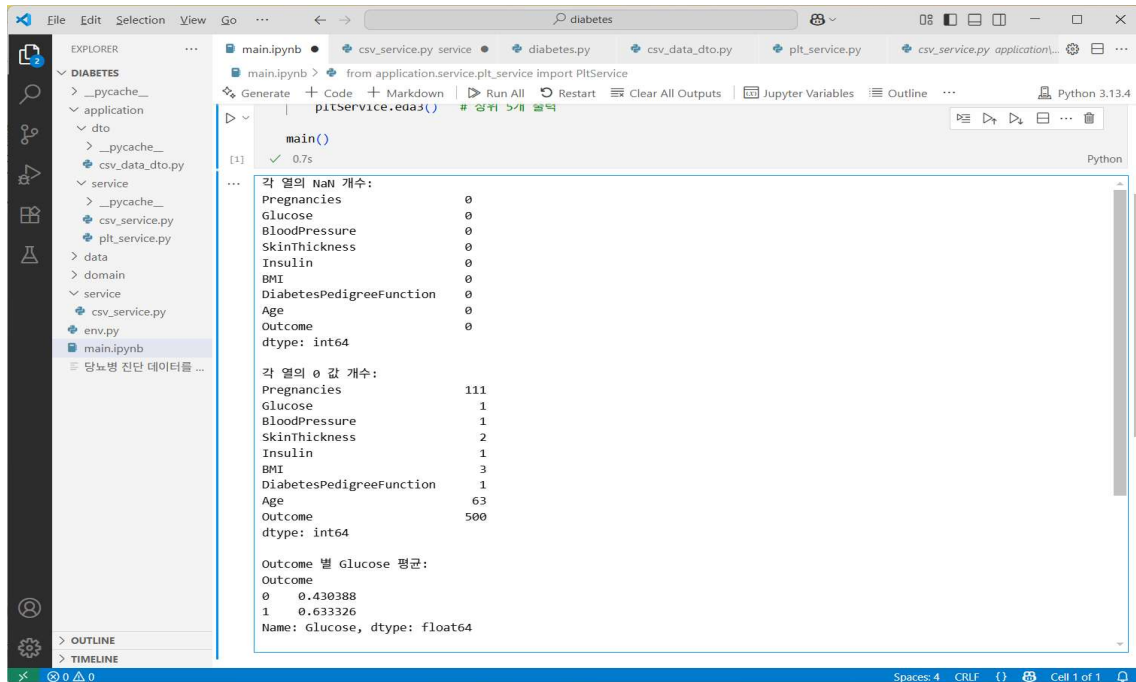


당뇨병 진단 데이터를 활용한 데이터 전처리 및 EDA 분석

26기_장상훈

전체 실행 화면 1



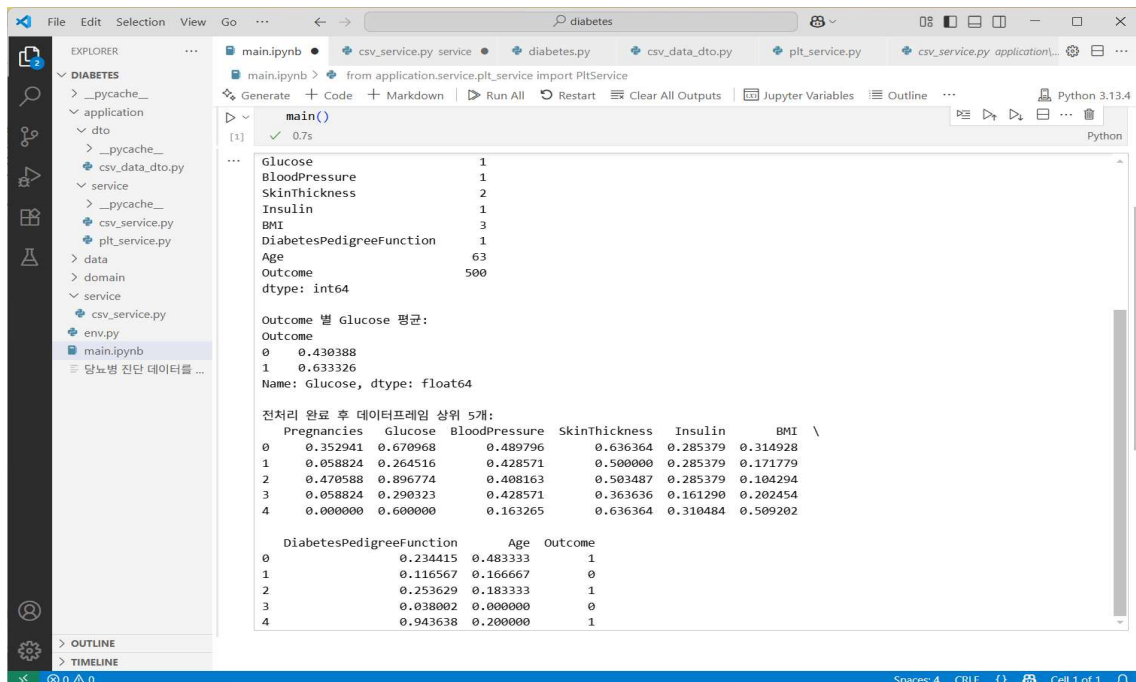
```
main()
[1] ✓ 0.7s

...
각 열의 NaN 개수:
Pregnancies      0
Glucose           0
BloodPressure     0
SkinThickness    0
Insulin          0
BMI              0
DiabetesPedigreeFunction  0
Age              0
Outcome          0
dtype: int64

각 열의 0 값 개수:
Pregnancies      111
Glucose           1
BloodPressure     1
SkinThickness     2
Insulin           1
BMI               3
DiabetesPedigreeFunction  1
Age              63
Outcome          500
dtype: int64

Outcome 별 Glucose 평균:
Outcome
0    0.430388
1    0.633326
Name: Glucose, dtype: float64
```

전체 실행 화면 2



```
main()
[1] ✓ 0.7s

...
Glucose           1
BloodPressure     1
SkinThickness     2
Insulin           1
BMI               3
DiabetesPedigreeFunction  1
Age              63
Outcome          500
dtype: int64

Outcome 별 Glucose 평균:
Outcome
0    0.430388
1    0.633326
Name: Glucose, dtype: float64

전처리 완료 후 데이터프레임 상위 5개:
Pregnancies  Glucose  BloodPressure  SkinThickness  Insulin  BMI \
0    0.352941  0.670968    0.489796    0.636364    0.285379  0.314928
1    0.058824  0.264516    0.428571    0.500000    0.285379  0.171779
2    0.470588  0.896774    0.408163    0.503487    0.285379  0.104294
3    0.058824  0.290323    0.428571    0.363636    0.161290  0.202454
4    0.000000  0.600000    0.163265    0.636364    0.310484  0.509202

DiabetesPedigreeFunction  Age  Outcome
0    0.234415    0.483333    1
1    0.116567    0.166667    0
2    0.253629    0.183333    1
3    0.038002    0.000000    0
4    0.943638    0.200000    1
```

EDA 1 : 결측치 개수 확인

```
각 열의 NaN 개수:
Pregnancies      0
Glucose           0
BloodPressure     0
SkinThickness     0
Insulin           0
BMI               0
DiabetesPedigreeFunction  0
Age              0
Outcome           0
dtype: int64

각 열의 0 값 개수:
Pregnancies      111
Glucose           1
BloodPressure     1
SkinThickness     2
Insulin           1
BMI               3
DiabetesPedigreeFunction  1
Age              63
Outcome           500
dtype: int64
```

EDA 2 : Outcome 별 Glucose 평균

```
Outcome 별 Glucose 평균:
Outcome
0      0.430388
1      0.633326
Name: Glucose, dtype: float64
```

EDA 3 : 전처리 완료 후 데이터프레임 상위 5개

전처리 완료 후 데이터프레임 상위 5개:

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI \
0	0.352941	0.670968	0.489796	0.636364	0.285379	0.314928
1	0.058824	0.264516	0.428571	0.500000	0.285379	0.171779
2	0.470588	0.896774	0.408163	0.503487	0.285379	0.104294
3	0.058824	0.290323	0.428571	0.363636	0.161290	0.202454
4	0.000000	0.600000	0.163265	0.636364	0.310484	0.509202

	DiabetesPedigreeFunction	Age	Outcome
0	0.234415	0.483333	1
1	0.116567	0.166667	0
2	0.253629	0.183333	1
3	0.038002	0.000000	0
4	0.943638	0.200000	1
