

Joining DataFrames (Concatenating & Merging), Reminders, and an Intro to Markdown

Concatenating two DataFrames:

pd.concat(<df1>, <df2>, <df3>)

df1					Result				
	A	B	C	D		A	B	C	D
0	A0	B0	C0	D0	0	A0	B0	C0	D0
1	A1	B1	C1	D1	1	A1	B1	C1	D1
2	A2	B2	C2	D2	2	A2	B2	C2	D2
3	A3	B3	C3	D3	3	A3	B3	C3	D3
df2					4	A4	B4	C4	D4
	A	B	C	D	5	A5	B5	C5	D5
4	A4	B4	C4	D4	6	A6	B6	C6	D6
5	A5	B5	C5	D5	7	A7	B7	C7	D7
6	A6	B6	C6	D6	8	A8	B8	C8	D8
7	A7	B7	C7	D7	9	A9	B9	C9	D9
df3					10	A10	B10	C10	D10
	A	B	C	D	11	A11	B11	C11	D11
8	A8	B8	C8	D8					
9	A9	B9	C9	D9					
10	A10	B10	C10	D10					
11	A11	B11	C11	D11					

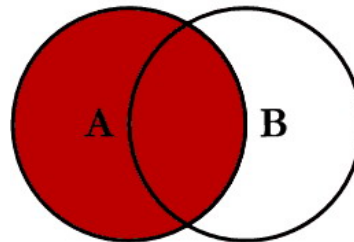
- Same data with same columns
- Like pasting them together

Merging two DataFrames:

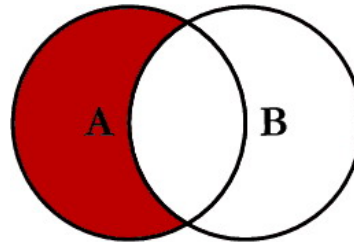
pd.merge(<df1>, <df2>, **on** = <col or list of cols to join on>, **how** = <join_type>)

- Different data with one or more columns in common

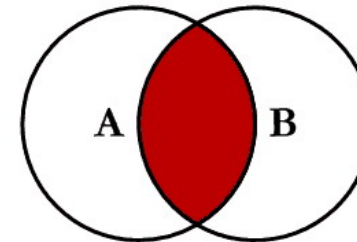
SQL JOINS



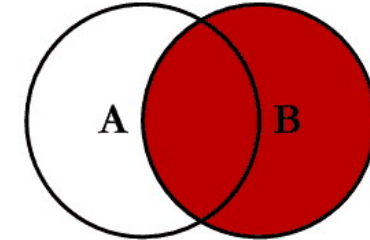
```
SELECT <select_list>  
FROM TableA A  
LEFT JOIN TableB B  
ON A.Key = B.Key
```



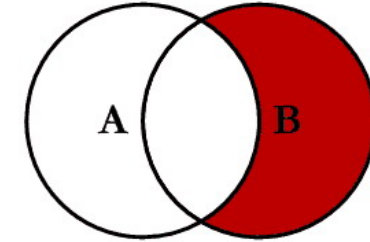
```
SELECT <select_list>  
FROM TableA A  
LEFT JOIN TableB B  
ON A.Key = B.Key  
WHERE B.Key IS NULL
```



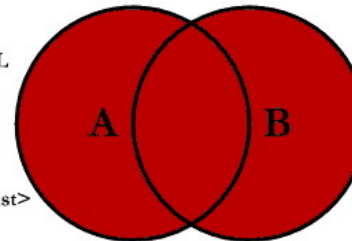
```
SELECT <select_list>  
FROM TableA A  
INNER JOIN TableB B  
ON A.Key = B.Key
```



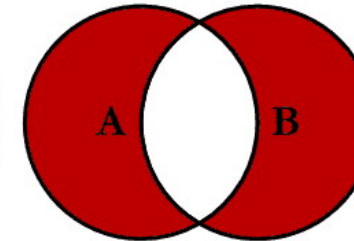
```
SELECT <select_list>  
FROM TableA A  
RIGHT JOIN TableB B  
ON A.Key = B.Key
```



```
SELECT <select_list>  
FROM TableA A  
RIGHT JOIN TableB B  
ON A.Key = B.Key  
WHERE A.Key IS NULL
```



```
SELECT <select_list>  
FROM TableA A  
FULL OUTER JOIN TableB B  
ON A.Key = B.Key
```



```
SELECT <select_list>  
FROM TableA A  
FULL OUTER JOIN TableB B  
ON A.Key = B.Key  
WHERE A.Key IS NULL  
OR B.Key IS NULL
```


Reminders

- Build upon your work in the same notebook each week (Github keeps a history)
- If the code in a cell did not run as expected, modify the code in that cell (not a new one)
- Remove any unused/un-useful cells
- ***Beware of the changing state of objects*** in your notebook
 - Example – if you create a df and drop 3 columns and then go back to add code to look at the head() in the same cell you are re-running the command to drop the 3 columns which are no longer there!

Markdown

- Markup language for formatting text
- Designed to be easily converted to HTML
- You can use HTML within markdown
- In Jupyter, markdown provides a way to *document* your workflow
- Improves reproducibility
- Look at your notebook with fresh eyes – if someone gave it to you and you knew nothing about the project, could you follow what is happening in each cell?

<https://github.com/adam-p/markdown-here/wiki/Markdown-Cheatsheet>

Text using Markdown syntax	Corresponding HTML produced by a Markdown processor	Text viewed in a browser
<p>Heading</p> <p>=====</p> <p>## Sub-heading</p> <p>Paragraphs are separated by a blank line.</p> <p>Two spaces at the end of a line produces a line break.</p> <p>Text attributes <i>_italic_</i>, **bold**, <code>`monospace`</code>.</p> <p>Horizontal rule:</p> <p>---</p> <p>Bullet list:</p> <ul style="list-style-type: none">* apples* oranges* pears <p>Numbered list:</p> <ol style="list-style-type: none">1. wash2. rinse3. repeat <p>A [link][example].</p> <p>[example]: http://example.com</p> <p>![Image](Image_icon.png "icon")</p> <p>> Markdown uses email-style > characters for blockquoting.</p> <p>Inline <abbr title="Hypertext Markup Language">HTML</abbr> is supported.</p>	<pre><h1>Heading</h1> <h2>Sub-heading</h2> <p>Paragraphs are separated by a blank line.</p> <p>Two spaces at the end of a line
 produces a line break.</p> <p>Text attributes italic, bold, <code>monospace</code>.</p> <p>Horizontal rule:</p> <hr /> <p>Bullet list:</p> apples oranges pears <p>Numbered list:</p> wash rinse repeat <p>A link.</p> <p></p> <blockquote> <p>Markdown uses email-style &gt; characters for blockquoting.</p> </blockquote> <p>Inline <abbr title="Hypertext Markup Language">HTML</abbr> is supported.</p></pre>	<h1>Heading</h1> <hr/> <h2>Sub-heading</h2> <hr/> <p>Paragraphs are separated by a blank line.</p> <p>Two spaces at the end of a line produces a line break.</p> <p>Text attributes <i>italic</i>, bold, <code>monospace</code> .</p> <p>Horizontal rule:</p> <hr/> <p>Bullet list:</p> <ul style="list-style-type: none">• apples• oranges• pears <p>Numbered list:</p> <ol style="list-style-type: none">1. wash2. rinse3. repeat <p>A link.</p>  <p>Markdown uses email-style > characters for blockquoting.</p> <p>Inline <code>HTML</code> is supported.</p> <p>https://en.wikipedia.org/wiki/Markdown</p>