

MODULE 3

DATA TRANSFORMATION

Data transformation is the process of converting data from one format or structure into another to make it more suitable for analysis, visualization, or other purposes. It involves cleaning, aggregating, merging, and restructuring data to make it more useful and meaningful. Data transformation is a crucial step in the data analysis process, as raw data often needs to be processed and formatted before it can be effectively analyzed or visualized.

Some common data transformation tasks include:

1. **Cleaning:** Removing or correcting errors, duplicates, and inconsistencies in the data.
2. **Filtering:** Selecting a subset of data based on specific criteria or conditions.
3. **Aggregating:** Combining multiple data points into a summary or aggregated value, such as calculating averages or totals.
4. **Joining/merging:** Combining data from multiple sources into a single dataset based on common keys or fields.
5. **Pivoting/un-pivoting:** Changing the structure of the data from wide to tall (or vice versa) to make it easier to analyze.
6. **Formatting:** Converting data into a standard format (e.g., date formatting) for consistency and compatibility with analysis tools.

Data transformation is often performed using tools like Tableau, Excel, Python, or SQL, depending on the complexity of the transformation and the requirements of the analysis.

DATA TRANSFORMATION IN TABLEAU:

To transform data in Tableau, you can use various features like calculated fields, data blending, data pivoting, and custom SQL. Here's a general overview of how you can perform data transformations in Tableau:

1. **Calculated Fields:** Use calculated fields to create new fields based on existing data. You can perform calculations, string operations, date calculations, and more.
2. **Data Blending:** Blend data from multiple sources by defining relationships between them. This can help you combine data for analysis.
3. **Data Pivoting:** Pivot your data to change the layout from wide to tall or vice versa. This can be useful when your data is not in the desired format for analysis.

4. Custom SQL: Write custom SQL queries to extract, transform, and load data into Tableau. This is useful for complex transformations that cannot be achieved using other methods.
5. Tableau Prep: Use Tableau Prep, a separate tool from Tableau, to visually prepare and clean your data before bringing it into Tableau for analysis.

Each of these methods has its strengths and use cases, so the best approach depends on your specific requirements and the structure of your data.

DATA VISUALISATION BEST PRACTICES:

1. Understand Your Audience: Consider who will be viewing your visualization and tailor it to their needs and knowledge level. Use appropriate terminology and simplify complex information for non-experts.
2. Keep it Simple: Avoid clutter and unnecessary elements. Use a clean and intuitive design to convey information effectively.
3. Choose the Right Chart Type: Select a chart type that best represents your data and makes it easy to understand. For example, use a bar chart for comparisons and a line chart for trends over time.
4. Use Color Wisely: Use color to highlight important information and create visual hierarchy. Avoid using too many colors or overly bright colors that can distract from the data.
5. Provide Context: Include labels, titles, and annotations to provide context and help users understand the data. Use tooltips to provide additional information when needed.
6. Ensure Accessibility: Make sure your visualization is accessible to all users, including those with visual impairments. Use accessible colors, provide alternative text for images, and use clear and simple language.
7. Test and Iterate: Test your visualization with real users to ensure it is effective and easy to understand. Iterate based on feedback to improve its usability and clarity.
8. Tell a Story: Use your visualization to tell a story or convey a message. Structure your visualization in a way that guides the viewer through the data and highlights key insights.
9. Consider Interactivity: Use interactive elements like filters, tooltips, and drill-downs to allow users to explore the data and gain deeper insights.
10. Use Consistent Design: Maintain a consistent design across your visualization to create a cohesive and professional look. Use consistent colors, fonts, and styles throughout.

By following these best practices, you can create data visualizations that are engaging, informative, and easy to understand.

CALCULATIONS IN TABLEAU:

In Tableau, calculations are essential for analyzing and presenting data. There are different types of calculations you can create:

1. Basic Calculations

Basic calculations allow you to create new fields based on your existing data.

- **Calculated Fields:** Create new data from your existing data using mathematical operations, logical statements, and other functions.

2. Table Calculations

Table calculations are transformations applied to the values in a visualization. They are computed based on the data in the view.

- **Running Total:** Sum of values up to the current point in the data.
- **Percent of Total:** Computes each value as a percentage of the total.

3. Level of Detail (LOD) Calculations

LOD calculations provide a way to compute values at the data source level and the visualization level. They allow you to control the level of granularity for your calculations.

4. Aggregate Calculations

These calculations involve aggregating your data in some way, like summing, averaging, or finding the maximum or minimum values.

5. String Calculations

String calculations allow you to manipulate text fields.

6. Date Calculations

Date calculations help manipulate date fields to perform tasks like extracting parts of a date or calculating the difference between dates.

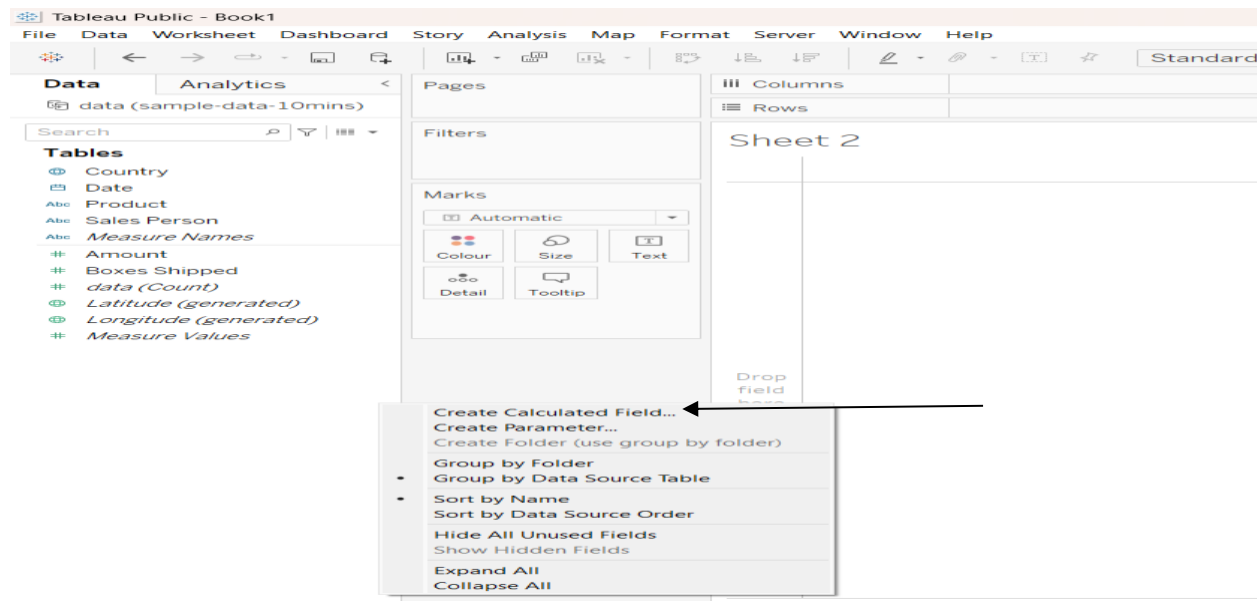
CREATING SIMPLE CALCULATIONS IN TABLEAU:

In Tableau, you can create simple calculations using calculated fields. Calculated fields allow you to perform basic arithmetic operations, string manipulation, date calculations, and more. Here's how you can create a simple calculation in Tableau:

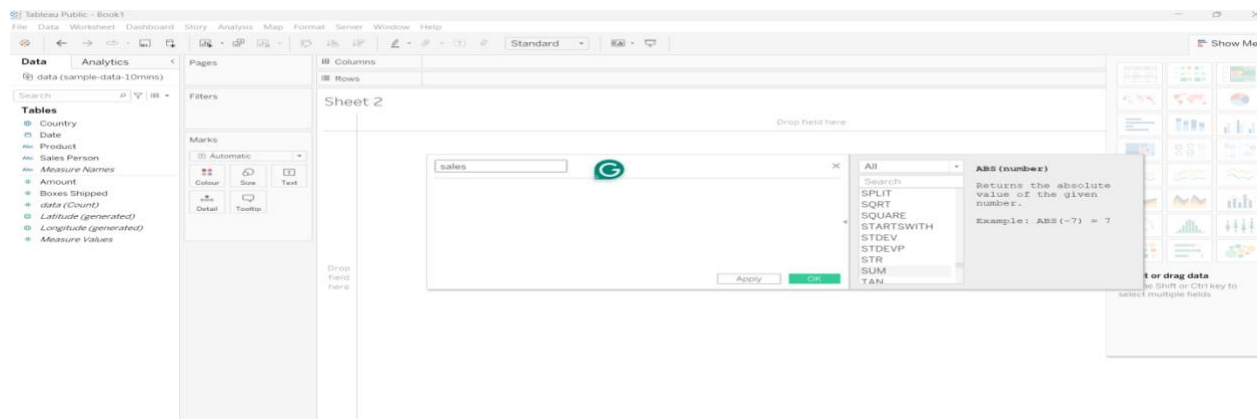
1. Open Tableau: Open Tableau Desktop and connect to your data source.
2. Navigate to the Data Source tab: Click on the Data Source tab at the bottom of the screen to start creating your calculated field.
3. Create a Calculated Field:
 - Right-click on a blank space in the Data Source tab.
 - Select "Create Calculated Field."
 - Enter a name for your calculated field.
 - In the calculation editor, write your calculation using the available functions and fields from your data source.
 - Click OK to save your calculated field.
4. Use the Calculated Field in your Visualizations:
 - Navigate back to a worksheet.
 - Drag the calculated field from the Data pane to the desired shelf (rows, columns, or marks).
 - Tableau will automatically apply the calculation to your visualization.

Here's an example of a simple calculation in Tableau that calculates the total sales amount:
SUM([Amount])

Step 1: click on create calculated field.

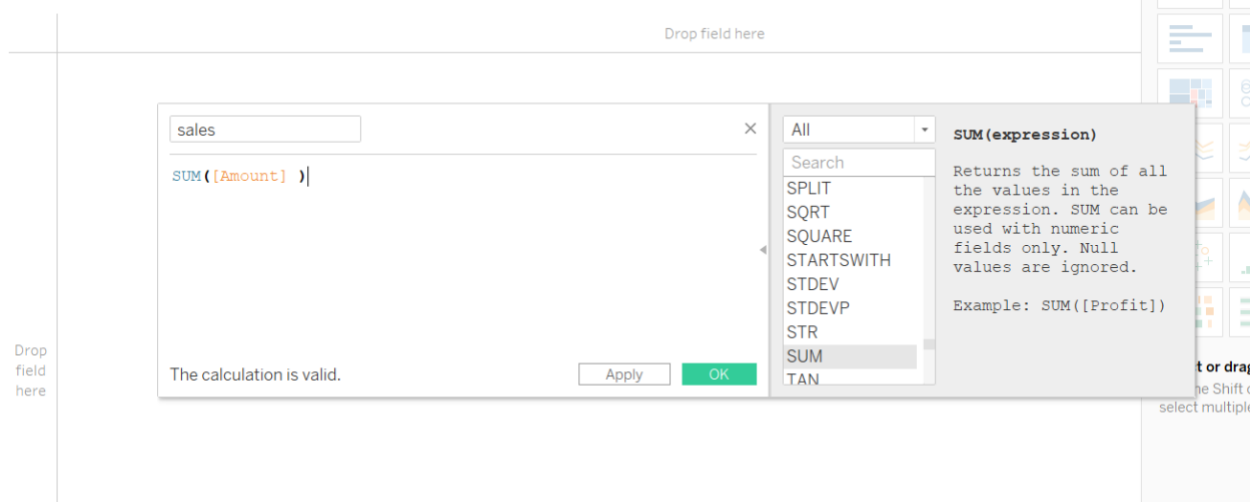


Step 2: Name the calculation as “sales” and select the sum function.

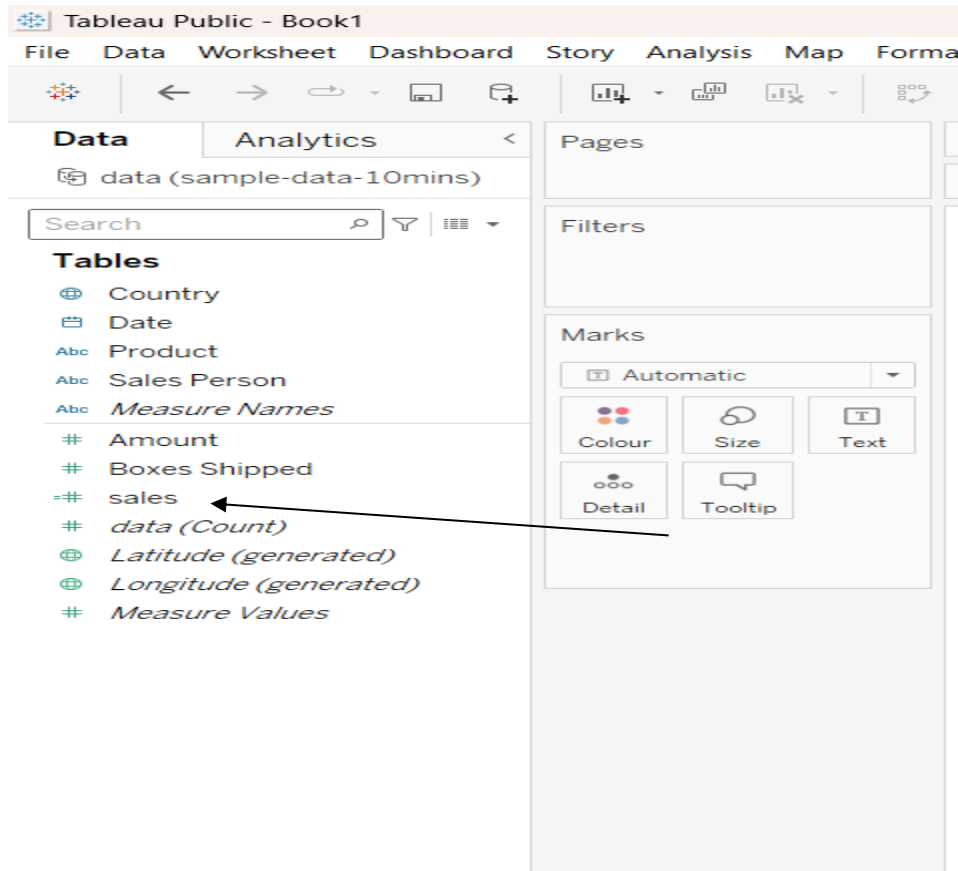


Step 3: sum the amount by adding Amount measure.

Sheet 2



Then our calculated field “sales” are created.



In this calculation, SUM([Amount]) calculates the sum of the sales from your data source. You can replace [Sales] with the actual field name from your data source.

You can create more complex calculations using other functions and operators as needed for your analysis.

USING TABLE CALCULATIONS:

Table calculations in Tableau are powerful ways to perform computations on the data in your visualization. They operate on the result of an existing query and allow you to transform and manipulate data within the context of your visualizations. Here’s an example to illustrate the use of table calculations in Tableau:

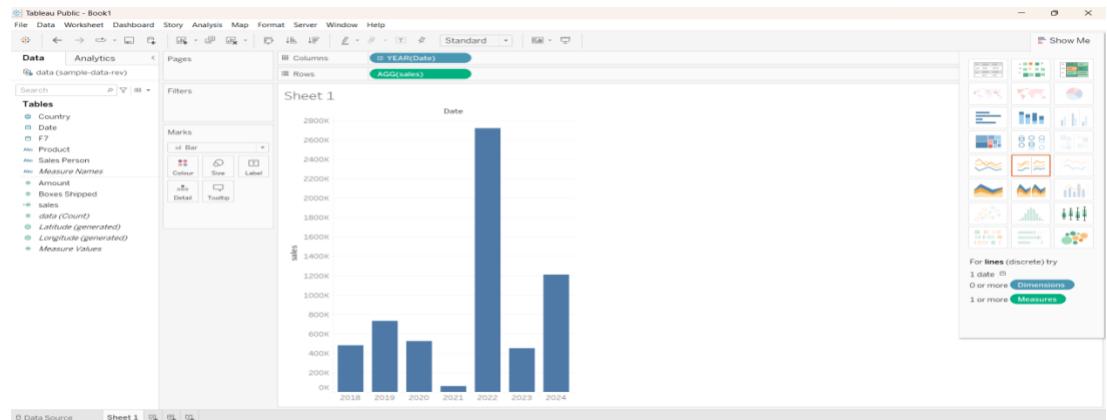
Example: Calculating Year-over-Year Growth

Let's say you have a dataset containing monthly sales data, and you want to calculate the year-over-year (YoY) growth in sales.

1. Connect to your Data: Connect to your data source in Tableau, which contains at least two fields: Date and Sales.

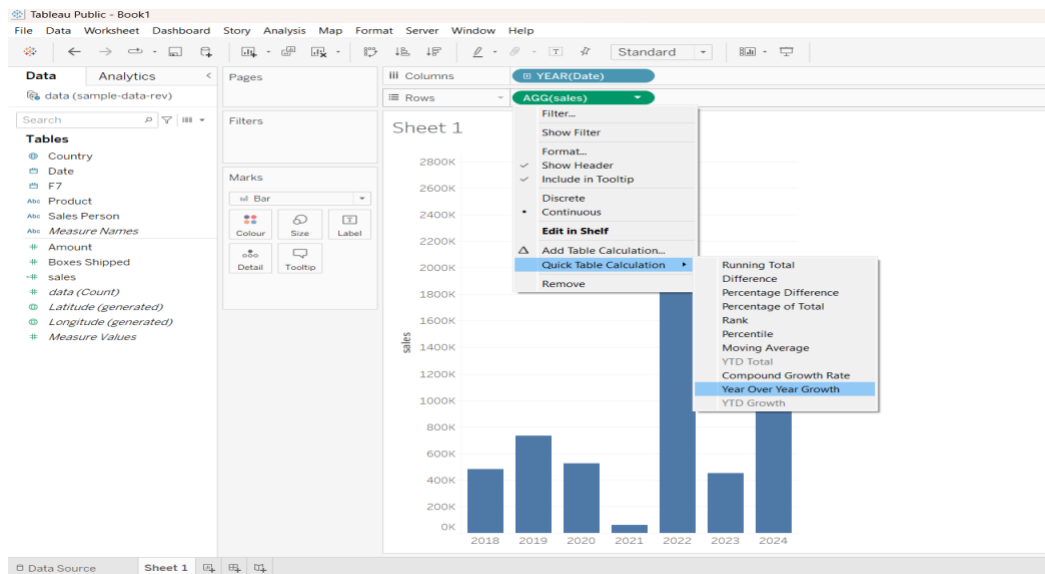
2. Create a Visualization:

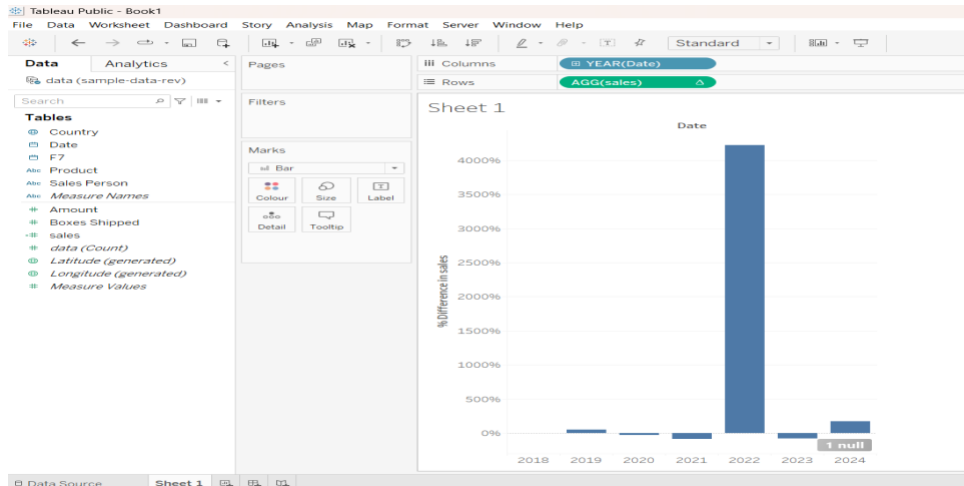
- Drag Date to the Columns shelf.
- Drag Sales to the Rows shelf.
- Change the Date field to display by month (right-click on the Date field, select "More", and then "Month").



3. Add a Table Calculation:

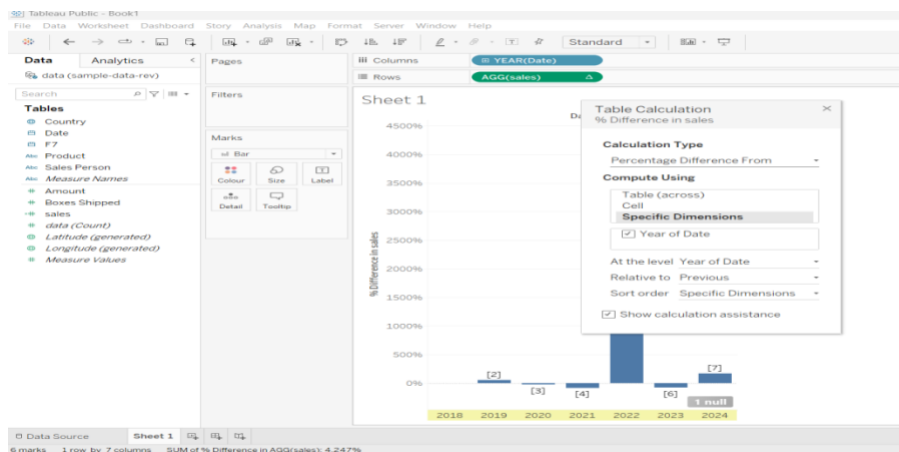
- Right-click on the Sales field in the Rows shelf.
- Select "Quick Table Calculation" and then choose "Year over Year Growth".





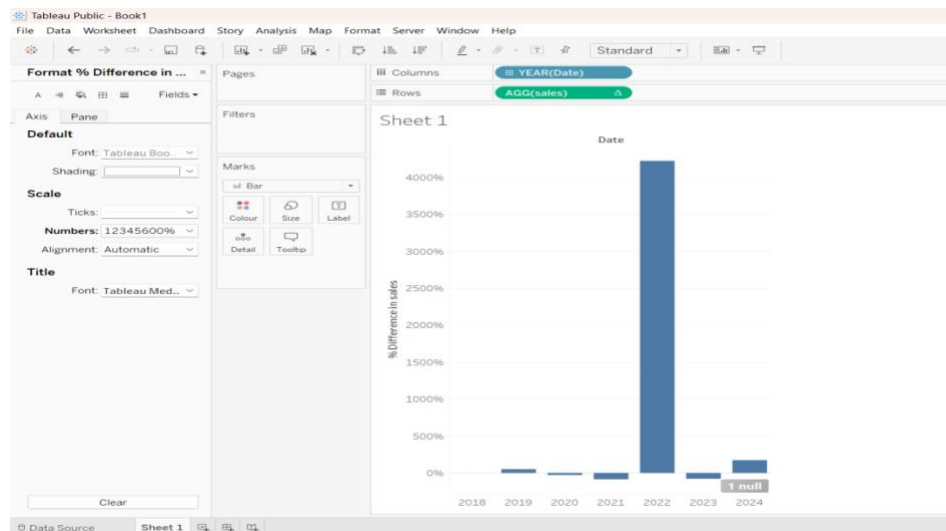
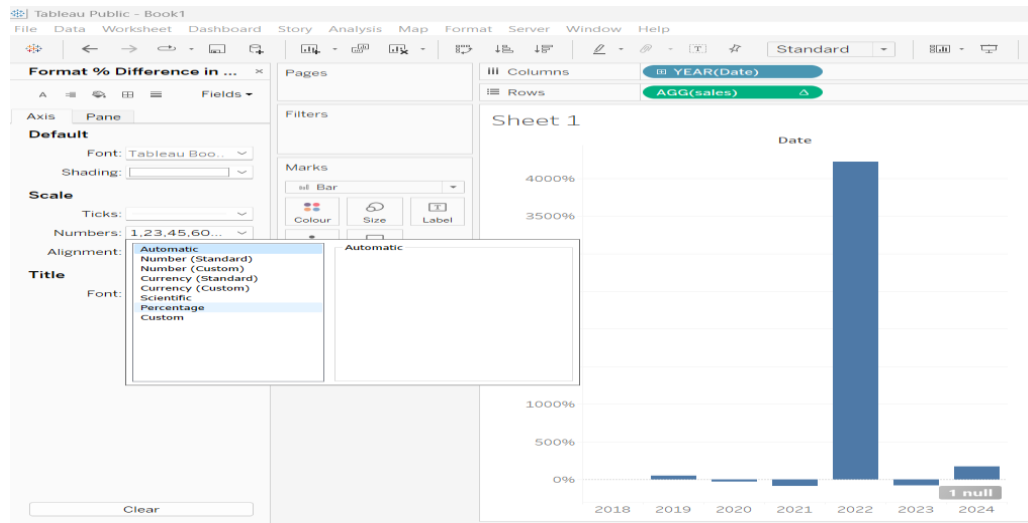
4. Customize the Calculation:

- If needed, you can customize the table calculation. Right-click on the Sales field again, select "Edit Table Calculation".
- In the dialog box, you can change the specific settings for the table calculation, such as the "Compute Using" option to ensure it computes along the correct dimension (e.g., Date).



5. Format the Result:

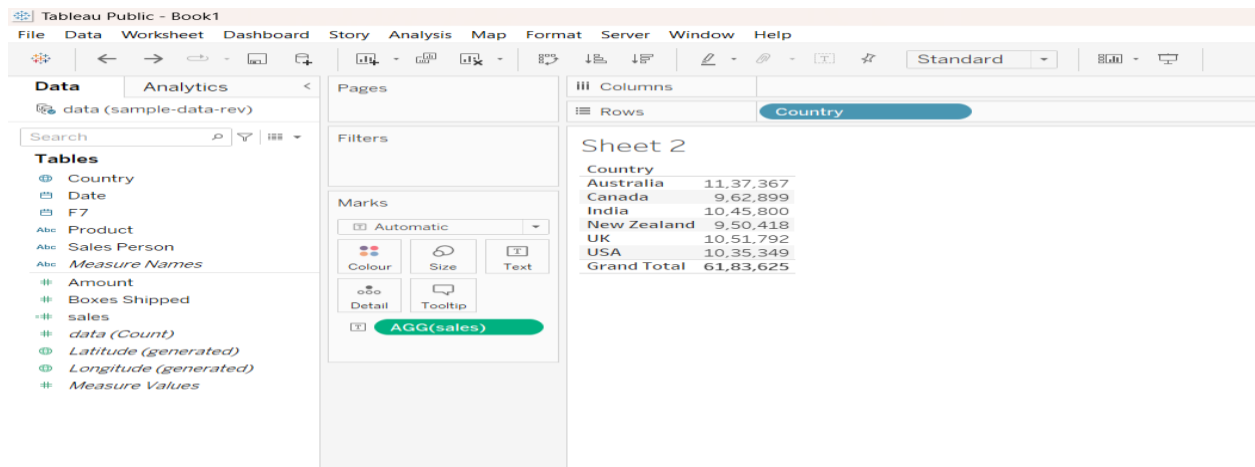
- You might want to format the result to show the growth as a percentage. Right-click on the Sales field, select "Format", and then choose the percentage format.



Example2: Calculating percentage of total sales. Here we want to see which country giving how much percentage of sales.

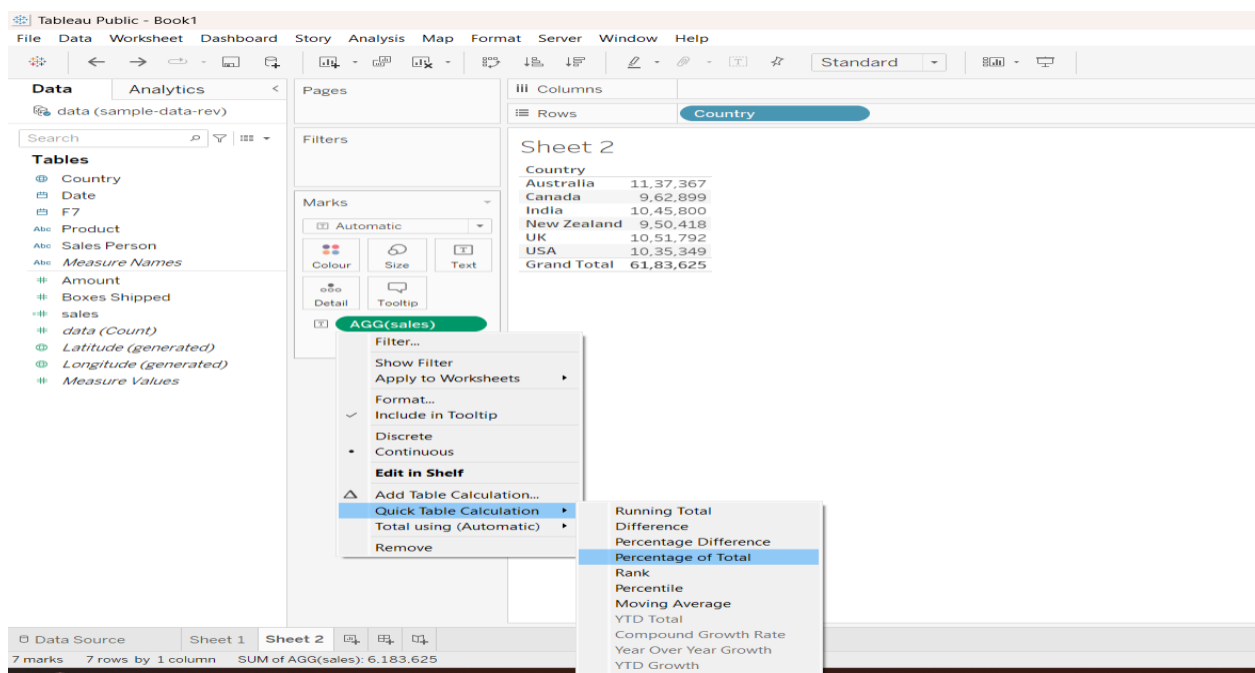
1. Connecting to Data and Creating the Initial Visualization:

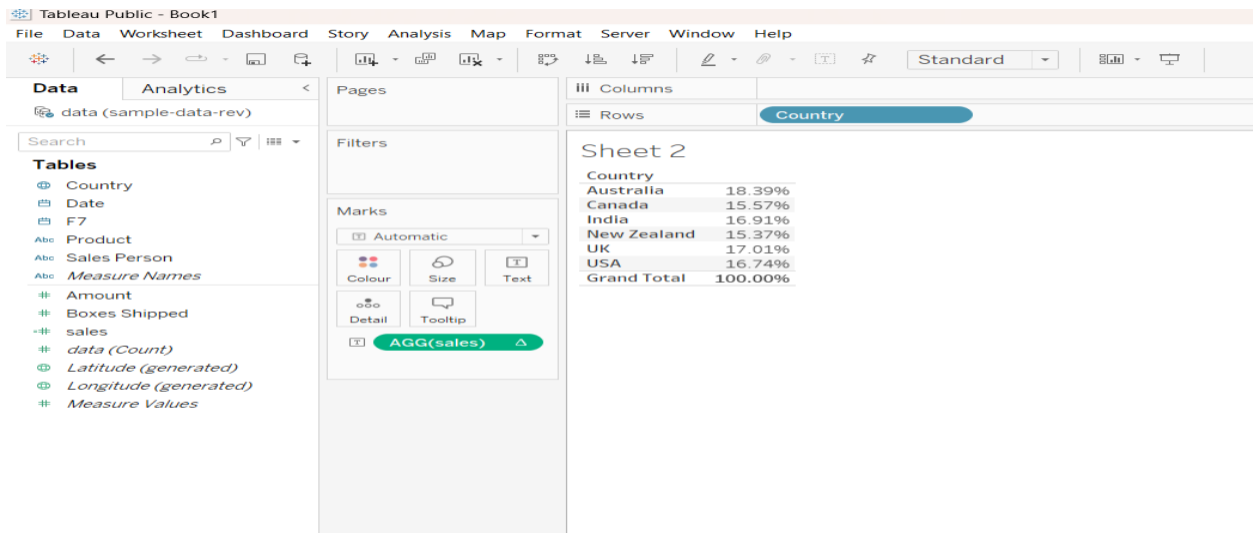
- Ensure your dataset is loaded into Tableau.
- Drag the Country field to the Columns shelf and the Sales field to the Rows shelf.



2. Adding and Customizing the Table Calculation:

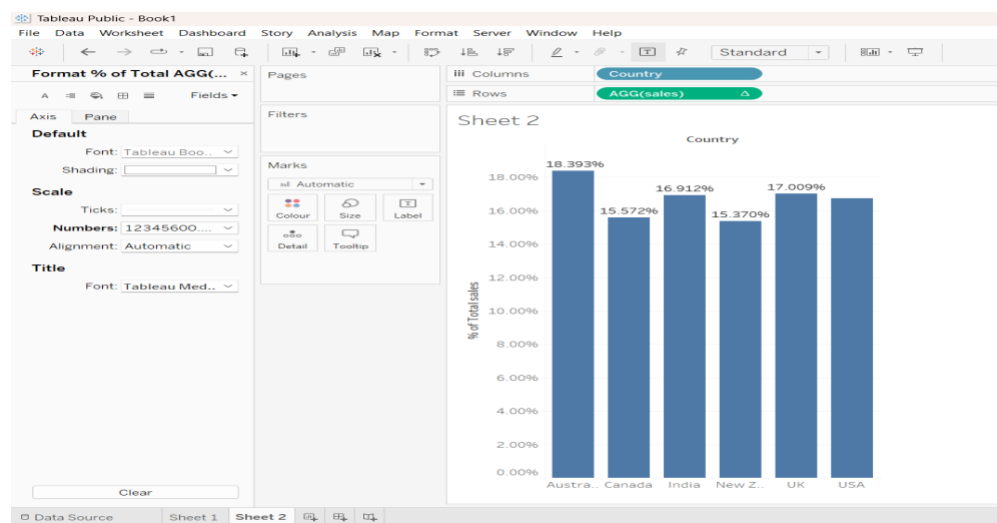
- Right-click on the Sales field in the Rows shelf, select "Quick Table Calculation", and choose "percentage of total".
- For further customization, right-click on the Sales field again, choose "Edit Table Calculation", and adjust the settings as needed, such as setting "Compute Using" to Date.





3. Formatting the Result:

- Format the field to display the results as percentages by right-clicking on the Sales field, selecting "Format", and choosing the percentage format.



This should give you a clear visualization of the year-over-year sales growth and percentage of total in Tableau using table calculations. You can apply similar steps for other types of calculations like moving averages, and running totals by selecting the appropriate quick table calculation and adjusting as necessary.