MICHAEL EBY

michaeleby1@gmail.com | GitHub | Blog | LinkedIn

Table of Contents

Universal search

Starburst Galaxy's **universal search** lets you locate your data entities in Galaxy.

Search overview

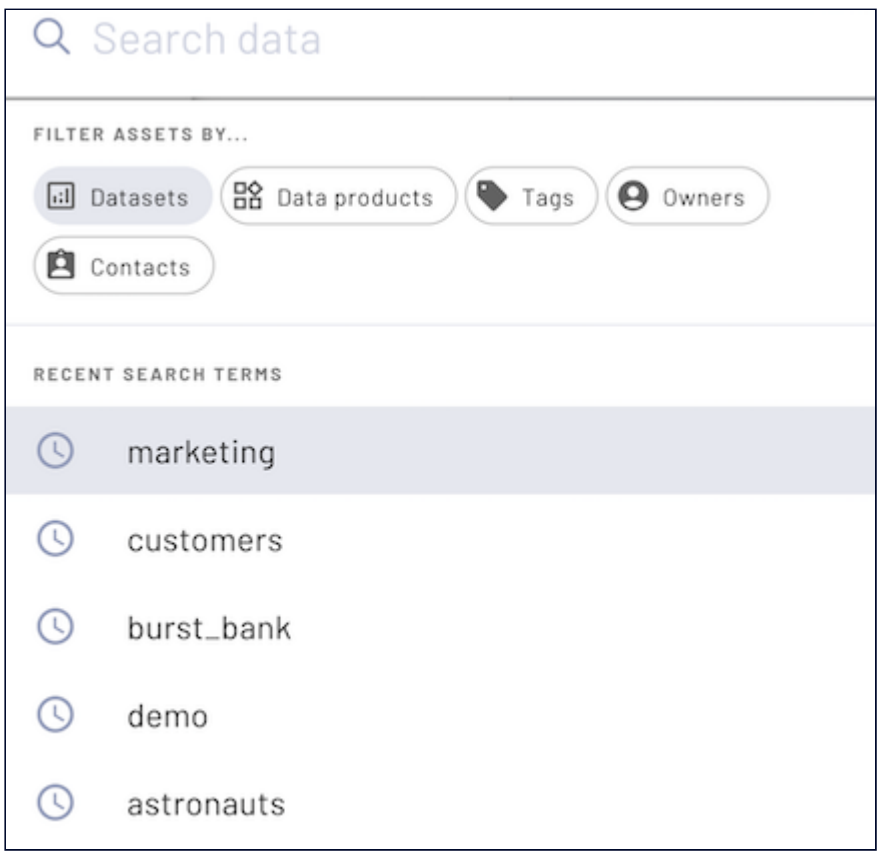
Use universal search to locate any data object by name or metadata name anywhere in your Starburst Galaxy account. You can search for the names of catalogs, schemas, tables, views, columns, data products, tags, owners, or contacts. Universal search parses Galaxy's cached description of your account, which means it does not require any cluster to be enabled or running.

The search type is a case-insensitive anchored search that matches entity names by trying to match the first characters of the search term to the first characters of various target strings. As a result, to locate an entity named `token`, the search string `tok` matches, but search string `ken` does not.

Using universal search

To begin searching:

1. Click the 🔍 magnifying glass icon in the banner of Starburst Galaxy to open the search field. As an alternative, you can open the search field with the key combination `Cmd + K` on MacOS, or `Ctrl + K` on Windows.



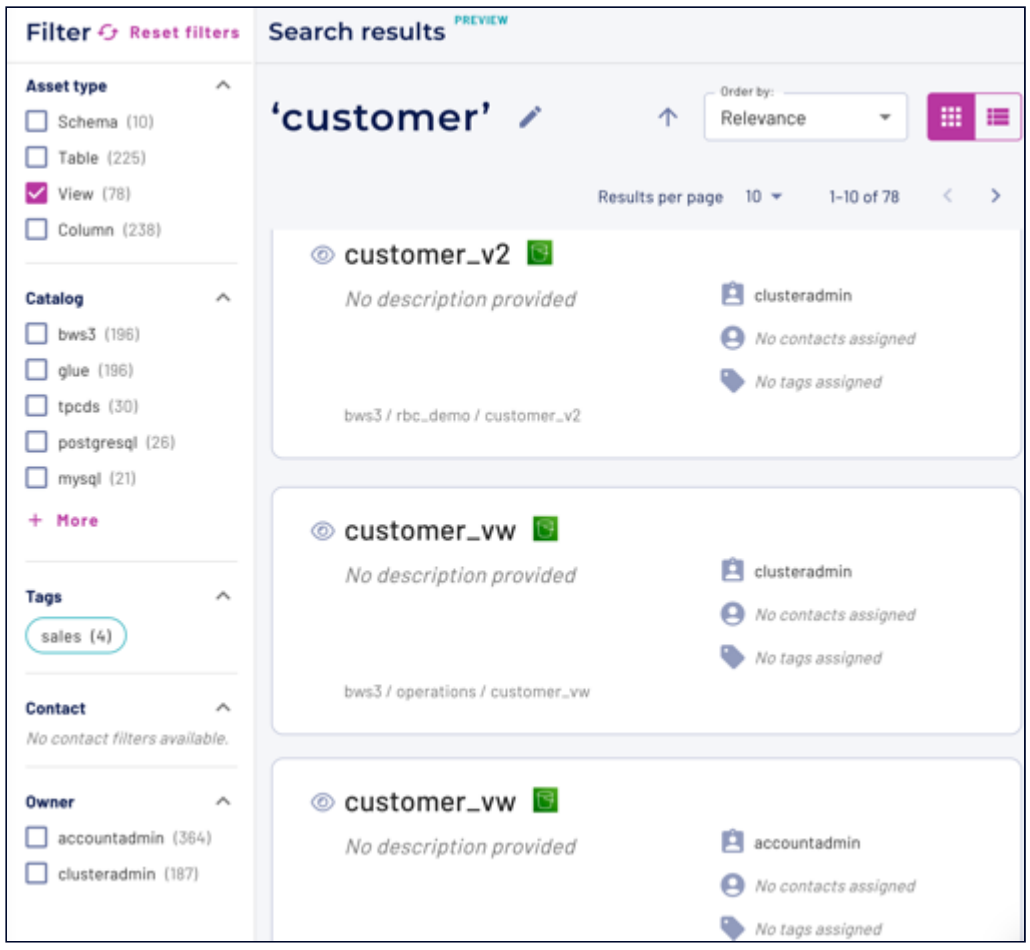
2. In the search field, enter your search term and select a filter category. By default, your search term is matched against any component of dataset names, including catalog, schema, table, view, and column names. You can also restrict the search to the names of data products or tags, as well as the ownership or contact metadata for datasets.
3. Entering a search term shows a truncated preview of the results.
4. Click the matching entity to navigate to that entity in the catalog explorer, or click **View all results** to open the **Search results pane**.

The five most recent search terms are listed in the **Recent search terms** section. Recent terms are saved in browser local storage. Clearing browser data clears recent search terms.

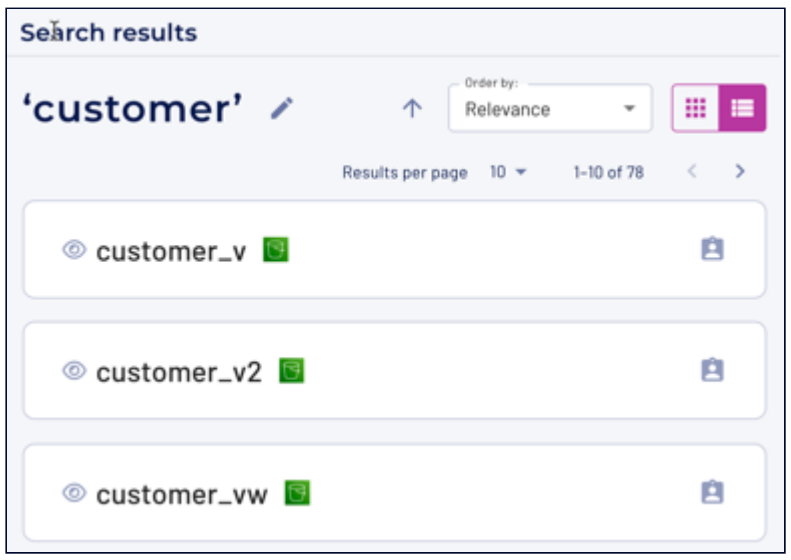
Search results pane

In the search results pane, narrow the scope of the search results using the **Filter** pane. Select one or more filter options to restrict the search results to match selected categories.

For example, to see only views related to your search term, select **View** in the **Asset type** filter category. The results are returned in the default ▦ grid mode:



To use list mode to show only the entity names matching your search term and filters, click the ☰ button:



Available filter categories include:

- **Asset type:** restrict the current search results to match names of catalogs, schemas, tables, views, columns, and/or data products.
- **Catalog:** restrict the current search results to match schema, table, view, or column names in specific catalogs.
- **Tags:** restrict the current search results to match entities that have one or more specified tag names.
- **Contact:** restrict the current search results to match usernames designated as contacts for the entity.
- **Owner:** restrict the current search results to match roles designated as entity owners.

Click the ✎ pencil icon next to the search term to edit the search query or start a new search.

Use the drop-down **Order by** menu to sort results by relevance, name, or description.

Click the **Results per page** drop-down to adjust the number of results you see per page to 5, 10, 25, or more if you have that many results.

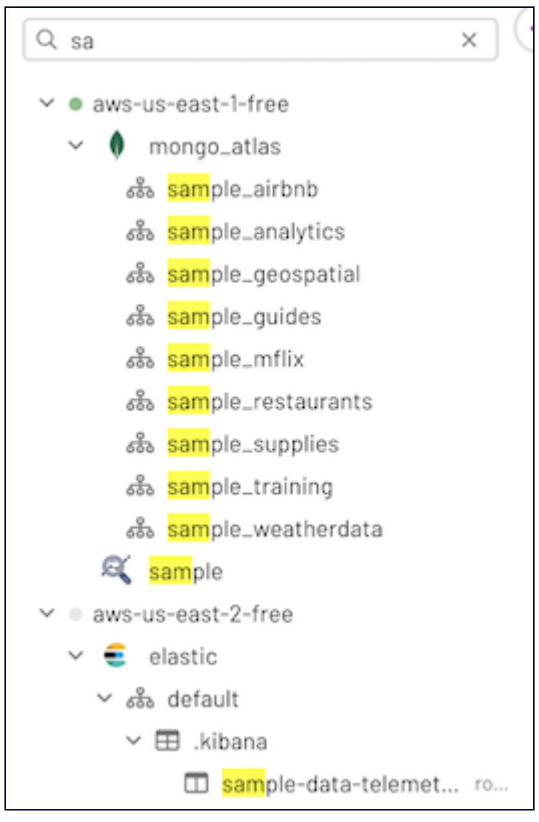
Click a tag in the Starburst Galaxy UI in places such as the catalog explorer and the tags pane to navigate directly to the search results page for that tag.

Search in the query editor

Search for an entity name in the cluster explorer pane of the query editor.

To search your data:

1. If the cluster explorer pane is not visible, click the ⊙ icon to open the pane.
2. Use the **Search data** field denoted by the 🔍 magnifying glass icon at the top of the cluster explorer pane.
3. Type your search term. As you type, entities that match what you are typing are revealed and highlighted in yellow:



In addition to providing search in the cluster explorer, Starburst Galaxy's implementation of universal search provides auto-completion for entity names when typing in the query editor pane.

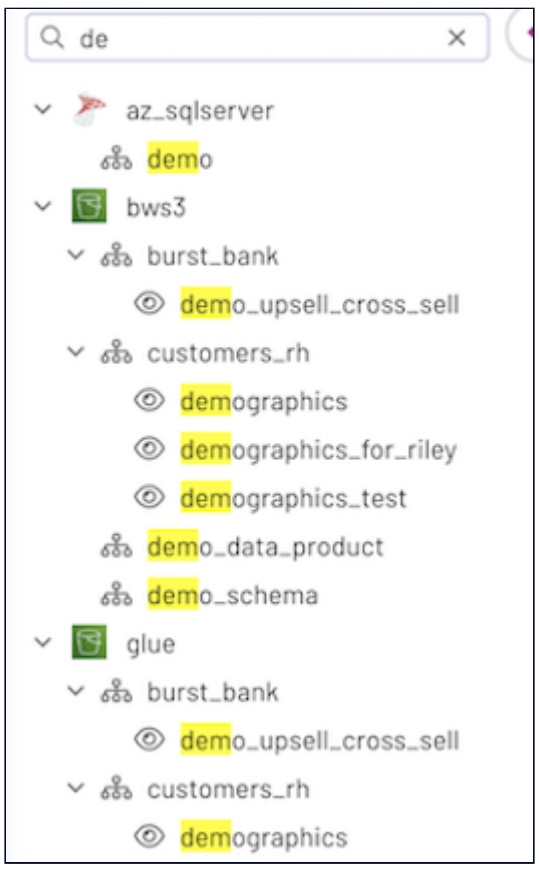
See Query editor > Auto-completion for further information on auto-completion.

Search in the catalog explorer

Search by entity name in the catalog explorer.

To search your data:

1. If the catalog explorer pane is not visible, click the icon to open the pane.
2. Use the **Search data** field denoted by the magnifying glass icon at the top of the catalog explorer pane.
3. Type your search term. As you type, entities that match what you are typing are revealed and highlighted in yellow:



AWS PrivateLink

> **Note:** Starburst Galaxy's support for AWS PrivateLink is a private preview feature. Contact Starburst support with questions or feedback.

General setup phases

To configure a Starburst Galaxy catalog to connect to an AWS data source that is protected with the AWS PrivateLink service, you must configure certain features of the AWS data source to prepare for the connection. There are two phases:

1. **On AWS:** Configure an AWS endpoint service in the AWS Management Console. See the step-by-step instructions for assistance.
2. **In Starburst Galaxy:** Contact your Starburst account team for support.

AWS PrivateLink Overview

Starburst Galaxy supports secure connections to AWS-hosted data sources that are protected with AWS PrivateLink.

AWS data sources can take advantage of the AWS PrivateLink service as one way to secure access without exposing the data source to the public internet. These data sources then operate within a virtual private cloud (VPC) within AWS. Starburst Galaxy also operates securely within its own VPC. Galaxy's support for PrivateLink-secured data sources provides a way to connect VPC to VPC securely within the AWS cloud.

Starburst Galaxy supports AWS PrivateLink for some catalogs. This page provides a general overview of Starburst Galaxy's support for AWS PrivateLink. It is not intended to be a comprehensive guide to creating and administering AWS PrivateLink.

Starburst Galaxy and AWS PrivateLink

With AWS PrivateLink, Starburst Galaxy and your AWS-hosted data service communicate with each other using VPC endpoints. Network traffic between your Galaxy VPC endpoint and your AWS VPC endpoint is secured using private IP addresses. Therefore, you do not need to use an internet gateway or a NAT gateway to connect your cluster to your data source.

> **Note:** You must configure a separate AWS PrivateLink connection for each catalog you would like to connect to Starburst Galaxy.

Contact your Starburst account team to create the VPC endpoint for your Galaxy cluster to use for communication with your AWS VPC endpoint.

AWS endpoint service requirements

Configure your AWS VPC endpoint in the AWS console as an endpoint service. Starburst Galaxy requires that you use a network load balancer to receive the incoming traffic from your Galaxy cluster. You must also create a target group that routes traffic from the cluster to the load balancer.

When you create your Starburst Galaxy cluster and configure a catalog, you must deploy your cluster in the same region as your AWS-hosted data service. Starburst Galaxy does not support cross-region connections with AWS PrivateLink.

Once configured, all traffic from to this data source is routed through AWS PrivateLink. You can federate your queries across multiple data sources in the same cluster that use PrivateLink.

Use Galaxy with Tableau

Welcome to Starburst Galaxy! In this topic, you will learn how to:

- Run simple queries from built-in catalogs.
- Create one or more catalogs from your organization's data sources.
- Connect Tableau to Starburst Galaxy.
- Use Starburst Galaxy to create a Tableau dashboard.

> **Before you begin:**
> This topic assumes the following things:
>
> - You have a Starburst Galaxy login.
> - You have access to one or more supported data sources.
> - You have access to Tableau.

Once you complete all the tasks in this topic, you are ready to share your first Galaxy-powered Tableau dashboard.

Introduction

Tableau is a popular analytics tool with powerful data visualization capabilities. You can access Starburst Galaxy clusters from the following Tableau products:

- Tableau Desktop
- Tableau Prep Builder
- Tableau Server
- Tableau Cloud (Online)

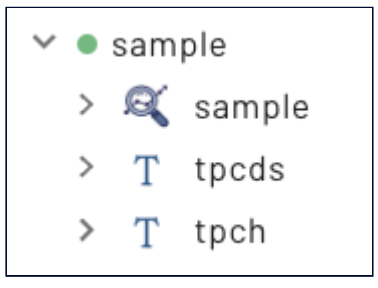
Today, you will learn how to build a dashboard in Tableau using Starburst Galaxy.

Run simple queries using the query editor

Running queries using the Starburst Galaxy query editor is easy.

You can access the query editor at any time by clicking **Query editor** from the **Query** option of the left-hand navigation menu.

Your Starburst Galaxy account comes pre-configured with sample data. Click the > expand icon next to the `sample` cluster to explore the sample catalogs:



From there, expand the `tpch` catalog. TPC-H is a popular sample database used for testing SQL queries and performing analytics. Each of the schemas in the `tpch` catalog differ only in the size of their dataset. For our purposes, we will use the `sf1` catalog, the base row size of the database.

In the query editor, type (or copy and paste) the following SQL statement:

```
SELECT *
FROM "tpch"."sf1"."customer"
ORDER BY acctbal DESC
LIMIT 10;
```

Copy

Click the ▶ **Run** button to execute the query. The result displays information about the ten customers in the `customer` table with the highest account balances.

Congratulations, you just executed your first query in Starburst Galaxy!

Connecting Tableau to Starburst Galaxy

You are now ready to connect Tableau to Starburst Galaxy.

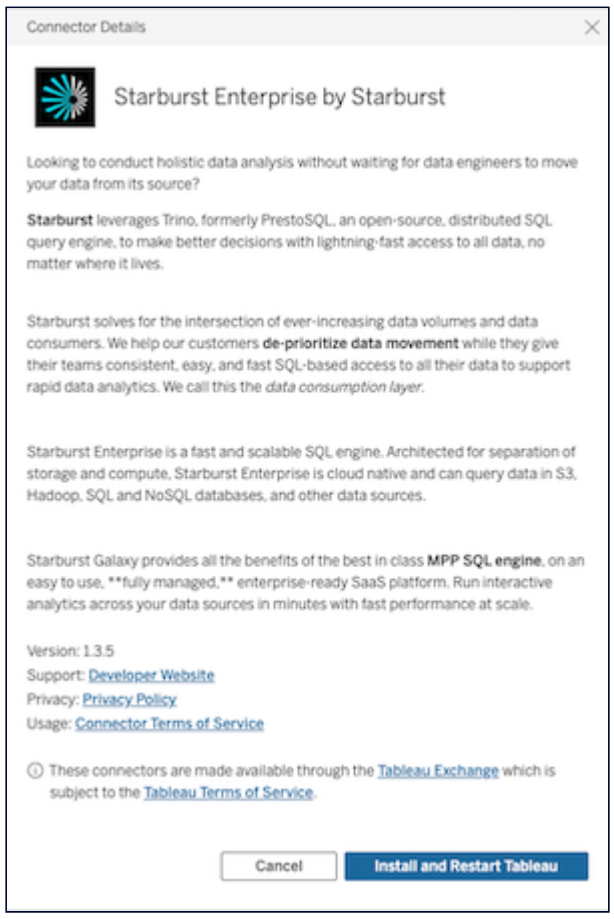
First, open the Tableau product you wish to use. The next steps depend on whether you are connecting to Galaxy via Tableau Desktop or Tableau Server.

Tableau Desktop

> **Note:** Make sure you install the latest Trino JDBC driver and save it to the correct folder.

In Tableau Desktop, take a look at the **Connect** pane on the left side of the **Start** page. Under the **To a Server** heading, click **More...** to see a menu of all the data sources you can connect to. Under **Additional Connectors**, find and select **Starburst Enterprise by Starburst**.

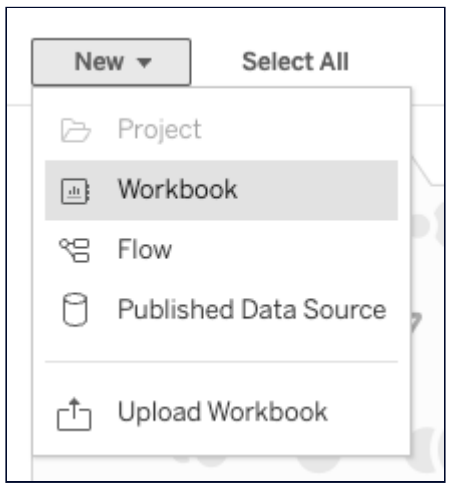
Since this is your first time connecting to Starburst Galaxy from Tableau, a pop-up menu prompts you to install the Starburst connection file. Click **Install and Restart Tableau**.



Once Tableau restarts, find and select **Starburst Enterprise by Starburst** in the same menu once again.

Tableau Server

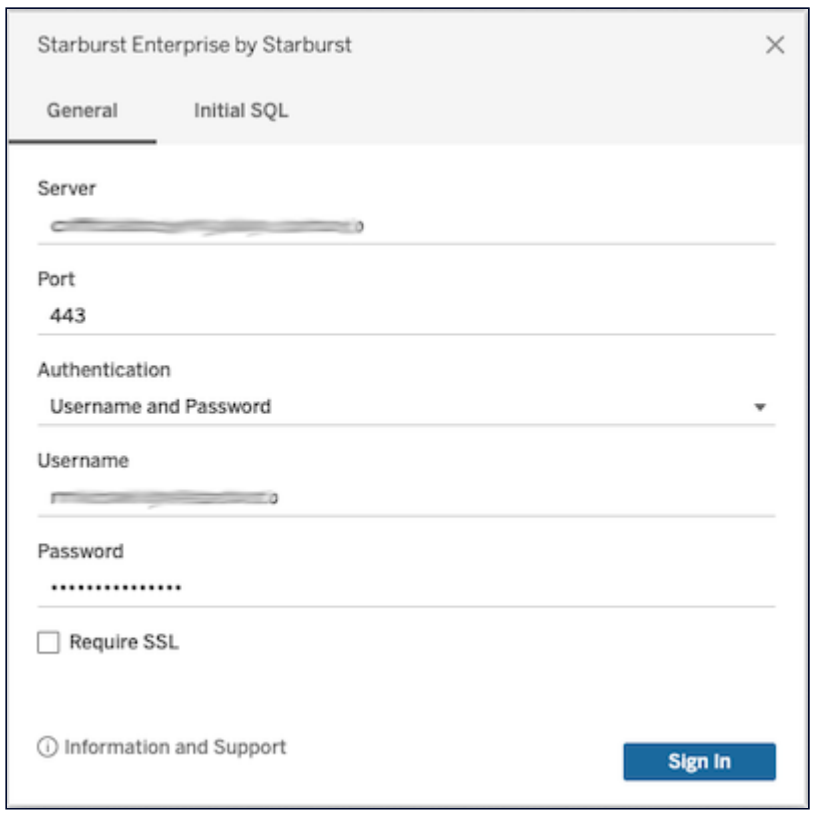
On Tableau Server, you must create a new workbook before you connect Tableau to Starburst Galaxy. Click **New**, then select **Workbook** from the drop-down menu.



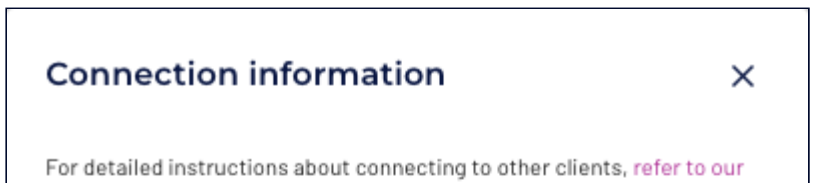
In the pop-up menu labelled **Connect to Data**, click the **Connectors** tab. Then, select **Starburst Enterprise by Starburst**.

Connection dialog

Now, you need to configure your connection to Starburst Galaxy in the Tableau connection dialog.



Starburst Galaxy makes this easy. Find your way back to Galaxy. Then, select **Clusters** from the left-hand navigation menu. In the clusters list, click the **Connection info** button for the `sample` cluster. In the **Select client** drop-down menu, choose **Tableau**.

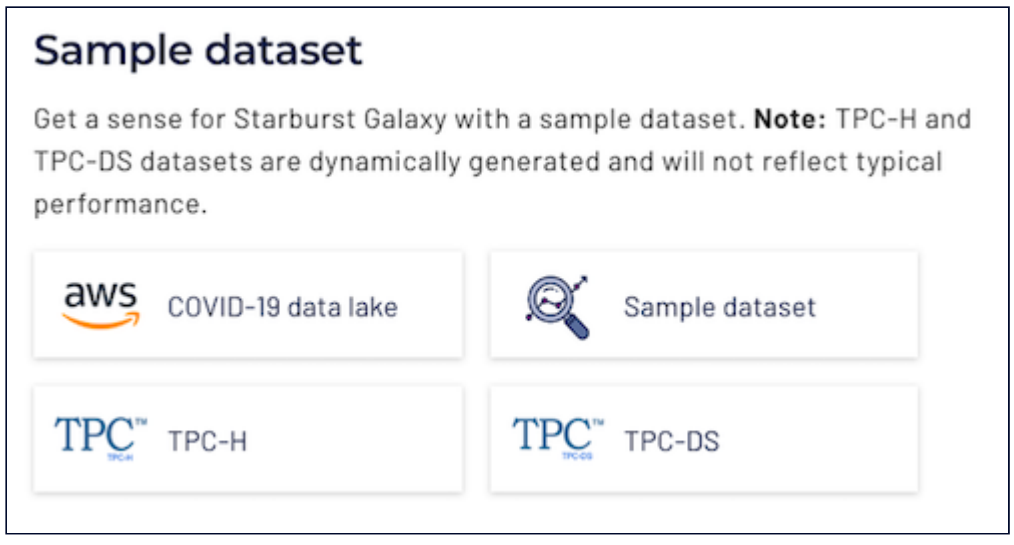


Copy the **User**, **Host**, and **Port** connection strings to your clipboard, and paste them into the **User**, **Host**, and **Port** fields in the Tableau connection dialog. Enter your password for Galaxy, and click **Sign In**.

Configure a new catalog

Before we begin building our Tableau dashboard, let's configure a new catalog in Starburst Galaxy. This will allow us to federate data between two Galaxy catalogs in Tableau. For demonstration purposes, we are going to create a copy of the `tpch` catalog in the `sample` cluster.

From the left-hand navigation menu in Galaxy, click **Catalogs**. In the **View Catalogs** pane, click **Create catalog**. In the **Select a data source** pane, scroll to the bottom of the page and choose **TPC-H**.



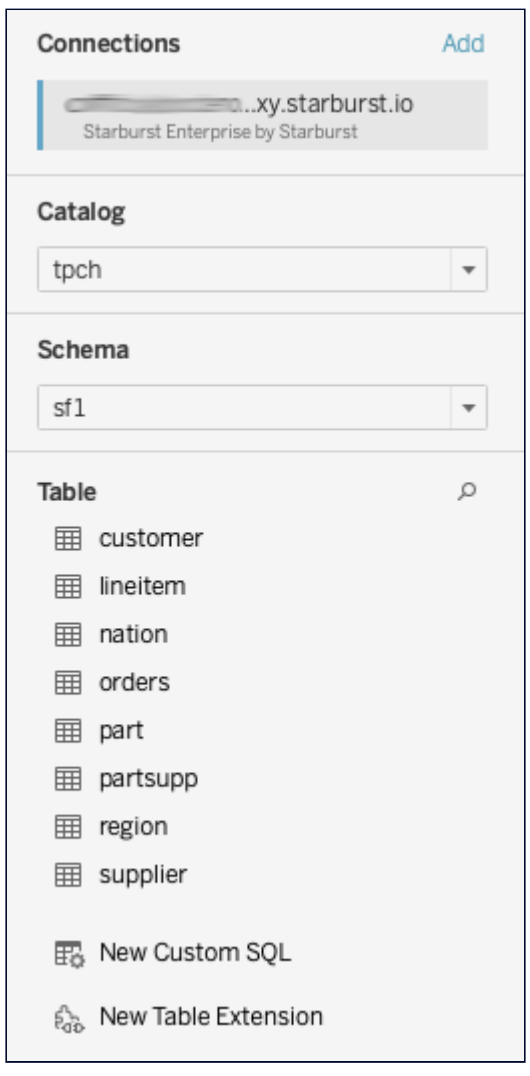
In the **Catalog name** field, name your catalog `tpch_copy`. Then, click **Connect catalog**. In the **Set permissions** pane, click **Save access controls**. In the **Add to cluster** pane, select the **sample** cluster from the drop-down menu. Click **Add to cluster** to finish setting up your catalog.

Create a simple Tableau dashboard

Once you have familiarized yourself with the query editor, established your connection from Tableau to Starburst Galaxy, and configured your new catalog, you are ready to create your first Tableau dashboard.

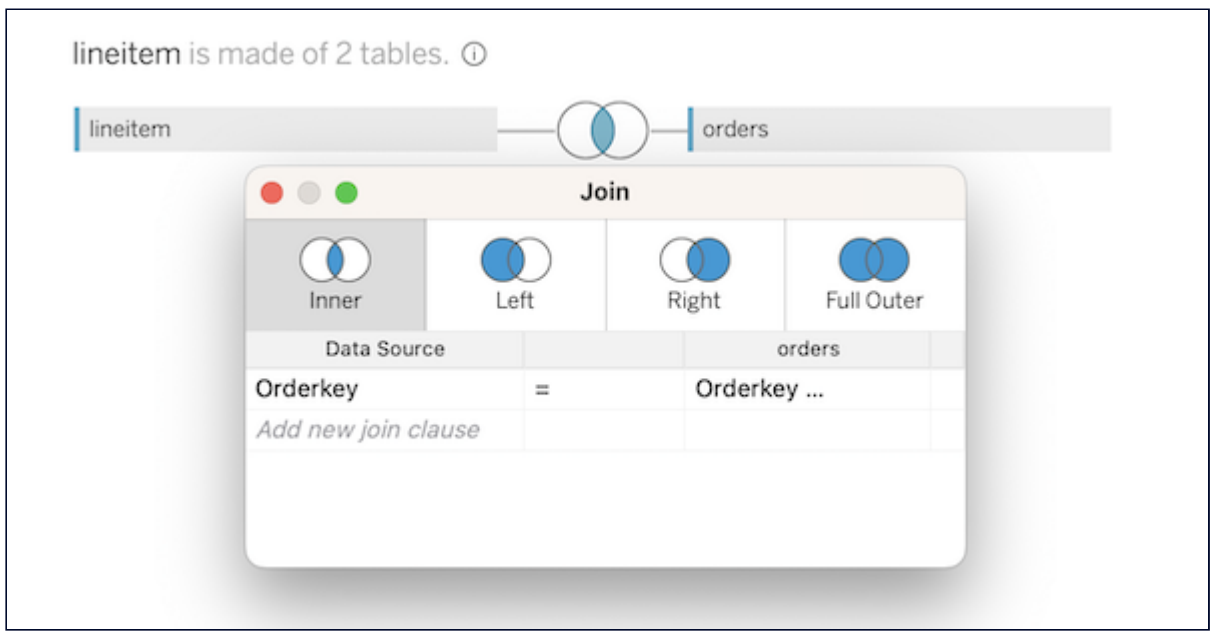
Set up your data sources

In Tableau, find the **Data Source** pane. From the **Catalog** drop-down menu, select `tpch`. From the **Schema** drop-down menu, select `sf1`.



Under **Table**, double click the **lineitem** table. This opens the canvas, where you may add more tables and join them (or relate them).

In the canvas, double-click the **lineitem** table. Now, drag the **orders** table next to it on the canvas. This creates a join between the two tables. To view or edit the join, click the two overlapping circles between the tables.

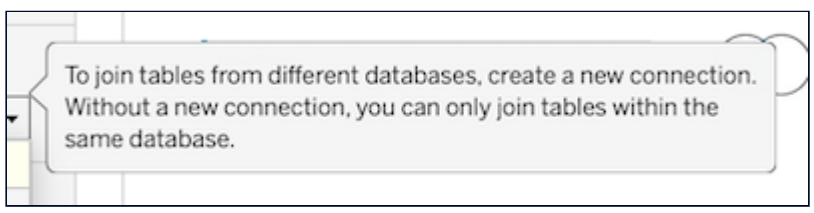


The **Join** window indicates that the **lineitem** table and the **orders** table are joined on the **OrderKey** column as an inner join.

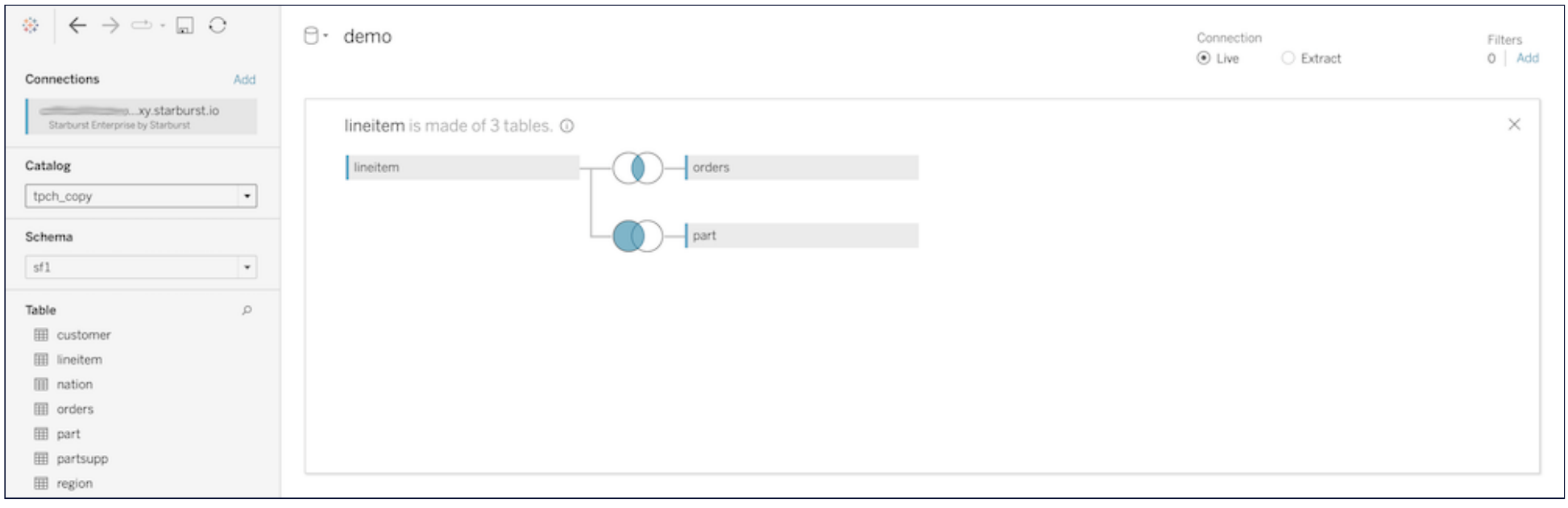
Federate your data sources

Now, let's demonstrate how easy it is to federate your data with Starburst Galaxy. We are going to join the `lineitem` table from the `tpch` catalog with the `part` table from the `tpch_copy` catalog. You can follow these steps to join tables from other platforms as well.

From the **Catalog** drop-down menu, select `tpch_copy`. You may receive the following tooltip message, which you can ignore:



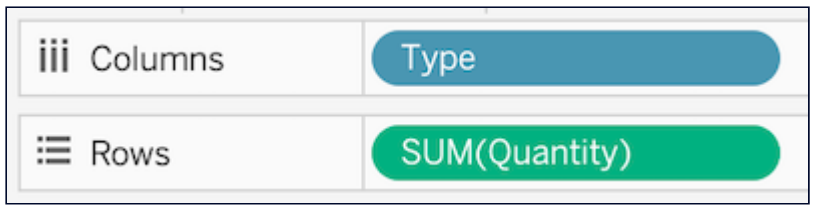
From the **Schema** drop-down menu, select `sf1`. Next, drag the **part** table onto the canvas. Then, open the **Join** window again by clicking the two overlapping circles between the **lineitem** table and the **part** table. Ensure that the tables are joined on the **PartKey** column. Select **Left** to configure the join as a left join.



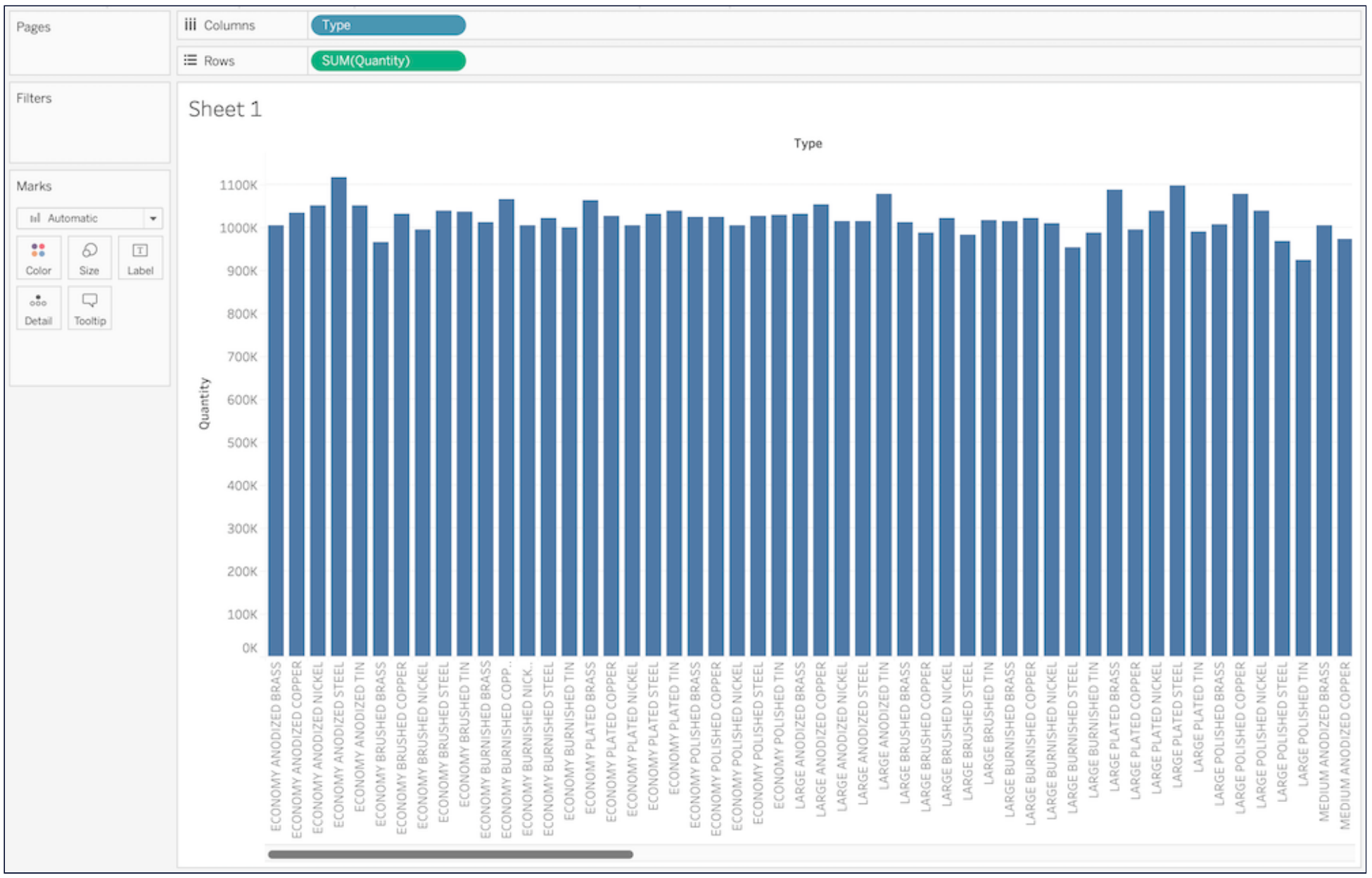
Create simple charts in Tableau

Now that you have configured your data sources, you are ready to create your first visualizations. First, open a new worksheet by clicking the **New Worksheet** button at the bottom of your workbook. The left-hand **Data** pane displays all the tables and columns you can use to build visualizations. The **Data** pane separates your data by placing dimensions (categorical data) on the top and measures (numerical data) on the bottom.

Let's create a simple bar chart to explore the `tpch` data's sales distribution by product type. Drag the **Type** dimension from the **part** table to the **Columns** shelf, and drag the **Quantity** measure from the **lineitem** table to the **Rows** shelf.

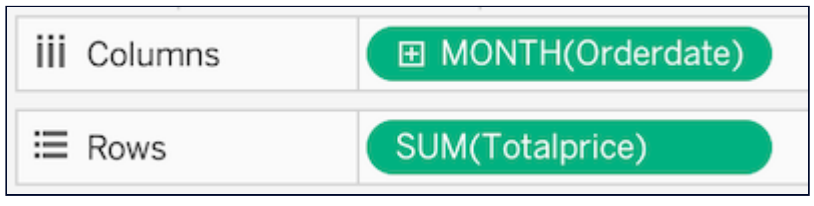


Your bar chart should look like this:

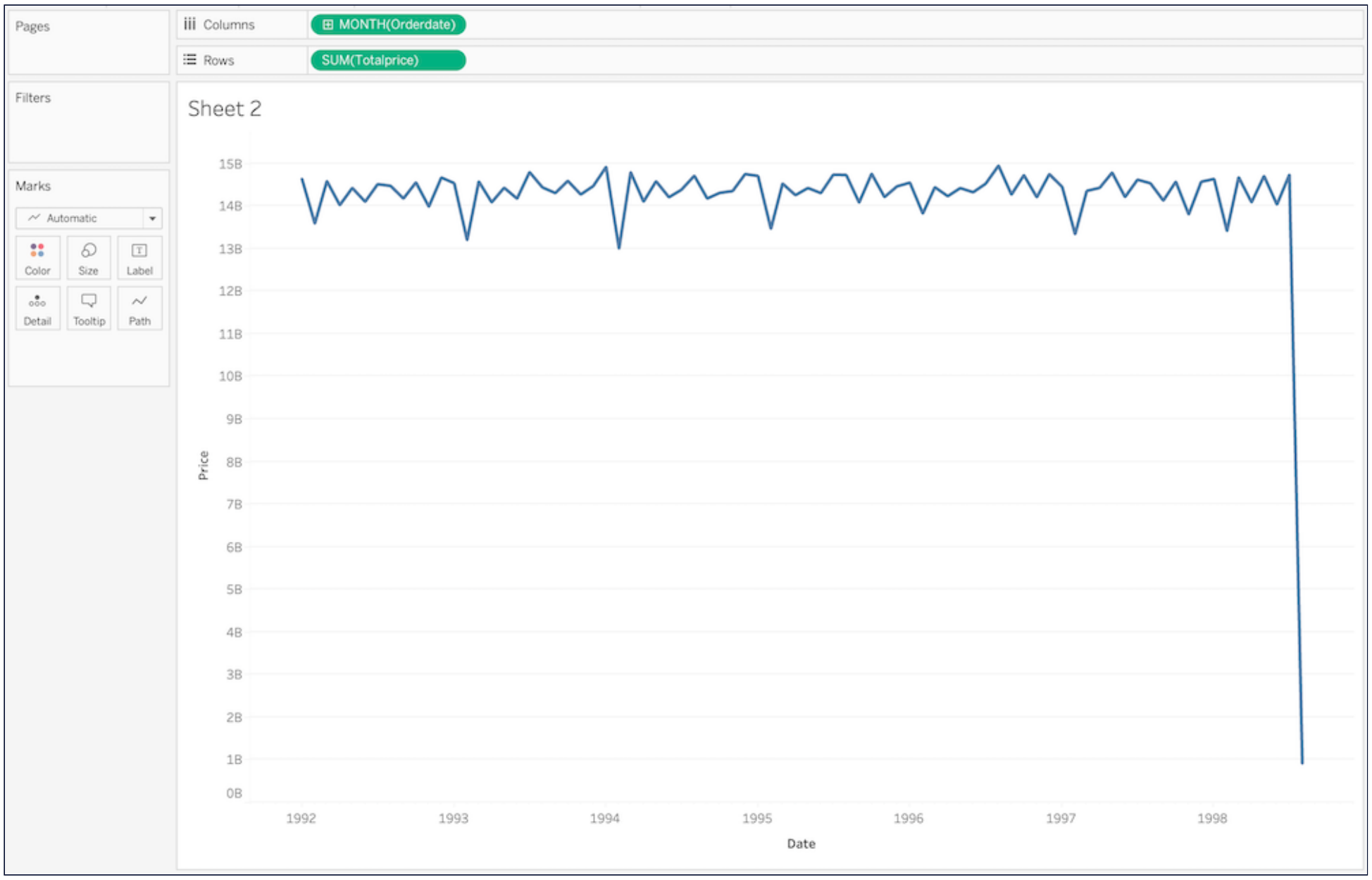


To rename the worksheet, double-click the worksheet name at the bottom of the workbook. Rename the worksheet to **Quantity Sold**.

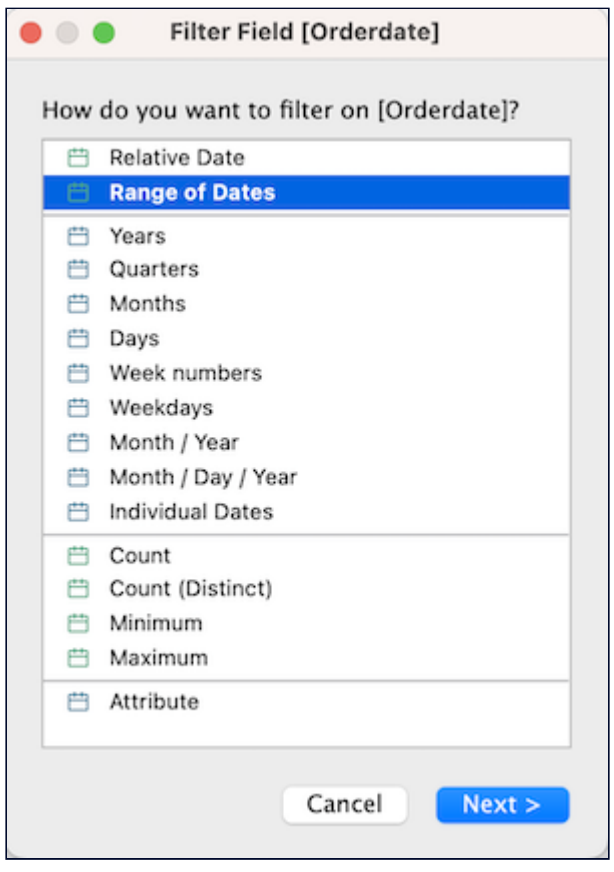
Next, let's create a simple line chart to show the total sales amount over time. Drag the **Orderdate** measure from the **orders** table to the **Columns** shelf, and drag the **Totalprice** measure from the **orders** table to the **Rows** shelf. In the **Columns** shelf, right click **Orderdate** and select **Month** from the drop-down menu.



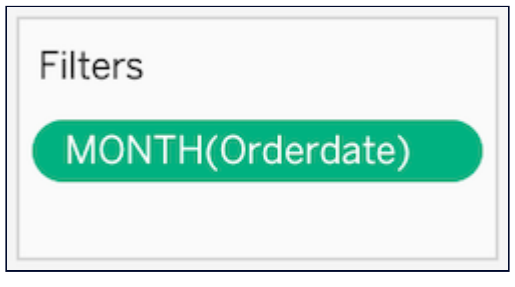
Your line chart should look like this:



Then, add a filter so you can select a specific date range for the chart. Drag the **Orderdate** measure from the **orders** table to **Filters**. In the **Filter Field** dialog box, choose **Range of Dates**, then click **Next >**. In the next dialog box, click **OK** to continue.



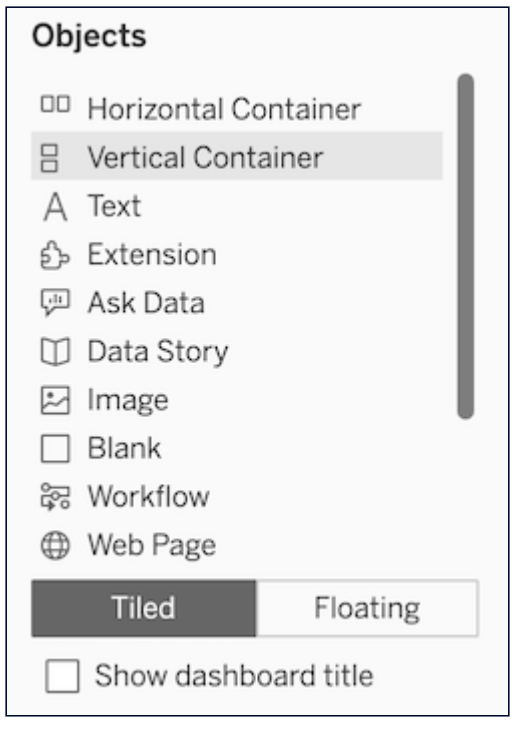
In the **Filters** shelf, right click **Orderdate** and select **Month** from the drop-down menu.



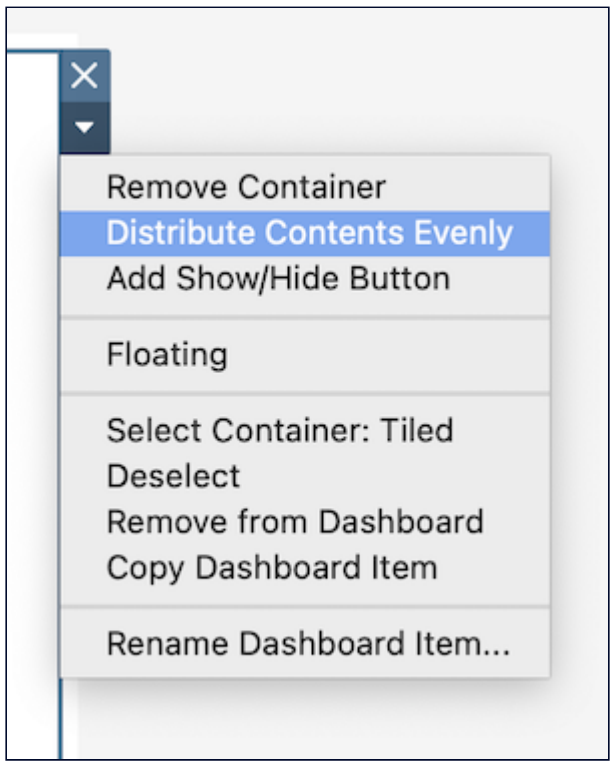
Rename the worksheet to **Total Sales**. To rename axis titles, double-click the axis title to open the **Edit Axis** dialog box. Enter a new title in the **Axis Title** section. Rename the Y axis to **Price**, then rename the X axis to **Date**.

Create your dashboard

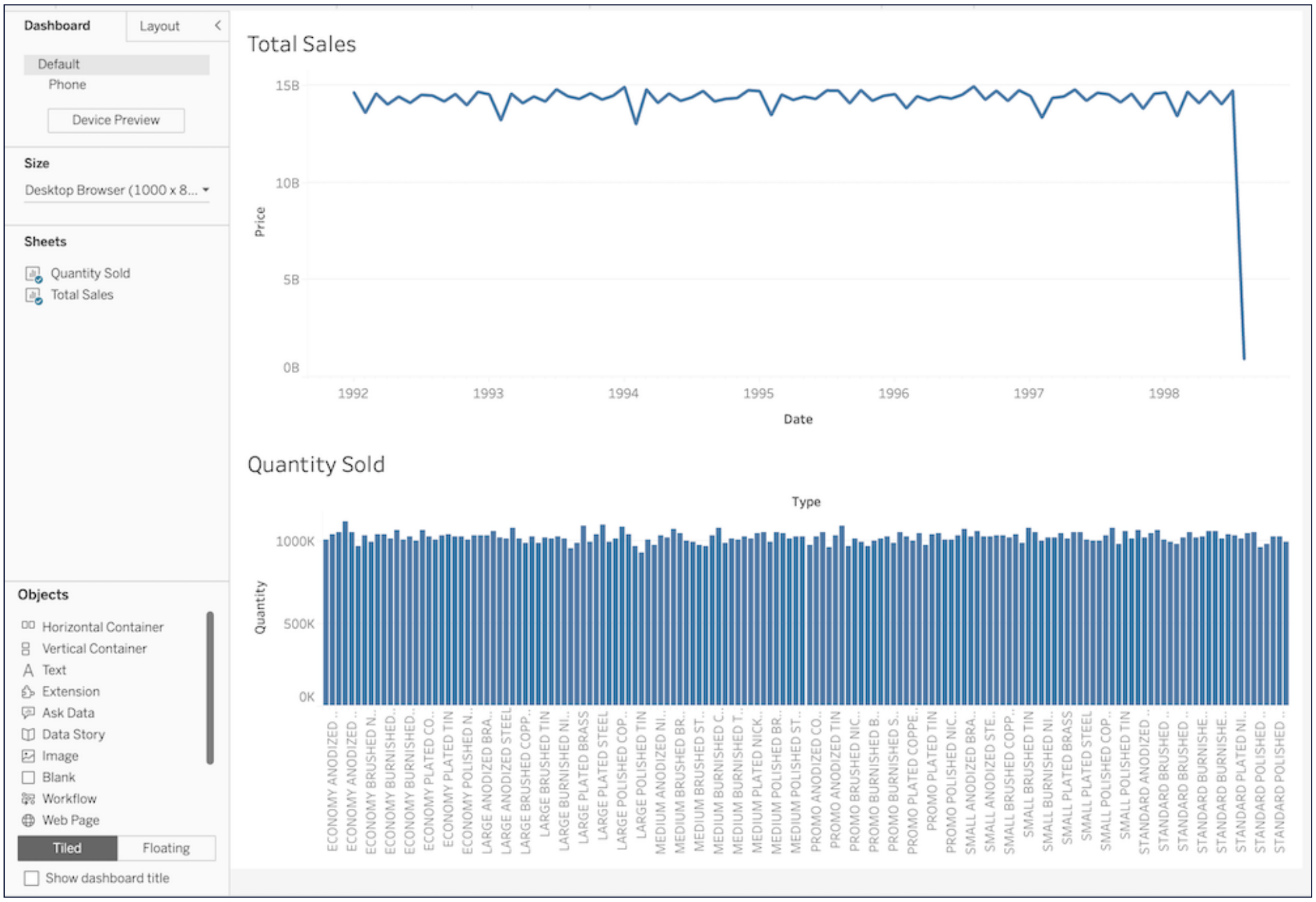
Now, you are ready to create your dashboard. Open a blank dashboard by clicking the **New Dashboard** button at the bottom of your workbook. Under the **Objects** list in the **Dashboard** pane, select **Vertical Container**. Drag the container onto the dashboard.



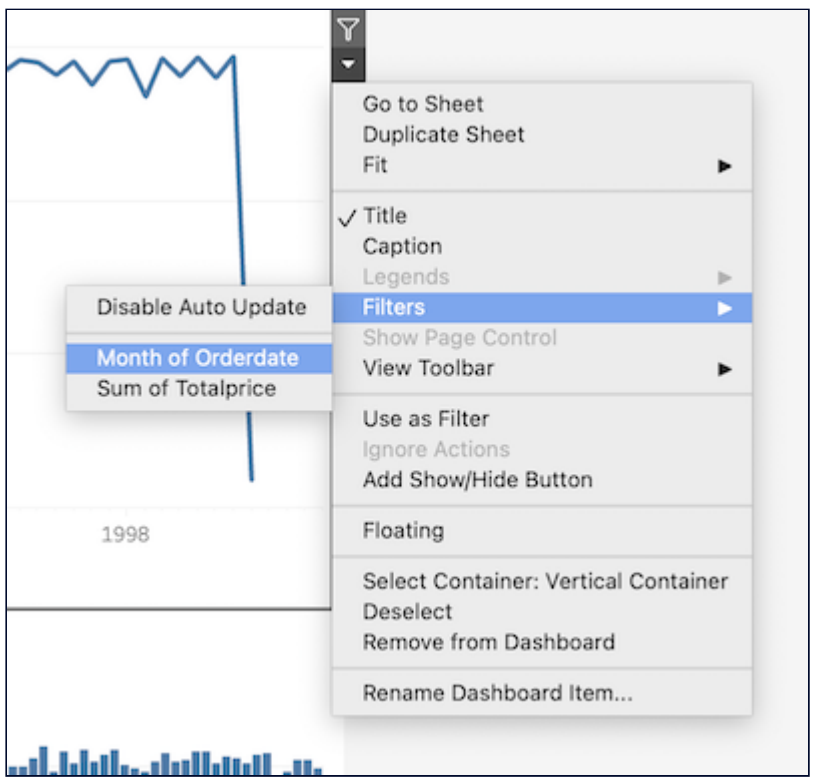
Select the empty layout container. From the container's shortcut drop-down menu, choose **Distribute Contents Evenly**.



Next, under the **Sheets** list in the **Dashboard** pane, select **Total Sales**. Drag the worksheet onto the dashboard. Then, drag **Quantity Sold** onto the dashboard below **Total Sales**.

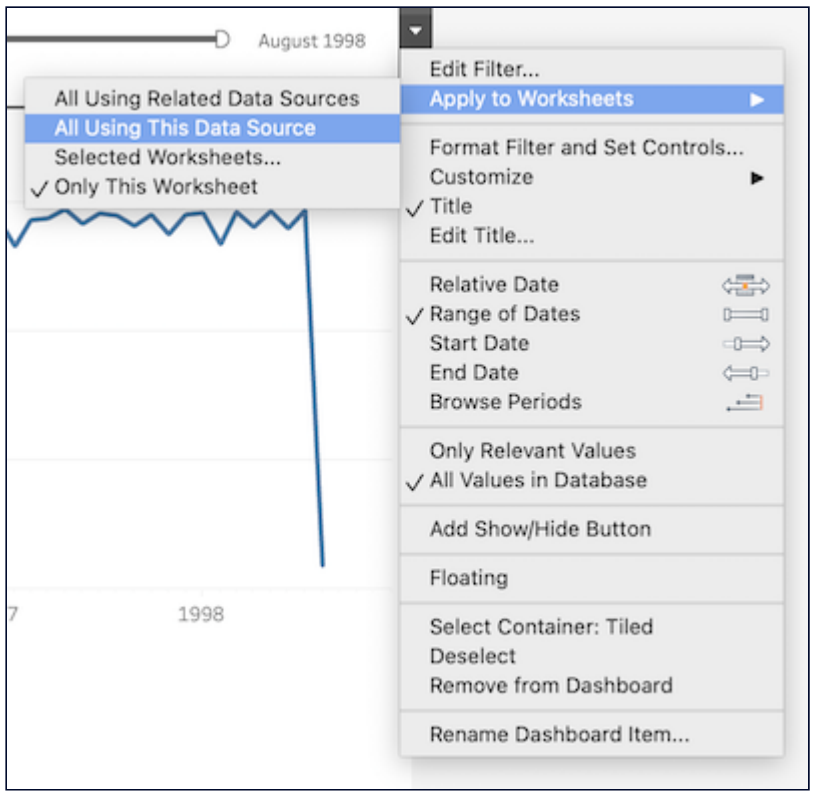


Now, select the **Total Price** sheet. From its shortcut drop-down menu, choose **Filters > Month of Orderdate**.



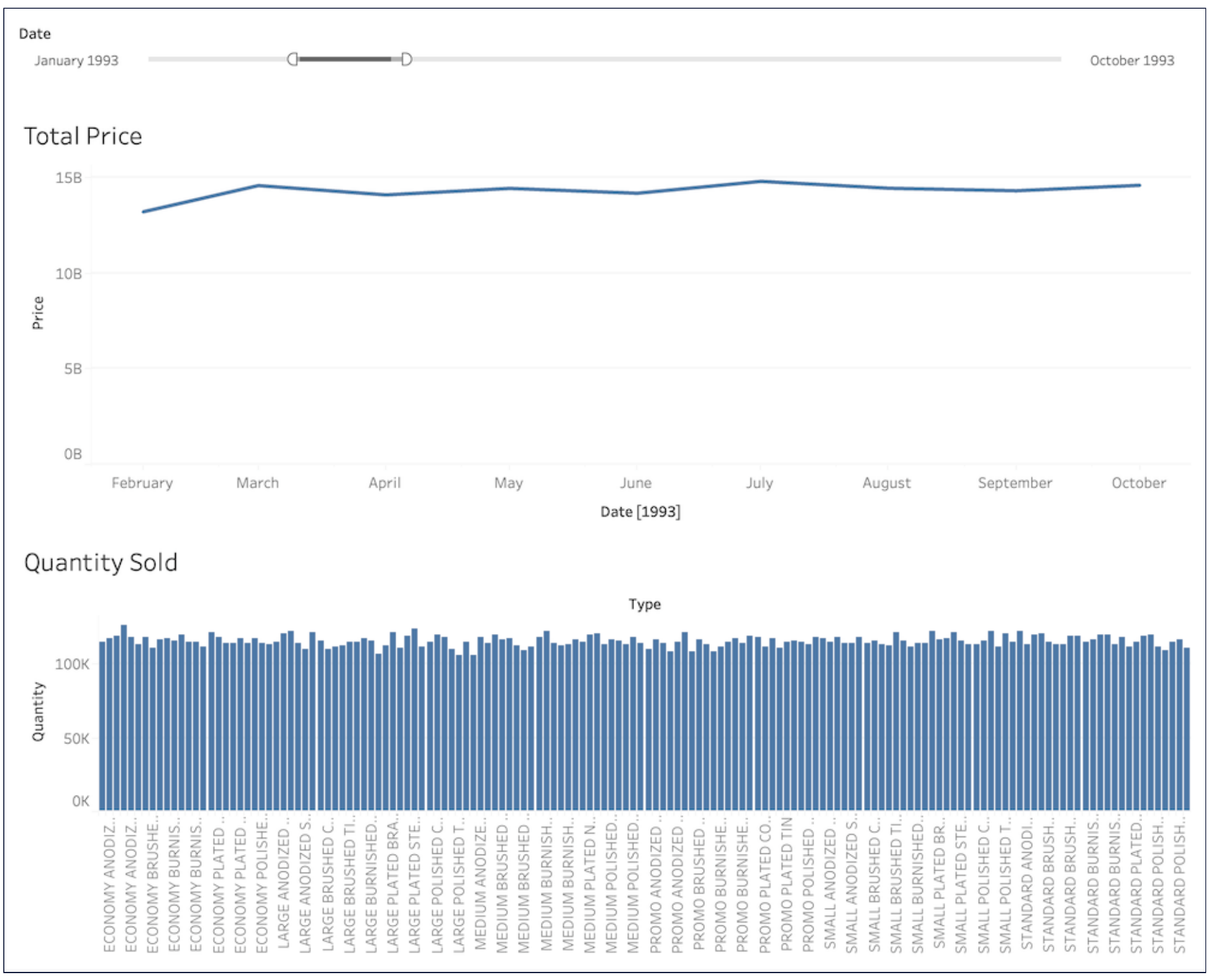
Drag the filter above the **Total Price** sheet. To rename the filter, right click the filter and choose **Edit Title**. Rename it **Date**.

As a final step, from the filter's shortcut-dropdown menu, choose **Apply to Worksheets > All Using This Data Source**.



Wow! Now you can view sales and product data for any date range by using the **Date** filter at the top.

Your dashboard should look like this:



Congrats, you just created your first Galaxy-powered Tableau dashboard!