

소속

상명대학교



공공데이터 분석을 통한
서울시 지하철
만성적자
해소 정책 제안

2조

팀명

파이썬 기반 빅데이터 분석을 통한 비즈니스 인사이트 역량제고 과정



서론

- 주제 선정 배경
- 분석 방향 및 목표

본론

- EDA
- 분석 내용 및 결과

결론

- 정책 제언
- 마무리

Part 1, 서론

Q.

우리가 매일같이 이용하는 지하철

당장 내일이라도 지하철 운영이
힘들다는 사실, 알고 계셨나요?



주제 선정 배경

지하철 적자 현황

적자 폭이 커져 당장 이달부터 직원들 월급 걱정까지 해야 한다던 서울교통공사가 결국 서울시에 2800억원 규모의 긴급 단기융자를 요청했다. 서울교통공사는 만성화된 적자 운영 구조로 매년 5000억원대 손실을 내왔고, 올 들어서는 코로나19 여파로 승객이 20% 이상 감소하고 방역비용 등이 증가하면서 이미 사상 최대 규모인 1조원에 달하는 순손실이 예상돼 왔다. 직원 복리후생을 줄이고, 내년도 채용 계획을 축소하는 등 비상 경영에 들어갔지만 당장 연말에 한꺼번에 만기가 도래하는 기업어음(CP) 5400억원을 해결할 방안도 요원하다.

적자는 늘어나는데 요금 인상도, 정부 지원도 이뤄지지 않는다면 가장 우려되는 부분은 안전에 대한 투자다. 1호선을 기준으로 이미 40년을 넘어선 서울지하철은 노후 전동차 교체나 철로 보수, 시설 개량 등 곳곳이 손봐야 할 곳 투성이다. 하루 평균 800만명 이상이 이용하는 서울지하철이 시민의 안전을 담보로 위험천만하게 달리는 일은 없어야 한다.

적자 만성화

전년도 1조원

초래하는 문제점

시민의 안전

지하철 재정난

2015년 이후 동결...운송원가 크게 밀돌아

지하철 요금 인상은 수년째 적자 상황을 면치 못하고 있는 서울교통공사(이하 공사)의 재정 상황을 얘기할 때 빠지지 않는 부분이다. 운수사업이 매출 대부분을 차지하고 있는 만큼 지하철 요금이 공사 수익을 결정짓기 때문이다.

특히 지하철 요금은 무임승차 손실과 함께 누적 공사 재정난의 원인 중 하나로 꼽힌다. 현행 지하철 요금(1250원)이 지하철 운송원가를 크게 밀돌고 있어서다. 2020년 기준 서울 지하철 운송원가는 2081원으로 승객 1명을 태울 때마다 830원가량의 적자를 보고 있다. 여기에 무임승차 등을 반영하는 평균운임을 고려하면 손해는 더 심각하다. 서울 지하철의 평균운임은 953원으로 1100원가량의 손해가 발생해 지하철이 운행될수록 적자가 커지는 구조다.

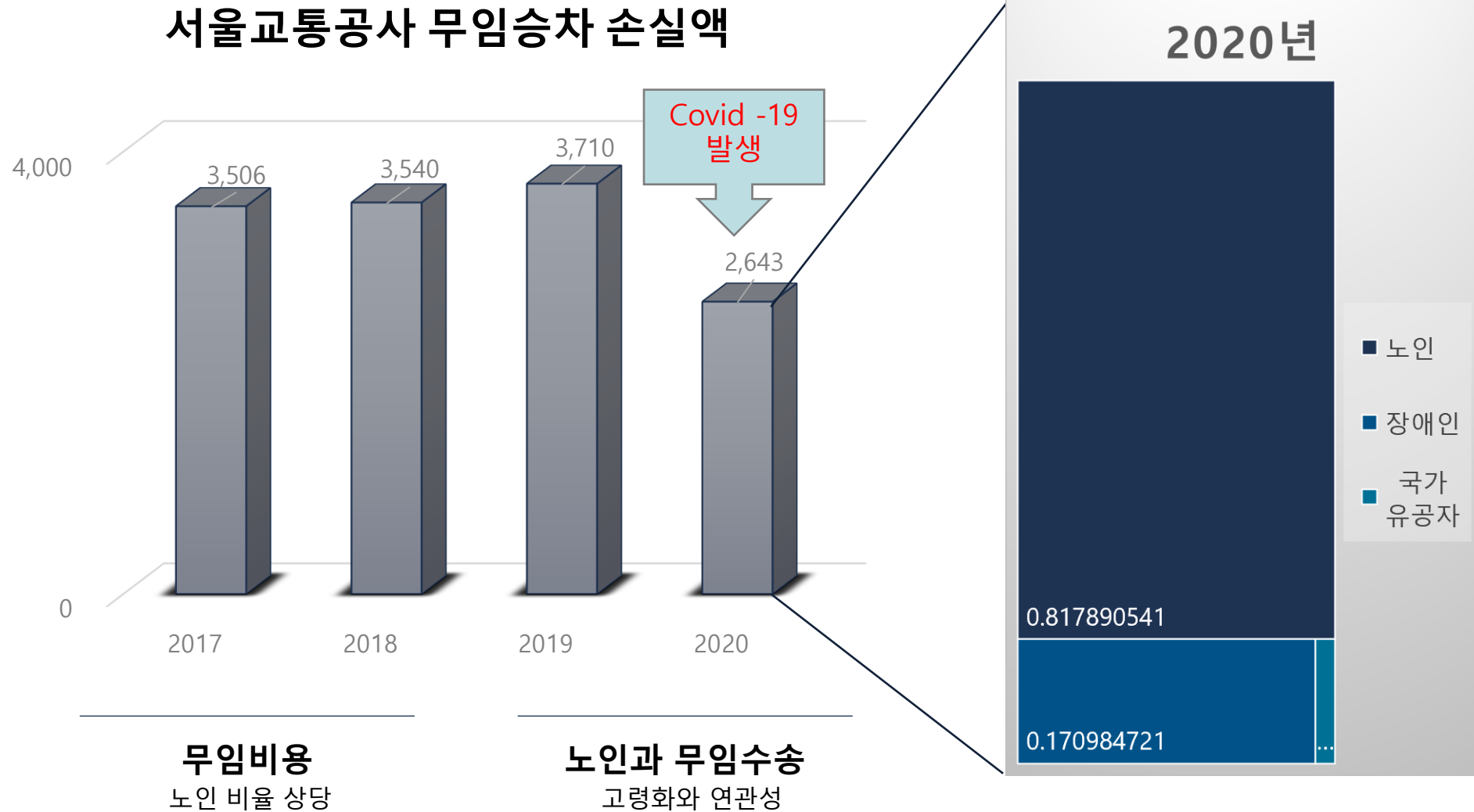
동결된 지하철 요금

7년째 제자리

재정난의 원인

무임손실과 함께

주제 선정 배경



분석 방향 및 목표

현 상황에 대한 각 입장들

빼앗길 위기 속 복지 혜택

고령화 사회 현실만으로 기존의 무상 교통 복지 혜택을 누리지 못할 수도 있는 상황

노인층

세 입장 모두 아우를 수 있는
**지속가능한 교통복지
정책 연구 필요**

공익서비스 제공

무임승차 제도는 우수한
교통복지 정책으로
중단할 수 없으며 근본적인
해결방안만이 필요

서울시
(현 정부)

서울
교통공사

매 순간
불어나고 있는 적자

만성적인 재정난과 운송 비용 부담으로
무임혜택 감소 정책 불가피

Part 2,

본론



데이터셋 생성

다양한 종류의 변수들을 지정

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
호선	역이름	무임승하차	무임승하차	유임승하차	출퇴근시간	출퇴근시간	역간거리(지	지역	행정동	노인인구수	노인인구수	평균연령	인구수(사	인구밀도	지역별통근인원비율(12세 이상)	지역내총생	자치구별	석유사
1호선	동대문	29.74611	256357.7	605461.6	201212.8	668491.1	0.6	종로구	창신동	4.927011	1432.333	48.7	1	6,431	0.624675168	#####	70.2	1351
1호선	동묘앞	46.74245	285406.9	325187.8	139446.5	459489.5	0.7	종로구	송인동	6.120118	1,690	44.7	1	6,431	0.624675168	#####	70.2	1351
1호선	서울역	13.29169	392607.2	2561172	965310.6	2019218	1.1	중구	회현동	3.862563	1,397	51.2	0	13,231	0.615770266	#####	73.1	71269
1호선	시청	13.3314	183237.7	1191244	522431.5	843980.1	1	중구	소곡동	11.25574	305	41.4	0	13,231	0.615770266	#####	73.1	71269
1호선	신설동	25.8447	225766.4	647783.7	301949.8	576345.1	0.9	동대문구	용신동	5.795593	5,900	44.4	3	24,762	0.627476874	7,180,082	80.7	56891
1호선	제기동	54.23527	609370.1	514197.7	269293.5	847797.5	1	동대문구	제기동	4.550992	5,844	47	3	24,762	0.627476874	7,180,082	80.7	56891
1호선	종각	11.97508	277174	2037416	833599.7	1506019	0.8	종로구	종로1.2.3.4	4.413413	1,998	50.9	1	6,431	0.624675168	#####	70.2	1351
1호선	종로3가	36.01087	643358.1	1143208	431128.8	1373562	0.9	종로구	종로1.2.3.4	4.413413	1,998	50.9	1	6,431	0.624675168	#####	70.2	1351
1호선	종로5가	34.52338	500271	948807.9	429670.4	1025502	0.8	종로구	종로5.6가	5.193732	1,053	46.8	1	6,431	0.624675168	#####	70.2	1351
1호선	청량리(서	39.50522	642270.9	983516.7	423827.8	1197891	0.8	동대문구	청량리동	3.613409	5,295	50.7	3	24,762	0.627476874	7,180,082	80.7	56891
2호선	강남	5.64266	308805.3	5163886	1629976	3858467	0.8	강남구	역삼1동	9.729279	3,716	41	3	13,616	0.660353369	#####	45.2	1844
2호선	강변(동서	9.98766	269361	2427577	#####	#####	0.9	광진구	구의동	7.004628	3768.5	42.65	2	20,666	0.650722907	5,970,256	76.9	64470
2호선	건대입구	7.213497	183076.3	2354893	723364.9	1821864	1.2	광진구	화양동	11.41461	2,354	37.1	2	20,666	0.650722907	5,970,256	76.9	64470
2호선	교대(법원	14.38169	323415.7	1925385	#####	#####	1.2	서초구	서초동	7.100753	3938.5	41.5	2	8,858	0.669376204	#####	45.2	14136
2호선	구로디지털	9.29336	327322.6	3194790	1297110	2233155	1.1	구로구	구로3동	7.038204	3,586	42.5	3	20,933	0.630238761	#####	84.2	8265

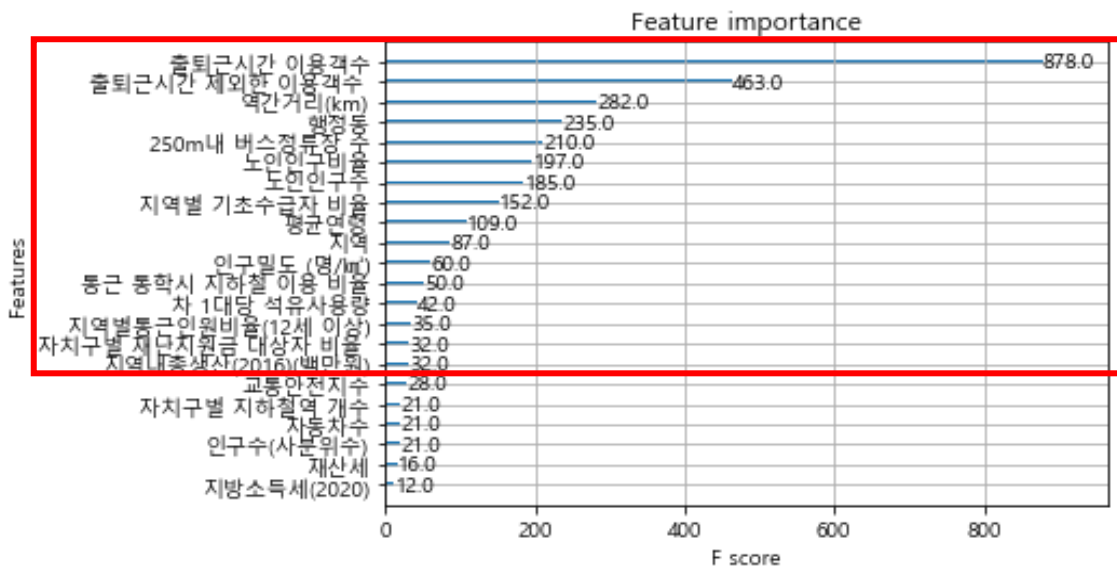
지역별 데이터

기초생활수급자	해당역이 속한 행정동(자치구) 별 기초생활수급자 대상자 수	인구밀도	해당역이 속한 자치구별 명/㎢당 인구밀도
노인인구수(비율)	해당역이 속한 행정동 별 65세 이상 인구수(비율)	지역별통근인 원비율(12세 이상)	해당역이 위치한 자치구별 통근 인원 비율
세금	해당역이 속한 2020년 기준 자치구별 세금 금액	지역내총생산 (2016)(백만원)	해당역이 위치한 2016년 자치구내의 총생산
평균연령	해당역이 속한 자치구 별 평균 연령	자치구별 지하철역 개수	해당역이 위치한 자치구별 지하철역 개수
자치구별 재난지원금	해당역이 속한 자치구별 재난지원금 대상자 비율	자동차수/석유 사용량(bbl)	해당역이 속한 자치구별 자동차 수/석유사용량(bbl)
교통안전지수	해당역이 속한 자치구별 교통 안전지수	통근 통학시 지 하철 이용 비율	해당역이 속한 자치구별 지하철 이용 비율

지하철 역 데이터

무임승하차 비율	해당역의 유임 승하차 승객 수 대비 무임 승하차 승객 수	역간거리	해당역과 다음역과의 거리
무(유)임승하차 수	2016년 1월부터 2022년 6월까지 해당역의 무(유)임 승하차 평균 승객 수	지역/행정동	해당역이 위치한 자치구 및 행정동
출퇴근시간 (제외한) 승객 수	오전 7~9시, 오후 6~8시 해당 역을 이용한(이용한) 승객 수	250m 내 버스정류장 수	해당역 250m 이내에 위치한 버스정류장 수

XGBoost 변수 중요도 확인

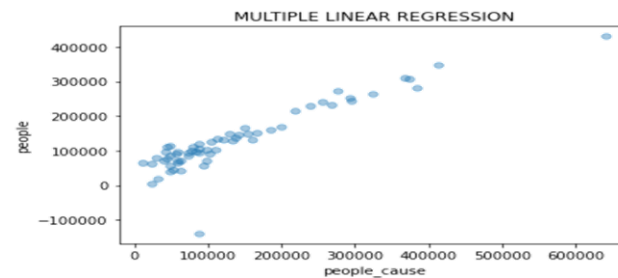


- Gain(분할에 각 변수를 사용할 때마다 감소한 평균 훈련 손실)을 기반으로 XGBoost 변수 중요도 확인
- F-score 상위 16개 기준으로 변수 선정

다중선형회귀분석

	coef	std err	t	P> t	[0.025	0.975]
출퇴근시간 제외한 이용객수	0.2275	0.030	7.674	0.000	0.169	0.286
출퇴근시간 이용객수	0.2645	0.044	5.960	0.000	0.177	0.352
유임승하차	-0.1429	0.030	-4.754	0.000	-0.202	-0.084
무임승하차비율	7065.9736	513.129	13.770	0.000	6056.059	8075.888
지역별통근인원비율(12세 이상)	-3.602e+05	5.07e+04	-7.111	0.000	-4.6e+05	-2.61e+05
평균연령	2572.1647	756.496	3.400	0.001	1083.268	4061.061
역간거리(km)	2.407e+04	6385.751	3.770	0.000	1.15e+04	3.66e+04

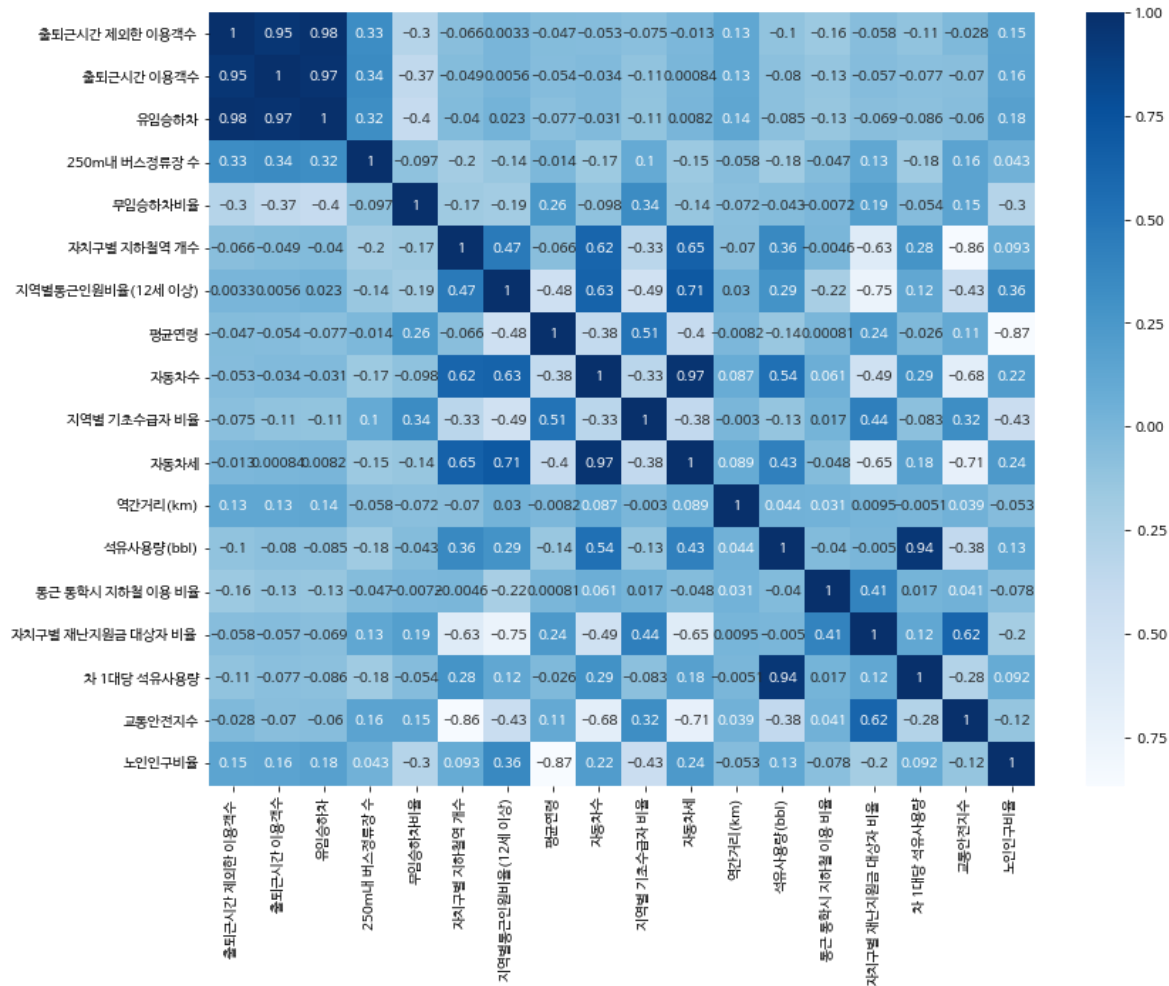
후진제거법 적용 p-value 0.05
이상인 변수들 순차적으로 제거



새 데이터셋 생성

Clustering(군집화)을 위한 중요 변수 추출 2

Coefficient of Correlation(상관계수) 및 다중공선성 확인



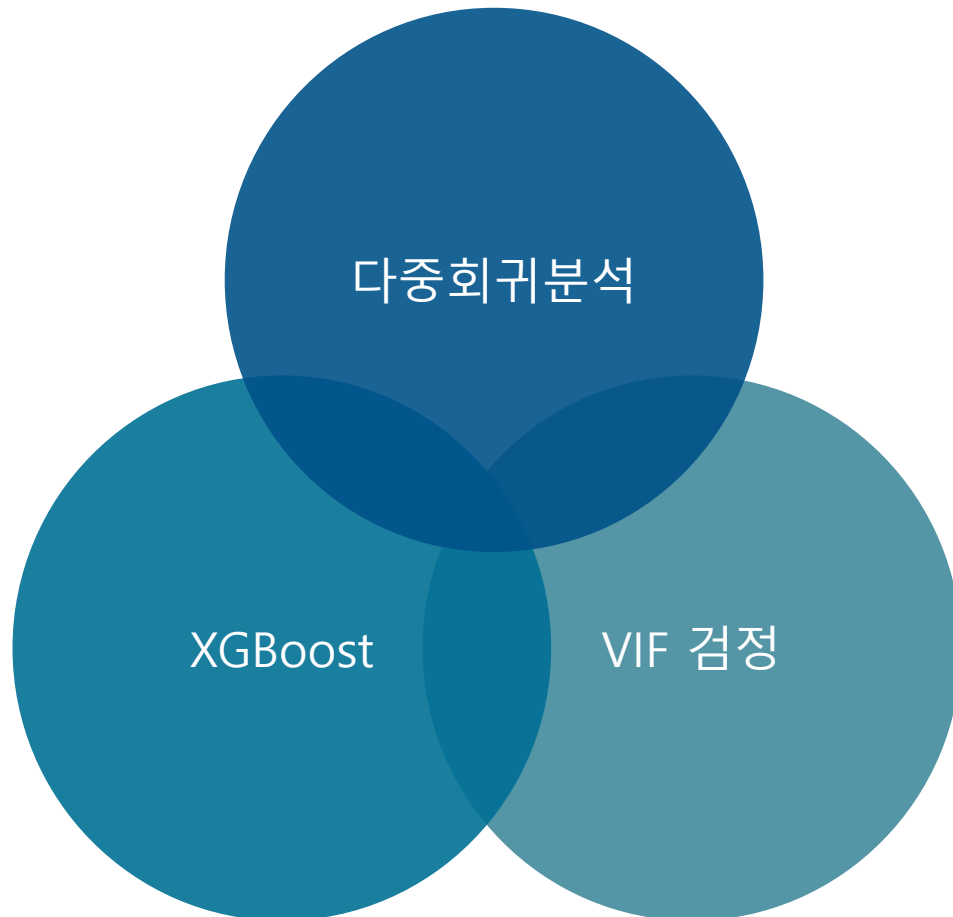
Target (무임승하차) 과의 상관계수가 0.1 이상인 변수들 확인

독립변수 간 높은 상관관계가 있으므로 VIF를 통해 다중공선성 확인
모든 변수의 VIF가 10미만이 되게 함

상관계수와 VIF를 고려한 변수 선정

VIF Factor		features	7	3.907821	자동차수
0	10354.786666	const	8	1.816911	지역별 기초수급자 비율
1	1.511885	출퇴근시간 이용객수	9	1.092384	역간거리(km)
2	1.243792	250m내 버스정류장 수	10	1.744876	통근 통학시 지하철 이용 비율
3	1.385677	무임승하차비율	11	7.783703	자치구별 재난지원금 대상자 비율
4	4.981158	자치구별 지하철역 개수	12	1.989909	차 1대당 석유사용량
5	5.000715	지역별통근인원비율(12세 이상)	13	6.121592	교통안전지수
6	7.388504	평균연령	14	5.563046	노인인구비율

최종변수 선정



2020 세금평균(자동차,재산,지방소득)

노인인구수

무임승하차

무임승하차 비율

유임승하차

출퇴근시간 이용객수

노인인구비율

지역별 기초수급자 비율

자치구별 재난지원금 대상자 비율

250m내 버스정류장 수

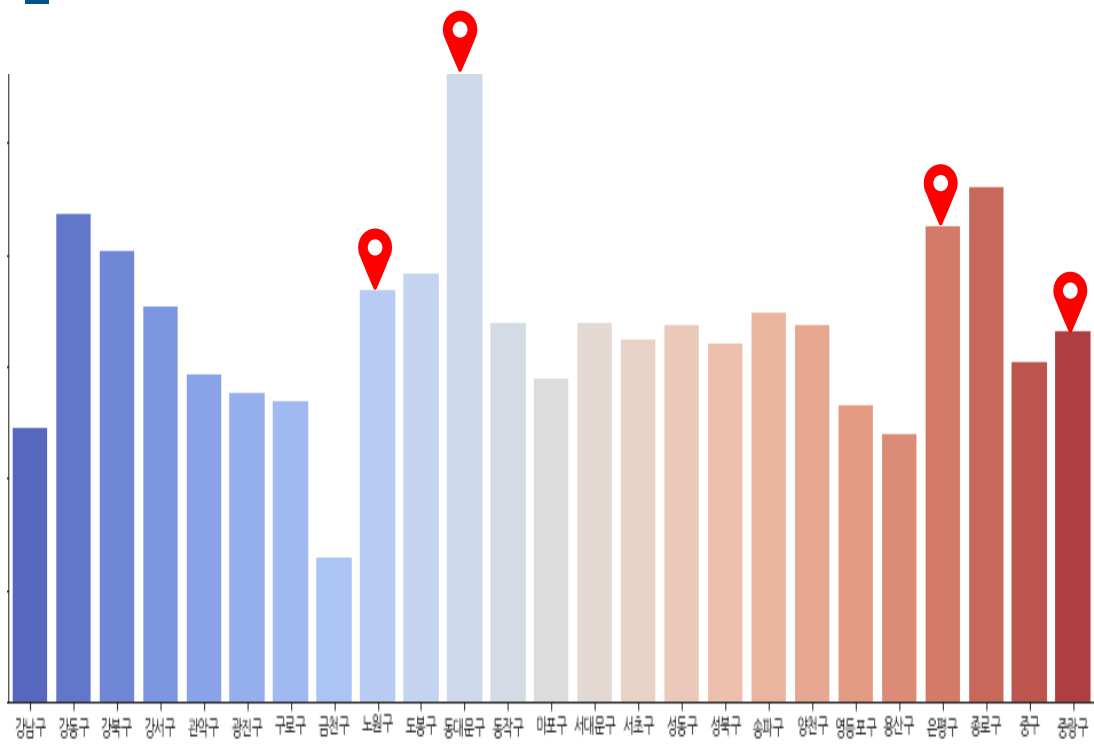
인구밀도 (명/km²)

평균연령

데이터셋 시각화

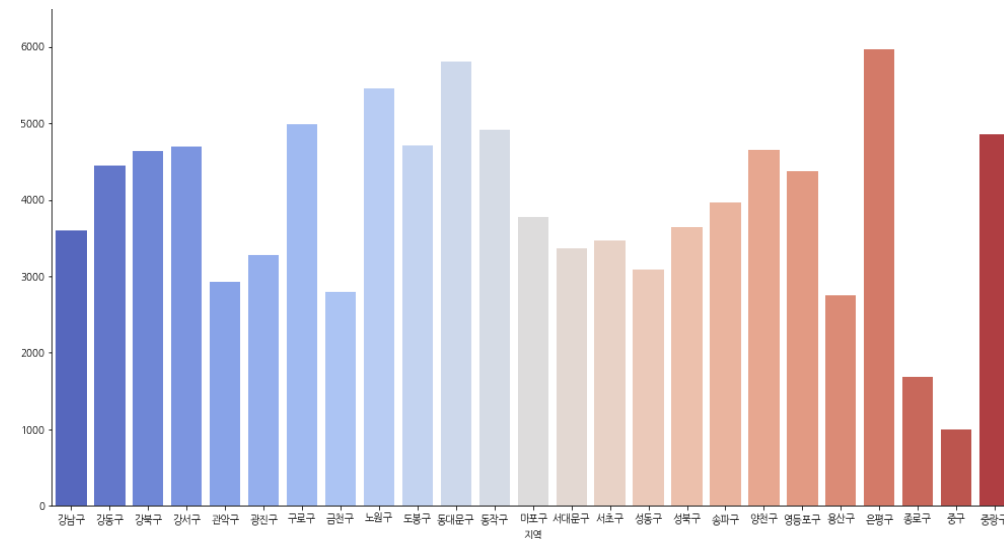
다각적인 시야로 데이터 살펴보기

지역별 특성파악

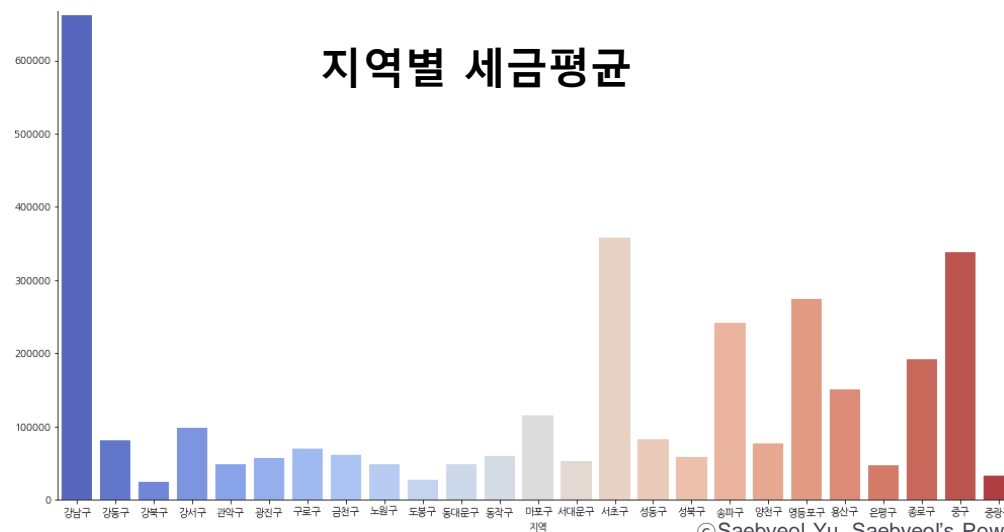


> > >

지역별 노인인구 수



지역별 세금평균



지역별 무임승하차비율

가나다 순으로 배열한
서울시 자치구별 무임승하차 비율 시각화

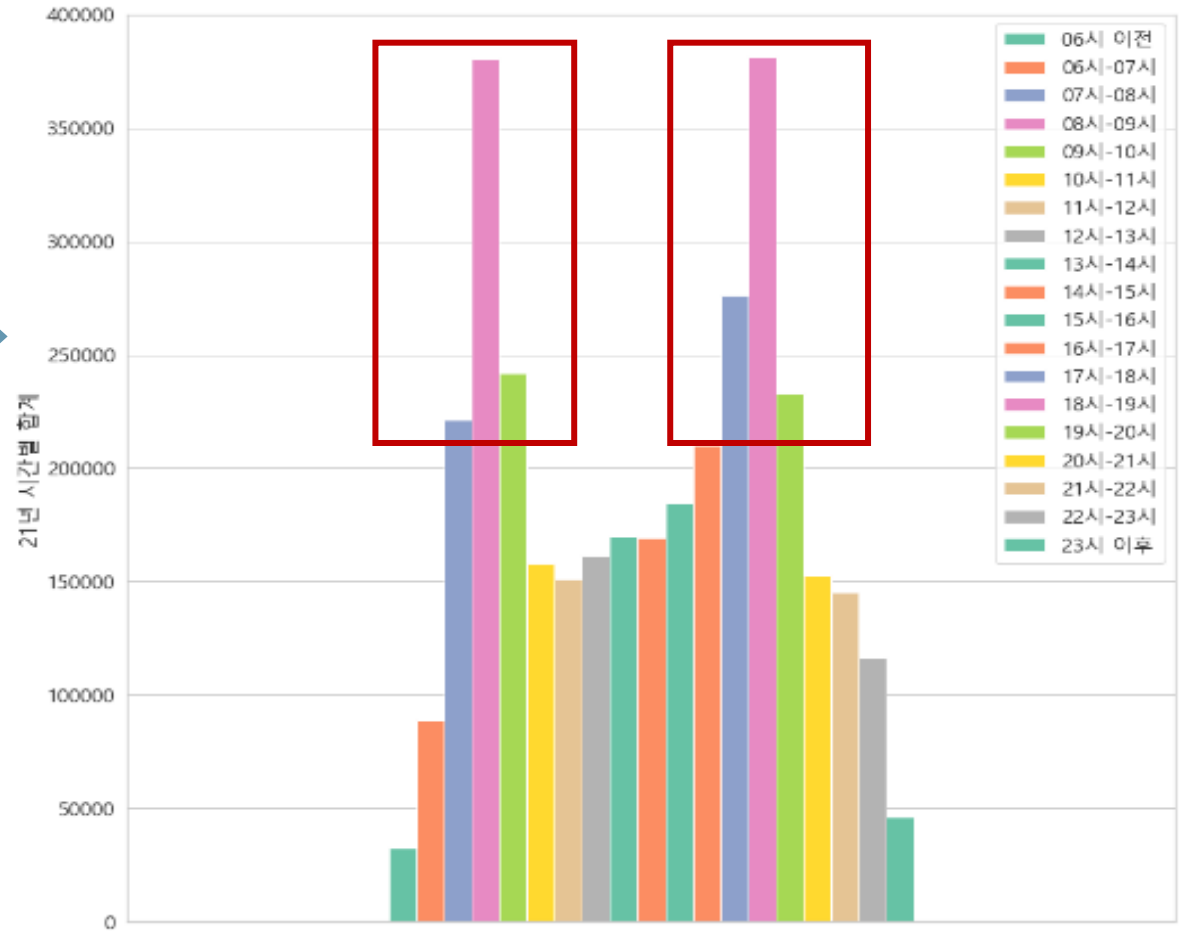
Part 2, 데이터셋 시각화

시간에 따라 변화하는 데이터

시간대별로 살펴본 데이터

	연번	일자	역번호	역명	구분	06시 이전	06시-07시	07시-08시	08시-09시	09시-10시	...	14시-15시	15시-16시	16시-17시	17시-18시	18시-19시	19시-20시	20시-21시	21시-22시	22시-23시	23시 이후	
	0	1	2021-07-01	150	서울역	승차	77	71.0	134.0	214.0	213.0	...	439.0	499.0	441.0	419.0	283.0	196.0	119.0	129.0	52.0	21
	1	2	2021-07-01	150	서울역	하차	65	235.0	249.0	265.0	361.0	...	316.0	365.0	370.0	240.0	184.0	128.0	87.0	63.0	37.0	16
	2	3	2021-07-01	151	시청	승차	23	28.0	23.0	35.0	47.0	...	195.0	296.0	230.0	193.0	136.0	74.0	65.0	42.0	23.0	8
	3	4	2021-07-01	151	시청	하차	48	86.0	115.0	155.0	181.0	...	131.0	105.0	77.0	67.0	53.0	18.0	18.0	12.0	4.0	4
	4	5	2021-07-01	152	종각	승차	28	27.0	40.0	45.0	96.0	...	397.0	513.0	428.0	294.0	175.0	174.0	151.0	136.0	67.0	

82183	82184	2022-05-31	2826	수진	합계 : 하차	11	41.0	37.0	72.0	77.0	...	132.0	140.0	138.0	93.0	61.0	42.0	36.0	18.0	15.0	10	
82184	82185	2022-05-31	2827	모란	합계 : 승차	31	37.0	46.0	61.0	78.0	...	151.0	143.0	142.0	101.0	74.0	56.0	37.0	23.0	12.0	4	
82185	82186	2022-05-31	2827	모란	합계 : 하차	14	35.0	42.0	65.0	68.0	...	85.0	86.0	90.0	76.0	66.0	25.0	19.0	8.0	13.0	3	
82186	82187	2022-05-31	2828	남위례	합계 : 승차	17	30.0	42.0	90.0	73.0	...	67.0	59.0	64.0	41.0	27.0	19.0	16.0	22.0	4.0	4	
82187	82188	2022-05-31	2828	남위례	합계 : 하차	9	44.0	46.0	39.0	44.0	...	62.0	68.0	94.0	82.0	74.0	48.0	23.0	26.0	18.0	19	



2021년도 지하철 승하차 수 데이터
시간대별로 분류하여 시각화

탐색적 데이터 분석 후 ,
인사이트 도출을 향한 과정에서

Clustering 선택 !

K- means Clustering

주어진 데이터를 k개의 클러스터로 묶는 알고리즘

STEP 1. 새로구성한데이터셋으로 군집화

	2020 세 금평균(자 동차, 재 산,지방소 득)	노인인구수	무임승하차	무임승하차 비율	유임승하차	출퇴근시간 이 용객수	노인인구 비율	지역별 기 초수급자 비율	자치 구별 재난 지원 금 대상 자 비율	250m 내 비 스정 류장 수	인구밀도 (명/km ²)	
0	192078.0	1432.333333	256357.72050	29.746110	6.054616e+05	201212.7912	0.204215	0.002775	70.2	5.0	6431.0	4
1	192078.0	1690.000000	285406.89250	46.742445	3.251878e+05	139446.4615	0.179684	0.003191	70.2	8.0	6431.0	4
2	338277.0	1397.000000	392607.18280	13.291690	2.561172e+06	965310.6264	0.258895	0.003666	73.1	10.0	13231.0	5
3	338277.0	305.000000	183237.73120	13.331403	1.191244e+06	522431.5495	0.088844	0.000388	73.1	7.0	13231.0	4
4	48699.0	5900.000000	225766.44080	25.844703	6.477837e+05	301949.8242	0.172545	0.004825	80.7	11.0	24762.0	4

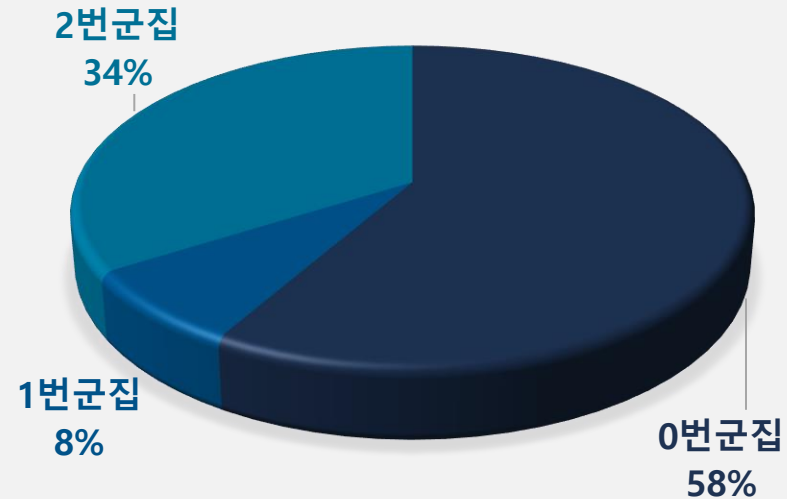
V V

STEP 2. silhouette_score를이용해 군집개수 선정



>
>

STEP 3. 정해진 k개로 군집화 진행



역이름	New
동대문	0
동묘앞	0
서울역	1
시청	2
신설동	0
...	...
종합운동장	0
중앙보훈병원	0
증미	0
한성백제	0
흑석(중앙대입구)	0

298 rows × 1 columns

군집별 시각화

지도 그래프로 역의 위치를 한눈에 파악

서울교통공사 관할 298개의 역 시각화

- 0번 군집
- 1번 군집
- 2번 군집

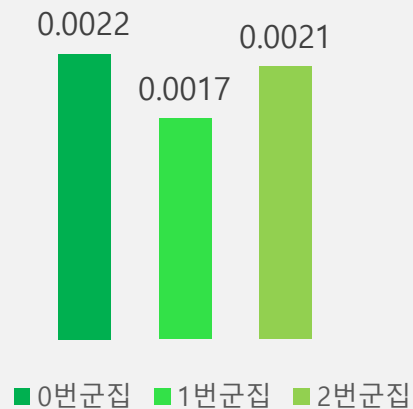


군집별 시각화

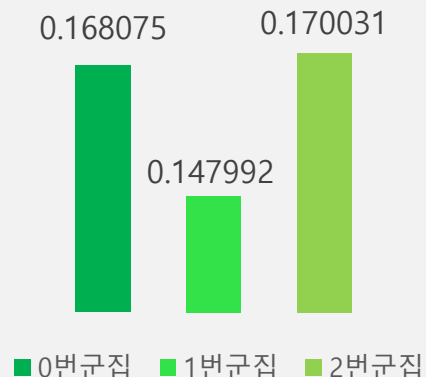
지도에는 군집별 역의 위치, 막대 그래프에는 군집별 각 변수들의 수치 확인



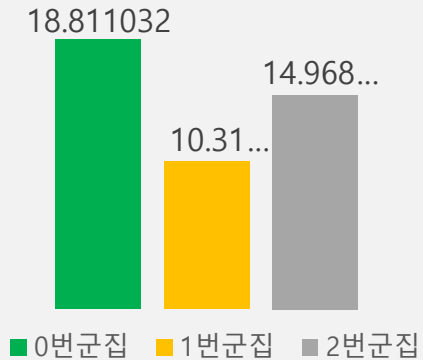
지역별 기초수급자 비율



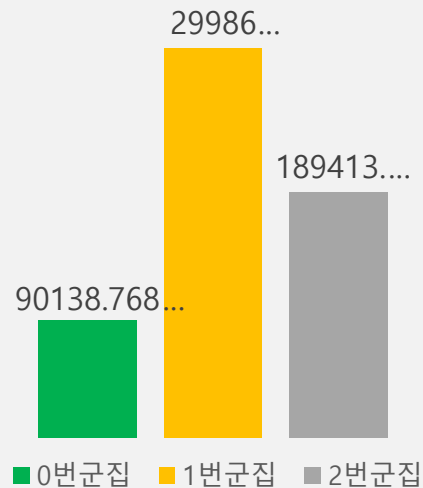
노인인구 비율



무임승하차 비율



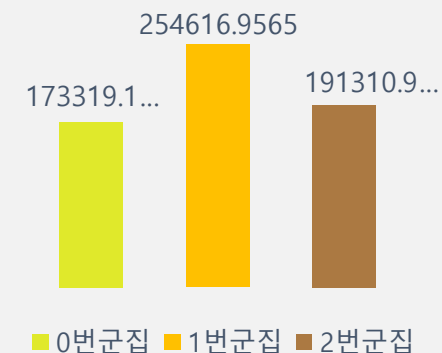
무임승하차 수



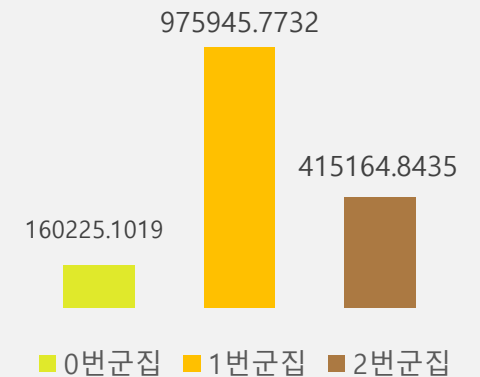
1번군집



2020세금평균
(자동차세, 재산세, 소득세)



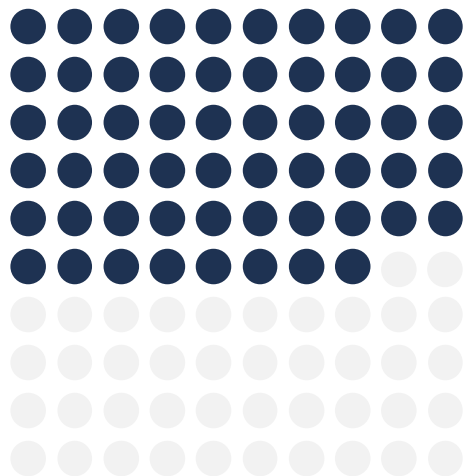
출퇴근 시간 이용객수



군집별 역 특성 확인

군집별 역의 개수, 이름, 위치 등 파악

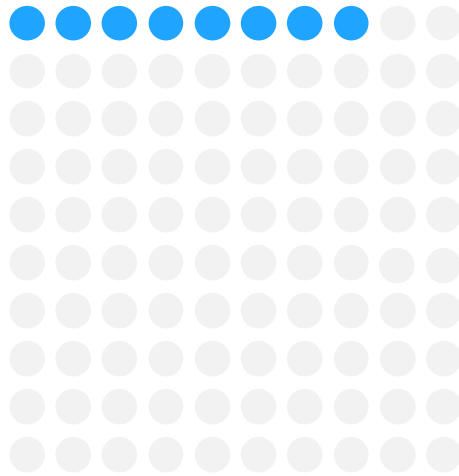
0번 군집



174개의 역

동대문, 동묘앞, 신설동,
제기동, 도림천, 상왕십리,
신답, 신당, 무악재, 약수,
오금, 옥수, 을지로3,4가,
가락시장, 독립문

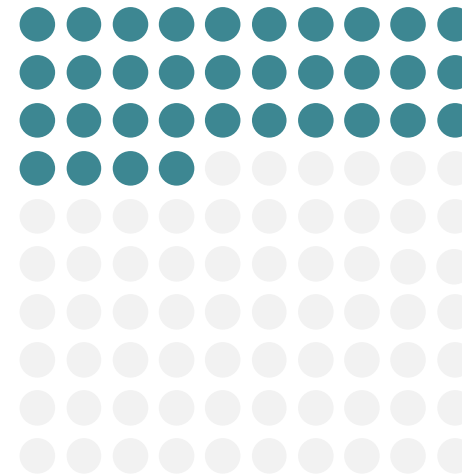
1번 군집



23개의 역

서울역, 종각, 강남, 강변, 건대입구, 교대,
구로디지털단지, 사당, 삼성,
서울대입구, 선릉, 신도림, 신림, 신촌

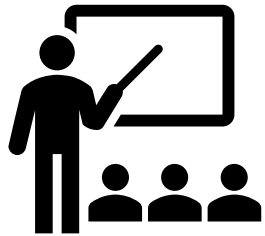
2번 군집



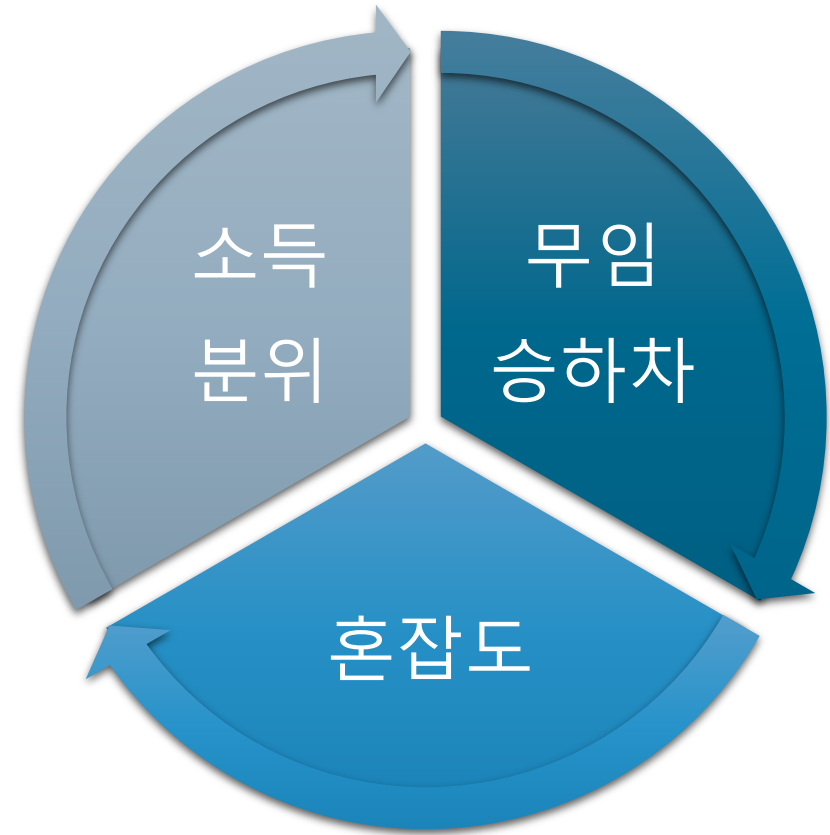
101개의 역

시청, 종로3가, 종로5가,
청량리, 구의, 낙성대, 당산,
동대문역사문화공원, 뚝섬,
문래, 방배, 봉천, 서초,
성수, 시청, 신대방,
영등포구청, 왕십리

효율적인 방안 모색



- 3가지 관점에서 지하철 재정난 문제점 접근
- Clustering으로 만들어진 군집을 활용한 정책 고안
- 각 입장 모두 수용 가능할 만한 정책 고안



Part 3, 결론



Part 3,

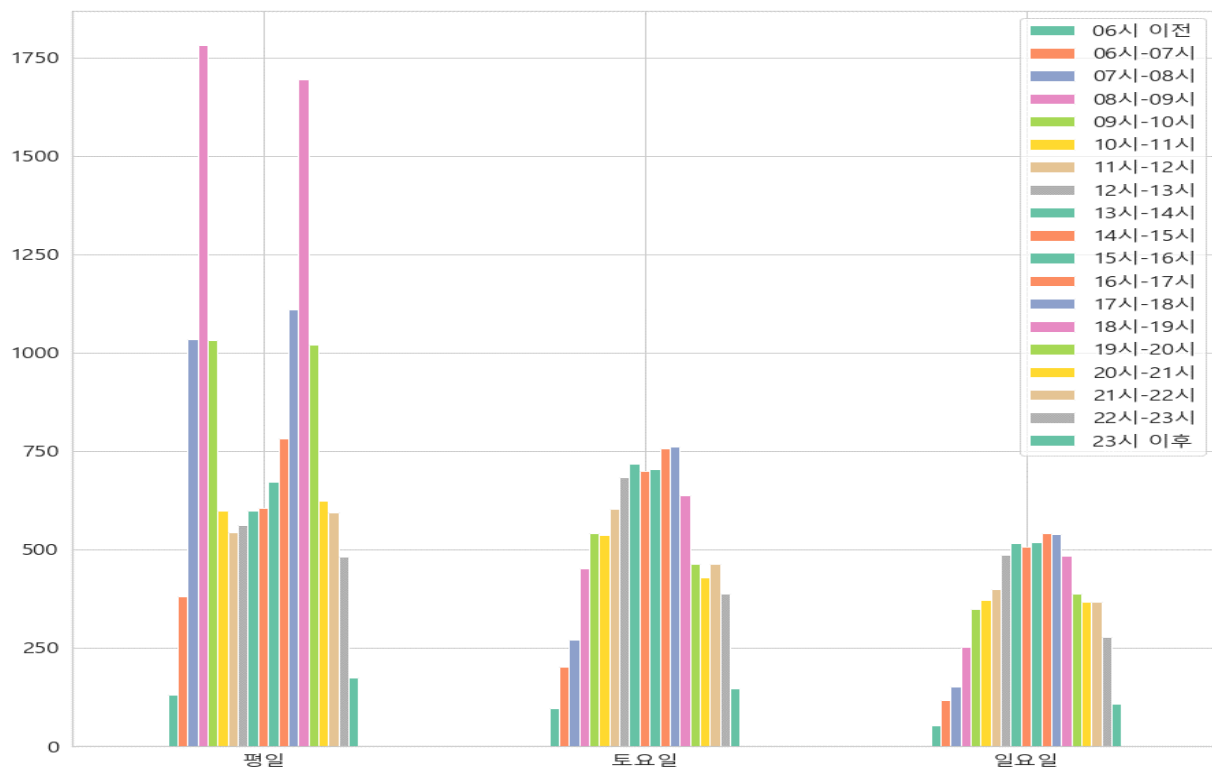
1안. 출퇴근 시간대 무임승차 혜택 제한정책

가격산정 및 기대효과

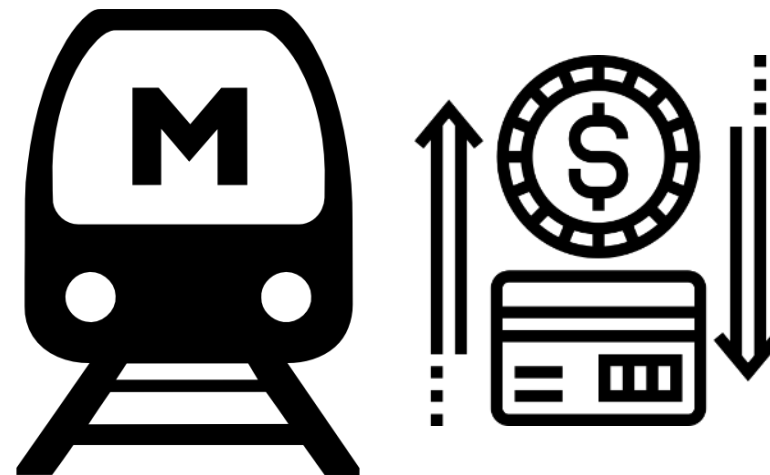
평일 출퇴근 시간대 혼잡도 개선 필요



오전 7~10시 / 오후 5~8시 무임승차 혜택 제거
모든 탑승 인원이 1250원 요금 지불



지하철 혼잡도 시각화 그래프



21년 7월 ~ 22년 5월 노인 승하차수: 324,372,125명
21년도 7월 ~ 22년 5월 평일 출퇴근시간대 노인 승하차수: 82,110,650명

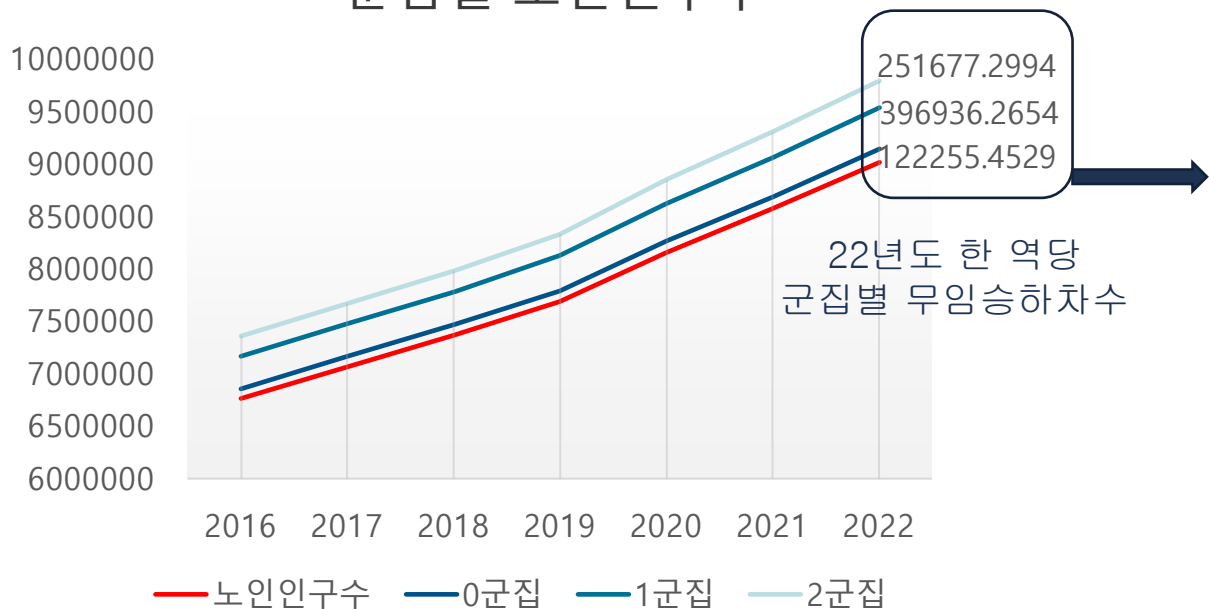
25%의 승객이 1250원의 돈을 내고 타게 됨

2안. 군집 당 소득분위별 가격차등화 정책

<2020년 5분위별 소득 데이터>

분위수	예측 금액	개인 소득
제1분위	180.80원	493.7
제2분위	293.52원	801.5
제3분위	466.01원	1,272.50
제4분위	662.04원	1,807.80
제5분위	1250원	3413.30

군집별 노인인구수



가격산정 및 기대효과

	0군집	1군집	2군집
1분위	90원	280원	180원
2분위	140원	450원	290원
3분위	220원	720원	460원
4분위	320원	1020원	650원
5분위	390원	1250원	790원
무임비용비율	1	3.24	2.06
역 개수	174개	23개	101개
기대효과	0.1856%	0.5952%	0.3792%

총 **34%** 개선 가능

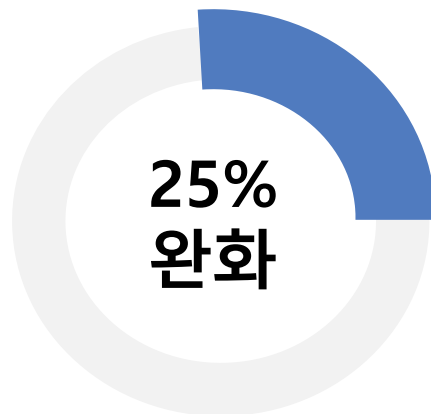
1,2안 혼합 정책

종합적 기대효과

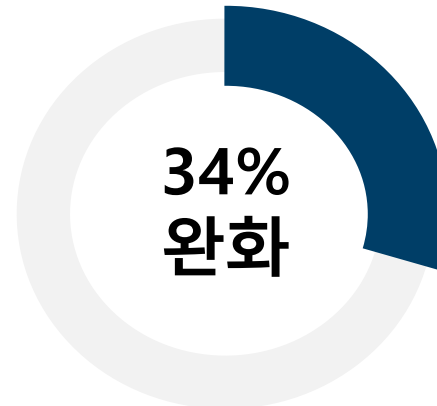
★ 한 역에서 한 분위에 아래와 같은 인원수가 탑승한다고 가정

	0군집	1군집	2군집	전체
무임비용	13,295,280,503 원	5,705,958,815 원	15,887,129,525 원	34,888,368,843 원
복구액	2,467,604,061 원	3,396,186,687 원	6,024,399,516 원	11,888,190,264 원
복구 비율	0.1856 %	0.5952 %	0.3792 %	0.340749386 %

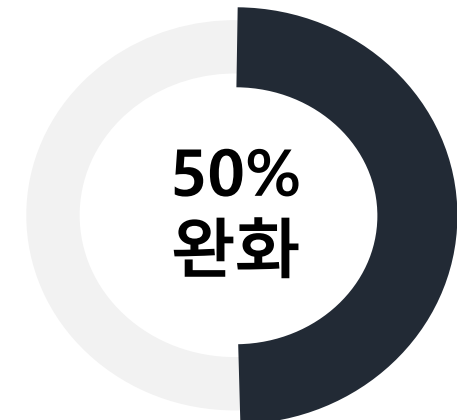
출퇴근 시간대
제한 정책



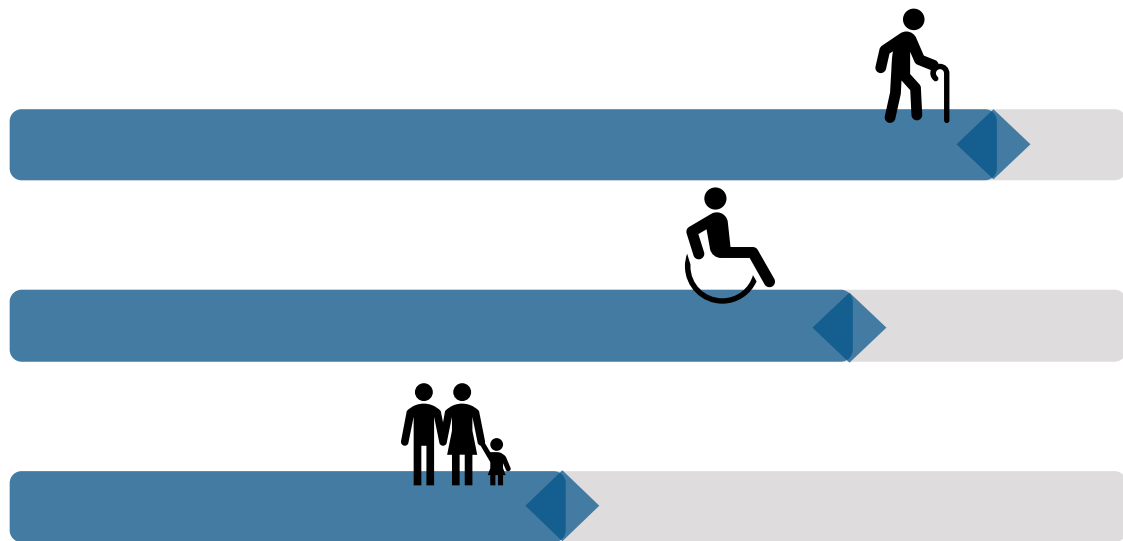
군집별 소득별
차등화 정책



혼합 정책



정책이 가지는 의의



무임수송은 지하철 재정난의 대표적인 원인,
노인의 비율이 가장 높은 것 또한 사실

그러나 노인의 무임혜택만을 타하는 것은 부당

우리의 정책은 **지하철 역, 소득분위, 시간대** 등
다양한 관점에서 바라보며 각 특성을 반영

노인과 지하철, **양측의 입장을 고려한** 정책 제안

참고문헌

데이터 출처

- 국가통계 포털
- 공공데이터 포털
- 서울 열린 데이터 광장

분석도구



PPT테마 출처

- 새별의파워포인트

<https://post.naver.com/viewer/postView.naver?volumeNo=29361216&memberNo=6532296>

참고문헌

[지하철역·노선 많은 동네가 집값도 비싸다? - 매일경제 \(mk.co.kr\)](http://mk.co.kr)

[\[초동시각\] '적자 산더미' 서울지하철 해결책 없나 - 아시아경제 \(asiae.co.kr\)](http://asiae.co.kr)

[무임승차 80%가 노인...'지하철 만성적자' 어떻게 메우나 - 머니투데이 \(mt.co.kr\)](http://mt.co.kr)

['4선' 오세훈 서울시장, 7년째 제자리걸음 지하철 요금 인상하나 - 인사이트코리아 \(insightkorea.co.kr\)](http://insightkorea.co.kr)

[1명 탈 때마다 500원 적자... "지하철 요금 조정 법제화해야" \(hankookilbo.com\)](http://hankookilbo.com)

[전기요금 올려도 올해 철도요금 인상없다...18조 적자는 '물음표' \(news1.kr\)](http://news1.kr)

http://m.naeil.com/m_news_view.php?id_art=396586

[오세훈 "서울교통공 적자, 무임승차 탓...정부가 지원해야"
출처 : https://www.sedaily.com/NewsView/22QAEBFCYQ](http://www.sedaily.com/NewsView/22QAEBFCYQ)

Q&A

“

감사합니다.

”