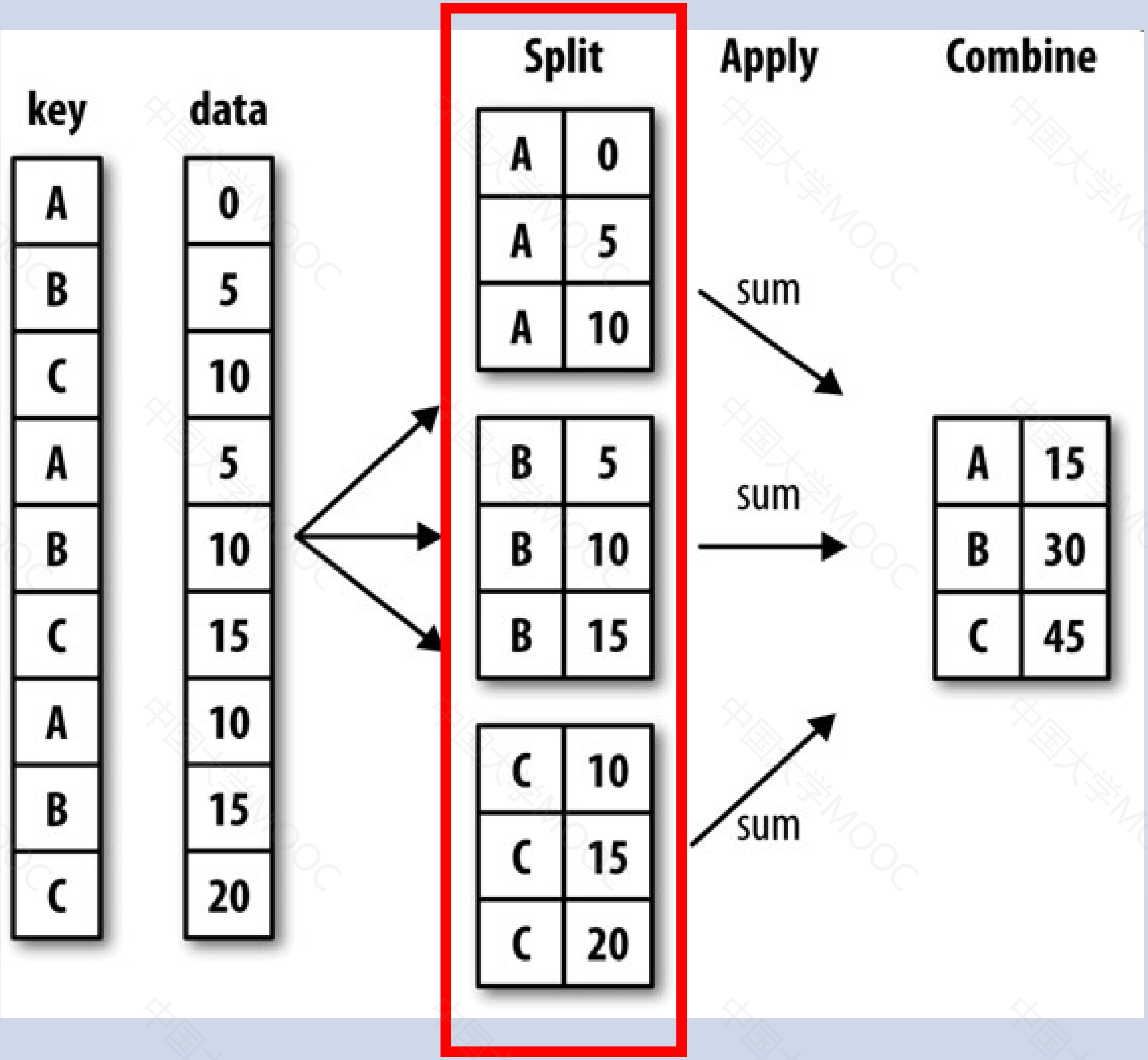
Python大数据分析

六、数据的分组和聚合运算



key	data
A	0
B	5
C	10
A	5
B	10
C	15
A	10
B	15
C	20

Split

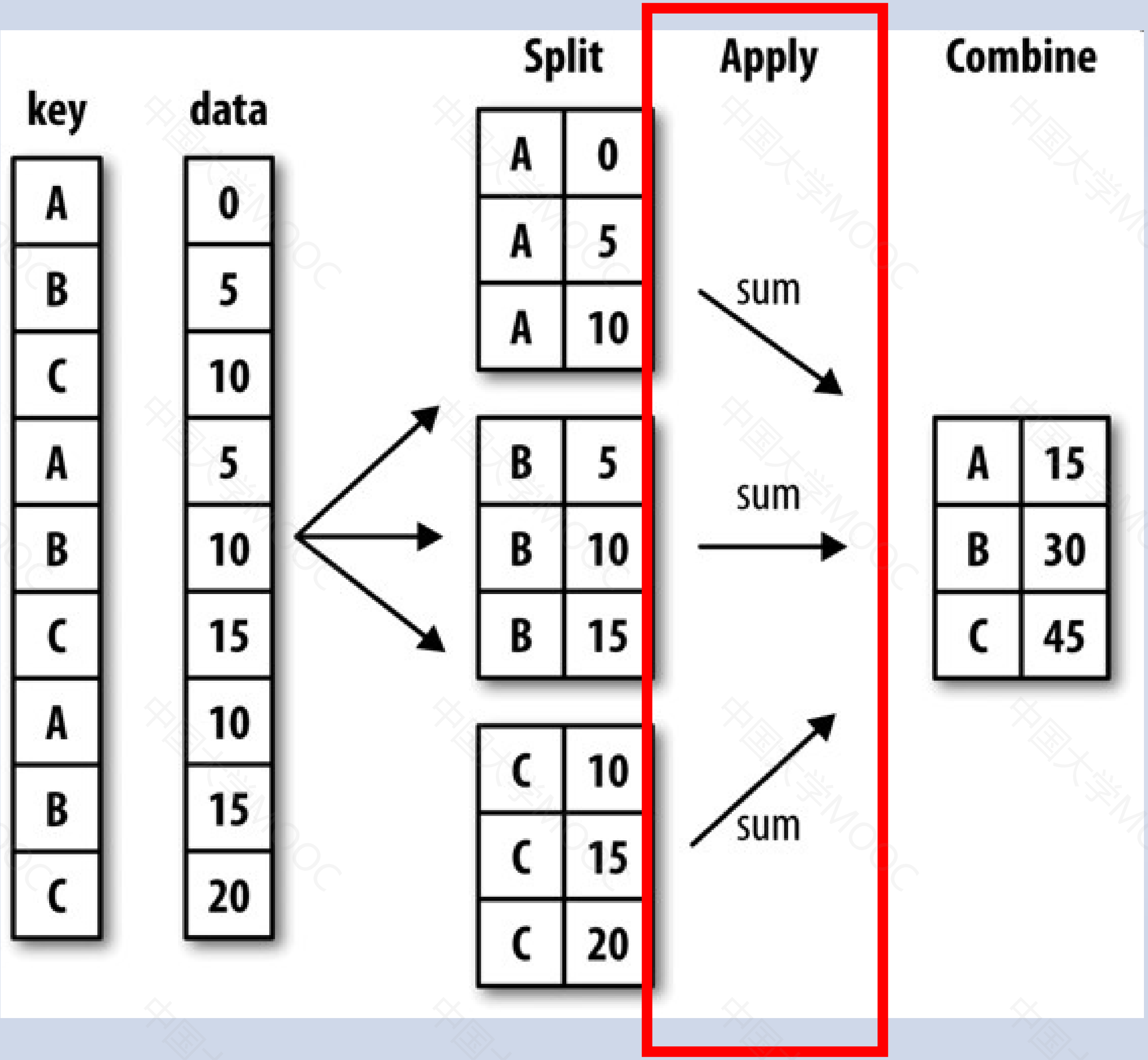
A	0
A	5
A	10

B	5
B	10
B	15

C	10
C	15
C	20

Combine

A	15
B	30
C	45



key	data
A	0
B	5
C	10
A	5
B	10
C	15
A	10
B	15
C	20

Split

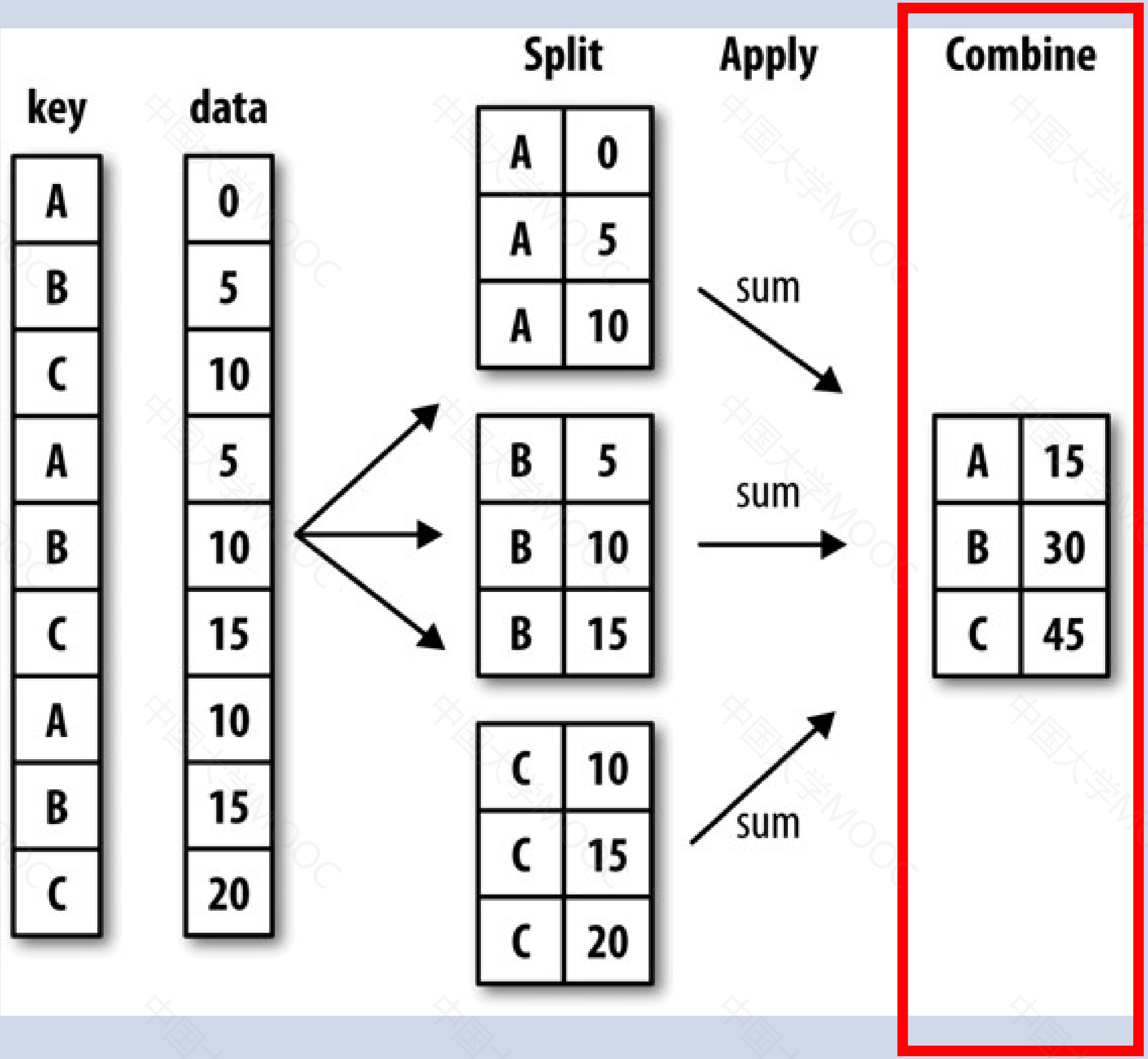
A	0
A	5
A	10

B	5
B	10
B	15

C	10
C	15
C	20

Combine

A	15
B	30
C	45



key	data
A	0
B	5
C	10
A	5
B	10
C	15
A	10
B	15
C	20

Split

A	0
A	5
A	10

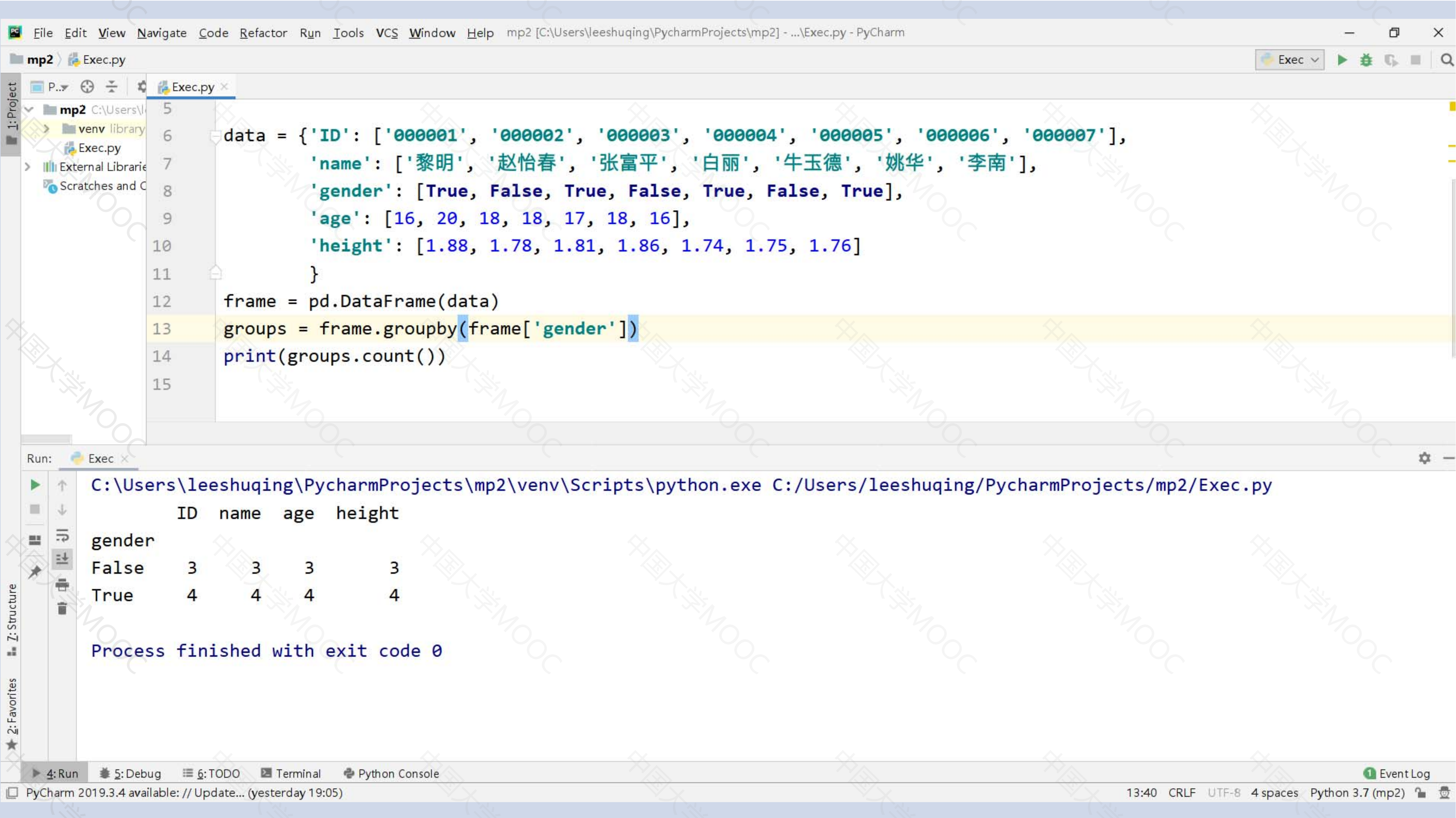
B	5
B	10
B	15

C	10
C	15
C	20

Apply

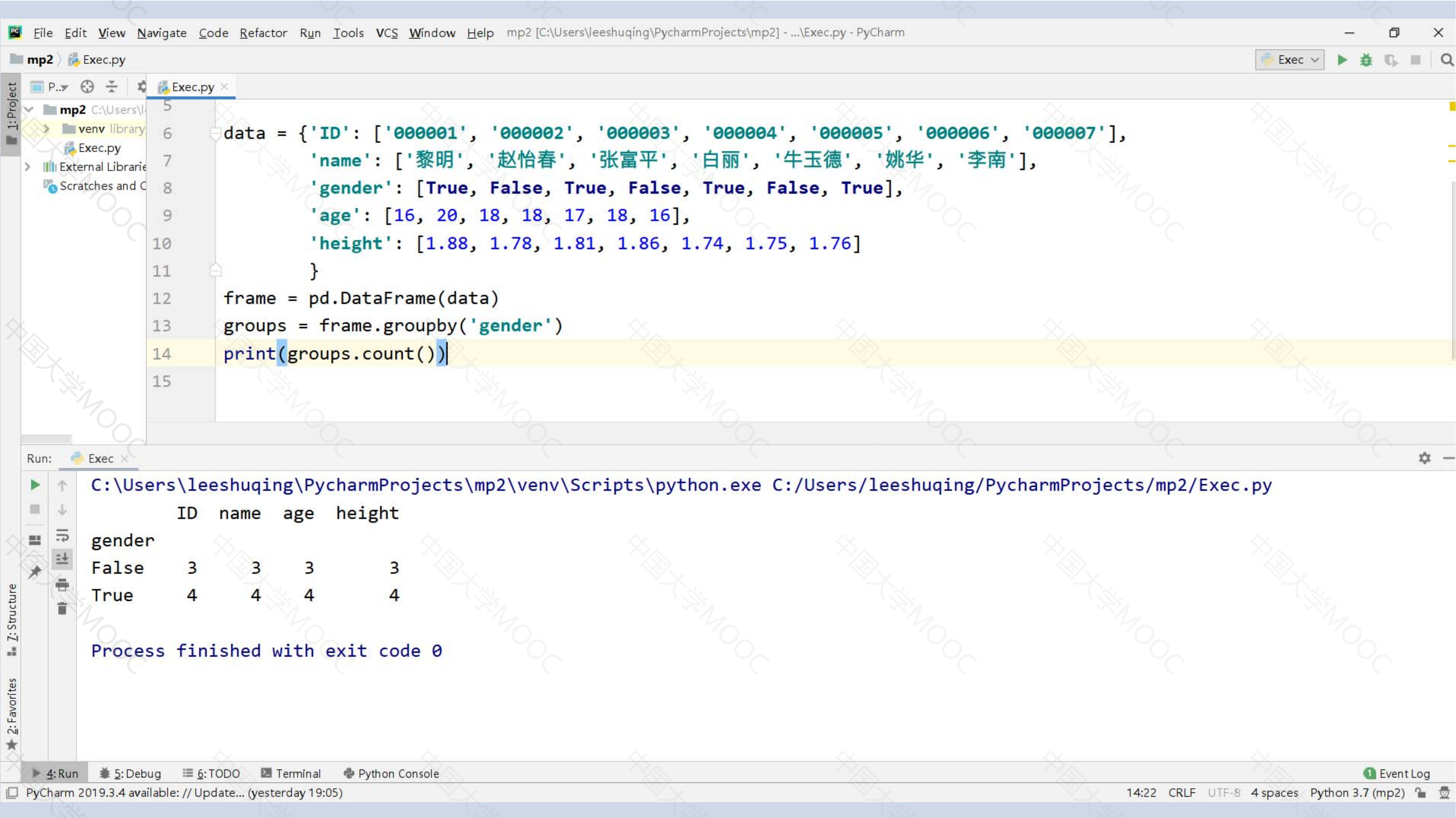
Combine

A	15
B	30
C	45


File Edit View Navigate Code Refactor Run Tools VCS Window Help
mp2 [C:\Users\leeshuqing\PycharmProjects\mp2] - ...\Exec.py - PyCharm
mp2 〉 Exec.py
Exec
1: Project
P..
Exec.py
mp2 C:\Users\l
venv library
Exec.py
External Librarie
Scratches and C
Run:
Exec
Z: Structure
2: Favorites
4: Run
5: Debug
6: TODO
Terminal
Python Console
Event Log
PyCharm 2019.3.4 available: // Update... (yesterday 19:05)
13:40 CRLF UTF-8 4 spaces Python 3.7 (mp2)

```python

data = {'ID': ['000001', '000002', '000003', '000004', '000005', '000006', '000007'],
        'name': ['黎明', '赵怡春', '张富平', '白丽', '牛玉德', '姚华', '李南'],
        'gender': [True, False, True, False, True, False, True],
        'age': [16, 20, 18, 18, 17, 18, 16],
        'height': [1.88, 1.78, 1.81, 1.86, 1.74, 1.75, 1.76]
        }
frame = pd.DataFrame(data)
groups = frame.groupby(frame['gender'])
print(groups.count())

```

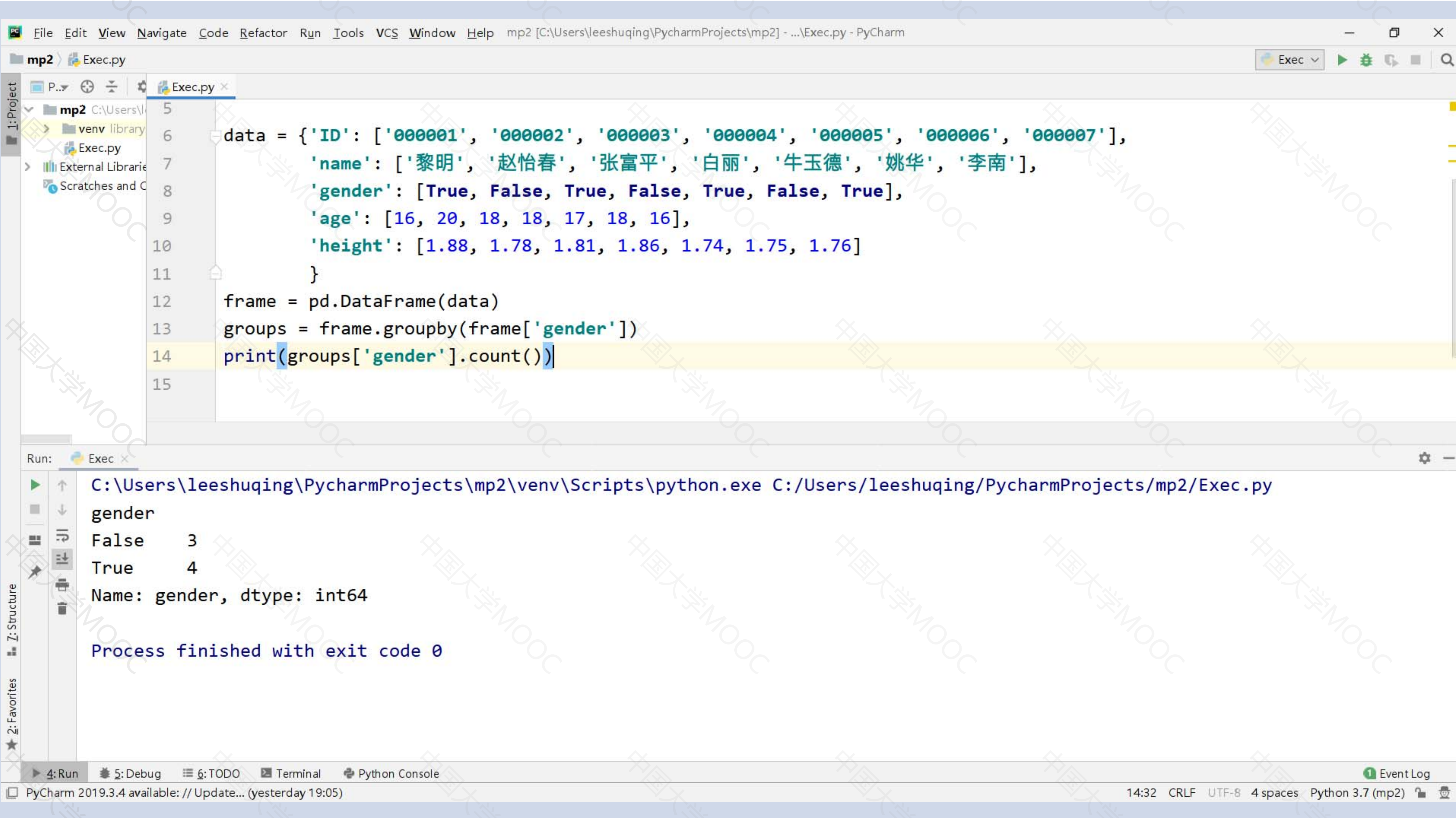
```
C:\Users\leeshuqing\PycharmProjects\mp2\venv\Scripts\python.exe C:/Users/leeshuqing/PycharmProjects/mp2/Exec.py
        ID  name  age  height
gender
False    3     3    3       3
True     4     4    4       4

Process finished with exit code 0
```



File Edit View Navigate Code Refactor Run Tools VCS Window Help
mp2 [C:\Users\leeshuqing\PycharmProjects\mp2] - ...\Exec.py - PyCharm
mp2
Exec.py
Exec
1: Project
mp2 C:\Users\l
venv library
Exec.py
External Librarie
Scratches and C
Run:
Exec
7: Structure
2: Favorites
4: Run
5: Debug
6: TODO
Terminal
Python Console
Event Log
PyCharm 2019.3.4 available: // Update... (yesterday 19:05)
14:22 CRLF UTF-8 4 spaces Python 3.7 (mp2)

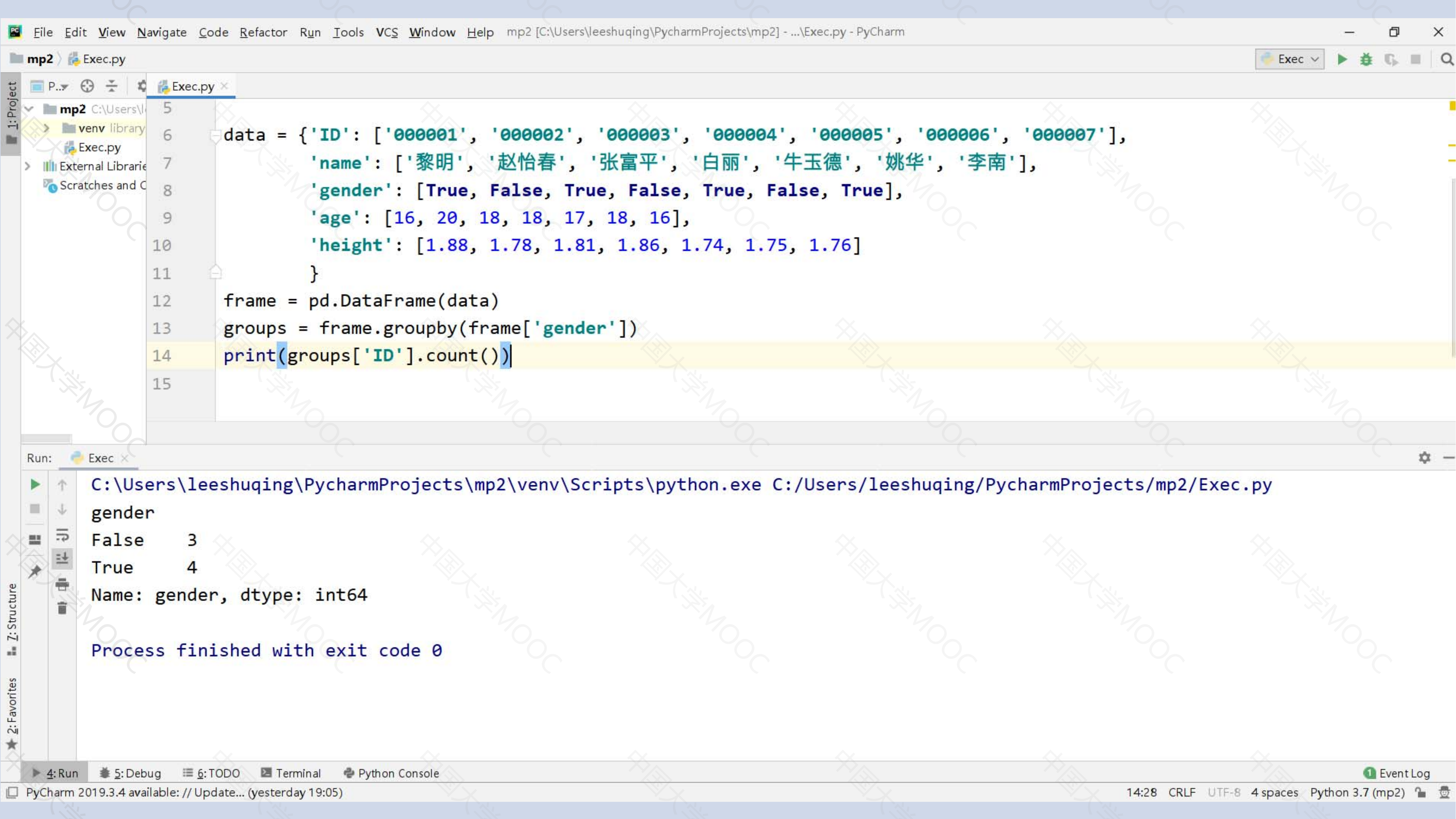
```python
data = {'ID': ['000001', '000002', '000003', '000004', '000005', '000006', '000007'],
        'name': ['黎明', '赵怡春', '张富平', '白丽', '牛玉德', '姚华', '李南'],
        'gender': [True, False, True, False, True, False, True],
        'age': [16, 20, 18, 18, 17, 18, 16],
        'height': [1.88, 1.78, 1.81, 1.86, 1.74, 1.75, 1.76]
        }
frame = pd.DataFrame(data)
groups = frame.groupby('gender')
print(groups.count())
```

```
C:\Users\leeshuqing\PycharmProjects\mp2\venv\Scripts\python.exe C:/Users/leeshuqing/PycharmProjects/mp2/Exec.py
        ID  name  age  height
gender
False    3     3    3       3
True     4     4    4       4

Process finished with exit code 0
```


File Edit View Navigate Code Refactor Run Tools VCS Window Help mp2 [C:\Users\leeshuqing\PycharmProjects\mp2] - ...\Exec.py - PyCharm
mp2 Exec.py
Exec
Exec.py
1: Project
mp2 C:\Users\l
venv library
Exec.py
External Librarie
Scratches and C
Run: Exec
7: Structure
2: Favorites
4: Run
5: Debug
6: TODO
Terminal
Python Console
Event Log
PyCharm 2019.3.4 available: // Update... (yesterday 19:05)
14:32 CRLF UTF-8 4 spaces Python 3.7 (mp2)

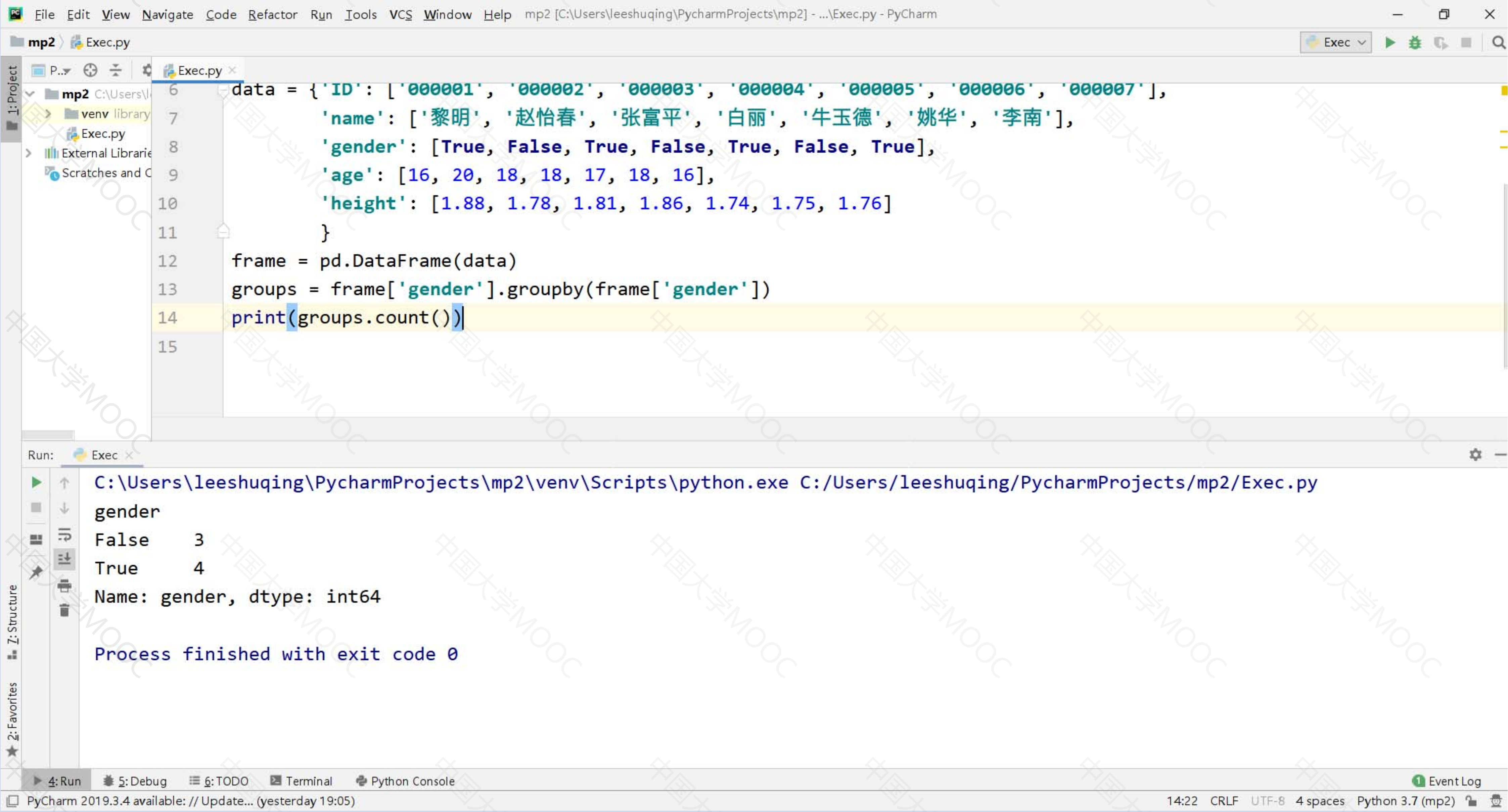
```python

data = {'ID': ['000001', '000002', '000003', '000004', '000005', '000006', '000007'],
        'name': ['黎明', '赵怡春', '张富平', '白丽', '牛玉德', '姚华', '李南'],
        'gender': [True, False, True, False, True, False, True],
        'age': [16, 20, 18, 18, 17, 18, 16],
        'height': [1.88, 1.78, 1.81, 1.86, 1.74, 1.75, 1.76]
        }
frame = pd.DataFrame(data)
groups = frame.groupby(frame['gender'])
print(groups['gender'].count())

```

```
C:\Users\leeshuqing\PycharmProjects\mp2\venv\Scripts\python.exe C:/Users/leeshuqing/PycharmProjects/mp2/Exec.py
gender
False    3
True     4
Name: gender, dtype: int64

Process finished with exit code 0
```


File Edit View Navigate Code Refactor Run Tools VCS Window Help mp2 [C:\Users\leeshuqing\PycharmProjects\mp2] - ...\Exec.py - PyCharm
mp2 Exec.py
Exec
1: Project
P...
mp2 C:\Users\l
venv library
Exec.py
External Librarie
Scratches and C
Exec.py
Run: Exec
7: Structure
2: Favorites
4: Run
5: Debug
6: TODO
Terminal
Python Console
Event Log
PyCharm 2019.3.4 available: // Update... (yesterday 19:05)
14:28 CRLF UTF-8 4 spaces Python 3.7 (mp2)

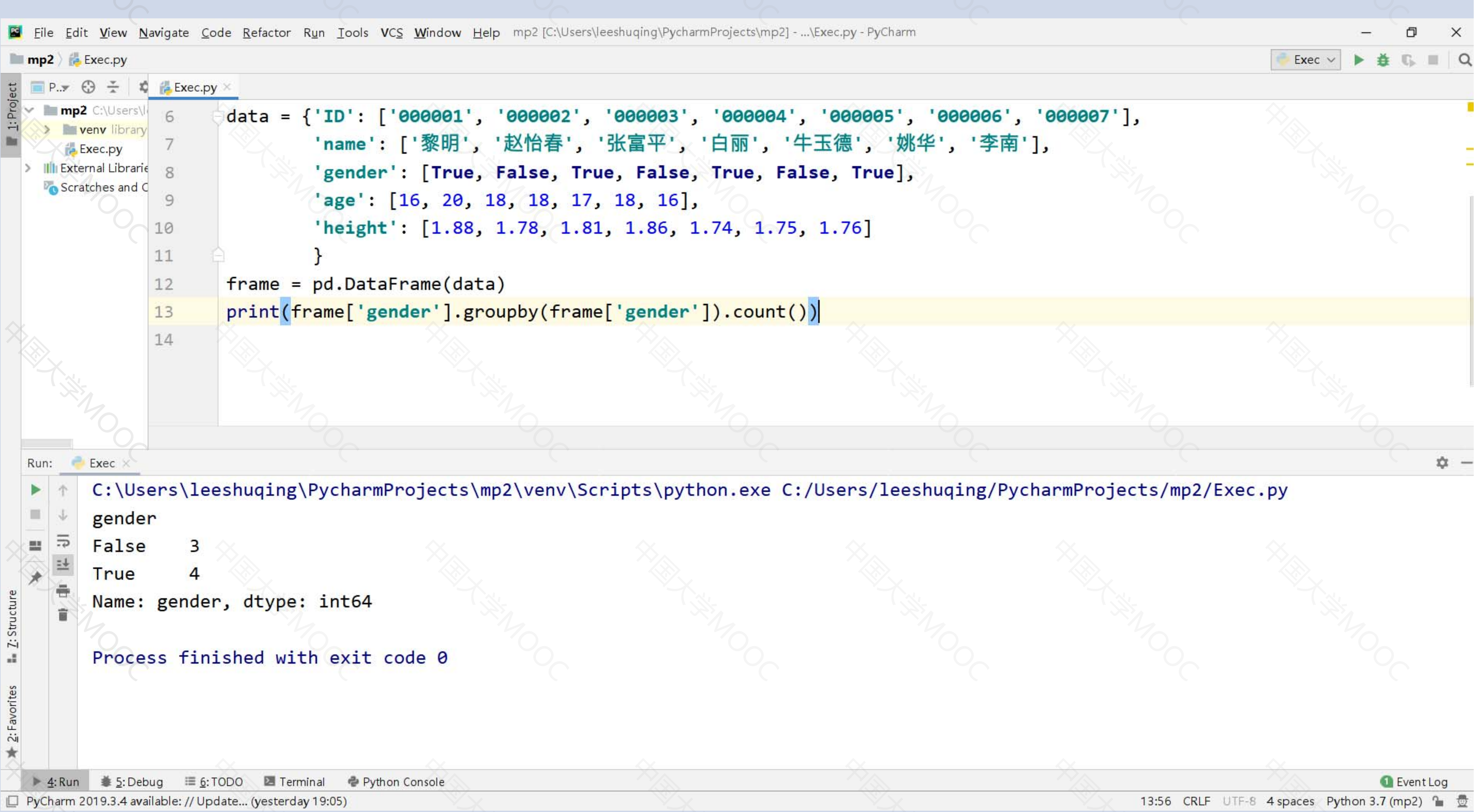
```python

data = {'ID': ['000001', '000002', '000003', '000004', '000005', '000006', '000007'],
        'name': ['黎明', '赵怡春', '张富平', '白丽', '牛玉德', '姚华', '李南'],
        'gender': [True, False, True, False, True, False, True],
        'age': [16, 20, 18, 18, 17, 18, 16],
        'height': [1.88, 1.78, 1.81, 1.86, 1.74, 1.75, 1.76]
        }
frame = pd.DataFrame(data)
groups = frame.groupby(frame['gender'])
print(groups['ID'].count())

```

```
C:\Users\leeshuqing\PycharmProjects\mp2\venv\Scripts\python.exe C:/Users/leeshuqing/PycharmProjects/mp2/Exec.py
gender
False    3
True     4
Name: gender, dtype: int64

Process finished with exit code 0
```


File Edit View Navigate Code Refactor Run Tools VCS Window Help mp2 [C:\Users\leeshuqing\PycharmProjects\mp2] - ...\Exec.py - PyCharm
mp2 Exec.py
Exec
mp2 C:\Users\l
venv library
Exec.py
External Librarie
Scratches and C
Run: Exec
4: Run 5: Debug 6: TODO Terminal Python Console
Event Log
PyCharm 2019.3.4 available: // Update... (yesterday 19:05)
14:22 CRLF UTF-8 4 spaces Python 3.7 (mp2)

```python
data = {'ID': ['000001', '000002', '000003', '000004', '000005', '000006', '000007'],
        'name': ['黎明', '赵怡春', '张富平', '白丽', '牛玉德', '姚华', '李南'],
        'gender': [True, False, True, False, True, False, True],
        'age': [16, 20, 18, 18, 17, 18, 16],
        'height': [1.88, 1.78, 1.81, 1.86, 1.74, 1.75, 1.76]
        }
frame = pd.DataFrame(data)
groups = frame['gender'].groupby(frame['gender'])
print(groups.count())
```

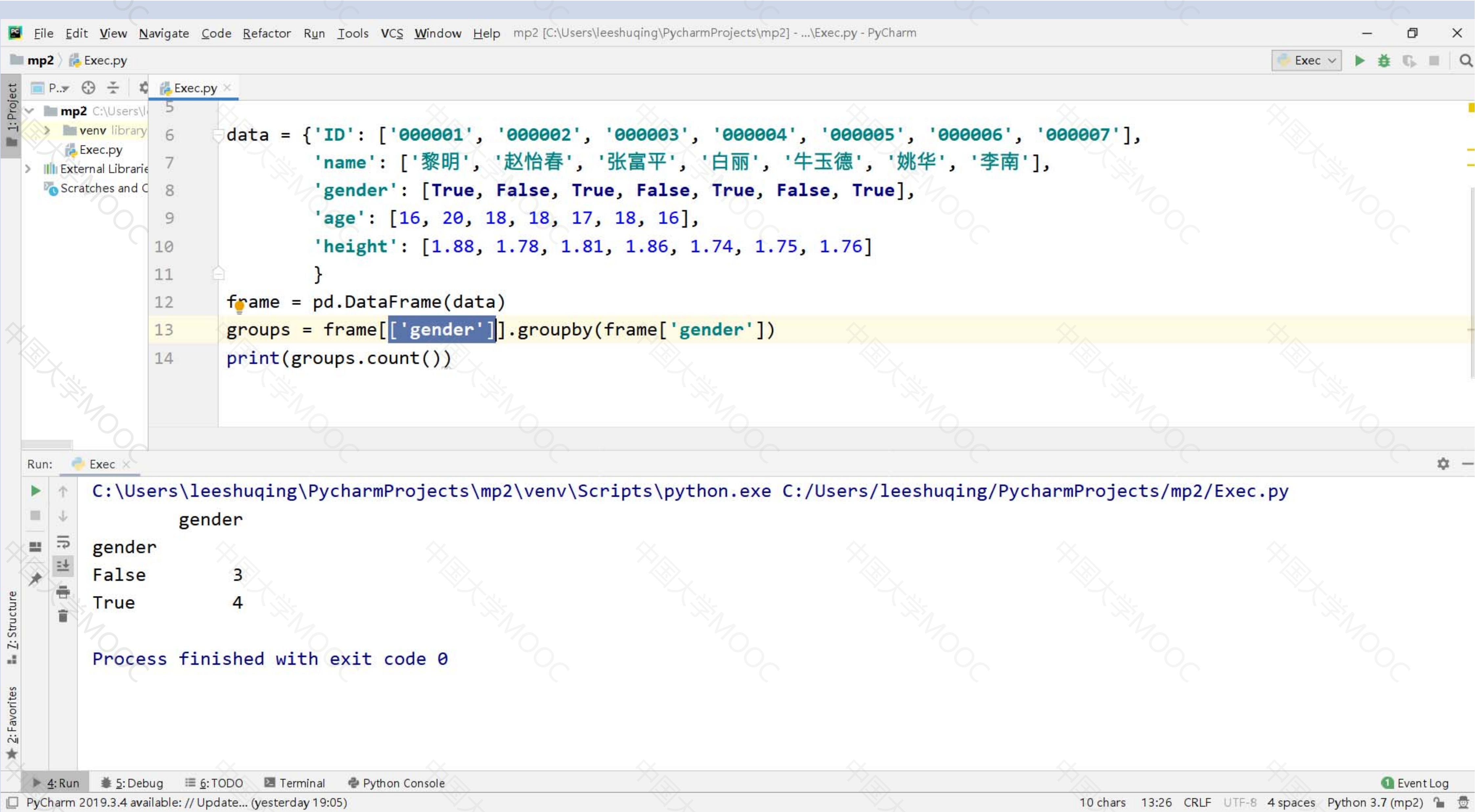
```
C:\Users\leeshuqing\PycharmProjects\mp2\venv\Scripts\python.exe C:/Users/leeshuqing/PycharmProjects/mp2/Exec.py
gender
False    3
True     4
Name: gender, dtype: int64

Process finished with exit code 0
```



File Edit View Navigate Code Refactor Run Tools VCS Window Help mp2 [C:\Users\leeshuqing\PycharmProjects\mp2] - ...\Exec.py - PyCharm
mp2 〉 Exec.py
Exec
Exec.py
mp2 C:\Users\l
venv library
Exec.py
External Librarie
Scratches and C
Run: Exec
4: Run
5: Debug
6: TODO
Terminal
Python Console
Event Log
PyCharm 2019.3.4 available: // Update... (yesterday 19:05)
13:56 CRLF UTF-8 4 spaces Python 3.7 (mp2)

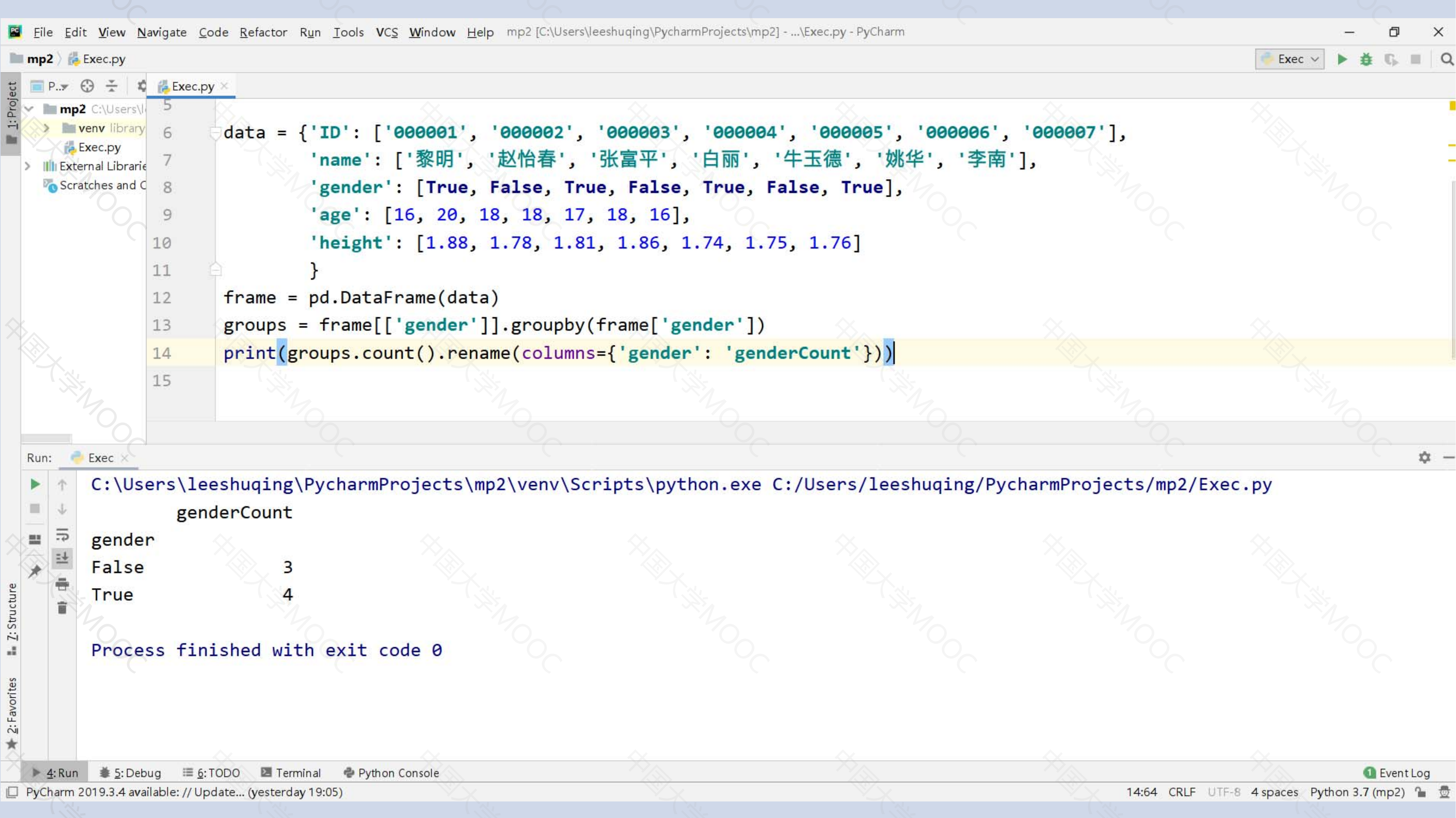
```python
data = {'ID': ['000001', '000002', '000003', '000004', '000005', '000006', '000007'],
        'name': ['黎明', '赵怡春', '张富平', '白丽', '牛玉德', '姚华', '李南'],
        'gender': [True, False, True, False, True, False, True],
        'age': [16, 20, 18, 18, 17, 18, 16],
        'height': [1.88, 1.78, 1.81, 1.86, 1.74, 1.75, 1.76]
        }
frame = pd.DataFrame(data)
print(frame['gender'].groupby(frame['gender']).count())
```

```
C:\Users\leeshuqing\PycharmProjects\mp2\venv\Scripts\python.exe C:/Users/leeshuqing/PycharmProjects/mp2/Exec.py
gender
False    3
True     4
Name: gender, dtype: int64

Process finished with exit code 0
```


File Edit View Navigate Code Refactor Run Tools VCS Window Help mp2 [C:\Users\leeshuqing\PycharmProjects\mp2] - ...\Exec.py - PyCharm
mp2 〉 Exec.py
Exec
1: Project
P...
Exec.py
mp2 C:\Users\l
venv library
Exec.py
External Librarie
Scratches and C
Run: Exec
7: Structure
2: Favorites
4: Run
5: Debug
6: TODO
Terminal
Python Console
Event Log
PyCharm 2019.3.4 available: // Update... (yesterday 19:05)
10 chars 13:26 CRLF UTF-8 4 spaces Python 3.7 (mp2)

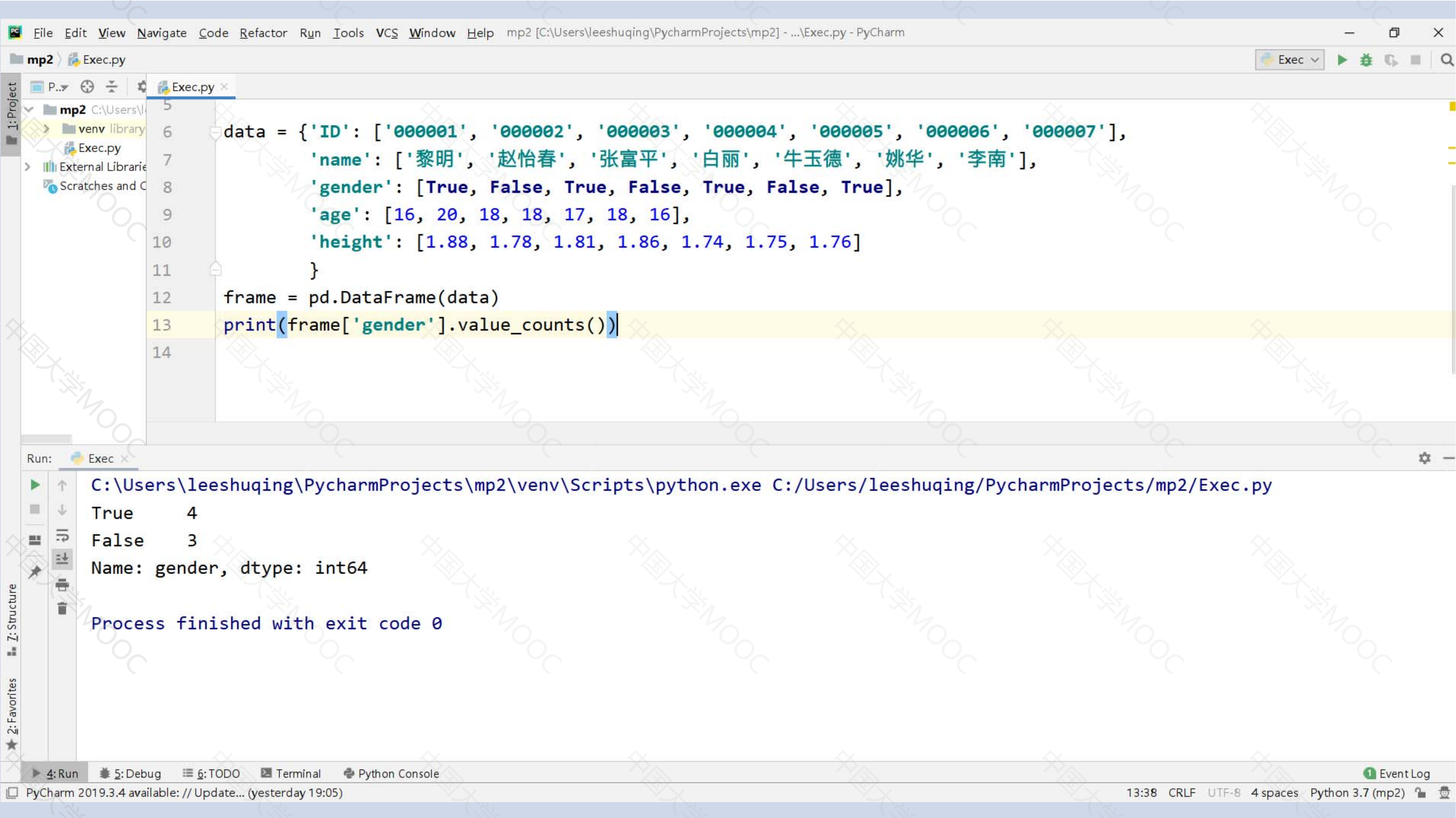
```python
data = {'ID': ['000001', '000002', '000003', '000004', '000005', '000006', '000007'],
        'name': ['黎明', '赵怡春', '张富平', '白丽', '牛玉德', '姚华', '李南'],
        'gender': [True, False, True, False, True, False, True],
        'age': [16, 20, 18, 18, 17, 18, 16],
        'height': [1.88, 1.78, 1.81, 1.86, 1.74, 1.75, 1.76]
        }
frame = pd.DataFrame(data)
groups = frame[['gender']].groupby(frame['gender'])
print(groups.count())
```

```
C:\Users\leeshuqing\PycharmProjects\mp2\venv\Scripts\python.exe C:/Users/leeshuqing/PycharmProjects/mp2/Exec.py
        gender
gender
False        3
True         4

Process finished with exit code 0
```


File Edit View Navigate Code Refactor Run Tools VCS Window Help mp2 [C:\Users\leeshuqing\PycharmProjects\mp2] - ...\Exec.py - PyCharm
mp2 Exec.py
Exec
Exec.py
mp2 C:\Users\l
venv library
Exec.py
External Librarie
Scratches and C
Run: Exec
4: Run 5: Debug 6: TODO Terminal Python Console Event Log
PyCharm 2019.3.4 available: // Update... (yesterday 19:05)
14:64 CRLF UTF-8 4 spaces Python 3.7 (mp2)

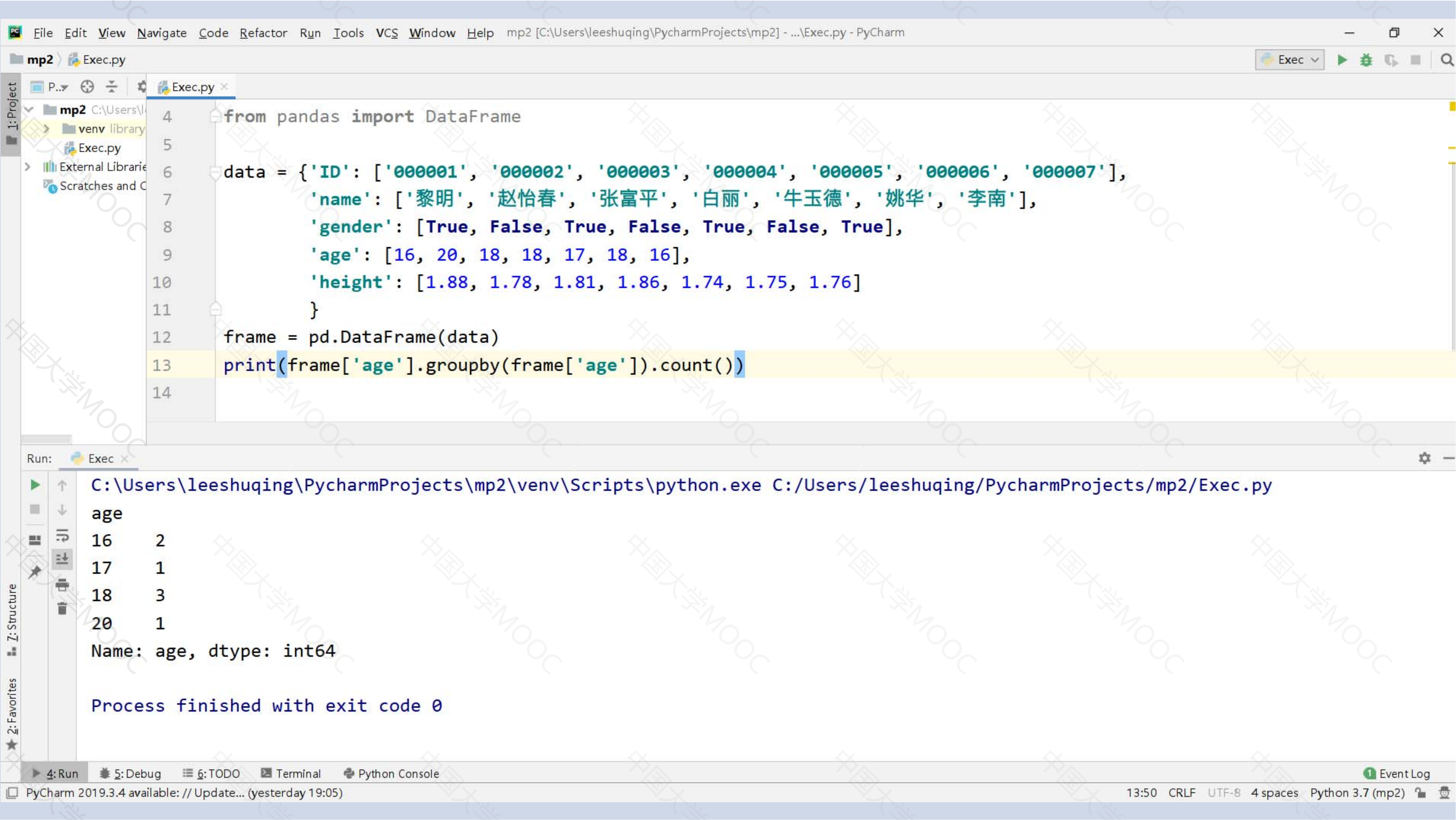
```python
data = {'ID': ['000001', '000002', '000003', '000004', '000005', '000006', '000007'],
        'name': ['黎明', '赵怡春', '张富平', '白丽', '牛玉德', '姚华', '李南'],
        'gender': [True, False, True, False, True, False, True],
        'age': [16, 20, 18, 18, 17, 18, 16],
        'height': [1.88, 1.78, 1.81, 1.86, 1.74, 1.75, 1.76]
        }
frame = pd.DataFrame(data)
groups = frame[['gender']].groupby(frame['gender'])
print(groups.count().rename(columns={'gender': 'genderCount'}))
```

```
C:\Users\leeshuqing\PycharmProjects\mp2\venv\Scripts\python.exe C:/Users/leeshuqing/PycharmProjects/mp2/Exec.py
        genderCount
gender
False             3
True              4

Process finished with exit code 0
```


File Edit View Navigate Code Refactor Run Tools VCS Window Help mp2 [C:\Users\leeshuqing\PycharmProjects\mp2] - ...\Exec.py - PyCharm
mp2 〉 Exec.py
Exec
1: Project
P..
mp2 C:\Users\l
venv library
Exec.py
External Librarie
Scratches and C
Exec.py
Run: Exec
Z: Structure
2: Favorites
4: Run
5: Debug
6: TODO
Terminal
Python Console
Event Log
PyCharm 2019.3.4 available: // Update... (yesterday 19:05)
13:38 CRLF UTF-8 4 spaces Python 3.7 (mp2)

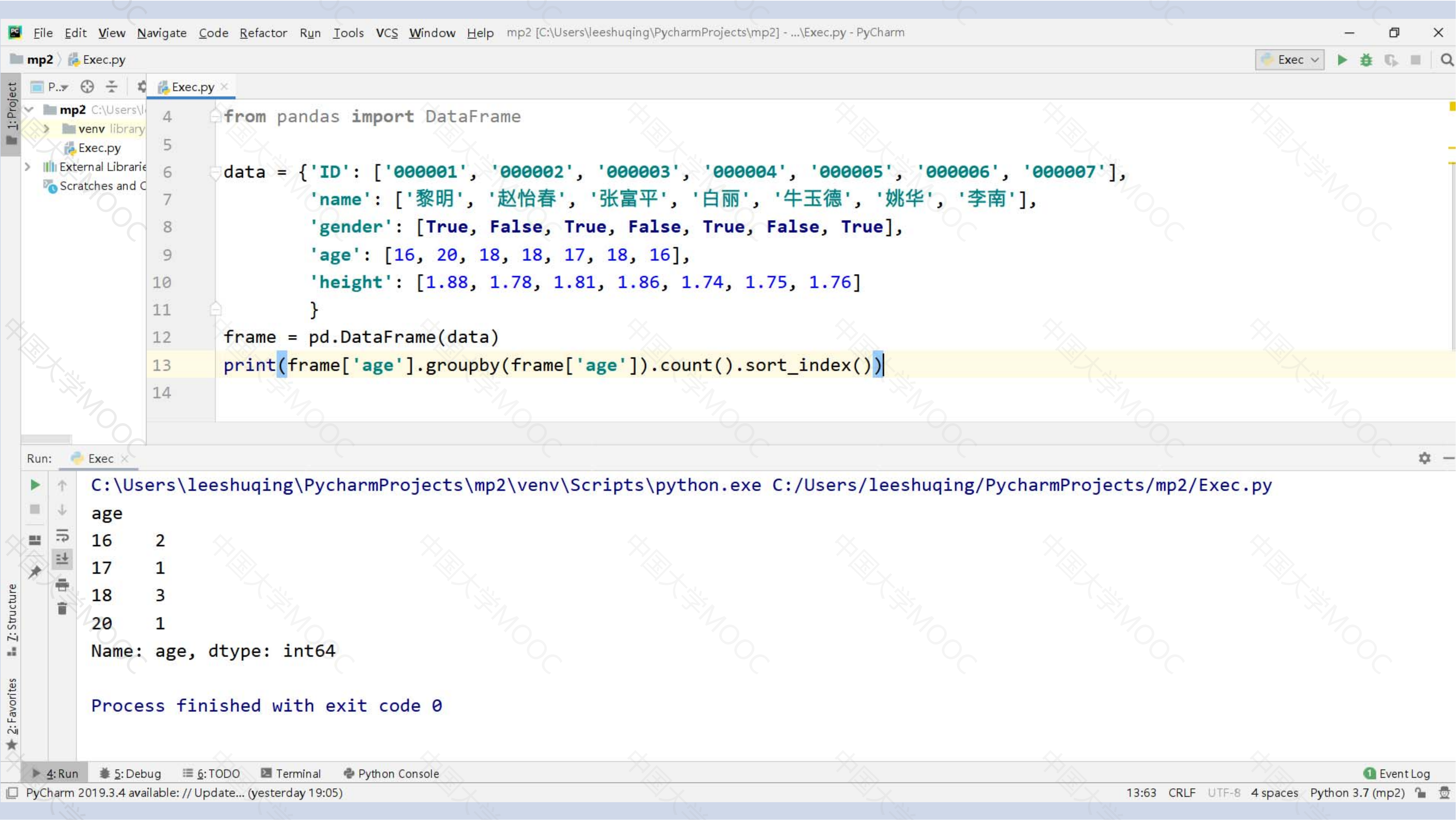
```python
data = {'ID': ['000001', '000002', '000003', '000004', '000005', '000006', '000007'],
        'name': ['黎明', '赵怡春', '张富平', '白丽', '牛玉德', '姚华', '李南'],
        'gender': [True, False, True, False, True, False, True],
        'age': [16, 20, 18, 18, 17, 18, 16],
        'height': [1.88, 1.78, 1.81, 1.86, 1.74, 1.75, 1.76]
        }
frame = pd.DataFrame(data)
print(frame['gender'].value_counts())
```

```
C:\Users\leeshuqing\PycharmProjects\mp2\venv\Scripts\python.exe C:/Users/leeshuqing/PycharmProjects/mp2/Exec.py
True     4
False    3
Name: gender, dtype: int64

Process finished with exit code 0
```


File Edit View Navigate Code Refactor Run Tools VCS Window Help mp2 [C:\Users\leeshuqing\PycharmProjects\mp2] - ...\Exec.py - PyCharm
mp2 〉 Exec.py
Exec
1: Project
P...
mp2 C:\Users\l
venv library
Exec.py
External Librarie
Scratches and C
Exec.py
Run: Exec
7: Structure
2: Favorites
4: Run
5: Debug
6: TODO
Terminal
Python Console
Event Log
PyCharm 2019.3.4 available: // Update... (yesterday 19:05)
13:50 CRLF UTF-8 4 spaces Python 3.7 (mp2)

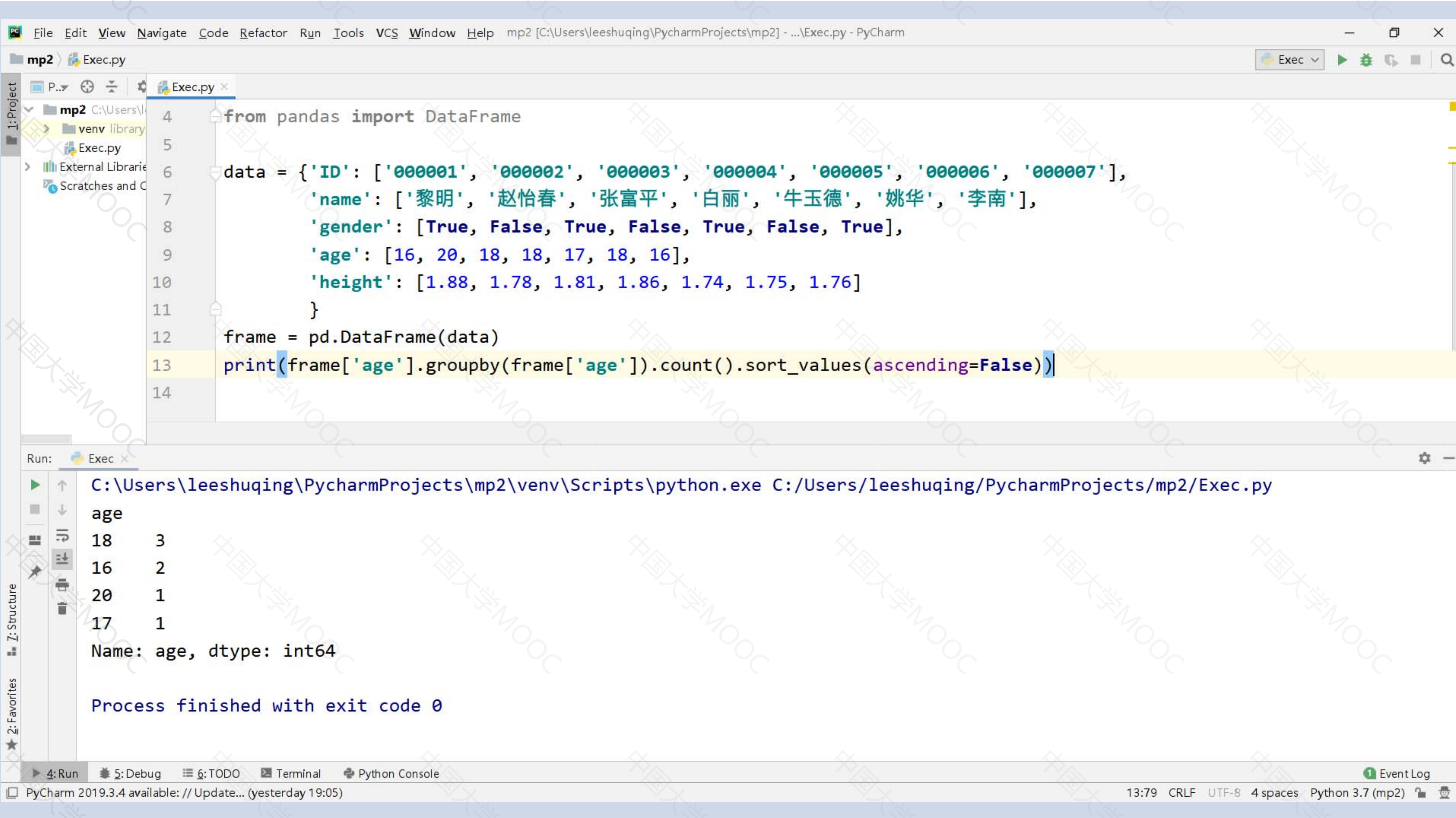
```python
from pandas import DataFrame

data = {'ID': ['000001', '000002', '000003', '000004', '000005', '000006', '000007'],
        'name': ['黎明', '赵怡春', '张富平', '白丽', '牛玉德', '姚华', '李南'],
        'gender': [True, False, True, False, True, False, True],
        'age': [16, 20, 18, 18, 17, 18, 16],
        'height': [1.88, 1.78, 1.81, 1.86, 1.74, 1.75, 1.76]
        }
frame = pd.DataFrame(data)
print(frame['age'].groupby(frame['age']).count())
```

```
C:\Users\leeshuqing\PycharmProjects\mp2\venv\Scripts\python.exe C:/Users/leeshuqing/PycharmProjects/mp2/Exec.py
age
16    2
17    1
18    3
20    1
Name: age, dtype: int64

Process finished with exit code 0
```


File Edit View Navigate Code Refactor Run Tools VCS Window Help mp2 [C:\Users\leeshuqing\PycharmProjects\mp2] - ...\Exec.py - PyCharm
mp2 〉 Exec.py
Exec
P..
Exec.py ×
mp2 C:\Users\l
venv library
Exec.py
External Librarie
Scratches and C
1: Project
Run: Exec ×
7: Structure
2: Favorites
4: Run
5: Debug
6: TODO
Terminal
Python Console
Event Log
PyCharm 2019.3.4 available: // Update... (yesterday 19:05)
13:63 CRLF UTF-8 4 spaces Python 3.7 (mp2)

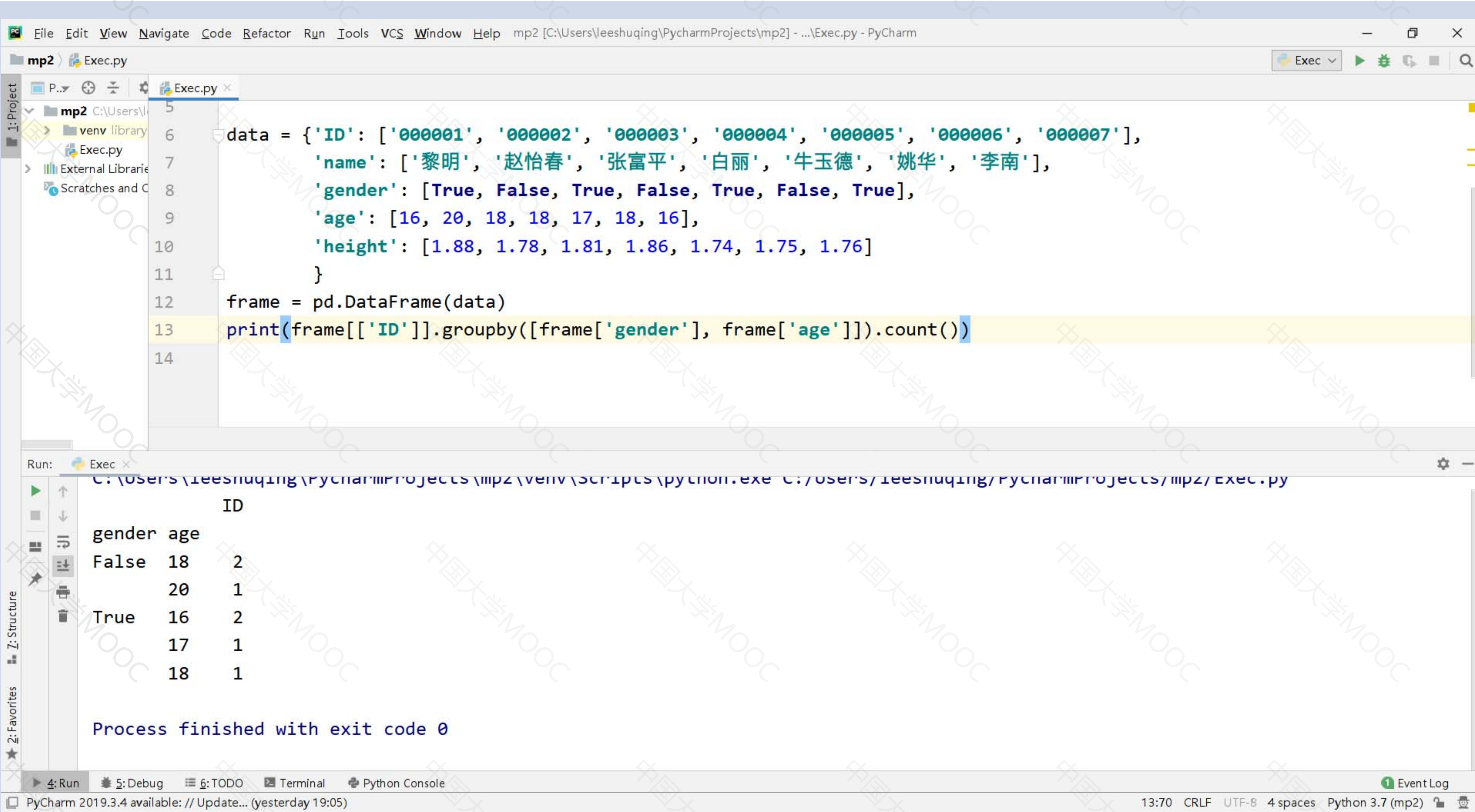
```python
from pandas import DataFrame

data = {'ID': ['000001', '000002', '000003', '000004', '000005', '000006', '000007'],
        'name': ['黎明', '赵怡春', '张富平', '白丽', '牛玉德', '姚华', '李南'],
        'gender': [True, False, True, False, True, False, True],
        'age': [16, 20, 18, 18, 17, 18, 16],
        'height': [1.88, 1.78, 1.81, 1.86, 1.74, 1.75, 1.76]
        }
frame = pd.DataFrame(data)
print(frame['age'].groupby(frame['age']).count().sort_index())
```

```
C:\Users\leeshuqing\PycharmProjects\mp2\venv\Scripts\python.exe C:/Users/leeshuqing/PycharmProjects/mp2/Exec.py
age
16    2
17    1
18    3
20    1
Name: age, dtype: int64

Process finished with exit code 0
```


File Edit View Navigate Code Refactor Run Tools VCS Window Help mp2 [C:\Users\leeshuqing\PycharmProjects\mp2] - ...\Exec.py - PyCharm
mp2 〉 Exec.py
Exec
Exec.py
mp2 C:\Users\l
venv library
Exec.py
External Librarie
Scratches and C
1: Project
Run: Exec
Z: Structure
2: Favorites
4: Run
5: Debug
6: TODO
Terminal
Python Console
Event Log
PyCharm 2019.3.4 available: // Update... (yesterday 19:05)
13:79 CRLF UTF-8 4 spaces Python 3.7 (mp2)

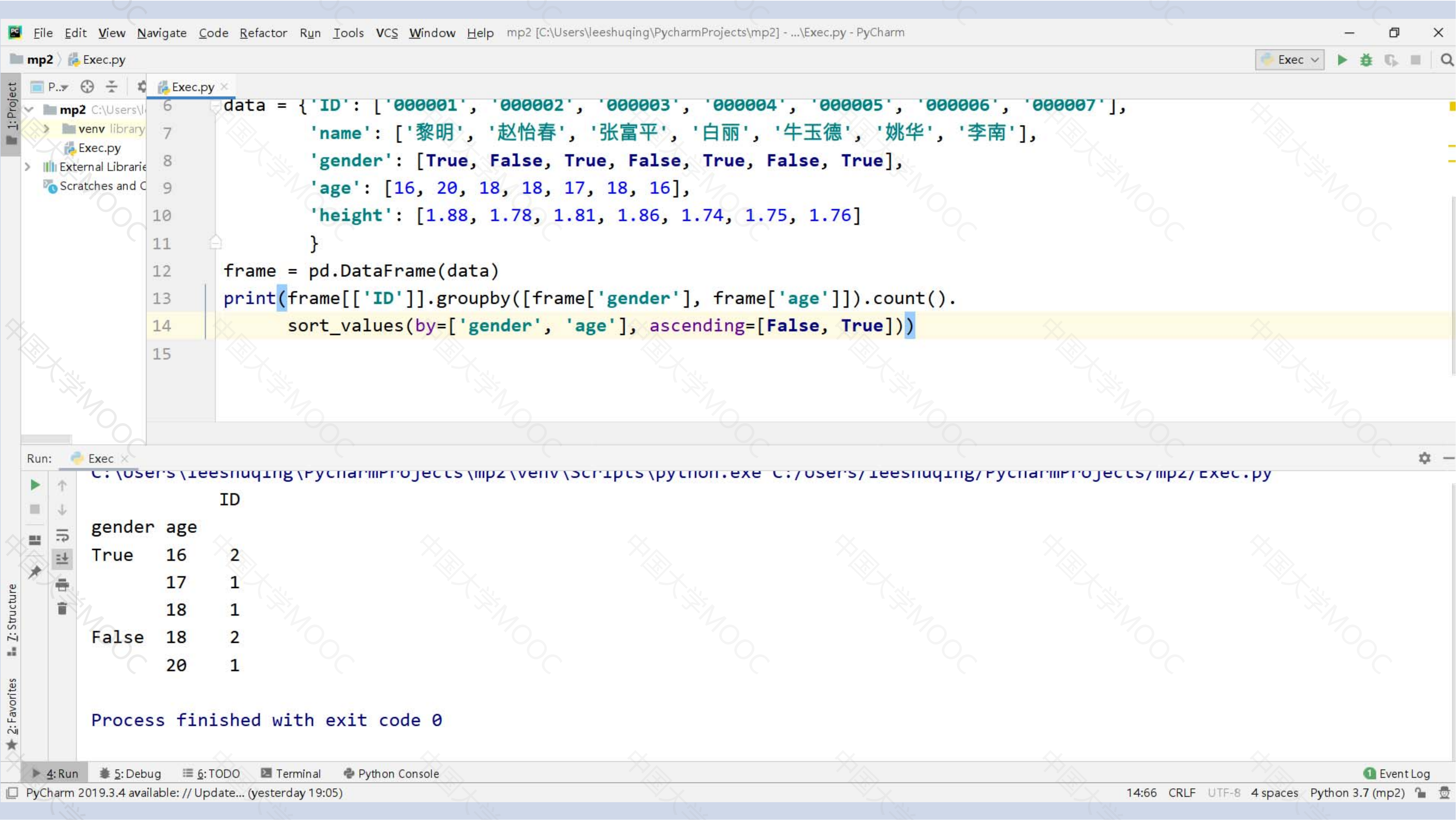
```python
from pandas import DataFrame

data = {'ID': ['000001', '000002', '000003', '000004', '000005', '000006', '000007'],
        'name': ['黎明', '赵怡春', '张富平', '白丽', '牛玉德', '姚华', '李南'],
        'gender': [True, False, True, False, True, False, True],
        'age': [16, 20, 18, 18, 17, 18, 16],
        'height': [1.88, 1.78, 1.81, 1.86, 1.74, 1.75, 1.76]
        }
frame = pd.DataFrame(data)
print(frame['age'].groupby(frame['age']).count().sort_values(ascending=False))
```

```
C:\Users\leeshuqing\PycharmProjects\mp2\venv\Scripts\python.exe C:/Users/leeshuqing/PycharmProjects/mp2/Exec.py
age
18    3
16    2
20    1
17    1
Name: age, dtype: int64

Process finished with exit code 0
```


File Edit View Navigate Code Refactor Run Tools VCS Window Help mp2 [C:\Users\leeshuqing\PycharmProjects\mp2] - ...\Exec.py - PyCharm
mp2 〉 Exec.py
Exec
1: Project
P..
Exec.py
mp2 C:\Users\l
venv library
Exec.py
External Librarie
Scratches and C
Run: Exec
7: Structure
2: Favorites
4: Run
5: Debug
6: TODO
Terminal
Python Console
Event Log
PyCharm 2019.3.4 available: // Update... (yesterday 19:05)
13:70 CRLF UTF-8 4 spaces Python 3.7 (mp2)

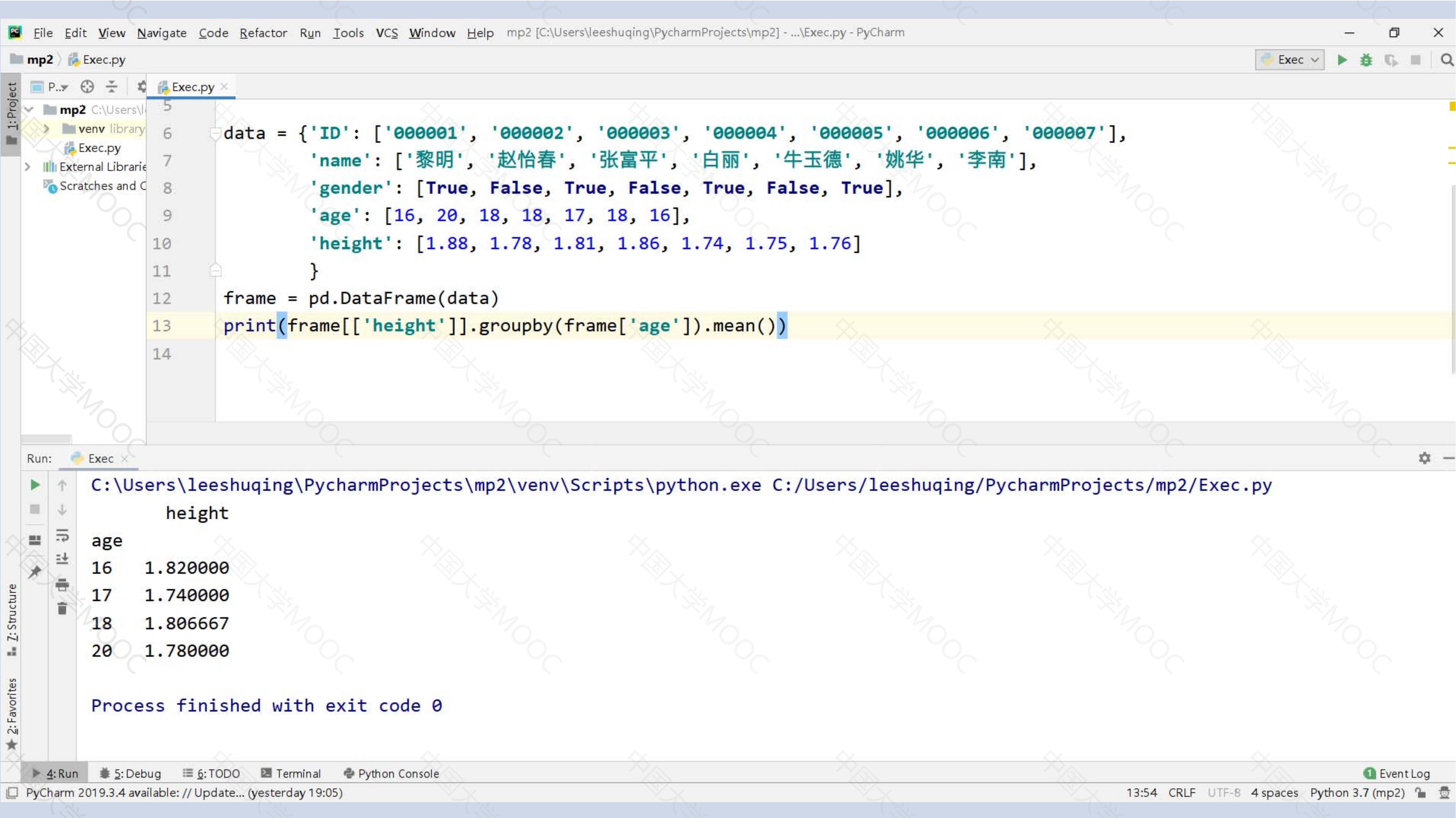
```

data = {'ID': ['000001', '000002', '000003', '000004', '000005', '000006', '000007'],
        'name': ['黎明', '赵怡春', '张富平', '白丽', '牛玉德', '姚华', '李南'],
        'gender': [True, False, True, False, True, False, True],
        'age': [16, 20, 18, 18, 17, 18, 16],
        'height': [1.88, 1.78, 1.81, 1.86, 1.74, 1.75, 1.76]
        }
frame = pd.DataFrame(data)
print(frame[['ID']].groupby([frame['gender'], frame['age']]).count())

```

```
C:\Users\leeshuqing\PycharmProjects\mp2\venv\Scripts\python.exe C:/Users/leeshuqing/PycharmProjects/mp2/Exec.py
            ID
gender age
False  18    2
       20    1
True   16    2
       17    1
       18    1

Process finished with exit code 0
```


File Edit View Navigate Code Refactor Run Tools VCS Window Help mp2 [C:\Users\leeshuqing\PycharmProjects\mp2] - ...\Exec.py - PyCharm
mp2 Exec.py
Exec
1: Project
P...
mp2 C:\Users\l
venv library
Exec.py
External Librarie
Scratches and C
Exec.py
Run: Exec
7: Structure
2: Favorites
4: Run
5: Debug
6: TODO
Terminal
Python Console
Event Log
PyCharm 2019.3.4 available: // Update... (yesterday 19:05)
14:66 CRLF UTF-8 4 spaces Python 3.7 (mp2)

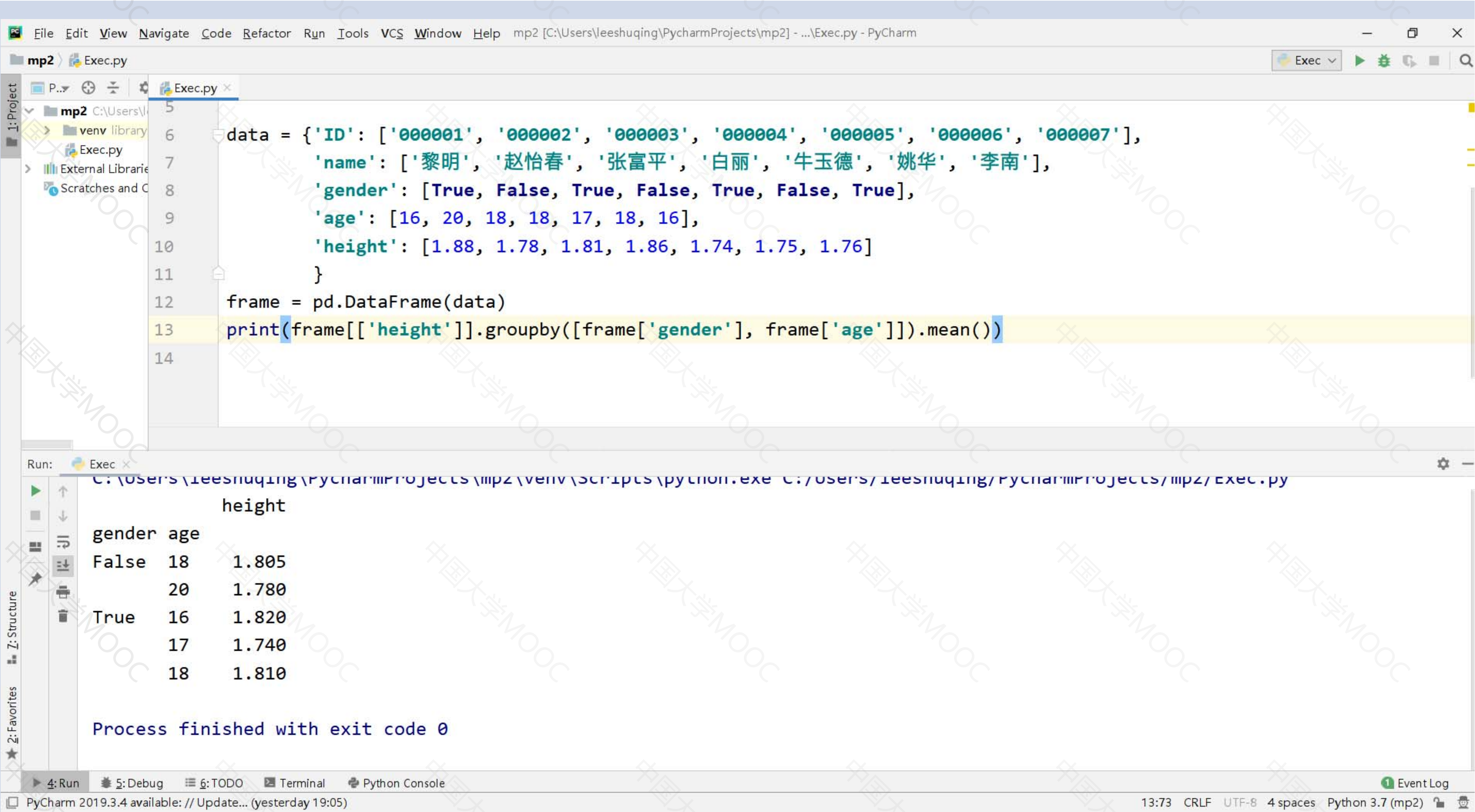
```python
data = {'ID': ['000001', '000002', '000003', '000004', '000005', '000006', '000007'],
        'name': ['黎明', '赵怡春', '张富平', '白丽', '牛玉德', '姚华', '李南'],
        'gender': [True, False, True, False, True, False, True],
        'age': [16, 20, 18, 18, 17, 18, 16],
        'height': [1.88, 1.78, 1.81, 1.86, 1.74, 1.75, 1.76]
        }
frame = pd.DataFrame(data)
print(frame[['ID']].groupby([frame['gender'], frame['age']]).count().
      sort_values(by=['gender', 'age'], ascending=[False, True]))
```

```
C:\Users\leeshuqing\PycharmProjects\mp2\venv\Scripts\python.exe C:/Users/leeshuqing/PycharmProjects/mp2/Exec.py
            ID
gender age    
True   16    2
       17    1
       18    1
False  18    2
       20    1

Process finished with exit code 0
```


File Edit View Navigate Code Refactor Run Tools VCS Window Help mp2 [C:\Users\leeshuqing\PycharmProjects\mp2] - ...\Exec.py - PyCharm
mp2 〉 Exec.py
Exec
1: Project
P..
Exec.py
mp2 C:\Users\l
venv library
Exec.py
External Librarie
Scratches and C
Run: Exec
7: Structure
2: Favorites
4: Run 5: Debug 6: TODO Terminal Python Console
Event Log
PyCharm 2019.3.4 available: // Update... (yesterday 19:05)
13:54 CRLF UTF-8 4 spaces Python 3.7 (mp2)

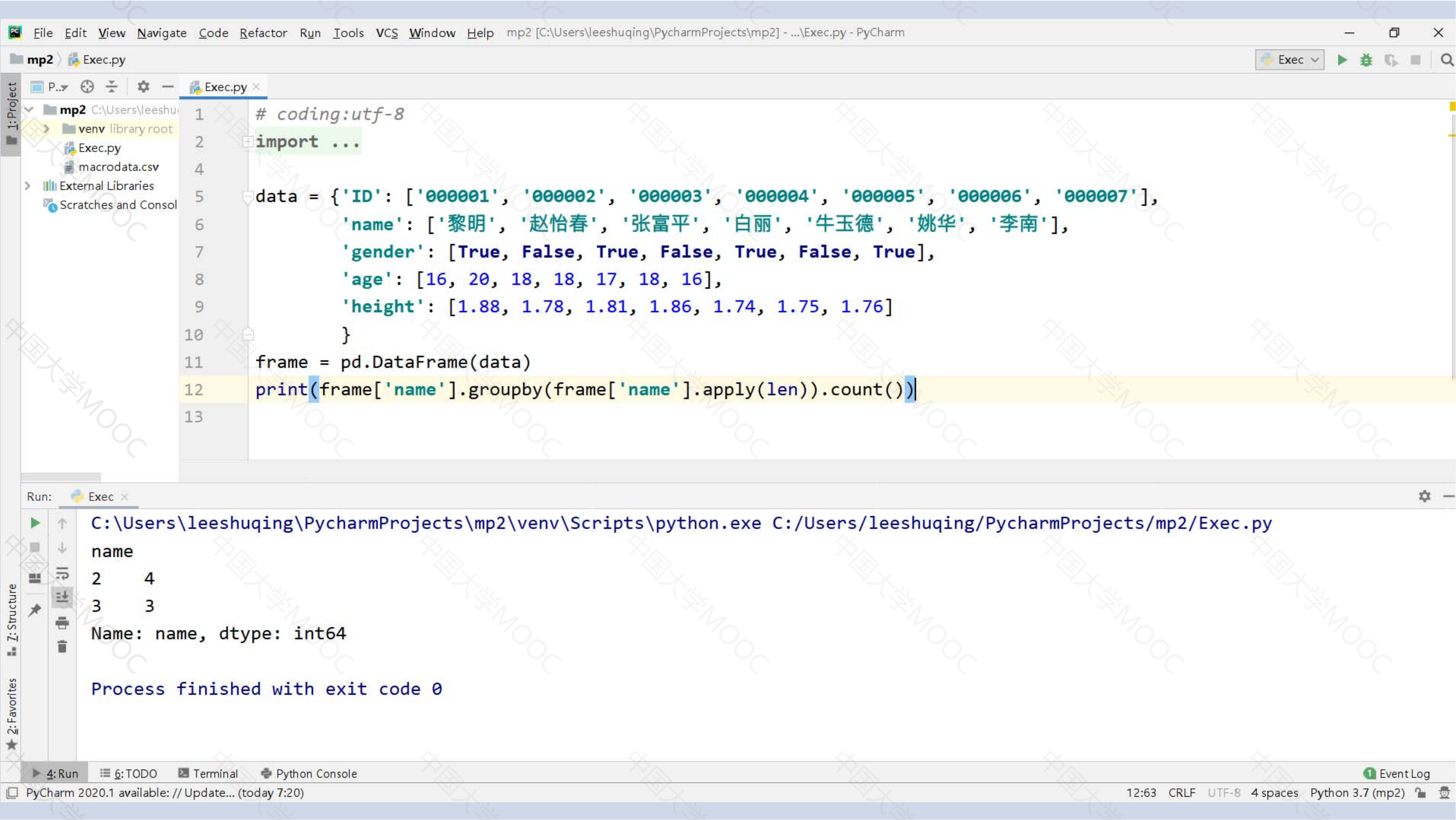
```python
data = {'ID': ['000001', '000002', '000003', '000004', '000005', '000006', '000007'],
        'name': ['黎明', '赵怡春', '张富平', '白丽', '牛玉德', '姚华', '李南'],
        'gender': [True, False, True, False, True, False, True],
        'age': [16, 20, 18, 18, 17, 18, 16],
        'height': [1.88, 1.78, 1.81, 1.86, 1.74, 1.75, 1.76]
        }
frame = pd.DataFrame(data)
print(frame[['height']].groupby(frame['age']).mean())
```

```
C:\Users\leeshuqing\PycharmProjects\mp2\venv\Scripts\python.exe C:/Users/leeshuqing/PycharmProjects/mp2/Exec.py
       height
age
16   1.820000
17   1.740000
18   1.806667
20   1.780000

Process finished with exit code 0
```


File Edit View Navigate Code Refactor Run Tools VCS Window Help mp2 [C:\Users\leeshuqing\PycharmProjects\mp2] - ...\Exec.py - PyCharm
mp2 〉 Exec.py
Exec
1: Project
mp2 C:\Users\l
venv library
Exec.py
External Librarie
Scratches and C
Run: Exec
4: Run
5: Debug
6: TODO
Terminal
Python Console
Event Log
PyCharm 2019.3.4 available: // Update... (yesterday 19:05)
13:73 CRLF UTF-8 4 spaces Python 3.7 (mp2)
Z: Structure
2: Favorites

```python

data = {'ID': ['000001', '000002', '000003', '000004', '000005', '000006', '000007'],
        'name': ['黎明', '赵怡春', '张富平', '白丽', '牛玉德', '姚华', '李南'],
        'gender': [True, False, True, False, True, False, True],
        'age': [16, 20, 18, 18, 17, 18, 16],
        'height': [1.88, 1.78, 1.81, 1.86, 1.74, 1.75, 1.76]
        }
frame = pd.DataFrame(data)
print(frame[['height']].groupby([frame['gender'], frame['age']]).mean())

```

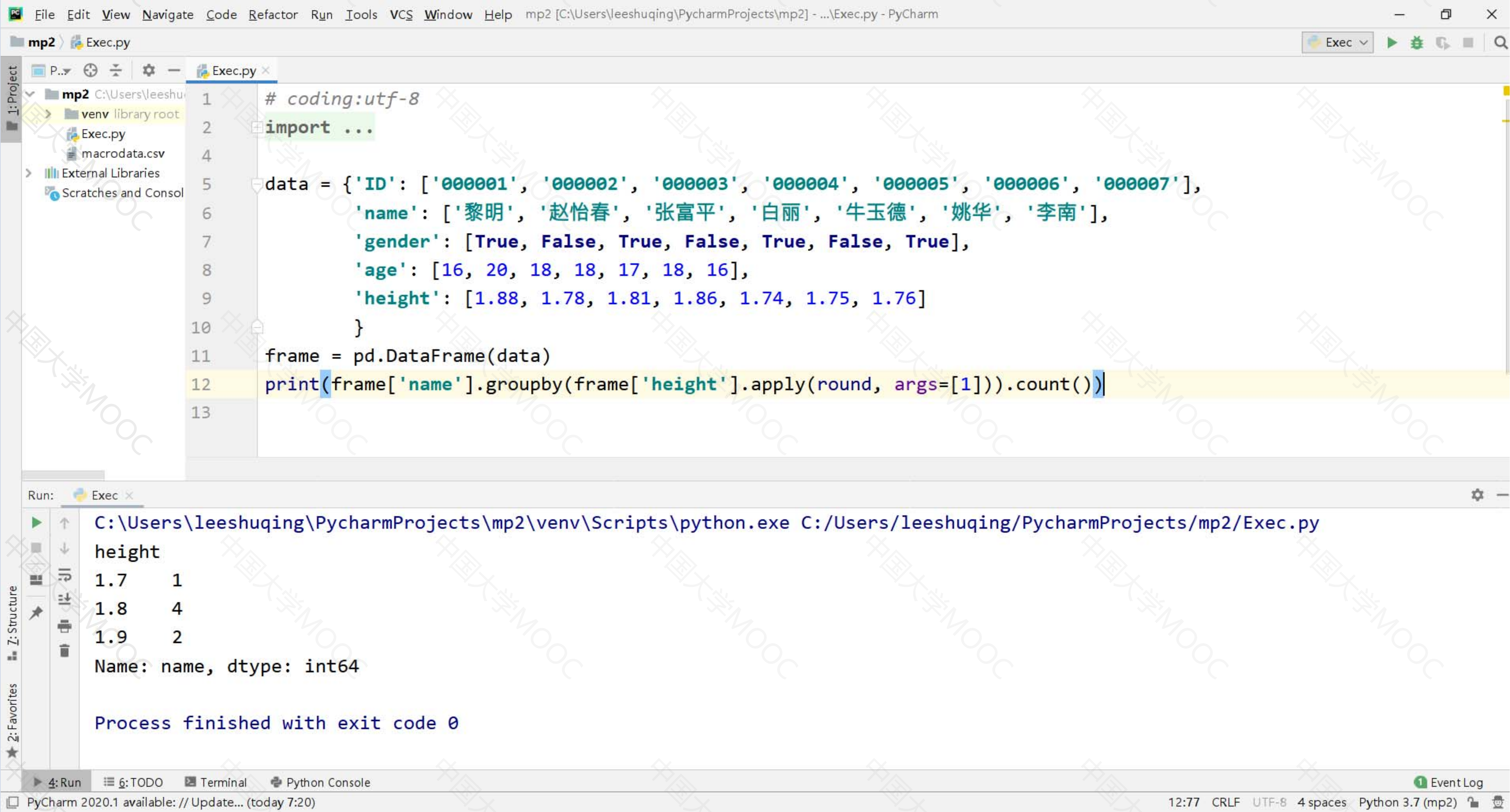
```
C:\Users\leeshuqing\PycharmProjects\mp2\venv\Scripts\python.exe C:/Users/leeshuqing/PycharmProjects/mp2/Exec.py
            height
gender age        
False  18    1.805
       20    1.780
True   16    1.820
       17    1.740
       18    1.810

Process finished with exit code 0
```



File Edit View Navigate Code Refactor Run Tools VCS Window Help mp2 [C:\Users\leeshuqing\PycharmProjects\mp2] - ...\Exec.py - PyCharm
mp2 Exec.py
Exec
1: Project
mp2 C:\Users\leeshu
venv library root
Exec.py
macrodata.csv
External Libraries
Scratches and Consol
Run: Exec
Z: Structure
2: Favorites
4: Run
6: TODO
Terminal
Python Console
Event Log
PyCharm 2020.1 available: // Update... (today 7:20)
12:63 CRLF UTF-8 4 spaces Python 3.7 (mp2)

```python
# coding:utf-8
import ...

data = {'ID': ['000001', '000002', '000003', '000004', '000005', '000006', '000007'],
        'name': ['黎明', '赵怡春', '张富平', '白丽', '牛玉德', '姚华', '李南'],
        'gender': [True, False, True, False, True, False, True],
        'age': [16, 20, 18, 18, 17, 18, 16],
        'height': [1.88, 1.78, 1.81, 1.86, 1.74, 1.75, 1.76]
        }
frame = pd.DataFrame(data)
print(frame['name'].groupby(frame['name'].apply(len)).count())
```

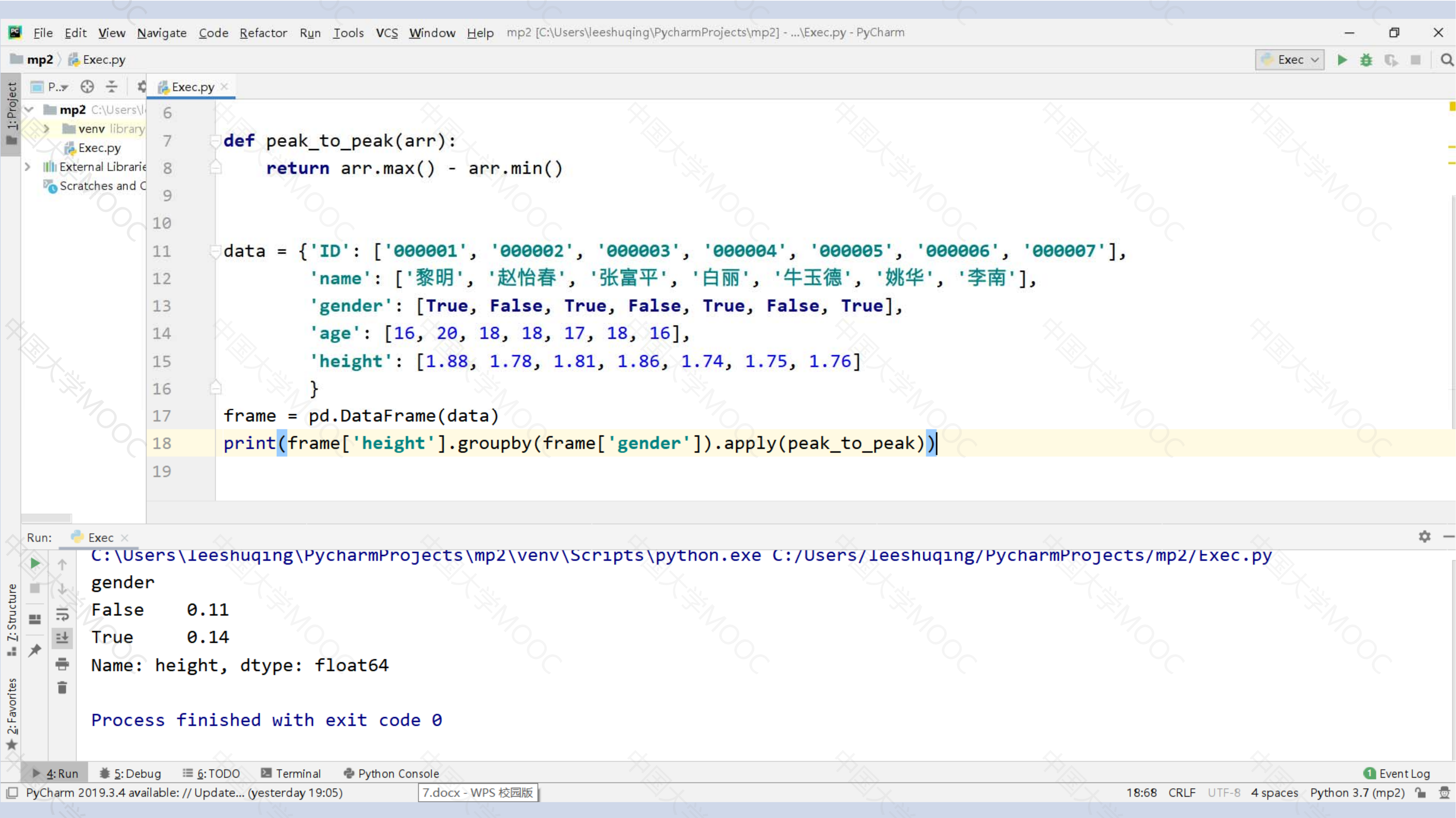
```
C:\Users\leeshuqing\PycharmProjects\mp2\venv\Scripts\python.exe C:/Users/leeshuqing/PycharmProjects/mp2/Exec.py
name
2    4
3    3
Name: name, dtype: int64

Process finished with exit code 0
```



File Edit View Navigate Code Refactor Run Tools VCS Window Help mp2 [C:\Users\leeshuqing\PycharmProjects\mp2] - ...\Exec.py - PyCharm
mp2 Exec.py
Exec
mp2 C:\Users\leeshu
venv library root
Exec.py
macrodata.csv
External Libraries
Scratches and Consol
Run: Exec
4: Run 6: TODO Terminal Python Console
Event Log
PyCharm 2020.1 available: // Update... (today 7:20)
12:77 CRLF UTF-8 4 spaces Python 3.7 (mp2)

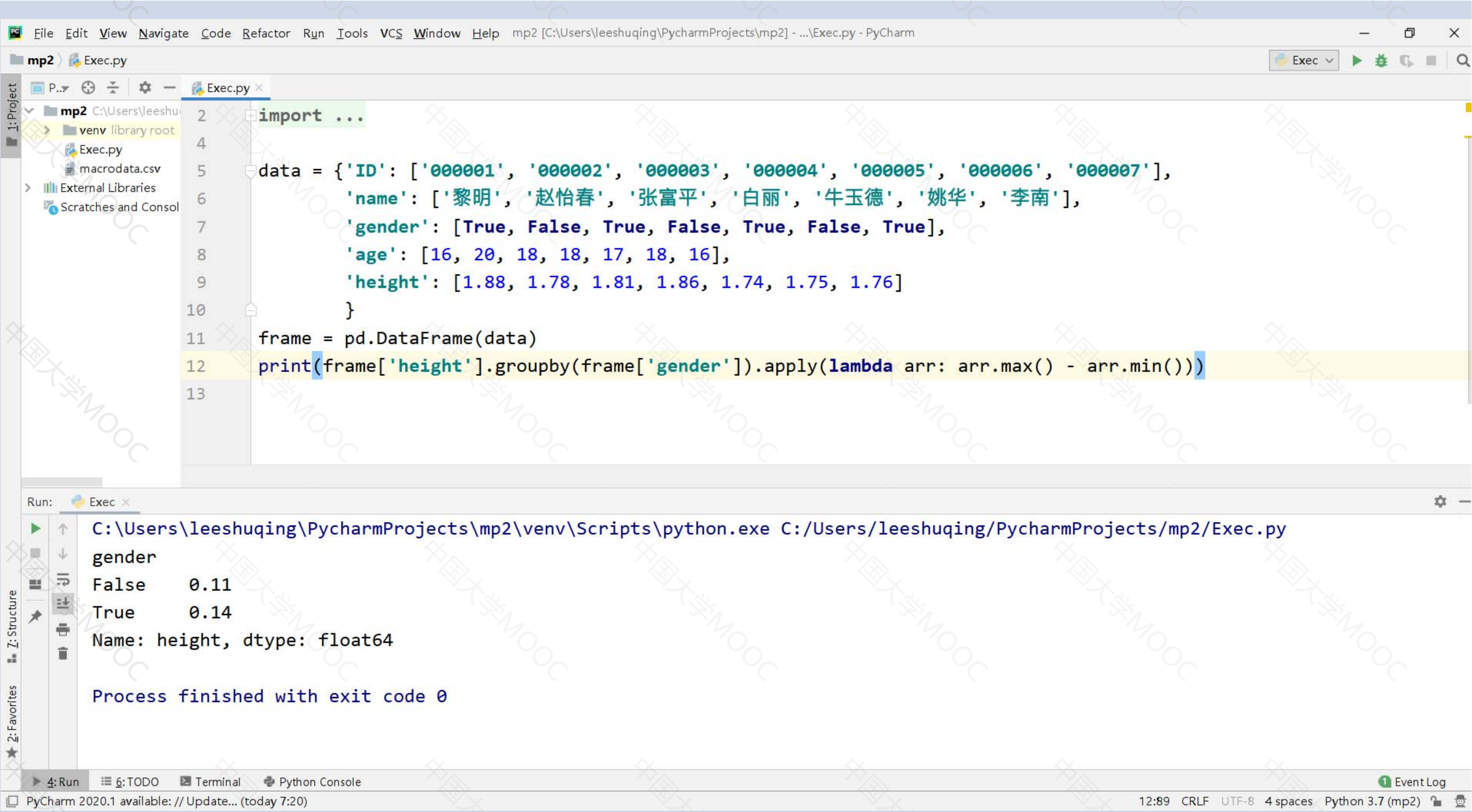
```python
# coding:utf-8
import ...

data = {'ID': ['000001', '000002', '000003', '000004', '000005', '000006', '000007'],
        'name': ['黎明', '赵怡春', '张富平', '白丽', '牛玉德', '姚华', '李南'],
        'gender': [True, False, True, False, True, False, True],
        'age': [16, 20, 18, 18, 17, 18, 16],
        'height': [1.88, 1.78, 1.81, 1.86, 1.74, 1.75, 1.76]
        }
frame = pd.DataFrame(data)
print(frame['name'].groupby(frame['height'].apply(round, args=[1])).count())
```

```
C:\Users\leeshuqing\PycharmProjects\mp2\venv\Scripts\python.exe C:/Users/leeshuqing/PycharmProjects/mp2/Exec.py
height
1.7    1
1.8    4
1.9    2
Name: name, dtype: int64

Process finished with exit code 0
```


File Edit View Navigate Code Refactor Run Tools VCS Window Help mp2 [C:\Users\leeshuqing\PycharmProjects\mp2] - ...\Exec.py - PyCharm
mp2 〉 Exec.py
Exec
Exec.py
mp2 C:\Users\l
venv library
Exec.py
External Librarie
Scratches and C
Run: Exec
4: Run 5: Debug 6: TODO Terminal Python Console Event Log
PyCharm 2019.3.4 available: // Update... (yesterday 19:05)
7.docx - WPS 校园版
18:68 CRLF UTF-8 4 spaces Python 3.7 (mp2)

```
def peak_to_peak(arr):
    return arr.max() - arr.min()


data = {'ID': ['000001', '000002', '000003', '000004', '000005', '000006', '000007'],
        'name': ['黎明', '赵怡春', '张富平', '白丽', '牛玉德', '姚华', '李南'],
        'gender': [True, False, True, False, True, False, True],
        'age': [16, 20, 18, 18, 17, 18, 16],
        'height': [1.88, 1.78, 1.81, 1.86, 1.74, 1.75, 1.76]
        }
frame = pd.DataFrame(data)
print(frame['height'].groupby(frame['gender']).apply(peak_to_peak))
```

```
C:\Users\leeshuqing\PycharmProjects\mp2\venv\Scripts\python.exe C:/Users/leeshuqing/PycharmProjects/mp2/Exec.py
gender
False    0.11
True     0.14
Name: height, dtype: float64

Process finished with exit code 0
```


File Edit View Navigate Code Refactor Run Tools VCS Window Help
mp2 [C:\Users\leeshuqing\PycharmProjects\mp2] - ...\Exec.py - PyCharm
mp2 Exec.py
Exec
1: Project
mp2 C:\Users\leeshu
venv library root
Exec.py
macrodata.csv
External Libraries
Scratches and Consol
Run: Exec
4: Run
6: TODO
Terminal
Python Console
Event Log
Z: Structure
2: Favorites
PyCharm 2020.1 available: // Update... (today 7:20)
12:89 CRLF UTF-8 4 spaces Python 3.7 (mp2)

```python
import ...

data = {'ID': ['000001', '000002', '000003', '000004', '000005', '000006', '000007'],
        'name': ['黎明', '赵怡春', '张富平', '白丽', '牛玉德', '姚华', '李南'],
        'gender': [True, False, True, False, True, False, True],
        'age': [16, 20, 18, 18, 17, 18, 16],
        'height': [1.88, 1.78, 1.81, 1.86, 1.74, 1.75, 1.76]
        }
frame = pd.DataFrame(data)
print(frame['height'].groupby(frame['gender']).apply(lambda arr: arr.max() - arr.min()))
```

```
C:\Users\leeshuqing\PycharmProjects\mp2\venv\Scripts\python.exe C:/Users/leeshuqing/PycharmProjects/mp2/Exec.py
gender
False    0.11
True     0.14
Name: height, dtype: float64

Process finished with exit code 0
```

一次不学多，下次再学