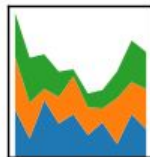
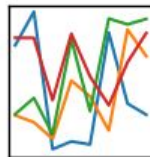


# Merging DataFrames

pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



# Introduction

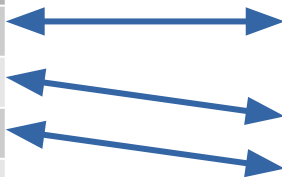
A join is used to combine data from two tables.

mlb

City	MLB_Team
Atlanta	Braves
Detroit	Tigers
Houston	Astros
Milwaukee	Brewers
Toronto	Blue Jays

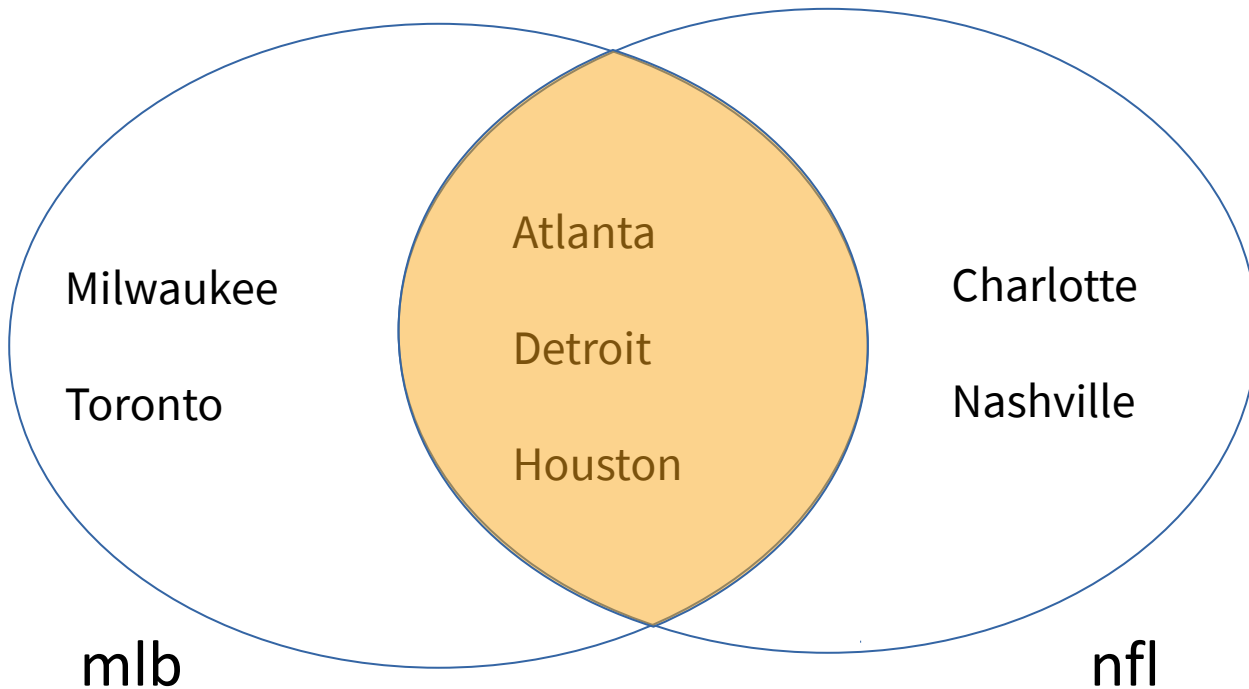
nfl

City	NFL_Team
Atlanta	Falcons
Charlotte	Panthers
Detroit	Lions
Houston	Texans
Nashville	Titans



# Inner Joins

Selects all indices appearing in table 1 AND table 2.



# Inner Join

City	MLB_Team		City	NFL_Team
Atlanta	Braves	↔	Atlanta	Falcons
Detroit	Tigers	↗	Charlotte	Panthers
Houston	Astros	↘	Detroit	Lions
Milwaukee	Brewers		Houston	Texans
Toronto	Blue Jays		Nashville	Titans

`pd.merge(mlb, nfl)` OR `pd.merge(mlb, nfl, how = 'inner')`  
OR `pd.merge(mlb, nfl, on = 'City')`

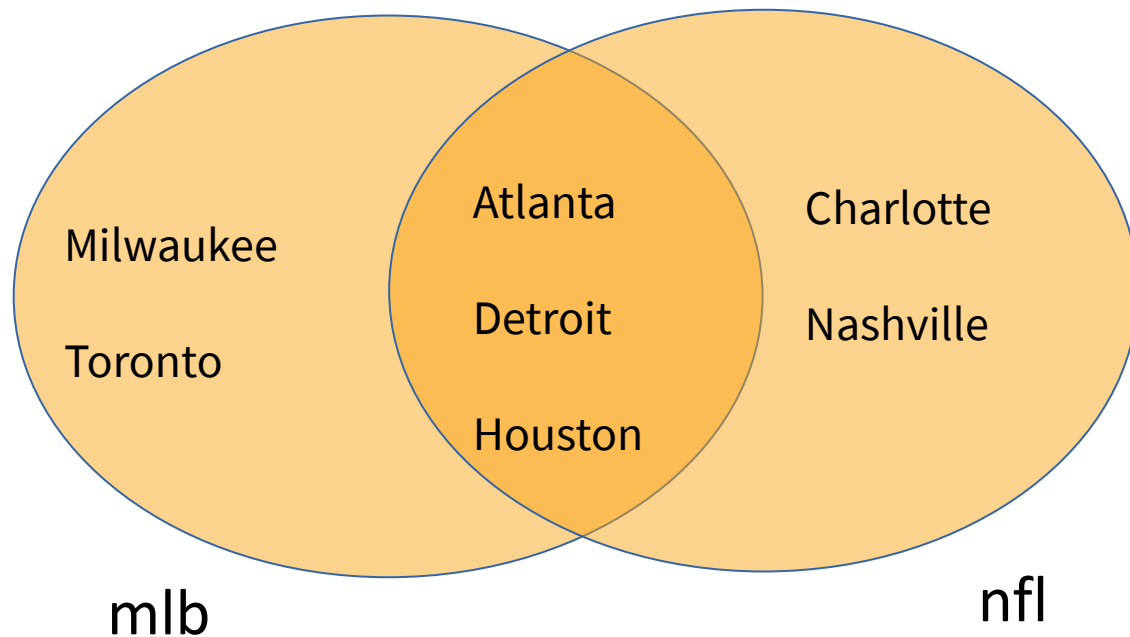
City	MLB_Team	NFL_Team
Atlanta	Braves	Falcons
Detroit	Tigers	Lions
Houston	Astros	Texans

# Joins in Pandas

- By default, *pd.merge()* will join on all columns that appear in both tables (with the exact same name)
- The exact column(s) to join on can be specified by passing a list to the 'on =' argument.
- If the column to join on has a different name in the different tables, it can be specified with a 'left\_on =' and a 'right\_on =' argument.

# Outer Join

Selects all indices appearing in table 1 OR table 2.  
Rows without a match are filled with NaN.



# Outer Join

City	MLB_Team
Atlanta	Braves
Detroit	Tigers
Houston	Astros
Milwaukee	Brewers
Toronto	Blue Jays

City	NFL_Team
Atlanta	Falcons
Charlotte	Panthers
Detroit	Lions
Houston	Texans
Nashville	Titans

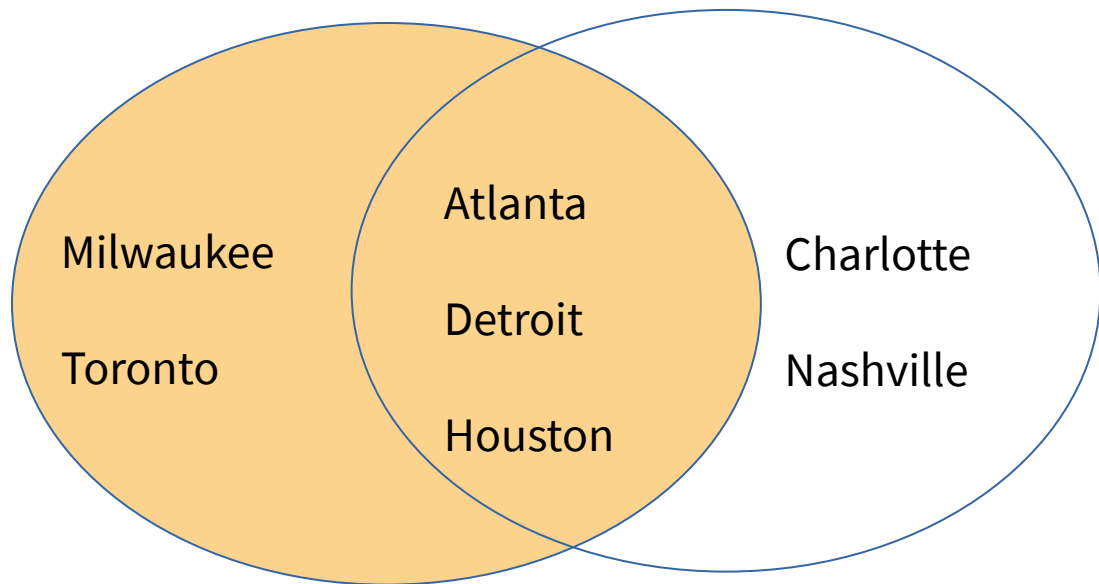
`pd.merge(mlb, nfl, how = 'outer')`

City	MLB_Team	NFL_Team
Atlanta	Braves	Falcons
Detroit	Tigers	Lions
Houston	Astros	Texans
Milwaukee	Brewers	<b>NaN</b>
Toronto	Blue Jays	<b>NaN</b>
Charlotte	<b>NaN</b>	Panthers
Nashville	<b>NaN</b>	Titans

# Left (Outer) Join

Selects all indices in the left table and any matching indices from the right table.

Rows from the left table without a match are filled with NaN.





# Left (Outer) Join

City	MLB_Team
Atlanta	Braves
Detroit	Tigers
Houston	Astros
Milwaukee	Brewers
Toronto	Blue Jays

City	NFL_Team
Atlanta	Falcons
Charlotte	Panthers
Detroit	Lions
Houston	Texans
Nashville	Titans

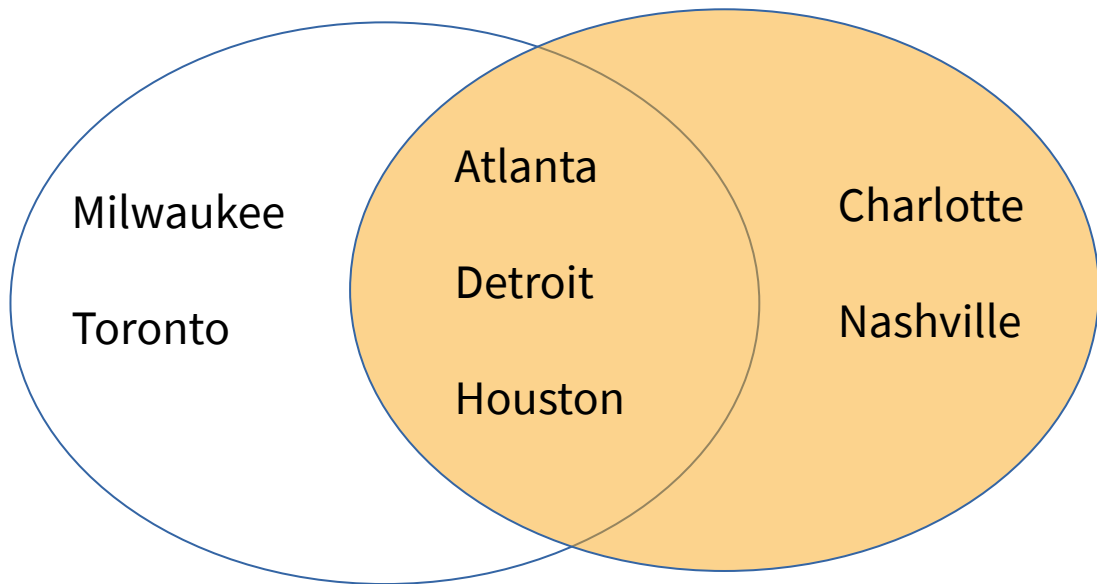
`pd.merge(mlb, nfl, how = 'left')`

City	MLB_Team	NFL_Team
Atlanta	Braves	Falcons
Detroit	Tigers	Lions
Houston	Astros	Texans
Milwaukee	Brewers	NaN
Toronto	Blue Jays	NaN

# Right (Outer) Join

Selects all indices in the right table and any matching indices from the left table.

Rows from the right table without a match are filled with NaN.



# Right (Outer) Join

City	MLB_Team
Atlanta	Braves
Detroit	Tigers
Houston	Astros
Milwaukee	Brewers
Toronto	Blue Jays

City	NFL_Team
Atlanta	Falcons
Charlotte	Panthers
Detroit	Lions
Houston	Texans
Nashville	Titans

```
pd.merge(mlb, nfl, how = 'right')
```

City	MLB_Team	NFL_Team
Atlanta	Braves	Falcons
Detroit	Tigers	Lions
Houston	Astros	Texans
Charlotte	<b>NaN</b>	Panthers
Nashville	<b>NaN</b>	Titans