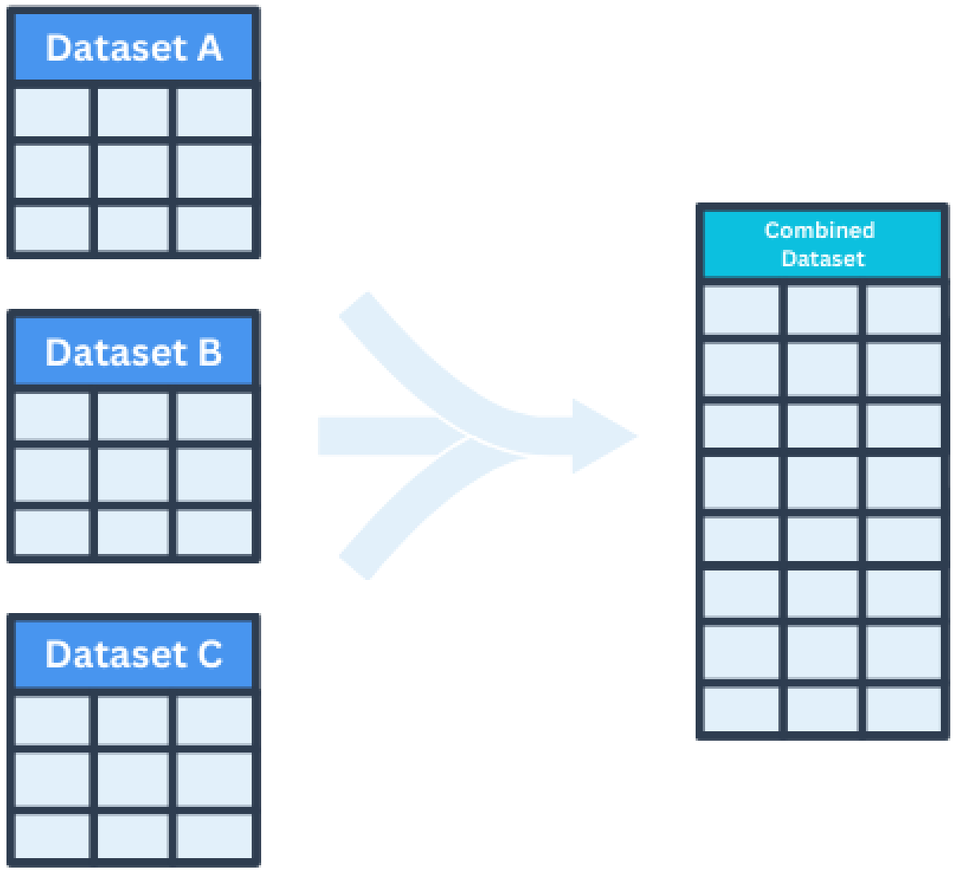
Appending and merging queries

INTERMEDIATE POWER QUERY IN EXCEL

Lyndsay Girard
Performance Analytics Consultant

Combining datasets in Power Query

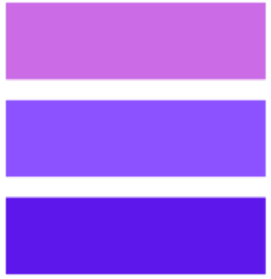


- Consolidates data from multiple sources
- Saves time by reducing manual manipulation (e.g., VLOOKUP)
- Enriches larger datasets by including additional columns from related tables

Principles of append vs. merge

Append

- Stacking datasets of a shared column structure vertically.



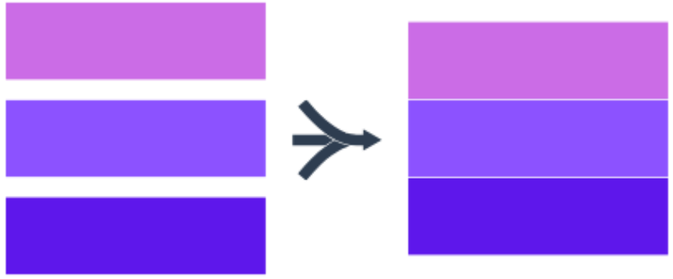
Merge

- Combining datasets horizontally based on common column(s).

Principles of append vs. merge

Append

- Stacking datasets of a shared column structure vertically.



Merge

- Combining datasets horizontally based on common column(s).

Append in practice

- Ensure that the columns in the tables being combined have the same:
  - Name
  - Data Type

Dataset A

Hospital	PatientID	A	B	C	F
A	1	0	0	10	0
A	2	10	1	10	0
A	3	10	0	0	1

Dataset B

Hospital	PatientID	A	B	C
B	4	10	1	10
B	5	0	1	1
B	6	1	10	1

Append in practice

- Ensure that the columns in the tables being combined have the same:
  - Name
  - Data Type

Append Output

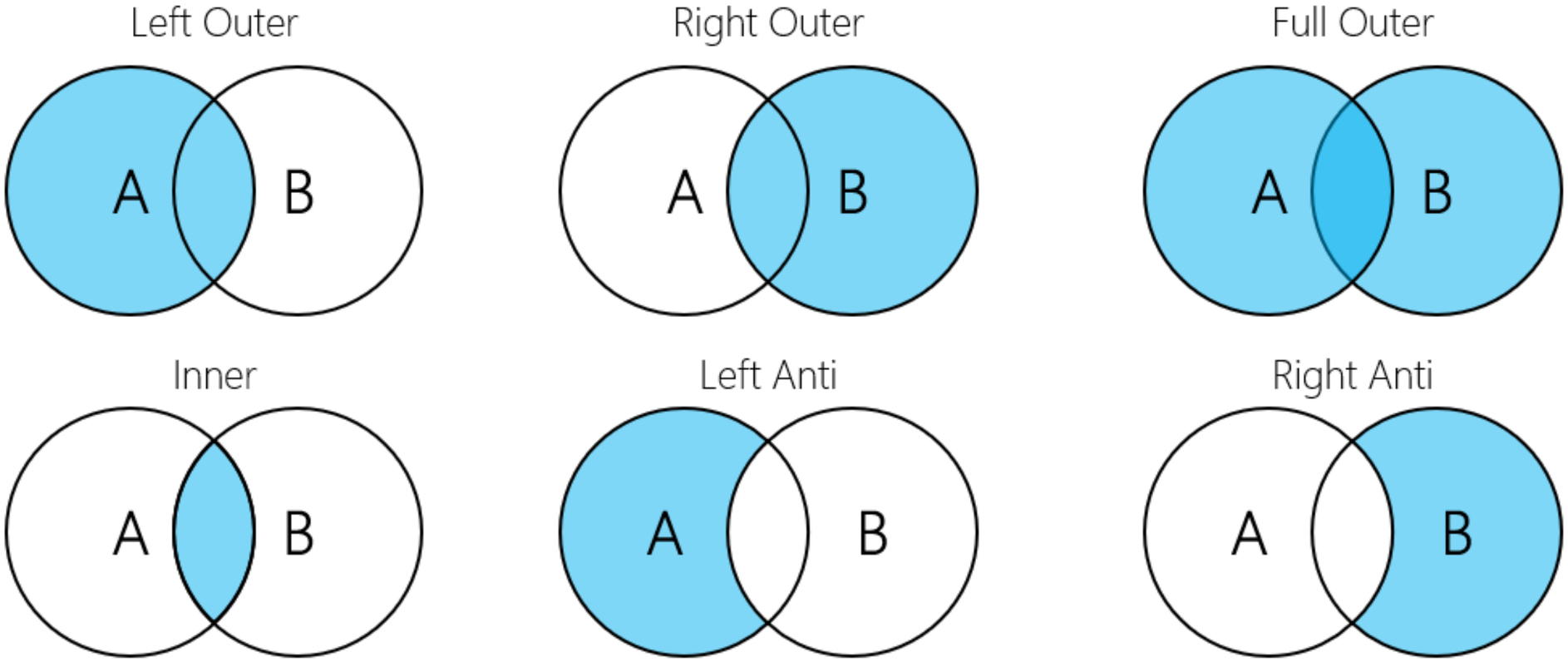
Hospital	PatientID	A	B	C	F
A	1	0	0	10	0
A	2	10	1	10	0
A	3	10	0	0	1
B	4	10	1	10	NULL
B	5	0	1	1	NULL
B	6	1	10	1	NULL

Diving deeper into the merge operation

- Merged data is relative to the current selected query
- Unmatched records may be included or excluded from output

Join types

- Type of join determines how records are combined or excluded based on matching criteria



Merge in practice: left outer join

- Ensure the shared column (key) to match on is of the same data type
- Preview the results to ensure that the merging logic is accurate

Dataset A

Hospital	PatientID	A	B	C	F
A	1	0	0	10	0
A	2	10	1	10	0
A	3	10	0	0	1

Dataset B

Hospital	PatientID	X	Y	Z
A	1	1	1	1
A	2	0	1	1
B	4	0	0	1
B	5	1	0	1

Merge in practice: left outer join

- Ensure the shared column (key) to match on is of the same data type
- Preview the results to ensure that the merging logic is accurate

Left Outer Join Output

Hospital	PatientID	A	B	C	F	X	Y	Z
A	1	0	0	10	0	1	1	1
A	2	10	1	10	0	0	1	1
A	3	10	0	0	1	*NULL*	*NULL*	*NULL*

Merge in practice: full outer join

- Ensure the shared column (key) to match on is of the same data type
- Preview the results to ensure that the merging logic is accurate

Full Outer Join Output

Hospital	PatientID	A	B	C	F	X	Y	Z
A	1	0	0	10	0	1	1	1
A	2	10	1	10	0	0	1	1
A	3	10	0	0	1	*NULL*	*NULL*	*NULL*
B	4	*NULL*	*NULL*	*NULL*	*NULL*	0	0	1
B	5	*NULL*	*NULL*	*NULL*	*NULL*	1	0	1

Special join conditions

- Exact Match: Identical matches between data elements
- Fuzzy Match: Allows for flexible and approximate matching

Dataset A

Department	A
Medicine	100
Surgery	50
Obstetrics	10

Dataset B

Department	B
Medicine	101
Med	95
medicine	100
OBS	8
Surgery	48
Med	99

Special join conditions

- Exact Match: Identical matches between data elements
- Fuzzy Match: Allows for flexible and approximate matching
  - Ignore case
  - Match by combining text parts

Dataset A

Department	A
Medicine	100
Surgery	50
Obstetrics	10

Dataset B

Department	B
Medicine	101
Med	95
medicine	100
OBS	8
Surgery	48
Med	99

Dataset

- Real-world hospital dataset (de-identified electronic health record data)
- Emergency department visits to the Yale New Haven Health System from March 2014 to July 2017
- Patient demographics and registration information, triage assessments, medications, and other related data

[1] journals.plos.org/plosone/article?id=10.1371/journal.pone.0201016

Let's practice!

INTERMEDIATE POWER QUERY IN EXCEL

Combining queries with Power Query

INTERMEDIATE POWER QUERY IN EXCEL

Lyndsay Girard
Performance Analytics Consultant

datacamp

Let's practice!

INTERMEDIATE POWER QUERY IN EXCEL