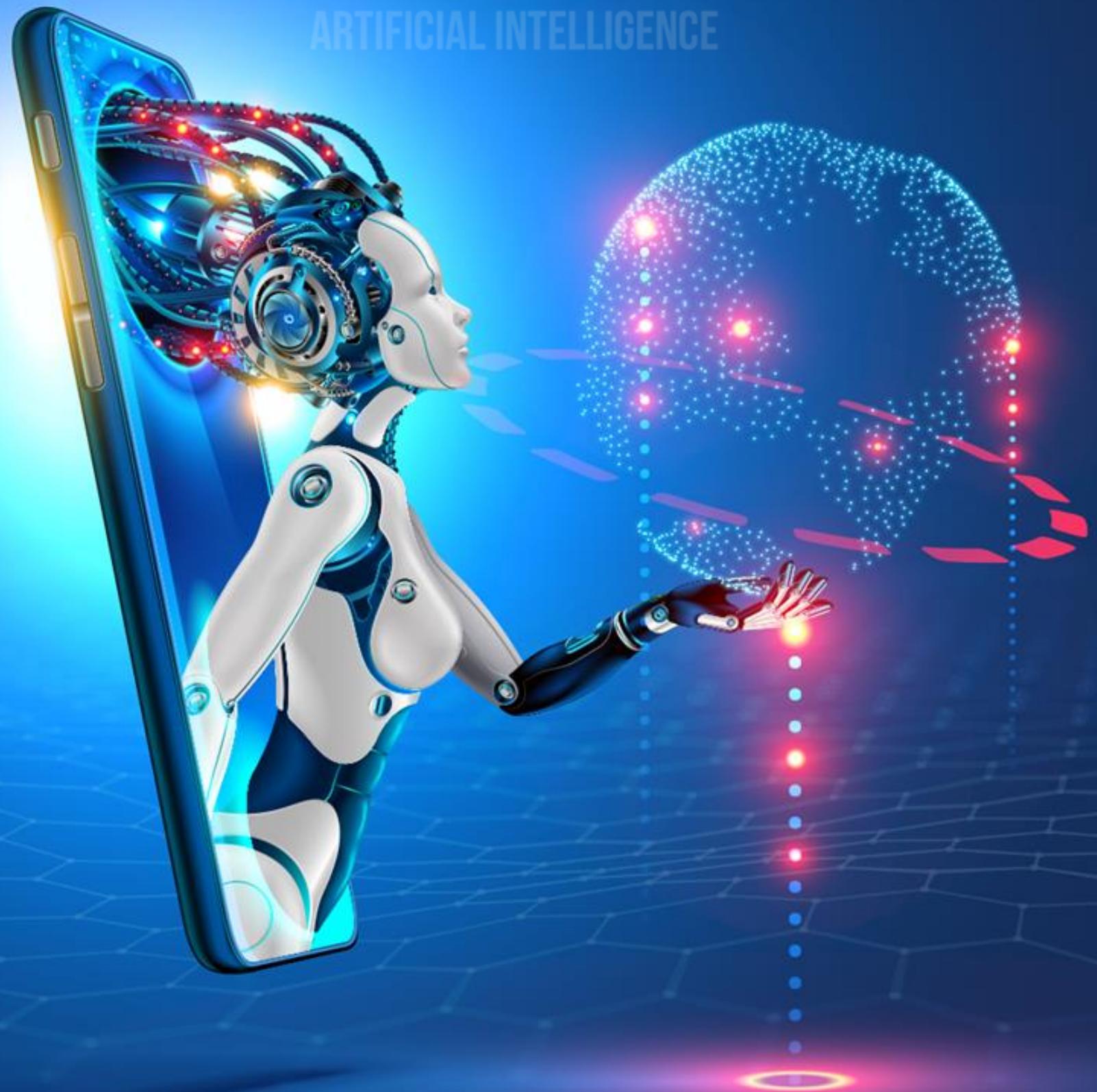
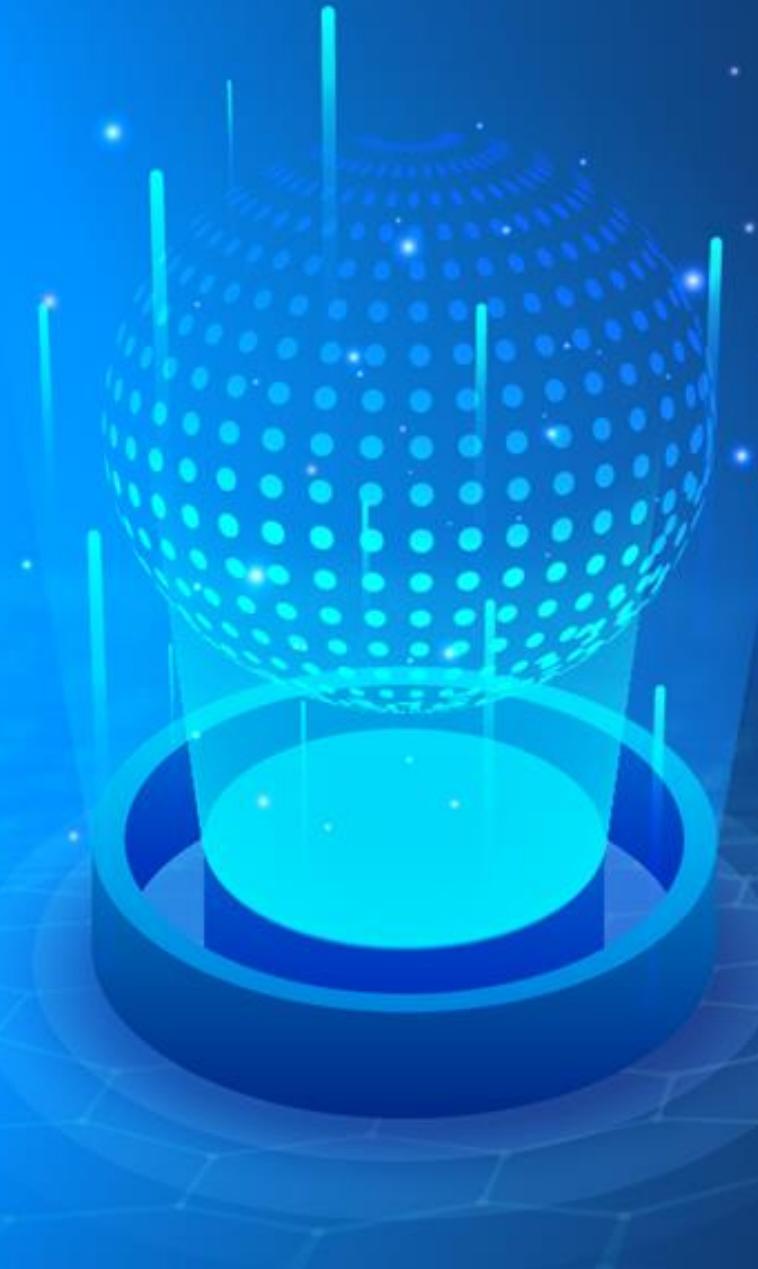


**DATA AND  
ARTIFICIAL INTELLIGENCE**



## Tableau Training



## 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.

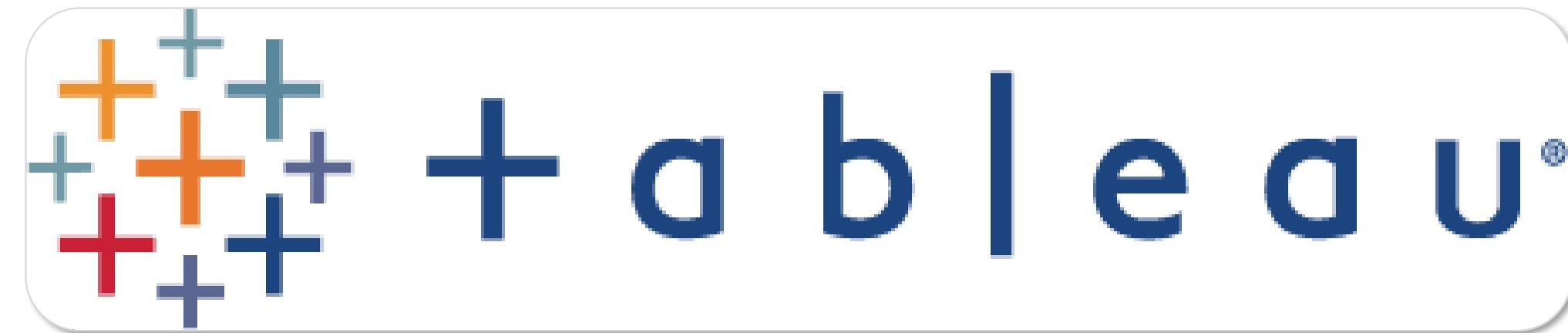
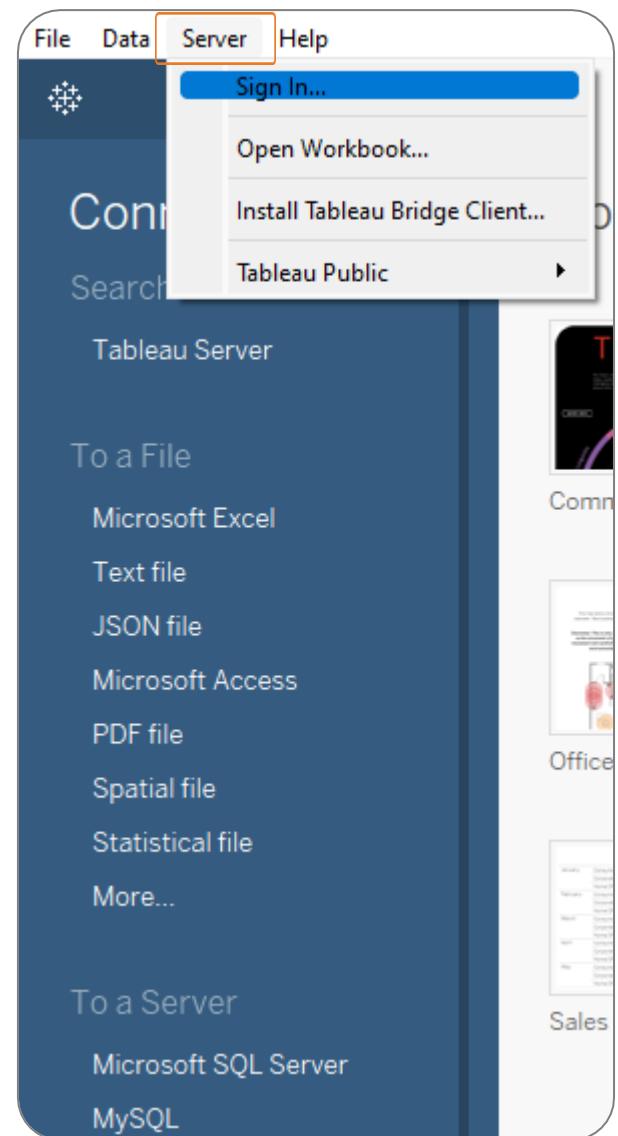


Tableau Server connection is used to publish workbooks and data sources.

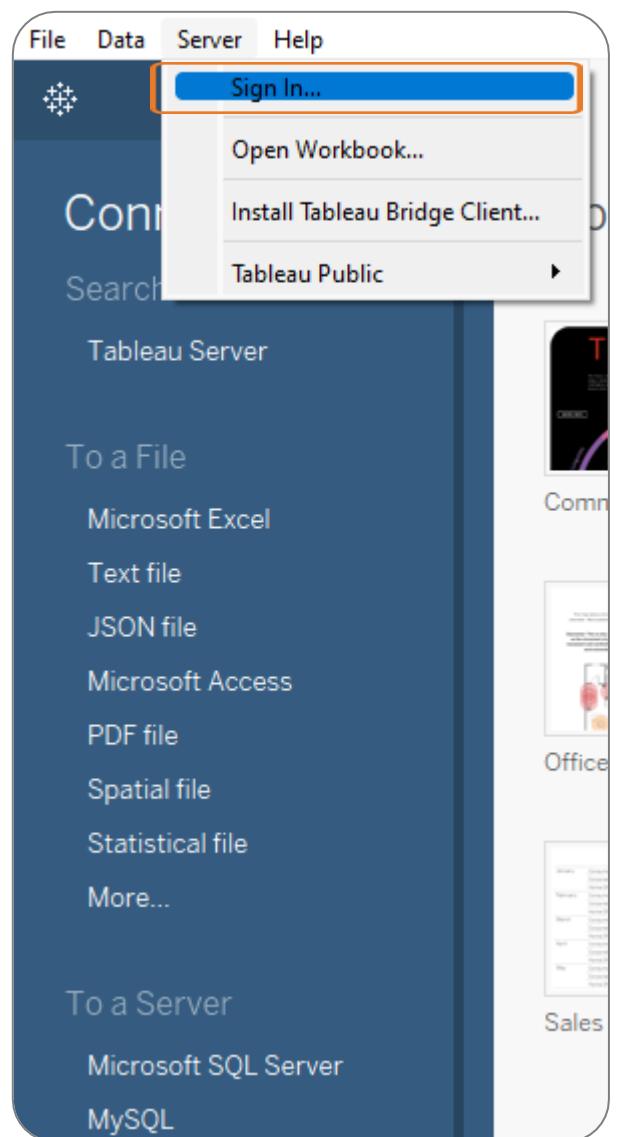
# Connecting to Tableau Server

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



# Connecting to Tableau Server

## Step 2: Select **Sign In** option



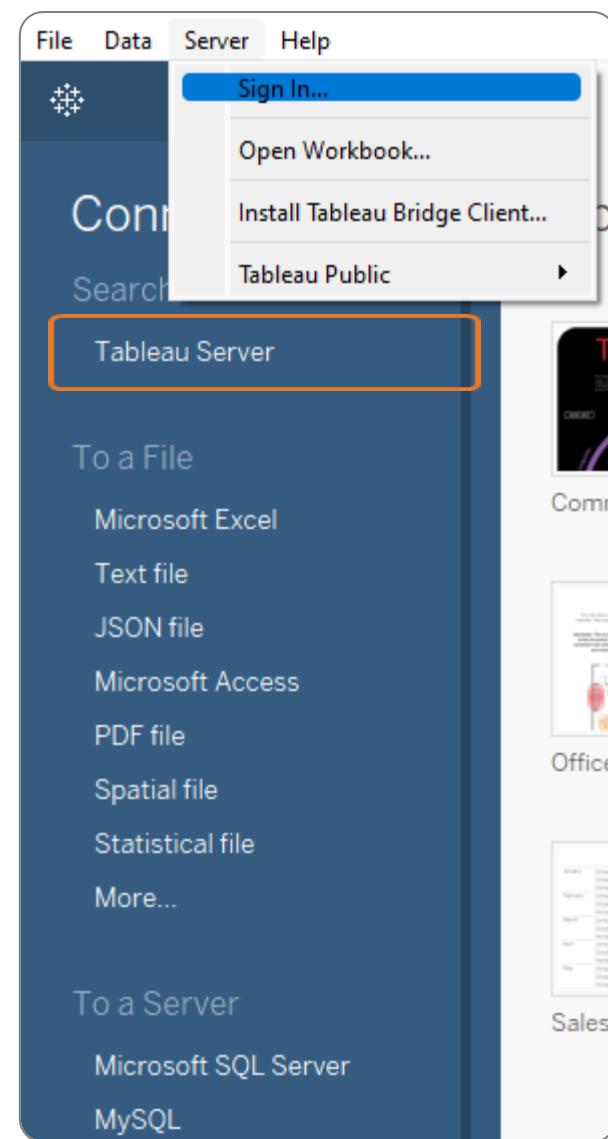
# Connecting to Tableau Server

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.

Search for Data

Content Type: Data Sources (3)

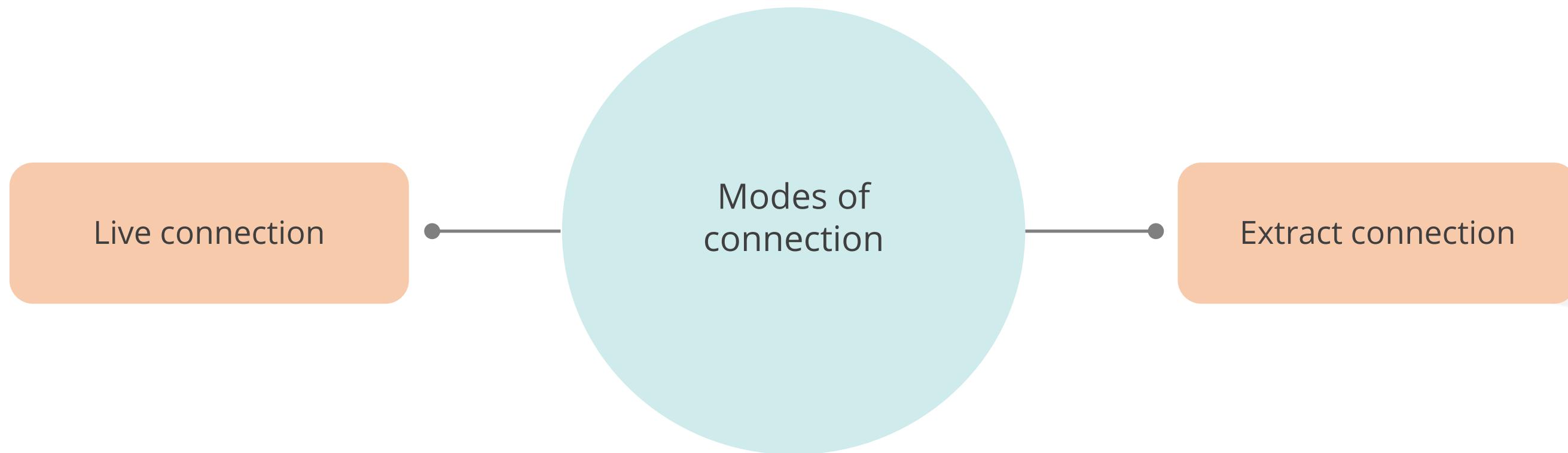
| ① | Type | Name                  | Workbooks | Live/Last Extract     | Connects To             | Location               | Owner        |
|---|------|-----------------------|-----------|-----------------------|-------------------------|------------------------|--------------|
|   | File | Sample - Superstore   | 1         | 1 Aug 2020, 6:45 PM   | Sample - Superstore.xls | My Projects - Training | RAMANATHAN S |
|   | File | Superstore Datasource | 0         | Live                  | Sample - Superstore.xls | Samples                | RAMANATHAN S |
|   | File | Project Tracker       | 0         | 16 Aug 2021, 10:38 AM | Project Tracker.xlsx    | Important              | RAMANATHAN S |

Connect

## Types of Connections

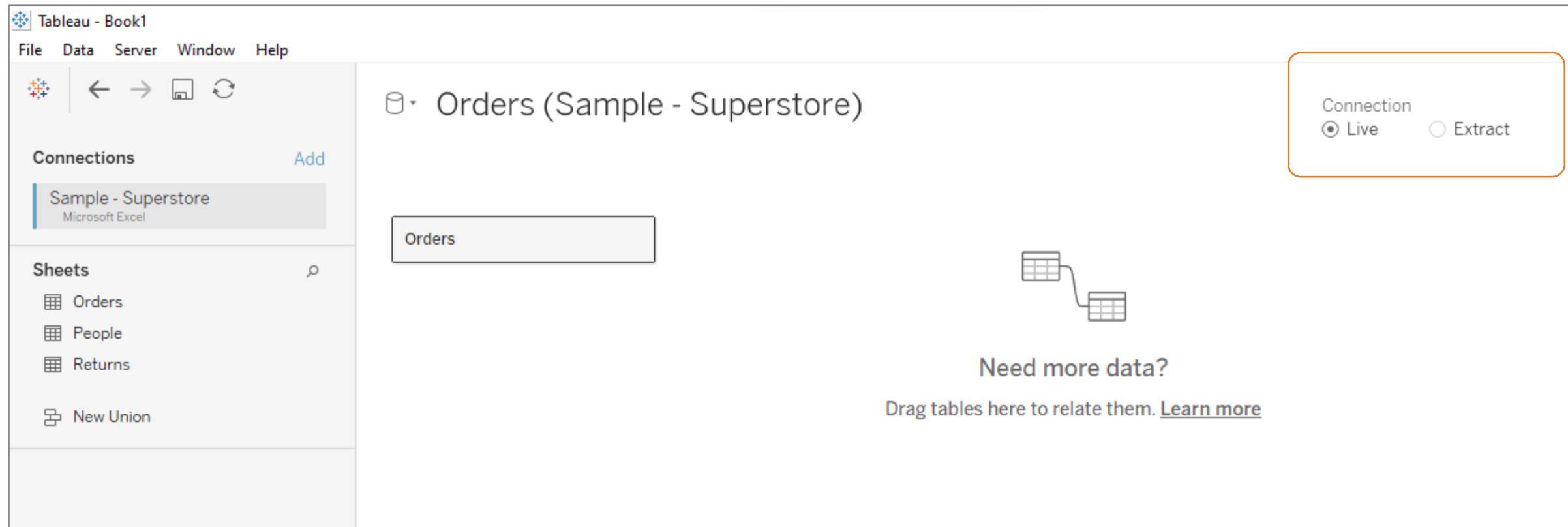
# Types of Connections

Tableau supports two types of connecting data sources.



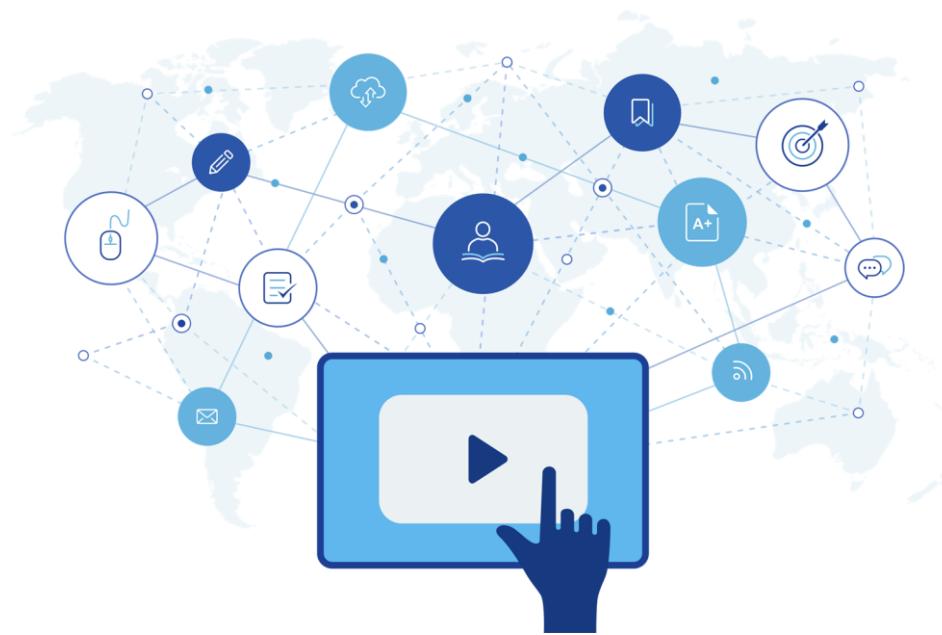
# Types of Connections

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



# Types of Connections

Live connection is used in the following scenarios:

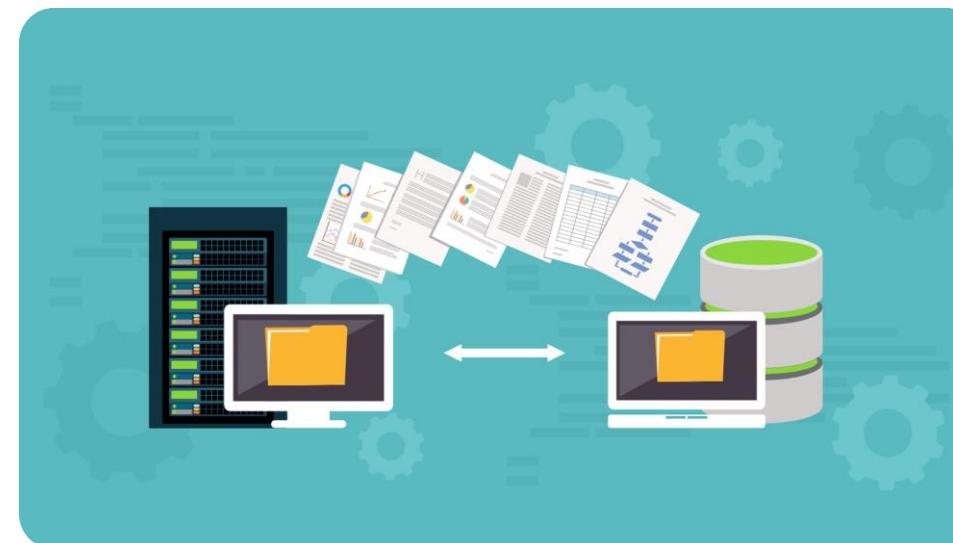


## Live Connection

- 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.

## Connecting to Different Data Sources

# Connecting to Different Data Sources

Popular data sources that can be connected to Tableau are:



Excel file



JSON file



Text file



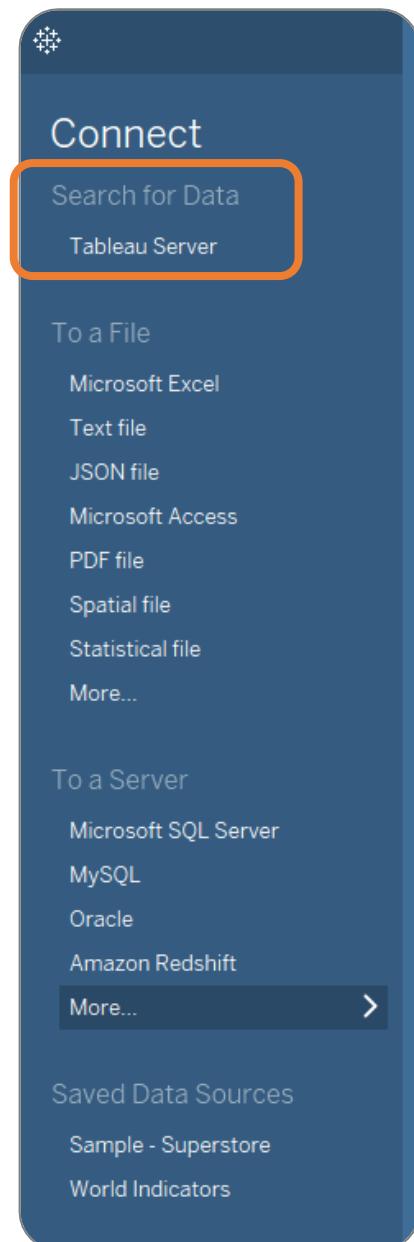
Spatial file



Statistical file

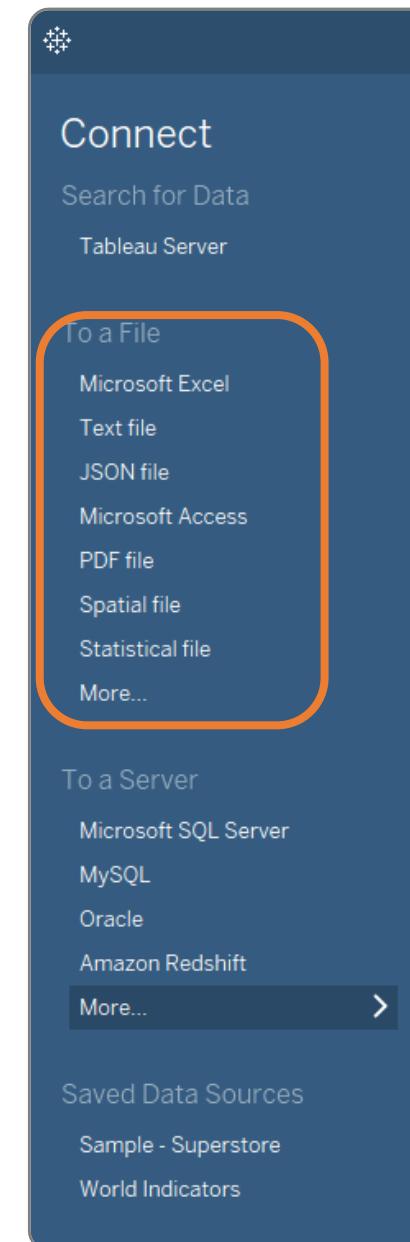
# Connecting to Different Data Sources

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



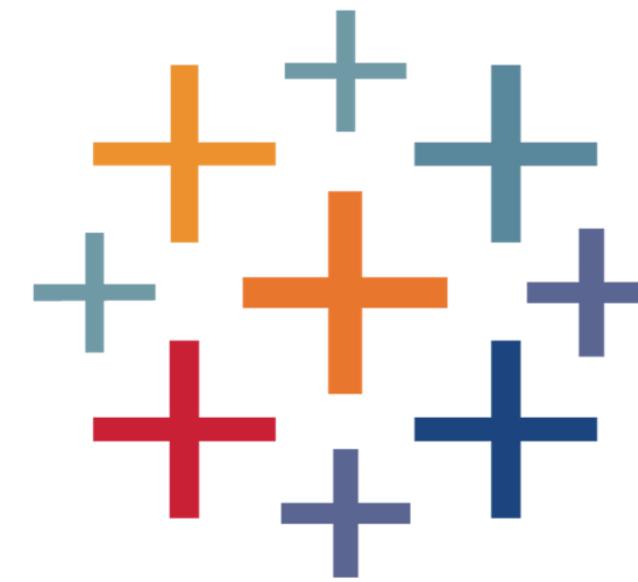
# Connecting to Different Data Sources

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



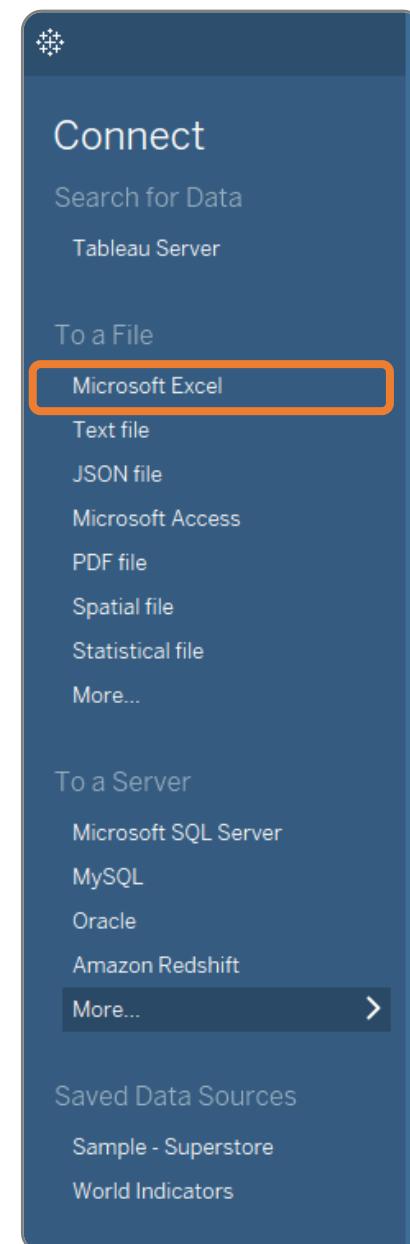
# Connecting to Different Data Sources

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



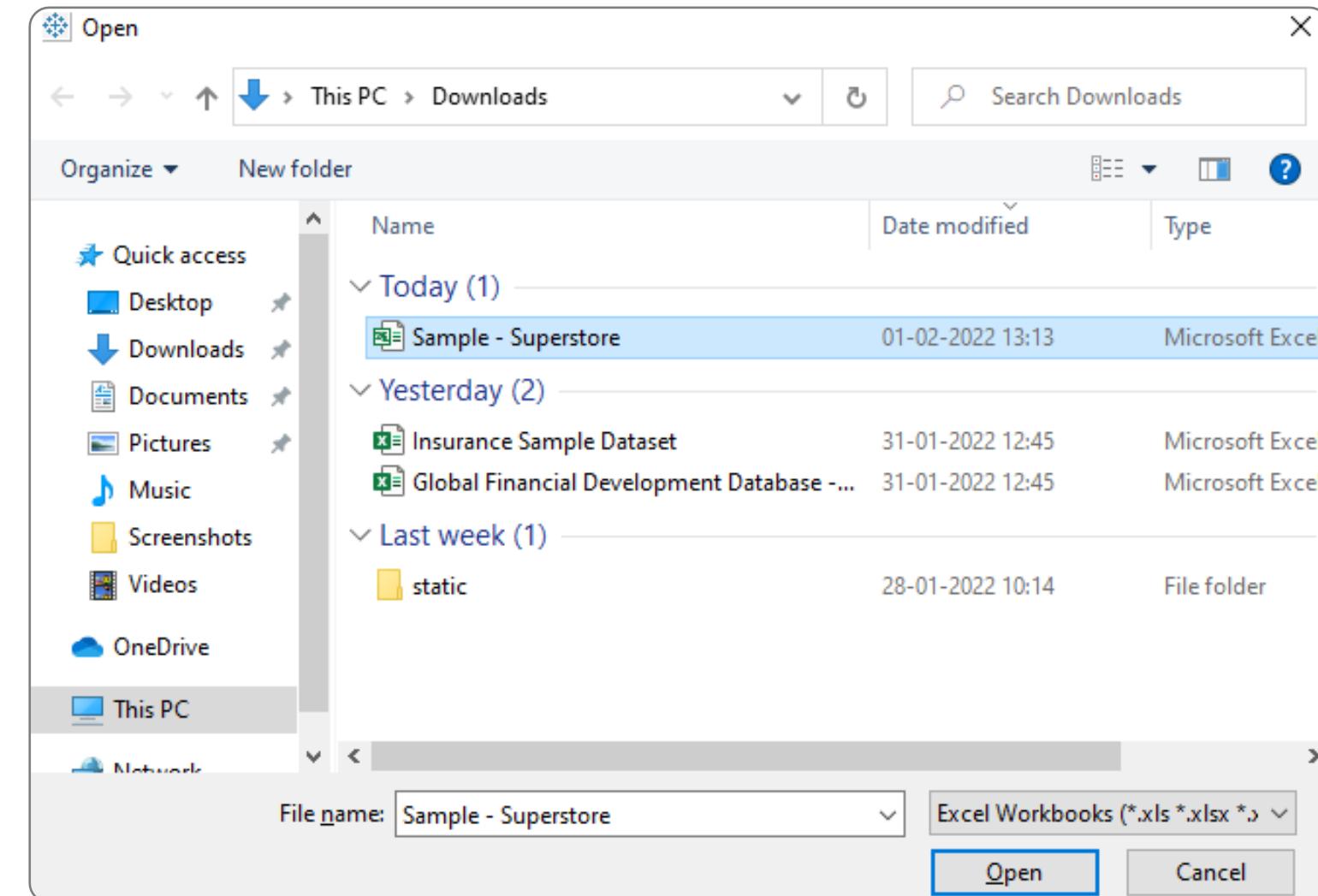
# Connecting to Different Data Sources

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



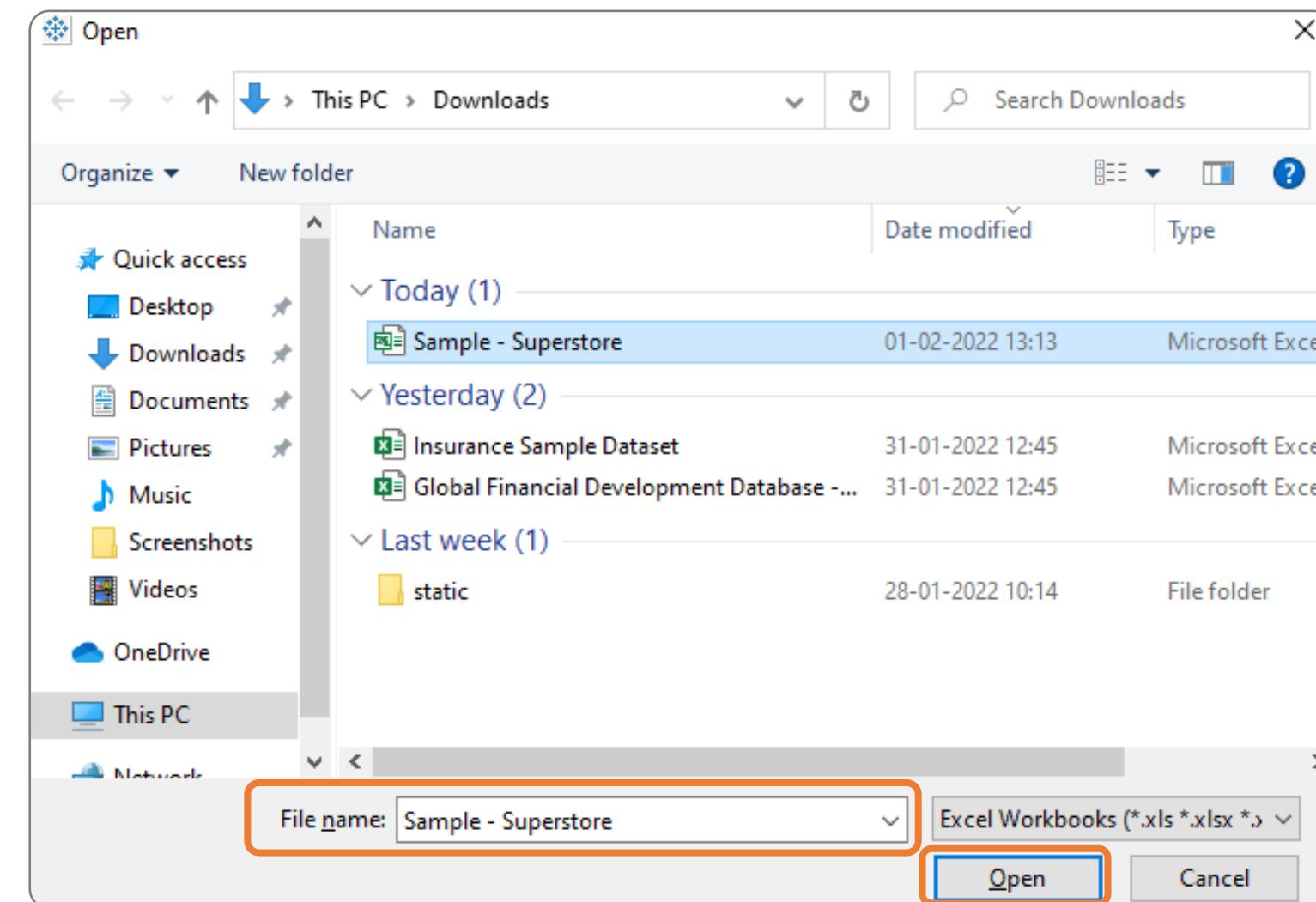
# Connecting to Different Data Sources

A pop-up window will open.



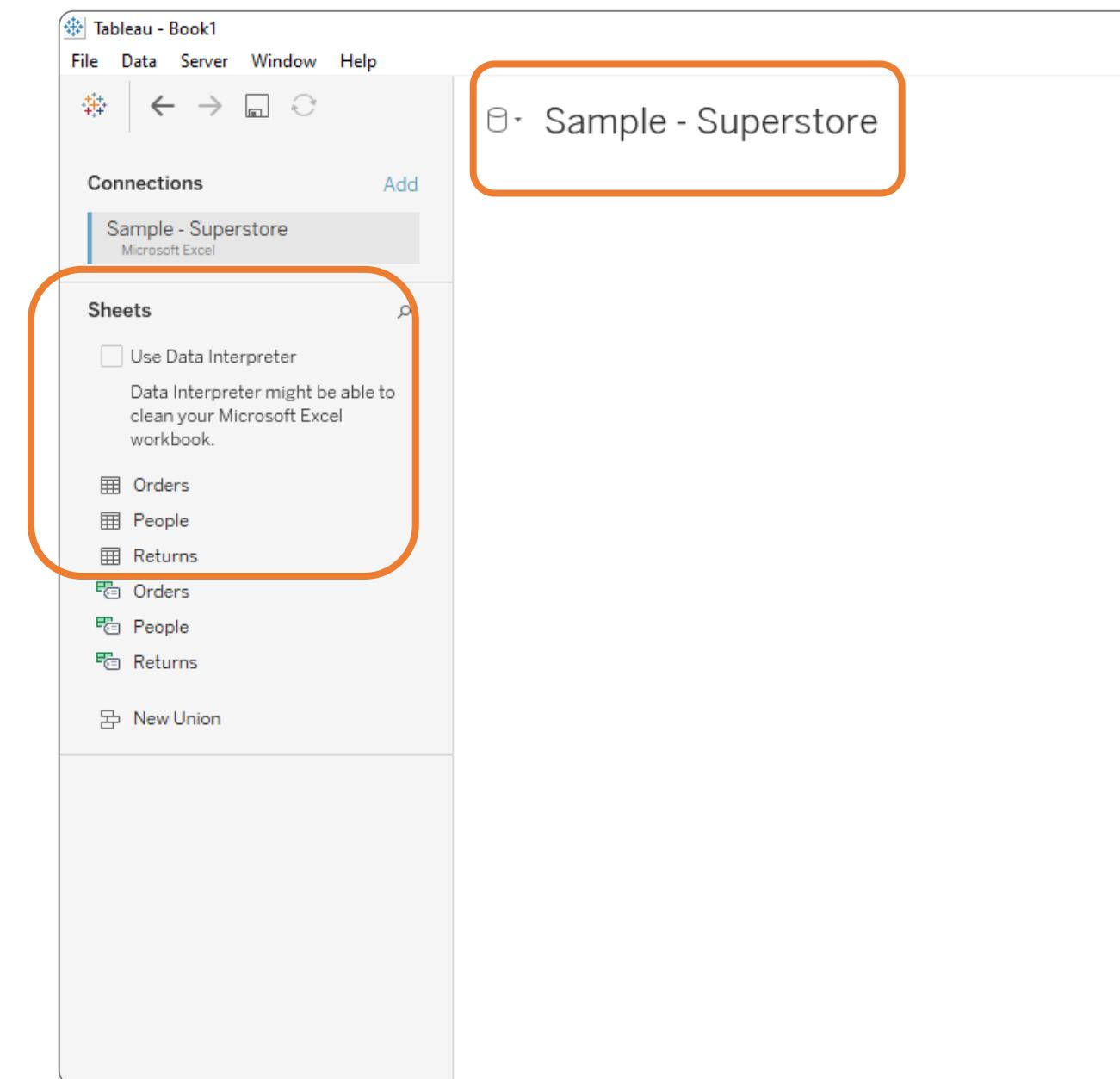
# Connecting to Different Data Sources

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



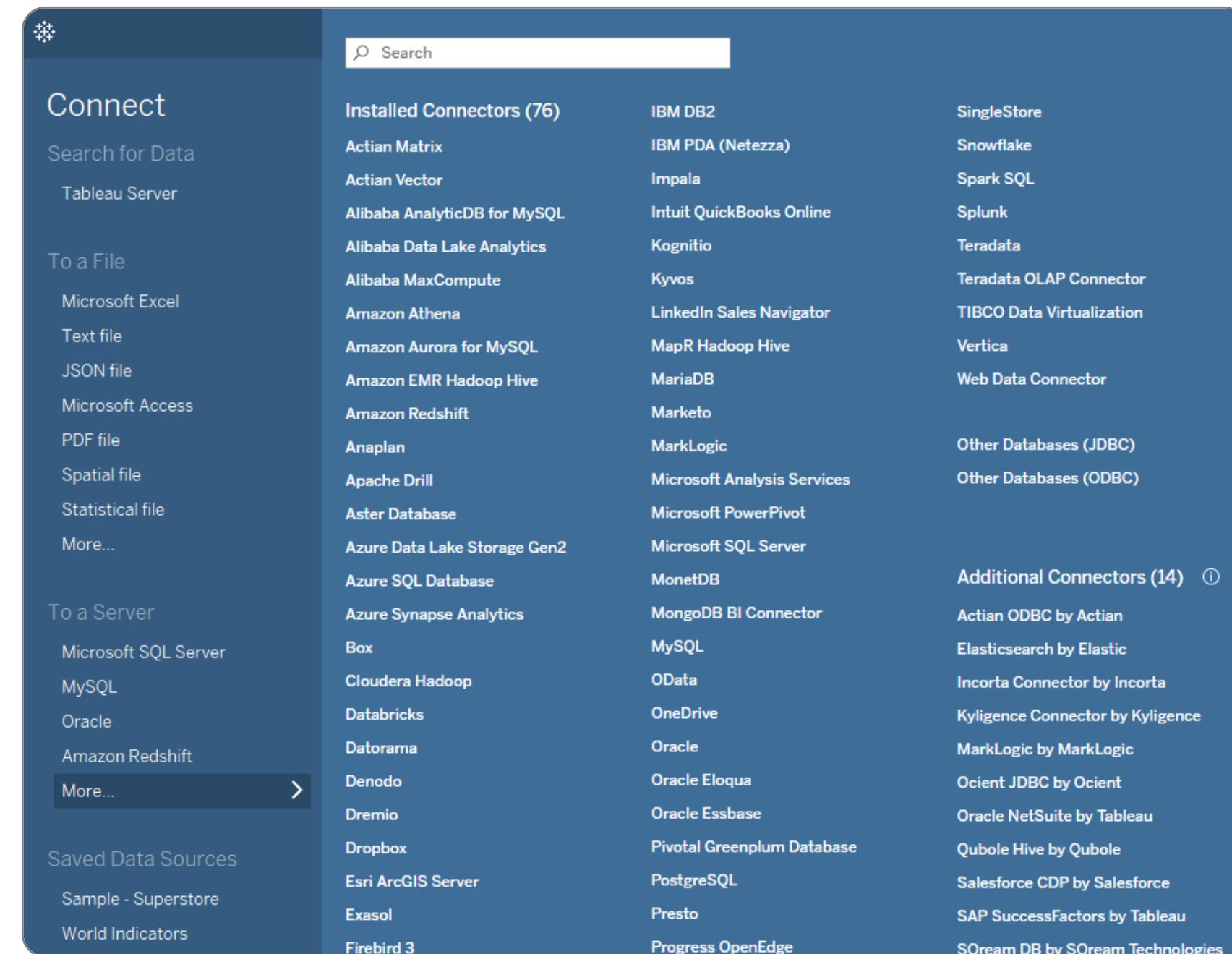
# Connecting to Different Data Sources

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.



The screenshot shows the Tableau Connect interface. On the left, there's a sidebar with options like 'Search for Data', 'Tableau Server', 'To a File' (listing Microsoft Excel, Text file, JSON file, Microsoft Access, PDF file, Spatial file, Statistical file, More...), 'To a Server' (listing Microsoft SQL Server, MySQL, Oracle, Amazon Redshift, More...), and 'Saved Data Sources' (listing Sample - Superstore, World Indicators). The main area is titled 'Installed Connectors (76)' and contains a grid of connector names. A search bar is at the top of this list. The connectors are grouped into columns: the first column has Actian Matrix, Actian Vector, Alibaba AnalyticDB for MySQL, Alibaba Data Lake Analytics, Alibaba MaxCompute, Amazon Athena, Amazon Aurora for MySQL, Amazon EMR Hadoop Hive, Amazon Redshift, Anaplan, Apache Drill, Aster Database, Azure Data Lake Storage Gen2, Azure SQL Database, Azure Synapse Analytics, Box, Cloudera Hadoop, Databricks, Datorama, Denodo, Dremio, Dropbox, Esri ArcGIS Server, Exasol, Firebird 3; the second column has IBM DB2, Impala, Intuit QuickBooks Online, Kyvos, LinkedIn Sales Navigator, MapR Hadoop Hive, MariaDB, Marketo, MarkLogic, Microsoft Analysis Services, Microsoft PowerPivot, Microsoft SQL Server, MonetDB, MongoDB BI Connector, MySQL, OData, OneDrive, Oracle, Oracle Eloqua, Oracle Essbase, Pivotal Greenplum Database, PostgreSQL, Presto, Progress OpenEdge; the third column has SingleStore, Snowflake, Spark SQL, Splunk, Teradata, Teradata OLAP Connector, TIBCO Data Virtualization, Vertica, Web Data Connector, Other Databases (JDBC), Other Databases (ODBC), Microsoft PowerPivot, Microsoft SQL Server, MonetDB, MongoDB BI Connector, MySQL, OData, OneDrive, Oracle, Oracle Eloqua, Oracle Essbase, Pivotal Greenplum Database, PostgreSQL, Presto, Progress OpenEdge, Other Databases (JDBC), Other Databases (ODBC); the fourth column has Additional Connectors (14) (with a link icon), Action ODBC by Actian, Elasticsearch by Elastic, Incorta Connector by Incorta, Kyligence Connector by Kyligence, MarkLogic by MarkLogic, Ocient JDBC by Ocient, Oracle NetSuite by Tableau, Qubole Hive by Qubole, Salesforce CDP by Salesforce, SAP SuccessFactors by Tableau, SOracle DB by SOracle Technologies.

| Installed Connectors (76)    |                              |                            |                                    |
|------------------------------|------------------------------|----------------------------|------------------------------------|
| Actian Matrix                | IBM DB2                      | SingleStore                |                                    |
| Actian Vector                | IBM PDA (Netezza)            | Snowflake                  |                                    |
| Alibaba AnalyticDB for MySQL | Impala                       | Spark SQL                  |                                    |
| Alibaba Data Lake Analytics  | Intuit QuickBooks Online     | Splunk                     |                                    |
| Alibaba MaxCompute           | Kognitio                     | Teradata                   |                                    |
| Amazon Athena                | Kyvos                        | Teradata OLAP Connector    |                                    |
| Amazon Aurora for MySQL      | LinkedIn Sales Navigator     | TIBCO Data Virtualization  |                                    |
| Amazon EMR Hadoop Hive       | MapR Hadoop Hive             | Vertica                    |                                    |
| Amazon Redshift              | MariaDB                      | Web Data Connector         |                                    |
| Anaplan                      | Marketo                      |                            |                                    |
| Apache Drill                 | MarkLogic                    | Other Databases (JDBC)     |                                    |
| Statistical file             | Microsoft Analysis Services  | Other Databases (ODBC)     |                                    |
| More...                      | Aster Database               |                            |                                    |
|                              | Azure Data Lake Storage Gen2 | Microsoft PowerPivot       |                                    |
|                              | Azure SQL Database           | Microsoft SQL Server       |                                    |
|                              | Azure Synapse Analytics      | MonetDB                    | Additional Connectors (14) ⓘ       |
|                              | Box                          | MongoDB BI Connector       | Action ODBC by Actian              |
|                              | Cloudera Hadoop              | MySQL                      | Elasticsearch by Elastic           |
|                              | Databricks                   | OData                      | Incorta Connector by Incorta       |
|                              | Datorama                     | OneDrive                   | Kyligence Connector by Kyligence   |
|                              | Denodo                       | Oracle                     | MarkLogic by MarkLogic             |
|                              | Dremio                       | Oracle Eloqua              | Ocient JDBC by Ocient              |
|                              | Dropbox                      | Oracle Essbase             | Oracle NetSuite by Tableau         |
|                              | Esri ArcGIS Server           | Pivotal Greenplum Database | Qubole Hive by Qubole              |
|                              | Exasol                       | PostgreSQL                 | Salesforce CDP by Salesforce       |
|                              | Firebird 3                   | Presto                     | SAP SuccessFactors by Tableau      |
|                              |                              | Progress OpenEdge          | SOracle DB by SOracle Technologies |

## Preparing Data for Analysis

# 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



## 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

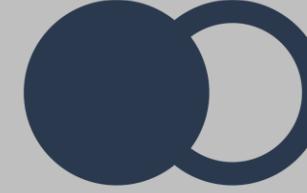


Removing duplicate

# Types of Joins

There are four types of joins:

Venn Diagram



Left Join

Venn Diagram



Right Join

Venn Diagram



Inner Join

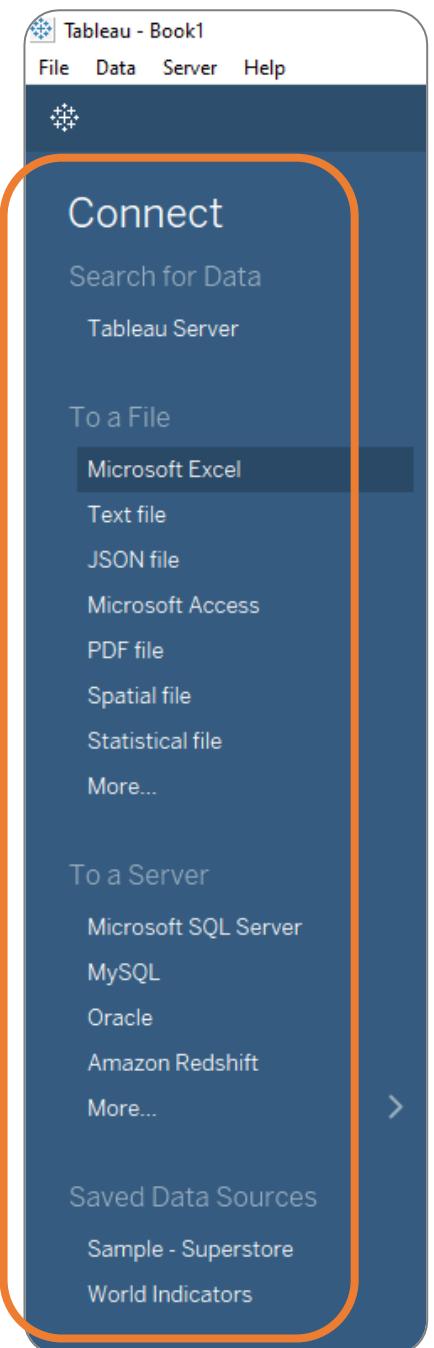
Venn Diagram



Full Outer Join

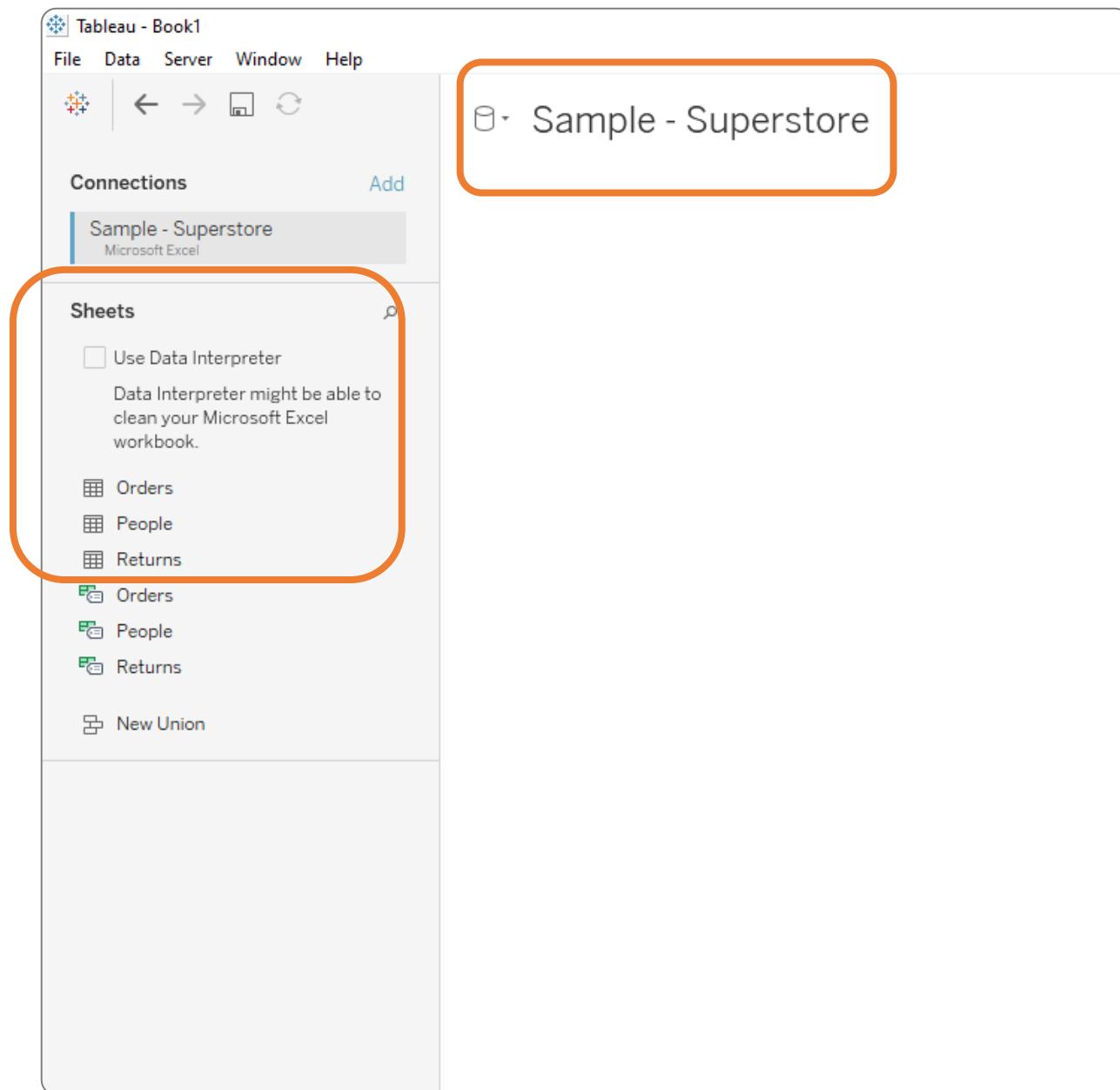
# Steps to Perform Joins

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

The screenshot shows the Tableau Data Source interface. On the left, the 'Connections' pane shows a single connection named 'Sample - Superstore' (Microsoft Excel). Below it, the 'Sheets' pane lists several tables: 'Orders', 'People', 'Returns', and three tables from the 'Orders' sheet: 'Orders', 'People', and 'Returns'. The 'Orders' table from the main sheet is highlighted with an orange border. The main workspace displays a tree view under the heading 'Orders (Sample - Superstore)', with the 'Orders' table also highlighted with an orange border. A small icon of two tables connected by a line is located in the bottom right corner of the workspace. Text at the bottom right says 'Need more data? Drag tables here to relate them.' followed by a link 'Learn more'.

# Steps to Perform Joins

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

The screenshot shows the Tableau Data Source interface. On the left, the 'Connections' pane shows a single connection named 'Sample - Superstore' (Microsoft Excel). Below it, the 'Sheets' pane lists several tables: 'Orders', 'People', 'Returns', and their respective 'Orders', 'People', and 'Returns' counterparts, along with 'New Union'. The 'Orders' table is currently selected, indicated by a blue border. A context menu is open over the 'Orders' table, with the 'Open...' option highlighted in blue. Other options in the menu include 'Rename', 'Remove', and 'Field names are in first row' (which is checked). To the right of the menu, there is a small icon of two tables connected by a line. Below the menu, the text 'Need more data?' and 'Drag tables here to relate them.' is visible, along with a link 'Learn more'.

# Steps to Perform Joins

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

The screenshot shows the Tableau Data Source interface. On the left, the 'Connections' pane displays a single connection named 'Sample - Superstore' (Microsoft Excel). Below it, the 'Sheets' pane lists several tables: 'Orders', 'People', 'Returns', 'Orders', 'People', 'Returns', and 'New Union'. The 'Returns' table from the second row is highlighted with an orange box. On the right, the main workspace shows a connection diagram for 'Orders+'. It indicates that 'Orders' is composed of two tables. A line connects the 'Orders' node to the 'Returns' node, which is also highlighted with an orange box. A tooltip states: 'Orders is made of 2 tables.'

# Steps to Perform Joins

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

Orders is made of 2 tables. ⓘ

The screenshot shows a software interface for performing joins. At the top, there are two rectangular boxes labeled 'Orders' and 'Returns'. Between them is a Venn diagram consisting of two overlapping circles, where the intersection area is shaded blue. Below this, a modal window titled 'Join' is open. It contains four options: 'Inner' (two overlapping circles), 'Left' (one circle with the other partially overlapping), 'Right' (one circle with the other partially overlapping), and 'Full Outer' (two separate circles). The 'Inner' option is selected. The modal also includes a 'Data Source' section with 'Order ID' and a 'Returns' section with 'Order ID (Returns)' separated by an equals sign. A note at the bottom says 'Add new join clause'.

Join

| Data Source | Returns              |
|-------------|----------------------|
| Order ID    | = Order ID (Returns) |

Add new join clause

# 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.

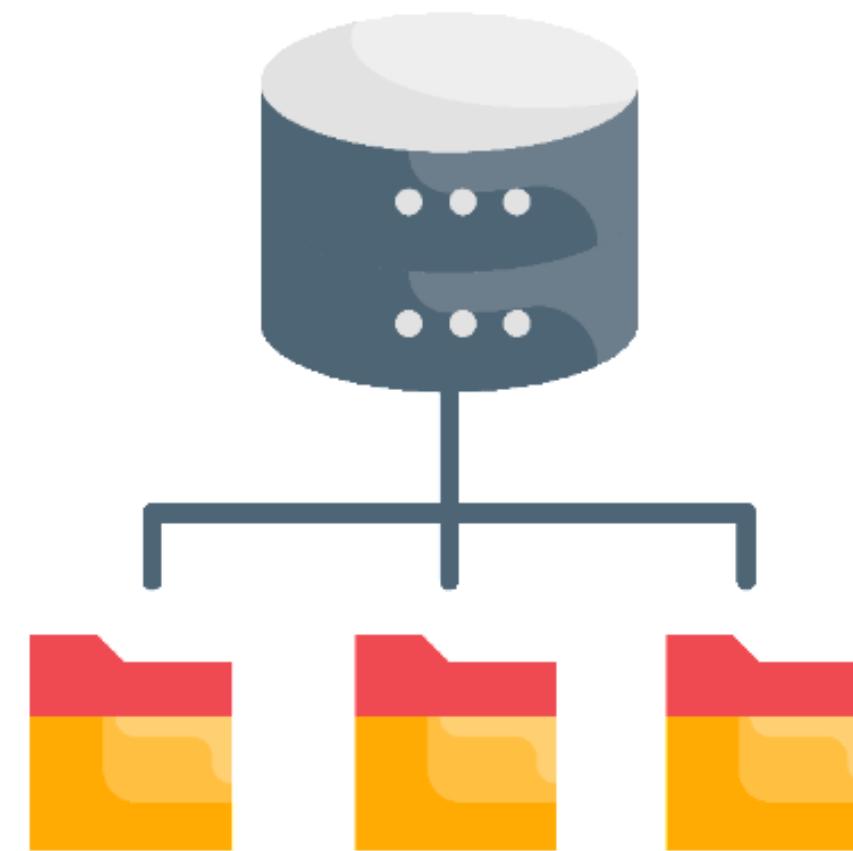
| Product Sales |       |                   |
|---------------|-------|-------------------|
| Product ID    | Sales | Number of Records |
| 6             | 985   | 2                 |
| 7             | 243   | 15                |
| 8             | 652   | 17                |
| 9             | 1123  | 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

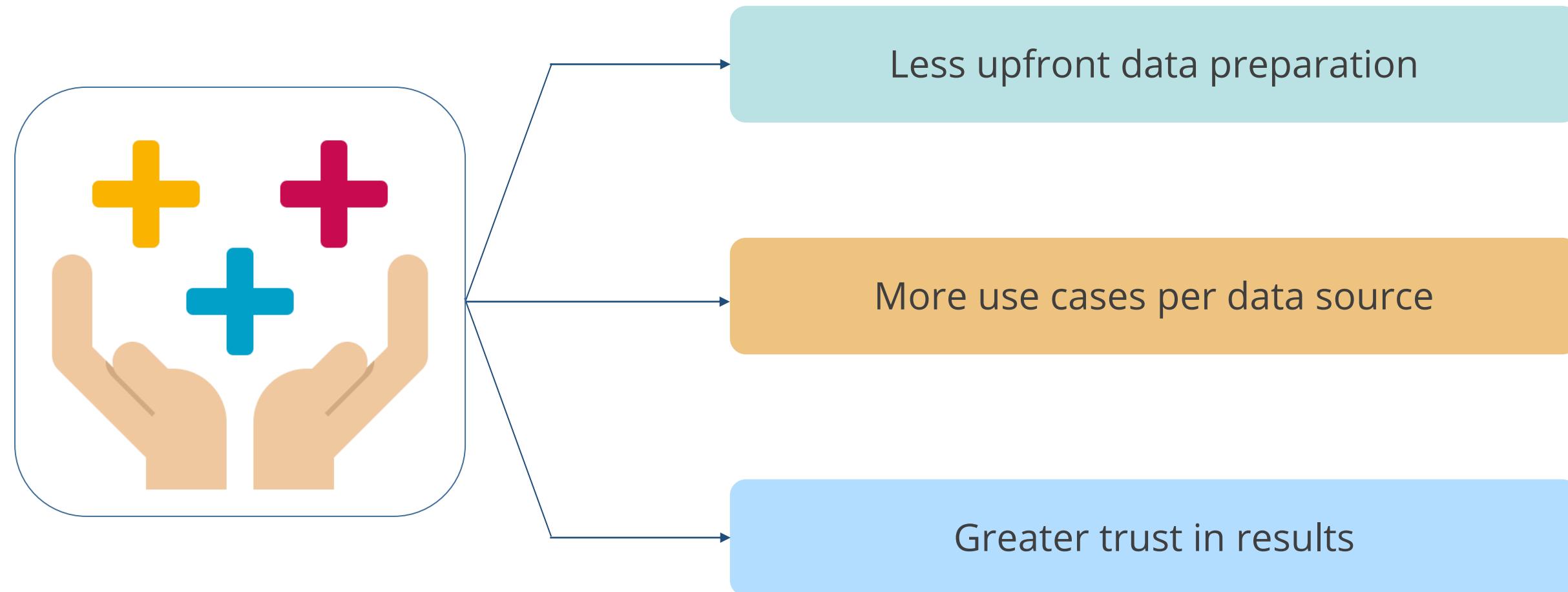
It is a simple and flexible way to link datasets together and is defined based on matching fields.



It can combine fields with non-unique values.

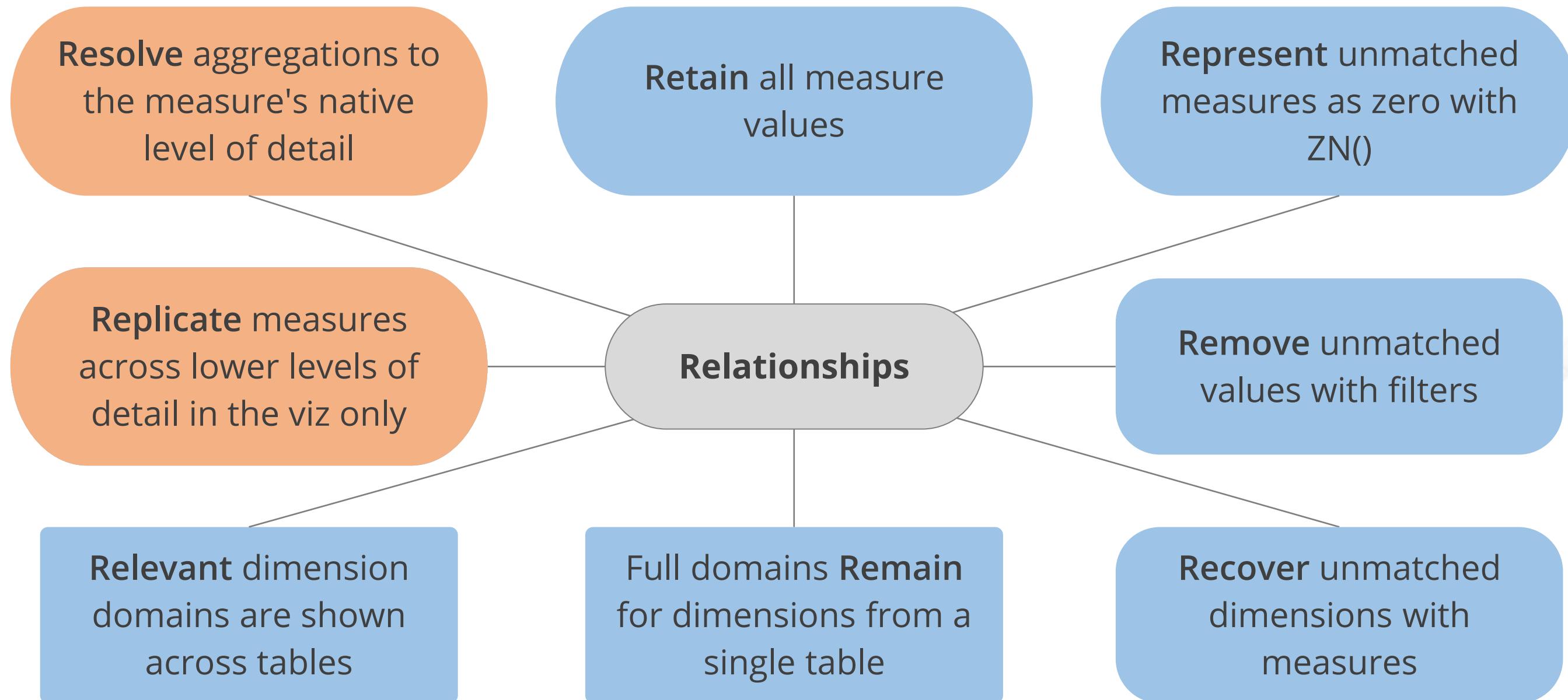
# Relationships

It ensures the following benefits:



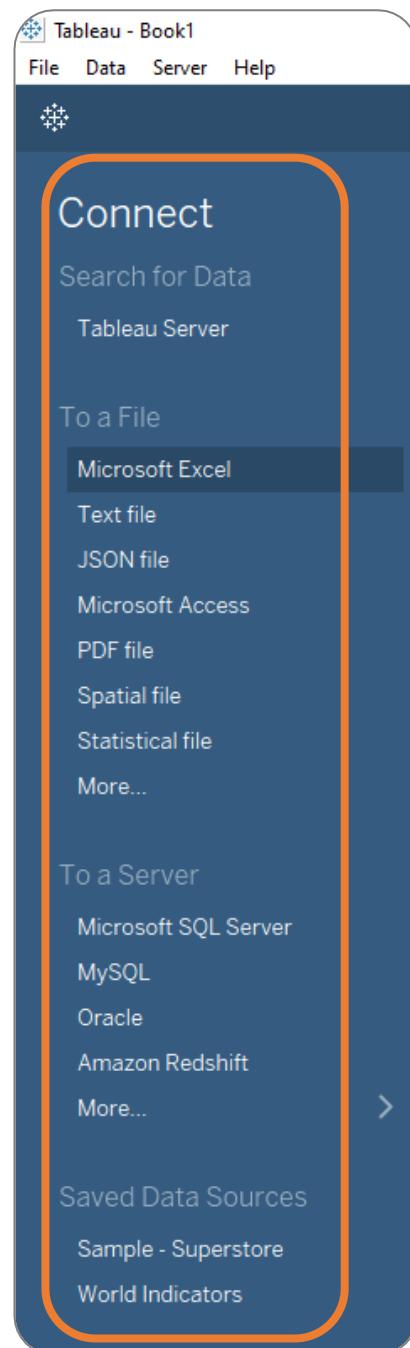
# Relationships

The figure below illustrates the 8Rs of relationship semantics:



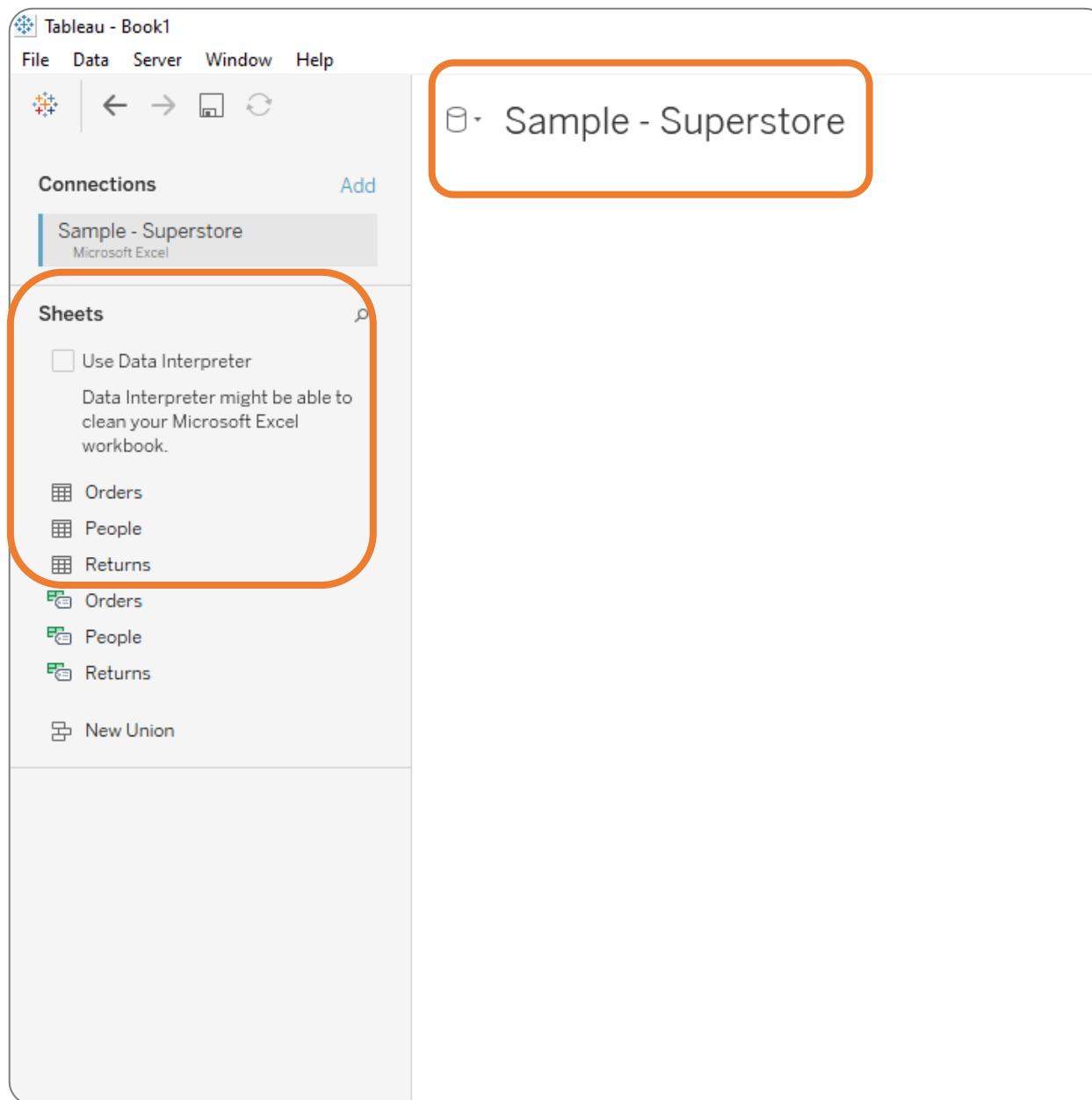
# Steps to Perform Relationships

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

The screenshot shows the Tableau Data Source interface. On the left, the 'Connections' section lists 'Sample - Superstore Microsoft Excel'. Below it, the 'Sheets' section lists several tables: 'Orders', 'People', 'Returns', and three tables from the 'Orders' database ('Orders', 'People', 'Returns'). The 'Orders' table from the main connection is highlighted with an orange border. The 'Orders' table from the 'Orders' database is also highlighted with an orange border. The main workspace displays the 'Orders (Sample - Superstore)' table, which contains a single row labeled 'Orders'. A 'Need more data?' section at the bottom right suggests dragging other tables to relate them.

# Steps to Perform Relationships

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

The screenshot shows the Power BI Relationships pane. On the left, there's a 'Connections' section with 'Sample - Superstore Microsoft Excel' selected. Below it is a 'Sheets' section with 'Orders', 'People', and 'Returns' listed. A red box highlights the area where the 'Orders' and 'Returns' tables are connected to the 'People' table. A yellow callout bubble points to this connection with the text: 'Drag the tables from sheet and drop it in the join area which creates Relationships by default'. In the center, there's a preview pane titled 'Orders — Returns' with a dropdown menu. It shows the relationship criteria: 'Orders' is joined with 'Returns' using the 'Order ID' field as the key. A blue callout bubble points to this area with the text: 'The criteria for the relationships to be defined here and the preview pane on the right side shows the sample data preview'. On the right, there's a preview pane showing a table with columns: '#', 'Orders', 'Abc Orders', and 'Orders'. The data is as follows:

| #      | Orders         | Abc Orders     | Orders     |
|--------|----------------|----------------|------------|
| Row ID | Order ID       | Order ID       | Order Date |
| 1      | CA-2016-152156 | CA-2016-152156 | 08-11-2016 |
| 2      | CA-2016-152156 | CA-2016-152156 | 08-11-2016 |
| 3      | CA-2016-138688 | CA-2016-138688 | 12-06-2016 |
| 4      | US-2015-108966 | US-2015-108966 | 11-10-2015 |
| 5      | US-2015-108966 | US-2015-108966 | 11-10-2015 |
| 6      | CA-2014-115812 | CA-2014-115812 | 09-06-2016 |
| 7      | CA-2014-115812 | CA-2014-115812 | 09-06-2016 |
| 8      | CA-2014-115812 | CA-2014-115812 | 09-06-2016 |

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 |

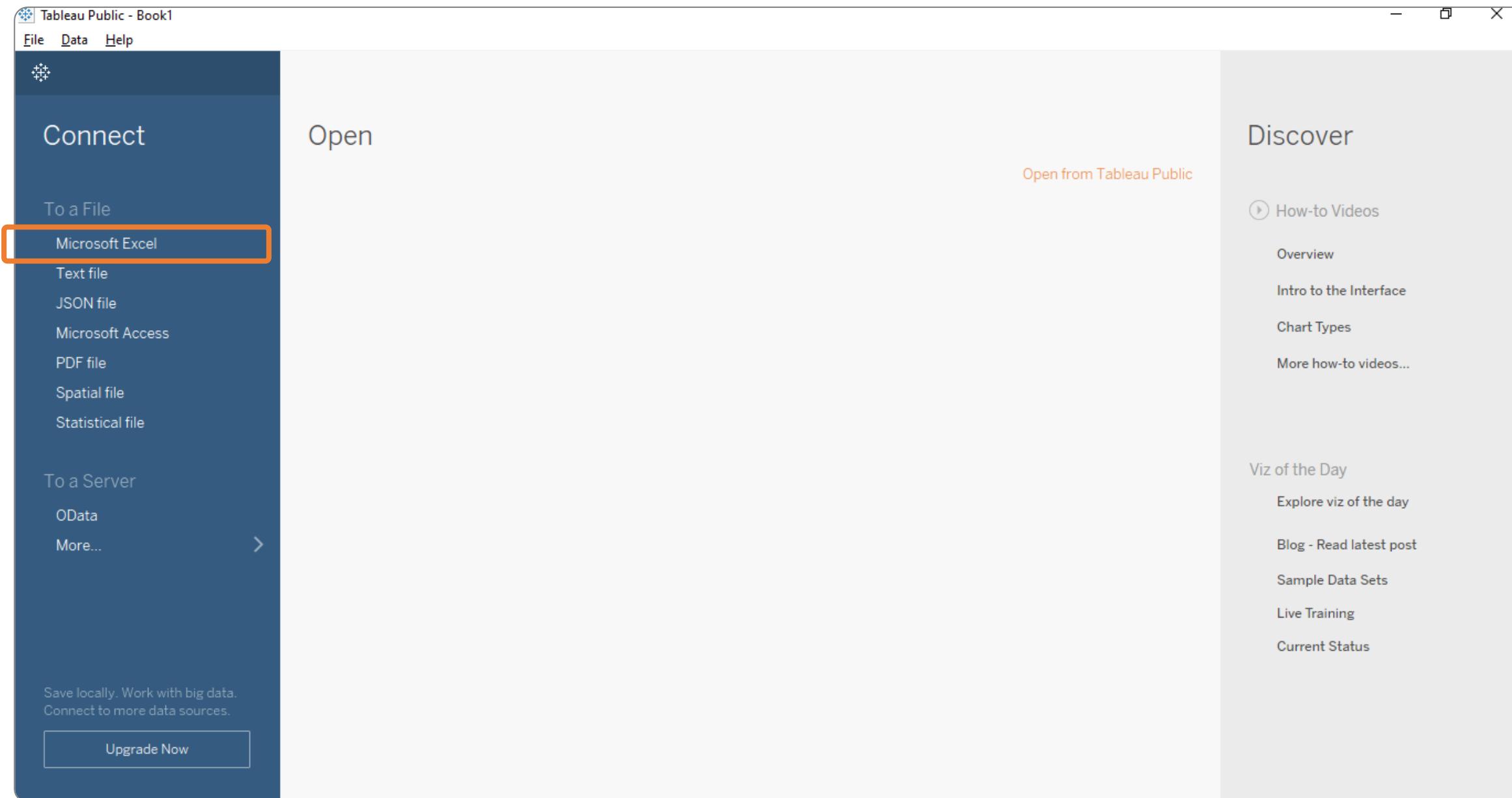
| Region  | Category        | Profit    | Sales     |
|---------|-----------------|-----------|-----------|
| Central | Furniture       | (\$2,871) | \$163,797 |
|         | Office Supplies | \$8,880   | \$167,026 |
|         | Technology      | \$33,697  | \$170,416 |

| 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 |

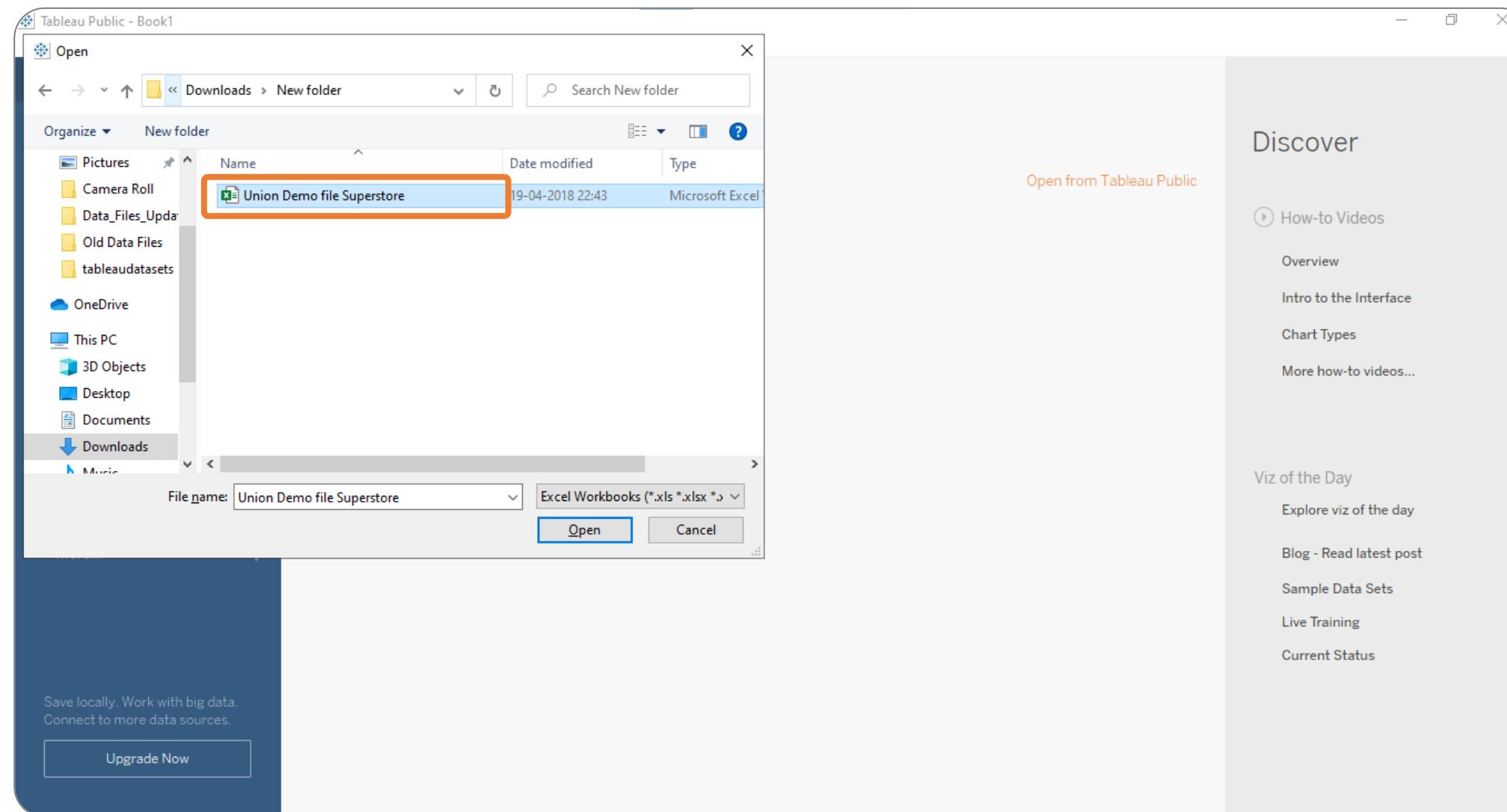
# Steps to Perform Union

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



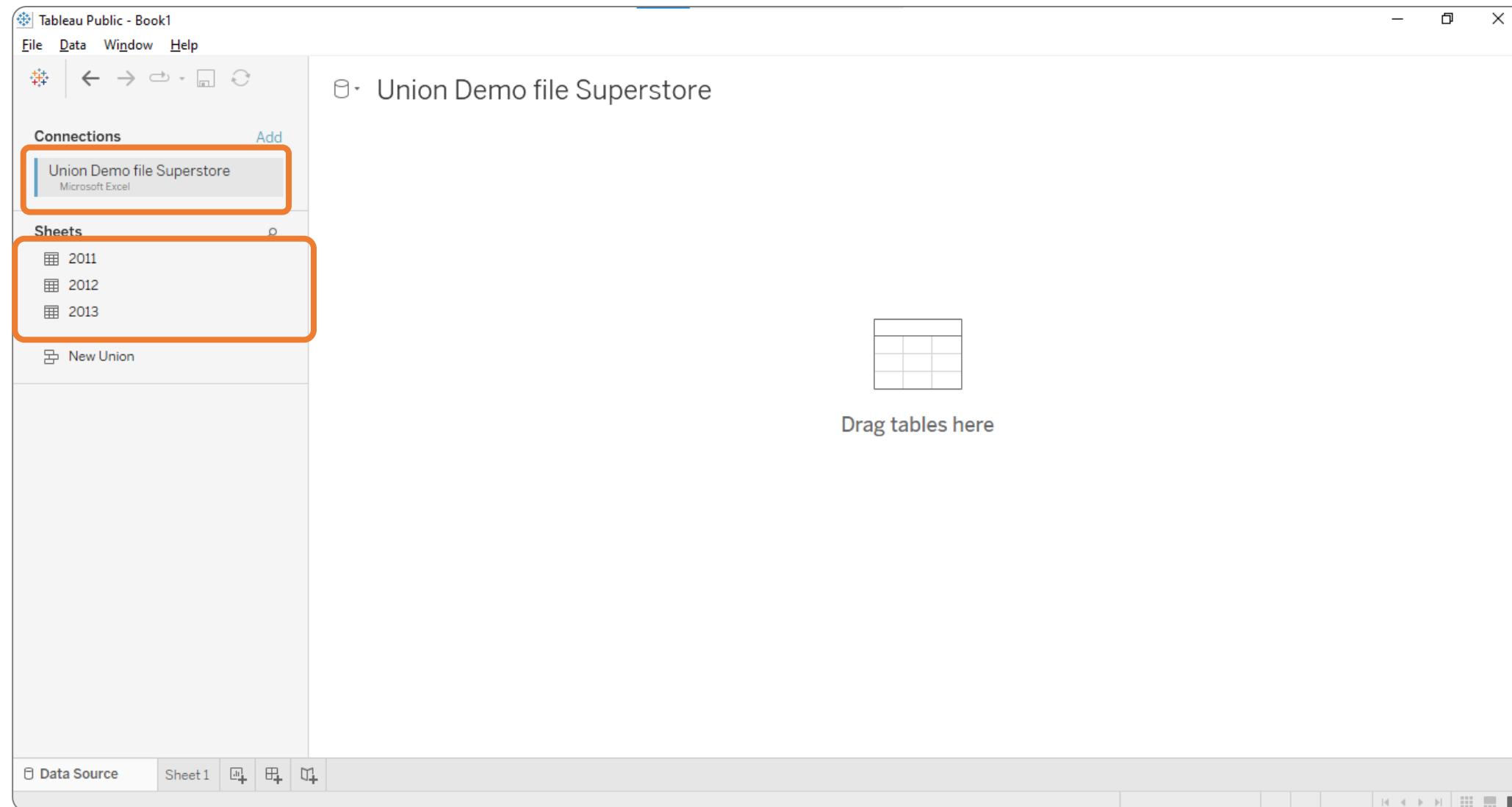
# Steps to Perform Union

## 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

The screenshot shows the Tableau Public interface with the following details:

- Connections:** Union Demo file Superstore (Microsoft Excel) is connected.
- Sheets:** The '2013' sheet is selected and highlighted with an orange border.
- Table Preview:** The '2013' table is displayed with 8 fields and 3994 rows. A portion of the data is shown:

| Country       | State      | Region | Category        | Sub-Category | Product Name    |
|---------------|------------|--------|-----------------|--------------|-----------------|
| United States | Idaho      | West   | Office Supplies | Paper        | Xerox Color Co  |
| United States | Arizona    | West   | Furniture       | Furnishings  | Electrix Incand |
| United States | Arizona    | West   | Office Supplies | Binders      | Cardinal Slant  |
| United States | Florida    | South  | Office Supplies | Binders      | Avery Trapezoid |
| United States | California | West   | Furniture       | Furnishings  | C-Line Magnet   |

# Steps to Perform Union

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

The screenshot shows the Tableau Public interface with the following details:

- Title Bar:** Tableau Public - Book1
- Menu Bar:** File, Data, Window, Help
- Connections:** Union Demo file Superstore (Microsoft Excel)
- Sheets:** 2011, 2012, 2013, New Union
- Current Sheet:** 2013 (Union Demo file Superstore)
- Tableau View:** A grid of data with columns: Country, State, Region, Category, Sub-Category, Product Name.
- Toolbars:** Top toolbar with zoom and refresh icons; bottom toolbar with data source, sheet, and other options.
- Bottom Panel:** Shows the name "2013" and fields: Type, Field Name, Physical Table, Remote Field... for the "Country" field.
- Central Area:** Displays the data grid with rows of data from the 2013 table.
- Pop-up:** A small orange-bordered box highlights the "Union" button in the top-left corner of the 2012 table area.
- Text Overlay:** "Need more data? Drag tables here to relate them. Learn more"

# Steps to Perform Union

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

The screenshot shows the Tableau Public interface with a union operation. On the left, the 'Connections' pane shows 'Union Demo file Superstore Microsoft Excel'. The 'Sheets' pane lists '2011', '2012', and '2013'. A large orange arrow points to the '2013' entry, which is highlighted with a red box. The main workspace displays a unioned table named '2013+ (Union Demo file Superstore)'. Below it, a smaller table named '2013' is shown with 10 fields and 6622 rows. The bottom navigation bar includes 'Data Source', 'Sheet 1', and various icons.

| 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    |

# Steps to Perform Union

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

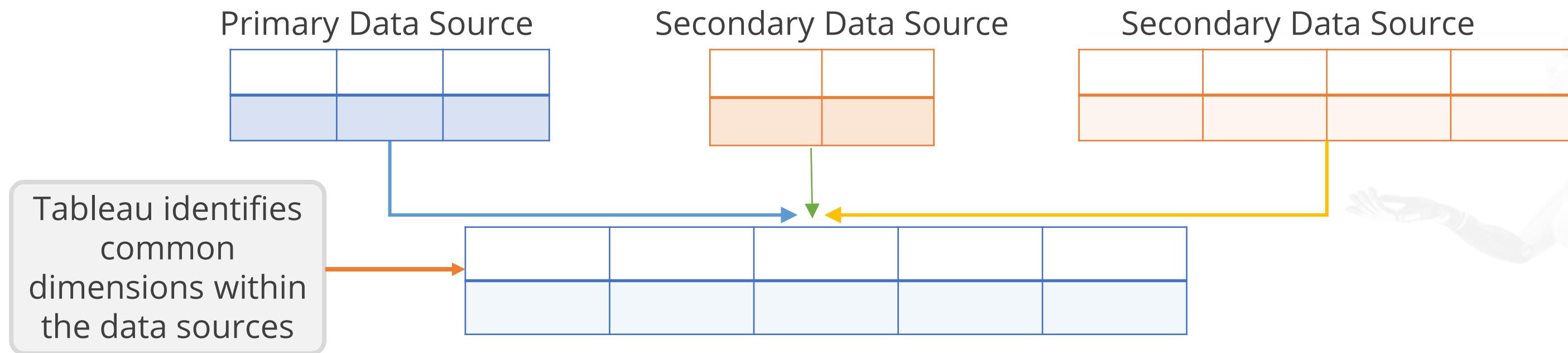
The screenshot shows the Tableau Public interface with a single data source named "Union Demo file Superstore". The data source is connected to Microsoft Excel and contains 10 fields and 6622 rows. The current view is a table titled "2013+ (Union Demo file Superstore)" which displays data from the year 2013. The columns include Country, State, Region, Category, Sub-Category, Product Name, Sales, Profit, Sheet, and Table Number. The data consists of various product sales across different categories like Furniture, Technology, and Office Supplies, with specific details such as Hon 94000 Series Round Table and VTECH DS6151.

| 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

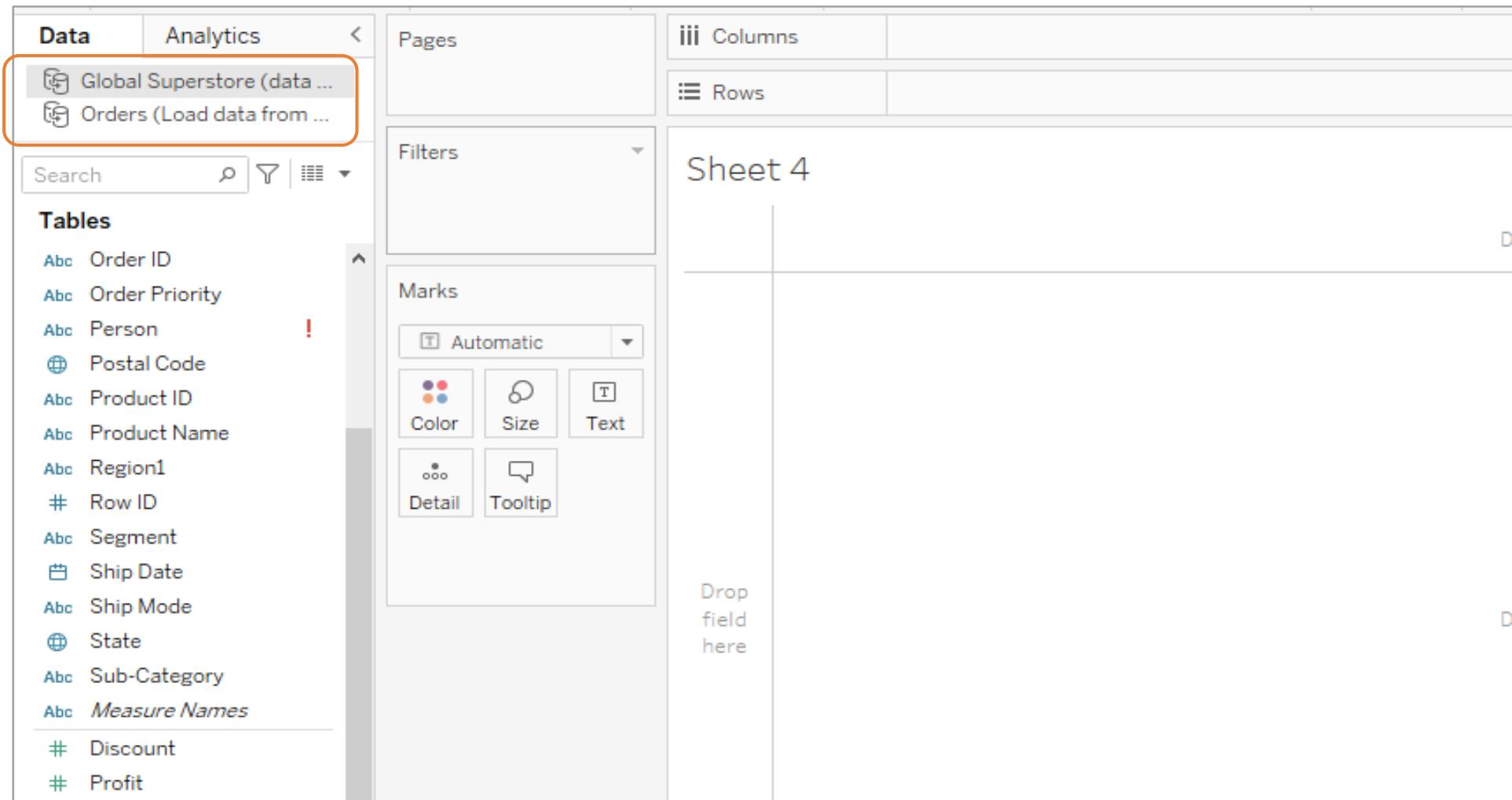
# Data Blending

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

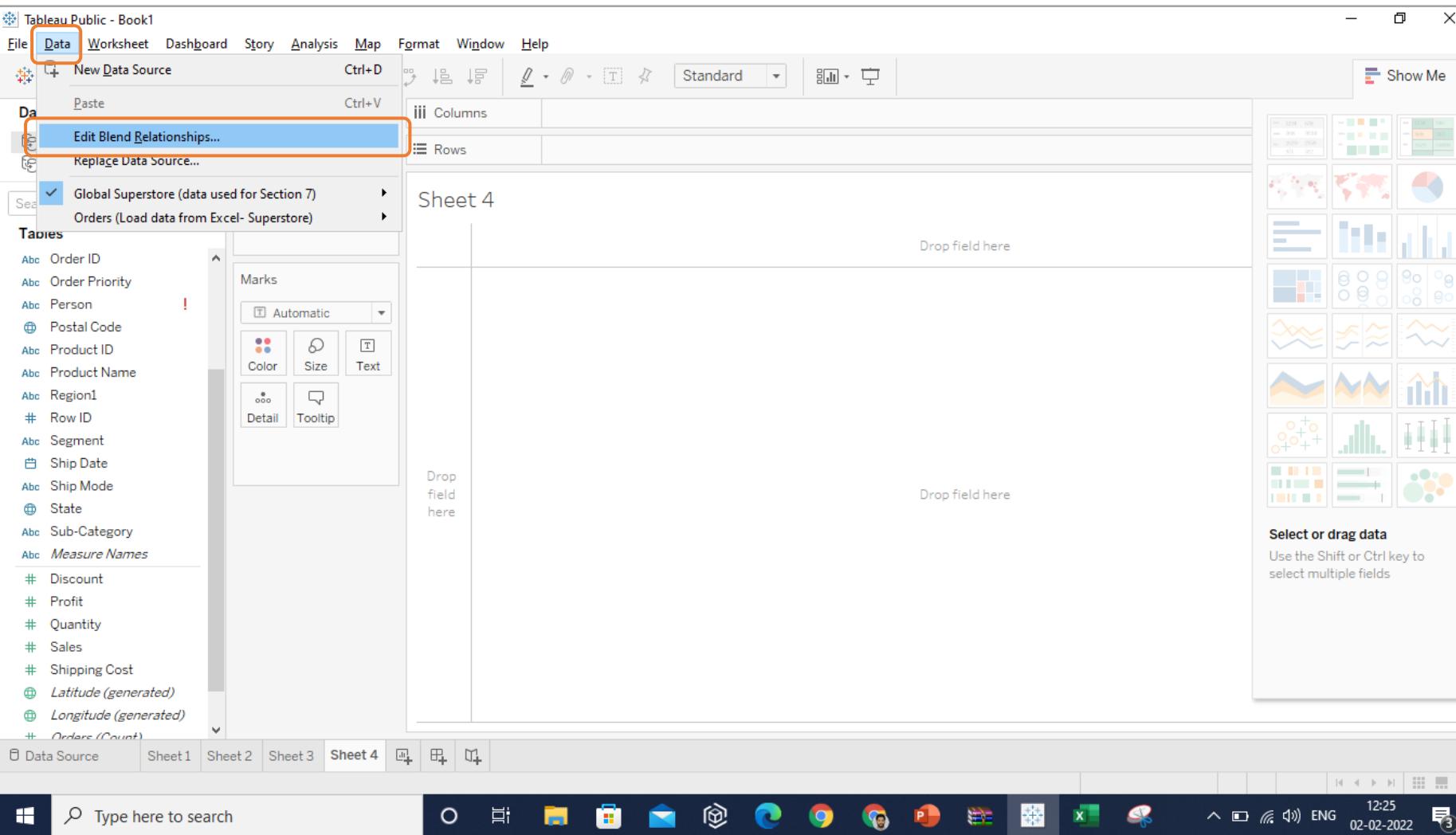


The screenshot shows the Tableau interface with the 'Data' tab selected. In the top left, there are two entries: 'Global Superstore (data ...)' and 'Orders (Load data from ...)', both highlighted with an orange border. Below these, the 'Tables' section lists various dimensions and measures from both datasets. The 'Marks' shelf on the right is set to 'Automatic' and includes options for Color, Size, Text, Detail, and Tooltip. The main workspace is titled 'Sheet 4' and contains a single text placeholder 'Drop field here'.

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

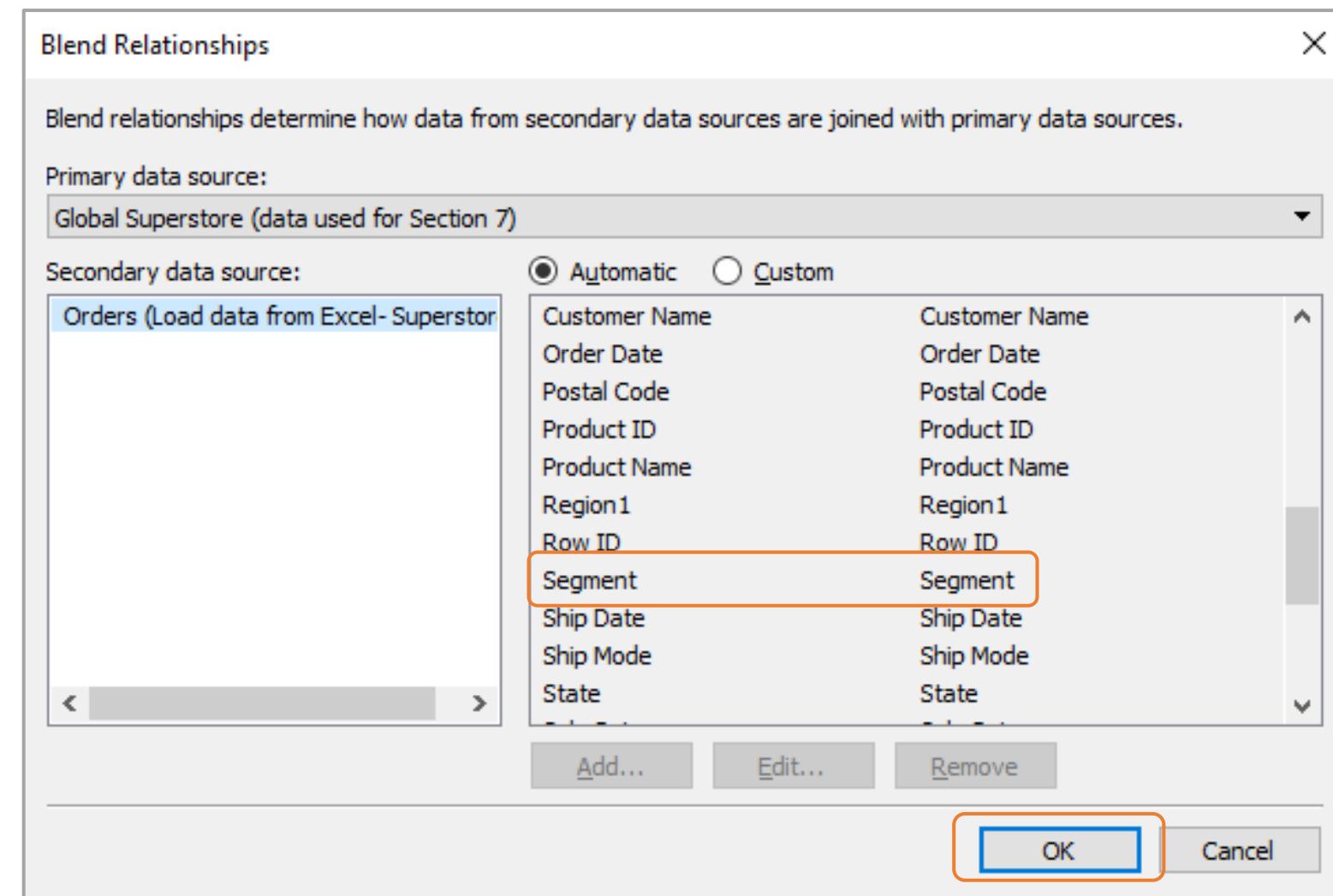
# Steps to Perform Data Blending

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



# Steps to Perform Data Blending

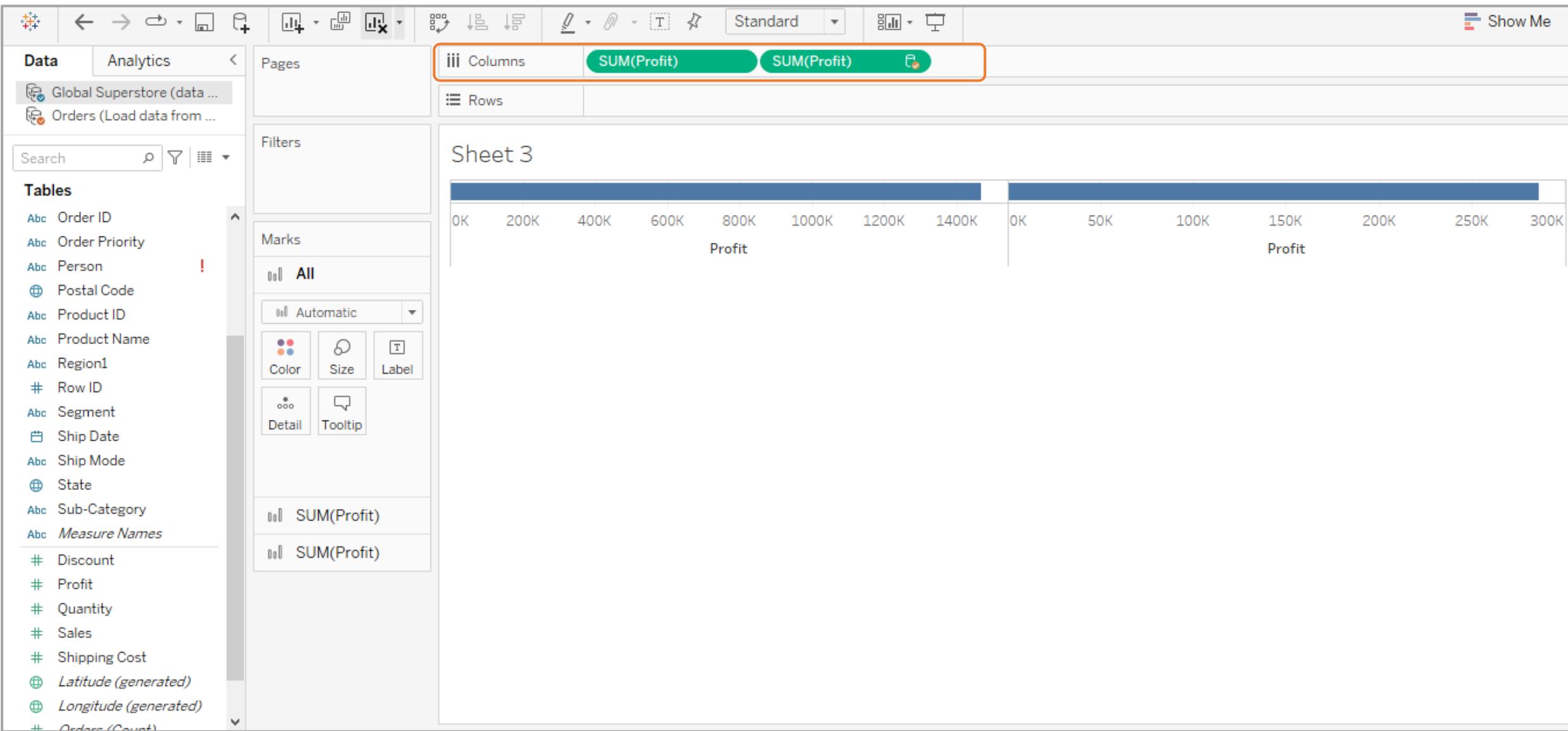
Step 3: In the **Edit Blend Relationship** dialog box, ensure that **Segment** is mapped with **Segment**



Click on **OK**

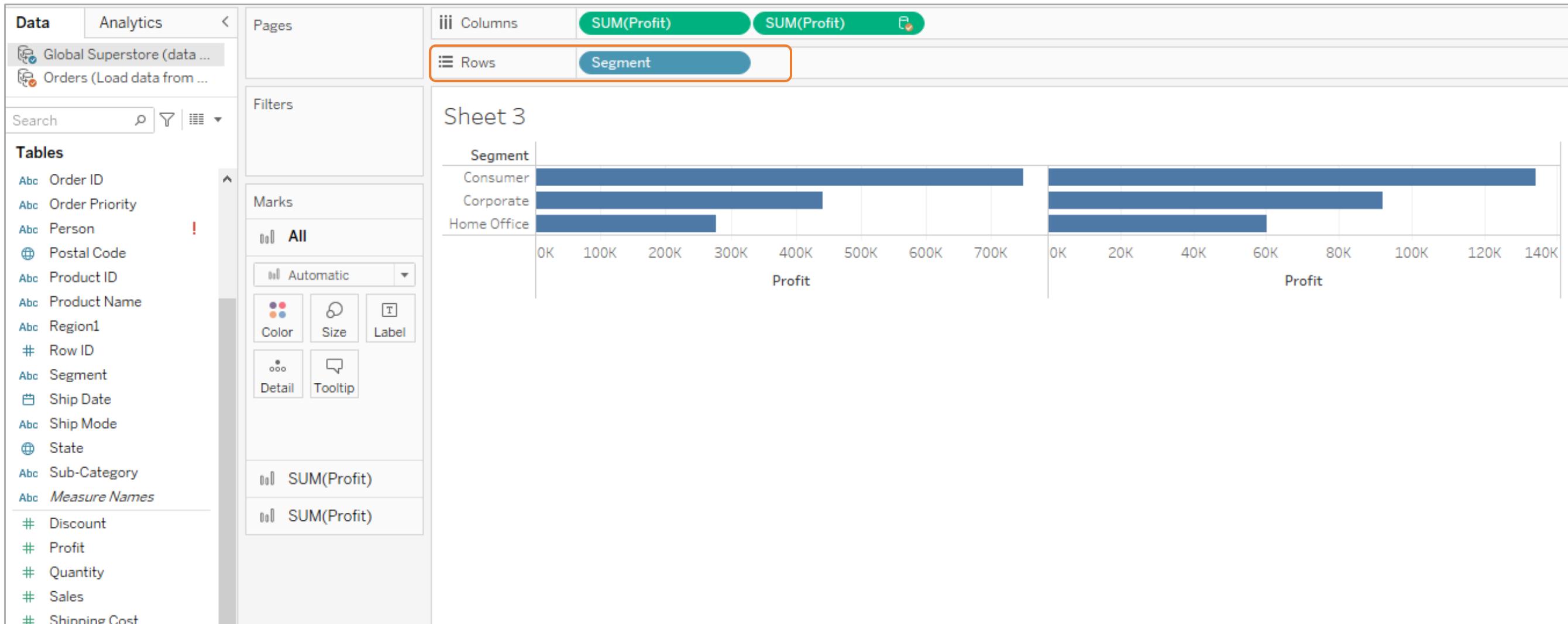
# Steps to Perform Data Blending

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.

## Aggregations

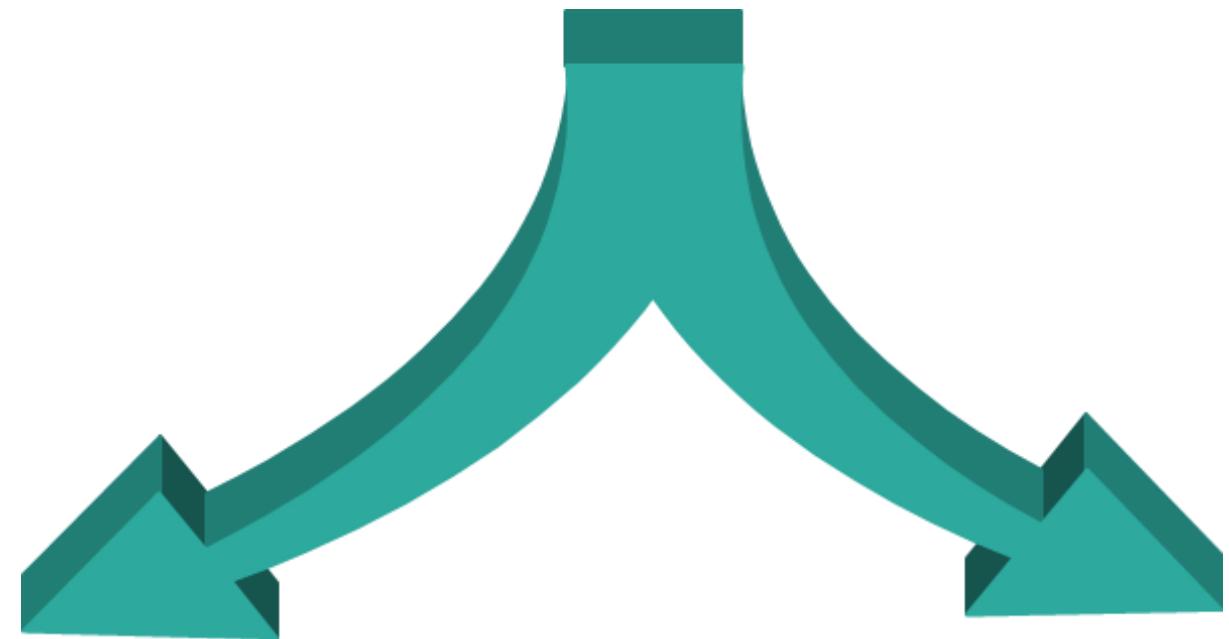
# Aggregations

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



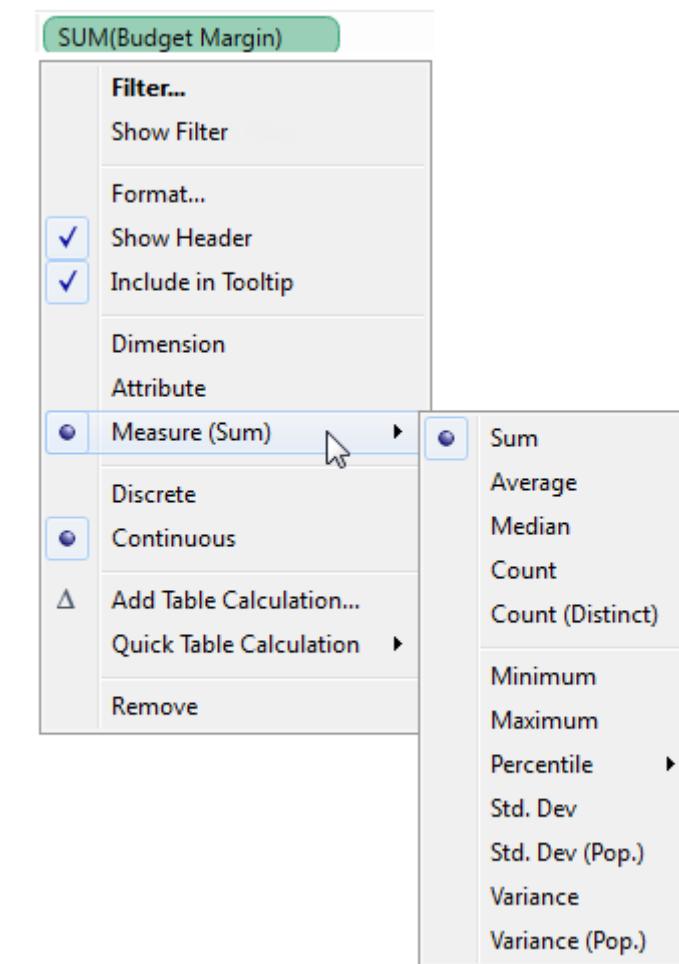
# Aggregations

Data aggregation is performed on:



# Aggregating Measures

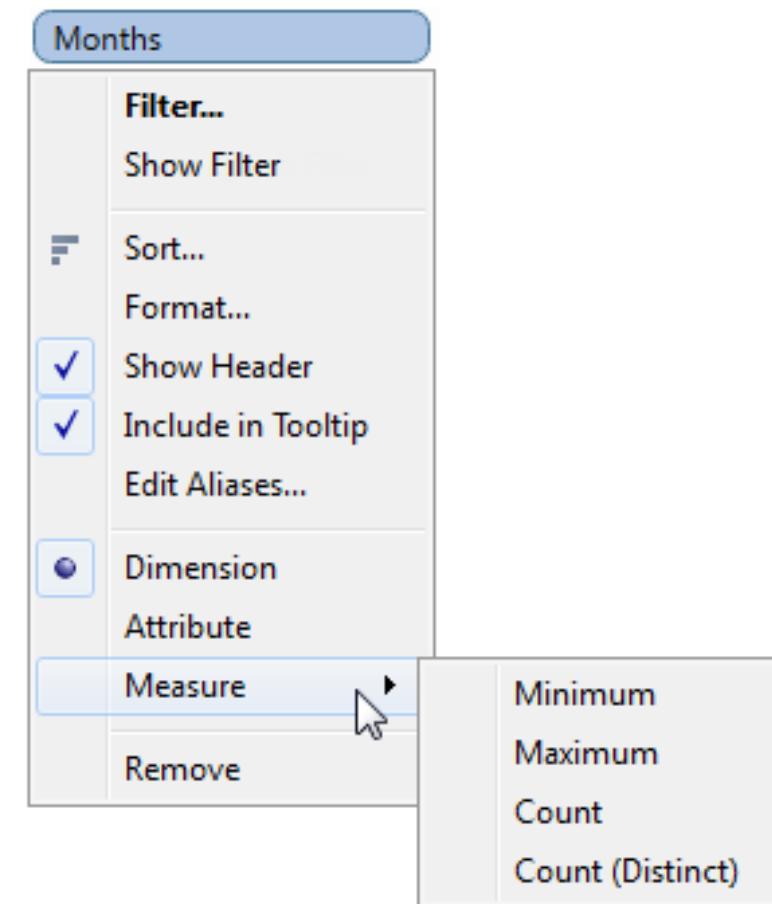
The aggregation of measures automatically aggregates their values when a measure is added to the view.



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

# Aggregating Dimensions

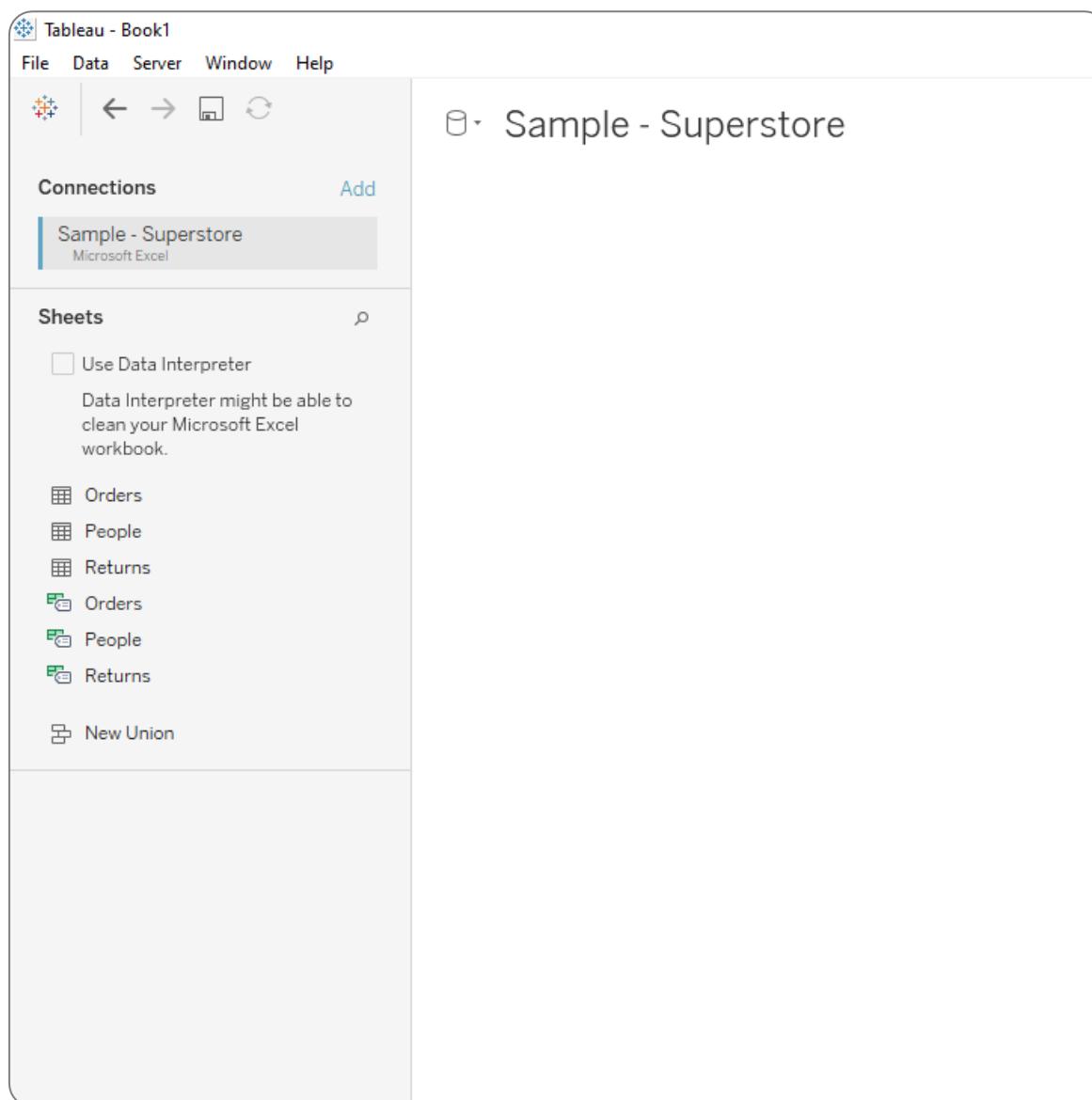
The aggregation of dimensions creates a new temporary measure column.



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

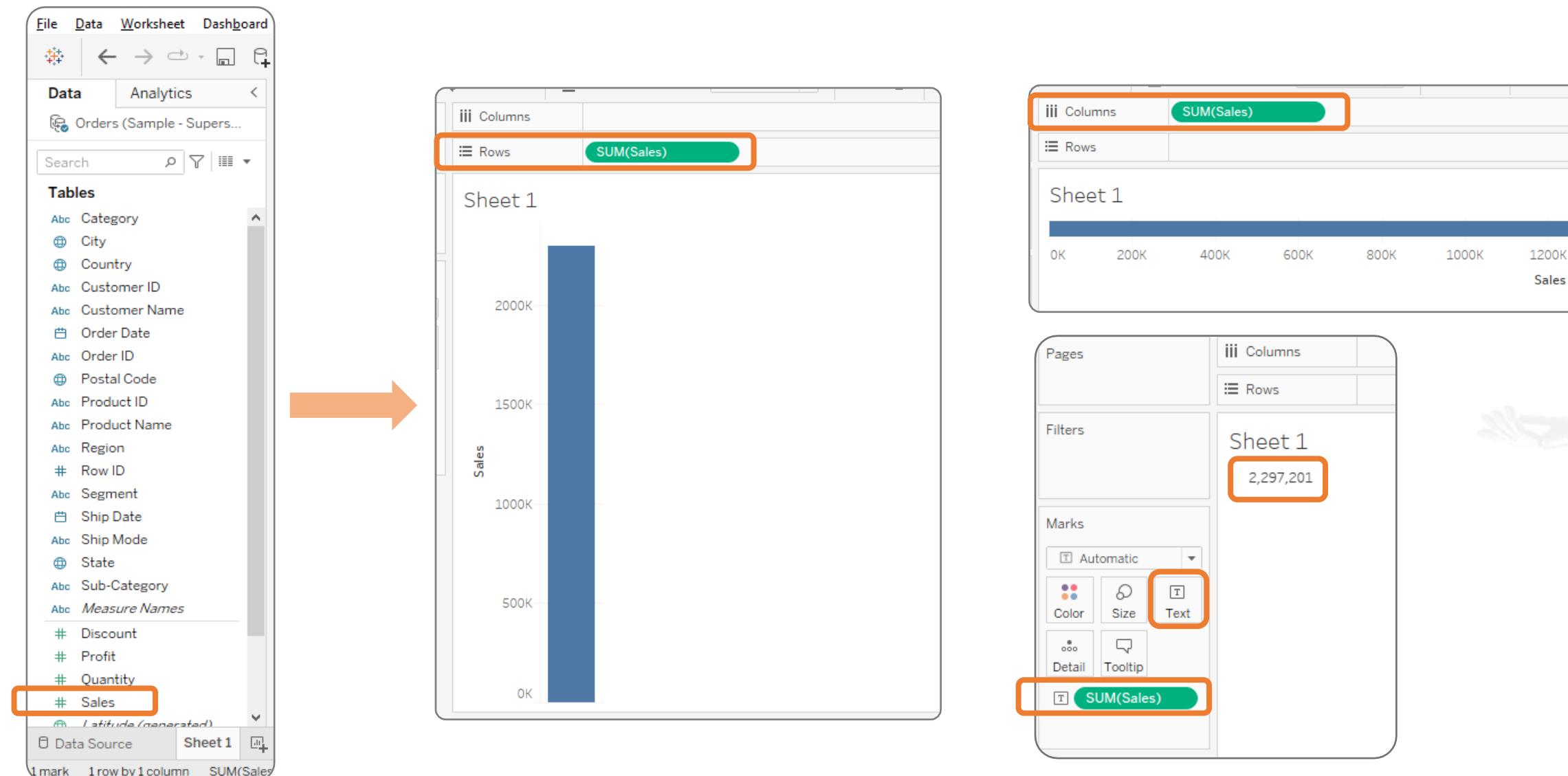
# Steps to Perform Aggregations

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



# Steps to Perform Aggregations

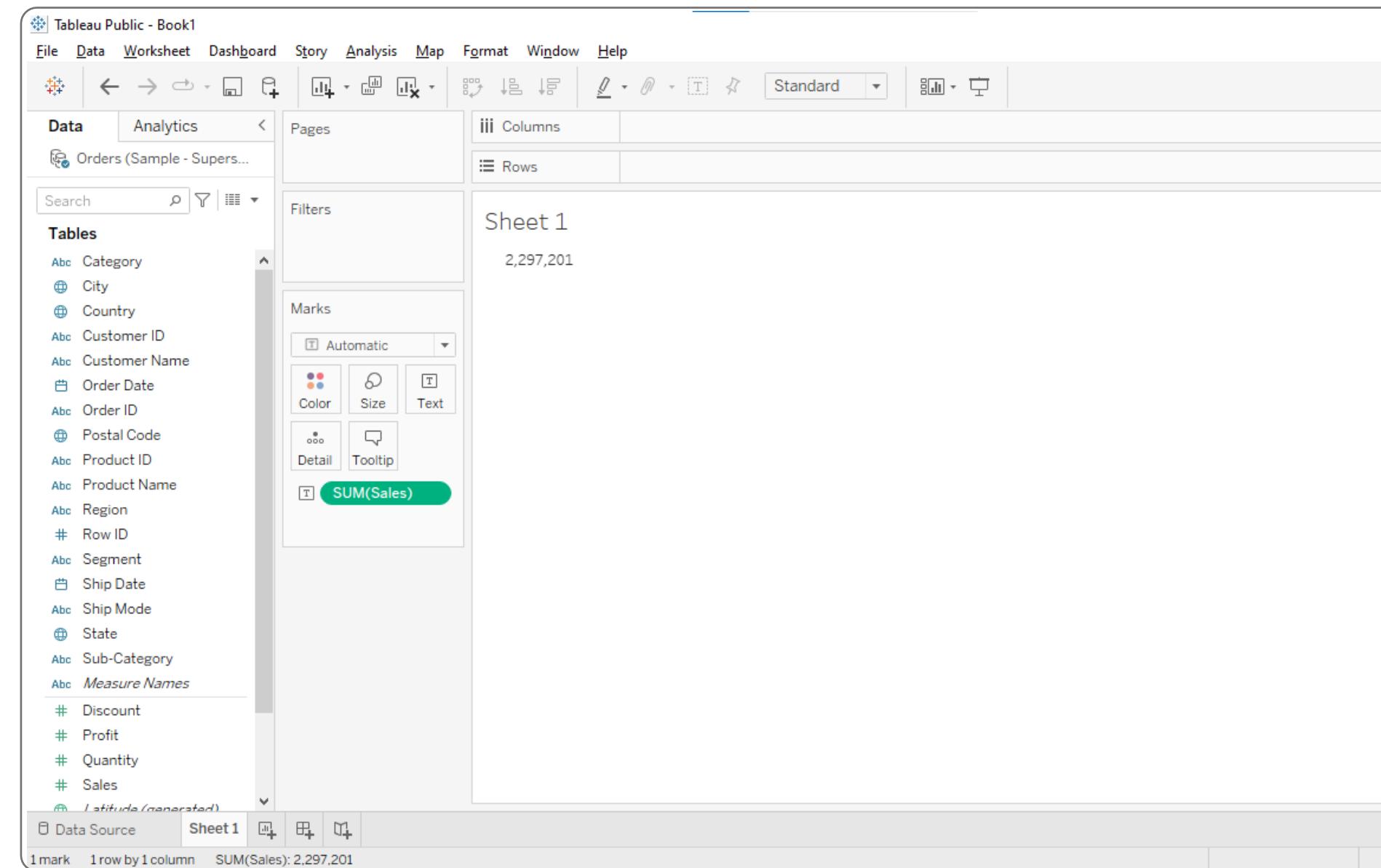
Aggregation can be performed by dragging **Sales** to **Rows**, **Columns**, or **Text** under **Marks**.



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.



The screenshot shows the Tableau Public interface with the following details:

- Top Bar:** File, Data, Worksheet, Dashboard, Story, Analysis, Map, Format, Window, Help.
- Data Pane:** Shows the "Orders (Sample - Superstore)" data source. The "Tables" section lists various dimensions and measures, including Category, City, Country, Customer ID, Customer Name, Order Date, Order ID, Postal Code, Product ID, Product Name, Region, Row ID, Segment, Ship Date, Ship Mode, State, Sub-Category, Measure Names, and Latitude (generated). The "Measure Names" section is currently selected, showing four measures: Discount, Profit, Quantity, and Sales. The "Sales" measure is highlighted with a green background.
- Marks Shelf:** Displays options for Color, Size, Text, Detail, and Tooltip. The "SUM(Sales)" measure is listed under the "Text" category.
- Sheet 1:** The title "Sheet 1" is at the top, followed by the value "2,297,201".
- Bottom Status Bar:** Shows "1 mark", "1 row by 1 column", and "SUM(Sales): 2,297,201".

# Steps to Perform Aggregations

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

The screenshot shows the Tableau Public interface. On the left, the 'Data' pane lists various dimensions and measures. In the center, the 'Marks' shelf has a single item: 'SUM(Sales)', which is highlighted with an orange border. The main workspace is labeled 'Sheet 1' and displays the value '2,297,201'. The top menu bar includes options like File, Data, Worksheet, Dashboard, Story, Analysis, Map, Format, Window, and Help.

# Steps to Perform Aggregations

## Step 2: Click on Measure (Sum)

The screenshot shows the Tableau Public interface with a context menu open over a measure named "SUM(Sales)". The menu is titled "Measure (Sum)" and includes options like Sum, Average, Median, Count, etc. The "Sum" option is selected. The background shows a dashboard with a single card displaying the value "2,297,201".

Tableau Public - Book1

File Data Worksheet Dashboard Story Analysis Map Format Window Help

Data Analytics

Orders (Sample - Supers...)

Search

Tables

- Abc Category
- Abc City
- Abc Country
- Abc Customer ID
- Abc Customer Name
- Abc Order Date
- Abc Order ID
- Abc Postal Code
- Abc Product ID
- Abc Product Name
- Abc Region
- # Row ID
- Abc Segment
- Abc Ship Date
- Abc Ship Mode
- Abc State
- Abc Sub-Category
- Abc Measure Names
- # Discount
- # Profit
- # Quantity
- # Sales
- Latitude (generated)

Marks

Automatic

Color Size Text

Detail Tooltip

SUM(Sales)

- Filter...
- Show Filter
- Format...
- Include in Tooltip
- Dimension
- Attribute
- Measure (Sum)
- Discrete
- Continuous
- Edit in Shelf
- Add Table Calculation...
- Quick Table Calculation...

Sheet 1

2,297,201

Show Me

For horizontal bars try

0 or more Dimensions

1 or more Measures

# Steps to Perform Aggregations

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

The screenshot shows the Tableau Public interface with the following details:

- Top Bar:** File, Data, Worksheet, Dashboard, Story, Analysis, Map, Format, Window, Help.
- Data Source:** Orders (Sample - Superstore).
- Sheet 1:** Displays the count 2,297,201.
- Marks Shelf:** SUM(Sales) is selected.
- Context Menu (Open at SUM(Sales)):**  A context menu is open, with the "Measure (Sum)" option highlighted and its submenu open. The submenu includes:
  - Sum (selected)
  - Average
  - Median
  - Count
  - Count (Distinct)
  - Discrete
  - Continuous
  - Edit in Shelf
  - Add Table Calculation...
  - Quick Table Calculation
- Show Me:** A panel on the right showing various visualization options.

# Steps to Perform Aggregations

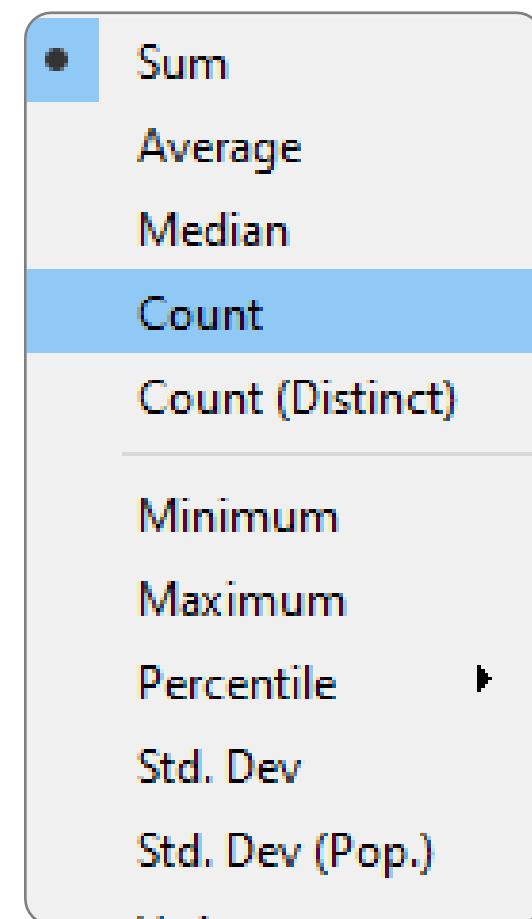
It will be changed to the desired aggregation.

The screenshot shows the Tableau Public interface with the following details:

- Top Bar:** File, Data, Worksheet, Dashboard, Story, Analysis, Map, Format, Window, Help.
- Data Source:** Orders (Sample - Superstore).
- Left Panel (Data View):** Shows various dimensions and measures. The measure **Sales** is highlighted with an orange box.
- Middle Panel (Marks View):** Shows the current mark type is **Automatic**. The **AVG(Sales)** button is highlighted with an orange box.
- Right Panel (Show Me):** Displays a grid of visualization preview icons. A horizontal bar chart icon is highlighted with an orange box.
- Bottom Status Bar:** 1 mark, 1 row by 1 column, SUM of AVG(Sales): 229.9.

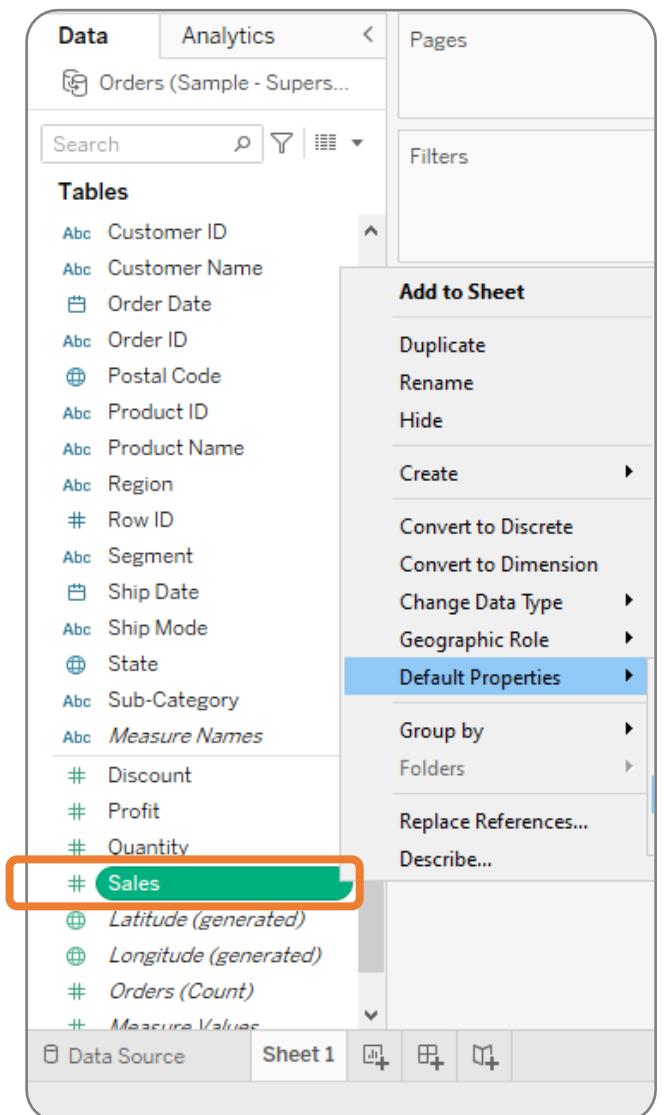
# Steps to Perform Aggregations

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



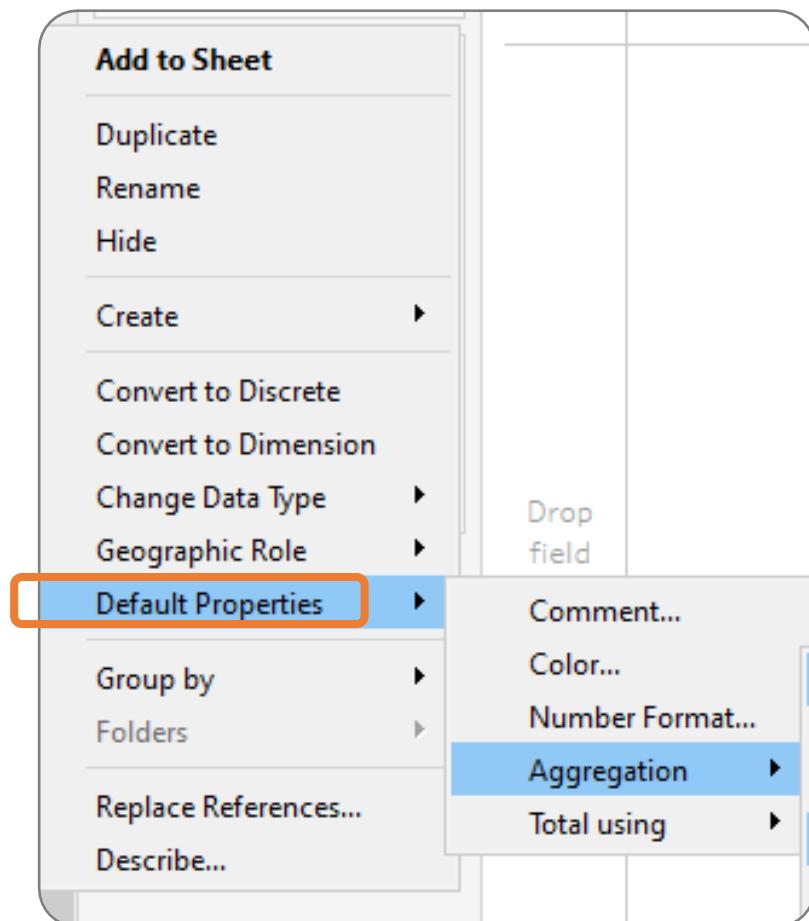
# Steps to Perform Aggregations

## Step 1: Click on the **Sales**



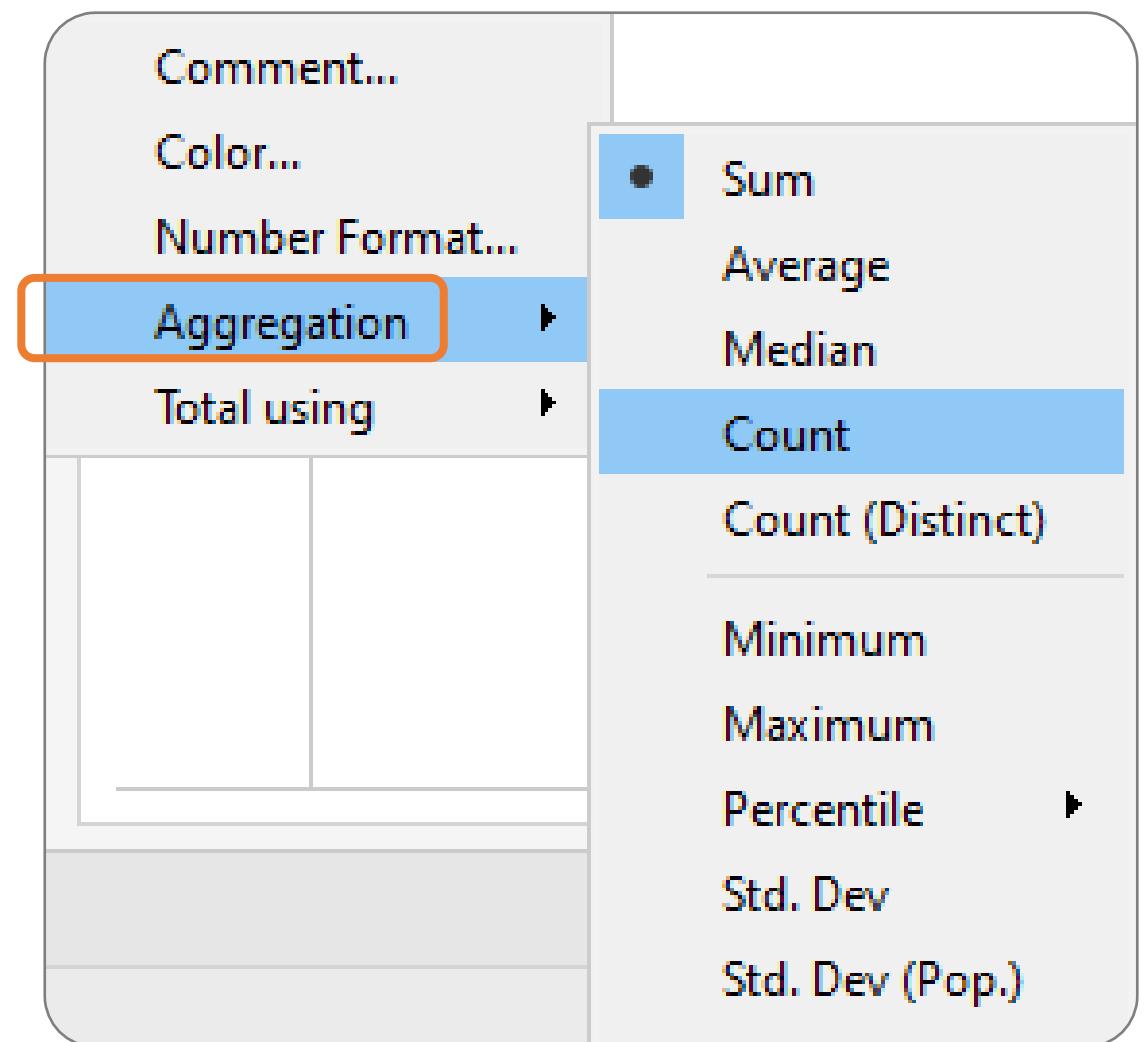
# Steps to Perform Aggregations

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



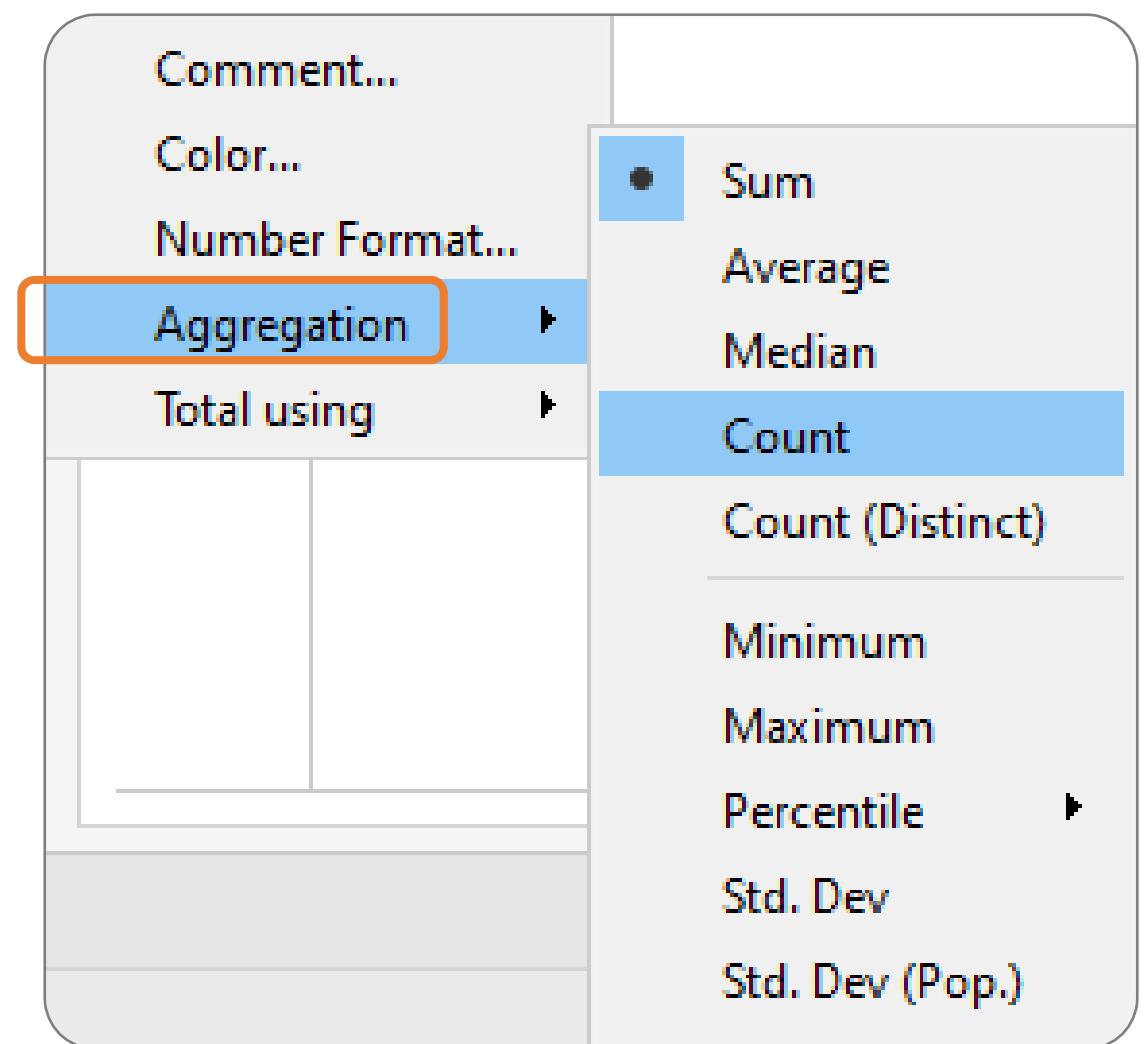
# Steps to Perform Aggregations

## Step 3: Click on Aggregation



# Steps to Perform Aggregations

Step 4: Select the desired aggregation to make it default



# DATA AND ARTIFICIAL INTELLIGENCE

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<br>Pivot<br>Product | #<br>Pivot<br>Fy2017 | #<br>Pivot<br>Fy2018 | #<br>Pivot<br>Fy2019 | #<br>Pivot<br>Fy2020 | #<br>Pivot<br>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<br>Pivot<br>Product | Abc<br>Pivot1<br>Pivot Field Names | #<br>Pivot1<br>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:

The screenshot shows the Microsoft Power BI desktop application. On the left, the 'Connections' pane displays a single connection named 'Sample Data (Microsoft Excel)'. Below it, the 'Sheets' pane lists several resources: 'Additional Resource List', 'Data Interpretation', 'Experience', 'Pivot', 'Res\_List', and 'New Union'. A note about 'Data Interpreter' is present, stating it might be able to clean the workbook. The main workspace is titled 'Sample\_Data' and contains a 'Pivot' visualization. The visualization interface includes a 'Need more data?' section with a 'Drag tables here to relate them.' message and a link to 'Learn more'. Below this is a table titled 'Pivot' with 6 fields and 7 rows. The table details the sales quantity for various products across five fiscal years (Fy2017 to Fy2021). The data is as follows:

| 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

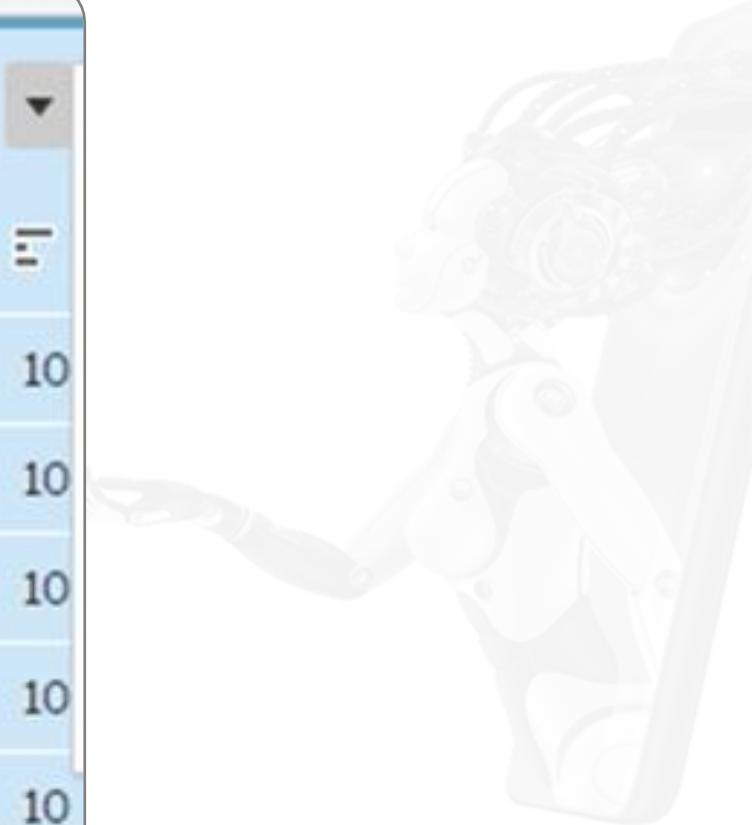
Though the data looks fine, the format is not so appropriate for Tableau to evaluate all items.

The screenshot shows the Microsoft Power BI desktop application. On the left, the 'Connections' pane shows a single connection named 'Sample Data (Microsoft Excel)'. The 'Sheets' pane lists several sheets: 'Additional Resource List', 'Data Interpretation', 'Experience', 'Pivot', 'Res\_List', and 'New Union'. The main workspace displays a 'Pivot' table titled 'Pivot' with 6 fields and 7 rows. The columns represent years from Fy2017 to Fy2021, and the rows list products: Phone, Television, PlayStation, Home Theater, Router, Smart Watch, and Laptop. The data values are as follows:

| 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

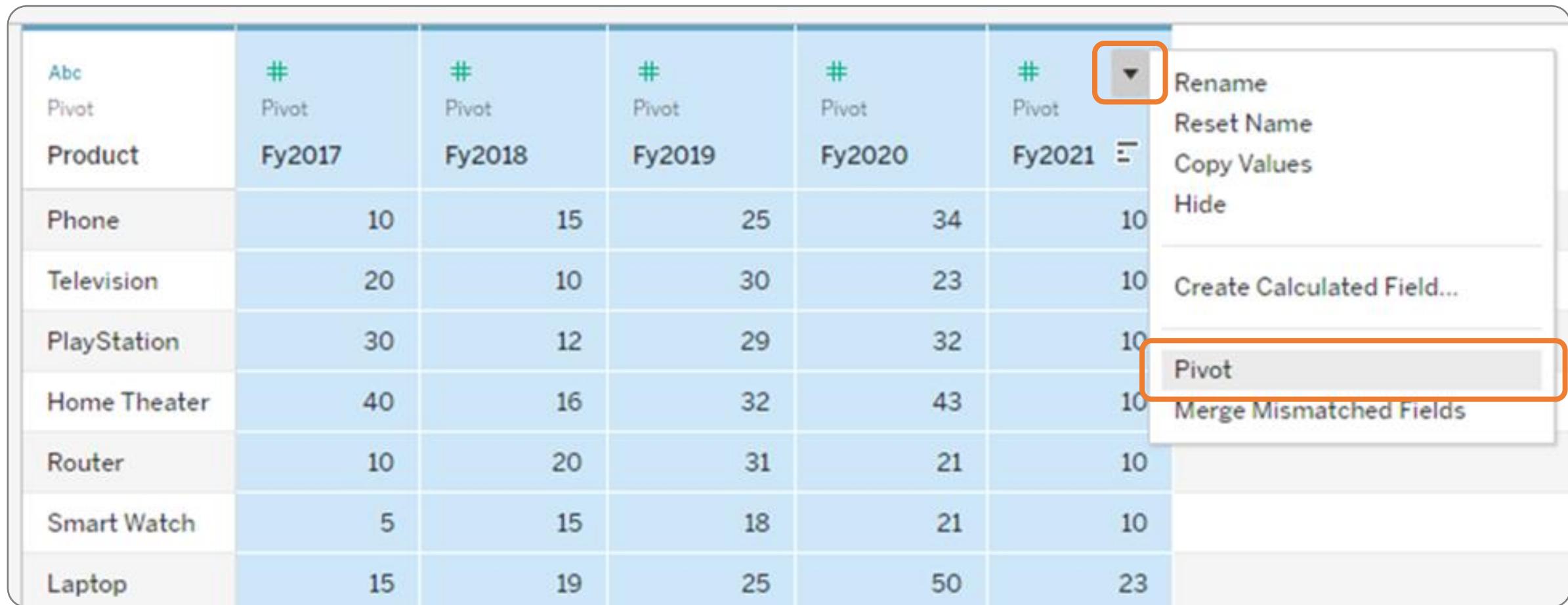
Step 1: Select all the columns in the file



| Abc<br>Pivot | #<br>Pivot | #<br>Pivot | #<br>Pivot | #<br>Pivot | #<br>Pivot | ▼ |
|--------------|------------|------------|------------|------------|------------|---|
| 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

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



The screenshot shows a spreadsheet with data for various products across five fiscal years (Fy2017 to Fy2021). A context menu is open over the 'Pivot' field in the Fy2021 column. The menu items are: Rename, Reset Name, Copy Values, Hide, Create Calculated Field..., Pivot (which is highlighted with an orange rectangle), and Merge Mismatched Fields.

| Abc<br>Pivot | #<br>Pivot | #<br>Pivot | #<br>Pivot | #<br>Pivot | #<br>Pivot | ▼ |
|--------------|------------|------------|------------|------------|------------|---|
| 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

The final output after pivoting will be shown as:

| Abc<br>Pivot | Abc<br>Pivot1     | #<br>Pivot1        |
|--------------|-------------------|--------------------|
| Product      | Pivot Field Names | 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                 |

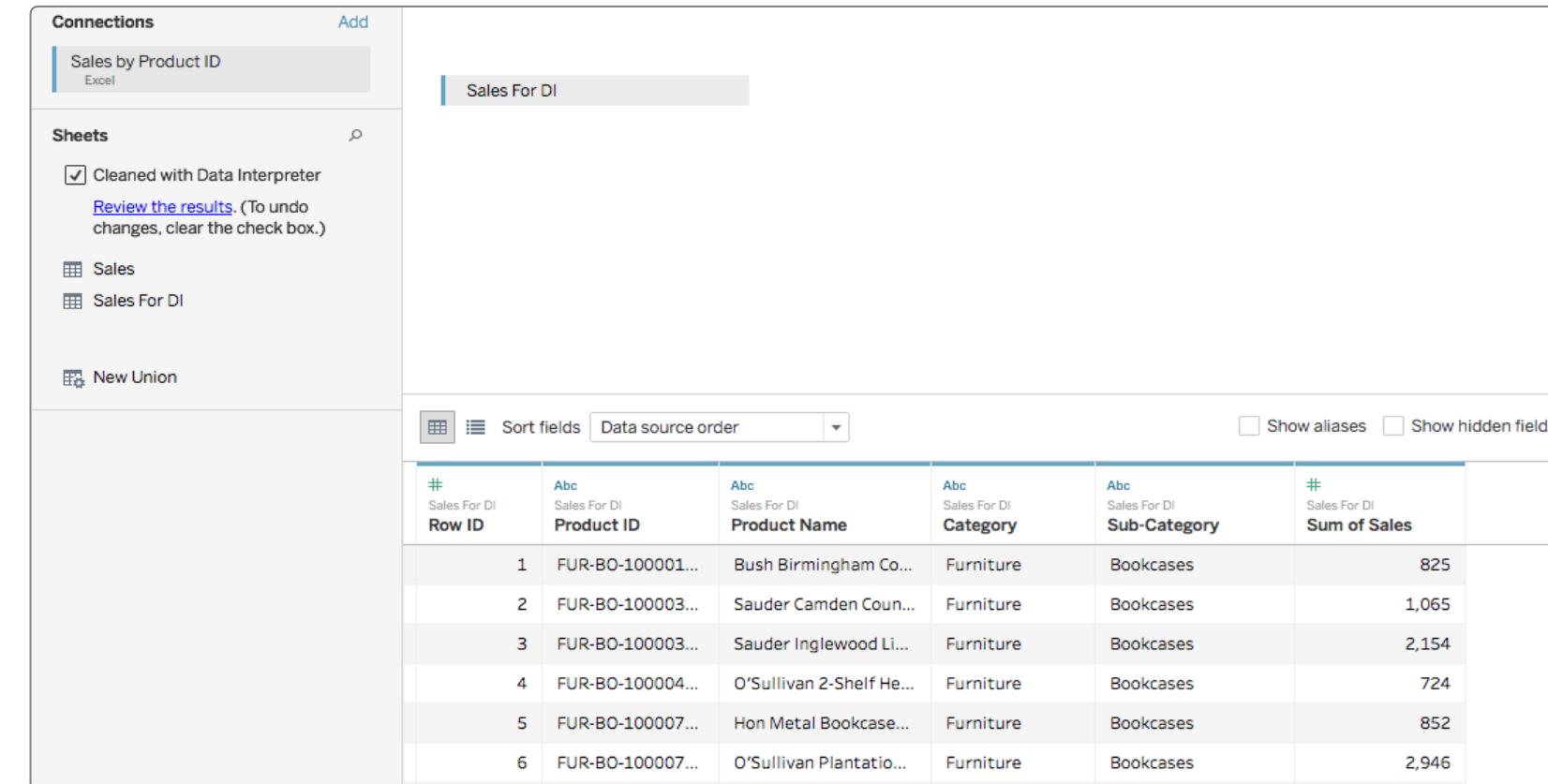


This is the most appropriate format for Tableau to evaluate all items.

## Data Interpreter

# Data Interpreter

A data interpreter automatically cleans and prepares data for analysis.



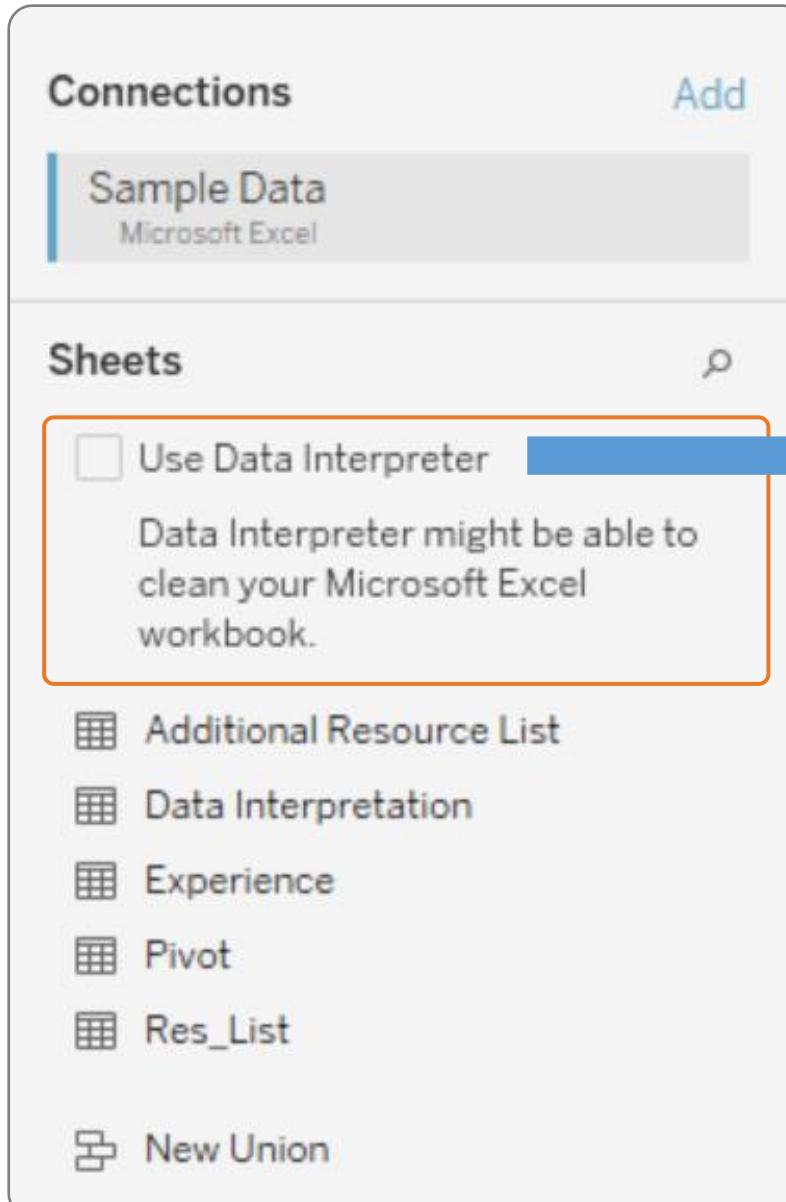
The screenshot shows a data interpretation interface. On the left, there's a sidebar with 'Connections' (Sales by Product ID, Excel) and 'Sheets' (Sales, Sales For DI). A note says 'Cleaned with Data Interpreter' with a link to 'Review the results'. Below are buttons for 'New Union' and a refresh icon. The main area is titled 'Sales For DI' and displays a table of sales data. The table has columns: Row ID, Product ID, Product Name, Category, Sub-Category, and Sum of Sales. The data shows six rows of furniture sales, all categorized under 'Bookcases'.

| #<br>Sales For DI<br>Row ID | Abc<br>Sales For DI<br>Product ID | Abc<br>Sales For DI<br>Product Name | Abc<br>Sales For DI<br>Category | Abc<br>Sales For DI<br>Sub-Category | #<br>Sales For DI<br>Sum of Sales |
|-----------------------------|-----------------------------------|-------------------------------------|---------------------------------|-------------------------------------|-----------------------------------|
| 1                           | FUR-BO-100001...                  | Bush Birmingham Co...               | Furniture                       | Bookcases                           | 825                               |
| 2                           | FUR-BO-100003...                  | Sauder Camden Coun...               | Furniture                       | Bookcases                           | 1,065                             |
| 3                           | FUR-BO-100003...                  | Sauder Inglewood Li...              | Furniture                       | Bookcases                           | 2,154                             |
| 4                           | FUR-BO-100004...                  | O'Sullivan 2-Shelf He...            | Furniture                       | Bookcases                           | 724                               |
| 5                           | FUR-BO-100007...                  | Hon Metal Bookcase...               | Furniture                       | Bookcases                           | 852                               |
| 6                           | FUR-BO-100007...                  | O'Sullivan Plantatio...             | Furniture                       | Bookcases                           | 2,946                             |

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.



The screenshot shows the 'Connections' pane in Tableau. At the top, there's a 'Connections' header and an 'Add' button. Below it, a 'Sample Data' connection is listed under 'Microsoft Excel'. In the 'Sheets' section, there's a checkbox labeled 'Use Data Interpreter'. A blue arrow points from this checkbox to a callout box. The callout box contains the text: 'Click on the data interpreter function to remove extraneous information'.

Connections Add

Sample Data Microsoft Excel

Sheets

Use Data Interpreter

Data Interpreter might be able to clean your Microsoft Excel workbook.

- Additional Resource List
- Data Interpretation
- Experience
- Pivot
- Res\_List
- New Union

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:

The screenshot shows the Microsoft Power BI interface with the 'Data Interpretation' feature enabled. On the left, under 'Connections', 'Sample Data (Microsoft Excel)' is selected. In the 'Sheets' pane, there is a checkbox labeled 'Use Data Interpreter' with a note: 'Data Interpreter might be able to clean your Microsoft Excel workbook.' Below this are several data sources: 'Additional Resource List', 'Data Interpretation', 'Experience', 'Pivot', 'Res\_List', and 'New Union'. The main area displays a table titled 'Data Interpretation' with 5 fields and 7 rows. The first row is a merged cell containing 'Abc Data Interpretation' and '# Data Interpretation'. The subsequent rows show data for columns F2 through F5. The table has a 'Table Details' sidebar on the left.

| Data Interpretation          |                          |                            |                            |                            |
|------------------------------|--------------------------|----------------------------|----------------------------|----------------------------|
|                              | F2                       | F3                         | F4                         | F5                         |
| Abc<br>Data Interpretation   | #<br>Data Interpretation | Abc<br>Data Interpretation | Abc<br>Data Interpretation | Abc<br>Data Interpretation |
| This is a sample merged cell |                          |                            |                            |                            |
| Name                         | null                     | Gender                     | Tool                       | Expertise                  |
| A                            | null                     | Male                       | Tableau                    | Novice                     |
| B                            | null                     | Male                       | Power BI                   | Proficient                 |
| C                            | null                     | Female                     | Qlik                       | Novice                     |
| D                            | null                     | Female                     | Tableau                    | Proficient                 |
| E                            | null                     | Male                       | Alteryx                    | Proficient                 |
| F                            | null                     | Female                     | Tableau                    | Expert                     |

# Data Interpreter

After enabling data interpreter:

The screenshot shows the Microsoft Excel ribbon interface. On the left, the 'Connections' section lists a connection to 'Sample Data' (Microsoft Excel). Below it, the 'Sheets' section shows several sheets: 'Additional Resource List', 'Data Interpretation', 'Experience', 'Pivot', 'Res\_List', and 'New Union'. A callout box highlights the 'Data Interpretation' sheet, which has a checked checkbox labeled 'Cleaned with Data Interpreter' and a link 'Review the results. (To undo changes, clear the check box.)'. To the right, a large data table is displayed under the heading 'Data Interpretation'. The table has 6 rows and 4 columns, with the first row serving as the header. The data is as follows:

| Name | Gender | Tool     | Expertise  |
|------|--------|----------|------------|
| A    | Male   | Tableau  | Novice     |
| B    | Male   | Power BI | Proficient |
| C    | Female | Qlik     | Novice     |
| D    | Female | Tableau  | Proficient |
| E    | Male   | Alteryx  | Proficient |
| F    | Female | Tableau  | Expert     |

# DATA AND ARTIFICIAL INTELLIGENCE

## Split

# Split

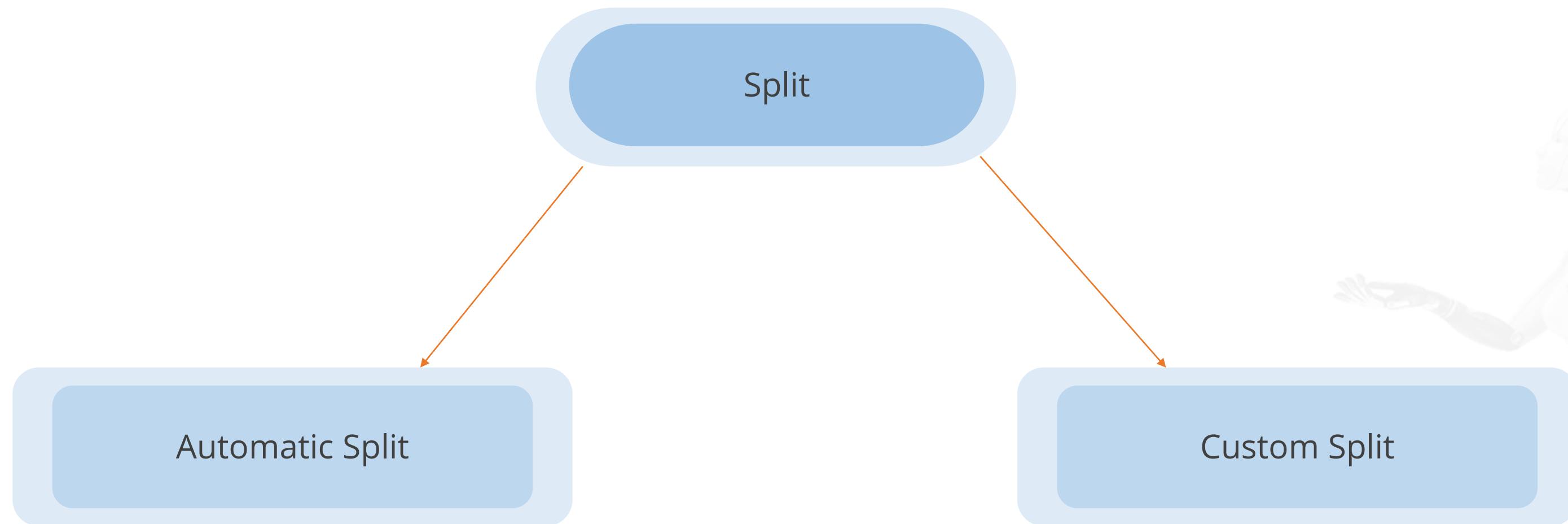
Split is a string function used to split a column into multiple columns.



Split is used to make data analysis easier.

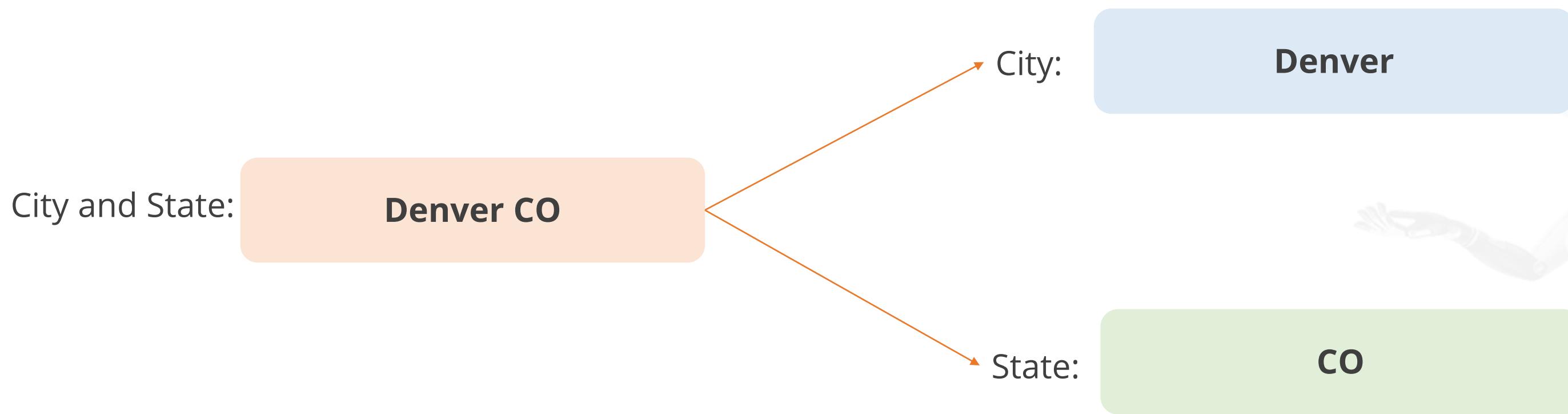
# Split

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.

The screenshot shows the Tableau Data Editor interface. A context menu is open over a data table. The menu options are: Rename, Copy Values, Hide, Aliases..., Create Calculated Field..., Create Group..., Split (which is highlighted with an orange border), Custom Split..., Pivot (select multiple fields), and Describe... . The data table has columns: Row ID, Sales, and Product ID. The Product ID column contains values like FUR-BO-10000112, FUR-BO-10000330, etc. The Sales column contains values like Furniture, Furniture, Furniture, Furniture, Furniture, Furniture, Furniture, Furniture, Furniture, Furniture.

| #  | Sales     | Product ID               |
|----|-----------|--------------------------|
| 1  | Furniture | FUR-BO-10000112          |
| 2  | Furniture | FUR-BO-10000330          |
| 3  | Furniture | FUR-BO-10000362          |
| 4  | Furniture | FUR-BO-10000468          |
| 5  | Furniture | FUR-BO-10000711          |
| 6  | Furniture | FUR-BO-10000780          |
| 7  | Furniture | FUR-BO-10001337          |
| 8  | Furniture | O'Sullivan 3-Shelf He... |
| 9  | Furniture | Bush Westfield Colle...  |
| 10 | Furniture | Sauder Mission Libra...  |

# Automatic Split

The following is the output produced by automatic split:

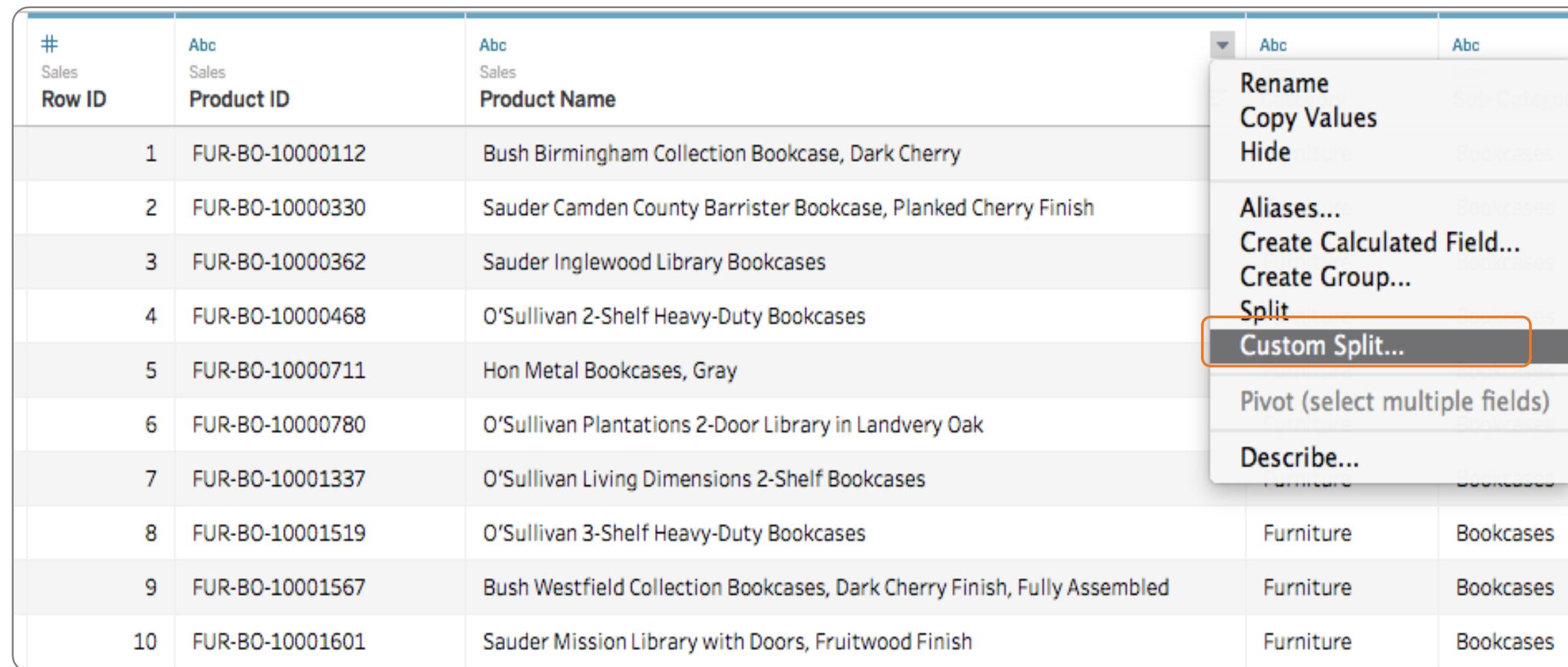


| #<br>Sales<br>Row ID | Abc<br>Sales<br>Product ID | =Abc<br>Calculation<br>Product ID - Split 1 | =#<br>Calculation<br>Product ID - Split 2 |
|----------------------|----------------------------|---|---|
| 1                    | FUR-BO-10000112            | BO  | 10000112                                  |
| 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.

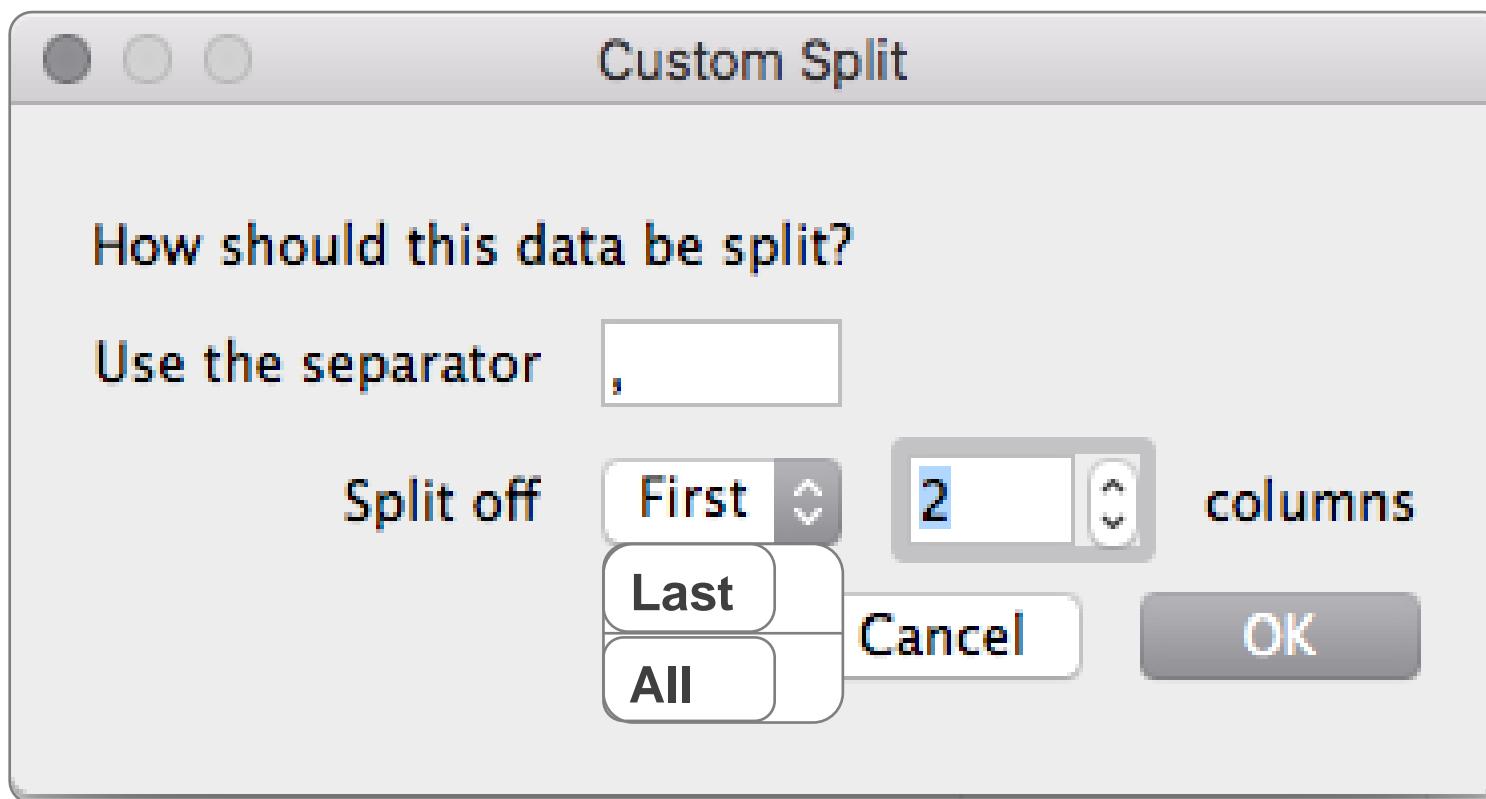
| #<br>Sales<br>Row ID | Abc<br>Sales<br>Product ID | Abc<br>Sales<br>Product Name   | Abc       | Abc       |
|----------------------|----------------------------|--|-----------|-----------|
| 1                    | FUR-BO-10000112            | Bush Birmingham Collection Bookcase, Dark Cherry                         |           |           |
| 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 |



A screenshot of a data grid interface. A context menu is open over the last row of the grid, specifically over the 'Product Name' column. The menu items are: Rename, Copy Values, Hide,Aliases..., Create Calculated Field..., Create Group..., Split, Custom Split..., Pivot (select multiple fields), and Describe... . The 'Custom Split...' option is highlighted with a red rectangle.

# 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:



| #      | Sales           | Sales  | Sales   | =Abc  | =Abc                   |
|--------|-----------------|--|---|---|------------------------|
| Row ID | Product ID      | Product Name   | Sales.Product Name                                    | Product Name - Split 1                                | Product Name - Split 2 |
| 1      | FUR-BO-10000112 | Bush Birmingham Collection Bookcase, Dark Cherry                         | Bush Birmingham Collection Bookcase                   | Bush Birmingham Collection Bookcase                   | Dark Cherry            |
| 2      | FUR-BO-10000330 | Sauder Camden County Barrister Bookcase, Planked Cherry Finish           | Sauder Camden County Barrister Bookcase               | Sauder Camden County Barrister Bookcase               | Planked Cherry Finish  |
| 3      | FUR-BO-10000362 | Sauder Inglewood Library Bookcases                                       | 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               | O'Sullivan 2-Shelf Heavy-Duty Bookcases               |                        |
| 5      | FUR-BO-10000711 | Hon Metal Bookcases, Gray  | Hon Metal Bookcases                                   | 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 | 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        | 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               | O'Sullivan 3-Shelf Heavy-Duty Bookcases               |                        |
| 9      | FUR-BO-10001567 | Bush Westfield Collection Bookcases, Dark Cherry Finish, Fully Assembled | Bush Westfield Collection Bookcases                   | Bush Westfield Collection Bookcases                   | Dark Cherry Finish     |
| 10     | FUR-BO-10001601 | Sauder Mission Library with Doors, Fruitwood Finish                      | Sauder Mission Library with Doors                     | 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.

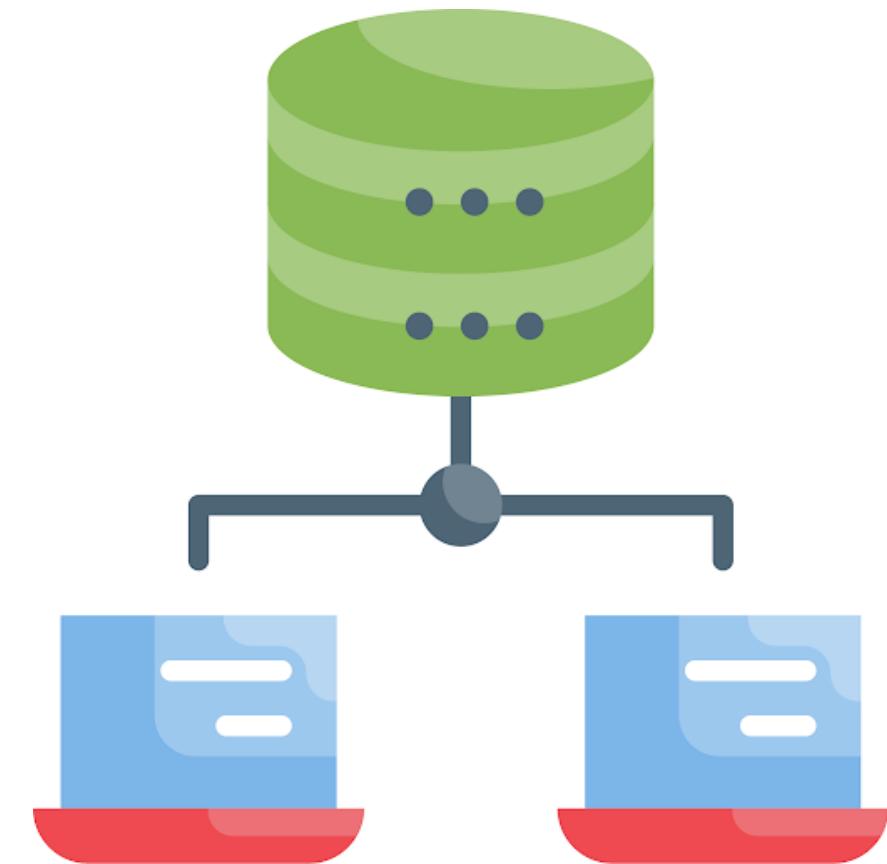
Custom aggregations can minimize the amount of data loaded into the extract.

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

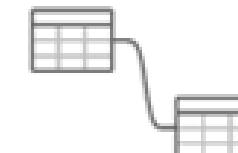
Step 1: Select **Extract** from the **Connection** radio buttons on the data source page

Orders (Sample - Superstore)

Connection  
 Live  Extract | [Edit](#) [Refresh](#)

Extract will include all data.

Orders



Need more data?  
Drag tables here to relate them. [Learn more](#)

# Creating Extracts with Multiple Tables

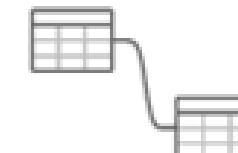
Step 2: Click on the **Edit** link to open the extract data dialog box

Orders (Sample - Superstore)

Connection  
 Live       Extract      [Edit](#) Refresh

Extract will include all data.

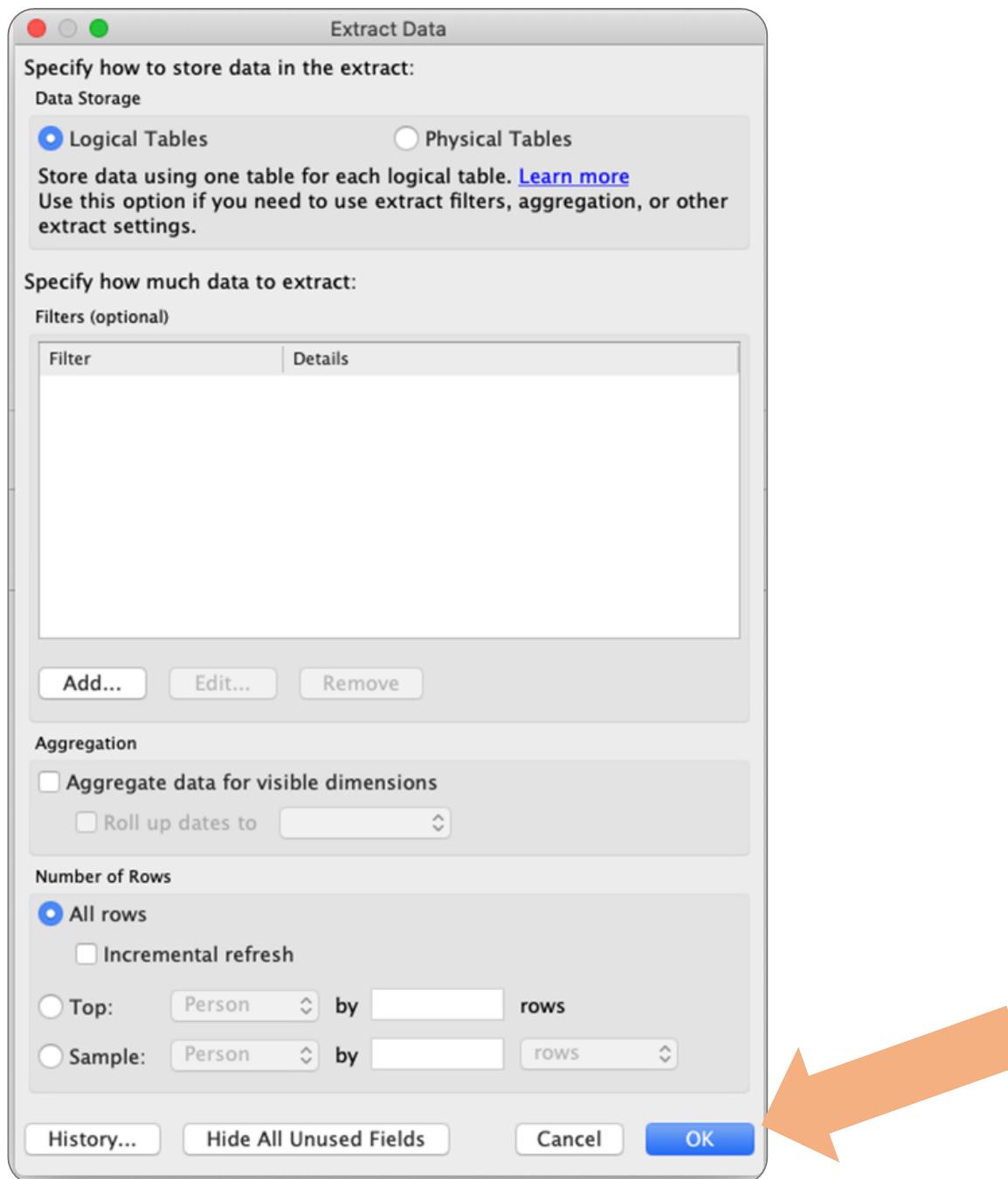
Orders



Need more data?  
Drag tables here to relate them. [Learn more](#)

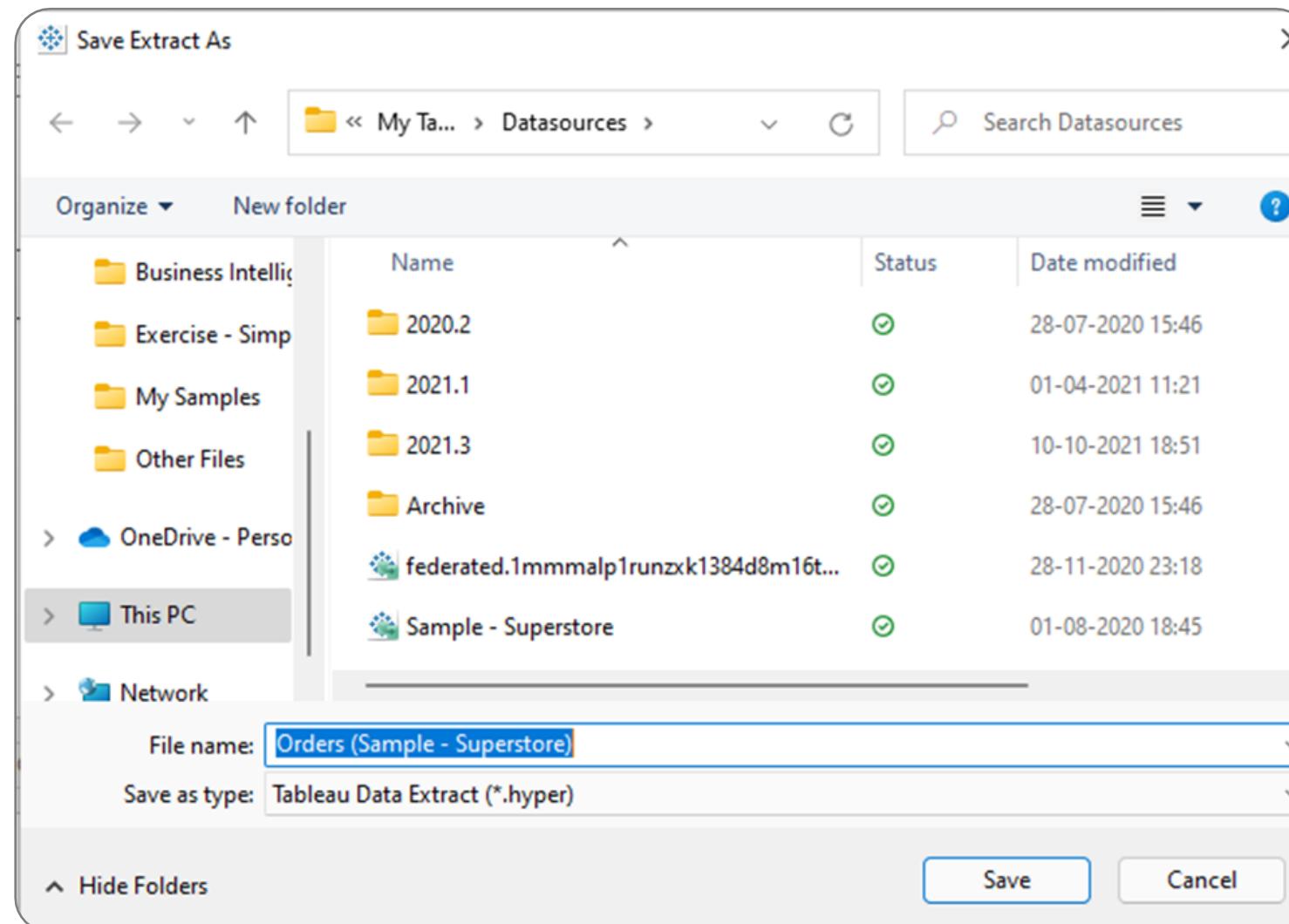
# Creating Extracts with Multiple Tables

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



# Creating Extracts with Multiple Tables

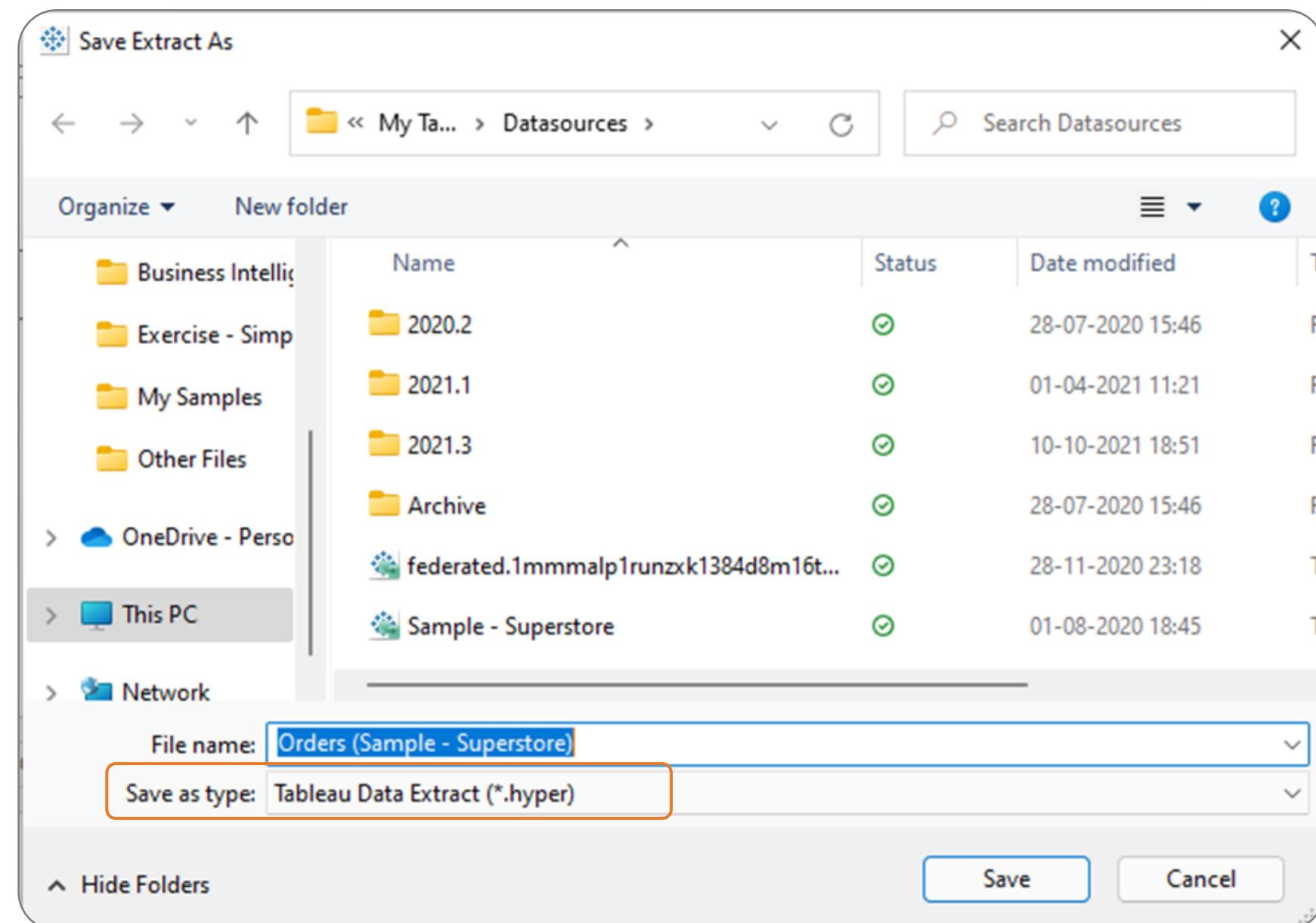
The extract will not start until the worksheet is moved.



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 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.





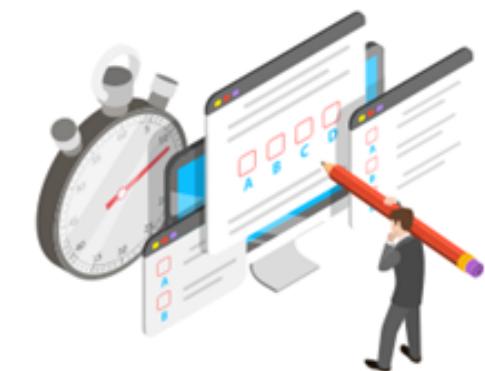
## Knowledge Check

## 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

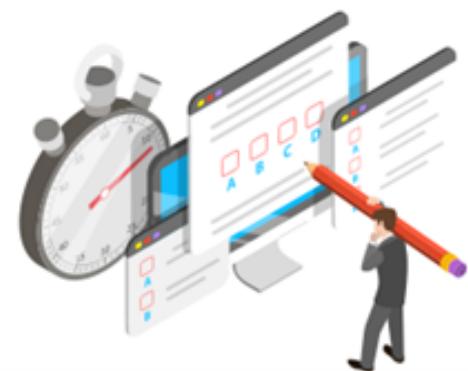


**Knowledge  
Check**

**1**

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

- A. Automatic split
- B. Split
- C. Custom split
- D. Delimiter



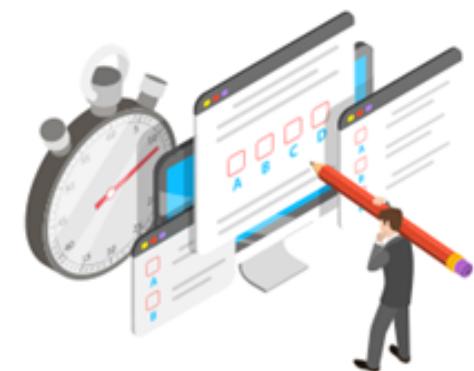
The correct answer is **C**

**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?**

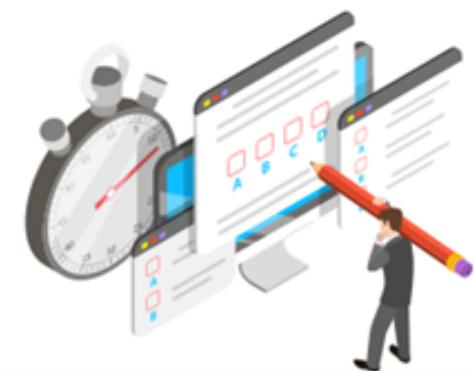
- A. Inner join
- B. Relationships
- C. Full outer join
- D. Cross database join



**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

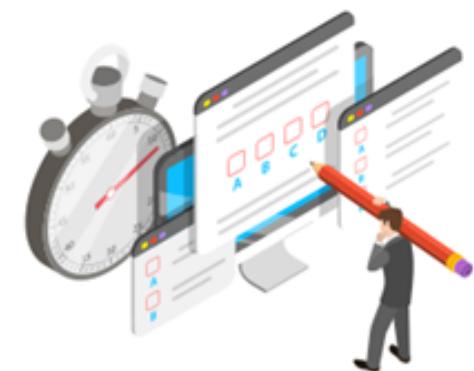


## Knowledge Check

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**

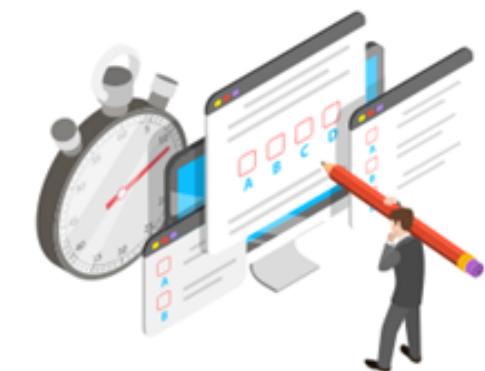
**A data interpreter is used for removing merged cells. It cleanses the data.**

**Knowledge  
Check**

**4**

**Which of the following formats store Tableau extracts?**

- A. .tde
- B. .xls
- C. .twb
- D. .hyper

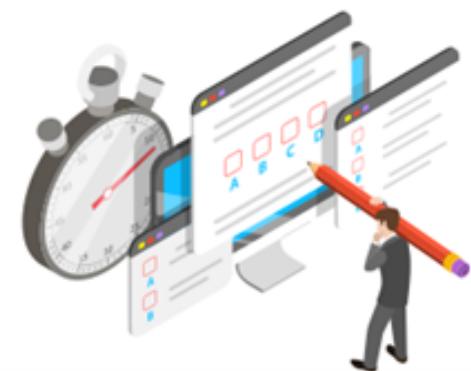


**Knowledge  
Check**

**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?**

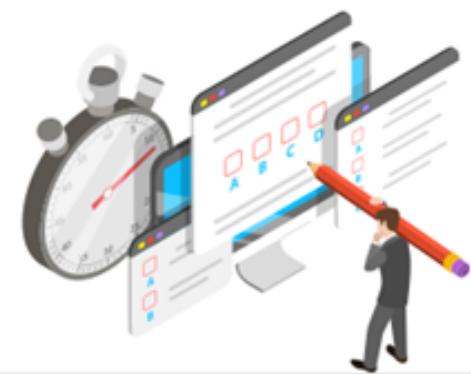
- A. Cross database join
- B. Join
- C. Data blending
- D. Relationships



**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.**