```
import pandas as pd
df = pd.read_csv(r"C:\Users\Jayditya\Downloads\DSBDA LAB\Lab\
Experiments\Datasets\13adult.csv")
df
```

```
       age     workclass  fnlwgt     education  educational-num  \
0       25       Private  226802          11th                7
1       38       Private   89814       HS-grad                9
2       28     Local-gov  336951    Assoc-acdm               12
3       44       Private  160323  Some-college               10
4       18             ?  103497  Some-college               10
...    ...           ...     ...           ...              ...
48837   27       Private  257302    Assoc-acdm               12
48838   40       Private  154374       HS-grad                9
48839   58       Private  151910       HS-grad                9
48840   22       Private  201490       HS-grad                9
48841   52  Self-emp-inc  287927       HS-grad                9

           marital-status         occupation relationship   race
gender  \
0           Never-married  Machine-op-inspct    Own-child  Black
Male
1      Married-civ-spouse    Farming-fishing      Husband  White
Male
2      Married-civ-spouse    Protective-serv      Husband  White
Male
3      Married-civ-spouse  Machine-op-inspct      Husband  Black
Male
4           Never-married                  ?    Own-child  White
Female
...                   ...                ...          ...    ...
..
48837  Married-civ-spouse       Tech-support         Wife  White
Female
48838  Married-civ-spouse  Machine-op-inspct      Husband  White
Male
48839             Widowed       Adm-clerical    Unmarried  White
Female
48840       Never-married       Adm-clerical    Own-child  White
Male
48841  Married-civ-spouse    Exec-managerial         Wife  White
Female

       capital-gain  capital-loss  hours-per-week native-country
income
0                 0             0              40  United-States
<=50K
1                 0             0              50  United-States
<=50K
2                 0             0              40  United-States
```

```
>50K
3              7688             0              40  United-States
>50K
4                 0             0              30  United-States
<=50K
...             ...           ...             ...            ...    ..
.
48837             0             0              38  United-States
<=50K
48838             0             0              40  United-States
>50K
48839             0             0              40  United-States
<=50K
48840             0             0              20  United-States
<=50K
48841         15024             0              40  United-States
>50K

[48842 rows x 15 columns]
```

```
df.groupby("gender")["age"].describe()
```

```
          count       mean        std   min   25%   50%   75%   max
gender
Female  16192.0  36.927989  14.137423  17.0  25.0  35.0  46.0  90.0
Male    32650.0  39.494395  13.412850  17.0  29.0  38.0  48.0  90.0
```

```
df.agg(['sum', 'min','max'])
```

```
         age                                                workclass
fnlwgt  \
sum  1887430  PrivatePrivateLocal-govPrivate?Private?Self-em...
9263575662
min       17                                                    ?
12285
max       90                                         Without-pay
1490400

                                                 education  educational-
num  \
sum  11thHS-gradAssoc-acdmSome-collegeSome-college1...
492234
min                                                   10th
1
max                                        Some-college
16

                                           marital-status  \
sum  Never-marriedMarried-civ-spouseMarried-civ-spo...
min                                           Divorced
```

```
max                                                    Widowed

                                                    occupation  \
sum  Machine-op-inspctFarming-fishingProtective-ser...
min                                                          ?
max                                      Transport-moving

                                                  relationship  \
sum  Own-childHusbandHusbandHusbandOwn-childNot-in-...
min                                                    Husband
max                                                       Wife

                                                         race  \
sum  BlackWhiteWhiteBlackWhiteWhiteBlackWhiteWhiteW...
min                                    Amer-Indian-Eskimo
max                                                    White

                                                       gender  capital-
gain  \
sum  MaleMaleMaleMaleFemaleMaleMaleMaleFemaleMaleMa...      52703821

min                                                    Female             0

max                                                      Male         99999


     capital-loss  hours-per-week  \
sum       4273788         1974310
min             0               1
max          4356              99

                                               native-country  \
sum  United-StatesUnited-StatesUnited-StatesUnited-...
min                                                          ?
max                                             Yugoslavia

                                                       income
sum  <=50K<=50K>50K>50K<=50K<=50K<=50K>50K<=50K<=50...
min                                                    <=50K
max                                                     >50K
```

```
df.groupby("marital-status")["age"].mean()
```

```
marital-status
Divorced                  43.159204
Married-AF-spouse         31.945946
Married-civ-spouse        43.353724
Married-spouse-absent     40.613057
Never-married             28.128064
Separated                 39.725490
```

```
Widowed                   59.377470
Name: age, dtype: float64
```

```
df.groupby("marital-status")["age"].median()
```

```
marital-status
Divorced                  42.0
Married-AF-spouse         30.0
Married-civ-spouse        42.0
Married-spouse-absent     40.0
Never-married             25.0
Separated                 39.0
Widowed                   60.0
Name: age, dtype: float64
```

```
df.groupby(["gender","marital-status"])["age"].std()
```

```
gender  marital-status
Female  Divorced                  10.794868
        Married-AF-spouse         12.342744
        Married-civ-spouse        11.402805
        Married-spouse-absent     13.019854
        Never-married             10.231671
        Separated                 10.757639
        Widowed                   11.657268
Male    Divorced                  10.161659
        Married-AF-spouse          6.336522
        Married-civ-spouse        12.080786
        Married-spouse-absent     12.631023
        Never-married              9.717602
        Separated                 10.811704
        Widowed                   14.216489
Name: age, dtype: float64
```

```
df.groupby("income")["age"].mean()
```

```
income
<=50K    36.872184
>50K     44.275178
Name: age, dtype: float64
```

```
df.groupby(["income","gender"])["age"].mean()
```

```
income  gender
<=50K   Female    36.278999
        Male      37.248548
>50K    Female    42.219333
        Male      44.641863
Name: age, dtype: float64
```

```
df.groupby("marital-status")["marital-status"].count()
```

```
marital-status
Divorced                  6633
Married-AF-spouse           37
Married-civ-spouse       22379
Married-spouse-absent      628
Never-married            16117
Separated                 1530
Widowed                   1518
Name: marital-status, dtype: int64
```

```
#Count number of records by category
#The value_counts() method counts the number of records for each
category in a column.
df["marital-status"].value_counts()
```

```
marital-status
Married-civ-spouse       22379
Never-married            16117
Divorced                  6633
Separated                 1530
Widowed                   1518
Married-spouse-absent      628
Married-AF-spouse           37
Name: count, dtype: int64
```

```
df["marital-status"].count()
```

```
48842
```

```
d = pd.read_csv(r"C:\Users\Jayditya\Downloads\DSBDA LAB\Lab\
Experiments\Datasets\13Iris.csv")
print('Iris-setosa')
setosa = d['Species'] == 'Iris-setosa'
print(d[setosa].describe())
print('\nIris-versicolor')
setosa = d['Species'] == 'Iris-versicolor'
print(d[setosa].describe())
print('\nIris-virginica')
setosa = d['Species'] == 'Iris-virginica'
print(d[setosa])
print(d[setosa].describe())
```

```
Iris-setosa
             Id  SepalLengthCm  SepalWidthCm  PetalLengthCm
PetalWidthCm
count  50.00000       50.00000     50.000000      50.000000
50.00000
mean   25.50000        5.00600      3.418000       1.464000
0.24400
std    14.57738        0.35249      0.381024       0.173511
0.10721
```

```
min     1.00000        4.30000      2.300000       1.000000
0.10000
25%    13.25000        4.80000      3.125000       1.400000
0.20000
50%    25.50000        5.00000      3.400000       1.500000
0.20000
75%    37.75000        5.20000      3.675000       1.575000
0.30000
max    50.00000        5.80000      4.400000       1.900000
0.60000

Iris-versicolor
              Id  SepalLengthCm  SepalWidthCm  PetalLengthCm
PetalWidthCm
count   50.00000      50.000000     50.000000      50.000000
50.000000
mean    75.50000       5.936000      2.770000       4.260000
1.326000
std     14.57738       0.516171      0.313798       0.469911
0.197753
min     51.00000       4.900000      2.000000       3.000000
1.000000
25%     63.25000       5.600000      2.525000       4.000000
1.200000
50%     75.50000       5.900000      2.800000       4.350000
1.300000
75%     87.75000       6.300000      3.000000       4.600000
1.500000
max    100.00000       7.000000      3.400000       5.100000
1.800000

Iris-virginica
      Id  SepalLengthCm  SepalWidthCm  PetalLengthCm  PetalWidthCm  \
100  101            6.3           3.3            6.0           2.5
101  102            5.8           2.7            5.1           1.9
102  103            7.1           3.0            5.9           2.1
103  104            6.3           2.9            5.6           1.8
104  105            6.5           3.0            5.8           2.2
105  106            7.6           3.0            6.6           2.1
106  107            4.9           2.5            4.5           1.7
107  108            7.3           2.9            6.3           1.8
108  109            6.7           2.5            5.8           1.8
109  110            7.2           3.6            6.1           2.5
110  111            6.5           3.2            5.1           2.0
111  112            6.4           2.7            5.3           1.9
112  113            6.8           3.0            5.5           2.1
113  114            5.7           2.5            5.0           2.0
114  115            5.8           2.8            5.1           2.4
115  116            6.4           3.2            5.3           2.3
```

116	117	6.5	3.0	5.5	1.8
117	118	7.7	3.8	6.7	2.2
118	119	7.7	2.6	6.9	2.3
119	120	6.0	2.2	5.0	1.5
120	121	6.9	3.2	5.7	2.3
121	122	5.6	2.8	4.9	2.0
122	123	7.7	2.8	6.7	2.0
123	124	6.3	2.7	4.9	1.8
124	125	6.7	3.3	5.7	2.1
125	126	7.2	3.2	6.0	1.8
126	127	6.2	2.8	4.8	1.8
127	128	6.1	3.0	4.9	1.8
128	129	6.4	2.8	5.6	2.1
129	130	7.2	3.0	5.8	1.6
130	131	7.4	2.8	6.1	1.9
131	132	7.9	3.8	6.4	2.0
132	133	6.4	2.8	5.6	2.2
133	134	6.3	2.8	5.1	1.5
134	135	6.1	2.6	5.6	1.4
135	136	7.7	3.0	6.1	2.3
136	137	6.3	3.4	5.6	2.4
137	138	6.4	3.1	5.5	1.8
138	139	6.0	3.0	4.8	1.8
139	140	6.9	3.1	5.4	2.1
140	141	6.7	3.1	5.6	2.4
141	142	6.9	3.1	5.1	2.3
142	143	5.8	2.7	5.1	1.9
143	144	6.8	3.2	5.9	2.3
144	145	6.7	3.3	5.7	2.5
145	146	6.7	3.0	5.2	2.3
146	147	6.3	2.5	5.0	1.9
147	148	6.5	3.0	5.2	2.0
148	149	6.2	3.4	5.4	2.3
149	150	5.9	3.0	5.1	1.8

	Species
100	Iris-virginica
101	Iris-virginica
102	Iris-virginica
103	Iris-virginica
104	Iris-virginica
105	Iris-virginica
106	Iris-virginica
107	Iris-virginica
108	Iris-virginica
109	Iris-virginica
110	Iris-virginica
111	Iris-virginica
112	Iris-virginica

```
113  Iris-virginica
114  Iris-virginica
115  Iris-virginica
116  Iris-virginica
117  Iris-virginica
118  Iris-virginica
119  Iris-virginica
120  Iris-virginica
121  Iris-virginica
122  Iris-virginica
123  Iris-virginica
124  Iris-virginica
125  Iris-virginica
126  Iris-virginica
127  Iris-virginica
128  Iris-virginica
129  Iris-virginica
130  Iris-virginica
131  Iris-virginica
132  Iris-virginica
133  Iris-virginica
134  Iris-virginica
135  Iris-virginica
136  Iris-virginica
137  Iris-virginica
138  Iris-virginica
139  Iris-virginica
140  Iris-virginica
141  Iris-virginica
142  Iris-virginica
143  Iris-virginica
144  Iris-virginica
145  Iris-virginica
146  Iris-virginica
147  Iris-virginica
148  Iris-virginica
149  Iris-virginica
             Id  SepalLengthCm  SepalWidthCm  PetalLengthCm
PetalWidthCm
count   50.00000       50.00000     50.000000      50.000000
50.00000
mean   125.50000        6.58800      2.974000       5.552000
2.02600
std     14.57738        0.63588      0.322497       0.551895
0.27465
min    101.00000        4.90000      2.200000       4.500000
1.40000
25%    113.25000        6.22500      2.800000       5.100000
1.80000
```

```
50%    125.50000         6.50000      3.000000       5.550000
2.00000
75%    137.75000         6.90000      3.175000       5.875000
2.30000
max    150.00000         7.90000      3.800000       6.900000
2.50000
```

```
d.groupby(["Species"])["SepalLengthCm"].mean()
```

```
Species
Iris-setosa        5.006
Iris-versicolor    5.936
Iris-virginica     6.588
Name: SepalLengthCm, dtype: float64
```

```
d.groupby(["Species"])["SepalLengthCm"].std()
```

```
Species
Iris-setosa        0.352490
Iris-versicolor    0.516171
Iris-virginica     0.635880
Name: SepalLengthCm, dtype: float64
```

```
d.groupby(["Species"])["SepalLengthCm"].describe()
```

```
                 count   mean       std  min    25%  50%  75%  max
Species
Iris-setosa       50.0  5.006  0.352490  4.3  4.800  5.0  5.2  5.8
Iris-versicolor   50.0  5.936  0.516171  4.9  5.600  5.9  6.3  7.0
Iris-virginica    50.0  6.588  0.635880  4.9  6.225  6.5  6.9  7.9
```

```
d.groupby(["Species"])["SepalLengthCm"].quantile(q=0.75)
```

```
Species
Iris-setosa        5.2
Iris-versicolor    6.3
Iris-virginica     6.9
Name: SepalLengthCm, dtype: float64
```

```
d.groupby(["Species"])["SepalLengthCm"].quantile(q=0.25)
```

```
Species
Iris-setosa        4.800
Iris-versicolor    5.600
Iris-virginica     6.225
Name: SepalLengthCm, dtype: float64
```