

INTRODUCTION TO TABLEAU: DATA VISUALIZATION WORKSHOP

WORKSHOP DETAILS

Instructor: Tanya Khanna

Email: tk759@scarletmail.rutgers.edu

Workshop Materials:

Github Link: <https://github.com/Tanya-Khanna/Data-Science-Workshop---Spring-2025---NBL->

WORKSHOP SCHEDULE

Introduction to Python Programming	February 3, 2025; 2 – 3:30 PM
Mastering Data Analysis: Pandas and Numpy	February 10, 2025; 2 – 3:30 PM
Introduction to Tableau: Visualizing Data Made Easy	February 17, 2025; 2 – 3:30 PM
Introduction to Machine Learning: Supervised Learning	February 24, 2025; 2 – 3:30 PM
Introduction to Machine Learning: Unsupervised Learning	March 3, 2025; 2 – 3:30 PM
Data-Driven Decision Making: A/B Testing and Statistical Hypothesis Testing	March 10, 2025; 2 – 3:30 PM
Demystifying Generative AI	March 24, 2025; 2 – 3:30 PM
Large Language Models: From Theory to Implementation	March 31, 2025; 2 – 3:30 PM
Generative AI Applications with AI Agents	April 7, 2025; 2 – 3:30 PM
Building Intelligent Recommendation Systems	April 14, 2025; 2 – 3:30 PM

SPRING 2025 WORKSHOP CALENDAR

TABLE OF CONTENTS

- WHAT IS DATA VISUALIZATION?
 - WHY IT MATTERS
 - KEY PRINCIPLES
- TABLEAU PUBLIC
- TABLEAU PUBLIC VS. TABLEAU DESKTOP
- DATA CONNECTION & EXPLORATION
- BASIC BUSINESS QUESTIONS
- INTERMEDIATE ANALYSIS
- ADVANCED ANALYTICS
- CROSS-DIMENSIONAL ANALYSIS AND INTERACTIVE DASHBOARD

WHAT IS DATA VISUALIZATION?

THE GRAPHICAL REPRESENTATION OF DATA AND INFORMATION
USES CHARTS, GRAPHS, MAPS, AND INTERACTIVE VISUALS
HELPS IN IDENTIFYING TRENDS, PATTERNS, AND OUTLIERS



Why Data Visualization Matters

1 Transform Raw Numbers

Raw numbers transform into instant insights through visual patterns, like turning a confusing spreadsheet into a clear story that your brain can grasp in seconds.

2 Faster Decision-Making

Visual patterns accelerate decision-making by letting you spot trends instantly, similar to how you can immediately tell if a line graph is trending up or down.

3 Better Communication

Complex data transforms into clear insights through visuals, like seeing a month's performance in a single glance rather than scanning hundreds of numbers.

4 Identify Patterns

Hidden patterns emerge visually as your eyes naturally detect relationships, outliers, and trends that would be invisible in rows of numbers.

Key Principles of Effective Data Visualization



Clarity

Present data without distortion or confusion



Purpose

Every visualization should answer a specific question



Simplicity

Remove unnecessary elements



Context

Provide relevant background information

INTRODUCTION TO TABLEAU PUBLIC

- Tableau Public is a free data visualization tool that allows users to create interactive visualizations and dashboards.
- Designed for public sharing—workbooks are saved to the Tableau Public cloud.
- Ideal for students, researchers, and professionals looking to share insights publicly.

Why Use Tableau Public?

- Free to use
- User-friendly drag-and-drop interface
- Wide range of visualizations (bar charts, maps, dashboards)
- Supports real-time interaction
- No coding required

DOWNLOAD & INSTALL TABLEAU PUBLIC

Step 1: Download

1. Go to <https://public.tableau.com/en-us/s/>
2. Click Download Tableau Public (available for Windows & Mac).
3. Enter your email address (required for account creation).
4. Click Download Now.

Step 2: Install

1. Open the downloaded file and follow installation instructions.
2. Accept the terms & conditions and complete the installation.

Step 3: Create an Account

- Why? Tableau Public requires an account to save and publish visualizations.
 - Visit <https://public.tableau.com> → Click Sign Up.
 - Enter email, name, and password → Click Create My Profile.
-  Tip: Without an account, you won't be able to save your work.

TABLEAU PUBLIC VS TABLEAU DESKTOP

Feature	Tableau Public (Free)	Tableau Desktop (Paid)
Cost	Free	Paid (Subscription-based)
Data Privacy	Workbooks are public	Workbooks are private
Data Sources	Limited (Excel, CSV, Google Sheets)	Wide range (SQL, Snowflake, AWS, APIs)
Save Options	Online (Tableau Public Cloud)	Local save or Tableau Server
Usage	Best for public sharing	Best for corporate & private analysis
Advanced Features	Limited (no Tableau Prep, no live connections)	Full advanced analytics & integration
Security	No password protection	Full data security & control

Which One Should You Use?

- Use Tableau Public if you are a student, researcher, or blogger sharing open data.
- Use Tableau Desktop if you need secure, enterprise-level reporting.

DATA CONNECTION & EXPLORATION

Types of Data You Can Connect To

-  Excel Files (.xlsx, .csv)
-  Google Sheets
-  Text Files (.txt, .tsv)

 Note: Tableau Public does NOT support live connections to databases (SQL, Snowflake, AWS, etc.), unlike Tableau Desktop.

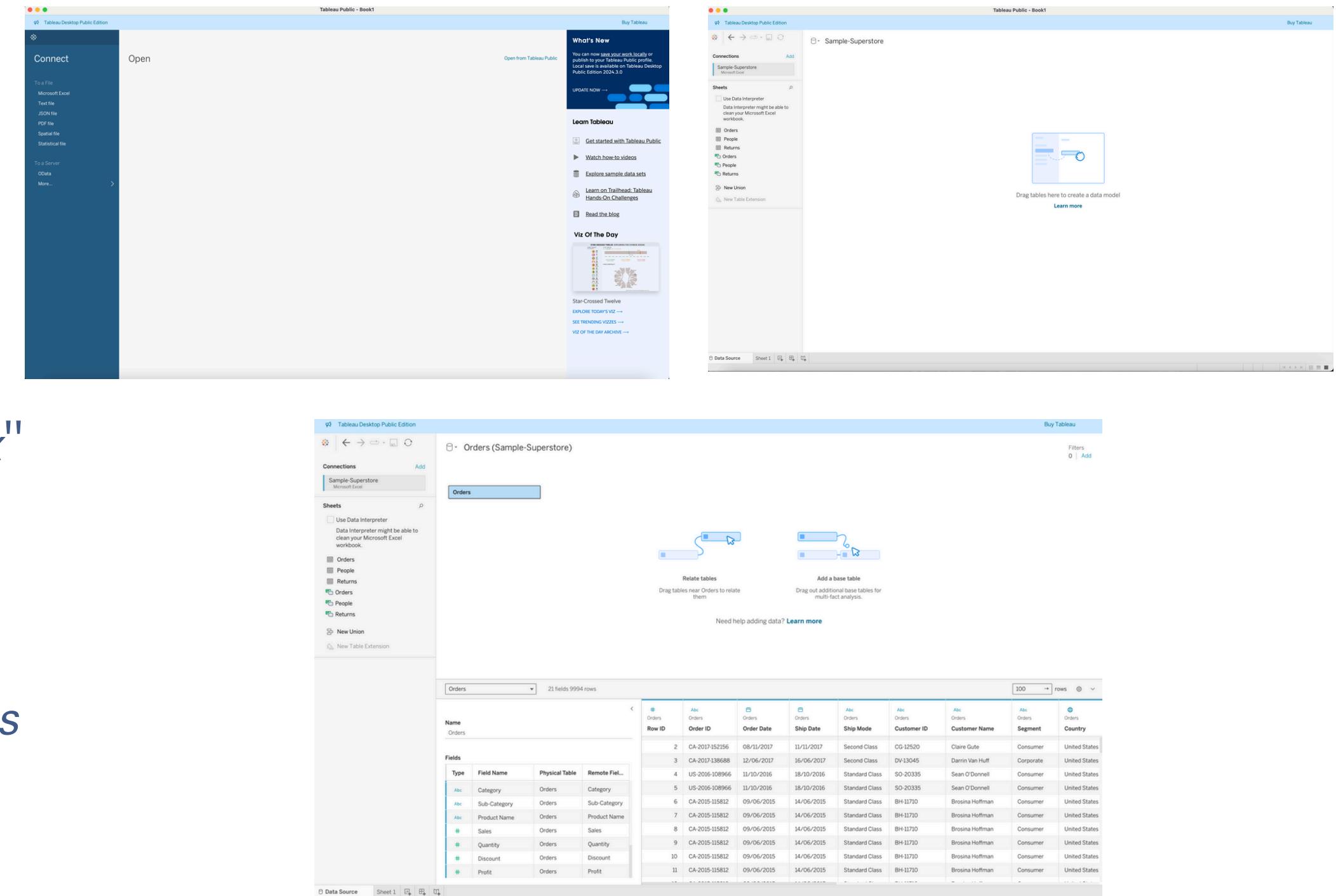
 Download the Sample - Superstore.xlsx dataset:

 [Get the dataset from here](#)

DATA CONNECTION & EXPLORATION

Load the Data into Tableau Public

- 1 Open the Tableau Public application in your computer.
 - 2 In "Connect", select "Microsoft Excel"
 - 3 Choose "Sample - Superstore.xlsx" and from your downloads.
 - 4 Drag the Orders sheet into the Data Source Canvas
- Tip: Tableau automatically detects relationships between tables.

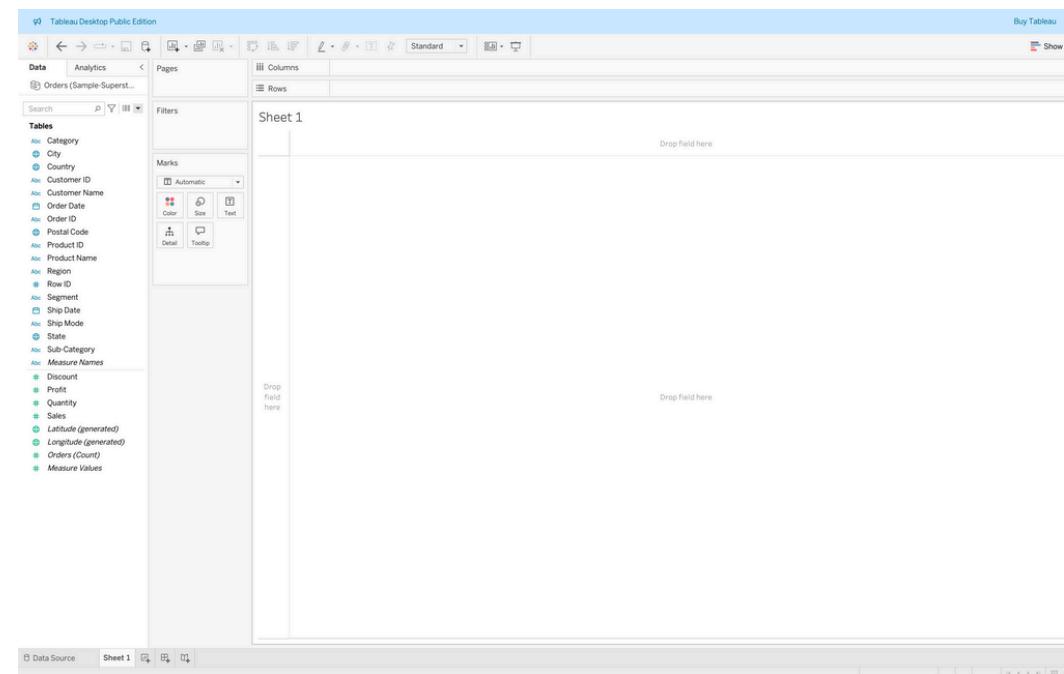


EXPLORING THE "SUPERSTORE" DATASET

The dataset contains sales data for an office supplies store across different locations.

Key Tables in the Dataset

- 📌 Orders → Main dataset containing sales transactions
- 📌 People → Sales team member information
- 📌 Returns → List of returned orders



Measures represent quantitative data for calculations. Key fields include Sales (total revenue), Profit (net earnings), Quantity (units sold), and Discount (percentage applied). Tableau auto-generates Latitude and Longitude for geographic mapping, Orders (Count) for order tracking, and Measure Names & Values for calculations.

Field Name	Data Type	Description
Category	Text	Main product category
City	Text	Customer's city
Country	Text	Customer's country
Customer ID	Text	Unique customer identifier
Customer Name	Text	Customer's full name
Order Date	Date	Date when order was placed
Order ID	Text	Unique order identifier
Postal Code	Number	Customer's postal code
Product ID	Text	Unique product identifier
Product Name	Text	Name of the purchased product
Region	Text	Region where the order was placed
Row ID	Number	Unique row identifier
Segment	Text	Customer segment (Consumer, Corporate, Home Office)
Ship Date	Date	Date when order was shipped
Ship Mode	Text	Type of shipping (Standard, First Class, Same Day)
State	Text	Customer's state
Sub-Category	Text	More specific product category

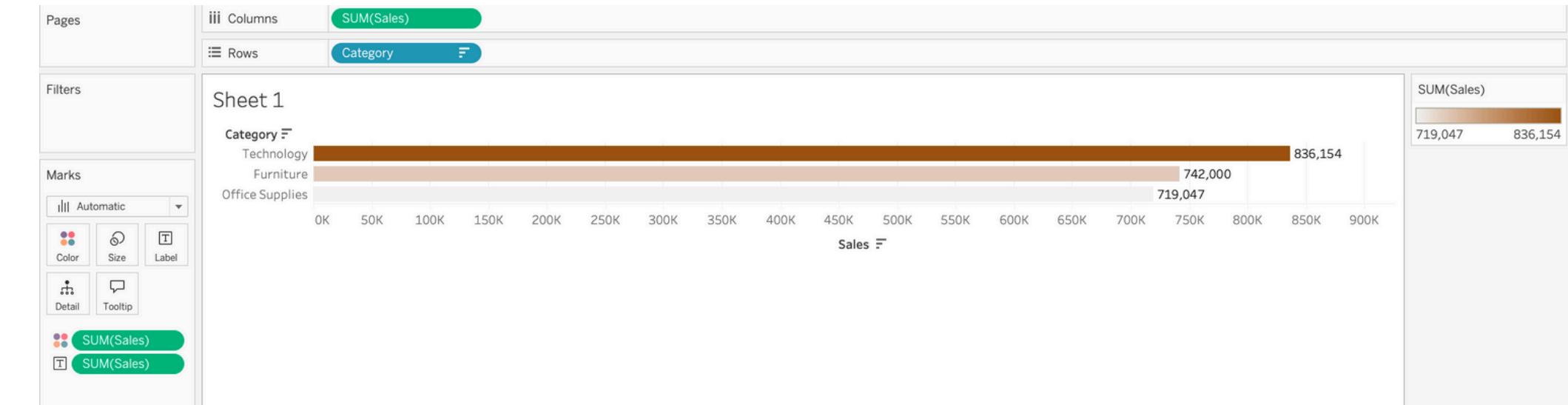
CHARTS, MAPS, AND INSIGHTS
AWAIT! LET'S UNLEASH THE
POWER OF TABLEAU!

Let's begin with the analysis and
visualizations!

BASIC BUSINESS QUESTIONS

Which product category generates the most revenue?

- Create a Bar Chart to Compare Revenue by Category:
Navigate to Sheet 1 and drag Category to the Rows shelf and drag Category to the Rows shelf.
- Click on the Sort Descending icon to arrange categories by revenue (highest to lowest).
- Drag Sales to the Color mark under the Marks pane.
- Change the color gradient to make higher sales more prominent.
- Drag Sales to the Label mark to display revenue values on the bars.
- The category with the longest bar (highest sales) is the top revenue-generating category.
- Compare the differences in sales between categories.
- Hover over bars for exact values.



BASIC BUSINESS QUESTIONS

What are our best and worst-performing regions?

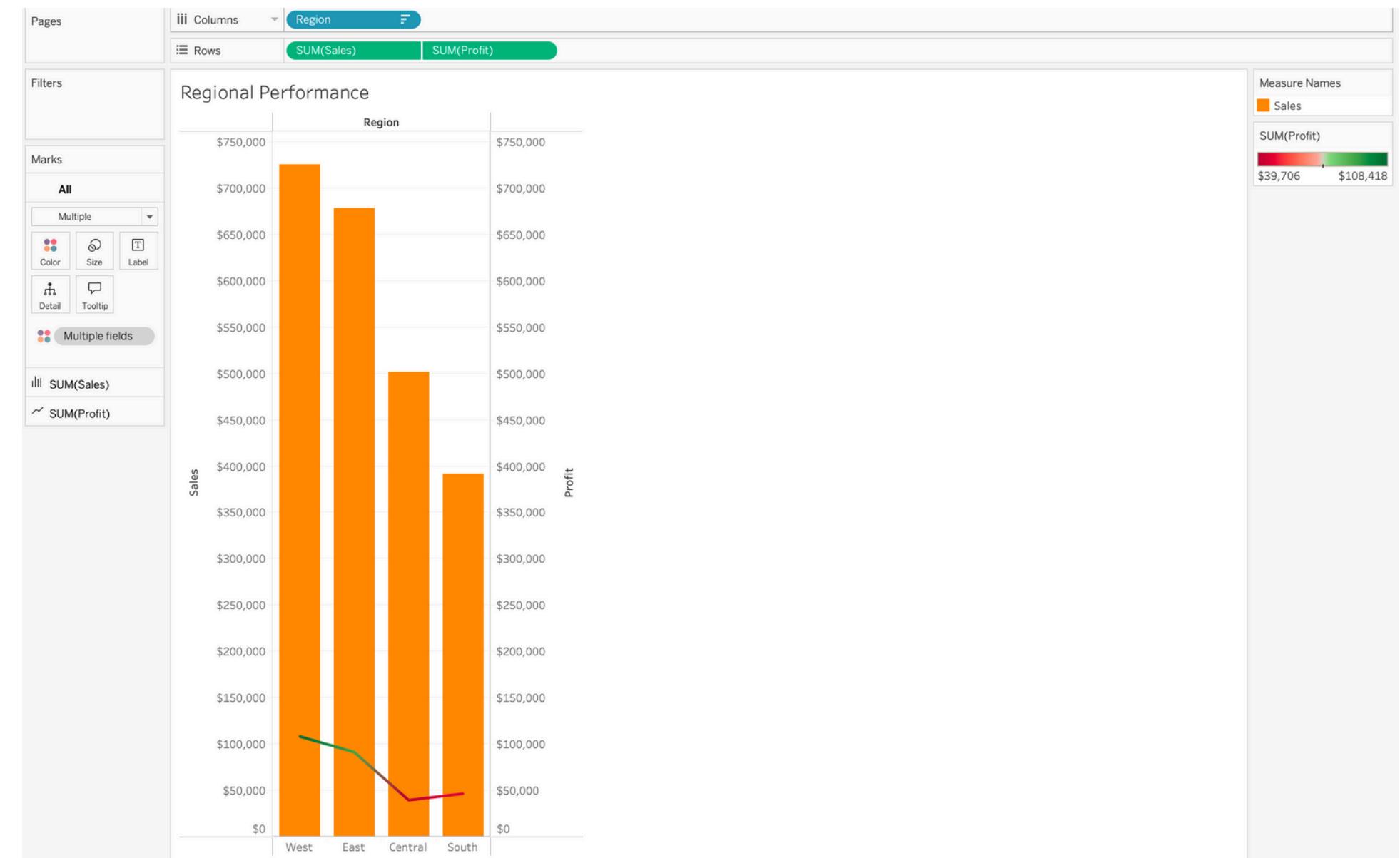
Drag Region to the Columns shelf.

Drag Sales to the Rows shelf.

Drag Profit to the Rows shelf (this will create two separate charts).

- 1 Click the drop-down menu on the Profit pill (in the Rows shelf).
- 2 Select Dual-Axis (this will merge the two charts onto one axis).
- 3 Right-click on the Profit Axis (right side of the chart) → Synchronize Axis.
- 4 Click on Sales in the Marks pane → Change to Bar Chart and click on Profit in the Marks pane → Change to Line Chart.

Drag Profit to the Color mark and select a Diverging Color Palette (Green for positive, Red for negative).



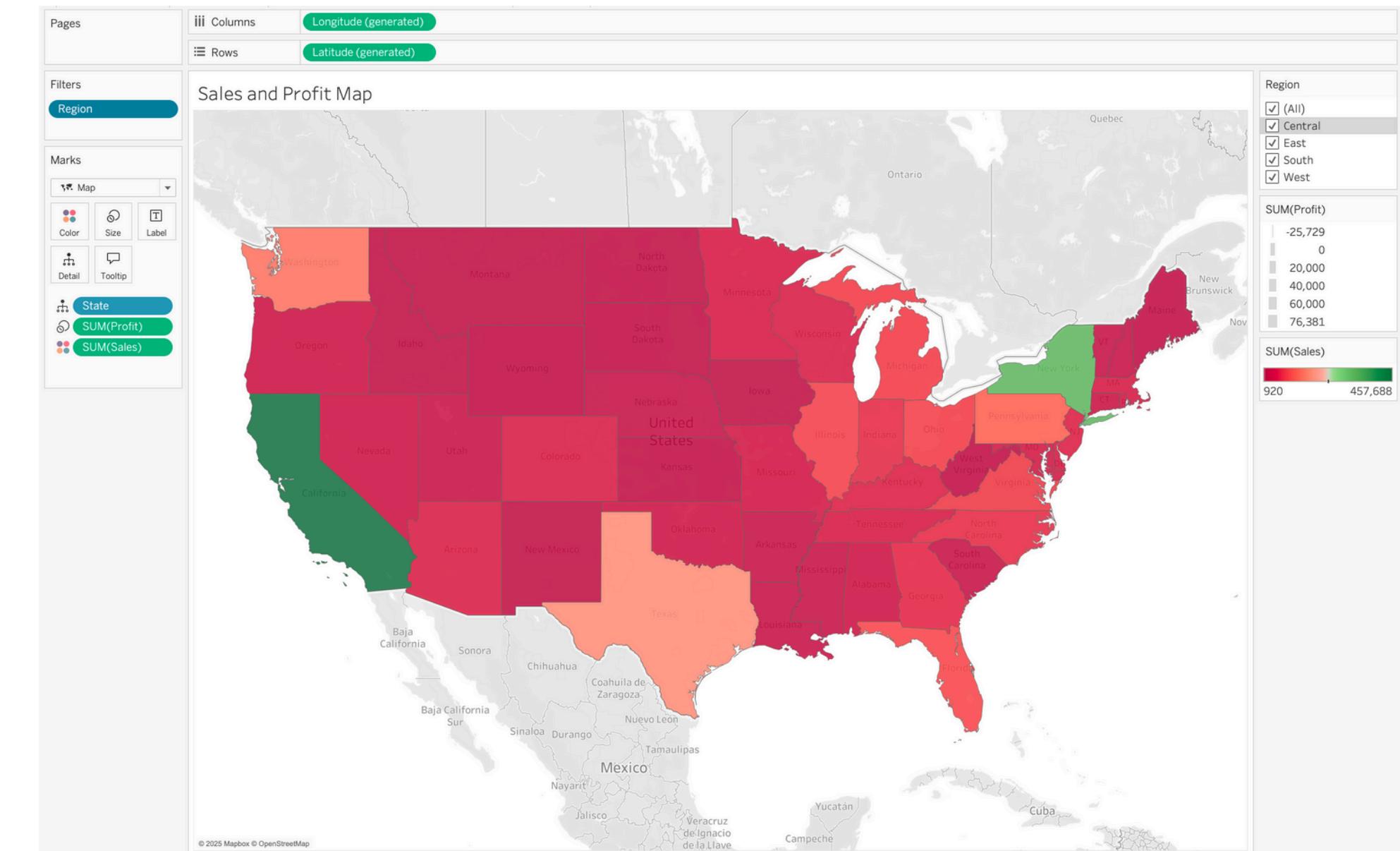
BASIC BUSINESS QUESTIONS

What are our best and worst-performing regions?

- 1 Drag State to the Columns shelf.
- 2 Drag Sales to the Size mark in the Marks pane.
- 3 Drag Profit to the Color mark in the Marks pane.
- 4 Click on the Marks drop-down menu → Select Map.
- 5 Click on Color → Choose a Diverging Color Scale:
 - Green for high profit
 - Red for low or negative profit

Add Region as a Filter

- 6 Drag Region to the Filters pane.
- 7 Click on the filter drop-down → Select Show Filter.

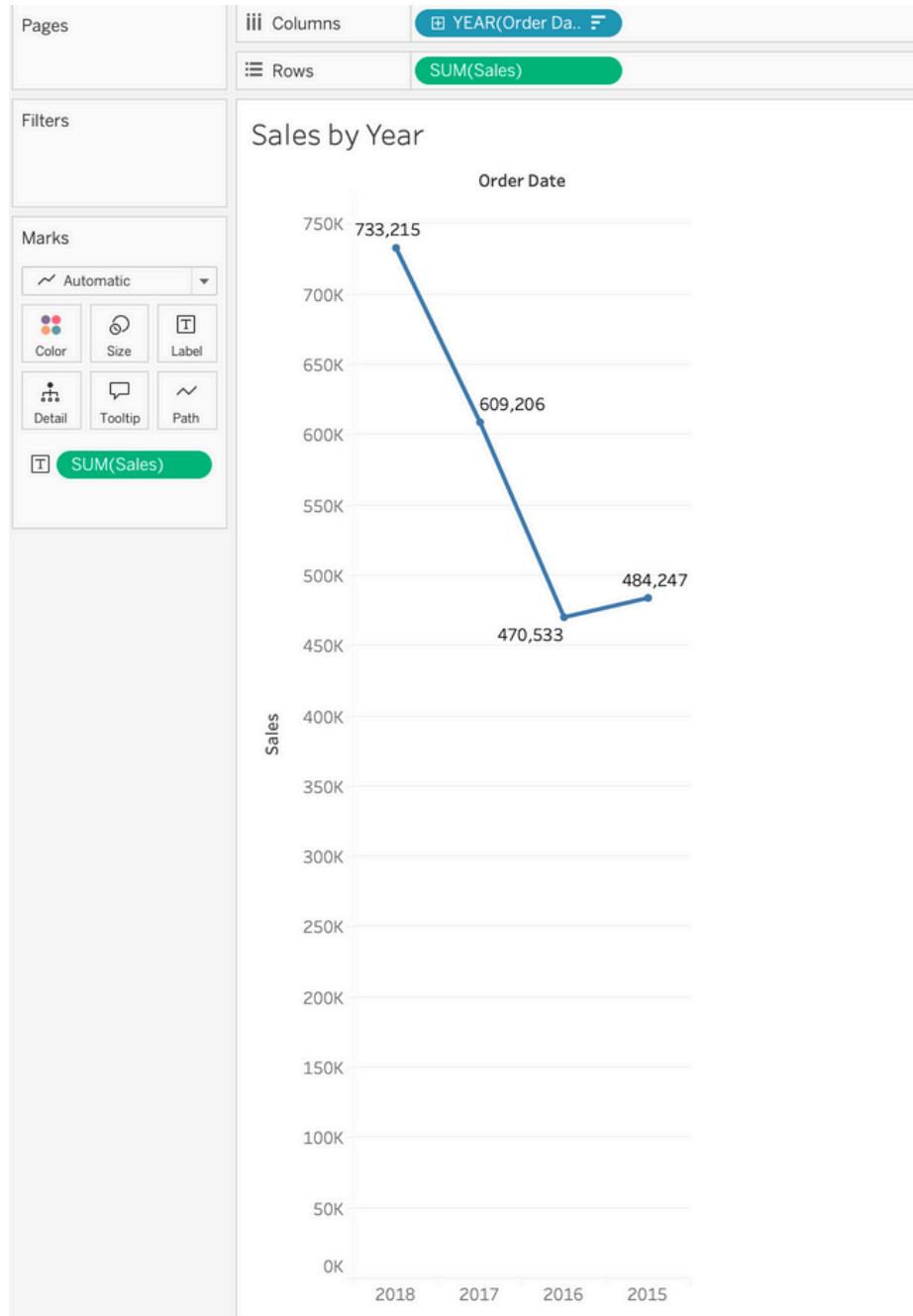


BASIC BUSINESS QUESTIONS

How do sales vary across different years/months?

Step 1: Analyze Sales Across Different Years

- 1 Click on New Worksheet (+) and rename it "Sales by Year".
- 2 Drag Order Date to the Columns shelf.
- 3 Drag Sales to the Rows shelf.
- 4 Right-click on Order Date in the Columns shelf → Select Convert to Discrete.
- 5 Right-click on Order Date again → Select Year.
- 6 Click on Sort Descending to display recent years first.
- 7 Drag Sales to the Label mark to display sales values.



Step 2: Drill Down to Monthly Sales for a Specific Year

- 1 Click on New Worksheet (+) and rename it "Sales by Month".
- 2 Drag Order Date to the Columns shelf.
- 3 Drag Sales to the Rows shelf.
- 4 Right-click on Order Date in the Columns shelf → Select Convert to Discrete.
- 5 Right-click on Order Date again → Select Month.
- 6 Click on the Filter pane → Drag Order Date to Filters.
- 7 In the Filter Order Date window → Select Years and click Next.
- 8 Check the box for the specific year you want to analyze → Click OK.
- 9 Click on the drop-down arrow in the Filters pane → Select Show Filter to allow users to change the year dynamically.
- 10 Drag Sales to the Label mark to display sales values.

INTERMEDIATE ANALYSIS

Which products have the highest profit margin?

Profit Margin is calculated as:

$$\text{Profit Margin} = (\text{Profit}/\text{Sales}) \times 100$$

We will create a calculated field to compute Profit Margin and visualize the top-performing products.

Step 1: Create a Calculated Field for Profit Margin

1 Click on Data → Create Calculated Field.

2 Name it "Profit Margin (%)".

3 Enter the formula:

$$([\text{Profit}] / [\text{Sales}]) * 100$$

Step 2: Create a Bar Chart for Product-Wise Profit Margin

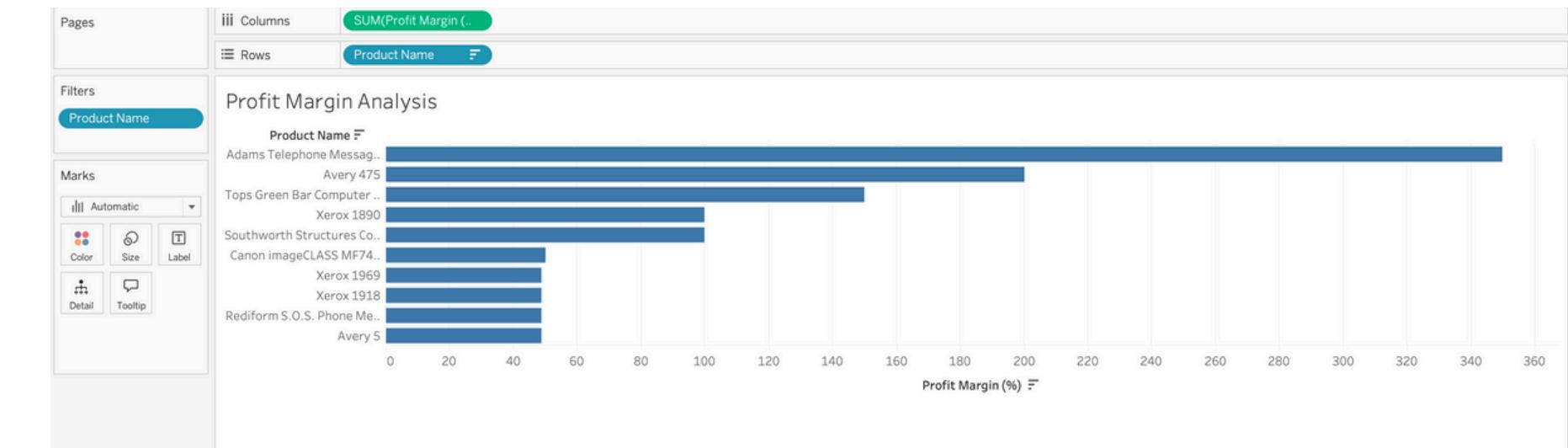
1 Drag Product Name to the Rows shelf.

2 Drag Profit Margin (%) to the Columns shelf.

4 Click Sort Descending to show highest-margin products first.

◆ Filter to Show Only Top Products:

- Drag Profit Margin (%) to the Filters pane.
- Select Top 10 by Profit Margin (%).



ADVANCED ANALYTICS

What's the customer lifetime value across different segments?

Customer Lifetime Value (CLV) measures the total revenue a business can expect from a single customer over their lifetime. It is calculated using:

$$CLV = \sum \text{Sales per Customer}$$

Step 1: Create a New Worksheet

- 1 Click on New Worksheet (+) and rename it "Customer Lifetime Value by Segment".

Step 2: Compute CLV for Each Customer

- 1 Drag Customer Name to the Rows shelf.
- 2 Drag Sales to the Columns shelf.
- 3 Click the Drop-down on SUM(Sales) → Select Quick Table Calculation → Running Total.
- 4 Drag Segment to the Color mark to differentiate customer segments.

Step 3: Aggregate CLV by Segment

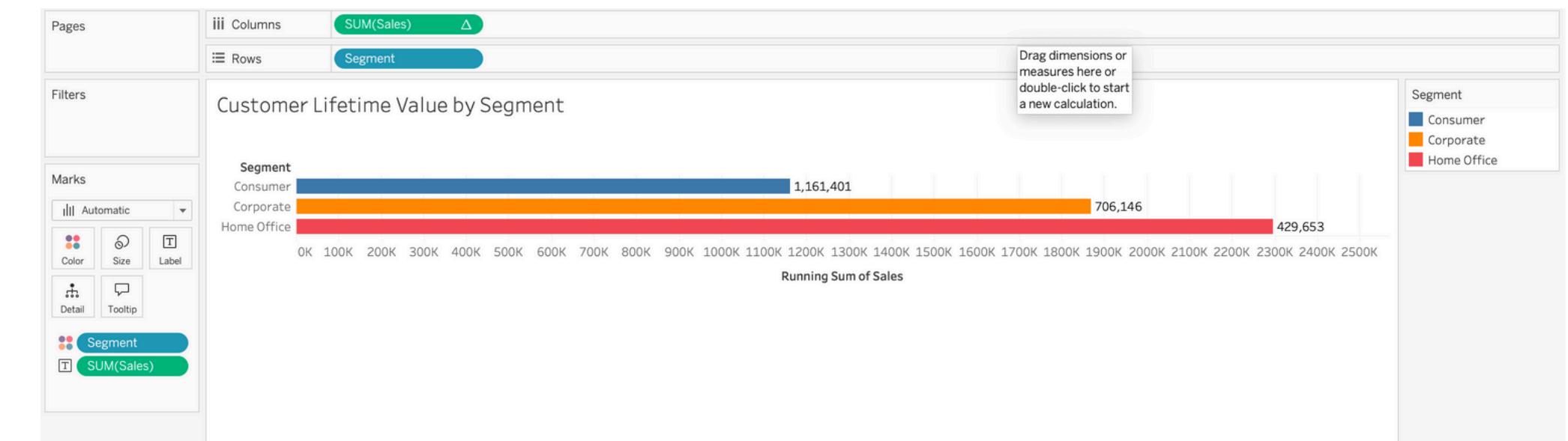
To compare CLV across segments: 1 Remove Customer Name from the Rows shelf.

- 2 Drag Segment to the Rows shelf.
- 3 Drag Sales to the Columns shelf → Convert to SUM.

Step 4: Enhance Visualization

✓ Show Exact CLV Values: Drag Sales to Label.

- Which segment has the highest CLV?



CROSS-DIMENSIONAL ANALYSIS

Creating an Interactive Dashboard for Customer Journey Analysis

To tell the story of a customer's journey, we will build:

- Customer Cohort Analysis (Tracking first purchase and retention rates).
- Customer Score (Evaluating Recency, Frequency, and Monetary factors).
- Bonus: Predictive Elements (Churn probability).

Step 1: Build a Customer Cohort Analysis

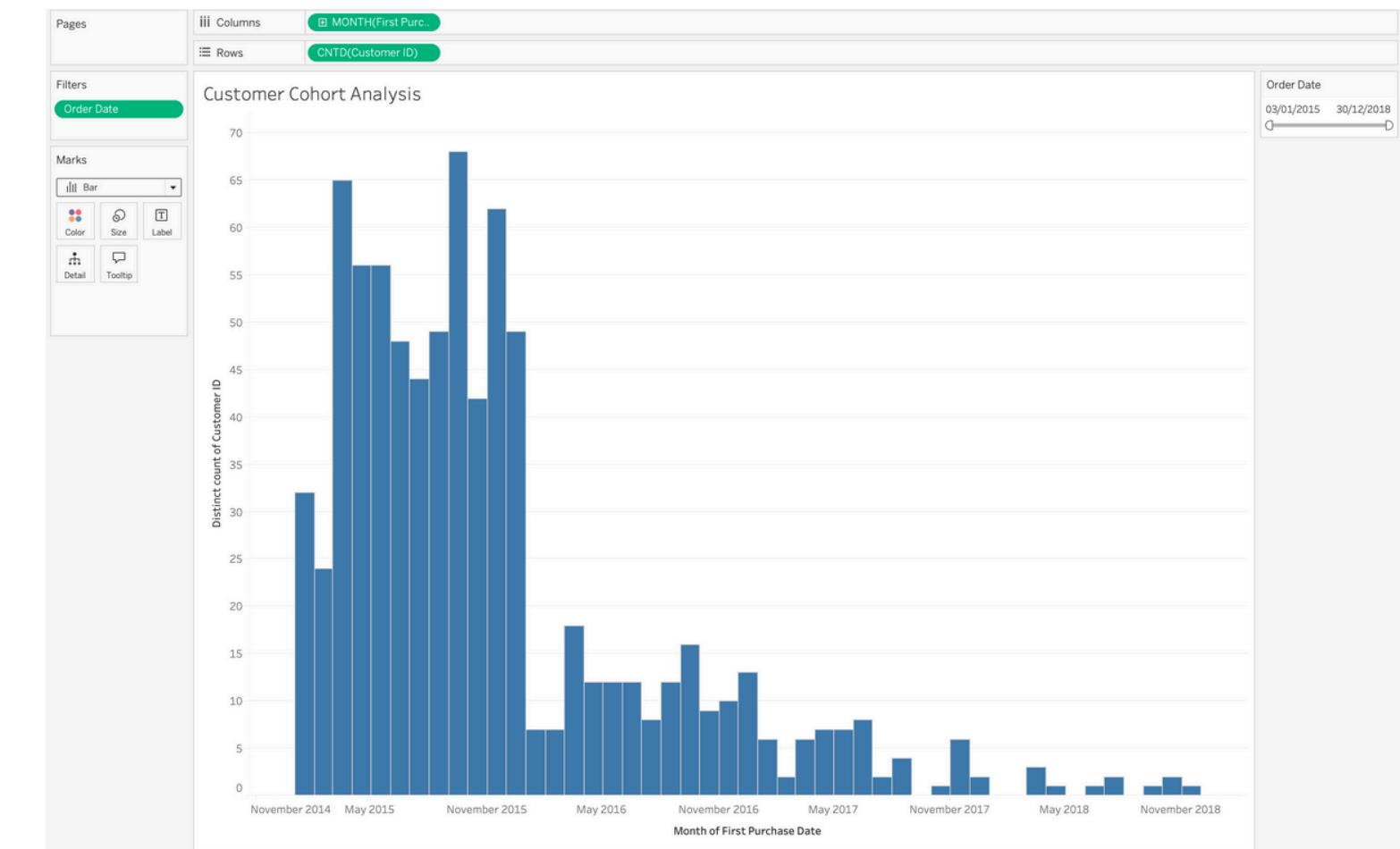
1 Group Customers by First Purchase Date

1. Click New Worksheet (+) → Rename it "Customer Cohort Analysis".
2. Click Data → Create Calculated Field → Name it "First Purchase Date".
 { FIXED [Customer ID]: MIN([Order Date]) }

This captures the first purchase date for each customer.

1. Drag First Purchase Date to the Columns shelf.
2. Drag Customer ID to the Rows shelf → Convert to Count (Distinct).
3. Drag Order Date to Filters
4. Click on Marks → Choose Bar Chart to show first purchase cohorts over time.

Now, customers are grouped by their first purchase month.



📌 Higher new customer acquisitions in 2015, then a drop after 2016.

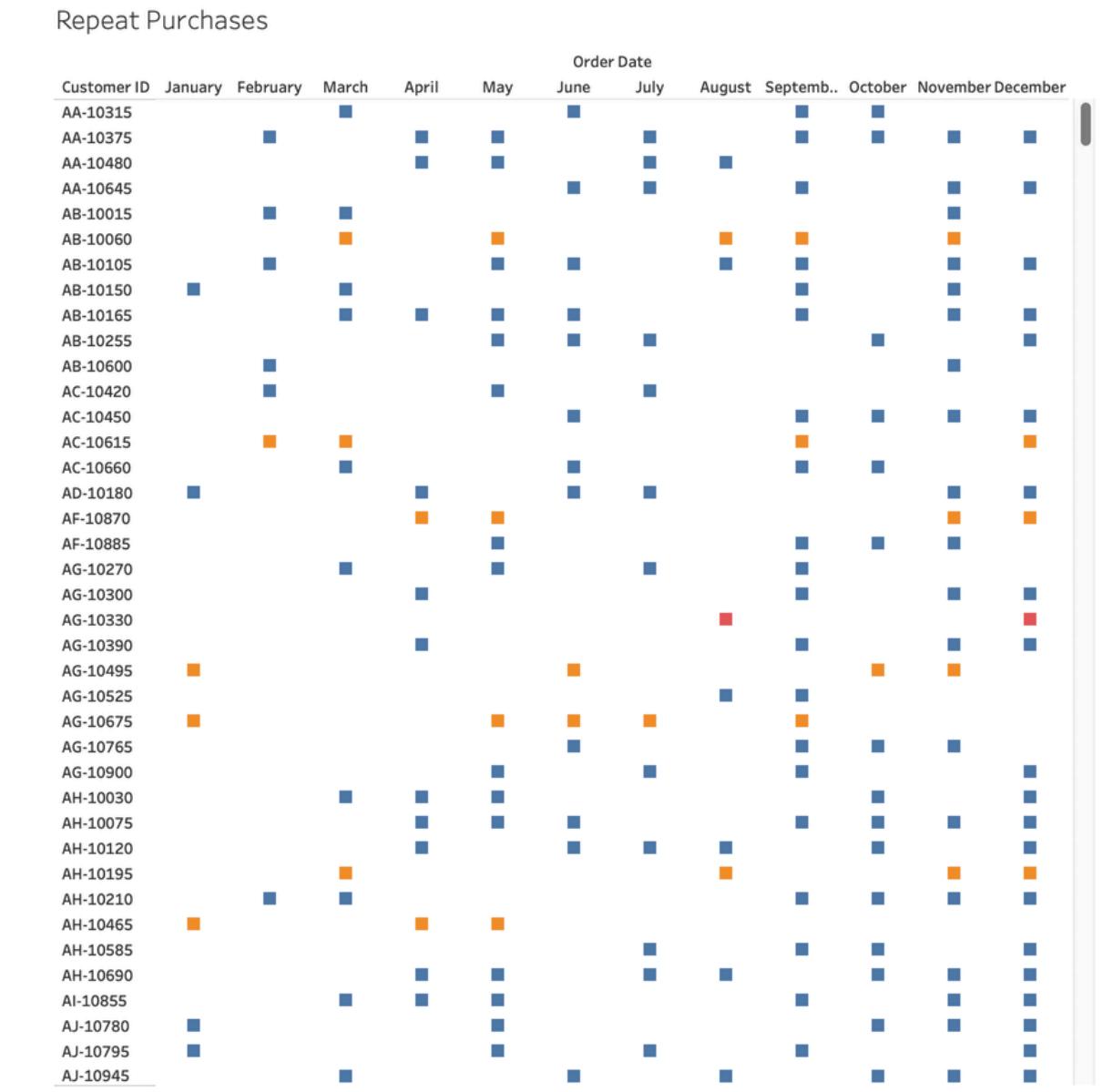
📌 Smaller cohorts in later years (2017-2018) suggest declining new customer growth.

CROSS-DIMENSIONAL ANALYSIS

Creating an Interactive Dashboard for Customer Journey Analysis

2 Track Repeat Purchase Behavior

1. Click New Worksheet (+) → Rename it "Repeat Purchases".
 2. Drag Order Date to the Columns shelf → Convert to Month.
 3. Drag Customer ID to the Rows shelf → Convert to Count (Distinct).
 4. Drag First Purchase Date to the Color mark (to segment customers by cohort)



CROSS-DIMENSIONAL ANALYSIS

Creating an Interactive Dashboard for Customer Journey Analysis

Step 2: Create a Customer Scores

- Recency → Days since last purchase.
- Frequency → Total orders placed.
- Monetary Value → Total revenue from the customer.
- Return Rate → Percentage of orders returned.
- Order Processing Time → Average time to ship.

1 Create a Scoring System

Create the following calculated fields:

Recency Score: DATEDIFF('day', { FIXED [Customer ID]: MAX([Order Date]) }, TODAY())

✓ Lower values mean more recent purchases (better score).

Frequency Score: { FIXED [Customer ID]: COUNT([Order ID]) }

✓ Higher values mean more frequent purchases (better score).

Monetary Score: { FIXED [Customer ID]: SUM([Sales]) }

✓ Higher revenue means higher value customers.

Order Processing Time: { FIXED [Customer ID]: AVG(DATEDIFF('day', [Order Date], [Ship Date])) }

✓ Longer processing times reduce the score.

CROSS-DIMENSIONAL ANALYSIS

Creating an Interactive Dashboard for Customer Journey Analysis

Return Rate:

🔗 Step 1: Join the "Orders" and "Returns" Tables

1 Go to the Data Source tab (bottom left).

2 Drag the Returns table next to the Orders table.

3 In the Join Configuration:

- Set the join condition as:
- Orders.[Order ID] = Returns.[Order ID (Returns)]
- Choose "Left Join" to keep all orders, even if they were not returned.

✓ Now, the Orders table includes return information.

📌 Step 2: Create the Return Rate Calculated Field

1 Click Data → Create Calculated Field.

2 Name it "Return Rate".

3 Enter this formula: $\text{SUM(IF [Returned] = "Yes" THEN 1 ELSE 0 END) / COUNT([Order ID])}$

✓ This formula calculates the percentage of returned orders out of total orders.

CROSS-DIMENSIONAL ANALYSIS

Creating an Interactive Dashboard for Customer Journey Analysis

Step 3: Predictive Elements

1 Predict Customer Churn Probability

- Create a Calculated Field called "Churn Risk":

IF [Recency Score] > 180 AND [Frequency Score] < 3 THEN "High Risk"

ELSEIF [Recency Score] > 90 THEN "Medium Risk"

ELSE "Low Risk"

END

✓ This flags customers likely to churn.

CROSS-DIMENSIONAL ANALYSIS

Creating an Interactive Dashboard for Customer Journey Analysis

Create a New Dashboard

- Click "New Dashboard" in the bottom toolbar
- Set size to "Automatic" or choose a fixed size (e.g., 1200 x 800)
- Name it "Customer Journey Analysis"

Add Cohort Analysis Section (Top Left)

- Drag the "Customer Cohort Analysis" worksheet onto the dashboard
- Add a title: "New Customer Acquisitions by Month"
- Add a date filter control for Order Date
- Size it to take up roughly the top left quarter

Add Retention Analysis Section (Top Right)

- Drag the "Repeat Purchases" worksheet next to the cohort analysis
- Add filter controls for:
 - cohort selector
 - customer ID
 - month of first purchase

CROSS-DIMENSIONAL ANALYSIS

Creating an Interactive Dashboard for Customer Journey Analysis

- Create Customer Scoring Section (Bottom Left)
- Create a new worksheet called "Customer Scoring Grid"
 - 1.Drag Frequency Score to the Columns (X-Axis). → Change it to "Dimension" or "Average"
 - 2.Drag Monetary Score to the Rows (Y-Axis). → Change it to "Dimension" or "Average".
 - 3.Drag Recency Score to the Size mark.
 - 4.Drag Return Rate to the Color mark.
 - 5.Click on the Marks drop-down → Select Circle for better visualization.
 - 6.Drag Customer ID to Detail in the Marks Pane.

Now, you have a scatter plot where:

- X-Axis: How often customers buy (Frequency Score).
- Y-Axis: How much they spend (Monetary Score).
- Size: How recent their last purchase was (Recency Score).
- Color: Return Rate (to identify high-returning customers).
- Add to dashboard below cohort analysis
- Click on the Filters pane (left side).
- Drag the following fields to the Filters shelf:
 - Recency Score
 - Frequency Score
 - Monetary Score
 - Return Rate
- Right-click on each filter → Select Show Filter.
- Move the filters to the right panel for easy access.

Now, users can filter customers based on scoring metrics dynamically.



CROSS-DIMENSIONAL ANALYSIS

Creating an Interactive Dashboard for Customer Journey Analysis

Add Churn Risk Section (Bottom Right)

- Create a new worksheet called "Churn Risk Analysis"
- 1 Drag Churn Risk to the Columns shelf.
- 2 Drag Customer ID to the Rows shelf → Convert to Count (Distinct)
- 3 Change Marks Type → Select Bar Chart
- 4 Drag Churn Risk to the Color mark to differentiate risk levels.
- 5 Click Sort Descending (so High-Risk appears first).
- Add it to the bottom right of the dashboard

Lastly, Adding an action filter allows you to click on a specific customer in your dashboard and view their entire journey across different visualizations.

TEST YOUR SKILLS!

Identifying Consistently Underperforming Products in Tableau Public

To find underperforming products, analyze products with low sales and/or negative profit over time. You can use three key methods:

- 1 Low Sales Products – Identify products with consistently low revenue.
- 2 Negative or Low Profit Products – Identify products that generate losses or very low profit margins.
- 3 Trend Analysis Over Time – Find products that have declined in performance over time.

Step 1: Identify Products with Low Sales

- 1 Click on New Worksheet (+) → Rename it "Low Sales Products".
- 2 Drag Product Name to the Rows shelf.
- 3 Drag Sales to the Columns shelf.
- 4 Click Sort Ascending (to list the lowest-selling products first).
- 5 Drag Category to the Color mark (to categorize products).
- 6 Apply a Top N Filter:
 - Click Filters → Drag Product Name to Filters.
 - Select By Field → Bottom 10 by SUM(Sales).

Now, you have a list of the lowest-selling products.

Step 2: Identify Products with Low or Negative Profit

- 1 Click on New Worksheet (+) → Rename it "Low Profit Products".
- 2 Drag Product Name to the Rows shelf.
- 3 Drag Profit to the Columns shelf.
- 4 Click Sort Ascending (to show the most unprofitable products first).
- 5 Apply a Profit Margin Filter:
 - Click Filters → Drag Profit Margin (%) (calculated earlier) to Filters.
 - Set Range from -100% to 5% (negative or low-margin products).

Now, you have a list of low-profit or loss-making products.

- 2 Drag all three sheets (Low Sales, Low Profit, and Underperforming Trends) onto the dashboard.
- 3 Add filters for Category, Region, and Year for dynamic exploration.

Now, you can interactively analyze which products consistently underperform.

TEST YOUR SKILLS!

Identifying Consistently Underperforming Products in Tableau Public

Step 3: Analyze Sales & Profit Trends Over Time

- 1 Click on New Worksheet (+) → Rename it "Underperforming Trends".
- 2 Drag Order Date to the Columns shelf.
- 3 Drag Sales and Profit to the Rows shelf.
- 4 Drag Product Name to Filters → Select Bottom 10 by SUM(Sales & Profit).
- 5 Click on Marks → Change to Line Chart.

 This shows how low-performing products have trended over time.

Step 4: Build an Interactive Dashboard

- 1 Click on New Dashboard (+).
 - 2 Drag all three sheets (Low Sales, Low Profit, and Underperforming Trends) onto the dashboard.
 - 3 Add filters for Category, Region, and Year for dynamic exploration.
-  Now, you can interactively analyze which products consistently underperform.

PUBLISHING YOUR TABLEAU DASHBOARD

Step 1: Ensure Your Dashboard is Ready

- Make sure your Tableau workbook is finalized and saved.
- If you're using private data, be aware that publishing on Tableau Public makes your data publicly accessible.

Step 2: Sign in to Tableau Public

- Open Tableau Desktop.
- Click "Server" > "Tableau Public" > "Save to Tableau Public".
- If you don't have an account, you will need to sign up for Tableau Public.

Step 3: Publish the Dashboard

- Once you log in, Tableau will upload your workbook to Tableau Public.
- It may take a few minutes depending on the size of your dashboard.

Step 4: Get the Shareable Link

- After uploading, Tableau Public will open your dashboard in the browser.
- You can copy the URL from the browser and share it with others.

Step 5: Embed or Download

- Tableau Public provides an Embed Code (under "Share") to insert the dashboard into a website or blog.
- Viewers can interact with your visualization directly from the browser.

Notes:

- Make sure your workbook is saved as .twbx (Tableau Packaged Workbook) before publishing.
- Tableau Public does not support live connections – ensure your data is static or extracted.
- If your data is sensitive, use Tableau Server or Tableau Cloud for private sharing.

THANK YOU.

