



데이터 기반의 도시문제 해결 플랫폼 COMPAS

고양시 공공자전거 스테이션 최적 위치 선정

백지민 엄윤경 이규복 이은솔 이형빈

Contents

01 / Problem Identification

- 연구 배경 및 목적
- 고양시 공공자전거 운영 현황
- 연구 방향

02 / Data Description

- 후보 스테이션 생성
- 학습 데이터 생성

03 / Data Analysis

- 군집화
- 그리드 별 점수 생성
- 거치대 수 계산
- 스테이션 최적화

04 / Results

- 스테이션 제안

매주 6시간 가량 팀 회의 및 구현
총 30일 가량 소요

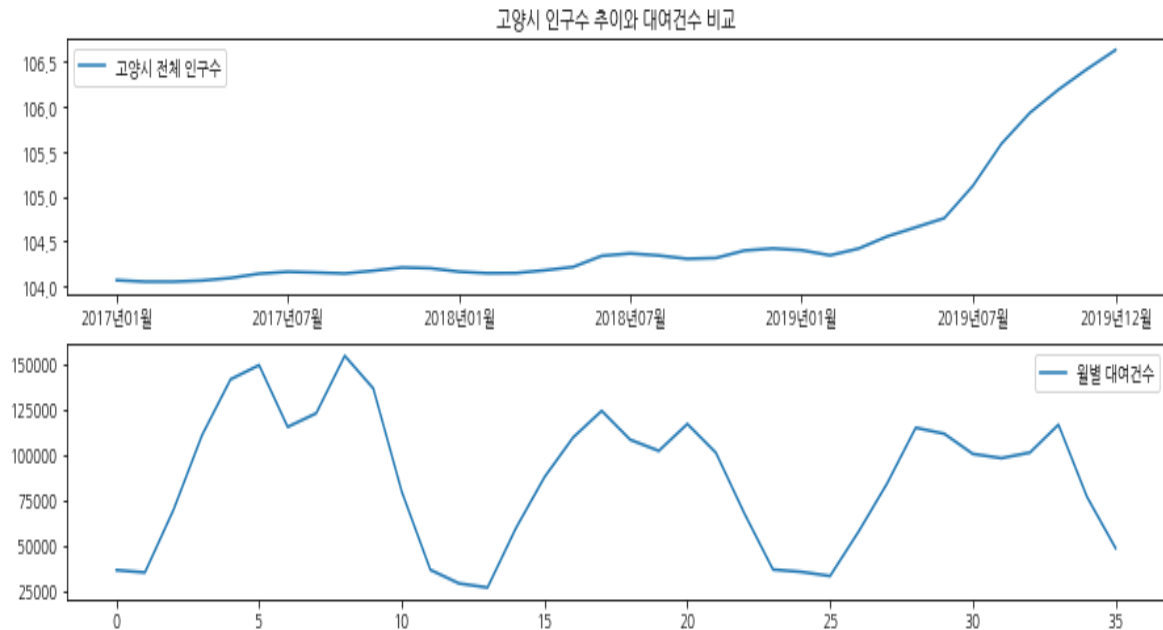
연구배경

- 고양시는 2010년 BOT(Build-Operation-Transfer) 방식의 공공자전거 서비스 피프틴 도입
- 약 161개 스테이션(자전거 보관소)과 1,700여대의 공공자전거로 시민들에게 서비스를 제공
- 시스템 노후화와 자전거 대여소 부족 등 시스템 투자에 대한 적극적 대응 미비로 이용자수 감소
- 수익창출의 어려움으로 사업 지속여부에 대한 고민
- 신규 택지개발 등으로 인한 도시화 지역과 인구의 증가로 기존 스테이션 위치에 대한 조정 필요

목적

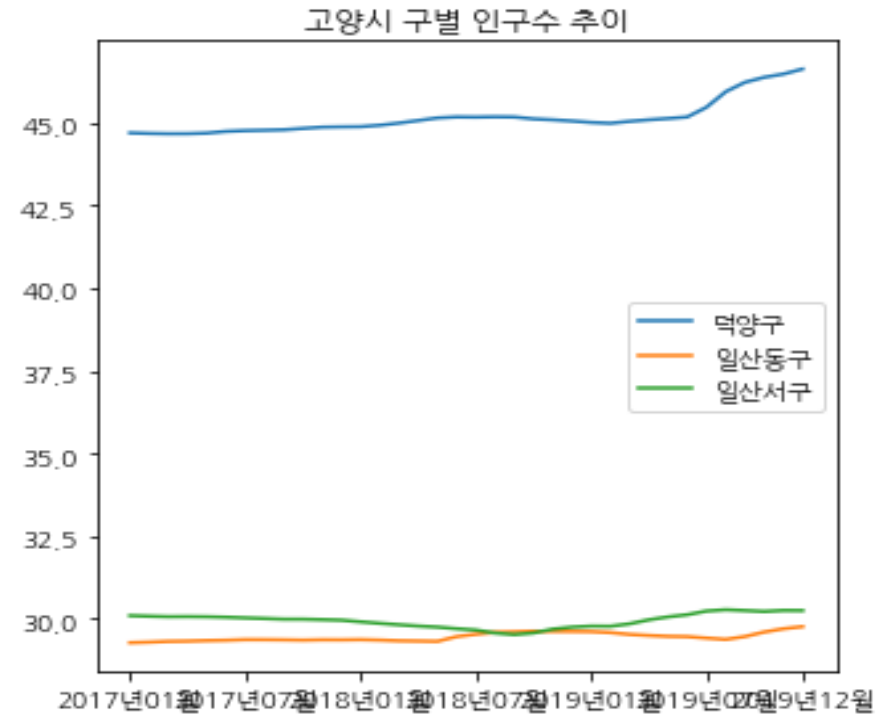
- 공공자전거 운영이력 데이터 및 공간 데이터를 활용하여 **자전거 스테이션의 최적 위치를 선정**하여 향후 시민들의 공공자전거(공유자전거) 사용에 대한 접근성 개선

인구수 추이와 월 별 대여건수 비교



- 고양시 인구는 2019년 이후로 급격한 증가세를 보이나 공공자전거 대여 건수는 감소하는 추세

고양시 구별 인구수 추이



- 덕양구 도시화지역과 일산동구 그리고 일산서구의 인구수 변화를 통하여 추가 스테이션의 필요성 증대

01

Problem Identification - 연구 방향

Solution1 현재 자전거 스테이션이 있는 곳은 어떤 특성을 가지고 있을까?

자전거 스테이션 별 주변 인구분포, 교통 접근성, 건물 연면적, 유형 별 건물 수와 자전거 대여량의 상관관계를 회귀분석을 통하여 조사

-> score가 0.47으로 유의미한 상관계수가 보이지 않았음

새로운 자전거스테이션 최적화 방안 필요

Solution2 무엇이 자전거스테이션을 존재하게 하는가?

자전거 스테이션이 있어야 하는 곳의 '입지'를 분석하자!

PHASE1. 후보 스테이션(그리드) 생성

분석기법: H3(고양시 전체를 육각타일들로 나누기)

PHASE2. 후보 그리드 별 점수화

- 그리드 별 특징이 담겨있는 학습 데이터 이용(해당 그리드 내에 어떤 것이 있는지)
- 그리드 최적의 군집 개수로 클러스터링
- 어떤 클러스터에서 어떤 클러스터로 이동했는지에 초점을 두고 후보 그리드 점수화

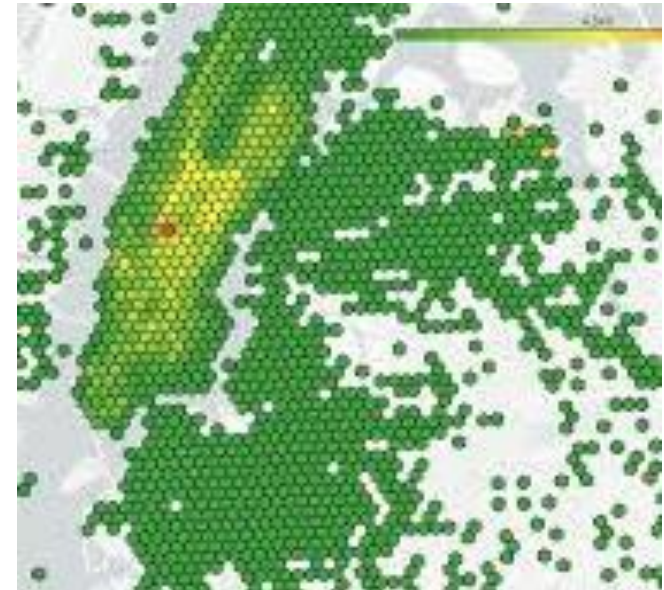
분석기법: Elbow Algorithm, 클러스터링, 연관 규칙 분석

PHASE3. 후보 그리드 중 점수가 상위 300개인 스테이션 추출

- > 최종 후보 스테이션 생성

PHASE4. 최종 후보 스테이션 각 각의 거치대 수 계산

분석기법: MinMaxScaler

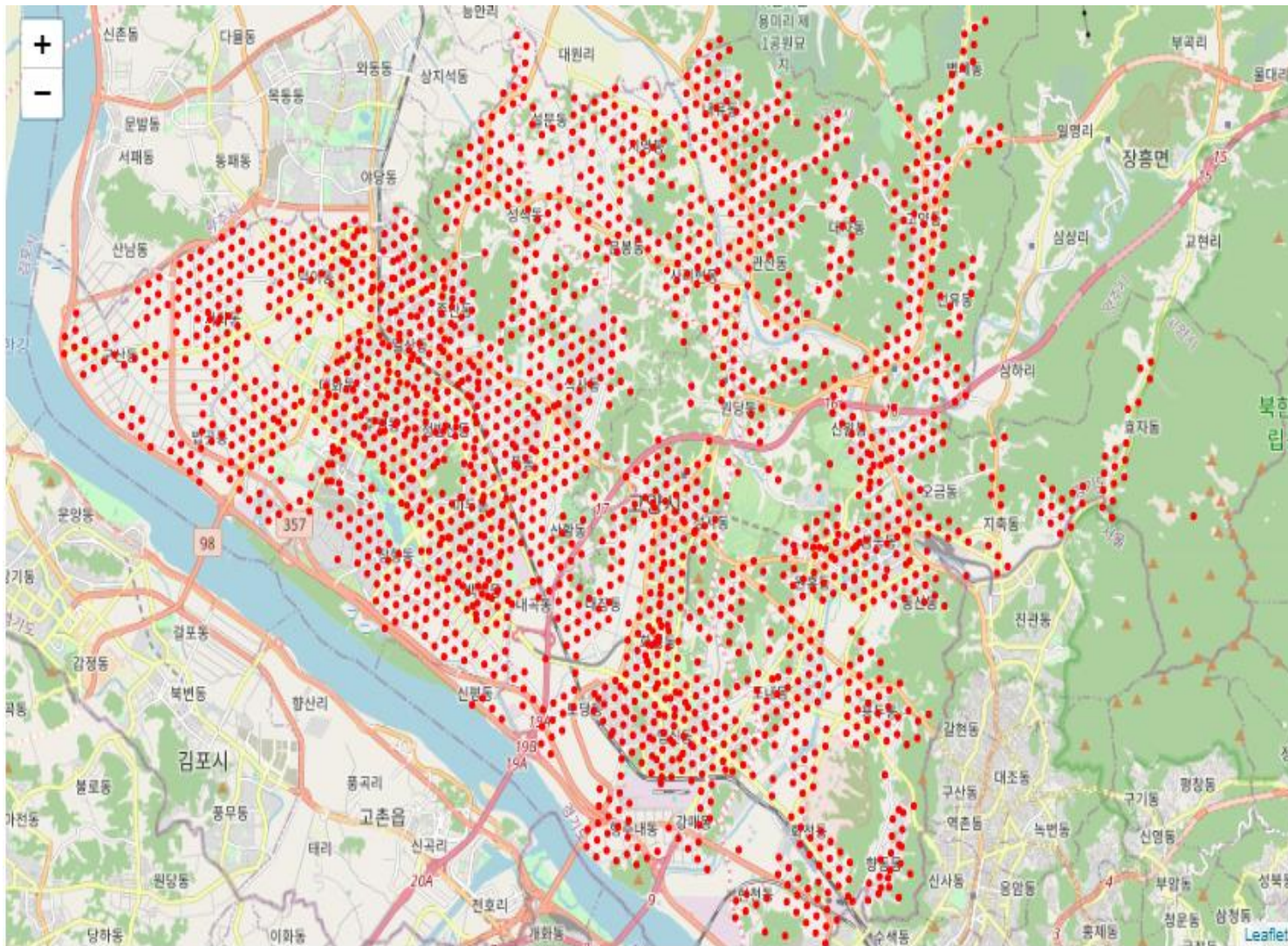
PHASE5. 도로 재배치를 통한 최종 후보 스테이션의 최적화

	전처리
공간시설	<ul style="list-style-type: none"> · 공간명 이상치 제거 ('? ? ? ? ?'으로 표기된 컬럼 제거) · 주차장 여부(0,1) 새로운 컬럼 생성
교통시설	<ul style="list-style-type: none"> · 지하철역별 승하차 인구 생성 · 지하철역별 동일 역 이름 통일 ('지축', '화전'은 '지축역', '화전(한국항공대)'로 통일)
일반 건물 분포도	<ul style="list-style-type: none"> · 연면적 0 초과인 데이터만 필터링
인구 분포도	<ul style="list-style-type: none"> · 인구분포가 0 초과인 것들만 필터링
도로명 주소 건물	<ul style="list-style-type: none"> · 건물 코드를 모두 한글 이름으로 변경 · 건물 여부(0,1) 새로운 컬럼 생성

※ 이후 분석모델마다 전처리 진행

02

Data Description - 그리드(후보스테이션) 생성



고양시 전체를 육각형 모양의 H3 모델로
나누어 후보 입지들을 기존 스테이션 위치에
덧붙여 추가

H3 알고리즘을 통한 1804개 블록과 기존
스테이션 164개의 블록을 합쳐 총 1968개
블록 생성

그리드의 중심 좌표 시각화

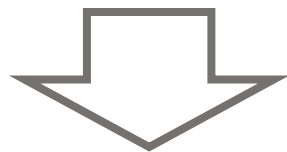
고양시 공공자전거 정책방향 수립을 위한 연구에서 이루어진 공공자전거 이용 이유에 대한 설문조사 결과로 건강 증진과 대체 교통의 대답이 많았음

따라서 공원, 주차장, 버스정류장, 지하철 등의 입지 데이터를 구하기 위하여 공원, 버스정류소_탑승인구합, 전철역_승차인구합, 전철역_하차인구합을 학습 데이터의 변수로 생성

학습데이터 변수	설명
인구분포	해당 그리드의 인구 분포
공원	해당 그리드에 있는 공원의 수
버스정류소_탑승인구합	해당 그리드에 있는 버스정류소들의 탑승인구합
전철역_승차인구합	해당 그리드에 있는 전철역들의 승차인구합
전철역_하차인구합	해당 그리드에 있는 전철역들의 하차인구합
주차장	해당 그리드에 있는 주차장의 수

학습데이터 변수	설명
건물 연면적합	해당 그리드에 있는 건물들의 연면적합
건물 총 개수	해당 그리드에 있는 건물들의 총 개수
건물 유형 X	해당 그리드에 있는 건물 X의 수 (163개의 건물 유형) (ex. 건물 유형이 아파트일 때 해당 그리드에 있는 아파트의 수)

서점(1종근,생미해당), 독서실, 기원, 노래연습장, 일반음식점, 휴게음식점, 제조업소, 수리점, 사진관, 표구점, 학원, 동물병원, 자동차영업소, 세탁소, 고시원 등 163가지의 건물 유형을 대분류로 분류하여 넣는 작업의 필요성



행정표준관리시스템의 '건축물 용도 코드'를 기준으로 분류

건물 대분류	세부 분류
비공동주택	단독주택, 다중주택, 다가구주택, 공간
공동주택	아파트, 연립주택, 다세대주택, 생활편익시설, 부대시설, 복리시설, 기숙사
제1종근린	마을공동구판장, 마을공동작업소, 조산원, 소매점, 휴게음식점, 의원, 침술원
제1종근린_공공시설	보건소, 경찰서, 파출소, 소방서, 우체국, 방송국, 기타공공시설, 공공시설, 공공도서관
제2종근린	서점(1종근,생미해당), 독서실, 기원, 노래연습장, 일반음식점, 휴게음식점, 제조업소, 수리점, 사진관, 표구점, 학원, 동물병원, 자동차영업소, 세탁소, 고시원
제2종근린_운동시설	체력단련장, 골프연습장, 기타운동시설, 당구장, 테니스장,
제2종근린_공연장	음악당, 연예장, 극장, 영화관, 기타공연장
제2종근린_사무소	금융업소, 사무소, 부동산중개업소, 기타사무소, 기타제2종근생활시설, 출판사
문화집회	기타문화및집회시설, 운동장, 기타전시장
문화집회_공연장	서어커스장, 비디오물감상실, 비디오물소극장
문화집회_동식물원	동물원, 식물원, 기타동,식물원, 작물재배사, 기타동식물관련시설
문화집회_전시장	과학관, 기념관
문화집회_관람장	체육관, 경마장

문화집회_집회장	기타집회장
종교시설	교회, 성당, 사찰, 기타종교집회장, 기타종교시설, 종교집회장,
판매시설	쇼핑센터, 소매점, 기타판매시설, 백화점, 시장, 기타소매시장, 대형점, 농수산물공판장
운수시설	여객자동차터미널, 공항시설, 기타운수시설
의료시설	기타의료시설, 의원, 한의원, 치과의원
창고시설	물류터미널, 집배송시설, 기타창고시설, 하역장
교육연구시설	학원, 중학교, 기타학교, 교육원, 연수원, 사무소, 기타교육원
노유자시설	영유아보육시설, 기타아동관련시설, 노인복지시설
위험물저장 및 처리시설	위험물제조소, 위험물저장소, 액화가스취급소, 기타위험물저장처리시설, 액화가스판매소
유해공장	기타공장
위락시설	단란주점, 단란주점 , 유흥주점, 기타위탁시설, 투전기업소, 무도장, 유원시설업의시설, 주점영업
관광휴게시설	공원, 기타관광휴게시설
숙박시설	기타숙박시설
자동차관련시설	주차장, 기타자동차관련시설, 세차장, 폐차장

02

Data Description - 학습 데이터 생성

	인구분포	건물연면적합	건물총개수	정류소승인	전철역 승차인구합	전철역 하차인구합	공동주택비동	공동주택	제1종근린종합	제1종근린 공공시설	제2종근린	제2종근린 운동시설	제2종근린 공원장	제2종근린사무소	문화집회	문화집회 공연장	문화집회 동식물원	문화집회 전시장	문화집회 관람장	문화집회 집회장	종교시설	판매시설	운수시설	의료시설	창고시설	교육연구시설	노유자시설	위험물저장및처리시설	유해공장	위락시설	관광휴게시설	숙박시설	자동차관련시설
0	0.0	0.00	4.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
1	72.0	491.56	58.0	0.0	0.0	0.0	28.0	1.0	4.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	
2	93.0	863.49	66.0	217.0	0.0	0.0	26.0	8.0	16.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	14.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
3	288.0	1055.93	136.0	958.0	0.0	0.0	72.0	8.0	15.0	0.0	0.0	1.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	14.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	
4	210.0	1218.26	84.0	236.0	0.0	0.0	36.0	16.0	13.0	0.0	2.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	12.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	
...
1963	2296.0	79197.57	12.0	15586.0	0.0	0.0	0.0	9.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
1964	1587.0	31635.55	17.0	13995.0	0.0	0.0	3.0	5.0	0.0	0.0	0.0	0.0	0.0	3.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
1965	1612.0	87112.67	8.0	14821.0	0.0	0.0	0.0	8.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
1966	3347.0	98056.66	8.0	3263.0	0.0	0.0	0.0	7.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
1967	616.0	8537.14	71.0	194.0	0.0	0.0	34.0	0.0	1.0	0.0	4.0	0.0	0.0	21.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	1.0	0.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

1968 rows × 33 columns

03

Data Analysis - 군집 개수 결정

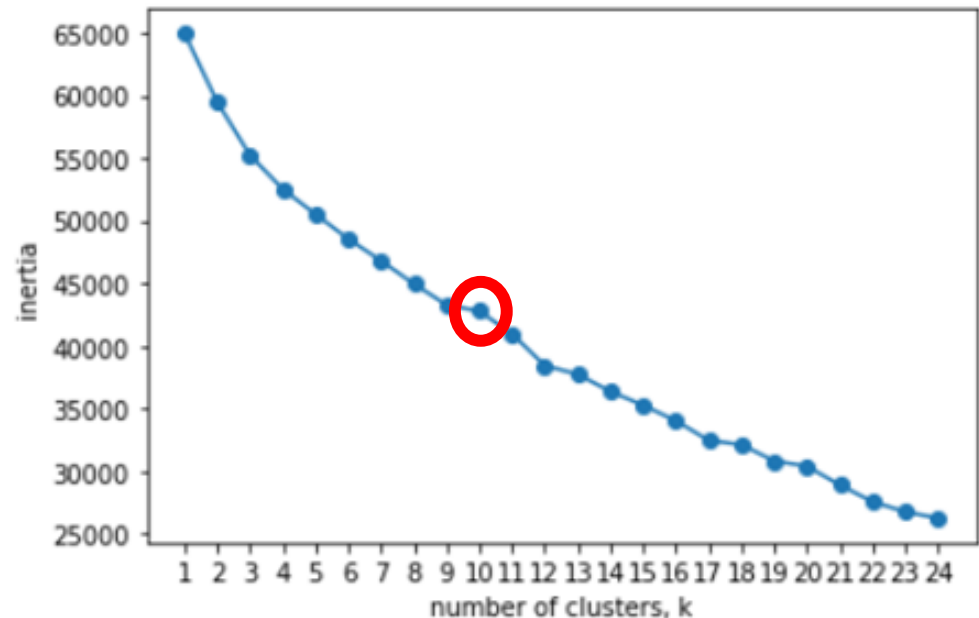
Elbow Algorithm을 이용한 최적 군집 개수 선정

Elbow Algorithm이란 클러스터링의 평가 척도 방법 중 하나로, 클러스터링 성능을 최적화하는 파라미터를 관찰

K-means를 통해 구한 군집 내에서, k값의 변화에 따른 SSE를 관찰 (SSE: 각 데이터로부터 그 데이터가 속한 그룹의 무게중심까지의 거리(오차)의 제곱 합)

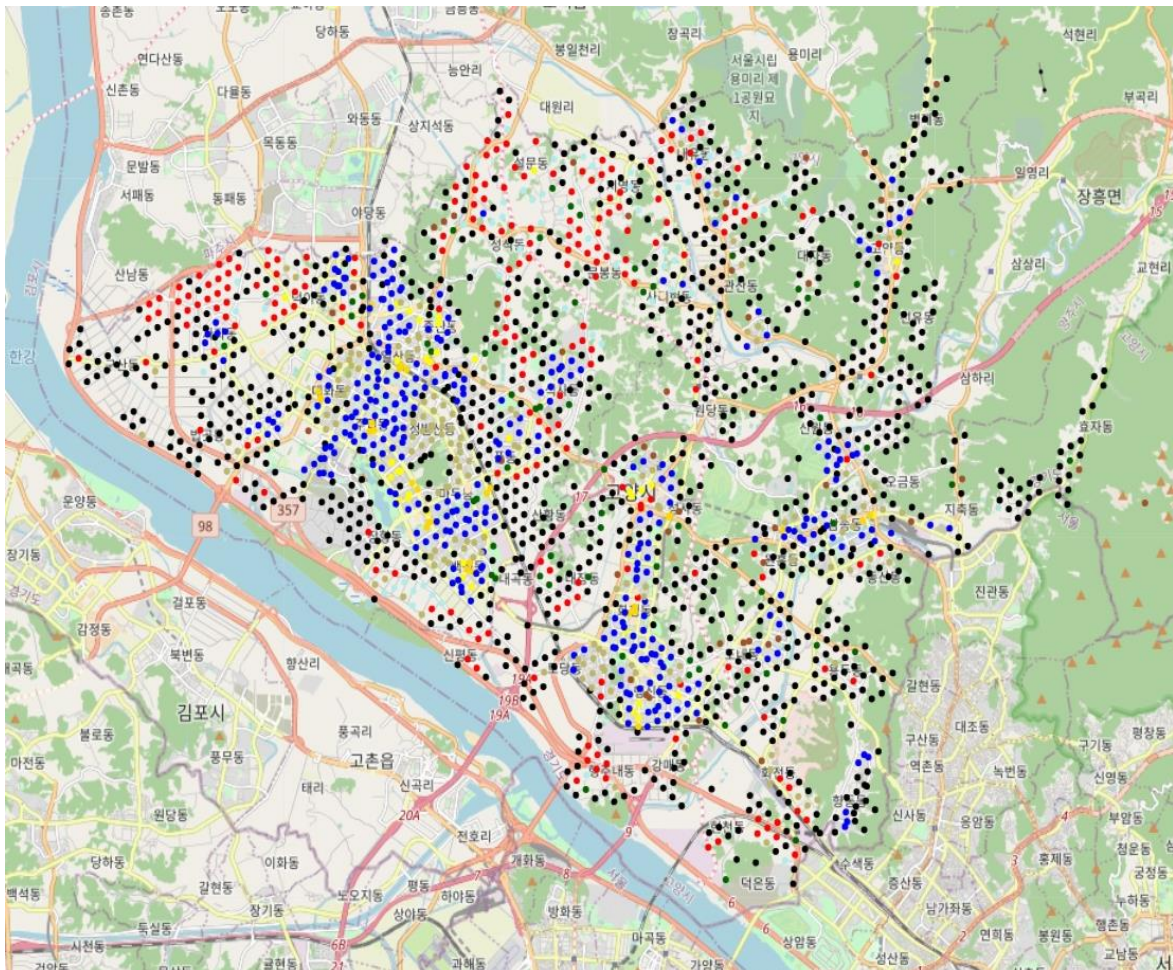
cost가 감소하다가 어느 순간 정체되는 지점을 클러스터 수로 설정

클러스터 범위를 1~25로 설정했을 때, 클러스터 개수가 k=10일 때 cost가 정체되는 것을 확인



03

Data Analysis - 군집화



-학습데이터를 정규화의 과정을 거쳐 10개의 군집으로
K-means 클러스터링

	인구분포	건물연면적 적합	건물총개 수	버스정류 소_탑승인 구합	전철역_승 차인구합	전철역_하 차인구합
분류						
0	-0.264951	-0.199158	-0.001151	-0.044509	-0.122258	-0.121380
1	-0.583065	-0.414072	0.913866	-0.207573	-0.122258	-0.121380
2	0.699666	1.142523	0.498480	0.312069	-0.122258	-0.121380
3	-0.512869	-0.357922	-0.436888	-0.181209	-0.102954	-0.102221
4	-0.434137	-0.373440	0.865558	-0.135449	-0.122258	-0.121380
5	0.325794	1.379990	0.049579	<u>5.641093</u>	<u>6.978297</u>	<u>6.942301</u>
6	-0.194220	-0.126317	-0.166493	-0.127579	-0.122258	-0.121380
7	0.499953	0.360235	0.193386	0.210199	-0.122258	-0.121380
8	<u>1.685031</u>	1.294591	-0.385697	0.152725	-0.106671	-0.106530
9	0.445117	-0.257650	2.229087	-0.021328	-0.104259	-0.104951

예) 5번 군집에 해당하는 입지는 다른 군집에 비해 교통량이 높다

예) 8번 군집에 해당하는 입지는 다른 군집에 비해 인구분포가 높다

03

Data Analysis - 군집 매핑

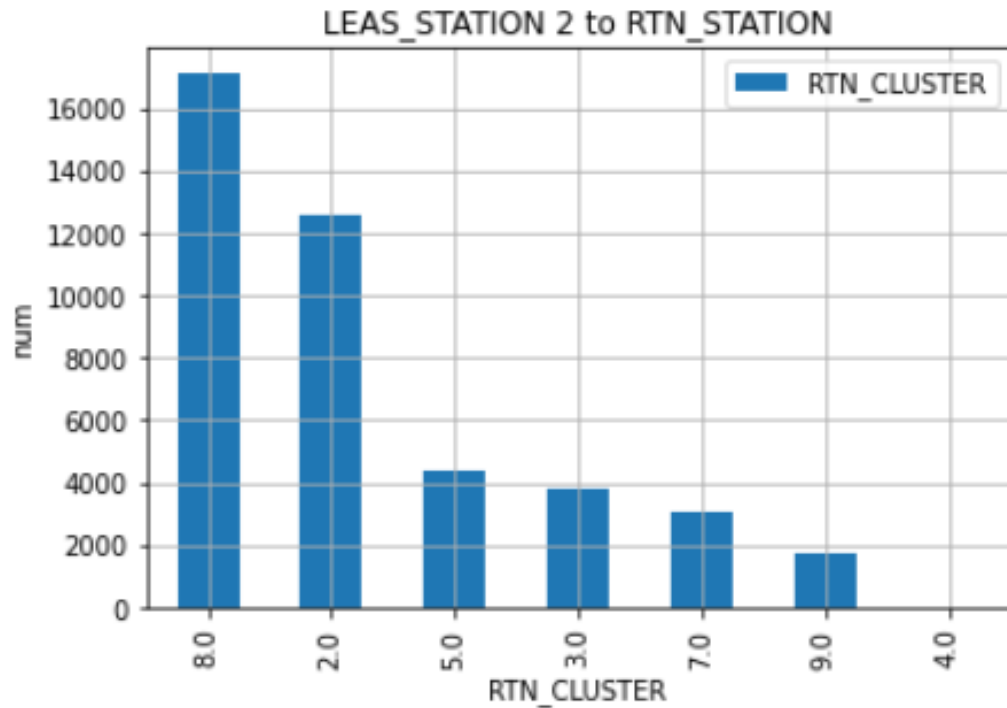
어떤 입지를 가진 스테이션(클러스터)에서 어떤 입지를 가진 스테이션(클러스터)으로 이동이 많았을까?

클러스터링된 후보 그리드에서 추출한 클러스터 값을 참고하여 운영이력에 있는 데이터에 각각 대여스테이션과 반납스테이션이 어느 클러스터에 속하는지 매핑
(운영이력에 RTN_STATION(반납스테이션)이 0인 이상치는 제거해준 뒤 매핑)

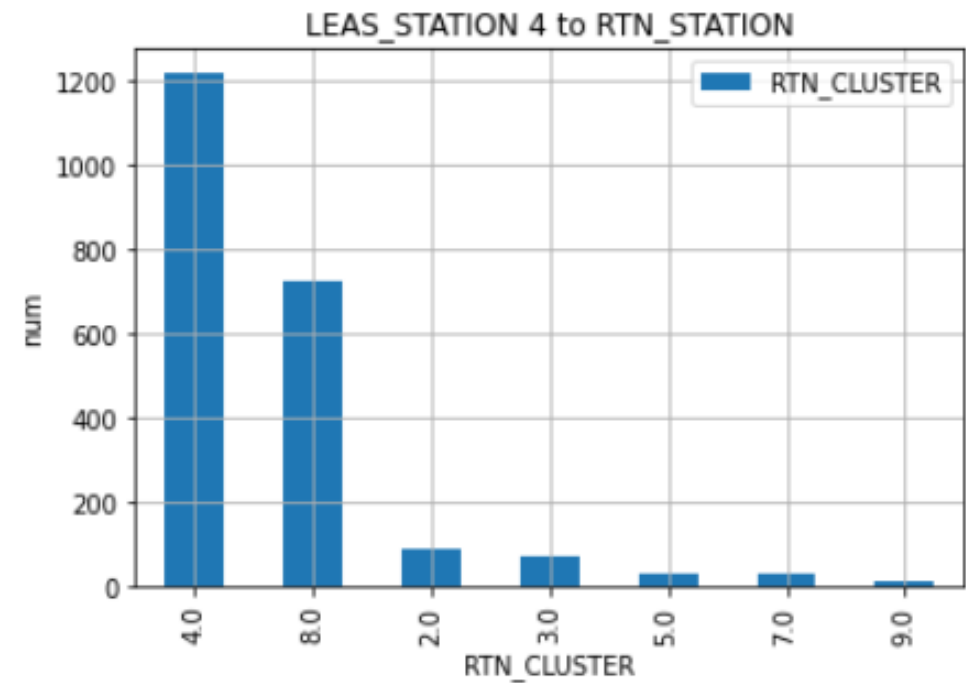
→
운영이력 데이터의 칼럼에 LEAS_CLUSTER(대여 스테이션이 속한 클러스터)와 RTN_CLUSTER(반납 스테이션이 속한 클러스터)를 추가한 예시

LEAS_CLUSTER	RTN_CLUSTER
5.0	8.0
5.0	7.0
5.0	2.0
5.0	7.0
5.0	8.0
...	...
5.0	8.0
5.0	8.0
5.0	8.0
5.0	5.0
5.0	8.0

클러스터 10개에 대한 대여스테이션과 반납 스테이션의 연결성 시각화 예시



- 대여 스테이션이 클러스터 2일 때 클러스터 8에 속한 스테이션에 가장 반납을 많이 함



- 대여 스테이션이 클러스터 4일 때 클러스터 4에 속한 스테이션에 가장 반납을 많이 함

대여스테이션과 반납 스테이션의 연결성에 대해 점수 부여

대여스테이션 분류 10가지 * 반납 스테이션 분류 10가지로 총 100가지의 점수표 생성

대여스테이션 분류 x에서 반납스테이션 분류 y로의 점수 ($x \rightarrow y$ 신뢰도)

(대여스테이션 분류가 x이고 반납스테이션 분류가 y인 운영이력의 개수) /

(대여스테이션 분류가 x인 운영이력의 총 개수) * 100

*분류 x에 속하는 스테이션의 운영이력이 존재하지 않는다면 0점으로 환산, 점수는 소수점 둘째 자리까지

ex) 19번째 행의 경우 대여스테이션 분류가 1이고 반납스테이션 분류가 9인 운영이력의 개수
/ 대여스테이션 분류가 1인 운영이력의 개수 * 100으로 계산

	LEAS_CLUSTER	RTN_CLUSTER	SCORE
0	0	0	0
1	0	1	0
2	0	2	0
3	0	3	0
4	0	4	0
5	0	5	0
6	0	6	0
7	0	7	0
8	0	8	0
9	0	9	0
10	1	0	0
11	1	1	32.03
12	1	2	8.25
13	1	3	0
14	1	4	1.1
15	1	5	0.13
16	1	6	24.9
17	1	7	13.66
18	1	8	6.43
19	1	9	13.49

*최빈 군집은 데이터가 적기 때문에 입력되지 않거나, 유동인구가 없으며, 도시화가 안된 군집으로 판단되어 제거

후보 그리드와 인근 1km~5km 사이에 있는 그리드에 점수환산표 기반 점수 부여(서울 공공자전거의 평균 이용거리를 참고)
지정 거리 (1km~5km) 내에 있는 그리드 간의 link 점수 평균값

빨간 점이 찍혀 있는 후보 그리드의 점수 산정 방법

12시에서 시작해 시계방향으로

5 → 2 2894점

4 → 2 68점

5 → 2 2894점

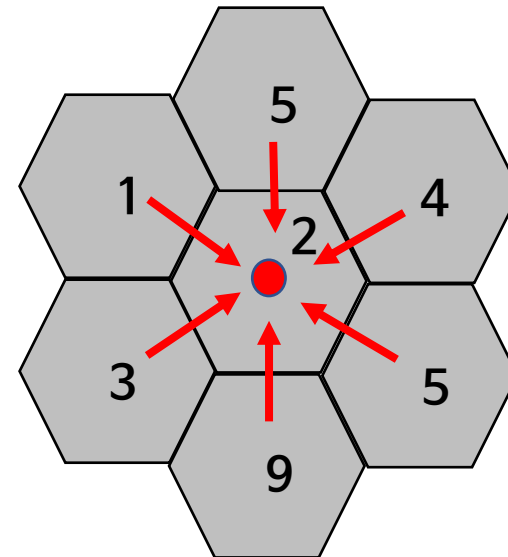
9 → 2 1855점

3 → 2 5731점

1 → 2 0점

다 더하면 13442이 되고 13442에서 6으로 나눈 평균은 2240.33. 따라서 가운데 빨간 점이 찍혀 있는 후보 그리드의 점수는 2240.33점이 됨

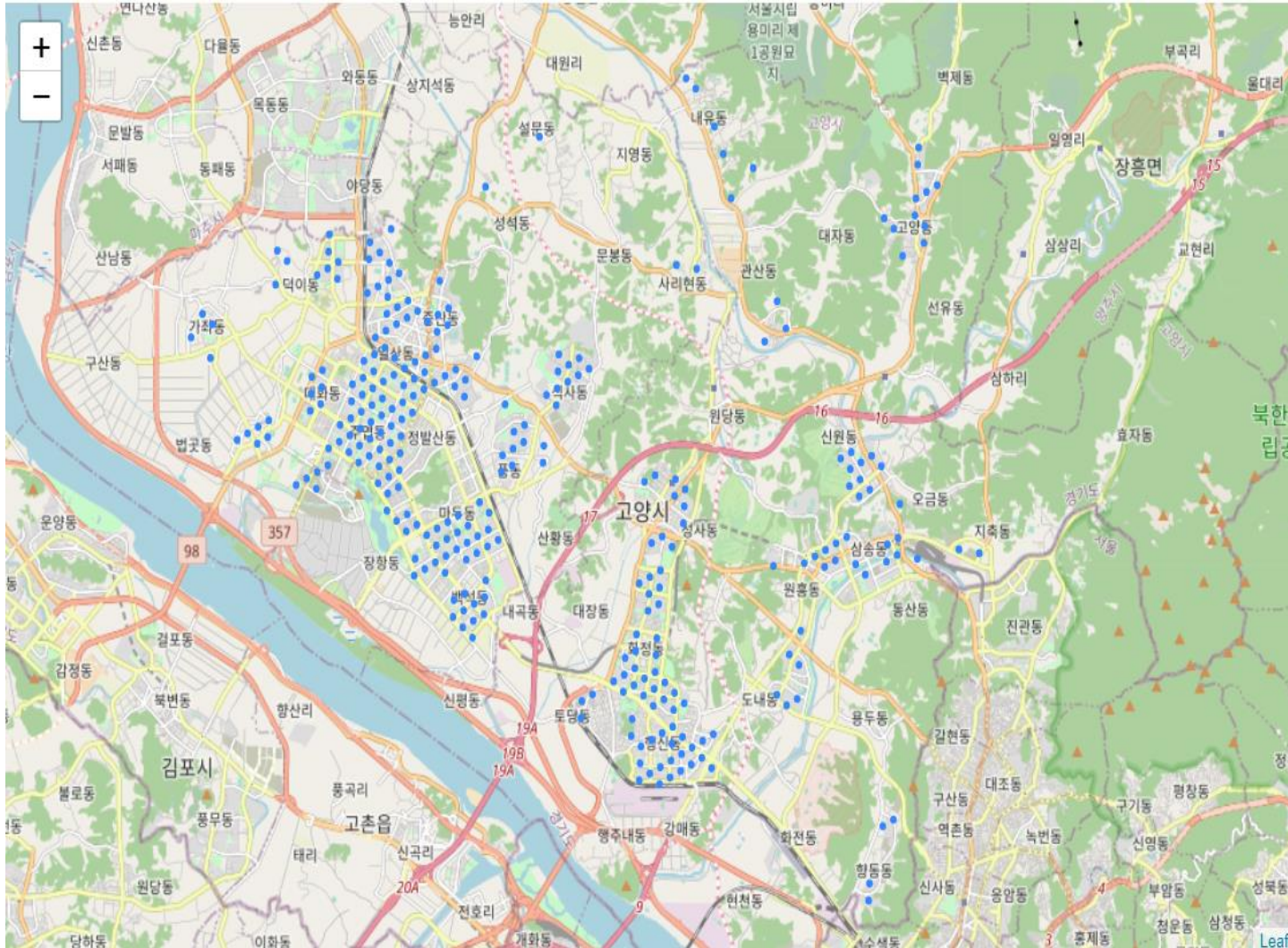
*각 그리드 안에 써있는 숫자는 각 그리드의 클러스터를 뜻함



LEAS_CLUSTER	RTN_CLUSTER	SCORE
0	2	0
1	2	0
2	2	8843
3	2	5731
4	2	68
5	2	2894
6	2	0
7	2	813
8	2	20783
9	2	1855

03

Data Analysis - 최종 후보 스테이션



군집화 및 점수 환산 이후, 다시 위경도
데이터와 연결하여 후보 스테이션 중
상위 300위를 선정

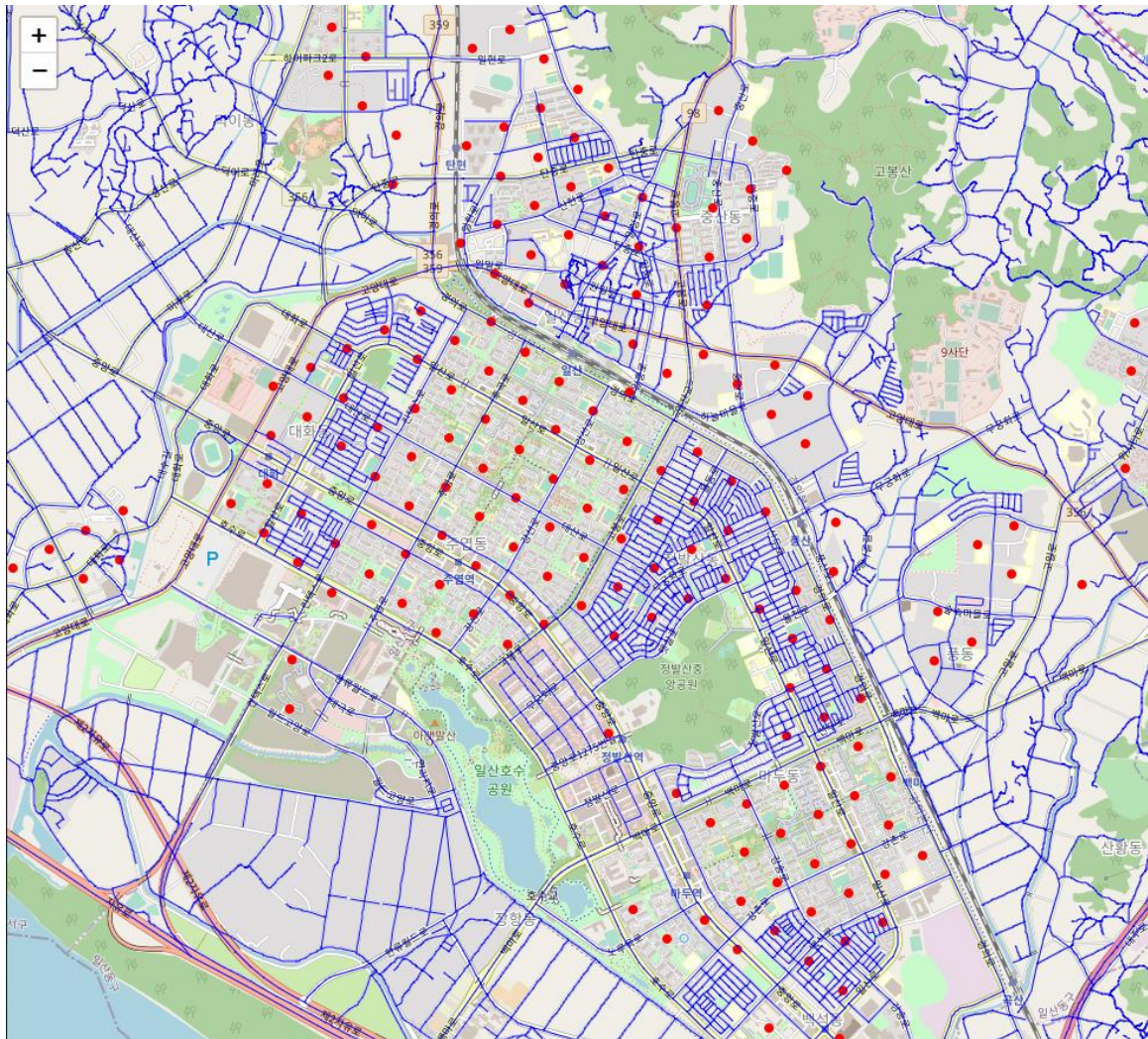
300개의 최종 후보 스테이션에 대한
시각화

스테이션 위치가 사용하기 편리한 장소라면, 거치대 수량은 실제 수요 반영 (단, 최소한의 운영을 위해 10대 이상, 대회에 언급된 바에 의해 30대 이하로 배치)

- 인구분포 변수(단위 면적당 해당 지역의 인구수)를 이용
- 300곳 중 최대 인구분포인 곳에 거치대 30대, 최소인 곳에 거치대 10대 배치
- MinMaxscaler를 이용하여 최종 거치대 수 결정

*MinMaxscaler란 다차원의 변수값들을 비교 분석하기 쉽게 하기위해 Scikit-Learn에서 제공하는 스케일러, 최대/최소값이 각각 1, 0이 되도록 스케일링하는 것이며, 모든 feature값이 0-1사이에 있도록 데이터를 재조정

스테이션 번호		거치대 수량
0	1	14
1	2	17
2	3	17
3	4	21
4	5	19
...
295	296	11
296	297	12
297	298	12
298	299	10
299	300	13



후보스테이션의 좌표를 지도위에 나타낸 결과

일부는 도로가 아닌 곳에 위치하고 있음

(파란선은 도로명주소_도로 파일에 존재하는 모든 도로, 빨간점은 최종후보스테이션 300개)

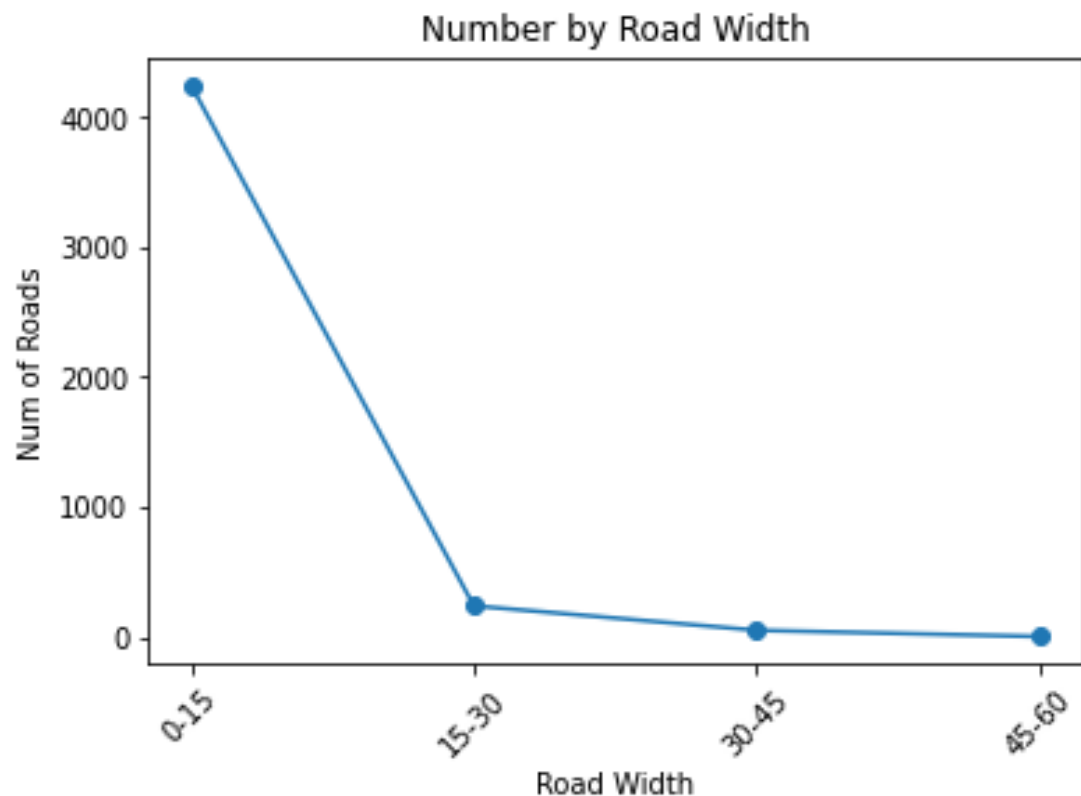
이처럼 스테이션이 도로가 아닌 곳에 위치하는 것을 방지하기 위해, 모든 후보 스테이션에 대하여 가장 가까운 도로로 재배치하는 최적화 과정 필요

스테이션이 존재할 수 없는 고속도로와 좁은 골목에 스테이션이 배치되지 않게 하기 위해 도로 별 도로 폭 고려

03

Data Analysis - 스테이션 최적화

도로 별 특징을 알아보기 위한 도로명주소_도로 데이터의 ROAD_BT(도로 폭) 범주화



왼쪽 그래프는 도로 폭 별 도로의 개수를 시각화한 것

- 가로축(도로 폭 범주): 0-15미만, 15-30미만, 30-45미만, 45-60미만
- 세로축: 해당 범주에 해당하는 도로의 개수

03

Data Analysis - 스테이션 최적화

가장 많은 개수를 차지하는 도로 폭이 15미만인 도로를 확대하여 확인한 결과, 대부분이 상가 건물과 건물 사이에 있는 길 (이차로, 자전거스테이션이 위치하기 어려운 좁은 길)





도로폭이 15-30미만인 도로의 경우 주로
사거리나 4차선 이상의 자전거 스테이션을
배치하기에 적합한 길

한편, 고속도로는 자전거 스테이션을 배치하기 부적합하므로 도로명에 '고속도로'가 포함된 도로는 제외

03

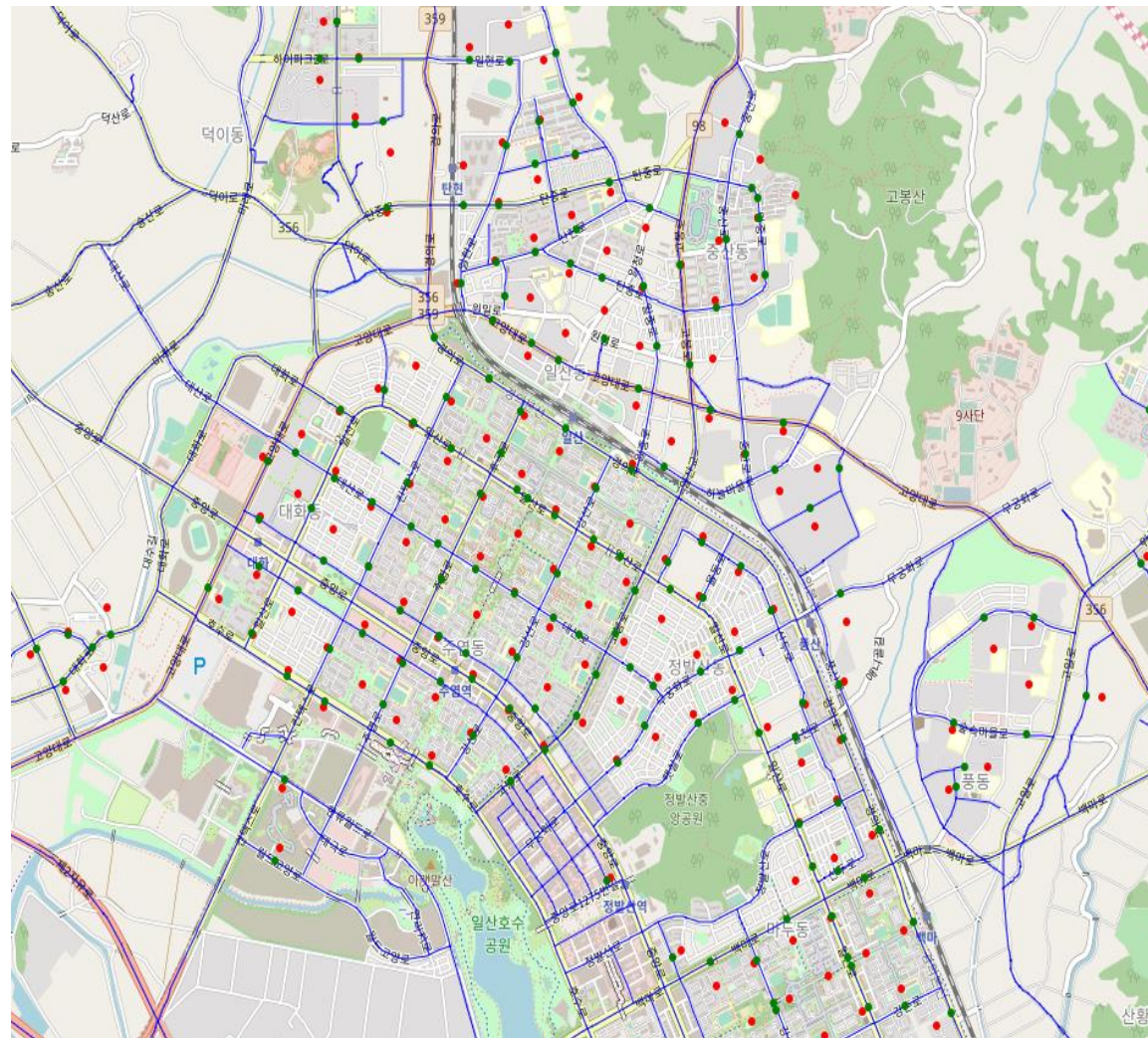
Data Analysis - 스테이션 최적화

최종적으로 도로폭이 15미만으로 좁은 도로와 고속도로를 제외한 도로에 대해서만 후보 스테이션을 재배치
(도로위의 모든 점에 대하여 후보스테이션과의 거리가 가장 짧은 점을 찍는 방식)

고속도로를 제외한 도로 중 도로폭이 15이상인 도로에 대하여 재배치한 결과, 모든 후보 스테이션이 스테이션이 존재할 수 있는 도로에 적절하게 재배치

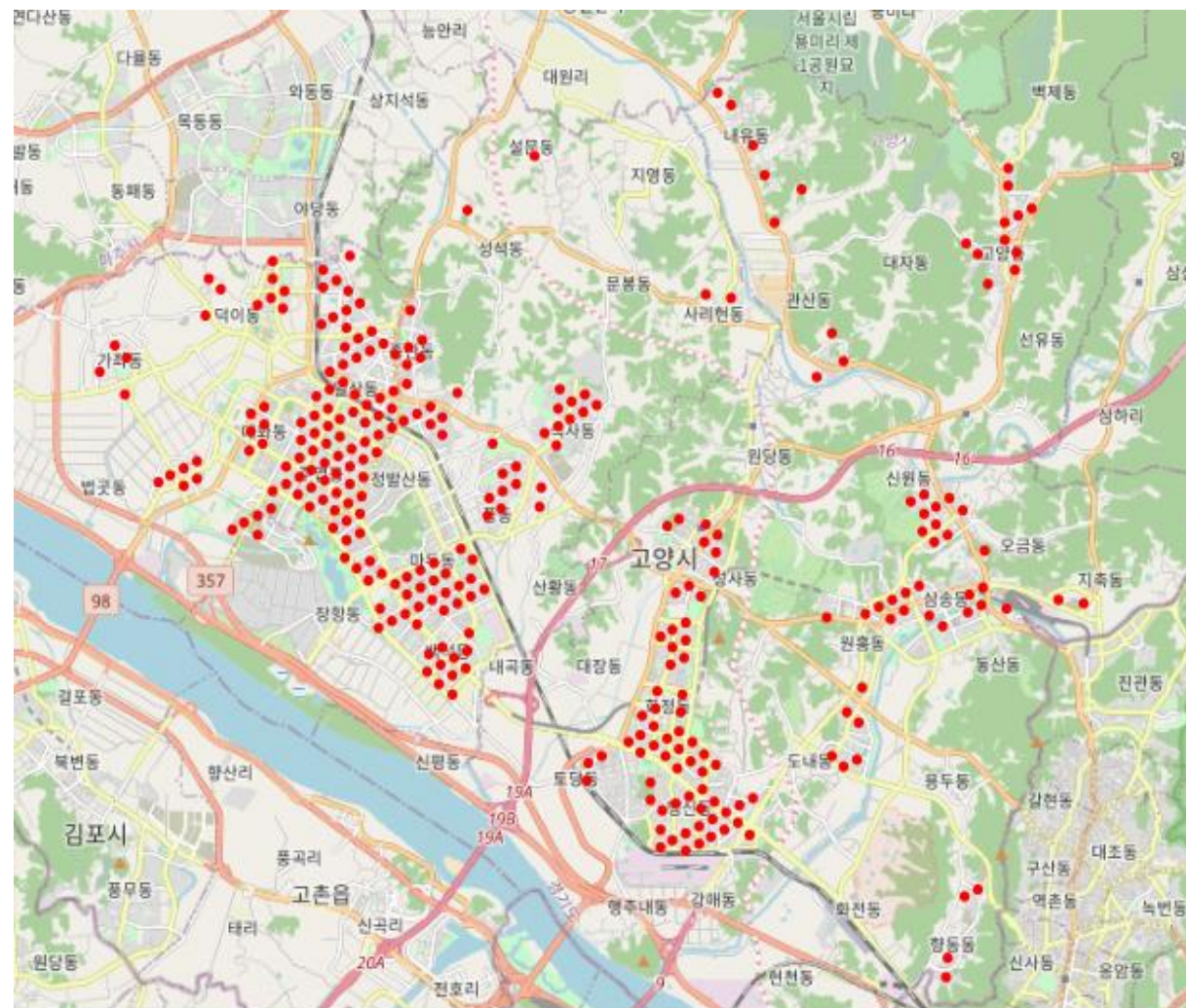
도로 배치 이전 → •

도로 배치 이후 → •



최적화된 최종 스테이션의 위경도 좌표 및 시각화

스테이션 번호	거치대 수량	X좌표(위도)	Y좌표(경도)
0	1	14	37.706757 126.893730
1	2	17	37.706161 126.896378
2	3	17	37.718452 126.902428
3	4	21	37.709007 126.901947
4	5	19	37.701008 126.898796
...
295	296	11	37.698582 126.738586
296	297	12	37.698582 126.738586
297	298	12	37.614212 126.834106
298	299	10	37.657944 126.896255
299	300	13	37.661606 126.890198



기대효과

2019년 이후 급격한 인구증가에 대응하는 300개의 자전거 스테이션 설치로 급증하는 수요 충족
최적화된 자전거 스테이션 재배치에 의한 고양시 공유 자전거 서비스 운영 효율성 극대화
수요의 증가와 그에 비례하는 수익의 상승 기대
친환경 교통수단 및 공유 서비스의 이용으로 사회에 긍정적 기여

제언

시민들의 이용을 이끌어내기 위해선 주기적인 정비와 관리 필요
자전거 도로 등과 같은 인프라 구축이 함께 이루어질 때 더욱 공유 자전거 서비스가 활성화 될 수 있음

아쉬운 점(제약사항)

호수공원 인근 자전거스테이션의 수요량이 높은 것을 확인하였으나, 해당 결과를 잘 반영할 수 있는 데이터의 부족
행사나 날씨 상황을 고려한 동적배치까지 추후 연구하면, 바람직할 것으로 예상됨

감사합니다

