



- 프로젝트 개요
 1.1 프로젝트 개요
 1.2 데이터 범위
- 2. 프로세싱2.1 데이터 수집 및 전처리 과정2.2 데이터 모델
- 3. 분석 결과
 - 지역 간 카페 점포 비교(1)
 - 지역 간 카페 점포 비교(2)
 - 카페와 feature 간 상관계수
 - 자치구 별 카페 feature 상관계수(1)
 - 자치구 별 카페 feature 상관계수(2)
- 4. 기대효과 및 한계점, 개선 사항
- 5. 개발후기 및 느낀점
- 6. 질의응답
- 7. 출처

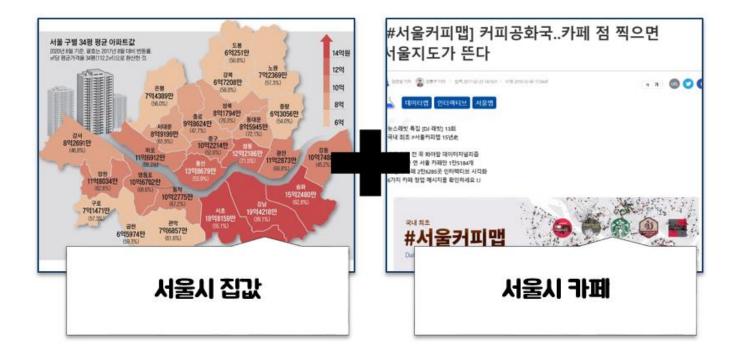
1. 프로젝트 개요



[주제 & 목표] 카페 입지 조건에 따른 서울시 자치구 간 상관 분석

[배경]

- '카페'로 지역 특성을 볼 수 있지 않을까?
- 고가형과 저가형 카페를 통해 지역간 불균형을 볼 수 있을까?



1. 데이터 범위: 강남3구 동북4구 선정 이유



(강남 3구) 강남, 서초, 송파

강남과 강북지역의 공동주택단지 생활 환경 비교 평가 연구

조미정**; 임영진***; 이명훈****

스타벅스 효과가 인근 표준지공시지가에 미치는 영향에 대한 연구

-서울 동북2권을 대상으로-

본 연구는 부동산 시장에서 발생하고 있는 스타벅스 입점에 따른 파급효과에 관심을 두었으며 이를 학문적인 근거 아래 학술적으로 증명하고자한다. 연구 대상 지역은 서울시 생활권 계획상 베드타운으로 제시된 동북 2권(성북구, 노원구, 도봉구, 강북구)을 범위로 설정하였으며, 그 지역에 위치한 스타벅스 카페와 이디야 카페를 중심으로 반경 300m 이내 토지의 표준지공시지가를 중심으로 연구하였다. 연구 목적은 첫 번째,

[동북 4구] 강북, 노윈, 도봉, 성북

1. 배경 및 목적

• 「도시재생 활성화 및 지원에 관한 특별법」(이하 "도시재생특별법"이라 한다.) 제13조(도시 재생 전략계획의 수립)에 따라 창동·상계지역이 활성화지역으로 지정(15,12,10)됨

동북2생활권의 간선버스 노선 수는 서북생활권에 비해 절반 수준

— 3개 생활권과 CE

지하철을 대신하는 간선버스 노선수 역시 최하위권

2호선이 지나지 않는 동북4구, 전철역 수도 최하위권

서울 각 권역을 통과하는 지하철노선도: 2호선이 지나지 않는 동북4구

■ 베드타운이었던 신시가지는 점차 노후화

• 동북부 320만의 입지적 중심이나 베드타운 (주거중심)으로 지역이미지가 쇠락



주최·주관 | **서울시 동북4구 도시재생협력지원센터**

협력 | 동북4구 행정협의회(성북, 강북, 도봉, 노원), 서울시 동북권 NPO센터

2. 프로세싱: 데이터 수집 및 전처리 과정



데이터 수집	데이터 추출	데이터 전처리 방법
아파트 애애가격	아파트 면적 아파트 거래금액 면적 당 거래금액(파생변수)	결측치 제거 이상치 제거 엔코딩(D4:0,G3:1) 구별 평균치
상권정보	카페 별 점포 수 (저가-빽다방,메가커피,컴포스 고가-스타벅스,커피빈,폴바셋)	평균치 계산 엔코딩(저가:0,고가:1)
유동인구수	동별 총 유동인구 수	이상치 제거 네이버API(주소변환)
지하철역 정보	카페와 지하철역 간의 거리(파생변수)	경도,위도 변환
버스정류장 정보	카페반경 160m 내 개수(파생변수)	경도,위도 변환

(2018~2020년, 서울시 기준)

2. 프로세싱: 데이터 모델



Input Output

D4, G3

아파트 면적 당 거래금액(파생변수) 카페 별 점포 수 동별 총 유동인구 수 카페와 지하철역 간의 거리(파생변수) 카페반경 160m 내 버스정류장 개수(파생변수)

저가형 카페 고가형 카페

D4:도시재생활성화 지역(구)

G3:강남3구

3. 데이터 분석

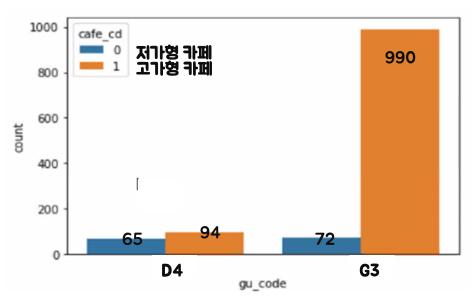


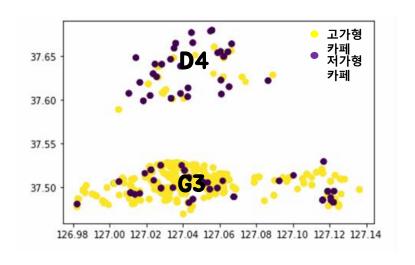
[Feature 정의]

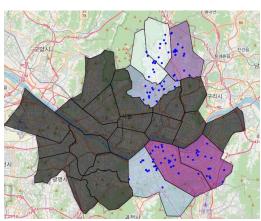
feature	정의
Ing	경도
lat	위도
year	데이터 연도
bus_cnt	카페 반경 160m 내에 있는 버스 정류장의 수
Hub_Dist	카페로부터 지하철역까지의 거리
cafe_cd	카페 유형 encoding(고가형:1, 저가형:0)
pop_avg	평균 유동인구 수
byear	아파트 건축연도
average_sizetrade	면적 당 평균 거래금액
gu_code	서울 지역별 encoding(D4구:0, G3구:1)

3. 데이터 분석: 자치구 간 카페 점포 수 비교(1)









- 저가형 카페 분포

- 고가형 카페 분포

3. 데이터 분석: 자치구 간 카페 점포 수 비교(2)





3. **데이터 분석**: 카페와 feature 간 상관계수



gu_co	average_sizetrade	byear	pop_avg	cafe_cd	HubDist	bus_cnt	year	lat	Ing	
0.0281	-0.068085	-0.225542	-0.009402	-0.053112	-0.022428	-0.306665	0.021520	-0.006193	1.000000	Ing
-0.9499	-0.699179	0.152287	0.077520	-0.349579	0.033124	0.119045	0.070362	1.000000	-0.006193	lat
-0.0667	0.189133	-0.033662	0.748386	-0.136847	0.007798	-0.000246	1.000000	0.070362	0.021520	year
-0.1925	-0.189776	0.201596	-0.016198	0.023901	0.016894	1.000000	-0.000246	0.119045	-0.306665	bus_cnt
-0.0114	0.021190	-0.076566	0.007692	-0.022581	1.000000	0.016894	0.007798	0.033124	-0.022428	HubDist
0.3636	0.275141	0.009806	-0.057371	1.000000	-0.022581	0.023901	-0.136847	-0.349579	-0.053112	cafe_cd
-0.0444	0.204599	-0.017264	1.000000	-0.057371	0.007692	-0.016198	0.748386	0.077520	-0.009402	pop_avg
-0.1432	-0.403069	1.000000	-0.017264	0.009806	-0.076566	0.201596	-0.033662	0.152287	-0.225542	byear
0.7391	1.000000	-0.403069	0.204599	0.275141	0.021190	-0.189776	0.189133	-0.699179	-0.068085	average_sizetrade
1.0000	0.739100	-0.143238	-0.044465	0.363623	-0.011414	-0.192514	-0.066751	-0.949999	0.028104	gu_code
0.0110	0.014686	0.006545	-0.043408	0.010178	0.015206	0.012414	-0.041658	-0.006119	0.016655	cafe_cnt

- 자치구와 양의 상관관계

- 면적당 아파트 거래 금액과 양의 상관관계

-Encoding-

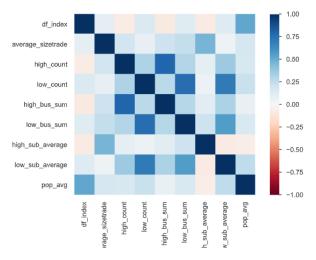
자치구: D4 -0, G3-1

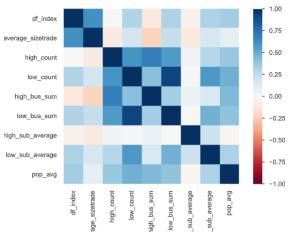
카페 : 저가 - 0, 고가-1

3.데이터분석: 자치구 별 카페 feature 상관계수(1)



〈Kendall 상관계수〉





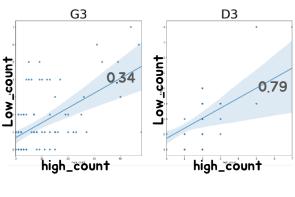
[16]:	corr=g3.corr()			
[17]:	print(corr)			
		average_sizetrade	high_count	low_count
	average_sizetrade	1.000000	0.152302	0.087590
	high_count	0.152302	1.000000	0.338915
	low_count	0.087590	0.338915	1.000000
	high_bus_sum	0.060388	0.900350	0.307803
	low_bus_sum	0.347714	0.243107	0.713672
	high_sub_average	0.647057	-0.049132	0.098698
	low_sub_average	0.008160	0.512126	0.609201
	pop_avg	0.204880	0.186569	0.231860

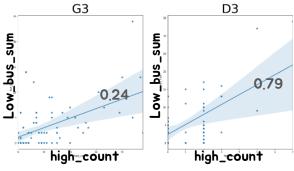
[22]:	corr=bed4.corr()			
[23]:	print(corr)			
		average_sizetrade	high_count	low_count h
	average_sizetrade	1.000000	-0.128806	0.082813
	high_count	-0.128806	1.000000	0.790772
	low_count	0.082813	0.790772	1.000000
	high_bus_sum	-0.301967	0.790257	0.627791
	low_bus_sum	0.104859	0.785778	0.953995
	high_sub_average	-0.165597	-0.014200	-0.030411
	low_sub_average	0.138375	0.210799	0.398292
	pop_avg	0.141883	0.428084	0.514784

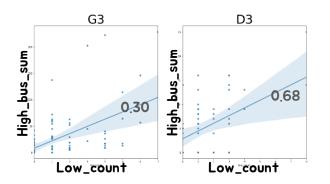
high_count - 고가형 카페 수 low_count - 저가형 카페 수 low_bus_~ - D4버스정류장 high_bus_~ - G3 버스정류장

3.데이터분석: 자치구 별 카페 feature 상관계수(2)









[16]:	corr=g3.corr()			
[17]:	print(corr)			
		average_sizetrade	high_count	low_count h
	average_sizetrade	1.000000	0.152302	0.087590
	high_count	0.152302	1.000000	0.338915
	low_count	0.087590	0.338915	1.000000
	high_bus_sum	0.060388	0.900350	0.307803
	low_bus_sum	0.347714	0.243107	0.713672
	high_sub_average	0.647057	-0.049132	0.098698
	low_sub_average	0.008160	0.512126	0.609201
	pop_avg	0.204880	0.186569	0.231860

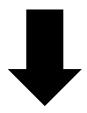
[22]:	corr=bed4.corr()			
[23]:	print(corr)			
		average sizetrade	high count	low count
	average_sizetrade	1.000000	-0.128806	0.082813
	high_count	-0.128806	1.000000	0.790772
	low_count	0.082813	0.790772	1.000000
	high_bus_sum	-0.301967	0.790257	0.627791
	low_bus_sum	0.104859	0.785778	0.953995
	high_sub_average	-0.165597	-0.014200	-0.030411
	low_sub_average	0.138375	0.210799	0.398292
	pop_avg	0.141883	0.428084	0.514784

high_count - 고가형 카페 수 low_count - 저가형 카페 수 low_bus_~ - D4버스정류장 high_bus_~ - G3 버스정류장

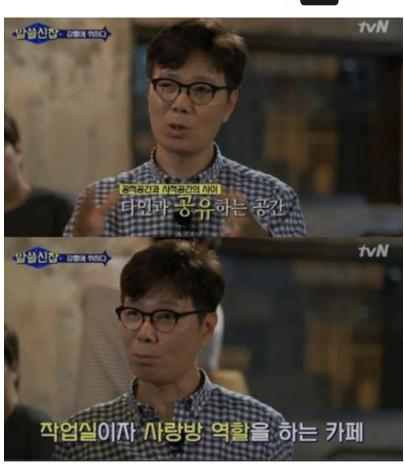
4. 기대 효과



- 서울 자치구 간 카페 점포 수 차이
- 서울 자치구 간 아파트 매매 가격의 차이
- 서울 자치구 간 대중교통 접근성 차이



- 카페의 가격 정책 별 입점 지역 제안 가능
- 문화공간으로서 <mark>카페 역할 확장</mark> -> 지역 균형 발전 요소로 사용 가능
- 자치구 간 불균형이 존재함
 - -> 정부에 지역 간 균형 잡힌 발전을 위한 개발 정책의 필요성을 제안



4. 한계점 및 향후 개선 사항



- 유동인구 데이터에서 인구통계학적 특성(예:나이, 성별 등)을 반영하지 못함
- 카페의 입점 위치를 결정하는 요인을 모두 고려하지 못함
 - → 향후 대학교의 수, 상가 개수 등 추가 및 분석 필요 (+ 요인)
 - → 연령, 카페 밀집도 추가 및 분석 필요 (요인)
- 유동인구에서 극단치로 높은 값을 보이는 송파구의 데이터를 제거한 후, 유동인구와 카페 점포 수에 차이가 있는지 추가적인 분석이 필요

 서울의 특정 구(G3와 D4) 의 비교를 통해 지역 간의 불균형을 분석하였으나, 분석범위를 확장하여 현 모델이 다른 지역에서도 적용 가능한지 확인을 못하였으며 실질적인 불균형 해소방안을 제안하지 못함.

5. 개발 후기 및 느낀점



이름	개발후기 및 느낀점
강주영	실제 데이터를 통해 전 처리 및 시각화를 해보니, 코드 실력이 따라와주지 않아 너무 힘들었습니다. 다음 프로젝트에는 가능하면 (비)전공자가 적절하게 섞여 프로젝트 할 수 있으면 좋겠습니다.
김민형	데이터로 하는 팀프로젝트는 처음이라서 많이 헤매기도 했고 논리적 설명이나 코딩부분에서 막 형서 힘들었던 경험도 종종 있었습니다. 그러나 팀원분들이 프로젝트에 진심을 다해서 임해주셨 기에 좋은 결말을 맺은 거 같습니다.
정길종	과정 중 첫 과제를 좋은 분들과 함께 할 수 있어 좋았습니다. 과제를 통해 배웠던 것을 복습하고 오류를 수정하는 과정에서 많은 공부가 되었습니다.
한유정	프로젝트를 처음 해볼 뿐만 아니라 코딩이나 서칭에 미숙하니깐 어려움이 많았습니다. 그래도 같이 하는 조원분들께서 이해해주고 제 의견에도 많은 관심을 주셔서 좀 더 재미있게 프로젝트 에 임할 수 있었던 것 같습니다. 그래도 다음 프로젝트 때에는 전공자와 비전공자가 함께 할 수 있었으면 좋겠습니다.

6. 질의응답





7. 출처



◆ 저가(컴포즈, 빽다방, 메가커피), 고가형(스타벅스, 커피빈, 폴바셋) 카페 자료 - 카페 위치

https://www.data.go.kr/data/15083033/fileData.do 공공데이터 포털,한국 소상공인시장진흥공단

- ◆ 아파트(매매)_실거래가 자료
 - 면적당 거래 금액

http://rtdown.molit.go.kr/ 국토교통부의 실거래가 공개시스템

- ◆ 서울시 우리마을가게 상권분석서비스(상권배후지-추정유동인구)
 - 행정구 별 유동인구

https://www.data.go.kr/data/15076417/fileData.do 공공데이터 포털,서울특별시

- ◆ 서울교통공사 노선별 지하철역 정보
 - 지하철 역 정보

http://data.seoul.go.kr/dataList/OA-15442/S/1/datasetView.do 서울교통공사 스마트정보처에서 제공

- ◆ 서울특별시 버스정류소 위치정보
 - 버스 정류장 정보

http://data.seoul.go.kr/dataList/OA-15067/S/1/datasetView.do 서울특별시 도시교통실 보행친화기획관 교통정보과에서 제공