|  |  |
| --- | --- |
| 교육 제목 | **Pandas 기본 및 활용** |
| 교육 일시 | 20211109 09:00 ~ 18:00 |
| 교육 장소 | 자택 온라인 수업 |
| **교육 내용** | |
| 오전 | import seaborn as sns  import pandas as pd  df = sns.load\_dataset('titanic')  df.head() #데이터 확인 후 가져오고자 하는 컬럼 확인 (sex, alive)  df\_sex = df.loc[:, ['sex', 'alive']]  #df\_sex.head()  # male의 생존자 수와 사망자수를 구함  male\_y = df\_sex[(df["sex"]=='male') & (df\_sex["alive"]== "yes")]['sex'].count()  male\_n = df\_sex[(df["sex"]=='male') & (df\_sex["alive"]== "no")]['sex'].count()  # female의 생존자 수와 사망자 수를 구함  female\_y = df\_sex[(df["sex"]=='female') & (df\_sex["alive"]== "yes")]['sex'].count()  female\_n = df\_sex[(df["sex"]=='female') & (df\_sex["alive"]== "no")]['sex'].count()  # 생존자 수와 사망수를 가지고 데이터 프레임 생성  df\_alive\_cnt = df.DataFrame({'male' : [male\_y, male\_n],  'female' : [female\_y, female\_n]}.  index=['생존자 수', '사망자 수'])  df\_alive\_cnt |
| 오후 | # 시도별 전출입 인구수 파일을 가져옴  df = pd.read.excel('./data/시도별 전출입 인구수.xlsx')  # excel 데이터 가져올 때 셀병함한 값 처리  df.fillna(method = 'ffill', inplace = True)  # 서울에서 다른 지역으로 이동한 자료만 추출  # 전출지가 서울특별시, 전입지는 서울특별시 외에 다른 지역  # mask = 전출지가 서울 & 전입지가 서울이 아닌 지역  # df\_seoul = df[mask]  df\_seoul = df(df['전출지별']=="서울특별시") & (df['전입지별']!='서울특별시')  df\_seoul # 추출된 자료 확인  # 전출지별 컬럼 삭제  df\_seoul = df\_seoul.drop  # 전입지별 컬럼명을 전입지로 변경  df\_seoul.rename({'전입지별 : 전입지'}, axis = 1, implace=True)  #df\_seoul  # 전입지 컬럼을 인덱스로 변경  df\_seoul.set\_index('전입지'.inplace = True)  df\_seoul.head() |