

INTRODUCTION

What is Data?

Data refers to raw facts, statistics, or information collected or stored in a structured or unstructured form. Data can take various forms, such as text, numbers, images, videos, and more. It is the foundation of all information and knowledge and is used in various fields for analysis, decision-making, and understanding trends and patterns.

Data can be categorized into two main types:

- **Structured Data:** This type of data is organized into a specific format, such as tables or databases, and is easily searchable and analyzable. Examples include spreadsheets, relational databases, and CSV files.
- **Unstructured Data:** Unstructured data lacks a specific format and can include text documents, social media posts, images, audio recordings, and more. Analyzing unstructured data often requires advanced techniques like natural language processing and image recognition.

Where to Find Data?

You can find data from various sources, depending on your specific needs:

- **Open Data Portals:** Many governments and organizations provide free access to a wide range of data through open data portals. Examples include Data.gov (United States) and data.gov.uk (United Kingdom).
- **Data Repositories:** Academic institutions, research organizations, and data enthusiasts often share datasets on platforms like Kaggle, GitHub, and the UCI Machine Learning Repository.
- **APIs (Application Programming Interfaces):** Some websites and services offer APIs that allow you to programmatically access and retrieve data. Examples include Twitter API, Google Maps API, and financial market APIs.
- **Web Scraping:** You can extract data from websites using web scraping tools and libraries like BeautifulSoup and Scrapy. However, be mindful of the website's terms of use and legal restrictions.

- **Surveys and Interviews:** You can conduct your own surveys or collect data through questionnaires and interviews.
- **IoT Devices:** Internet of Things (IoT) devices generate vast amounts of data that can be used for various purposes.
- **Commercial Data Providers:** Some companies specialize in selling datasets for specific industries, such as market research, finance, and healthcare.

Data Visualization

Data visualization is the graphical representation of data and information. It uses visual elements like charts, graphs, and maps to help viewers understand trends, outliers, and patterns in data. Data visualization is important because it allows people to see and understand data more easily than they would with raw numbers.

There are many types of data visualizations, including:

1. **Bar charts:** Used to compare different categories of data.
2. **Line charts:** Show trends over time or relationships between variables.
3. **Pie charts:** Display parts of a whole and are useful for showing percentages.
4. **Scatter plots:** Show the relationship between two variables.
5. **Heat maps:** Use color to represent data values in a matrix.
6. **Tree maps:** Display hierarchical data using nested rectangles.
7. **Dashboards:** Combine multiple visualizations into a single interface for comprehensive data analysis.

Effective data visualization involves choosing the right type of visualization for the data and the audience, using appropriate colors and labels, and providing context to help viewers interpret the data correctly.

Applications of Data Visualization:

Data visualization has a wide range of applications across various industries and fields. Some common uses include:

1. **Business Intelligence:** Data visualization is used to analyze business data, identify trends, and make informed decisions. It helps businesses understand their performance, customer behavior, and market trends.

2. **Financial Analysis:** In finance, data visualization is used to analyze stock prices, market trends, and financial performance. It helps investors and financial analysts make informed decisions about investments.
3. **Healthcare:** Data visualization is used in healthcare to analyze patient data, track disease outbreaks, and monitor public health trends. It helps healthcare professionals make informed decisions about patient care and public health interventions.
4. **Marketing:** Marketers use data visualization to analyze customer behavior, track marketing campaigns, and measure the effectiveness of marketing strategies. It helps them make data-driven decisions to improve marketing performance.
5. **Education:** In education, data visualization is used to track student performance, identify learning trends, and improve educational outcomes. It helps educators tailor their teaching methods to individual student needs.
6. **Research:** Researchers use data visualization to analyze research data, visualize scientific findings, and communicate research results. It helps researchers communicate complex ideas and findings to a wider audience.
7. **Operations Management:** Data visualization is used in operations management to track production processes, analyze supply chain data, and optimize operations. It helps organizations improve efficiency and reduce costs.
8. **Urban Planning:** Urban planners use data visualization to analyze population trends, traffic patterns, and environmental data. It helps them make informed decisions about urban development and infrastructure planning.

These are just a few examples of the many applications of data visualization. Its versatility and effectiveness make it a valuable tool in almost every industry for making sense of complex data and driving informed decision-making.

Foundations for Building Data Visualizations:

Creating effective data visualizations requires a strong foundation in several key areas:

- **Data Analysis:** Before creating visualizations, you should thoroughly analyze your data to understand its structure, relationships, and any patterns or trends. Exploratory data analysis (EDA) techniques can help with this.

- **Statistical Knowledge:** Understanding basic statistics is essential for making meaningful interpretations of data. Concepts like mean, median, standard deviation, and correlation are commonly used in data visualization.
- **Domain Knowledge:** Having knowledge of the specific domain or subject matter related to your data is crucial for creating contextually relevant visualizations. It helps you ask the right questions and provide valuable insights.
- **Visualization Tools:** Familiarize yourself with data visualization tools and libraries such as matplotlib, Seaborn, ggplot2, D3.js, and Tableau. Each tool has its strengths and can be used for different types of visualizations.
- **Design Principles:** Study design principles, including color theory, typography, and visual hierarchy, to create visually appealing and effective visualizations. Avoid common pitfalls like misleading visualizations.
- **Interactivity:** Learn how to add interactive elements to your visualizations to engage users and allow them to explore the data. This can be achieved using tools like JavaScript, Python libraries, or dedicated visualization software.

Creating Your First Visualization:

To create your first data visualization, follow these general steps:

- **Select Your Data:** Choose a dataset that aligns with your goals and interests. Ensure that the data is clean and well-structured.
- **Define Your Objective:** Clearly define what you want to communicate or explore with your visualization. Are you looking to show trends, comparisons, or distributions?
- **Choose the Right Visualization Type:** Select a visualization type that suits your data and objectives. Common types include bar charts, line charts, scatter plots, histograms, and pie charts.
- **Prepare and Transform Data:** Preprocess your data as needed. This may involve aggregating, filtering, or transforming the data to fit the chosen visualization.
- **Create the Visualization:** Use a suitable tool or library to create your visualization. Customize it with labels, colors, and other design elements.
- **Interactivity (Optional):** If appropriate, add interactive features to your visualization to allow users to interact with the data.
- **Test and Iterate:** Review your visualization for accuracy and clarity. Seek feedback from others and make improvements as necessary.

- **Publish or Share:** Once you are satisfied with your visualization, publish it on a platform, embed it in a report, or share it with your intended audience.
- **Document and Explain:** Provide context and explanations for your visualization. Clearly communicate what the viewer should take away from it.
- **Maintain and Update:** If the data changes or new insights emerge, update your visualization accordingly.

MODULE 1

INTRODUCTION TO TABLEAU

TABLEAU:

Tableau is a visual analytics platform that is revolutionizing the way we use data to solve problems by enabling individuals and organisations to make the most of their data.

Tableau is a great data visualization and business intelligence application that can be used to report and analyse massive amounts of data. Salesforce purchased Tableau in June 2019, an American firm founded in 2003. It enables users to build various charts, graphs, maps, dashboards, and stories for visualising and analysing data in order to aid in business choices. Tableau offers several unique and fascinating features that make it one of the most popular business intelligence (BI) applications.

Use of Tableau:

Tableau is the fastest and most powerful visualization tool. It is very easy to use. There are no complex formulas like Excel and other visualization tools. It provides the features like cleaning, organizing, and visualizing data, it easier to create interactive visual analytics in the form of dashboards. These dashboards make it easier for non-technical analysts and end-users to convert data into understandable ones.

Tableau Features

1. Tableau supports powerful data discovery and exploration that enables users to answer important questions in seconds
2. No prior programming knowledge is needed; users without relevant experience can start immediately with creating visualizations using Tableau
3. It can connect to several data sources that other BI tools do not support. Tableau enables users to create reports by joining and blending different datasets
4. Tableau Server supports a centralized location to manage all published data sources within an organization

Values in Tableau

There are two types of values in the tableau:

- **Dimensions:** Values that are discrete(which can not change with respect to time) in nature called Dimension in tableau. Example: city name, product name, country name.
- **Measures:** Values that are continuous(which can change with respect to time) in nature called Measure in tableau. Example: profit, sales, discount, population.

Advantages of Tableau

- Quick calculation- All the calculations on the tableau done by the backend, so it is relatively faster than any other tool.
- Interactive dashboards– Tableau dashboards are very interactive and easy to draw.
- No manual calculation- All the calculations are done by the tableau only. There is no manual calculation, but in some specific cases, we used calculated fields for calculation.
- A large amount of data- Tableau can handle a large amount of data. Different types of visualization can be created with a large amount of data without impacting the performance of the dashboards.

Disadvantages of Tableau

- High Cost- tableau is a paid tool for visualization, and it is a reason why people are not using tableau so much.
- Static and single value parameters- Tableau's parameters are static and always single value can be selected using a parameter. Whenever the data gets changed, these parameters need to be updated manually every time.
- Limited Data Preprocessing- Tableau is strictly a visualization tool. Tableau Desktop allows you to do very basic preprocessing.

DATA VISUALIZATION BEST PRACTICES:

Data visualization best practices can help you create more effective and engaging visualizations. Here are some key principles:

1. **Simplicity:** Keep your visualizations simple and focused. Avoid clutter and unnecessary elements.
2. **Clarity:** Ensure that your visualizations are easy to understand. Use clear labels, titles, and legends.
3. **Consistency:** Use consistent colors, fonts, and styles across your visualizations for a cohesive look.
4. **Accuracy:** Ensure that your data is accurate and correctly represented in your visualizations. Avoid distorting or misrepresenting data.
5. **Relevance:** Choose the most appropriate type of visualization for your data and the message you want to convey.
6. **Interactivity:** Use interactive elements like tooltips and filters to allow users to explore the data in more detail.
7. **Accessibility:** Ensure that your visualizations are accessible to all users, including those with disabilities. Use appropriate color schemes and provide alternative text for images.
8. **Storytelling:** Use your visualizations to tell a story and guide the viewer through the data, highlighting key insights and trends.
9. **Feedback:** Seek feedback from others to improve your visualizations. Pay attention to how users interact with your visualizations and make adjustments accordingly.
10. **Iteration:** Don't be afraid to iterate on your visualizations. Experiment with different approaches to find the most effective way to present your data. By following these best practices, you can create visualizations that are not only visually appealing but also informative and impactful.

GETTING STARTED WITH TABLEAU:

Getting started with Tableau software is a great way to create data visualizations quickly and efficiently. Here are the steps to get started, including connecting your data to Tableau, creating basic charts..

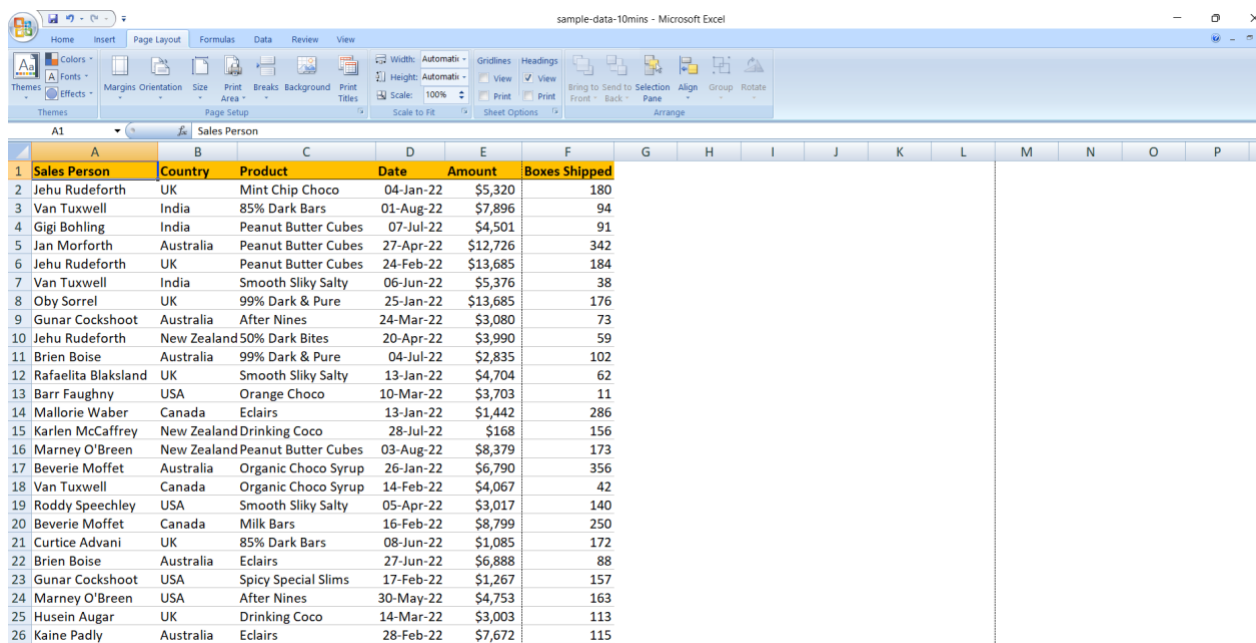
Download and Install Tableau:

First, you'll need to download and install **Tableau Desktop** or **Tableau Public** (a free version). Follow the installation instructions provided on the Tableau website for your specific operating system.

Link: <https://www.tableau.com/products/public/download>

Prepare Your Data:

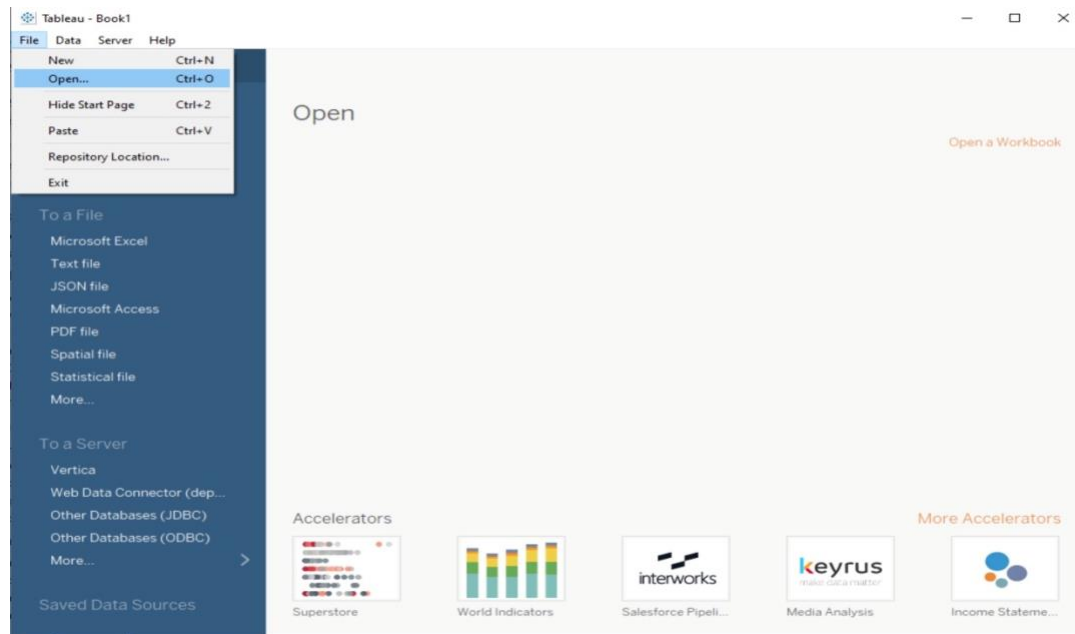
Before connecting your data to Tableau, ensure that your data is in a suitable format. Common data file formats that Tableau supports include **Excel (.xlsx)**, **CSV (.csv)**, and **text files (.txt)**. Make sure your data is organized with headers for each column.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Sales Person	Country	Product	Date	Amount	Boxes Shipped										
2	Jehu Rudeforth	UK	Mint Chip Choco	04-Jan-22	\$5,320	180										
3	Van Tuxwell	India	85% Dark Bars	01-Aug-22	\$7,896	94										
4	Gigi Bohling	India	Peanut Butter Cubes	07-Jul-22	\$4,501	91										
5	Jan Morforth	Australia	Peanut Butter Cubes	27-Apr-22	\$12,726	342										
6	Jehu Rudeforth	UK	Peanut Butter Cubes	24-Feb-22	\$13,685	184										
7	Van Tuxwell	India	Smooth Silky Salty	06-Jun-22	\$5,376	38										
8	Oby Sorrel	UK	99% Dark & Pure	25-Jan-22	\$13,685	176										
9	Gunar Cockshoot	Australia	After Nines	24-Mar-22	\$3,080	73										
10	Jehu Rudeforth	New Zealand	50% Dark Bites	20-Apr-22	\$3,990	59										
11	Brien Boise	Australia	99% Dark & Pure	04-Jul-22	\$2,835	102										
12	Rafaelita Blaksland	UK	Smooth Silky Salty	13-Jan-22	\$4,704	62										
13	Barr Faughny	USA	Orange Choco	10-Mar-22	\$3,703	11										
14	Mallorie Waber	Canada	Eclairs	13-Jan-22	\$1,442	286										
15	Karlen McCaffrey	New Zealand	Drinking Coco	28-Jul-22	\$168	156										
16	Marney O'Brien	New Zealand	Peanut Butter Cubes	03-Aug-22	\$8,379	173										
17	Beverie Moffet	Australia	Organic Choco Syrup	26-Jan-22	\$6,790	356										
18	Van Tuxwell	Canada	Organic Choco Syrup	14-Feb-22	\$4,067	42										
19	Roddy Speechley	USA	Smooth Silky Salty	05-Apr-22	\$3,017	140										
20	Beverie Moffet	Canada	Milk Bars	16-Feb-22	\$8,799	250										
21	Curtice Advani	UK	85% Dark Bars	08-Jun-22	\$1,085	172										
22	Brien Boise	Australia	Eclairs	27-Jun-22	\$6,888	88										
23	Gunar Cockshoot	USA	Spicy Special Slims	17-Feb-22	\$1,267	157										
24	Marney O'Brien	USA	After Nines	30-May-22	\$4,753	163										
25	Husein Augar	UK	Drinking Coco	14-Mar-22	\$3,003	113										
26	Kaine Padly	Australia	Eclairs	28-Feb-22	\$7,672	115										

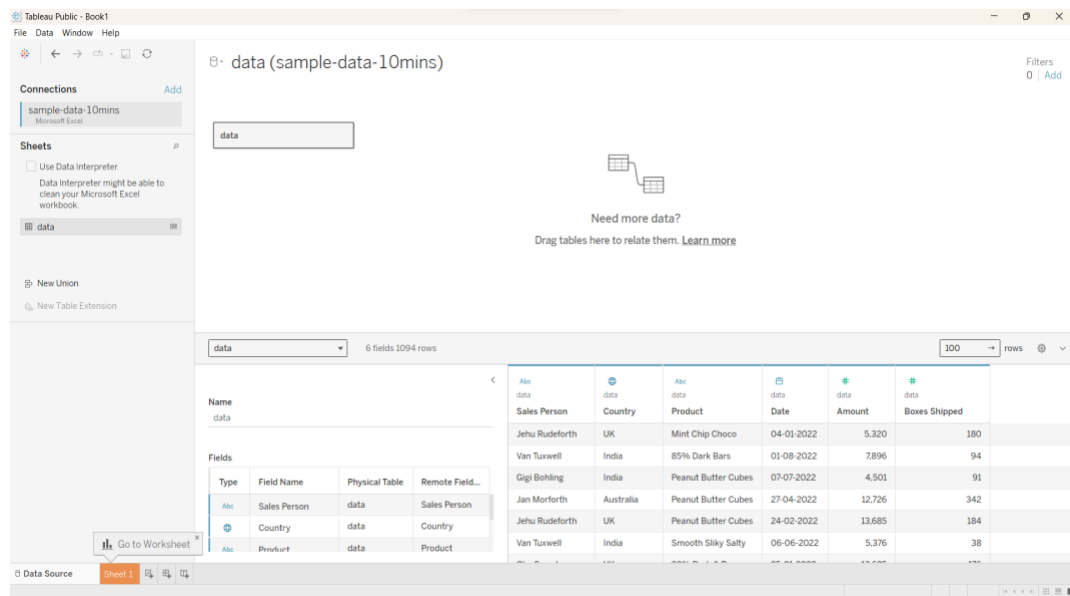
Connect Your Data to Tableau or connecting to the tutorial dataset:

1. Launch **Tableau Desktop**.
2. Go to "**File**" Menu and then click on "**Open**".

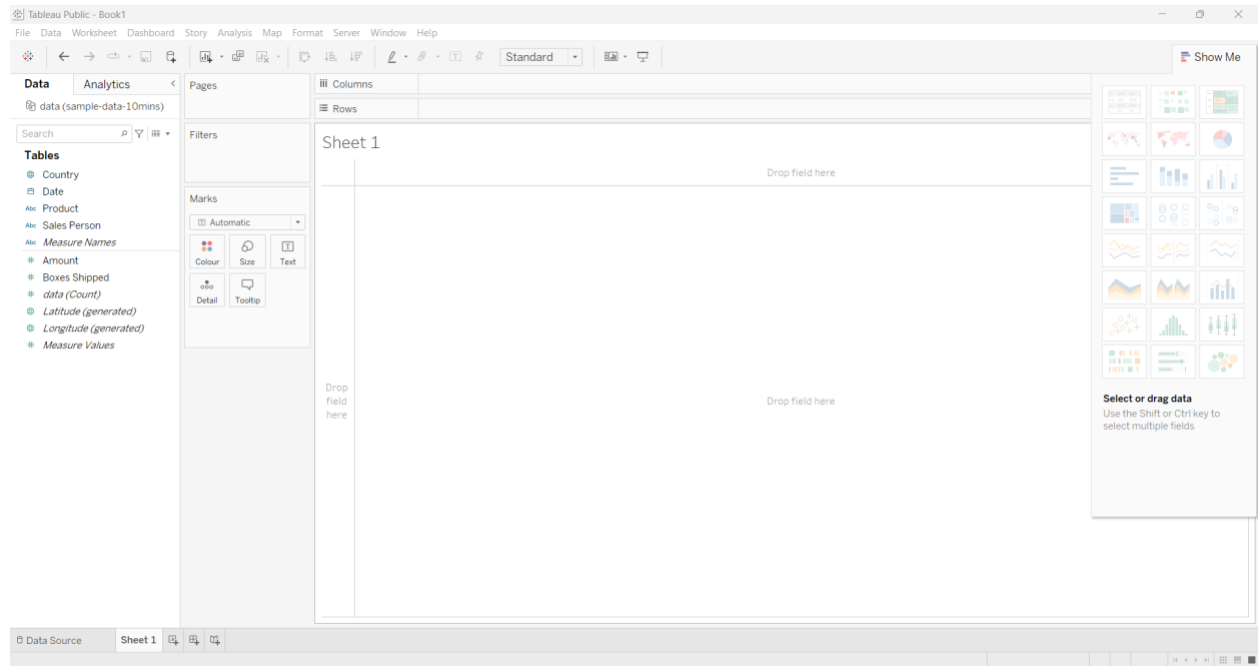


3. Choose the data source type (e.g., Excel, CSV, text file) and Select the data file and click "Open".

4. Drag any table into working area.



5. Click on Worksheet(Sheet1).

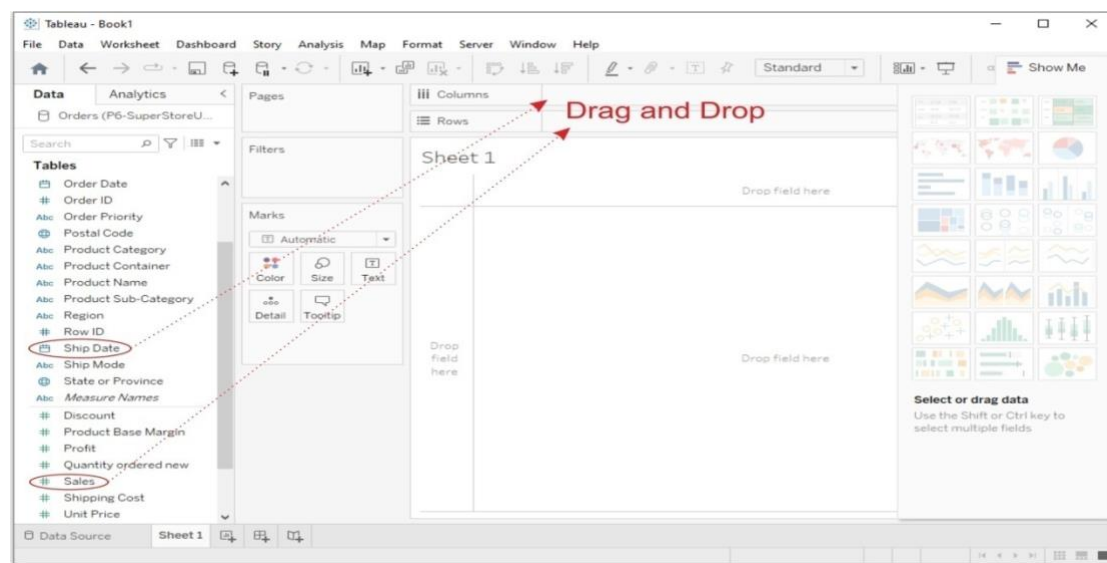


CREATING BASIC CHARTS:

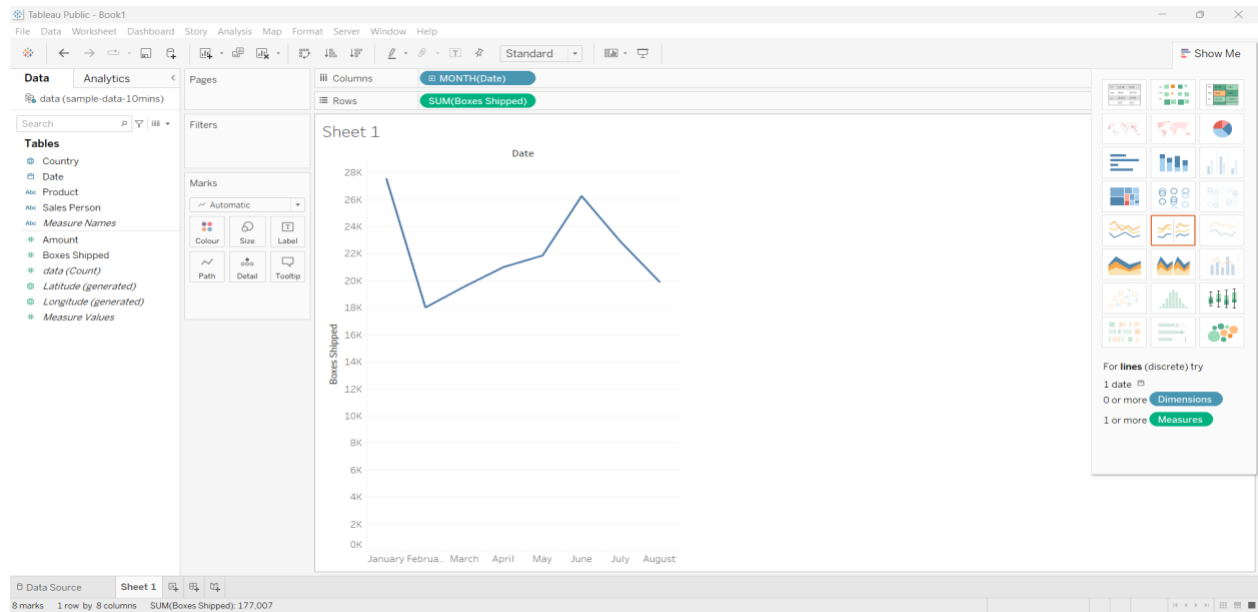
Now, let's create some basic charts using Tableau:

a. Line Chart:

1. From the "**Data Source pane**", drag and drop the date field to the **Columns shelf** and a numeric field (e.g., date, boxesshipped) to the **Rows shelf**.

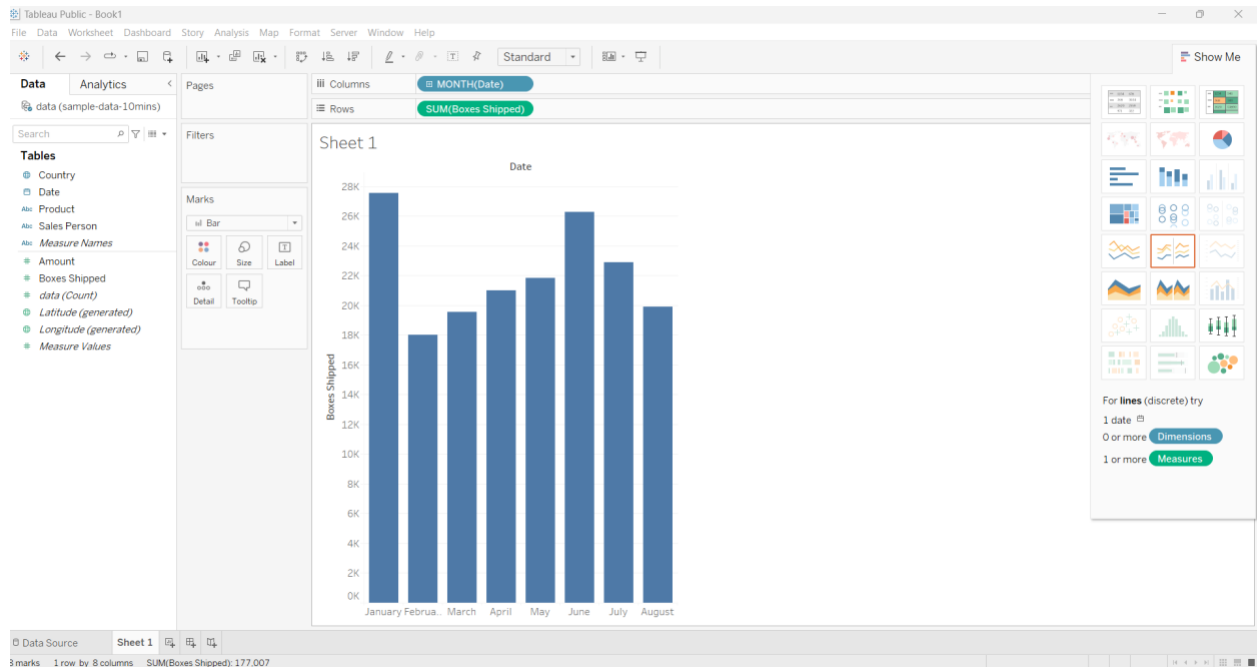


2. Then **Tableau** will automatically create a line chart. You can customize it by adding labels, titles, and formatting.



b. Bar Chart:

1. Drag and drop a **categorical field** (e.g., date, boxes shipped) to the Columns shelf and a **numeric field** to the Rows shelf.



FILTERING AND SORTING DATA:

Filtering:

Data filtering is the process of selecting and displaying a subset of data based on specified criteria. It allows you to focus on specific information within a dataset while temporarily hiding the rest. Filtering can be done based on various conditions, such as values, dates, or categories, to extract relevant insights and simplify data analysis. In Tableau, for example, you can apply filters to your visualizations to show only the data that meets certain criteria, helping you to explore and understand your data better.

Sorting:

Sorting is the process of arranging data in a specific order, typically based on the values of one or more columns. It allows you to organize data in a meaningful way, making it easier to analyze and interpret. In sorting, you can arrange data in ascending or descending order, depending on your requirements. For example, you can sort a list of names alphabetically or sort sales data from highest to lowest. Sorting helps in identifying patterns, trends, and outliers in the data, making it an important aspect of data analysis and visualization.

In Tableau, filtering and sorting data are essential techniques for exploring and analyzing your data. Here's how you can do it:

Filtering Data:

1. Basic Filtering:

- Drag a field from the data pane to the Filters shelf.
- Choose the values you want to include or exclude.

2. Quick Filters:

- Right-click on a field in the view and select "Show Filter."
- A filter control will be added to the view, allowing users to interactively filter the data.

3. Filtering with Conditions:

- Use the "Filter" option in the context menu or drag a field to the Filters shelf.

- Choose "Custom" to define specific conditions for filtering.

4. Top N Filters:

- Use this to filter the view to show the top or bottom N items based on a measure.
- Right-click on a measure in the view, go to "Quick Table Calculation," and select "Top N."

Sorting Data:

1. Sorting within a Field:

- Right-click on a field in the view and select "Sort."
- Choose either ascending or descending order.

2. Sorting by a Different Field:

- Drag a field from the data pane and drop it onto the field you want to sort by.
- Choose the "Sort" option and select the desired sorting order.

3. Manual Sorting:

- Drag a field to the Rows or Columns shelf.
- Click on the field's drop-down menu and select "Sort."
- Choose "Manual" and drag items to rearrange them.

4. Sorting by a Calculation:

- Create a calculated field that defines the sorting logic.
- Use this calculated field to sort the data in your view.

By using these filtering and sorting techniques, you can easily navigate and analyze your data in Tableau to gain valuable insights.

