```
import numpy as np
import pandas as pd
df=pd.read_csv('customer5.txt',header=None)
df.columns = ['id', 'fname', 'lname', 'age', 'prof', 'loc',
'additional_column']

print(df)
```

```
         id     fname    lname  age                            prof
loc
0   4000001  Kristina    Chung   55                           Pilot
india  \
1   4000002     Paige     Chen   74                         Teacher
uk
2   4000003    Sherri   Melton   34                      Firefighter
us
3   4000004  Gretchen     Hill   66     Computer hardware engineer
china
4   4000005     Karen  Puckett   74                          Lawyer
africa
..      ...       ...      ...  ...                             ...
...
68  4000069   Melanie   Hewitt   47              Real estate agent
us
69  4000070  Marianne   Branch   53                          Judge
us
70  4000071   Natalie   Walton   24  Recreation and fitness worker
us
71  4000072  Caroline  O'Brien   44    Computer support specialist
us
72  4000073    Arlene     Case   62                        Musician
us

    additional_column
0               45000
1               45987
2               78956
3               98570
4               45680
..                ...
68             456987
69             258974
70            7896254
71             478921
72             159879

[73 rows x 7 columns]
```

1. Find Row count

```
print(df.count())
```

```
id                   73
fname                73
lname                73
age                  73
prof                 73
loc                  73
additional_column    73
dtype: int64
```

1. Remove duplicates rows and find total row count

```
df1=df.drop_duplicates()
print(df1.count())
```

```
id                   73
fname                73
lname                73
age                  73
prof                 73
loc                  73
additional_column    73
dtype: int64
```

1. Age maximum 10 fname,lname,prof,loc

```
df2=df.sort_values(by='age',ascending=False).head(10)
[['fname','lname','prof','loc','age']]
print(df2)
```

```
       fname      lname                              prof     loc  age
4      Karen    Puckett                           Lawyer  africa   74
1      Paige       Chen                          Teacher      uk   74
37      Beth     Walton                       Firefighter      uk   73
16      Neal   Lawrence      Computer support specialist   india   72
48    Harvey  Underwood           Engineering technician      uk   70
66      Luis     Hinton                 Childcare worker      us   69
67     Allan      Marsh                          Athlete      us   67
43  Patricia     Mangum                   Civil engineer      uk   67
19   Crystal     Powers           Engineering technician   india   67
61      Evan      Grant  Agricultural and food scientist      uk   66
```

1. Age minimum 5 employees fname,lname,prof,loc

```
df3=df.sort_values(by="age").head(5)
[['fname','lname','prof','loc','age']]
print(df3)
```

```
       fname      lname                              prof loc  age
51   Shirley    Merritt                          Reporter  uk   21
53    Judith     Cooper                         Economist  uk   22
70   Natalie     Walton  Recreation and fitness worker  us   24
31  Gretchen  Goldstein         Engineering technician  uk   24
34   Shelley      Weeks                          Reporter  uk   25
```

1. Each location count [count desc order]

```
df4=df.groupby(by='loc')['loc'].count().sort_values(ascending=False)
print(df4)
```

```
loc
uk           37
india        21
us           11
africa        1
australia     1
china         1
ireland       1
Name: loc, dtype: int64
```

1. Each age group count [age desc order]

```
df5=df.groupby(by='age')['age'].count().sort_values(ascending=False)
print(df5)
```

```
age
27    4
44    4
39    3
53    3
52    3
47    3
45    3
67    3
66    3
65    3
42    3
63    3
40    2
74    2
37    2
24    2
55    2
56    1
64    1
62    1
69    1
70    1
72    1
60    1
59    1
73    1
58    1
21    1
54    1
50    1
22    1
```

```
43    1
41    1
38    1
35    1
34    1
33    1
31    1
28    1
26    1
25    1
49    1
Name: age, dtype: int64
```

8.Each profession count [count desc order]

```
df6=df.groupby(by='prof')['prof'].count().sort_values(ascending=False)
print(df6)
```

```
prof
Musician                          5
Childcare worker                  5
Doctor                            4
Computer support specialist       3
Real estate agent                 3
Lawyer                            3
Engineering technician            3
Economist                         3
Writer                            3
Civil engineer                    3
Carpenter                         3
Computer hardware engineer        2
Pilot                             2
Teacher                           2
Firefighter                       2
Secretary                         2
Reporter                          2
Athlete                           2
Artist                            2
Agricultural and food scientist   2
Police officer                    2
Veterinarian                      1
Therapist                         1
Social worker                     1
Recreation and fitness worker     1
Accountant                        1
Physicist                         1
Photographer                      1
Pharmacist                        1
Judge                             1
Actor                             1
Financial analyst                 1
```

```
Environmental scientist            1
Dancer                             1
Computer software engineer         1
Human resources assistant          1
Name: prof, dtype: int64
```

1. India work

A. Row count

```
df7=df.loc[df['loc']=='india']
print(df7.count())
```

```
id                   21
fname                21
lname                21
age                  21
prof                 21
loc                  21
additional_column    21
dtype: int64
```

B. Each profession count [count desc order]

```
df8=df7.groupby(by="prof")['prof'].count()
print(df8)
```

```
prof
Accountant                     1
Artist                         1
Carpenter                      3
Childcare worker               1
Computer support specialist    1
Dancer                         1
Doctor                         1
Engineering technician         1
Environmental scientist        1
Financial analyst              1
Lawyer                         1
Musician                       3
Pharmacist                     1
Pilot                          1
Therapist                      1
Writer                         2
Name: prof, dtype: int64
```

C. Age mxm 3 employees fname,lname,age,prof

```
df9=df7.sort_values(by='age',ascending=False).head(3)
[['fname','lname','age','prof']]
print(df9)
```

```
      fname     lname  age                         prof
16     Neal  Lawrence   72  Computer support specialist
19  Crystal    Powers   67       Engineering technician
21     Eric    Steele   66                       Doctor
```

D. Age minimum 3 employees fname,lname,age,prof

```
df10=df7.sort_values(by='age').head(3)[['fname','lname','age','prof']]
print(df10)
```

```
       fname    lname  age    prof
11     Sandy   Raynor   26  Writer
25    Marian  Solomon   27  Lawyer
23  Franklin     Vick   28  Dancer
```

E. age above 40 full data

```
df11=df7.loc[df7['age']>=40]
print(df11)
```

```
         id     fname       lname  age                         prof
loc
0   4000001  Kristina       Chung   55                        Pilot
india  \
7   4000008     Hazel      Bender   63                   Carpenter
india
9   4000010   Dolores  McLaughlin   60                      Writer
india
10  4000011   Francis    McNamara   47                   Therapist
india
12  4000013    Marion        Moon   41                   Carpenter
india
13  4000014      Beth     Woodard   65                    Musician
india
14  4000015     Julia       Desai   49                    Musician
india
15  4000016    Jerome     Wallace   52                  Pharmacist
india
16  4000017      Neal    Lawrence   72  Computer support specialist
india
17  4000018      Jean     Griffin   45            Childcare worker
india
18  4000019  Kristine   Dougherty   63           Financial analyst
india
19  4000020   Crystal      Powers   67      Engineering technician
india
21  4000022      Eric      Steele   66                      Doctor
india
22  4000023    Wesley      Teague   42                   Carpenter
india
24  4000025    Claire   Gallagher   42                    Musician
```

```
india
26  4000027    Marcia       Walsh   64                    Accountant
india

    additional_column
0               45000
7              400000
9               45000
10              85000
12              25000
13              75000
14              26000
15              36000
16              95000
17              65000
18              25000
19              55555
21               3685
22              58789
24              75000
26              95000
```

F. age range 30 to 40 [profession count]

```
df12=df7.loc[(df7['age']>=30)&(df7['age']<=40)].groupby(by='age')
['age'].count()
print(df12)
```

```
age
39    2
Name: age, dtype: int64
```

1. us work

```
df13=df.loc[df['loc']=='us']
print(df13)
```

```
         id     fname    lname  age                          prof loc

2   4000003    Sherri   Melton   34                    Firefighter  us
\
63  4000064    Calvin     Diaz   65                        Athlete  us

64  4000065    Eugene   Graham   52                 Police officer  us

65  4000066    Vickie  Watkins   55    Computer support specialist  us

66  4000067      Luis   Hinton   69                Childcare worker  us

67  4000068     Allan    Marsh   67                        Athlete  us
```

```
68  4000069   Melanie   Hewitt   47              Real estate agent  us

69  4000070  Marianne   Branch   53                          Judge  us

70  4000071   Natalie   Walton   24  Recreation and fitness worker  us

71  4000072  Caroline  O'Brien   44    Computer support specialist  us

72  4000073    Arlene     Case   62                       Musician  us


    additional_column
2               78956
63             960000
64              75205
65            1700000
66             456985
67             478962
68             456987
69             258974
70            7896254
71             478921
72             159879
```

A. Row count

```
print(df13.count())
```

```
id                   11
fname                11
lname                11
age                  11
prof                 11
loc                  11
additional_column    11
dtype: int64
```

B. Each age group count

```
df14=df13.groupby(by='age')['age'].count()
print(df14)
```

```
age
24    1
34    1
44    1
47    1
52    1
53    1
55    1
62    1
```

```
65    1
67    1
69    1
Name: age, dtype: int64
```

C. Each profession count [count desc]

```
df15=df13.groupby(by='prof')
['prof'].count().sort_values(ascending=False)
print(df15)
```

```
prof
Athlete                          2
Computer support specialist      2
Childcare worker                 1
Firefighter                      1
Judge                            1
Musician                         1
Police officer                   1
Real estate agent                1
Recreation and fitness worker    1
Name: prof, dtype: int64
```

D. Civil engineer dept and age above 30

```
df16=df13.loc[(df13['prof']=='Civil engineer') &
(df13['age']>=30)].count()[['fname','lname','prof','age']]
print(df16)
```

```
fname    0
lname    0
prof     0
age      0
dtype: int64
```