

2022-2 빅데이터 캡스톤 디자인

네이버 리뷰 기반 송파구 일식집 추천 서비스

정보통계학과 20180278 이민주

보건관리학과 20180178 이다희

CONTENTS

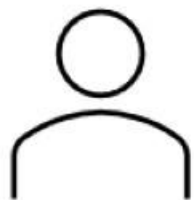
01 연구 선정 배경 및 필요성

02 콘텐츠 기반 필터링

03 협업 필터링

04 하이브리드 필터링

Summary



USER ID 및
USER 선호 정보 입력

협업 필터링을 이용 → 유저가 선호할 거 같은 음식점 10개 추천



컨텐츠 기반 필터링을 이용 → 유저가 선호할 거 같은 음식점과 비슷한 음식점을 각각 5개씩 추천



01.

연구 선정 배경 및 필요성

01 연구 선정 배경

왜 송파구인가?



인구수가 가장 많은 행정구역

서울시에서 주민등록 세대수가
가장 많은 행정구역,
서울시 인구의 6.9%가 송파구 거주



7000여개의 음식점

공공데이터 포털에 등록된
7125개 음식점



유원지

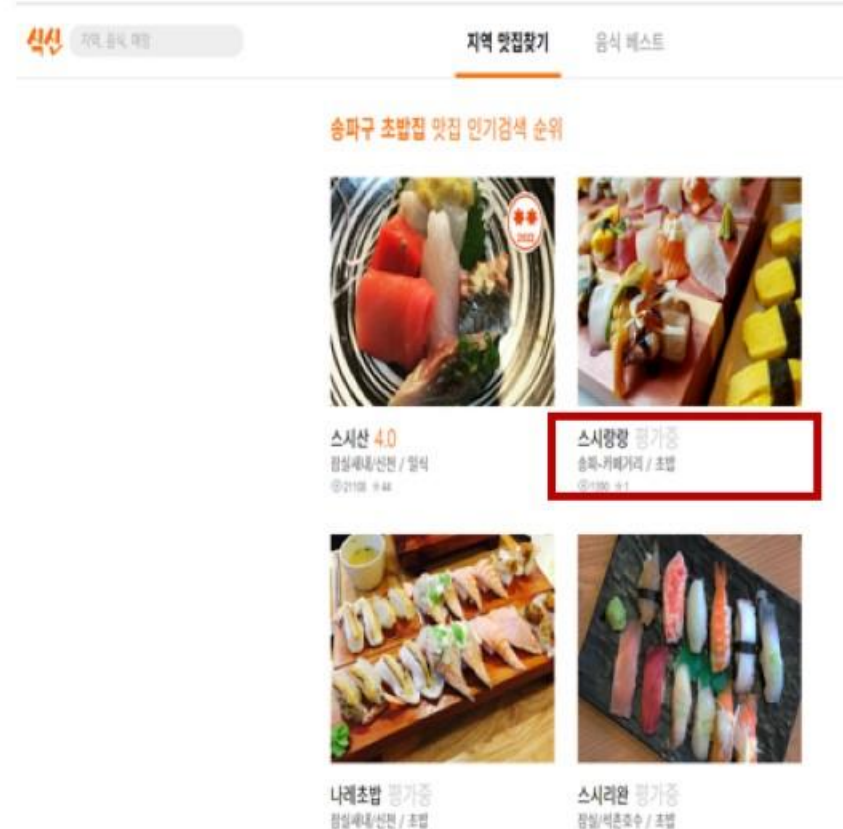
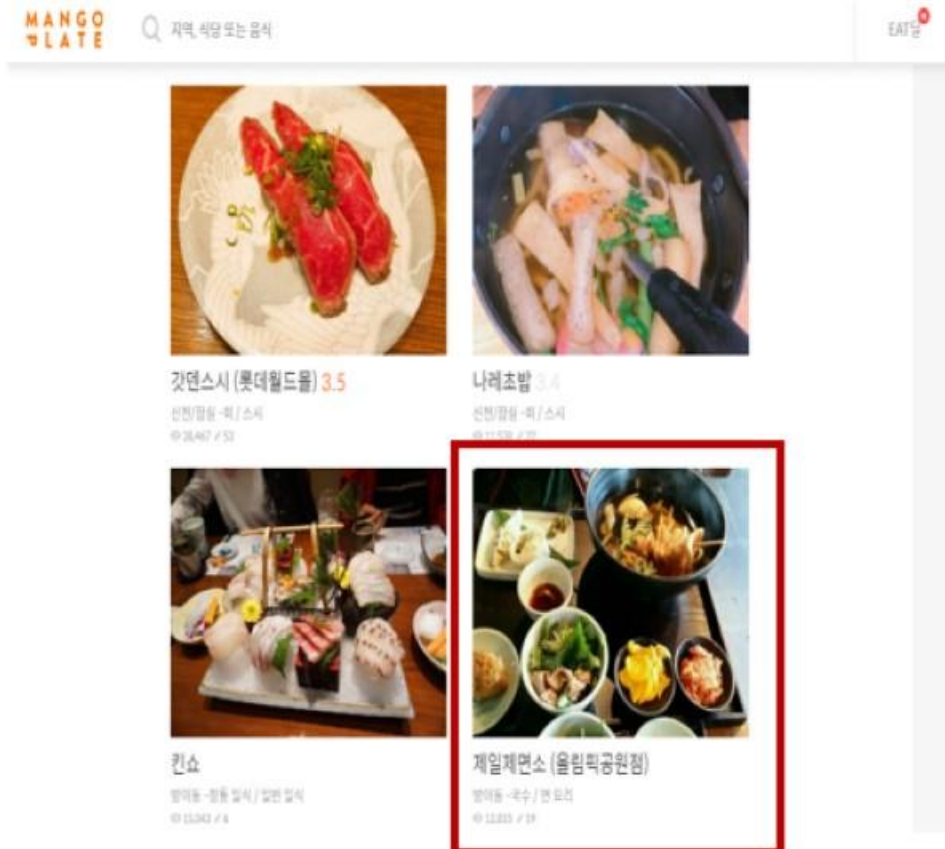
유동인구가 항상 많은
롯데월드, 롯데타워,
종합운동장, 올림픽공원

1-2. 연구의 필요성

기존 음식점 추천 시스템에 문제점이 존재

1. 망고 플레이트 / 식신의 경우

: '송파구 초밥집 검색' 결과 → 검색과 무관한 결과가 도출 / 다수의 식당에서 평점 리뷰가 없음

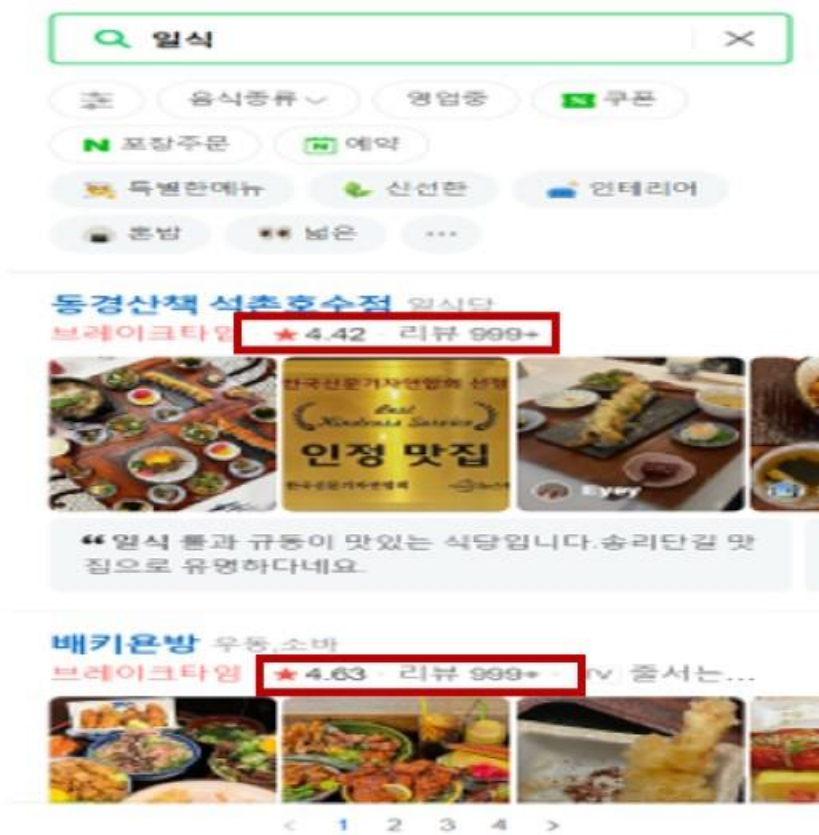
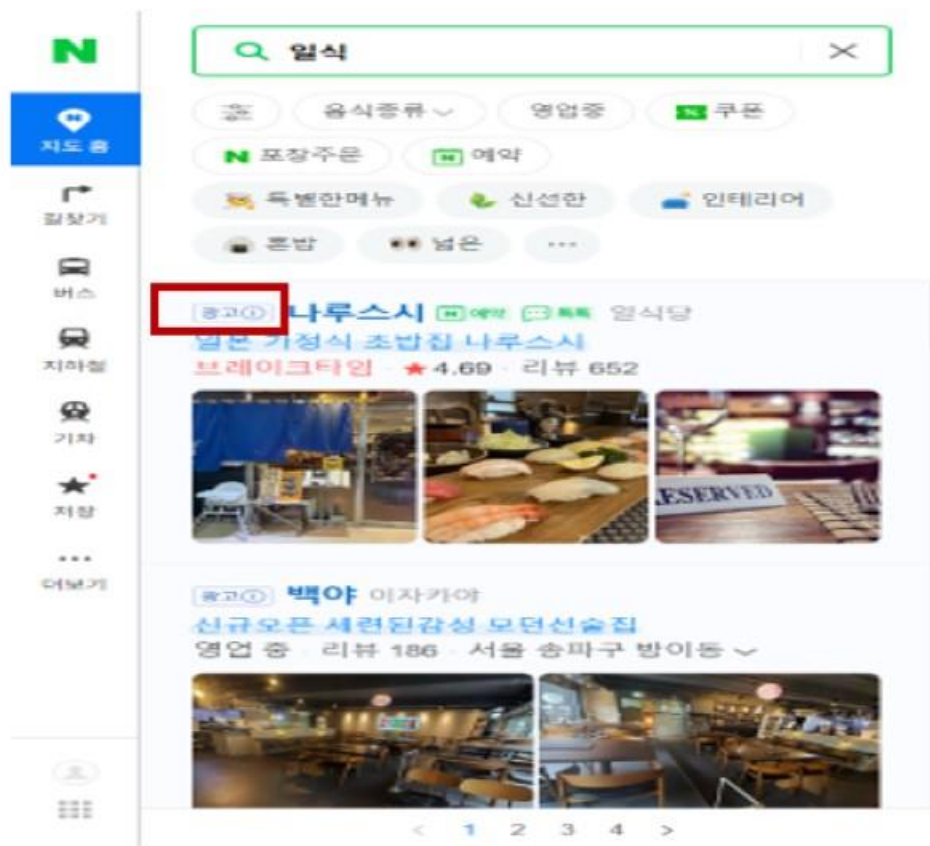


1-2. 연구의 필요성

기존 음식점 추천 시스템에 문제점이 존재

2. 네이버의 경우

: 광고하는 음식점 상위에 위치 / 좋은 별점이 아닌 리뷰가 많은 음식점으로 추천하는 경향



02.

컨텐츠 기반 필터링
















2-1. 데이터 수집

1. 공공데이터 포털 송파구 음식점 자료 7125개 데이터 활용

: 일식집과 무관한 식당, 배달전문점, 방문자 리뷰가 없는 곳 제외 -> 총 285개의 데이터 수집

2. 네이버 방문자 리뷰의 모든 카테고리를 $\text{항목} \div \text{평가횟수}$ 로 점수화하여 항목으로 추가

✓ 717회 (662명 참여)

| | | | | | |
|--|-----|---|-----|---|----|
|  "음식이 맛있어요" | 605 |  "가성비가 좋아요" | 215 |  "특별한 날 가기 좋아요" | 21 |
|  "재료가 신선해요" | 385 |  "인테리어가 멋져요" | 89 |  "매장이 넓어요" | 17 |
|  "친절해요" | 328 |  "양이 많아요" | 80 |  "화장실이 깨끗해요" | 7 |
|  "매장이 청결해요" | 265 |  "특별한 메뉴가 있어요" | 78 |  "뷰가 좋아요" | 6 |
|  "혼밥하기 좋아요" | 249 |  "주차하기 편해요" | 42 |  "단체모임 하기 좋아요" | 3 |

2-2. 전처리 및 EDA

1. 가격

: 가격대 라벨링 후 LabelEncoding하여 범주화

```
# 가격 label encoding
from sklearn.preprocessing import LabelEncoder
```

```
[ ] le = LabelEncoder()
le.fit(res['가격대'])
labels = le.transform(res['가격대'])
print('인코딩 결과:', labels)
```

```
인코딩 결과: [ 4 13  1 13  8  3 13 12 13 12 13 13 12 13 13  4  4  0 13  1 12  0 12  0
11  3  0  5  8 13  0 12 12  5  8 13 11  1  5  1  5 11 13  1  8  1  7  0
 7 13  0  0  0 13 10  4  4 13 12 13 12  1  2 13 13 13 12  2  5  2  9 13
 0 13 13 13  0 13 12  0 13 12 13 12  0  5 11  4  2 13  2  1 13 13  0 12
13  1  1  1 13  1  0  7 13 13  0  5  7  2  8  9  5 11  0 13  2 10  1  1
 5 10  4 11  2  6  0 12  1  8  1  4  0  6  4  2  0  3  4  6  8  3  6  9
 0  0  7 13  1  6 13 12  0  6  8  4 13 13 13 13 12 12 12 12 12  1  8 13
 6  0  2 11  2  1  9  4 12  4 12  5  5  1  2  6  0  8  0 13  1  0  0 11
 1  0  7 13  1  5 13  0 13  2  2 13 12  1  7  8  8  0  6  1  8 10  6  8
 3  7  9  3  0  1  9  5  9  1  0 13 12  0  0 13 13 13 13 13 13 13 13
12  0 13 13 13 13 13 13  0  0  0  3  0 13  0  7  3  8 13 13  0 13 13 13
13  1  0  0  0 12 13 13  0 13 13  6  3  9  0 12 13 13 13 13  0]
```

2. 별점

: StreamRating 적용

```
[ ] # 별점 stream rating
res['s_별점'] = res['별점'] - (res['별점'] - 4.0) *
    * np.power(2, -np.log10(res['전체리뷰수']))
```

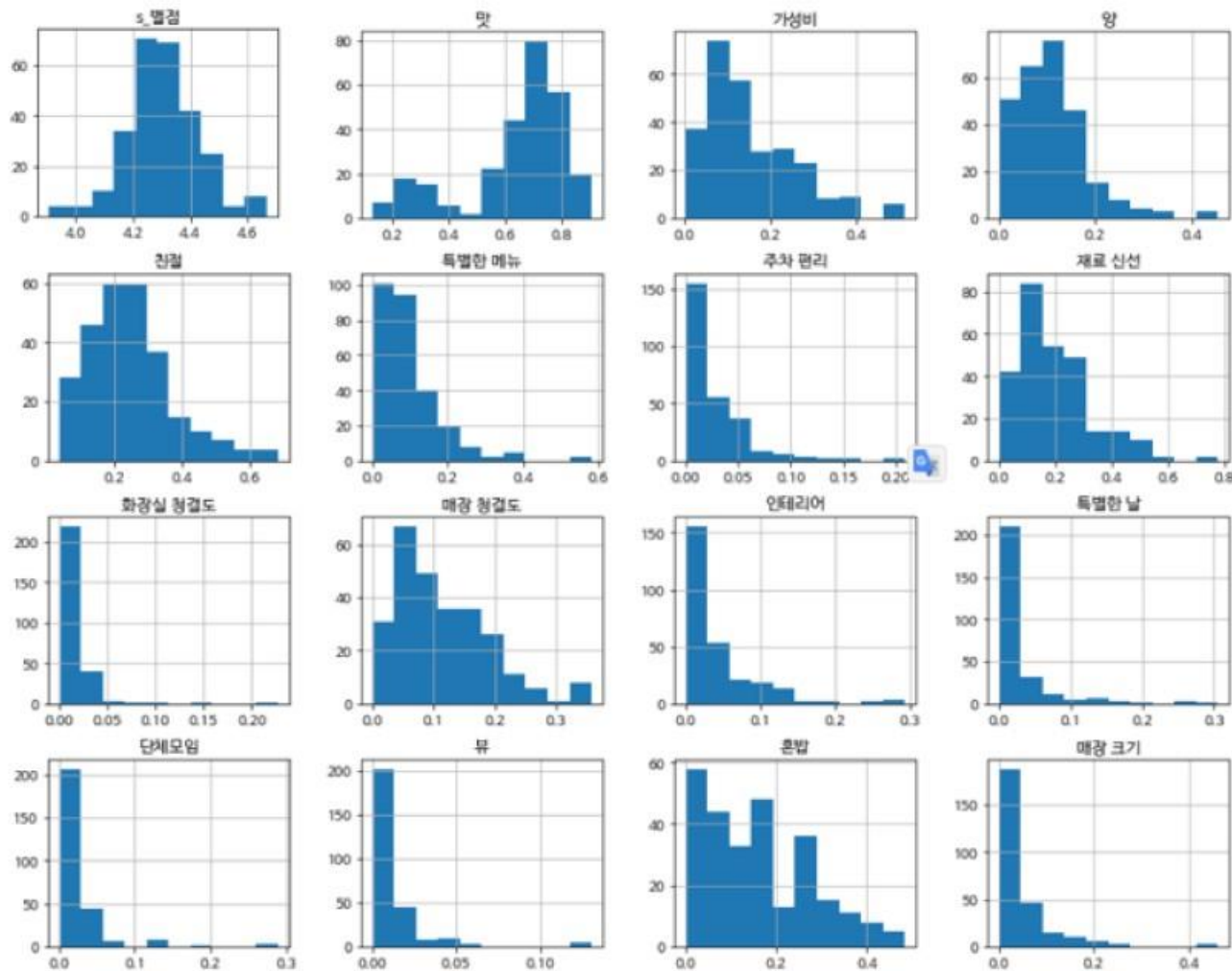
```
[ ] res
```

| | 업소명 | 위치 | 분류 | 가격 | 별점 | 전체리뷰수 | 가격대 | s_별점 |
|-----|--------------|-----|-------|-------|------|-------|-----|----------|
| 0 | 4.5평 우동집 | 오금동 | 우동,소바 | 4500 | 4.44 | 98 | 4 | 4.329329 |
| 1 | 가고야 | 잠실동 | 밥류 | 8000 | 4.36 | 33 | 13 | 4.234344 |
| 2 | 가락정초밥 | 가락동 | 초밥 | 12500 | 4.63 | 131 | 1 | 4.484796 |
| 3 | 가일 | 방이동 | 밥류 | 8000 | 4.53 | 145 | 13 | 4.411522 |
| 4 | 가일 | 방이동 | 사시미 | 38000 | 4.53 | 145 | 8 | 4.411522 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 280 | 호니도니 가락점 | 가락동 | 밥류 | 9500 | 4.31 | 218 | 13 | 4.248706 |
| 281 | 호또맨 | 문정동 | 라면 | 8500 | 4.37 | 242 | 13 | 4.299107 |
| 282 | 호또맨 | 문정동 | 밥류 | 8500 | 4.37 | 242 | 13 | 4.299107 |
| 283 | 홍대돈부리 롯데월드몰점 | 신천동 | 밥류 | 8000 | 4.45 | 938 | 13 | 4.392656 |
| 284 | 회랑 초밥이랑 | 송파동 | 초밥 | 12000 | 4.47 | 112 | 0 | 4.356441 |

285 rows × 8 columns

2-2. 전처리 및 EDA

3. EDA



1 # 가격대 파악

2 res['가격대'].value_counts()

| | |
|---------------------|----|
| 7500원 이상 10000원 미만 | 74 |
| 10000원 이상 12500원 미만 | 48 |
| 5000원 이상 7500원 미만 | 28 |
| 12500원 이상 15000원 미만 | 27 |
| 15000원 이상 17500원 미만 | 16 |
| 35000원 이상 40000원 미만 | 13 |
| 20000원 이상 25000원 미만 | 13 |
| 30000원 이상 35000원 미만 | 10 |
| 25000원 이상 30000원 미만 | 10 |
| 17500원 이상 20000원 미만 | 9 |
| 40000원 이상 45000원 미만 | 8 |
| 50000원 이상 | 6 |
| 1이상 5000원 미만 | 5 |
| 45000원 이상 50000원 미만 | 4 |

Name: 가격대, dtype: int64

최종 데이터셋

1 res

| | 업소명 | 위치 | 분류 | 가격 | 맛 | 가성비 | 양 | 친절 | 특별한 메뉴 | 주차 편리 | ... | 화장실 청결도 | 매장 청결도 | 인테리어 | 특별한 날 | 단체모임 | 뷰 | 혼밥 | 매장 크기 | 가격대 | s_별점 |
|-----|--------------|-----|-------|-------|----------|----------|----------|----------|----------|----------|-----|----------|----------|----------|----------|----------|----------|----------|----------|-----|----------|
| 0 | 4.5평 우동집 | 오금동 | 우동,소바 | 4500 | 0.750000 | 0.236111 | 0.069444 | 0.083333 | 0.138889 | 0.000000 | ... | 0.000000 | 0.055556 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.416667 | 0.000000 | 4 | 4.329329 |
| 1 | 가고야 | 잠실동 | 밥류 | 8000 | 0.625000 | 0.208333 | 0.125000 | 0.375000 | 0.041667 | 0.208333 | ... | 0.000000 | 0.125000 | 0.041667 | 0.041667 | 0.000000 | 0.000000 | 0.333333 | 0.000000 | 13 | 4.234344 |
| 2 | 가락정초밥 | 가락동 | 초밥 | 12500 | 0.851796 | 0.290419 | 0.110778 | 0.447605 | 0.103293 | 0.061377 | ... | 0.010479 | 0.357784 | 0.122754 | 0.026946 | 0.002994 | 0.008982 | 0.336826 | 0.023952 | 1 | 4.484796 |
| 3 | 가월 | 방이동 | 밥류 | 8000 | 0.863795 | 0.129220 | 0.169965 | 0.435390 | 0.142026 | 0.009313 | ... | 0.019790 | 0.245634 | 0.292200 | 0.064028 | 0.142026 | 0.023283 | 0.012806 | 0.461001 | 13 | 4.411522 |
| 4 | 가월 | 방이동 | 사시미 | 38000 | 0.863795 | 0.129220 | 0.169965 | 0.435390 | 0.142026 | 0.009313 | ... | 0.019790 | 0.245634 | 0.292200 | 0.064028 | 0.142026 | 0.023283 | 0.012806 | 0.461001 | 8 | 4.411522 |
| ... | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| 266 | 호니도니 가락점 | 가락동 | 밥류 | 9500 | 0.678082 | 0.102740 | 0.164384 | 0.116438 | 0.054795 | 0.000000 | ... | 0.020548 | 0.047945 | 0.006849 | 0.006849 | 0.006849 | 0.000000 | 0.171233 | 0.034247 | 13 | 4.248706 |
| 267 | 호도면 | 문정동 | 라면 | 8500 | 0.806723 | 0.067227 | 0.109244 | 0.210084 | 0.067227 | 0.008403 | ... | 0.000000 | 0.042017 | 0.025210 | 0.008403 | 0.008403 | 0.016807 | 0.268908 | 0.008403 | 13 | 4.299107 |
| 268 | 호도면 | 문정동 | 밥류 | 8500 | 0.806723 | 0.067227 | 0.109244 | 0.210084 | 0.067227 | 0.008403 | ... | 0.000000 | 0.042017 | 0.025210 | 0.008403 | 0.008403 | 0.016807 | 0.268908 | 0.008403 | 13 | 4.299107 |
| 269 | 홍대돈부리 롯데월드몰점 | 신전동 | 밥류 | 8000 | 0.730496 | 0.127660 | 0.127660 | 0.195035 | 0.081560 | 0.023050 | ... | 0.023050 | 0.054965 | 0.005319 | 0.017730 | 0.003546 | 0.053191 | 0.200355 | 0.021277 | 13 | 4.392656 |
| 270 | 희랑 초밥이랑 | 송파동 | 초밥 | 12000 | 0.644737 | 0.236842 | 0.065789 | 0.223684 | 0.026316 | 0.013158 | ... | 0.000000 | 0.092105 | 0.000000 | 0.013158 | 0.000000 | 0.000000 | 0.105263 | 0.013158 | 0 | 4.356441 |

271 rows x 21 columns

```
1 print(res.shape)
2 res.info()
```

```
(271, 21)
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 271 entries, 0 to 270
Data columns (total 21 columns):
#   Column      Non-Null Count  Dtype
---  -
0   업소명      271 non-null    object
1   위치        271 non-null    object
2   분류        271 non-null    object
3   가격        271 non-null    int64
4   맛          271 non-null    float64
5   가성비비     271 non-null    float64
6   양          271 non-null    float64
7   친절        271 non-null    float64
8   특별한 메뉴 271 non-null    float64
9   주차 편리   271 non-null    float64
10  재료 신선   271 non-null    float64
11  화장실 청결도 271 non-null    float64
12  매장 청결도 271 non-null    float64
13  인테리어    271 non-null    float64
14  특별한 날    271 non-null    float64
15  단체모임     271 non-null    float64
16  뷰          271 non-null    float64
17  혼밥        271 non-null    float64
18  매장 크기    271 non-null    float64
19  가격대       271 non-null    int64
20  s_별점      271 non-null    float64
dtypes: float64(16), int64(2), object(3)
memory usage: 44.6+ KB
```

2-3. 추천시스템

| | 기본 추천시스템 | 항목에 가중치 부여한 추천시스템 |
|------|-------------------|--------------------|
| NDCG | 0.795527607422254 | 0.8552332068423639 |

→ NDCG 값이 더 높기에 <항목에 가중치를 부여한 추천시스템>을 추천 방법으로 선정

2-3. 추천시스템

1. Input 값을 받아 사용자가 원하는 음식 종류만 추천 받을 수 있도록 함 → 음식 종류 초밥으로 한정

```
1 print('다음 중 원하는 음식 종류를 선택해주세요')
2 print('사시미 / 초밥 / 밥류(덮밥, 카레, 탕동 등) / 우동,소바 / 가스류 / 라면')
3
4
5 print('~'*5)
6 print(' ')
7
8 i =input('원하는 음식 종류는 무엇인가요? ',)
9 print(i,'으로 선택되었습니다')
```

다음 중 원하는 음식 종류를 선택해주세요
사시미 / 초밥 / 밥류(덮밥, 카레, 탕동 등) / 우동,소바 / 가스류 / 라면

원하는 음식 종류는 무엇인가요? 초밥
초밥 으로 선택되었습니다

```
1 res_cat = res[res['분류']==i] # 초밥집만 추출한 데이터프레임
```

```
1 res_cat.reset_index(drop=True, inplace= True) # 인덱스 재지정
```

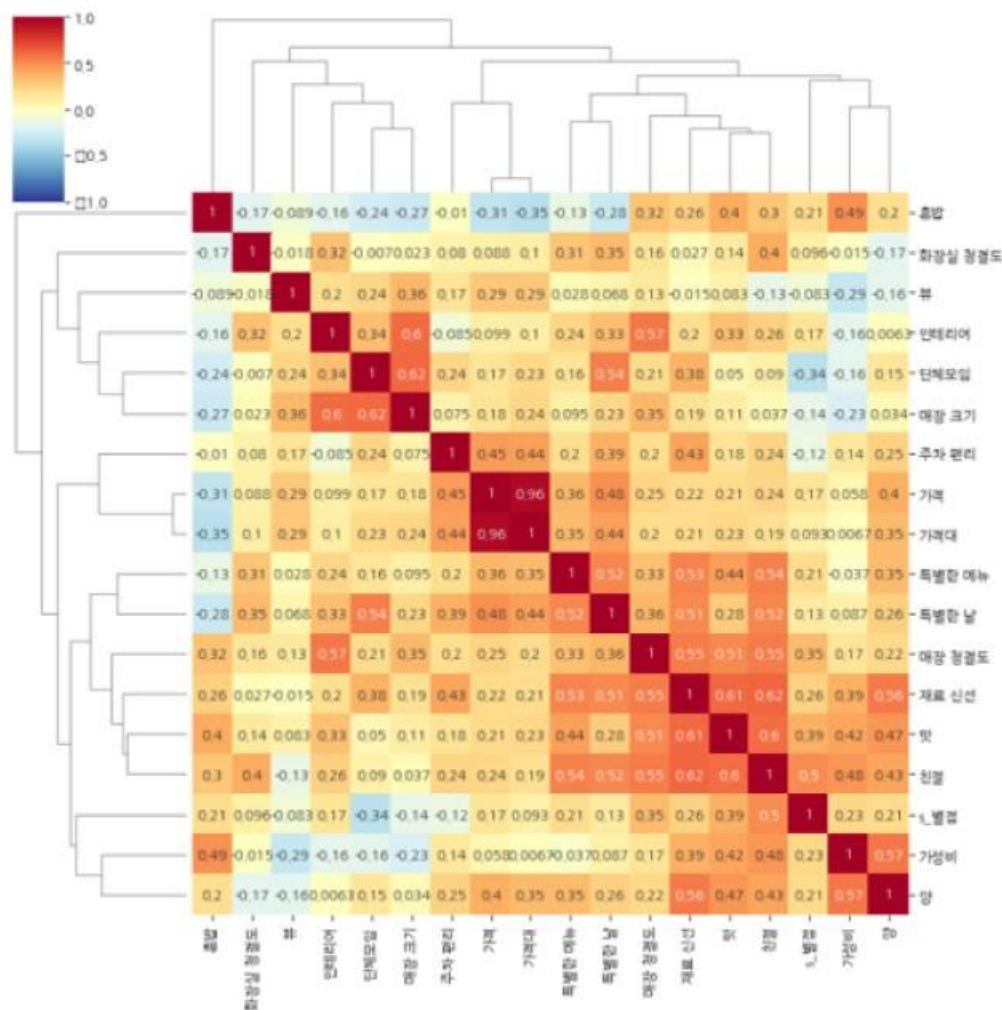
```
1 print(res_cat.shape)
2 res_cat.head(5)
```

(57, 21)

| | 업소명 | 위치 | 분류 | 가격 | 맛 | 가성비 | 양 | 친절 | 특별한 메뉴 | 주차 편의 | ... | 화장실 청결도 | 매장 청결도 | 인테리어 | 특별한 날 | 단체모임 | 뷰 | 혼밥 | 매장 크기 | 가격대 | s_별점 | |
|---|-------------|-----|----|-------|----------|----------|----------|----------|----------|----------|-----|----------|----------|----------|----------|----------|----------|----------|----------|-----|----------|--|
| 0 | 가락정초밥 | 가락동 | 초밥 | 12500 | 0.851796 | 0.290419 | 0.110778 | 0.447605 | 0.103293 | 0.061377 | ... | 0.010479 | 0.357784 | 0.122754 | 0.026946 | 0.002994 | 0.008982 | 0.336826 | 0.023952 | 1 | 4.484796 | |
| 1 | 가일 | 방이동 | 초밥 | 18000 | 0.863795 | 0.129220 | 0.169965 | 0.435390 | 0.142026 | 0.009313 | ... | 0.019790 | 0.245634 | 0.292200 | 0.064028 | 0.142026 | 0.023283 | 0.012806 | 0.461001 | 3 | 4.411522 | |
| 2 | 기맨정스시 문정동본점 | 문정동 | 초밥 | 11000 | 0.702479 | 0.239669 | 0.049587 | 0.297521 | 0.033058 | 0.000000 | ... | 0.000000 | 0.165289 | 0.008264 | 0.008264 | 0.033058 | 0.024793 | 0.206612 | 0.082645 | 0 | 4.182527 | |
| 3 | 김태완 스시 송파본점 | 송파동 | 초밥 | 21000 | 0.730964 | 0.277496 | 0.120135 | 0.336717 | 0.128596 | 0.050761 | ... | 0.023689 | 0.184433 | 0.027073 | 0.043993 | 0.023689 | 0.013536 | 0.138748 | 0.153976 | 5 | 4.259189 | |
| 4 | 나레초밥 거여본점 | 거여동 | 초밥 | 13000 | 0.716390 | 0.226519 | 0.090239 | 0.189687 | 0.042357 | 0.034991 | ... | 0.025783 | 0.156538 | 0.047882 | 0.038674 | 0.011050 | 0.005525 | 0.093923 | 0.092081 | 1 | 4.304823 | |

2-3. 추천시스템

2. 가중치 부여를 위한 가중치 도출 → 별점과 항목 간의 상관관계 도출 후 값에 1을 더해 가중치 설정 ★



```
1 # 가중치 부여
2
3 # 상관관계으로 나온 값에 +1
4 # 차별성을 두기 위해서 소수둘째자리까지 곱함
5
6 weight_df['맛'] = weight_df['맛']*1.39
7 weight_df['가성비'] = weight_df['가성비']*1.23
8 weight_df['양'] = weight_df['양']*1.21
9 weight_df['친절'] = weight_df['친절']*1.5
10 weight_df['특별한 메뉴'] = weight_df['특별한 메뉴']*1.21
11 weight_df['주차 편의'] = weight_df['주차 편의']*1.12
12 weight_df['재료 신선'] = weight_df['재료 신선']*1.26
13 weight_df['화장실 청결도'] = weight_df['화장실 청결도']*1.1
14 weight_df['매장 청결도'] = weight_df['매장 청결도']*1.35
15 weight_df['인테리어'] = weight_df['인테리어']*1.17
16 weight_df['특별한 날'] = weight_df['특별한 날']*1.13
17 weight_df['단체모임'] = weight_df['단체모임']*1.34
18 weight_df['뷰'] = weight_df['뷰']*1.08
19 weight_df['혼밥'] = weight_df['혼밥']*1.21
20 weight_df['매장 크기'] = weight_df['매장 크기']*1.14
```


2-3. 추천시스템

3. 가중치를 부여한 항목들을 더한 total 컬럼 생성

```
1 # 각 요소의 점수를 더한 total 컬럼 생성
```

2

```
3 weight df['total'] = weight df['맛'] + weight df['가성비'] + weight df['양'] + weight df['치결'] + weight df['특별한 메뉴'] + weight df['주차 편리'] + weight df['재료 신선'] + weight df['화장실 청결도'] +
```

```
weight_df['매장 청결도'] + weight_df['인테리어'] + weight_df['특별화 단'] + weight_df['단체모임'] + weight_df['뷰'] + weight_df['온도'] + weight_df['매장 크기']
```

```
1 weight_df.head(10)
```

| | 업소명 | 가격 | 가격대 | 맛 | 가성비 | 양 | 친절 | 특별한 메뉴 | 주차 | 편리 | 재료 신선 | 화장실 청결도 | 매장 청결도 | 인테리어 | 특별한 날 | 단체모임 | 뷰 | 혼밥 | 매장 크기 | s_별점 | total |
|---|-------------|-------|-----|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|-------|
| 0 | 가락정초밥 | 12500 | 1 | 1.183997 | 0.357216 | 0.134042 | 0.671407 | 0.124985 | 0.068743 | 0.684701 | 0.011527 | 0.483009 | 0.143623 | 0.030449 | 0.004012 | 0.009701 | 0.407560 | 0.027305 | 4.484796 | 4.342275 | |
| 1 | 가일 | 18000 | 3 | 1.200675 | 0.158941 | 0.205658 | 0.653085 | 0.171851 | 0.010431 | 0.661537 | 0.021769 | 0.331607 | 0.341874 | 0.072352 | 0.190314 | 0.025146 | 0.015495 | 0.525541 | 4.411522 | 4.586275 | |
| 2 | 기맨정스시 문정동본점 | 11000 | 0 | 0.976446 | 0.294793 | 0.060000 | 0.446281 | 0.040000 | 0.000000 | 0.260331 | 0.000000 | 0.223140 | 0.009669 | 0.009339 | 0.044298 | 0.026777 | 0.250000 | 0.094215 | 4.182527 | 2.735289 | |
| 3 | 김태완 스시 송파본점 | 21000 | 5 | 1.016041 | 0.341320 | 0.145364 | 0.505076 | 0.155601 | 0.056853 | 0.439188 | 0.026058 | 0.248985 | 0.031675 | 0.049712 | 0.031743 | 0.014619 | 0.167885 | 0.175533 | 4.259189 | 3.405651 | |
| 4 | 나레초밥 거여본점 | 13000 | 1 | 0.995783 | 0.278619 | 0.109190 | 0.284530 | 0.051252 | 0.039190 | 0.343425 | 0.028361 | 0.211326 | 0.056022 | 0.043702 | 0.014807 | 0.005967 | 0.113646 | 0.104972 | 4.304823 | 2.680792 | |
| 5 | 나레초밥 잠실새내점 | 13000 | 1 | 1.021090 | 0.244834 | 0.097488 | 0.234597 | 0.068815 | 0.026540 | 0.280664 | 0.000000 | 0.095972 | 0.022180 | 0.010711 | 0.019052 | 0.005118 | 0.223649 | 0.037820 | 4.318588 | 2.388531 | |
| 6 | 나루스시 | 50000 | 11 | 1.137273 | 0.458840 | 0.356552 | 0.963950 | 0.276897 | 0.210658 | 0.679373 | 0.013793 | 0.342790 | 0.055016 | 0.286928 | 0.046207 | 0.003386 | 0.022759 | 0.028589 | 4.535755 | 4.883009 | |
| 7 | 난 스시 | 13000 | 1 | 1.022059 | 0.108529 | 0.142353 | 0.264706 | 0.035588 | 0.000000 | 0.333529 | 0.000000 | 0.079412 | 0.034412 | 0.033235 | 0.039412 | 0.000000 | 0.071176 | 0.000000 | 4.324758 | 2.164412 | |
| 8 | 낭만 스시 | 13000 | 1 | 1.015769 | 0.416308 | 0.176846 | 0.403846 | 0.093077 | 0.025846 | 0.484615 | 0.000000 | 0.384231 | 0.099000 | 0.017385 | 0.061846 | 0.008308 | 0.316462 | 0.052615 | 4.231170 | 3.556154 | |
| 9 | 동경산책 석촌호수점 | 13000 | 1 | 1.044240 | 0.065710 | 0.092922 | 0.365609 | 0.240384 | 0.003740 | 0.183005 | 0.009182 | 0.254674 | 0.179699 | 0.062254 | 0.015659 | 0.014424 | 0.066661 | 0.049482 | 4.359550 | 2.647646 | |

2-3. 추천시스템

4. 항목을 모두 반영하여 코사인 유사도 계산

NDCG 결과 확인을 위해 1차적으로 음식점을 '스시센세이' 로 선정

```
1 from sklearn.metrics.pairwise import cosine_similarity  
  
1 c_sim2 = cosine_similarity(df2, df2)  
2 print(c_sim2.shape)  
3 print(c_sim2[:2])
```

(57, 57)

```
[[1. 0.95639174 0.96878012 0.83489578 0.97519693 0.96262033  
0.65192998 0.95089342 0.99641995 0.97059953 0.97197629 0.96725047  
0.77619668 0.97462367 0.97102391 0.77635715 0.96172429 0.90195782  
0.96123257 0.75529334 0.9672048 0.97948543 0.98137023 0.92988532  
0.979984 0.96958899 0.94281747 0.95697701 0.97983856 0.95714718  
0.87079681 0.85887218 0.81671205 0.87108674 0.90940854 0.8026328  
0.83166687 0.83831742 0.85493301 0.97256046 0.97971741 0.95345025  
0.92078475 0.63905639 0.9742706 0.92711483 0.95425572 0.94932506  
0.94649107 0.80031248 0.62605753 0.99893316 0.8795732 0.94242024  
0.95876185 0.9246333 0.95062026]  
[0.95639174 1. 0.88234337 0.947093 0.94137148 0.92810836  
0.83145955 0.91936516 0.95742941 0.93817965 0.97544332 0.97969744  
0.90303865 0.94041733 0.97922796 0.90887066 0.97396748 0.94325701  
0.92906715 0.88586929 0.97470769 0.8924725 0.94687205 0.90040285  
0.89328373 0.97848218 0.96341074 0.99134788 0.89309964 0.86988799  
0.84688878 0.7735132 0.79816931 0.89713722 0.93298729 0.7826106  
0.85992385 0.7541359 0.77226708 0.93614382 0.89373479 0.96661347  
0.89230438 0.81753679 0.93752014 0.98134827 0.9926856 0.86658197  
0.96721601 0.9171781 0.80725923 0.95679234 0.94489401 0.9945289  
0.92292182 0.97844219 0.86346953]]
```

```
1 df2_sim_sorted_ind = c_sim2.argsort()[::-1]  
2 print(df2_sim_sorted_ind[:1])
```

```
[[ 0 51  0 17 22 24 40 21 20  4 13 44 39 10 14  9 25  2 11 20  5 16 18 54  
29 27  1 46 41  7 56 47 48 26 53 23 45 55 42 34 52 33 30 31 38 37  3 36  
32 35 49 15 12 19  6 43 50]]
```

```
1 def find_sim_res(weight_df, sorted_ind, res_name, top_n=10):  
2  
3     res_name = weight_df[weight_df['업소명'] == res_name]  
4  
5     res_index = res_name.index.values  
6     similar_indexes = sorted_ind[res_index, :(top_n)]  
7  
8     print(similar_indexes)  
9     similar_indexes = similar_indexes.reshape(-1)  
10  
11     return weight_df.iloc[similar_indexes]
```

```
1 similar_restaurant2 = find_sim_res(weight_df, df2_sim_sorted_ind, '스시센세이', 20) # NDCG 결과 확인용으로 1차적으로 스시센세이로 음식점 선정  
2  
3 res_cat_sim = weight_df[weight_df['업소명'] == '스시센세이']  
4 res_cat_sim[['업소명', '가격', '맛', '가성비', '양', '친절', '특별한 메뉴', '주차 관리', '재료 신선', '화장실 청결도', '매장 청결도', '인테리어', '특별한 날', '단체모임', '뷰', '혼밥', '매장 크기', '가격대', 's_별점', 'total']]
```

```
[[25 14 11 16 20 41 10 26 48 22 45 17  4 13 44 55 53  8  9  5]]
```

| | 업소명 | 가격 | 맛 | 가성비 | 양 | 친절 | 특별한 메뉴 | 주차 관리 | 재료 신선 | 화장실 청결도 | 매장 청결도 | 인테리어 | 특별한 날 | 단체모임 | 뷰 | 혼밥 | 매장 크기 | 가격대 | s_별점 | total |
|----|-------|-------|--------|--------|-------|-----|--------|-------|-------|---------|--------|--------|-------|-------|-------|-------|-------|-----|----------|--------|
| 25 | 스시센세이 | 15000 | 1.1815 | 0.0615 | 0.121 | 0.3 | 0.0605 | 0.056 | 0.567 | 0.0 | 0.2025 | 0.0585 | 0.0 | 0.067 | 0.054 | 0.363 | 0.057 | 2 | 4.310399 | 3.1495 |

2-3. 추천시스템

5. Total 내림차순 sorting으로 음식점 추천

```

1 similar_restaurant2 = similar_restaurant2[['업소명', '가격', '맛', '가성비', '양', '친절', '특별한 메뉴', '주차 편리', '재료 신선', '화장실 청결도', '매장 청결도', '인테리어', '특별한 날', '단체모임', '뷰', '온밥', '매장 크기', '가격대', 's_별점', 'total']]
2
3 # 가중치(total) 기준 sorting
4 similar_restaurant2 = similar_restaurant2.sort_values(by=['total'], axis=0, ascending=False)
5
6 # 2차 추천 결과
7 similar_restaurant2

```

| | 업소명 | 가격 | 맛 | 가성비 | 양 | 친절 | 특별한 메뉴 | 주차 편리 | 재료 신선 | 화장실 청결도 | 매장 청결도 | 인테리어 | 특별한 날 | 단체모임 | 뷰 | 온밥 | 매장 크기 | 가격대 | s_별점 | total |
|----|------------|-------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|-----|----------|----------|
| 53 | 태인일식 | 19000 | 0.904337 | 0.118554 | 0.102048 | 0.542169 | 0.145783 | 0.053976 | 0.637590 | 0.026506 | 0.227711 | 0.126867 | 0.299518 | 0.387470 | 0.039036 | 0.043735 | 0.192289 | 3 | 4.164493 | 3.847590 |
| 8 | 낭만 스시 | 13000 | 1.015769 | 0.416308 | 0.176846 | 0.403846 | 0.093077 | 0.025846 | 0.484615 | 0.000000 | 0.384231 | 0.099000 | 0.017385 | 0.061846 | 0.008308 | 0.316462 | 0.052615 | 1 | 4.231170 | 3.556154 |
| 10 | 롤스시 | 16000 | 1.028961 | 0.287532 | 0.094286 | 0.915584 | 0.031429 | 0.000000 | 0.294545 | 0.028571 | 0.245455 | 0.015195 | 0.029351 | 0.000000 | 0.014026 | 0.298571 | 0.014805 | 2 | 4.505346 | 3.298312 |
| 14 | 서가참치&스시 | 15000 | 1.184659 | 0.195682 | 0.275000 | 0.443182 | 0.137500 | 0.012727 | 0.458182 | 0.000000 | 0.168750 | 0.026591 | 0.025682 | 0.015227 | 0.000000 | 0.220000 | 0.038864 | 2 | 4.379537 | 3.202045 |
| 25 | 스시센세이 | 15000 | 1.181500 | 0.061500 | 0.121000 | 0.300000 | 0.060500 | 0.056000 | 0.567000 | 0.000000 | 0.202500 | 0.058500 | 0.000000 | 0.067000 | 0.054000 | 0.363000 | 0.057000 | 2 | 4.310399 | 3.149500 |
| 45 | 참치의미 | 19000 | 0.810833 | 0.085417 | 0.134444 | 0.479167 | 0.336111 | 0.062222 | 0.420000 | 0.045833 | 0.243750 | 0.048750 | 0.125556 | 0.093056 | 0.045000 | 0.033611 | 0.174167 | 3 | 4.399872 | 3.137917 |
| 20 | 스시항랑 | 15000 | 1.180977 | 0.443910 | 0.127368 | 0.552632 | 0.118271 | 0.000000 | 0.341053 | 0.008271 | 0.101504 | 0.035188 | 0.033985 | 0.010075 | 0.000000 | 0.127368 | 0.025714 | 2 | 4.442048 | 3.106316 |
| 11 | 분장지 | 15000 | 1.070815 | 0.291556 | 0.170296 | 0.366667 | 0.170296 | 0.024889 | 0.494667 | 0.024444 | 0.200000 | 0.043333 | 0.066963 | 0.049630 | 0.000000 | 0.044815 | 0.033778 | 2 | 4.257578 | 3.052148 |
| 55 | 해조회초밥 | 18000 | 0.787667 | 0.328000 | 0.282333 | 0.550000 | 0.161333 | 0.037333 | 0.546000 | 0.000000 | 0.135000 | 0.000000 | 0.000000 | 0.089333 | 0.000000 | 0.040333 | 0.076000 | 3 | 4.331255 | 3.033333 |
| 17 | 스시 오하요 | 14000 | 1.019333 | 0.283273 | 0.066000 | 0.372727 | 0.110000 | 0.006788 | 0.526909 | 0.006667 | 0.229091 | 0.056727 | 0.047939 | 0.008121 | 0.013091 | 0.198000 | 0.006909 | 1 | 4.484094 | 2.951576 |
| 16 | 스시 153 | 15000 | 1.005772 | 0.270000 | 0.137724 | 0.414634 | 0.078699 | 0.081951 | 0.399512 | 0.044715 | 0.131707 | 0.019024 | 0.009187 | 0.032683 | 0.026341 | 0.177073 | 0.027805 | 2 | 4.279855 | 2.856829 |
| 22 | 스시리완 | 14000 | 0.840864 | 0.334074 | 0.119506 | 0.388889 | 0.119506 | 0.041481 | 0.311111 | 0.040741 | 0.183333 | 0.000000 | 0.013951 | 0.066173 | 0.026667 | 0.224074 | 0.084444 | 1 | 4.184824 | 2.794815 |
| 41 | 젠 | 16000 | 1.107971 | 0.071304 | 0.146135 | 0.463768 | 0.175362 | 0.005411 | 0.225217 | 0.005314 | 0.221739 | 0.141304 | 0.060048 | 0.006473 | 0.020870 | 0.058454 | 0.011014 | 2 | 4.437671 | 2.720386 |
| 4 | 나레초밥 거여본점 | 13000 | 0.995783 | 0.278619 | 0.109190 | 0.284530 | 0.051252 | 0.039190 | 0.343425 | 0.028361 | 0.211326 | 0.056022 | 0.043702 | 0.014807 | 0.005967 | 0.113646 | 0.104972 | 1 | 4.304823 | 2.680792 |
| 44 | 진미 참치 | 13000 | 0.866706 | 0.072353 | 0.085412 | 0.441176 | 0.213529 | 0.039529 | 0.340941 | 0.012941 | 0.238235 | 0.000000 | 0.000000 | 0.047294 | 0.012706 | 0.298941 | 0.000000 | 1 | 4.364944 | 2.669765 |
| 13 | 사월의 초밥 | 13000 | 1.001618 | 0.135662 | 0.071176 | 0.507353 | 0.124559 | 0.049412 | 0.389118 | 0.016176 | 0.129044 | 0.017206 | 0.058162 | 0.019706 | 0.000000 | 0.106765 | 0.041912 | 1 | 4.333122 | 2.667868 |
| 9 | 동강산책 석촌로수점 | 13000 | 1.044240 | 0.065710 | 0.092922 | 0.365609 | 0.240384 | 0.003740 | 0.183005 | 0.009182 | 0.254674 | 0.179699 | 0.062254 | 0.015659 | 0.014424 | 0.066661 | 0.049482 | 1 | 4.359550 | 2.647646 |
| 48 | 정담초밥 문정점 | 15000 | 0.870186 | 0.101653 | 0.030000 | 0.322314 | 0.060000 | 0.067107 | 0.229091 | 0.022727 | 0.273347 | 0.106364 | 0.042025 | 0.074752 | 0.008926 | 0.070000 | 0.259091 | 2 | 4.250737 | 2.537583 |
| 26 | 스시아지 | 16000 | 0.900132 | 0.043348 | 0.074626 | 0.290749 | 0.149251 | 0.088811 | 0.344141 | 0.048458 | 0.172467 | 0.041233 | 0.024890 | 0.023612 | 0.000000 | 0.106608 | 0.100441 | 2 | 4.178748 | 2.408767 |
| 5 | 나레초밥 잠실새내점 | 13000 | 1.021090 | 0.244834 | 0.097488 | 0.234597 | 0.068815 | 0.026540 | 0.280664 | 0.000000 | 0.095972 | 0.022180 | 0.010711 | 0.019052 | 0.005118 | 0.223649 | 0.037820 | 1 | 4.318588 | 2.388531 |

2-3. 추천시스템

6. NDCG 평점 지표 적용

정의한 similar index 기반 '스시센세이'와 관련된 relevant item 도출

```
1 # relevant_item
2 relevant_item2 =similar_restaurant2[['업소명','가격','가격대','s_별점','total']]
3 relevant_item2.head(10)
```

| | 업소명 | 가격 | 가격대 | s_별점 | total |
|----|---------|-------|-----|----------|----------|
| 53 | 태인일식 | 19000 | 3 | 4.164493 | 3.847590 |
| 8 | 낭만 스시 | 13000 | 1 | 4.231170 | 3.556154 |
| 10 | 롤스시 | 16000 | 2 | 4.505346 | 3.298312 |
| 14 | 서가참치&스시 | 15000 | 2 | 4.379537 | 3.202045 |
| 25 | 스시센세이 | 15000 | 2 | 4.310399 | 3.149500 |
| 45 | 참치의미 | 19000 | 3 | 4.399872 | 3.137917 |
| 20 | 스시랑랑 | 15000 | 2 | 4.442048 | 3.106316 |
| 11 | 룬참치 | 15000 | 2 | 4.257578 | 3.052148 |
| 55 | 해조회초밥 | 18000 | 3 | 4.331255 | 3.033333 |
| 17 | 스시 오하요 | 14000 | 1 | 4.484094 | 2.951576 |

2-3. 추천시스템

6. NDCG 평점 지표 적용

정의한 `relevant_value` 식을 적용하여 recommend item 추출

```
1 #18000원 뺀값 전체 로그화 하여 recommend item 구하기
2
3 relevant_item2['relevant_value'] = (1/np.log(abs(relevant_item2['가격']-18005))) * round(relevant_item2['s_별점'],3)
4
5 recommend_item2 = relevant_item2.sort_values('relevant_value',ascending=False)
6 recommend_item2.head(10)
```

| | 업소명 | 가격 | 가격대 | s_별점 | total | relevant_value |
|----|---------|-------|-----|----------|----------|----------------|
| 55 | 해조회초밥 | 18000 | 3 | 4.331255 | 3.033333 | 2.691002 |
| 45 | 참치의미 | 19000 | 3 | 4.399872 | 3.137917 | 0.637428 |
| 53 | 태인일식 | 19000 | 3 | 4.164493 | 3.847590 | 0.603238 |
| 10 | 롤스시 | 16000 | 2 | 4.505346 | 3.298312 | 0.592498 |
| 41 | 젠 | 16000 | 2 | 4.437671 | 2.720386 | 0.583686 |
| 20 | 스시랑랑 | 15000 | 2 | 4.442048 | 3.106316 | 0.554693 |
| 26 | 스시아지 | 16000 | 2 | 4.178748 | 2.408767 | 0.549623 |
| 14 | 서가참치&스시 | 15000 | 2 | 4.379537 | 3.202045 | 0.546951 |
| 17 | 스시 오하요 | 14000 | 1 | 4.484094 | 2.951576 | 0.540547 |
| 25 | 스시센세이 | 15000 | 2 | 4.310399 | 3.149500 | 0.538210 |

relevant value 수식 ★

$$\frac{1}{\text{np.log}(\text{abs}(\text{추천대상 가격} - \text{비교대상 가격} + 5))} \times \text{round}(\text{steamRating}, 3)$$

2-3. 추천시스템

6. NDCG 평점 지표 적용

NDCG 계산

```
1 relevant_item_2 = list(relevant_item2['업소명'].head(10))
2 recommend_item_2 = list(recommend_item2['업소명'].head(10))
3
4 k = len(recommend_item_2)
5
6 cg = []
7 for item in recommend_item_2:
8     if item in relevant_item_2:
9         cg.append(1)
10    else:
11        cg.append(0)
12 cg
```

```
[1, 1, 1, 1, 0, 1, 0, 1, 1, 1]
```

```
1 #DCG
2 discount = np.log2(np.arange(k) + 2)
3 discount
```

```
array([1., 1.5849625, 2., 2.32192809, 2.5849625,
       2.80735492, 3., 3.169925, 3.32192809, 3.45943162])
```

```
1 # discounted cg 구하기
2 dcg = np.sum(np.divide(cg, discount))
3 dcg
```

```
3.8233731975204703
```

```
1 #IDCG
2 k_for_icg = min(k, len(relevant_item_2))
3 icg = np.zeros(k)
4 for i in range(k_for_icg):
5     icg[i] += 1
6 icg
```

```
array([1., 1., 1., 1., 1., 1., 1., 1., 1., 1.])
```

```
1 # ideal dcg 구하기
2 idcg = np.sum(np.divide(icg, discount))
3 idcg
```

```
4.543559338088345
```

```
1 # NDCG
2 ndcg = dcg / idcg
3 ndcg
```

```
0.841492960259019
```

NDCG 값 0.84 도출

03.

협업 필터링

3-1. 데이터 수집

1. 음식점별 리뷰 목록 '추천 순' 에서 상위 10명의 리뷰 작성자에 한해 데이터 수집

[< 가일](#)

[홈](#) [소식](#) [메뉴](#) [리뷰](#) [사진](#)

14

언어회 67

하이볼 26

라면 25

광어 21

만족도 449

분위기 118

서비스 83

음식량



복승e
리뷰 212 · 사진 108

[팔로우](#)



소문 듣고 왔는데 정말 장난 아니네요.!!! 언어두께 찍으려고 젓가락과 손가락을 동원해보았는데 보이시나요? □ 정말 두꺼워서 입 안에 가득 언어 맛을 ... ▾


단체모임 하기 좋아요

매장이 넓어요

친절해요

음식이 맛있어요

특별한 메뉴가 있어요



복승e
리뷰 212 · 사진 108
팔로잉 0 · 팔로워 0

[팔로우](#)

[전체](#) | [사진](#)

[최신순 ▾](#)

[전체 211](#)

[카페 63](#)

[한식 28](#)

[일식 18](#)


[양식 11](#)

[아시아/퓨전 음식 6](#)

[중식 6](#)

[분식](#)

우정초밥 서울시 중암동



중암의 빛입니다 ☺ 취소 자리 예약을 노리시는게 효율적(?)입니다.!!! 처음에는 어머니, 두번째는 가족, 세번째는 애인까지 ㅋㅋㅋㅋ 우정초밥은 꼭 가보세요!👍


특별한 메뉴가 있어요

음식이 맛있어요

가성비가 좋아요

친절해요

유소바 서울시 대현동



타다키 소바 이후에 덴푸라 소바도 먹었는데 말 없이 먹기만 할 정도로.. 정말 맛있었습니다 ☺

가성비가 좋아요

친절해요

재료가 신선해요

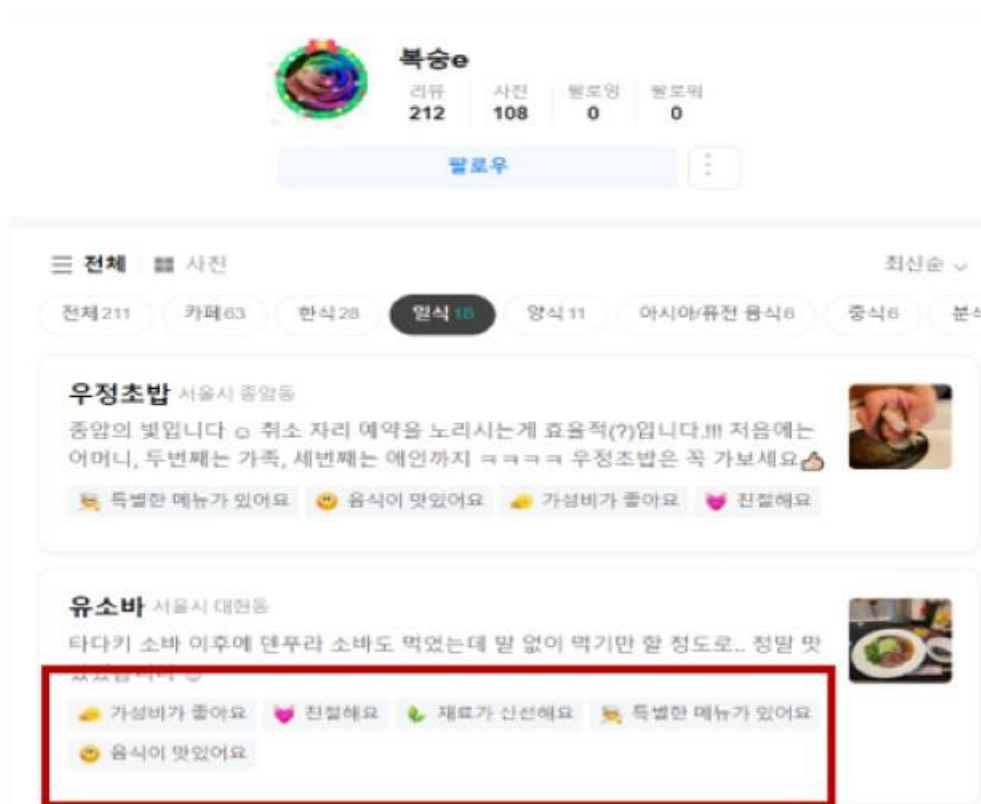
특별한 메뉴가 있어요

음식이 맛있어요

3-1. 데이터 수집

2. 유저가 네이버 영수증 리뷰에 투표한 항목에 상관관계 분석에서 얻은 가중치를 적용하여 합산

→ 유저의 음식점 평점 생성 ★



| | A | B | C | D | E | F |
|----|---------|------|-------------|---------------|------|---------|
| 1 | 가중치목록 | | | 가락정초밥 | 가일 | 기맨정스시 문 |
| 2 | 맛 | 1.39 | 안녕하세요 젤라예요 | 6.59 | | |
| 3 | 가성비 | 1.23 | sunacjil | 6.3 | | |
| 4 | 앙 | 1.21 | 포르테70 | 4.06 | | |
| 5 | 친절 | 1.5 | origju | 5.07 | | |
| 6 | 특별한메뉴 | 1.21 | 랑구 K입맛 | 3.95 | | |
| 7 | 주차 편리 | 1.12 | 니스의태양 | =B2+B8+B5+B14 | | |
| 8 | 재료 신선 | 1.26 | 쿵쿵쿵 | 2.89 | | |
| 9 | 화장실 청결도 | 1.1 | soletart | 3.86 | | |
| 10 | 매장 청결도 | 1.35 | 사나387 | 5.45 | | |
| 11 | 인테리어 | 1.17 | 또용잉89 | 6.27 | | |
| 12 | 특별한 날 | 1.13 | 이선179 | 5.21 | | |
| 13 | 단체모임 | 1.34 | ehg**** | 6.29 | | |
| 14 | 뷰 | 1.08 | melloclever | 3.83 | | |
| 15 | 혼밥 | 1.21 | 방귀대장뽕뽕이야 | 6.59 | | |
| 16 | 매장크기 | 1.14 | 엘리스58 | 6.49 | | |
| 17 | | | 와이알알 | 4.15 | | |
| 18 | | | 괘남송연98 | 2.76 | | |
| 19 | | | 돈없는만수르 | 5.09 | | |
| 20 | | | 봄이좋아Bomom | 6.59 | | |
| 21 | | | 데미지444 | 5.47 | | |
| 22 | | | 복숭e | | 6.58 | |
| 23 | | | MINILIM31 | | 6.3 | |
| 24 | | | 김민준14 | | 6.61 | |

3-1 데이터셋 수집

704명의 유저 별 음식점 평가 점수 matrix 생성

```
[ ] R=RI.iloc[:,0:]
R
```

| res_name | 가락정초밥 | 가일 | 기맨정스시 | 문정동본점 | 김태완 | 스시송파본점 | 나레초밥 | 거여본점 | 나레초밥 | 잠실새내점 | 나루스시 | 난스시 | 낭만스시 | 동경산책 | 석촌호수점 | ... | 참치현 | 채원일식당 | 청담초밥 | 문정점 | 정화초밥 | 초밥식당호시점 | 초이스시 | 큐스시 | 태인일식 | 토리가 | 화진 |
|----------|-------|-----|-------|-------|-----|--------|------|------|------|-------|------|-----|------|------|-------|-----|-----|-------|------|-----|------|---------|------|-----|------|-----|-----|
| User_ID | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1 | NaN | NaN | | 2.65 | | NaN | NaN | | NaN | NaN | NaN | NaN | NaN | | NaN | ... | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| 2 | NaN | NaN | | NaN | | NaN | NaN | | NaN | NaN | 6.35 | NaN | NaN | | NaN | ... | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| 3 | NaN | NaN | | NaN | | NaN | NaN | | NaN | NaN | NaN | NaN | NaN | | NaN | ... | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| 4 | NaN | NaN | | NaN | | NaN | NaN | | NaN | NaN | NaN | NaN | NaN | | NaN | ... | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| 5 | NaN | NaN | | NaN | | NaN | NaN | | NaN | NaN | NaN | NaN | NaN | | NaN | ... | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| ... | ... | ... | | ... | | ... | ... | | ... | ... | ... | ... | ... | | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 700 | NaN | NaN | | NaN | | 1.23 | NaN | | NaN | NaN | NaN | NaN | NaN | | NaN | ... | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| 701 | NaN | NaN | | NaN | | NaN | NaN | | NaN | NaN | NaN | NaN | NaN | | NaN | ... | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| 702 | NaN | NaN | | NaN | | NaN | NaN | | NaN | NaN | NaN | NaN | NaN | | NaN | ... | NaN | NaN | 5.0 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| 703 | NaN | NaN | | NaN | | NaN | NaN | | NaN | NaN | 6.30 | NaN | NaN | | NaN | ... | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| 704 | NaN | NaN | | NaN | | NaN | NaN | | NaN | NaN | NaN | NaN | NaN | | NaN | ... | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |

704 rows × 56 columns

3-2. SDG 예측 행렬 생성

유저 별 음식점 평가 점수 matrix와 랜덤 값으로 입력된 P,Q matrix를 이용-> SGD 예측 행렬 생성

```
1 R = np.array(R)
2 R
array([[ nan,  nan, 2.65, ..., nan,  nan,  nan],
       [ nan,  nan,  nan, ..., nan,  nan,  nan],
       [ nan,  nan,  nan, ..., nan,  nan,  nan],
       ...,
       [ nan,  nan,  nan, ..., nan,  nan,  nan],
       [ nan,  nan,  nan, ..., nan,  nan,  nan],
       [ nan,  nan,  nan, ..., nan,  nan,  nan]])
```

```
1 num_users, num_items = R.shape
2 K=50
3
4 # P와 Q 매트릭스의 크기를 지정하고 정규분포를 가진 random한 값으로 입력합니다.
5 np.random.seed(1)
6 P = np.random.normal(scale=1./K, size=(num_users, K))
7 Q = np.random.normal(scale=1./K, size=(num_items, K))
```

```
1 from sklearn.metrics import mean_squared_error
2
3 def get_rmse(R, P, Q, non_zeros):
4     error = 0
5     # 두개의 분해된 행렬 P와 Q.T의 내적으로 예측 R 행렬 생성
6     full_pred_matrix = np.dot(P, Q.T)
7
8     # 실제 R 행렬에서 널이 아닌 값의 위치 인덱스 추출하여 실제 R 행렬과 예측 행렬의 RMSE 추출
9     x_non_zero_ind = [non_zero[0] for non_zero in non_zeros]
10    y_non_zero_ind = [non_zero[1] for non_zero in non_zeros]
11    R_non_zeros = R[x_non_zero_ind, y_non_zero_ind]
12    full_pred_matrix_non_zeros = full_pred_matrix[x_non_zero_ind, y_non_zero_ind]
13
14    mse = mean_squared_error(R_non_zeros, full_pred_matrix_non_zeros)
15    rmse = np.sqrt(mse)
16
17    return rmse
```

```
1 # R > 0 인 행 위치, 열 위치, 값을 non_zeros 리스트에 저장.
2 non_zeros = [ (i, j, R[i,j]) for i in range(num_users) for j in range(num_items) if R[i,j] > 0 ]
3
4 steps=5000
5 learning_rate=0.01
6 r_lambda=0.01
7
8 # SGD 기법으로 P와 Q 매트릭스를 계속 업데이트.
9 for step in range(steps):
10     for i, j, r in non_zeros:
11         # 실제 값과 예측 값의 차이인 오류 값 구함
12         eij = r - np.dot(P[i, :], Q[j, :].T)
13         # Regularization을 반영한 SGD 업데이트 공식 적용
14         P[i,:] = P[i,:] + learning_rate*(eij * Q[j, :] - r_lambda*P[i,:])
15         Q[j,:] = Q[j,:] + learning_rate*(eij * P[i, :] - r_lambda*Q[j,:])
16
17     rmse = get_rmse(R, P, Q, non_zeros)
18     if (step % 500) == 0 :
19         print("### iteration step : ", step, " rmse : ", rmse)
```

```
### iteration step : 0 rmse : 4.45899724280908
### iteration step : 500 rmse : 0.017952997111696234
### iteration step : 1000 rmse : 0.015736415056583748
### iteration step : 1500 rmse : 0.014322519050925907
### iteration step : 2000 rmse : 0.013399093112126103
### iteration step : 2500 rmse : 0.012768343564222026
### iteration step : 3000 rmse : 0.012321095171656228
### iteration step : 3500 rmse : 0.011994537043343325
### iteration step : 4000 rmse : 0.011750572745613596
### iteration step : 4500 rmse : 0.011564979107147232
```

```
1 pred_matrix = np.dot(P, Q.T)
2 print('예측 행렬:\n', np.round(pred_matrix, 3))
```

```
예측 행렬:
[[-0.090  0.226  2.644 ...  0.702 -0.063  0.128]
 [-0.034 -0.992  0.957 ... -0.26  0.865  1.256]
 [-0.162  1.109  0.10 ...  0.354 -0.851 -1.398]
 ...
 [-0.704 -0.604  1.554 ... -0.45  0.857  1.204]
 [ 0.02 -0.929  0.974 ... -0.207  0.819  1.234]
 [-0.026 -1.096  0.265 ...  0.728  0.588  0.397]]
```

```
1 print('경사하강 rating 값 예측 RMSE : ', rmse)
```

```
경사하강 rating 값 예측 RMSE : 0.011421977955467143
```

RMSE 값 0.0114 도출

최종 데이터셋

경사 하강 행렬 분해 예측 값 적용한 최종 유저 matrix 생성

```
1 # 예측값 넣은 행렬
2 rating_pred_matrix = RR_T
3 rating_pred_matrix
```

| res_name | 가락정초밥 | 가일 | 기원정스시 | 문정등본점 | 김태완 스시 | 승파본점 | 나래초밥 | 거여본점 | 나래초밥 | 잠실새내점 | 나루스시 | 난 스시 | 낭만 스시 | 동경산책 | 석촌호수점 | ... | 참치현 | 채원 | 일식당 | 청담초밥 | 문정점 | 청화초밥 | 초밥식당호시절 | 초이스시 | 큐 스시 | 태연일식 | 토리가 | 최항 | 초밥미랑 |
|----------|-------|-------|-------|-------|--------|-------|------|-------|------|-------|-------|-------|-------|------|-------|-----|-------|-------|-----|-------|-------|-------|---------|-------|-------|-------|-----|-------|------|
| user_id | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1 | -0.10 | 0.23 | | 2.64 | | 0.70 | | 0.53 | | -0.06 | 0.46 | 0.43 | 0.12 | | 0.42 | _ | -1.07 | 0.76 | | 0.79 | 0.45 | -0.02 | 0.09 | -0.07 | 0.70 | -0.06 | | 0.13 | |
| 2 | -0.03 | -0.99 | | 0.96 | | 3.01 | | -0.01 | | 1.16 | 6.34 | -0.40 | -0.90 | | 1.93 | _ | -1.07 | 1.02 | | 2.50 | -0.02 | 1.18 | -0.08 | 0.52 | -0.26 | 0.87 | | 1.26 | |
| 3 | -0.16 | 1.11 | | 0.18 | | -1.85 | | 0.04 | | 0.58 | -1.42 | 2.11 | 0.20 | | 1.04 | _ | 0.37 | -0.35 | | 0.30 | 0.67 | 1.30 | -0.72 | -0.47 | 0.35 | -0.85 | | -1.40 | |
| 4 | -0.21 | -0.05 | | -0.02 | | -0.11 | | 0.15 | | -0.03 | 0.02 | -0.12 | -0.30 | | 0.20 | _ | -0.16 | -0.20 | | -0.09 | 0.54 | -0.18 | -0.23 | -0.04 | 0.33 | 0.05 | | 0.02 | |
| 5 | 0.26 | 0.87 | | -0.26 | | -0.74 | | 0.43 | | 0.68 | -0.75 | 0.82 | 0.49 | | 0.59 | _ | 0.03 | -0.24 | | 0.51 | -0.05 | 1.45 | 0.10 | -0.25 | -0.55 | -0.12 | | -0.57 | |
| ... | - | - | - | - | | - | | - | | - | - | - | - | | - | - | - | - | | - | - | - | - | - | - | - | - | - | |
| 700 | 0.72 | 1.68 | | 0.85 | | 1.23 | | 0.13 | | 2.15 | 0.32 | 0.61 | 0.80 | | 1.24 | _ | 0.31 | 0.54 | | 0.38 | 1.26 | 2.33 | 0.71 | -0.34 | 1.35 | -0.11 | | 0.57 | |
| 701 | 0.68 | 1.75 | | 0.69 | | 0.40 | | 0.09 | | -0.34 | -0.51 | -0.70 | 0.06 | | 0.49 | _ | 0.56 | 0.04 | | 0.49 | 1.07 | 0.06 | 0.12 | 0.07 | -0.01 | 0.61 | | 0.63 | |
| 702 | -0.78 | -0.68 | | 1.55 | | 0.45 | | -0.88 | | 0.39 | 2.42 | 0.24 | 0.45 | | 1.74 | _ | -1.30 | 1.97 | | 4.99 | -0.67 | 0.10 | -0.06 | -0.40 | -0.45 | 0.86 | | 1.28 | |
| 703 | 0.02 | -0.93 | | 0.97 | | 3.03 | | -0.02 | | 1.09 | 6.29 | -0.38 | -0.93 | | 1.93 | _ | -1.05 | 1.02 | | 2.44 | -0.01 | 1.13 | -0.08 | 0.55 | -0.29 | 0.82 | | 1.23 | |
| 704 | -0.03 | -1.10 | | 0.27 | | 0.59 | | 0.27 | | 1.30 | 0.87 | 0.12 | 0.31 | | 0.12 | _ | -0.20 | 0.53 | | 0.26 | -0.17 | 0.76 | 0.05 | -0.06 | 0.73 | 0.59 | | 0.40 | |

704 rows x 56 columns

3-3. 추천시스템

1. 경사 하강 예측 matrix 값과 유저 별 음식점 평가 matrix (결측 채워지지 않은 값) 이용해 추천

```
1 # 협업 필터링
2
3 def get_ungone_rest(rating_matrix, User_Id):
4     user_rating = rating_matrix.loc[User_Id,:]
5     already_gone = user_rating[user_rating > 0].index.tolist()
6
7     restaurant_list = rating_matrix.columns.tolist()
8
9     ungone_list = [restaurant for restaurant in restaurant_list if restaurant not in already_gone]
10
11     return ungone_list
```

```
1 def recomm_res_by_userid(pred_df, User_Id, unseen_list, top_n=10):
2     recomm_res = pred_df.loc[User_Id, unseen_list].sort_values(ascending=False)[:top_n]
3     return recomm_res
```

```
1 userid = int(input('user id를 입력하세요: '))
```

user id를 입력하세요: 270

```
1 # 사용자가 가지않은 식당명 추출
2 ungone_list = get_ungone_rest(rating_matrix, userid) #유저 270번
3
4 # 아이템 기반의 인접 이웃 협업 필터링으로 영화 추천
5 recomm_res = recomm_res_by_userid(rating_pred_matrix, userid, ungone_list, top_n=20) # 유저 270번
6
7 # 평균 데이터를 DataFrame으로 생성.
8 recomm_restaurants = pd.DataFrame(data=recomm_res.values, index=recomm_res.index, columns=['pred_rating'])
```

```
1 recomm_restaurants.reset_index(inplace=True)
```

[94] recomm_restaurants

| | res_name | pred_rating |
|----|----------------|-------------|
| 0 | 나레초밥 잠실새내점 | 2.51 |
| 1 | 스시 어도락 | 2.31 |
| 2 | 정 스시 | 2.20 |
| 3 | 스시메시 | 1.84 |
| 4 | 여항 스시&사시미 | 1.50 |
| 5 | 사월의 초밥 | 1.49 |
| 6 | 가일 | 1.45 |
| 7 | 스시 오하요 | 1.44 |
| 8 | 나레초밥 거여본점 | 1.39 |
| 9 | 낭만 스시 | 1.36 |
| 10 | 가락정초밥 | 1.29 |
| 11 | 스시유메 | 1.19 |
| 12 | 진가와 잠실롯데월드몰점 | 0.95 |
| 13 | 초이스시 | 0.91 |
| 14 | 미스터 스시 가락시장직영점 | 0.82 |
| 15 | 성전 | 0.80 |
| 16 | 스시센세이 | 0.77 |
| 17 | 야나기스시 | 0.75 |

→ 추천리스트

3-3. 추천시스템

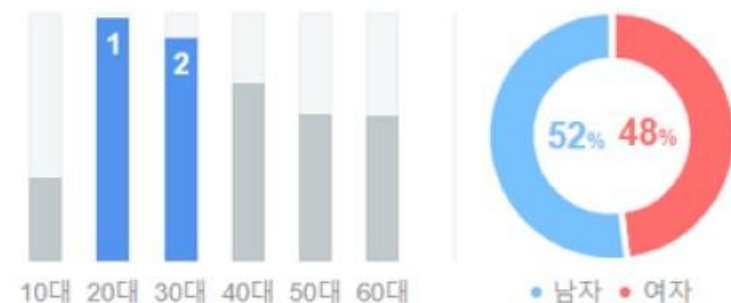
2. 추천 시스템 발전 : 데이터랩 정보 사용 → 유저의 특성 반영하여 추천 ★

2-1. 데이터랩 연령별, 시간대별 인기도 정보 사용 - 결측치 : 최빈값 대체

금요일 시간대별 검색 인기도



연령별 / 성별 검색 인기도



```
print(dl['나이'].mode())  
print(dl['시간대'].mode())
```

```
dl['나이'] = dl['나이'].fillna('30대')  
dl['시간대'] = dl['시간대'].fillna('저녁')  
dl.head()
```

```
0    20대  
dtype: object  
0    저녁  
dtype: object
```

| | res_name | 위치 | 나이 | 시간대 |
|---|-------------|-----|-----|-----|
| 0 | 가락정초밥 | 가락동 | 40대 | 저녁 |
| 1 | 가일 | 방이동 | 20대 | 저녁 |
| 2 | 기맨정스시 문정등본점 | 문정동 | 30대 | 저녁 |
| 3 | 김태완 스시 송파본점 | 송파동 | 20대 | 저녁 |
| 4 | 나레초밥 거여본점 | 거여동 | 40대 | 저녁 |

3-3. 추천시스템

2. 추천 시스템 발전 : 데이터랩 정보 사용 → 유저의 특성 반영하여 추천

2-2. Input 으로 유저 선호 정보 받음

```
▶ print('유저 정보를 입력해주세요')  
  
print(' ')  
print('--' *5)  
print(' ')  
  
a = input('선호하는 식당 위치 (ex. 가락동...) : ', )  
b = input('나시대 (ex. 10대...) : ', )  
c = input('선호하는 시간대 (ex. 아침, 점심, 저녁, 밤) : ', )
```

👤 유저 정보를 입력해주세요

선호하는 식당 위치 (ex. 가락동...) : 가락동
나시대 (ex. 10대...) : 20대
선호하는 시간대 (ex. 아침, 점심, 저녁, 밤) : 저녁

3-3. 추천시스템

2. 추천 시스템 발전 : 데이터랩 정보 사용 → 유저의 특성 반영하여 추천

2-3. 유저의 선호 정보와 교집합이 가장 많은 식당을 우선으로 추천하고,
교집합의 개수가 같을 시에는 pred_rating이 높은 순으로 추천

```
1 recomm_restaurants_dl = recomm_restaurants_dl.sort_values(by=['count', 'pred_rating'], ascending = [False, False])  
2 recomm_restaurants_dl
```

| | res_name | pred_rating | 위치 | 나이 | 시간대 | count |
|----|----------------|-------------|-----|-----|-----|-------|
| 0 | 나레초밥 잠실새내점 | 2.51 | 잠실동 | 20대 | 저녁 | 2 |
| 1 | 스시 어도락 | 2.31 | 가락동 | 30대 | 저녁 | 2 |
| 2 | 정 스시 | 2.20 | 가락동 | 30대 | 저녁 | 2 |
| 5 | 사월의 초밥 | 1.49 | 가락동 | 30대 | 저녁 | 2 |
| 6 | 가일 | 1.45 | 방이동 | 20대 | 저녁 | 2 |
| 3 | 스시메시 | 1.84 | 잠실동 | 30대 | 저녁 | 1 |
| 4 | 어항 스시&사시미 | 1.50 | 문정동 | 30대 | 저녁 | 1 |
| 7 | 스시 오하요 | 1.44 | 삼전동 | 30대 | 저녁 | 1 |
| 8 | 나레초밥 거여본점 | 1.39 | 거여동 | 40대 | 저녁 | 1 |
| 9 | 낭만 스시 | 1.36 | 풍납동 | 40대 | 저녁 | 1 |
| 10 | 가락정초밥 | 1.29 | 가락동 | 40대 | 저녁 | 0 |
| 11 | 스시유메 | 1.19 | 방이동 | 20대 | 저녁 | 0 |
| 12 | 진가와 잠실롯데월드몰점 | 0.95 | 신천동 | 20대 | 점심 | 0 |
| 13 | 초이스시 | 0.91 | 잠실동 | 20대 | 저녁 | 0 |
| 14 | 미스터 스시 가락시장직영점 | 0.82 | 가락동 | 60대 | 저녁 | 0 |

3-3. 추천시스템

유저 270번에게 최종적으로 추천된 음식점 20개

```
1 # 최종 유저 추천 시스템
2 recomm_restaurants_dl
```

| res_name | pred_rating | 위치 | 나이 | 시간대 |
|------------|-------------|-----|-----|-----|
| 나레초밥 잠실새내점 | 2.51 | 잠실동 | 20대 | 저녁 |
| 스시 어도락 | 2.31 | 가락동 | 30대 | 저녁 |
| 정 스시 | 2.20 | 가락동 | 30대 | 저녁 |
| 사월의 초밥 | 1.49 | 가락동 | 30대 | 저녁 |
| 가일 | 1.45 | 방이동 | 20대 | 저녁 |
| 스시메시 | 1.84 | 잠실동 | 30대 | 저녁 |
| 어항 스시&사시미 | 1.50 | 문정동 | 30대 | 저녁 |
| 스시 오하요 | 1.44 | 삼전동 | 30대 | 저녁 |
| 나레초밥 거여본점 | 1.39 | 거여동 | 40대 | 저녁 |
| 낭만 스시 | 1.36 | 풍납동 | 40대 | 저녁 |
| 가락정초밥 | 1.29 | 가락동 | 40대 | 저녁 |



| | | | | |
|----------------|------|-----|-----|----|
| 가락정초밥 | 1.29 | 가락동 | 40대 | 저녁 |
| 스시유메 | 1.19 | 방이동 | 20대 | 저녁 |
| 진가와 잠실롯데월드몰점 | 0.95 | 신천동 | 20대 | 점심 |
| 초이스시 | 0.91 | 잠실동 | 20대 | 저녁 |
| 미스터 스시 가락시장직영점 | 0.82 | 가락동 | 60대 | 저녁 |
| 성전 | 0.80 | 가락동 | 30대 | 저녁 |
| 스시센세이 | 0.77 | 신천동 | 30대 | 저녁 |
| 야나기스시 | 0.75 | 잠실동 | 40대 | 저녁 |
| 젠 | 0.74 | 송파동 | 20대 | 점심 |
| 난 스시 | 0.70 | 문정동 | 30대 | 저녁 |

04.

하이브리드 추천

4-1 적용 방법

협업 필터링을 통해 유저에게 추천된 식당과 비슷한 식당을 5개씩 콘텐츠 기반 필터링으로 추천 ★

[87] # 유저에게 추천된 식당

```
res_user_id = recomm_restaurants_dl.index.values  
res_user_id
```



```
array(['나레초밥 잠실새내점', '스시 어도락', '정 스시', '사월의 초밥', '가일', '스시메시',  
      '어항 스시&사시미', '스시 오하요', '나레초밥 거여본점', '낭만 스시', '가락정초밥', '스시유메',  
      '진가와 잠실롯데월드몰점', '초이스시', '미스터 스시 가락시장직영점', '성전', '스시센세이', '야나기스시',  
      '젠', '난 스시'], dtype=object)
```

[88] df = pd.DataFrame(columns = ['유저식당', '추천식당', 'total'])

```
for i in res_user_id:  
    similar_restaurant2 = find_sim_res(weight_df, df2_sim_sorted_ind, i, 6) # 결과 확인하기 위해 1차적으로 스시센세이로 선정  
  
    similar_restaurant_name = similar_restaurant2['업소명'].values[1:]  
    similar_restaurant_total = similar_restaurant2['total'].values[1:]  
  
    # 가중치(total) 기준 sorting  
    #similar_restaurant2 = similar_restaurant2.sort_values(by=['total'], axis=0, ascending=False)  
  
    df = df.append(pd.DataFrame([[i, similar_restaurant_name, similar_restaurant_total]], columns=['유저식당', '추천식당', 'total']))  
  
df.set_index(['유저식당'], inplace = True)
```

최종 결과

Multi index 적용 + total로 내림차순 정렬하여 최종 결과 추천

| rank | 유저식당 | 추천식당 | total |
|------|----------------|----------------|----------|
| 1 | 나레초밥 잠실새내점 | 나레초밥 거여본점 | 2.680792 |
| | | 사월의 초밥 | 2.667868 |
| | | 토리가 | 2.340577 |
| | | 스시 준 | 2.337343 |
| | | 난 스시 | 2.164412 |
| 2 | 스시 어도락 | 김태완 스시 송파본점 | 3.405651 |
| | | 청화초밥 | 2.572372 |
| | | 성전 | 2.555614 |
| | | 진가와 잠실롯데월드몰점 | 2.542754 |
| | | 미스터 스시 가락시장직영점 | 2.322047 |
| 3 | 정 스시 | 스시유메 | 3.545120 |
| | | 스시선 방이점 | 3.470672 |
| | | 스시를 담다 | 3.374737 |
| | | 기원정스시 문정동본점 | 2.735289 |
| | | 스시이타 | 2.563953 |
| 4 | 스시메시 | 나레초밥 잠실새내점 | 2.388531 |
| | | 토리가 | 2.340577 |
| | | 스시 준 | 2.337343 |
| | | 난 스시 | 2.164412 |
| | | 지평선초밥 가락본점 | 1.685556 |
| 5 | 어항 스시&샤시미 | 아나기스시 | 1.415112 |
| | | 스시진 | 1.131747 |
| | | 열애초밥 거여역점 | 1.045884 |
| | | 왕스시 | 0.987968 |
| | | 아카시 문정점 | 0.713469 |
| 6 | 사월의 초밥 | 스시 오하요 | 2.951576 |
| | | 나레초밥 거여본점 | 2.680792 |
| | | 진미 참지 | 2.669765 |
| | | 동경산책 석촌호수점 | 2.647646 |
| | | 스시 준 | 2.337343 |
| 7 | 가일 | 스시엔 | 4.891129 |
| | | 참지현 | 4.244091 |
| | | 태안일식 | 3.847590 |
| | | 참지의미 | 3.137917 |
| | | 론참지 | 3.052148 |
| 8 | 스시 오하요 | 스시리완 | 2.794815 |
| | | 나레초밥 거여본점 | 2.680792 |
| | | 진미 참지 | 2.669765 |
| | | 사월의 초밥 | 2.667868 |
| | | 이화초밥 | 2.601298 |
| 9 | 나레초밥 거여본점 | 스시 오하요 | 2.951576 |
| | | 스시리완 | 2.794815 |
| | | 사월의 초밥 | 2.667868 |
| | | 나레초밥 잠실새내점 | 2.388531 |
| | | 스시 준 | 2.337343 |
| 10 | 낭만 스시 | 가락정초밥 | 4.342275 |
| | | 조이스시 | 3.933171 |
| | | 스시 오하요 | 2.951576 |
| | | 스시리완 | 2.794815 |
| | | 나레초밥 거여본점 | 2.680792 |
| 11 | 가락정초밥 | 조이스시 | 3.933171 |
| | | 낭만 스시 | 3.556154 |
| | | 스시선 방이점 | 3.470672 |
| | | 스시 오하요 | 2.951576 |
| | | 스시리완 | 2.794815 |
| 12 | 스시유메 | 스시선 방이점 | 3.470672 |
| | | 스시를 담다 | 3.374737 |
| | | 정 스시 | 3.284815 |
| | | 기원정스시 문정동본점 | 2.735289 |
| | | 스시이타 | 2.563953 |
| 13 | 진가와 잠실롯데월드몰점 | 나루스시 | 4.883009 |
| | | 초밥식당호시절 | 2.845133 |
| | | 성전 | 2.555614 |
| | | 미스터 스시 가락시장직영점 | 2.322047 |
| | | 스시 어도락 | 2.039675 |
| 14 | 조이스시 | 가락정초밥 | 4.342275 |
| | | 낭만 스시 | 3.556154 |
| | | 스시선 방이점 | 3.470672 |
| | | 스시 오하요 | 2.951576 |
| | | 스시리완 | 2.794815 |
| 15 | 미스터 스시 가락시장직영점 | 김태완 스시 송파본점 | 3.405651 |
| | | 청화초밥 | 2.572372 |
| | | 성전 | 2.555614 |
| | | 진가와 잠실롯데월드몰점 | 2.542754 |
| | | 스시 어도락 | 2.039675 |
| 16 | 성전 | 김태완 스시 송파본점 | 3.405651 |
| | | 청화초밥 | 2.572372 |
| | | 진가와 잠실롯데월드몰점 | 2.542754 |
| | | 미스터 스시 가락시장직영점 | 2.322047 |
| | | 스시 어도락 | 2.039675 |
| 18 | 아나기스시 | 스시아지 | 2.408767 |
| | | 어항 | 1.896033 |
| | | 지평선초밥 가락본점 | 1.685556 |
| | | 스시진 | 1.131747 |
| | | 열애초밥 거여역점 | 1.045884 |
| 19 | 첸 | 스시랑랑 | 3.106316 |
| | | 론참지 | 3.052148 |
| | | 스시 153 | 2.856829 |
| | | 청담초밥 문정점 | 2.537583 |
| | | 스시아지 | 2.408767 |
| 20 | 난 스시 | 사월의 초밥 | 2.667868 |
| | | 나레초밥 잠실새내점 | 2.388531 |
| | | 토리가 | 2.340577 |
| | | 스시 준 | 2.337343 |
| | | 스시메시 | 1.799714 |

2022-2 빅데이터 캡스톤디자인

**THANK
YOU,**