|  |  |
| --- | --- |
| 교육 제목 | AI 프로젝트 기반 S/W 전문가 양성과정 |
| 교육 일시 | 21.11.10 |
| 교육 장소 | local |
| **교육 내용** | |
| 오전 | 1. seaborn    1. import seaborn as sns    2. %load 경로.py    3. ./JungilWoong/python/Data/예제/seaborn.ipynb    4. plt = df\_1.plot(kind="area") seaborn = sns.plot이름    5. plt = plt.plot(index, values) seaborn = sns.regplot(x='변수', y='변수', data=titanic)    6. sns.heatmap(cbar=True) = 컬러바    7. sns.plot(x='x', y='y', hue='변수', data=데이터, ax=fig.add\_subplot(1, 1, 1))  hue=컬럼 한번 더 분리(소분류) 2. folium    1. zip() 여러 개를 하나로 묶어    2. folium.ipynb 둘러보기 3. 데이터 사전처리    1. df\_age = df.dropna(subset = ["age"]) parameter = subset = [“column”] : 해당 축의 null 제거    2. .fillna(method="ffill")[825:830] method = "f"fill은 앞, "b"fill은 뒤의 값으로 채움    3. 실습 sns.titanic    4. 중복값 제거 df[“c1”].duplicated() # [“컬럼”]의 중복 값 확인 df.duplicated() # 인덱스별 중복 값 확인    5. drop\_duplicates() : 행 기준 drop\_duplicates(subset = [“c1”, “c2”]) : 컬럼 기준으로    6. print(df\_mpg["origin"].dtypes = 데이터 타입 살펴보기 df\_mpg["origin"].dtypes = 주소만 return    7. df\_mpg["컬럼명"].astype("자료형") = 자료형으로 변환 |
| ,오후 | 1. 데이터 표준화2    1. df[“컬럼”].unique() 하나씩만 보여주기 때문에 domain이 갖춰지지 않은 value를 찾을 수 있음    2. df["컬럼"].replace("x", np.nan, inplace = True) df컬럼의 “x”를 null로 변환    3. 결측치 처리 -> 데이터 타입 변경 -> 단위 환산    4. np.histogram(카테고리, bins = 개수) 카테고리 데이터 구간을 나눠주고 개수를 설정 함    5. pd.cut(x=DF[“컬럼”], bins = hp\_bin    6. pd.get\_dummies(DF[“col”]) 문자를 컴퓨터가 인식할 수 있는 숫자로 바꾸는데, 이전에 bin으로 잘랐던 category 정보가 column으로, value는 0 or 1 2. sklearn 실습    1. from sklearn import preprocessing as ppc    3. 있었다… 3. 정규화    1. 0~1 사이로 만들어서 작업 4. 시계열 데이터    1. read\_csv(“./stockdata.csv”)    2. DF[“새컬럼명”]pd.to\_datetime(DF[“컬럼”])    3. 실습 |