프로젝트 1: Modeling Preprocessing

2022254004 안성인

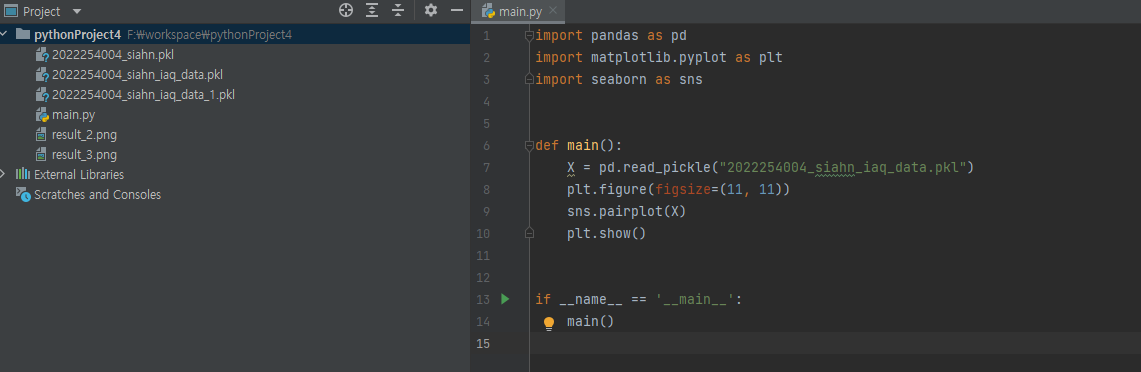
1. Linear Regression에서 종속변수와 독립변수(복수 개 추천)의 선정  
   및 내용 설명

독립변수 : TEMP\_C(섭씨), TEMP\_F(화씨), HUMI(습도),  
**CO2(이산화탄소), TVOC(총휘발성유기화합물),**  
PM\_10(미세먼지), PM\_2\_5(초미세먼지), PM\_1\_0(극초미세먼지)

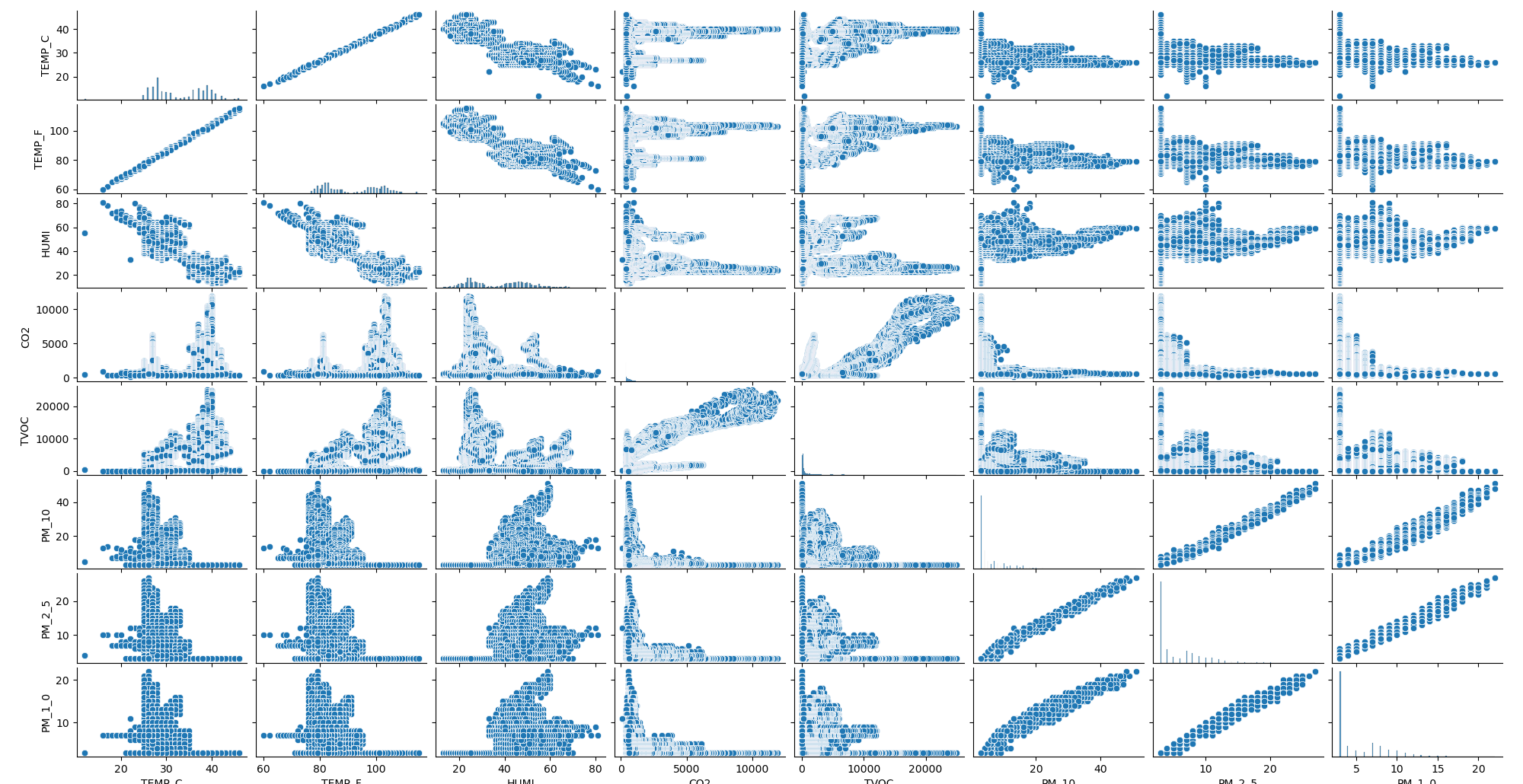
종속변수 : SCORE(공기질)

해당 프로젝트에서 추정하고자 하는 데이터는 공기질 점수(SCORE)로  
이를 위해 공기질 점수 데이터를 종속변수로 두고 해당 데이터를 얻기 위해  
공기질 점수 계산을 해주는 기기에서 같이 전송하는 센서데이터들을 독립변수로 하여  
독립변수들과 종속변수간의 관계를 추정합니다.

1. 아래 pairplot의 결과 및 내용 설명

   
피클 파일은 독립변수만을 저장하여 독립변수간의 관계를 보기 위해 위와 같이 작업하여 동작 시킨 결과 다음과 같은 결과가 나왔습니다.

독립변수간의 관계(동작 결과)가 위와 같이 나오게 되었는데 위 결과에 대한 분석으로는



이산화탄소 농도(CO2)와 총 휘발성 유기화합물(TVOC)은 유사한 데이터 파형을 보이고   
미세먼지 관련 데이터들 또한 유사한 데이터 파형을 보이는데  
이산화탄소 농도 데이터와 총 휘발성 유기화합물 데이터들과 미세먼지 관련 데이터들  
사이에는 반비례 관계가 성립하는 것으로 보입니다.

또한 온도 데이터(섭씨, 화씨)가 높아질수록 습도가 낮아지는 경향인 것을 확인할 수  
있습니다.