대구맛집 과제

2022235027 민현기

1. 대구의 7개 구(중구, 동구, 서구, 남구, 북구, 수성구, 달서구)과 1개 군(달성군)의 맛집은 각각 몇 개인가?

○ requests와 urllib를 활용하여 api 호출용 주소를 작성하고, 원하는 지역에 대한 리스트를 작성한 뒤 get 방식으로 요청하고 json으로 return 받음.

- 중구: 181 개
- 동구: 137 개
- 서구: 63 개
- 남구: 57 개
- 북구: 95 개
- 수성구: 131 개
- 달서구: 168 개
- 달성군: 93 개

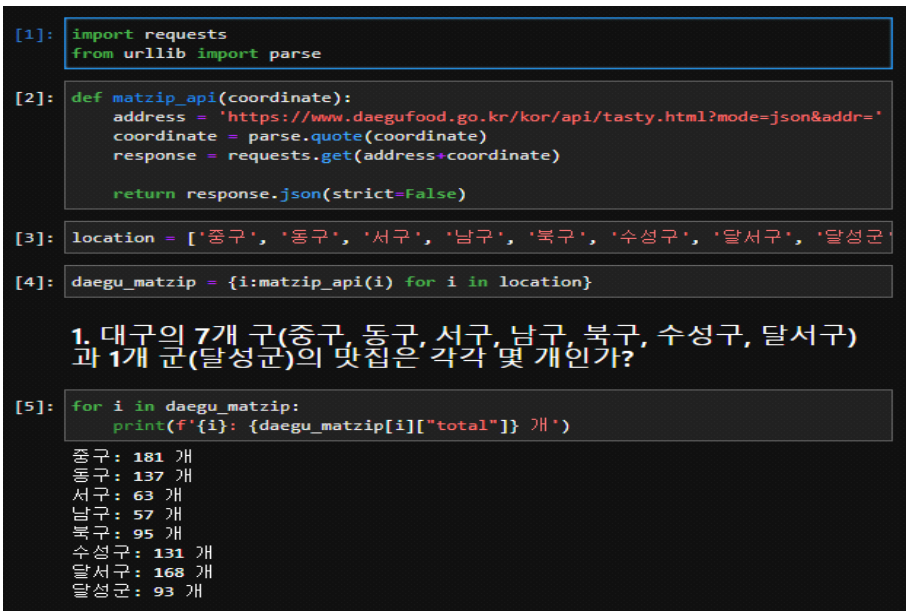

그림 1 api를 호출한 뒤 결과를 json으로 리턴받는 함수 및 맛집개수 파악

2. 오픈 api에 포함된 "설명"(SMPL_DESC)을 분석하여, 맛집 설명에서 가장 유의미하게 많이 등장하는 5개 단어를 제시하라.

○ 토픽모델링(LDA)을 활용하여 맛집 설명글을 분석할 수 있도록 LDA 클래스 작성

○ 토픽모델링 수행 중 토픽의 개수는 이미 맛집 설명으로 주제가 통일되었으므로 1개로 설정함

○ 불용어의 경우 초기에 설정하지 않고 토픽모델링을 수행한 뒤, 문서에 대한 기여도가 높은 단어에 대하여 실제 맛집에 대한 설명으로 적절한지 판단함

■ 예를 들어, '요리', '음식점' 등의 단어는 맛집이라는 주제에 대해 당연하게 나올 수 있는 단어이므로 제외하였고, '수', '곳', '등' 등의 단어는 부적절한 표현으로 제외함

- 최종 선정 단어: '전문점', '맛', '위치', '전통', '직접'

```
import numpy as np
import pandas as pd
import re
from konlpy.tag import Okt
import gensim
from gensim import corpora

def make_df(location):
    df = []
    for i in location:
        local_df = pd.DataFrame(daegu_matzip[i]['data'])
        local_df['LOCATION'] = i
        df.append(local_df)

    return pd.concat(df).reset_index(drop=True)

df = make_df(location)
```

그림 2 초기 return받는 json에서 필요한 정보가 들어있는 'data' 딕셔너리 키를 활용한 데이터프레임 생성함수

```
class CustomLDA:
    def __init__(self, pandas_text_column, stop_words, num_topics=1, num_words=5):
        clean = re.compile("[^ㄱ-힣0-9 %]")
        self.texts = pandas_text_column.apply(lambda x: clean.sub('', str(x)))
        self.tagger = Okt()
        self.stop_words = stop_words
        self.num_topics = num_topics
        self.num_words = num_words

    def nouns_tokenizer(self, text):
        word_token = self.tagger.nouns(text)
        self.result = [word for word in word_token if word not in self.stop_words]
        return self.result

    def LDA(self):
        nouns_tokenizer = self.nouns_tokenizer
        texts = self.texts.apply(nouns_tokenizer)
        dictionary = corpora.Dictionary(texts)
        corpus = [dictionary.doc2bow(text) for text in texts]
        NUM_TOPICS = self.num_topics
        ldamodel = gensim.models.ldamodel.LdaModel(corpus, num_topics = NUM_TOPICS, id2word=dictionary, passes=15)
        topics = ldamodel.print_topics(num_words=self.num_words)
        return topics
```

- 처음엔 불용어를 설정하지 않고, 결과를 돌려보며 리스트에 나오는 불용어를 하나씩 제거하며 최종적으로 5개의 단어를 선정함

```
stop_words = ['수', '요리', '곳', '등', '음식점', '한우', '사용', '음식']
CustomLDA(pandas_text_column=df['SMPL_DESC'], stop_words=stop_words, num_topics=1).LDA()
```

```
[(0, '0.043*"전문점" + 0.012*"맛" + 0.010*"위치" + 0.008*"전통" + 0.008*"직접"')]
```

그림 3 LDA 클래스. 명사를 추출하고 불용어를 제거하는 클래스 안 nonus_tokenizer 함수와 LDA를 수행하는 함수 작성

3. 오픈 api에 포함된 "메뉴"(MNU)를 분석하여, 대구 맛집들이 제공하는 가장 흔한 메뉴 3개를 제시하라. 그리고 이들의 평균 가격 또한 함께 제시하라.

○ 각 식당별 메뉴와 가격이 포함된 정보에서 이를 분리시키기 위하여 정규식을 활용하였으며, 분리시킨 정보를 활용한 데이터프레임 생성 함수 작성

■ 분석용 데이터프레임은 각 식당별 메뉴의 품목과, 가격에 대한 정보만을 나타냄

```
def make_menu_df(text):
    text = text.strip('\r\n').split('\r\n')

    pattern_price = re.compile('[0-9,[0-9]+(?=원)')
    pattern_menu = re.compile('^(.*?)(?= \d+,)')

    menu = pd.Series([pattern_menu.findall(i) for i in text]).apply(lambda x: ''.join(x) if x else '없음')
    # menu = pd.Series(text).apply(CustomLDA(pd.Series(text[0]), stop_words).nouns_tokenizer).apply(lambda x: ''.join(x))
    price = pd.Series([pattern_price.findall(i) for i in text]).apply(lambda x: int(x[0].replace(',', '')) if x else np.nan)

    return pd.DataFrame(zip(menu, price), columns=['품목', '가격'])
```

```
def menu_concat(df):
    concat_list_df = [make_menu_df(df['MNU'][i]) for i in range(len(df['MNU']))]
    return pd.concat(concat_list_df).reset_index(drop=True), concat_list_df
```

```
menu_df, store_df = menu_concat(df)
```

그림 4 품목 가격 통합 데이터프레임 생성 함수

○ 데이터 프레임에서 품목에 대한 정보를 가져온 뒤, 2번 문제에서 적용한 LDA를 수행하여 품목별 토픽에 대한 대표적인 음식 3개를 호출

○ 이 때, 토픽의 수는 식당 메뉴라는 토픽으로 정해져 있으므로 1개로 설정하였으며, 2번 문제를 수행한 방식과 동일하게 불용어 처리를 진행하여 대표 메뉴 3개를 선정함

- 최종 선정 메뉴: '한우', '불고기', '정식

menu_df

	품목	가격
0	복어매운탕	9000.0
1	복어지리	9000.0
2	복어수육	45000.0
3	모듬회	20000.0
4	껍질무침	15000.0
...	...	...
6618	해물파전	13000.0
6619	하와이안 피자	32000.0
6620	베조 피자	34000.0
6621	베지 스파게티	15000.0
6622	까르보나라	15000.0

6623 rows × 2 columns

그림 5 품목 가격 통합 데이터 프레임

```
stop_words_menu = ['']
CustomLDA(menu_df['품목'], stop_words=stop_words_menu, num_topics=1, num_words=3).LDA()
```

```
[(0, '0.016*"한우" + 0.016*"불고기" + 0.013*"정식"')]
```

그림 6 최종 선정 메뉴

○ 최종 선정된 메뉴가 포함된 서브 데이터프레임을 생성하여, 해당 메뉴에 대한 평균 가격 계산

menu_df[menu_df['품목'].str.contains('한우')]

	품목	가격
14	디너 한우 채끝등심 코스	75000.0
198	한우 불고기정식	11000.0
199	한우 갈비탕	13000.0
200	한우 갈비찜정식	19000.0
203	한우한마리(300g)	45000.0
...	...	...
6351	한우갈비살(100g)	17000.0
6528	한우모듬세트(1인분 100g)	5800.0
6531	한우 토시살(1인분 100g)	13500.0
6532	한우 안창살(1인분 100g)	13500.0
6573	순한우우거지찌개	11000.0

309 rows × 2 columns

그림 7 한우 데이터프레임

menu_df[menu_df['품목'].str.contains('불고기')]

	품목	가격
70	닭불고기	17000.0
126	불고기퀘사디아	14900.0
198	한우 불고기정식	11000.0
222	오리불고기	25000.0
276	은복불고기(1인분)	13000.0
...	...	...
6554	버섯불고기(1인분)	11000.0
6580	청둥오리불고기	40000.0
6596	메기불고기	40000.0
6608	흑염소 돌판불고기(2인분)	40000.0
6612	오리백숙불고기	45000.0

242 rows × 2 columns

그림 8 불고기 데이터프레임

menu_df[menu_df['품목'].str.contains('정식')]

	품목	가격
40	정담정식	13000.0
139	정식	6000.0
198	한우 불고기정식	11000.0
200	한우 갈비찜정식	19000.0
281	특선연요리정식(1인)	20000.0
...	...	...
6341	돈까스정식	11900.0
6344	어린이정식	6900.0
6494	정강희특정식	15000.0
6515	고등어정식(2인이상)	7000.0
6582	연잎밥 장아찌정식	15000.0

252 rows × 2 columns

그림 9 정식 데이터프레임

- 한우가 포함된 음식의 평균 가격: 22,697 원
- 불고기가 포함된 음식의 평균 가격: 17,207 원
- 정식이 포함된 음식의 평균 가격: 17,028 원

4. 본인이 제시한 2번, 3번 결과에 부합하는 맛집을 최종적으로 최대 3곳 선정하여 제시하라. (논리적 근거를 명확히 밝히라.)

○ 최종적인 맛집을 선정하기 위하여, 몇 가지 가정을 설정한 뒤 이를 반영할 수 있는 맛집 점수를 계산하는 계산식을 만들어 음식점 별로 비교하여, 최종 맛집을 선정함.

○ 2번 문제와 3번 문제에 대한 결과를 반영할 수 있도록 가정을 설정하고, 가정에 부합하는 경우 가중치를 주는 방식으로 계산함.

○ 가정

- 음식 가격이 비싼만큼 맛이 뛰어나 사람들이 좋아하고, 가격이 싼 만큼 가성비가 뛰어나 사람들이 좋아한다(음식 가격을 기준으로 맛집 점수 모델을 작성).
- 3번 문제의 결과 선정된 메뉴는 사람들의 선호도가 반영된 결과이다(사람들은 한우, 불고기, 정식을 특히 좋아한다).
- 식당 메뉴가 다양할수록 사람들이 좋아할만한 음식이 존재할 확률이 높다.

○ 가정에 대한 가중치 적용 방식

- 식당 메뉴 평균 가격이 카테고리별 전체 평균 가격에서 높거나 낮은 경우 가중치를 부여함
- 3가지 음식(한우, 불고기, 정식)의 요소가 포함되는 메뉴를 가진 식당의 경우 각각 추가적으로 가중치를 부여함.
- 특히, 맛있는(비싼)음식이 더 효과를 받을 수 있도록 3가지 음식의 합계에서 각 음식의 가격의 비율만큼의 가중치를 추가함(한 음식 메뉴가 중복된 경우 가장 큰 것을 적용)

○ 최종 맛집을 계산하는 계산식

$$\left|\log\left(0.1+\frac{\text{카테고리별 개별식당 가격 점수}}{\text{카테고리별 전체식당 평균 가격}}\right)\right|$$

- 카테고리별 전체 식당의 평균 가격보다 높은 경우 해당 음식점의 개별 음식 메뉴에 대한 가중치를 추가로 더함(한우:40%, 불고기:30%, 정식:30% 추가)

$$\text{한우가중치} = \frac{\text{한우평균가격}}{\text{한우평균가격}+\text{불고기평균가격}+\text{정식평균가격}} \fallingdotseq 40\%$$

$$\text{불고기가중치} = \frac{\text{불고기평균가격}}{\text{한우평균가격}+\text{불고기평균가격}+\text{정식평균가격}} \fallingdotseq 30\%$$

$$\text{정식가중치} = \frac{\text{정식평균가격}}{\text{한우평균가격}+\text{불고기평균가격}+\text{정식평균가격}} \fallingdotseq 30\%$$

- 카테고리별 전체 식당의 평균 가격보다 낮은 경우 해당 음식점의 개별 음식 메뉴에 대한 가중치를 추가로 할인함

$$\text{한우가중치} = \frac{1}{2}\times\frac{\text{한우평균가격}}{\text{한우평균가격}+\text{불고기평균가격}+\text{정식평균가격}} \fallingdotseq 20\%$$

$$\text{불고기가중치} = \frac{1}{2}\times\frac{\text{불고기평균가격}}{\text{한우평균가격}+\text{불고기평균가격}+\text{정식평균가격}} \fallingdotseq 15\%$$

$$\text{정식가중치} = \frac{1}{2}\times\frac{\text{정식평균가격}}{\text{한우평균가격}+\text{불고기평균가격}+\text{정식평균가격}} \fallingdotseq 15\%$$

- log 함수가 1보다 작은 경우 값이 급격하게 낮아지는 경향이 있으므로, 이를 보정하기 위하여 평균가격보다 가격이 낮은 경우의 가중치를 절반으로 낮춤

○ 구현

- 음식점 별 카테고리 확인 후, 각 카테고리별 가격을 확인할 수 있는 딕셔너리 작성 후 카테고리별 음식점 평균 가격 저장

```
df['FD_CS'].value_counts()
```

```
한식            700
양식             41
일식             40
디저트/베이커리       38
중식             36
세계요리           34
전통차/커피전문점      19
퓨전/뷔페          11
특별한 술집          6
Name: FD_CS, dtype: int64
```

그림 10 맛집 카테고리 확인

```
### 카테고리별전체식당평균가격
category_dict = {'한식': [],
                 '양식': [],
                 '세계요리': [],
                 '일식': [],
                 '중식': [],
                 '디저트/베이커리': [],
                 '전통차/커피전문점': [],
                 '특별한 술집': [],
                 '퓨전/뷔페': []}

for i in range(len(store_df)):
    category = df['FD_CS'][i]
    category_dict[category].append(store_df[i]['가격'].mean())
```

```
category_dict_mean_price = {}

for i in category_dict:
    category_dict_mean_price[i] = np.nanmean(category_dict[i])
```

```
category_dict_mean_price
```

```
{'한식': 19841.000682209422,
 '양식': 23945.30892549185,
 '세계요리': 17996.95079718436,
 '일식': 37225.54934381857,
 '중식': 19790.487213403878,
 '디저트/베이커리': 5355.0066000066,
 '전통차/커피전문점': 6808.876137712702,
 '특별한 술집': 13831.349206349207,
 '퓨전/뷔페': 36110.416666666664}
```

그림 11 각 음식 카테고리별 가격 확인

- 음식점 메뉴의 개별 가격을 가지고 와서, 해당 음식점 메뉴의 평균가격이 음식점이 속한 카테고리의 평균 가격보다 높은지 낮은지 확인함
- 그 후, 음식점의 가격이 높은 경우 음식점의 개별 메뉴가 ['한우', '불고기', '정식']을 포함하는 경우에 해당 음식 가격의 가중치를 더하고, 반대로 낮은 경우 가중치를 포함하여 할인함.

```
: ### 카테고리별개별식당가격점수

adjust_store_score = []

for i in range(len(store_df)):
    store_mean = store_df[i]['가격'].mean()
    adjust_price = []
    category = df['FD_CS'][i]

    for menu, price in zip(store_df[i]['품목'], store_df[i]['가격']):
        if store_mean >= category_dict_mean_price[category]:
            if '한우' in menu:
                adjust_price.append(price + price*0.4)
            elif ('불고기' or '정식') in menu:
                adjust_price.append(price + price*0.3)
            else:
                adjust_price.append(price)
        else:
            if '한우' in menu:
                adjust_price.append(price - price*0.2)
            elif ('불고기' or '정식') in menu:
                adjust_price.append(price - price*0.15)
            else:
                adjust_price.append(price)

    adjust_price = np.nanmean(adjust_price)
    adjust_store_score.append([category, adjust_price])
```

```
C:\Users\user\AppData\Local\Temp\ipykernel_11652\358722784.py:26: RuntimeWarning: Mean of empty slice
  adjust_price = np.nanmean(adjust_price)
```

그림 12 가중치가 부여된 음식점별 최종 점수를 산정하는 코드

- 최종 맛집 점수를 계산할 수 있는 함수를 작성

```
def calc_score(score, category):
    return abs(np.log((score/category_dict_mean_price[category])+0.1))
```

그림 13 최종 맛집 점수 계산 함수

- 만들어진 개별 음식점별 계산된 점수와, 음식점 카테고리별 평균 가격을 활용하여 맛집 점수를 계산한 뒤 음식점별 내림차순 정렬

```
rank_df = pd.DataFrame(adjust_store_score, columns=['category', 'score'])
```

```
rank_df['rank_score'] = [calc_score(rank_df['score'][i], rank_df['category'][i]) for i in range(len(rank_df))]
```

```
rank_df.sort_values(by=['rank_score'], ascending=False)[:5]
```

	category	score	rank_score
225	퓨전/뷔페	2000.000000	1.861845
661	세계요리	90420.000000	1.633971
623	한식	2083.333333	1.584738
682	한식	91666.666667	1.551822
491	한식	89600.000000	1.529507

그림 14 5개의 맛집 점수 상위권 맛집 추출

- 이 중, 가장 상위의 225번 음식점의 경우 음식점의 메뉴 가격이 제대로 입력되지 않아 데이터의 오차가 있음을 확인하여, 최종적으로 하위 661번 음식점, 623번 음식점, 682번 음식점을 선정함

- 최종 선정 음식점:

아트리움(수성구, 세계요리),
윤옥연할매떡볶이(수성구, 한식)
신기동대게나라(달서구, 한식)

```
print(df.iloc[661, :])
print(df.iloc[661, :]['MNU'])
```

```
cnt                                                  129
OPENDATA_ID                                          147
GNG_CS                                 대구광역시 수성구 범어동 143-19
FD_CS                                              세계요리
BZ_NM                                              아트리움
TLNO                                        053-754-3111
MBZ_HR                                      11:00 - 23:00
SEAT_CNT                                           120석
PKPL                                                50대
HP                                      www.atriumkorea.co.kr/
PSB_FRN                                 일본어 기타 (인도, 네팔)
BKN_YN                                              가능
INFN_FCL                                           불가능
BRFT_YN                                            불가능
DSSRT_YN                                            가능
MNU          (런치) 아트리움런치코스 : 안심스테이크 74,800원 채끝등심스테이크 64,900...
SMPL_DESC    주택가의 전원같은 별장, 편안한 분위기가 있는 음식점으로 아름다운 전경, 야경과 함...
SBW                                 지하철 2호선 수성구청역 1번 출구 약 530m
BUS                          버스 정류장은 `도시철도수성구청역북편1` 정류장이 가장 가깝습니다.
LOCATION                                            수성구
Name: 661, dtype: object
(런치) 아트리움런치코스 : 안심스테이크 74,800원 채끝등심스테이크 64,900원 안심스테이크 45,100원
탑스터코스 : 라이브탑스터와안심 89,100원
(디너) 아트리움코스 : 라이브탑스터와안심 119,900원 안심스테이크 99,000원
탑스터코스 : 라이브탑스터와안심 89,100원 캐나다산라이브탑스터 64,900
스페셜코스 : 안심스테이크 79,200원 채끝등심스테이크 70,400원
```

그림 15 맛집 1위 아트라움 정보

```
print(df.iloc[623, :])
print(df.iloc[623, :]['MNU'])
```

```
cnt                                                   91
OPENDATA_ID                                          505
GNG_CS                                 대구광역시 수성구 수성동4가 1120-2
FD_CS                                              한식
BZ_NM                                      윤옥연할매떡볶이
TLNO                                        053-756-7597
MBZ_HR                                      10:00 - 22:00
SEAT_CNT                                           32석
PKPL                                      4대외 노상주차
HP                                                 없음
PSB_FRN                                 영어 기타 (인도, 네팔)
BKN_YN                                  전화주문 가능,자리예약불가
INFN_FCL                                           불가능
BRFT_YN                                            불가능
DSSRT_YN                                           불가능
MNU          떡볶이 1,000원 \r\n만두 1,000원 \r\n오뎅 1,000원 \r\n김밥 ...
SMPL_DESC    대구특미음식 [매운떡볶이]선정 업체\r\n대구에서 한때 `신천할매떡볶이`로 불리우던...
SBW                                 지하철 2호선 대구은행역 1번 출구 약 810m
BUS                          버스 정류장은 `수성4가동주민센터앞` 정류장이 가장 가깝습니다.
LOCATION                                            수성구
Name: 623, dtype: object
떡볶이 1,000원
만두 1,000원
오뎅 1,000원
김밥 2,500원
라면 4,000원
순대 3,000원
```

그림 16 맛집 2위 윤옥연할매떡볶이 정보

```
print(df.iloc[682, :])
print(df.iloc[682, :]['MNU'])
```

```
cnt                                                          19
OPENDATA_ID                                                1649
GNG_CS                                  대구광역시 달서구 월성동 1317-1
FD_CS                                                        한식
BZ_NM                                                 신기동대게나라
TLNO                                               053-628-2277
MBZ_HR                                            11:30 ~ 23:30
SEAT_CNT                                               96석(룸3)
PKPL                                                       50대
HP                                               pf.kakao.com/_kpSnT
PSB_FRN                                      중국어 &lt;BR&gt;
BKN_YN                                                       가능
INFN_FCL                                                   불가능
BRFT_YN                                                    불가능
DSSRT_YN                                                     가능
MNU           대게(1kg) 96,000원\r\n킹크랩(1kg) 139,000원\r\n랍스타(1k...
SMPL_DESC     신기동 대표의 이름을 걸고 대구 최고의 대게, 킹크랩, 랍스타를 착한 가격에 제공하...
SBW                          지하철 1호선 월배역 4번 출구에서 도보로 약 454m 거리.
BUS                          버스 정류장은 `태왕아너스베스트앞` 정류장이 가장 가깝습니다.
LOCATION                                                   달서구
Name: 682, dtype: object
대게(1kg) 96,000원
킹크랩(1kg) 139,000원
랍스타(1kg) 87,000원
홍게(1kg) 69,000원
대게세트 144,000원 ~ 336,000원
물회 15,000원
```

그림 17 맛집 3위 신기동 대게나라 정보