

공공 Wi-Fi 설치와 효율성 분석

인구 밀도와 인터넷 접속률과의 상관관계

산업보안학과 32230324 김가람

Dankook univ.



과목명 : 데이터사이언스

담당교수 : 오세종

학과 : 산업보안학과

학번 : 32230324

이름: 김가람

제출일 : 2024/12/13

요약

■ 분석주제

인터넷 접속률, 인구 밀도와 공공 Wi-Fi 사용량의 관계 및 설치 효율성 분석

■ 데이터셋

- 서울특별시_공공와이파이 AP별 사용량
- 서울특별시_공공와이파이 서비스 위치 정보
- 서울특별시_인터넷 이용빈도 및 주 평균 인터넷 이용시간 통계
- 서울특별시_인구밀도(구별) 통계

■ 분석도구

평균, 표준편차, 사분위수, 상관계수, 선 그래프, 막대 그래프, 상자 그림, 산점도 및 회귀선, 트리맵, 버블 차트, 지도 시각화, 독립표본 T-검정

■ 분석내용 요약

- 공공 Wi-Fi 설치 현황과 이용 수요
 - Wi-Fi 설치 수
 - Wi-Fi 사용량
 - Wi-Fi 설치 수와 사용량의 관계
 - Wi-Fi 설치 위치 및 사용량 지도
- 공공 Wi-Fi 설치와 인터넷 사용 시간의 관계 분석
 - 인터넷 사용 시간
 - Wi-Fi 설치 수와 인터넷 사용 시간의 관계
- 공공 Wi-Fi 설치와 인구 밀도의 관계 분석
 - 인구 밀도와 설치 수의 관계
 - 인구 밀도와 사용량의 관계
 - 인구 밀도, 설치 수, 사용량의 관계

■ 결론 요약

서울시 공공 Wi-Fi는 전체적으로 고르게 분포되어 있으나 설치 수와 사용량 간의 상관성이 약하며, Wi-Fi 설치 수와 인터넷 사용 시간 간에는 강한 상관성이 확인 되었지만, 인구 밀도와는 약한 음의 상관성을 보였다. 또 인구 밀도가 높다고 해서 사용량이 높진 않음을 알 수 있었다. 따라서 단순히 공공 Wi-Fi의 설치 수를 늘리는 것이 아니라 지역별 특성, 유동 인구, 시설 활용도 등을 중심으로 효율적인 설치가 이루어져야 한다.

■ 이 분석의 장점

- 평균, 표준편차, 상관계수 등 통계적 분석부터 지도 시각화, 트리맵, 버블 차트 등 고급 시각화까지 활용하여 데이터를 다각도로 분석함.
- 공공 Wi-Fi 설치 위치나 수, 사용량 분석에서만 그치지 않고 이들과 인터넷 이용률, 인구 밀도의 관계도 종합적으로 분석함.
- 데이터의 일관성을 위해 2023년이나 2021년의 통계만 사용하거나 두 데이터셋 모두에 있는 Wi-Fi만 추출하고, '과천시'라는 특이값을 제거함.
- 결과 해석에만 그치지 않고, 해당 결과가 나타나게 된 원인도 분석함.

[분석 주제 및 데이터셋 소개]

본 보고서의 주제는 ‘인터넷 접속률과 공공 Wi-Fi 사용량의 관계 및 설치 효율성 분석’으로, 공공 Wi-Fi가 지역별 인터넷 접속률에 어떤 영향을 미치는지 분석하고, 이용 수요에 따른 설치 효율성을 평가하고자 한다. 또 추가적인 분석을 통해 지역별 인구 밀도가 공공 Wi-Fi 사용량에 영향을 미치는지도 알아보고자 한다.

본 분석에 사용될 데이터셋은 총 5개로, 각 데이터셋은 서울특별시의 공공 Wi-Fi 사용량, 설치 위치, 지역별 인터넷 접속률, 인구 밀도에 관한 정보를 제공하며, 자세한 내용은 아래와 같다.

1. 서울 특별시_공공와이파이 AP별 사용량

이 데이터셋은 서울특별시 내 각 공공 Wi-Fi 접속 지점(AP)의 사용량 정보를 포함하고 있으며, 고정형, 마을버스, 시내버스 총 세 가지 형태로 나뉘어 있다. 고정형에는 공원, 도서관, 공공기관, 버스정류장 등에 설치된 공공 Wi-Fi의 사용량이, 마을버스와 시내버스에는 각 버스에 설치된 공공 Wi-Fi의 사용량이 포함되어 있으며, 각 데이터셋은 관리번호, 자치구, AP별 이용량(GB)와 같은 요소로 구성되어 있다.



1	관리번호	자치구	AP별 이용량(GB)
2	WF181037	성동구	13990.43
3	WF181190	은평구	12011.05
4	WF191009	은평구	10940.32
5	WF181038	성동구	10244.12
6	WF181035	성동구	9946.631
7	WF191010	은평구	9687.848
8	WF190048	강남구	9396.758
9	WF190888	영등포구	7983.709
10	WF190785	성북구	7752.215

2. 서울특별시_공공와이파이 서비스 위치 정보

이 데이터셋은 서울특별시 내 공공 Wi-Fi 서비스의 설치 위치 정보를 포함하고 있으며, 관리번호, 자치구, 와이파이명, 도로명주소, 상세주소, 설치위치, 설치년도, 실내외구분, Y좌표, X좌표 등의 요소로 구성되어 있다.

1	관리번호	자치구	와이파이명	도로명주소	상세주소	설치위치	설치유형	설치기관	서비스구분	망종류	설치년도	설내외구분	wifi접속환	Y좌표	X좌표	작업일자
2	ARI00001	서대문구	서울아리수서소문로	1본관 1F	복도	7-1-3. 공공서울시(AP 공공WiFi)	자가망_수	2024	실내		37.56192	126.9668	#####			
3	ARI00003	서대문구	서울아리수서소문로	1본관 2F	복도	7-1-3. 공공서울시(AP 공공WiFi)	자가망_수	2024	실내		37.56192	126.9668	#####			
4	ARI00004	서대문구	서울아리수서소문로	1본관 2F	본부장실	7-1-3. 공공서울시(AP 공공WiFi)	자가망_수	2024	실내		37.56192	126.9668	#####			
5	ARI00005	서대문구	서울아리수서소문로	1본관 2F	부본부장실	7-1-3. 공공서울시(AP 공공WiFi)	자가망_수	2024	실내		37.56192	126.9668	#####			
6	ARI00006	서대문구	서울아리수서소문로	1본관 1F	식당	7-1-3. 공공서울시(AP 공공WiFi)	자가망_수	2024	실내		37.56192	126.9668	#####			
7	ARI00008	서대문구	서울아리수서소문로	1본관 2F	경영혁신과	7-1-3. 공공서울시(AP 공공WiFi)	자가망_수	2024	실내		37.56192	126.9668	#####			
8	ARI00009	서대문구	서울아리수서소문로	1본관 2F	소회의실	7-1-3. 공공서울시(AP 공공WiFi)	자가망_수	2024	실내		37.56192	126.9668	#####			
9	ARI00010	서대문구	서울아리수서소문로	1본관 3F	복도	7-1-3. 공공서울시(AP 공공WiFi)	자가망_수	2024	실내		37.56192	126.9668	#####			
10	ARI00011	서대문구	서울아리수서소문로	1본관 3F	생산부	7-1-3. 공공서울시(AP 공공WiFi)	자가망_수	2024	실내		37.56192	126.9668	#####			

3. 서울특별시_인터넷 이용빈도 및 주 평균 인터넷 이용시간 통계

이 데이터셋은 서울특별시 내 인터넷 이용빈도(일주일에 1회 이상, 한달에 1회 이상 등)와 주 평균 인터넷 이용시간(1시간 미만, 1시간 이상, 3시간 이상, 7시간 이상 등)을 기준으로 분석한 정보를 연도별로 제공하고 있다.

4. 서울특별시_인구밀도(구별) 통계

이 데이터셋은 서울특별시 각 구별 인구 수와 면적에 대한 통계 정보를 연도별로 담고 있다. 이를 통해 시간에 따른 각 구별 인구 변화와 면적 대비 인구 밀도의 변동을 분석할 수 있다.

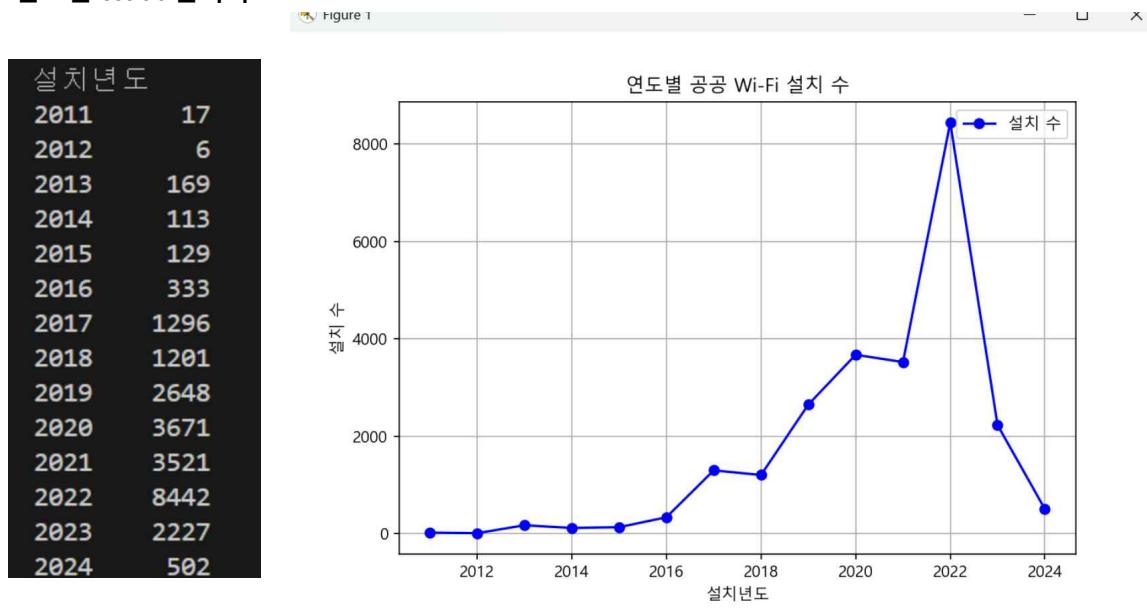
	A	B	C	D	E	F	G	H	I	J	K
1	동별(1)	동별(2)	2007	2007	2007	2008	2008	2008	2009	2009	2009
2	동별(1)	동별(2)	인구 (명)	면적 (km ²)	인구밀도 (인구 (명))	면적 (km ²)	인구밀도 (인구 (명))	면적 (km ²)	인구밀도 (인구 (명))	면적 (km ²)	인구밀도 (인구 (명))
3	합계	종로구	173843	23.91	7271	179754	23.91	7519	177543	23.91	742
4	합계	중구	137435	9.96	13798	138811	9.96	13936	137861	9.96	1384
5	합계	용산구	248362	21.87	11358	251043	21.87	11480	251200	21.87	1148
6	합계	성동구	341620	16.85	20277	322679	16.85	19151	316064	16.85	1875
7	합계	광진구	386367	17.05	22657	387254	17.06	22698	386513	17.06	2265
8	합계	동대문구	385825	14.2	27168	381663	14.2	26876	374277	14.2	2635

[탐색적 데이터 분석 내용]

1. 공공 Wi-Fi 설치 현황과 이용 수요

1) Wi-Fi 설치 수

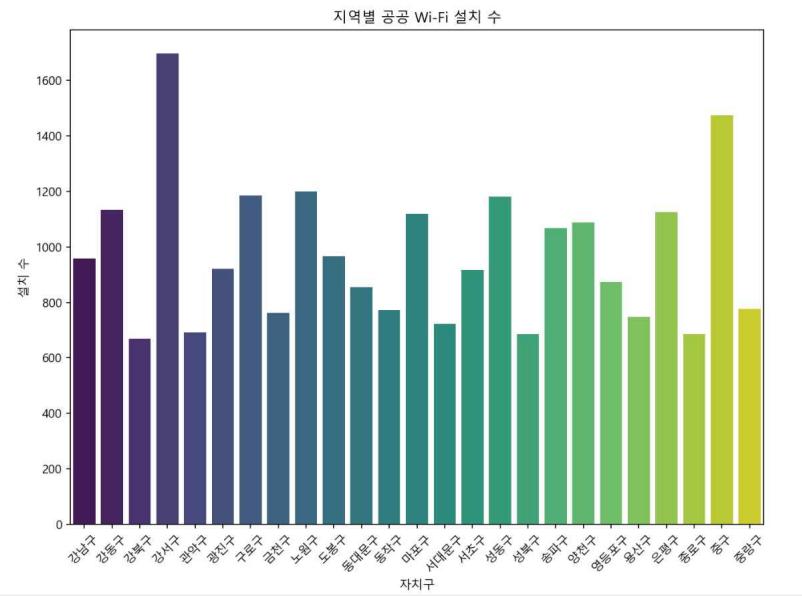
연도별 Wi-Fi 설치 수



- 2011년부터 2024년까지 공공 Wi-Fi 설치 수가 급격히 증가하는 추세를 보인다.
- 특히 2022년에 설치 수가 폭발적으로 증가했다.
→ 코로나 19 펜데믹으로 인해 비대면 활동과 공공 Wi-Fi에 대한 수요가 급증했을 것으로 추정
- 2023년과 2024년에는 설치 수가 감소하였으나, 여전히 많은 양의 Wi-Fi가 설치되었다.
→ 2023년에 설치가 감소한 이유는 이미 설치된 공공 Wi-Fi가 상당히 많기 때문으로 추정
→ 2024년의 데이터는 올해 데이터이기에 완전한 데이터가 아닐 가능성성이 높음

지역구별 Wi-Fi 설치 수

자치구	설치 수
강남구	958
강동구	1134
강북구	668
강서구	1697
관악구	692
광진구	922
구로구	1185
금천구	761
노원구	1199
도봉구	966
동대문구	856
동작구	773
마포구	1119
서대문구	722
서초구	917
성동구	1181
성북구	686
송파구	1067
양천구	1089
영등포구	874
용산구	748
은평구	1125
종로구	685
중구	1474
중랑구	777



- Wi-Fi가 가장 많이 설치된 지역은 ‘강서구’로 총 1697개가 설치되었다.
→ 강서구는 서울 내 대규모 주거 지역과 지하철, 김포국제공항 등 다양한 교통 시설을 포함하고 있어 Wi-Fi 설치 수가 매우 높은 지역인 것으로 추정
- 그 뒤로는 ‘중구’ 1474개, ‘노원구’ 1199개가 있다.
→ 중구는 서울의 상업 중심지로, 유동 인구가 많고 다양한 활동이 이루어지는 지역이라 Wi-Fi 설치의 필요성이 컸을 것으로 추정
→ 노원구는 인구 밀도가 높고 교육 및 주거 시설이 집중된 지역이기 때문에 추정
- 가장 적게 설치된 지역은 ‘강북구’로 총 668개가 설치되었다.
→ 강북구는 서울 외곽 지역으로, 유동 인구와 상업 활동이 상대적으로 적기 때문에 추정
- 그 뒤로는 ‘종로구’ 685개, ‘성북구’ 686개가 있다.
→ 종로구는 관광지와 역사적 지역으로 특정 구역에만 집중 설치되고, 전체 설치 수는 적은 편
→ 성북구는 상대적으로 작은 규모와 낮은 유동 인구로 인해 설치 수가 적은 것으로 추정

지역구별 Wi-Fi 설치 비율

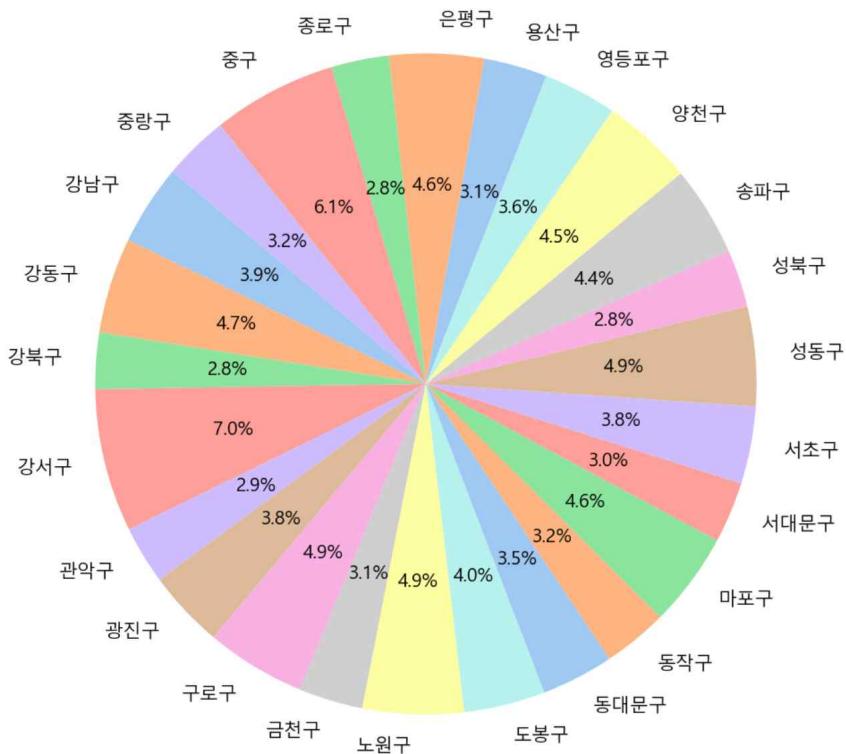
- 총 Wi-Fi 설치 수

```
>>> total_install_count  
24275
```

- Wi-Fi 설치 비율

자치구		마포구	4.609681
강남구	3.946447	서대문구	2.974253
강동구	4.671473	서초구	3.777549
강북구	2.751802	성동구	4.865088
강서구	6.990731	성북구	2.825953
관악구	2.850669	송파구	4.395469
광진구	3.798146	양천구	4.486097
구로구	4.881565	영등포구	3.600412
금천구	3.134912	용산구	3.081359
노원구	4.939238	은평구	4.634398
도봉구	3.979403	종로구	2.821833
동대문구	3.526262	중구	6.072091
동작구	3.184346	중랑구	3.200824

지역별 공공 Wi-Fi 설치 비율

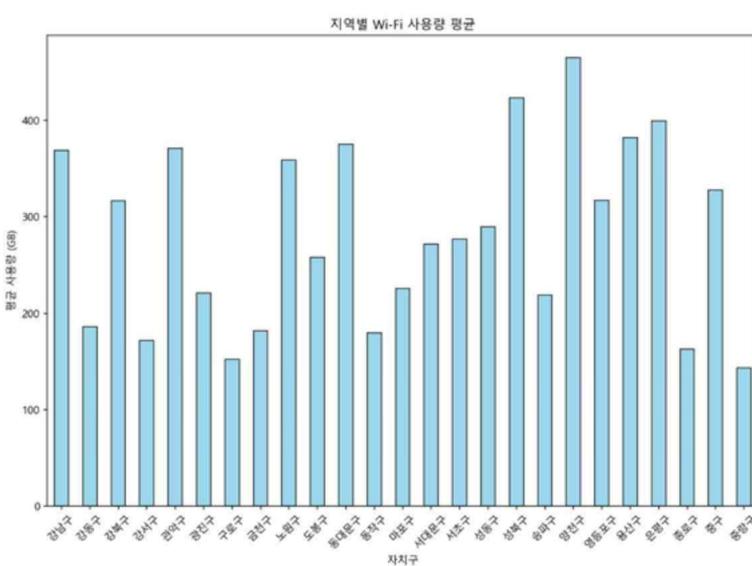


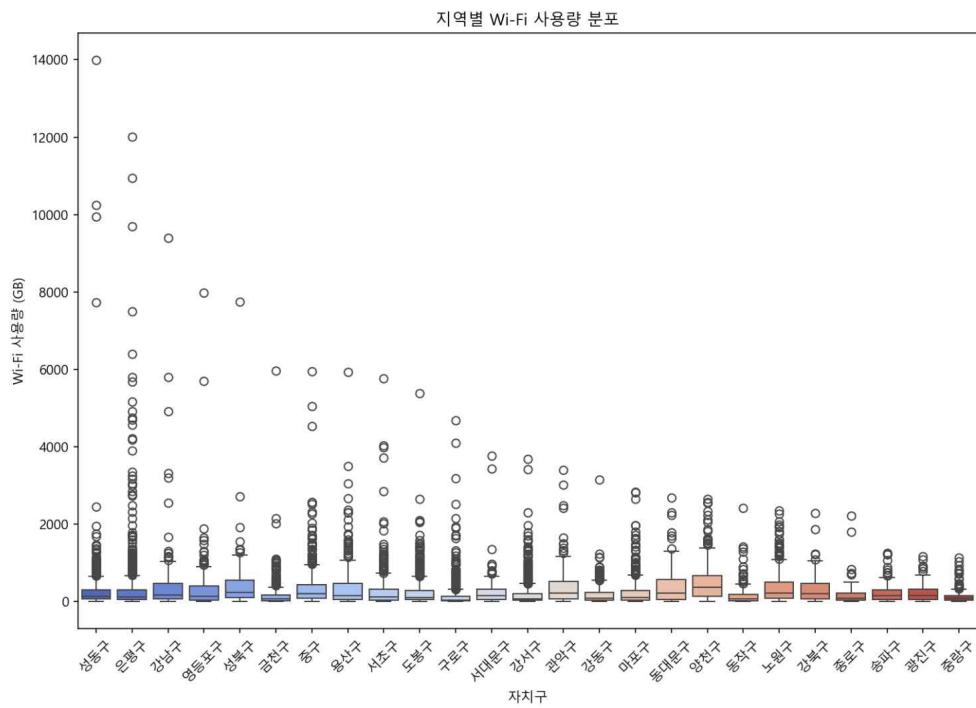
- 전체 공공 Wi-Fi 설치 수는 총 24275개이다.
- ‘강서구’가 전체 설치 수의 약 7.0%로 가장 높은 비율을 차지하고 있다.
- 그 뒤로는 ‘중구’ 6.1%, ‘노원구’ 4.9%, ‘성동구’ 4.9%가 있다.
- 가장 낮은 비율을 차지하는 지역은 ‘강북구’ 2.8%, ‘종로구’ 2.8%, ‘성북구’ 2.8%이다.

2) Wi-Fi 사용량

지역구별 사용량 통계 계산

자치구	평균	중앙값	분산	표준편차	절사평균
강남구	369.046137	170.635161	5.982154e+05	773.443854	241.723498
강동구	186.320661	86.963394	7.199782e+04	268.324089	134.524494
강북구	316.652587	201.810404	1.123960e+05	335.255143	266.145177
강서구	171.646215	69.092556	8.180317e+04	286.012532	110.908060
관악구	371.136541	213.787302	2.167847e+05	465.601393	285.288075
평진구	221.140406	144.579921	5.048067e+04	224.679634	184.952004
구로구	152.383066	40.897073	1.335205e+05	365.404635	75.521414
금천구	182.089531	65.521736	1.810609e+05	425.512530	101.743580
노원구	359.198126	214.224578	1.586722e+05	398.336853	283.477769
도봉구	258.004992	103.296991	1.813420e+05	425.842640	167.327588
동대문구	375.156396	209.252626	2.083170e+05	456.417590	289.577256
동작구	179.852352	67.030865	9.820676e+04	313.379579	108.866252
마포구	225.635078	100.470248	1.203503e+05	346.915420	153.197124
서대문구	271.803826	156.389297	2.162545e+05	465.031675	187.863191
서초구	276.747635	114.822111	2.715817e+05	521.135056	172.409031
성동구	289.725477	137.335435	6.546801e+05	809.123031	181.888231
성북구	423.675501	236.927801	4.664242e+05	682.952578	323.748639
송파구	219.003704	146.119220	5.607410e+04	236.799713	174.713224
양천구	465.195991	367.320989	1.846562e+05	429.716427	404.178503
영등포구	317.108899	130.861826	4.313413e+05	656.765797	208.919374
용산구	382.260078	152.247045	3.748396e+05	612.241474	254.189563
은평구	399.458411	120.771327	1.070773e+06	1034.781518	185.243083
종로구	162.948137	84.602513	6.168525e+04	248.365147	118.900204
중구	327.850462	207.620491	2.004252e+05	447.688783	250.626935
중랑구	143.121692	74.987828	4.029999e+04	200.748566	95.686313





- 평균 사용량이 가장 높은 지역은 '양천구'로, 465.195991 GB이고, 가장 낮은 지역은 '중랑구'로, 143.121692 GB이다.
 - 양천구는 지역 주민의 공공 Wi-Fi 의존도가 높거나 유동 인구가 많기 때문으로 추정
 - 중랑구는 공공 Wi-Fi의 활용도가 낮거나 설치된 공공 Wi-Fi들의 위치가 효과적이지 않을 가능성이 있음
- 중앙값은 데이터의 중앙값을 나타내며, '양천구'가 367.320989 GB로 가장 높고, '중랑구'가 74.987828 GB로 가장 낮다.
- 분산 및 표준편차가 높은 지역은 '은평구', '성동구', '강남구'이며, 낮은 지역은 중랑구, 광진구, 송파구이다.
 - 높은 지역들은 Wi-Fi 사용량의 변동이 매우 큰 지역으로 최대 사용량을 봤을 때, 특정 위치에서의 Wi-Fi 의존도가 매우 높음을 알 수 있음
 - 낮은 지역들은 사용량이 비교적 균일하게 분포되어 있음
- 상하위 10% 데이터를 제거하고 평균을 계산한 절사평균은 '양천구'가 404.170503 GB로 가장 높고, 중랑구가 95.686313으로 가장 낮다.

3) Wi-Fi 설치 수와 사용량의 관계

- * 사용량 데이터셋과 위치 데이터셋 모두에 존재하는 공공 Wi-Fi만 추출 후 진행
- * 지역별 설치 수가 다르므로 사용량은 평균을 내어 사용

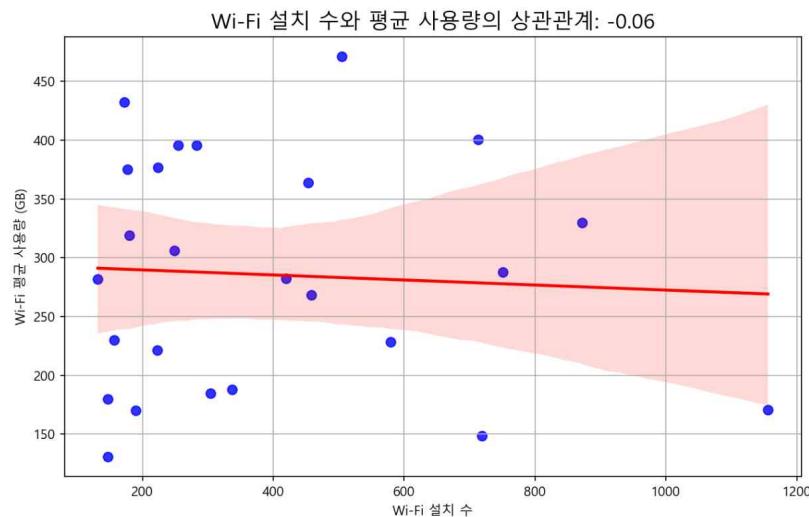
지역구별 Wi-Fi 설치 수와 사용량 합계의 상관관계

- 지역구별 공공 Wi-Fi 설치 수 및 사용량

	자치구	설치 수	평균 사용량
0	강남구	255	395.559160
1	강동구	337	187.769296
2	강북구	180	319.348609
3	강서구	1156	170.589090
4	관악구	224	376.689702
5	광진구	157	229.831202
6	구로구	719	148.432434
7	금천구	304	184.705619
8	노원구	453	364.080996
9	도봉구	458	268.224120
10	동대문구	177	375.163599
11	동작구	147	179.993406
12	마포구	579	228.683227
13	서대문구	132	281.983456
14	서초구	419	282.511559
15	성동구	751	287.797065
16	성북구	172	432.618934
17	송파구	223	221.279288
18	양천구	505	471.043980
19	영등포구	249	306.072442
20	용산구	283	395.534647
21	은평구	713	400.708816
22	종로구	190	169.897226
23	중구	872	329.989002
24	중랑구	147	131.005109

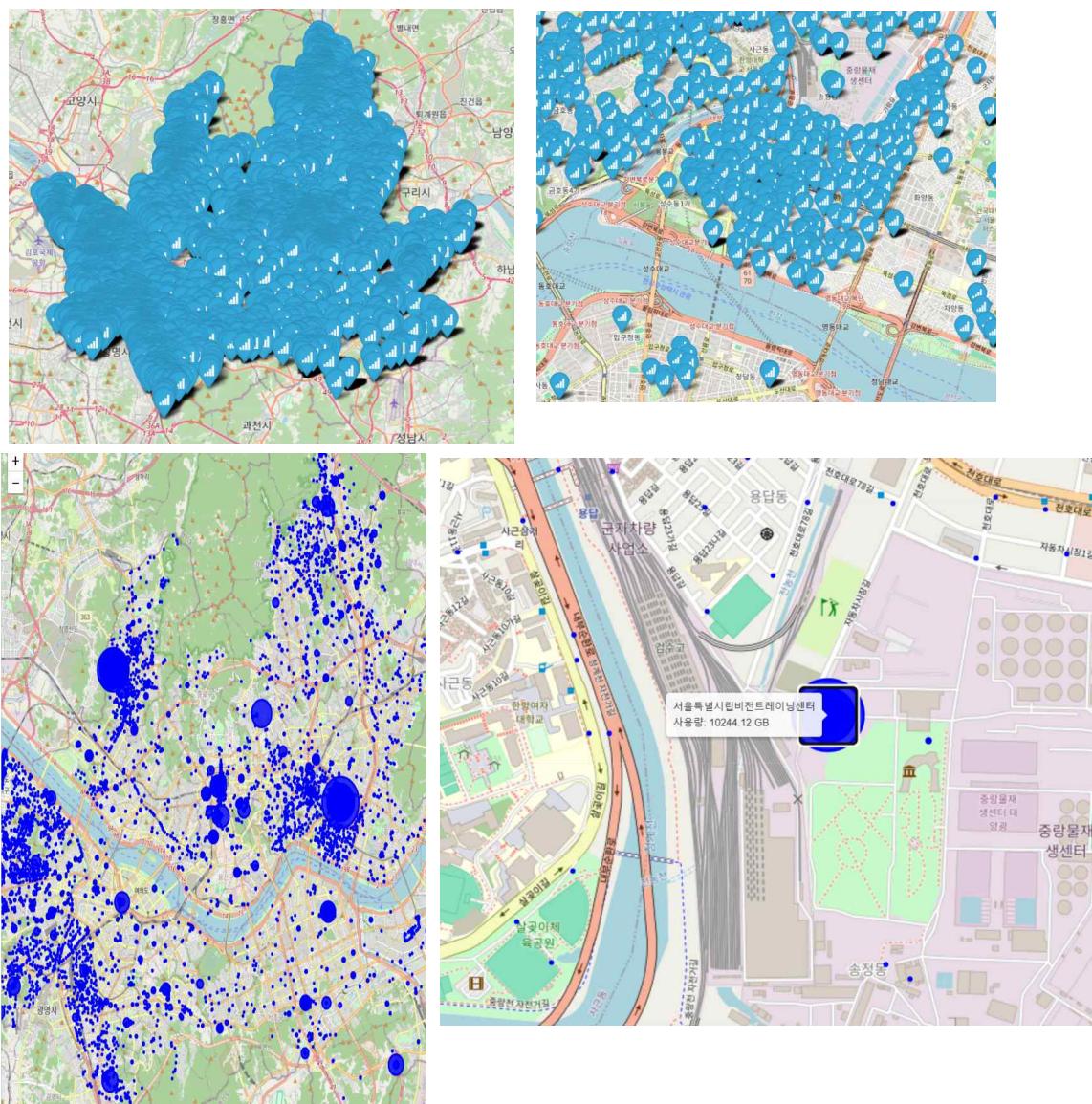
- 상관계수

```
>>> cor
np.float64(-0.05909152470096412)
...
```



- 상관계수 값이 -0.06으로 공공 Wi-Fi 설치 수와 평균 사용량 간의 상관관계가 거의 없다.
- 점들이 회귀선 주변에 고르게 분포되어 있지 않고 설치 수와 사용량 간의 일관된 패턴이 보이지 않는다.
- 회귀선이 거의 평평하게 나타나는 것으로 보아 평균 사용량과 설치 수 간의 연관성이 미미함을 알 수 있다.

4) Wi-Fi 설치 위치 및 사용량 지도



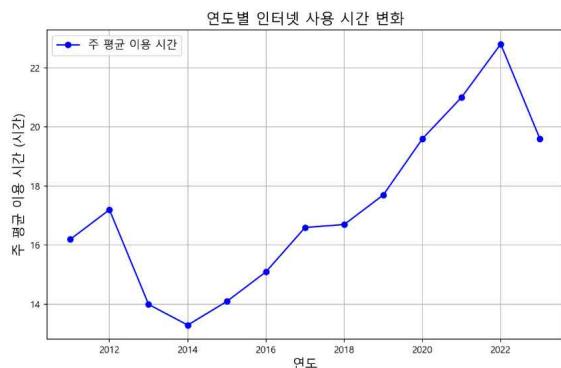
- 공공 Wi-Fi가 서울 전역에 고르게 분포되어 있다.
- 특정 위치에서 사용량이 높게 나타나는 것을 볼 수 있다.
→ 트레이닝센터, 청소년드림센터, 직업재활원, 병원 등 특정 목적을 위해 방문하는 사람들이 밀집하는 시설에서 공공 Wi-Fi의 활용도가 높음

2. 공공 Wi-Fi 설치와 인터넷 사용 시간의 관계 분석

1) 인터넷 사용 시간

연도별 인터넷 사용 시간

연도	주 평균 이용 시간(시간)
1 2011	16.2
2 2012	17.2
3 2013	14.0
4 2014	13.3
5 2015	14.1
6 2016	15.1
7 2017	16.6
8 2018	16.7
9 2019	17.7
10 2020	19.6
11 2021	21.0
12 2022	22.8
13 2023	19.6



- 2011년~2014년은 인터넷 사용 시간이 소폭 감소하였다.
→ 2014년에는 13.3시간으로 최저치
- 2015년~2019년은 인터넷 사용 시간이 꾸준히 증가하였다.
→ 스마트폰 보급 확대와 모바일 콘텐츠 활성화로 인해 증가한 것으로 추정
- 2020년~2022년에는 인터넷 사용 시간이 급격히 증가하였다.
→ 특히, 2022년에는 22.8시간으로 크게 증가
→ 코로나19 팬데믹으로 인한 재택근무, 온라인 강의, 스트리밍 서비스 등 비대면 활동이 증가의 영향을 받았을 것으로 추정
- 2023년에는 다시 감소
→ 사회적 거리두기 완화로 인한 오프라인 활동 증가가 영향을 미쳤을 것으로 추정

2) Wi-Fi 설치 수와 인터넷 사용 시간의 관계

* 2023년까지의 인터넷 사용 시간만 존재하기에 공공 Wi-Fi 누적 설치 수는 2023년까지만 활용

연도별 공공 Wi-Fi 설치 수와 인터넷 사용 시간의 상관관계

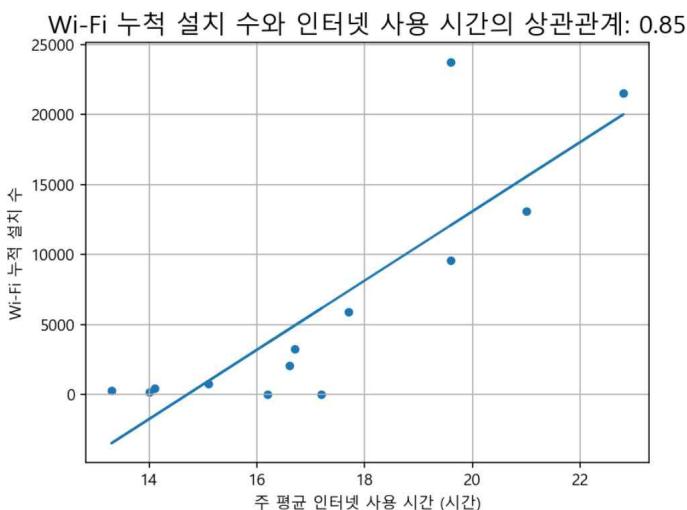
- 연도별 공공 Wi-Fi 누적 설치 수

연도	누적 설치 수
0 2011	17
1 2012	23
2 2013	192
3 2014	305
4 2015	434
5 2016	767
6 2017	2063
7 2018	3264
8 2019	5912
9 2020	9583
10 2021	13104
11 2022	21546
12 2023	23773
13 2024	24275
□	

- 2024년까지 총 24275개의 공공 Wi-Fi가 설치되었다.
- 2011년~2015년은 누적 설치 수가 상대적으로 낮다.
- 2016년~2019년은 누적 설치 수의 증가량이 증가했다.
→ 특히, 2017년부터 설치 속도가 급격히 증가하여 2019년에는 5912개를 달성했다.
- 2020년~2023년은 누적 설치 수의 증가량이 더욱 증가했다.
→ 2020년, 9583개, 2023년은 23773개로 약 2.5배 증가
→ 코로나19 팬데믹으로 인해 공공 Wi-Fi 필요성이 더욱 강조된 것으로 추정
- 2024년에는 24275개로 증가량이 줄었다.
→ 이미 설치된 공공 Wi-Fi가 많고, 아직 2024년이기에 완전한 데이터가 아닐 수 있음

- 상관계수

```
>>> correlation
np.float64(0.8478608528831205)
```



- 상관계수 값이 0.85로 공공 Wi-Fi 설치 수와 인터넷 사용 시간 사이의 상관성이 상당히 높다.
→ 설치된 공공 Wi-Fi 수가 증가할수록 주 평균 인터넷 사용 시간도 증가
→ 공공 Wi-Fi 설치가 인터넷 사용 시간 증가의 주요 요인 중 하나

3. 공공 Wi-Fi 설치와 인구 밀도의 관계 분석)

1) 인구 밀도와 설치 수의 관계

인구 밀도와 공공 Wi-Fi 설치 수의 상관관계

- 자치구별 인구 밀도 (연도별)

	자치구	2007	2008	2009	2010	2011	...	2018	2019	2020	2021	2022	2023
1	종로구	7271	7519	7426	7582	7421	...	6817	6769	6649	6431	6365	6292
2	중구	13798	13936	13841	14177	14212	...	13618	13704	13517	13231	13131	13232
3	용산구	11358	11480	11487	11759	11857	...	11209	11213	11188	10852	10668	10386
4	성동구	20277	19151	18759	18808	18325	...	18771	18327	17822	17359	17136	16930
5	광진구	22657	22698	22653	22786	22661	...	21746	21507	21105	20666	20586	20582
6	동대문구	27168	26876	26356	26713	26653	...	25630	25537	25114	24762	24873	25315
7	중랑구	23320	23399	23232	23371	23175	...	22067	21736	21603	21188	21092	20948
8	성북구	19431	19597	19721	20260	20125	...	18218	18505	18190	17909	17983	17828
9	강북구	14772	14644	14566	14823	14780	...	13683	13462	13202	12820	12614	12414
10	도봉구	18285	18196	17988	17908	17774	...	16531	16245	15852	15465	15205	14987
11	노원구	17536	17524	17360	17367	17160	...	15468	15162	14872	14530	14335	14191
12	은평구	15601	15607	15855	16600	16782	...	16415	16308	16292	16061	15838	15849
13	서대문구	20288	19356	19126	18965	18434	...	18331	18335	18322	17908	18129	18196
14	마포구	16771	16472	16434	16771	16720	...	16199	16180	16011	15878	15745	15727
15	양천구	29221	29183	29104	29028	29049	...	26894	26559	26301	25882	25509	25236
16	강서구	13676	14086	13983	14013	13901	...	14567	14438	14140	13986	13862	13723
17	구로구	22132	22282	22347	22497	22806	...	21793	21837	21496	20933	20794	20656
18	금천구	20316	20396	20240	20349	20326	...	19510	19341	19035	18808	18649	18517
19	영등포구	17876	18056	17983	18141	17892	...	16441	16334	16595	16331	16215	16205
20	동작구	25375	25136	24954	25312	25296	...	25032	25003	24555	24113	23870	23826
21	관악구	18625	18664	18510	18592	18478	...	17587	17496	17241	16891	16951	16838
22	서초구	8766	8767	9172	9362	9340	...	9326	9261	9132	8858	8696	8774
23	강남구	14396	14307	14405	14604	14501	...	13859	13929	13773	13616	13522	13932
24	송파구	18616	19971	20347	20460	20382	...	19883	20156	19896	19599	19616	19483
25	강동구	19083	19285	19920	20214	20359	...	17564	17909	18869	18970	18871	18841

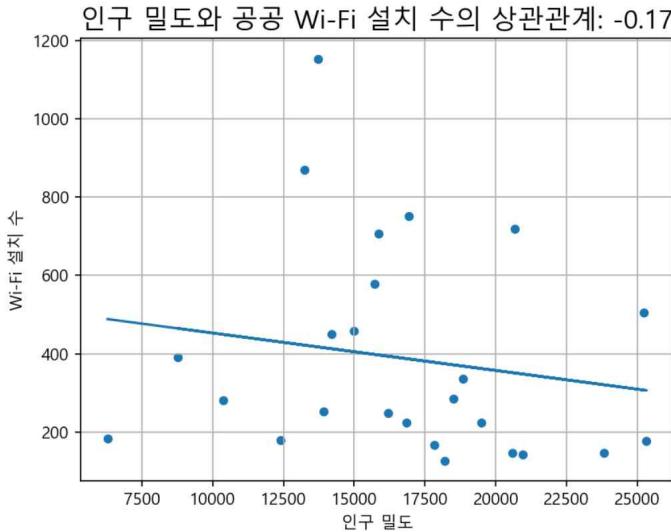
- 2007년부터 2023년까지 서울시 자치구별 인구 밀도의 변화로 전반적으로 완만한 감소세를 보이고 있다.

→ 서울의 인구가 경기도 (수도권 외곽)으로 분산되거나 저출산 및 고령화의 영향으로 도심 인구가 줄어든 것을 원인으로 추정

- 2023년 기준, 양천구, 동작구, 중랑구가 상대적으로 높은 인구 밀도를 보인다.
- 대규모 아파트 단지 등이 발달된 주거지 중심의 지역구일 것으로 추정
- 2023년 기준, 종로구, 관악구, 용산구가 상대적으로 낮은 인구 밀도를 보인다.
- 업무 중심지나 상업지역이 발달된 지역구로 주거지의 비율이 낮을 것으로 추정

- 상관계수

```
>>> cor_density  
np.float64(-0.16681562061474128)
```



- 인구 밀도와 공공 Wi-Fi 설치 수의 상관계수는 -0.17로 상관관계가 크지 않다.

→ 음의 상관성을 띠는 것으로 보아 인구 밀도가 클수록 공공 Wi-Fi 설치 수가 많은 것은 아니라는 점을 알 수 있음

- 인구 밀도가 낮은 일부 지역에서 Wi-Fi 설치 수가 높게 나타남
→ 관광지가 발달한 곳이거나 유동 인구가 많은 지역구일 것으로 추정
- 인구 밀도가 높은 일부 지역에서 Wi-Fi 설치 수가 낮게 나타남
→ 주거지, 업무 중심지나 상업지역이 발달되어 공공 Wi-Fi보다 개인 Wi-Fi나 상업적 네트워크 이용이 많은 지역구일 것으로 추정

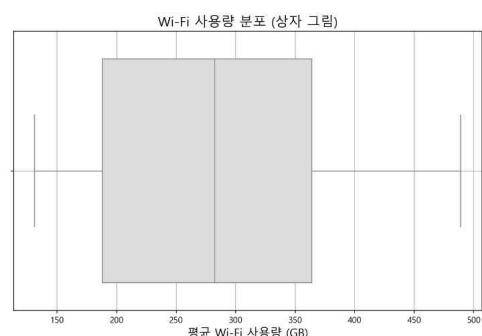
2) 인구 밀도와 사용량의 관계

* 공공 Wi-Fi 사용량 통계가 2021년 기준이므로 2021년 인구 밀도만 사용

공공 Wi-Fi 사용량의 사분위수, 표준편차

- 사분위수

```
>>> quartiles  
0.25    187.769296  
0.50    282.511559  
0.75    364.080996  
Name: 평균 사용량, dtype: float64  
>>> std_dev  
np.float64(97.67282474535463)
```



- 1사분위수 (25%)는 187.77 GB이다.
→ 하위 25%는 187.77 GB 이하의 평균 사용량을 가짐
- 2사분위수 (50%, 중앙값)는 282.51 GB이다.
→ 절반의 자치구는 282.51 GB 이하, 나머지 절반은 그 이상의 평균 사용량을 가짐
- 3사분위수 (75%)는 364.08 GB이다.
→ 상위 25%는 364.08 GB 이상의 평균 사용량을 가짐
→ 1사분위수보다 중앙값(2사분위수)에 가까운 것으로 보아 상위 50% 데이터가 하위 50% 데이터보다 비교적 덜 분산된 것을 알 수 있음
- 표준편차는 97.67 GB이다.
→ 평균값으로부터 약 \pm 97.67 GB 정도 분산되어 있음
→ 비교적 일정하게 분포되어 있음
- 상자 그림 상에서 특이값이 나타나지 않는다.
→ 대부분의 데이터가 사분범위 안에 포함됨
- 상자의 중앙이 중앙값에 가깝다.
→ 좌우 균형이 잡힌 분포임

인구 고밀도 지역과 저밀도 지역 간 사용량 차이 분석

- * 상위 50%는 고밀도, 하위 50%는 저밀도 지역으로 설정
- * 유의수준 0.05로 검정

- 독립 검정
- 고밀도 지역과 저밀도 지역은 다른 지역들이므로 독립된 표본이다.

- 정규성 검정

```
>>> pvalue = stats.shapiro(high_density['평균 사용량']).pvalue
>>> pvalue
np.float64(0.28754426810901834)
>>> pvalue = stats.shapiro(low_density['평균 사용량']).pvalue
>>> pvalue
np.float64(0.40939897699071703)
```

- 고밀도 지역과 저밀도 지역 모두 pvalue가 > 0.05 이상이므로 정규성을 만족한다.
→ 독립표본 T-검정 사용 가능

- 등분산성 검정

```
>>> pvalue
np.float64(0.22941970382047275)
```

- pvalue가 0.05 이상이므로 equal_var = True

- 결과

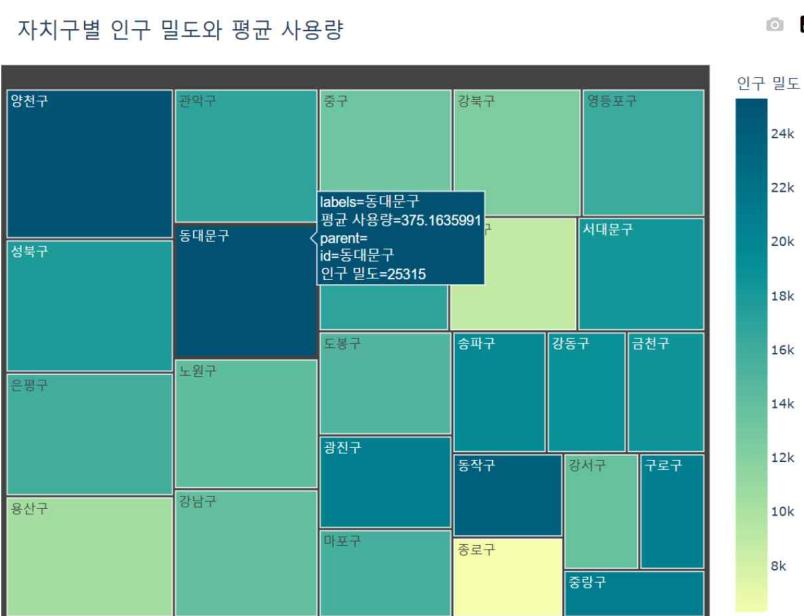
```
>>> result  
TtestResult(statistic=np.float64(-0.6957222384404502), pvalue=np.float64(0.49357335444229766), df=np.float64(23.0))
```

- pvalue가 약 0.4935로 유의수준 0.05보다 크다.

→ 고밀도 지역과 저밀도 지역의 공공 Wi-Fi 사용량은 유의미한 차이가 없다고 결론 내릴 수 있음

→ 공공 Wi-Fi 사용량이 인구 밀도와 크게 관련되지 않음을 알 수 있음

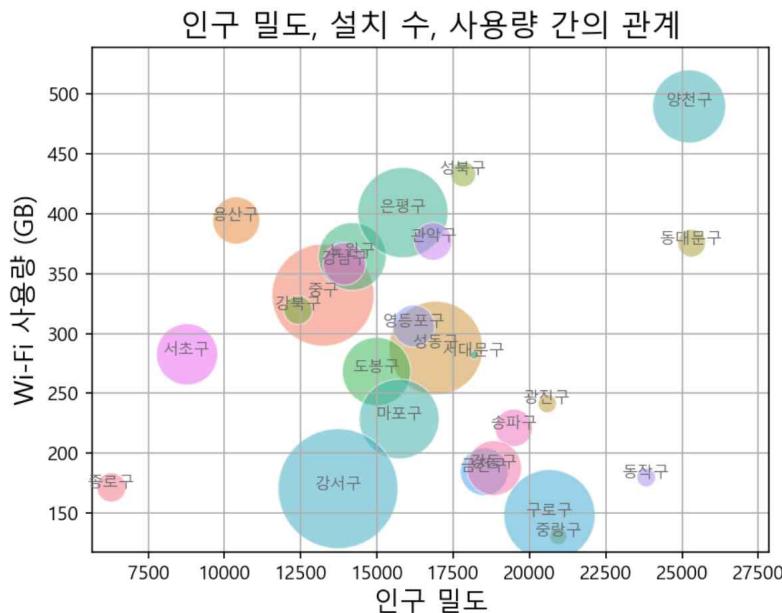
자치구별 인구 밀도와 평균 사용량



- 각 박스의 크기는 공공 Wi-Fi 평균 사용량, 색상은 인구 밀도를 나타낸다.
→ 색상이 진할수록 인구 밀도가 높고, 연한수록 인구 밀도가 낮음
- 동작구, 중랑구 등은 색상이 진하지만 박스의 크기는 작다.
→ 인구 밀도가 높지만 공공 Wi-Fi 평균 사용량은 작음
- 인구 밀도가 높은 지역의 공공 Wi-Fi 평균 사용량이 반드시 높다고 할 수 없음

3) 인구 밀도, 설치 수, 사용량의 관계

시각화



- x축은 인구 밀도, y축은 공공 Wi-Fi 사용량, 버블 크기는 공공 Wi-Fi 설치 수이다.
- 양천구 : 인구 밀도도 높고 공공 Wi-Fi 사용량도 많으며 설치 수도 비교적 많다.
 - 공공 Wi-Fi의 효율성이 높다고 볼 수 있음
 - 공공 Wi-Fi가 주로 사람들이 많이 모이는 장소에 설치되었고 이러한 시설들이 잘 분포되어 있을 것으로 추정
- 강서구 : 공공 Wi-Fi 설치 수는 많지만 인구 밀도가 낮고 평균 사용량도 낮다.
 - 공공 Wi-Fi의 효율성이 낮다고 볼 수 있음
 - 김포 공항과 여러 산업시설로 인해 공공 Wi-Fi 설치 수는 많지만 이들이 효율적으로 사용되지 못한 것으로 추정
- 중랑구, 동작구 : 인구 밀도가 높지만 공공 Wi-Fi 사용량이 낮고 설치 수도 적다.
 - 공공시설보다 주거 중심지라서 공공 Wi-Fi보다 개인 Wi-Fi를 사용하는 경우가 더 많을 것으로 추정
- 서초구, 종로구 : 인구 밀도와 공공 Wi-Fi 사용량이 모두 낮고 설치 수도 적다.
 - 서초구는 고소득층 중심의 주거 지역이 많아 개인 Wi-Fi 사용률이 높아 공공 Wi-Fi에 대한 의존도가 낮을 것으로 추정
 - 종로구는 관광지와 행정 중심지의 특성이 강하기 때문으로 추정

[결론]

본 보고서에서는 다양한 분석 도구를 활용해 서울특별시 내의 공공 Wi-Fi 설치 수 및 사용량, 인터넷 이용 시간, 인구 밀도 간의 상호 관계를 다각도로 살펴보았다. 서울특별시 내에는 공공 Wi-Fi가 전체적으로 고르게 분포되어 있지만 설치 수와 사용량 간의 연관성이 약하다는 점을 확인할 수 있었다. 특히, 강서구는 공공 Wi-Fi 설치 수가 가장 많았지만 인구 밀도가 낮고 평균 사용량도 낮아 효율성이 떨어지는 것으로 나타났다. 반대로 양천구는 인구 밀도, 설치 수, 사용량 모두 높아 공공 Wi-Fi 효율성이 높은 것으로 나타났다. 이를 통해 단순히 설치 수를 늘리는 것이 아니라 지역별 수요와 특성을 고려한 설치가 필요함을 알 수 있다.

공공 Wi-Fi 누적 설치 수와 인터넷 평균 사용 시간 간에는 강한 상관관계(0.85)가 있음을 확인할 수 있었다. 이는 Wi-Fi 설치가 증가함에 따라 인터넷 사용 시간이 증가했음을 보여준다. 특히, 코로나 19 시기 동안 비대면 활동의 확산으로 인터넷 사용 시간과 Wi-Fi 설치 수가 함께 급격히 증가했으나 사회적 거리두기가 완화된 이후로는 감소한 것으로 보아 이러한 상관관계가 환경적 요인에 의해 영향을 받을 수 있음을 알 수 있다.

반면, 인구 밀도와 공공 Wi-Fi 설치 수 간에는 상관관계(-0.17)가 거의 없고, 음의 상관관계가 나타남을 확인할 수 있었다. 또한 고밀도 지역과 저밀도 지역의 사용량에서도 큰 차이가 나타나지 않은 것으로 보아 인구 밀도가 높은 지역이라고 해서 공공 Wi-Fi 설치가 더 많이 이루어지거나 더 많이 사용되지는 않음을 알 수 있다.

이러한 분석 결과에서 알 수 있듯이 단순히 공공 Wi-Fi 설치 수를 늘리는 것이 아니라 지역별 특성과 수요를 기반으로 효율적인 설치가 이루어져야 한다. 또 유동 인구와 시설 활용도를 중심으로 설치 지역을 선정해야 하며 개인 Wi-Fi 사용이 높은 지역에서는 공공 Wi-Fi 의존도가 낮을 수 있음을 고려해야 한다.

[별첨 : Source Code]

<캡처본 / 텍스트로는 뒤에 있습니다!>

```
1 import pandas as pd
2 from scipy import stats
3 import numpy as np
4 import matplotlib.pyplot as plt
5 import webbrowser
6 import seaborn as sns
7 from scipy.stats import trim_mean
8 import folium
9 import geokakao as gk
10 import plotly.express as px
11 import matplotlib.image as mpimg
12 import plotly.graph_objects as go
13
14 # 한글
15 plt.rcParams['font.family'] = 'Malgun Gothic'
16 plt.rcParams['axes.unicode_minus'] = False
17
18 # 데이터셋 로드
19 Usage = pd.read_csv('C:/Users/r1b04/OneDrive/바탕 화면/김가람/2학년 2학기/데이터사이언스/기말고사/서울특별시_공공와이파이.csv')
20 Location = pd.read_csv('C:/Users/r1b04/OneDrive/바탕 화면/김가람/2학년 2학기/데이터사이언스/기말고사/서울시_공공와이파이.csv')
21 Population = pd.read_csv('C:/Users/r1b04/OneDrive/바탕 화면/김가람/2학년 2학기/데이터사이언스/기말고사/인구밀도.csv')
22 Time = pd.read_csv('C:/Users/r1b04/OneDrive/바탕 화면/김가람/2학년 2학기/데이터사이언스/기말고사/인터넷 이용시간.csv')
# 1. 공공 Wi-Fi 설치 현황과 이용 수요의 관계 분석

# 1) Wi-Fi 설치 수

# 데이터 전처리
# 특이값 (과천시) 제거
Location = Location[Location['자치구'] != '과천시']
Location

# 연도별
install_count_by_year = Location.groupby('설치년도').size()
install_count_by_year

# 지역별(구) - 가나다순 정렬
install_count_by_region = Location['자치구'].value_counts().sort_index()
install_count_by_region

# 총합
total_install_count = Location.shape[0]
total_install_count

# 전체 설치 수에서 각 지역의 비율
install_percentage_by_region = install_count_by_region / total_install_count * 100
install_percentage_by_region

# 시각화 : 설치 연도별 현황 (선 그래프)
plt.figure(figsize = (10, 6))
plt.plot(install_count_by_year.index, install_count_by_year.values, marker = 'o', color = 'b', label = '설치 수')
plt.title('연도별 공공 Wi-Fi 설치 수')
plt.xlabel('설치년도')
plt.ylabel('설치 수')
plt.grid(True)
plt.legend()
plt.show()

# 시각화 : 지역별 설치 현황 (막대 그래프)
plt.figure(figsize = (12, 8))
sns.barplot(x = install_count_by_region.index, y = install_count_by_region.values, palette = 'viridis')
plt.title('지역별 공공 Wi-Fi 설치 수')
plt.xlabel('자치구')
plt.ylabel('설치 수')
plt.xticks(rotation = 45)
plt.show()
```

```

# 시각화 : 지역별 설치 비율 (원 그래프)
plt.figure(figsize = (8, 8))
plt.pie(install_percentage_by_region, labels = install_percentage_by_region.index, autopct = '%1.1f%%', startangle = 140, colors = sns.color_palette('pastel'))
plt.title('지역별 공공 Wi-Fi 설치 비율')
plt.show()

# 2) Wi-Fi 사용량

# 데이터 전처리
# 둘다값 (과천시) 제거
Usage = Usage[Usage['자치구'] != '과천시']

# 지역별 사용량 통계 계산 - 가나다순 정렬
usage_stats = Usage.groupby('자치구')['AP별 이용량(GB)'].agg(
    평균 = 'mean',
    중앙값 = 'median',
    분산 = 'var',
    표준편차 = 'std',
    절사평균 = lambda x: trim_mean(x, 0.1)
).sort_index()
usage_stats

# 시각화 : 지역별 사용량 평균 (막대 그래프)
plt.figure(figsize = (12, 8))
usage_stats['평균'].plot(kind = 'bar', color = 'skyblue', alpha = 0.8, edgecolor = 'black')
plt.title('지역별 Wi-Fi 사용량 평균')
plt.xlabel('자치구')
plt.ylabel('평균 사용량 (GB)')
plt.xticks(rotation = 45)
plt.show()

# 시각화 : 지역별 사용량 분포 (상자 그림)
plt.figure(figsize = (12, 8))
sns.boxplot(x = '자치구', y = 'AP별 이용량(GB)', data = Usage, palette = 'coolwarm')
plt.title('지역별 Wi-Fi 사용량 분포')
plt.xlabel('자치구')
plt.ylabel('Wi-Fi 사용량 (GB)')
plt.xticks(rotation = 45)
plt.show()

# 3) Wi-Fi 설치 수와 사용량의 관계

# 데이터 전처리
# 두 데이터셋에 모두 존재하는 Wi-Fi만 추출
common_ids = set(Location['관리번호']) & set(Usage['관리번호'])
common_ids

filtered_location_data = Location[Location['관리번호'].isin(common_ids)]
filtered_usage_data = Usage[Usage['관리번호'].isin(common_ids)]

# 가나다순 정렬
filtered_location_data = filtered_location_data.sort_values(by = '자치구')
filtered_usage_data = filtered_usage_data.sort_values(by = '자치구')
filtered_location_data
filtered_usage_data

# 지역별 Wi-Fi 설치 수
install_counts = filtered_location_data.groupby('자치구').size().reset_index(name = '설치 수')
install_counts

# 지역별 설치 수가 다르므로 사용량은 평균을 내어 사용
usage_mean = filtered_usage_data.groupby('자치구')['AP별 이용량(GB)'].mean().reset_index(name = '평균 사용량')
usage_mean

# 두 변수 병합
cor_data = pd.merge(install_counts, usage_mean, on = '자치구').dropna()
cor_data

```

```

# 상관계수 계산
cor = cor_data['설치 수'].corr(cor_data['평균 사용량'])
cor

# 시각화 : Wi-Fi 설치 수와 사용량 합계의 상관관계
plt.figure(figsize = (10, 6))
sns.regplot(
    x = '설치 수',
    y = '평균 사용량',
    data = cor_data,
    scatter_kws = {'color' : 'blue', 's' : 50},
    line_kws = {'color' : 'red'}
)
plt.title(f'Wi-Fi 설치 수와 평균 사용량의 상관관계: {cor:.2f}', fontsize = 16)
plt.xlabel('Wi-Fi 설치 수')
plt.ylabel('Wi-Fi 평균 사용량 (GB)')
plt.grid(True)
plt.show()

# 4) Wi-Fi 설치 위치 및 사용량 지도

# 데이터 병합
merged_data = pd.merge(filtered_location_data, filtered_usage_data, on = '관리번호', how = 'inner')
merged_data = merged_data.rename(columns = {'자치구_x' : '자치구'})
merged_data = merged_data[['자치구', '관리번호', '와이파이명', '도로명주소', '설치년도', 'X좌표', 'Y좌표', 'AP별 이용량(GB)']].dropna()

# 도로명주소를 주소로 변경 ex) 서소문로 51 -> 서울 서대문구 서소문로 51
merged_data['도로명주소'] = '서울' + ' ' + merged_data['자치구'] + ' ' + merged_data['도로명주소']
merged_data = merged_data.rename(columns = {'도로명주소' : '주소'})
merged_data

# 서울 중심 좌표 (기준점)
center = [37.5665, 126.9780]

# Folium 지도 생성
map = folium.Map(location = center, zoom_start = 11)

```

```

# 설치 위치
for _, row in merged_data.iterrows():
    folium.Marker(location = (row['Y좌표'], row['X좌표']),
                  icon = folium.Icon(color = 'blue', icon = 'signal')).add_to(map)

map.save('map.html')
webbrowser.open('map.html')

map = folium.Map(location = center, zoom_start = 12)

for _, row in merged_data.iterrows():
    folium.CircleMarker(
        location = (row['Y좌표'], row['X좌표']),
        radius = row['AP별 이용량(GB)'] / 400,
        color = 'blue',
        fill = True,
        fill_opacity = '80%',
        tooltip = f'{row['와이파이명']}<br>사용량: {row['AP별 이용량(GB)']:.2f} GB'
    ).add_to(map)

map.save('map.html')
webbrowser.open('map.html')

# 2. 인터넷 사용 시간과 공공 Wi-Fi 사용량의 관계 분석

# 1) 인터넷 사용 시간

filtered_time = Time[['시점', '주 평균(시간)']].rename(columns = {'시점' : '연도', '주 평균(시간)' : '주 평균 이용 시간(시간)'})
filtered_time = filtered_time[filtered_time['연도'].str.isnumeric()]
filtered_time['연도'] = pd.to_numeric(filtered_time['연도'], errors = 'coerce')
filtered_time['주 평균 이용 시간(시간)'] = pd.to_numeric(filtered_time['주 평균 이용 시간(시간)'], errors = 'coerce')
filtered_time = filtered_time.dropna()
filtered_time

# 시각화 : 연도별 인터넷 사용 시간 변화 (선 그래프)
plt.figure(figsize = (10, 6))
plt.plot(filtered_time['연도'], filtered_time['주 평균 이용 시간(시간)'], marker = 'o', linestyle = '--', color = 'blue', label = '주 평균 이용 시간')
plt.title('연도별 인터넷 사용 시간 변화', fontsize = 16)
plt.xlabel('연도', fontsize = 14)
plt.ylabel('주 평균 이용 시간(시간)', fontsize = 14)
plt.grid(True)
plt.legend(fontsize = 12)
plt.show()

# 2) Wi-Fi 설치 수와 인터넷 사용 시간의 관계

# 연도별 Wi-Fi 누적 설치 수

wifi_year_total = Location[['설치년도']].copy()
wifi_year_total['설치년도'] = pd.to_numeric(wifi_year_total['설치년도'], errors = 'coerce')
wifi_year_total = wifi_year_total.groupby('설치년도').size().cumsum().reset_index()
wifi_year_total.columns = ['연도', '누적 설치 수']
wifi_year_total = wifi_year_total.dropna()
wifi_year_total

# 병합
combined_data = pd.merge(filtered_time, wifi_year_total, on='연도', how='inner')
combined_data

# 상관분석
correlation = combined_data['주 평균 이용 시간(시간)'].corr(combined_data['누적 설치 수'])
correlation

```

```

# 시각화
combined_data.plot.scatter(x = '주 평균 이용 시간(시간)',
                           y = '누적 설치 수')
m, b = np.polyfit(combined_data['주 평균 이용 시간(시간)'], combined_data['누적 설치 수'], 1)
plt.plot(combined_data['주 평균 이용 시간(시간)'], m * np.array(combined_data['주 평균 이용 시간(시간)']) + b)
plt.title(f'Wi-Fi 누적 설치 수와 인터넷 사용 시간의 상관관계: {correlation:.2f}', fontsize = 16)
plt.xlabel('주 평균 인터넷 사용 시간(시간)')
plt.ylabel('Wi-Fi 누적 설치 수')
plt.grid(True)
plt.show()

# 3. 공공 Wi-Fi 설치와 지역별 인구 밀도의 관계 분석

# 1) 인구 밀도와 설치 수의 관계

# 설치 수 데이터 (설치년도 2023년까지만)
install_data = merged_data[merged_data['설치년도'] <= 2023]
install_data = install_data.groupby('자치구').size().reset_index(name = '설치 수')
install_data

# 인구 밀도 데이터
density_data = Population.drop(columns = ['동별(1)'])
density_data = density_data.rename(columns = {'동별(2)' : '자치구'})
density_data = density_data.iloc[1:]
density_data

# 2023년 인구 밀도만 추출
density_2023 = density_data[['자치구', '2023']]
density_2023 = density_2023.rename(columns = {'2023' : '인구 밀도'})
density_2023

# 병합
merged_density = pd.merge(density_2023, install_data, on = '자치구', how = 'inner')
merged_density['인구 밀도'] = pd.to_numeric(merged_density['인구 밀도'], errors='coerce')
merged_density['설치 수'] = pd.to_numeric(merged_density['설치 수'], errors='coerce')
merged_density

# 상관분석
cor_density = merged_density[['인구 밀도', '설치 수']].corr().iloc[0, 1]
cor_density

# 시각화 : 인구 밀도와 Wi-Fi 설치 수의 관계
merged_density.plot.scatter(x = '인구 밀도',
                            y = '설치 수')
m, b = np.polyfit(merged_density['인구 밀도'], merged_density['설치 수'], 1)
plt.plot(merged_density['인구 밀도'], m * np.array(merged_density['인구 밀도']) + b)
plt.title(f'인구 밀도와 공공 Wi-Fi 설치 수의 상관관계: {cor_density:.2f}', fontsize = 16)
plt.xlabel('인구 밀도')
plt.ylabel('Wi-Fi 설치 수')
plt.grid(True)
plt.show()

```

```

# 2) 인구 밀도와 사용량의 관계

# 인구 밀도와 사용량의 데이터 분포와 변동성 분석
# 사분위수, 사분범위, 표준편차

# 2021년 인구 밀도만 (Wi-Fi 사용량 통계가 2021년 기준임)
density_2021 = density_data[['자치구', '2021']]
density_2021 = density_2021.rename(columns = {'2021' : '인구 밀도'})
density_2021

usage_data = merged_data
usage_data = usage_data.groupby('자치구')[['AP별 이용량(GB)']].mean().reset_index(name = '평균 사용량')
usage_data

# 병합
density_usage_data = pd.merge(density_2021, usage_data, on = '자치구', how = 'inner')
density_usage_data['인구 밀도'] = pd.to_numeric(density_usage_data['인구 밀도'], errors='coerce')
density_usage_data['평균 사용량'] = pd.to_numeric(density_usage_data['평균 사용량'], errors='coerce')
density_usage_data

# 분석
quartiles = density_usage_data['평균 사용량'].quantile([0.25, 0.5, 0.75])
std_dev = density_usage_data['평균 사용량'].std()
quartiles
std_dev

# 시각화 : Wi-Fi 사용량 분포 (상자 그림)
plt.figure(figsize = (10, 6))
sns.boxplot(x = '평균 사용량',
            data = density_usage_data,
            palette = 'coolwarm')
plt.title('Wi-Fi 사용량 분포', fontsize = 16)
plt.xlabel('평균 Wi-Fi 사용량 (GB)', fontsize = 14)
plt.grid(True)
plt.show()

# 고밀도 지역과 저밀도 지역 간 사용량 차이 분석
median_density = density_usage_data['인구 밀도'].median()
high_density = density_usage_data[density_usage_data['인구 밀도'] >= median_density].reset_index()
low_density = density_usage_data[density_usage_data['인구 밀도'] < median_density].reset_index()

# 1. 독립 검정
# 고밀도 지역과 저밀도 지역은 다른 지역들이므로 독립된 표본임

# 2. 정규성 검정
pvalue = stats.shapiro(high_density['평균 사용량']).pvalue
pvalue
# pvalue = 0.28로 0.05 이상 -> 정규성 만족
pvalue = stats.shapiro(low_density['평균 사용량']).pvalue
pvalue
# pvalue = 0.40으로 0.05 이상 -> 정규성 만족
# 독립표본 T-검정 사용 가능

# 3. 등분산성 검정
pvalue = stats.levene(high_density['평균 사용량'], low_density['평균 사용량']).pvalue
pvalue
# pvalue = 0.22로 0.05 이상 -> equal_var = True

result = stats.ttest_ind(high_density['평균 사용량'], low_density['평균 사용량'], equal_var = True)
result

```

```

# 시작화 : Wi-Fi 사용량 분포 (트리맵)
treemap_data = density_usage_data[['자치구', '평균 사용량', '인구 밀도']]
fig = px.treemap(
    data_frame = treemap_data,
    path = ['자치구'], # 계층 구조
    values = '평균 사용량', # 타일 면적
    color = '인구 밀도', # 색상
    color_continuous_scale = 'Bluyl'
)

fig.update_layout(
    margin_t = 50, margin_l = 25, margin_r = 25, margin_b = 25,
    width = 800, height = 600,
    title_text = '자치구별 인구 밀도와 평균 사용량',
    title_font_size = 20
)

fig.write_html('C:/Users/r1b04/OneDrive/바탕 화면/김가람/2학년 2학기/데이터 사이언스/기말고사/treemap.html')
webbrowser.open('C:/Users/r1b04/OneDrive/바탕 화면/김가람/2학년 2학기/데이터 사이언스/기말고사/treemap.html')

# 3) 인구 밀도, 설치 수, 사용량의 관계

# 설치 수
install_data

# 시작화 (버블 차트)
bubble_data = pd.merge(density_usage_data, install_data, on = '자치구', how = 'inner')

sns.scatterplot(
    data = bubble_data,
    x = '인구 밀도',
    y = '평균 사용량',
    size = '설치 수',
    sizes = (20, 4000),
    hue = '자치구',
    alpha = 0.5,
    legend = False
)

plt.xlim(bubble_data['인구 밀도'].min() * 0.9, bubble_data['인구 밀도'].max() * 1.1)
plt.ylim(bubble_data['평균 사용량'].min() * 0.9, bubble_data['평균 사용량'].max() * 1.1)

for i in range(bubble_data.shape[0]):
    plt.text(
        x = bubble_data['인구 밀도'].iloc[i],
        y = bubble_data['평균 사용량'].iloc[i],
        s = bubble_data['자치구'].iloc[i],
        horizontalalignment = 'center',
        size = 'small',
        color = 'dimgray'
    )

plt.title('인구 밀도, 설치 수, 사용량 간의 관계', fontsize = 16)
plt.xlabel('인구 밀도', fontsize = 14)
plt.ylabel('Wi-Fi 사용량 (GB)', fontsize = 14)
plt.grid(True)
plt.show()

```

<복사본>

```
import pandas as pd
from scipy import stats
import numpy as np
import matplotlib.pyplot as plt
import webbrowser
import seaborn as sns
from scipy.stats import trim_mean
import folium
import geokakao as gk
import plotly.express as px
import matplotlib.image as mpimg
import plotly.graph_objects as go

# 한글
plt.rcParams['font.family'] = 'Malgun Gothic'
plt.rcParams['axes.unicode_minus'] = False

# 데이터셋 로드
Usage = pd.read_csv('C:/Users/rlb04/OneDrive/바탕 화면/김가람/2학년
2학기/데이터사이언스/기말고사/서울특별시_공공와이파이 AP별 사용량_10_13_2021/서울특별시 공공와이파이
AP별 사용량_고정형(20210501_20211013).csv', encoding = 'cp949')
Location = pd.read_csv('C:/Users/rlb04/OneDrive/바탕 화면/김가람/2학년
2학기/데이터사이언스/기말고사/서울시 공공와이파이 서비스 위치 정보.csv', encoding = 'cp949')
Population = pd.read_csv('C:/Users/rlb04/OneDrive/바탕 화면/김가람/2학년
2학기/데이터사이언스/기말고사/인구밀도.csv')
Time = pd.read_csv('C:/Users/rlb04/OneDrive/바탕 화면/김가람/2학년
2학기/데이터사이언스/기말고사/인터넷 이용시간.csv')

# 1. 공공 Wi-Fi 설치 현황과 이용 수요의 관계 분석

# 1) Wi-Fi 설치 수

# 데이터 전처리
# 특이값 (과천시) 제거
Location = Location[Location['자치구'] != '과천시']
Location

# 연도별
install_count_by_year = Location.groupby('설치년도').size()
install_count_by_year
```

```

# 지역별(구) - 가나다순 정렬
install_count_by_region = Location['자치구'].value_counts().sort_index()
install_count_by_region

# 총합
total_install_count = Location.shape[0]
total_install_count

# 전체 설치 수에서 각 지역의 비율
install_percentage_by_region = install_count_by_region / total_install_count * 100
install_percentage_by_region

# 시각화 : 설치 연도별 현황 (선 그래프)
plt.figure(figsize = (10, 6))
plt.plot(install_count_by_year.index, install_count_by_year.values, marker = 'o', color = 'b', label = '설치 수')
plt.title('연도별 공공 Wi-Fi 설치 수')
plt.xlabel('설치년도')
plt.ylabel('설치 수')
plt.grid(True)
plt.legend()
plt.show()

# 시각화 : 지역별 설치 현황 (막대 그래프)
plt.figure(figsize = (12, 8))
sns.barplot(x = install_count_by_region.index, y = install_count_by_region.values, palette = 'viridis')
plt.title('지역별 공공 Wi-Fi 설치 수')
plt.xlabel('자치구')
plt.ylabel('설치 수')
plt.xticks(rotation = 45)
plt.show()

# 시각화 : 지역별 설치 비율 (원 그래프)
plt.figure(figsize = (8, 8))
plt.pie(install_percentage_by_region, labels = install_percentage_by_region.index, autopct = '%.1f%%', startangle = 140, colors = sns.color_palette('pastel'))
plt.title('지역별 공공 Wi-Fi 설치 비율')
plt.show()

# 2) Wi-Fi 사용량

# 데이터 전처리
# 특이값 (과천시) 제거

```

```

Usage = Usage[Usage['자치구'] != '과천시']

# 지역별 사용량 통계 계산 - 가나다순 정렬
usage_stats = Usage.groupby('자치구')['AP별 이용량(GB)'].agg(
    평균 = 'mean',
    중앙값 = 'median',
    분산 = 'var',
    표준편차 = 'std',
    절사평균 = lambda x: trim_mean(x, 0.1)
).sort_index()
usage_stats

# 시각화 : 지역별 사용량 평균 (막대 그래프)
plt.figure(figsize = (12, 8))
usage_stats['평균'].plot(kind = 'bar', color = 'skyblue', alpha = 0.8, edgecolor = 'black')
plt.title('지역별 Wi-Fi 사용량 평균')
plt.xlabel('자치구')
plt.ylabel('평균 사용량 (GB)')
plt.xticks(rotation = 45)
plt.show()

# 시각화 : 지역별 사용량 분포 (상자 그림)
plt.figure(figsize = (12, 8))
sns.boxplot(x = '자치구', y = 'AP별 이용량(GB)', data = Usage, palette = 'coolwarm')
plt.title('지역별 Wi-Fi 사용량 분포')
plt.xlabel('자치구')
plt.ylabel('Wi-Fi 사용량 (GB)')
plt.xticks(rotation = 45)
plt.show()

# 3) Wi-Fi 설치 수와 사용량의 관계

# 데이터 전처리
# 두 데이터셋에 모두 존재하는 Wi-Fi만 추출
common_ids = set(Location['관리번호']) & set(Usage['관리번호'])
common_ids

filtered_location_data = Location[Location['관리번호'].isin(common_ids)]

```

```

filtered_usage_data = Usage[Usage['관리번호'].isin(common_ids)]

# 가나다순 정렬
filtered_location_data = filtered_location_data.sort_values(by = '자치구')
filtered_usage_data = filtered_usage_data.sort_values(by = '자치구')
filtered_location_data
filtered_usage_data

# 지역별 Wi-Fi 설치 수
install_counts = filtered_location_data.groupby('자치구').size().reset_index(name = '설치 수')
install_counts

# 지역별 설치 수가 다르므로 사용량은 평균을 내어 사용
usage_mean = filtered_usage_data.groupby('자치구')['AP별 이용량(GB)'].mean().reset_index(name = '평균 사용량')
usage_mean

# 두 변수 병합
cor_data = pd.merge(install_counts, usage_mean, on = '자치구').dropna()
cor_data

# 상관계수 계산
cor = cor_data['설치 수'].corr(cor_data['평균 사용량'])

# 시각화 : Wi-Fi 설치 수와 사용량 합계의 상관관계
plt.figure(figsize = (10, 6))
sns.regplot(
    x = '설치 수',
    y = '평균 사용량',
    data = cor_data,
    scatter_kws = {'color' : 'blue', 's' : 50},
    line_kws = {'color' : 'red'}
)
plt.title(f'Wi-Fi 설치 수와 평균 사용량의 상관관계: {cor:.2f}', fontsize = 16)
plt.xlabel('Wi-Fi 설치 수')
plt.ylabel('Wi-Fi 평균 사용량 (GB)')
plt.grid(True)

```

```
plt.show()

# 4) Wi-Fi 설치 위치 및 사용량 지도

# 데이터 병합
merged_data = pd.merge(filtered_location_data, filtered_usage_data, on = '관리번호', how = 'inner')
merged_data = merged_data.rename(columns = {'자치구_x' : '자치구'})
merged_data = merged_data[['자치구', '관리번호', '와이파이명', '도로명주소', '설치년도', 'X좌표', 'Y좌표',
'AP별 이용량(GB)']].dropna()

# 도로명주소를 주소로 변경 ex) 서소문로 51 -> 서울 서대문구 서소문로 51
merged_data['도로명주소'] = '서울' + ' ' + merged_data['자치구'] + ' ' + merged_data['도로명주소']
merged_data = merged_data.rename(columns = {'도로명주소' : '주소'})
merged_data

# 서울 중심 좌표 (기준점)
center = [37.5665, 126.9780]

# Folium 지도 생성
map = folium.Map(location = center, zoom_start = 11)

# 설치 위치
for _, row in merged_data.iterrows():

    folium.Marker(location = (row['Y좌표'], row['X좌표']),

icon = folium.Icon(color = 'blue', icon = 'signal')).add_to(map)

map.save('map.html')
webbrowser.open('map.html')

map = folium.Map(location = center, zoom_start = 12)

for _, row in merged_data.iterrows():

    folium.CircleMarker(
        location = (row['Y좌표'], row['X좌표']),
```

```

radius = row['AP별 이용량(GB)'] / 400,
color = 'blue',
fill = True,
fill_opacity = '80%',

tooltip = f'{row['와이파이명']}<br>사용량: {row['AP별 이용량(GB)']:.2f} GB'

).add_to(map)

map.save('map.html')
webbrowser.open('map.html')

# 2. 인터넷 사용 시간과 공공 Wi-Fi 사용량의 관계 분석

# 1) 인터넷 사용 시간

filtered_time = Time[['시점', '주 평균(시간)']].rename(columns = {'시점' : '연도', '주 평균(시간)' : '주 평균 이용 시간(시간)'})
filtered_time = filtered_time[filtered_time['연도'].str.isnumeric()]
filtered_time['연도'] = pd.to_numeric(filtered_time['연도'], errors = 'coerce')
filtered_time['주 평균 이용 시간(시간)'] = pd.to_numeric(filtered_time['주 평균 이용 시간(시간)'], errors = 'coerce')
filtered_time = filtered_time.dropna()
filtered_time

# 시각화 : 연도별 인터넷 사용 시간 변화 (선 그래프)
plt.figure(figsize = (10, 6))
plt.plot(filtered_time['연도'], filtered_time['주 평균 이용 시간(시간)'], marker = 'o', linestyle = '-', color = 'blue', label = '주 평균 이용 시간')
plt.title('연도별 인터넷 사용 시간 변화', fontsize = 16)
plt.xlabel('연도', fontsize = 14)
plt.ylabel('주 평균 이용 시간 (시간)', fontsize = 14)
plt.grid(True)
plt.legend(fontsize = 12)
plt.show()

```

```
# 2) Wi-Fi 설치 수와 인터넷 사용 시간의 관계
```

```
# 연도별 Wi-Fi 누적 설치 수
```

```
wifi_year_total = Location[['설치년도']].copy()
wifi_year_total['설치년도'] = pd.to_numeric(wifi_year_total['설치년도'], errors = 'coerce')
wifi_year_total = wifi_year_total.groupby('설치년도').size().cumsum().reset_index()
wifi_year_total.columns = ['연도', '누적 설치 수']
wifi_year_total = wifi_year_total.dropna()
wifi_year_total
```

```
# 병합
```

```
combined_data = pd.merge(filtered_time, wifi_year_total, on='연도', how='inner')
combined_data
```

```
# 상관분석
```

```
correlation = combined_data['주 평균 이용 시간(시간)'].corr(combined_data['누적 설치 수'])
correlation
```

```
# 시각화
```

```
combined_data.plot.scatter(x = '주 평균 이용 시간(시간)',
```

```
y
= '누적 설치 수')
m, b = np.polyfit(combined_data['주 평균 이용 시간(시간)'], combined_data['누적 설치 수'], 1)
plt.plot(combined_data['주 평균 이용 시간(시간)'], m * np.array(combined_data['주 평균 이용 시간(시간)']) +
b)
plt.title(f'Wi-Fi 누적 설치 수와 인터넷 사용 시간의 상관관계: {correlation:.2f}', fontsize = 16)
plt.xlabel('주 평균 인터넷 사용 시간 (시간)')
plt.ylabel('Wi-Fi 누적 설치 수')
plt.grid(True)
plt.show()
```

```
# 3. 공공 Wi-Fi 설치와 지역별 인구 밀도의 관계 분석
```

```
# 1) 인구 밀도와 설치 수의 관계
```

```
# 설치 수 데이터 (설치년도 2023년까지만)
```

```

install_data = merged_data[merged_data['설치년도'] <= 2023]
install_data = install_data.groupby('자치구').size().reset_index(name = '설치 수')
install_data

# 인구 밀도 데이터
density_data = Population.drop(columns = ['동별(1)'])
density_data = density_data.rename(columns = {'동별(2)' : '자치구'})
density_data = density_data.iloc[1:]
density_data

# 2023년 인구 밀도만 추출
density_2023 = density_data[['자치구', '2023']]
density_2023 = density_2023.rename(columns = {'2023' : '인구 밀도'})
density_2023

# 병합
merged_density = pd.merge(density_2023, install_data, on = '자치구', how = 'inner')
merged_density['인구 밀도'] = pd.to_numeric(merged_density['인구 밀도'], errors='coerce')
merged_density['설치 수'] = pd.to_numeric(merged_density['설치 수'], errors='coerce')
merged_density

# 상관분석
cor_density = merged_density[['인구 밀도', '설치 수']].corr().iloc[0, 1]
cor_density

# 시각화 : 인구 밀도와 Wi-Fi 설치 수의 관계
merged_density.plot.scatter(x = '인구 밀도',
y
= '설치 수')
m, b = np.polyfit(merged_density['인구 밀도'], merged_density['설치 수'], 1)
plt.plot(merged_density['인구 밀도'], m * np.array(merged_density['인구 밀도']) + b)
plt.title(f'인구 밀도와 공공 Wi-Fi 설치 수의 상관관계: {cor_density:.2f}', fontsize = 16)
plt.xlabel('인구 밀도')
plt.ylabel('Wi-Fi 설치 수')
plt.grid(True)
plt.show()

# 2) 인구 밀도와 사용량의 관계

```

```

# 인구 밀도와 사용량의 데이터 분포와 변동성 분석
# 사분위수, 사분범위, 표준편차

# 2021년 인구 밀도만 (Wi-Fi 사용량 통계가 2021년 기준임)
density_2021 = density_data[['자치구', '2021']]
density_2021 = density_2023.rename(columns = {'2021' : '인구 밀도'})
density_2021

usage_data = merged_data
usage_data = usage_data.groupby('자치구')[['AP별 이용량(GB)']].mean().reset_index(name = '평균 사용량')
usage_data

# 병합
density_usage_data = pd.merge(density_2021, usage_data, on = '자치구', how = 'inner')
density_usage_data['인구 밀도'] = pd.to_numeric(density_usage_data['인구 밀도'], errors='coerce')
density_usage_data['평균 사용량'] = pd.to_numeric(density_usage_data['평균 사용량'], errors='coerce')
density_usage_data

# 분석
quartiles = density_usage_data['평균 사용량'].quantile([0.25, 0.5, 0.75])
std_dev = density_usage_data['평균 사용량'].std()
quartiles
std_dev

# 시각화 : Wi-Fi 사용량 분포 (상자 그림)
plt.figure(figsize = (10, 6))
sns.boxplot(x = '평균 사용량',

data = density_usage_data,

palette = 'coolwarm')
plt.title('Wi-Fi 사용량 분포', fontsize = 16)
plt.xlabel('평균 Wi-Fi 사용량 (GB)', fontsize = 14)
plt.grid(True)
plt.show()

# 고밀도 지역과 저밀도 지역 간 사용량 차이 분석
median_density = density_usage_data['인구 밀도'].median()

```

```

high_density = density_usage_data[density_usage_data['인구 밀도'] >= median_density].reset_index()
low_density = density_usage_data[density_usage_data['인구 밀도'] < median_density].reset_index()

# 1. 독립 검정
# 고밀도 지역과 저밀도 지역은 다른 지역들이므로 독립된 표본임

# 2. 정규성 검정
pvalue = stats.shapiro(high_density['평균 사용량']).pvalue
pvalue
# pvalue = 0.28로 0.05 이상 -> 정규성 만족
pvalue = stats.shapiro(low_density['평균 사용량']).pvalue
pvalue
# pvalue = 0.40으로 0.05 이상 -> 정규성 만족
# 독립표본 T-검정 사용 가능

# 3. 등분산성 검정
pvalue = stats.levene(high_density['평균 사용량'], low_density['평균 사용량']).pvalue
pvalue
# pvalue = 0.22로 0.05 이상 -> equal_var = True

result = stats.ttest_ind(high_density['평균 사용량'], low_density['평균 사용량'], equal_var = True)
result

# 시각화 : Wi-Fi 사용량 분포 (트리맵)
treemap_data = density_usage_data[['자치구', '평균 사용량', '인구 밀도']]
fig = px.treemap(
    data_frame = treemap_data,
    path = ['자치구'], # 계층 구조
    values = '평균 사용량', # 타일 면적
    color = '인구 밀도', # 색상
    color_continuous_scale = 'Bluy'
)
fig.update_layout(
    margin_t = 50, margin_l = 25, margin_r = 25, margin_b = 25,
    width = 800, height = 600,
)

```

```

title_text = '자치구별 인구 밀도와 평균 사용량',
title_font_size = 20
)

fig.write_html('C:/Users/r1b04/OneDrive/바탕 화면/김가람/2학년
2학기/데이터사이언스/기말고사/treemap.html')
webbrowser.open('C:/Users/r1b04/OneDrive/바탕 화면/김가람/2학년
2학기/데이터사이언스/기말고사/treemap.html')

# 3) 인구 밀도, 설치 수, 사용량의 관계

# 설치 수
install_data

# 시각화 (버블 차트)
bubble_data = pd.merge(density_usage_data, install_data, on = '자치구', how = 'inner')

sns.scatterplot(
    data = bubble_data,
    x = '인구 밀도',
    y = '평균 사용량',
    size = '설치 수',
    sizes = (20, 4000),
    hue = '자치구',
    alpha = 0.5,
    legend = False
)

plt.xlim(bubble_data['인구 밀도'].min() * 0.9, bubble_data['인구 밀도'].max() * 1.1)
plt.ylim(bubble_data['평균 사용량'].min() * 0.9, bubble_data['평균 사용량'].max() * 1.1)

for i in range(bubble_data.shape[0]):

```

```
plt.text(  
    x=  
    bubble_data['인구 밀도'].iloc[i],  
    y=  
    bubble_data['평균 사용량'].iloc[i],  
    s=  
    bubble_data['자치구'].iloc[i],  
  
    horizontalalignment = 'center',  
    size = 'small',  
    color = 'dimgray'  
)  
  
plt.title('인구 밀도, 설치 수, 사용량 간의 관계', fontsize = 16)  
plt.xlabel('인구 밀도', fontsize = 14)  
plt.ylabel('Wi-Fi 사용량 (GB)', fontsize = 14)  
plt.grid(True)  
plt.show()
```