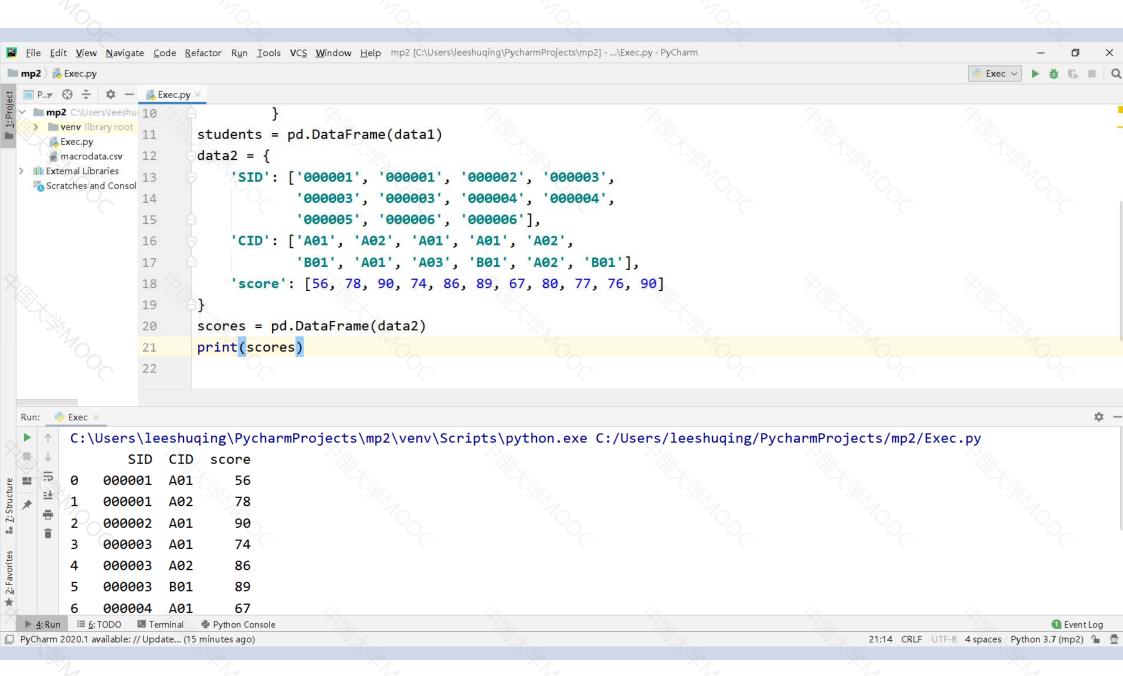
## Python大数据分析

七、数据的连接

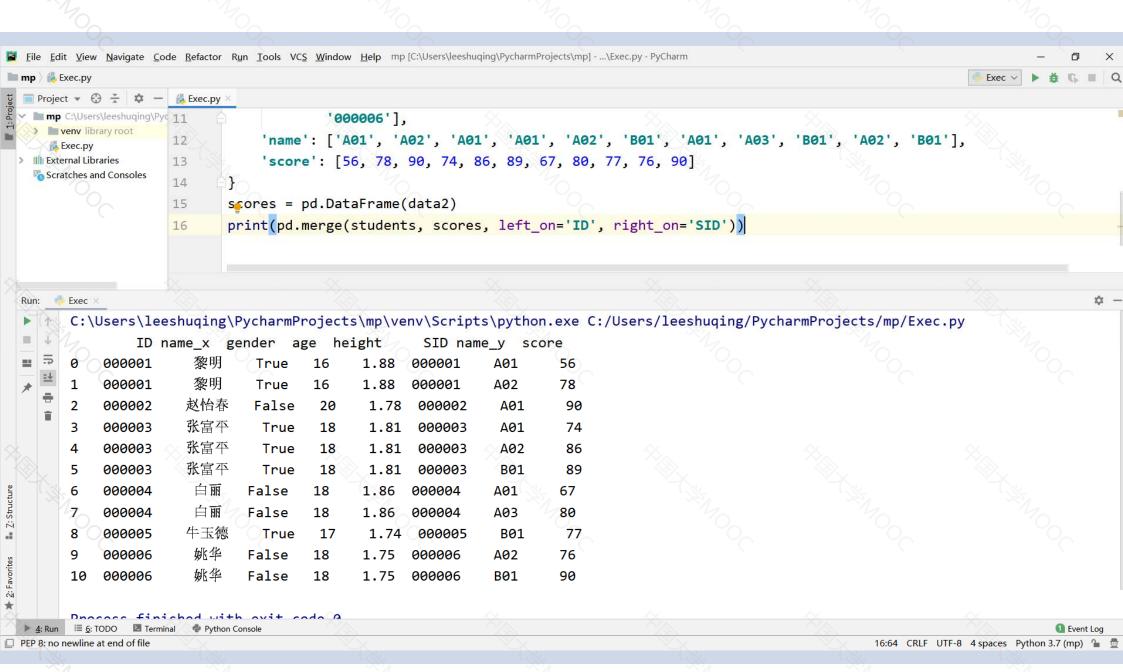
## 成绩数据

SID	CID	Score
000001	A01	56
000001	A02	78
000002	A01	90
000003	A01	74
000003	A02	86
000003	B01	89
000004	A01	67
000004	A03	80
000005	B01	77
000006	A02	76
000006	B01	90



```
🔟 File Edit View Navigate Code Refactor Run Iools VCS Window Help mp2 [C:\Users\leeshuqing\PycharmProjects\mp2] - ...\Exec.py - PyCharm
                                                                                                                                    Exec
    P... 🗘 😤 Exec.py
   mp2 C:\Users\| 13
                           'SID': ['000001', '000001', '000002', '000003', '000003', '000003', '000004', '000004', '000005', '000006',
   > wenv library
                                   '000006'],
     Exec.py
  III External Librarie 15
                           'CID': ['A01', 'A02', 'A01', 'A01', 'A02', 'B01', 'A01', 'A03', 'B01', 'A02', 'B01'],
   Scratches and C
                           'score': [56, 78, 90, 74, 86, 89, 67, 80, 77, 76, 90]
              16
              17
                      scores = pd.DataFrame(data2)
              18
                      print(pd.merge(students, scores, left_on='ID', right_on='SID'))
              19
              20
        Exec
        C:\Users\leeshuqing\PycharmProjects\mp2\venv\Scripts\python.exe C:/Users/leeshuqing/PycharmProjects/mp2/Exec.py
                 ID name gender age height
                                                     SID CID score
                      黎明
            000001
                               True
                                      16
                                             1.88
                                                    000001 A01
                                                                      56
                      黎明
            000001
                               True
                                       16
                                             1.88
                                                    000001 A02
                                                                      78
                     赵怡春
                              False
            000002
                                       20
                                              1.78
                                                    000002
                                                             A01
                                                                      90
                     张富平
            000003
                                True
                                       18
                                              1.81
                                                     000003
                                                             A01
                                                                      74
                     张富平
            000003
                                True
                                       18
                                              1.81
                                                     000003
                                                             A02
                                                                      86
                     张富平
                                              1.81
            000003
                                True
                                       18
                                                     000003
                                                             B01
                                                                      89
                       白丽
                                             1.86
                                                    000004
        6
            000004
                              False
                                       18
                                                             A01
                                                                      67
                       白丽
                             False
            000004
                                       18
                                             1.86
                                                    000004
                                                             A03
                                                                      80
                     牛玉德
                                True
                                                    000005
                                                             B01
            000005
                                       17
                                              1.74
                                                                      77
                      姚华
                                             1.75
             000006
                              False
                                                    000006
                                                             A02
                                                                      76
                                       18
                      姚华
                              False
                                                             B01
        10
            000006
                                       18
                                             1.75
                                                    000006
                                                                      90
                 ■ Terminal
                          Python Console
                                                                                                                                             C Event Log
                                                                                                                   19:64 CRLF UTF-8 8 spaces* Python 3.7 (mp2) 🦫 👨
```

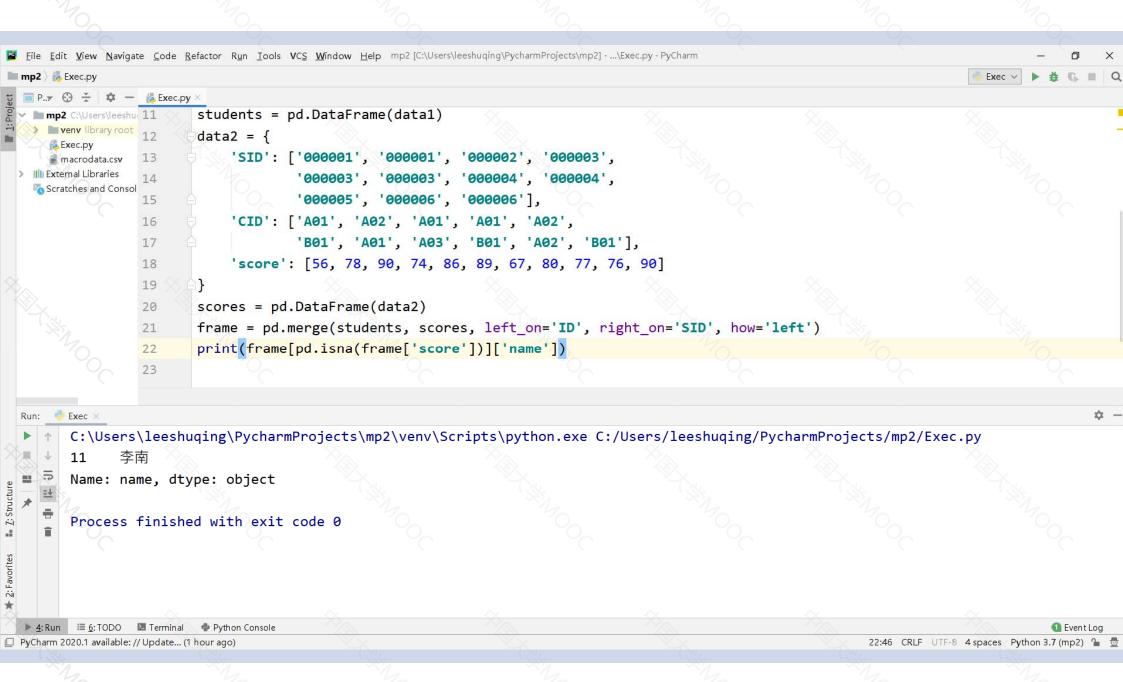
```
📱 Eile Edit View Navigate Code Refactor Run Iools VCS Window Help mp2 [C:\Users\leeshuqing\PycharmProjects\mp2] - ...\Exec.py - PyCharm
   P.. ▼ 🕀 😤 🗢 🎏 Exec.py ×
                         students = pd.DataFrame(datal)
   mp2 C:\Users\leeshu 11
   > env library root 12
                         data2 = {
     Exec.py
                              'ID': ['000001', '000001', '000002', '000003', '000003',
                  13
     macrodata.csv
  III External Libraries
                  14
                                      '000003', '000004', '000004', '000005', '000006',
   Scratches and Consol
                  15
                                      '000006'],
                              'CID': ['A01', 'A02', 'A01', 'A01', 'A02',
                  16
                                       'B01', 'A01', 'A03', 'B01', 'A02', 'B01'],
                  17
                              'score': [56, 78, 90, 74, 86, 89, 67, 80, 77, 76, 90]
                  18
                  19
                         scores = pd.DataFrame(data2)
                  20
                         print(pd.merge(students, scores))
                  21
                  22
        Exec
        C:\Users\leeshuqing\PycharmProjects\mp2\venv\Scripts\python.exe C:/Users/leeshuqing/PycharmProjects/mp2/Exec.py
                 ID name gender age height CID score
                      黎明
            000001
                               True
                                       16
                                             1.88 A01
                                                             56
                      黎明
            000001
                               True
                                       16
                                             1.88
                                                    A02
                                                             78
            000002 赵怡春
                               False
                                        20
                                              1.78 A01
                                                              90
            000003
                     张富平
                                              1.81 A01
                                True
                                       18
                                                              74
                     张富平
            000003
                                              1.81 A02
                                                              86
                                True
                                       18
            000003 张富平
                                True
                                       18
                                              1.81 B01
                                                              89
                       白丽
            000004
                              False
                                       18
                                             1.86
                                                    A01
                                                             67
                       白丽
            000004
                              False
                                       18
                                             1.86
                                                    A03
                                                             80
                     十工油
             DODDODE
                                              1 7/
                                                     DA1
                                                              77
                 Terminal
                         Python Console
                                                                                                                                              C Event Log
                                                                                                                     21:34 CRLF UTF-8 4 spaces Python 3.7 (mp2) 🧣 💆
```



```
🔟 File Edit View Navigate Code Refactor Run Iools VCS Window Help mp2 [C:\Users\leeshuqing\PycharmProjects\mp2] - ...\Exec.py - PyCharm
                                                                                                                                    Exec
    P... 🗘 😤 Exec.py
   mp2 C:\Users\| 13
                           'SID': ['000001', '000001', '000002', '000003', '000003', '000003', '000004', '000004', '000005', '000006',
   > wenv library
                                   '000006'],
     Exec.py
  III External Librarie 15
                           'CID': ['A01', 'A02', 'A01', 'A01', 'A02', 'B01', 'A01', 'A03', 'B01', 'A02', 'B01'],
   Scratches and C
                           'score': [56, 78, 90, 74, 86, 89, 67, 80, 77, 76, 90]
              16
              17
                      scores = pd.DataFrame(data2)
              18
                      print(pd.merge(students, scores, left_on='ID', right_on='SID'))
              19
              20
        Exec
        C:\Users\leeshuqing\PycharmProjects\mp2\venv\Scripts\python.exe C:/Users/leeshuqing/PycharmProjects/mp2/Exec.py
                 ID name gender age height
                                                     SID CID score
                      黎明
            000001
                               True
                                      16
                                             1.88
                                                    000001 A01
                                                                      56
                      黎明
            000001
                               True
                                       16
                                             1.88
                                                    000001 A02
                                                                      78
                     赵怡春
                              False
            000002
                                       20
                                              1.78
                                                    000002
                                                             A01
                                                                      90
                     张富平
            000003
                                True
                                       18
                                              1.81
                                                     000003
                                                             A01
                                                                      74
                     张富平
            000003
                                True
                                       18
                                              1.81
                                                     000003
                                                             A02
                                                                      86
                     张富平
                                              1.81
            000003
                                True
                                       18
                                                     000003
                                                             B01
                                                                      89
                       白丽
                                             1.86
                                                    000004
        6
            000004
                              False
                                       18
                                                             A01
                                                                      67
                       白丽
                             False
            000004
                                       18
                                             1.86
                                                    000004
                                                             A03
                                                                      80
                     牛玉德
                                True
                                                    000005
                                                             B01
            000005
                                       17
                                              1.74
                                                                      77
                      姚华
                                             1.75
             000006
                              False
                                                    000006
                                                             A02
                                                                      76
                                       18
                      姚华
                              False
                                                             B01
        10
            000006
                                       18
                                             1.75
                                                    000006
                                                                      90
                 ■ Terminal
                          Python Console
                                                                                                                                             C Event Log
                                                                                                                   19:64 CRLF UTF-8 8 spaces* Python 3.7 (mp2) 🦫 👨
```

```
🔟 File Edit View Navigate Code Refactor Run Iools VCS Window Help mp2 [C:\Users\leeshuqing\PycharmProjects\mp2] - ...\Exec.py - PyCharm
                                                                                                                                     Exec
             Exec.py
    mp2 C:\Users\l
                                    '000006'],
     venv library
                           'CID': ['A01', 'A02', 'A01', 'A01', 'A02', 'B01', 'A01', 'A03', 'B01', 'A02', 'B01'],
      Exec.py
               15
   IIII External Librarie 16
                           'score': [56, 78, 90, 74, 86, 89, 67, 80, 77, 76, 90]
    Scratches and C
               17
                      scores = pd.DataFrame(data2)
               18
                       print(pd.merge(students, scores, left on='ID', right on='SID', how='left'))
               19
               20
       Exec >
         C:\Users\leeshuqing\PycharmProjects\mp2\venv\Scripts\python.exe C:/Users/leeshuqing/PycharmProjects/mp2/Exec.py
                 ID name gender age height
                                                      SID CID score
                       黎明
         0
             000001
                               True
                                       16
                                              1.88
                                                     000001
                                                              A01
                                                                     56.0
                       黎明
             000001
                               True
                                       16
                                              1.88
                                                     000001
                                                              A02
                                                                     78.0
                      赵怡春
             000002
                               False
                                        20
                                               1.78
                                                      000002
                                                               A01
                                                                     90.0
                      张富平
             000003
                                True
                                        18
                                               1.81
                                                      000003
                                                               A01
                                                                     74.0
             000003
                     张富平
                                True
                                        18
                                               1.81
                                                      000003
                                                               A02
                                                                     86.0
                      张富平
             000003
                                               1.81
                                True
                                        18
                                                      000003
                                                               B01
                                                                     89.0
                       白丽
             000004
                              False
                                       18
                                              1.86
                                                     000004
                                                              A01
                                                                     67.0
                       白丽
                              False
             000004
                                       18
                                              1.86
                                                     000004
                                                              A03
                                                                     80.0
                      牛玉德
             000005
                                True
                                        17
                                               1.74
                                                      000005
                                                               B01
                                                                     77.0
                       姚华
                              False
                                       18
                                              1.75
                                                     000006
             000006
                                                              A02
                                                                     76.0
                       姚华
                              False
                                              1.75
                                                     000006
             000006
                                       18
                                                              B01
                                                                     90.0
         10
                       李南
                                              1.76
             000007
                                       16
                               True
                                                        NaN
                                                             NaN
                                                                      NaN
                 ■ Terminal
                          Python Console
                                                                                                                                              Event Log
19:76 CRLF UTF-8 8 spaces* Python 3.7 (mp2) 🦫 👨
```

```
🖺 Eile Edit View Navigate Code Refactor Run Jools VCS Window Help mp2 [C:\Users\leeshuqing\PycharmProjects\mp2] - ...\Exec.py - PyCharm
mp2 \ Exec.py
    P..y ⊕ 😤 🌣 — 🐔 Exec.py ×
                          students = pd.DataFrame(data1)
    mp2 C:\Users\leeshu 11
   > env library root 12
                           data2 = {
      Exec.py
                                'SID': ['000001', '000001', '000002', '000003',
      macrodata.csv
                  13
  > IIII External Libraries
                                         '000003', '000003', '000004', '000004',
    Scratches and Consol
                                         '000005', '000006', '000006'],
                   15
                                'CID': ['A01', 'A02', 'A01', 'A01', 'A02',
                   16
                                         'B01', 'A01', 'A03', 'B01', 'A02', 'B01'],
                   17
                                'score': [56, 78, 90, 74, 86, 89, 67, 80, 77, 76, 90]
                   18
                   19
                          scores = pd.DataFrame(data2)
                   20
                          frame = pd.merge(students, scores, left_on='ID', right_on='SID', how='left')
                   21
                           print(frame[pd.isna(frame['score']) == True]['name'])
                   22
         C:\Users\leeshuqing\PycharmProjects\mp2\venv\Scripts\python.exe C:/Users/leeshuqing/PycharmProjects/mp2/Exec.py
                李南
         11
         Name: name, dtype: object
         Process finished with exit code 0
         Python Console
PyCharm 2020.1 available: // Update... (59 minutes ago)
                                                                                                                         22:54 CRLF UTF-8 4 spaces Python 3.7 (mp2) 🦫 💆
```



```
🖺 File Edit View Navigate Code Refactor Run Jools VCS Window Help mp2 [C:\Users\leeshuqing\PycharmProjects\mp2] - ...\Exec.py - PyCharm
                                                                                                                                        Exec v
             Exec.py
    mp2 C:\Users\| 15
                                     '000006'].
    > wenv library
                            'course': ['A01', 'A02', 'A01', 'A01', 'A02', 'B01', 'A01', 'A03', 'B01', 'A02', 'B01'],
      Exec.py
   III External Librarie 17
                            'score': [56, 78, 90, 74, 86, 89, 67, 80, 77, 76, 90]
    Scratches and C
               18
                       scores = pd.DataFrame(data2)
               19
                       scores.index = [0, 0, 1, 2, 2, 2, 3, 3, 4, 5, 5]
               20
                       print(students.join(scores))
               21
               22
        Exec ×
  Run:
         C:\Users\leeshuqing\PycharmProjects\mp2\venv\Scripts\python.exe C:/Users/leeshuqing/PycharmProjects/mp2/Exec.py
                 ID name gender age height
                                                      SID course score
                      黎明
            000001
                                                     000001
                                                                       56.0
                               True
                                       16
                                              1.88
                                                                A01
                      黎明
                                              1.88
                                                                A02
            000001
                               True
                                       16
                                                     000001
                                                                       78.0
                     赵怡春
            000002
                               False
                                               1.78
                                                     000002
                                                                        90.0
                                        20
                                                                 A01
                     张富平
                                              1.81
            000003
                                True
                                        18
                                                     000003
                                                                 A01
                                                                        74.0
                     张富平
                                              1.81 000003
            000003
                                True
                                        18
                                                                 A02
                                                                        86.0
                     张富平
            000003
                                True
                                        18
                                               1.81
                                                     000003
                                                                 B01
                                                                        89.0
                      白丽
            000004
                              False
                                       18
                                              1.86
                                                     000004
                                                                A01
                                                                       67.0
                      白丽
                              False
                                                    000004
            000004
                                       18
                                              1.86
                                                                A03
                                                                       80.0
                     牛玉德
                                True
            000005
                                        17
                                               1.74
                                                     000005
                                                                 B<sub>0</sub>1
                                                                        77.0
                      姚华
                              False
            000006
                                       18
                                              1.75
                                                     000006
                                                                A02
                                                                       76.0
                      姚华
            000006
                              False
                                       18
                                              1.75
                                                     000006
                                                                B01
                                                                       90.0
                      李南
            000007
                               True
                                       16
                                              1.76
                                                        NaN
                                                                NaN
                                                                        NaN

    ■ Terminal

                           Python Console
                                                                                                                                                 Event Log
21:29 CRLF UTF-8 8 spaces* Python 3.7 (mp2) 🦫 👨
```

```
🖺 Eile Edit View Navigate Code Refactor Run Jools VCS Window Help mp2 [C:\Users\leeshuqing\PycharmProjects\mp2] - ...\Exec.py - PyCharm
                                                                                                                                         Exec
mp2 Exec.py
    'CID': ['A01', 'A02', 'A01', 'A01', 'A02',
    mp2 C:\Users\leeshu, 16
   > env library root 17
                                        'B01', 'A01', 'A03', 'B01', 'A02', 'B01'],
      Exec.py
                               'score': [56, 78, 90, 74, 86, 89, 67, 80, 77, 76, 90]
                   18
      macrodata.csv
  > IIII External Libraries
                   19
    Scratches and Consol
                          scores = pd.DataFrame(data2)
                   20
                          scores.index = [0, 0, 1, 2, 2, 2, 3, 3, 4, 5, 5]
                   21
                          print(scores.join(students))
                   22
                   23
        Exec
         C:\Users\leeshuqing\PycharmProjects\mp2\venv\Scripts\python.exe C:/Users/leeshuqing/PycharmProjects/mp2/Exec.py
                     CID score
                                        ID name gender age height
                SID
            000001
                     A01
                                   000001
                                              黎明
                                                      True
                                                              16
                                                                     1.88
                                              黎明
                                                                     1.88
            000001
                     A02
                                   000001
                                                      True
                                                              16
                                            赵怡春
                     A01
                                                      False
                                                                      1.78
            000002
                                   000002
                                                               20
                                            张富平
            000003
                     A01
                                   000003
                                                       True
                                                               18
                                                                      1.81
                                            张富平
            000003
                     A02
                                   000003
                                                       True
                                                               18
                                                                      1.81
                                            张富平
                                   000003
            000003
                      B01
                                                       True
                                                               18
                                                                      1.81
                                              白丽
                                                                     1.86
            000004
                      A01
                                   000004
                                                     False
                                                              18
                                              白丽
                                                     False
            000004
                      A03
                                   000004
                                                              18
                                                                     1.86
            000005
                      B01
                                            牛玉德
                                                                      1.74
                                   000005
                                                       True
                                                               17
                                              姚华
            000006
                      A02
                                                     False
                                                                     1.75
                                   000006
                                                              18
                                              姚华
                                                     False
                      B01
                                                                     1.75
            000006
                                   000006
                                                              18

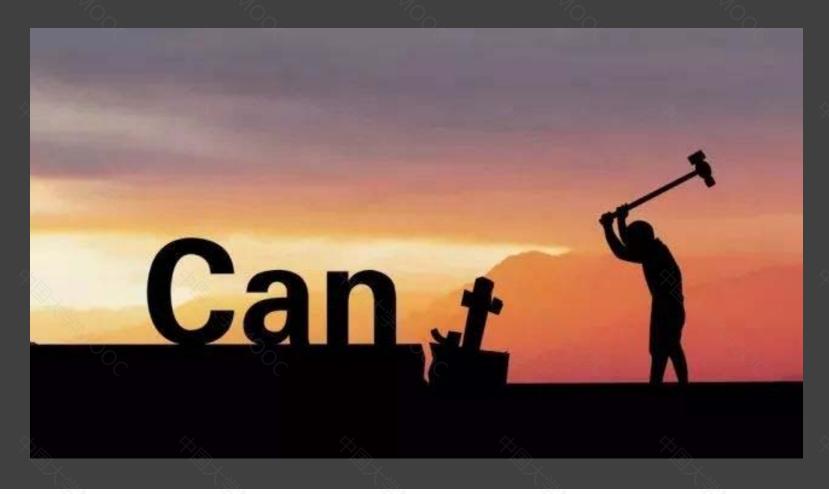
<u>6</u>: TODO

                  ■ Terminal
                           Python Console
                                                                                                                                                  Event Log
PyCharm 2020.1 available: // Update... (today 7:20)
                                                                                                                         22:29 CRLF UTF-8 4 spaces Python 3.7 (mp2) 🦫 💆
```

```
🖺 Eile Edit View Navigate Code Refactor Run Jools VCS Window Help mp2 [C:\Users\leeshuqing\PycharmProjects\mp2] - ...\Exec.py - PyCharm
mp2 Exec.py
  📺 P..▼ 🚱 😤 💠 — 🐔 Exec.py ×
   mp2 C:\Users\leeshu 13
                                'SID': ['000001', '000001', '000002', '000003', '000003', '000003', '000004', '000004', '000005', '000006',
   venv library root
                                         '000006'],
      Exec.py
      macrodata.csv
                                'CID': ['A01', 'A02', 'A01', 'A01', 'A02', 'B01', 'A01', 'A03', 'B01', 'A02', 'B01'],
  > IIII External Libraries
                                'score': [56, 78, 90, 74, 86, 89, 67, 80, 77, 76, 90]
    Scratches and Consol 16
                   17
                           scores = pd.DataFrame(data2)
                   18
                           frame = pd.merge(students, scores, left on='ID', right on='SID')
                   19
                           print(frame[['score']].groupby(frame['name']).mean())
                   20
         C:\Users\leeshuqing\PycharmProjects\mp2\venv\Scripts\python.exe C:/Users/leeshuqing/PycharmProjects/mp2/Exec.py
                score
         name
         姚华
                   83.0
         张富平
                    83.0
         牛玉德
                    77.0
         白丽
                   73.5
         赵怡春
                    90.0
         黎明
                   67.0
         Process finished with exit code 0
  ▶ 4: Run ≡ 6: TODO 🛂 Terminal 🕏 Python Console
                                                                                                                                                     C Event Log
PEP 8: no newline at end of file
                                                                                                                           20:54 CRLF UTF-8 4 spaces Python 3.7 (mp2) 🦫 💆
```

```
🖺 Eile Edit View Navigate Code Refactor Run Jools VCS Window Help mp2 [C:\Users\leeshuqing\PycharmProjects\mp2] - ...\Exec.py - PyCharm
mp2 Exec.py
  📺 P..▼ 🚱 😤 💠 — 🐔 Exec.py ×
   mp2 C:\Users\leeshu 13
                                'SID': ['000001', '000001', '000002', '000003', '000003', '000003', '000004', '000004', '000005', '000006',
    > wenv library root
                                         '000006'],
      Exec.py
      macrodata.csv
                                'CID': ['A01', 'A02', 'A01', 'A01', 'A02', 'B01', 'A01', 'A03', 'B01', 'A02', 'B01'],
  > IIII External Libraries
                                'score': [56, 78, 90, 74, 86, 89, 67, 80, 77, 76, 90]
    Scratches and Consol 16
                   17
                           scores = pd.DataFrame(data2)
                   18
                           frame = pd.merge(students, scores, left on='ID', right on='SID')
                   19
                           print(frame.groupby(frame['name'])[['score']].mean())
                   20
         C:\Users\leeshuqing\PycharmProjects\mp2\venv\Scripts\python.exe C:/Users/leeshuqing/PycharmProjects/mp2/Exec.py
                score
         name
         姚华
                   83.0
         张富平
                    83.0
         牛玉德
                    77.0
         白丽
                   73.5
         赵怡春
                    90.0
         黎明
                   67.0
         Process finished with exit code 0
  ▶ 4: Run ≡ 6: TODO 🛂 Terminal 🕏 Python Console
                                                                                                                                                     C Event Log
PEP 8: no newline at end of file
                                                                                                                           20:54 CRLF UTF-8 4 spaces Python 3.7 (mp2) 🦫 💆
```

## 只有多练习:)



## 一次不学多,下次再学