

데이터 분석

전통주 잘 어울리는 음식 군집화

- 데이터 수집
 - 네이버 전통백과 참고
- 데이터 전처리
 - csv 파일에서 잘 어울리는 음식 컬럼 가져오기
 - KoNLPy(한국어 정보처리를 위한 파이썬 패키지)를 활용하여 메뉴 단어의 형태소 분석
 - NLP(Natural Language Processing: 자연어 처리)란 텍스트에서 의미있는 정보를 분석, 추출하고 이해하는 일련의 기술집합 ex) 애플의 Siri, 구글 번역기
 - KoNLPy는 한국어 형태소를 효과적으로 추출하기 위한 오픈소스 파이썬 패키지
 - KoNLPy에서는 품사를 태깅하기 위한 클래스들을 제공하는데 Kkma, Komoran, Hannanum, Okt, Mecab 등이 있음.
 - 그 중 Kkma (꼬꼬마) 클래스는 실행 속도가 가장 느리다는 단점이 있지만 다른 클래스들에 비해 더 잘게 품사를 쪼개줌.
 - 우리가 분석하고자 하는 음식의 경우 데이터의 개수가 많다고 할 수 없고 같은 재료를 쓰더라도 구이, 찌개, 볶음 등 바リエ이션이 다양하기 때문에 더 잘게 쪼개줄 필요가 있다고 생각이 들어 kkma 클래스 사용

```

noun_list = []
for content in (df['잘 어울리는 음식']): # 잘 어울리는 음식 컬럼들 데이터 찾기
    nouns = kkma.nouns(content) # 데이터 돌면서 명사 형태로 쪼개놓기
    noun_list.append(nouns) # 쪼개 놓은 명사들 리스트에 넣기

df['nouns'] = noun_list
df.head()

```

✓ 0.5s

| | 전통주명 | 잘 어울리는 음식 | nouns |
|---|------------|-------------------|---------------|
| 0 | 구름을 벗삼아 | 가리비 | [가리비] |
| 1 | 해창 생막걸리 6도 | 간재미 찜 | [간재미, 찜] |
| 2 | 도문대작 생막걸리 | 갈비찜 | [갈비찜] |
| 3 | 이동 생 쌀 막걸리 | 갈비찜 | [갈비찜] |
| 4 | 등정춘 | 갈비찜, 단맛이 약간 있는 음식 | [갈비찜, 단맛, 음식] |

- 품사 쪼개기가 다 되었으면 군집화를 시켜주기 위해 상대적으로 중요한 단어에 가중치를 주어야 함. 따라서 상대적으로 희귀하면서도 빈도가 높은 것을 찾아 주는 TF-IDF 알고리즘을 사용하여 가중치 부여하고 벡터값을 만들어준다.

```

from sklearn.feature_extraction.text import TfidfVectorizer # 단어의 가중치를 부여하는 TF-IDF 알고리즘으로 벡터화

text = [" ".join(noun) for noun in df['nouns']]

tfidf_vectorizer = TfidfVectorizer(min_df=5, ngram_range=(1, 5)) # min_df는 전체 문서에서 해당 단어가 나오는 빈도 수, 즉 단어가 5개 이상 나온 것들만 추출
# ngram_range는 단어 묶음에 대한 범위, 즉 1개에서 5개까지 쪼개진 명사를 한 묶음으로 취급
tfidf_vectorizer.fit(text) # 벡터 출력하기 위해 학습시킴
vector = tfidf_vectorizer.transform(text).toarray() # 어레이로 만들
print(vector)

```

✓ 0.3s

```

[[0. 0. 0. ... 0. 0. 0.]
 [0. 0. 0. ... 0. 0. 0.]
 [0. 1. 0. ... 0. 0. 0.]
 ...
 [0. 0. 0. ... 0. 0. 0.]
 [0. 0. 0. ... 0. 0. 0.]
 [0. 0. 0. ... 0. 0. 0.]]

```

• 데이터 분석

◦ 가중치 벡터들간의 유사도를 판단해주기 위해서 DBSCAN 군집화 알고리즘 사용

- 군집화 알고리즘에는 대표적으로 K-means와 DBSCAN이 있다. K-means는 몇 개의 군집으로 나뉘야 할 지 미리 정해야 하는데 이를 정하는 것은 어렵고 특이값의 경우에는 걸러주는 것이 편한데 거르지 못함
- 따라서 두 가지 어려움을 해소할 수 있는 DBSCAN 알고리즘을 사용하여 군집화하여 결과를 얻음

```

from sklearn.cluster import DBSCAN # 군집화를 하기 위한 DBSCAN 알고리즘
import numpy as np

vector = np.array(vector)
model = DBSCAN(eps=0.4, min_samples=5, metric='cosine') # eps는 벡터들을 묶는 원의 반지름. 크기가 작을수록 좁은 범위를 택함.
# min_samples는 핵심 군집인지 판단하기 위한 기준 여기서의 원 안에 5개 미만의 벡터가 있는 경우 불필요한 군집이라 판단하여 제외시킴. 유사도는 코사인 유사도 채택
result = model.fit_predict(vector) # 학습시킴

df['result'] = result
✓ 0.6s

for cluster_num in set(result):
    if (cluster_num == -1 or cluster_num == 0): # 결과값이 -1이나 0인 경우 잡음 데이터이기 때문에 패스
        continue
    else:
        print("cluster num : {}".format(cluster_num)) # 결과값 출력
        temp_df = df[df['result'] == cluster_num]
        for menu in temp_df['잘 어울리는 음식']:
            print(menu)
        print('-----')
✓ 0.1s

```

- 결과

```
cluster num : 1  
감자탕, 김치찌개, 매운 제육볶음  
닭볶음탕  
닭볶음탕  
매콤한 요리(꼬막무침, 닭볶음탕)  
쭈꾸미 볶음  
쭈꾸미 볶음
```

```
-----  
cluster num : 2  
곱창 전골, 연포탕  
두부전골  
두부전골  
맑은 소고기국, 미역국, 계란국  
버섯 전골  
버섯 전골류  
버섯구이  
버섯요리, 송이버섯 전골  
버섯전골  
소고기 구이  
소고기 구이  
소고기 구이  
소고기 소금구이  
소고기 전골  
소고기류  
소고기류
```

```
-----  
cluster num : 3  
과일 샐러드  
샐러드  
샐러드  
샐러드  
샐러드  
샐러드  
샐러드  
샐러드, 두부  
샐러드, 카르파초, 가벼운 튀김류  
샐러드, 해산물, 케이크, 마카롱  
애플파이, 과실 샐러드
```



```
-----  
cluster num : 6  
민물 매운탕, 보쌈  
민물고기 매운탕, 수육, 보쌈  
보쌈  
보쌈  
보쌈, 족발  
-----  
cluster num : 7  
생선찜  
생선찜  
양념을 강하게 하지 않은 생선찜  
양념한 생선찜  
양념한 생선찜  
-----  
cluster num : 8  
스테이크  
스테이크  
스테이크  
양꼬치, 향정살, 스테이크  
함박 스테이크, 파스타  
-----  
cluster num : 9  
장어구이  
장어구이  
장어구이  
장어구이  
장어구이  
장어구이, 디저트, 치즈  
-----  
cluster num : 10  
홍어, 오징어  
홍어무침  
홍어무침  
홍어무침  
홍어무침, 오징어무침  
홍어삼합  
홍어삼합  
-----
```

전통주점 메뉴 군집화

- 데이터 수집
 - 더숏닷컴, 네이버 지도 크롤링
- 과정은 전통주 잘 어울리는 음식과 동일

```
cluster num : 1
해물 순두부찌개 죽조림 아롱사태스지수육 칼뱅이 등레해물파전 장땡이 바섯스지전골 보들탕 곰소 짓갈과 육임자 두부 주인암 디쉬
김치찌개 정식 된장찌개 정식 순두부찌개 정식 계란찜 정식 모듬전 고추장불고기정식
김치육배기 된장육배기 순두부찌개 해물계란찜 푸꾸미볶음
다람쥐 순두부찌개
김치찌개 정식 된장찌개 정식 고추장 불고기정식 순두부찌개 정식 해물계란찜 정식
막걸리 김치전 해물파전 해물부추전 심마니전 쥐포 계란말이 계란탕 순두부찌개 두부김치 돼지찌개 참치찌개 알탕 굴뱅이 푸꾸미볶음
모듬전 해물파전 동등전 김치전 부추전 녹두빈대떡 김치찌개 순두부찌개 해물계란찜 어묵탕 알밤막걸리 옥수수막걸리
순두부찌개, 김치 치즈 전 스틱, 불고기메일 썸, 육회 타르타르, 모듬어묵탕, 베이컨 감자전, 부추 새우 전 스틱, 맥적구이, 떡갈비 지짐, 토마토 범벅, 팽이버섯 튀김, 어란 슬라이스, 명란구이, 수제약과, 틀깨라면
새우 부추튀김, 토종 닭 무침, 호랑이 육전, 된장수육 큰상, 마라 순두부찌개
-----
cluster num : 2
하얀연꽃막걸리 꿀막걸리 알밤막걸리
단체체험- 영인연잎주빚기 (15인~) 개인체험- 한지공예 등만들기 체험 단체체험- 막걸리소믈리에Class 단체체험- 누룩전만들기(15인~) 개인체험 - 쌀누룩임욕제 만들기
막걸리
얼음막걸리
막걸리 우리술 Signature Cocktail
말갈차림 1인 담은 꽃장 오늘의막걸리3잔
막걸리 프리미엄막걸리
천원막걸리
대표막걸리 도문대작막걸리
파전, 막걸리
막걸리, 파전
막걸리, 감자전 등
퓨전 한식, 막걸리 칵테일 등
막걸리, 두부, 도토리묵
귀빈모듬전, 막걸리 등
-----
cluster num : 3
이달의메뉴
안산 최대 지역 전통주 / 제철 메뉴
전통주 와인
계절숙성회, 해산물, 제철메뉴
제철 생선 구이
제철 재료로 일주일에 한번 바뀌는 한식 코스요리
메뉴변동, 우리술
```

```

-----
cluster num : 4
번게기탕 두부탕 돼지감치찌개 코다리조림 닭볶음탕 가지미구이 감치구이
김치찌개 대 매운돼지갈비찜 대 목은지갈비찜 대 목은지닭볶음탕 굴보쌈 대
닭다리완 술살 닭볶음탕 대창술국 육회+카나페
한우 도가니 어묵전골, 웅치기 닭, 늘 닭볶음탕, 늘 떡볶이
숙주 들국 닭볶음탕, 베이컨 김치전
-----
cluster num : 5
모듬전 등대전 녹두전 김치전 감자전 등대탕 메밀전병 닭발
누룽지감자전 두부김치 콩김치볶음 무해닭발 굴뱅이 무침 해물파전 매생이굴전 정구지전 김치전 아채고추전 돌미나리전 메밀전병 제육볶음 오돌배 볶음 낚지볶음
해물파전 육전파전이 녹두전 부추전 메밀전병
베이컨감자전 오짜렐라김치전 굴뱅이소면 메밀전병
한돈 보쌈, 한우 육회, 치즈 감자전, 등심튀김, 메밀전병, 해물파전, 감태밥
-----
cluster num : 6
육회 아롱사태수육 소고기찌개 참소라무침 푸꾸미볶음 치즈감자전 닭발편육 청어알두부쌈
탈콩 오일 파스타 청어알 파스타 항정살 보쌈 미나리 파스타 치즈감자전 삼계 닭튀김 두부크림을 곁들인 갑오징어 버터구이 문어 세비체
치즈감자전 파리고추특살구이 전통주 쫄면고기 제주막창순대 고기김치전 간재미무침 투블한우육전 오징어뿔초전 수제닭발편육 목은지참치말이 외할머니호박찌개
치즈감자전, 반상세트
순살 닭볶음탕, 크림새우볶음, 치즈감자전, 스텔 고추장 찌개
-----
cluster num : 7
한우육회와 참문어
한돈안심 직화구이 트리플리조프 참란찜닭밥 스키치에그 사천식 닭냉채 닭은 바지락국 두부구이 1++한우육회와 김부각
육회김밥 육회 한우육전
우니와 단새우, 제철조개술찜, 2+한우육회, 오물렛 감자전, 명란구이
한우육회, 닭묵은지찌개
-----
cluster num : 8
오징어무침 도토리무침 굴뱅이무침 옥사발 가오리찜
오징어순대 모듬전 매콤짜먹전 베이컨김치치즈전 불고기치즈파전 치즈감자전 허니치즈고구마전 해물파전 갑오징어숙회 굴뱅이무침 시골돼지예호박찌개 삼겹두부김치
국산전두부와볶음김치 *SIGNATURE 전당모듬전 *BEST 옛날모듬전 체다들볶감자전 전당찜떡볶이 굴뱅이무침 한돈육전 흥어무침 오짜렐라김치전 내맘대로세가지전 송명섬막걸리 해창막걸리9도 칠곡신통막걸리 금정산성막걸리 복-
A코스 B코스 C코스 D코스 과일 굴뱅이무침-대 굴뱅이무침-소 굴뱅이 클회 오삼볶음 치킨너겟 과일화채 육포 마른오징어 노가리 모듬소세지볶음
등래파전, 등래고등집, 굴뱅이무침
-----

```