|  |  |
| --- | --- |
| 교육 제목 | **python 기본** |
| 교육 일시 | 2021.11.10 |
| 교육 장소 | 영우글로벌러닝 |
| **교육 내용** | |
| 오전 | seaborn 라이브러리  seaborn은 Matplotlib의 기능과 스타일을 확장한 파이썬 시각화 도구의 고급버전  hue= ‘구분’ 옵션을 stripplot() 함수 또는 swarmplot() 함수에 추가하면 열의 데이터 값을 구분하여 출력.  Folium 라이브러리 – 지도 활용  Folium 라이브러리의 map()함수를 이용하면 간단하게 기도객체를 만들 수 있음. 지도화면은  고정된 것이 아니고 줌 기능과 이동 기능이 모두 가능  데이터 사전처리  누락데이터 확인  Seaborn 라이브러리의 .head() 함수와 .info()함수를 이용하여Nan값을 확인  누락데이터를 찾는 직접적인 방법  isnull(), notnull() 각 유효한 데이터가 존재하면 True, False를 리턴  누락데이터를 제거  df\_ = df.dropna(subset=['age'],how='any',axis=0)  중복 데이터 확인  duplicated()메소드를 이용 전에 나온 행들과 비교하여 중복되는지 여부를 확인  True 혹은 False를 리턴 |
| 오후 | 데이터 표준화  단위 환산  같은 데이터셋 안에서 서로 다른 측정 단위를 사용한다면, 전체 데이터의 일관성 문제가 발생  측정 단위를 동일하게 맞춰 주어야함  자료형 반환  숫자가 문자열로 저장된 경우 숫자형으로 변환 해야함  범주형 데이터 처리  구간 분할  연속 데이터를 그대로 사용하지 않고 일정 구간으로 나눠서 분석하는 것이 효율 적일 때가 있음  bin= ‘값’을 통해 값만큼 나눔  더미변수  카테고리를 나타내는 범주형 데이터를 데이터 회귀분석 등 머신러닝 알고리즘에 사용할 수 없는  경우가 있는데 컴퓨터가 인식 가능한 값(0,1)으로 변환 해야함  정규화  각 변수에 들어 있는 숫자의 데이터의 상대적 크기 차이 때문에 머신러닝 분석 결과가 달라질 수 있음, 숫자 데이터의 상대적인 크기 차이를 제거할 필요가 있음  시계열 데이터  주식, 환율 등 금융 데이터를 다루기 위해 개발된 판다스는 시계열 데이터를 다루는 기능이 있음 |