



# Code States Project 1

## 프로젝트 결과 작성 가이드(초안)

# 외식 산업 예비 창업자를 위한 데이터 시각화 및 분석 프로그래밍

15 - 윤종률

# 목차

- 01. 프로젝트 개요
- 02. 프로젝트 구성
- 03. 프로젝트 수행 절차 및 방법
- 04. 프로젝트 수행 결과
- 05. 자체 평가 의견

# 1. 프로젝트 개요

---

## ▶ [프로젝트 개요]

1. 외식 관련 기사를 수집하고 이를 통해 키워드 및 현황 분석을 계획
2. 외식 산업에서 음식점 업종, 평균 영업 시간, 사업장 크기, 1일 평균 방문객 등의 데이터를 제공해서 시각화를 계획
3. 데이터를 시각화하여 외식 창업을 시작하려는 사람에게 정보를 제공해주어 도움이 되고자 프로젝트를 설계

## 02. 프로젝트 구성

---

### ▶ [프로젝트 구성]

- 담당 업무 : 데이터 수집, 가공 및 적재, 시각화

훈련생	담당 업무
윤종률	▶ 데이터 수집, 가공 및 적재, 시각화

---

### 03. 프로젝트 수행 절차 및 방법

#### ▶ [프로젝트 수행 절차 및 방법]



## ▶ 데이터 수집

크롤링을 실패해서 데이터 수집을 위해  
하나의 기사를 선택해 엑셀파일에 직접  
입력

Date	Title	text
2022/12/26	코로나 여지나해	코로나19 여파에도 지역 내 프랜차이즈 기업점 수가 급증한 것으로 나타났다.코로나 장기화 속에 위험 부담이 큰 개인창업보다는 상대적으로 안정적인 프랜차이즈에 창업 수요
	1일 평균방문객	2023-01-11 오후 9:56 Microsoft Excel 실효표... 60KB
	news	2023-01-12 오전 3:14 Microsoft Excel 실효표... 4KB
	매출액	2023-01-11 오후 10:04 Microsoft Excel 실효표... 12KB
	불용어리스트	2023-01-12 오전 4:05 Microsoft Excel 실효표... 7KB
	사업장 크기	2023-01-11 오후 10:09 Microsoft Excel 실효표... 15KB
	운영형태	2023-01-11 오후 10:15 Microsoft Excel 실효표... 2KB
	음식점 업종별	2023-01-11 오후 10:11 Microsoft Excel 실효표... 16KB
	주요 고객 현황	2023-01-11 오후 10:16 Microsoft Excel 실효표... 32KB
	평균 영업시간	2023-01-11 오후 10:12 Microsoft Excel 실효표... 27KB

## 04. 프로젝트 수행 결과

### 데이터 적재

The screenshot shows the DBeaver interface with a project named 'CSVD8파일'. The left sidebar lists various database tables under the '데이터베이스' (Database) folder. The main pane displays the '설정' (Settings) for the 'CSVD8파일' connection, including 'Origin: local', '연결 시간: 23.1.12 오전 5:57', and '연결 유형: Development'. Below the settings, a table lists the database tables and their types.

테이블명	테이블 타입	테이블 설명
1일 평균방문객	TABLE	
news	TABLE	
매출액	TABLE	
불용어리스트	TABLE	
사업장 크기	TABLE	
운영형태	TABLE	
음식점 업종별	TABLE	
주요 고객 현황	TABLE	
평균 영업시간	TABLE	

The screenshot shows the DBeaver interface with a SQL query executed against the 'sale' table. The query is: `SELECT * FROM sale s`. The results are displayed in a table with 6 columns: '특성별(1)', '특성별(2)', '특성별(3)', '특성별(4)', '방문 고객수', and '방문 고객및배달테이크아웃수'. The results show 16 rows of data. Below the results, the 'Script' editor shows the same query: `SELECT * FROM sale s`.

특성별(1)	특성별(2)	특성별(3)	특성별(4)	방문 고객수	방문 고객및배달테이크아웃수
전체	소계	소계	소계	방문 고객수	20명 미만 (%)
전체	소계	소계	소계	방문 고객수	20~30명 미만 (%)
전체	소계	소계	소계	방문 고객수	30~50명 미만 (%)
전체	소계	소계	소계	방문 고객수	50~100명 미만 (%)
전체	소계	소계	소계	방문 고객수	100명 이상 (%)
전체	소계	소계	소계	방문 고객수	평균 (명)
전체	소계	소계	소계	배달 수	없음 (%)
전체	소계	소계	소계	배달 수	10인분 미만 (%)
전체	소계	소계	소계	배달 수	10~20인분 미만 (%)
전체	소계	소계	소계	배달 수	20~30인분 미만 (%)
전체	소계	소계	소계	배달 수	30~50인분 미만 (%)
전체	소계	소계	소계	배달 수	50인분 이상 (%)
전체	소계	소계	소계	배달 수	평균 (인분)
전체	소계	소계	소계	테이크아웃 수	없음 (%)
전체	소계	소계	소계	테이크아웃 수	10인분 미만 (%)
전체	소계	소계	소계	테이크아웃 수	10~20인분 미만 (%)

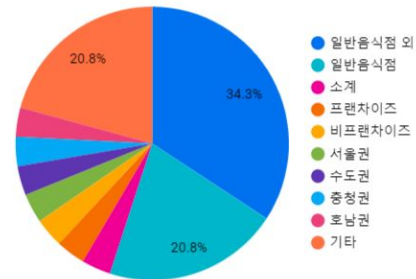


## 04. 프로젝트 수행 결과

### 데이터 시각화

특성별(1)	시점 ▾
1. 업종별	
2. 지역별	
3. 매출액규모별	
4. 운영형태별	
5. 전체	

특성별(1)에 해당하는 범위가 업종, 지역, 매출액규모, 운영형태로 나뉜다.



특성별(3)	Record Count ▾
1. 소계	300
2. 기타 음식점업	400
3. 한식	420
4. 중식	440
5. 일식	460
6. 서양식	480
7. 기타 외국식	500
8. 출장·이동음식점업	520
9. 주점업	540
10. 비알콜 음료점업	560
11. 기관 구내식당업	578



# 04. 프로젝트 수행 결과

## 데이터 시각화



방문고객및배달테이크... 원데이터 ▾	
1. 방문 고객수	4,267.7
2. 테이크아웃 수	3,334.9
3. 배달 수	3,241.1

1 - 3 / 3 < >



방문고객및배달테이크... Record Count ▾	
1. 없음 (%)	58
2. 10인분 미만 (%)	58
3. 10~20인분 미만 (%)	58
4. 20~30인분 미만 (%)	58
5. 30~50인분 미만 (%)	58
6. 50인분 이상 (%)	58
7. 평균 (인분)	56
8. 20명 미만 (%)	29
9. 20~30명 미만 (%)	29
10. 30~50명 미만 (%)	29
11. 50~100명 미만 (%)	29
12. 100명 이상 (%)	29

1 - 13 / 13 < >

## 04. 프로젝트 수행 결과

---

### ▶ 데이터 시각화

데이터 스튜디오 : [https://datastudio.google.com/s/ksePbl45\\_As](https://datastudio.google.com/s/ksePbl45_As)

## 05. 자체 평가 의견

### ▶ [자체 평가 의견 작성 가이드]

1. 데이터를 수집하는 것 부터 많은 난관을 겪었다. 예시에 나와있는 사이트 빅카인드에서 크롤링을 하려고 검색을 했는데 생각보다 데이터의 양도 많았고, 그것을 크롤링 하는 방법이 숙지가 안되다 보니 시간만 날리다 결국은 포기하고 계획으로만 남게 되었습니다. 그리고 알맞는 데이터를 찾는 것이 너무 어려웠고, 지금도 맞는 데이터인지 확신이 들지 않습니다. (코치분들이 피드백 해주실 때 크롤링이나 셀레니움을 쓰는 방법에 대해 자세히 알려주시면 좋겠습니다. 그래서 꼭 보완해서 데이터 수집하는 연습을 하고 싶습니다.)
2. 실무에서 활용 가능하냐고 물어보면 어렵다고 할 수 있을 것 같습니다. 웹페이지를 제작하고 꾸미는 것에 대해 공부를 하지 않아서 예시로 보여준 사이트들을 봐도 구현을 할 엄두도 나지 않아서 하지 못했는데 공부 할 수 있는 사이트들 중에서 간단한 웹사이트보단 구체적이고 실생활에서도 쓸 수 있는 웹사이트를 만드는 곳을 알려주시면 실습을 하거나 클론 코딩을 통해 공부해서 제것으로 만드는 연습을 하고 싶습니다.
3. 공부를 하다보니 데이터를 시각화하려고 해도 제가 원하는 값들만 선택해서 시각화가 되지 않아 많은 어려움을 겪었습니다. 지금도 데이터 스튜디오를 통해서 원하는 값들을 보여주지 못해 속상했습니다. 태블로를 공부해서 더욱 시각화를 할 수 있도록 노력하겠습니다.

## 05. 자체 평가 의견

---

### ▶ [자체 평가 의견 작성 가이드]

4. 기사들을 수집해서 기사 내용을 형태소로 분할 하고, 가장 많이 사용한 단어를 분석하려고 했는데, 도중에 에러가 나서 하지 못했습니다. 왜 구현이 안됐는지 궁금하고, 그것을 해결 할 수 있도록 많이 공부해서 시도 해보려고 합니다.

마지막장에 오류 구문을 올려 놓도록 하겠습니다.

5. cp1을 통해서 많이 부족하다는 것을 느꼈고, 데이터 엔지니어를 할 수 있을지 많이 고민이 됐습니다. 다른 프로젝트 주제를 할까 많이 고민도 했지만, 취업을 위해서 선택한 만큼 더욱 노력해서 하도록 하겠습니다.

## 05. 자체 평가 의견

### ▶ [자체 평가 의견 작성 가이드]

```
#함수 만들기
def freq_word(wds):
    fr_w = dict()
    for wd in wds:
        try : fr_w[wd] = fr_w[wd]+1#같은 거 찾으면 +1
        except : fr_w[wd] = 1 #그대로 진행
    fre_df = pd.DataFrame.from_dict(fr_w,orient='index')#딕셔너리를 데이터 프레임으로 변환
    fre_df.columns = ['frequency']
    fre_df = fre_df.sort_values(['frequency'],ascending=False)#오름차순으로 정렬
    return fr_w, fre_df
```

freq\_word(wds)

-----  
TypeError Traceback (most recent call last)  
<ipython-input-58-bd4a6d8bac19> in freq\_word(wds)

```
10     for wd in wds:
--> 11         try : fr_w[wd] = fr_w[wd]+1#같은 거 찾으면 +1
12         except : fr_w[wd] = 1 #그대로 진행
```

TypeError: unhashable type: 'list'

During handling of the above exception, another exception occurred:

TypeError Traceback (most recent call last)

1 frames

<ipython-input-58-bd4a6d8bac19> in freq\_word(wds)

```
10     for wd in wds:
11         try : fr_w[wd] = fr_w[wd]+1#같은 거 찾으면 +1
--> 12         except : fr_w[wd] = 1 #그대로 진행
13     fre_df = pd.DataFrame.from_dict(fr_w,orient='index')#딕셔너리를 데이터 프레임으로 변환
14     fre_df.columns = ['frequency']
```