

Find and Recommend Royal floor

팀 명 노유현

팀 원 노유현

지도교수 이슬

멘 토

개발 동기 및 목적

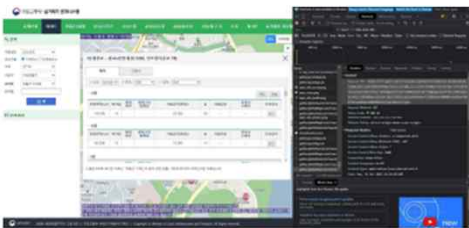
현재 사회는 아파트가 집으로서 굉장히 선호되는 사회이다. 그만큼 아파트의 가격도 많이 오르는 추세인데 아파트의 경우에는 빌라나 전원주택과는 다르게 층수가 존재한다. 층수에 따라서 아파트의 가격이 달라지게 되는데 여기서 우리는 흔히 로열층이란 곳을 찾을 수 있다. 실제로 아파트의 거래 내역을 살펴보고 데이터마이닝 기법을 통해서 아파트의 로열층 변화 내역을 살펴보고자 하고 더 나아가다면 아파트의 로열층을 예측하거나 로열층을 추천해 줄 수 있을 것이라 생각했다.

개발 내용

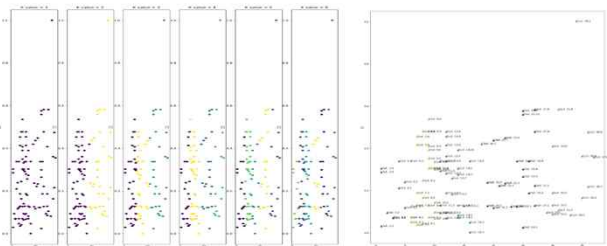
Process



먼저 국토 교통부 실거래가 공개시스템 웹사이트에서 찾고자 하는 아파트 단지의 거래내역을 크롤링을 했다. Json파일로 크롤링한 데이터를 가지고 먼저 필요 없는 내용들을 삭제 및 필요한 내용인 월, 층, 가격 정보만 빼내어서 따로 엑셀 파일로 정리 하였다. 이후 데이터 전처리를 통해서 가격에서 필요 없는, 를 없애고 가격의 크기가 너무 크기 Minmaxscaler method를 사용하여 크기를 작게 만들었다.



이후 k-means 클러스터링 기법을 통해서 데이터 마이닝을 적용하였다. 클러스터링은 k 값을 여러 번 바꾸어 돌려서 가장 좋은 값이 나오도록 하는 k 값을 찾도록 했다. 이후 클러스터링 한것을 시각화 시켜서 그래프로 표현했다.



다음으로는 SVD를 사용하여 또 다른 결과 값을 얻고자 했다. 결과 값을 기반으로 heatmap으로 표현했고 이때 나온 층 수를 기반으로 다른 층수를 추천해보고자 했다.

주요기술

1. Crawling

Postman, Python을 이용하여 국토 교통부의 아파트 실거래가 거래내역 공개시스템에서 데이터를 크롤링했다. Json 파일을 postman에서 param과 header를 알아내어서 Python에서 필요한 내용을 크롤링하여 엑셀파일로 저장시켰다.

2. K-means Clustering

클러스터링을 통해서 아파트의 층 별 거래가로 군집화 시켜서 로열층을 찾고자했다.

3. SVD 특이값 분해

특이값 분해(Singular Value Decomposition, SVD)란 행렬분해(Matrix Factorization) 방식 중 하나이다. 이 방법은 $m \times n$ 행렬의 고유값과 고유벡터를 이용한 고유분해(eigen decomposition)의 일반화 된 방식이다.

결과 및 분석

Clustering과 SVD를 진행한 결과 값을 모두 얻을 수 있었다. K-means Clustering의 경우에는 그래프에서 확인 할 수 있듯이 비슷한 층 수 별로 군집화가 이루어진 것을 확인 할 수 있다. 처음 예상하기로는 비슷한 가격대의 군집화가 이루어져서 저층에서의 인기 층수 그리고 고층에서의 인기 층수를 확인할 수있을 것이라 예상했지만 실제 군집화가 된 것은 비슷한 층수로 묶인 것을 확인 했다. 이때 중심이 어디인지 그래프에 표현하지 않았지만 비슷한 층수에서도 중심을 기준으로 군집화 된 것이라고 생각 할 수 있을 것 같다. 이를 통해서 나누어진 그룹들 사이에서의 로열층을 찾아 볼 수 있다고 생각한다. cluster_id가 2인 것들을 묶어서 표로 나타낸 것인데 중심이 되는 층수는 11,12 층인 것을 확인 할 수 있다. 다음으로는 SVD를 이용하여 만든 층수 추천 알고리즘이다. Data set을 가지고 만든 pivot table로 SVD를 진행한 뒤 만들어진 matrix를 heatmap으로 표현 했을 때 heatmap이 대칭적으로 만들어 졌다는 것을 확인 할 수 있었다

층수	가	가	가
1	10	14	112000
2	9	16	108000
13	8	12	120000
14	8	11	108000
15	8	12	118000
21	9	12	108000
26	7	13	101000
30	7	12	105000
34	6	11	104000
35	6	12	114000
39	3	12	110000
45	3	14	102000
48	2	11	90000
49	2	16	115000
50	2	14	108000
54	2	14	80000
66	1	11	88000
78	1	12	87000
83	1	14	88000
88	1	13	84000
90	1	16	85000
93	1	14	85000
94	1	16	79000
96	1	18	89700
100	1	12	84000

