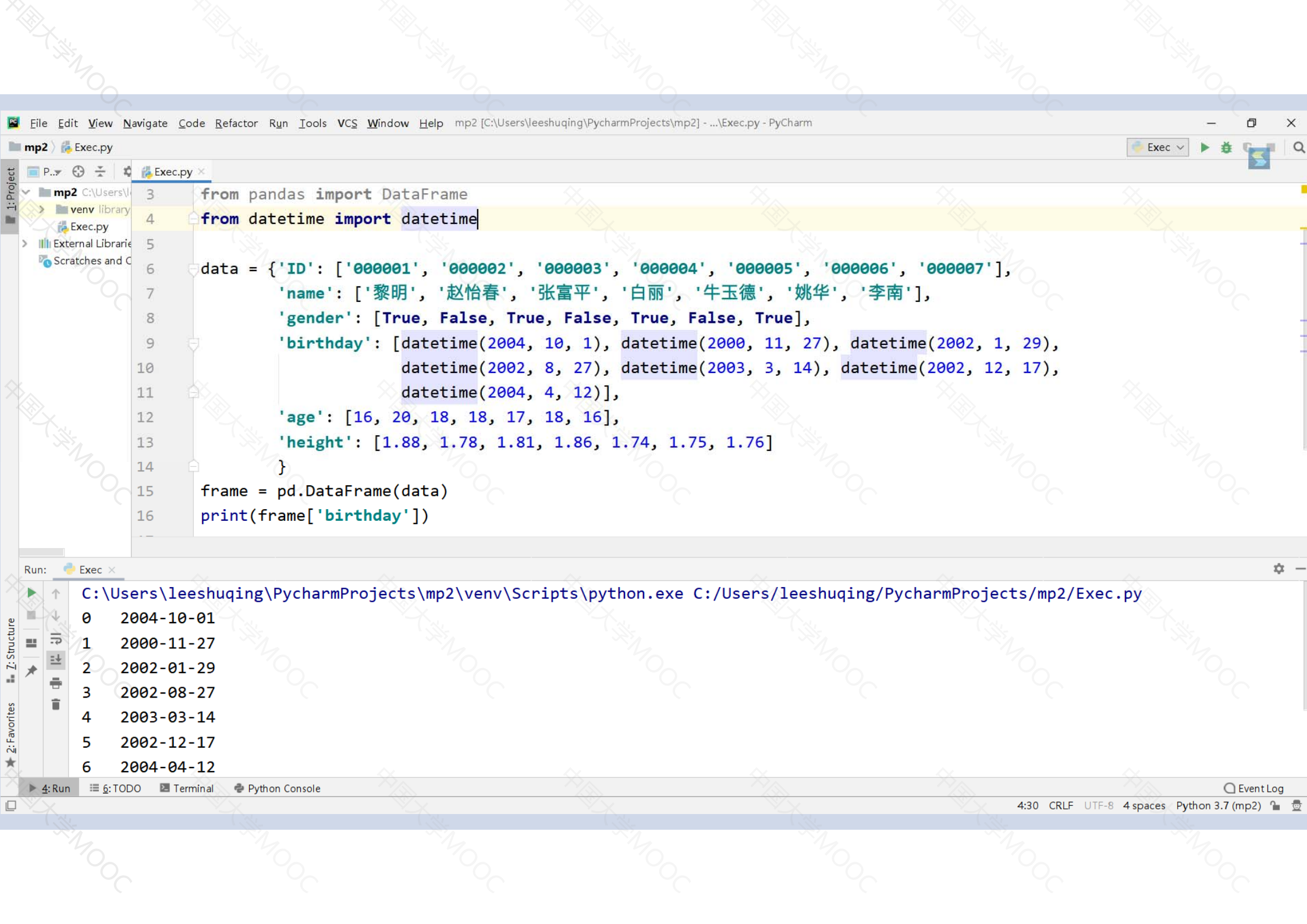


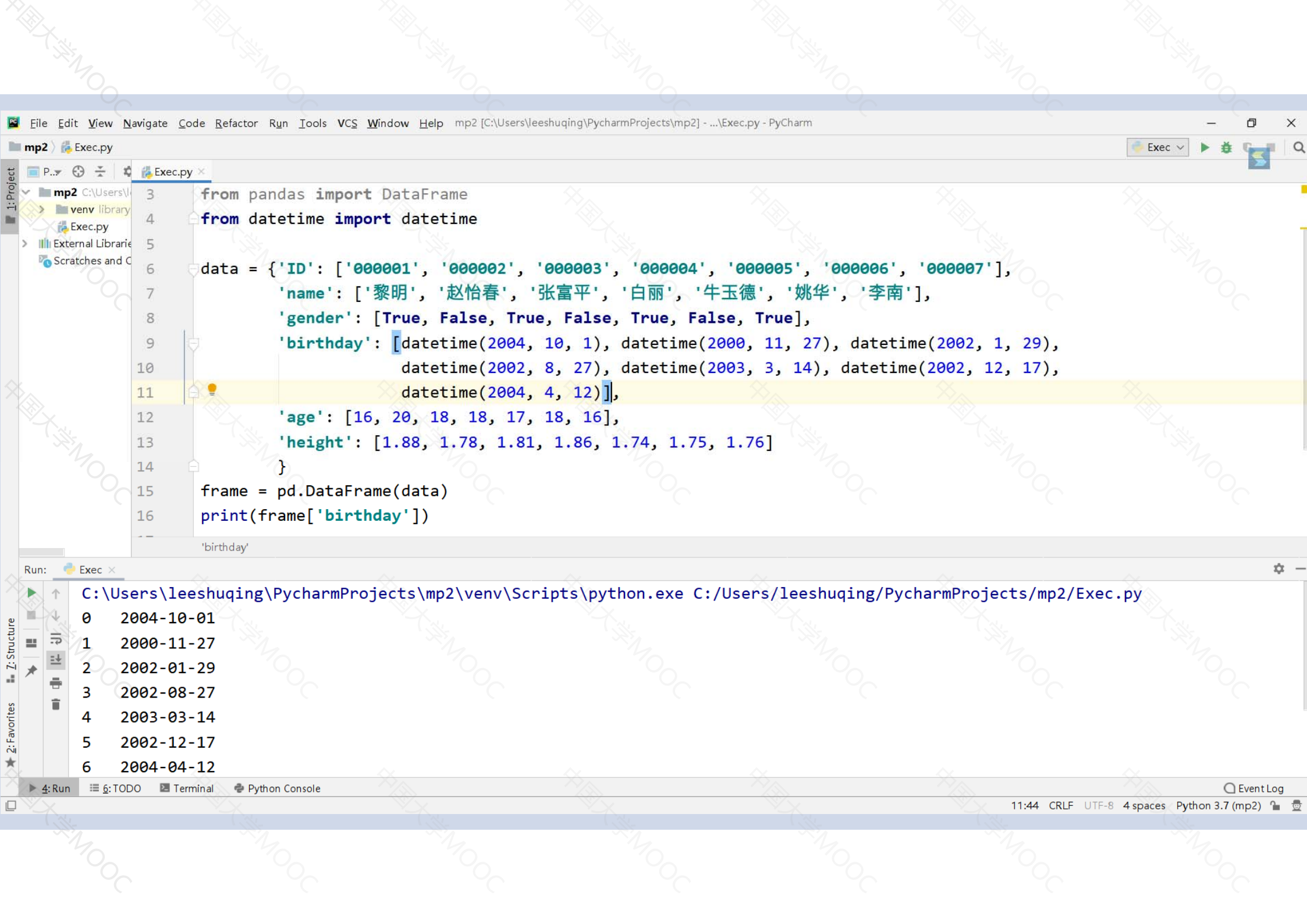
Python大数据分析

一、时间列的基本操作



```
3 from pandas import DataFrame
4 from datetime import datetime
5
6 data = {'ID': ['000001', '000002', '000003', '000004', '000005', '000006', '000007'],
7         'name': ['黎明', '赵怡春', '张富平', '白丽', '牛玉德', '姚华', '李南'],
8         'gender': [True, False, True, False, True, False, True],
9         'birthday': [datetime(2004, 10, 1), datetime(2000, 11, 27), datetime(2002, 1, 29),
10                     datetime(2002, 8, 27), datetime(2003, 3, 14), datetime(2002, 12, 17),
11                     datetime(2004, 4, 12)],
12         'age': [16, 20, 18, 18, 17, 18, 16],
13         'height': [1.88, 1.78, 1.81, 1.86, 1.74, 1.75, 1.76]
14     }
15 frame = pd.DataFrame(data)
16 print(frame['birthday'])
```

```
C:\Users\leeshuqing\PycharmProjects\mp2\venv\Scripts\python.exe C:/Users/leeshuqing/PycharmProjects/mp2/Exec.py
0    2004-10-01
1    2000-11-27
2    2002-01-29
3    2002-08-27
4    2003-03-14
5    2002-12-17
6    2004-04-12
```



```
3 from pandas import DataFrame
4 from datetime import datetime
5
6 data = {'ID': ['000001', '000002', '000003', '000004', '000005', '000006', '000007'],
7         'name': ['黎明', '赵怡春', '张富平', '白丽', '牛玉德', '姚华', '李南'],
8         'gender': [True, False, True, False, True, False, True],
9         'birthday': [datetime(2004, 10, 1), datetime(2000, 11, 27), datetime(2002, 1, 29),
10                     datetime(2002, 8, 27), datetime(2003, 3, 14), datetime(2002, 12, 17),
11                     datetime(2004, 4, 12)],
12         'age': [16, 20, 18, 18, 17, 18, 16],
13         'height': [1.88, 1.78, 1.81, 1.86, 1.74, 1.75, 1.76]
14     }
15 frame = pd.DataFrame(data)
16 print(frame['birthday'])
```

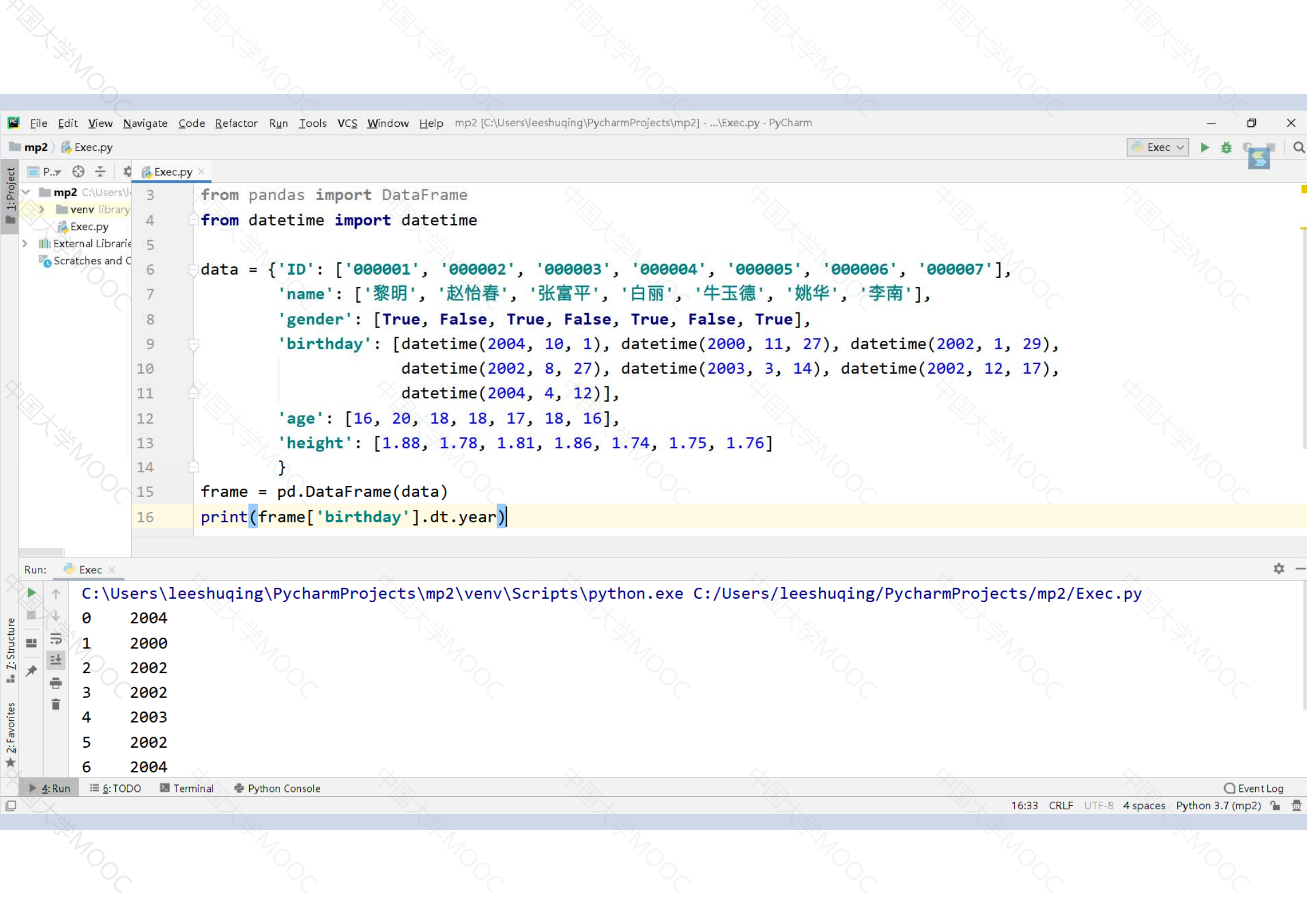
'birthday'

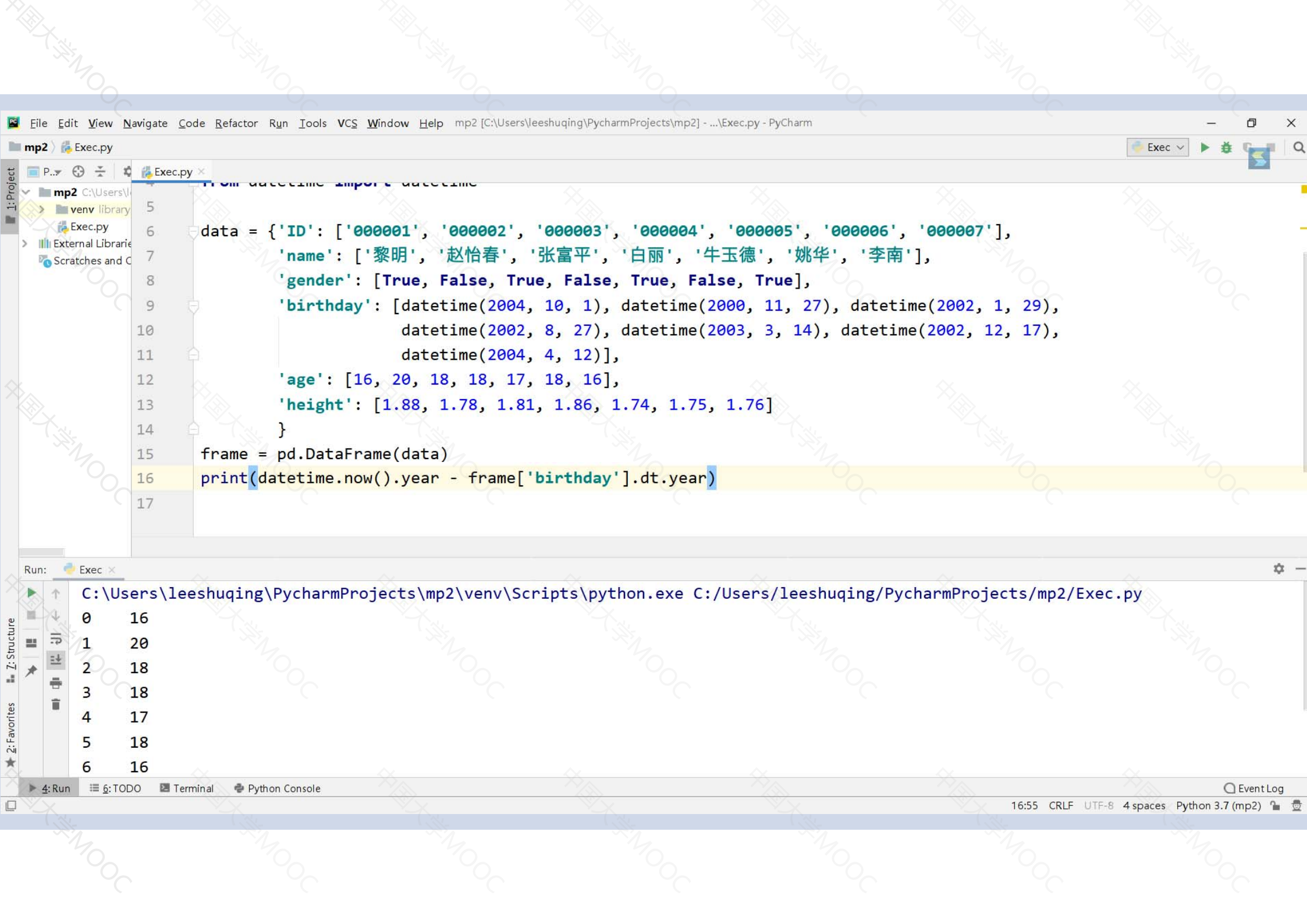
C:\Users\leeshuqing\PycharmProjects\mp2\venv\Scripts\python.exe C:/Users/leeshuqing/PycharmProjects/mp2/Exec.py

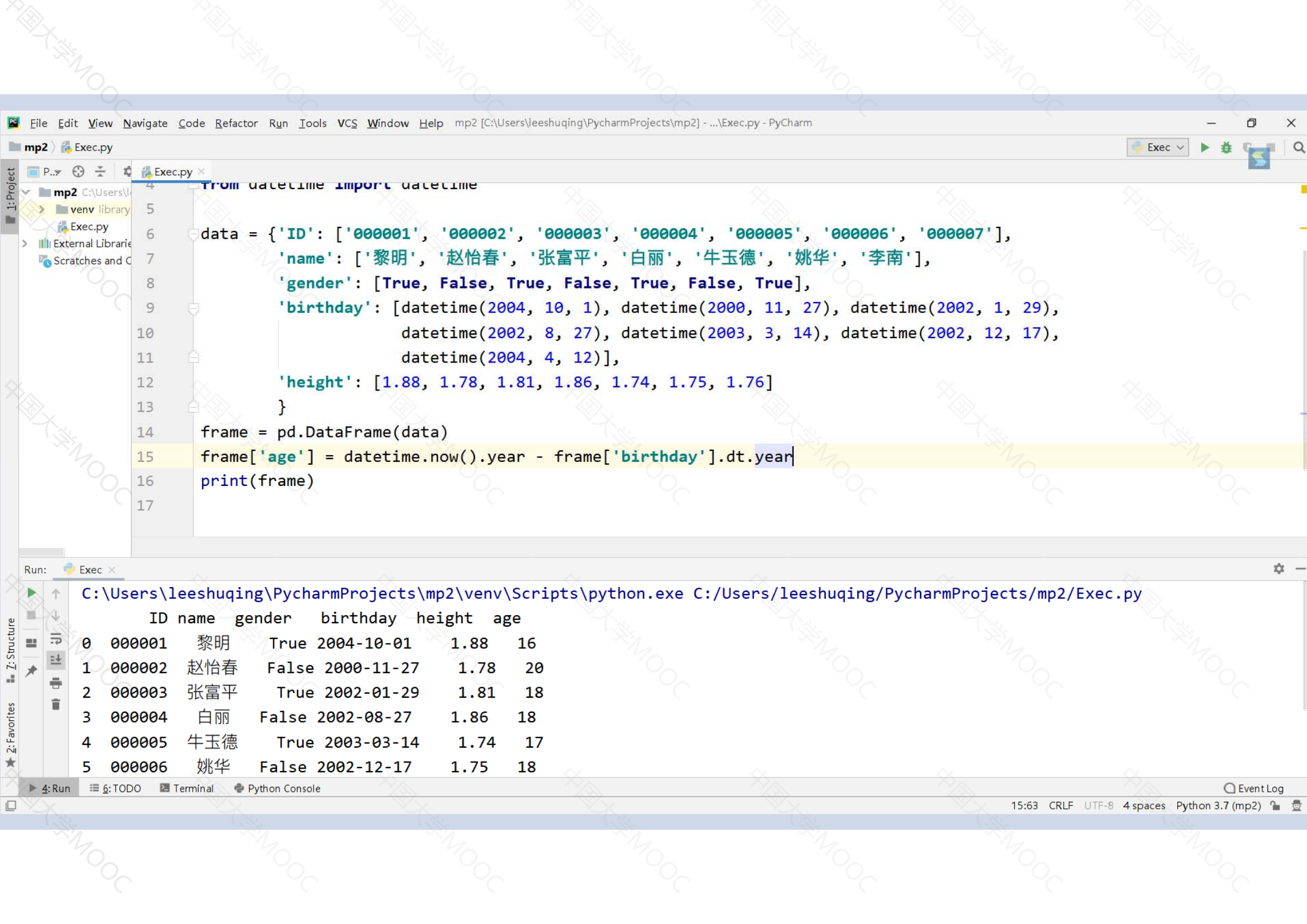
```
0  2004-10-01
1  2000-11-27
2  2002-01-29
3  2002-08-27
4  2003-03-14
5  2002-12-17
6  2004-04-12
```

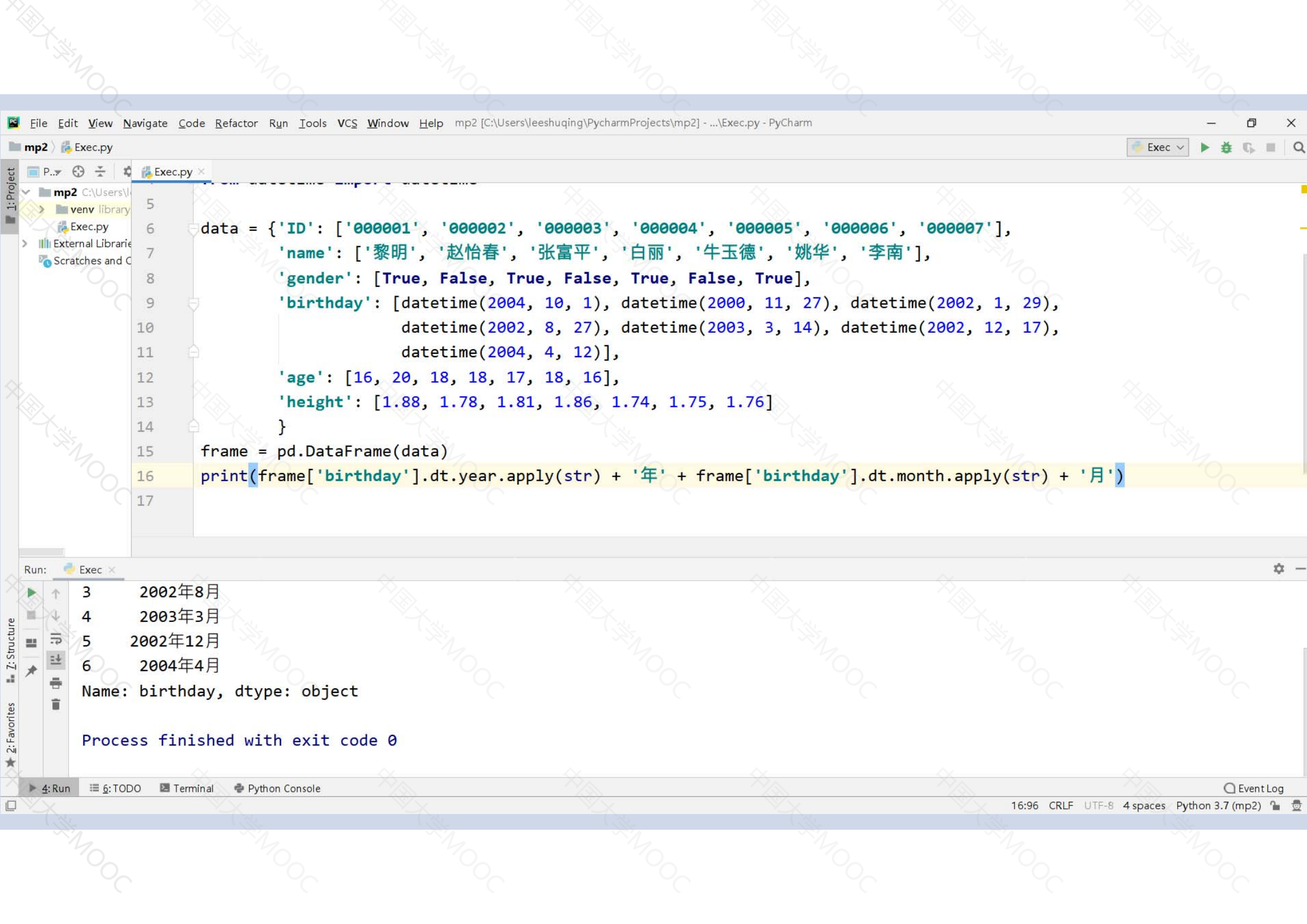
Event Log

11:44 CRLF UTF-8 4 spaces Python 3.7 (mp2)





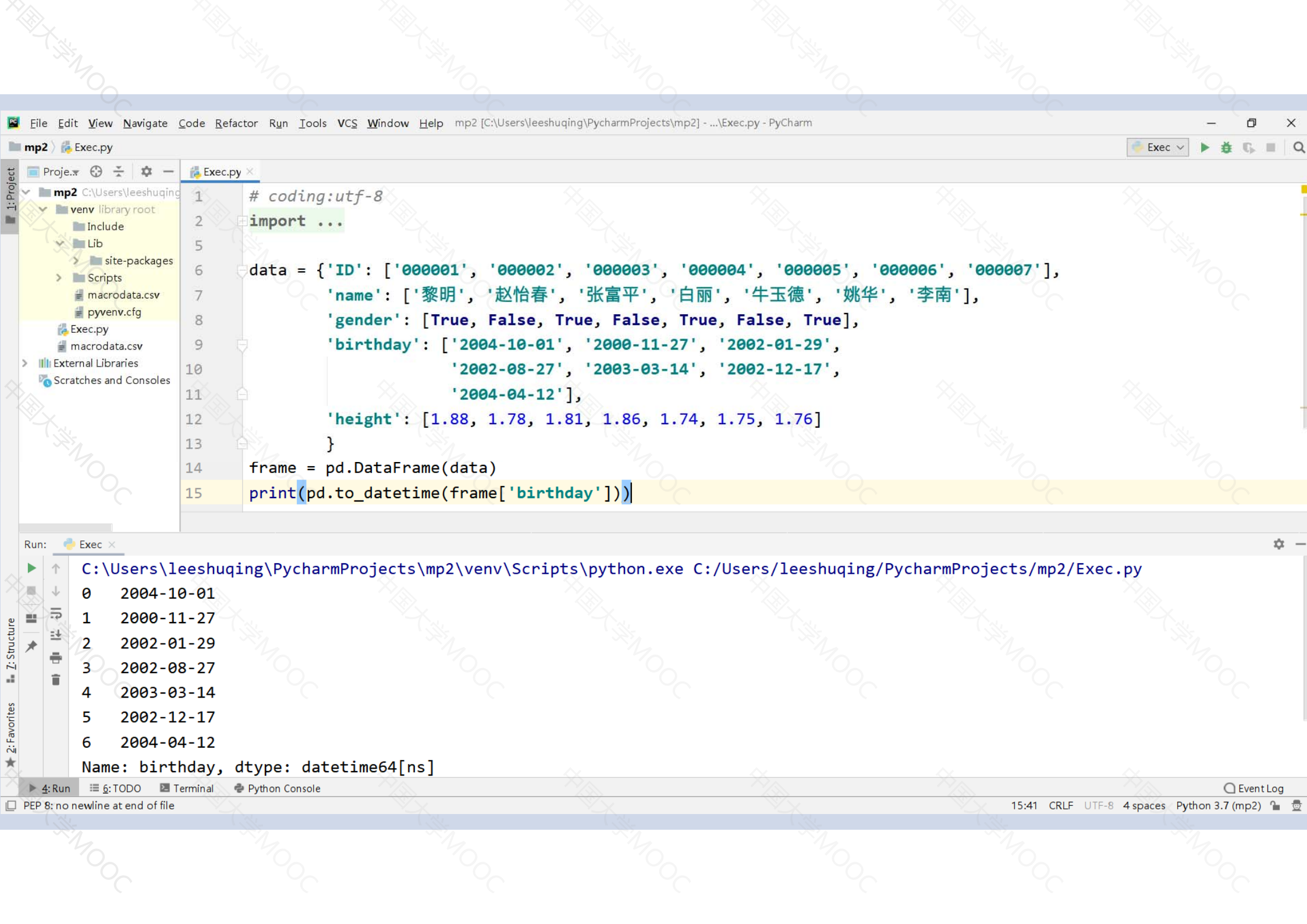




```
5
6 data = {'ID': ['000001', '000002', '000003', '000004', '000005', '000006', '000007'],
7         'name': ['黎明', '赵怡春', '张富平', '白丽', '牛玉德', '姚华', '李南'],
8         'gender': [True, False, True, False, True, False, True],
9         'birthday': [datetime(2004, 10, 1), datetime(2000, 11, 27), datetime(2002, 1, 29),
10                     datetime(2002, 8, 27), datetime(2003, 3, 14), datetime(2002, 12, 17),
11                     datetime(2004, 4, 12)],
12         'age': [16, 20, 18, 18, 17, 18, 16],
13         'height': [1.88, 1.78, 1.81, 1.86, 1.74, 1.75, 1.76]
14     }
15 frame = pd.DataFrame(data)
16 print(frame['birthday'].dt.year.apply(str) + '年' + frame['birthday'].dt.month.apply(str) + '月')
17
```

```
3    2002年8月
4    2003年3月
5    2002年12月
6    2004年4月
Name: birthday, dtype: object

Process finished with exit code 0
```

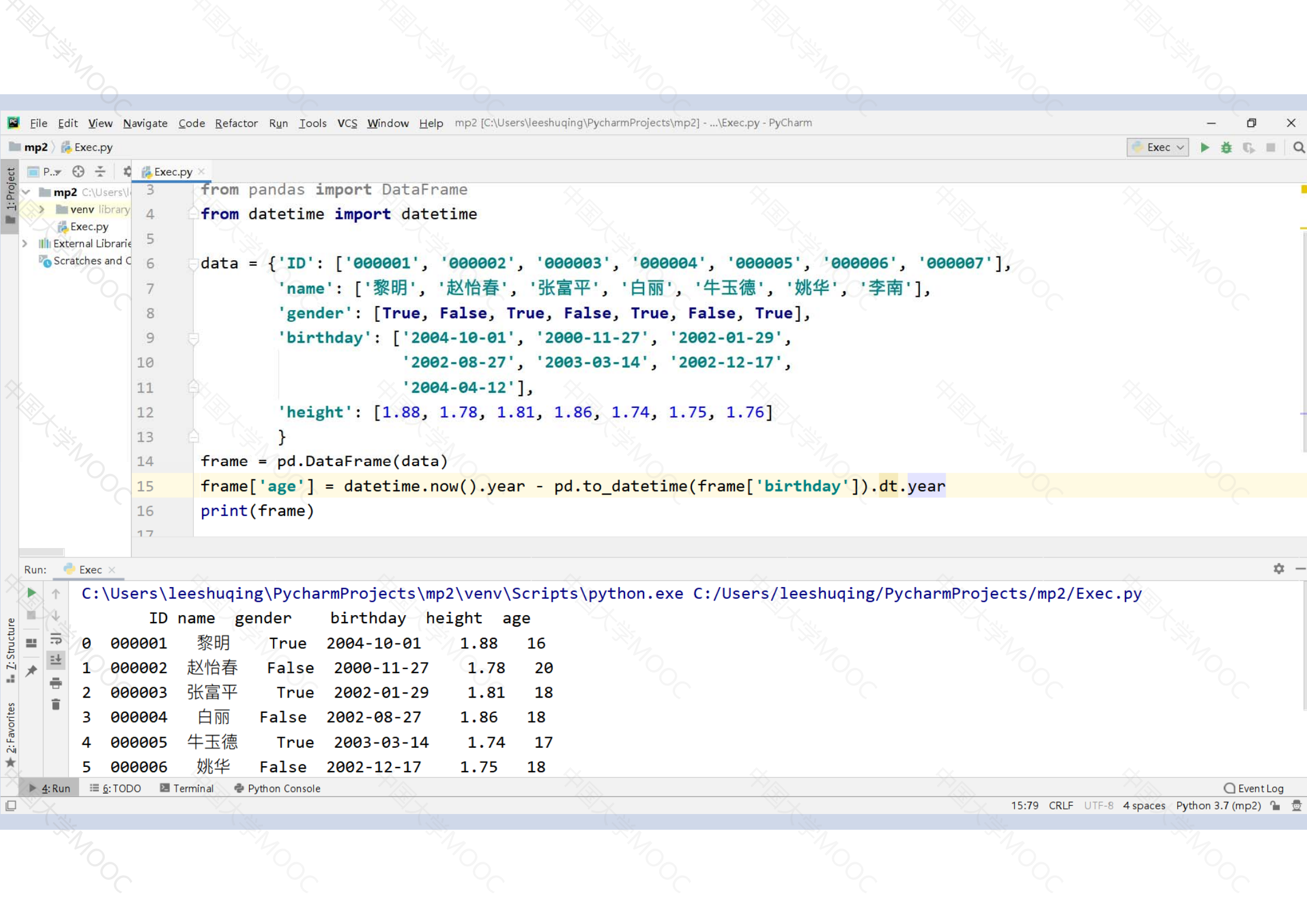



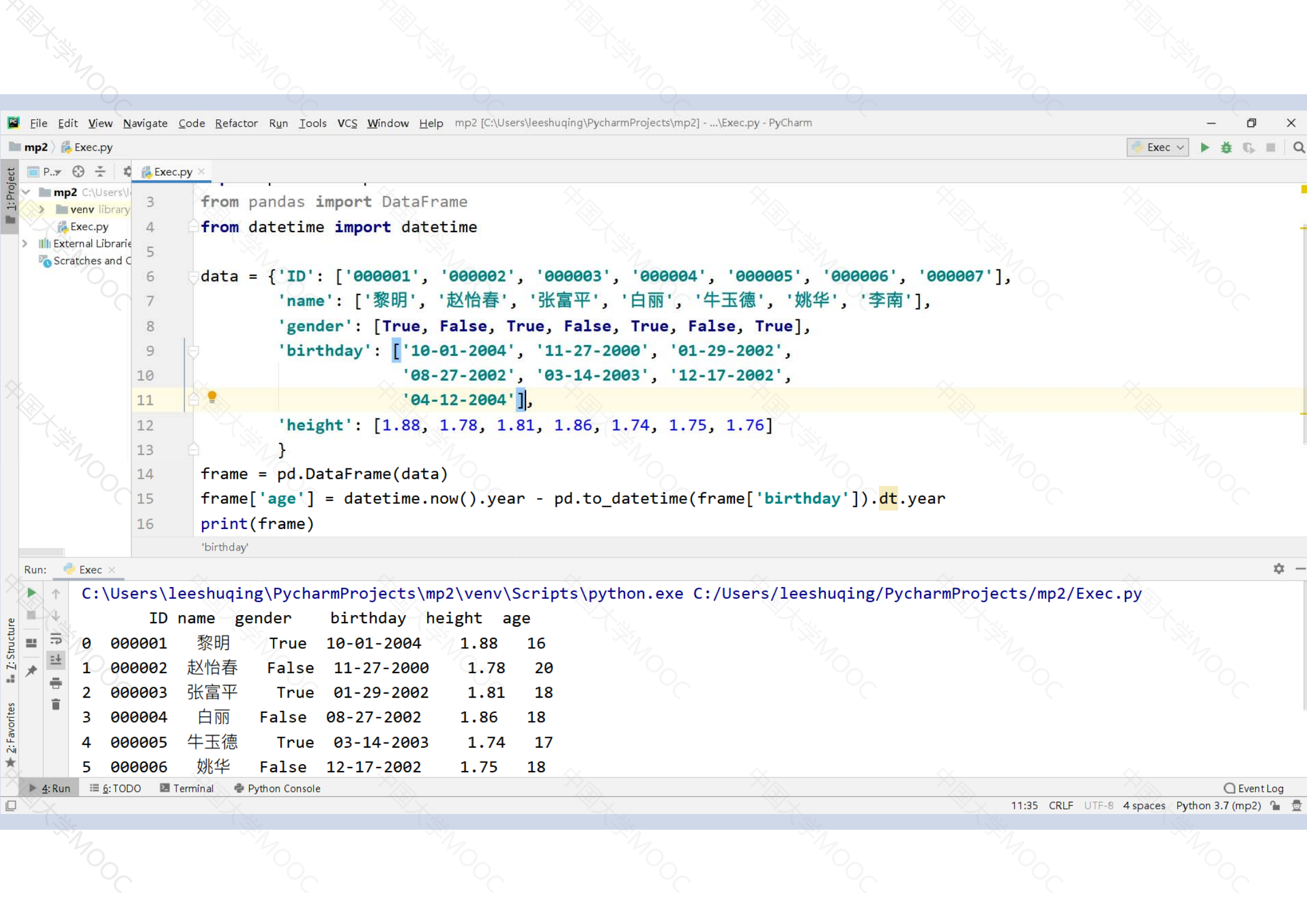
```
# coding:utf-8
import ...

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],
        'birthday': ['2004-10-01', '2000-11-27', '2002-01-29',
                     '2002-08-27', '2003-03-14', '2002-12-17',
                     '2004-04-12'],
        'height': [1.88, 1.78, 1.81, 1.86, 1.74, 1.75, 1.76]}

frame = pd.DataFrame(data)
print(pd.to_datetime(frame['birthday']))
```

```
C:\Users\leeshuqing\PycharmProjects\mp2\venv\Scripts\python.exe C:/Users/leeshuqing/PycharmProjects/mp2/Exec.py
0    2004-10-01
1    2000-11-27
2    2002-01-29
3    2002-08-27
4    2003-03-14
5    2002-12-17
6    2004-04-12
Name: birthday, dtype: datetime64[ns]
```



```
3 from pandas import DataFrame
4 from datetime import datetime
5
6 data = {'ID': ['000001', '000002', '000003', '000004', '000005', '000006', '000007'],
7         'name': ['黎明', '赵怡春', '张富平', '白丽', '牛玉德', '姚华', '李南'],
8         'gender': [True, False, True, False, True, False, True],
9         'birthday': ['10-01-2004', '11-27-2000', '01-29-2002',
10                    '08-27-2002', '03-14-2003', '12-17-2002',
11                    '04-12-2004'],
12         'height': [1.88, 1.78, 1.81, 1.86, 1.74, 1.75, 1.76]
13     }
14 frame = pd.DataFrame(data)
15 frame['age'] = datetime.now().year - pd.to_datetime(frame['birthday']).dt.year
16 print(frame)
```

'birthday'

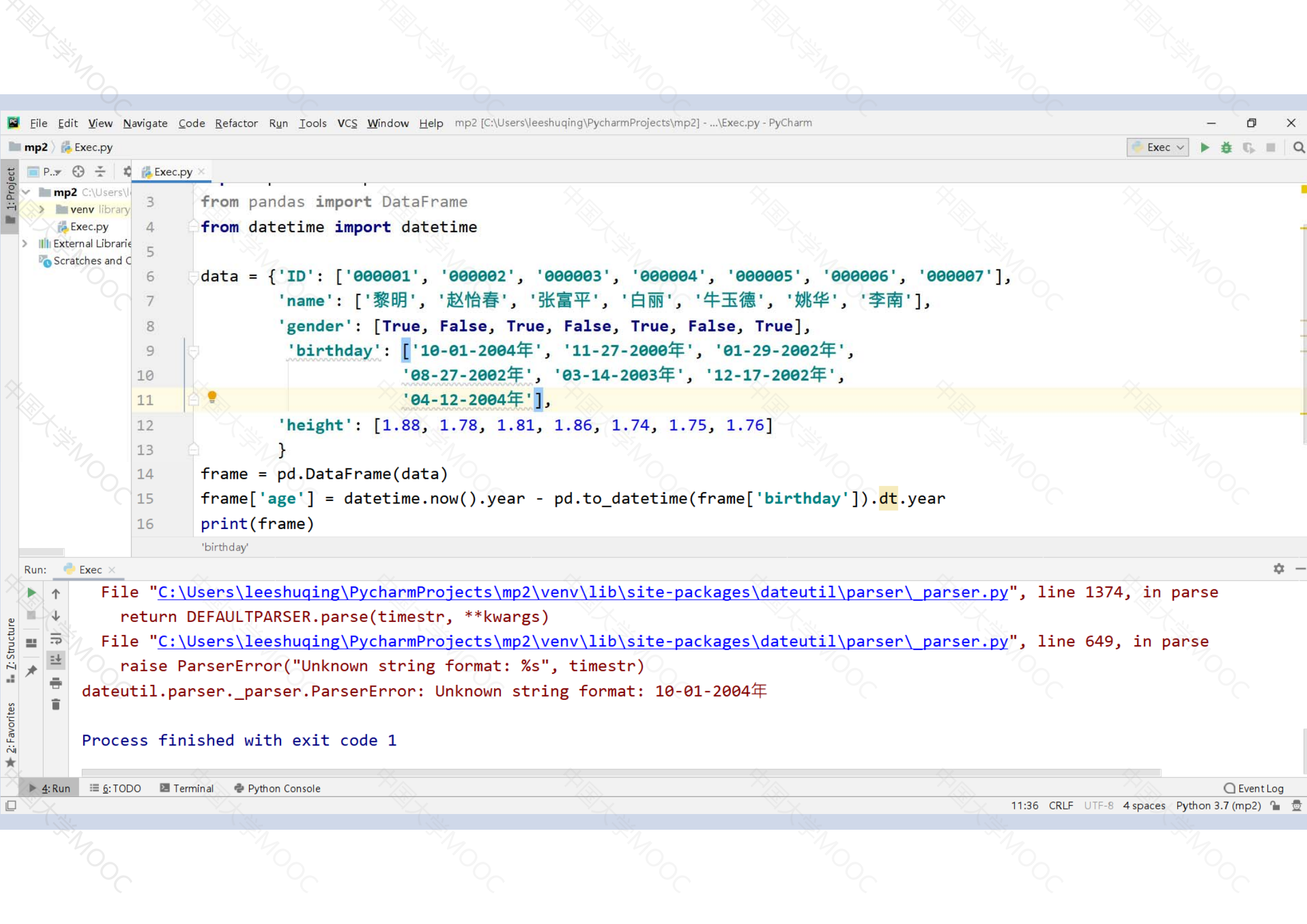
Run: Exec

C:\Users\leeshuqing\PycharmProjects\mp2\venv\Scripts\python.exe C:/Users/leeshuqing/PycharmProjects/mp2/Exec.py

	ID	name	gender	birthday	height	age
0	000001	黎明	True	10-01-2004	1.88	16
1	000002	赵怡春	False	11-27-2000	1.78	20
2	000003	张富平	True	01-29-2002	1.81	18
3	000004	白丽	False	08-27-2002	1.86	18
4	000005	牛玉德	True	03-14-2003	1.74	17
5	000006	姚华	False	12-17-2002	1.75	18

Event Log

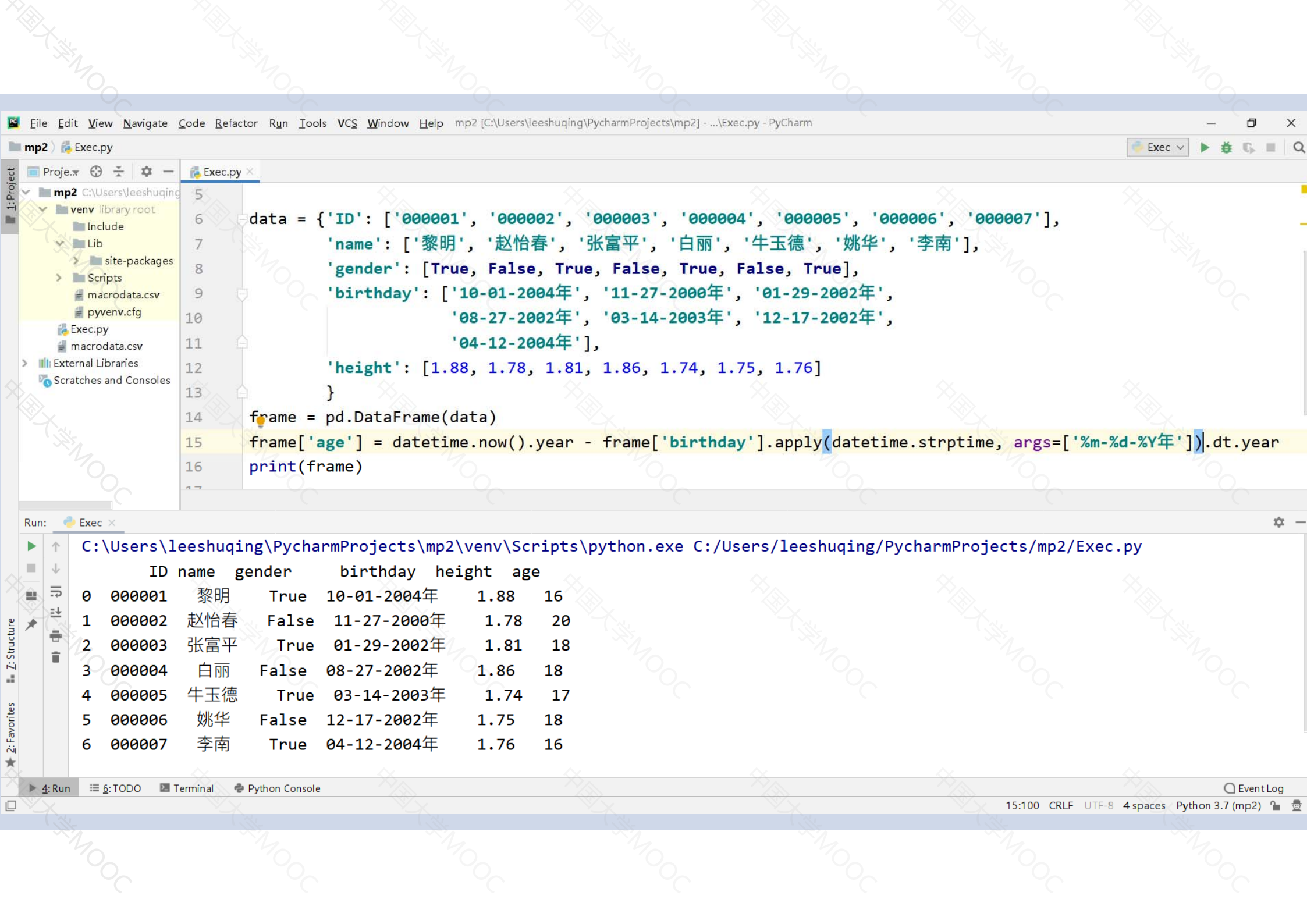
11:35 CRLF UTF-8 4 spaces Python 3.7 (mp2)

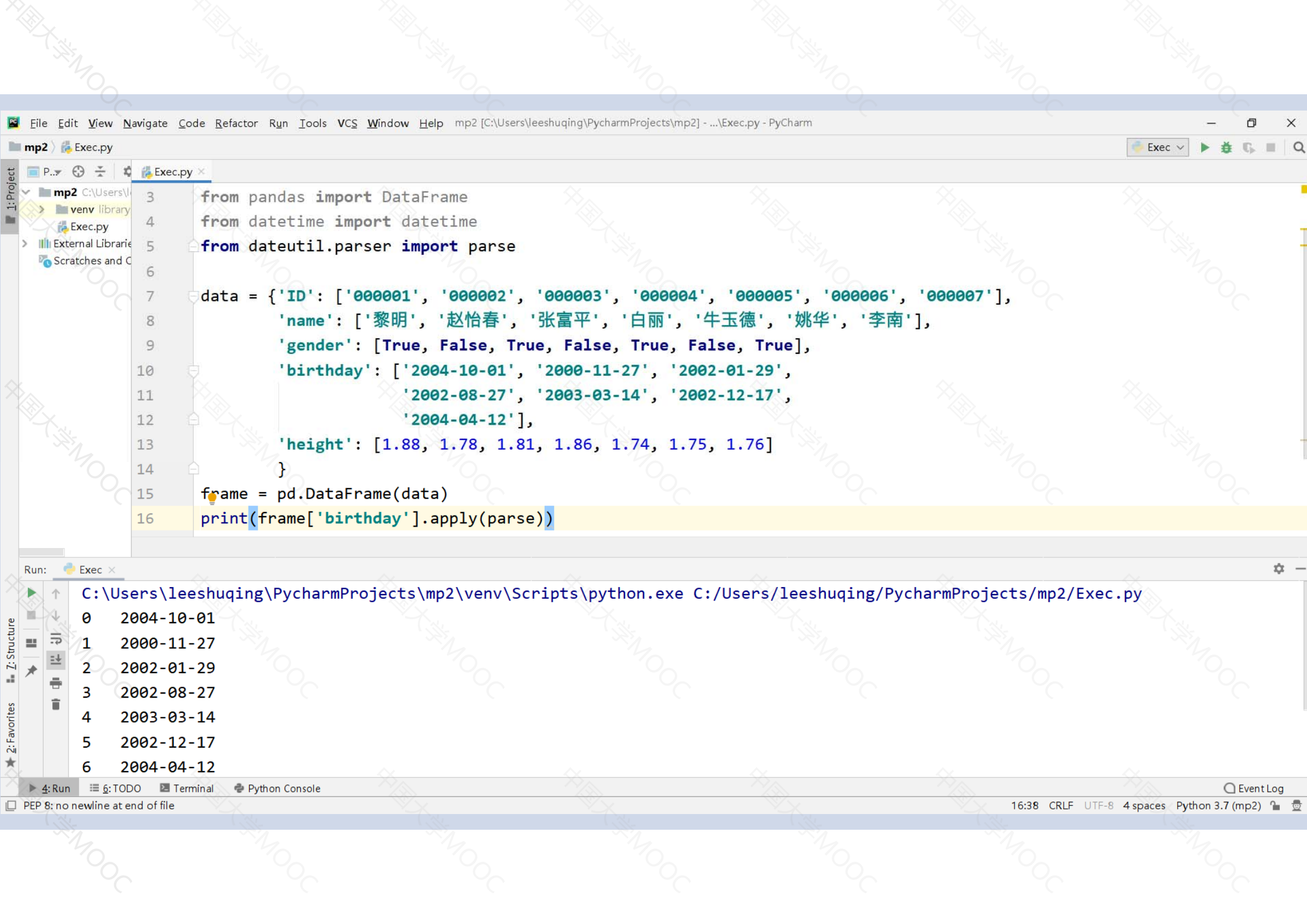


```
3 from pandas import DataFrame
4 from datetime import datetime
5
6 data = {'ID': ['000001', '000002', '000003', '000004', '000005', '000006', '000007'],
7         'name': ['黎明', '赵怡春', '张富平', '白丽', '牛玉德', '姚华', '李南'],
8         'gender': [True, False, True, False, True, False, True],
9         'birthday': ['10-01-2004年', '11-27-2000年', '01-29-2002年',
10                    '08-27-2002年', '03-14-2003年', '12-17-2002年',
11                    '04-12-2004年'],
12         'height': [1.88, 1.78, 1.81, 1.86, 1.74, 1.75, 1.76]}
13
14 frame = pd.DataFrame(data)
15 frame['age'] = datetime.now().year - pd.to_datetime(frame['birthday']).dt.year
16 print(frame)
```

```
File "C:\Users\leeshuqing\PycharmProjects\mp2\venv\lib\site-packages\dateutil\parser\_parser.py", line 1374, in parse
    return DEFAULTPARSER.parse(timestr, **kwargs)
File "C:\Users\leeshuqing\PycharmProjects\mp2\venv\lib\site-packages\dateutil\parser\_parser.py", line 649, in parse
    raise ParserError("Unknown string format: %s", timestr)
dateutil.parser._parser.ParserError: Unknown string format: 10-01-2004年

Process finished with exit code 1
```

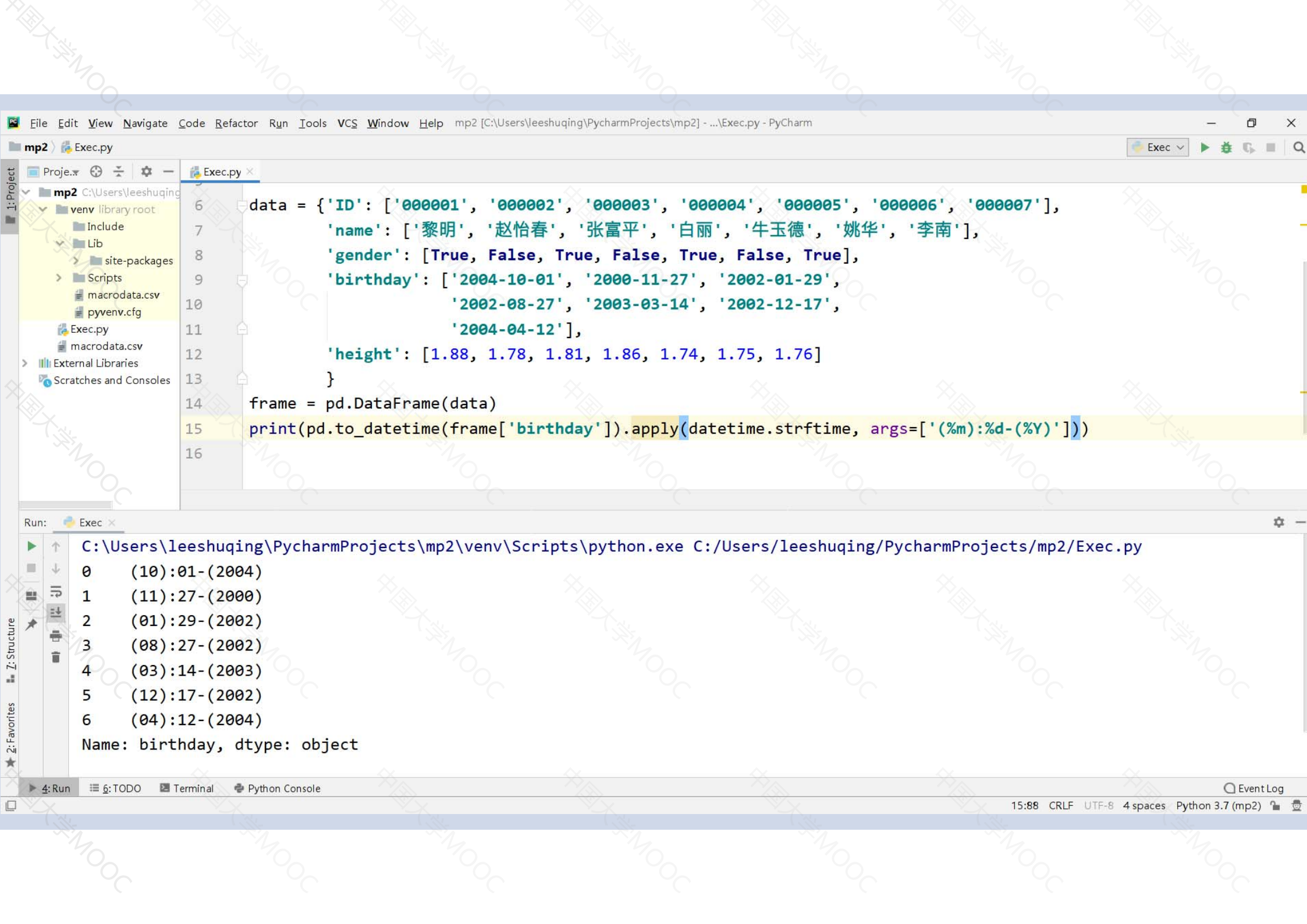


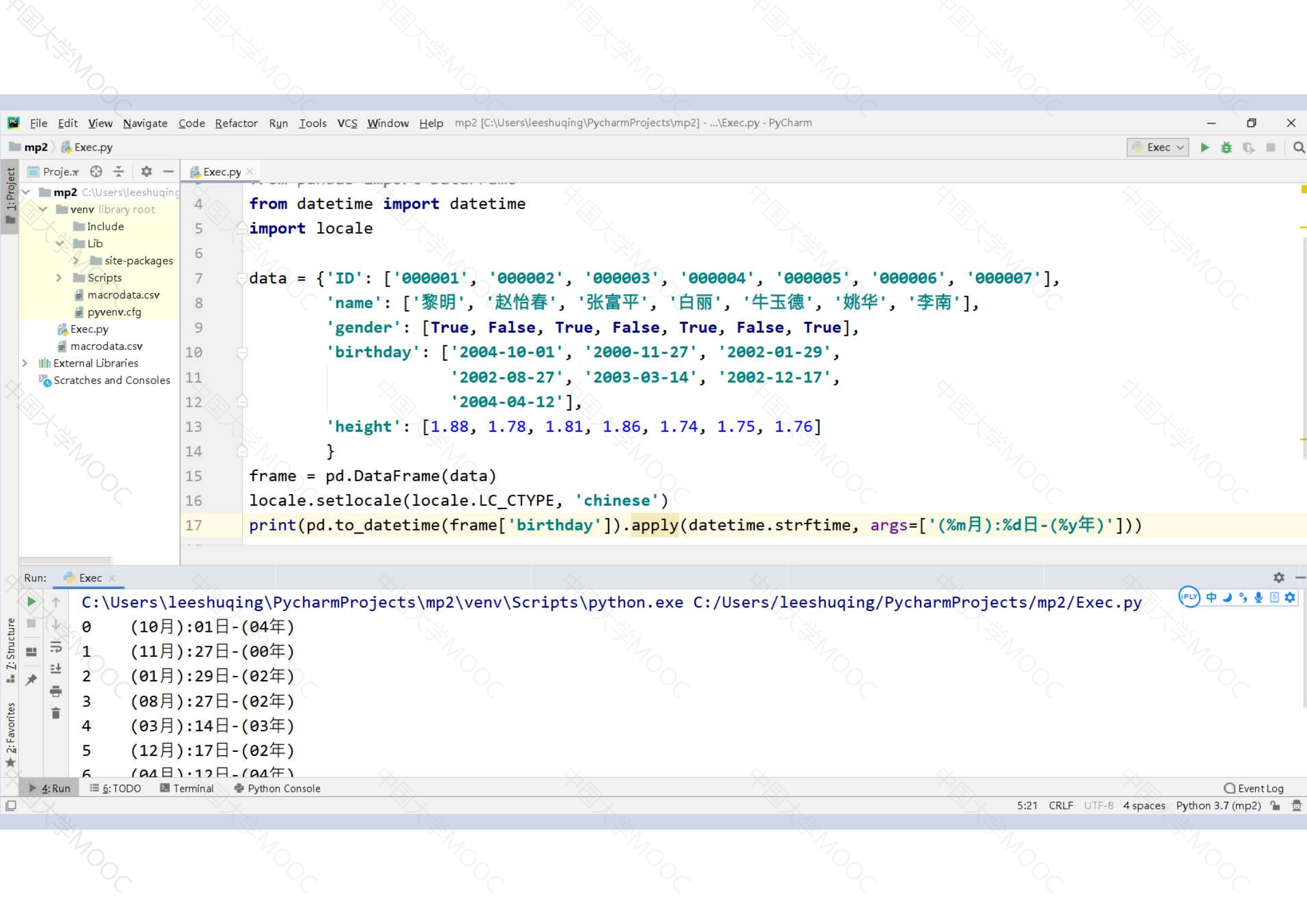


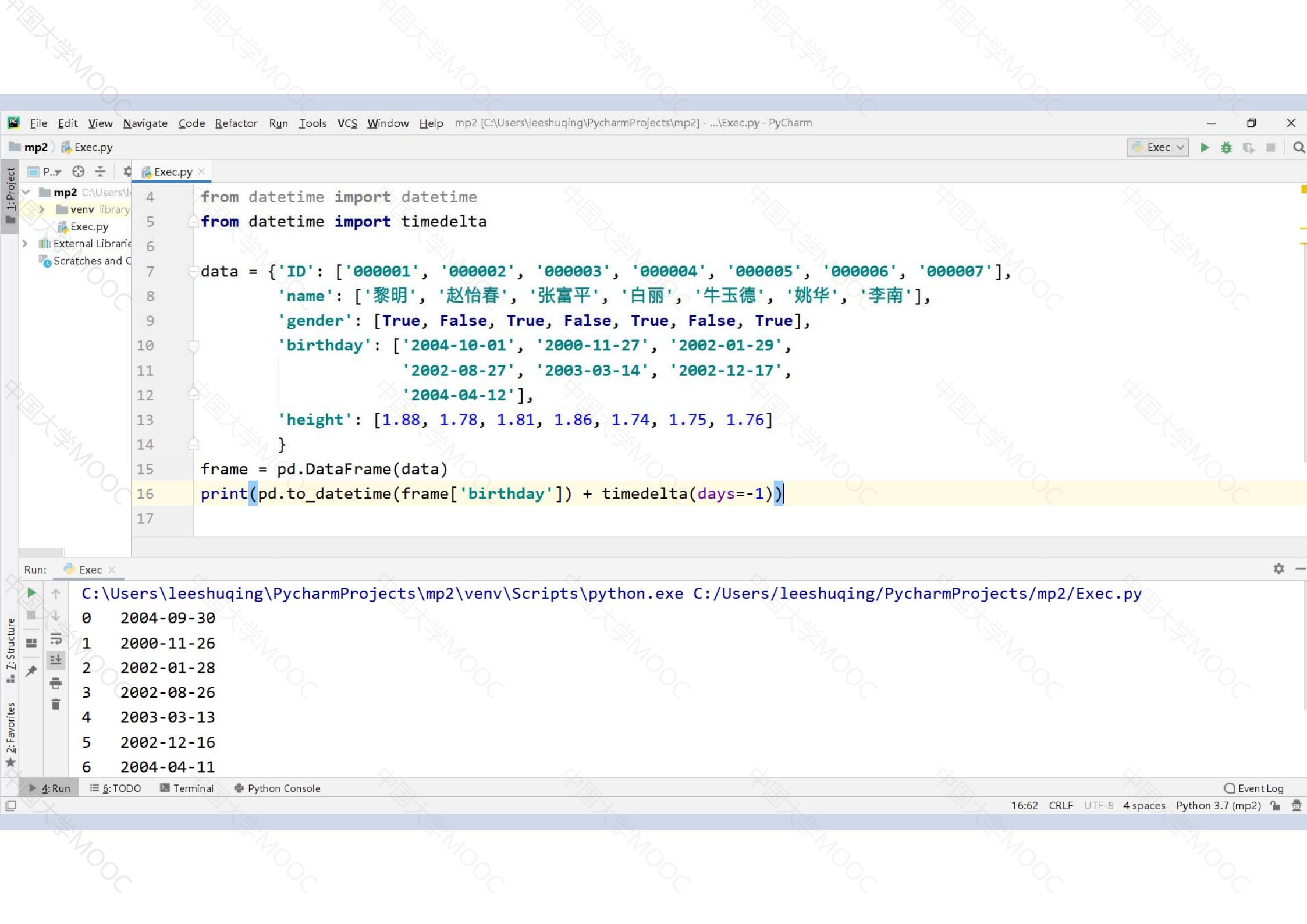
```
3 from pandas import DataFrame
4 from datetime import datetime
5 from dateutil.parser import parse
6
7 data = {'ID': ['000001', '000002', '000003', '000004', '000005', '000006', '000007'],
8         'name': ['黎明', '赵怡春', '张富平', '白丽', '牛玉德', '姚华', '李南'],
9         'gender': [True, False, True, False, True, False, True],
10        'birthday': ['2004-10-01', '2000-11-27', '2002-01-29',
11                    '2002-08-27', '2003-03-14', '2002-12-17',
12                    '2004-04-12'],
13        'height': [1.88, 1.78, 1.81, 1.86, 1.74, 1.75, 1.76]}
14
15 frame = pd.DataFrame(data)
16 print(frame['birthday'].apply(parse))
```

C:\Users\leeshuqing\PycharmProjects\mp2\venv\Scripts\python.exe C:/Users/leeshuqing/PycharmProjects/mp2/Exec.py

```
0  2004-10-01
1  2000-11-27
2  2002-01-29
3  2002-08-27
4  2003-03-14
5  2002-12-17
6  2004-04-12
```



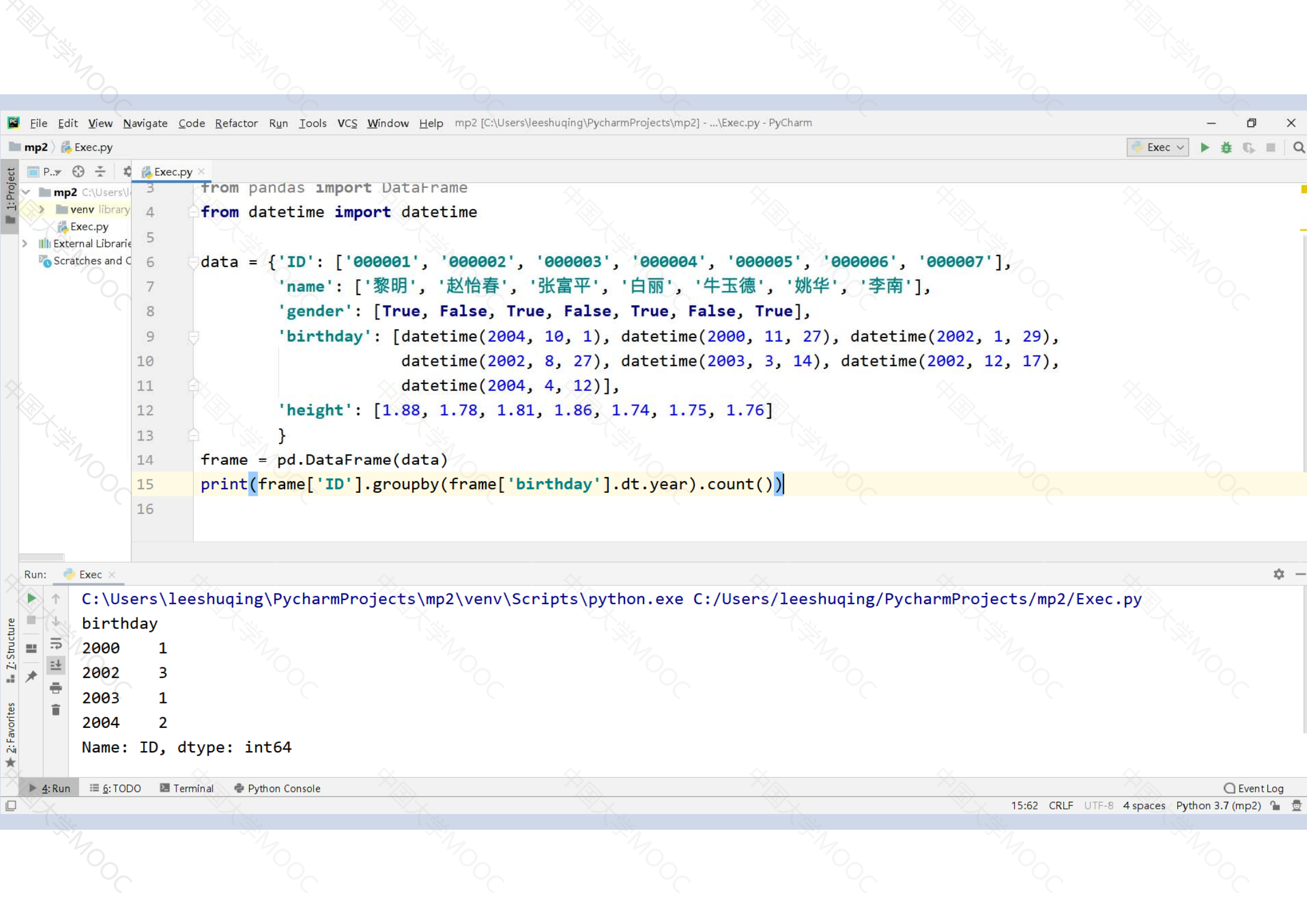


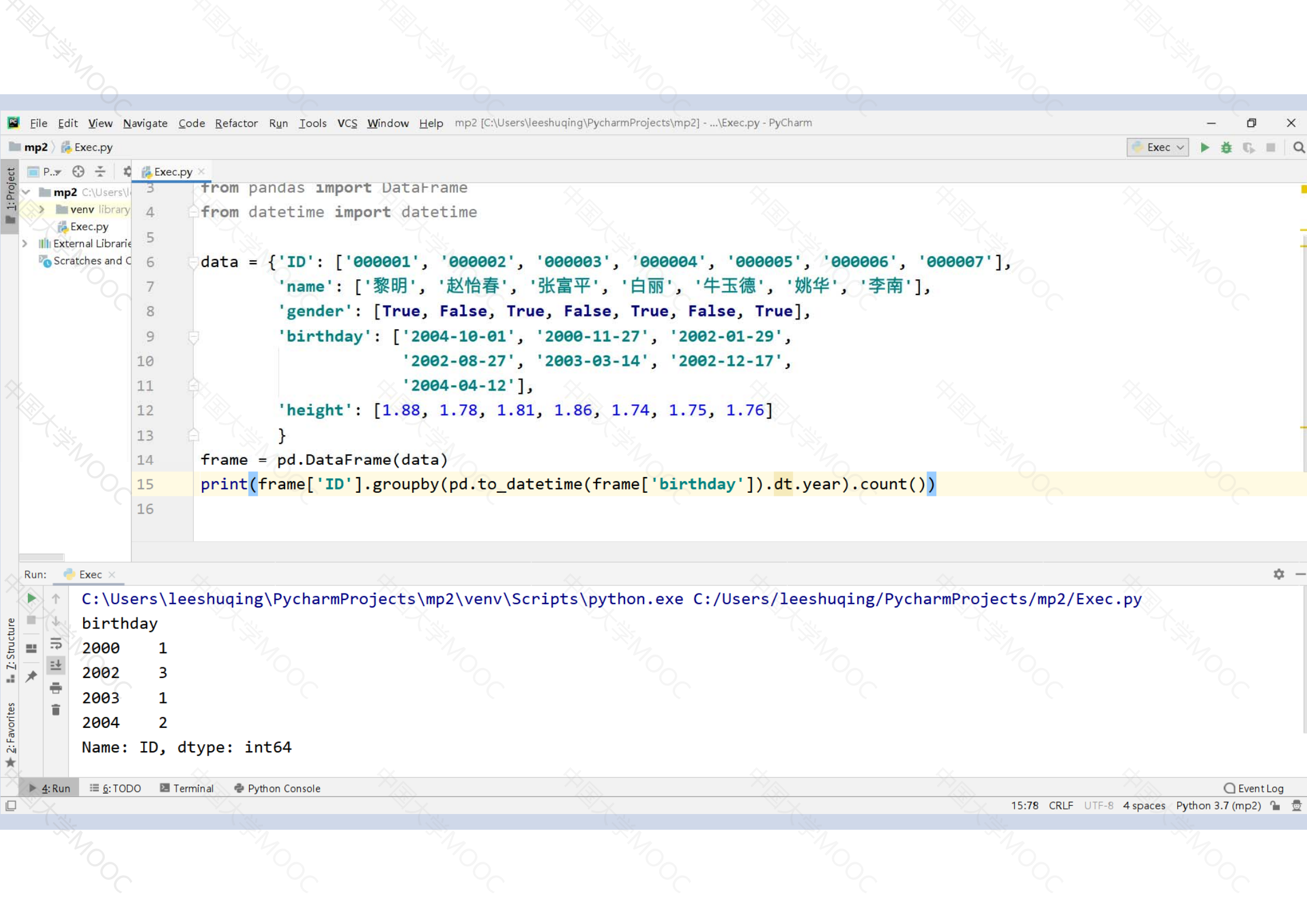


```
4 from datetime import datetime
5 from datetime import timedelta
6
7 data = {'ID': ['000001', '000002', '000003', '000004', '000005', '000006', '000007'],
8         'name': ['黎明', '赵怡春', '张富平', '白丽', '牛玉德', '姚华', '李南'],
9         'gender': [True, False, True, False, True, False, True],
10        'birthday': ['2004-10-01', '2000-11-27', '2002-01-29',
11                    '2002-08-27', '2003-03-14', '2002-12-17',
12                    '2004-04-12'],
13        'height': [1.88, 1.78, 1.81, 1.86, 1.74, 1.75, 1.76]}
14
15 frame = pd.DataFrame(data)
16 print(pd.to_datetime(frame['birthday']) + timedelta(days=-1))
17
```

C:\Users\leeshuqing\PycharmProjects\mp2\venv\Scripts\python.exe C:/Users/leeshuqing/PycharmProjects/mp2/Exec.py

```
0  2004-09-30
1  2000-11-26
2  2002-01-28
3  2002-08-26
4  2003-03-13
5  2002-12-16
6  2004-04-11
```



一次不学多，下次再学