

# 서울특별시

## 외부 요인에 따른 매매가 비교

# 목차 페이지

---

01

프로젝트 개요

02

프로젝트 팀 구성 및 역할

03

프로젝트 수행 절차 및 방법

04

프로젝트 수행 과정

05

프로젝트 수행 결과



# 01

## 프로젝트 개요

# 프로젝트 개요

## Keyword 01

서울 부동산에 영향을 끼치는 요인은 무엇이 있을까?

## Keyword 02

부동산에 영향을 미치는 요인에 대한 가설 설정

GOAL

## Keyword 03

활용 자료 및 장비 (PYTHON, VSCODE, 서울시 열린데이터 광장, 공공데이터 등)

## Keyword 04

서울 부동산에 영향을 미치는 요인 결과

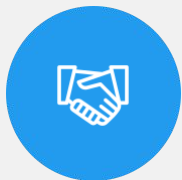


# 02

## 프로젝트 팀 구성 및 역할

# 프로젝트 팀 구성 및 역할

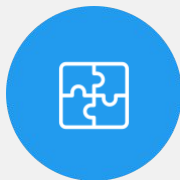
임병남



팀장

공공데이터 수집 및 분석  
데이터 전처리 및 가공  
상관관계 분석 및 인사이트 도출  
발표자

이원빈



팀원

공공데이터 수집 및 분석  
데이터 전처리 및 가공  
상관관계 분석 및 인사이트 도출  
발표자료 제작

우병준



팀원

공공데이터 수집 및 분석  
데이터 전처리 및 가공  
상관관계 분석 및 인사이트 도출  
발표자료 제작



# 03

## 프로젝트 수행 절차 및 방 법

# 프로젝트 수행 절차 및 방법

구분	기간
프로젝트 기획 분석 내용 가설계	10/31(목) - 11/1(금)
데이터 수집 및 가공	11/1(금) - 11/5(화)
데이터 분석 및 전처리	11/3(일) - 11/6(수)
인사이트 도출	11/5(화) - 11/7(수)
산출물 정리, 발표 자료 제작	11/7(목) - 11/8(금)

01

프로젝트 기획 및 주제선정  
계획서 작성  
데이터 수집 절차 설계

02

외부 데이터 수집  
데이터 가공  
서울시 공공데이터, 국토교통부 데이터 참조

03

데이터 정제 및 정규화



# 프로젝트 수행 절차 및 방법

구분	기간
프로젝트 기획 분석 내용 가설계	10/31(목) - 11/1(금)
데이터 수집 및 가공	11/1(금) - 11/5(화)
데이터 분석 및 전처리	11/3(일) - 11/6(수)
인사이트 도출	11/5(화) - 11/7(수)
산출물 정리, 발표 자료 제작	11/7(목) - 11/8(금)

04

상관관계 분석 및  
인사이트 도출

05

발표자료 제작  
프로젝트 성찰지 작성



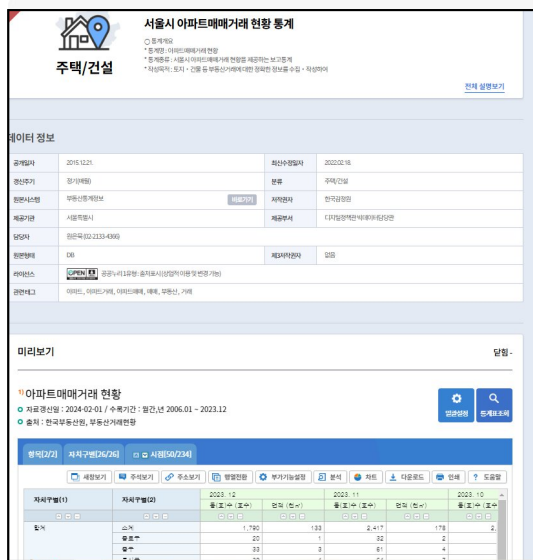
# 04

## 프로젝트 수행 과정

## 데이터 원천 파악 및 데이터 수집 작업

서울시 공공데이터에서  
서울시 아파트 매매거래 현황 통계 자료를  
csv파일로 수집

<https://data.seoul.go.kr/dataList/datasetTotalList.do>  
<https://www.data.go.kr/data/15047124/fileData.do>



# 프로젝트 수행 과정

## 데이터 전처리

```
import pandas as pd
data = pd.read_csv('...')
data.head(3)
```

✓ 0.2s

점수	자치구	자치구명	점수연도	자치구명	물건금액(만원)	건물면적(m²)
1	2022	관진구	105000	75.03		
15	2022	종로구	40200	59.40		
16	2022	성동구	110000	25.33		
38	2022	서대문구	53459	54.02		
62	2022	성북구	58100	114.66		
...	...	...	...	...		
66307	2022	송파구	95547	59.96		
66308	2022	구로구	37000	83.02		
66319	2022	서초구	155000	101.32		
66320	2022	송파구	69905	59.96		
66326	2022	서초구	42300	52.47		

[13030 rows x 4 columns]  
 <class 'pandas.core.frame.DataFrame'  
 Index: 13030 entries, 1 to 66326  
 Data columns (total 4 columns):  
 # Column Non-Null Count Dtype  
 ---  
 0 점수연도 13030 non-null int64  
 1 자치구명 13030 non-null object  
 2 물건금액(만원) 13030 non-null int64  
 3 건물면적(m²) 13030 non-null float64  
 dtypes: float64(1), int64(2), object(1)  
 memory usage: 509.0+ KB  
 ...  
 25% 2022.0 3.950000e+04 39.600000  
 50% 2022.0 7.500000e+04 59.970000  
 75% 2022.0 1.220000e+05 84.930000  
 max 2022.0 1.450000e+06 301.470000

3 rows x 21 columns

Output is truncated. View as a [scrollable element](#) or open in a [text editor](#). Adjust cell output settings...

info로 정보 확인 후  
 결측치와 데이터 타입을 확인  
 describe를 사용해서 정수형 데이터를 확인

# 프로젝트 수행 과정

## 데이터 정제 및 시각화

자치구명을 기준으로 건물 면적을 평 수로 변환

평 수를 기준으로 평당가격을 책정하여  
기준매매가를 설정시켜  
데이터를 정규화

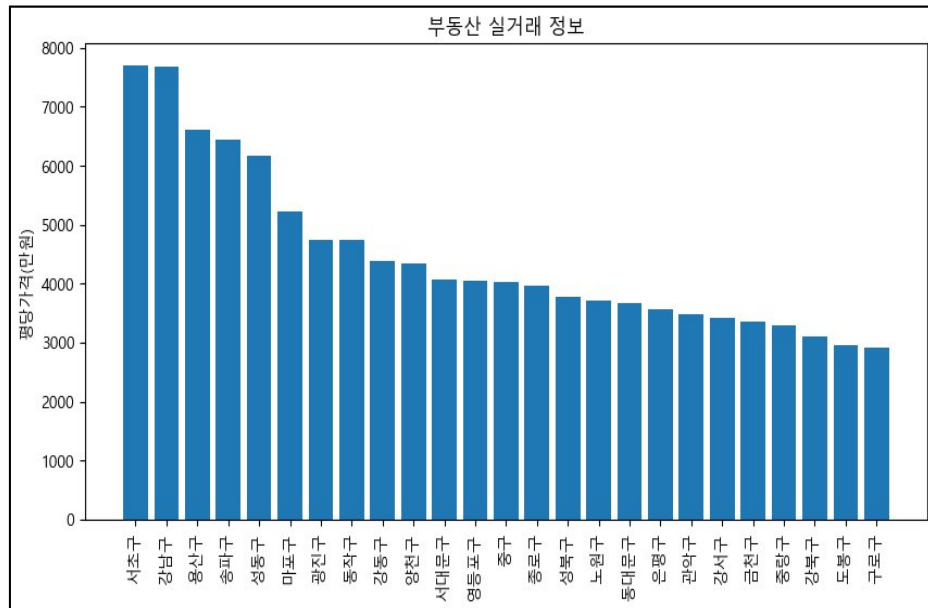
	접수연도	자치구명	물건금액(만원)	건물면적(㎡)	평 수	평당가격(만원)
1	2022	광진구	105000	75.03	22.736364	4618.152739
15	2022	종량구	40200	59.40	18.000000	2233.333333
16	2022	성동구	110000	25.33	7.675758	14330.833004
38	2022	서대문구	53459	54.02	16.369697	3265.729359
62	2022	성북구	58100	114.66	34.745455	1672.161172
...	...	...	...	...	...	...
66307	2022	송파구	95547	59.96	18.169697	5258.590727
66308	2022	구로구	37000	83.02	25.157576	1470.729945
66319	2022	서초구	155000	101.32	30.703030	5048.361627
66320	2022	송파구	69905	59.96	18.169697	3847.339893
66326	2022	서초구	42300	52.47	15.900000	2660.377358

13030 rows × 6 columns

# 프로젝트 수행 과정

## 데이터 정제 및 시각화

자치구명	접수연도	물건금액(만원)	건물면적(㎡)	평 수	평당가격(만원)
서초구	2022.0	213213.541322	87.757672	26.593234	7700.282084
강남구	2022.0	211000.905923	85.674925	25.962098	7674.664125
용산구	2022.0	215472.367647	101.916360	30.883746	6617.721205
송파구	2022.0	150778.963839	78.527636	23.796253	6450.090994
성동구	2022.0	150605.643045	78.924121	23.916400	6160.860748
마포구	2022.0	108522.447013	69.433815	21.040550	5218.657317
광진구	2022.0	105964.700704	72.570282	21.990994	4751.952731
동작구	2022.0	84774.690265	59.842513	18.134095	4743.194563
강동구	2022.0	86321.604938	62.290265	18.875838	4389.463866
양천구	2022.0	98368.518931	75.280290	22.812209	4342.837334
서대문구	2022.0	84251.285354	69.504848	21.062075	4079.923200
영등포구	2022.0	83602.259684	59.002123	17.879431	4049.405334
중구	2022.0	74822.233010	59.495485	18.028935	4034.691102
종로구	2022.0	69167.094444	53.598444	16.241953	3955.637398
성북구	2022.0	78207.743879	70.764068	21.443657	3781.384881
노원구	2022.0	65673.368243	60.033716	18.192035	3713.654381
동대문구	2022.0	69077.976827	60.381016	18.297278	3672.730888
은평구	2022.0	66815.177706	63.897641	19.362922	3555.679411
관악구	2022.0	63025.968109	60.343872	18.286022	3474.392775
강서구	2022.0	65851.178994	62.556893	18.956634	3424.268471
금천구	2022.0	46203.038278	48.155670	14.592627	3362.112205
중랑구	2022.0	49640.805604	51.036130	15.465494	3300.624124
강북구	2022.0	50716.060284	56.046277	16.983720	3105.625411
도봉구	2022.0	56424.204276	65.445891	19.832088	2953.911328
구로구	2022.0	51004.389027	55.109065	16.699717	2919.360805



# 프로젝트 수행 과정

자치구명	범죄 발생 건수	대비율	평당가격	월_평균_소득_금액	생활용품_지출_총_금액	교육_지출_총_금액	인구밀도 (명/㎡)	인구 (명)	버스승차객수	지하철승차객수	마트개수	고등학교개수
0	종로구	1.046526	1.198678e+07	3725859.0	1.607569e+09	4.017572e+09	14543.222222	16912.333333	66224995	77401702	16	14
1	중구	0.690003	1.222634e+07	3513603.0	7.613716e+09	2.627210e+10	18616.750000	16348.125000	45503810	78865473	57	11
2	용산구	0.705426	2.005370e+07	3971498.0	1.568267e+10	1.311866e+09	15373.235294	27445.176471	50338521	20010296	22	10
3	성동구	0.727894	1.866927e+07	3734499.0	1.614918e+09	2.827612e+09	22131.277778	32026.000000	35118053	45692050	14	7
4	광진구	0.690246	1.439986e+07	3192971.0	1.937717e+09	2.971485e+09	24183.187500	43906.500000	31840472	23353031	26	9
5	동대문구	0.685828	1.112949e+07	3030256.0	1.734690e+09	2.504125e+09	25205.266667	47146.800000	62655773	17992368	28	11
6	중랑구	0.742151	1.000189e+07	2598785.0	1.640601e+09	2.547978e+09	23818.117647	45898.823529	49759622	17943368	13	10
7	성북구	0.773008	1.145874e+07	3281773.0	1.168155e+09	3.608064e+09	24818.857143	42093.714286	75959717	12954299	30	13
8	강북구	0.827331	9.410986e+06	2547460.0	1.183999e+09	1.432918e+09	23025.785714	42528.857143	58432859	9001341	21	7
9	도봉구	0.732835	8.951246e+06	2785111.0	1.015555e+09	2.049020e+09	25696.400000	41865.200000	38187812	23775519	15	9
10	노원구	0.677875	1.125350e+07	3195375.0	1.676249e+09	7.833240e+09	22216.050000	50801.400000	51409495	52859796	19	25
11	은평구	0.762260	1.077479e+07	2912739.0	1.930021e+09	4.210251e+09	23585.294118	55364.941176	62673190	50654457	19	18
12	서대문구	0.676074	1.236340e+07	3417706.0	8.783018e+08	3.158566e+09	22371.666667	42607.200000	73712784	9466208	18	7
13	마포구	0.687012	1.581411e+07	3832436.0	2.815424e+09	8.303955e+09	23221.411765	44186.470588	63786743	92408668	27	9

## 모든 데이터 병합

자치구명 = 공통컬럼

검거율 평당가격 월평균소득

교육지출액 인구밀도

인구 버스승차객수 지하철승차객수

대형마트개수 고등학교개수

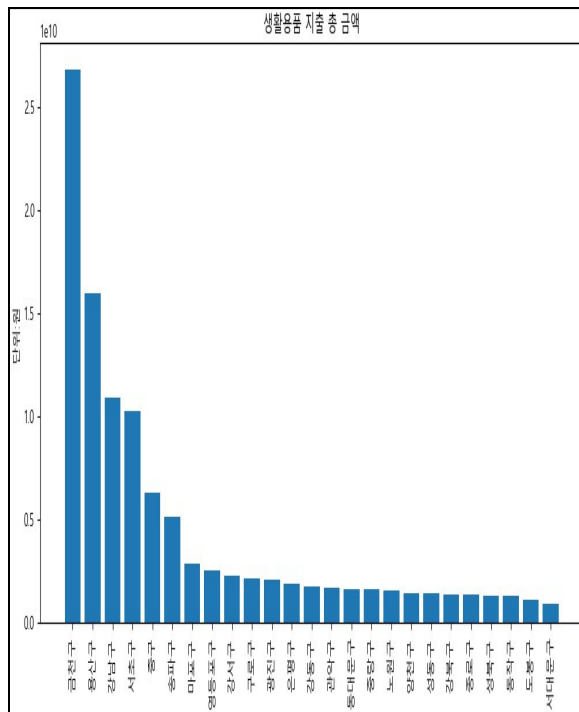
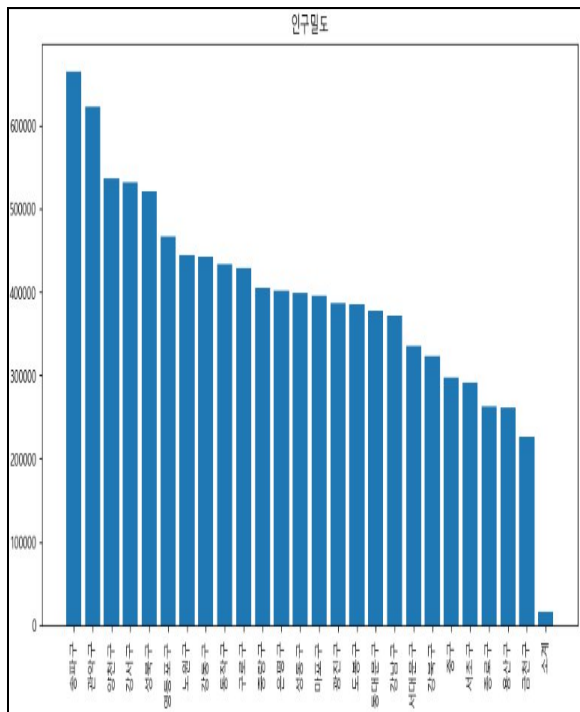
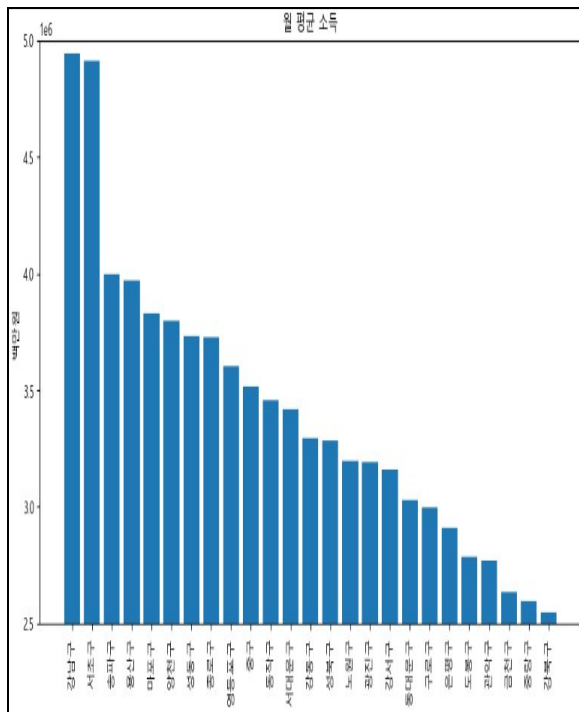
컬럼을 모두 병합





# 프로젝트 수행 과정

## 가설 확인 시각화





# 05

## 프로젝트 수행 결과

# 프로젝트 수행 결과

가설



가설

서울에서 매매가가  
가장 높은 자치구는 어디일까?  
비싼 이유는 무엇일까?

&gt;

세부 가설



세부 가설

유동인구의 영향일까?  
대중교통 활성화의 영향일까?  
여가생활시설의 영향일까?  
교육시설의 영향일까?

&gt;

추론



추론

서초구와 강남구의  
월 수입이 높고, 인구밀도가 낮으며 생  
활용품 지출 금액이 높기때문에 매매  
가가 비싼가?

&gt;

결론



결론

서초, 강남구 월 수입과 생활용품지출  
금액이 높으나  
인구 밀도가 낮다는 점  
매매가에 영향을 미친다는 결론