

Plug into your Data First !!

Data Connectivity & Published Data

Bijoy Kalpa Hazarika, Lead Solution Engineer
bijoy.hazarika@salesforce.com
He/Him



Herman Cokro, Lead Solution Engineer
hcokro@salesforce.com
He/Him

Hands-On Guide

Contents

Connecting to Data.....	3
Getting the Data Ready.....	8
Plan the Data Source.....	13
Filter Data from Data Sources.....	28
Extract Your Data.....	31
Save Customization as a Saved Data Source.....	34
Publish Your Data Source.....	36
Conclusion.....	41

Connecting to Data

Before you begin your analysis, you must connect to your data and then set up the data source. Tableau supports connecting to a wide variety of data, stored in a variety of places.

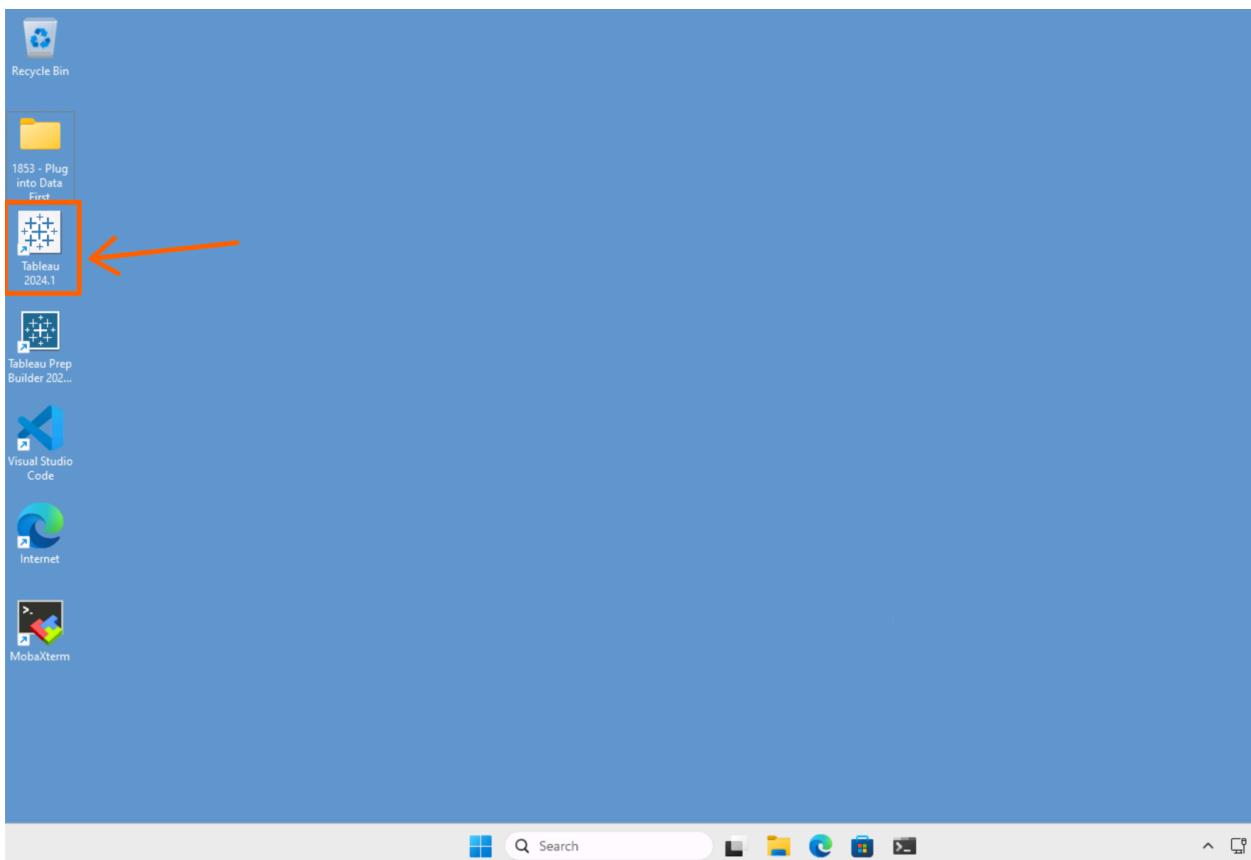
For example, your data might be stored on your computer in a spreadsheet or a text file, or in a big data, relational, or cube (multidimensional) database on a server in your enterprise. Or, you might connect to public domain data available on the web such as U.S. Census Bureau information, or to a cloud database source, such as Google Analytics, Amazon Redshift, or Salesforce.

1. Start your VM instance

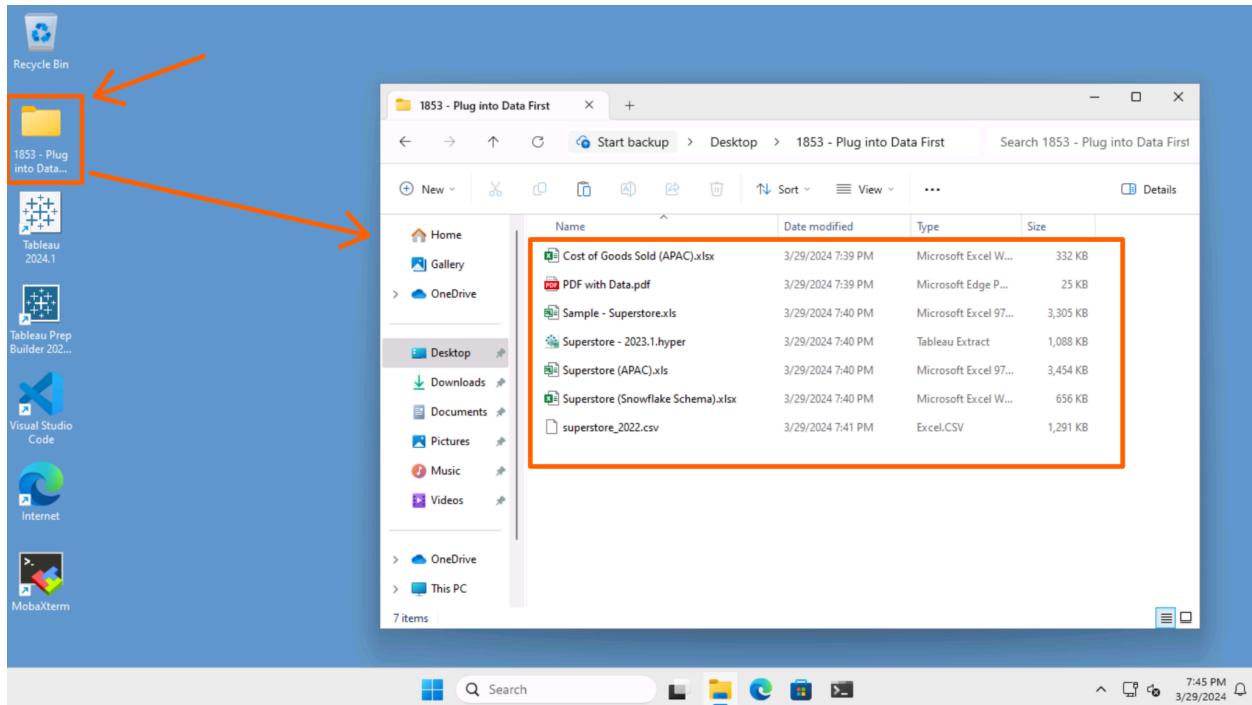
2. Start Tableau Desktop

You can find Tableau Desktop is already pre-installed in the VM.

- Go to Desktop
- Double click on Tableau 2024.1 icon in the desktop to launch Tableau Desktop.



3. You can also find the data sets that we are going to use for today's hands-on exercise on the folder '1853 - Plug into Data First' located in the desktop.



4. When you launch Tableau Desktop, the data connectors that are available to you are listed on the Connect pane, which is the left pane on the Start page.



For supported files and databases, Tableau provides native connectors that are built for and optimized for those types of data. If your file or database type is listed under Connect, use this native connector to connect to your data.

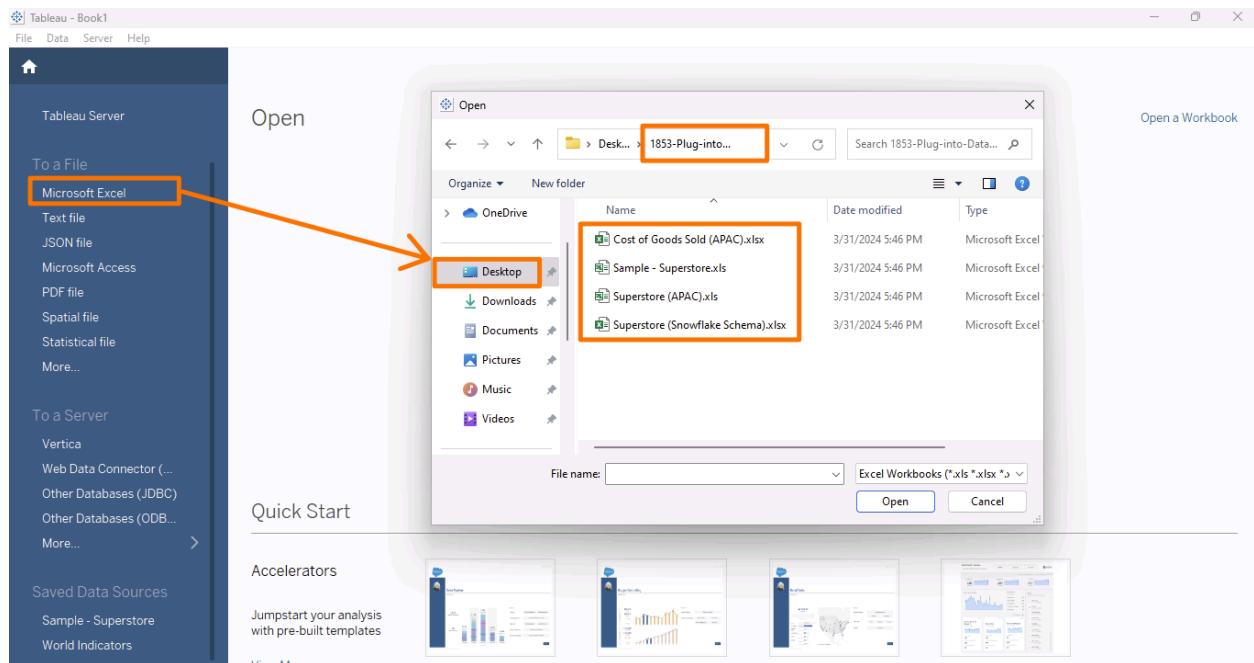


If your file or database type isn't listed, you might have the option of creating your own connection using Other Databases (JDBC), Other Databases (ODBC), a Web Data Connector, or a Connector Plugin built using the Tableau Connector SDK. Tableau provides limited support for connections that you create using these options.

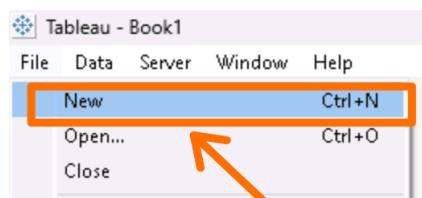
Sort by Name (a-z) ▾			
Installed Connectors (70)	Firebird 3	OData	Teradata
Action Vector	Google Analytics	OneDrive and SharePoint Online	Teradata OLAP Connector
Alibaba AnalyticDB for MySQL	Google BigQuery	Oracle	TIBCO Data Virtualization
Alibaba Data Lake Analytics	Google BigQuery (JDBC)	Oracle Eloqua (deprecated)	Vertica
Alibaba MaxCompute	Google Cloud SQL	Oracle Essbase	Web Data Connector (deprecated)
Amazon Athena	Google Drive	Pivotal Greenplum Database	
Amazon Aurora for MySQL	Hortonworks Hadoop Hive	PostgreSQL	Other Databases (JDBC)
Amazon EMR Hadoop Hive	IBM DB2	Presto	Other Databases (ODBC)
Amazon Redshift	IBM PDA (Netezza)	Progress OpenEdge	
Anaplan (deprecated)	Impala	Qubole Presto (deprecated)	
Apache Drill (deprecated)	Intuit QuickBooks Online (deprecated)	Salesforce	Additional Connectors (34) ⓘ
Azure Data Lake Storage Gen2	Kyvos	Salesforce Data Cloud	Action JDBC by Actian
Azure SQL Database	LinkedIn Sales Navigator (deprecated)	SAP HANA	Action ODBC by Actian
Azure Synapse Analytics	MariaDB	SAP NetWeaver Business Warehouse	Agiloft by Agiloft
Box	Marketing Cloud Intelligence	SAP Sybase ASE	Altinity Connector for ClickHouse by Altinity Inc
Cloudera Hadoop	Marketo (deprecated)	SAP Sybase IQ	Amazon DocumentDB by Amazon
Databricks	MarkLogic (deprecated)	ServiceNow ITSM (deprecated)	Amazon S3 by Tableau
Denodo (deprecated)	Microsoft SQL Server	SharePoint Lists (deprecated)	BI Connector by Guidanz Inc
Dremio	Microsoft SQL Server Analysis Services	SingleStore	Couchbase Analytics by Couchbase
Dropbox	MonetDB	Snowflake	Data Cloud by Salesforce
Esri	MongoDB BI Connector	Spark SQL	Data Virtuality JDBC by Data Virtuality
Exasol	MySQL	Splunk	

The data connectors supported by your copy of Tableau Desktop are determined by the version you purchased. For more information, see the list of [data connectors](#) (Link opens in a new window) on the Tableau website.

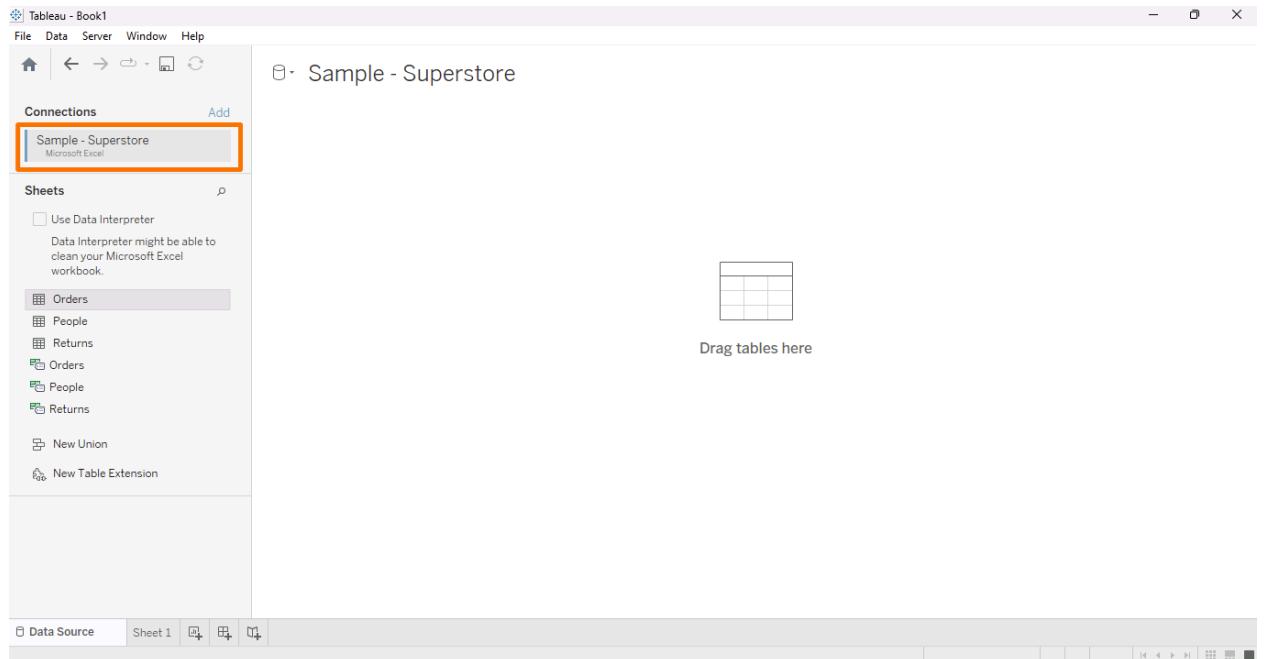
5. This step describes how to connect Tableau to Microsoft Excel file data and set up the data source. Tableau connects to .xls and .xlsx files.
 - a. After you [open Tableau](#), under Connect, [click Microsoft Excel](#).
 - b. Select the Excel workbook you want to connect to, and then click Open.
 - c. Note: You can also find the data sets that we are going to use for today's hands-on exercise on the folder '1853 - Plug into Data First' located in the desktop.



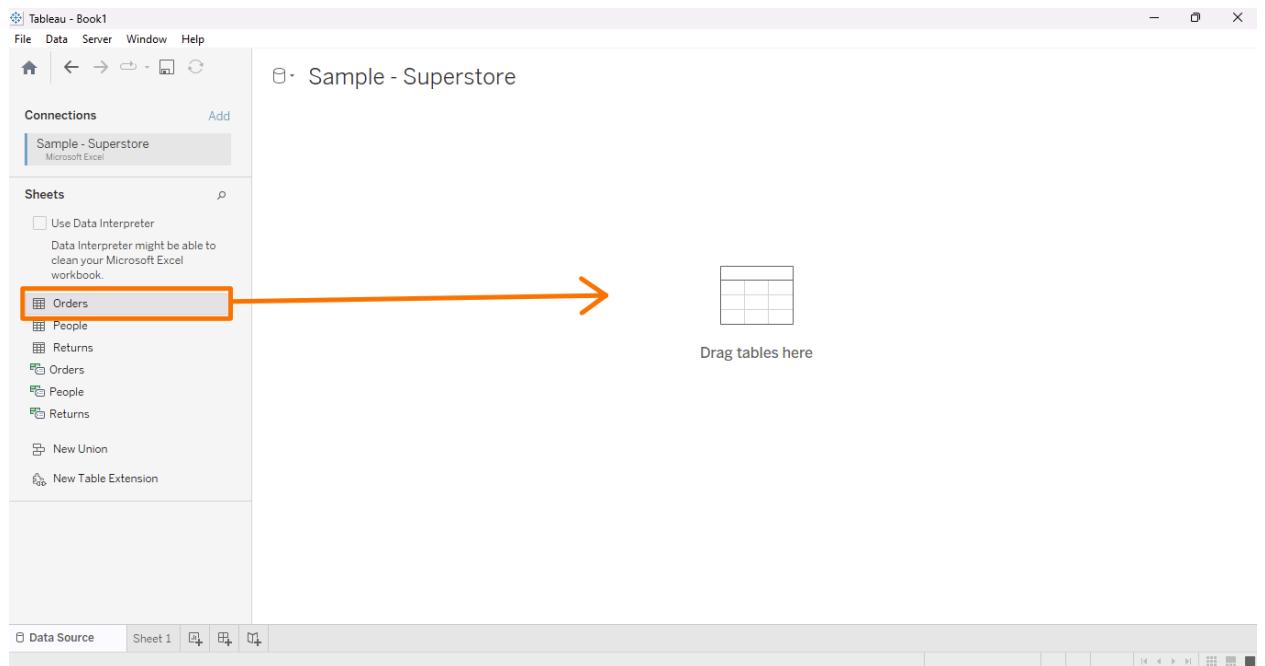
6. After connecting to the first data set (Excel), try to connect to another different data sources:
 - o CSV files **To connect to a .csv file, use the **Text file** connector.
 - o PDF
 - o Spatial files (.zip)



7. On the data source page, check if the connection is showing the correct data source and data source type.



8. Select the table that you want to use, drag and drop the table to the empty data source canvas (you can also double-click the table)



Getting the Data Ready

Data can be organized in various ways. To take advantage of Tableau Desktop, Tableau recommends that you connect to data that is formatted for analysis. Specifically, data that is:

- as granular as possible rather than aggregated (such as daily weather data rather than monthly averages)
- organized like a database table (rather than a column-oriented table such as a crosstab)
- stripped of extraneous information (anything that's not the data and its headers)

When data is structured for analysis, it's much easier to ask and answer questions. Tableau can aggregate raw data to the desired level, rather than being restricted by the aggregations already present in the data. Groups and hierarchies can be created as needed, and calculations can be performed in the flow of analysis.

Tableau Desktop has basic cleaning options and the Data Interpreter. Tableau Prep may be necessary for more complex formatting issues.

If Tableau detects that it can help optimize your data source for analysis, it prompts you to use Data Interpreter.

Tableau Data Interpreter can detect sub-tables that you can use and remove unique formatting that might cause problems later on in your analysis. For more information, see [Clean Data from Excel, CSV, PDF, and Google Sheets with Data Interpreter](#).

The following Tableau Desktop-specific sections highlight and provide suggestions for resolving some common formatting or issues that can make analyzing your data difficult.

1. If you want, you can also rename the data source and/or the table name. You can select the default data source name at the top of the page, and then enter a unique data source name for use in Tableau. For example, use a naming convention that helps other users of the data source figure out which data source to connect to.

Tableau - Book1

File Data Server Window Help

Connections Add

Sample - Superstore Microsoft Excel

Sheets

Use Data Interpreter

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

Orders People Returns Orders People Returns

New Union New Table Extension

Go to Worksheet

Orders (Sample - Superstore)

Orders 21 fields 9994 rows

Name Orders

#	Abc Orders	Orders	Orders	Abc Orders	Abc Orders
Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID
1	CA-2019-152156	11/8/2019	11/11/2019	Second Class	CG-12520
2	CA-2019-152156	11/8/2019	11/11/2019	Second Class	CG-12520
3	CA-2019-138688	6/12/2019	6/16/2019	Second Class	DV-13045
4	US-2018-108966	10/11/2018	10/18/2018	Standard Class	SO-20335
5	US-2018-108966	10/11/2018	10/18/2018	Standard Class	SO-20335
6	CA-2017-115812	6/9/2017	6/14/2017	Standard Class	BH-11710
7	CA-2017-115812	6/9/2017	6/14/2017	Standard Class	BH-11710
8	CA-2017-115812	6/9/2017	6/14/2017	Standard Class	BH-11710
9	CA-2017-115812	6/9/2017	6/14/2017	Standard Class	BH-11710

100 rows

Fields

Type	Field Name	Physical Table	Remote File...
#	Row ID	Orders	Row ID
Abc	Order ID	Orders	Order ID
	Order Date	Orders	Order Date
	Ship Date	Orders	Ship Date
Abc	Ship Mode	Orders	Ship Mode

Data Source Sheet 1

2. You can also view and inspect the fields information such as the data type (numeric, string, geographical location, date, etc.), the field name, and the physical location.

Tableau - Book1

File Data Server Window Help

Connections Add

Sample - Superstore Microsoft Excel

Sheets

Use Data Interpreter

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

Orders People Returns Orders People Returns

New Union New Table Extension

Go to Worksheet

Orders (Sample - Superstore)

Orders 21 fields 9994 rows

Name Orders

#	Abc Orders	Orders	Orders	Abc Orders	Abc Orders
Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID
1	CA-2019-152156	11/8/2019	11/11/2019	Second Class	CG-12520
2	CA-2019-152156	11/8/2019	11/11/2019	Second Class	CG-12520
3	CA-2019-138688	6/12/2019	6/16/2019	Second Class	DV-13045
4	US-2018-108966	10/11/2018	10/18/2018	Standard Class	SO-20335
5	US-2018-108966	10/11/2018	10/18/2018	Standard Class	SO-20335
6	CA-2017-115812	6/9/2017	6/14/2017	Standard Class	BH-11710
7	CA-2017-115812	6/9/2017	6/14/2017	Standard Class	BH-11710
8	CA-2017-115812	6/9/2017	6/14/2017	Standard Class	BH-11710
9	CA-2017-115812	6/9/2017	6/14/2017	Standard Class	BH-11710

100 rows

Fields

Type	Field Name	Physical Table	Remote File...
#	Row ID	Orders	Row ID
Abc	Order ID	Orders	Order ID
	Order Date	Orders	Order Date
	Ship Date	Orders	Ship Date
Abc	Ship Mode	Orders	Ship Mode

Data Source Sheet 1

3. This shows the preview of the data from the table that you are connecting to.

The screenshot shows the Tableau desktop interface with the 'Orders' sheet selected. The preview pane displays the first 100 rows of the 'Orders' table, which contains 21 fields and 9994 rows. The columns are labeled: Row ID, Order ID, Order Date, Ship Date, Ship Mode, and Customer ID. The entire preview area is highlighted with an orange border.

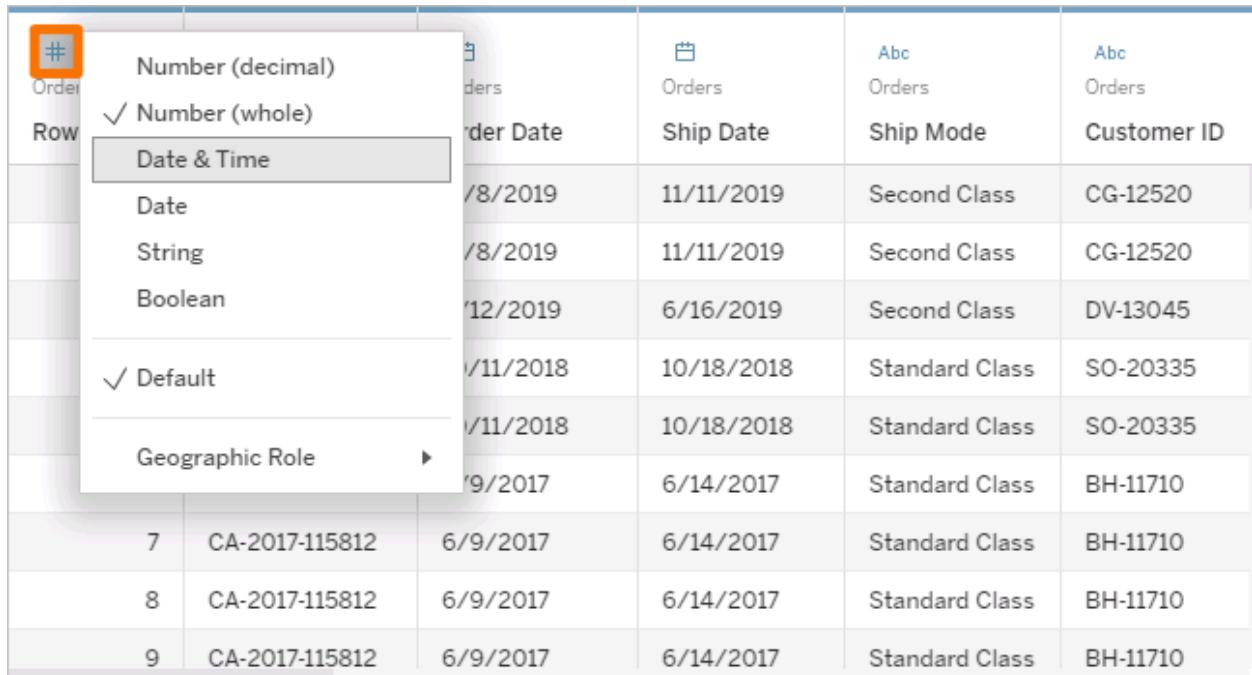
#	Abc Orders	Orders	Orders	Orders	Abc Orders
Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID
1	CA-2019-152156	11/8/2019	11/11/2019	Second Class	CG-12520
2	CA-2019-152156	11/8/2019	11/11/2019	Second Class	CG-12520
3	CA-2019-138688	6/12/2019	6/16/2019	Second Class	DV-13045
4	US-2018-108966	10/11/2018	10/18/2018	Standard Class	SO-20335
5	US-2018-108966	10/11/2018	10/18/2018	Standard Class	SO-20335
6	CA-2017-115812	6/9/2017	6/14/2017	Standard Class	BH-11710
7	CA-2017-115812	6/9/2017	6/14/2017	Standard Class	BH-11710
8	CA-2017-115812	6/9/2017	6/14/2017	Standard Class	BH-11710
9	CA-2017-115812	6/9/2017	6/14/2017	Standard Class	BH-11710

4. Data can be organized in various ways. To take advantage of Tableau Desktop, Tableau recommends that you connect to data that is formatted for analysis. Before starting building visualizations and analysis, it is a good practice to inspect if the data type already corresponds to the data within the field.

This screenshot shows the same Tableau desktop interface as above, but with the data types for each column highlighted by an orange border. The 'Type' column in the Fields section and the corresponding columns in the preview table are all highlighted.

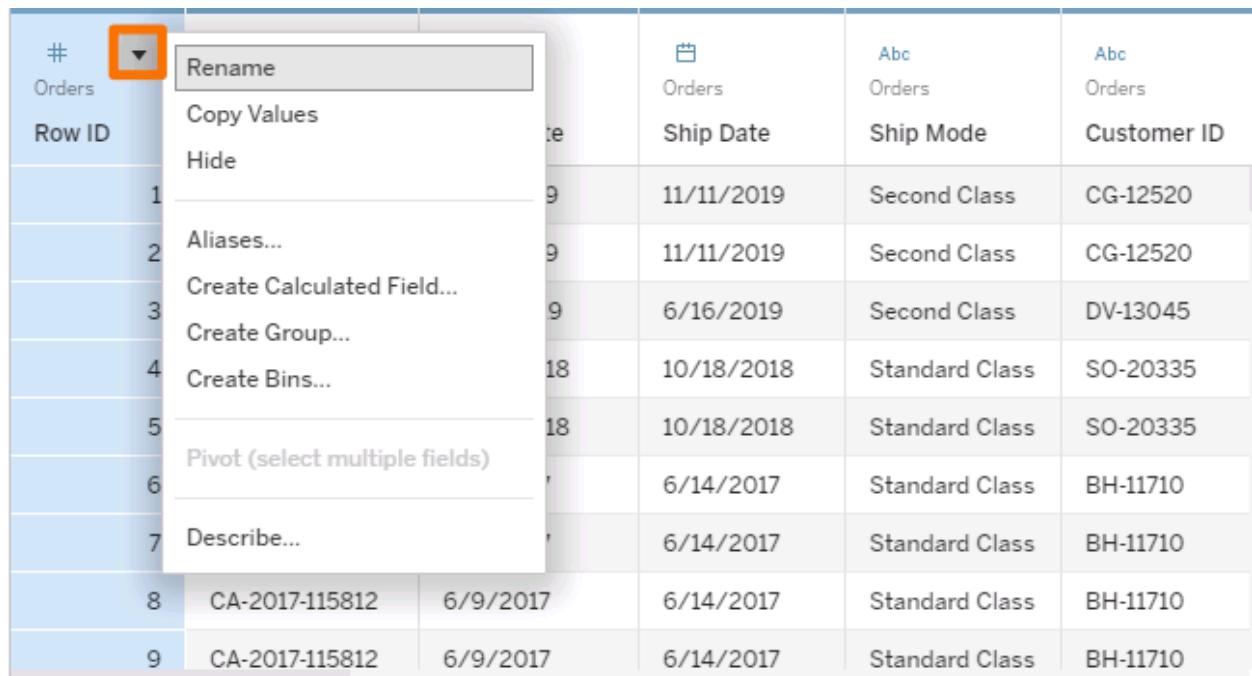
#	Abc Orders	Orders	Orders	Orders	Abc Orders
Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID
1	CA-2019-152156	11/8/2019	11/11/2019	Second Class	CG-12520
2	CA-2019-152156	11/8/2019	11/11/2019	Second Class	CG-12520
3	CA-2019-138688	6/12/2019	6/16/2019	Second Class	DV-13045
4	US-2018-108966	10/11/2018	10/18/2018	Standard Class	SO-20335
5	US-2018-108966	10/11/2018	10/18/2018	Standard Class	SO-20335
6	CA-2017-115812	6/9/2017	6/14/2017	Standard Class	BH-11710
7	CA-2017-115812	6/9/2017	6/14/2017	Standard Class	BH-11710
8	CA-2017-115812	6/9/2017	6/14/2017	Standard Class	BH-11710
9	CA-2017-115812	6/9/2017	6/14/2017	Standard Class	BH-11710

5. If the data type is not accurately assigned, or if you want to make changes to the data type, you can do so by clicking on the data type symbol on the top of the field header.
- Click on the data type symbol, on top of the field header.
 - Select the appropriate data type.



#	Orders	Orders	Abc	Abc
Row ID	Order Date	Ship Date	Orders	Orders
	/8/2019	11/11/2019	Second Class	CG-12520
	/8/2019	11/11/2019	Second Class	CG-12520
	/12/2019	6/16/2019	Second Class	DV-13045
	/11/2018	10/18/2018	Standard Class	SO-20335
	/11/2018	10/18/2018	Standard Class	SO-20335
	/9/2017	6/14/2017	Standard Class	BH-11710
7	CA-2017-115812	6/9/2017	Standard Class	BH-11710
8	CA-2017-115812	6/9/2017	Standard Class	BH-11710
9	CA-2017-115812	6/9/2017	Standard Class	BH-11710

6. You can rename the field name by clicking on the dropdown menu beside the data type symbol.



#	Orders	Orders	Abc	Abc
Row ID	Order Date	Ship Date	Orders	Orders
1	9	11/11/2019	Second Class	CG-12520
2	9	11/11/2019	Second Class	CG-12520
3	9	6/16/2019	Second Class	DV-13045
4	18	10/18/2018	Standard Class	SO-20335
5	18	10/18/2018	Standard Class	SO-20335
6	18	6/14/2017	Standard Class	BH-11710
7	18	6/14/2017	Standard Class	BH-11710
8	CA-2017-115812	6/9/2017	Standard Class	BH-11710
9	CA-2017-115812	6/9/2017	Standard Class	BH-11710

7. If you have string fields in your data that contain multiple distinct pieces of information (for example, the first and last name of a customer) you may be able to split the values into separate fields.

You can use split options in Tableau Desktop to separate the values based on a separator (delimiter) or a repeated pattern of values present in each row.

The screenshot shows the Tableau desktop interface with a data source named 'Orders+ (Sample - Superstore)' connected to 'Sample - Superstore' via Microsoft Excel. The context menu for a selected field is open, and the 'Split' option is highlighted with a red box. The menu also includes other options like Rename, Copy Values, Hide, Aliases..., Create Calculated Field..., Create Group..., and Custom Split... . Below the menu, a preview of the data shows columns for Row ID, Order ID, Name, and Fields. The 'Fields' section displays four columns: Type, Field Name, Physical Table, and Remote The data preview shows 21 rows of order information.

Type	Field Name	Physical Table	Remote ...
#	Row ID	Orders	Row ID
Abc	Order ID	Orders	Order ID
	Order Date	Orders	Order Date
	Ship Date	Orders	Ship Date

Tableau Desktop—but not web editing in the browser—has a menu option for automatic or custom splits. These are based on the SPLIT string function, which can also be used manually in a calculation for complete control over the split.

Not all data sources support SPLIT. You can tell if your data supports the SPLIT function by checking for the Split and Custom Split menu options.

8. You can use split options in Tableau Desktop to separate the values based on a separator (delimiter) or a repeated pattern of values present in each row. In the example of Order ID, the common separator is a dash (-) between the country code, year, and the unique ID.

	# Orders	# Discount	# Profit	# Order ID - Split 1	# Order ID - Split 2	# Order ID - Split 3
2	0.000000	41.91		CA	2019	152156
3	0.000000	219.58		CA	2019	152156
2	0.000000	6.87		CA	2019	138688
5	0.450000	-383.03		US	2018	108966
2	0.200000	2.52		US	2018	108966
7	0.000000	14.17		CA	2017	115812
1	0.000000	1.97		CA	2017	115812
5	0.200000	90.72		CA	2017	115812

Plan the Data Source

At the center of Tableau is your data. How successful you are with exploring data, answering questions, and building visualizations for yourself or others to consume all depend on the underlying data.

If your goal is to do some quick exploration or ad-hoc analysis, you might be able to hop in, connect to some data, drag and drop a bit to build some vizs, and hop out with the information you need. But if your goal is to create an analysis or data source that will be used more than once, it's best to think through and plan your data source.

If your data comes from one table, you can [connect to your data](#) to create the data source, drag the table onto the canvas, and then start building your view. But if your data is spread across multiple tables—or across multiple databases—you'll need to combine it. Combining data happens on the Data Source Page.

Relationships are the default way to combine data in Tableau. Relationships are a dynamic, flexible way to combine data from multiple tables for analysis. If necessary, tables can also be joined or unioned. Data sources can also be blended. Let's take a look at the options for how data can be combined in Tableau and some of the situations when each method is useful.

Relationships	<p>Establish the potential for joins between two tables based on related fields. Does not merge data together to create a new, fixed table. During analysis, queries the relevant tables automatically using the contextually-appropriate joins to generate a custom table of data for that analysis.</p> <p>Maintains the appropriate level of detail, does not lose data, keeps appropriate aggregations, and handles nulls.</p>
Joins	<p>Merge two tables of data based on a join clause and join type to form a new, fixed table of data. Often used to add new columns of data across the same basic row structure.</p> <p>May cause data loss with some join types if fields are not present in all tables. May cause data duplication if tables are at different levels of detail.</p>
Unions	<p>Merge two or more tables of data to form a new, fixed table of data. Used to append (add new rows of) data across the same basic column structure.</p>
Blends	<p>Work across two or more separate Tableau data sources. Data remains separate. Tableau queries the data sources independently and visualizes the results together in the view, based on the linking fields established for that sheet. Mimics the behavior of a left join and may filter data from secondary data sources.</p>

Relationship (logical tables)

Relating is a method for working with data from multiple tables based on shared fields—columns—between those tables. Establishing a relationship informs Tableau how rows can be connected across tables. Tableau holds that information but does not immediately bring the rows together. Instead, when a visualization is created, the fields involved in the analysis are traced back through the relationship and the appropriate data is returned from its original table.

Because the data is never fully brought together into a single table outside the context of a visualization, relationships are useful when the data is at different levels of detail, or granularity. For example, if you need to work with data about daily rainfall in one table but hourly temperatures in another.

Tableau supports relationships between tables in the same database and between tables in different databases.

1. To combine tables using relationships:

- Drag and drop the first table (Orders table) into the empty data source canvas.
- Drag and drop the second table (People table) into the data source canvas. When you see the "noodle" between the two tables, drop that table.

The screenshot shows the Tableau Data Source canvas. On the left, the 'Connections' pane shows a single connection to 'Sample - Superstore'. The 'Sheets' pane lists several sheets: 'Orders', 'People', 'Returns', 'Orders', 'People', and 'Returns'. The 'People' sheet is highlighted with a red box and has an orange arrow pointing from it towards the 'Orders' table. The 'Orders' table is also highlighted with an orange box. In the center, the 'Orders (Sample - Superstore)' pane displays the 'Orders' table with 21 fields and 9994 rows. A preview of the data is shown in a grid. On the right, there are tabs for 'Live' and 'Extract' connections, and a 'Filters' section. Below the preview, a note says 'Need more data? Drag tables here to relate them.' with a link to 'Learn more'.

2. The Edit Relationship dialog box opens. Tableau automatically attempts to create the relationship based on existing key constraints and matching fields to define the relationship. If it can't determine the matching fields, you'll need to select them.

The screenshot shows the 'Edit Relationship' dialog box. It displays two tables: 'Orders' and 'People'. A relationship is being established between 'Orders' and 'People'. The 'Orders' table has a dropdown menu showing 'Orders — People'. A red box highlights this dropdown, and a red arrow points down to the 'Relationships' section. This section contains three fields: 'Orders' (with dropdown 'Abc Region'), 'Operator' (with dropdown '='), and 'People' (with dropdown 'Abc Region (People)'). Below this, there's a note about the difference between relationships and joins, a 'Performance Options' section, and a note about query optimization. To the right, a preview of the joined data is shown in a grid. The preview table has columns 'Person' and 'Region (People)'. Data rows include Anna Andreadi (West), Chuck Magee (East), Kelly Williams (Central), and Cassandra Brandow (South).

3. To change the fields:

- Select a field pair, and then click in the list of fields below to select a new pair of matching fields.
- (optional) To add multiple field pairs: After you select the first pair, click Close, and then click Add more fields.

The screenshot shows the Tableau relationship editor interface. A dropdown menu is open, listing several fields: Customer Name, Order ID, Product ID, Product Name, and Region. The 'Region' option is selected, indicated by a checkmark. Below the dropdown is a search bar with the placeholder 'Edit Calculation...'. At the bottom of the editor, there are two dropdown menus for 'Orders' and 'People', both set to 'Region'. A button labeled '(+) Add more fields' is located at the bottom left. A help link 'How do relationships differ from joins? Learn more' is visible at the bottom right.

The screenshot shows the Tableau relationship editor interface. A dropdown menu is open, listing fields: Region, Person, and Region (People). The 'Region (People)' option is selected, indicated by a checkmark. Below the dropdown is a search bar with the placeholder 'Edit Calculation...'. At the bottom of the editor, there are two dropdown menus for 'Orders' and 'People', both set to 'Region (People)'. A button labeled '(+) Add more fields' is located at the bottom left. A help link 'How do relationships differ from joins? Learn more' is visible at the bottom right. A section titled 'Performance Options' is expanded, containing a note about optimization settings.

4. If no constraints are detected, a Many-to-many relationship is created and referential integrity is set to Some records match.

<

How do relationships differ from joins? [Learn more](#)

The screenshot shows the Tableau Data Source setup for a many-to-many relationship. On the left, under 'Orders', there is a dropdown menu for 'Region' with the value 'Abc Region'. In the center, the 'Operator' section shows a dropdown menu with '=' selected, and a tooltip provides information about performance options. On the right, under 'People', there is a dropdown menu for 'Region (People)' with the value 'Abc Region (People)'. Below these dropdowns, there is a note about performance options: 'These settings help Tableau optimize queries during analysis. The default settings are recommended, if you aren't sure what to choose. [Learn more](#)'.

5. These default settings are a safe choice and provide the most flexibility for your data source. The default settings support full outer joins and optimize queries by aggregating table data before forming joins during analysis. All column and row data from each table becomes available for analysis.

▼ **Performance Options**

These settings help Tableau optimize queries during analysis. The default settings are recommended, if you aren't sure what to choose. [Learn more](#)

The screenshot shows the 'Cardinality' and 'Referential Integrity' settings for the relationship. Under 'Cardinality', both 'Orders' and 'People' are set to 'Many'. Under 'Referential Integrity', both 'Orders' and 'People' are set to 'Some records match'.

- Add more tables following the same steps, as needed.

The screenshot shows the Tableau Data Source interface. On the left, the 'Sheets' pane lists 'Orders', 'People', and 'Returns'. The 'Returns' sheet is selected and highlighted with an orange box. An orange arrow points from this selection towards the relationships diagram. The main area displays a relationship diagram where 'Orders' is connected to 'Returns'. Below the diagram, a table shows the data for the 'Orders — Returns' relationship. The table has two columns: 'Returned' and 'Order ID (Returns)'. The data consists of eight rows, all of which have 'Yes' in the 'Returned' column and a unique order ID in the 'Order ID (Returns)' column.

Returned	Order ID (Returns)
Yes	CA-2017-100762
Yes	CA-2017-100867
Yes	CA-2017-102652
Yes	CA-2017-102652
Yes	CA-2017-102652

Joins (physical tables)

While similar, joins and relationships behave differently in Tableau, and are defined in different layers of the data model. You create relationships between logical tables at the top-level, logical layer of your data source. You create joins between physical tables in the physical layer of your data source.

Joins merge data from two tables into a single table before your analysis begins. Merging the tables together can cause data to be duplicated or filtered from one or both tables; it can also cause NULL rows to be added to your data if you use a left, right, or full outer join. When analyzing joined data, you need to make sure that you correctly handle the effects of the join on your data.

Note: When duplication or the filtering effects of a join might be desirable, use joins to merge tables together instead of relationships.

A relationship describes how two independent tables relate to each other but doesn't merge the tables together. This avoids the data duplication and filtering issues that might occur in a join and can make working with your data easier.

RELATIONSHIPS	JOINS
Defined between logical tables in the Relationship canvas (logical layer)	Defined between physical tables in the Join/Union canvas (physical layer)
Don't require you to define a join type	Require join planning and join type
Act like containers for tables that are joined or unioned	Are merged into their logical table
Only data relevant to the viz is queried. Cardinality and referential integrity settings can be adjusted to optimize queries.	Run as part of every query
Level of detail is at the aggregate for the viz	Level of detail is at the row level for the single table
Join types are automatically formed by Tableau based on the context of analysis. Tableau determines the necessary joins based on the measures and dimensions in the viz.	Join types are static and fixed in the data source, regardless of analytical context. Joins and unions are established prior to analysis and don't change.
Rows aren't duplicated	Merged table data can result in duplication
Unmatched records are included in aggregates, unless explicitly excluded	Unmatched records are omitted from the merged data
Create independent domains at multiple levels of detail	Support scenarios that require a single table of data, such as extract filters and aggregation

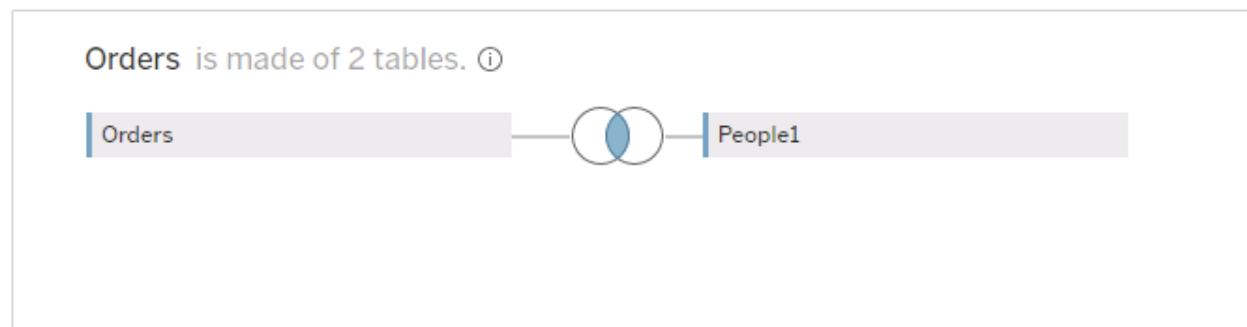
- To create a join, connect to the relevant data source or sources.

These can be in the same data source (such as tables in a database or sheets in an Excel spreadsheet) or different data sources (this is known as a cross-database join). If you combine tables using a cross-database join, Tableau colors the tables in the canvas and the columns in the data grid to show you which connection the data comes from.

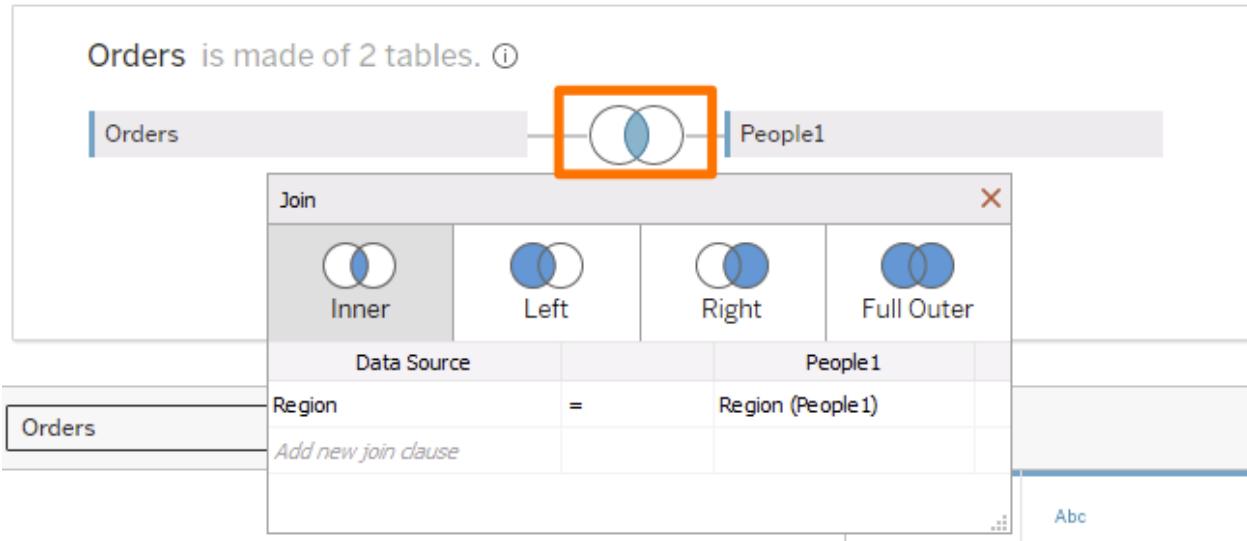
- Drag the first table (table Orders) into the canvas.
- Select Open from the menu (right-click on table Orders), or double-click the first table to open the join canvas (physical layer).
- Add a table (table People in the example) by drag and drop it into the physical layer canvas.

The screenshot shows the Tableau Data Source interface. On the left, under 'Connections', there is a single connection named 'Sample - Superstore'. Under 'Sheets', several tables are listed: 'Orders', 'People', 'Returns', 'Orders', 'People', and 'Returns'. The 'People' table is highlighted with an orange box and has an orange arrow pointing towards the 'Orders' table in the main pane. The main pane displays the 'Orders' table, which is composed of one table ('Orders'). Below this, the 'Fields' section shows three fields: Row ID, Order ID, and Order Date, all mapped to the 'Orders' table. To the right, a preview of the data is shown in a grid format with 100 rows.

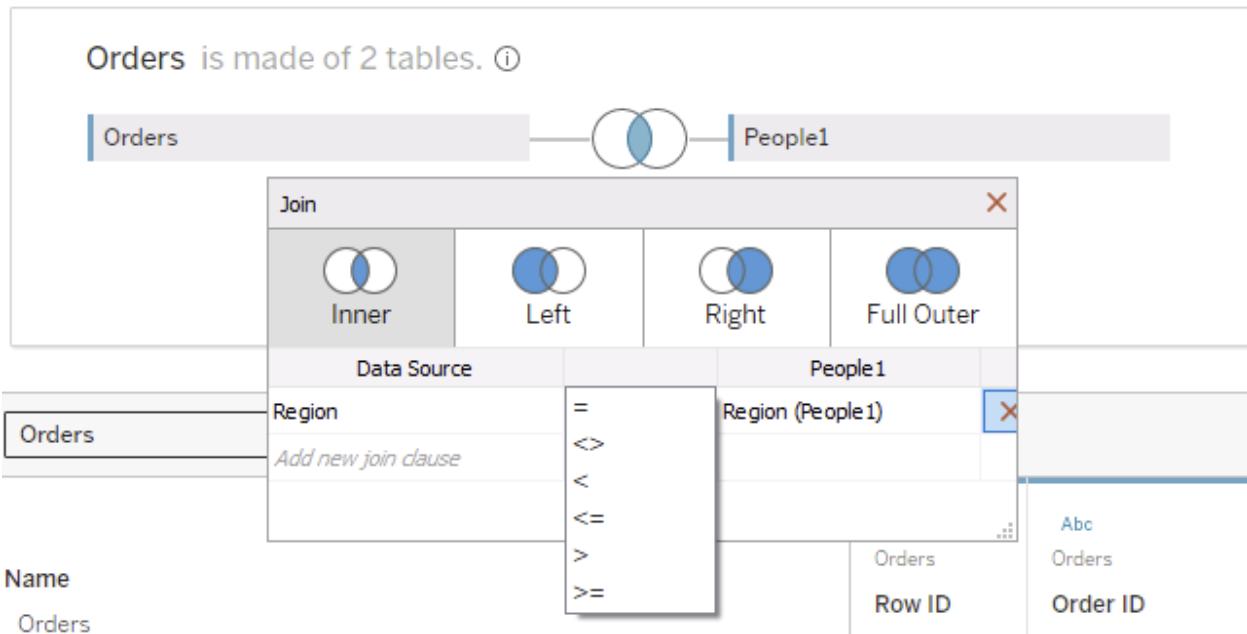
#	Abc Orders	Orders	Orders	Abc Orders	Abc Orders
Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID
5	US-2018-108966	10/11/2018	10/18/2018	Standard Class	SO-20335
13	CA-2020-114412	4/15/2020	4/20/2020	Standard Class	AA-10480
15	US-2018-118983	11/22/2018	11/26/2018	Standard Class	HP-14815
16	US-2018-118983	11/22/2018	11/26/2018	Standard Class	HP-14815
17	CA-2017-105893	11/11/2017	11/18/2017	Standard Class	PK-19075
22	CA-2019-137330	12/9/2019	12/13/2019	Standard Class	KB-16585
23	CA-2019-137330	12/9/2019	12/13/2019	Standard Class	KB-16585



2. Click the join icon to configure the join.



3. Add one or more join clauses by selecting a field from one of the available tables used in the data source, choosing a join operator, and a field from the added table.



4. When finished, close the join dialog and join canvas.

Joins are defined by their type as well as the join clause.

Join types

In general, there are four types of joins that you can use in Tableau: inner, left, right, and full outer. If you aren't sure what join type you want to use to combine data from multiple tables, you should use relationships.

Join Type	Result
Inner 	<p>When you use an inner join to combine tables, the result is a table that contains values that have matches in both tables.</p> <p>When a value doesn't match across both tables, it is dropped entirely.</p>
Left 	<p>When you use a left join to combine tables, the result is a table that contains all values from the left table and corresponding matches from the right table.</p> <p>When a value in the left table doesn't have a corresponding match in the right table, you see a null value in the data grid.</p>
Right 	<p>When you use a right join to combine tables, the result is a table that contains all values from the right table and corresponding matches from the left table.</p> <p>When a value in the right table doesn't have a corresponding match in the left table, you see a null value in the data grid.</p>
Full outer 	<p>When you use a full outer join to combine tables, the result is a table that contains all values from both tables.</p> <p>When a value from either table doesn't have a match with the other table, you see a null value in the data grid.</p>
Union 	<p>Though union is not a type of join, union is another method for combining two or more tables by appending rows of data from one table to another. Ideally, the tables that you union have the same number of fields, and those fields have matching names and data types. For more information about union, see Union Your Data.</p>

Not all databases support all join types. If an option is unavailable in the join dialog, it is likely due to a constraint from your data source.

Union Your Data

You can union your data to combine two or more tables by appending values (rows) from one table to another. To union your data in Tableau data source, the tables must come from the same connection.

If your data source supports union, the New Union option displays in the left pane of the data source page after you connect to your data. Supported connectors may vary between Tableau Desktop and Tableau Server and Tableau Cloud.

The screenshot shows the Tableau Data Source page for 'Sample - Superstore'. On the left, under 'Connections', 'Sample - Superstore' is listed. Under 'Sheets', several tables are listed: 'Orders', 'People', 'Returns', 'Orders' (selected), 'People', and 'Returns'. At the bottom of the left sidebar, there is a button labeled 'New Union' which is highlighted with a red box. The main pane is titled 'Sample - Superstore'.

For best results, the tables that you combine using a union must have the same structure. That is, each table must have the same number of fields, and related fields must have matching field names and data types.

For example, suppose you have the following customer purchase information stored in three tables, separated by month. The table names are "May2016," "June2016," and "July2016."

May2016				June2016				July2016			
DAY	CUSTOMER	PURCHASES	TYPE	DAY	CUSTOMER	PURCHASES	TYPE	DAY	CUSTOMER	PURCHASES	TYPE
4	Lane	5	Credit	1	Lisa	3	Credit	2	Mario	2	Credit
10	Chris	6	Credit	28	Isaac	4	Cash	15	Wei	1	Cash
28	Juan	1	Credit	28	Sam	2	Credit	21	Jim	7	Cash

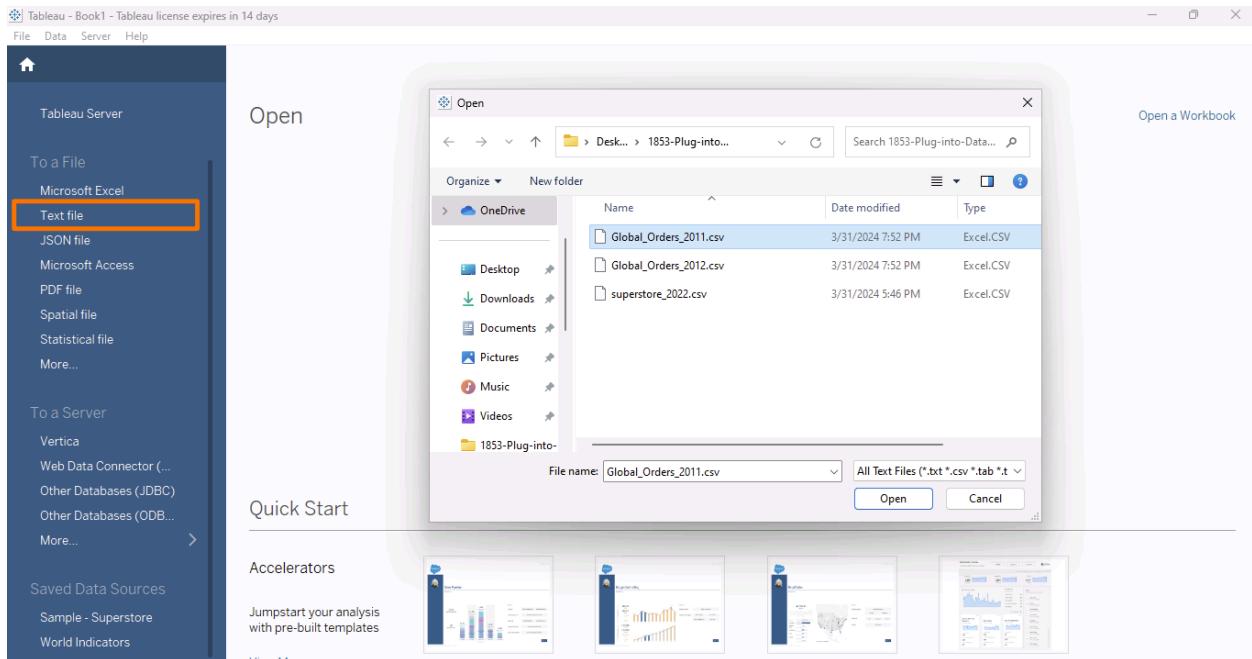
A union of these tables creates the following single table that contains all rows from all tables.

Union

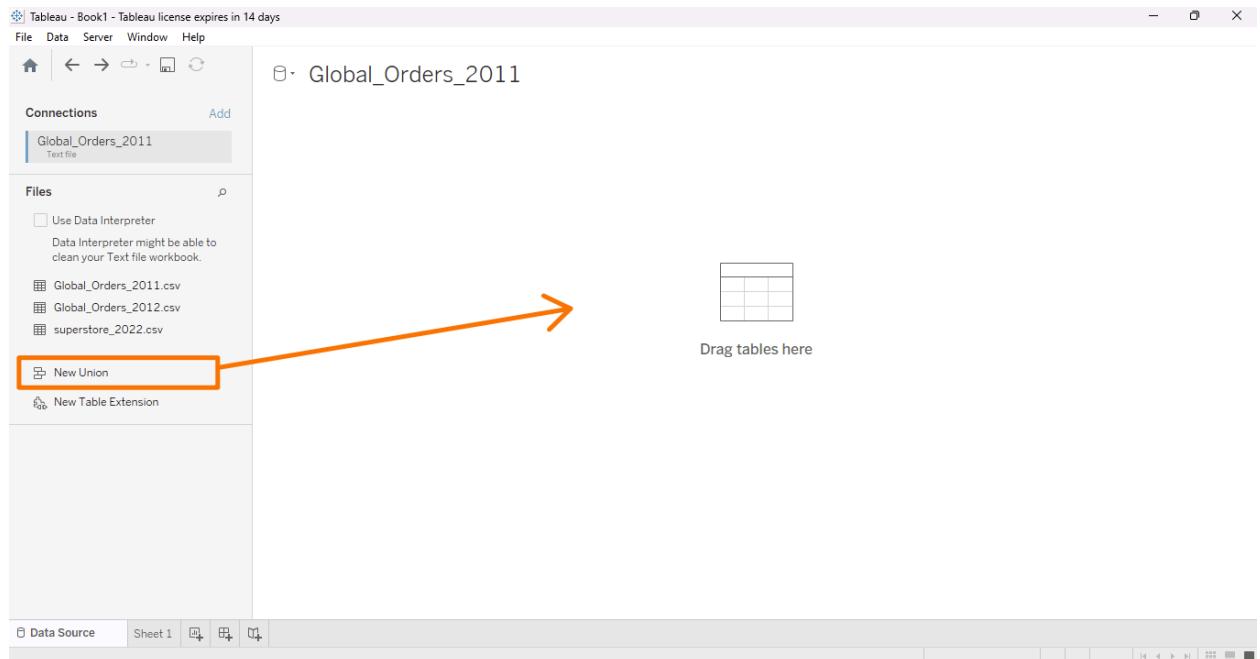
DAY	CUSTOMER	PURCHASES	TYPE
4	Lane	5	Credit
10	Chris	6	Credit
28	Juan	1	Credit
1	Lisa	3	Credit
28	Isaac	4	Cash
28	Sam	2	Credit
2	Mario	2	Credit
15	Wei	1	Cash
21	Jim	7	Cash

Use this method to manually union distinct tables. This method allows you to drag individual tables from the left pane of the Data Source page and into the Union dialog box.

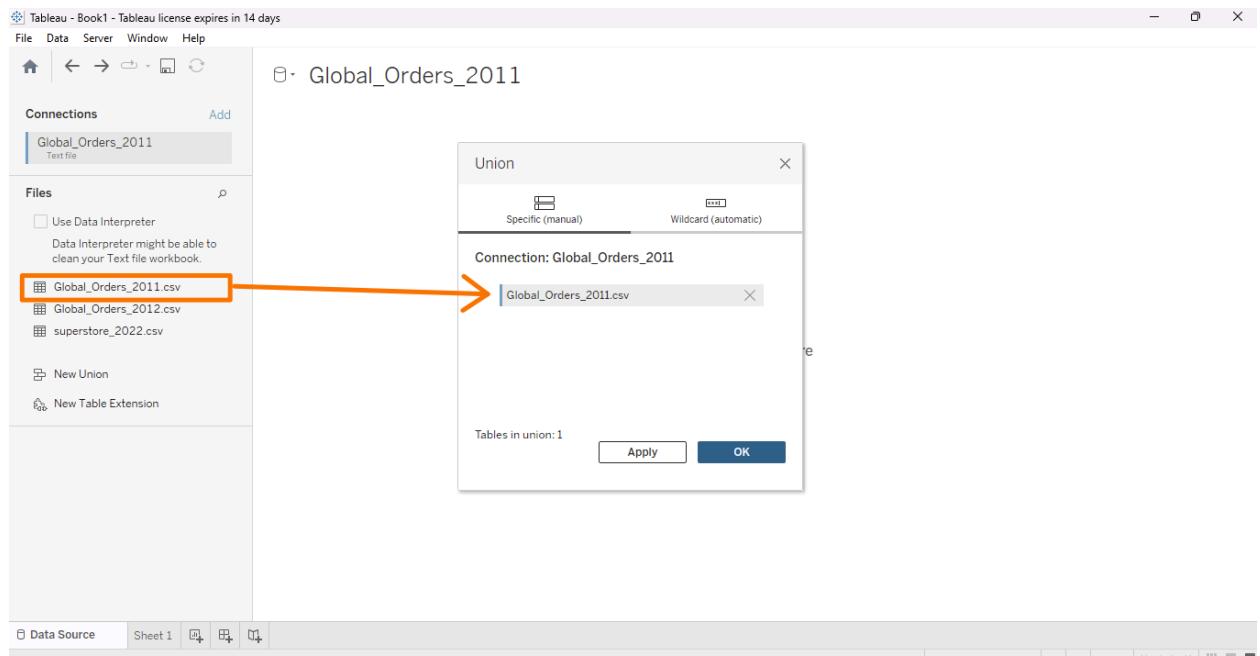
1. Connect to CSV data sources first.
 - a. We are going to use the Text file connector to add **Global_Orders_2011.csv** and **Global_Orders_2012.csv** files into Tableau Desktop.



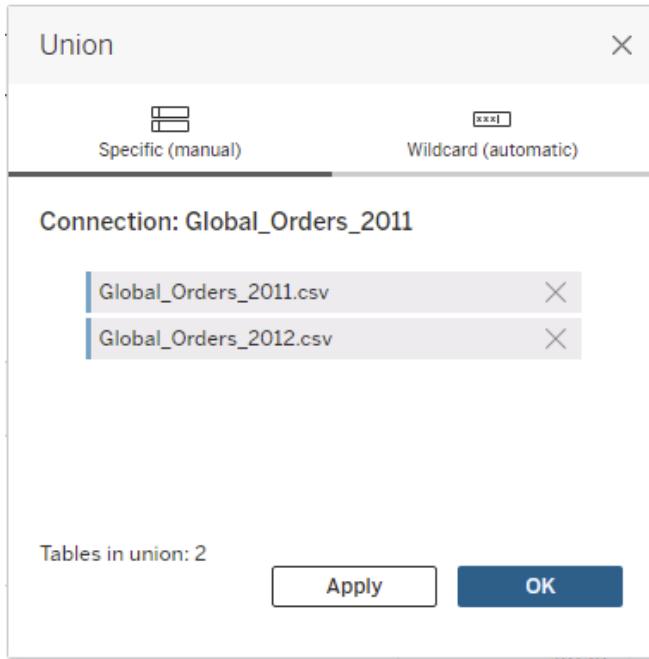
2. On the data source page, double-click New Union to set up the union.



3. Drag a table from the left pane to the Union dialog box.



4. Select another table from the left pane and drag it directly below the first table.



5. Click Apply or OK to union.
6. Rename the union to **Global Orders**.

Global_Orders_2011

Connection Live Extract Filters 0 | Add

Global Orders is made of 2 tables. ⓘ

Union

Name	Global Orders		
Type	Field Name	Physical Table	Remote Fie...
#	Row ID	Union	Row ID
Abc	Order ID	Union	Order ID
abc	Order_Date	Union	Order_Date
abc	Ship Date	Union	Ship Date

Fields

Type	Field Name	Physical Table	Remote Fie...
#	Row ID	Union	Row ID
Abc	Order ID	Union	Order ID
abc	Order_Date	Union	Order_Date
abc	Ship Date	Union	Ship Date

Global Orders 25 fields 19960 rows 100 → rows ⚙️

	# Union	# Union	# Union	Abc Union	Abc Union
	Discount	Profit	Shipping Cost	Order Priority	Table Name
2	0.000000	106.14	35.460	Medium	Global_Orders_2011.csv
4	0.000000	26.76	5.110	Medium	Global_Orders_2011.csv
1	0.000000	2.61	0.820	Medium	Global_Orders_2011.csv
1	0.000000	2.10	0.510	Medium	Global_Orders_2011.csv
1	0.000000	12.72	2.000	Medium	Global_Orders_2011.csv
2	0.000000	14.58	4.360	Medium	Global_Orders_2011.csv
2	0.000000	46.02	52.460	High	Global_Orders_2011.csv
1	0.000000	8.49	13.700	High	Global_Orders_2011.csv

There are many ways to combine data tables, each with their own preferred scenarios and nuances.

Relate	<p>Use when combining data from different levels of detail.</p> <ul style="list-style-type: none">• Requires matching fields between two logical tables. Multiple matching field pairs can define the relationship.• Automatically uses correct aggregations and contextual joins based on how fields are related and used in the viz.• Supports many-to-many and outer joins.• Relationships are consistent for the entire workbook and can be published.• Can be published, but you can't relate published data sources.• Can't define relationships based on calculated fields or geographic fields.• Using data source filters limits join culling benefits of relationships.
Join	<p>Use when you want to add more columns of data across the same row structure.</p> <ul style="list-style-type: none">• Requires common fields between two physical tables.• Requires establishing a join clause and a join type.• Can join on a calculation.• Joined physical tables are merged into a single logical table with a fixed combination of data.• May cause data loss if fields or values aren't present in all tables (dependent on join types used).• May cause data duplication if fields are at different levels of detail.• Can use data source filters.
Union	<p>Use when you want to add more rows of data with the same column structure.</p> <ul style="list-style-type: none">• Based on matching columns between two tables.• Unioned physical tables are merged into a single logical table with a fixed combination of data.

Filter Data from Data Sources

You can create filters on a data source, thereby reducing the amount of data in the data source. This feature is available to Creators when authoring in Tableau Cloud and Tableau Server.

Data source filters can be useful for restricting the data users can see when you publish a workbook or data source. When you publish a data source to Tableau Server, the data source and any associated files or extracts are transported in entirety to the Server. As you publish a data source you can define access permissions for downloading or modifying the data source, and you can also choose the users and groups who can remotely issue queries through Tableau Server against that data source. When users have query permission and no download permission, you can share a rich data model having calculated fields, aliases, groups, sets and more—but only for querying.

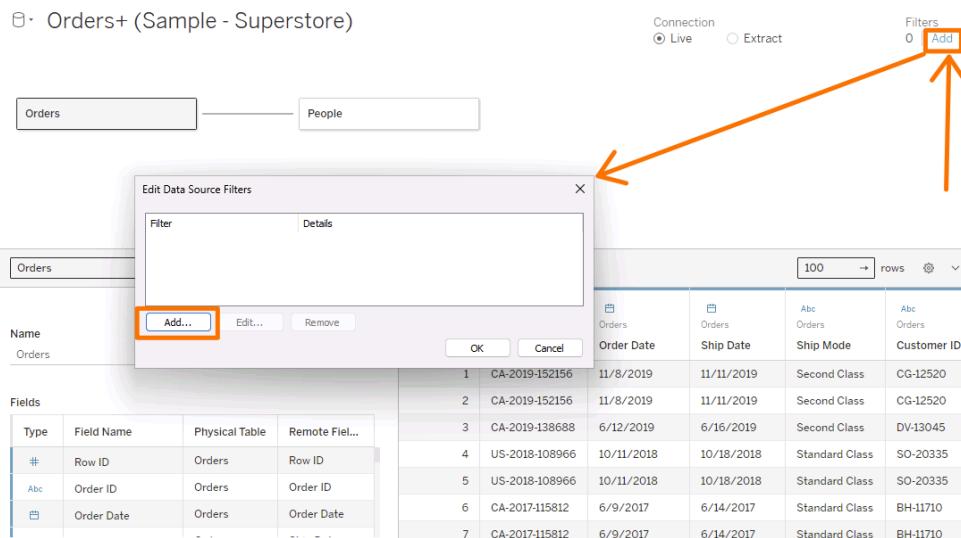
Furthermore, users who query published data sources will never be able to see or modify any data source filters present on the originally published data source, but all of the users' queries will be subject to those data source filters. This is a great way to offer a restricted subset of your data, for example by filtering dimensions for specific users and groups, or by defining data source filters based on a fixed or relative date range. This is often useful for data security, and it also allows you to manage performance of the remote database which Tableau Server will ultimately query on a user's behalf. For systems that rely heavily on partitions or indexing, data source filters may yield tremendous control over the performance of queries issued by Tableau.

1. Create a data source filter.

The primary way to create a data source filter is from the data source page.

- On the data source page, click **Add** in the Filters section in the upper-right corner of the page.
- To create a data source filter on a worksheet, right-click (control-click on a Mac) the data source and choose **Edit Data Source Filters**.

Whether you start from the Data Source page or from a worksheet, you see an Edit Data Source Filter dialog box, listing any existing data source filters.



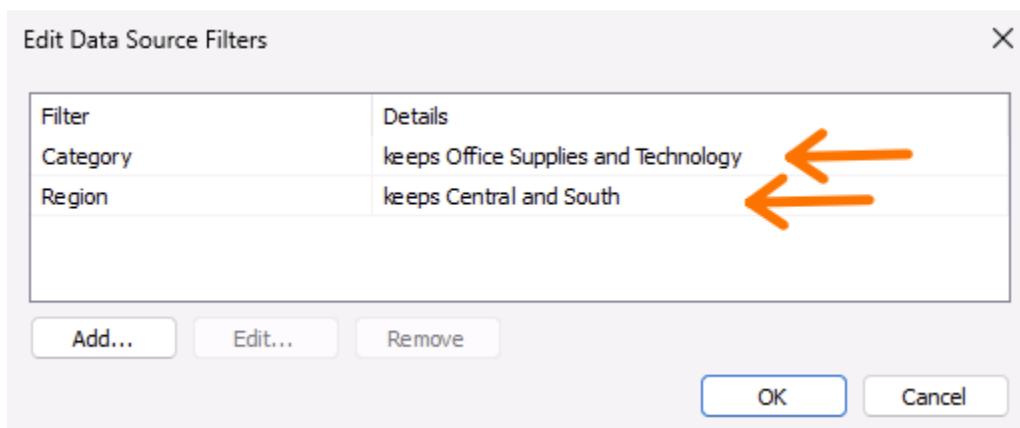
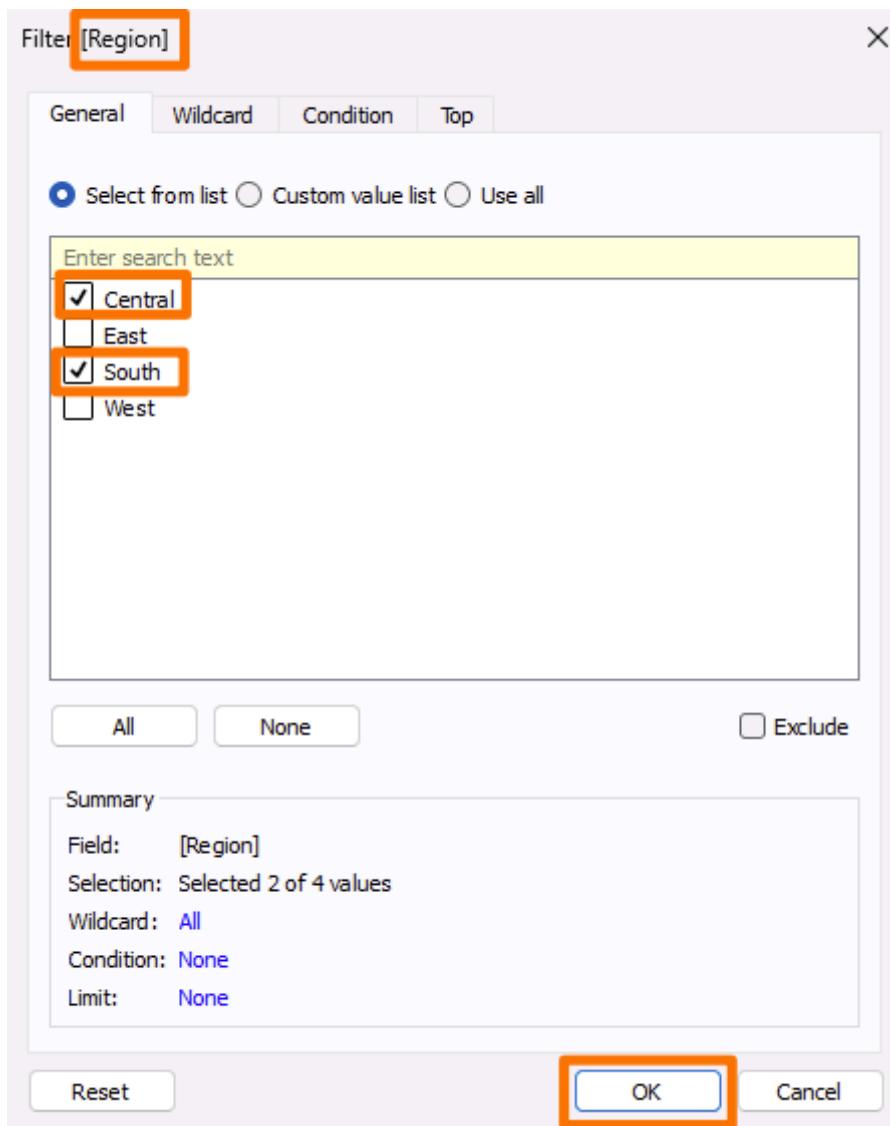
2. Click **Add** to open an Add Filter dialog box listing all fields in the data source.

The screenshot shows the Power BI desktop interface. At the top, there's a title bar with 'Orders+ (Sample - Superstore)' and connection settings ('Live' selected). On the right, it says 'Filters 0 | Add'. Below the title bar, there's a 'Edit Data' pane on the left containing sections for 'Orders', 'Name', and 'Fields', with a table preview showing columns like Order ID, Order Date, Ship Date, and Ship Mode. Overlaid on this is the 'Add Filter' dialog box, which has a search bar at the top and a list of fields below it. The 'Category' field is highlighted with a blue selection bar. At the bottom of the dialog are 'OK' and 'Cancel' buttons.

3. Click to select a field to filter; then specify how the field should be filtered, just as you would for a field on the Filters shelf.

This screenshot shows the 'Filter [Category]' dialog box. The 'General' tab is selected. The 'Select from list' radio button is selected. In the main area, there's a list box with three items: 'Furniture' (unchecked), 'Office Supplies' (checked with a blue selection bar), and 'Technology' (checked with a blue selection bar). Below the list are buttons for 'All', 'None', and 'Exclude'. At the bottom, there's a 'Summary' section with details: Field: [Category], Selection: Selected 2 of 3 values, Wildcard: All, Condition: None, Limit: None. The 'OK' button is highlighted with a blue selection bar at the bottom right.

4. To add an additional data source filter, repeat this procedure.



If you create an extract from a data source that already has data source filters in place, those filters are automatically recommended as extract filters, and will appear in the Extract dialog. Those recommended filters are not required to be part of the Extract filter list, and can safely be removed without affecting the existing set of data source filters.

Extract Your Data

A data extract is a subset of information that is saved separately from the original dataset. It serves two purposes: to enhance performance and to utilize Tableau features that may not be available or supported in the original data. By creating a data extract, you can effectively reduce the overall data volume by applying filters and setting other limitations.

Once a data extract is created, it can be refreshed with the latest data from the original source. During the refresh process, you have the flexibility to choose between a full refresh, which replaces all existing content in the extract, or an incremental refresh, which only includes new rows since the previous refresh.

Benefits of Extracts

- Handling large datasets: Extracts can handle massive amounts of data, even reaching billions of rows. This allows users to work with extensive datasets efficiently.
- Improved performance: Interacting with views that utilize extract data sources results in better performance compared to views connected directly to the original data. Extracts optimize query performance, resulting in faster data analysis and visualization.
- Enhanced functionality: Extracts provide access to additional Tableau functionality that may not be available or supported by the original data source.
For instance, users can leverage extracts to compute Count Distinct, enabling more advanced calculations and analysis.
- Offline data access (Tableau Desktop): Extracts allow for offline access to data. This means that even when the original data source isn't available, users can still save, manipulate, and work with the data locally.

There are multiple options available within your Tableau workflow to create an extract, but the main approach is explained below.

- After you connect to your data and set up the data source on the Data Source page, in the upper-right corner, select Extract.

Tableau - Book1 - Tableau license expires in 14 days

File Data Server Window Help

Connections Add

Sample - Superstore Microsoft Excel

Sheets

- Orders
- People
- Returns
- Orders
- People
- Returns
- New Union
- New Table Extension

Orders

Orders+ (Sample - Superstore)

Connection Live Extract

Filters 2 | Edit

Orders 23 fields 3130 rows

#	Orders	Abc Orders	Orders	Orders	Abc Orders	Abc Orders
Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	
5	US-2018-108966	10/11/2018	10/18/2018	Standard Class	SO-20335	
13	CA-2020-114412	4/15/2020	4/20/2020	Standard Class	AA-10480	
15	US-2018-118983	11/22/2018	11/26/2018	Standard Class	HP-14815	
16	US-2018-118983	11/22/2018	11/26/2018	Standard Class	HP-14815	
17	CA-2017-105893	11/11/2017	11/18/2017	Standard Class	PK-19075	
22	CA-2019-137330	12/9/2019	12/13/2019	Standard Class	KB-16585	
23	CA-2019-137330	12/9/2019	12/13/2019	Standard Class	KB-16585	

Data Source Sheet 1

- And then select the Edit link to open the Extract Data dialog box.

Orders+ (Sample - Superstore)

Extract Data

Connection Live Extract [Edit](#) Refresh

Extract will contain subset of data.

Data Storage Logical Tables [Edit](#)

Filters 2 filters applied [Edit](#)

Aggregation No data aggregated [Edit](#)

Number of Rows All rows [Edit](#)

Incremental Refresh Not enabled [Edit](#)

Learn more about setting up data extracts.

History... Hide All Unused Fields Cancel Save Settings

Orders 100 rows

#	Orders	Abc Orders	Abc Orders
Date	Ship Date	Ship Mode	Customer ID
2018	10/18/2018	Standard Class	SO-20335
2020	4/20/2020	Standard Class	AA-10480
2018	11/26/2018	Standard Class	HP-14815
2018	11/26/2018	Standard Class	HP-14815
2017	11/11/2017	11/18/2017	Standard Class
22	CA-2019-137330	12/9/2019	12/13/2019
23	CA-2019-137330	12/9/2019	12/13/2019

3. Under **Data Storage**, select either Logical Tables or Physical Tables. For assistance with this step see the [Data Storage](#) section.
4. Expand **Filters** to set up filters to limit how much data gets extracted based on fields and their values.
5. Select **Aggregate data for visible dimensions** to aggregate the measures using their default aggregation.

Extract Data

[Expand All](#)

▼ **Data Storage**

Logical Tables
Store data using one table for each logical table.
Use this option if you need to use extract filters, aggregation, or other extract settings.

Physical Tables
Store data using one table for each physical table.

▼ **Filters**

[ABC Category excludes Furniture and Null](#)  

[ABC Region keeps Central and South](#)  

[Add Filter ▾](#)

▼ **Aggregation**

Aggregate data for visible dimensions

Roll up dates to Month ▾

[Learn more about setting up data extracts.](#)

[History...](#) [Hide All Unused Fields](#) [Cancel](#) **Save Settings**

Save Customization as a Saved Data Source

If you've created a data connection that you might want to use with other workbooks or share with colleagues, you can export (save) the data source to a file. You might want to do this also if you've added joined tables, default properties, or custom fields—such as groups, sets, calculated fields, and binned fields—to the Data pane.

In this exercise, you will learn how to save a data source to a TDS format (Tableau Data Source):



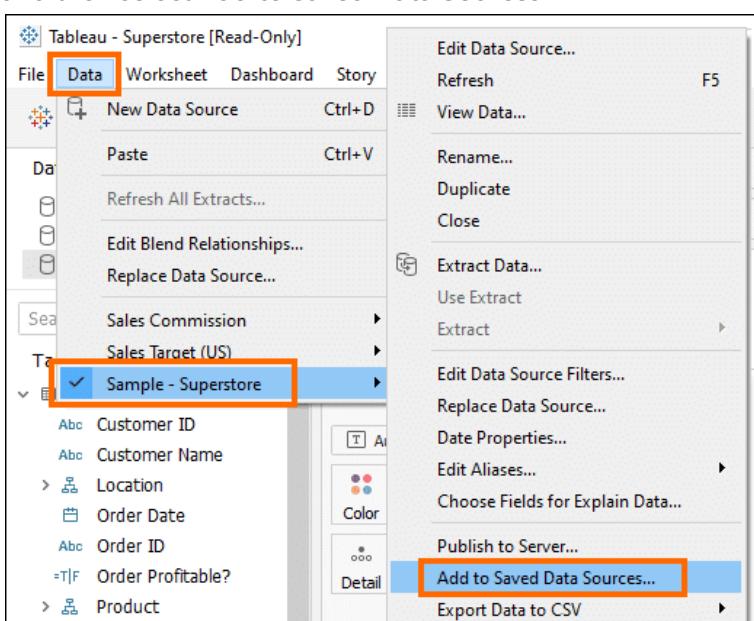
Data Source (.tds) - contains only the information you need to connect to the data source, including the following:

- Data source type
- Connection information specified on the data source page; for example, database server address, port, location of local files, tables
- Groups, sets, calculated fields, bins
- Default field properties; for example, number formats, aggregation, and sort order

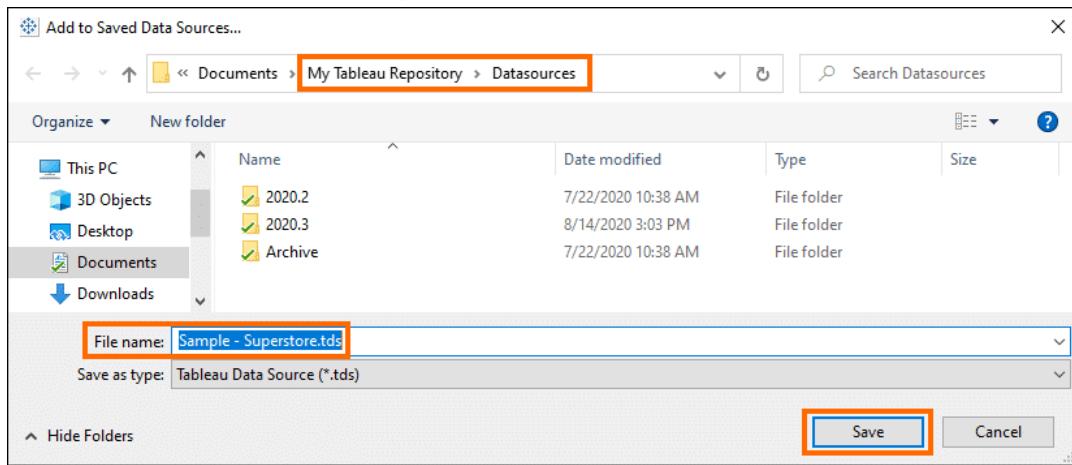
Use this format if everyone who will use the data source has access to the underlying file or database defined in the connection information. For example, the underlying data is a CSV file on your computer, and you are the only person who will use it; or the data is hosted on a cloud platform, and your colleagues all have the same access you do.

Save a data source

1. In Tableau Desktop, open the workbook that has the connection to the data you want to save as a file.
2. At the top of the Data pane, right-click (Control-click on Mac) the name of the data source, and then select **Add to Saved Data Sources**.



3. Enter a filename, select the file type (.tds), and then click **Save**.



Connect to your data source

1. By default, Tableau saves .tds files to the Datasources folder under your Tableau repository. When you use the default location, you can connect to the data source on the Connect pane.



2. If you specified a different location, you can connect to the data source by selecting File > Open and navigating to it.

Publish Your Data Source

Suppose you create a view that exposes a new range of questions in the data you're using, and you want to share this analysis with other people using this data. Or maybe you are your team's Data Steward, in charge of building the data models approved for use by analysts, and meeting your organization's requirements for security, compliance, performance, and so on.

You can share your work with the rest of your team by publishing it to Tableau Server or Tableau Cloud. After it's published, you and your team can access it through your web browser or the Tableau mobile app. Publishing data sources can also help you to centralize data management.

Why publish

You can publish data sources and workbooks when you want to widen the audience for your data analysis within your organization. By publishing you can begin to do the following:

- Collaborate and share with others
Allow people in your organization to view, interact with, download, subscribe to, share, edit, and save published views, even if they do not use Tableau Desktop. Incorporate views into blog posts or websites.
- Centralize data and database driver management
Create and publish data models that everyone can use. Centralized data management allows for sharing a single source for your Tableau data. All workbooks connected to the published data reflect updates to it.
In addition, when you publish and connect to data on the server, people connecting to the data from Tableau Desktop do not need to install and maintain database drivers on their own computers.
- Support mobility
Access your data from a different computer or location, through a web browser or the Tableau Mobile iOS app. Sign in to your organization's Tableau Server from a private network offsite.

Who can publish

To publish to Tableau Server or Tableau Cloud, your server or site administrator must grant you the following capabilities:

- A site role of Creator (formerly Publisher) on the site you're publishing to.
- View and Save capabilities set to Allowed on the project into which you publish.

If you use Tableau Desktop and are not sure whether you can publish to a server, or you are having trouble publishing, see your Tableau administrator. If you're an administrator, see [Content Access and Ownership](#) in the Tableau Server help (or the [Tableau Cloud version](#)) for more information about site roles and permissions.

When you are ready to make a data source available to other users, you can publish it to Tableau Server or Tableau Cloud. If the data source is in a workbook that you published to Tableau Server or Tableau Cloud, you can make it available by saving it, provided it's an embedded Excel or text file.

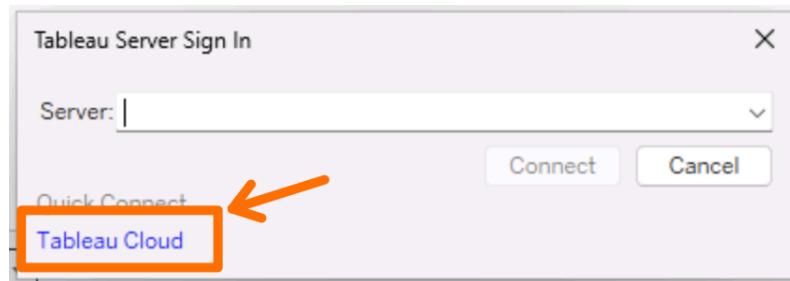
The following steps give an overview of the publishing workflow you will use regardless of the type of data or the server you publish to.

1. If you're not already signed in to Tableau Server or Tableau Cloud, sign in now.

Sign-in to the Tableau Server or Tableau Cloud site where you want to publish your data source.

The screenshot shows the Tableau desktop interface. The main window displays the 'Orders+' data source from the 'Sample - Superstore' sample data. On the left, the 'Sheets' pane shows the 'Orders' sheet selected. The top navigation bar has 'Server' highlighted with a red box, and the 'Sign In...' option under it is also highlighted with a red box. The bottom of the window shows the 'Data Source' tab selected.

2. We are using a Tableau Cloud site for this exercise. Click on Tableau Cloud link on the sign-in window.



3. Enter this credential to sign-in:

- Username = **workshopusertableau@gmail.com**
- Password = **Salesforce.12345**



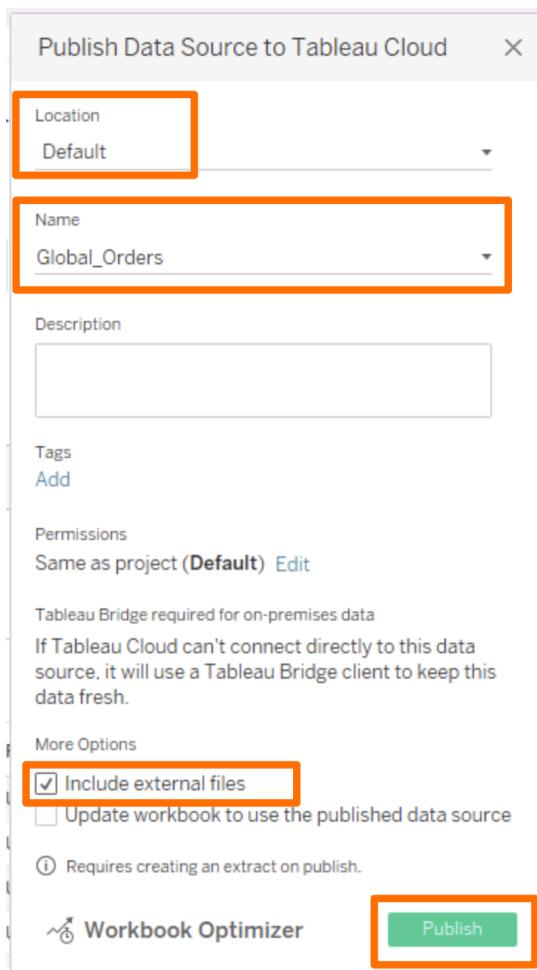
4. Select Server > Publish Data Source.

If your workbook is connected to multiple data sources, select the one you want from the Publish Data Sources submenu

Name	Type	Field Name	Physical Table	Remote File...
#	Row ID	Union	Row ID	
Abc	Order ID	Union	Order ID	
	Order_Date	Union	Order_Date	
	Ship Date	Union	Ship Date	

5. In the **Publish Data Source** dialog box, do the following:

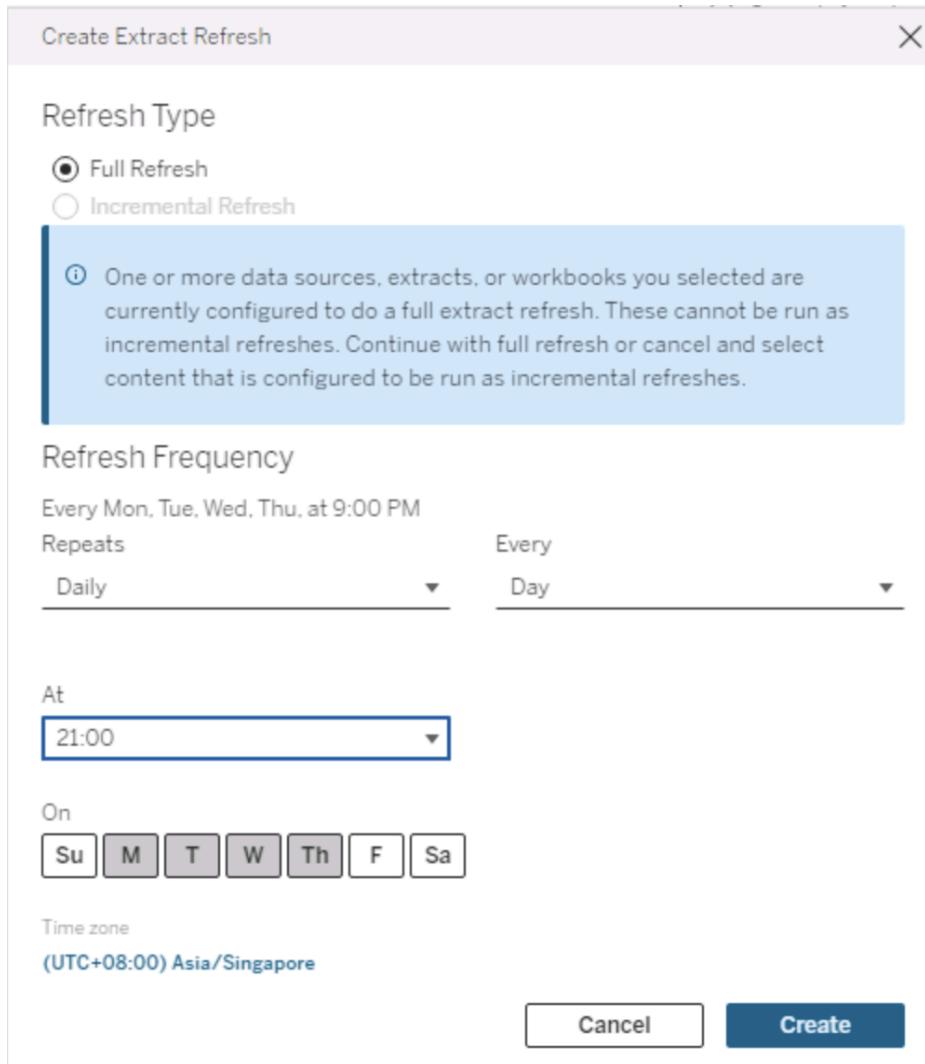
- For **Location**, select the project you want to publish to and enter the data source name.
 - We use the “**Default**” project as our location.
- For **Name**, add a name that will help you and other users find the data source.
 - You can use your unique identifier such as your name, your initial, etc.
- For **Description** and **Tags**, add a description and tags that will help you and other users find it. Separate tags using either a comma or a space. To add a tag that contains a space, put it in quotation marks (e.g., “Sales Quotes”).
- If you are publishing file-based data that is on a Windows mapped drive, or using images that will not be available from the server, select **Include external files**.
When you include external files, copies of the files are put on the server as part of your data source. Copies of files are also put on the server and included as part of the data source when you publish extracts of multi-connection data sources that contain a connection to file-based data such as Excel.
- For **Permissions**, accept the default project settings.
Generally a site administrator manages permissions on the server. If you think your data source is an exception, work with your administrator to determine the best course of action, and see [Set Permissions as You Publish a Data Source or Workbook](#).



6. (Optional) Set up a refresh schedule on the server or Tableau Cloud.

You can schedule refresh tasks directly on Tableau Cloud for extracts of cloud-hosted data.

- a. When you're signed in to Tableau Cloud, select Explore from the left navigation pane, and then, depending on the type of content you want to refresh, select All Workbooks or All Data Sources from the drop-down menu.
- b. Select the check box for the workbook or data source you want to refresh, and then select: Actions > Refresh Extracts.



Conclusion

This hands-on Tableau guide has equipped you with the essential knowledge and practical skills needed to confidently publish your data sources. By following these step-by-step instructions and best practices, you've learned how to seamlessly integrate your data into Tableau and make it accessible to your team or organization.

As you embark on your data journey, remember that the power of Tableau lies not only in its robust analytics capabilities but also in its ability to democratize data and drive informed decision-making.

Armed with this guide, you're now well-positioned to leverage Tableau's publishing features to unlock new insights and drive meaningful change within your organization.

Happy analyzing!