

지속 가능한 지하철 운영 지하철 광고 위치 추천을 통한 수익증가

김현동 신유빈 장순규 홍재민

2025. 6.11

목차

- 01 변경사항
- 02 연구 설계
- 03 데이터 전처리
- 04 클러스터링
- 05 결과
- 06 아쉬운 점

변경사항

↻ 지속되는 적자, 반복되는 악순환

적자 증가 → 운영 비용 상승 → 고령화로 인한 무임승차 증가 → 다시 적자 증가

적자 증가 → 우대권 조정 시도 → 사회적 반발 및 실패 → 다시 적자 증가

이러한 순환 구조 속에서 기존의 방식(요금 인상)이나 우대권 비율 조정만으로는 **문제 해결에 한계가 있다고 판단**

🎯 새로운 접근: 적자를 줄이기보다 '간접적인 수익을 늘리자'

적자를 줄이기 위한 단순한 접근은 지하철이 가진 복지적 기능을 약화시킬 수 있다.

데이터와 지하철역의 유동인구와 인프라를 활용한 **간접적 수익 창출 상승시킬 방법**을 모색하고자 한다.

이는 복지적 역할을 유지하면서도 지속가능한 운영 구조를 만드는 대안적 접근이 될 수 있다.

변경 사항

광고 데이터 클러스터링 -> 광고 데이터 수집의 어려움

광고데이터(단가, 제품종류, 광고 타겟, 계약금 등) 은 쉽게 공개하지 않음



지하철 데이터와 매치 -> 주 목표로 변경

각각의 변수들을 지하철 데이터와 매치 하겠다는 기존 목표와 달리 지하철역 데이터를 메인으로 설계



광고에 적합한 역 추천 -> 추천 방식의 변경이 필요

추천 모델이 아닌 광고를 설치하기 적합한 역을 찾는것을 목표로 변경

변경 사항

지하철역 클러스터링

지하철역 데이터 (유동인구, 광고단가, 광고시설물수, 평균대기시간 등) 을 사용한 클러스터링



클러스터 분석 , K-means를 사용

클러스터를 분석, 해석 하여 광고의 적합한 클러스터 찾기



광고시설 추가에 적합한 역 추천

최종 추천역 선정

연구 설계

1.역들을 분류하기 위해 필요한 데이터 전처리

2.클러스터링을 통해 역들을 분류

3.최종 광고 추천역 선정 및 광고전략 구성

데이터 전처리

수집해야할 데이터

광고를 몇 명이 접하는가?

→ 승객 총인원수: 유동인구가 많을수록 광고 도달 범위 증가

얼마나 오랫동안 광고를 접하는가?

→ 평균대기시간: 체류시간이 길수록 광고 노출 시간 증가

→ 환승역일 경우, 환승률에 따라 가중치 부여

광고 단가는 얼마인가?

→ 호선별 광고단가: 운영 비용 대비 효율성 판단 기준

광고 공간은 얼마나 남아 있는가?

→ 시설물면적 비율 (설치면적 / 전체면적): 광고 포화도를 나타냄

데이터 전처리

사용한 데이터

서울교통공사_1_8호선 역별 일별 승객유형별 수송인원(환승유입인원 포함) 정보

<https://www.data.go.kr/data/15104835/fileData.do>

서울교통공사_역사면적정보

<https://www.data.go.kr/data/15060120/fileData.do>

서울교통공사_역사별 광고시설 현황 정보

<https://www.data.go.kr/data/15112954/fileData.do>

서울교통공사_환승역거리 소요시간 정보

<https://www.data.go.kr/data/15044419/fileData.do>

호선별 광고단가 정보

<http://www.springbeat.co.kr/03/0101.php>

<https://goodadlab.co.kr/subway>

광고면적 정보

<https://jungsungad.com/65W>

<https://www.yeonsu.go.kr/main/>

데이터 전처리

서울교통공사_1_8호선 역별 일별 승객유형별 수송인원(환승유입인원 포함) 정보

	날짜	호선	역번호	역명	승객유형	승차인원	환승유입인
1	2023-01-01	1호선	150	서울역	일반	27121	14887
2	2023-01-01	1호선	150	서울역	어린이	331	161
3	2023-01-01	1호선	150	서울역	중고생	10	7
4	2023-01-01	1호선	150	서울역	청소년	611	384
5	2023-01-01	1호선	150	서울역	우대권	3454	1867
6	2023-01-01	1호선	150	서울역	직원	134	27
7	2023-01-01	1호선	150	서울역	영어 일반	163	78
8	2023-01-01	1호선	150	서울역	일어 일반	91	31
9	2023-01-01	1호선	150	서울역	중국어 일반	71	13
10	2023-01-01	1호선	150	서울역	영어 어린이	22	3
11	2023-01-01	1호선	150	서울역	일어 어린이	5	0
12	2023-01-01	1호선	150	서울역	중국어 어린이	10	0

데이터 전처리

서울교통공사_1_8호선 역별 일별 승객유형별 수송인원(환승유입인원 포함) 정보

1. 기본 집계 및 불필요 열 제거

날짜 및 연번 제거
역/호선/승객유형 기준으로 승차인원, 환승유입인원 합산

2. 승객유형 통합 및 제거

공통, 국가유공자 등 0.01% 미만의 승객유형 삭제
'어린이', '중고생', '청소년' → 미성년자,
'영어/중국어/일어' → '외국인'
'경로' → '우대권', '직원' → '일반' 으로 통합

3. 환승비율, 외국인비율, 장애인비율 계산

파생변수를 위한 환승비율과 분석에 사용할 가능성이 있는
외국인비율과 장애인 비율 생성

4. 병합을 위한 괄호추가 및 괄호 안 내용 제거

청량리(서울시립대입구) → 청량리
중복 역명은 호선번호 포함하여 구분 서울역(1), 서울역(4)

5. 광고단가 변수 생성

호선번호 기준으로 광고단가 설정
1,4,5,8 호선 = 30, 2호선 = 40,
3호선 = 35, 6호선 = 22

데이터 전처리

역명	총인원	환승비율	일반_총인	우대권_총	미성년자_	외국인비율	장애인비율	광고단가
서울역(1)	31076244	39.66	26374879	3827114	567108	0.6	0.39	30
시청(1)	15793785	43.36	13429113	1885868	298127	0.78	0.36	30
종각	23922725	45.07	20474174	2921538	381536	0.28	0.33	30
종로3가(1)	14536978	35.17	8865432	5361029	138013	0.38	0.81	30
종로5가	15570508	43.92	10070813	5062904	198260	0.86	0.67	30
동대문(1)	6746495	37.36	4287531	2242311	126101	0.48	0.86	30
신설동(1)	7766773	37.66	5477228	2083057	135577	0.23	0.69	30
제기동	10626789	42.94	4734211	5682339	88195	0.06	1.08	30
청량리	12617598	41.52	6987791	5296429	201580	0.07	0.98	30
동묘앞(1)	5918633	40.25	2771514	2899748	142471	0.36	1.42	30
시청(2)	12767970	32.67	11392426	1029580	207421	0.85	0.23	40
을지로입구	24364559	33.75	21873904	1840126	335759	1.07	0.22	40
을지로3가	12242820	32.58	10660054	1316645	149834	0.7	0.25	40

데이터 전처리

서울교통공사_역사별 광고시설 현황 정보

A	B	C	D	E	F	G	H
연번	호선	역번호	역명	광고시설물명	광고시설물 수량	광고대행사	데이터기준일자
1	1	150	서울역(1)	역구내조명광고	13	브랜드발전소	2024-11-28
2	1	150	서울역(1)	역구내포스터	0		2024-11-28
3	1	150	서울역(1)	멀티비전	1	나스미디어	2024-11-28
4	1	150	서울역(1)	디지털포스터	26	브랜드발전소	2024-11-28
5	1	150	서울역(1)	기둥조명	4	전홍	2024-11-28
6	1	150	서울역(1)	승강장안전문 조명(민자제외)	44	에스엘엠엔씨	2024-11-28
7	1	150	서울역(1)	승강장매립형	0		2024-11-28
8	1	150	서울역(1)	공간활용 프로모션	0	전홍	2024-11-28
9	1	150	서울역(1)	승강장음성광고	6	양진텔레콤	2024-11-28
10	1	150	서울역(1)	디지털종합안내도_디지털	11	싸인텔레콤	2024-11-28
11	1	150	서울역(1)	디지털종합안내도_인쇄식	0	싸인텔레콤	2024-11-28
12	1	150	서울역(1)	디지털종합안내도_조명	0	싸인텔레콤	2024-11-28
13	1	150	서울역(1)	스마트 정보 안내시스템	0		2024-11-28
14	1	150	서울역(1)	디지털보드	0	씨에스와이커두	2024-11-28
15	1	150	서울역(1)	지하철 출입구 상단벽면	0	씨제이씨지브0	2024-11-28
16	1	150	서울역(1)	승강장안전문 광고(민자)_일반형	0	유진메트로컴	2024-11-28
17	1	150	서울역(1)	승강장안전문 광고(민자)_난간형	0	유진메트로컴	2024-11-28
18	1	150	서울역(1)	승강장안전문 광고(민자)_동영상	0	유진메트로컴	2024-11-28
19	1	150	서울역(1)	역구내행선안내게시기	0		2024-11-28

데이터 전처리

서울교통공사_역사별 광고시설 현황 정보

1. 기본 집계 및 불필요 열 제거

연번 제거, 역 기준으로 대합실광고와 승강장광고로 나눠
시설물 수량 합산

2. 광고 시설물 면적 매핑

광고시설물을 승강장 광고와 대합실광고로 분리하여
면적 계산

3. 광고 사용 면적 계산

면적 × 수량 → 대합실사용면적
승강장면적 × 수량 → 승강장사용면적

4. 병합을 위한 괄호추가 및 괄호 안 내용 제거

청량리(서울시립대입구) → 청량리
중복 역명은 호선번호 포함하여 구분 서울역(1), 서울역(4)

데이터 전처리

광고시설물명	광고 유형	평균 크기 (mm)	면적 (m제곱)	출처
DID	디지털 패널	1,920 × 1,080	2.07	정성기획
가로기	배너형 광고	600 × 1,800	1.08	연수구청
기둥사이벽면	고정 벽면 광고	2,000 × 3,000	6.00	정성기획
대형 벽면	대형 벽면 광고	5,000 × 10,000	50.00	정성기획
벽면	고정 벽면 광고	2,000 × 3,000	6.00	정성기획
벽면(DID)	디지털 패널	1,920 × 1,080	2.07	정성기획
벽면/DID	디지털 패널	1,920 × 1,080	2.07	정성기획
벽면/DID/스크린도어	디지털 패널	1,920 × 1,080	2.07	정성기획
스크린도어	스크린도어 광고	2,000 × 1,300	2.60	정성기획
스크린도어(DID)	디지털 패널	1,920 × 1,080	2.07	정성기획
승강장 안전문	스크린도어 광고	2,000 × 1,300	2.60	정성기획
승강장 안전문/스크린도어	스크린도어 광고	2,000 × 1,300	2.60	정성기획
승강장안전문	스크린도어 광고	2,000 × 1,300	2.60	정성기획
유리벽면	고정 벽면 광고	2,000 × 3,000	6.00	정성기획

정성기획, 연수구청

원본데이터에 있는 광고시설물의 면적을 **추론**하는데 근거가 됨

데이터 전처리

광고시설물명	매핑된 광고 유형	추론 근거 키워드	추론 면적 (m²)
공간활용 프로모션	대형 벽면 광고	프로모션 = 대형 벽면 활용	50.00
기둥조명	조명형 광고	조명, 기둥 = 조명박스	2.16
디지털보드	디지털 패널	DID 계열	2.07
디지털종합안내도_디지털	디지털 패널	DID 계열	2.07
디지털종합안내도_인쇄식	고정 벽면 광고	안내도 + 인쇄 = 벽면	6.00
디지털종합안내도_조명	조명형 광고	안내도 + 조명	2.16
디지털포스터	디지털 패널	DID 계열	2.07
멀티비전	디지털 패널	멀티비전 = DID 계열	2.07
스마트 정보 안내시스템	디지털 패널	스마트 + 정보 + DID	2.07
승강장매립형	고정 벽면 광고	매립형 벽면	6.00
승강장안전문 광고(민자)_난간형	스크린도어 광고	승강장안전문	2.60
승강장안전문 광고(민자)_동영상	디지털 패널	동영상 = DID 계열	2.07
승강장안전문 광고(민자)_일반형	스크린도어 광고	일반형 = 스크린도어	2.60
승강장안전문 조명(민자제외)	스크린도어 광고	조명 + 안전문	2.60
승강장음성광고	비시각 매체 (면적 0 처리)	음성 전용, 면적 없음	0.00
역구내조명광고	조명형 광고	조명 광고	2.16
역구내포스터	고정 벽면 광고	포스터 = 벽면	6.00
역구내행선안내게시기	디지털 패널	안내 + 디지털 기능	2.07
제2기 교통카드 RF단말기 광고	디지털 패널	단말기 = 디지털 기기	2.07
지하철 출입구 상단벽면	고정 벽면 광고	벽면 키워드	6.00

수집한 면적 데이터를 기준으로

서울교통공사_역사별 광고시설 현황 정보

광고시설물명 열의 존재하는 값들의 면적을 매핑

데이터 전처리

역명	대합실사용면적	승강장사용면적
가락시장(3)	71.07	0
가락시장(8)	202.14	36
가산디지털단지	313.47	0
강남	371.24	720.68
강남구청	257.64	24
강동	179.1	18
강동구청	91.56	24
강변	79.77	592.8
개롱	105.12	24
개화산	45.12	24
거여	117.63	36
건대입구(2)	199.94	176.8
건대입구(7)	209.31	12
경복궁	99.63	10.4
경찰병원	54.54	0
고덕	113.61	30
고려대	169.95	18
고속터미널(3)	203.99	31.2
고속터미널(7)	224.07	30
공덕(5)	282.9	0
공덕(6)	176.19	36
공릉	105.72	36
광나루	113.7	18
광명사거리	166.05	18
광화문	344.25	0
광흥창	130.14	0
교대(2)	524.61	107.08

서울역(1)	115.38	114.4
서울역(4)	95.64	36.4
서초	170.49	0
석계	179.61	36
석촌	65.97	24
선릉	263.43	720.68
성수	152.52	0
성신여대입구	73.5	20.8
송정	103.26	0
송파	145.26	30
수락산	135.09	0
수서	90.33	7.8
수유	280.17	36.4
수진	61.35	30
숙대입구	205.38	10.4
송실대입구	144.18	30
시청(1)	75.9	114.4
시청(2)	232.2	49.4
신금호	70.98	0
신길	41.85	18
신정	115.38	114.4

데이터 전처리

서울교통공사_역사면적정보, 서울교통공사_환승역거리 소요시간 정보

연번	호선	역명	대합실면적	승강장면적
1	1	서울역	8725	2080
2	1	시청	6863	3784
3	1	종각	7737.43	2672.81
4	1	종로3가	5628	3683
5	1	종로5가	7822	2208
6	1	동대문	3400	2090
7	1	동묘앞	5700.56	4194.19
8	1	신설동	4164	3076
9	1	제기동	5105	2815
10	1	청량리	5354	1771
11	2	시청	9008	3553
12	2	을지로입구	7362.26	3046.4
13	2	을지로3가	4522	2664
14	2	을지로4가	4121	2501
15	2	동대문역사	4972	1568
16	2	신당	3701	2061

연번	호선	환승역명	환승노선	환승거리	환승소요시간
1	1	서울역	4호선	159	2:13
2	1	서울역	공항철도	309	4:18
3	1	서울역	경의중앙선	164	2:17
4	1	시청	2호선	101	1:24
5	1	종로3가	3호선	118	1:38
6	1	종로3가	5호선	312	4:20
7	1	동대문	4호선	194	2:42
8	1	동묘앞	6호선	96	1:20
9	1	신설동	2호선	159	2:13
10	1	신설동	우이신설선	129	1:48
11	1	청량리	경의중앙선	165	2:18
12	1	청량리	경춘선	203	2:49
13	1	청량리	수인분당선	180	2:30
14	2	시청	1호선	101	1:24
15	2	을지로3가	3호선	149	2:04
16	2	을지로4가	5호선	77	1:04
17	2	동대문역사	4호선	45	0:38
18	2	동대문역사	5호선	234	3:15

데이터 전처리

서울교통공사_역사면적정보, 서울교통공사_환승역거리 소요시간 정보

1. 기본 집계 및 불필요 열 제거

연번 제거 역 기준으로 대합실광고와 승강장광고로 나눠
시설물 수량 합산

2. 고려하지 않을 환승 노선 정리

1~8호선의 외의 환승노선은 고려하지 않음

3. 병합을 위한 괄호추가 및 괄호 안 내용 제거

청량리(서울시립대입구) → 청량리
중복 역명은 호선번호 포함하여 구분 서울역(1), 서울역(4)

데이터 전처리

역명	대합실면적	승강장면적	환승소요시간_초
서울역(1)	8725	2080	133
시청(1)	6863	3784	84
종각	7737.43	2672.81	0
종로3가(1)	5628	3683	98
종로5가	7822	2208	0
동대문(1)	3400	2090	162
동묘앞(1)	5700.56	4194.19	80
신설동(1)	4164	3076	133
제기동	5105	2815	0
청량리	5354	1771	0
시청(2)	9008	3553	84
을지로입구	7362.26	3046.4	0
을지로3가	4522	2664	124
을지로4가	4121	2501	64
동대문역사	4972	1568	38
신당(2)	3701	2061	202
상왕십리	4041	1917	0
왕십리(2)	7275	2602	72
한양대	3807	2167	0
뚝섬	4961	3423	0
성수	6807	2478	0
건대입구(2)	3969	2574	64
구의	3236	2348	0
강변	1791	2362	0

데이터 전처리

병합 및 파생변수 생성

1. 역명을 기준으로 병합

3개의 csv 파일에 모두 존재하는 역들의 한하여 역명을 기준으로 병합

2. 대합실 면적, 승강장 면적 사용비율 파생변수 생성

대합실사용면적 / 대합실면적 = 대합실사용면적비율

승강장사용면적 / 승강장면적 = 승강장사용면적비율

3. 평균 대기시간 파생변수 생성

기본 대기시간 60초에 환승시간이 존재하는 역들의 한하여 환승비율의 비례해 평균대기시간을 증가

'평균대기시간' = (1 - '환승비율') * 60 + '환승비율' * (60 + '환승소요시간')

4. 병합을 위한 괄호추가 및 괄호 안 내용 제거

청량리(서울시립대입구) → 청량리

중복 역명은 호선번호 포함하여 구분 서울역(1), 서울역(4)

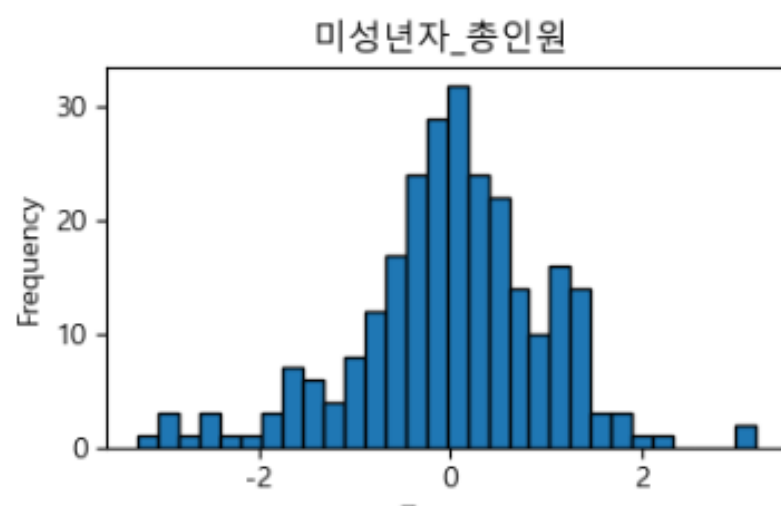
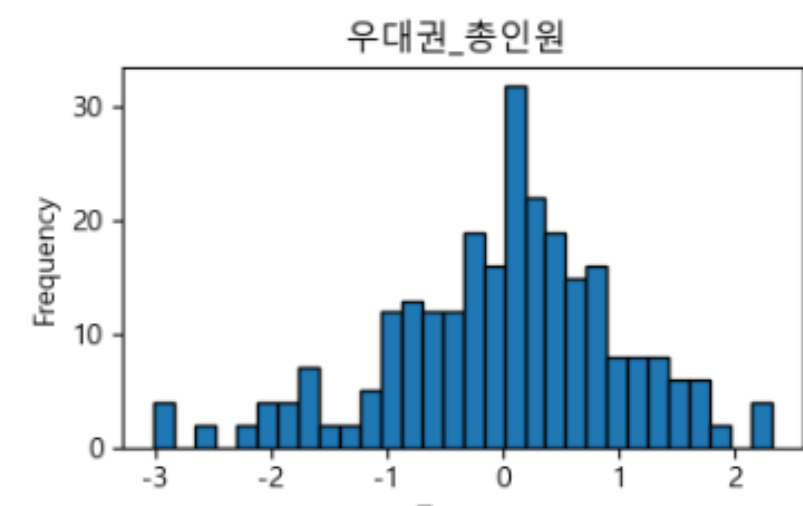
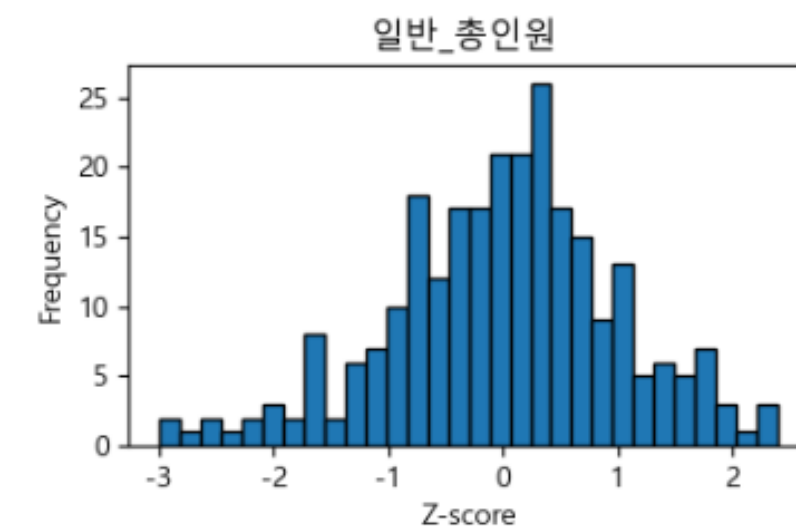
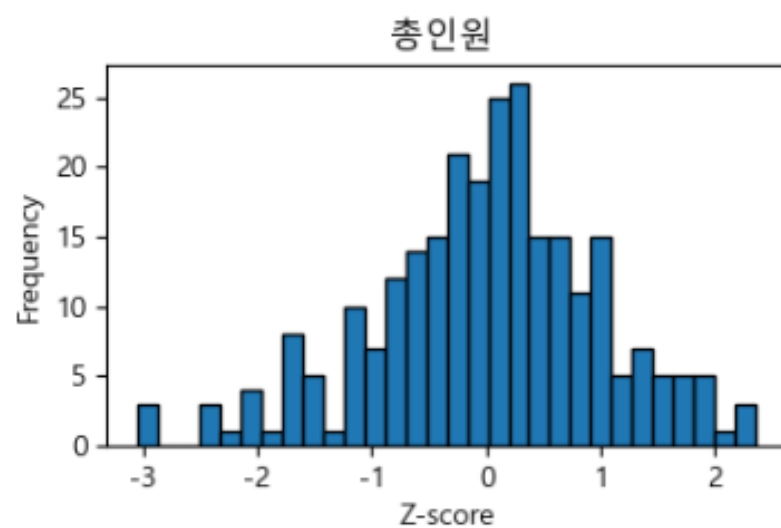
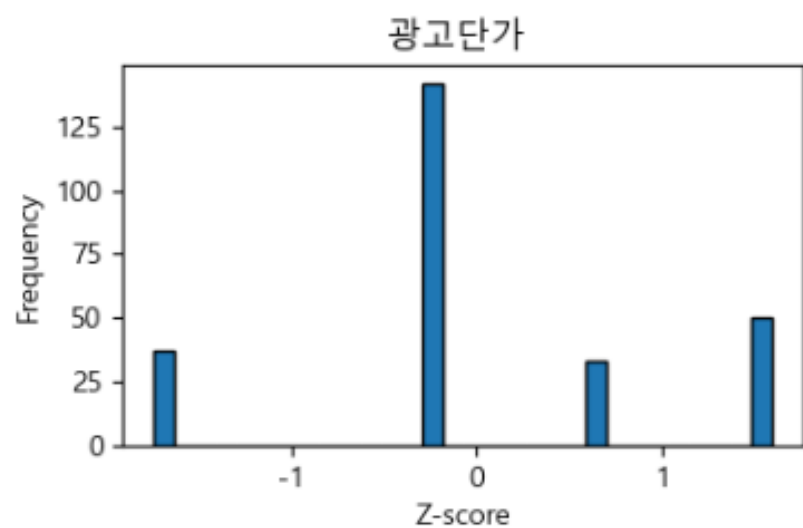
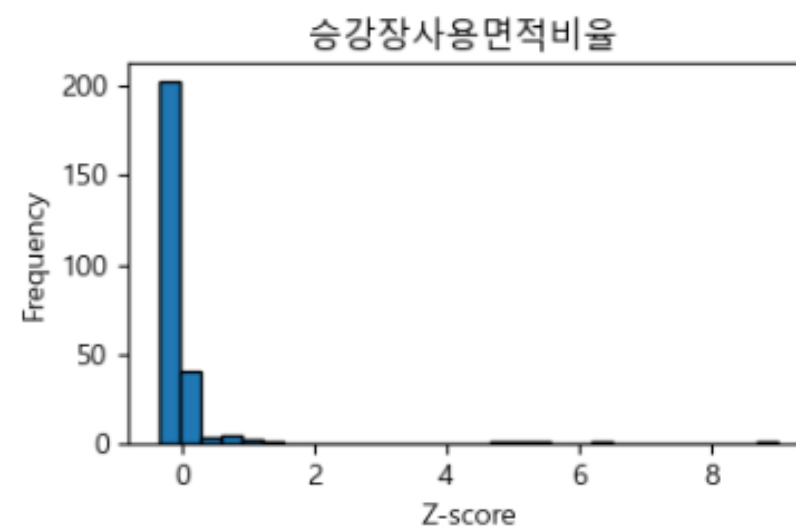
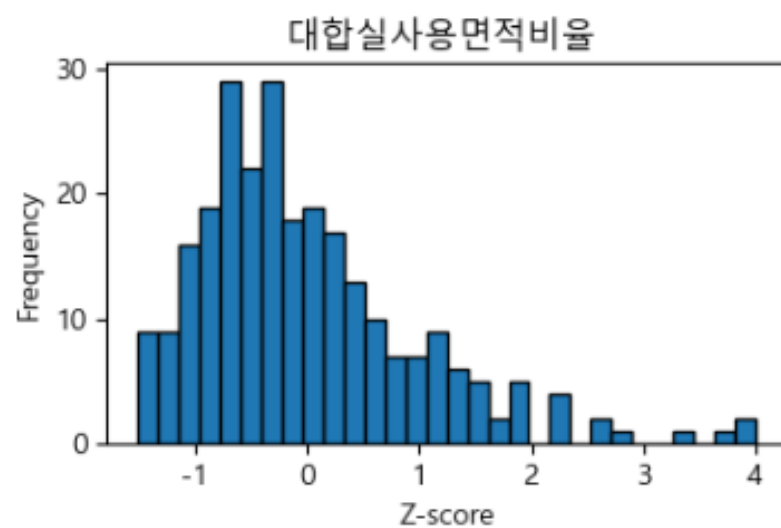
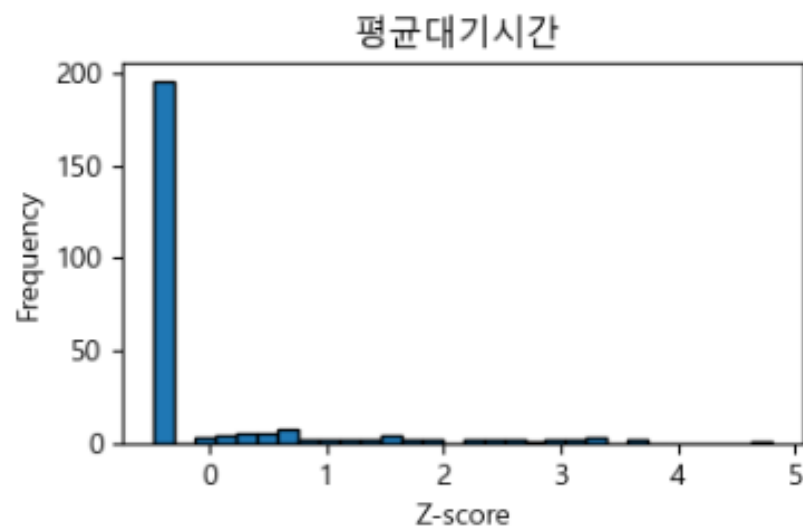
데이터 전처리

최종 데이터 263개의 역 / 표준화, 이상치 처리, 변수 선택 전

역명	평균인기	평균대기시간	승인현	일선_승인	주내권_승	비승인사_	내일승사현	승승승사현	외국인미출	승배인미출
서울역(1)	30	113	31076244	26374879	3827114	567108	0.013224	0.055	0.6	0.39
시청(1)	30	96	15793785	13429113	1885868	298127	0.011059	0.030233	0.78	0.36
종각	30	60	23922725	20474174	2921538	381536	0.020755	0.013619	0.28	0.33
종로3가(1)	30	94	14536978	8865432	5361029	138013	0.029403	0.031062	0.38	0.81
종로5가	30	60	15570508	10070813	5062904	198260	0.034675	0.003533	0.86	0.67
동대문(1)	30	121	6746495	4287531	2242311	126101	0.041815	0.004976	0.48	0.86
동묘앞(1)	30	92	5918633	2771514	2899748	142471	0.015756	0	0.36	1.42
신설동(1)	30	110	7766773	5477228	2083057	135577	0.037327	0	0.23	0.69
제기동	30	60	10626789	4734211	5682339	88195	0.021973	0.002771	0.06	1.08
청량리	30	60	12617598	6987791	5296429	201580	0.012865	0.008809	0.07	0.98
시청(2)	40	87	12767970	11392426	1029580	207421	0.025777	0.013904	0.85	0.23
을지로입구	40	60	24364559	21873904	1840126	335759	0.037647	0.236568	1.07	0.22
을지로3가	40	100	12242820	10660054	1316645	149834	0.017932	0.270526	0.7	0.25
을지로4가	40	78	6594063	5078624	1338185	85051	0.019124	0.004158	0.98	0.42
동대문역사	40	69	7006053	5681697	922077	277860	0.017287	0.072959	1.43	0.35
신당(2)	40	124	7456120	5619975	1620691	166654	0.029692	0	0.11	0.54
상왕십리	40	60	7313081	5832718	1225000	216962	0.025145	0	0.06	0.47
왕십리(2)	40	79	7653100	6488862	837189	293925	0.018812	0.008993	0.08	0.35
한양대	40	60	6038262	5518043	248431	256707	0.019866	0.005999	0.07	0.18
뚝섬	40	60	12822362	11415967	1065227	289134	0.021709	0	0.13	0.27
성수	40	60	21887525	19475681	1829505	452417	0.022406	0	0.24	0.35
거대이구(1)	40	77	17817163	15583072	1473725	604217	0.050375	0.068687	0.12	0.25

데이터 전처리

Z-score 표준화



데이터 전처리

이상치

Z-score 기준 ± 3 을 기준으로한 이상치가 존재함

대합실사용면적비율 이상치 (4개)

삼성 3.9751

강남 3.9974

교대(2) 3.4198

지축 3.7774

평균대기시간 이상치 (9개)

신당(2) 3.0786

대림(2) 3.2451

노원(4) 3.3561

서울역(4) 3.2451

종로3가(5) 3.5782

신당(6) 3.1341

노원(7) 3.2451

대림(7) 3.5782

연신내 4.7995

승강장사용면적비율 이상치 (6개)

을지로입구 4.7823

을지로3가(2) 5.5171

강변 5.0940

삼성 9.0014

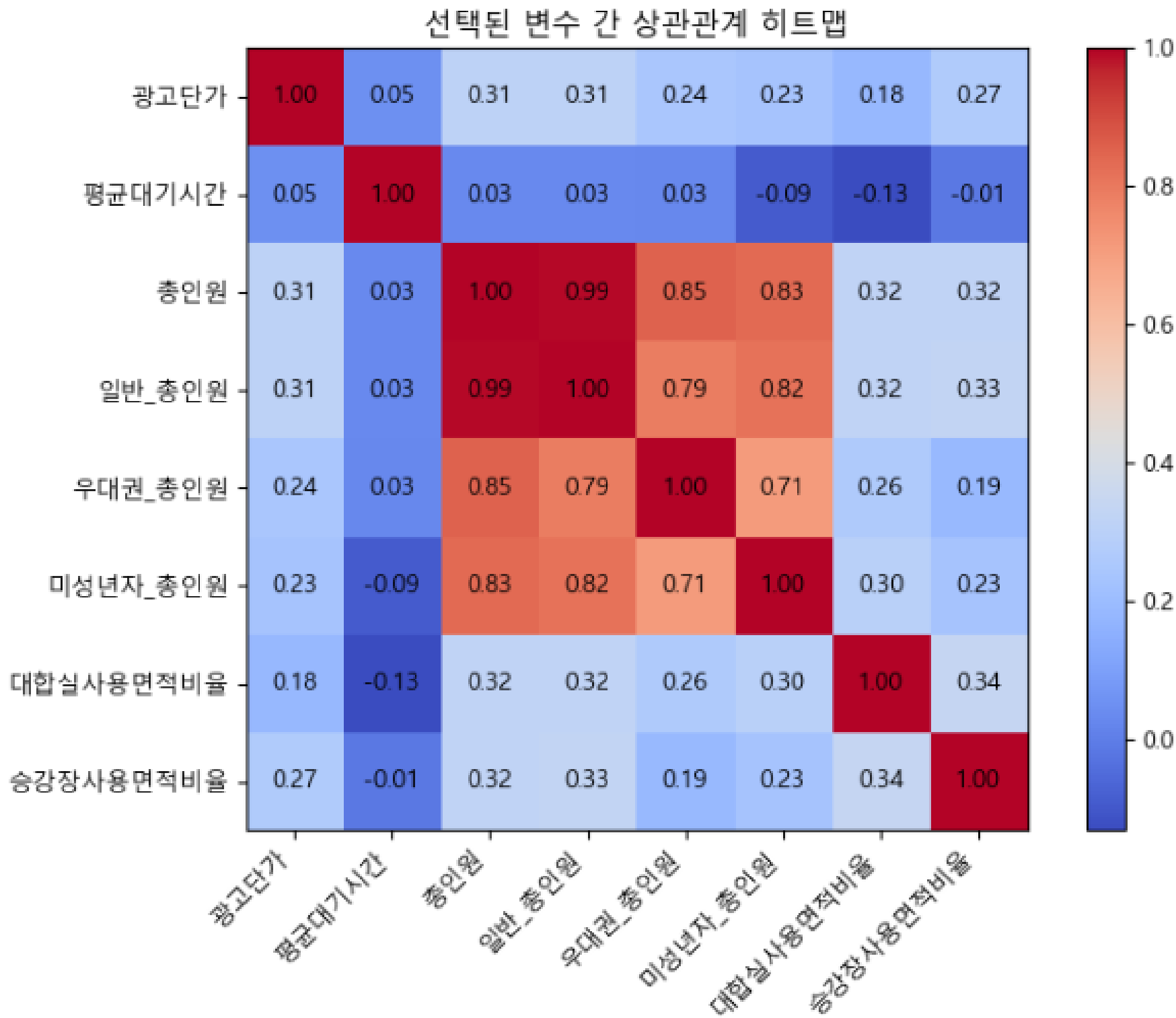
선릉 6.3477

강남 6.2322

데이터 전처리

상관관계 분석

피어슨 상관계수 사용



클러스터링 - 최종 변수

1. 일반_총인원

인구수 관련 변수중 가장 소비력이 높은 구간을 선택

2. 대합실사용면적비율

승강장사용면적비율의 이상치는 승강장사용면적이0인 역들을 해결하지 못하면 조절할 수 없다고 판단

3. 평균대기시간

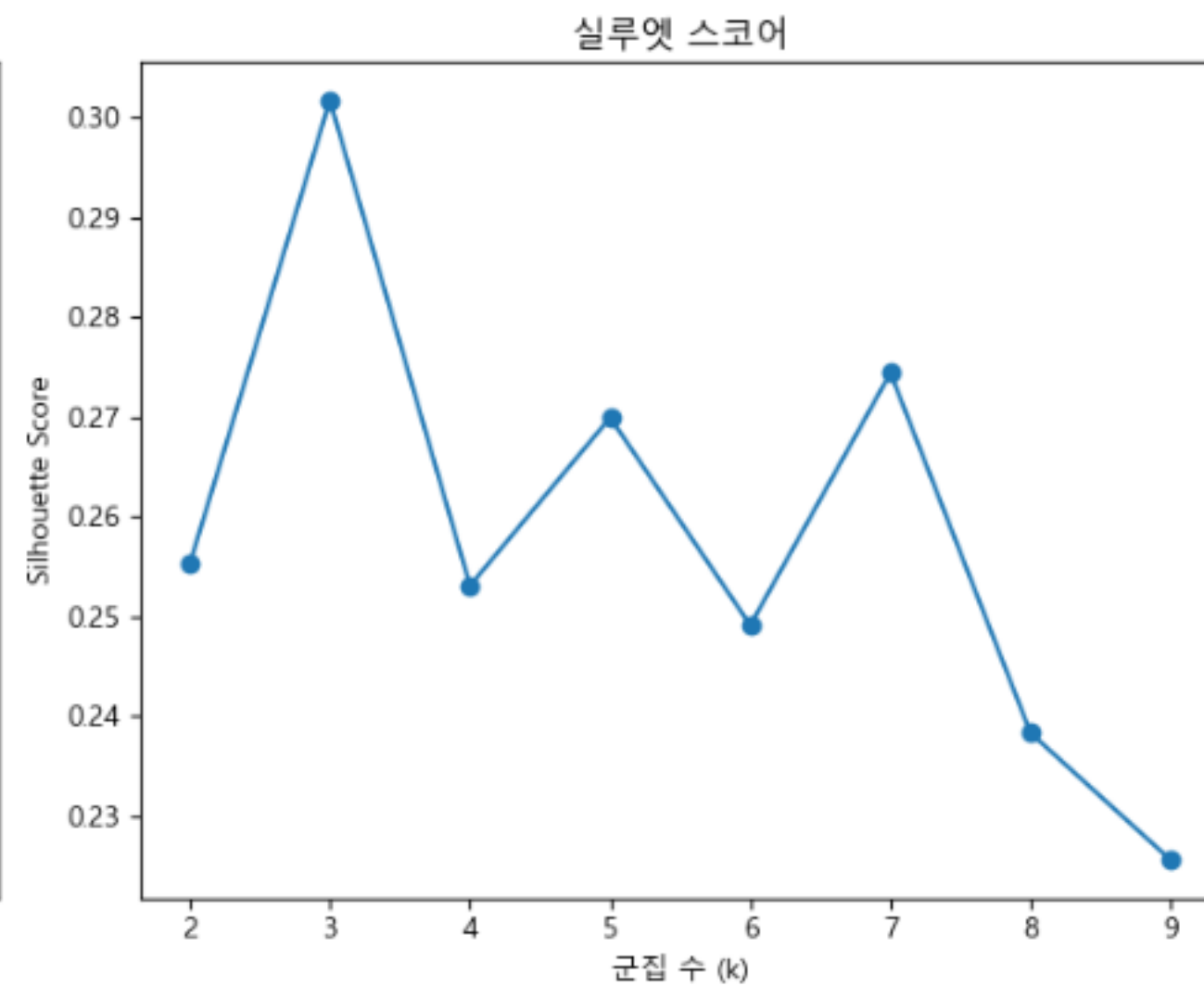
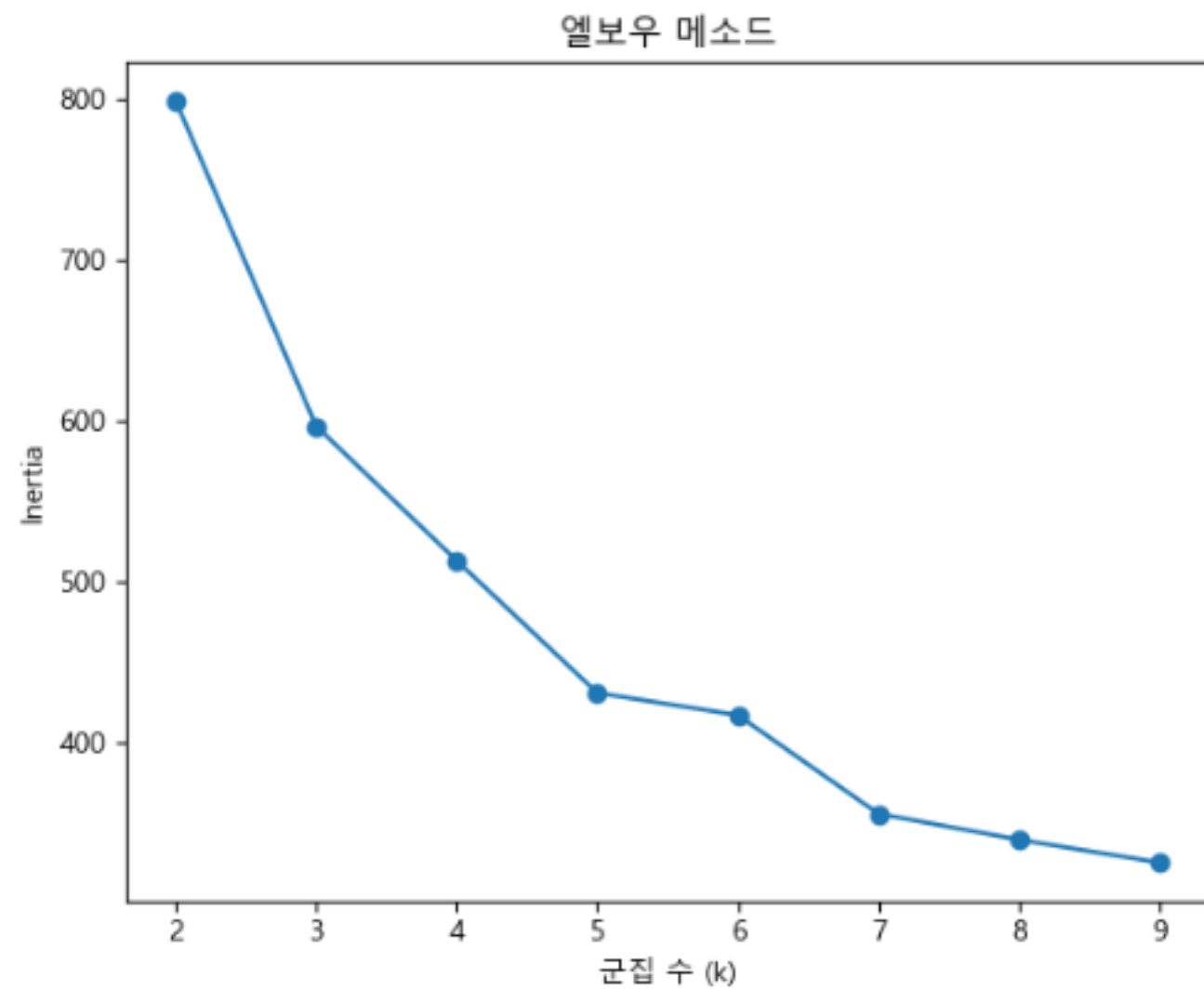
4. 광고단가

클러스터링

클러스터 개수 설정

실루엣 스코어, 엘보우 메소드 사용하여 적합 클러스터 개수 설정

일반_총인원, 대합실사용면적비율, 평균대기시간, 광고단가 모두 사용할 경우 -> 만족스럽지 못함

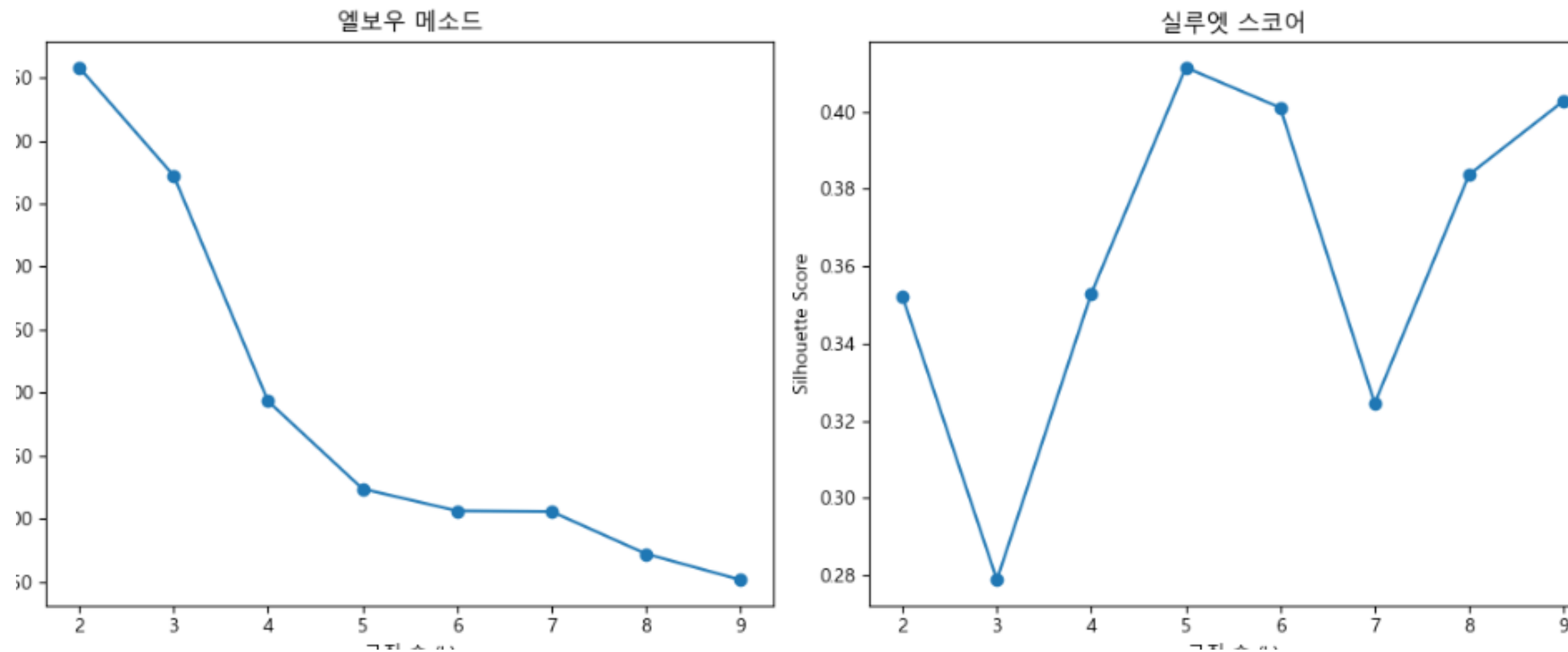


클러스터링

클러스터 개수 설정

실루엣 스코어, 엘보우 메소드 사용하여 적합 클러스터 개수 설정

일반_총인원, 평균대기시간, 광고단가 사용할 경우 -> 조금 나아지긴 했지만 만족스럽지 못함

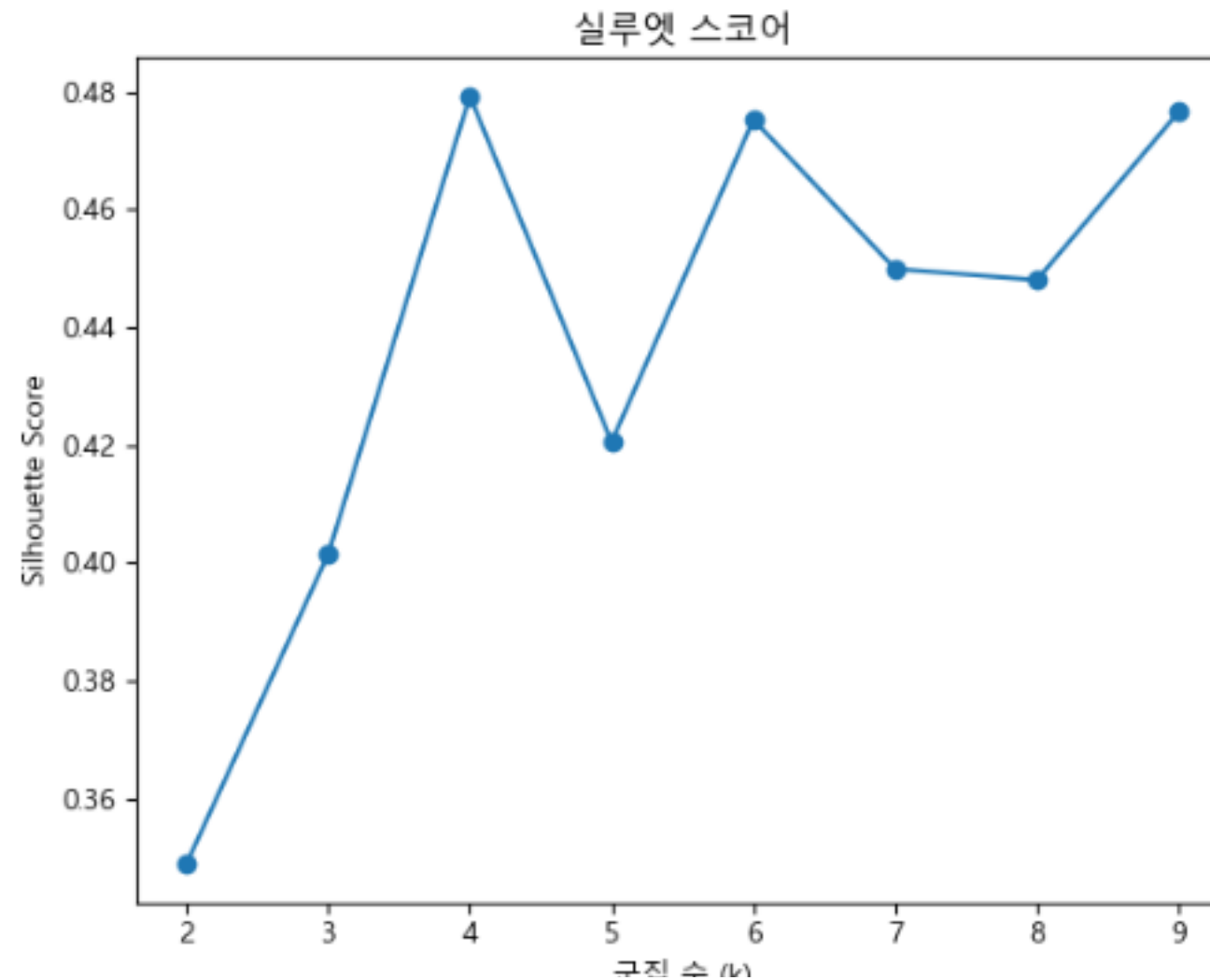
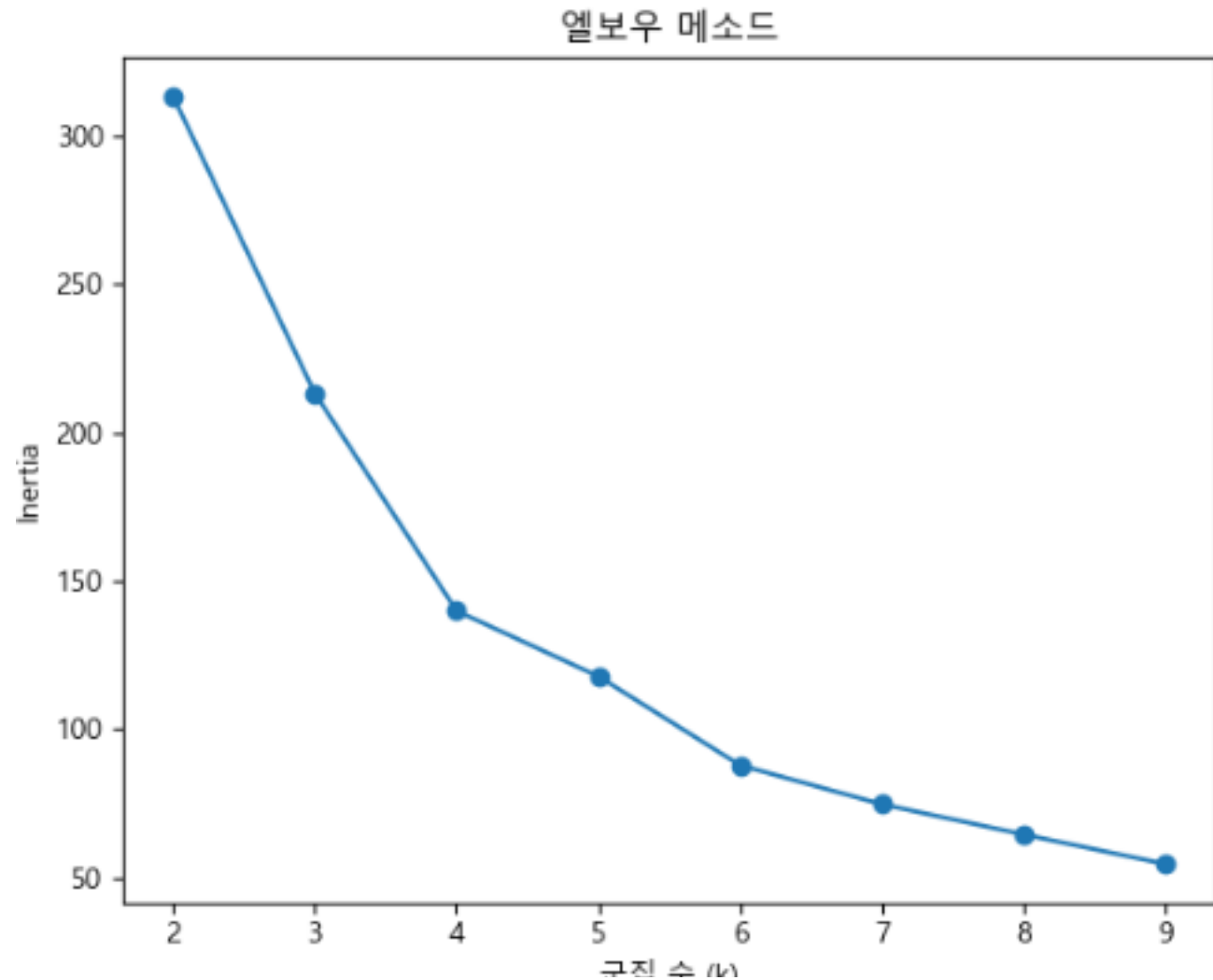


클러스터링

클러스터 개수 설정

실루엣 스코어, 엘보우 메소드 사용하여 적합 클러스터 개수 설정

일반_총인원, 광고단가 사용할 경우 -> 실루엣 스코어는 가장 높았으나, 변수 수가 적어 클러스터링 결과가 만족스럽지 못할것이라 예상가능



클러스터링

클러스터 결과 확인 및 시각화

클러스터 0 역 목록 (총 24개)

서울역(1), 시청(1), 종각, 종로3가(1), 종로5가, 동대문, 신설동(1), 등등

클러스터 2 역 목록 (총 15개)

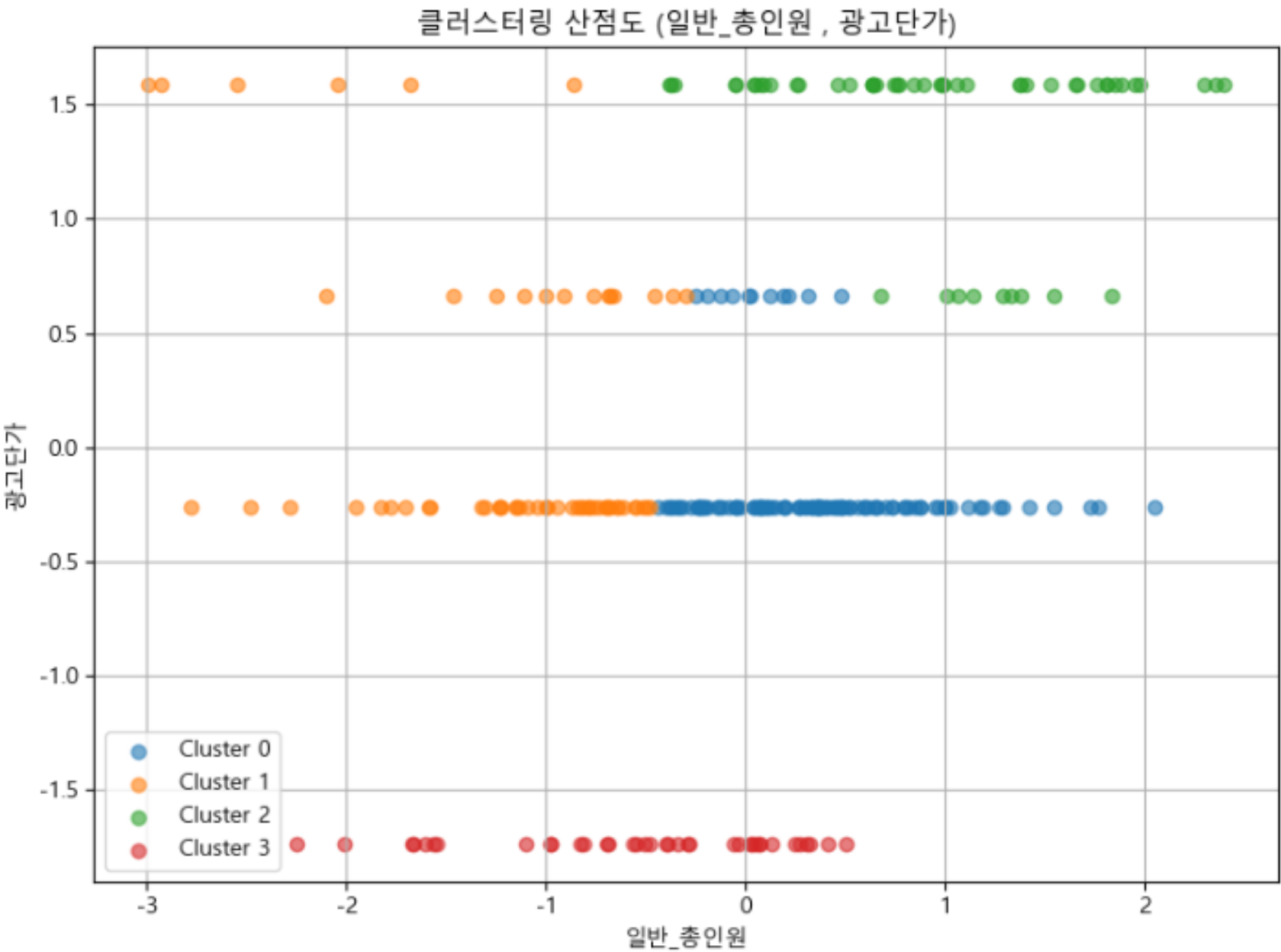
강남, 교대(2), 서초, 방배, 사당(2), 이수(4), 남성, 송실대입구, 등등

소속 클러스터간 극단적인 차이를 보이는 역이 존재

예 : 서울역(1) vs 대림(7)

인원 차이: 약 2.53백만 명

클러스터	일반_총인원 평균	광고단가 평균	해석 요약
0	0.35	-0.17	유동인구는 다소 많고, 광고단가는 낮음 → 가성비 좋은 역
1	-1.17	0.12	유동인구는 적고, 광고단가는 높음 → 광고주 입장에서 비효율적, 비추천
2	0.99	1.43	유동인구와 광고단가 모두 매우 높음 → 핵심 프리미엄 역
3	-0.54	-1.74	유동인구와 광고단가 모두 낮음 → 저가 광고 전략 가능



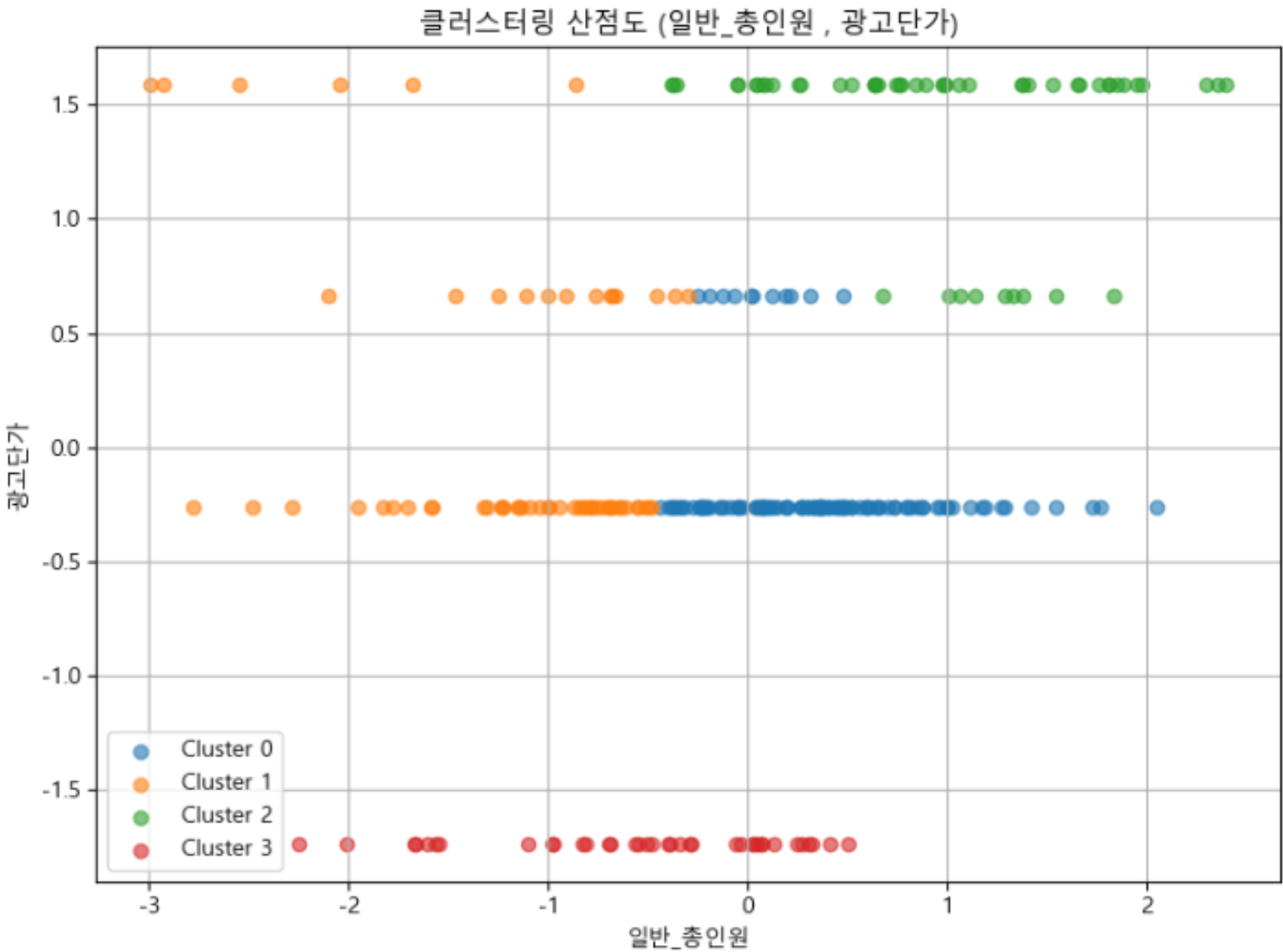
클러스터링

클러스터 결과 확인 및 시각화

광고사 데이터를 기반으로 임의로 설정한
광고단가가 4종류밖에 없고 변수의 수가 너무
적어서, 분석 결과의 신뢰도가 낮아졌다

이 상태로 분석을 종료하기에는 해석력이 부족

클러스터	일반_총인원 평균	광고단가 평균	해석 요약
0	0.35	-0.17	유동인구는 다소 많고, 광고단가는 낮음 → 가성비 좋은 역
1	-1.17	0.12	유동인구는 적고, 광고단가는 높음 → 광고주 입장에서 비효율적, 비추천
2	0.99	1.43	유동인구와 광고단가 모두 매우 높음 → 핵심 프리미엄 역
3	-0.54	-1.74	유동인구와 광고단가 모두 낮음 → 저가 광고 전략 가능



클러스터링

클러스터링 결과가 과도하게 단순화되는 문제 해결하기 위해
각 클러스터별 대합실사용면적비율, 평균대기시간을 변수로 서브 클러스터 생성
각 클러스터 모두 실루엣 스코어, 엘보우 메소드 기준 가장 적합한 k=3 로 kmeans 클러스터링 후
각각 광고추천 **클러스터, 서브 클러스터 조합 선정 (01 , 12, 22, 31)**

클러스터0 서브클러스터	대합실사용면적비율	평균대기시간
0	-0.363	-0.384
1	-0.238	2.454
2	1.190	-0.327

서브클러스터 1: 대기시간이 길고 공간 효율이 보통임

클러스터2 서브클러스터	대합실사용면적비율	평균대기시간
0	2.330	-0.308
1	-0.144	-0.243
2	-0.422	2.769

서브클러스터 2: 대기시간이 가장 길고 공간 효율도 가장 높음

클러스터1 서브클러스터	대합실사용면적비율	평균대기시간
0	-0.788	-0.360
1	1.387	-0.424
2	-0.222	1.959

서브클러스터 2: 대기시간이 길고 공간 효율이 양호함

클러스터3 서브클러스터	대합실사용면적비율	평균대기시간
0	-0.360	-0.410
1	-0.433	1.350
2	1.579	-0.474

서브클러스터 1: 대기시간이 길고 공간 효율이 높음

결과

클러스터링 결과를 토대로 최종 광고추천(광고시설물개발추천) 지하철역 선정 조건 두가지

이전에 정한 조건을 만족하는 서브클러스터 내에서 '평균대기시간 ↑ & 포화도 ↓' 기준 상위 15개 역

종로3가(5), 대림(7), 노원(4), 서울역(4), 노원(7), 동대문(1), 동대문(4), 잠실(8), 서울역(1), 신설동(1), 종로3가(3), 고속터미널(7), 시청(1), 을지로3가(3), 약수(3), 동대문역사문화공원(5), 충정로(5), 삼각지(4), 신설동(2), 동묘앞(1), 청구(5), 연신내, 대림(2), 신당(2), 충정로(2), 고속터미널(3), 잠실(2), 을지로3가(2), 신당(6), 약수(6), 삼각지(6), 동묘앞(6), 청구(6), 불광(6), 공덕(6)

서브클러스터 없이 대합실사용면적비율 은 하위 25% 대기시간은 상위 25% 인 역들

서울역(4), 서울역(1), 종로3가(3), 시청(1), 을지로3가(3), 영등포구청(5), 오금(3), 을지로4가(5), 연신내, 충정로(2), 고속터미널(3), 신당(6), 동묘앞(6)

결과

두가지 조건을 모두 만족하는 최종 추천역들

일반적으로 알려져 있지만 여전히 발전 가능성 있는 역들
서울역(1), 서울역(4), 을지로3가(3),
시청(1), 고속터미널(3)

-
- 1 종로3가(3) (0,1)**
광고단가가 상대적으로 낮음, 유동인구 많음
상대적으로 광고 경쟁 약함
광고에 많은 돈을 투자할 수 없는 제품들에게 추천
 - 2 연신내 (2, 2)**
대기시간 가장 김 (4.8), 면적비도 낮음
3호선+6호선 환승, 경기북부 거점
단점: 광고단가 높음 → 예산 충분할 때 타겟
몰입도가 높은 영상광고 추천
 - 3 신당 (2, 2)**
유동인구는 적지만 체류시간 길다.
단점: 광고단가 높음 → 예산 충분할 때 타겟
몰입도가 높은 영상광고 추천

아쉬운 점

데이터 수집 단계

정확한 광고 단가 정보를 확보하지 못해, 역별·시설별 단가를 세분화하기 어려웠다.

이를 보완하기 위해 호선별 평균 광고 단가를 임의 기준으로 설정했지만, 실제 가격과의 괴리로 인해 분석 결과의 정확성과 신뢰도에 한계가 있었다.

데이터 전처리 단계

환승역의 평균 대기시간을 처리하는 과정에서, 환승 통합역에 적절한 가중치를 부여하지 못했다.

그 결과, 대기시간 변수의 분포가 올바르게 못하고 왜곡되는 문제가 발생했으며, 이는 클러스터링 결과에도 영향을 미칠 가능성이 있다.

결론 및 모델링 단계

초기 클러스터링 결과의 해석력이 부족하여 서브 클러스터링을 시도했지만 여전히 의문이 남아있다.

또한, 광고 효율성을 직접 평가할 수 있는 수익·노출 등의 정량적 지표가 없어, 클러스터 결과를 객관적으로 검증할 수단이 부족했다.