

# Joining data in pandas

In this notebook, we'll use pandas to join some relational data:

- `../data/country-codes.csv` -- a table of ISO country codes and country names
- `../data/country-population.csv` -- country population data from the U.N.

👉 [Read more about the `merge` method for joining dataframes](#)

```
In [8]: import pandas as pd
```

When we read in the CSVs, we need to make sure that pandas doesn't parse the ISO codes as numbers, because we want to keep any leading zeroes. So in addition to the path to the CSV, we'll also use an argument called `dtype` to specify that the `code` columns need to be parsed as a string.

👉 You can find more information on the `dtype` argument [in the documentation for the `read\_csv\(\)` method](#)

```
In [9]: country_codes = pd.read_csv('../data/country-codes.csv', dtype={'code': str})
```

```
In [10]: country_codes.head()
```

```
Out[10]:
```

	code	country
0	108	Burundi
1	174	Comoros
2	262	Djibouti
3	232	Eritrea
4	231	Ethiopia

```
In [11]: country_pop = pd.read_csv('../data/country-population.csv', dtype={'code': str})
```

```
In [12]: country_pop.head()
```

```
Out[12]:
```

	code	pop2000	pop2001	pop2002	pop2003	pop2004	pop2005	pop2006	pop2007	pop2008	pop2009
0	108	6401.0	6556.0	6742.0	6953.0	7182.0	7423.0	7675.0	7940.0	8212.0	8483.0
1	174	542.0	556.0	569.0	583.0	597.0	612.0	626.0	642.0	657.0	671.0
2	262	718.0	733.0	746.0	759.0	771.0	783.0	796.0	809.0	823.0	836.0
3	232	3393.0	3497.0	3615.0	3738.0	3859.0	3969.0	4067.0	4153.0	4233.0	4313.0
4	231	66537.0	68492.0	70497.0	72545.0	74624.0	76727.0	78851.0	81000.0	83185.0	85400.0

## Join the data with the country codes lookup table

To join data in pandas, we can use the `merge()` method. At minimum, you need to hand this method the two dataframes to join, plus specify the name of the column to join on. (If the columns have different

names, you can use the `left_on` and `right_on` arguments -- the "left" dataframe is the first one you hand to the `merge` method.)

```
In [13]: merged = pd.merge(country_pop,
                           country_codes,
                           on='code')
```

```
In [14]: merged.head()
```

Out[14]:

	code	pop2000	pop2001	pop2002	pop2003	pop2004	pop2005	pop2006	pop2007	pop2008	pop2009
0	108	6401.0	6556.0	6742.0	6953.0	7182.0	7423.0	7675.0	7940.0	8212.0	8483.0
1	174	542.0	556.0	569.0	583.0	597.0	612.0	626.0	642.0	657.0	671.0
2	262	718.0	733.0	746.0	759.0	771.0	783.0	796.0	809.0	823.0	835.0
3	232	3393.0	3497.0	3615.0	3738.0	3859.0	3969.0	4067.0	4153.0	4233.0	4308.0
4	231	66537.0	68492.0	70497.0	72545.0	74624.0	76727.0	78851.0	81000.0	83185.0	85400.0