

# 데이터분석\_프로젝트

---

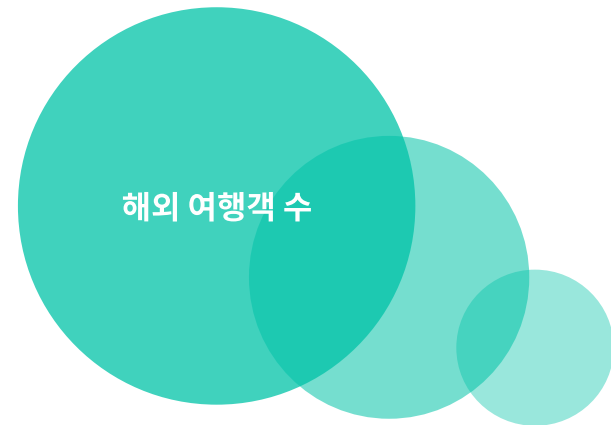
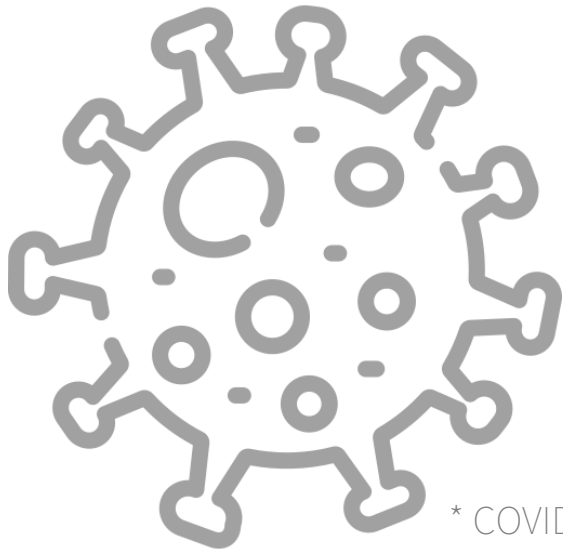
정효인

# EDA (Exploratory data analysis)

---

: 데이터를 분석하기 전에 그래프나 통계적인 방법으로  
자료를 직관적으로 바라보는 과정

# 끝나지 않는 \*코로나 바이러스의 확산



전년 동월대비 해외여행객  
93.9% 감소(3월 기준)

이에 따라, 갈 곳을 잃은  
여행객들

MISSION

**올여름, 코로나로 갈 곳을 잃은 여행객들을 위해  
휴가지 선정에 도움 가치 있는 데이터 분석**

# 플랫폼 선정

| Hotels.com

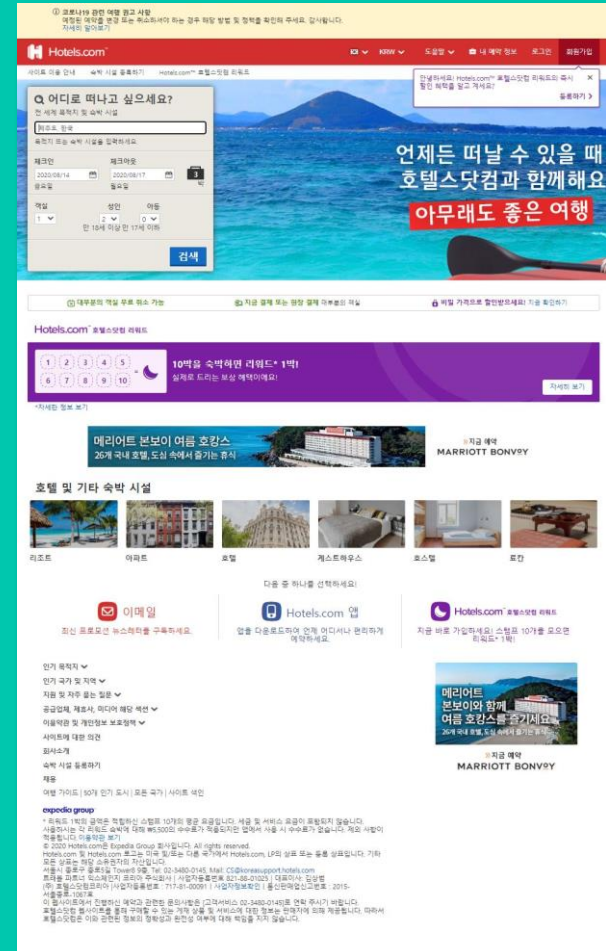


| 여행지 : 제주, 부산

코로나로 지친 여행객들을 위해  
상쾌한 공기를 마실 수 있는 **제주도와 인기 여행지인 부산**을 선정

| 여행 기간 : 8/14~8/17 (3박 4일)

광복절이 휴일로 끼여있어,  
부담없이 떠나기 좋은 **연휴기간을 여행날로** 선정



#호텔스닷컴 홈페이지

# 데이터 구축

## | 사용한 도구들



엑셀



파이썬



주피터



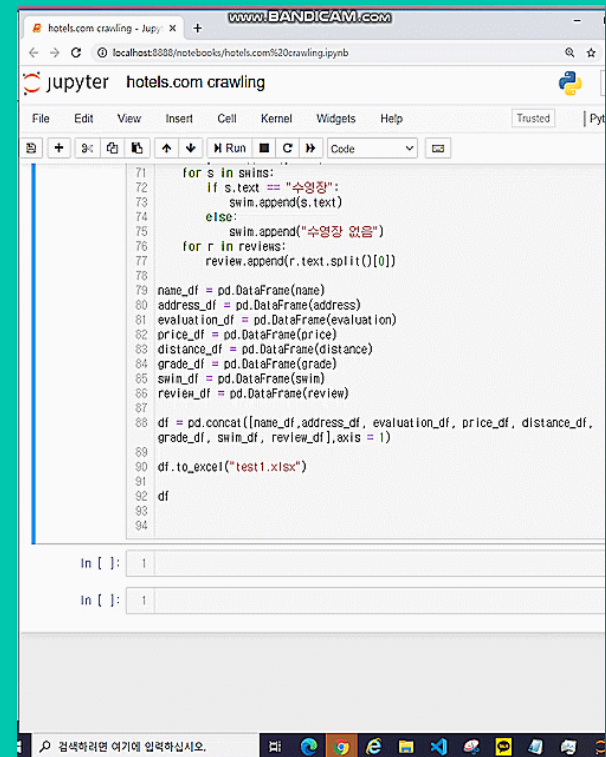
아나콘다

## | 변수 설정

웹사이트 내에서 필요한 정보나,  
유의미한 관계가 있을 것 같은 항목들을 직관적으로 선정  
(호텔명, 주소, 지역, 평점, 등급, 가격, 거리, 수영장, 후기 개수)

## | 데이터 정제 과정

- 1) 반복문을 활용한 자동스크롤을 통해,  
크롤링을 위한 **충분한 페이지 길이 확보**
- 2) Selenium 패키지(css\_selector, xpath)를 활용하여,  
각 변수 별 데이터 크롤링
- 3) 추출한 데이터들을 엑셀로 옮겨 정제한 후,  
다시 파이썬으로 데이터 프레임 구축



#구현한\_자동스크롤모습

# 데이터 구축

## | 데이터 크기

9개의 변수에 해당하는  
제주(269개), 부산(270개), 총539개의 데이터 추출

## | 특이사항

제주 파일의 공항까지의 거리와,  
부산 파일의 도심까지의 거리를  
통합파일에서 '거리'로 통합하고  
변수 '지역'으로 구분

지역		호텔명	주소	평점	가격	거리	등급	수영장	후기 개수
0	1	제주 아시아 펜션 (Jeu Asia Pension)	애월읍 가문동길 79, 제주시, 제주특별자치도, 63050, 대한민국	2.0	70,728	9.7	2.5	0	1
1	1	뷰티풀하우스 (Beautyfull House)	이아도로253번길 2, 서귀포시, 63543, 대한민국	2.0	114,544	30.0	2.5	1	1
2	2	호텔 뽐 (Hotel Bom)	충무대로34번길 11-32, 서구, 부산광역시, 부산광역시, 49269, 대한민국	2.0	147,273	9.3	2.5	0	1
3	2	엠 모텔 송도 (M Motel Songdo)	송도해변로125번길 9, 서구, 부산광역시, 49269, 대한민국	2.0	231,818	9.2	2.5	0	1
4	1	제주 알프스호텔 (Jeu Alps Hotel)	신산로 79, 제주시, 제주특별자치도, 63274, 대한민국	3.0	61,362	3.9	2.5	0	2
...	...	...	...	...	...	...	...	...	...
534	2	그랜드 엘시티 레지던스 (Grand LCT Residence)	달맞이길 30 현대마크동, 한도마크 동, 부산광역시, 부산광역시, 48099, 대한민국	10.0	2,495,455	9.9	4.0	0	1
535	2	아바호텔 (Ava Hotel)	서구 송도해변로 183, 부산광역시, 49269, 대한민국	10.0	218,182	9.2	2.5	0	2
536	2	티एस 호텔 (TS Hotel)	중앙대로 1931, 부산광역시, 부산광역시, 46230, 대한민국	10.0	127,274	12.0	2.5	0	1
537	2	행복한 공유 (Happy sharing)	중구 대청로 121-1, 부산광역시, 48932, 대한민국	10.0	271,819	6.2	2.5	0	2
538	2	굿데이메이트 호스텔 (Good Day Mate Hostel)	해운대구 구남로21번길 9-22, 부산광역시, 48095, 대한민국	10.0	854,545	9.2	2.5	0	1

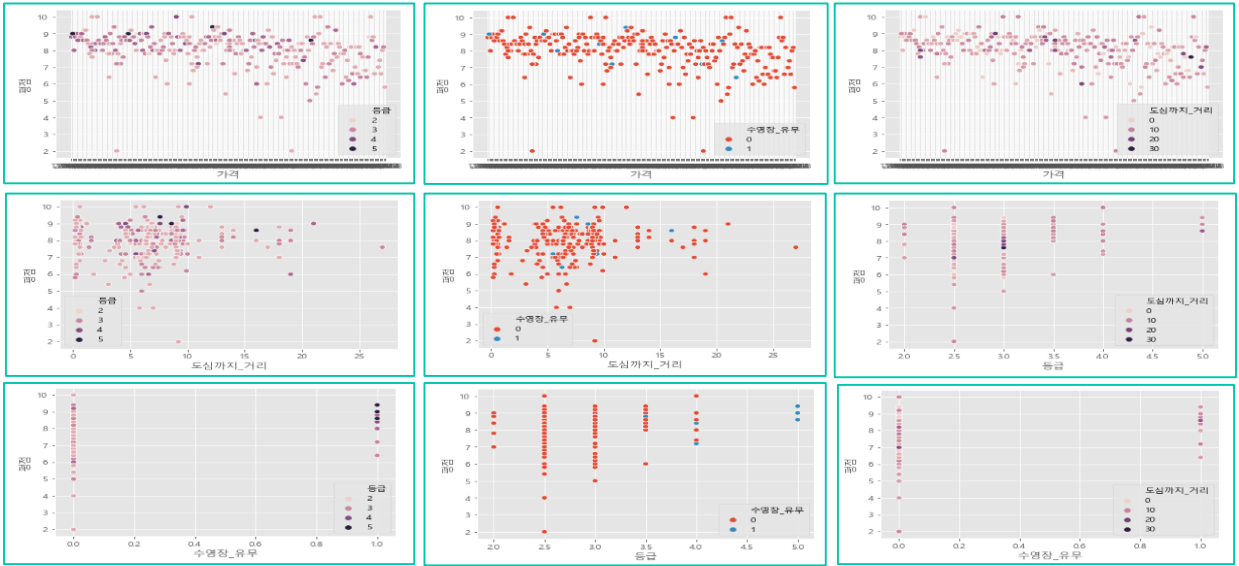
539 rows x 9 columns

539 rows × 9 columns

지역	호텔명	주소	평점	가격	거리	등급	수영장 유무	후기 개수
0	1	제주 아시아펜션을 가문동길 79, 제주시, 제	2.0	70,728	9.7	2.5	0	1
1	1	뷰티풀하우스(미아도로253번길 2, 서귀포시,	2.0	114,544	30.0	2.5	1	1
2	2	호텔 뽐 송도해변로 11-32, 서구, 제	2.0	147,273	9.3	2.5	0	1
3	2	엠 모텔 송도해변로 125번길 9, 서구, 제	2.0	231,818	9.2	2.5	0	1
4	1	제주 알프스(신산로 79, 제주시, 제주특별	3.0	61,362	3.9	2.5	0	2
5	1	호텔 엘시티 레지던스 30, 제주시, 제주특별	10.0	2,495,455	9.9	4.0	0	1
6	1	아바호텔 송도해변로 183, 제주시, 제	10.0	218,182	9.2	2.5	0	2
7	1	티에스 호텔 중앙대로 1931, 부산광역시,	10.0	127,274	12.0	2.5	0	1
8	1	행복한 공유 중구 대청로 121-1, 부산	10.0	271,819	6.2	2.5	0	2
9	1	굿데이메이트 호스텔 해운대구 구남로21	10.0	854,545	9.2	2.5	0	1
10	2	그랜드엘시티레지던스현대마크동, 한	10.0	2,495,455	9.9	4.0	0	1
11	2	아바호텔 서구 송도해변로 183, 부산	10.0	218,182	9.2	2.5	0	2
12	2	티에스 호텔 중앙대로 1931, 부산	10.0	127,274	12.0	2.5	0	1
13	2	행복한 공유 중구 대청로 121-1, 부산	10.0	271,819	6.2	2.5	0	2
14	2	굿데이메이트 호스텔 해운대구 구남로21	10.0	854,545	9.2	2.5	0	1
15	2	그랜드엘시티레지던스현대마크동, 한	10.0	2,495,455	9.9	4.0	0	1
16	2	아바호텔 서구 송도해변로 183, 부산	10.0	218,182	9.2	2.5	0	2
17	2	티에스 호텔 중앙대로 1931, 부산	10.0	127,274	12.0	2.5	0	1
18	2	행복한 공유 중구 대청로 121-1, 부산	10.0	271,819	6.2	2.5	0	2
19	2	굿데이메이트 호스텔 해운대구 구남로21	10.0	854,545	9.2	2.5	0	1
20	2	그랜드엘시티레지던스현대마크동, 한	10.0	2,495,455	9.9	4.0	0	1
21	2	아바호텔 서구 송도해변로 183, 부산	10.0	218,182	9.2	2.5	0	2
22	2	티에스 호텔 중앙대로 1931, 부산	10.0	127,274	12.0	2.5	0	1
23	2	행복한 공유 중구 대청로 121-1, 부산	10.0	271,819	6.2	2.5	0	2
24	2	굿데이메이트 호스텔 해운대구 구남로21	10.0	854,545	9.2	2.5	0	1
25	2	그랜드엘시티레지던스현대마크동, 한	10.0	2,495,455	9.9	4.0	0	1
26	2	아바호텔 서구 송도해변로 183, 부산	10.0	218,182	9.2	2.5	0	2
27	2	티에스 호텔 중앙대로 1931, 부산	10.0	127,274	12.0	2.5	0	1
28	2	행복한 공유 중구 대청로 121-1, 부산	10.0	271,819	6.2	2.5	0	2
29	2	굿데이메이트 호스텔 해운대구 구남로21	10.0	854,545	9.2	2.5	0	1
30	2	그랜드엘시티레지던스현대마크동, 한	10.0	2,495,455	9.9	4.0	0	1
31	2	아바호텔 서구 송도해변로 183, 부산	10.0	218,182	9.2	2.5	0	2
32	2	티에스 호텔 중앙대로 1931, 부산	10.0	127,274	12.0	2.5	0	1
33	2	행복한 공유 중구 대청로 121-1, 부산	10.0	271,819	6.2	2.5	0	2
34	2	굿데이메이트 호스텔 해운대구 구남로21	10.0	854,545	9.2	2.5	0	1
35	2	그랜드엘시티레지던스현대마크동, 한	10.0	2,495,455	9.9	4.0	0	1

#완성된데이터프레임  
파이선(위),엑셀(아래)

#Seaborn\_scatterplot

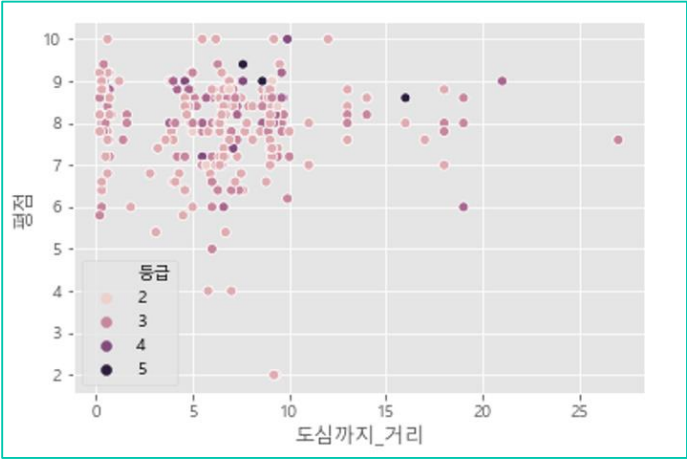


주어진 데이터에서 변수간 대략적인 상관관계를  
알아보기 위해 산포도 그래프를 활용

호텔 선정에 가장 크게 영향을 끼칠 것 같은 ‘평점’ 을 기준으로(y축)  
모든 변수를 x축에 순차적으로 대입하여 변수간 관계를 시각화



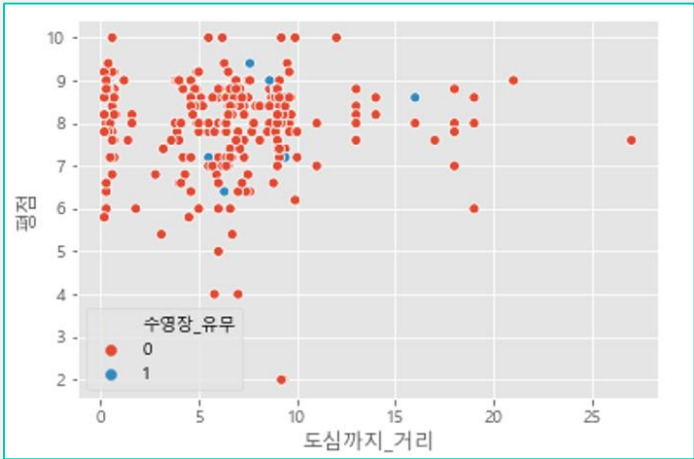
#평점&거리의 관계가등급에 미치는 영향



도심까지의 거리가 가까울 수록 호텔 수가 많다

- 1) 도심과의 거리와 평점은 뚜렷한 관계x  
(But, 호텔의 등급과 평점은 정비례 관계)
- 2) 사람들은 호텔의 등급에 상관없이  
도심과의 거리가 가까우면 높은 평점을 부여

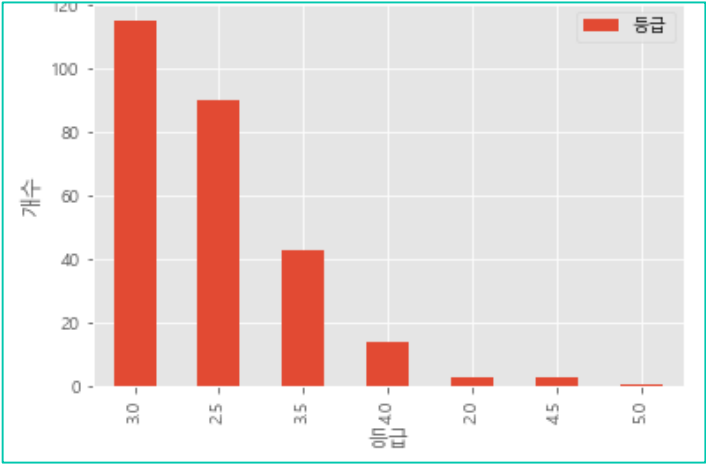
#평점&거리의 관계가수영장유/무에 미치는 영향



도심까지의 거리가 가까울수록 수영장을 보유할 가능성 ↑

- 1) 도심과의 거리가 멀 때, 오히려 수영장이 없는 호텔에  
숙박할 가능성이 있다.
- 2) 수영장을 보유한 호텔은 도심과의 거리가 멀어도 높은  
평점을 받는다.

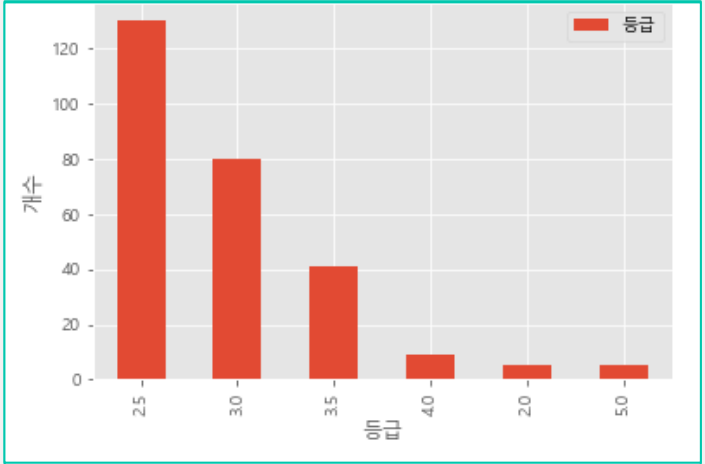
#등급별호텔수\_제주



제주도에는 3등급 호텔이 가장 많다

- 1) 3등급 호텔의 수요에 따른 공급으로 판단  
사람들은 3.0, 2.5, 3.5 수준의 호텔을 가장 선호할 것이다
- 2) 제주도에는 3.0, 2.5, 3.5 등급의 호텔이 270개 중,  
248개로 약 92퍼센트를 차지한다(91.85%)

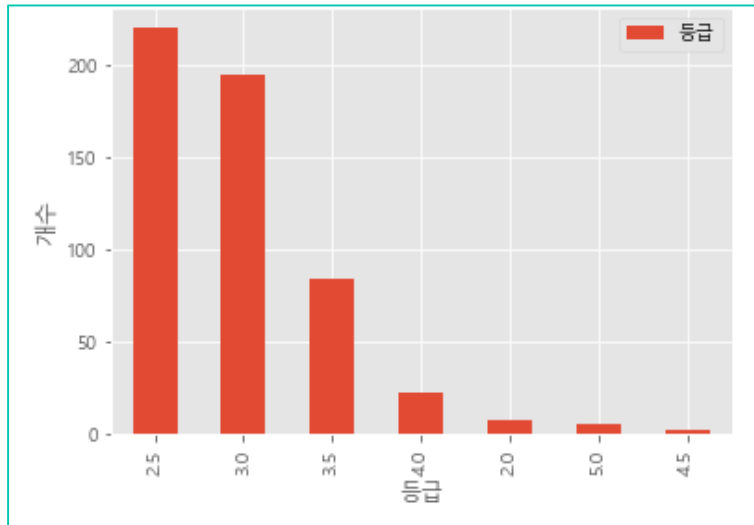
#등급별호텔수\_부산



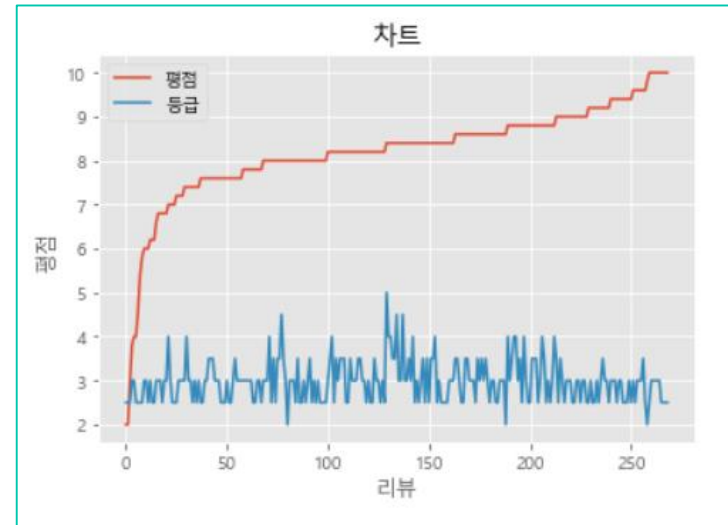
부산에는 2.5등급 호텔이 가장 많다

- 1) 제주도와 비슷하게 3점대 전후의 호텔이 가장 많다
- 2) 부산에는 2.5, 3.0, 3.5 등급의 호텔이 270개 중,  
251개로 약 93퍼센트를 차지한다(92.96%)

#등급별 호텔수\_통합



#등급별 호텔수\_평점추가



## 결론

/

- 1) 제주도와 부산에는 2.5~3.5 등급의 호텔이 90% 이상 존재한다
- 2) 평점이 높을수록, 호텔의 등급이 높은 것은 아니다.
- 3) 위 내용을 통해 국내의 다른 지역의 호텔 수와 비교해 본다면, 2.5~3.5등급의 호텔이 가장 많을 것으로 추측할 수 있다.

오피스를 잘 활용하자!  
**앞으로의 분석은...**



2일만에 프로젝트 하나 너무 힘들어요...

하나의 분석 툴을 고집하는게 아니라 다양한 상황에 맞춰서,  
적재적소에 활용할 줄 아는 **능력 있는 개발자**가 되었으면!

감사합니다

---

정효인