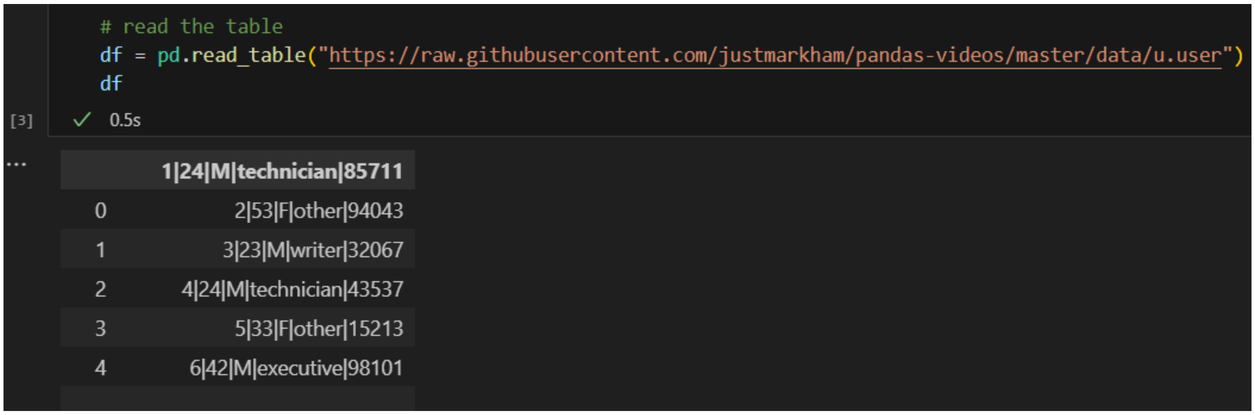
# ***Practical three pandas***

Link of table

https://raw.githubusercontent.com/justmarkham/pandas-videos/master/data/u.user

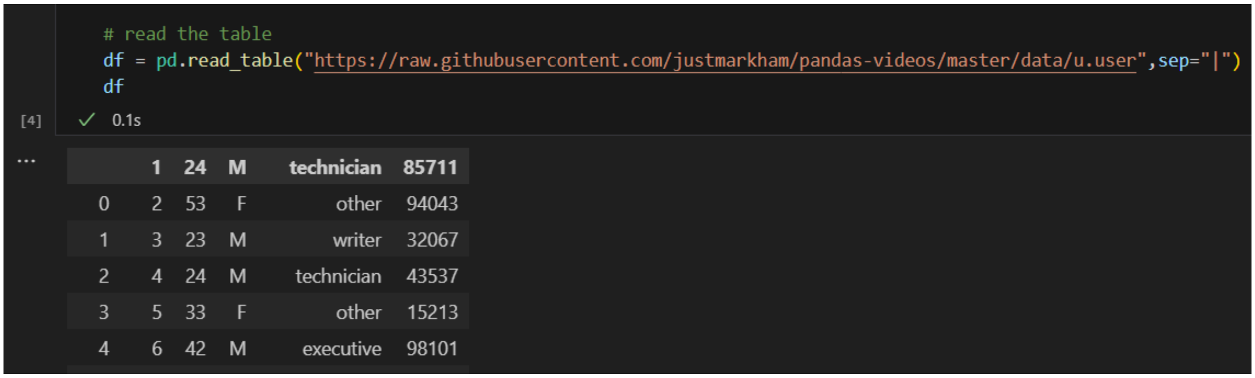


```
# read the table
df = pd.read_table("https://raw.githubusercontent.com/justmarkham/pandas-videos/master/data/u.user")
df
```

```
[3]  ✓ 0.5s

      1|24|M|technician|85711
0          2|53|F|other|94043
1         3|23|M|writer|32067
2     4|24|M|technician|43537
3          5|33|F|other|15213
4      6|42|M|executive|98101
```

1- لو تلاحظ انه كدا عباره عن كولوم واحد بس و لازق ف بعضه لازم نفرق بينهم طب
ازاي أولا نشوف العلامه ال بينهم و بعدين نعمل
Sep = “sign”

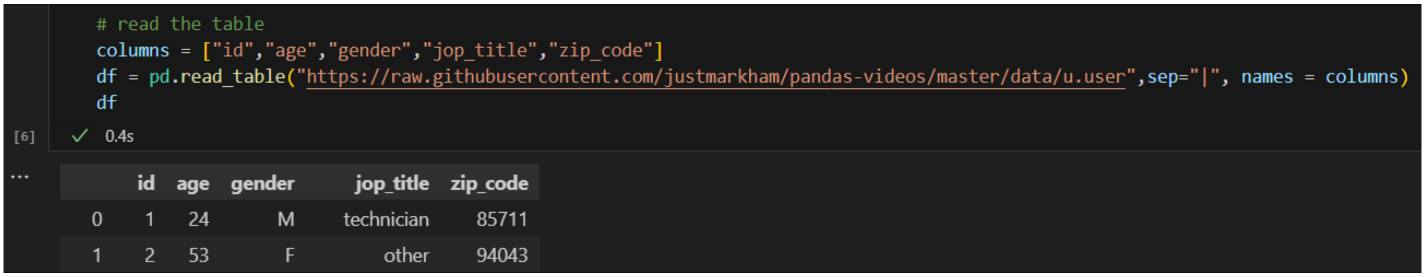


```
# read the table
df = pd.read_table("https://raw.githubusercontent.com/justmarkham/pandas-videos/master/data/u.user",sep="|")
df
```

```
[4]  ✓ 0.1s

   1  24  M  technician  85711
0  2  53  F       other  94043
1  3  23  M      writer  32067
2  4  24  M  technician  43537
3  5  33  F       other  15213
4  6  42  M   executive  98101
```

كدا اتعدلت بس لو تلاحظ ان أسماء الكولوم هي عباره عن داتا ف يلا نسمي الكولوم



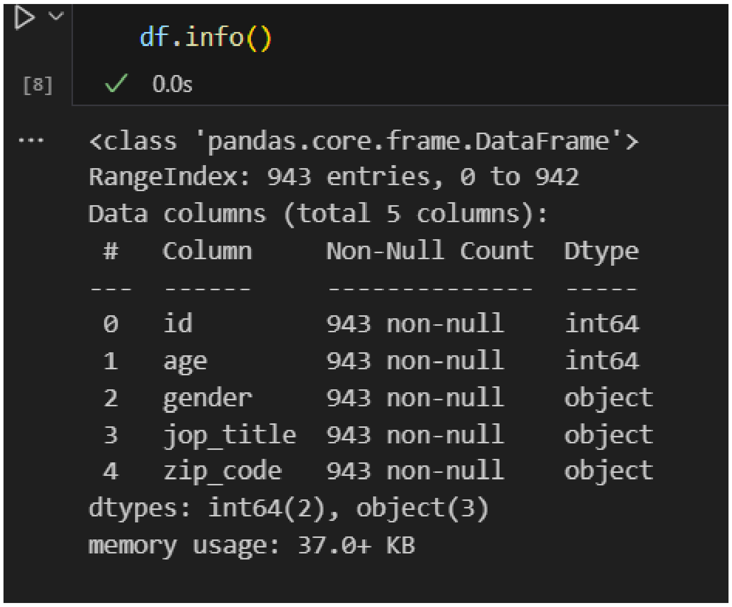
```
# read the table
columns = ["id","age","gender","jop_title","zip_code"]
df = pd.read_table("https://raw.githubusercontent.com/justmarkham/pandas-videos/master/data/u.user",sep="|", names = columns)
df
```

```
[6]  ✓ 0.4s

   id  age  gender   jop_title  zip_code
0   1   24       M  technician     85711
1   2   53       F       other     94043
```

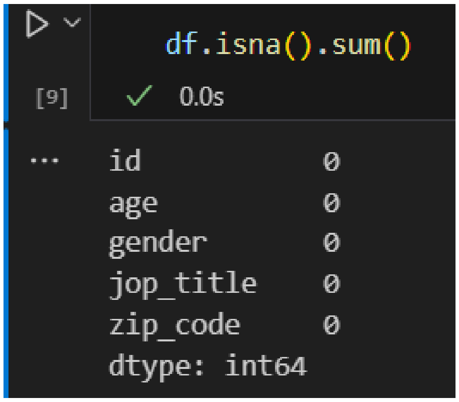
كدا الداتا جاهزه يلاا check data cleaning

1- Missing values



```
df.info()
[8]  ✓ 0.0s
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 943 entries, 0 to 942
Data columns (total 5 columns):
 #   Column     Non-Null Count  Dtype 
---  ------     --------------  ----- 
 0   id         943 non-null    int64 
 1   age        943 non-null    int64 
 2   gender     943 non-null    object
 3   jop_title  943 non-null    object
 4   zip_code   943 non-null    object
dtypes: int64(2), object(3)
memory usage: 37.0+ KB
```

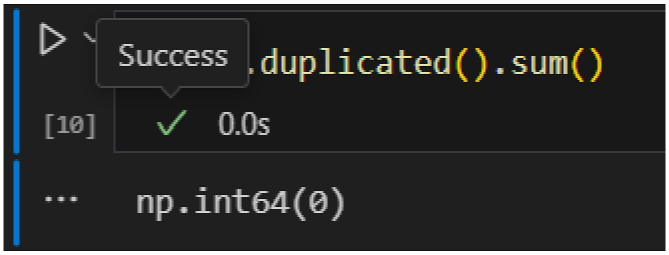


```
df.isna().sum()
[9]  ✓ 0.0s
```

```
id           0
age          0
gender       0
jop_title    0
zip_code     0
dtype: int64
```

No missing value

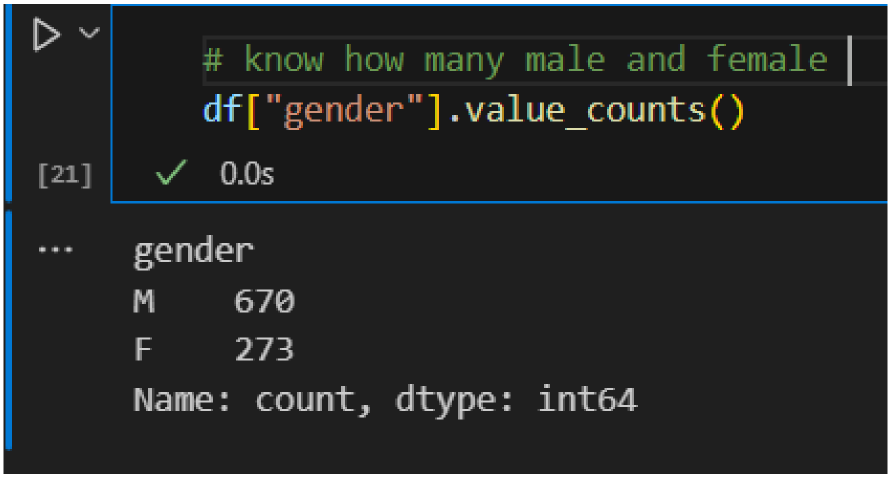
2- Check duplicates



```
.duplicated().sum()
[10]  ✓ 0.0s
```

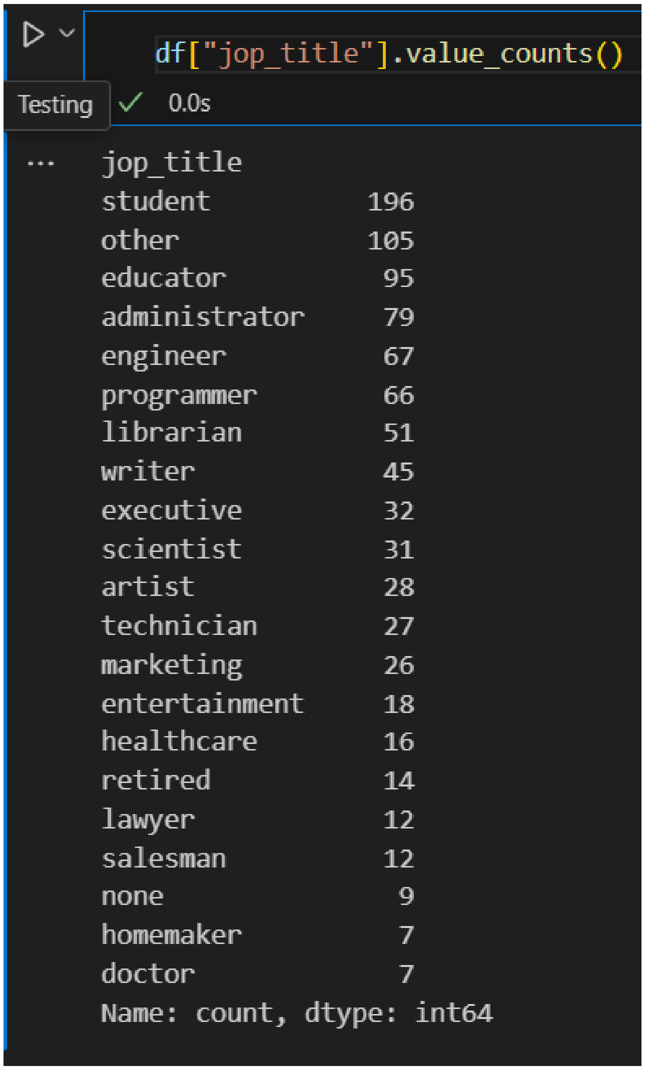
```
np.int64(0)
```

No duplicated rows



```python
# know how many male and female
df["gender"].value_counts()
```

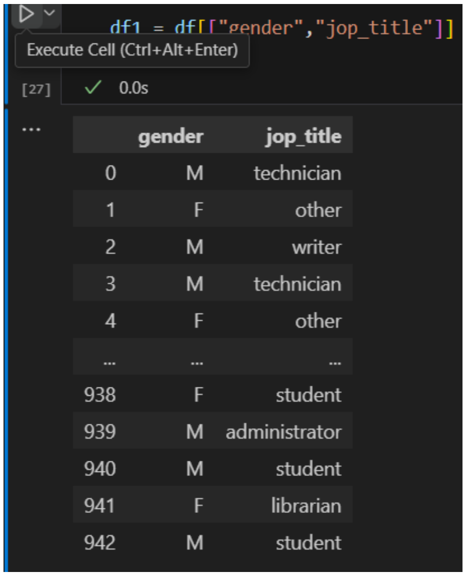
```
gender
M    670
F    273
Name: count, dtype: int64
```



```python
df["jop_title"].value_counts()
```

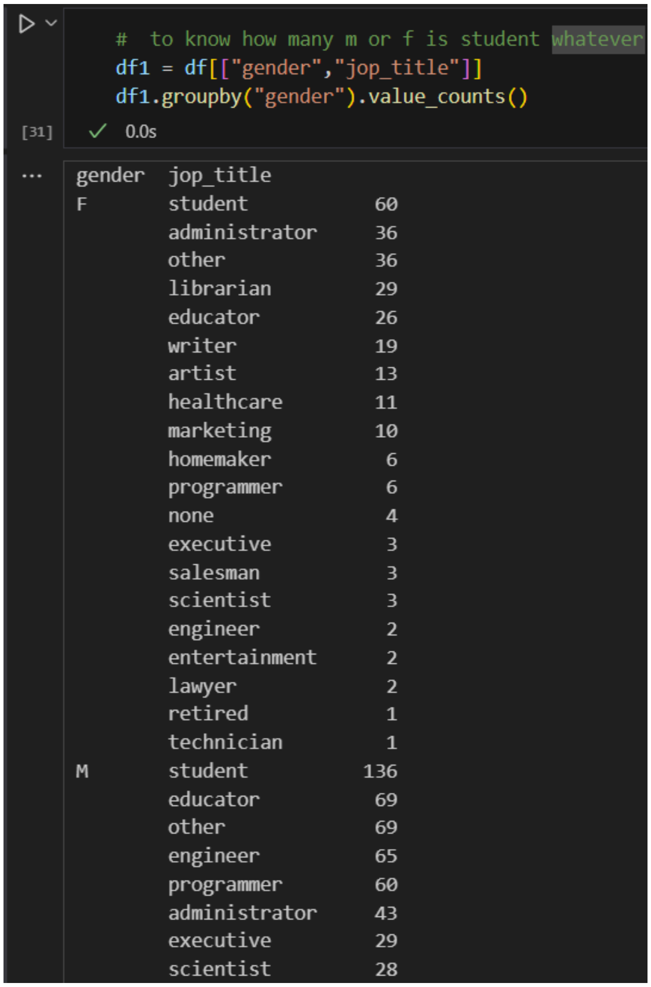
```
jop_title
student          196
other            105
educator          95
administrator     79
engineer          67
programmer        66
librarian         51
writer            45
executive         32
scientist         31
artist            28
technician        27
marketing         26
entertainment     18
healthcare        16
retired           14
lawyer            12
salesman          12
none               9
homemaker          7
doctor             7
Name: count, dtype: int64
```

عاوزين نعرف الشباب اكتر ف انهو وظيفه و البنات اكتر ف أي



```
df1 = df[["gender","jop_title"]]
```

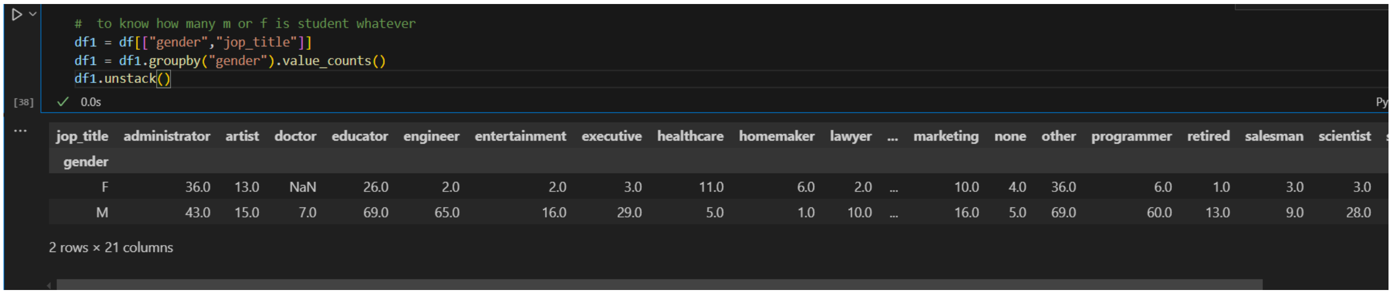
|  | gender | jop_title |
|---|---|---|
| 0 | M | technician |
| 1 | F | other |
| 2 | M | writer |
| 3 | M | technician |
| 4 | F | other |
| ... | ... | ... |
| 938 | F | student |
| 939 | M | administrator |
| 940 | M | student |
| 941 | F | librarian |
| 942 | M | student |



```
#  to know how many m or f is student whatever
df1 = df[["gender","jop_title"]]
df1.groupby("gender").value_counts()
```

```
gender  jop_title
F       student           60
        administrator     36
        other             36
        librarian         29
        educator          26
        writer            19
        artist            13
        healthcare        11
        marketing         10
        homemaker          6
        programmer         6
        none               4
        executive          3
        salesman           3
        scientist          3
        engineer           2
        entertainment      2
        lawyer             2
        retired            1
        technician         1
M       student          136
        educator          69
        other             69
        engineer          65
        programmer        60
        administrator     43
        executive         29
        scientist         28
```

يلا نحسن الشكل شويه



```
#  to know how many m or f is student whatever
df1 = df[["gender","jop_title"]]
df1 = df1.groupby("gender").value_counts()
df1.unstack()
```

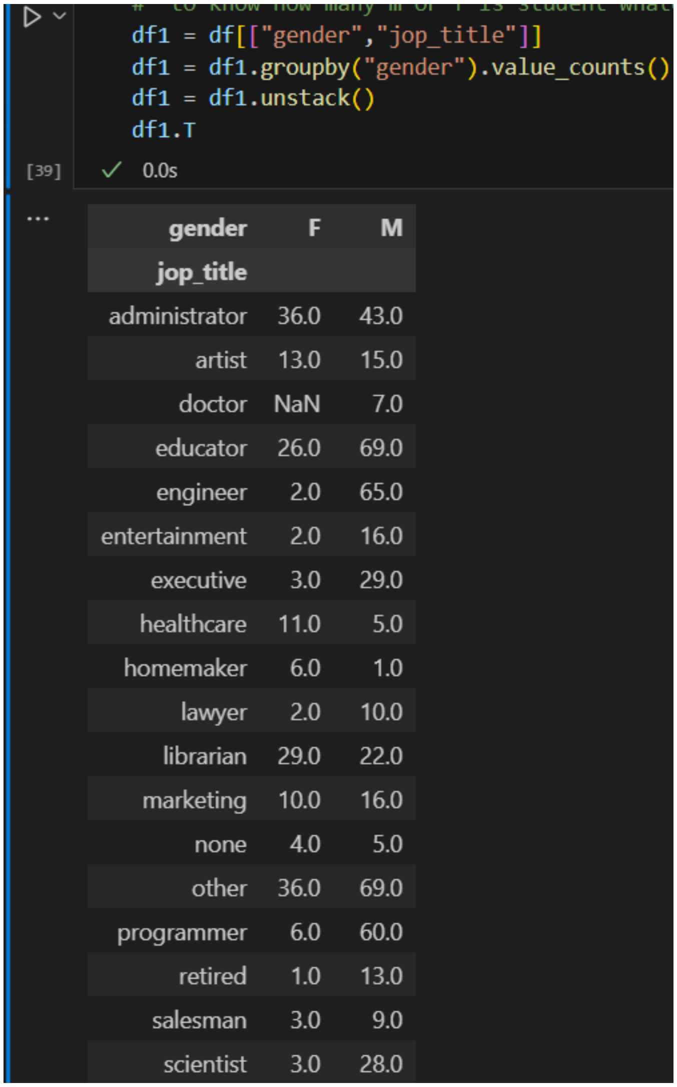
[38] ✓ 0.0s

| jop_title | administrator | artist | doctor | educator | engineer | entertainment | executive | healthcare | homemaker | lawyer | ... | marketing | none | other | programmer | retired | salesman | scientist |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| gender | | | | | | | | | | | | | | | | | | |
| F | 36.0 | 13.0 | NaN | 26.0 | 2.0 | 2.0 | 3.0 | 11.0 | 6.0 | 2.0 | ... | 10.0 | 4.0 | 36.0 | 6.0 | 1.0 | 3.0 | 3.0 |
| M | 43.0 | 15.0 | 7.0 | 69.0 | 65.0 | 16.0 | 29.0 | 5.0 | 1.0 | 10.0 | ... | 16.0 | 5.0 | 69.0 | 60.0 | 13.0 | 9.0 | 28.0 |

2 rows × 21 columns

بس بردو حاسسها رخمه اعمل

Transpose

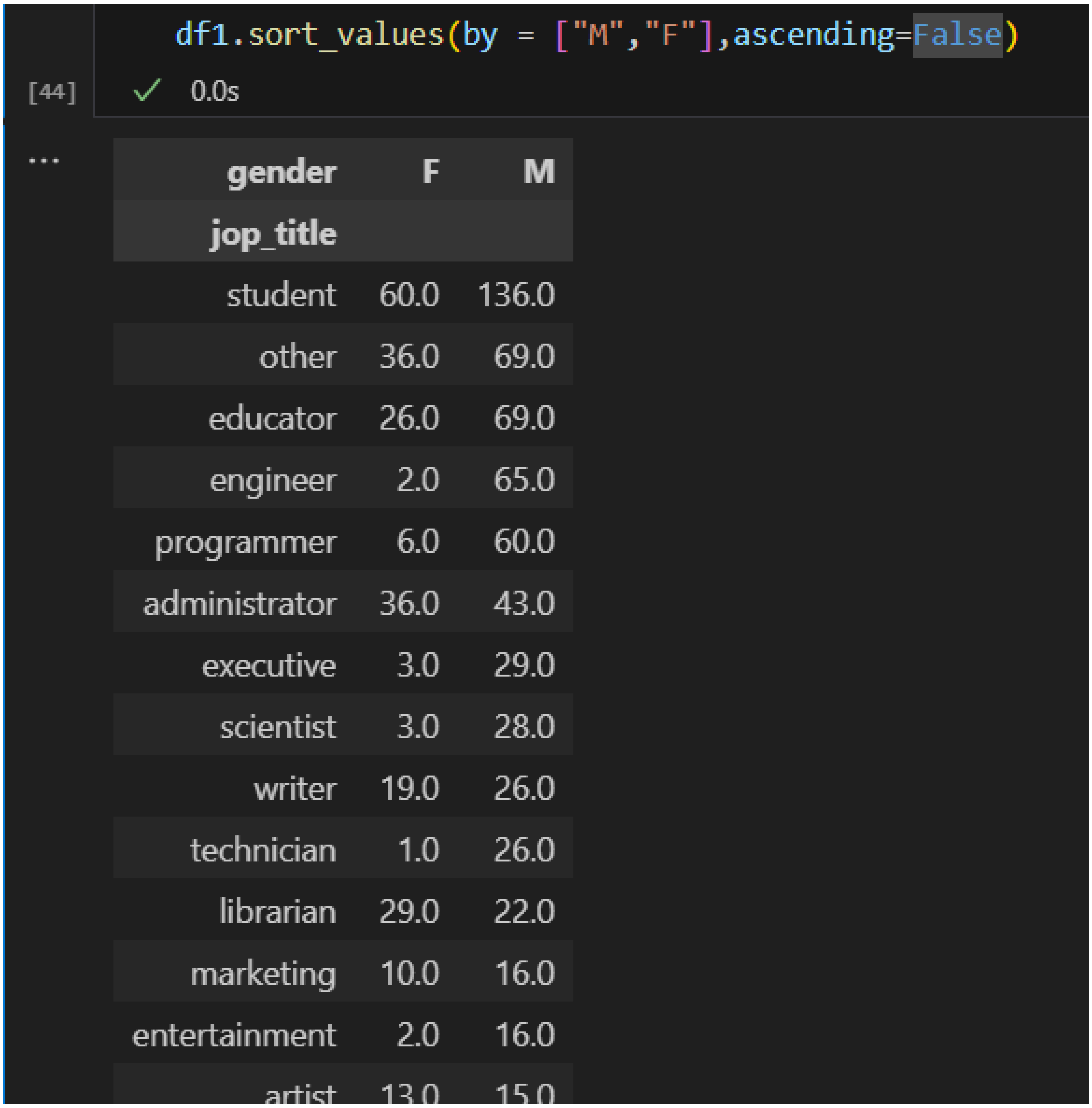


```
#  to know how many m or f is student what
df1 = df[["gender","jop_title"]]
df1 = df1.groupby("gender").value_counts()
df1 = df1.unstack()
df1.T
```

[39] ✓ 0.0s

| gender | F | M |
|---|---|---|
| jop_title | | |
| administrator | 36.0 | 43.0 |
| artist | 13.0 | 15.0 |
| doctor | NaN | 7.0 |
| educator | 26.0 | 69.0 |
| engineer | 2.0 | 65.0 |
| entertainment | 2.0 | 16.0 |
| executive | 3.0 | 29.0 |
| healthcare | 11.0 | 5.0 |
| homemaker | 6.0 | 1.0 |
| lawyer | 2.0 | 10.0 |
| librarian | 29.0 | 22.0 |
| marketing | 10.0 | 16.0 |
| none | 4.0 | 5.0 |
| other | 36.0 | 69.0 |
| programmer | 6.0 | 60.0 |
| retired | 1.0 | 13.0 |
| salesman | 3.0 | 9.0 |
| scientist | 3.0 | 28.0 |

Sort



```
df1.sort_values(by = ["M","F"],ascending=False)
```

[44] ✓ 0.0s

```
       gender     F      M
     jop_title
       student  60.0  136.0
         other  36.0   69.0
      educator  26.0   69.0
      engineer   2.0   65.0
    programmer   6.0   60.0
 administrator  36.0   43.0
     executive   3.0   29.0
     scientist   3.0   28.0
        writer  19.0   26.0
    technician   1.0   26.0
     librarian  29.0   22.0
     marketing  10.0   16.0
 entertainment   2.0   16.0
        artist  13.0   15.0
```