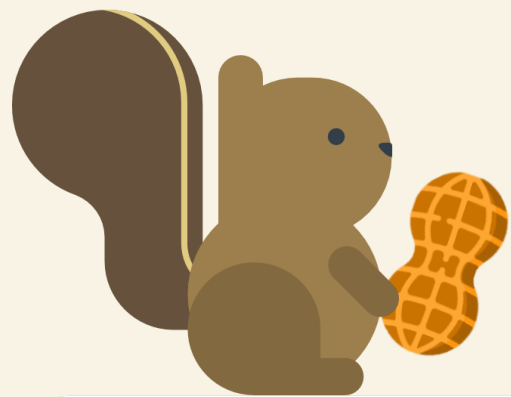




다람쥐버스 신규 노선 제안



박서영

김민주

이윤희

이지연

황정현



Index



주제선정



전처리 및 EDA



행정동 분석



다음주 예고



주제선정

다람쥐버스 소개

아침마다 직장인들의 출근 전쟁...

직장인들의 출근을 돕기 위해 나타난 다람쥐 버스의 정체는?



주제선정



전처리
및 EDA



행정동 분석



다음주 예고

다람쥐버스란?

다람쥐가 쳇바퀴를 돌 듯이 짧은 구간을 반복 운행하는 버스

오전7~9시 버스 혼잡이 극심한 구간에서만 운행

국토교통부 주관 '2018 지속가능 교통도시평가'

교통정책 분야 최우수 정책으로 선정!

다람쥐버스 소개

얼마나 노선이 짧길래 다람쥐람?



버스 번호	다람쥐버스가 해결한 혼잡 원인
8221	답십리동·장안동은 지하철없는 섬 같은 공간(...)
8331	3·5·8호선 이용 시 2회 환승노선을 무환승으로 연결
8441	대규모 개발지임에도 도시철도 없음
8551	동서로 가는 1·2·7호선을 남북으로 연결
8552	대규모 재개발로 인구 늘었으나 도시철도 없음
8761	동서로 가는 6·9호선을 남북으로 연결
8771	남북으로 가는 3·6호선을 동서로 연결

지하철 역과의 연결성 · 통근 종착지인 사업체 밀집지역으로의 연결성 고려

기존 노선 혼잡도 완화로 버스의 수송 분담률 높이는 선순환 효과!



주제선정



전처리
및 EDA



행정동 분석



다음주 예고

주제 선정 배경

이렇게 인기가 높은 다람쥐버스... 나도 타고 싶어! (정말?)



주제선정



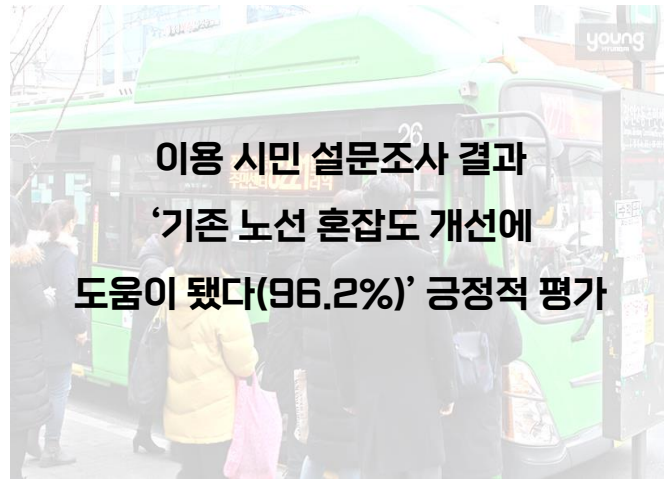
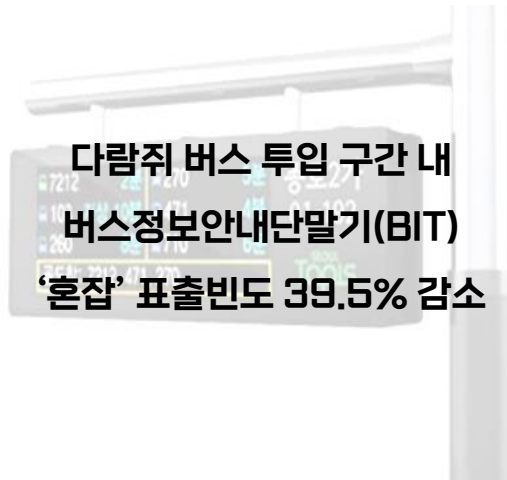
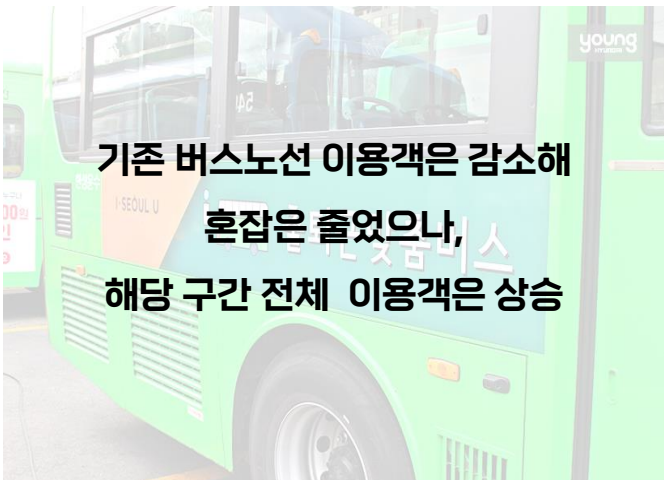
전처리
및 EDA



행정동 분석



다음주 예고



출근 시간대 다람쥐버스의 효과가 크다고 파악



다람쥐버스의 수요가 높을 곳을 선정해 신규 노선을 제안해보자!



분석 계획

2주차

다람쥐버스의 현황 분석

버스 수요와 연관된 변수에 대한 데이터 분석

행정동 단위로 지도 시각화 & 출근시간 버스 수요가 필요한 곳 시각적 이해

가 보자~



4주차

공간회귀 등 다양한 모델링 통해

통계적으로 버스 수요가 필요한 지역 예측

거리 최적화를 통한 노선 제안



선대팀 화이팅



주제선정



전처리
및 EDA



행정동 분석

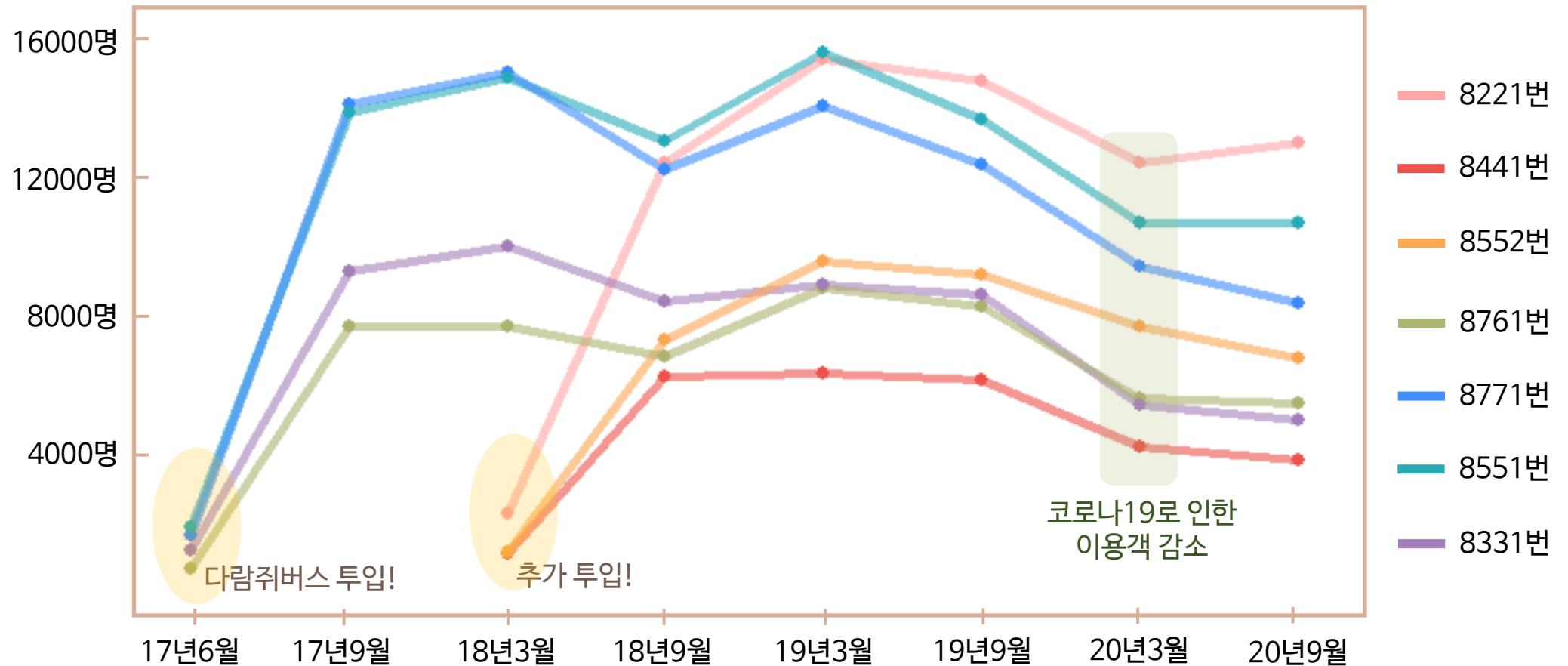


다음주 예고



다람쥐버스 월별 이용현황

노선 도입이후 2020년 9월까지 하차인원 기준



주제선정

전처리
및 EDA

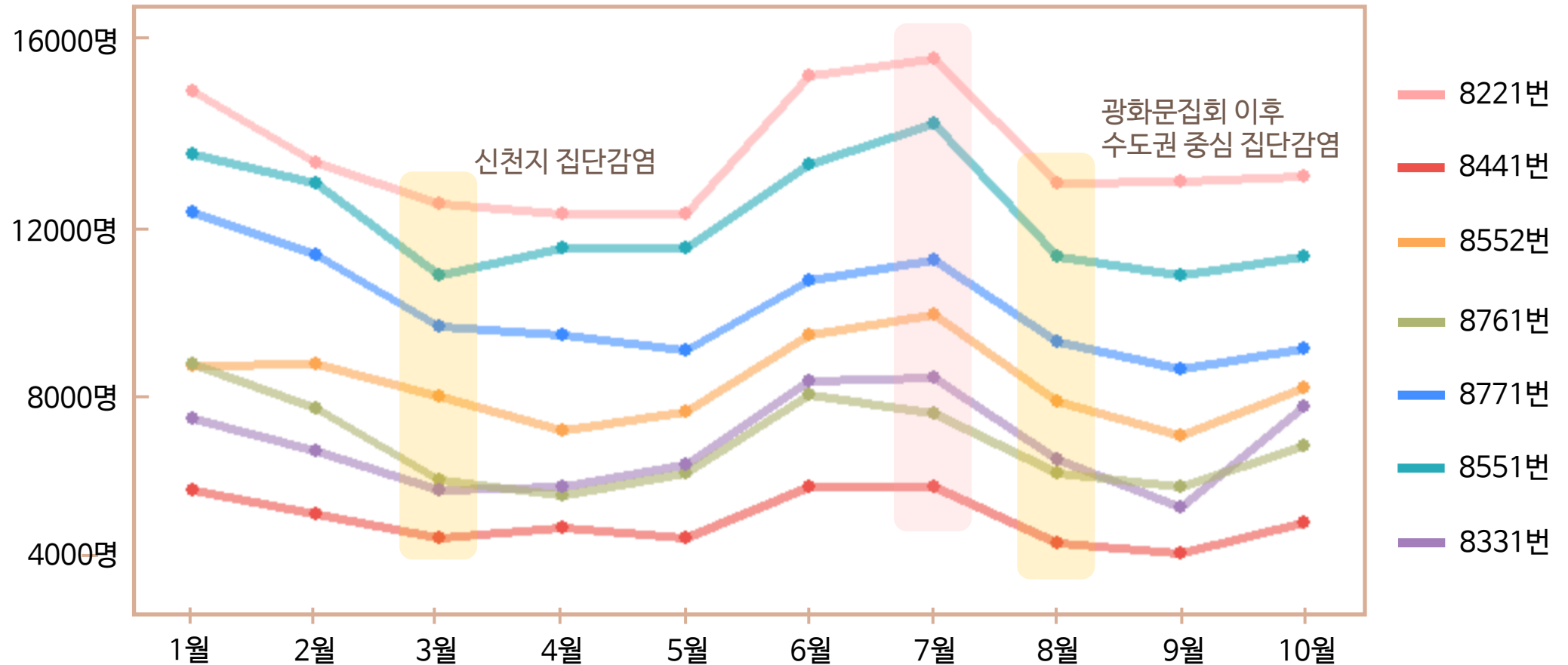
행정동 분석

다음주 예고



다람쥐버스 월별 이용현황

하차인원 기준 코로나 이후 2020년의 상황



주제선정

전처리
및 EDA

행정동 분석

다음주 예고



시각화 준비

전국 행정동 경계 파일(shapefile) : 데이터프레임으로 변환해 서울시만 뽑아 시각화

Z_SOP_BND_ADM_DONG_PG.shp



주제선정



전처리
및 EDA



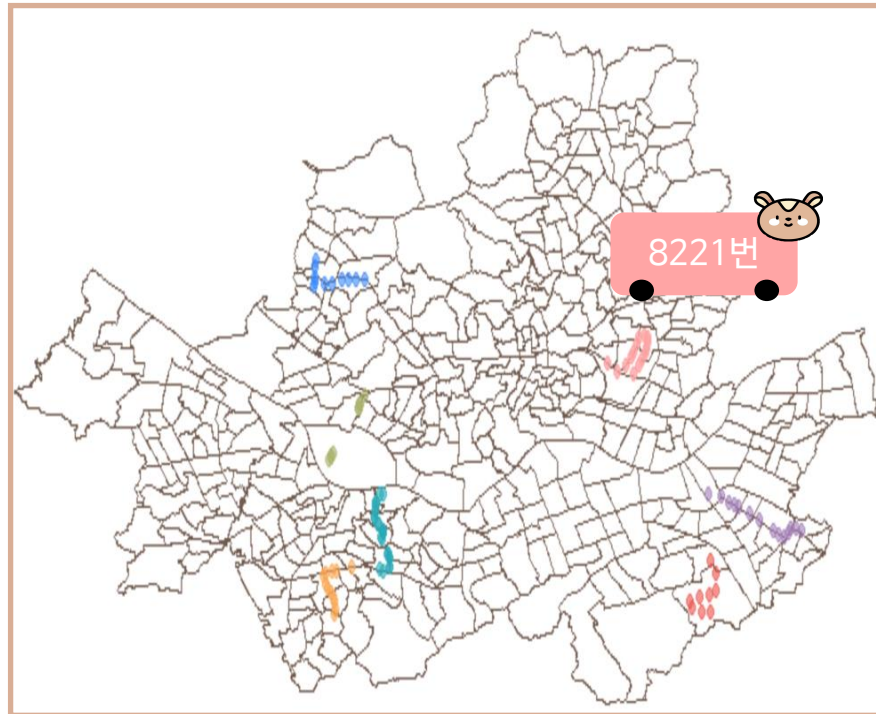
행정동 분석



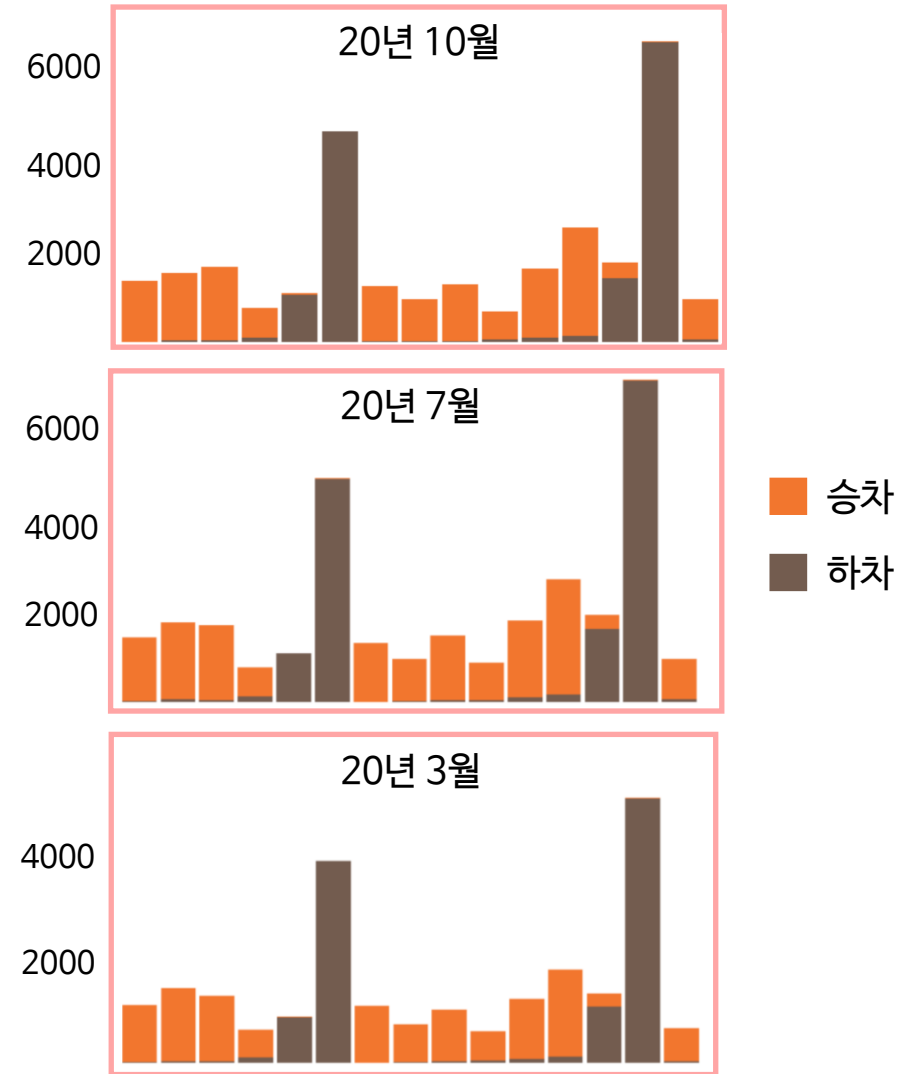
다음주 예고

다람쥐버스 정류장별 이용현황

7개 다람쥐버스 정류장별 승하차 확인



시기별 그래프 모양이 모두 비슷하군!!!



x=정류장, y=이용객

주제선정

전처리
및 EDA

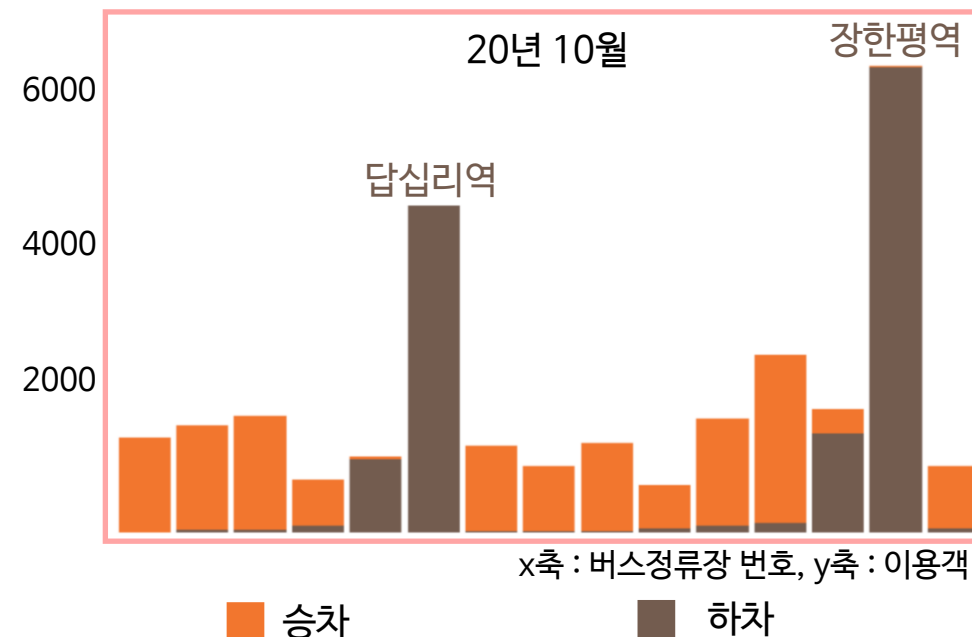
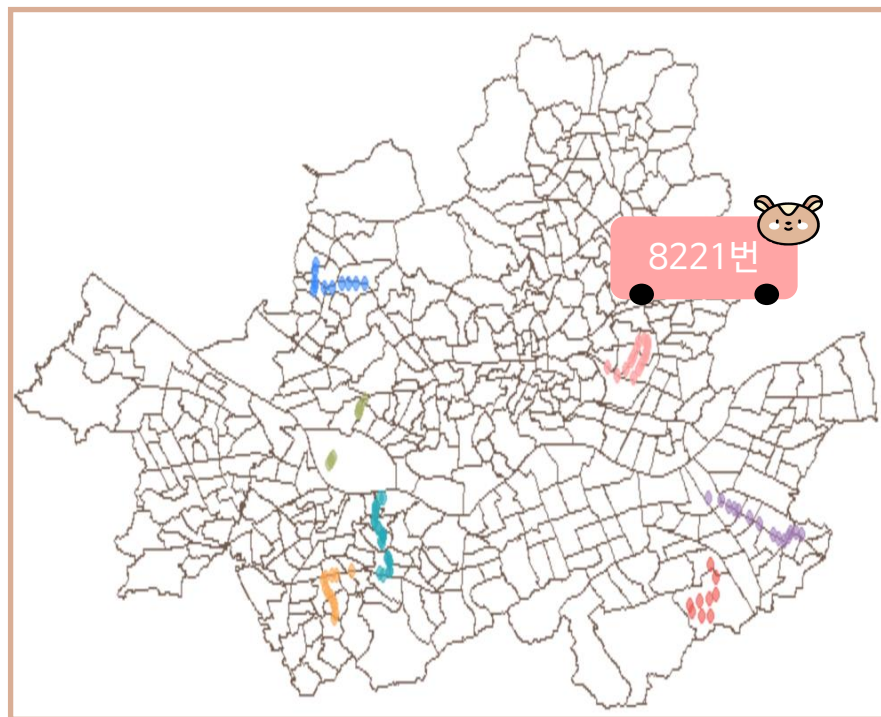
행정동 분석

다음주 예고



다람쥐버스 정류장별 이용현황

7개 다람쥐버스 정류장별 승하차 확인



장한평역, 답십리역 순으로 이용객이 많음

하차인원이 많은 정류장은 모두 지하철 역 근처!



주제선정



전처리
및 EDA



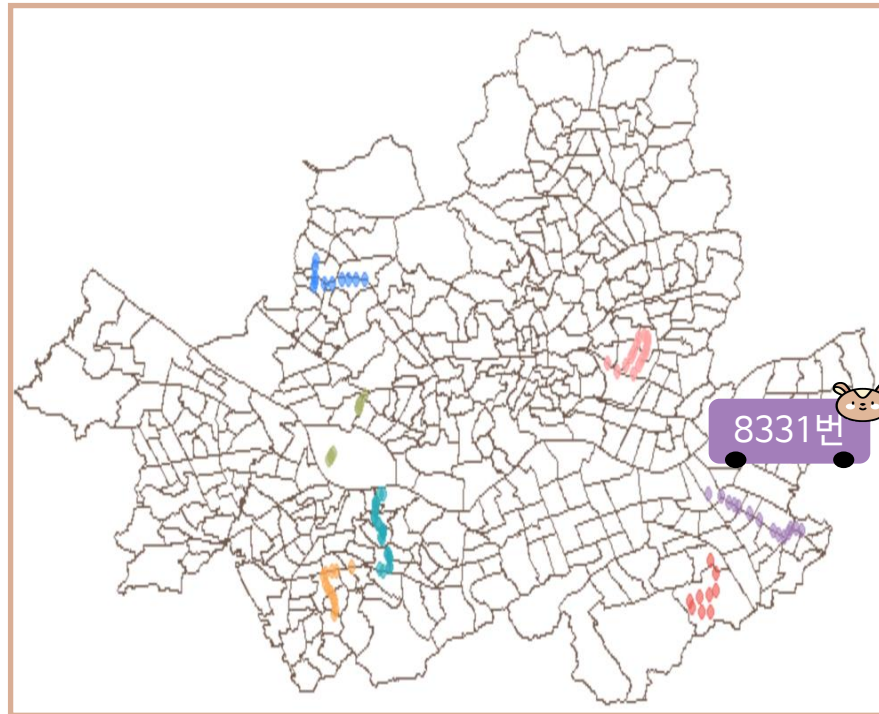
행정동 분석



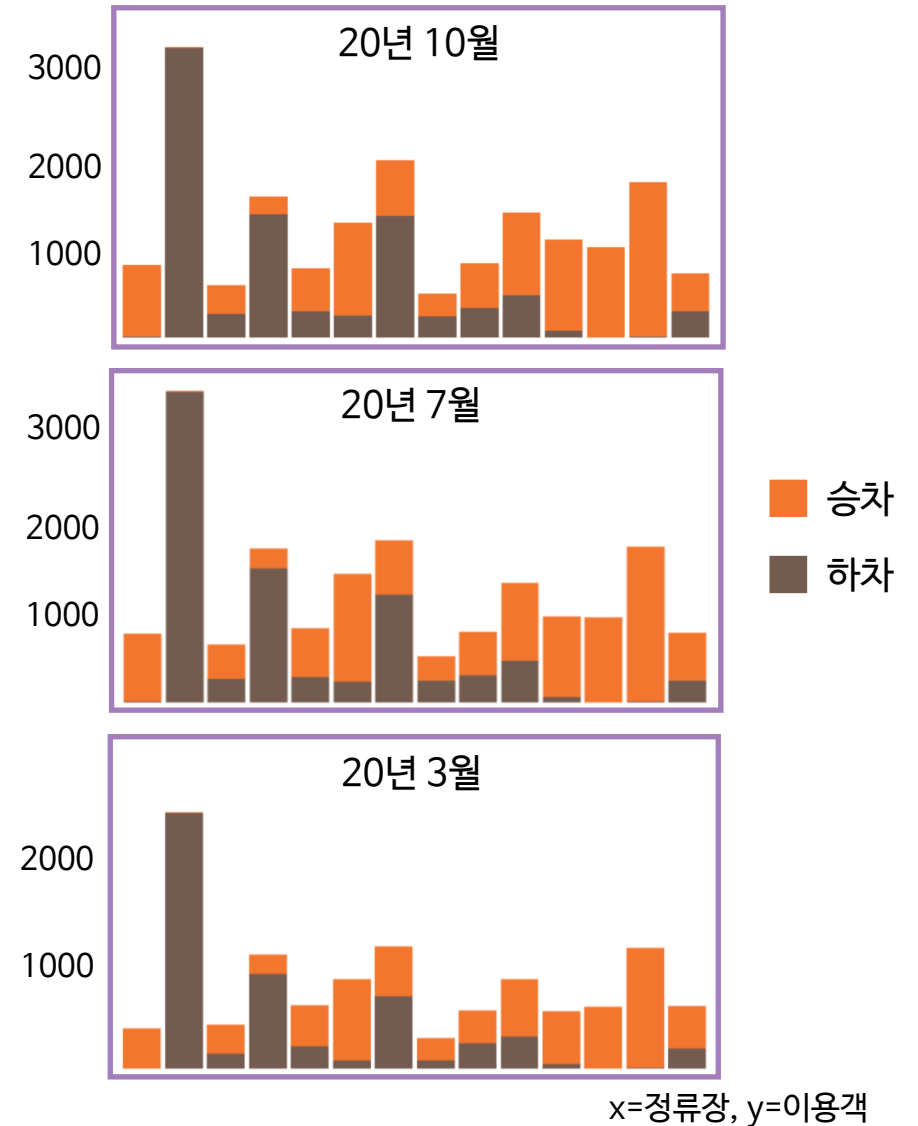
다음주 예고

다람쥐버스 정류장별 이용현황

7개 다람쥐버스 정류장별 승하차 확인



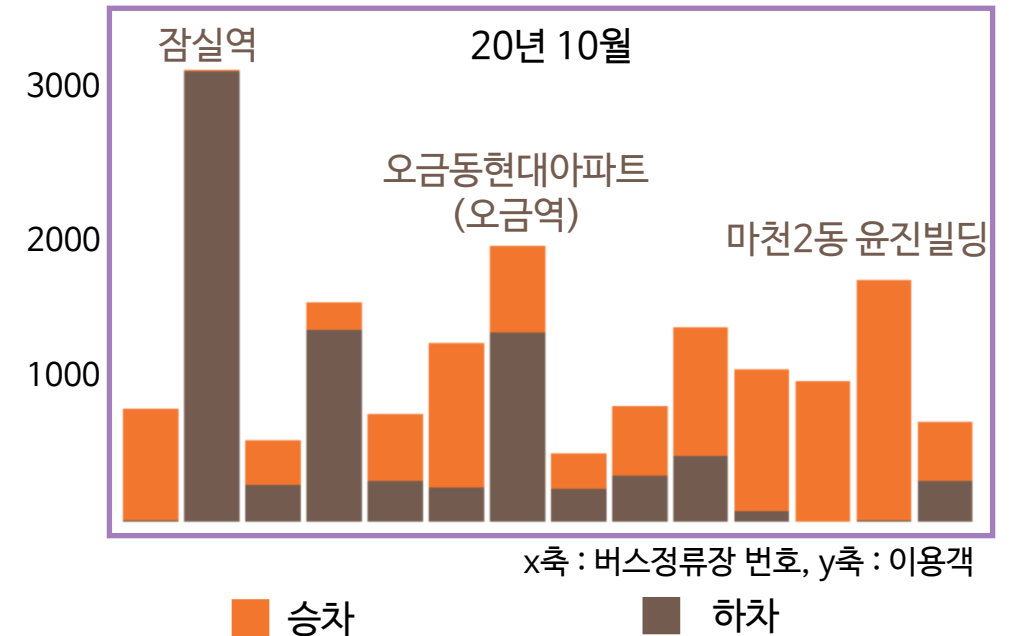
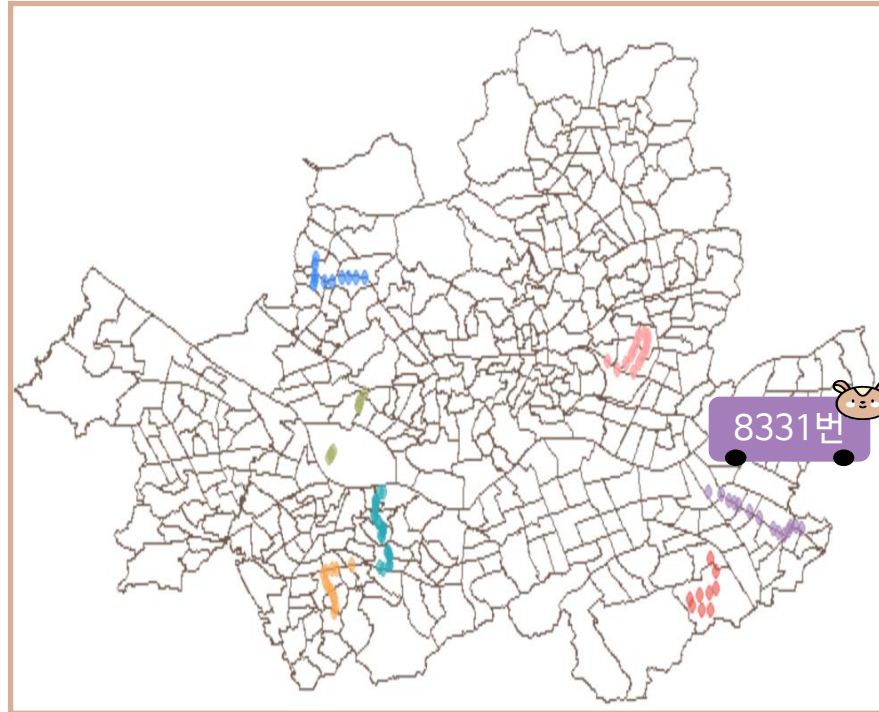
이번에도 그래프 모양이 모두 비슷하군!?!



- 주제선정
- 전처리 및 EDA
- 행정동 분석
- 다음주 예고

다람쥐버스 정류장별 이용현황

7개 다람쥐버스 정류장별 승하차 확인



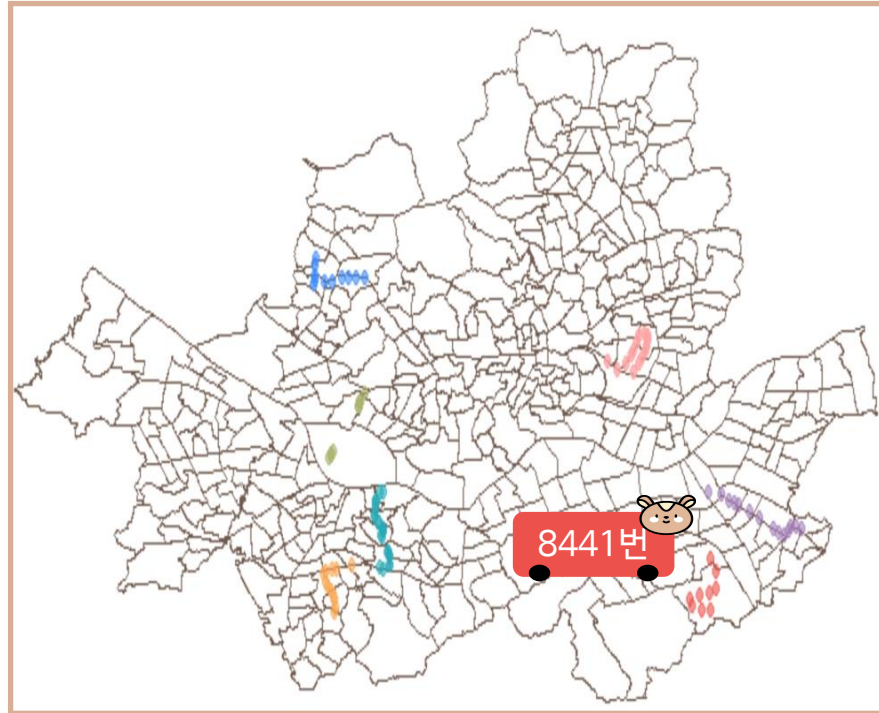
잠실역, 오금동현대아파트, 마천2동윤진빌딩 순으로 이용객 많음

지하철 역 근처의 정류장에서 하차가 많고, 주택가 근처 정류장에서 승차가 많은 편!

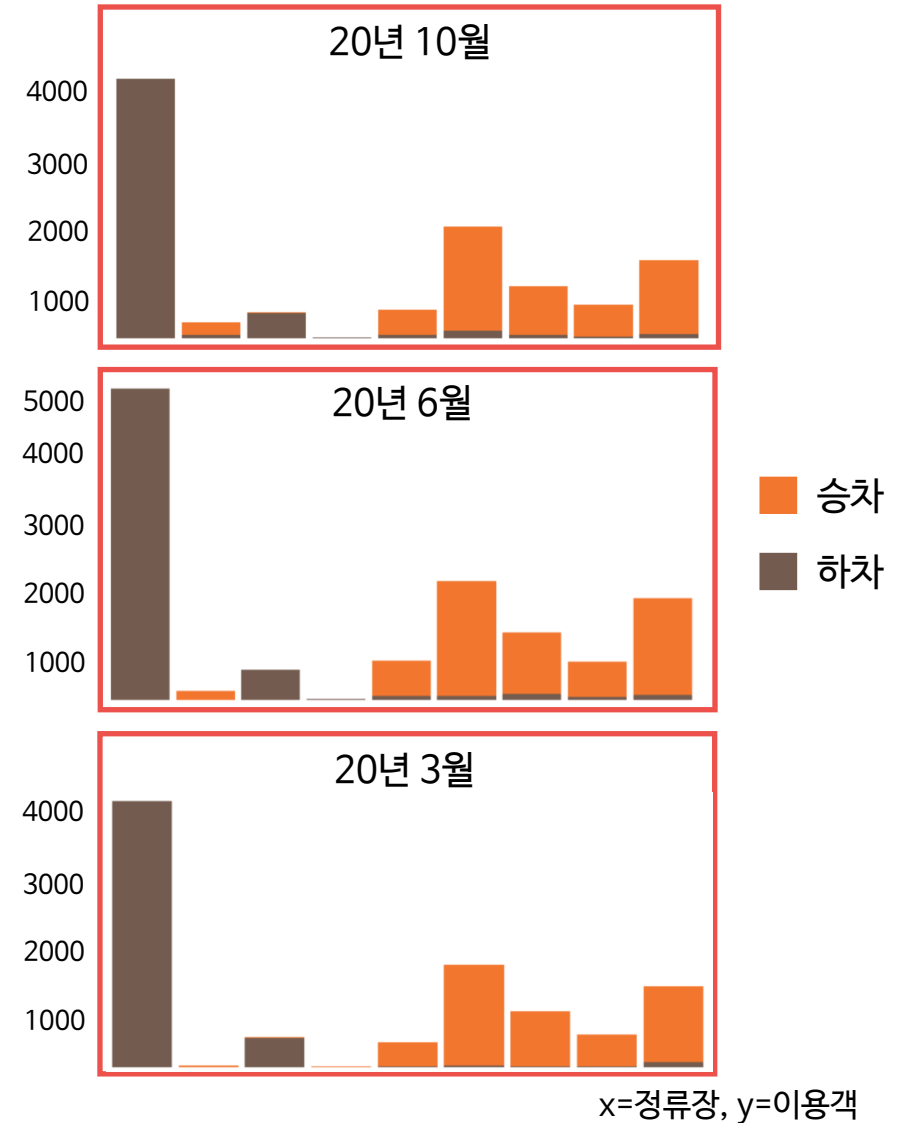
- 주제선정
- 전처리 및 EDA
- 행정동 분석
- 다음주 예고

다람쥐버스 정류장별 이용현황

7개 다람쥐버스 정류장별 승하차 확인



앞으로는 10월 자료 하나만 보자...



주제선정



전처리
및 EDA



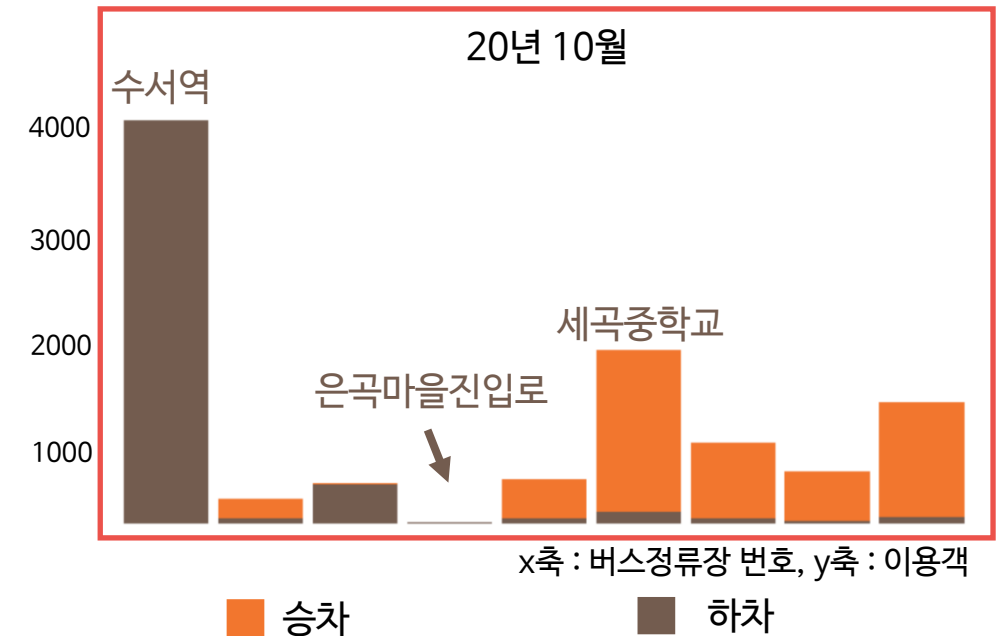
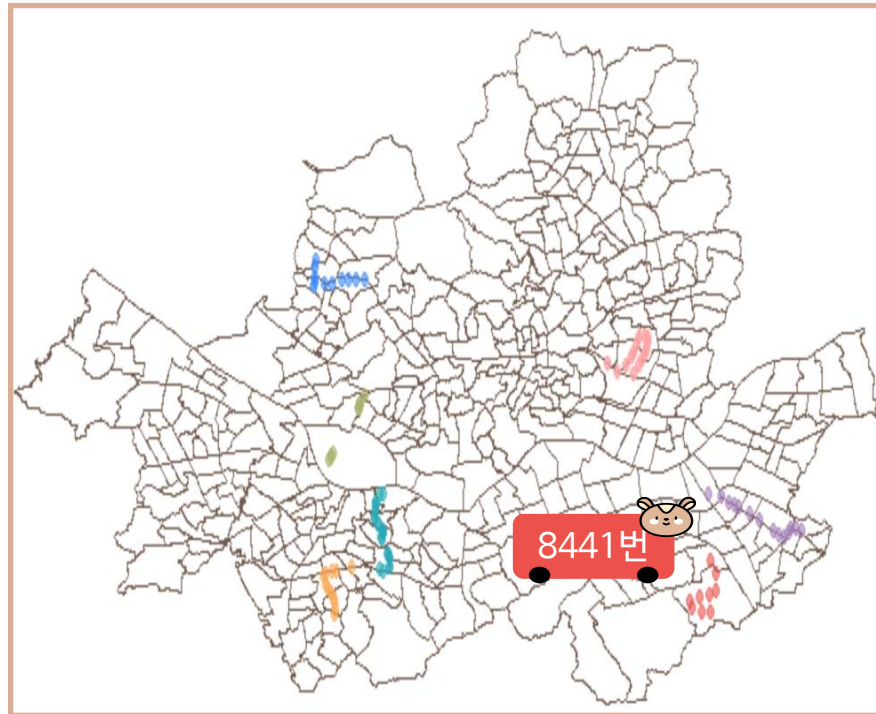
행정동 분석



다음주 예고

다람쥐버스 정류장별 이용현황

7개 다람쥐버스 정류장별 승하차 확인



이용객과 하차 모두 **수서역**이 가장 많았음

이외의 정류장은 이용객이 적은 편 (**은곡마을진입로** 특히 적음)

주제선정

전처리
및 EDA

행정동 분석

다음주 예고



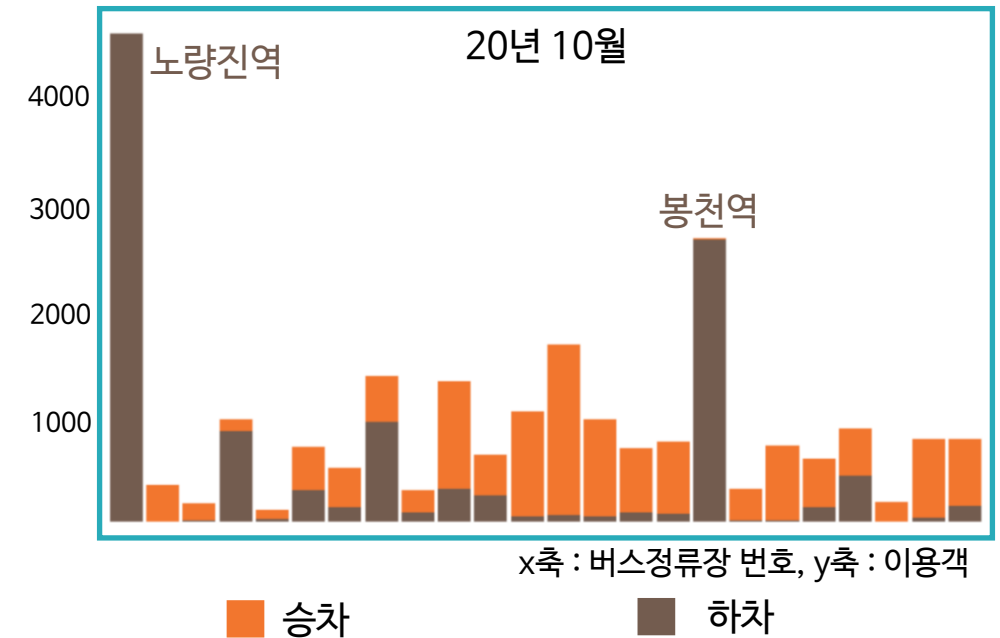
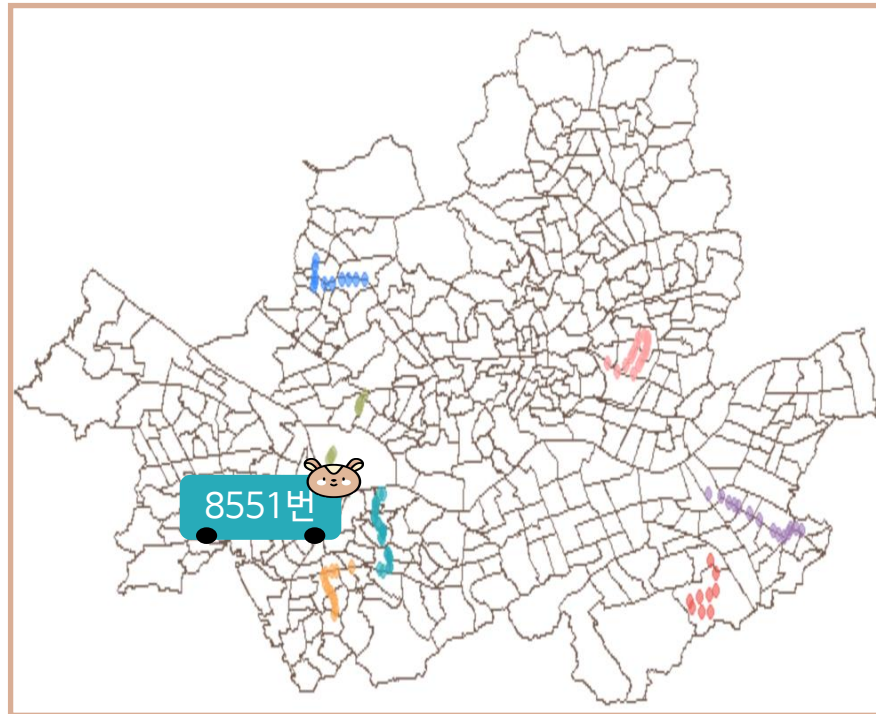
다람쥐버스 정류장별 이용현황

 주제선정

 전처리
및 EDA

 행정동 분석

 다음주 예고



노량진역, 봉천역이 이용객(모두 하차) 가장 많음

다람쥐 버스 노선 중 정류장 수 가장 많음



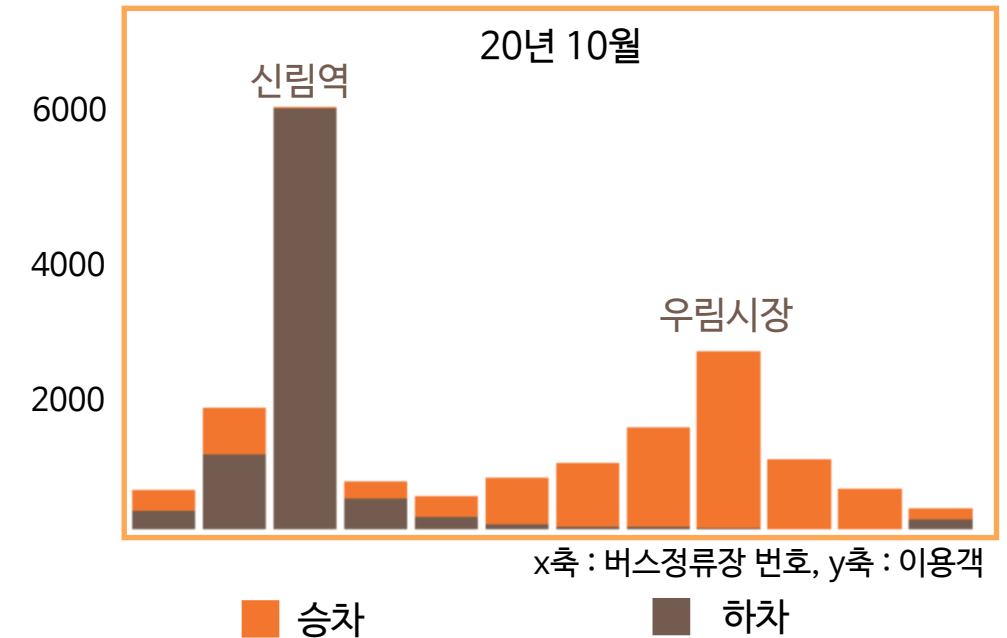
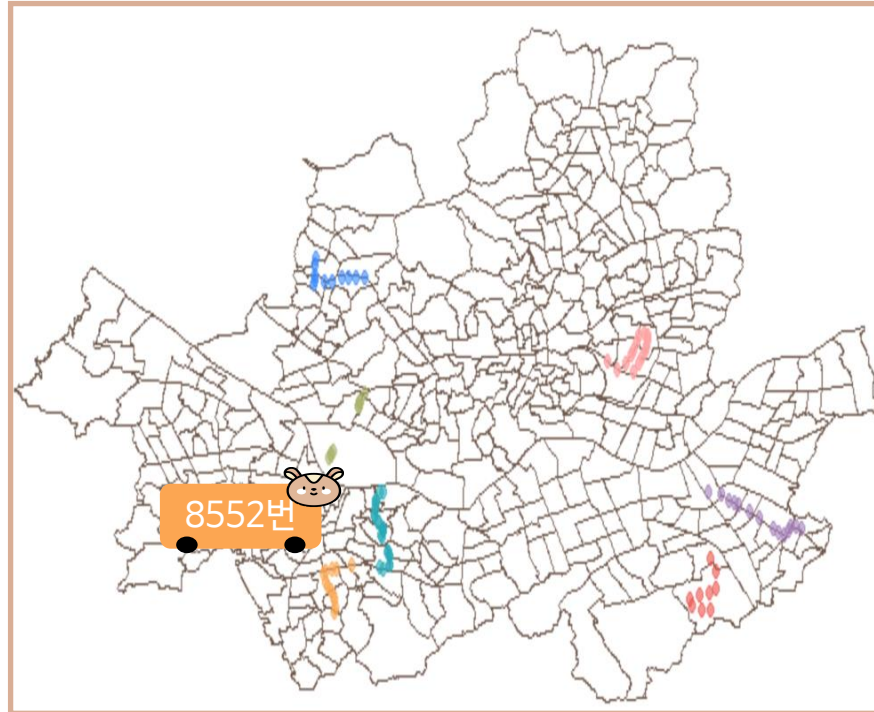
다람쥐버스 정류장별 이용현황

 주제선정

 전처리
및 EDA

 행정동 분석

 다음주 예고

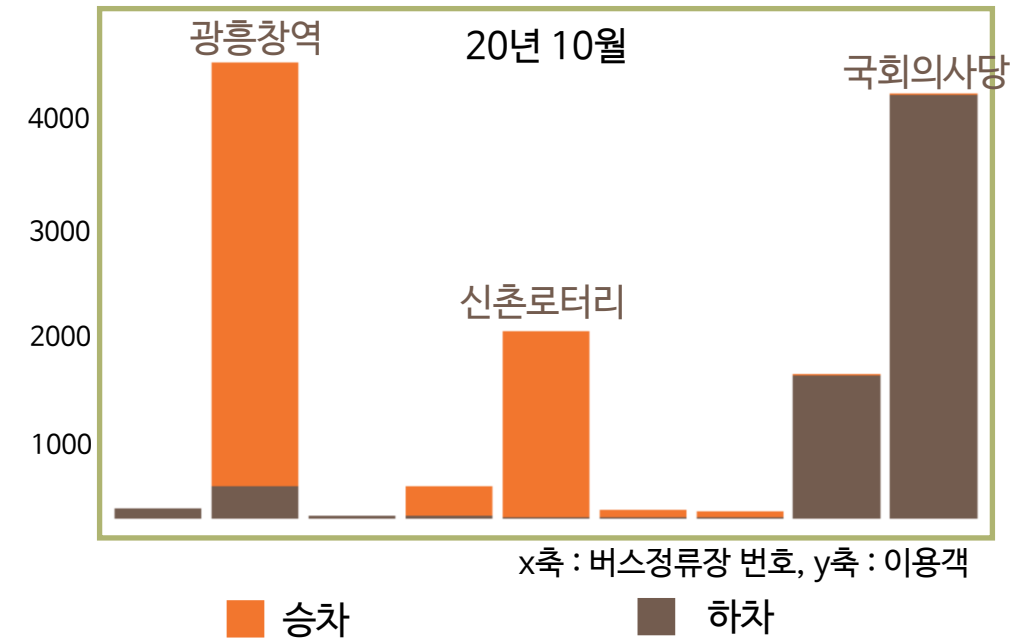
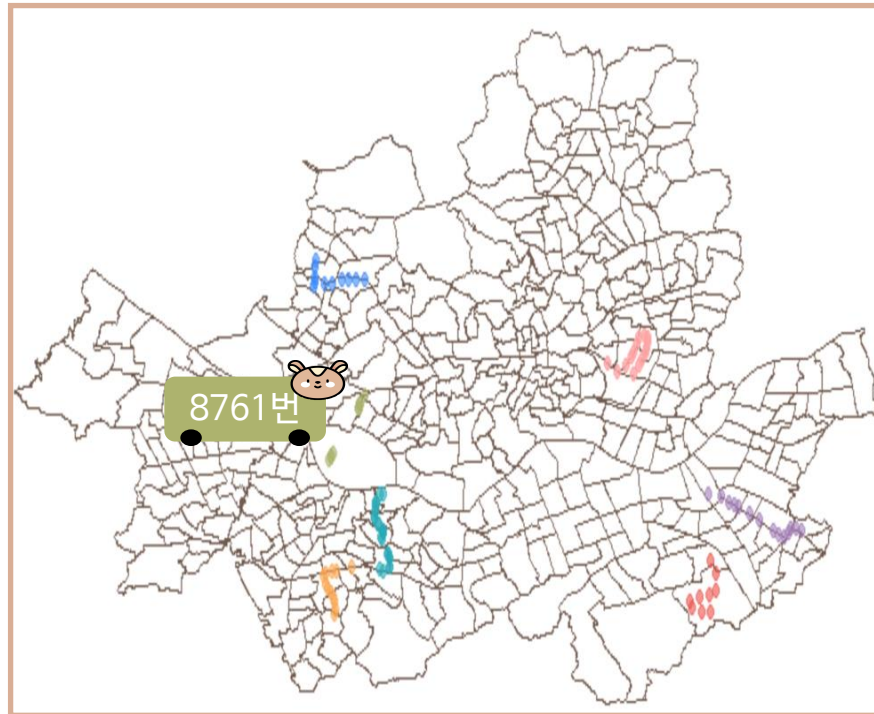


신림역이 이용객(모두 하차) 가장 많음

우림시장, 남강중고등학교입구, 세이브마트앞 순으로 승차 인원 수 높음

다람쥐버스 정류장별 이용현황

- 주제선정
- 전처리 및 EDA
- 행정동 분석
- 다음주 예고



승차는 **광흥창역**, 하차는 **국회의사당**에 집중되어 있음

특정정류장의 존재감이 유난히 큰 편



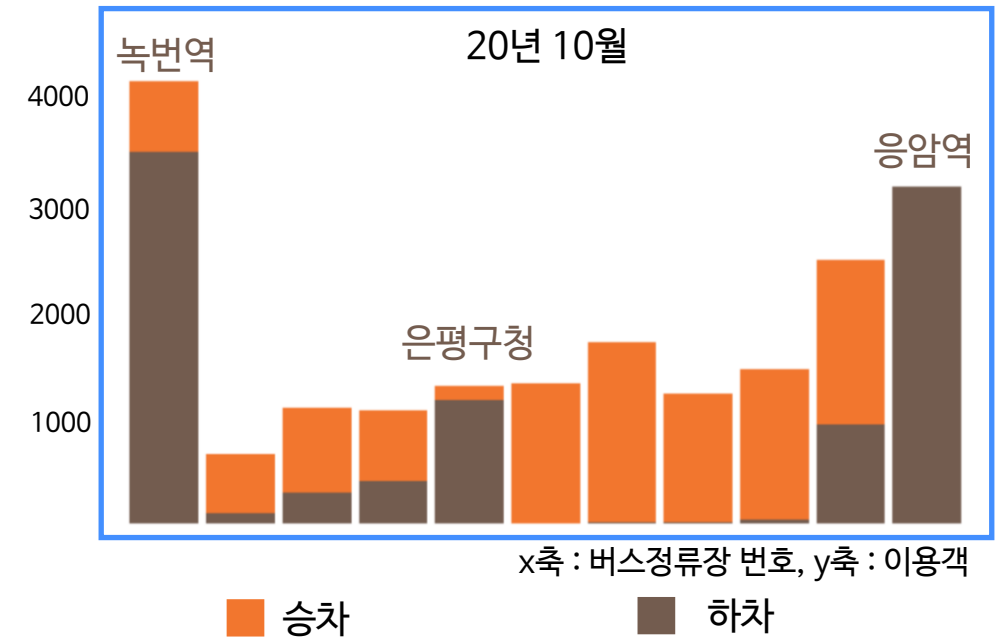
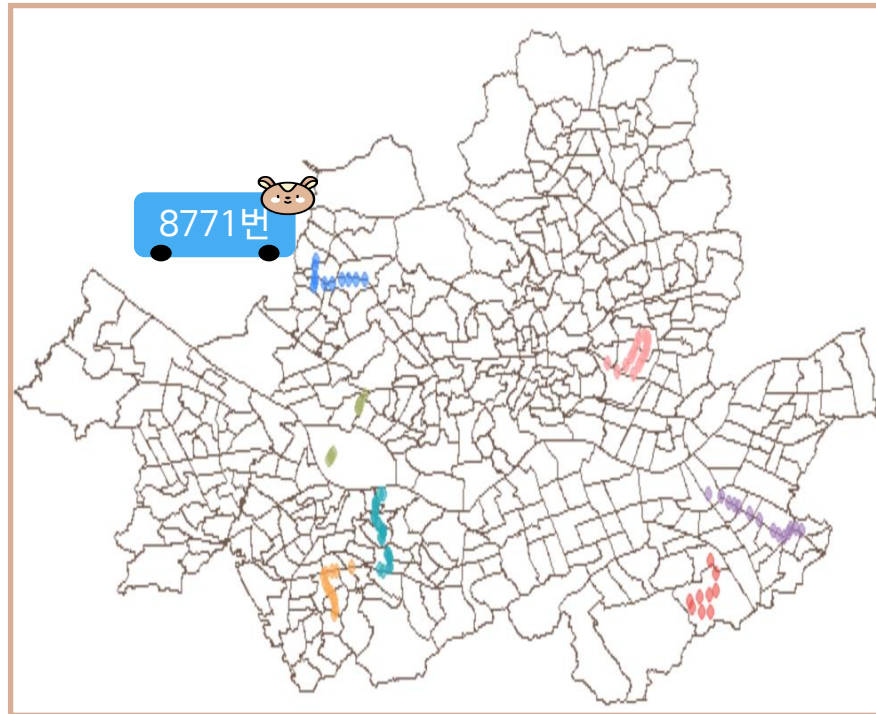
다람쥐버스 정류장별 이용현황

주제선정

전처리
및 EDA

행정동 분석

다음주 예고



녹번역과 응암역 이용객과 하차 매우 많음

은평구청은 지하철역과 연결되어 있진 않지만 하차가 많은 편!



다람쥐버스 정류장별 이용현황



주제선정

전처리
및 EDA

행정동 분석



다음주 예고

다람쥐버스 이용객의 월별 차이는 있었으나
정류장 이용에는 월별 차이가 거의 없음

이용객과 하차는 **지하철역 근처 정류장**에 집중되는 경향 있었음

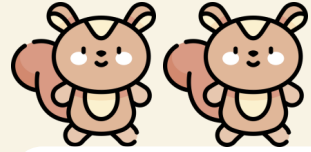
x축 : 버스정류장 번호, y축 : 이용객

승차

하차

녹번역과 응암역 이용객과 하차 매우 많음

은평구청은 지하철역과 연결되어 있진 않지만 하차가 많은 편!



전처리 및 EDA

사용 데이터 소개

서울시 버스노선별 정류장별 시간대별 승하차 인원(월단위)

“2020년 버스노선별 정류장별 시간대별 승하차 인원 정보(09월).csv”

x9 DATA SETS
39059 obs x 55 var

사용년월	노선번호	노선명	표준버스 정류장ID	버스정류장 ARS번호	역명	00시승차 총이용객수	00시하차 총이용객수	...	23시승차 총이용객수	23시하차 총이용객수	등록일자
202009	470	470번(상암차 고지~안골마 을)	1E+08	1001	종로2가 사거리	38	66	...	130	78	20201003
202009	N37	N37번(송파공 영차고지~진 관공영차고지)	1E+08	1001	종로2가 사거리	0	0	...	0	0	20201003
202009	741	741번(진관차 고지~현인릉 입구)	1E+08	1001	종로2가 사거리	63	108	...	133	69	20201003
202009	N37	N37번(진관공 영차고지~송 파공영차고지)	1E+08	1001	종로2가 사거리	0	0	...	0	0	20201003
202009	100	100번(하계동 ~용산구청)	1E+08	1002	창경궁.서울대 학교병원	0	0	...	19	2	20201003
202009	107	107번(송현동 ~동대문)	1E+08	1002	창경궁.서울대 학교병원	2	0	...	5	21	20201003

서울소재 버스정류장의 ARS번호는
01001 ~ 25999까지

다람쥐버스 운영 시간을 고려해
7시 · 8시의 승하차만!

주제선정

전처리
및 EDA

행정동 분석

다음주 예고

전처리

역명 · 버스정류장ARS번호 기준으로 그룹화

2020년 월별 정류장별 시간대별 승하차 인원정보 데이터 완성! (1월부터 9월까지)



주제선정



전처리
및 EDA



행정동 분석



다음주 예고

승차 인원 정보

역명	버스정류장 ARS번호	7시승차 총이용객평균	8시승차 총이용객평균
논현역	18503	7388	10978
가산디지털단지역	22587	6963	13464
삼익아파트앞	18220	6069	4984
개봉한진아파트.기점	17477	5554	3165
청림동푸르지오아파트 121동앞	21503	4203	3751
일원역	23546	4111	1872
금호타운아파트	21327	3818	3938
가산디지털단지역1호선	18797	3177	5550

하차 인원 정보

역명	버스정류장 ARS번호	7시하차 총이용객평균	8시하차 총이용객평균
길음전철역	8560	10426	10141
신림역4번출구	21916	6661	8123.333
가산디지털단지역	18834	6399	8812
신림역	21651	5654	6224
구로디지털단지역 환승센터	17014	5536.8	6959.8
당산역.지하철2호선	19645	5151	5400
회기역종점	6546	4977	5202
파리바게트. 낙성대지하철	21572	4842	7076

EDA

정류장별 시간별 총 이용객 평균 비교

	역명	7시승차 총이용객평균	8시승차 총이용객평균
지하철역	논현역	7388	10978
	일원역	4111	1872
주거지역	금호타운아파트	3818	3938
	삼익아파트앞	6069	4984

시간대 평균



역명	승차 총이용객평균
논현역	8970.5
일원역	2991.5
금호타운아파트	3878
삼익아파트앞	5526.5

	역명	7시하차 총이용객평균	8시하차 총이용객평균
지하철역	신림역4번출구	6661	8123.333
	길음전철역	10426	10141
주거지역	송화시장	3890	4201
	목2동주민센터	2994.5	2710.5

시간대 평균



역명	하차 총이용객평균
신림역4번출구	8970.5
길음전철역	2991.5
송화시장	3878
목2동주민센터	5526.5

주제선정

전처리
및 EDA

행정동 분석

다음주 예고

EDA

2020년 월별 정류장별 승하차 평균 순위

승차기준 순위

순위	1월	4월	8월
1	논현역	논현역	가산디지털단지역 (18503)
2	가산디지털단지역 (18503)	가산디지털단지역 (18503)	논현역
3	삼익아파트앞	삼익아파트앞	삼익아파트앞
4	개봉한진아파트.기점	매봉역	매봉역
5	낙성대역	개봉한진아파트.기점	개봉한진아파트.기점
6	매봉역	가산디지털단지역 1호선	가산디지털단지역 1호선
7	금호타운아파트	금호타운아파트	금호타운아파트
8	가산디지털단지역 1호선	독산역	구로디지털단지역
9	당산역.지하철2호선	당산역.지하철2호선	당산역.지하철2호선
10	서울대입구역	청림동푸르지오 아파트121동앞	낙성대역

하차기준 순위

순위	1월	4월	8월
1	길음전철역	길음전철역	길음전철역
2	신림역4번출구	신림역4번출구	신림역4번출구
3	가산디지털단지역 (18834)	가산디지털단지역 (18834)	가산디지털단지역 (18834)
4	경희대의료원	구로디지털단지역 환승센터	경희대의료원
5	구로디지털단지역 환승센터	파리바게트. 낙성대지하철	구로디지털단지역 환승센터
6	파리바게트.낙성대지 하철	신림역	신림역
7	신림역	당산역.지하철2호선	파리바게트. 낙성대지하철
8	당산역.지하철2호선	경희대의료원	당산역.지하철2호선
9	회기역종점	합정역	합정역
10	합정역	회기역종점	회기역종점



순위에서 약간의 변동만 있을 뿐 정류장은 비슷함



주제선정



전처리
및 EDA



행정동 분석

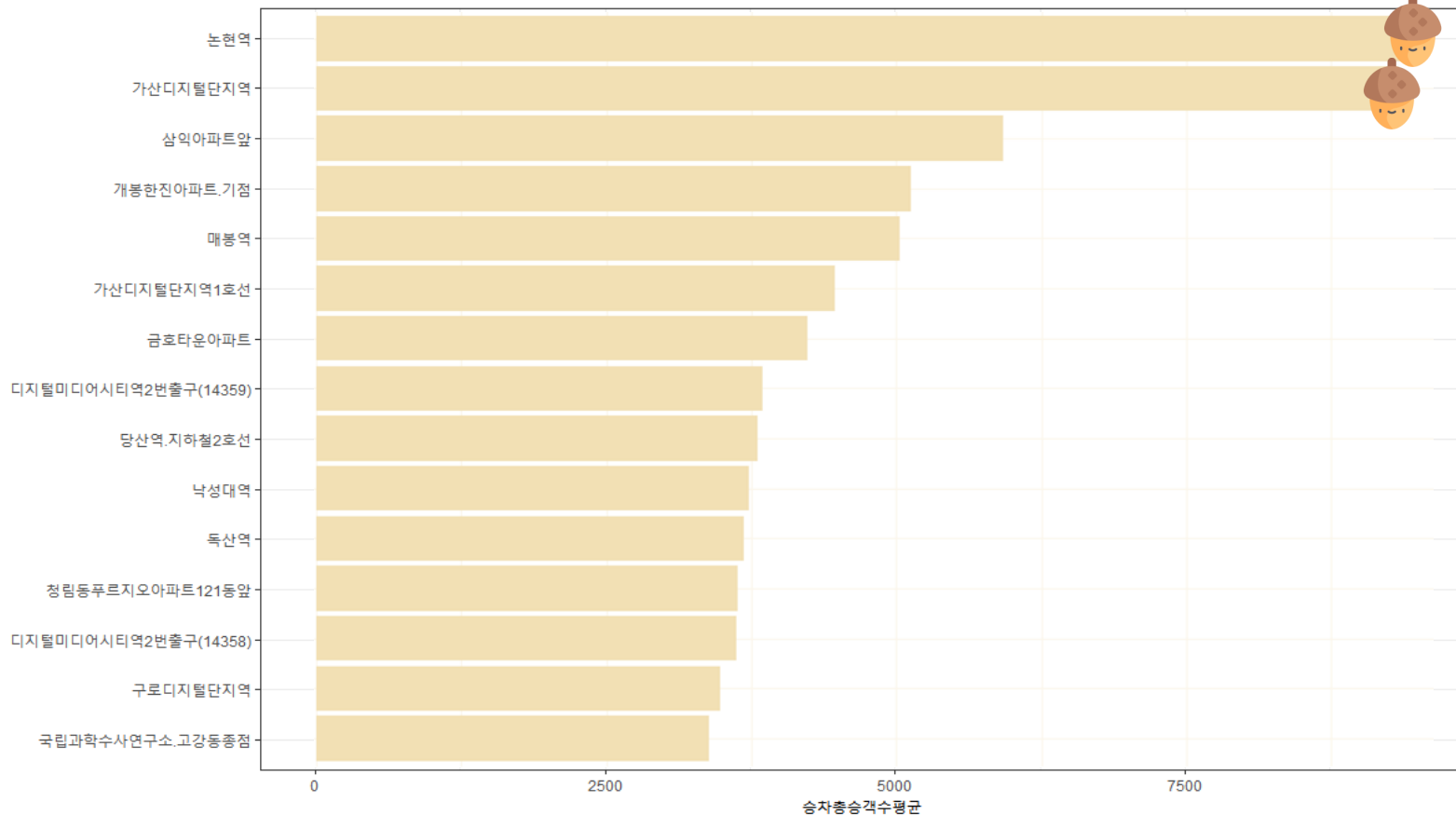


다음주 예고

EDA

2020년 정류장별 승차평균 상위 15개

“버스 정류장 별 승차 TOP 15”



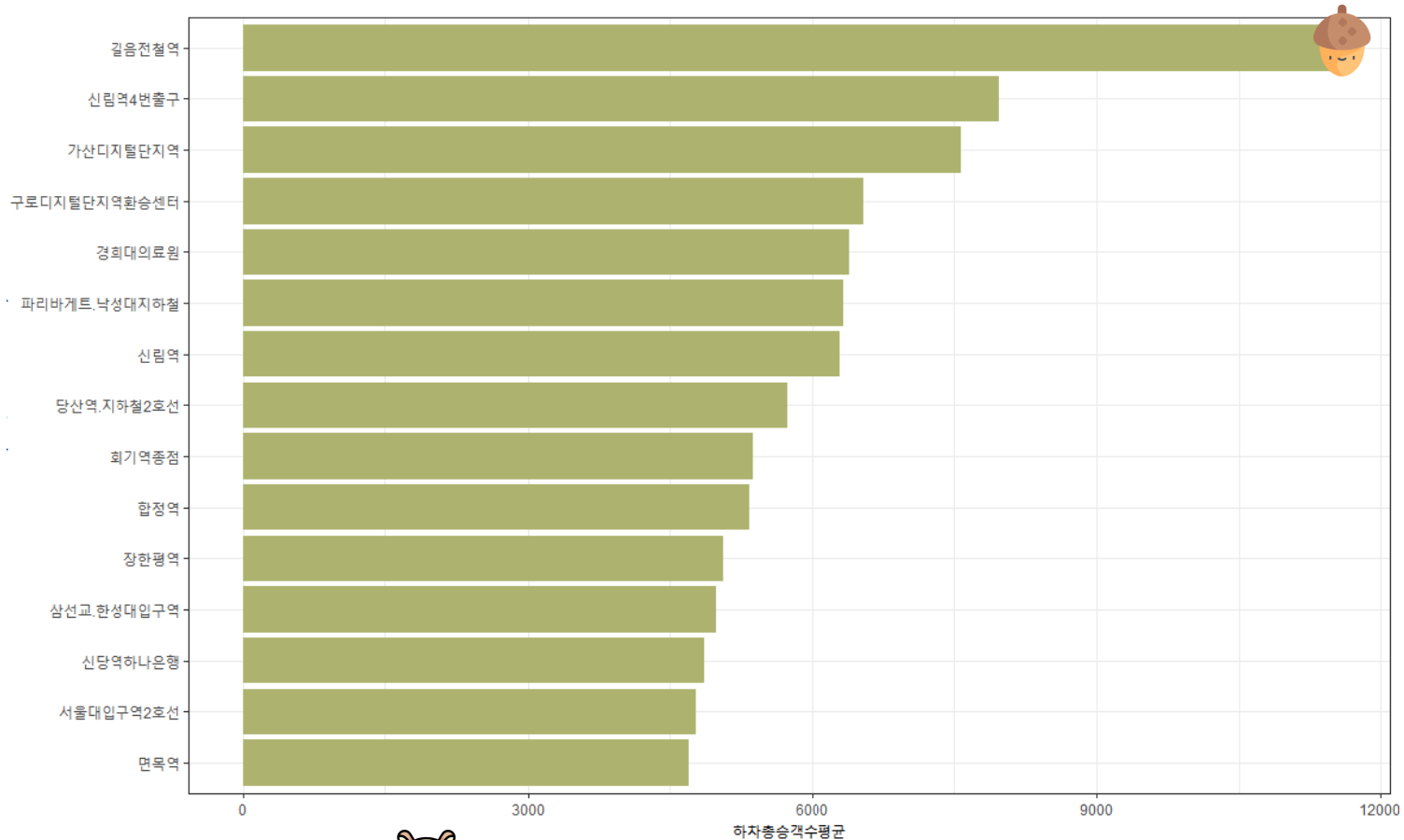
	역명
1	논현역
2	가산디지털단지역
3	삼익아파트앞
4	개봉한진아파트.기점
5	매봉역
6	가산디지털단지역1호선
7	금호타운아파트
8	디지털미디어시티역2번출구
9	당산역.지하철2호선
10	낙성대역
11	독산역
12	청림동푸르지오아파트121동앞
13	디지털미디어시티역2번출구
14	구로디지털단지역
15	국립과학수사연구소.고강동종점

주로 지하철역과 아파트단지에 고르게 분포

EDA

2020년 정류장별 하차평균 상위 15개

“버스 정류장 별 하차 TOP 15”



	역명
1	길음전철역
2	신림역4번출구
3	가산디지털단지역
4	구로디지털단지역환승센터
5	경희대의료원
6	파리바게트.낙성대지하철
7	신림역
8	당산역.지하철2호선
9	회기역종점
10	합정역
11	장한평역
12	삼선교.한성대입구역
13	신당역하나은행
14	서울대입구역2호선
15	면목역

지하철역에 가기 위해 버스를 이용하는 경우가 많아 보임

주제선정

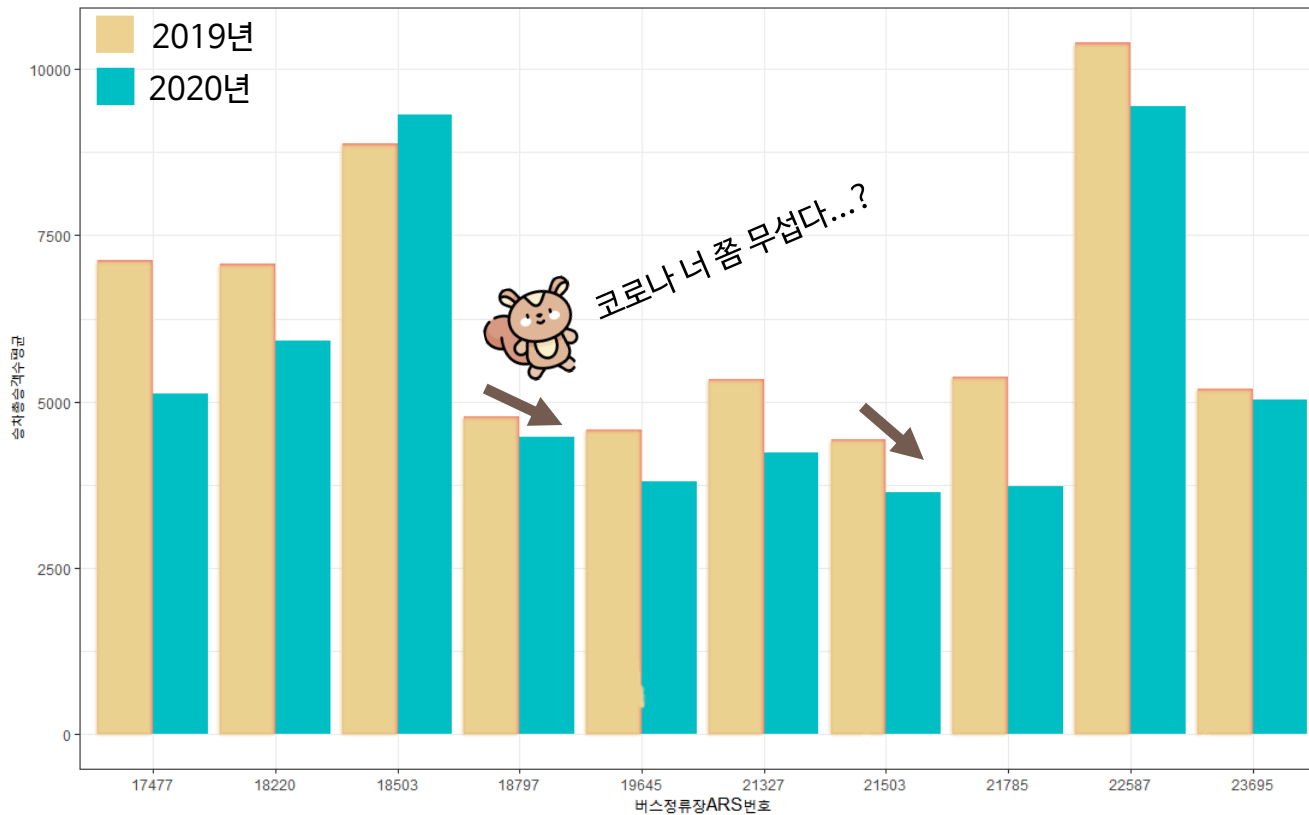
전처리
및 EDA

행정동 분석

다음주 예고

EDA

코로나로 인한 재택근무 및 비대면 수업이 정류장별 **승차**에 미친 영향



2019년 2020년 정류장 순위

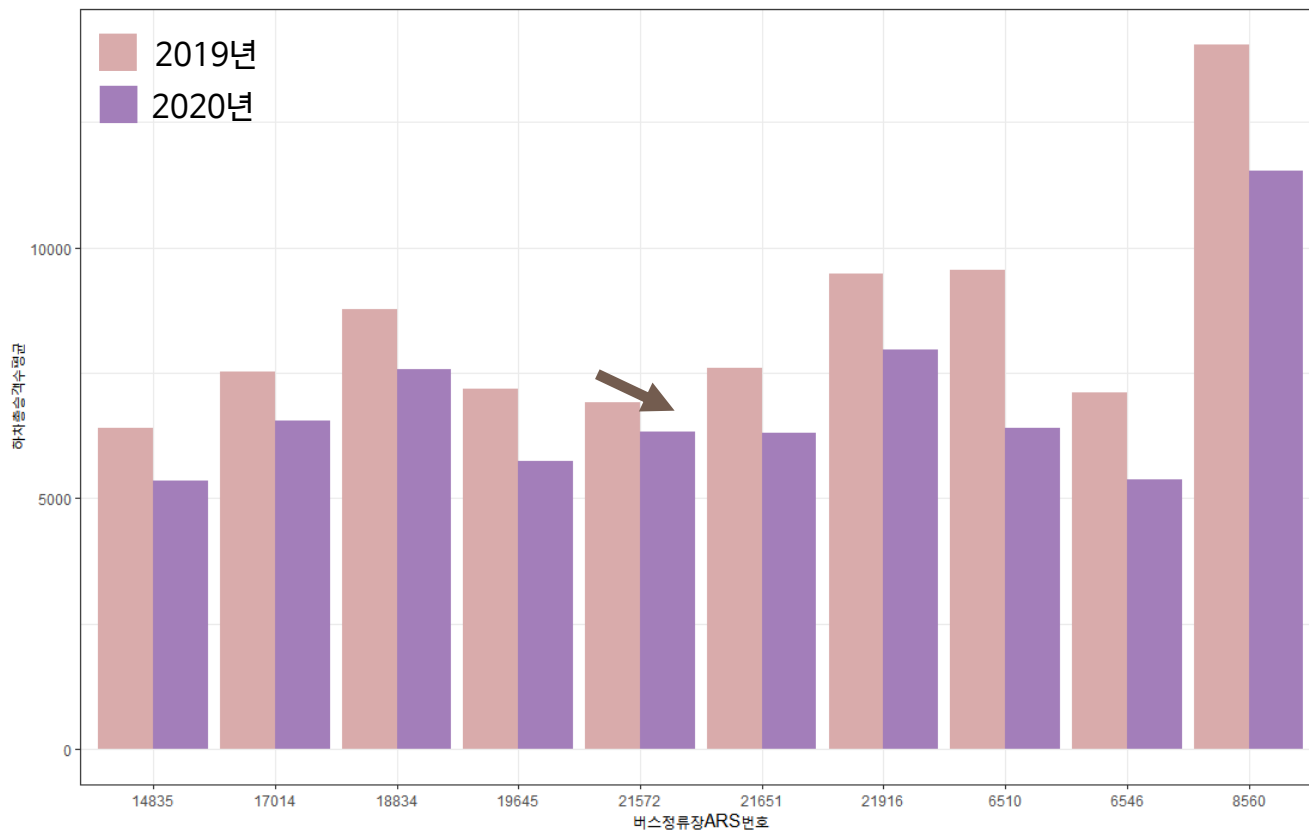
순위	2019년	2020년
1	논현역	논현역
2	가산디지털단지역	가산디지털단지역
3	개봉한진아파트.기점	삼익아파트앞
4	삼익아파트앞	개봉한진아파트.기점
5	낙성대역	매봉역
6	금호타운아파트	가산디지털단지역 1호선



이용객은 감소했지만 이용하는 정류장은 변동 없음!

EDA

코로나로 인한 재택근무 및 비대면 수업이 정류장별 **하차**에 미친 영향



2019년 2020년 정류장 순위

순위	2019년	2020년
1	길음전철역	길음전철역
2	경희대의료원	신림역4번출구
3	신림역4번출구	가산디지털단지역
4	가산디지털단지역	구로디지털단지역 환승센터
5	신림역	경희대의료원
6	구로디지털단지역 환승센터	파리바게트, 낙성대지하철



하차기준으로도 이용객은 감소했지만 이용하는 정류장은 변동 없음!

전처리

카카오 API 주소검색으로 2020년 정류장별 승하차 평균 데이터와 위치좌표를 병합

승차정류장 위치좌표

역명	버스정류장 ARS번호	승차평균	X좌표	Y좌표
논현역	22587	9436.388	127.0219091	37.50931936
가산디지털단지역	18503	9323.666	126.8830204	37.48106947
삼익아파트앞	18220	5924.611	126.909455	37.462545

행정동 병합

정류장별 승차총이용객수평균(행정동별)

역명	버스정류장 ARS번호	승차평균	행정동
논현역	22587	9436.388	반포1동
가산디지털단지역	18503	9323.666	가산동
삼익아파트앞	18220	5924.611	시흥4동

하차정류장 위치좌표

역명	버스정류장 ARS번호	하차평균	X좌표	Y좌표
길음전철역	8560	11524.33	127.0241163	37.60408552
신림역4번출구	21916	7966.537	126.9290486	37.48391302
가산디지털단지역	18834	7569.166	126.8831128	37.48128853

정류장별 하차총이용객수평균(행정동별)

역명	버스정류장 ARS번호	하차평균	행정동
길음전철역	8560	11524.333	길음1동
신림역4번출구	21916	7966.537	서원동
가산디지털단지역	18834	7569.166	가산동

주제선정

전처리
및 EDA

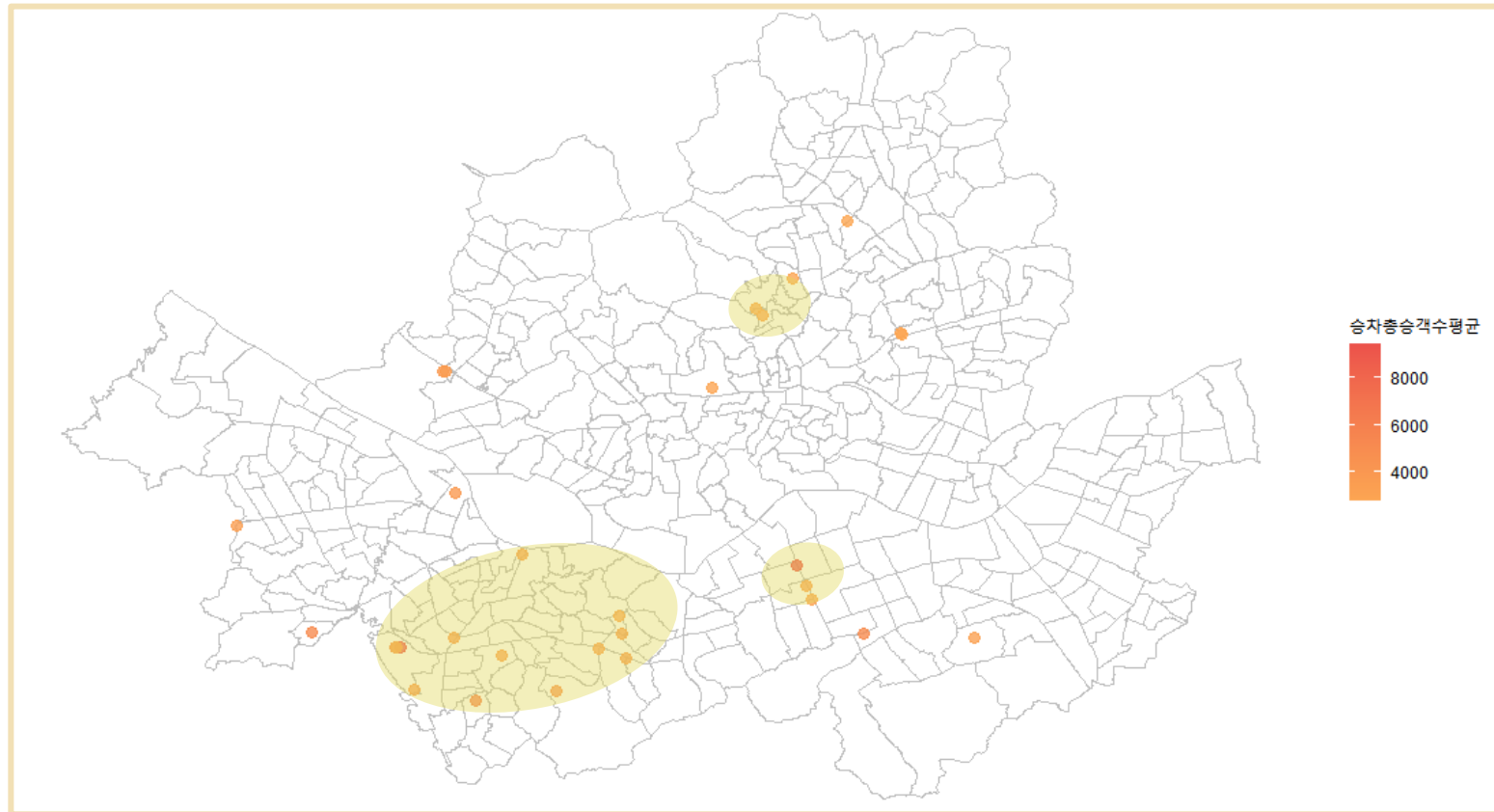
행정동 분석

다음주 예고

EDA

승차평균 상위 30개 정류장 위치

“버스 정류장 별 승차 TOP 30”



주제선정

전처리
및 EDA

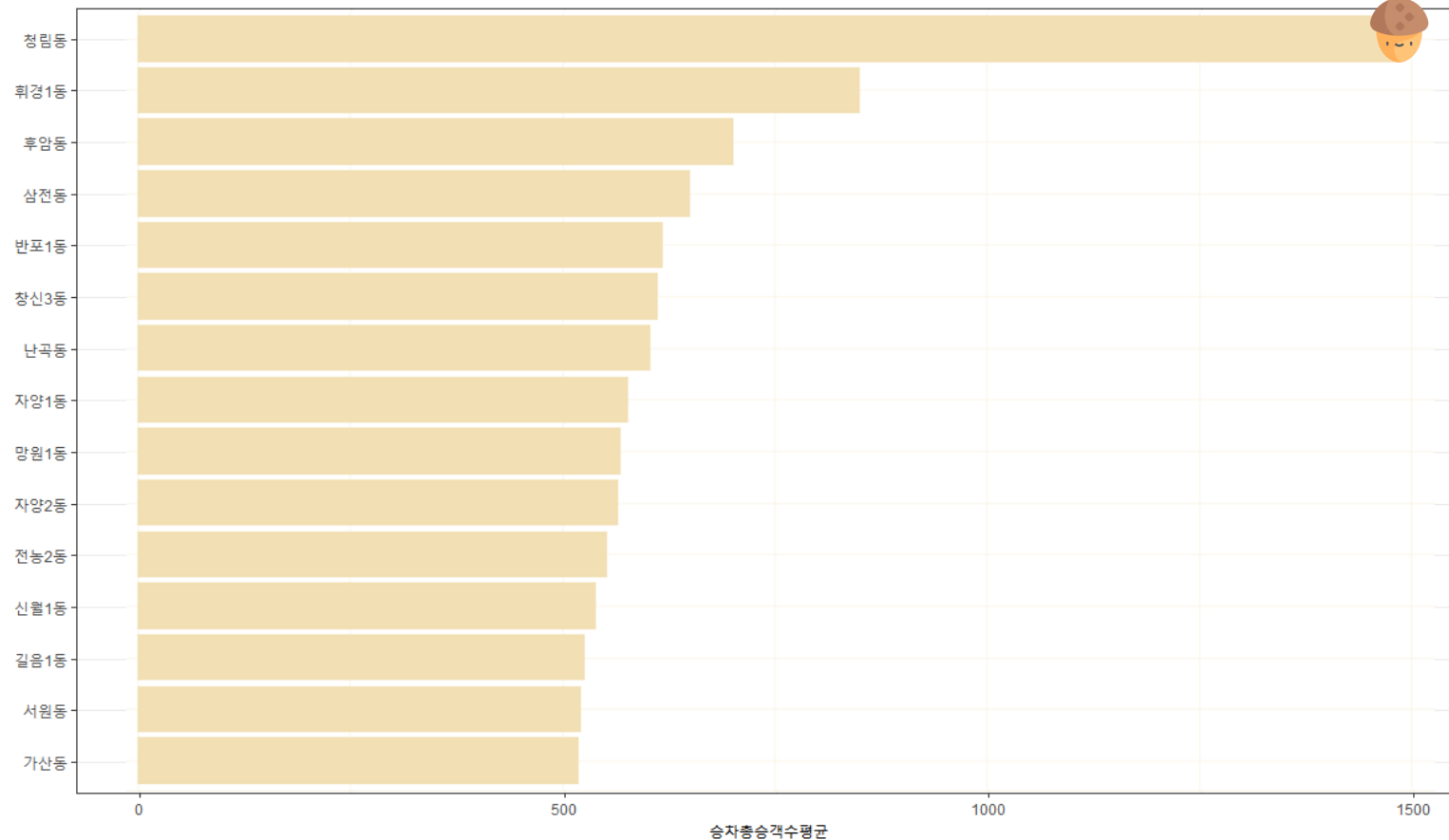
행정동 분석

다음주 예고

EDA

행정동별 승차평균 상위 15개

“행정동별 승차 Top15”



주제선정

전처리
및 EDA

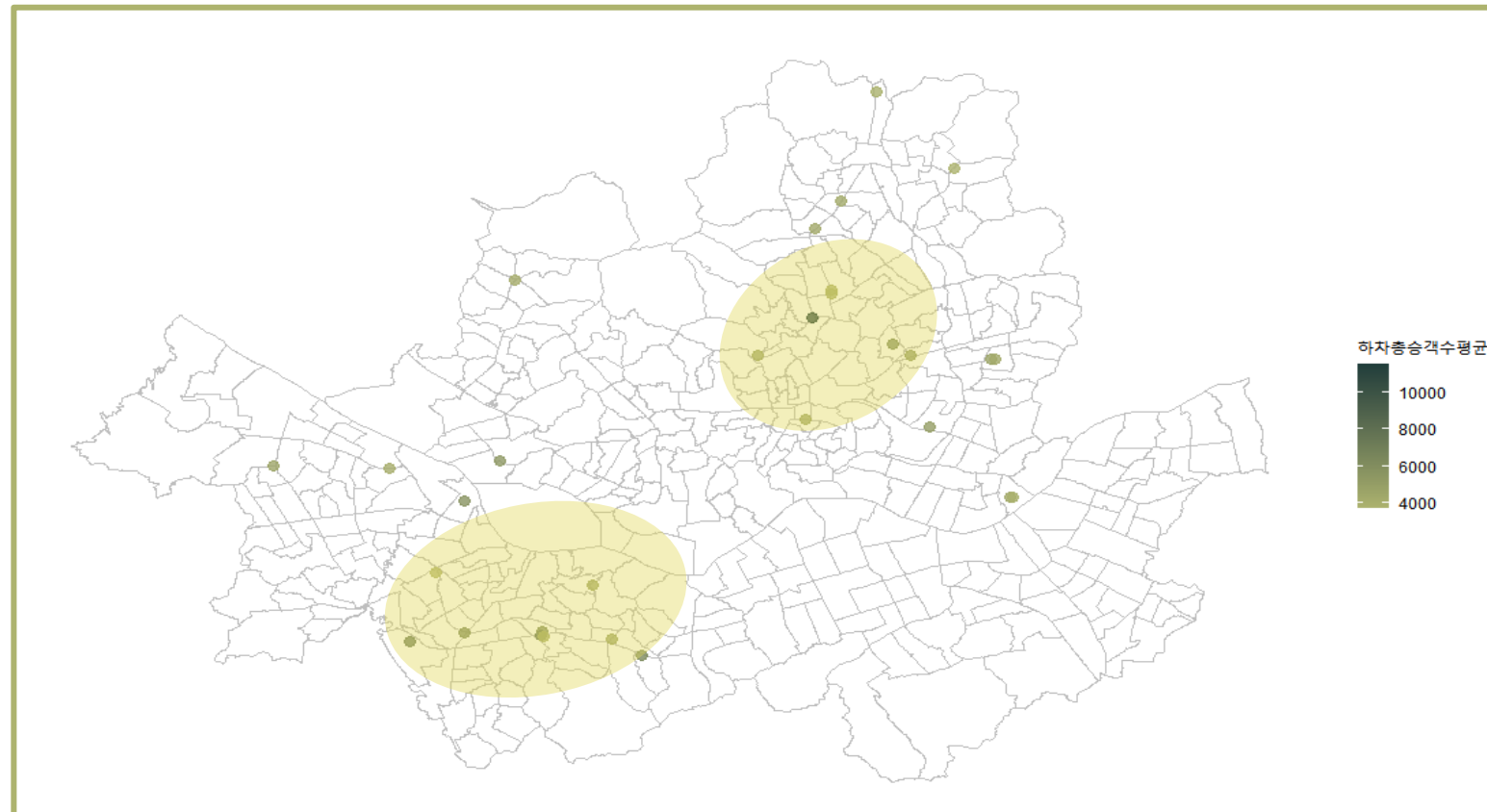
행정동 분석

다음주 예고

EDA

하차평균 상위 30개 정류장 위치

“버스 정류장 별 하차 TOP 30”



주제선정

전처리
및 EDA

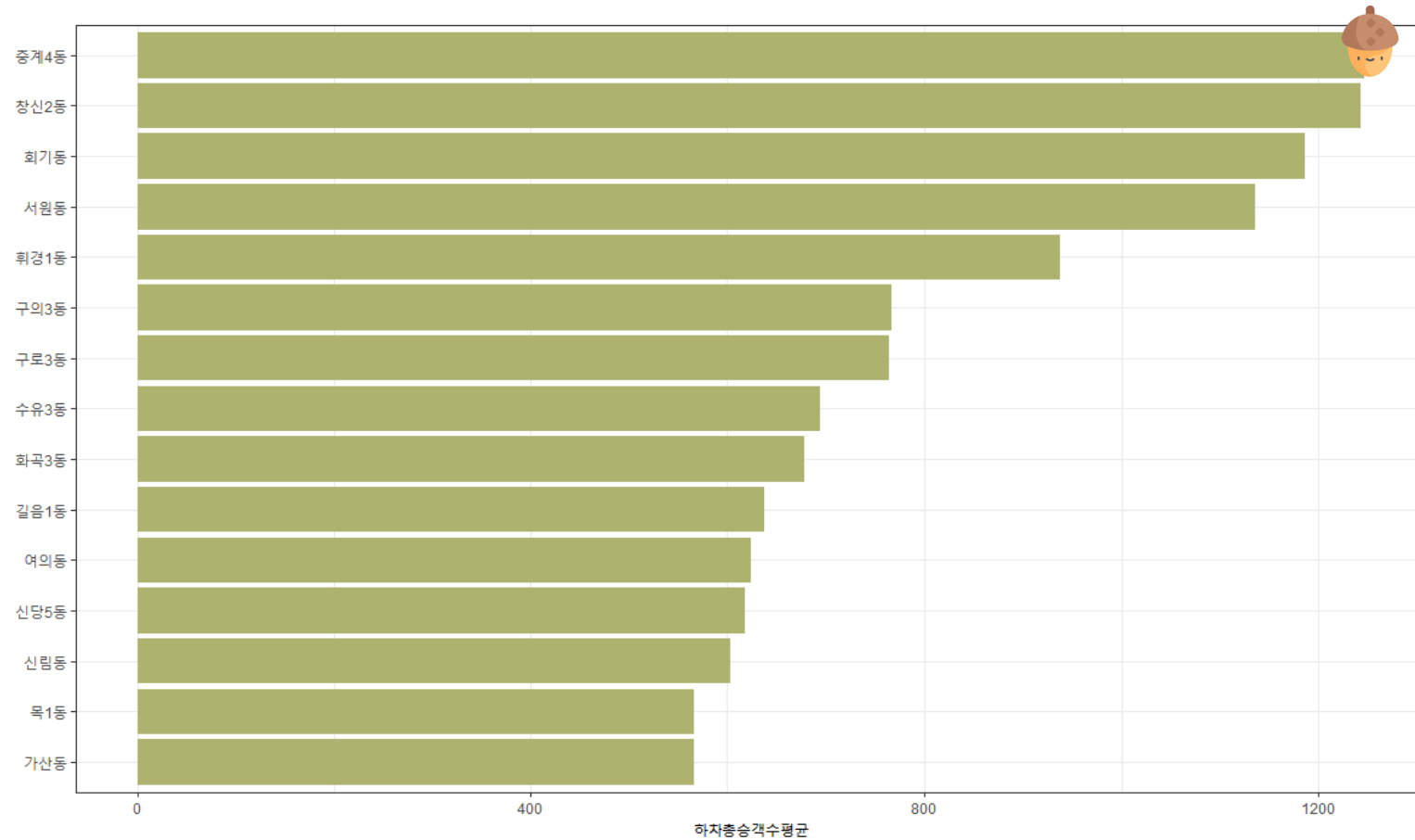
행정동 분석

다음주 예고

EDA

행정동별 하차평균 상위 15개

“행정동별 하차 Top15”



주제선정

전처리
및 EDA

행정동 분석

다음주 예고

여기서 잠깐!!

이제부터는 서울시 내 출근시간대 버스의 수요를 예측하기 위해 수요에 영향을 주는 변수에 관한 데이터를 전처리 및 EDA 함



주제선정



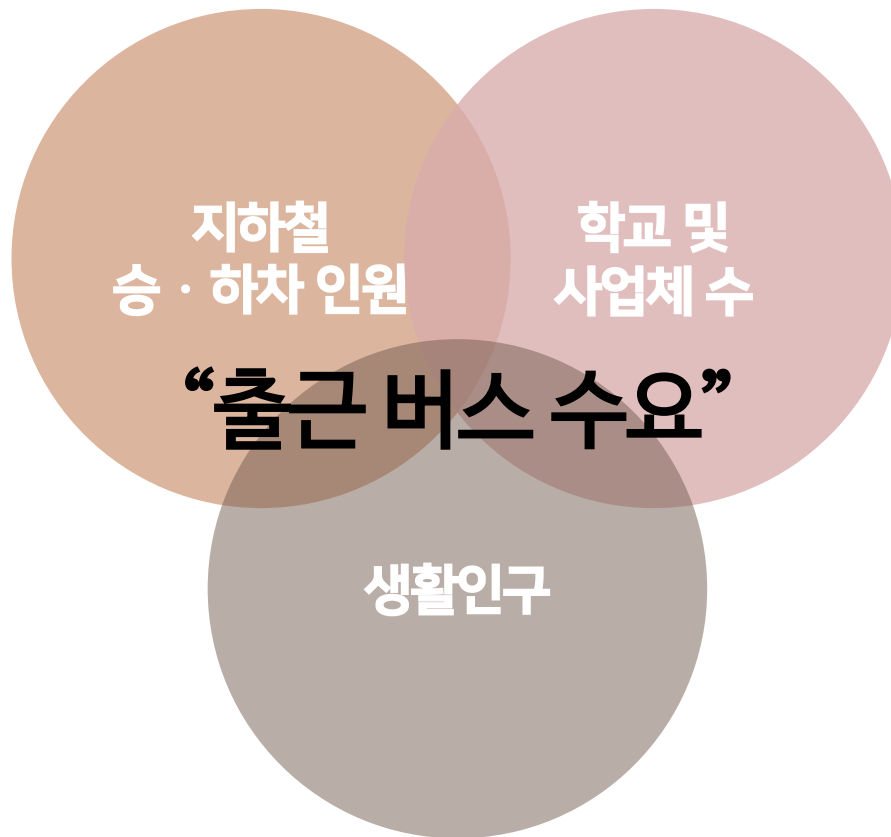
전처리
및 EDA



행정동 분석



다음주 예고



세 가지 요인으로 서울시 내 행정동의 특성을 파악하고
출근시간대 버스 수요가 필요한 지역을 시각적으로 확인할 예정

데이터 소개 및 전처리

서울시 내 지하철 역 주소 파악을 위해 철도 데이터 포털의 ‘수도권 지하철 주소 데이터’ 이용

“수도권 N호선 주소 데이터.csv”

철도운영기관	선명	역명	지번주소	도로명주소
코레일	1호선	소요산	경기도 동두천시 상봉암동 126-3	경기도 동두천시 평화로 2925
코레일	1호선	동두천	경기도 동두천시 동두천동 245-210	경기도 동두천시 평화로 2539
코레일	1호선	보산	경기도 동두천시 보산동 422	경기도 동두천시 평화로 2539
...
코레일	1호선	용산	서울특별시 동작구 노량진동 67-2	서울특별시 동작구 노량진로 151
코레일	1호선	노량진	서울특별시 영등포구 신길동 1368-3	서울특별시 영등포구 여의대방로 300
코레일	1호선	대방	서울특별시 영등포구 신길동 41-1	서울특별시 영등포구 영등포로 327

× 15 DATA SETS

1-9호선 및
경의중앙선 · 경춘선 · 공항철도
분당선 · 신분당선 · 우이신설

주제선정

전처리
및 EDA

행정동 분석

다음주 예고

데이터 소개 및 전처리

지하철 주소 데이터의 지번주소가 '서울특별시'로 시작하는 지하철역만 추출 후 하나의 데이터셋으로 저장

“수도권 N호선 주소 데이터.csv”

철도운영기관	선명	역명	지번주소	도로명주소	선명	역명	지번주소
코레일	1호선	소요산	경기도 동두천시 상봉암동 126-3	경기도 동두천시 평화로 2925	1호선	서울역	서울특별시 용산구 동자동 14-151
코레일	1호선	동두천	경기도 동두천시 동두천동 245-210	경기도 동두천시 평화로 2539	2호선	합정	서울특별시 마포구 서교동 393
코레일	1호선	보산	경기도 동두천시 보산동 422	경기도 동두천시 평화로 2539	3호선	강남	서울특별시 강남구 역삼동 858
...
코레일	1호선	용산	서울특별시 동작구 노량진동 67-2	서울특별시 동작구 노량진로 151	9호선	여의도	서울특별시 영등포구 여의도동 3
코레일	1호선	노량진	서울특별시 영등포구 신길동 1368-3	서울특별시 영등포구 여의대방로 300	분당	왕십리	서울특별시 성동구 행당동 192
코레일	1호선	대방	서울특별시 영등포구 신길동 41-1	서울특별시 영등포구 영등포로 327	우이신설	정릉	서울특별시 성북구 정릉동 139-80



주제선정



전처리
및 EDA



행정동 분석



다음주 예고

데이터 소개 및 전처리

카카오 API를 이용해 지번주소를 행정동 및 위도, 경도로 추출 후 이들을 변수로 추가

“수도권 N호선 주소 데이터.csv”

선명	역명	지번주소
1호선	서울역	서울특별시 용산구 동자동 14-151
2호선	합정	서울특별시 마포구 서교동 393
3호선	강남	서울특별시 강남구 역삼동 858
...
9호선	여의도	서울특별시 영등포구 여의도동 3
분당	왕십리	서울특별시 성동구 행당동 192
우이신설	정릉	서울특별시 성북구 정릉동 139-80

선명	역명	지번주소	행정동	경도	위도
1호선	서울역	서울특별시 용산구 동자동 14-151	회현동	126.9726	37.5569
2호선	합정	서울특별시 마포구 서교동 393	서교동	126.9149	37.55012
3호선	강남	서울특별시 강남구 역삼동 858	역삼1동	127.0282	37.50125
...
9호선	여의도	서울특별시 영등포구 여의도동 3	여의동	126.9269	37.52817
분당	왕십리	서울특별시 성동구 행당동 192	행당1동	127.0606	37.56122
우이신설	정릉	서울특별시 성북구 정릉동 139-80	정릉1동	127.0135	37.6032

주제선정

전처리
및 EDA

행정동 분석

다음주 예고

데이터 소개 및 전처리

오전 7시-9시의 서울시 내 지하철 역의 승·하차 인원을 파악하기 위해 ‘역 별 승·하차 인원 데이터’ 이용

“지하철 역 별 시간대 별 승·하차 인원.csv”

39888 obs x 52 var

사용월	호선명	지하철역	04-05시 승차인원	04-05시 하차인원	...	03-04시 하차인원
202009	1호선	소요산	483	11	...	0
202009	1호선	동두천	126	0	...	0
202009	1호선	보산	713	9	...	0
...
201501	분당선	용산	60	3	...	0
201501	분당선	노량진	145	5	...	0
201501	분당선	대방	253	0	...	0

2015.01-2020.09
(월별)

시간대 별
승·하차 인원

주제선정

전처리
및 EDA

행정동 분석

다음주 예고

데이터 소개 및 전처리

2020년 1월에서 2020년 9월까지의 7시-9시 승·하차 데이터로 축소 후, 앞서 전처리한 주소 데이터와 병합해
최종 지하철 데이터셋 완성

“지하철 역 별 승·하차 인원_최종.csv”



267 obs x 7 var

역명	지번주소	행정동	경도	위도	승차 인원 평균	하차 인원 평균
도봉	서울특별시 도봉구 도봉동 639	도봉2동	127.0449	37.6752	17479.06	6495.111
방학	서울특별시 도봉구 방학동 728	방학1동	127.0444	37.66593	26802.5	12679.06
창동	서울특별시 도봉구 창동 135-1	창4동	127.0477	37.65338	41205.94	13360.5
...
녹천	서울특별시 도봉구 창동 57-11	창1동	127.0509	37.64564	17482.94	3584.556
월계	서울특별시 노원구 월계동 263-1	월계2동	127.0589	37.63507	19272.72	6558.278
광운대	서울특별시 노원구 월계동 85	월계1동	127.0623	37.62341	29890.67	7160.722

총 267 개 역

7 - 9시
승·하차 인원 평균



주제선정



전처리
및 EDA



행정동 분석

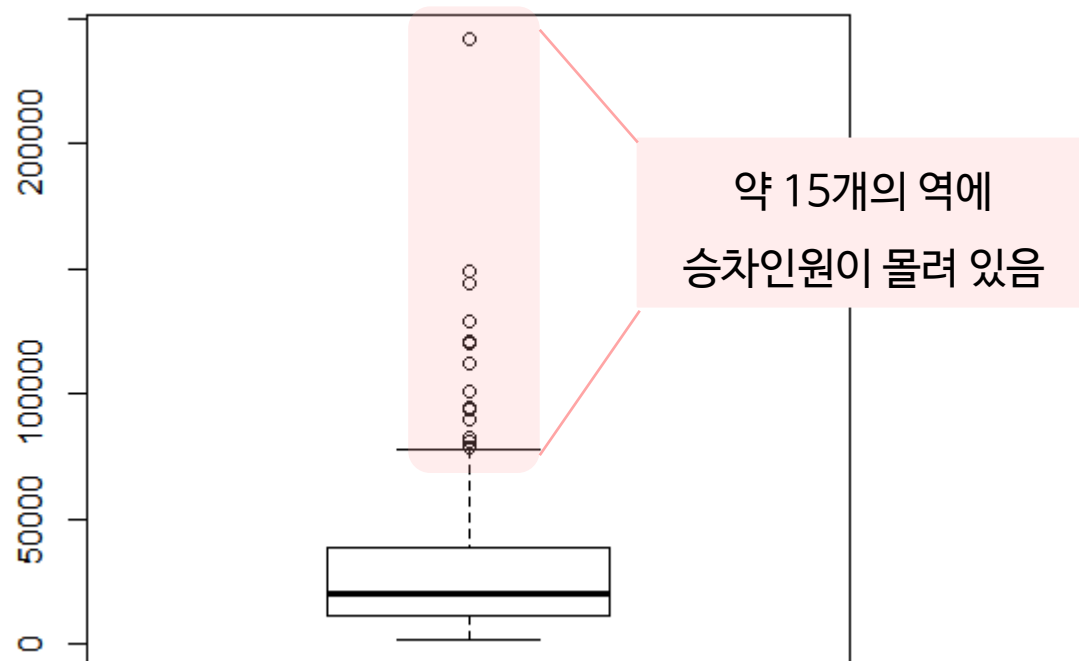


다음주 예고

EDA

서울시 내 출근 시간대에 승차가 많은 지하철 시각화

“지하철 역 별 승차 인원분포”



Min	2,103
1분위수	11,709
3분위수	38,232
Max	241,162

주제선정

전처리
및 EDA

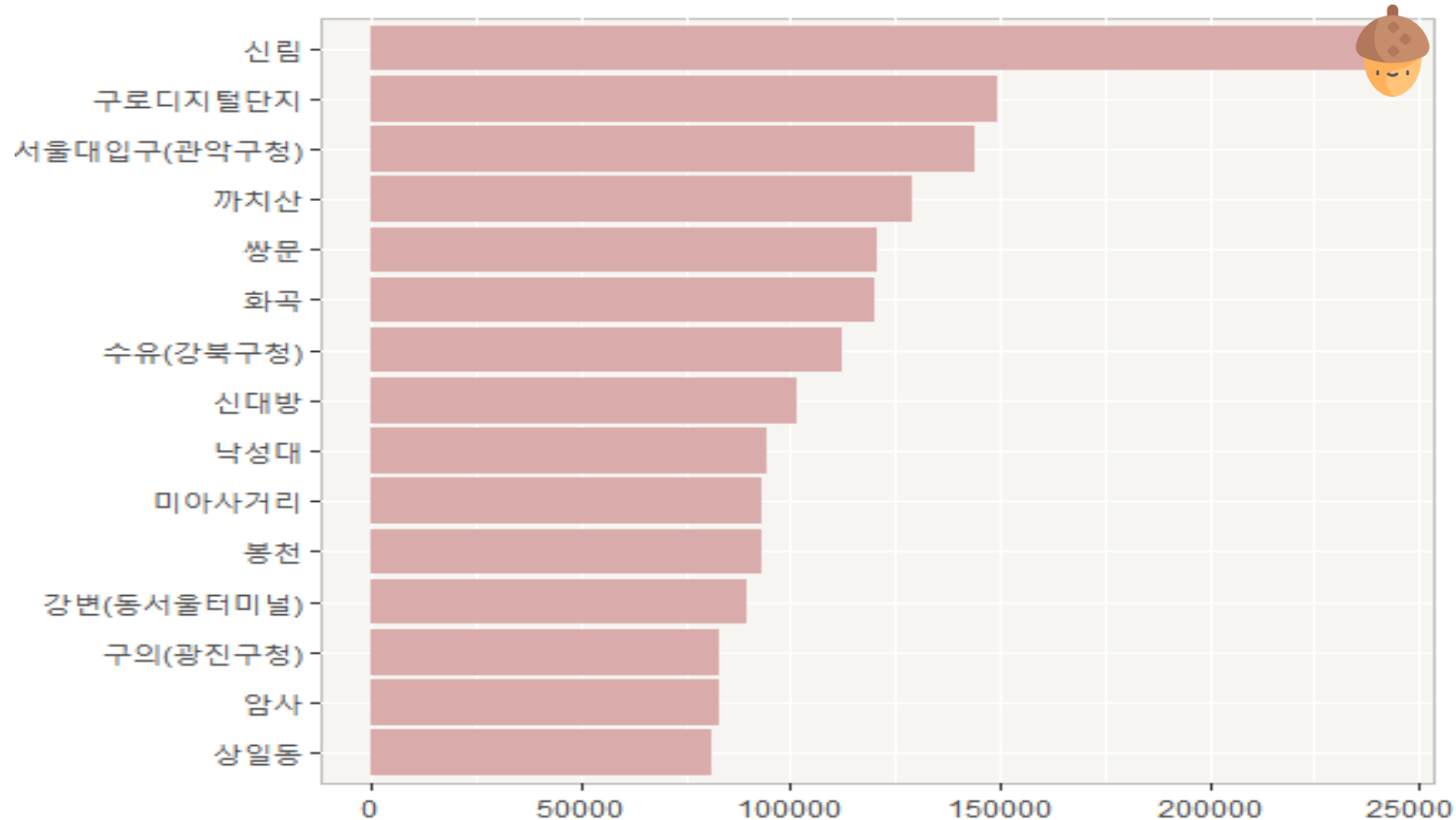
행정동 분석

다음주 예고

EDA

서울시 내 출근 시간대에 승차가 많은 지하철 시각화

“지하철 역 별 승차 TOP 15”



주제선정

전처리
및 EDA

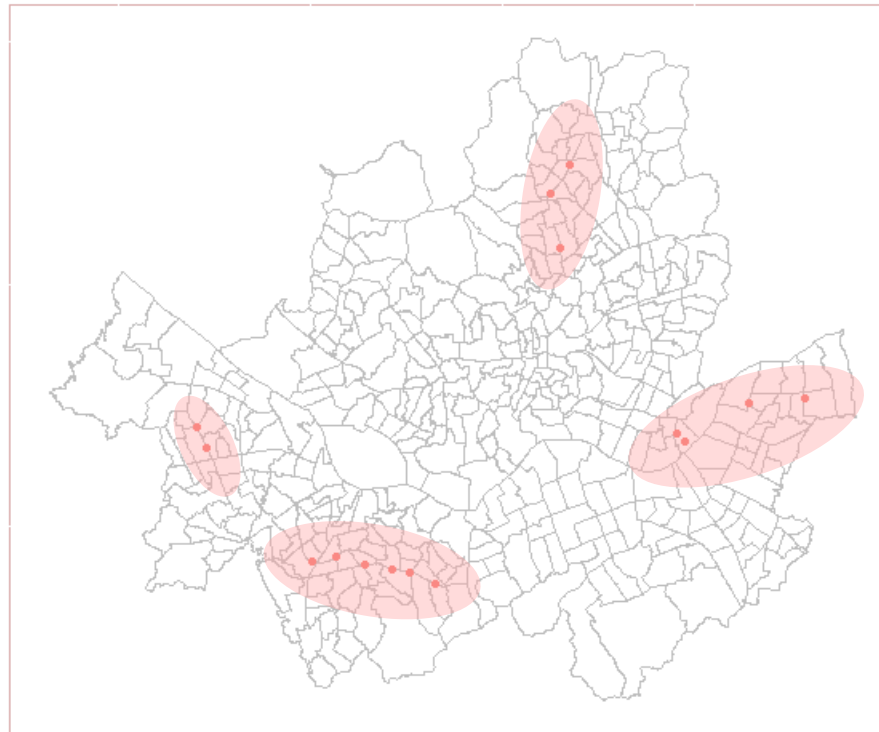
행정동 분석

다음주 예고

EDA

서울시 내 출근 시간대에 승차가 많은 지하철 시각화

“지하철 역 별 승차 TOP 15”



주로 관악구 · 강서구 · 강북구 · 강동구에 승차인원이 많은 지하철 역 집중됨



주제선정



전처리
및 EDA



행정동 분석

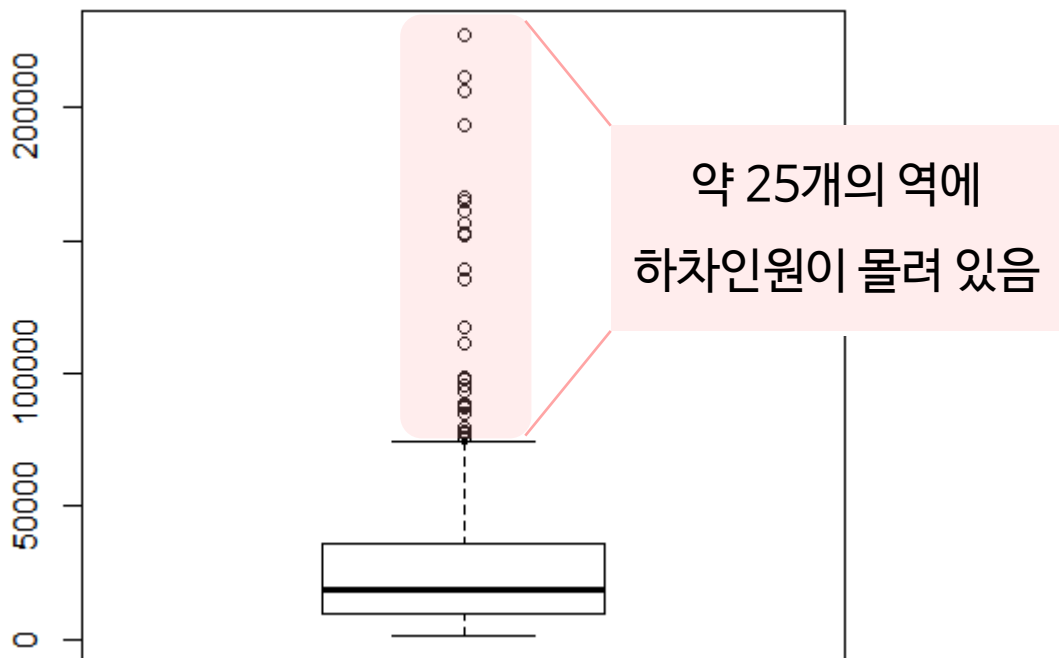


다음주 예고

EDA

서울시 내 출근 시간대에 하차가 많은 지하철 시각화

“지하철 역 별 하차 인원 분포”



Min	1,434
1분위수	9,413
3분위수	35,459
Max	227,112

주제선정

전처리
및 EDA

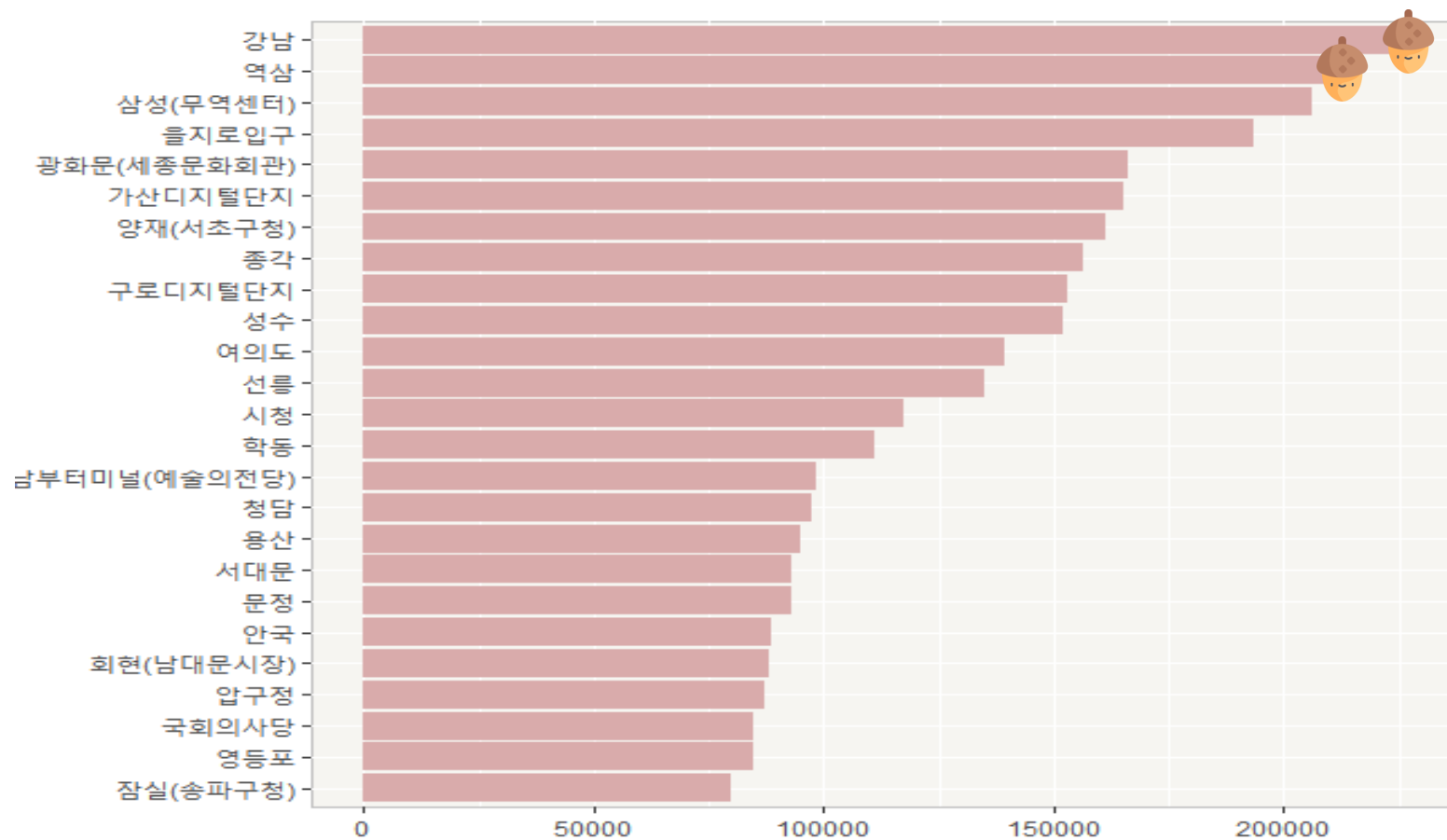
행정동 분석

다음주 예고

EDA

서울시 내 출근 시간대에 하차가 많은 지하철 시각화

“지하철 역 별 하차 TOP 25”



주제선정

전처리
및 EDA

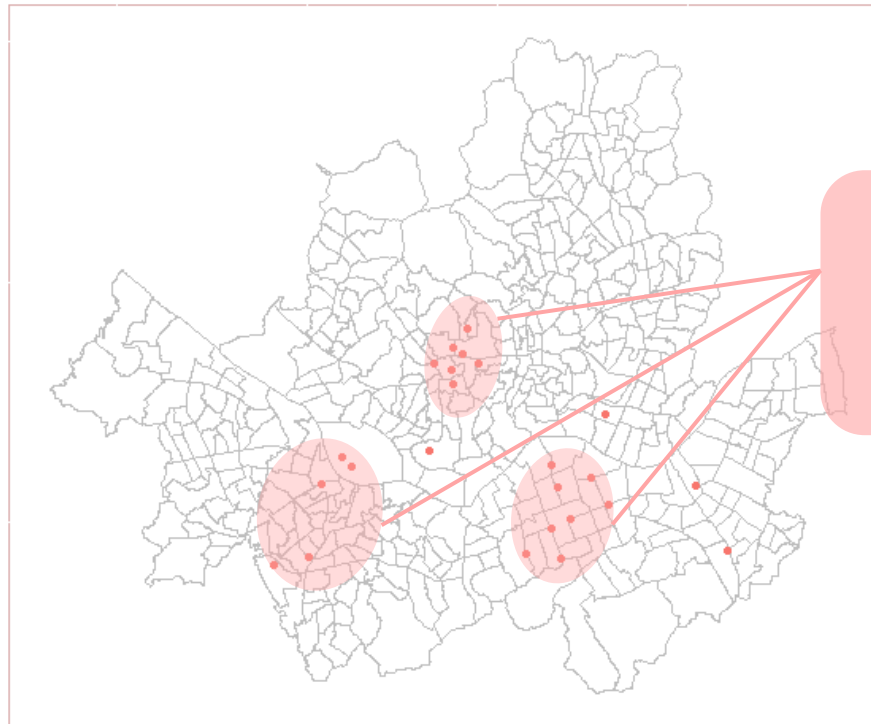
행정동 분석

다음주 예고

EDA

서울시 내 출근 시간대에 하차가 많은 지하철 시각화

“지하철 역 별 하차 TOP 25”



이 지역으로 출근하는
사람이 많음을 추측
(사업체 데이터 와도 같이 볼 예-정)

주로 강남구 · 종로구 · 영등포구에 하차인원이 많은 지하철 역 집중됨

주제선정

전처리
및 EDA

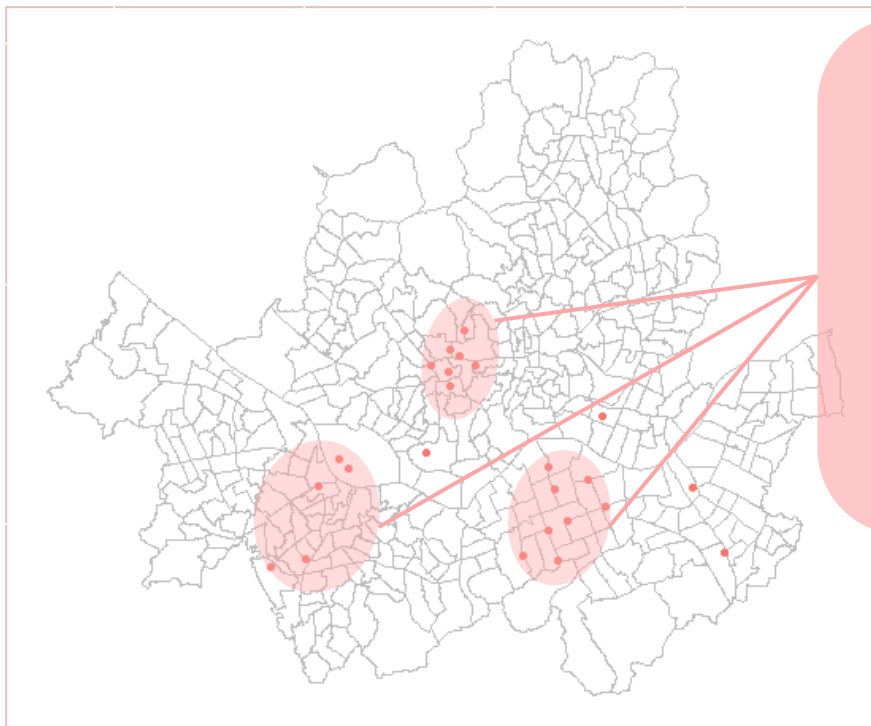
행정동 분석

다음주 예고

EDA

서울시 내 출근 시간대에 하차가 많은 지하철 시각화

“지하철 역 별 하차 TOP 25”



지하철과 연결성을 높이고
환승의 용이성을 높이려는
다람쥐 버스의 취지를 고려하면,
신규 노선 제안 시
이 지역을 염두 해야겠군!



주로 강남구 · 종로구 · 영등포구에 하차인원이 많은 지하철 역 집중됨

주제선정

전처리
및 EDA

행정동 분석

다음주 예고

사용 데이터 소개

교육통계서비스(KESS)의 학교/학과별 데이터셋 (2020년)

학제, 학교명, 시도, 시군구, 본분교, 주야, 주소, 학위과정구분, 입학지원자/입학자 수, 재적학생(재학생, 휴학생) 수, 교직원 수 등

“ 2020 고등교육기관(대학) 학교별 입학정원
입학지원 재적 재학 휴학 외국인유학생 졸업 교원.csv ”

1900 obs x 78var

학제	학교명	주소	...	입학자전체계	재적생전체계	재학생전체계
대학교	강릉원주대학교	강원도 강릉시 죽헌길 7	...	1,444	7,989	5,775
대학교	강릉원주대학교	강원도 원주시 흥업면 남원로 150	...	510	2,873	2,061
...
대학교	성균관대학교	서울특별시 종로구 성균관로 25-2	...	4,087	25,916	19,310
일반대학원	성균관대학교 일반대학원	서울특별시 종로구 성균관로 25-2	...	2,738	5,412	4,946
특수대학원	성균관대학교 교육대학원	서울특별시 종로구 성균관로 25-2	...	55	150	136
특수대학원	성균관대학교 디자인대학원	서울특별시 종로구 성균관로 25-2	...	16	35	22
...

주제선정

전처리
및 EDA

행정동 분석

다음주 예고

데이터 전처리

서울시 내 대학교 관측치만 547개...



대학원 관측치도 포함되어 있으므로 학제 구분을 기준으로 **대학교 및 전문대학(2,3,4년제)**만 추출

포함

학제	학교명
교육대학	서울교육대학교
전공대학	국제예술대학교
전공대학	백석예술대학교
전공대학	정화예술대학교
전공대학	정화예술대학교 (제2캠퍼스)
각종대학(대학)	한국예술종합학교
기능대학	한국폴리텍대학교 (정수캠퍼스)
기능대학	한국폴리텍대학교 (강서캠퍼스)

제외

학제	학교명
기술대학	정석대학(한진그룹 사내대학)
사내대학(대학)	KDB금융대학교(...?)
사내대학(대학)	SPC식품과학대학
방송통신대학	한국방송통신대학교
사이버대학(대학)	서울사이버대학교
사이버대학(대학)	세종사이버대학교
사이버대학(대학)	경희사이버대학교
...	...

학제로 구분한 결과 총 59개 대학교 추출 (주야 구분 모두 주간 대학)



주제선정



전처리
및 EDA



행정동 분석



다음주 예고

데이터 전처리

카카오 API 주소 검색으로 학교별 행정동, 위도, 경도 추출



주제선정



전처리
및 EDA



행정동 분석



다음주 예고

학교명	주소
서울대학교	서울특별시 관악구 관악로 1
한국체육대학교	서울특별시 송파구 양재대로 1239
서울과학기술대학교	서울특별시 노원구 공릉로 232
서울시립대학교	서울특별시 동대문구 서울시립대로 163
가톨릭대학교	서울특별시 서초구 반포대로 222
가톨릭대학교	서울특별시 종로구 창경궁로 296-12
감리교신학대학교	서울특별시 서대문구 독립문로 56
...	...

카카오지도 API 주소 검색



행정동	위도	경도
대학동	37.46804	126.9593
오륜동	37.51990	127.1320
공릉2동	37.63308	127.0768
전농2동	37.58258	127.0600
반포4동	37.50239	127.0059
해화동	37.58490	127.0048
천연동	37.56792	126.9619
...

데이터 전처리

행정동별로 그룹화하여 대학교 수와 대학교 재학생 수 계산



주제선정



전처리
및 EDA



행정동 분석



다음주 예고

학교명	주소	행정동
서울대학교	서울특별시 관악구 관악로 1	대학동
한국체육대학교	서울특별시 송파구 양재대로 1239	오륜동
서울과학기술대학교	서울특별시 노원구 공릉로 232	공릉2동
서울시립대학교	서울특별시 동대문구 서울시립대로 163	전농2동
가톨릭대학교	서울특별시 서초구 반포대로 222	반포4동
가톨릭대학교	서울특별시 종로구 창경궁로 296-12	혜화동
감리교신학대학교	서울특별시 서대문구 독립문로 56	천연동
...

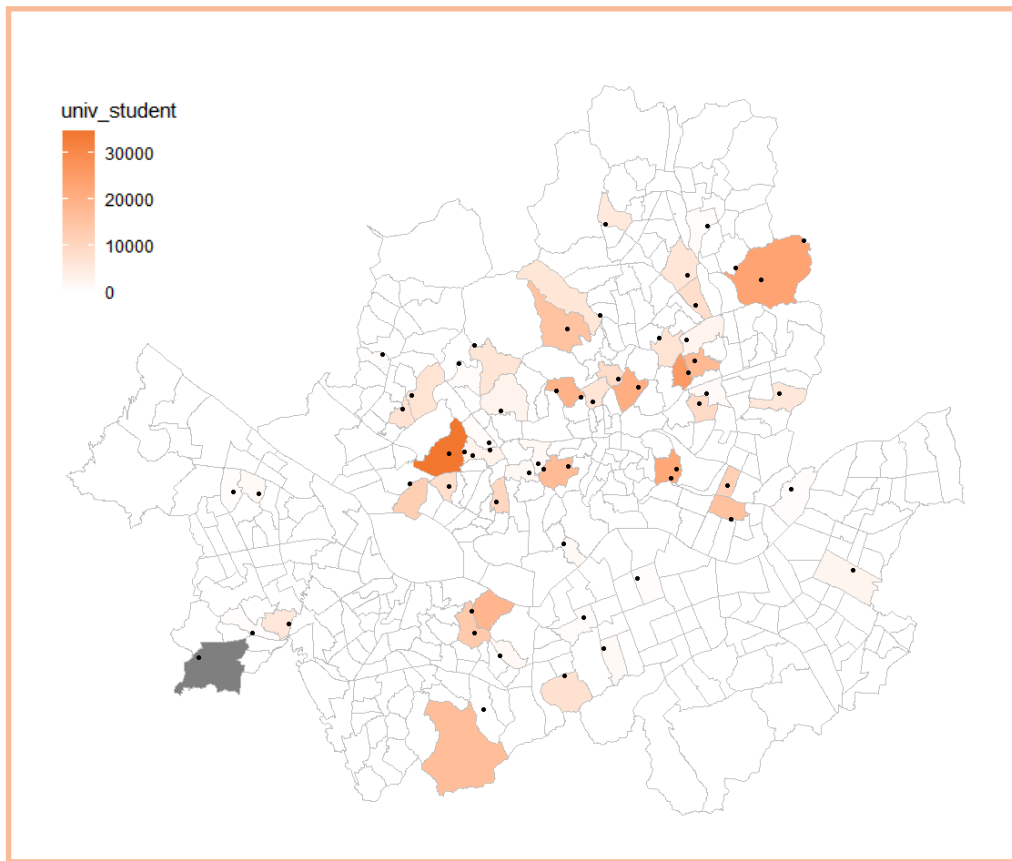
행정동별로
그룹화



행정동	대학교 수	재적학생 수
사직동	0	0
삼청동	0	0
부암동	1	6544
...
혜화동	2	19541
소공동	0	0
회현동	1	1435
...

EDA

서울시내 대학교 분포 지도



행정동	대학교 수	대학재학생 수
신촌동	2	34,622
회기동	1	25,875
공릉2동	3	23,245
사근동	2	22,389
안암동	1	20,822
혜화동	1	19,541
흑석동	1	18,747
이문1동	1	17,438
필동	2	17,340



대학교 수가 가장 많은 공릉2동보다 재학생 수가 많은 **신촌동·회기동** 통학 유동인구 높을 것으로 예상됨



주제선정

전처리
및 EDA

행정동 분석



다음주 예고

사용 데이터 소개

서울특별시 교육청 열린서울교육정보 연도별 학교현황 2020년 상반기 학교일람표

소재구, 학교명, 남녀, 주야, 주소, 학급수, 학생수(학년별) 등

“200040기준_학교일람표_고등학교.csv”

320 obs x 43var

소재구	학교명	설립	남녀	주소	학급수 계	학생수 계	...
종로구	경기상업고등학교	공립	남녀공학	서울특별시 종로구 자하문로 136	30	284	...
종로구	경북고등학교	공립	남자	서울특별시 종로구 자하문로 28가길 9	35	726	...
종로구	경신고등학교	사립	남자	서울특별시 종로구 혜화로 74	26	565	...
종로구	대신고등학교	사립	남자	서울특별시 종로구 계동길 84-10	29	725	...
종로구	덕성여자고등학교	사립	여자	서울특별시 종로구 사직로 9	26	584	...
종로구	동성고등학교	사립	남자	서울특별시 종로구 율곡로 3길 50	27	637	...
종로구	배화여자고등학교	사립	여자	서울특별시 종로구 대학로 156	29	791	...
...

주제선정

전처리
및 EDA

행정동 분석

다음주 예고

데이터 전처리

카카오 API 주소 검색으로 학교별 행정동, 위도, 경도 추출
& 행정동 별로 그룹화하여 행정동 별 고등학교 수, 고등학생 수 계산



주제선정



전처리
및 EDA



행정동 분석



다음주 예고

학교명	주소	행정동
경기상업고등학교	서울특별시 종로구 자하문로 136	사직동
경북고등학교	서울특별시 종로구 자하문로 28가길 9	삼청동
경신고등학교	서울특별시 종로구 혜화로 74	부암동
대신고등학교	서울특별시 종로구 사직로 9	...
덕성여자고등학교	서울특별시 종로구 율곡로 3길 50	혜화동
동성고등학교	서울특별시 종로구 대학로 156	소공동
배화여자고등학교	서울특별시 종로구 필운대로1길 34	회현동
...

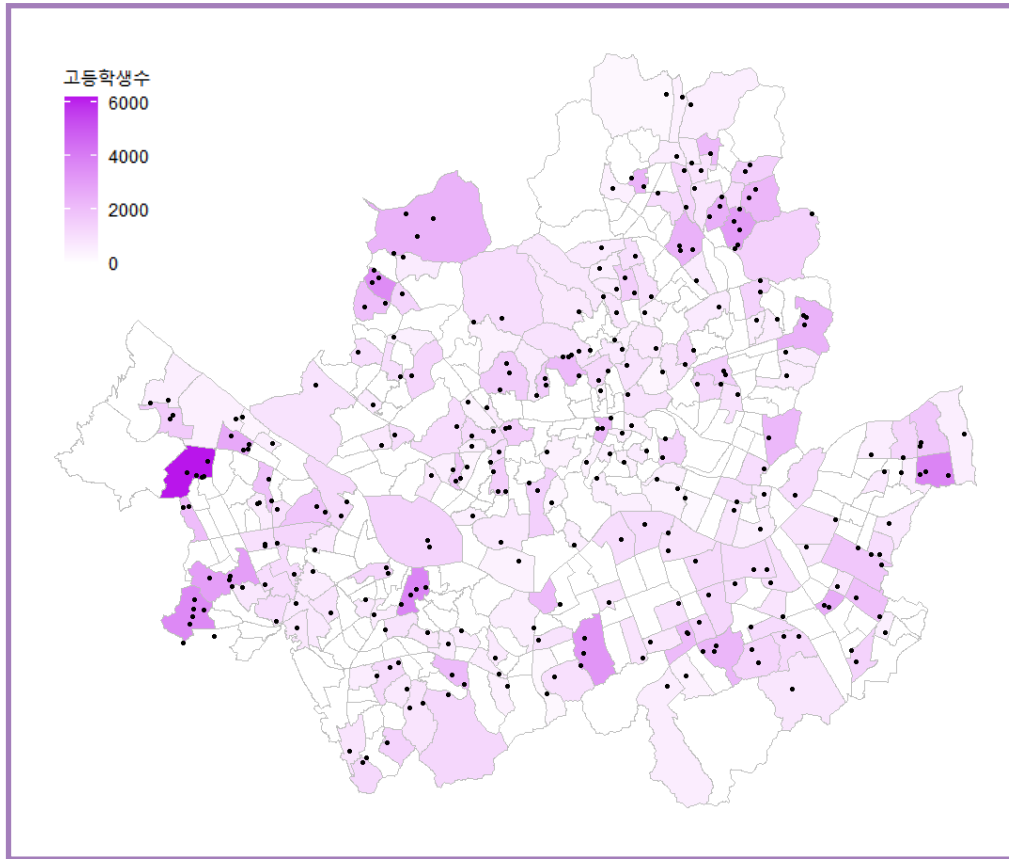
행정동별로
그룹화



행정동	고등학교수	고등학생수
사직동	0	0
삼청동	1	637
부암동	1	429
...
혜화동	4	2,183
소공동	2	1,563
회현동	0	0
...

EDA

서울시내 고등학교 분포 지도



소재구	행정동	고등학교 수	고등학생 수
강서구	발산1동	8	6,198
강동구	상일동	5	3,802
동작구	대방동	4	3,769
구로구	수궁동	6	3,658
은평구	갈현2동	5	3,568
서초구	서초3동	3	3,319
노원구	하계1동	3	3,116
양천구	신정3동	4	3,040
강서구	등촌3동	3	2,719



서울시내에 고르게 분포하는 편임

주제선정

전처리
및 EDA

행정동 분석

다음주 예고

사용 데이터 소개

서울시 사업체 현황을 구·동별로 제공하는 지정·조사통계 (2018년)

“서울시 사업체 및 종사자 밀도(동별) 통계.csv”

424 obs x 8 var

자치구	행정동	사업체수	종사자수	사업체 밀도	종사자 밀도	인구밀도	종사자 밀도비
종로구	가회동	766	14,100	1,419	26,111	8,419	310.16
종로구	종로1.2.3.4가동	15,603	109,824	6,640	46,734	3,601	1,297.85
종로구	종로5·6가동	7,461	24,713	12,435	41,188	9,328	441.54
종로구	교남동	393	4,008	1,123	11,451	30,989	36.95
...
강동구	암사1동	1,916	5,756	1,878	5,643	37,020	15.24
강동구	천호2동	3,736	14,933	2,380	9,511	23,199	41
강동구	길동	5,172	19,657	2,383	9,059	21,738	41.67

* 사업체 : 영리·비영리를 불문하고 개개의 상점, 사무소, 영업소, 은행, 학교, 병원, 여관, 식당, 학원, 교회, 사찰, 공공기관, 사회복지시설 등과 같이 일정한 장소에서 단일소유권 또는 단일통제 하에 재화의 생산·판매, 서비스 제공 등 산업활동을 영위하고 있는 모든 경영단위

* 사업체밀도 = 사업체(개)/총 면적(km²), 종사자밀도 = 종사자(수)/총 면적(km²), 인구밀도 = 인구(수)/총 면적(km²), 종사자밀도비 = 종사자밀도/인구밀도*100

주제선정

전처리
및 EDA

행정동 분석

다음주 예고

데이터 전처리

행정동코드와 join 하여 id column 추가

주제선정

전처리
및 EDA

행정동 분석

다음주 예고



행정동코드	시군구	읍면동
1101060	가회동	가회동
1101061	종로1·2·3·4가동	종로1·2·3·4가동
1101063	종로5·6가동	종로5·6가동
...
1125073	관악구	신사동
1125074	강남구	신사동
...



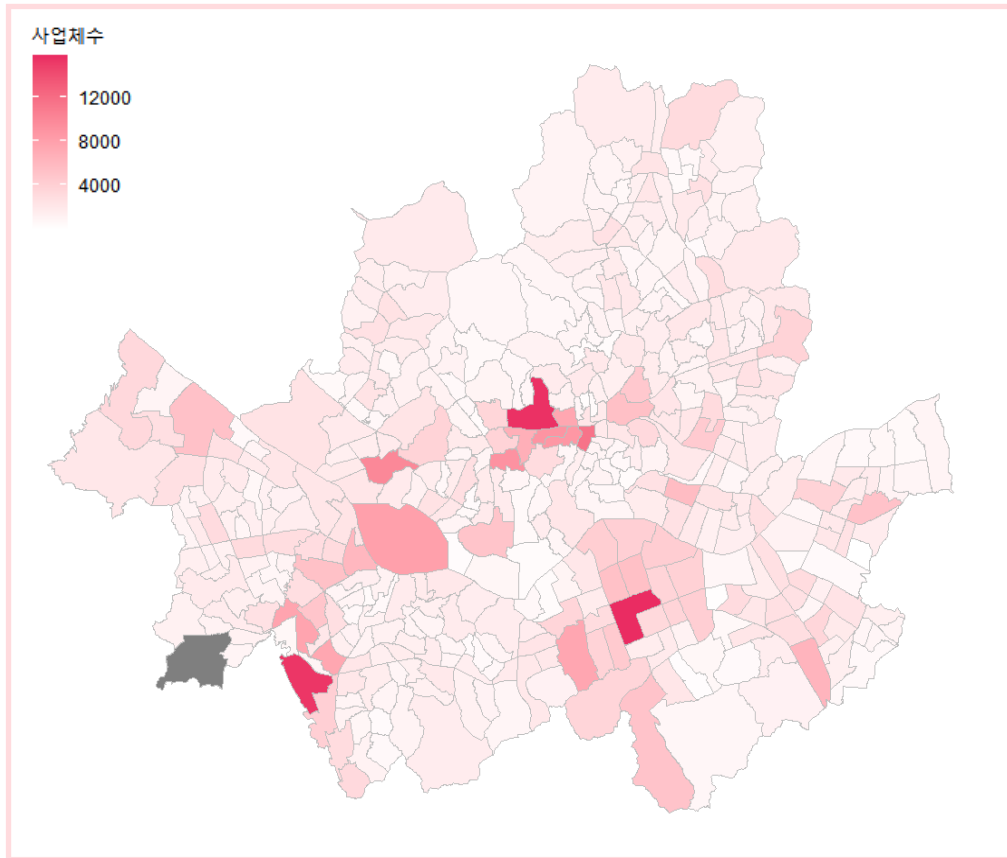
자치구	행정동	id	사업체수	종사자수	...
종로구	가회동	1101060	766	14,100	...
종로구	종로1.2.3.4가동	1101061	15,603	109,824	...
종로구	종로5·6가동	1101063	7,461	24,713	...
종로구	교남동	1101058	393	4,008	...
...
강동구	암사1동	1125072	1,916	5,756	...
강동구	천호2동	1125073	3,736	14,933	...
강동구	길동	1125074	5,172	19,657	...

서울시에 신사동이 2군데 (강남구와 관악구) 있기 때문에...
행정동과 자치구 모두 key 값으로 두고 join

id 컬럼 추가!

EDA

행정동 별 사업체 수 지도 시각화



자치구	행정동	사업체 수
강남구	역삼1동	15,787
종로구	종로1·2·3·4가동	15,603
금천구	가산동	15,347
중구	신당동	11,568
마포구	서교동	9,958
중구	회현동	9,122
중구	을지로동	9,028
중구	광희동	8,744
영등포구	여의동	8032

 **역삼1동, 종로1·2·3·4가동, 가산동**에 사업체 수 압도적

중구는 여러 행정동에 사업체가 많이 분포함

주제선정

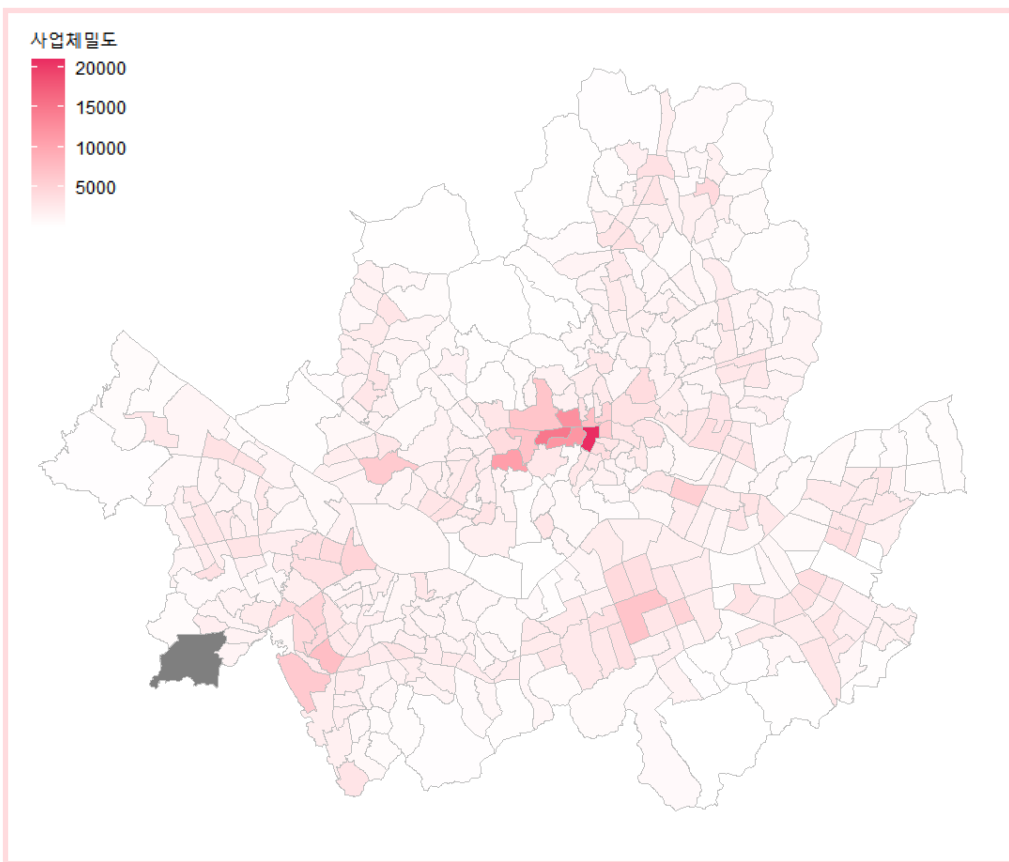
전처리
및 EDA

행정동 분석

다음주 예고

EDA

행정동 별 사업체 밀도 지도 시각화



자치구	행정동	사업체 밀도
중구	신당동	21,033
중구	을지로동	15,047
종로구	종로5·6가동	12,435
중구	광희동	11,816
중구	회현동	10,860
구로구	구로3동	7,320
종로구	창신1동	7,287
중구	명동	6,833
강남구	역삼1동	6,718



면적 대비 사업체 수가 가장 많은 곳은 **중구, 종로구**
신당동, 을지로동, 광희동 부근에 많은 사업체 밀집됨



주제선정

전처리
및 EDA

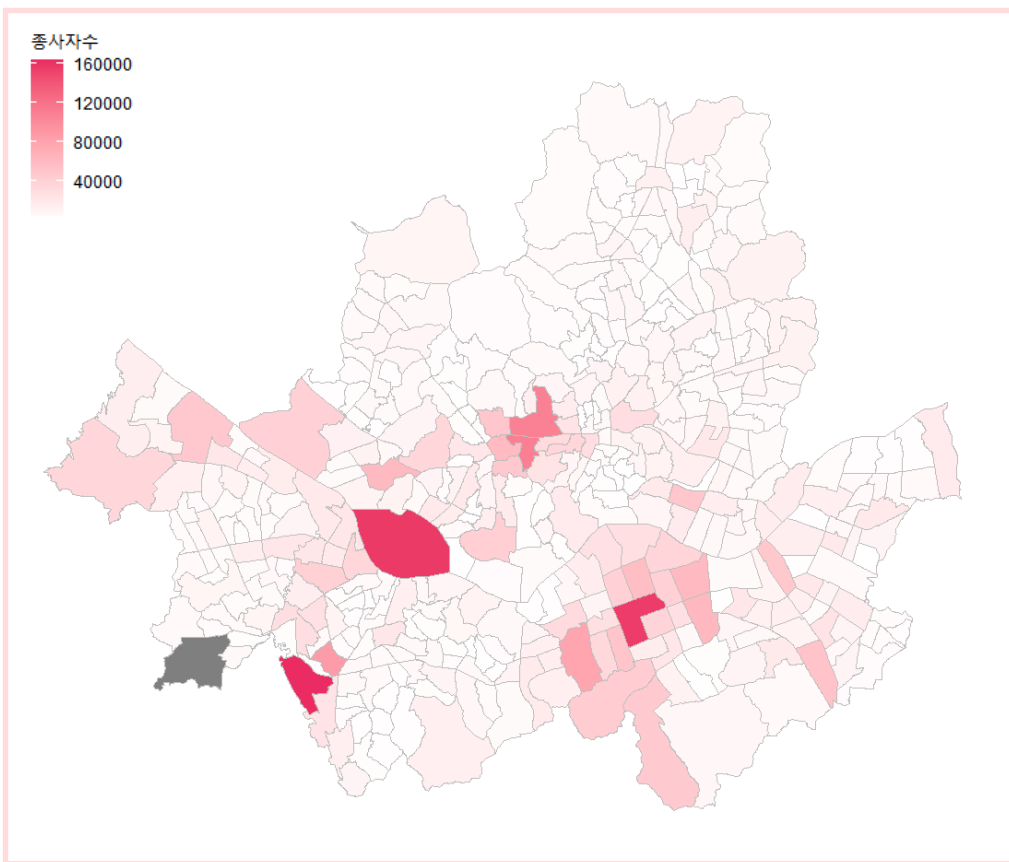
행정동 분석



다음주 예고

EDA

행정동 별 종사자 수 지도 시각화



자치구	행정동	종사자 수
금천구	가산동	164,278
영등포구	여의동	157,954
강남구	역삼1동	156,700
중구	명동	110,677
종로구	종로1·2·3·4가동	109,824
구로구	구로3동	87,533
서초구	서초3동	78,942
강남구	대치2동	62,612
마포구	서교동	62,184

대체로 사업체 수 높은 동과 비슷하지만 **가산디지털단지**와 **여의도**에 종사자가 많이 분포함

주제선정

전처리
및 EDA

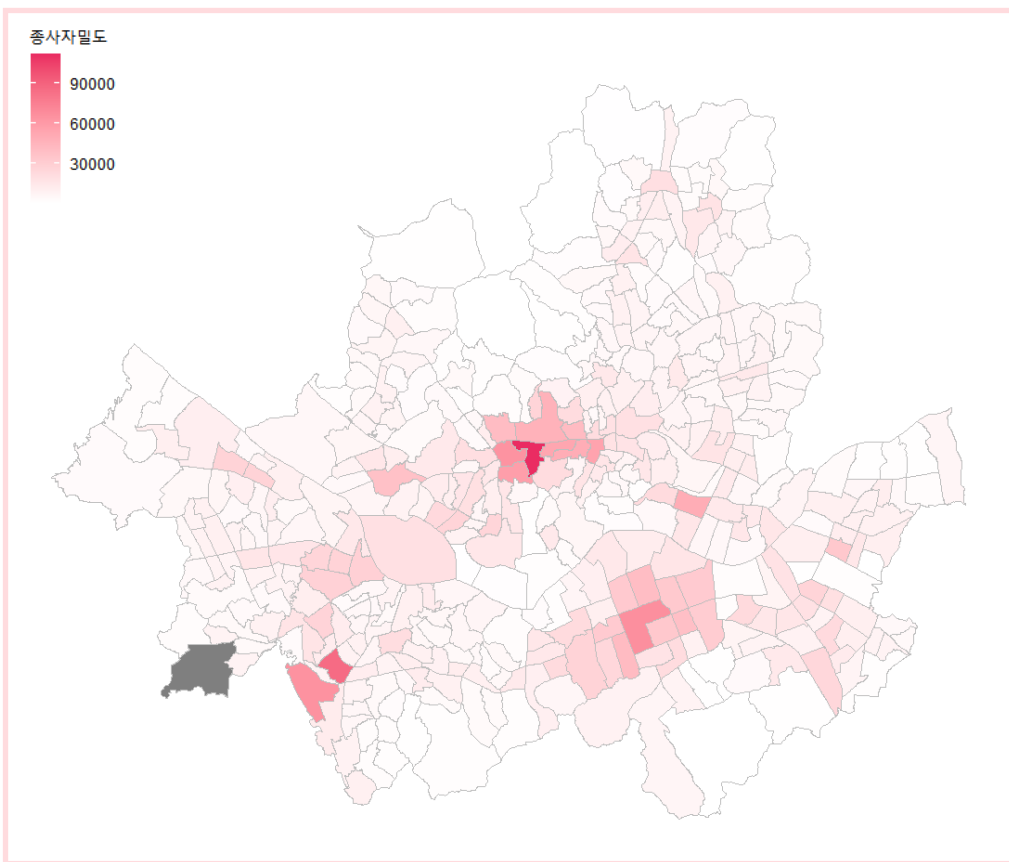
행정동 분석



다음주 예고

EDA

행정동 별 종사자 밀도 지도 시각화



자치구	행정동	종사자 밀도
중구	명동	111,795
구로구	구로3동	85,817
강남구	역삼1동	66,681
금천구	가산동	65,190
중구	소공동	64,200
중구	회현동	58,052
중구	신당동	55,465
중구	을지로동	53,485
중구	광희동	49,641

중구에 종사자 밀도 높은 행정동 많이 분포



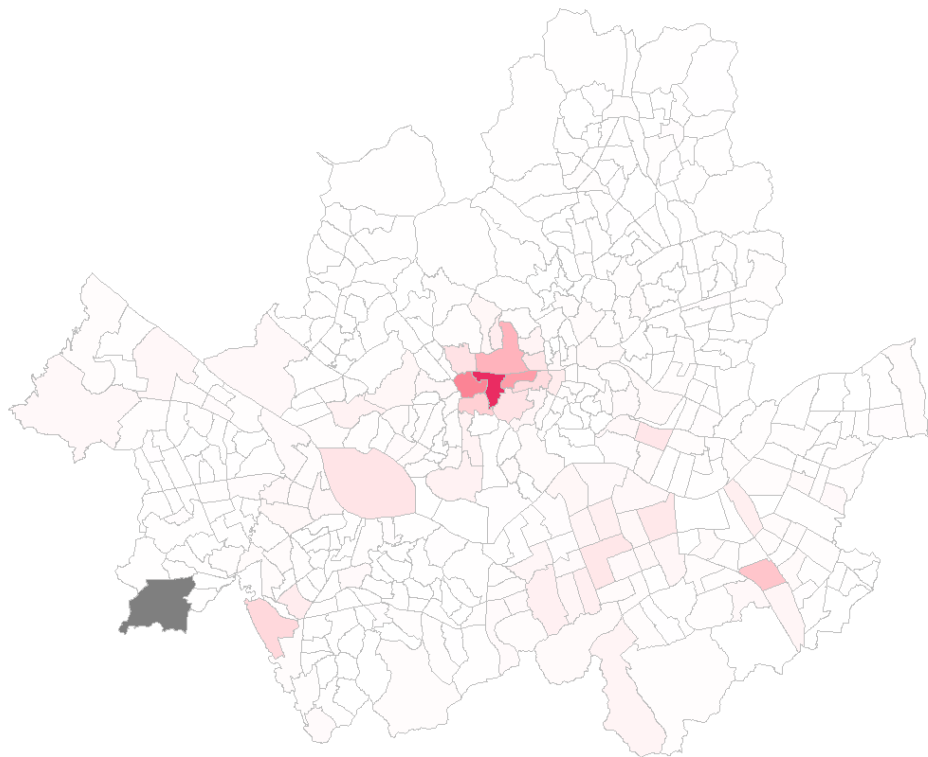
면적 대비 종사자 수는 명동과 구로3동이 가장 높음



EDA

행정동 별 종사자 밀도비 지도 시각화

* 종사자밀도비 = 종사자밀도/인구밀도*100



자치구	행정동	종사자 밀도비
중구	명동	3189.54
중구	소공동	2077.32
중구	을지로동	1688.11
종로구	종로1·2·3·4가동	1297.85
송파구	가락1동	995.70
중구	회현동	819.29
금천구	가산동	671.65
중구	광희동	606.47
종로구	사직동	504.54

인구밀도 대비 종사자 인구밀도는 **명동**이 압도적 높음**중구 · 종로구**의 종사자밀도비가 가장 높음

주제선정

전처리
및 EDA

행정동 분석

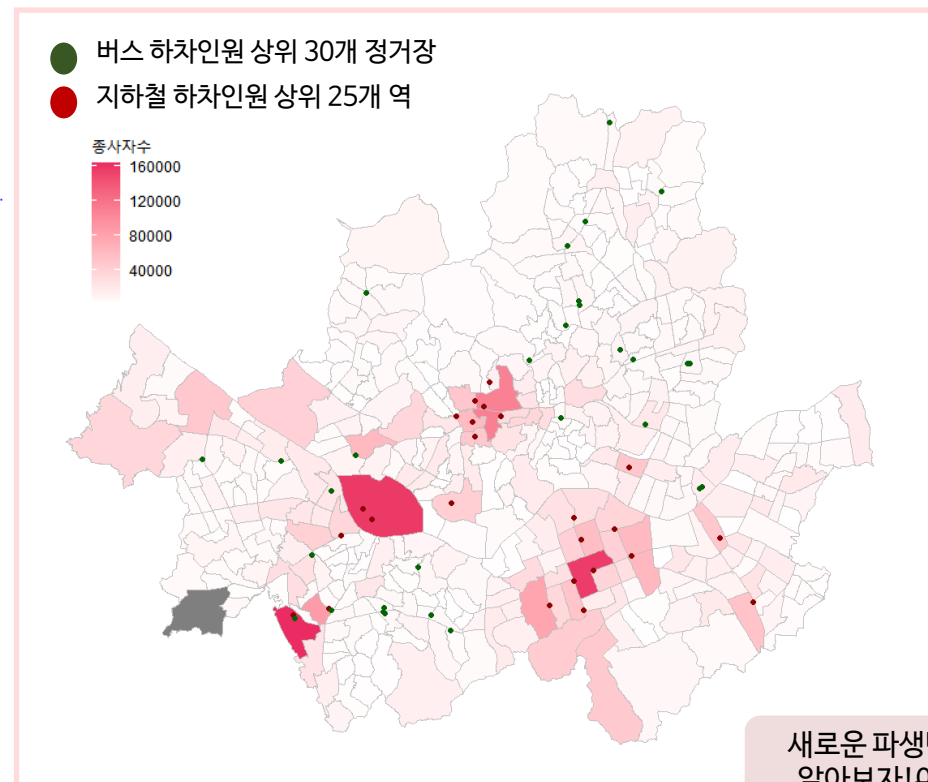
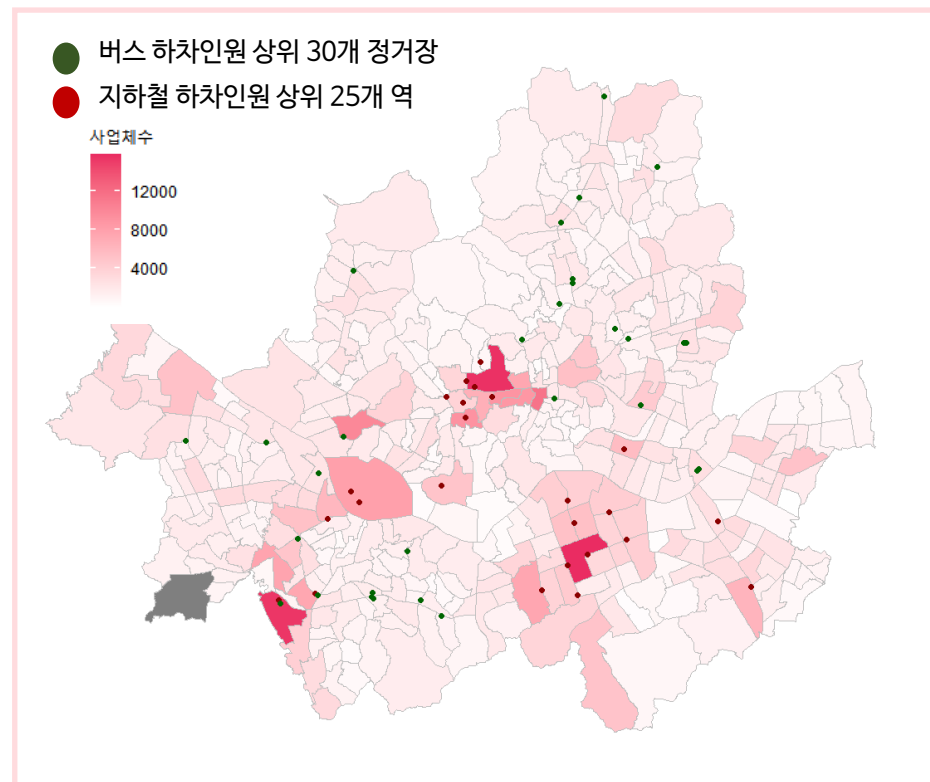


다음주 예고



EDA

행정동 별 사업체 · 종사자 수 지도 + 버스·지하철 하차 인원 수 높은 역·정거장 분포



새로운 파생변수로
알아보자! 아마도

지하철 하차 인원 수 높은 역은 사업체 · 종사자 수 높은 강남구, 종로구, 영등포구에 분포

버스 하차 인원 수 높은 정거장은 사업체 분포와 직접적 연관성 없어 보임... 버스-지하철 환승구간 예상됨



사용 데이터 소개

서울시 특정시점에 서울의 특정 지역(행정동)에 존재하는 모든 인구수를 일 · 시간대 · 나잇대별로 제공하는 데이터

: 202001~202010까지 10개월치 데이터 사용

“LOCAL_PEOPLE_DONG_202001.csv”

315456 obs x 32 var

기준일ID	시간대구분	행정동코드	총생활인구수	남자0세부터9세 생활인구수	남자10세부터 14세생활인구 수	...	여자70세이상 생활인구수
20200101	0	11620625	24188.58	870.1487	439.6648	...	1501.712
20200101	0	11140605	6716.533	89.0565	36.2822	...	303.8293
20200101	0	11470640	46603.62	2164.969	1254.327	...	3226.197
20200101	0	11290780	9589.181	390.6835	244.4328	...	697.1448
...
20200131	23	11590605	34441.96	1054.075	489.391	...	2785.423
20200131	23	11215760	17221.93	535.6532	264.8733	...	1155.236
20200131	23	11200580	15466.15	629.6817	386.2273	...	1376.087



주제선정



전처리
및 EDA



행정동 분석



다음주 예고



데이터 전처리

남녀 구분된 인구수를 하나로 통합하고, 다람쥐버스 운행시간대(7시~9시)만 추출



주제선정



전처리
및 EDA



행정동 분석



다음주 예고

“LOCAL_PEOPLE_DONG_202001.csv”

315456 obs x 32 var

기준일ID	시간대구분	행정동코드	총생활인구수	남자0세부터9세 생활인구수	남자10세부터 14세생활인구 수	...	여자70세이상 생활인구수
20200101	0	11620	3.58	870.1487	439.6648	...	501.712
20200101	0	11140	533	89.0565	36.2822	...	3.8293
20200101	0	11470	3.62	2164.969			
20200101	0	11290	181	390.6835	00_09	10_14	...
...
20200131	23	11590	1.96	1054.075	489.391	...	2785.423
20200131	23	11215760	17221.93	535.6532	264.8733	...	1155.236
20200131	23	11200580	15466.15	629.6817	386.2273	...	1376.087



데이터 전처리

7자리 행정동 코드(join의 기준) id 변수 추가, 행정동별 면적 변수 추가



주제선정



전처리
및 EDA



행정동 분석



다음주 예고

LOCAL_PEOPLE_DONG_202001.csv

기준일ID	시간대구분	행정동코드	...	id	면적
20200101	7	11440610	...	1114061	0.43
20200101	7	11680655	...	1123066	1.02
20200101	7	11230536	...	1106081	1.61
20200101	7	11440730	...	1114073	2.07
...
20200131	9	11305660	...	1109075	3.28
20200131	9	11215870	...	1105062	1.02
20200131	9	11290685	...	1108068	0.58

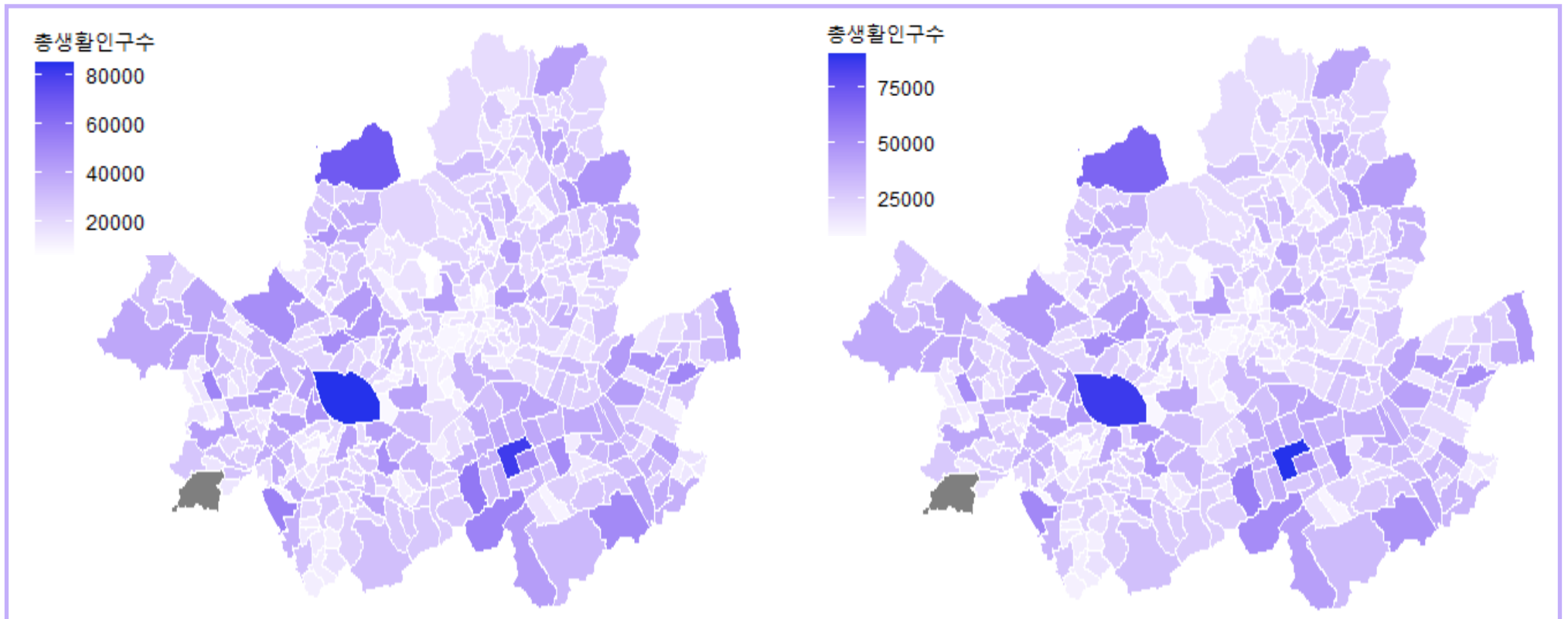
추가!



EDA

월요일 7시 행정동별 총생활인구수 2020년 7월 2주차, 10월 2주차 시각화

주별, 월별 차이는 거의 없다고 보임



주제선정



전처리
및 EDA



행정동 분석



다음주 예고



EDA

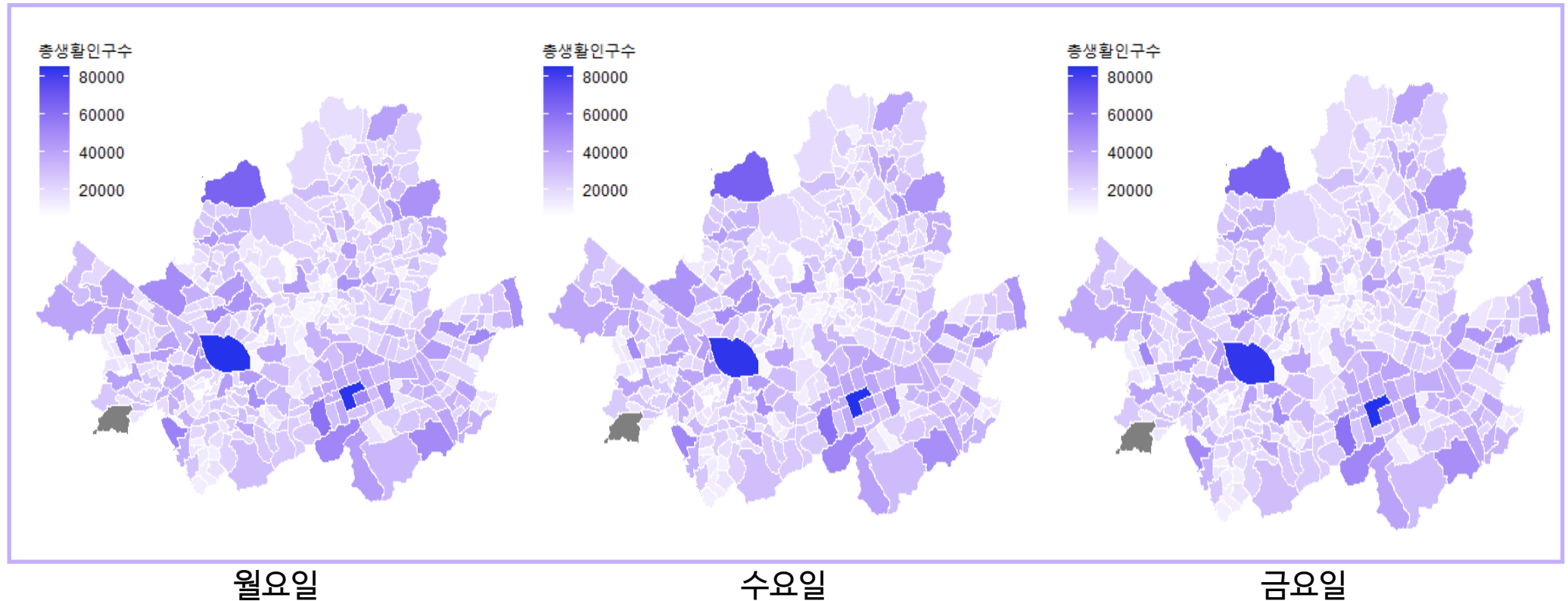
월-수-금 7시 행정동별 총생활인구수 10월 3주차 시각화
요일별 차이는 거의 없다고 보임 복불아님 다 다른 그래프

주제선정

전처리
및 EDA

행정동 분석

다음주 예고





EDA

월-수-금 7시 행정동별 총생활인구수 10월 3주차 시각화
요일별 차이는 거의 없다고 보임 복불아님 다 다른 그래프



주제선정



전처리
및 EDA



행정동 분석



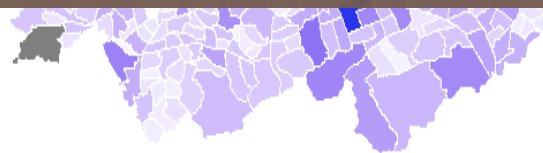
다음주 예고

총생활인구수
80000
60000
40000
20000

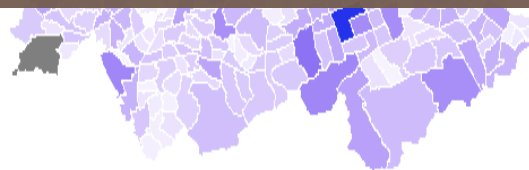
총생활인구수
80000
60000
40000
20000

총생활인구수
80000
60000
40000
20000

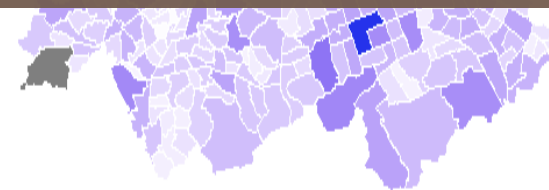
요일별, 주별, 월별 차이가 없다고 판단해
20200101~20201031 데이터를 2020년 일평균으로 전처리!



월요일



수요일

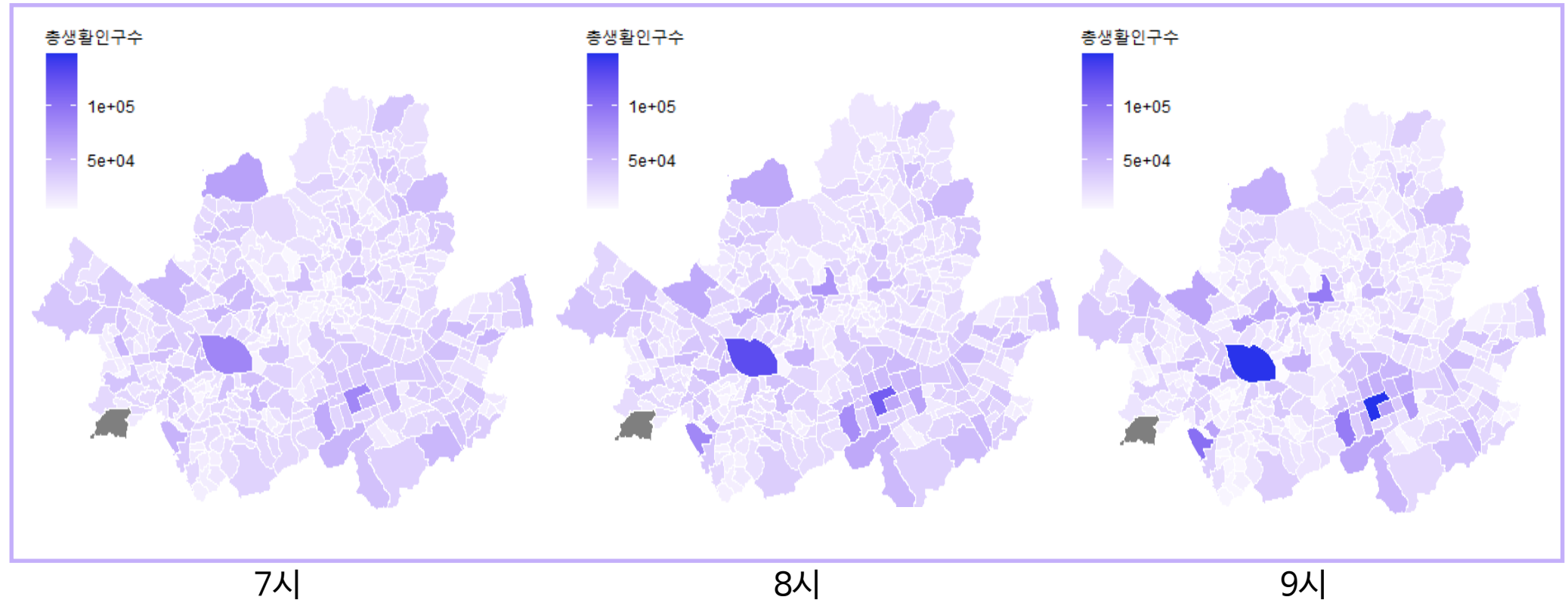


금요일



EDA

7-8-9시 행정동별 총생활인구수 시각화(gradient 범위 맞춰주기)



주제선정

전처리
및 EDA

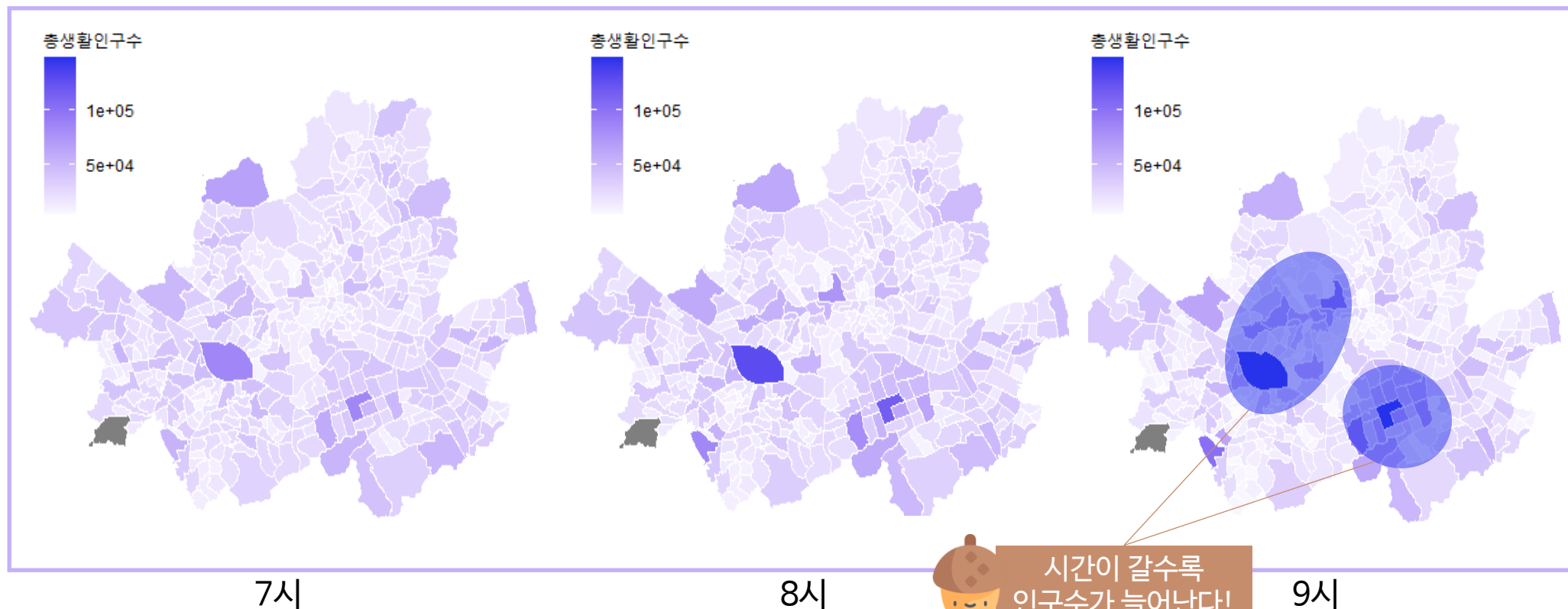
행정동 분석

다음주 예고



EDA

7-8-9시 행정동별 총생활인구수 시각화(gradient 범위 맞춰주기)



주제선정

전처리
및 EDA

행정동 분석

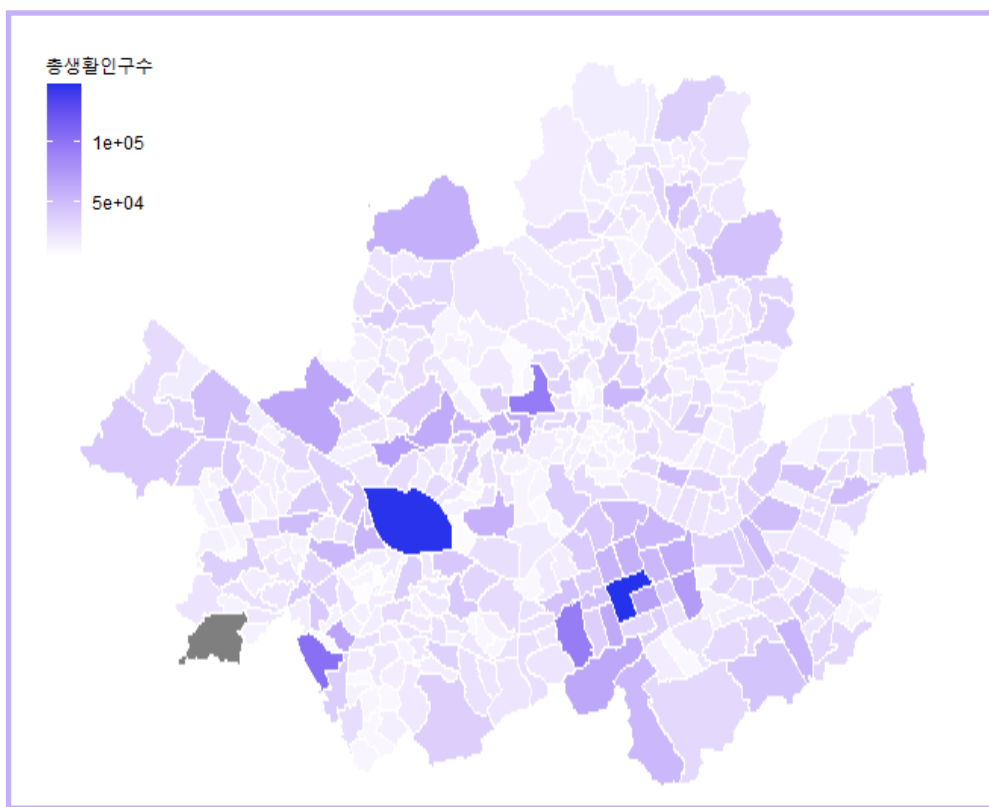


다음주 예고



EDA

행정동별 총생활인구수 9시 시각화, 총생활인구수 상위 9개 행정동



자치구	행정동	총생활인구수
강남구	역삼1동	147937.3
영등포구	여의동	147204
금천구	가산동	103948.3
종로구	종로1·2·3·4가동	95248.83
서초구	서초3동	94332.95
강남구	대치2동	72205.46
마포구	서교동	68948.52
강남구	역삼2동	68406.4
구로구	구로3동	66199.71



주제선정

전처리
및 EDA

행정동 분석



다음주 예고



EDA

행정동별 총생활인구수 9시 시각화, 행정동별 종사자 수 시각화



주제선정



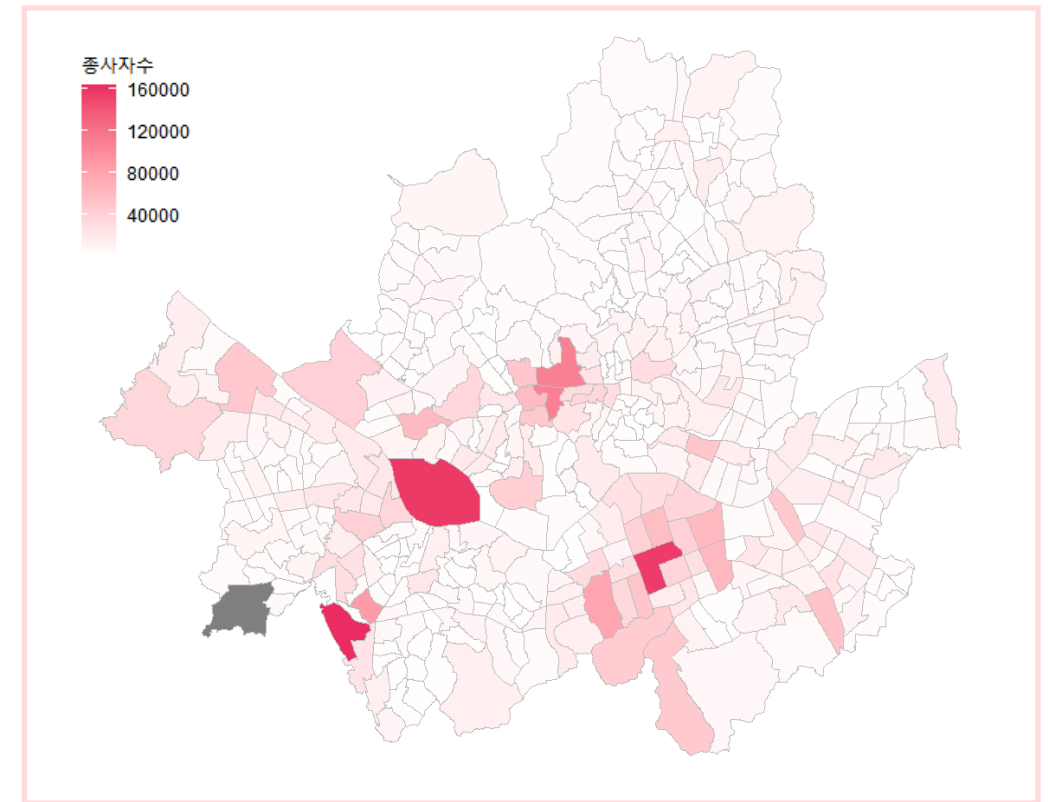
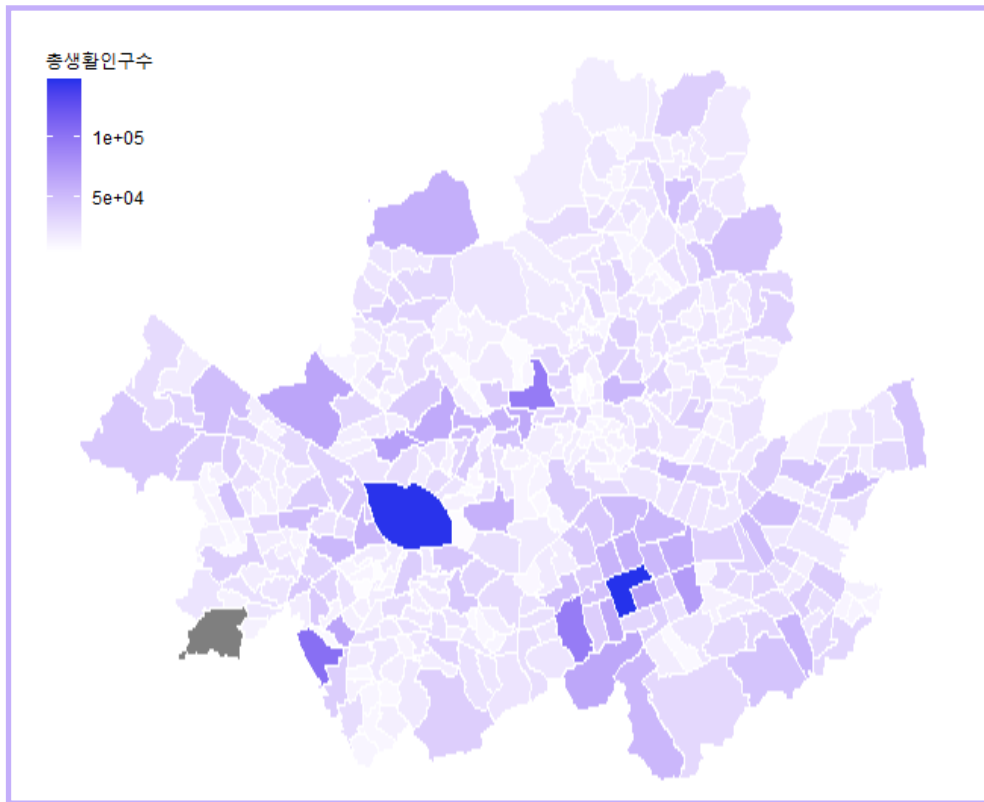
전처리
및 EDA



행정동 분석



다음주 예고



거의 흡사함



EDA

종사자 수 상위 9개 행정동, 총생활인구수 상위 9개 행정동
명동, 역삼2동을 제외하고 모두 겹침



주제선정

전처리
및 EDA

행정동 분석



다음주 예고

자치구	행정동	총생활인구수
강남구	역삼1동	147937.3
영등포구	여의동	147204
금천구	가산동	103948.3
종로구	종로1·2·3·4가동	95248.83
서초구	서초3동	94332.95
강남구	대치2동	72205.46
마포구	서교동	68948.52
강남구	역삼2동	68406.4
구로구	구로3동	66199.71

자치구	행정동	종사자 수
금천구	가산동	164,278
영등포구	여의동	157,954
강남구	역삼1동	156,700
중구	명동	110,677
종로구	종로1·2·3·4가동	109,824
구로구	구로3동	87,533
서초구	서초3동	78,942
강남구	대치2동	62,612
마포구	서교동	62,184



EDA

종사자 수 상위 9개 행정동, 총생활인구수 상위 9개 행정동
명동, 역삼2동을 제외하고 모두 겹침



주제선정

전처리
및 EDA

행정동 분석



다음주 예고

구	행정동	총생활인구수	구	행정동	총생활인구수
강남구	역삼1동	147937.3	금천구	가산동	164,278
영등포구	영등포1동	147,894	영등포구	영등포1동	157,954
금천구	가산동	103948.3	강남구	역삼1동	156,700
종로구	종로1·2·3·4가동	109,677	종로구	종로1·2·3·4가동	109,824
서초구	서초3동	94332.95	구로구	구로3동	87,533
강남구	대치2동	72205.46	서초구	서초3동	78,942
마포구	서교동	68948.52	강남구	대치2동	62,612
강남구	역삼2동	68406.4	마포구	서교동	62,184
구로구	구로3동	66199.71			

직장인은 보통 9시까지 출근이므로, 사람들이
7시~ 9시 사이 해당 지역으로 출근했다고 예측할 수 있다.



EDA

종사자 수 상위 9개 행정동, 총생활인구수 상위 9개 행정동
명동, 역삼2동을 제외하고 모두 견침

행정동별 7시-9시 생활인구 차이를 본다면
어느 지역에서 어느 지역으로 출근했는지
추측할 수 있지 않을까?

7시-9시 사이 해당 지역으로 출근했다고 예측할 수 있다.

강남구	대치2동	72205.40	구로구	구로3동	66199.71
마포구	서교동	68948.52	서초구	서초3동	78,942
강남구	역삼2동	68406.4	강남구	대치2동	62,612
구로구	구로3동	66199.71	마포구	서교동	62,184



주제선정

전처리
및 EDA

행정동 분석



다음주 예고



EDA

종사자 수 상위 9개 행정동, 총생활인구수 상위 9개 행정동
명동, 역삼2동을 제외하고 모두 견치



주제선정

전처리
및 EDA

행정동 분석



다음주 예고

행정동별 7시-9시 생활인구 차이를 본다면
어느 지역에서 어느 지역으로 출근했는지
추측할 수 있지 않을까?



직장인 출근시간대 7시-9시까지만 출근하므로, 사람들이
7시-9시 사이 해당 지역으로 출근했다고 예측할 수 있다.



일단은 출근시간대 전체를 고려해
시각화하기 위해 시간대도 통합하여
평균값으로 처리해보자!

강남구	대치2동	72205.40	마포구	서교동	68948.52
영등포구	영등포1동	103948.3	강남구	역삼2동	68406.4
금천구	가산동	164,278	구로구	구로3동	66199.71
강남구	역삼1동	157,954	마포구	서교동	62,184
영등포구	영등포2동	156,700			
금천구	가산동	109,824			
강남구	대치2동	87,533			
영등포구	영등포1동	78,942			
금천구	가산동	62,612			
강남구	역삼1동	62,184			



EDA

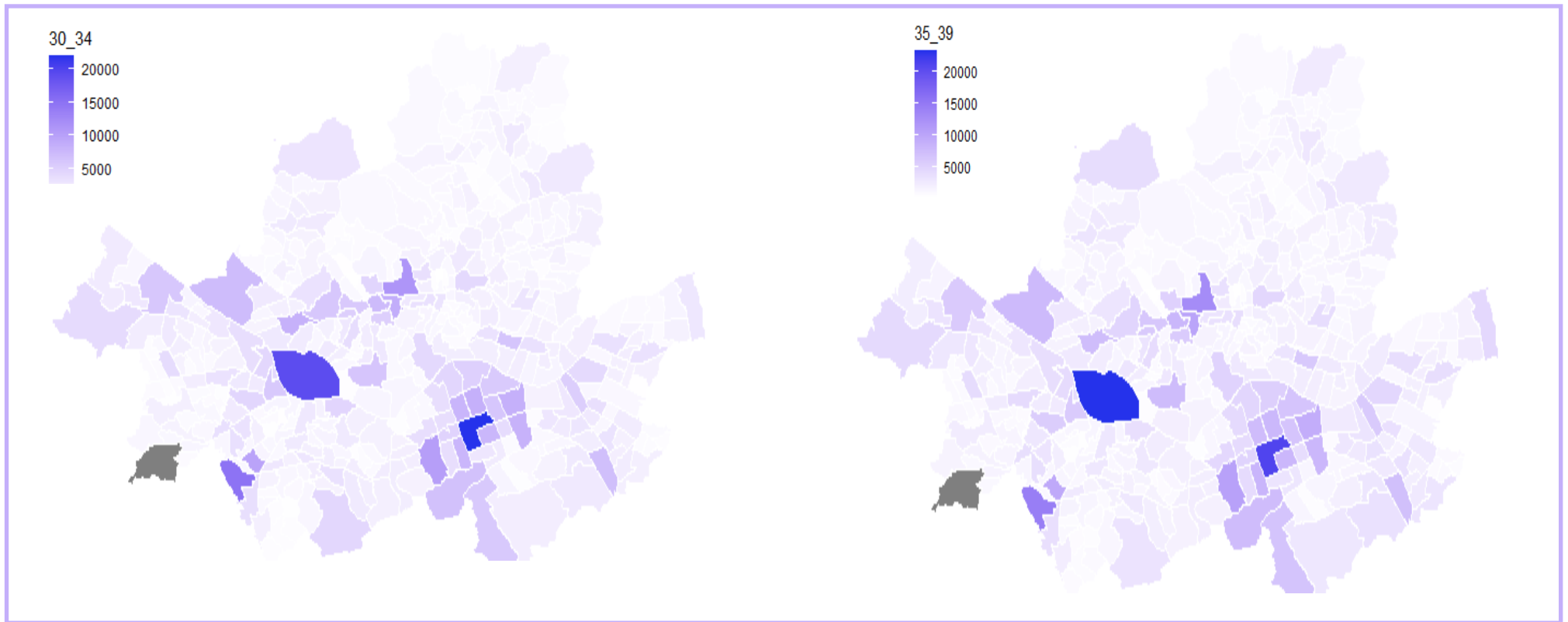
행정동별 30~34세 생활인구, 35~39세 생활인구 시각화

 주제선정

 전처리
및 EDA

 행정동 분석

 다음주 예고





데이터 전처리

같은 나잇대 시각화 양상은 비슷하게 나타나므로 14세 이하는 제외하고 나이 변수도 재범주화



주제선정

전처리
및 EDA

행정동 분석



다음주 예고

LOCAL_PEOPLE_DONG_202001.csv

기준일ID	시간대구분	행정동코드	총생활인구수	00_09	10_14	...	70_
20200101	0	11620625	24188.58	870.1487	439.6648	...	501.712
20200101	0	11140605	6716.533	89.0565	36.2822	...	13.8293
20200101	0	11470640	46603.62	2164.969	10대 20대 ...		
20200101	0	11290780	9589.181	390.6835			
...
20200131	23	11590605	34441.96	1054.075	489.391	...	2785.423
20200131	23	11215760	17221.93	535.6532	264.8733	...	1155.236
20200131	23	11200580	15466.15	629.6817	386.2273	...	1376.087



데이터 전처리

같은 나잇대 시각화 양상은 비슷하게 나타나므로 14세 이하는 제외하고 나이 변수도 재범주화



주제선정



전처리
및 EDA



행정동 분석



다음주 예고

LOCAL PEOPLE DONG 202001.csv

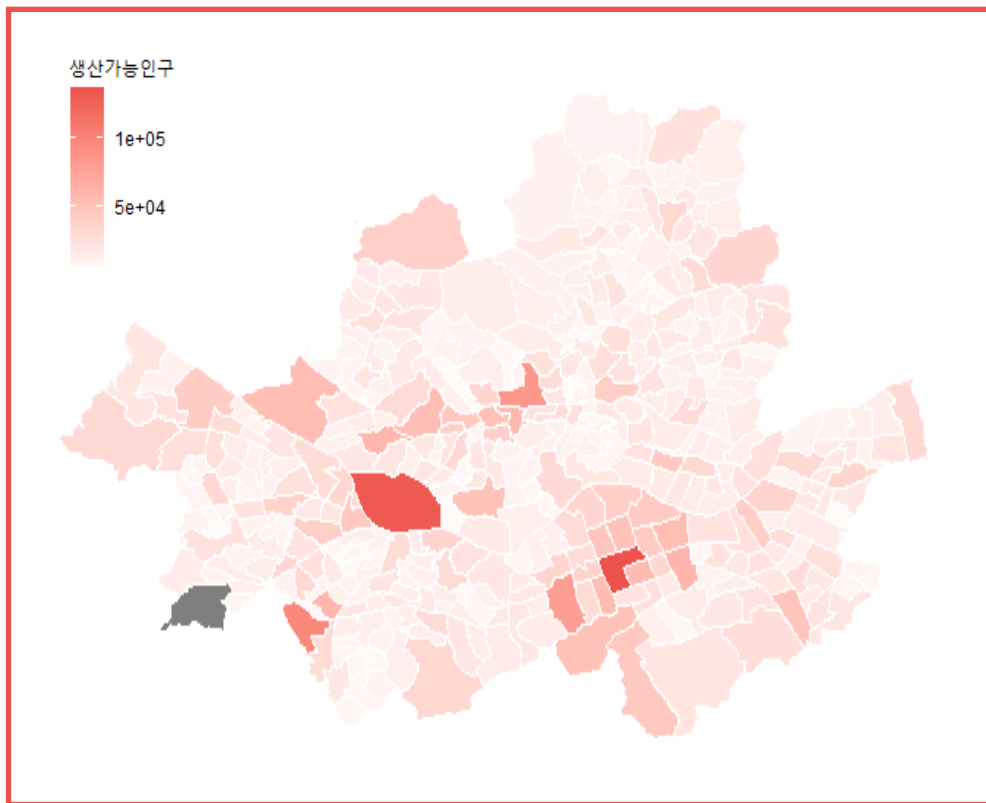
나잇대별 차이를 보고 싶어 재범주화했지만...
양상에 큰 변화가 없어 생산가능인구/외 인구로 통합!

20200131	23	11590605	34441.96	1054.075	489.391	...	2785.423
20200131	23	11215760	17221.93	535.6532	264.8733	...	1155.236
20200131	23	11200580	15466.15	629.6817	386.2273	...	1376.087



EDA

행정동별 생산가능인구수 시각화



자치구	행정동	생산가능인구
강남구	역삼1동	136202
영등포구	여의동	130625.9
금천구	가산동	97215.85
종로구	종로1·2·3·4가동	85100.95
서초구	서초3동	80441.65
강남구	대치2동	60703.59
구로구	구로3동	60004.76
마포구	서교동	59997.05
중구	명동	58841.36



종사자 수와 상위 9개 행정동 목록이 일치함

생산가능인구 : 경제활동을 할 수 있는 연령의 인구로, 15~64세 인구를 이른다.



주제선정

전처리
및 EDA

행정동 분석

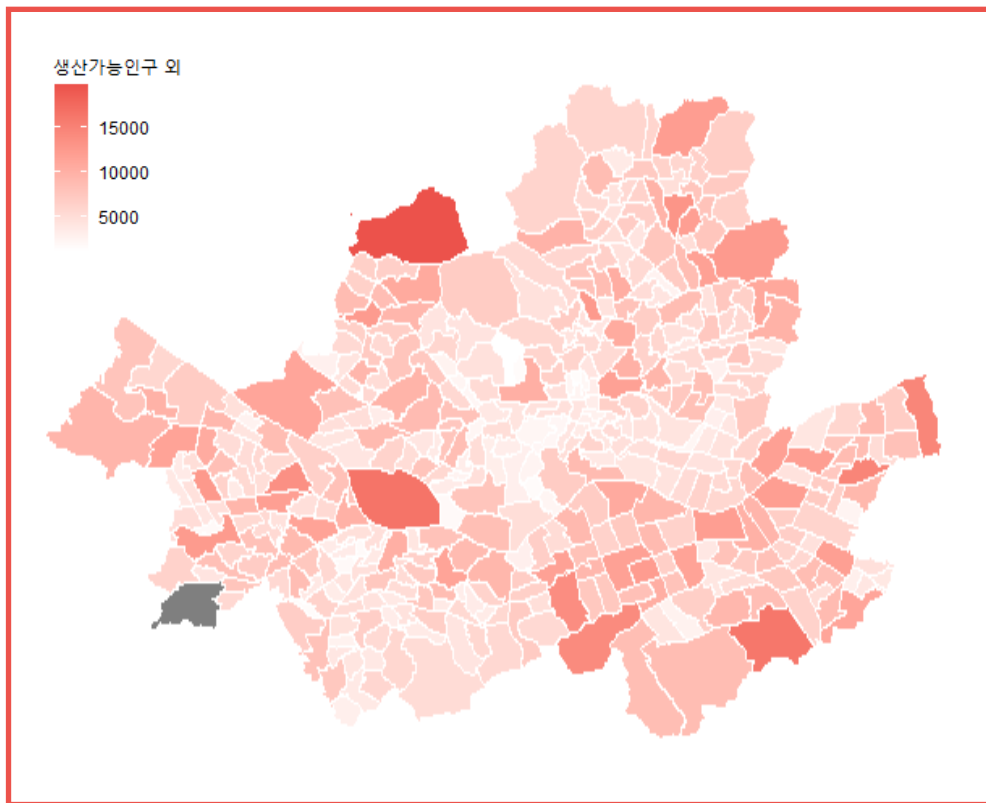


다음주 예고



EDA

행정동별 생산가능인구 외 연령대 시각화



자치구	행정동	외 연령대
은평구	진관동	19884.84
영등포구	의동	16578.07
구로구	오류2동	16507.13
강남구	세곡동	16316.53
강동구	길동	14834.89
강동구	강일동	14707.09
서초구	양재1동	14167.32
서초구	서초3동	13891.3
양천구	목5동	13570.23



생산가능인구와는 상위 9개 행정동 목록이 상이함
주거인구 등 다른 데이터셋과의 비교가 필요함



주제선정

전처리
및 EDA

행정동 분석



다음주 예고



행정동 분석

행정동 기준 데이터

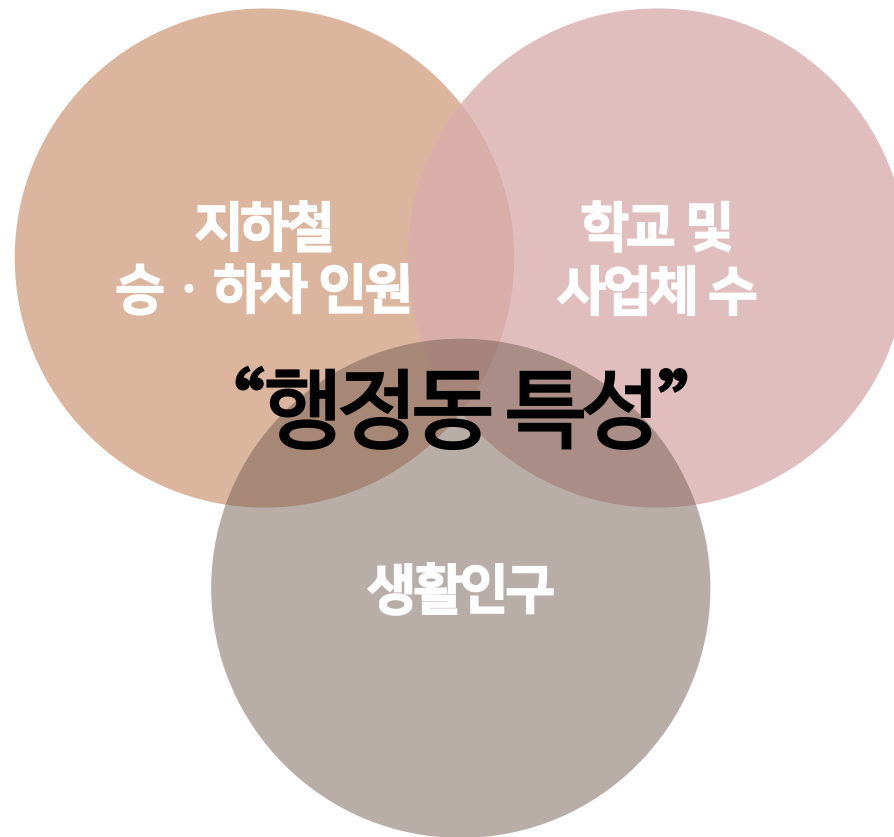
각 데이터로 살펴 본 지역별 특성을 종합적으로 알아보기 위해 행정동 클러스터링을 진행

 주제선정

 전처리
및 EDA

 행정동 분석

 다음주 예고



지금까지 살펴 본 데이터들을 종합적으로 고려해
출근 버스의 수요가 필요한 지역을 1차적으로 알아보고자 함 (feat. K-means 클러스터링)

행정동 기준 데이터

각 데이터로 살펴 본 지역별 특성을 종합적으로 알아보기 위해 행정동 클러스터링을 진행

“행정동 별 현황.csv”

424 obs x 29 var

id	행정동	대학교 수	대학 재학생 수	...	50세 밀도	60세 이상 밀도
1101053	사직동	0	0	...	3400.64	4068.758
1101054	삼청동	0	0	...	604.3477	829.4968
1101055	부암동	1	6544	...	1099.723	1844.409
...
1125072	암사1동	0	0	...	4225.871	7070.02
1125073	천호2동	0	0	...	4560.845	7183.149
1125074	길동	0	0	...	3636.513	5858.399



지금까지의 데이터를 행정동 기준으로 합침



주제선정



전처리
및 EDA



행정동 분석



다음주 예고

행정동 기준 데이터

각 데이터로 살펴 본 지역별 특성을 종합적으로 알아보기 위해 행정동 클러스터링을 진행

“행정동 별 현황.csv”

행정동
사직동
삼청동
부암동
...
암사1동
천호2동
길동

총 424개 (항동 제외)

27 var

대학교 수

대학재학생 수

대학재적생 수

고등학교 수

고등학생 수

사업체 수

종사자 수

사업체 밀도

종사자 밀도

총 생활 인구 수 / 밀도

10대 수 / 밀도

20대 수 / 밀도

30대 수 / 밀도

40대 수 / 밀도

50대 수 / 밀도

60세 이상 수 / 밀도

지하철 승차 평균

지하철 하차 평균

버스 승차 평균

버스 하차 평균





클러스터링

각 데이터로 살펴 본 지역별 특성을 종합적으로 알아보기 위해 행정동 클러스터링을 진행함

“행정동 별 현황.csv”

27 var

EDA에서 대표적 변수였던

‘대학 재학생 수’, ‘종사자 수’, ‘지하철 하차 평균’, ‘20/30/40대 수’로

클러스터링 해서 종합적으로 행정동의 특성을 파악해보자!



주제선정



전처리
및 EDA



행정동 분석



다음주 예고

암사1동

천호2동

길동

총 424개 (항동 제외)

지하철 승차 평균

지하철 하차 평균

버스 승차 평균

버스 하차 평균

10대 수 / 밀도

20대 수 / 밀도

30대 수 / 밀도

40대 수 / 밀도

50대 수 / 밀도

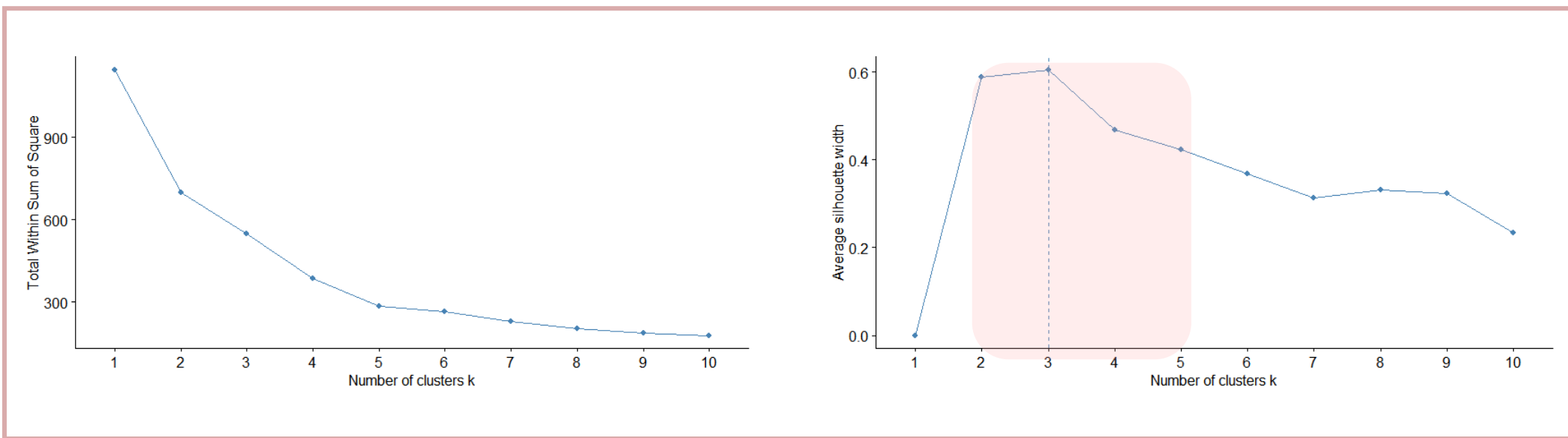
60세 이상 수 / 밀도



클러스터링

각 데이터로 살펴 본 지역별 특성을 종합적으로 알아보기 위해 행정동 클러스터링을 진행함

“클러스터 수 찾기”



총 네 개의 변수를 scaling 한 후,

WSS값과 Silhouette값, OBS값 을 고려해 클러스터의 수를 4로 결정



주제선정



전처리
및 EDA



행정동 분석



다음주 예고



클러스터링

각 데이터로 살펴 본 지역별 특성을 종합적으로 알아보기 위해 행정동 클러스터링을 진행함

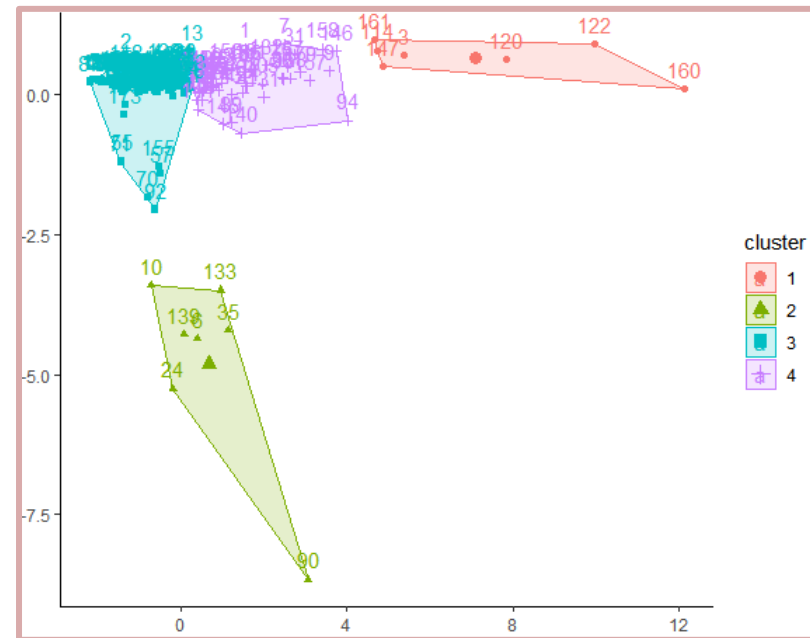
주제선정

전처리
및 EDA

행정동 분석

다음주 예고

“클러스터 PLOT”

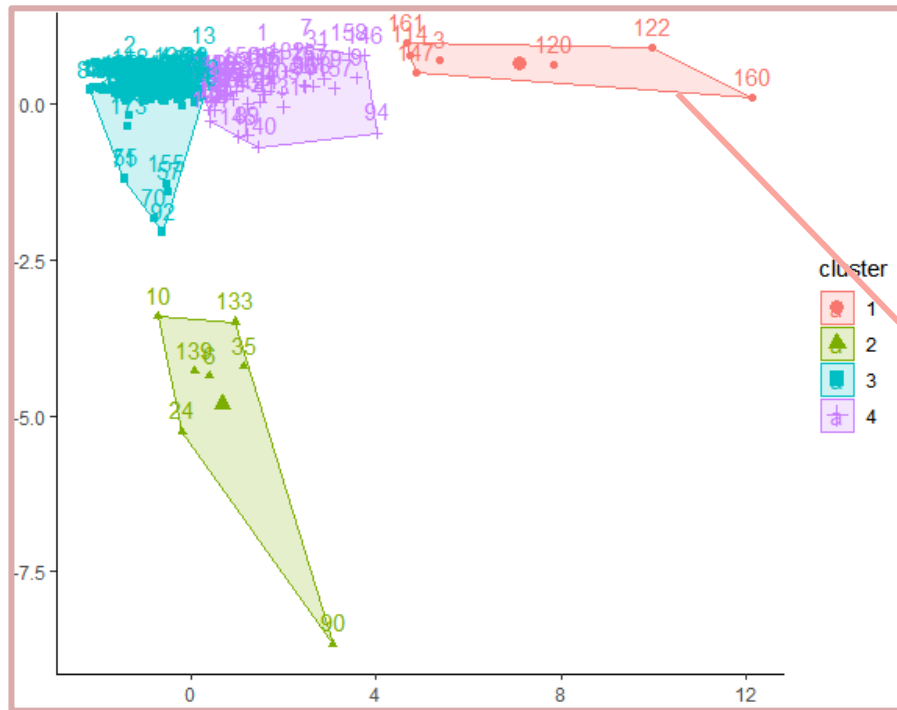


클러스터 1, 2에 소수로 묶인 곳을 확인
(클러스터 3, 4에 다수가 묶인 것은 한계..)

클러스터링

각 데이터로 살펴 본 지역별 특성을 종합적으로 알아보기 위해 행정동 클러스터링을 진행함

“클러스터1 특성”



행정동	대학 재학생 수	지하철 하차 평균	종사자수	20대	30대	40대
종로1·2·3·4가동	0	92821.265	109824	9782.833	13033.43	12731.5
구로3동	0	152954.167	87533	8991.502	11878.36	9094.37
가산동	0	113184.194	164278	14022.21	17142.87	13379.28
여의동	0	71098.771	157954	14193.5	25507.09	22655.78
서초3동	0	87797.472	78942	10379.43	13026.74	12871.59
역삼1동	0	145902.111	156700	24249.46	28857.69	19888.17
역삼2동	0	211266.778	38585	8479.21	11089.14	11689.38

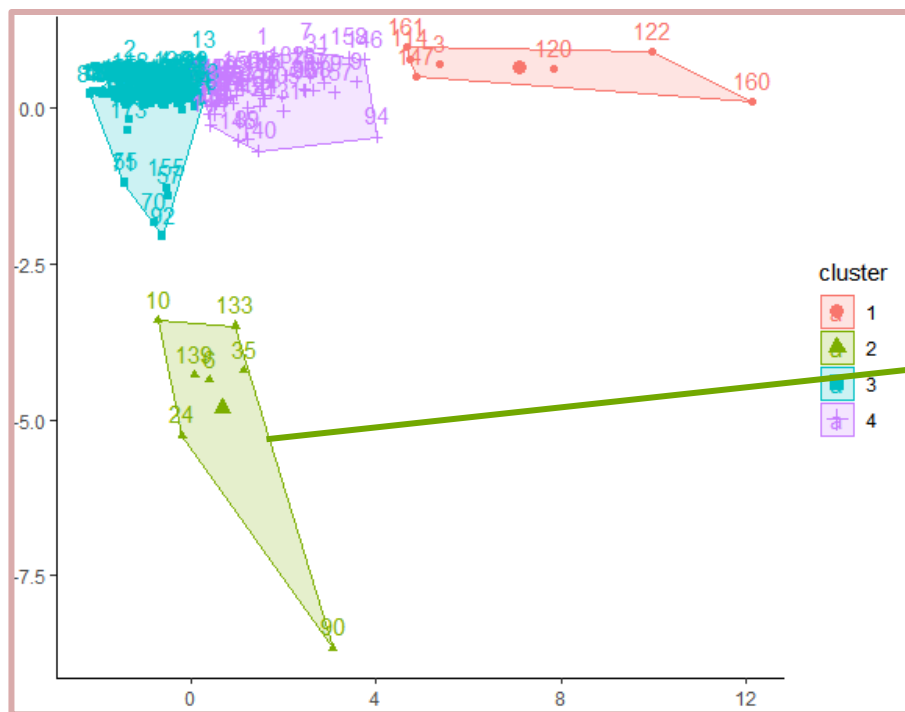
종사자수와 20,30 대 생활 인구수, 지하철 하차 평균이 많은 지역
→ 출근지로 인한 인구 집중 지역이므로 다람쥐 노선 고려 필요



클러스터링

각 데이터로 살펴 본 지역별 특성을 종합적으로 알아보기 위해 행정동 클러스터링을 진행함

“클러스터 2 특성”

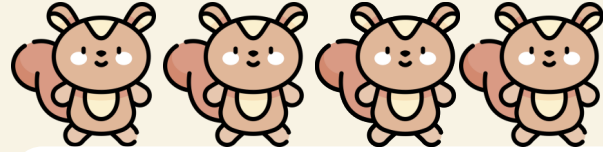


행정동	대학 재학생 수	지하철 하차 평균	종사자수	20대	30대	40대
혜화동	19541	73908.11	8855	7657.449	3364.079	3674.256
필동	17340	54910.28	24414	2290.657	1952.487	2079.725
사근동	22389	19590.39	11260	7885.405	3423.372	2809.596
화양동	15764	27544.63	16052	13249.1	5772.242	3724.713
신촌동	34622	28366.56	35949	17850.13	8432.084	6251.069
상도1동	13613	11224.17	8761	10882.77	6403.071	6049.463
흑석동	18747	11615.94	10950	6964.653	4698.224	4848.418

대학재학생 수와 20대 생활 인구수, 지하철 하차 평균이 많은 지역

→ 대학교로 인한 인구 집중 지역이므로 다람쥐 노선 고려 필요





다음주 예고

다음주 예고



주제선정



전처리
및 EDA



상관분석



다음주 예고



**2주 동안 다람쥐버스 못지않게 EDA 쳇바퀴에 갇혀
열심히 플랏을 찍어낸 (귀여운) 선대 팀..**

다음주 예고



주제선정



전처리
및 EDA



상관분석



다음주 예고



**4주차에는 통계적 기법을 활용해
지금까지 얻은 인사이트를 구체화 시켜볼 것!**

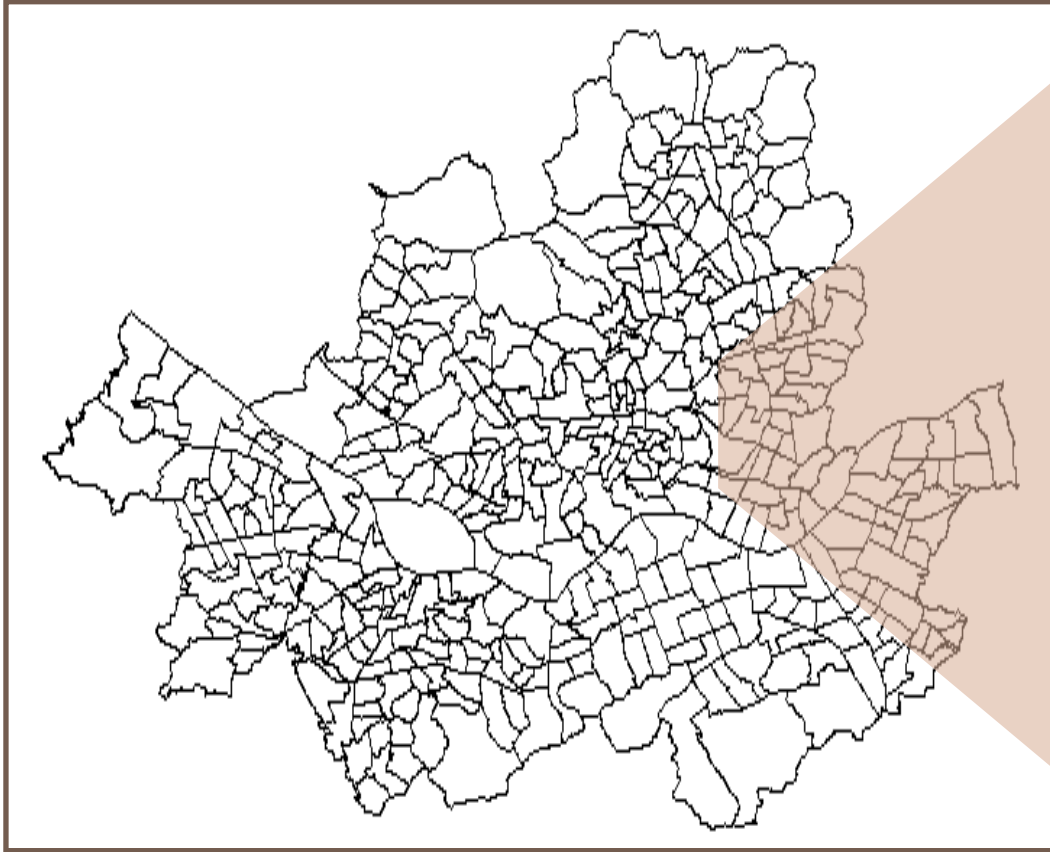
다음주 예고

 주제선정

 전처리
및 EDA

 상관분석

 다음주 예고



공간 데이터는
기본적으로 **공간자기상관성**을 지니기에
기본 회귀가정을 위배!

➡ 자기상관성을 가정하지 않는
모델들이 필요!

다음주 예고



주제선정



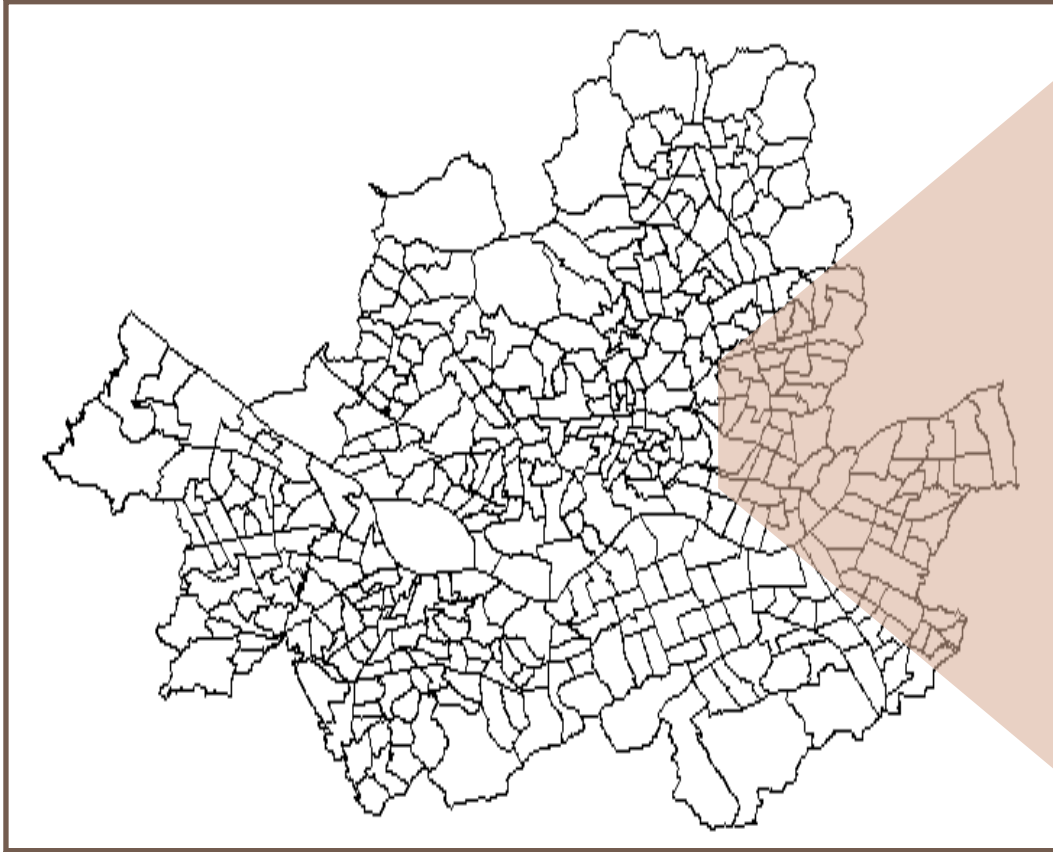
전처리
및 EDA



상관분석



다음주 예고



따라서 **공간자기상관성을 검정**하고,
GAM, 공간회귀 모델 등을 통해
버스 수요 지역을 추출한 뒤,
노선 최적화를 통해 노선을 건설할 예정!
(다음 주 선대팀에게 물어보세요-!)

해낼 수 있게 힘을 주세요..

