|  |  |
| --- | --- |
| 교육 제목 | **데이터 사전처리,데이터프레임의 다양한 응용(함수매핑)** |
| 교육 일시 | 20211001 |
| 교육 장소 | 비대면 |
| **교육 내용** | |
|  | **데이터 사전처리**  **-**누락 데이터 처리  -누락 데이터 확인: isnull(): 누락데이터면 True, 유효=False  notnull(): 유효한 데이터 존재T, 누락=False    -누락 데이터 제거: dropna()    -누락 데이터 치환: .fillna()    중복 데이터 처리  -중복 데이터 확인: duplicated()  -중복 데이터 제거: drop\_duplicates( )  -특정 열을 기준으로 중복 행을 제거: drop\_duplicates(subset=['\_','\_'])    데이터 표준화  -단위 환산: mpg\_to\_kpl , round():소수점 아래 둘째자리 반올림    -자료형 변환: unique()자료형 확인 ,replace()  Df[“origin”].astype[‘category’] = 문자열을 범주형으로 변환  Df[“origin”].astype[‘str’] = 범주형을 문자열로 다시 변환    범주형 (카테고리) 데이터 처리  -구간분할:연속데이터를 그대로 처리하는 것보단 나눠서 분석 하는게 더 효율적      -더미 변수: 컴퓨터가 인식가능한 입력 값으로 변환 (0과로 표현되는 더미변수사용)    -정규화  -시계열 데이터 |
|  | **데이터프레임의 다양한 응용**  함수 매핑  -개별원소에 함수 매핑  - 시리즈의 원소에 함수 매핑:Series객체.apply(매핑함수)  - 데이터프레임 원소에 함수 매핑: dataFrame객체.apply(매핑함수)    -시리즈 객체에 함수 매핑  -데이터프레임의 각 열에 함수 매핑:  dataFrame객체.apply(매핑함수,axis=0)  -데이터프레임의 각 행에 함수 매핑:  dataFrame객체.apply(매핑함수,axis=1) |