|  |  |
| --- | --- |
| 교육 제목 | **Pandas 와 Numpy를 통한 데이터 처리** |
| 교육 일시 | 21년 10월 1일 |
| 교육 장소 | YGL - 비대면 수업 (자택) |
| **교육 내용** | |
| 오전 | 데이터 사전 처리  누락된 데이터(NaN)이 많으면 데이터의 질이 떨어져서 그걸 없애주는 사전 처리를 거쳐야 한다.  -df[열/행].isnull()  -df.dropna(axis=1,thresh=숫자): 숫자개 이상이 널값이 있는 axis=1열을 제거  -df['열이름'].duplicated() : 특정열의 중복 체크  -df.drop\_duplicates() : 특정 행의 중복행(들)을 제거  -범주형 처리 1-2 참고  -더미변수 교재 참고  -객체.get\_dummies  -fit\_transform |
| 오후 | 데이터 분석과정  데이터 수집 ->데이터 정제 ->기술 통계 분석 및 탐색적 분석 ->미래 예측  Numpy  -다차원 배열이나 행렬과 수학 함수를 지원한다  -벡터화 연산을 이용하여 간단한 토드로도 복잡한 선형 대수 연산을 수행  -배열 인덱싱을 사용한 질의 기능을 이용하여 간단한 코드로도 복잡한 수식을 계산  - import numpy. 또는 import numy as np로 사용한다  -메소드  \*ndarray creation : 배열을 만들어준다  \*numpy.aragne([시작부분/숫자]stop,[step,]dtype=None)  \*ndarray.shape : 배열의 안의 각 차원들의 크기를 알려준다.  \*ndarray.size : 배열 안의 요소의 갯수를 알려준다  \*ndarray.dtype : 요소의 자료형을 알려준다.  리스트를 이용해서 배열 만드는법  리스트 명 = [요소,요소,요소...]  inf와 NaN, 초기값을 설정한 배열 생성  ndarray broadcast란 차원이 다른 배열도 계산가능하다  객체.reshape(행,열)로 배열의 모양을 바꿀수도 있따  데이터 프레임의 다양한 응용  -함수 매핑  \*개별 원소에 함수 매핑: 시리즈 객체[행 이나 열].apply(함수) 메소드를 통해서 객체 안에 있는 요소에 하나하나 함수에 적용이 된다.  -데이터프레임 객체.applymap(함수) : 함수를 데이터프레임의 모든 원소에 적용 열이 기본적으로 들어가나 행을 넣고 싶으면 (함수, axis=1)을 하면 된다. |