3rd of August 2020

In [1]:

```
import pandas as pd
import numpy as np
import os
import sys
```

In [2]:

```
sys.version
```

Out[2]:

```
'3.7.6 (default, Jan  8 2020, 20:23:39) [MSC v.1916 64 bit (AMD64)]'
```

In [3]:

```
pd.__version__
```

Out[3]:

```
'1.0.1'
```

In [4]:

```
eng = pd.read_csv("eng.csv")
```

In [5]:

```
state = pd.read_csv("state.csv")
```

In [6]:

```
party = pd.read_csv("party.csv.csv")
```

In [7]:

```
liberal = pd.read_csv("liberal_arts.csv")
```

In [8]:

```
ivies = pd.read_csv("ivies.csv")
```

Concatenating DataFrames

In [9]:

```
dfs = [state, eng, liberal, ivies, party]
```

In [10]:

```
for df in dfs:
    print(df.shape)
```

```
(175, 4)
(19, 4)
(47, 4)
(8, 4)
(20, 4)
```

In [11]:

```
1 pd.concat([ivies, eng])
```

Out[11]:

	School Name	School Type	Starting Median Salary	Mid-Career Median Salary
0	Dartmouth College	Ivy League	$58,000.00	$134,000.00
1	Princeton University	Ivy League	$66,500.00	$131,000.00
2	Yale University	Ivy League	$59,100.00	$126,000.00
3	Harvard University	Ivy League	$63,400.00	$124,000.00
4	University of Pennsylvania	Ivy League	$60,900.00	$120,000.00
5	Cornell University	Ivy League	$60,300.00	$110,000.00
6	Brown University	Ivy League	$56,200.00	$109,000.00
7	Columbia University	Ivy League	$59,400.00	$107,000.00
0	Massachusetts Institute of Technology (MIT)	Engineering	$72,200.00	$126,000.00
1	California Institute of Technology (CIT)	Engineering	$75,500.00	$123,000.00
2	Harvey Mudd College	Engineering	$71,800.00	$122,000.00
3	Polytechnic University of New York, Brooklyn	Engineering	$62,400.00	$114,000.00
4	Cooper Union	Engineering	$62,200.00	$114,000.00
5	Worcester Polytechnic Institute (WPI)	Engineering	$61,000.00	$114,000.00
6	Carnegie Mellon University (CMU)	Engineering	$61,800.00	$111,000.00
7	Rensselaer Polytechnic Institute (RPI)	Engineering	$61,100.00	$110,000.00
8	Georgia Institute of Technology	Engineering	$58,300.00	$106,000.00
9	Colorado School of Mines	Engineering	$58,100.00	$106,000.00
10	Stevens Institute of Technology	Engineering	$60,600.00	$105,000.00
11	Illinois Institute of Technology (IIT)	Engineering	$56,000.00	$97,800.00
12	Wentworth Institute of Technology	Engineering	$53,000.00	$96,700.00
13	Virginia Polytechnic Institute and State Unive...	Engineering	$53,500.00	$95,400.00
14	South Dakota School of Mines & Technology	Engineering	$55,800.00	$93,400.00
15	New Mexico Institute of Mining and Technology ...	Engineering	$51,000.00	$93,400.00
16	Rochester Institute of Technology (RIT)	Engineering	$48,900.00	$84,600.00
17	Embry-Riddle Aeronautical University (ERAU)	Engineering	$52,700.00	$80,700.00
18	Tennessee Technological University	Engineering	$46,200.00	$80,000.00

In [12]:

```
pd.concat([ivies, eng]).shape
```

Out[12]:

```
(27, 4)
```

In [13]:

```
ivies.shape[0] + eng.shape[0]
```

Out[13]:

```
27
```

In [14]:

```
pd.concat(dfs) #->  this contains duplicate school names
```

Out[14]:

	School Name	School Type	Starting Median Salary	Mid-Career Median Salary
0	University of California, Berkeley	State	$59,900.00	$112,000.00
1	University of Virginia (UVA)	State	$52,700.00	$103,000.00
2	Cal Poly San Luis Obispo	State	$57,200.00	$101,000.00
3	University of California at Los Angeles (UCLA)	State	$52,600.00	$101,000.00
4	University of California, San Diego (UCSD)	State	$51,100.00	$101,000.00
...	...	...	...	...
15	University of New Hampshire (UNH)	Party	$41,800.00	$78,300.00
16	West Virginia University (WVU)	Party	$43,100.00	$78,100.00
17	University of Tennessee	Party	$43,800.00	$74,600.00
18	Ohio University	Party	$42,200.00	$73,400.00
19	Florida State University (FSU)	Party	$42,100.00	$73,000.00

269 rows × 4 columns

In [15]:

```
set(party['School Name']).difference(state['School Name'])
```

Out[15]:

```
{'Randolph-Macon College'}
```

In [16]:

```
'Randolph-Macon College' in liberal['School Name'].values
```

Out[16]:

```
True
```

In [17]:

```
1 liberal['School Name'].values
```

Out[17]:

```
array(['Bucknell University', 'Colgate University', 'Amherst College',
       'Lafayette College', 'Bowdoin College',
       'College of the Holy Cross', 'Occidental College',
       'Washington and Lee University', 'Swarthmore College',
       'Davidson College', 'Carleton College', 'Williams College',
       'Pomona College', 'Wesleyan University (Middletown, Connecticut)',
       'Bates College', 'Union College', 'University of Richmond',
       'Vassar College', 'Middlebury College', 'Mount Holyoke College',
       'Franklin and Marshall College', 'DePauw University',
       'St. Olaf College', 'Colby College', 'Gettysburg College',
       'Siena College', 'Smith College', 'Hamilton College',
       'Randolph-Macon College', 'Wellesley College',
       'Denison University', 'Oberlin College',
       'University of Puget Sound', 'Colorado College (CC)',
       'Reed College', 'Gustavus Adolphus College', 'Whitman College',
       'Ursinus College', 'Juniata College', 'Wittenberg University',
       'Grinnell College', 'Skidmore College', 'Moravian College',
       'Lewis & Clark College', 'Fort Lewis College',
       'Thomas Aquinas College', 'Evergreen State College'], dtype=object)
```

In [18]:

```
1 liberal
```

Out[18]:

	School Name	School Type	Starting Median Salary	Mid-Career Median Salary
0	Bucknell University	Liberal Arts	$54,100.00	$110,000.00
1	Colgate University	Liberal Arts	$52,800.00	$108,000.00
2	Amherst College	Liberal Arts	$54,500.00	$107,000.00
3	Lafayette College	Liberal Arts	$53,900.00	$107,000.00
4	Bowdoin College	Liberal Arts	$48,100.00	$107,000.00
5	College of the Holy Cross	Liberal Arts	$50,200.00	$106,000.00
6	Occidental College	Liberal Arts	$51,900.00	$105,000.00
7	Washington and Lee University	Liberal Arts	$53,600.00	$104,000.00
8	Swarthmore College	Liberal Arts	$49,700.00	$104,000.00
9	Davidson College	Liberal Arts	$46,100.00	$104,000.00
10	Carleton College	Liberal Arts	$47,500.00	$103,000.00
11	Williams College	Liberal Arts	$51,700.00	$102,000.00
12	Pomona College	Liberal Arts	$48,600.00	$101,000.00
13	Wesleyan University (Middletown, Connecticut)	Liberal Arts	$46,500.00	$97,900.00
14	Bates College	Liberal Arts	$47,300.00	$96,500.00
15	Union College	Liberal Arts	$47,200.00	$95,800.00
16	University of Richmond	Liberal Arts	$48,600.00	$94,600.00
17	Vassar College	Liberal Arts	$46,000.00	$94,600.00
18	Middlebury College	Liberal Arts	$47,700.00	$94,200.00
19	Mount Holyoke College	Liberal Arts	$42,400.00	$94,100.00
20	Franklin and Marshall College	Liberal Arts	$49,100.00	$92,800.00
21	DePauw University	Liberal Arts	$41,400.00	$88,300.00

	School Name	School Type	Starting Median Salary	Mid-Career Median Salary
22	St. Olaf College	Liberal Arts	$45,300.00	$86,200.00
23	Colby College	Liberal Arts	$46,400.00	$85,800.00
24	Gettysburg College	Liberal Arts	$44,700.00	$85,800.00
25	Siena College	Liberal Arts	$45,500.00	$85,200.00
26	Smith College	Liberal Arts	$44,000.00	$83,900.00
27	Hamilton College	Liberal Arts	$49,200.00	$83,700.00
28	Randolph-Macon College	Liberal Arts	$42,600.00	$83,600.00
29	Wellesley College	Liberal Arts	$42,800.00	$83,500.00
30	Denison University	Liberal Arts	$42,000.00	$83,500.00
31	Oberlin College	Liberal Arts	$43,400.00	$81,600.00
32	University of Puget Sound	Liberal Arts	$46,600.00	$81,500.00
33	Colorado College (CC)	Liberal Arts	$38,500.00	$81,400.00
34	Reed College	Liberal Arts	$40,500.00	$81,100.00
35	Gustavus Adolphus College	Liberal Arts	$44,500.00	$80,600.00
36	Whitman College	Liberal Arts	$43,500.00	$80,100.00
37	Ursinus College	Liberal Arts	$42,100.00	$80,000.00
38	Juniata College	Liberal Arts	$41,800.00	$78,900.00
39	Wittenberg University	Liberal Arts	$39,200.00	$78,200.00
40	Grinnell College	Liberal Arts	$42,600.00	$76,600.00
41	Skidmore College	Liberal Arts	$41,600.00	$74,600.00
42	Moravian College	Liberal Arts	$42,500.00	$74,400.00
43	Lewis & Clark College	Liberal Arts	$38,900.00	$72,600.00
44	Fort Lewis College	Liberal Arts	$42,000.00	$69,800.00
45	Thomas Aquinas College	Liberal Arts	$41,500.00	$67,500.00
46	Evergreen State College	Liberal Arts	$39,500.00	$63,900.00

In [19]:

```
1 pd.concat(dfs).duplicated(subset=['School Name'], keep='first')
```

Out[19]:

```
0     False
1     False
2     False
3     False
4     False
      ...
15     True
16     True
17     True
18     True
19     True
Length: 269, dtype: bool
```

In [20]:

```
1 pd.concat(dfs)[pd.concat(dfs).duplicated(subset=['School Name'], keep='first')]
```

Out[20]:

	School Name	School Type	Starting Median Salary	Mid-Career Median Salary
0	University of Illinois at Urbana-Champaign (UIUC)	Party	$52,900.00	$96,100.00
1	University of Maryland, College Park	Party	$52,000.00	$95,000.00
2	University of California, Santa Barbara (UCSB)	Party	$50,500.00	$95,000.00
3	University of Texas (UT) - Austin	Party	$49,700.00	$93,900.00
4	State University of New York (SUNY) at Albany	Party	$44,500.00	$92,200.00
5	University of Florida (UF)	Party	$47,100.00	$87,900.00
6	Louisiana State University (LSU)	Party	$46,900.00	$87,800.00
7	University of Georgia (UGA)	Party	$44,100.00	$86,000.00

In [21]:

```
1 pd.concat(dfs)[pd.concat(dfs).duplicated(subset=['School Name'], keep='first')].shape
```

Out[21]:

```
(20, 4)
```

In [22]:

```
1 pd.concat(dfs).drop_duplicates(subset='School Name')
```

Out[22]:

	School Name	School Type	Starting Median Salary	Mid-Career Median Salary
0	University of California, Berkeley	State	$59,900.00	$112,000.00
1	University of Virginia (UVA)	State	$52,700.00	$103,000.00
2	Cal Poly San Luis Obispo	State	$57,200.00	$101,000.00
3	University of California at Los Angeles (UCLA)	State	$52,600.00	$101,000.00
4	University of California, San Diego (UCSD)	State	$51,100.00	$101,000.00
...	...	...	...	...
3	Harvard University	Ivy League	$63,400.00	$124,000.00
4	University of Pennsylvania	Ivy League	$60,900.00	$120,000.00
5	Cornell University	Ivy League	$60,300.00	$110,000.00

In [23]:

```
1 schools = pd.concat(dfs).drop_duplicates(subset='School Name')
```

In [24]:

```
1 schools
```

Out[24]:

	School Name	School Type	Starting Median Salary	Mid-Career Median Salary
0	University of California, Berkeley	State	$59,900.00	$112,000.00
1	University of Virginia (UVA)	State	$52,700.00	$103,000.00
2	Cal Poly San Luis Obispo	State	$57,200.00	$101,000.00
3	University of California at Los Angeles (UCLA)	State	$52,600.00	$101,000.00
4	University of California, San Diego (UCSD)	State	$51,100.00	$101,000.00
...	...	...	...	...
3	Harvard University	Ivy League	$63,400.00	$124,000.00
4	University of Pennsylvania	Ivy League	$60,900.00	$120,000.00
5	Cornell University	Ivy League	$60,300.00	$110,000.00
6	Brown University	Ivy League	$56,200.00	$109,000.00
7	Columbia University	Ivy League	$59,400.00	$107,000.00

249 rows × 4 columns

The dupliacted index issue

In [25]:

```
schools.loc[0]
```

Out[25]:

	School Name	School Type	Starting Median Salary	Mid-Career Median Salary
0	University of California, Berkeley	State	$59,900.00	$112,000.00
0	Massachusetts Institute of Technology (MIT)	Engineering	$72,200.00	$126,000.00
0	Bucknell University	Liberal Arts	$54,100.00	$110,000.00
0	Dartmouth College	Ivy League	$58,000.00	$134,000.00

pd.concat() does not discard the original index of the dataframes being concatenated

In [26]:

```
schools.index.duplicated()
```

Out[26]:

```
array([False, False, False, False, False, False, False, False, False,
       False, False, False, False, False, False, False, False, False,
       False, False, False, False, False, False, False, False, False,
       False, False, False, False, False, False, False, False, False,
       False, False, False, False, False, False, False, False, False,
       False, False, False, False, False, False, False, False, False,
       False, False, False, False, False, False, False, False, False,
       False, False, False, False, False, False, False, False, False,
       False, False, False, False, False, False, False, False, False,
       False, False, False, False, False, False, False, False, False,
       False, False, False, False, False, False, False, False, False,
       False, False, False, False, False, False, False, False, False,
       False, False, False, False, False, False, False, False, False,
       False, False, False, False, False, False, False, False, False,
       False, False, False, False, False, False, False, False, False,
       False, False, False, False, False, False, False, False, False,
       False, False, False, False, False, False, False, False, False,
       False, False, False, False, False, False, False, False, False,
       False, False, False, False, False, False, False, False, False,
       False, False, False, False,  True,  True,  True,  True,  True,
        True,  True,  True,  True,  True,  True,  True,  True,  True,
        True,  True,  True,  True,  True,  True,  True,  True,  True,
        True,  True,  True,  True,  True,  True,  True,  True,  True,
        True,  True,  True,  True,  True,  True,  True,  True,  True,
        True,  True,  True,  True,  True,  True,  True,  True,  True,
        True,  True,  True,  True,  True,  True,  True,  True,  True,
        True,  True,  True,  True,  True,  True,  True,  True,  True,
        True,  True,  True,  True,  True,  True])
```

In [27]:

```
#schools.loc[0:2] -> this won't work anymore because multiple rows have the same index
```

In [28]:

```
1 schools.iloc[0]
```

Out[28]:

```
School Name                 University of California, Berkeley
School Type                                              State
Starting Median Salary                             $59,900.00
Mid-Career Median Salary                          $112,000.00
Name: 0, dtype: object
```

In [29]:

```
1 schools.iloc[0:2] # this works because it is indexing by position and not by label
```

Out[29]:

	School Name	School Type	Starting Median Salary	Mid-Career Median Salary
0	University of California, Berkeley	State	$59,900.00	$112,000.00
1	University of Virginia (UVA)	State	$52,700.00	$103,000.00

In [30]:

```
1 #the first fix to our schools.loc[0:2] problem
```

In [31]:

```
1 schools.reset_index(drop =True, inplace=True)
```

In [32]:

```
schools.index.duplicated()
```

Out[32]:

```
array([False, False, False, False, False, False, False, False, False,
       False, False, False, False, False, False, False, False, False,
       False, False, False, False, False, False, False, False, False,
       False, False, False, False, False, False, False, False, False,
       False, False, False, False, False, False, False, False, False,
       False, False, False, False, False, False, False, False, False,
       False, False, False, False, False, False, False, False, False,
       False, False, False, False, False, False, False, False, False,
       False, False, False, False, False, False, False, False, False,
       False, False, False, False, False, False, False, False, False,
       False, False, False, False, False, False, False, False, False,
       False, False, False, False, False, False, False, False, False,
       False, False, False, False, False, False, False, False, False,
       False, False, False, False, False, False, False, False, False,
       False, False, False, False, False, False, False, False, False,
       False, False, False, False, False, False, False, False, False,
       False, False, False, False, False, False, False, False, False,
       False, False, False, False, False, False, False, False, False,
       False, False, False, False, False, False, False, False, False,
       False, False, False, False, False, False, False, False, False,
       False, False, False, False, False, False, False, False, False,
       False, False, False, False, False, False, False, False, False,
       False, False, False, False, False, False, False, False, False,
       False, False, False, False, False, False, False, False, False,
       False, False, False, False, False, False, False, False, False,
       False, False, False, False, False, False, False, False, False,
       False, False, False, False, False, False, False, False, False,
       False, False, False, False, False, False])
```

In [33]:

```
schools.index.duplicated().sum()
```

Out[33]:

```
0
```

In [34]:

```
schools.loc[0:2]
```

Out[34]:

	School Name	School Type	Starting Median Salary	Mid-Career Median Salary
0	University of California, Berkeley	State	$59,900.00	$112,000.00
1	University of Virginia (UVA)	State	$52,700.00	$103,000.00
2	Cal Poly San Luis Obispo	State	$57,200.00	$101,000.00

In [35]:

```
#second/alternative fix to the problem
```

In [36]:

```
1 pd.concat(dfs).drop_duplicates(subset=['School Name'])
```

Out[36]:

	School Name	School Type	Starting Median Salary	Mid-Career Median Salary
0	University of California, Berkeley	State	$59,900.00	$112,000.00
1	University of Virginia (UVA)	State	$52,700.00	$103,000.00
2	Cal Poly San Luis Obispo	State	$57,200.00	$101,000.00
3	University of California at Los Angeles (UCLA)	State	$52,600.00	$101,000.00
4	University of California, San Diego (UCSD)	State	$51,100.00	$101,000.00
...	...	...	...	...
3	Harvard University	Ivy League	$63,400.00	$124,000.00
4	University of Pennsylvania	Ivy League	$60,900.00	$120,000.00
5	Cornell University	Ivy League	$60,300.00	$110,000.00
6	Brown University	Ivy League	$56,200.00	$109,000.00
7	Columbia University	Ivy League	$59,400.00	$107,000.00

249 rows × 4 columns

In [37]:

```
1 pd.concat(dfs, ignore_index=True).drop_duplicates(subset=['School Name'])
```

Out[37]:

	School Name	School Type	Starting Median Salary	Mid-Career Median Salary
0	University of California, Berkeley	State	$59,900.00	$112,000.00
1	University of Virginia (UVA)	State	$52,700.00	$103,000.00
2	Cal Poly San Luis Obispo	State	$57,200.00	$101,000.00
3	University of California at Los Angeles (UCLA)	State	$52,600.00	$101,000.00
4	University of California, San Diego (UCSD)	State	$51,100.00	$101,000.00
...	...	...	...	...
244	Harvard University	Ivy League	$63,400.00	$124,000.00
245	University of Pennsylvania	Ivy League	$60,900.00	$120,000.00
246	Cornell University	Ivy League	$60,300.00	$110,000.00
247	Brown University	Ivy League	$56,200.00	$109,000.00
248	Columbia University	Ivy League	$59,400.00	$107,000.00

249 rows × 4 columns

In [38]:

```
1 pd.concat(dfs, ignore_index=True).drop_duplicates(subset=['School Name']).index.duplica
```

Out[38]:

```
0
```

Enforcing Unique Indices

Goal: What if we wanted to preserve the index and force Uniqueness

In [39]:

```
1 ivies.set_index('School Name')
```

Out[39]:

School Name	School Type	Starting Median Salary	Mid-Career Median Salary
Dartmouth College	Ivy League	$58,000.00	$134,000.00
Princeton University	Ivy League	$66,500.00	$131,000.00
Yale University	Ivy League	$59,100.00	$126,000.00
Harvard University	Ivy League	$63,400.00	$124,000.00
University of Pennsylvania	Ivy League	$60,900.00	$120,000.00
Cornell University	Ivy League	$60,300.00	$110,000.00
Brown University	Ivy League	$56,200.00	$109,000.00
Columbia University	Ivy League	$59,400.00	$107,000.00

In [40]:

```
1 ivies2 = ivies.set_index('School Name')
```

In [41]:

```
1 eng.set_index('School Name')
```

Out[41]:

School Name	School Type	Starting Median Salary	Mid-Career Median Salary
Massachusetts Institute of Technology (MIT)	Engineering	$72,200.00	$126,000.00
California Institute of Technology (CIT)	Engineering	$75,500.00	$123,000.00
Harvey Mudd College	Engineering	$71,800.00	$122,000.00
Polytechnic University of New York, Brooklyn	Engineering	$62,400.00	$114,000.00
Cooper Union	Engineering	$62,200.00	$114,000.00
Worcester Polytechnic Institute (WPI)	Engineering	$61,000.00	$114,000.00
Carnegie Mellon University (CMU)	Engineering	$61,800.00	$111,000.00
Rensselaer Polytechnic Institute (RPI)	Engineering	$61,100.00	$110,000.00
Georgia Institute of Technology	Engineering	$58,300.00	$106,000.00
Colorado School of Mines	Engineering	$58,100.00	$106,000.00
Stevens Institute of Technology	Engineering	$60,600.00	$105,000.00
Illinois Institute of Technology (IIT)	Engineering	$56,000.00	$97,800.00
Wentworth Institute of Technology	Engineering	$53,000.00	$96,700.00
Virginia Polytechnic Institute and State University (Virginia Tech)	Engineering	$53,500.00	$95,400.00
South Dakota School of Mines & Technology	Engineering	$55,800.00	$93,400.00
New Mexico Institute of Mining and Technology (New Mexico Tech)	Engineering	$51,000.00	$93,400.00
Rochester Institute of Technology (RIT)	Engineering	$48,900.00	$84,600.00
Embry-Riddle Aeronautical University (ERAU)	Engineering	$52,700.00	$80,700.00
Tennessee Technological University	Engineering	$46,200.00	$80,000.00

In [42]:

```
1 eng2 = eng.set_index('School Name')
```

In [43]:

```
1 eng2.head()
```

Out[43]:

School Name	School Type	Starting Median Salary	Mid-Career Median Salary
Massachusetts Institute of Technology (MIT)	Engineering	$72,200.00	$126,000.00
California Institute of Technology (CIT)	Engineering	$75,500.00	$123,000.00
Harvey Mudd College	Engineering	$71,800.00	$122,000.00
Polytechnic University of New York, Brooklyn	Engineering	$62,400.00	$114,000.00
Cooper Union	Engineering	$62,200.00	$114,000.00

In [44]:

```
1 pd.concat([ivies2, eng2])
```

Out[44]:

School Name	School Type	Starting Median Salary	Mid-Career Median Salary
Dartmouth College	Ivy League	$58,000.00	$134,000.00
Princeton University	Ivy League	$66,500.00	$131,000.00
Yale University	Ivy League	$59,100.00	$126,000.00
Harvard University	Ivy League	$63,400.00	$124,000.00
University of Pennsylvania	Ivy League	$60,900.00	$120,000.00
Cornell University	Ivy League	$60,300.00	$110,000.00
Brown University	Ivy League	$56,200.00	$109,000.00
Columbia University	Ivy League	$59,400.00	$107,000.00
Massachusetts Institute of Technology (MIT)	Engineering	$72,200.00	$126,000.00
California Institute of Technology (CIT)	Engineering	$75,500.00	$123,000.00
Harvey Mudd College	Engineering	$71,800.00	$122,000.00
Polytechnic University of New York, Brooklyn	Engineering	$62,400.00	$114,000.00
Cooper Union	Engineering	$62,200.00	$114,000.00
Worcester Polytechnic Institute (WPI)	Engineering	$61,000.00	$114,000.00
Carnegie Mellon University (CMU)	Engineering	$61,800.00	$111,000.00
Rensselaer Polytechnic Institute (RPI)	Engineering	$61,100.00	$110,000.00
Georgia Institute of Technology	Engineering	$58,300.00	$106,000.00
Colorado School of Mines	Engineering	$58,100.00	$106,000.00
Stevens Institute of Technology	Engineering	$60,600.00	$105,000.00
Illinois Institute of Technology (IIT)	Engineering	$56,000.00	$97,800.00
Wentworth Institute of Technology	Engineering	$53,000.00	$96,700.00
Virginia Polytechnic Institute and State University (Virginia Tech)	Engineering	$53,500.00	$95,400.00
South Dakota School of Mines & Technology	Engineering	$55,800.00	$93,400.00
New Mexico Institute of Mining and Technology (New Mexico Tech)	Engineering	$51,000.00	$93,400.00
Rochester Institute of Technology (RIT)	Engineering	$48,900.00	$84,600.00
Embry-Riddle Aeronautical University (ERAU)	Engineering	$52,700.00	$80,700.00
Tennessee Technological University	Engineering	$46,200.00	$80,000.00

In [45]:

```
1 pd.concat([ivies2, eng2], verify_integrity=True)
```

Out[45]:

School Name	School Type	Starting Median Salary	Mid-Career Median Salary
Dartmouth College	Ivy League	$58,000.00	$134,000.00
Princeton University	Ivy League	$66,500.00	$131,000.00
Yale University	Ivy League	$59,100.00	$126,000.00
Harvard University	Ivy League	$63,400.00	$124,000.00
University of Pennsylvania	Ivy League	$60,900.00	$120,000.00
Cornell University	Ivy League	$60,300.00	$110,000.00
Brown University	Ivy League	$56,200.00	$109,000.00
Columbia University	Ivy League	$59,400.00	$107,000.00
Massachusetts Institute of Technology (MIT)	Engineering	$72,200.00	$126,000.00

In [46]:

```
1 random_eng_school = eng2.sample()
```

In [47]:

```
1 random_eng_school
```

Out[47]:

School Name	School Type	Starting Median Salary	Mid-Career Median Salary
Worcester Polytechnic Institute (WPI)	Engineering	$61,000.00	$114,000.00

In [48]:

```
1 ivies2.append(random_eng_school)
```

Out[48]:

School Name	School Type	Starting Median Salary	Mid-Career Median Salary
Dartmouth College	Ivy League	$58,000.00	$134,000.00
Princeton University	Ivy League	$66,500.00	$131,000.00
Yale University	Ivy League	$59,100.00	$126,000.00
Harvard University	Ivy League	$63,400.00	$124,000.00
University of Pennsylvania	Ivy League	$60,900.00	$120,000.00
Cornell University	Ivy League	$60,300.00	$110,000.00
Brown University	Ivy League	$56,200.00	$109,000.00
Columbia University	Ivy League	$59,400.00	$107,000.00
Worcester Polytechnic Institute (WPI)	Engineering	$61,000.00	$114,000.00

In [49]:

```
1 ivies2 = ivies2.append(random_eng_school)
```

In [50]:

```
1 #pd.concat([ivies2, eng2], verify_integrity=True)
```

Creating Multiple Indices with concat()

How about a MultiIndex?

In [51]:

```
1 pd.concat([ivies, eng])
```

Out[51]:

	School Name	School Type	Starting Median Salary	Mid-Career Median Salary
0	Dartmouth College	Ivy League	$58,000.00	$134,000.00
1	Princeton University	Ivy League	$66,500.00	$131,000.00
2	Yale University	Ivy League	$59,100.00	$126,000.00
3	Harvard University	Ivy League	$63,400.00	$124,000.00
4	University of Pennsylvania	Ivy League	$60,900.00	$120,000.00
5	Cornell University	Ivy League	$60,300.00	$110,000.00
6	Brown University	Ivy League	$56,200.00	$109,000.00
7	Columbia University	Ivy League	$59,400.00	$107,000.00
0	Massachusetts Institute of Technology (MIT)	Engineering	$72,200.00	$126,000.00

In [52]:

```
1 pd.concat([ivies, eng], keys=['ivyleague_schools', 'engineering schools'])
2 #the keys parameter allows us to name each dataframe index that is about to be concater
```

Out[52]:

		School Name	School Type	Starting Median Salary	Mid-Career Median Salary
ivyleague_schools	**0**	Dartmouth College	Ivy League	$58,000.00	$134,000.00
	1	Princeton University	Ivy League	$66,500.00	$131,000.00
	2	Yale University	Ivy League	$59,100.00	$126,000.00
	3	Harvard University	Ivy League	$63,400.00	$124,000.00
	4	University of Pennsylvania	Ivy League	$60,900.00	$120,000.00
	5	Cornell University	Ivy League	$60,300.00	$110,000.00
	6	Brown University	Ivy League	$56,200.00	$109,000.00
	7	Columbia University	Ivy League	$59,400.00	$107,000.00
engineering schools	**0**	Massachusetts Institute of Technology (MIT)	Engineering	$72,200.00	$126,000.00

In [53]:

```
1 new_df = pd.concat([ivies, eng], keys=['ivyleague_schools', 'engineering schools'])
```

In [54]:

```
1 type(new_df.index)
```

Out[54]:

```
pandas.core.indexes.multi.MultiIndex
```

In [55]:

```
1 type(pd.concat([ivies, eng]).index)
```

Out[55]:

```
pandas.core.indexes.numeric.Int64Index
```

In [56]:

```
1 new_df.head()
```

Out[56]:

		School Name	School Type	Starting Median Salary	Mid-Career Median Salary
ivyleague_schools	**0**	Dartmouth College	Ivy League	$58,000.00	$134,000.00
	1	Princeton University	Ivy League	$66,500.00	$131,000.00
	2	Yale University	Ivy League	$59,100.00	$126,000.00
	3	Harvard University	Ivy League	$63,400.00	$124,000.00
	4	University of Pennsylvania	Ivy League	$60,900.00	$120,000.00

In [57]:

```
1 new_df.loc[('ivyleague_schools',3)]
```

Out[57]:

```
School Name                 Harvard University
School Type                         Ivy League
Starting Median Salary             $63,400.00
Mid-Career Median Salary          $124,000.00
Name: (ivyleague_schools, 3), dtype: object
```

In [58]:

```
1 type(new_df.loc[('ivyleague_schools',3)])
```

Out[58]:

```
pandas.core.series.Series
```

Column Axis Concatenation

In [59]:

```
1 schools
```

Out[59]:

	School Name	School Type	Starting Median Salary	Mid-Career Median Salary
0	University of California, Berkeley	State	$59,900.00	$112,000.00
1	University of Virginia (UVA)	State	$52,700.00	$103,000.00
2	Cal Poly San Luis Obispo	State	$57,200.00	$101,000.00
3	University of California at Los Angeles (UCLA)	State	$52,600.00	$101,000.00
4	University of California, San Diego (UCSD)	State	$51,100.00	$101,000.00
...	...	...	...	...
244	Harvard University	Ivy League	$63,400.00	$124,000.00
245	University of Pennsylvania	Ivy League	$60,900.00	$120,000.00
246	Cornell University	Ivy League	$60,300.00	$110,000.00
247	Brown University	Ivy League	$56,200.00	$109,000.00
248	Columbia University	Ivy League	$59,400.00	$107,000.00

249 rows × 4 columns

In [60]:

```
1 ivies.sort_values(by=['Starting Median Salary'], ascending=False)
```

Out[60]:

	School Name	School Type	Starting Median Salary	Mid-Career Median Salary
1	Princeton University	Ivy League	$66,500.00	$131,000.00
3	Harvard University	Ivy League	$63,400.00	$124,000.00
4	University of Pennsylvania	Ivy League	$60,900.00	$120,000.00
5	Cornell University	Ivy League	$60,300.00	$110,000.00
7	Columbia University	Ivy League	$59,400.00	$107,000.00
2	Yale University	Ivy League	$59,100.00	$126,000.00
0	Dartmouth College	Ivy League	$58,000.00	$134,000.00
6	Brown University	Ivy League	$56,200.00	$109,000.00

In [61]:

```
ivies.sort_values(by=['Starting Median Salary'], ascending=False)[:5].reset_index(drop=
```

Out[61]:

	School Name	School Type	Starting Median Salary	Mid-Career Median Salary
0	Princeton University	Ivy League	$66,500.00	$131,000.00
1	Harvard University	Ivy League	$63,400.00	$124,000.00
2	University of Pennsylvania	Ivy League	$60,900.00	$120,000.00
3	Cornell University	Ivy League	$60,300.00	$110,000.00
4	Columbia University	Ivy League	$59,400.00	$107,000.00

In [62]:

```
ivies.reset_index()
```

Out[62]:

	index	School Name	School Type	Starting Median Salary	Mid-Career Median Salary
0	0	Dartmouth College	Ivy League	$58,000.00	$134,000.00
1	1	Princeton University	Ivy League	$66,500.00	$131,000.00
2	2	Yale University	Ivy League	$59,100.00	$126,000.00
3	3	Harvard University	Ivy League	$63,400.00	$124,000.00
4	4	University of Pennsylvania	Ivy League	$60,900.00	$120,000.00
5	5	Cornell University	Ivy League	$60,300.00	$110,000.00
6	6	Brown University	Ivy League	$56,200.00	$109,000.00
7	7	Columbia University	Ivy League	$59,400.00	$107,000.00

In [63]:

```
eng.sort_values(by=['Starting Median Salary'], ascending=False)[:5].reset_index(drop=Tr
```

Out[63]:

	School Name	School Type	Starting Median Salary	Mid-Career Median Salary
0	California Institute of Technology (CIT)	Engineering	$75,500.00	$123,000.00
1	Massachusetts Institute of Technology (MIT)	Engineering	$72,200.00	$126,000.00
2	Harvey Mudd College	Engineering	$71,800.00	$122,000.00
3	Polytechnic University of New York, Brooklyn	Engineering	$62,400.00	$114,000.00
4	Cooper Union	Engineering	$62,200.00	$114,000.00

In [64]:

```
ivies3 = ivies.sort_values(by=['Starting Median Salary'], ascending=False)[:5].reset_ir
```

In [65]:

```
eng3 = eng.sort_values(by=['Starting Median Salary'], ascending=False)[:5].reset_index(
```

In [66]:

```
pd.concat([ivies3, eng3], axis=0)
```

Out[66]:

	School Name	School Type	Starting Median Salary	Mid-Career Median Salary
0	Princeton University	Ivy League	$66,500.00	$131,000.00
1	Harvard University	Ivy League	$63,400.00	$124,000.00
2	University of Pennsylvania	Ivy League	$60,900.00	$120,000.00
3	Cornell University	Ivy League	$60,300.00	$110,000.00
4	Columbia University	Ivy League	$59,400.00	$107,000.00
0	California Institute of Technology (CIT)	Engineering	$75,500.00	$123,000.00
1	Massachusetts Institute of Technology (MIT)	Engineering	$72,200.00	$126,000.00
2	Harvey Mudd College	Engineering	$71,800.00	$122,000.00
3	Polytechnic University of New York, Brooklyn	Engineering	$62,400.00	$114,000.00
4	Cooper Union	Engineering	$62,200.00	$114,000.00

In [67]:

```
1 pd.concat([ivies3, eng3], axis=1)
```

Out[67]:

	School Name	School Type	Starting Median Salary	Mid-Career Median Salary	School Name	School Type	Starting Median Salary	Mid-Ca Median Salary
0	Princeton University	Ivy League	$66,500.00	$131,000.00	California Institute of Technology (CIT)	Engineering	$75,500.00	$123,0(
1	Harvard University	Ivy League	$63,400.00	$124,000.00	Massachusetts Institute of Technology (MIT)	Engineering	$72,200.00	$126,0(
2	University of Pennsylvania	Ivy League	$60,900.00	$120,000.00	Harvey Mudd College	Engineering	$71,800.00	$122,0(
3	Cornell University	Ivy League	$60,300.00	$110,000.00	Polytechnic University of New York, Brooklyn	Engineering	$62,400.00	$114,0(
4	Columbia University	Ivy League	$59,400.00	$107,000.00	Cooper Union	Engineering	$62,200.00	$114,0(

The append() Method: A special case of concat()

In [68]:

```
1 liberal.head()
```

Out[68]:

	School Name	School Type	Starting Median Salary	Mid-Career Median Salary
0	Bucknell University	Liberal Arts	$54,100.00	$110,000.00
1	Colgate University	Liberal Arts	$52,800.00	$108,000.00
2	Amherst College	Liberal Arts	$54,500.00	$107,000.00
3	Lafayette College	Liberal Arts	$53,900.00	$107,000.00
4	Bowdoin College	Liberal Arts	$48,100.00	$107,000.00

In [69]:

```
1 liberal.append(party)
```

Out[69]:

	School Name	School Type	Starting Median Salary	Mid-Career Median Salary
0	Bucknell University	Liberal Arts	$54,100.00	$110,000.00
1	Colgate University	Liberal Arts	$52,800.00	$108,000.00
2	Amherst College	Liberal Arts	$54,500.00	$107,000.00
3	Lafayette College	Liberal Arts	$53,900.00	$107,000.00
4	Bowdoin College	Liberal Arts	$48,100.00	$107,000.00
...	...	...	...	...
15	University of New Hampshire (UNH)	Party	$41,800.00	$78,300.00
16	West Virginia University (WVU)	Party	$43,100.00	$78,100.00
17	University of Tennessee	Party	$43,800.00	$74,600.00

In [70]:

```
1 pd.concat([liberal, party])
```

Out[70]:

	School Name	School Type	Starting Median Salary	Mid-Career Median Salary
0	Bucknell University	Liberal Arts	$54,100.00	$110,000.00
1	Colgate University	Liberal Arts	$52,800.00	$108,000.00
2	Amherst College	Liberal Arts	$54,500.00	$107,000.00
3	Lafayette College	Liberal Arts	$53,900.00	$107,000.00
4	Bowdoin College	Liberal Arts	$48,100.00	$107,000.00
...	...	...	...	...
15	University of New Hampshire (UNH)	Party	$41,800.00	$78,300.00
16	West Virginia University (WVU)	Party	$43,100.00	$78,100.00
17	University of Tennessee	Party	$43,800.00	$74,600.00
18	Ohio University	Party	$42,200.00	$73,400.00
19	Florida State University (FSU)	Party	$42,100.00	$73,000.00

67 rows × 4 columns

append is a DataFrame instance method
append only operates along the index axis(row)

Concat() on different columns

In [71]:

```
1 ivies.head()
```

Out[71]:

	School Name	School Type	Starting Median Salary	Mid-Career Median Salary
0	Dartmouth College	Ivy League	$58,000.00	$134,000.00
1	Princeton University	Ivy League	$66,500.00	$131,000.00
2	Yale University	Ivy League	$59,100.00	$126,000.00
3	Harvard University	Ivy League	$63,400.00	$124,000.00
4	University of Pennsylvania	Ivy League	$60,900.00	$120,000.00

In [72]:

```
1 eng.head()
```

Out[72]:

	School Name	School Type	Starting Median Salary	Mid-Career Median Salary
0	Massachusetts Institute of Technology (MIT)	Engineering	$72,200.00	$126,000.00
1	California Institute of Technology (CIT)	Engineering	$75,500.00	$123,000.00
2	Harvey Mudd College	Engineering	$71,800.00	$122,000.00
3	Polytechnic University of New York, Brooklyn	Engineering	$62,400.00	$114,000.00
4	Cooper Union	Engineering	$62,200.00	$114,000.00

add STEM column

In [73]:

```
1 eng.copy()
```

Out[73]:

	School Name	School Type	Starting Median Salary	Mid-Career Median Salary
0	Massachusetts Institute of Technology (MIT)	Engineering	$72,200.00	$126,000.00
1	California Institute of Technology (CIT)	Engineering	$75,500.00	$123,000.00
2	Harvey Mudd College	Engineering	$71,800.00	$122,000.00
3	Polytechnic University of New York, Brooklyn	Engineering	$62,400.00	$114,000.00
4	Cooper Union	Engineering	$62,200.00	$114,000.00
5	Worcester Polytechnic Institute (WPI)	Engineering	$61,000.00	$114,000.00
6	Carnegie Mellon University (CMU)	Engineering	$61,800.00	$111,000.00
7	Rensselaer Polytechnic Institute (RPI)	Engineering	$61,100.00	$110,000.00
8	Georgia Institute of Technology	Engineering	$58,300.00	$106,000.00
9	Colorado School of Mines	Engineering	$58,100.00	$106,000.00
10	Stevens Institute of Technology	Engineering	$60,600.00	$105,000.00
11	Illinois Institute of Technology (IIT)	Engineering	$56,000.00	$97,800.00
12	Wentworth Institute of Technology	Engineering	$53,000.00	$96,700.00
13	Virginia Polytechnic Institute and State Unive...	Engineering	$53,500.00	$95,400.00
14	South Dakota School of Mines & Technology	Engineering	$55,800.00	$93,400.00
15	New Mexico Institute of Mining and Technology ...	Engineering	$51,000.00	$93,400.00
16	Rochester Institute of Technology (RIT)	Engineering	$48,900.00	$84,600.00
17	Embry-Riddle Aeronautical University (ERAU)	Engineering	$52,700.00	$80,700.00
18	Tennessee Technological University	Engineering	$46,200.00	$80,000.00

In [74]:

```
1 eng4 = eng.copy()
```

In [75]:

```
1 eng4['STEM'] = True
```

In [76]:

```
1 eng4.head()
```

Out[76]:

	School Name	School Type	Starting Median Salary	Mid-Career Median Salary	STEM
0	Massachusetts Institute of Technology (MIT)	Engineering	$72,200.00	$126,000.00	True
1	California Institute of Technology (CIT)	Engineering	$75,500.00	$123,000.00	True
2	Harvey Mudd College	Engineering	$71,800.00	$122,000.00	True
3	Polytechnic University of New York, Brooklyn	Engineering	$62,400.00	$114,000.00	True
4	Cooper Union	Engineering	$62,200.00	$114,000.00	True

In [77]:

```
1 pd.concat([ivies, eng4])
```

Out[77]:

	School Name	School Type	Starting Median Salary	Mid-Career Median Salary	STEM
0	Dartmouth College	Ivy League	$58,000.00	$134,000.00	NaN
1	Princeton University	Ivy League	$66,500.00	$131,000.00	NaN
2	Yale University	Ivy League	$59,100.00	$126,000.00	NaN
3	Harvard University	Ivy League	$63,400.00	$124,000.00	NaN
4	University of Pennsylvania	Ivy League	$60,900.00	$120,000.00	NaN
5	Cornell University	Ivy League	$60,300.00	$110,000.00	NaN
6	Brown University	Ivy League	$56,200.00	$109,000.00	NaN
7	Columbia University	Ivy League	$59,400.00	$107,000.00	NaN
0	Massachusetts Institute of Technology (MIT)	Engineering	$72,200.00	$126,000.00	True

In [78]:

```
pd.concat([ivies, eng4], join='inner')
```

Out[78]:

	School Name	School Type	Starting Median Salary	Mid-Career Median Salary
0	Dartmouth College	Ivy League	$58,000.00	$134,000.00
1	Princeton University	Ivy League	$66,500.00	$131,000.00
2	Yale University	Ivy League	$59,100.00	$126,000.00
3	Harvard University	Ivy League	$63,400.00	$124,000.00
4	University of Pennsylvania	Ivy League	$60,900.00	$120,000.00
5	Cornell University	Ivy League	$60,300.00	$110,000.00
6	Brown University	Ivy League	$56,200.00	$109,000.00
7	Columbia University	Ivy League	$59,400.00	$107,000.00
0	Massachusetts Institute of Technology (MIT)	Engineering	$72,200.00	$126,000.00

Skill Challenge

1.Concatenate the liberal and state schools into a new dataframe. How many unique school names are there?.

In [79]:

```
new_df = pd.concat([liberal, state]).drop_duplicates(subset=['School Name'])
```

In [80]:

```
liberal.shape, state.shape
```

Out[80]:

```
((47, 4), (175, 4))
```

In [81]:

```
new_df['School Name'].size
```

Out[81]:

```
222
```

In [82]:

```
new_df['School Name'].nunique()
```

Out[82]:

```
222
```

2. What is the average mediian starting salary created in the dataframe above?.

In [83]:

```
1 new_df.columns
```

Out[83]:

```
Index(['School Name', 'School Type', 'Starting Median Salary',
       'Mid-Career Median Salary'],
      dtype='object')
```

In [84]:

```
1 new_df.head()
```

Out[84]:

	School Name	School Type	Starting Median Salary	Mid-Career Median Salary
0	Bucknell University	Liberal Arts	$54,100.00	$110,000.00
1	Colgate University	Liberal Arts	$52,800.00	$108,000.00
2	Amherst College	Liberal Arts	$54,500.00	$107,000.00
3	Lafayette College	Liberal Arts	$53,900.00	$107,000.00
4	Bowdoin College	Liberal Arts	$48,100.00	$107,000.00

In [85]:

```
1 new_df['Starting Median Salary'].replace('\$|,', '', regex=True).astype(float).mean()
```

Out[85]:

```
44469.36936936937
```

In [86]:

```
1 new_df = new_df.astype(str).replace('\$|,', '', regex=True)
```

In [87]:

```
new_df
```

Out[87]:

	School Name	School Type	Starting Median Salary	Mid-Career Median Salary
0	Bucknell University	Liberal Arts	54100.00	110000.00
1	Colgate University	Liberal Arts	52800.00	108000.00
2	Amherst College	Liberal Arts	54500.00	107000.00
3	Lafayette College	Liberal Arts	53900.00	107000.00
4	Bowdoin College	Liberal Arts	48100.00	107000.00
...	...	...	...	...
170	Austin Peay State University	State	37700.00	59200.00
171	Pittsburg State University	State	40400.00	58200.00
172	Southern Utah University	State	41900.00	56500.00
173	Montana State University - Billings	State	37900.00	50600.00
174	Black Hills State University	State	35300.00	43900.00

222 rows × 4 columns

In [88]:

```
new_df.iloc[:, 2].astype(float).mean()
```

Out[88]:

```
44469.36936936937
```

3. Create a short dataframe that shows the top 3 liberal arts and state schools that produces the highest(mid-career) earning graduates. show the School Name and Mid-Career Median Salary Columns from each dataset, side by side, i.e horizontally.
Bonus: nest the columns labels with 'Liberal Arts' and 'State' labels

In [89]:

```
lib2 = liberal.sort_values(by=['Mid-Career Median Salary'], ascending=False)\
       .iloc[:3, [0, 3]].reset_index(drop=True)
```

In [90]:

```
state2 = state.sort_values(by=['Mid-Career Median Salary'], ascending=False)\
       .iloc[:3, [0, 3]].reset_index(drop=True)
```

In [91]:

```
pd.concat([lib2, state2], axis=1, keys=['Liberal Arts', 'States'])
```

Out[91]:

	Liberal Arts		States	
	School Name	Mid-Career Median Salary	School Name	Mid-Career Median Salary
0	Wesleyan University (Middletown, Connecticut)	$97,900.00	University of California, Davis	$99,600.00
1	Bates College	$96,500.00	University of Colorado - Boulder (UCB)	$97,600.00
2	Union College	$95,800.00	University of California, Irvine (UCI)	$96,700.00

The merge() Method

In [92]:

```
regions = pd.read_csv("regions.csv")
```

In [93]:

```
regions.head()
```

Out[93]:

	School Name	Region
0	Massachusetts Institute of Technology (MIT)	Northeastern
1	California Institute of Technology (CIT)	California
2	Harvey Mudd College	California
3	Polytechnic University of New York, Brooklyn	Northeastern
4	Cooper Union	Northeastern

In [94]:

```
regions.shape
```

Out[94]:

```
(269, 2)
```

The merge method is more sensitive to what is in the dataset unlike the concat() method

In [95]:

```
1 schools.head()
```

Out[95]:

	School Name	School Type	Starting Median Salary	Mid-Career Median Salary
0	University of California, Berkeley	State	$59,900.00	$112,000.00
1	University of Virginia (UVA)	State	$52,700.00	$103,000.00
2	Cal Poly San Luis Obispo	State	$57,200.00	$101,000.00
3	University of California at Los Angeles (UCLA)	State	$52,600.00	$101,000.00
4	University of California, San Diego (UCSD)	State	$51,100.00	$101,000.00

In [96]:

```
1 regions.head()
```

Out[96]:

	School Name	Region
0	Massachusetts Institute of Technology (MIT)	Northeastern
1	California Institute of Technology (CIT)	California
2	Harvey Mudd College	California
3	Polytechnic University of New York, Brooklyn	Northeastern
4	Cooper Union	Northeastern

In [97]:

```
1 schools.shape
```

Out[97]:

```
(249, 4)
```

In [98]:

```
1 pd.merge(schools, regions)
```

Out[98]:

	School Name	School Type	Starting Median Salary	Mid-Career Median Salary	Region
0	University of California, Berkeley	State	$59,900.00	$112,000.00	California
1	University of Virginia (UVA)	State	$52,700.00	$103,000.00	Southern
2	Cal Poly San Luis Obispo	State	$57,200.00	$101,000.00	California
3	University of California at Los Angeles (UCLA)	State	$52,600.00	$101,000.00	California
4	University of California, San Diego (UCSD)	State	$51,100.00	$101,000.00	California
...	...	...	...	...	...
264	Harvard University	Ivy League	$63,400.00	$124,000.00	Northeastern

In [99]:

```
1 np.nan in pd.merge(schools,regions)['School Name']
```

Out[99]:

```
False
```

In [100]:

```
1 np.nan in schools
```

Out[100]:

```
False
```

In [101]:

```
1 pd.merge(regions, schools)
```

Out[101]:

	School Name	Region	School Type	Starting Median Salary	Mid-Career Median Salary
0	Massachusetts Institute of Technology (MIT)	Northeastern	Engineering	$72,200.00	$126,000.00
1	California Institute of Technology (CIT)	California	Engineering	$75,500.00	$123,000.00
2	Harvey Mudd College	California	Engineering	$71,800.00	$122,000.00
3	Polytechnic University of New York, Brooklyn	Northeastern	Engineering	$62,400.00	$114,000.00
4	Cooper Union	Northeastern	Engineering	$62,200.00	$114,000.00
...	...	...	...	...	...
264	Austin Peay State University	Southern	State	$37,700.00	$59,200.00

pd.merge(schools, regions) is the same as pd.merge(schools, regions, on='School Name')
-> This is so because the merge() method looks for what is common in each dataset being matched and in this case is the "School Name" column

The left-on and right on parameters

In [102]:

```
1 mid_career = pd.read_csv("mid_career.csv")
```

In [103]:

```
1 mid_career.head()
```

Out[103]:

	school_name	Mid-Career 10th Percentile Salary	Mid-Career 25th Percentile Salary	Mid-Career 75th Percentile Salary	Mid-Career 90th Percentile Salary
0	Massachusetts Institute of Technology (MIT)	$76,800.00	$99,200.00	$168,000.00	$220,000.00
1	California Institute of Technology (CIT)	NaN	$104,000.00	$161,000.00	NaN
2	Harvey Mudd College	NaN	$96,000.00	$180,000.00	NaN
3	Polytechnic University of New York, Brooklyn	$66,800.00	$94,300.00	$143,000.00	$190,000.00
4	Cooper Union	NaN	$80,200.00	$142,000.00	NaN

In [104]:

```
1 schools.head()
```

Out[104]:

	School Name	School Type	Starting Median Salary	Mid-Career Median Salary
0	University of California, Berkeley	State	$59,900.00	$112,000.00
1	University of Virginia (UVA)	State	$52,700.00	$103,000.00
2	Cal Poly San Luis Obispo	State	$57,200.00	$101,000.00
3	University of California at Los Angeles (UCLA)	State	$52,600.00	$101,000.00
4	University of California, San Diego (UCSD)	State	$51,100.00	$101,000.00

pd.merge(schools, mid_career) -> this would give an error because there is no common column name in either of the datasets

In [105]:

```
1 pd.merge(schools, mid_career, left_on='School Name', right_on='school_name')
```

Out[105]:

	School Name	School Type	Starting Median Salary	Mid-Career Median Salary	school_name	Mid-Career 10th Percentile Salary	Mid-Career 25th Percentile Salary	Mid-Career 75th Percentile Salary	Mid-Ca 90th Percen Salary
0	University of California, Berkeley	State	$59,900.00	$112,000.00	University of California, Berkeley	$59,500.00	$81,000.00	$149,000.00	$201,0
1	University of Virginia (UVA)	State	$52,700.00	$103,000.00	University of Virginia (UVA)	$52,200.00	$71,800.00	$146,000.00	$215,0
2	Cal Poly San Luis Obispo	State	$57,200.00	$101,000.00	Cal Poly San Luis Obispo	$55,000.00	$74,700.00	$133,000.00	$178,0
3	University of California at Los Angeles	State	$52,600.00	$101,000.00	University of California at Los Angeles	$51,300.00	$72,500.00	$139,000.00	$193,0

The reason why the above line of code worked is because of the left_on and right_on parameters

In [106]:

```
1 pd.merge(mid_career, schools, left_on='school_name', right_on='School Name')
```

Out[106]:

	school_name	Mid-Career 10th Percentile Salary	Mid-Career 25th Percentile Salary	Mid-Career 75th Percentile Salary	Mid-Career 90th Percentile Salary	School Name	School Type	Starting Median Salary
0	Massachusetts Institute of Technology (MIT)	$76,800.00	$99,200.00	$168,000.00	$220,000.00	Massachusetts Institute of Technology (MIT)	Engineering	$72,200.00
1	California Institute of Technology (CIT)	NaN	$104,000.00	$161,000.00	NaN	California Institute of Technology (CIT)	Engineering	$75,500.00
2	Harvey Mudd College	NaN	$96,000.00	$180,000.00	NaN	Harvey Mudd College	Engineering	$71,800.00
	Polytechnic					Polytechnic		

In [107]:

```
1 pd.merge(schools, mid_career, left_on='School Name', right_on='school_name').drop('scho
```

Out[107]:

	School Name	School Type	Starting Median Salary	Mid-Career Median Salary	Mid-Career 10th Percentile Salary	Mid-Career 25th Percentile Salary	Mid-Career 75th Percentile Salary	Mid-Career 90th Percentile Salary
0	University of California, Berkeley	State	$59,900.00	$112,000.00	$59,500.00	$81,000.00	$149,000.00	$201,000.00
1	University of Virginia (UVA)	State	$52,700.00	$103,000.00	$52,200.00	$71,800.00	$146,000.00	$215,000.00
2	Cal Poly San Luis Obispo	State	$57,200.00	$101,000.00	$55,000.00	$74,700.00	$133,000.00	$178,000.00
3	University of California at Los Angeles (UCLA)	State	$52,600.00	$101,000.00	$51,300.00	$72,500.00	$139,000.00	$193,000.00

Inner Vs Outer joins

the how parameter

In [108]:

```
1 ivies.merge(regions)# same as pd.merge(ivies, regions)
```

Out[108]:

	School Name	School Type	Starting Median Salary	Mid-Career Median Salary	Region
0	Dartmouth College	Ivy League	$58,000.00	$134,000.00	Northeastern
1	Princeton University	Ivy League	$66,500.00	$131,000.00	Northeastern
2	Yale University	Ivy League	$59,100.00	$126,000.00	Northeastern
3	Harvard University	Ivy League	$63,400.00	$124,000.00	Northeastern
4	University of Pennsylvania	Ivy League	$60,900.00	$120,000.00	Northeastern
5	Cornell University	Ivy League	$60,300.00	$110,000.00	Northeastern
6	Brown University	Ivy League	$56,200.00	$109,000.00	Northeastern
7	Columbia University	Ivy League	$59,400.00	$107,000.00	Northeastern

In [109]:

```
1 pd.merge(ivies, regions, how='inner')# the default value for the how parameter is 'inne
```

Out[109]:

	School Name	School Type	Starting Median Salary	Mid-Career Median Salary	Region
0	Dartmouth College	Ivy League	$58,000.00	$134,000.00	Northeastern
1	Princeton University	Ivy League	$66,500.00	$131,000.00	Northeastern
2	Yale University	Ivy League	$59,100.00	$126,000.00	Northeastern
3	Harvard University	Ivy League	$63,400.00	$124,000.00	Northeastern
4	University of Pennsylvania	Ivy League	$60,900.00	$120,000.00	Northeastern
5	Cornell University	Ivy League	$60,300.00	$110,000.00	Northeastern
6	Brown University	Ivy League	$56,200.00	$109,000.00	Northeastern
7	Columbia University	Ivy League	$59,400.00	$107,000.00	Northeastern

In [110]:

```
pd.merge(ivies, regions, how='outer')#same as union of set
```

Out[110]:

	School Name	School Type	Starting Median Salary	Mid-Career Median Salary	Region
0	Dartmouth College	Ivy League	$58,000.00	$134,000.00	Northeastern
1	Princeton University	Ivy League	$66,500.00	$131,000.00	Northeastern
2	Yale University	Ivy League	$59,100.00	$126,000.00	Northeastern
3	Harvard University	Ivy League	$63,400.00	$124,000.00	Northeastern
4	University of Pennsylvania	Ivy League	$60,900.00	$120,000.00	Northeastern
...	...	...	...	...	...
264	Austin Peay State University	NaN	NaN	NaN	Southern
265	Pittsburg State University	NaN	NaN	NaN	Midwestern
266	Southern Utah University	NaN	NaN	NaN	Western
267	Montana State University - Billings	NaN	NaN	NaN	Western
268	Black Hills State University	NaN	NaN	NaN	Midwestern

269 rows × 5 columns

In [111]:

```
pd.merge(ivies, regions, how='outer')['Starting Median Salary'].hasnans
```

Out[111]:

```
True
```

In [112]:

```
np.count_nonzero(pd.merge(ivies, regions, how='outer')['Starting Median Salary'].isna()
```

Out[112]:

```
261
```

In [113]:

```
a = {1,2,3}
b = {4,5,6}
```

In [114]:

```
c = {1,3,4,6}
```

In [115]:

```
a.union(b)
```

Out[115]:

```
{1, 2, 3, 4, 5, 6}
```

In [116]:

```
a.intersection(b)
```

Out[116]:

```
set()
```

In [117]:

```
a.intersection(c)
```

Out[117]:

```
{1, 3}
```

In [118]:

```
b.intersection(c)
```

Out[118]:

```
{4, 6}
```

Left Vs Right joins

In [119]:

```
pd.merge(ivies, regions, how='left')
```

Out[119]:

	School Name	School Type	Starting Median Salary	Mid-Career Median Salary	Region
0	Dartmouth College	Ivy League	$58,000.00	$134,000.00	Northeastern
1	Princeton University	Ivy League	$66,500.00	$131,000.00	Northeastern
2	Yale University	Ivy League	$59,100.00	$126,000.00	Northeastern
3	Harvard University	Ivy League	$63,400.00	$124,000.00	Northeastern
4	University of Pennsylvania	Ivy League	$60,900.00	$120,000.00	Northeastern
5	Cornell University	Ivy League	$60,300.00	$110,000.00	Northeastern
6	Brown University	Ivy League	$56,200.00	$109,000.00	Northeastern
7	Columbia University	Ivy League	$59,400.00	$107,000.00	Northeastern

In [120]:

```
1 pd.merge(ivies, regions, how='right')
```

Out[120]:

	School Name	School Type	Starting Median Salary	Mid-Career Median Salary	Region
0	Dartmouth College	Ivy League	$58,000.00	$134,000.00	Northeastern
1	Princeton University	Ivy League	$66,500.00	$131,000.00	Northeastern
2	Yale University	Ivy League	$59,100.00	$126,000.00	Northeastern
3	Harvard University	Ivy League	$63,400.00	$124,000.00	Northeastern
4	University of Pennsylvania	Ivy League	$60,900.00	$120,000.00	Northeastern
...	...	...	...	...	...
264	Austin Peay State University	NaN	NaN	NaN	Southern
265	Pittsburg State University	NaN	NaN	NaN	Midwestern
266	Southern Utah University	NaN	NaN	NaN	Western
267	Montana State University - Billings	NaN	NaN	NaN	Western
268	Black Hills State University	NaN	NaN	NaN	Midwestern

269 rows × 5 columns

In [121]:

```
1 pd.merge(ivies, regions, how='right').equals(pd.merge(ivies, regions, how='outer'))
```

Out[121]:

```
True
```

From the above line of code, it is safe to say that "how='outer'" is the same as "how='right'"

One-to-One and One-to-Many joins

1-1 join: happens when each record in a dataframe is associated with one record in another dataframe.

In [122]:

```
1 ivies.head()
```

Out[122]:

	School Name	School Type	Starting Median Salary	Mid-Career Median Salary
0	Dartmouth College	Ivy League	$58,000.00	$134,000.00
1	Princeton University	Ivy League	$66,500.00	$131,000.00
2	Yale University	Ivy League	$59,100.00	$126,000.00
3	Harvard University	Ivy League	$63,400.00	$124,000.00
4	University of Pennsylvania	Ivy League	$60,900.00	$120,000.00

In [123]:

```
1 regions.head()
```

Out[123]:

	School Name	Region
0	Massachusetts Institute of Technology (MIT)	Northeastern
1	California Institute of Technology (CIT)	California
2	Harvey Mudd College	California
3	Polytechnic University of New York, Brooklyn	Northeastern
4	Cooper Union	Northeastern

In [124]:

```
1 pd.merge(schools, mid_career, left_on='School Name', right_on='school_name', suffixes=(
```

Out[124]:

	School Name	School Type	Starting Median Salary	Mid-Career Median Salary	school_name	Mid-Career 10th Percentile Salary	Mid-Career 25th Percentile Salary	Mid-C 75th Percer Salary
0	University of California, Berkeley	State	$59,900.00	$112,000.00	University of California, Berkeley	$59,500.00	$81,000.00	$149,(
1	University of Virginia (UVA)	State	$52,700.00	$103,000.00	University of Virginia (UVA)	$52,200.00	$71,800.00	$146,(
2	Cal Poly San Luis Obispo	State	$57,200.00	$101,000.00	Cal Poly San Luis Obispo	$55,000.00	$74,700.00	$133,(
3	University of California at Los Angeles (UCLA)	State	$52,600.00	$101,000.00	University of California at Los Angeles (UCLA)	$51,300.00	$72,500.00	$139,(
4	University of California, San Diego (UCSD)	State	$51,100.00	$101,000.00	University of California, San Diego (UCSD)	$51,700.00	$75,400.00	$131,(
...	...	...	...	...	...	...	...	
264	Harvard University	Ivy League	$63,400.00	$124,000.00	Harvard University	$54,800.00	$86,200.00	$179,(
265	University of Pennsylvania	Ivy League	$60,900.00	$120,000.00	University of Pennsylvania	$55,900.00	$79,200.00	$192,(
266	Cornell University	Ivy League	$60,300.00	$110,000.00	Cornell University	$56,800.00	$79,800.00	$160,(
267	Brown University	Ivy League	$56,200.00	$109,000.00	Brown University	$55,400.00	$74,400.00	$159,(
268	Columbia University	Ivy League	$59,400.00	$107,000.00	Columbia University	$50,300.00	$71,900.00	$161,(

269 rows × 9 columns

In [125]:

```
1 schools.merge(mid_career, left_on='School Name', right_on='school_name', suffixes=('_x'
```

Out[125]:

	School Name	School Type	Starting Median Salary	Mid-Career Median Salary	school_name	Mid-Career 10th Percentile Salary	Mid-Career 25th Percentile Salary	Mid-C 75th Percer Salary
0	University of California, Berkeley	State	$59,900.00	$112,000.00	University of California, Berkeley	$59,500.00	$81,000.00	$149,(
1	University of Virginia (UVA)	State	$52,700.00	$103,000.00	University of Virginia (UVA)	$52,200.00	$71,800.00	$146,(
2	Cal Poly San Luis Obispo	State	$57,200.00	$101,000.00	Cal Poly San Luis Obispo	$55,000.00	$74,700.00	$133,(
3	University of California at Los Angeles (UCLA)	State	$52,600.00	$101,000.00	University of California at Los Angeles (UCLA)	$51,300.00	$72,500.00	$139,(
4	University of California, San Diego (UCSD)	State	$51,100.00	$101,000.00	University of California, San Diego (UCSD)	$51,700.00	$75,400.00	$131,(
...	...	...	...	...	...	...	...	
264	Harvard University	Ivy League	$63,400.00	$124,000.00	Harvard University	$54,800.00	$86,200.00	$179,(
265	University of Pennsylvania	Ivy League	$60,900.00	$120,000.00	University of Pennsylvania	$55,900.00	$79,200.00	$192,(
266	Cornell University	Ivy League	$60,300.00	$110,000.00	Cornell University	$56,800.00	$79,800.00	$160,(
267	Brown University	Ivy League	$56,200.00	$109,000.00	Brown University	$55,400.00	$74,400.00	$159,(
268	Columbia University	Ivy League	$59,400.00	$107,000.00	Columbia University	$50,300.00	$71,900.00	$161,(

269 rows × 9 columns

1-Many joins

In [126]:

```
1 state
```

Out[126]:

	School Name	School Type	Starting Median Salary	Mid-Career Median Salary
0	University of California, Berkeley	State	$59,900.00	$112,000.00
1	University of Virginia (UVA)	State	$52,700.00	$103,000.00
2	Cal Poly San Luis Obispo	State	$57,200.00	$101,000.00
3	University of California at Los Angeles (UCLA)	State	$52,600.00	$101,000.00
4	University of California, San Diego (UCSD)	State	$51,100.00	$101,000.00
...	...	...	...	...
170	Austin Peay State University	State	$37,700.00	$59,200.00
171	Pittsburg State University	State	$40,400.00	$58,200.00
172	Southern Utah University	State	$41,900.00	$56,500.00
173	Montana State University - Billings	State	$37,900.00	$50,600.00
174	Black Hills State University	State	$35,300.00	$43,900.00

175 rows × 4 columns

In [127]:

```
1 state['School Name'].is_unique
```

Out[127]:

```
True
```

In [128]:

```
regions[regions['School Name'].isin(state['School Name'])]
```

Out[128]:

	School Name	Region
19	University of Illinois at Urbana-Champaign (UIUC)	Midwestern
20	University of Maryland, College Park	Southern
21	University of California, Santa Barbara (UCSB)	California
22	University of Texas (UT) - Austin	Southern
23	State University of New York (SUNY) at Albany	Northeastern
...	...	...
264	Austin Peay State University	Southern
265	Pittsburg State University	Midwestern
266	Southern Utah University	Western
267	Montana State University - Billings	Western
268	Black Hills State University	Midwestern

194 rows × 2 columns

In [129]:

```
regions[regions['School Name'].isin(state['School Name'])].loc[:, 'School Name'].value_
```

Out[129]:

```
University of Mississippi                          2
University of Illinois at Urbana-Champaign (UIUC)  2
West Virginia University (WVU)                     2
University of New Hampshire (UNH)                  2
University of Iowa (UI)                            2
                                                  ..
Stony Brook University                             1
University of Virginia (UVA)                       1
University of Arizona                              1
Cal Poly San Luis Obispo                           1
State University of New York (SUNY) at Farmingdale 1
Name: School Name, Length: 175, dtype: int64
```

NOTE: The two cells above shows that some school names occured twice

Merging states and regions one to many.

In [130]:

```
1 pd.merge(state, regions, how='inner', on='School Name').sort_values(by='School Name')
```

Out[130]:

	School Name	School Type	Starting Median Salary	Mid-Career Median Salary	Region
169	Appalachian State University	State	$40,400.00	$69,100.00	Southern
58	Arizona State University (ASU)	State	$47,400.00	$84,100.00	Western
57	Arizona State University (ASU)	State	$47,400.00	$84,100.00	Western
184	Arkansas State University (ASU)	State	$38,700.00	$63,300.00	Southern
48	Auburn University	State	$45,400.00	$84,700.00	Southern
...	...	...	...	...	...
107	West Virginia University (WVU)	State	$43,100.00	$78,100.00	Southern
108	West Virginia University (WVU)	State	$43,100.00	$78,100.00	Southern
175	Western Carolina University	State	$36,900.00	$66,600.00	Southern
131	Western Michigan University (WMU)	State	$42,300.00	$73,800.00	Midwestern
122	Western Washington University	State	$42,700.00	$75,400.00	Western

194 rows × 5 columns

In [131]:

```
pd.merge(state, regions, how='inner', on='School Name').sort_values(by='School Name').c
```

Out[131]:

	School Name	School Type	Starting Median Salary	Mid-Career Median Salary	Region
169	Appalachian State University	State	$40,400.00	$69,100.00	Southern
58	Arizona State University (ASU)	State	$47,400.00	$84,100.00	Western
184	Arkansas State University (ASU)	State	$38,700.00	$63,300.00	Southern
48	Auburn University	State	$45,400.00	$84,700.00	Southern
189	Austin Peay State University	State	$37,700.00	$59,200.00	Southern
...	...	...	...	...	...
117	Wayne State University	State	$42,800.00	$76,100.00	Midwestern
	West Virginia University				

In [132]:

```
#OR
pd.merge(state, regions.drop_duplicates())# this line of code gives the same result as
```

Out[132]:

	School Name	School Type	Starting Median Salary	Mid-Career Median Salary	Region
0	University of California, Berkeley	State	$59,900.00	$112,000.00	California
1	University of Virginia (UVA)	State	$52,700.00	$103,000.00	Southern
2	Cal Poly San Luis Obispo	State	$57,200.00	$101,000.00	California
3	University of California at Los Angeles (UCLA)	State	$52,600.00	$101,000.00	California
4	University of California, San Diego (UCSD)	State	$51,100.00	$101,000.00	California
...	...	...	...	...	...
170	Austin Peay State University	State	$37,700.00	$59,200.00	Southern
171	Pittsburg State University	State	$40,400.00	$58,200.00	Midwestern
172	Southern Utah University	State	$41,900.00	$56,500.00	Western
173	Montana State University - Billings	State	$37,900.00	$50,600.00	Western
174	Black Hills State University	State	$35,300.00	$43,900.00	Midwestern

175 rows × 5 columns

In [133]:

```
regions['School Name'].is_unique
```

Out[133]:

```
False
```

In [134]:

```
mid_career['school_name'].is_unique
```

Out[134]:

```
False
```

In [135]:

```
pd.merge(state, mid_career.drop_duplicates(), left_on='School Name', right_on='school_r
```

Out[135]:

	School Name	School Type	Starting Median Salary	Mid-Career Median Salary	school_name	Mid-Career 10th Percentile Salary	Mid-Career 25th Percentile Salary	Mid-Career 75th Percentile Salary	Mid-Caree 90th Percentile Salary
0	University of California, Berkeley	State	$59,900.00	$112,000.00	University of California, Berkeley	$59,500.00	$81,000.00	$149,000.00	$201,000.0
1	University of Virginia (UVA)	State	$52,700.00	$103,000.00	University of Virginia (UVA)	$52,200.00	$71,800.00	$146,000.00	$215,000.0
2	Cal Poly San Luis Obispo	State	$57,200.00	$101,000.00	Cal Poly San Luis Obispo	$55,000.00	$74,700.00	$133,000.00	$178,000.0
	University of				Universitv of				

Many-to-Many joins

These are joins that occur when we have duplicates in the key columns of both dataframes that are being merged

Survey data-> What is the prestige of an ivy or eng degree

In [136]:

```
survey = pd.DataFrame({'School Type':['Ivy League', 'Ivy League', 'Engineering', 'Engir
                       'Prestige':['High', 'Good', 'Good', 'Okay'],
                       'Respondent':[1, 2, 3, 4]
                      })
```

In [137]:

```
1 survey
```

Out[137]:

	School Type	Prestige	Respondent
0	Ivy League	High	1
1	Ivy League	Good	2
2	Engineering	Good	3
3	Engineering	Okay	4

In [138]:

```
1 pd.merge(ivies, survey)
```

Out[138]:

	School Name	School Type	Starting Median Salary	Mid-Career Median Salary	Prestige	Respondent
0	Dartmouth College	Ivy League	$58,000.00	$134,000.00	High	1
1	Dartmouth College	Ivy League	$58,000.00	$134,000.00	Good	2
2	Princeton University	Ivy League	$66,500.00	$131,000.00	High	1
3	Princeton University	Ivy League	$66,500.00	$131,000.00	Good	2
4	Yale University	Ivy League	$59,100.00	$126,000.00	High	1
5	Yale University	Ivy League	$59,100.00	$126,000.00	Good	2

In [139]:

```
1 survey = survey.append(pd.Series(data={'School Type':'Ivy League', 'Prestige':'Very Hig
```

In [140]:

```
1 survey
```

Out[140]:

	School Type	Prestige	Respondent
0	Ivy League	High	1
1	Ivy League	Good	2
2	Engineering	Good	3
3	Engineering	Okay	4
4	Ivy League	Very High	5

In [141]:

```
1 pd.merge(ivies, survey)
```

Out[141]:

	School Name	School Type	Starting Median Salary	Mid-Career Median Salary	Prestige	Respondent
0	Dartmouth College	Ivy League	$58,000.00	$134,000.00	High	1
1	Dartmouth College	Ivy League	$58,000.00	$134,000.00	Good	2
2	Dartmouth College	Ivy League	$58,000.00	$134,000.00	Very High	5
3	Princeton University	Ivy League	$66,500.00	$131,000.00	High	1
4	Princeton University	Ivy League	$66,500.00	$131,000.00	Good	2
5	Princeton University	Ivy League	$66,500.00	$131,000.00	Very High	5

Merging by Index

In [142]:

```
1 ivies4 = ivies.set_index('School Name')
```

In [143]:

```
1 regions2 = regions.set_index('School Name')
```

In [144]:

```
1 ivies4.head()
```

Out[144]:

School Name	School Type	Starting Median Salary	Mid-Career Median Salary
Dartmouth College	Ivy League	$58,000.00	$134,000.00
Princeton University	Ivy League	$66,500.00	$131,000.00
Yale University	Ivy League	$59,100.00	$126,000.00
Harvard University	Ivy League	$63,400.00	$124,000.00
University of Pennsylvania	Ivy League	$60,900.00	$120,000.00

In [145]:

```
1 regions2.head()
```

Out[145]:

School Name	Region
Massachusetts Institute of Technology (MIT)	Northeastern
California Institute of Technology (CIT)	California
Harvey Mudd College	California
Polytechnic University of New York, Brooklyn	Northeastern
Cooper Union	Northeastern

In [146]:

```
1 #pd.merge(ivies4, regions2) would give an error
```

In [147]:

```
pd.merge(ivies4, regions2, left_index=True, right_index=True)
```

Out[147]:

School Name	School Type	Starting Median Salary	Mid-Career Median Salary	Region
Brown University	Ivy League	$56,200.00	$109,000.00	Northeastern
Columbia University	Ivy League	$59,400.00	$107,000.00	Northeastern
Cornell University	Ivy League	$60,300.00	$110,000.00	Northeastern
Dartmouth College	Ivy League	$58,000.00	$134,000.00	Northeastern
Harvard University	Ivy League	$63,400.00	$124,000.00	Northeastern
Princeton University	Ivy League	$66,500.00	$131,000.00	Northeastern
University of Pennsylvania	Ivy League	$60,900.00	$120,000.00	Northeastern
Yale University	Ivy League	$59,100.00	$126,000.00	Northeastern

Merging by index vs column

In [148]:

```
regions2.head()
```

Out[148]:

School Name	Region
Massachusetts Institute of Technology (MIT)	Northeastern
California Institute of Technology (CIT)	California
Harvey Mudd College	California
Polytechnic University of New York, Brooklyn	Northeastern
Cooper Union	Northeastern

In [149]:

```
1 ivies4.head()
```

Out[149]:

School Name	School Type	Starting Median Salary	Mid-Career Median Salary
Dartmouth College	Ivy League	$58,000.00	$134,000.00
Princeton University	Ivy League	$66,500.00	$131,000.00
Yale University	Ivy League	$59,100.00	$126,000.00
Harvard University	Ivy League	$63,400.00	$124,000.00
University of Pennsylvania	Ivy League	$60,900.00	$120,000.00

In [150]:

```
1 pd.merge(ivies4, regions, left_index=True, right_on='School Name')
```

Out[150]:

	School Type	Starting Median Salary	Mid-Career Median Salary	School Name	Region
86	Ivy League	$58,000.00	$134,000.00	Dartmouth College	Northeastern
87	Ivy League	$66,500.00	$131,000.00	Princeton University	Northeastern
88	Ivy League	$59,100.00	$126,000.00	Yale University	Northeastern
89	Ivy League	$63,400.00	$124,000.00	Harvard University	Northeastern
90	Ivy League	$60,900.00	$120,000.00	University of Pennsylvania	Northeastern
91	Ivy League	$60,300.00	$110,000.00	Cornell University	Northeastern
92	Ivy League	$56,200.00	$109,000.00	Brown University	Northeastern
93	Ivy League	$59,400.00	$107,000.00	Columbia University	Northeastern

The join Method

In [151]:

```
1 ivies4.join(regions2)
```

Out[151]:

School Name	School Type	Starting Median Salary	Mid-Career Median Salary	Region
Brown University	Ivy League	$56,200.00	$109,000.00	Northeastern
Columbia University	Ivy League	$59,400.00	$107,000.00	Northeastern
Cornell University	Ivy League	$60,300.00	$110,000.00	Northeastern
Dartmouth College	Ivy League	$58,000.00	$134,000.00	Northeastern
Harvard University	Ivy League	$63,400.00	$124,000.00	Northeastern
Princeton University	Ivy League	$66,500.00	$131,000.00	Northeastern
University of Pennsylvania	Ivy League	$60,900.00	$120,000.00	Northeastern
Yale University	Ivy League	$59,100.00	$126,000.00	Northeastern

In [152]:

```
1 pd.merge(ivies4, regions2, left_index=True, right_index=True)
```

Out[152]:

School Name	School Type	Starting Median Salary	Mid-Career Median Salary	Region
Brown University	Ivy League	$56,200.00	$109,000.00	Northeastern
Columbia University	Ivy League	$59,400.00	$107,000.00	Northeastern
Cornell University	Ivy League	$60,300.00	$110,000.00	Northeastern
Dartmouth College	Ivy League	$58,000.00	$134,000.00	Northeastern
Harvard University	Ivy League	$63,400.00	$124,000.00	Northeastern
Princeton University	Ivy League	$66,500.00	$131,000.00	Northeastern
University of Pennsylvania	Ivy League	$60,900.00	$120,000.00	Northeastern
Yale University	Ivy League	$59,100.00	$126,000.00	Northeastern

The immediate 2 cells above gives the same result

In [153]:

```
1 ivies
```

Out[153]:

	School Name	School Type	Starting Median Salary	Mid-Career Median Salary
0	Dartmouth College	Ivy League	$58,000.00	$134,000.00
1	Princeton University	Ivy League	$66,500.00	$131,000.00
2	Yale University	Ivy League	$59,100.00	$126,000.00
3	Harvard University	Ivy League	$63,400.00	$124,000.00
4	University of Pennsylvania	Ivy League	$60,900.00	$120,000.00
5	Cornell University	Ivy League	$60,300.00	$110,000.00
6	Brown University	Ivy League	$56,200.00	$109,000.00
7	Columbia University	Ivy League	$59,400.00	$107,000.00

In [154]:

```
1 regions2.head()
```

Out[154]:

	Region
School Name	
Massachusetts Institute of Technology (MIT)	Northeastern
California Institute of Technology (CIT)	California
Harvey Mudd College	California
Polytechnic University of New York, Brooklyn	Northeastern
Cooper Union	Northeastern

In [155]:

```
ivies.join(regions2, on='School Name')
```

Out[155]:

	School Name	School Type	Starting Median Salary	Mid-Career Median Salary	Region
0	Dartmouth College	Ivy League	$58,000.00	$134,000.00	Northeastern
1	Princeton University	Ivy League	$66,500.00	$131,000.00	Northeastern
2	Yale University	Ivy League	$59,100.00	$126,000.00	Northeastern
3	Harvard University	Ivy League	$63,400.00	$124,000.00	Northeastern
4	University of Pennsylvania	Ivy League	$60,900.00	$120,000.00	Northeastern
5	Cornell University	Ivy League	$60,300.00	$110,000.00	Northeastern
6	Brown University	Ivy League	$56,200.00	$109,000.00	Northeastern
7	Columbia University	Ivy League	$59,400.00	$107,000.00	Northeastern

Anything that can be done with the merge() method can also be done with the join method

Skill Challenge

1.

In [156]:

```
liberal.head()
```

Out[156]:

	School Name	School Type	Starting Median Salary	Mid-Career Median Salary
0	Bucknell University	Liberal Arts	$54,100.00	$110,000.00
1	Colgate University	Liberal Arts	$52,800.00	$108,000.00
2	Amherst College	Liberal Arts	$54,500.00	$107,000.00
3	Lafayette College	Liberal Arts	$53,900.00	$107,000.00
4	Bowdoin College	Liberal Arts	$48,100.00	$107,000.00

In [157]:

```
regions.head()
```

Out[157]:

	School Name	Region
0	Massachusetts Institute of Technology (MIT)	Northeastern
1	California Institute of Technology (CIT)	California
2	Harvey Mudd College	California
3	Polytechnic University of New York, Brooklyn	Northeastern
4	Cooper Union	Northeastern

In [158]:

```
dfm = pd.merge(liberal, regions)
```

In [159]:

```
dfm.columns
```

Out[159]:

```
Index(['School Name', 'School Type', 'Starting Median Salary',
       'Mid-Career Median Salary', 'Region'],
      dtype='object')
```

In [160]:

```
dfm.Region.value_counts()
```

Out[160]:

```
Northeastern    25
Midwestern       8
Western          7
Southern         5
California       3
Name: Region, dtype: int64
```

The Northeastern Region has the highest number of liberal arts schools

In [161]:

```
liberal.shape
```

Out[161]:

```
(47, 4)
```

2.

In [162]:

```
mid_career.set_index('school_name', inplace=True)
```

In [163]:

```
1 mid_career.head()
```

Out[163]:

school_name	Mid-Career 10th Percentile Salary	Mid-Career 25th Percentile Salary	Mid-Career 75th Percentile Salary	Mid-Career 90th Percentile Salary
Massachusetts Institute of Technology (MIT)	$76,800.00	$99,200.00	$168,000.00	$220,000.00
California Institute of Technology (CIT)	NaN	$104,000.00	$161,000.00	NaN
Harvey Mudd College	NaN	$96,000.00	$180,000.00	NaN
Polytechnic University of New York, Brooklyn	$66,800.00	$94,300.00	$143,000.00	$190,000.00
Cooper Union	NaN	$80,200.00	$142,000.00	NaN

In [164]:

```
1 mid_career.shape
```

Out[164]:

```
(269, 4)
```

3.

In [165]:

```
1 pd.merge(dfm, mid_career, left_on='School Name', right_on='school_name')
```

Out[165]:

	School Name	School Type	Starting Median Salary	Mid-Career Median Salary	Region	Mid-Career 10th Percentile Salary	Mid-Career 25th Percentile Salary	Mid-Career 75th Percentile Salary	Mid-Caree 90th Percentile Salary
0	Bucknell University	Liberal Arts	$54,100.00	$110,000.00	Northeastern	$62,800.00	$80,600.00	$156,000.00	$251,000.0
1	Colgate University	Liberal Arts	$52,800.00	$108,000.00	Northeastern	$60,000.00	$76,700.00	$167,000.00	$265,000.0
2	Amherst College	Liberal Arts	$54,500.00	$107,000.00	Northeastern	NaN	$84,900.00	$162,000.00	Na
3	Lafayette College	Liberal Arts	$53,900.00	$107,000.00	Northeastern	$70,600.00	$79,300.00	$144,000.00	$204,000.0
4	Bowdoin College	Liberal Arts	$48,100.00	$107,000.00	Northeastern	NaN	$74,600.00	$146,000.00	Na

In [166]:

```
#Is it 1-1
```

In [167]:

```
left_key = mid_career.index
```

In [168]:

```
right_key = dfm['School Name'] # liberal arts inner join regions
```

In [169]:

```
1 left_key[left_key.isin(right_key)].value_counts()
```

Out[169]:

```
Randolph-Macon College                          2
College of the Holy Cross                       1
Wittenberg University                           1
Moravian College                                1
Bates College                                   1
Reed College                                    1
Colgate University                              1
Skidmore College                                1
Bucknell University                             1
Middlebury College                              1
Washington and Lee University                   1
Bowdoin College                                 1
Wellesley College                               1
Amherst College                                 1
Juniata College                                 1
Colby College                                   1
Carleton College                                1
Williams College                                1
Union College                                   1
Evergreen State College                         1
Davidson College                                1
Wesleyan University (Middletown, Connecticut)   1
Grinnell College                                1
Siena College                                   1
Occidental College                              1
Swarthmore College                              1
Whitman College                                 1
Hamilton College                                1
Pomona College                                  1
Mount Holyoke College                           1
Vassar College                                  1
St. Olaf College                                1
DePauw University                               1
Smith College                                   1
Thomas Aquinas College                          1
Ursinus College                                 1
Oberlin College                                 1
Lewis & Clark College                           1
Gustavus Adolphus College                       1
University of Richmond                          1
Lafayette College                               1
Gettysburg College                              1
University of Puget Sound                       1
Denison University                              1
Franklin and Marshall College                   1
Colorado College (CC)                           1
Fort Lewis College                              1
Name: school_name, dtype: int64
```

In [170]:

```
1 right_key[right_key.isin(right_key)].value_counts()
```

Out[170]:

```
Randolph-Macon College                          2
College of the Holy Cross                       1
Wittenberg University                           1
Bates College                                   1
Reed College                                    1
Moravian College                                1
Colgate University                              1
Skidmore College                                1
Bucknell University                             1
Middlebury College                              1
Washington and Lee University                   1
Siena College                                   1
Grinnell College                                1
Amherst College                                 1
Juniata College                                 1
Colby College                                   1
Carleton College                                1
Williams College                                1
Union College                                   1
Evergreen State College                         1
Davidson College                                1
Wesleyan University (Middletown, Connecticut)   1
Bowdoin College                                 1
Wellesley College                               1
Occidental College                              1
Ursinus College                                 1
Whitman College                                 1
Hamilton College                                1
Pomona College                                  1
Mount Holyoke College                           1
Vassar College                                  1
St. Olaf College                                1
Smith College                                   1
Thomas Aquinas College                          1
Swarthmore College                              1
Lewis & Clark College                           1
DePauw University                               1
Gustavus Adolphus College                       1
University of Richmond                          1
Lafayette College                               1
Gettysburg College                              1
University of Puget Sound                       1
Denison University                              1
Franklin and Marshall College                   1
Colorado College (CC)                           1
Oberlin College                                 1
Fort Lewis College                              1
Name: School Name, dtype: int64
```

In [171]:

```
1 left_key[left_key.isin(right_key)].nunique()
```

Out[171]:

47

In [172]:

```
1 left_key[left_key.isin(right_key)].is_unique
```

Out[172]:

False

In [173]:

```
1 left_key[left_key.isin(right_key)].duplicated().size
```

Out[173]:

48

In [ ]:

```
1
```