Import Pandas library

```
import pandas as pd
```

Read the dataset from this address in a data frame 'users'

```
# your code here
df=pd.read_csv("https://raw.githubusercontent.com/justmarkham/DAT8/ma
```

Display the first 15 entries from 'users'

Follow link (ctrl + click)

```
# your code here
print(df.iloc[:15])
```

```
    user_id  age gender     occupation zip_code
0         1   24      M     technician    85711
1         2   53      F          other    94043
2         3   23      M         writer    32067
3         4   24      M     technician    43537
4         5   33      F          other    15213
5         6   42      M      executive    98101
6         7   57      M  administrator    91344
7         8   36      M  administrator    05201
8         9   29      M        student    01002
9        10   53      M         lawyer    90703
10       11   39      F          other    30329
11       12   28      F          other    06405
12       13   47      M       educator    29206
13       14   45      M      scientist    55106
14       15   49      F       educator    97301
```

Display the last 10 entries from 'users'

```
# your code here
print(df.iloc[-10:])
```

```
     user_id  age gender     occupation zip_code
933      934   61      M       engineer    22902
934      935   42      M         doctor    66221
935      936   24      M          other    32789
936      937   48      M       educator    98072
937      938   38      F     technician    55038
938      939   26      F        student    33319
939      940   32      M  administrator    02215
940      941   20      M        student    97229
941      942   48      F      librarian    78209
942      943   22      M        student    77841
```

What is the number of observations in the dataset?

Follow link (ctrl + click)

```
# your code here
print(df.count())
```

```
user_id       943
age           943
gender        943
occupation    943
zip_code      943
dtype: int64
```

What is the number of attributes in the dataset?

```
# your code here
print(len(df.columns))
```

```
5
```

How is the dataset indexed?

```
# your code here
print(df.index)
```

```
RangeIndex(start=0, stop=943, step=1)
```

What is the data type of each column?

```
# your code here
#type(df.columns)
df.dtypes
```

Follow link (ctrl + click)

```
user_id        int64
age            int64
gender        object
occupation    object
zip_code      object
dtype: object
```

Print only the occupation column

```
# your code here
print(df.occupation.to_string(index=False))
```

```
     technician
          other
         writer
     technician
          other
      executive
  administrator
  administrator
```

```
       student
        lawyer
         other
         other
      educator
     scientist
      educator
entertainment
   programmer
         other
     librarian
     homemaker
        writer
        writer
        artist
        artist
      engineer
      engineer
     librarian
        writer
   programmer
        student
         artist
        student
        student
administrator
     homemaker
        student
        student
         other
entertainment
     scientist
      engineer
administrator
     librarian
   technician
   programmer
     marketing
     marketing
administrator
        student
         writer
```

Follow link (ctrl + click)

```
educator
student
programmer
executive
programmer
librarian
none
programmer
```

How many different occupations are in this dataset?

```
# your code here
print(len(df.occupation.unique()))
```

```
21
```

Follow link (ctrl + click)

Summarize the DataFrame.

```
# your code here
print(df.describe())
```

```
          user_id         age
count  943.000000  943.000000
mean   472.000000   34.051962
std    272.364951   12.192740
min      1.000000    7.000000
25%    236.500000   25.000000
50%    472.000000   31.000000
75%    707.500000   43.000000
max    943.000000   73.000000
```

Summarize only the occupation column

```
# your code here
```

```
print(df.occupation.describe())
```

```
count         943
unique         21
top       student
freq          196
Name: occupation, dtype: object
```

What is the mean age of users?

```
# your code here
print(df.age.mean())
```

```
34.05196182396607
```

Follow link (ctrl + click)

What is the mean age per occupation

```
# your code here
print(df.groupby(by='occupation').mean('age').age)
```

```
occupation
administrator    38.746835
artist           31.392857
doctor           43.571429
educator         42.010526
engineer         36.388060
entertainment    29.222222
executive        38.718750
healthcare       41.562500
homemaker        32.571429
lawyer           36.750000
librarian        40.000000
marketing        37.615385
none             26.555556
other            34.523810
programmer       33.121212
```

```
retired          63.071429
salesman         35.666667
scientist        35.548387
student          22.081633
technician       33.148148
writer           36.311111
Name: age, dtype: float64
```

For each occupation, calculate the minimum and maximum ages

```
# your code here
print('Maximum ages:')
print(df.groupby(by='occupation').max('age').age)
print('Minimum ages:')
print(df.groupby(by='occupation').min('age').age)
```

Follow link (ctrl + click)

```
Maximum ages:
occupation
administrator    70
artist           48
doctor           64
educator         63
engineer         70
entertainment    50
executive        69
healthcare       62
homemaker        50
lawyer           53
librarian        69
marketing        55
none             55
other            64
programmer       63
retired          73
salesman         66
scientist        55
student          42
```

```
technician       55
writer           60
Name: age, dtype: int64
Minimum ages:
occupation
administrator    21
artist           19
doctor           28
educator         23
engineer         22
entertainment    15
executive        22
healthcare       22
homemaker        20
lawyer           21
librarian        23
marketing        24
none             11
other            13
programmer       20
retired          51
salesman         18
scientist        23
student           7
technician       21
writer           18
Name: age, dtype: int64
```

Follow link (ctrl + click)

For each combination of occupation and gender, calculate the mean age

```
# your code here
print(df.groupby(by=['occupation','gender']).mean('age').age)
```

```
occupation     gender
administrator  F         40.638889
               M         37.162791
artist         F         30.307692
```

```
                M         32.333333
doctor          M         43.571429
educator        F         39.115385
                M         43.101449
engineer        F         29.500000
                M         36.600000
entertainment   F         31.000000
                M         29.000000
executive       F         44.000000
                M         38.172414
healthcare      F         39.818182
                M         45.400000
homemaker       F         34.166667
                M         23.000000
lawyer          F         39.500000
                M         36.200000
librarian       F         40.000000
                M         40.000000
marketing       F         37.200000
                M         37.875000
none            F         36.500000
                M         18.600000
other           F         35.472222
                M         34.028986
programmer      F         32.166667
                M         33.216667
retired         F         70.000000
                M         62.538462
salesman        F         27.000000
                M         38.555556
scientist       F         28.333333
                M         36.321429
student         F         20.750000
                M         22.669118
technician      F         38.000000
                M         32.961538
writer          F         37.631579
                M         35.346154
Name: age, dtype: float64
```

Follow link (ctrl + click)

Follow link (ctrl + click)