
# SECP2753 – DATA MINING

# SEMESTER 2 2021/2022

---

## GROUP 2

## ASSIGNMENT 2

## Members:

1. **GROUP LEADER**: NAYLI NABIHAH JASNI (A20EC0105)
2. MIKHEL ADAM BIN MUHAMMAD EZRIN (A20EC0237)
3. MUHAMMAD DINIE HAZIM BIN AZALI (A20EC0084)
4. MADINA SURAYA BINTI ZHARIN (A20EC0203)

# TABLE OF CONTENTS

# USER MANUAL

# 1.0 INTRODUCTION TO TABLEAU

## 1.1 About Tableau

Tableau is a visual analytics platform that is revolutionizing the way we use data to solve problems by enabling individuals and organizations to get the most out of their data. It is also one of the most powerful and fastest growing data visualization tools available in the Business Intelligence Industry.

Tableau began in 2003 as a computer science project at Stanford with the goal of improving the flow of analysis and making data more accessible to people through visualization. Tableau's foundational technology, VizQL, was developed and patented by co-founders Chris Stolte, Pat Hanrahan, and Christian Chabot. VizQL visually expresses data by translating drag-and-drop actions into data queries through an intuitive interface.

## 1.2 Tableau Main Products

Tableau offers several softwares in their product suite related to visual analytics that vary in functionality for different use cases. The main products offered are as below:

### 1.2.1 Tableau Prep

Tableau Prep is a tool in the Tableau product suite designed to make preparing your data easy and intuitive. Use Tableau Prep Builder to combine, shape, and clean your data for analysis in Tableau.

### 1.2.2 Tableau Desktop

In Tableau Desktop, you can connect to data and create workbooks, stories, and dashboards. You can publish workbooks and data sources to your Tableau site to share insights with others. You can view, interact with, and edit those published views in your Tableau site.

### 1.2.3 Tableau Server & Tableau Online

Tableau Online is the sharing tool which is used as high-performance data repositories whereas Tableau Server is a sharing tool which is used to share and create visualizations for organizations with full security.

### 1.2.4 Tableau Mobile

Tableau Mobile is a mobile companion app for Tableau Online and Tableau Server that allows you to access your Tableau site from anywhere. Tableau Mobile, which is available for Android and iOS, allows you to interact with your site's content and discover data insights even when you're offline.

# 2.0 USER GUIDE

## 2.1 Installation of Tableau

### 2.1.1 Public

Tableau public is the free version of the software for users without a paid subscription. This is a good option for a total beginner who wants to experiment with most of Tableau's features without having to pay the annual fee. Below are the step to install Tableau public:

1. Open a Google Search and type “Tableau Public” in the search bar.

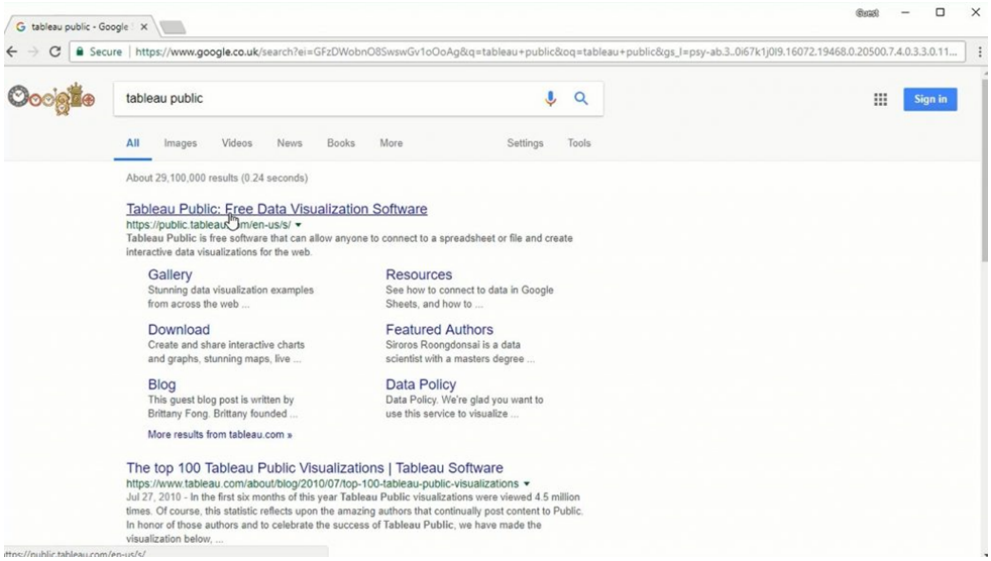



Figure 1: Search “Tableau Public” in Google Search

2. The first result that you can see will point to Tableau’s website at www.tableau.com. Simply click on the link and it will direct you to the Tableau Public domain.

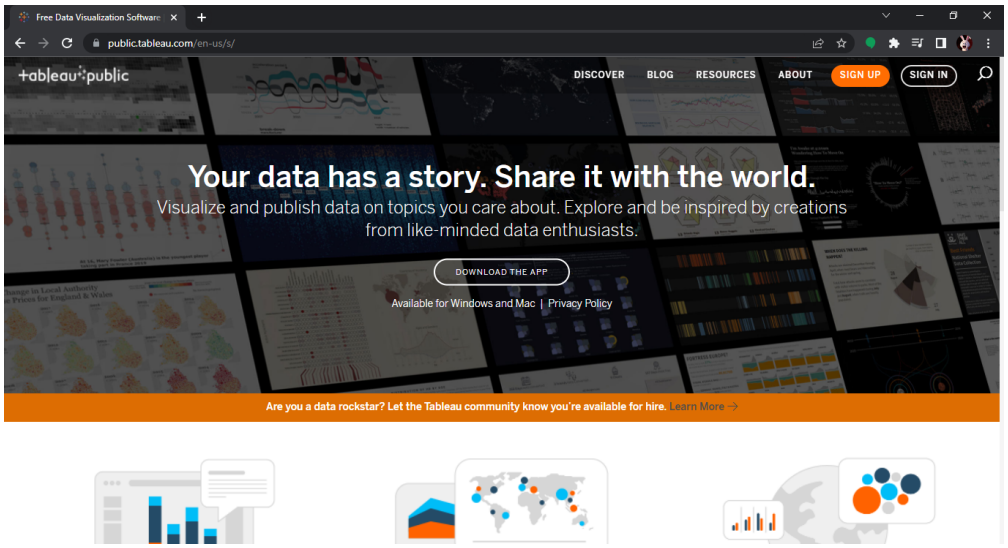



Figure 2: Display of the Tableau’s website

3. At the main page, you will see a bar that will prompt you to download the app. All you have to do now is click on it and it will bring you to another page. Then, also you will see a bar that will prompt you to download the app. Just click on it and a form will show up.

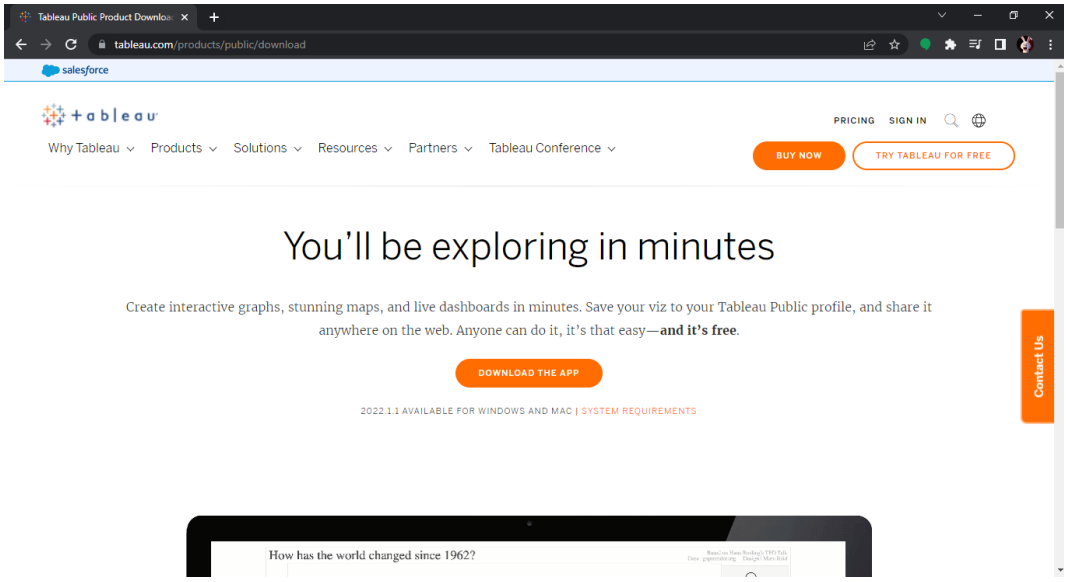



Figure 3: Display of the download page

4. A form that you will see after clicking on "Download The App".

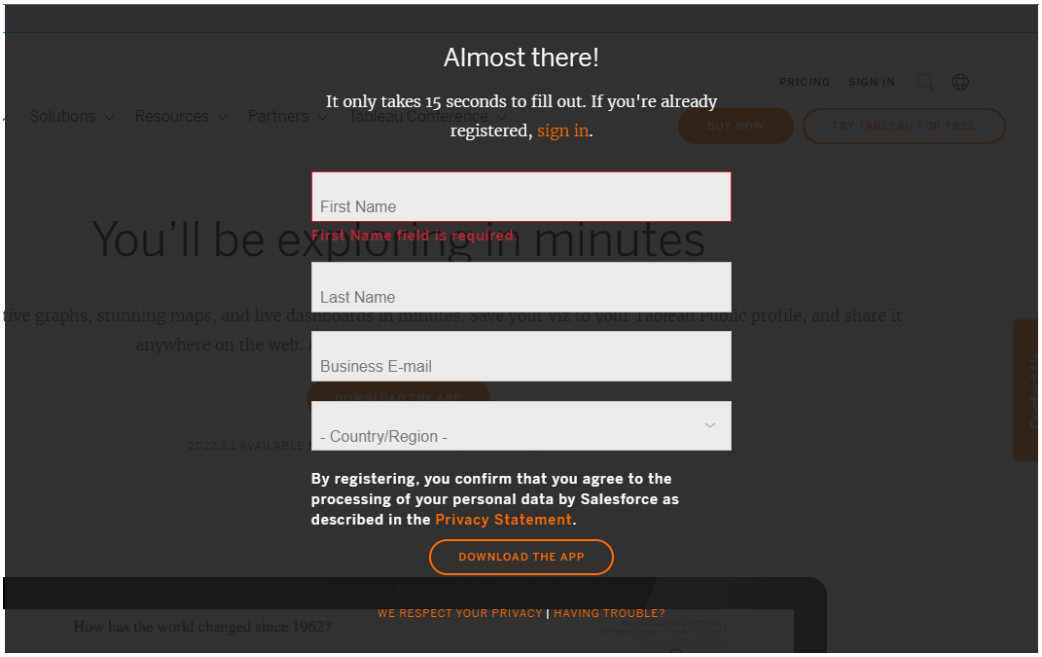



Figure 4: The form in order to download the app

5. Fill in the form and click on the "Download The App" button. The '.exe' file will automatically download on your computer.


Figure 5: .exe file will be downloaded

6. Open the '.exe' file to begin your installation. However, you need to agree to Tableau's terms and conditions before starting installing.

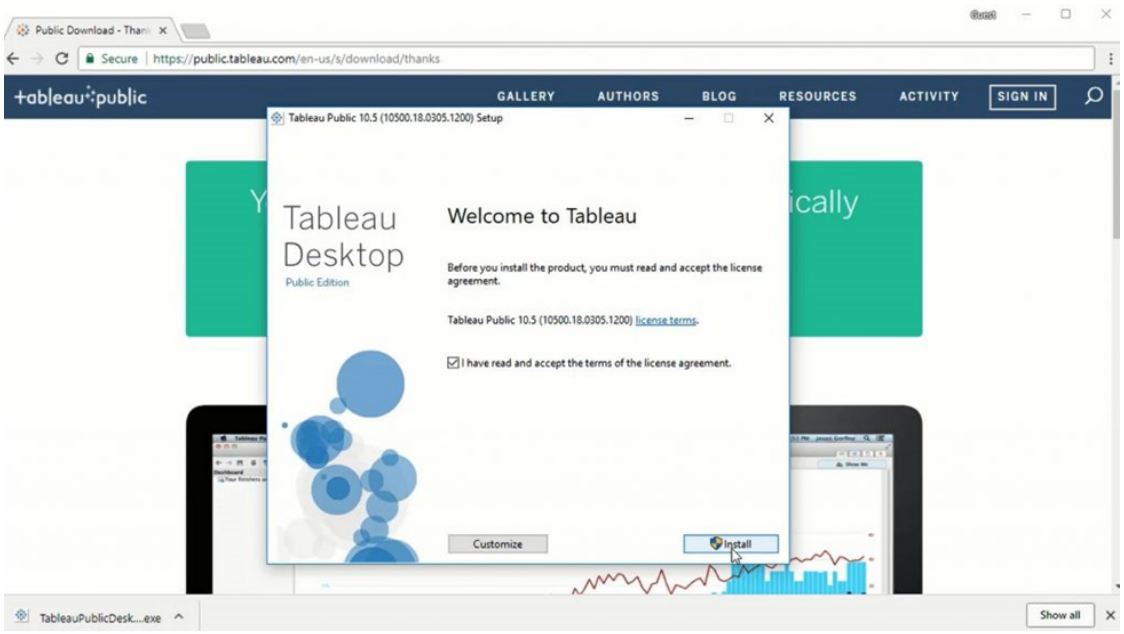



Figure 6: Installation of Tableau

7. Once the process starts, all you have to do is wait for the program to start running on your machine.

### 2.1.2 Desktop

Tableau Desktop is meant for someone with a career in Business Intelligence and Data warehousing such as Analysts and BI Professionals. This software is not free because you need to pay monthly for $70. Below are the tutorial on how to install Tableau Desktop:

1. Open the https://www.tableau.com/support/releases and select Tableau Desktop or Tableau Prep Builder from the list.

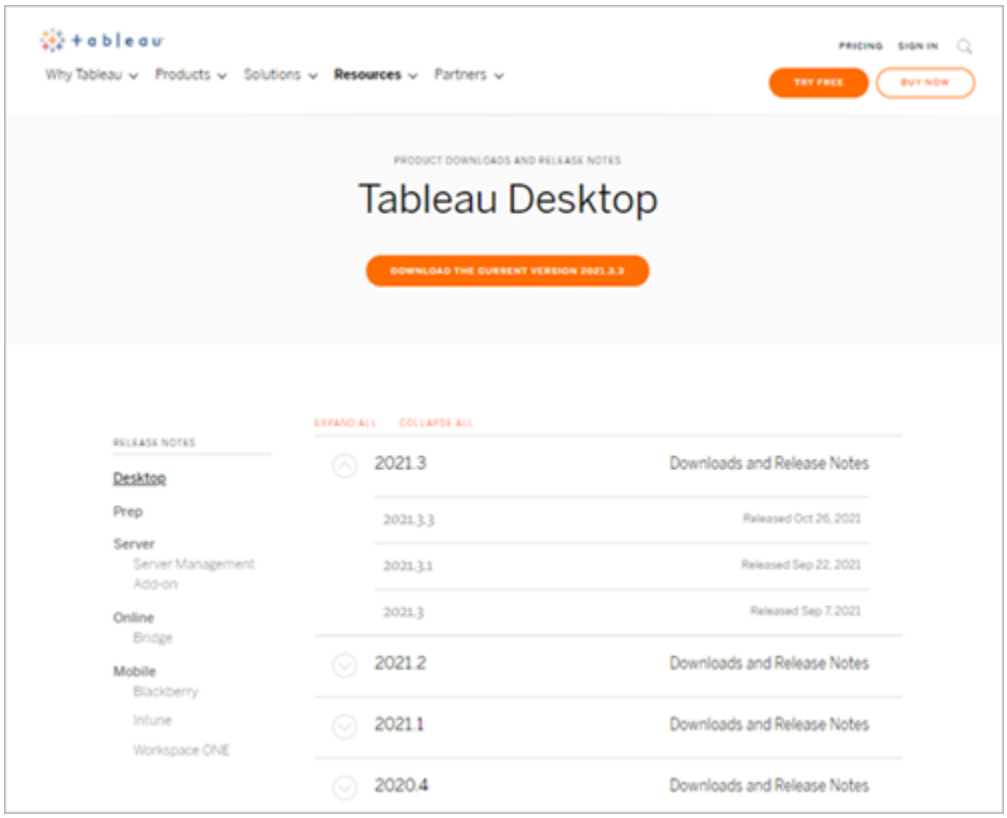



Figure 7: Display of the website

2. Choose the version you want to install and click on it. In the Download Files section of the page, click on the installer for your operating system to download it.

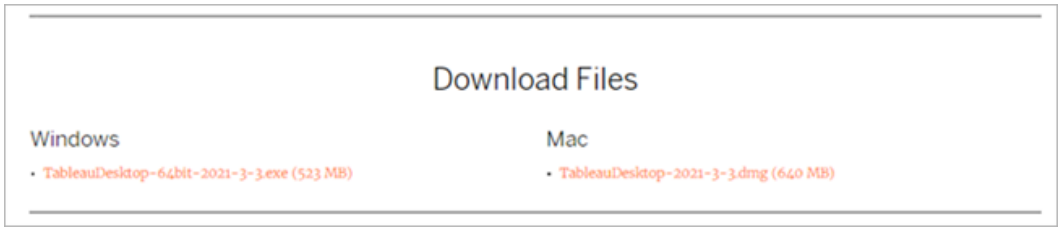



Figure 8: Types of Tableau's version

3. For first time user:

a. Sign in to the [http://customer.tableau.com/](http://customer.tableau.com/) on the Tableau website using your email address and password. The information you need to sign into the portal for the first time is sent to you in a welcome email from Tableau after you purchase the product.

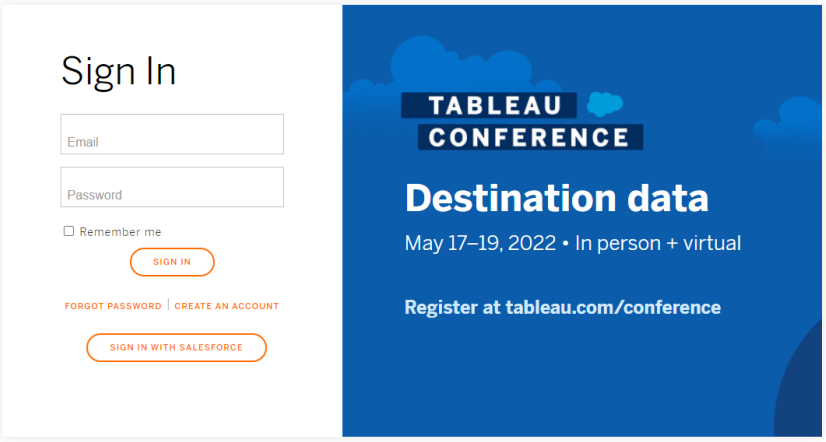



Figure 9: Sign in to Tableau

b. At the top of the Welcome to the Customer Portal page, click Licenses.

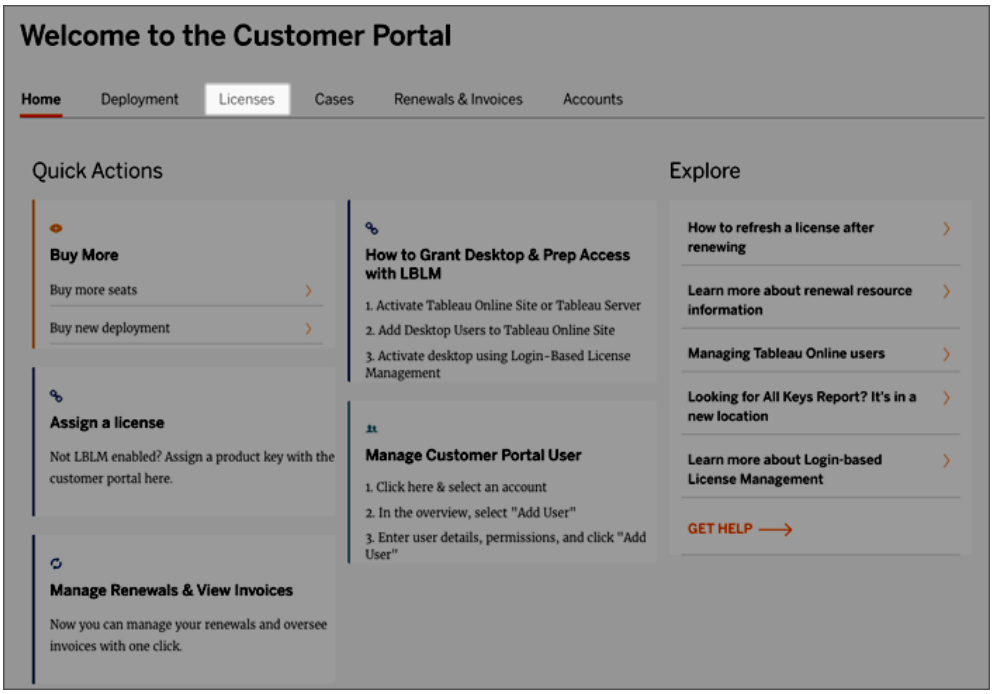



Figure 10: Display after successfully sign in

c. A table opens, listing any product keys assigned to you. Copy your product key from the Key Name column in the table. Keep this key ready and secure.

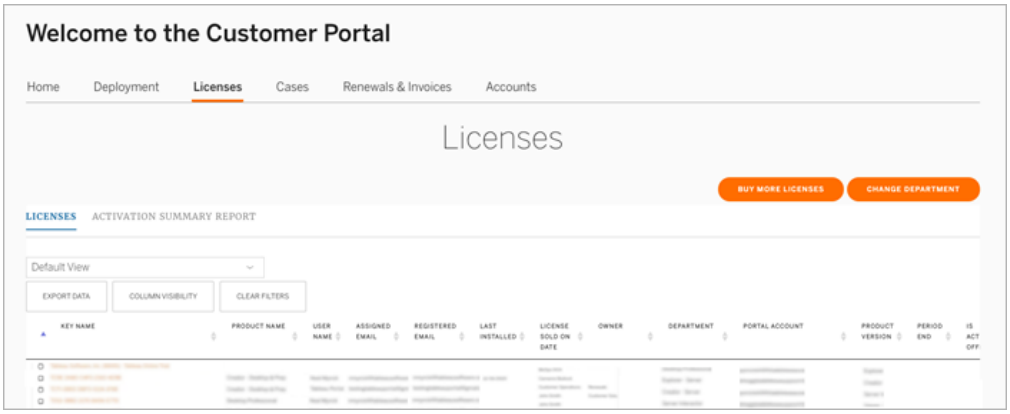



Figure 11: Insert your product key for first time user

4. For existing user:
    a. Open your existing version of Tableau Desktop or Tableau Prep Builder.
    b. From the top menu, click Help > Manage Product Keys.

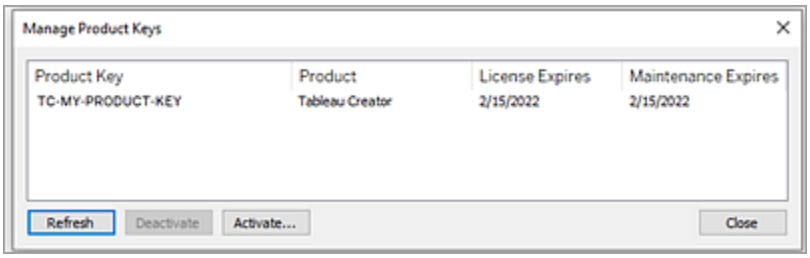



Figure 12: Insert your product key for existing user

    c. Copy your product key from the Product Key column in the table. Keep this key ready and secure.

### 2.1.3 Student

You can apply for Tableau's visualization software through its Students Program. To be eligible for a free license through the Tableau for Students programme, you must be actively enrolled in a K-12 authorized academic institution. Students from all over the world can apply and be

prepared to verify their claims by providing documents to SheerID, a third-party verification service. Here, I will explain on how to install Tableau for students:

1. Go to the Tableau website and fill the form https://www.tableau.com/academic/students.

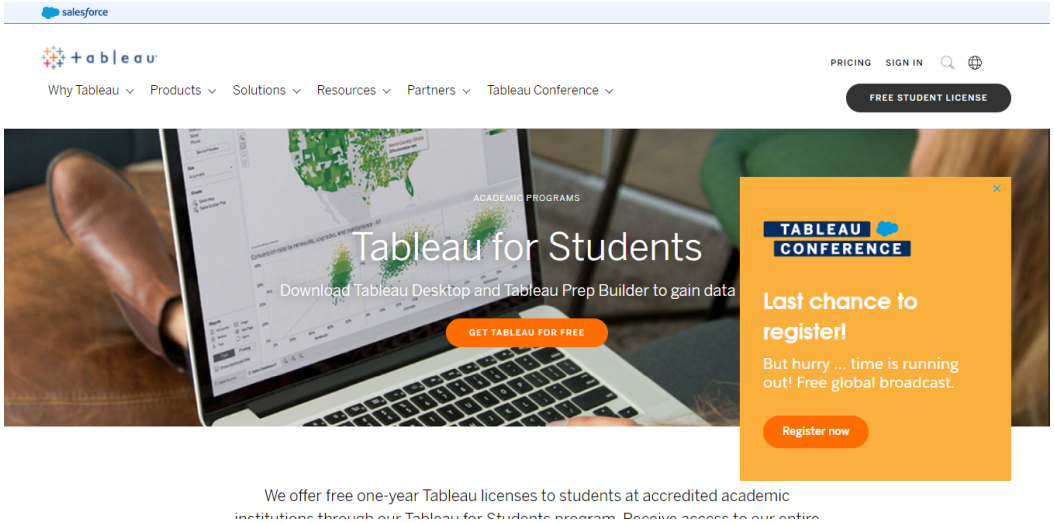



Figure 13: Display of the website

2. Click on the "Free Student License" button and enter all the information required. Then, click on Verify Student Status.

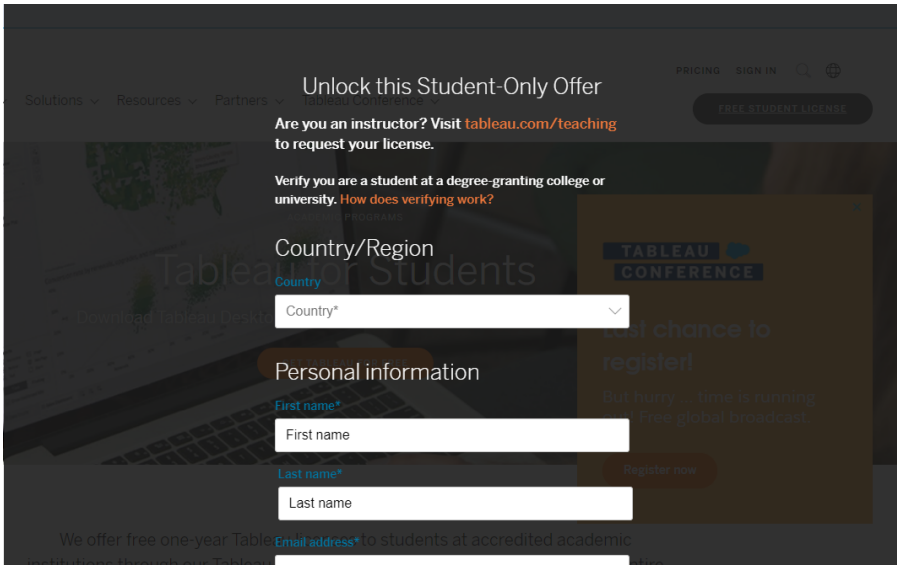



Figure 14: Form to get a free student license

3. You need to upload documentation as a verification such as ID card, transcript or other school issued document. Then, click Submit For Review.

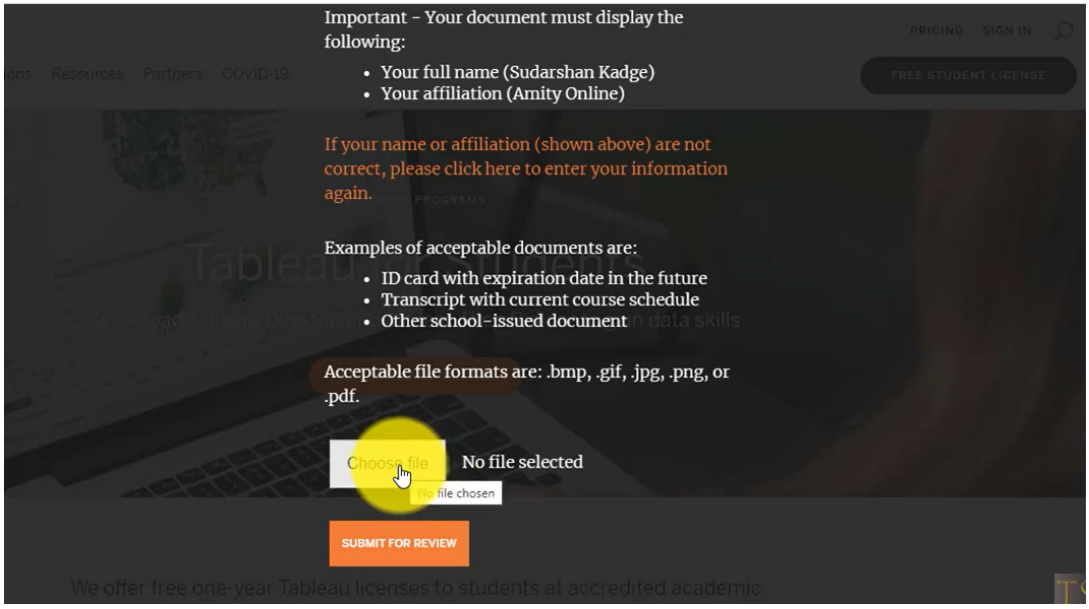



Figure 15: Click submit after fill in the form

4. The information you submitted will be verified by a third party service called SheerID. Once SheerID confirms that your document is authentic, Tableau will approve your request for a student license valid for 1 year.
5. Then, the Tableau team will get in touch with you by email and send you the product key or license for activating the software.
6. You will also get access to download the Tableau Academic suite which contains Tableau Desktop, Tableau Prep, and Tableau Public.
7. You may also download the latest version of the software when you visit Tableau Desktop here. When you click that link, an automatic download of the software will display on your screen. So, Save the file on your computer download folder.
8. After download, locate the Tableau file on your computer download folder and double-click on it to run and install it on your desktop.
9. Enter the product license to activate the software and begin to use the software for 1 year.

## 2.2 Data Source

A Tableau data source is the link between your source data and Tableau. It's simply the sum of your data (as a live connection or an extract), the connection details, the names of tables or

sheets containing data, and the customizations you apply to data in Tableau. Customizations include how data is merged as well as metadata like computations, renamed fields, and default formatting.

## 2.2.1 Connect Data Source

Multiple data connections to different databases or files might be found in a Tableau data source. Connection information comprises details on how to connect to your data, such as database server name and server sign-in information, as well as where the data is located, such as a file name and path or a network location.

### 2.2.1.1 Server

Tableau allows us to connect with a different type of server such as Tableau Server, Microsoft SQL Server, SQL Server, MySQL and many more. In order to connect with another server, you need to download a driver for your chosen server type.

#### 2.2.1.1.1 Tableau

1. Click on Tableau Server at the Start Page.

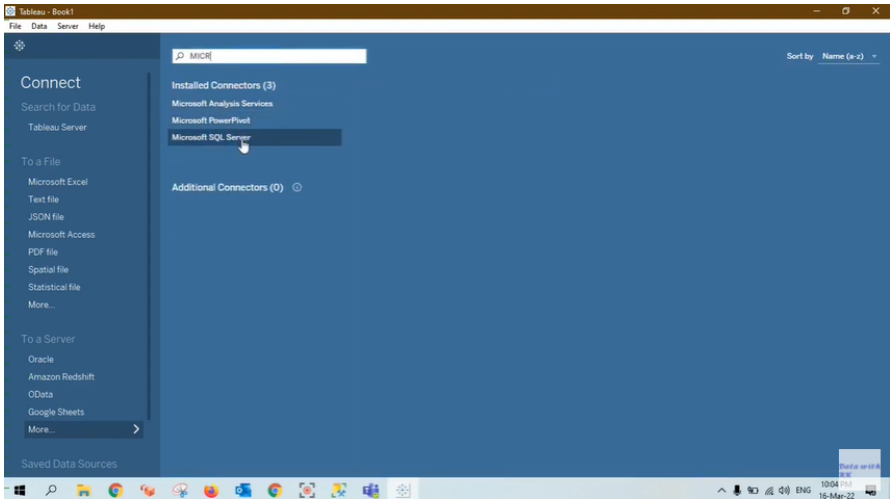



Figure 16: Click Tableau Server

2. Next, it will ask you to fill your server, username and password.


Figure 17: Fill your server, username and password

3. Lastly, you will be connected with your Tableau Server.

#### 2.2.1.1.2 Microsoft SQL

1. First, click on the Microsoft SQL Server at the Start Page.

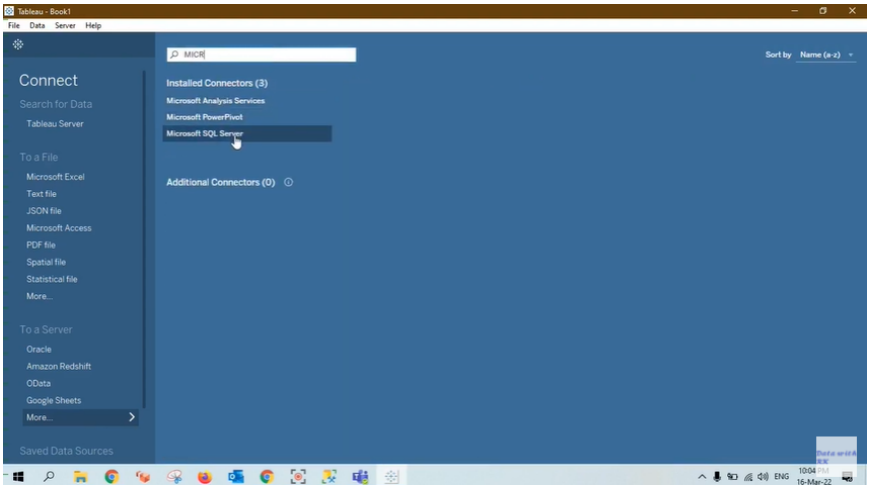



Figure 18: Click on Microsoft SQL

2. Then, a dialog box will show up and ask you to insert your Microsoft SQL server name. After you insert your server name, click on the Sign In button.

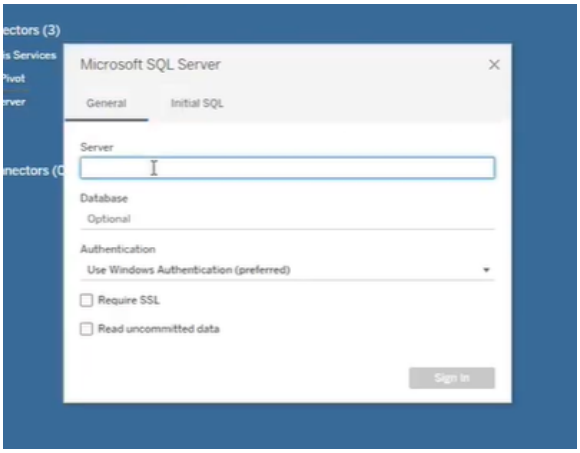



Figure 19: Insert your server name

3. Finally, you will be connected with your Microsoft SQL server.

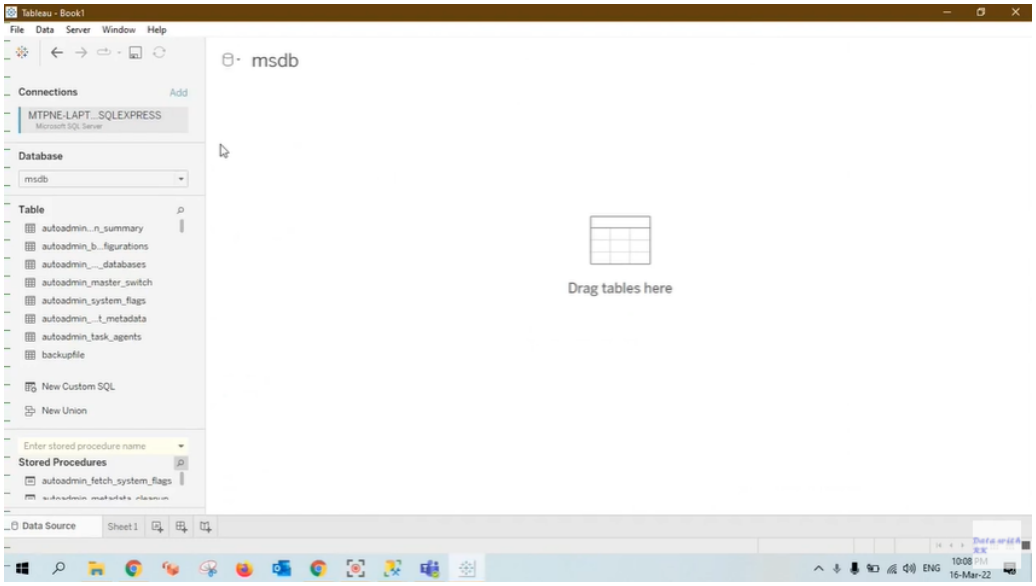



Figure 20: Display after successfully connected to Microsoft SQL server

### 2.2.1.1.3 MySQL

1. Under the Server part at Start Page, click on the MySQL.

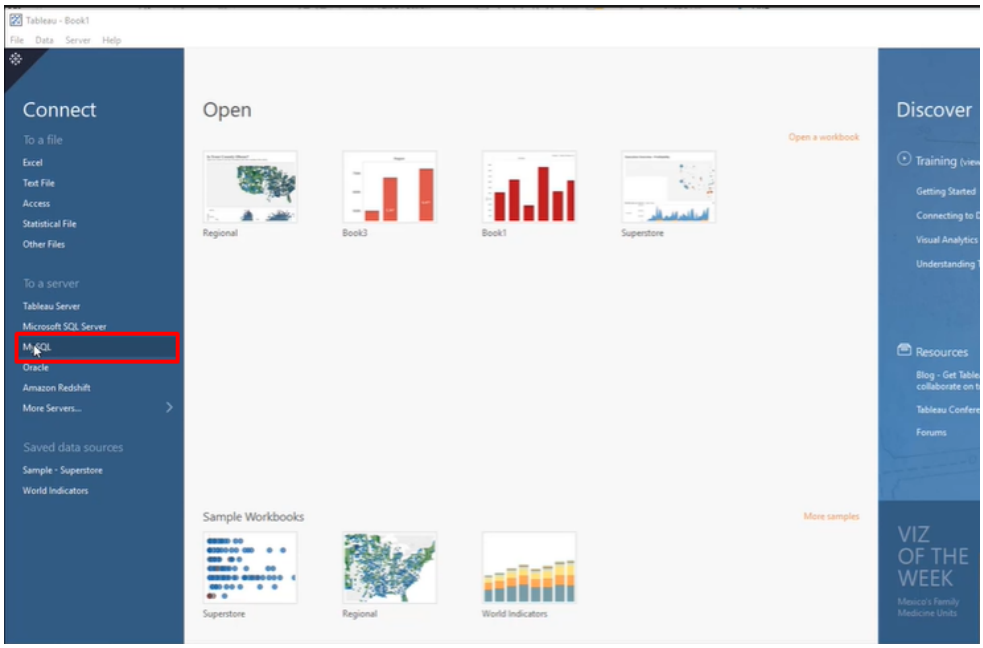



Figure 21: Click on MySQL

2. Next, a dialog box will show up and ask you to fill your MySQL server, username and password.

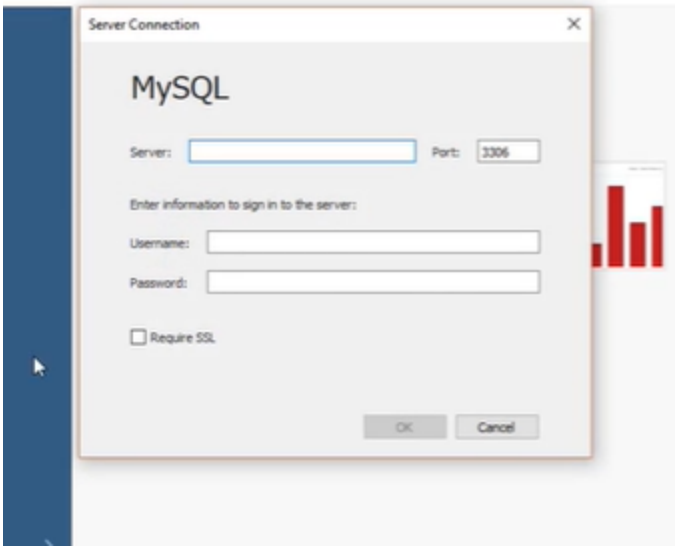



Figure 22: Insert your server, username and password

3. Finally, you will be connected with your MySQL server.

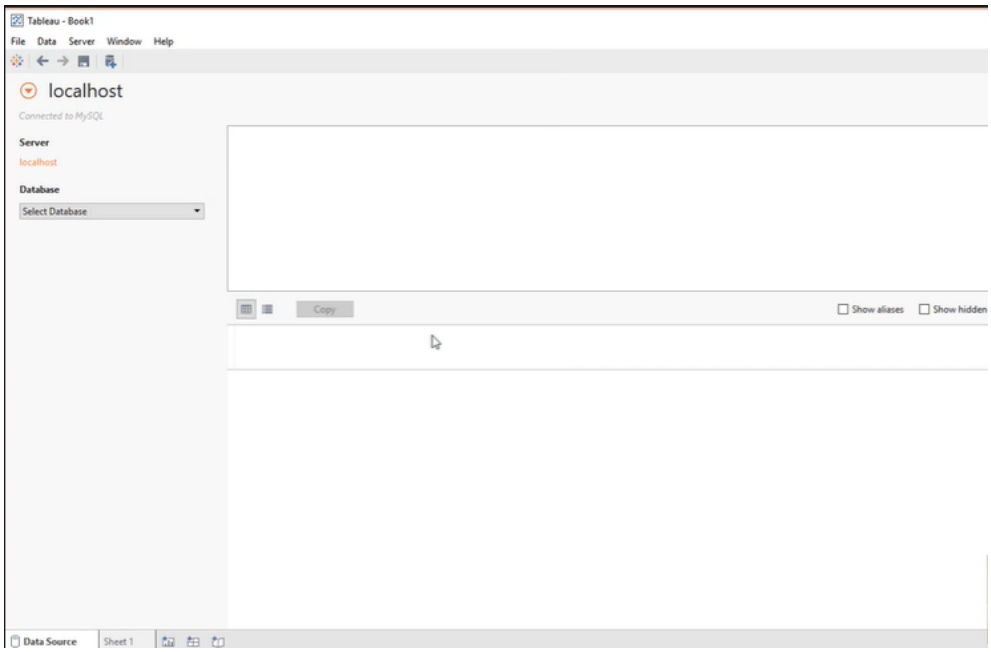



Figure 23: Display after successfully connected to MySQL server

### 2.2.1.1.4 Oracle

1. Just like the other, you need to click on Oracle under the server at the Start Page.

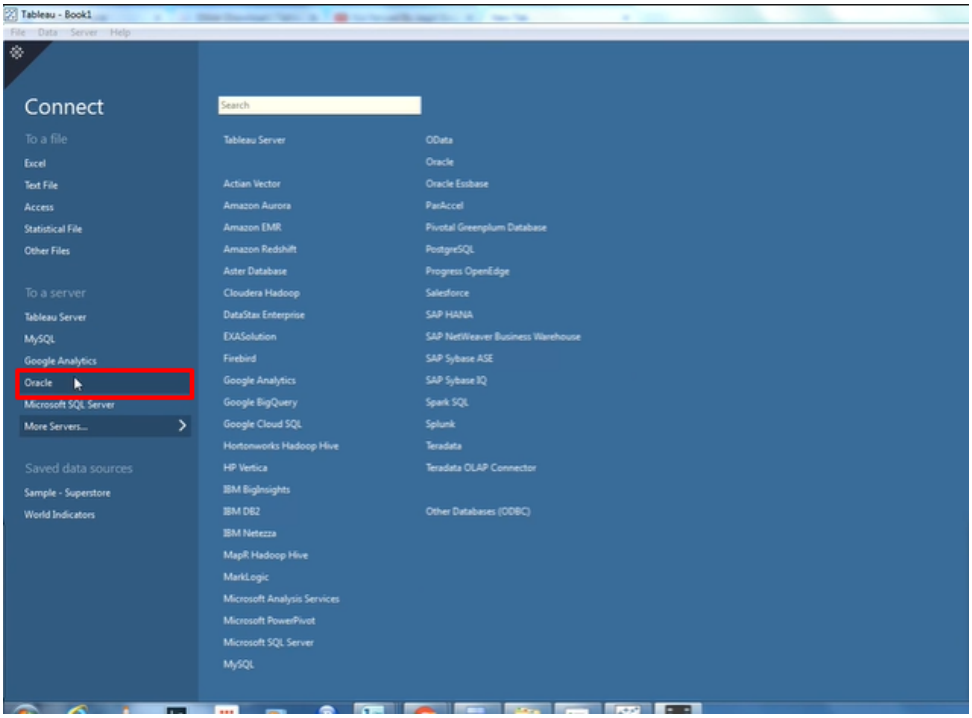



Figure 24: Click on Oracle server

2. It will ask you to fill in your server, service, username and password.

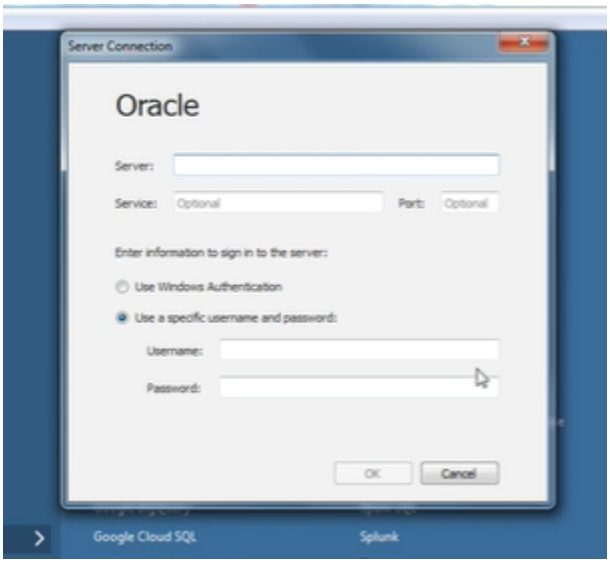



Figure 25: Insert your server, service, username and password

3. After you fill in all the required things, you will be successfully connected with your Oracle server.

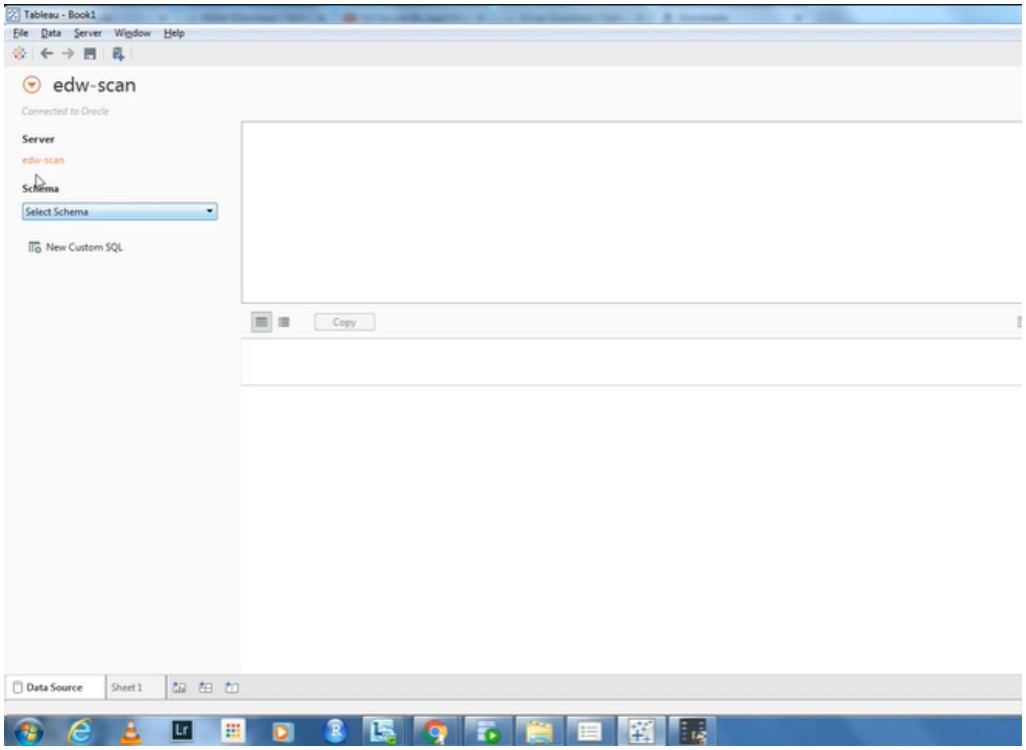



Figure 26: Display after successfully connected to Oracle server

### 2.2.1.1.5 Amazon Redshift

1. First of all, click on the Microsoft Redshift at the Start Page.

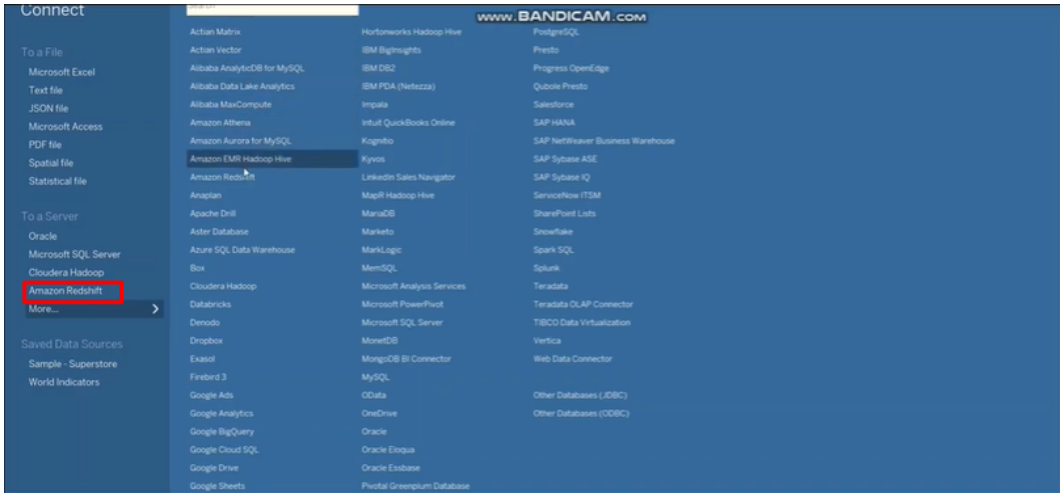



Figure 27: Click on Amazon Redshift server

2. Then, a dialog box will show up and ask you to insert your Microsoft Redshift server, database, username and password.


Figure 28: Insert your Microsoft Redshift server, database, username and password

3. Lastly, you will be connected to your Microsoft Redshift server.

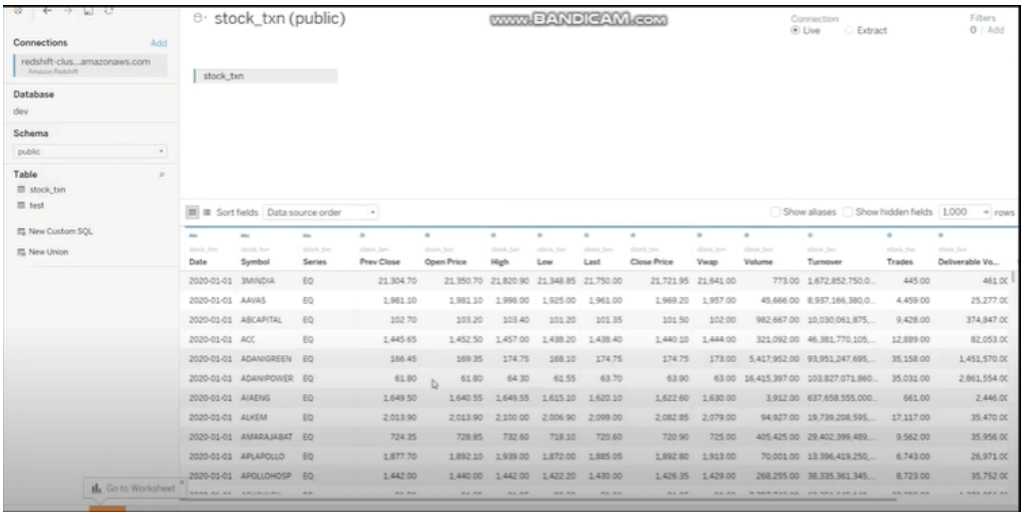



Figure 29: Display after successfully connected to Microsoft Redshift server

### 2.2.1.2 File

In addition, Tableau also allows us to connect with different types of database files. For example, Text File, Spatial File, Excel File etc.

#### 2.2.1.2.1 Excel File

1. First click on Microsoft Excel at the Start Page.

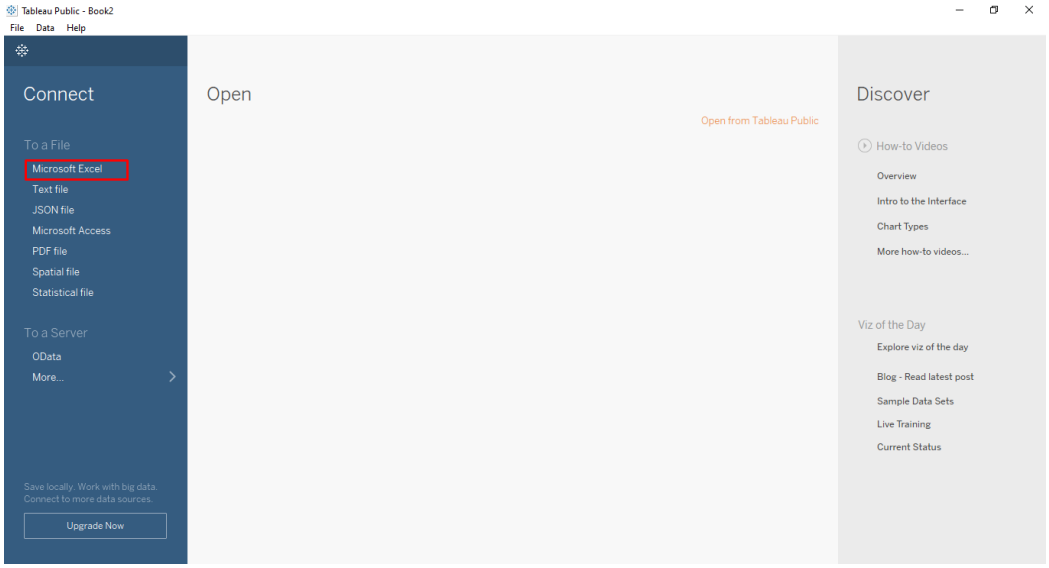



Figure 30: Click on Microsoft Excel file

2. Then, choose any of your database files (.xlsx) from your computer.

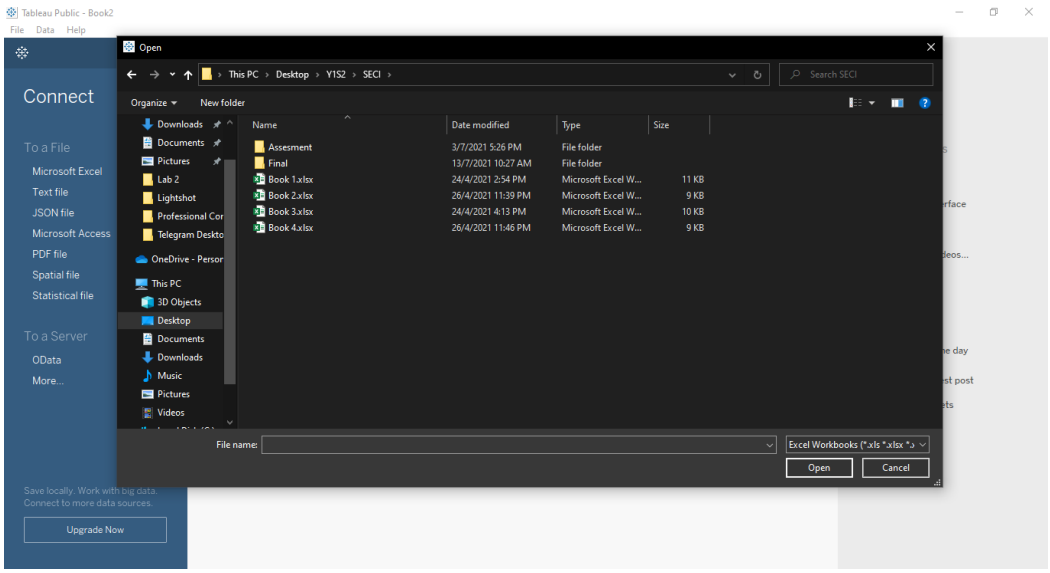



Figure 31: Choose your .xlsx file from your computer

### 2.2.1.2.2 Text File

1. Click on the Text file at the Start Page.

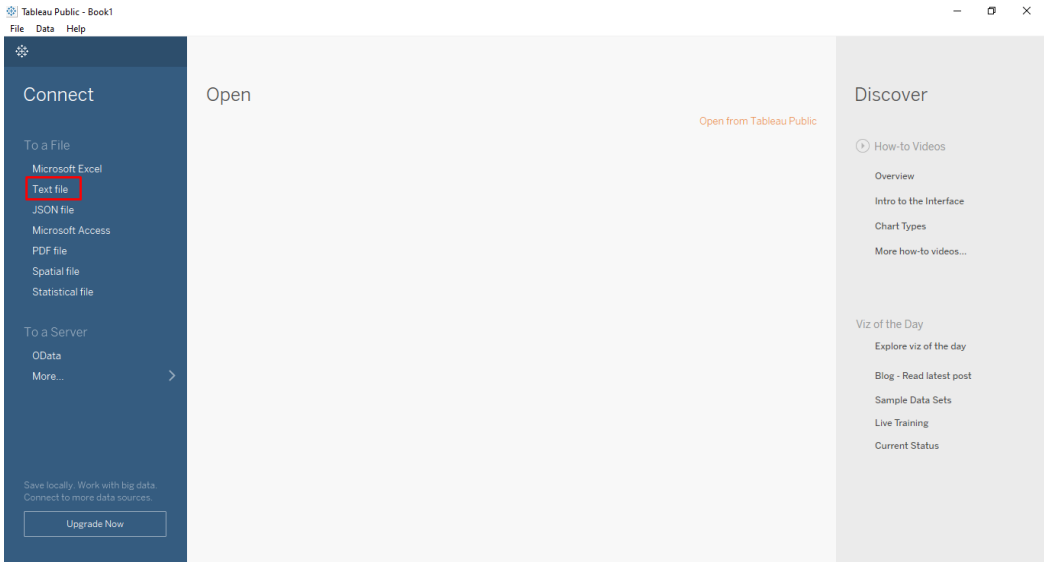



Figure 31: Click on Text file

2. After that, choose your .txt file from your computer.

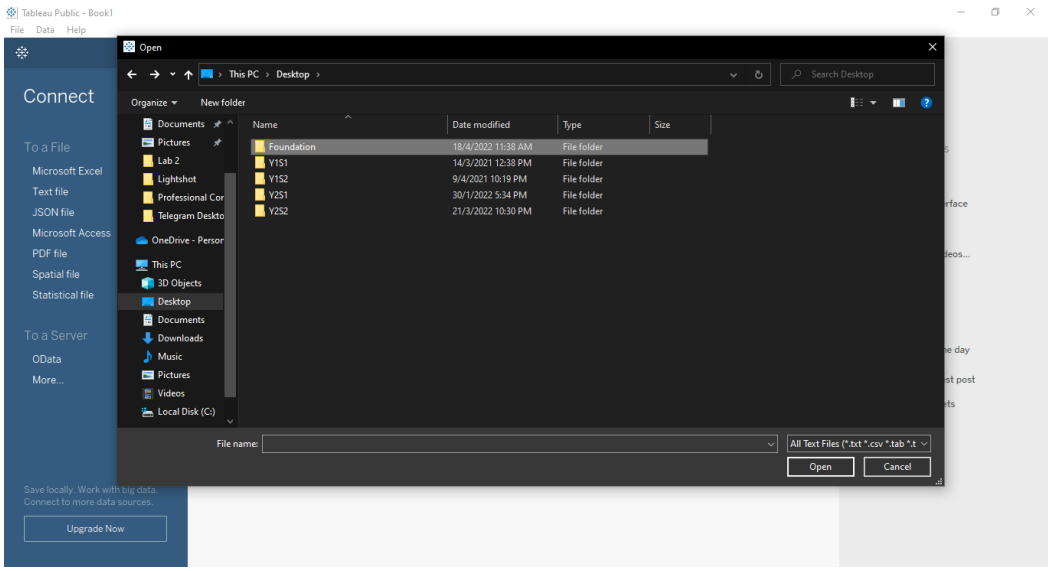



Figure 32: Choose your .txt file from your computer

## 2.2.1.2.3 PDF File

1. First of all, click on the PDF File at the Start Page.

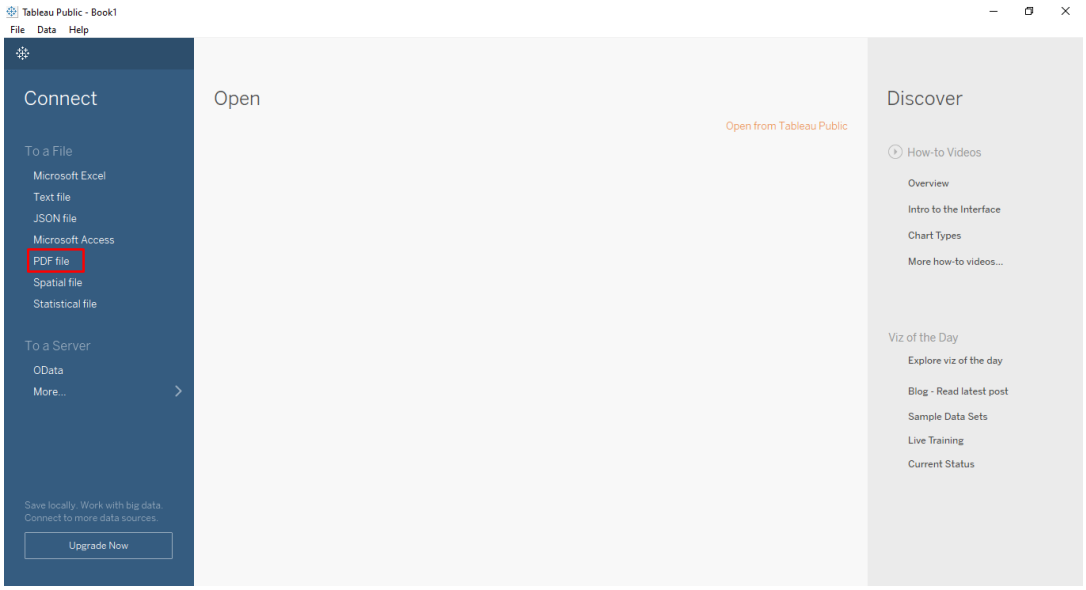



Figure 33: Click on PDF file

2. After that, choose your PDF file from your computer.

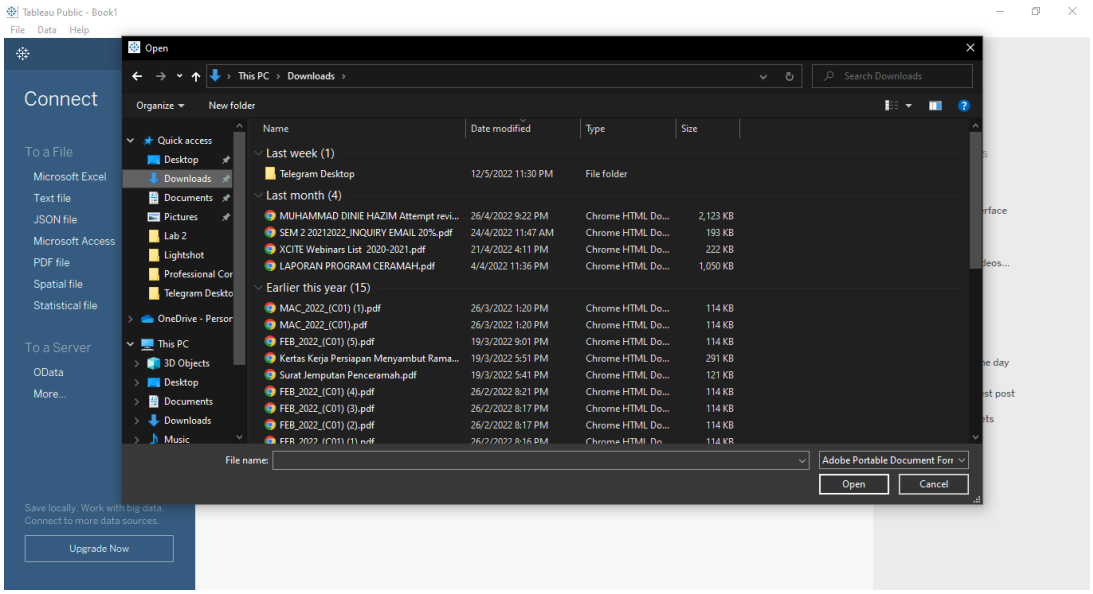



Figure 34: Choose your .pdf file from your computer

### 2.2.1.2.4 Statistical File

1. Click on the Statistical file at the Start Page.

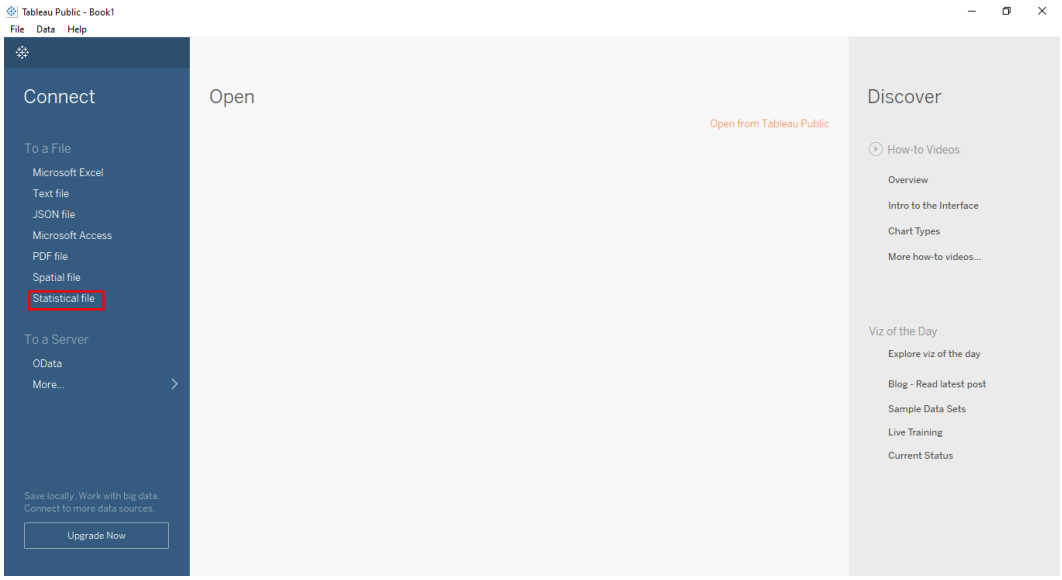



Figure 35: Click on Statistical file

2. Choose your .sav, .sas7bdat, .rda or .rdata file from your computer.

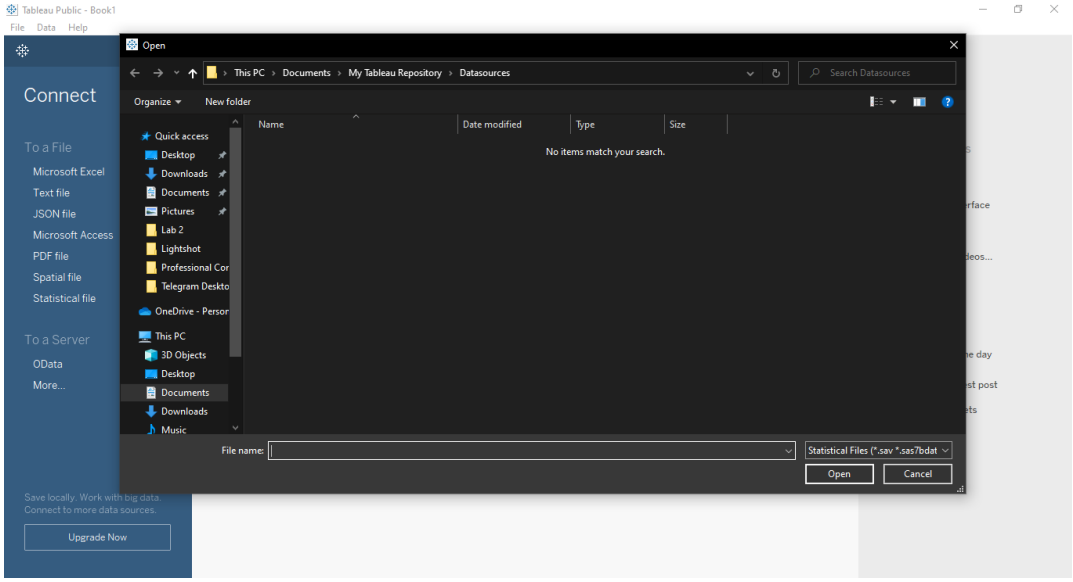



Figure 36: Choose your Statistical file from your computer

### 2.2.1.2.5 Spatial File

1. First, click on the Spatial file at the Start Page. Then, choose your .geojson, .json, .kml, .mif etc file from your computer.

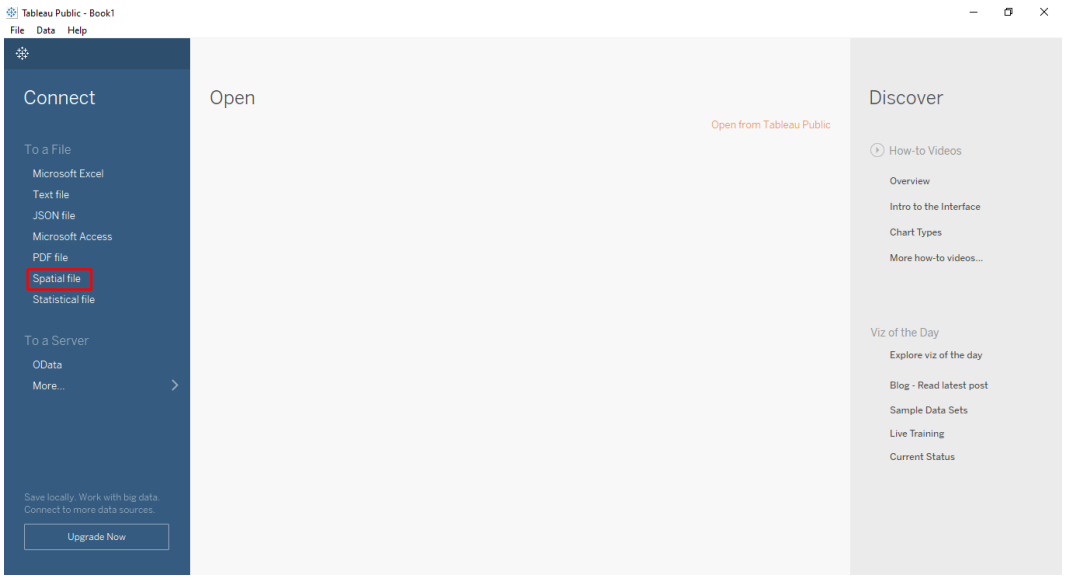



Figure 37: Click on Spatial file

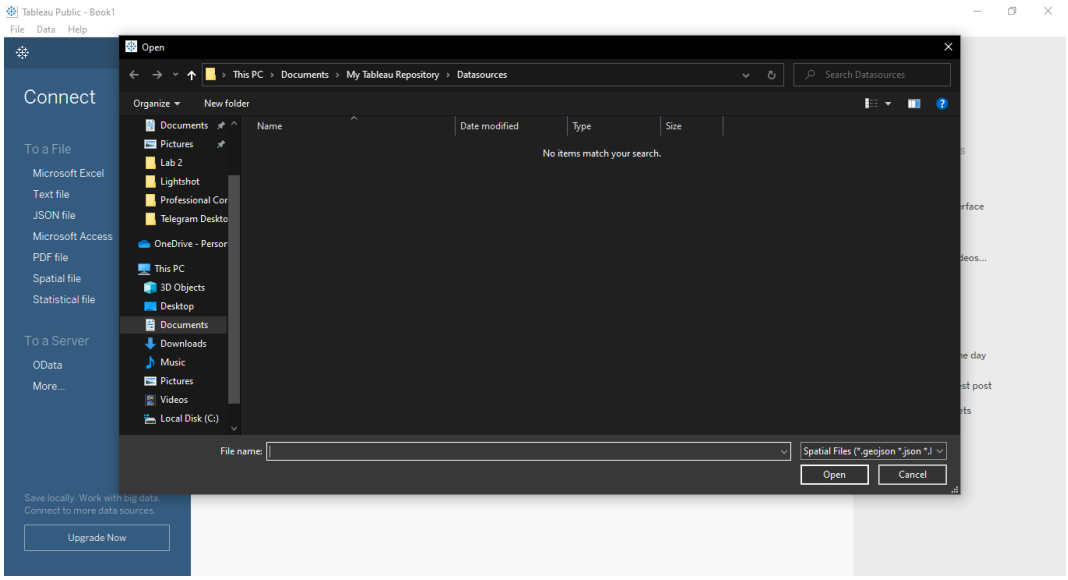



Figure 38: Choose your Spatial file from your computer

## 2.2.2 Manage Data Source

### 1) Edit

a. On the Data menu, select a data source, and then select Edit Data Source.

b. On the data source page, make the changes to the data source.

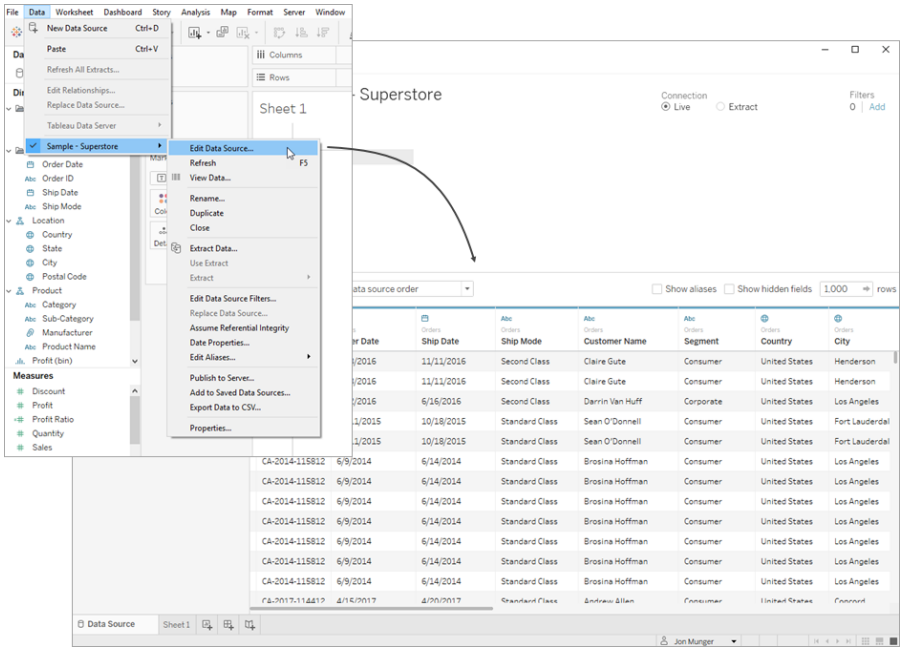



Figure 39: How to edit your data source

### 2) Refresh

a. If you are connected to a data source that has been modified, you can immediately update Tableau Desktop with the changes by selecting a data source on the Data menu and then selecting Refresh.

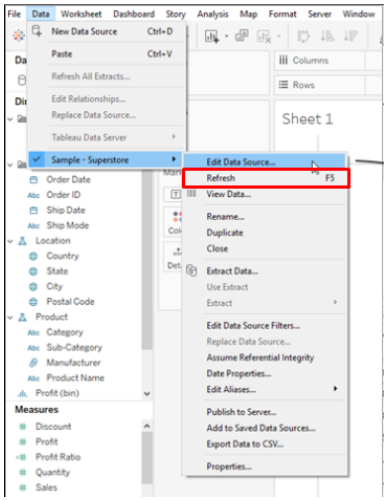



Figure 40: How to refresh your data source

**3) Replace**

a. Open a workbook that connects to the original data source.
b. Select Data > New Data Source and then connect to the new data source.
c. On the Data Source page, drag a table to the canvas to set up the data source (if this is not automatically done for you).
d. Go to the sheet tab and select Data > Replace Data Source.

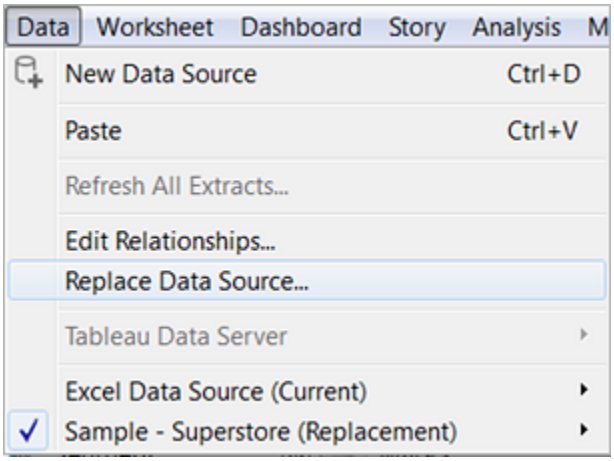



Figure 41: How to replace your data source

e. In the Replace Data Source dialog box, select the Current data source and the Replacement data source.

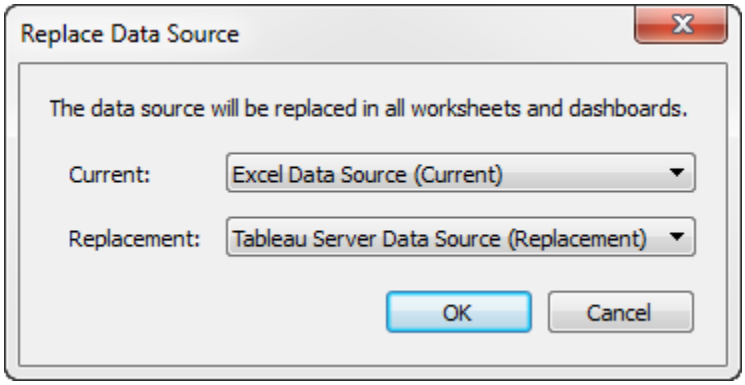



Figure 42: Select your current data source

f. When finished, click OK.

**4) Export/Save**

a. In Tableau Desktop, open the workbook that has the connection to the data you want to save as a file.
b. 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.
c. Enter the file name, select the file type (.tds or .tdsx), and then click Save.

**5) Upgrade**

a. If you have workbooks that were created before Tableau Desktop 8.2 that use Microsoft Excel or text file data sources, or you are using the Excel or text file legacy connection option, you have the option to upgrade the data sources in your workbook. By upgrading your Excel and text file data sources you can take advantage of better data interpretation, and compatibility on the Mac. If you are considering upgrading an existing workbook to use the new connection, review the tables below to compare how your data is processed and displayed between the legacy and default connection types.

b. If there are certain capabilities of the Jet driver that your data relies on, you might need to use the legacy connection to get the expected results. Workbooks that were created before Tableau Desktop 8.2 that use Excel and text file data sources automatically use the legacy connection. To create a new workbook that uses the legacy connection, in Tableau Desktop, navigate to your Excel or text file data, click the drop-down arrow on Open, and then select Open with Legacy Connection.

**File formats and attributes**

The following tables show some examples of the types of Excel and text files, and tables that you can connect to in Tableau when using the legacy and default connection types.

**Excel**

| Your Data | Legacy Connection | Default Connection |
|---|---|---|
| .xlsb file format | Allows connections to Excel data in .xlsb file format. | Does not allow connections to Excel data in .xlsb format. Connections to Excel data in .xls or .xlsx file formats can be used instead. |
| Excel built-in named ranges | Allows connections to built-in named ranges. | Built-in named ranges are hidden. |
| Hidden sheet | Allows connections to a table in a hidden sheet. | Hides a table in a hidden sheet. |
| Very hidden sheet | Allows connections to a table in a very hidden sheet. | Hides a table in a very hidden sheet. |

| | | |
|---|---|---|
| Table that contain charts | Allows connections to tables that contain Excel charts. However, the table does not contain any values. | Hides tables that contain Excel charts. |
| Empty table | Allows connections to an empty table. | Hides an empty table. |
| Table headers | Treats the first row of a table as a header. | Automatically detects whether the first row of a table is a header. |
| Table width | Limits the table width to 255 columns. | No table width constraints. |
| Empty columns and rows | Empty columns and rows are visible. | Empty columns and rows are hidden because they do not contain any values. |

**Text**

| Your Data | Legacy Connection | Default Connection |
|---|---|---|
| File name length | Does not allow connections to a file whose file name is more than 64 characters long. | No length constraints for file names. |
| Multiple period in the file name | Does not allow connections to a file whose file name contains multiple periods. | No period constraints for file names. |
| Number of columns in the file | Uses the number of fields present in the first row to determine the number of columns in the file. | Scans the entire file and uses the most common number of fields in a row to determine the number of columns in the file. |

| | | |
|---|---|---|
| Headers rows | Does not automatically detect whether the first row in the file is a header.<br><br>Note: You can manually specify that the first row in the file is a header. | Automatically detects whether the first row in the file is a header.<br><br>Note: You can manually override the detection. |
| Empty columns | Empty columns are visible. | Empty columns are hidden because they do not contain any values. |

**Characters and formatting**

The following tables show some examples of how your Excel and text file data might be displayed in Tableau when using the legacy and default connection types.

**Excel**

| Your Data | Legacy Connection | Default Connection |
|---|---|---|
| Number of characters in the field name | Field names are truncated to 64 characters. | No constraint on the number of characters in a field name. |
| Special characters in field names | Special characters are not allowed in field names. For example, special characters like quotation marks and periods are converted to number signs. Brackets are converted to parentheses. | Special characters are allowed in field names. |
| Leading and trailing spaces in field names | Leading and trailing spaces in field names are allowed. | Leading and trailing spaces in field names are automatically removed from column names. |

| | | |
|---|---|---|
| Duplicate field names | For duplicate field names, an index number is appended onto the field name.<br><br>For example, Test Scores1. | For duplicate field names, a space and an index number are appended to the field name.<br>For example, Test Scores 1. |
| Excel cell formatting | Supports cell formatting configured using the Format Cells option in Excel. | Does not support cell formatting configured using the Format Cells option in Excel. |
| Precision with currency values | Fields that contain currency values are represented with a maximum precision of four digits | Full precision of currency values is represented. |

**Text**

| Your Data | Legacy Connection | Default Connection |
|---|---|---|
| Special characters in file names | Special characters are not allowed in file names.<br><br>For example, special characters like periods are converted to number signs. | Special characters are allowed. |
| Characters in header field names | Automatically removes leading spaces from header field names. | Automatically removes leading and trailing spaces from header field names. |
| Field separators | Recognizes only commas as field separators.<br><br>Note: You can manually override the detection. | Automatically detects the field separator.<br><br>Note: You can manually override the detection. |

| | | |
|---|---|---|
| Text enclosed with quotation marks | Quotation marks used to enclose text values remain visible. | Quotation marks used to enclose text values are automatically hidden. |
| ANSI and OEM code pages | ANSI and OEM code pages are supported. | ANSI an OEM code pages are automatically converted to be cross-platform compatible. |
| BOM code pages | Inconsistently supports BOM code pages. | Supports BOM code pages. |

**Data type detection**

The following tables show some examples of how data type is detected and how certain values are displayed in Tableau when using the legacy and default connection types.

**Excel**

| Data Type Detection | Legacy Connection | Default Connection |
|---|---|---|
| Columns | Data type of a column is determined by the first 8 rows.<br><br>Note: After the data type of a column in the table has been determined, it cannot be changed. | Data type of a column is determined by the 95% of the first 10000 rows.<br><br>Note: You can manually override the detection. |
| Date values without time | Date values are assigned a datetime data type. | Date values without time are assigned a date data type. Date values with time are assigned a datetime data type. |
| Number values | All number values are represented as real numbers. | Number values without decimal points are represented as integers. |

| | | |
|---|---|---|
| Null values | If a column contains a null cell, the data type for the column is automatically designated as a string data type. | Null cells do not affect data type detection. |
| Reference errors or empty cells | If a column contains cells with reference errors or empty cells, the entire column is interpreted as a string data type. | Reference errors or cells with no values do not affect the data type detection. |
| Overriding data type detection | After the data type of a column in the table has been determined, it cannot be changed. | Data type of a column can be changed after the automatic detection. |
| Time value precision | The smallest measurement of time values is whole seconds. | The smallest measurement of time values is fractional seconds. |
| DecimalSymbol and CurrencyDecimalSymbol schema.ini values | Both DecimalSymbol and CurrencyDecimalSymbol schema.ini values are recognized. | DecimalSymbol and CurrencyDecimalSymbol scheam.ini values are recognized. However if both values are used, DecimalSymbol takes precedence. |
| Cells formatted as text | The data type of a column is detected as a string data type when cells are formatted as text using the Format Cells option in Excel. | Does not support cell formatting configured using the Format Cells option in Excel. |

**Text**

| Data Type Detection | Legacy Connection | Default Connection (8.2 and later) |
|---|---|---|
| Columns | Data type of a column is determined by the first 25 rows. | Data type of a column is determined by the first 1024 rows. |
| Boolean (True/False) values | Boolean values are assigned the string data type. | Boolean values are assigned the boolean data type. |
| Values that become null | Spaces in a cell, whether enclosed in quotation marks or not, are treated as null values. Columns with null values are detected as the string data type. | Two field separators in a row are treated as a null value. Null values are ignored during data type detection. |

**Data source connection properties**

The following table shows examples of the differences in data source connection properties in Tableau when using the legacy and default connection types.

| Property | Legacy Connection | Default Connection |
|---|---|---|
| Custom SQL | Custom SQL is allowed. | Does not allow the use of Custom SQL. |
| Join type | Allows left, right, and inner join types. | Allows left, right, inner, and full outer join types. |
| Join operators | Allows equal to (=), greater than (>), greater than or equal to (>=), less than (<), less than or equal to (<=), and not equal to (<>) join operators. | Allows equal to (=) join operat |

**6) Close**

a. You can close a data source at any time. Doing so does not modify the data source. Instead, it disconnects Tableau from the data so that you can no longer query it. Additionally, the data source is cleared from the Data pane and all open worksheets associated with the data source are cleared. If you accidentally close a data source, use the Undo button to reopen it.

b. Right-click (control-click on a Mac) the data source at the top of the Data pane and select Close.

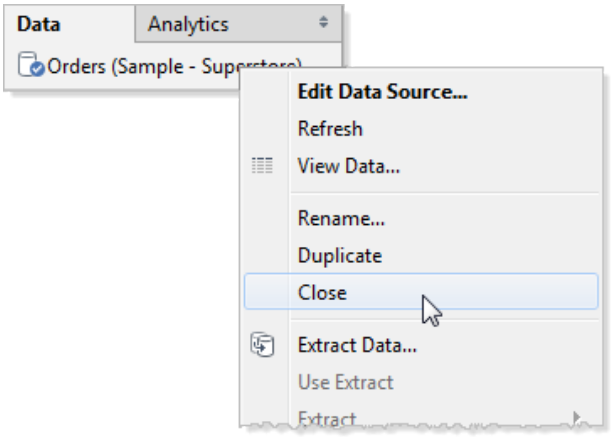



Figure 43: How to close data source in Mac OS

c. Select a data source on the Data menu and then select Close.

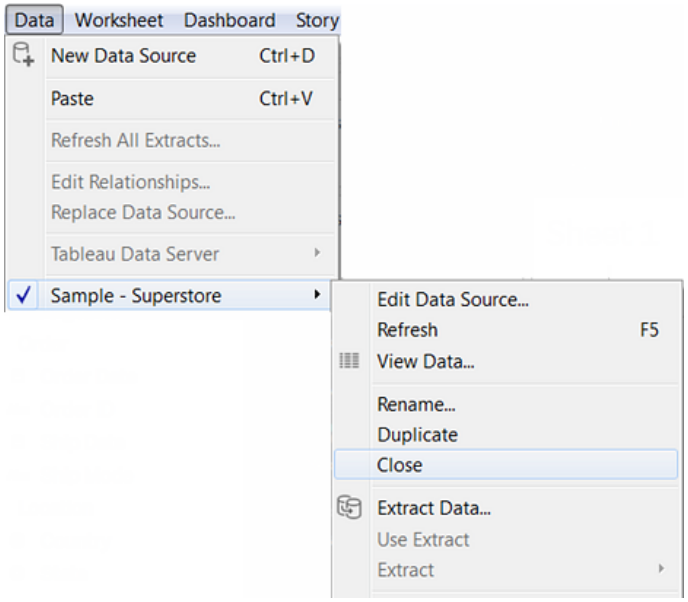



Figure 44: How to close data source in Windows OS

### 2.2.3 Publish Data Source

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. Below these steps you can find supplemental information for authentication types and using Tableau Bridge.

1. 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.

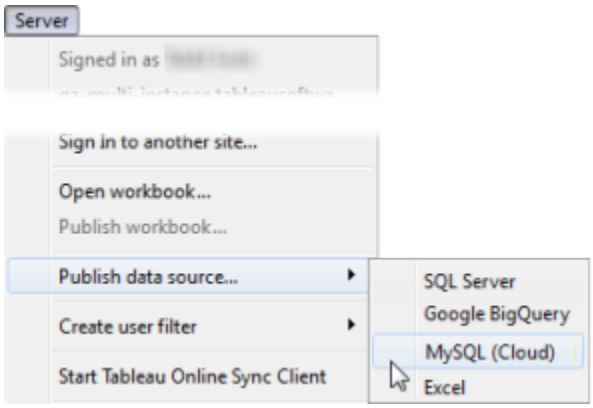



Figure 45: How to publish your data source

2. If you're not already signed in to Tableau Server or Tableau Online, sign in now.
3. In the Publish Data Source dialog box, do the following:
   a. For Project, select the project you want to publish to and enter the data source name.
   b. 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").
   c. For Refresh Schedule, if publishing an extract, you have the option to select a refresh schedule to ensure that your extract data stays fresh. If you choose not to select a schedule here, you can select a schedule from Tableau Server after publishing.
   d. 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.
   e. For Authentication, if you need to provide credentials to access your data, you can specify how authentication should be handled when the data is published to the server.

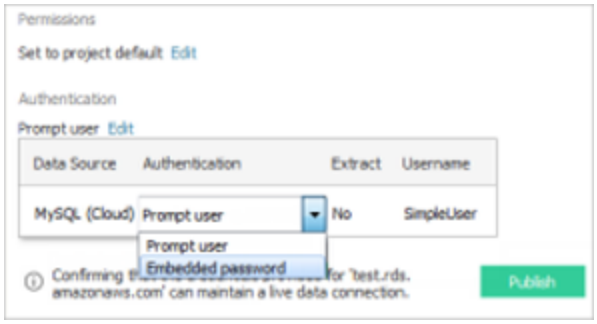



Figure 46: Published data source authentication

The options available for accessing the data source depend on the type of data you publish and whether you are publishing to Tableau Server or Tableau Online.

Information appears at the bottom of the dialog box to let you know whether you need to take further action, such as adding Tableau Online to your data provider's authorized list.

4. 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.

   If you do not want to publish the external files to the server, change the connection information so that the data source references a full UNC path. For example, rather than connecting to D:\datasource.xls, you would connect to \\filesrv\datasource.xls.

5. By default, during the publishing process, Tableau updates the workbook connection to use the new published data source. It also closes the local data source.

   To continue using the local data source instead, clear the Update workbook to use the published data source check box.

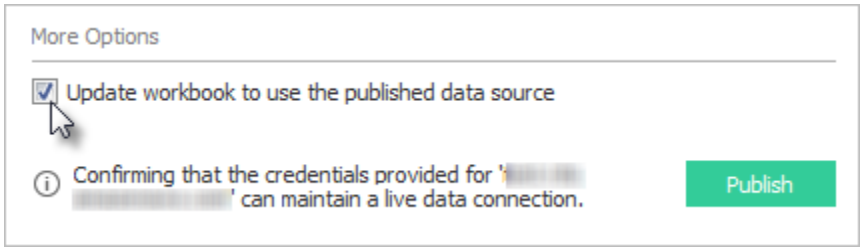



Figure 47: Update your workbook

6. Click Publish.

   After publishing is complete, your web browser opens Ask Data for the data source, where you can ask questions to automatically create vizzes.

7. (Optional) Set up a refresh schedule on the server.

### 2.2.4 Connect Multiple Data Sources

1. Connect to the primary data source as usual.
2. Create a Worksheet. This Worksheet will utilize data from your first data source.
3. Create a new Worksheet.
4. Using the top menu bar, select Data > New Data Source.
5. Connect to a new data source (and if necessary, drag the tables into the (empty) data relationship window). If you like, you can navigate back to the most recently created Worksheet to see fields from the new data source available for use.
6. Repeat steps 3 through 5 to add as many data sources & worksheets as desired.

## 2.3 Start Project

### 2.3.1 Start Page

There are three panes  provided in the start page such as follows:

- **Connect** to your data
- **Open** your most recently used workbooks
- **Discover** and explore the content produced by the Tableau community


Figure 48: Starting page

## 1) Connect

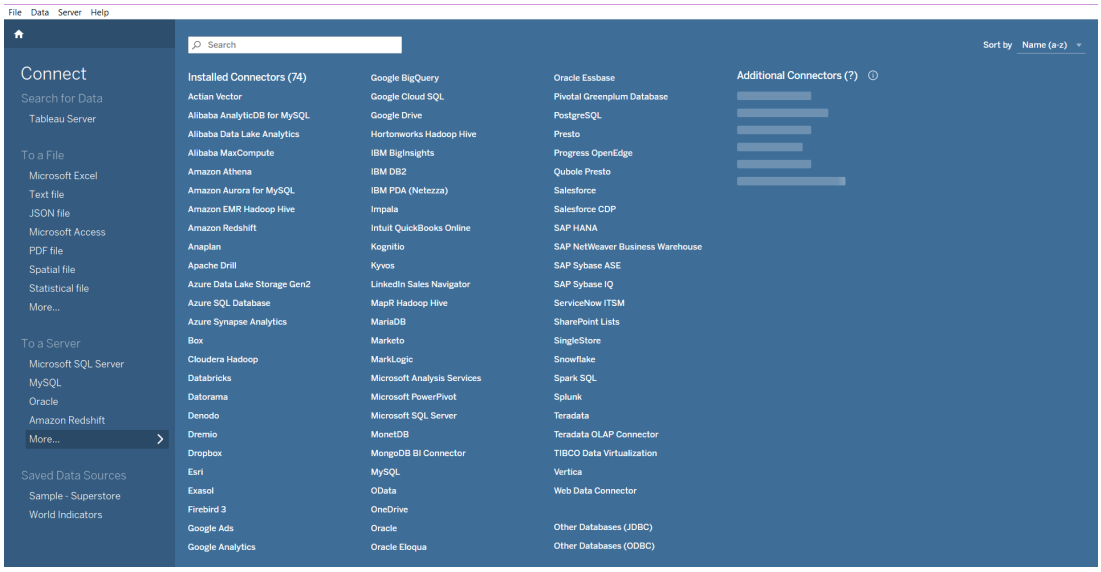



Figure 49: Connect pane

- **To a File:** connects to data stored in Excel, text files, access files and statistical files (such as SAS, SPSS and R).
- **To a Server:** connects to data stored in databases. For instance, Tableau server, Microsoft SQL Server, MySQL, and Oracle.

- **Saved Data Sources:** Open previously saved data sources in My Tableau Repository directory. Besides, sample data sources to explore Tableau Desktop functionality. **Sample-Superstore** data source provides examples in the Tableau Desktop documentation.

## 2) Open

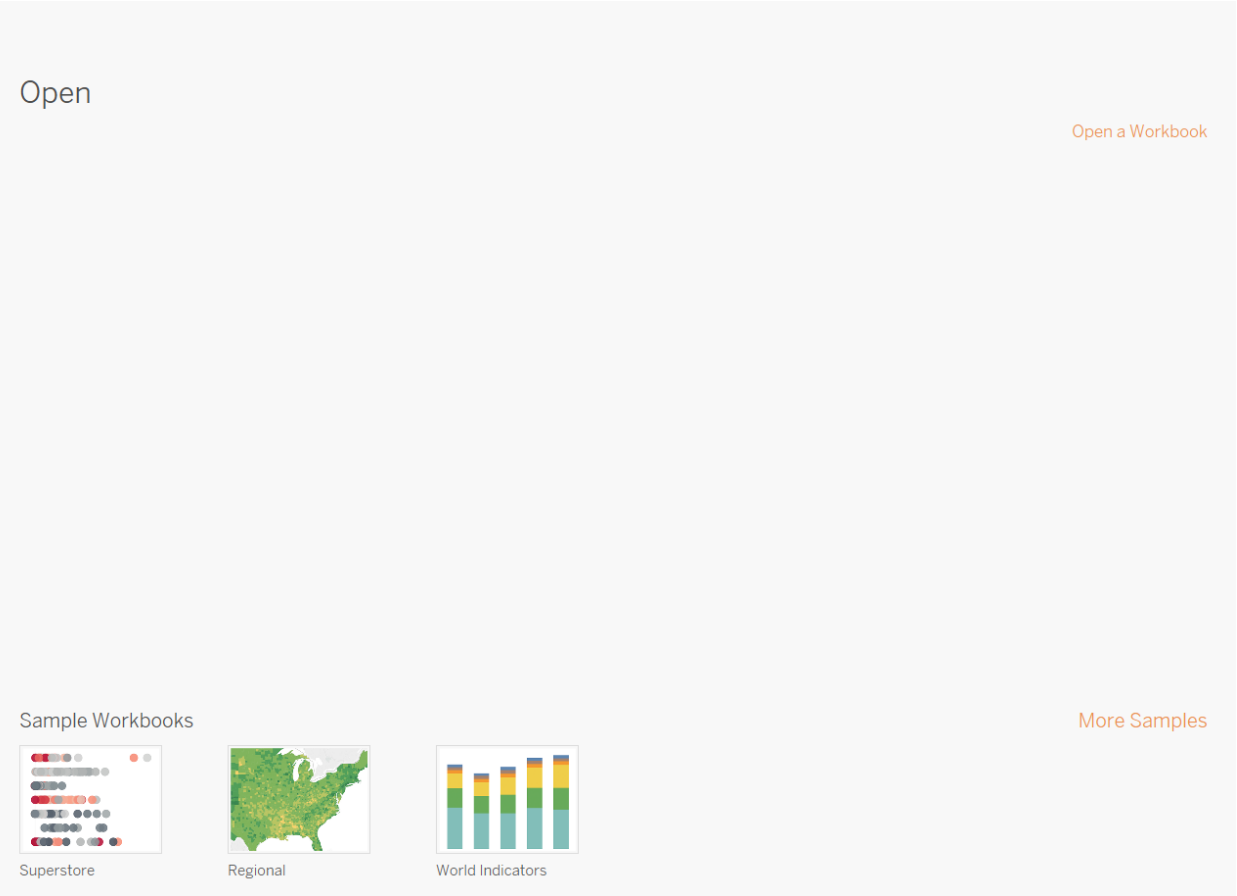



Figure 50: Open pane

- **Open recently opened workbooks:** After creating and saving workbooks, most recently used workbooks will appear here. If the one you search for is not there, click the Open a Workbook link to find other workbooks that are saved to your computer. For someone who just downloaded Tableau, the recently opened workbooks might not be there.
- **Pin workbooks:** You can pin workbooks by clicking the pin icon at the top-left corner of the workbook that appears at the thumbnail. You can also remove the pinned by clicking “x” that appears after hovering over the workbook thumbnail.
- **Sample Workbooks:** Explore some sample workbooks for some ideas.

## 3) Discover

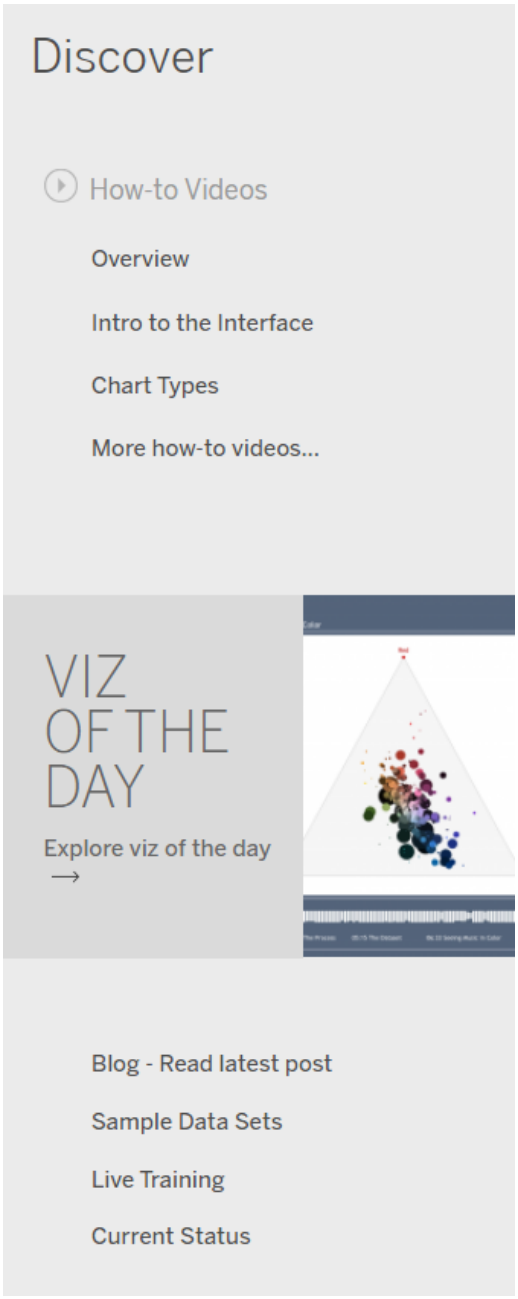



Figure 51: Discover pane

You can explore some popular views in Tableau Public, read blog posts and news about Tableau, find some training videos and tutorials, and many more.

### 2.3.2 Data Source Page

Before doing analysis, you might want to update the data source after making a connection to the data. This will lead you to the data source page. You can also click the **Data Source** tab from your workbook.

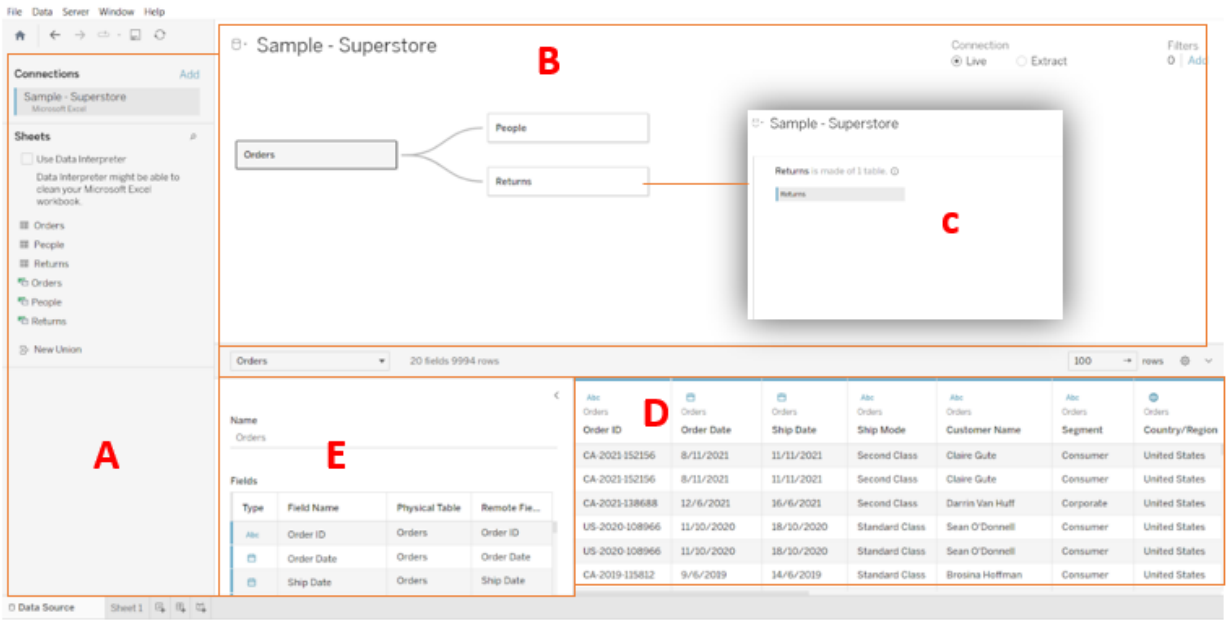



Figure 52: Data source page

**A. Left pane:** Display connected data source with details of your data. You can also add more connections to create cross-database joins. It does not display cube (multidimensional data).

i. **For file-based data:** Left pane displays filename and the worksheets in the file.

ii. **For relational-based data:** Left pane displays the server, database or scheme, and tables in the database.

**B. Canvas (logical layer):** Create relationships between logical tables. This is the default view.

**C. Canvas (physical layer):** Double-click the table in the logical layer to the physical layer, where you can add joins and unions between tables.

**D. Data grid:** Display first 1000 rows of data from data source. You can also modify the data like sorting, hiding fields, renaming fields, reset field names, create calculations, change column, or row sort, or adding aliases.

In the **physical layer,** you can see merged data based on the joins and union.

In the **logical layer**, you will see data from the selected table.

Click a column and drag the mouse to the selected columns to select multiple fields in the grid.

**E. Metadata grid:** Display fields in data source in rows. By this, you can perform routine management tasks, such as renaming fields or hiding multiple fields at once. If connected to cube or some extract only data, it will display it by default.

## 2.3.3 Workspace

The Tableau consists of **menus, toolbar, Data pane, cards, shelves, and sheets (worksheets, dashboards, or stories).**

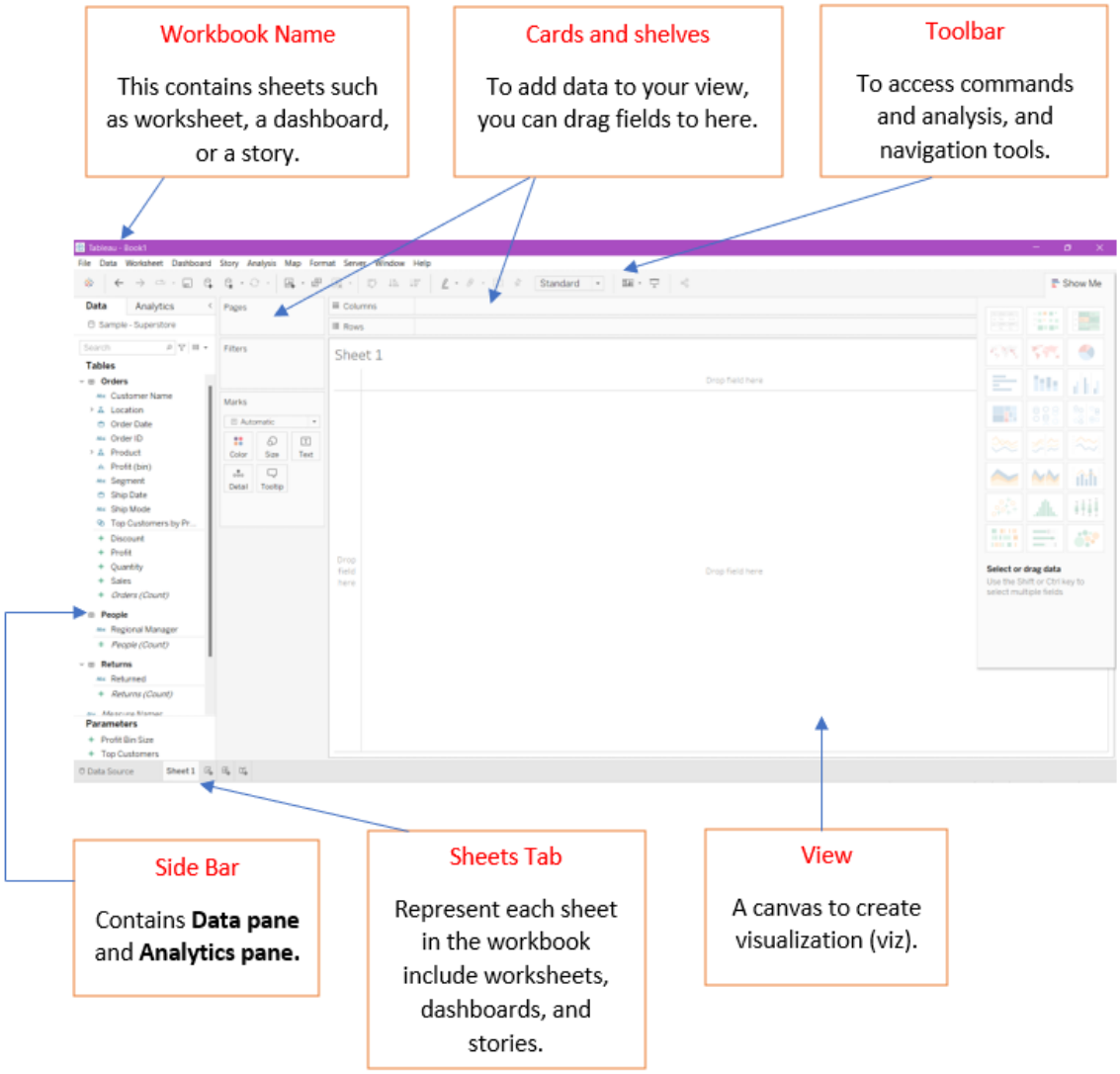
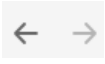
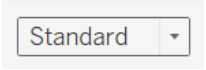



Figure 53: Tableau workspace

Toolbar Button Reference

| Toolbar Button | Description |
|---|---|
|  | Navigate to the **Start Page.** |
|  | **Undo or Redo:** Reverses or repeats work unlimited times. |
|  | **Save** your work. |

| | |
|---|---|
|  | New data source. Opens the Connect pane to create a new connection or open a saved connection. |
|  | **Pause Auto Updates:** Controls if Tableau can update the view when changes are made. Can use a drop-down menu to automatically update the entire sheets or can just use filters. |
|  | **Run update:** Runs a manual query of the data to update the view with changes when automatic update is turned off. Can use a drop-down menu to automatically update the entire sheets or can just use filters. |
|  | **New worksheet:** Create a new blank worksheet, use drop-down menu to create a new worksheet, dashboard, or story. |
|  | **Duplicate** worksheet. |
|  | **Clears:** Clear current worksheet. Can use a drop-down menu to clear specific parts. |
|  | **Swap:** Moves the fields on the Rows shelf to the Column shelf. Alsi the Hide Empty button. |
|  | **Sort Ascending/Descending:** Sort selected field based on the measures in the view. |
|  | **Highlight:** Highlight selected sheet. Can use a drop-down menu to define how values are highlighted. |
|  | **Group members:** Creates groups by combining selected values. Multiple dimensions can use the drop-down menu to specify specific dimensions or across all dimensions. |
|  | **Show Mark Labels:** Switches between showing or hiding mark labels for current sheet. |
|  | **Fix Axes:** Switch between locked axes which only shows a specific range and dynamic axis that can adjust range based on minimum and maximum values in view. |
|   | **Fit:** Specifies how the view should be sized within the window. |
|  | **Show/Hide Cards:** Select cards that you want to show in the drop-down menu. |
|  | **Presentation mode:** Switch between hiding and showing everything except the view. |
|  | **Share workbook with others:** Publish your workbook to Tableau Server or Tableau Online. |

| | |
|---|---|
| Show Me | **Show Me:** Helps you choose a view type by highlighting view types that work best the field types in your data. An orange outline shows the recommended chart type best match for you. |

Show and Hide the SideBar (Data Pane)

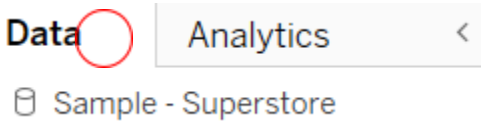



Figure 54: Button to show or hide SideBar

Status Bar Information

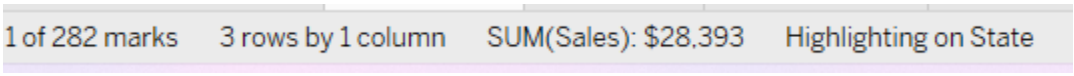



Figure 55: Status bar

You can hide the status bar by selecting **Window > Show Status Bar**.

Areas of the Data Pane

After you connect your data and set up the data source, the data source connections and fields appear on the left side of the workbook in the Data Pane. Current data source connections appear at the top of the Data Pane. Visualization can be built by adding fields from the Data Pane to the View.

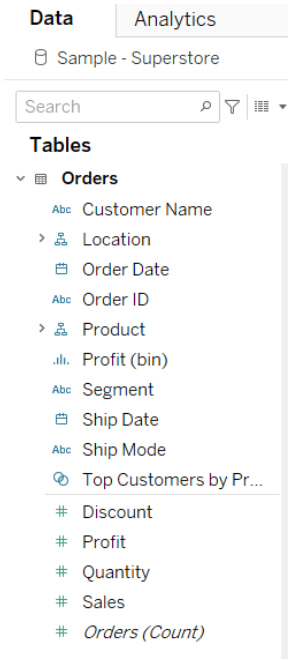


Figure 56: List of data pane fields

You can toggle between **Data** and **Analytics** panes in the worksheet. The **Data** pane includes:

- **Dimension fields:** Contain qualitative values such as names, dates, or geographical data. This affects the level of detail in the view.
- **Measure fields:** Contain numeric, quantitative values that can be measured. You can calculate and aggregate them.
- **Calculated fields:** Create new fields to meet your requirements using calculation and then save them as part of the data source.
- **Sets:** Subset of data that you define.
- **Parameters:** Values that can be used as placeholders in formulas, or replace constant values in calculated fields and filters.

The **Analytics** pane is used to add an analytics object to the view by choose and drag to the view:

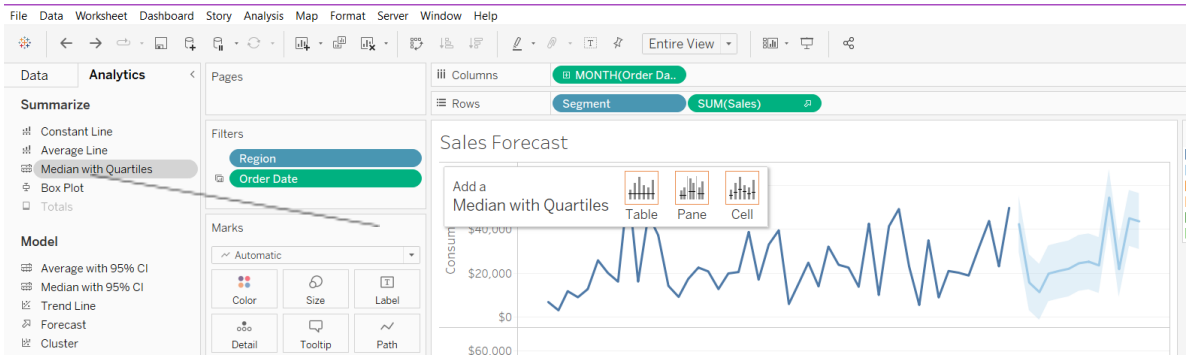



Figure 57: Analytic pane

The following can be dragged from the Analytics pane and dropped in the view. You can also **Edit**, **Remove** or **Format** those.

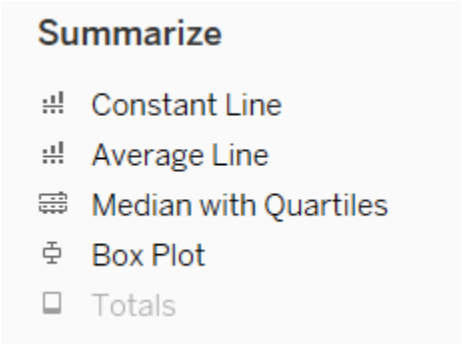



Figure 58: Summarize part in analytic pane

- **Constant Line:** Add constant line for a specific measure, for all measures, or for date dimensions.

- **Average Line:** Click on a resulting average line and choose a different aggregation, such as Total or Sum.
- **Median with Quartiles:** Add one or more sets of median line and distribution bands to the view. You can also add a median with quartiles for a specific measure or for all measures. The distribution bands are computed as quartiles (the middle two quartiles are shaded).
- **Box Plot:** Add boxplots for a specific measure or for all measures.
- **Totals:** Adds Subtotals, Column Grand Totals, and Row Grand Totals.

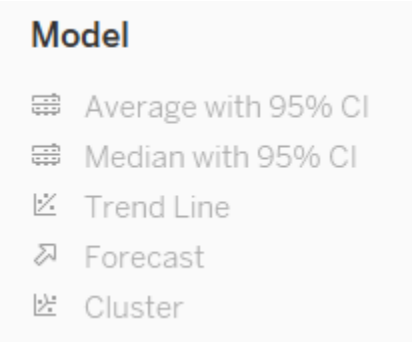



Figure 59: Model part in analytic pane

- **Average with 95% CI:** Shades the region in which the population average will fall 95% of the time.
- **Median with 95% CI:** Shades the region in which the population median will fall 95% of the time.
- **Trend Line:** Adds one or more trend lines to the view. The drop gives options: **Linear, Logarithmic, Exponential,** and **Polynomial.**
- **Forecast:** Only possible when there is at least one measure in the view. The view cannot contain table calculations, disaggregated measures, percent calculations, grand totals of subtotals, and date values with aggregation set to Exact Date or null.

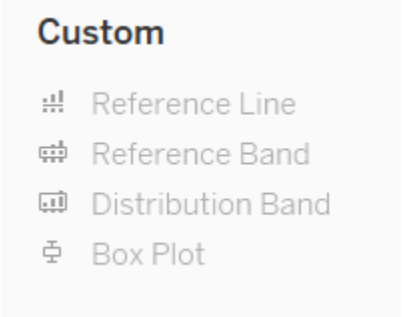



Figure 60: Custom part in analytic pane

- **Reference Line:** Add reference lines for a specific measure or for all measures in the view.

- **Reference Band:** Add reference band for a specific measure or for all measures in the view.
- **Distribution Band:** Add distribution band for a specific measure or for all measures in the view.
- **Box Plot:** Custom box plot.

Shelves and Cards Reference

Few options you can do such as drag fields and double-click from the Data pane onto cards and shelves.


Figure 61: Drop fields space

1. **Columns and Rows shelves**

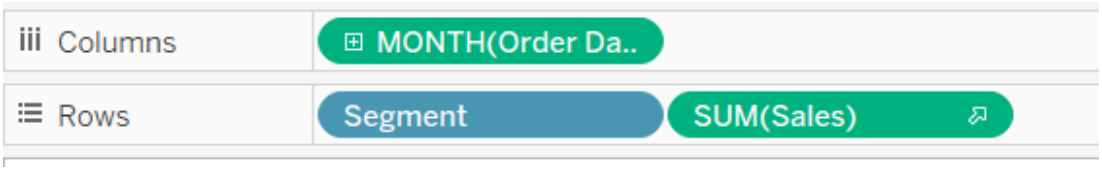



Figure 62: Columns and Rows shelves

The headers of that dimension are created when you place a dimension on the Rows (Horizontal axis) or Columns (Vertical axis). You can hide or show a row or column by right clicking it.

2. **Marks card**

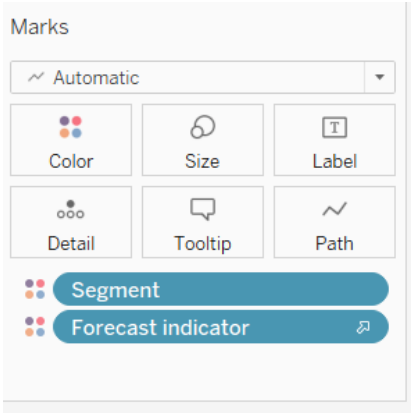



Figure 63: Marks card

This is a key element for visual analysis. From the picture above, two different fields have been dragged to different properties.

3. **Filters shelf**

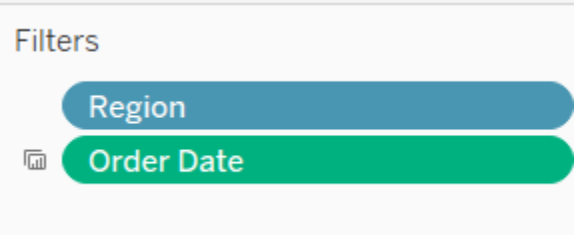



Figure 64: Filters shelf

Allows you to choose which data to include and exclude. You can do an internal filter to filter data based on the fields that make up the columns and rows of the table. External filters on the other hand filter data using fields that don't contribute headers or axes to the table.

4. **Pages shelf**

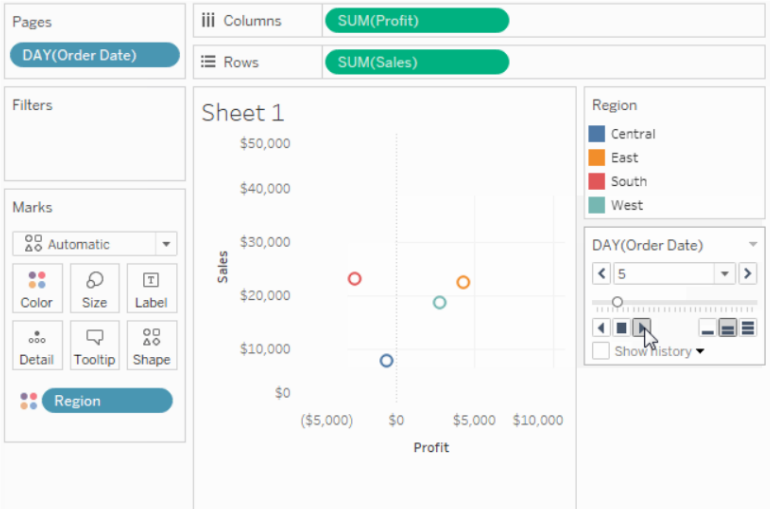



Figure 65: Pages shelf

Break a view into a series of pages so you can better analyze how a specific field affects the rest of the data in a view. You can also jump to the specific page that you want.

View Area

Data views are displayed in a table. It is a collection of rows and columns which consists as follows:

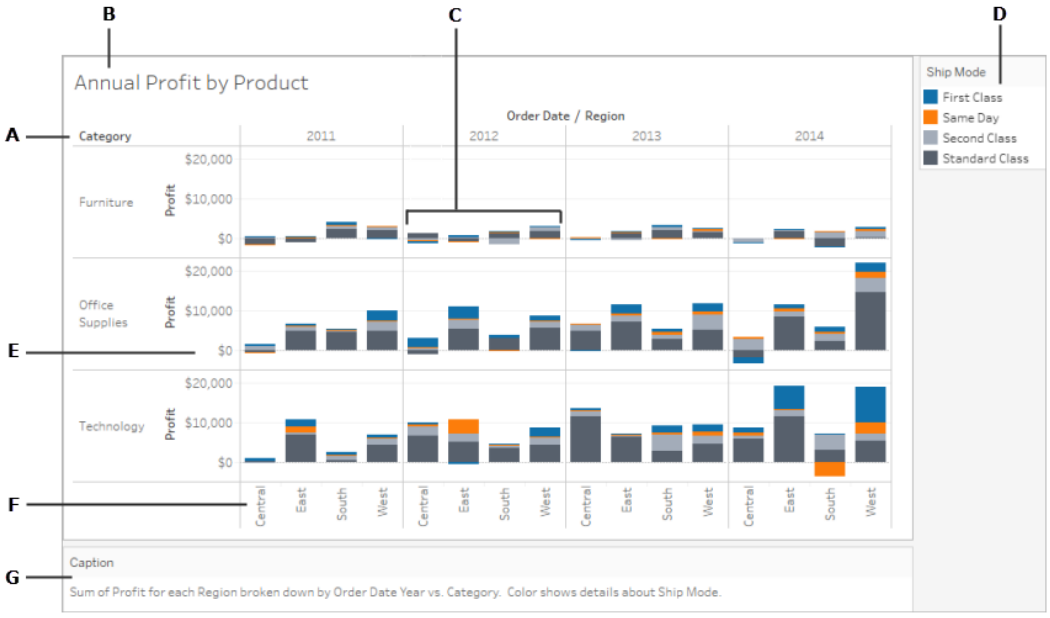



Figure 66: View Area

A. **Field Labels:** Describe members of the field.
B. **Titles:** Name you give to your worksheet, dashboard, or story.
C. **Marks:** Represents the intersection of the fields (dimensions and measures) included in the view. It can be represented using lines, bars, shapes, maps, and so on.
D. **Legends:** Describes how data encoded.
E. **Axes:** Created when you add a measure. The fields contain quantitative or numerical information to the view.
F. **Headers**: Name of each field.
G. **Captions:** Text describing the data in the view.
H. **Tooltips**: Additional data details when you hover one or more marks in the view.

## 2.3.4 Dashboards

A dashboard is a collection of several views which let you compare a variety of data simultaneously. All the views are compiled at a dashboard so that you don't have to navigate through separate worksheets.

Same as worksheets, you can see dashboards from the tabs at the bottom of the workbook. When you update a worksheet, it will reflect the dashboard along with the data source.

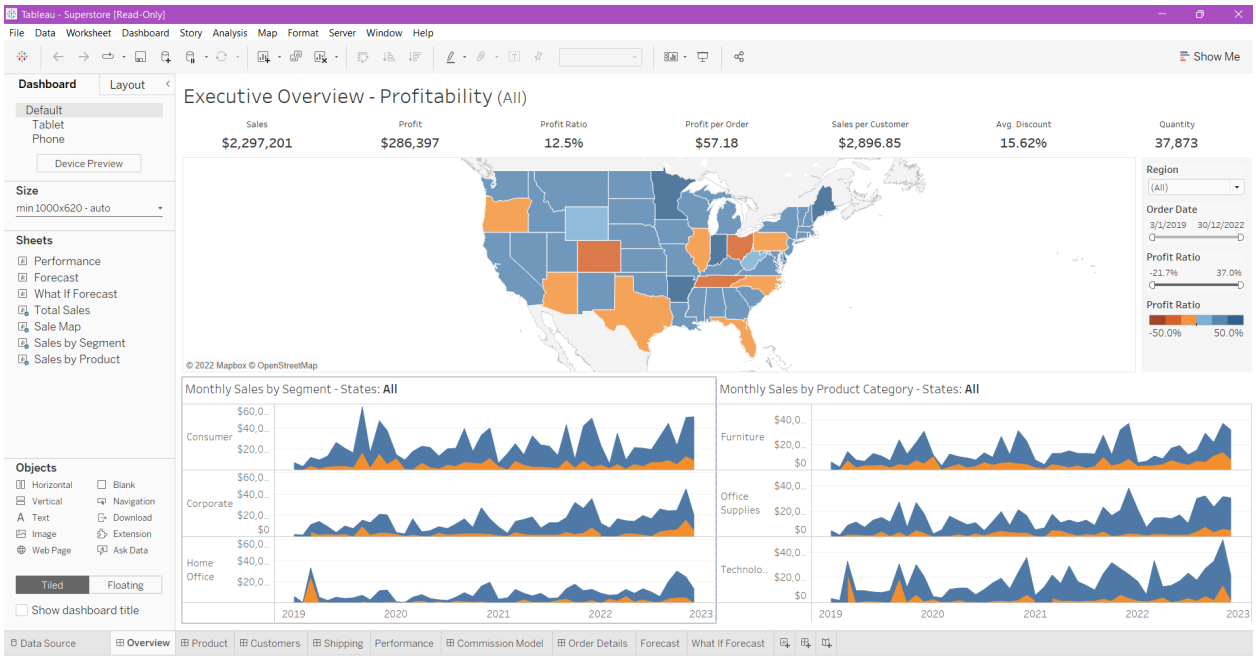



Figure 67: Dashboard

You can also manage your view, such as sizing at the dashboard from the **Dashboard** pane.

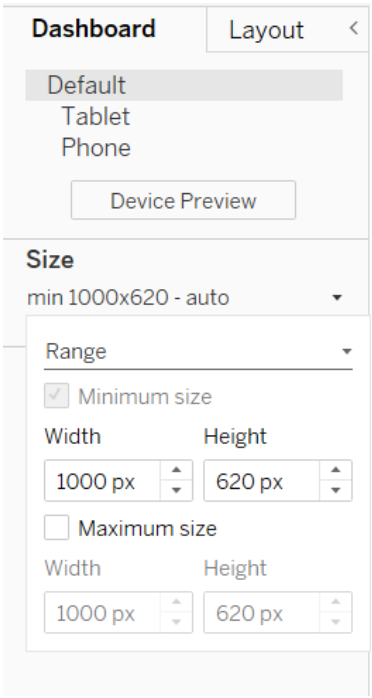



Figure 67: Edit your dashboard view

Create a Dashboard

1. Click the **New Dashboard** icon at the bottom of the workbook.

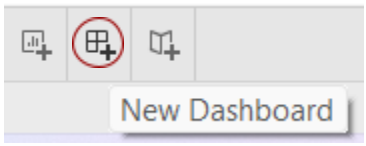



Figure 68: Create new dashboard

2. Drag and drop view from the **Sheets** list at the **Dashboard** pane.

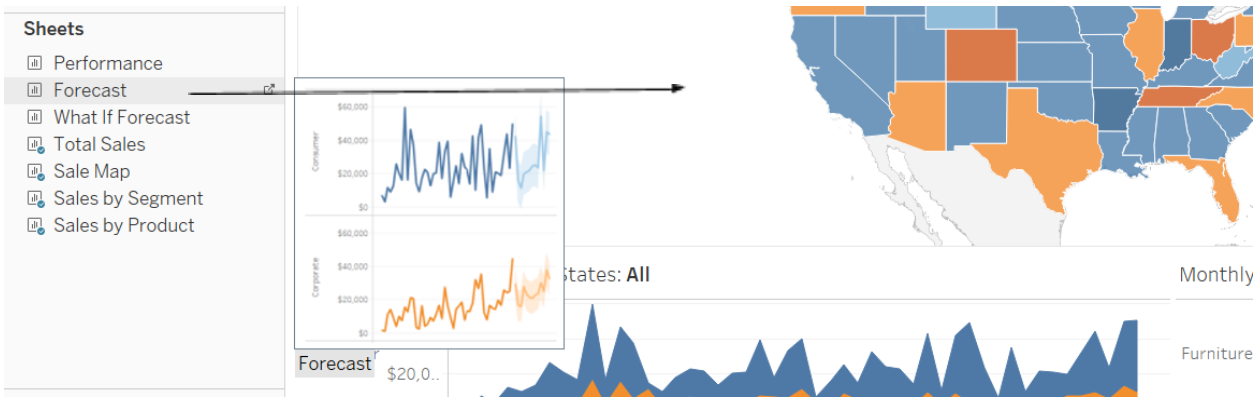



Figure 69: Drag and Drop selected Sheets

Add and Configure a Show/Hide Button

1. Select a dashboard object.
2. From the button **More Options** at the corner of the object, select **Add Show/Hide Button**.

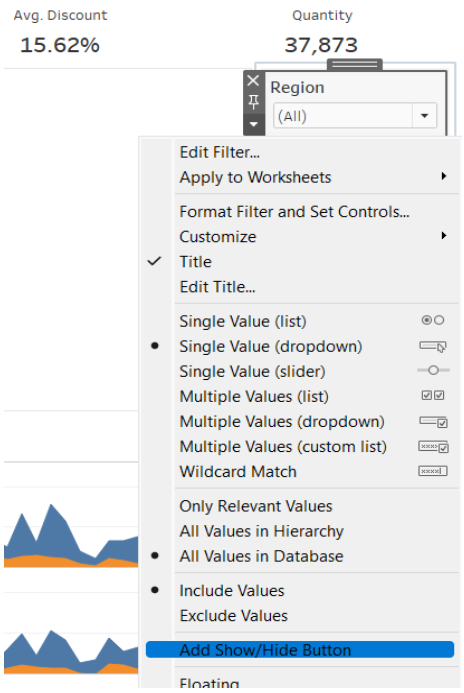



Figure 70: Click Add Show/Hide Button

3. From the button menu, choose **Edit Button**.

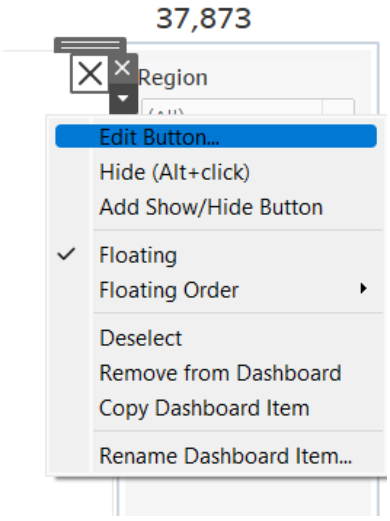



Figure 71: Click Edit Button

4. You can set those options as follows:

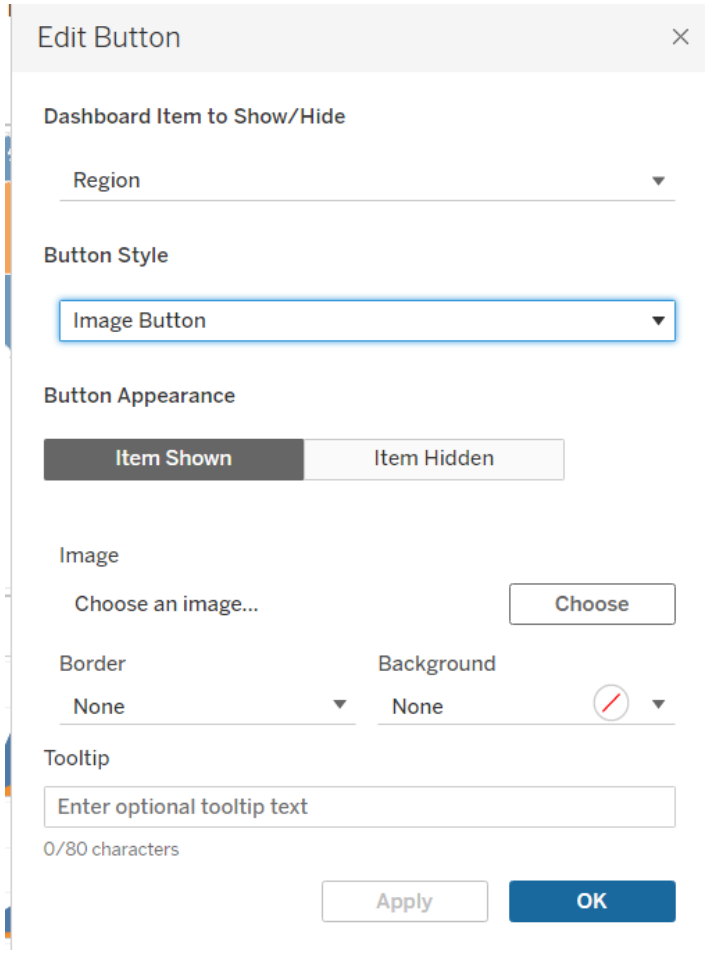



Figure 72: Your edit option

- **Dashboard Item to Show/Hide:** Specifies targeted object.
- **Button Style:** Choose either image or text displays for the button.
- **Button Appearance:** Specifies how the button looks when the item is both shown or hidden.
- **Tooltip text:** Provides explanatory text when you hover the button.

5. You can also drag the button to a different location or resize it.

Size and Layout on Dashboard

There are three dashboard size options such as:

1. Fixed size (default): It becomes scrollable if the dashboard is larger than the window.
2. Range: Dashboard scales between minimum and maximum sizes that you specify.
3. Automatic: Dashboard automatically resizes to fill the window.

You can set the size from here:

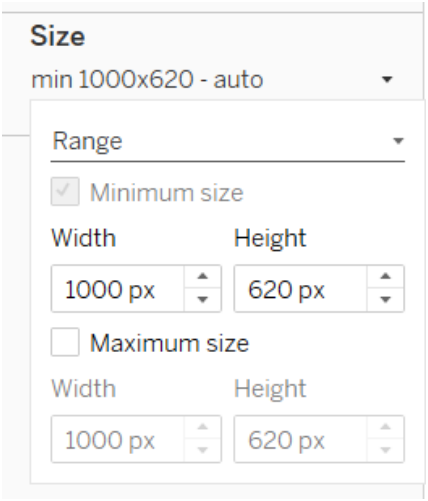



Figure 73: Set the size of Dashboard

Layout the containers either horizontally or vertically.

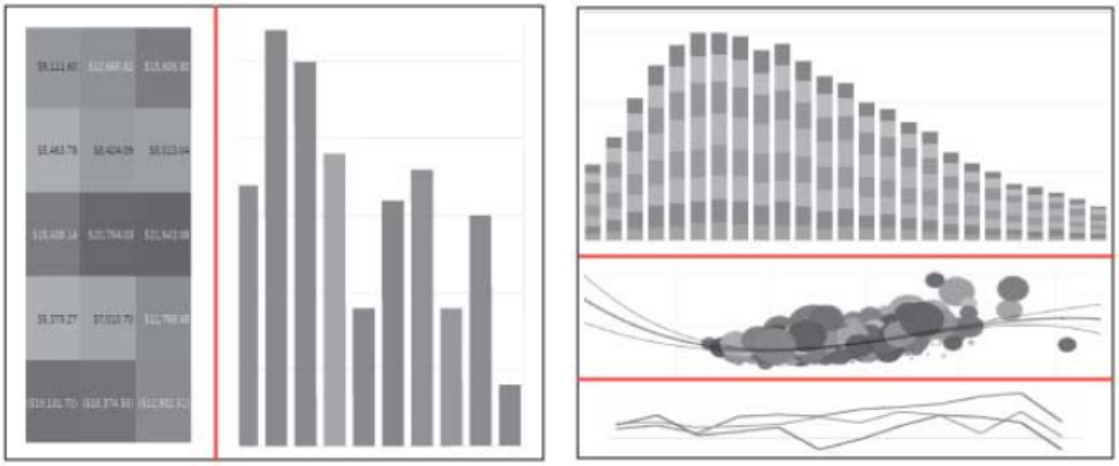


Figure 74: Choose your preferred layout

Add a layout container by following these steps:

1. Under Objects on the **Dashboard** pane, select **Horizontal** or **Vertical**.
2. Drag the container to the dashboard.

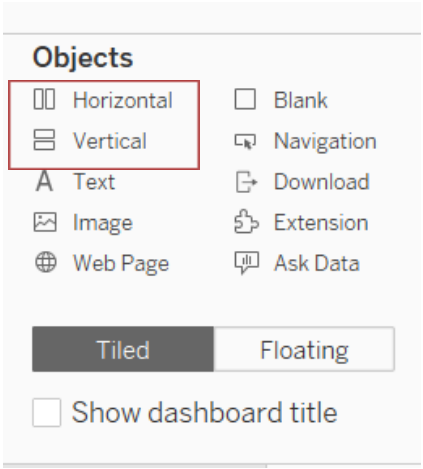



Figure 75: Choose Horizontal or Vertical

3. Add views and objects to the layout container created.

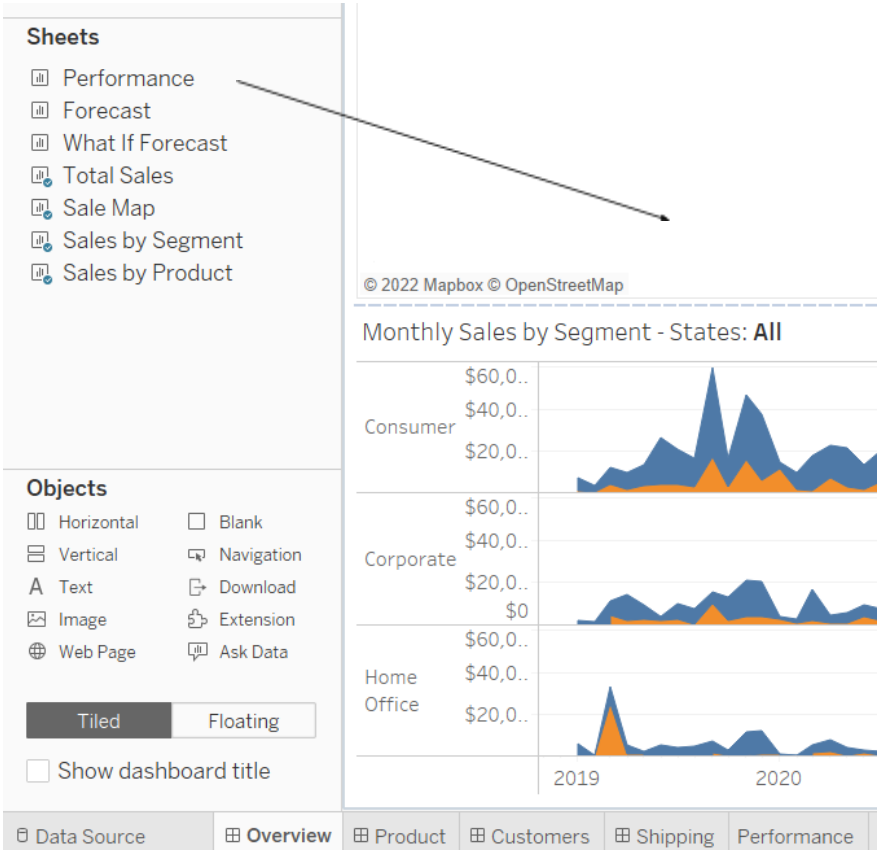



Figure 76: Choose your sheets to view in the layout

4. You can also distribute it evenly from its shortcut menu.

Float or tile a new item

1. Under Objects in the Dashboard pane, click the layout option you want to use, either Floating or Tiled.

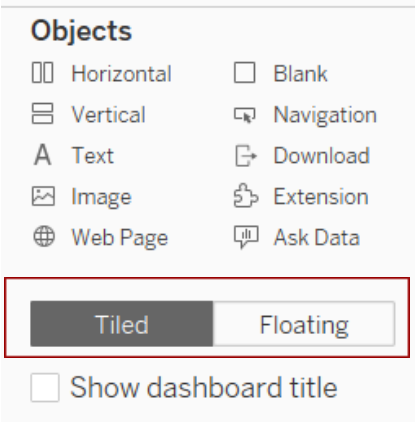



Figure 77: Select Tiled or Floating

2. Drag the view or object onto the dashboard.

Layout Pane

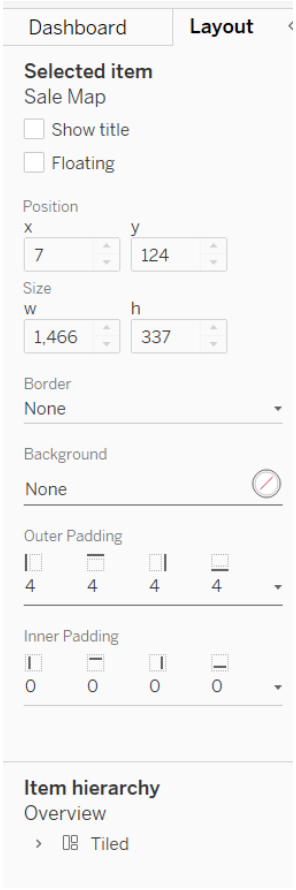



Figure 78: Layout pane

You can done these by referring the **Layout** pane:

1. Position the item by updating the position x and y in pixels.
2. Add outer or inner padding.
3. Arrange item hierarchy.
4. Change background.
5. Add padding and border.
6. Get your data set into a tableau and examine the data needed for visualization on the dashboard.
7. Start creating an object in the dashboard by drag and drop the items from the **Data** pane into the **Columns** and **Rows** pane at the worksheet.

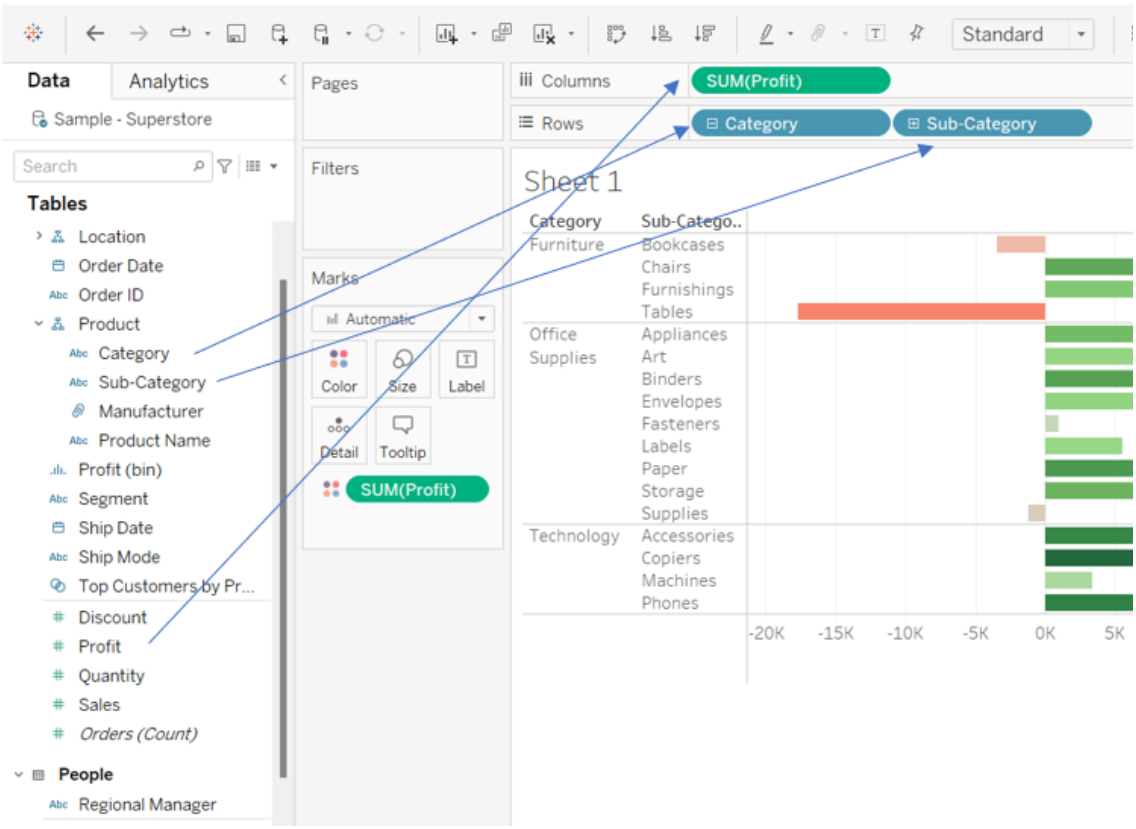



Figure 79: Drag and Drop items

8. Choose your preferred graph by clicking the Show Me button at the top right corner.
9. Drag and press CTRL (for Windows) or command (for Mac) at the same time from data in Columns or Rows to whichever Marks you want for a better visualization of the graph.

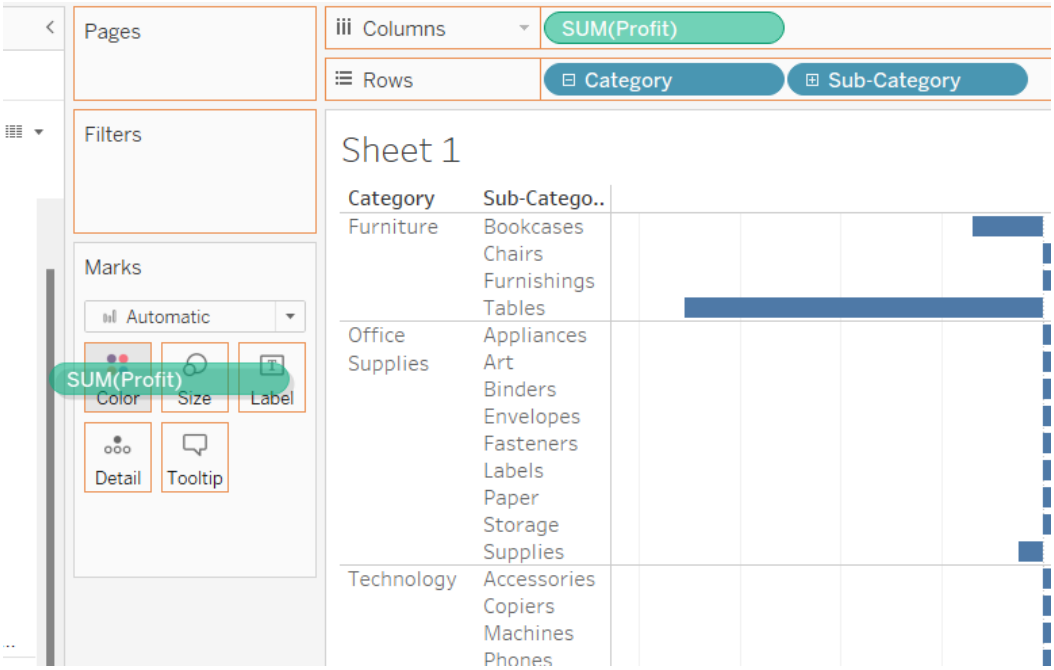



Figure 80: Better visualization

10. Adjust the Marks according to your needs by pressing the icons.
11. Changing your worksheet's name will automatically update the title of the graph.
12. Add new worksheets by clicking the New Worksheet icon at the sheet tab and repeat from step one by using different data and visualization
13. After having several worksheets, you can start building your dashboard.
14. Add a new dashboard by clicking the New Dashboard button at the sheet tab.
15. Before adding your sheets, you can choose how you want to organize your dashboard by choosing some options at the Objects section at the Dashboard pane.

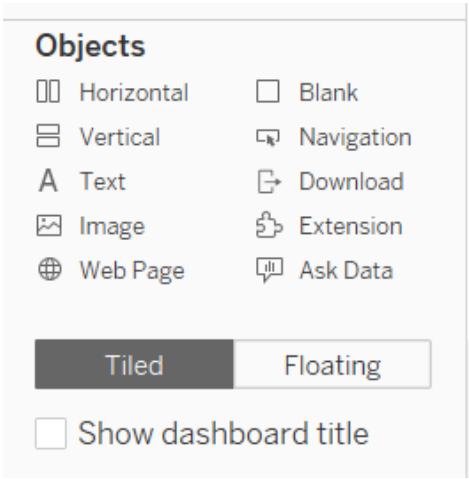



Figure 81: Objects section

16. Drag and drop your Sheets from the Sheets section at the Dashboard pane to the dashboard in a good position.

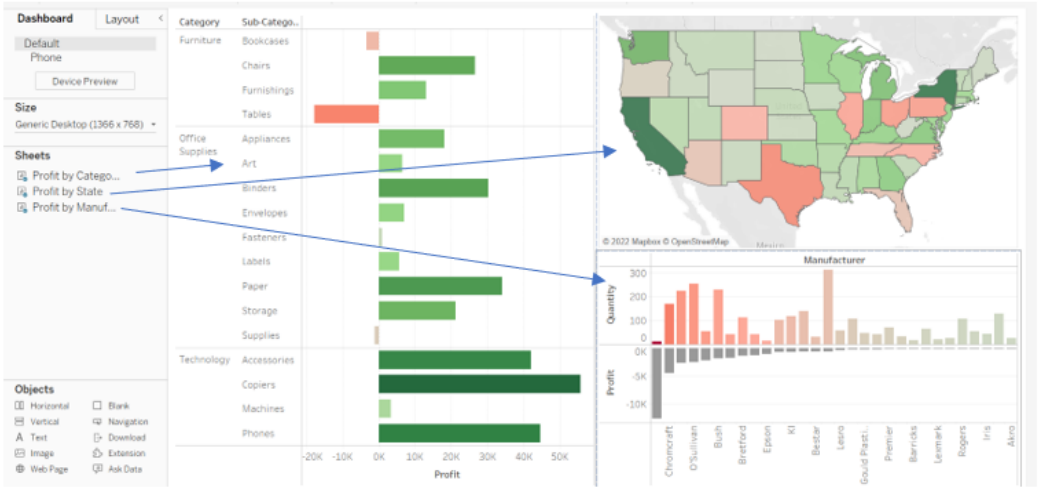



Figure 82: Drag and drop your Sheets from the Sheets section

17. Click Use Filter options after selecting the object to help you update the same data of the others graph. The other information related to your selection will be emphasized on the other graphs.

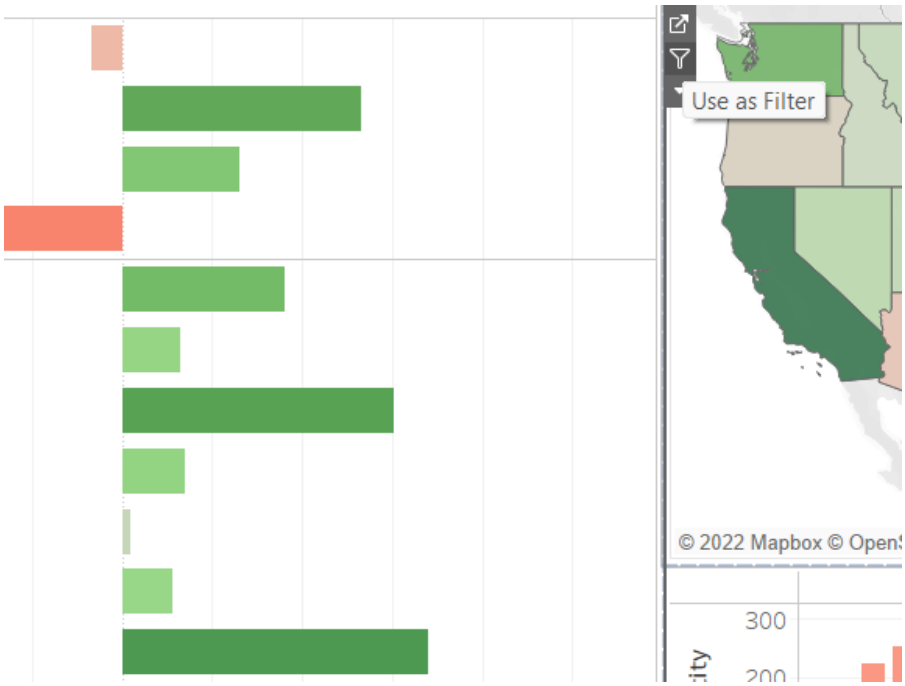



Figure 83: Updated data

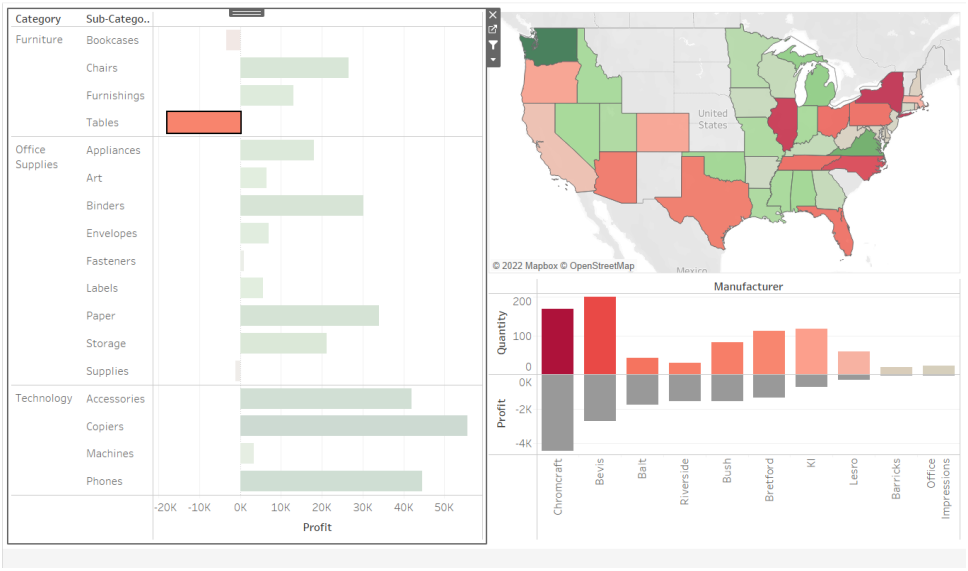



Figure 84: Updated data on other graphs

18. To edit an object, you can select the object and click the Go To Sheet button to make some changes to your viz.

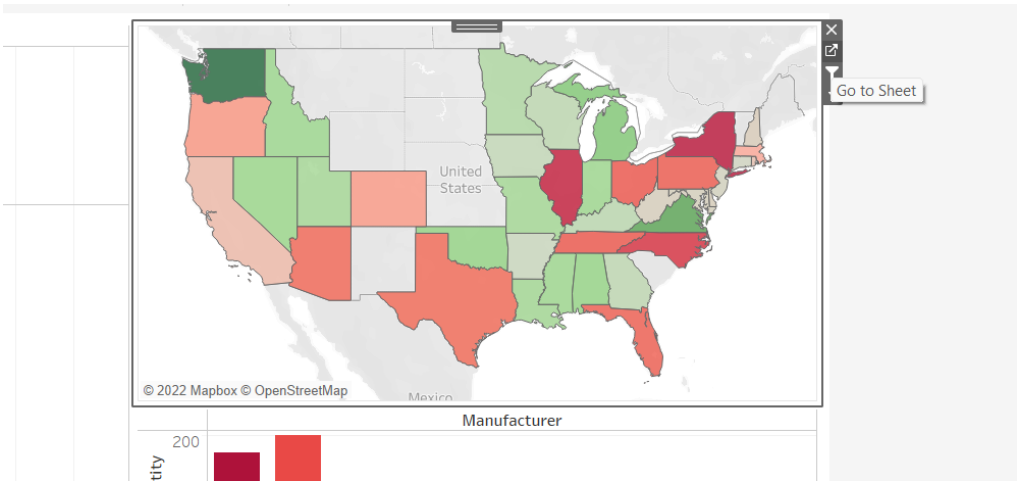



Figure 85: Edit an object

19. Add filters from the Table section in the Data pane to your viz to make it more presentable.

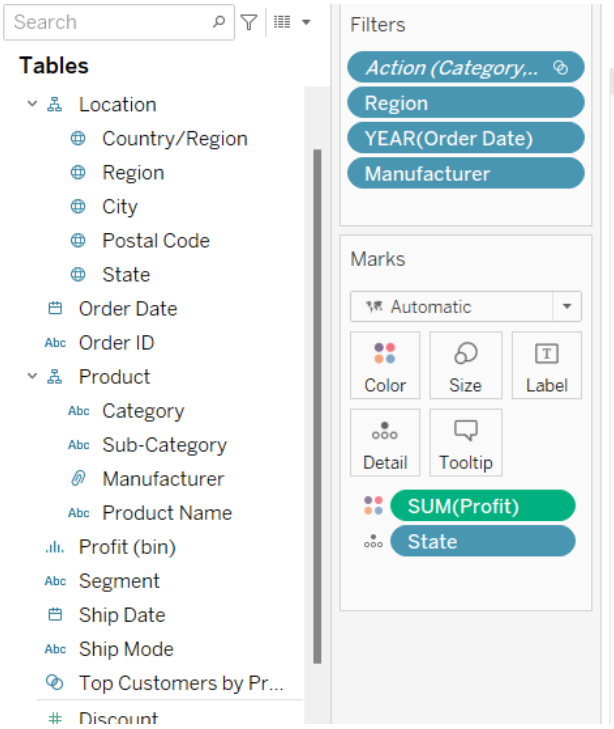



Figure 86: Add filters

20. Rename your dashboard.

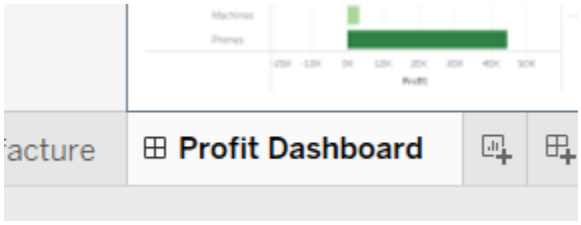



Figure 87: Rename your dashboard

21. You can add filters to adjust your viz.

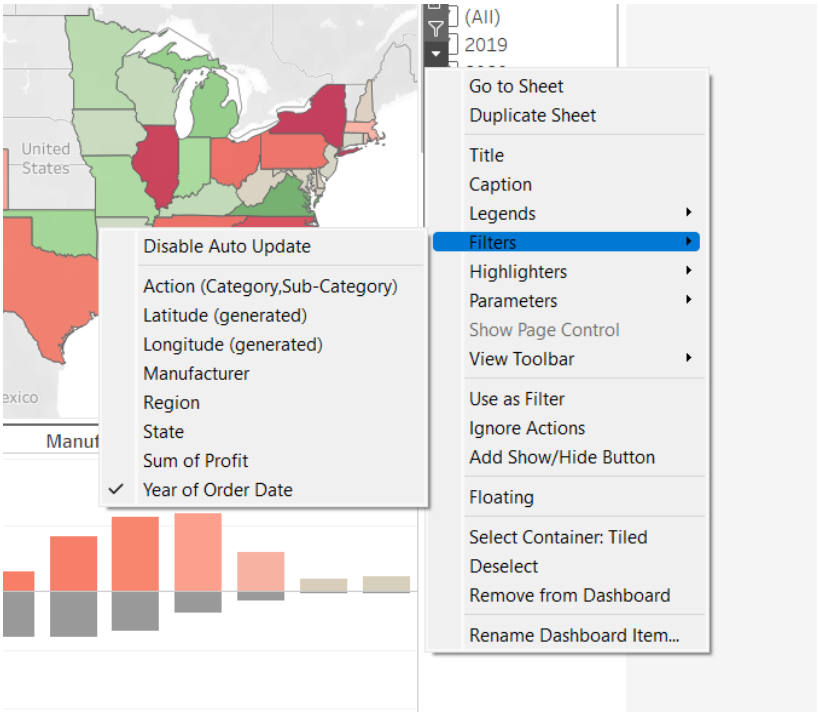



Figure 88: Adjust your viz

22. Add the title of your dashboard by drag and drop text from the Objects section to any parts of your dashboard.

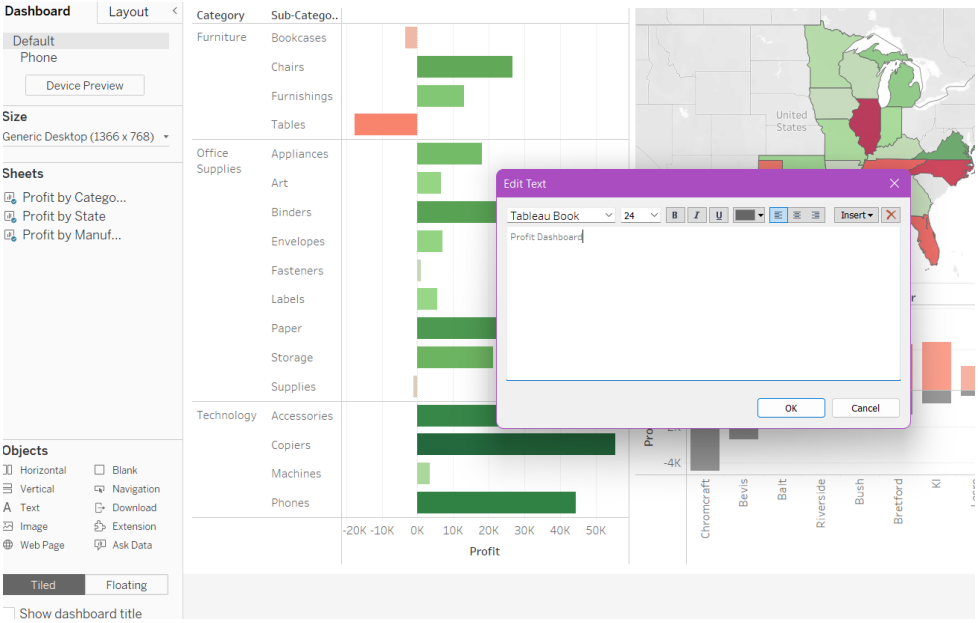



Figure 89: Add title

23. Now you can start presenting your dashboard but always make sure it always looks presentable.

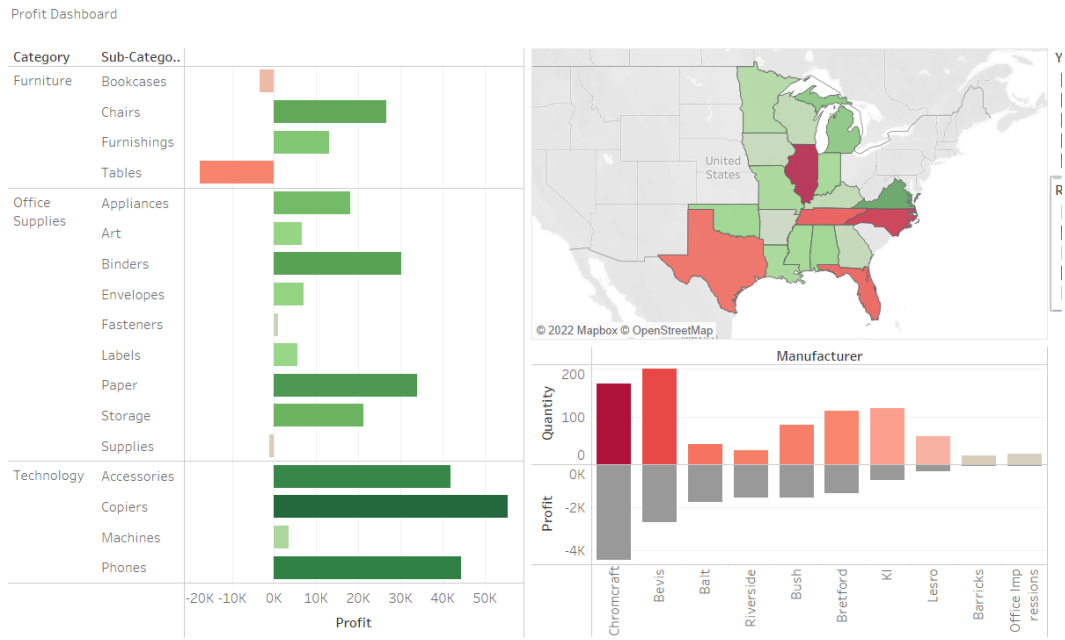



Figure 90: Presentable dashboard

## 2.3.5 Graph & Chart

### 2.3.5.1 Bar Charts

Bar charts help you to highlight differences between categories, show trends and outliers, and reveal historical highs and lows at a glance. It is useful for data that can split into many categories.

1. Go to the new worksheet.

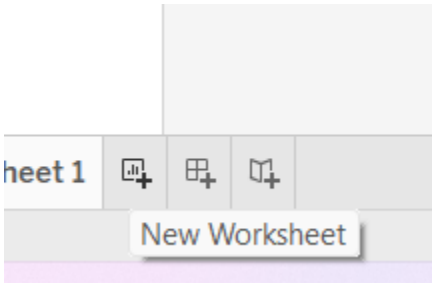



Figure 91: Create new worksheet

2. Drag tables from the Data pane, into columns and rows.

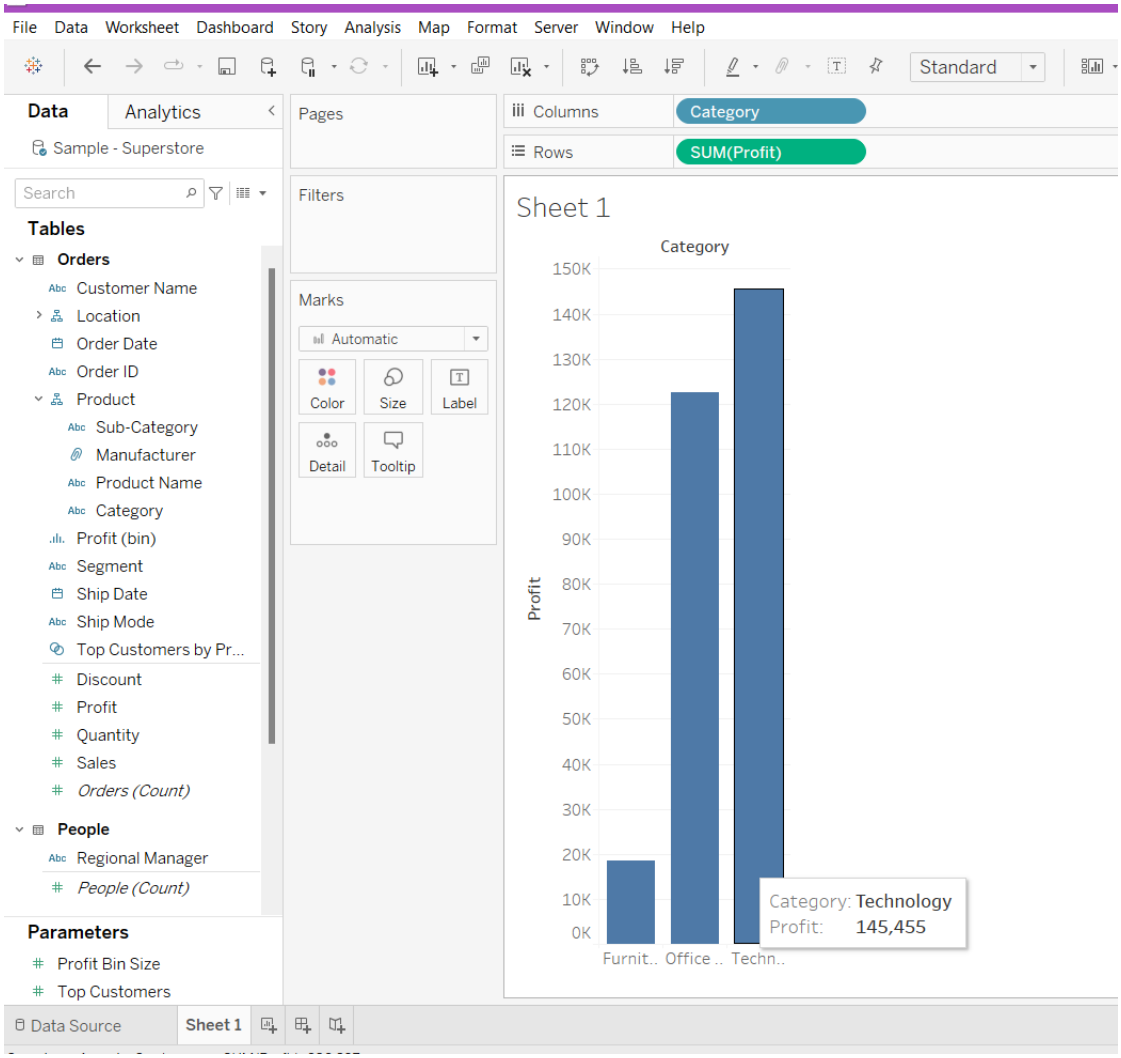



Figure 92: Drag tables from data pane

3. If the bar chart does not appear, you can go to **Marks** and select the **Bar** option.

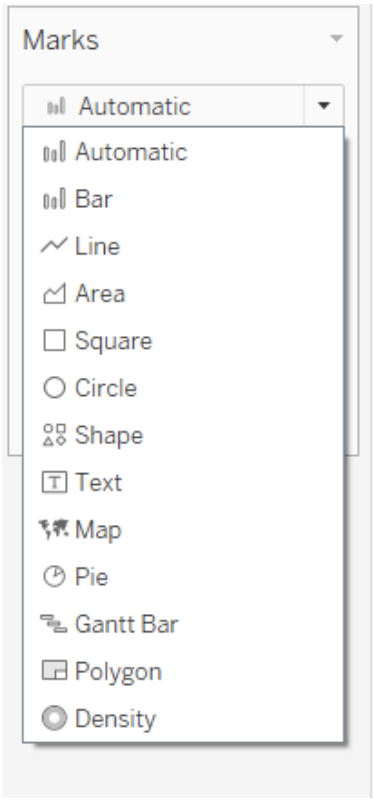



Figure 93: Select bar

### 2.3.5.2 Line Charts

Line graph connects several distinct data points to present as one continuous evolution.

1. Go to the new Worksheet.

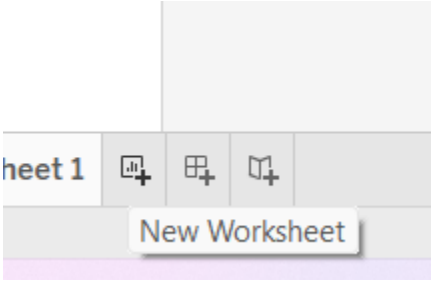



Figure 94: Create new worksheet

2. Drag tables from the Data pane, into columns and rows.

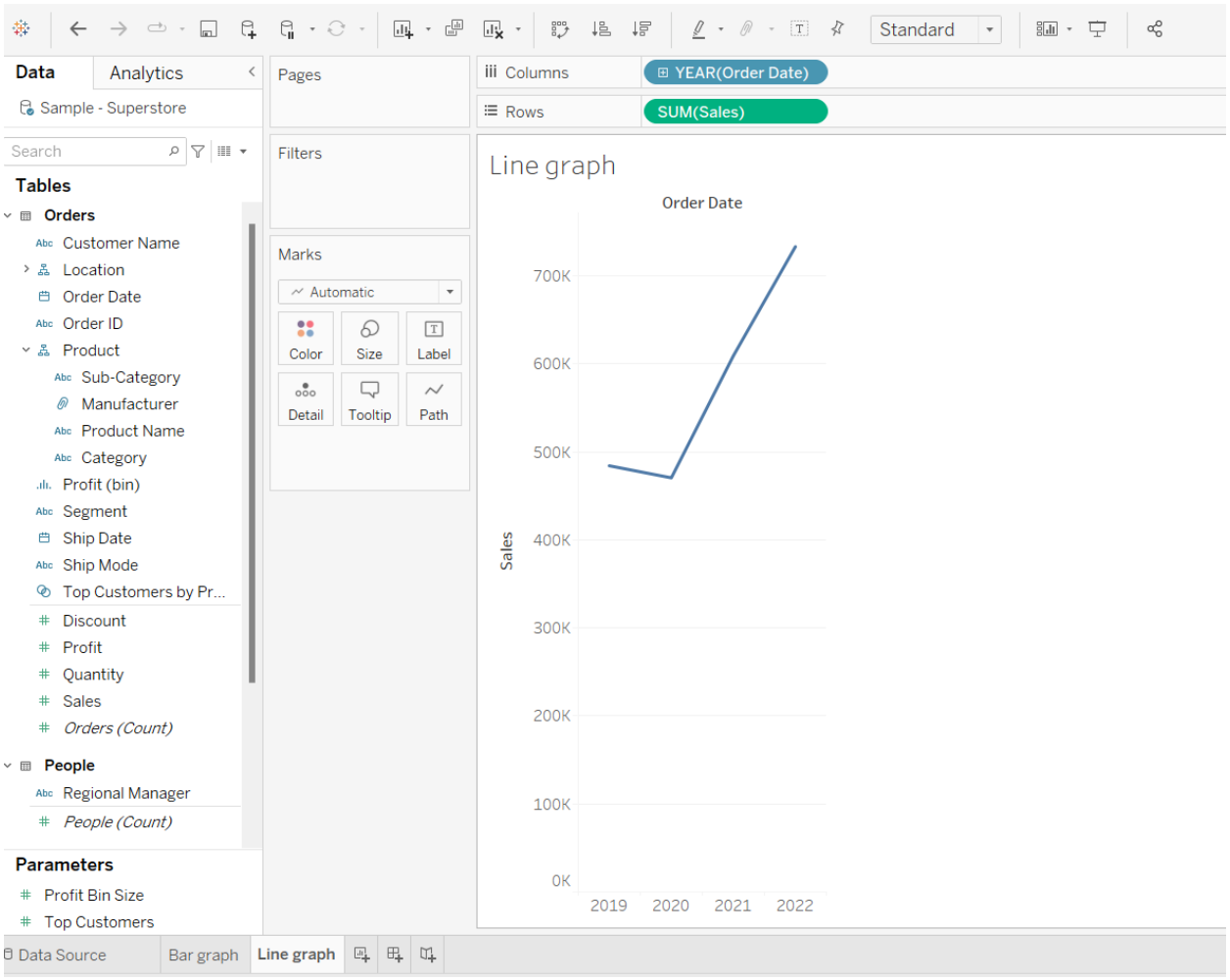



Figure 95: Drag tables from the Data pane

3. If the bar chart does not appear, you can go to **Marks** and select the **Line** option.

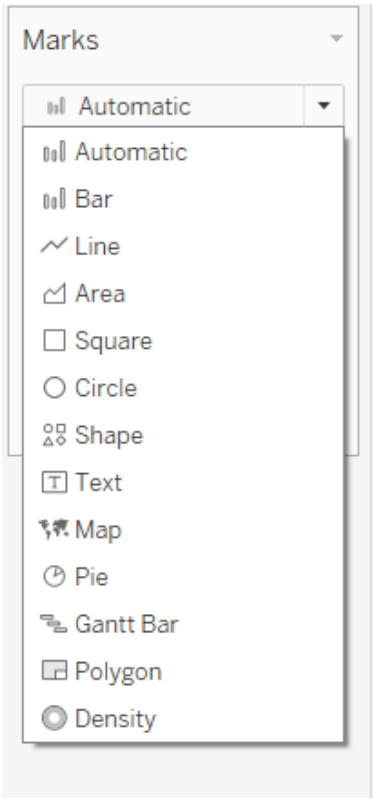



Figure 96: Select Line

### 2.3.5.3 Pareto Chart

The Pareto chart consists of both bar and line graphs. It is to identify the contribution of members present in a field. The Measure shelf values are manipulated differently.

1. Go to the new Worksheet.

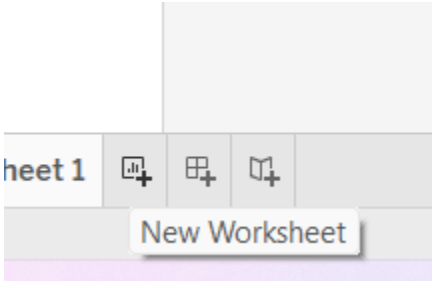



Figure 97: Create new worksheet

2. Drag tables from the Data pane, into columns and rows.

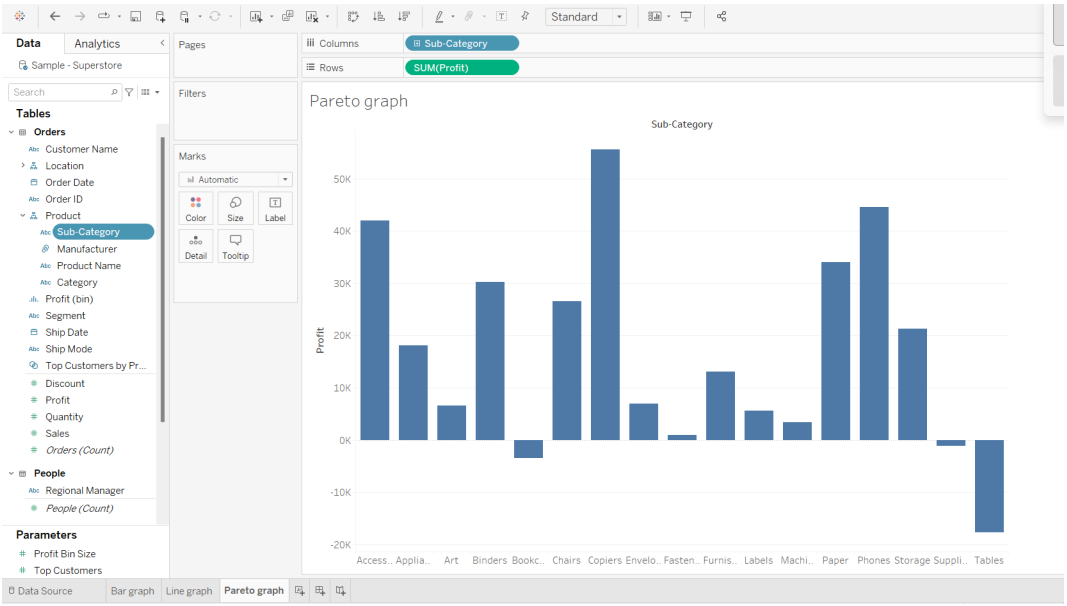



Figure 98: Drag tables from the Data pane

3. Right-click on the column's pill and choose the Sort option from the drop-down list.

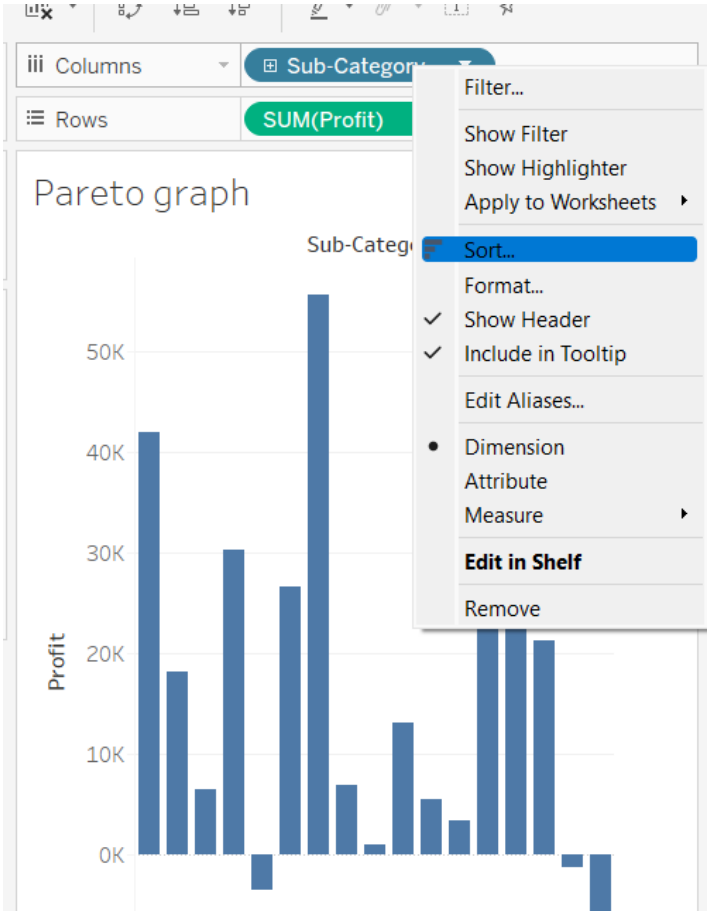



Figure 99: Click on Sort

4. A Sort window will appear.

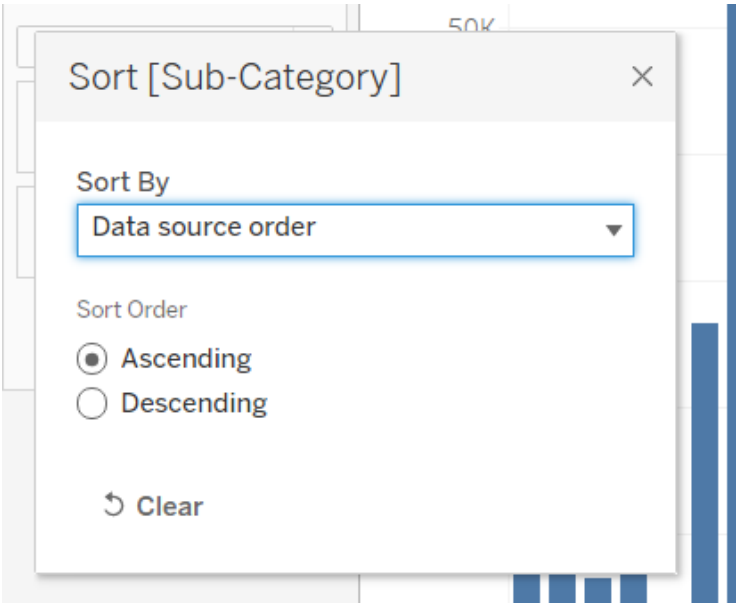



Figure 100: Sort window

5. Choose your desired option and this will appear. Choosing the best option results in the best viz of your graph.

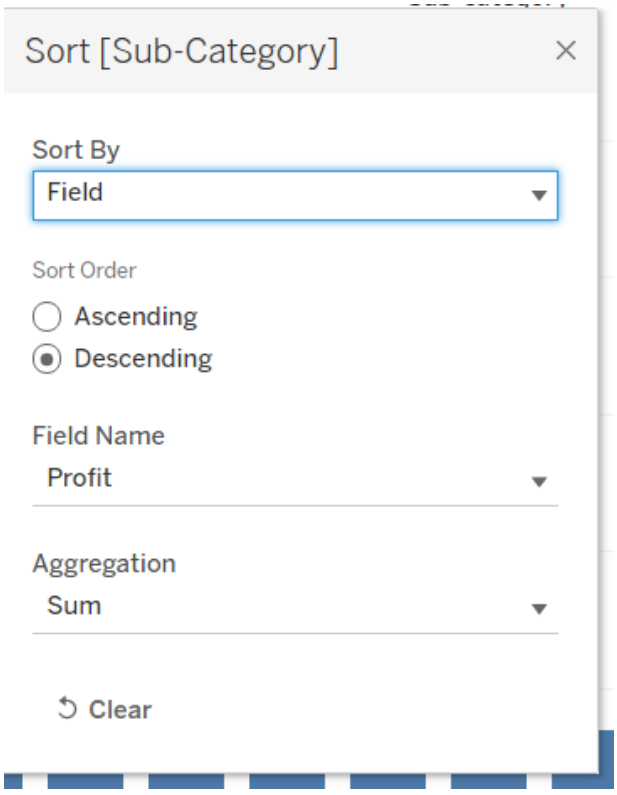



Figure 101: Choose your desired option

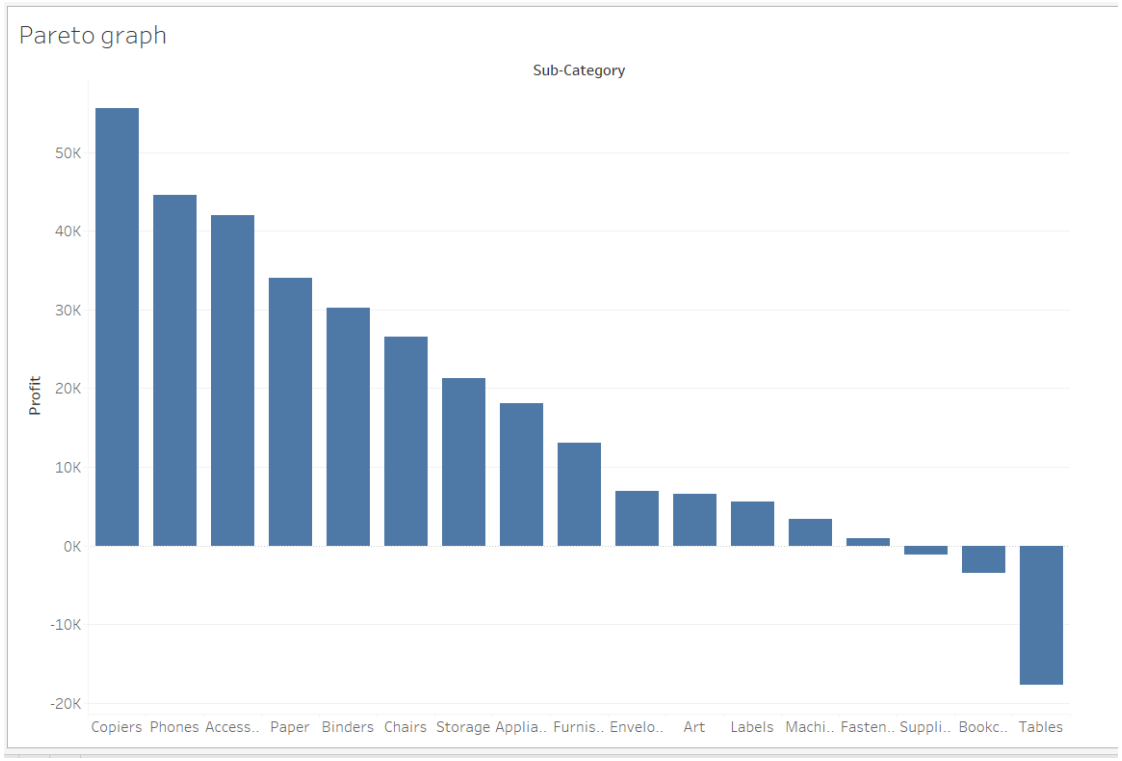



Figure 102: Bar graph

6. Selecting other pills from the Data pane into rows give you this kind of viz.

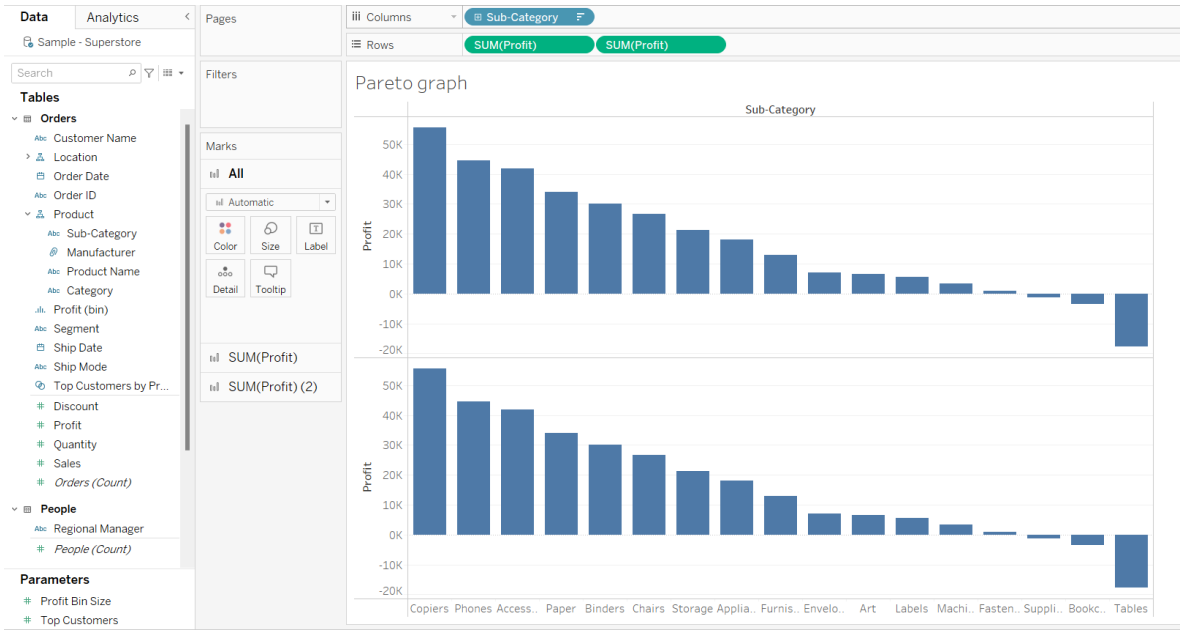



Figure 103: Selecting other pills

7. Selecting the **Dual Axis** option from the newly added data into rows gives you a dotted graph.

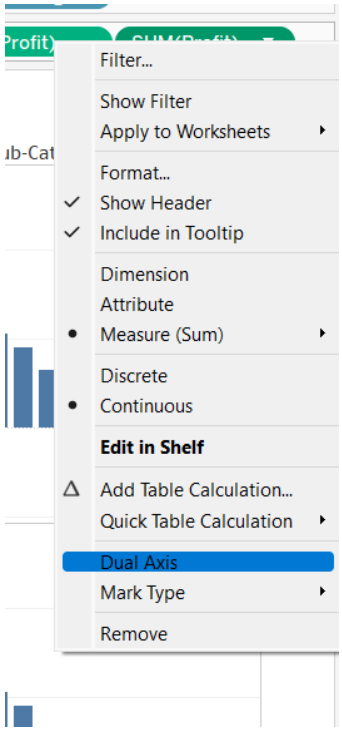



Figure 104: Click on Dual Axis

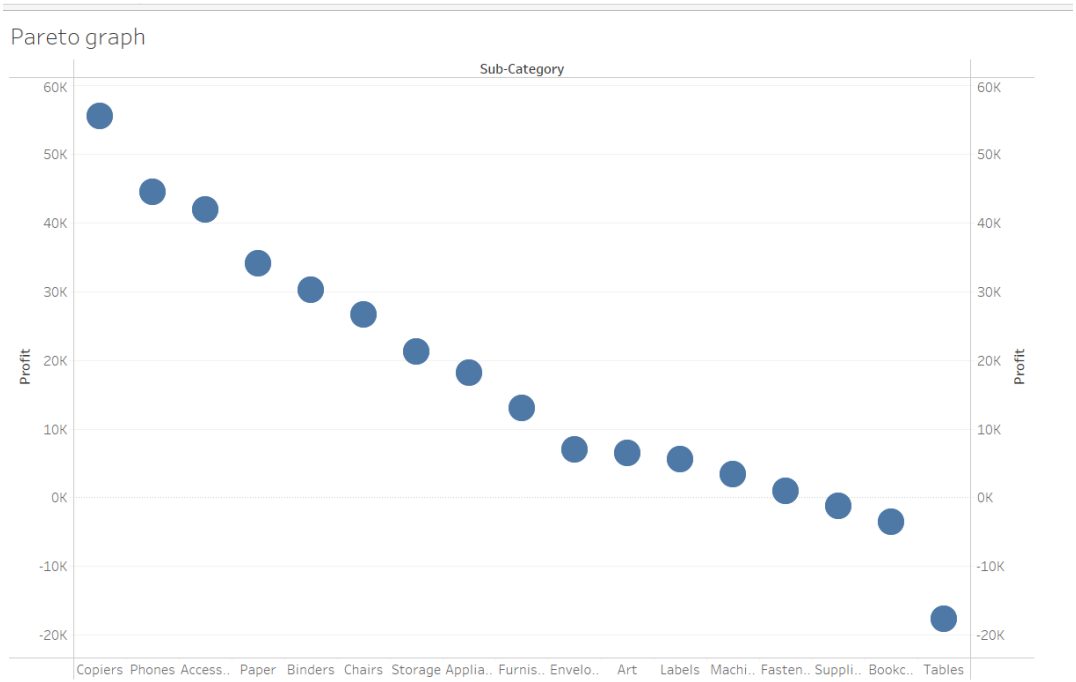



Figure 105: Plotting

9. Go to the Marks Card and choose your desired data set for your rows. Select **Bar** as chart type from the list.

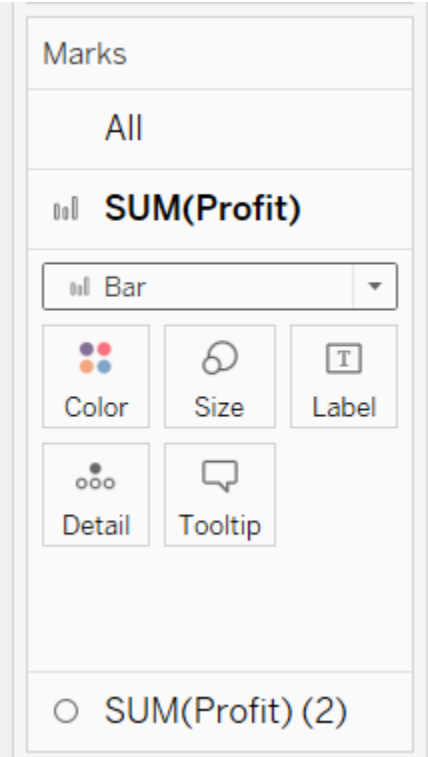



Figure 106: Marks Card

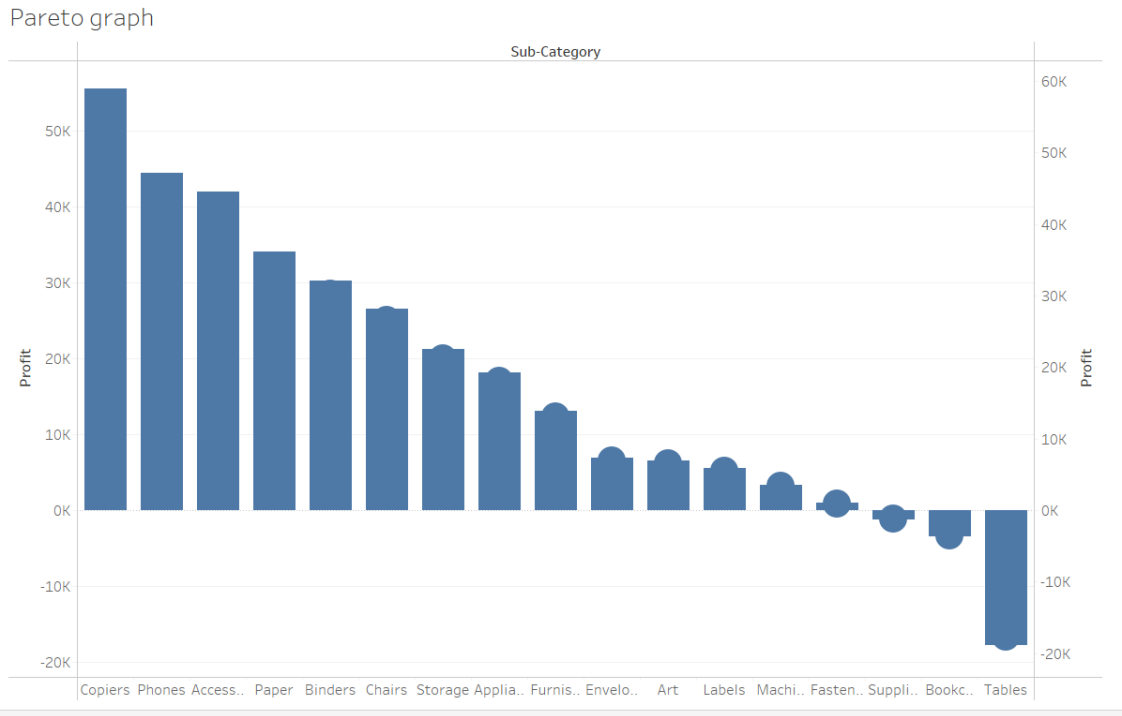



Figure 107: Bar graph

10. Choose your desired data set for your rows. Select **Line** as chart type from the list.

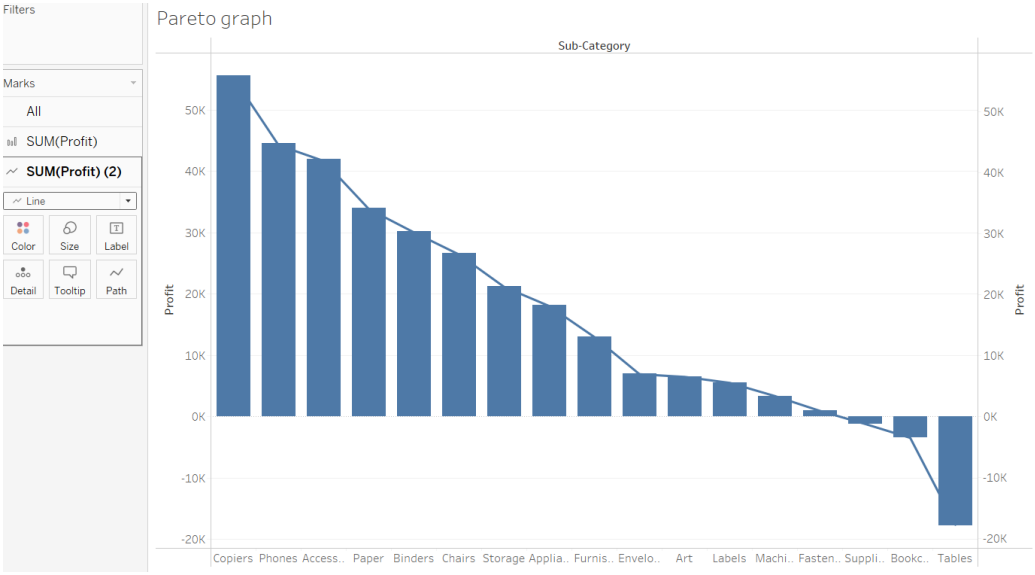



Figure 108: Select Line as chart type

11. Choose one data from rows, right click and choose **Add Table Calculation** from the list.

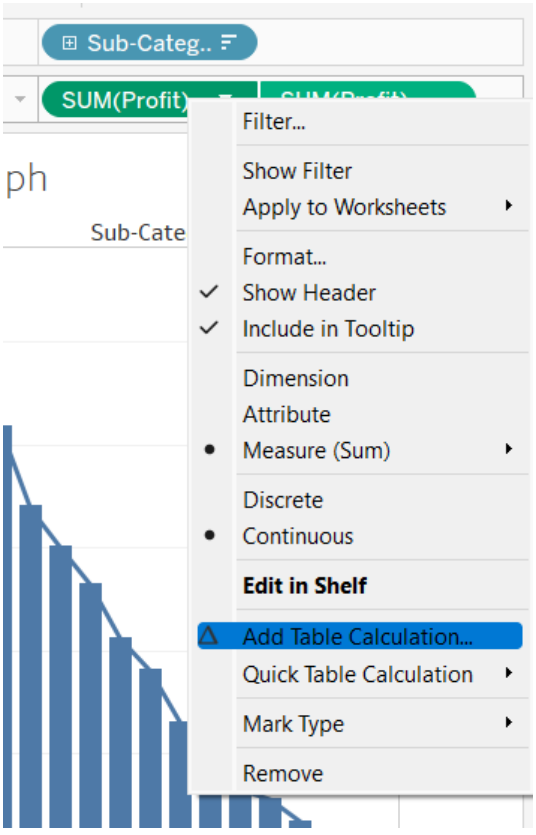



Figure 109: Click on Add Table Calculation

12. Primary Calculation Type window will appear.

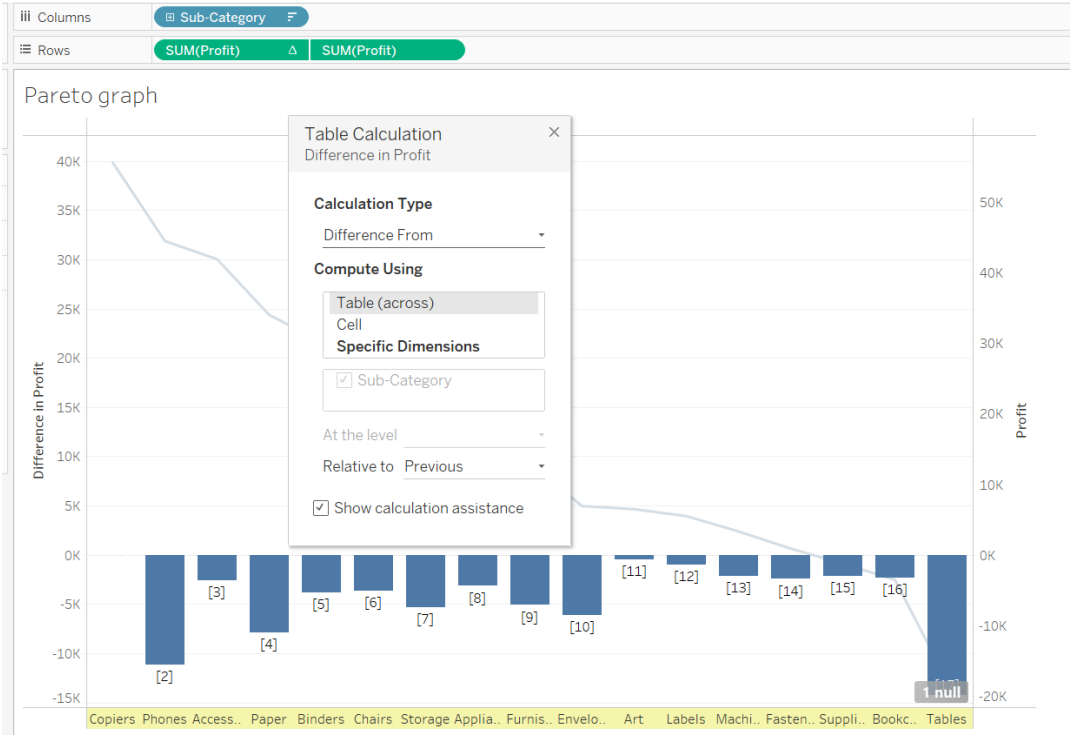



Figure 110: Primary Calculation Type window

13. Select options based on the picture below and close the window.

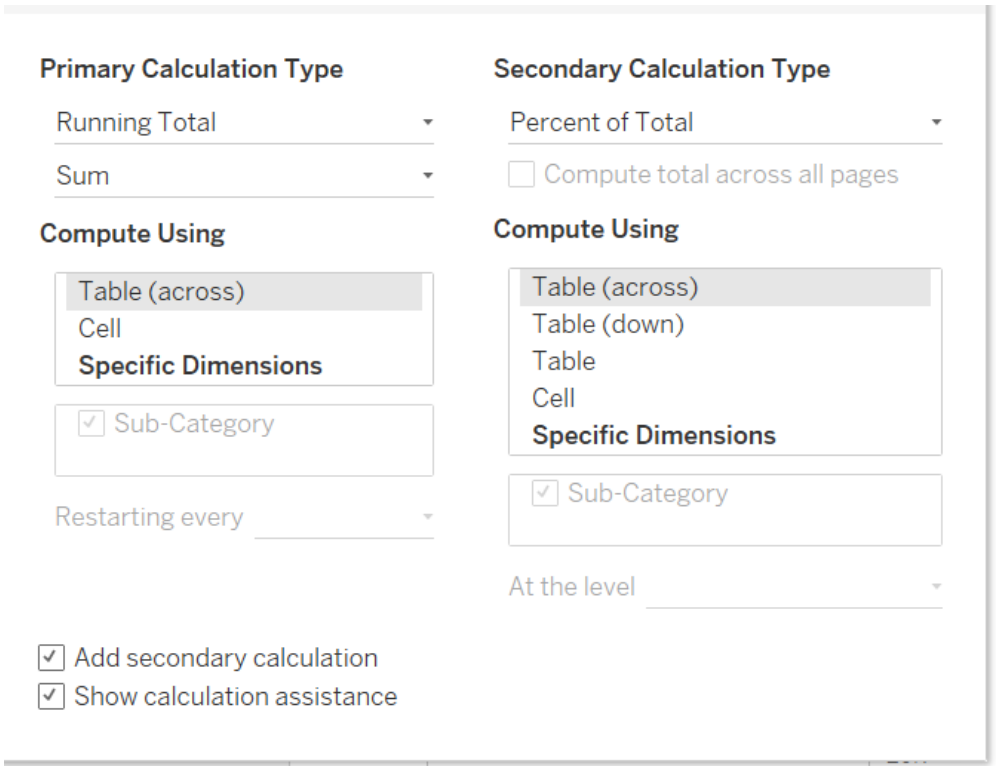



Figure 111: Select your option

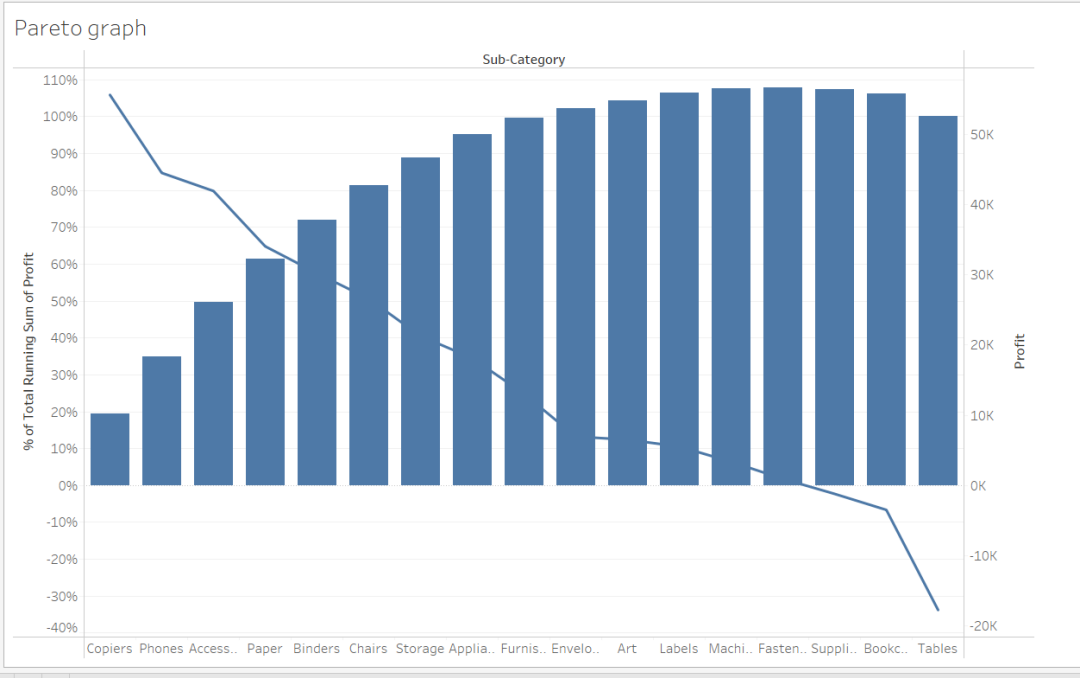



Figure 112: Pareto graph

14. Go to the last Marks card and click on the Color icon. Choose your desired color to represent your Line graph.

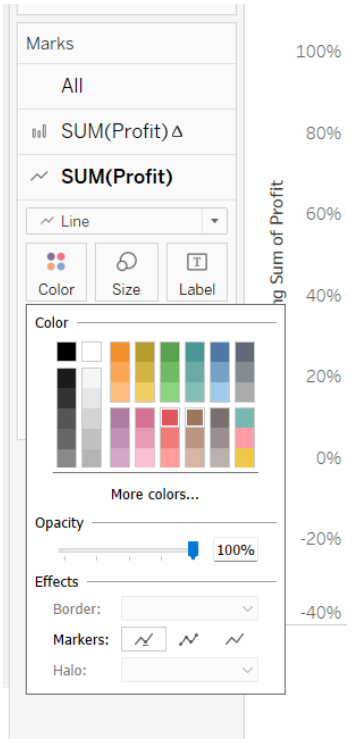



Figure 113: Choose color for your line chart

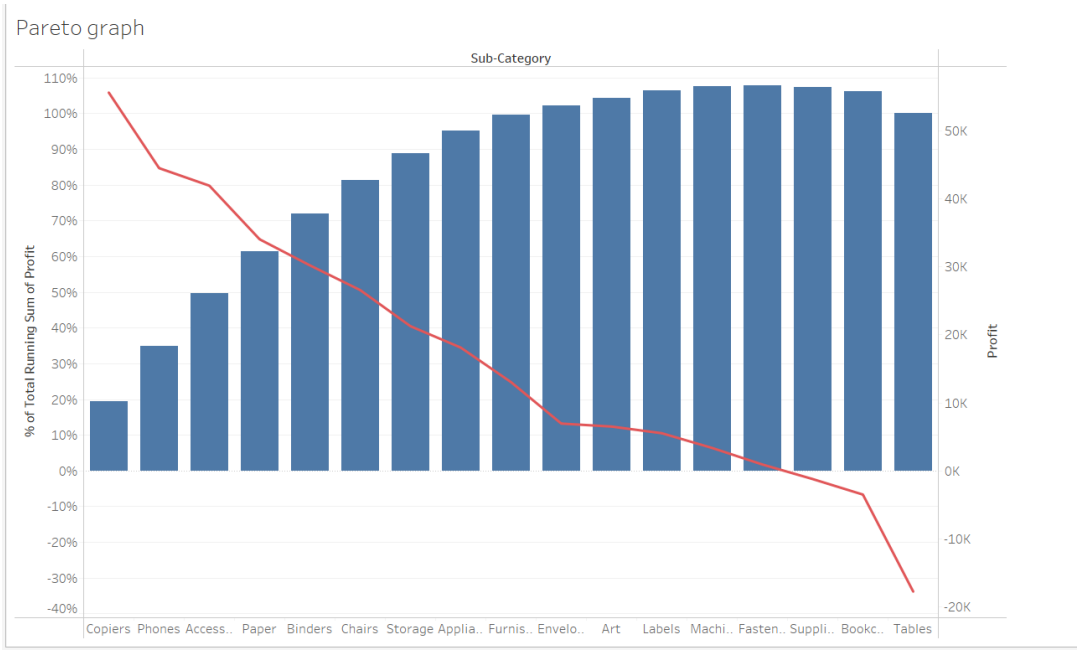



Figure 114: Final result of Pareto graph

## 2.3.5.4 Area Chart

Area charts represent any quantitative data over time.

1. Go to new Worksheet
2. Hold the control key at your keyboard and select data from the Data pane.

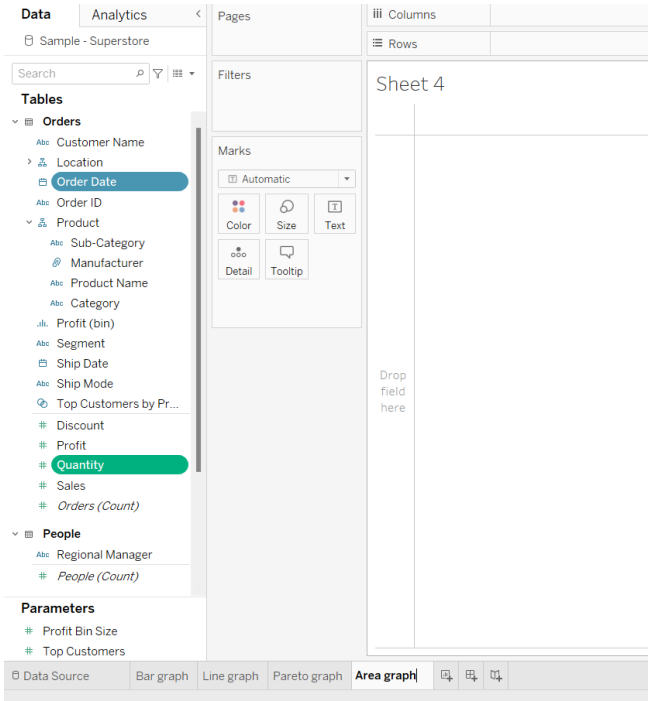



Figure 115: Select data

3. Select the **Show Me** option and choose the **Area Chart** icon.


Figure 116: Choose Area chart icon

4. Select data again from the dimension pane and add Color to it from the Marks card.

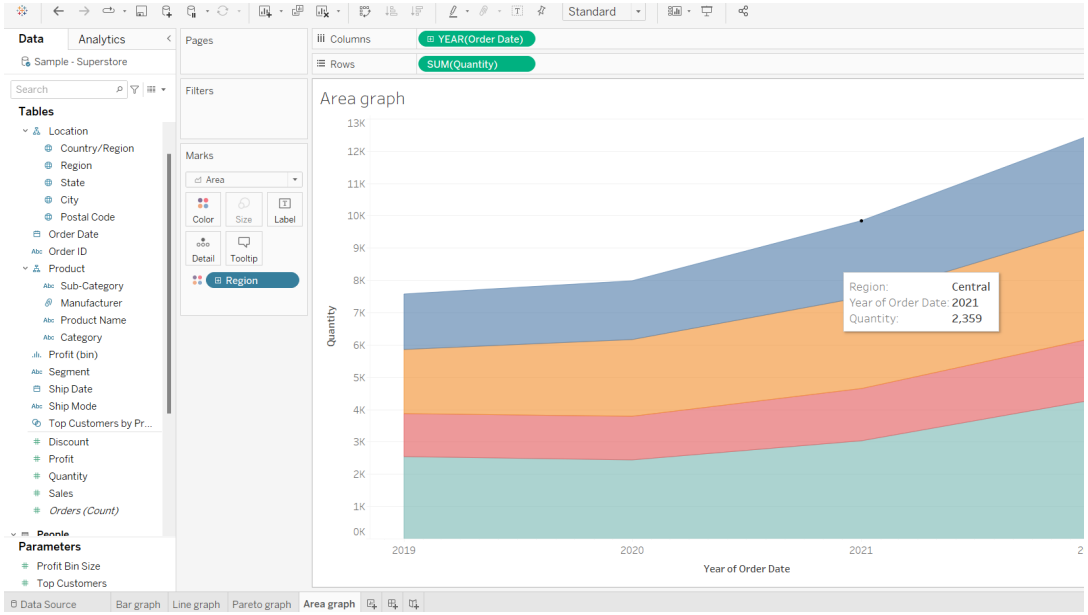



Figure 117: Area chart

## 2.3.5.5 Histogram

Histograms show how your data is distributed across distinct groups. Grouping your data into categories (bins) and plotting the number of records in a category as a vertical bar, you can see where the majority of your data falls in.

1. Go to a new Worksheet.
2. Select the data form measures pane and click on histogram from the Show Me button.

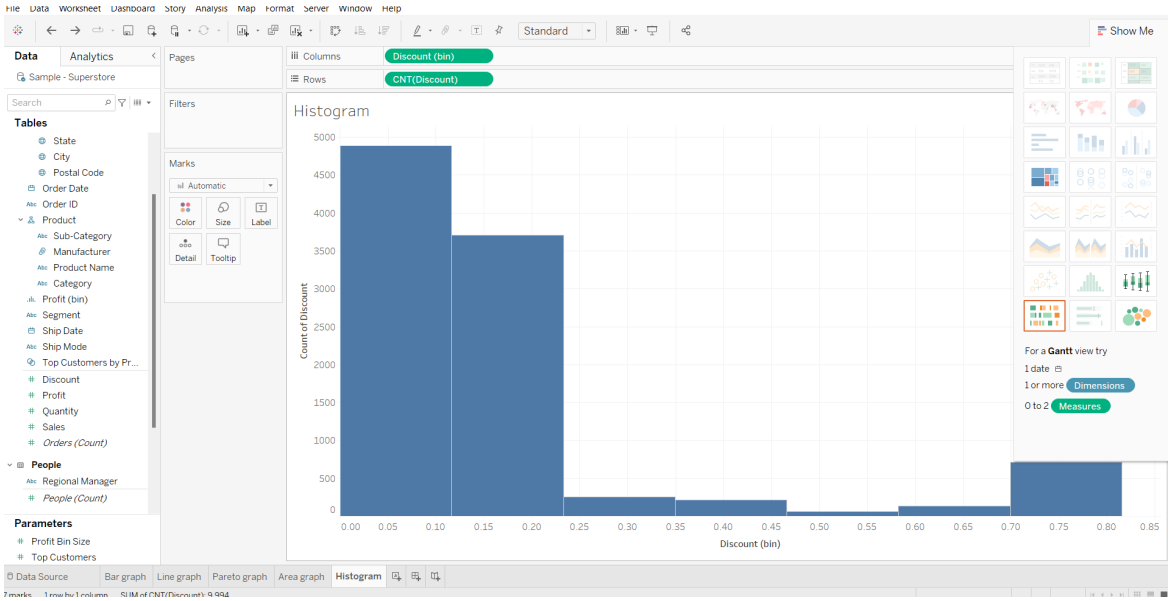



Figure 118: Click on histogram

### 2.3.5.6 Pie Charts

Pie charts are useful to add detail to other viz. Angle of the pre determines the measured values. Different colors represent members in a dimension.

1. Go to a new Worksheet.
2. Select data while pressing the control button on your keyboard from the data pane and click on Show Me button, choose Pie Chart.

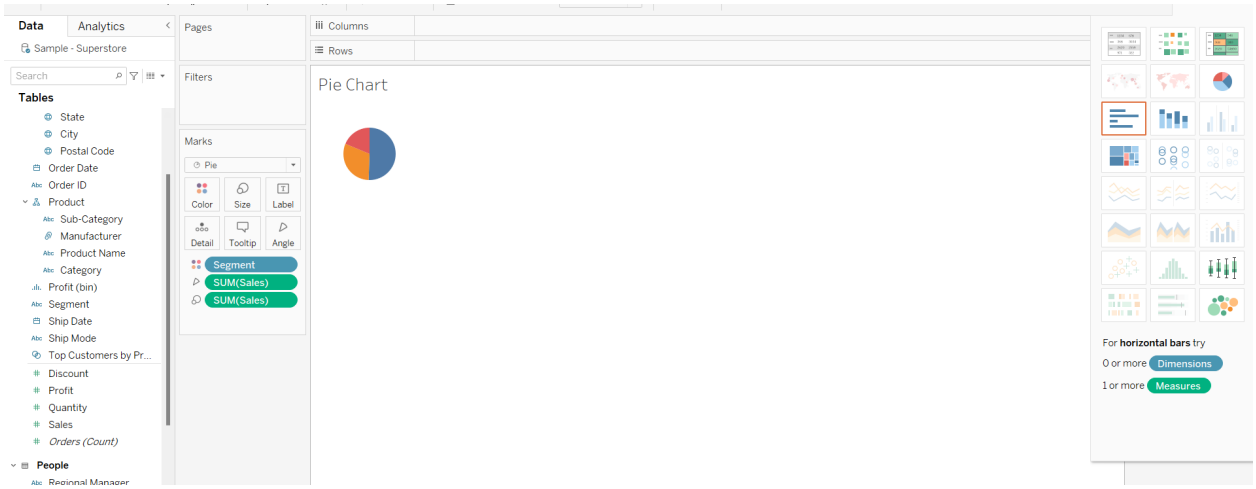



Figure 119: Click on pie chart

### 2.3.5.7 Tree Maps

Tree maps relate different segments of your data to the whole. It shows how individual data points fit in a hierarchy. Each rectangle subdivides into smaller rectangles, or sub-branches, based on the whole proportion. It also best shows what categories most of the data falls in.

1. Go to the new Worksheet.
2. Drag and drop the measure pane's data to the Marks card two times. One for color, another one for size.

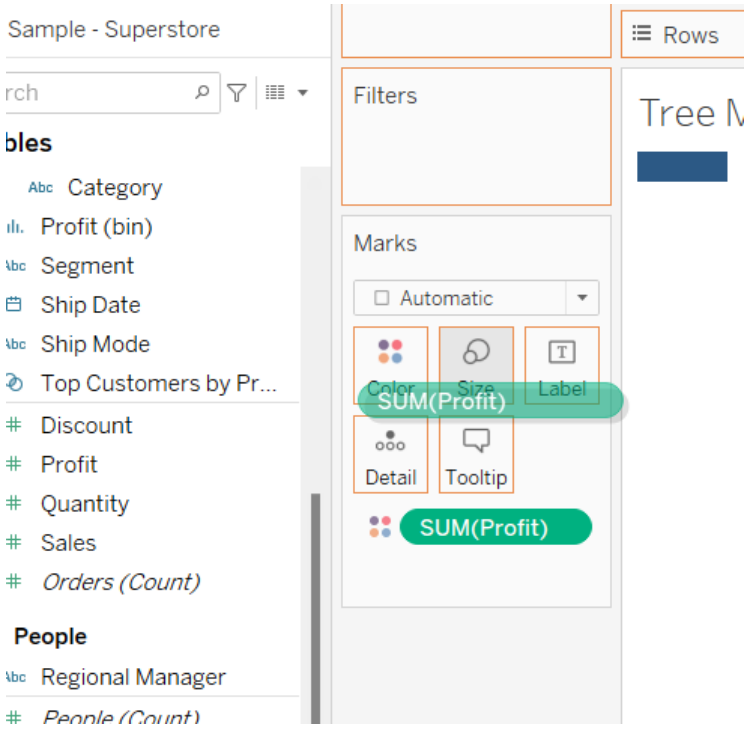



Figure 120: Select variable to drag

3. Drag and drop from dimesion's pane data to the Label shelf at Marks card and choose the chart type as TreeMap from the Show Me button.

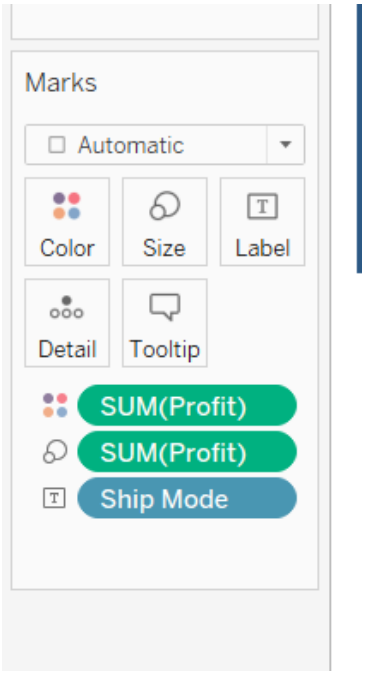



Figure 121: Drag variable to Marks card

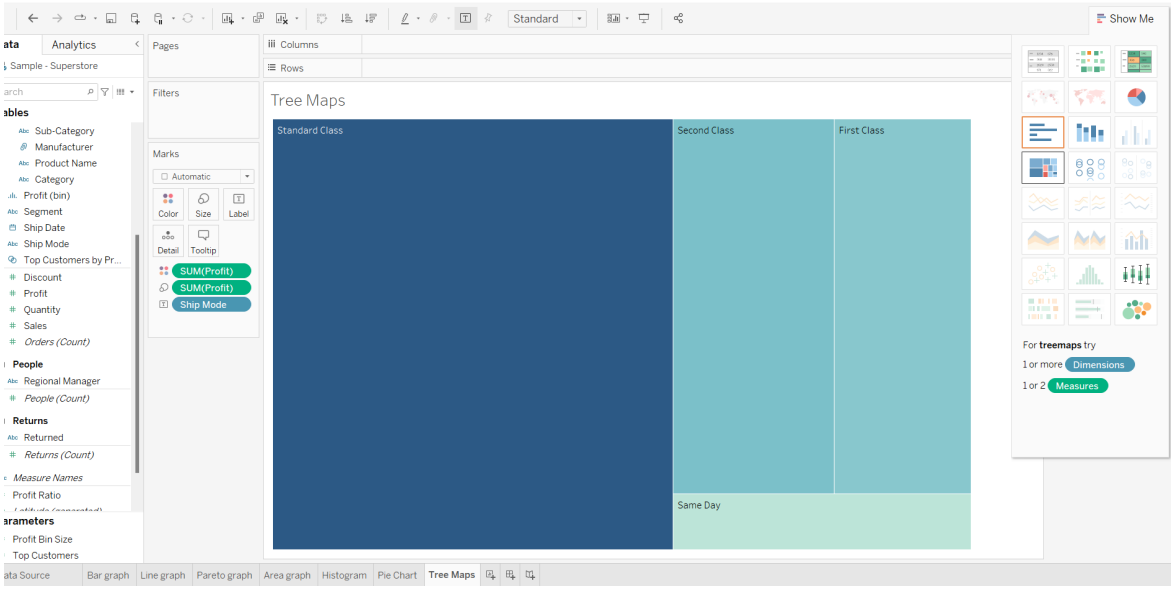



Figure 122: Tree Map

### 2.3.5.8 Scatter Plots

Scatter plots can give you a sense of trends, concentrations, and outliers that lead to a deeper investigation of your data. It presents lots of distinct data points on a single chart. It shows if one variable is a good predictor to the other, or if they tend to change it independently.

1. Go to the new Worksheet.
2. Drag data from dimension to the columns and measure pane to the rows and it will create a scatter plot by default.

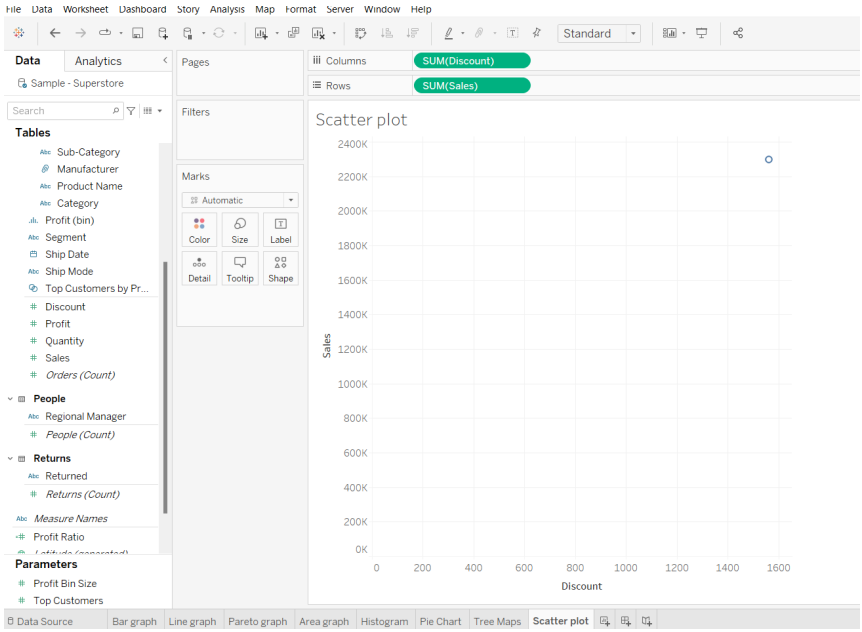



Figure 123: Select variables into columns and rows field

3. Select another data from the dimension pane into the Color icon at the Marks card.

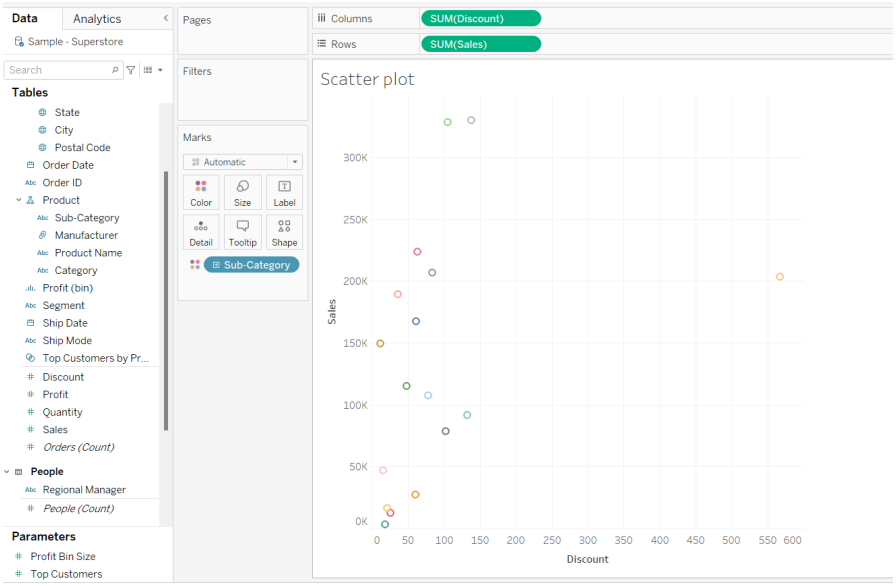



Figure 124: Drag data to Color icon

## 2.3.5.9 Bubble Charts

Bubble charts add great detail to scatter plots or maps. Varying size and color creates more compelling charts that represent large volumes of data at once. Bubbles add more detail to the traditional two-axis chart, highlighting the relationship between three or more variables, without overwhelming the viewer.

1. Go to a new Worksheet.
2. Hold the Control Key while clicking on data from the Data pane.
3. Click the Show Me option and choose the Bubbles icon.

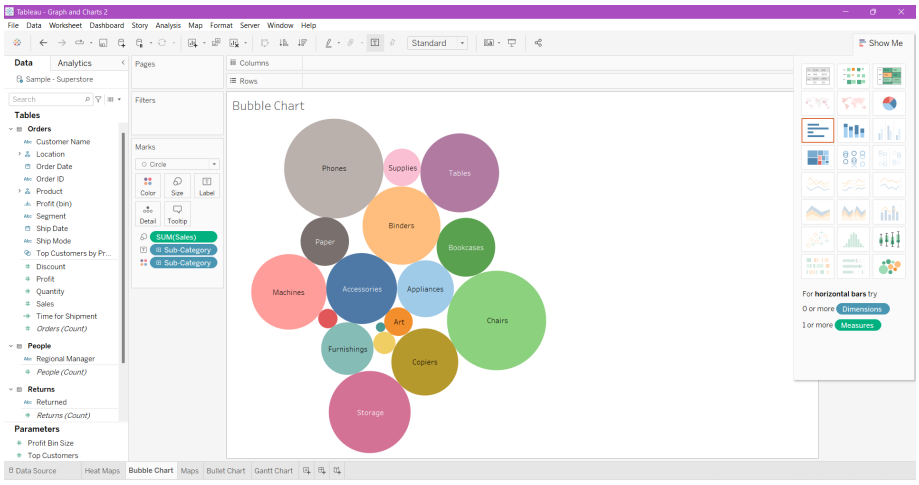



Figure 125: Click on Bubble icon on Show Me

### 2.3.5.10 Heat Maps

Heat maps are a great way to compare data across two or more categories using color. It shows where the intersection of categories is strongest and weakest. It can pack hundreds of comparisons into a small area.

1. Go to a new Worksheet.
2. Hold the Control key while selecting data from the Data pane and choose Heat Map icon from the Show Me button.

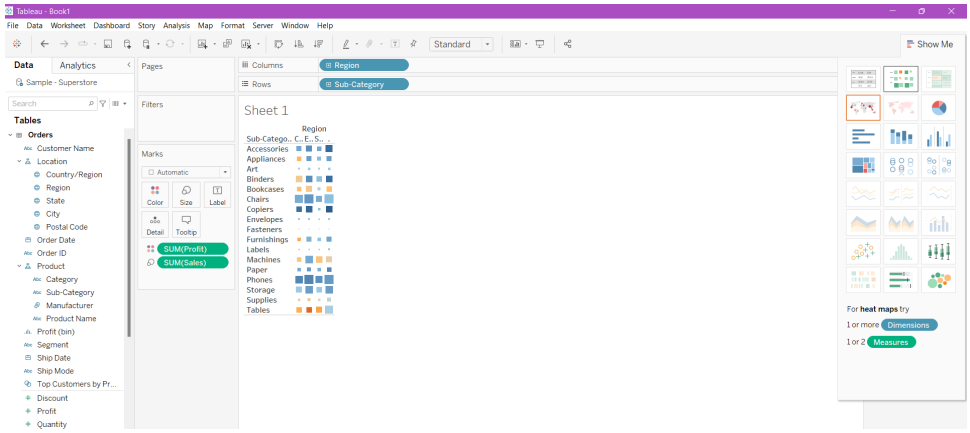



Figure 126: Heat Map

### 2.3.5.11 Maps

Maps highlight geographic trends in a format which everyone knows and understands. It is a simple and compelling way to show how location correlates with trends in your data.

1. Go to a new Worksheet.
2. In the Data pane, select data and the third data to the Size on the Marks card.

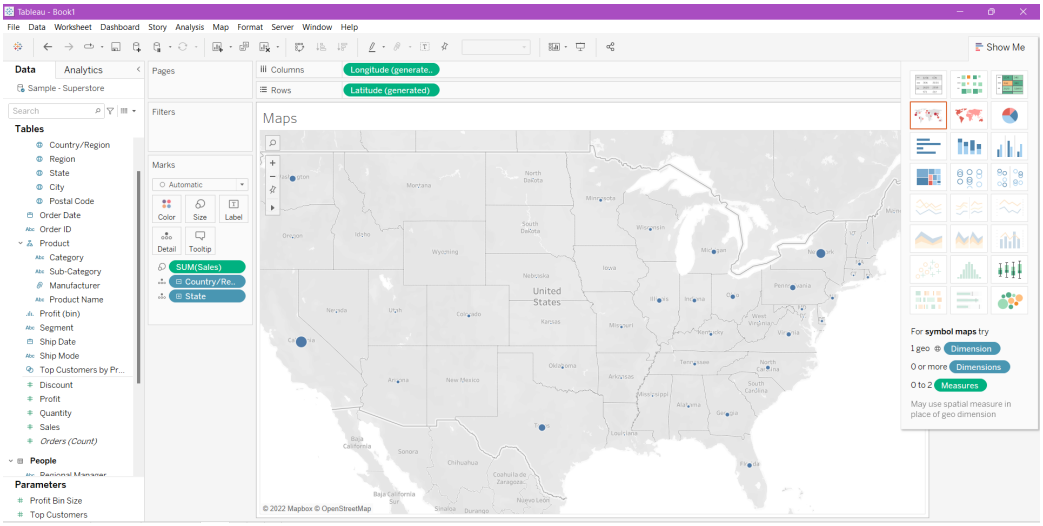



Figure 127: Choose map

3. From here, choose **Map**, and click on **Background Layers**.

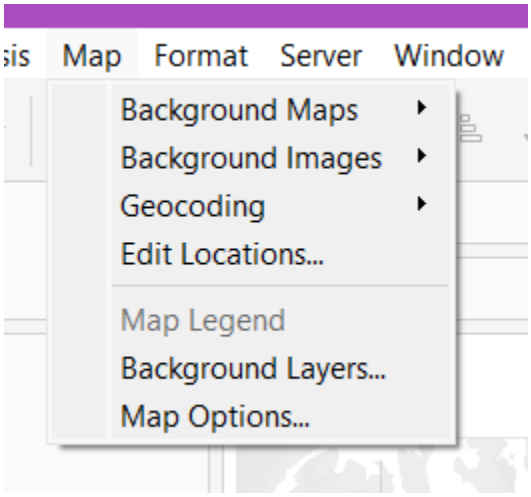



Figure 128: Click on background layers

4. From the Map Layers pane, choose your desired style and Map Layers or follow the picture down here.

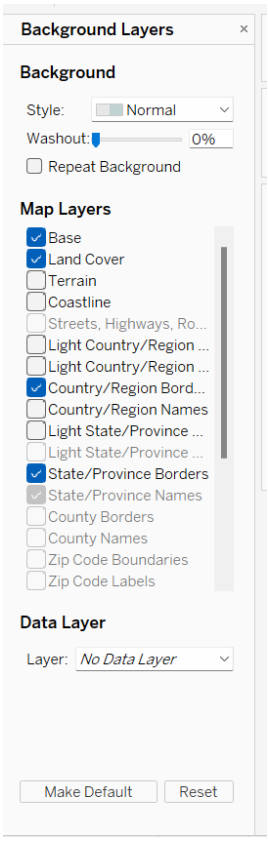



Figure 129: Choose style

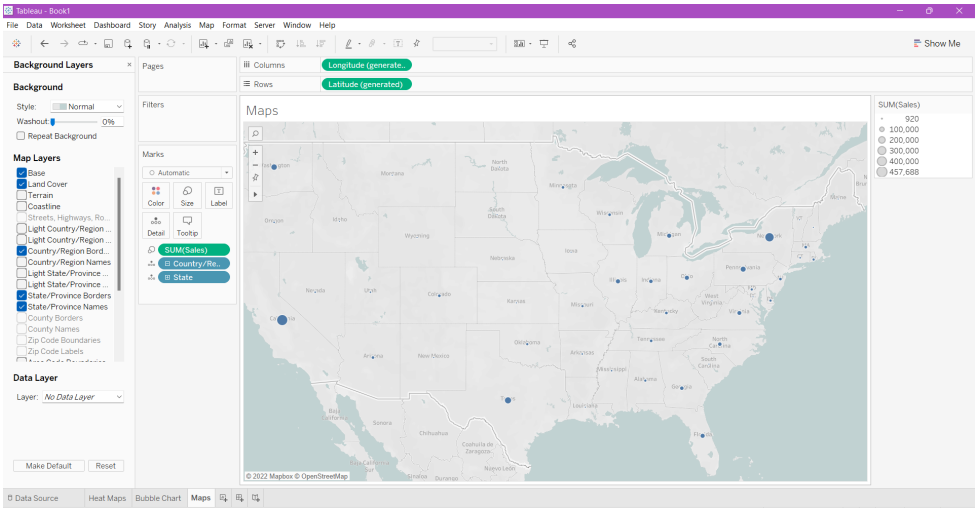



Figure 130: Map

## 2.3.5.12 Bullet Charts

Bullet charts comparing measures to show goal's progress. It is a variation of a bar chart.

1. Drag and drop Dimension from the Data pane into the column shelf.
2. Drag and drop Measure from the Data pane into the rows shelf.
3. Press Control Key while clicking on data from the Measure data pane and choose bullet charts from the Show Me button.

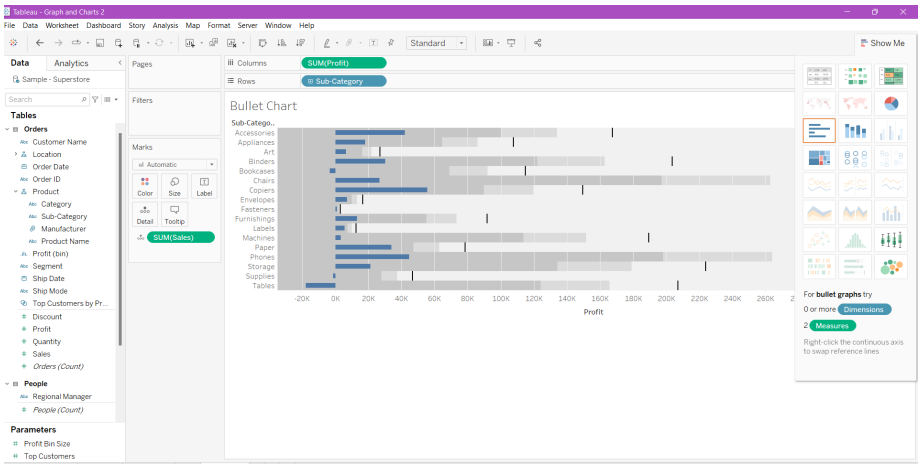



Figure 131: Bullet Chart

## 2.3.5.13 Gantt Charts

Gantt charts illustrate the start and finish dates of steps or the deadlines (time series).

1. Go to the new Worksheet.
2. Click on the drop down button in the Mark card and choose Gantt Bar.

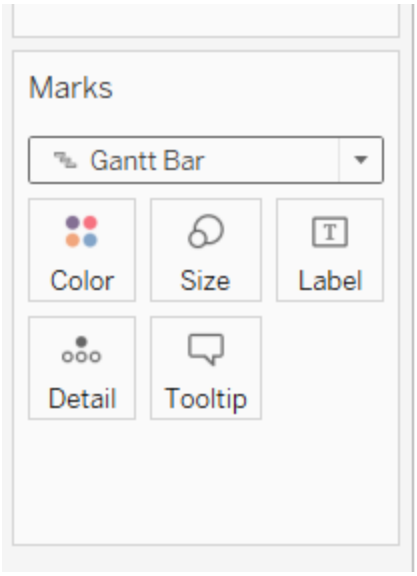



Figure 132: Choose Gantt Bar

3. Drag data about Date into columns.
4. Right click the data in the column and select the Day.

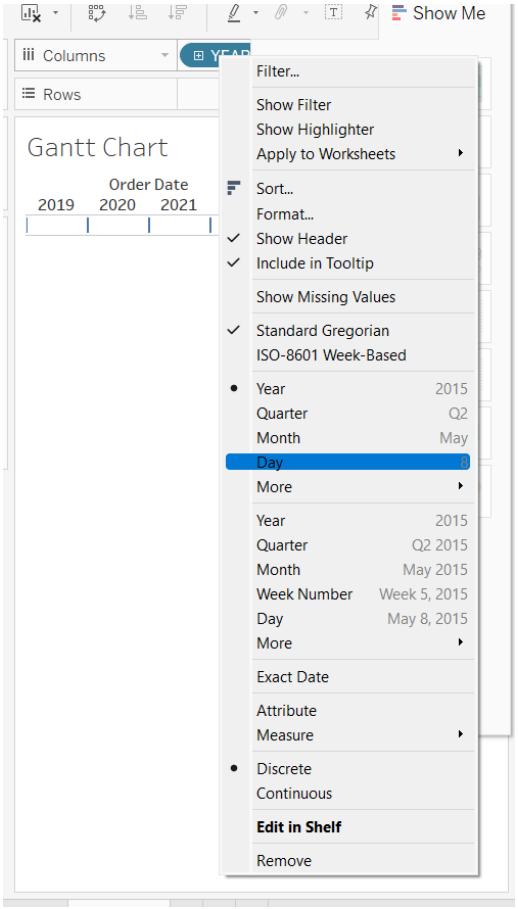



Figure 133: Select filters

5. Click on Analysis in the Menu bar and choose **Create Calculated Field**.

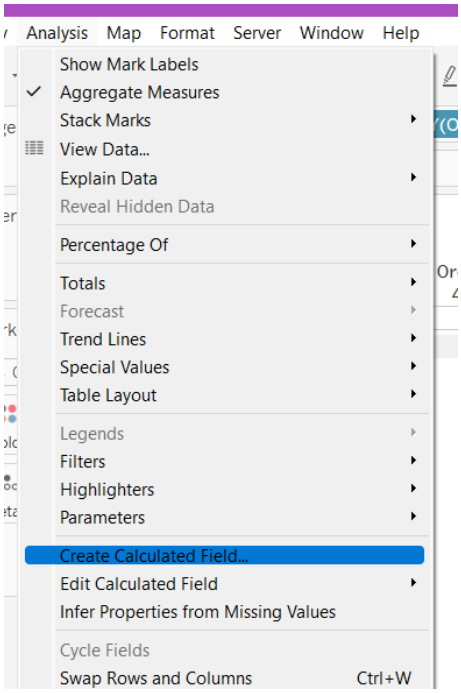



Figure 134: Click on the create calculated field

6. Enter the name of the calculated field and enter your formula.

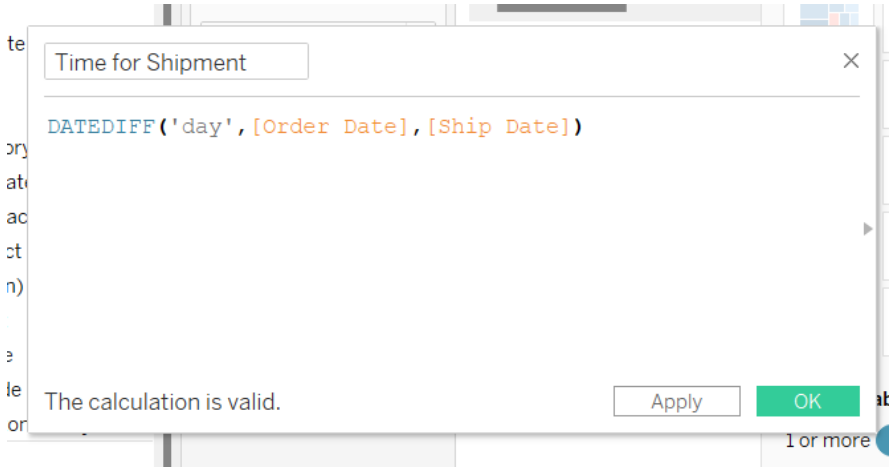



Figure 135: Calculated field dialog box

7. Click on OK.
8. Drag data into Row, and a calculated field created just now into a size icon at the Marks card.

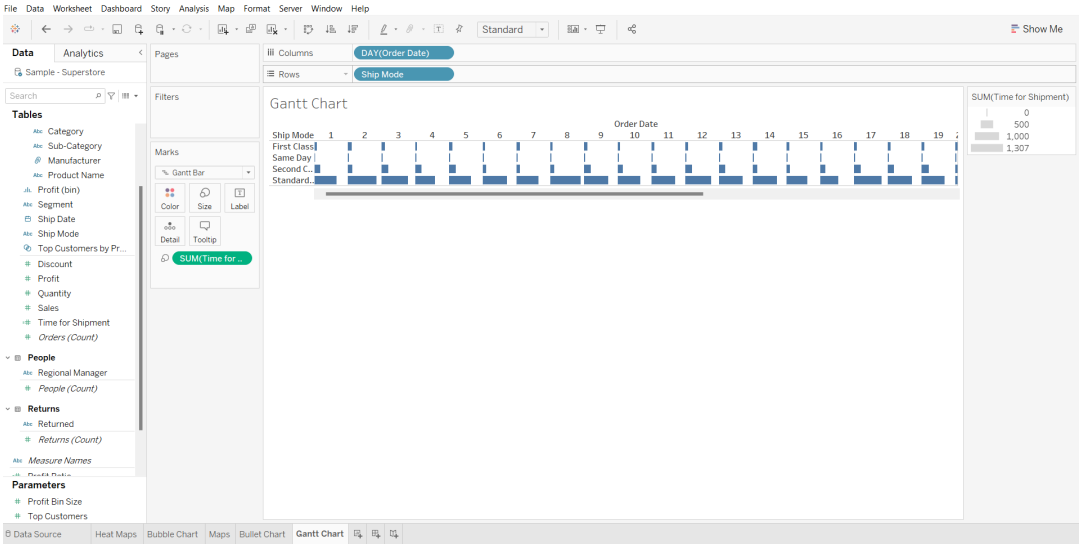



Figure 136: Gantt Chart

## 2.3.5.14 Box and Whiskers Plots

Box-and-whiskers plots (boxplot) are a common way to show data distribution. The **box** holds median, 1st and 3rd quartiles, and the **whiskers** represent data 1st and 3rd quartiles. Whiskers also can show the maximum and the minimum points in the data. Measure skewed and outliers also could be represented.

1. Create a new Worksheet.
2. Drag data into Columns (dimension) and Rows (measure). Here, a vertical axis and a bar chart is displayed by default.
3. Drag another data into Columns. Now you can see a two-level hierarchy of dimensions

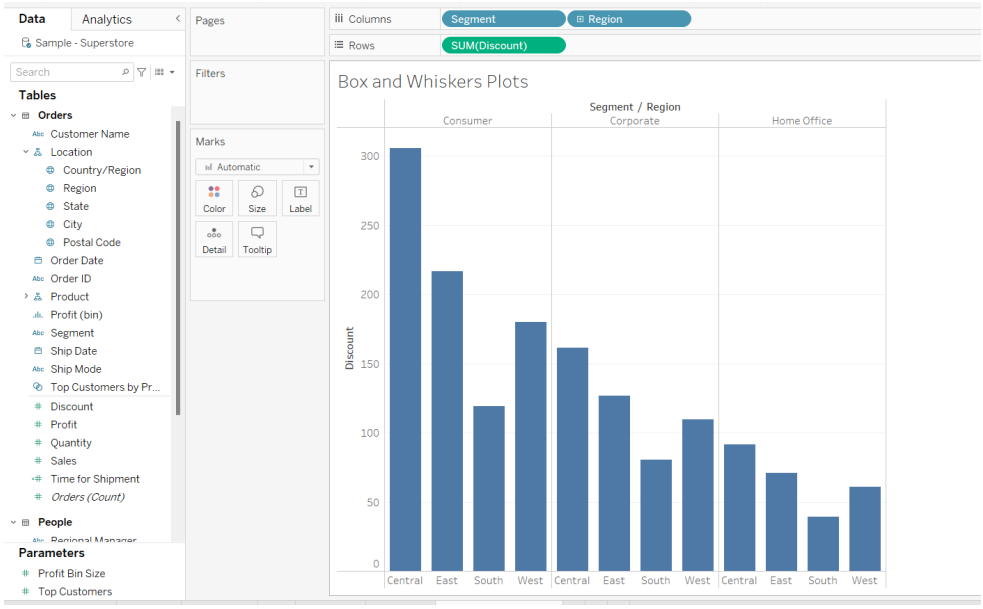



Figure 137: Default bar chart

4. Click Show Me and select box-and-whiskers plot chart type.


Figure 138: Choose box-and-whiskers chart

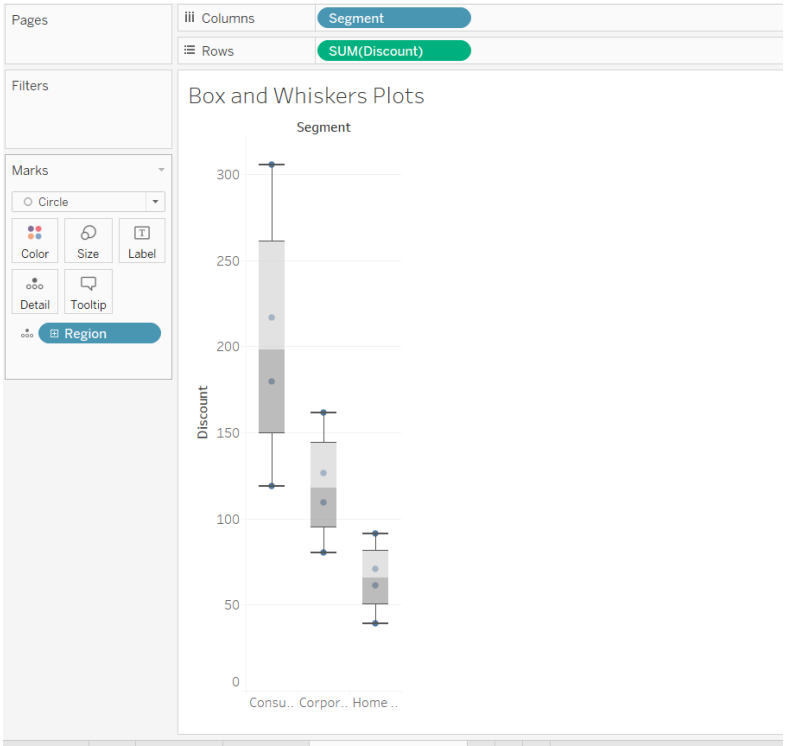



Figure 139: Display a Box-and-whiskers Chart

5. Drag back data from Marks card back to Columns.

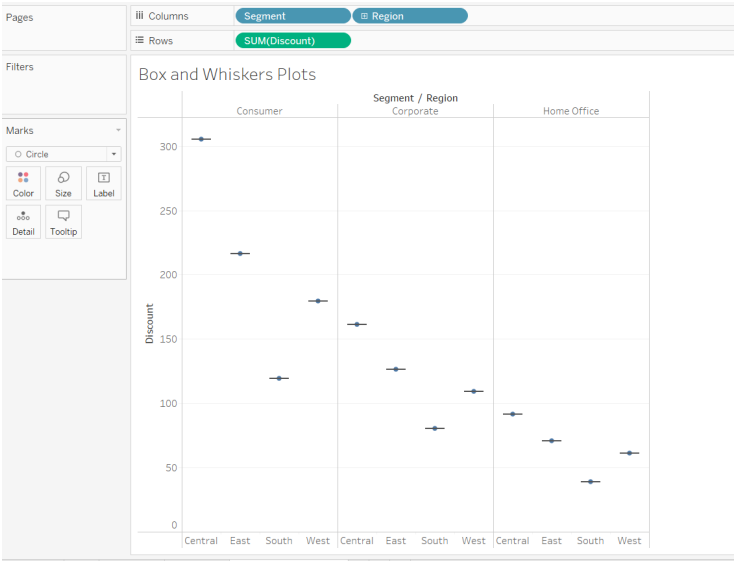



Figure 140: Drag data from Marks card to Columns field

7. You can aggregate and disaggregate data by clicking on **Analysis** in the menu and choose **Aggregate Measure.**

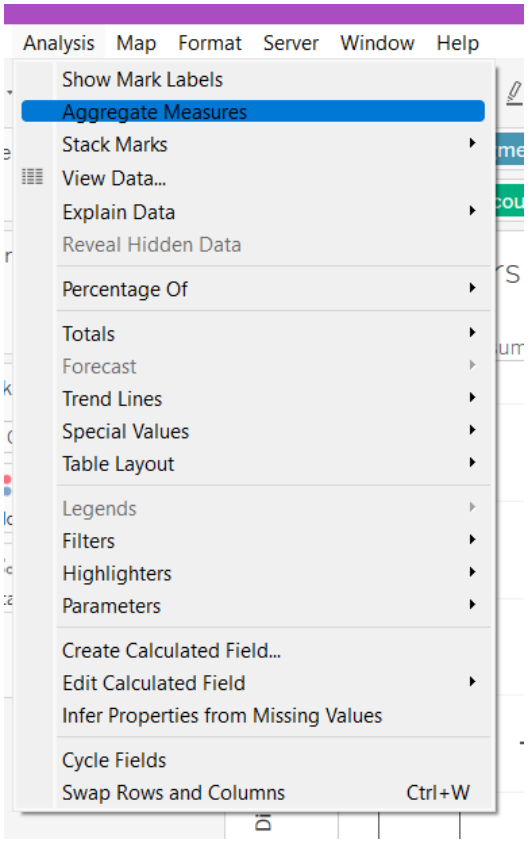



Figure 141: Choose Aggregate Measures

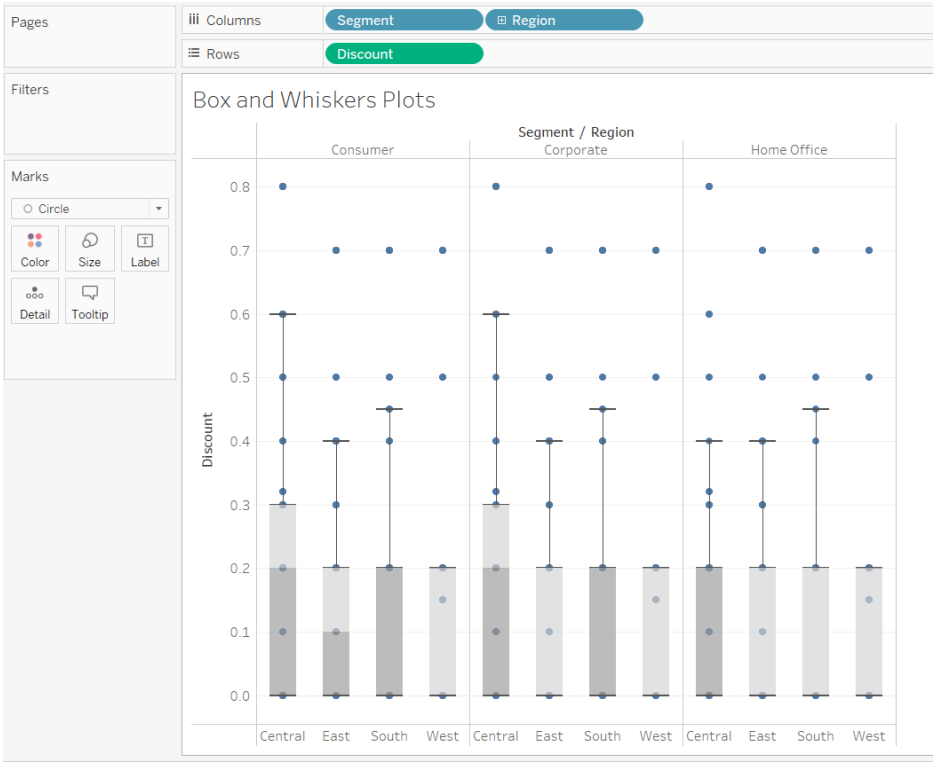



Figure 142: Completed Box-and-whiskers Chart

## 2.3.5.15 Waterfall Charts

Waterfall charts are in analogous orientation and flow. It analyzes the cumulative effect of Measure and see how it increases or decreases.

1. Right click Measures data, then *Create* > *Calculated Field*.

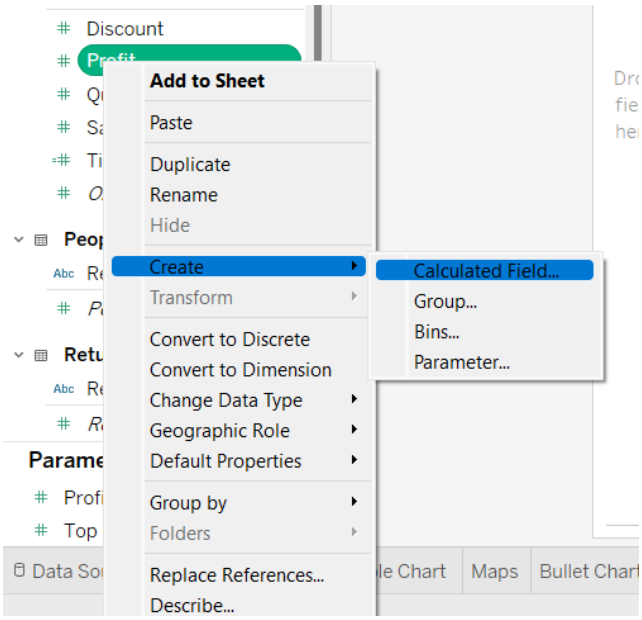



Figure 143: Find Calculated Field option

2. Create a Calculated Field as follows:

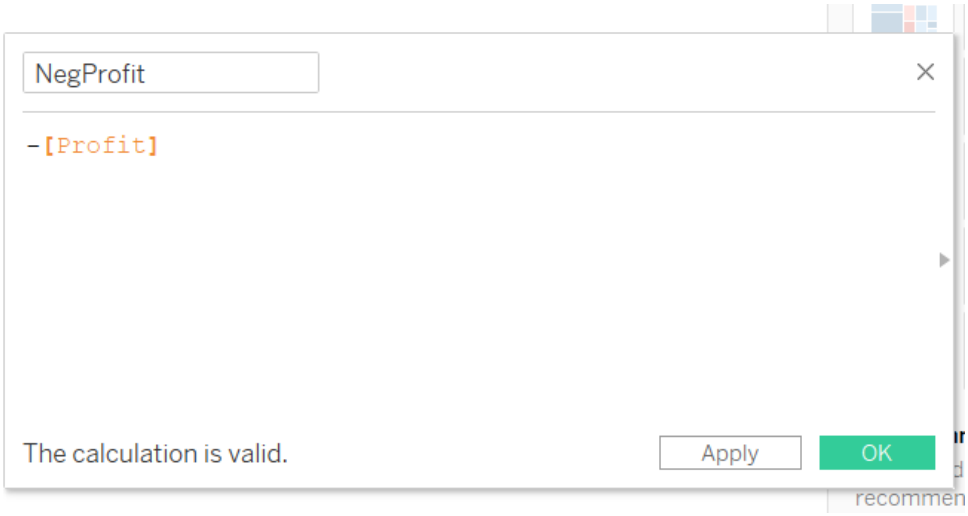



Figure 144: Calculated Field dialog box

3. Change Mark Type to Gantt Bar.

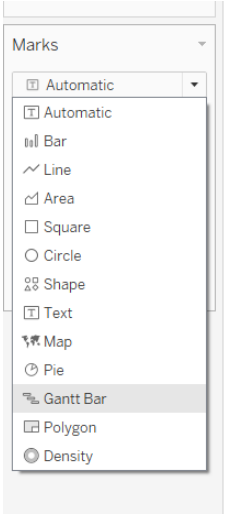



Figure 145: Picking a Gantt Bar

4. Drag NegProfit created just now into Marks card Size. A negative value will extend downwards while positive upwards.
5. Place data into Column and Row. Right-Click data in rows and choose Quick Table Calculation > Running Total.

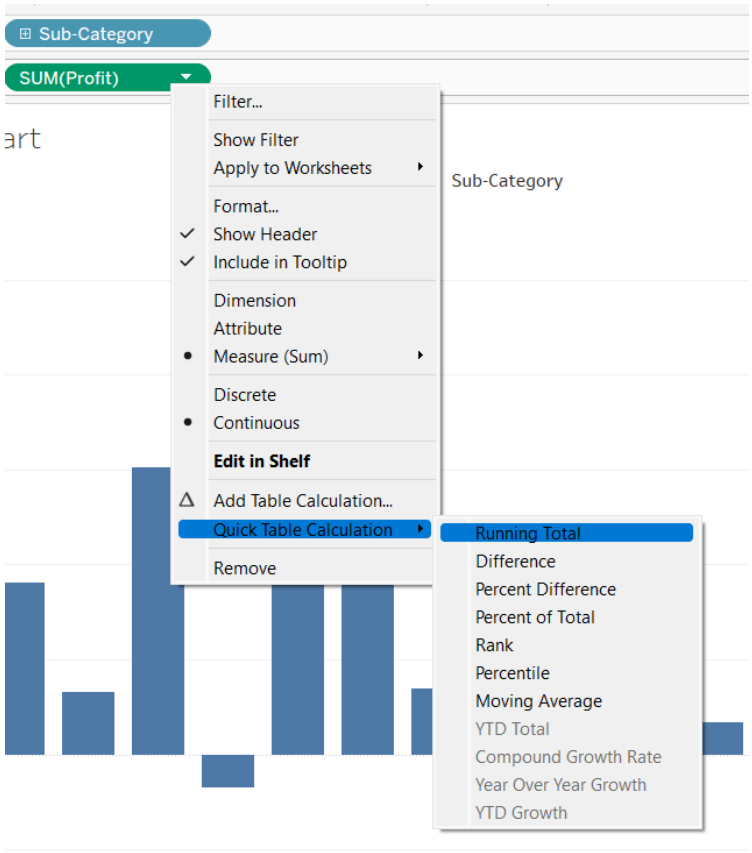



Figure 146: Click on Running Total in Quick Table Calculation option

6. You can also add colors by adding data into Color in Mark's shelf.

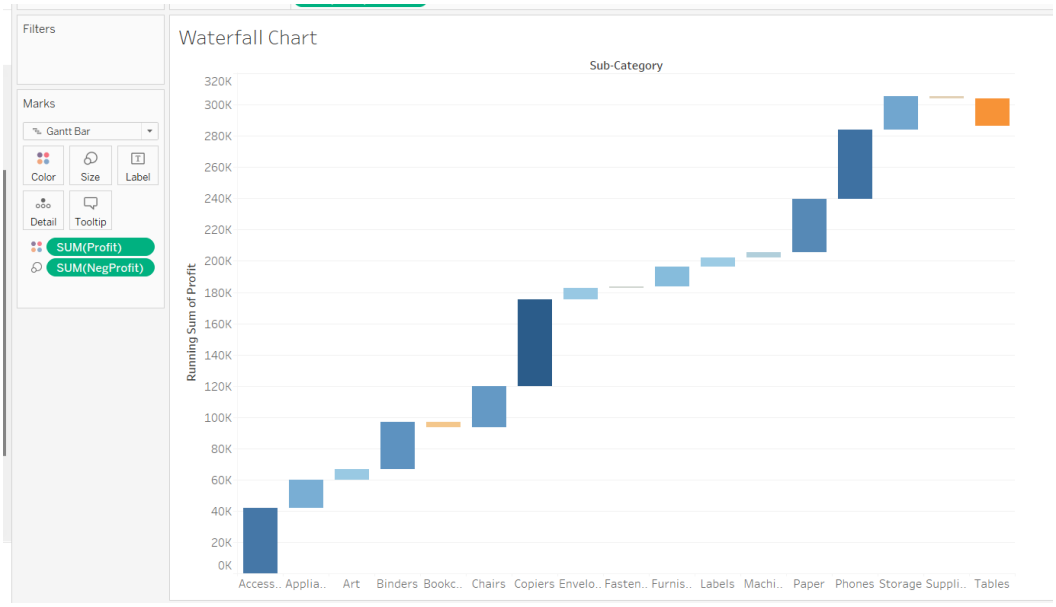



Figure 147: Add data to Color in Marks shelf

## 2.3.6 Save/Export Work

### 2.3.6.1 Packaged Workbooks

You can save your works in a simple way but these packaged workbooks are better for a large project. Packaged workbooks reference external resources which contain the workbook with a copy of any local file data source and background images. The workbook no longer links to the original source. It is being saved with a .twbx file extension.

Create a .twbx with file-based data sources

1. File > Save As
2. Specify file name
3. Select **Tableau Packaged Workbook** on the Save as type.

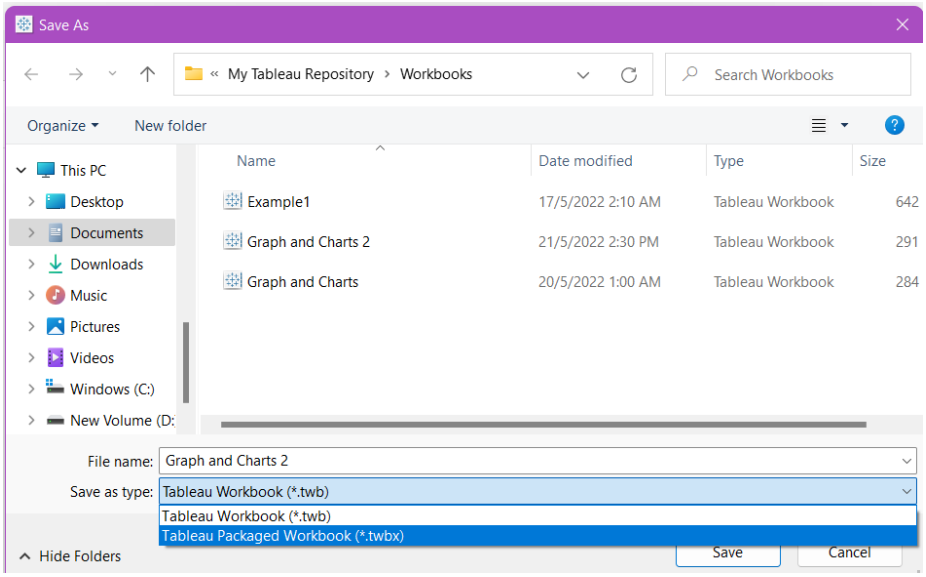



Figure 148: Choosing save as type

4. Click Save.

Following are the files that can be included in the packaged workbook.

- Background images
- Custom geocoding
- Custom shapes
- Local cube files
- Microsoft Access files
- Microsoft Excel files
- Tableau extract files
- Text files

### 2.3.6.2 Export as Image/PDF/Presentation

Image

For Tableau Desktop on macOS, a TIFF(Tagged Image File Format) image is copied to the clipboard while Windows uses a BMP (Bitmap) image.

1. Select **Worksheet > Copy > Image.**
2. In its dialog box, select the element you want to include in the image.
3. Click copy
4. Open the target application and paste it.

You can also create an image file, you can choose BMP, JPEG, PNG, or SVG image file formats on macOS or Windows.

1. Select **Worksheet > Export > Image**.
2. From its dialog box, select the element included in the image.
3. Click Save.
4. Specify a file location, name, and format in the Save Image dialog box and click **Save**.

Powerpoint

1. Select **File > Export as Powerpoint**.
2. Select sheets you want to include in it.

PDF

1. Select **File > Page Setup.**
2. Choose **File > Print to PDF.**

## 2.4 Data Preprocessing

### 2.4.1 Data Cleaning

The process of correcting or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data from a dataset is known as data cleaning. There are endless possibilities for data to be duplicated or mislabeled when combining multiple data sources. Even if the data is correct, outcomes and algorithms are unreliable if the data is incorrect. Because the processes vary from dataset to dataset, there is no absolute way to prescribe the exact steps in the data cleaning process. However, creating a template for your data cleaning process will ensure that you do it correctly every time.

Even though there is no one absolute way to perform data cleaning, you may follow these general steps to perform data cleaning:

1. Remove any unnecessary observations from your dataset, such as duplicates or irrelevant observations. Duplicate observations are most likely to occur during the data collection process. Duplicate data can be created when you combine data sets from multiple sources, scrape data, or receive data from clients or multiple departments.

2. Fix structural errors. When you measure or transfer data and notice strange naming conventions, typos, or incorrect capitalization, you have structural errors. Mislabeled categories or classes can result from these inconsistencies. "N/A" and "Not Applicable," for example, may both appear, but they should be analyzed as one category.

3. Remove any unwanted outliers. There will frequently be one-off observations that do not appear to fit within the data you are analyzing at first glance. If you have a good reason to remove an outlier, such as incorrect data entry, it will help the data you're working with perform better.

4. Handle missing data. Many algorithms will not accept missing values, so you can't ignore them. There are several options for dealing with missing data.

    a) You can drop observations with missing values as a first option, but this will cause you to lose or lose information, so be aware of this before you do so.

    b) Another option is to fill in missing values based on other observations; however, you risk losing data integrity because you're working with assumptions rather than actual observations.

c) As a third option, you could change how the data is used to navigate null values more effectively.

5. Validation and quality assurance. As part of basic validation, you should be able to answer these questions at the end of the data cleaning process:

a) Does the data make sense?
b) Does the data follow the appropriate rules for its field?
c) Does it prove or disprove your working theory, or bring any insight to light?
d) Can you find trends in the data to help you form your next theory?
e) If not, is that because of a data quality issue?

Most of these steps can be done by clicking the triple dot icon on a field, hovering over the "Clean" dropdown menu, then selecting the appropriate action as shown below.

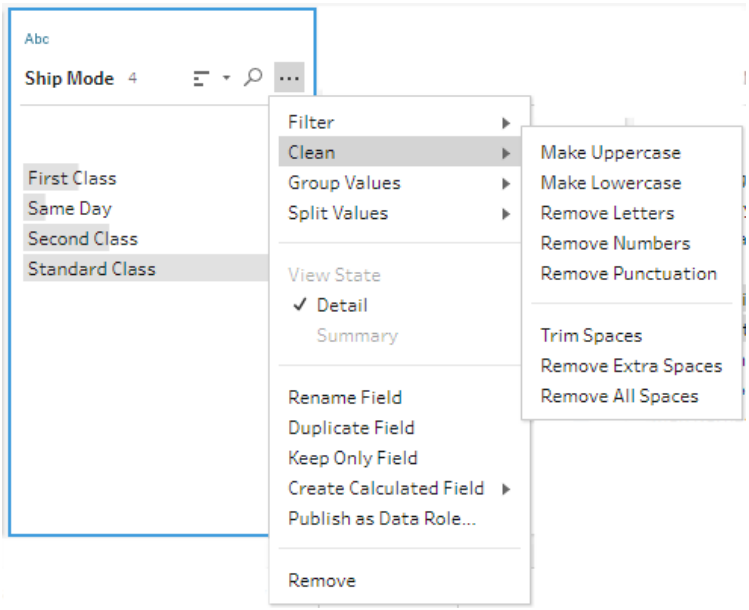



Figure 149: Click on Clean option

## 2.4.2 Data Transformation

The process of converting data from one format or structure to another is known as data transformation. Data transformation, also known as data wrangling or data munging, is the process of transforming and mapping data from one "raw" data form into another for warehousing and analysis.

Tableau offers two ways to transform data on-the-go, that is split and pivot transformations. Below are examples of using these transformations on a certain data set.

### 2.4.2.1 Split transformations

It allows us to automatically split a single column into several columns from a single delimited field in our data. The procedure is as simple as a right-click in Tableau Desktop.

To create a split, begin by connecting to a new data source. You'll see the data window after you've connected, where you can create joins between your connected tables. You'll see something like this: Columns and rows have been reorganized.

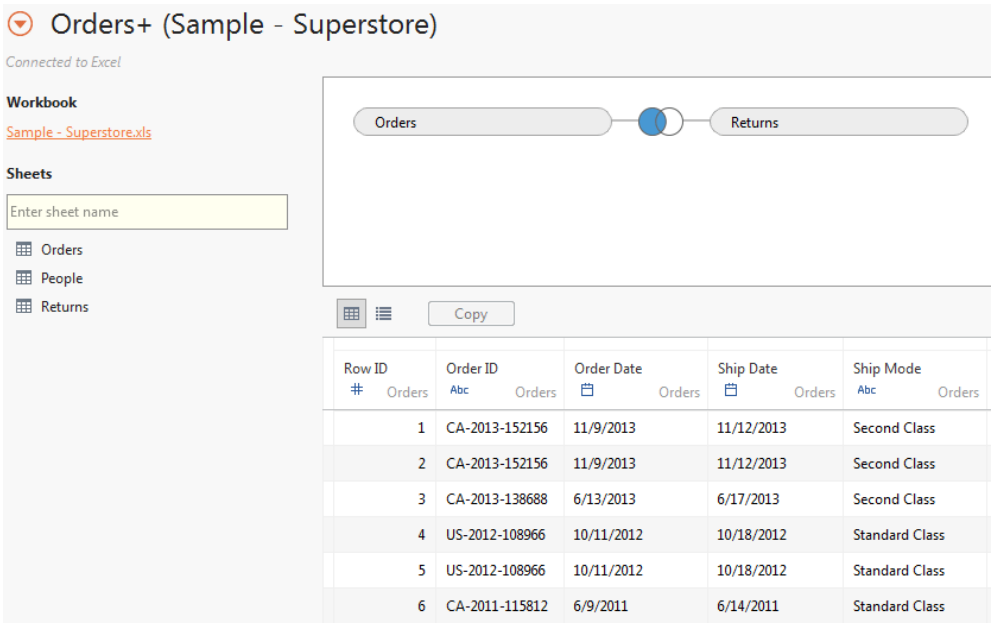



Figure 150: Data source page with sample data

The state, year, and order ID number are separated in the **Order ID** column by two hyphens. The new split and pivot options are available by selecting the dropdown menu next to **Order ID**.

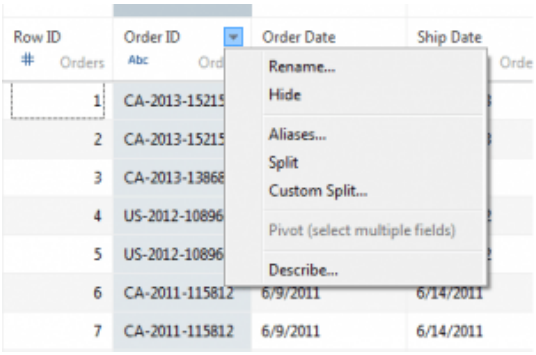



Figure 151: Split and Pivot options

Tableau Desktop recognises the hyphens in the field and treats them as delimiters by simply selecting the split option. The single column is then split into three using a split transformation. You can rename new columns after the split if you want. Below is how the example looks after the split.

| Order ID | Order ID - Split 1 | Order ID - Split 2 | Order ID - Split 3 |
|---|---|---|---|
| CA-2013-152156 | CA | 2013 | 152156 |
| CA-2013-152156 | CA | 2013 | 152156 |
| CA-2013-138688 | CA | 2013 | 138688 |
| US-2012-108966 | US | 2012 | 108966 |
| US-2012-108966 | US | 2012 | 108966 |

Figure 152: After completing the split

### 2.4.2.2 Custom Split Transformation

Tableau also has a custom split option. For instance, referring to the previous example, you might only want to split the State information from our original column and keep the rest. The Custom Split window allows you to specify the delimiter character and the first or last column number.

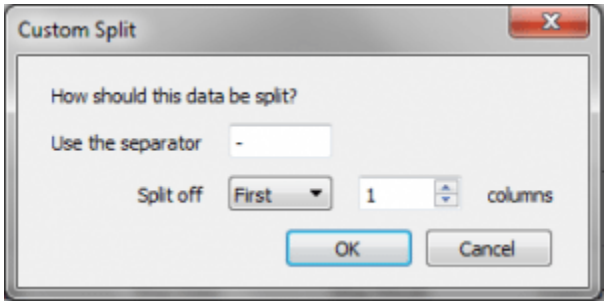



Figure 153: Customizing split

Simply specify a split of the first column in the example to get a new column that identifies states in the data.

| Row ID<br># Orders | Order ID<br>Abc Orders | Order ID - Split 1<br>=Abc |
|---|---|---|
| 1 | CA-2013-152156 | CA |
| 2 | CA-2013-152156 | CA |
| 3 | CA-2013-138688 | CA |
| 4 | US-2012-108966 | US |
| 5 | US-2012-108966 | US |

Figure 154: Customized split

The data window is not the only window that can be split. You can use the split function on the go after you've created a dashboard.

Right-click the axis to use the Order ID column. The function of splitting will be available.

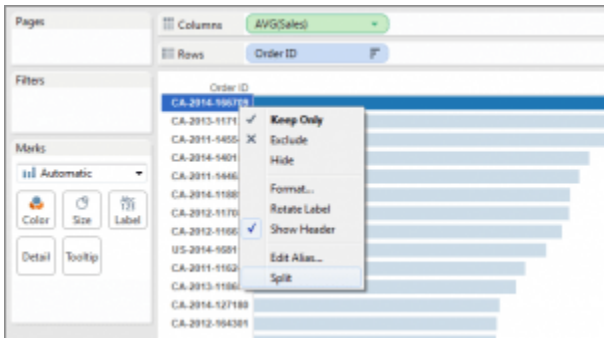


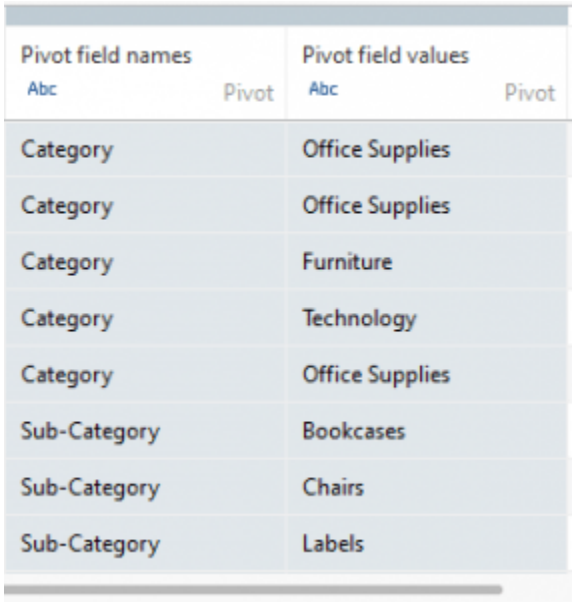

Figure 155: Finding the split option from the axis

### 2.4.2.3 Pivot Transformations

Split transformations can be used to break a single column while pivot transformations can be used to combine multiple columns into one. Pivot field names and values combine multiple columns into two.

Start by selecting the columns you want to transform from the Data Source window. Press Ctrl + right-click on one of the columns. The Category and Sub-category columns are selected in the example below.

The Pivot window appears when you right-click a column. In the Pivot field values column, Tableau Desktop combines data from the selected columns. These results are displayed in the Pivot field names category column.



| Pivot field names Abc Pivot | Pivot field values Abc Pivot |
|---|---|
| Category | Office Supplies |
| Category | Office Supplies |
| Category | Furniture |
| Category | Technology |
| Category | Office Supplies |
| Sub-Category | Bookcases |
| Sub-Category | Chairs |
| Sub-Category | Labels |

Figure 156: Pivot transformation

## 2.4.3 Data Discretization

The process of converting continuous data attribute values into a finite set of intervals and associating a specific data value with each interval is known as data discretization. In Tableau, You can convert discrete to continuous or continuous to discrete measures. You can also convert date and other numeric dimensions to discrete or continuous formats.

To convert discrete to continuous or continuous to discrete measures, Select Discrete or Continuous from the drop-down menu as shown below. When the field is continuous, it is green; when it is discrete, it is blue.

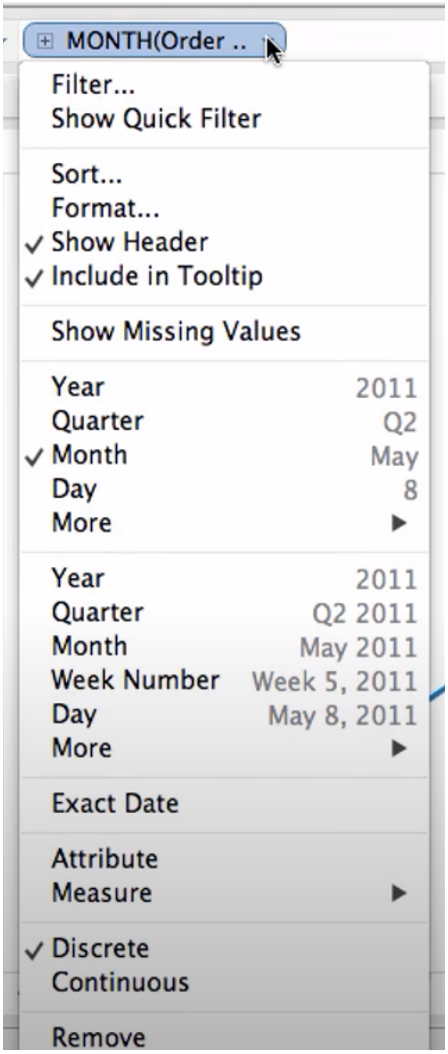



Figure 157: Choosing Discrete function

Date fields can be converted from discrete to continuous. Continuous dates will have an axis similar to a measure, while discret dates will act as labels.

To change a Date field from discrete to continuous or continuous to discrete, right-click it and select one of the options from the context menu.

Right-click a Date field in the Data pane and select Convert to Discrete or Convert to Continuous to convert it (and thus determine the default result when you drag it into a view).

## 2.5 Analysis

### 2.5.1 Clustering

The process of finding similar groups of any objects so that the data can be grouped into clusters is known as clustering. Clustering is an example of an unsupervised Machine Learning Algorithm. Clustering is used in data mining as it can help to split the data into diverse subsets. Each subsets consist of data similar to one another and these subsets are known as clusters.

Below are the steps in how to do a clustering analysis in Tableau.

#### 2.5.1.1 Create Cluster

To create clusters inside a view in Tableau, do follow these steps.

1. Create a view. To create a view, select at least one measure for each column and row. This can be done by manually dragging the variables to the column and row.

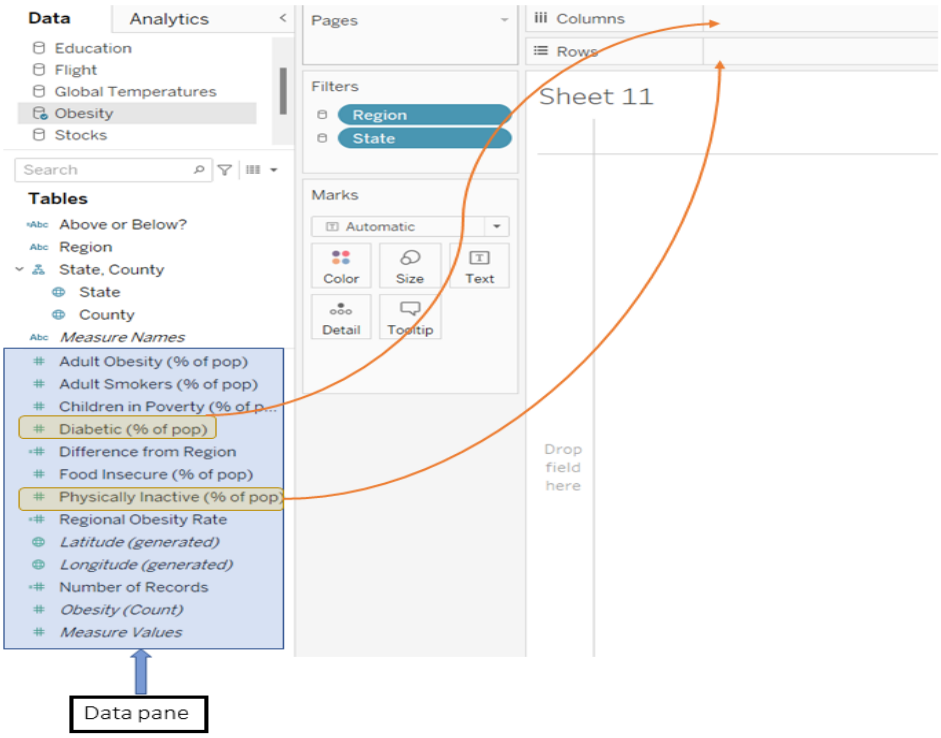



Figure 158: Inserting variables into the view

It should look as if like this after the column and row has been filled, however the charts or graphs can be varied according to Tableau settings or according to your own choice.

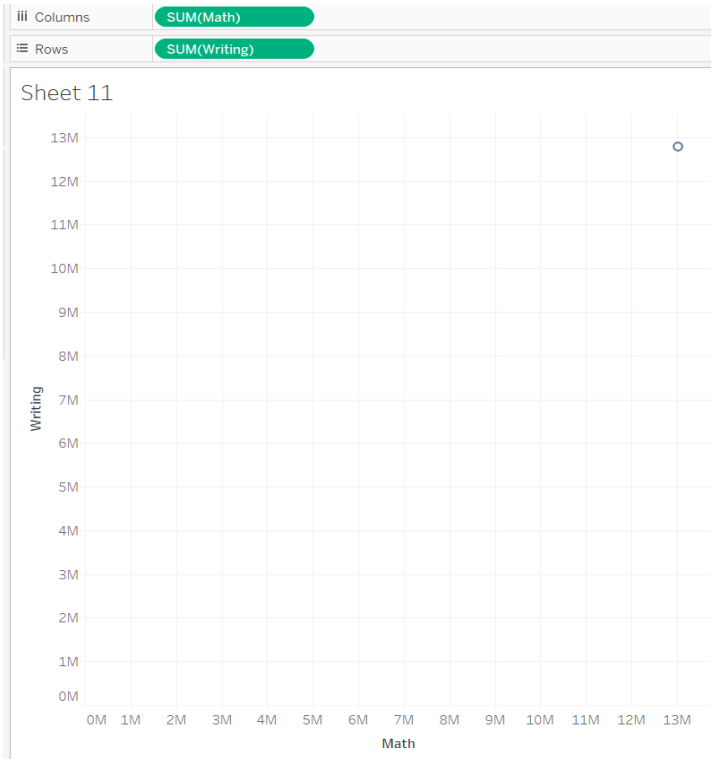



Figure 159: View inside the view after inserting variables

In some cases, the view will only show one point. To change this, just follow the step shown below.

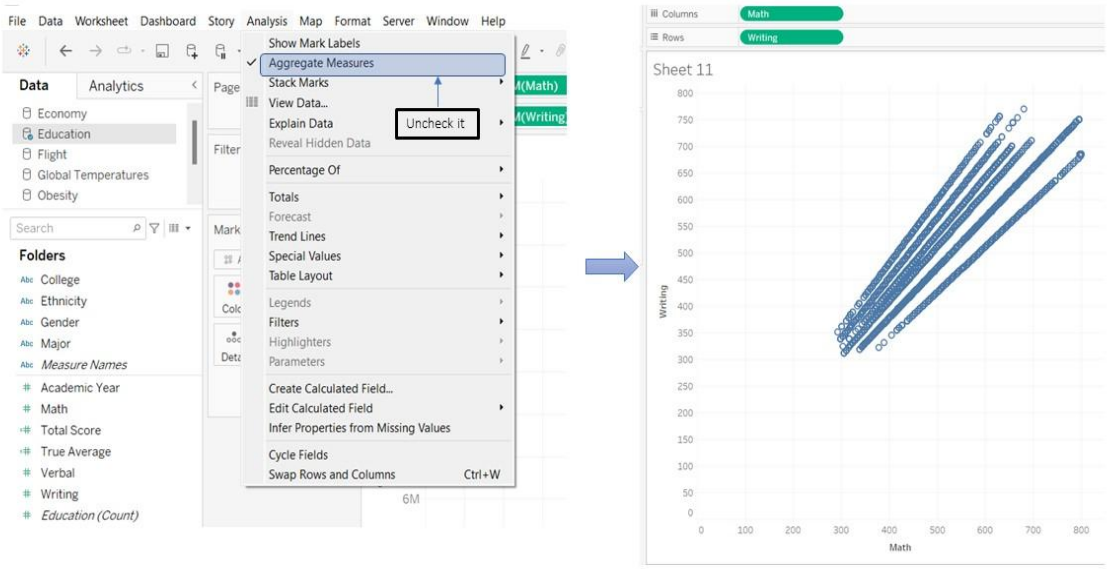



Figure 160: Changing the way of displaying the plots

2. Create cluster(s). Below are the steps shown on how to create a cluster from view.

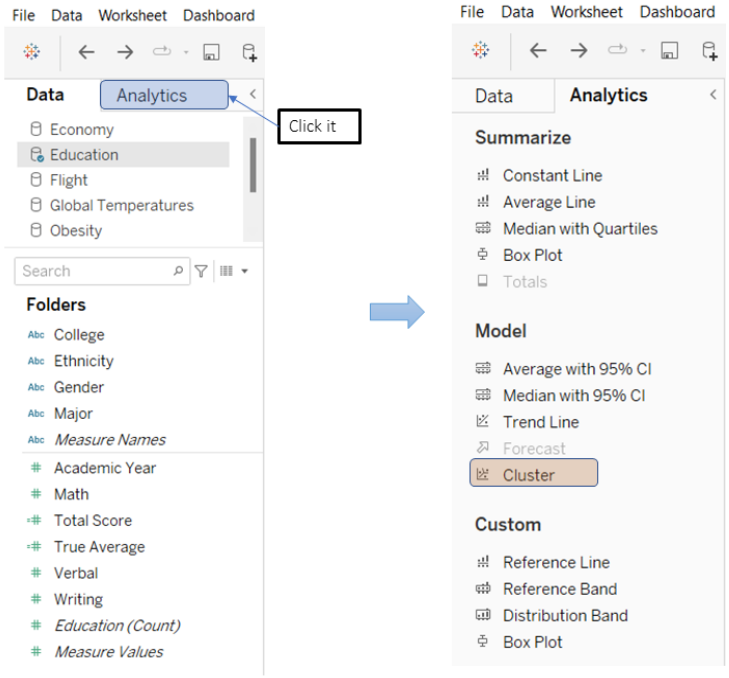



Figure 161: Finding cluster function

By clicking on the 'Analytics' tab, under the 'Model' category, there is a cluster function that can be found. To apply the cluster function, it either can be dragged into the view or by double clicking on the cluster function.

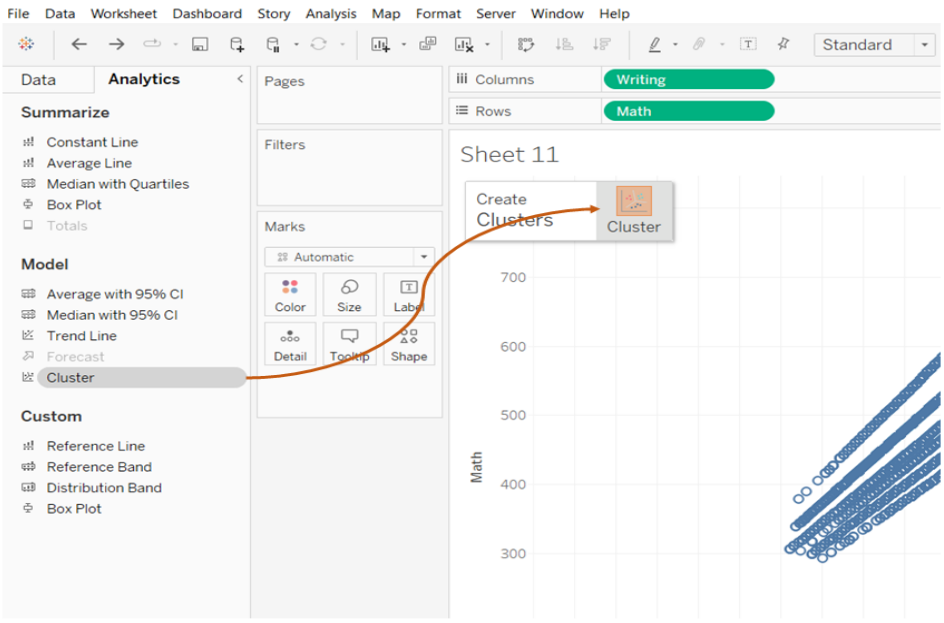



Figure 162: Drag cluster function into the view

After that, a small window will pop up on the screen, which is a dialog box, where the cluster can be customized in there.

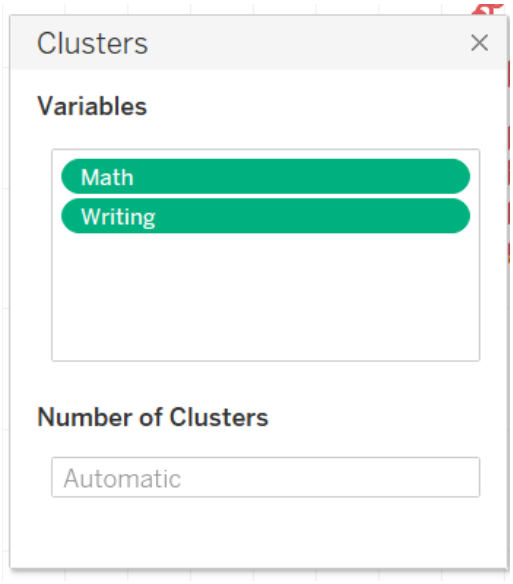



Figure 163: Customizing cluster

3. Customize the cluster. These are some of the list of customization that can be made in the dialog box:

    - Drag new fields from the Data pane into the Variables area of the Clusters' dialog box. To remove the fields from the Variables area, just drag it out.
    - Specify the number of clusters (between 2 and 50). If the value is not specified, Tableau will automatically create up to 25 clusters.

4. After all the customization has been done, click on the X in the upper-right corner of the Clusters dialog box to close it.

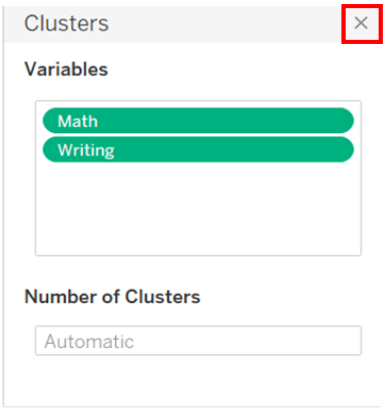



Figure 164: Exit cluster dialog box

### 2.5.1.2 Edit Cluster

To edit the cluster created, right-click (if using Windows) or control-click (on a Mac), on a **Clusters** field on **Color** and select **Edit clusters**.

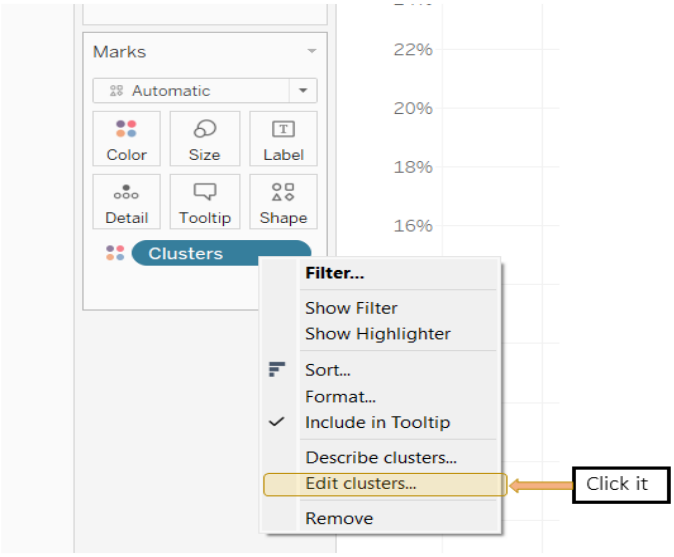



Figure 165: Finding edit cluster option

Each cluster can be renamed, however, users need to drag the Clusters field to the **Data** pane and save it as a **group** (see 2.5.1.3).

Right-click the cluster group and select **Edit Group** to make changes to each cluster.

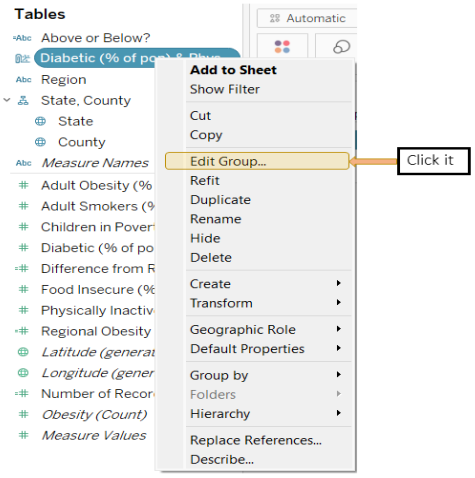



Figure 166: Specified the category to be edited

Select a cluster group in the list and click **Rename** to change the cluster name.

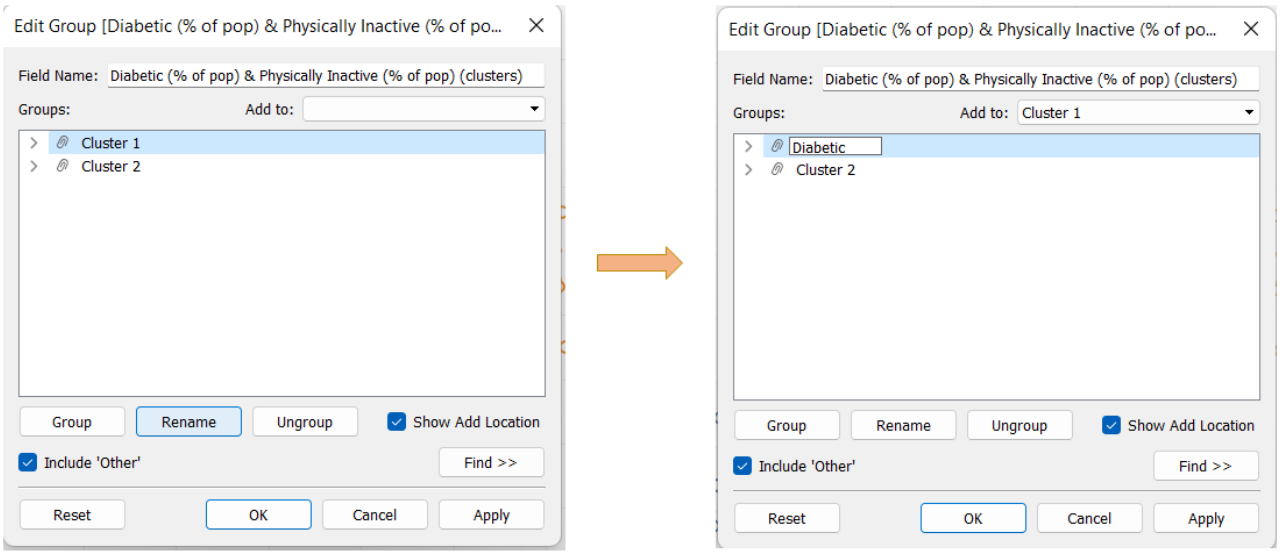



Figure 167: Rename cluster

### 2.5.1.3 Create a Group From Cluster Results

To create a group from cluster results, these are the steps that should be followed.

1. Make sure that you have created a cluster.
2. From the **Marks** pane, drag the **Clusters** to the **Data** pane.

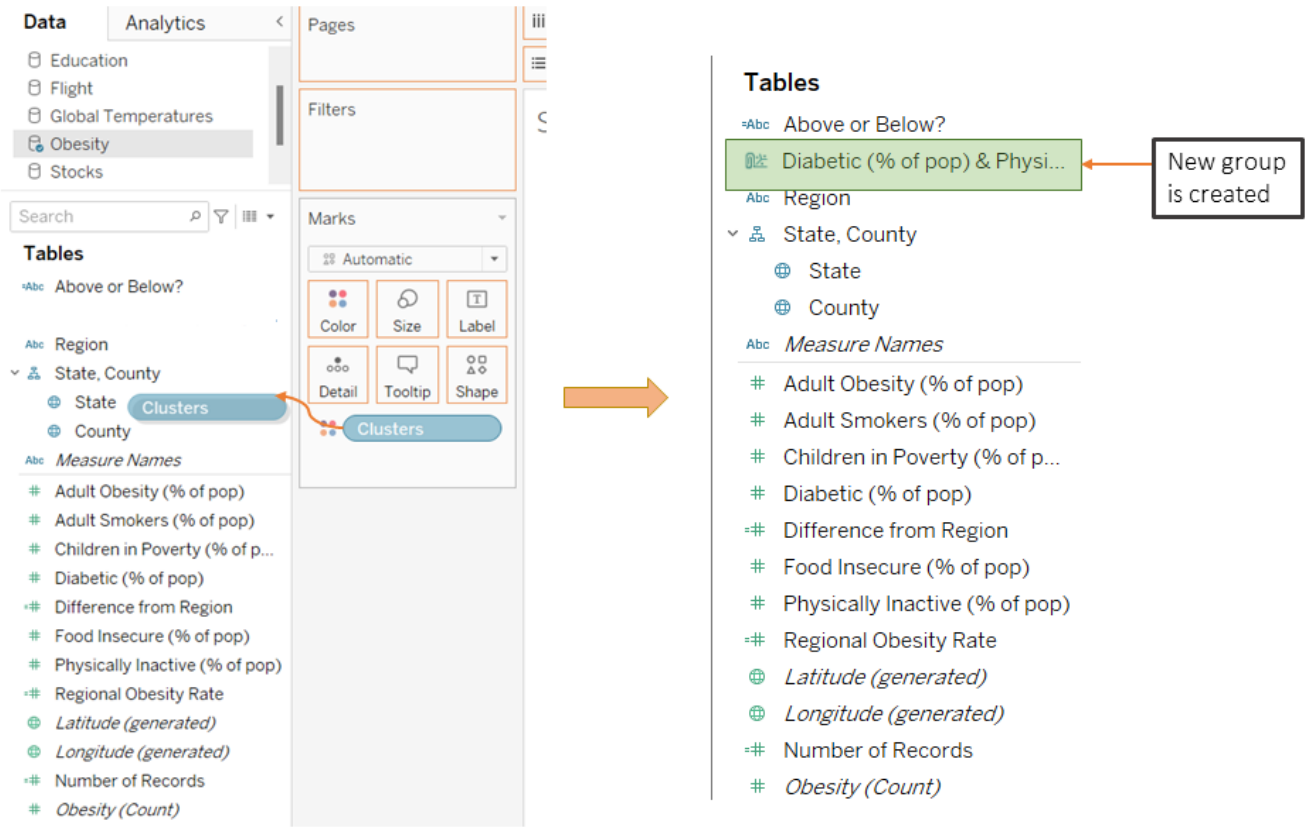



Figure 168: Drag the Clusters to the Data pane

That is how users can create a group from cluster results.

### 2.5.1.4 Refit Saved Clusters

In Tableau, when a Clusters field is saved as a group, it will be saved with its analytic model. The saved cluster groups can be used in other worksheets and workbooks,even so, they are not automatically refreshed.


Figure 169: Example of cluster field

Above is the example after a saved cluster group and its analytic model is applied to a different worksheet. As it can be seen, there are some marks which were not included in the clustering (the gray marks).

Therefore, these are the steps to refit a saved cluster.

1. Right-click a cluster group in the Data pane, then click **Refit**.

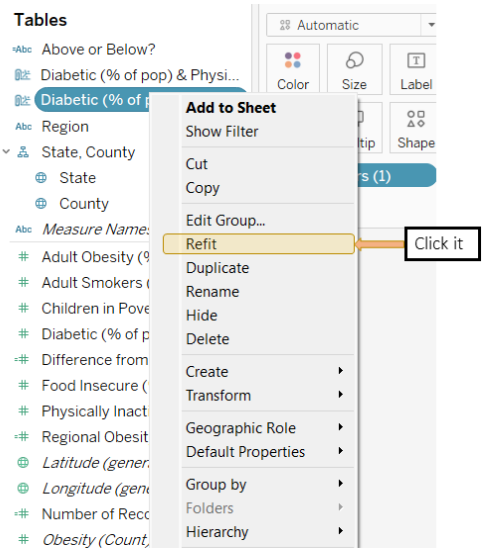



Figure 170:Click Refit

Then, Tableau will automatically refit the cluster back so there will be no more not clustered data.

Here is an example of updated clustering after refitting the saved cluster:

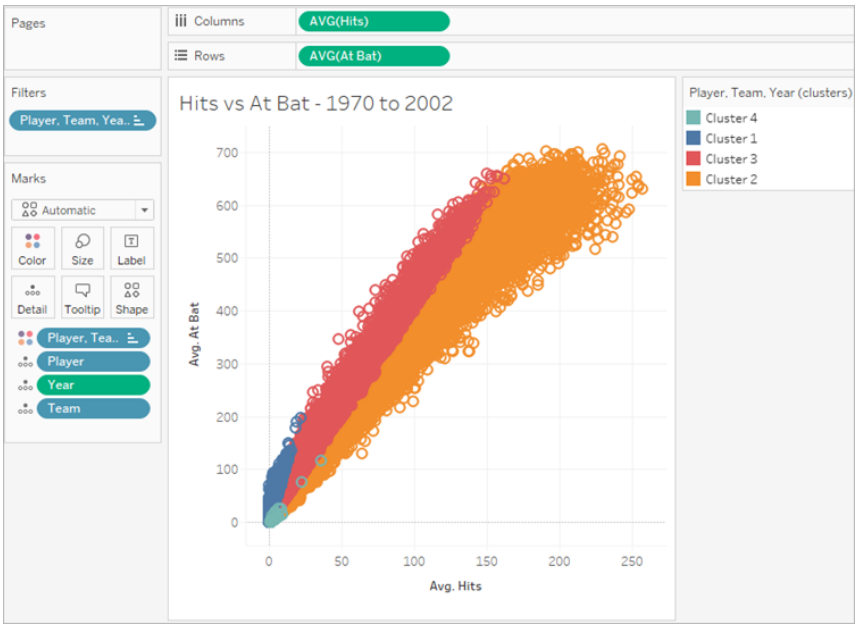
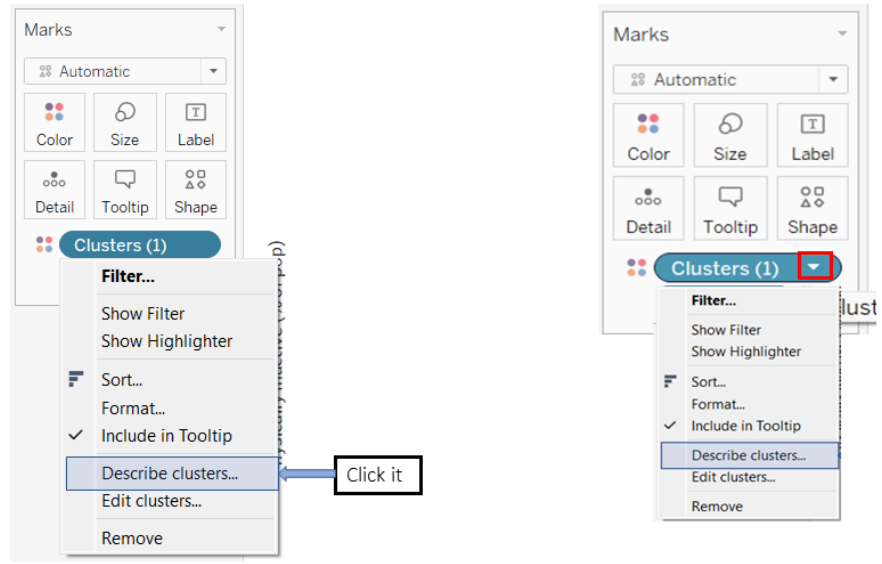



Figure 171: Updated clustering

### 2.5.1.5 Get Information on Statistics Models

To obtain all the information about the model that Tableau has computed for clustering, please follow the steps shown.

1. Right-click on **Clusters** on the **Marks** card or click on the small downward arrow at the right end of **Clusters** and select **Describe Clusters**. The information in the Describe Clusters dialog box is read-only, however, it can be copied and pasted into a writable document by clicking on **Copy to Clipboard** at the left-bottom of the Describe Clusters dialog box.




(a) By clicking right-click on mouse

(b) By clicking right-end arrow

Figure 172: Click Describe clusters

2. Then, this is an example of the Describe Clusters dialog box.

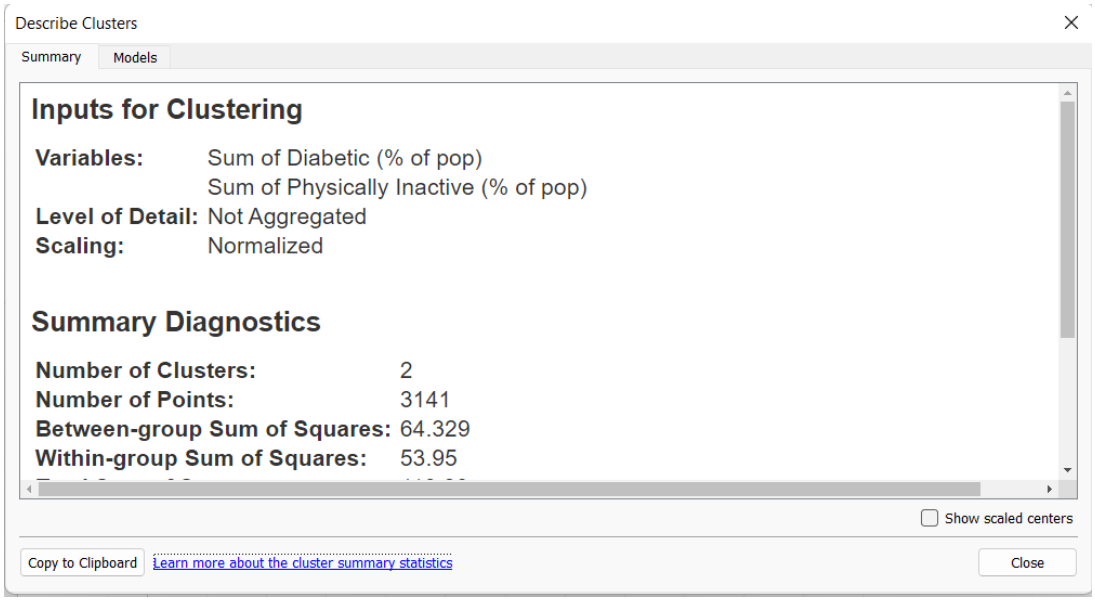



Figure 173: Describe Clusters dialog box

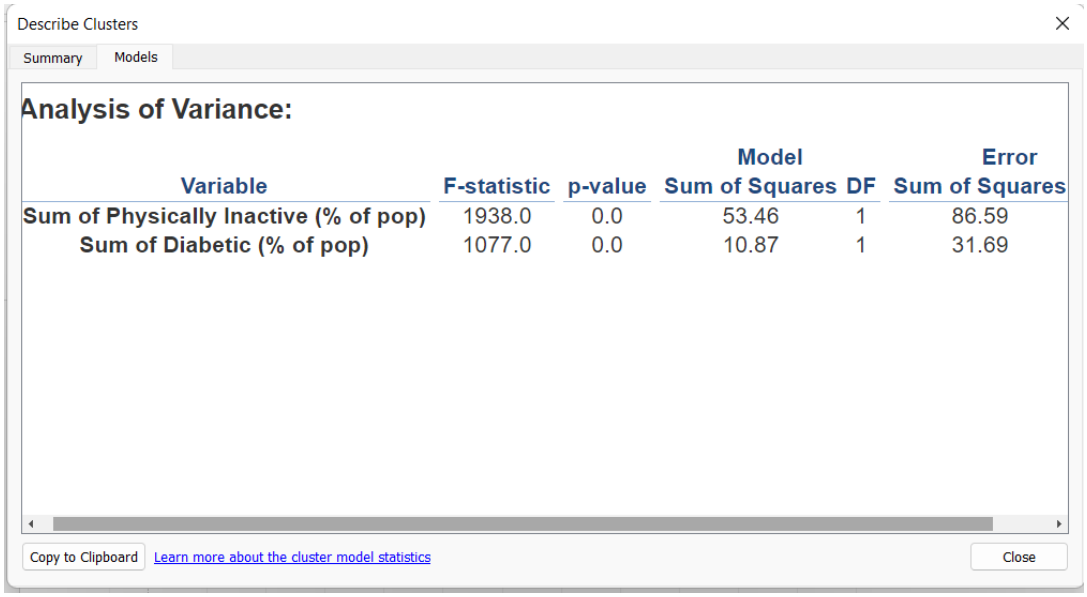



Figure 174: Describe Clusters dialog box

2.5.1.6 Describe Clusters - Summary Tab

| **Inputs for Clustering** | |
|---|---|
| Variables | Identifies all the fields Tableau uses to compute the clusters |
| Level of Details | Identifies all the fields that contribute to the view's level of detail (field that determine the level of aggregation) |
| Scaling | Identifies the scaling method used for pre-processing. Normalized is the only scaling method Tableau uses currently. The formula for this method is known as min-max normalization, which is **(x - min(x)) / (max(x) - min (x))** |
| **Summary Diagnostics** | |
| Number of Clusters | Number of individual clusters in the clustering |
| Number of Points | The number of marks in the view |
| Between-group Sum of Squares | A metric that measures cluster separation as the sum of squared distances between each cluster's center (average value) and the data set's center, weighted by the number of data points given to the cluster. The higher the value, the better the cluster separation. |
| Within-group Sum of Squares | The sum of squared distances between the center of each cluster and the individual marks in the cluster is used to measure cluster cohesion. The lesser the value, the closer the clusters are together. |
| Total Sum of Squares | The sum of squares between groups and within groups is added together. The proportion of variation explained by the model is calculated as (between-group sum of squares)/(total sum of squares). Values range from 0 to 1, with higher values indicating a better model. However, because you may improve this ratio simply by increasing the number of clusters, using this statistic to compare a five-cluster model to a three-cluster model could be misleading. |
| **Cluster Statistics** | |
| # Items | Number of the marks that is present in the cluster |
| Centers | Average value for each cluster (displayed for numeric items) |
| Most Common | The most common value in each cluster (shown only for categorical items) |

### 2.5.1.7 Describe Clusters - Model Tab

Analysis of variance (ANOVA) is a set of statistical models and processes for examining variation within and across groups or clusters of observations. Analysis of variance is computed per variable in this scenario, and the resulting analysis of the variance table can be used to discover which variables are most efficient at separating clusters.

This is a list of relevant analysis of variance statistics for clustering used in Tableau:

| Analysis | Description |
|---|---|
| F-statistic | The F-statistic for one-way ANOVA is the fragment of variance explained by a variable. It is the ratio of the between-group variance and the total variance.<br>The smaller the F-statistic, the worse the corresponding variable is distinguishing between the clusters. |
| p-value | It is the probability that the F-distribution of all possible values of the F-statistic takes on a value greater than the real F-statistic for a variable.<br>If the p-value falls short on a specific significant level, then the null hypothesis can be denied.<br>The degrees of freedom for this F- distribution are (k - 1, N - k), where k is the number of clusters and N is the number of items (rows) clustered.<br>The higher the p-value, the less the expected values of the elements of the corresponding variable differ among the clusters. |
| Model Sum of Squares and Degrees of Freedom | The Model Sum of Squares is the ratio of the between-group sum of squares to the model degrees of freedom. The between-group sum of squares is a measure of the variation between cluster means.<br>If the cluster means are far from each other (far from the overall mean), this value will be large.<br>The model has k-1 degrees of freedom, where k is the number of clusters |
| Error Sum of Squares and Degrees of Freedom | The ratio of within-group sum of squares to the error degrees of freedom. The within-group sum of squares measures the variation between observations within each cluster.The error has N-k degrees of freedom, where N is the total number of observations (rows) clustered and k is the number of clusters.<br>The Error Sum of Squares can be used as the overall Mean Square Error, assuming that each cluster center represents the "truth" for each cluster. |

## 2.5.2 Forecasting

### 2.5.2.1 Create a Forecast

To create a forecast, the view must consist of at least one **date** dimension and one measure.

1. Create a view for the forecast

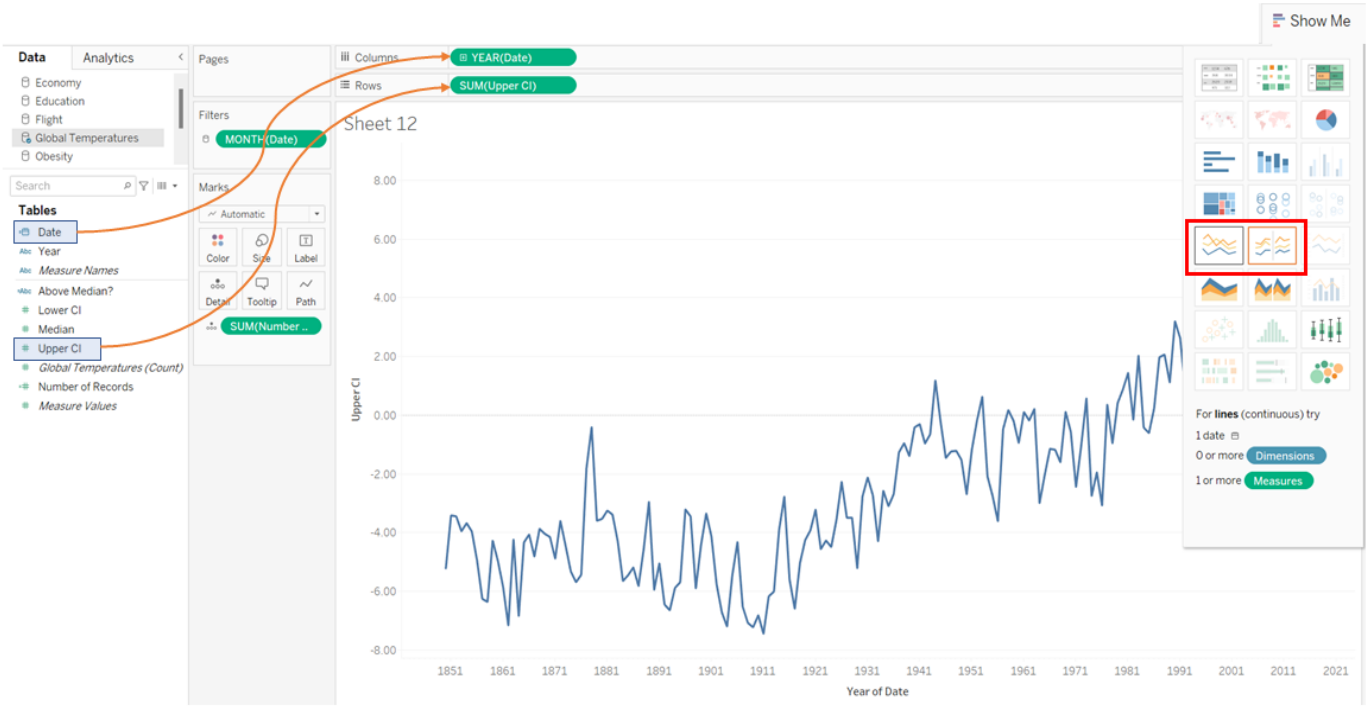



Figure 175: Create a view

By dragging the data from **Data pane** to column and row (make sure that it consists of **at least one date dimension** and one measure) and choose either graphs or charts (according to the suitability) under the **Show M**e tab.

2. Drag the **Forecast** from the **Analytics tab** to the view to create the forecast.

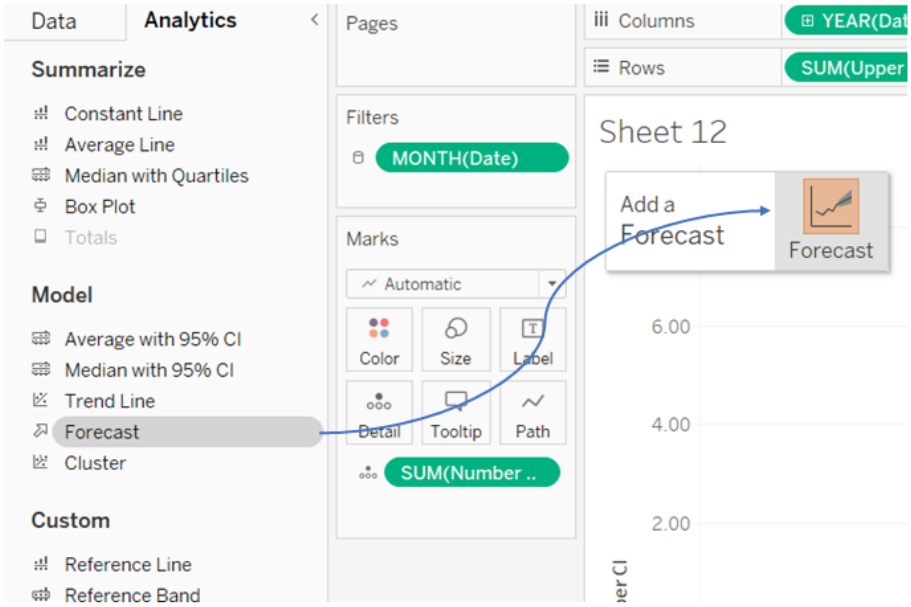



Figure 176: Drag the Forecast from the Analytics tab

It should look like the example below after **Forecast** is applied in the view.

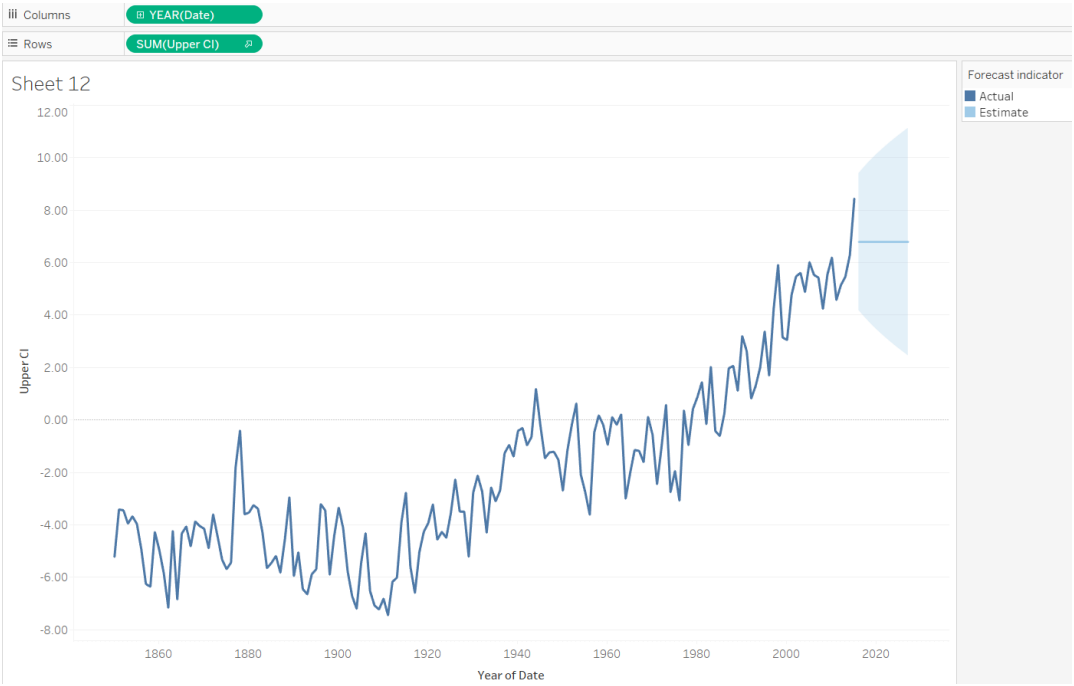



Figure 177: Result

Therefore, this is how users can create a forecast.

## 2.5.2.2 Labels in a Forecast

There are a few things that are featured in a forecast. One of them is to insert labels in the view. These are the label features of Tableau .

By clicking the **'Label'** box in **Marks** Tab, users can show the label on the forecast and can choose the way the label will be displayed.

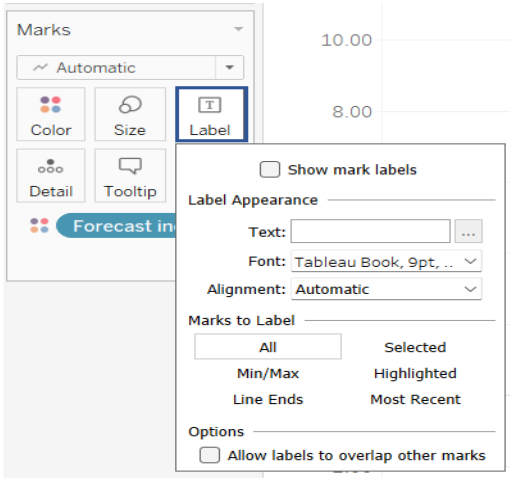



Figure 178: Choose the way the label will be displayed

By clicking on the **Show mark labels**, all the data in the forecast will be displayed in the line graph. This is the example of it:

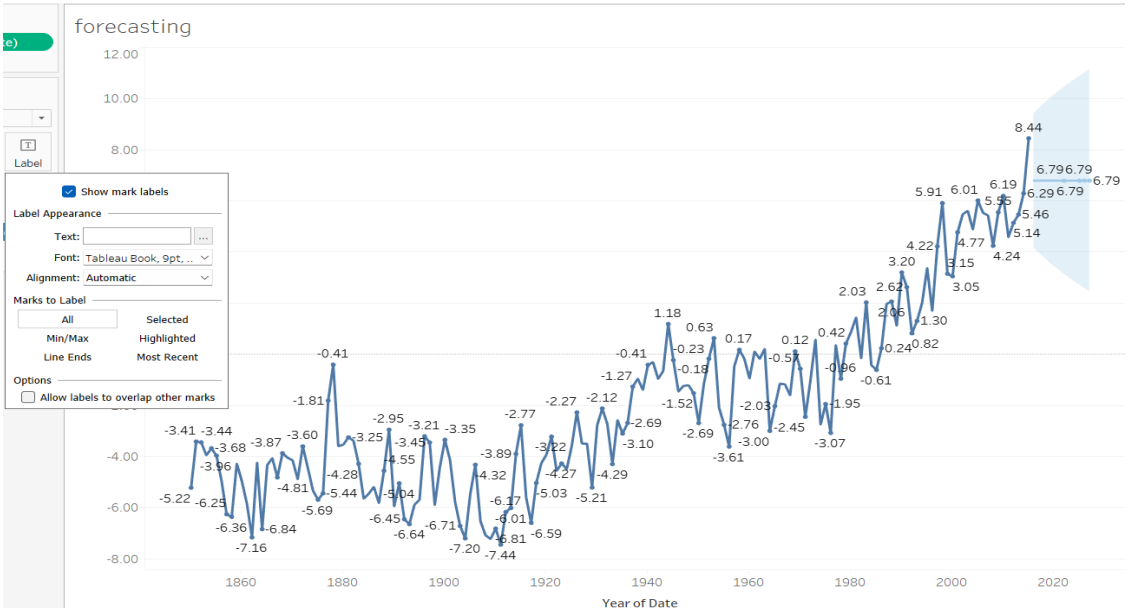



Figure 179: Displayed forecast in line graph

To select the specific label, users can choose from the choices available at **Marks to Label**. Note that **to choose from Marks to Label**, make sure that the **'Show mark labels'** is checked. Below is the example if **Min/Max** label is chosen.

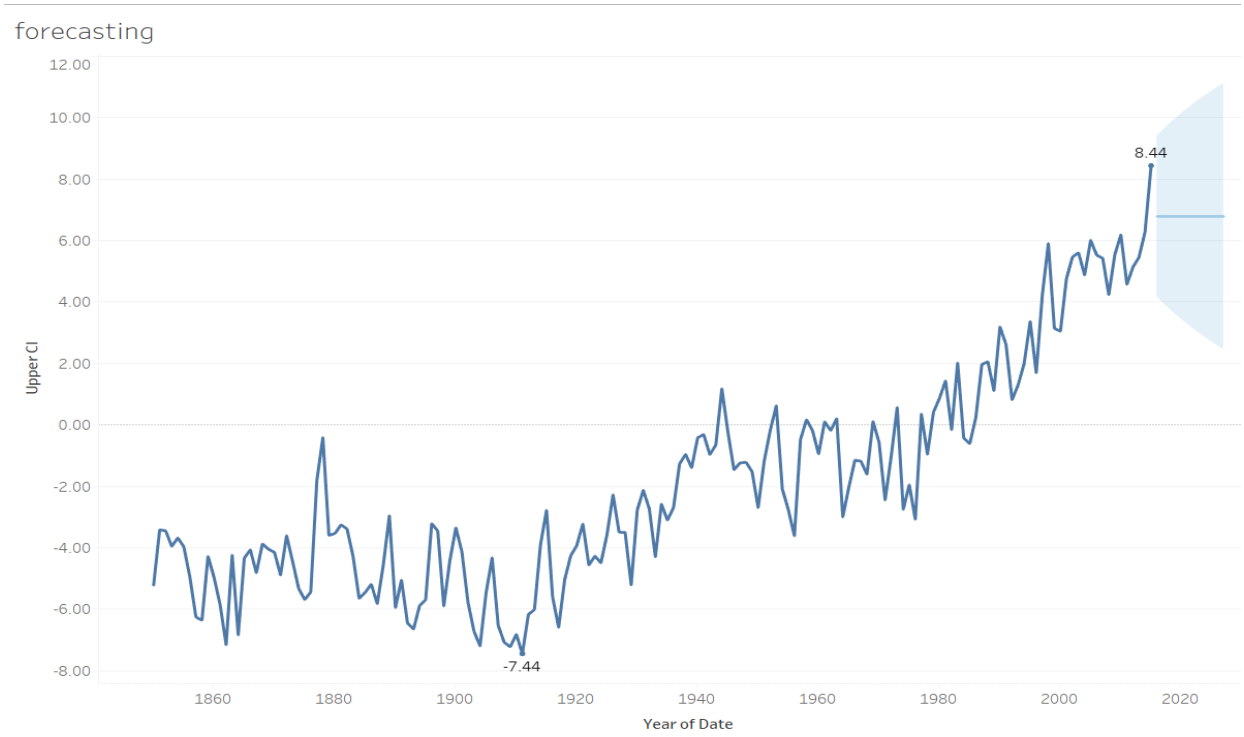



Figure 180: Min/Max label

### 2.5.2.3 Prediction Intervals

From the forecast shown in the view, Tableau has shaded the area in which the estimated values are computed. These values are by default shown as 95% prediction intervals for the forecast. This means that the model has determined that there is 95% chances of success of the predicted value to be obtained for the forecast period.

The prediction bands can be adjusted to a few confidence level percentiles. Therefore, below are the steps to change the confidence level percentile of the forecast.

1. Right-click on the view (model) then users can see a few options that can be chosen from it. Drag the cursor to **Forecast** and select **'Forecast Options…'**

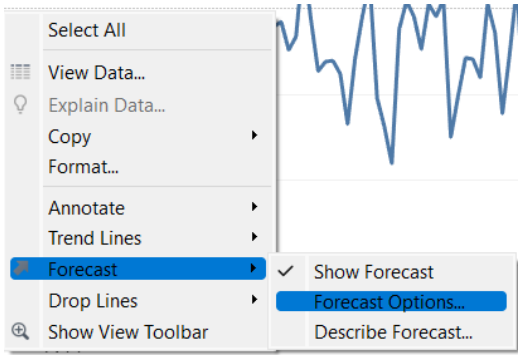



Figure 181: Select Forecast Options

Then a forecast dialog box will appear like shown below.

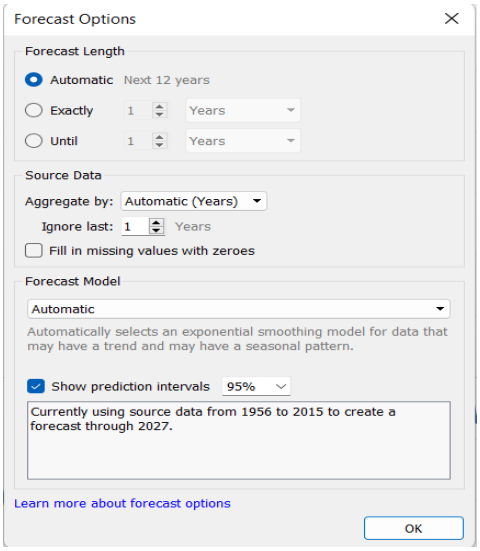



Figure 182: Forecast dialog box

2. On the bottom part of the Forecast dialog box, there is a section where a checkbox of **Show prediction intervals**. Beside it, there is a ribbon tab of confidence level percentile choices. That is where the confidence level percentile can be manipulated.

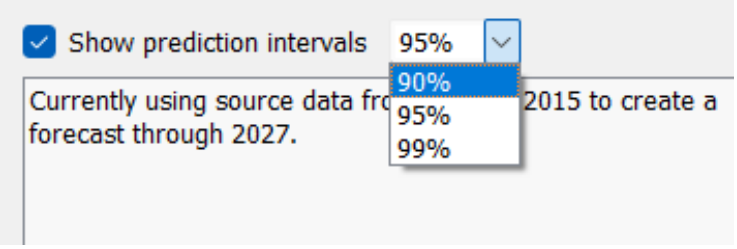



Figure 183: Choose your confidence level

So, that is how users can change the prediction intervals in Tableau.

### 2.5.2.4 Enhancing Forecasts

Users must also consider to verify the quality and precision of the forecast made. This can be done in a few steps which will be visualized below.

1. Drag another instance of the forecast measure from the Data **pane** to the **Detail** shelf on the **Marks** card.

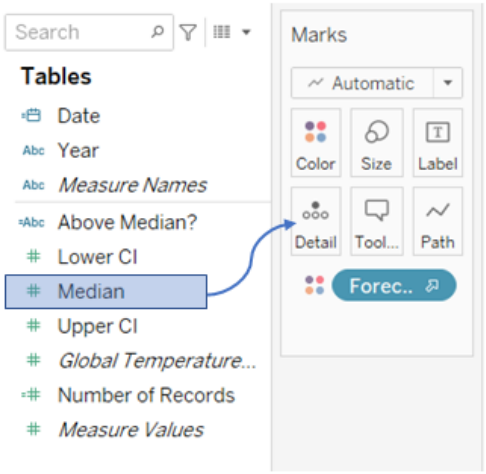



Figure 184: Drag another instance

2. Right-click on the new field to open the menu, then choose one of the available options.

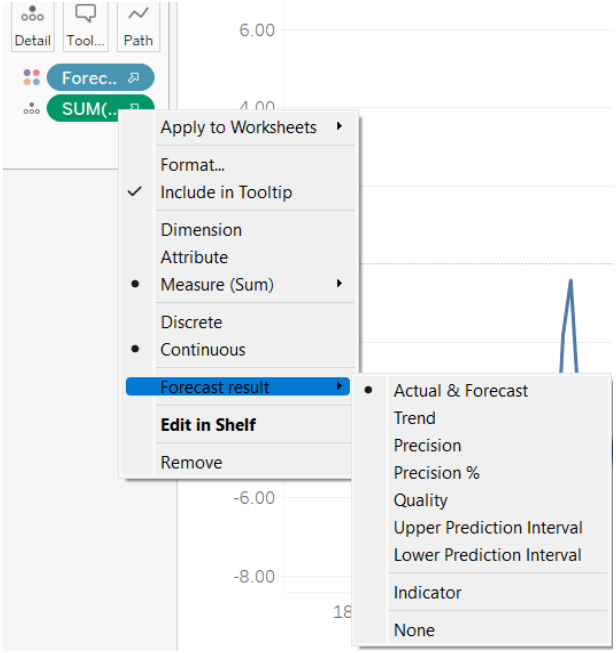



Figure 185: Choose one of the available options

3. This is a list of the description of these options:

| Options | Description |
|---|---|
| Actual & Forecast | Display the actual data extended by forecasted data. |
| Trend | Display the forecast value with the seasonal component removed |
| Precision | Display the prediction interval distance from the forecast value for the configured confidence level |
| Precision % | Display the precision as a percentage |
| Quality | Display the quality of the forecast, on a scale of 0 (worst) to 100 (best) |
| Upper Prediction Interval | Shows the value above which the true future value will lie confidence level percent of the time assuming a high quality model |
| Lower Prediction Interval | Shows 90, 95, or 99 confidence level below the forecast value |
| Indicator | Show the string **Actual** for rows that were already on the worksheet when forecasting was inactive and **Estimate** for rows that were added when forecasting was activated |
| None | Hide the forecast data for this measure |

## 2.5.3 Regression

### 2.5.3.1 Create a Regression

Creating regression analysis does not require many steps to accomplish it. Therefore, these are the steps that should be followed.

1. Firstly, users need to create a line graph. Since it has been shown in 2.3, this step will not be detailed as the others. To do this drag any variables (at least two) into the view and choose a line graph on the **Show Me** ribbon tab. Below is the example of the line graph made.

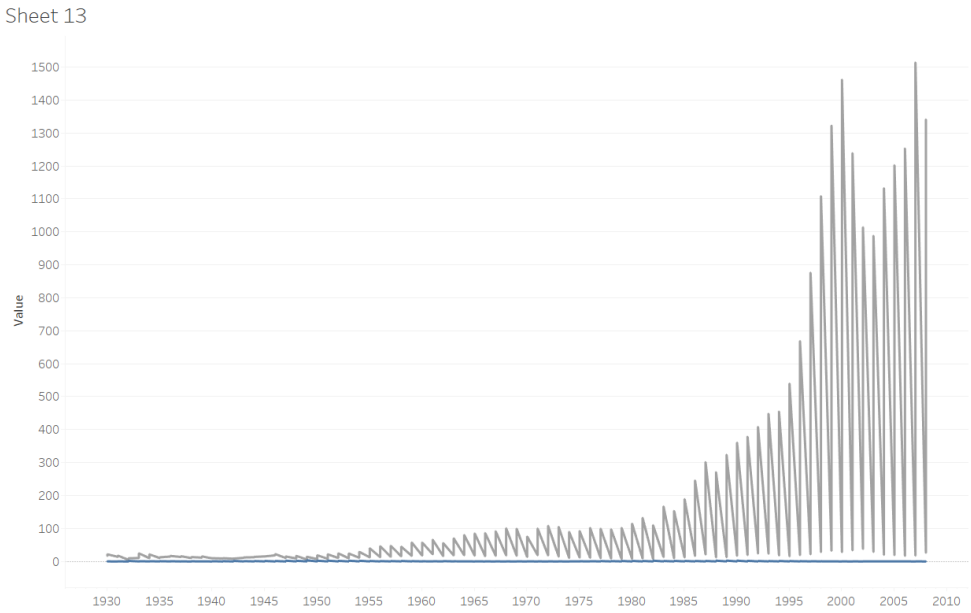



Figure 186: Create line graph

2. To add a regression line, users can find **Trend Line** in the **Analytics** tab. Drag the Trend Line in the view and select the desired trend line. To make it simple, this tutorial will illustrate linear regression.

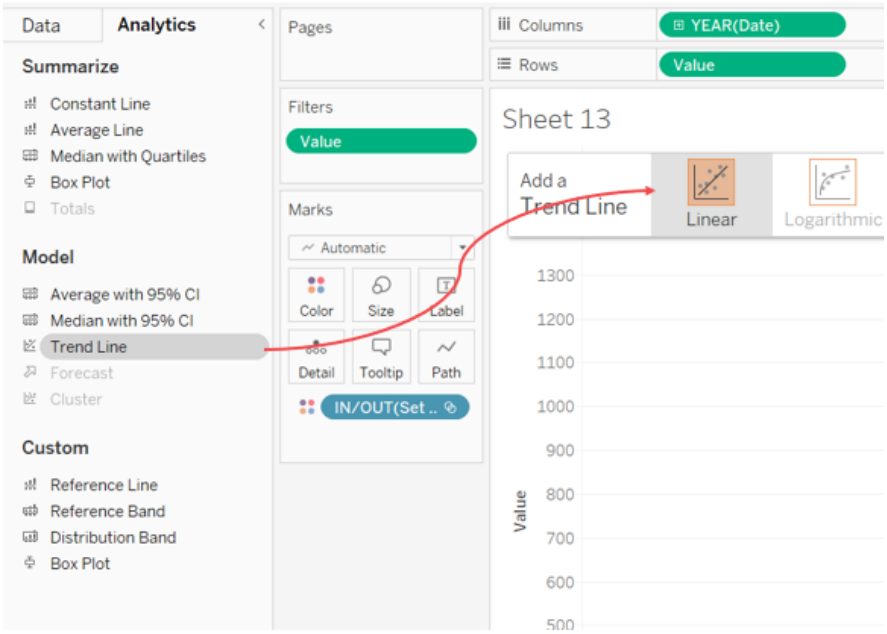



Figure 187: Add regression line

3. The result should look something like this.

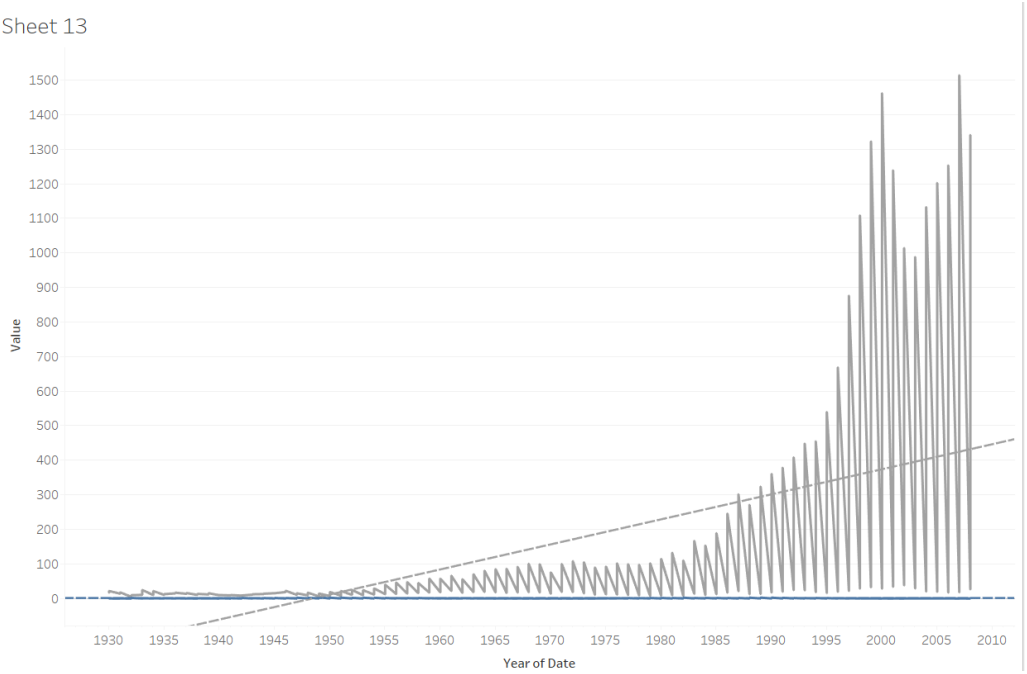



Figure 188: Result

Notice that there is an extra linear line on the graph representing the regression.

4. Tableau has already included the important values that were calculated during the formation of the trend line. When the cursor is brought to the trend line, the values will be displayed at the side of the cursor like shown below.

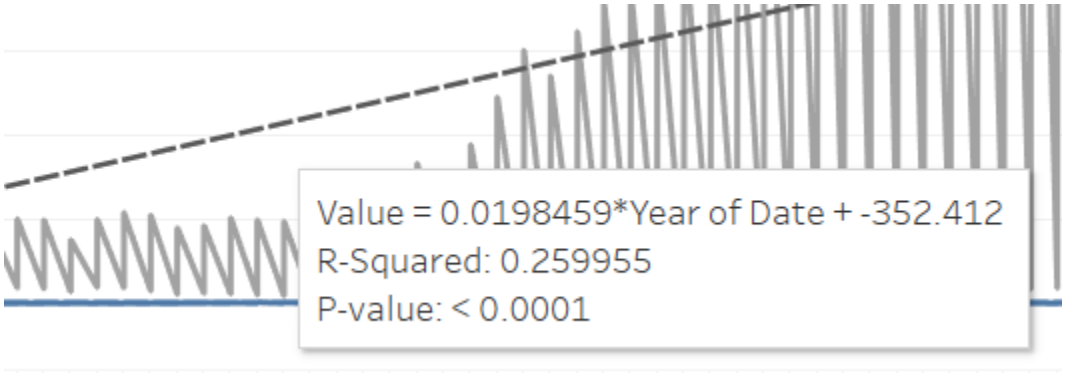



Figure 189: Final result

### 2.5.3.2 Edit a Regression

There are a few different features that Tableau has included to edit the trend line made by the users. The users can see the features just by right-clicking the line then, choose the **Edit All Trend Lines** and you can see the features that can be edited.

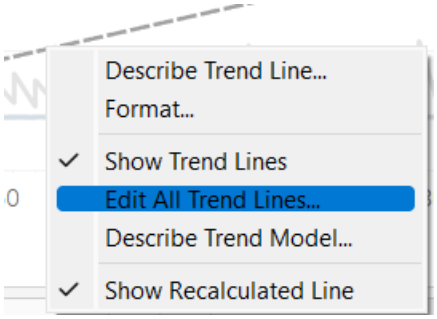



Figure 190: Click Edit All Trend Lines

After clicking on the **Edit All Trend Lines**, below is the display of the things that can be edited.

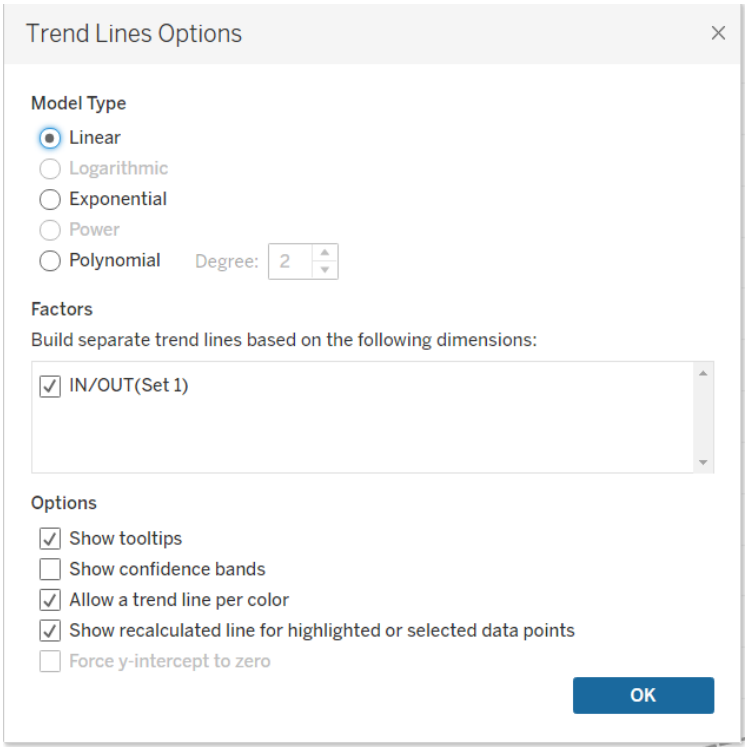



Figure 191: Things that can be edited

### 2.5.3.3 Add Confidence Bands on a Regression

Confidence band represents the lines on the probability plot or fitted plot which depict the upper and lower confidence bounds for all points on a fitted line within the range of data. Tableau has calculated the confidence bands for the users and it is optional for the users either to display it or hide it from the graph. By default, Tableau has been programmed to hide the confidence bands unless the users want to display it. Therefore, this is the steps on how to add confidence bands the graph made.

1. Right-click on the trend line made and choose **Edit All Trend Lines**.

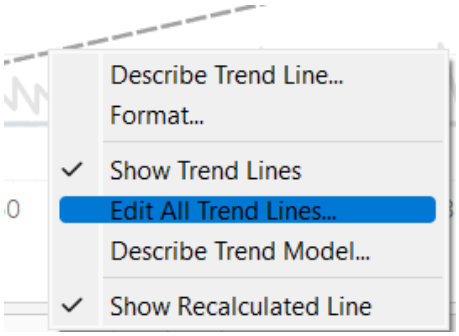



Figure 192: Click Edit All Trend Lines

2. From the **Edit All Trend Lines** dialog box, on the Options section, make sure that the **Show confidence bands** choice is checked (clicked in the small box). So, that is how users can display the confidence bands on the graph. Below is the example on the step and the result of it.

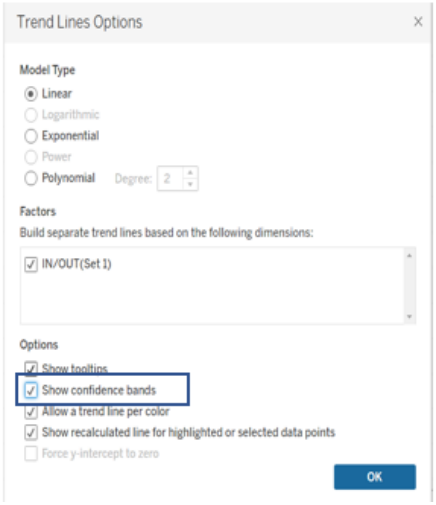



Figure 193: Example on the step and result

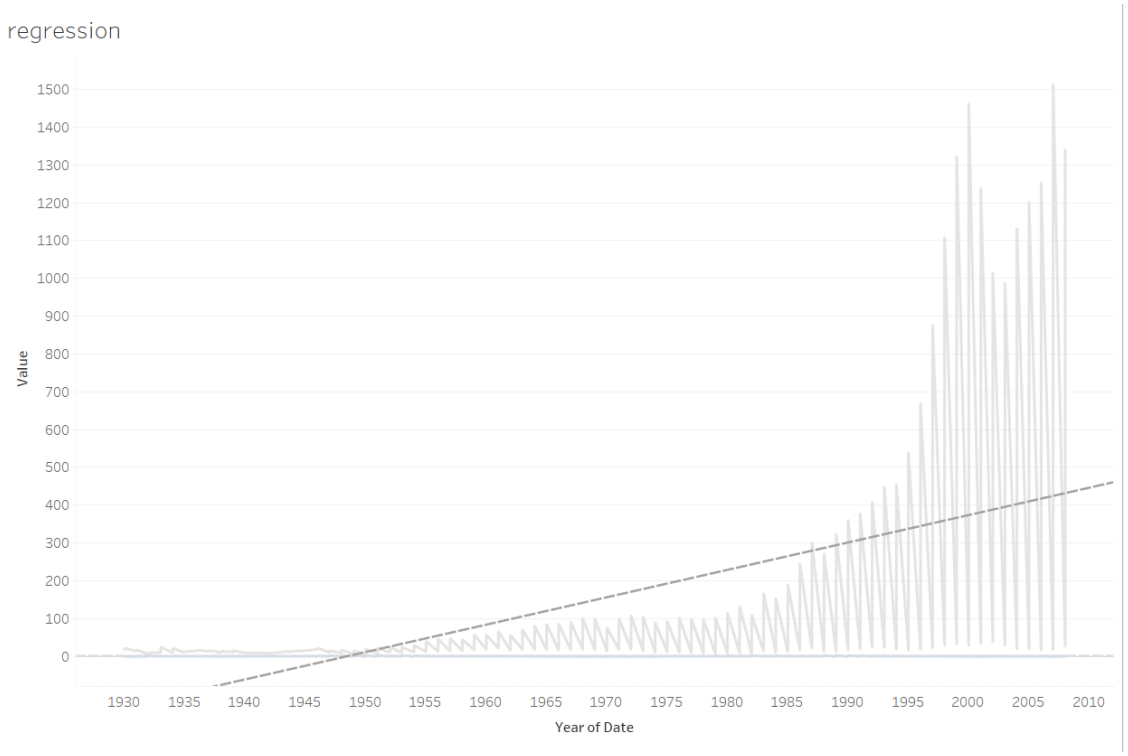



Figure 194: Before adding confidence bands

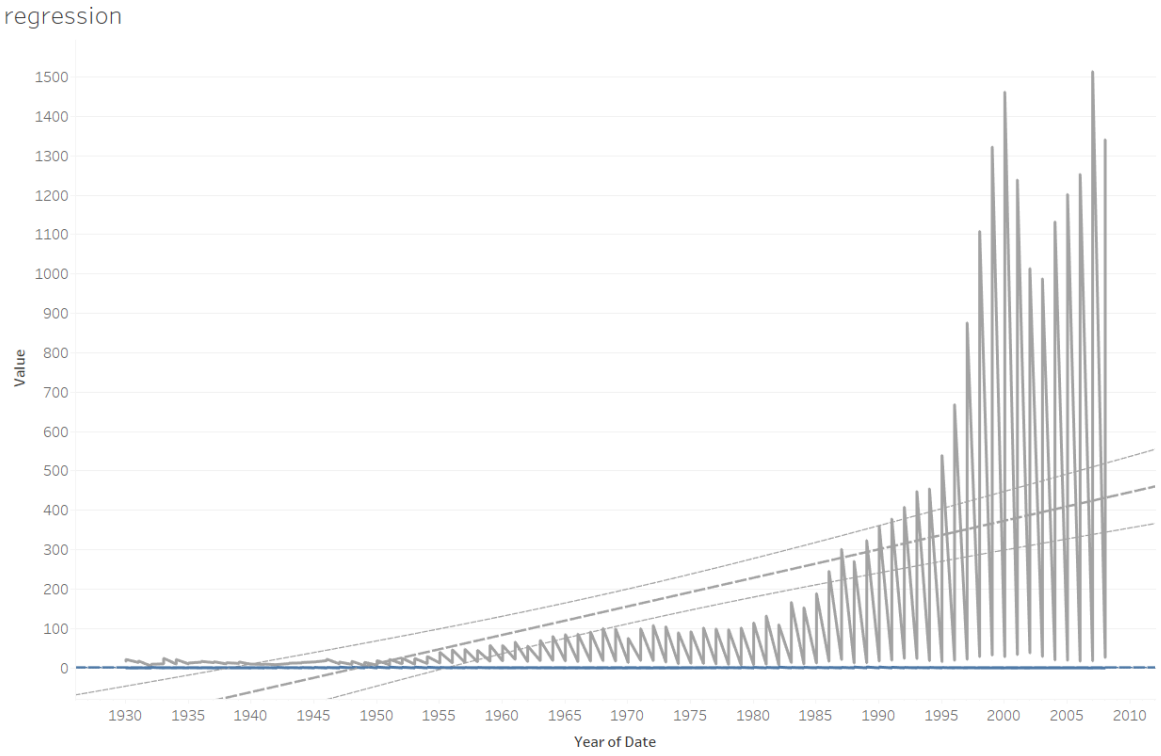



Figure 195: After adding confidence bands

### 2.5.3.4 Description of Trend Lines

Tableau has also included all the calculations made during the formation of the trend line. Some of the calculations include Sum Squared Error (SSE), R-Squared, p-value, ANOVA, and etc. To show the description of the trend line, just right-click on the trend line and choose **Describe Trend Lines**. From there, a dialog box of the trend line descriptions will be popped-up and it can be copied into a written-type document by clicking on the **Copy** button at the left-bottom of the dialog box.

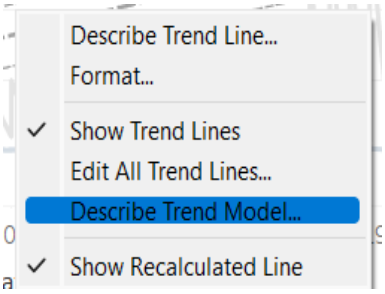



Figure 195: Click Describe Trend Lines

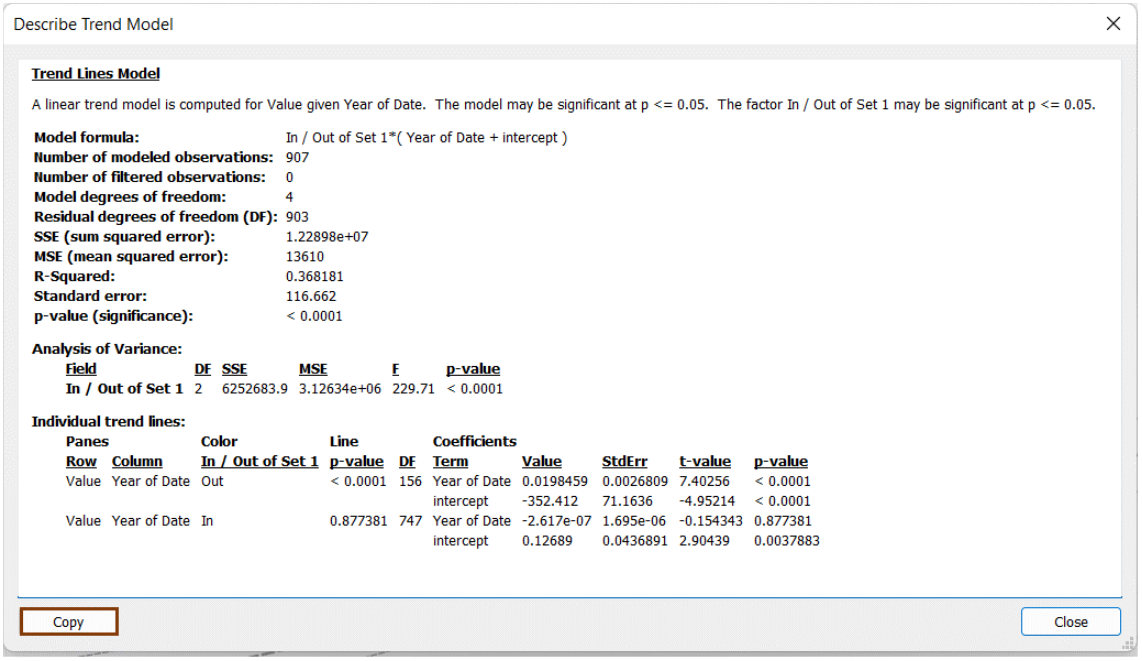



Figure 196: Describe Trend Line dialog box

Below is the list of what all of the values meant in the description dialog box.

| Analysis | Description |
|---|---|
| Model Formula | Formula for the full trend line model |
| Number of Modeled Observations | The number of rows used in the view |
| Number of Filtered Observations | The number of observation excluded from the model |
| Model Degrees of Freedom | The number of parameters needed to completely specify the model |
| Residue Degrees of Freedom (DF) | For a fixed model, this value is defined as the number of observations minus the number of parameters estimated in the model |
| Sum Squared Error (SSE) | The errors are the difference between the observed value and the value predicted by the model. In the Analysis of Variance table, this column is actually the difference between the SSE of the simpler model in that particular row and the full model, which uses all the factors. This SSE also corresponds to the sum of the differences squared of the predicted values from the smaller model and the full model |
| Mean Squared Error (MSE) | The value produced when the SSE quantity divided by its corresponding degrees of freedom |
| R-Squared | R-squared is a measure of how well the data fits the linear model |
| Standard Error | The square root of the MSE of the full model |
| P-value (significance) | The probability that an F random variable with the above degrees of freedom exceeds the observed F in this row of the Analysis of Variance table |
| ANOVA | A table lists information for each factor in the trend line model. The values are a comparison of the model without the factor in question to the entire model, which includes all factors |
| Individual Trend Lines | This table provides information about each trend line in the view individually |

## 2.5.4 Anomaly or Outliers Detection

### 2.5.4.1 Create Outliers for Line Graph

To perform outlier detection obviously requires the user to have a graph or chart. From there, this manual will guide users on how to apply outlier detection (Anomaly) on the graph or chart. To make it easier, this section will show how to detect outliers in line graphs.

1. Create a line graph. Users need to drag at least one variable for each column and row in the view. For this example, a time variable and a value variable will be used.

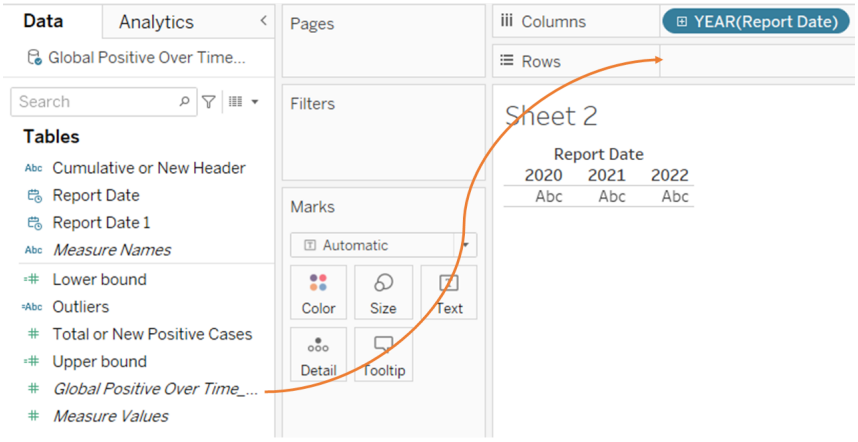



Figure 197: Drag a variable

This is how the line graph would look like.

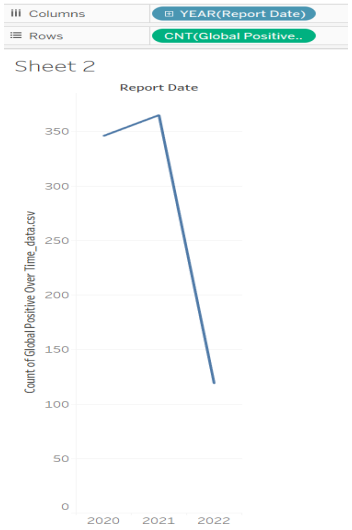



Figure 198: Line graph

Since it looks like too little data in the line graph, the user can change the time variable into the desired time measurement such as year, month, day, hours, and more. So, below is how to change the time measurement.

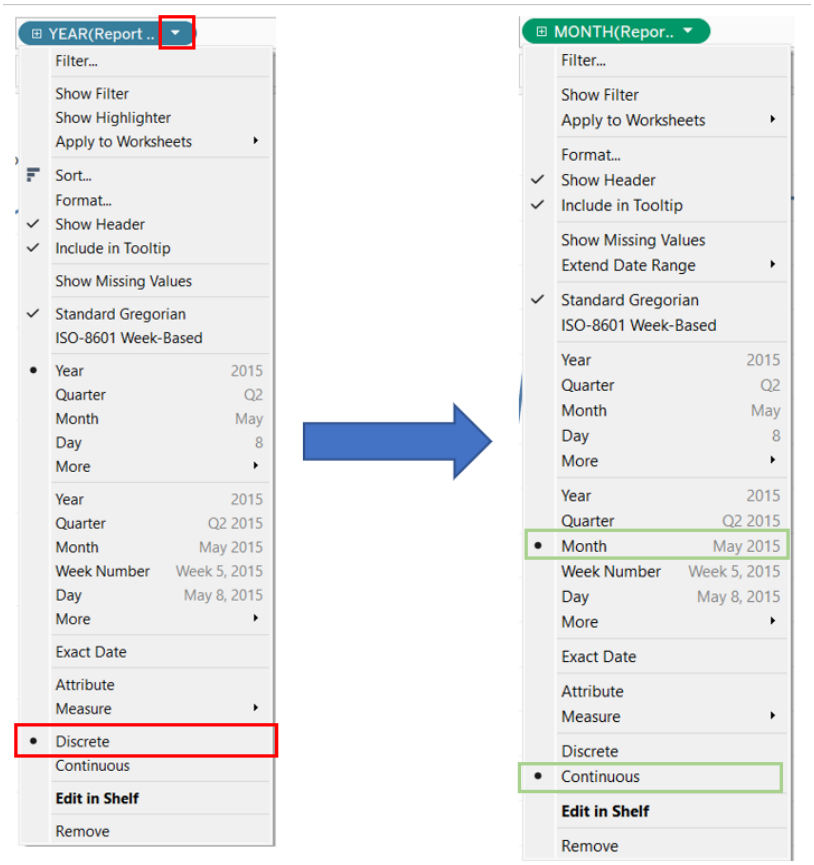



Figure 199: Change Discrete to Continuous

After that the line graph should look better to work with.

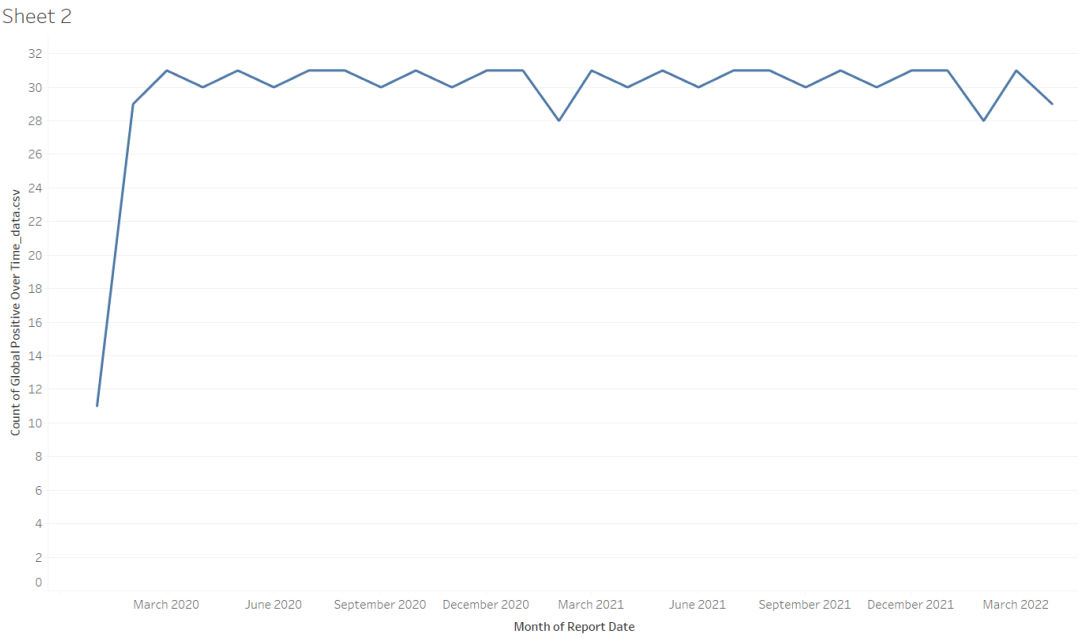



Figure 200: Better line graph

2. Set the average, lower bound and upper bound for the graph created. To create these fields, users need to use the **Create** function and choose **Calculate Field**. Then, a dialog box will appear for the user to put the formula used, so Tableau can calculate it automatically. Take note that these formulas will use WINDOW_AVG function and WINDOW_STDEV function and all variables used in this is the variable that is **not** a time measured variable.

   To create an average field, users can use this formula.


Figure 201: Create average

Basically, the formula is **WINDOW_AVG(sum(variable))**.
To create the upper bound, users can use this formula shown below.


Figure 202: Formula

The formula will consist of the average calculated above and the standard deviation of the variable. The formula is **[Average field name] + WINDOW_STDEV(sum(variable))**.

Similar to the upper bound formula, the lower bound formula is **[Average field name] - WINDOW_STDEV(sum(variable))**.

3. Next, copy and paste the same variable (the one that is not the time measured variable) into the row, to create another graph in the view.

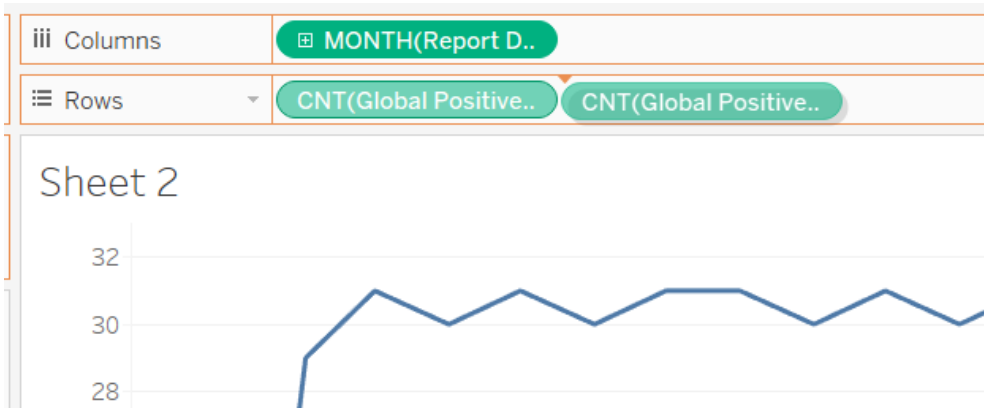



Figure 203: Copy variable

Users also need to change the chart to dual axis chart, by right-clicking on the right y-axis and selecting dual axis. Do not forget to synchronize the axis of the charts or otherwise the chart will look odd. For the second line graph, do change the chart type to circle (or any preferable chart type as long as it is not the same as the first one).

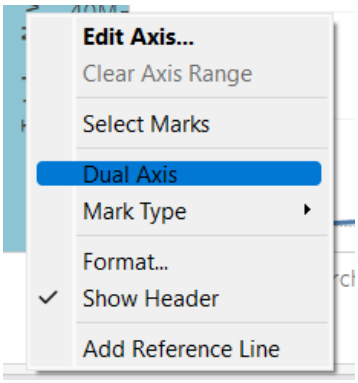



Figure 204: Click Dual Axis

Choosing Dual Axis will produce a view like below.

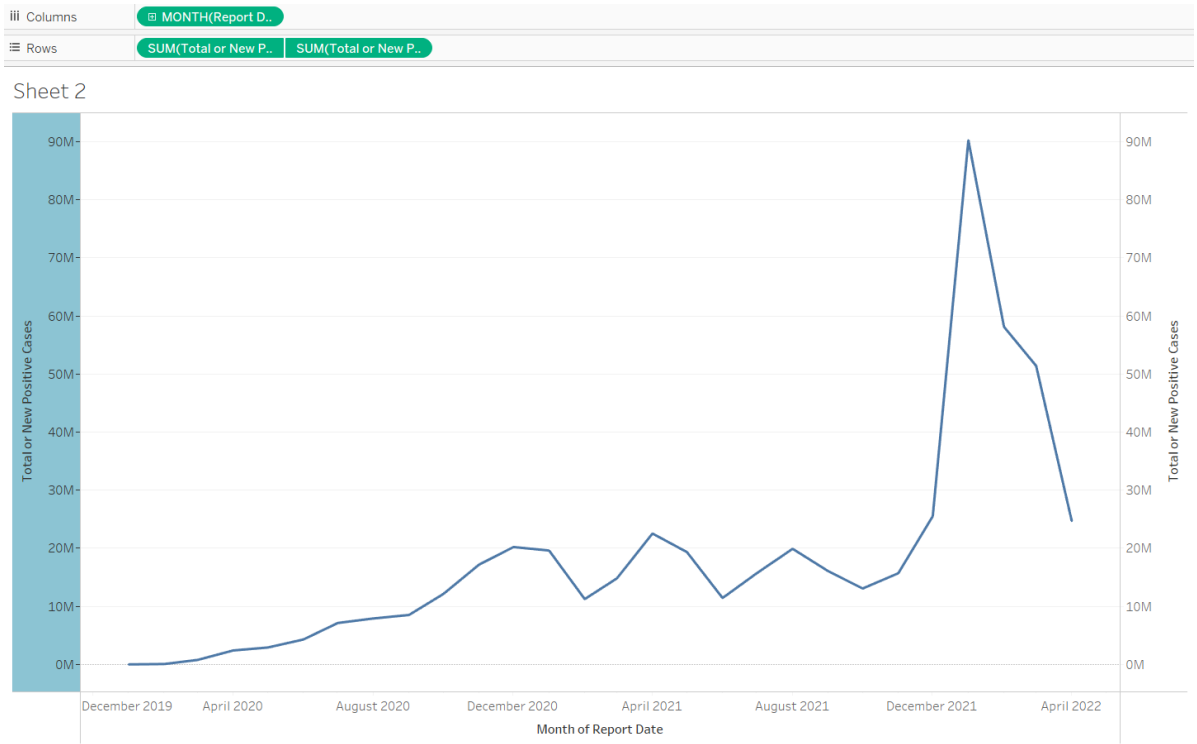



Figure 205: Dual axis chart

Next is to synchronize the axis.

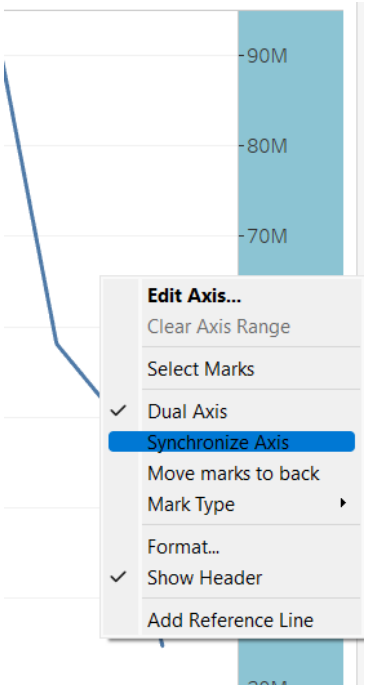



Figure 206: Click Synchronize Axis

Since the axes are synchronized, users can also choose to hide the other axis by unchecking the button **Show Header**.

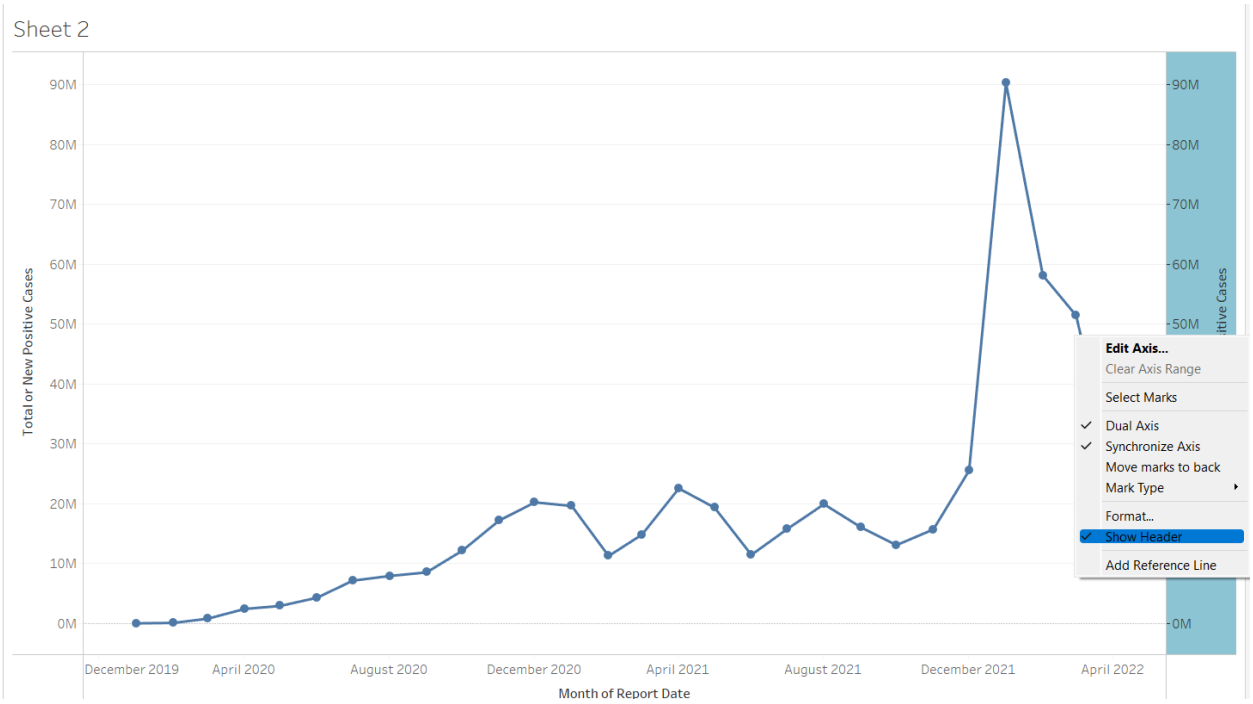



Figure 207: Click on show header to hide it

The view will look like this after that.

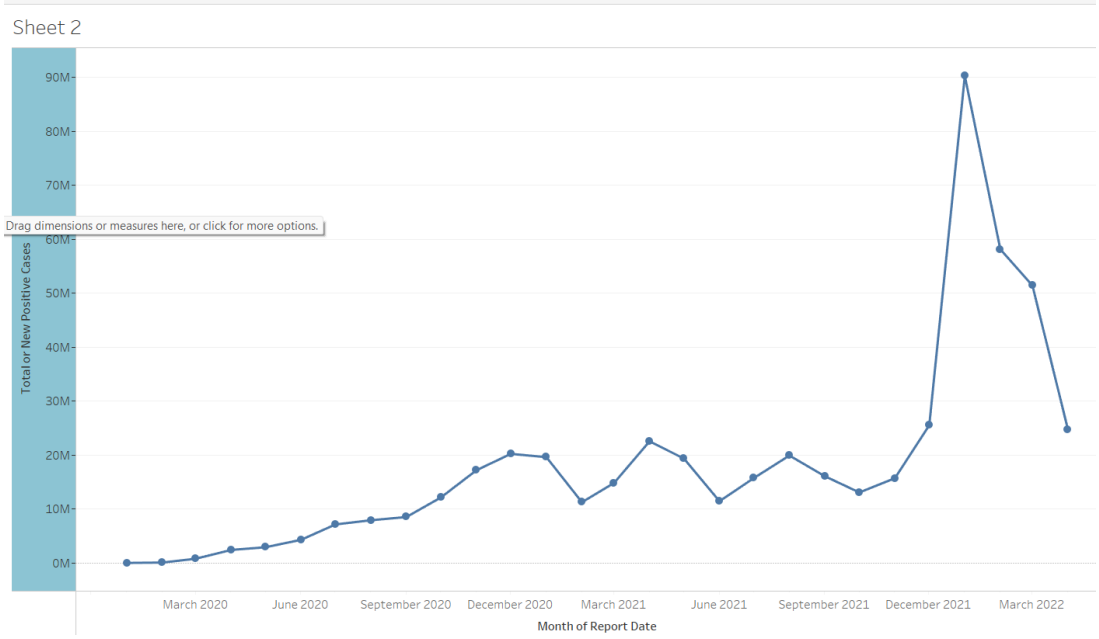
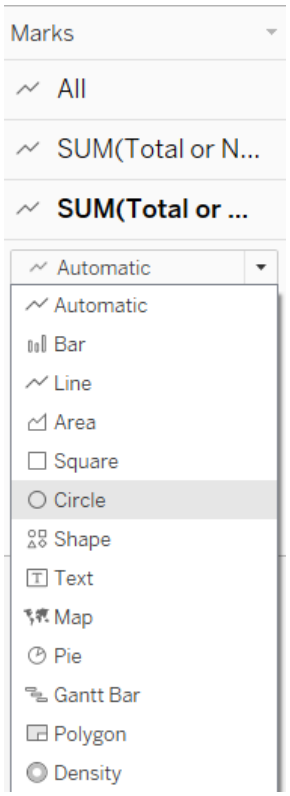
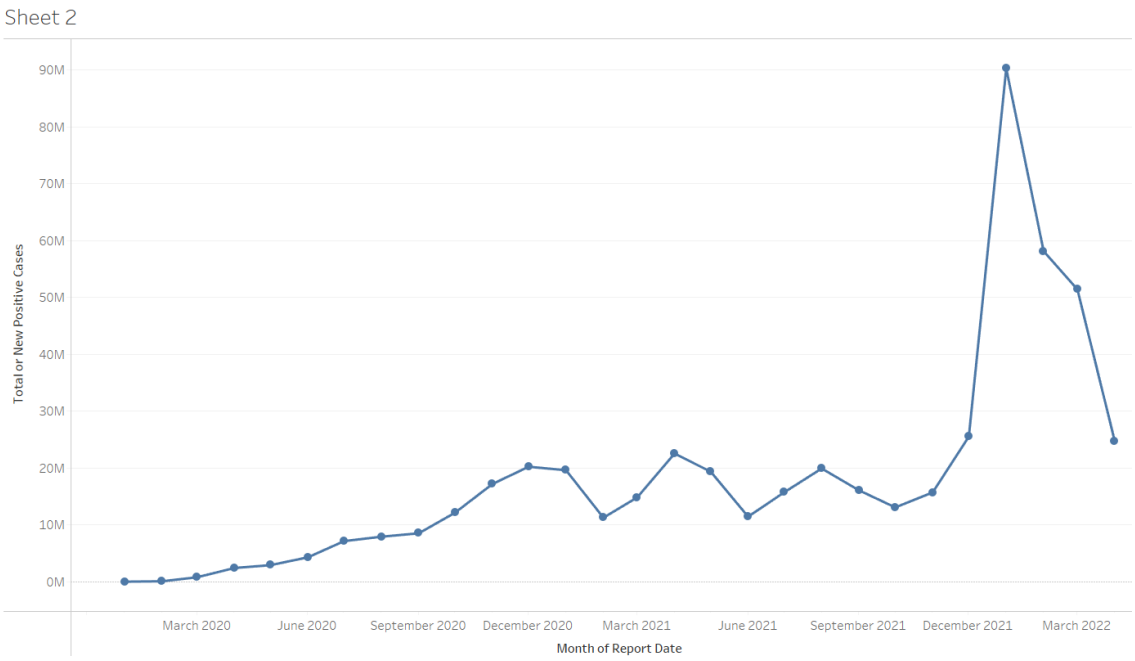



Figure 208: Better view

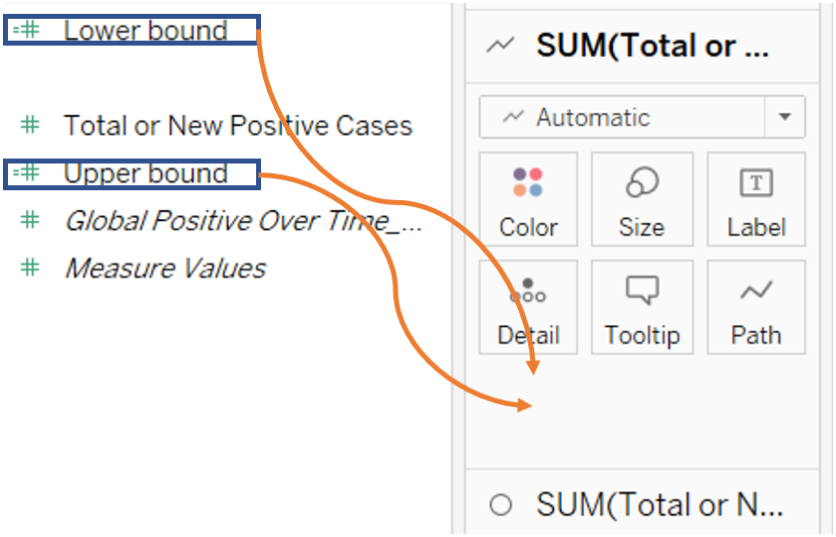
Changing the chart type to circle at **Marks** tab.



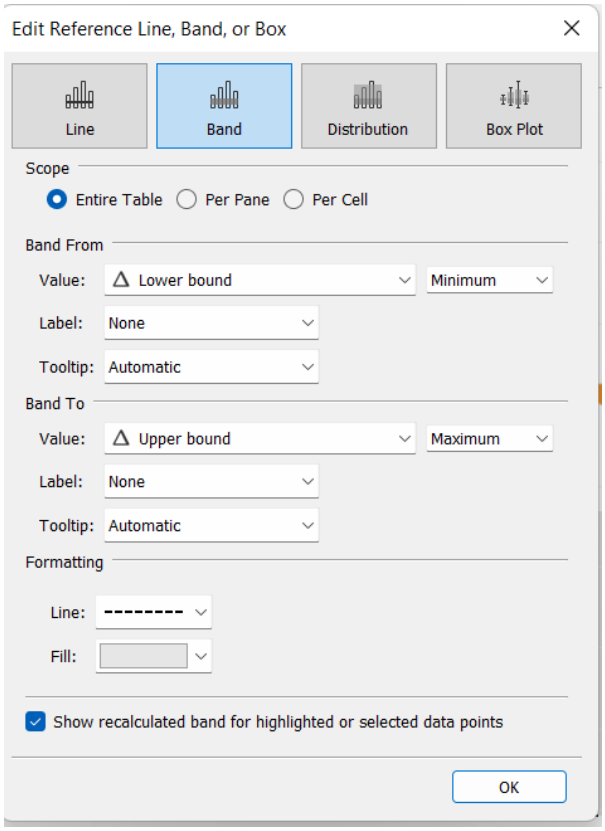
Then finally, the view will display the chart like shown below,



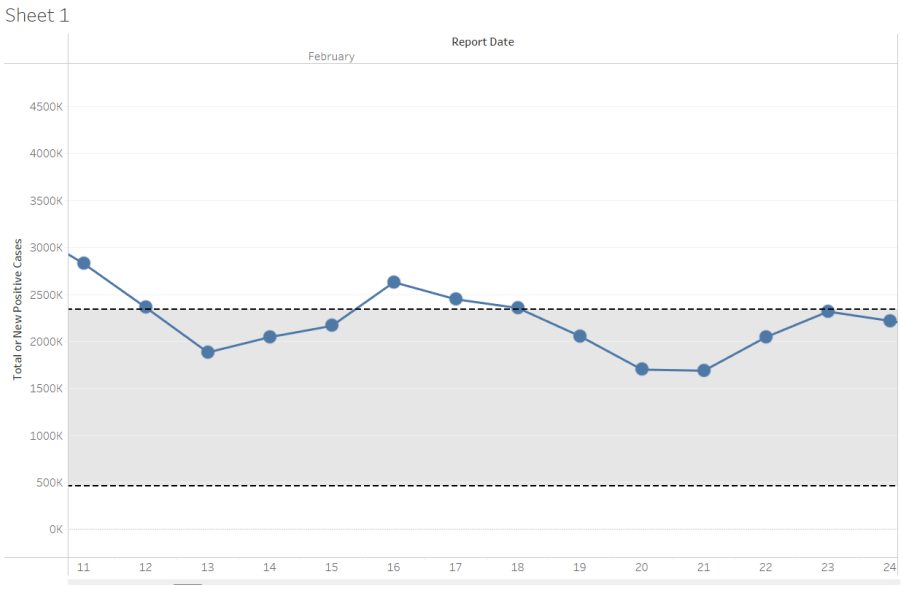

4. Add the upper and lower bound into the view by dragging it to the **Marks** tab.



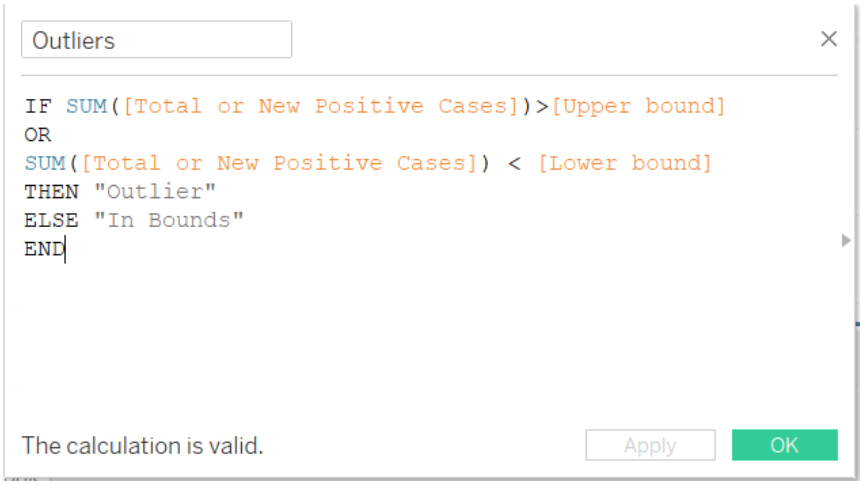
5. Right-click on the non-time measured variable axis and select **Add Reference Line**. Do follow the settings set below.



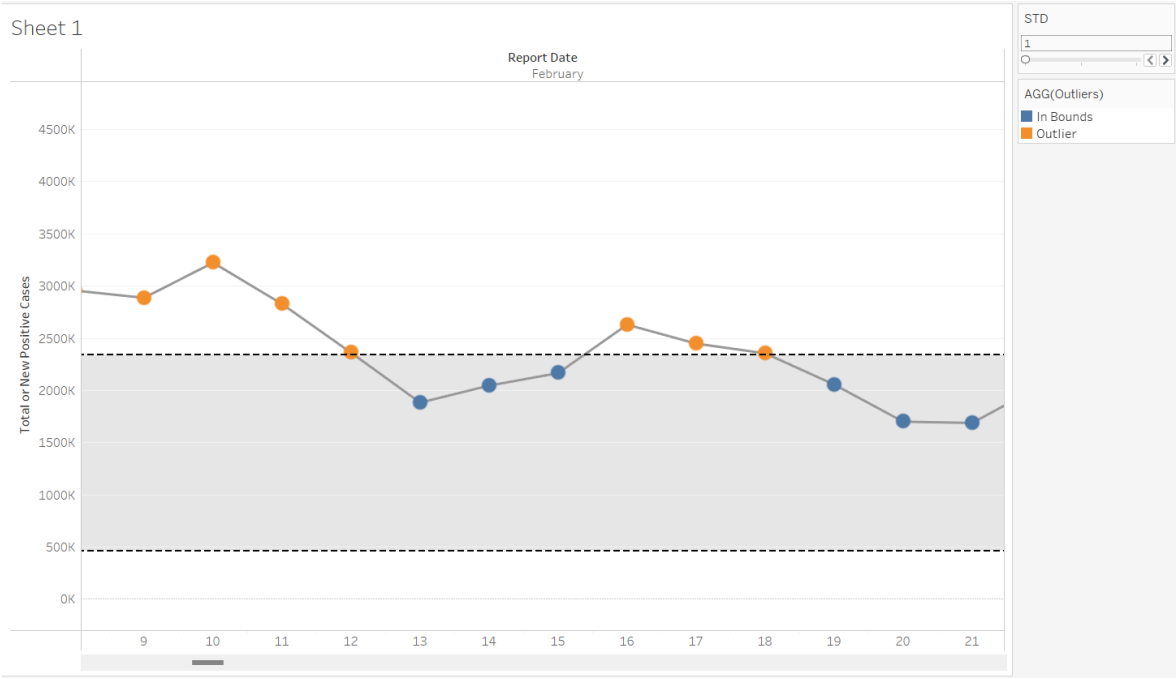

Then, the graph will be updated like shown.



6. The next step is where the outlier will be applied in the graph. To apply the outliers, users can create a Boolean calculated field.




The THEN and ELSE statement can be ignored if the users found it as something that they do not need because Tableau can already differentiate the outliers of the data automatically. Below is the result after applying the outlier. So, that is how users can create an outlier.



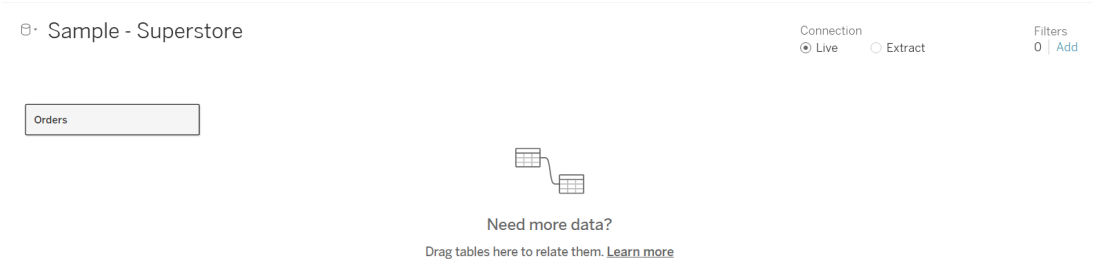

Note: As it can be seen, on the figure above, there is a tab of **'STD'** on top-right of the figure. This is a parameter, which can be set by the users if they want to.

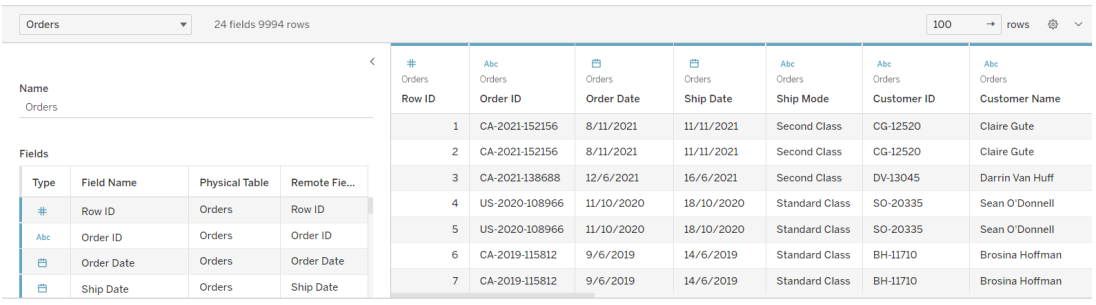
## 2.5.5 Market Basket Analysis (Association Rule Mining)

### 2.5.5.1 Set the Dataview Parameter

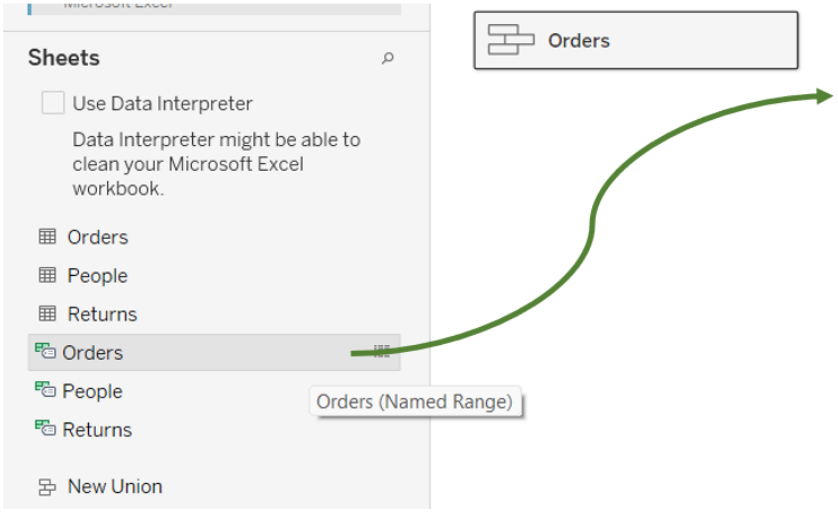
Setting the parameters first is an important step when performing a market basket analysis. In Tableau, there are a few steps that should be performed to set the dataview parameters. Hence, below are the steps to do so.

1. On the dataset tab, drag a table into the data area. In this case, the table is named ‘Orders’.

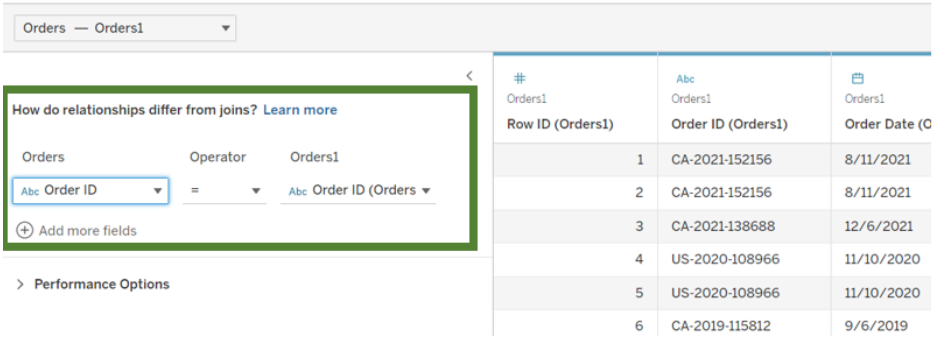



2. Make a self-join on the table (in this case ‘Orders’ table) by dragging a second instance of the orders table to the data area. Then, select any attribute in the table (we use Order ID in this case) as the first relationship field. Set the operator as equal (symbol: =).

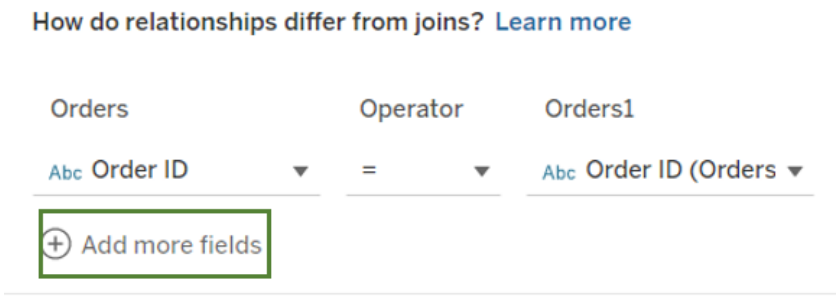


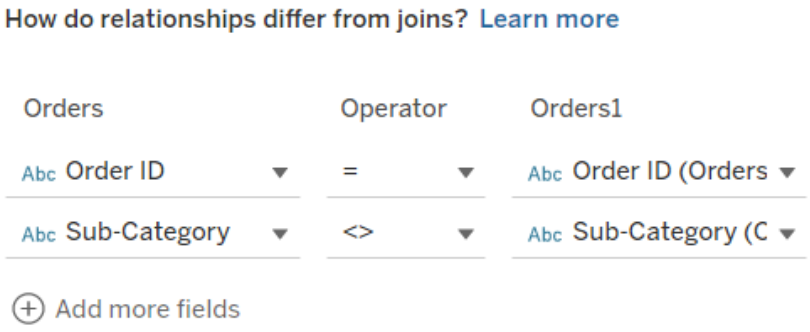





Please make sure that the relationship field is filled or else it is going to be an error during this process.

3. Then add the second relationship field. In this example, the attribute used in the second relationship field is Sub-Category. Set the operator as not equal (symbol: < >).



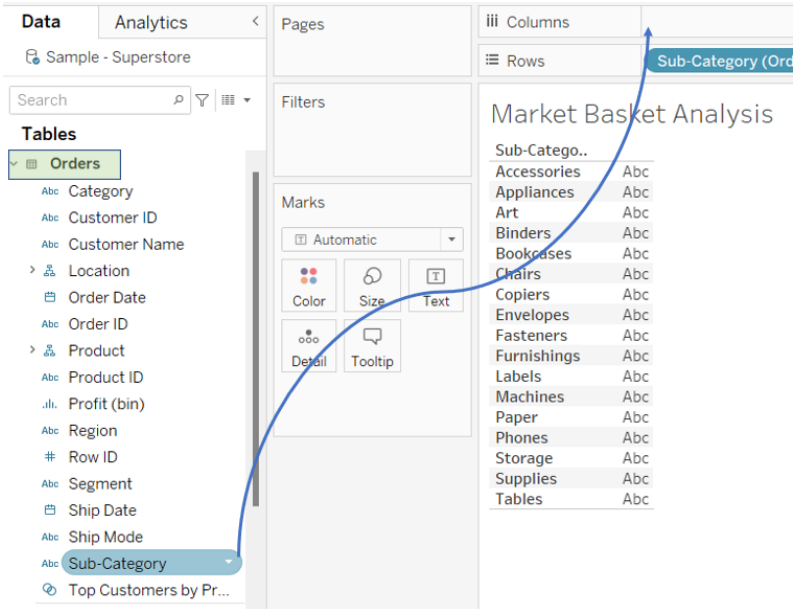



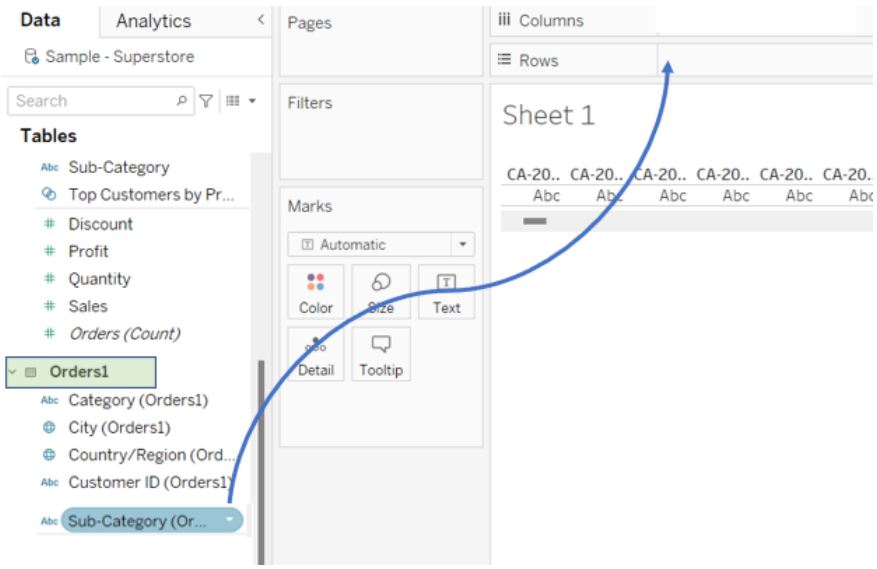

4. Then, users can move on to the first worksheet which means that the process of setting the dataview parameter ends here.

## 2.5.5.2 Create Data Visualization

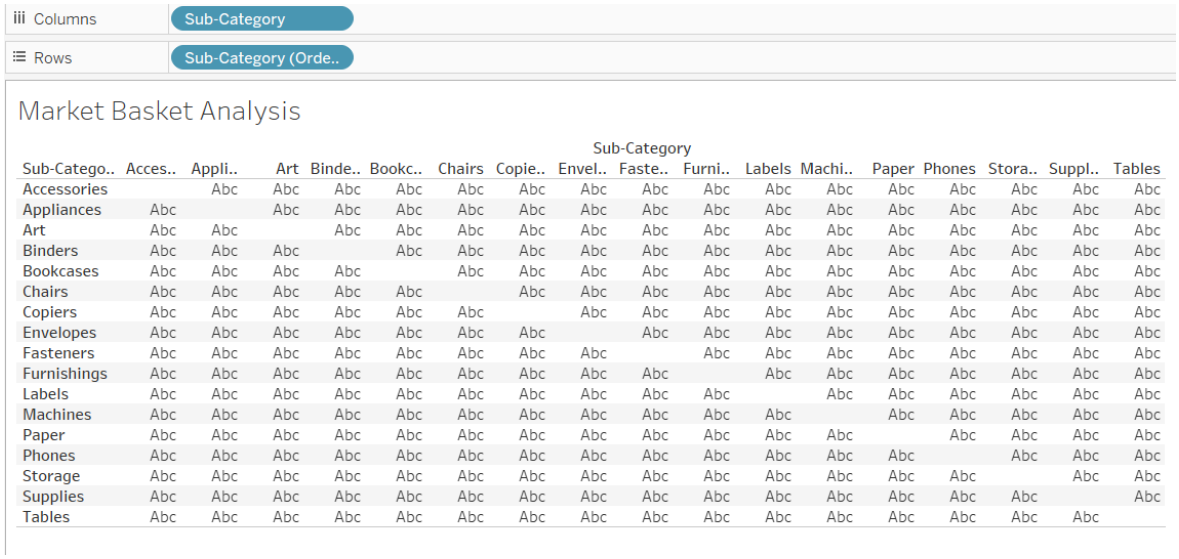
These are the steps on how to visualize the data for market basket analysis.

1. From the relationship created at 2.5.5.1, choose one of the two attributes used based on the relationship, and drag the same attribute from different tables to the columns and rows field. In this example, the used attribute is Sub-Category from table ‘Orders’ and Sub-Category from table ‘Orders1’.



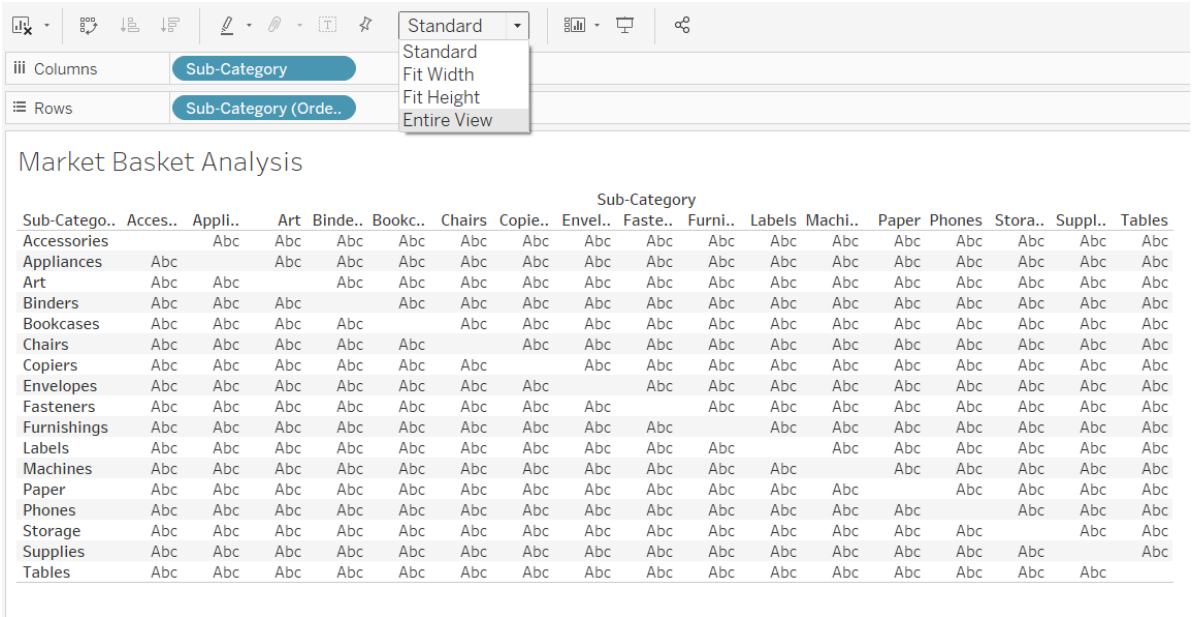




Then, after applying both attributes, the view should look like this.

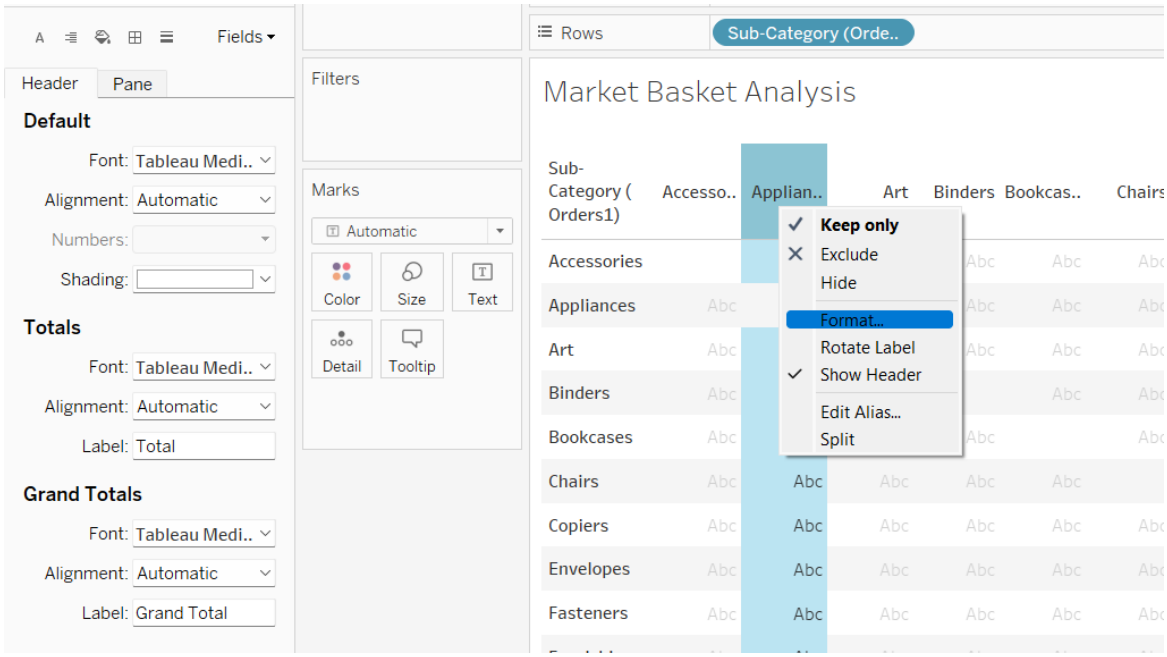


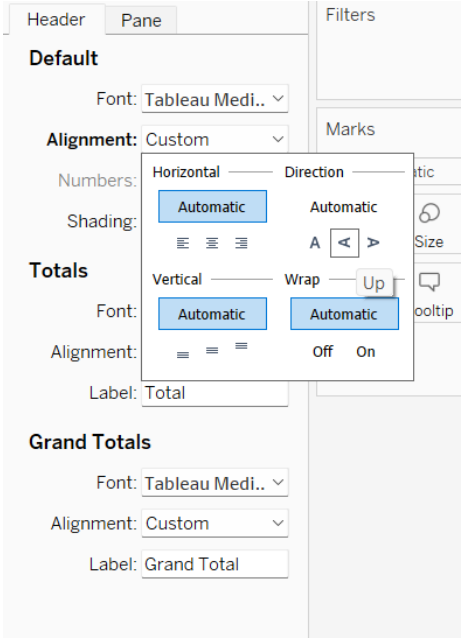

2. You can choose to ignore this step if you find it unnecessary since this step shows how to make the figure in the view look well-structured.

   Firstly, change the view from **Standard** to **Entire View**.




Then, rotate one of the title axes to make it look nicer.



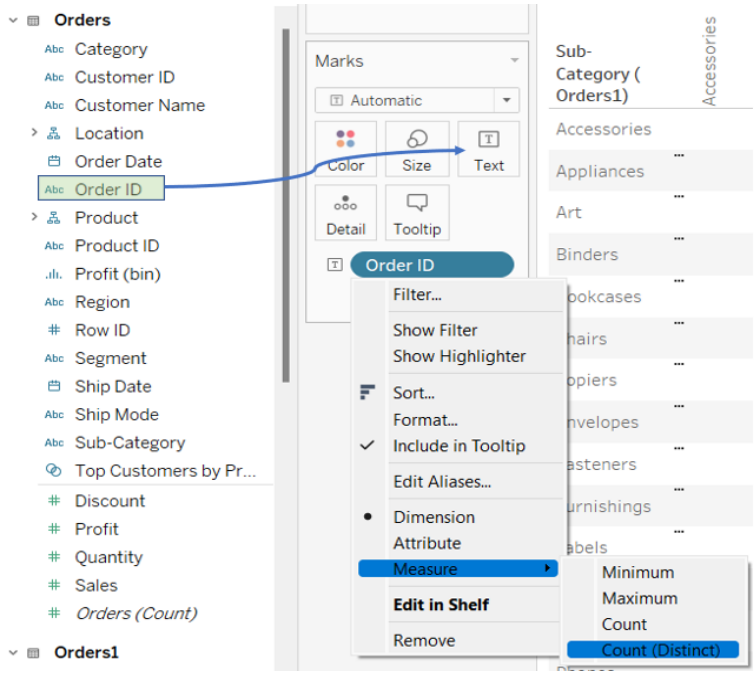




The end result will look like shown below.

Market Basket Analysis

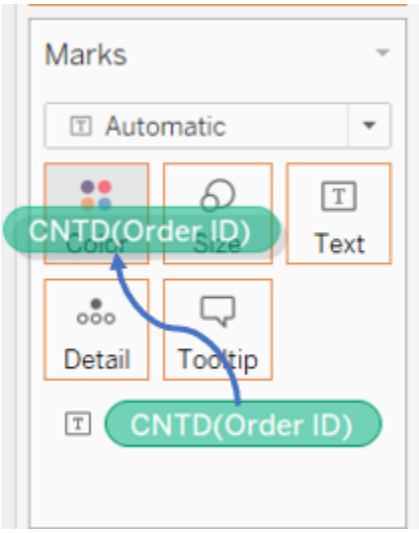
| Sub-Category ( Orders1) | Accessories | Appliances | Art | Binders | Bookcases | Chairs | Copiers | Envelopes | Fasteners | Furnishings | Labels | Machines | Paper | Phones | Storage | Supplies | Tables |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Accessories | | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc |
| Appliances | Abc | | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc |
| Art | Abc | Abc | | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc |
| Binders | Abc | Abc | Abc | | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc |
| Bookcases | Abc | Abc | Abc | Abc | | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc |
| Chairs | Abc | Abc | Abc | Abc | Abc | | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc |
| Copiers | Abc | Abc | Abc | Abc | Abc | Abc | | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc |
| Envelopes | Abc | Abc | Abc | Abc | Abc | Abc | Abc | | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc |
| Fasteners | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc |
| Furnishings | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | | Abc | Abc | Abc | Abc | Abc | Abc | Abc |
| Labels | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | | Abc | Abc | Abc | Abc | Abc | Abc |
| Machines | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | | Abc | Abc | Abc | Abc | Abc |
| Paper | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | | Abc | Abc | Abc | Abc |
| Phones | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | | Abc | Abc | Abc |
| Storage | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | | Abc | Abc |
| Supplies | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | | Abc |
| Tables | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | Abc | |

3. Drag the Order ID field from the ‘Order’ table into **“Text”** in the **Marks** tab. Set it as a **Count (Distinct)** measure.

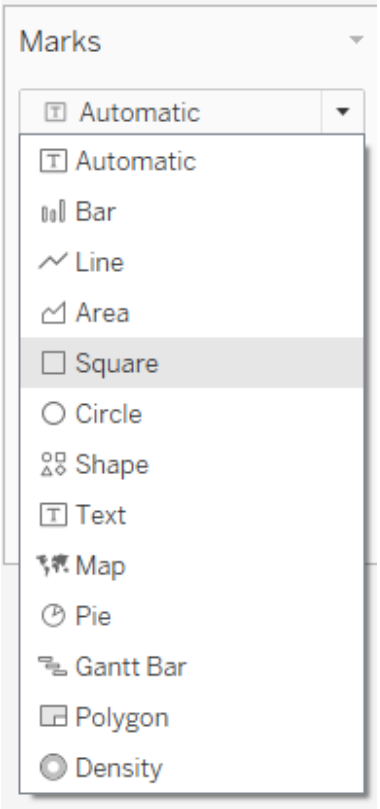


4. Pick a copy of Order ID and drag it to the **'Color'** in the **Marks** tab, then, set it as a measure.

   Note: Do not forget while dragging the 'Order ID' attribute to the 'Color' field, press **ctrl** on your keyboard.




   Then, adjust it to square to see it clearly.



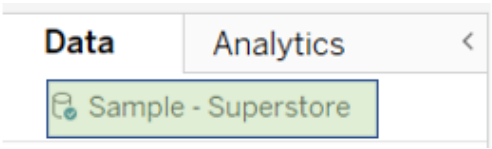
The view now should look like the figure below.

Market Basket Analysis

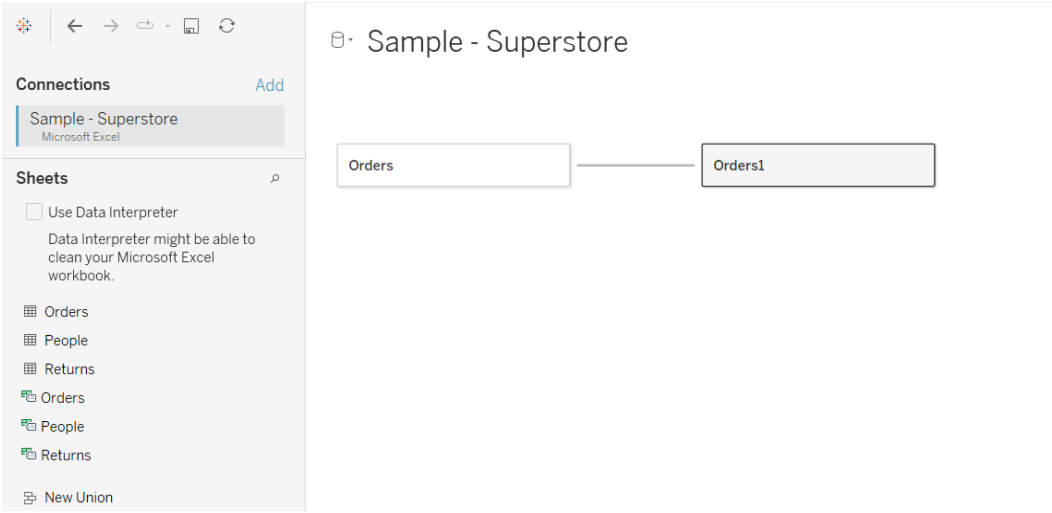
| Sub-Category ( Orders1) | Sub-Category: Accessories | Appliances | Art | Binders | Bookcases | Chairs | Copiers | Envelopes | Fasteners | Furnishings | Labels | Machines | Paper | Phones | Storage | Supplies | Tables |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Null | 196 | 112 | 251 | 420 | 56 | 175 | 18 | 67 | 55 | 273 | 94 | 24 | 365 | 200 | 219 | 43 | 95 |
| Accessories | | 60 | 89 | 161 | 29 | 64 | 6 | 32 | 27 | 115 | 47 | 22 | 153 | 116 | 103 | 25 | 44 |
| Appliances | 60 | | 59 | 130 | 13 | 47 | 9 | 21 | 21 | 80 | 21 | 9 | 109 | 76 | 56 | 24 | 30 |
| Art | 89 | 59 | | 150 | 34 | 82 | 7 | 27 | 30 | 100 | 47 | 12 | 150 | 124 | 100 | 29 | 38 |
| Binders | 161 | 130 | 150 | | 56 | 119 | 19 | 54 | 62 | 198 | 82 | 30 | 275 | 199 | 199 | 43 | 71 |
| Bookcases | 29 | 13 | 34 | 56 | | 21 | 2 | 11 | 7 | 30 | 21 | 6 | 48 | 42 | 37 | 10 | 9 |
| Chairs | 64 | 47 | 82 | 119 | 21 | | 10 | 29 | 26 | 99 | 36 | 17 | 133 | 89 | 83 | 19 | 31 |
| Copiers | 6 | 9 | 7 | 19 | 2 | 10 | | 4 | 3 | 12 | 6 | 2 | 20 | 11 | 12 | 4 | 2 |
| Envelopes | 32 | 21 | 27 | 54 | 11 | 29 | 4 | | 11 | 32 | 11 | 6 | 57 | 41 | 38 | 5 | 12 |
| Fasteners | 27 | 21 | 30 | 62 | 7 | 26 | 3 | 11 | | 41 | 15 | 8 | 59 | 37 | 37 | 11 | 10 |
| Furnishings | 115 | 80 | 100 | 198 | 30 | 99 | 12 | 32 | 41 | | 57 | 27 | 177 | 148 | 138 | 31 | 44 |
| Labels | 47 | 21 | 47 | 82 | 21 | 36 | 6 | 11 | 15 | 57 | | 8 | 80 | 58 | 61 | 18 | 17 |
| Machines | 22 | 9 | 12 | 30 | 6 | 17 | 2 | 6 | 8 | 27 | 8 | | 28 | 22 | 21 | 1 | 6 |
| Paper | 153 | 109 | 150 | 275 | 48 | 133 | 20 | 57 | 59 | 177 | 80 | 28 | | 175 | 178 | 45 | 54 |
| Phones | 116 | 76 | 124 | 199 | 42 | 89 | 11 | 41 | 37 | 148 | 58 | 22 | 175 | | 117 | 22 | 52 |
| Storage | 103 | 56 | 100 | 199 | 37 | 83 | 12 | 38 | 37 | 138 | 61 | 21 | 178 | 117 | | 31 | 48 |
| Supplies | 25 | 24 | 29 | 43 | 10 | 19 | 4 | 5 | 11 | 31 | 18 | 1 | 45 | 22 | 31 | | 11 |
| Tables | 44 | 30 | 38 | 71 | 9 | 31 | 2 | 12 | 10 | 44 | 17 | 6 | 54 | 52 | 48 | 11 | |

5. As it can be seen, up until step 4, the final figure inside the view looks like the data is being mirrored. It can either be removed manually or just go back to the part where we set the dataview. The steps are shown below.

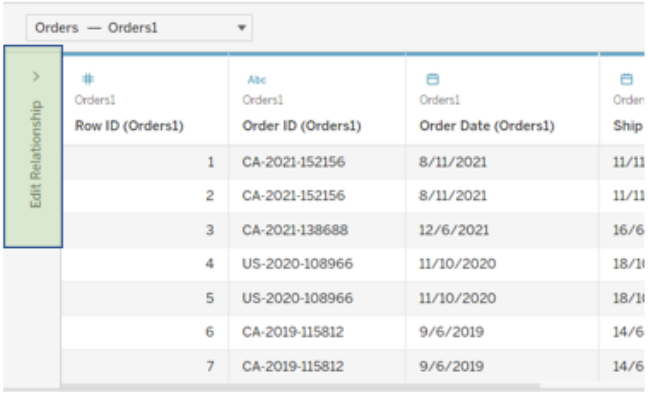
   Double clicked on the data source to display the data source page.



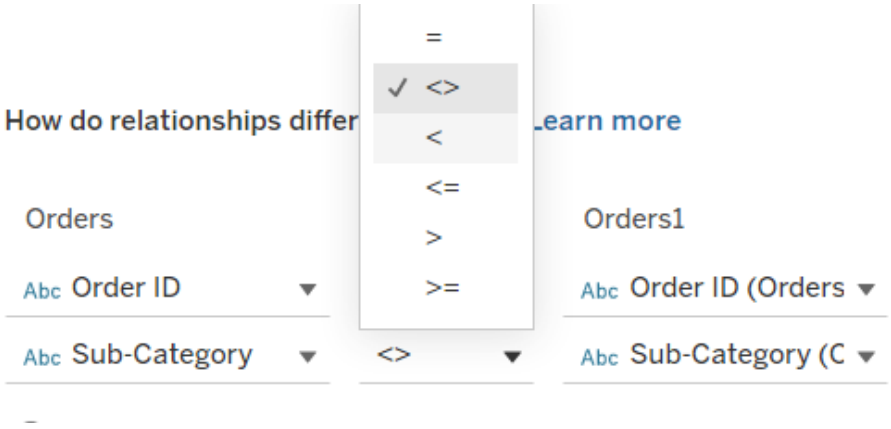
This is the data source page.



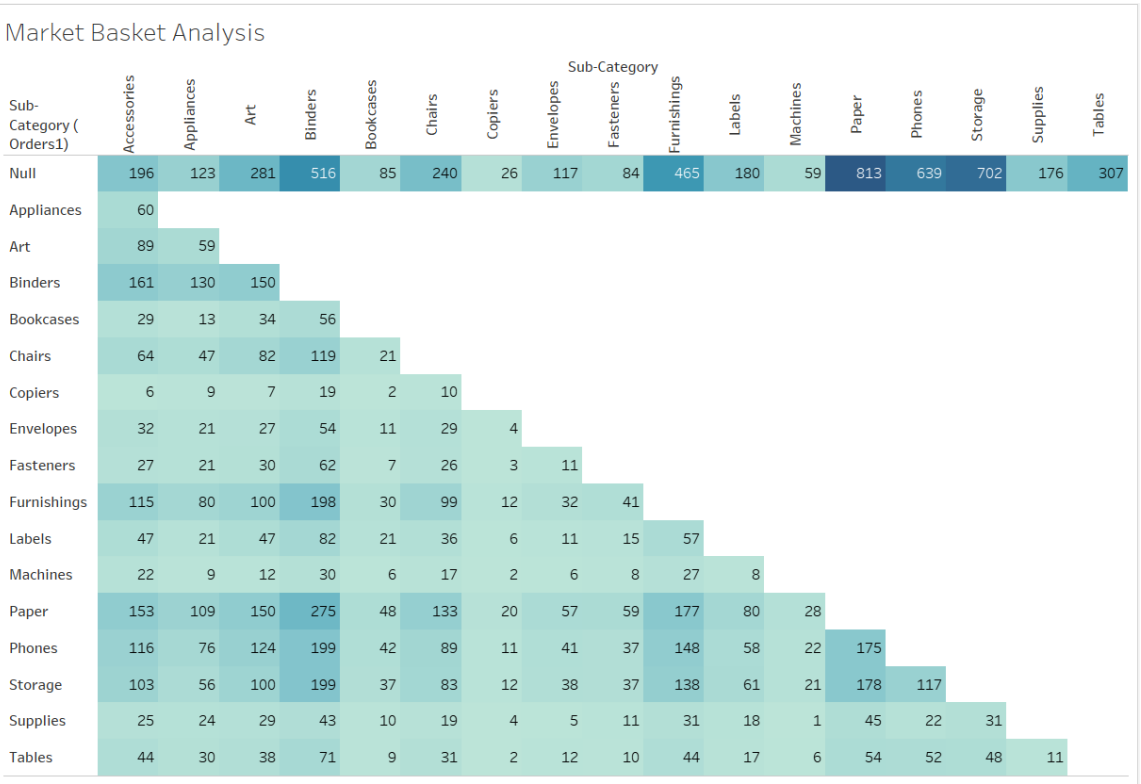

Notice that, below the relationship diagram, there is a tab called **Edit Relationship**. Click the tab.

Change the not equal relationship to less than (<).




Then, go back to the view to see the difference.




Note that now, the mirrored values are now disappeared and only display one value for each column and rows.

Users can also drag other attributes into the Marks view, to see the effects that that particular attribute brings to this analysis. Therefore, this is how users can perform a Market Basket Analysis.

# 3.0 ADVANTAGES OF TABLEAU

## 3.0 Advantage of Tableau

As mentioned, Tableau is a very popular data visualization tool for creating visually appealing and simple dashboards and reports. As a result, many people want to learn Tableau and take advantage of its unique features to complete their tasks. With that being said, let's take a look at some of Tableau's unique features and see what makes it so superior.

1. An excellent data visualization tool

It's the most advanced data visualization software available. As a result, it makes it easier to create excellent, interactive dashboards and reports that provide more information than reports created with spreadsheets or other tools. Complex data blending, computations, and dashboard creation are all possible with it. This is why it is the best data visualization tool available.

2. Fastest way to create interactive visualizations

The biggest benefit of Tableau is that it allows you to create stunning dashboards in a short amount of time. With the help of quick drag and drop functions, the user can quickly create a number of interactive reports. It can quickly create dashboards with a variety of variables.

3. Easy to implement

Tableau offers a variety of visualization options to improve the user experience. This tool is easier to learn and implement than Python, Business Objects, or Domo. The most significant benefit of this tool is that it can be learned and mastered by anyone who has no prior coding experience.

4. Working with huge data sets is made easy

Tableau's ability to work with large amounts of data is one of the main reasons for its use. It can effortlessly handle millions of data rows. Users can create a variety of data visualizations from Big Data without affecting the speed and performance of the dashboards. It also has the ability to connect to a variety of data sources, including SQL and others.

5. Mobile device companion app support

As we all know, smartphone usage is increasing by the day. As a result, having a tool that allows mobile users to access and view reports and dashboards is critical. Tableau is one such tool that recognises the platform a user is using and optimizes reports accordingly, providing the best viewing option to the user.

6. Supports multiple scripting languages

To solve the performance problem or speed up computations, users can use the R and Python programming languages. These measures assist in reducing the load on Tableau. Tableau, on the other hand, does not support Python as a native language. However, Python's visual packages can be imported.

# 4.0 DISADVANTAGES OF TABLEAU

## 4.0 Disadvantage of Tableau

Tableau is a very well-known software in the Business Intelligence (BI) field and for data visualization for its amazing functionalities. In spite of all of the supreme features and advantages that Tableau successfully provides, Tableau still has some limitations and drawbacks. Therefore, we are going to explore the limitations and the drawbacks of Tableau that can be a problem for certain individuals.

1. Expensive

Tableau might not be as expensive as some of data visualization softwares such as softwares produced by Oracle or IBM, however, despite it, Tableau can also be considered as costly for the users. Tableau can cost up to $70 per month for each user. Tableau may have some different products according to who their subscribers are, however it is still pricey considering that some of the other products have limitations. Small and medium companies may be affected by the cost of subscribing to Tableau since the price is calculated according to the product they use and the number of the users of the product since the price stated is for per user.

2. Security Issues

Even if it's not all, most data is confidential and must be protected from any types of threats such as being stolen or hacked. Regardless of Tableau's concern for their clients' information safety, Tableau has never provided centralized data-level security for their clients. Tableau only provides row-level security, which specifies that every user should have his/her own account to access data. With a large number of accounts being created in Tableau, it can escalate the possibility of the system being hacked.

3. Bad Versioning

It can be seen that only the recent versions of Tableau support revision history, while for older versions, it seems like it is an impossible task for them.

4. Absence of Scheduling Options

Users are not provided with the option of refreshing with the help of automatic scheduling. Since Tableau does not provide this function, users need to manually refresh it rather than it is done automatically.

5. Shortfall in Data Modeling and Data Dictionary Capabilities

This can be related to Data Analysts job scope as data modeling is one of the key importance in classification analysis.

6. No custom visual import

Compared to the other visualization tools, developers can import the visuals rather than need to recreate it all over again. When using Tableau, users must recreate the visuals all over again since Tableau only gives the users the freedom to import only the data and not the visuals. Therefore, it may cost the users a lot of their time being wasted just to reconstruct the visuals again in Tableau.

7. Awful After-Sales Support

From the past years and up until now, numerous messages from the Tableau forum can be seen made by the users complaining that Tableau software lacks proper after-sales maintenance. On multiple occasions, when users/customers faced a problem with the software and reached out to the support team, the support team did not solve the matter by investigating the problem's root and eliminating it. Most of the time, the best solution the support team advises is to purchase a feature, which will compensate for their software's shortcomings.

8. Working with Uncleaned Data is Impractical

Tableau focuses mainly on data visualization and does not work with uncleaned data. In order to use Tableau efficiently, users need to do proper data cleaning in the underlying database first before performing any data visualization.

# 5.0 REFERENCE

## 5.0 Reference


*Connecting to Multiple Data Sources Without Joining or Blending | Tableau Software.* (2016, August 30). Knowledge Base. Retrieved May 23, 2022, from https://kb.tableau.com/articles/howto/connecting-multiple-data-sources-without-joining-or-blending

*Connecting to Tableau Server connections.* (2014, September 19). [Video]. Youtube. https://www.youtube.com/watch?v=pABgrUSRuBo

*Data Mining Cluster Analysis. (n.d.).* Javatpoint. Retrieved May 10, 2022, from https://www.javatpoint.com/data-mining-cluster-analysis

*Forecasting in Tableau*. (2016, October 23). [Video]. YouTube. https://www.youtube.com/watch?v=_sRjbbWhcCg

GeeksforGeeks. (2021, October 22). *Data Mining - Cluster Analysis.* Retrieved May 11, 2022, from https://www.geeksforgeeks.org/data-mining-cluster-analysis/

*How to in Tableau in 5 mins: Find Outliers in Time Series Data*. (2021, September 8). [Video]. YouTube. https://www.youtube.com/watch?v=9hhesez762g

*Install Tableau Desktop - Tableau.* (n.d.). Tableau Help. Retrieved May 23, 2022, from https://help.tableau.com/current/desktopdeploy/en-us/desktop_deploy_download_and_install.htm

*Linear Regression Model in Tableau | Data Science in Tableau*. (2019, December 25). [Video]. YouTube. https://www.youtube.com/watch?v=S7jaK34GanE

*Manage Data Sources - Tableau*. (n.d.). Tableau Help. Retrieved May 23, 2022, from https://help.tableau.com/current/pro/desktop/en-us/supported_datasources.htm

*Market Basket Analysis in Tableau (Alternate Approach)*. (2018, November 24). [Video]. YouTube. https://www.youtube.com/watch?v=rzLBsZcPWmc

*Market Basket Analysis in Tableau.* (2018, November 18). [Video]. YouTube. https://www.youtube.com/watch?v=IXZyPbahDWs

*Publish a Data Source - Tableau.* (n.d.). Tableau Help. Retrieved May 23, 2022, from https://help.tableau.com/current/pro/desktop/en-us/publish_datasources.htm

Rodriguez, R. (2021, July 9). *How to Conduct a Market Basket Analysis in Tableau.* Tech Business Guide. Retrieved May 18, 2022, from https://techbusinessguide.com/how-to-create-a-market-basket-analysis-in-tableau/

Shaptunova, Y. (2021, September 16). *Pros and Cons of Tableau Software for Data Visualization* [Review]. SaM Solutions. Retrieved May 21, 2022, from https://www.sam-solutions.com/blog/tableau-software-review-pros-and-cons-of-a-bi-solution-for-data-visualization/

*Start Page.* (n.d.). Tableau. Retrieved May 23, 2022, from https://help.tableau.com/current/pro/desktop/en-us/environment_startpage.htm

*Tableau Charts: How & when to use different tableau charts. Edureka. (2021, December 19). Retrieved May 23, 2022, from* https://www.edureka.co/blog/tableau-charts/#:~:text=Tableau%20Charts%3A%20Bar%20Charts&text=You%20can%20quickly%20highlight%20differences,Go%20to%20a%20new%20worksheet

*Tableau - Connecting With Different Database.* (2018, January 22) [Video]. Youtube. https://www.youtube.com/watch?v=vRu_2nZXBPI

*Tableau: How to acquire the free one year student version of Tableau.* (2019, August 21). [Video]. Youtube. https://www.youtube.com/watch?v=giY8J1wfnH0

To, K. (2022, March 29). *Tableau Limitations and Top 4 Alternatives.* The Holistics Blog. Retrieved May 21, 2022, from https://www.holistics.io/blog/tableau-limitations-and-top-alternatives/

*What are the 8 Amazing Benefits of Using Tableau For Your Projects?* (n.d.). KnowledgeHut. Retrieved May 21, 2022, from https://www.knowledgehut.com/blog/business-intelligence-and-visualization/tableau-advantages-disadvantages

*What Are the Pros & Cons of Tableau?* (2020, December 14). Thinklytics. Retrieved May 21, 2022, from https://thinklytics.com/what-are-the-pros-cons-of-tableau/