

에이콘

2팀

PROJECT

아매방

아파트 매매 예측 플랫폼



CONTENTS

I. 개요

1. 팀원 소개
2. 주제 선정 배경
3. 수행 절차

II. 데이터 수집

1. 데이터 전처리

III. 데이터 분석

1. 모델링 생성
2. 결과 및 해석

IV. 결론

1. 웹 시연
2. 코드 분석
3. 피드백 및 느낀점

아매방

아파트 매매 예측 플랫폼



이수*

팀장

데이터 수집과 전처리, 모델 생성

홍성*

팀원

데이터 수집과 전처리, 모델 생성

장영*

부팀장

프론트엔드, AWS

박주*

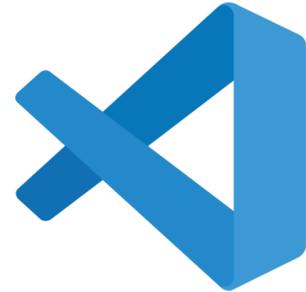
팀원

백엔드, 지도 API 구현

이기*

팀원

데이터 수집과 시각화



I.개요 - 주제 선정 배경



zibang 아파트 매매/전월세/신축분양 빌라, 투룸+ 신축분양/매매/전월세 원룸 전월세 오피스텔 도시형생활주택/전월세 창업/사무실 임대/매매

매매/전·월세 신축분양 인구흐름

강남역 검색 조건을 설정해주세요. [입력 필터]

필터 모두 초기화

면적 전체

전체	10평 이하	10평대	20평대
	30평대	40평대	50평대
		60평대	60평 이상

준공년도 전체

전체	5년 이내	10년 이내
	15년 이내	20년 이내
		25년 이내

12㎡ 이하 30㎡ 이하 32㎡ 이하

30㎡ 이하 2㎡ 이하 10㎡ 이하

30㎡대 40㎡대 20㎡대 90㎡ 이상

현장

층수 중복선택이 가능합니다.

전체 1층 2층 3층 4층 5층 6층 7층 이상 반지층 옥탑방

방구조 중복선택이 가능합니다.

주방 분리형(1.5룸) 복층 투룸 쓰리룸 이상

세대수

전체 100세대 이하 100세대 이상 500세대 이상 1000세대 이상

주차대수

상관없음 세대당 2대 이상 세대당 1대 이상

거래종류

전체 중개 직거래

추가 옵션 중복선택이 가능합니다.

주차가능 단기임대 풀옵션 엘리베이터 베란다

보안/안전시설 360° VR 비대면계약

i 풀옵션 항목은 세탁기, 냉장고, 에어컨, 가스레인지 또는 인덕션입니다.

100/85 평균전세 2억1400

100/82 유독외계 3억1400

2K레크티 유독외계 3억1400

21 유독외계 3억1400

32 유독외계 3억1400

35 유독외계 3억1400

38 유독외계 3억1400

41 유독외계 3억1400

44 유독외계 3억1400

47 유독외계 3억1400

50 유독외계 3억1400

53 유독외계 3억1400

56 유독외계 3억1400

59 유독외계 3억1400

62 유독외계 3억1400

65 유독외계 3억1400

68 유독외계 3억1400

71 유독외계 3억1400

74 유독외계 3억1400

77 유독외계 3억1400

80 유독외계 3억1400

83 유독외계 3억1400

86 유독외계 3억1400

89 유독외계 3억1400

92 유독외계 3억1400

95 유독외계 3억1400

98 유독외계 3억1400

101 유독외계 3억1400

104 유독외계 3억1400

107 유독외계 3억1400

110 유독외계 3억1400

113 유독외계 3억1400

116 유독외계 3억1400

119 유독외계 3억1400

122 유독외계 3억1400

125 유독외계 3억1400

128 유독외계 3억1400

131 유독외계 3억1400

134 유독외계 3억1400

137 유독외계 3억1400

140 유독외계 3억1400

143 유독외계 3억1400

146 유독외계 3억1400

149 유독외계 3억1400

152 유독외계 3억1400

155 유독외계 3억1400

158 유독외계 3억1400

161 유독외계 3억1400

164 유독외계 3억1400

167 유독외계 3억1400

170 유독외계 3억1400

173 유독외계 3억1400

176 유독외계 3억1400

179 유독외계 3억1400

182 유독외계 3억1400

185 유독외계 3억1400

188 유독외계 3억1400

191 유독외계 3억1400

194 유독외계 3억1400

197 유독외계 3억1400

200 유독외계 3억1400

203 유독외계 3억1400

206 유독외계 3억1400

209 유독외계 3억1400

212 유독외계 3억1400

215 유독외계 3억1400

218 유독외계 3억1400

221 유독외계 3억1400

224 유독외계 3억1400

227 유독외계 3억1400

230 유독외계 3억1400

233 유독외계 3억1400

236 유독외계 3억1400

239 유독외계 3억1400

242 유독외계 3억1400

245 유독외계 3억1400

248 유독외계 3억1400

251 유독외계 3억1400

254 유독외계 3억1400

257 유독외계 3억1400

260 유독외계 3억1400

263 유독외계 3억1400

266 유독외계 3억1400

269 유독외계 3억1400

272 유독외계 3억1400

275 유독외계 3억1400

278 유독외계 3억1400

281 유독외계 3억1400

284 유독외계 3억1400

287 유독외계 3억1400

290 유독외계 3억1400

293 유독외계 3억1400

296 유독외계 3억1400

299 유독외계 3억1400

302 유독외계 3억1400

305 유독외계 3억1400

308 유독외계 3억1400

311 유독외계 3억1400

314 유독외계 3억1400

317 유독외계 3억1400

320 유독외계 3억1400

323 유독외계 3억1400

326 유독외계 3억1400

329 유독외계 3억1400

332 유독외계 3억1400

335 유독외계 3억1400

338 유독외계 3억1400

341 유독외계 3억1400

344 유독외계 3억1400

347 유독외계 3억1400

350 유독외계 3억1400

353 유독외계 3억1400

356 유독외계 3억1400

359 유독외계 3억1400

362 유독외계 3억1400

365 유독외계 3억1400

368 유독외계 3억1400

371 유독외계 3억1400

374 유독외계 3억1400

377 유독외계 3억1400

380 유독외계 3억1400

383 유독외계 3억1400

386 유독외계 3억1400

389 유독외계 3억1400

392 유독외계 3억1400

395 유독외계 3억1400

398 유독외계 3억1400

401 유독외계 3억1400

404 유독외계 3억1400

407 유독외계 3억1400

410 유독외계 3억1400

413 유독외계 3억1400

416 유독외계 3억1400

419 유독외계 3억1400

422 유독외계 3억1400

425 유독외계 3억1400

428 유독외계 3억1400

431 유독외계 3억1400

434 유독외계 3억1400

437 유독외계 3억1400

440 유독외계 3억1400

443 유독외계 3억1400

446 유독외계 3억1400

449 유독외계 3억1400

452 유독외계 3억1400

455 유독외계 3억1400

458 유독외계 3억1400

461 유독외계 3억1400

464 유독외계 3억1400

467 유독외계 3억1400

470 유독외계 3억1400

473 유독외계 3억1400

476 유독외계 3억1400

479 유독외계 3억1400

482 유독외계 3억1400

485 유독외계 3억1400

488 유독외계 3억1400

491 유독외계 3억1400

494 유독외계 3억1400

497 유독외계 3억1400

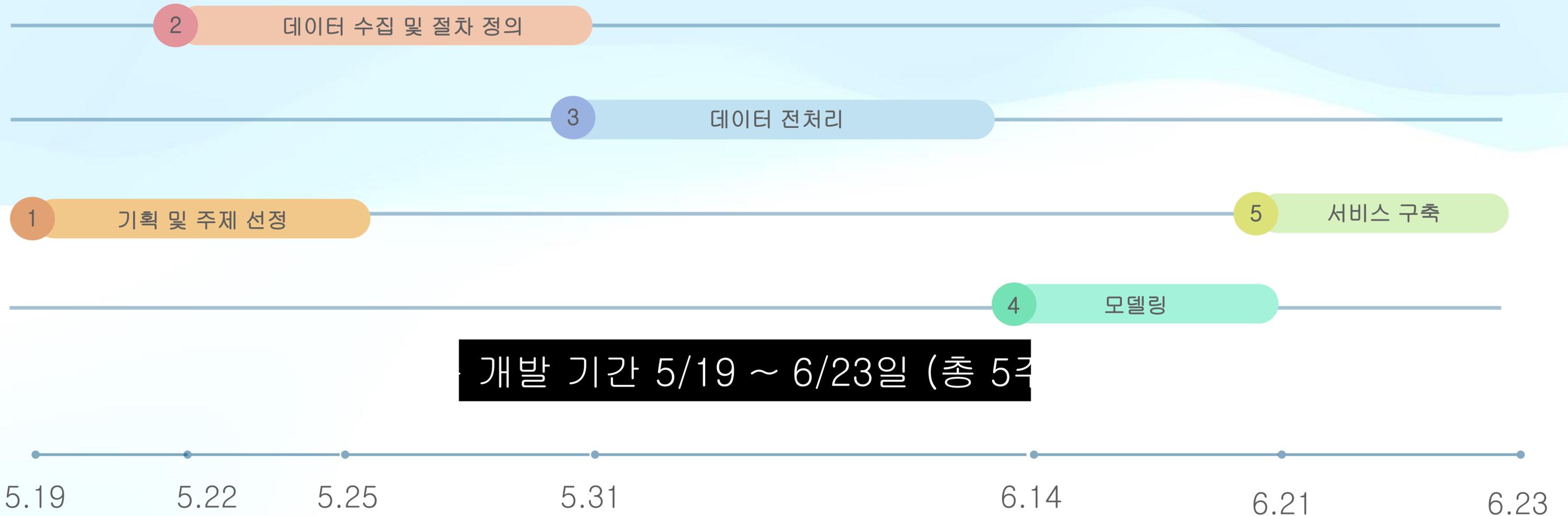
500 유독외계 3억1400



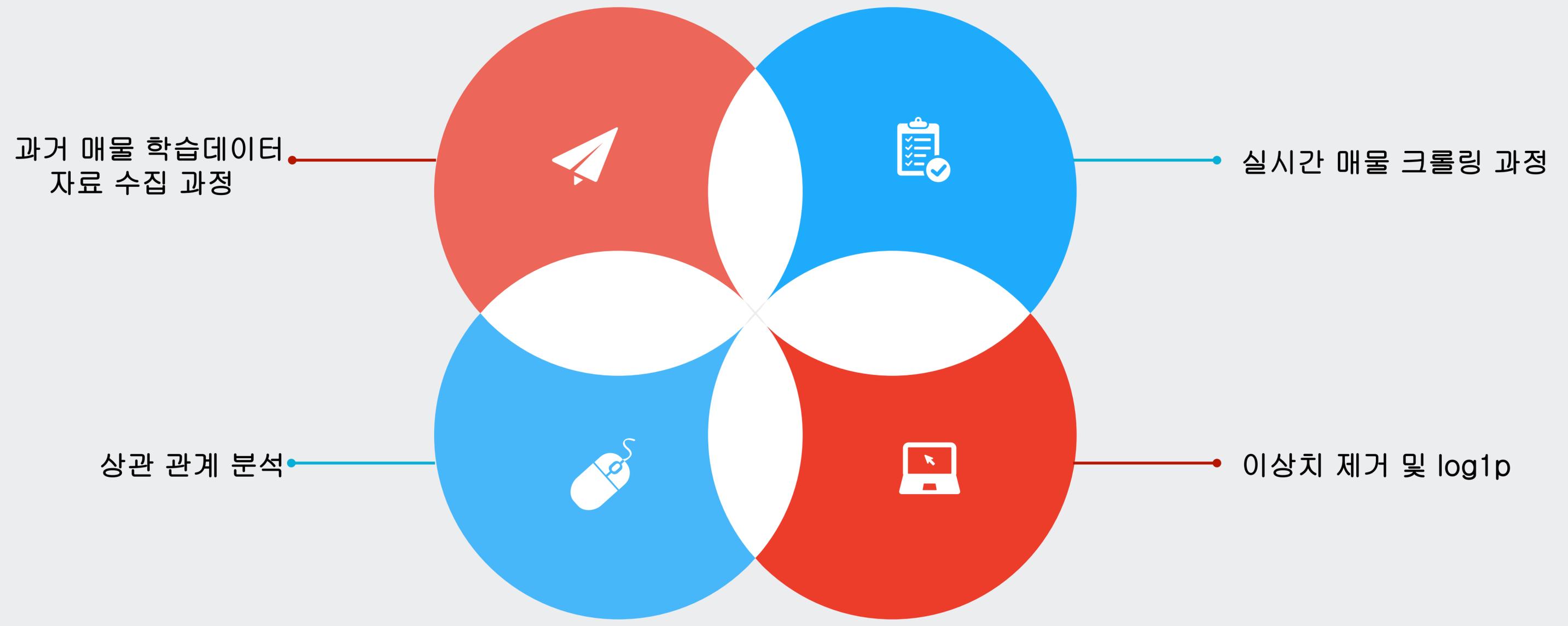
The image shows a screenshot of the Zibang real estate search interface. The interface includes a search bar at the top, a navigation menu with categories like '아파트' (Apartment), '빌라, 투룸+' (Villa, 2+ rooms), '원룸' (Studio), '오피스텔' (Office텔), and '창업/사무실' (Startup/Office). Below the search bar, there are filters for '매매/전·월세' (Buy/Sale/Rent), '신축분양' (New construction), and '인구흐름' (Population flow). The main area is a map showing various real estate listings with price tags. On the right side, there are several filter panels: '층수' (Floor count), '방구조' (Room structure), '세대수' (Number of units), '주차대수' (Number of parking spaces), '거래종류' (Transaction type), and '추가 옵션' (Additional options). A large white text overlay is centered on the map, reading '기존 사이트에서는 제한된 검색만 제공' (Only limited search is provided on the existing site).



체육시설, 스타벅스, 백화점이 인접한 매물 검색



II. 데이터수집 - 데이터 전처리



II. 데이터수집

과거 매물 학습 데이터 자료 수집 과정

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1	접수연도	자치구코드	자치구명	법정동코드	법정동명	지번구분	지번구분명	본번	부번	건물명	계약일	물건금액(원)	건물면적(㎡)	토지면적(㎡)	층	권리구분	취소일	건축년도	건물용도	신고구분	신고한 개업공인중개사	시군구	
2	2023	11560	영등포구	10200	영등포동1	1	대지	113	1	여의도씨티	2E+07	13300	12.59	0	13			2013	아파트	중개거래	서울 영등포구		
3	2023	11740	강동구	10900	천호동	1	대지	432	7	아스하임	2E+07	23000	12.88	0	19			2021	아파트	중개거래	서울 강동구		
4	2023	11305	강북구	10100	미아동	1	대지	217	29	청한에버뉴	2E+07	33000	72.66	36.82	2			2003	연립다세대	직거래			
5	2023	11590	동작구	10900	신대방동	1	대지	711	0	보라매파크	2E+07	68000	84.73	0	3			2002	아파트	직거래			
6	2023	11710	송파구	10700	가락동	1	대지	79	7	(79-7)	2E+07	12400	29.52	38.86	3			1990	오피스텔	중개거래	서울 강북구, 서울 송파구		
7	2023	11410	서대문구	11900	북가좌동	1	대지	75	139	태양아트빌	2E+07	40000	32.91	17	2			1992	연립다세대	직거래			
8	2023	11530	구로구	10200	구로동	1	대지	103	6	구로 오네	2E+07	17500	21.39	30.55	9			2022	오피스텔	중개거래	서울 구로구		
9	2023	11230	동대문구	10700	청량리동	1	대지	235	6	힐스테이트	2E+07	26400	21.2	29.24	16			2023	오피스텔	중개거래	서울 동대문구		
10	2023	11740	강동구	10900	천호동	1	대지	291	60	291에비뉴	2E+07	38000	37.07	23.14	2			2017	연립다세대	직거래			
11	2023	11710	송파구	10700	가락동	1	대지	913	0	헬리오시티	2E+07	115000	39.86	0	3			2018	아파트	중개거래	서울 강동구, 서울 송파구		
12	2023	11680	강남구	10100	역삼동	1	대지	832	33	성우스타스	2E+07	21400	32.03	41.12	4			2004	오피스텔	중개거래	서울 강남구		
13	2023	11440	마포구	12600	중동	1	대지	398	0	울트라월드	2E+07	95000	84.94	0	14			2006	아파트	중개거래	서울 마포구		
14	2023	11500	강서구	10100	염창동	1	대지	97	33	다담빌	2E+07	25400	49.59	25.56	6			2010	연립다세대	중개거래	서울 강서구		
15	2023	11350	노원구	10300	공릉동	1	대지	115	0	태릉현대빌	2E+07	106000	115.98	159.77	2			2004	연립다세대	중개거래	서울 노원구		
16	2023	11305	강북구	10300	수유동	1	대지	358	5	새로움빌	2E+07	24300	45.75	27.49	2			2016	연립다세대	중개거래	서울 강북구		
17	2023	11230	동대문구	10400	전농동	1	대지	6	0	전농우성	2E+07	57000	59.78	0	11			1992	아파트	중개거래	서울 동대문구		
18	2023	11110	종로구	11300	필운동						2E+07	45400	19.82	49.6				1960	단독다가	중개거래	서울 종로구		
19	2023	11320	도봉구	10500	쌍문동	1	대지	723	0	신원2	2E+07	34500	60	0	14			1998	아파트	중개거래	서울 도봉구		
20	2023	11710	송파구	10700	가락동	1	대지	140	0	가락(1차)빌	2E+07	109900	59.92	0	6			1997	아파트	중개거래	서울 송파구		
21	2023	11350	노원구	10600	중계동	1	대지	502	1	중계그린1	2E+07	52000	49.5	0	11			1990	아파트	중개거래	서울 노원구		
22	2023	11680	강남구	10600	대치동	1	대지	974	0	대치현대	2E+07	210500	85	0	4			1999	아파트	중개거래	서울 강남구		
23	2023	11380	은평구	10300	불광동	1	대지	448	53	조이빌	2E+07	23850	30.06	24.15	5			2015	연립다세대	중개거래	서울 은평구		
24	2023	11500	강서구	10900	방화동	1	대지	184	44	헤리티지	2E+07	29500	36.53	23.15	4			2015	연립다세대	중개거래	서울 강서구		
25	2023	11305	강북구	10100	미아동	1	대지	1363	0	수유지웰C	2E+07	9000	15.89	13.04	5			2015	연립다세대	중개거래	서울 강북구, 서울 서대문구		
26	2023	11305	강북구	10100	미아동	1	대지	1363	0	수유지웰C	2E+07	11000	19.74	16.2	3			2015	연립다세대	중개거래	서울 강북구, 서울 서대문구		

[2666598 rows x 20 columns]

II. 데이터수집

과거 매물 학습 데이터 자료 수집 과정

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1	접수연도	자치구코드	자치구명	법정동코드	법정동명	지번구분	지번구분명	본번	부번	건물명	계약일	물건금액(원)	건물면적(㎡)	토지면적(㎡)	층	권리구분	취소일	건축년도	건물용도	신고구분	신고한 개업공인중개사	시군구	
2	2023	1156	영등포구	10200	영등포동1	1	대지	113	1	여의도씨티	2E+07	13300	12.59	0	13			2013	아파트	중개거래	서울 영등포구		
3	2023	1174	강동구	10900	천호동	1	대지	432	7	아스하임	2E+07	23000	12.88	0	19			2021	아파트	중개거래	서울 강동구		
4	2023	1130	강북구	10100	미아동	1	대지	217	29	청한에버뉴	2E+07	33000	72.66	36.82	2			2003	연립다세대	직거래			
5	2023	1159	동작구	10900	신대방동	1	대지	711	0	보라매파크	2E+07	68000	84.73	0	3			2002	아파트	직거래			
6	2023	1171	송파구	10700	가락동	1	대지	79	7	(79-7)	2E+07	12400	29.52	38.86	3			1990	오피스텔	중개거래	서울 강북구, 서울 송파구		
7	2023	1141	서대문구	11900	북가좌동	1	대지	75	139	태양아트빌	2E+07	40000	32.91	17	2			1992	연립다세대	직거래			
8	2023	1153	구로구	10200	구로동	1	대지	103	6	구로 오네	2E+07	17500	21.39	30.55	9			2022	오피스텔	중개거래	서울 구로구		
9	2023	1123	동대문구	10700	청량리동	1	대지	235	6	힐스테이트	2E+07	26400	21.2	29.24	16			2023	오피스텔	중개거래	서울 동대문구		
10	2023	1174	강동구	10900	천호동	1	대지	291	60	291에비뉴	2E+07	38000	37.07	23.14	2			2017	연립다세대	직거래			
11	2023	1171	송파구	10700	가락동	1	대지	913	0	헬리오시티	2E+07	115000	39.86	0	3			2018	아파트	중개거래	서울 강동구, 서울 송파구		
12	2023	1168	강남구	10100	역삼동	1	대지	832	33	성우스타워	2E+07	21400	32.03	41.12	4			2004	오피스텔	중개거래	서울 강남구		
13	2023	1144	마포구	12600	중동	1	대지	398	0	울트라월드	2E+07	95000	84.94	0	14			2006	아파트	중개거래	서울 마포구		
14	2023	1150	강서구	10100	염창동	1	대지	97	33	다담빌	2E+07	25400	49.59	25.56	6			2010	연립다세대	중개거래	서울 강서구		
15	2023	1135	노원구	10300	공릉동	1	대지	115	0	태릉현대힐	2E+07	106000	115.98	159.77	2			2004	연립다세대	중개거래	서울 노원구		
16	2023	1130	강북구	10300	수유동	1	대지	358	5	새로움빌	2E+07	24300	45.75	27.49	2			2016	연립다세대	중개거래	서울 강북구		
17	2023	1123	동대문구	10400	전농동	1	대지	6	0	전농우성	2E+07	57000	59.78	0	11			1992	아파트	중개거래	서울 동대문구		
18	2023	1111	종로구	11300	필운동						2E+07	45400	19.82	49.6				1960	단독다가	중개거래	서울 종로구		
19	2023	1132	도봉구	10500	쌍문동	1	대지	723	0	신원2	2E+07	34500	60	0	14			1998	아파트	중개거래	서울 도봉구		
20	2023	1171	송파구	10700	가락동	1	대지	140	0	가락(1차)	2E+07	109900	59.92	0	6			1997	아파트	중개거래	서울 송파구		
21	2023	1135	노원구	10600	중계동	1	대지	502	1	중계그린1	2E+07	52000	49.5	0	11			1990	아파트	중개거래	서울 노원구		
22	2023	1168	강남구	10600	대치동	1	대지	974	0	대치현대	2E+07	210500	85	0	4			1999	아파트	중개거래	서울 강남구		
23	2023	1138	은평구	10300	불광동	1	대지	448	53	조이빌	2E+07	23850	30.06	24.15	5			2015	연립다세대	중개거래	서울 은평구		
24	2023	1150	강서구	10900	방화동	1	대지	184	44	헤리티지	2E+07	29500	36.53	23.15	4			2015	연립다세대	중개거래	서울 강서구		
25	2023	1130	강북구	10100	미아동	1	대지	1363	0	수유지웰0	2E+07	9000	15.89	13.04	5			2015	연립다세대	중개거래	서울 강북구, 서울 서대문구		
26	2023	1130	강북구	10100	미아동	1	대지	1363	0	수유지웰0	2E+07	11000	19.74	16.2	3			2015	연립다세대	중개거래	서울 강북구, 서울 서대문구		

[2666598 rows x 20 columns]

II. 데이터수집

과거 매물 학습 데이터 자료 수집 과정

	A	B	C	D	자치구명	법정동명	본번	부번	건물명	계약일	물건금액(만원)	건물면적(㎡)	건물용도	T	U	V	W
1	접수연도	자치구코드	자치구명	법정동코드	자치구명	법정동명	본번	부번	건물명	계약일	물건금액(만원)	건물면적(㎡)	건물용도	신고구분	신고한	개업공인중개사	시군구
2	2023	11560	영등포구	10200	영등포구	길음동	1284	0	길음뉴타운	20221231	71000	59.99	아파트	중개거래	서울	영등포구	
3	2023	11740	강동구	10900	강남구	대치동	316	0	은마	20221231	184000	76.79	아파트	중개거래	서울	강동구	
4	2023	11305	강북구	10100	동작구	상도동	531	0	힐스테이트	20221231	115000	84.61	아파트	직거래			
5	2023	11590	동작구	10900	성북구	석관동	410	0	래미안아트	20221231	69500	59.98	아파트	직거래			
6	2023	11710	송파구	10700	성북구	석관동	410	0	래미안아트	20221231	69500	59.98	아파트	중개거래	서울	강북구, 서울	송파구
7	2023	11410	서대문구	11900	동대문구	용두동	792	0	래미안 허브	20221231	97000	84.97	아파트	직거래			
8	2023	11530	구로구	10200	동대문구	용두동	792	0	래미안 허브	20221231	97000	84.97	아파트	중개거래	서울	구로구	
9	2023	11230	동대문구	10700	서초구	서초동	1682	0	서초래미안	20221231	186000	84.97	아파트	중개거래	서울	동대문구	
10	2023	11740	강동구	10900	서초구	서초동	1682	0	서초래미안	20221231	186000	84.97	아파트	직거래			
11	2023	11710	송파구	10700	동대문구	용두동	792	0	래미안 허브	20221231	80000	59.98	아파트	중개거래	서울	강동구, 서울	송파구
12	2023	11680	강남구	10100	동대문구	용두동	11	1	청량리역 허브	20221231	90452	84.87	아파트	중개거래	서울	강남구	
13	2023	11440	마포구	12600	동대문구	용두동	11	1	청량리역 허브	20221231	90452	84.87	아파트	중개거래	서울	마포구	
14	2023	11500	강서구	10100	동작구	신대방동	710	0	경남교수	20221231	86500	84.94	아파트	중개거래	서울	강서구	
15	2023	11350	노원구	10300	동작구	신대방동	710	0	경남교수	20221231	86500	84.94	아파트	중개거래	서울	노원구	
16	2023	11305	강북구	10300	구로구	신도림동	692	0	디큐브시티	20221231	85000	59.88	아파트	중개거래	서울	강북구	
17	2023	11230	동대문구	10400	송파구	문정동	150	0	올림픽훼밀	20221231	138000	84.71	아파트	중개거래	서울	동대문구	
18	2023	11110	종로구	11300	송파구	문정동	150	0	올림픽훼밀	20221231	138000	84.71	아파트	중개거래	서울	종로구	
19	2023	11320	도봉구	10500	송파구	방이동	89	0	올림픽선수	20221231	185000	100.82	아파트	중개거래	서울	도봉구	
20	2023	11710	송파구	10700	송파구	방이동	89	0	올림픽선수	20221231	185000	100.82	아파트	중개거래	서울	송파구	
21	2023	11350	노원구	10600	중랑구	면목동	1502	0	현대	20221231	69250	84.09	아파트	중개거래	서울	노원구	
22	2023	11680	강남구	10600	송파구	신천동	17	0	파크리오	20221231	172000	84.9	아파트	중개거래	서울	강남구	
23	2023	11380	은평구	10300	송파구	신천동	17	0	파크리오	20221231	172000	84.9	아파트	중개거래	서울	은평구	
24	2023	11500	강서구	10900	동대문구	이문동	225	0	대림e-편한	20221231	76500	84.97	아파트	중개거래	서울	강서구	
25	2023	11305	강북구	10100	동대문구	이문동	225	0	대림e-편한	20221231	76500	84.97	아파트	중개거래	서울	강북구, 서울	서대문구
26	2023	11305	강북구	10100	서초구	서초동	1312	3	롯데캐슬클	20221231	194000	84.98	아파트	중개거래	서울	강북구, 서울	서대문구

[334600 rows x 9 columns]

II. 데이터수집

과거 매출 학습 데이터 자료 수집 과정



```
api_key = "6900291ad1abb5455af2b904dde0193f"
max_retries = 100

def addr_to_lat_lon(addr):
    url = 'https://dapi.kakao.com/v2/local/search/address.json?query={address}'.format(addr)
    headers = {"Authorization": "KakaoAK " + api_key}
    retries = 0
    while retries < max_retries:
        try:
            result = json.loads(str(requests.get(url, headers=headers).text))
            if 'documents' in result and len(result['documents']) > 0:
                match_first = result['documents'][0]['address']
                return float(match_first['y']), float(match_first['x'])
            else:
                print("Warning: No result found for address:", addr)
                return None, None
        except (IndexError, KeyError):
            print("Error: Failed to retrieve address:", addr)
            retries += 1
            time.sleep(1) # wait for 1 second before retrying
    return None, None
```

경도위도
(37.495409757977, 126.950873258%)
(37.5205366173754, 127.10615089%)
(37.5781737388209, 127.04220384%)
(37.6060457872052, 127.06361443%)
(37.6075574804896, 127.01704870%)
(37.5753483700000, 127.01704870%)
(37.5762581100000, 127.01704870%)
(37.6045229100000, 127.01704870%)
(37.5711000000000, 127.01704870%)

- 금리.csv
- 매매_지수.csv
- 물가지수.csv
- 한강공원 주차장 정보.csv

```
from haversine import haversine

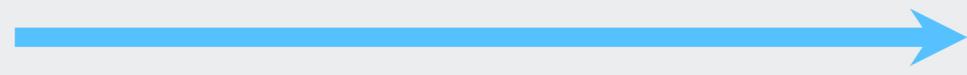
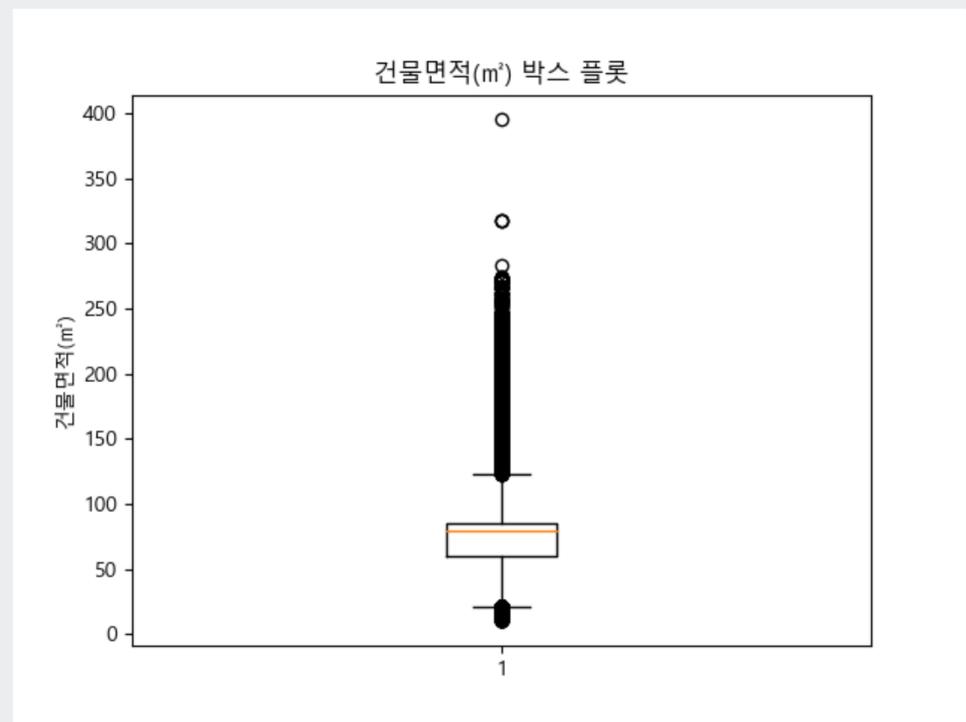
for s in range(len(df2)):
    hangang=df2['경도위도'][s].split(',')
    hangang_list=list(map(float, hangang))
    hangang_pos=(hangang_list[0], hangang_list[1])
    dist=haversine(house_pos, hangang_pos, unit='m')
    # print(dist)
    if dist < 1000:
        hangang_cnt+=1
    if dist < hangang_dist:
        hangang_dist=dist
    hangang_type=df2['한강최소거리'][s]
```

한강시민공원의 경우
수평적 거리에서는 가까울수록 가격이 상승하는 만큼 긍정적인
효과들이 많다고 하였다.

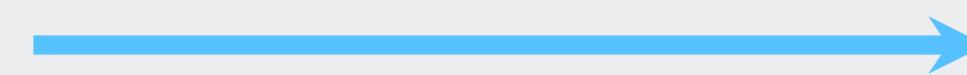
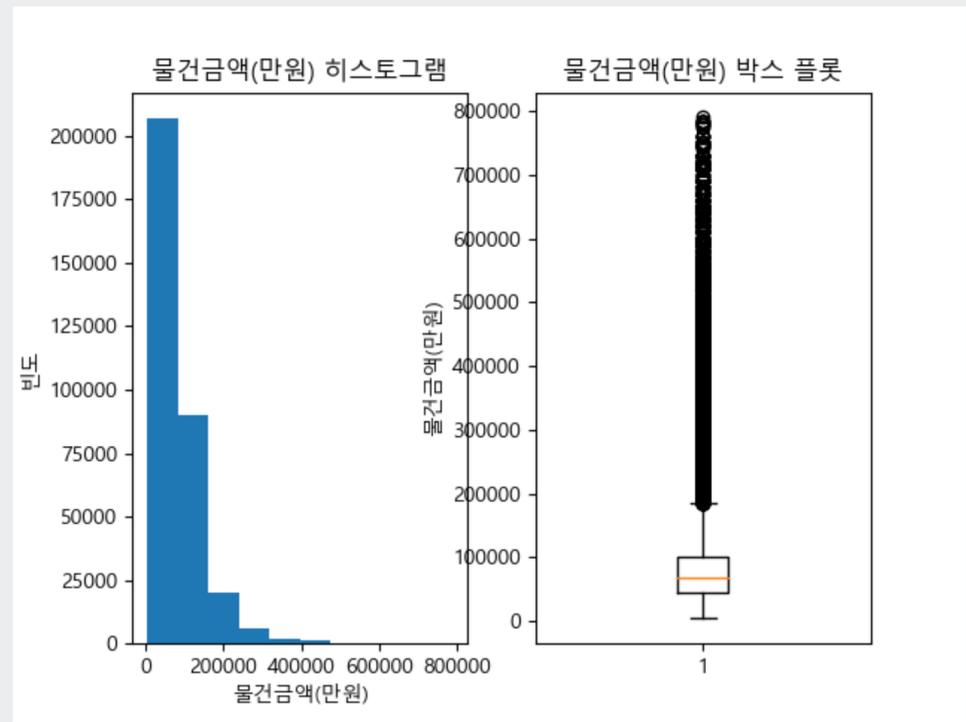
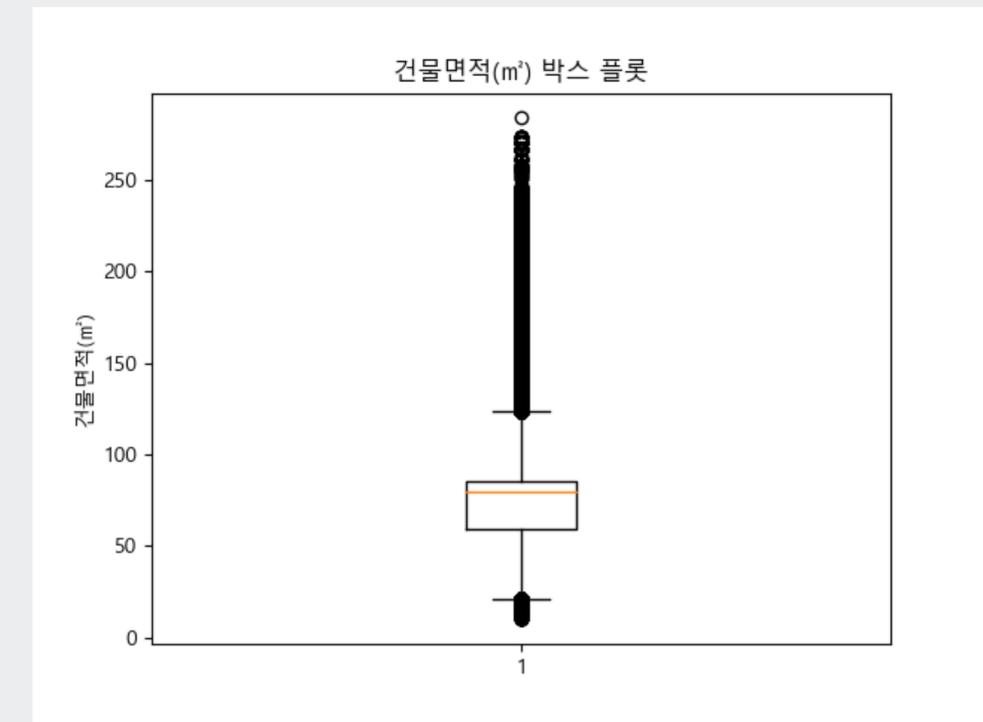
한강공원최물가지수	매매지수	금리	
2779.468	108.47	98.44151	3.25
1073.77	108.47	93.85927	3.25

II. 데이터수집

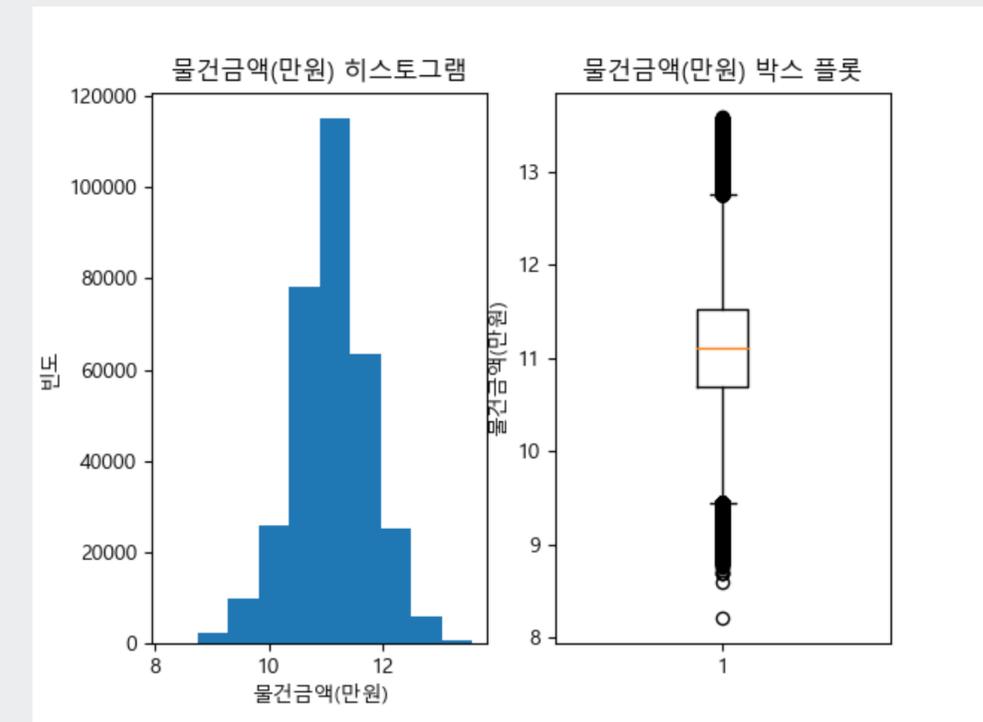
이상치 제거 및 log1p



건물 금액 80억 이
하
건물 면적 100평 이
하



종속변수 정규화



II. 데이터수집

모델의 정확성

이상치 제거
80억, 면적 100평 이상 제거

왜 $\log_{10} p$?

$\log_{10} p$ 정규화를 통해 종속 변수 정규화

II. 데이터수집

과거 매출 학습 데이터 자료 수집 과정

```
data['자치구명'] = le.fit_transform(data['자치구명'])
자치구명 = le.classes_
data['법정동명'] = le.fit_transform(data['법정동명'])
법정동명 = le.classes_
```

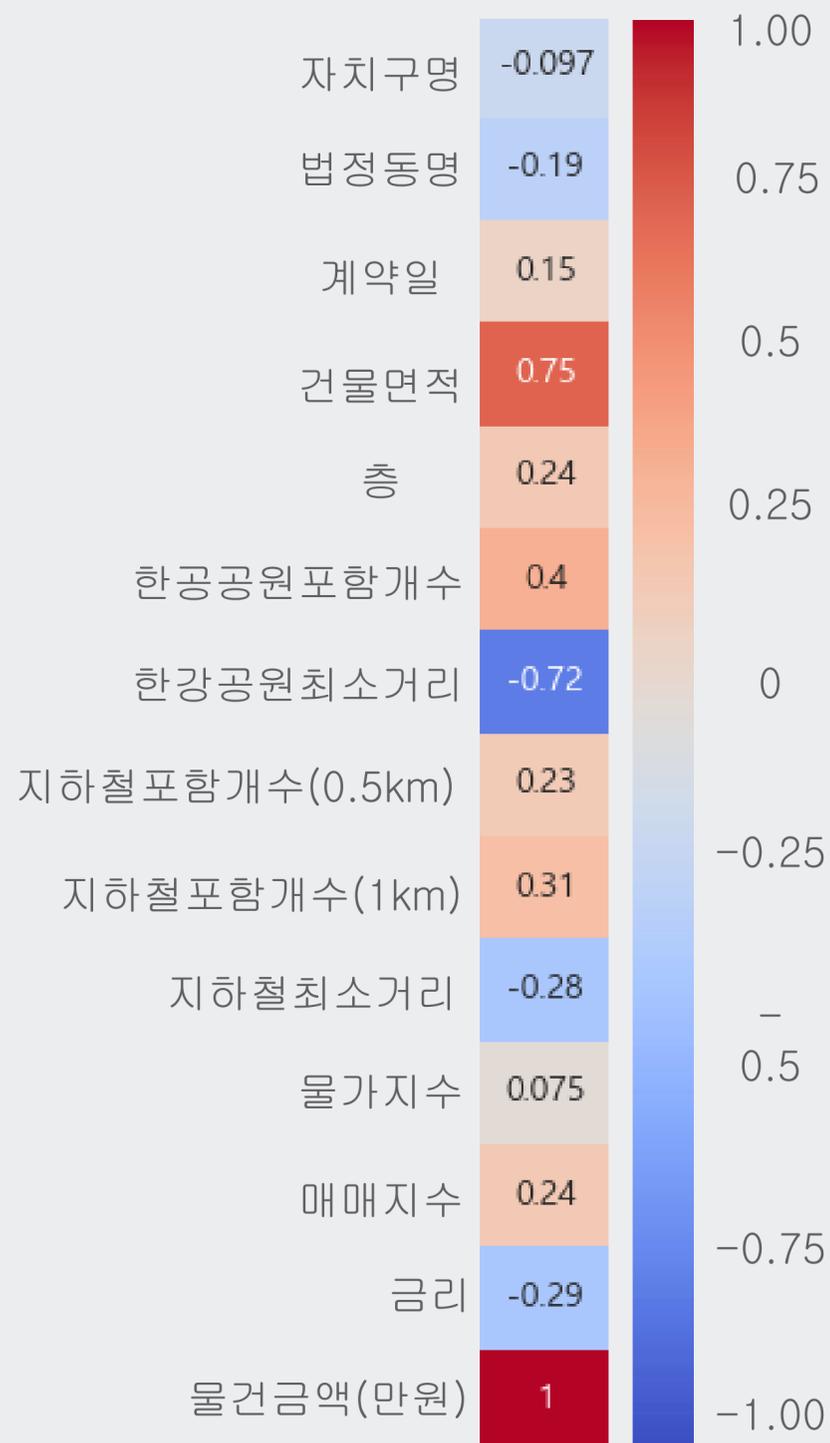
```
뷰웁웁 = J6·CJ92262
```

범주형 변수 라벨링

자치구	법정동
강남구	가락동
강동구	가리봉동
강북구	가산동
강서구	가양동
관악구	갈월동
광진구	갈현동
구로구	강일동
금천구	개봉동
노원구	개포동
도봉구	거여동
동대문구	견지동
동작구	고덕동

II. 데이터수집

독립변수와 종속변수 간의 상관관계 확인



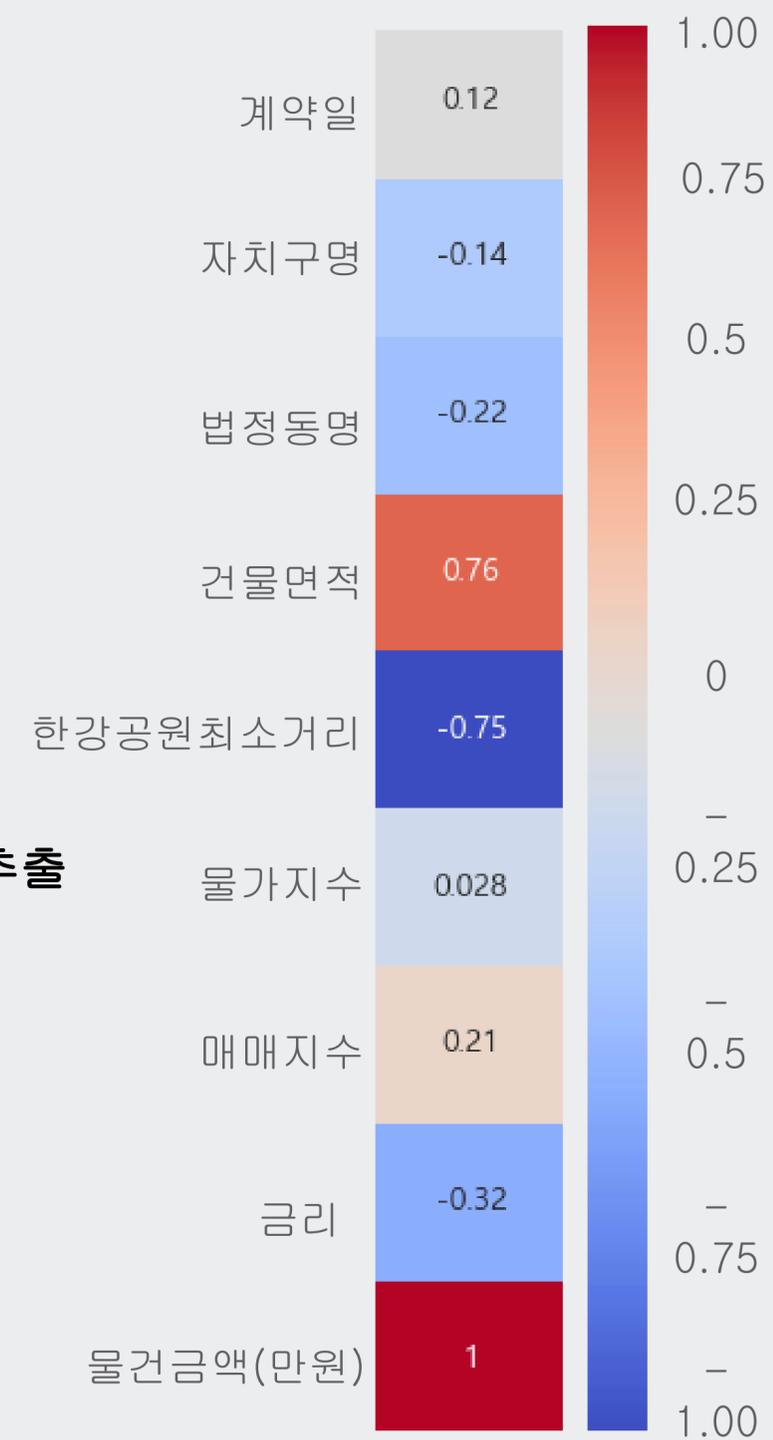
1. 변수 설정

(모든 독립변수에 대한 상관계수 Heatmap 출력)



2. 상관계수 확인 후 학습에 사용할 독립변수를 추출

(필요 변수 추출 후 Heatmap 출력)



II. 데이터수집 - 통계

실시간 매물 크롤링 과정

```
down_url = 'https://m.land.naver.com/map/getRegionList?cortarNo=1100000000&mycortarNo=1100000000'
r = requests.get(down_url, data={"sameAddressGroup": "false"}, headers={
    "Accept-Encoding": "gzip",
    "Host": "new.land.naver.com",
    "Referer": "https://m.land.naver.com/",
    "Sec-Fetch-Dest": "empty",
    "Sec-Fetch-Mode": "cors",
    "Sec-Fetch-Site": "same-origin",
    "User-Agent": "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/114.0
})
```

25개구 경도, 위도 파악

```
{
  "code": "success",
  "hasPaidPreSale": false,
  "more": true,
  "TIME": false,
  "z": 12,
  "page": 1,
  "body": [
    {
      "atclNm": "빌라",
      "atclStatCd": "R0",
      "rletTpCd": "C02",
      "uprRletTpCd": "C03",
      "rletTpNm": "빌라",
      "tradTpNm": "월세",
      "vrfcTpCd": "SITE",
      "flrInfo": "1/2",
      "prc": 10000,
      "rentPrc": 220,
      "hanPrc": "1억",
      "spc1": "63",
      "spc2": "58.73",
      "direction": "남향",
      "atclCfmYmd": "23.06.20.",
      "repImgUrl": "/20230620_139/land_naver_1687260541007Gjr0M_JPEG/1f9f1053-0f5c-11ee-92e7-a0369ffdebe4_03.jpg",
      "repImgTpCd": "SITEV",
      "repImgThumb": "f130_98",
      "lat": 37.502324,
      "lng": 127.063513,
      "atclFetrDesc": "신축, 신혼부부추천, 방3화2, 주차면리, 휘문중, 고 배정",
      "tagList": ["2년이내", "용자금없는", "1층"],
      "bidNm": "",
      "minute": 0,
      "sameAddrCnt": 3,
      "sameAddrDirectCnt": 0,
      "sameAddrHash": "21C02B2N85e00ec19d76c9a9661646696bfc31573e875279b4388bc2ca15ef91be2615a8",
      "sameAddrMaxPrc": "1억",
      "sameAddrMaxPrc2": "220",
      "sameAddrMinPrc": "1억",
      "sameAddrMinPrc2": "220",
      "cpNm": "NEONET",
      "cpCnt": 1,
      "rletNm": "에이블빌딩 무중산중개",
      "directTradYn": "N",
      "minMviFee": 0,
      "maxMviFee": 0,
      "etRoomCnt": 0,
      "tradePriceHan": "",
      "tradeRentPrice": 0,
      "tradeCheckedByOwner": false,
      "cpLinkV0": {
        "cpId": "NEONET",
        "mobileArticleLinkTypeCode": "NONE",
        "mobileBmsInspectPassYn": "Y",
        "pcArticleLinkUseAtArticleTitle": false,
        "pcArticleLinkUseAtCpName": false,
        "mobileArticleLinkUseAtArticleTitle": false,
        "mobileArticleLinkUseAtCpName": false
      },
      "dtlAddrYn": "N",
      "dtlAddr": "",
      "atclNo": "2326114686",
      "cortarNo": "1168010300",
      "atclNm": "빌라",
      "atclStatCd": "R0",
      "rletTpCd": "C02",
      "uprRletTpCd": "C03",
      "rletTpNm": "빌라",
      "tradTpCd": "B1",
      "tradTpNm": "전세",
      "vrfcTpCd": "SITE",
      "flrInfo": "5/5",
      "prc": 55000,
      "rentPrc": 0,
      "hanPrc": "5억 5,000",
      "spc1": "67",
      "spc2": "57.67",
      "direction": "남서향",
      "atclCfmYmd": "23.06.20.",
      "repImgUrl": "/20230620_221/land_naver_1687266371353RbN7g_JPEG/d50d48302aac2bb035990aa16124d3ad.JPG",
      "repImgTpCd": "SITE",
      "repImgThumb": "f130_98",
      "lat": 37.47447,
      "lng": 127.054959,
      "atclFetrDesc": "통4화2 관리 잘 된 집, 방 많고 넓은 집 원하시는 분께 약 입니다",
      "tagList": ["10년이내", "용자금없는", "탑층"],
      "bidNm": "",
      "minute": 0,
      "sameAddrCnt": 1,
      "sameAddrDirectCnt": 0,
      "cpId": "SERVE",
      "cpNm": "무중산채브",
      "cpCnt": 1,
      "rletNm": "명인공인중개사사무소",
      "directTradYn": "N",
      "minMviFee": 0,
      "maxMviFee": 0,
      "etRoomCnt": 0,
      "tradePriceHan": "",
      "tradeRentPrice": 0,
      "tradeCheckedByOwner": false,
      "cpLinkV0": {
        "cpId": "SERVE",
        "mobileArticleUrl": "http://m.serve.co.kr/naver/rd.asp?UID=",
        "mobileArticleLinkTypeCode": "CPNAMEONLY",
        "mobileBmsInspectPassYn": "Y",
        "pcArticleLinkUseAtArticleTitle": false,
        "pcArticleLinkUseAtCpName": false,
        "mobileArticleLinkUseAtArticleTitle": false,
        "mobileArticleLinkUseAtCpName": true
      },
      "dtlAddrYn": "N",
      "dtlAddr": "",
      "atclNo": "2326085433",
      "cortarNo": "1168011400",
      "atclNm": "다가구",
      "atclStatCd": "R0",
      "rletTpCd": "C03",
      "uprRletTpCd": "C03",
      "rletTpNm": "단독/다가구",
      "tradTpCd": "B1",
      "tradTpNm": "전세",
      "vrfcTpCd": "SITE",
      "flrInfo": "2/3",
      "prc": 36000,
      "rentPrc": 0,
      "hanPrc": "3억 6,000",
      "spc1": "68",
      "spc2": "68.25",
      "direction": "남향"
    }
  ]
}
```

경도, 위도 파악한 데이터로 매물 데이터에 접근하여 JSON 데이터 추출

```
=['atclNm', 'tradTpNm', 'flrInfo', 'prc', 'hanPrc', 'spc1', 'spc2', 'lat', 'lng', 'tagList', '지역구']
```

	atclNm	tradTpNm	flrInfo	prc	hanPrc	spc1	spc2	lat	lng	tagList	지역구	rentPrc
0	빌라	전세	02월 05일	30000	3억	31	23.48	37.5093	127.0308	['4년이내',	강남구	0
1	다가구	월세	중/4	1000	1,000	34	33.06	37.49928	127.0411	['25년이상	강남구	67
2	다가구	월세	03월 03일	2000	2,000	33	33	37.5146	127.0417	['25년이상	강남구	85
3	빌라	전세	05월 06일	20000	2억	38	28.37	37.50906	127.0243	['25년이상	강남구	0

필요 변수만 지정하여 추출



II. 데이터수집 - 통계



III. 분석 및 결과

각 모델별 학습 시도 및 성능 비교

최적의 파라미터: {'learning_rate': 0.1, 'max_depth': 10, 'n_estimators': 1000}

```
# XGBRegressor 모델 정의
model = XGBRegressor()

# 탐색할 파라미터 그리드 설정 : 탐색할 값의 후보들을 리스트로 지정
param_xgb={
    'max_depth': [10, 30, 50],
    'learning_rate': [0.1, 0.01, 0.001],
    'n_estimators':[200, 300, 500, 1000]
}

grid_search = GridSearchCV(estimator=model, param_grid=param_xgb, cv=5, scoring='neg_mean_squared_error', refit=True, n_jobs=6,
grid_search.fit(x_train, y_train)

# 최적의 파라미터 출력
print('최적의 파라미터:', grid_search.best_params_)
print('점수: ',grid_search.best_score_)
```

기존 파라미터의 학습 오차율 : 12 %



GridSearchCV를 이용한
파라미터의 학습 오차율 : 6 %

III. 분석 및 결과

RandomForestRegressor (전통적 모델)

n_estimators=100

일반	보안	자세히	이전 버전
RanForRegmodel_n_estimators_100.h5			
파일 형식:	H5 파일(.h5)		
연결 프로그램:	앱 선택		<input type="button" value="변경(C)..."/>
위치:	C:\Users\whsung\OneDrive\바탕 화면\자료		
크기:	1.51GB (1,630,859,953 바이트)		
디스크 할당 크기:	1.51GB (1,630,863,360 바이트)		

VS

GridSearchCV

일반	보안	자세히	이전 버전
RanForRegmodel_n_estimators_1000.h5			
파일 형식:	H5 파일(.h5)		
연결 프로그램:	앱 선택		<input type="button" value="변경(C)..."/>
위치:	C:\Users\whsung\OneDrive\바탕 화면\자료		
크기:	16.2GB (17,461,118,275 바이트)		
디스크 할당 크기:	16.2GB (17,461,121,024 바이트)		

오차율 차이는 **0.01%**

III. 분석 및 결과

RandomForestRegressor (전통적 모델)



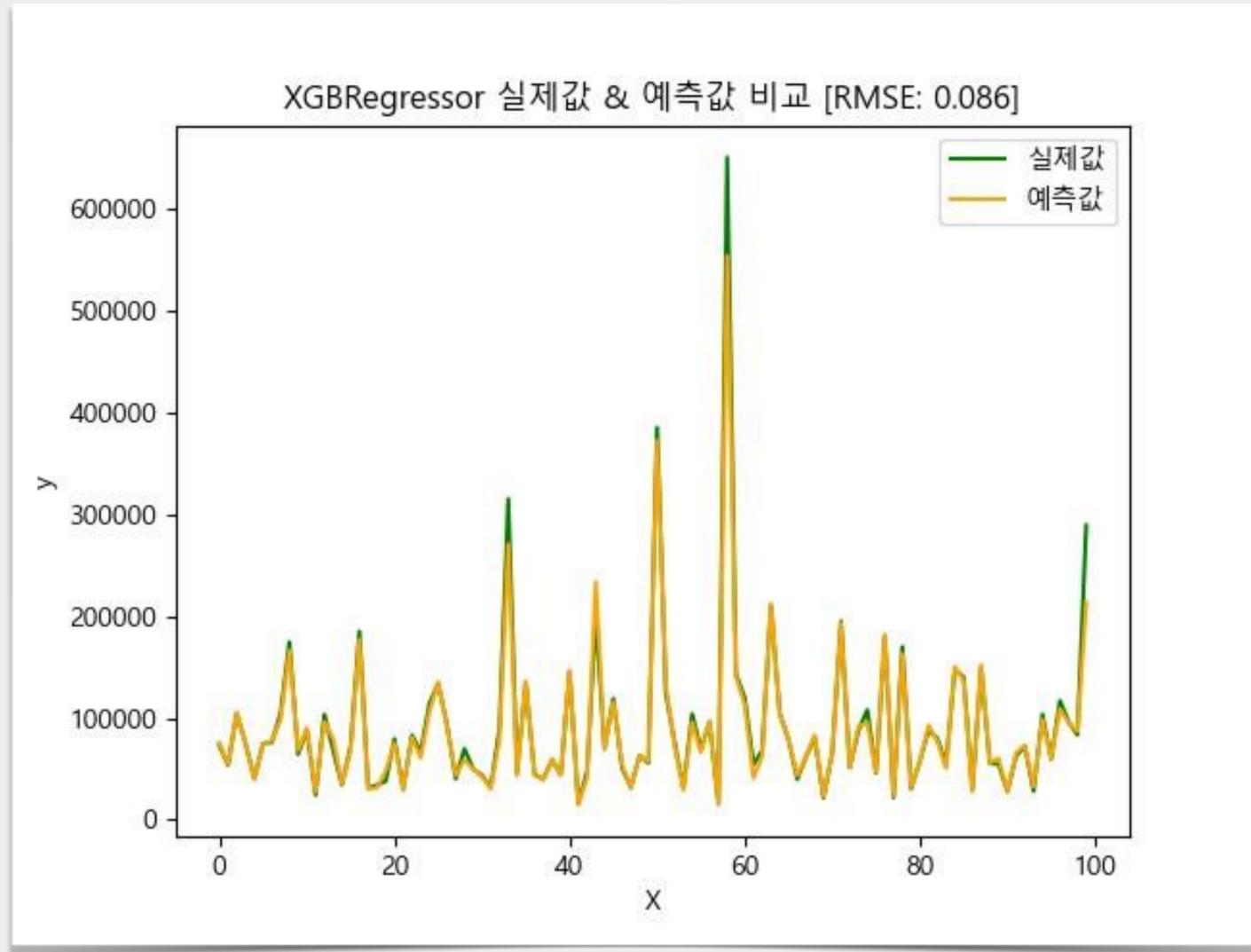
모델 용량이 클 수록 긴 로딩 시간 필요

0.01% 오차율 < 시스템 부하

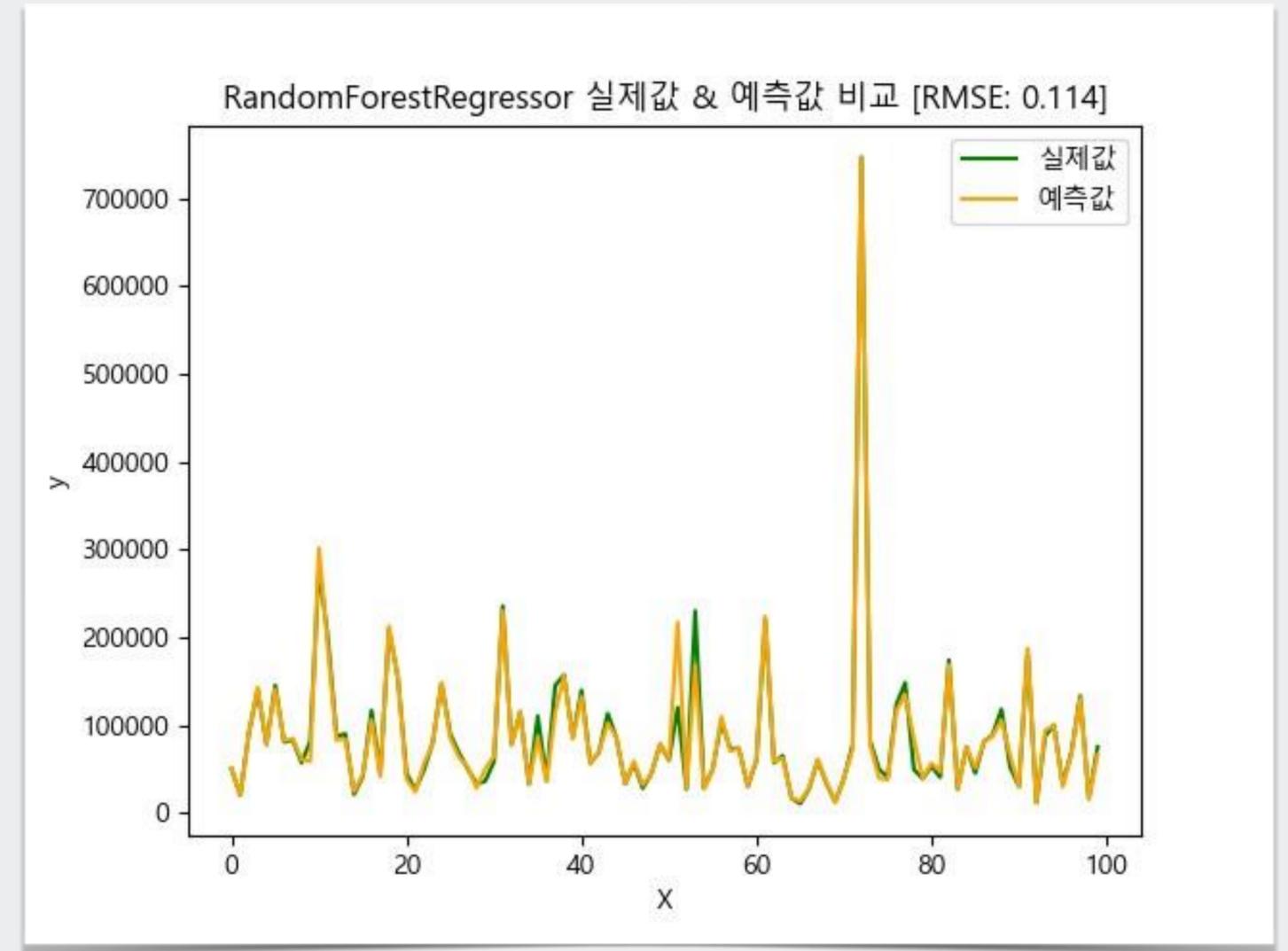
n_estimators=100 모델 선택

III. 분석 및 결과

실제값과 예측값 비교



극단값을 제외한, 값 예측이 비교적 정확



예측값과 실제값의 차이가 크게 나지
않음

III. 분석 및 결과

LSTM(Deep Learning)

datetime 타입으로 형변
화

```
x_data['계약일'] = pd.to_datetime(data['계약일'], format='%Y%m%d')
```

```
# LSTM 모델 구성
model = keras.Sequential([
    keras.layers.LSTM(32, input_shape=(x_train.shape[1], 1), return_sequences=True, activation='tanh'),
    keras.layers.LSTM(16, return_sequences=True, activation='tanh'),
    keras.layers.LSTM(1, activation='linear')
])
```

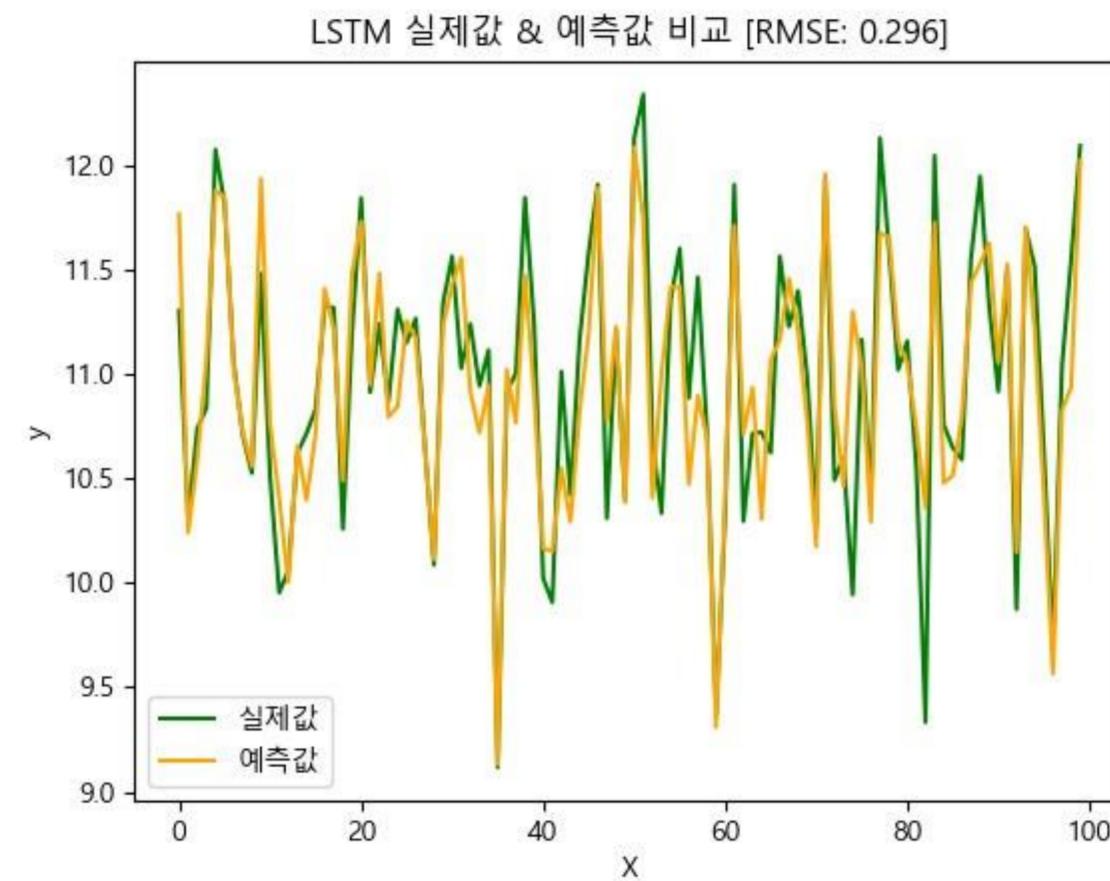
```
model.summary()
```

```
# 모델 컴파일
```

```
model.compile(optimizer=tf.keras.optimizers.Adam(learning_rate=0.01), loss='mse')
```

```
# 모델 학습
```

```
model.fit(x_train, y_train, epochs=100, batch_size=64, verbose=1)
```



III. 분석 및 결과

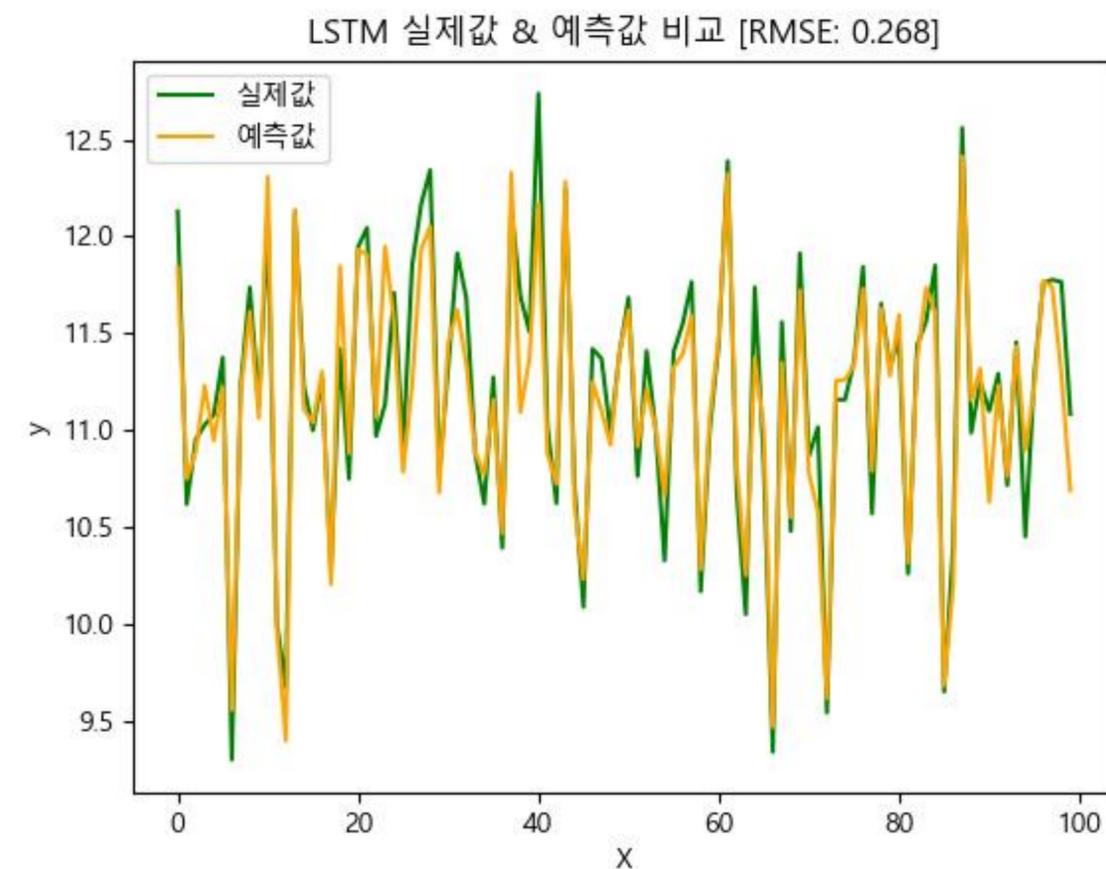
LSTM(Deep Learning)

```
# LSTM 모델 구성
model = keras.Sequential([
    keras.layers.LSTM(64, input_shape=(x_train.shape[1], 1), return_sequences=True),
    keras.layers.LSTM(32, return_sequences=True),
    keras.layers.LSTM(16, return_sequences=True),
    keras.layers.LSTM(8, return_sequences=True),
    keras.layers.LSTM(1, activation='linear')
])

model.summary()

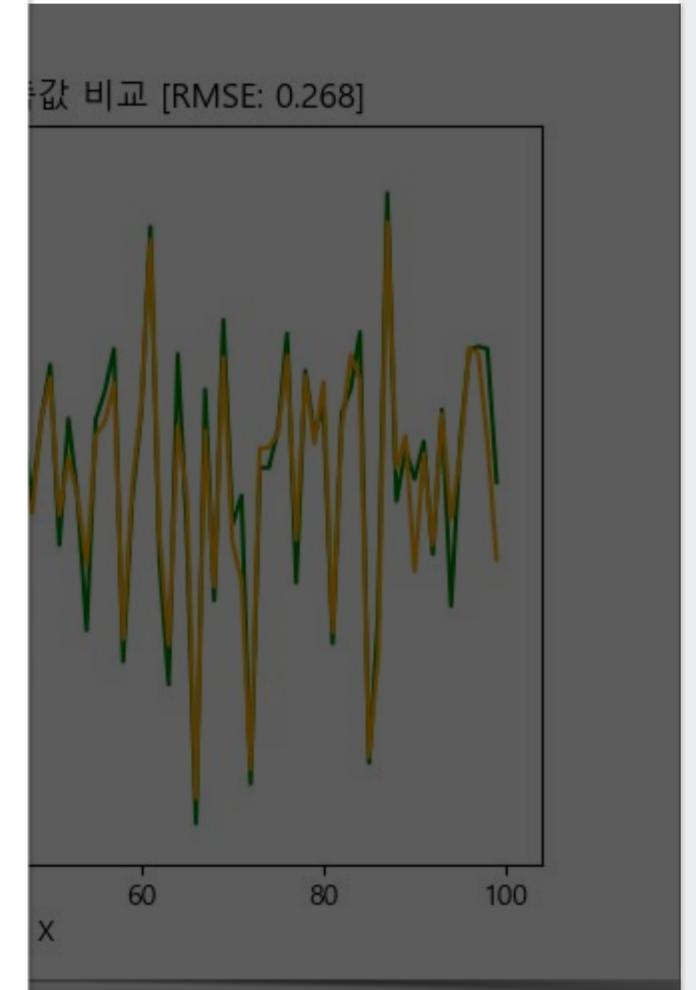
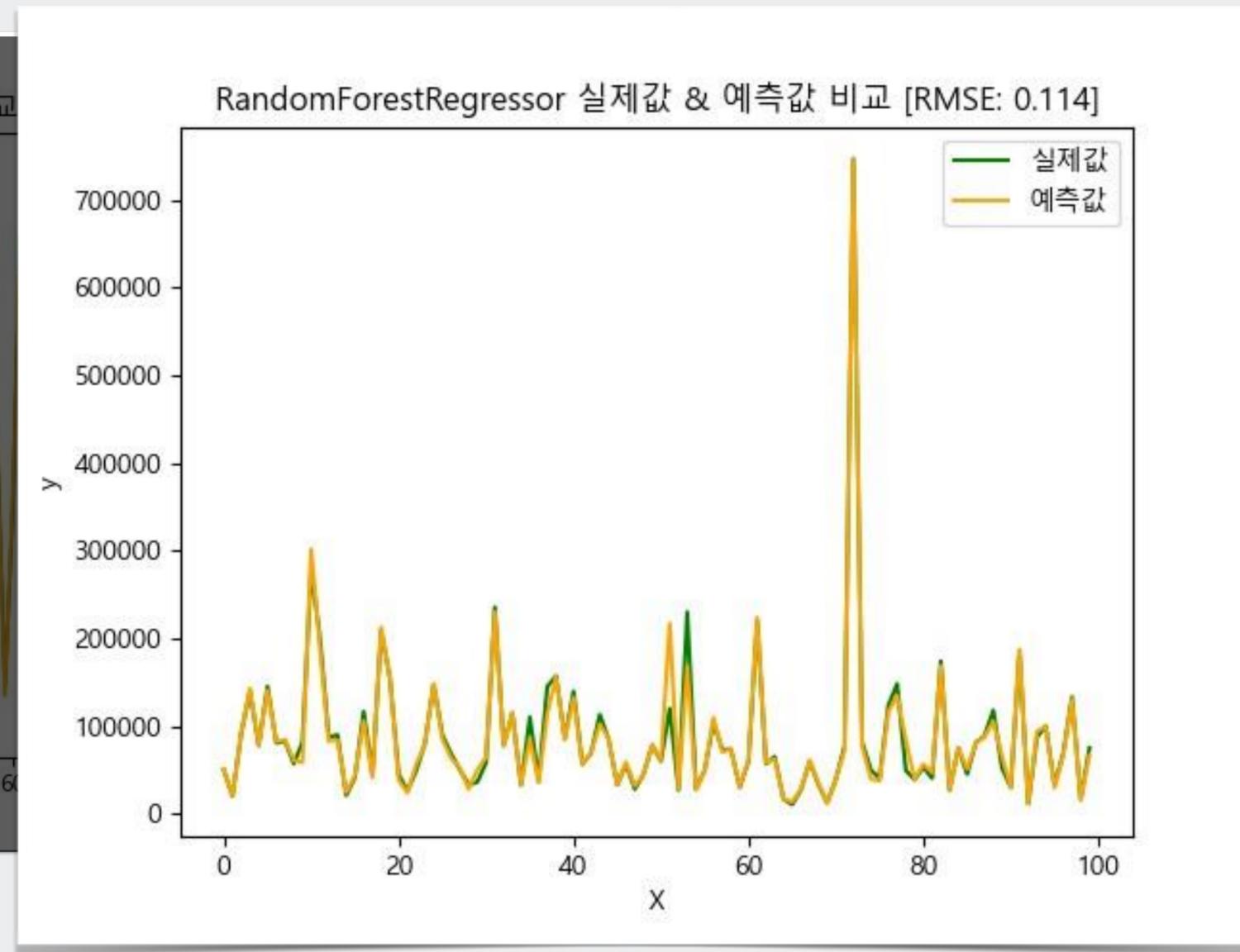
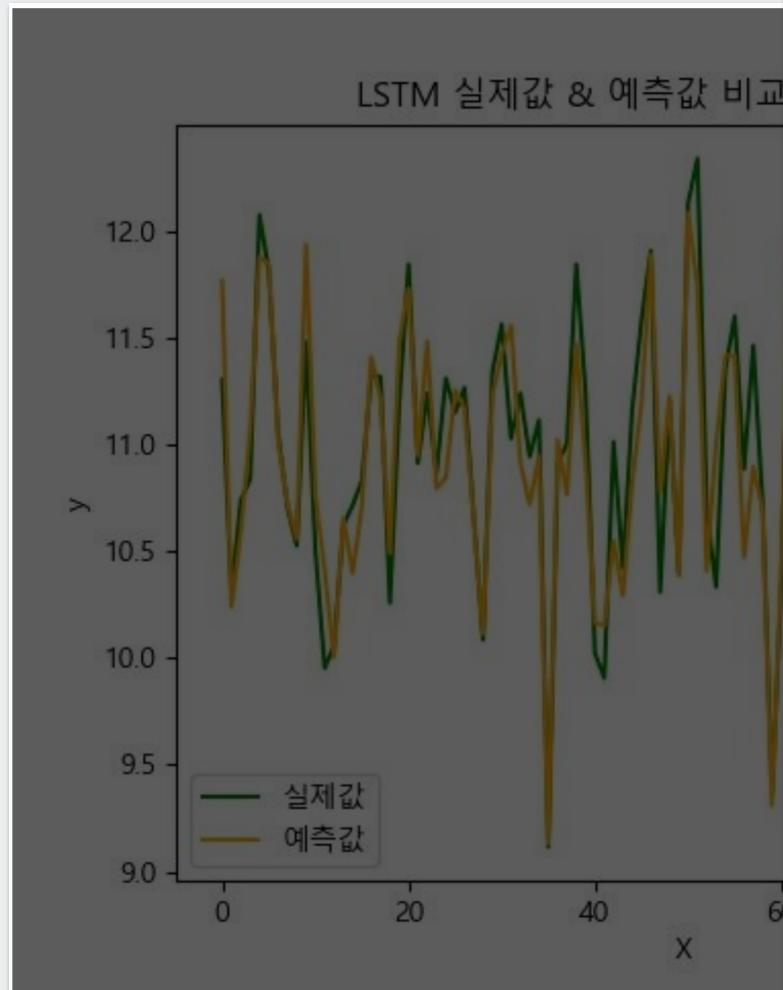
# 모델 컴파일
model.compile(optimizer=tf.keras.optimizers.Adam(learning_rate=0.01), loss='mse')

# 모델 학습
model.fit(x_train, y_train, epochs=50, batch_size=128, verbose=1)
pickle.dump(model, open('deeplearn.h5', 'wb'))
```



III. 분석 및 결과

LSTM(Deep Learning)



~~DEEP LEARNING~~ Machine learning

III. 분석 및 결과

RMSE & R-squared 값 이용

RandomForestRegressor 선택 이유

```
XGBRegressor 모델 RMSE & R-squared  
RMSE: 0.08414911244361757  
R-squared: 0.9844988878784471
```

XGBRegressor

```
RandomForestRegressor 모델 RMSE & R-squared  
RMSE: 0.11489435240733711  
R-squared: 0.9709111470054737
```

RandomForestRegressor

LOSS 값

xgboost vs RandomForestRegressor

실제 예측 결과

xgboost vs RandomForestRegressor

III. 분석 및 결과

RMSE & R-squared 값 이용

RandomForestRegressor 선택 이유

```
XGBRegressor 모델 RMSE & R-squared  
RMSE: 0.08414911244361757  
R-squared: 0.9844988878784471
```

XGBRegressor

```
RandomForestRegressor 모델 RMSE & R-squared  
RMSE: 0.11489435240733711  
R-squared: 0.9709111470054737
```

RandomForestRegressor

LOSS 값

xgboost vs RandomForestRegressor

실제 예측 결과

xgboost vs RandomForestRegressor

III. 분석 및 결과

구명 ▼

동명 ▼

6 평

예상가격 조회

내가 원하는 조건의 매물가격은?

만원

초기화

매물 추천

편의시설

헬스장
 백화점 · 대형마트
 스타벅스

원하는 조건을 선택해주세요

통계 정보

원하는 조건을 선택해주세요

웹 화면

III. 분석 및 결과

백엔드 주요 코드

강남구 ▾
대치동 ▾
15 평

ML 모델 ▾

예상가격 조회

내가 원하는 조건의 매물가격은?

_____ 만 원

초기화

원하는 아파트 매매 조건 선택 후 예상가격 조회 클릭

```

// 추천 매물 리스트 출력
function renderRecommendedProperties(properties) {
  let listItemsHTML = '';
  //console.log(properties.length)
  if (properties.length === 0){
    listItemsHTML = '<div class="text-center subtitle">조건에 맞는 매물이 없습니다.</div>'
    document.querySelector('#recom_list').innerHTML = listItemsHTML;
  } else {
    properties.forEach((property, index) => {
      // 가격 문자열 수정
      let formattedPrice = property.price;
      let lastChar = formattedPrice.slice(-1);
      if (lastChar === "억") {
        formattedPrice += "원";
      } else if (lastChar === "0") {
        formattedPrice += "만원";
      }
      let tagIndex=index+1;
      listItemsHTML += `
      <a href="#" class="list-group-item list-group-item-action" onclick="convenient(${property.lat}, ${property.lng})">
        <div class="d-flex w-100 justify-content-between">
          <p class="mb-1 maintitle">${tagIndex}.${property.title}</p>
          <small class="text-body-secondary h4">${formattedPrice}</small>
        </div>
        <div class='d-flex'>
          <span class="mb-1 h6">${property.desc}</span><span class='ms-auto maintitle'>${property.spec}평</span>
        </div>
      </a><hr>
      `;
    });
    document.querySelector('#recom_list').innerHTML = listItemsHTML;
  }
}

```

III. 분석 및 결과

백엔드 주요 코드

```

var select2 = ""
var select3 = ""
if ($("#flexCheckDefault1").is(':checked')) {
  select1 = $("#flexCheckDefault1").val();
}
if ($("#flexCheckDefault3").is(':checked')) {
  select2 = $("#flexCheckDefault3").val();
}
if ($("#flexCheckDefault4").is(':checked')) {
  select3 = $("#flexCheckDefault4").val();
}
$.ajax({
  url: 'convenient',
  type: 'GET',
  data: {
    'lat': lat,
    'lng': lng,
    'health': select1,
    'mart': select2,
    'starbucks': select3,
  },
  dataType: 'json',
  success: function(response) {
    // 체크박스에 따라 편의시설 좌표와 정보 생성되게 할 것.
    if (response.length > 0){
      reset_marker()
      move(response)
      create_convenient_marker(response)
      closeOverlay()
      make_overlay2(response[13].name, response[13].addr, response[13].price, response[13].spec, response[0][0], response[0][1])
    }
  },
  error: function(){
    alert('Error occurred while fetching data from the server');
    console.log(lat, lng)
  }
});

```

매물 리스트 클릭 시 실행되는
함수

```

function make_overlay(name, addr, lat, lng){
  // 마커 위에 커스텀오버레이를 표시합니다
  // 마커를 중심으로 커스텀 오버레이를 표시하기 위해 css를 이용해 위치를 설정했습니다
  custom_overlay = new kakao.maps.CustomOverlay({
    map: map,
    position: new kakao.maps.LatLng(lat, lng) ,
    clickable: true // 이 설정으로 인해 마우스 클릭으로 인한 지도 이벤트 발생을 막을 수 있다.
  });

  // 일반 string type으로 html문 content 생성할 경우 click 이벤트 처리가 되지 않음.
  var content = document.createElement('div');
  content.style.cssText = 'background: white; position: absolute; left: 0; bottom: 40px; width: 286px; height: 80px; margin-left: -14';

  var title = document.createElement('div');
  title.innerHTML = name;
  title.style.cssText = 'padding: 5px 0 0 10px; height: 30px; background: #eee; border-bottom: 1px solid #ddd; font-size: 18px; font-';

  var info = document.createElement('div');
  info.style.cssText = 'width: 286px; height: 80px; border-radius: 5px; border-bottom: 2px solid #ccc; border-right: 1px solid #ccc;';

  var body = document.createElement('div');
  body.style.cssText = 'position: relative; overflow: hidden;';

  var closeBtn = document.createElement('button');
  closeBtn.onclick = function () {
    custom_overlay.setMap(null);
  };
  closeBtn.style.cssText = "position: absolute; top: 7px; right: 10px; color: #888; width: 20px; height: 20px; background: url('https://t1.daumcdn.net/mapdata/allcountry/icon/';

  var ellipsis = document.createElement('div');
  ellipsis.appendChild(document.createTextNode(addr));
  ellipsis.style.cssText = 'text-align: center; overflow: auto; text-overflow: ellipsis; white-space: nowrap; font-size: 13px;';

  // content가 가장 바깥 div 태그
  content.appendChild(info);
  info.appendChild(title);
  title.appendChild(closeBtn);
  info.appendChild(body);
}

```

생성된 마커 클릭 시 보여지는 오버레이
함수

III. 분석 및 결과

백엔드 주요 코드

편의시설 관련 마커와 오버레이 생성 시 필요한 데이터 전처리 시 사용된 코드

```
df["체육시설"] = df["체육시설"].apply(literal_eval)  
df["스타벅스"] = df["스타벅스"].apply(literal_eval)  
df["백화점/대형마트"] = df["백화점/대형마트"].apply(literal_eval)
```

Syntax를 구조화 하는 AST모듈에 존재 ←

String 타입이었던 JSON 데이터들을 JSON으로 변환

III. 분석 및 결과



프로젝트 결과물에 대한 FEED BACK

주택 등 부동산 매매의 경우 ML(머신러닝)을 학습하여 예측하는 방식보다는 주로 전통적 방식의 모델 학습을 통하여 가격을 예측하는 것이 더 정확한 예측값을 추정 할 수 있습니다. 이에, ML방식보다 전통적 방식으로 프로젝트를 진행하였고, 이에따라 과정 상 배운 ML 방식을 활용 할 수 없었음에 아쉬움이 있었습니다. 또한, Vue.js를 활용하여 웹페이지에 화면을 구현하고 싶었으나 Django와의 연동의 어려움을 느껴 구현할 수 없었기에 추후 보완하려 합니다.

프로젝트 평가 의견과 느낀점

수업시간 내 배운 것들을 통해 다양한 기술을 습득하였고, 프로젝트 진행으로 학습한 것들을 다시 한번 복습하고, 발전시켰습니다. 더 나아가 커리큘럼 상, 배우지 못했던 기술들도 찾아 익힐 수 있었습니다. 그러나, 원했지만 구현 하지 못하였던 기술들이 있었기에 처음에 구상 했던 것보다는 완성도가 부족하게 되었습니다. 이를 보완하기 위하여, 프로젝트가 종료된 이후에도 지속적으로 학습하여, 차후에 원하던 기술까지 구현하여 보완하고자 합니다.



이고은,최열, “도시공원 유형 및 입지적 특성이 공동주택가격에 미치는 영향”,
929page

서울시 부동산 실거래가 정보(서울 열린데이터 광장) “<https://data.seoul.go.kr/>”

-2018 ~2022 년 서울시 아파트 매매가 데이터

한강공원 주차장 정보 (서울 열린데이터 광장) “<https://data.seoul.go.kr/>”

- 서울시 한강 공원 위치를 파악5. 소비자물가지수, 주택매매가격지수, 기준금리
(KOSIS)

매물 데이터 (네이버 부동산 크롤링 2023.06.14 기준)

“<https://land.naver.com/>”

- 예측 결과에 맞는 아파트 매물 추천을 위한 데이터 수집

KAKAOMAP API “<https://apis.map.kakao.com/>”

