|  |  |
| --- | --- |
| 교육 제목 | 판다스 |
| 교육 일시 | 2021-10-01 |
| 교육 장소 | 집(비대면) |
| **교육 내용** | |
| 오전 | isnull(누락일경우 True, 유효할경우 False), notnull(누락일 경우 False, 유효할경우 True) 을 사용하여 누락데이터 확인.  dropna(axis=0 or 1) 을 사용하여 누락데이터를 제거  평균값을 구한 뒤 .fillna를 사용하여 평균값으로 누락데이터 변경  duplicated() 메소드를 통하여 행의 중복여부를 확인 중복된다면 Ture 처음나온다면 False를 반환  drop\_duplicated() 메소드를 통하여 중복자료 제거 |
| 오후 | pandas와 matplotlib의 기반인 numpy관련 개념 학습  관련 문제 풀이 및 복습  데이터의 표준화  단위 환산 : 같은 데이터셋 안에서 서로 다른 측정 단위를 사용한다면 데이터의 일관성 측면에서 문제가 생기므로 단위를 동일하게 맞춰주는것(마일, 야드, 온스 -> 미터, 평, 그램)  자료형 변환 : 숫자가 문자열로 저장된 경우 숫자형(int or float)으로 변환.  dtype or info() 를 사용하여 열의 자료형 확인 후 적절한 자료형으로 변환.  구간 분할 : 구간(bin)으로 나누어 분석하여 효율성을 높힘.  pandas - cut()을 이용하여 여러 구간으로 나누고 범주형 데이터처리  머신러닝 알고리즘에 바로 사용 할 수 없는 경우 더미 변수를 사용하여 특성의 유무를 확인 0과1 로만 구성되여 one hot encording 이라 부름. |