|  |  |
| --- | --- |
| 교육 제목 | Pandas-data |
| 교육 일시 | 11.10 |
| 교육 장소 | YGL |
| **교육 내용** | |
| 오전 | titanic = sns.load\_dataset('titanic')  sns.set\_style('darkgrid')  #그래프 생성  fig = plt.figure(figsize=(15,5))  ax1 = fig.add\_subplot(1,2,1)  ax2 = fig.add\_subplot(1,2,2)  #선형회귀선 표시  sns.regplot(x='age',  y='fare',  data=titanic,  ax=ax1)  #선형회귀선 미표시  sns.regplot(x='age',  y= 'fare',  data=titanic,  ax=ax2,  fit\_reg=False)  plt.show() |
| 오후 | #df.info()  df['deck'].value\_counts(dropna=False)  df.head().isnull().sum(axis=0)  df.head().notnull  #데이터의 널 여부 확인  df.isnull().sum(axis=0)  # 누락 데이터가 400개 이상인 컬럼 삭제  df\_t = df.dropna(axis=1, thresh=400)  df\_t.head()  # age 컬럼의 NAN 처리  #제거 방법  df\_age = df.dropna(subset=['age','deck'],how= 'any',axis = 0 )  print(len(df),len(df\_age))  #치환방법 :평균  mean-age = df['age'].fillna(df['age'].mean(axis=0))  mean\_age.head(10)  df.head(10) |