

# Announcement

- **HW2** due on Apr 9th (Fri) 11:59pm
- **Midterm**
  - 4/21 (Wed) 9:00 - 10:15 am
  - Offline (preferred) - under review by CS department
    - Instructions will be announced soon
- **TAs**
  - contact: [cose471.sp21.staff@gmail.com](mailto:cose471.sp21.staff@gmail.com)
- **HW1**
  - We are now grading your homework. Grades will be published soon.

## [Pandas] merge

- Basic syntax for joining two dataframes `df` and `df2`
  - `df.merge(df2)`
- Output is another dataframe

### pandas.DataFrame.merge

```
DataFrame.merge(right, how='inner', on=None, left_on=None, right_on=None, left_index=False, right_index=False,  
sort=False, suffixes=('_x', '_y'), copy=True, indicator=False, validate=None) \[source\]
```

Merge DataFrame or named Series objects with a database-style join.

The join is done on columns or indexes. If joining columns on columns, the DataFrame indexes *will be ignored*. Otherwise if joining indexes on indexes or indexes on a column or columns, the index will be passed on. When performing a cross merge, no column specifications to merge on are allowed.

## Types of Joins

- Typical usage 

```
df.merge(df2,  
        how = "inner",  
        left_on = "column_label_in_df",  
        right_on = "column_label_in_df2")
```
- “inner” can be replaced by “outer” or “left” or “right”

## Inner Join

s		t	
<u>id</u>	<u>name</u>	<u>id</u>	<u>breed</u>
0	Apricot		
1	Boots	1	persian
2	Cally	2	ragdoll
4	Eugene	4	bengal
		5	persian

Only pairs of matching rows appear in the result.

## Inner Join

s		t	
<u>id</u>	<u>name</u>	<u>id</u>	<u>breed</u>
0	Apricot	1	persian
1	Boots	2	ragdoll
2	Cally	4	bengal
4	Eugene	5	persian

<u>s.id</u>	<u>name</u>	<u>t.id</u>	<u>breed</u>
1	Boots	1	persian
2	Cally	2	ragdoll
4	Eugene	4	bengal

Only pairs of matching rows appear in the result.

## Left (Outer) Join

s		t	
<u>id</u>	<u>name</u>	<u>id</u>	<u>breed</u>
0	Apricot		
1	Boots	1	persian
2	Cally	2	ragdoll
4	Eugene	4	bengal
		5	persian

Every row in the first table appears in the result, matching or not.

## Left (Outer) Join

s		t	
<u>id</u>	<u>name</u>	<u>id</u>	<u>breed</u>
0	Apricot	1	persian
1	Boots	2	ragdoll
2	Cally	4	bengal
4	Eugene	5	persian

<u>s.id</u>	<u>name</u>	<u>t.id</u>	<u>breed</u>
0	Apricot		
1	Boots	1	persian
2	Cally	2	ragdoll
4	Eugene	4	bengal

Missing values are null.

Every row in the first table appears in the result, matching or not.

## Right (Outer) Join

s

<u>id</u>	<u>name</u>
0	Apricot
1	Boots
2	Cally
4	Eugene

t

<u>id</u>	<u>breed</u>
1	persian
2	ragdoll
4	bengal
5	persian

s.id   name   t.id   breed

1	Boots	1	persian
2	Cally	2	ragdoll
4	Eugene	4	bengal
		5	persian

Every row in the second table appears in the result, matching or not.



# Outer Join

s		t	
<u>id</u>	<u>name</u>	<u>id</u>	<u>breed</u>
0	Apricot	1	persian
1	Boots	2	ragdoll
2	Cally	4	bengal
4	Eugene	5	persian

<u>s.id</u>	<u>name</u>	<u>t.id</u>	<u>breed</u>
0	Apricot		
1	Boots	1	persian
2	Cally	2	ragdoll
4	Eugene	4	bengal
		5	persian

Every row in both tables appears, matching or not.

# Example

## elections

	Year	Candidate	Party	Popular vote	Result	%
0	1824	Andrew Jackson	Democratic-Republican	151271	loss	57.210122
1	1824	John Quincy Adams	Democratic-Republican	113142	win	42.789878
2	1828	Andrew Jackson	Democratic	642806	win	56.203927
3	1828	John Quincy Adams	National Republican	500897	loss	43.796073
4	1832	Andrew Jackson	Democratic	702735	win	54.574789

## presidents

	Date of birth	President	Birthplace	State of birth	In office
0	February 22, 1732	George Washington	Westmoreland County	Virginia	(1st) April 30, 1789 – March 4, 1797
1	October 30, 1735	John Adams	Braintree	Massachusetts	(2nd) March 4, 1797 – March 4, 1801
2	April 13, 1743	Thomas Jefferson	Shadwell	Virginia	(3rd) March 4, 1801 – March 4, 1809
3	March 16, 1751	James Madison	Port Conway	Virginia	(4th) March 4, 1809 – March 4, 1817
4	April 28, 1758	James Monroe	Monroe Hall	Virginia	(5th) March 4, 1817 – March 4, 1825

```
elections.merge(presidents,  
                 how = "inner",  
                 left_on = "Candidate", right_on = "President")
```

**left\_on** : *label or list, or array-like*

Column or index level names to join on in the left DataFrame. Can also be an array or list of arrays of the length of the left DataFrame. These arrays are treated as if they are columns.

**right\_on** : *label or list, or array-like*

Column or index level names to join on in the right DataFrame. Can also be an array or list of arrays of the length of the right DataFrame. These arrays are treated as if they are columns.

# Example

## elections

	Year	Candidate	Party	Popular vote	Result	%
0	1824	Andrew Jackson	Democratic-Republican	151271	loss	57.210122
1	1824	John Quincy Adams	Democratic-Republican	113142	win	42.789878
2	1828	Andrew Jackson	Democratic	642806	win	56.203927
3	1828	John Quincy Adams	National Republican	500897	loss	43.796073
4	1832	Andrew Jackson	Democratic	702735	win	54.574789

## presidents

	Date of birth	President	Birthplace	State of birth	In office
0	February 22, 1732	George Washington	Westmoreland County	Virginia	(1st) April 30, 1789 – March 4, 1797
1	October 30, 1735	John Adams	Braintree	Massachusetts	(2nd) March 4, 1797 – March 4, 1801
2	April 13, 1743	Thomas Jefferson	Shadwell	Virginia	(3rd) March 4, 1801 – March 4, 1809
3	March 16, 1751	James Madison	Port Conway	Virginia	(4th) March 4, 1809 – March 4, 1817
4	April 28, 1758	James Monroe	Monroe Hall	Virginia	(5th) March 4, 1817 – March 4, 1825

```
elections.merge(presidents,  
                 how = "inner",  
                 left_on = "Candidate", right_on = "President")
```

	Year	Candidate	Party	Popular vote	Result	%	Date of birth	President	Birthplace	State of birth	In office
0	1824	Andrew Jackson	Democratic-Republican	151271	loss	57.210122	March 15, 1767	Andrew Jackson	Waxhaws Region	South/North Carolina	(7th) March 4, 1829 – March 4, 1837
1	1828	Andrew Jackson	Democratic	642806	win	56.203927	March 15, 1767	Andrew Jackson	Waxhaws Region	South/North Carolina	(7th) March 4, 1829 – March 4, 1837
2	1832	Andrew Jackson	Democratic	702735	win	54.574789	March 15, 1767	Andrew Jackson	Waxhaws Region	South/North Carolina	(7th) March 4, 1829 – March 4, 1837
3	1824	John Quincy Adams	Democratic-Republican	113142	win	42.789878	July 11, 1767	John Quincy Adams	Braintree	Massachusetts	(6th) March 4, 1825 – March 4, 1829
4	1828	John Quincy Adams	National Republican	500897	loss	43.796073	July 11, 1767	John Quincy Adams	Braintree	Massachusetts	(6th) March 4, 1825 – March 4, 1829
5	1836	Martin Van Buren	Democratic	763291	win	52.272472	December 5, 1782	Martin Van Buren	Kinderhook	New York	(8th) March 4, 1837 – March 4, 1841

# Example (Right Join)

## elections

	Year	Candidate	Party	Popular vote	Result	%
0	1824	Andrew Jackson	Democratic-Republican	151271	loss	57.210122
1	1824	John Quincy Adams	Democratic-Republican	113142	win	42.789878
2	1828	Andrew Jackson	Democratic	642806	win	56.203927
3	1828	John Quincy Adams	National Republican	500897	loss	43.796073
4	1832	Andrew Jackson	Democratic	702735	win	54.574789

## presidents

	Date of birth	President	Birthplace	State of birth	In office
0	February 22, 1732	George Washington	Westmoreland County	Virginia	(1st) April 30, 1789 – March 4, 1797
1	October 30, 1735	John Adams	Braintree	Massachusetts	(2nd) March 4, 1797 – March 4, 1801
2	April 13, 1743	Thomas Jefferson	Shadwell	Virginia	(3rd) March 4, 1801 – March 4, 1809
3	March 16, 1751	James Madison	Port Conway	Virginia	(4th) March 4, 1809 – March 4, 1817
4	April 28, 1758	James Monroe	Monroe Hall	Virginia	(5th) March 4, 1817 – March 4, 1825

```
elections.merge(presidents,  
                 how = "right",  
                 left_on = "Candidate", right_on = "President")
```

	Year	Candidate	Party	Popular vote	Result	%	Date of birth	President	Birthplace	State of birth	In office
0	1824.0	Andrew Jackson	Democratic-Republican	151271.0	loss	57.210122	March 15, 1767	Andrew Jackson	Waxhaws Region	South/North Carolina	(7th) March 4, 1829 – March 4, 1837
1	1828.0	Andrew Jackson	Democratic	642806.0	win	56.203927	March 15, 1767	Andrew Jackson	Waxhaws Region	South/North Carolina	(7th) March 4, 1829 – March 4, 1837
2	1832.0	Andrew Jackson	Democratic	702735.0	win	54.574789	March 15, 1767	Andrew Jackson	Waxhaws Region	South/North Carolina	(7th) March 4, 1829 – March 4, 1837
3	1824.0	John Quincy Adams	Democratic-Republican	113142.0	win	42.789878	July 11, 1767	John Quincy Adams	Braintree	Massachusetts	(6th) March 4, 1825 – March 4, 1829
4	1828.0	John Quincy Adams	National Republican	500897.0	loss	43.796073	July 11, 1767	John Quincy Adams	Braintree	Massachusetts	(6th) March 4, 1825 – March 4, 1829
5	1836.0	Martin Van Buren	Democratic	763291.0	win	52.272472	December 5, 1782	Martin Van Buren	Kinderhook	New York	(8th) March 4, 1837 – March 4, 1841
6	1840.0	Martin Van Buren	Democratic	1128854.0	loss	46.948787	December 5, 1782	Martin Van Buren	Kinderhook	New York	(8th) March 4, 1837 – March 4, 1841

# Example (Left Join)

elections

	Year	Candidate	Party	Popular vote	Result	%
0	1824	Andrew Jackson	Democratic-Republican	151271	loss	57.210122
1	1824	John Quincy Adams	Democratic-Republican	113142	win	42.789878
2	1828	Andrew Jackson	Democratic	642806	win	56.203927
3	1828	John Quincy Adams	National Republican	500897	loss	43.796073
4	1832	Andrew Jackson	Democratic	702735	win	54.574789

presidents

	Date of birth	President	Birthplace	State of birth	In office
0	February 22, 1732	George Washington	Westmoreland County	Virginia	(1st) April 30, 1789 – March 4, 1797
1	October 30, 1735	John Adams	Braintree	Massachusetts	(2nd) March 4, 1797 – March 4, 1801
2	April 13, 1743	Thomas Jefferson	Shadwell	Virginia	(3rd) March 4, 1801 – March 4, 1809
3	March 16, 1751	James Madison	Port Conway	Virginia	(4th) March 4, 1809 – March 4, 1817
4	April 28, 1758	James Monroe	Monroe Hall	Virginia	(5th) March 4, 1817 – March 4, 1825

```
elections.merge(presidents,  
                how = "left",  
                left_on = "Candidate", right_on = "President")
```

	Year	Candidate	Party	Popular vote	Result	%	Date of birth	President	Birthplace	State of birth	In office
0	1824	Andrew Jackson	Democratic-Republican	151271	loss	57.210122	March 15, 1767	Andrew Jackson	Waxhaws Region	South/North Carolina	(7th) March 4, 1829 – March 4, 1837
1	1824	John Quincy Adams	Democratic-Republican	113142	win	42.789878	July 11, 1767	John Quincy Adams	Braintree	Massachusetts	(6th) March 4, 1825 – March 4, 1829
2	1828	Andrew Jackson	Democratic	642806	win	56.203927	March 15, 1767	Andrew Jackson	Waxhaws Region	South/North Carolina	(7th) March 4, 1829 – March 4, 1837
3	1828	John Quincy Adams	National Republican	500897	loss	43.796073	July 11, 1767	John Quincy Adams	Braintree	Massachusetts	(6th) March 4, 1825 – March 4, 1829
4	1832	Andrew Jackson	Democratic	702735	win	54.574789	March 15, 1767	Andrew Jackson	Waxhaws Region	South/North Carolina	(7th) March 4, 1829 – March 4, 1837
5	1832	Henry Clay	National Republican	484205	loss	37.603628	NaN	NaN	NaN	NaN	NaN
6	1832	William Wirt	Anti-Masonic	100715	loss	7.821583	NaN	NaN	NaN	NaN	NaN

# Example (Outer Join)

elections

	Year	Candidate	Party	Popular vote	Result	%
0	1824	Andrew Jackson	Democratic-Republican	151271	loss	57.210122
1	1824	John Quincy Adams	Democratic-Republican	113142	win	42.789878
2	1828	Andrew Jackson	Democratic	642806	win	56.203927
3	1828	John Quincy Adams	National Republican	500897	loss	43.796073
4	1832	Andrew Jackson	Democratic	702735	win	54.574789

presidents

	Date of birth	President	Birthplace	State of birth	In office
0	February 22, 1732	George Washington	Westmoreland County	Virginia	(1st) April 30, 1789 – March 4, 1797
1	October 30, 1735	John Adams	Braintree	Massachusetts	(2nd) March 4, 1797 – March 4, 1801
2	April 13, 1743	Thomas Jefferson	Shadwell	Virginia	(3rd) March 4, 1801 – March 4, 1809
3	March 16, 1751	James Madison	Port Conway	Virginia	(4th) March 4, 1809 – March 4, 1817
4	April 28, 1758	James Monroe	Monroe Hall	Virginia	(5th) March 4, 1817 – March 4, 1825

```
elections.merge(presidents,  
                how = "outer",  
                left_on = "Candidate", right_on = "President")
```

	Year	Candidate	Party	Popular vote	Result	%	Date of birth	President	Birthplace	State of birth	In office
0	1824.0	Andrew Jackson	Democratic-Republican	151271.0	loss	57.210122	March 15, 1767	Andrew Jackson	Waxhaws Region	South/North Carolina	(7th) March 4, 1829 – March 4, 1837
1	1828.0	Andrew Jackson	Democratic	642806.0	win	56.203927	March 15, 1767	Andrew Jackson	Waxhaws Region	South/North Carolina	(7th) March 4, 1829 – March 4, 1837
2	1832.0	Andrew Jackson	Democratic	702735.0	win	54.574789	March 15, 1767	Andrew Jackson	Waxhaws Region	South/North Carolina	(7th) March 4, 1829 – March 4, 1837
3	1824.0	John Quincy Adams	Democratic-Republican	113142.0	win	42.789878	July 11, 1767	John Quincy Adams	Braintree	Massachusetts	(6th) March 4, 1825 – March 4, 1829
4	1828.0	John Quincy Adams	National Republican	500897.0	loss	43.796073	July 11, 1767	John Quincy Adams	Braintree	Massachusetts	(6th) March 4, 1825 – March 4, 1829
5	1832.0	Henry Clay	National Republican	484205.0	loss	37.603628	NaN	NaN	NaN	NaN	NaN
6	1844.0	Henry Clay	Whig	1300004.0	loss	49.250523	NaN	NaN	NaN	NaN	NaN

# Example (Outer Join)

elections

	Year	Candidate	Party	Popular vote	Result	%
0	1824	Andrew Jackson	Democratic-Republican	151271	loss	57.210122
1	1824	John Quincy Adams	Democratic-Republican	113142	win	42.789878
2	1828	Andrew Jackson	Democratic	642806	win	56.203927
3	1828	John Quincy Adams	National Republican	500897	loss	43.796073
4	1832	Andrew Jackson	Democratic	702735	win	54.574789

presidents

	Date of birth	President	Birthplace	State of birth	In office
0	February 22, 1732	George Washington	Westmoreland County	Virginia	(1st) April 30, 1789 – March 4, 1797
1	October 30, 1735	John Adams	Braintree	Massachusetts	(2nd) March 4, 1797 – March 4, 1801
2	April 13, 1743	Thomas Jefferson	Shadwell	Virginia	(3rd) March 4, 1801 – March 4, 1809
3	March 16, 1751	James Madison	Port Conway	Virginia	(4th) March 4, 1809 – March 4, 1817
4	April 28, 1758	James Monroe	Monroe Hall	Virginia	(5th) March 4, 1817 – March 4, 1825

```
elections.merge(presidents,  
                how = "outer",  
                left_on = "Candidate", right_on = "President")
```

**left\_index** : *bool, default False*

Use the index from the left DataFrame as the join key(s). If it is a MultiIndex, the number of keys in the other DataFrame (either the index or a number of columns) must match the number of levels.

**right\_index** : *bool, default False*

Use the index from the right DataFrame as the join key. Same caveats as left\_index.