```
import pandas as pd
from sklearn import datasets
```

```
df = pd.read_csv('PR3.csv')
print(df.head(30),sep="\n")
```

```
   AGE_GROUP  INCOME
0      31-50     607
1      31-50     280
2      31-50       7
3     51- 70     559
4      18-30     521
5      31-50     122
6      31-50     304
7      18-30     202
8     51- 70     247
9      18-30     229
10     31-50     374
11     31-50     625
12    51- 70      80
13     18-30     649
14     31-50     339
15     18-30     376
16    51- 70     280
17     31-50     670
18     31-50     305
19     18-30     376
20     18-30     205
21     31-50     525
22    51- 70     720
23     18-30     277
24    51- 70      48
25     31-50     381
26     18-30     621
27    51- 70     557
28    51- 70     354
29     18-30     481
```

```
#unique values
print(df.AGE_GROUP.unique())
```

```
['31-50' '51- 70' '18-30']
```

```
#aggregate functions
print(df.groupby(df.AGE_GROUP).count())
```

```
            INCOME
AGE_GROUP
18-30           10
31-50           12
51- 70           8
```

```
print(df.groupby(df.AGE_GROUP).min())
```

```
            INCOME
AGE_GROUP
18-30          202
31-50            7
51- 70          48
```

```
print(df.groupby(df.AGE_GROUP).max())
```

```
            INCOME
AGE_GROUP
18-30          649
31-50          670
51- 70         720
```

```
print(df.groupby(df.AGE_GROUP).mean())
```

```
             INCOME
AGE_GROUP
18-30       393.700
31-50       378.250
51- 70      355.625
```

```
print(df.groupby(df.AGE_GROUP).std())
```

```
               INCOME
AGE_GROUP
18-30       168.126308
31-50       201.235784
51- 70      239.882672
```

```
print(df.groupby(df.AGE_GROUP).describe())
```

```
           INCOME
            count     mean         std    min     25%    50%    75%    max
AGE_GROUP
18-30        10.0  393.700  168.126308  202.0  241.00  376.0  511.0  649.0
31-50        12.0  378.250  201.235784    7.0  298.00  356.5  545.5  670.0
51- 70        8.0  355.625  239.882672   48.0  205.25  317.0  557.5  720.0
```

```
#iris
data = datasets.load_iris()
df = pd.DataFrame(data.data,columns=data.feature_names)
df['species'] = pd.Series(data.target)
print(df.head())
```

```
   sepal length (cm)  sepal width (cm)  petal length (cm)  petal width (cm)  \
0                5.1               3.5                1.4               0.2
1                4.9               3.0                1.4               0.2
2                4.7               3.2                1.3               0.2
3                4.6               3.1                1.5               0.2
4                5.0               3.6                1.4               0.2

   species
0        0
1        0
2        0
3        0
4        0
```

```
print(df.species.unique())
```

```
[0 1 2]
```

```
print(df.groupby(df.species))
```

```
<pandas.core.groupby.generic.DataFrameGroupBy object at 0x0000023EE39C29A0>
```

```
print(df.groupby(df.species).max())
```

```
         sepal length (cm)  sepal width (cm)  petal length (cm)  \
species
0                      5.8               4.4                1.9
1                      7.0               3.4                5.1
2                      7.9               3.8                6.9

         petal width (cm)
species
0                     0.6
1                     1.8
2                     2.5
```

```
print(df.groupby(df.species).min())
```

```
         sepal length (cm)  sepal width (cm)  petal length (cm)  \
species
0                      4.3               2.3                1.0
1                      4.9               2.0                3.0
2                      4.9               2.2                4.5

         petal width (cm)
species
0                     0.1
1                     1.0
2                     1.4
```

.

```
print(df.groupby(df.species).mean())
```

```
         sepal length (cm)  sepal width (cm)  petal length (cm)  \
species
0                    5.006             3.428              1.462
1                    5.936             2.770              4.260
2                    6.588             2.974              5.552

         petal width (cm)
species
0                   0.246
1                   1.326
2                   2.026
```

```
print(df.groupby(df.species).std())
```

```
         sepal length (cm)  sepal width (cm)  petal length (cm)  \
species
0                 0.352490          0.379064           0.173664
1                 0.516171          0.313798           0.469911
2                 0.635880          0.322497           0.551895

         petal width (cm)
species
0                0.105386
1                0.197753
2                0.274650
```

```
print(df.groupby(df.species)["sepal length (cm)"].describe())
```

```
         count   mean       std  min    25%  50%  75%  max
species
0         50.0  5.006  0.352490  4.3  4.800  5.0  5.2  5.8
1         50.0  5.936  0.516171  4.9  5.600  5.9  6.3  7.0
2         50.0  6.588  0.635880  4.9  6.225  6.5  6.9  7.9
```

```
print(df.groupby(df.species)["sepal width (cm)"].describe())
```

```
         count   mean       std  min    25%  50%    75%  max
species
0         50.0  3.428  0.379064  2.3  3.200  3.4  3.675  4.4
1         50.0  2.770  0.313798  2.0  2.525  2.8  3.000  3.4
2         50.0  2.974  0.322497  2.2  2.800  3.0  3.175  3.8
```