

**공공자전거 대여소
최적입지선정 프로젝트**

Contents

01

개요

프로젝트 선정 배경
활용 언어 및 프로그램

02

팀 구성 및 역할

03

수행 과정

사전 기획

04

수행 결과

화면 설계

05

자체 평가

01

개요
OUTLINE

개요

주제 선정 이유

서울 공공자전거(이하 따릉이) 대여소 위치가
최적의 장소라 가정.

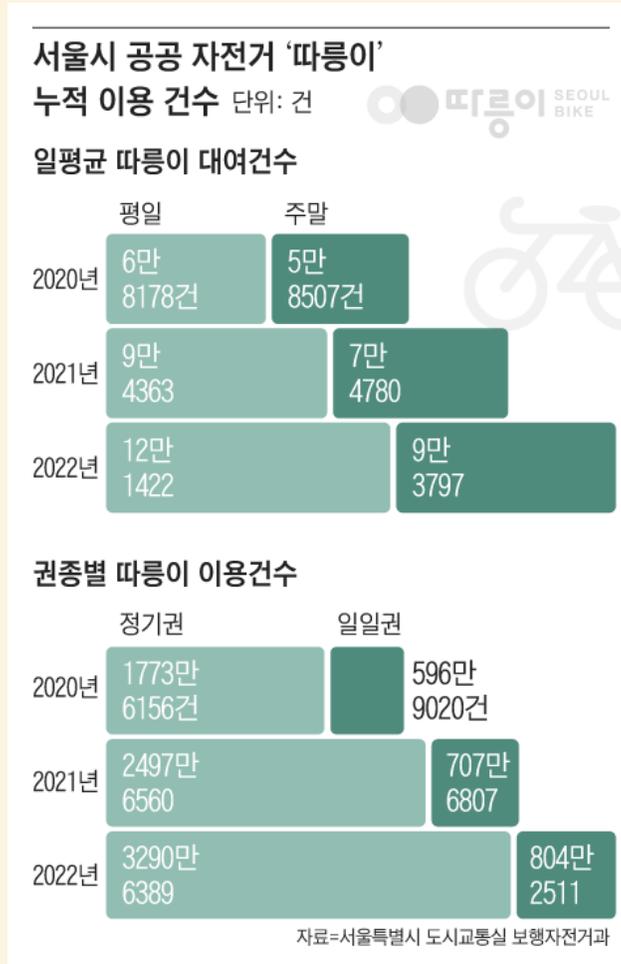
이를 기반으로 제주도에 공공자전거 대여소 입지 선정
장소를 추천한다.

대여소를 추가적으로 설치함으로써 관광객 유입에
도움이 될 수 있다.



개요

분석 배경



서울시 공공 자전거 '따릉이'
2022년 기준 이용건수 4094만 달성
전년(3205만 건) 대비 28% 증가

Chart.js를 이용한 시각화

① 공공자전거 월별 이용건수(18년~22년)

공공자전거 월별 이용건수(18~22년)

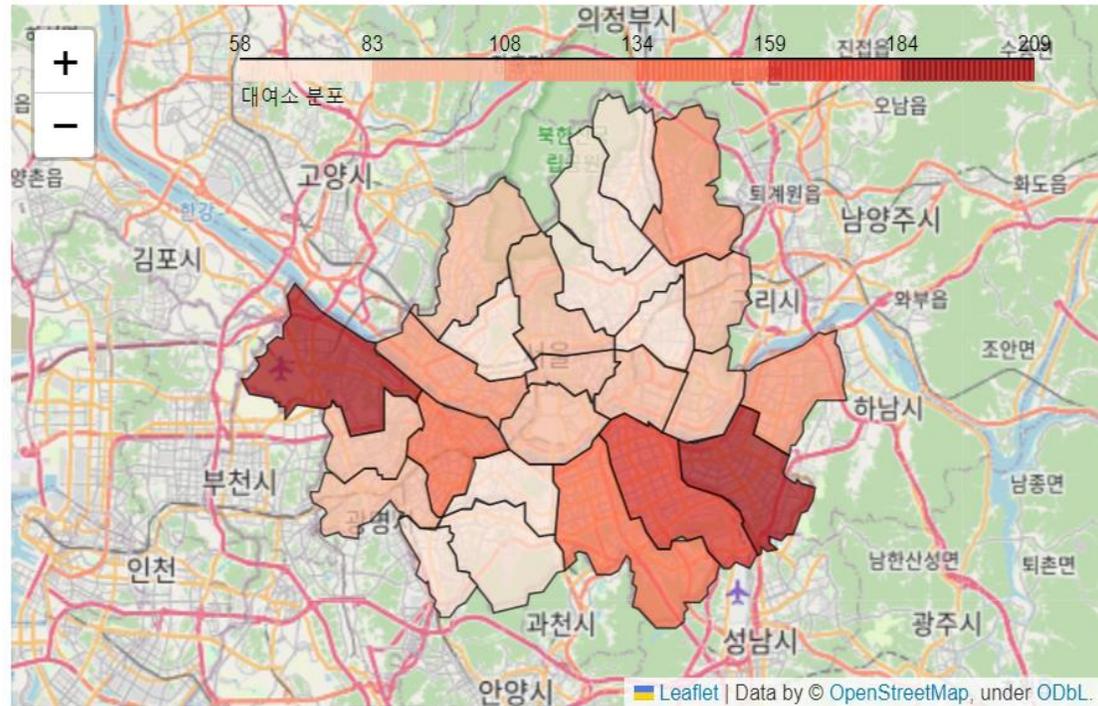


개요

데이터 시각화

② 유동인구 데이터 분석 및 수집 서울시 대여소 현황 및 유동인구 시각화

(1-1) 서울시 대여소 현황



(1-2) 서울시 유동인구 현황

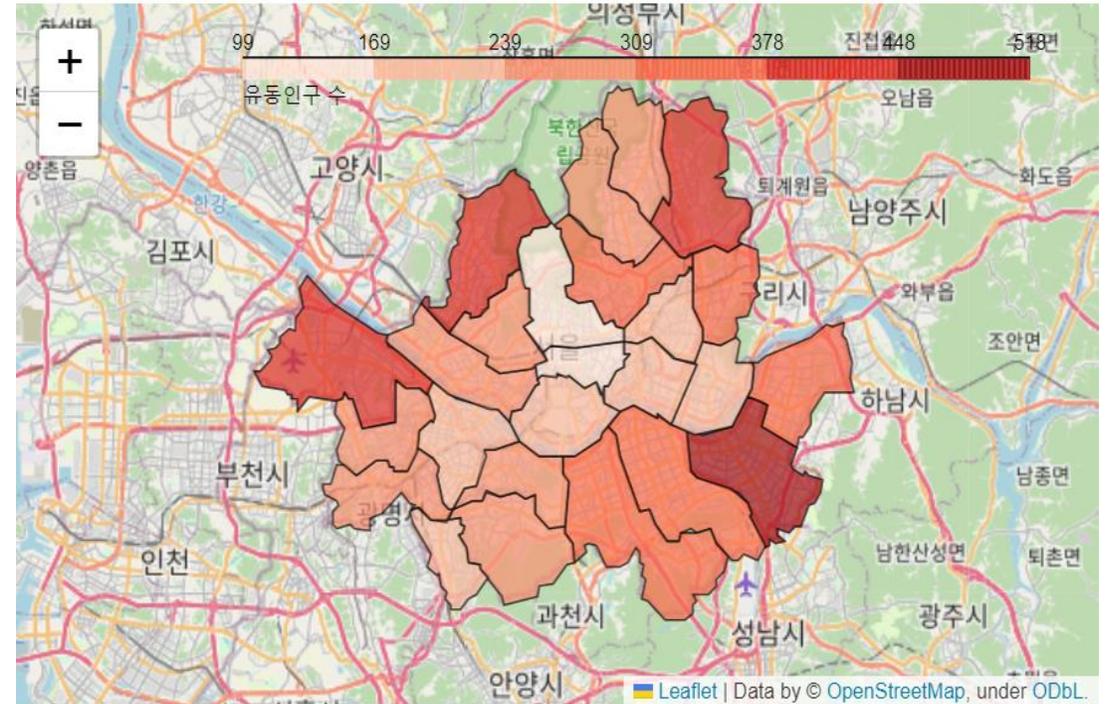
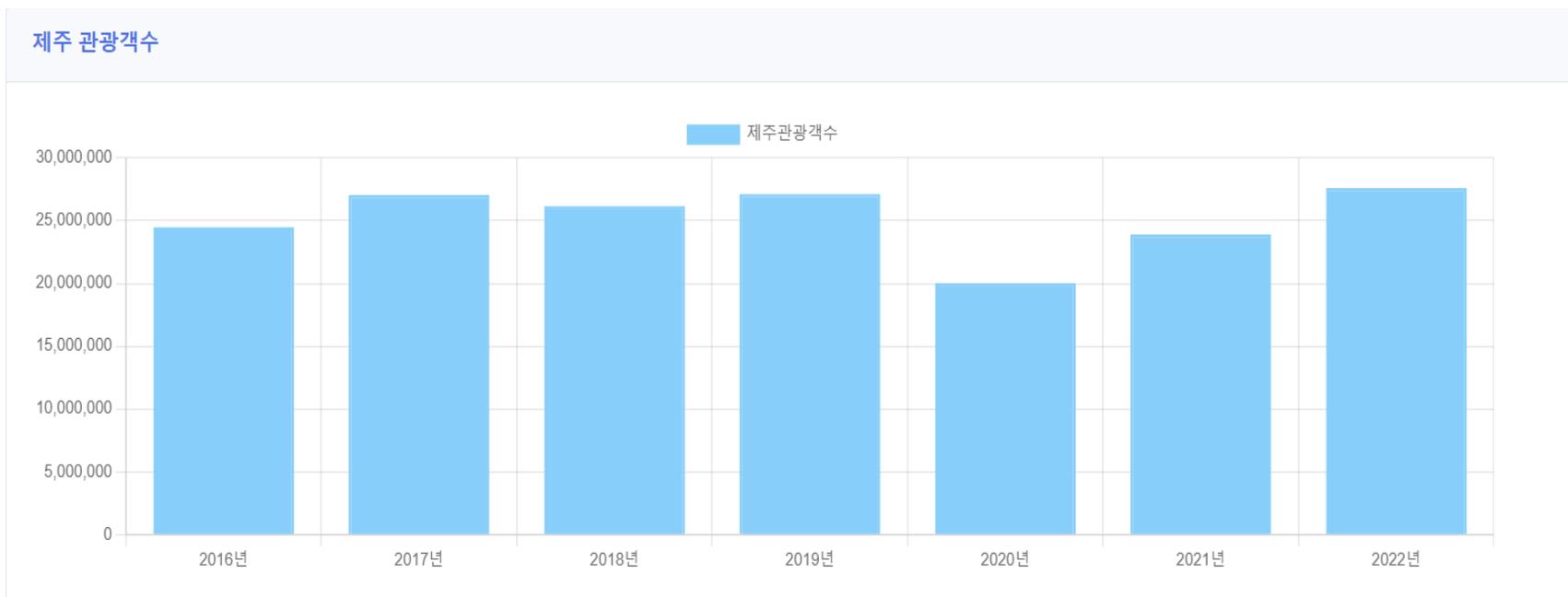


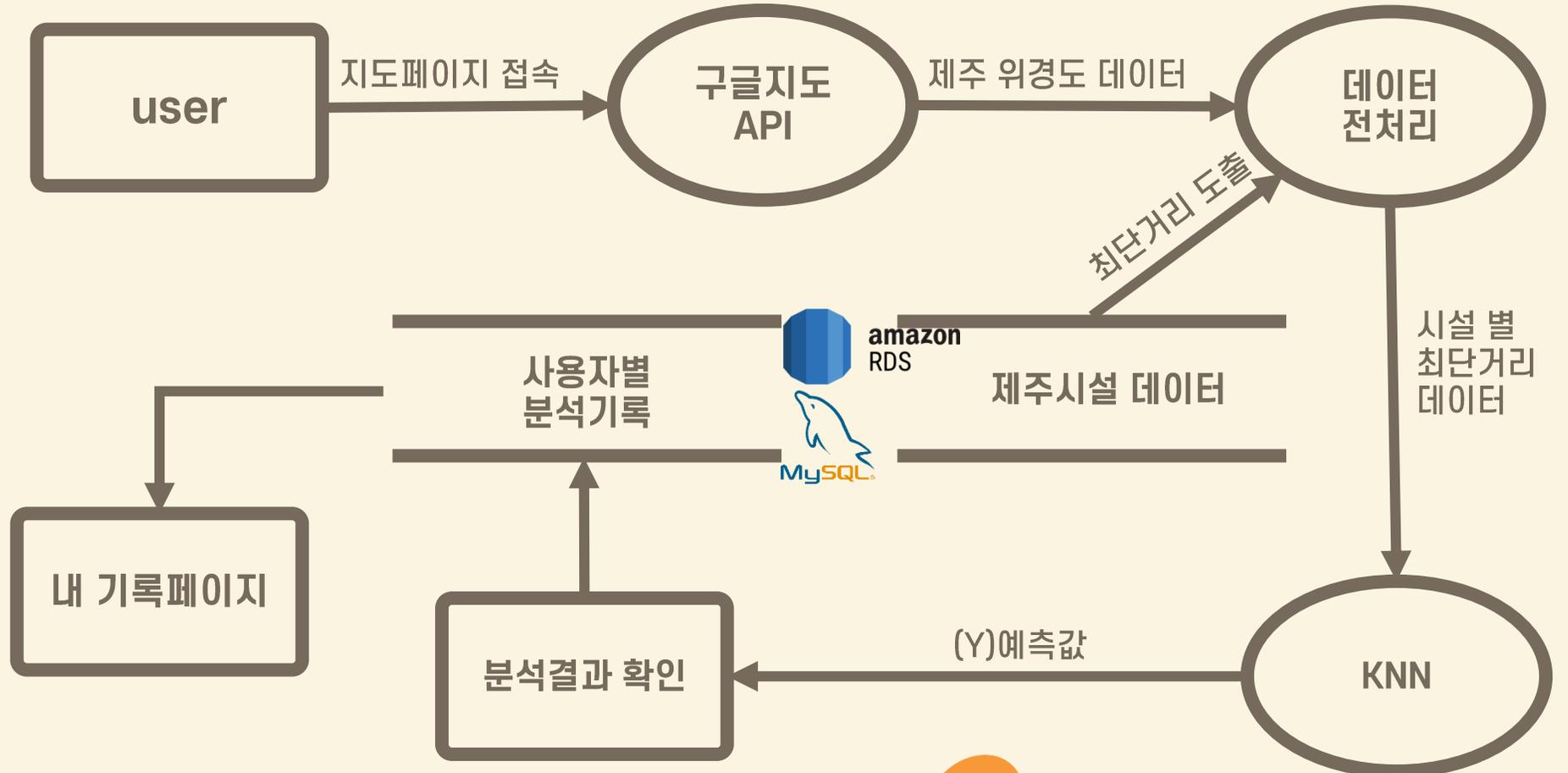
Chart.js를 이용한 시각화

③ 연도별 제주관광객수(2016년~2022년)

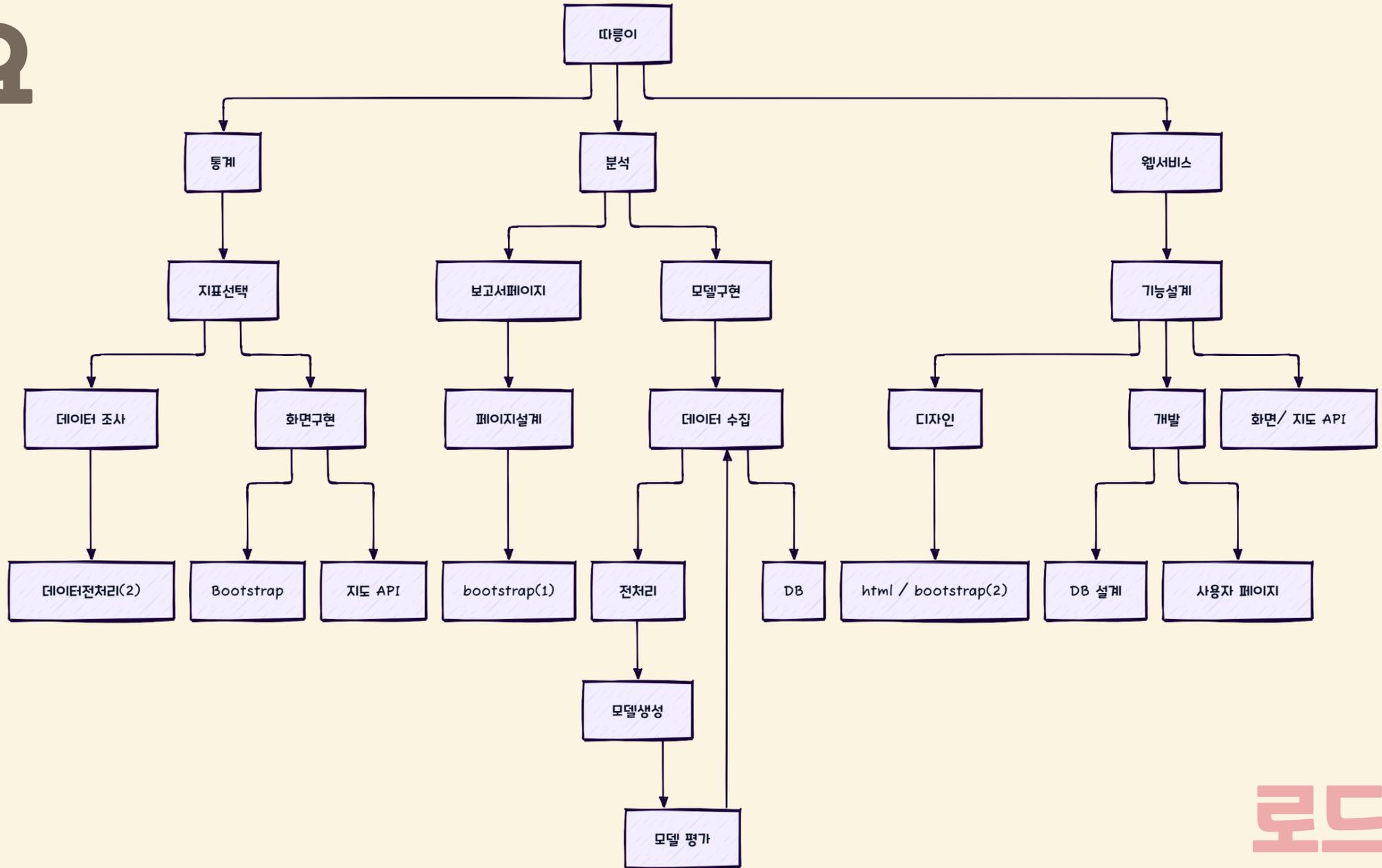


개요

활용 프로그램 및 언어



개요



02

팀원 구성 및 역할

MEMBER COMPOSITION

팀원 구성 & 역할

훈련생	역할	담당업무
박준*	팀장	데이터 전처리, 웹페이지 제작, 발표
이강*	부팀장	데이터 전처리 및 분석, 분석결과 페이지 제작, 발표
김연*	팀원	데이터 전처리, 웹페이지 제작
김유*	팀원	데이터 전처리 및 분석, ppt
박창*	팀원	데이터 전처리, 웹페이지 제작
윤지*	팀원	데이터 분석 및 모델 설계, 지도페이지 제작
현재*	팀원	데이터 전처리, 웹 페이지 제작

수행 과정

5월 25일 (목)

데이터 수집 및 가공

- 데이터 전처리 - 스케일링
- 데이터 시각화

6월 5일 (월)

서비스 구축

- 분석 페이지 제작

5월 22일 (월)

사전계획

- 프로젝트 기획 및 주제 선정
- 데이터 흐름도, 로드맵 작성
- 데이터 변수 설정

5월 30일 (화)

데이터 모델링

- 데이터 분석
 - 데이터 수집
 - 데이터 전처리
 - 데이터 시각화
- 모델 설계

6월 15일 (목)

발표준비

- PPT 제작

03 수행 과정

PROCESS

수행 과정

데이터 변수 설정

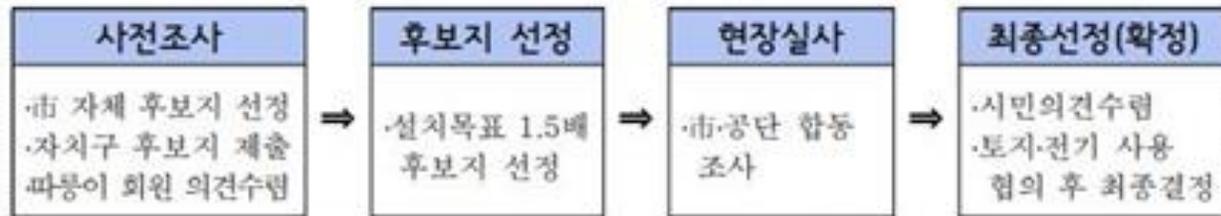
○ 대여소 설치기준 및 절차

<대여소 선정기준 및 절차>

▶ 대여소 설치 기준

- 기존 서울공공자전거의 이용형태에 따른 대여소 상호 연계성 고려
- 대중교통·다중이용시설(관공서 및 공공기관 등)·주요생활거점 위주
- 이용편의성 뿐만 아니라 보행·교통·인근 상권을 고려한 장소
- 횡단보도·버스정류장·건물 출입구 등 보행 안전 및 차로 등 주변환경 고려
- 대여소 설치로 인한 점포의 상행위 방해 예방 및 민원해소 고려

▶ 대여소 설치 절차



대학교, 관광지, 자전거도로, 정류장, 공원, 유동인구

수행 과정

데이터 수집

1. 학습을 위한 데이터 수집 - 서울 대여소, 시설

① 대여소와 거리계산을 위한 피쳐 데이터 수집

대여소 번호	대여소 명	위도	경도	대여 건수
102	망원역 1번	37.55565	126.9106	82308
103	망원역 2번	37.55495	126.9108	66924
104	합정역 1번	37.55063	126.915	37588
105	합정역 5번	37.55001	126.9148	26127
106	합정역 7번	37.54865	126.9128	78176
107	신한은행 가	37.55751	126.9185	44629
108	서교동 사	37.55275	126.9186	41233
109	제일빌딩 9	37.54769	126.92	32738
111	상수역 2번	37.54787	126.9235	31266
112	극동방송국	37.5492	126.9232	30013
113	홍대입구역	37.55744	126.9238	96075
114	홍대입구역	37.55706	126.9244	65305
115	마스타 빌딩	37.55893	126.9271	20758

[서울시 대여소별 대여건수]

	장소명	경도	위도	주제
0	서울시립대학교	127.06	37.58342	대학
1	서울여자간호대학교	126.9474	37.59748	대학
2	서울여자대학교	127.0905	37.62811	대학
3	서일대학교	127.1002	37.58671	대학
4	성공회대학교	126.8252	37.48774	대학
5	성균관대학교	126.9936	37.58823	대학
6	성신여자대학교	127.0221	37.5913	대학
7	세종대학교	127.0736	37.5518	대학
8	숙명여자대학교	126.9647	37.54648	대학
9	숭실대학교	126.9569	37.49635	대학
10	송의여자대학교	126.9874	37.5569	대학
11	연세대학교	126.94	37.56508	대학
12	이화여자대학교	126.9468	37.56186	대학
13	인덕대학교	127.0551	37.63109	대학
14	장로회신학대학교	127.1039	37.55012	대학

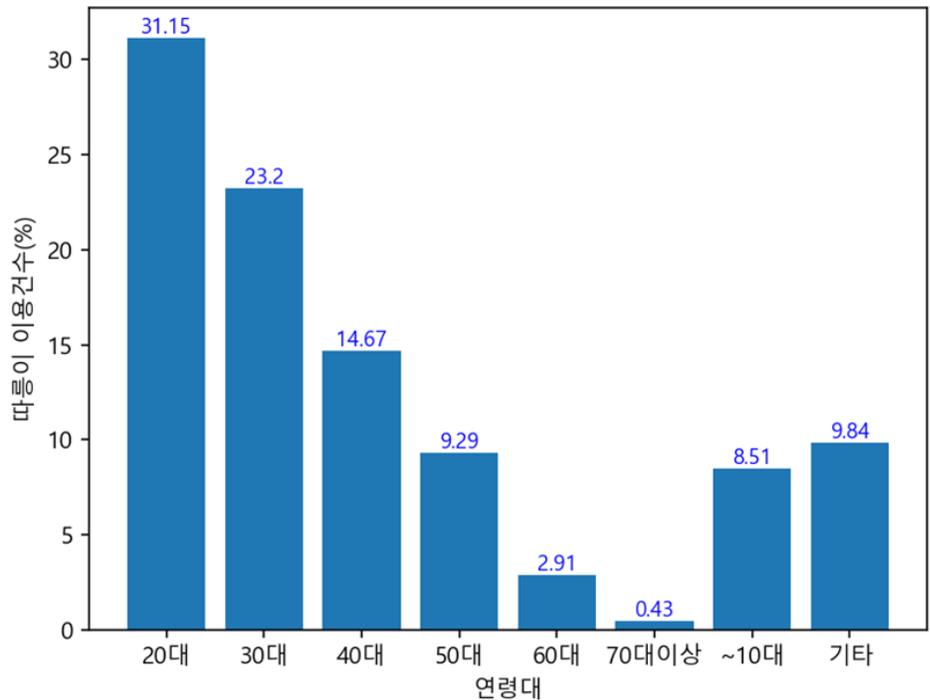
[서울시의 지하철역, 관광명소, 대학교, 자전거도로, 공원 위경도 데이터]

수행 과정

데이터 수집

1. 학습을 위한 데이터 수집 - 서울 유동인구

② 유동인구 데이터 분석 및 수집
- '따릉이' 이용 건수와 연령대 분석



행정구역별	나이대별	총이동
강남구	20	33089
강남구	30	33711
강남구	40	25961
강동구	20	23535
강동구	30	25640
강동구	40	17035
강북구	20	17150
강북구	30	12612
강북구	40	8978

20~49세 이용자가 전체 이용수의 69% 차지
→ 유동인구 범위 20대~40대로 제한

수행 과정

데이터 수집

1. 학습을 위한 데이터 전처리 - 서울시 각 시설별 최단거리

② 모델 학습용 데이터프레임 전처리 결과물

대여소번호	장소명	자전거도	지하철역	공원	관광지	대학	자치구	대여 건수	총이동
102	망원역 1번	0.011605	0.000632	0.008885	0.013891	0.016294	마포구	82308	70069
103	망원역 2번	0.012312	0.001292	0.008173	0.014109	0.015811	마포구	66924	70069
104	합정역 1번	0.018186	0.000739	0.005236	0.014488	0.010494	마포구	37588	70069
105	합정역 5번	0.018614	0.000268	0.004682	0.015106	0.010543	마포구	26127	70069
106	합정역 7번	0.018823	0.000816	0.002324	0.017315	0.01243	마포구	78176	70069
107	신한은행	0.016019	0.005199	0.011977	0.00679	0.011287	마포구	44629	70069
108	서교동 사	0.018967	0.004886	0.00934	0.011012	0.007898	마포구	41233	70069
109	제일빌딩	0.023565	0.002937	0.008614	0.015596	0.005328	마포구	32738	70069
111	상수역 2번	0.025885	0.000633	0.006689	0.015282	0.001824	마포구	31266	70069
112	극동방송국	0.024745	0.001526	0.006319	0.013931	0.00218	마포구	30013	70069
113	홍대입구역	0.020809	0.000712	0.007702	0.005857	0.009086	마포구	96075	70069
114	홍대입구역	0.021512	0.000839	0.00702	0.006393	0.008635	마포구	65305	70069
115	마스타 빌	0.023405	0.004102	0.007443	0.006337	0.010632	서대문구	20758	52561
116	일진아이몰	0.02241	0.008513	0.006082	0.004936	0.012897	서대문구	38224	52561
117	홍은사거리	0.018399	0.003599	0.011565	0.019679	0.006251	서대문구	74624	52561
118	광흥창역 2	0.032449	0.00034	0.004863	0.008964	0.006549	마포구	21116	70069
119	서강나루	0.033296	0.002363	0.006768	0.011291	0.006614	마포구	38465	70069

[독립변수]

Euclidean 거리계산 방식을 이용하여 서울시 약 3000개 대여소와 각 대여소로부터 5가지 시설별 최단거리를 도출.

자치구별 유동인구를 추가

[종속변수]

대여소별 대여건수

수행 과정

데이터 수집

2. 예측을 위한 데이터 수집 - 제주

- 정류장, 관광명소, 대학교, 자전거도로, 공원
5가지 시설데이터 취합

공공 데이터 / 카카오 API

장소명	위도	경도	주제
0 장작평사와친수공원	33.94058	126.3239	공원
1 한원리소공원	33.33795	126.1964	공원
2 근린공원	33.24984	126.3218	공원
3 제원어린이공원	33.48737	126.4864	공원
4 하모체육공원	33.21882	126.2524	공원
5 녹남못꽃동산	33.21932	126.2697	공원
6 제주오름공원	33.48588	126.5124	공원
7 환상숲	33.32256	126.2632	공원
8 기적의공원	33.48024	126.4825	공원
9 설문대할망테마공원	33.3627	126.8354	공원
10 제주승마공원	33.40551	126.4066	공원
11 44호어린이공원	33.50404	126.4648	공원
12 감산마을체육공원	33.25811	126.3542	공원
13 월림리소공원	33.34795	126.2563	공원
14 올레길쉼터	33.3218	126.778	공원
15 가시리체육공원	33.35326	126.7748	공원
16 청수소공원	33.31137	126.2471	공원

수행 과정

데이터 전처리

거리 스케일링 여부

제주, 서울변수의
분포, 밀도가 상이하여
분석에 용이하도록
거리 이용

유동인구 스케일링

제주 - 서울 간
유동인구 단위의 차이로
표준화 진행

결측치 처리 이상치 제거

결측치가 포함된
행 삭제.

이상치 제거 후
정확도의 유의미한 변화
확인

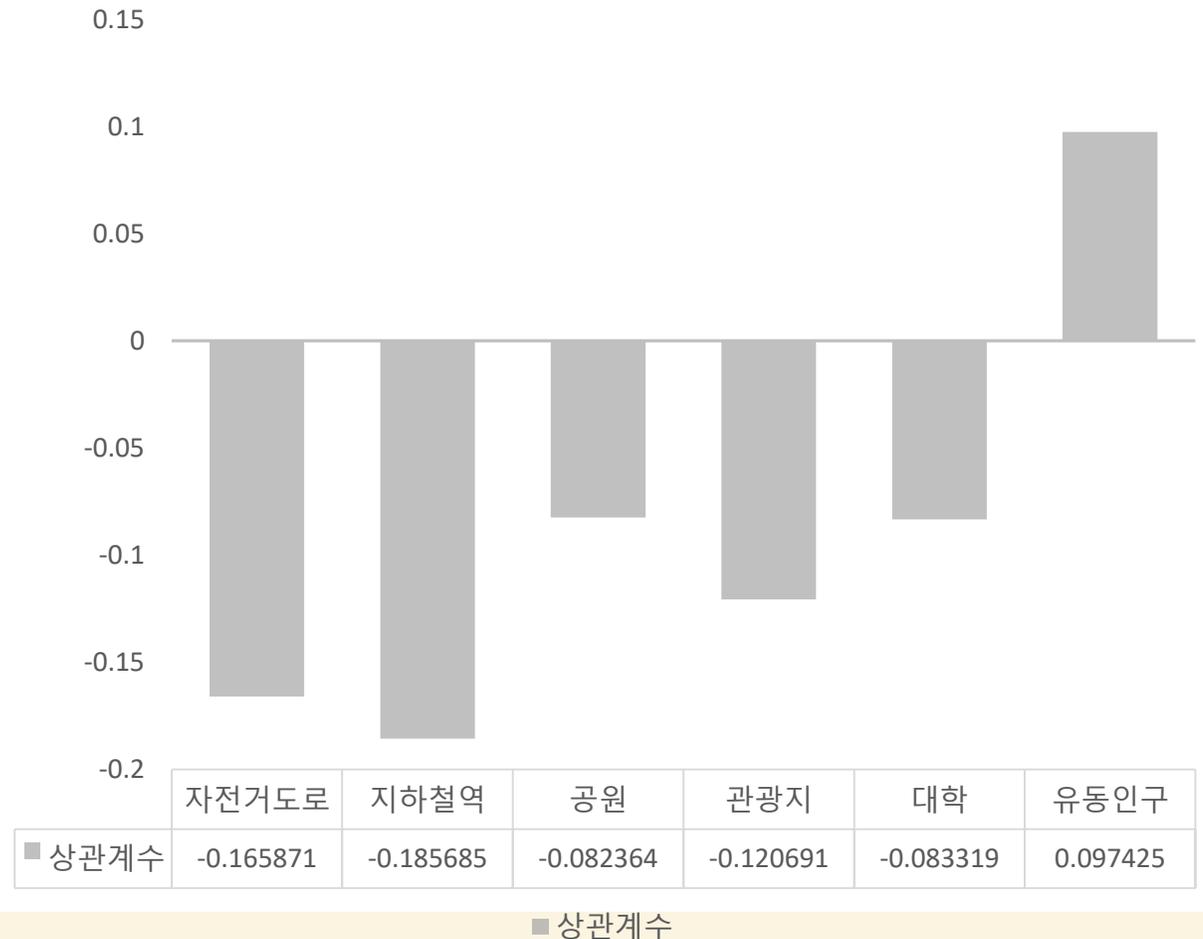
수행 과정

알고리즘 탐색

1. 상관관계 확인

지하철역, 자전거도로, 관광지, 유동인구, 대학, 공원순으로 대여건수에 약하게 영향을 미쳤다.

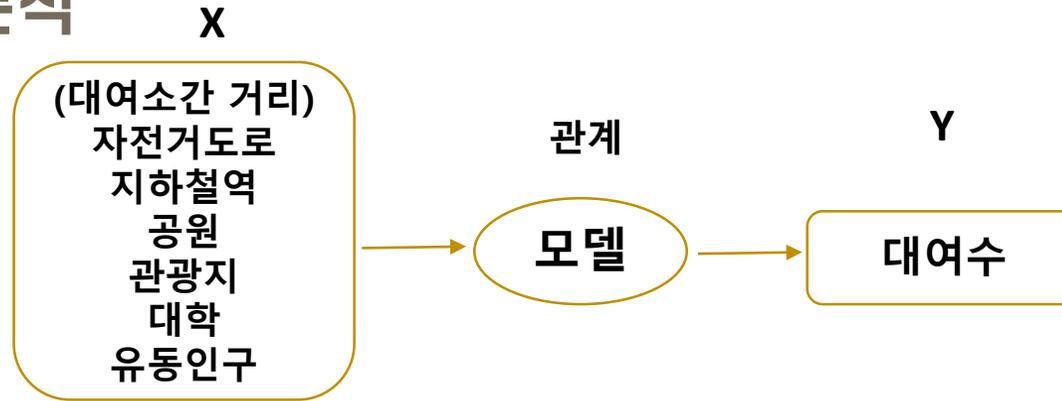
상관관계



수행 과정

알고리즘 탐색

2. 다중선형회귀모델 분석



결과

```
OLS Regression Results
=====
Dep. Variable:          y          R-squared:                0.076
Model:                 OLS        Adj. R-squared:           0.074
Method:                Least Squares  F-statistic:              36.29
Date:                  Tue, 20 Jun 2023  Prob (F-statistic):       1.91e-42
Time:                  16:07:16      Log-Likelihood:           -30106.
No. Observations:     2646          AIC:                     6.023e+04
Df Residuals:         2639          BIC:                     6.027e+04
Df Model:              6
Covariance Type:      nonrobust
```

독립변수들과 종속변수간에
인과관계가 있으나 설명력이 7%로
약하다.

→ 모델이 유의하기 때문에 입지추천에
대여수를 점수화하는 로직이 아닌
범주화하는 방법을 검토해보았다.

수행 과정

알고리즘 탐색

- 평가 범주 설정

추천: 상위10% / 보통: 10%~80% 구간 / 부적합: 하위20%

- 훈련데이터 설정

학습데이터 약 2700개 → 훈련데이터 개수 80%로 설정

모델	정확도(Train)	정확도(Test)	비고
LogisticRegression	70%	68 %	'적합' 예측 못 함
K-NN(n=3)	80%	65%	예측비율이 가장 적합
Random Forest	100%	67.1%	'적합' 예측률이 낮음
Xgboost(depth=5)	100%	60%	'적합' 예측률이 낮음
MLPClassifier(tensorflow)	71.4%	68.6%	정확도는 준수하나 '적합'을 예측하지 못함

* 유동인구 제외

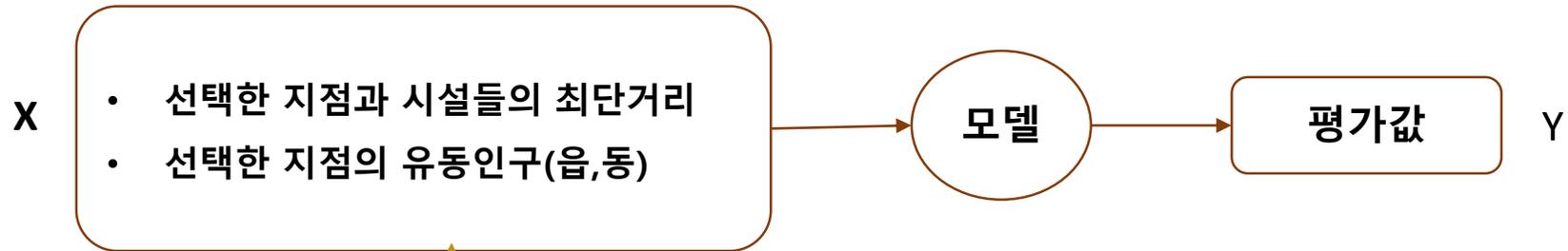
모델들을 제주도 적용해본 결과 K-NN모델이 가장 적합하게 예측해 K-NN을 선택했다.

수행 과정

모델 적용

예측 과정

서울시기반 모델로 제주도의 임의의 지점 대여소 입지 평가



제주 시설 위치 데이터

제주 유동인구 데이터

구좌읍	33.43883071923182	126.72435149519629
오동동	33.4290359729518	126.53582145040565
중문동	33.35257600789905	126.46334504480042

마커 없애기

제출

수행 결과

모델 적용

문제점1

[현상]
임의 지점 선택 시 전반적으로
서울시의 피쳐보다 높은 값 도출

[원인]
제주도는 서울보다 문화시설의
밀집도가 낮음

자전거도로를 지도에서 보면 해안가,시
근방에 몰려있음

[조치 : 거리 조정]

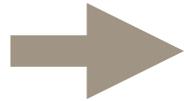
	자전거도로	지하철역	공원	관광지	대학
평균	0.0277978	0.007254	0.010248	0.016132	0.019612
최대	0.08199885	0.05469	0.034336	0.075114	0.080194
최소	0.00023716	8.25E-05	0.000183	0.000267	0.000118

```
# * @desc 거리 조정  
result_data['자전거도로']=result_data['자전거도로']*1.15  
result_data['지하철역']=result_data['지하철역']*0.7  
result_data['공원']=result_data['공원']*0.55  
result_data['관광지']=result_data['관광지']*0.5  
result_data['대학']=result_data['대학']*0.45
```

문제점2

[현상]
독립변수에 유동인구 추가시
좁은 구간 내 심해지는 편차

[원인]
제주 관광지 유동인구 비율이 많고,
지역간 유동인구 편차가 커
영향도가 커짐



유동인구
피쳐 삭제

- 서울시 유동인구와 제주도 유동인구 특성이 다름을 인지
- 표준화를 한 데이터를 축소해 사용하는 것이
유의미하지 않을 수도 있어 삭제.

04 수행 결과

RESULT

페이지 시연

[http://bestjejubike](http://bestjejubike.shop)
.shop

05 자체 평가

FEEDBACK

자체 평가

- ◆ 모델을 활용하기까지 데이터수집, 모델 성능확인, 실사용 여부 등 예측하지 못한 변수가 많았지만 창의력과 기지를 발휘해 결과를 성공적으로 이끌어낼 수 있었다.
- ◆ 프로젝트 개발 동안 부트스트랩의 적극적인 사용으로 UX/UI를 최적화했다.
또한 git과 slack 을 통한 협업이 이루어져 수월하게 진행되었으며 AWS 서버로 서비스를 했다.
- ◆ 데이터 분석을 할 때 충분한 조사와 결정을 위한 시각화가 충분하지 못했던 점이 아쉬웠으며, 이후 보완 과정에선 최적입지 추천 불가능한 지역에 대한 제한과 그 지역에 대한 특성을 반영하고자 한다.

