

데이터분석캡스톤디자인

---

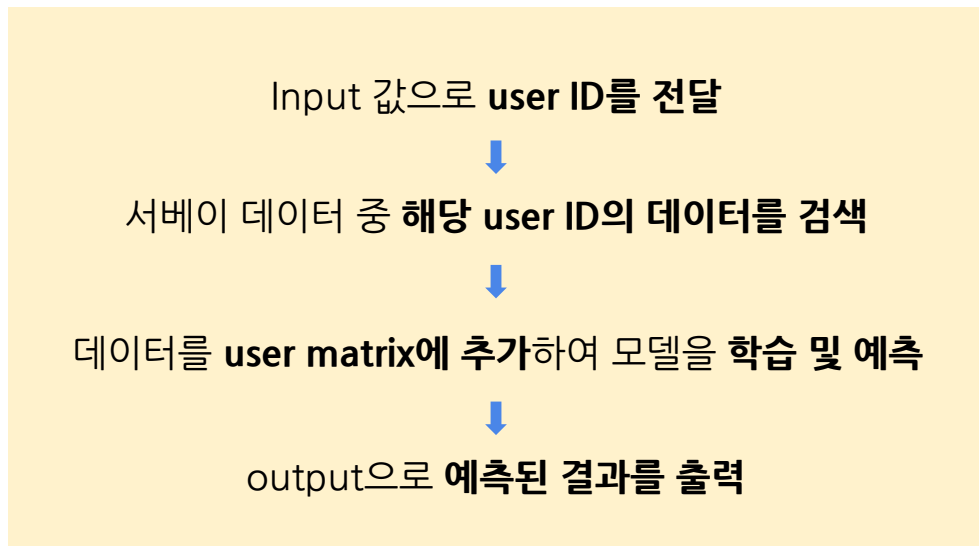
# 14주차 수행보고

**Khupid** 조

산업경영공학과 김동혁  
관광학과 류연주  
산업경영공학과 유정수

## 이번 주 한 일

1. 최종 평가 데이터 확보를 위한 설문지 작성
2. 영화: 현재 상영작 추천을 위한 데이터 추가 수집
3. 서베이 데이터를 통해 추천을 하기 위한 로직 설계



# 데이터 확보를 위한 설문지 작성

## 서울 데이트 코스 추천 알고리즘 테스트를 위한 설문

안녕하세요. 저희는 경희대학교 2020학년도 1학기 경희대학교 데이터분석캡스톤디자인을 수강하는 KHUPID팀입니다.

본 설문은 KHUPID팀의 프로젝트 "취향 분석을 통한 서울 데이트 코스 추천 알고리즘 개발"에 따른 내용이며, 유저의 취향을 분석하여 사용자 맞춤 서울 데이트 코스(맛집 - 가보기 좋은 곳 - 영화 추천)를 추천하기 위해 귀하의 취향을 파악할 수 있는 간단한 질문들을 포함하고 있습니다.

설문 응답은 단순히 프로젝트를 위해서만 사용되며, 설문에 사용하신 ID와 이메일 주소는 프로젝트가 종료된 직후(2020년 07월 10일) 파기됩니다.

\* 필수항목

설문에 사용할 ID를 입력하세요. \*

내 답변

설문 결과를 받을 이메일 주소를 입력하세요. \*

내 답변

**1차 설문지:** 총 30명의 설문 대상자들에게 데이터 조사용 설문 발송

- 2차 설문지 발송을 위한 email 질문
- 형용사별 감정의 강도 질문
- 장소 방문 경험 질문
- 데이트를 하고자하는 행정구 질문

**2차 설문지:** 1차 설문에서 확보한 개인별 데이터를 분석하여 맞춤형 데이트 코스 결과 발송, 추천 코스에 대한 **만족도** 설문

MOS 방식을 사용해

3.5 / 5 이상이 최종 목표

# 형용사 선정

선택한 형용사가 들어있는 클러스터에 점수를 부여하기위함

단어들에 대한 호감도는?

아래 단어들에 대해 본인이 느끼는 호감의 강도를 선택해주세요. \*

\*긍정적일수록 5, 부정적일수록 0을 선택해주세요

	0	1	2	3	4	5
이국적인	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
가격이 싼	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
스릴있는	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
고전적인	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
감성적인	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

아래 단어들에 대해 본인이 느끼는 호감의 강도를 선택해주세요. [이국적인]	아래 단어들에 대해 본인이 느끼는 호감의 강도를 선택해주세요. [가격이 싼]
4	5
4	4
3	3
5	5
5	0
5	5

분야별 클러스터를 대표하는 단어를 도출한 뒤,  
비슷한 의미를 지닌 단어는 묶음

맛집/카페  
cluster

데이트명소  
cluster

영화  
cluster

'감성적인' : ['group\_69', '22', 'group\_69'],

```

{
  #공통 특성 (19)
  '이국적인' : ['group_93', '24', 'group_43'],
  '가격이 싼' : ['group_45', '48', 'group_42'],
  '스릴있는' : ['group_74', '29', 'group_55'],
  '고전적인' : ['group_1', '35', 'group_25'],
  '감성적인' : ['group_69', '22', 'group_69'],
  '전망좋은' : ['group_22', '11', 'group_53'],
  '자연적인' : ['group_42', '38', 'group_47'],
  '모험적인' : ['group_60', '42', 'group_77'],
  '차분한' : ['group_13', '88', 'group_22'],
  '예쁜' : ['group_107', '86', 'group_85'],
  '문화적인' : ['', '51', 'group_56'],
  '친절한' : ['group_3', '43', ''],
  '심플한' : ['', '52', 'group_23'],
  '신나는' : ['group_19', '85', ''],
  '감동적인' : ['group_54', '', 'group_4'],
  '종교적인' : ['', '26', 'group_16'],
  '다채로운' : ['group_109', '', 'group_17'],
  '깔끔한' : ['group_66', '', 'group_36'],
  '숨겨진' : ['group_116', '', 'group_87'],

```

```

#먹을 곳 (5)
'자극적인' : ['group_15', '', ''],
'짬조름한' : ['group_34', '', ''],
'무제한의' : ['group_11', '', ''],
'정성스러운' : ['group_20', '', ''],
'전문적인' : ['group_39', '', ''],

```

```

#갈 곳 (5)
'평범한' : ['', '98', ''],
'의미있는' : ['', '59', ''],
'신비한' : ['', '82', ''],
'시골벽적한' : ['', '9', ''],
'흥미로운' : ['', '20', ''],

```

```

#볼 곳 (5)
'웅장한' : ['', '', 'group_98'],
'눈물나는' : ['', '', 'group_76'],
'비현실적인' : ['', '', 'group_8'],
'동물의' : ['', '', 'group_67'],
'십대의' : ['', '', 'group_30']
}

```

2개 이상의 분야에 속한 단어들

단일 분야에 속한 단어들

# 장소

NCF 모델에 사용하기 위해 사용자가 소비해본 아이템을 조사

아래의 장소/영화들 중 내가 가본 곳/본 것은?


1) 서울 내의 맛집/카페, 2) 대표적인 데이트 명소, 3) 영화에 대해 각 문항당 10가지의 옵션 중 응답자 본인이 실제로 방문했던 장소를 선택해주세요.


- \* 서울 내의 맛집/카페의 경우 지점(ex. 연남점, 한남점)을 구분하지 않습니다. 따라서 지점 구분없이 방문한 경험이 있다면 선택해주세요.
- \* 복수 선택이 가능합니다.
- \* 무응답이 가능합니다.


설문에 사용할 ID를 입력하세요.	1) 서울 맛집 / 카페
dhyeok1996	정돈, 명동교자, 교대이층집, 매화반점
Yepp	미분당, 정돈, 광장시장 육회자매집, 진주집
hhhhhhhhh	미분당, 정돈, 브루클린더버거조인트, 광장시장 육회자매집, 명동교자, 교대이층집
Zain	소이연남, 광장시장 육회자매집
thwisWkd	브루클린더버거조인트, 소이연남, 광장시장 육회자매집, 명동교자, 이태리총각, 젤라띠젤라띠
tbvjwodnraos	정돈, 하동관, 광장시장 육회자매집, 자니엄플링, 명동교자, 육전식당, 젤라띠젤라띠
123vcorn	
이현구	테일라커피, 광장시장 육회자매집, 명동교자, 육전식당
해지니	
와리	자니엄플링, 명동교자


맛집/카페, 데이트 명소, 영화 각각 상위 10개의 장소를 골라  
방문 유/무를 질문

1) 서울 맛집 / 카페



☐ 테일라커피



☐ 미분당



☐ 정돈


☐ 브루


2) 대표적인 데이트 명소



☐ 경복궁



☐ 남산서울타워



☐ 용산 전쟁기념관

3) 영화


☐ 어거스트 러쉬


☐ 클릭


☐ 이월스


☐ 다이버전트 시리즈 : 얼라전트

# 추천 받을 행정구역

추천받을 구역은?

본인이 장소를 추천받을 행정구역을 선택해주세요. \*

\*복수응답이 가능합니다.

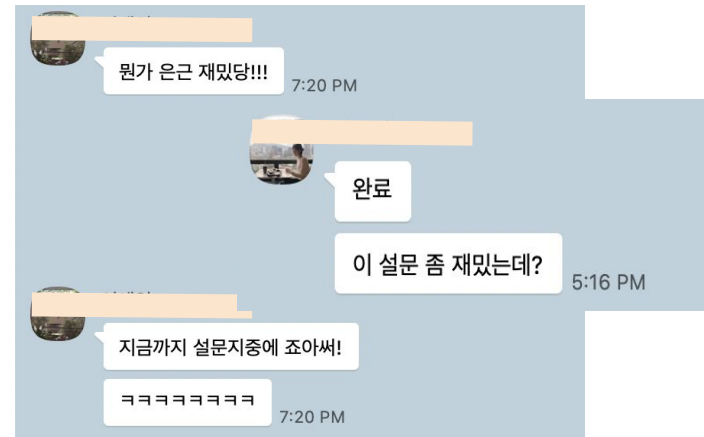
- ☐ 종로구
- ☐ 중구
- ☐ 용산구
- ☐ 성동구
- ☐ 광진구
- ☐ 동대문구
- ☐ 중랑구
- ☐ 성북구
- ☐ 강북구
- ☐ 도봉구

설문에 사용할 ID를 입력하세요.	본인이 장소를 추천받을 행정구역을 선택해주세요.
dhyeok1996	강남구
Yepp	종로구, 중구, 용산구, 성동구, 동대문구, 성북구, 은평구, 서대문구, 마포구, 강남구, 송파구
hhhhhhhhh	종로구, 중구, 용산구, 성동구, 광진구, 서대문구, 영등포구, 동작구, 관악구, 서초구, 강남구, 송파구
Zain	광진구, 동대문구, 중랑구, 성북구, 서초구, 강남구, 송파구, 강동구
thwlsWkd	종로구, 중구, 용산구, 동대문구, 성북구, 강북구, 도봉구, 노원구, 마포구
tbvjwodnraos	중구, 용산구, 마포구, 영등포구, 송파구
123vcom	종로구, 중구, 용산구, 중랑구, 성북구, 마포구, 강남구
이현구	종로구, 서대문구, 마포구, 강서구, 송파구
헤지니	영등포구, 서초구, 강남구, 송파구
와리	종로구, 중구, 광진구, 노원구
rr_sky	노원구, 은평구, 마포구, 강남구
minho0968	종로구, 중구, 용산구, 노원구, 은평구, 마포구, 영등포구, 서초구, 강남구, 송파구
wjdt1912	강남구, 송파구

사용자에게 추천받을 행정 구역을 선택하도록 함.



prediction score가 가장 높은 장소 중에서 지역구를 필터링해 결과값을 이메일로 전송



# 현재 상영작 추가 크롤링

## 현재상영영화

상영영화

영화제

리뷰

이미지

15

결백

네티즌

★★★★★

7.72

참여 1,749명

예매율 30.95%

개요

드라마

110분

2020.06.10 개봉

감독

박상현

출연

신혜선, 배종옥, 허준호

예매하기

포토보기

예고편

메이킹

2

온워드: 온워드

네티즌

★★★★★

개요

판타지, 모험

110분

2020.06.10 개봉

감독

댄 스캅론

출연

롱 홀랜드

예매하기

15

사라진

네티즌

★★★★★

개요

미스터리, 공포

110분

2020.06.10 개봉

감독

정진영

출연

조진웅, 배

예매하기

현재 상영작(154)과  
영화에 대한 리뷰(11836) 크롤링

df_movie			
	name	url	
0	결백	https://movie.naver.com/movie/bi/mi/basic.nhn?...	
1	온워드: 단 하루의 기적	https://movie.naver.com/movie/bi/mi/basic.nhn?...	
2	사라진 시간	https://movie.naver.com/movie/bi/mi/basic.nhn?...	
3	침입자	https://movie.naver.com/movie/bi/mi/basic.nhn?...	
4	슈퍼스타 두루	https://movie.naver.com/movie/bi/mi/basic.nhn?...	
...	...	...	
149	창문 넘어 도망친 100세 노인	https://movie.naver.com/movie/bi/mi/basic.nhn?...	
150	첫사랑	https://movie.naver.com/movie/bi/mi/basic.nhn?...	
151	캐리	https://movie.naver.com/movie/bi/mi/basic.nhn?...	
152	프랭크	https://movie.naver.com/movie/bi/mi/basic.nhn?...	
153	플라이	https://movie.naver.com/movie/bi/mi/basic.nhn?...	
154 rows x 2 columns			

df_movie_review			
	mid	user	rate
0	0	허현(heoh****)	6
1	0	황금마차(goga****)	1
2	0	케구리(jhw6****)	2
3	0	김스(volk****)	1
4	0	밀고나(mjh5****)	1
...	...	...	...
5	153	안경침이(ayan****)	8
6	153	개릭(got****)	9
7	153	뉴존(newz****)	10
8	153	무탈 Watanabe(yaso****)	8
9	153	효정바라기(taeg****)	8
11836 rows x 4 columns			

## 서베이 데이터를 통해 개인화된 결과를 추천해주기 위한 내부 로직 설계

- i. userID를 인풋으로 넣음
- ii. survey\_result\_matrix 내의 userID의 결과 row 를 추출
- iii. 형용사 응답 column 잘라오기
- iv. 맛집, 데이트 명소, 영화 user\_matrix를 가져오고
- v. userID로 row 추가
- vi. user가 각각 클러스터 별로 준 점수를 매칭하여 row 추가
- vii. 추가한 row 정규화
- viii. user\_matrix와 item\_matrix의 코사인 유사도를 구해 contents-based result 저장
- ix. 장소 선택 응답 column 잘라오기
- x. 맛집, 데이트 명소, 영화 ncf\_matrix에 user-id - place-id를 매칭하여 각각 row추가
- xi. ncf 모델에 넣고 ncf result 저장
- xii. contents-based result와 ncf result를 곱해 최종 prediction이 높은 순서로 정렬
- xiii. 지역구 리스트 내에 있는 곳들을 loop로 돌며 통해 맛집,데이트 명소 모두 충족시키는 지역구가 있는 지역구와 장소를 선정



## 다음 주 할 일

1. 설문 데이터로 NCF 학습
2. 설문 결과 처리 알고리즘 코딩
3. 설문 결과 이메일 회신 및 만족도 취합
4. 최종 발표 준비
  - 데모영상 촬영
  - 발표자료 준비
  - 보고서 준비