

데이터분석캡스톤디자인

---

# 5주차 수행보고

**Khupid** 조

산업경영공학과 김동혁  
관광학과 류연주  
산업경영공학과 유정수

## 동혁-수행보고



장소 정보 + 리뷰 크롤링 완료

DiningCode\_df.shape

(2360, 9)

DiningCode\_review\_df.shape

(19684, 4)

2360개의 식당

19684개의 리뷰

DiningCode\_df.head()

	name	address	category	main_mn	price	opng_tm	rating	rvw_cnt
--	------	---------	----------	---------	-------	---------	--------	---------

0	중앙해장	DiningCode_review_df.head(10)						
1	브루클린더 버거조인트							
2	농민백암순대분점							
3	다윤타너 청담점							
4	퓨즈							

	res_name	user_name	rating	review
0	중앙해장	rimnim	4	"깔끔한 해장국으로 속 까지 편안하게"보통 내가 접한 양선지해장국은 너무 매콤하거나...
1	중앙해장	명정한험스터	4	음식맛은 국물은 얼큰 깔끔 속이 풀리는 맛이고 곱창은 곱도 차있고 씹으면 고소, 특...
2	중앙해장	달리는진저맨	4	상성역 맛집 탐급 중 하나인 중앙해장일요일 다섯시쯤 방문. 웨이팅은 없었다. 첫 인상...
3	중앙해장	ye	5	- 수요일식회 맛집, 친구 추천으로 방문. 웨이팅 순번 적고 가게에서 이름 호명될때...
4	중앙해장	Kunwoo Kim	5	퇴근길 아내와 저녁식사어제 먹은 술이 확 내려갑니다깔끔한 맛 강추!!우육탕 고기가 ...
5	중앙해장	일상다반사	4	매장 뒷편으로 주차장 있음항상 사람 많고 대기 시간 엄청남음식 맛은 매우 훌륭하나 ...
6	중앙해장	Kaanilm	3	주매뉴인 곱창전골은 모르겠으나내장탕은 가격이 비전가에 비해서 맛이 상당한 편은 아니...
7	중앙해장	hange	4	곱창전골이 유명하대서 웨이팅감수하고 일찍 방문했어요 내부가 넓어서 생각보다 일찍 차...
8	중앙해장	조이	5	일단 가보고 드셔보세요. 곱창을 좋아하거나 곱창전골을 좋아하거나 얼큰한 해장국을 좋아...
9	중앙해장	KW	4	곱창전골로 유명한 곳이라던데 확실히 곱창전골이 맛있었다. 하지만 버릇이 없고 사람이...



장소 정보 + 리뷰 크롤링 중

len(Mango\_urls)

4358

4358개의 식당

```
def crawling_reviews_MangoPlate(name, link):
    reviews = pd.DataFrame(columns=['res_name', 'user_name', 'rat
```

```
    try:
        driver.get(link)
    except:
        return None
```

```
    while True:
        try:
            click_more = driver.find_element_by_text('더보기')
            click_more.click()
        except:
            break
```

```
    time.sleep(1)
```

```
    try:
        html = driver.page_source
        soup = BeautifulSoup(html, 'html.parser')
        for user in soup.find('div', class_='user'):
            tmp = user.find('span', class_='user_name')
            user_name = tmp.get_text()
            user_rating = int(re.findall(r'\d+', user_rating)[0])
            user_review = user.find('div', class_='review').get_text()
            reviews = reviews.append({'res_name': name, 'user_name': user_name, 'rating': user_rating, 'review': user_review}, ignore_index=True)
```

```
    except:
        return None
```

```
    return reviews
```

### 망고플레이트

```
M_restaurants_df = pd.DataFrame(columns=['name', 'address'])
M_review_df = pd.DataFrame(columns=['res_name', 'user_name', 'rating', 'review'])
```

```
with open("../data/restaurants_MangoPlate.txt", "r") as f:
    text = f.read()
    Mango_urls = text.split()
```

Mango\_urls

```
[ 'https://www.mangoplate.com/restaurants/0-jWS81VhN',
  'https://www.mangoplate.com/restaurants/DfK9o3D1ITgg',
  'https://www.mangoplate.com/restaurants/ZT0NvVaxck',
  'https://www.mangoplate.com/restaurants/zfARB-g1NKRx',
  'https://www.mangoplate.com/restaurants/cmdrCTi09Mdk',
  'https://www.mangoplate.com/restaurants/LcJmG1zNzID',
  'https://www.mangoplate.com/restaurants/XcJmG1zNzID',
  'https://www.mangoplate.com/restaurants/0FwTUKRD6F',
  'https://www.mangoplate.com/restaurants/T_wllrv3Ba6',
  'https://www.mangoplate.com/restaurants/nj3Hk9TB47d',
  'https://www.mangoplate.com/restaurants/NMH5sAD0D01h',
  'https://www.mangoplate.com/restaurants/WluttekMvNOS',
  'https://www.mangoplate.com/restaurants/a8c0oYimUji1',
  'https://www.mangoplate.com/restaurants/NOVaeYukjT51',
  'https://www.mangoplate.com/restaurants/Bwf13T_W5eN_',
  'https://www.mangoplate.com/restaurants/00s1Jaxz55',
  'https://www.mangoplate.com/restaurants/0P21XS1izr',
  'https://www.mangoplate.com/restaurants/H0_nC5TGgT0c',
  'https://www.mangoplate.com/restaurants/Hcf9P64EBD',
  'https://www.mangoplate.com/restaurants/0FwTUKRD6F' ]
```

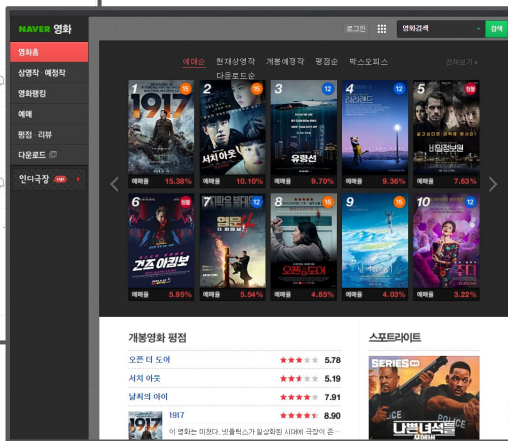
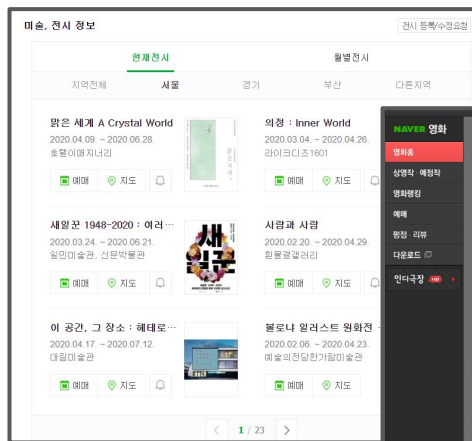
## 정수-수행보고

### 전시 오픈갤러리 & 네이버 전시정보

장소 정보 + 리뷰 크롤링 중

### 영화 왓차 & 네이버 영화

장소 정보 + 리뷰 크롤링 완료



data.to_csv('opengallery_raw.csv')										
data										
	title	artists	location	term	open_at	close_at	day	explain	price	
0	(VERY ART) ARTIST TALK 유투브 동시중계 (4/18 SAT)	Sen Chungin	서울특별시 중로구 우정국로 45-14 (수송동)	2020-04-18 ~ ...	10:00 ~ 18:00...	10:00 ~ 18:00...	10:00 ~ 18:00...	OCI미술관에서 선정 초대개개인전 (VERY ART)의 아티스트 토크를 진행합니다....	~	~
1	모네에서 세잔까지	Sen Chungin	서울특별시 서초구 남부순환로 2406 (서초동) 예술의전당 한가 람미술관	2020-01-17 ~ ...	10:00 ~ 19:00...	10:00 ~ 19:00...	10:00 ~ 19:00...	이번 모네에서 세잔까지: 국립미술 관 미술관 인상파 걸작들은 전시 와 문화 예술의...	~	~
2	9 to 0 (Nine to Zero)	Sen Chungin	서울특별시 마포구 독막로9길 34 (서교동)	2020-04-03 ~ ...	11:00 ~ 20:00...	11:00 ~ 20:00...	11:00 ~ 20:00...	[전시요약]아트아트, 2020.04.03 ~ 2020.04.19선과 선을 이어 독독...	~	~
3	볼로냐 일러스트 원화전 2019	Sen Chungin	서울특별시 서초구 남부순환로 2406 (서초동) 예술의전당 한가 람미술관	2020-02-06 ~ ...	10:00 ~ 19:00...	10:00 ~ 19:00...	10:00 ~ 19:00...	<볼로냐 일러스트 원화전>은 1967 년부터 시작하여 2019년 53회째 전 시를 진행...	~	~
4	자원이 지닌 태도의 힘	Sen Chungin	서울특별시 중로구 삼정로7길	2020-04-03 ~ ...	11:00 ~ 20:00...	11:00 ~ 20:00...	11:00 ~ 20:00...	자원이 지닌 태도의 힘	~	~

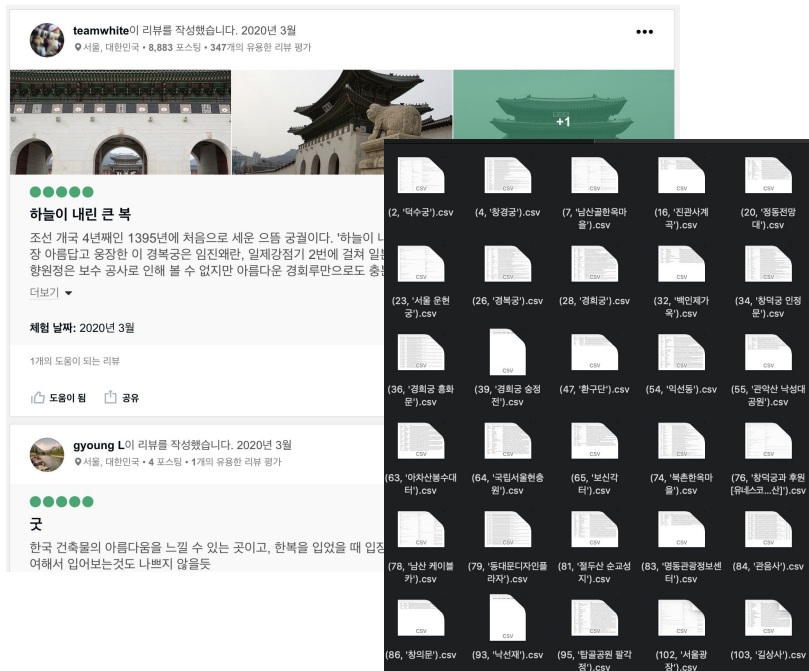
data_watcha										
	title	year	country	genre	runtime	release_date	description	rating	price	
4	Deepwater Horizon	2016	미국	액션/드라마/역사/스릴러/재난	1시간 47분	2016년 4월 20일	미국 루이지애나주 앞바다 멕시코만 석유 시추선 '딥워터 호라이즌'이 폭발한다.	7.6	~	~
665	Pirates of the Caribbean: The Curse of the Black Pearl	2003	미국	액션/판타지/모험	2시간 23분	2003년 7월 11일	매력 넘치는 해적 캡틴 잭 스패로우(조니 뎀 분)에게 수정처럼 맑고 투명한 카리비...	7.8	~	~
666	Thor	2011	미국	액션/모험/판타지/SF	1시간 52분	2011년 4월 15일	신의 세계 '아스가르드'의 후계자로 강력한 파워를 지닌 천둥의 신 '토르' 평소...	7.7	~	~
667	不能說的秘密	2007	대만	로맨스/판타지/드라마	1시간 41분	2007년 7월 13일	예술학교로 전학 온 상룡(주절룡)은 아버지의 영향을 받아 피아노에 천부적인 소질을...	7.6	~	~
668	Harry Potter and the Prisoner of Azkaban	2004	영국, 미국	판타지/미스터리/가족/모험/액션	2시간 21분	2004년 6월 17일	13세가 된 해리 포터(다aniel 래드클리프)는 또 한번의 여름 방학을 이모 가족인 더...	7.9	~	~
669	My Sassy Girl	2001	한국	코미디/로맨틱/코미디/로맨스	2시간 2분	2001년 11월 15일	건우는 그녀와 함께 같은 지하철을 타게된다. 취해서 비틀거리지만 않았다면 정말 매력...	7.8	~	~

전시는 리뷰가 많지 않아 고민 중

영화의 경우 장르 별로 더 수집 예정

## 연주-수행보고

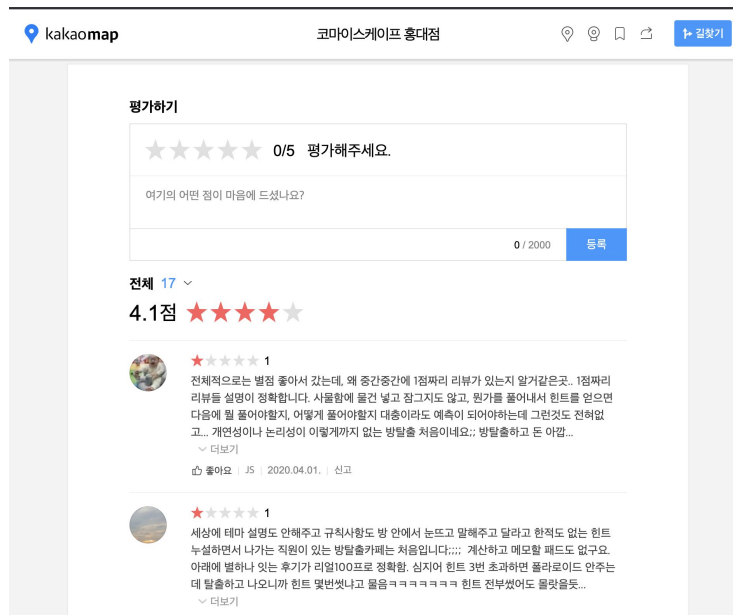
### TripAdvisor 장소 정보 + 리뷰 크롤링 완료



### Kakao API 장소 정보 + 리뷰 크롤링 중

```
In [2]: region='서울'
spot_ls = ['만화카페', '불림장', '당구장', '멀티방', '방탈출', '보드카페', '코인노래방', '공원']
spot_kind = ['a', 'b', 'c', 'd', 'e', 'f', 'g', 'h']

In [3]: place_df = pd.DataFrame(columns=['place_id', 'name', 'location', 'tags', 'open_time', 'open_day'])
```



전처리 계획

1. 장소정보에 대한 field 형식 통일 (갈 곳, 먹을 것, 볼 것)

필드명	이름	자료형	설명	예시
장소ID	p_id			
장소명	p_name	str	상호명	경복궁
소재지 (주소)	address	str	장소가 속해있는 전체 주소지	서울특별시 종로구 세종로 사직로 161
행정구	district	category	소재지의 행정구	종로구
영업시간	monday-sun day	tuple	요일별(월-일) 오픈시간과 마감시간	월(12:00,18:00), 화(12:00, 21:00) 일(null, null)
카테고리	category	category	대분류(갈곳, 먹을 것, 볼 것)별 소분류 카테고리	(갈)노래방/공원 등 (볼) 전시/영화 등 (먹)한식/중식 등
가격	price	int	단위(원) / 성인 기준 / 1인 기준	15000
평점	p_rate	float	장소에 대한 총평점	3.5

## 전처리 계획

### 2. 리뷰정보에 대한 field 형식 통일 (갈 곳, 먹을 것, 볼 것)

필드명	이름	자료형	설명	예시
장소ID	p_id	String	장소 아이디	G001
사용자ID	u_id	String	사용자 아이디	파잘알동짱
평점	u_rate	float	사용자별 장소 평가점수	4.5
리뷰	review	String	사용자별 장소 리뷰	주말에 tv에 나온 곳을 보고 꽃힌 친구덕에 가봤습니다 ㅋㅋㅋ 전부터 궁금했는데 드디어! 맛집 답게 ...