Prepared by Asif Bhat

Merging Dataframes

In [1]:
```
import numpy as np
import pandas as pd
from IPython.core.interactiveshell import InteractiveShell
InteractiveShell.ast_node_interactivity = "all"
```

Merging Dataframes

In [2]:
```
emp1= pd.DataFrame({'empid':['E90','E87'],
                    'Name':['Asif','Basit']})
emp1


emp2= pd.DataFrame({'empid':['E22','E74','E90'],
                    'Name':['Minil','Akash','Asif']})
emp2


address= pd.DataFrame({'empid':['E87','E22','E49'],
                    'City':['Mumbai','Banglore','Pune'] ,
                    'State':['Maharashtra','Karnataka','Maharashtra']})
address
```

Out[2]:

	empid	Name
0	E90	Asif
1	E87	Basit

Out[2]:

	empid	Name
0	E22	Minil
1	E74	Akash
2	E90	Asif

Out[2]:

	empid	City	State
0	E87	Mumbai	Maharashtra
1	E22	Banglore	Karnataka
2	E49	Pune	Maharashtra

In [3]:
```
#Concat Dataframes
# Method-1
employees = pd.concat([emp1,emp2])
employees
employees=employees.reset_index(drop=True)
employees
```

Out[3]:

	empid	Name
0	E90	Asif
1	E87	Basit
0	E22	Minil
1	E74	Akash
2	E90	Asif

Out[3]:

	empid	Name
0	E90	Asif
1	E87	Basit
2	E22	Minil
3	E74	Akash
4	E90	Asif

In [4]:
```
#Concat Dataframe
# Method-2
employees1 = emp1.append(emp2)
employees1
employees1=employees1.reset_index(drop=True)
employees1
```

Out[4]:

	empid	Name
0	E90	Asif
1	E87	Basit
0	E22	Minil
1	E74	Akash
2	E90	Asif

Out[4]:

	empid	Name
0	E90	Asif
1	E87	Basit
2	E22	Minil
3	E74	Akash
4	E90	Asif

In [40]:
```
employees2= pd.concat([emp1,emp2],ignore_index=True)
employees2
```

Out[40]:

	empid	Name
0	E90	Asif
1	E87	Basit
2	E22	Minil
3	E74	Akash
4	E90	Asif

In [41]:
```
# Check for duplicates after concatenation
duplicates = employees[employees.duplicated()]
duplicates
```

Out[41]:

	empid	Name
4	E90	Asif

In [42]:
```
# Remove duplicates
employees.drop_duplicates(keep='last',inplace=True)
employees
```

Out[42]:

	empid	Name
1	E87	Basit
2	E22	Minil
3	E74	Akash
4	E90	Asif

In [43]:
```
employees.reset_index(drop=True,inplace=True)
employees
```

Out[43]:

	empid	Name
0	E87	Basit
1	E22	Minil
2	E74	Akash
3	E90	Asif

In [44]:
```
employees
address
```

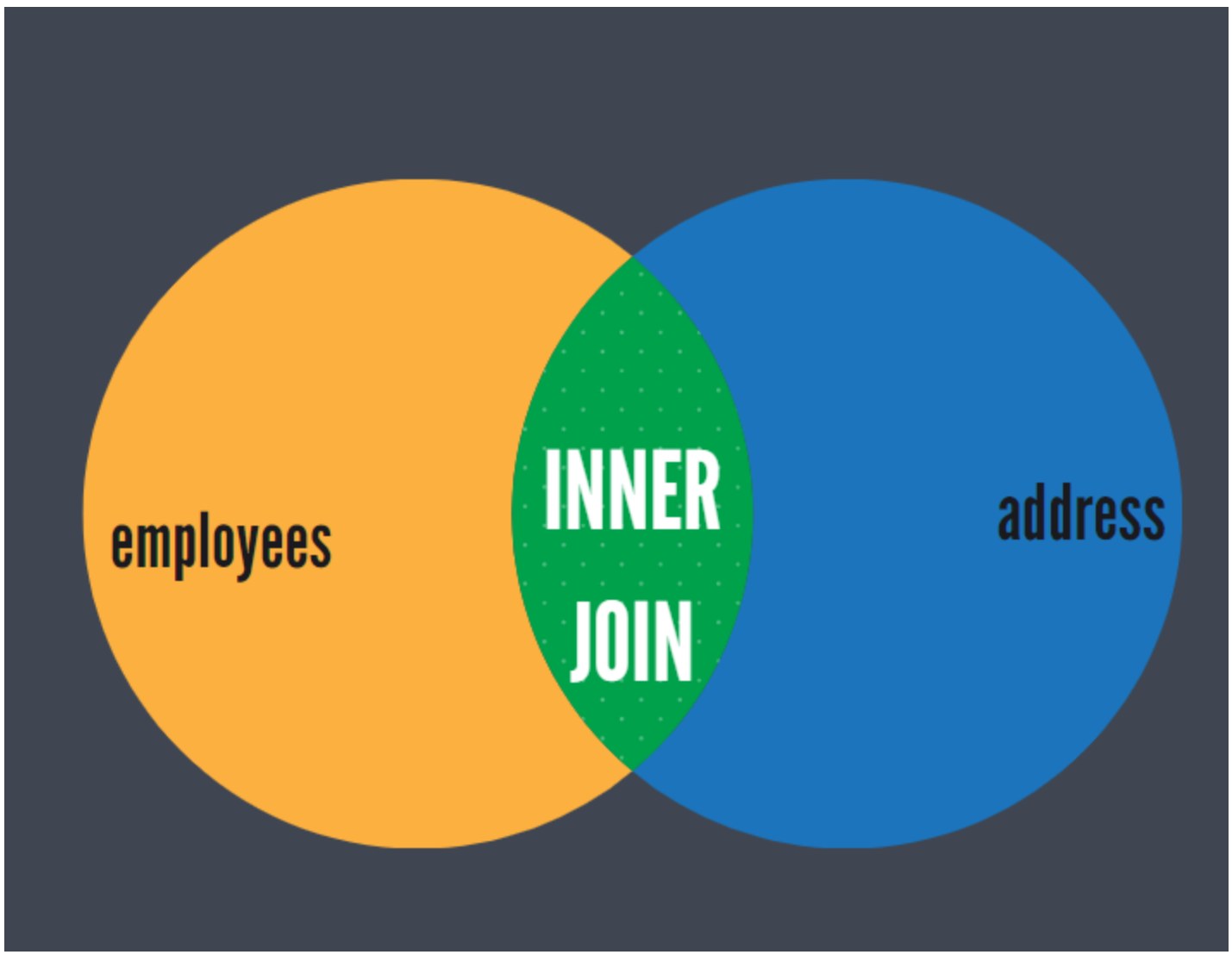
Out[44]:

	empid	Name
0	E87	Basit
1	E22	Minil
2	E74	Akash
3	E90	Asif

Out[44]:

	empid	City	State
0	E87	Mumbai	Maharashtra
1	E22	Banglore	Karnataka
2	E49	Pune	Maharashtra

Inner Join



In [20]:
```
# Inner Join
inner = pd.merge(employees,address,on='empid')
inner
```

Out[20]:

	empid	Name	City	State
0	E87	Basit	Mumbai	Maharashtra
1	E22	Minil	Banglore	Karnataka

In [21]:
```
# Inner Join
inner = pd.merge(employees,address,on='empid',how='inner')
inner
```

Out[21]:

	empid	Name	City	State
0	E87	Basit	Mumbai	Maharashtra
1	E22	Minil	Banglore	Karnataka

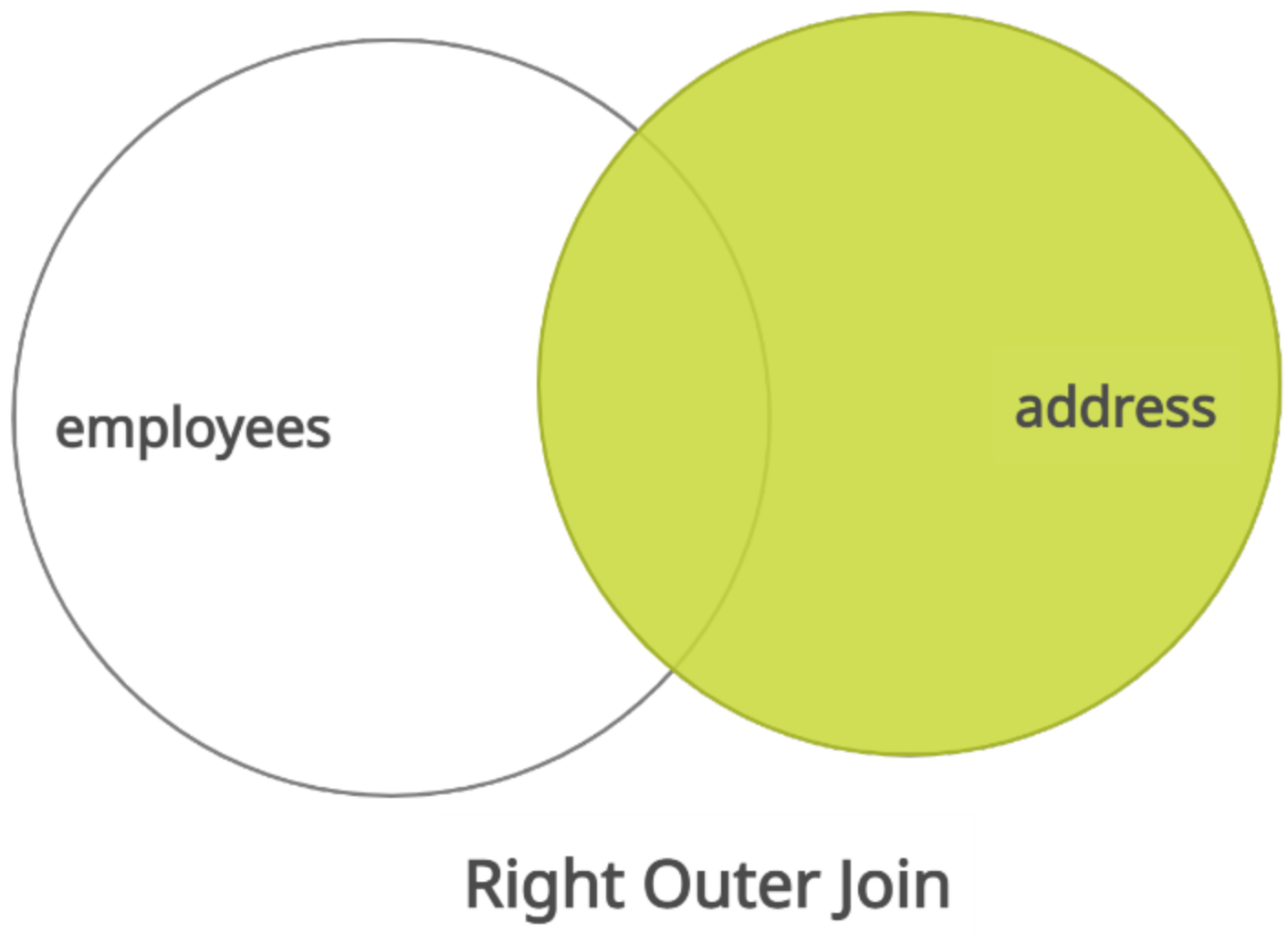
Left Outer Join

In [22]:
```
# Left Outer Join
left = pd.merge(employees,address,on='empid',how='left')
left
```

Out[22]:

	empid	Name	City	State
0	E87	Basit	Mumbai	Maharashtra
1	E22	Minil	Banglore	Karnataka
2	E74	Akash	NaN	NaN
3	E90	Asif	NaN	NaN

Right Outer Join

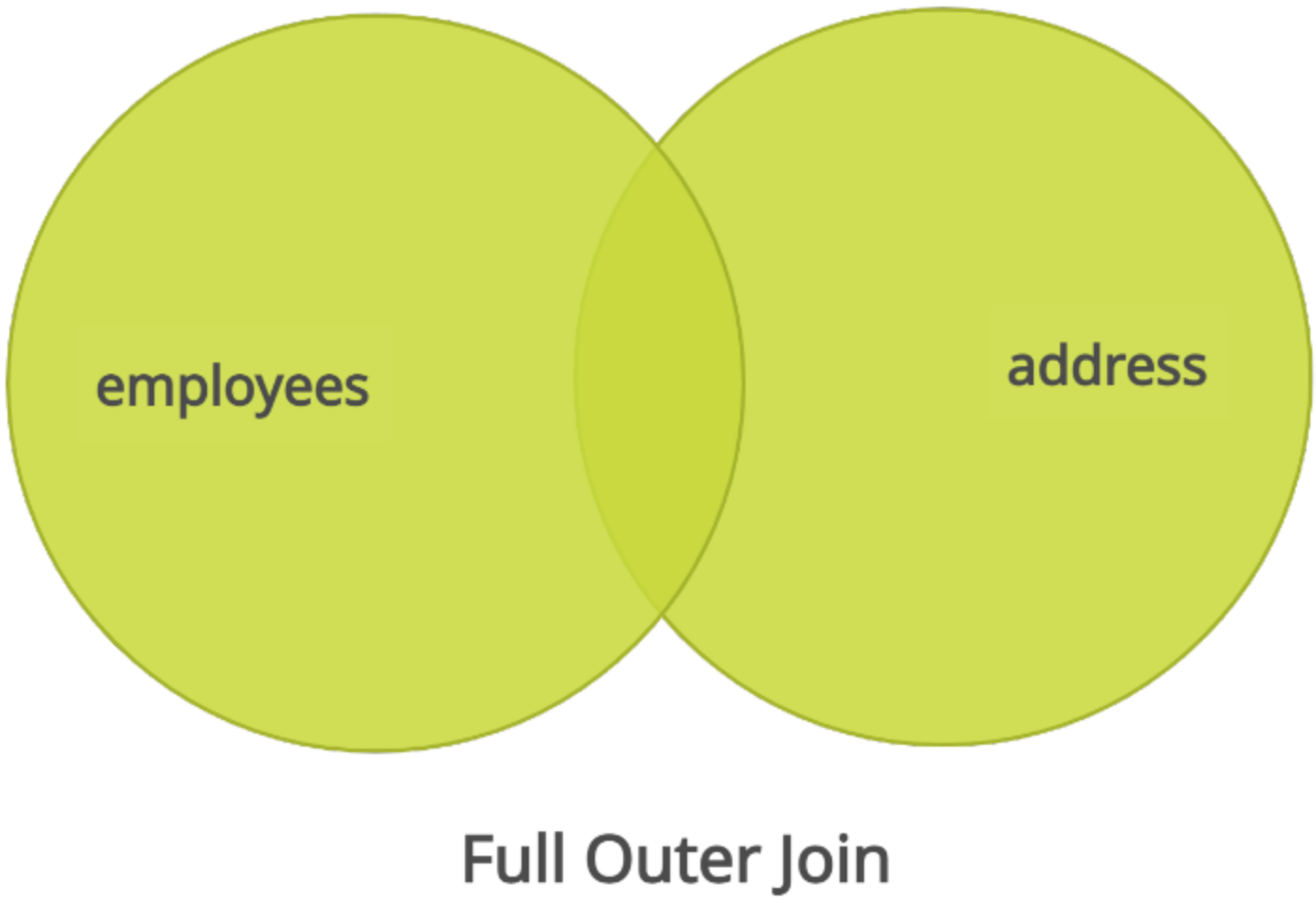


In [48]:
```
# Right Outer Join
right = pd.merge(employees,address,on='empid',how='right')
right
```

Out[48]:

	empid	Name	City	State
0	E87	Basit	Mumbai	Maharashtra
1	E22	Minil	Banglore	Karnataka
2	E49	NaN	Pune	Maharashtra

Full Outer Join



In [24]:
```
employees
address
```

Out[24]:

	empid	Name
0	E87	Basit
1	E22	Minil
2	E74	Akash
3	E90	Asif

Out[24]:

	empid	City	State
0	E87	Mumbai	Maharashtra
1	E22	Banglore	Karnataka
2	E49	Pune	Maharashtra

In [25]:
```
#Full outer join
full = pd.merge(employees,address,how='outer',on='empid')
full
```

Out[25]:

	empid	Name	City	State
0	E87	Basit	Mumbai	Maharashtra
1	E22	Minil	Banglore	Karnataka
2	E74	Akash	NaN	NaN
3	E90	Asif	NaN	NaN
4	E49	NaN	Pune	Maharashtra

In [26]:
```
# Salary Data
sal = pd.DataFrame({'empid':['E87','E22','E74','E90','E49'],
                    'salary':['$10,000','$30,000','$20,000','$60,000','$90,000']})
sal
```

Out[26]:

	empid	salary
0	E87	$10,000
1	E22	$30,000
2	E74	$20,000
3	E90	$60,000
4	E49	$90,000

In [27]:
```
# Add Salary details ( Merge sal & full dataframes)
employee_details = pd.merge(full,sal,how='inner',on='empid')
employee_details
```

Out[27]:

	empid	Name	City	State	salary
0	E87	Basit	Mumbai	Maharashtra	$10,000
1	E22	Minil	Banglore	Karnataka	$30,000
2	E74	Akash	NaN	NaN	$20,000
3	E90	Asif	NaN	NaN	$60,000
4	E49	NaN	Pune	Maharashtra	$90,000

In [28]:
```
# Employee experience
exp = pd.DataFrame({'employee_id':['E87','E22','E74','E90','E49'],
                    'experience':['5 years','3 years','7 years','2 years','10 years']})
exp
```

Out[28]:

	employee_id	experience
0	E87	5 years
1	E22	3 years
2	E74	7 years
3	E90	2 years
4	E49	10 years

In [29]:
```
# Add employee experinece using merge()
# As column names are different (empid , employee_id) we have to use left_on & right_on parameter
pd.merge(employee_details,exp,left_on=['empid'],right_on=['employee_id'],how='inner')
```

Out[29]:

	empid	Name	City	State	salary	employee_id	experience
0	E87	Basit	Mumbai	Maharashtra	$10,000	E87	5 years
1	E22	Minil	Banglore	Karnataka	$30,000	E22	3 years
2	E74	Akash	NaN	NaN	$20,000	E74	7 years
3	E90	Asif	NaN	NaN	$60,000	E90	2 years
4	E49	NaN	Pune	Maharashtra	$90,000	E49	10 years

In [30]:
```
# Add employee experinece using merge() and drop the duplicate column (employee_id)
pd.merge(employee_details,exp,left_on=['empid'],right_on=['employee_id'],how='inner').drop('employee_id',axis=1)
```

Out[30]:

	empid	Name	City	State	salary	experience
0	E87	Basit	Mumbai	Maharashtra	$10,000	5 years
1	E22	Minil	Banglore	Karnataka	$30,000	3 years
2	E74	Akash	NaN	NaN	$20,000	7 years
3	E90	Asif	NaN	NaN	$60,000	2 years
4	E49	NaN	Pune	Maharashtra	$90,000	10 years

In [31]:
```
#Save the dataframe
employee_details=pd.merge(employee_details,exp,
                          left_on=['empid'],
                          right_on=['employee_id'],
                          how='inner').drop('employee_id',axis=1)
employee_details
```

Out[31]:

	empid	Name	City	State	salary	experience
0	E87	Basit	Mumbai	Maharashtra	$10,000	5 years
1	E22	Minil	Banglore	Karnataka	$30,000	3 years
2	E74	Akash	NaN	NaN	$20,000	7 years
3	E90	Asif	NaN	NaN	$60,000	2 years
4	E49	NaN	Pune	Maharashtra	$90,000	10 years

End