|  |  |
| --- | --- |
| 교육 제목 |  |
| 교육 일시 | 2021년 11월 16일 화요일 |
| 교육 장소 | 영우글로벌러닝 2층 |
| **교육 내용** | |
| 오전 | 1. Part 2. 데이터 입출력    1. # 라이브러리 불러오기       1. import pandas as pd    2. # 파일경로를 찾고, 변수 file\_path에 저장       1. file\_path = './data/read\_csv\_sample.csv'    3. # read\_csv() 함수로 데이터프레임 변환. 변수 df1에 저장       1. df1 = pd.read\_csv(file\_path)       2. print(df1)       3. display(df1)       4. print('\n')    4. # read\_csv() 함수로 데이터프레임 변환. 변수 df2에 저장. header=None 옵션       1. df2 = pd.read\_csv(file\_path, header=None)       2. print(df2)       3. display(df2)       4. print('\n')    5. # read\_csv() 함수로 데이터프레임 변환. 변수 df3에 저장. index\_col=None 옵션       1. df3 = pd.read\_csv(file\_path, index\_col=None)       2. print(df3)       3. display(df3)       4. print('\n')    6. # read\_csv() 함수로 데이터프레임 변환. 변수 df4에 저장. index\_col='c0' 옵션       1. df4 = pd.read\_csv(file\_path, index\_col='c0')       2. print(df4)       3. display(df4) |
| 오후 | 1. # tables 리스트의 원소를 iteration하면서 각각 화면 출력    1. for i in range(len(tables)):    2. # print("tables[%s]" % i)    3. print(f'tables{i}')    4. print(tables[i])    5. print('\n') 2. # 'name' 열을 인덱스로 지정    1. df.set\_index(['name'], inplace=True)    2. print(df) 3. 데이터 프레임 생성    1. df = pd.DataFrame( {"a" : [4 ,5, 6], "b" : [7, 8, 9], "c" : [10, 11, 12]}, index = [1, 2, 3]) Specify values for each column.    2. df = pd.DataFrame( [[4, 7, 10], [5, 8, 11], [6, 9, 12]], index=[1, 2, 3], columns=['a', 'b', 'c']) Specify values for each row. a b    3. df = pd.DataFrame( {"a" : [4 ,5, 6], "b" : [7, 8, 9], "c" : [10, 11, 12]}, index = pd.MultiIndex.from\_tuples( [('d',1),('d',2),('e',2)], names=['n','v']))  Create DataFrame with a MultiIndex 4. 데이터 프레임 리쉐이프    1. df.sort\_values('mpg') Order rows by values of a column (low to high). df.sort\_values('mpg',ascending=False) Order rows by values of a column (high to low).    2. df.rename(columns = {'y':'year'}) Rename the columns of a DataFrame df.sort\_index() Sort the index of a DataFrame    3. df.reset\_index() Reset index of DataFrame to row numbers, moving index to columns.    4. df.drop(columns=['Length','Height']) Drop columns from DataFrame 5. 서브셋 관측치 처리    1. df[df.Length > 7] Extract rows that meet logical criteria.    2. df.drop\_duplicates() Remove duplicate rows (only considers columns). df.head(n) Select first n rows.    3. df.tail(n) Select last n rows    4. df.iloc[10:20] Select rows by position.    5. df.loc[:,'x2':'x4'] Select all columns between x2 and x4 (inclusive). df.iloc[:,[1,2,5]] Select columns in positions 1, 2 and 5 (first column is 0). df.loc[df['a'] > 10, ['a','c']] Select rows meeting logical condition, and only the specific columns |