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    A
2    B
3    C
4    D
5    E
6    F
dtype: object
```

```
def make_df(cols, ind):
    """Quickly make a DataFrame"""
    data = {c: [str(c) + str(i) for i in ind]
            for c in cols}
    return pd.DataFrame(data, ind)

# example DataFrame
make_df('ABC', range(3))
```

```
    A   B   C
0  A0  B0  C0
1  A1  B1  C1
2  A2  B2  C2
```

```
df1 = make_df('AB', [1, 2])
df2 = make_df('AB', [3, 4])
display('df1', 'df2', 'pd.concat([df1, df2])')
```

```
df1
    A   B
1  A1  B1
```

```
2  A2  B2

df2
    A   B
3  A3  B3
4  A4  B4

pd.concat([df1, df2])
    A   B
1  A1  B1
2  A2  B2
3  A3  B3
4  A4  B4
```

```
df3 = make_df('AB', [0, 1])
df4 = make_df('CD', [0, 1])
display('df3', 'df4', "pd.concat([df3, df4], axis=1)")
```

```
df3
    A   B
0  A0  B0
1  A1  B1

df4
    C   D
0  C0  D0
1  C1  D1

pd.concat([df3, df4], axis=1)
    A   B   C   D
0  A0  B0  C0  D0
1  A1  B1  C1  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])
    A   B
0  A0  B0
1  A1  B1
0  A2  B2
1  A3  B3
```

```
display('x', 'y', 'pd.concat([x, y], ignore_index=True)')
```

```
x
    A   B
0  A0  B0
1  A1  B1

y
    A   B
0  A2  B2
1  A3  B3

pd.concat([x, y], ignore_index=True)
    A   B
0  A0  B0
1  A1  B1
2  A2  B2
3  A3  B3
```

Adding MultiIndex keys

```
display('x', 'y', "pd.concat([x, y], keys=['x', 'y'])")
```

```
x
    A   B
0  A0  B0
1  A1  B1

y
    A   B
0  A2  B2
1  A3  B3

pd.concat([x, y], keys=['x', 'y'])
      A   B
x 0  A0  B0
  1  A1  B1
y 0  A2  B2
  1  A3  B3
```

Concatenation with joins

```
df5 = make_df('ABC', [1, 2])
df6 = make_df('BCD', [3, 4])
display('df5', 'df6', 'pd.concat([df5, df6])')
```

df5

	A	B	C
1	A1	B1	C1
2	A2	B2	C2

df6

	B	C	D
3	B3	C3	D3
4	B4	C4	D4

```
pd.concat([df5, df6])
```

	A	B	C	D
1	A1	B1	C1	NaN
2	A2	B2	C2	NaN
3	NaN	B3	C3	D3
4	NaN	B4	C4	D4

```
display('df5', 'df6',
        "pd.concat([df5, df6], join='inner')")
```

df5

	A	B	C
1	A1	B1	C1
2	A2	B2	C2

df6

	B	C	D
3	B3	C3	D3
4	B4	C4	D4

```
pd.concat([df5, df6], join='inner')
```

	B	C
1	B1	C1
2	B2	C2
3	B3	C3
4	B4	C4

```
display('df5', 'df6',
        "pd.concat([df5, df6], join='outer')")
```

df5

	A	B	C
1	A1	B1	C1
2	A2	B2	C2

df6

	B	C	D

```
3  B3  C3  D3
4  B4  C4  D4
```

```
pd.concat([df5, df6], join='outer')
     A   B   C    D
1   A1  B1  C1  NaN
2   A2  B2  C2  NaN
3  NaN  B3  C3   D3
4  NaN  B4  C4   D4
```

The append() method

```
display('df1', 'df2', 'df1.append(df2)')
```

```
df1
    A   B
1  A1  B1
2  A2  B2

df2
    A   B
3  A3  B3
4  A4  B4

df1.append(df2)
    A   B
1  A1  B1
2  A2  B2
3  A3  B3
4  A4  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
1    Male   21                  15                      81
2  Female   20                  16                       6
3  Female   23                  16                      77
4  Female   31                  17                      40
```

```
df1=pd.read_csv('/content/2.csv')
```

```
df1.head()
```

```
   Gender  Age  Annual Income (k$)  Spending Score (1-100)
0  Female   23                  18                      94
```

```
1    Male   64                  19                       3
2  Female   30                  19                      72
3    Male   67                  19                      14
4  Female   35                  19                      99
```

```
df.append(df1)
```

```
   Gender  Age  Annual Income (k$)  Spending Score (1-100)
0    Male   19                  15                      39
1    Male   21                  15                      81
2  Female   20                  16                       6
3  Female   23                  16                      77
4  Female   31                  17                      40
5  Female   22                  17                      76
6  Female   35                  18                       6
0  Female   23                  18                      94
1    Male   64                  19                       3
2  Female   30                  19                      72
3    Male   67                  19                      14
4  Female   35                  19                      99
5  Female   58                  20                      15
6  Female   24                  20                      77
7    Male   37                  20                      13
```

```
df.append(df1,ignore_index=True)
```

```
    Gender  Age  Annual Income (k$)  Spending Score (1-100)
0     Male   19                  15                      39
1     Male   21                  15                      81
2   Female   20                  16                       6
3   Female   23                  16                      77
4   Female   31                  17                      40
5   Female   22                  17                      76
6   Female   35                  18                       6
7   Female   23                  18                      94
8     Male   64                  19                       3
9   Female   30                  19                      72
10    Male   67                  19                      14
11  Female   35                  19                      99
12  Female   58                  20                      15
13  Female   24                  20                      77
14    Male   37                  20                      13
```

```
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()
```

```
   Gender           Position  Level  Salary
0  Female   Business Analyst      1   45000
1    Male  Junior Consultant      2   50000
2  Female  Senior Consultant      3   60000
3    Male            Manager      4   80000
4  Female    Country Manager      5  110000
```

```
df1.append(df3)
```

```
   Gender   Age  Annual Income (k$)  ...           Position Level
Salary
0  Female  23.0                18.0  ...                NaN   NaN
NaN
1    Male  64.0                19.0  ...                NaN   NaN
NaN
2  Female  30.0                19.0  ...                NaN   NaN
NaN
3    Male  67.0                19.0  ...                NaN   NaN
NaN
4  Female  35.0                19.0  ...                NaN   NaN
NaN
5  Female  58.0                20.0  ...                NaN   NaN
NaN
6  Female  24.0                20.0  ...                NaN   NaN
NaN
7    Male  37.0                20.0  ...                NaN   NaN
NaN
0  Female   NaN                 NaN  ...   Business Analyst   1.0
45000.0
1    Male   NaN                 NaN  ...  Junior Consultant   2.0
50000.0
2  Female   NaN                 NaN  ...  Senior Consultant   3.0
60000.0
3    Male   NaN                 NaN  ...            Manager   4.0
80000.0
4  Female   NaN                 NaN  ...    Country Manager   5.0
110000.0
5  Female   NaN                 NaN  ...     Region Manager   6.0
150000.0
6  Female   NaN                 NaN  ...            Partner   7.0
200000.0
7    Male   NaN                 NaN  ...     Senior Partner   8.0
300000.0
8  Female   NaN                 NaN  ...            C-level   9.0
500000.0
9    Male   NaN                 NaN  ...                CEO  10.0
1000000.0
```

```
[18 rows x 7 columns]
```

```
df3.append(df1)
```

```
   Gender           Position  ...  Annual Income (k$)  Spending Score
(1-100)
0  Female   Business Analyst  ...                 NaN
NaN
1    Male  Junior Consultant  ...                 NaN
NaN
2  Female  Senior Consultant  ...                 NaN
NaN
3    Male            Manager  ...                 NaN
NaN
4  Female    Country Manager  ...                 NaN
NaN
5  Female     Region Manager  ...                 NaN
NaN
6  Female            Partner  ...                 NaN
NaN
7    Male     Senior Partner  ...                 NaN
NaN
8  Female            C-level  ...                 NaN
NaN
9    Male                CEO  ...                 NaN
NaN
0  Female                NaN  ...                18.0
94.0
1    Male                NaN  ...                19.0
3.0
2  Female                NaN  ...                19.0
72.0
3    Male                NaN  ...                19.0
14.0
4  Female                NaN  ...                19.0
99.0
5  Female                NaN  ...                20.0
15.0
6  Female                NaN  ...                20.0
77.0
7    Male                NaN  ...                20.0
13.0

[18 rows x 7 columns]
```

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

```
        Age Annual Income (k$)  ...    Salary Spending Score (1-100)
0        23                 18  ...  no value                    94
1        64                 19  ...  no value                     3
```

```
2        30                  19  ...  no value                    72
3        67                  19  ...  no value                    14
4        35                  19  ...  no value                    99
5        58                  20  ...  no value                    15
6        24                  20  ...  no value                    77
7        37                  20  ...  no value                    13
0  no value           no value  ...     45000              no value
1  no value           no value  ...     50000              no value
2  no value           no value  ...     60000              no value
3  no value           no value  ...     80000              no value
4  no value           no value  ...    110000              no value
5  no value           no value  ...    150000              no value
6  no value           no value  ...    200000              no value
7  no value           no value  ...    300000              no value
8  no value           no value  ...    500000              no value
9  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)
```

```
    Gender  Age  Annual Income (k$)  Spending Score (1-100)
0     Male   19                  15                      39
1     Male   21                  15                      81
2   Female   20                  16                       6
3   Female   23                  16                      77
4   Female   31                  17                      40
5   Female   22                  17                      76
6   Female   35                  18                       6
7   Female   23                  18                      94
8     Male   64                  19                       3
9   Female   30                  19                      72
10    Male   67                  19                      14
11  Female   35                  19                      99
12  Female   58                  20                      15
13  Female   24                  20                      77
14    Male   37                  20                      13
```

```
pd.concat([df,df3])
```

```
   Gender   Age  Annual Income (k$)  ...           Position Level
Salary
0    Male  19.0                15.0  ...                  NaN   NaN
NaN
```

```
1    Male  21.0                15.0  ...                NaN   NaN
NaN
2  Female  20.0                16.0  ...                NaN   NaN
NaN
3  Female  23.0                16.0  ...                NaN   NaN
NaN
4  Female  31.0                17.0  ...                NaN   NaN
NaN
5  Female  22.0                17.0  ...                NaN   NaN
NaN
6  Female  35.0                18.0  ...                NaN   NaN
NaN
0  Female   NaN                 NaN  ...   Business Analyst   1.0
45000.0
1    Male   NaN                 NaN  ...  Junior Consultant   2.0
50000.0
2  Female   NaN                 NaN  ...  Senior Consultant   3.0
60000.0
3    Male   NaN                 NaN  ...            Manager   4.0
80000.0
4  Female   NaN                 NaN  ...    Country Manager   5.0
110000.0
5  Female   NaN                 NaN  ...     Region Manager   6.0
150000.0
6  Female   NaN                 NaN  ...            Partner   7.0
200000.0
7    Male   NaN                 NaN  ...     Senior Partner   8.0
300000.0
8  Female   NaN                 NaN  ...            C-level   9.0
500000.0
9    Male   NaN                 NaN  ...                CEO  10.0
1000000.0

[17 rows x 7 columns]
```

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

```
         Gender   Age  Annual Income (k$)  ...           Position
Level     Salary
xls 1 0    Male  19.0                15.0  ...                NaN
NaN         NaN
      1    Male  21.0                15.0  ...                NaN
NaN         NaN
      2  Female  20.0                16.0  ...                NaN
NaN         NaN
      3  Female  23.0                16.0  ...                NaN
NaN         NaN
      4  Female  31.0                17.0  ...                NaN
NaN         NaN
```

```
      5  Female  22.0                17.0  ...                NaN
NaN        NaN
      6  Female  35.0                18.0  ...                NaN
NaN        NaN
xls 2 0  Female   NaN                 NaN  ...   Business Analyst
1.0    45000.0
      1    Male   NaN                 NaN  ...  Junior Consultant
2.0    50000.0
      2  Female   NaN                 NaN  ...  Senior Consultant
3.0    60000.0
      3    Male   NaN                 NaN  ...            Manager
4.0    80000.0
      4  Female   NaN                 NaN  ...    Country Manager
5.0   110000.0
      5  Female   NaN                 NaN  ...     Region Manager
6.0   150000.0
      6  Female   NaN                 NaN  ...            Partner
7.0   200000.0
      7    Male   NaN                 NaN  ...     Senior Partner
8.0   300000.0
      8  Female   NaN                 NaN  ...            C-level
9.0   500000.0
      9    Male   NaN                 NaN  ...                CEO
10.0  1000000.0

[17 rows x 7 columns]
```

```
concatdf.loc["xls 1"]
```

```
   Gender   Age  Annual Income (k$)  ...  Position Level  Salary
0    Male  19.0                15.0  ...       NaN   NaN     NaN
1    Male  21.0                15.0  ...       NaN   NaN     NaN
2  Female  20.0                16.0  ...       NaN   NaN     NaN
3  Female  23.0                16.0  ...       NaN   NaN     NaN
4  Female  31.0                17.0  ...       NaN   NaN     NaN
5  Female  22.0                17.0  ...       NaN   NaN     NaN
6  Female  35.0                18.0  ...       NaN   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
0    Male  19.0               15.0  ...   Business Analyst     1
45000
1    Male  21.0               15.0  ...  Junior Consultant     2
50000
```

```
2  Female  20.0               16.0  ...  Senior Consultant     3
60000
3  Female  23.0               16.0  ...            Manager     4
80000
4  Female  31.0               17.0  ...    Country Manager     5
110000
5  Female  22.0               17.0  ...     Region Manager     6
150000
6  Female  35.0               18.0  ...            Partner     7
200000
7     NaN   NaN                NaN  ...     Senior Partner     8
300000
8     NaN   NaN                NaN  ...            C-level     9
500000
9     NaN   NaN                NaN  ...                CEO    10
1000000

[10 rows x 8 columns]
```

```
pd.concat([df,df3], join="inner")
```

```
   Gender
0    Male
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
9    Male
```

```
pd.concat([df,df3], join="outer")
```

```
   Gender   Age  Annual Income (k$)  ...           Position Level
Salary
0    Male  19.0                15.0  ...                NaN   NaN
NaN
1    Male  21.0                15.0  ...                NaN   NaN
NaN
2  Female  20.0                16.0  ...                NaN   NaN
NaN
3  Female  23.0                16.0  ...                NaN   NaN
```

```
NaN
4  Female  31.0                17.0  ...                NaN   NaN
NaN
5  Female  22.0                17.0  ...                NaN   NaN
NaN
6  Female  35.0                18.0  ...                NaN   NaN
NaN
0  Female   NaN                 NaN  ...   Business Analyst   1.0
45000.0
1    Male   NaN                 NaN  ...  Junior Consultant   2.0
50000.0
2  Female   NaN                 NaN  ...  Senior Consultant   3.0
60000.0
3    Male   NaN                 NaN  ...            Manager   4.0
80000.0
4  Female   NaN                 NaN  ...    Country Manager   5.0
110000.0
5  Female   NaN                 NaN  ...     Region Manager   6.0
150000.0
6  Female   NaN                 NaN  ...            Partner   7.0
200000.0
7    Male   NaN                 NaN  ...     Senior Partner   8.0
300000.0
8  Female   NaN                 NaN  ...            C-level   9.0
500000.0
9    Male   NaN                 NaN  ...                CEO  10.0
1000000.0

[17 rows x 7 columns]
```