

안녕하세요^^

AIVLE '서울시 생활정보 기반 대중교통 수요 분석' 과정에 오신 여러분을 환영합니다.

- 본 과정에서는 실제 사례와 데이터를 기반으로 문제를 해결하는 전체 과정을 자기 주도형 실습으로 진행해볼 예정입니다.
- 앞선 교육과정을 정리하는 마음과 지금까지 배운 내용을 바탕으로 문제 해결을 해볼게요!
- 미니 프로젝트를 통한 문제 해결 과정 'A에서 Z까지', 지금부터 시작합니다!

데이터 분석부터 먼저 시작해보겠습니다.

"구별 업종 등록 데이터" 를 확인해 보도록 하겠습니다

In [244...

필요 라이브러리부터 설치합니다.

%pip install pandas seaborn

```
Requirement already satisfied: pandas in c:\users\user\anaconda3\lib\site-packages (2.0.3)
Requirement already satisfied: seaborn in c:\users\user\anaconda3\lib\site-packages (0.13.2)
Requirement already satisfied: python-dateutil>=2.8.2 in c:\users\user\anaconda3\lib\site-packages (from pandas) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in c:\users\user\anaconda3\lib\site-packages (from pandas) (2023.3.post1)
Requirement already satisfied: tzdata>=2022.1 in c:\users\user\anaconda3\lib\site-packages (from pandas) (2023.3)
Requirement already satisfied: numpy>=1.21.0 in c:\users\user\anaconda3\lib\site-packages (from pandas) (1.24.3)
Requirement already satisfied: matplotlib!=3.6.1,>=3.4 in c:\users\user\anaconda3\lib\site-packages (from seaborn) (3.8.3)
Requirement already satisfied: contourpy>=1.0.1 in c:\users\user\anaconda3\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (1.0.5)
Requirement already satisfied: cycycler>=0.10 in c:\users\user\anaconda3\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (0.11.0)
Requirement already satisfied: fonttools>=4.22.0 in c:\users\user\anaconda3\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (4.25.0)
Requirement already satisfied: kiwisolver>=1.3.1 in c:\users\user\anaconda3\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (1.4.4)
Requirement already satisfied: packaging>=20.0 in c:\users\user\anaconda3\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (23.1)
Requirement already satisfied: pillow>=8 in c:\users\user\anaconda3\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (10.0.1)
Requirement already satisfied: pyparsing>=2.3.1 in c:\users\user\anaconda3\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (3.0.9)
Requirement already satisfied: six>=1.5 in c:\users\user\anaconda3\lib\site-packages (from python-dateutil>=2.8.2->pandas) (1.16.0)
Note: you may need to restart the kernel to use updated packages.
```

기본전제

- 처음에 제공되는 기본/추가 데이터는 'raw_data' 폴더에 있습니다.

[기본 데이터]

- 1.4 business_type_202401.csv

[데이터 소개]

- 서울 시 구별 등록 업종 상위 10개 데이터(제공 가능한 최신 버전)
 - 택시 운송업 (개인택시 등록자 주거지)
 - 한식일반음식점업
 - 용달 화물자동차 운송업(개인용달 등록자 주거지)
 - 부동산 중개 및 대리업
 - 두발미용업
 - 커피전문점
 - 남녀용 겔옷 및 셔츠 도매업
 - 기타주점업
 - 일반 교과 학원
 - 한식 육류요리 전문점

[변수 소개]

- 다수 업종 상위 1~10

1.데이터 불러오기

모든 미니 프로젝트의 시작은 '데이터 불러오기' 부터라고 할 수 있습니다.

- KeyPoint : 불러오고자 하는 데이터에 따라 자유롭게 변수로 지정할 수 있다.

데이터 프레임을 불러오고 변수로 저장(여기서는 **CSV** 기준으로 진행)

- csv : pd.read_csv("파일이름.csv")
- txt : pd.read_csv("파일이름.csv", sep="구분자")
- xlsx : pd.read_excel('파일이름.xlsx')
- pickle : pd.read_pickle("파일이름.pkl")

[참고] pickle은 파이썬의 모든 객체를 파일로 저장할 수 있는 방법으로 DataFrame,List,Dict

등 모든 객체 저장 가능(특히 sklearn라이브러리를 통해 모델을 학습시키고, 저장할 때 많이 사용)

[실습문제1] 데이터 로딩

- 'business_type_202401.csv'파일을 'seoul_business' 변수에 저장하고 그 데이터를 확인하세요.
 - 데이터 파일 로딩시 참고 사항
 - 구분자(sep)는 ',' 입니다
 - UTF-8 인코더를 사용해 주세요

```
In [245... # 아래에 실습코드를 작성하고 결과를 확인합니다.
import pandas as pd

seoul_business = pd.read_csv('1.4 business_type_202401.csv', sep=',', encoding = 'UTF-8')
```

```
In [246... # 데이터 프레임의 Shape을 확인합니다.
seoul_business.shape
```

```
Out[246]: (451, 22)
```

2.기본 정보 확인 및 클렌징

- 데이터 클렌징 : 결측치, 이상치 등을 제거하여 데이터 분석 결과가 왜곡 되는 문제를 방지하기 위한 정제 과정

[실습문제2] 기본 정보 확인하기

- 'seoul_business' 데이터의 정보를 확인해보세요.
- 'describe', 'info', 'head' 등 전부 활용해 보겠습니다.

```
In [247... # 아래에 실습코드를 작성하고 결과를 확인합니다.
# info()

seoul_business.info()
```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 451 entries, 0 to 450
Data columns (total 22 columns):
#   Column                Non-Null Count  Dtype
---  -
0   자치구                451 non-null   object
1   동                    451 non-null   object
2   다수업종 상위10       451 non-null   object
3   다수업종 상위10.1     451 non-null   object
4   다수업종 상위10.2     451 non-null   object
5   다수업종 상위10.3     451 non-null   object
6   다수업종 상위10.4     451 non-null   object
7   다수업종 상위10.5     451 non-null   object
8   다수업종 상위10.6     451 non-null   object
9   다수업종 상위10.7     451 non-null   object
10  다수업종 상위10.8     451 non-null   object
11  다수업종 상위10.9     451 non-null   object
12  다수업종 상위10.10    451 non-null   object
13  다수업종 상위10.11    451 non-null   object
14  다수업종 상위10.12    451 non-null   object
15  다수업종 상위10.13    451 non-null   object
16  다수업종 상위10.14    451 non-null   object
17  다수업종 상위10.15    451 non-null   object
18  다수업종 상위10.16    451 non-null   object
19  다수업종 상위10.17    451 non-null   object
20  다수업종 상위10.18    451 non-null   object
21  다수업종 상위10.19    451 non-null   object
dtypes: object(22)
memory usage: 77.6+ KB

```

In [248...

```

# 아래에 실습코드를 작성하고 결과를 확인합니다.
# describe()
seoul_business.describe().T

```

Out[248]:

| | count | unique | top | freq |
|--------------|-------|--------|-----|------|
| 자치구 | 451 | 27 | 송파구 | 28 |
| 동 | 451 | 425 | 소계 | 25 |
| 다수업종 상위10 | 451 | 226 | 71 | 7 |
| 다수업종 상위10.1 | 451 | 280 | 11 | 7 |
| 다수업종 상위10.2 | 451 | 176 | 24 | 11 |
| 다수업종 상위10.3 | 451 | 305 | 104 | 6 |
| 다수업종 상위10.4 | 451 | 177 | 56 | 8 |
| 다수업종 상위10.5 | 451 | 180 | 56 | 9 |
| 다수업종 상위10.6 | 451 | 142 | 40 | 11 |
| 다수업종 상위10.7 | 451 | 205 | 54 | 9 |
| 다수업종 상위10.8 | 451 | 121 | 46 | 12 |
| 다수업종 상위10.9 | 451 | 180 | 47 | 10 |
| 다수업종 상위10.10 | 451 | 131 | 13 | 18 |
| 다수업종 상위10.11 | 451 | 239 | 16 | 8 |
| 다수업종 상위10.12 | 451 | 69 | - | 82 |
| 다수업종 상위10.13 | 451 | 147 | - | 82 |
| 다수업종 상위10.14 | 451 | 127 | 18 | 14 |
| 다수업종 상위10.15 | 451 | 189 | 20 | 10 |
| 다수업종 상위10.16 | 451 | 113 | 4 | 15 |
| 다수업종 상위10.17 | 451 | 222 | 42 | 10 |
| 다수업종 상위10.18 | 451 | 107 | 17 | 16 |
| 다수업종 상위10.19 | 451 | 230 | 45 | 9 |

In [249...

```
# 아래에 실습코드를 작성하고 결과를 확인합니다.
# head()
seoul_business.head()
```

Out[249]:

| | 자치구 | 동 | 다수업종상위 10 | 다수업종상위 10.1 | 다수업종상위 10.2 | 다수업종상위 10.3 | 다수업종상위 10.4 | 다수업종상위 10.5 | 다수업종상위 10.6 | 다수업종상위 10.7 | 다수업종상위 10.8 | 다수업종상위 10.9 | 다수업종상위 10.10 | 다수업종상위 10.11 |
|---|-----|-----|--------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|-----------------|-----------------|
| 0 | 자치구 | 동 | 택시운송업 | 택시운송업 | 한식음식점업 | 한식음식점업 | 용달화물자동차운송업 | 용달화물자동차운송업 | 부동산중개대리업 | 부동산중개대리업 | 두발미용업 | 두발미용업 | 커피전문점 | 커피전문점 |
| 1 | 자치구 | 동 | 사업체수 | 종사자수 | 사업체수 | 종사자수 | 사업체수 | 종사자수 | 사업체수 | 종사자수 | 사업체수 | 종사자수 | 사업체수 | 종사자수 |
| 2 | 합계 | 합계 | 47,176 | 76,701 | 29,225 | 100,183 | 28,494 | 29,356 | 24,013 | 42,636 | 18,080 | 36,214 | 16,255 | 57,3 |
| 3 | 종로구 | 소계 | 215 | 215 | 1,548 | 5,912 | 618 | 646 | 549 | 889 | 261 | 480 | 948 | 3,6 |
| 4 | 종로구 | 사직동 | 7 | 7 | 230 | 982 | 12 | 12 | 57 | 96 | 28 | 66 | 124 | 5 |

In [250...]

```
# 아래에 실습코드를 작성하고 결과를 확인합니다.
# tail()
seoul_business.tail()
```

Out[250]:

| | 자치구 | 동 | 다수업종상위 10 | 다수업종상위 10.1 | 다수업종상위 10.2 | 다수업종상위 10.3 | 다수업종상위 10.4 | 다수업종상위 10.5 | 다수업종상위 10.6 | 다수업종상위 10.7 | 다수업종상위 10.8 | 다수업종상위 10.9 | 다수업종상위 10.10 | 다수업종상위 10.11 | 다수업종상위 10.12 | 다수업종상위 10.13 |
|-----|-----|------|--------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|-----------------|-----------------|-----------------|-----------------|
| 446 | 강동구 | 성내3동 | 120 | 120 | 135 | 390 | 73 | 73 | 104 | 163 | 83 | 113 | 54 | 133 | 10 | 86 |
| 447 | 강동구 | 둔촌2동 | 117 | 436 | 63 | 249 | 89 | 89 | 61 | 108 | 62 | 75 | 34 | 85 | 10 | 788 |
| 448 | 강동구 | 암사1동 | 209 | 372 | 71 | 195 | 153 | 153 | 67 | 107 | 82 | 114 | 29 | 59 | 3 | 6 |
| 449 | 강동구 | 천호2동 | 157 | 187 | 152 | 509 | 144 | 151 | 97 | 188 | 83 | 168 | 63 | 212 | 4 | 17 |
| 450 | 강동구 | 길동 | 1,441 | 1,557 | 152 | 436 | 217 | 225 | 146 | 237 | 121 | 193 | 71 | 228 | 6 | 126 |

[실습문제3] 데이터 확인 및 처리

- head 와 tail 을 보았을때, 어느 데이터만 가져와야 할지 생각 해 봅시다.
- 두번째 행부터 데이터가 시작된다.

In [251...

```
# 아래에 실습코드를 작성하고 결과를 확인합니다.
# header=1
seoul_business = pd.read_csv('1.4 business_type_202401.csv', sep=",", encoding = "UTF-8", head
seoul_business
```

Out[251]:

| | 자치구 | 동 | 택시 운송 업 | 택시 운송 업.1 | 한식 일반 음식 점업 | 한식 일 반 음식 점업.1 | 용달 화물 자동차 운송업 | 용달 화물 자동차 운송업.1 | 부동 산 중 개 및 대리 업 | 부동 산 중 개 및 대리 업.1 | 두발 미용 업 | 두발 미용 업.1 | 커피 전문 점 |
|-----|-----|------|---------------|-----------------|----------------------|----------------------------|------------------------|--------------------------|-----------------------------|-------------------------------|---------------|-----------------|---------------|
| 0 | 자치구 | 동 | 사업 체수 | 종사 자수 | 사업 체수 | 종사 자수 | 사업 체수 | 종사 자수 | 사업 체수 | 종사 자수 | 사업 체수 | 종사 자수 | 사업 체수 |
| 1 | 합계 | 합계 | 47,176 | 76,701 | 29,225 | 100,183 | 28,494 | 29,356 | 24,013 | 42,636 | 18,080 | 36,214 | 16,255 |
| 2 | 종로구 | 소계 | 215 | 215 | 1,548 | 5,912 | 618 | 646 | 549 | 889 | 261 | 480 | 948 |
| 3 | 종로구 | 사직동 | 7 | 7 | 230 | 982 | 12 | 12 | 57 | 96 | 28 | 66 | 124 |
| 4 | 종로구 | 삼청동 | 5 | 5 | 42 | 145 | 4 | 4 | 9 | 10 | - | - | 66 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 445 | 강동구 | 성내3동 | 120 | 120 | 135 | 390 | 73 | 73 | 104 | 163 | 83 | 113 | 54 |
| 446 | 강동구 | 둔촌2동 | 117 | 436 | 63 | 249 | 89 | 89 | 61 | 108 | 62 | 75 | 34 |
| 447 | 강동구 | 암사1동 | 209 | 372 | 71 | 195 | 153 | 153 | 67 | 107 | 82 | 114 | 29 |
| 448 | 강동구 | 천호2동 | 157 | 187 | 152 | 509 | 144 | 151 | 97 | 188 | 83 | 168 | 63 |
| 449 | 강동구 | 길동 | 1,441 | 1,557 | 152 | 436 | 217 | 225 | 146 | 237 | 121 | 193 | 71 |

450 rows × 22 columns

```
In [252... # 아래에 실습코드를 작성하고 결과를 확인합니다.
# DataFrame 출력 시 모든 열을 표시하도록 설정
pd.set_option('display.max_columns', None)
```

```
In [253... # 아래에 실습코드를 작성하고 결과를 확인합니다.
# head()
```



```
seoul_business.head()
```

Out[253]:

| | 자치구 | 동 | 택시 운송 업 | 택시 운송 업.1 | 한식 일반 음식 점업 | 한식 일 반 음식 점업.1 | 용달 화물 자 동 차 운 송 업 | 용달 화물 자 동 차 운 송 업.1 | 부동 산 중 개 및 대리 업 | 부동 산 중 개 및 대리 업.1 | 두발 미용 업 | 두발 미용 업.1 | 커피 전문 점 | 커피 전 점 |
|---|-------------|-------------|---------------|-----------------|----------------------|----------------------------|--|--|-----------------------------------|-------------------------------------|---------------|-----------------|---------------|--------------|
| 0 | 자치구 | 동 | 사업 체수 | 종사 자수 | 사업 체수 | 종사자 수 | 사업 체수 | 종사 자수 | 사업 체수 | 종사 자수 | 사업 체수 | 종사 자수 | 사업 체수 | 종 자 |
| 1 | 합 계 | 합 계 | 47,176 | 76,701 | 29,225 | 100,183 | 28,494 | 29,356 | 24,013 | 42,636 | 18,080 | 36,214 | 16,255 | 57,3 |
| 2 | 종 로 구 | 소 계 | 215 | 215 | 1,548 | 5,912 | 618 | 646 | 549 | 889 | 261 | 480 | 948 | 3,6 |
| 3 | 종 로 구 | 사 직 동 | 7 | 7 | 230 | 982 | 12 | 12 | 57 | 96 | 28 | 66 | 124 | 5 |
| 4 | 종 로 구 | 삼 청 동 | 5 | 5 | 42 | 145 | 4 | 4 | 9 | 10 | - | - | 66 | 2 |

In [254...

```
# 아래에 실행코드를 작성하고 결과를 확인합니다.
# 구별 등록 업종 중 버스정류장 설치에 중요한 업종 데이터만 불러오기
# 예 : 대중교통 이동수요가 높을 것으로 예상되는 업종, 종사자수가 많은 업종
seoul_business = seoul_business[['자치구', '동', '한식 일반 음식점업', '한식 일반 음식점업.1',
                                   '남녀용 겔옷 및 셔츠 도매업', '남녀용 겔옷 및 셔츠 도매업.1'],
```

In [255...

```
seoul_business.head()
```

Out[255]:

| | 자치구 | 동 | 한식 일반 음식 점업 | 한식 일반 음식 점업.1 | 커피 전문 점 | 커피 전문 점.1 | 일반 교과 학원 | 일반 교과 학원.1 | 한식 육류 요리 전문 점 | 한식 육류 요리 전문 점.1 | 남녀 용 겉 옷 및 셔츠 도매 업 | 남녀 용 겉 옷 및 셔츠 도매 업.1 | 부 동산 개 및 대 리 업 | 부 동산 개 대 업 |
|---|-----|-----|----------------------|------------------------|---------------|-----------------|----------------|------------------|---------------------------|-----------------------------|---|---|----------------------------------|------------------------|
| 0 | 자치구 | 동 | 사업 체수 | 종사자 수 | 사업 체수 | 종사 자수 | 사업 체수 | 종사 자수 | 사업 체수 | 종사 자수 | 사업 체수 | 종사 자수 | 사업 체수 | 종 자 |
| 1 | 합계 | 합계 | 29,225 | 100,183 | 16,255 | 57,322 | 11,911 | 50,773 | 11,109 | 44,653 | 13,388 | 51,116 | 24,013 | 42,6 |
| 2 | 종로구 | 소계 | 1,548 | 5,912 | 948 | 3,613 | 110 | 382 | 437 | 1,982 | 400 | 962 | 549 | 8 |
| 3 | 종로구 | 사직동 | 230 | 982 | 124 | 525 | 9 | 34 | 45 | 220 | 2 | 18 | 57 | |
| 4 | 종로구 | 삼청동 | 42 | 145 | 66 | 243 | - | - | 7 | 33 | 2 | 3 | 9 | |

In [256...

```
# 아래에 실습코드를 작성하고 결과를 확인합니다.
# '동' 열이 '소계'인 행만을 필터링 (구별 데이터만 필요함)
```

```
seoul_business = seoul_business.loc[seoul_business['동'] == '소계']
```

In [257...

```
seoul_business
```

Out[257]:

| | 자치구 | 동 | 한식 일반음식점업 | 한식 일반음식점업.1 | 커피 전문점 | 커피 전문점.1 | 일반 교과학원 | 일반 교과학원.1 | 한식 육류요리전문점 | 한식 육류요리전문점.1 | 남녀용 겉옷및셔츠도매업 | 남녀용 겉옷및셔츠도매업.1 | 부동 산개및대리업 | 부동 산개및대리업.1 |
|-----|------|----|--------------|----------------|-----------|-------------|------------|--------------|---------------|-----------------|-----------------|-------------------|--------------|----------------|
| 2 | 종로구 | 소계 | 1,548 | 5,912 | 948 | 3,613 | 110 | 382 | 437 | 1,982 | 400 | 962 | 549 | 889 |
| 20 | 중구 | 소계 | 1,588 | 6,414 | 869 | 3,542 | 57 | 175 | 443 | 2,121 | 10,810 | 24,248 | 576 | 1,152 |
| 36 | 용산구 | 소계 | 857 | 2,987 | 634 | 2,076 | 104 | 573 | 300 | 1,184 | 54 | 317 | 791 | 1,214 |
| 53 | 성동구 | 소계 | 956 | 2,785 | 547 | 1,859 | 258 | 1,494 | 400 | 1,545 | 149 | 1,572 | 801 | 1,453 |
| 71 | 광진구 | 소계 | 963 | 2,861 | 590 | 1,887 | 413 | 1,646 | 467 | 1,678 | 89 | 687 | 871 | 1,313 |
| 87 | 동대문구 | 소계 | 1,236 | 3,566 | 494 | 1,472 | 302 | 984 | 442 | 1,509 | 211 | 976 | 842 | 1,208 |
| 102 | 종랑구 | 소계 | 991 | 2,479 | 381 | 1,037 | 288 | 1,045 | 499 | 1,504 | 63 | 284 | 738 | 1,104 |
| 119 | 성북구 | 소계 | 973 | 2,822 | 457 | 1,384 | 464 | 1,755 | 365 | 1,334 | 85 | 454 | 846 | 1,175 |
| 140 | 강북구 | 소계 | 965 | 2,663 | 358 | 1,095 | 220 | 740 | 362 | 1,176 | 40 | 94 | 589 | 867 |
| 154 | 도봉구 | 소계 | 672 | 2,009 | 288 | 837 | 351 | 1,162 | 302 | 943 | 22 | 43 | 515 | 794 |
| 169 | 노원구 | 소계 | 866 | 2,650 | 535 | 1,748 | 735 | 3,281 | 428 | 1,557 | 30 | 123 | 714 | 1,053 |
| 189 | 은평구 | 소계 | 881 | 2,585 | 457 | 1,296 | 551 | 1,793 | 476 | 1,639 | 26 | 74 | 1,134 | 1,614 |
| 206 | 서대문구 | 소계 | 863 | 2,694 | 370 | 1,323 | 361 | 1,268 | 311 | 1,088 | 22 | 68 | 667 | 956 |

| | 자치구 | 동 | 한식 일반음식점업 | 한식 일반음식점업.1 | 커피 전문점 | 커피 전문점.1 | 일반 교과학원 | 일반 교과학원.1 | 한식 육류요리전문점 | 한식 육류요리전문점.1 | 남녀용 겉옷및셔츠도매업 | 남녀용 겉옷및셔츠도매업.1 | 부동 산개및대리업 | 부동 산개및대리업.1 |
|-----|------|----|--------------|----------------|-----------|-------------|------------|--------------|---------------|-----------------|-----------------|-------------------|--------------|----------------|
| 221 | 마포구 | 소계 | 1,411 | 4,705 | 1,369 | 4,475 | 469 | 1,855 | 701 | 2,928 | 96 | 1,209 | 1,149 | 1,903 |
| 238 | 양천구 | 소계 | 831 | 2,800 | 402 | 1,375 | 1,284 | 5,804 | 281 | 1,111 | 25 | 52 | 922 | 1,364 |
| 257 | 강서구 | 소계 | 1,325 | 4,584 | 690 | 2,530 | 672 | 2,347 | 581 | 2,274 | 78 | 700 | 1,311 | 2,352 |
| 278 | 구로구 | 소계 | 1,048 | 3,215 | 421 | 1,384 | 336 | 1,111 | 428 | 1,496 | 114 | 923 | 789 | 1,213 |
| 294 | 금천구 | 소계 | 867 | 2,619 | 383 | 1,215 | 153 | 487 | 300 | 1,042 | 200 | 2,575 | 671 | 1,381 |
| 305 | 영등포구 | 소계 | 1,738 | 6,476 | 838 | 3,295 | 288 | 1,082 | 494 | 2,151 | 76 | 1,220 | 986 | 1,566 |
| 324 | 동작구 | 소계 | 765 | 2,322 | 486 | 1,628 | 407 | 1,463 | 331 | 1,177 | 20 | 199 | 883 | 1,483 |
| 340 | 관악구 | 소계 | 1,226 | 3,689 | 581 | 2,049 | 354 | 1,230 | 422 | 1,628 | 31 | 190 | 1,010 | 2,136 |
| 362 | 서초구 | 소계 | 1,564 | 6,712 | 1,012 | 3,897 | 812 | 3,812 | 445 | 2,338 | 157 | 1,563 | 1,457 | 2,758 |
| 381 | 강남구 | 소계 | 2,415 | 11,486 | 1,624 | 7,273 | 1,426 | 9,020 | 827 | 5,008 | 384 | 8,320 | 2,337 | 6,508 |
| 404 | 송파구 | 소계 | 1,575 | 5,760 | 975 | 3,372 | 876 | 4,160 | 639 | 2,781 | 147 | 2,830 | 1,677 | 3,190 |
| 432 | 강동구 | 소계 | 1,101 | 3,388 | 546 | 1,660 | 620 | 2,104 | 428 | 1,459 | 59 | 1,433 | 1,188 | 1,990 |

In [258...

```
# 아래에 실습코드를 작성하고 결과를 확인합니다.
# '동' 열 제거
seoul_business = seoul_business.drop('동', axis=1)
```

```
In [259... # 아래에 실습코드를 작성하고 결과를 확인합니다.
# head()
seoul_business.head()
```

Out[259]:

| | 자치구 | 한식 일반 음식 점업 | 한식 일반 음식 점업.1 | 커피 전문 점 | 커피 전문 점.1 | 일반 교과 학원 | 일반 교과 학원.1 | 한식 육류 요리 전문 점 | 한식 육류 요리 전문 점.1 | 남녀용 겉옷 셔츠 도 매업 | 남녀용 겉옷 셔츠 도 매업.1 | 부동 산 개 대리 업 | 부동 산 개 대리 업.1 |
|----|-----|----------------------|------------------------|---------------|-----------------|----------------|------------------|---------------------------|-----------------------------|----------------------------|------------------------------|-------------------------|---------------------------|
| 2 | 종로구 | 1,548 | 5,912 | 948 | 3,613 | 110 | 382 | 437 | 1,982 | 400 | 962 | 549 | 889 |
| 20 | 중구 | 1,588 | 6,414 | 869 | 3,542 | 57 | 175 | 443 | 2,121 | 10,810 | 24,248 | 576 | 1,152 |
| 36 | 용산구 | 857 | 2,987 | 634 | 2,076 | 104 | 573 | 300 | 1,184 | 54 | 317 | 791 | 1,214 |
| 53 | 성동구 | 956 | 2,785 | 547 | 1,859 | 258 | 1,494 | 400 | 1,545 | 149 | 1,572 | 801 | 1,453 |
| 71 | 광진구 | 963 | 2,861 | 590 | 1,887 | 413 | 1,646 | 467 | 1,678 | 89 | 687 | 871 | 1,313 |

```
In [260... # 해당 데이터프레임을 csv 파일로 저장하세요.
seoul_business.to_csv('seoul_business.csv', index=False)
```

3.데이터 분석하기

- KeyPoint : 데이터의 형태를 살펴보고 다양한 분석기법을 통해 모델링에 적합하도록 정제요소를 선별할 수 있다.
 - 데이터들의 패턴 탐색
 - 변수들간의 관계 파악

```
In [261... # 시각화 한글폰트 설정
import seaborn as sns
import matplotlib.pyplot as plt

plt.rc('font', family='Malgun Gothic')
sns.set(font="Malgun Gothic", # "NanumGothicCoding",
        rc={"axes.unicode_minus":False}, # 마이너스 부호 깨짐 현상 해결
        style='darkgrid')
```

[실습문제4] 데이터 분포 알아보기

- 여러 변수들을 다양한 그래프로 그려보고 인사이트를 도출해보세요.

In [262... seoul_business.head()

Out[262]:

| | 자치구 | 한식 일반 음식 점업 | 한식 일반 음식 점업.1 | 커피 전문 점 | 커피 전문 점.1 | 일반 교과 학원 | 일반 교과 학원.1 | 한식 육류 요리 전문 점 | 한식 육류 요리 전문 점.1 | 남녀 용 겉옷 및 셔츠 도 매업 | 남녀 용 겉옷 및 셔츠 도 매업.1 | 부 동 산 개 대 리 업 | 부 동 산 개 대 리 업.1 |
|----|-----|----------------------|------------------------|---------------|-----------------|----------------|------------------|---------------------------|-----------------------------|-------------------------------------|---------------------------------------|---------------------------------|-----------------------------------|
| 2 | 종로구 | 1,548 | 5,912 | 948 | 3,613 | 110 | 382 | 437 | 1,982 | 400 | 962 | 549 | 889 |
| 20 | 중구 | 1,588 | 6,414 | 869 | 3,542 | 57 | 175 | 443 | 2,121 | 10,810 | 24,248 | 576 | 1,152 |
| 36 | 용산구 | 857 | 2,987 | 634 | 2,076 | 104 | 573 | 300 | 1,184 | 54 | 317 | 791 | 1,214 |
| 53 | 성동구 | 956 | 2,785 | 547 | 1,859 | 258 | 1,494 | 400 | 1,545 | 149 | 1,572 | 801 | 1,453 |
| 71 | 광진구 | 963 | 2,861 | 590 | 1,887 | 413 | 1,646 | 467 | 1,678 | 89 | 687 | 871 | 1,313 |

```
In [263... # 데이터 스케일링
import numpy as np
from sklearn.preprocessing import minmax_scale
```

```
In [264... for column in seoul_business.columns[1:]:
    # 쉼표 제거 후 숫자로 변환
    seoul_business[column] = seoul_business[column].str.replace(',', '').astype(float)

# 종사자 수 구분 # 데이터 스케일링
for column in seoul_business.columns[1::]:
    mean = np.mean(seoul_business[column])
    std = np.std(seoul_business[column])
    seoul_business[column] = (seoul_business[column] - mean) / std
```

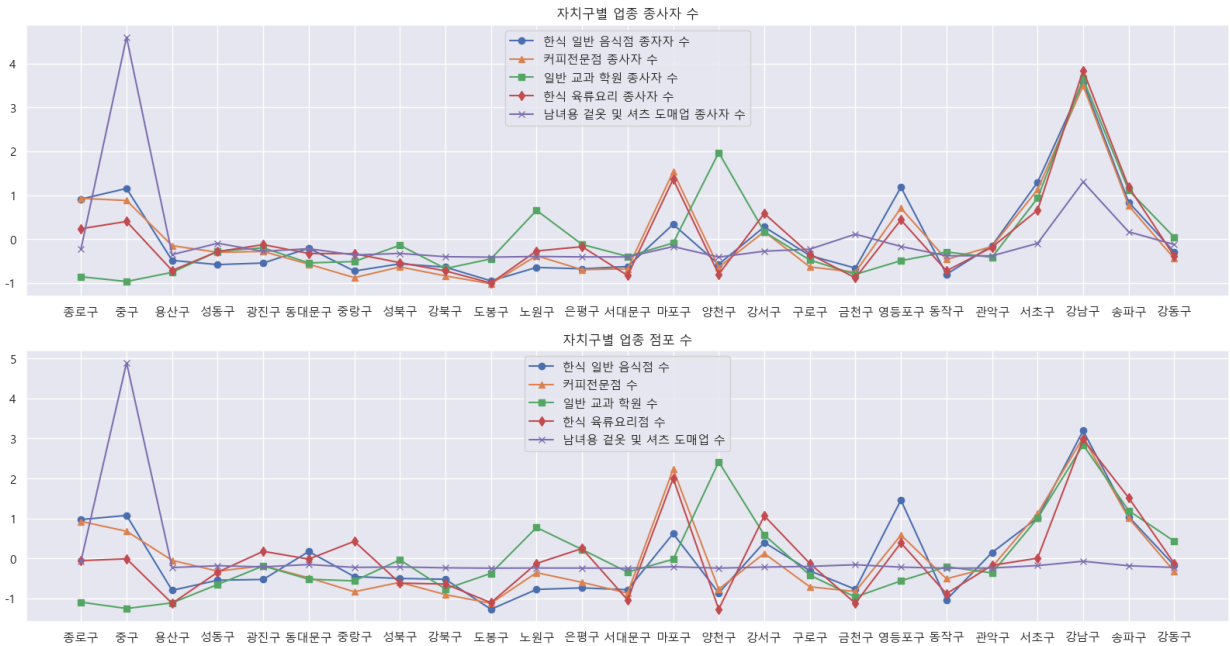
```
In [267... # 아래에 실습코드를 작성하고 결과를 확인합니다.
plt.figure(figsize=(20, 10))

plt.subplot(2, 1, 1)
plt.title('자치구별 업종 종사자 수')
plt.plot(seoul_business['자치구'], seoul_business['한식 일반 음식점업.1'], marker='o', label='한식 일반 음식점업.1')
plt.plot(seoul_business['자치구'], seoul_business['커피전문점.1'], marker='^', label='커피전문점.1')
plt.plot(seoul_business['자치구'], seoul_business['일반 교과 학원.1'], marker='s', label='일반 교과 학원.1')
plt.plot(seoul_business['자치구'], seoul_business['한식 육류요리 전문점.1'], marker='d', label='한식 육류요리 전문점.1')
plt.plot(seoul_business['자치구'], seoul_business['남녀용 겉옷 및 셔츠 도매업.1'], marker='x', label='남녀용 겉옷 및 셔츠 도매업.1')
plt.grid(True)
plt.legend()

plt.subplot(2, 1, 2)
plt.title('자치구별 업종 점포 수')
plt.plot(seoul_business['자치구'], seoul_business['한식 일반 음식점업'], marker='o', label='한식 일반 음식점업')
plt.plot(seoul_business['자치구'], seoul_business['커피전문점'], marker='^', label='커피전문점')
plt.plot(seoul_business['자치구'], seoul_business['일반 교과 학원'], marker='s', label='일반 교과 학원')
```

```
plt.plot(seoul_business['자치구'], seoul_business['한식 육류요리 전문점'], marker='d', label='한식 육류요리 전문점')
plt.plot(seoul_business['자치구'], seoul_business['남녀용 겔옷 및 셔츠 소매업'], marker='x', label='남녀용 겔옷 및 셔츠 소매업')
plt.grid(True)
plt.legend()

plt.show()
```



In [199...

위 차트를 통해 알게된 사실을 정리해봅시다.

- # 1. 강남구가 각 점포 및 종사자 수가 가장 많음 (유동 인구가 많아 경제 좋음) (직장 및 직장인도
- # 2. 양천구가 교과 학원 수 및 종사자가 강남구를 제외 하고 가장 높음 (아마 교육열이 높은 목동이
- # 3. 마포구가 강남구 다음으로 커피매점 수가 가장 많음 (아마 상암도 같은 방송국이 있어서 그런거
- # 4. 영등포구와 중구가 강남구 다음으로 한식 음식점 수 및 종사자가 많음 (아마 중구는 한국의 역사
- # (영등포구는 예비 공무원들이 공부하는 노량진이 있어서
- # 5. 금천구 대부분 매장 수 및 종사자가 적음? (가산디지털단지가 있어 직장인들이 많을 텐데 왜?)