

24-2 EDA 프로젝트

빅데이터를 활용한 서울 심야버스 노선 신설 방안



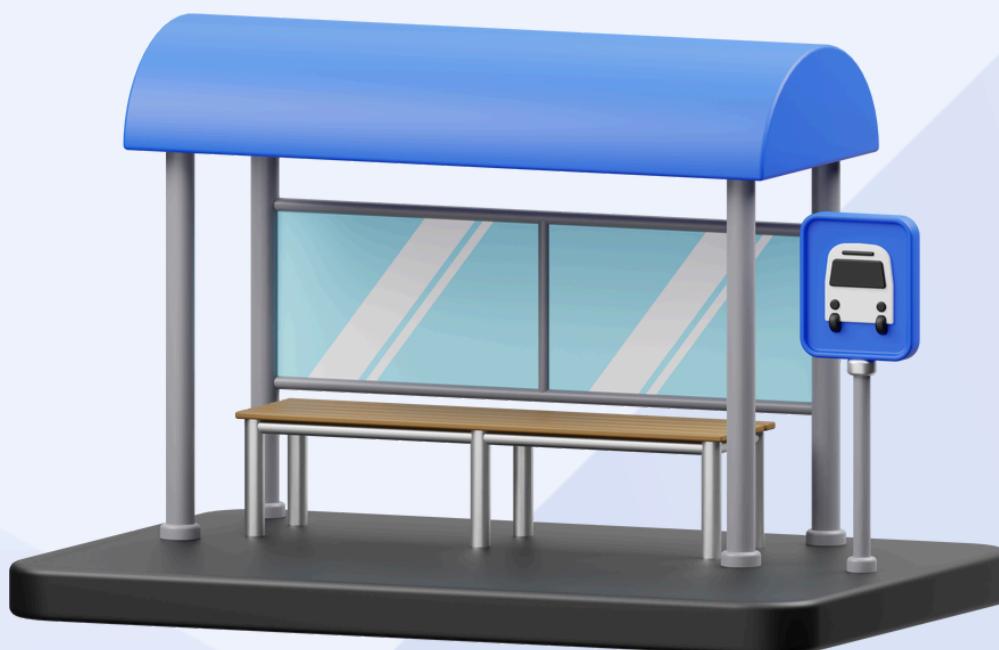
24-2 EDA 교통팀 (정시도착)

11기 고현아

12기 김민규 김은희 추준식

목차

DATA SCIENCE LAB
24-2 EDA PROJECT



01 서론

- 프로젝트 배경 및 목적
- 선행 연구

02 데이터 개요

- 활용 데이터 소개
- 데이터 주요 변수

03 데이터 전처리

- 결측치 처리
- 변수 변화, 스케일링, 그룹화
- 통합 Dataset 생성

04 탐색적 데이터 분석

- 기술 통계 및 수치 자료 시각화
- 공간 자료 분석 및 시각화

05 모델링 및 가설 검정

- 다중선형회귀모형 분석
- 기타 통계적 분석

06 신규 노선 결정 방안

- 분석 및 결정 과정

07 결론 및 향후 과제

- 의의 및 한계, 향후 연구 방향

08 참고 문헌 및 질의응답

- 참고 문헌
- 질의 응답

| 서론

"심야 시간에도 마음 편히 집으로~" 서울시, 심야 버스 '올빼미버스' 확대 운행한다

정지원 기자 | 0 승인 2022.04.05 13:20

**오는 18일부터 확대 운행...빅데이터 활용해 유동인구·이동실태 파악
심야 시간 '통행량 집중지역'과 '이동 사각지역' 발생...'신설 노선'으로 문제 해결
올빼미버스...주요 상권의 경제 활성화로 이어질 것**

데이터안심구역 **서비스안내**

주요 분석사례 **음영 지역 채우고, '과밀 노선' 나눠 담고**

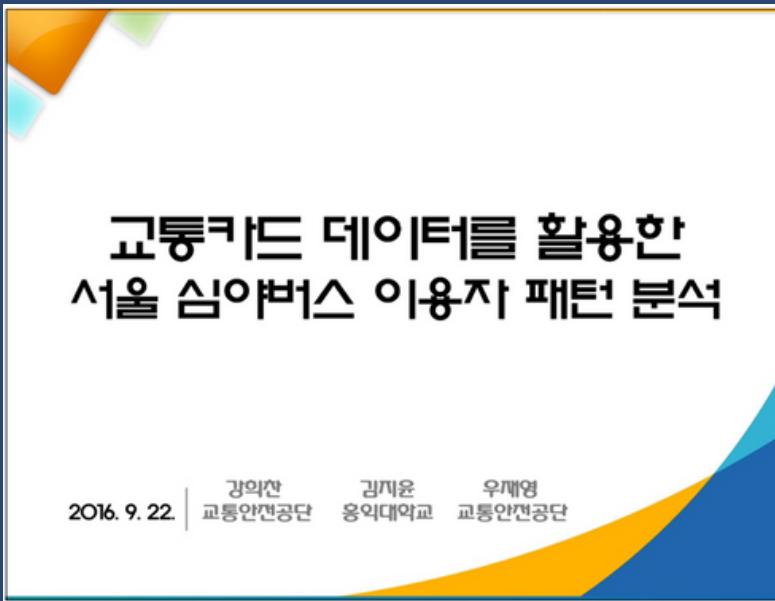
심야버스 노선 확충은 지난 2016년 서남권 심야버스로 계획되었던 N65번 버스 개통 이후 6년 만의 일이다. 이번 노선 확충안을 살펴보면, 버스가 탑승객으로 꽉 차 승하차에 어려움을 겪었던 과밀 노선의 혼잡을 나눠 담고, 수요가 많았지만 심야버스를 이용하기 어려웠던 지역을 중심으로 노선을 늘렸다.

**2022년 심야 버스 노선이 14개로 증가함에도 불구하고,
노선 증설 및 개선 필요성이 꾸준히 대두됨.**



| 서론

✓ 1. 교통카드 데이터를 활용한 서울 심야버스 이용자 패턴 ✓



- 진행 분석
 - 승하차 패턴 분석 → 권역별/시간대별
 - 통행 분석 → 권역별/시간대별
 - O/D분석 (승하차량) → 권역별/시간대별
 - 승차량 대비 하차량 분석

- 얻은 인사이트
 - 해당 연구의 결론 (의의 및 한계)
: 노선이 없는 지역에 대해서도 지속적 수요 분석 필요



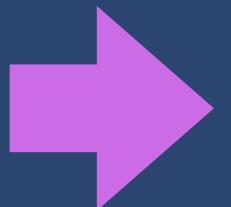
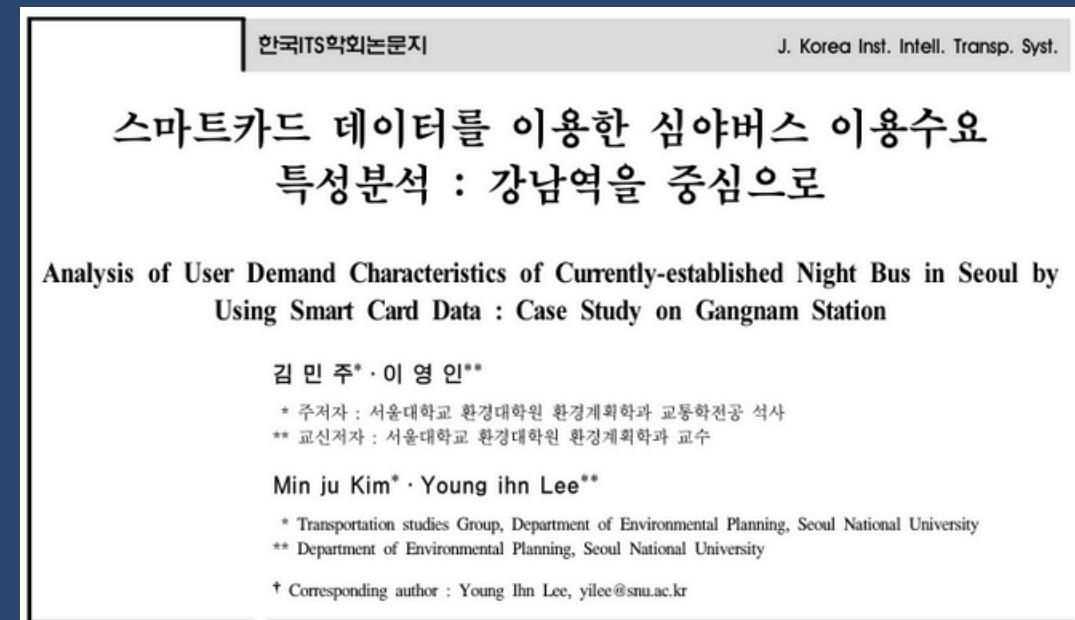
수요는 있으나 노선이 없는 지역을 우선적으로 파악하여
해당 지역에 노선을 신설하려고 하는
팀의 방향성과 부합함.

| 서론

✓ 2. 스마트카드 데이터를 이용한 심야버스 이용수요 특성분석 ✓

- 얻은 인사이트

- 서울시 전체를 공간적 단위로 나눈 데이터를 사용해야 함
→ **행정동 및 권역별**로 서울시의 데이터를 나눔
- 주요 정류소간 기종점 파악 필요함
→ 심야버스 신설 노선의 기종점은 노선이 없는 지역을 우선적으로 파악하므로, 서울시 인구의 **권역별 방향성**으로 수요를 파악한 후, **심야버스 노선이 없는 권역**을 노선의 기종점으로 삼음
- 각 행정동 및 권역이 갖는 **지역적 특성** 고려해야 함



| 서론

✓ 3. 택시 통행 패턴 기반 심야 버스 노선 계획 방안에 관한 연구 ✓

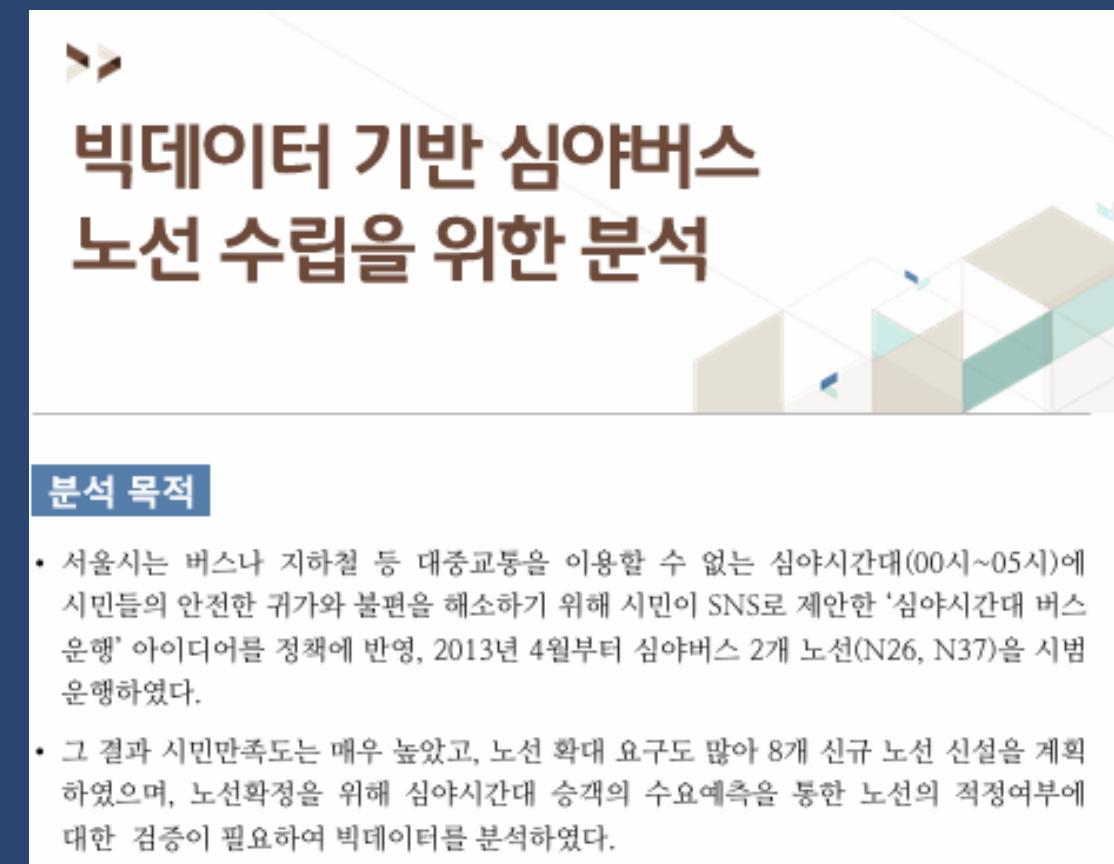


• 택시 통행 데이터 시간대 지정

- 평일 심야시간대 (00-05시)
→ 수집한 데이터의 지정 시간대 (00-05시)와 동일
- '탑승시간' 기준 택시 데이터 수집
→ 서울 생활이동 데이터는 '도착 시간' 기준 데이터 수집
- 심야버스의 편익 / 비용을 산출할 수 있는 정보 고려 필요
→ 심야버스 노선이 도출되고 난 후, 가용하다면 편익 / 비용을 계산하여 수치적으로 노선의 현실성을 검증할 필요성 있음

| 서론

✓ 4. 빅데이터 기반 심야버스 노선 수립을 위한 분석 ✓



- 얻은 인사이트

- 활용데이터 중 KT 유동인구 데이터의 요일 기준
: 주중 / 토 / 일 00시~05시
-> 시간별 분석시, '주중 / 주말'이 아닌
'주중 / 토 / 일' 00시~05시 적용
- 노선 검증 시, 유동인구 분포 및 구역별 밀도 시각화 진행
-> 유동인구 분포 시각화에서 더 나아가

유동인구 밀도의 수치적 시각화 필요

| 데이터 개요

01

서울시 버스 노선별 정류장별 승하차 인원 정보

교통데이터

활용사례(갤러리) 등록 URL 복사 목록 이동

서울시 버스노선별 정류장별 승하차 인원 정보

교통카드(선후불교통카드)를 이용한 서울버스 노선별/정류장별 승하차인원을 나타내는 정보입니다.
일별 버스노선마다 각 정류장에 송/하차한 데이터의 합입니다. (일단위)
서울버스 대상 (서울시내, 서울광역, 서울마을)
※ Sheet 서비스는 마지막 25일치 데이터만 서비스 합니다. (* 데이터 적재는 매일 3일전 데이터를 경신합니다.)

전체 설명보기

02

서울 생활이동 인구 (행정동별)

데이터 내려받기

행정동 단위

2024년 ▾ 1월 2월 3월 4월 5월 6월 7월 8월 9월 10월 11월 12월

자치구 단위

2024년 ▾ 1월 2월 3월 4월 5월 6월 7월 8월 9월 10월 11월 12월

데이터 안내

서울생활이동 데이터 설명서 행정동 코드 정보 자치구 코드 정보

03

행정동 구분 코드 및 중심 위/경도 좌표

A	B	C	D	E	F	G	H	I	J
시구동	시도명	시군구명	읍면동명	경도	위도	sidocode	gungucode	dongcode	행정동코드
서울특별시 서울특별시 종로구	청운효자동	126.9704	37.58466	110000	110100	1101072	1111050		
서울특별시 서울특별시 종로구	사직동	126.9701	37.57411	110000	110100	1101053	1111050		
서울특별시 서울특별시 종로구	삼청동	126.9811	37.58801	110000	110100	1101054	1111050		
서울특별시 서울특별시 종로구	부암동	126.9626	37.5967	110000	110100	1101055	1111050		
서울특별시 서울특별시 종로구	평창동	126.9693	37.61397	110000	110100	1101056	1111050		
서울특별시 서울특별시 종로구	무악동	126.959	37.57774	110000	110100	1101057	1111050		
서울특별시 서울특별시 종로구	교남동	126.9642	37.57105	110000	110100	1101058	1111050		
서울특별시 서울특별시 종로구	가회동	126.9866	37.58268	110000	110100	1101060	1111060		



04

행정동 경계 Polygon 데이터 (geojson 형식)

대한민국 행정동 경계(admdongkor)

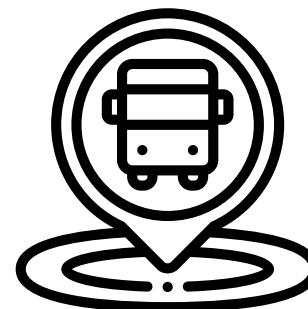
- 저작권 관련 사항은 본 문서의 끝부분에 기술하였으니 참고하시기 바랍니다.
- 정확하고 자세한 행정동 경계가 필요한 분들은 아래의 링크들에서 다운받으시기 바랍니다.
 - <https://business.juso.go.kr/addlink/adresInfoProv/guidance/provdAdresInfo.do>
 - 2022년 5월부터 도로명주소지도 사이트에서도 행정동 경계를 배포하고 있습니다. 한 달에 한번 업데이트 됩니다.
 - 위의 링크에서 '구역의 도형'을 신청하여 받으면 됩니다.
 - 지도 별로 분할되어 있으며, TL_SCCO_GEMD 파일이 행정동 경계입니다.
 - 이 곳에서 배포하는 경계보다 훨씬 세밀하고 정확하며, 토풀로지 정합성도 거의 잘 맞춰져 있습니다.
 - <http://data.nsdi.go.kr/dataset/20171206ds00001>
 - 국가공간정보 포털 사이트에서도 행정동 경계를 받을 수 있습니다.
 - 1년에 한 번 정도 업데이트 되는 것 같습니다.

| 데이터 개요

01

서울시 버스 노선별, 정류장별 승하차 인원 정보

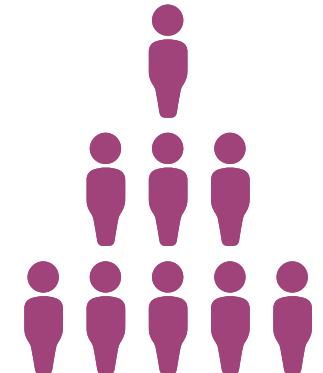
심야 버스 정류장의 위치를 파악하고
승하차 인원 수를 파악하기 위해 사용
(정류장 위치 확인을 위해 정류장 위치 정보도 사용)



02

서울 생활이동 인구 (행정동별)

심야 시간대 특정 지역 간 유동 인구 추이와
그에 따른 이동 수단 수요를 파악하기 위해 사용
(행정동 구분은 21년 6월 기준)



03

행정동 구분 코드 및 중심 위/경도 좌표

생활이동 인구와 같은 데이터들을 시각화하는 과정에서
해당 행정동의 중심지를 통해 표현하기 위해 사용



04

행정동 경계 Polygon 데이터 (geojson 형식)

공간 자료를 시각화하고 버스 정류장 등
특정 지점의 소속 행정동을 확인하는 용도로 사용

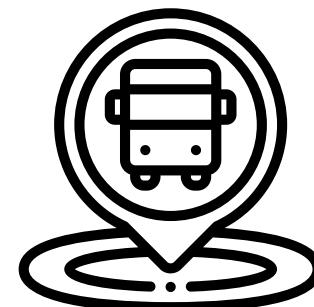


| 데이터 개요

01

서울시 버스 노선별 정류장별 승하차 인원 정보

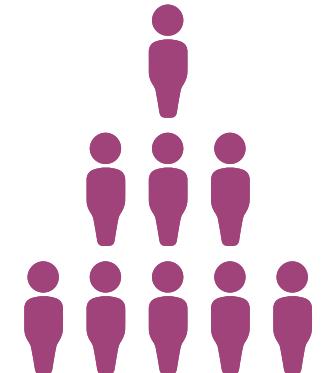
노선번호, 표준버스정류장 ID,
1시간 단위 정류장 승/하차 총승객수



02

서울 생활이동 인구 (행정동별)

출발/도착 행정동 코드, 성별 및 나이, 이동 유형,
평균 이동시간, 이동인구 (3명 미만 시 마스킹)



03

행정동 구분 코드 및 중심 위/경도 좌표

시/구/동 이름, 중심 지점 위/경도, 행정동 코드 7자리



04

행정동 경계 Polygon 데이터 (geojson 형식)

행정동 이름 및 행정동 코드 7자리,
행정동 경계 (MultiPolygon) 좌표 집합



데이터 전처리

서울시 버스노선별 정류장별 시간대별 승하차 인원 정보

<https://data.seoul.go.kr/dataList/OA-12913/S/1/datasetView.do>

1) 월별 이용객수 -> 연간 이용객수로 그룹화 계산

2) 00-05시 승차/하차/승하차 수 파생변수로 생성

3) 정류장 좌표 정보 + 행정동 경계 정보를 통해

정류장의 소속 행정동 및 좌표 정보 결합

month	bus_num	station_id	station_name	00_in	00_out	01_in	01_out	02_in	02_out	03_in	03_out	04_in	04_out	05_in	05_out
2	202301	N37	100000001 종로2가사거리(00089)	0	0	0	0	91	72	186	100	89	31	0	0
3	202301	N37	100000001 종로2가사거리(00032)	21	14	163	146	96	85	0	0	0	0	0	0
5	202301	N31	100000002 창경궁.서울대학교병원(00013)	6	1	2	1	0	0	0	0	0	0	0	0
7	202301	N31	100000002 창경궁.서울대학교병원(00089)	0	0	0	0	1	2	0	2	1	0	0	0
22	202301	N31	100000003 명륜3가.성대입구(00088)	0	0	0	0	0	18	4	2	1	3	1	0

bus_num	station_id	00_in	00_out	01_in	01_out	02_in	02_out	03_in	03_out	...	05_in	05_out	night_in	night_out	night_total	STTN_NM	CRDNT_X	CRDNT_Y	geometry	adm_nm	
0	N13	100000190	605	307	837	888	469	565	356	529	...	0	0	2523	2460	4983	종로중부새마을금고	127.006591	37.572417	POINT (127.00659 37.57242)	서울특별시 종로구 종로 5-6가동
157	N15	100000396	342	170	1728	1426	1699	2095	1237	1271	...	0	0	5625	5539	11164	종로6가.동대문종합시장	127.004463	37.571011	POINT (127.00446 37.57101)	서울특별시 종로구 종로 5-6가동
158	N15	100000397	932	439	1534	714	1129	701	915	979	...	0	0	4924	3162	8086	종로6가.동대문종합시장	127.005714	37.571127	POINT (127.00571 37.57113)	서울특별시 종로구 종로 5-6가동
291	N16	100000118	574	1172	1073	3954	863	2181	627	1915	...	0	0	3324	9785	13109	동대문	127.008413	37.571726	POINT (127.00841 37.57173)	서울특별시 종로구 종로 5-6가동
292	N16	100000119	88	364	177	742	147	381	168	331	...	0	0	852	1997	2849	충신동	127.006279	37.574022	POINT (127.00628 37.57402)	서울특별시 종로구 종로 5-6가동
...		
2031	N75	120000055	0	0	29	3629	15	3066	12	3182	...	13	902	77	12752	12829	서울대벤처타운역.삼성동시장	126.932960	37.472713	POINT (126.93296 37.47271)	서울특별시 관악구 삼성동
2042	N75	121000019	2186	420	2159	770	1199	663	1040	490	...	0	0	7689	3038	10727	고속터미널	127.005129	37.506290	POINT (127.00513 37.50629)	서울특별시 서초구 반포 4동
2043	N75	121000020	1547	959	2855	1389	2372	659	1695	936	...	396	260	9598	5057	14655	고속터미널	127.003924	37.505846	POINT (127.00392 37.50585)	서울특별시 서초구 반포 4동
2075	N75	121000302	191	385	244	604	218	369	116	378	...	2	70	800	2064	2864	반포대교남단.한강시민공원입구	126.999468	37.507811	POINT (126.99947 37.50781)	서울특별시 서초구 반포 2동
2076	N75	121000303	428	139	638	220	647	242	454	148	...	0	0	2473	836	3309	반포대교남단.한강시민공원입구	127.000390	37.506944	POINT (127.00039 37.50694)	서울특별시 서초구 반포 3동

2076 rows x 22 columns

데이터 전처리

행정동 단위 서울 생활 이동, 인구 (내국인)

<https://data.seoul.go.kr/dataList/OA-14991/S/1/datasetView.do#>

서울 생활이동

서울 생활이동 인구란?

서울 안에서 이동하거나 서울 외부에서 서울로 오고 간 이동으로 통근, 통학, 쇼핑, 여가 등 행정수요를 유발하는 모든 이동

서울 생활이동 현황 (2024.06 기준)

대상연월	요일	도착시간	출발 행정동 코드	도착 행정동 코드	성별	나이	이동유형	평균 이동	이동인구(합)
202301	일	0	1101053	1101053	F	20	EE	66	39.07
202301	일	0	1101053	1101053	F	20	EH	138	*
202301	일	0	1101053	1101053	F	20	HE	36	*
202301	일	0	1101053	1101053	F	25	EE	24	*
202301	일	0	1101053	1101053	F	25	EH	14	4.08
202301	일	0	1101053	1101053	F	25	HH	1	*
202301	일	0	1101053	1101053	F	25	WW	48	*
202301	일	0	1101053	1101053	F	30	EH	37	10.78
202301	일	0	1101053	1101053	F	30	HH	49	*
202301	일	0	1101053	1101053	F	30	WH	57	4.73
202301	일	0	1101053	1101053	F	25	EE	29	6.05

생활이동 데이터 : 출발/도착 행정동 코드, 도착 시간대, 나이, 이동 인구 등
 → 3명 미만으로 * 표기된 수치는 1로, 00~05시 시간대 자료에 대해 처리

생활인구 데이터 : 행정동 코드, 해당 시간대별/나이대별 총 생활 인구
 → 월별 통합하여 데이터 분석에 활용, 00~05시 시간대 자료에 대해 처리

행정동 단위 서울 생활인구(내국인)

서울시가 보유한 공공데이터와 통신데이터로 측정한 특정시점에 서울의 특정 지역에 존재하는 모 든인구수 정보※ Sheet, Open API는 최근 2개월 데이터만 제공합니다.

파일내려받기

NO	파일명	내려받기
1	LOCAL_PEOPLE_DONG_202406.zip	
2	LOCAL_PEOPLE_DONG_202405.zip	
3	LOCAL_PEOPLE_DONG_202404.zip	
4	LOCAL_PEOPLE_DONG_202403.zip	
5	LOCAL_PEOPLE_DONG_202402.zip	

A	B	C	D	E	F
1 기준일ID	시간대구분	행정동코드	총생활인구	남자0세부	남자10세부
2 20220701	0	11740560	20846.92	818.9991	607.4987
3 20220701	0	11320681	15736.86	638.648	427.2095
4 20220701	0	11710642	33484.93	590.556	259.1215
5 20220701	0	11215820	26227.69	556.0031	339.0859
6 20220701	0	11230545	31948.47	890.8261	425.6721
7 20220701	0	11305535	21858.19	473.9707	317.9279
8 20220701	0	11470630	16028.44	605.2261	671.1383
9 20220701	0	11620715	13055.54	603.6797	296.3384

| 데이터 전처리

B	C	D	E	F	G	H	I	J
시도명	시군구명	읍면동명	경도	위도	sidocode	gungucoddongcode	행정동코드	
서울특별시 종로구	청운효자동	126.97	37.5847	11000	11010	1101072	11110515	
서울특별시 종로구	사직동	126.97	37.5741	11000	11010	1101053	11110530	
서울특별시 종로구	삼청동	126.981	37.588	11000	11010	1101054	11110540	
서울특별시 종로구	부암동	126.963	37.5967	11000	11010	1101055	11110550	
서울특별시 종로구	평창동	126.969	37.614	11000	11010	1101056	11110560	
서울특별시 종로구	무악동	126.959	37.5777	11000	11010	1101057	11110570	
서울특별시 종로구	교남동	126.964	37.5711	11000	11010	1101058	11110580	
서울특별시 종로구	가회동	126.987	37.5827	11000	11010	1101060	11110600	
서울특별시 종로구	종로1·2·3·4	126.99	37.5751	11000	11010	1101061	11110615	

대한민국 전체의 지명, 경도, 위도, 행정동 코드 7자리, 8자리가
모두 포함된 데이터에서 '서울특별시'에 해당하는 자료만 담긴 xlsx 생성

컬럼 명	컬럼 의미
시도명	서울시의 시/도 이름
시군구명	서울시의 군/구 이름
읍면동명	서울시의 읍/면/동 이름
경도	해당 지역의 경도 좌표
위도	해당 지역의 위도 좌표
sidocode	시/도를 나타내는 고유 코드
gungucode	군/구까지 나타내는 고유 코드
dongcode	동까지 나타내는 고유 코드
행정동코드	동코드 7자리를 변환한 8자리 코드

| 데이터 전처리

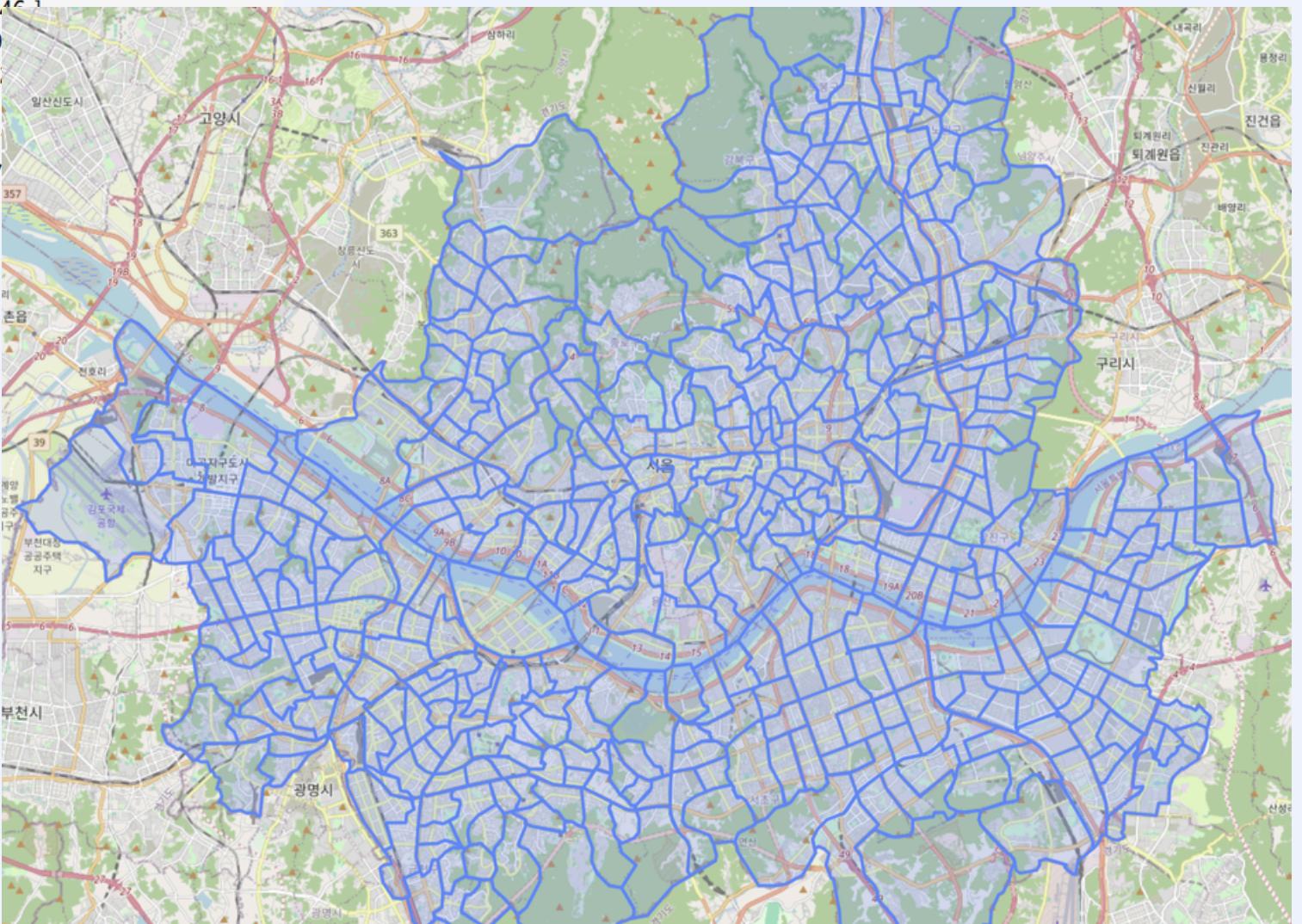
대한민국 행정동 경계 Polygon 데이터

<https://github.com/vuski/admdongkor/tree/master/ver20210701>

```
{
  "type": "FeatureCollection",
  "name": "HangJeongDong_ver20210401",
  "crs": { "type": "name", "properties": { "name": "urn:ogc:def:crs:OGC:1.3:CRS84" } },
  "features": [
    { "type": "Feature", "properties": { "OBJECTID": 1, "adm_nm": "서울특별시 종로구 사직동", "adm_cd": "1101053", "adm_cd2": "1111053000", "sgg": "11110", "sido": "11", "sidonm": "서울특별시", "temp": "종로구 사직동", "sggnm": "종로구" }, "geometry": { "type": "MultiPolygon", "coordinates": [ [ [ [ 126.976888842748167, 37.575650779448786 ], [ 126.977034498877501, 37.569194530054553 ], [ 126.975974728212492, 37.569336299425764 ], [ 126.975374709912543, 37.569315567021562 ], [ 126.974331935623255, 37.569261800517531 ], [ 126.969048370018541, 37.568194417708334 ], [ 126.968544936033837, 37.568427679612761 ], [ 126.966649959821197, 37.569491655206583 ], [ 126.966281750244846, 37.569700734798701 ], [ 126.966097327080405, 37.569856509723699 ], [ 126.965728529225771, 37.570183936115114 ], [ 126.965926998221278, 37.570318805686199 ], [ 126.96601094018429, 37.571548395577466 ], [ 126.963659220521961, 37.575174660660373 ], [ 126.963086004345101, 37.576485920015543 ], [ 126.962840990511978, 37.57666158609274 ], [ 126.962810410472628, 37.579448809656775 ], [ 126.967424315843317, 37.579601537124489 ], [ 126.967421763026508, 37.57926352144161 ], [ 126.96743006184597, 37.579192577998604 ], [ 126.967457090095607, 37.578975250585437 ], [ 126.968066046996256, 37.578246780467872 ], [ 126.9689551169, 37.577935262340283 ], [ 126.969212842969057, 37.577935299309388 ], [ 126.969414538865792, 37.578121124142164 ], [ 126.969664426694706, 37.578531136682 ], [ 126.969667219148718, 37.578736205134931 ], [ 126.969668773533087, 37.578992879009881 ], [ 126.969669499103631, 37.57911252674959 ], [ 126.96990457361, 37.579301753628734 ], [ 126.97135197544759, 37.57951327793981 ], [ 126.973819257844539, 37.579372140302631 ], [ 126.973917363383421, 37.57848707304101 ], [ 126.973939619980882, 37.578240429978088 ], [ 126.974331538357575, 37.575749906299862 ], [ 126.975803789978045, 37.575649468824203 ], [ 126.976888842748167, 37.575650779448786 ] ] ] }
  ] }
```

서울시 polygon 데이터만 추출하여 사용

OBJECTID	adm_nm	adm_cd	adm_cd2	sgg	sido	sidonm	temp	sggnm	geometry
0	1	서울특별시 종로구 사직동	1101053	1111053000	11110	11	서울특별시	종로구 사직동	종로구 MULTIPOLYGON (((126.97689 37.57565, 126.97703 37.57565, 126.97703 37.57555, 126.97689 37.57555, 126.97689 37.57565)))
1	2	서울특별시 종로구 삼청동	1101054	1111054000	11110	11	서울특별시	종로구 삼청동	종로구 MULTIPOLYGON (((126.98269 37.59507, 126.98337 37.59507, 126.98337 37.59498, 126.98269 37.59498, 126.98269 37.59507)))
2	3	서울특별시 종로구 부암동	1101055	1111055000	11110	11	서울특별시	종로구 부암동	종로구 MULTIPOLYGON (((126.97585 37.59656, 126.97359 37.59656, 126.97359 37.59647, 126.97585 37.59647, 126.97585 37.59656)))
3	4	서울특별시 종로구 평창동	1101056	1111056000	11110	11	서울특별시	종로구 평창동	종로구 MULTIPOLYGON (((126.97507 37.63139, 126.97649 37.63139, 126.97649 37.63129, 126.97507 37.63129, 126.97507 37.63139)))
4	5	서울특별시 종로구 무악동	1101057	1111057000	11110	11	서울특별시	종로구 무악동	종로구 MULTIPOLYGON (((126.96067 37.58080, 126.96281 37.58080, 126.96281 37.58070, 126.96067 37.58070, 126.96067 37.58080)))
...
419	420	서울특별시 강동구 둔촌1동	1125070	1174069000	11740	11	서울특별시	강동구 둔촌1동	강동구 MULTIPOLYGON (((127.14580 37.52199, 127.14564 37.52199, 127.14564 37.52183, 127.14580 37.52183, 127.14580 37.52199)))
420	421	서울특별시 강동구 둔촌2동	1125071	1174070000	11740	11	서울특별시	강동구 둔촌2동	강동구 MULTIPOLYGON (((127.15355 37.53400, 127.15357 37.53400, 127.15357 37.53384, 127.15355 37.53384, 127.15355 37.53400)))
421	422	서울특별시 강동구 암사1동	1125072	1174057000	11740	11	서울특별시	강동구 암사1동	강동구 MULTIPOLYGON (((127.14447 37.55477, 127.14319 37.55477, 127.14319 37.55461, 127.14447 37.55461, 127.14447 37.55477)))
422	423	서울특별시 강동구 천호2동	1125073	1174061000	11740	11	서울특별시	강동구 천호2동	강동구 MULTIPOLYGON (((127.13458 37.54712, 127.13115 37.54712, 127.13115 37.54696, 127.13458 37.54696, 127.13458 37.54712)))
423	424	서울특별시 강동구 길동	1125074	1174068500	11740	11	서울특별시	강동구 길동	강동구 MULTIPOLYGON (((127.14857 37.54578, 127.14871 37.54578, 127.14871 37.54562, 127.14857 37.54562, 127.14857 37.54578)))



| 탐색적 데이터 분석

각 권역을 이어주는 N버스의 특징 분석



=> N버스의 이름 구성 : N + 출발 권역 + 도착 권역

ex) N37은 3번 (강동, 송파)에서 출발하여
7번 (은평, 마포, 서대문)까지 가는 버스

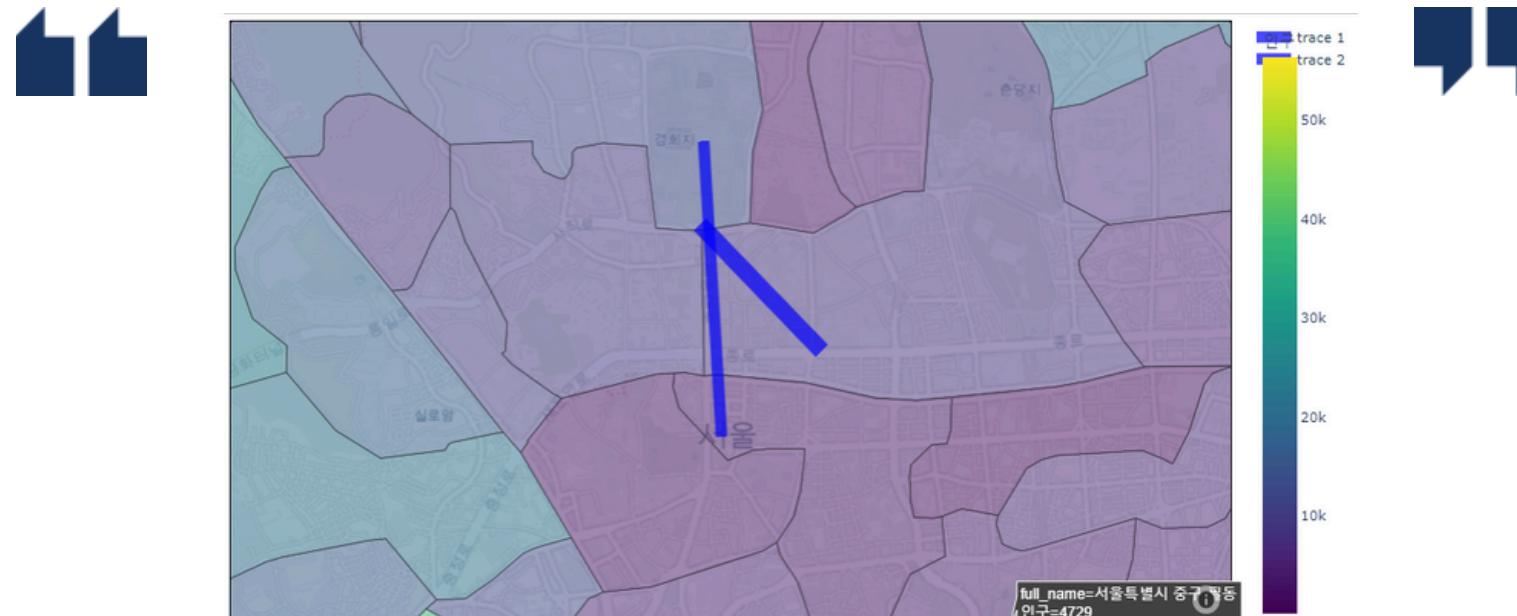
시행되는 N 버스 14개의 실제 노선 경로 분포를 시각화를 통해 확인

예) N64 노선 시각화



| 탐색적 데이터 분석

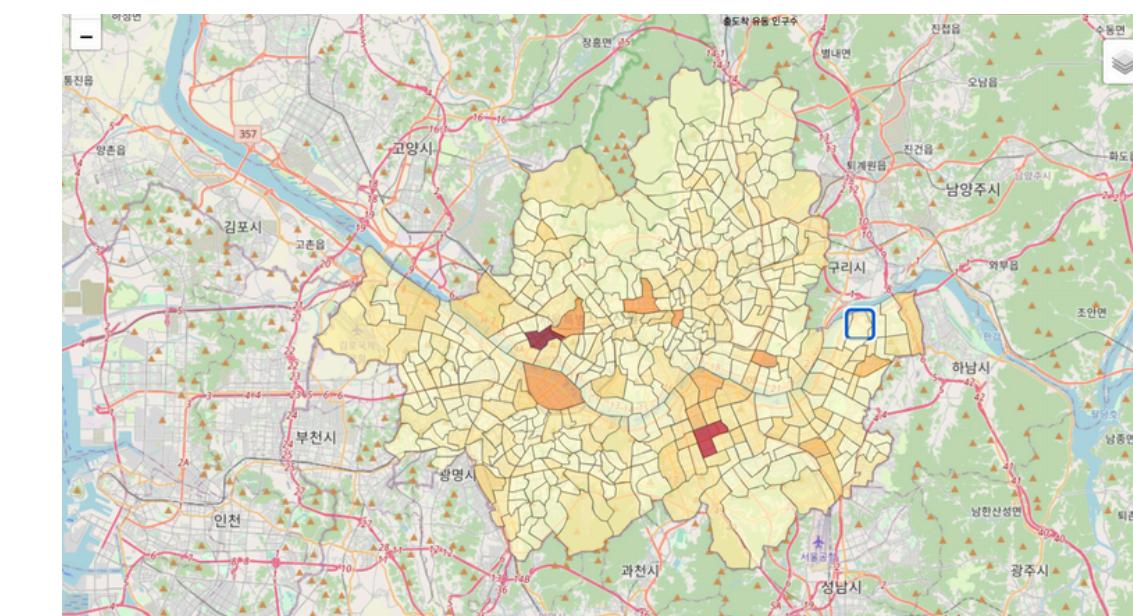
각 월별 출발, 도착 행정동 연결 시각화



=> 1월 00시~05시의 출발 행정동 코드와 도착 행정동 코드의 위치, 경도를 선으로 연결하고, 이동인구의 수를 굵기로 표현함

- 선: 출발 위도/경도, 도착 위도/경도 연결
- 굵기: 이동 인구 수
- 서울생활이동데이터의 1~12월 & 00시-05시 데이터 통합
- 서울 시민의 '이동방향성'과 '이동인구 수'를 파악하여 중요도를 매기기 위함

각 행정동별 유동인구 시각화

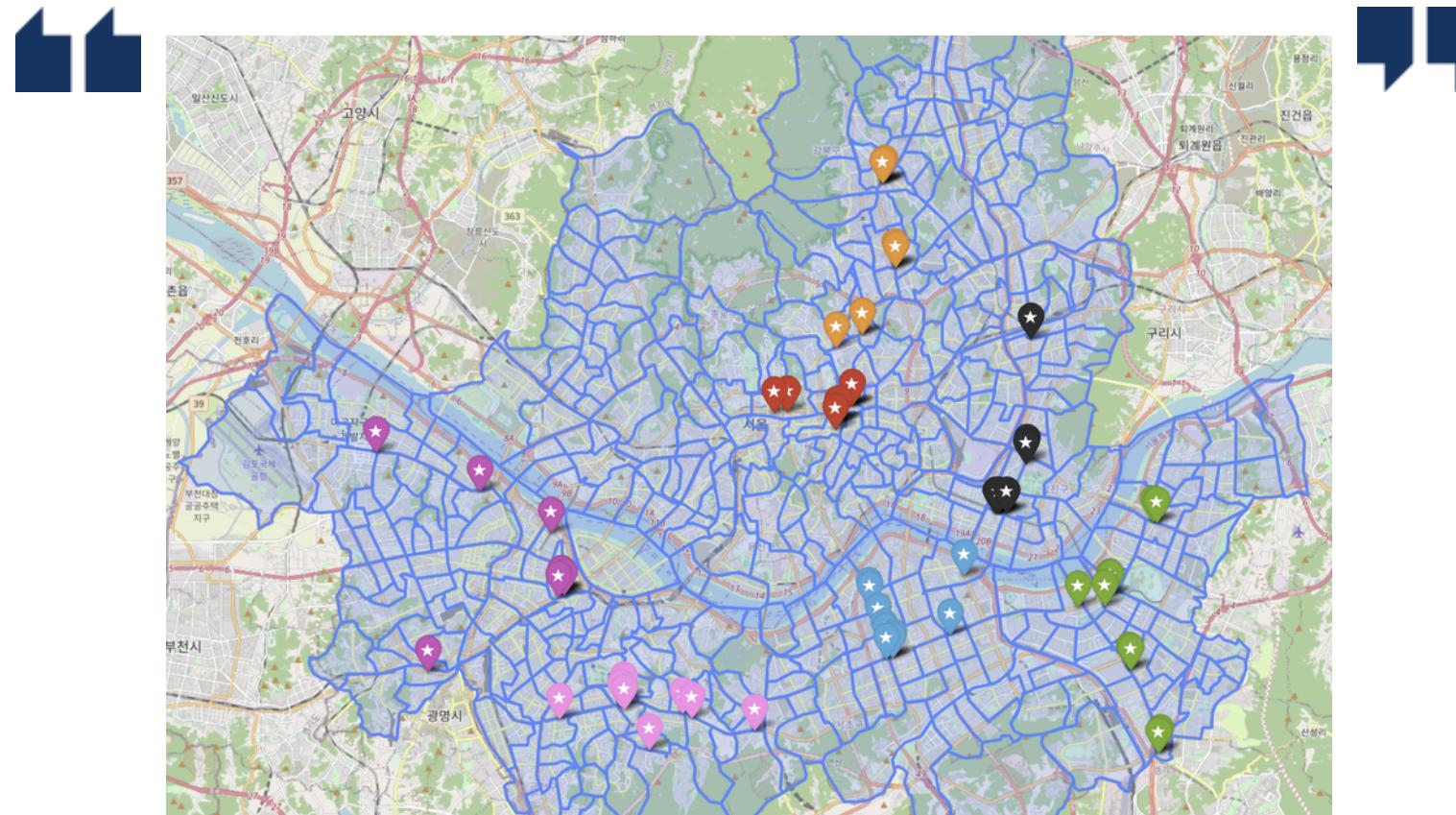


=> 서울 생활이동 데이터에서 각 행정동별로 출발 / 도착 / 합계를 이동 인구를 기준으로 하여 합산함

- 진하게 색칠된 부분은 유동인구가 많다는 것을 의미

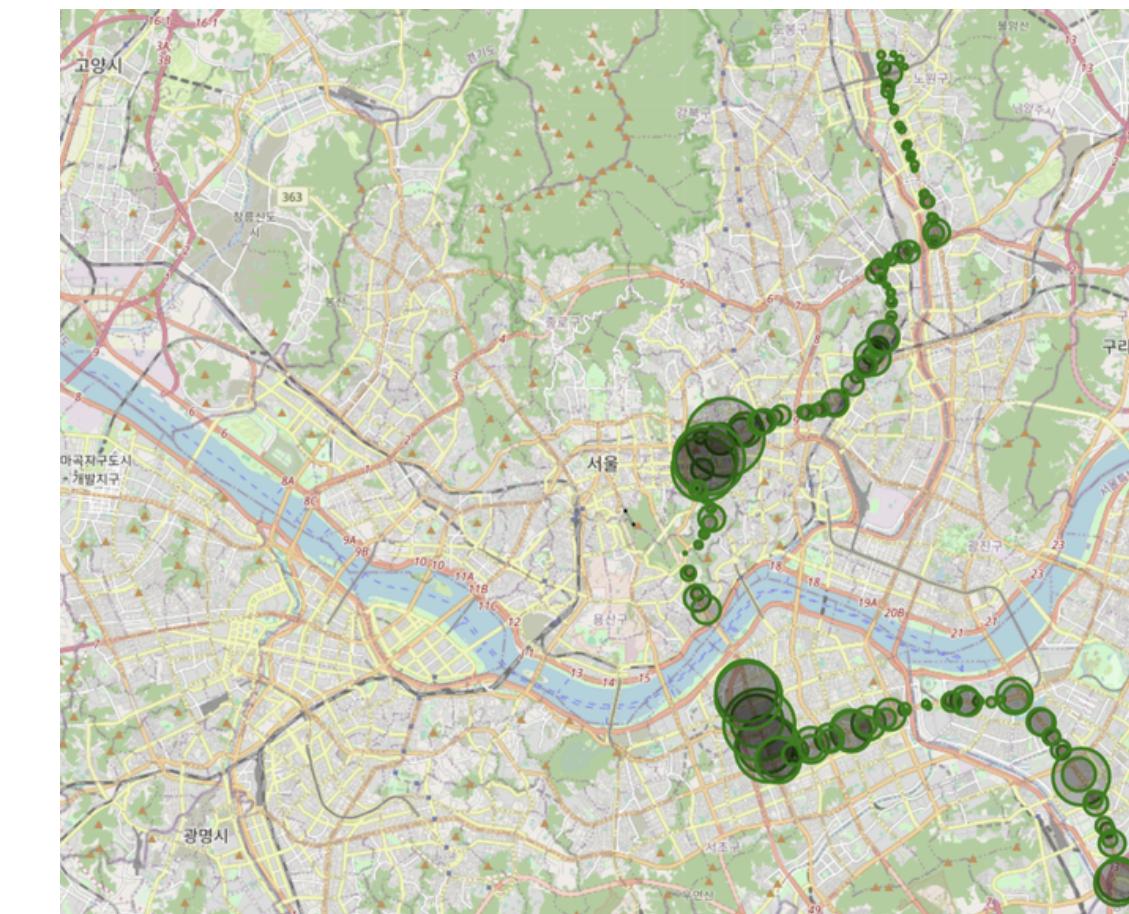
| 탐색적 데이터 분석

권역별 버스정류장 승/하차 이용객 수 상위 지점



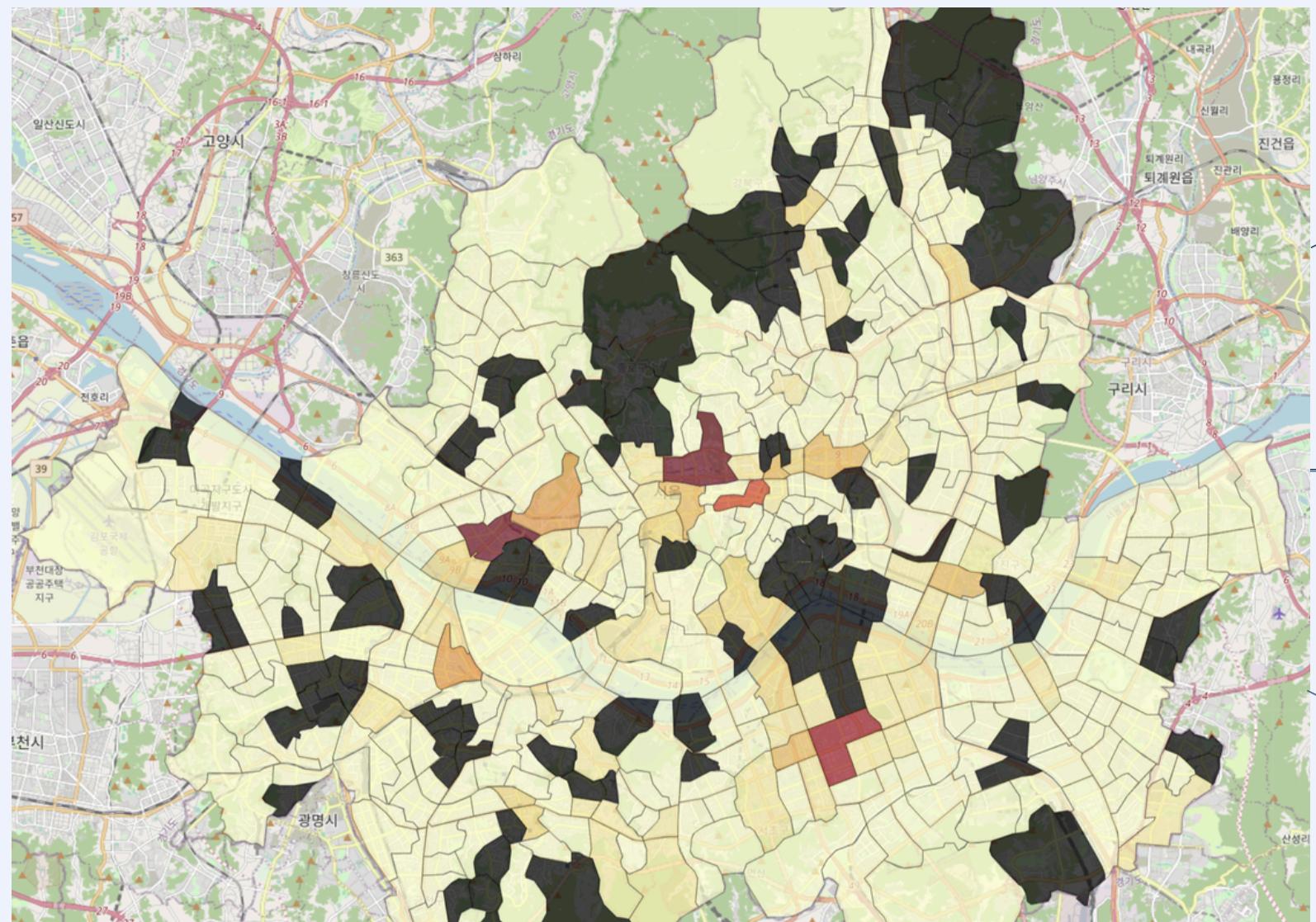
=> 정류장의 위치를 기준으로 0~6권역으로 구분한 후,
각 권역에서의 이용객 수 상위 10개소를 추출해
각기 다른 색으로 시각화함.
(팝업 및 툴팁을 통해 수치, 정류소명을 확인 가능함.)

노선별 정류장별 승/하차 이용객 수 시각화



=> 특정 노선을 중심으로 정류장의 승차 / 하차 / 승하차
인원 수를 원형으로 표시함.
(원의 크기는 해당 지점의 이용객 수에 비례함.)

| 모델링 및 가설 검정



행정동별 심야버스 총 승/하차인원

● : 정류장 없는 지점! = 97곳

가정

심야버스 이용량은 행정동의 면적, 주거인구, 점포수, 생활인구수 등의 영향을 받을 것이다.

목표 1

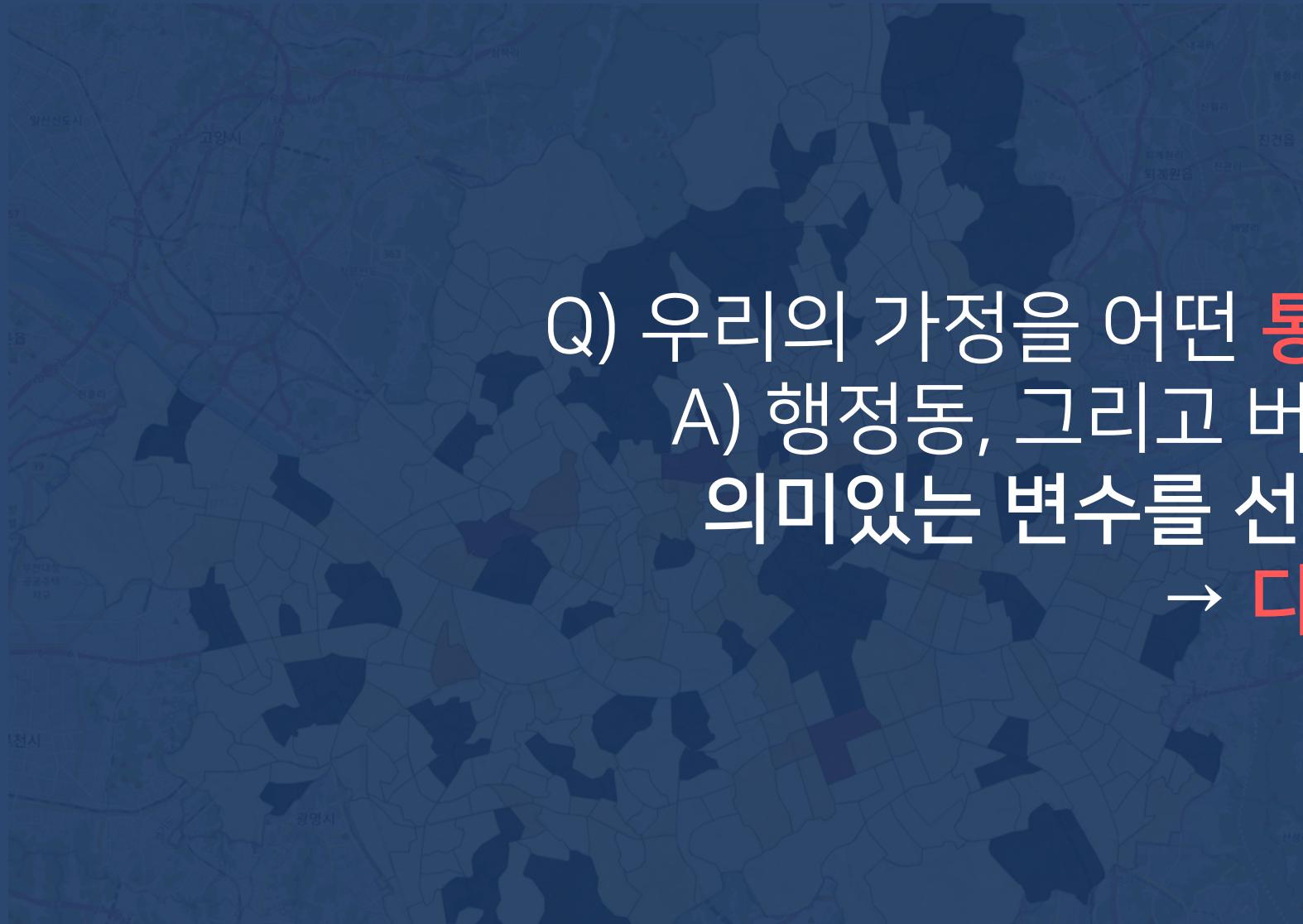
심야버스 정류장이 **존재하는** 행정동에 대해
실제 이용량과 모델 예측값의 차이가 큰 지점을 분석한다.

목표 2

심야버스 정류장이 **존재하지 않는** 행정동에 대해
예상 이용량이 큰 지점을 중심으로 신규 노선을 설정한다.



| 모델링 및 가설 검정



행정동별 심야버스 총 승/하차인원

● : 정류장 없는 지점!

가정

심야버스 이용량은 행정동의 면적, 주거인구, 점포수, 생활인구수 등의 영향을 받을 것이다.

Q) 우리의 가정을 어떤 **통계적 방법**을 통해 검증할 수 있을까?

A) 행정동, 그리고 버스 이용량과 관련된 여러 변수 중
의미있는 변수를 선별하여 이용량을 추정하고 싶다.

→ **다중선형회귀모형!**

목표 2

심야버스 정류장이 존재하지 않는 행정동에 대해
예상 이용량이 큰 지점을 중심으로 신규 노선을 설정한다.

모델링 및 가설 검정

사용 데이터

X 서울 생활인구 (총생활인구수)

서울 생활인구

서울 생활인구 현황 (2024.07.30. 기준)

일일평균 생활인구 **10,686**천명

24시간 평균 내국인 **10,169**천명

24시간 평균 정기체류 외국인 **364**천명

24시간 평균 단기체류 외국인 **153**천명

Home > 서울빅데이터 > 서울 생활인구

서울시 행정구역 (동별, 구별) 통계-면적

서울시 행정구역 (동별) 통계

서울시 행정구역 (구별) 통계

서울시 행정구역 (동별) 통계

서울시 행정구역 (구별) 통계

서울시 버스 노선별 정류장별 승하차 인원 정보

서울시 버스 노선별 정류장별 승하차 인원 정보

교통

서울시 상권분석 서비스-지역분석 (인구수)

서울시 상권분석 서비스-지역분석 (인구수)

서울시 상권분석 서비스-지역분석 (점포수)

서울시 상권분석 서비스-지역분석 (점포수)

X 서울시 행정구역 (동별, 구별) 통계-면적

활용사례(갤러리) 등록 URL 복사

서울시 행정구역 (동별) 통계

○ 통계개요
* 통계명 : 행정구역(동별)
* 통계종류 : 서울시 각 자치구 동별 행정구역 현황
* 작성목적 : 서울시 행정구역 현황을 파악하여, 지방도시의 변화

활용사례(갤러리) 등록 URL 복사

서울시 행정구역 (구별) 통계

○ 통계개요
* 통계명 : 행정구역(구별)
* 통계종류 : 서울시 자치구별 행정구역 현황을 제공하는 일반 통계
* 작성목적 : 서울시 행정구역 현황을 파악하여, 지방도시의 변화

2022년 4분기 2023년 4분기

생활밀집업종	2022년 4분기		2023년 4분기		
	전체 점포수	프랜차이즈 점포수	전체 점포수	프랜차이즈 점포수	
서울시 전체	외식업/ 전체	147,899	25,033	122,866	148,530
종로구	외식업/ 전체	6,989	961	6,028	7,030
중구	외식업/ 전체	6,540	1,121	5,419	6,653
용산구	외식업/ 전체	5,436	671	4,765	5,508
성동구	외식업/ 전체	4,635	710	3,925	4,782

외식업 → 서비스업 → 소매업

X 서울시 상권분석 서비스-지역분석 (인구수)

X 서울시 상권분석 서비스-지역분석 (점포수)

X 서울시 상권분석 서비스-지역분석 (점포수)

| 모델링 및 가설 검정

회귀모형 적합

X

Y

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	시구동	시	구	동	길단위유동인구	주거인구	직장인구	면적	외식업점포수	소매업점포수	서비스업점포수	총생활인구수	버스이용량
2	서울특별시	서울특별시	종로구	사직동	34919	73	412	1.23	1001	777	990	56298192.99	15687
3	서울특별시	서울특별시	종로구	삼청동	6255	19	42	1.49	266	334	95	12495240.43	591
4	서울특별시	서울특별시	종로구	교남동	108234	125	24	0.35	125	172	136	16116442.24	5720
5	서울특별시	서울특별시	종로구	가회동	27231	77	207	0.54	269	226	142	16711981.85	856
6	서울특별시	서울특별시	종로구	종로1·2·3·4·7	58795	29	401	2.35	2460	6477	1845	140089525.8	101018
7	서울특별시	서울특별시	종로구	종로5·6가동	112461	89	260	0.6	549	1398	372	38223279.33	20771
8	서울특별시	서울특별시	종로구	이화동	48404	110	70	0.78	397	313	130	52625068.53	5791
9	서울특별시	서울특별시	종로구	혜화동	83739	140	57	1.12	666	481	396	72280958.42	6377
10	서울특별시	서울특별시	종로구	창신1동	99860	191	44	0.31	215	1491	193	18463706.29	44385

- 학습 데이터 : 2022년 4분기 (10월~12월)에 해당하는 자료 → 인구 / 점포수 수치 갱신 주기에 따른 범위 축소
- 심야버스가 지나가지 않는 정류장 ($Y = 0$) 자료는 적합 시 사용하지 않음.
- 변수 선택법으로는 **단계적 선택법** (전진 선택법 + 후진 소거법 반복) 사용.
- SPSS** 소프트웨어 활용 회귀모형 분석 및 잔차 분석 진행.

모델링 및 가설 검정

회귀모형 분석 (SPSS, 단계적 선택법)

모형	R	R 제곱	수정된 R 제곱	추정값의 표준오차	모형 요약 ^d					
					통계량 변화량			유의확률 F 변화량	Durbin-Watson	
R 제곱 변화량	F 변화량	자유도1	자유도2							
1	.709 ^a	.502	.501	8187.210	.502	326.703	1	324	<.001	
2	.719 ^b	.517	.514	8078.514	.015	9.777	1	323	.002	
3	.728 ^c	.530	.526	7977.355	.013	9.244	1	322	.003	1.722

a. 예측자: (상수), 외식업점포수

b. 예측자: (상수), 외식업점포수, 소매업점포수

c. 예측자: (상수), 외식업점포수, 소매업점포수, 면적

d. 종속변수: 버스이용량

입력/제거된 변수 ^a			
모형	입력된 변수	제거된 변수	방법
1	외식업점포수		단계선택 (기준: 입력에 대한 F의 확률 $\leq .050$, 제거에 대한 F의 확률 $\geq .100$).
2	소매업점포수		단계선택 (기준: 입력에 대한 F의 확률 $\leq .050$, 제거에 대한 F의 확률 $\geq .100$).
3	면적		단계선택 (기준: 입력에 대한 F의 확률 $\leq .050$, 제거에 대한 F의 확률 $\geq .100$).

a. 종속변수: 버스이용량

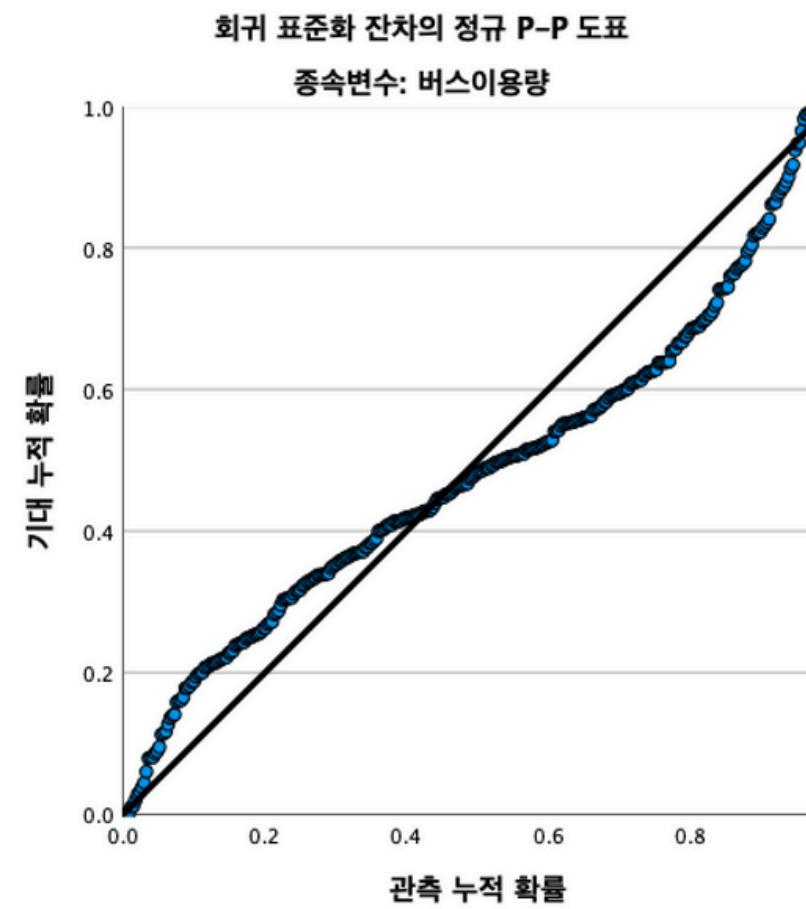
모형	비표준화 계수			표준화 계수 베타	t	유의확률	공선성 통계량	
	B	표준화 오류	베타				공차	VIF
1 (상수)	-1470.562	675.528			-2.177	.030		
외식업점포수	23.290	1.288	.709		18.075	<.001	1.000	1.000
2 (상수)	-1798.962	674.783			-2.666	.008		
외식업점포수	20.105	1.629	.612		12.343	<.001	.609	1.641
소매업점포수	2.079	.665	.155		3.127	.002	.609	1.641
3 (상수)	-827.998	738.912			-1.121	.263		
외식업점포수	20.914	1.630	.636		12.828	<.001	.593	1.686
소매업점포수	2.025	.657	.151		3.084	.002	.609	1.643
면적	-868.619	285.697	-.118		-3.040	.003	.964	1.037

a. 종속변수: 버스이용량

- 단계적 선택법 적용 결과 **외식업점포수** → **소매업점포수** → **면적** 순으로 모형에 변수가 추가됐으며, 제거된 변수는 없음.
- Adjusted R-square 값은 약 0.53이며, 최종 모형의 경우 **모든 변수의 p-value가 0.05보다 작음.** (유의한 변수임.)

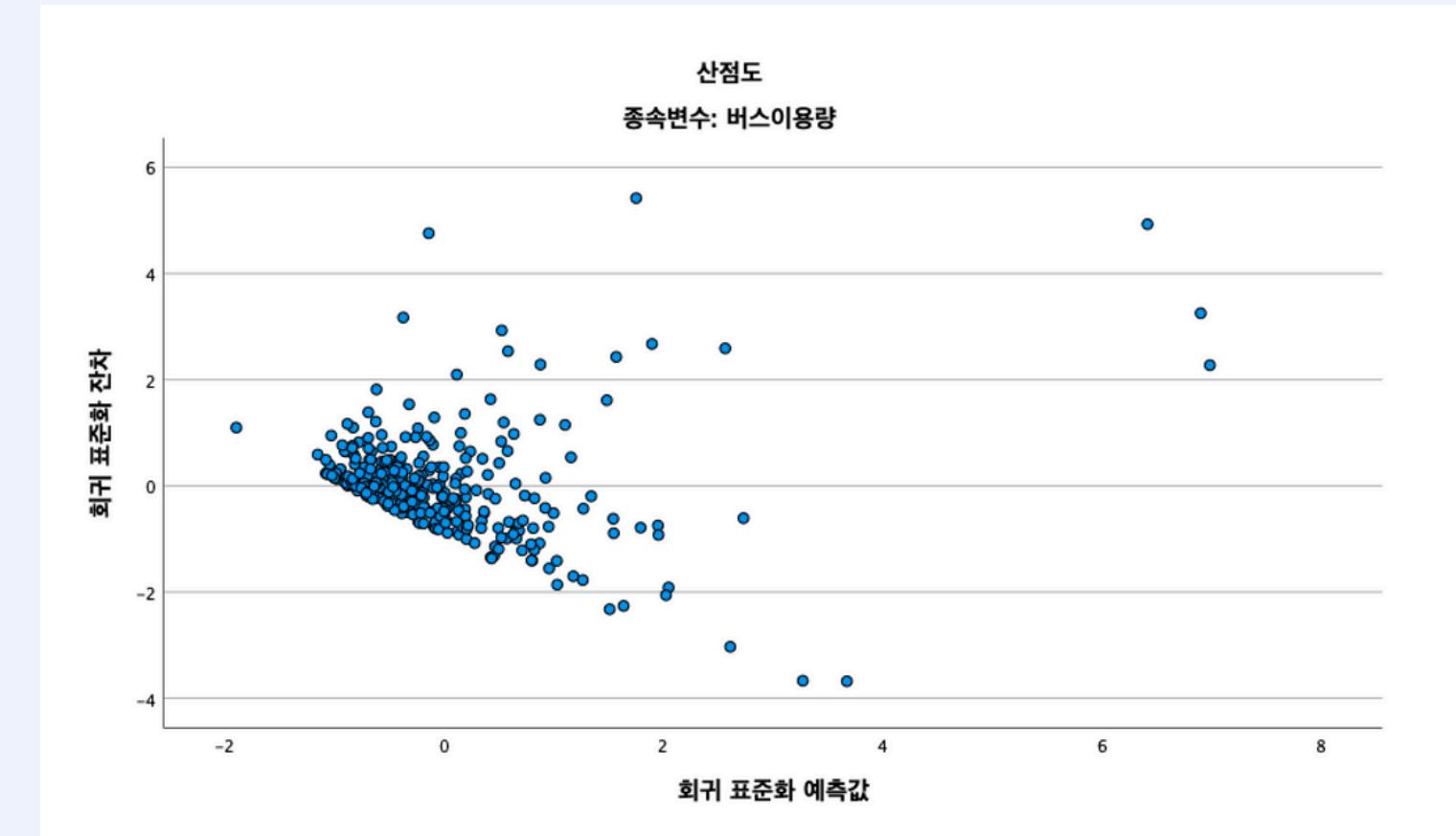
모델링 및 가설 검정

회귀모형 분석 (SPSS, 단계적 선택법)



모형	R	R 제곱	수정된 R 제곱	추정값의 표준오차	통계량 변화량				유의확률 F 변화량	Durbin-Watson
					R 제곱 변화량	F 변화량	자유도1	자유도2		
1	.709 ^a	.502	.501	8187.210	.502	326.703	1	324	<.001	
2	.719 ^b	.517	.514	8078.514	.015	9.777	1	323	.002	
3	.728 ^c	.530	.526	7977.355	.013	9.244	1	322	.003	1.722

- a. 예측자: (상수), 외식업점포수
- b. 예측자: (상수), 외식업점포수, 소매업점포수
- c. 예측자: (상수), 외식업점포수, 소매업점포수, 면적
- d. 종속변수: 버스이용량



- 정규 P-P Plot을 통해 어느 정도 정규성을 만족하며, Durbin-Watson 통계량이 2에 근접함을 통해 잔차 독립성을 보임.
- 잔차 그림의 경우 많은 예측값들이 특정 지점에 몰려있는 모습을 보이나, 어느 정도 등분산성을 (패턴 X) 만족하고 있음.

| 모델링 및 가설 검정

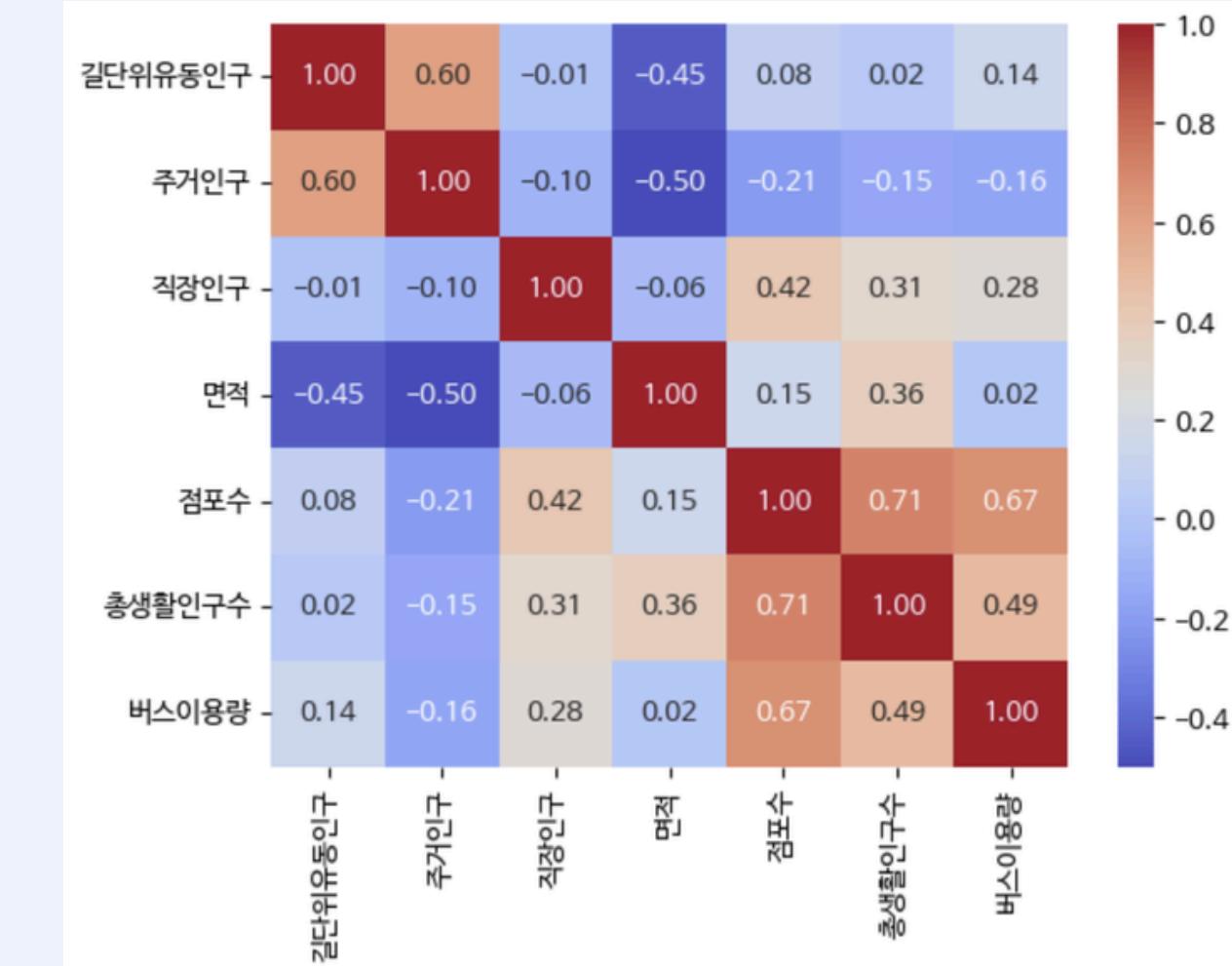
그러나!

- 독립변수 간 높은 상관성을 탐지하여 설명력이 완전하지 않음.
- 사용하는 데이터 자체의 신뢰성에 문제가 있음.

서울시 상권분석서비스(상주인구-상권)

서울시 상권 영역 내 상주인구 정보를 제공합니다.

1년중 4분기에 한번 데이터를 업데이트 하여 다음 해 1, 2, 3분기의 값이 동일합니다.



행정구역	2021년 4분기			2022년 4분기			2023년 4분기		
	길단위 유동인구	주거 인구	직장 인구	길단위 유동인구	주거 인구	직장 인구	길단위 유동인구	주거 인구	직장 인구
광진구	54,438	203	48	54,051	200	48	54,686	200	48
동대문구	73,855	240	56	74,771	226	56	75,328	226	56

신규 노선 결정 방안

HOME > 라이프 > 생활

"심야 시간에도 마음 편히 집으로~" 서울시, 심야 버스 '올빼미버스' 확대 운행한다

정지원 기자 | 승인 2022.04.05 13:20



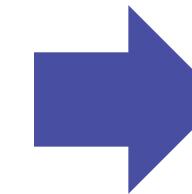
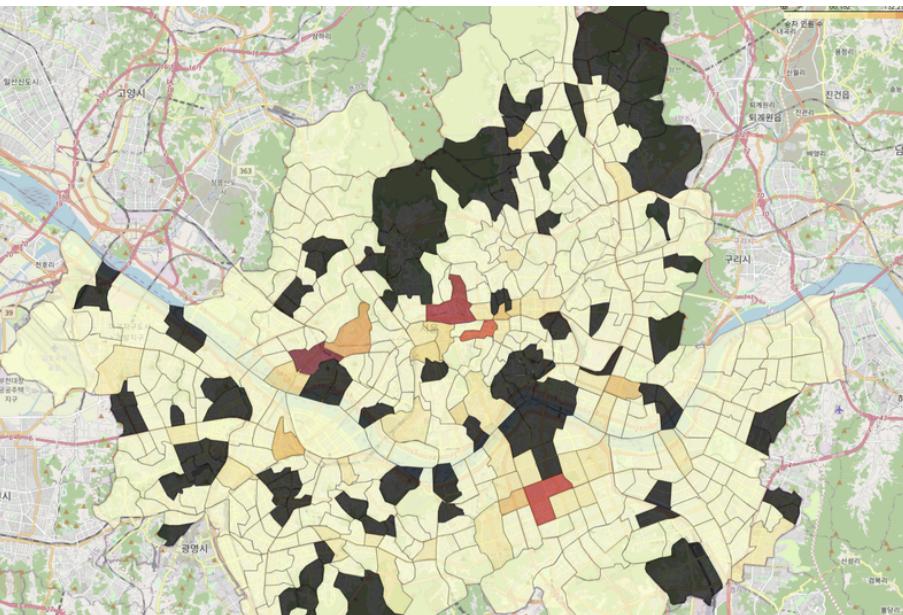
※ 심야시간 유동인구 집중 지역

- 1) 강남 (강남구 역삼1동)
- 2) 여의도 (영등포구 여의동)
- 3) 홍대 (마포구 서교동)
- 4) 원도심 (중구 신당동)
- 5) 건대 (광진구 화양동)

- 어느 지역의 수요에 집중해야 하는가? **집중지역과 사각지역!**
 - 통행량 집중지역 : 상업 / 업무 시설이 밀집한 도심 및 부도심 지역으로, 지하철 2호선 구간 중심으로 분포
 - 집중지역 → 주거지역으로 이동할 때, 수요가 존재함에도 심야버스가 지나치는 '**이동 사각지역**' 발생
- **도심과 부도심을 연계하는 노선을 확대하였음**에도 불구하고, 현 생활이동 데이터를 통해 **여전히 사각지대가 존재**하는 것을 인지함.

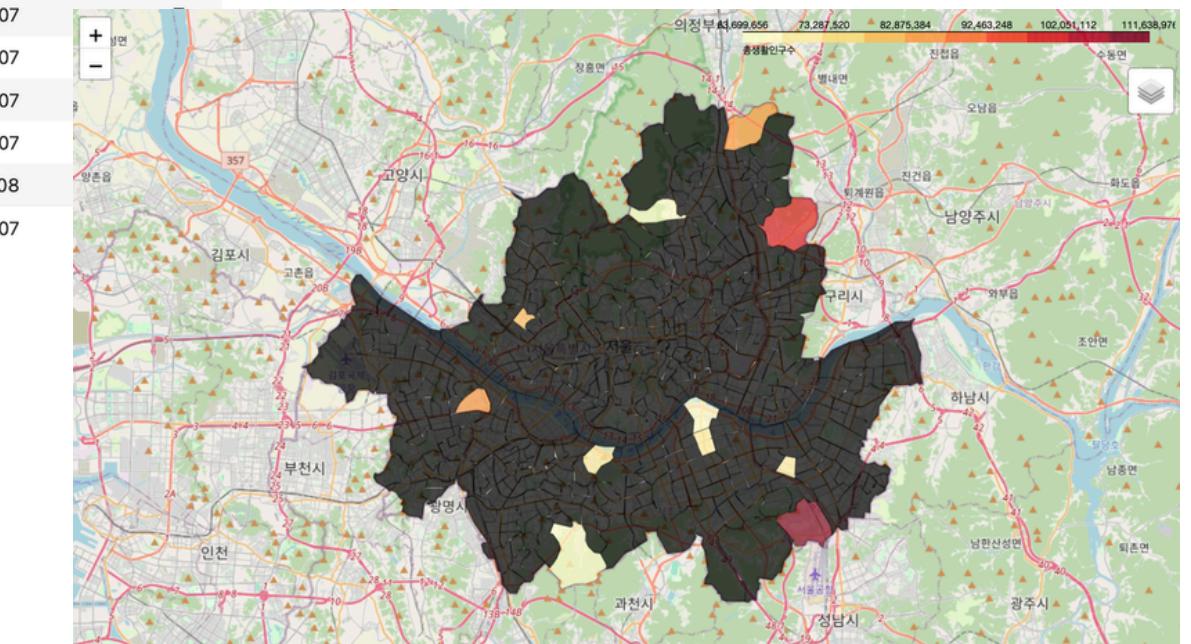
신규 노선 결정 방안

심야시간대 버스정류장이 존재하지 않는 행정동 시각화 (97개)



생활인구 데이터의 총생활인구수의 Q3 이상인 11개 행정동 도출 및 시각화

행정동코드	시구동	위도	경도	총생활인구수	총생활인구수_평균비교
142	11305660	서울특별시 강북구 인수동	37.637641	127.003963	6.605773e+07
161	11350600	서울특별시 노원구 공릉2동	37.631614	127.094228	1.020507e+08
168	11350630	서울특별시 노원구 상계1동	37.683268	127.065333	8.719409e+07
205	11410720	서울특별시 서대문구 북가좌2동	37.581004	126.913501	8.246107e+07
226	11470550	서울특별시 양천구 목5동	37.536311	126.881027	8.689748e+07
309	11590605	서울특별시 동작구 흑석동	37.506296	126.963763	7.108110e+07
335	11620735	서울특별시 관악구 대학동	37.453475	126.948482	6.369966e+07
359	11680531	서울특별시 강남구 논현2동	37.515206	127.036218	7.162780e+07
360	11680545	서울특별시 강남구 암구정동	37.530267	127.033484	6.599047e+07
374	11680700	서울특별시 강남구 세곡동	37.471075	127.104348	1.116390e+08
392	11710610	서울특별시 송파구 삼전동	37.501720	127.091906	7.251483e+07



- 2023년 버스 정류장 승하차량 데이터 분석
 - 서울시 내 행정동 중 '심야버스 운행 정류장'이 없는 지역 시각화 (검은색)
 - 총 97개 행정동 존재
- 2023년 생활인구 데이터 (행정동 총생활인구수)
 - 이 중 버스정류장이 없었던 행정동 중 생활 인구 데이터 총생활인구수의 Q3 이상인 11개 행정동 도출

- 11개 행정동
 - 심야버스 정류장이 없는 동시에 생활 인구가 서울 전체 행정동 분포의 Q3 이상에 해당하는 행정동
 - 하지만, 이것만으로 저 지역들로의 이동이 실제로 많은지 확인할 수는 없음.
 - 확인을 위해 '주요 도심과의 이동량'을 도출해야 함.

신규 노선 결정 방안

[주요 도심과의 이동량 확인 방법]

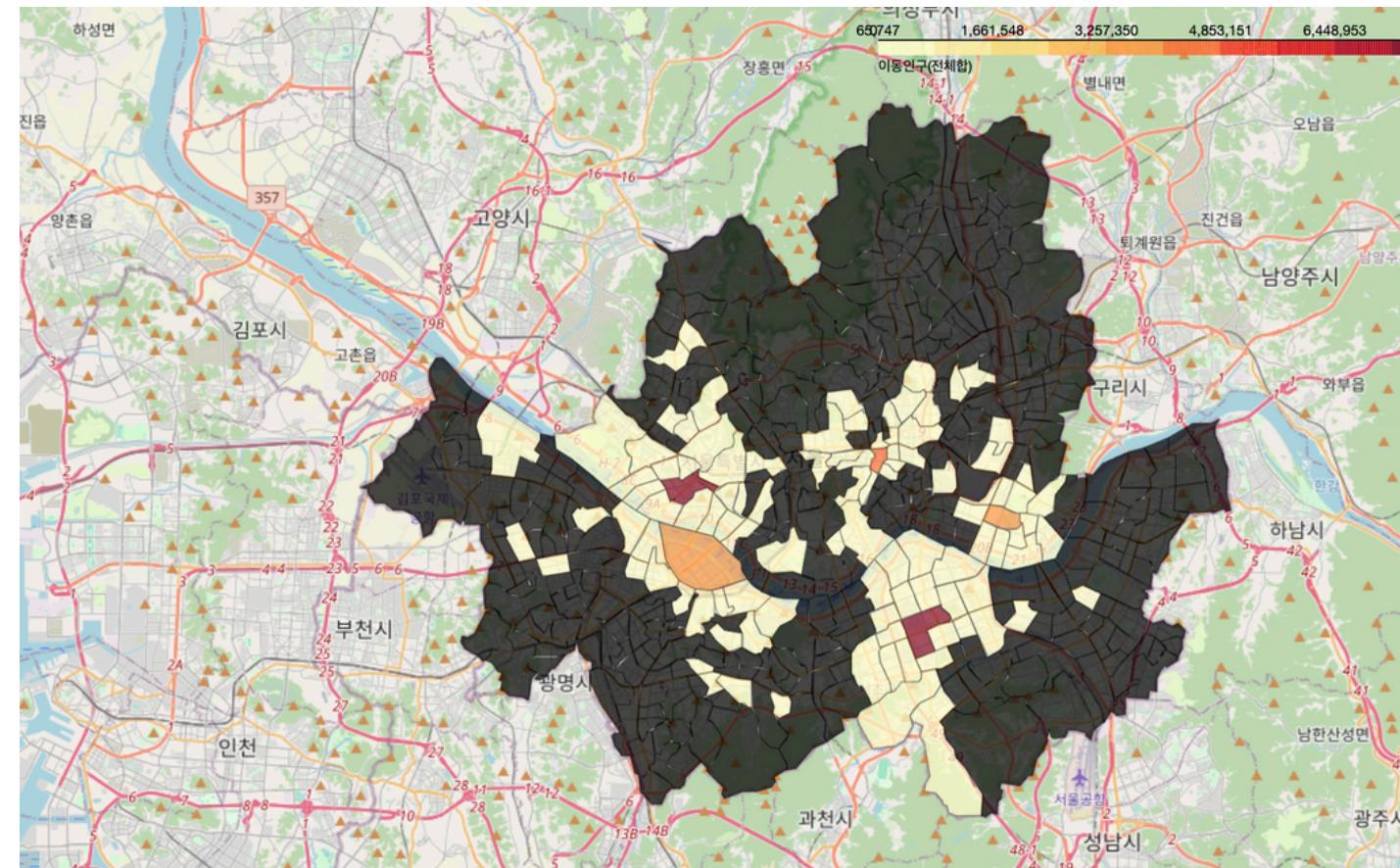
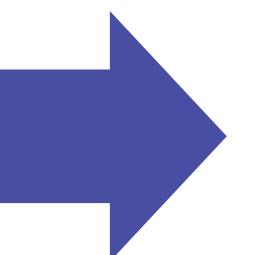
각 행정동별 주요 도심과의 총 이동량 합산

A	B
행정동	이동인구(전체합)
1	
2 서울특별시 금천구 가산동	55043.97
3 서울특별시 양천구 목3동	24563.67
4 서울특별시 서대문구 연희동	256086.88
5 서울특별시 서대문구 홍제3동	19047.96
6 서울특별시 성북구 장위1동	29371.68
7 서울특별시 성동구 금호2-3가동	41894.18
8 서울특별시 강동구 명일2동	8665.08
9 서울특별시 영등포구 신길5동	14416.68
10 서울특별시 송파구 거여1동	9857.94
11 서울특별시 서대문구 북가좌1동	32966.06
12 서울특별시 성동구 용답동	49664.66
13 서울특별시 노원구 하계1동	31099.34
14 서울특별시 성동구 옥수동	38557.36
15 서울특별시 강서구 가양1동	75857
16 서울특별시 동대문구 휘경2동	34601.07
17 서울특별시 서초구 방배2동	26603.03
18 서울특별시 동작구 노량진2동	77607.44
19 서울특별시 강서구 발산1동	33209.17
20 서울특별시 노원구 중계1동	10890.35
21 서울특별시 용산구 원효로2동	62316.19
22 서울특별시 광진구 군자동	276859
23 서울특별시 용산구 원효로1동	38830.45

출도착 순서 관계없이
각 행정동별 주요 도심과의 총 이동량 합산

[도심 순환선 구성 방안]

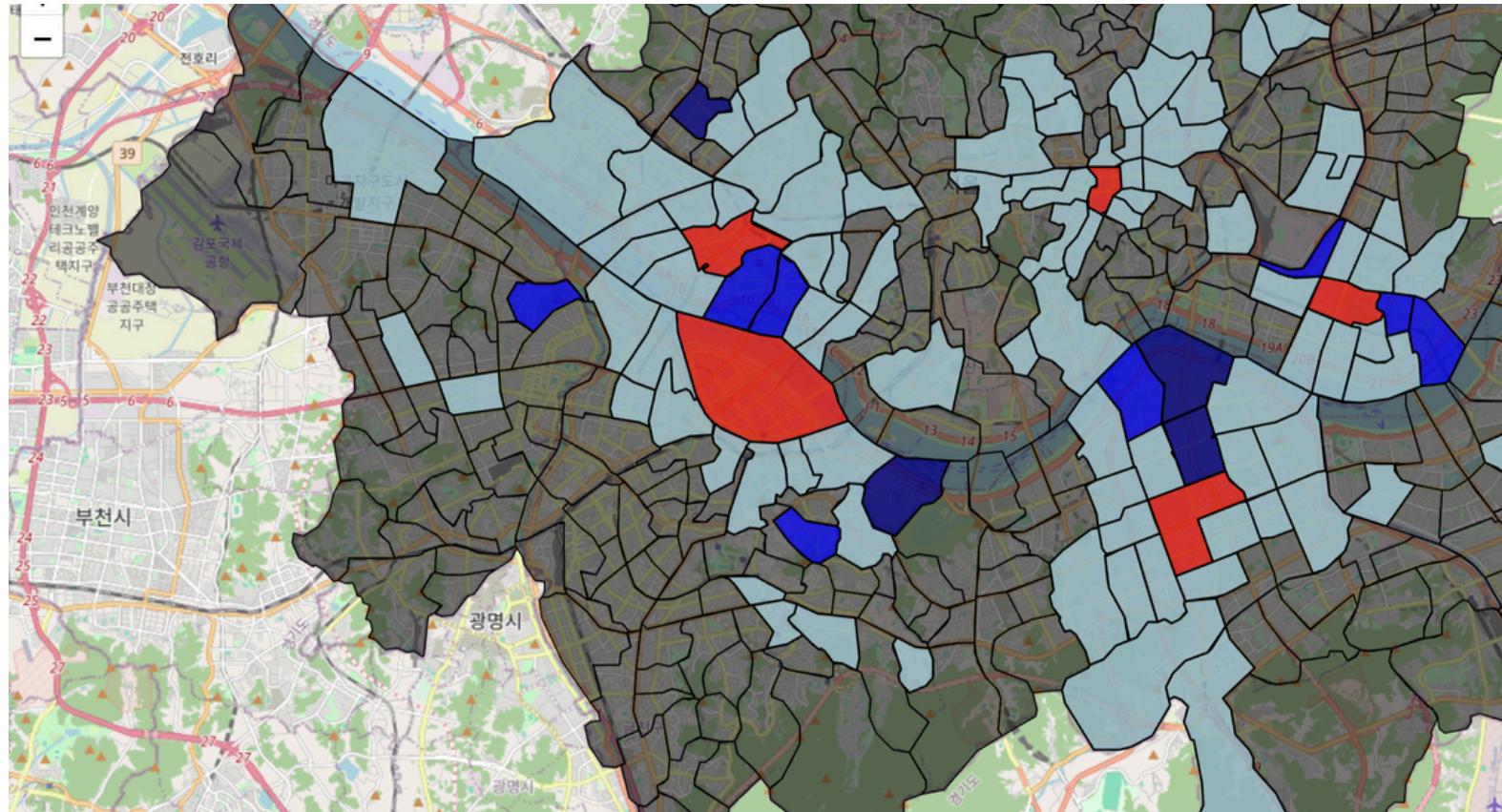
주요 도심과의 총 이동량 Q3 이상인 행정동 106개



- **빨강+주황** : 주요 도심
- **노랑** : 총 이동량 Q3 이상
- 검정 : 총 이동량 Q3 미만

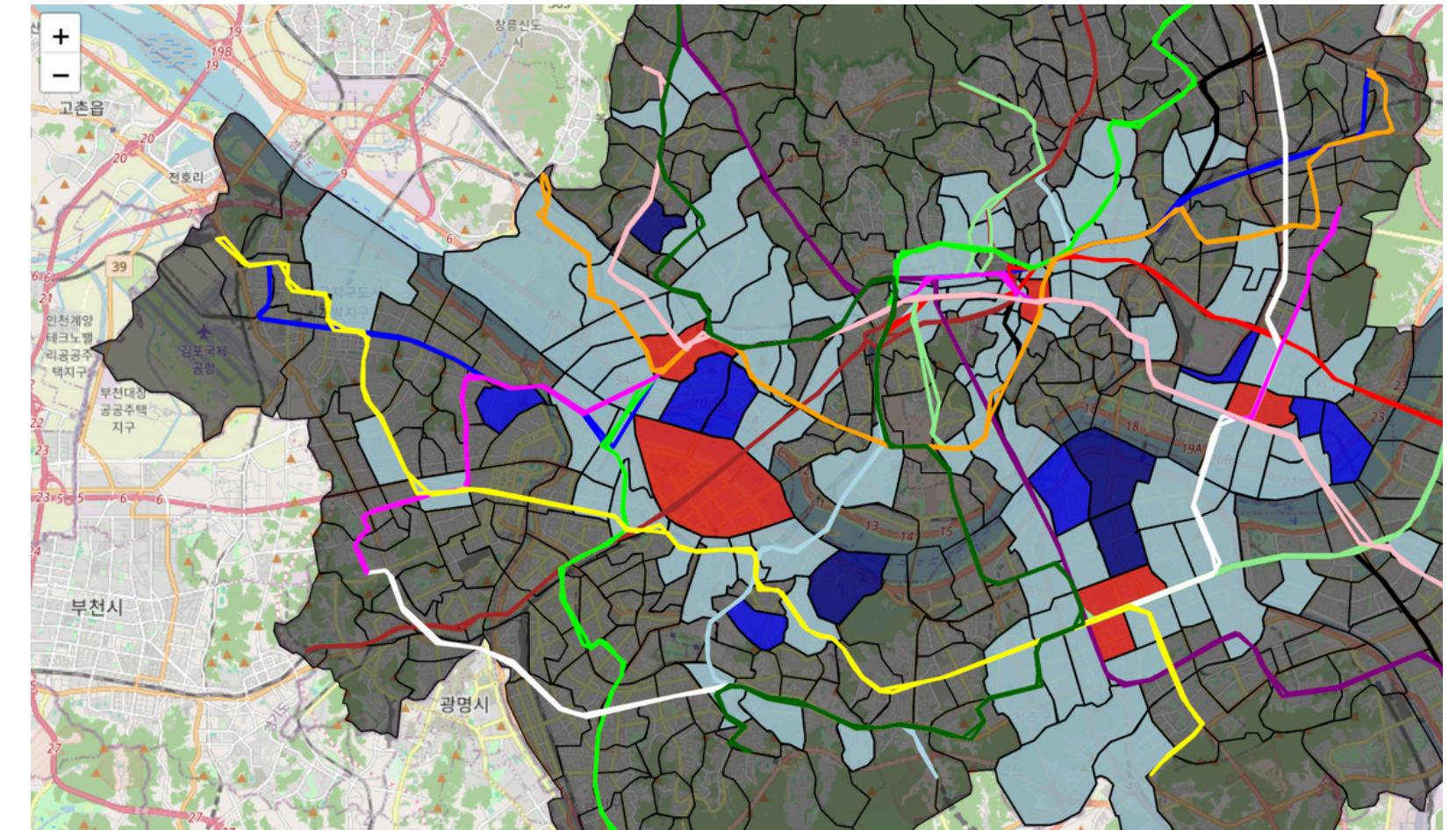
| 신규 노선 결정 방안

주요 도심과의 총 이동량 및 사각지대 중요도



- **빨간색** : 주요 도심 (홍대, 종로, 여의동, 건대, 강남)
- **파란색** : 심야버스 정류장이 없는 행정동
- **남색** : 파란색 + 생활인구 Q3 ↑ (사각지대)
=> 해당 지역 연결 필요성 ↑

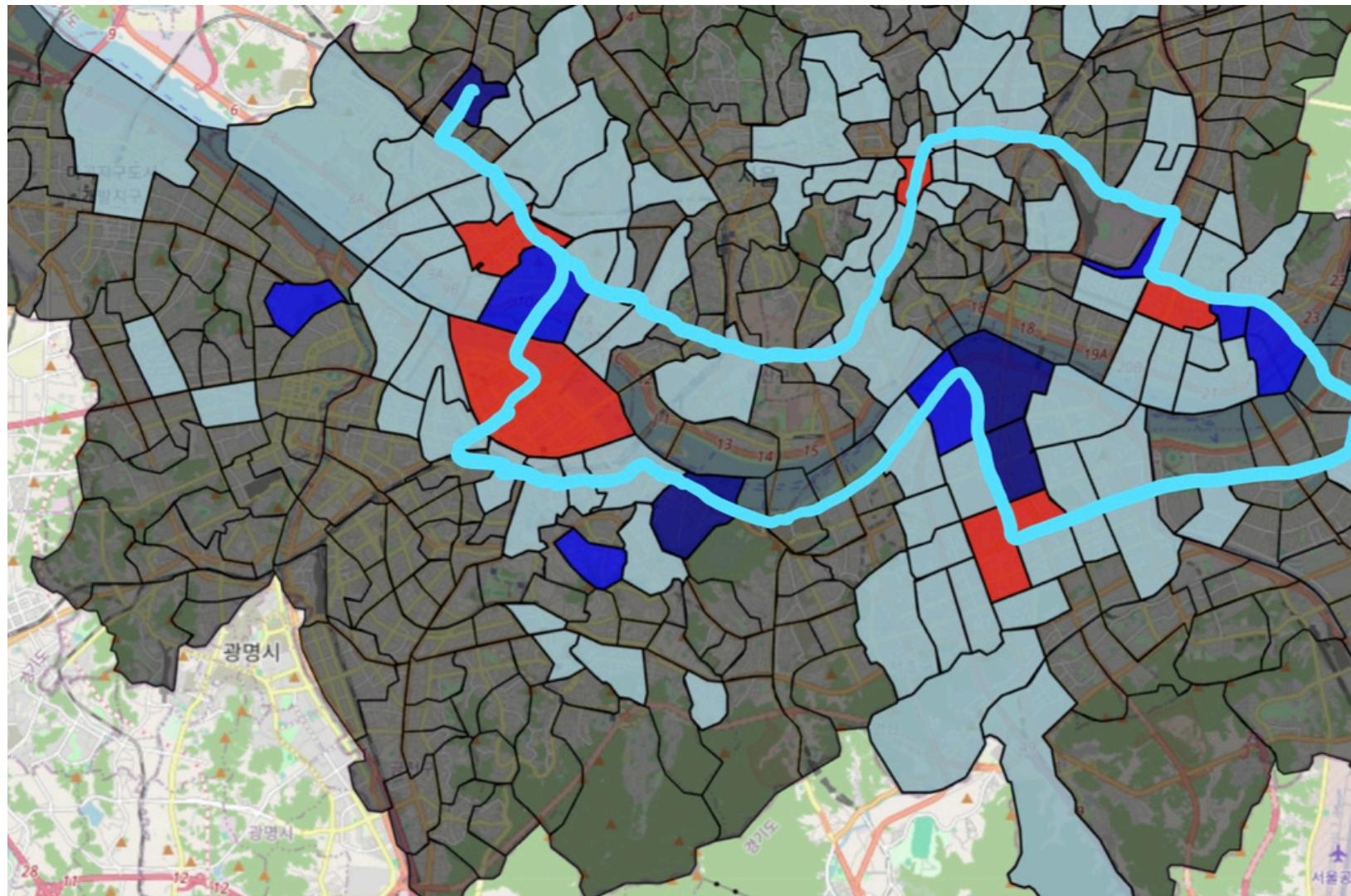
기존 노선도와의 연계성 파악



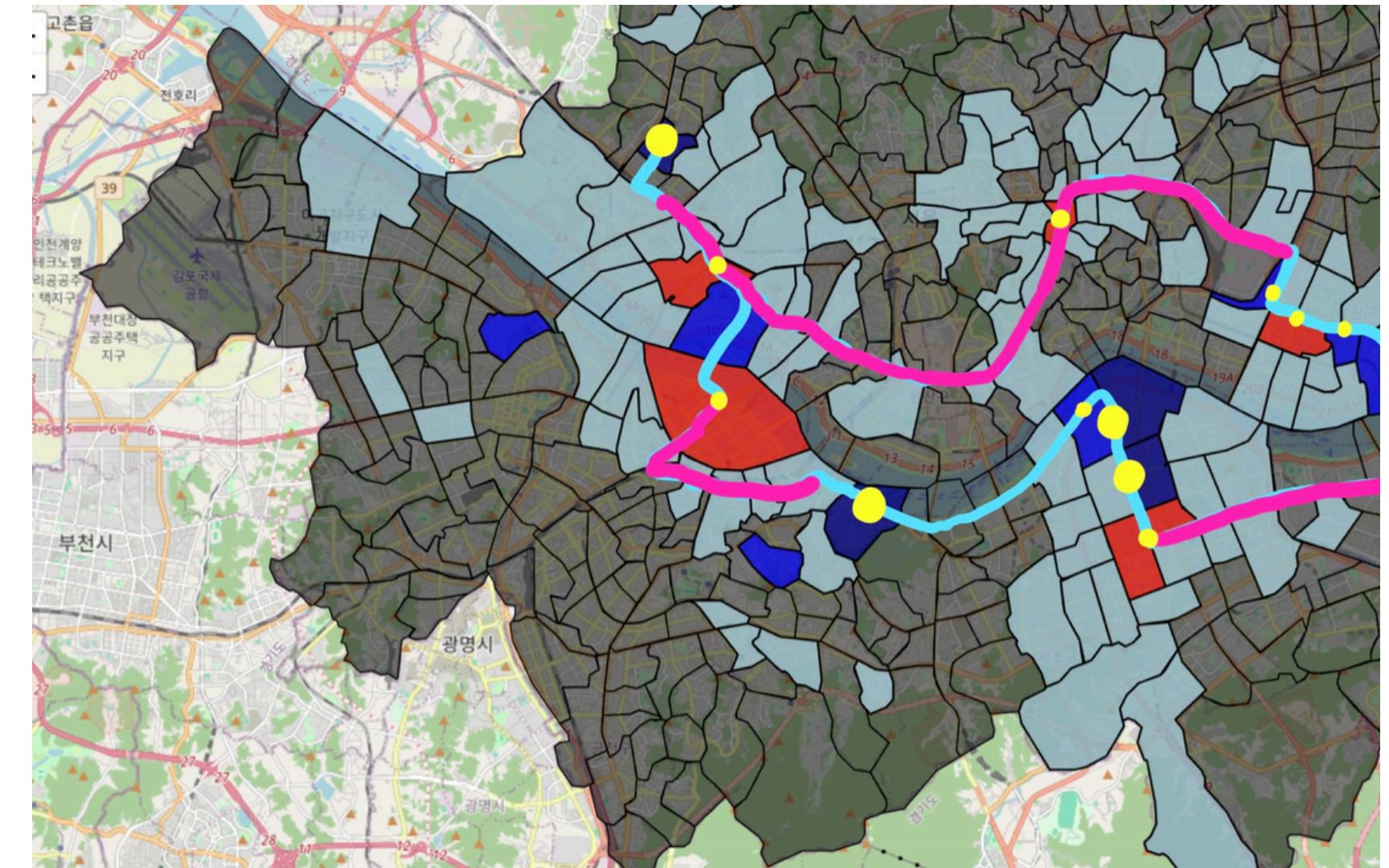
- 다양한 색의 선 → 기존 심야 버스 노선들
- 기존 노선과 일부 중복되는 노선을 만드는 이유
: **환승 연계성** + **승객 분산** → 버스 혼잡도 완화

| 신규 노선 결정 방안

최종 신규 노선



- **하늘색 선**: 주요 지역 (행정동)을 연결한 신규 노선
- 굴곡 노선 최소화 → 선형성 보장
- 주요 도심과 중요 사각지대의 연결성



- **큰 노란 원** : 중요 사각지대 (남색) 대표 지점
- **작은 노란 원** : 주요 도심 및 정류장이 없는 행정동 대표 지점
- **분홍색 선** : 기존 노선과 경로가 겹치는 신규 경로

| 결론 및 향후 과제

의의 및 한계, 향후 연구 방향

의의

수치적 증명을 통한 주요 사각지대 발견 → 사각지대 수요 충족

도심 순환선 개발 → 기존 방사형 노선과의 연결을 통한 환승 용이성

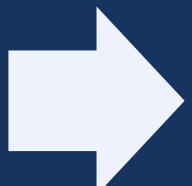
공공 빅데이터의 다각적 분석 → 향후 정부 교통 정책에 활용 가능

한계

행정동 간 노선 제안 → 도로까지 고려한 구체적 노선 계획 미비

택시 데이터 부재 → 심야 시간대 이동 실수요 파악 부족

노선의 편익 / 비용 계산의 어려움 → 현실성 파악 미흡

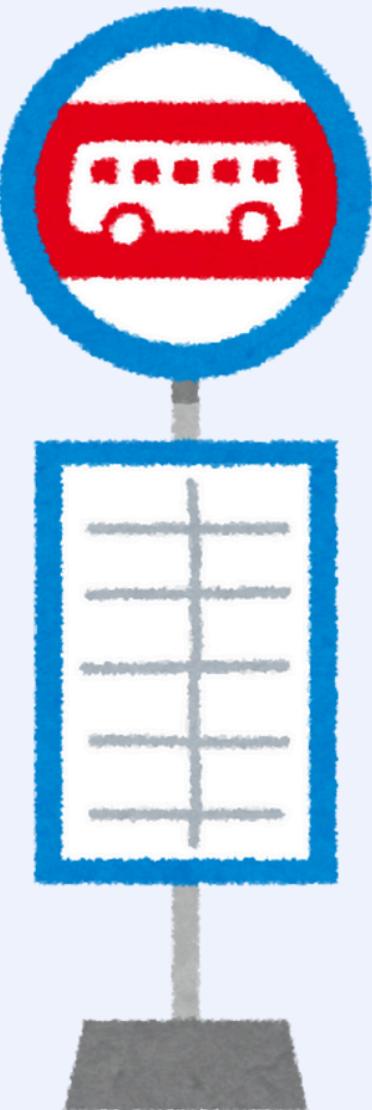


다양하고 세부적인 데이터셋을 활용하여 시간대별/권역별 특성에 맞는 구체적인 노선 선정 필요

| 참고 문헌 및 질의 응답

1. 강희찬 외 2명. (2016.09.22-23). 교통카드 데이터를 활용한 서울 심야버스 이용자 패턴. 대한교통학회, 제75회 학술발표회.
2. 도승일 외 3명. (2023.10.11-12). 택시통행 패턴 기반 심야버스 노선 계획 방안에 관한 연구.
대한교통학회 및 대한국토도시계획학회 공동학술대회.
3. 김민주, 이영인. (2017). 스마트카드 데이터를 이용한 심야버스 이용수요 특성 분석. 한국ITS학회논문지, 16(1), 101-116.
4. 정석호, 이정욱. (2019). 서울시 심야버스정책은 어떻게 성공적으로 도입되었는가?: 정치적 관리의 관점을 중심으로.
지방정부연구, 23(3), 261-283.
5. 서울열린데이터광장. (2023). 서울생활이동_메뉴얼_v3.1.
<https://data.seoul.go.kr/dataVisual/seoul/seoulLivingMigration.do>
6. 손정수. (2019). 빅데이터를 활용한 교통정책에 관한 연구: 서울시 심야(올빼미)버스 중심으로. 한국정책학회 학술대회.
7. 고승렬. (2021.11.10-11). 교통카드 빅데이터를 활용한 광역버스 노선 신설 방안 연구. 대한교통학회 제85회 학술발표회.
8. 서울특별시 빅데이터 캠퍼스. (2020). 빅데이터 기반 심야버스 노선 수립을 위한 분석.
https://bigdata.seoul.go.kr/noti/selectNoti.do?r_id=P260&bbs_seq=357&ac_type=A4&sch_type=
9. 정지원 기자, "심야 시간에도 마음 편히 집으로~" 서울시, 심야 버스 '올빼미버스' 확대 운행한다
<http://www.nwtn.co.kr/news/articleView.html?idxno=367284>
10. 서울시 상권분석 서비스, <https://golmok.seoul.go.kr/stateArea.do>
11. 김철수. (2021). 도시 발전과 환경 문제에 관한 연구. 환경과학연구, 34(2), 123-145.
12. 문화체육관광부. (2019). 문화정책 기본 방향.

| 참고 문헌 및 질의 응답



Questions?

E.O.D.

빅데이터를 활용한 서울 심야버스 노선 신설 방안