

PANDAS TEST

총 10문항 × 각 5 소문항 = 50문항

2025. 11. 05. AM09:10 ~ AM11:40

구글 클래스룸에 제출하세요.^^

[1] DataFrame 기초 생성 및 속성 이해

- 1-1. people_basic 데이터를 DataFrame으로 생성하시오.
- 1-2. 컬럼명과 인덱스를 각각 출력하시오.
- 1-3. 행 수와 열 수를 각각 출력하시오.
- 1-4. 나이의 평균과 중앙값을 구하시오.
- 1-5. 나이가 30 이상인 사람만 선택하시오.

[2] 데이터 입출력 및 컬럼/인덱스 조작

- 2-1. students.csv 파일을 불러오시오.
- 2-2. 상위 5개 행을 출력하시오.
- 2-3. 컬럼명을 ['이름','국어','영어','수학']으로 변경하시오.
- 2-4. 인덱스를 이름 컬럼으로 설정하시오.
- 2-5. 다시 인덱스를 초기화(reset_index)하시오.

[3] 조건 필터링과 정렬

- 3-1. mpg가 30 이상인 행만 선택하시오.
- 3-2. manufacturer가 'audi'이고 year가 1999인 행만 선택하시오.
- 3-3. horsepower 기준 내림차순 정렬하시오.
- 3-4. origin별 평균 mpg를 계산하시오.
- 3-5. mpg가 상위 5%에 해당하는 차량만 출력하시오.

[4] 결측치 처리

- 4-1. 컬럼별 결측치 개수를 구하시오.
- 4-2. age 결측치를 평균값으로 대체하시오.
- 4-3. 결측치가 한 개라도 포함된 행을 삭제하시오.
- 4-4. 결측치 여부(True/False)를 표시한 DataFrame을 만드시오.
- 4-5. 결측치 처리 전후의 describe() 결과를 비교하시오.

[5] 그룹화(groupby)와 집계

- 5-1. department별 평균 salary를 계산하시오.
- 5-2. gender별 최대 salary를 구하시오.
- 5-3. (department, rank) 기준으로 그룹화 후 salary 평균을 계산하시오.
- 5-4. agg() 사용해 salary에 대해 평균, 중앙값, 표준편차 동시에 계산하시오.
- 5-5. department별 데이터 개수를 구하고 오름차순으로 정렬하시오.

[6] 피벗(pivot_table)과 멀티인덱스

- 6-1. city를 행, gender를 열로 하여 sales 평균을 계산하시오.
- 6-2. 피벗 결과의 NaN 값을 0으로 채우고 정수형(int)으로 변환하시오.
- 6-3. 각 행의 총합을 나타내는 'total' 컬럼을 추가하시오.
- 6-4. melt() 이용하여 long 형태로 변환하시오.
(id_vars=['city','gender']).
- 6-5. long 데이터에서 city와 gender별 평균 value를 계산하시오.

[7] 병합(merge)과 결합(concat)

- 7-1. user_info와 user_score를 id 기준으로 내부 조인(inner join)하시오.
- 7-2. 컬럼명 충돌 시 suffix를 적용하여 병합하시오.
(예: suffixes=('_L','_R')).
- 7-3. id별 평균 score를 구하시오.
- 7-4. user_info와 평균 score 데이터를 left join 하시오.
- 7-5. 두 데이터를 행 방향으로 결합(concat)하시오 (ignore_index=True).

[8] apply / lambda / 사용자 정의 함수

- 8-1. score1의 제곱근을 계산하시오.
- 8-2. 행 단위(axis=1)로 score1+score2 합을 계산하시오.
- 8-3. name과 age를 결합한 full_name(예: 'Kim_25') 컬럼을 생성하시오.
- 8-4. age가 30 이상이면 '성인', 아니면 '청년'으로 구분하는 새 컬럼을 생성하시오.
- 8-5. 수치형 컬럼 평균을 반환하는 사용자 함수를 정의하고 결과를 구하시오.

[9] 고급 데이터 분석 응용

- 9-1. city별 평균 score를 계산하시오.
- 9-2. score 상위 10%에 해당하는 행만 선택하시오.
- 9-3. cut()으로 age를 3구간으로 나누고 각 구간의 income 평균을 계산하시오.
- 9-4. 수치형 컬럼 간 상관계수(corr)를 계산하시오.
- 9-5. 상관계수의 절댓값이 가장 큰 변수 쌍을 찾으시오.

[10] 인덱스와 멀티인덱스

- 10-1. class와 number를 인덱스로 설정하시오.
- 10-2. subject별 mid 평균을 계산하시오.
- 10-3. 인덱스를 초기화(reset_index)하시오.
- 10-4. 인덱스 이름을 ('반','번호')로 설정하시오.
- 10-5. (class, subject) 기준 mid 평균을 pivot_table로 구성하시오.