토익 진단평가 데이터 다듬기

단계1 : 데이터 전처리

- 이번 시간에는 파이썬을 활용한 AI모델링에서 전처리에 대해 실습해보겠습니다.
- 전처리란 데이터 분석을 수행하기 전에 데이터를 적절한 형태로 가공하는 과정을 말합니다.
- 머신러닝과 AI모델링에서 60~70% 차지하는 부분이 바로 전처리 파트입니다.
- 굉장히 시간과 노력이 많이 투입되며, 어려운 부분일 수 있습니다.
- 데이터의 이상치나 결측치를 제거하거나, 변수를 수치형으로 변환하거나, 표준화나 정규화를 적용하는 등의 작업이 있습니다.
- 데이터가 깨끗이 정리되지 않으면 머신러닝/AI 성능을 장담할 수 없습니다.
- 데이터 전처리에 심혈을 기울여주시기 바라며, 이론보다 실습이 더 많은 시간과 노력이 투자되어야 합니다.

0.미션 확인

- 개요
  - 응시할 토익점수를 예측해주는 모델을 생성하기 위한 데이터를 전처리하려고 합니다.
  - 이를 위해 주어진 데이터는 응시자별 3회차씩의 토익점수와 당시 학습방식 등의 정보입니다.
  - 관련 토익점수 및 학습정보 데이터를 분석하여 미니프로젝트 1차 이후에 예측 모델을 개발하여 토익 응시생들에게 도움이 되는 솔루션을 개발하려고 합니다.
- 전처리 단계 목표
  - 3회차(3행)의 데이터를 하나의 분석단위(1행)로 만들기 위해 응시자별 데이터를 집계합니다.
  - 응시자는 하나의 ID로 구분됩니다.
  - 응시회차는 seq로 구분됩니다. 모든 응시자가 1,2,3 의 회차 값을 갖고 있습니다.
- 학습목차 0) 미션 내용 확인 1) 환경설정 : 라이브러리 불러오기 및 파일 읽어오기 2) 데이터 프레임 탐색 3) 데이터 전처리 수행 (불필요 컬럼 삭제, Null 처리, 중복값 제거, 한 행으로 합치기 등) 4) 결과 저장하기

1.환경설정

(1) 라이브러리 불러오기

- **세부 요구사항**
  - 기본적으로 필요한 라이브러리를 import 하도록 코드를 작성해주세요.
  - 필요하다고 판단되는 라이브러리를 추가하겠습니다.

In [766…
```
#[문제] pandas, numpy 라이브러리를 임포트하세요.
```

In [767…
```
import pandas as pd
import numpy as np
```

(2) 데이터 불러오기

- 읽어올 데이터 파일
  - 학습 데이터 : data04.xls
  - 엑셀 파일이므로 pd.read_excel 함수를 이용해서 불러 옵니다.
- 다음과 같이 데이터를 불러와주세요.
  - 주피터랩 실행
    - 제공된 압축파일 '미프 1차_토익'을 다운받아 압축을 해제합니다.
    - anaconda의 root directory(C:\Users\) 에 '미프 1차_토익' 폴더를 만들고, 복사합니다.
    - '1.전처리_교육생용'과 '2.데이터_탐색_교육생용' 실습 파일을 열어주세요.

1) 주피터랩 실행

- '미프 1차_토익' 폴더에 필요한 파일들을 넣고, 본 파일을 열거나, 별도 경로를 지정해서 데이터를 읽어올 수 있습니다.

In [768…
```
#[문제] '미프 1차_토익' 폴더에서 본 파일인 '1.전처리_교육생용'을 열어주세요.
```

2) 데이터 읽어오기

In [769…
```
#[문제] data04.xlsx 파일을 Pandas read_excel 함수를 이용하여 읽고 data변수에 저장하세요.
```

In [770…
```
# 읽어들일 파일명 : data04.xlsx
# Pandas read_excel 함수 활용
# 결과 : data 저장
data = pd.read_excel('data04.xlsx')
file = 'data04.xlsx'
```

In [771…
```
#[문제] 읽어온 데이터프레임을 확인하고, 상위 10개 행만 보여주세요.
```

In [772…
```
data.head(10)
```

Out[772]:

	ID	Seq	Gender	Birth_Year	LC_Score	RC_Score	Total Score	학습목표	학습방법	강의학습교재유형	학습빈도	기출문제공부횟수	취약분야인지여부	토익모의테스트횟수	Student ID
0	1	1	M	1973	181	173	354	자기계발	참고서	일반적인영어텍스트기반교재	주3-4회	6.0	알고있지않음	6	student1
1	1	2	M	1973	227	213	440	자기계발	오프라인강의	뉴스/이슈기반교재	주1-2회	3.0	알고있음	5	student1
2	1	3	M	1973	345	336	681	승진	온라인강의	영상교재	주5-6회	7.0	알고있음	10	student1
3	2	1	F	1982	330	290	620	자기계발	오프라인강의	뉴스/이슈기반교재	매일(주7회)	8.0	알고있지않음	19	student2
4	2	2	F	1982	354	339	693	승진	온라인강의	영상교재	주5-6회	2.0	알고있음	15	student2
5	2	3	F	1982	380	368	748	승진	온라인강의	뉴스/이슈기반교재	주5-6회	4.0	알고있음	14	student2

	ID	Seq	Gender	Birth_Year	LC_Score	RC_Score	Total Score	학습목표	학습방법	강의학습교재유형	학습빈도	기출문제공부횟수	취약분야인지여부	토익모의테스트횟수	Student ID
6	3	1	F	1995	367	309	676	취업	온라인강의	영상교재	매일(주 7회)	9.0	알고있지않음	7	student3
7	3	2	F	1995	396	365	761	자기계발	온라인강의	영상교재	주 3-4회	7.0	알고있지않음	6	student3
8	3	3	F	1995	416	382	798	자기계발	참고서	일반적인영어텍스트기반교재	주 1-2회	4.0	알고있음	4	student3
9	4	1	M	1987	470	285	755	자기계발	온라인강의	뉴스/이슈기반교재	주 1-2회	7.0	알고있지않음	4	student4

2. 데이터프레임 탐색

- **세부 요구사항**
  - data의 형태, 기초통계량, 정보 등을 확인합니다.
  - 특히 .info() 를 통해서 각 변수별 데이터타입이 적절한지 확인합니다.
  - 다양한 함수로 데이터를 탐색해주세요.

In [773…

```
#[문제] 전체 데이터의 행,열 개수를 확인하기
```

In [774…
```
data.shape
```

Out[774]:
```
(1500, 15)
```

In [775…
```
#[문제] 전체 데이터의 하위 5개 행 확인하기
```

In [776…
```
data.tail()
```

Out[776]:

	ID	Seq	Gender	Birth_Year	LC_Score	RC_Score	Total Score	학습목표	학습방법	강의학습교재유형	학습빈도	기출문제공부횟수	취약분야인지여부	토익모의테스트횟수	S
1495	499	2	F	1990	378	326	704	승진	온라인강의	뉴스/이슈기반교재	주5-6회	6.0	알고있지않음	12	stud
1496	499	3	F	1990	422	370	792	자기계발	오프라인강의	비즈니스시뮬레이션(Role Play)	주3-4회	4.0	알고있음	7	stud
1497	500	1	M	1984	169	188	357	자기계발	참고서	일반적인영어텍스트기반교재	주3-4회	8.0	알고있지않음	2	stud
1498	500	2	M	1984	172	190	362	자기계발	참고서	뉴스/이슈기반교재	매일(주7회)	10.0	알고있음	16	stud
1499	500	3	M	1984	235	226	461	승진	오프라인강의	비즈니스시뮬레이션(Role Play)	주5-6회	7.0	알고있음	15	stud

In [777…
```
#[문제] 전체 데이터의 모든 변수 확인하기
```

In [778…
```
data.columns
```

Out[778]:
```
Index(['ID', 'Seq', 'Gender', 'Birth_Year', 'LC_Score', 'RC_Score',
       'Total Score', '학습목표', '학습방법', '강의 학습 교재 유형', '학습빈도', '기출문제 공부 횟수',
       '취약분야 인지 여부', '토익 모의테스트 횟수', 'Student ID'],
      dtype='object')
```

In [779…
```
#[문제] 전체 데이터 정보 확인
```

In [780…
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1500 entries, 0 to 1499
Data columns (total 15 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   ID           1500 non-null   int64  
 1   Seq          1500 non-null   int64  
 2   Gender       1500 non-null   object 
 3   Birth_Year   1500 non-null   int64  
 4   LC_Score     1500 non-null   int64  
 5   RC_Score     1500 non-null   int64  
 6   Total Score  1500 non-null   int64  
 7   학습목표         1500 non-null   object 
 8   학습방법         1500 non-null   object 
 9   강의 학습 교재 유형  1500 non-null   object 
 10  학습빈도         1500 non-null   object 
 11  기출문제 공부 횟수   1497 non-null   float64
 12  취약분야 인지 여부   1500 non-null   object 
 13  토익 모의테스트 횟수  1500 non-null   int64  
 14  Student ID   1500 non-null   object 
dtypes: float64(1), int64(7), object(7)
memory usage: 175.9+ KB
```

In [781…
```
#[문제] 각 열별 Null 데이터 값의 개수를 확인해주세요.
```

In [782…
```
data.isna().sum()
```

Out[782]:
```
ID              0
Seq             0
Gender          0
Birth_Year      0
LC_Score        0
RC_Score        0
Total Score     0
학습목표            0
학습방법            0
강의 학습 교재 유형     0
학습빈도            0
기출문제 공부 횟수      3
취약분야 인지 여부      0
토익 모의테스트 횟수     0
Student ID      0
dtype: int64
```

In [783…
```
#[문제] 데이터의 통계정보를 확인해주세요.
```

In [784…
```
data.describe().T
```

Out[784]:

	count	mean	std	min	25%	50%	75%	max
ID	1500.0	250.500000	144.385415	1.0	125.75	250.5	375.25	500.0
Seq	1500.0	2.000000	0.816769	1.0	1.00	2.0	3.00	3.0
Birth_Year	1500.0	1992.906000	8.218893	1973.0	1986.75	1992.5	2000.00	2007.0
LC_Score	1500.0	340.079333	86.807523	105.0	279.00	335.0	404.00	495.0
RC_Score	1500.0	340.164667	87.143890	84.0	280.00	337.0	406.00	495.0
Total Score	1500.0	680.260667	159.110652	250.0	564.00	687.0	800.00	990.0
기출문제 공부 횟수	1497.0	5.286573	2.797303	1.0	3.00	5.0	8.00	10.0
토익 모의테스트 횟수	1500.0	9.784000	5.324181	1.0	5.00	9.0	14.00	20.0

3.데이터 전처리 수행

- **세부 요구사항**
  - 여기부터는 스스로 실습해봅시다.
  - ID를 기준으로 3회차(3행) 데이터를 1개의 분석단위(1행)으로 만들어야 합니다.
    - (1) 개인정보 데이터와 각 회차별 학습정보 데이터 분리
    - (2) 개인정보의 중복값 제거
    - (3) 각 회차별 정보를 한 행으로 만들기
    - (4) (2)와 (3)을 합치기 (merge, ID 기준)
    - (5) 레이블 만들기
  - 전처리 결과를 다시 한번 정리해봅니다.

(1) 컬럼삭제 및 값변경, 개인정보와 각 회차별 학습정보 분리

In [785…
```
#[문제] data의 컬럼별 데이터 타입을 출력하세요.
```

In [786…
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1500 entries, 0 to 1499
Data columns (total 15 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   ID           1500 non-null   int64
 1   Seq          1500 non-null   int64
 2   Gender       1500 non-null   object
 3   Birth_Year   1500 non-null   int64
 4   LC_Score     1500 non-null   int64
 5   RC_Score     1500 non-null   int64
 6   Total Score  1500 non-null   int64
 7   학습목표         1500 non-null   object
 8   학습방법         1500 non-null   object
 9   강의 학습 교재 유형  1500 non-null   object
 10  학습빈도         1500 non-null   object
 11  기출문제 공부 횟수   1497 non-null   float64
 12  취약분야 인지 여부   1500 non-null   object
 13  토익 모의테스트 횟수  1500 non-null   int64
 14  Student ID   1500 non-null   object
dtypes: float64(1), int64(7), object(7)
memory usage: 175.9+ KB
```

In [787…

```
#[문제] data에서 'Student ID' 컬럼을 삭제하세요.
```

In [788…

```
data.head()
```

Out[788]:

	ID	Seq	Gender	Birth_Year	LC_Score	RC_Score	Total Score	학습목표	학습방법	강의학습교재유형	학습빈도	기출문제공부횟수	취약분야인지여부	토익모의테스트횟수	Student ID
0	1	1	M	1973	181	173	354	자기계발	참고서	일반적인영어텍스트기반교재	주3-4회	6.0	알고있지않음	6	student1
1	1	2	M	1973	227	213	440	자기계발	오프라인강의	뉴스/이슈기반교재	주1-2회	3.0	알고있음	5	student1
2	1	3	M	1973	345	336	681	승진	온라인강의	영상교재	주5-6회	7.0	알고있음	10	student1
3	2	1	F	1982	330	290	620	자기계발	오프라인강의	뉴스/이슈기반교재	매일(주7회)	8.0	알고있지않음	19	student2
4	2	2	F	1982	354	339	693	승진	온라인강의	영상교재	주5-6회	2.0	알고있음	15	student2

In [789…

```
# axis=1 옵션, 컬럼단위 삭제(drop함수)
# inplace=True 옵션, data 데이터프레임에 저장
data.drop('Student ID', axis=1, inplace=True)
data.head()
```

Out[789]:

	ID	Seq	Gender	Birth_Year	LC_Score	RC_Score	Total Score	학습목표	학습방법	강의학습교재유형	학습빈도	기출문제공부횟수	취약분야인지여부	토익모의테스트횟수
0	1	1	M	1973	181	173	354	자기계발	참고서	일반적인영어텍스트반 교재	주 3-4 회	6.0	알고있지않음	6
1	1	2	M	1973	227	213	440	자기계발	오프라인강의	뉴스/이슈기반 교재	주 1-2 회	3.0	알고있음	5
2	1	3	M	1973	345	336	681	승진	온라인강의	영상교재	주 5-6 회	7.0	알고있음	10
3	2	1	F	1982	330	290	620	자기계발	오프라인강의	뉴스/이슈기반 교재	매일 (주 7 회)	8.0	알고있지않음	19
4	2	2	F	1982	354	339	693	승진	온라인강의	영상교재	주 5-6 회	2.0	알고있음	15

In [790…
```
#[문제] 14개컬럼에서 13개컬럼으로 줄어든 것을 확인해주세요.
```

In [791…
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1500 entries, 0 to 1499
Data columns (total 14 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   ID           1500 non-null   int64
 1   Seq          1500 non-null   int64
 2   Gender       1500 non-null   object
 3   Birth_Year   1500 non-null   int64
 4   LC_Score     1500 non-null   int64
 5   RC_Score     1500 non-null   int64
 6   Total Score  1500 non-null   int64
 7   학습목표         1500 non-null   object
 8   학습방법         1500 non-null   object
 9   강의 학습 교재 유형  1500 non-null   object
 10  학습빈도         1500 non-null   object
 11  기출문제 공부 횟수   1497 non-null   float64
 12  취약분야 인지 여부   1500 non-null   object
 13  토익 모의테스트 횟수  1500 non-null   int64
dtypes: float64(1), int64(7), object(6)
memory usage: 164.2+ KB
```

In [792…
```
#[문제] data 데이터프레임에서 '기출문제 공부 횟수' 컬럼의 Null 값을 '' --> '0'으로 변경하세요.
```

In [793…
```
# fillna 함수
# 대상컬럼 : '기출문제 공부 횟수'
data['기출문제 공부 횟수'].fillna(0, inplace=True)
```

In [794…
```
#[문제] '기출문제 공부 횟수' 컬럼의 값 0으로 변경 확인
```

In [795…
```
data['기출문제 공부 횟수'].value_counts()
```

Out[795]:
```
기출문제 공부 횟수
2.0     178
3.0     176
4.0     162
6.0     160
5.0     151
8.0     139
9.0     138
7.0     137
1.0     134
10.0    122
0.0       3
Name: count, dtype: int64
```

In [796…
```
#[문제] 개인정보 데이터와 토익시험 학습정보 데이터를 2개의 데이터 프레임인 df1, df2으로 각각 분
```

In [797…
```
# 개인정보 데이터는 features1, 토익시험 학습정보 데이터는 features2로 분리해주세요.
# df1(개인정보 데이터)에 포함될 features : 'ID', 'Gender', 'Birth_Year'
# df2(토익시험 학습정보 데이터)에 포함될 features : 'ID','Seq', 'LC_Score', 'RC_Score', 'Total

df1 = data[['ID', 'Gender', 'Birth_Year']]
df2 = data[['ID','Seq', 'LC_Score', 'RC_Score', 'Total Score', '학습목표', '학습방법', '강의 학
```

In [798…
```
df1
```

Out[798]:

	ID	Gender	Birth_Year
0	1	M	1973
1	1	M	1973
2	1	M	1973
3	2	F	1982
4	2	F	1982
...	...	...	...
1495	499	F	1990
1496	499	F	1990
1497	500	M	1984
1498	500	M	1984
1499	500	M	1984

1500 rows × 3 columns

In [799…
```
df1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1500 entries, 0 to 1499
Data columns (total 3 columns):
 #   Column      Non-Null Count  Dtype
---  ------      --------------  -----
 0   ID          1500 non-null   int64
 1   Gender      1500 non-null   object
 2   Birth_Year  1500 non-null   int64
dtypes: int64(2), object(1)
memory usage: 35.3+ KB
```

In [800…
```
df2.head()
```

Out[800]:

	ID	Seq	LC_Score	RC_Score	Total Score	학습 목표	학습 방법	강의 학습 교재 유형	학습 빈도	기출문제 공부 횟수	취약분야 인지 여부	토익 모의테스트 횟수
0	1	1	181	173	354	자기계발	참고서	일반적인 영어 텍스트 기반 교재	주3-4회	6.0	알고있지 않음	6
1	1	2	227	213	440	자기계발	오프라인 강의	뉴스/이슈 기반 교재	주1-2회	3.0	알고있음	5
2	1	3	345	336	681	승진	온라인강의	영상 교재	주5-6회	7.0	알고있음	10
3	2	1	330	290	620	자기계발	오프라인 강의	뉴스/이슈 기반 교재	매일(주 7회)	8.0	알고있지 않음	19
4	2	2	354	339	693	승진	온라인강의	영상 교재	주5-6회	2.0	알고있음	15

(2) 개인정보의 중복값 제거

In [801…

```
#[문제] 개인정보 데이터 'df1'의 중복된 행을 제거해주세요.
```

In [802…

```
# drop_duplicates 함수 활용
# 제거된 결과는 원본 데이터프레임 'df1'에 바로 적용
df1.drop_duplicates(subset='ID', keep='first', inplace=True)
#df1= df1.drop_duplicates()
```

```
C:\Users\User\AppData\Local\Temp\ipykernel_13836\309843842.py:3: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/
indexing.html#returning-a-view-versus-a-copy
  df1.drop_duplicates(subset='ID', keep='first', inplace=True)
```

In [803…

```
df1
```

Out[803]:

	ID	Gender	Birth_Year
0	1	M	1973
3	2	F	1982
6	3	F	1995
9	4	M	1987
12	5	M	1994
...	...	...	...
1485	496	M	2006
1488	497	F	1988
1491	498	M	2006
1494	499	F	1990
1497	500	M	1984

500 rows × 3 columns

In [804…
```
df1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 500 entries, 0 to 1497
Data columns (total 3 columns):
 #   Column      Non-Null Count  Dtype
---  ------      --------------  -----
 0   ID          500 non-null    int64
 1   Gender      500 non-null    object
 2   Birth_Year  500 non-null    int64
dtypes: int64(2), object(1)
memory usage: 15.6+ KB
```

(3) 각 회차별 정보를 한 행으로 만들기

- 우리는 3차시(Seq == 3)를 기준으로 1,2차시 정보를 집계하여 한 행으로 만들어 봅니다.

In [805…
```
#[문제] 토익시험 학습정보 데이터 'df2'에서 각 ID별 차시(Seq)가 3인 행을 선택하여 새로운 데이터
```

In [806…
```
# 3차시 데이터 : ['Seq'] == 3
# loc 함수 활용 : 특정 행, 특정 행과 열, 그리고 조건에 따라 필터링된 행을 선택하는 함수
temp = df2.loc[df2['Seq'] == 3]
```

In [807…
```
temp
```

Out[807]:

	ID	Seq	LC_Score	RC_Score	Total Score	학습 목표	학습 방법	강의 학습 교재 유형	학습 빈도	기출문제 공부 횟수	취약 분야 인지 여부	토익 모의 테스트 횟수
2	1	3	345	336	681	승진	온라인강의	영상 교재	주5-6회	7.0	알고 있음	10
5	2	3	380	368	748	승진	온라인강의	뉴스/이슈 기반 교재	주5-6회	4.0	알고 있음	14
8	3	3	416	382	798	자기계발	참고서	일반적인 영어 텍스트 기반 교재	주1-2회	4.0	알고 있음	4
11	4	3	495	397	892	승진	온라인강의	뉴스/이슈 기반 교재	주3-4회	9.0	알고 있음	8
14	5	3	398	437	835	자기계발	온라인강의	영상 교재	주3-4회	6.0	알고 있음	4
...	...	...	...	...	...	...	...	...	...	...	...	...
1487	496	3	364	336	700	자기계발	온라인강의	일반적인 영어 텍스트 기반 교재	매일(주 7회)	10.0	알고 있음	13
1490	497	3	187	252	439	승진	온라인강의	비즈니스 시뮬레이션(Role Play)	매일(주 7회)	9.0	알고 있음	17
1493	498	3	255	167	422	자기계발	오프라인강의	일반적인 영어 텍스트 기반 교재	주1-2회	0.0	알고 있음	4
1496	499	3	422	370	792	자기계발	오프라인강의	비즈니스 시뮬레이션(Role Play)	주3-4회	4.0	알고 있음	7
1499	500	3	235	226	461	승진	오프라인강의	비즈니스 시뮬레이션(Role Play)	주5-6회	7.0	알고 있음	15

500 rows × 12 columns

In [808…
```
#[문제] temp 데이터프레임의 열이름 중 'LC_Score', 'RC_Score', 'Total Score'를 각각 '3st_LC_Sco
```

In [809…
```
# rename 함수 활용, temp에 저장
# rename 함수 : 데이터프레임의 행 인덱스 또는 열 이름을 변경하는 데 사용
```

```
temp = temp.rename(columns={'LC_Score':'3st_LC_Score',
                            'RC_Score':'3st_RC_Score',
                            'Total Score':'3st_Total_Score'})
```

In [810…

```
temp
```

Out[810]:

	ID	Seq	3st_LC_Score	3st_RC_Score	3st_Total_Score	학습목표	학습방법	강의학습교재유형	학습빈도	기출문제공부횟수	취약분야인지여부	토익모의테스트횟수
2	1	3	345	336	681	승진	온라인강의	영상교재	주 5-6 회	7.0	알고있음	10
5	2	3	380	368	748	승진	온라인강의	뉴스/이슈기반교재	주 5-6 회	4.0	알고있음	14
8	3	3	416	382	798	자기계발	참고서	일반적인 영어 텍스트 반 기교재	주 1-2 회	4.0	알고있음	4
11	4	3	495	397	892	승진	온라인강의	뉴스/이슈기반교재	주 3-4 회	9.0	알고있음	8
14	5	3	398	437	835	자기계발	온라인강의	영상교재	주 3-4 회	6.0	알고있음	4
...	...	...	...	...	...	...	...	...	...	...	...	...
1487	496	3	364	336	700	자기계발	온라인강의	일반적인 영어 텍스트 반 기교재	매일 (주 7 회)	10.0	알고있음	13
1490	497	3	187	252	439	승진	온라인강의	비즈니스 시뮬레이션 (Role Play)	매일 (주 7 회)	9.0	알고있음	17

	ID	Seq	3st_LC_Score	3st_RC_Score	3st_Total_Score	학습목표	학습방법	강의학습교재유형	학습빈도	기출문제공부횟수	취약분야인지여부	토익모의테스트횟수
1493	498	3	255	167	422	자기계발	오프라인강의	일반적인영어텍스트반	주 1-2 회	0.0	알고있음	4
1496	499	3	422	370	792	자기계발	오프라인강의	비즈니스시뮬레이션 (Role Play)	주 3-4 회	4.0	알고있음	7
1499	500	3	235	226	461	승진	오프라인강의	비즈니스시뮬레이션 (Role Play)	주 5-6 회	7.0	알고있음	15

500 rows × 12 columns

In [811…
```
#[문제] 토익시험 학습정보 데이터 'df2'의  각 ID별 차시(Seq)가 1인 행을 선택하여 새로운 데이터프
```

In [812…
```
df2.head()
```

Out[812]:

	ID	Seq	LC_Score	RC_Score	Total Score	학습 목표	학습 방법	강의 학습 교재 유형	학습 빈도	기출문제 공부 횟수	취약분야 인지 여부	토익 모의테스트 횟수
0	1	1	181	173	354	자기계발	참고서	일반적인 영어 텍스트 기반 교재	주3-4회	6.0	알고 있지 않음	6
1	1	2	227	213	440	자기계발	오프라인 강의	뉴스/이슈 기반 교재	주1-2회	3.0	알고 있음	5
2	1	3	345	336	681	승진	온라인강의	영상 교재	주5-6회	7.0	알고 있음	10
3	2	1	330	290	620	자기계발	오프라인 강의	뉴스/이슈 기반 교재	매일(주 7회)	8.0	알고 있지 않음	19
4	2	2	354	339	693	승진	온라인강의	영상 교재	주5-6회	2.0	알고 있음	15

In [813…

```
# 1차시 데이터 : ['Seq'] == 1
# loc 함수 활용 : 특정 행, 특정 행과 열, 그리고 조건에 따라 필터링된 행을 선택하는 함수
temp1 = df2.loc[df2['Seq'] == 1]
```

In [814…

```
temp1
```

Out[814]:

	ID	Seq	LC_Score	RC_Score	Total Score	학습 목표	학습 방법	강의 학습 교재 유형	학습 빈도	기출 문제 공부 횟수	취약 분야 인지 여부	토익 모의테스트 횟수
0	1	1	181	173	354	자기계발	참고서	일반적인 영어 텍스트 기반 교재	주3-4회	6.0	알고 있지 않음	6
3	2	1	330	290	620	자기계발	오프라인 강의	뉴스/이슈 기반 교재	매일 (주 7회)	8.0	알고 있지 않음	19
6	3	1	367	309	676	취업	온라인강의	영상 교재	매일 (주 7회)	9.0	알고 있지 않음	7
9	4	1	470	285	755	자기계발	온라인강의	뉴스/이슈 기반 교재	주1-2회	7.0	알고 있지 않음	4
12	5	1	273	372	645	승진	오프라인 강의	비즈니스 시뮬레이션(Role Play)	주5-6회	3.0	알고 있지 않음	13
...	...	...	...	...	...	...	...	...	...	...	...	...
1485	496	1	347	315	662	자기계발	온라인강의	뉴스/이슈 기반 교재	주1-2회	7.0	알고 있지 않음	1
1488	497	1	112	250	362	자기계발	온라인강의	영상 교재	주5-6회	4.0	알고 있지 않음	10
1491	498	1	252	150	402	자기계발	온라인강의	영상 교재	주3-4회	6.0	알고 있지 않음	15
1494	499	1	371	324	695	자기계발	오프라인 강의	비즈니스 시뮬레이션(Role Play)	주1-2회	0.0	알고 있지 않음	5
1497	500	1	169	188	357	자기계발	참고서	일반적인 영어 텍스트 기반 교재	주3-4회	8.0	알고 있지 않음	2

500 rows × 12 columns

In [815…
```
#[문제] 'temp1'에는 df2['ID','LC_Score','RC_Score','Total Score'] 컬럼만 불러와 주세요.
```

In [816…
```
temp1 = temp1[['ID','LC_Score','RC_Score','Total Score']]
```

In [817…
```
temp1
```

Out[817]:

	ID	LC_Score	RC_Score	Total Score
0	1	181	173	354
3	2	330	290	620
6	3	367	309	676
9	4	470	285	755
12	5	273	372	645
...	...	...	...	...
1485	496	347	315	662
1488	497	112	250	362
1491	498	252	150	402
1494	499	371	324	695
1497	500	169	188	357

500 rows × 4 columns

In [818…
```
#[문제] temp1 데이터프레임의 열이름 중 'LC_Score', 'RC_Score', 'Total Score'를 각각 '1st_LC_Sc
```

In [819…
```
# rename 함수 활용, temp1에 저장
temp1 = temp1.rename(columns={'LC_Score':'1st_LC_Score',
                              'RC_Score':'1st_RC_Score',
                              'Total Score':'1st_Total_Score'})
```

In [820…
```
temp1
```

Out[820]:

	ID	1st_LC_Score	1st_RC_Score	1st_Total_Score
0	1	181	173	354
3	2	330	290	620
6	3	367	309	676
9	4	470	285	755
12	5	273	372	645
...	...	...	...	...
1485	496	347	315	662
1488	497	112	250	362
1491	498	252	150	402
1494	499	371	324	695
1497	500	169	188	357

500 rows × 4 columns

In [821…
```
#[문제] 토익시험 학습정보 데이터 'df2'에서 각 ID별 차시(Seq)가 2인 행을 선택하여 새로운 데이터
```

In [822…
```
# 2차시 데이터 : ['Seq'] == 2
# loc 함수 활용 : 특정 행, 특정 행과 열, 그리고 조건에 따라 필터링된 행을 선택
temp2 = df2.loc[df2['Seq'] == 2]
```

In [823…
```
temp2
```

Out[823]:

	ID	Seq	LC_Score	RC_Score	Total Score	학습 목표	학습 방법	강의 학습 교재 유형	학습 빈도	기출문제 공부 횟수	취약 분야 인지 여부	토익 모의 테스트 횟수
1	1	2	227	213	440	자기계발	오프라인 강의	뉴스/이슈 기반 교재	주1-2회	3.0	알고 있음	5
4	2	2	354	339	693	승진	온라인강의	영상 교재	주5-6회	2.0	알고 있음	15
7	3	2	396	365	761	자기계발	온라인강의	영상 교재	주3-4회	7.0	알고 있지 않음	6
10	4	2	495	341	836	자기계발	온라인강의	영상 교재	주1-2회	7.0	알고 있지 않음	7
13	5	2	314	426	740	자기계발	오프라인 강의	비즈니스 시뮬레이션(Role Play)	주5-6회	8.0	알고 있지 않음	10
...	...	...	...	...	...	...	...	...	...	...	...	...
1486	496	2	349	321	670	자기계발	참고서	뉴스/이슈 기반 교재	주3-4회	3.0	알고 있지 않음	4
1489	497	2	120	251	371	자기계발	오프라인 강의	영상 교재	주3-4회	5.0	알고 있지 않음	9
1492	498	2	254	158	412	자기계발	온라인강의	뉴스/이슈 기반 교재	주5-6회	8.0	알고 있지 않음	18
1495	499	2	378	326	704	승진	온라인강의	뉴스/이슈 기반 교재	주5-6회	6.0	알고 있지 않음	12
1498	500	2	172	190	362	자기계발	참고서	뉴스/이슈 기반 교재	매일(주 7회)	10.0	알고 있음	16

500 rows × 12 columns

In [824…
```
#[문제] 'temp2'에는 df2['ID','LC_Score','RC_Score','Total Score'] 컬럼만 불러와 주세요.
```

In [825…
```
temp2 = temp2[['ID','LC_Score','RC_Score','Total Score']]
```

In [826…
```
temp2
```

Out[826]:

	ID	LC_Score	RC_Score	Total Score
1	1	227	213	440
4	2	354	339	693
7	3	396	365	761
10	4	495	341	836
13	5	314	426	740
…	…	…	…	…
1486	496	349	321	670
1489	497	120	251	371
1492	498	254	158	412
1495	499	378	326	704
1498	500	172	190	362

500 rows × 4 columns

In [827…
```
#[문제] temp2 데이터프레임의 열이름 중 'LC_Score', 'RC_Score', 'Total Score'를 각각 '2st_LC_Sc
```

In [828…
```
# rename 함수 활용, temp2에 저장 및 확인
temp2 = temp2.rename(columns={'LC_Score':'2st_LC_Score',
                              'RC_Score':'2st_RC_Score',
                              'Total Score':'2st_Total_Score'})
```

In [829…
```
temp2
```

Out[829]:

	ID	2st_LC_Score	2st_RC_Score	2st_Total_Score
1	1	227	213	440
4	2	354	339	693
7	3	396	365	761
10	4	495	341	836
13	5	314	426	740
...	...	...	...	...
1486	496	349	321	670
1489	497	120	251	371
1492	498	254	158	412
1495	499	378	326	704
1498	500	172	190	362

500 rows × 4 columns

In [830…
```
#[문제] 3회차 토익시험 정보가 모두 포함된 'temp' 데이터 + 1차시 토익시험점수 'temp1' 데이터을
```

In [831…
```
# 'temp'와 'temp1'을 'ID' 열 기준 조인하여 'score_merged_data1'에 저장
# merge 함수 할용, how='outer'
score_merged_data1 = pd.merge(temp, temp1, on='ID', how='outer')
```

In [832…
```
score_merged_data1
```

Out[832]:

	ID	Seq	3st_LC_Score	3st_RC_Score	3st_Total_Score	학습목표	학습방법	강의학습교재유형	학습빈도	기출문제공부횟수	취약분야인지여부	토익모의테스트횟수	1st_LC_Sc
0	1	3	345	336	681	승진	온라인강의	영상교재	주5-6회	7.0	알고있음	10	
1	2	3	380	368	748	승진	온라인강의	뉴스/이슈기반교재	주5-6회	4.0	알고있음	14	
2	3	3	416	382	798	자기계발	참고서	일반적인영어텍스트기반교재	주1-2회	4.0	알고있음	4	
3	4	3	495	397	892	승진	온라인강의	뉴스/이슈기반교재	주3-4회	9.0	알고있음	8	
4	5	3	398	437	835	자기계발	온라인강의	영상교재	주3-4회	6.0	알고있음	4	
...	...	...	...	...	...	...	...	...	...	...	...	...	
495	496	3	364	336	700	자기계발	온라인강의	일반적인영어텍스트기반교재	매일(주7회)	10.0	알고있음	13	
496	497	3	187	252	439	승진	온라인강의	비즈니스시뮬레이션(Role Play)	매일(주7회)	9.0	알고있음	17	
497	498	3	255	167	422	자기	오프라	일반적인영어	주1-	0.0	알고	4	

	ID	Seq	3st_LC_Score	3st_RC_Score	3st_Total_Score	학습목표	학습방법	강의학습교재유형	학습빈도	기출문제공부횟수	취약분야인지여부	토익모의테스트횟수	1st_LC_Sc
						계발	인강의	텍스트기반교재	2회		있음		
498	499	3	422	370	792	자기계발	오프라인강의	비즈니스시뮬레이션(Role Play)	주3-4회	4.0	알고있음	7	
499	500	3	235	226	461	승진	오프라인강의	비즈니스시뮬레이션(Role Play)	주5-6회	7.0	알고있음	15	

In [833…
```
#[문제] 'score_merged_data1'과 'temp2' 데이터를 합쳐서 'score_merged_data2'에 할당하세요.
```

In [834…
```
# 'score_merged_data1'과 'temp2'을 'ID' 열을 기준으로 조인하여 'score_merged_data2'에 저장
# merge 함수 활용, how='outer'
score_merged_data2 = pd.merge(score_merged_data1, temp2, on='ID', how='outer')
```

In [835…
```
score_merged_data2
```

Out[835]:

	ID	Seq	3st_LC_Score	3st_RC_Score	3st_Total_Score	학습목표	학습방법	강의학습교재유형	학습빈도	기출문제공부횟수	취약분야인지여부	토익모의테스트횟수	1st_LC_Sc
0	1	3	345	336	681	승진	온라인강의	영상교재	주5-6회	7.0	알고있음	10	
1	2	3	380	368	748	승진	온라인강의	뉴스/이슈기반교재	주5-6회	4.0	알고있음	14	
2	3	3	416	382	798	자기계발	참고서	일반적인영어텍스트기반교재	주1-2회	4.0	알고있음	4	
3	4	3	495	397	892	승진	온라인강의	뉴스/이슈기반교재	주3-4회	9.0	알고있음	8	
4	5	3	398	437	835	자기계발	온라인강의	영상교재	주3-4회	6.0	알고있음	4	
...	...	...	...	...	...	...	...	...	...	...	...	...	
495	496	3	364	336	700	자기계발	온라인강의	일반적인영어텍스트기반교재	매일(주7회)	10.0	알고있음	13	
496	497	3	187	252	439	승진	온라인강의	비즈니스시뮬레이션(Role Play)	매일(주7회)	9.0	알고있음	17	
497	498	3	255	167	422	자기	오프라	일반적인영어	주1-	0.0	알고	4	

	ID	Seq	3st_LC_Score	3st_RC_Score	3st_Total_Score	학습목표	학습방법	강의학습교재유형	학습빈도	기출문제공부횟수	취약분야인지여부	토익모의테스트횟수	1st_LC_Sc
						계발	인강의	텍스트기반교재	2회		있음		
498	499	3	422	370	792	자기계발	오프라인강의	비즈니스시뮬레이션(Role Play)	주3-4회	4.0	알고있음	7	
499	500	3	235	226	461	승진	오프라인강의	비즈니스시뮬레이션(Role Play)	주5-6회	7.0	알고있음	15	

In [836…
```
#[문제] score_merged_data2 데이터를 확인하세요.
```

In [837…
```
score_merged_data2.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 500 entries, 0 to 499
Data columns (total 18 columns):
 #   Column           Non-Null Count  Dtype
---  ------           --------------  -----
 0   ID               500 non-null    int64
 1   Seq              500 non-null    int64
 2   3st_LC_Score     500 non-null    int64
 3   3st_RC_Score     500 non-null    int64
 4   3st_Total_Score  500 non-null    int64
 5   학습목표             500 non-null    object
 6   학습방법             500 non-null    object
 7   강의 학습 교재 유형      500 non-null    object
 8   학습빈도             500 non-null    object
 9   기출문제 공부 횟수       500 non-null    float64
 10  취약분야 인지 여부       500 non-null    object
 11  토익 모의테스트 횟수      500 non-null    int64
 12  1st_LC_Score     500 non-null    int64
 13  1st_RC_Score     500 non-null    int64
 14  1st_Total_Score  500 non-null    int64
 15  2st_LC_Score     500 non-null    int64
 16  2st_RC_Score     500 non-null    int64
 17  2st_Total_Score  500 non-null    int64
dtypes: float64(1), int64(12), object(5)
memory usage: 70.4+ KB
```

(4) (2)개인정보 데이터와 (3)토익시험 학습정보 합치기

In [838…
```
#[문제] 개인정보 데이터 'df1'과 토익시험 학습정보 'score_merged_data2'를 ID 기준으로 합치고 'b
```

In [839…
```
df1
```

Out[839]:

	ID	Gender	Birth_Year
0	1	M	1973
3	2	F	1982
6	3	F	1995
9	4	M	1987
12	5	M	1994
...	...	...	...
1485	496	M	2006
1488	497	F	1988
1491	498	M	2006
1494	499	F	1990
1497	500	M	1984

500 rows × 3 columns

In [840…
```
score_merged_data2
```

Out[840]:

	ID	Seq	3st_LC_Score	3st_RC_Score	3st_Total_Score	학습목표	학습방법	강의학습교재유형	학습빈도	기출문제공부횟수	취약분야인지여부	토익모의테스트횟수	1st_LC_Sc
0	1	3	345	336	681	승진	온라인강의	영상교재	주5-6회	7.0	알고있음	10	
1	2	3	380	368	748	승진	온라인강의	뉴스/이슈기반교재	주5-6회	4.0	알고있음	14	
2	3	3	416	382	798	자기계발	참고서	일반적인영어텍스트기반교재	주1-2회	4.0	알고있음	4	
3	4	3	495	397	892	승진	온라인강의	뉴스/이슈기반교재	주3-4회	9.0	알고있음	8	
4	5	3	398	437	835	자기계발	온라인강의	영상교재	주3-4회	6.0	알고있음	4	
...	...	...	...	...	...	...	...	...	...	...	...	...	
495	496	3	364	336	700	자기계발	온라인강의	일반적인영어텍스트기반교재	매일(주7회)	10.0	알고있음	13	
496	497	3	187	252	439	승진	온라인강의	비즈니스시뮬레이션(Role Play)	매일(주7회)	9.0	알고있음	17	
497	498	3	255	167	422	자기	오프라	일반적인영어	주1-	0.0	알고	4	

	ID	Seq	3st_LC_Score	3st_RC_Score	3st_Total_Score	학습목표	학습방법	강의학습교재유형	학습빈도	기출문제공부횟수	취약분야인지여부	토익모의테스트횟수	1st_LC_Sc
						계발	인강의	텍스트기반교재	2회		있음		
498	499	3	422	370	792	자기계발	오프라인강의	비즈니스시뮬레이션(Role Play)	주3-4회	4.0	알고있음	7	
499	500	3	235	226	461	승진	오프라인강의	비즈니스시뮬레이션(Role Play)	주5-6회	7.0	알고있음	15	

In [841…
```
# merge 함수, 합쳐진 데이터는 'baseline_data'로 저장
baseline_data = pd.merge(score_merged_data2, df1, on='ID', how='outer')
```

In [842…
```
baseline_data
```

Out[842]:

	ID	Seq	3st_LC_Score	3st_RC_Score	3st_Total_Score	학습목표	학습방법	강의학습교재유형	학습빈도	기출문제공부횟수	취약분야인지여부	토익모의테스트횟수	1st_LC_Sc
0	1	3	345	336	681	승진	온라인강의	영상교재	주5-6회	7.0	알고있음	10	
1	2	3	380	368	748	승진	온라인강의	뉴스/이슈기반교재	주5-6회	4.0	알고있음	14	
2	3	3	416	382	798	자기계발	참고서	일반적인영어텍스트기반교재	주1-2회	4.0	알고있음	4	
3	4	3	495	397	892	승진	온라인강의	뉴스/이슈기반교재	주3-4회	9.0	알고있음	8	
4	5	3	398	437	835	자기계발	온라인강의	영상교재	주3-4회	6.0	알고있음	4	
...	...	...	...	...	...	...	...	...	...	...	...	...	
495	496	3	364	336	700	자기계발	온라인강의	일반적인영어텍스트기반교재	매일(주7회)	10.0	알고있음	13	
496	497	3	187	252	439	승진	온라인강의	비즈니스시뮬레이션(Role Play)	매일(주7회)	9.0	알고있음	17	
497	498	3	255	167	422	자기	오프라	일반적인영어	주1-	0.0	알고	4	

	ID	Seq	3st_LC_Score	3st_RC_Score	3st_Total_Score	학습목표	학습방법	강의학습교재유형	학습빈도	기출문제공부횟수	취약분야인지여부	토익모의테스트횟수	1st_LC_Sc
						계발	인강의	텍스트기반교재	2회		있음		
498	499	3	422	370	792	자기계발	오프라인강의	비즈니스시뮬레이션(Role Play)	주3-4회	4.0	알고있음	7	
499	500	3	235	226	461	승진	오프라인강의	비즈니스시뮬레이션(Role Play)	주5-6회	7.0	알고있음	15	

(5) 레이블 만들기

- **세부 요구사항**
  - 시험 2회차와 3회차의 Score 차이를 구하여 분석하고 싶습니다.
  - 이를 계산해서 'Score_diff_total'이라는 변수로 추가해봅시다.
  - 레이블을 만드는 것은 데이터를 의미있는 방식으로 구분하거나 식별하는 것을 말합니다.

In [843…
```
#[문제] baseline_data 데이터프레임에서 2차시와 3차시의 시험점수 차이를 'Score_diff_total'이라는
```

In [844…
```
# baseline_data의 'Score_diff_total' 변수 = '3st_Total_Score'에서 '2st_Total_Score'를 마이너스
baseline_data['Score_diff_total'] = baseline_data['3st_Total_Score'] - baseline_data['2st_Tota
```

In [845…
```
#[문제] baseline_data 확인해주세요.
```

In [846…
```
baseline_data
```

Out[846]:

	ID	Seq	3st_LC_Score	3st_RC_Score	3st_Total_Score	학습목표	학습방법	강의학습교재유형	학습빈도	기출문제공부횟수	...	토익모의테스트횟수	1st_LC_Sc
0	1	3	345	336	681	승진	온라인강의	영상교재	주5-6회	7.0	...	10	1
1	2	3	380	368	748	승진	온라인강의	뉴스/이슈기반교재	주5-6회	4.0	...	14	3
2	3	3	416	382	798	자기계발	참고서	일반적인영어텍스트기반교재	주1-2회	4.0	...	4	3
3	4	3	495	397	892	승진	온라인강의	뉴스/이슈기반교재	주3-4회	9.0	...	8	4
4	5	3	398	437	835	자기계발	온라인강의	영상교재	주3-4회	6.0	...	4	2
...	...	...	...	...	...	...	...	...	...	...	...	...	
495	496	3	364	336	700	자기계발	온라인강의	일반적인영어텍스트기반교재	매일(주7회)	10.0	...	13	3
496	497	3	187	252	439	승진	온라인강의	비즈니스시뮬레이션(Role Play)	매일(주7회)	9.0	...	17	1
497	498	3	255	167	422	자기	오프라	일반적인영어	주1-	0.0	...	4	2

	ID	Seq	3st_LC_Score	3st_RC_Score	3st_Total_Score	학습목표	학습방법	강의학습교재유형	학습빈도	기출문제공부횟수	...	토익모의테스트횟수	1st_LC_Sc
						계발	인강의	텍스트기반교재	2회				
498	499	3	422	370	792	자기계발	오프라인강의	비즈니스시뮬레이션(Role Play)	주3-4회	4.0	...	7	
499	500	3	235	226	461	승진	오프라인강의	비즈니스시뮬레이션(Role Play)	주5-6회	7.0	...	15	

4.데이터셋 저장하기

- **세부 요구사항**
  - to_csv를 이용하여 전처리된 데이터셋을 저장하세요.
  - 저장할 파일의 확장자는 .csv 입니다.

In [847…
```
#[문제] 전처리된 데이터프레임 'baseline_data'를 CSV 파일로 저장합니다.
```

In [848…
```
# 파일 : 'data04_baseline.csv'
# to_csv 함수 활용, index=False
baseline_data.to_csv('data04_baseline_1_csv', index=False)
```

In [849…
```
#[문제] 제대로 저장이 잘 되었는지 'data04_baseline.csv' 파일을 불러와 확인해보세요.
```

In [850…
```
data4 =  pd.read_csv('data04_baseline_1_csv')
data4
```

Out[850]:

	ID	Seq	3st_LC_Score	3st_RC_Score	3st_Total_Score	학습목표	학습방법	강의학습교재유형	학습빈도	기출문제공부횟수	...	토익모의테스트횟수	1st_LC_Sc
0	1	3	345	336	681	승진	온라인강의	영상교재	주5-6회	7.0	...	10	1
1	2	3	380	368	748	승진	온라인강의	뉴스/이슈기반교재	주5-6회	4.0	...	14	3
2	3	3	416	382	798	자기계발	참고서	일반적인영어텍스트기반교재	주1-2회	4.0	...	4	3
3	4	3	495	397	892	승진	온라인강의	뉴스/이슈기반교재	주3-4회	9.0	...	8	4
4	5	3	398	437	835	자기계발	온라인강의	영상교재	주3-4회	6.0	...	4	2
...	...	...	...	...	...	...	...	...	...	...	...	...	
495	496	3	364	336	700	자기계발	온라인강의	일반적인영어텍스트기반교재	매일(주7회)	10.0	...	13	3
496	497	3	187	252	439	승진	온라인강의	비즈니스시뮬레이션(Role Play)	매일(주7회)	9.0	...	17	1
497	498	3	255	167	422	자기	오프라	일반적인영어	주1-	0.0	...	4	2

	ID	Seq	3st_LC_Score	3st_RC_Score	3st_Total_Score	학습목표	학습방법	강의학습교재유형	학습빈도	기출문제공부횟수	...	토익모의테스트횟수	1st_LC_Sc
						계발	인강의	텍스트기반교재	2회				
498	499	3	422	370	792	자기계발	오프라인강의	비즈니스시뮬레이션(Role Play)	주3-4회	4.0	...	7	3
499	500	3	235	226	461	승진	오프라인강의	비즈니스시뮬레이션(Role Play)	주5-6회	7.0	...	15	1

In [51]:
```
## 고생 정말 많으셨습니다!!
## 실습시간이 남으신 분은 '중급'용 파일에 도전해보세요.
```