自定义数据 - 合并

导入相关库

```
1  import pandas as pd
```

创建数据

```
raw_data_1 = {
        'subject_id': ['1', '2', '3', '4', '5'],
        'first_name': ['Alex', 'Amy', 'Allen', 'Alice',
'Ayoung'],
        'last_name': ['Anderson', 'Ackerman', 'Ali',
'Aoni', 'Atiches']}

raw_data_2 = {
        'subject_id': ['4', '5', '6', '7', '8'],
        'first_name': ['Billy', 'Brian', 'Bran', 'Bryce',
'Betty'],
        'last_name': ['Bonder', 'Black', 'Balwner',
'Brice', 'Btisan']}

raw_data_3 = {
        'subject_id': ['1', '2', '3', '4', '5', '7', '8',
'9', '10', '11'],
        'test_id': [51, 15, 15, 61, 16, 14, 15, 1, 61, 16]}
```

将上述的数据框分别命名为data1, data2, data3

```
data1 = pd.DataFrame(raw_data_1, columns = ['subject_id',
'first_name', 'last_name'])
data2 = pd.DataFrame(raw_data_2, columns = ['subject_id',
'first_name', 'last_name'])
data3 = pd.DataFrame(raw_data_3, columns = ['subject_id',
'test_id'])
```

将data1和data2两个数据框按照行的维度进行合并，命名为all_data

```
1  all_data = pd.concat([data1, data2], axis=0)
2  all_data
```

```
1  <tr style="text-align: right;">
2    <th></th>
3    <th>subject_id</th>
4    <th>first_name</th>
5    <th>last_name</th>
6  </tr>
```

```html
<tr>
  <th>0</th>
  <td>1</td>
  <td>Alex</td>
  <td>Anderson</td>
</tr>
<tr>
  <th>1</th>
  <td>2</td>
  <td>Amy</td>
  <td>Ackerman</td>
</tr>
<tr>
  <th>2</th>
  <td>3</td>
  <td>Allen</td>
  <td>Ali</td>
</tr>
<tr>
  <th>3</th>
  <td>4</td>
  <td>Alice</td>
  <td>Aoni</td>
</tr>
<tr>
  <th>4</th>
  <td>5</td>
  <td>Ayoung</td>
  <td>Atiches</td>
</tr>
<tr>
  <th>0</th>
  <td>4</td>
  <td>Billy</td>
```

```
35      <td>Bonder</td>
36   </tr>
37   <tr>
38      <th>1</th>
39      <td>5</td>
40      <td>Brian</td>
41      <td>Black</td>
42   </tr>
43   <tr>
44      <th>2</th>
45      <td>6</td>
46      <td>Bran</td>
47      <td>Balwner</td>
48   </tr>
49   <tr>
50      <th>3</th>
51      <td>7</td>
52      <td>Bryce</td>
53      <td>Brice</td>
54   </tr>
55   <tr>
56      <th>4</th>
57      <td>8</td>
58      <td>Betty</td>
59      <td>Btisan</td>
60   </tr>
```

将data1和data2两个数据框按照列的维度进行合并，命名为all_data_col

```python
all_data_col = pd.concat([data1, data2], axis=1)
all_data_col
```

```html
<tr style="text-align: right;">
  <th></th>
  <th>subject_id</th>
  <th>first_name</th>
  <th>last_name</th>
  <th>subject_id</th>
  <th>first_name</th>
  <th>last_name</th>
</tr>
```

```
<tr>
  <th>0</th>
  <td>1</td>
  <td>Alex</td>
  <td>Anderson</td>
  <td>4</td>
  <td>Billy</td>
  <td>Bonder</td>
</tr>
<tr>
  <th>1</th>
  <td>2</td>
  <td>Amy</td>
  <td>Ackerman</td>
  <td>5</td>
  <td>Brian</td>
  <td>Black</td>
</tr>
<tr>
  <th>2</th>
  <td>3</td>
  <td>Allen</td>
  <td>Ali</td>
  <td>6</td>
  <td>Bran</td>
  <td>Balwner</td>
</tr>
<tr>
  <th>3</th>
  <td>4</td>
  <td>Alice</td>
  <td>Aoni</td>
  <td>7</td>
  <td>Bryce</td>
```

```
35      <td>Brice</td>
36  </tr>
37  <tr>
38      <th>4</th>
39      <td>5</td>
40      <td>Ayoung</td>
41      <td>Atiches</td>
42      <td>8</td>
43      <td>Betty</td>
44      <td>Btisan</td>
45  </tr>
```

打印data3

```
1  data3
```

```
1  <tr style="text-align: right;">
2      <th></th>
3      <th>subject_id</th>
4      <th>test_id</th>
5  </tr>
```

```
<tr>
  <th>0</th>
  <td>1</td>
  <td>51</td>
</tr>
<tr>
  <th>1</th>
  <td>2</td>
  <td>15</td>
</tr>
<tr>
  <th>2</th>
  <td>3</td>
  <td>15</td>
</tr>
<tr>
  <th>3</th>
  <td>4</td>
  <td>61</td>
</tr>
<tr>
  <th>4</th>
  <td>5</td>
  <td>16</td>
</tr>
<tr>
  <th>5</th>
  <td>7</td>
  <td>14</td>
</tr>
<tr>
  <th>6</th>
  <td>8</td>
  <td>15</td>
```

```
35    </tr>
36    <tr>
37      <th>7</th>
38      <td>9</td>
39      <td>1</td>
40    </tr>
41    <tr>
42      <th>8</th>
43      <td>10</td>
44      <td>61</td>
45    </tr>
46    <tr>
47      <th>9</th>
48      <td>11</td>
49      <td>16</td>
50    </tr>
```

按照subject_id的值对all_data和data3作合并

```
1    pd.merge(all_data, data3, on='subject_id')
```

```html
<tr style="text-align: right;">
  <th></th>
  <th>subject_id</th>
  <th>first_name</th>
  <th>last_name</th>
  <th>test_id</th>
</tr>
```

```
<tr>
  <th>0</th>
  <td>1</td>
  <td>Alex</td>
  <td>Anderson</td>
  <td>51</td>
</tr>
<tr>
  <th>1</th>
  <td>2</td>
  <td>Amy</td>
  <td>Ackerman</td>
  <td>15</td>
</tr>
<tr>
  <th>2</th>
  <td>3</td>
  <td>Allen</td>
  <td>Ali</td>
  <td>15</td>
</tr>
<tr>
  <th>3</th>
  <td>4</td>
  <td>Alice</td>
  <td>Aoni</td>
  <td>61</td>
</tr>
<tr>
  <th>4</th>
  <td>4</td>
  <td>Billy</td>
  <td>Bonder</td>
  <td>61</td>
```

```
35  </tr>
36  <tr>
37    <th>5</th>
38    <td>5</td>
39    <td>Ayoung</td>
40    <td>Atiches</td>
41    <td>16</td>
42  </tr>
43  <tr>
44    <th>6</th>
45    <td>5</td>
46    <td>Brian</td>
47    <td>Black</td>
48    <td>16</td>
49  </tr>
50  <tr>
51    <th>7</th>
52    <td>7</td>
53    <td>Bryce</td>
54    <td>Brice</td>
55    <td>14</td>
56  </tr>
57  <tr>
58    <th>8</th>
59    <td>8</td>
60    <td>Betty</td>
61    <td>Btisan</td>
62    <td>15</td>
63  </tr>
```

对data1和data2按照subject_id作连接

```
pd.merge(data1, data2, on='subject_id')
```

```
<tr style="text-align: right;">
  <th></th>
  <th>subject_id</th>
  <th>first_name_x</th>
  <th>last_name_x</th>
  <th>first_name_y</th>
  <th>last_name_y</th>
</tr>
```

```
<tr>
  <th>0</th>
  <td>4</td>
  <td>Alice</td>
  <td>Aoni</td>
  <td>Billy</td>
  <td>Bonder</td>
</tr>
<tr>
  <th>1</th>
  <td>5</td>
  <td>Ayoung</td>
  <td>Atiches</td>
  <td>Brian</td>
  <td>Black</td>
</tr>
```

找到 data1 和 data2 合并之后的所有匹配结果

```
1  pd.merge(data1, data2, on='subject_id', how='outer')
```

```html
1  <tr style="text-align: right;">
2    <th></th>
3    <th>subject_id</th>
4    <th>first_name_x</th>
5    <th>last_name_x</th>
6    <th>first_name_y</th>
7    <th>last_name_y</th>
8  </tr>
```

```html
<tr>
  <th>0</th>
  <td>1</td>
  <td>Alex</td>
  <td>Anderson</td>
  <td>NaN</td>
  <td>NaN</td>
</tr>
<tr>
  <th>1</th>
  <td>2</td>
  <td>Amy</td>
  <td>Ackerman</td>
  <td>NaN</td>
  <td>NaN</td>
</tr>
<tr>
  <th>2</th>
  <td>3</td>
  <td>Allen</td>
  <td>Ali</td>
  <td>NaN</td>
  <td>NaN</td>
</tr>
<tr>
  <th>3</th>
  <td>4</td>
  <td>Alice</td>
  <td>Aoni</td>
  <td>Billy</td>
  <td>Bonder</td>
</tr>
<tr>
  <th>4</th>
```

```
      <td>5</td>
      <td>Ayoung</td>
      <td>Atiches</td>
      <td>Brian</td>
      <td>Black</td>
</tr>
<tr>
      <th>5</th>
      <td>6</td>
      <td>NaN</td>
      <td>NaN</td>
      <td>Bran</td>
      <td>Balwner</td>
</tr>
<tr>
      <th>6</th>
      <td>7</td>
      <td>NaN</td>
      <td>NaN</td>
      <td>Bryce</td>
      <td>Brice</td>
</tr>
<tr>
      <th>7</th>
      <td>8</td>
      <td>NaN</td>
      <td>NaN</td>
      <td>Betty</td>
      <td>Btisan</td>
</tr>
```