

안녕하세요!

이 프로젝트는 리뷰 분석을 통한 추천시스템 관련 프로젝트입니다.



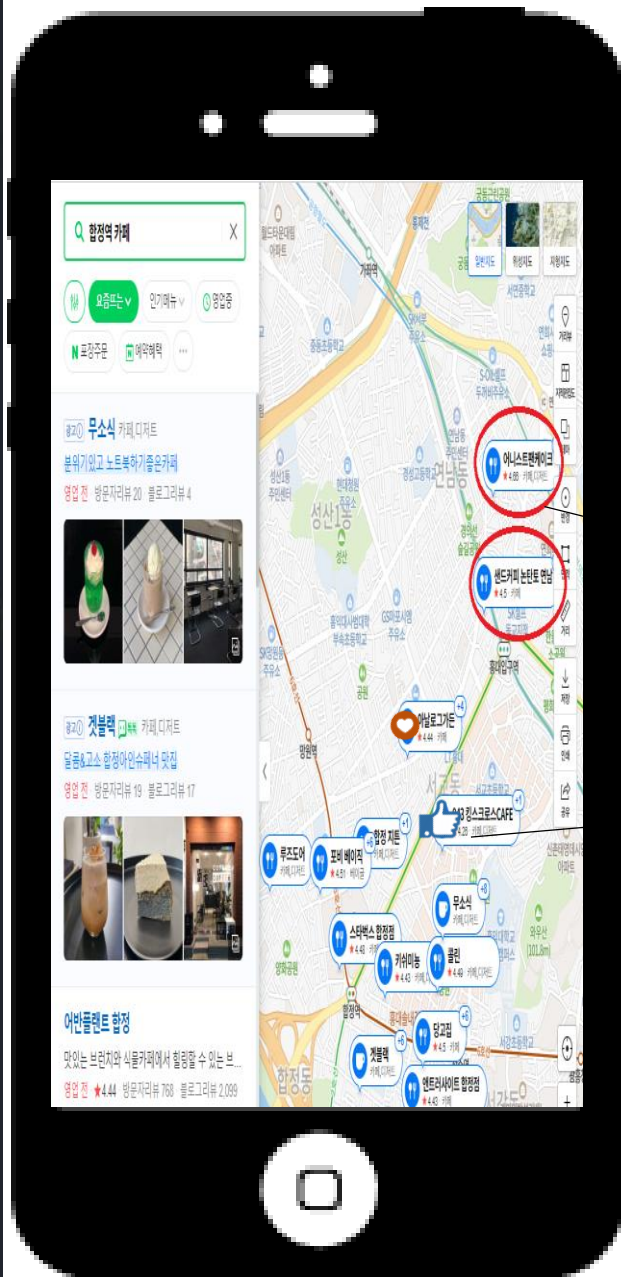
café-dafé

Recommender system (by **Naver-map**)

Team1.
이시우, 홍은지

Project motivation

,Why?



'검색을 통해서 커피를 마시러 가는 시대에 살고 있습니다'

'합정역 카페'라고 검색을 하면, 망원, 상수, 홍대, 합정 결과가 모두 출력

네이버 지도에 있는 저장 기능을 통해서 '다음에 가봐야지'라는 생각으로 저장해두지만, 저장해둔 지역을 가야하는 날에는 모든 선택지가 애매합니다.

우리는 분명하게 자신이 가고 싶은 카페가 있지만,
찾는 것은 쉽지않은 과정입니다.

이러한 고민을 해보신 분들을 위해서 이 프로젝트를 설계했습니다.

<데이터 크롤링>

1	cafe_name	quary	region	evaluation
2	umoe	합정	서교동	인테리어가 감각적이라고 느꼈습니다 사장님이 친절하시고 디저트가 맛있어요 소르베도 재미있습니다
3	umoe	합정	서교동	커피와 디저트 모두 너무 맛있어요 가게 분위기도 원목 가구를 써서 그런지 단정하고 대화하기 좋았어요. 의자는 참 무겁더라고요. 집어넣고 가기엔 쉽지 않았네요. 늦게까지 영업하
4	umoe	합정	서교동	프렌치토스트 소문대로 존맛탱
5	어반플랜트 합정	합정	합정동	자연에 둘러싸여 있는듯한 색다른 기분도 느낄수있었고~ 시그니처 오믈렛 정말 맛있었어요!! 스파게티류는 면류가 없어서 아쉬웠지만~ 음식맛은 괜춘! 무엇보다 분위기가 너무 좋
6	어반플랜트 합정	합정	합정동	직원분이 테이블링을 도와주셔서 웨이팅을 하다보면 이름을 불러주세요~ 음료도 맛있었고 가성비가 좋았던 것 같아요 개인적으로 매장 내에 식물이 많아서 출냄새가 나서 좋았어
7	어반플랜트 합정	합정	합정동	카페 인테리어도 예쁘고 이름에 걸맞게 식물들이 많아서 좋았습니다. 특히 건물 외부에도 자리를 많이 만들어두어서 날이 좋을 때에 이용하면 좋을 것 같아요. 브런치 메뉴의 금액도
8	포비 베이직	합정	합정동	이제 서울에 꽤나 많은 포비 매장이 생겼지만 역시나 최고는 포비 베이직 합정 매장입니다!!! 통유리 건물 을 통째로 사용하고 있어 건물 외관도 예쁘고, ...
9	포비 베이직	합정	합정동	요기 빛도 너무 좋고 인테리어도 이쁘구 참새도 있음!!! 참새 너무 익숙하게 빵 얻어먹는 ㅋㅋㅋㅋㅋ 커피 맛있구, 러스크 가격 좀 사악하지만 정말 매우 흔하고 아는 그맛인데 돈주
10	포비 베이직	합정	합정동	주말이나 저녁에 가보다가 평일 낮에 방문하니 너무 좋더라고요! 그것도 봄!! 여유 있고 정말 ~ 커피는 말할 것도 없이 맛있구~ 피클로를 좋아합니다. 베이글도 크림치즈 없이도 넘
11	어썸마운틴	합정	동교동	500m에서 등반 후 먹는 흑임자치즈케이크와 아아 입니다 아이스아메리카노가 산미 없고 고소하고 묵직한 맛이 납니다 그리고 흑임자치즈케이크 산모양으로 겉모습은 퍼석할거 같
12	어썸마운틴	합정	동교동	카페가 이쁘고 감각적입니다! 말차파운드는 못먹었지만 얼그레이도 맛있었어요! 토마토바질에이드 기대했는데 생각이랑 다른맛이었지만 좋았어요!! 바질 좋아해서 진짜 괜찮았어
13	어썸마운틴	합정	동교동	맨날 강아지랑 산책할때마다 지나가는데 사람이 엄청 많습니다 처음에 생겼을때는 사람이 많지 않아서 강아지랑 지나가다가 들어오고 그랬었는데 이제는... 너무나 핫플이 되어버
14	943 킹스크로스CAFE	합정	서교동	해리포터 매니아라면 너무나 좋아할 카페입니다 음료도 비싸지만 맛있고 구경하기 재밌어요~
15	944 킹스크로스CAFE	합정	서교동	해리포터를 좋아하는 사람과 함께하면 더 좋은 곳!! 해리포터 음악이 계속 나와서 좋았어요.
16	945 킹스크로스CAFE	합정	서교동	홍대에서 어떤 카페를 갈까 친구들과 고민을 하던 중 해리포터 카페로 유명해서 선택하게 되었습니다. 저도 다른 사람들과 다르지않게 리뷰를 검색하고 결정을 하는 편인데 맛에 더
17	샌드커피 논탄토 연남	합정	동교동?	카이막이 먹고 싶어서 찾아갔는데ㅠㅠ 저녁에 갔더니 품절이더라고요!! 아쉽긴 했지만 카페가 조용하니 좋아서 커피라도 마시고 왔습니다! 체즈베는 처음 접해보는데 생각보다 먹
18	샌드커피 논탄토 연남	합정	동교동?	존맛 너무 잘 먹었어요 토스트는 더 구워 달라고 할 걸 그랬어요! 더 바삭하게 좋아서골든크림라떼는 생각보다 라떼에서 산미가 느껴졌어요 제 취향은 아니었는데 그래도 카이막 넘
19	샌드커피 논탄토 연남	합정	동교동?	시원하고 쾌적하고 메뉴 신기 핸드드립 커피 과테말라 좋아요 다음엔 터키식 카이막도 먹어볼게요
20	클린	합정	서교동	갑자기 걷는데 비가내려서 들어왔어요. 빗소리 들으면서 맥주 한잔 마시고 있는데 넘나 좋아요??
21	클린	합정	서교동	플라워카페 너무 이쁘요 들어갈때 꽃향기도 좋고 2층도 있고 테라스도 있고 바나나주스 진짜 맛있었어요 딸기 케이크는 조금 더 달았으면 좋겠어요 굿굿
22	클린	합정	서교동	인절미 크루아상 대존맛탱구리구리ㅠ 향기나는 꽃들에 둘러싸여 디저트 먹으면서 기분전환 지대로 했어용?????
23	차곡파티세리	합정	서교동	에스프레소 맛집 원두 기름이 잘 어우러지는 신선한 원두로 갓내린 커피 초콜릿은 더 맛있다!!! 에스프레 소 좋아하시는 분의 추천으로 발견한 커피 맛...
24	차곡파티세리	합정	서교동	케이크가 정말 너무 맛있어요!!! 녹차쉬폰인데 녹차 크림이 정말 진하고 맛있더라고요! 진짜 케이크 완전 강추합니다!!
25	차곡파티세리	합정	서교동	음료도 맛있구 케이크가 넘 귀여워요~~ 셀카찍기 좋은 예쁜 곳^^
26	브런치가	합정	동교동	크로크햄 치즈와 치킨퀘사디아를 주문했습니다. 치킨 브리토를 먹으려고 했지만 품질이라고 브리토와 비슷한 퀘사디아를 추천해주셔서 먹었는데 맛있더군요.....!! 감자튀김도 많이
27	브런치가	합정	동교동	음식도 맛있고, 무엇보다 일찍 열어서 저희처럼 일찍 만나는 사람에게 딱이었어요~^^

How,

프로젝트 진행

Task1. 크롤링한 데이터 전처리

Goal : 정확한 처리를 위해서는 불필요한 단어 제거 및 띄어쓰기 수정

Task2. 분석한 정보를 통한 카페 추천

Goal : 문장끼리 유사도 비교를 코사인 유사도를 이용한다.

Task3. 제작된 설문지 내용을 이용하기

설문지 제작, 결과 csv로 받기

Task1. 크롤링한 데이터 전처리

Goal : 정확한 처리를 위해서는 불필요한 단어 제거 및 띄어쓰기 수정

<Re모듈을 이용한 한글 이외의 단어 제거>

re.compile('[^가-힣]')을 사용할 때 string으로 변환해줘야합니다.

```
1 def extract_word(text):
2     hangul = re.compile('[^가-힣]')
3     result = hangul.sub(' ', text)
4     return result
```

<명사, 동사, 형용사 기반으로 문장 추출>

```
1 for i, document in enumerate(df['evaluation']):
2     clean_words = []
3     for word in okt.pos(document, stem=True): #어간 추출
4         if word[1] in ['Noun', 'Verb', 'Adjective']: #명사, 동사, 형용사
5             clean_words.append(word[0])
6     # print(clean_words)
7     document = ' '.join(clean_words)
8     print(document)
9     df['evaluation'][i] = document
```


Task1. 크롤링한 데이터 전처리

Goal : 정확한 처리를 위해서는 불필요한 단어 제거 및 띄어쓰기 수정

한글 띄어쓰기 해결하기

PyKoSpacing!

한국어 띄어쓰기 패키지로 띄어쓰기가 되어있지 않은 문장을 띄어쓰기를 한 문장으로 변환해주는 패키지

```
1 # !pip install git+https://github.com/haven-jeon/PyKoSpacing.git
```

Py-Hanspell

네이버 한글 맞춤법 검사기를 바탕으로 만들어진 패키지로써 맞춤법을 고친 문장을 반환하도록 해줍니다.

```
[296] 1 from pykospadding import Spacing
      2 spacing = Spacing()
```

```
[297] 1 from hanspell import spell_checker
      2 for i in range(1,len(df)):
      3     spelled_sent = spell_checker.check(df['evaluation'][i])
      4     spell = spelled_sent.checked
      5     df['evaluation'][i] = spacing(spell)
      6     print(df['evaluation'][i])
```

커피와 디저트 모두 너무 맛있어요 가게 분위기도 원목 가구를 써서 그런지 단정하고 대화하기 좋았어요 의자는 참 무겁더라고요 집어넣고 가기엔 쉽지 않았네요 늦게까지 프렌치 토스트 소문대로 존 맛탱 자연에 둘러싸여 있는 듯한 색다른 기분도 느낄 수 있었고 시그니처 오믈렛 정말 맛 있었어요 스파게티류는 면류가 없어서 아쉬웠지만 음식 맛은 괜춘 무엇보다 분위기가 나 직원분이 테이블 리를 도와 주셔서 웨이팅을 하다 보면 이름을 불러주세요 음료도 맛있었고 구성비가 좋았던 것 같아 요 개인적으로 매장 내에 식물이 많아서 풀 냄새가 나 카페 인테리어도 예쁘고 이름에 걸 맞게 식물들이 많아서 좋았습니다 특히 건물 외부에도 자리를 많이 만들어 두어서 날이 좋을 때 매 이용하면 좋을 것 같아 요 브런치 메 이제 서울에 꽤나 많은 포비 매장이 생겼지만 역시나 최고는 포비 베이직 합정 매장입니다 통유리 건물을 통째로 사용하고 있어 건물 외관도 예쁘고 요기 빛도 너무 좋고 인테리어도 이쁘고 참새도 있음 참새 너무 익숙하게 빵 얼어먹는 커피 맛있고 러스크 가격 좀 사악하지만 정말 매우 흔하고 아는 그 맛인데 돈 주고 주말이나 저녁에 가보다가 평일 낮에 방문하니 너무 좋더라고요 그것도 봄 여유 있고 정말 커피는 말 할 것도 없이 맛있고 피콜로를 좋아합니다 베이글도 크림치즈 없이도 에서 등반 후 먹는 흑임자 치즈케이크와 아입니다 아이스 아메리카노가 산 미 없고 고소하고 묵직한 맛이 납니다 그리고 흑임자 치즈케이크 산 모양으로 걸 모습은 퍼석할 카페가 이쁘고 감각적입니다 말 차 파운드는 못 먹었지만 얼 그레이도 맛있었어요 토마토 바질 에이드 기대했는데 생각이랑 다른 맛이었지만 좋았어요 바질 좋아해서 진짜 맨날 감아지랑 산책할 때마다 지나가는데 사람이 엄청 많습니다 처음에 생겼을 때는 사람이 많지 않아서 감아지랑 지나가다가 들어오고 그랬었는데 이제는 너무나 핫 플이

Task2. 분석한 정보를 통한 카페 추천

Goal : 설문지에 들어온 정보를 이용해서 코사인 유사도를 통한 카페 추천

```
[ ] 1 from sklearn.metrics.pairwise import cosine_similarity
```

```
[ ] 1 cos_sim = cosine_similarity(features, features, dense_output=True)
```

외부이라는 단어가 부정적일때 어떻게 처리해야하는가?

```
[ ] 1 similarity_simple_pair = cosine_similarity(features[54] , features)
    2 print(similarity_simple_pair)
```

```
[[0.         0.19902117 0.         0.         0.         0.2478933
 0.         0.11503755 0.08121169 0.         0.08688439 0.14338644
 0.         0.         0.         0.05864385 0.         0.13193135
 0.         0.         0.         0.22641424 0.         0.
 0.         0.         0.         0.         0.12565542 0.
 0.         0.         0.         0.         0.         0.15269649
 0.         0.16827738 0.06098155 0.10457015 0.12500522 0.25242716]
```



```
1 # 유사도 높은 순으로 정렬하기
2 (-similarity_simple_pair).argsort()
```

설문지를 통해서,
원하는 카페 정보를
입력해주면?

Find your favorite cafe

원문지 설명

네이버 지도에서 어디역으로 찾고싶으세요?

☐ 압정역

☐ 용산역

☐ 부정구정역

섹션 1 다음 3 섹션(어떤 카페를 찾으세요?)으로 이동

3 중 2 섹션

찾고 싶은 카페가 어떤동(ex.서교동)이었으면 좋겠나요?

근처의 카페를 찾기를 원하시면 네이버 지도를 커서 위치를 한번 확인해보세요!

찾고 싶은 카페가 어떤동이었던 좋겠나요? *

☐ 동교동

☐ 서교동

☐ 압정동

섹션 2 다음 다음 섹션으로 진행하기

3 중 3 섹션

어떤 카페를 찾으세요?

ex) 따스한 햇빛 속에서 야외에서 마시고 싶습니다.

어떤 카페를 찾으세요? *

정문형 텍스트

당신의 이름을 알려주세요! *

단말형 텍스트

합정역, 햇빛 따스한 외부에서 커피 한잔

✓
0초



```
1 index = (-similarity_simple_pair).argsort()
2 print('{}님, 검색 문장 : "{}" 에 대한 카페 추천 시스템의 결과입니다\n'.format(df['who']
3 for i in range(1,6):
4     order = index[0][i]
5     print('당신을 위한 추천 카페는 "{}"'.format(df['cafe_name'][order]))
```

📄 eden님, 검색 문장 : "햇빛 따스하다 외부 커피 한잔" 에 대한 카페 추천 시스템의 결과입니다

당신을 위한 추천 카페는 "허밍벨라"
당신을 위한 추천 카페는 "빈브라더스 합정"
당신을 위한 추천 카페는 "어반플랜트 합정"
당신을 위한 추천 카페는 "차곡파티세리"
당신을 위한 추천 카페는 "umoae"