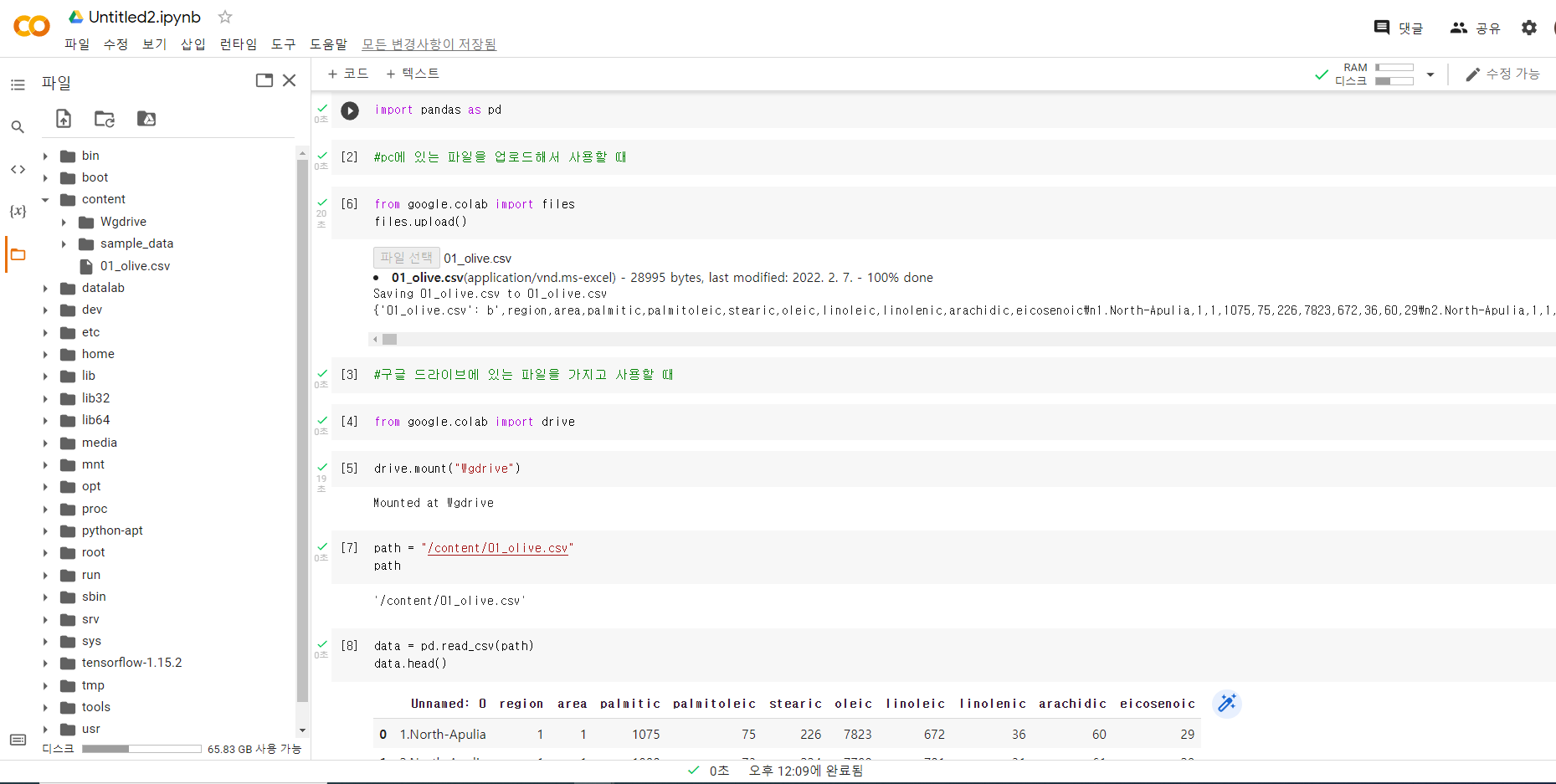
2022-02-08

Pandas(자료 불러오고 저장하기)

* colab으로 데이터 여는 방법
* data = pd.read\_csv(“data/ (tap) , header = , sep = “ ,”, names = [“ “,..], skiprows = ) 🡪 tap누르면 자동으로 뜸

1. header = None : 위에 컬럼까지 같이 내보냄
2. sep = “,” : 디폴트값 ,으로 구분
3. names = [“ “,” “,…] 컬럼 이름 변경
4. skiprows = 3 : 위에서부터 3건너 띠고 내보냄

* csv 파일 : 값, 구분자(“,”) 🡪 read\_csv : 읽기 / to\_csv :내보내기
* 빵구난 값들이거나, 원하는 형식이 아닌 것들을 정리 -> dropna, fillna 🡪 줄 단위로 apply, 벡터 (컬럼선택) df
* 컬럼이름에 대한 변경 : rename EX) data.rename(columns = {“컬럼이름” : “바꾸고 싶은 이름”}, inplace = True) 🡪 이름 대소문자 공백 정확하게 체크! 대소문자 체크하기 싫다면 data.columns[숫자] 이용
* 이름 변경 후 unique한지 체크 EX) data.shape[0] == len(data["바꾼이름"].unique())
* set.index : 컬럼을 인덱스로 변환할 때 EX) data.set\_index("id\_area", inplace=True)
* excel 파일 : sheet name! EX) data\_ex = pd.read\_excel(path, sheet\_name ="Sheet1")
* csv, excel의 공통점 : 일단 불러오면 뭐라도 데이터가 있음.. 불러오면 DataFrame이 생성됨
* 파일이 여러 개가 있을 때 일괄 처리 : glob, 파일을 개별 DF ,문자열을 변수명 vars( ) 🡪 glob, sys, os 각각 import 해줘야 함
  1. 파일 1개를 DF 1개로 개별적으로 다 생성 EX) many\_file\_path = "01\_data/06\_manyfile/\*.csv" file\_path\_list = glob.glob(many\_file\_path)
  2. 여러 개의 파일을 다 합쳐서 1개 통합 파일 생성