

「2020 공공 빅데이터 분석 공모전」

발품 팔기 전에 바로방에서 찾아보자!

바로방

2030의 서울에서 최적의 집 찾기 프로젝트

바로방팀

- 권서영, 김기원, 박수정, 유연승, 최수진





목차

01 주제 및 개요

주제 선정 배경

기대효과

02 제작 과정

웹사이트 형태 기획

공공데이터 수집, 크롤링

EDA, 데이터 전처리

부동산 추천 시스템 제작

웹사이트 제작

03 웹사이트 결과 시연

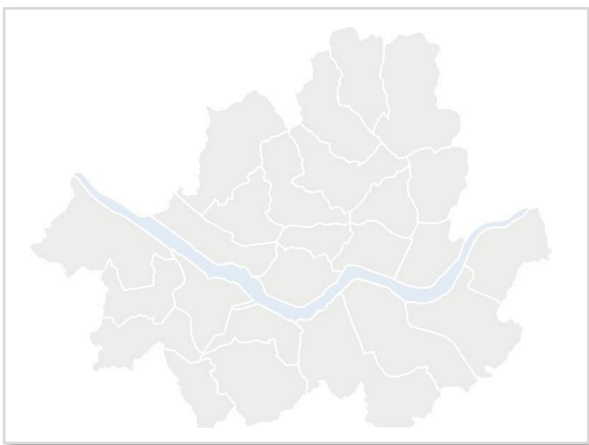
04 한계 및 향후 보완점

01 주제 및 개요

주제 선정 배경

“취직 또는 취업 준비로 서울에 집을 구해야 하는 2030세대”

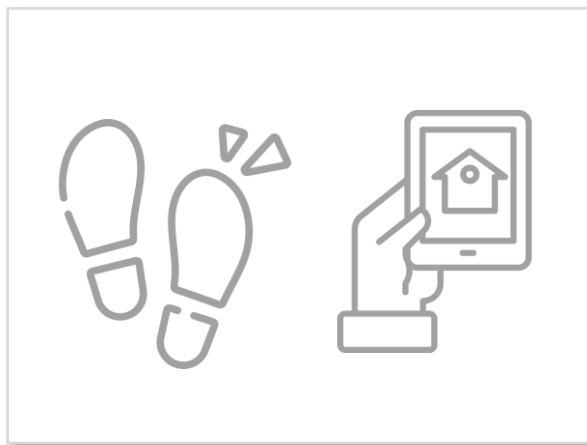
어디서부터 찾지?



넓은 서울에서 어느 지역부터
찾아봐야 할지 막막했던 경험

직장 근처가 좋을까?
아니면 조금 거리를 두고 구하는 게 좋을 까?

언제까지 찾아야하지?



직접 발품을 팔거나, 부동산 관련
앱으로 시간을 들여 찾아봐야 함

많은 매물을 일일이 비교할 수 없는 한계점

지역별 보증금/월세는?



실거래가 정보가 공시 되어 있지만,
일반 시민이 집을 구할 때 참고용
으로 활용하기는 어려움

01 주제 및 개요

주제 선정 배경



실거래가
기반
집 추천

2030 타겟
원룸/투룸
특화

부동산
매물 정보
연동

서울의 집을 추천해주는 웹사이트

부동산 실거래가 기반 집 추천과
추천 집의 현재 부동산 매물 정보까지

01 주제 및 개요

기대효과



공공 기관

- 공공 행정 서비스의 일환으로 집 추천 웹사이트 제공으로 국민 편의성 증진에 기여
- 공공 기관 및 행정 서비스 이미지 및 만족도 향상에 도움

일반 2030 시민

- 추천된 집 위주로 집을 구하면 됨으로 발품, 부동산 앱을 찾아보는 데에 소모 되는 불필요한 시간 감소
- 조건에 맞는 칩 추천으로, 집 찾기 편의성 증대



02 제작 과정

02 제작과정

웹사이트 형태 기획

홈페이지 가안

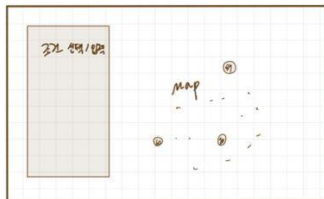
1. 홈 화면

- 회원가입(이름, 성별, 사는곳, 나이)
- 로그인

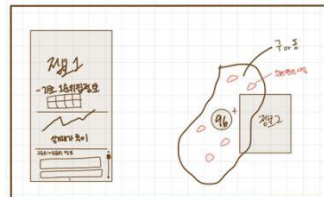
환영합니다!

여러분들에게

구성 초안



↓ 평양과 1891 (보여주기)



1. 사용자 입력란

DB 적재용 입력정보

나이 성별, 전, 구할 때 중요 요소

오피스텔 □ 연립주택 □ 원룸/투룸

매물 검색

3. 가용자금 선택(가격 기준은 가격 통계 낸 이후 설정)
(전세 경우) 가용가능 자금(전세대출금 포함) 약 _____만원

(월세 경우)

1) 가용가능 보증금 약 _____만원

2) 가능한 월세(가격 기준은 통계를 낸 뒤 결정)
(원룸 체크 경우) □ 30만원이하 □ 30~40만원 □ 40~50만원 □ 50~60만원 □ 60~70만원 □ 70만원 이상

(연립주택 체크 경우) □ 20~30만원 □ 30~40만원 □ 40~50만원 □ 50~60만원 □ 60~70만원 □ 70만원 이상

(오피스텔 체크 경우) □ 20~30만원 □ 30~40만원 □ 40~50만원 □ 50~60만원 □ 60~70만원 □ 70만원 이상

4. 건축연수

2. 매물 추천

1. 크리스탈 오피스텔
5억, 5평, 1995년식
직장소요시간 30분

2. 크라임진 오피스텔
3억, 4평, 1994년식
직장소요시간 1시간 20분

3. 밀키공주 오피스텔
4억5천, 10평, 1990년식
직장소요시간 50분

4. Wish Kim 오피스텔
7억, 8평, 1993년식
직장소요시간 3분

검색 결과에 만족하시나요????? ☆☆☆☆

얼마나 평점 받아도 좋을듯.

물스장 □ 세탁소 □ ...음...

웹사이트 구성 초안

최종 목표했던 웹사이트 형태

1. 홈 화면

- 주택 형태 선택 (오피스텔, 빌라, 단독다가구)
- 전/월세 선택
- 보증금/월세 입력
- 직장 주소 입력
- 선호 편의시설 선택 (최대 3개)

2. 디테일 화면

- 지도에 추천 집과 편의시설 정보 띄우기
- 실제 부동산 매물
- 기존 거래되었던 가격 추이 통계정보

02 제작과정

공공데이터 수집

1. 부동산 실거래가 정보

수집처: 국토교통부 실거래가 공개시스템 (<http://rtdown.molit.go.kr/>)

종류: 서울에서 자취방을 구하는 2030세대 타겟에 맞추어 서울 전/월세 연립, 단독 다가구, 오피스텔 데이터 수집

기간: 2015.01.01~2020.09.30

2. 편의시설

수집처: 공공 데이터 포털 (<https://www.data.go.kr/>) - 데이터명: 소상공인 시장진흥공단_상가(상권)정보

종류: 세탁업, 다이소, 실내운동시설

수집처: 서울 열린 데이터 광장 (<https://data.seoul.go.kr/>) - 데이터명: 서울특별시 체력단련장업 인허가 정보

종류: 실내운동시설 (공공 데이터 포털에서 수집한 실내운동시설 데이터가 다소 미비하여 보완 용도로 수집)

3. 지하철역

서울 열린 데이터 광장 (<https://data.seoul.go.kr/>) - 데이터명: 서울특별시 노선 별 지하철역 정보(신규)

02 제작과정

데이터 수집 - API, 네이버 부동산 크롤링

4. 위도, 경도 - 카카오API

부동산 거래, 편의시설, 지하철역 데이터에 위도, 경도 컬럼 추가

(부동산, 편의시설, 지하철역 데이터가 각각 별도의 데이터이므로, 위도, 경도 기준으로 매칭하여 활용하기 위한 목적)

5. 네이버 부동산 크롤링

크롤링 코드 이용하여 네이버 부동산에 올라온 부동산 실 매물 데이터 구 별로 나누어 크롤링

크롤링 된 데이터에 위도, 경도 컬럼 추가 - 카카오API

부동산 실거래가 데이터에 실제 매물 데이터를 매칭/연동하여, 최신 데이터 제공 용도

02 제작과정

데이터 전처리

1. 부동산 실거래가 정보

오피스텔(208947): 도로명 외 주소 컬럼 삭제, 그 외 불필요한 컬럼 삭제

단독 다가구(762504): 도로명 결측치 73841개 행 삭제(관악구 주로 분포, 지번 주소는 단독 다가구의 모든 데이터가 마스킹되어있어 활용 불가)

연립(503838) : 도로명 외 주소 컬럼 삭제, 그 외 불필요한 컬럼 삭제

→ 전체 데이터에서 중복된 데이터 삭제

→ 각 매물 별 최신 거래 데이터 구분 컬럼 추가(최신 거래일 경우 1, 아닌 경우 0)

- 주소, 건물명, 전월세, 면적(범주형), 층(범주형) 컬럼이 모두 같은 매물 중 최신 거래 구별

2. 네이버 부동산 매물

주소 결측치, 같은 매물이지만 각기 다른 공인중개사에서 업로드한 중복 매물 데이터 삭제

면적 단위가 공급면적, 계약면적 등으로 나뉘어져 있어, 전용면적 수치를 추출하여 통일

부동산 실거래가 정보 데이터와 도로명주소 기준으로 매칭하여 최종 웹페이지에서 불러올 수 있도록 함

02 제작과정

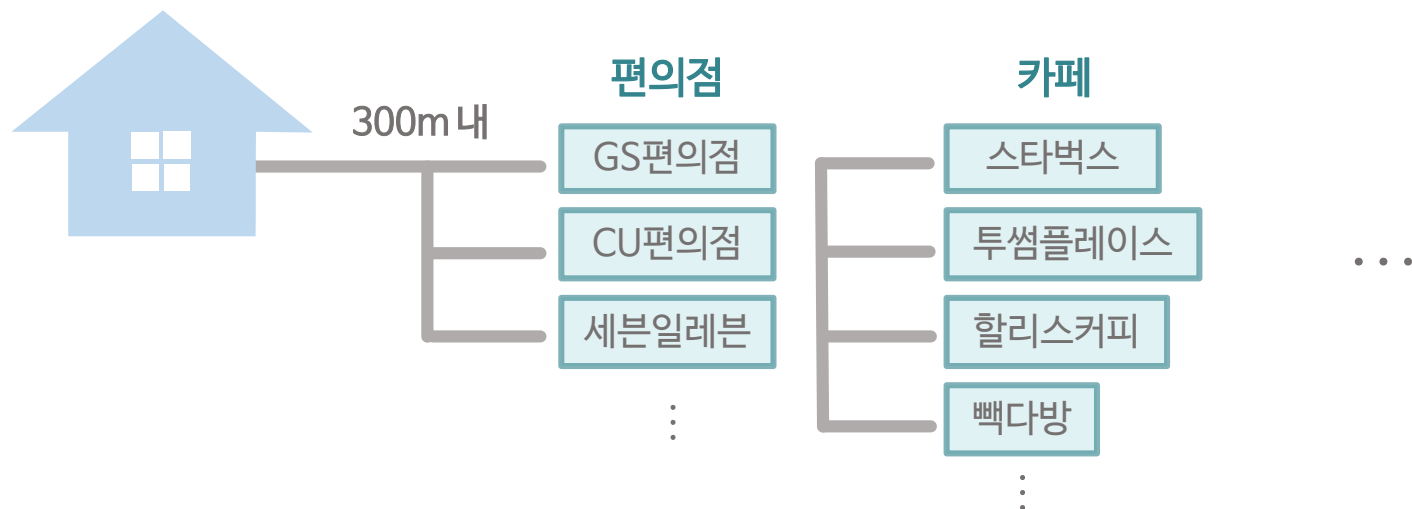
데이터 전처리 - 반경 300m 내 편의시설

3-1. 편의시설 - 1차 전처리

부동산 거래 데이터에 위도, 경도 기준으로 주변 반경 300m 내 편의시설 정보(점포명, 위/경도, 주소) 수집 및
매물 당 편의시설 각각의 개수 컬럼 추가

카카오API 이용: 대형마트, 편의점, 문화시설, 카페, 병원, 음식점

파이썬 거리계산 함수 이용(공공데이터): 세탁업, 다이소, 실내운동시설



02 제작과정

데이터 전처리

3-2. 편의시설 - 2차 전처리

사용자가 선택한 주소 근처 편의시설 개수가 매물의 추천 점수에 반영되므로
사용자에게 좀 더 유용한 편의시설로 개수를 제한하여 점수에 반영할 수 있도록,
마트 중 대형마트, 편의점 중 5개 프랜차이즈, 카페 중 10개 프랜차이즈 외 데이터 삭제
(해당 기준은 브랜드 평판 지수, 설문조사 브랜드 순위 등을 참고해 결정함)

하나의 부동산 매물에 편의시설이 과도하게 많이 붙는 경우가 있어 이상치로 추정하고
해당되는 편의시설을 지도에서 일일이 검색하여, 존재하지 않는 편의시설 행 삭제

```
main_data2['다이소_cnt'].value_counts().sort_index()
```

executed in 9ms, finished 10:40:35 2020-10-13

0	10257
1	1786
2	256
3	33
4	7
5	3

```
data_cnt['세탁소_cnt'].value_counts().sort_index()
```

executed in 11ms, finished 11:42:35 2020-10-13

0	772
1	738
2	1230
3	1561
⋮	
16	15
17	22
18	5
19	3
21	1
23	1

Q 1월커피전문점브랜드 순위는?



그래픽 dizzo

자료 한국기업평판연구소(2020년 1월 기준)

02 제작과정

데이터 전처리 - 반경 300m/500m 내 지하철역

4. 지하철역

부동산 거래 데이터에 위도, 경도 기준으로 지하철역 관련 새로운 컬럼을 추가하여 추천 매물 점수에 반영될 수 있도록 함

초/역세권 컬럼 추가

: 주변 300m 내 지하철역 존재 시 초역세권(2), 주변 500m 내 지하철역 존재 시 역세권(1), 지하철역이 없는 경우(0)

지하철역명 컬럼 추가

: 주변 500m내 지하철역의 이름이 담긴 컬럼 추가

환승역 여부 컬럼 추가

: 주변 500m내 지하철역이 환승역인 경우(1), 환승역이 아닌 경우(0)

02 제작과정

추천 매물 시스템

1. 구별 평균 속도 계산(대중교통 기준)

사용자가 입력한 직장 주소와 모든 매물 간에 대중교통 소요 시간 계산

→ 실시간으로 대중교통 소요 시간 검색 사용 시, 웹사이트 상에서 추천 서비스 제공 속도 느림

→ 이점 보완을 위해 구별 평균 속도를 미리 산출해,

직장과 매물 간 대중교통 소요 시간을 계산하여 매물 추천 점수에 활용

구별 평균 속도 산출 과정

각 구마다 랜덤으로 매물 샘플 위치를 50개씩 선정(총 1,250개)

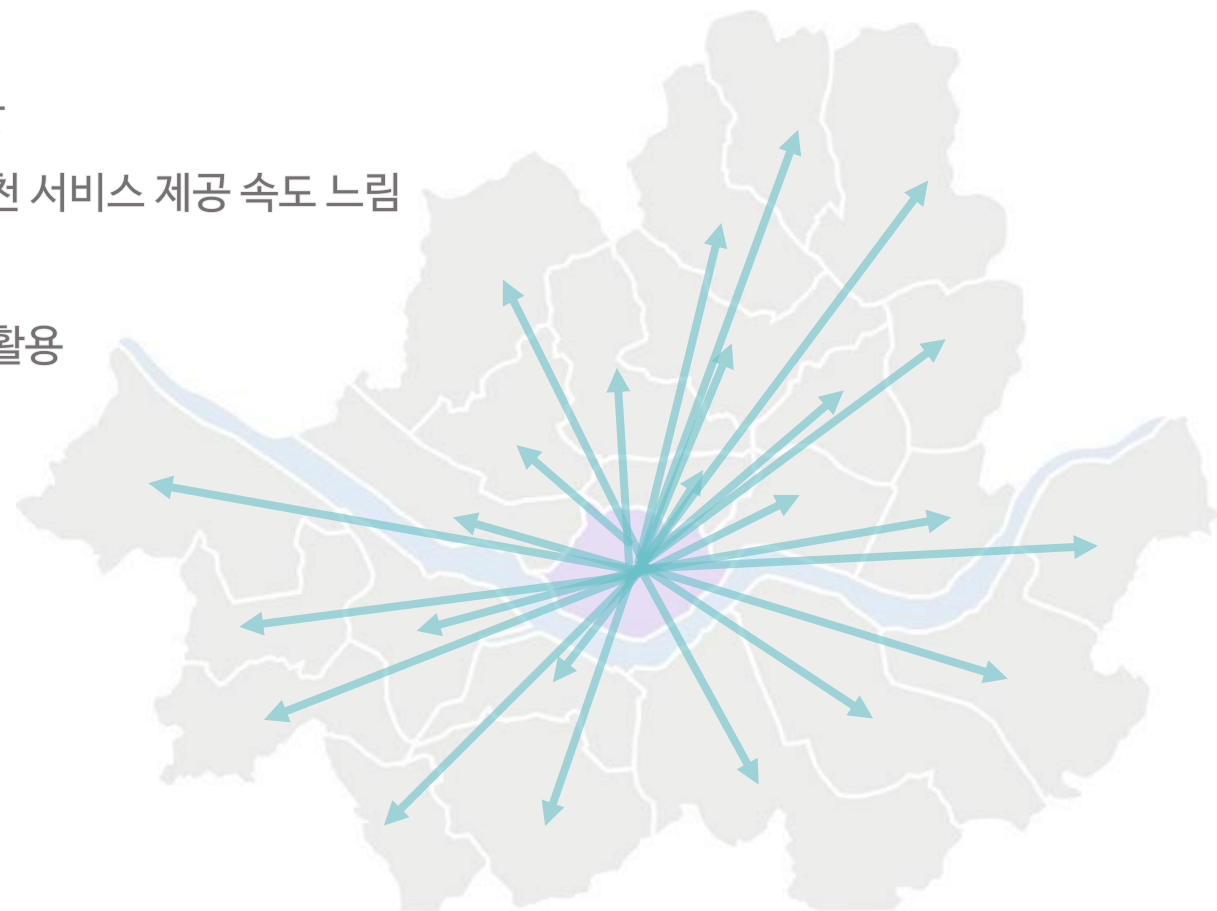
→ 구별 50개 샘플 중 25개는 평일 출근시간(오전 8시),

나머지 25개는 평일 퇴근시간(오후 6시) 계산에 사용

한 구의 중심부에서 1,250개의 샘플 위치로 가는 **거리**와

대중교통 이동 **시간** 추출 (구글 대중교통 길찾기 API 이용)

→추출된 거리와 시간으로 $\text{속력} = \frac{\text{거리}}{\text{시간}}$ 공식을 통해 구별 **평균 속도** 산출



02 제작과정

추천 매물 시스템

2. 가중치와 점수 설정(점수가 높은 매물을 최종 검색 결과에 노출)

편의시설 개수 별 점수(10점 만점)

: 사용자는 편의시설을 0개~3개까지 선택할 수 있고, 선택 개수에 따라 점수 계산 방법이 달라짐

편의시설 3개 선택 = 편의시설 당 3점+1점(기본값) / 2개 선택 = 5점씩 / 1개 선택 = 10점 / 0개 선택 = 0점

→ 편의시설 점수가 10점 이상이면 10점, 아닐 경우 개수에 비례해서 점수 계산

대중교통 소요 시간 점수(10점 만점)

: 직장 주소(사용자가 입력)와 매물 간 거리 계산 후, 구별 평균 속도를 이용해 매물 별 소요 시간 도출 (시간 = 거리/속력)

→ 10분 이하면 10점, 1시간 반 이상이면 0점, 사이 시간은 소요시간에 반비례해서 점수 계산

지하철(5점 만점)

: 초역세권 / 역세권 / 역세권이 아닌 경우에 따라 다르게 점수 계산 후 → 2.5점 만점으로 환산

: 환승역인 경우 50점, 아니면 10점으로 점수 계산 후 → 2.5점 만점으로 환산

02 제작과정

추천 매물 시스템 - 최종 테이블 ERD

주택 유형별 실거래가 정보 (3개)

오피스텔			
PK/FK	Logical Name	Name	Type
	address	도로명주소	VARCHAR
	rent	전/월세 구분	VARCHAR
	size_cut	면적 범주	VARCHAR
	size	면적	FLOAT
	deposit	보증금	INT
	pay	월세	INT
	year	건축년도	FLOAT
	contract	계약일	VARCHAR
	road	도로명	VARCHAR
	name	건물명	VARCHAR
	floor_cut	층 범주	VARCHAR
	floor	층	INT
	area	구	VARCHAR
	recent	최신 여부	INT
	latitude	위도	FLOAT
	longitude	경도	FLOAT
PK	criteria	위도, 경도 병합 문자	VARCHAR
	station_area	초/역세권 여부	INT
	station_name	500m 내 지하철명	VARCHAR
	transfer	환승역 여부	INT
	hos_num	300m 내 병원 개수	INT
	mart_num	300m 내 마트 개수	INT
	fast_num	300m 내 패스트푸드점 개수	INT
	cafe_num	300m 내 카페 개수	INT
	cul_num	300m 내 문화시설 개수	INT
	con_num	300m 내 편의점 개수	INT
	laun_num	300m 내 세탁소 개수	INT
	da_num	300m 내 다이소 개수	INT
	gym_num	300m 내 실내운동시설 개수	INT

편의시설 상세 정보 (9개)

카페			
PK/FK	Logical Name	Name	Type
	house_type	주택 유형	VARCHAR
FK	criteria	위도, 경도 병합 문자	VARCHAR
	con_name	카페 이름	VARCHAR
	con_address	카페 도로명주소	VARCHAR
	con_latitude	카페 위도	VARCHAR
	con_longitude	카페 경도	VARCHAR

편의점			
PK/FK	Logical Name	Name	Type
	house_type	주택 유형	VARCHAR
FK	criteria	위도, 경도 병합 문자	VARCHAR
	con_name	편의점 이름	VARCHAR
	con_address	편의점 도로명주소	VARCHAR
	con_latitude	편의점 위도	VARCHAR
	con_longitude	편의점 경도	VARCHAR

⋮

...

02 제작과정

추천 매물 시스템 - 최종 테이블 ERD

네이버 부동산 매물

네이버 부동산 매물			
PK/FK	Logical Name	Name	Type
	n_address	도로명주소	VARCHAR
	n_rent	전/월세 구분	VARCHAR
	n_size	전용 면적	FLOAT
	n_size_pro	공급 면적	FLOAT
	n_deposit	보증금	INT
	n_pay	월세	INT
PK	n_road	도로명	VARCHAR
PK	n_road_some	도로명 일부(길/로까지만)	VARCHAR
	n_name	건물명	VARCHAR
	n_floor	해당층/총층	VARCHAR
	n_date	입주가능일	VARCHAR
	n_latitude	위도	FLOAT
	n_longitude	경도	FLOAT
	n_loan	융자금	VARCHAR
	n_side	건물 방향	VARCHAR
	n_cost_incloud	관리비 포함 종류	VARCHAR
	n_cost	관리비	VARCHAR
	n_double_floor	복층 여부	VARCHAR
	n_room_bath	방수/욕실수	VARCHAR
	n_heat	난방	VARCHAR
	n_parking	총 주차대수	VARCHAR
	n_households	총 세대수	VARCHAR
	n_area_type	지역 용도	VARCHAR
	n_security	보안시설 종류	VARCHAR
	n_other_con	기타 시설	VARCHAR
	n_parking_incloud	주차 포함 여부	VARCHAR
	n_window_veranda	방범창/베란다	VARCHAR
	house_type	주택 유형	VARCHAR

구별 평균 속도

구별 평균 속도			
PK/FK	Logical Name	Name	Type
PK	area	구	VARCHAR
	speed	평균 속도	FLOAT

부동산 거래 데이터 테이블에
네이버 부동산 매물을 도로명으로 매칭

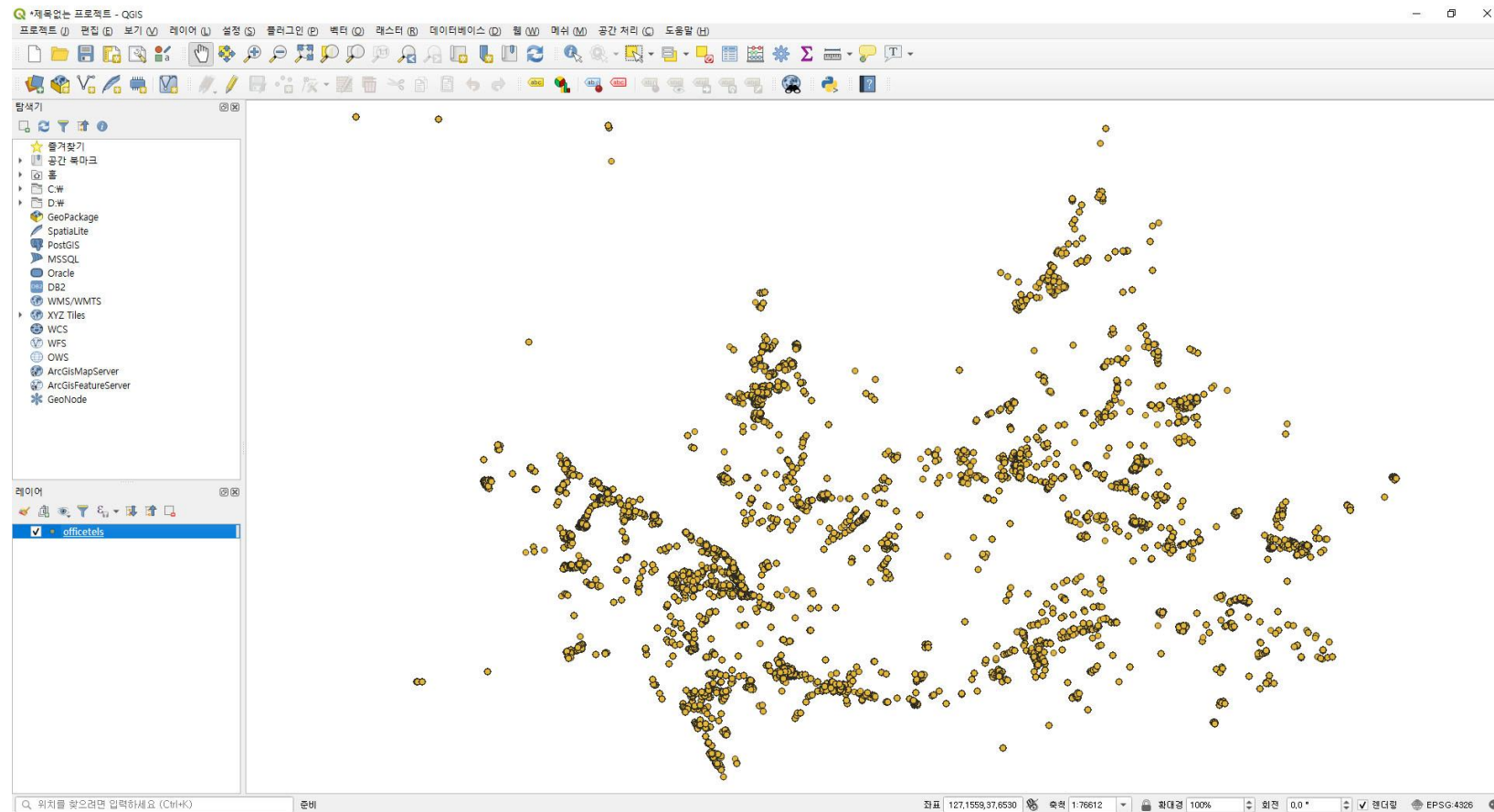
구별 평균 속도는 추천 매물 시스템 코드가 실행되면
해당 테이블에서 값을 불러옴

02 제작과정

웹사이트에 DB 연동

QGIS

공간 정보 오픈 소스 프로그램.
모든 데이터를 지도 상에 표시하고,
위치 정보 함수를 사용하기 위해
공간 정보 데이터로 변경



02 제작과정

웹사이트에 DB 연동

PostgreSQL

공간 데이터베이스 구축

데이터를 불러올 때

속도 단축을 위해

SQL언어 사용

The screenshot shows the pgAdmin 4 web interface. On the left, the 'Browser' pane displays a tree view of the database structure, including 'Tables (24)' and 'Trigger Functions'. The 'Query Editor' pane shows a SQL query: `select * from officetels`. The 'Data Output' pane displays the results of the query in a table format.

gid	fid	address	rent	size	deposit	pay	year	contract	road
[PK] integer	numeric	character varying (254)	character varying (254)	numeric	bigint	bigint	numeric	character varying (254)	character varying (254)
1	1	서울특별시 강서구 방화동 개...	전세	000000000001	13000	0	000000000000	2017-10-19	개화동로27가길 4
2	2	서울특별시 강서구 방화동 개...	전세	000000000001	13000	0	000000000000	2017-08-22	개화동로27가길 4
3	3	서울특별시 강서구 방화동 개...	전세	000000000001	13000	0	000000000000	2017-03-06	개화동로27가길 4
4	4	서울특별시 강서구 방화동 개...	전세	000000000001	12500	0	000000000000	2016-10-03	개화동로27가길 4
5	5	서울특별시 강서구 방화동 개...	전세	000000000001	13000	0	000000000000	2016-09-08	개화동로27가길 4
6	6	서울특별시 강서구 방화동 개...	전세	000000000001	13000	0	000000000000	2016-08-29	개화동로27가길 4
7	7	서울특별시 강서구 방화동 개...	전세	000000000001	13000	0	000000000000	2016-07-13	개화동로27가길 4
8	8	서울특별시 강서구 방화동 개...	월세	000000000001	500	60	000000000000	2020-04-28	개화동로27가길 4
9	9	서울특별시 강서구 방화동 개...	월세	000000000001	13000	10	000000000000	2020-01-16	개화동로27가길 4
10	10	서울특별시 강서구 방화동 개...	월세	000000000001	7000	20	000000000000	2020-01-09	개화동로27가길 4
11	11	서울특별시 강서구 방화동 개...	월세	000000000001	5000	35	000000000000	2019-12-27	개화동로27가길 4
12	12	서울특별시 강서구 방화동 개...	월세	000000000001	1000	57	000000000000	2019-12-05	개화동로27가길 4

Successfully run. Total query runti



03 웹사이트 결과 시연

03 웹사이트 결과 시연

웹사이트 - Django, Javascript 활용



‘바로방’ 홈페이지

홈 화면

상단바

: 홈 화면으로 돌아갈 수 있는 HOME 메뉴 및 데이터를 관리하는 ADMIN 메뉴 존재

검색 조건 설정

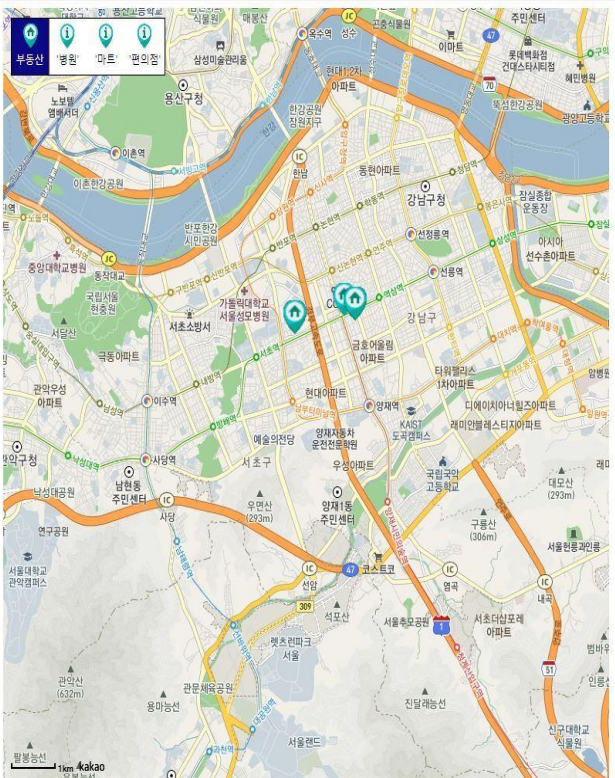
: 사용자로부터 부동산 유형, 전월세 여부, 보증금, 월세, 직장주소, 편의시설을 입력 받아 검색

→ 월세 입력 칸은 전월세 여부에서 월세를 선택할 경우에만 생성

→ 조건을 모두 입력하지 않을 경우 메시지 출력

03 웹사이트 결과 시연

웹사이트 - Django, Javascript 활용



검색 결과

서울특별시 서초구 서초동 강남대로 381

서울특별시 강남구 역삼동 테헤란로4길 34

서울특별시 강남구 역삼동 테헤란로4길 34

서울특별시 강남구 역삼동 테헤란로4길 34

서울특별시 강남구 역삼동 테헤란로4길 34

서울특별시 서초구 서초동 서초대로 316

서울특별시 서초구 서초동 서초대로 316

서울특별시 서초구 서초동 서초대로 316

서울특별시 서초구 서초동 서초대로 316

디테일 화면

지도


: 입력한 직장주소를 중심 좌표로 주변 매물 표시

매물 목록

: 지도에 표시된 매물들의 주소 목록으로,
사용자가 선택한 조건에 따라 가중치를 계산하여
점수가 높은 상위 10개 매물 표시

03 웹사이트 결과 시연

웹사이트 - Django, Javascript 활용



검색 결과

서울특별시 서초구 서초동 강남대로 381

건물세	보증금	세	거래년도	면적	지하철 정보
월세	1000	100	2009-10-17	42.00	([강남역 3호선] 강남역 신분당선)

서울특별시 강남구 역삼동 테헤란로4길 34

서울특별시 강남구 역삼동 테헤란로4길 34

서울특별시 강남구 역삼동 테헤란로4길 34

서울특별시 강남구 역삼동 테헤란로4길 34

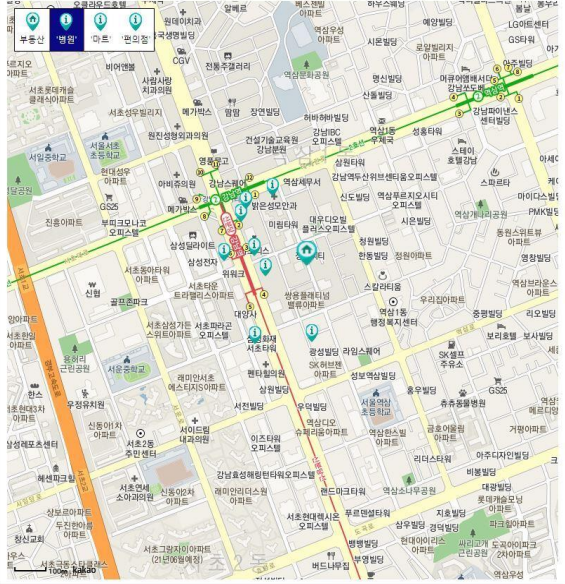
디테일 화면

매물 목록 설명

: 매물 주소를 클릭하면 상세 정보가 표시되고,
해당 매물을 중심으로 지도 이동

편의시설 카테고리

: 사용자가 선택한 편의시설이 카테고리 상으로 표시
→ 카테고리를 클릭하여 편의시설별 위치 확인 가능



검색 결과

서울특별시 서초구 서초동 강남대로 381

서울특별시 강남구 역삼동 테헤란로4길 34

건물세	보증금	세	거래년도	면적	지하철 정보
월세	3500	13	2018-06-21	30.96	([강남역 3호선] 강남역 신분당선)

서울특별시 강남구 역삼동 테헤란로4길 34

서울특별시 강남구 역삼동 테헤란로4길 34

서울특별시 강남구 역삼동 테헤란로4길 34

서울특별시 강남구 역삼동 테헤란로4길 34


서울특별시 서초구 서초동 서초대로 316

서울특별시 서초구 서초동 서초대로 316

서울특별시 서초구 서초동 서초대로 316

03 웹사이트 결과 시연

웹사이트 - Django, Javascript 활용



서울특별시 강남구 역삼동 테헤란로4길 34

서울특별시 강남구 역삼동 테헤란로4길 34

서울특별시 서초구 서초동 서초대로 316

서울특별시 서초구 서초동 서초대로 316

서울특별시 서초구 서초동 서초대로 316

서울특별시 서초구 서초동 서초대로 316

전월세	보증금	세	거래년도	면적	지하철 정보
월세	1000	115	2019-02-11	29.61	[교대역 3호선] 교대역 2호선

네이버 매물

주소	전월세	계약면적	전용면적	보증금	세	건물이름	층	계약일
서울시 서초구 서초대로 316	월세	24.12	None	1000	90	하림인터내셔널 1동	6/15층	1개월 이내 협의가능
서울시 서초구 서초대로 316	월세	29.61	None	1000	120	하림인터내셔널 1동	중/15층	즉시 입주 협의가능
서울시 서초구 서초대로 316	월세	24.12	None	1000	90	하림인터내셔널 1동	13/15층	1개월 이내 협의가능

디테일 화면

네이버 매물 목록

: 네이버 매물이 존재할 경우 매물 목록 아래에 전월세, 면적, 보증금 등의 상세 정보를 포함한 네이버 매물 목록 생성

➔ 매물이 존재하지 않을 경우에는 목록 대신 존재하지 않는다는 메시지 출력



04 한계 및 향후 보완점

04 한계 및 향후 보완점

한계 및 보완점



실시간 정보 반영 불가

→ 실시간 부동산 매물 정보,
편의시설 정보 등을
가져 올 수 있도록 개선 필요



부동산 동향 관련 추가 정보 제공 미흡

→ 실거래가 추이,
전/월세 거래 동향 등을
그래프 형태로 제공 필요



가중치 점수의 타당성 부족

→ 사용자 평점 및 후기를 통해
가중치 지속 보완 필요



사용자 로그 추적 및 서비스 평점 기능 미비

→ 로그인을 통한 로그 데이터 수집,
추천 관련 평점 및 후기 기능 구현을 통
해 향후 데이터 분석을 통한 개선 필요

사용자 UI 개선 필요

감사합니다.

