Pandas Python Library for Data Science

Import Libraries

In [1]:
```
import numpy as np              # import numpy as np for convenience
import pandas as pd             # import pandas as pd for convenience
```

Load Data

In [2]:
```
df=pd.read_csv("What_does_aid_to_Africa_finance.csv")

df.head(3)
```

Out[2]:

	countryc	year	agrgdp	popn	infmort	schprim	schsec	grtdsbp	grlndsbp	aiddsbp	...	
0	Burkina Faso	1970	35.44188862	5633000.0	141.3999939	13	1	13.3182802200317	1.02303504943848	14.3413200378418	...	1.797693134€
1	Burkina Faso	1970	35.44188862	5633000.0	141.3999939	13	1	13.3182802200317	1.02303504943848	14.3413200378418	...	1.797693134€
2	Burkina Faso	1971	36.16739069	5740700.0	139.1999969	13.6	1.2	16.7043991088867	0.655763506889343	17.3601703643799	...	1.797693134€

3 rows × 50 columns

Splitting the Data

In [3]:
```
df_new=df.copy()
df1=df_new.sample(frac=0.25,random_state=0)
df_new=df_new.drop(df1.index)

df2=df_new.sample(frac=0.25,random_state=0)
df_new=df_new.drop(df2.index)

df3=df_new.sample(frac=0.25,random_state=0)

df4=df_new.drop(df3.index)
```

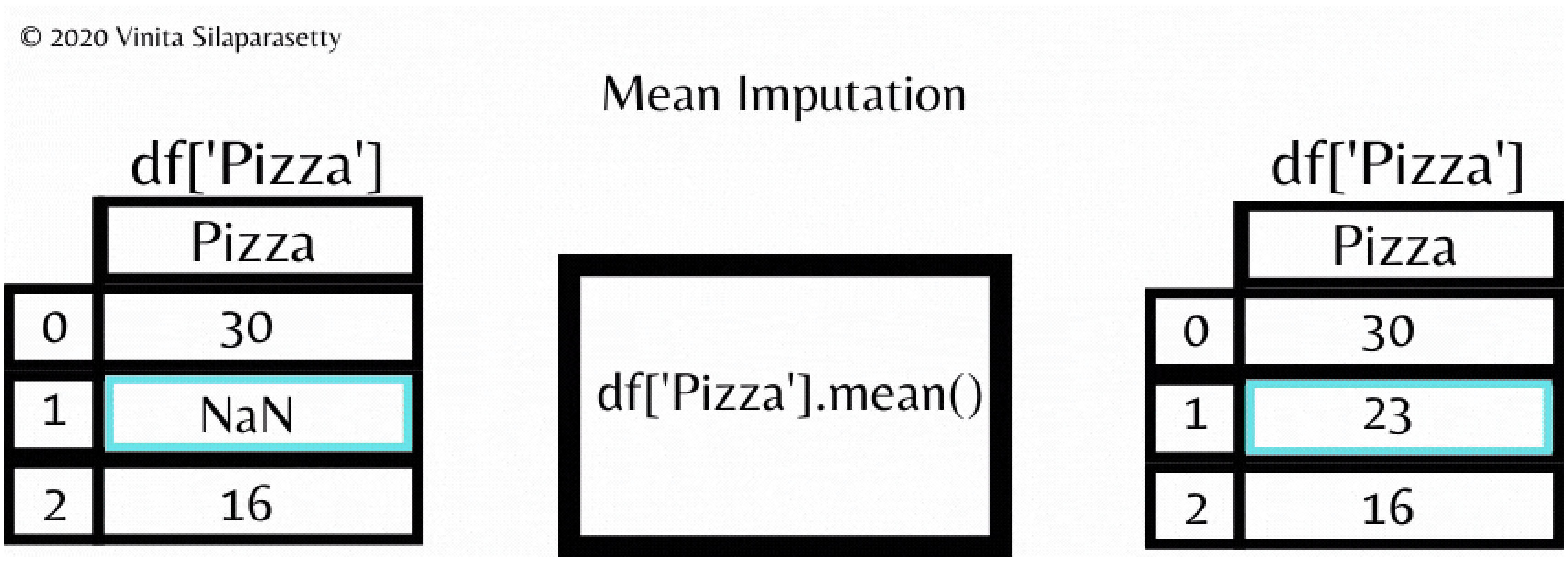
Handle Missing Values

Detect Missing Values

In [4]:
```
print(df3.isnull().sum())
```

```
countryc      0
year          0
agrgdp        0
popn          1
infmort       0
schprim       0
schsec        0
grtdsbp       0
grlndsbp      0
aiddsbp       0
totexpp       0
agexpp        0
enexpp        0
indexpp       0
tacexpp       0
eduexpp       0
hthexpp       0
prirepp       0
curexpp       0
capexpp       0
gdnpp         0
d0            0
cnlnagp       0
cnlnenp       0
cnlninp       0
cnlntacp      0
cnlnedup      0
cnlnhthp      0
cnlnothp      0
dgrtdsbp      0
dgrlndsbp     0
daiddsbp      0
dtotexpp      0
dagexpp       0
denexpp       0
dindexpp      0
dtacexpp      0
deduexpp      0
dhthexpp      0
dothexpp      0
dcurexpp      0
dcapexpp      0
dprirepp      0
dcnlnagp      0
dcnlnenp      0
dcnlninp      0
dcnlntacp     0
dcnlnedup     0
dcnlnhthp     0
dcnlnothp     0
dtype: int64
```

Input Missing Values



In [5]:
```
df3['popn'].mean()
```

Out[5]: 12570541.658536585

In [6]:
```
df3.isnull()
```

Out[6]:

	countryc	year	agrgdp	popn	infmort	schprim	schsec	grtdsbp	grlndsbp	aiddsbp	...	dcurexpp	dcapexpp	dprirepp	dcnlnagp	dcnlnenp
237	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False
244	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False
299	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False
87	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False
91	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False
260	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False
14	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False
157	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False
207	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False
160	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False
255	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False
9	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False
138	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False
181	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False
88	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False
33	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False
102	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False
164	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False
100	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False
104	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False
275	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False
171	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False
289	False	False	False	True	False	False	False	False	False	False	...	False	False	False	False	False
180	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False
61	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False
110	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False
281	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False
75	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False
43	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False
54	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False
213	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False
71	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False
16	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False
203	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False
153	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False
161	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False
273	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False
279	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False
40	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False
50	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False
193	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False

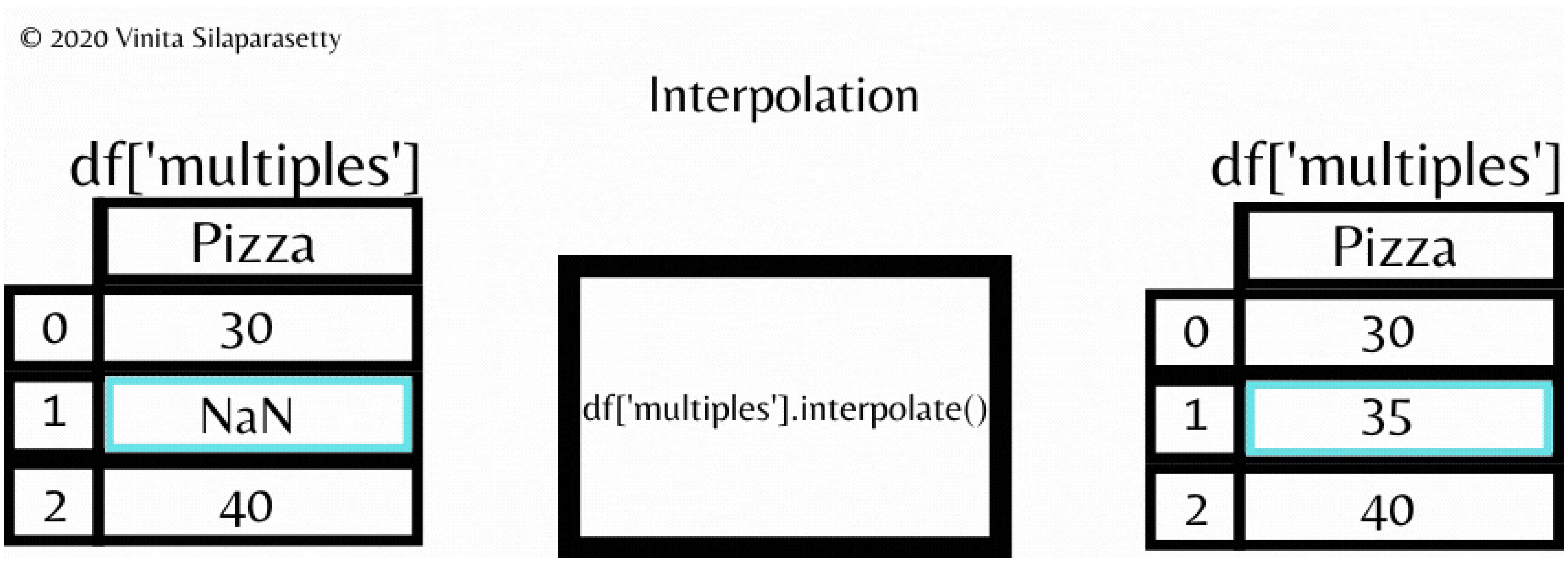
42 rows × 50 columns

In [7]:
```
df3['popn'].fillna(df3['popn'].mean(),inplace=True)
```

In [8]:
```
print(df3.isnull().sum())
```

```
countryc     0
year         0
agrgdp       0
popn         0
infmort      0
schprim      0
schsec       0
grtdsbp      0
grlndsbp     0
aiddsbp      0
totexpp      0
agexpp       0
enexpp       0
indexpp      0
tacexpp      0
eduexpp      0
hthexpp      0
prirepp      0
curexpp      0
capexpp      0
gdnpp        0
d0           0
cnlnagp      0
cnlnenp      0
cnlninp      0
cnlntacp     0
cnlnedup     0
cnlnhthp     0
cnlnothp     0
dgrtdsbp     0
dgrlndsbp    0
daiddsbp     0
dtotexpp     0
dagexpp      0
denexpp      0
dindexpp     0
dtacexpp     0
deduexpp     0
dhthexpp     0
dothexpp     0
dcurexpp     0
dcapexpp     0
dprirepp     0
dcnlnagp     0
dcnlnenp     0
dcnlninp     0
dcnlntacp    0
dcnlnedup    0
dcnlnhthp    0
dcnlnothp    0
dtype: int64
```

Interpolate Missing Values



In [9]:
```
# detect missing values
print(df1.isnull().sum())
```

```
countryc     0
year         0
agrgdp       0
popn         1
infmort      0
schprim      0
schsec       0
grtdsbp      0
grlndsbp     0
aiddsbp      0
totexpp      0
agexpp       0
enexpp       0
indexpp      0
tacexpp      0
eduexpp      0
hthexpp      0
prirepp      0
curexpp      0
capexpp      0
gdnpp        0
d0           0
cnlnagp      0
cnlnenp      0
cnlninp      0
cnlntacp     0
cnlnedup     0
cnlnhthp     0
cnlnothp     0
dgrtdsbp     0
dgrlndsbp    0
daiddsbp     0
dtotexpp     0
dagexpp      0
denexpp      0
dindexpp     0
dtacexpp     0
deduexpp     0
dhthexpp     0
dothexpp     0
dcurexpp     0
dcapexpp     0
dprirepp     0
dcnlnagp     0
dcnlnenp     0
dcnlninp     0
dcnlntacp    0
dcnlnedup    0
dcnlnhthp    0
dcnlnothp    0
dtype: int64
```

In [10]:
```
# interpolate missing values
df1['popn'].fillna(df1['popn'].interpolate(),inplace=True)
```

In [11]:
```
print(df1.isnull().sum())
```

```
countryc     0
year         0
agrgdp       0
popn         0
infmort      0
schprim      0
schsec       0
grtdsbp      0
grlndsbp     0
aiddsbp      0
totexpp      0
agexpp       0
enexpp       0
indexpp      0
tacexpp      0
eduexpp      0
hthexpp      0
prirepp      0
curexpp      0
capexpp      0
gdnpp        0
d0           0
cnlnagp      0
cnlnenp      0
cnlninp      0
cnlntacp     0
cnlnedup     0
cnlnhthp     0
cnlnothp     0
dgrtdsbp     0
dgrlndsbp    0
daiddsbp     0
dtotexpp     0
dagexpp      0
denexpp      0
dindexpp     0
dtacexpp     0
deduexpp     0
dhthexpp     0
dothexpp     0
dcurexpp     0
dcapexpp     0
dprirepp     0
dcnlnagp     0
dcnlnenp     0
dcnlninp     0
dcnlntacp    0
dcnlnedup    0
dcnlnhthp    0
dcnlnothp    0
dtype: int64
```

Assume that, a study has been conducted on the effects of infant mortality rate on each of the variables in the data frame. Detect missing values in df2 & decide on the best method to handle them

In [12]:
```
df2['popn'].fillna(df2['popn'].interpolate(),inplace=True)
```

In [13]:
```
df2.isnull().sum()
```

Out[13]:
```
countryc     0
year         0
agrgdp       0
popn         0
infmort      0
schprim      0
schsec       0
grtdsbp      0
grlndsbp     0
aiddsbp      0
totexpp      0
agexpp       0
enexpp       0
indexpp      0
tacexpp      0
eduexpp      0
hthexpp      0
prirepp      0
curexpp      0
capexpp      0
gdnpp        0
d0           0
cnlnagp      0
cnlnenp      0
cnlninp      0
cnlntacp     0
cnlnedup     0
cnlnhthp     0
cnlnothp     0
dgrtdsbp     0
dgrlndsbp    0
daiddsbp     0
dtotexpp     0
dagexpp      0
denexpp      0
dindexpp     0
dtacexpp     0
deduexpp     0
dhthexpp     0
dothexpp     0
dcurexpp     0
dcapexpp     0
dprirepp     0
dcnlnagp     0
dcnlnenp     0
dcnlninp     0
dcnlntacp    0
dcnlnedup    0
dcnlnhthp    0
dcnlnothp    0
dtype: int64
```

Combining Data

Joining

In [14]:
```
df5=df1.join(df2,lsuffix="_left")
df5
```

Out[14]:

	countryc_left	year_left	agrgdp_left	popn_left	infmort_left	schprim_left	schsec_left	grtdsbp_left	grlndsbp_left	aiddsbp_lef
223	Lesotho	1983	23.91304348	1483270.0	98	106.8	21	87.4762725830078	20.0394096374512	107.51570129394
150	Gambia, The	1988	31.22936246	841250.0	140.7799988	65	15.25	113.245697021484	31.0677890777588	144.31340026855
226	Lesotho	1986	21.14252061	1603960.0	92	109	23.4	66.7168197631836	10.6219997406006	77.33881378173
296	Malawi	1976	39.20110669	5409980.0	179.8	56	4	14.338809967041	15.4878797531128	29.82670021057
52	Botswana	1994	5.199306759	1420270.0	55.39999898	117	53	58.7042083740234	16.0785007476807	74.78270721435
...	...	...	...	...	...	...	...	...	...	..
20	Burkina Faso	1988	48.94457166	8534390.0	107.8	33.5	6.5	35.4728813171387	10.2006902694702	45.67356872558
46	Botswana	1988	7.060807251	1195140.0	56.6	111.75	35.75	135.776702880859	30.8545207977295	166.63119506835
158	Kenya	1970	33.29286623	11498000.0	102	58	9	9.42126178741455	9.48890781402588	18.91016960144
230	Lesotho	1990	19.96355858	1783000.0	84.6	105	25	67.9856872558594	26.8740100860596	94.85970306396
179	Kenya	1991	28.14106137	24015140.0	61.4	93	28	29.552059173584	21.131010055542	50.68307113647

75 rows × 100 columns

Concatenation

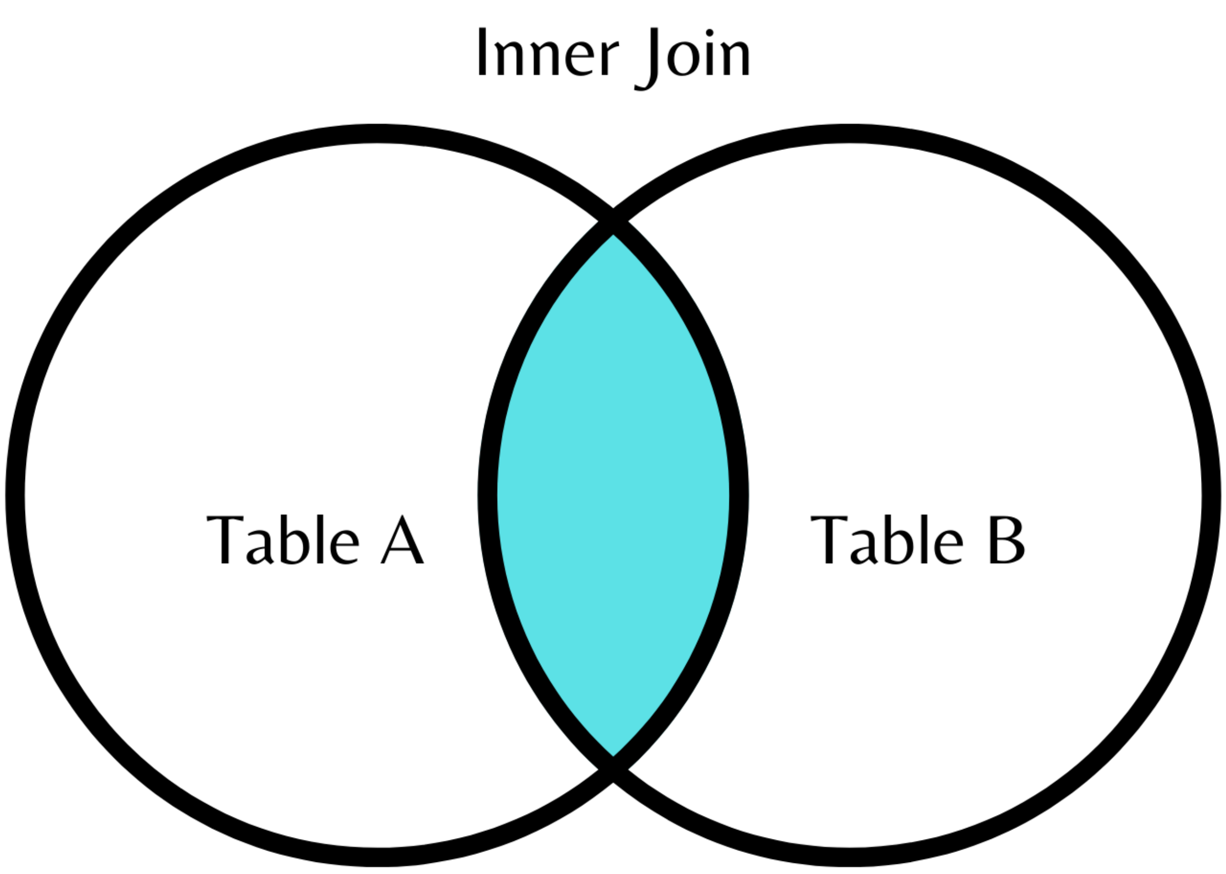
In [15]:
```
df6=pd.concat([df1,df2],axis=0)
df6
```

Out[15]:

	countryc	year	agrgdp	popn	infmort	schprim	schsec	grtdsbp	grlndsbp	aiddsbp	...
223	Lesotho	1983	23.91304348	1483270.0	98	106.8	21	87.4762725830078	20.0394096374512	107.515701293945	...
150	Gambia, The	1988	31.22936246	841250.0	140.7799988	65	15.25	113.245697021484	31.0677890777588	144.313400268555	...
226	Lesotho	1986	21.14252061	1603960.0	92	109	23.4	66.7168197631836	10.6219997406006	77.3388137817383	...
296	Malawi	1976	39.20110669	5409980.0	179.8	56	4	14.338809967041	15.4878797531128	29.8267002105713	...
52	Botswana	1994	5.199306759	1420270.0	55.39999898	117	53	58.7042083740234	16.0785007476807	74.7827072143555	...
...	...	...	...	...	...	...	...	...	...	...	...
240	Madagascar	1974	34.22234966	7408570.0	163.2	94	12	16.4736003875732	9.76997661590576	26.243579864502	...
256	Madagascar	1990	32.30721538	11672000.0	101.1600006	87	17	42.6463584899902	13.9560403823853	56.6024017333984	...
98	Ethiopia	1988	49.15748278	47643232.0	129.4	34.25	13.5	20.3594608306885	5.55403518676758	25.9134902954102	...
23	Burkina Faso	1991	34.66403162	9269910.0	104.2	37	8	38.1004219055176	15.298939704895	53.3993606567383	...
191	Liberia	1977	31.87008374	1708160.0	167	60.66666667	18.33333333	20.3501396179199	29.2217292785644	49.5718688964844	...

131 rows × 50 columns

Advanced Joins



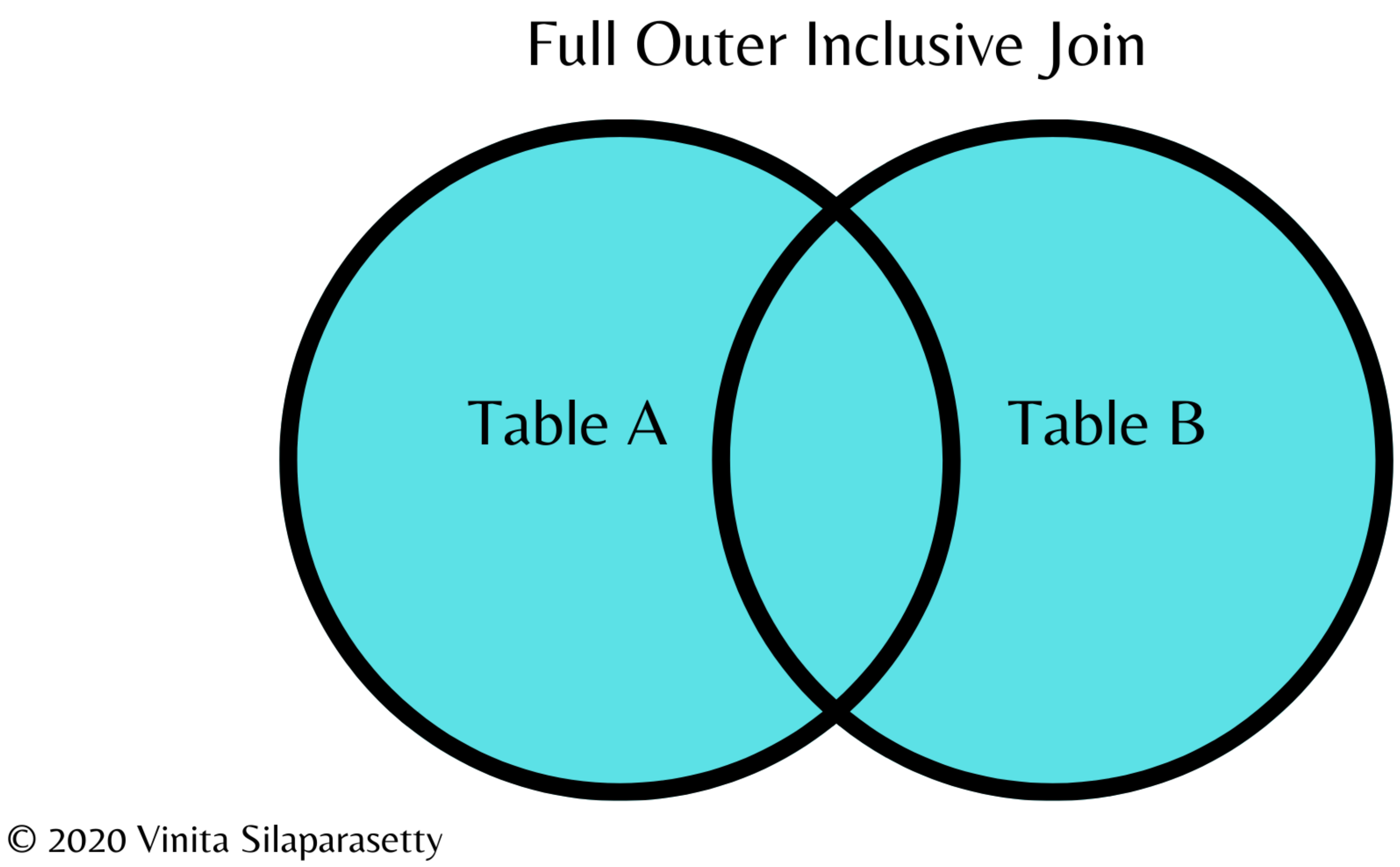
In [16]:

```
df7=pd.merge(df1,df2,on='countryc')
df7
```

Out[16]:

	countryc	year_x	agrgdp_x	popn_x	infmort_x	schprim_x	schsec_x	grtdsbp_x	grlndsbp_x	aiddsbp_x	...	
0	Lesotho	1983	23.91304348	1483270.0	98	106.8	21	87.4762725830078	20.0394096374512	107.515701293945	...	
1	Lesotho	1983	23.91304348	1483270.0	98	106.8	21	87.4762725830078	20.0394096374512	107.515701293945	...	1.797
2	Lesotho	1983	23.91304348	1483270.0	98	106.8	21	87.4762725830078	20.0394096374512	107.515701293945	...	1.797
3	Lesotho	1983	23.91304348	1483270.0	98	106.8	21	87.4762725830078	20.0394096374512	107.515701293945	...	1.797
4	Lesotho	1983	23.91304348	1483270.0	98	106.8	21	87.4762725830078	20.0394096374512	107.515701293945	...	
...	...	...	...	...	...	...	...	...	...	...	...	
356	Ghana	1970	46.51883327	8614000.0	110.5999985	64	14	8.47124862670898	16.3063297271728	24.7775802612305	...	
357	Ghana	1970	46.51883327	8614000.0	110.5999985	64	14	8.47124862670898	16.3063297271728	24.7775802612305	...	
358	Ghana	1970	46.51883327	8614000.0	110.5999985	64	14	8.47124862670898	16.3063297271728	24.7775802612305	...	-
359	Ghana	1970	46.51883327	8614000.0	110.5999985	64	14	8.47124862670898	16.3063297271728	24.7775802612305	...	
360	Ghana	1970	46.51883327	8614000.0	110.5999985	64	14	8.47124862670898	16.3063297271728	24.7775802612305	...	1.797

361 rows × 99 columns

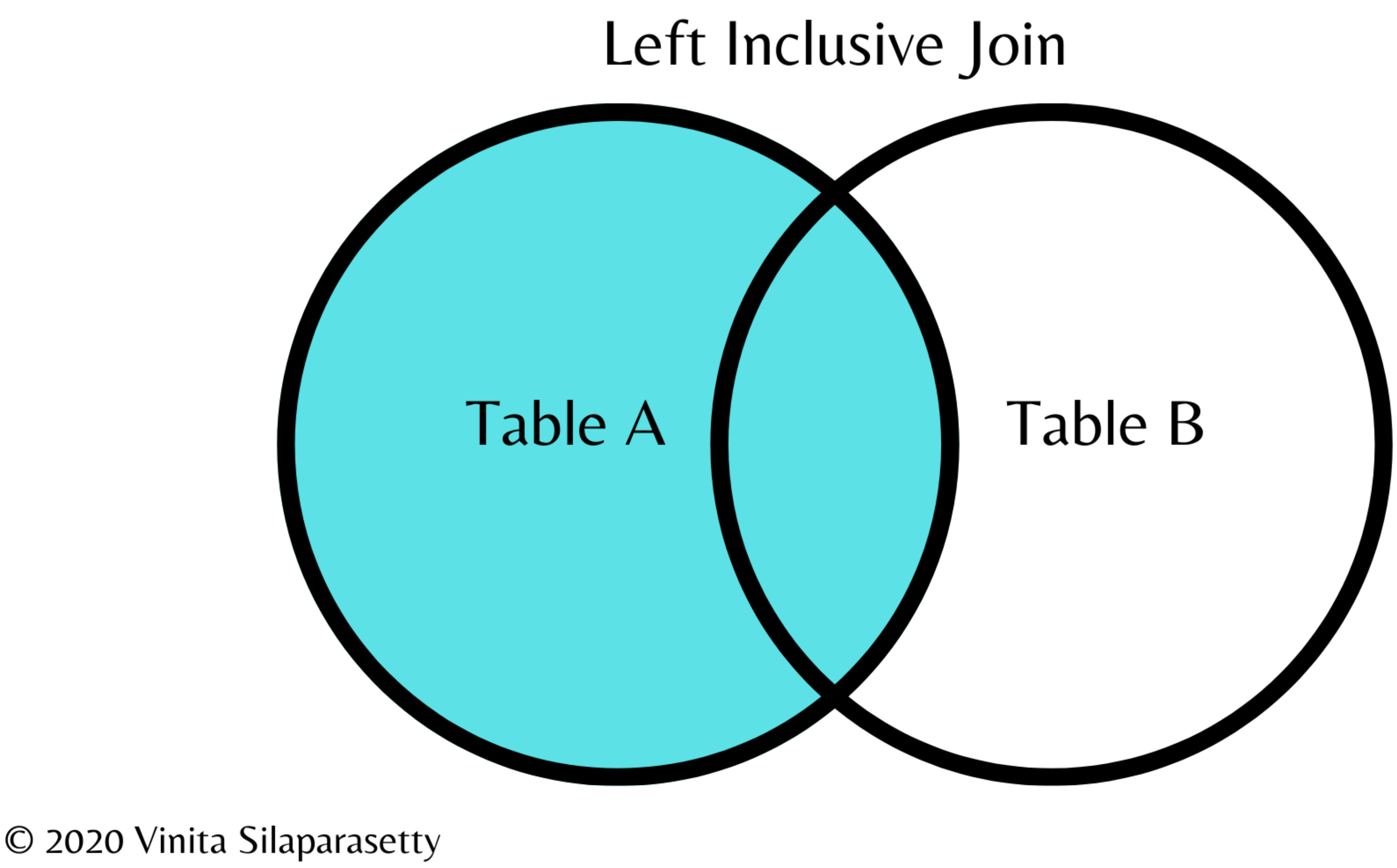


In [17]:
```
df8=pd.merge(df1,df2,how='outer')
df8
```

Out[17]:

	countryc	year	agrgdp	popn	infmort	schprim	schsec	grtdsbp	grlndsbp	aiddsbp	...
0	Lesotho	1983	23.91304348	1483270.0	98	106.8	21	87.4762725830078	20.0394096374512	107.515701293945	...
1	Gambia, The	1988	31.22936246	841250.0	140.7799988	65	15.25	113.245697021484	31.0677890777588	144.313400268555	...
2	Lesotho	1986	21.14252061	1603960.0	92	109	23.4	66.7168197631836	10.6219997406006	77.3388137817383	...
3	Malawi	1976	39.20110669	5409980.0	179.8	56	4	14.338809967041	15.4878797531128	29.8267002105713	...
4	Botswana	1994	5.199306759	1420270.0	55.39999898	117	53	58.7042083740234	16.0785007476807	74.7827072143555	...
...	...	...	...	...	...	...	...	...	...	...	...
126	Madagascar	1974	34.22234966	7408570.0	163.2	94	12	16.4736003875732	9.76997661590576	26.243579864502	...
127	Madagascar	1990	32.30721538	11672000.0	101.1600006	87	17	42.6463584899902	13.9560403823853	56.6024017333984	...
128	Ethiopia	1988	49.15748278	47643232.0	129.4	34.25	13.5	20.3594608306885	5.55403518676758	25.9134902954102	...
129	Burkina Faso	1991	34.66403162	9269910.0	104.2	37	8	38.1004219055176	15.298939704895	53.3993606567383	...
130	Liberia	1977	31.87008374	1708160.0	167	60.66666667	18.33333333	20.3501396179199	29.2217292785644	49.5718688964844	...

131 rows × 50 columns



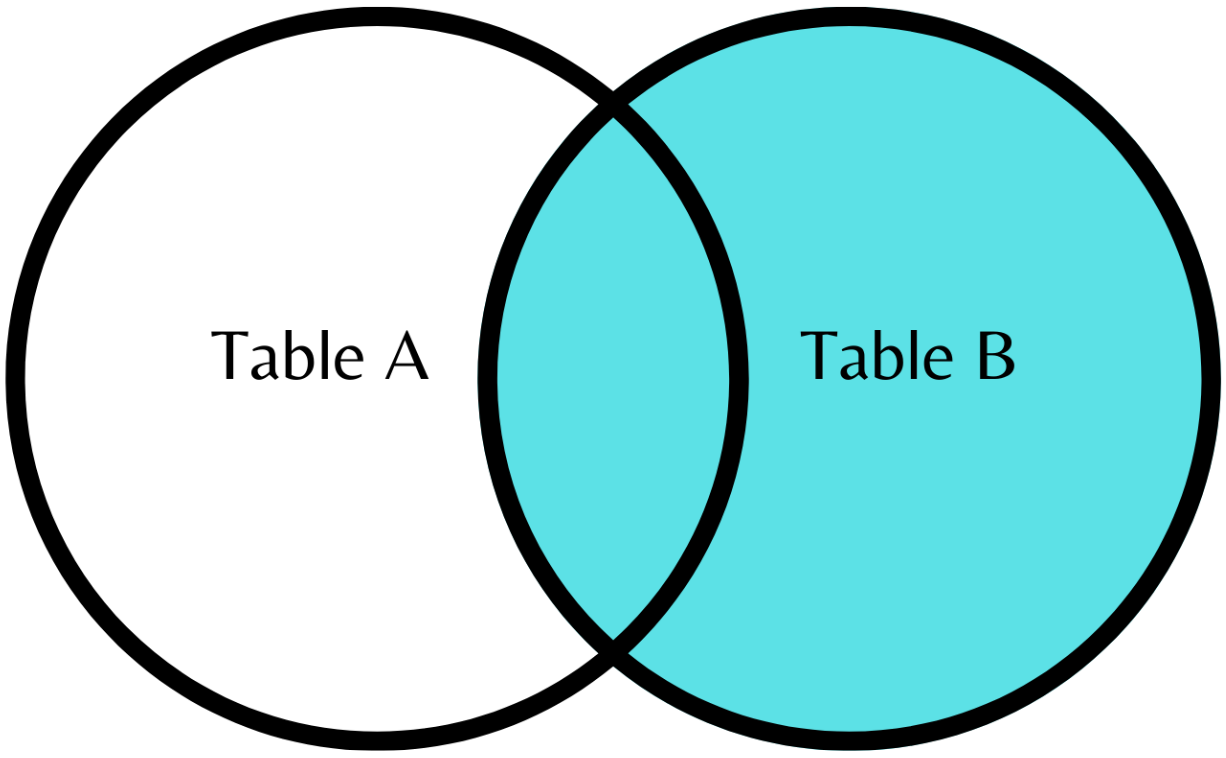
In [18]:
```
df9=pd.merge(df1,df2,how='left')
df9
```

Out[18]:

	countryc	year	agrgdp	popn	infmort	schprim	schsec	grtdsbp	grlndsbp	aiddsbp	...	
0	Lesotho	1983	23.91304348	1483270.0	98	106.8	21	87.4762725830078	20.0394096374512	107.515701293945	...	1.79769313
1	Gambia, The	1988	31.22936246	841250.0	140.7799988	65	15.25	113.245697021484	31.0677890777588	144.313400268555	...	1.79769313
2	Lesotho	1986	21.14252061	1603960.0	92	109	23.4	66.7168197631836	10.6219997406006	77.3388137817383	...	1.79769313
3	Malawi	1976	39.20110669	5409980.0	179.8	56	4	14.338809967041	15.4878797531128	29.8267002105713	...	-3.285
4	Botswana	1994	5.199306759	1420270.0	55.39999898	117	53	58.7042083740234	16.0785007476807	74.7827072143555	...	
...	...	...	...	...	...	...	...	...	...	...	...	
70	Burkina Faso	1988	48.94457166	8534390.0	107.8	33.5	6.5	35.4728813171387	10.2006902694702	45.6735687255859	...	1.79769313
71	Botswana	1988	7.060807251	1195140.0	56.6	111.75	35.75	135.776702880859	30.8545207977295	166.631195068359	...	160.9
72	Kenya	1970	33.29286623	11498000.0	102	58	9	9.42126178741455	9.48890781402588	18.9101696014404	...	1.79769313
73	Lesotho	1990	19.96355858	1783000.0	84.6	105	25	67.9856872558594	26.8740100860596	94.8597030639648	...	-0.9964
74	Kenya	1991	28.14106137	24015140.0	61.4	93	28	29.552059173584	21.131010055542	50.6830711364746	...	-4.143

75 rows × 50 columns

Right Inclusive Join



In [19]:
```
df10=pd.merge(df1,df2,how='right')
df10
```

Out[19]:

	countryc	year	agrgdp	popn	infmort	schprim	schsec	grtdsbp	
0	Kenya	1980	32.59223808	16560000.0	72.40000153	115	20	26.8336296081543	17.75713
1	Ghana	1973	48.97463727	9388140.0	106	68.2	27.8	7.54381990432739	7.859230
2	Liberia	1990	1.79769313486232e+308	2435000.0	176.8	1.79769313486232e+308	1.79769313486232e+308	38.6235008239746	16.28816
3	Madagascar	1981	33.07584521	8951460.0	134	1.79769313486232e+308	1.79769313486232e+308	17.8668098449707	25.69960
4	Ethiopia	1989	48.50468489	49337260.0	126.8	34	14	15.0112895965576	3.978796
5	Lesotho	1988	24.40058125	1689570.0	88.2	107	24.2	67.3578262329102	17.2617
6	Kenya	1984	33.91489131	19302140.0	64.8	102.2	20.8	20.7450504302978	10.90307
7	Ethiopia	1987	49.64547916	46087060.0	132	34.5	13	13.8647003173828	4.371497
8	Mauritius	1984	14.406639	1011330.0	26.4	106.6	46.8	30.8346195220947	23.71894
9	Lesotho	1980	23.58414239	1367000.0	108.4000015	102	18	109.822998046875	14.00716
10	Liberia	1974	32.28125548	1560820.0	175.4	60.8	15.6	20.0104808807373	12.83183
11	Ghana	1990	47.85769483	14870000.0	82.74000092	77	37	39.5492897033691	28.96863
12	Ethiopia	1974	56.27994714	32098120.0	152.6	22.4	5.6	7.50680685043335	3.329279
13	Burkina Faso	1979	34.21393892	6797540.0	123	18	3	56.8610992431641	12.25341
14	Liberia	1984	35.45335614	2140800.0	148.6	48	22	58.6903495788574	33.41609
15	Madagascar	1986	36.78584035	10287560.0	116.7200012	1.79769313486232e+308	1.79769313486232e+308	16.711820602417	26.50024
16	Liberia	1971	24.425	1426840.0	179.4000015	57.2	11.4	27.8442993164062	7.067080
17	Burkina Faso	1974	36.48014145	6075700.0	133	15.4	1.8	38.6305809020996	6.662034
18	Cameroon	1973	30.79282681	7021850.0	115.6	93.8	10.6	18.3287105560303	11.28155
19	Mauritius	1983	13.8038255	1003930.0	27.2	103.2	47.6	33.3487701416016	30.45845
20	Madagascar	1991	32.98051479	12054150.0	97.08000031	79	15	31.4275207519531	19.66687
21	Liberia	1973	29.92056487	1514530.0	178.2	59.6	14.2	23.8042697906494	2.479192
22	Lesotho	1979	30.48723898	1328910.0	112.600001	104	17	83.5996475219726	14.09477
23	Lesotho	1995	10.08687856	1980000.0	75.59999847	1.79769313486232e+308	1.79769313486232e+308	48.4444389343262	13.53030
24	Gambia, The	1979	31.23783032	622460.0	161.799998	42	12	69.366943359375	47.23722
25	Ethiopia	1973	56.27994714	31273540.0	153.8	20.8	5.2	4.12413215637207	2.609184
26	Mauritius	1973	19.9317554	859470.0	51.6	101.8	35.4	36.8324813842773	15.52149
27	Botswana	1990	5.456680968	4370235.0	55.8	114	42	125.423896789551	14.68628
28	Burkina Faso	1985	37.93019376	7881000.0	112.2	29	5	27.2893295288086	7.285047
29	Kenya	1971	31.37739887	11903370.0	100	65.4	9.8	10.6624298095703	9.919784
30	Liberia	1995	1.79769313486232e+308	2733000.0	171.8000031	1.79769313486232e+308	1.79769313486232e+308	45.2506408691406	
31	Kenya	1986	33.04255174	20683850.0	63.6	98.2	22.16666667	22.578649520874	8.994091
32	Gambia, The	1977	33.64667747	585370.0	167	33	11	53.4062881469727	31.28305
33	Cameroon	1988	23.9429277	10835870.0	71.4	102.75	25.25	20.7212104797363	15.42134
34	Liberia	1985	36.52958877	2199000.0	146.4	48	22	39.3754997253418	19.3882
35	Burkina Faso	1993	34.90956072	9804460.0	101.8000005	38	8	39.2027893066406	14.15841
36	Burkina Faso	1989	31.72235372	8770560.0	106.6	35	7	28.7813892364502	11.1896
37	Liberia	1981	31.56142828	1941970.0	155.8000031	48	22	58.0791893005371	38.80352
38	Burkina Faso	1978	36.08594394	6639120.0	125	17	2	41.1725006103516	12.00825
39	Ghana	1989	48.96733723	14425360.0	85.16000061	74	39	22.296459197998	40.31914
40	Lesotho	1991	12.096718	1820770.0	82.8	105	24	59.3357810974121	19.9711
41	Mauritius	1981	14.34112949	981170.0	30	96.4	49.2	57.4679794311524	43.15922
42	Liberia	1975	26.59668835	1609000.0	172.6	62	17	24.1224708557129	15.0526
43	Gambia, The	1978	30.6446491	603960.0	164.399999	37	11	67.2458572387695	57.74449
44	Ghana	1984	49.24249984	12167740.0	94.8	76.8	40.2	15.3000497817993	12.2099
45	Ethiopia	1984	48.82879875	42152320.0	148.2	34.8	11.2	9.63077259063721	2.928190
46	Botswana	1984	7.441225106	1037290.0	61.2	102.2	27	120.155899047852	24.16527
47	Mauritius	1990	12.10414774	1057000.0	20.4	109	53	48.7851715087891	64.72013
48	Kenya	1990	29.51903784	23354000.0	61.8	95	22.16666667	58.4292602539062	27.14094

	countryc	year	agrgdp	popn	infmort	schprim	schsec	grtdsbp	
49	Mauritius	1977	19.66080402	913710.0	38	109	46	33.2450904846191	24.76597
50	Ghana	1995	46.27585233	17075000.0	72.73999786	80	40	20.2940006256104	26.37949
51	Madagascar	1974	34.22234966	7408570.0	163.2	94	12	16.4736003875732	9.769976
52	Madagascar	1990	32.30721538	11672000.0	101.1600006	87	17	42.6463584899902	13.95604
53	Ethiopia	1988	49.15748278	47643232.0	129.4	34.25	13.5	20.3594608306885	5.554035
54	Burkina Faso	1991	34.66403162	9269910.0	104.2	37	8	38.1004219055176	15.2989
55	Liberia	1977	31.87008374	1708160.0	167	60.66666667	18.33333333	20.3501396179199	29.22172

56 rows × 50 columns

Generate a new data frame which contains information only from those African countries with the same "Agriculture as a share of GDP"

In [20]:
```python
df11=pd.merge(df1,df2,on='agrgdp')
df11
```

Out[20]:

	countryc_x	year_x	agrgdp	popn_x	infmort_x	schprim_x	schsec_x	grtdsbp_x	
0	Ethiopia	1979	56.27994714	36696848.0	153	37	10	5.80811786651611	4.587
1	Ethiopia	1979	56.27994714	36696848.0	153	37	10	5.80811786651611	4.587
2	Ethiopia	1980	56.27994714	37717000.0	155	34	8	7.98918008804321	2.262
3	Ethiopia	1980	56.27994714	37717000.0	155	34	8	7.98918008804321	2.262
4	Liberia	1991	1.79769313486232e+308	2483450.0	188.4	1.79769313486232e+308	1.79769313486232e+308	70.8964462280274	0.022293
5	Liberia	1991	1.79769313486232e+308	2483450.0	188.4	1.79769313486232e+308	1.79769313486232e+308	70.8964462280274	0.022293
6	Ethiopia	1971	1.79769313486232e+308	29698260.0	156.4000015	17.6	4.4	2.73099207878113	2.6246
7	Ethiopia	1971	1.79769313486232e+308	29698260.0	156.4000015	17.6	4.4	2.73099207878113	2.6246

8 rows × 99 columns

Sorting

Sort values by a single column

In [21]:
```python
df1.sort_values(by=['agrgdp'],ascending=True)
```

Out[21]:

	countryc	year	agrgdp	popn	infmort	schprim	schsec	grtdsbp	
81	Ethiopia	1971	1.79769313486232e+308	29698260.0	156.4000015	17.6	4.4	2.73099207878113	2.62460
205	Liberia	1991	1.79769313486232e+308	2483450.0	188.4	1.79769313486232e+308	1.79769313486232e+308	70.8964462280274	0.0222931
285	Mauritius	1992	10.82725922	1081000.0	18	107	57	32.6669311523438	35.577
282	Mauritius	1989	12.32418189	1048560.0	21.6	109.2	51.6	44.3517112731934	42.4502
234	Lesotho	1994	13.69297806	1938930.0	77.39999898	1.79769313486232e+308	1.79769313486232e+308	43.487491607666	22.0319
...	...	...	...	...	...	...	...	...	
90	Ethiopia	1980	56.27994714	37717000.0	155	34	8	7.98918008804321	2.2627
118	Ghana	1982	57.34115279	11366410.0	98	78.4	40.6	8.70774555206299	13.2136
101	Ethiopia	1991	59.13332578	52954000.0	121.6	25	12	20.6525192260742	2.57341
46	Botswana	1988	7.060807251	1195140.0	56.6	111.75	35.75	135.776702880859	30.8545
286	Mauritius	1993	9.715403179	1097000.0	17.35999997	106	59	31.8793106079102	24.5195

75 rows × 50 columns

Sort values by row labels

In [22]:
```
df1.sort_index(axis=0,ascending=True)
```

Out[22]:

	countryc	year	agrgdp	popn	infmort	schprim	schsec	grtdsbp	grlndsbp	aiddsbp	...	
5	Burkina Faso	1973	34.83428571	5958700.0	135	14.8	1.6	25.9791507720947	3.87817406654358	29.8573207855225	...	1.79769313
7	Burkina Faso	1975	34.27100776	6202000.0	131	16	2	30.4486293792725	7.36860179901123	37.8172302246094	...	5.940
8	Burkina Faso	1976	34.80431988	804215.0	129	15	2	24.3181304931641	8.15182018280029	32.4699592590332	...	0.03461
12	Burkina Faso	1980	33.24267254	6962000.0	121	18	3	41.1754417419434	14.5752201080322	55.7506484985352	...	-3.462
15	Burkina Faso	1983	31.9042673	7490710.0	115.4	24.6	4.2	27.2162609100342	9.1533613204956	36.3696212768555	...	-10.63
...	...	...	...	...	...	...	...	...	...	...	...	
285	Mauritius	1992	10.82725922	1081000.0	18	107	57	32.6669311523438	35.577751159668	68.2446670532226	...	47.52
286	Mauritius	1993	9.715403179	1097000.0	17.35999997	106	59	31.8793106079102	24.5195999145508	56.3989105224609	...	-61.58
294	Malawi	1974	41.17239788	5087140.0	185.4	51.6	4	8.73302841186524	15.7701101303101	24.5031394958496	...	0.467
295	Malawi	1975	37.23468769	5244000.0	182.6	53.8	4	10.5895004272461	22.165210723877	32.754711151123	...	0.5157
296	Malawi	1976	39.20110669	5409980.0	179.8	56	4	14.338809967041	15.4878797531128	29.8267002105713	...	-3.285

75 rows × 50 columns

Sort the values in df3 according to the least 'Secondary School Enrolment Rate'. Select the best method for sorting to solve this challenge.

In [23]:

```
df3.sort_values(by=['schsec'],ascending=True)
```

Out[23]:

	countryc	year	agrgdp	popn	infmort	schprim	schsec	
4	Burkina Faso	1973	34.83428571	5.958700e+06	135	14.8	1.6	25.9791
244	Madagascar	1978	32.22944414	8.251580e+06	146	100	1.79769313486232e+308	16.9368
289	Mauritius	1996	9.393333333	1.257054e+07	1.79769313486232e+308	1.79769313486232e+308	1.79769313486232e+308	1.7976931348
207	Liberia	1993	1.79769313486232e+308	2.596430e+06	190.600001	1.79769313486232e+308	1.79769313486232e+308	50.0916
203	Liberia	1989	1.79769313486232e+308	2.391540e+06	165.2	1.79769313486232e+308	1.79769313486232e+308	31.9660
160	Kenya	1972	35.19454123	1.232986e+07	98	72.8	10.6	22.445
213	Lesotho	1973	42.52275683	1.131220e+06	128.2	97.8	10.6	38.4818
138	Gambia, The	1976	37.72430669	5.666900e+05	169.4	31	11	34.9581
102	Ethiopia	1992	64.4192191	5.479000e+07	119	23	11	21.1613
161	Kenya	1973	35.46295124	1.277806e+07	94.8	80.2	11.4	12.9125
104	Ethiopia	1994	56.9741607	5.489000e+07	114.200002	24.3	11.7	14.2693
237	Madagascar	1971	24.30771479	6.901230e+06	176.5999985	91	12	20.3827
100	Ethiopia	1990	49.26866451	5.118000e+07	124.2	31	13	19.3059
164	Kenya	1976	37.90313747	1.425500e+07	85.2	97	15	18.6398
153	Gambia, The	1991	28.27984753	9.646900e+05	134.1199951	65	16	81.0120
61	Cameroon	1977	33.64508393	7.920570e+06	102	101	16	24.3732
33	Botswana	1975	31.60036166	7.551000e+05	80.8	72	16	105.040
260	Madagascar	1994	38.99940568	1.324875e+07	90.36000061	89	18	21.136
157	Gambia, The	1995	1.79769313486232e+308	1.113000e+06	125.9599991	76	19	36.8104
9	Burkina Faso	1977	34.31152713	6.486870e+06	127	16	2	27.812
171	Kenya	1983	34.21715536	1.859766e+07	65.4	105.4	20.6	21.9400
40	Botswana	1982	12.55695077	9.669400e+05	64	96.6	23	148.528
193	Liberia	1979	34.27732326	1.819140e+06	161.4000041	71	23	38.9656
255	Madagascar	1989	32.93639906	1.130174e+07	105.2400009	101	23	22.0431
71	Cameroon	1987	23.9858232	1.053526e+07	74	102.5	24.5	17.5125
181	Kenya	1993	31.22035451	2.534740e+07	60	91	25	22.8666
180	Kenya	1992	28.50164438	2.467985e+07	61	92	27	28.8002
43	Botswana	1985	6.496773488	1.075000e+06	59.8	105	29	115.636
14	Burkina Faso	1982	32.11751268	7.308230e+06	117	22.4	3.8	35.7688
75	Cameroon	1991	24.25394355	1.182539e+07	63.6	96.33333333	32	35.7696
110	Ghana	1974	51.13838759	9.621420e+06	105	69.6	32.4	9.25522
299	Malawi	1979	39.64158617	5.947940e+06	171.4000041	59	4	45.1732
16	Burkina Faso	1984	32.90579486	7.681280e+06	113.8	26.8	4.6	28.1816
279	Mauritius	1986	15.25835866	1.024540e+06	24.8	109.8	47.4	46.8331
275	Mauritius	1982	15.26946108	9.938500e+05	28	99.8	48.4	37.610
273	Mauritius	1980	12.3697388	9.660000e+05	32	93	50	44.5064
281	Mauritius	1988	13.05016417	1.040330e+06	22.8	109.4	50.2	39.6210
50	Botswana	1992	5.088348271	1.353060e+06	55	116	52	88.8960
87	Ethiopia	1977	56.27994714	3.475976e+07	149	24	7	4.89760
54	Cameroon	1970	31.36392206	6.506000e+06	125.8000031	89	7	20.725
91	Ethiopia	1981	56.27994714	3.877237e+07	157	34.2	8.8	8.39179
88	Ethiopia	1978	56.27994714	3.475976e+07	151	29	9	5.58737

42 rows × 50 columns

Selection

Select columns by their names

In [24]:
```
df1[['countryc','year']]
```

Out[24]:

	countryc	year
223	Lesotho	1983
150	Gambia, The	1988
226	Lesotho	1986
296	Malawi	1976
52	Botswana	1994
...	...	...
20	Burkina Faso	1988
46	Botswana	1988
158	Kenya	1970
230	Lesotho	1990
179	Kenya	1991

75 rows × 2 columns

Select columns by their index

In [25]:
```
df1[df1.columns[1:8]].head()
```

Out[25]:

	year	agrgdp	popn	infmort	schprim	schsec	grtdsbp
223	1983	23.91304348	1483270.0	98	106.8	21	87.4762725830078
150	1988	31.22936246	841250.0	140.7799988	65	15.25	113.245697021484
226	1986	21.14252061	1603960.0	92	109	23.4	66.7168197631836
296	1976	39.20110669	5409980.0	179.8	56	4	14.338809967041
52	1994	5.199306759	1420270.0	55.39999898	117	53	58.7042083740234

Slicing

In [26]:
```
df.iloc[:3]
```

Out[26]:

	countryc	year	agrgdp	popn	infmort	schprim	schsec	grtdsbp	grlndsbp	aiddsbp	...	
0	Burkina Faso	1970	35.44188862	5633000.0	141.3999939	13	1	13.3182802200317	1.02303504943848	14.3413200378418	...	1.7976931348
1	Burkina Faso	1970	35.44188862	5633000.0	141.3999939	13	1	13.3182802200317	1.02303504943848	14.3413200378418	...	1.7976931348
2	Burkina Faso	1971	36.16739069	5740700.0	139.1999969	13.6	1.2	16.7043991088867	0.655763506889343	17.3601703643799	...	1.7976931348

3 rows × 50 columns

In [27]:
```
df.loc[:3]
```

Out[27]:

	countryc	year	agrgdp	popn	infmort	schprim	schsec	grtdsbp	grlndsbp	aiddsbp	...	
0	Burkina Faso	1970	35.44188862	5633000.0	141.3999939	13	1	13.3182802200317	1.02303504943848	14.3413200378418	...	1.7976931348
1	Burkina Faso	1970	35.44188862	5633000.0	141.3999939	13	1	13.3182802200317	1.02303504943848	14.3413200378418	...	1.7976931348
2	Burkina Faso	1971	36.16739069	5740700.0	139.1999969	13.6	1.2	16.7043991088867	0.655763506889343	17.3601703643799	...	1.7976931348
3	Burkina Faso	1972	37.51058767	5848380.0	137	14.2	1.4	20.9176502227783	2.97720909118652	23.8948593139648	...	1.7976931348

4 rows × 50 columns

Select only those values in df4 that relate to the concessionary loans to various sectors in Africa

In [28]:
```
df4[df4.columns[23:29]].head()
```

Out[28]:

	cnlnenp	cnlninp	cnlntacp	cnlnedup	cnlnhthp	cnlnothp
0	0	0.320718288421631	0.00690389983355999	0	0	0.289336889982224
1	0	0.320718288421631	0.00690389983355999	0	0	0.289336889982224
2	0	0.317928194999695	0.00292750005610287	0	0	0.124712198972702
3	0	0.185248598456383	0.567323684692383	0	0	0.616943776607513
13	0.703820884227753	2.31013488769531	0.713891804218292	0.320209890604019	0.201189801096916	3.35064506530762

Grouping

Group by multiple columns

In [29]:
```
df1.groupby(['countryc','year']).groups
```

Out[29]:
```
{('Botswana', 1976): Int64Index([34], dtype='int64'),
 ('Botswana', 1987): Int64Index([45], dtype='int64'),
 ('Botswana', 1988): Int64Index([46], dtype='int64'),
 ('Botswana', 1994): Int64Index([52], dtype='int64'),
 ('Burkina Faso', 1973): Int64Index([5], dtype='int64'),
 ('Burkina Faso', 1975): Int64Index([7], dtype='int64'),
 ('Burkina Faso', 1976): Int64Index([8], dtype='int64'),
 ('Burkina Faso', 1980): Int64Index([12], dtype='int64'),
 ('Burkina Faso', 1983): Int64Index([15], dtype='int64'),
 ('Burkina Faso', 1988): Int64Index([20], dtype='int64'),
 ('Burkina Faso', 1990): Int64Index([22], dtype='int64'),
 ('Burkina Faso', 1994): Int64Index([26], dtype='int64'),
 ('Cameroon', 1971): Int64Index([55], dtype='int64'),
 ('Cameroon', 1975): Int64Index([59], dtype='int64'),
 ('Cameroon', 1979): Int64Index([63], dtype='int64'),
 ('Cameroon', 1980): Int64Index([64], dtype='int64'),
 ('Cameroon', 1982): Int64Index([66], dtype='int64'),
 ('Cameroon', 1989): Int64Index([73], dtype='int64'),
 ('Cameroon', 1990): Int64Index([74], dtype='int64'),
 ('Ethiopia', 1971): Int64Index([81], dtype='int64'),
 ('Ethiopia', 1979): Int64Index([89], dtype='int64'),
 ('Ethiopia', 1980): Int64Index([90], dtype='int64'),
 ('Ethiopia', 1982): Int64Index([92], dtype='int64'),
 ('Ethiopia', 1991): Int64Index([101], dtype='int64'),
 ('Gambia, The', 1971): Int64Index([133], dtype='int64'),
 ('Gambia, The', 1974): Int64Index([136], dtype='int64'),
 ('Gambia, The', 1975): Int64Index([137], dtype='int64'),
 ('Gambia, The', 1982): Int64Index([144], dtype='int64'),
 ('Gambia, The', 1988): Int64Index([150], dtype='int64'),
 ('Gambia, The', 1990): Int64Index([152], dtype='int64'),
 ('Ghana', 1970): Int64Index([106], dtype='int64'),
 ('Ghana', 1972): Int64Index([108], dtype='int64'),
 ('Ghana', 1975): Int64Index([111], dtype='int64'),
 ('Ghana', 1982): Int64Index([118], dtype='int64'),
 ('Ghana', 1986): Int64Index([122], dtype='int64'),
 ('Ghana', 1993): Int64Index([129], dtype='int64'),
 ('Kenya', 1970): Int64Index([158], dtype='int64'),
 ('Kenya', 1978): Int64Index([166], dtype='int64'),
 ('Kenya', 1985): Int64Index([173], dtype='int64'),
 ('Kenya', 1987): Int64Index([175], dtype='int64'),
 ('Kenya', 1988): Int64Index([176], dtype='int64'),
 ('Kenya', 1991): Int64Index([179], dtype='int64'),
 ('Kenya', 1994): Int64Index([182], dtype='int64'),
 ('Lesotho', 1972): Int64Index([212], dtype='int64'),
 ('Lesotho', 1974): Int64Index([214], dtype='int64'),
 ('Lesotho', 1975): Int64Index([215], dtype='int64'),
 ('Lesotho', 1976): Int64Index([216], dtype='int64'),
 ('Lesotho', 1981): Int64Index([221], dtype='int64'),
 ('Lesotho', 1983): Int64Index([223], dtype='int64'),
 ('Lesotho', 1984): Int64Index([224], dtype='int64'),
 ('Lesotho', 1985): Int64Index([225], dtype='int64'),
 ('Lesotho', 1986): Int64Index([226], dtype='int64'),
 ('Lesotho', 1987): Int64Index([227], dtype='int64'),
 ('Lesotho', 1990): Int64Index([230], dtype='int64'),
 ('Lesotho', 1994): Int64Index([234], dtype='int64'),
 ('Liberia', 1970): Int64Index([184], dtype='int64'),
 ('Liberia', 1976): Int64Index([190], dtype='int64'),
 ('Liberia', 1987): Int64Index([201], dtype='int64'),
 ('Liberia', 1991): Int64Index([205], dtype='int64'),
 ('Madagascar', 1972): Int64Index([238], dtype='int64'),
 ('Madagascar', 1973): Int64Index([239], dtype='int64'),
 ('Madagascar', 1975): Int64Index([241], dtype='int64'),
 ('Madagascar', 1980): Int64Index([246], dtype='int64'),
 ('Madagascar', 1984): Int64Index([250], dtype='int64'),
 ('Madagascar', 1987): Int64Index([253], dtype='int64'),
 ('Madagascar', 1988): Int64Index([254], dtype='int64'),
 ('Madagascar', 1995): Int64Index([261], dtype='int64'),
 ('Malawi', 1974): Int64Index([294], dtype='int64'),
 ('Malawi', 1975): Int64Index([295], dtype='int64'),
 ('Malawi', 1976): Int64Index([296], dtype='int64'),
 ('Mauritius', 1970): Int64Index([263], dtype='int64'),
 ('Mauritius', 1978): Int64Index([271], dtype='int64'),
 ('Mauritius', 1989): Int64Index([282], dtype='int64'),
 ('Mauritius', 1992): Int64Index([285], dtype='int64'),
 ('Mauritius', 1993): Int64Index([286], dtype='int64')}
```

Calculate the aggregate of a group

In [30]:
```
df1.groupby(['countryc','year']).agg(np.mean)
```
Out[30]:

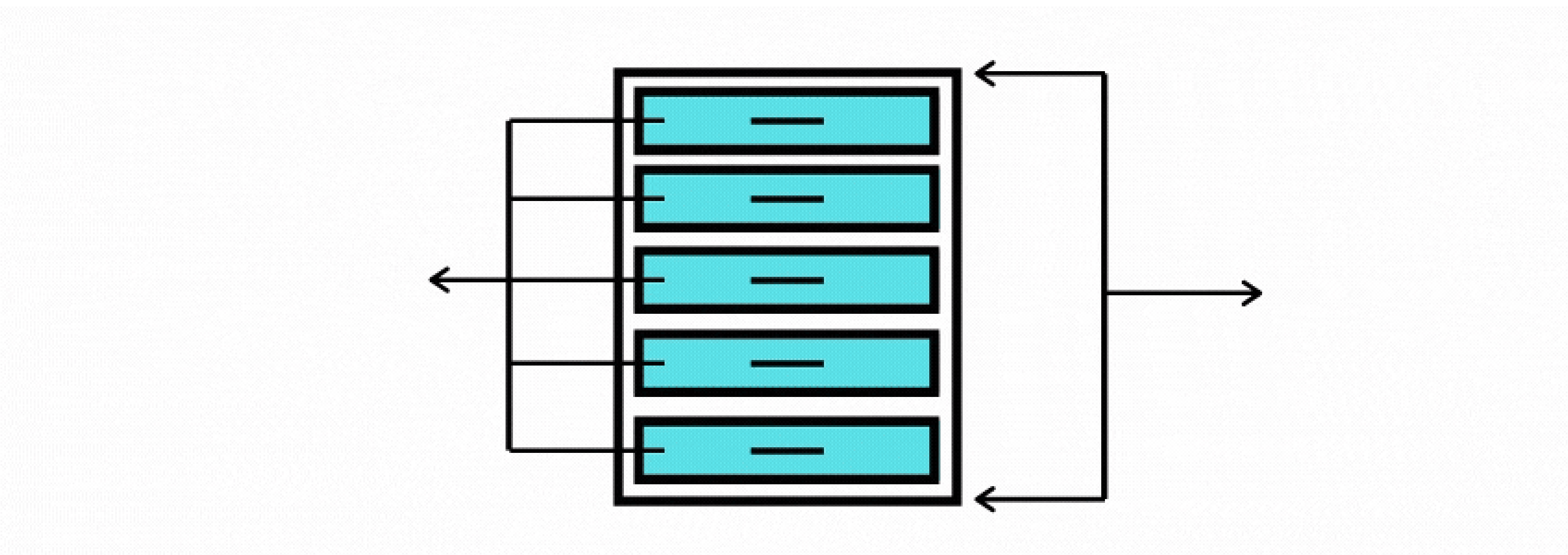
		popn
countryc	**year**	
Botswana	**1976**	782650.0
	1987	1154280.0
	1988	1195140.0
	1994	1420270.0
Burkina Faso	**1973**	5958700.0
...	**...**	...
Mauritius	**1970**	829000.0
	1978	930800.0
	1989	1048560.0
	1992	1081000.0
	1993	1097000.0

75 rows × 1 columns

Group the values in df4 by their 'GDP per Capita in 1995'

In [31]:
```
df4.groupby(['gdnpp']).groups
```
```
 '348.727386474609': Int64Index([13], dtype='int64'),
 '348.815185546875': Int64Index([232], dtype='int64'),
 '349.856109619141': Int64Index([251], dtype='int64'),
 '349.927886962891': Int64Index([218], dtype='int64'),
 '370.860687255859': Int64Index([132], dtype='int64'),
 '372.9169921875': Int64Index([145], dtype='int64'),
 '379.727691650391': Int64Index([134], dtype='int64'),
 '380.071899414062': Int64Index([177], dtype='int64'),
 '395.732604980469': Int64Index([130], dtype='int64'),
 '424.085784912109': Int64Index([135], dtype='int64'),
 '442.639709472656': Int64Index([28], dtype='int64'),
 '453.507690429688': Int64Index([124], dtype='int64'),
 '463.250885009766': Int64Index([143], dtype='int64'),
 '463.472290039062': Int64Index([128], dtype='int64'),
 '473.528686523438': Int64Index([123], dtype='int64'),
 '479.167999267578': Int64Index([170], dtype='int64'),
 '481.151611328125': Int64Index([121], dtype='int64'),
 '486.431396484375': Int64Index([249], dtype='int64'),
 '497.304412841797': Int64Index([127], dtype='int64'),
 '504.190612792969': Int64Index([119], dtype='int64'),
```

Binning Data



In [32]:
```
df.shape
```
Out[32]: (301, 50)

In [33]:
```
pd.qcut(df['popn'],q=7).value_counts()
```

Out[33]:
```
(16616600.0, 56404000.0]       43
(2652272.857, 7886652.857]     43
(463999.999, 953097.143]       43
(10664697.143, 16616600.0]     42
(7886652.857, 10664697.143]    42
(1377285.714, 2652272.857]     42
(953097.143, 1377285.714]      42
Name: popn, dtype: int64
```

Bin the data in df2 according to the year

In [34]:
```
df2['year'].shape
```

Out[34]:
```
(56,)
```

In [35]:
```
pd.qcut(df2['year'],q=7).value_counts()
```

Out[35]:
```
(1970.999, 1974.0]     11
(1988.286, 1990.0]      9
(1984.429, 1988.286]    8
(1978.0, 1981.0]        8
(1990.0, 1995.0]        7
(1981.0, 1984.429]      7
(1974.0, 1978.0]        6
Name: year, dtype: int64
```