#### Concatenation of NumPy Arrays

```
import numpy as np
x = [1, 2, 3]
y = [4, 5, 6]
z = [7, 8, 9]
np.concatenate([x, y, z])
array([1, 2, 3, 4, 5, 6, 7, 8, 9])
x = [[1, 2],
      [3, 4]]
np.concatenate([x, x], axis=1)
array([[1, 2, 1, 2],
        [3, 4, 3, 4]])
import pandas as pd
ser1 = pd.Series(['A', 'B', 'C'], index=[1, 2, 3])
ser2 = pd.Series(['D', 'E', 'F'], index=[4, 5, 6])
pd.concat([ser1, ser2])
1
      Α
2
      В
      C
3
4
      D
5
      Ε
6
      F
dtype: object
def make df(cols, ind):
    """Quickly make a DataFrame"""
    data = \{c: [str(c) + str(i) \text{ for } i \text{ in } ind]\}
              for c in cols}
    return pd.DataFrame(data, ind)
# example DataFrame
make df('ABC', range(3))
    Α
         В
             C
0
  Α0
        B0 C0
        B1 C1
1
  Α1
2 A2
        B2
            C2
df1 = make_df('AB', [1, 2])
df2 = make_df('AB', [3, 4])
display('df1', 'df2', 'pd.concat([df1, df2])')
df1
         В
    Α
1 A1 B1
```

```
2 A2 B2
df2
    Α
         В
   А3
        В3
   Α4
        B4
pd.concat([df1, df2])
    Α
         В
   Α1
1
        В1
2
   A2
        B2
3
   А3
        В3
   Α4
        В4
df3 = make_df('AB', [0, 1])
df4 = make_df('CD', [0, 1])
display('df3', 'df4', "pd.concat([df3, df4], axis=1)")
df3
    Α
         В
   Α0
        B0
1 A1
        B1
df4
    C
         D
   C0
0
        D0
  C1
        D1
pd.concat([df3, df4], axis=1)
              C
    Α
         В
   Α0
        B0
             C0
                 D0
  Α1
             C1
1
        В1
                 D1
```

#### **Duplicate indices**

One important difference between np.concatenate and pd.concat is that Pandas concatenation preserves indices, even if the result will have duplicate indices! Consider this simple example:

```
x = make_df('AB', [0, 1])
y = make_df('AB', [2, 3])
y.index = x.index  # make duplicate indices!
display('x', 'y', 'pd.concat([x, y])')

x
          A      B
0     A0     B0
1     A1     B1

y
          A      B
```

```
0 A2
       B2
1 A3 B3
pd.concat([x, y])
        В
   Α0
       B0
1
  Α1
       B1
  Α2
       B2
  А3
       В3
display('x', 'y', 'pd.concat([x, y], ignore_index=True)')
Х
        В
    Α
0
  Α0
       B0
  Α1
       В1
1
У
        В
    Α
  Α2
0
       B2
  А3
       В3
pd.concat([x, y], ignore_index=True)
  Α0
0
       B0
  Α1
1
       В1
2
  Α2
       B2
  А3
       В3
Adding MultiIndex keys
display('x', 'y', "pd.concat([x, y], keys=['x', 'y'])")
Χ
    Α
        В
0
       B0
   Α0
1
  Α1
       В1
У
    Α
        В
0
  Α2
       B2
  А3
       В3
pd.concat([x, y], keys=['x', 'y'])
      Α
          В
     Α0
x 0
         B0
     Α1
         B1
     Α2
         B2
y 0
     А3
         В3
```

Concatenation with joins

```
df5 = make_df('ABC', [1, 2])
df6 = make_df('BCD', [3, 4])
display('df5', 'df6', 'pd.concat([df5, df6])')
df5
    Α
        В
            C
1 A1
           C1
       В1
2 A2
       B2
           C2
df6
       C
    В
            D
3
   В3
       C3
           D3
4 B4
      C4
           D4
pd.concat([df5, df6])
     Α
         В
            C
        В1
    Α1
1
            C1
               NaN
2
        B2
           C2
    Α2
                NaN
3
  NaN
       В3
           С3
                 D3
  NaN
       В4
           C4
                 D4
display('df5', 'df6',
        "pd.concat([df5, df6], join='inner')")
df5
    Α
        В
           C
       B1 C1
1 A1
2 A2
       B2
          C2
df6
    В
       C
            D
       С3
3
   В3
           D3
4 B4
      C4
           D4
pd.concat([df5, df6], join='inner')
    В
      C
   В1
      C1
1
2
   B2
      C2
3
  В3
      С3
4
  В4
      C4
display('df5', 'df6',
        "pd.concat([df5, df6], join='outer')")
df5
    Α
        В
           C
1 A1
       В1
          C1
2 A2
       B2
           C2
df6
    В
        C
            D
```

```
В3
       C3
           D3
4 B4
      C4
           D4
pd.concat([df5, df6], join='outer')
     Α
         В
             C
1
    Α1
        В1
            C1
                NaN
        B2
            C2
    A2
                NaN
3
  NaN
        B3
            C3
                 D3
  NaN
        В4
            C4
                 D4
The append() method
display('df1', 'df2', 'df1.append(df2)')
df1
    Α
        В
1 A1
       B1
2 A2
       B2
df2
        В
    Α
3 A3
       В3
4 A4
       В4
df1.append(df2)
    Α
        В
   Α1
1
       В1
2
  Α2
       B2
  А3
       В3
4
  Α4
       B4
df=pd.read csv('/content/1.csv')
df.shape
(7, 4)
df.head()
   Gender
           Age Annual Income (k$)
                                     Spending Score (1-100)
0
     Male
            19
                                 15
                                                          39
     Male
            21
                                 15
                                                          81
1
2
   Female
            20
                                 16
                                                           6
  Female
            23
                                                          77
                                 16
  Female
            31
                                 17
                                                          40
df1=pd.read_csv('/content/2.csv')
df1.head()
           Age Annual Income (k$)
                                     Spending Score (1-100)
   Gender
0 Female
            23
                                 18
                                                          94
```

```
Male
                                    19
             64
                                                                3
1
2
   Female
             30
                                    19
                                                               72
     Male
                                    19
3
             67
                                                               14
   Female
             35
                                    19
                                                               99
df.append(df1)
   Gender
                  Annual Income (k$)
                                        Spending Score (1-100)
            Age
                                    15
0
     Male
             19
                                                               39
                                    15
1
     Male
             21
                                                               81
2
                                    16
                                                                6
   Female
             20
3
   Female
             23
                                    16
                                                               77
   Female
             31
                                    17
                                                               40
5
                                    17
   Female
             22
                                                               76
6
   Female
             35
                                    18
                                                                6
             23
                                    18
                                                               94
0
   Female
1
                                                                3
     Male
             64
                                    19
2
   Female
                                    19
                                                               72
             30
3
                                    19
     Male
             67
                                                               14
4
   Female
             35
                                    19
                                                               99
5
                                    20
                                                               15
   Female
             58
6
   Female
             24
                                    20
                                                               77
7
             37
                                    20
                                                               13
     Male
df.append(df1,ignore index=True)
                   Annual Income (k$)
                                         Spending Score (1-100)
    Gender
             Age
0
      Male
              19
                                     15
                                                                39
      Male
              21
                                     15
                                                                81
1
2
              20
                                     16
    Female
                                                                 6
3
    Female
              23
                                     16
                                                                77
4
    Female
              31
                                     17
                                                                40
5
    Female
              22
                                     17
                                                                76
6
    Female
              35
                                     18
                                                                 6
7
              23
                                                                94
    Female
                                     18
8
      Male
              64
                                     19
                                                                 3
9
    Female
              30
                                     19
                                                                72
10
      Male
              67
                                     19
                                                                14
                                     19
                                                                99
11
    Female
              35
12
    Female
              58
                                     20
                                                                15
13
    Female
              24
                                     20
                                                                77
                                     20
                                                                13
14
      Male
              37
df1.keys()
Index(['Gender', 'Age', 'Annual Income (k$)', 'Spending Score (1-
100)'], dtype='object')
df3=pd.read csv('/content/3.csv')
df3.shape
```

# (10, 4)

# df3.head()

| 0<br>1<br>2<br>3<br>4 | Gender<br>Female<br>Male<br>Female<br>Male<br>Female | Junio<br>Senio | ness Ana<br>r Consul<br>r Consul | alyst<br>ltant<br>ltant<br>nager | Level 1 2 3 4 5 | Sala<br>450<br>500<br>600<br>800<br>1100 | 00<br>00<br>00<br>00 |       |
|-----------------------|--|----------------|----------------------------------|----------------------------------|-----------------|--|----------------------|-------|
| df1                   | .append  | (df3)          |                                  |                                  |                 |  |                      |       |
| C n 1                 | Gender   | Age            | Annual                           | Incom                            | e (k\$)         |  | Position             | Level |
|                       |  | 23.0           |                                  |                                  | 18.0            |  | NaN                  | NaN   |
| 1<br>NaN              | Male   | 64.0           |                                  |                                  | 19.0            |  | NaN                  | NaN   |
|                       | Female   | 30.0           |                                  |                                  | 19.0            |  | NaN                  | NaN   |
| 3                     | Male   | 67.0           |                                  |                                  | 19.0            |  | NaN                  | NaN   |
| NaN<br>4<br>NaN       | Female   | 35.0           |                                  |                                  | 19.0            |  | NaN                  | NaN   |
| 5                     | Female   | 58.0           |                                  |                                  | 20.0            |  | NaN                  | NaN   |
| NaN<br>6              | Female   | 24.0           |                                  |                                  | 20.0            |  | NaN                  | NaN   |
| NaN<br>7              | Male   | 37.0           |                                  |                                  | 20.0            |  | NaN                  | NaN   |
| NaN<br>0              | I<br>Female  | NaN            |                                  |                                  | NaN             |  | Business Analyst     | 1.0   |
| 1                     | Male   | NaN            |                                  |                                  | NaN             |  | Junior Consultant    | 2.0   |
| 2                     | 000.0<br>Female                                      | NaN            |                                  |                                  | NaN             |  | Senior Consultant    | 3.0   |
| 600<br>3              | 000.0<br>Male  | NaN            |                                  |                                  | NaN             |  | Manager              | 4.0   |
|                       | 00.0<br>Female                                       | NaN            |                                  |                                  | NaN             |  | Country Manager      | 5.0   |
|                       | 000.0<br>Female                                      | NaN            |                                  |                                  | NaN             |  | Region Manager       | 6.0   |
|                       | 0000.0<br>Female                                     | NaN            |                                  |                                  | NaN             |  | Partner              | 7.0   |
| 200                   | 0.000  |                |                                  |                                  |                 |  |                      |       |
| 7<br>300              | Male<br>0000.0                                       | NaN            |                                  |                                  | NaN             |  | Senior Partner       | 8.0   |
| 8<br>500              | Female   | NaN            |                                  |                                  | NaN             |  | C-level              | 9.0   |
| 9                     | Male<br>00000.0                                      | NaN            |                                  |                                  | NaN             |  | CE0                  | 10.0  |

### [18 rows x 7 columns]

### df3.append(df1)

| Gender                     | Position          | <br>Annual Income (k\$) | Spending Score |
|----------------------------|-------------------|-------------------------|----------------|
| (1-100)<br>0 Female<br>NaN | Business Analyst  | <br>NaN                 |                |
| nan<br>1 Male<br>NaN       | Junior Consultant | <br>NaN                 |                |
|                            | Senior Consultant | <br>NaN                 |                |
| 3 Male<br>NaN              | Manager           | <br>NaN                 |                |
| 4 Female<br>NaN            | Country Manager   | <br>NaN                 |                |
| 5 Female<br>NaN            | Region Manager    | <br>NaN                 |                |
| 6 Female<br>NaN            | Partner           | <br>NaN                 |                |
| 7 Male<br>NaN              | Senior Partner    | <br>NaN                 |                |
| 8 Female<br>NaN            | C-level           | <br>NaN                 |                |
| 9 Male<br>NaN              | CE0               | <br>NaN                 |                |
| 0 Female<br>94.0           | NaN               | <br>18.0                |                |
| 1 Male 3.0                 | NaN               | <br>19.0                |                |
| Female 72.0                | NaN               | <br>19.0                |                |
| 3 Male<br>14.0             | NaN               | <br>19.0                |                |
| 4 Female 99.0              | NaN               | <br>19.0                |                |
| 5 Female<br>15.0           | NaN               | <br>20.0                |                |
| 6 Female 77.0              | NaN               | <br>20.0                |                |
| 7 Male<br>13.0             | NaN               | <br>20.0                |                |
| [10                        | 7                 |                         |                |

### [18 rows x 7 columns]

# df1.append(df3,sort=True).fillna("no value")

|   | Age Annual | Income (k\$) | <br>Salary   | Spending | Score | (1-100) |
|---|------------|--------------|--------------|----------|-------|---------|
| 0 | 23         | 18           | <br>no value |          |       | 94      |
| 1 | 64         | 19           | <br>no value |          |       | 3       |

```
2
          30
                               19
                                                                          72
                                         no value
3
                               19
                                                                          14
          67
                                         no value
4
          35
                               19
                                         no value
                                                                          99
5
          58
                               20
                                         no value
                                                                          15
6
                                                                          77
          24
                               20
                                         no value
7
          37
                               20
                                         no value
                                                                          13
0
   no value
                                             45000
                                                                   no value
                        no value
1
   no value
                                             50000
                                                                   no value
                        no value
2
   no value
                        no value
                                             60000
                                                                   no value
3
   no value
                                             80000
                        no value
                                                                   no value
4
   no value
                        no value
                                            110000
                                                                   no value
5
   no value
                        no value
                                            150000
                                                                   no value
6
                                                                   no value
   no value
                        no value
                                            200000
7
   no value
                        no value
                                            300000
                                                                   no value
8
   no value
                        no value
                                            500000
                                                                   no value
   no value
                        no value
                                             1e+06
                                                                   no value
[18 rows x 7 columns]
Newdf=df.append([df,df1,df3],ignore index=True) #combine all
dataframes
Newdf.shape
(32, 7)
Combining the dataframe using concat()
pd.concat([df,df1],ignore index=True)
                                         Spending Score (1-100)
    Gender
                   Annual Income (k$)
             Age
0
      Male
              19
                                     15
                                                                39
                                     15
1
      Male
              21
                                                                81
2
                                     16
                                                                 6
    Female
              20
3
              23
                                                                77
    Female
                                     16
4
                                     17
                                                                40
    Female
              31
5
    Female
              22
                                     17
                                                                76
6
    Female
              35
                                     18
                                                                 6
7
    Female
              23
                                     18
                                                                94
8
      Male
              64
                                     19
                                                                 3
9
                                                                72
    Female
              30
                                     19
10
                                     19
      Male
              67
                                                                14
11
    Female
              35
                                     19
                                                                99
                                                                15
12
    Female
              58
                                     20
13
                                                                77
    Female
              24
                                     20
14
      Male
              37
                                     20
                                                                13
pd.concat([df,df3])
                                                         Position Level
   Gender
             Age Annual Income (k$)
Salary
0
     Male
            19.0
                                   15.0
                                                               NaN
                                                                     NaN
NaN
```

| 1 Male<br>NaN<br>2 Female<br>NaN<br>3 Female | 21.0<br>20.0<br>23.0<br>31.0 | 15.<br>16.<br>16.<br>17. | 9       | NaN<br>NaN<br>NaN | NaN<br>NaN |
|--|------------------------------|--------------------------|---------|-------------------|------------|
| 2 Female<br>NaN<br>3 Female                  | 23.0                         | 16.                      |         |                   |            |
| 3 Female                                     |                              |                          | 9       | NaN               | NaN        |
|  | 31.0                         | 17                       |         |                   | NaN        |
| NaN<br>4 Female                              |                              | 17.                      | 9       | NaN               | NaN        |
| NaN<br>5 Female<br>NaN                       | 22.0                         | 17.                      | <b></b> | NaN               | NaN        |
| 6 Female                                     | 35.0                         | 18.                      | 9       | NaN               | NaN        |
| 0 Female<br>45000.0                          | NaN                          | Na                       | ٠       | Business Analyst  | 1.0        |
| 1 Male 5000.0                                | NaN                          | Na                       | ١       | Junior Consultant | 2.0        |
| 2 Female 60000.0                             | NaN                          | Na                       | ٠       | Senior Consultant | 3.0        |
| 3 Male<br>80000.0                            | NaN                          | Na                       | ١       | Manager           | 4.0        |
| 4 Female 110000.0                            | NaN                          | Na                       | ٠       | Country Manager   | 5.0        |
| 5 Female 150000.0                            | NaN                          | Na                       | ١       | Region Manager    | 6.0        |
| 6 Female 200000.0                            | NaN                          | Na                       | ١       | Partner           | 7.0        |
| 7 Male 300000.0                              | NaN                          | Na                       | ١       | Senior Partner    | 8.0        |
| 8 Female 500000.0                            | NaN                          | Na                       | ١       | C-level           | 9.0        |
| 9 Male 1000000.0                             | NaN                          | Na                       | ١       | CE0               | 10.0       |
| [17] mayer x                                 |                              | _                        |         |                   |            |

### [17 rows x 7 columns]

concatdf=pd.concat([df,df3],keys=["xls 1","xls 2"])
concatdf

|       |   | Gender | Age  | Annual Income | (k\$) | <br>Position |
|-------|---|--------|------|---------------|-------|--------------|
| Level |   | Salary |      |               |       |              |
| xls 1 | 0 | Male   | 19.0 |               | 15.0  | <br>NaN      |
| NaN   |   | NaN    |      |               |       |              |
|       | 1 | Male   | 21.0 |               | 15.0  | <br>NaN      |
| NaN   |   | NaN    |      |               |       |              |
|       | 2 | Female | 20.0 |               | 16.0  | <br>NaN      |
| NaN   |   | NaN    |      |               |       |              |
|       | 3 | Female | 23.0 |               | 16.0  | <br>NaN      |
| NaN   |   | NaN    |      |               |       |              |
|       | 4 | Female | 31.0 |               | 17.0  | <br>NaN      |
| NaN   |   | NaN    |      |               |       |              |

```
Female
                  22.0
                                        17.0
                                                                   NaN
NaN
            NaN
                  35.0
                                        18.0
         Female
                                                                   NaN
NaN
            NaN
xls 2 0
         Female
                                                     Business Analyst
                   NaN
                                         NaN
                                               . . .
1.0
       45000.0
                                                    Junior Consultant
      1
           Male
                   NaN
                                         NaN
2.0
       50000.0
        Female
                   NaN
                                         NaN
                                               . . .
                                                    Senior Consultant
3.0
       60000.0
           Male
                   NaN
                                         NaN
                                                              Manager
                                               . . .
4.0
       80000.0
      4 Female
                   NaN
                                         NaN
                                               . . .
                                                      Country Manager
      110000.0
5.0
         Female
                   NaN
                                         NaN
                                                       Region Manager
6.0
      150000.0
         Female
                   NaN
                                         NaN
                                                               Partner
7.0
      200000.0
                                                       Senior Partner
           Male
                   NaN
                                         NaN
      7
                                               . . .
8.0
      300000.0
         Female
                                                              C-level
                   NaN
                                         NaN
9.0
      500000.0
           Male
                   NaN
                                         NaN
                                                                   CE0
                                               . . .
10.0
      1000000.0
[17 rows x 7 columns]
concatdf.loc["xls 1"]
   Gender
                  Annual Income (k$)
                                             Position Level
                                                               Salary
            Age
                                        . . .
0
     Male
           19.0
                                 15.0
                                                   NaN
                                                         NaN
                                                                  NaN
                                        . . .
1
     Male
           21.0
                                 15.0
                                                   NaN
                                                         NaN
                                                                  NaN
                                        . . .
  Female 20.0
                                 16.0
                                        . . .
                                                   NaN
                                                         NaN
                                                                  NaN
  Female 23.0
                                 16.0
                                                   NaN
                                                         NaN
                                                                  NaN
                                        . . .
4
   Female 31.0
                                                   NaN
                                                         NaN
                                                                  NaN
                                 17.0
                                        . . .
5
   Female
           22.0
                                 17.0
                                                   NaN
                                                         NaN
                                                                  NaN
   Female
           35.0
                                                                  NaN
                                 18.0
                                                   NaN
                                                         NaN
[7 rows x 7 columns]
concatdf=pd.concat([df,df3],keys=["xls 1","xls 2"], axis=1)
concatdf
    xls 1
                                                         xls 2
   Gender
            Age Annual Income (k$)
                                                      Position Level
                                       . . .
Salary
     Male
           19.0
                                15.0
                                       ... Business Analyst
                                                                    1
45000
     Male
           21.0
                                15.0 ...
                                            Junior Consultant
                                                                    2
50000
```

```
2 Female
                                     ... Senior Consultant
           20.0
                                16.0
                                                                   3
60000
  Female 23.0
                                16.0
3
                                                      Manager
                                                                   4
80000
   Female 31.0
                                17.0
                                             Country Manager
                                                                   5
                                      . . .
110000
   Female 22.0
                                17.0
                                              Region Manager
                                                                   6
150000
   Female
           35.0
                                18.0
                                                      Partner
                                                                   7
                                      . . .
200000
      NaN
            NaN
                                 NaN
                                              Senior Partner
                                                                   8
                                      . . .
300000
                                                      C-level
      NaN
                                                                   9
            NaN
                                 NaN
                                      . . .
500000
                                                          CE0
      NaN
            NaN
                                 NaN
                                                                  10
1000000
[10 rows x 8 columns]
pd.concat([df,df3], join="inner")
   Gender
     Male
0
1
     Male
2
  Female
3
   Female
4
   Female
5
   Female
6
   Female
0
   Female
1
     Male
2
   Female
3
     Male
4
   Female
5
   Female
6
   Female
7
     Male
8
   Female
     Male
pd.concat([df,df3], join="outer")
   Gender
            Age Annual Income (k$) ...
                                                      Position Level
Salary
     Male 19.0
                                 15.0
                                                           NaN
                                                                  NaN
NaN
     Male 21.0
                                 15.0
                                                           NaN
                                                                  NaN
1
                                      . . .
NaN
2
  Female 20.0
                                 16.0
                                                           NaN
                                                                  NaN
NaN
3 Female 23.0
                                 16.0
                                      . . .
                                                           NaN
                                                                  NaN
```

| NaN                        |      |      |                       |      |
|----------------------------|------|------|-----------------------|------|
| 4 Female<br>NaN            | 31.0 | 17.0 | <br>NaN               | NaN  |
| 5 Female<br>NaN            | 22.0 | 17.0 | <br>NaN               | NaN  |
| 6 Female                   | 35.0 | 18.0 | <br>NaN               | NaN  |
| NaN<br>0 Female<br>45000.0 | NaN  | NaN  | <br>Business Analyst  | 1.0  |
| 1 Male 5000.0              | NaN  | NaN  | <br>Junior Consultant | 2.0  |
| 2 Female 60000.0           | NaN  | NaN  | <br>Senior Consultant | 3.0  |
| 3 Male<br>80000.0          | NaN  | NaN  | <br>Manager           | 4.0  |
| 4 Female 110000.0          | NaN  | NaN  | <br>Country Manager   | 5.0  |
| 5 Female<br>150000.0       | NaN  | NaN  | <br>Region Manager    | 6.0  |
| 6 Female 200000.0          | NaN  | NaN  | <br>Partner           | 7.0  |
| 7 Male 300000.0            | NaN  | NaN  | <br>Senior Partner    | 8.0  |
| 8 Female 500000.0          | NaN  | NaN  | <br>C-level           | 9.0  |
| 9 Male 1000000.0           | NaN  | NaN  | <br>CE0               | 10.0 |

[17 rows x 7 columns]