

KT AIVLE School

1차 미니프로젝트

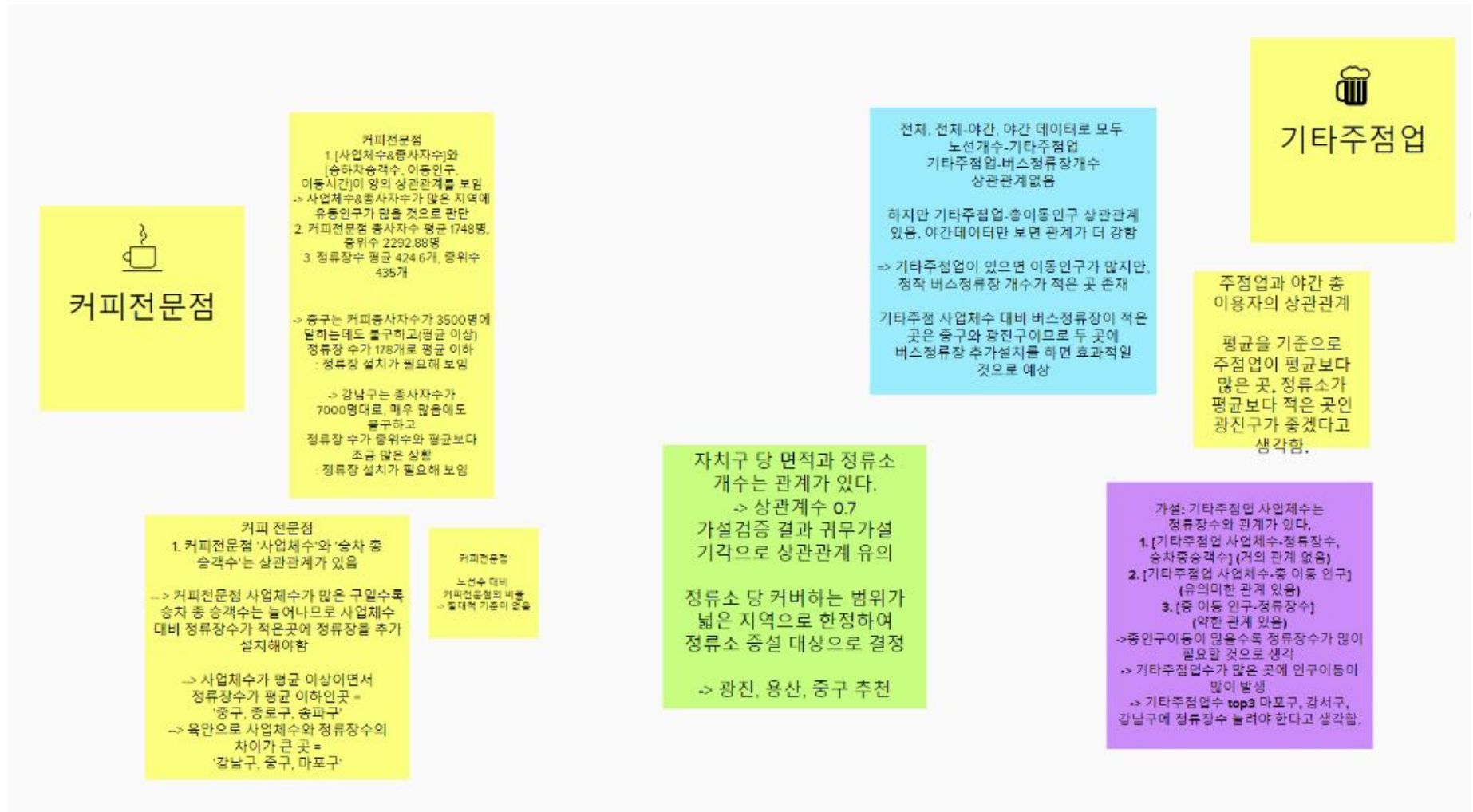
AI 수도권 2반 7조



팀원 별 분석 진행



팀원 별 분석 진행



가설 수립

- ✓ 가설 1 : 커피전문점 종사자수가 많은 지역에 더 많은 정류장이 필요할 것이다.
- ✓ 가설 2 : 밤에 술을 많이 마시니, 기타주점 사업체수와 버스정류장 수가 비례관계에 있을 것이다.
- ✓ 가설 3 : 자치구 당 면적과 버스정류장 개수는 서로 상관이 있을 것이다.

커피전문점

커피전문점과 정류장 수의 관련성

1. 커피전문점 분석 계기

[1-4 업종 분석 실습] 종사자수가 가장 많은 Top3 직종

1. 한식일반음식점업 2. 택시운송업 3. 커피전문점

=> 커피 전문점수가 많은 자치구로 많은 인구가 오갈 것으로 예상하였고,
커피전문점이 많은 지역에 정류장이 많이 필요할 것이라는 가설을 세움

2. 1-1~1-4 실습에서 주요 feature만 선정

[도착 시군구 코드/자치구/정류장수/노선수/승차총승객수/하차총승객수/평균 이동 시간(분)/

이동인구(합)/인구/세대/세대당인구/65세이상고령자/커피전문점 사업체수/커피전문점 /종사자수]

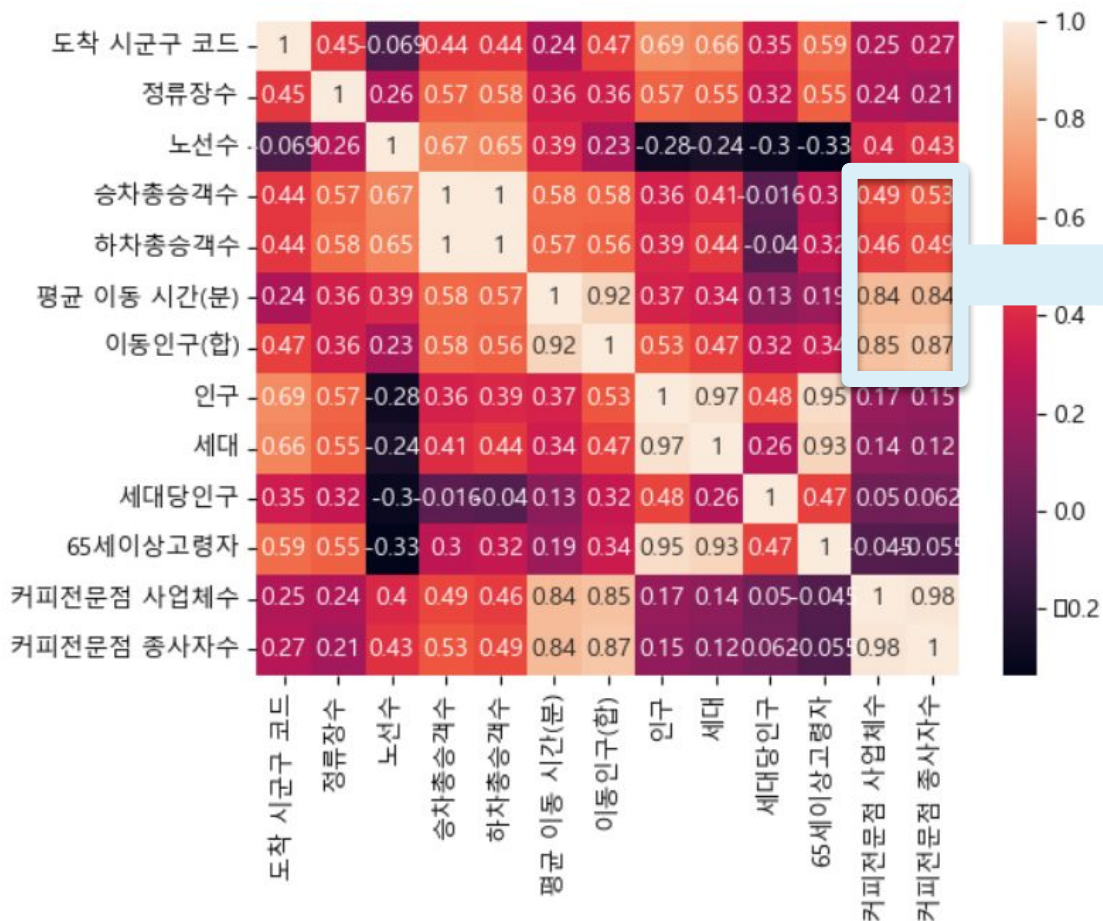
커피전문점과 정류장 수의 관련성

3. 도착시군구코드 및 자치구 기준 데이터 merge

	도착 시군구 코드	자치 구	정류장 수	노선 수	승차총승객 수	하차총승객 수	평균 이동 시간 (분)	이동인구(합)	인구	세대	세대당 인구	65세이상고령자	커피전문점 업체수	커피전문점 종사자수
0	11010	종로구	349	106	5468935	5205394	9356120	19385709.60	153684	73766	1.96	28064	948	3613
1	11020	중구	178	103	3776675	3598932	9306020	19715322.61	131450	63767	1.92	24742	869	3542
2	11030	용산구	323	67	4189977	4086474	8514380	14707082.87	236518	111207	2.00	39461	634	2076
3	11040	성동구	435	58	2807280	2746738	8887140	17449273.95	291609	134533	2.12	46896	547	1859
4	11050	광진구	269	41	2749448	2753411	8625540	17947278.99	353110	169483	2.01	52604	590	1887
5	11060	동대문구	306	73	5240565	5115379	8838130	16288077.28	352592	169695	1.99	62963	494	1472

커피전문점과 정류장 수의 관련성

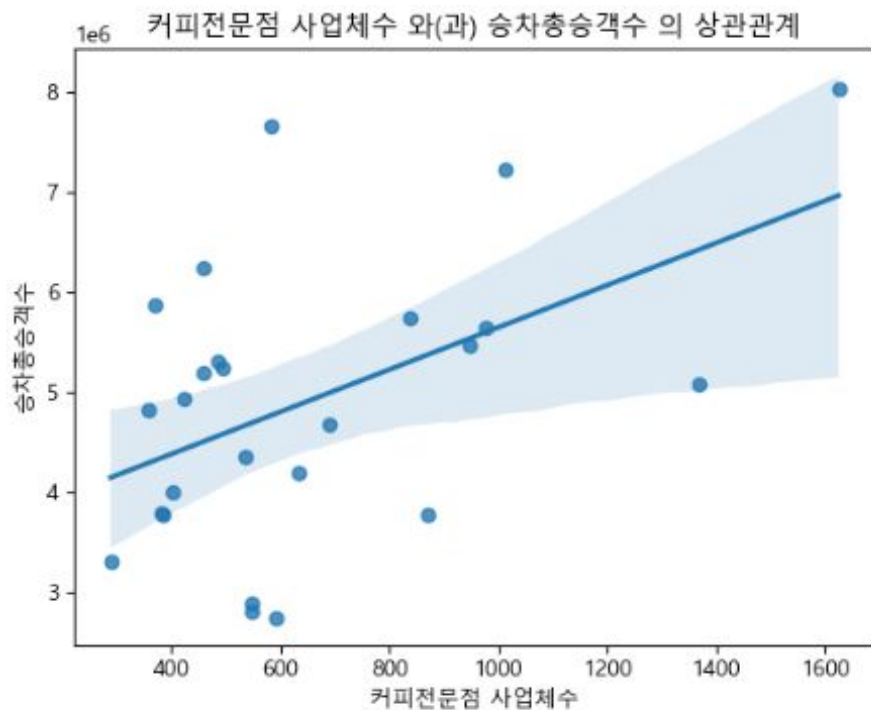
4. 해당 데이터의 heatmap 분석 결과



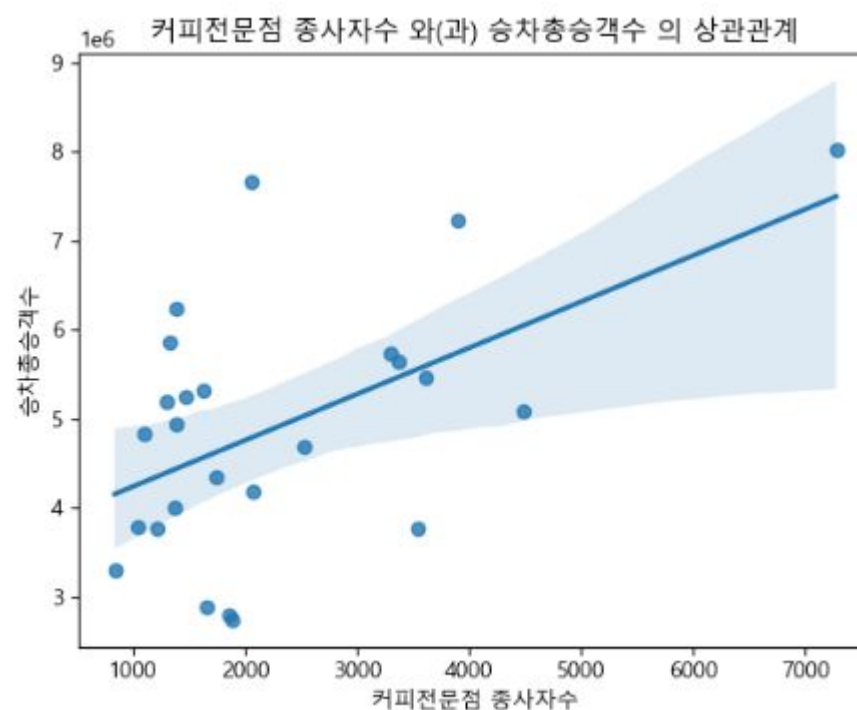
커피전문점 사업체수/종사자수와
승하차승객수, 이동시간, 이동인구에
유의미한 상관관계가 있는 것으로 보임

이변량 분석

1.1. (1) 커피전문점 사업체수 & 종사자수 - 승차총승객수 상관관계



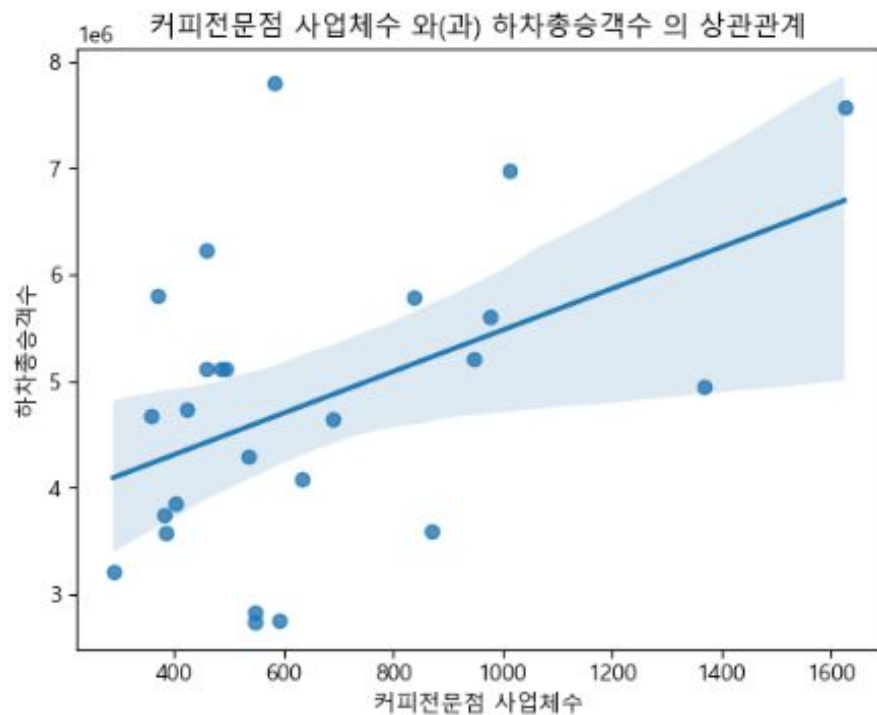
상관계수 : 0.4881692366349319, p-value : 0.013295050925791798



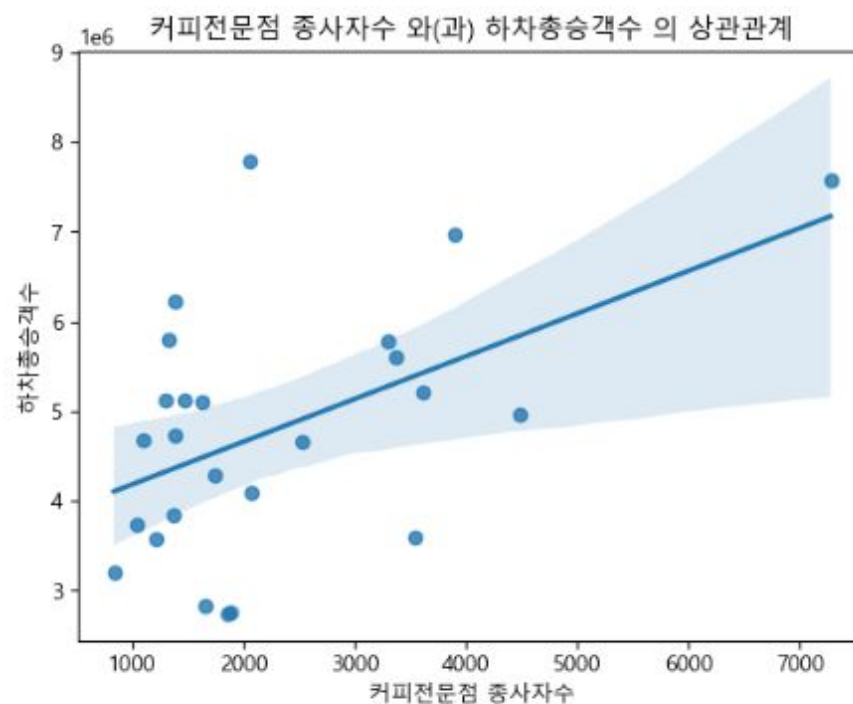
상관계수 : 0.5303859652398508, p-value : 0.006383736312117321

이변량 분석

1.1. (2) 커피전문점 사업체수 & 종사자수 - 하차총승객수 상관관계



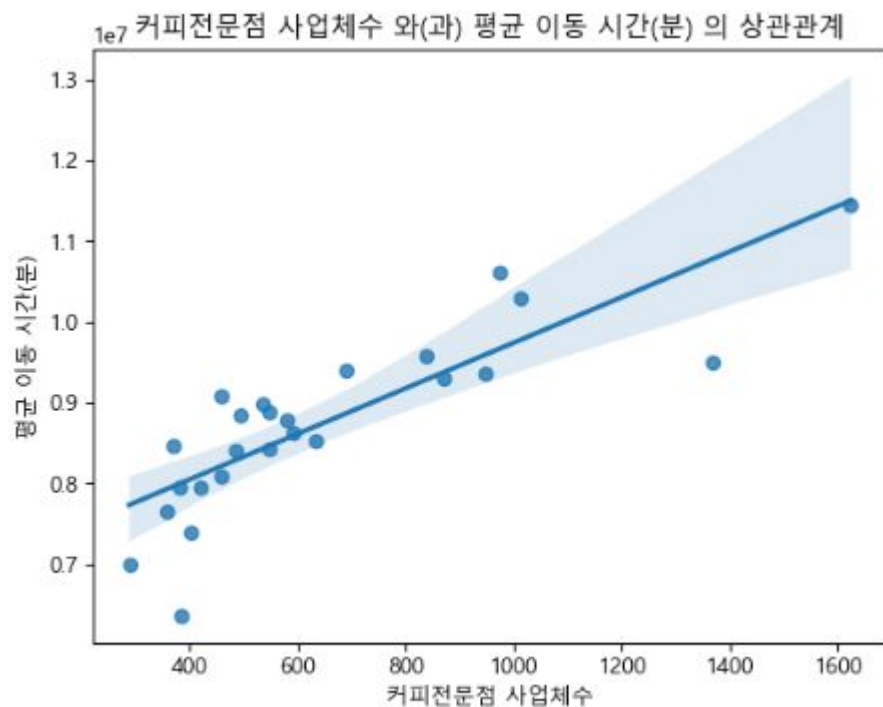
상관계수 : 0.4571394442668359, p-value : 0.021593978868975967



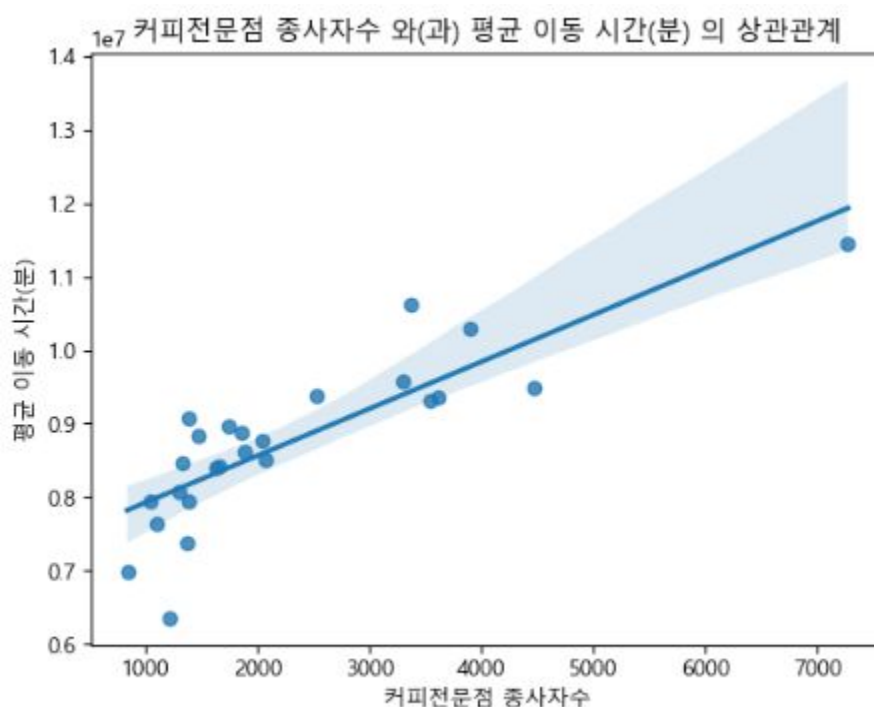
상관계수 : 0.4942655091300091, p-value : 0.012024441735292326

이변량 분석

1.2 커피전문점 사업체수 & 종사자수 - 평균 이동 시간(분) 상관관계



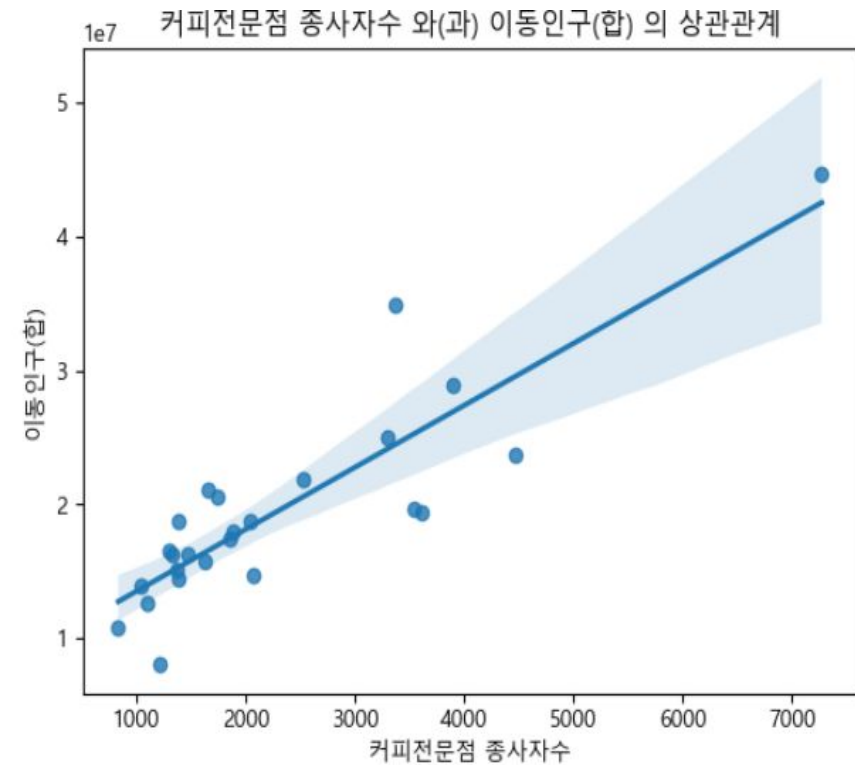
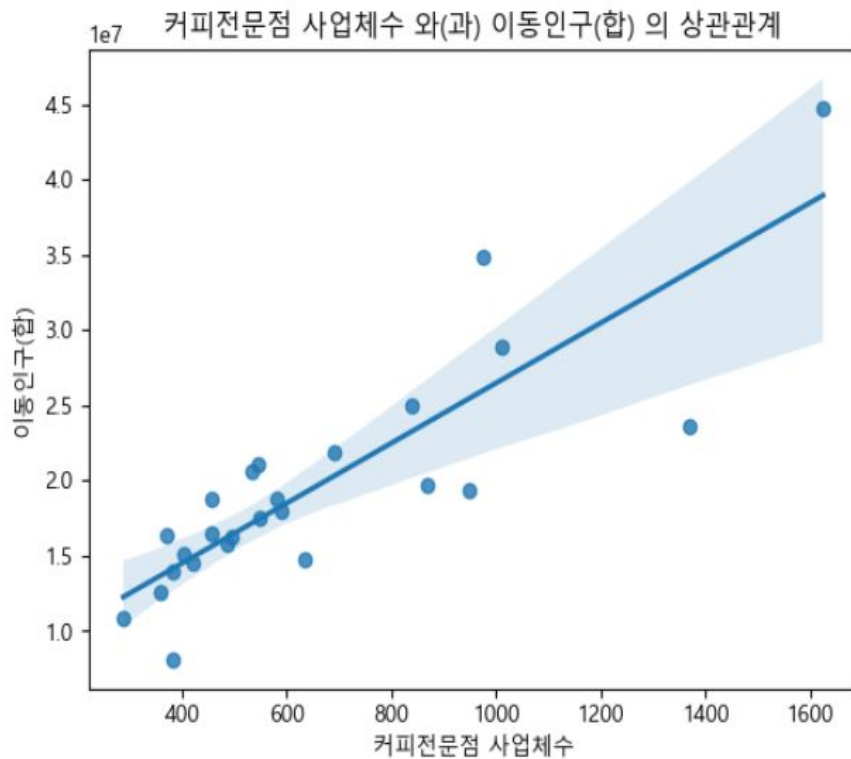
상관계수 : 0.836271130797342, p-value : 1.92543519525677e-07



상관계수 : 0.8357724842802066, p-value : 1.988802094603936e-07

이변량 분석

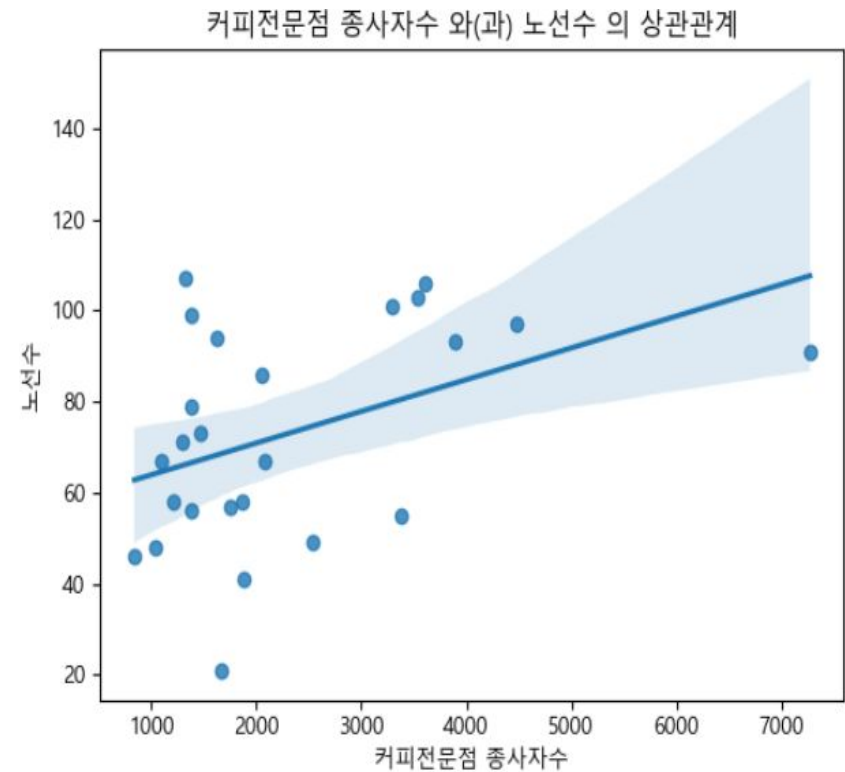
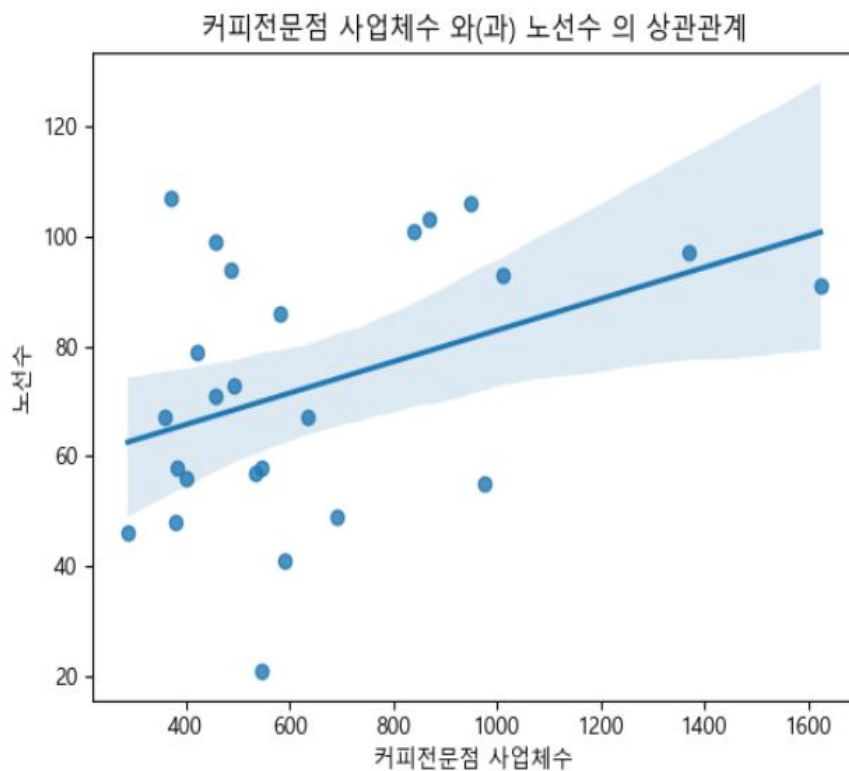
1.3 커피전문점 사업체수 & 종사자수 - 이동인구(합) 상관관계



상관계수 : 0.8546507402767236, p-value : 5.384215881994175e 상관계수 : 0.8736562838487263, p-value : 1.184607858712529e-08

이변량 분석

1.4. (1) 커피전문점 사업체수 & 종사자수 상관관계 - 노선수

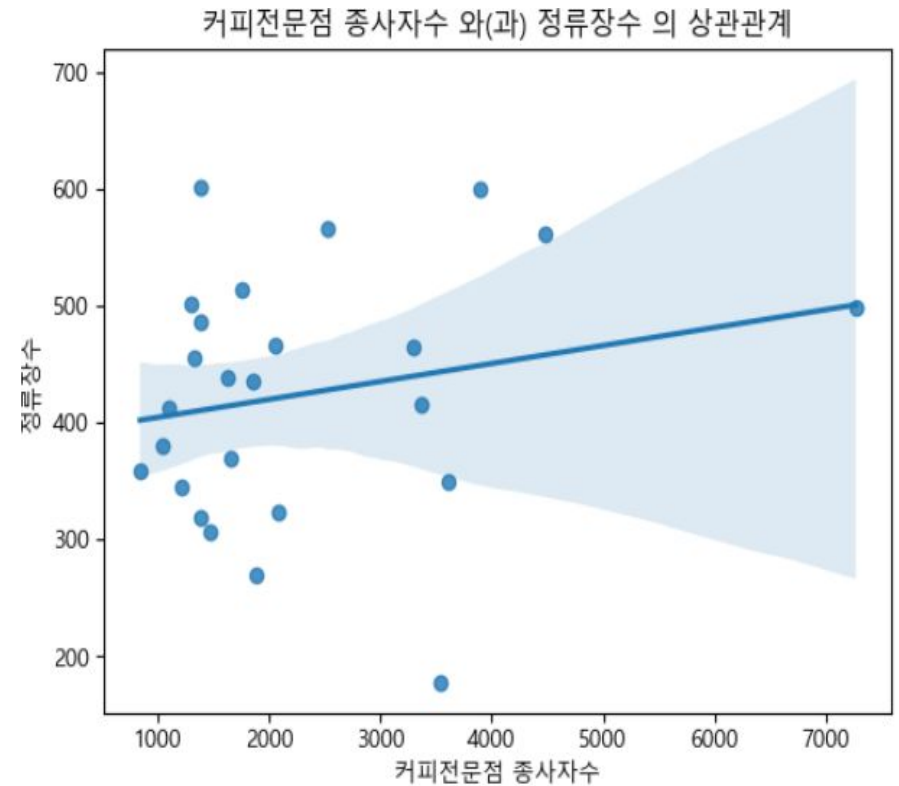
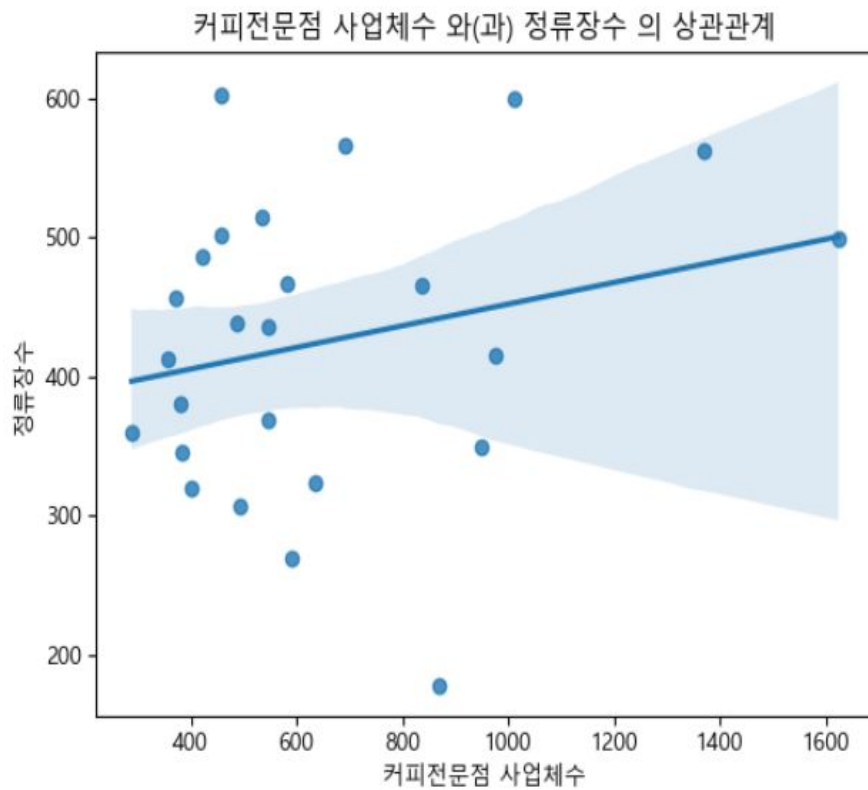


상관계수 : 0.39527392470061734, p-value : 0.05050405934174307

상관계수 : 0.42518048108643286, p-value : 0.03410811239601253

이변량 분석

1.4. (2) 커피전문점 사업체수 & 종사자수 상관관계 - 정류장수



상관계수 : 0.24064271378327762, p-value : 0.24656211167736675 상관계수 : 0.20980792269322665, p-value : 0.3141301127923003

분석 결과

- [커피전문점 사업체수 및 종사자 수]와 [승하차 승객수]
: 약 0.5 정도의 양의 상관관계 (p-value 0.02 이하)
- [커피전문점 사업체수 및 종사자 수]와 [평균 이동시간 및 이동인구의 수]
: 0.8 이상의 강한 상관관계 (p-value 유의미하게 작음)

즉, 커피전문점 사업체 수와 종사자의 수가 많을 수록, 해당 지역의 유동인구가 많을 것이라고 판단

-> 커피전문점 종사자수를 기준으로 정류장 수를 파악해보자

분석 결과

1. 중구

도착 시군구 코드	자치구	정류장 수	노선 수	승차총승객 수	하차총승객 수	평균 이동 시간 (분)	이동인구 (합)	인구	세대	세대당인구	65세이상고령자	커피전문점 사업체수	커피전문점 종사자수	
1	11020	중구	178	103	3776675	3598932	9306020	19715322.61	131450	63767	1.92	24742	869	3542
4	11050	광진구	269	41	2749448	2753411	8625540	17947278.99	353110	169483	2.01	52604	590	1887
5	11060	동대문구	306	73	5240565	5115379	8838130	16288077.28	352592	169695	1.99	62963	494	1472
14	11150	양천구	319	56	4009172	3851173	7379950	15069643.56	449378	181765	2.45	69953	402	1375
2	11030	용산구	323	67	4189977	4086474	8514380	14707082.87	236518	111207	2.00	39461	634	2076

서초구, 마포구 : 종사자수 3000~4000명, 정류장 수 600개

=> 커피전문점 종사자수와 정류장 수 모두 높음

중구 : 종사자수 3500명, 그러나 정류장 수가 178개로 적은 상황

=> 중구에 추가 정류장이 필요

분석 결과

2. 강남구

	도착 시군구 코드	자치구	정류장 수	노선 수	승차총승객 수	하차총승객 수	평균 이동 시간 (분)	이동인구 (합)	인구	세대	세대당인구	65세이상고령자	커피전문점 사업체 수	커피전문점 종사자 수
22	11230	강남구	499	91	8030483	7569213	11437560	44684651.43	537338	234612	2.27	79575	1624	7273
13	11140	마포구	562	97	5088461	4957596	9494790	23639807.30	378629	180702	2.04	55168	1369	4475
21	11220	서초구	600	93	7221330	6977950	10291660	28911318.58	413677	169895	2.41	61404	1012	3897
0	11010	종로구	349	106	5468935	5205394	9356120	19385709.60	153684	73766	1.96	28064	948	3613
1	11020	중구	178	103	3776675	3598932	9306020	19715322.61	131450	63767	1.92	24742	869	3542

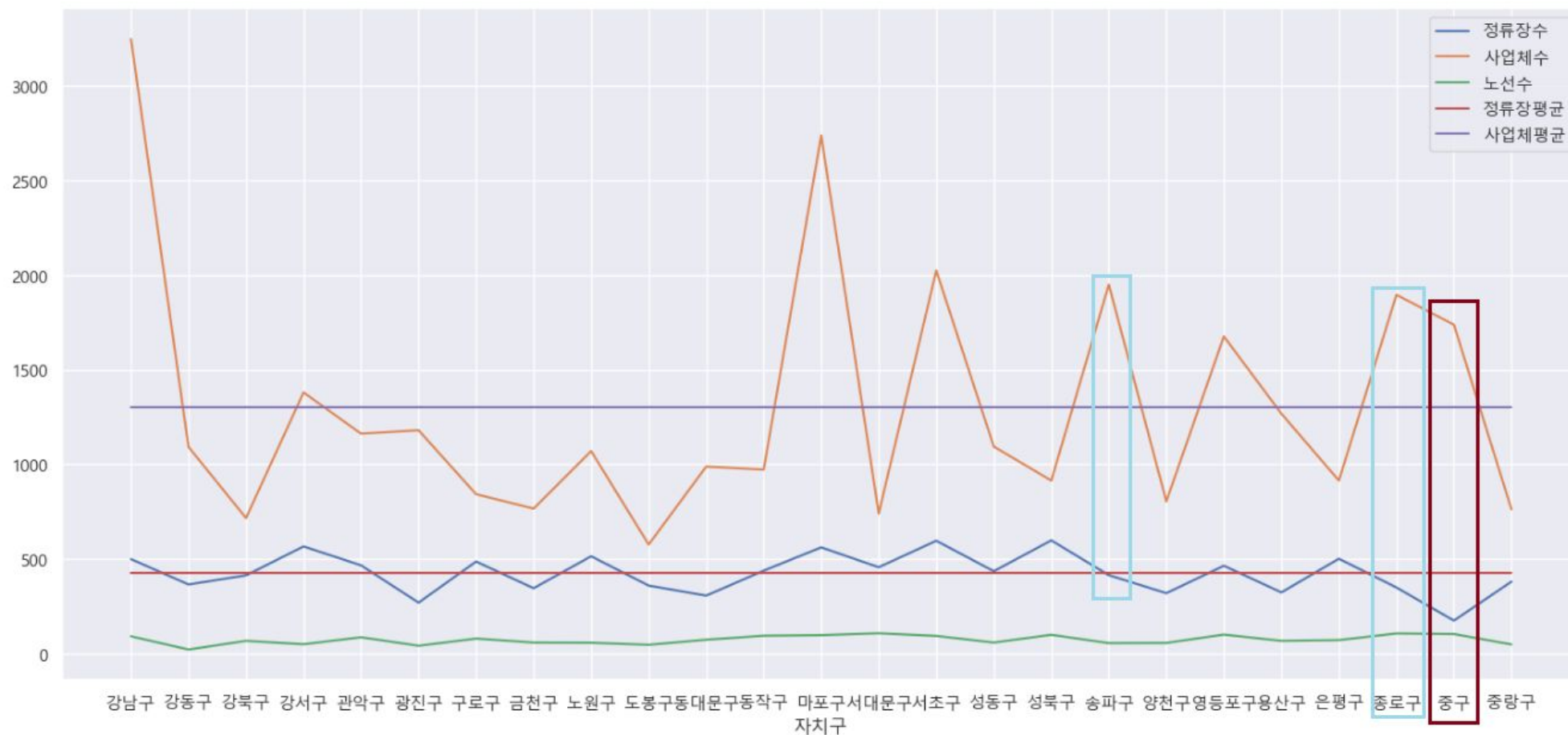
강남구 : 커피전문점 종사자수 7000명 = 중위수 약 1700, 평균 약 2300에 비해 매우 큰 상황

그러나, 정류장 수 = 중위수와 평균 420~430 보다 조금 더 많은 상황

=> 커피전문점 종사자가 많은
즉, **유동인구가 많은 강남구에 정류장 추가 증설이 필요**

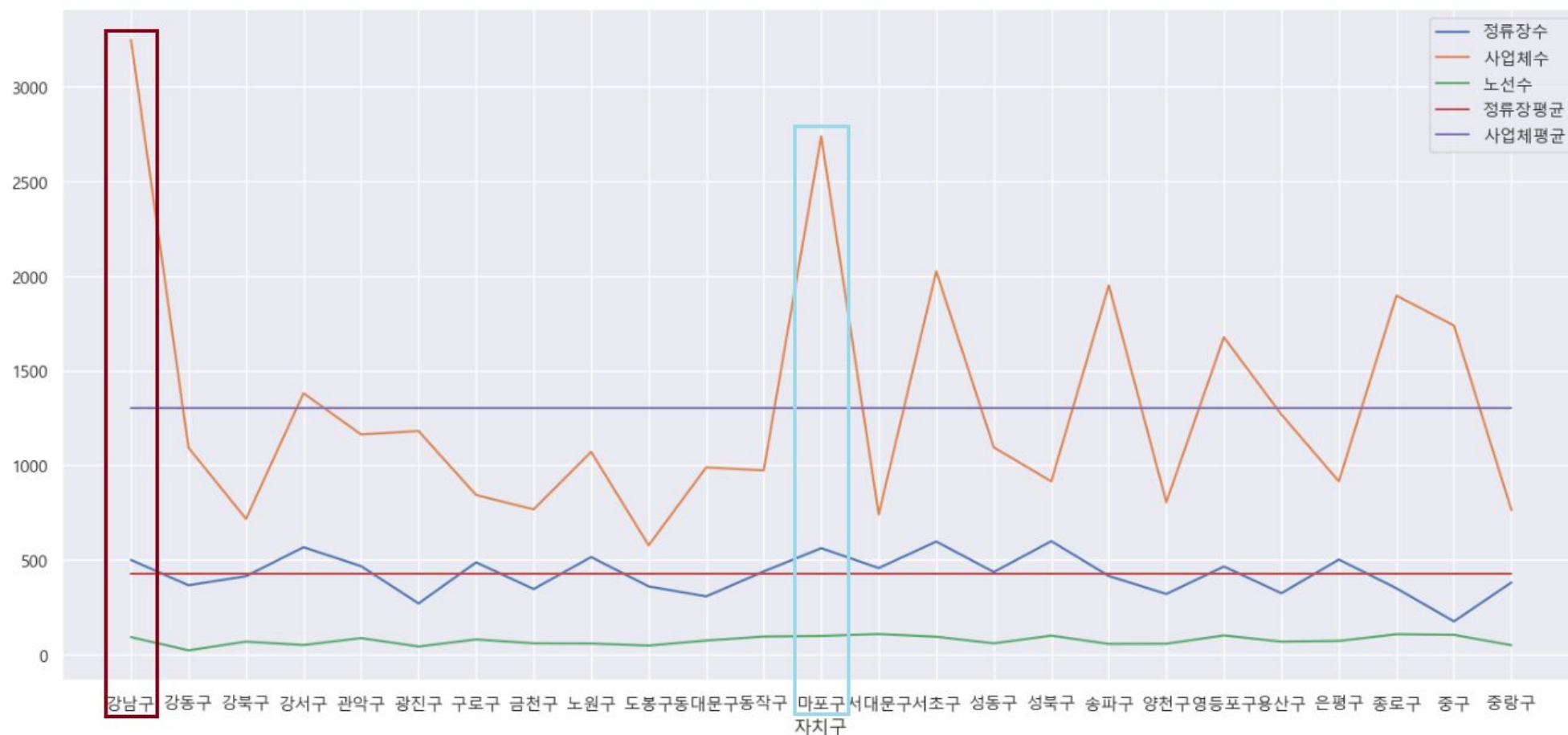
분석 결과

커피전문점 ‘사업체수’가 평균 이상이면서 ‘정류장수’가 평균 이하인 구



분석 결과

커피전문점 ‘사업체수’와 ‘정류장수’의 차이가 커보이는 구



기타주점업

개요

✓ 기타주점업과 관련된 데이터를 선택한 이유

- 밤에 술을 많이 마시니, 자차를 운전을 해서 이동하지 않을 것이라고 생각
- 버스의 탑승률이 올라간다
- 정류장 개수에 영향을 줄 수 있을 수 있겠다.

개요

✓ 순서

- 주간만 포함되어있는 데이터였기에 기타주점업 vs 정류장 개수의 분석결과가 상관관계가 거의 없는 것으로 나왔음
- 술을 주로 밤에 마시니, 야간 데이터가 더 유효하다고 생각, 추출 후 접목
- 기타주점사업체수와 총 이동인구수에 비례 관계를 발견
- 총 이동인구수와 승하차승객수와 비례관계
- 승하차승객수와 정류장 수 비례관계
- 고로, 기타주점사업체가 많은 곳에 정류장의 개수가 많아야 할 것이다!
- 부족한 곳에 정류장을 추가로 세워주면 되겠구나!

기타주점업

1. 1-2에서 야간데이터(0~4)만 따로 추출
 - a. 총 야간 이동시간(분), 총 야간 이동인구(합)
 - b. 평균 야간 이동시간(분), 평균 야간 이동인구(합)

2. 1-5 데이터프레임에 merge

	총 야간 이동시 간(분)	총 야간 이 동인구 (합)	평균 야간 이동시간 (분)	평균 야간 이동인구 (합)	자 치 구	세대	인구총 합계	남자합 계	여자합 계	한국인 합계	...	승차총 승객수	하차총 승객수	승차평균승 객수	하차평균승 객수	도착 시군 구 코드	총 이동 시간	총 이동인구 (합)	평균 이동 시간(분)	이동인구 (합)	종사 자수
0	697590	413951.53	21.538533	12.781015	종로구	73766	153684	74018	79666	144275	...	5468935	5205394	131.889620	125.534028	11010	8538540	19343343.60	20.826215	47.180038	16494
1	777630	549760.32	21.554133	15.238104	중구	63767	131450	63869	67581	122173	...	3776675	3598932	121.020124	115.324511	11020	8496210	19673357.11	20.477532	47.416648	40018
2	786770	529868.37	22.570715	15.200768	용산구	111207	236518	114571	121947	222169	...	4189977	4086474	137.891694	134.485421	11030	7700390	14664517.37	20.803600	39.618091	11449

기타주점업

3. 정류장과 필요한 요소간 Heatmap 작성



가설 검증 과정

■ 상관계수

```
print(spst.pearsonr(df['노선개수'],df['기타주점업']))
print(spst.pearsonr(df['기타주점업'],df['버스정류장개수']))
print(spst.pearsonr(df['기타주점업'],df['총 이동 인구']))
```

주간 데이터

```
PearsonRResult(statistic=0.14732845674186884, pvalue=0.4821937415104064)
PearsonRResult(statistic=0.36697158696218213, pvalue=0.07115719485759615)
PearsonRResult(statistic=0.446980486978646, pvalue=0.02508264851131259)
```

```
print(spst.pearsonr(df['노선개수'],df['기타주점업']))
print(spst.pearsonr(df['기타주점업'],df['버스정류장개수']))
print(spst.pearsonr(df['기타주점업'],df['총 이동 인구']))
```

전체 데이터

```
PearsonRResult(statistic=0.14732845674186884, pvalue=0.4821937415104064)
PearsonRResult(statistic=0.36697158696218213, pvalue=0.07115719485759615)
PearsonRResult(statistic=0.4520919275069978, pvalue=0.023274505465892152)
```

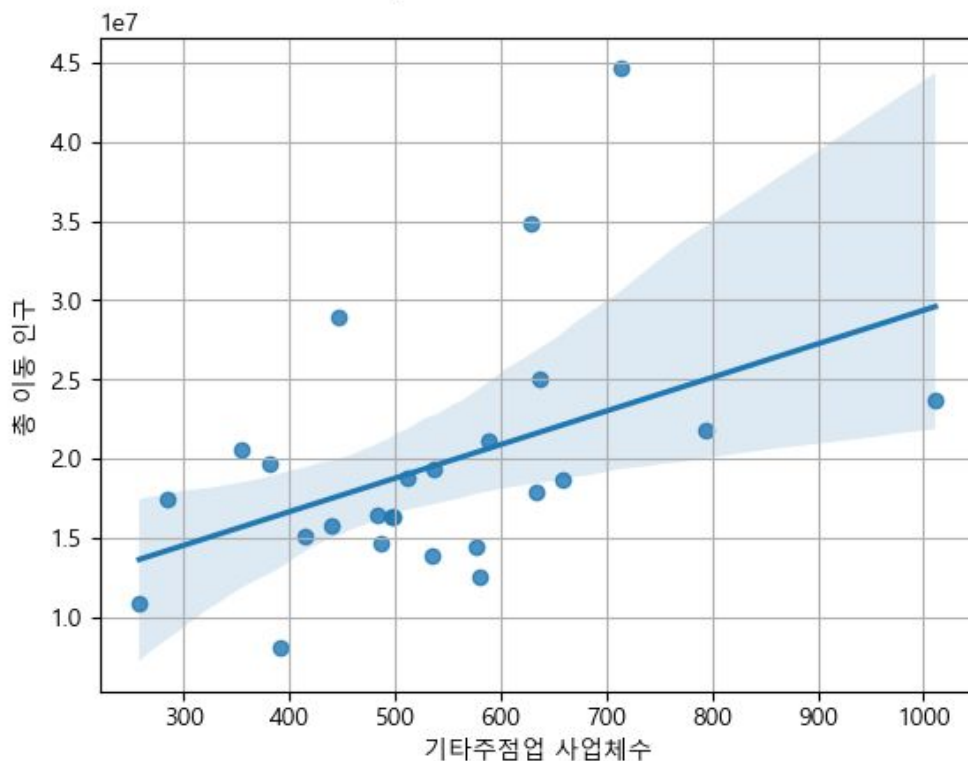
```
print(spst.pearsonr(df['노선개수'],df['기타주점업']))
print(spst.pearsonr(df['기타주점업'],df['버스정류장개수']))
print(spst.pearsonr(df['기타주점업'],df['총 이동 인구']))
```

야간 데이터

```
PearsonRResult(statistic=0.14732845674186884, pvalue=0.4821937415104064)
PearsonRResult(statistic=0.36697158696218213, pvalue=0.07115719485759615)
PearsonRResult(statistic=0.5590418346960516, pvalue=0.0036717238657614027)
```

이변량 분석

✓ 기타주점업 사업체수 - 총 이동 인구수 상관관계

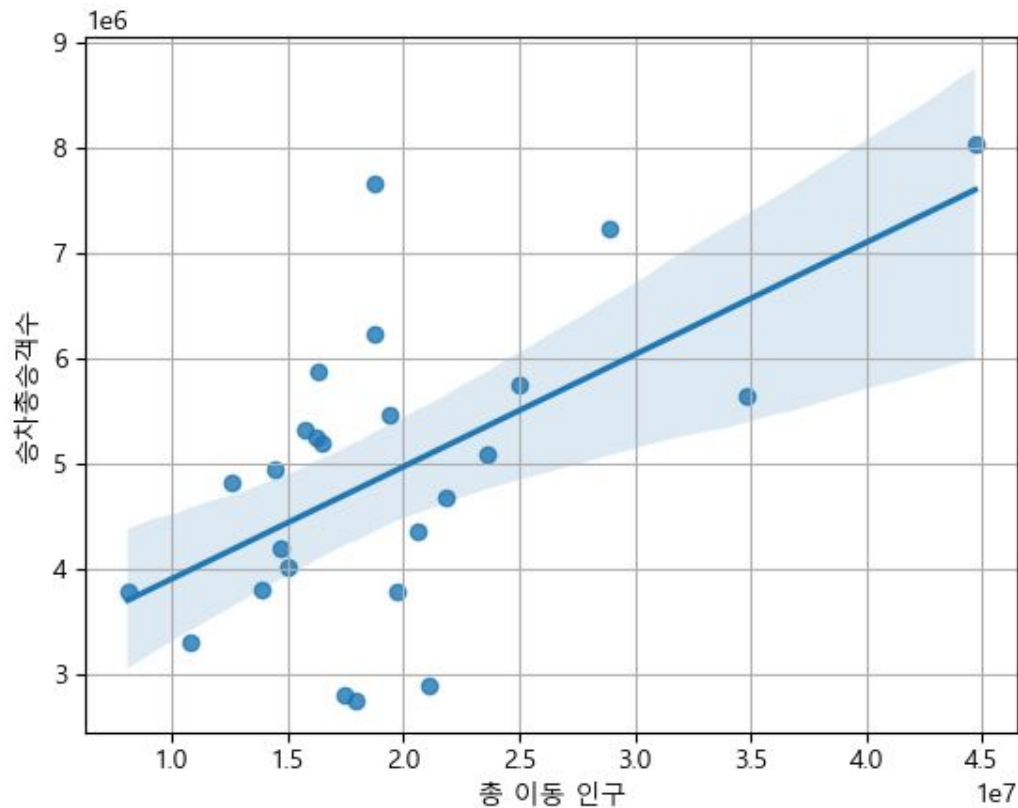


```
sns.regplot(df ["기타주점업 사업체수"], df["총 이동 인구"])
plt.xlabel("기타주점업 사업체수")
plt.ylabel("총 이동 인구")
plt.grid()
plt.show()
result = spst.pearsonr(df["기타주점업 사업체수"],df["총 이동 인구"])
print(f'상관계수 : {result[0]}, p-value : {result[1]}')
```

상관계수 : 0.446980486978646, p-value : 0.02508264851131259

이변량 분석

✓ 총이동인구 - 승차총승객수 상관관계



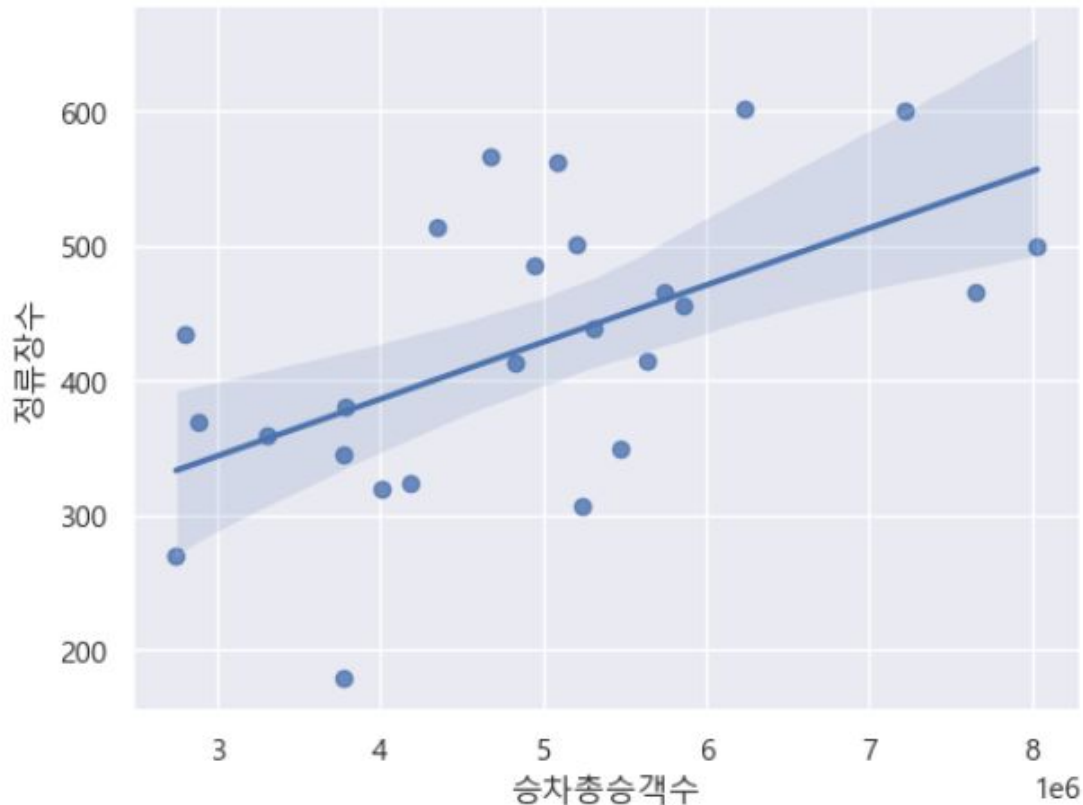
```
sns.regplot(df ["총 이동 인구"], df["승차총승객수"])
plt.xlabel("총 이동 인구")
plt.ylabel("승차총승객수")
plt.grid()
plt.show()
result = spst.pearsonr(df["총 이동 인구"],df["승차총승객수"])
print(f'상관계수 : {result[0]}, p-value : {result[1]}')
```

상관계수 : 0.5765529811137536, p-value : 0.002554753067686566

이변량 분석

✓ 승차총승객수와 상관관계 - 정류장 수

승차총승객수 VS 정류장수

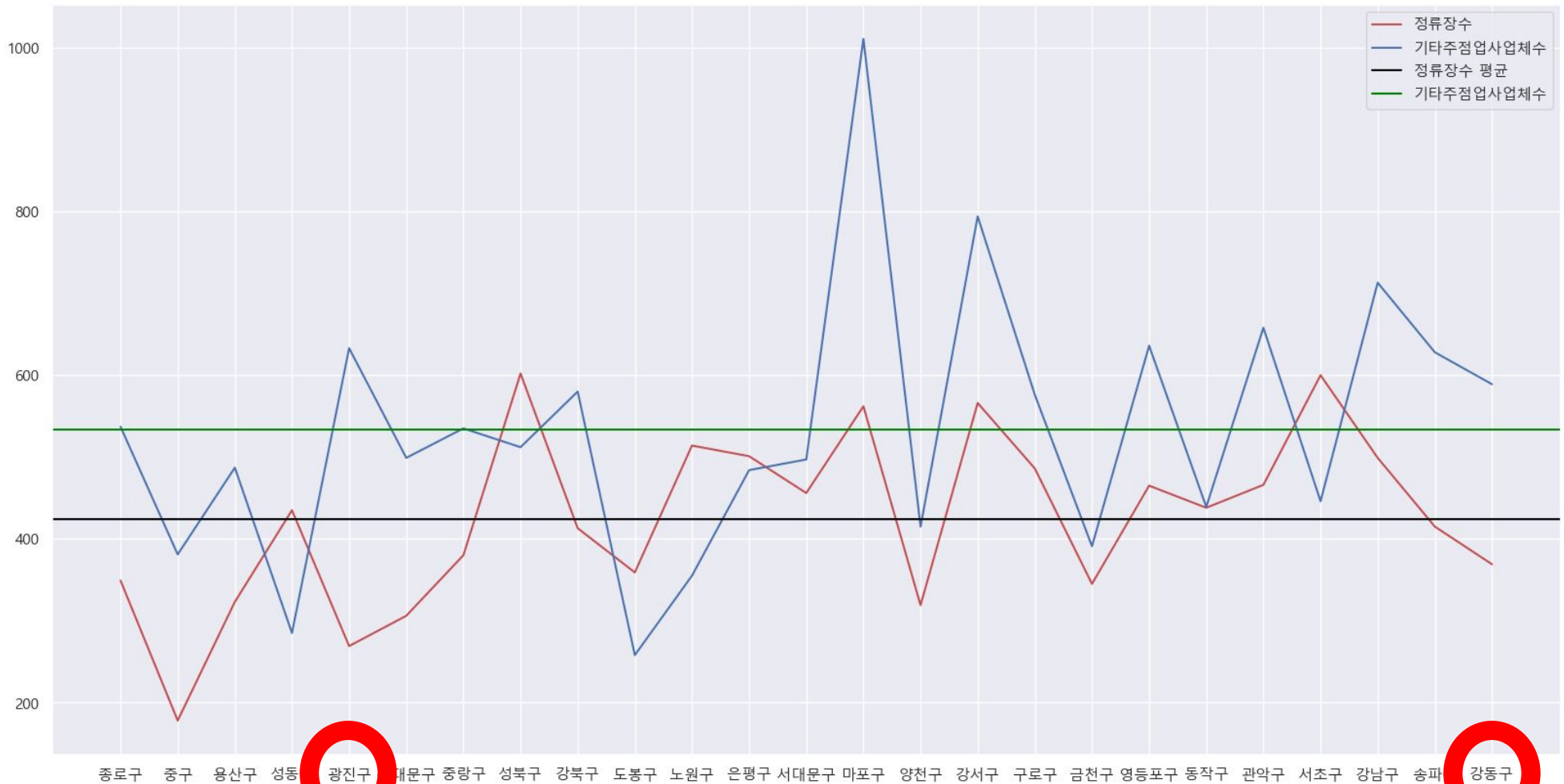


```
sns.regplot(x = '승차총승객수', y = '정류장수', data = m4_data)
plt.show()
# 승차총승객수와 정류장수는 어느정도 비례 관계 인 것 같아
print(spst.pearsonr(m4_data['승차총승객수'], m4_data['정류장수']))
```

PearsonRResult(statistic=0.5666813370974585, pvalue=0.003141977397117322)

분석결과

기타주점업 '사업체수'가 평균 이상이면서 '정류장수'가 평균 이하인 구



면적

면적과 정류장 수의 관련성

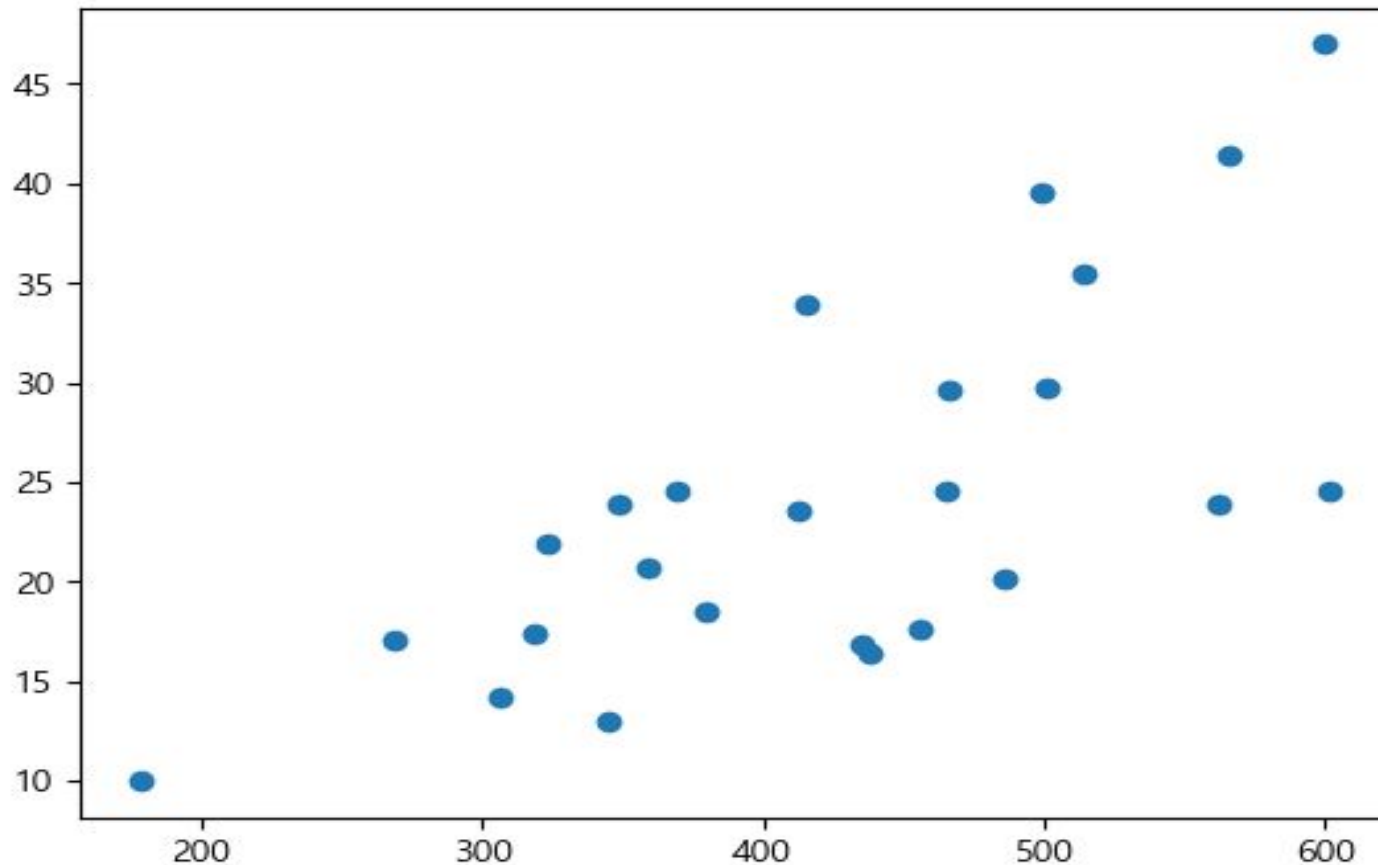
- 면적과 관련된 데이터를 선택한 이유

‘버스 정류장’은 보통 일반적인 간격과 유동인구를 고려하여 설치되기 때문에 면적이 넓으면 정류소가 많을 것이고, 적으면 정류소가 적을 것으로 예상하였음. 이에 따라 각 자치구별 면적 데이터와 정류장 수의 관계성을 파악해보고자 함.

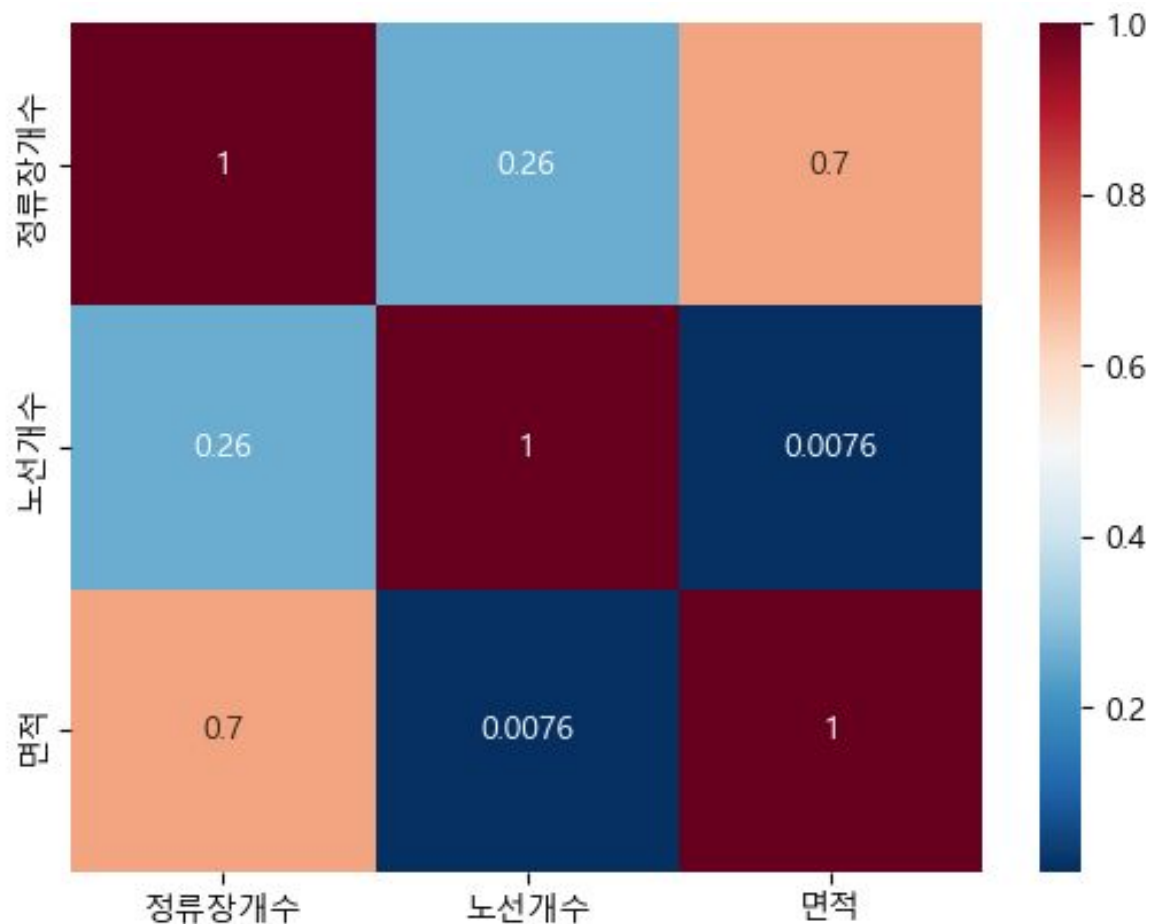
Scatter plot

```
plt.scatter(total['정류장개수'], total['면적'])
```

```
<matplotlib.collections.PathCollection at 0x27681665850>
```



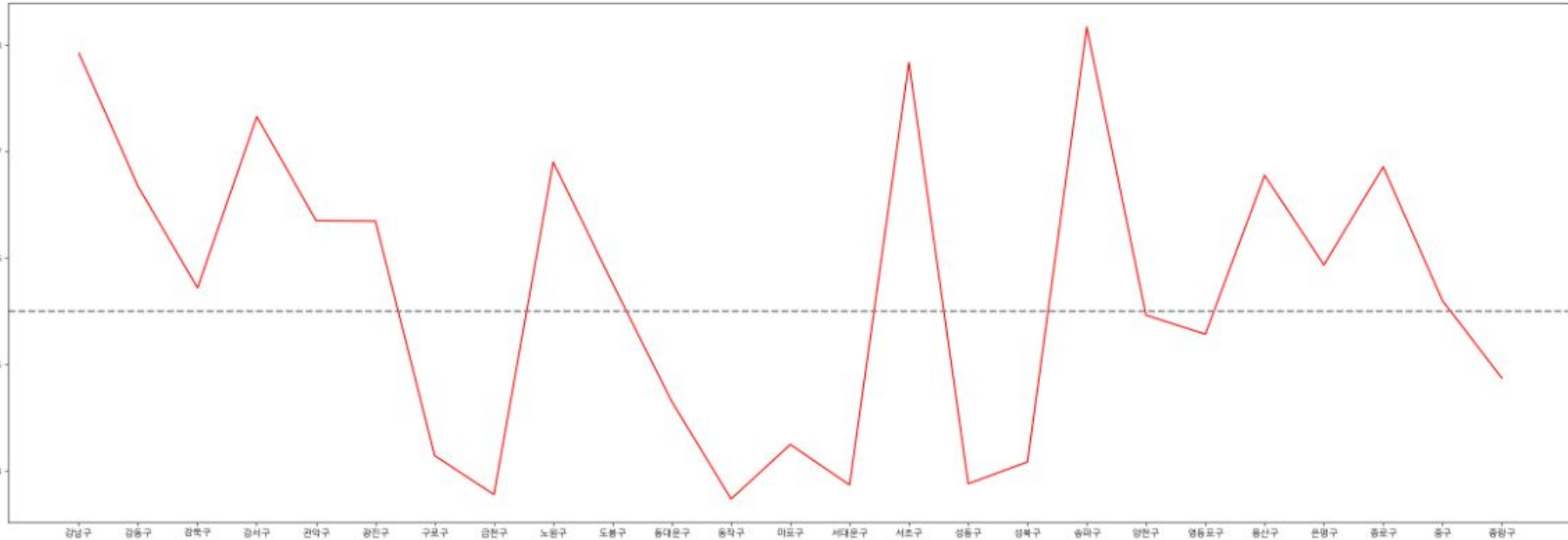
히트맵



```
# 0.05 이하이므로 귀무가설 기각. 상관관계 유의
pearsonr(total['정류장개수'],total['면적'])
```

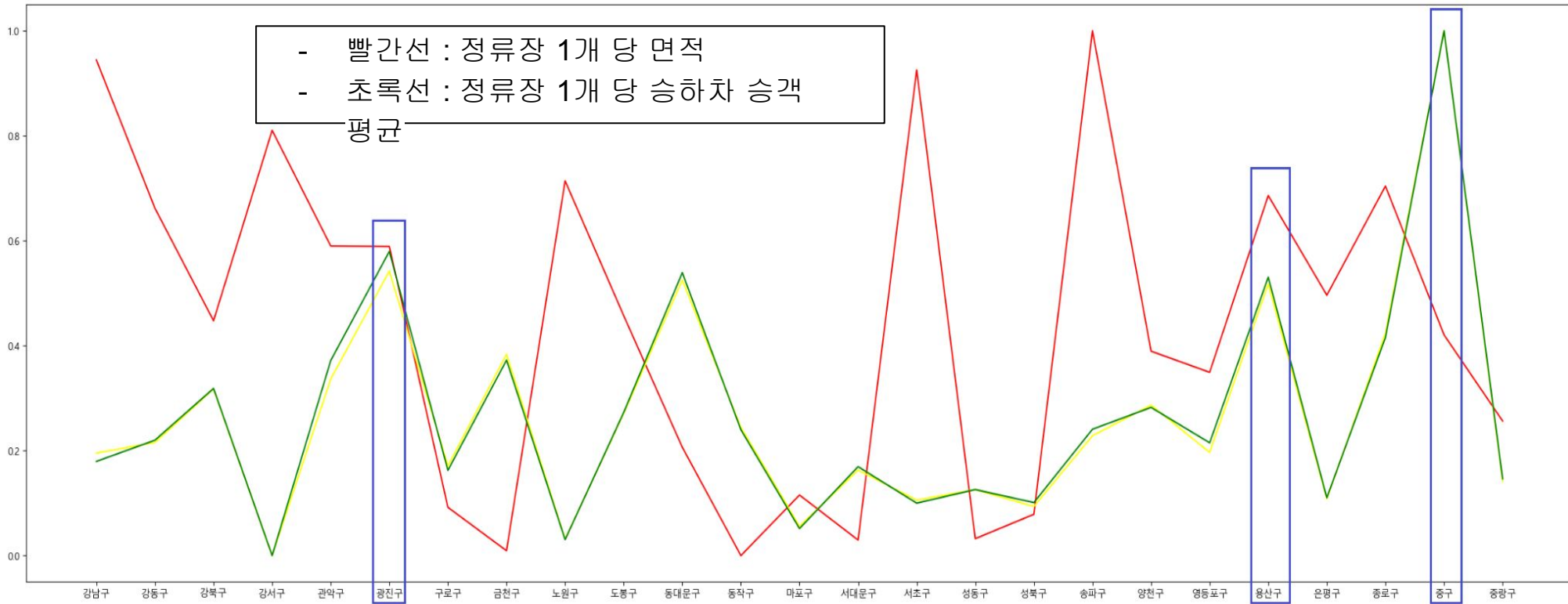
```
PearsonRResult(statistic=0.7016294576894639, pvalue=9.298655202703015e-05)
```

정류장 1개 당 면적 범위(km²)



- 점선 위 부분은 상대적으로 정류장 1개 당 면적 범위가 넓음 (서초구, 송파구 등)
- 점선 아래 부분은 정류장 1개 당 면적 범위가 좁음. (금천구, 서대문구 등)
- 따라서 점선 위 부분 지역에 초점을 맞춰 정류소 추가 설치 장소를 탐색

정류장 1개 당 면적 범위(km²)와 승하차 승객 비교



- 중구 : 정류장 1개당 승하차 승객수가 타 자치구보다 약 3배 많음(중위수 기준)
- 용산구, 광진구 : 정류장 1개당 승하차 승객수 & 면적범위 모두 top5

-> 정류장 증설을 통하여 정류장 1개가 커버하는 면적과 승하차 승객수를 줄임으로써
보다 쾌적한 대중교통 환경 조성

결론

버스 정류장의 추가 설치가 필요한 자치구

[커피전문점 기준]

중구, 강남구

[기타주점업 기준]

광진구, 강동구

[면적 기준]

중구, 용산구, 광진구

kt

 AIVLE

 AIVLE
Let's make it possible