

# ‘영화관-인구수’ 상관관계 데이터 분석

영화관은 무슨 기준으로 생길까?

2022.11~, 김성지

# 순서

1. 가설 설정
2. 데이터 수집
3. 데이터 전처리
4. 시각화

# 1. 가설 설정

“**영화관**은 무슨 기준으로 생길까?”

1. 고정 지역 인구 / 2. 유동 지역 인구

# 1. 가설 설정

**상관관계** : 두 변수 간에 일정한 관계가 있음

**인과관계** : 원인A  $\rightarrow$  결과 B

## 2. 데이터 수집 (데이터기준, 출처)

영화관 데이터(2022.10)

### 1. 고정인구

: 주민등록 인구 데이터

### 2. 유동인구

: 전철 승하차 데이터

행정안전부

- 주민등록 인구데이터

서울시 열린 데이터광장

- 서울시 영화관 데이터

- 서울시 역별 승하차 데이터

국토교통부레일포털

- 지하철 역사 데이터

# 3. 데이터 전처리 (전)

서울시 전철 승하차 데이터 + 전국 전철 역사 데이터 = subway.csv

	A	B	C	D	E	F
1	사용일자	노선명	역명	승차총승객	하차총승객	등록일자
2	20220901	2호선	신대방	26990	26203	20220904
3	20220901	2호선	신림	61313	58762	20220904
4	20220901	2호선	봉천	26070	23873	20220904
5	20220901	2호선	서울대입구	51225	50545	20220904
6	20220901	2호선	낙성대(강)	30179	29402	20220904

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	역번호	역사명	노선번호	노선명	영문역사명	한자역사명	환승역구분	환승노선번호	환승노선명	역위도	역경도	운영기관명	역사도로명	역사전화번호	데이터기준일자	
2	101	판암	S3001	대전 도시	Panam	板岩	일반역			36.3169	127.4583	대전교통공	대전광역시	042-539-3	#####	
3	102	신흥	S3001	대전 도시	Sinheung	新興	일반역			36.31965	127.4488	대전교통공	대전광역시	042-539-3	#####	
4	103	대동	S3001	대전 도시	Dae-dong	大洞	일반역			36.32953	127.4428	대전교통공	대전광역시	042-539-3	#####	
5	104	대전	S3001	대전 도시	Daejeon	大田	일반역			36.33158	127.4331	대전교통공	대전광역시	042-539-3	#####	
6	105	중앙로	S3001	대전 도시	Jungangro	中央路	일반역			36.32866	127.4258	대전교통공	대전광역시	042-539-3	#####	
7	106	주요천	S3001	대전 도시	Ju-yocheon	主路川	일반역			36.32403	127.4100	대전교통공	대전광역시	042-539-3	#####	

### 3. 데이터 전처리 (과정)

	A	B	C	D	E	F
1	역번호	역명	역위도	역경도	역사도로명행정구역	
2	A01	서울	37.54912	126.9705	서울특별시용산구	
3	A02	공덕	37.54284	126.9513	서울특별시마포구	
4	A03	홍대입구	37.55774	126.9265	서울특별시마포구	
5	A04	디지털미디어	37.57683	126.899	서울특별시마포구	
6	A042	마곡나루	37.56582	126.8276	서울특별시강서구	

전국

+

	A	B	C
1	역명	승차총승객수	하차총승객수
2	4.19민주묘지	28082	27022
3	가농	48395	46177
4	가락시장	111685	116869
5	가산디지털단지	384069	402920
6	가양	146923	141822

서울(기준점)

	A	B	C	D	E	F	G	H
1	역번호	역명	역위도	역경도	역사도로명행정구역	승차총승객수	하차총승객수	
2	A02	공덕	37.54284	126.9513	서울특별시마포구	260898	265454	
3	A03	홍대입구	37.55774	126.9265	서울특별시마포구	541944	581567	
4	A04	디지털미디어	37.57683	126.899	서울특별시마포구	210985	211200	
5	A042	마곡나루	37.56582	126.8276	서울특별시강서구	146468	143058	
6	A05	김포공항	37.56212	126.8015	서울특별시강서구	186451	178736	

# 3. 데이터 전처리 (과정)

필요한 데이터 추출 → 데이터 그룹화 → 최종 데이터 추출

## 가. 전처리 데이터 병합

1. 타입 변경 및 특수문자 제거
2. 중복 제거
3. 역명을 기준으로 inner 병합

결측치 처리, 문제 가능성 있음

## 나. 데이터 그룹화

1. 그룹화
2. 행정구역 기준 연산

## 다. 최종데이터 추출

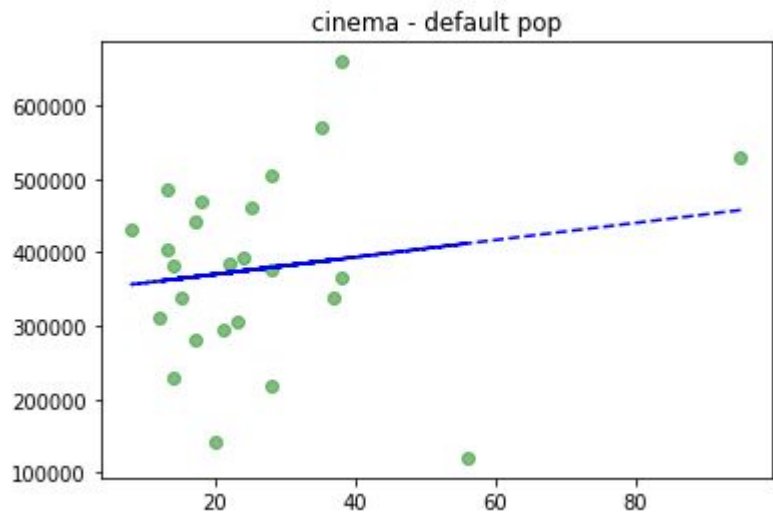
1. 행정구역(구) 기준으로 병합



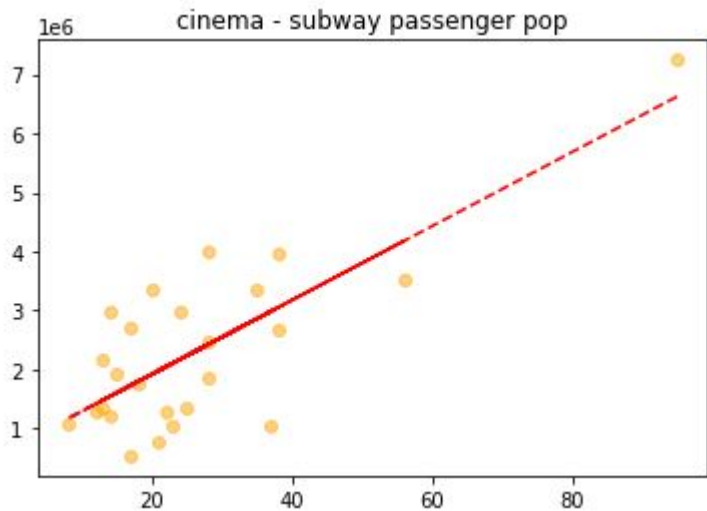
### 3. 데이터 전처리 (후)

	A	B	C	D	E
1	행정구역	상영관개수	총인구수	역사개수	총승객수
2	강남구	95	529946	32	7251659
3	강동구	25	460413	12	1359361
4	강북구	21	294429	11	755862
5	강서구	35	570011	22	3358883
6	관악구	13	487201	9	1358155
7	광진구	37	337900	11	1034741
8	구로구	24	394212	13	2989980
9	금천구	14	230156	4	1197447
10	노원구	28	505058	16	2457976

### 3. 데이터 분석 및 시각화

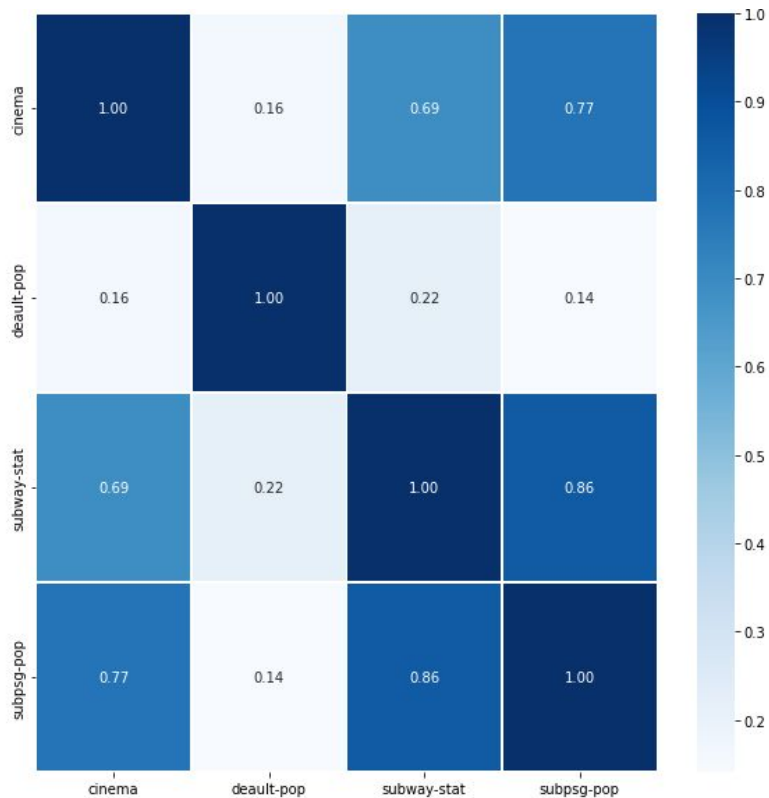


고정인구



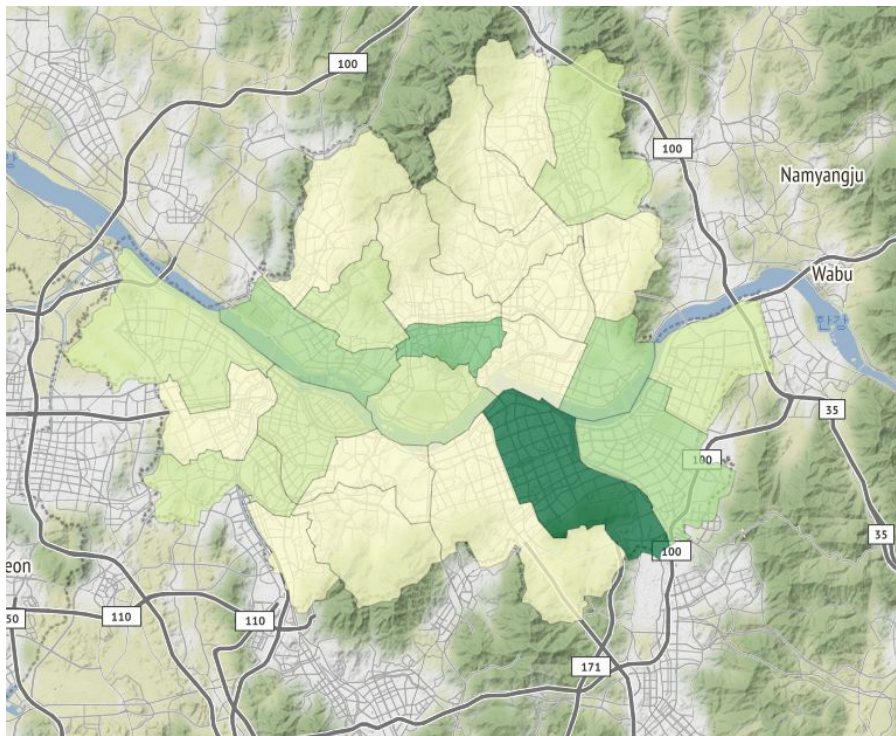
유동인구

### 3. 데이터 분석 및 시각화

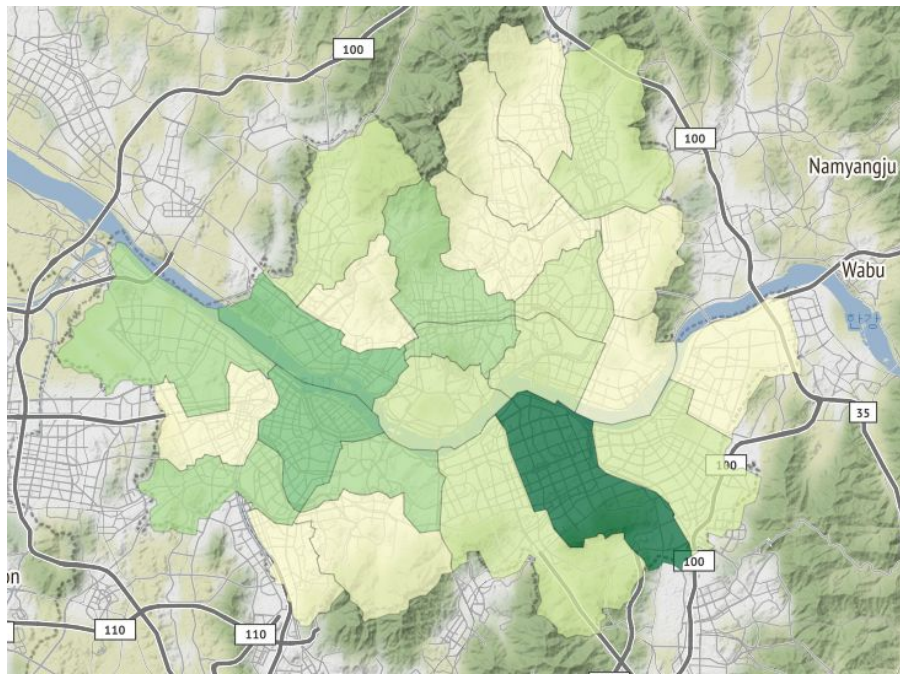


**히트맵** : 1에 가까울 수록 상관관계에 가까움

### 3. 데이터 분석 및 시각화



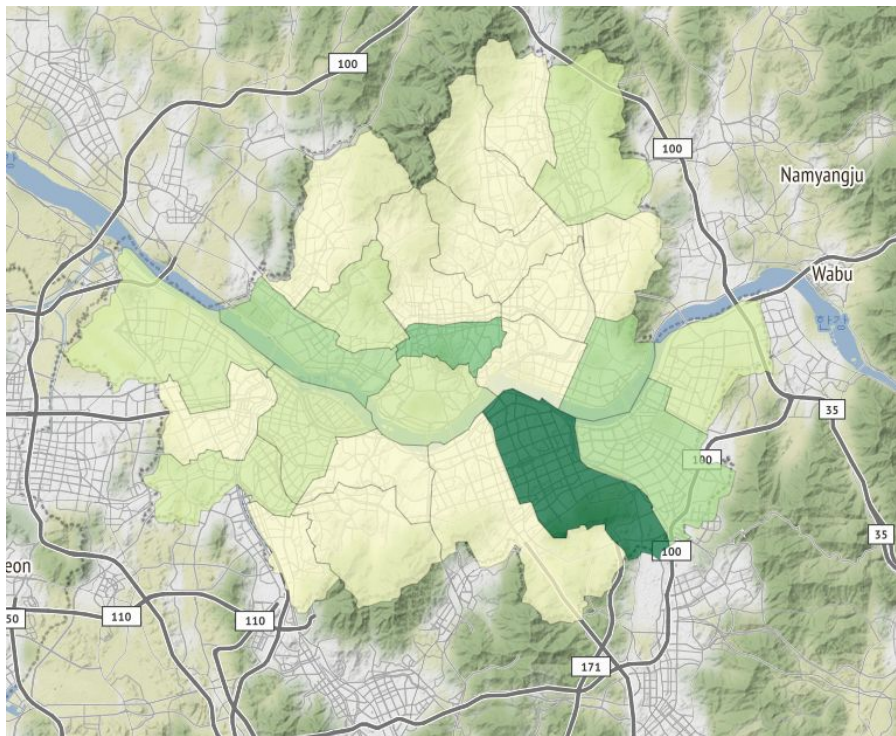
상영관수



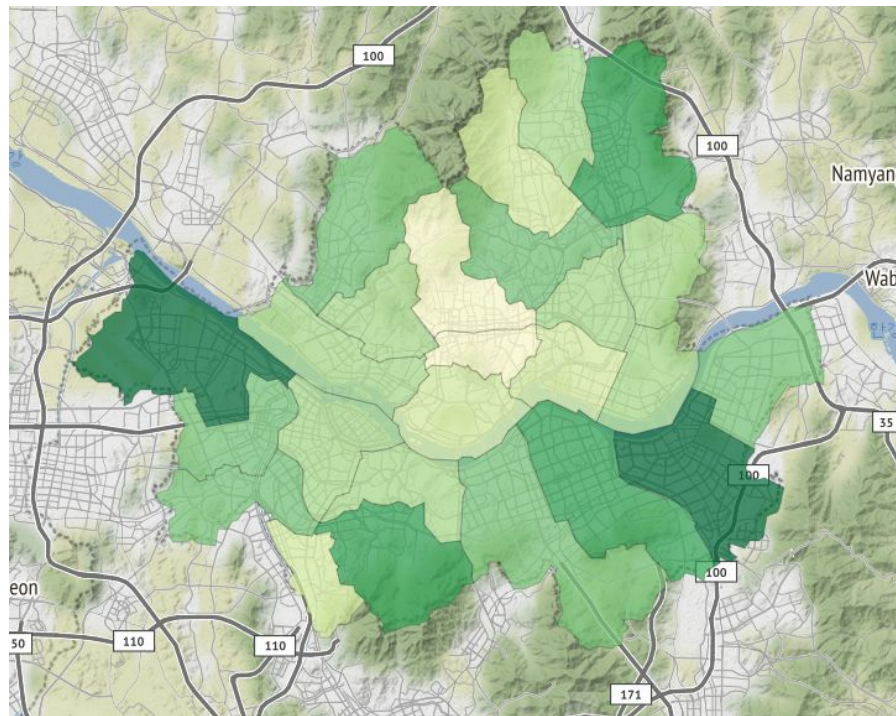
유동인구수-승객수



### 3. 데이터 분석 및 시각화



상영관수



고정인구수

# ‘상영관수-유동인구수’는 높은 양의 상관관계

<https://github.com/sseongji/datanalysis-cinema-population>