Tableau Training

simplilearn

DATA AND
ARTIFICIAL INTELLIGENCE

Connecting to Various Data Sources and Preparing Data

Learning Objectives

By the end of this lesson, you will be able to:

- Describe how to connect data to Tableau
- List the tools offered by Tableau to prepare data for analysis
- Demonstrate how Tableau data engine optimizes performance
- Create and use data extracts in Tableau

A Day in the Life of a Data Analyst

You are working as a data analyst in an organization that is looking for the data of users to gain insights for maintaining the consistency of a product's performance.

The goal is to create and use data extracts.
The team is also looking for a method for optimizing the performance of Tableau data engine and also for connecting data to Tableau.

To achieve all of the above, along with some additional features, you will be learning a few concepts in this lesson and listing out the tools that will help find a solution for the given scenario.

Connecting to Tableau Server

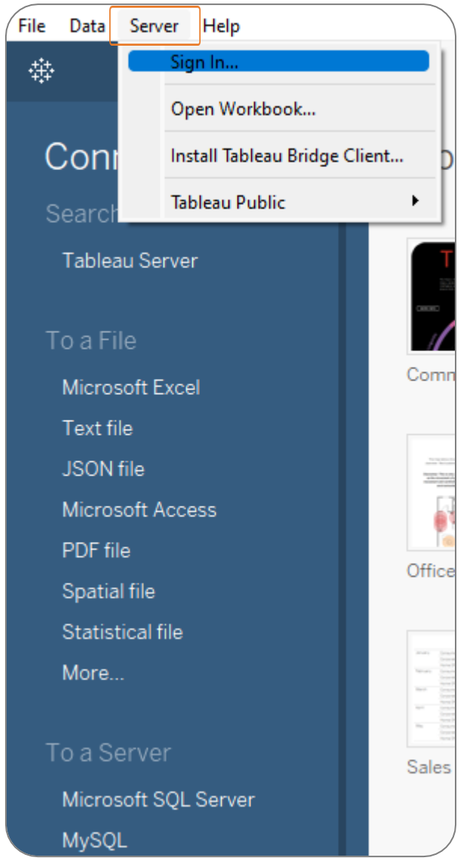
Connecting to Tableau Server

Tableau Desktop connects with Tableau Server for schedule-based data sources that are updated in a server.

Tableau Server connection is used to publish workbooks and data sources.

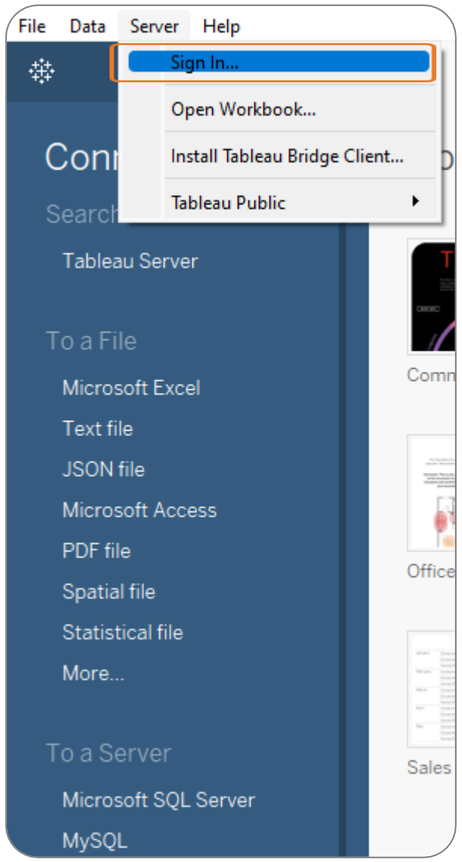
Connecting to Tableau Server

Step 1: Click on the **Server** tab



Connecting to Tableau Server

Step 2: Select **Sign In** option

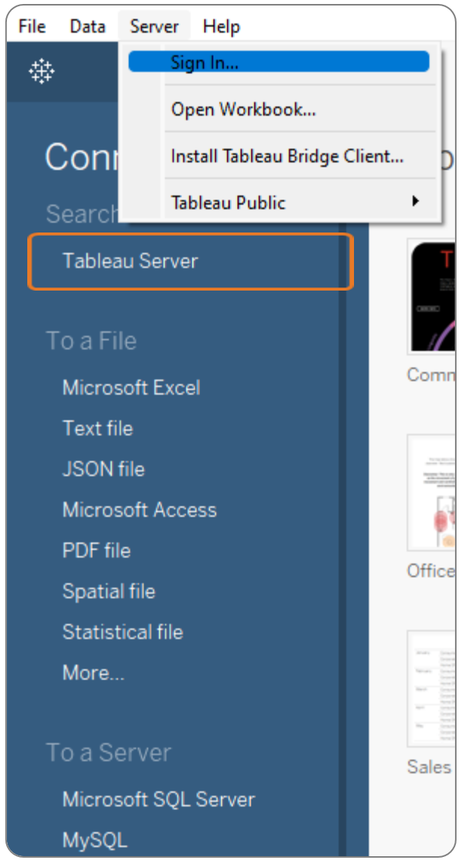


Connecting to Tableau Server

Step 3: Sign in with the credentials to connect to the server

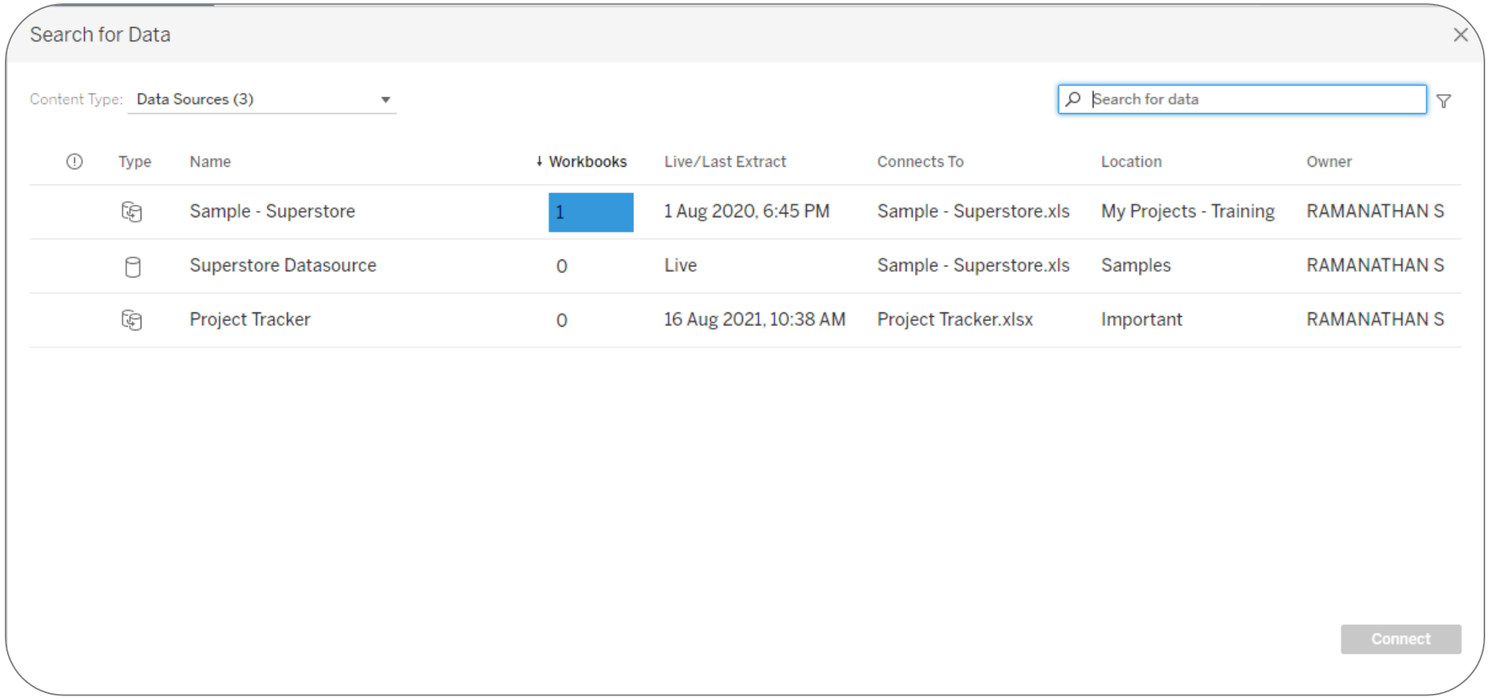
Connecting to Tableau Server

To connect to data sources published in the Tableau Server, select Tableau Server option in the start page.



Connecting to Tableau Server

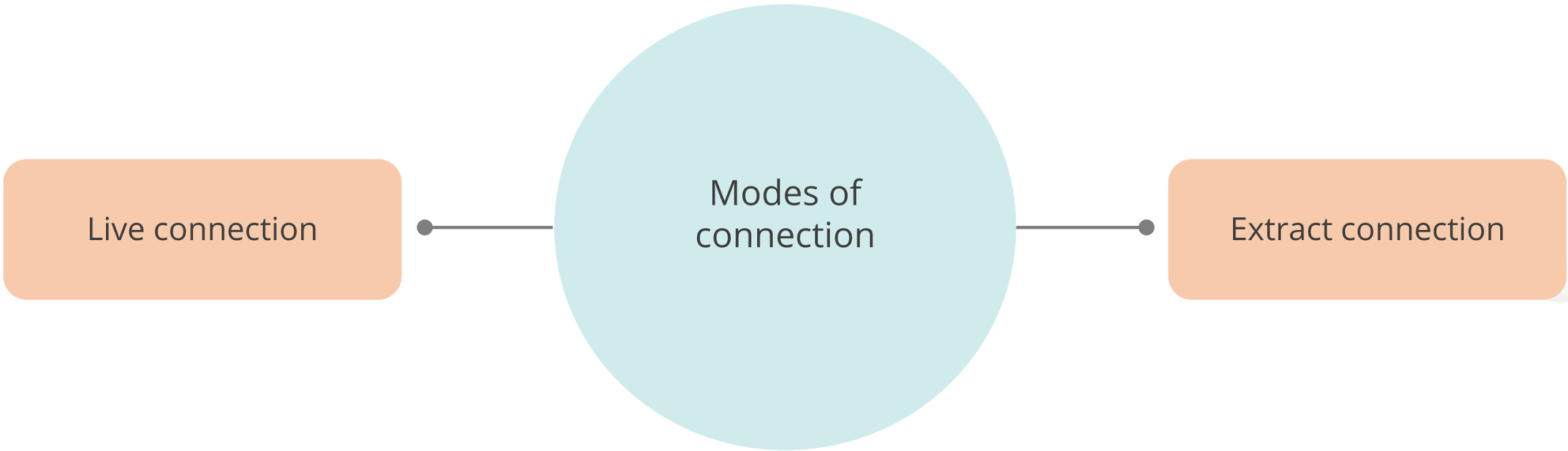
The published data sources will be available for selection.



Types of Connections

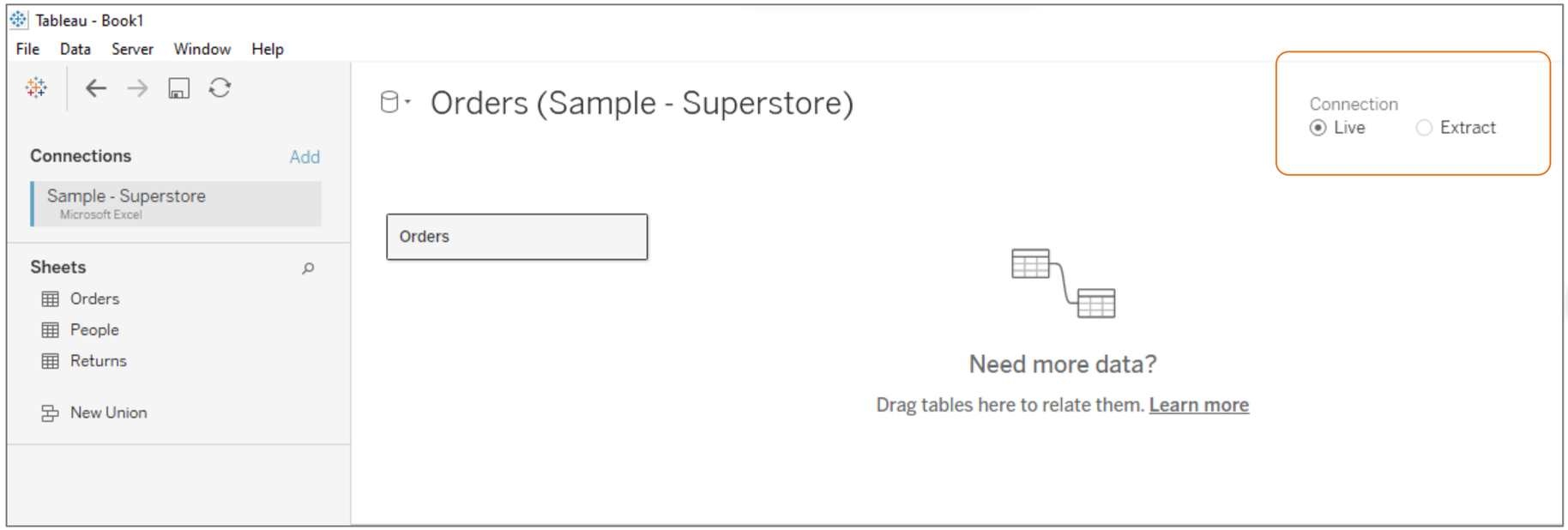
Types of Connections

Tableau supports two types of connecting data sources.



Types of Connections

The type of connection is selected on the data source page.



Types of Connections

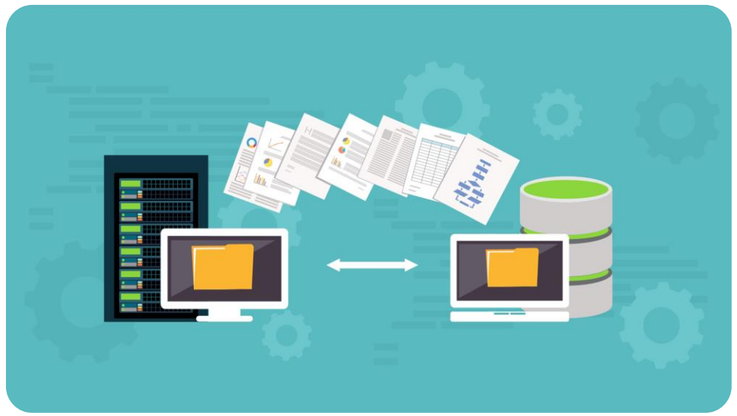
Live connection is used in the following scenarios:

Live Connection

- The dataset is smaller.
- It is used when real-time data is needed.
- The dashboard will reflect updates to the base data immediately.

Types of Connections

Extract connection is used in the following scenarios:



Extract Connection

- It supports larger dataset.
- The performance is faster than the live connection.
- It does not need real-time data in dashboards.
- The updates to the base data reflect only after the extract refresh.

DATA AND
ARTIFICIAL INTELLIGENCE

Connecting to Different Data Sources

Connecting to Different Data Sources

Popular data sources that can be connected to Tableau are:

Excel file

JSON file

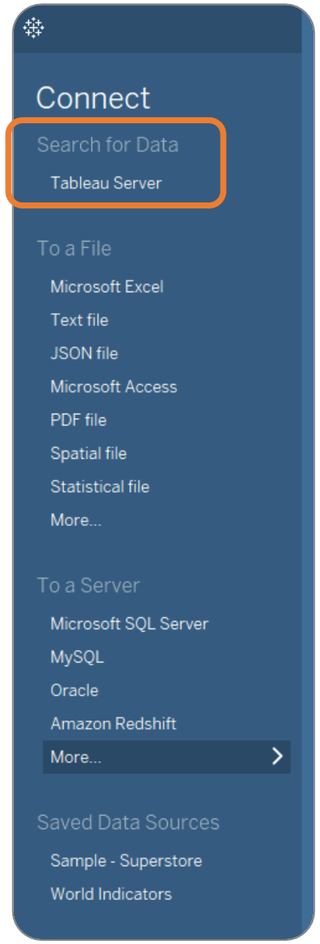
Statistical file

Text file

Spatial file

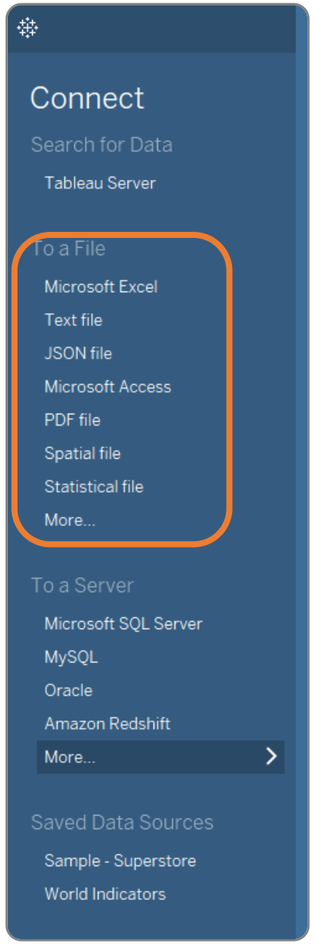
Connecting to Different Data Sources

Tableau Server, listed under **Search for Data**, can be used to connect to the extracted data.



Connecting to Different Data Sources

There are various options mentioned under **To a File** in the **Connect** pane.

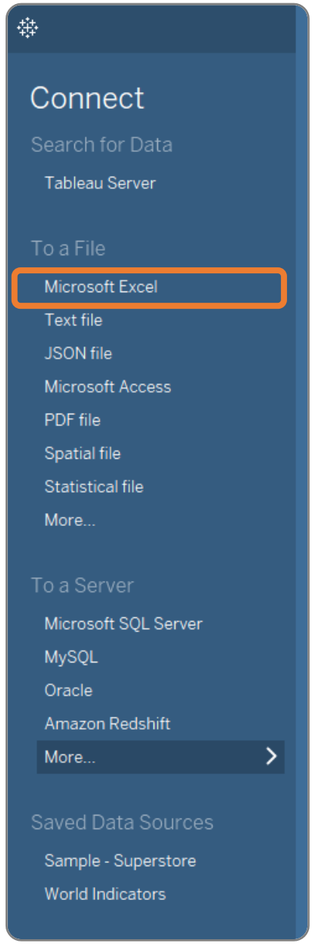


Connecting to Different Data Sources

Let’s understand this by connecting to a Microsoft Excel file.

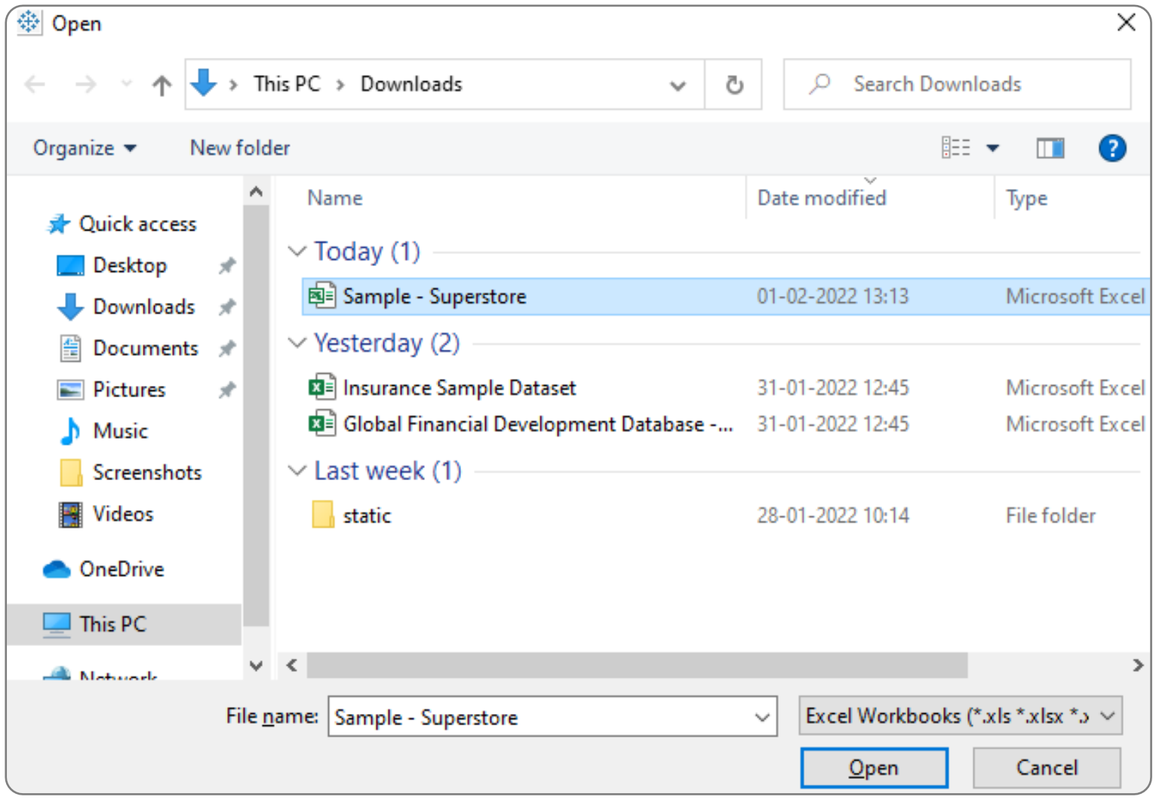
Connecting to Different Data Sources

Step 1: Select **Microsoft Excel** from the **Connect** pane



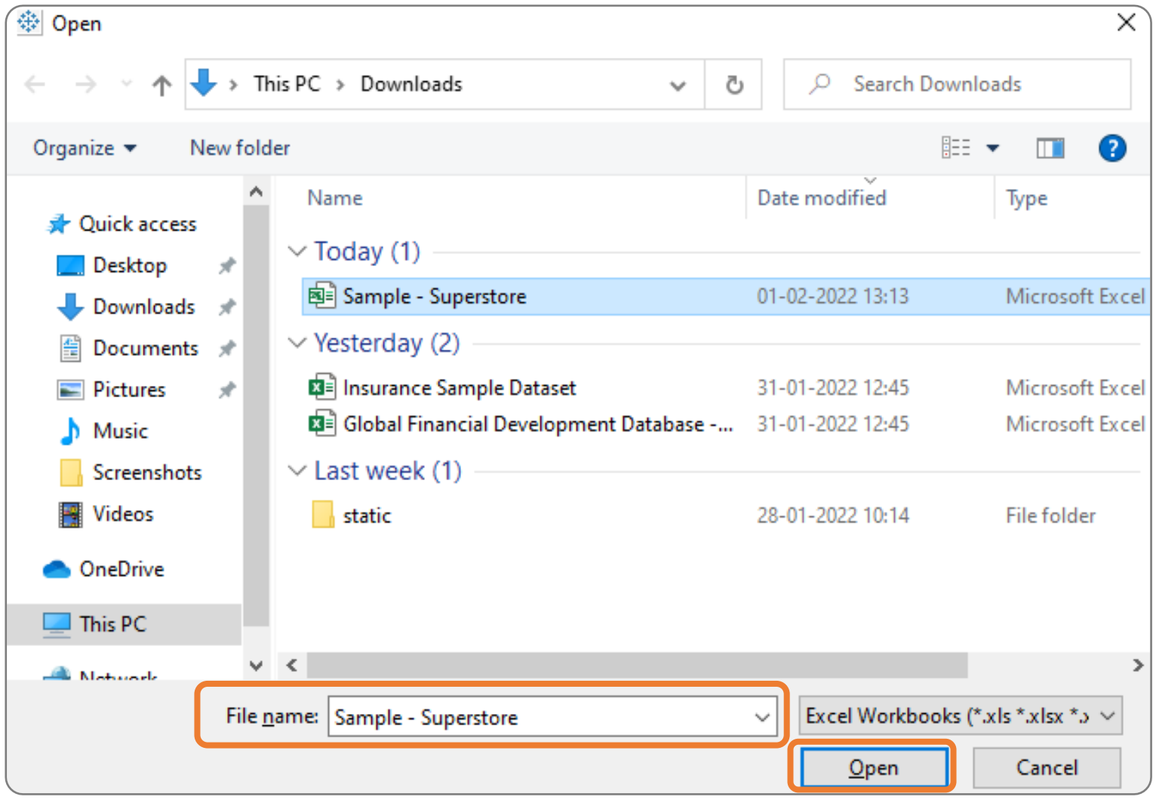
Connecting to Different Data Sources

A pop-up window will open.



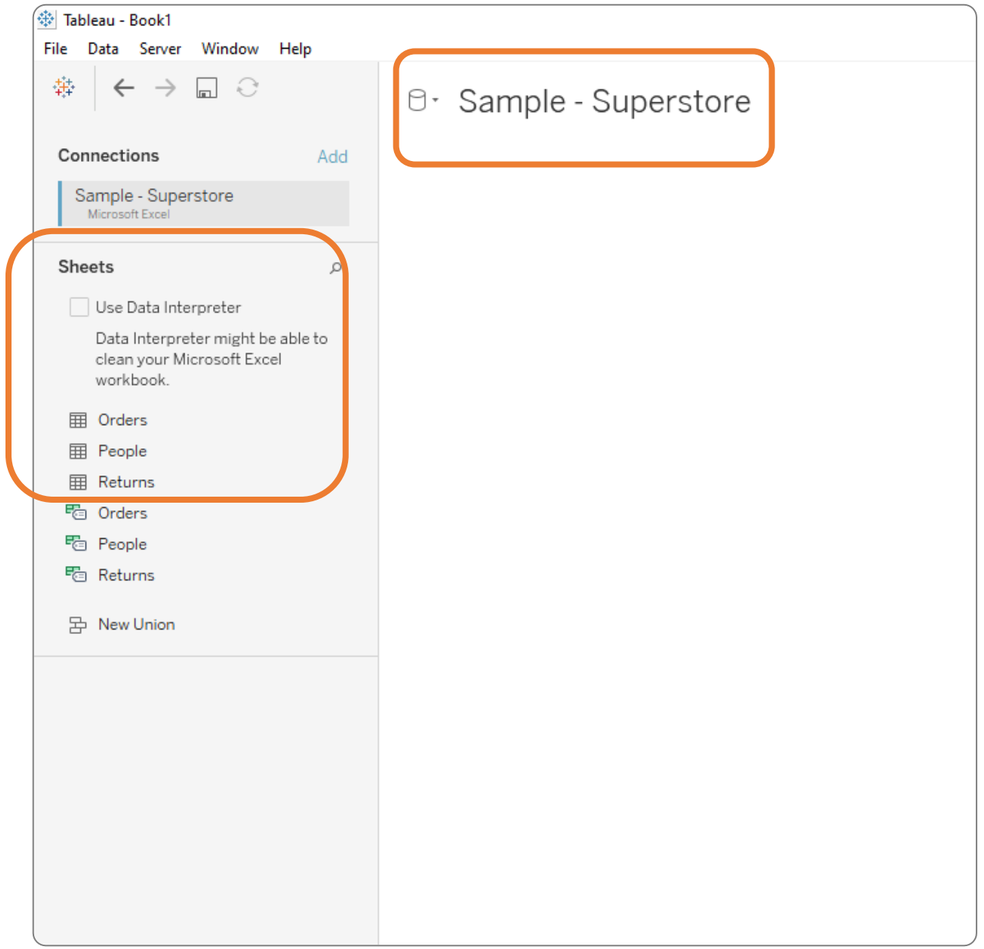
Connecting to Different Data Sources

Step 2: Select the **Sample-Superstore** dataset



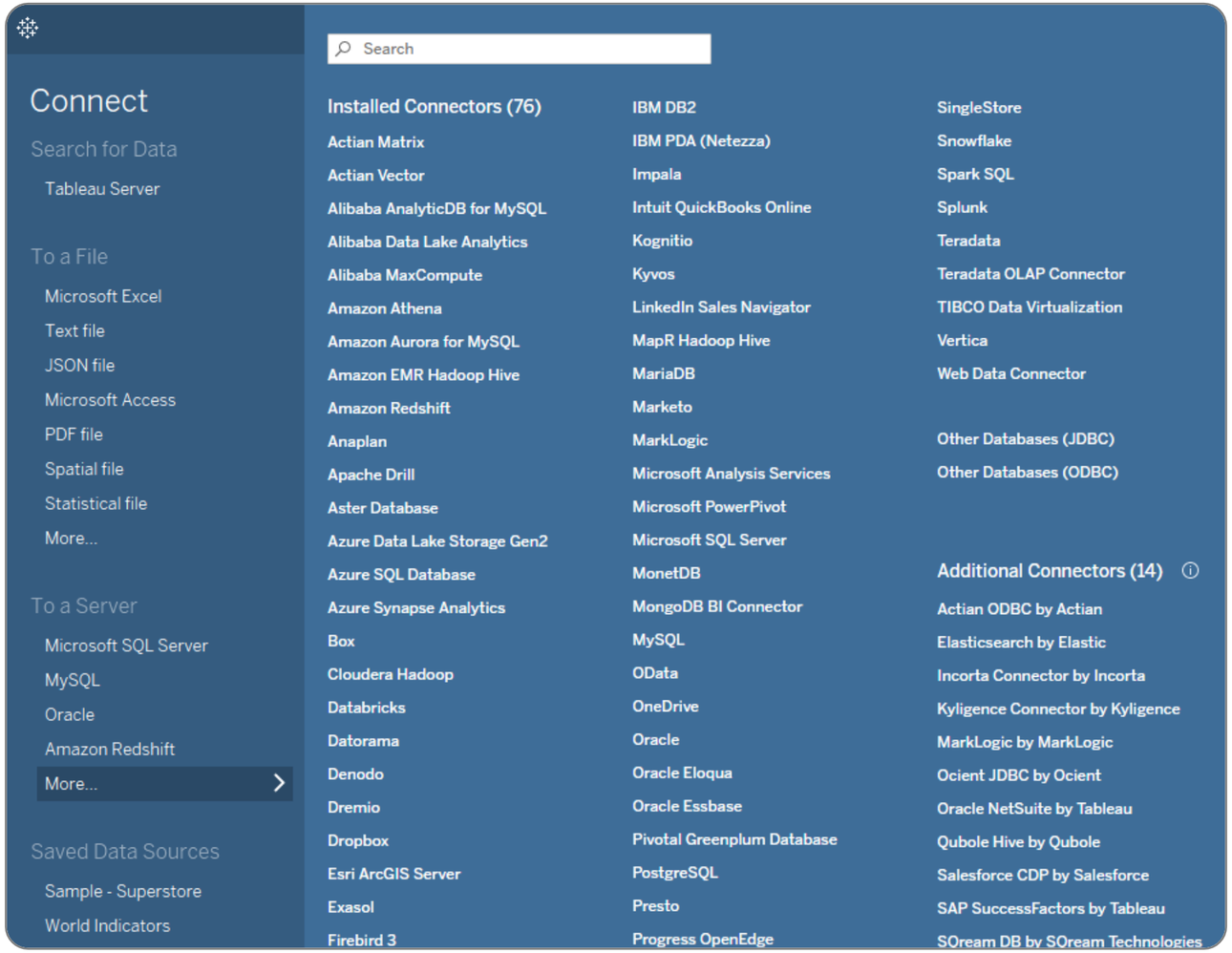
Connecting to Different Data Sources

The **Data Source** page will be opened which displays the **Sample-Superstore** dataset.



Connecting to Different Data Sources

Tableau has built-in native connections optimized for various data sources.



DATA AND
ARTIFICIAL INTELLIGENCE

Preparing Data for Analysis

simplilearn

Preparing Data for Analysis

Data preparation is the process of preparing data for visualization.

Preparing Data for Analysis

The following built-in tools are used to prepare data:

01	Joins	05	Aggregations
02	Relationships	06	Pivot
03	Unions	07	Data Interpreter
04	Data Blending	08	Split

DATA AND
ARTIFICIAL INTELLIGENCE

Joins

Joins

A join merges two tables side by side related by specific fields.

A join results in a virtual table extended horizontally by adding columns.

Joins

Joins are used to combine data for desired aspects, such as:

Filtering

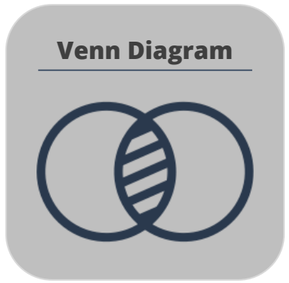
Removing duplicate

Types of Joins

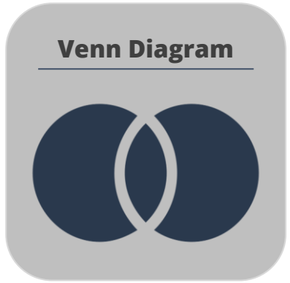
There are four types of joins:

Left Join

Right Join



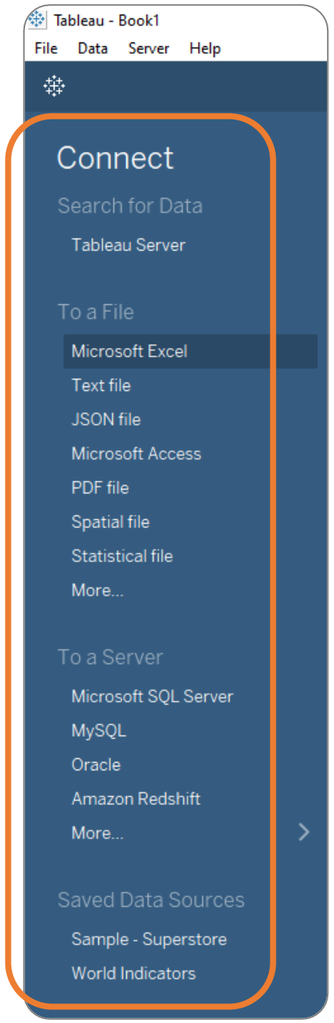
Inner Join



Full Outer Join

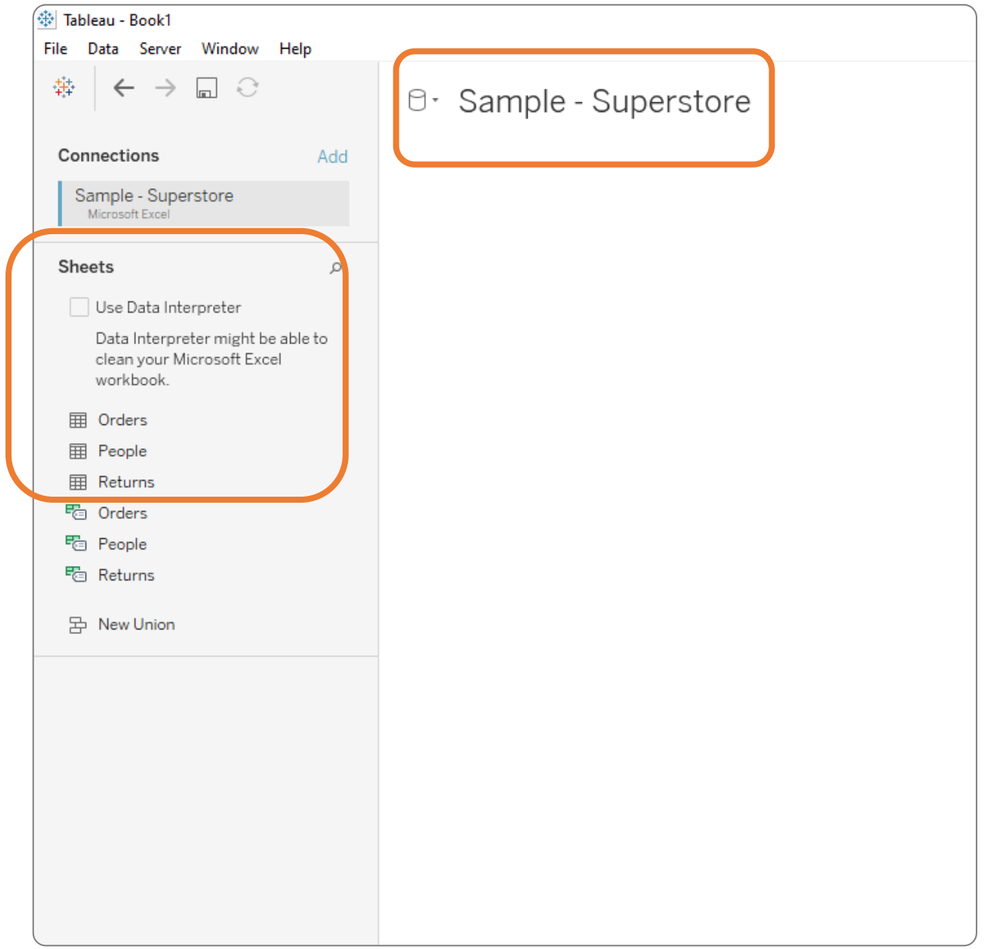
Steps to Perform Joins

Step 1: Connect to the **Sample-Superstore** dataset from the **Connect** pane



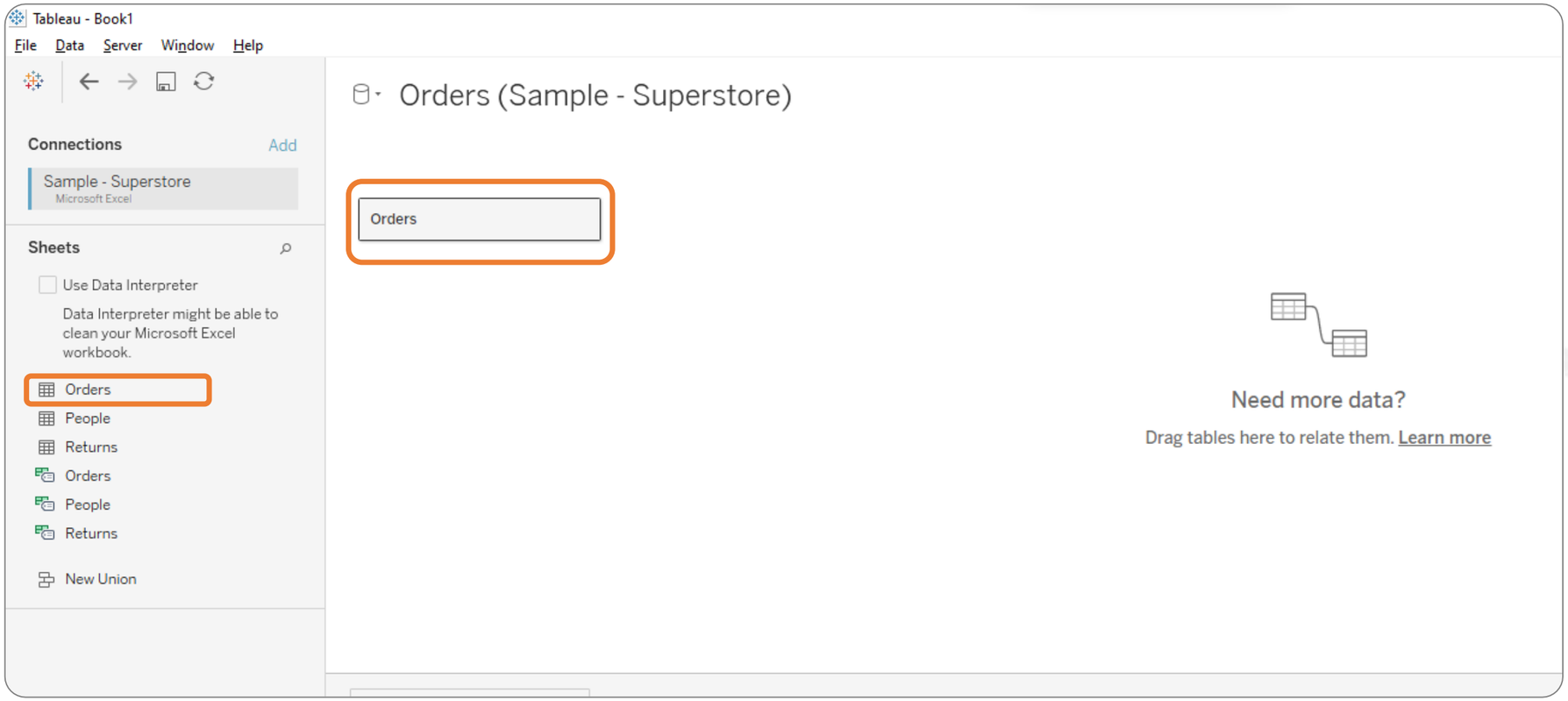
Steps to Perform Joins

The data source page will show the connections and the three **Sheets** in the file.



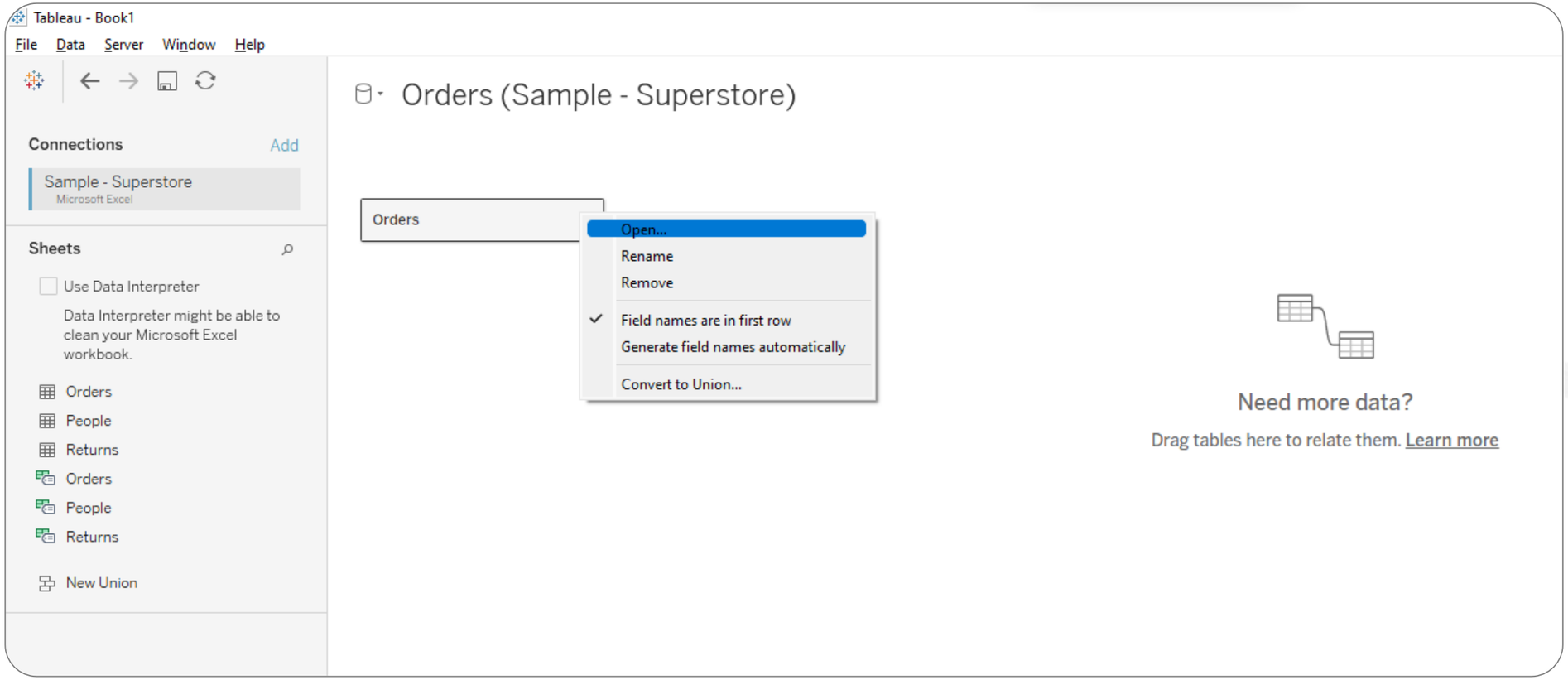
Steps to Perform Joins

Step 2: Drag **Orders** table to the canvas



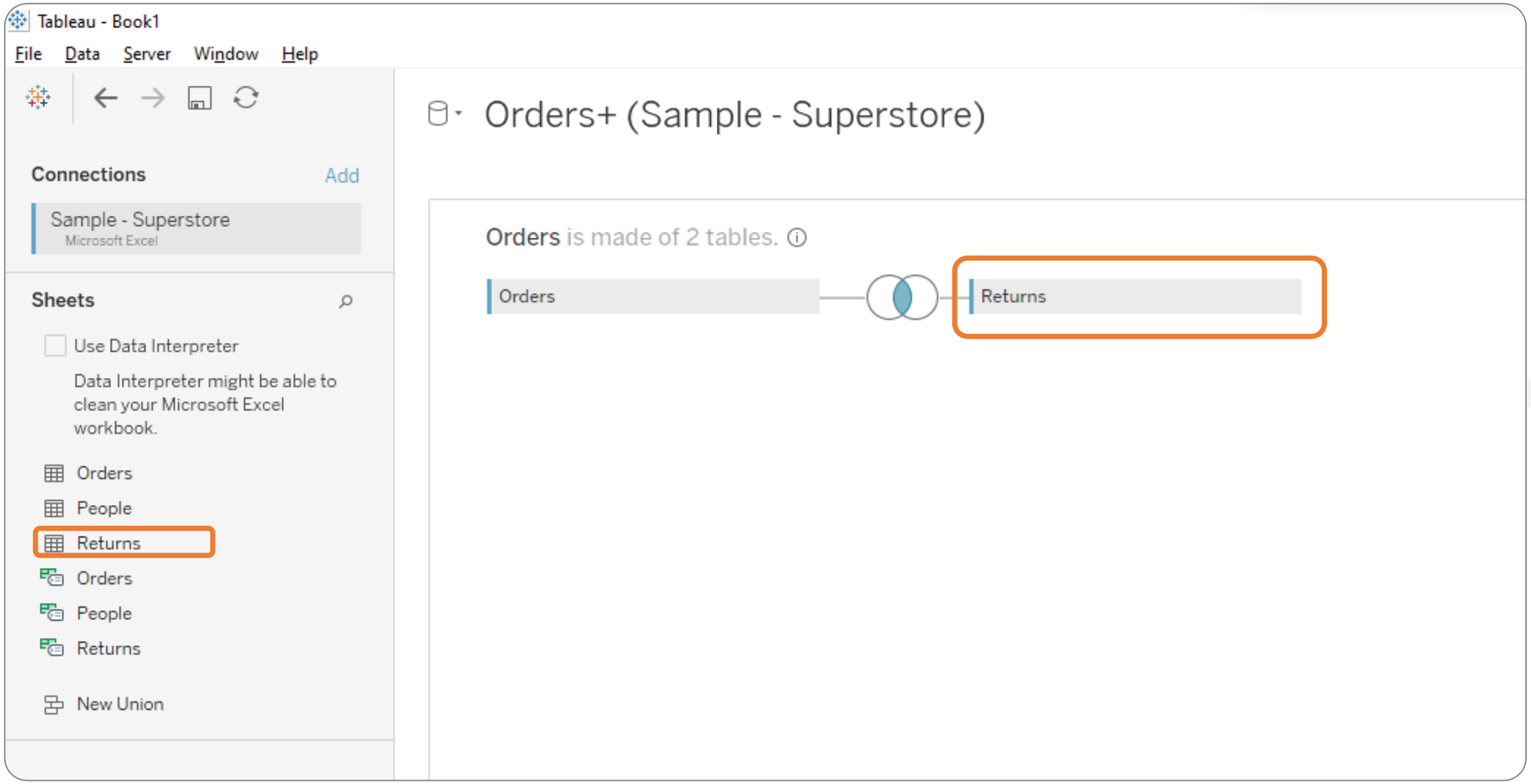
Steps to Perform Joins

Step 3: Right-click on **Orders** and then click **Open**



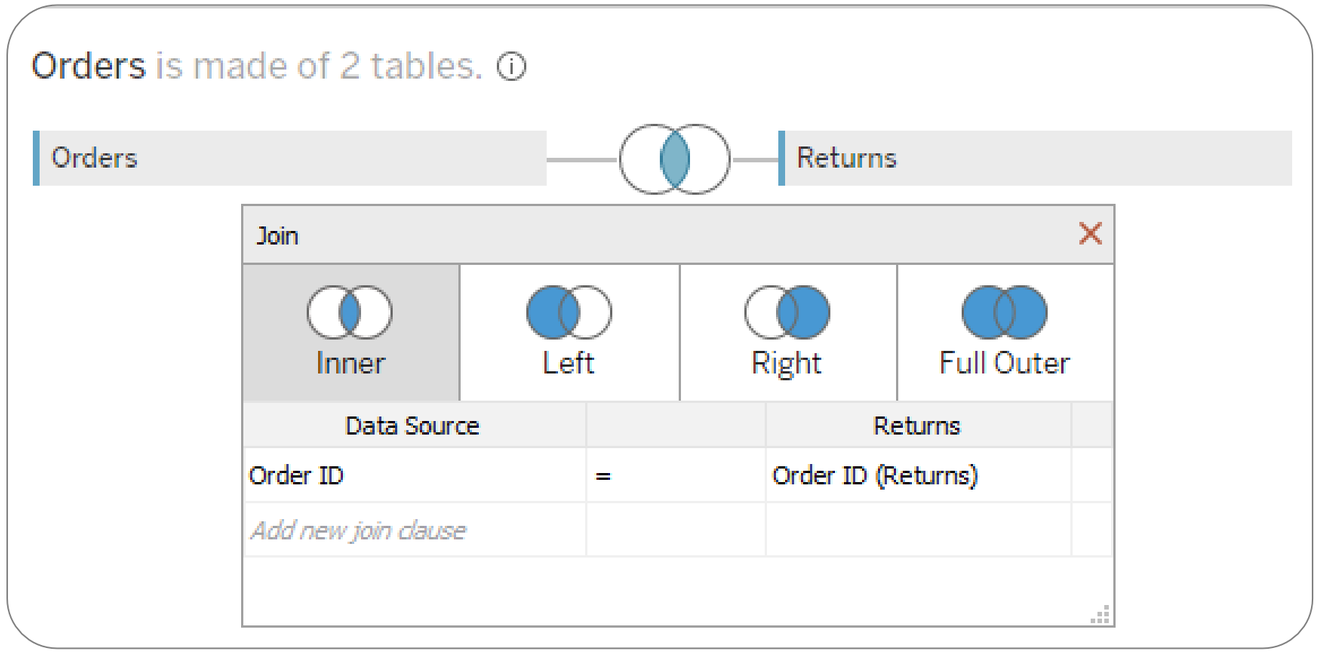
Steps to Perform Joins

Step 4: Drag **Returns** table to the canvas to create an inner join by default



Steps to Perform Joins

Step 5: Choose a **Venn Diagram** to select the join criteria and type



Joins: Example

Analysis of product sales from two different files:

Product Sales: Left table

Product Sales		
Product ID	Sales	Number of Records
6	985	2
7	243	15
8	652	17
9	1123	25
10	851	33
11	665	15
12	458	12
13	784	9
14	965	8
15	1033	29

Product Profits: Right table

Product Profits		
Product ID	Profits	Number of Records
1	246	15
2	61	12
3	163	9
4	281	8
5	213	29
6	166	22
7	115	6
8	196	41
9	241	13
10	258	21

The product ID field serves as the primary key for joining data from the two sets.

Left Join: Example

The result of the left join will contain all values from the LEFT table and match the RIGHT table.

Product Sales		
Product ID	Sales	Number of Records
6	985	2
7	243	15
8	652	17
9	1123	25
10	851	33
11	665	15
12	458	12
13	784	9
14	965	8
15	1033	29

Product Profits		
Product ID	Profits	Number of Records
1	246	15
2	61	12
3	163	9
4	281	8
5	213	29
6	166	22
7	115	6
8	196	41
9	241	13
10	258	21

Right Join: Example

The result of the right join will contain all values from the RIGHT table and match the LEFT table.

Product Sales		
Product ID	Sales	Number of Records
6	985	2
7	243	15
8	652	17
9	1123	25
10	851	33
11	665	15
12	458	12
13	784	9
14	965	8
15	1033	29

Product Profits		
Product ID	Profits	Number of Records
1	246	15
2	61	12
3	163	9
4	281	8
5	213	29
6	166	22
7	115	6
8	196	41
9	241	13
10	258	21

Inner Join: Example

The result of the inner join will include common data present in both datasets.

Product Sales		
Product ID	Sales	Number of Records
6	985	2
7	243	15
8	652	17
9	1123	25
10	851	33
11	665	15
12	458	12
13	784	9
14	965	8
15	1033	29

Product Profits		
Product ID	Profits	Number of Records
1	246	15
2	61	12
3	163	9
4	281	8
5	213	29
6	166	22
7	115	6
8	196	41
9	241	13
10	258	21

Full Outer Join: Example

The result of the full outer join will include all values from both tables.

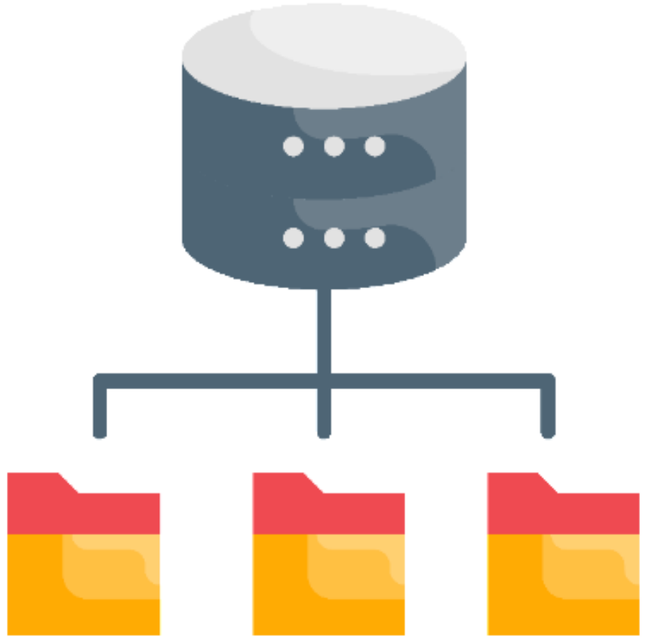
Product Sales		
Product ID	Sales	Number of Records
6	985	2
7	243	15
8	652	17
9	1123	25
10	851	33
11	665	15
12	458	12
13	784	9
14	965	8
15	1033	29

Product Profits		
Product ID	Profits	Number of Records
1	246	15
2	61	12
3	163	9
4	281	8
5	213	29
6	166	22
7	115	6
8	196	41
9	241	13
10	258	21

Relationships

Relationships

It is a method to combine multiple datasets, and it supports many-to-many cardinality.



Relationships

It is a simple and flexible way to link datasets together and is defined based on matching fields.

It can combine fields with non-unique values.

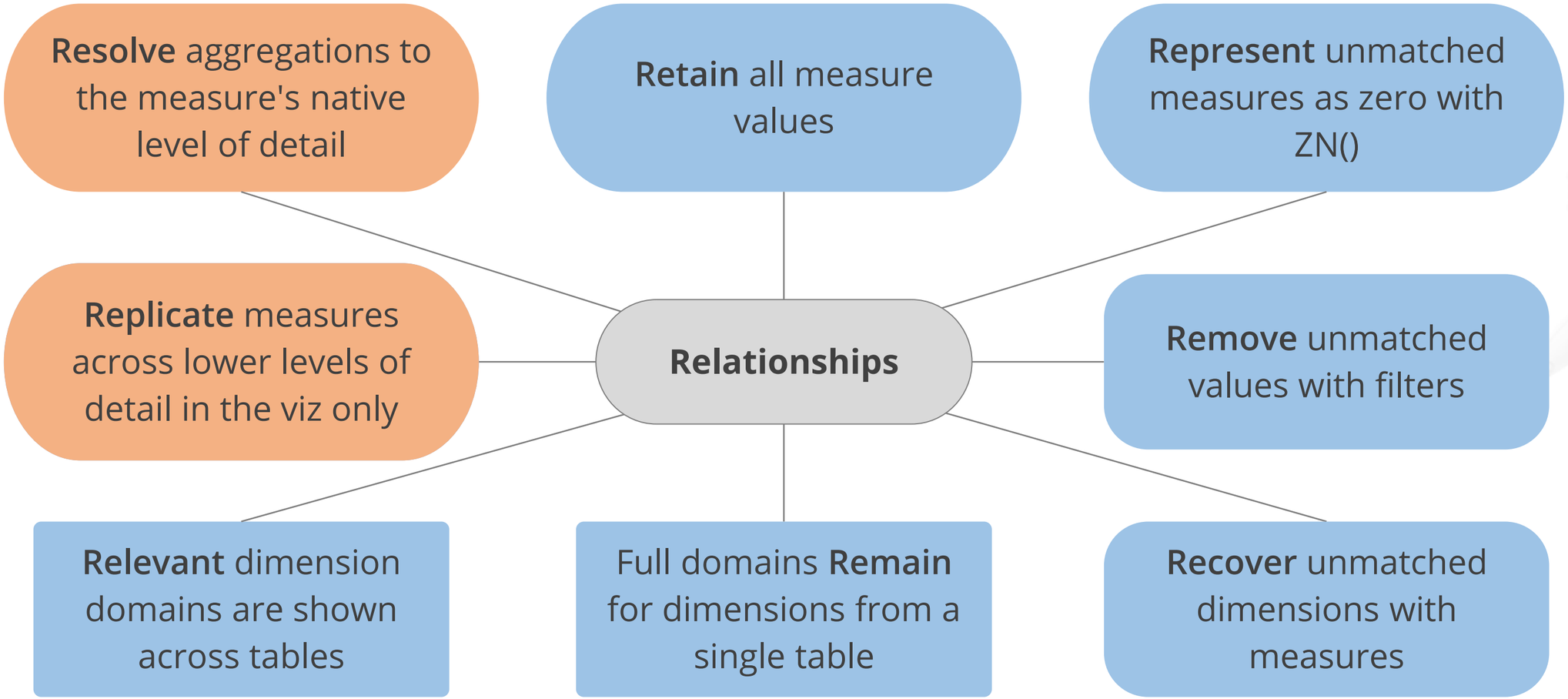
Relationships

It ensures the following benefits:

- Less upfront data preparation
- More use cases per data source
- Greater trust in results

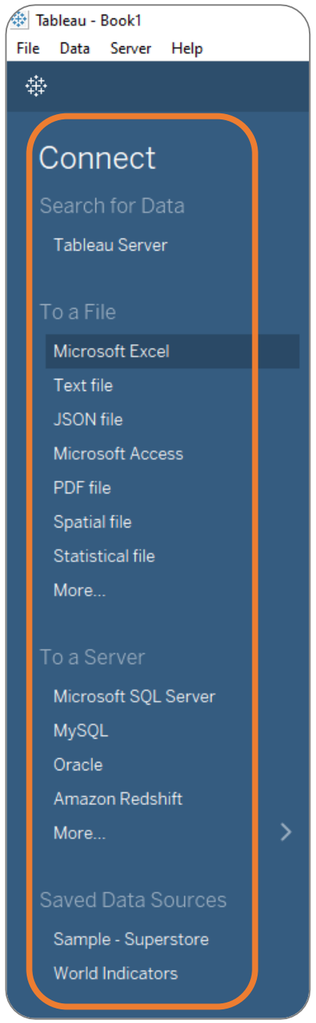
Relationships

The figure below illustrates the 8Rs of relationship semantics:



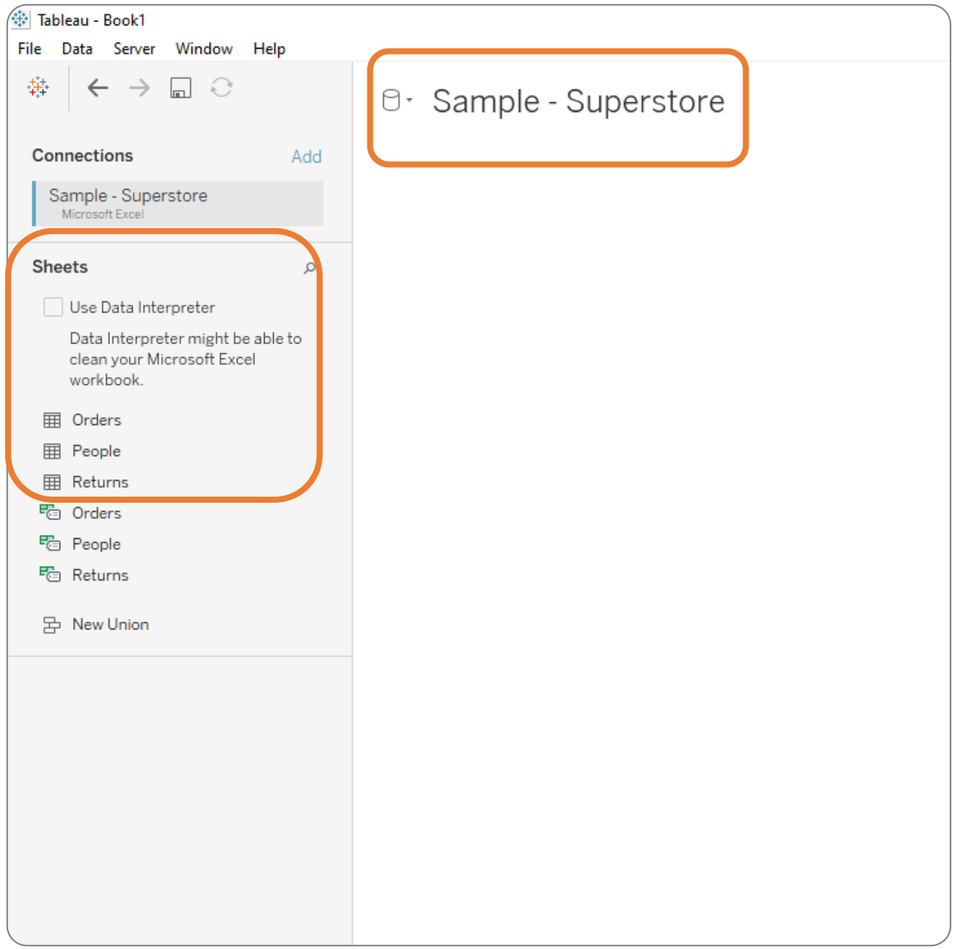
Steps to Perform Relationships

Step 1: Connect to the **Sample-Superstore** dataset from the **Connect** pane



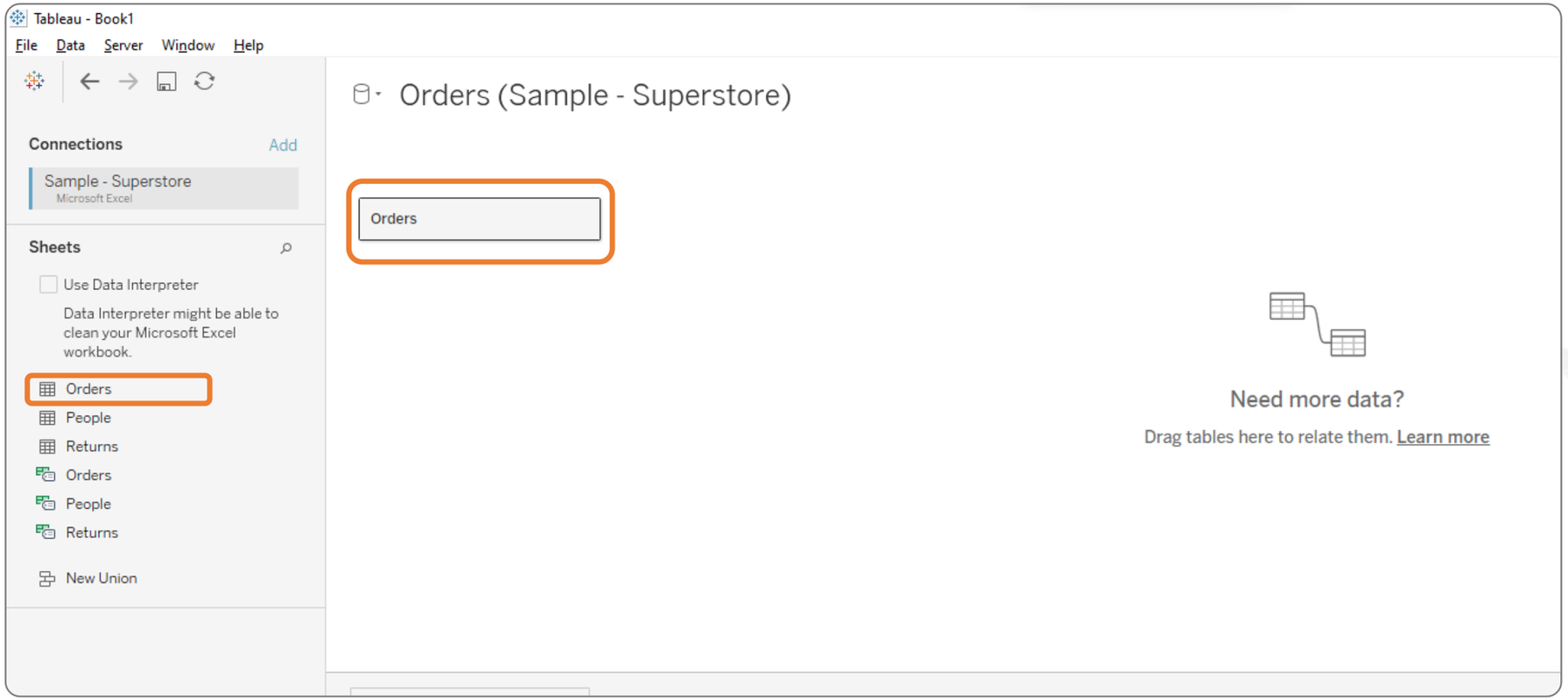
Steps to Perform Relationships

The data source page will show the connections and the three **Sheets** in the file.



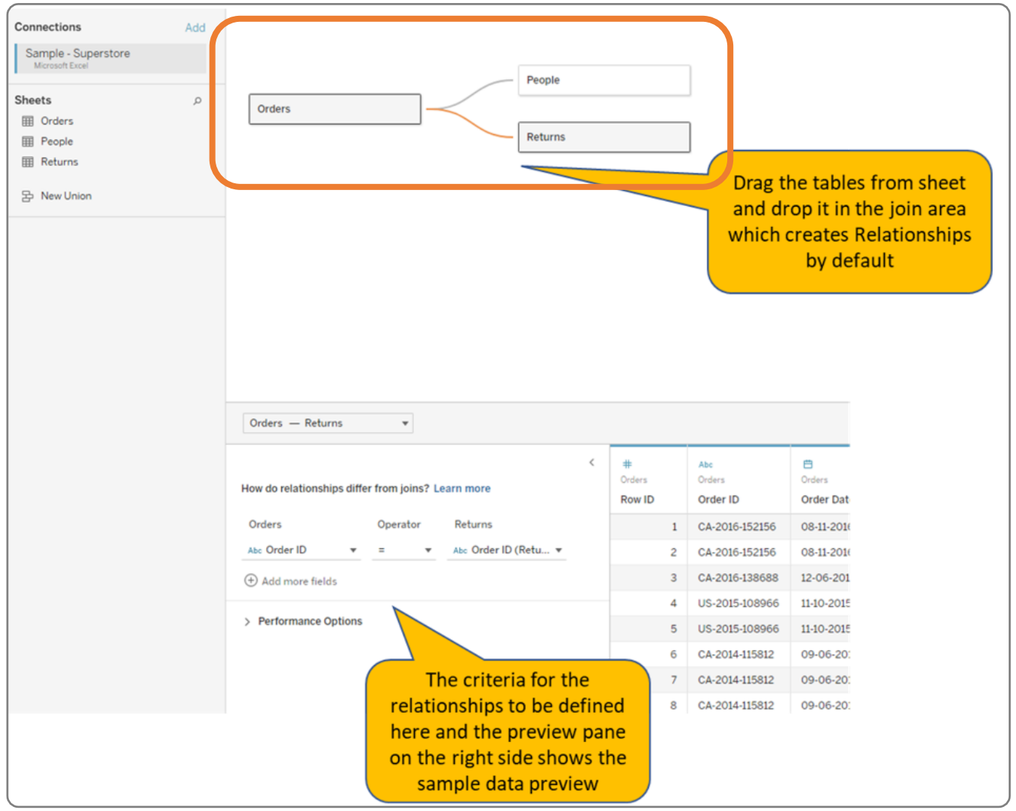
Steps to Perform Relationships

Step 2: Drag **Orders** table to the canvas



Steps to Perform Relationships

Step 3: Drag **People** and **Returns** table to the canvas that creates Relationships by default



The **Relationships** are created between the tables of the dataset.

Union

Union

The **Union** feature in Tableau helps assemble data from multiple small files into one large file.

Region	Category	Profit	Sales
West	Furniture	$11,505	$252,613
	Office Supplies	$52,610	$220,853
	Technology	$44,304	$251,992

Region	Category	Profit	Sales
East	Furniture	$3,046	$208,291
	Office Supplies	$41,015	$205,516
	Technology	$47,462	$264,974

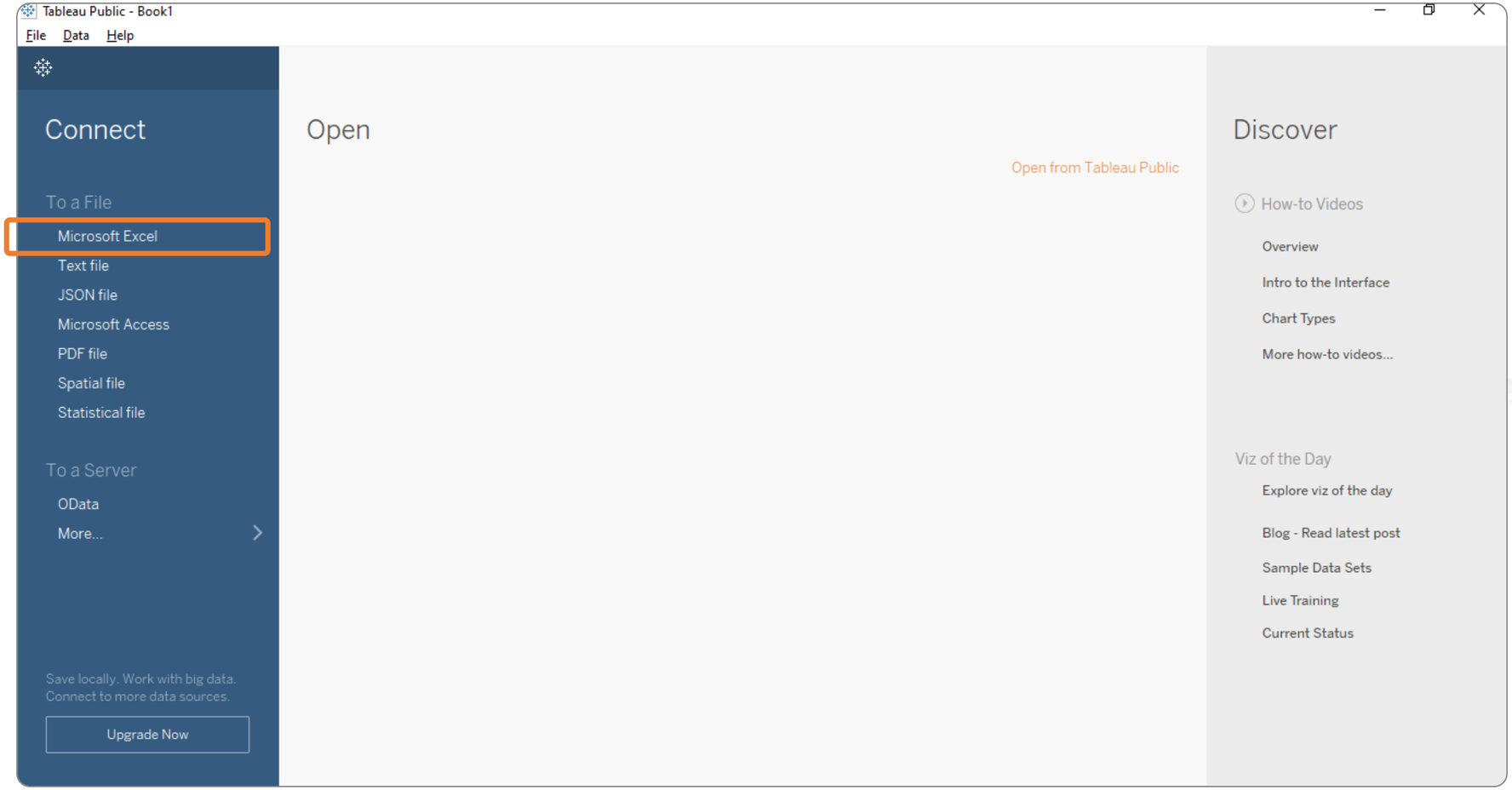
Region	Category	Profit	Sales
Central	Furniture	($2,871)	$163,797
	Office Supplies	$8,880	$167,026
	Technology	$33,697	$170,416

Region	Category	Profit	Sales
South	Furniture	$6,771	$117,299
	Office Supplies	$19,986	$125,651
	Technology	$19,992	$148,772

Region	Category	Profit	Sales
Central	Furniture	($2,871)	$163,797
	Office Supplies	$8,880	$167,026
	Technology	$33,697	$170,416
East	Furniture	$3,046	$208,291
	Office Supplies	$41,015	$205,516
	Technology	$47,462	$264,974
South	Furniture	$6,771	$117,299
	Office Supplies	$19,986	$125,651
	Technology	$19,992	$148,772
West	Furniture	$11,505	$252,613
	Office Supplies	$52,610	$220,853
	Technology	$44,304	$251,992

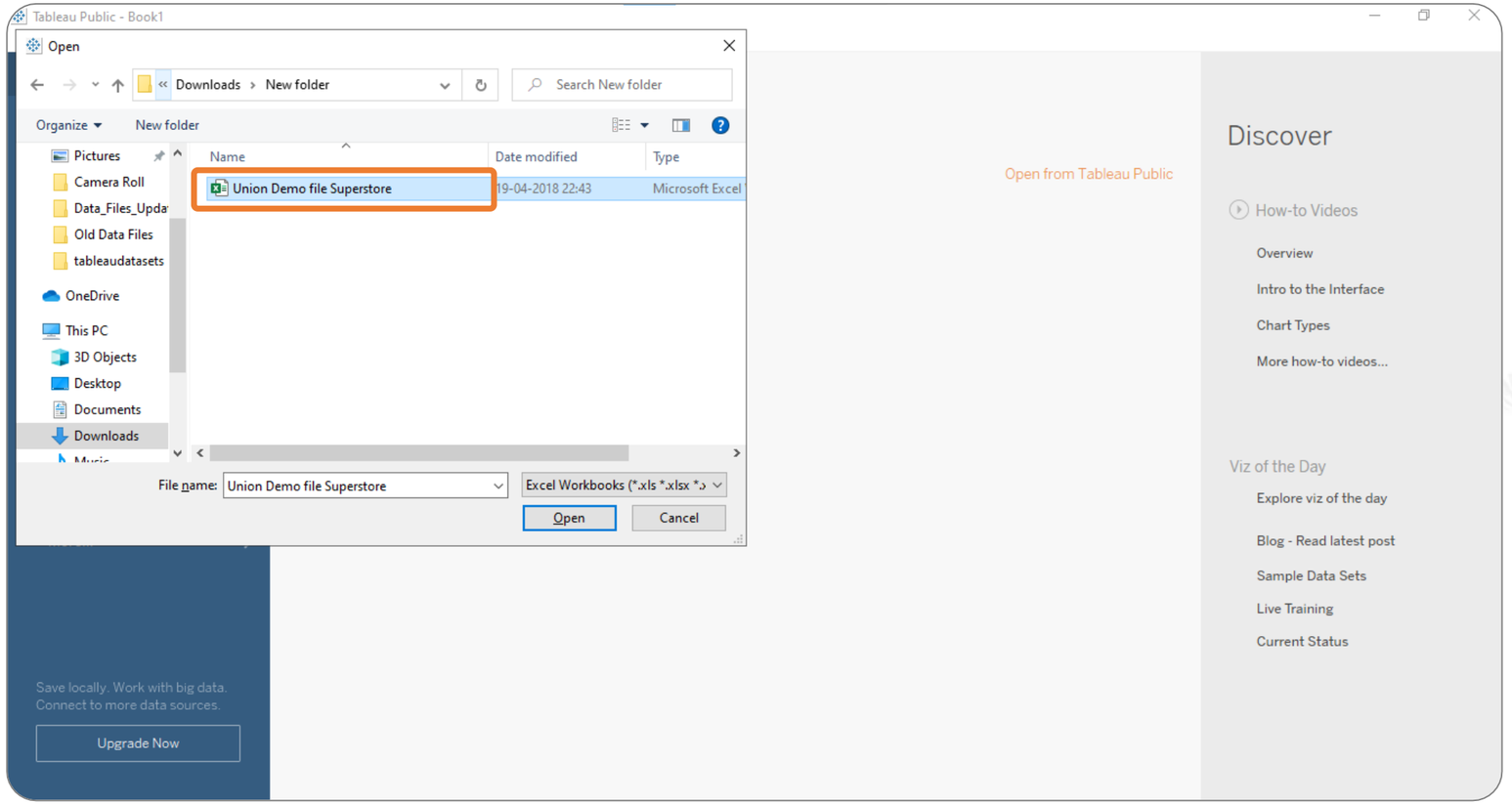
Steps to Perform Union

Step 1: Open Tableau Public and select **Microsoft Excel**



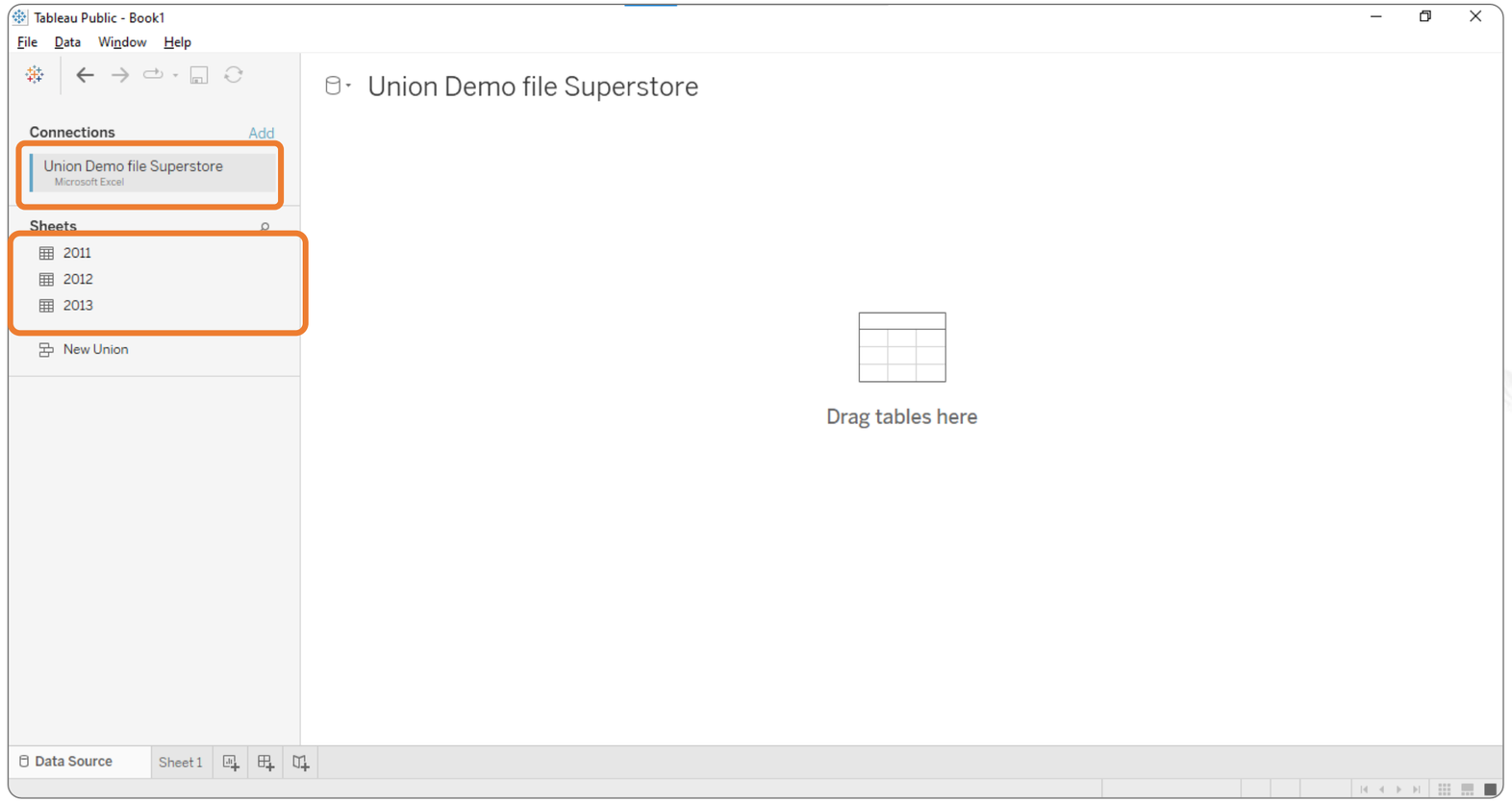
Steps to Perform Union

Step 2: Select the **Union Demo file Superstore** dataset



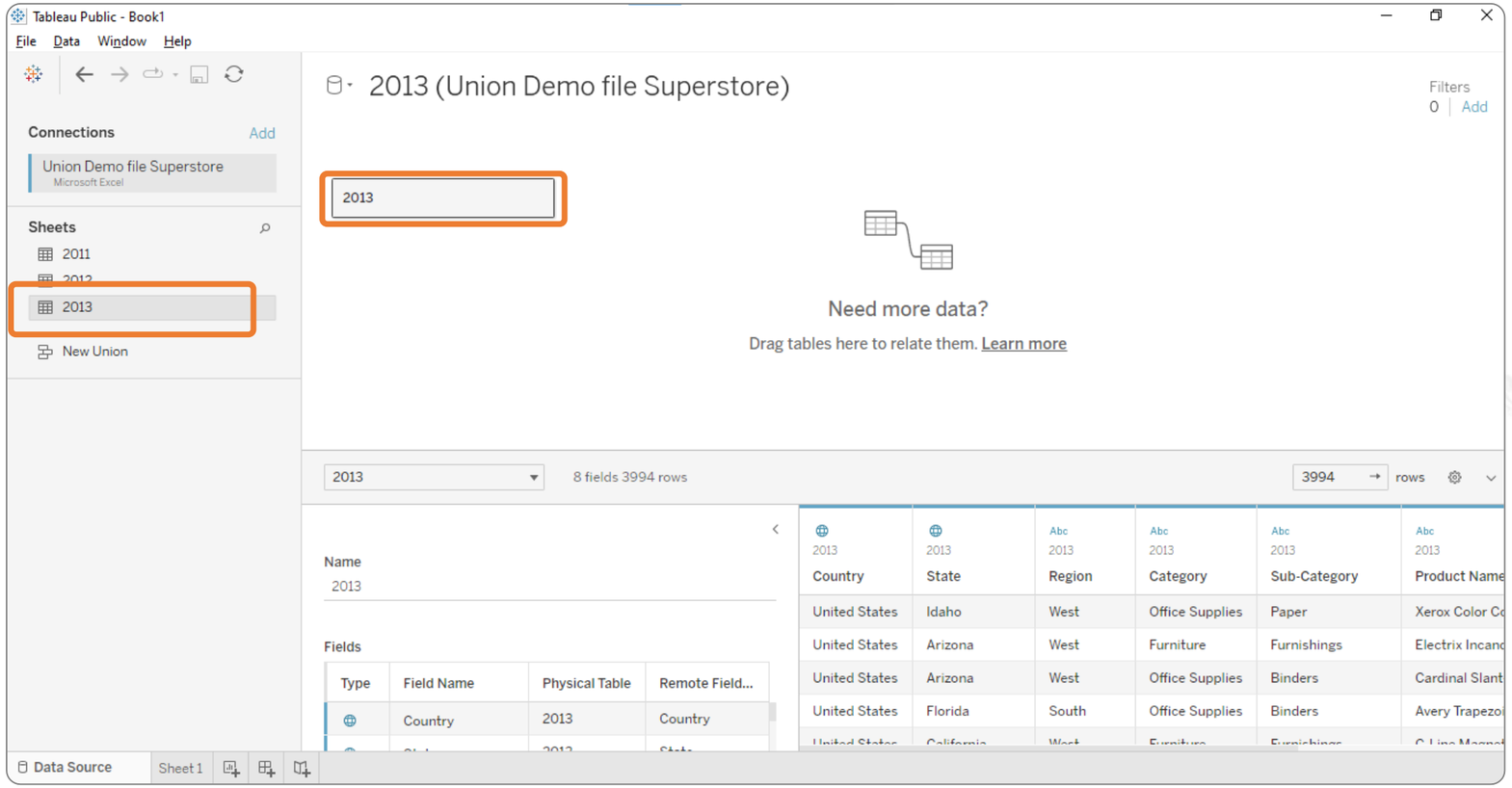
Steps to Perform Union

The **Data Source** page will be opened which displays the **Union Demo file Superstore** dataset.



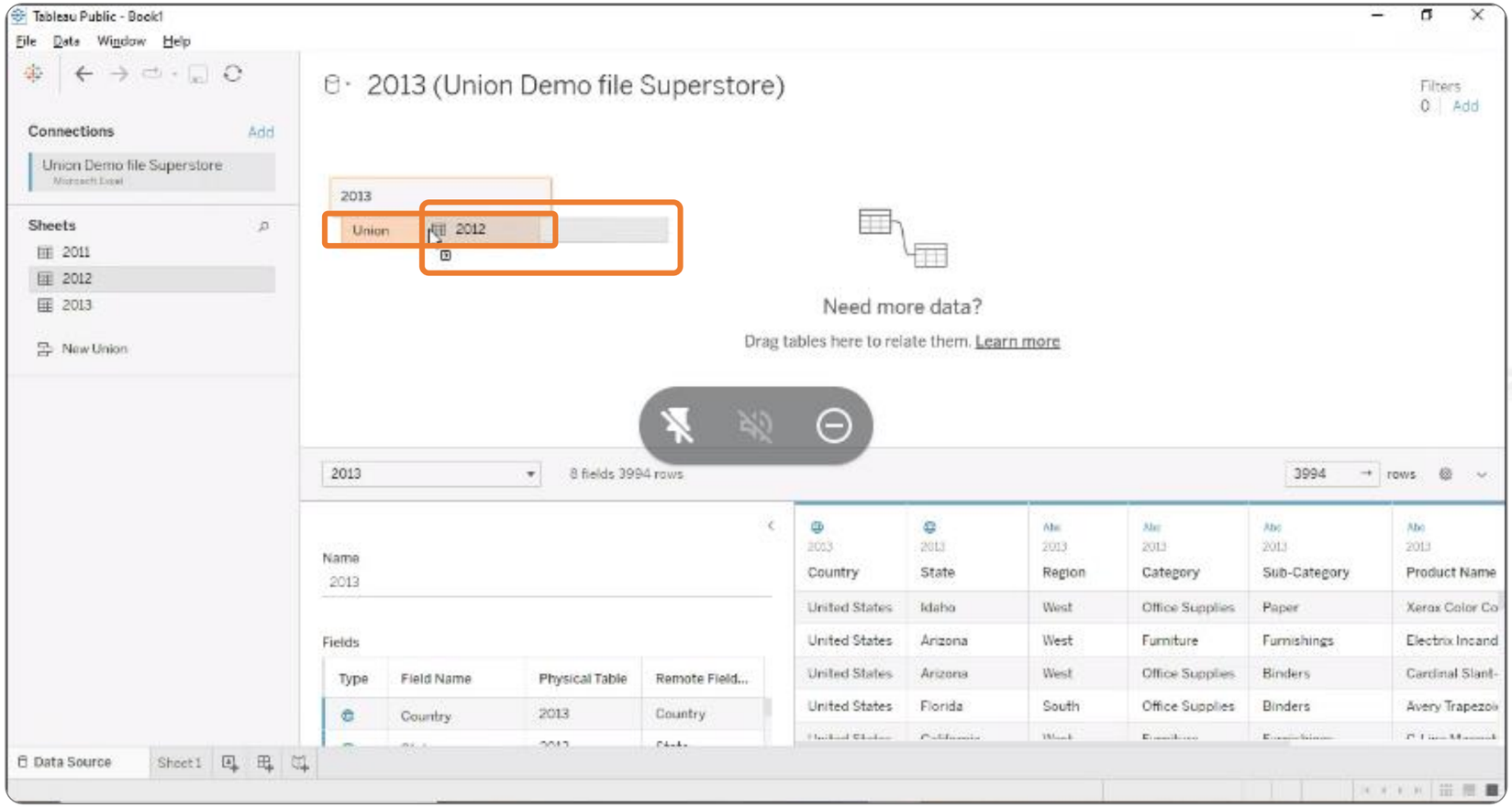
Steps to Perform Union

Step 3: Drag **2013** table to the canvas



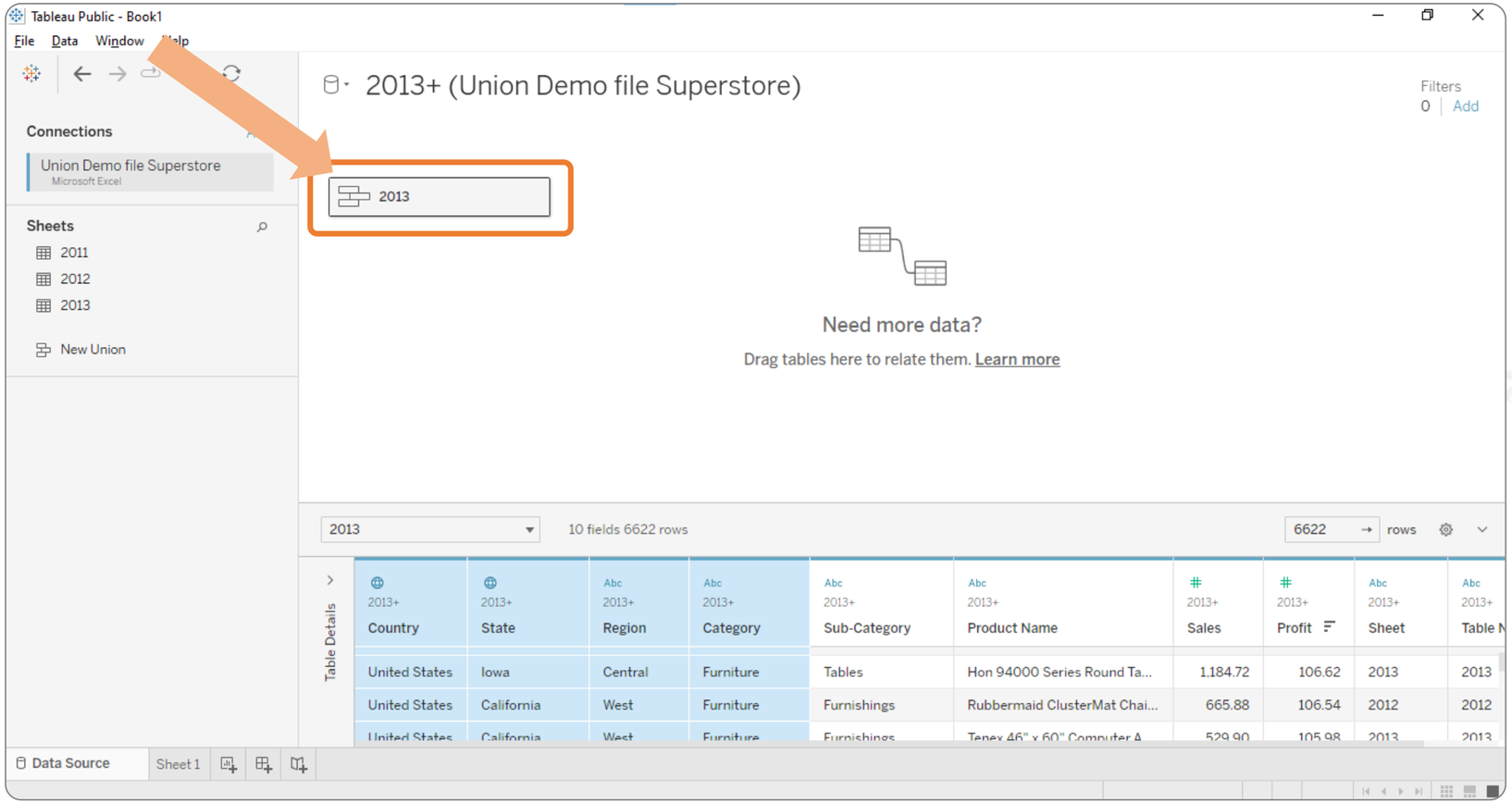
Steps to Perform Union

Step 4: Drag **2012** table to overlap with **2013** until **Union** pop-up shows



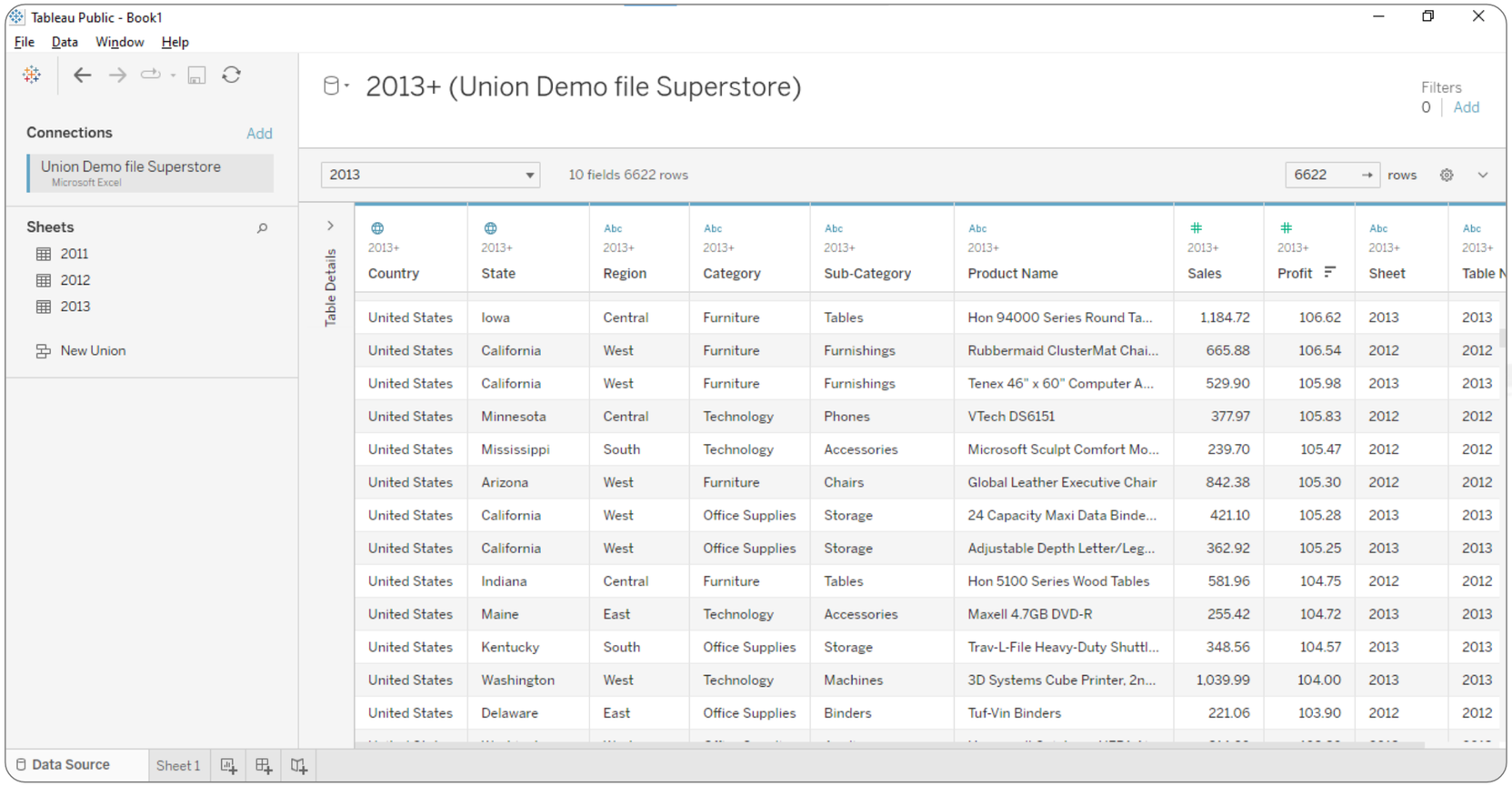
Steps to Perform Union

If the union is performed, a symbol will appear next to the **2013** table.



Steps to Perform Union

The final output after performing the union will be shown as:

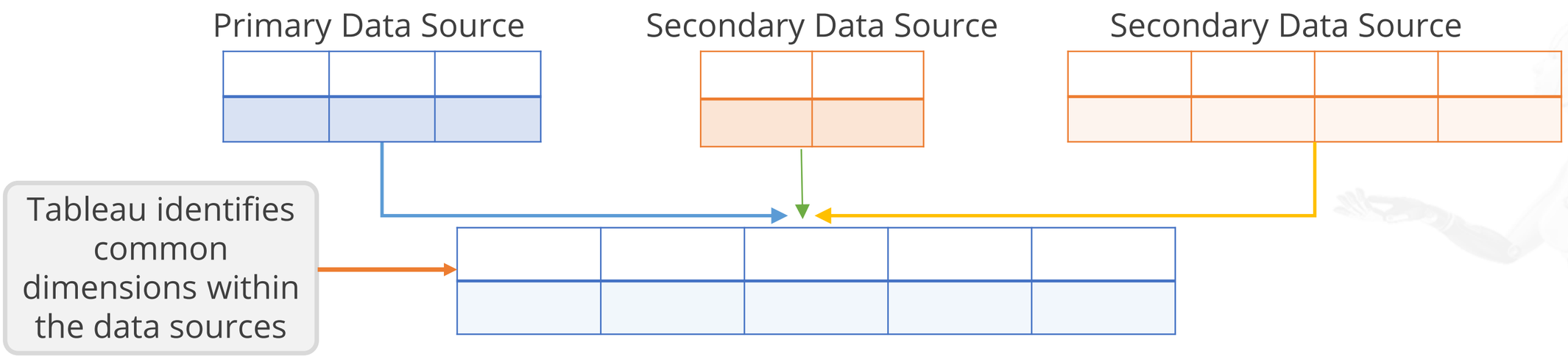


Country	State	Region	Category	Sub-Category	Product Name	Sales	Profit	Sheet	Table N
United States	Iowa	Central	Furniture	Tables	Hon 94000 Series Round Ta...	1,184.72	106.62	2013	2013
United States	California	West	Furniture	Furnishings	Rubbermaid ClusterMat Chai...	665.88	106.54	2012	2012
United States	California	West	Furniture	Furnishings	Tenex 46" x 60" Computer A...	529.90	105.98	2013	2013
United States	Minnesota	Central	Technology	Phones	VTech DS6151	377.97	105.83	2012	2012
United States	Mississippi	South	Technology	Accessories	Microsoft Sculpt Comfort Mo...	239.70	105.47	2012	2012
United States	Arizona	West	Furniture	Chairs	Global Leather Executive Chair	842.38	105.30	2012	2012
United States	California	West	Office Supplies	Storage	24 Capacity Maxi Data Binde...	421.10	105.28	2013	2013
United States	California	West	Office Supplies	Storage	Adjustable Depth Letter/Leg...	362.92	105.25	2013	2013
United States	Indiana	Central	Furniture	Tables	Hon 5100 Series Wood Tables	581.96	104.75	2012	2012
United States	Maine	East	Technology	Accessories	Maxell 4.7GB DVD-R	255.42	104.72	2013	2013
United States	Kentucky	South	Office Supplies	Storage	Trav-L-File Heavy-Duty Shuttl...	348.56	104.57	2013	2013
United States	Washington	West	Technology	Machines	3D Systems Cube Printer, 2n...	1,039.99	104.00	2013	2013
United States	Delaware	East	Office Supplies	Binders	Tuf-Vin Binders	221.06	103.90	2012	2012

Data Blending

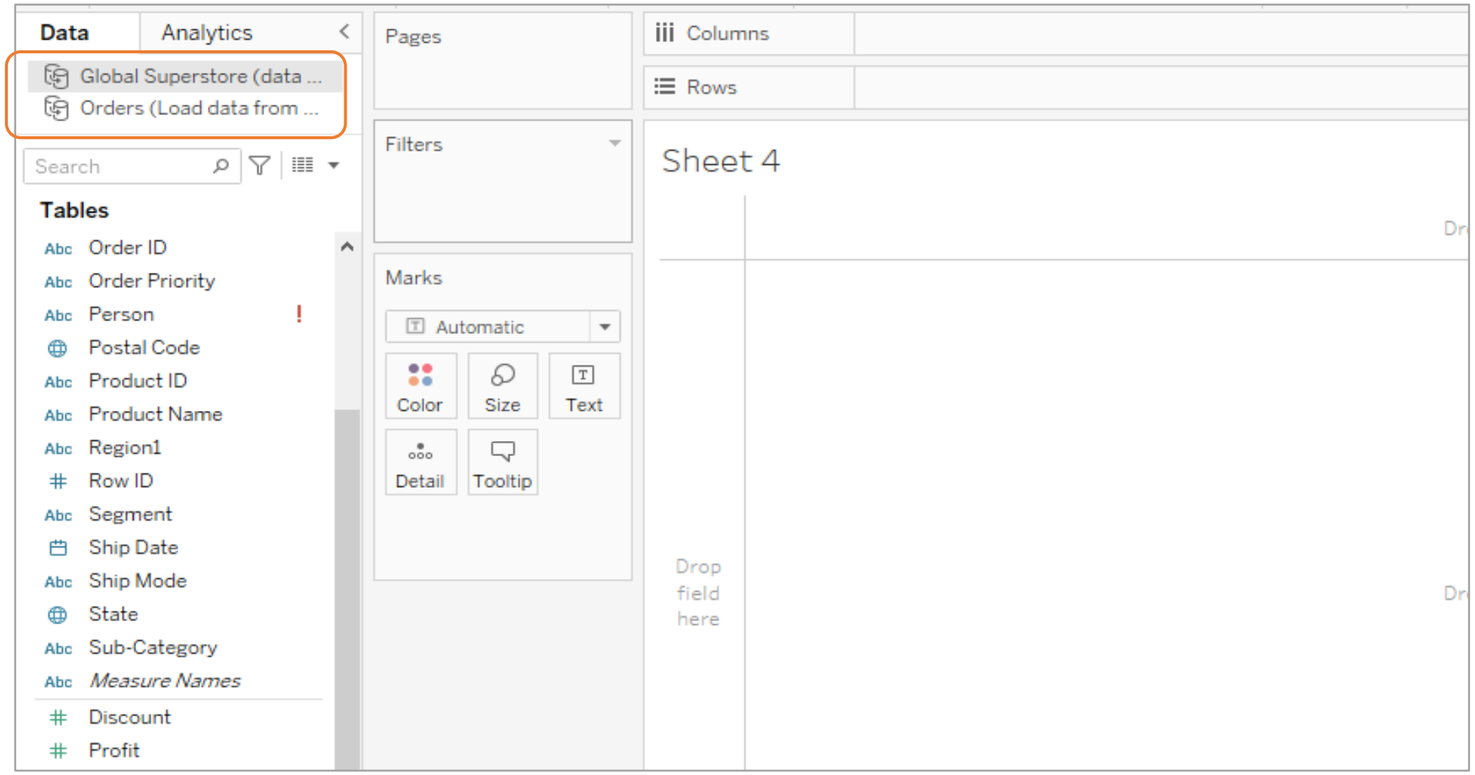
Data Blending

Blending is a method of combining related data from multiple sources in a single view to analyze it.



Steps to Perform Data Blending

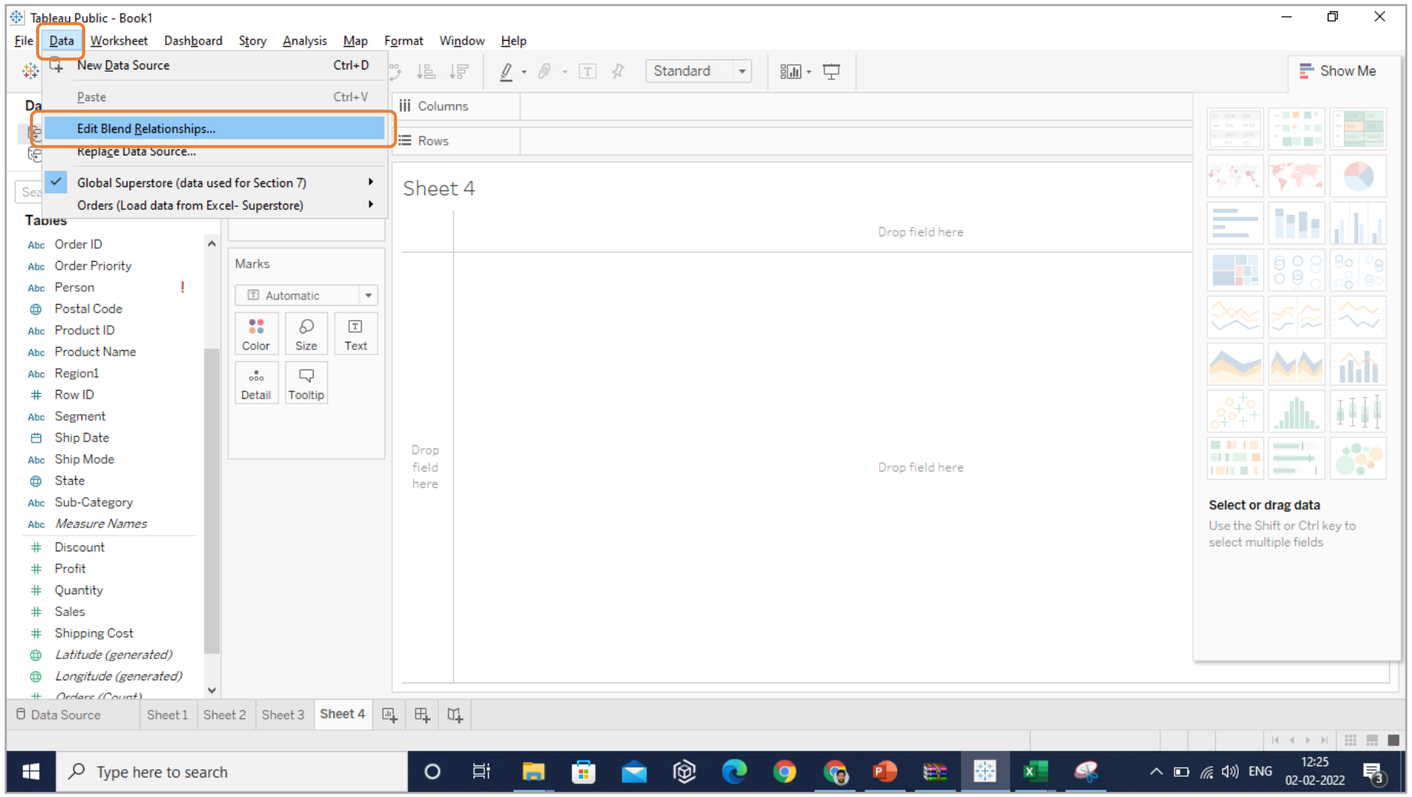
Step 1: Connect the datasets **Load data from Excel-Superstore** and **Global Superstore** to Tableau



Use the **orders** table from both the datasets

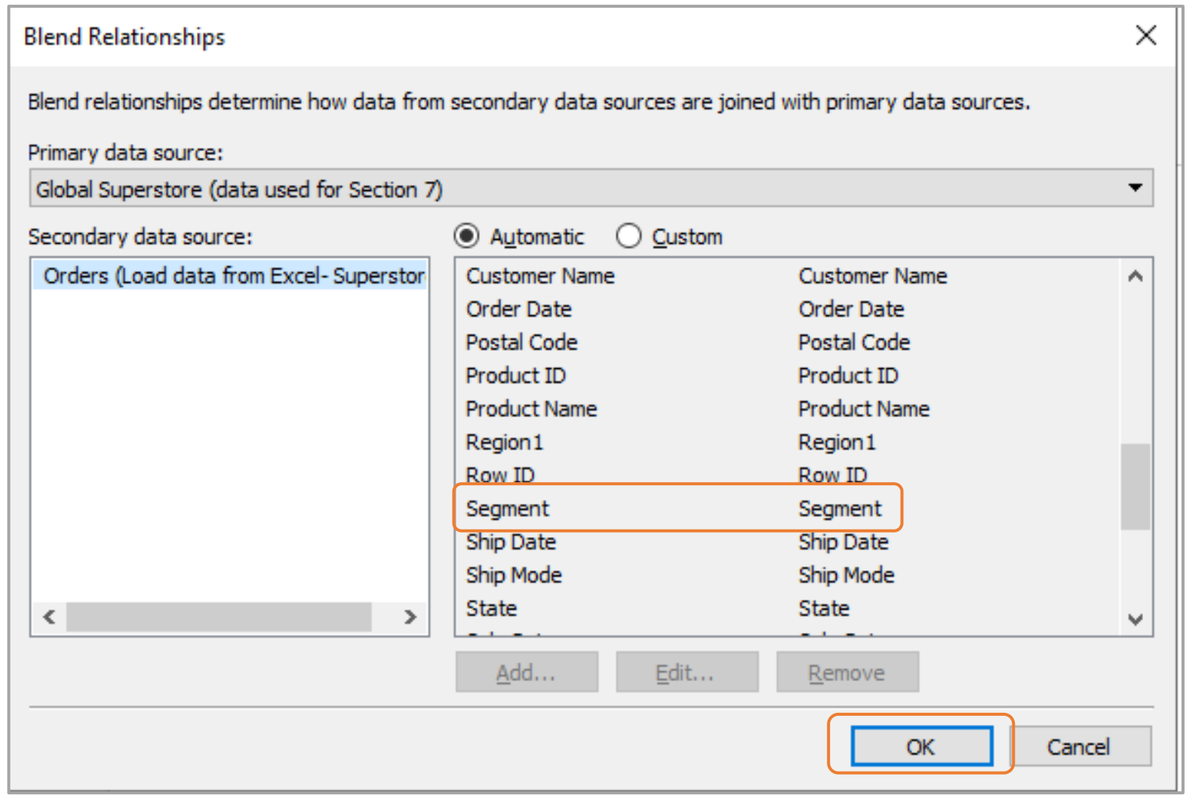
Steps to Perform Data Blending

Step 2: Click on the **Data** tab and select **Edit Blend Relationship**



Steps to Perform Data Blending

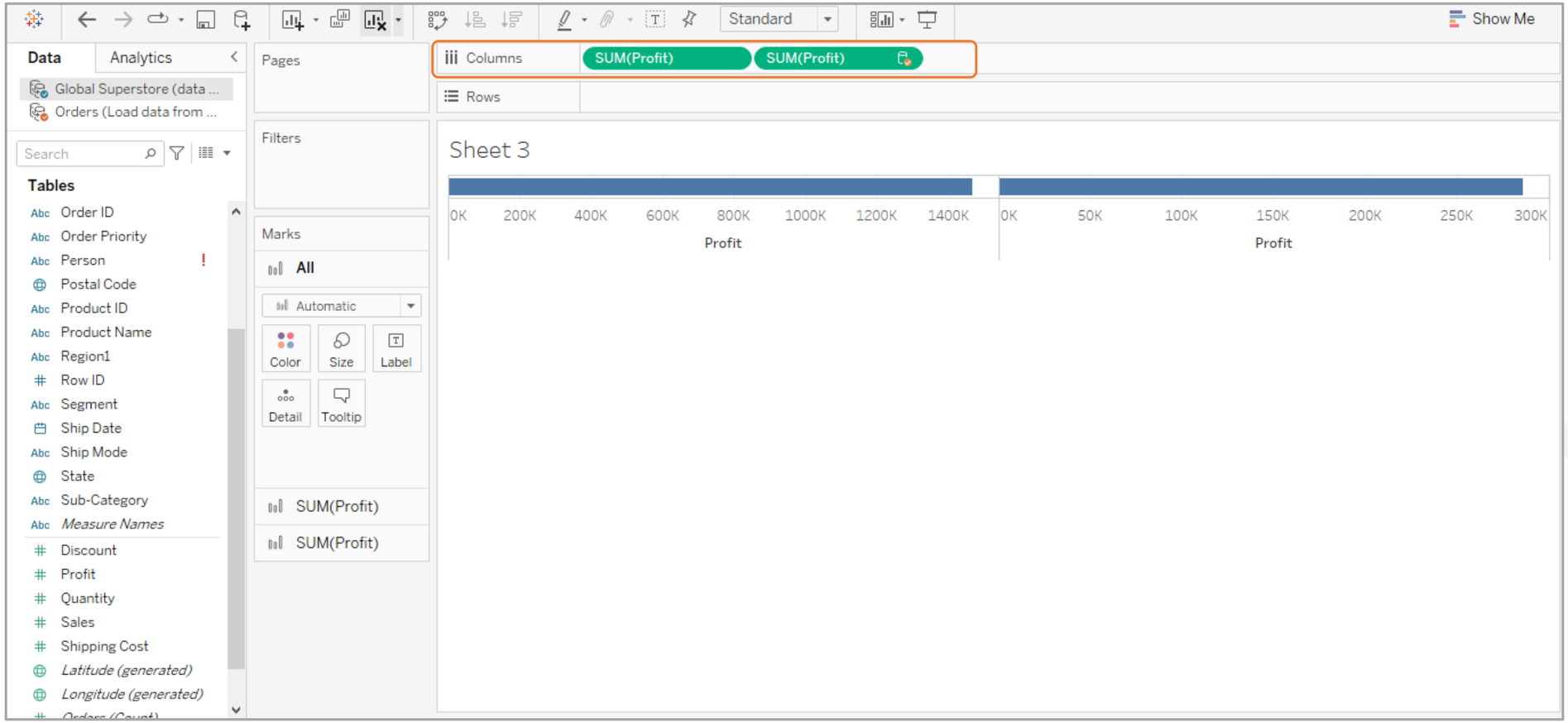
Step 3: In the **Edit Blend Relationship** dialog box, ensure that **Segment** is mapped with **Segment**



Click on **OK**

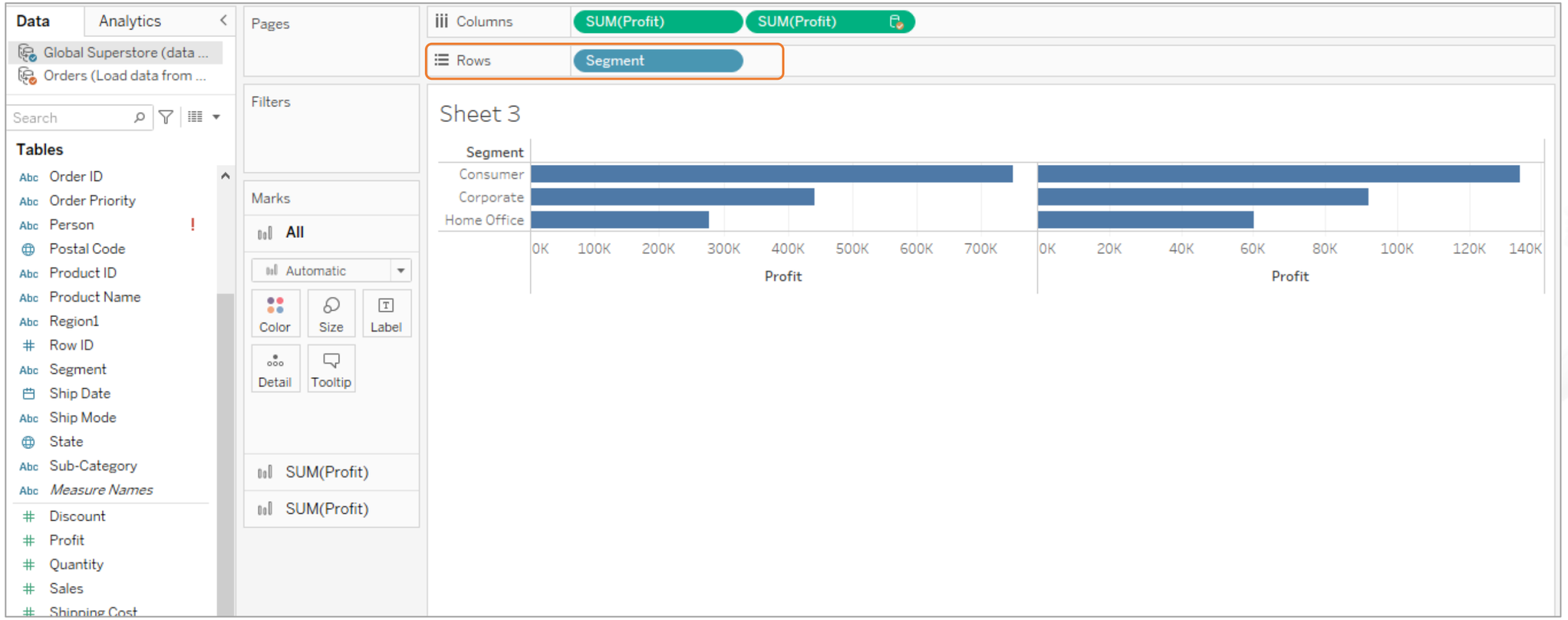
Steps to Perform Data Blending

Step 4: Drag **Profit** from both the datasets to **Columns**



Steps to Perform Data Blending

Step 5: Drag **Segment** from Global Superstore dataset to **Rows**



Now we can see segment wise profit from both the datasets with the help of data blending

Blend vs. Join

When to use a Blend:

- Blend is used to combine data from different databases that do not support cross-table joins.
- Data within different sources are at different levels of detail.
- Joins result in duplicate rows.
- It involves a large amount of data.

When to use a Join:

- The data format is consistent across all sources.
- It involves relatively small amounts of data.
- Data within each source is at the same level of detail.

DATA AND
ARTIFICIAL INTELLIGENCE

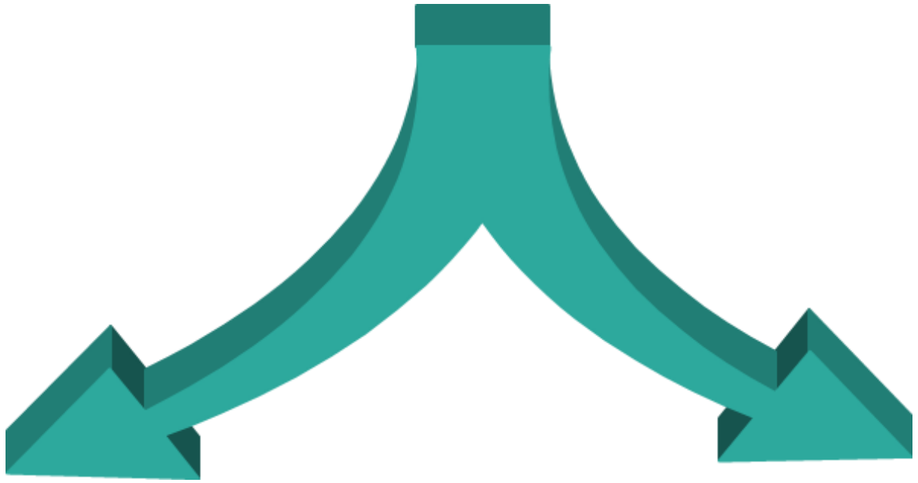
Aggregations

Aggregations

Aggregation is the process of converting a set of values into a single value.

Aggregations

Data aggregation is performed on:

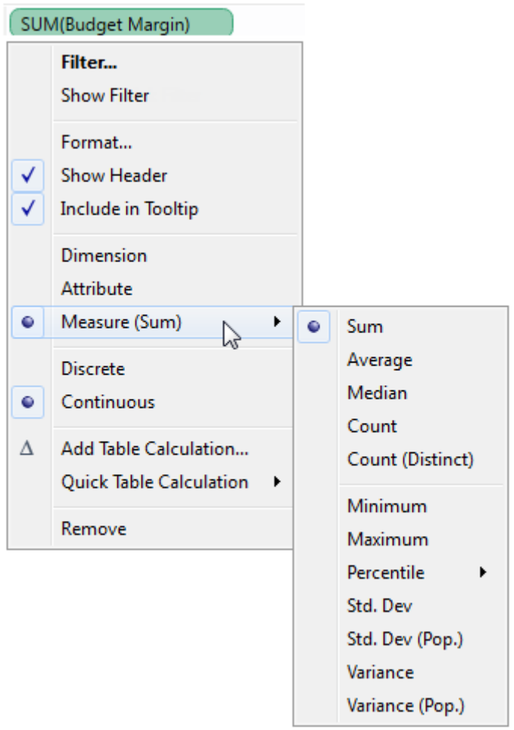


Measures

Dimensions

Aggregating Measures

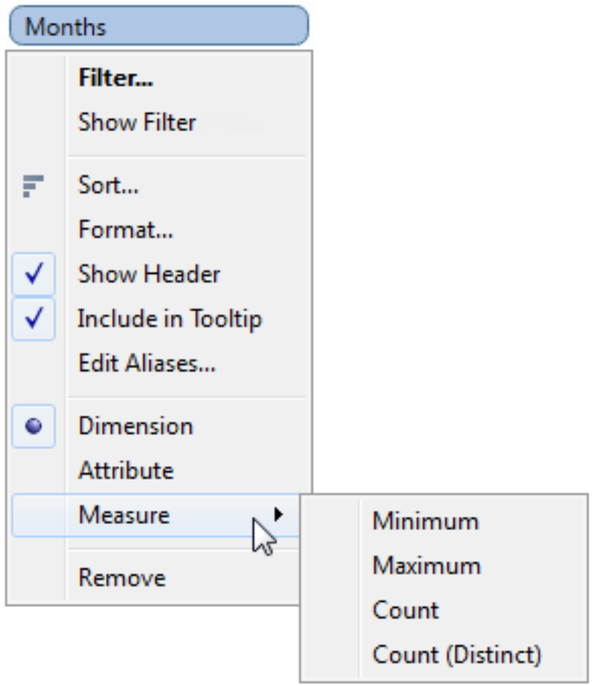
The aggregation of measures automatically aggregates their values when a measure is added to the view.



Measures can be aggregated only for relational data sources in Tableau.

Aggregating Dimensions

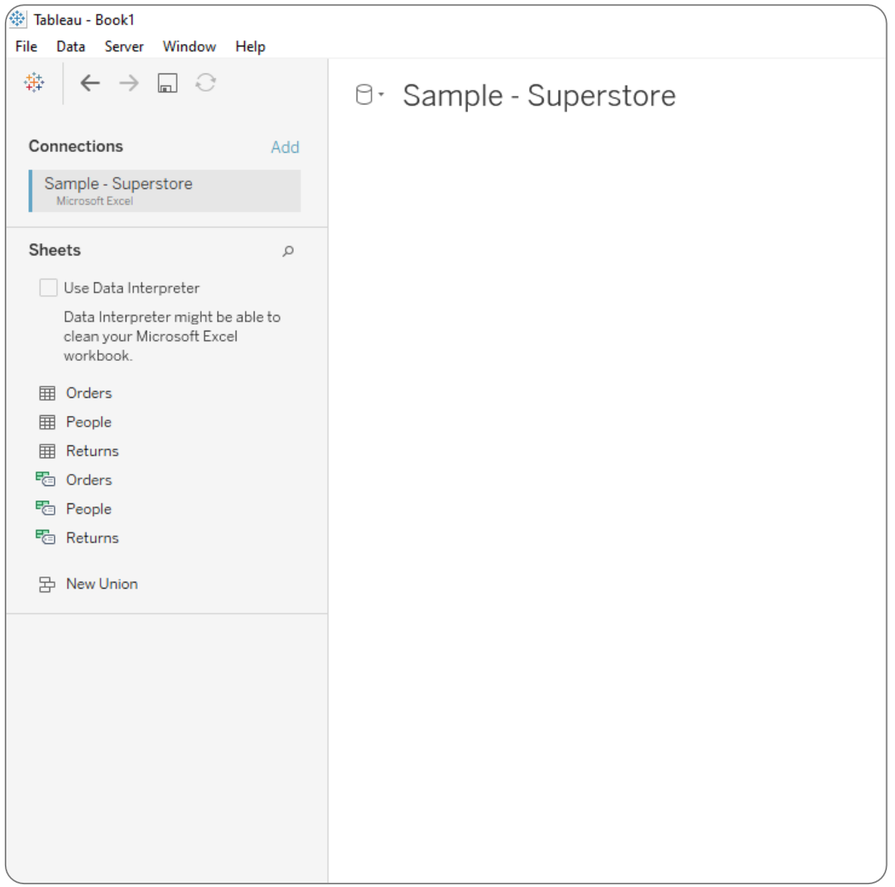
The aggregation of dimensions creates a new temporary measure column.



Dimensions can be aggregated as **Minimum**, **Maximum**, **Count**, or **Count (Distinct)**.

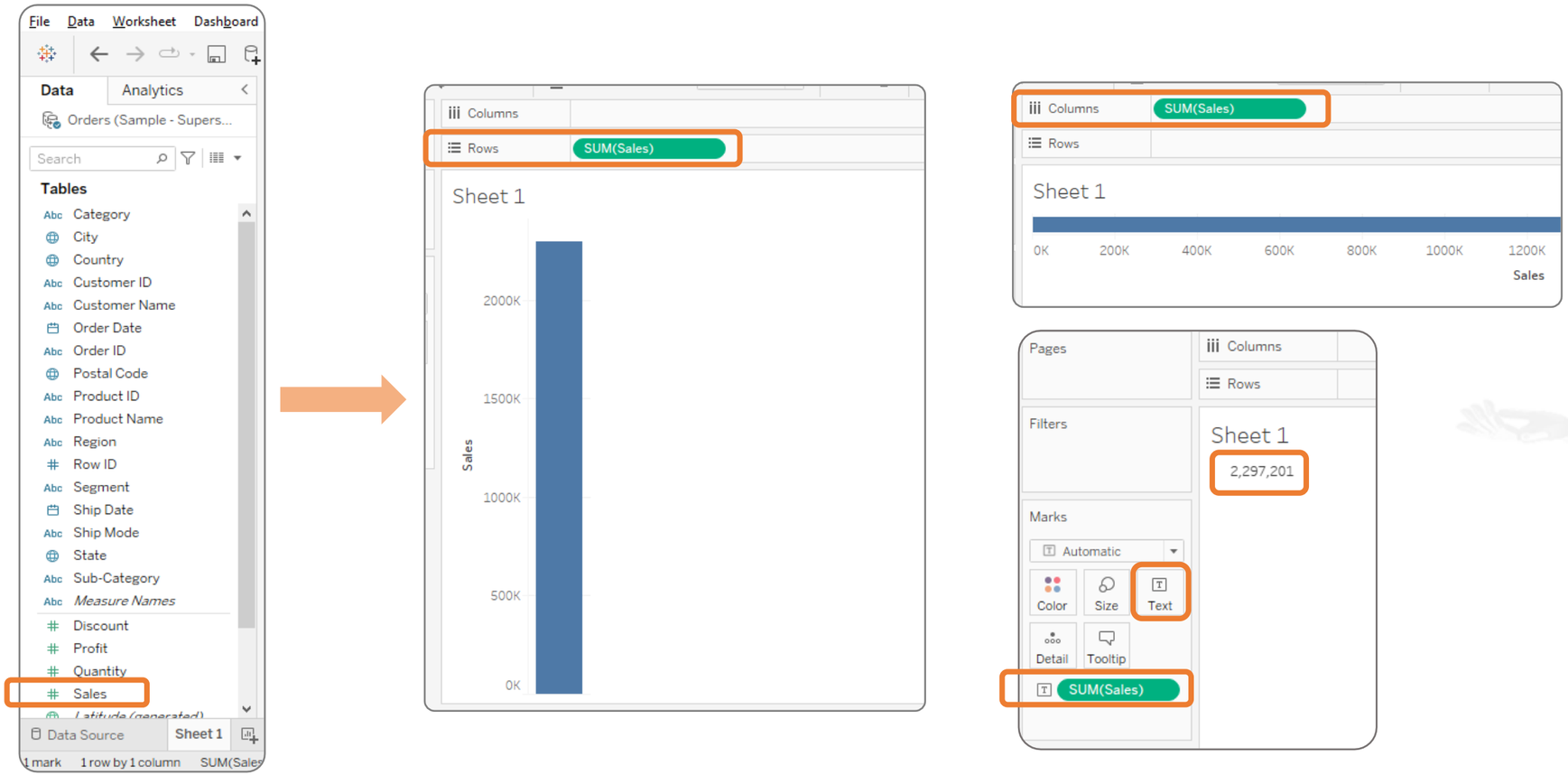
Steps to Perform Aggregations

Let us understand this better with the help of an example using **Sample-Superstore** dataset:



Steps to Perform Aggregations

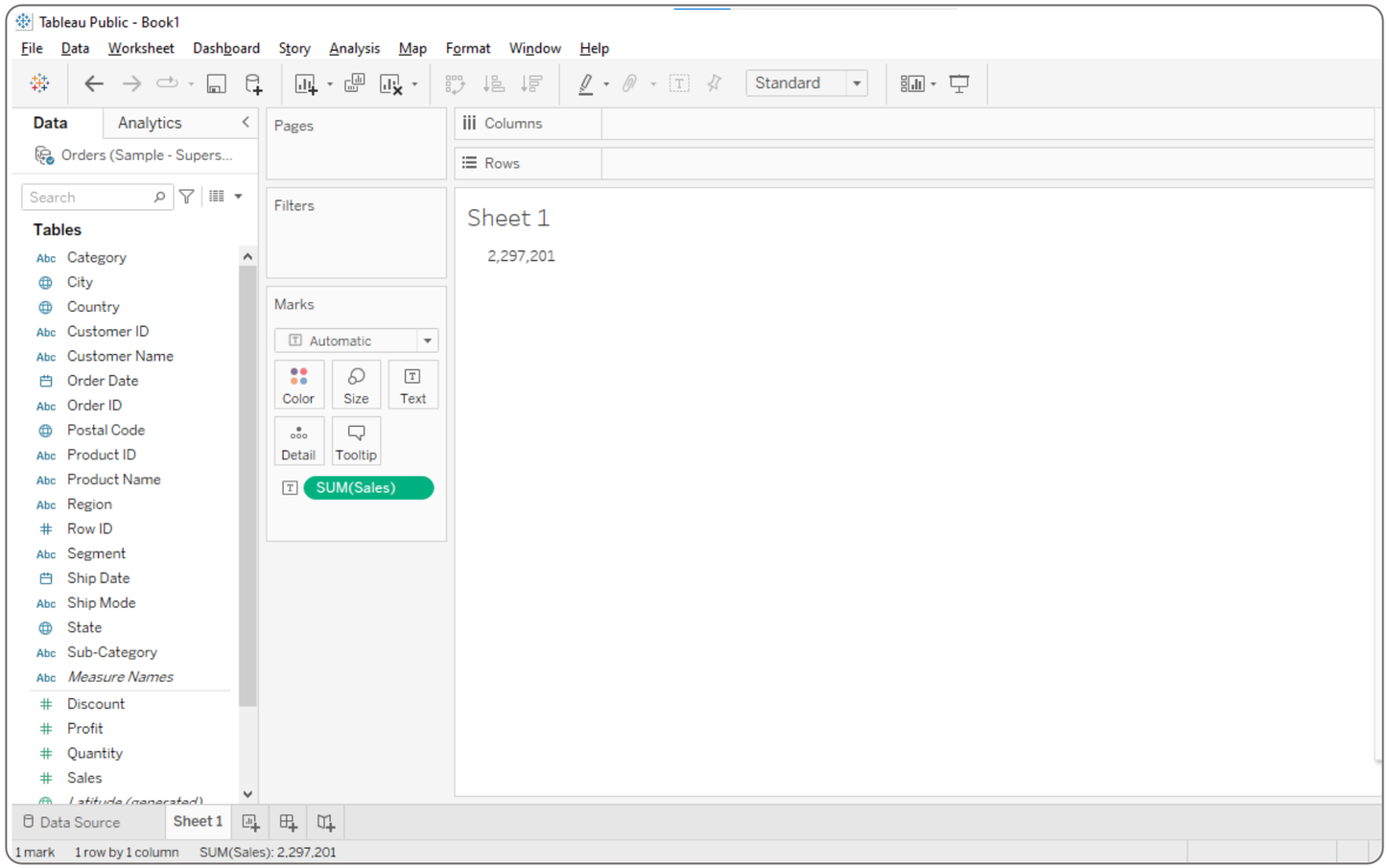
Aggregation can be performed by dragging **Sales** to **Rows**, **Columns,** or **Text** under **Marks**.



The default aggregation is **SUM** and will not change if the sales are dragged to rows, columns, or marks.

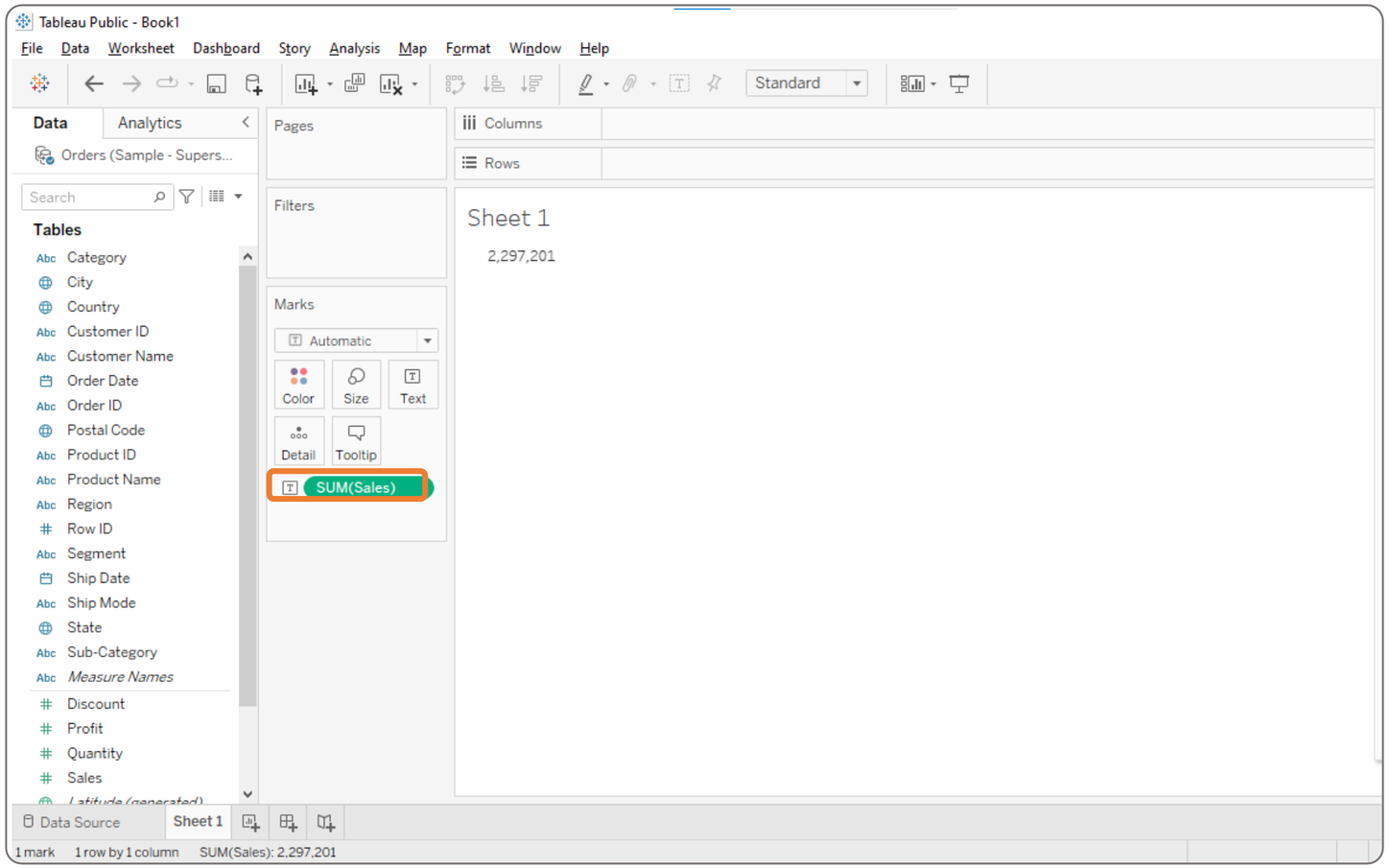
Steps to Perform Aggregations

Let’s learn how to change the aggregation type of any measure.



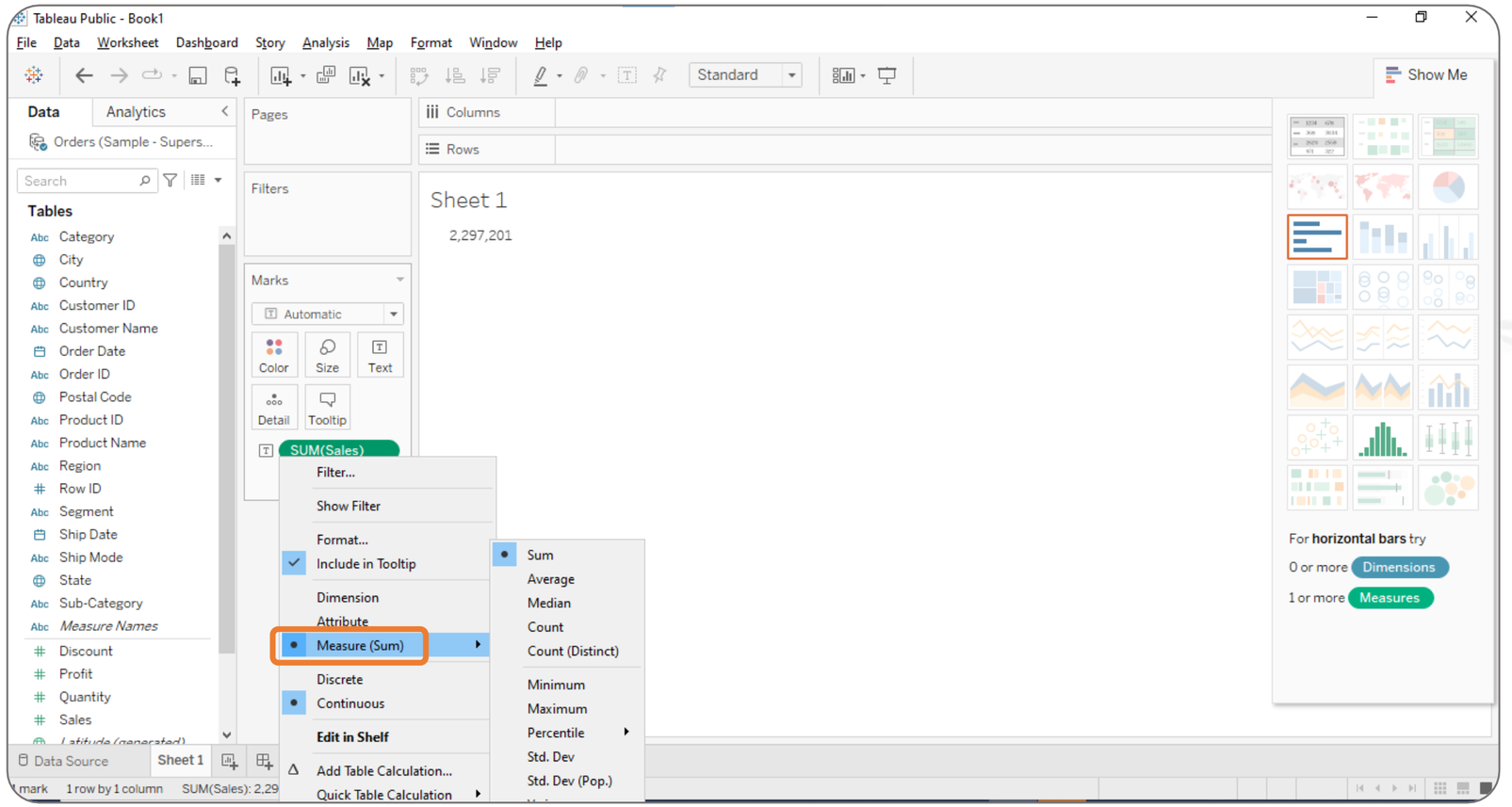
Steps to Perform Aggregations

Step 1: Right-click on **SUM(Sales)**



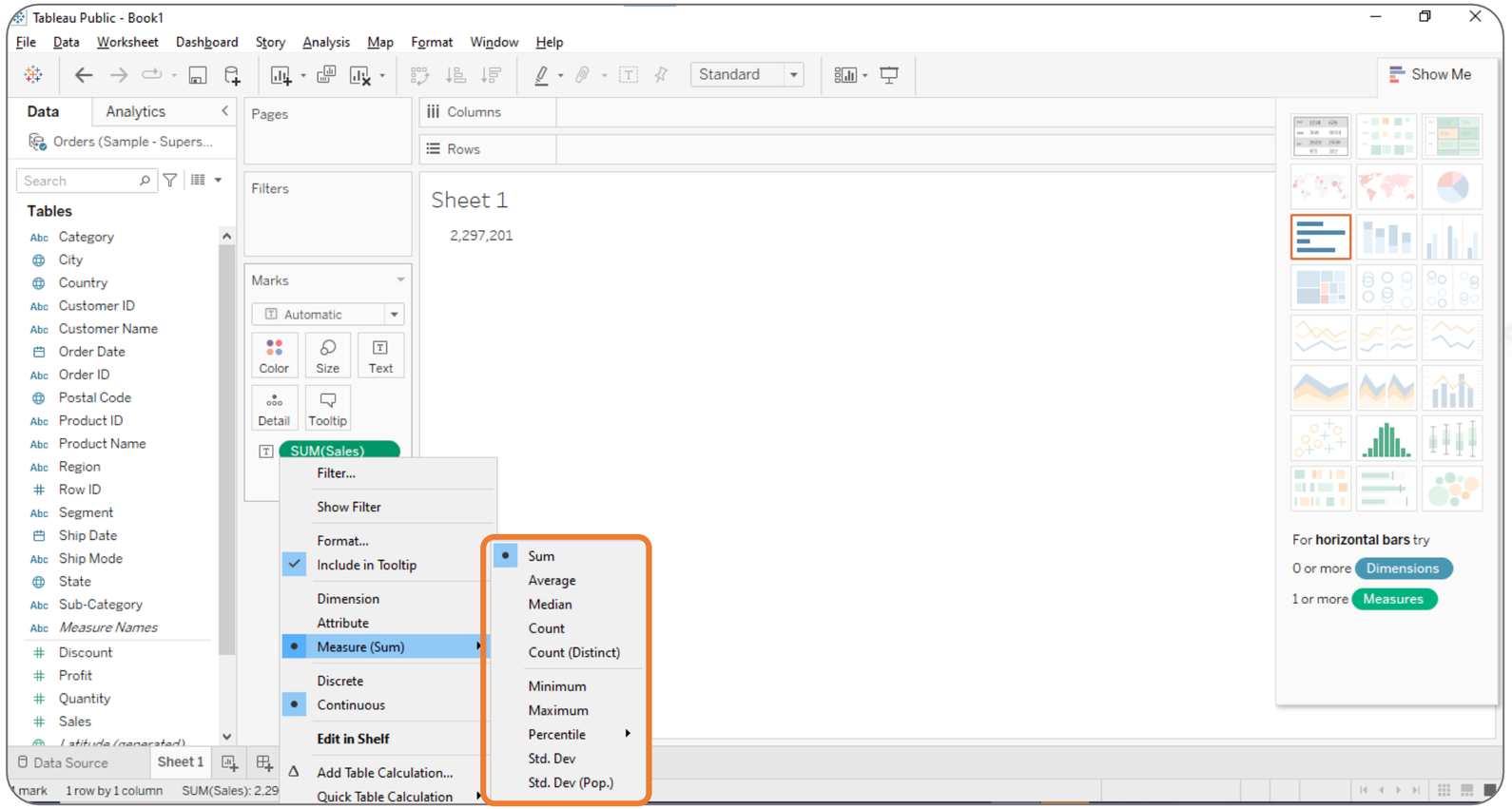
Steps to Perform Aggregations

Step 2: Click on **Measure (Sum)**



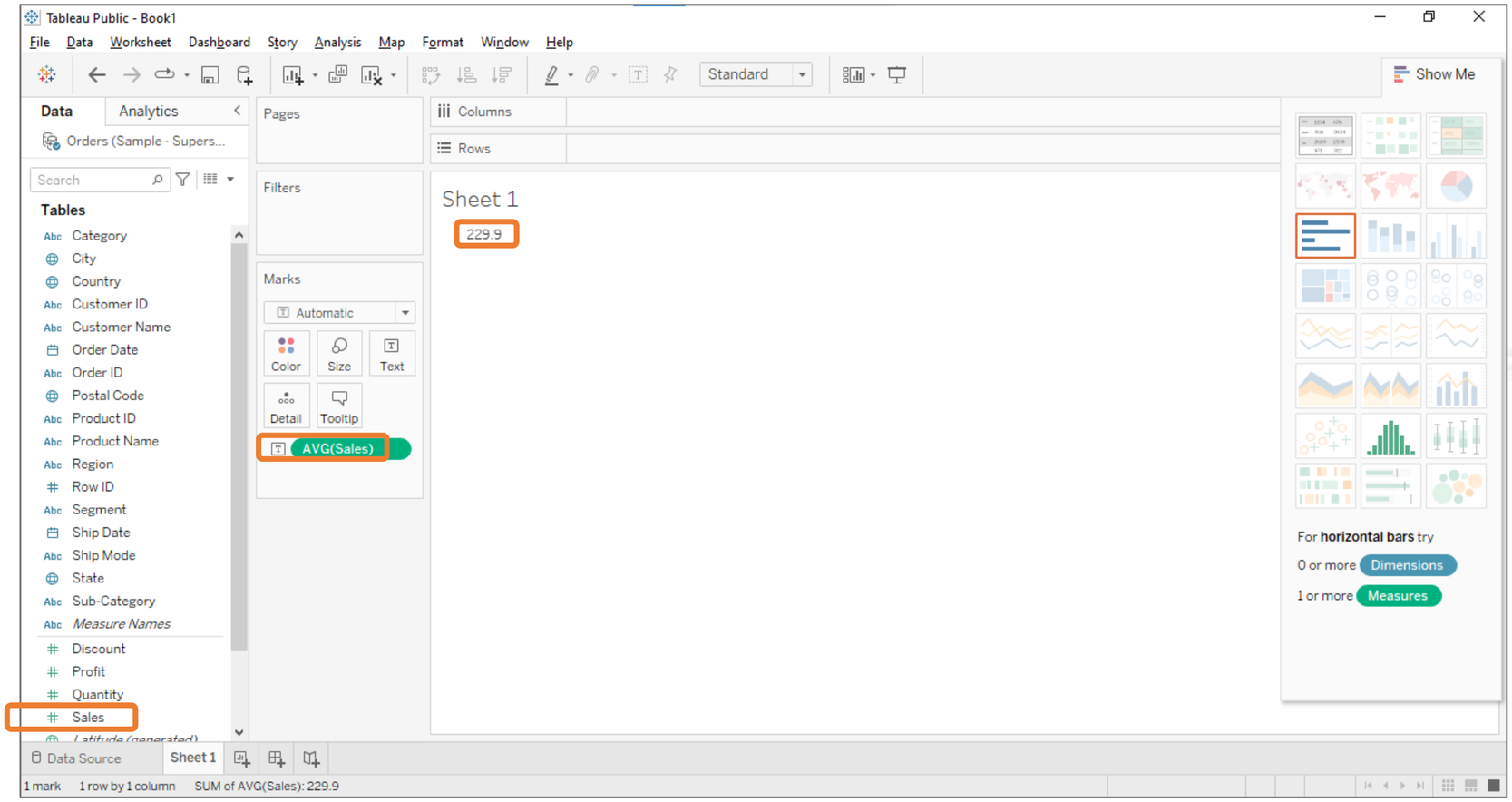
Steps to Perform Aggregations

Step 3: Select the desired aggregation from **Measure (Sum)** list



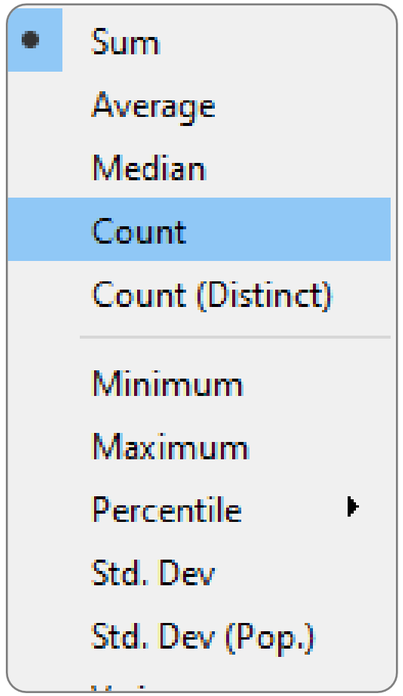
Steps to Perform Aggregations

It will be changed to the desired aggregation.



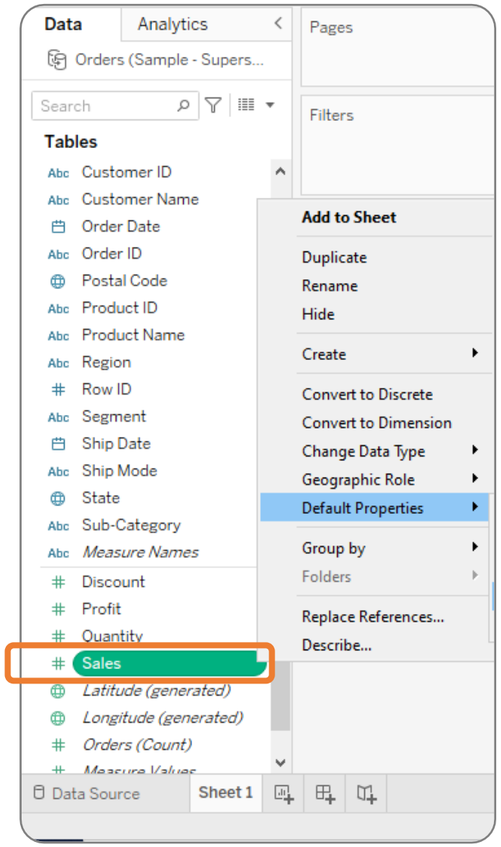
Steps to Perform Aggregations

Let’s learn how to change the default aggregation of any measure.



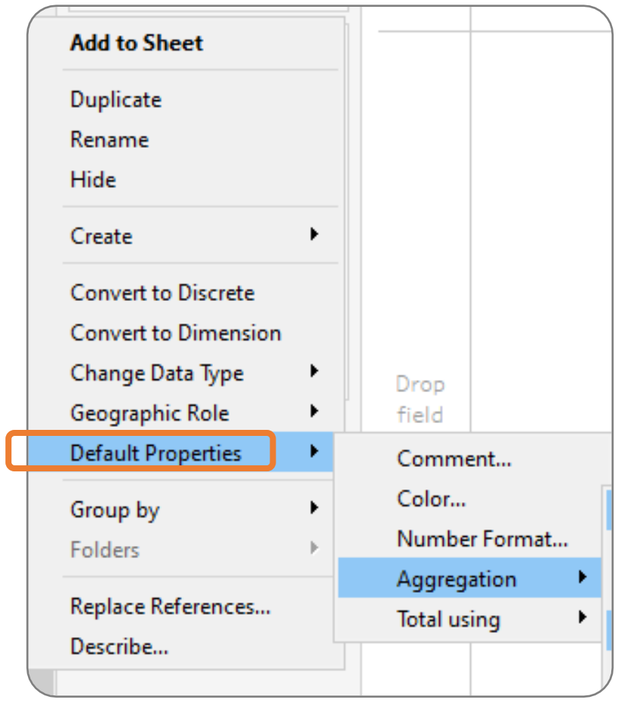
Steps to Perform Aggregations

Step 1: Click on the **Sales**



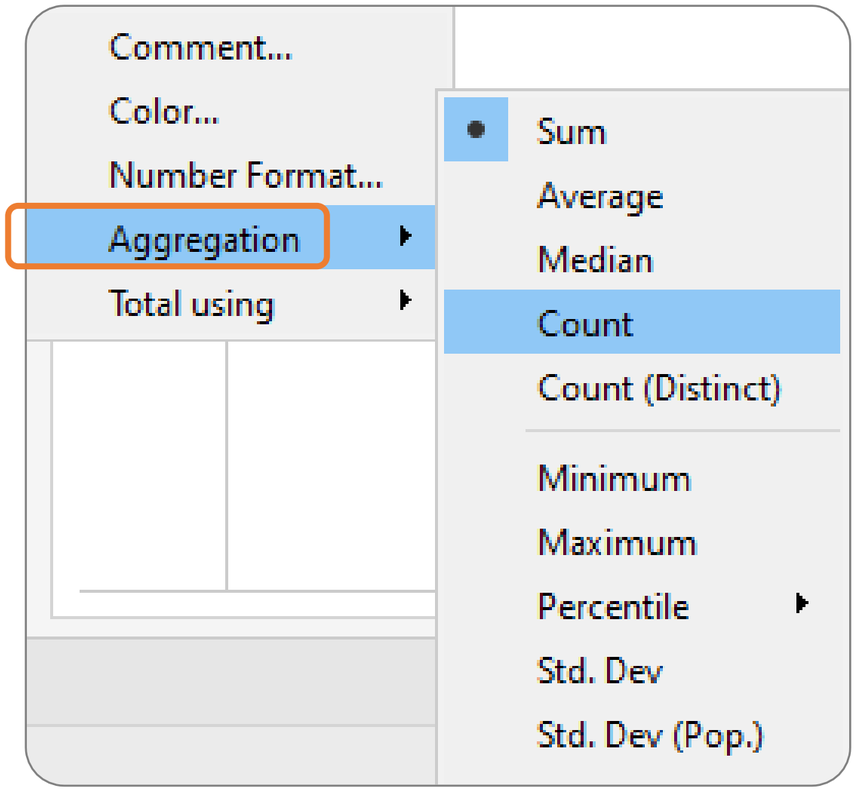
Steps to Perform Aggregations

Step 2: Click on the **Default Properties**



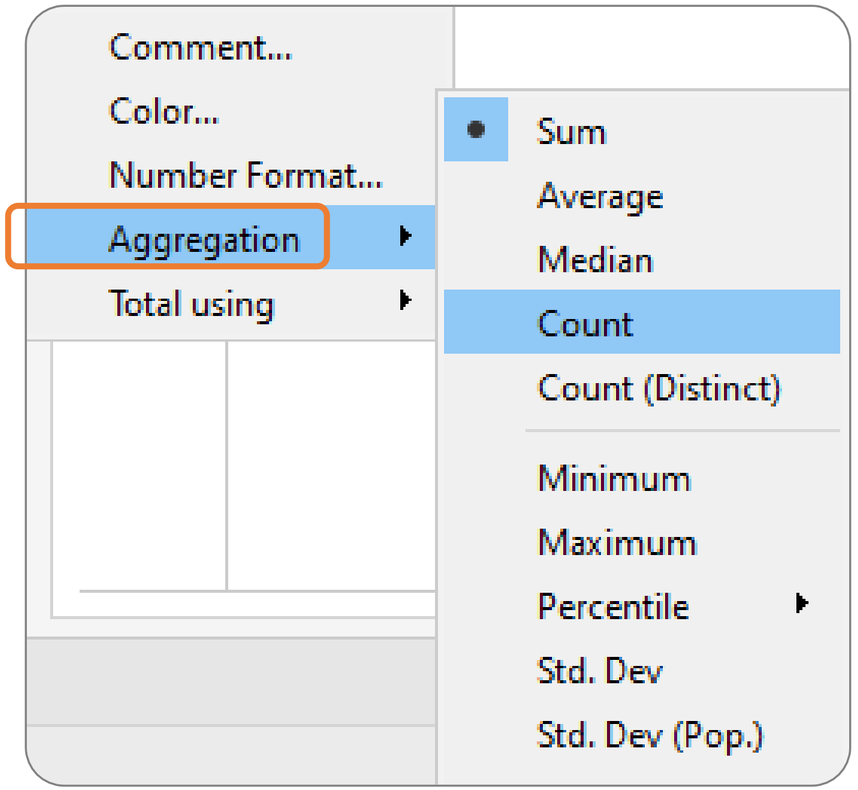
Steps to Perform Aggregations

Step 3: Click on **Aggregation**



Steps to Perform Aggregations

Step 4: Select the desired aggregation to make it default



Pivot

Pivot

The pivot function in Tableau allows selecting the columns that need to be manipulated and formatting them into a typical dataset ready for analysis.

Pivot

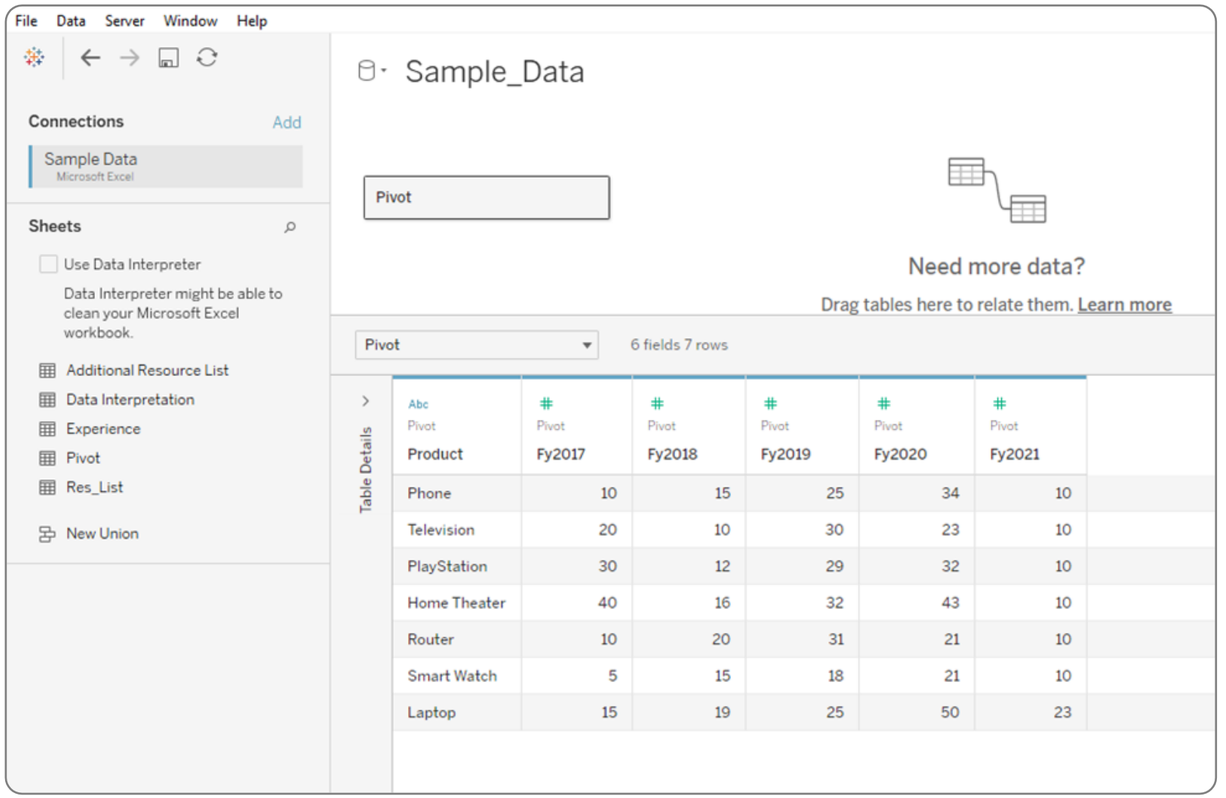
BI tool prefers data to be tall instead of wide, which means that one must pivot data from columns to rows so that the tool evaluates the data properly.

Abc Pivot Product	# Pivot Fy2017	# Pivot Fy2018	# Pivot Fy2019	# Pivot Fy2020	# Pivot Fy2021
Phone	10	15	25	34	10
Television	20	10	30	23	10
PlayStation	30	12	29	32	10
Home Theater	40	16	32	43	10
Router	10	20	31	21	10
Smart Watch	5	15	18	21	10
Laptop	15	19	25	50	23

Abc Pivot Product	Abc Pivot1 Pivot Field Names	# Pivot1 Pivot Field Values
Phone	Fy2017	10
Phone	Fy2018	15
Phone	Fy2019	25
Phone	Fy2020	34
Phone	Fy2021	10
Television	Fy2017	20
Television	Fy2018	10
Television	Fy2019	30
Television	Fy2020	23

Steps to Perform Pivot

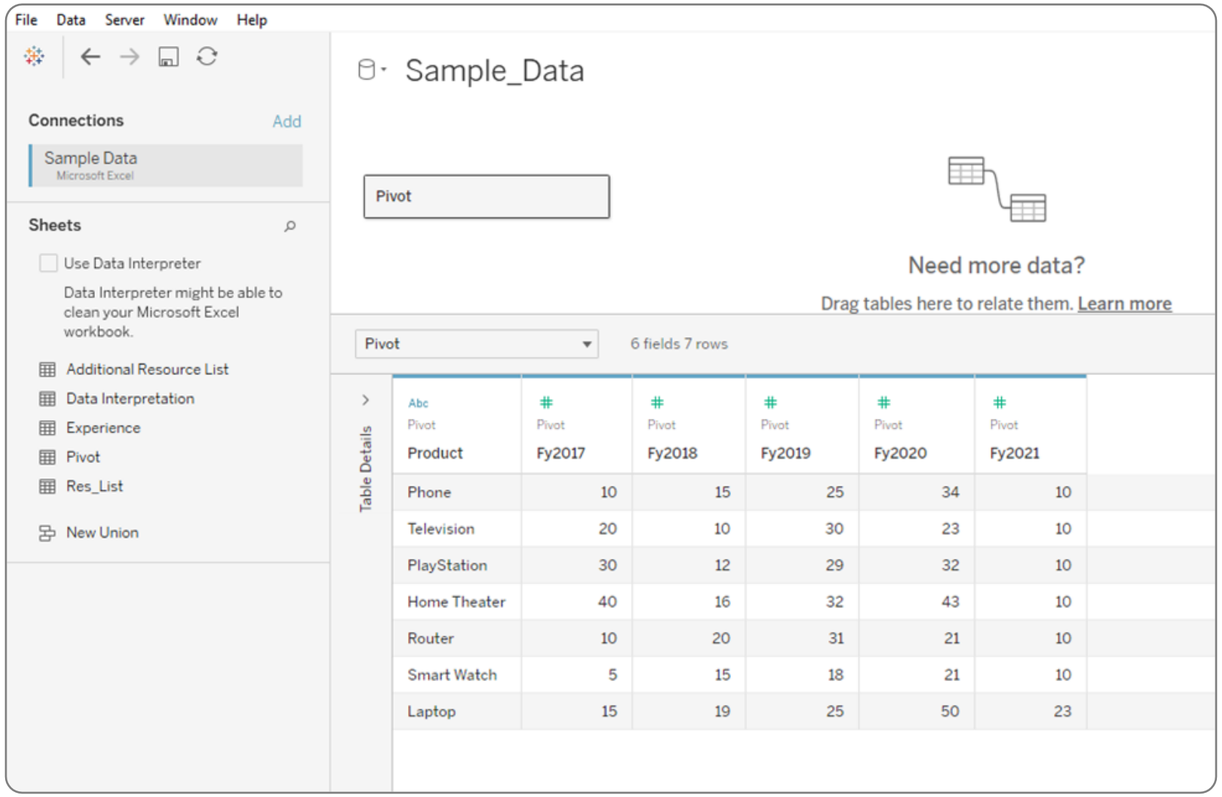
Let us understand this with the help of an example. The sales quantity for each year is updated in separate columns across each product in the file below:



Product	Fy2017	Fy2018	Fy2019	Fy2020	Fy2021
Phone	10	15	25	34	10
Television	20	10	30	23	10
PlayStation	30	12	29	32	10
Home Theater	40	16	32	43	10
Router	10	20	31	21	10
Smart Watch	5	15	18	21	10
Laptop	15	19	25	50	23

Steps to Perform Pivot

Though the data looks fine, the format is not so appropriate for Tableau to evaluate all items.



Abc Pivot Product	# Pivot Fy2017	# Pivot Fy2018	# Pivot Fy2019	# Pivot Fy2020	# Pivot Fy2021
Phone	10	15	25	34	10
Television	20	10	30	23	10
PlayStation	30	12	29	32	10
Home Theater	40	16	32	43	10
Router	10	20	31	21	10
Smart Watch	5	15	18	21	10
Laptop	15	19	25	50	23

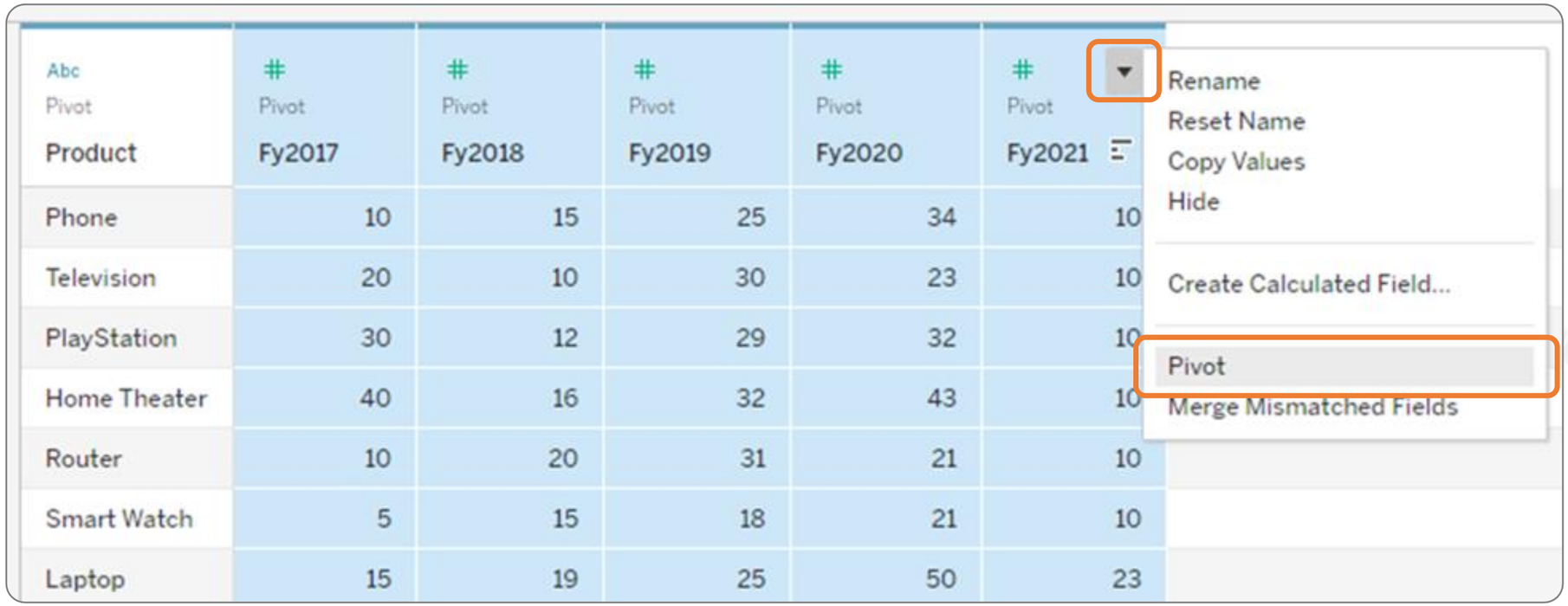
Steps to Perform Pivot

Step 1: Select all the columns in the file

Abc Pivot Product	# Pivot Fy2017	# Pivot Fy2018	# Pivot Fy2019	# Pivot Fy2020	# Pivot Fy2021
Phone	10	15	25	34	10
Television	20	10	30	23	10
PlayStation	30	12	29	32	10
Home Theater	40	16	32	43	10
Router	10	20	31	21	10
Smart Watch	5	15	18	21	10
Laptop	15	19	25	50	23

Steps to Perform Pivot

Step 2: Select **Pivot** from the pop-up menu



Abc Pivot Product	# Pivot Fy2017	# Pivot Fy2018	# Pivot Fy2019	# Pivot Fy2020	# Pivot Fy2021
Phone	10	15	25	34	10
Television	20	10	30	23	10
PlayStation	30	12	29	32	10
Home Theater	40	16	32	43	10
Router	10	20	31	21	10
Smart Watch	5	15	18	21	10
Laptop	15	19	25	50	23

Steps to Perform Pivot

The final output after pivoting will be shown as:

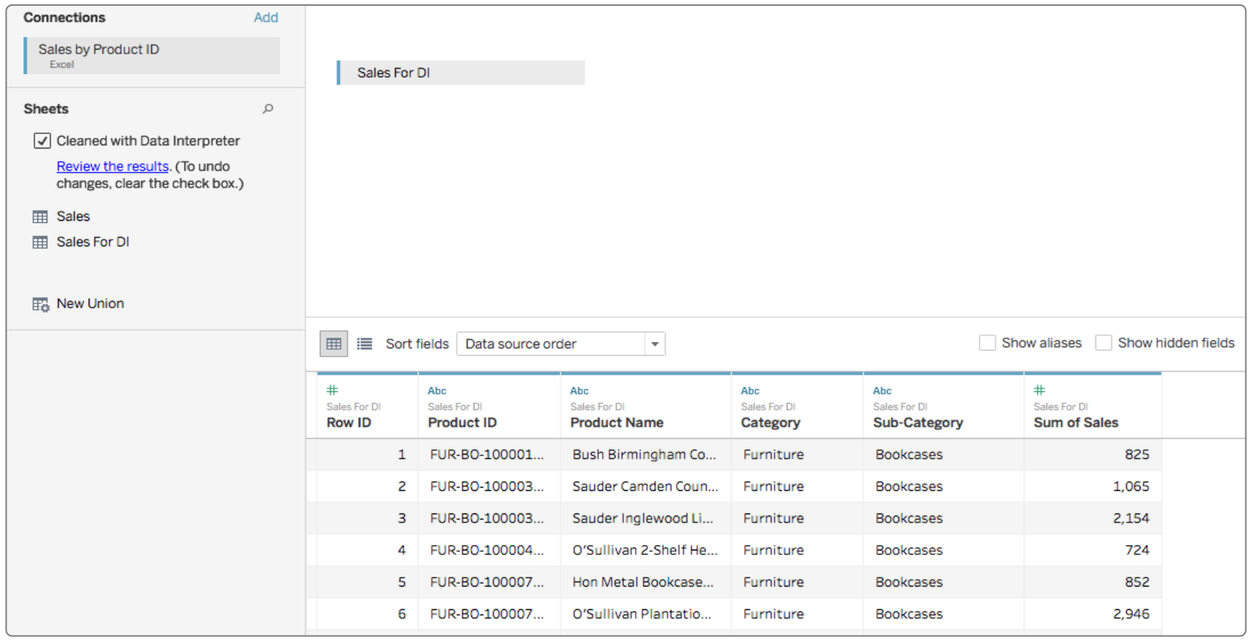
Abc Pivot Product	Abc Pivot1 Pivot Field Names	# Pivot1 Pivot Field Values
Phone	Fy2017	10
Phone	Fy2018	15
Phone	Fy2019	25
Phone	Fy2020	34
Phone	Fy2021	10
Television	Fy2017	20
Television	Fy2018	10
Television	Fy2019	30
Television	Fy2020	23

This is the most appropriate format for Tableau to evaluate all items.

Data Interpreter

Data Interpreter

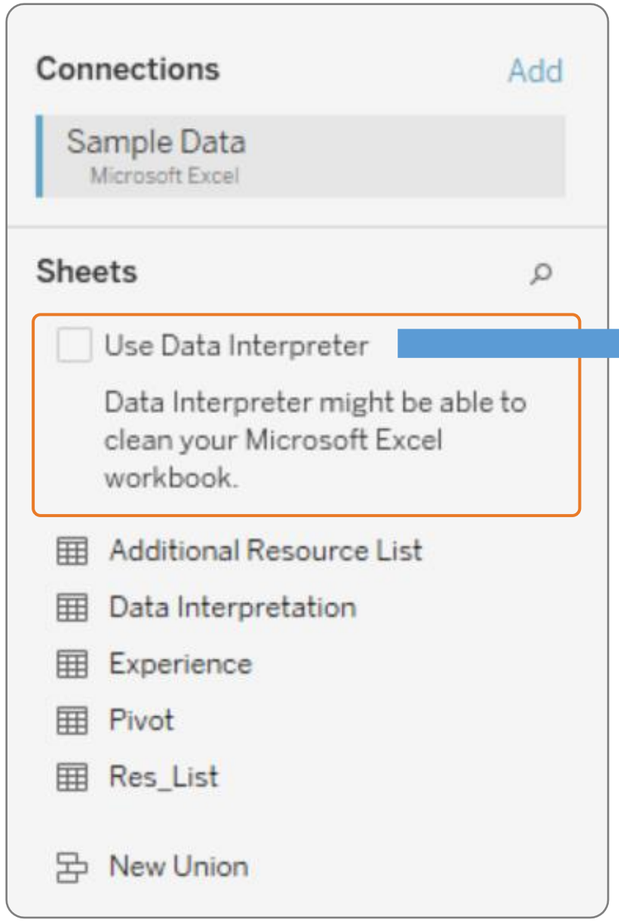
A data interpreter automatically cleans and prepares data for analysis.



It can detect titles, notes, footers, and empty cells and bypass them to identify the actual fields and values in a dataset.

Data Interpreter

Tableau activates the data interpreter option whenever a unique format or extraneous information is detected.



Click on the data interpreter function to remove extraneous information

Data Interpreter

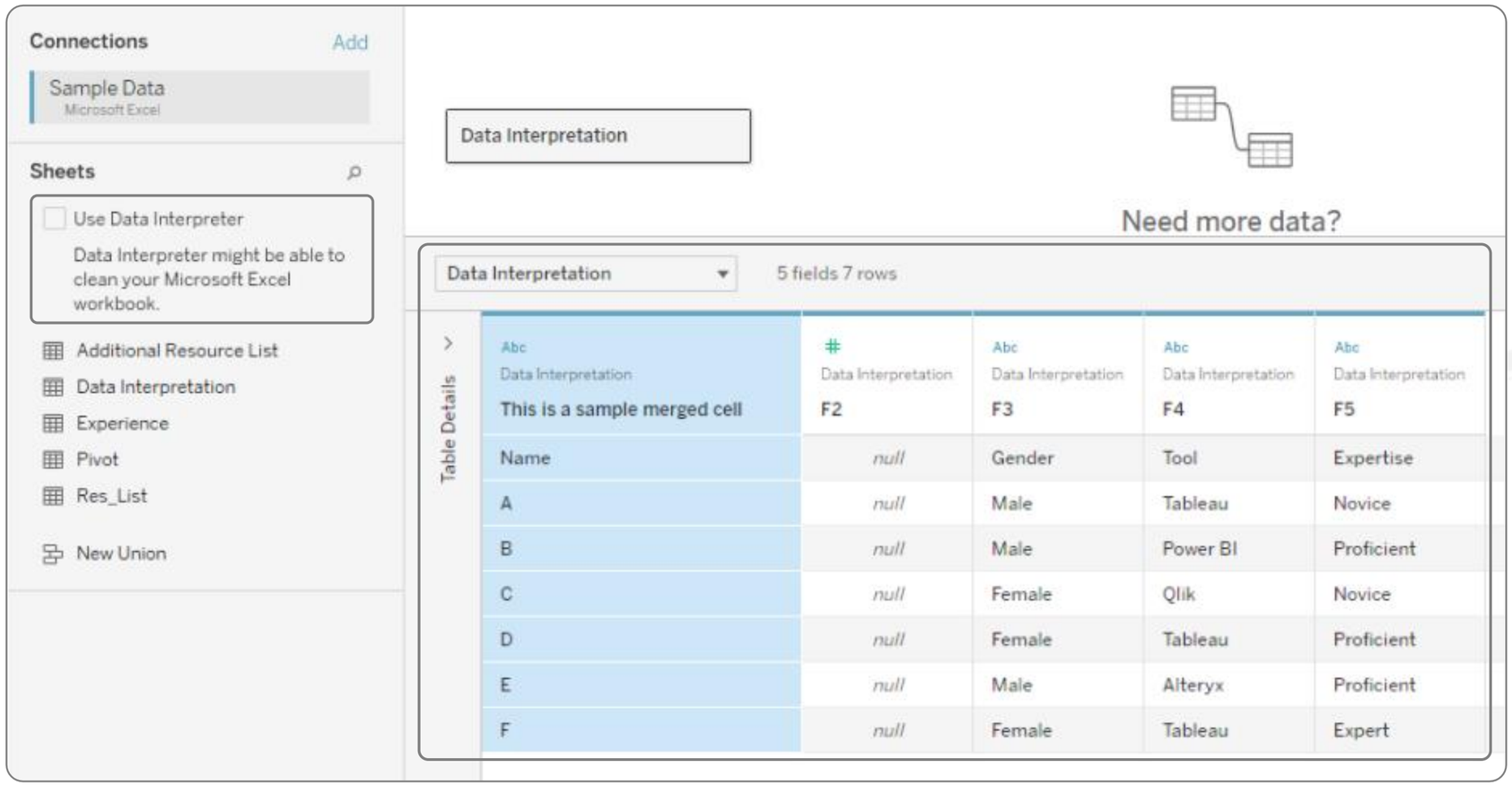
Consider the given table with merged cells, blank rows, and blank columns:

	A	B	C	D	E
1	This is a sample merged cell				
2					
3	Name		Gender	Tool	Expertise
4	A		Male	Tableau	Novice
5	B		Male	Power BI	Proficient
6	C		Female	Qlik	Novice
7	D		Female	Tableau	Proficient
8	E		Male	Alteryx	Proficient
9	F		Female	Tableau	Expert

Poses many challenges during visualization

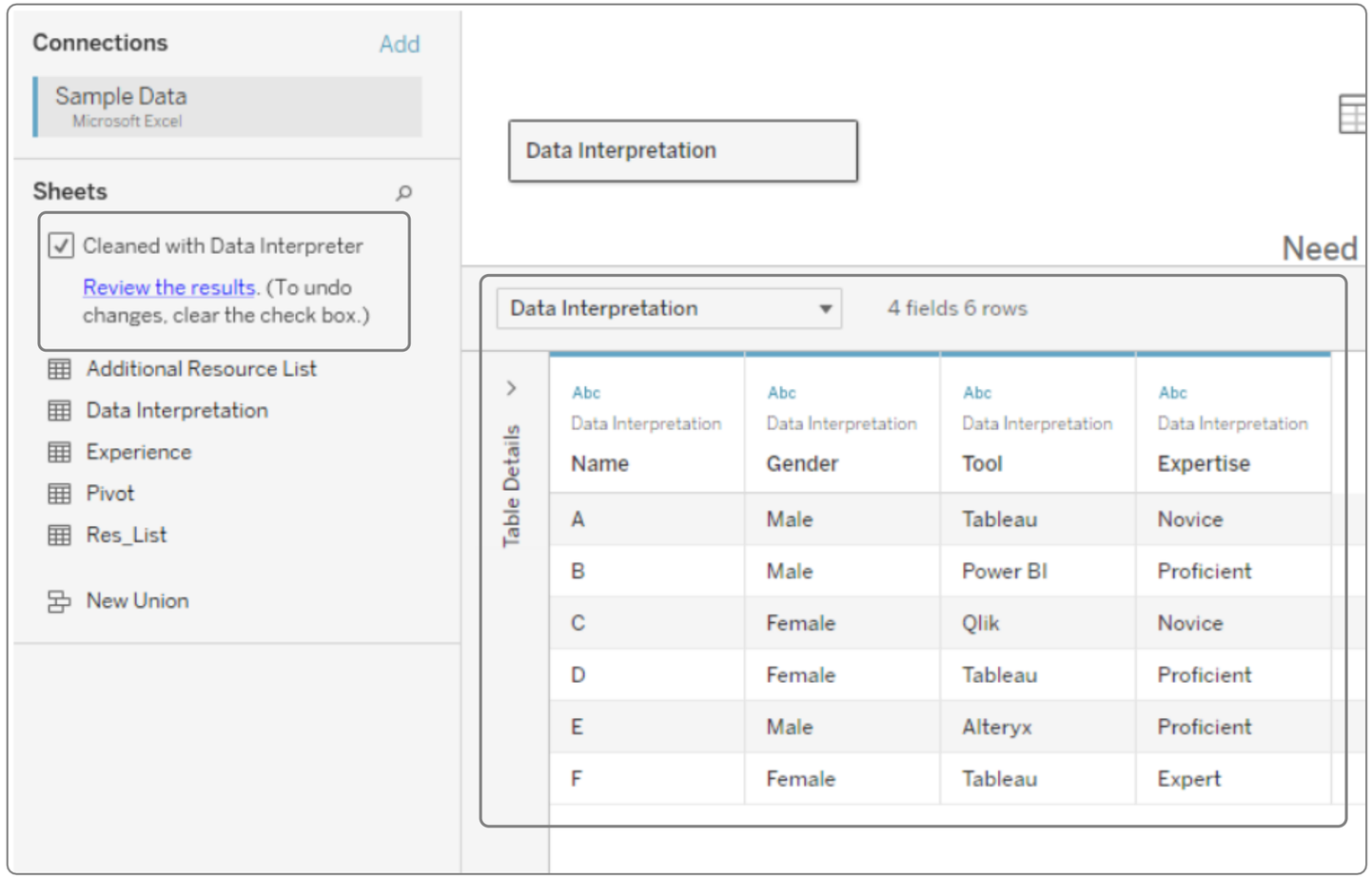
Data Interpreter

Before enabling data interpreter:



Data Interpreter

After enabling data interpreter:



Split

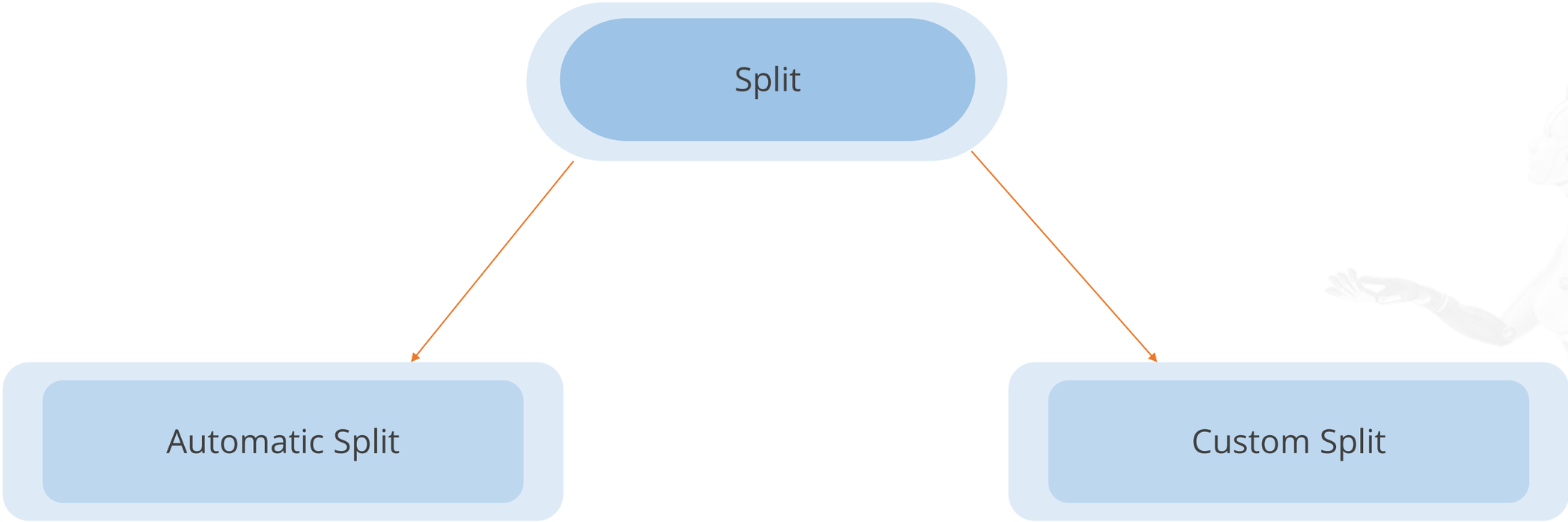
Split

Split is a string function used to split a column into multiple columns.

Split is used to make data analysis easier.

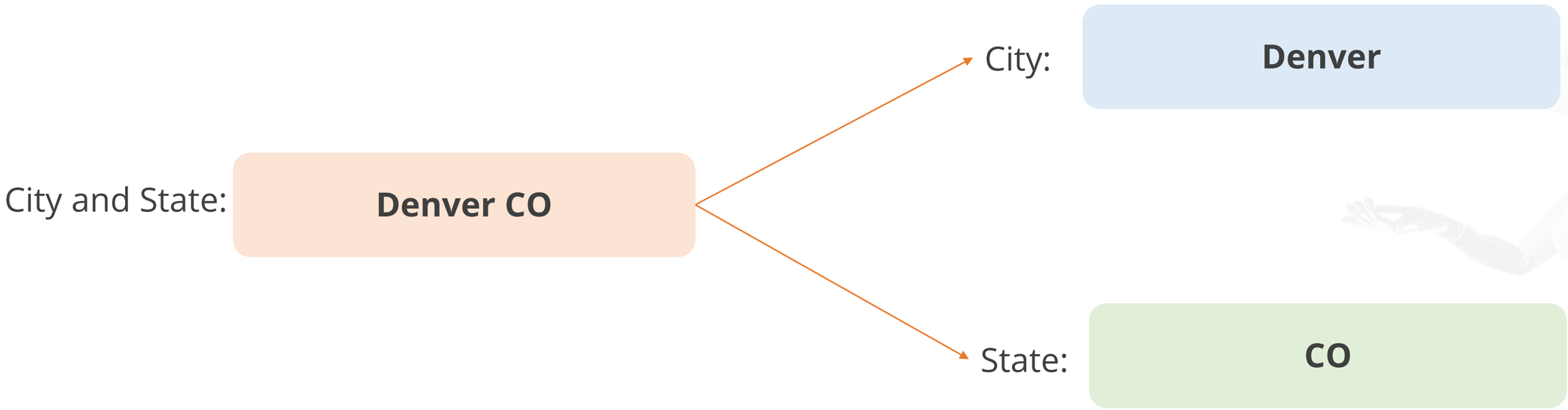
Split

Tableau offers two functions to perform the data preparation step:



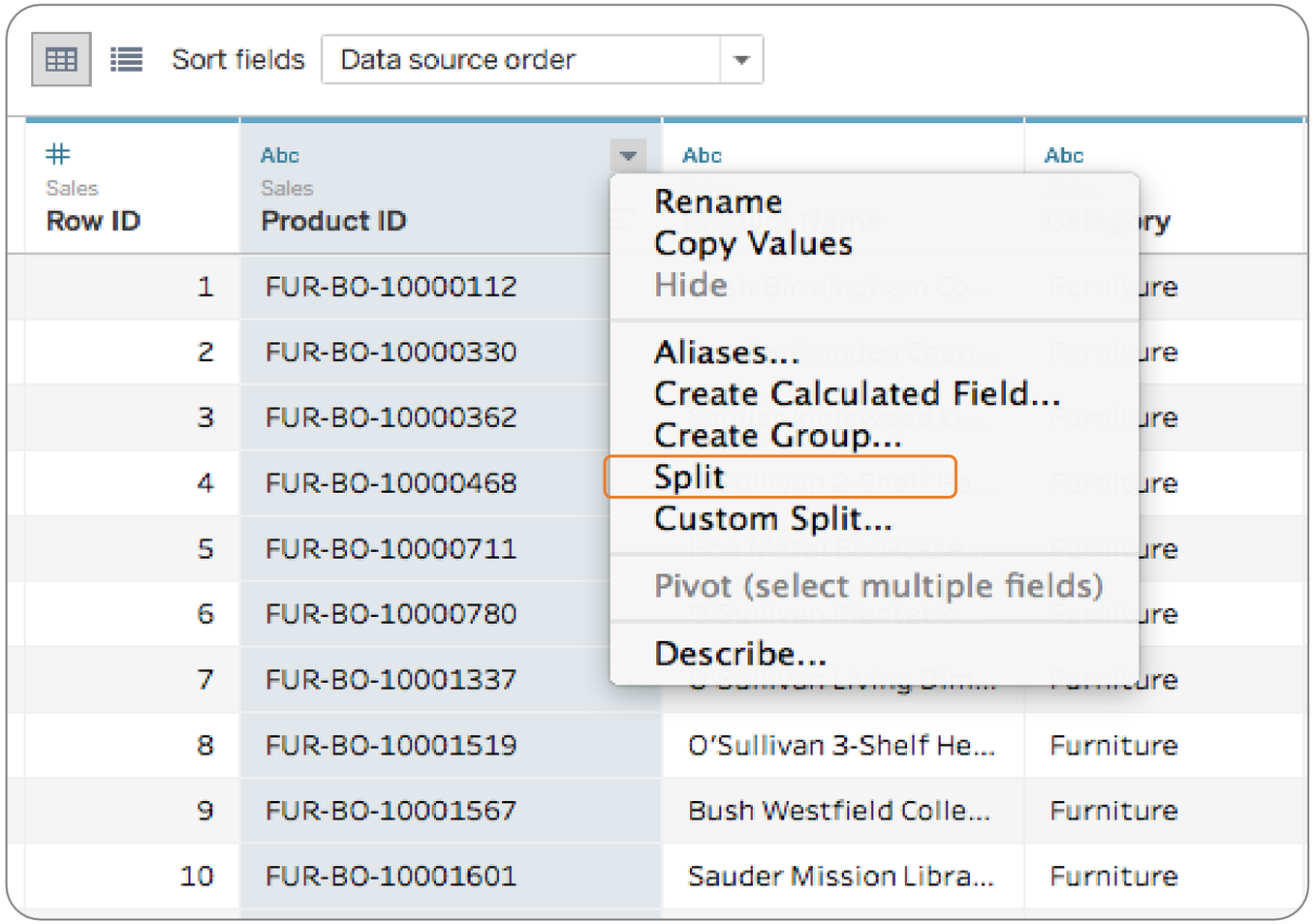
Split

Data splitting from one field to multiple columns are used in data preparation, for example:



Automatic Split

A string field can be split automatically based on a common separator detected by Tableau.



Automatic Split

The following is the output produced by automatic split:

Sort fields: Data source order

# Sales Row ID	Abc Sales Product ID	=Abc Calculation Product ID - Split 1	=# Calculation Product ID - Split 2
1	FUR-BO-10000112	BO	10000112
2	FUR-BO-10000330	BO	10000330
3	FUR-BO-10000362	BO	10000362
4	FUR-BO-10000468	BO	10000468
5	FUR-BO-10000711	BO	10000711
6	FUR-BO-10000780	BO	10000780
7	FUR-BO-10001337	BO	10001337
8	FUR-BO-10001519	BO	10001519
9	FUR-BO-10001567	BO	10001567
10	FUR-BO-10001601	BO	10001601

Custom Split

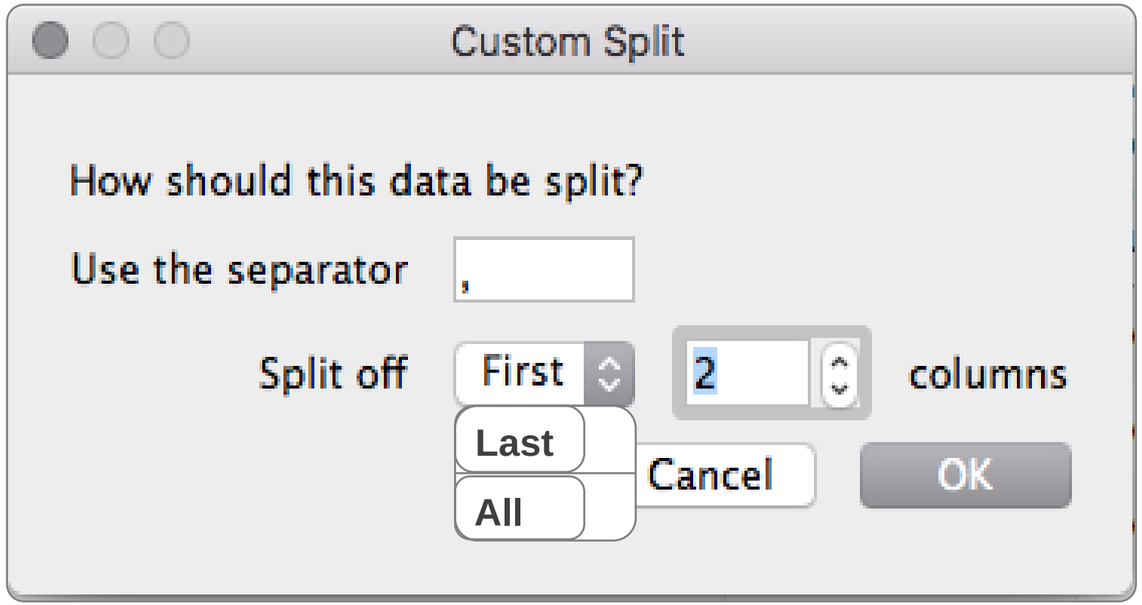
The custom split can separate a string field into a maximum of ten new fields based on a separator within the original field.

# Sales Row ID	Abc Sales Product ID	Abc Sales Product Name	Abc	Abc
1	FUR-BO-10000112	Bush Birmingham Collection Bookcase, Dark Cherry		
2	FUR-BO-10000330	Sauder Camden County Barrister Bookcase, Planked Cherry Finish		
3	FUR-BO-10000362	Sauder Inglewood Library Bookcases		
4	FUR-BO-10000468	O'Sullivan 2-Shelf Heavy-Duty Bookcases		
5	FUR-BO-10000711	Hon Metal Bookcases, Gray		
6	FUR-BO-10000780	O'Sullivan Plantations 2-Door Library in Landvery Oak		
7	FUR-BO-10001337	O'Sullivan Living Dimensions 2-Shelf Bookcases		
8	FUR-BO-10001519	O'Sullivan 3-Shelf Heavy-Duty Bookcases	Furniture	Bookcases
9	FUR-BO-10001567	Bush Westfield Collection Bookcases, Dark Cherry Finish, Fully Assembled	Furniture	Bookcases
10	FUR-BO-10001601	Sauder Mission Library with Doors, Fruitwood Finish	Furniture	Bookcases

Rename
Copy Values
Hide
Aliases...
Create Calculated Field...
Create Group...
Split
Custom Split...
Pivot (select multiple fields)
Describe...

Custom Split

The following screenshot represents the final step:



Splits in the values can be chosen as per the following:

- The separator's first occurrences
- The separator's last occurrences
- The separator’s all occurrences

Custom Split

The following is the output produced by custom split:

Sort fields: Data source order

# Sales Row ID	Abc Sales Product ID	Abc Sales Product Name	=Abc Calculation Product Name - Split 1	=Abc Calculation Product Name - Split 2
1	FUR-BO-10000112	Bush Birmingham Collection Bookcase, Dark Cherry	Bush Birmingham Collection Bookcase	Dark Cherry
2	FUR-BO-10000330	Sauder Camden County Barrister Bookcase, Planked Cherry Finish	Sauder Camden County Barrister Bookcase	Planked Cherry Finish
3	FUR-BO-10000362	Sauder Inglewood Library Bookcases	Sauder Inglewood Library Bookcases	
4	FUR-BO-10000468	O'Sullivan 2-Shelf Heavy-Duty Bookcases	O'Sullivan 2-Shelf Heavy-Duty Bookcases	
5	FUR-BO-10000711	Hon Metal Bookcases, Gray	Hon Metal Bookcases	Gray
6	FUR-BO-10000780	O'Sullivan Plantations 2-Door Library in Landvery Oak	O'Sullivan Plantations 2-Door Library in Landvery Oak	
7	FUR-BO-10001337	O'Sullivan Living Dimensions 2-Shelf Bookcases	O'Sullivan Living Dimensions 2-Shelf Bookcases	
8	FUR-BO-10001519	O'Sullivan 3-Shelf Heavy-Duty Bookcases	O'Sullivan 3-Shelf Heavy-Duty Bookcases	
9	FUR-BO-10001567	Bush Westfield Collection Bookcases, Dark Cherry Finish, Fully Assembled	Bush Westfield Collection Bookcases	Dark Cherry Finish
10	FUR-BO-10001601	Sauder Mission Library with Doors, Fruitwood Finish	Sauder Mission Library with Doors	Fruitwood Finish

Creating Extracts with Multiple Tables

Tableau Data Extracts

Tableau includes functionality for creating data extracts that are compact.

Custom filters can limit the scope of information included in an extract.

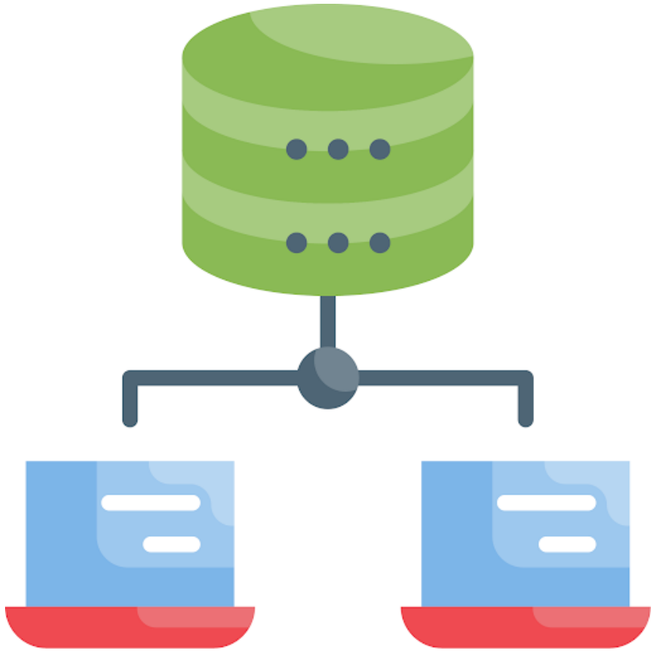
Custom aggregations can minimize the amount of data loaded into the extract.

Extract refreshes can be executed on schedule with a Tableau Server.

Extract can also be published to the Tableau Server for sharing.

Tableau Data Extracts

Extract connections are used in larger datasets.



It can be used if there is no need for real-time data in dashboards.

Tableau Data Extracts

Tableau extracts use the **.hyper** format which is an improved data engine.

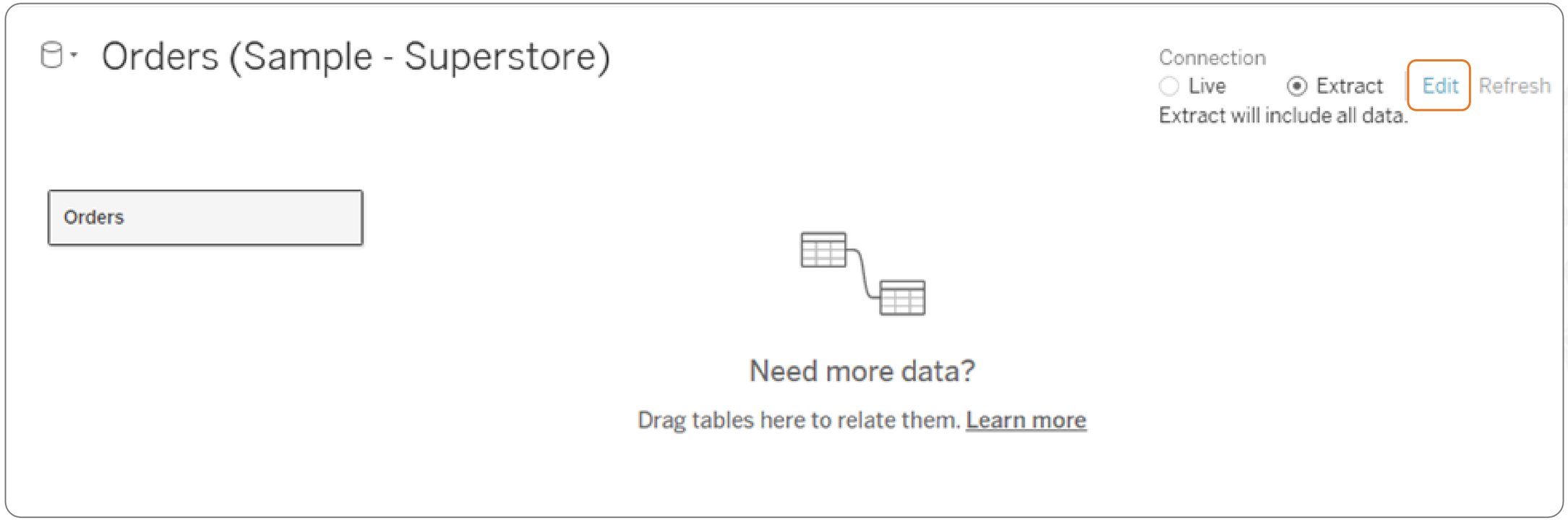
- It can handle any amount of data.
- It supports fast analytics.
- It supports query performance.

Creating Extracts with Multiple Tables

Step 1: Select **Extract** from the **Connection** radio buttons on the data source page

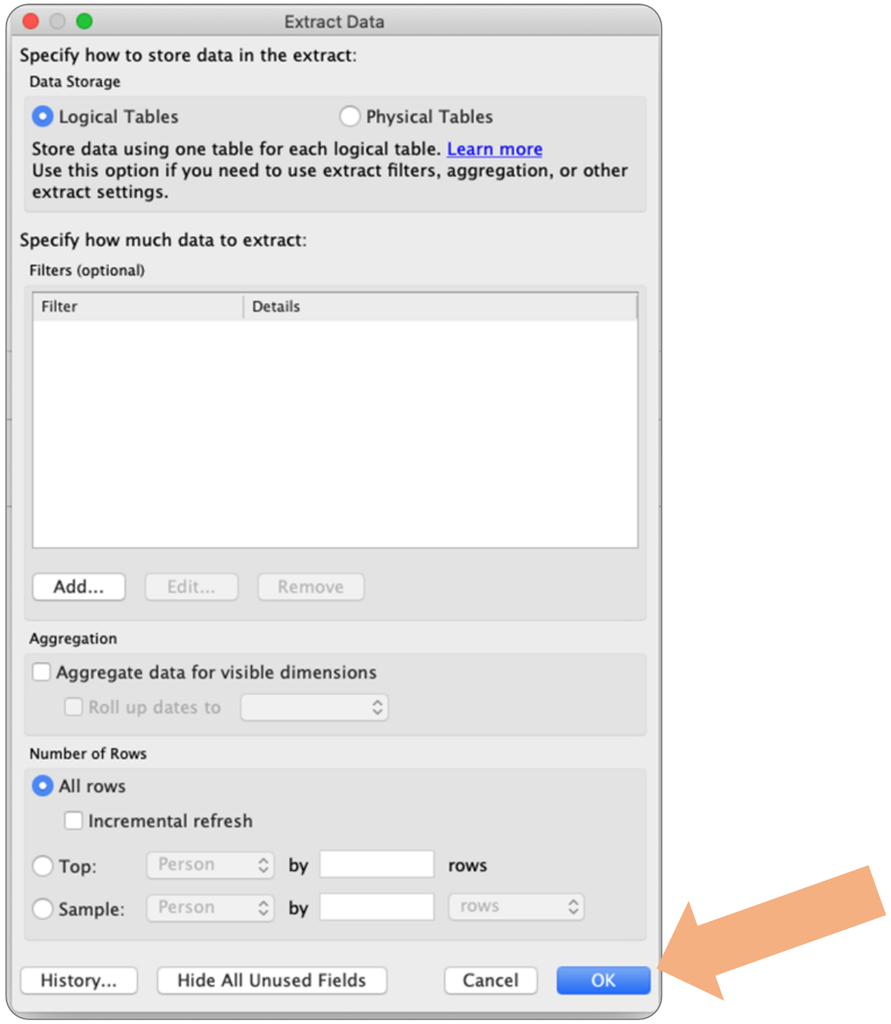
Creating Extracts with Multiple Tables

Step 2: Click on the **Edit** link to open the extract data dialog box



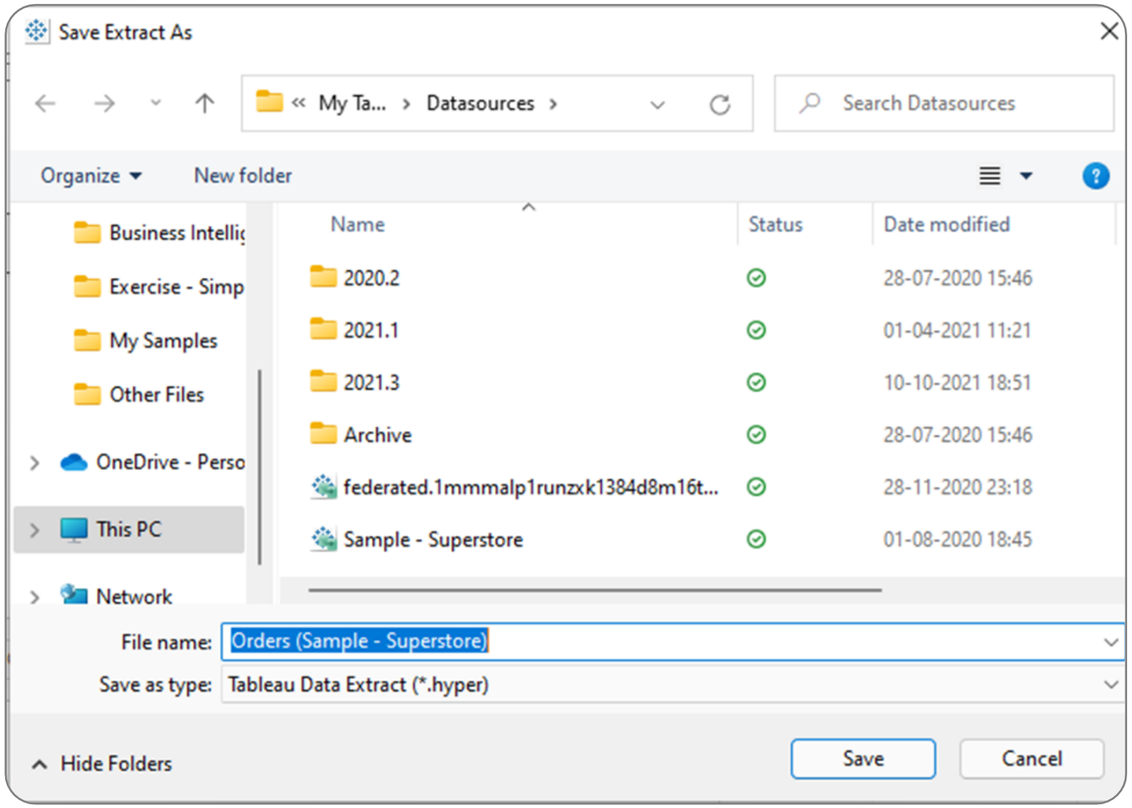
Creating Extracts with Multiple Tables

Step 3: The next step is to click **OK**



Creating Extracts with Multiple Tables

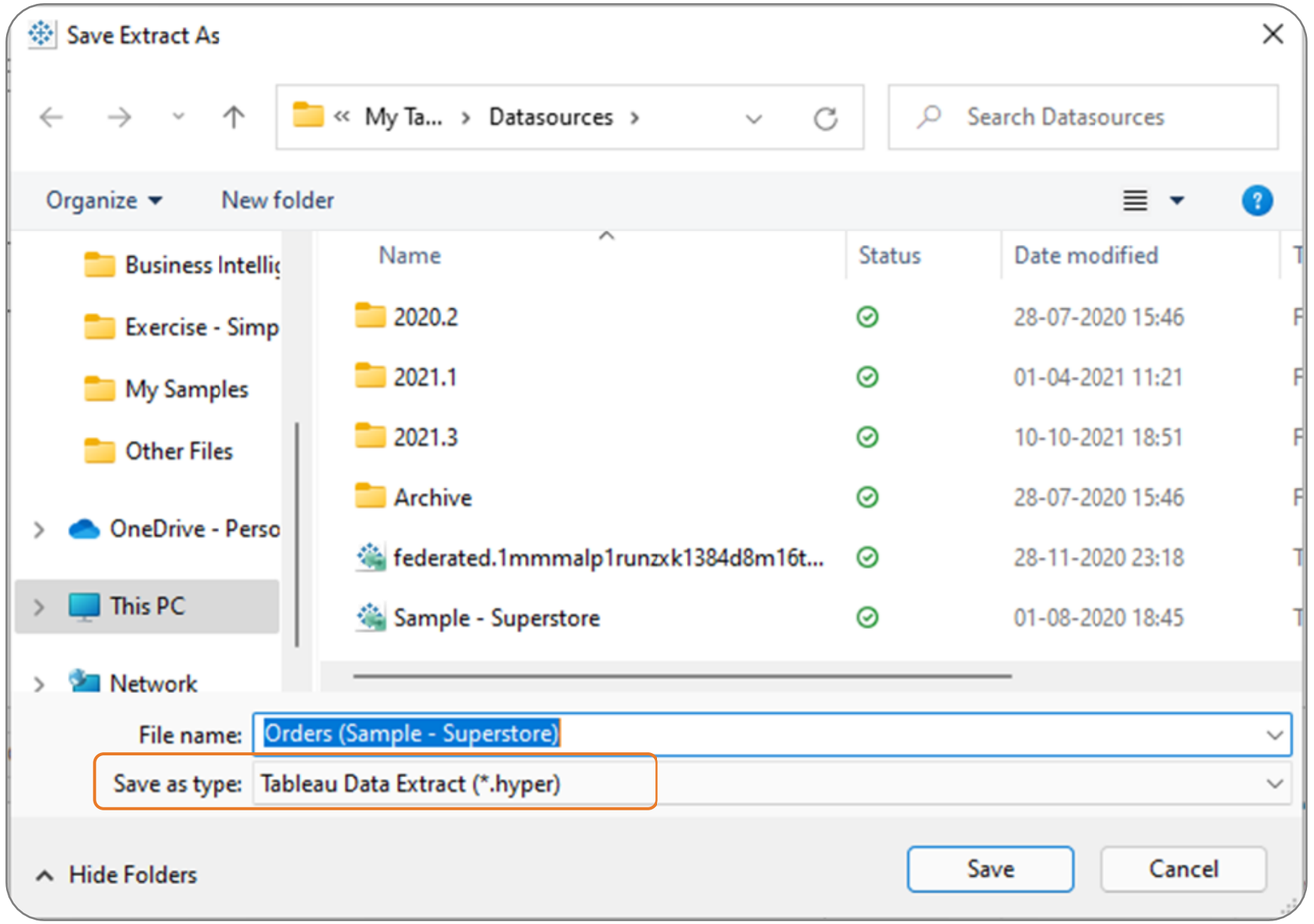
The extract will not start until the worksheet is moved.



Select the location to save the file.

Creating Extracts with Multiple Tables

The extract will be saved in **.hyper** format.



Assisted Practice: Joins and Data Extract

Duration: 20 minutes

Problem statement:

Rollin Ford, the regional manager of a retail chain, has to redesign distribution and production strategies. He wants to look at each state's profits along with the managers' names. He also wants to see the number of products returned by customers in each state. You must combine the data that is stored across different tables and create a visualization to satisfy the above requirement. To enhance the performance, create an extract for this data source. Rollin wants refresh to happen incrementally based on the date.

ASSISTED PRACTICE

Assisted Practice Guidelines

Steps to follow:

Step 1: Combine the Orders, Returns, and People tables

Step 2: Use inner join to combine Orders and People tables

Step 3: Use the left join between Orders and Returns table

Step 4: Create the required view and enable mark labels

Step 5: Create an incremental refresh extract using the order date field

ASSISTED PRACTICE

Assisted Practice: Data Blending

Duration: 20 minutes

Problem statement:

A manager wants to look at the vendor price for these products to calculate the profit margin after the discount. The customer analysis dataset stores the category information, and the sales inventory analysis dataset stores the vendor item price information. You are required to blend data from two datasets and create a visualization that represents the vendor item price for each product category.

ASSISTED PRACTICE

Assisted Practice Guidelines

Steps to follow:

Step 1: Blend data from two data sources based on the common link

Step 2: Add Category Name dimension from Customer Analysis data source and Item Price measure from Sales Inventory Analysis data source

Step 3: Analyze the category-wise item price

Step 4: Color the view by Dept Id

ASSISTED PRACTICE

Key Takeaways

- Data is rarely clean, tidy, and complete, and it is typically not in one file or even one type of file.
- A good BI professional takes the time to understand data prior to using it for analysis with a BI tool like Tableau.
- Tableau offers a suite of functions that help the user automate as much of the data preparation as possible.
- Tableau can point to several different types of files, from a simple Excel file to a custom, complex SQL database.
- Data extracts in Tableau can boost performance, get refreshed automatically, and limit the scope of data.

Knowledge Check

Knowledge Check

1

Which of the following can separate a string field into ten new fields?

A. Automatic split

B. Split

C. Custom split

D. Delimiter

Knowledge Check

1

Which of the following can separate a string field into ten new fields?

A. Automatic split

B. Split

C. Custom split

D. Delimiter

The correct answer is **C**

The custom split can separate a string field into a maximum of ten new fields based on a separator within the original field.

Knowledge Check

2

Which form of modeling supports many-to-many cardinality?

A. Inner join

B. Relationships

C. Full outer join

D. Cross database join

Knowledge Check

2

Which form of modeling supports many-to-many cardinality?

A. Inner join

B. Relationships

C. Full outer join

D. Cross database join

The correct answer is **B**

The newer form of the connectivity method in Tableau supports many-to-many cardinality relationships.

Knowledge Check

3

_____ is used for removing merged cells in Tableau.

A. Pivot

B. Data interpreter

C. Split

D. Null value treatment

Knowledge Check

3

_____ is used for removing merged cells in Tableau.

A. Pivot

B. Data interpreter

C. Split

D. Null value treatment

The correct answer is **B**

A data interpreter is used for removing merged cells. It cleanses the data.

Knowledge Check

4

Which of the following formats store Tableau extracts?

A. .tde

B. .xls

C. .twb

D. .hyper

Knowledge Check

4

Which of the following formats store Tableau extracts?

A. .tde

B. .xls

C. .twb

D. .hyper

The correct answer is **D**

The current hyperdata engine of Tableau supports .hyper. The older version of Tableau supported .tde.

Knowledge Check

5

Which of the following approach connects two datasets at two levels of detail?

A. Cross database join

B. Join

C. Data blending

D. Relationships

Knowledge Check

5

Which of the following approach connects two datasets at two levels of detail?

A. Cross database join

B. Join

C. Data blending

D. Relationships

The correct answer is **C**

Data blending connects two datasets at two levels of detail, whereas joins won't work even if the data is at different levels of detail.