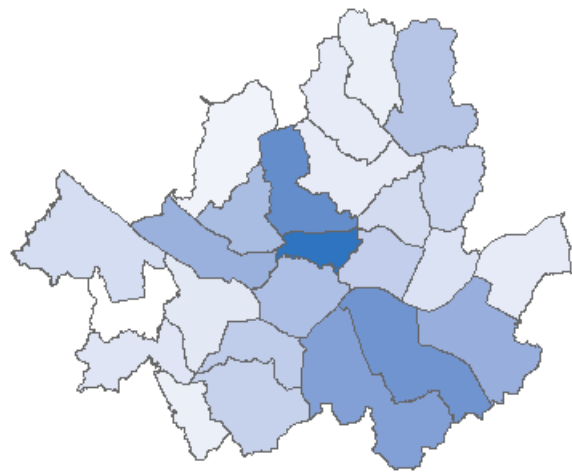
1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
2015				2016				2017				2018				2019			

## 02 생존분석을 통한 생존확률 제시

- 2015년에 개업한 카페 대상
- Cox-PHM 모형을 통한 위험률 및 생존확률의 변화 예측



## 1) 전체 행정구에 대한 폐업률 분석 [정확도 = 90.8%]

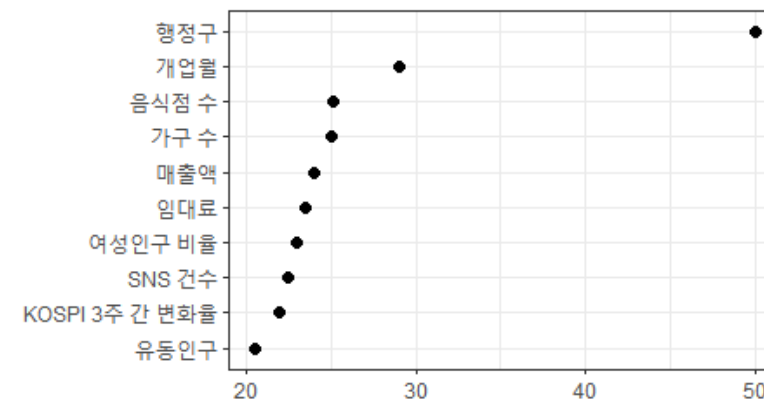


### 모델링

Logistic  
Ridge/Lasso  
SVM  
**Random Forest**  
XGboost

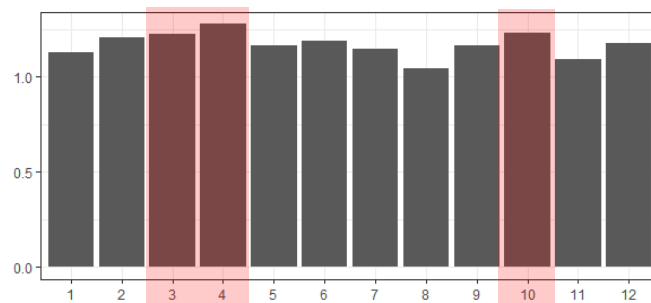


### 변수 중요도 그래프



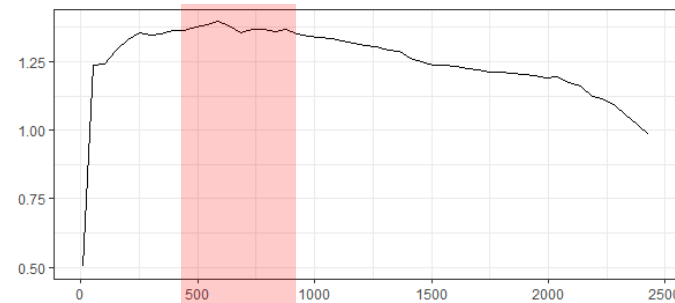
### Partial Dependence Plot

#### [개업 월]에 따른 생존확률



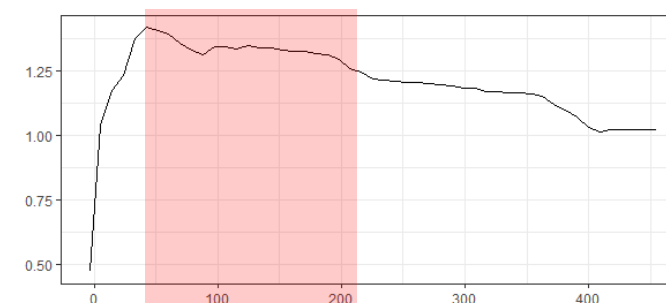
» 3월, 4월 또는 10월에 개업한 카페의  
생존률이 높음

#### [총 면적]에 따른 생존확률



» 총면적이 약 500~2000m<sup>2</sup>일 때 카페의  
생존률이 높음

#### [주변 음식점 수]에 따른 생존확률

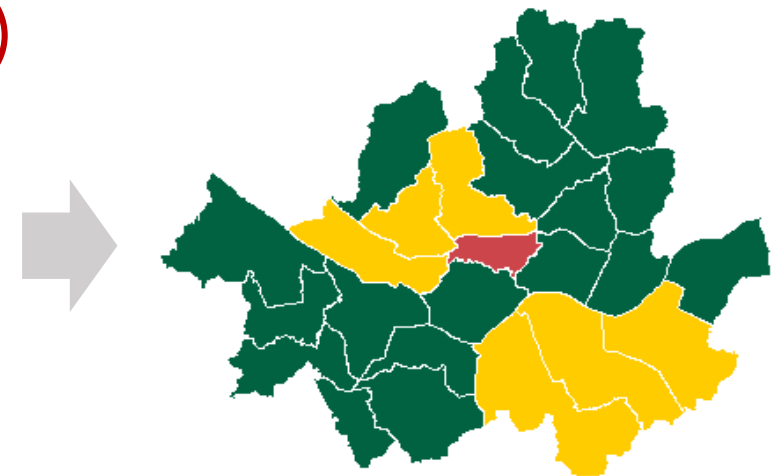
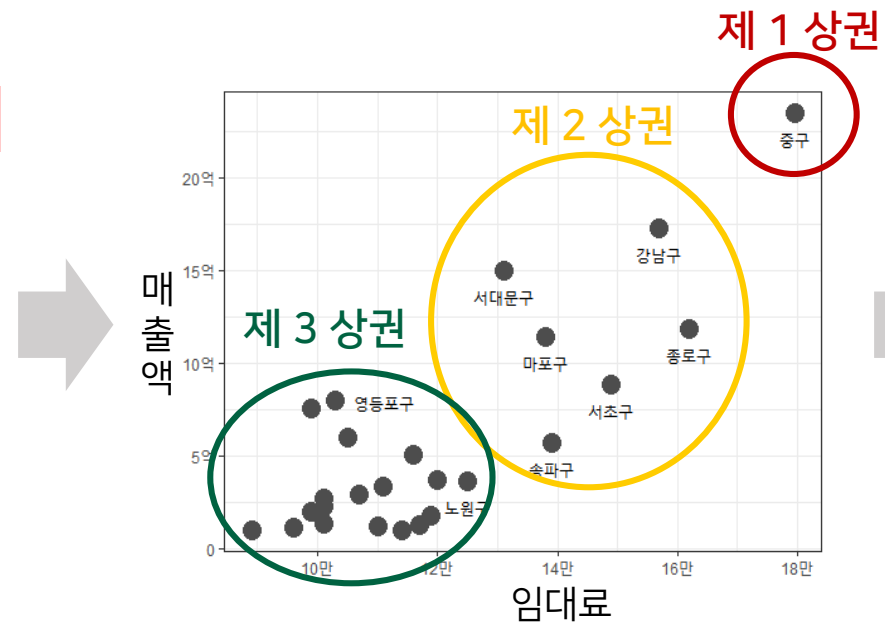
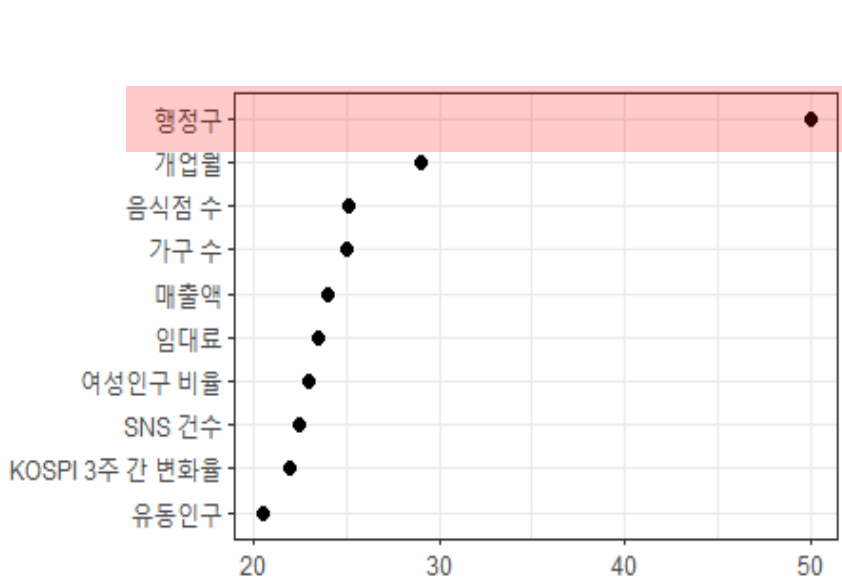


» 주변 음식점 수가 약 500~800개인  
곳의 카페 생존률이 높음

폐업률 예측에 가장  
중요한 변수 = 행정구

소상공인에게 가장 중요한  
임대료, 매출액을 이용하여  
행정구별로 클러스터링 실시

3개 군집으로 분류



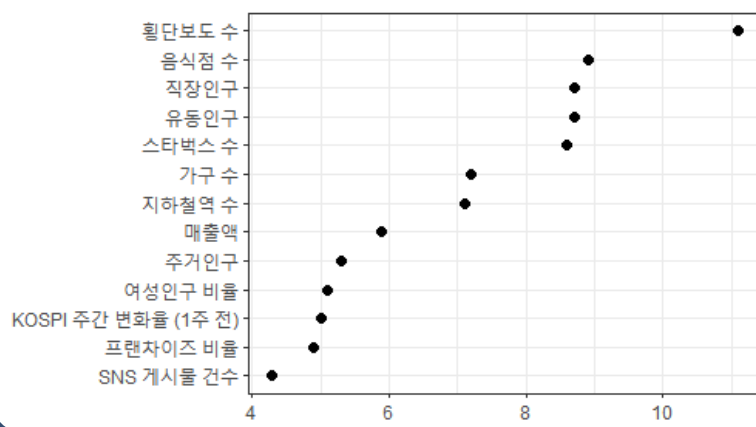
2) 클러스터링 후 각 상권 별 폐업률 분석

제1 상권 지역 [Accuracy = 93.2%]

(중구)

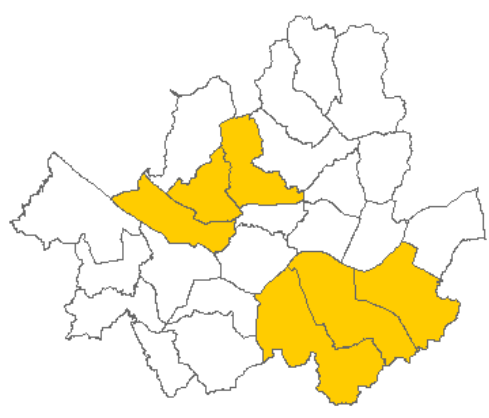


변수 중요도

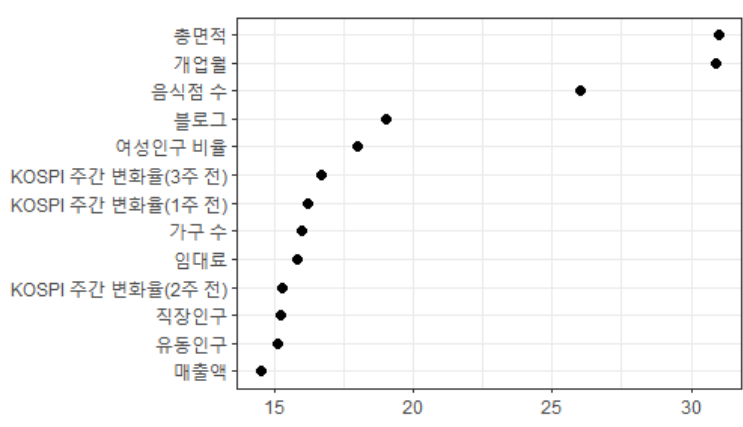


제2 상권 지역 [Accuracy = 88.9%]

(강남구, 서초구, 송파구, 종로구, 마포구, 서대문구)

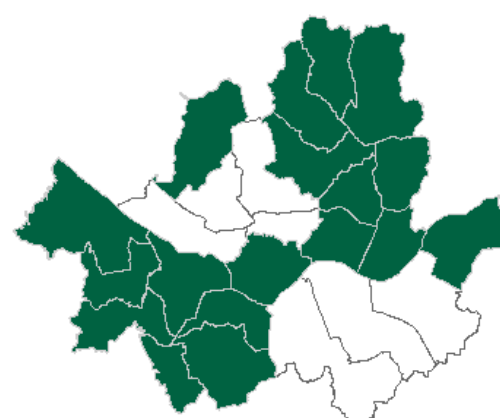


변수 중요도

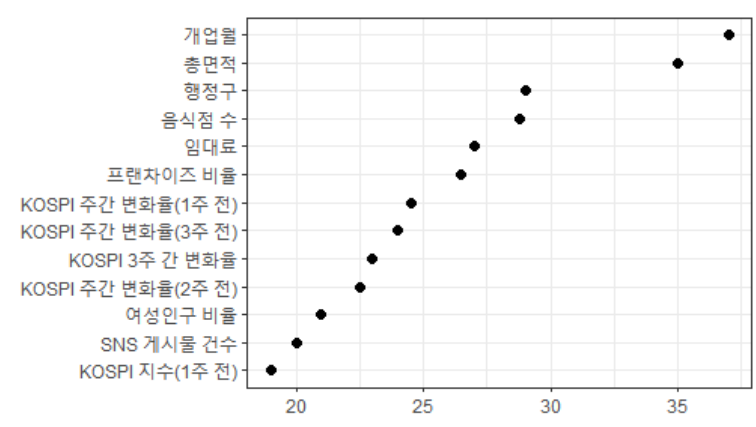


제3 상권 지역 [Accuracy = 90.2%]

(강북구, 성북구, 은평구, 강서구, 관악구, 성동구, ...)



변수 중요도



중요 변수의 생존률에 대한 영향

총 면적  
근처 새롭게 생긴 카페

근처 스타벅스 수

횡단보도

코스피 연속감소

관공서 수

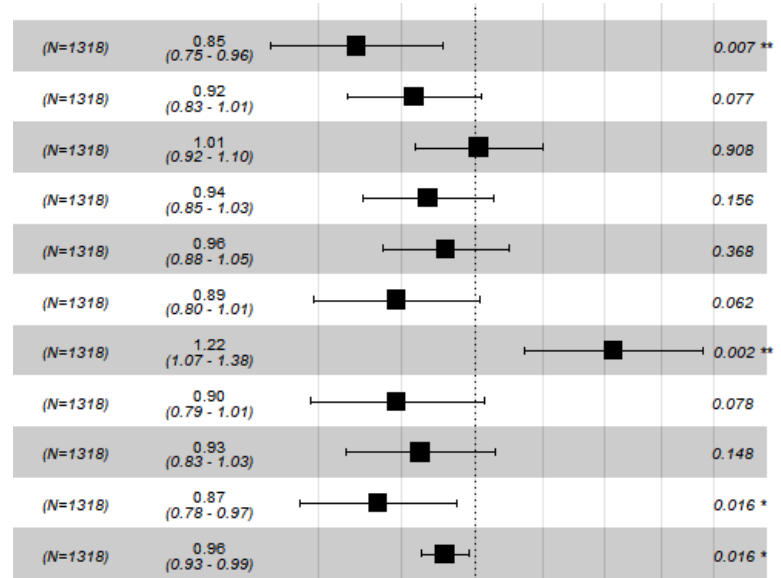
일반 병원 수

종합 병원 수

대학교 수

슈퍼 수

주차장 수



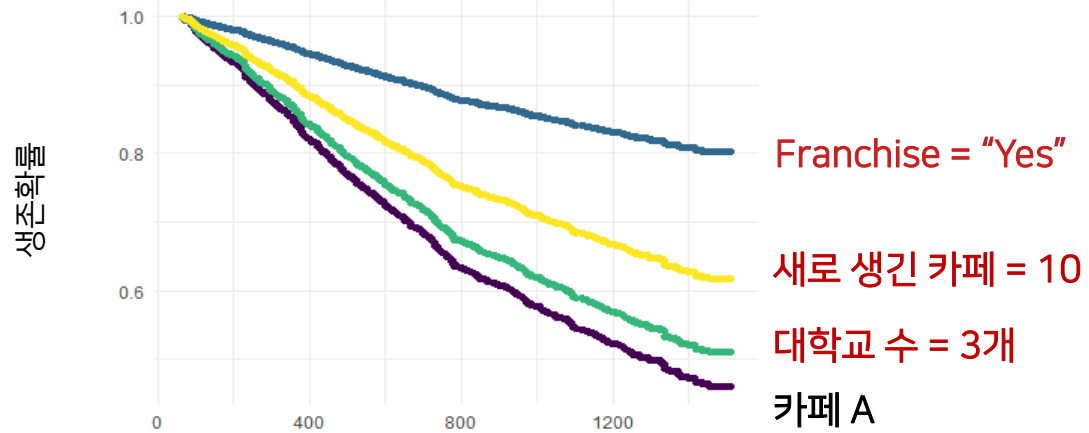
생존률 증가

생존률 감소

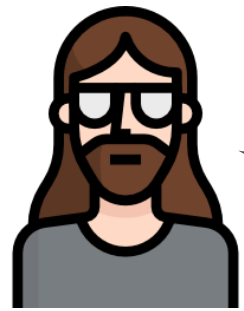
중요 변수에 따른 생존 확률의 변화

- 시뮬레이션을 통해 중요 변수의 값에 따른 카페 생존률의 변화를 비교  
[카페 A]

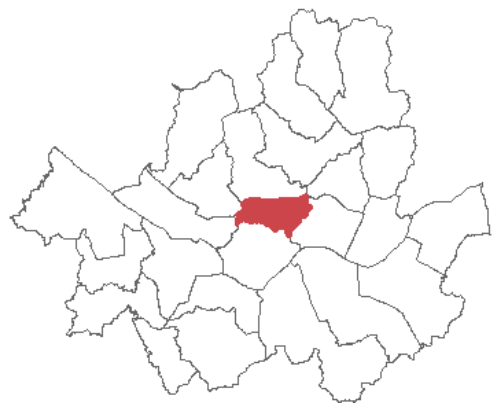
행정구	총 면적	...	근처 새롭게 생긴 카페 수	프랜차이즈	대학교 수
강남구	173 m <sup>2</sup>		1 개	No	0 개





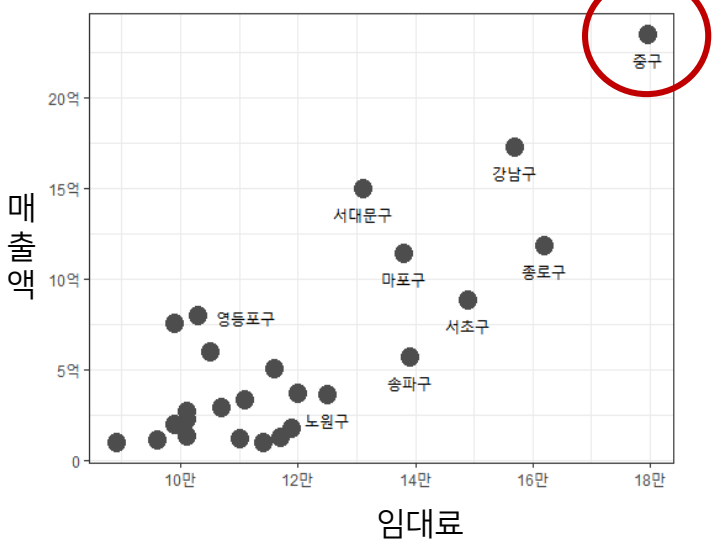


서울시에 **초대규모 자본**으로 카페를  
창업하고 싶은 임씨에게 최적 위치를 추천한다면?

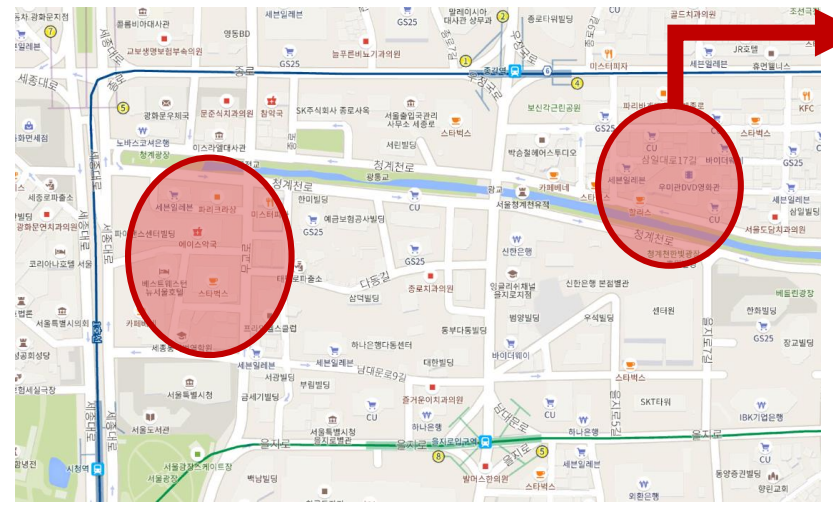


임대료, 매출액이 가장 높은 '중구'를 선정

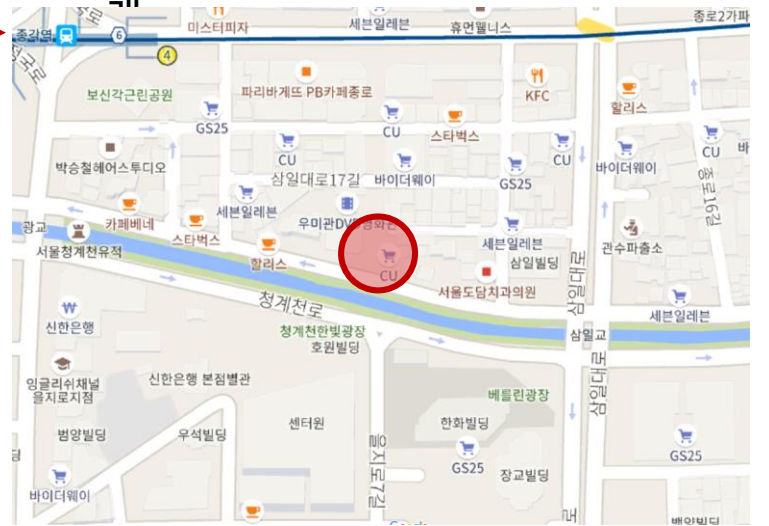
제 1 상권



행정동별 직장인구 수 40,000 ~ 45,000  
-> 무교동, 관철동

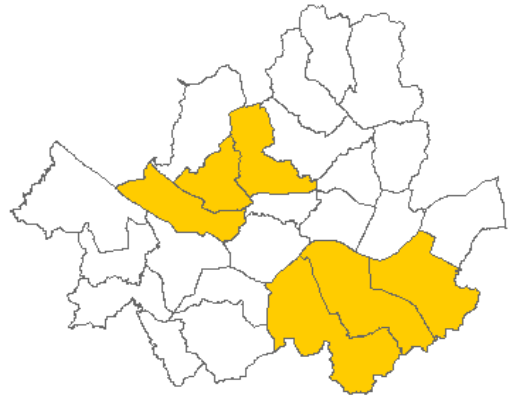


같은 도로명 내 주변 카페 수 2-3개, 주변 스타  
벅스 수 1개, 주변 50m이내 횡단보도 수 1-3

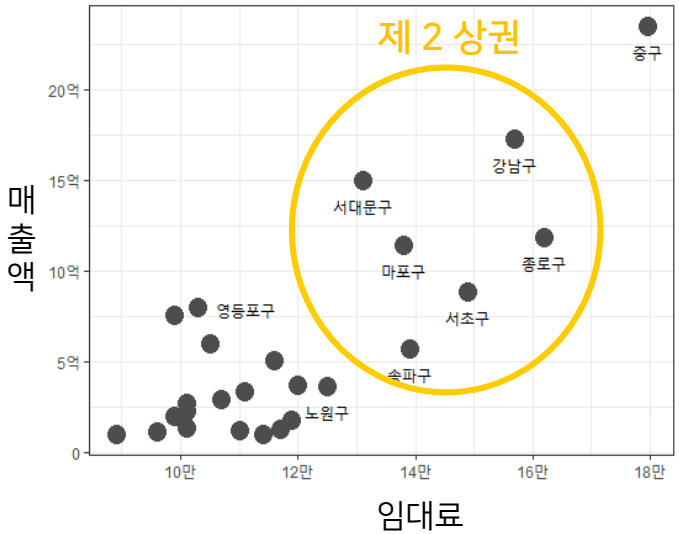




서울시에 **대규모 자본**으로 카페를 창업하고 싶은 김씨에게 최적 위치를 추천한다면?



제2 상권 중 폐업률이 낮은 종로구 선정



도로명별 주변 음식점 수 30~40개  
→ 창경궁로 26, 27길



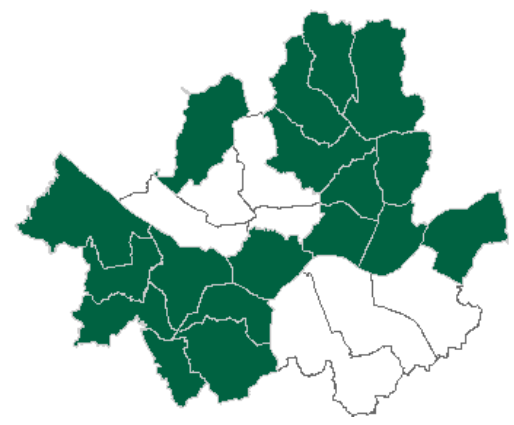
총면적이 약 50~120m<sup>2</sup>  
4, 5월 또는 10월에 개업 추천







서울시에 **소규모 자본**으로 카페를 창업하고 싶은 문씨에게 최적 위치를 추천한다면?



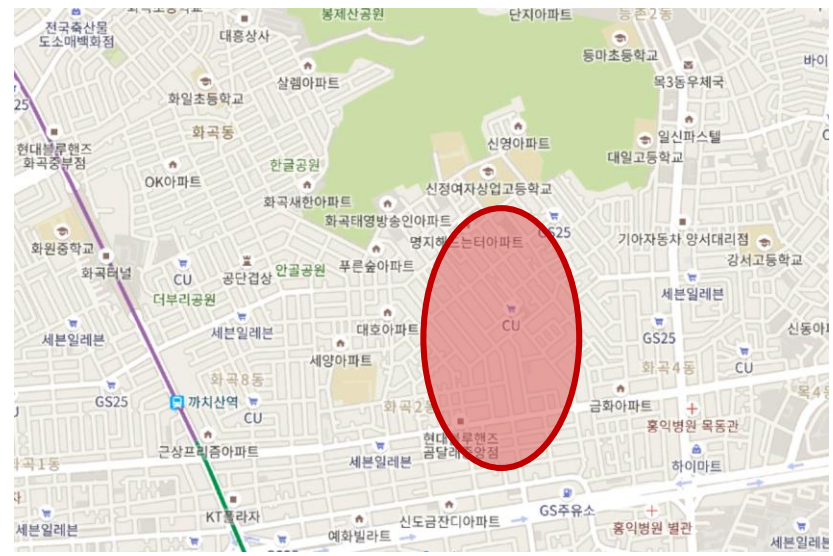
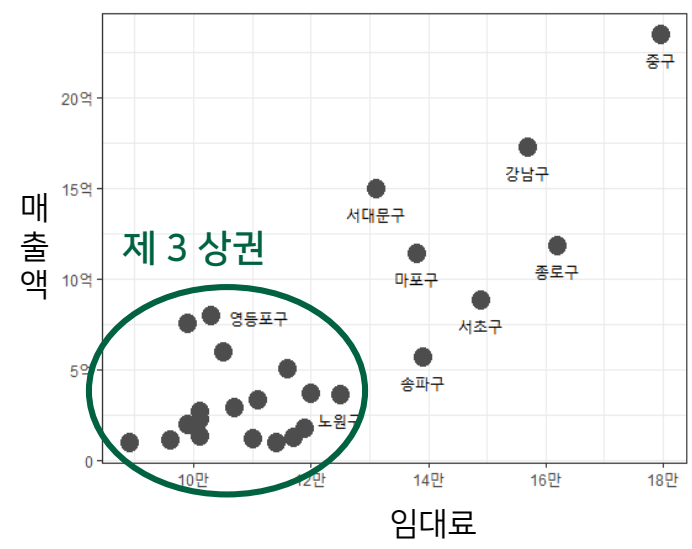
제 3상권 구 중 폐업률이 가장 낮은 강서구 추천



주변 카페 중 프랜차이즈 비율이 약 30~ 40%  
주변 음식점 수 50개 내외인 지역을 선정



총면적이 약 200~300m<sup>2</sup>  
4, 5월 또는 10월에 개업 추천



## [ CASE 1 ]

상암동에 다음과 같은 카페를 차린  
김씨의 폐업 위험도는?

- 서초구
- 2019년 2월 개업
- "지하 1층"
- 총 면적 " $3.04 m^2$ "
- 근처 "181개"의 음식점
- 근처 새롭게 생긴 카페 수는 "3개"
- 여성인구 비율 : 0.52



## [ CASE 2 ]

돈암동 카페거리에 다음과 같은  
카페를 차린 임씨의 폐업 위험도는?

- 성북구 위치
- 2018년 4월 개업
- "1층"
- 총 면적 " $45.11 m^2$ "
- 근처 "437개"의 음식점
- 근처 새롭게 생긴 카페 수는 "4개"
- 여성인구 비율 : 0.53

위험확률 **상**



위험확률 **하**



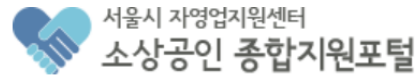
## 차별점

- 최대한 **다양한 변수**를 활용하여, 창업 전 여러가지를 고려할 수 있게 도와줌
- **카페 밀집도** 변수를 활용하여 주변 경쟁 과열 상황을 반영했고, **주변 음식점 수** 변수를 활용 상승 효과를 반영함
- SNS 변수 활용하여 시시각각 변화하는 상권의 **입소문 효과**를 반영 가능함
- 유동인구, 주변 카페 수, 횡단보도 등의 변수로 **발달상권 / 골목상권**을 구분 짓고자 함
- 자본에 따라 행정구를 세분화해서 모델링을 했기 때문에 행정구에 특성마다 폐업률에 중요한 요인을 각각 다르게 반영할 수 있음



## 한계점

- 카페 인테리어, 가격, 사용하는 원두 종류 등을 반영하지 못함
- 블로그 건수 대신 20대가 많이 사용하는 SNS(인★그램 등)를 활용한다면 더욱 정확한 카페 트렌드를 반영할 수 있을 것
- 매출액, 임대료 데이터가 행정동 단위로 존재하여 카페 고유의 값을 반영하지 못함



01

## 교육 자료 활용



- 서울시 자영업지원센터 '소상공인 창업 아카데미' 창업 교육 지원 정책에 활용
- 서울시 '소상공인 정책연구센터' 에서 활용

02

## 최적 입지 추천



- '우리마을 가게 상권분석서비스'에서 이미 서비스하고 있지만, **지역적인 특성만 반영한다는 한계점**
- 본 상권 분석에서는 **카페 고유의 특성과 주변 상권 특성(주변 경쟁 상황 등)도 반영하여 보다 정확도 높게 활용**

03

## 폐업 위험 대비



- 이미 개업한 카페 중 폐업 위험도가 높은 카페에 대하여 추천을 통해 지역 인테리어 업체와 **무료 리모델링 행사**
- 공사 전/후 사진을 SNS에 공개하여 카페와 지역 인테리어 업체 모두 홍보 효과

데이터 목록

데이터명	출처
서울시 강남구 전체 위생업소 현황 서울시 강동구 전체 위생업소 현황 : (25개 행정구) 서울시 중랑구 전체 위생업소 현황	서울 열린데이터 광장
소상공인시장진흥공단 상가업소 정보	공공데이터 포털
서울시 도로명 주소 정보	서울 열린데이터 광장
서울시 횡단보도 위치정보	
서울시 공영주차장 안내 정보	
서울시 역코드로 지하철역 위치 조회	
서울시 우리마을가게 상권분석서비스(상권배후지-집객시설)	
서울시 우리마을가게 상권분석서비스(상권-추정매출)	서울시 우리마을가게 상권분석서비스
지역분석-인구수(유동 인구, 주거인구, 직장인구), 소득/가구수, 임대시세	
KOSPI Composite Index	Yahoo Finance
Naver 블로그 건수	NAVER(웹 크롤링)

참고문헌

- 백다예(2016), 커피 전문점의 입지특성과 업종특성이 점포 수익성에 미치는 영향
- 임소연(2017) 골목상권 내 외식업종 점포의 월 매출액 예측 모형에 관한 연구
- 정동규(2017), 발달상권과 골목상권에 위치한 음식점 생존과 폐업 비교 분석
- 최민정(2017), 딥러닝 알고리즘을 이용한 제주도 관광객 수 예측
- 이다영(2018), 시공간적 패턴에 기반한 서울시 공공자전거 수요 예측

분석도구

- R - dplyr, caret, randomForest, range, h2o, ROSE, RANN, xgboost, e1071, glmnet, geosphere 등
- Python - Numpy , Python, re, Selenium, BeautifulSoup 등