

Numpy & Pandas 과목 평가

- 시험시간 : 3시간50분 (14:00 ~ 17:50), 종료시간 이후 제출 시 불인정
- 시험 구성 및 배점: Numpy 5문항, Pandas 3문항, 문제 별 배점 상이, 100점 만점
- 책, 수업시간 작성 소스코드 참고 허용 (단, Chat GPT와 같은 대화형 AI 사용 금지)
- 답안 제출 방법
 - Markdown으로 문제 번호 및 문제 내용 기재 후 코드 셀 추가하여 답안 코드 작성
 - 답안 파일을 "이름.ipynb"으로 하여 하기 메일로 제출
 - 제출처: jihyunsoo@softsociety.net

※ 부정행위(소스코드 공유 등) 적발 시 0점 처리

■ Numpy

1. 다음 보기와 같이 0이상 10미만의 3x3 모양의 정수 난수를 생성하고 각 행의 첫 번째 값으로 정렬하세요. (5점)

[[3 2 0]	→	[[2 1 6]
[9 6 9]		[3 2 0]
[2 1 6]]		[9 6 9]]

2. 8x8 모양의 정수 0으로 채워진 행렬을 만들고 다음과 인덱싱과 슬라이싱을 이용하여 다음과 같은 체크보드 패턴을 만드세요. (5점)

```
[[0 1 0 1 0 1 0 1]
 [1 0 1 0 1 0 1 0]
 [0 1 0 1 0 1 0 1]
 [1 0 1 0 1 0 1 0]
 [0 1 0 1 0 1 0 1]
 [1 0 1 0 1 0 1 0]
 [0 1 0 1 0 1 0 1]
 [1 0 1 0 1 0 1 0]]
```

3. [0,1) 사이의 값을 갖는 크기 10의 랜덤 벡터를 만들고 최대값을 0으로 바꾸는 코드를 만드세요. (5점)

■ Numpy

4. 경계가 1이고 내부가 0인 5X5 모양의 2차원 배열을 만드세요. (5점)

```
[[1 1 1 1 1]
 [1 0 0 0 1]
 [1 0 0 0 1]
 [1 0 0 0 1]
 [1 1 1 1 1]]
```

5. 0이상 10미만의 난수 정수를 3x3모양으로 생성 후 2번째 열 값을 기준으로 정렬하세요. (seed값 0으로 고정) (5점)

```
[[5 0 3]
 [3 7 9]
 [3 5 2]] → [[5 0 3]
 [3 5 2]
 [3 7 9]]
```

■ Pandas

■ 월드컵 출전선수 골 기록 데이터 분석 (30점)

1. 데이터 파일: worldcupgoals.csv

2. 분석 내용

- ① 주어진 전체 기간의 각 나라별 골 득점수 합계 상위 5개 국가와 그 득점수를 출력하시오
- ② 주어진 전체기간동안 골득점을 한 선수가 가장 많은 나라 상위 5개 국가와 그 선수 숫자를 출력하시오.
- ③ Years 컬럼은 년도 -년도 형식으로 구성 되어있고, 각 년도는 4자리 숫자이다. 각 년도 표기가 4자리 숫자로 안된 케이스가 존재한다. 해당 건은 몇 건인지 출력하시오. (hint: for문 이용)
- ④ 3번 문제에서 발생한 예외 케이스(년도 표기가 4자리 숫자가 아닌 경우)를 제외한 새로운 데이터 프레임을 생성하고 전체 데이터 행의 숫자를 출력하시오. (hint: 년도 표기가 4자리 숫자가 아닌 경우를 검사하는 함수를 만든 후 apply 함수를 이용하여 year 컬럼에 적용)
- ⑤ 월드컵 출전횟수를 나타내는 'LenCup' 컬럼을 추가하고 4회 출전한 선수의 숫자를 구하시오.
- ⑥ yugoslavia 국가의 월드컵 출전횟수가 2회인 선수들의 숫자를 구하시오.
- ⑦ 2002년도에 출전한 전체 선수 수를 구하시오.
- ⑧ 월드컵 출전 횟수가 1회 뿐인 선수들 중에서 가장 많은 득점을 올렸던 선수는 누구이며 몇 골을 넣었는지 구하시오.

■ Pandas

■ 전 세계 행복도 지수 데이터 분석 (25점)

1. 데이터 파일: happiness.csv

2. 분석 내용

- ① 각 연도별 행복랭킹 10위안에 드는 나라의 행복점수의 평균을 구하시오.
- ② 각 연도의 행복랭킹 50위 이내의 나라들의 각각의 행복점수 평균을 데이터프레임으로 표시하시오.
- ③ 각 연도별 하위 행복점수의 하위 5개 국가의 평균 행복점수를 구하시오.
- ④ 2019년 데이터를 추출하고 해당 데이터의 상대 GDP 평균 이상의 나라들과 평균 이하의 나라들의 행복점수 평균을 각각 구하고 그 차이 값을 출력하시오
- ⑤ 각 연도의 부패에 대한인식을 내림차순 정렬했을 때 상위 20개 국가의 부패에 대한인식의 평균을 구하시오.
- ⑥ 2018년도 행복랭킹 50위 이내에 포함됐다가 2019년 50위 밖으로 밀려난 국가의 숫자를 구하시오.
- ⑦ 2018년,2019년 모두 기록이 있는 나라들 중 연도별 행복점수가 가장 증가한 나라와 그 증가 수치는?

■ Pandas

■ 유튜브 데이터 분석 (20점)

1. 데이터 파일: youtube_rank.csv

2. 분석 내용

① 컬럼의 이름을 다음과 같이 변경하시오

■ title : 채널명, category : 카테고리, subscriber : 구독자수, view : 조회수, video : 영상개수

② '구독자수' 컬럼과 '조회수' 컬럼의 타입을 정수형(int64)로 변경하시오 ('억','만' 한글을 삭제 후 숫자로 변경)

③ '영상개수' 컬럼의 ','와 '개'문자를 제거하고 정수형(int32)로 변경하시오.

④ 인기 채널 확인하기

■ '구독자수'와 '조회수'가 많은 5개의 채널을 각각 출력하시오.

⑤ '카테고리'가 [음악/댄스/가수]인 채널의 '구독자수' 상위 5개를 출력하시오.

⑥ '카테고리가' [TV/방송]이거나 [게임]인 채널의 개수를 출력하시오.

⑦ '채널명'에 'KBS'가 포함된 채널 명을 출력하시오.

⑧ '카테고리'별 '구독자수', '조회수', '영상개수'의 평균을 출력하시오.