|  |  |
| --- | --- |
| 교육 제목 | **판다스 기본 : 시리즈,데이터프레임,연산,외부데이터 가져오기** |
| 교육 일시 | 2021년 11월 08일 |
| 교육 장소 | 영우글로벌러닝 학동역 |
| **교육 내용** | |
| 내용 | 2021.11.08 10일차     1. 판다스 입문    1. 시리즈       1. 인덱스 구조 : 원소의 순서와 주소를 저장.       2. 인덱스를 잘 활용하면 데이터 값의 탐색, 정렬, 선택, 결합 등 데이터 조작을 쉽게 할 수 있다.       3. 인덱스의 종류          1. 정수형 위치 인덱스 : 대괄호안에 숫자입력          2. 인덱스 이름 또는 인덱스 라벨 : 대괄호안에 이름과 함께 따옴표를 입력.       5. 시리즈에 접근    2. 데이터 프레임 : 2차원 배열. R의 데이터프레임에서 유래.       1. 데이터프레임의 열은 시리즈 객체. 시리즈를 열벡터라고하면, 데이터프레임은 여러 개의 열벡터들이 같은 행 인덱스를 기준으로 줄지어 결합된 2차원 벡터 또는 행렬.       2. 데이터프레임은 행과 열을 나타내기위해 두가지 종류의 주소를 사용. 행 인덱스와 열 이름으로 구분       3. 데이터프레임의 각 열은 공통의 속성을 갖는 일련의 데이터를 나타냄.       4. 각 행은 개별 관측 대상에 대한 다양한 속성 데이터들의 모음인 레코드.    3. 데이터프레임 만들기       1. 같은 길이의 배열 여러개가 필요. 데이터 프레임은 여러개의 시리즈를 모아 놓은 집합.       2. 판다스 DataFrame() 함수를 사용. 여러 개의 리스트를 원소로 갖는 딕셔너리를 함수에 전달하는 방식을 주로 활용 2. 시리즈 연산    1. 산술 연산 : 행렬 인덱스를 기준으로 모든 원소를 정렬하고,  동일한 위치에 있는 원소끼리 일대일 대응시켜서 연산처리한다. 이때 원소가 없으면 NaN으로 처리한다. 3. 데이터프레임 연산    1. 데이터프레임은 여러 시리즈가 한데 모인 것이므로, 시리즈 연산을 확장하는 개념으로 이해한다. 먼저 행렬 인덱스는 기준으로 정렬하고, 일대일 대응되는 원소끼리 연산한다.    3. 데이터프레임 vs. 숫자       1. 데이터프레임에 수를 연산하면, 모든 원소에 대해 수를 연산한다. 4. seaborn 모듈에서 데이터 가져오기    1. import seqborn as sns    2. titanic = sns.load\_dataset('titanic') 5. 데이터입출력    1. 외부파일 읽기       1. csv       2. json       3. html       4. ms\_excel       5. SQL       6. HDF5 format |