

## COSE102 과제6 1차 문제 해결 과정

### 1. 제공된 파일 분석

- 20231231.csv 파일은 '영역', '유형', '표준점수', '남자', '여자' 열로 구성되어 있으며, 각 과목의 표준점수별 남학생 수 및 여학생 수 데이터가 정리되어 있음.

### 2. 각 py 파일의 구성

#### 2-1. module\_a.py

- read\_sunung 함수는 주어진 csv 파일의 데이터를 읽어 DataFrame으로 반환함.
- get\_subjects 함수는 반환된 데이터에서 영역 및 유형을 추출하여 list로 반환함.
- process\_subject\_data 함수는 선택된 과목에 따라 데이터를 필터링하여 남학생/여학생 데이터를 병합하여 DataFrame으로 반환함.

#### 2-2. module\_b.py

- create\_graph 함수는 입력된 DataFrame 데이터와 영역명, 세부 과목명을 기반으로 그래프를 생성하여 출력함.

#### 2-3. main.py

- module\_a의 read\_sunung 함수를 호출하여 csv 파일에서 데이터를 읽어 온 뒤 DataFrame 형식으로 저장함.
- module\_a의 get\_subjects 함수를 호출하여 csv 파일에서 영역 및 유형을 추출하여 list로 저장함.
- choose\_subject 함수는 사용자가 분석할 과목의 영역 및 유형을 선택함. main\_subjects의 영역명을 출력하여 사용자에게서 원하는 영역명을 입력받음. 선택된 영역(main\_choice)이 세부 과목이 존재하는 영역일 경우 해당 세부 과목을 출력하여 사용자에게서 원하는 세부 과목명을 입력받아 sub\_choice에 저장함.
- module\_a의 process\_subject\_data 함수를 호출하여 선택된 과목에 따라 데이터를 선택한 뒤 DataFrame 형식으로 저장함.
- module\_b의 create\_graph 함수를 호출하여 추출된 데이터를 기반으로 그래프를 생성하고 출력함.

### 3. 그래프 구성

- 표준점수별로 남학생의 분포와 여학생의 분포를 명확히 표현하기 위해 선그래프가 적절하다고 판단함.
- 그래프의 제목으로 어떤 과목의 분포인지 나타냄.
- 표준점수의 최댓값 및 최솟값을 기준으로 x축을 나타내며, 10 단위로 표시함.