

과제보고

2 0 2 1 . 0 1 . 1 1 ~ 1 3

문지현

CONTENTS

01 과제

02 작성한 코드

03 또 다른 시도

04 배운점 및 느낀점

02

총 환자수 구하기 위한 함수

작성한 코드

```
#각 csv파일의 총 포인트 갯수 구하는 함수
def get_number_of_points(csvPath):
    data_df=pd.read_csv(csvPath)

    return data_df.count()[0]
```

각 라벨별 데이터 수 구하기 위한 함수

```
#각각 라벨별 데이터 수 구하고, 리스트형으로 반환
def get_number_of_labels(csvPath):
    data_df=pd.read_csv(csvPath)
    data_df.columns=['X_pointer','Y_pointer','Z_pointer','label']
    df=data_df.sort_values('label')
    i=0
    list1=[]
    while i<15:
        if (df['label']==i).any():
            Label_num=df['label']==i
            list1.insert(i,int(df.loc[Label_num,'label'].value_counts()))
        else:
            list1.insert(i,0)
        i+=1
    return list1
```

02

작성한 코드

메인함수

```
[4] > import numpy as np
      a=np.zeros(15)
      a
      array([0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.])

[5] > a.astype(np.int32)
      array([0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0])
```

numpy array
자료형 변환

`b.astype(np.int32)`

```
#csv파일 불러오기
if __name__ == "__main__":
    csvList = glob.glob("C:/Users\moonj/pointcloud_data/*.csv")

    #총 환자수
    print('총환자수:'+str(len(csvList))+'명')

    a=0 #환자별 총 포인트갯수 다 더할 변수
    #환자별 총 포인트 갯수
    for csvFile in csvList:
        nPoints = get_number_of_points(csvFile)
        a+=nPoints

    #환자별 평균 포인트개수 출력
    print('평균 포인트 개수:'+str(a/len(csvList))+'개')

    b=np.zeros(15) #라벨별 데이터 수 구한 리스트 합할 넘파이 어레이
    for csvFile in csvList:
        Label_Data = np.array(get_number_of_labels(csvFile))
        b=b+Label_Data #넘파이 어레이로 라벨별 데이터 합치기
    print(b)
```

02

작성한 코드 문제점

도출 결과

```
PS C:\Users\moonj\OneDrive\바탕 화면\testproject> python work.py
총환자수:896명
평균 포인트 개수:80724.20758928571개
[5.2453933e+07 9.5852000e+05 1.0055390e+06 1.3400540e+06 8.8202300e+05
 1.1761670e+06 2.4136640e+06 2.1134950e+06 2.8460000e+03 8.9737200e+05
 9.9179500e+05 1.3527390e+06 8.7928400e+05 1.1757330e+06 2.5236770e+06]
```

```
b.astype(np.int32)
```

```
Label_Data = np.array(int(get_number_of_labels(csvFile)))
```

```
Traceback (most recent call last):
  File "c:/Users/moonj/OneDrive/바탕 화면/testproject/work.py", line 51, in <module>
    Label_Data = np.array(int(get_number_of_labels(csvFile)))
TypeError: int() argument must be a string, a bytes-like object or a number, not 'list'
```

타입에러

문제점 : 데이터가 많은데 반복문 코드로 작성해서
실행하는데 시간이 오래걸리는 문제가 발생

02

작성한 코드 문제점

저장공간 지정하고 삽입해야함.

array 타입을 변환하고, 값 삽입

```
[2] ▶ M4
import numpy as np
a=np.zeros(15)
a

array([0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.])

[4] ▶ M4
a.astype(np.int32)

array([0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0])

[5] ▶ M4
np.insert(a,3,1)

array([0., 0., 0., 1., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.])
```

array 값 삽입하고, 타입 변환

```
[1] ▶ M4
import numpy as np
a=np.zeros(15)
a

array([0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.])

[2] ▶ M4
np.insert(a,3,1)

array([0., 0., 0., 1., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.])

[4] ▶ M4
a.astype(np.int32)

array([0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0])
```

정수형으로 라벨별 데이터 수 구함.

```
b=np.zeros(15) #라벨별 데이터 수 구한 리스트 합할 넘파이 어레이
c=np.zeros(15)
b=b.astype(np.int32)
c=c.astype(np.int32)
for csvFile in csvList:
    Label_Data = np.array(get_number_of_labels(csvFile)) #CSV파일 불러와서 넘파이 어레이로 만들기
    c = Label_Data.astype(np.int32) #넘파이 어레이 정수형으로 만들기
    b=b+c #넘파이 어레이로 라벨별 데이터 합치기
print(b)
```

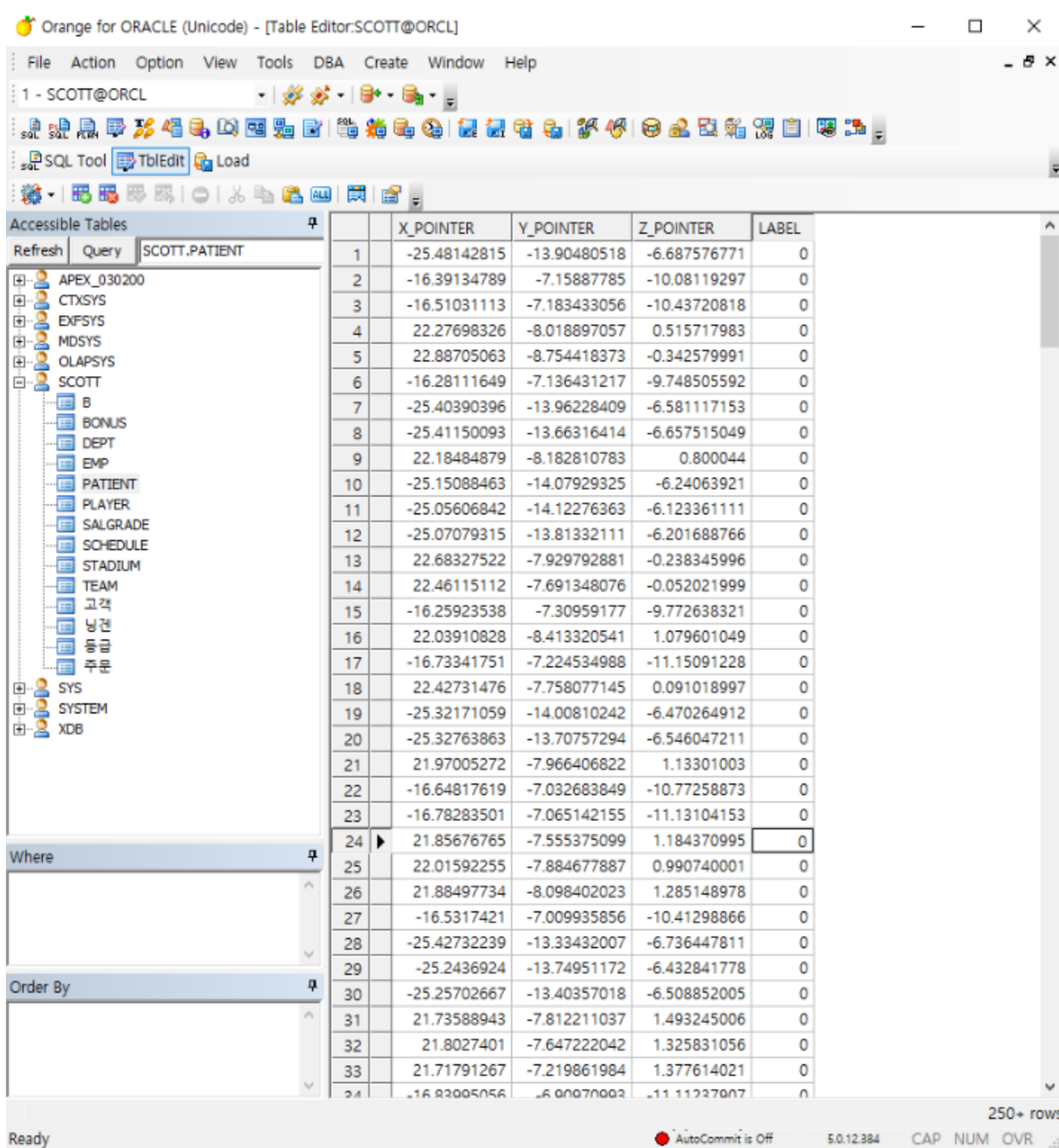
```
(base) C:\Users\moonj\OneDrive\바탕 화면\testproject>C:/Users/moonj/anaconda3/python.exe "c:/Users/moonj/OneDrive/바탕 화면/testproject/work1.py"
총환자수:896명
평균 포인트 개수:80724.20758928571개
[52453933  958520 1005539 1340054  882023 1176167 2413664 2113495
 2846  897372  991795 1352739  879284 1175733 2523677]
```

03

DB에 올려서 추출해보기?

오라클 DB에 CSV파일 LOAD하는 방법 :
<https://geniusjo.tistory.com/entry/SQLLoader-csv%E%8C%8C%EC%9D%BC-import>

또 다른 시도



The screenshot shows the Oracle SQL Developer interface with the SCOTT.PATIENT table selected. The table contains 34 rows of data with columns X_POINTER, Y_POINTER, Z_POINTER, and LABEL. The LABEL column values range from 0 to 15, with 0 appearing 16 times and 15 appearing once.

	X_POINTER	Y_POINTER	Z_POINTER	LABEL
1	-25.48142815	-13.90480518	-6.687576771	0
2	-16.39134789	-7.15887785	-10.08119297	0
3	-16.51031113	-7.183433056	-10.43720818	0
4	22.27698326	-8.018897057	0.515717983	0
5	22.88705063	-8.754418373	-0.342579991	0
6	-16.28111649	-7.136431217	-9.748505592	0
7	-25.40390396	-13.96228409	-6.581117153	0
8	-25.41150093	-13.66316414	-6.657515049	0
9	22.18484879	-8.182810783	0.800044	0
10	-25.15088463	-14.07929325	-6.24063921	0
11	-25.05606842	-14.12276363	-6.123361111	0
12	-25.07079315	-13.81332111	-6.201688766	0
13	22.68327522	-7.929792881	-0.238345996	0
14	22.46115112	-7.691348076	-0.052021999	0
15	-16.25923538	-7.30959177	-9.772638321	0
16	22.03910828	-8.413320541	1.079601049	0
17	-16.73341751	-7.224534988	-11.15091228	0
18	22.42731476	-7.758077145	0.091018997	0
19	-25.32171059	-14.00810242	-6.470264912	0
20	-25.32763863	-13.70757294	-6.546047211	0
21	21.97005272	-7.966406822	1.13301003	0
22	-16.64817619	-7.032683849	-10.77258873	0
23	-16.78283501	-7.065142155	-11.13104153	0
24	21.85676765	-7.555375099	1.184370995	0
25	22.01592255	-7.884677887	0.990740001	0
26	21.88497734	-8.098402023	1.285148978	0
27	-16.5317421	-7.009935856	-10.41298866	0
28	-25.42732239	-13.33432007	-6.736447811	0
29	-25.2436924	-13.74951172	-6.432841778	0
30	-25.25702667	-13.40357018	-6.508852005	0
31	21.73588943	-7.812211037	1.493245006	0
32	21.8027401	-7.647222042	1.325831056	0
33	21.71791267	-7.219861984	1.377614021	0
34	-16.82005056	-6.90070002	-11.11227007	15

라벨별 데이터 갯수 추출

```
14 SELECT LABEL,COUNT(*)
15 FROM PATIENT
16 GROUP BY LABEL
17 ORDER BY LABEL;
```

Result

Grid Result Server Output Text Output

	LABEL	COUNT(*)
1	0	52454534
2	1	958530
3	2	1005549
4	3	1340084
5	4	882031
6	5	1176183
7	6	2413702
8	7	2113521
9	8	2846
10	9	897378
11	10	991804
12	11	1352771
13	12	879301
14	13	1175747
15	14	2523723
16	15	2162082

건수가 다름....

```
(base) C:\Users\moonj\OneDrive\바탕 화면\testproject>C:/Users/moonj/anaconda3/python.exe "c:/Users/moonj/OneDrive/바탕 화면/testproject/work1.py"  
총환자수:896명  
평균 포인트 개수:80724.20758928571개  
[52453933  958520  1005539  1340054  882023  1176167  2413664  2113495  
 2846  897372  991795  1352739  879284  1175733  2523677]
```

14 ▶ SELECT LABEL,COUNT(*)
15 FROM PATIENT
16 GROUP BY LABEL
17 ORDER BY LABEL;

Result

Grid Result Server Output Text Output

	LABEL	COUNT(*)
1	0	52454534
2	1	958530
3	2	1005549
4	3	1340084
5	4	882031
6	5	1176183
7	6	2413702
8	7	2113521
9	8	2846
10	9	897378
11	10	991804
12	11	1352771
13	12	879301
14	13	1175747
15	14	2523723
16	15	2162082

로드할 건수: ALL

생략 건수: 0

허용 오류 수: ALL

바인드 배열: 64 행, 최대 256000 바이트

계속: 지정 사항 없음

사용된 경로: 규약

04

배운점

배운점 및 느낀점

1. glob 함수
2. dataframe의 .any() , .all () 함수
3. 이 외 함수 작성 방법 etc

느낀점

pandas와 numpy 모듈 공부와 코드 작성 연습을 많이 할 것