# Occupation

#### Introduction:

Special thanks to: https://github.com/justmarkham for sharing the dataset and materials.

### Step 1. Import the necessary libraries

import pandas as pd

Step 2. Import the dataset from this <u>address</u>.

```
url = "https://raw.githubusercontent.com/thieu1995/csv-files/main/data/pandas/u.user"
```

Step 3. Assign it to a variable called users.

```
users = pd.read_csv(url, sep="|")
users.head()
```

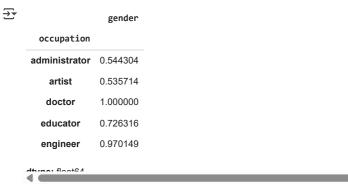


Step 4. Discover what is the mean age per occupation

38.718750 executive healthcare 41.562500 32.571429 homemaker lawyer 36.750000 librarian 40.000000 marketing 37.615385 26.555556 other 34.523810 programmer 33.121212 63.071429 retired salesman 35.666667 35.548387 scientist 22.081633 student technician 33.148148 writer 36.311111 Name: age, dtype: float64

Step 5. Discover the Male ratio per occupation and sort it from the most to the least

```
male_ratio = users[users['gender'] == 'M']
male_ratio = male_ratio.groupby('occupation')['gender'].count() / users.groupby('occupation')['gender'].count()
male_ratio.head()
```



## Step 6. For each occupation, calculate the minimum and maximum ages

```
age_stats = users.groupby("occupation")["age"].agg(['min', 'max'])
print(age_stats)
₹
                     min max
     occupation
     administrator
                      21
                      19
                           48
     artist
     doctor
                      28
                           64
                      23
                           63
     educator
     engineer
                      22
                           70
     entertainment
                      15
                           50
     executive
                      22
                           69
     healthcare
                      22
                           62
     homemaker
                      20
                           50
     lawyer
                      21
                           53
     librarian
                      23
                           69
     marketing
                      24
                           55
                      11
                           55
     none
     other
                      13
                           64
                           63
     programmer
                      20
     retired
                      51
                           73
                      18
                           66
     salesman
                           55
     scientist
                      23
     student
                      7
                           42
     technician
                      21
                           55
min = users.groupby("occupation")['age'].min()
print(min)
→ occupation
     administrator
                       21
     artist
                       19
                       28
     doctor
     educator
                       23
     engineer
                       22
     entertainment
                       15
     executive
                       22
     healthcare
                       22
     homemaker
                       20
     lawyer
                       21
     librarian
     marketing
                       24
     none
                       11
     other
                       13
     programmer
                       20
     retired
                       51
     salesman
                       18
     scientist
                       23
     student
     technician
                       21
                       18
     writer
     Name: age, dtype: int64
max = users.groupby("occupation")['age'].max()
print(max)
→ occupation
     {\tt administrator}
                       70
                       48
     artist
                       64
     doctor
     educator
                       63
     engineer
                       70
```

entertainment

```
executive
                  69
healthcare
                  62
homemaker
                  50
lawyer
librarian
marketing
                  55
                  55
none
                  64
other
programmer
                  63
retired
                  73
salesman
                  66
scientist
                  55
student
                  42
technician
                  55
writer
Name: age, dtype: int64
```

## Step 7. For each combination of occupation and gender, calculate the mean age

```
combination = users.groupby(['occupation', 'gender'])['age'].mean()
print(combination)
    occupation
                    gender
     administrator
                              40.638889
                    Μ
                              37.162791
                              30.307692
                              32.333333
     doctor
                              43.571429
     educator
                              39.115385
                              43.101449
     engineer
                              29.500000
                              36.600000
                    Μ
     entertainment
                              31.000000
                              29.000000
     executive
                              44.000000
                              38.172414
     healthcare
                              39.818182
                              45.400000
     homemaker
                              34.166667
                              23.000000
                              39.500000
     lawver
                    Μ
                              36.200000
     librarian
                              40.000000
                    Μ
                              40.000000
     marketing
                              37.200000
                              37.875000
                              36.500000
                              18.600000
     other
                              35.472222
                              34.028986
                              32,166667
     programmer
                              33.216667
     retired
                              70.000000
                    Μ
                              62,538462
     salesman
                              27.000000
                              38.555556
     scientist
                              28.333333
                              36.321429
```

20.750000 22.669118

38.000000 32.961538

37.631579

35.346154

Step 8. For each occupation present the percentage of women and men

```
gender_counts = users.groupby(['occupation', 'gender']).size()
total_counts = users.groupby('occupation').size()
gender_percentage = gender_counts / total_counts * 100
print(gender_percentage)
→ occupation
     administrator
                               45.569620
                               54.430380
     artist
                    F
                               46.428571
                               53.571429
     doctor
                              100.000000
                               27.368421
```

student

writer

educator

technician

Μ

Name: age, dtype: float64

.39 F W		
	М	72.631579
engineer	F	2.985075
	М	97.014925
entertainment	F	11.111111
	М	88.88889
executive	F	9.375000
	М	90.625000
healthcare	F	68.750000
	М	31.250000
homemaker	F	85.714286
	М	14.285714
lawyer	F	16.666667
•	М	83.333333
librarian	F	56.862745
	М	43.137255
marketing	F	38.461538
_	М	61.538462
none	F	44.44444
	М	55.55556
other	F	34.285714
	М	65.714286
programmer	F	9.090909
	М	90.909091
retired	F	7.142857
	М	92.857143
salesman	F	25.000000
	M	75.000000
scientist	F	9.677419
	М	90.322581
student	F	30.612245
	М	69.387755
technician	F	3.703704
	М	96.296296
writer	F	42.22222
	М	57.777778
dtype: float64		

Start coding or generate with AI.