Combining data tables

CONNECTING DATA IN TABLEAU

Lis Sulmont

Head of Curriculum Expansion, DataCamp

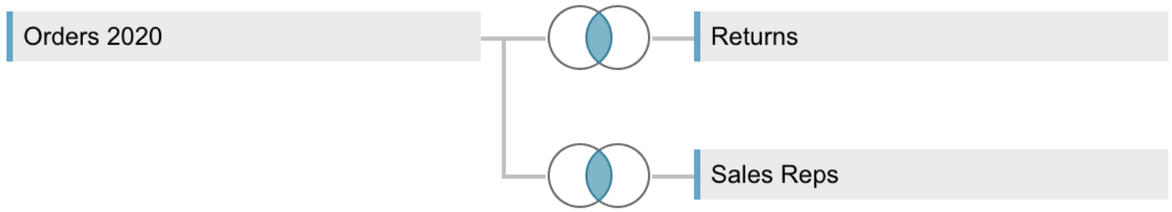
datacamp

Combining data

Superstore

Filters
0 | Add

Orders 2020 is made of 3 tables.

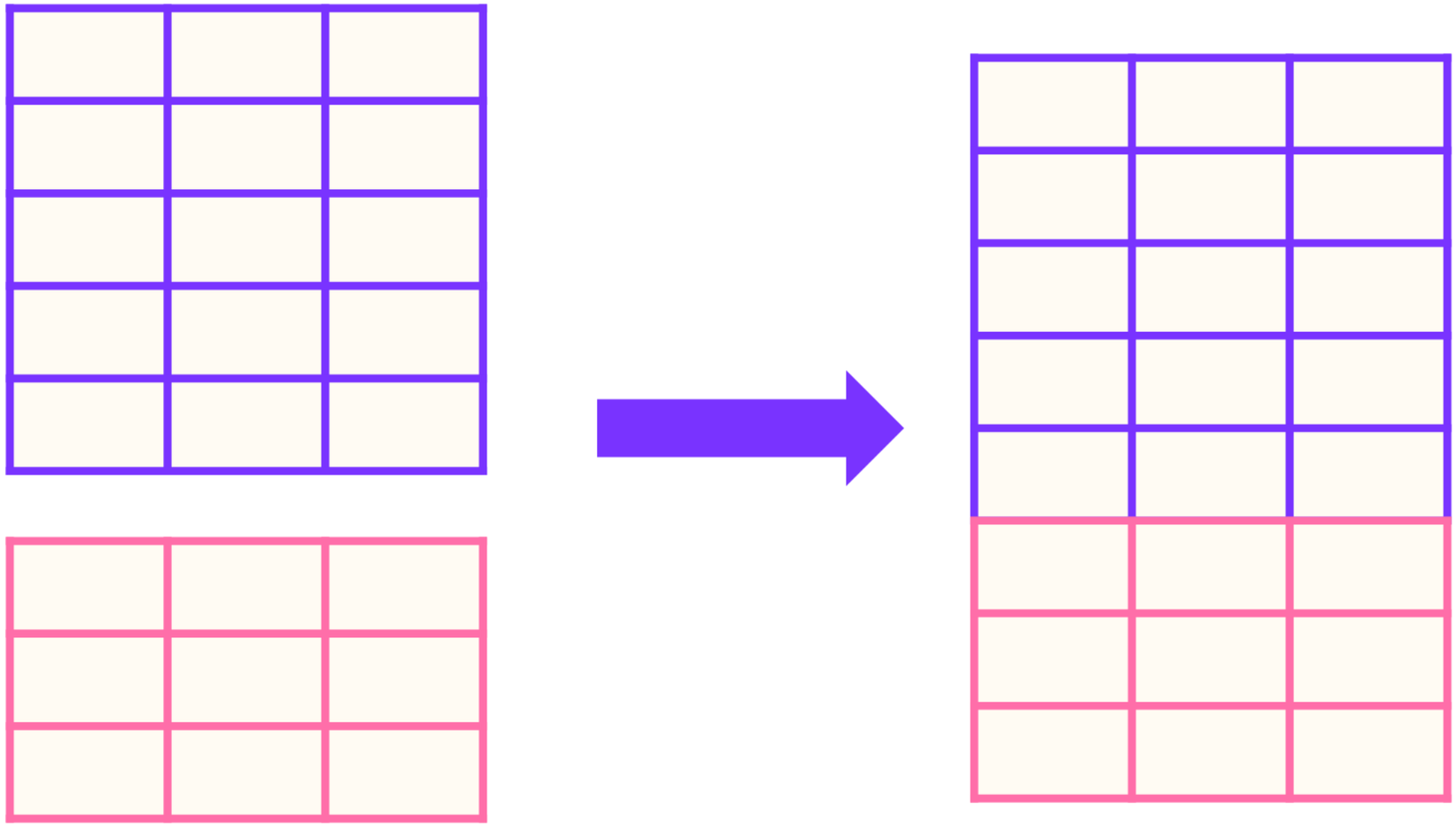


Sort fields: Data source order

Show aliases | Show hidden fields | 289 row

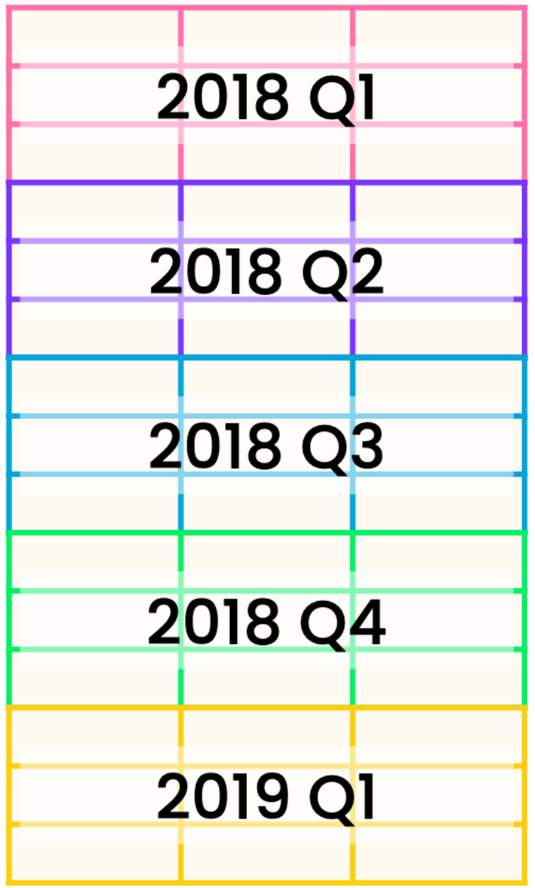
Orders 2020 Row ID	Orders 2020 Order ID	Orders 2020 Order Date	Orders 2020 Ship Date	Orders 2020 Ship Mode	Orders 2020 Customer ID	Orders 2020 Customer Name	Orders 2020 Segment	Orders 2020 Country/Region	Orders 2020 City
4910	CA-2020-127306	1/14/2020	1/18/2020	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Johnson
4911	CA-2020-127306	1/14/2020	1/18/2020	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Johnson
4912	CA-2020-127306	1/14/2020	1/18/2020	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Johnson

Unions

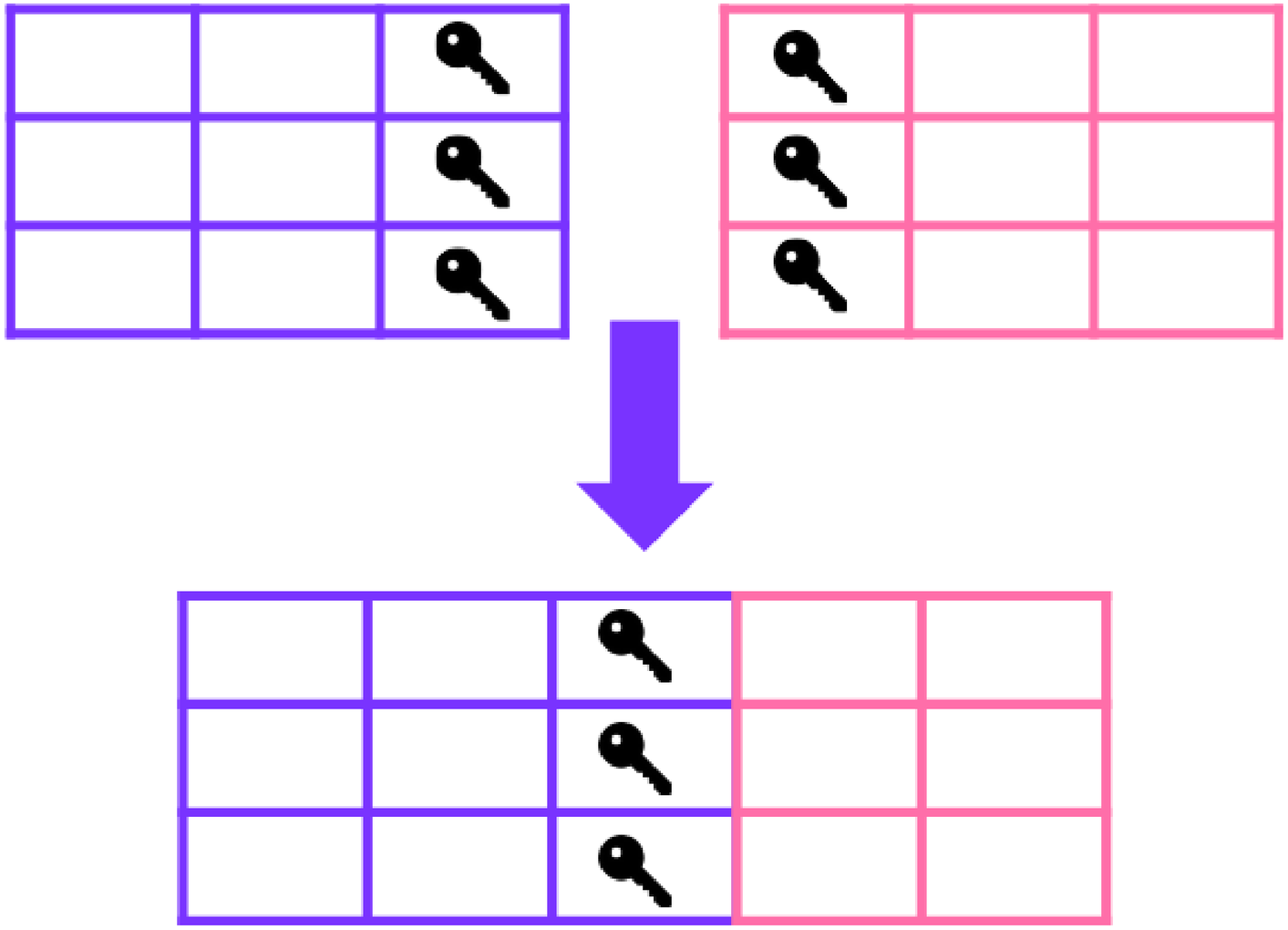


Union example

Divvy_Trips_2018_Q1.zip	Jan 24th 2020, 10:07:56 am
Divvy_Trips_2018_Q2.zip	Jan 24th 2020, 10:07:57 am
Divvy_Trips_2018_Q3.zip	Jan 24th 2020, 10:08:00 am
Divvy_Trips_2018_Q4.zip	Jan 24th 2020, 10:08:05 am
Divvy_Trips_2019_Q1.zip	Jan 24th 2020, 10:08:06 am
Divvy_Trips_2019_Q2.zip	Jan 24th 2020, 10:08:05 am
Divvy_Trips_2019_Q3.zip	Jan 24th 2020, 10:08:06 am
Divvy_Trips_2019_Q4.zip	Jan 24th 2020, 10:08:07 am
Divvy_Trips_2020_Q1.zip	May 26th 2020, 07:17:43 pm



Joins



Joins example

Table 1: Employees

Employee Name	Dept	Employee Email	Location ID
Tom	HR	tom@company.com	1
Sarah	Marketing	sarah@company.com	2
Jane	Sales	jane@company.com	2
Marty	Sales	marty@company.com	1
Bob	Product	bob@company.com	remote

Table 2: Offices

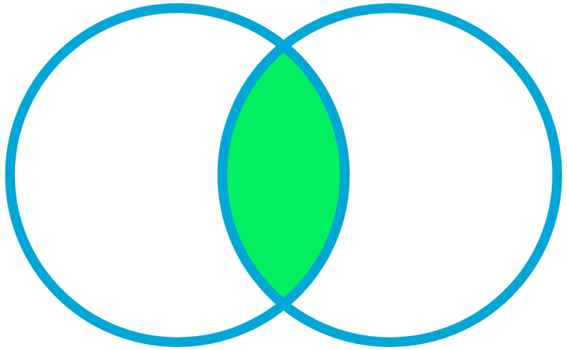
Location ID	Address
1	20 W 34th St, New York, NY 10001, USA
2	Martelarenlaan 38 3010 Kessel-Lo, Belgium
3	207 Old Street, London EC1V 9NRk UK

Joins example

Result after join on Location ID

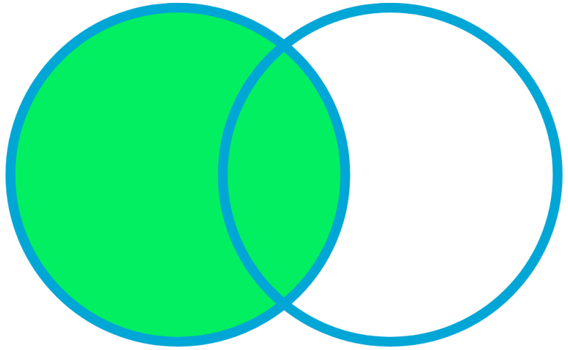
Employee Name	Dept	Employee Email	Location ID	Address
Tom	HR	tom@company.com	1	20 W 34th St, New York, NY 10001, USA
Sarah	Marketing	sarah@company.com	2	Martelarenlaan 38 3010 Kessel-Lo, Belgium
Jane	Sales	jane@company.com	2	Martelarenlaan 38 3010 Kessel-Lo, Belgium
Marty	Sales	marty@company.com	1	20 W 34th St, New York, NY 10001, USA
Bob	Product	bob@company.com	remote	*NULL*

Inner Join



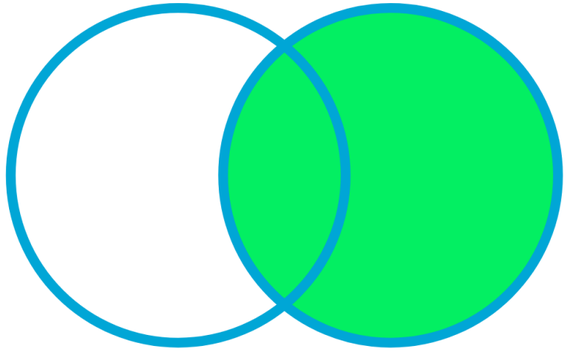
Returns matched rows only

Left Join



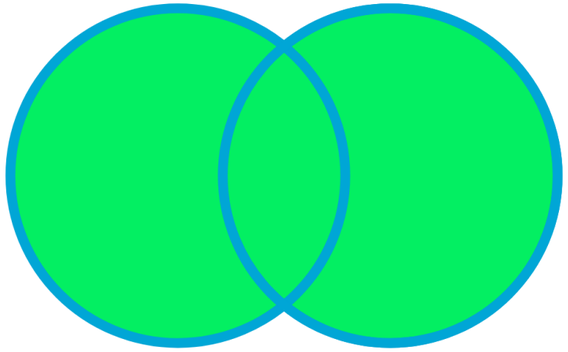
Returns matched rows and left table's rows

Right Join



Returns matched rows and right table's rows

Full Join



Returns matched rows and both tables' rows

Table 1: Employees

Employee Name	Dept	Employee Email	Location ID
Tom	HR	tom@company.com	1
Sarah	Marketing	sarah@company.com	2
Jane	Sales	jane@company.com	2
Marty	Sales	marty@company.com	1
Bob	Product	bob@company.com	remote

Table 2: Offices

Location ID	Address
1	20 W 34th St, New York, NY 10001, USA
2	Martelarenlaan 38 3010 Kessel-Lo, Belgium
3	207 Old Street, London EC1V 9NRk UK

Result after join on Location ID

Employee Name	Dept	Employee Email	Location ID	Address
Tom	HR	tom@company.com	1	20 W 34th St, New York, NY 10001, USA
Sarah	Marketing	sarah@company.com	2	Martelarenlaan 38 3010 Kessel-Lo, Belgium
Jane	Sales	jane@company.com	2	Martelarenlaan 38 3010 Kessel-Lo, Belgium
Marty	Sales	marty@company.com	1	20 W 34th St, New York, NY 10001, USA
Bob	Product	bob@company.com	remote	*NULL*

Superstore dataset

- `Orders 2016`
- `Orders 2017`
- `Orders 2018`
- `Orders 2019`
- `Orders 2020`
- `Returns`
- `Sales Reps`

Let's practice!

CONNECTING DATA IN TABLEAU

Unioning tables

CONNECTING DATA IN TABLEAU

Lis Sulmont

Head of Curriculum Expansion, DataCamp

datacamp

Let's practice!

CONNECTING DATA IN TABLEAU

Joining tables

CONNECTING DATA IN TABLEAU

Lis Sulmont

Head of Curriculum Expansion, DataCamp

datacamp

Let's practice!

CONNECTING DATA IN TABLEAU

Relationships

CONNECTING DATA IN TABLEAU

Lis Sulmont

Head of Curriculum Expansion,
DataCamp

datacamp

Relationships, part 1: Introducing new data modeling in Tableau

SHARE

BETHANY LYONS

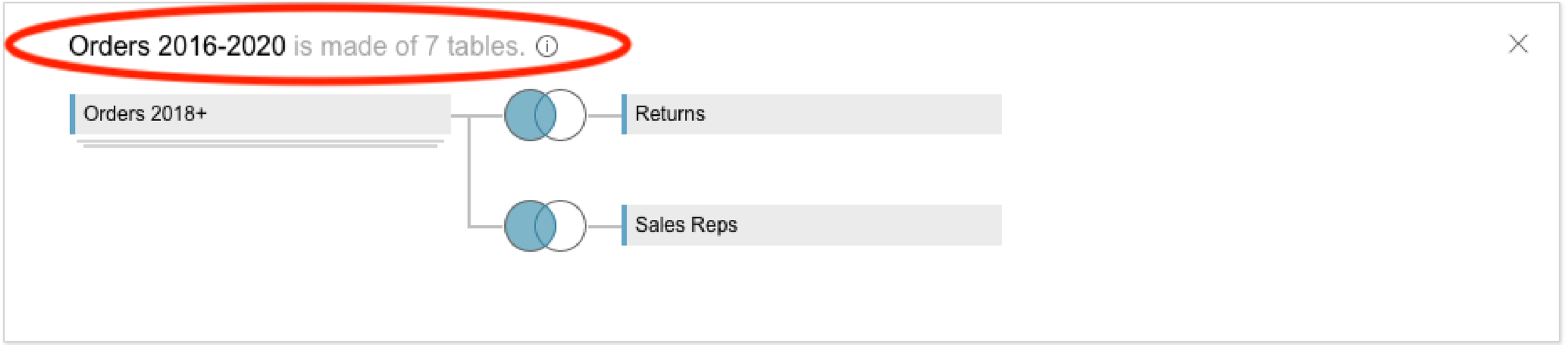
SENIOR PRODUCT MANAGER, TABLEAU

MAY 11, 2020

[1] https://www.tableau.com/about/blog/2020/5/relationships-part-1-meet-new-tableau-data-model

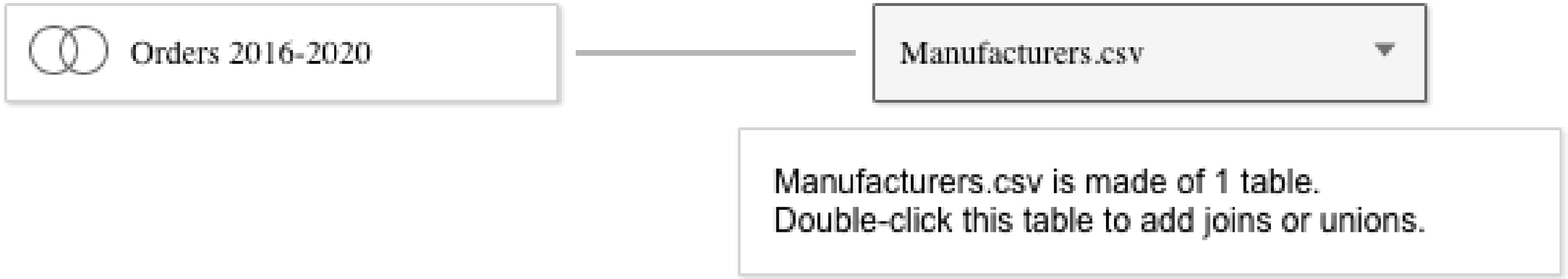
Relationships

A relationship describes how two tables relate to each other, based on common fields, but does not merge the tables together or append fields.



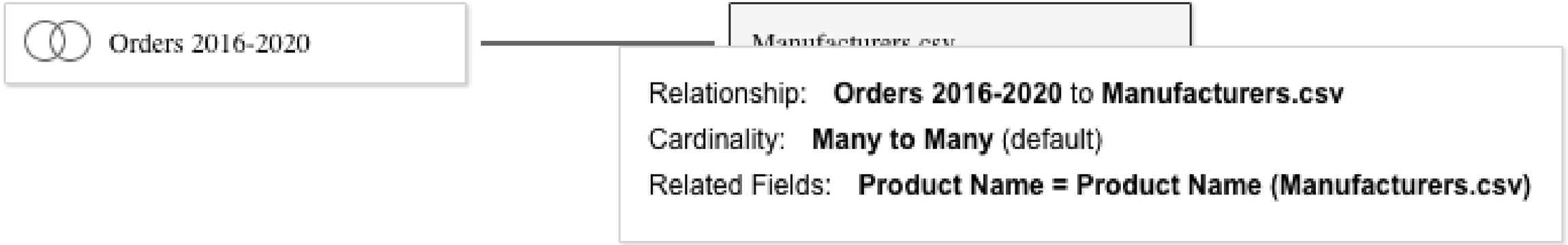
Relationships

A relationship describes how two tables relate to each other, based on common fields, but does not merge the tables together or append fields.



Relationships

A relationship describes how two tables relate to each other, based on common fields, but does not merge the tables together or append fields.



Relationships vs joins

Relationships

- You only need to select matching field(s), no join type
- Tableau automatically selects the join type based on the context
- Unmatched rows are kept
- Relationships reduce upfront data preparation, such as custom SQL code or database views

Relationships are more flexible and more dynamic.

Joins

- Choose a type of join: left, right, inner, full outer
- Certain rows may not be returned depending on the join type
- Changing join type mid-analysis will impact the work you've done

Let's practice!

CONNECTING DATA IN TABLEAU

Establishing a relationship

CONNECTING DATA IN TABLEAU

Lis Sulmont

Head of Curriculum Expansion, DataCamp

datacamp

Let's practice!

CONNECTING DATA IN TABLEAU

Extracts

CONNECTING DATA IN TABLEAU

Lis Sulmont

Head of Curriculum Expansion,
DataCamp

datacamp

Tableau file types

- **Packaged Workbook,** `.twbx` : workbook and supporting files including the data source
- **Workbook,** `.twb` : just the workbook
- **Extract,** `.hyper` or `.tde` : local copy of parts or all of the data
- **Data Source,** `.tds` : connection information and modifications to data

Extracts

A local copy of the entire or a subset of data.

- Performance improvements
  - Supports very large data
  - Faster because leverages the Hyper database engine
- Retain work done to enrich and extend the original data set
- Work offline

Live connections

Data source has a direct connection to underlying data

Comparison

Live connections

- Real-time updates
- Slower because connection to external databases and data files

Extracts

- Need to refresh data to get updates
- Generally faster performance with Tableau's Hyper

Let's practice!

CONNECTING DATA IN TABLEAU

Creating extracts

CONNECTING DATA IN TABLEAU

Lis Sulmont

Head of Curriculum Expansion, DataCamp

datacamp

Let's practice!

CONNECTING DATA IN TABLEAU