IMEN574: Programming for Data Science

Hustar Homework #1 서울시 아파트 실거래 데이터 전처리

포항공과대학교 산업경영공학과

조현재

email: present@postech.ac.kr

Contents

- 1. Introduction
- 2. Data collection & preprocessing
- 3. Data description
- 4. Future works

Data collection

데이터 수집



- 네이버 부동산
 - 아파트 특성



- 국토교통부
 - 기간별 아파트 실거래가



Data description

데이터 분석: 네이버 부동산

```
> dim(apt_info)
[1] 43001 27
> apt_info %>% select(아파트.코드) %>% unique() %>% nrow()
[1] 8299
```

- 총 아파트 정보(개별 아파트, 전용면적): 43,001개
- 아파트 관련 변수 수: 27개
- 서울시 개별 아파트 단지 수: 8,299단지

〈변수 정보〉

```
> str(apt_info)
                                                                      <To-Do>
'data.frame': 43001 obs. of 27 variables:
               "서울특별시" "서울특별시" "서울특별시" "서울특별시" ...
          : chr "종로구" "종로구" "종로구" "종로구" ...
          : logi NA NA NA NA NA NA ...
              "청운벽산빌리지" "청운벽산빌리지" "청운벽산빌리지" "청운벽산빌리지" ...
$ 아파트.코드: int 12076 12076 12076 12076 12076 12076 12076 12076 12076 12076 12076 ...
        : num 82.9 83.6 137 141.4 140.5 ...
         : chr "계단식" "계단식" "계단식" "계단식" ...
          : int 3 3 4 4 4 4 5 5 5 5 ...
$ toilet
          : int 1122222222...
$ 세대.수
         : int 9999999999...
$ y_c
          : chr "벽산" "벽산" "벽산" "벽산" ...
$ park
          : int NA ...
$ per_park : num NA ...
$ heat mat : chr "도시가스" "도시가스" "도시가스" "도시가스" ...
         : int NA ...
         : int NA ...
$ f_high
          : int 3 3 3 3 3 3 3 3 3 3 ...
$ f_1ow
               "서울청운초등학교" "서울청운초등학교" "서울청운초등학교" "서울청운초등학교" ...
$ 도로명
         : chr "서울시 종로구 청운동 1" "서울시 종로구 청운동 1" "서울시 종로구 청운동 1" "서울시 종로구 청운동 1" ...
        : chr "서울시 종로구 자하문로36길 16-14" "서울시 종로구 자하문로36길 16-14" "서울시 종로구 자하문로36길 16-14"
종로구 자하문로36길 16-14" ...
$ 위도
          : num 37.6 37.6 37.6 37.6 37.6 ...
```

1. 데이터를 불러와 출력 경로: ./Dataset/서울특별시 unique apt list.csv

2. 아파트 내에서 unique한 아파트 단지(아파트 코드) 수를 확인할 것

: num 127 127 127 127 127 ...

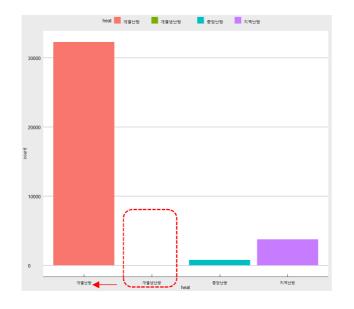
데이터 전처리: 네이버 부동산

이상치(Outlier) 확인

1. Numeric 변수 정보

>0 >0					⟨1500	⟨100		
> apt_info %>% select_if(negate(is.character)) %>% select(-0	파트.코드, -위도, -경도) %>% summary()	V 5	nank	per_park	용적률	거폐육	f high	f low
Mode:logical Min. : 0.00 Min. :0.000 Min. :0.0	00 Min. : 1.0 Min. : 1.000			Min. : 0.020	Min. : 2.0			Min. :-1.000
NA's:43001 1st Qu.: 59.22 1st Qu.:2.000 1st Qu.:1.0 Median : 80.26 Median :3.000 Median :2.0								
Mean : 83.38 Mean :2.845 Mean :1.6	34 Mean : 244.9 Mean : 3.672	Mean :2003	Mean : 299.8	Mean : 1.193	Mean : 324.3	Mean : 40.99	Mean :12.02	Mean : 8.719
3rd Qu.: 97.06	00 3rd Qu.: 208.0 3rd Qu.: 3.000 00 Max. :9510.0 Max. :124.000	Max. :2019	Max. :12456.0	Max. :11.950		Max. :2457.00		

Character 변수 정보



Numeric 변수 전처리

- 전용면적, 방 수가 0인 데이터 제거
- 용적률과 건폐율 수치가 비이상적인 데이터 제거

Character 변수 전처리

• 개별냉난방은 개별난방으로 변경 (의미 동일)

<To-Do>

1. 위 기준을 적용하여 이상치 데이터를 제거/수정할 것



Data collection

데이터 수집: 국토교통부 실거래가

▼ 조건별 검색 국토교통부 실거래가 공개시스템을 이용하시면 쉽고 편리하게 이용하실 수 있습니다.
<조건별 자료제공 이용시 유의사항>
□ 본 서비스에서 제공하는 정보는 법적인 효력이 없으므로 참고용으로만 활용하시기 바라며, 외부 공개시에는 반드시 신고일 기준으로 집계되는 공식통계를 이용하여 주시기 바랍니다. □ 신고정보가 실시간 변경, 해제되어 제공시점에 따라 공개건수 및 내용이 상이할 수 있는 점 참고하시기 바랍니다. □ 본 자료는 계약일 기준입니다. (※ 7월 계약, 8월 신고건 → 7월 거래건으로 제공) □ 시도별 자료제공 계약일자 범위를 최대 1년으로 개선하였으나 이용에 참고하시기 바랍니다.
● 계약일자 2014-01-01 ● ~ 2019-12-31 ● 파일구분 EXCEL ▼ ● 실거래가구분 아파트(매매) ▼ ● 주소구분 ●지번주소 ●도로명주소 ● 시도 서울특별시 ▼ ● 유소구분 ● 지번주소 ● 조례▼ ● 면적 전체 ▼ ● 금액선택 (만원) ~ (만원)

〈변수 정보〉

> str(trading_all)

<To-Do>

1. 국토교통부 실거래가 데이터를 불러와 데이터 개관을 확인 경로: ./Dataset/서울특별시 매매가 info.csv



데이터 병합: 국토교통부 실거래가

- 데이터 병합 -> join key가 필요
- Join key 후보
 - 1. 아파트 이름+전용면적 크기
 - ① 다른 지역의 동일한 아파트 이름인 경우 (코드번호 다름)

→ 사용하기 어려움

2. <u>주소+전용면적 크기</u>

<To-Do>

1. 주소 + 전용면적을 합친 데이터를 기준으로 국토교통부 데이터와 네이버 부동산 데이터를 병합할 것



데이터 병합: 국토교통부 실거래가

<To-Do>

1. 방 개수, 화장실 개수에 대해서 NA값에 대해 방개수는 평균값으로, 화장실 개수는 최빈값(mode)으로 Impute할 것



Data Description

최종 데이터셋

I. 데이터 크기

> dim(final_info)
[1] 454709 39

- 원본 거래 데이터 수: 569,331개
- 병합 후 데이터 수: 454,709개 (Loss 20.1%)

II. 데이터 요약

Min. : 2014 Min. : 1.000 Length:454709 Length:454709 Mode:logical Length:454709 Length:454709 Min. : 3 lst Qu.: 2015 1st Qu.: 4.000 Class : character Mode : character Class : character Mode : character Mode : character Mode : character Mode : character Class : character Mode : character Mode : character Mode : character Mode : character Class : character	> summary(final,	_info)							
1st Qu.: 2015 1st Qu.: 4.00 Class : character Mode : characte	YYYY	MM	도	지역	구	법정동	단지명	아파트.코드	
Mean : 2016 Mean : 7.000 Mode : character Mode : character Mode : character Median : 3426 Mean : 2063 3rd Qu.: 2018 3rd Qu.: 9,000 Max. : 12019 Max. : 12.000	Min. :2014	Min. : 1.000	Length:454709	Length:454709	Mode:logical	Length:454709	Length:45470)9 Min. :	3
Mean : 2016 Mean : 6.515 3rd Qu.: 2018 3rd Qu.: 2019 Max. : 12.000 Max. : 12.001 Max. : 12.001 Max. : 12.015 Max.	1st Qu.:2015	1st Qu.: 4.000	Class :character	Class :character	NA's:454709	Class :characte	r Class:chara	acter 1st Qu.:	842
3rd Qu.: 2018	Median :2016	Median : 7.000	Mode :character	Mode :character		Mode :characte	r Mode :chara	acter Median :	3426
Max. : 12.000 Max. : 1.000 Min. : 1.0	Mean :2016	Mean : 6.515						Mean : 2	20603
area room toilet 현관구조 세대.수 dong birth_year con_n park Min. :10.32 Min. :1.000 Min. :1.000 Length:454709 Min. :1 1st Qu.: 59.87 1st Qu.: 3.000 1st Qu.: 1.000 Class :character 1st Qu.: 261 1st Qu.: 3.00 Hedian :2.001 Mode :character Median : 7.00 Median :2.001 Mode :character Median : 64.2 Median : 7.00 Median :2.001 Mode :character Median : 67.0 Median :2.001 Max. :2.001 Max. :2.001 Max. :2.001 Max. :2.001 Max. :2.019 Max.	3rd Qu.:2018	3rd Qu.: 9.000						3rd Qu.: 2	22570
Min. : 10.32 Min. : 11.000 Min. : 11.000 Length:454709 Min. : 1 Min. : 1.00 Min. : 11.970 Length:454709 Min. : 1 St Qu.: 1.95.87 Ist Qu.: 1.000 Length:454709 Mode : character Median : 3.000 Median : 2.000 Mode character Median : 67 Mean : 79.70 Mean : 2.966 Mean : 1.659 Mean : 1.018 Mean : 11.77 Mean : 2.001 Mode : character Median : 67 Mean : 2.966 Mean : 1.659 Mean : 1.018 Mean : 11.77 Mean : 2.001 Mode : character Median : 67 Mean : 2.966 Mean : 1.659 Mean : 1.018 Mean : 11.77 Mean : 2.001 Mode : character Median : 67 Mean : 2.966 Mean : 1.150 Mean : 1.150 Mean : 2.966 Mean : 1.150 Mean : 1.1	Max. :2019	Max. :12.000						Max. :12	6215
Min. : 10.32 Min. : 11.000 Min. : 11.000 Length:454709 Min. : 1 Min. : 1.00 Min. : 11.970 Length:454709 Min. : 1 St Qu.: 1.95.87 Ist Qu.: 1.000 Length:454709 Mode : character Median : 3.000 Median : 2.000 Mode character Median : 67 Mean : 79.70 Mean : 2.966 Mean : 1.659 Mean : 1.018 Mean : 11.77 Mean : 2.001 Mode : character Median : 67 Mean : 2.966 Mean : 1.659 Mean : 1.018 Mean : 11.77 Mean : 2.001 Mode : character Median : 67 Mean : 2.966 Mean : 1.659 Mean : 1.018 Mean : 11.77 Mean : 2.001 Mode : character Median : 67 Mean : 2.966 Mean : 1.150 Mean : 1.150 Mean : 2.966 Mean : 1.150 Mean : 1.1									
1st Qu.; 59.87	area	room	toilet	현관구조	세대.수	dong l	oirth_year	con_n	park
Median : 84.57 Median : 3.000 Median : 2.000 Mode : character Median : 642 Median : 7.00 Median : 2001 Mode : character Median : 67 Mean : 79.70 Mean : 2.966 Mean : 1.659 Mean : 1.108 Mean : 11.77 Mean : 2001 Mode : character Median : 67 Mean : 11.77 Mean : 2001 Max. : 37d Qu. : 2007 Mean : 124.00 Max. : 37d Qu. : 124.00 Max. : 37d Qu. : 124.00 Max.	Min. : 10.32	Min. :1.000	Min. :1.000 l	ength:454709	Min. : 1	Min. : 1.00	Min. :1970 L	.ength:454709	Min. : 1
Mean : 79,70 Mean : 2,966 Mean : 1.659 Mean : 1018 Mean : 11.77 Mean : 2001 Mean : 115 3rd Qu.: 2000 Max. : 124.00 Max. : 124.00 Max. : 2019 Mean : 115 Max. : 124.02 Mean : 124.00 Mex. : 2019 Mex. : 124.00 Mex. : 124.00 Mex. : 2019 Mex. : 124.00 Mex. : 124.01 Mex. :	1st Qu.: 59.87	1st Qu.:3.000	1st Qu.:1.000 (lass :character	1st Qu.: 261	1st Qu.: 3.00	1st Qu.:1996 C	lass :character	1st Qu.: 278
3rd Qu.: 84.97 3rd Qu.: 3.000 3rd Qu.: 2.000 3rd Qu.: 2.000 3rd Qu.: 15.00 3rd Qu.: 2007 3rd Qu.: 141 Max. : 424.32 Max. : 8.000 Max. : 5.000 Max. : 5.000 Max. : 9510 Max. : 124.00 Max. : 2019 Max. : 1245 Max. : 124.72 Max. : 2124.00 Max. : 2019 Max. : 124.72 Max. : 124.72 Max. : 124.00 Max. : 2019 Max. : 124.72 Max. : 124.72 Max. : 124.00 Max. : 2019 Max. : 124.72 Max. : 124.72 Max. : 124.00 Max. : 2019 Max. : 124.72 Max. : 124.72 Max. : 124.00 Max. : 2019 Max. : 124.72 Max. : 124.00 Max. : 124.00 Max. : 1200 Min. : 1.00 Length: 454709 Min. : 3.15 Qu.: 10.00 Min. : 1.00 Length: 454709 Min. : 1.00 Min. : 1.00 Length: 454709 Min. : 1.00 Median : 1.14 Max. : 11.130 Mode : character Median : 257 Median : 257 Median : 121.00 Median : 18.00 Median : 12.00 Mode : character Median : 3 Mean : 24.68 Mean : 19.11 Mean : 12.13 Mean : 1	Median : 84.57	Median :3.000	Median :2.000 M	lode :character	Median : 642	Median : 7.00	Median :2001 M	lode :character	Median : 674
Max. : :424.32 Max. : :8.000 Max. : :5.000 Max. : :5.000 Max. : :5.000 Max. : :124.00 Max. : :2019 Max. : :1245 Max. : :124.00 Max. : :2019 Max. : :1245 Max. : :124.00 Max. : :2019 Max. : :1372 Max. : :124.00 Max. : :2019 Max. : :1372 Max. : :1372 Max. : :124.00 Max. : :2019 Max. : :100 Length: :454709 Min. : :2 Min. : : :2.00 Min. : : :2.00 Min. : : :1.00 Length: :454709 Min. : :3 Min. : :2.00 Min. : : :2.00 Min. : : :1.00 Length: :454709 Min. : :3 Max. : :1.10 Mode : character Ist Qu.: : :25 Ist Qu.: :18.00 Ist Qu.: :15.00 Ist Qu.: : :1.00 Mode : character Median : :257 Median : :21.00 Median : :12.00 Mode : character Median : :3 Mean : :287 Mean : :24.68 Mean : :19.11 Mean : :12.13 Mean : :3 Max. : :1477 Max. : :96.00 Max. : :69.00 Max. : :54.00 Max. : :3 Max. : :3 Max. : :1477 Max. : :96.00 Max. : :69.00 Max. : :54.00 Max. : :3 Max. : :3 Max. : :3 Max. : :3 Max. : :4.00 Min. : : :4.000 Min. : : :4.000 Min. : : :5.00 Min. : : :55.74 Min. : :6.78 Max. : :3 Max. : :17.0 Mode : character Median : :4.000 Min. : :0.00 Min. : :2.646 Min. : :55.74 Min. : :6.78 Mean : :127.0 Mode : character Median : :4.000 Min. : :0.00 Median : :596.144 Median : :970.38 Median : 322.94 Mean : :17.0 Max. : :17.0 Mean : :17.0 Mean : :57.32 Mean : :57.32 Mean : :77.75 Mean : :17.0 Mean : :77.75 Mean : :17.0 Mean : :77.75 Mean : :17.0 Mean : :77.0 Mean : :77.	Mean : 79.70	Mean :2.966	Mean :1.659		Mean :1018	Mean : 11.77	Mean :2001		Mean : 1154
per_park		3rd Qu.:3.000	3rd Qu.:2.000		3rd Qu.:1335		3rd Qu.:2007		3rd Qu.: 1415
per_park	Max. :424.32	Max. :8.000	Max. :5.000		Max. :9510	Max. :124.00	Max. :2019		Max. :12456
Min. : 0.020 Length:454709 Length:454709 Min. : 2 Min. : 2.00 Min. : 2.00 Min. : 1.00 Length:454709 Min. : 3 1st Qu.: 0.970 (lass :character Median : 1.130 Mode :character Median : 1.130 Mode :character Median : 1.130 Mode :character Median : 1.144 Mode : 1.1950 Max. : 11.950 Max. : 11.950 Max. : 13721 전도 지원주소 Price Min. : 126.8 Length:454709 Min. : 700 Max. : 1477 Max. : 96.00 Max. : 169.00 Max. : 54.00 Max. : 55.74 Min. : 6.78 Median : 127.0 Mode :character Median : 8.00 Median : 15.00 Max. : 155.74 Min. : 6.78 Median : 127.0 Mode :character Median : 9.418 Mean : 1532 Mean : 770.75 Mean : 1572.7 Mean : 1580.7 Mean : 1580.									
1st Qu.: 0.970 Class :character Median : 1.130 Mode :character Median : 1.130 Mode :character Median : 1.144 Mode : 1.130 Mode :character Median : 1.144 Mode : 1.130 Mode :character Median : 257 Median : 21.00 Median : 12.00 Median : 12.00 Mode : 12.0	per_park	heat	heat_mat	용적률	건폐율	f_high	f_low	초등학교	위도
Median : 1.130 Mode :character Median : 257 Median : 21.00 Median : 12.00 Mode :character Median : 3 Mean : 1.144 Mean : 1.145 Mean : 1.145 Mean : 287 Mean : 24.68 Mean : 19.11 Mean : 12.13 Mean : 12.13 Mean : 3 3rd Qu.: 12.90 Max. : 11.950 Max. : 11.950 Max. : 13721 전도 기반주소 Price floor age_C dis_sub dis_park dis_ele Min. : 126.8 Length:454709 Min. : 700 Min. : -4.000 Min. : 0.00 Min. : 2.646 Min. : 55.74 Min. : 6.78 1st Qu.: 126.90 Mode :character Median : 45900 Median : 8.000 Median : 15.00 Median : 596.144 Median : 970.38 Median : 322.94 Mean : 127.0 Mode :character Median : 45900 Median : 8.000 Median : 15.00 Median : 596.144 Median : 970.38 Median : 322.94 Mean : 127.2 Max. : 127.2 Max. : 70000 Max. : 69.000 Max. : 49.00 Max. : 49.00 Max. : 5583.588 Max. : 3268.24 Max. : 1810.04 Min. : 2.587 Min. : 24.62 Min. : 25.87 Min. : 24.62 Min. : 25.14 Min. : 35.34 Min. : 13.49 Median : 436.379 Median : 505.46 Median : 157.49 Mean : 1634.26 Median : 1839.61 Mean : 470.248 Mean : 575.54 Mean : 1580.67 Mean : 1843.09 Mean : 1931.14 Mean : 1931.14 Median : 470.248 Mean : 575.54 Mean : 1580.67 Mean : 1843.09 Mean : 1931.14 Mean : 147.00 Median : 12.00 Max. : 69.00 Max. : 69.00 Max. : 1849.60 Median : 1839.61 Mean : 470.248 Mean : 575.54 Mean : 1586.67 Mean : 1843.09 Mean : 1931.14 Mean : 1470.04 Median : 1470.00 Min. : 2487.65 Mean : 1843.09 Mean : 1931.14 Mean : 1470.00 Median : 120.00 Median : 1840.00 Median : 1839.61 Mean : 470.248 Mean : 575.54 Mean : 1586.67 Mean : 1843.09 Mean : 1931.14 Mean : 470.248 Mean : 575.54 Mean : 1586.67 Mean : 1843.09 Mean : 1931.14 Mean : 1470.00 Median : 120.00	Min. : 0.020	Length:454709	Length:454709	Min. : 2	Min. : 2.0	0 Min. : 2.00	Min. : 1.00	Length:454709	Min. :37.4
Mean : 1.144 3rd Qu.: 1.290 Max. :11.950 Max. :11.950 Na's :13721 정도 지번주소 price floor age_c dis_sub dis_park Median : 126.8 Length:454709 Min. : 700 Median : 127.0 Median : 127.0 Median : 127.0 Median : 127.0 Mean : 287 Mean : 24,62 Mean : 287 Mean : 19.11 Mean : 12.13 Mean : 3 Ard Qu.: 28.00 3rd Qu.: 15.00 3rd Qu.: 15.00 Max. : 13721 Max. : 1477 Max. : 96.00 Max. : 69.00 Max. : 54.00 Max. : 54.00 Max. : 13721 Mexim Mexim : 127.0 Min. : -4.000 Min. : -0.00 Min. : 0.00 Min. : 2.646 Min. : 55.74 Min. : 6.78 1st Qu.: 126.8 Length: 454709 Min. : 40.00 Min. : -4.000 Min. : 0.00 Min. : 2.646 Min. : 55.74 Min. : 6.78 1st Qu.: 127.0 Median : 127.0 Median : 57329 Mean : 9.418 Mean : 15.32 Mean : 772.755 Mean : 1035.23 Mean : 337.12 Max. : 127.2 Max. : 700000 Max. : 69.000 Max. : 49.00 Max. : 5583.588 Max. : 3268.24 Max. : 1810.04 dis_mid dis_high dis_univ dis_muse dis_Gu Min. : 2.587 Min. : 24.62 Min. : 25.14 Min. : 35.34 Min. : 13.49 1st Qu.: 289.105 1st Qu.: 329.44 1st Qu.: 967.51 Ist Qu.: 1037.49 1st Qu.: 1203.52 Median : 436.379 Median : 505.46 Median : 1574.95 Median : 1634.26 Median : 1839.61 Mean : 470.248 Mean : 575.54 Mean : 1860.67 Mean : 1843.09 Mean : 1931.14 3rd Qu.: 614.382 3rd Qu.: 762.57 3rd Qu.: 12499.60 3rd Qu.: 2487.65 3rd Qu.: 2187.106	1st Qu.: 0.970	Class :charact	er Class:charact	ter 1st Qu.: 225	1st Qu.:18.0	0 1st Qu.:15.00	1st Qu.: 8.00	Class :characte	er 1st Qu.:37.5
3rd Qu.: 1.290 Max. :11.950 Max. :14.77 Max. :96.00 Max. :69.00 Max. :54.00 Max. :54.00 Max. :3 NA's :13721 전도 지번주소 price floor age_c dis_sub dis_park dis_ele Min. :126.8 Length:454709 Min. : 700 Min. :-4.000 Min. : 0.00 Min. : 2.646 Min. : 55.74 Min. : 6.78 1st Qu.:126.9 Class :character 1st Qu.: 34000 1st Qu.: 5.000 1st Qu.: 391.007 1st Qu.: 634.04 1st Qu.: 214.92 Median :127.0 Mode :character Mean : 57329 Mean : 9.418 Mean : 15.32 Mean : 772.755 Mean : 1035.23 Mean : 337.12 3rd Qu.:127.1 3rd Qu.: 67000 Max. :69.000 Max. :49.00 Max. :5583.588 Max. :3268.24 Max. :1810.04 dis_mid dis_high dis_univ dis_muse dis_Gu Min. : 2.587 Min. : 24.62 Min. : 25.14 Min. : 35.34 Min. : 13.49 1st Qu.: 289.105 1st Qu.: 329.44 1st Qu.: 967.51 Ist Qu.: 1037.49 1st Qu.: 1203.52 Median : 436.379 Median : 505.46 Median : 155.54 Mean : 1634.26 Median : 1839.61 Mean : 470.248 Mean : 575.54 Mean : 1860.67 Mean : 1843.09 Mean : 1931.14 3rd Qu.: 614.382 3rd Qu.: 762.57 3rd Qu.: 2499.66 3rd Qu.: 2487.65 3rd Qu.: 2287.016	Median : 1.130	Mode :charact	er Mode :charact	ter Median : 257	Median :21.0	0 Median :18.00	Median :12.00	Mode :characte	er Median :37.5
Max. :1477 Max. :96.00 Max. :69.00 Max. :54.00 Max. :3 NA's :13721 장도 지번주소 price floor age_c dis_sub dis_park dis_ele Min. :126.8 Length:454709 Min. : 700 Min. :-4.000 Min. : 0.00 Min. : 2.646 Min. : 55.74 Min. : 6.78 1st Qu.:126.9 Class :character Median :127.0 Mode :character Median : 46900 Median : 8.000 Median : 15.00 Median : 596.144 Median : 970.38 Median : 322.94 Mean : 127.0 Mode :character Median : 67000 Median : 9.418 Mean : 15.32 Mean : 772.755 Mean : 1035.23 Mean : 337.12 Max. :127.2 Max. :700000 Max. :69.000 Max. :49.00 Max. :5583.588 Max. :3268.24 Max. :1810.04 dis_mid dis_high dis_univ dis_muse dis_Gu Min. : 2.587 Min. : 24.62 Min. : 25.14 Min. : 35.34 Min. : 13.49 1st Qu.: 289.105 1st Qu.: 329.44 1st Qu.: 296.51 1st Qu.: 1037.49 1st Qu.: 239.46 Median : 436.379 Median : 505.46 Median : 157.95 Median : 1634.26 Median : 1839.61 Mean : 470.248 Mean : 575.54 Mean : 1860.67 Mean : 1843.09 Mean : 1931.14 3rd Qu.: 614.382 3rd Qu.: 762.57 3rd Qu.: 72495.65 3rd Qu.: 21870.16							Mean :12.13		
NA's :13721 경도 지번주소 price floor age_c dis_sub dis_park dis_ele Min. :126.8 Length:454709 Min. : 700 Min. :-4.000 Min. : 0.00 Min. : 2.646 Min. : 55.74 Min. : 6.78 1st Qu.:126.9 Class :character 1st Qu.: 34000 1st Qu.: 5.000 1st Qu.: 10.00 1st Qu.: 391.007 1st Qu.: 634.04 1st Qu.: 214.92 Median :127.0 Mode :character Median : 46900 Median : 8.000 Median :15.00 Median : 596.144 Median : 970.38 Median : 322.94 Mean : 127.1	3rd Qu.: 1.290			3rd Qu.: 308	3rd Qu.:26.0	0 3rd Qu.:23.00	3rd Qu.:15.00		3rd Qu.:37.6
경도 지번주소 price floor age_C dis_sub dis_park dis_ele Min. :126.8 Length:454709 Min. : 700 Min. : -4.000 Min. : 0.00 Min. : 2.646 Min. : 55.74 Min. : 6.78 1st Qu.:126.9 Class :character Median :127.0 Mode :character Median : 127.0 Mode :character Mean : 127.1 Mean : 57329 Mean : 9.418 Mean : 15.32 Mean : 772.755 Max. :127.2 Max. : 70000 Max. :69.000 Max. :49.00 Median : 436.82 dis_mid dis_high dis_mir di				Max. :1477	Max. :96.0	0 Max. :69.00	Max. :54.00		Max. :37.6
Min. : 126.8 Length:454709 Min. : 700 Min. : -4.000 Min. : 0.00 Min. : 2.646 Min. : 55.74 Min. : 6.78 1st Qu.:126.9 Class :character									
1st Qu.:126.9 Class :character	경도	지번주소		floor	age_c		dis_park	dis_ele	
Median :127.0 Mode :character Median : 46900 Median : 8.000 Median : 15.00 Median : 596.144 Median : 970.38 Median : 322.94 Mean :127.0 Mean : 57329 Mean : 9.418 Mean : 15.32 Mean : 772.755 Mean : 1035.23 Mean : 337.12 3rd Qu.:127.1 3rd Qu.:67000 3rd Qu.:20.00 3rd Qu.:20.00 3rd Qu.: 946.882 3rd Qu.: 1357.78 3rd Qu.: 435.89 Max. :700000 Max. :700000 Max. :69.000 Max. :49.00 Max. :5583.588 Max. :3268.24 Max. :1810.04 dis_mid dis_high dis_muse dis_Gu Min. : 2.587 Min. : 24.62 Min. : 25.14 Min. : 35.34 Min. : 13.49 1st Qu.: 289.105 1st Qu.: 329.44 1st Qu.: 967.51 1st Qu.: 1037.49 1st Qu.: 1203.52 Median : 436.379 Median : 505.46 Median : 1574.95 Median : 1634.26 Median : 1839.61 Mean : 470.248 Mean : 575.54 Mean : 1260.67 Mean : 1237.95 Median : 1931.14 3rd Qu.: 62.57 3rd Qu.: 2499.60 3rd Qu.: 2487.65 3rd Qu.: 2510.16									
Mean :127.0 Mean : 57329 Mean : 9.418 Mean : 15.32 Mean : 772.755 Mean : 1035.23 Mean : 337.12 3rd Qu.:127.1 3rd Qu.: 67000 3rd Qu.: 13.000 3rd Qu.: 20.00 3rd Qu.: 946.882 3rd Qu.: 1357.78 3rd Qu.: 1435.89 Max. :127.2 Max. :700000 Max. :69.000 Max. :49.00 Max. :5583.588 Max. :3268.24 Max. :1810.04 dis_mid dis_mid dis_univ dis_muse dis_Gu Min. : 24.62 Min. : 25.14 Min. : 13.49 1st Qu.: 289.105 1st Qu.: 329.44 1st Qu.: 967.51 1st Qu.: 1037.49 1st Qu.: 1203.52 Median : 436.379 Median : 505.46 Median : 1534.26 Median : 1839.61 Mean : 470.248 Mean : 575.54 Mean : 1843.09 Mean : 1931.14 3rd Qu.: 614.382 3rd Qu.: 762.57 3rd Qu.: 2499.60 3rd Qu.: 2487.65 3rd Qu.: 2210.16		Class :characte		1st Qu.: 5.000	1st Qu.:10.00	1st Qu.: 391.00	7 1st Qu.: 634	1.04 1st Qu.: 21	4.92
3rd Qu.:127.1		Mode :characte							
Max. :127.2 Max. :700000 Max. :69.000 Max. :49.00 Max. :5583.588 Max. :3268.24 Max. :1810.04 dis_mid dis_high dis_univ dis_muse dis_Gu Min. : 2.587 Min. : 24.62 Min. : 25.14 Min. : 35.34 Min. : 13.49 1st Qu.: 289.105 1st Qu.: 329.44 1st Qu.: 967.51 Ist Qu.: 1037.49 1st Qu.: 1203.52 Median : 436.379 Median : 505.46 Median :1574.95 Median :1634.26 Median :1839.61 Mean : 470.248 Mean : 575.54 Mean :1860.67 Mean :1843.09 Mean :1931.14 3rd Qu.: 614.382 3rd Qu.: 762.57 3rd Qu.: 2499.60 3rd Qu.: 2487.65 3rd Qu.: 2510.16									
dis_mid dis_high dis_univ dis_muse dis_Gu Min. : 2.587 Min. : 24.62 Min. : 25.14 Min. : 35.34 Min. : 13.49 1st Qu.: 289.105 1st Qu.: 329.44 1st Qu.: 967.51 1st Qu.:1037.49 1st Qu.:1203.52 Median : 436.379 Median : 505.46 Median :1574.95 Median :1634.26 Median :1839.61 Mean : 470.248 Mean : 575.54 Mean :1860.67 Mean :1843.09 Mean :1931.14 3rd Qu.: 614.382 3rd Qu.: 762.57 3rd Qu.:2499.60 3rd Qu.:2487.65 3rd Qu.:2510.16									
Min. : 2.587 Min. : 24.62 Min. : 25.14 Min. : 35.34 Min. : 13.49 1st Qu.: 289.105 1st Qu.: 329.44 1st Qu.: 967.51 1st Qu.:1037.49 1st Qu.:1203.52 Median : 436.379 Median : 505.46 Median :1574.95 Median :1634.26 Median :1839.61 Mean : 470.248 Mean : 575.54 Mean :1860.67 Mean :1843.09 Mean :1931.14 3rd Qu.: 614.382 3rd Qu.: 762.57 3rd Qu.:2499.60 3rd Qu.:2487.65 3rd Qu.:2510.16	Max. :127.2		Max. :700000	Max. :69.000	Max. :49.00	Max. :5583.58	8 Max. :3268	3.24 Max. :181	.0.04
Min. : 2.587 Min. : 24.62 Min. : 25.14 Min. : 35.34 Min. : 13.49 1st Qu.: 289.105 1st Qu.: 329.44 1st Qu.: 967.51 1st Qu.:1037.49 1st Qu.:1203.52 Median : 436.379 Median : 505.46 Median :1574.95 Median :1634.26 Median :1839.61 Mean : 470.248 Mean : 575.54 Mean :1860.67 Mean :1843.09 Mean :1931.14 3rd Qu.: 614.382 3rd Qu.: 762.57 3rd Qu.:2499.60 3rd Qu.:2487.65 3rd Qu.:2510.16									
1st Qu.: 289.105									
Median : 436,379 Median : 505.46 Median :1574.95 Median :1634.26 Median :1839.61 Mean : 470.248 Mean : 575.54 Mean :1860.67 Mean :1843.09 Mean :1931.14 3rd Qu.: 614.382 3rd Qu.: 762.57 3rd Qu.:2499.60 3rd Qu.:2487.65 3rd Qu.:2510.16									
Mean : 470.248 Mean : 575.54 Mean :1860.67 Mean :1843.09 Mean :1931.14 3rd Qu.: 614.382 3rd Qu.: 762.57 3rd Qu.:2499.60 3rd Qu.:2487.65 3rd Qu.:2510.16									
3rd Qu.: 614.382 3rd Qu.: 762.57 3rd Qu.:2499.60 3rd Qu.:2487.65 3rd Qu.:2510.16									
Max. :2130.155 Max. :2837.36 Max. :7111.54 Max. :6839.08 Max. :6521.69									
	Max. :2130.19	55 Max. :2837	.36 Max. :7111.	54 Max. :6839	.08 Max. :6	521.69			

Ⅲ. 변수 정보

변수 특성	변수명
시계열	거래년월 (YYYY, MM)
주택특성	면적, 방/화장실 갯수, 현관구조
이웃특성	용적률, 건폐율, 최고층, 최저층, 세대 수, 단지 크기, (세대 당) 주차 대수, 난방 방식(개별/중앙), 연료
입지특성	법정동
종속변수	거래가격
보충 특성	도, 지역, 구, 단지명, 아파트 코드, 지번주소, 위경도, 아파트 준공년도, 건축사 명, 주변 초등학교

<To-Do>

1. 데이터 전처리 후 최종 생성된 데이터 출력

POHANG UNIVERSITY OF SCIENCE AND TECHNOLOGY

과제 제출

- 데이터 전처리에 사용한 코드파일(.ipynb 형식)
- 코드에 대한 설명(.ipynb 파일 내 주석을 이용하여 설명할 것)
- 데이터 전처리 후 생성된 최종 데이터 셋

위 결과들을 압축하여 #1_이름.zip 형식으로 LMS에 제출할 것 (과제 기한: 2020-09-08 23:59)



Thank You ©