

Tableau Lecture 1: Intro to Data Visualization & Tableau

Agenda

- Part A : Introduction to Data Visualization and Tableau
 - Discover what data visualization is, and how we can use it to better understand data.
 - We acquaint ourselves with the history of data visualization and get introduced to Tableau. Then, we get an idea about the various Tableau product offerings.
 - We also learn about the pros and cons of Tableau.
 - Finally, we see how to download and install Tableau Public.
- Part B : Tableau Public GUI and Basics
 - In this section, you will get a brief idea about the Tableau UI components, different data types and various fields in Tableau.

Netflix Viz & Dashboard Demonstration

Dataset : [Netflix Movies & TV shows](#)

Demonstration :

<https://drive.google.com/file/d/1efZEbXNxJo46WL64cs6ON7gaevvsilta/view?usp=sharing>

Tableau Public Dashboard :

https://public.tableau.com/app/profile/tino3819/viz/NetflixDashboard_16521889408740/Dashboard1

What is Business Intelligence (BI)

BI combines business analytics, data mining, data visualization, data tools and infrastructure, and best practices to help organizations to make more data-driven decisions.

Why BI?

It helps companies make better decisions by showing present and historical data within their business context.

What is Data Visualization?

- It is the graphical representation of information and data.
- By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data.
- Humans are good at deriving knowledge from visualizations.

Brief History

- Until the early 21st century, Database, Excel, Access etc were used to produce numbers and data.
- The main idea behind Tableau's creation was to make the database industry interactive and comprehensive.
- Tableau is a popular data visualization and business intelligence tool used for reporting and analyzing vast volumes of data.

- Tableau was founded by Pat Hanrahan, Christian Chabot, and Chris Stolte from Stanford University in 2003.

Over the years..

- Tableau has been named a Leader in the Gartner Magic Quadrant for Analytics & Business Intelligence Platforms for the 10th consecutive year.
- Tableau Software has a market capitalization of \$14.61 billion and generates \$982.95 million in revenue each year.



Pros of Tableau

- Quick and interactive visualizations
- Easy to use for non programmers
- High performance
- Mobile friendly
- Extensive customer resources (Tableau Community)

- Working with different data sources
- Easy to upgrade

Cons of Tableau

- Focuses primarily on visualization and cannot work with uncleaned data. In order to efficiently use Tableau, you need to do proper data cleaning in the underlying database first.
- Lacks data modeling and data dictionary capabilities for Data Analysts. This means that you've to separately maintain your metrics definitions elsewhere.
- Lack of version control and collaboration when building data logic and dashboard.

Tableau Product Suite

- Tableau Prep
- Tableau Desktop
- Tableau Server
- Tableau Online
- Tableau Reader
- Tableau Public

References:

- [Comparison of Product Suite](#)
- [Tableau Pricing](#)
- [Understanding License types of Tableau](#)
- [Tableau Desktop vs Tableau Public](#)

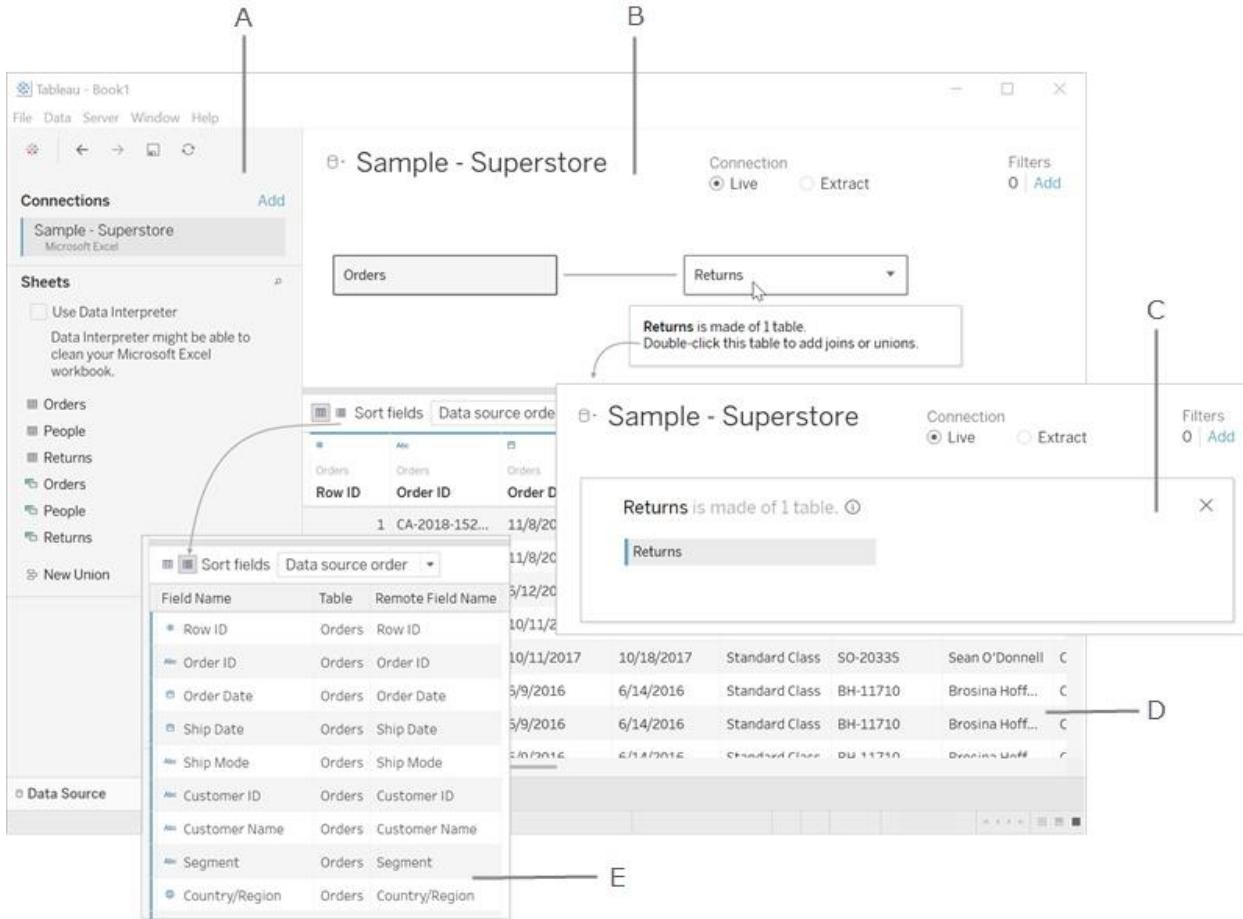
Tableau Public Installation

- Pre-read Document: [link](#)
- Tableau Public download link : [link](#)
- Tableau Public Desktop installation steps: [link](#)

Tableau Public GUI

Loading dataset in Tableau and Data Source view

Dataset: [link](#)



- Left pane - Displays the connected data source and other details about your data.
- Canvas: logical layer - The canvas opens with the logical layer, where you can create relationships between logical tables.
- Canvas: physical layer - Double-click a table in the logical layer to go to the physical layer of the canvas, where you can add joins and unions between tables.
- Data grid - Displays first 1,000 rows of the data contained in the Tableau data source.
- Metadata grid - Displays the fields in your data source as rows.

Note: Canvas will be explained in future lectures.

Reference: https://help.tableau.com/current/pro/desktop/en-us/environment_datasource_page.htm

Data Fields

- After you connect to your data and set up the data source with Tableau, the data source connections and fields appear on the left side of the workbook in the Data pane.
- The term "fields" refers to columns.
- When you connect to a new data source :
 - Each field is automatically assigned a Data Type (such as integer, string, date).

- Tableau assigns each field in the data source as dimension or measure in the Data pane, depending on the type of data the field contains.
 - Discrete Dimension or Continuous Measure (more common)
 - Continuous Dimension or Discrete Measure (less common)

Reference:

- https://help.tableau.com/current/pro/desktop/en-us/datafields_typesandroles.htm
- https://help.tableau.com/current/pro/desktop/en-us/datafields_understanddatawindow.htm

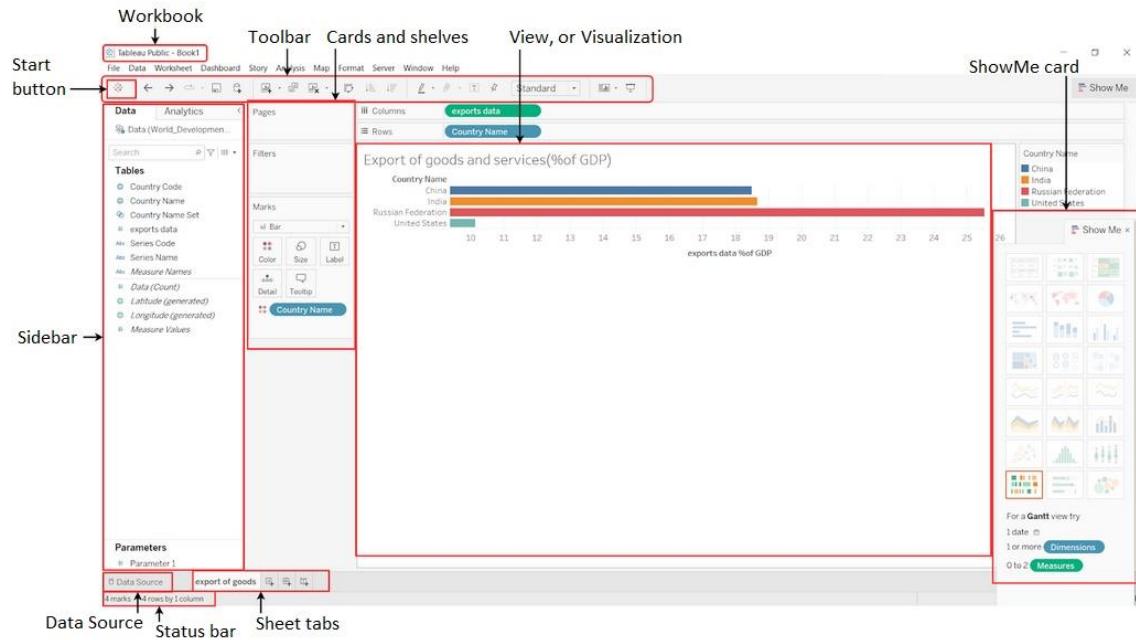
Data Type

- Tableau expresses fields and assigns data types automatically.
- If the data source appoints the data type, Tableau will use that data type.
- If the data source doesn't individually assign a data type, Tableau will assign one

Tableau Data Types

Data Type	Icon
String values (Text)	
Integer values (Numbers)	
Date values (DD/MM/YYYY or MM/DD/YYYY)	
Date & Time values	
Boolean values (True or False; relational)	
Geographic values (Region, Postal code etc.)	
Cluster group or mixed values	

[Tableau View page](#)



Measure:

- A measure is a field that is a dependent variable; that is, its value is a function of one or more dimensions.
- Tableau treats any field containing numeric (quantitative) information as a measure.

Dimension:

- Dimension is a field that can be considered an independent variable.
- By default, Tableau treats any field containing qualitative, categorical information as a dimension.

Rule of Thumb :

Generally,

- the measure is the number;
- the dimension is what you “slice and dice” the number by.

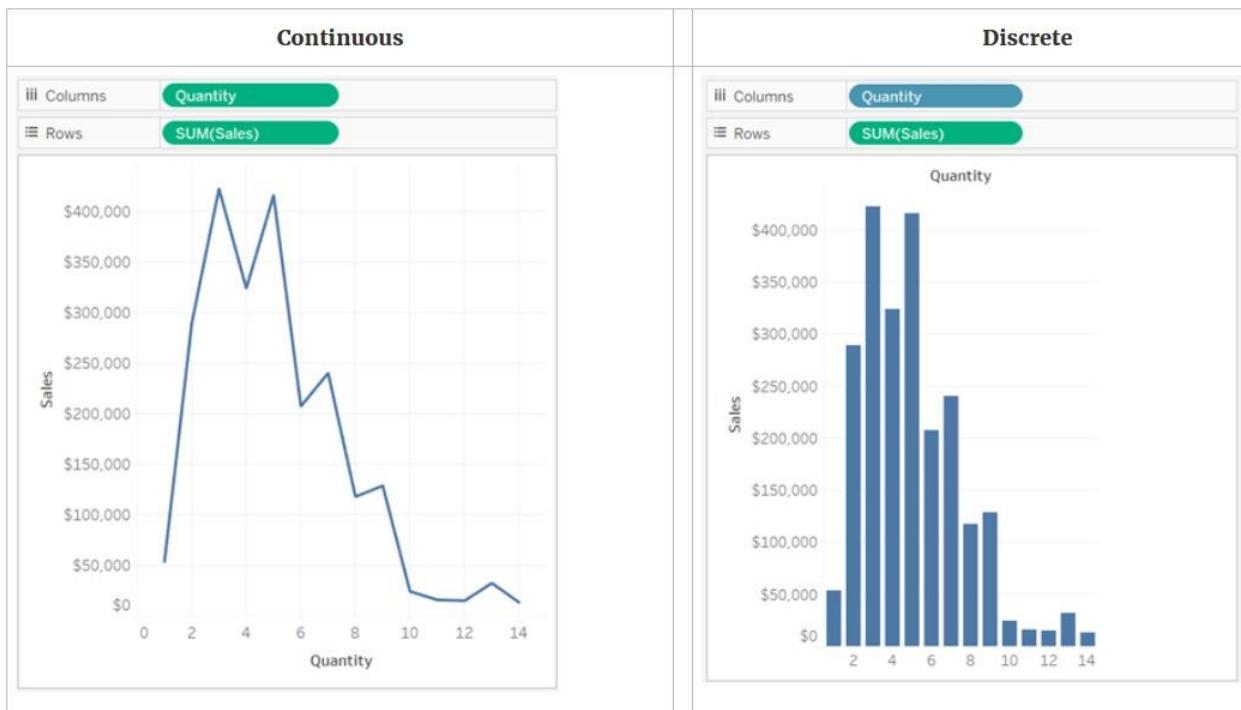
Discrete and Continuous fields

- Tableau represents data differently in the view depending on whether the field is discrete (blue), or continuous (green).
- Continuous and Discrete are mathematical terms.
 - Continuous means "forming an unbroken whole, without interruption"
 - Discrete means "individually separate and distinct"

- Green measures **SUM(Profit)** and dimensions **YEAR(Order Date)** are continuous. Continuous field values are treated as an infinite range. Generally, continuous fields add axes to the view.
- Blue measures **SUM(Profit)** and dimensions **Product Name** are discrete. Discrete values are treated as finite. Generally, discrete fields add headers to the view.

Rule of Thumb :

- Discrete fields draw headers; Continuous fields draw axes.
- Discrete fields can be sorted; Continuous fields cannot.
- Blue color field indicates Discrete Field
- Green color field indicates Continuous Field



- In the example on the left (above), because the Quantity field is set to Continuous, it creates a horizontal axis along the bottom of the view.
 - The green background and the axis help you to see that it's a continuous field.
- In the example on the right (above), the Quantity field has been set to Discrete. It creates horizontal headers instead of an axis.
 - The blue background and the horizontal headers help you to see that it's discrete.
- In both examples, the Sales field is set to Continuous.
 - It creates a vertical axis because it is Continuous and it's been added to the Rows shelf.
 - If it was on the Columns shelf, it would create a horizontal axis.

- The green background and aggregation function (in this case, SUM) help to indicate that it's a Measure.
- The absence of an aggregation function in the Quantity field name helps to indicate that it's a Dimension.

Converting Measure to Dimension

- You can convert a field from a measure to a dimension in the current view.
- If you want the change to affect all future uses of the field in the workbook, you can convert a field in the Data pane from a measure to a dimension.

To convert a measure to a dimension in the Data pane, do either of the following :

- Click and drag the field from the measures area in the Data pane and drop it into a dimensions area (above the line).
- Right-click (control-click on a Mac) the measure in the Data pane and select Convert to Dimension.

If you place a field that you converted from a measure to a dimension on a shelf, it now produces headers instead of an axis.

Business problem 1:

Dataset : [sample superstore](#)

Determine the total sales value of each category.

- Use the Orders table
- Drag Category to Columns
- Drag Sales to Rows
- Switch to Entire View
- Click on Label and check "Show marks labels"

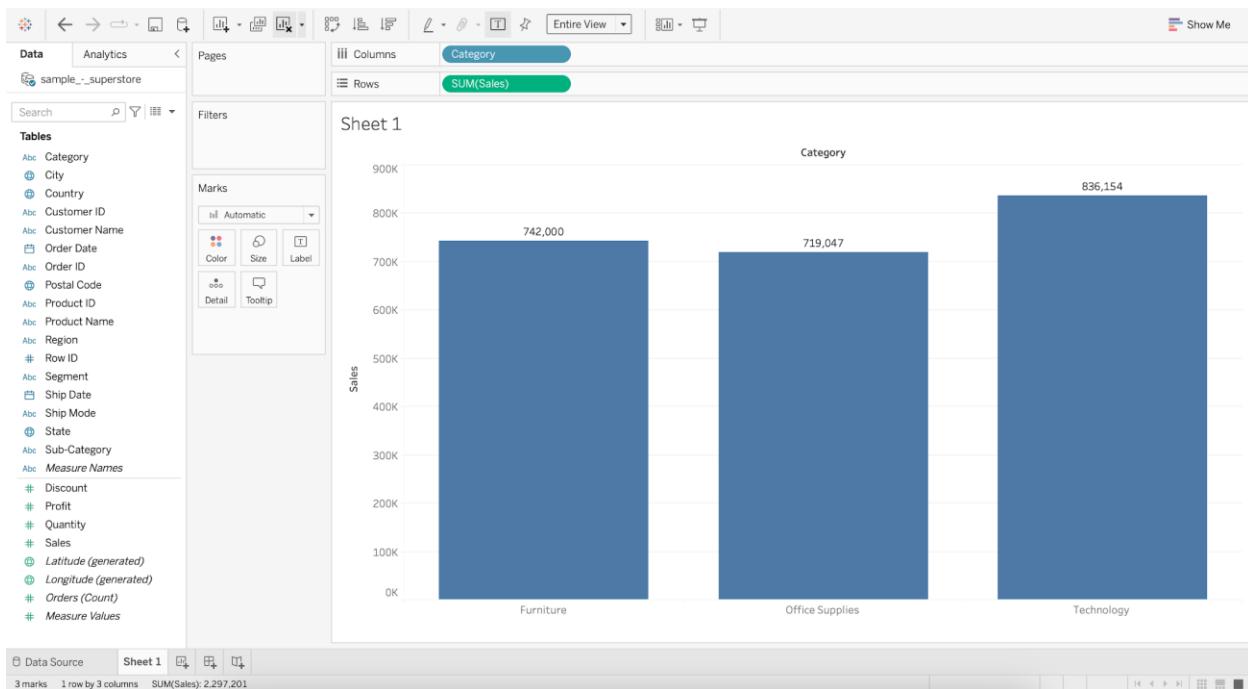


Tableau Lecture 2: Basic Charts and Operations

Agenda

- Data Connection in Tableau
- Aggregation on Dimensions and Measures
- Basic Charts
- Cleaning and Formatting data for analysis
- Introduction to Summer Olympics dataset

What is Visual Analytics?

- It is the use of sophisticated tools and processes to analyze datasets using visual representations of the data.

Why is it important?

- Users get actionable insights which in turn help organizations make better, data-driven decisions.
- Allows users without data science skills or experience to combine, manipulate, and explore large, dynamic, multi-dimensional, and multi-sourced datasets.

Difference between Visual Analytics and Data Visualization

Data Visualization

Visual Analytics

Graphical/Visual depiction of data to help people better understand the patterns, relationships, trends, and other meaningful insights in datasets.	The use of an analytics program to perform advanced analysis of complex datasets allowing users to explore and interact with dynamic visualizations.

Tableau Live and Extract Connections

Live and Extracts are two ways you can make the data connection to the Tableau.

Live connection	Extract connection
Live allows you real-time data.	Extract is kind of a batch which needs to be refreshed from time to time to get the updated data.
Whatever changes are done at the data source will be directly available to Tableau Desktop (professional).	Any changes made in the data source won't reflect in the report immediately. It will be reflected when the extract is refreshed.

Note:

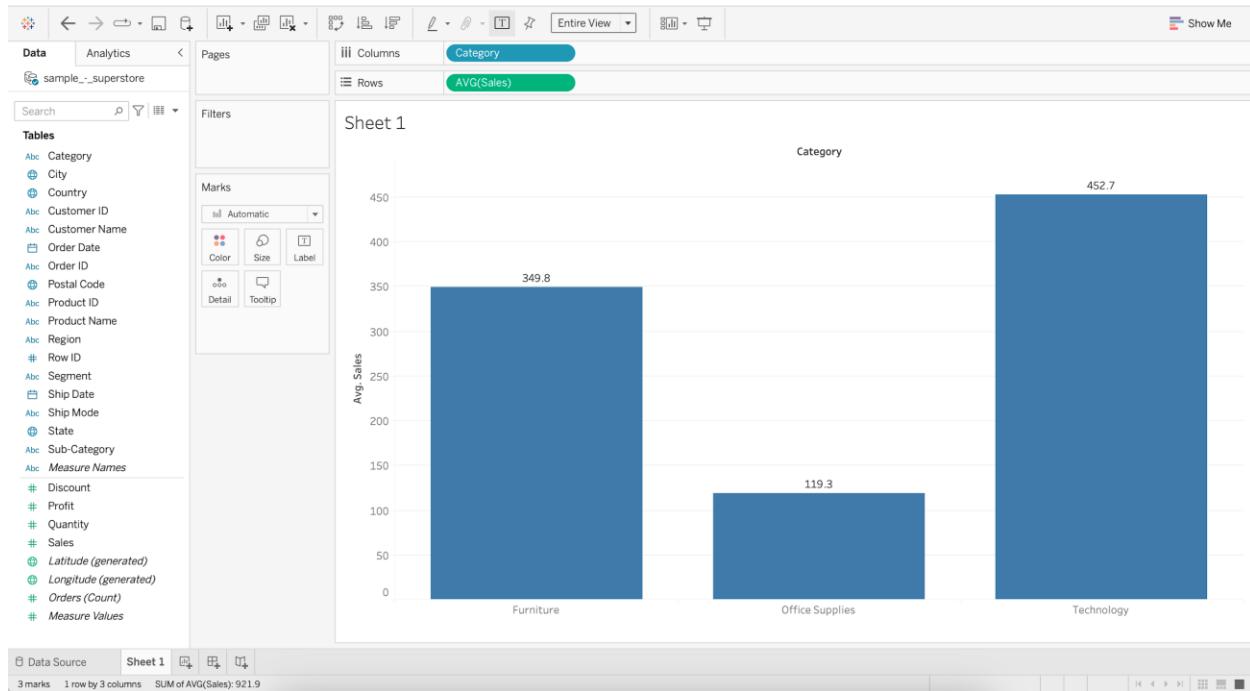
- Tableau Public only supports extract connections.
- Tableau Desktop (professional) supports both live and extract connections.

Types of Aggregation on measures and dimensions :

- We can add aggregation to dimension and measures in Tableau.
- [Types of aggregation for measures](#)
- We can convert a dimension to a measure and perform aggregation on it.
- [Types of aggregation for dimension](#)
- If you want to remove the default aggregation for fields in the view than click on analysis and clear the aggregate measure option.

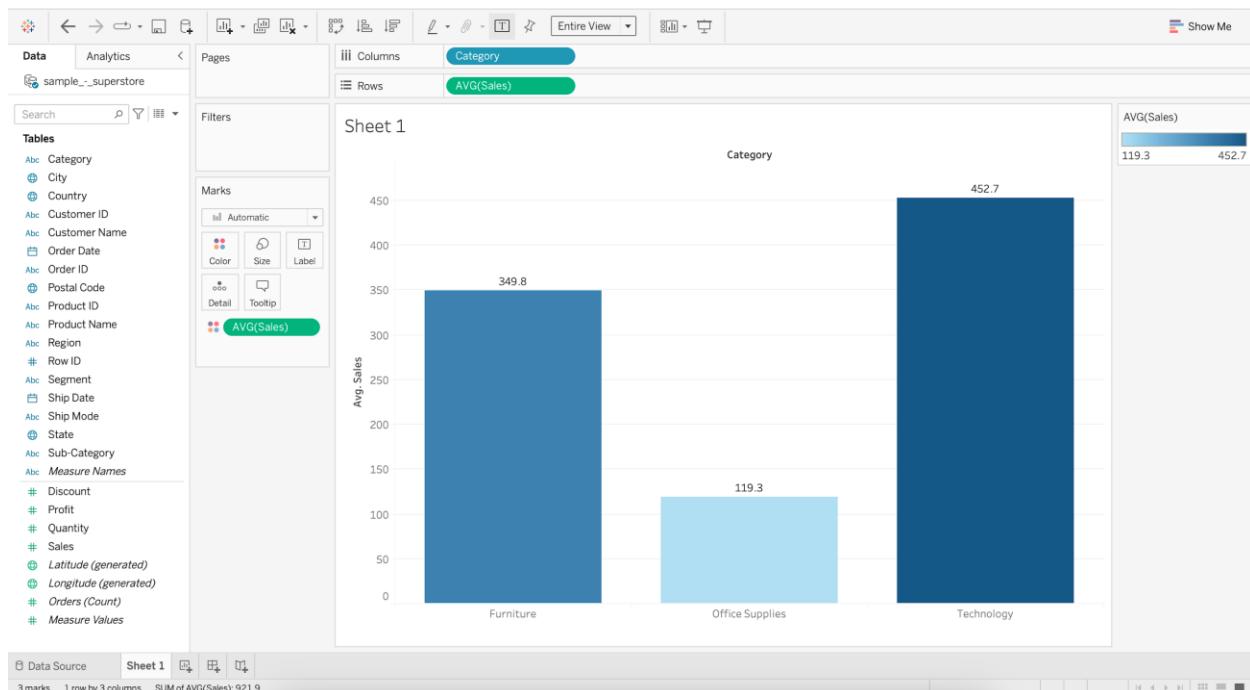
Business problem 1:

- Determine the average sales value of each category.
 - Drag Category to Columns
 - Drag Sales to Rows
 - Change aggregation of Sales from Sum to Average

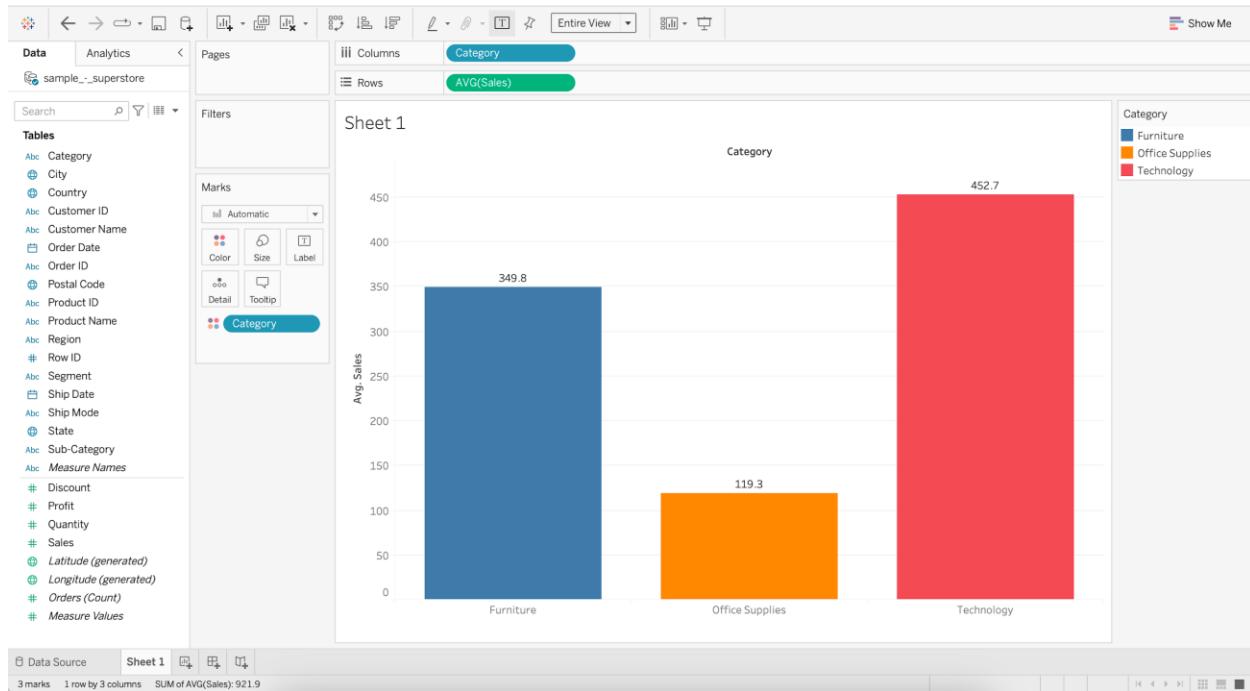


We can demonstrate color marks card shelf.

- If we drag Sales onto the color marks card shelf it'll assign a single color with different shades of that color.

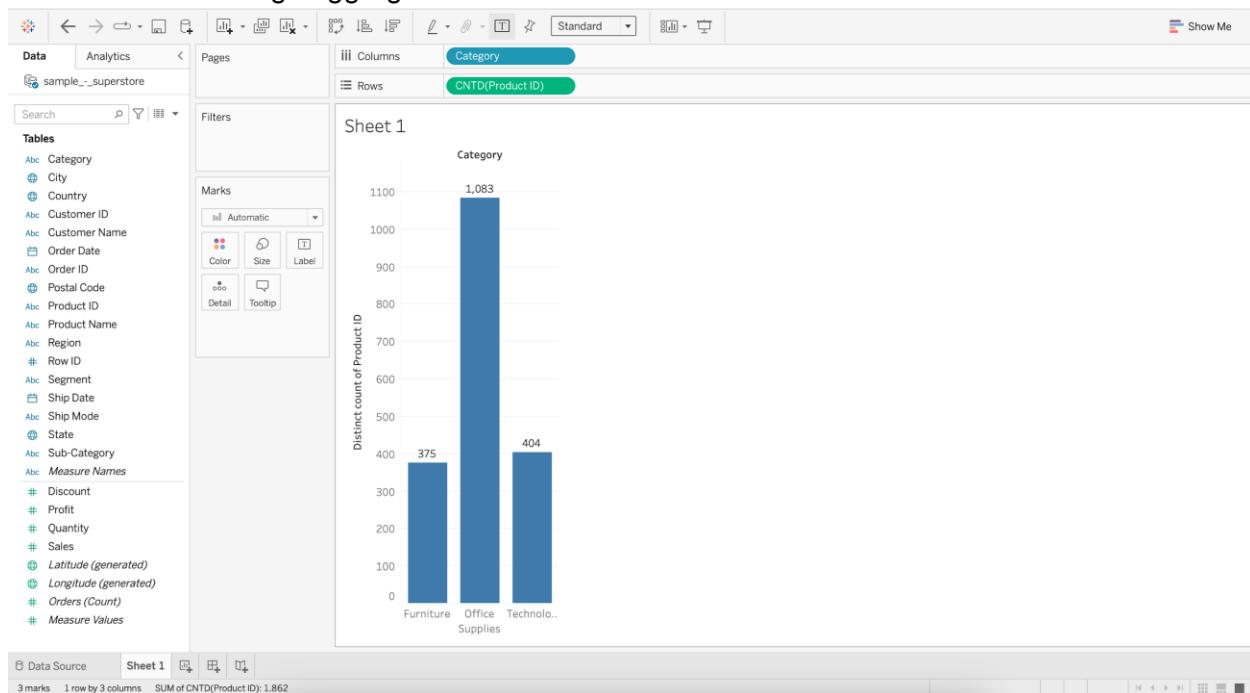


- If we drag the Category to the color marks card shelf we'll get different color for each category.



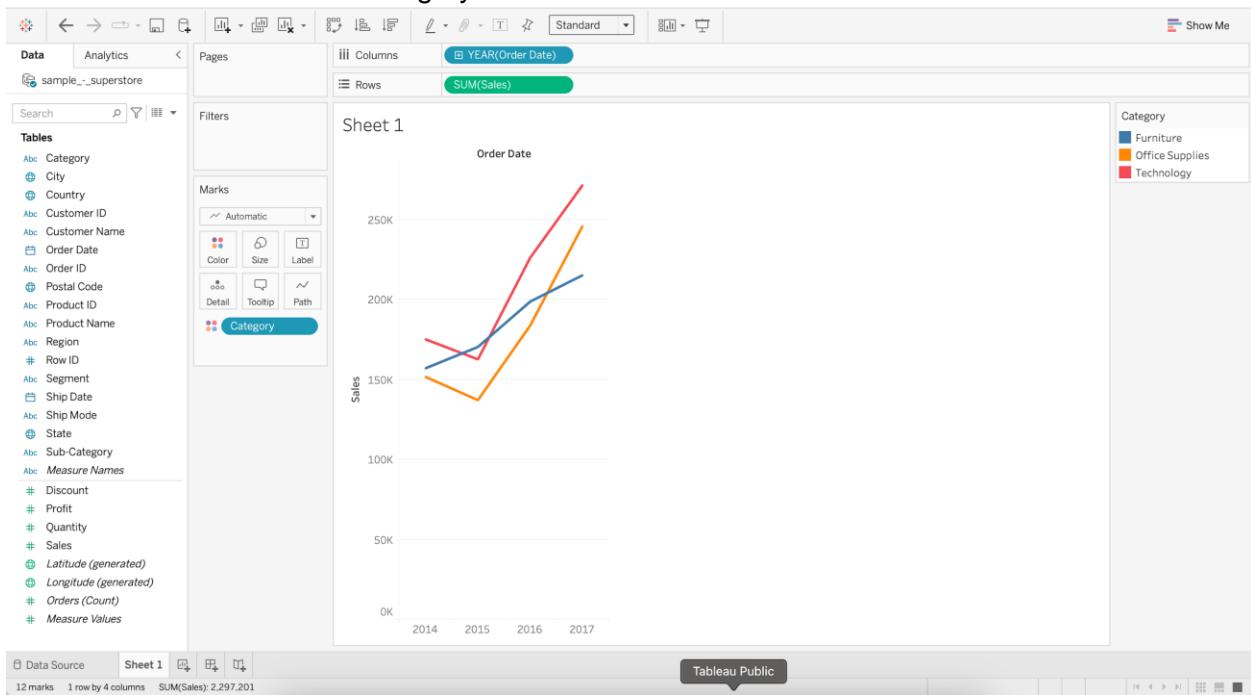
Business problem 2:

- Determine total number of products within each Category.
 - Drag Category to Columns
 - Drag Product ID to Rows
 - Change aggregation of Product ID to Count Distinct



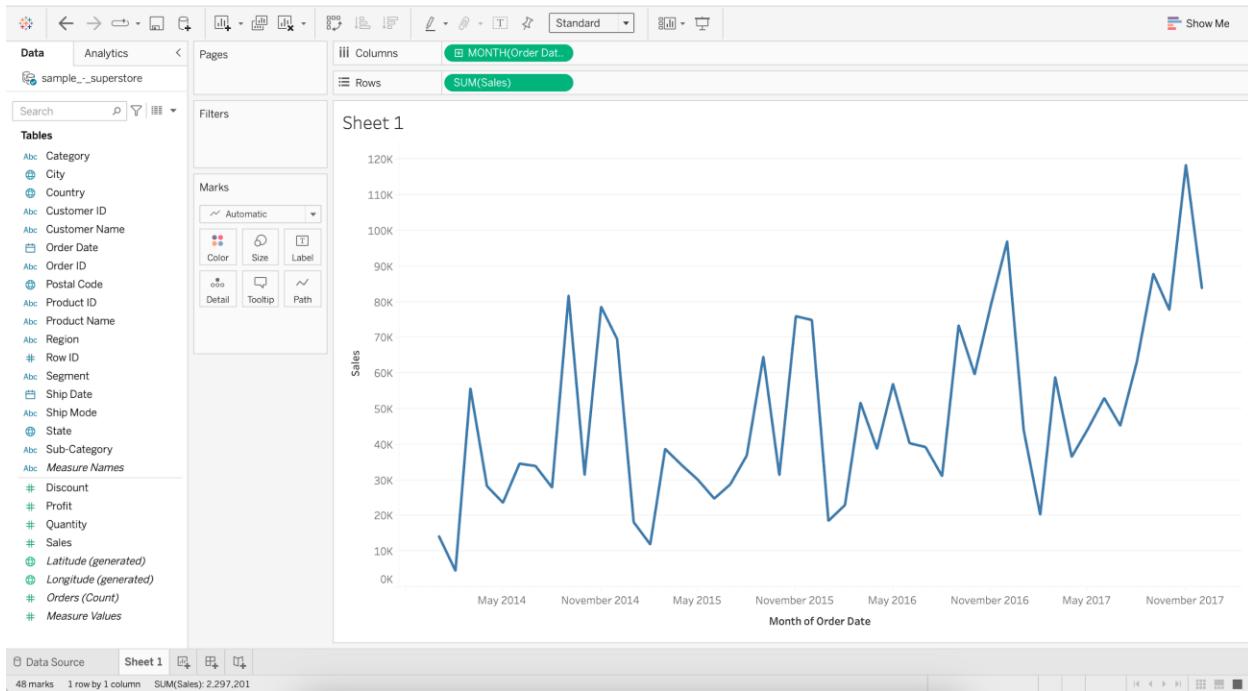
Line chart

- They connect individual data points in a view. They provide a simple way to visualize a sequence of values and are useful when you want to see trends over time.
- **Discrete line chart**
 - **Business problem 3** : Find product category that has highest sales for most of the years.
 - Drag Order Date to Columns
 - Drag Sales to Rows
 - Add Category to Color



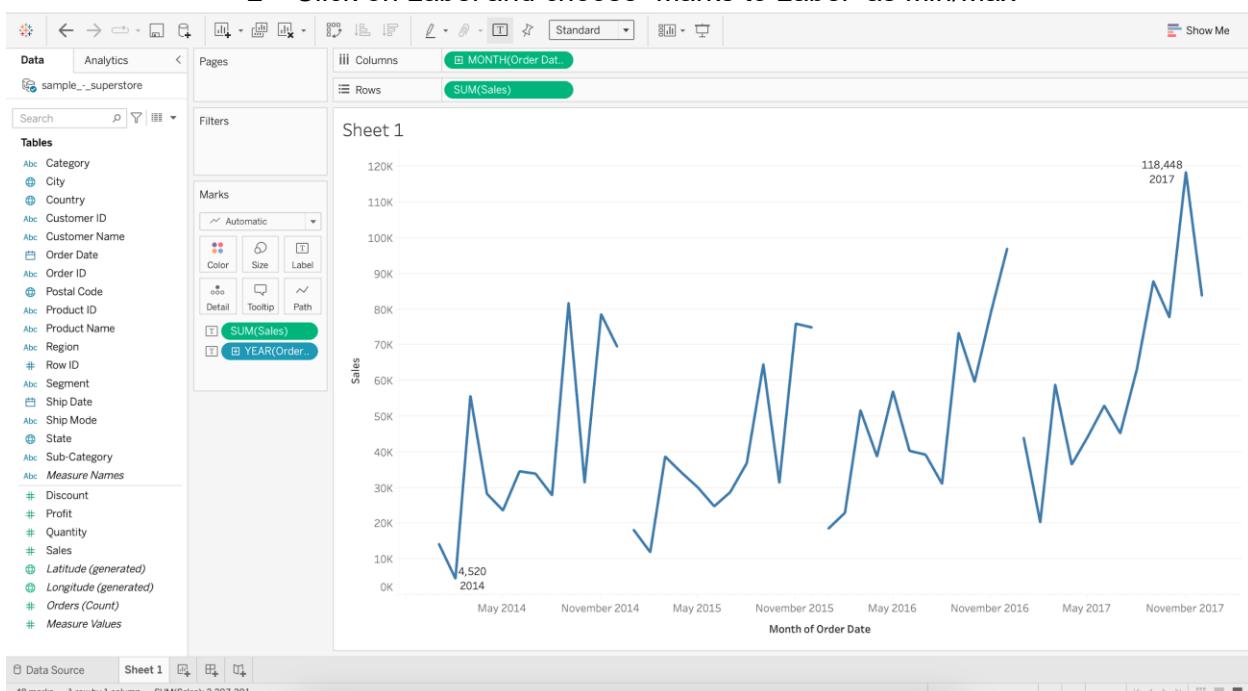
• Continuous line chart

- **Business problem 4** : Find the year and month that had the highest and lowest sales.
 - Drag Order Date to Columns
 - Drag Sales to Rows
 - Switch from YEAR (E.g. 2015) to MONTH (E.g. May 2015)



- Annotating the Line chart with min and max Sales.

- Add Sales to Label
- Add Order Date to Label
- Click on Label and choose “Marks to Label” as Min/Max

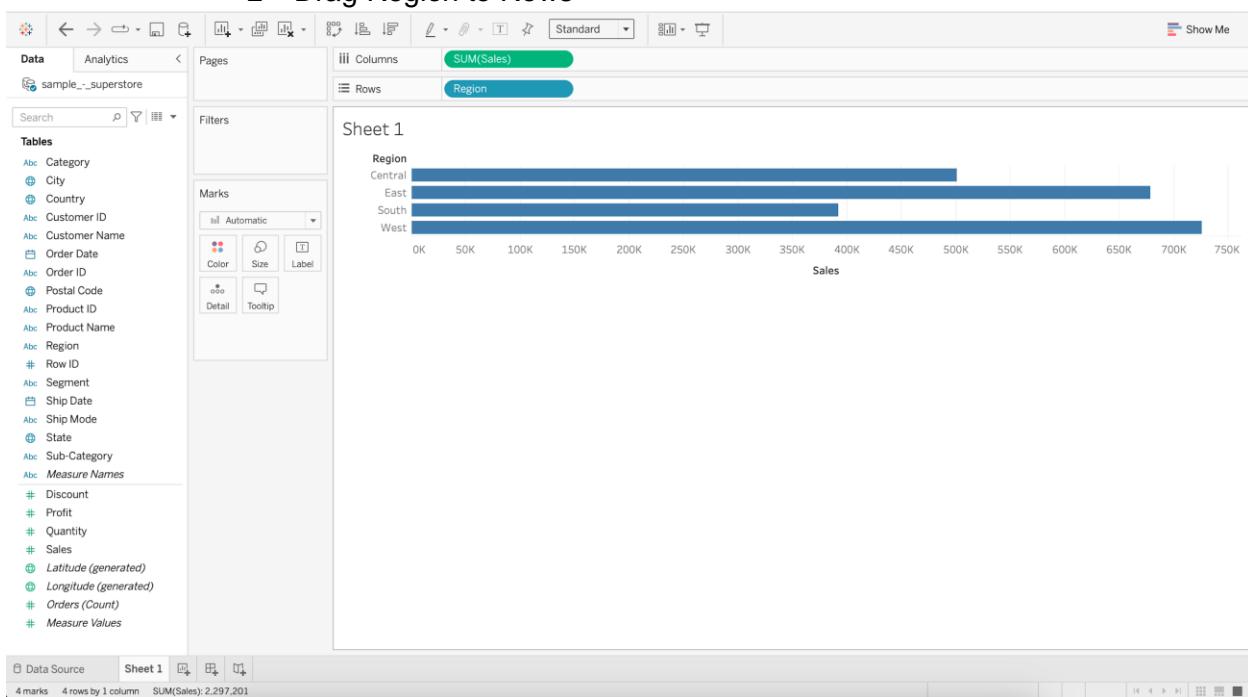


Bar chart

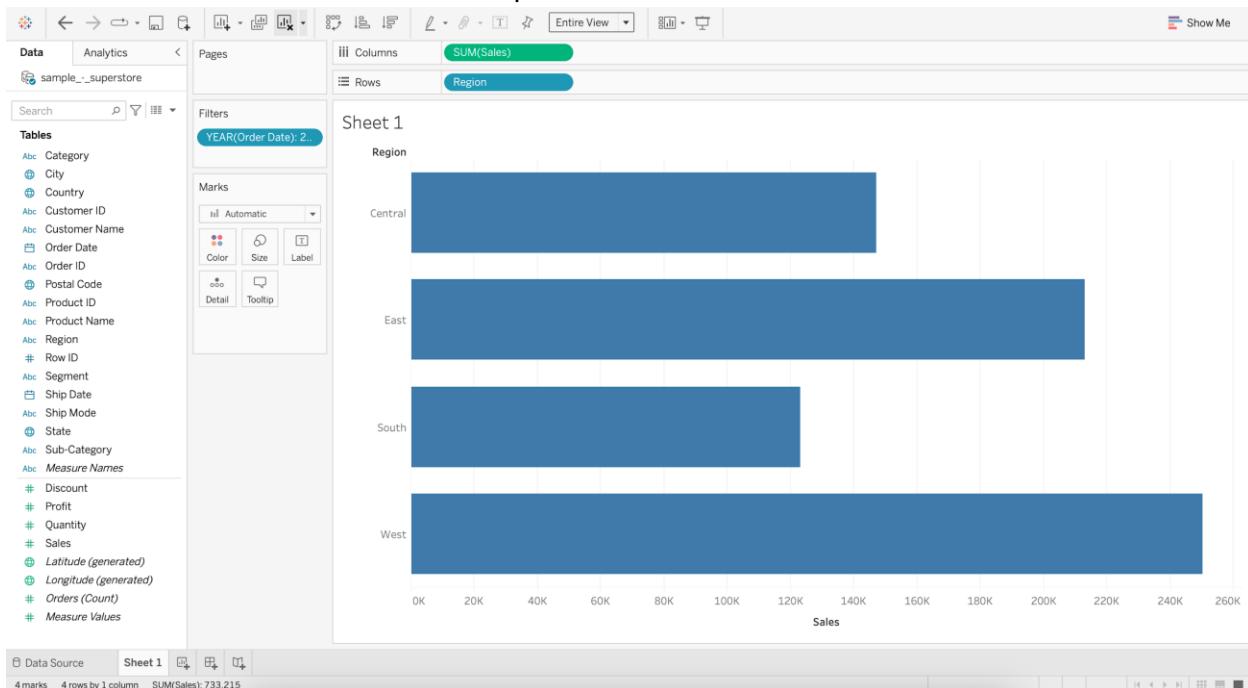
- It is a chart with rectangular bars where lengths and heights are proportional to the value that they represent.

- **Horizontal bar chart**

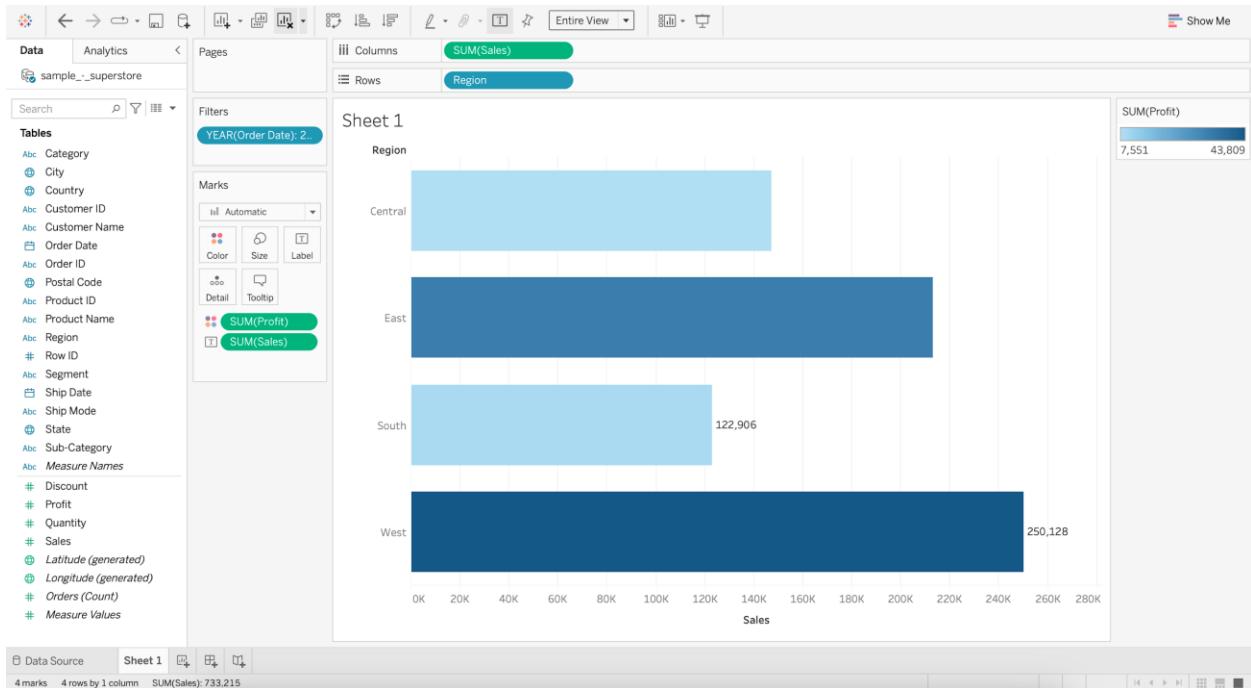
- It is a bar graph that represents data horizontally, dimensions are present on the vertical axis and data values on the horizontal axis.
- **Business problem 5 :** Find sales and profit by region for the year 2017.
 - Drag Sales to Columns
 - Drag Region to Rows



- Put Order Date into Filters
- Select Years and press next
- Select 2017 and press OK

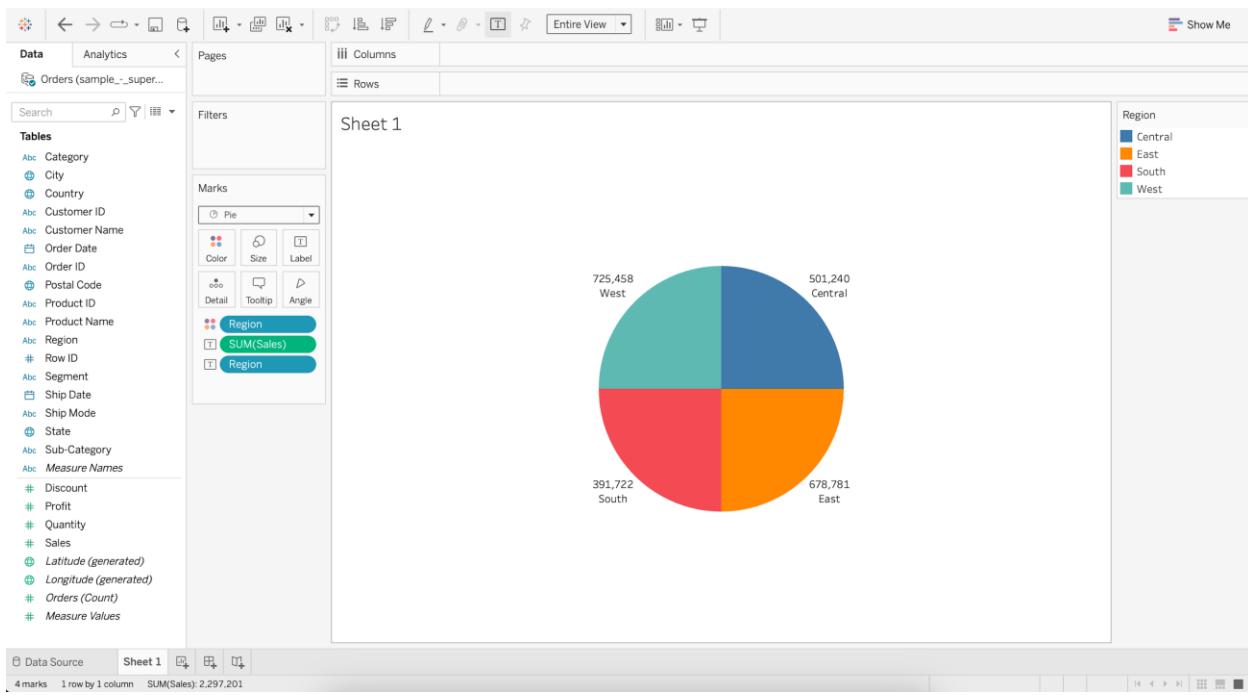


- Add Profit to Color
- Add Sales to Label

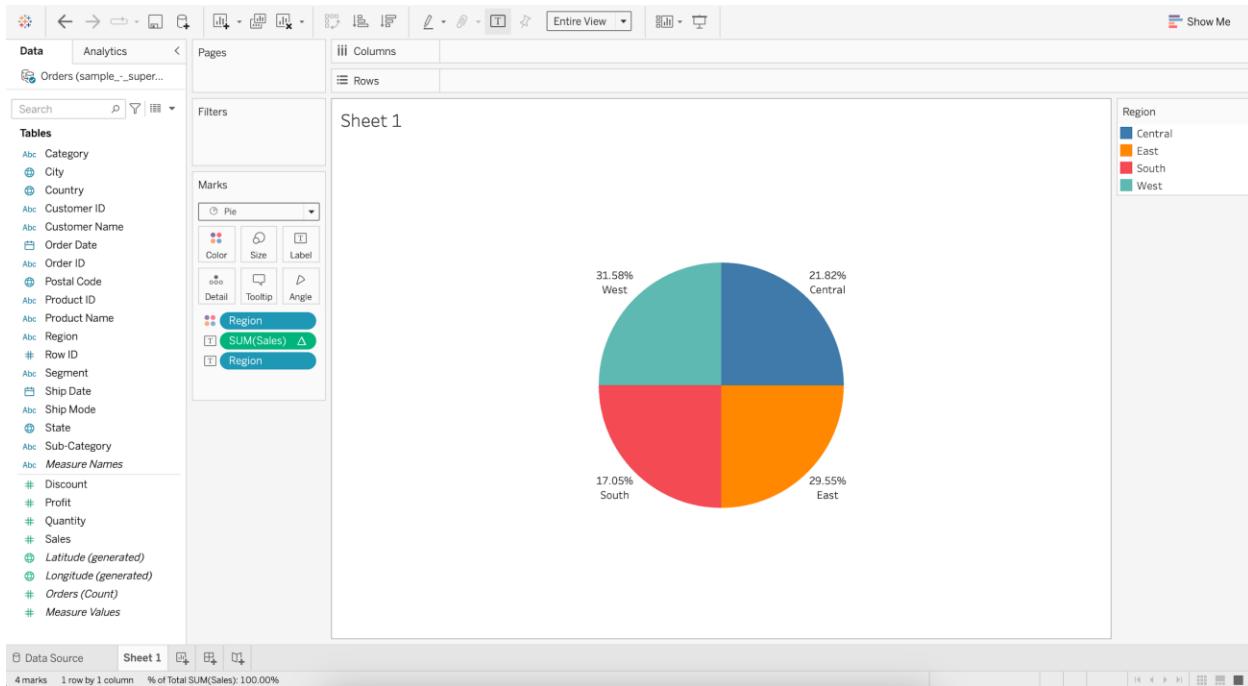


Pie chart

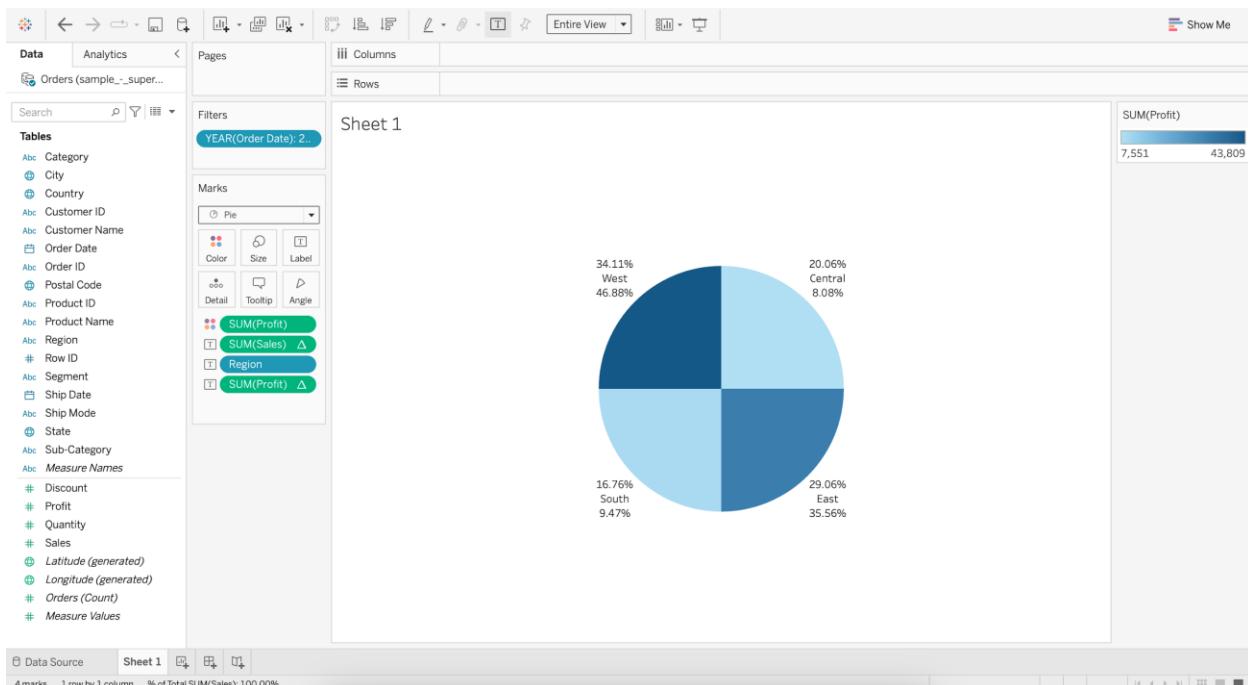
- It represents data as a slice of a circle with different sizes and colors.
 - **Business problem 6 :** Show relative percentage of sales and profit by region for the year 2017.
 - Drag Sales to the Sheet
 - Change Marks from Automatic to Pie
 - Add Region to Color
 - Add Region to Label
 - Switch to Entire View



- For SUM(Sales), go to Quick Table Calculation
- Select Percent of Total

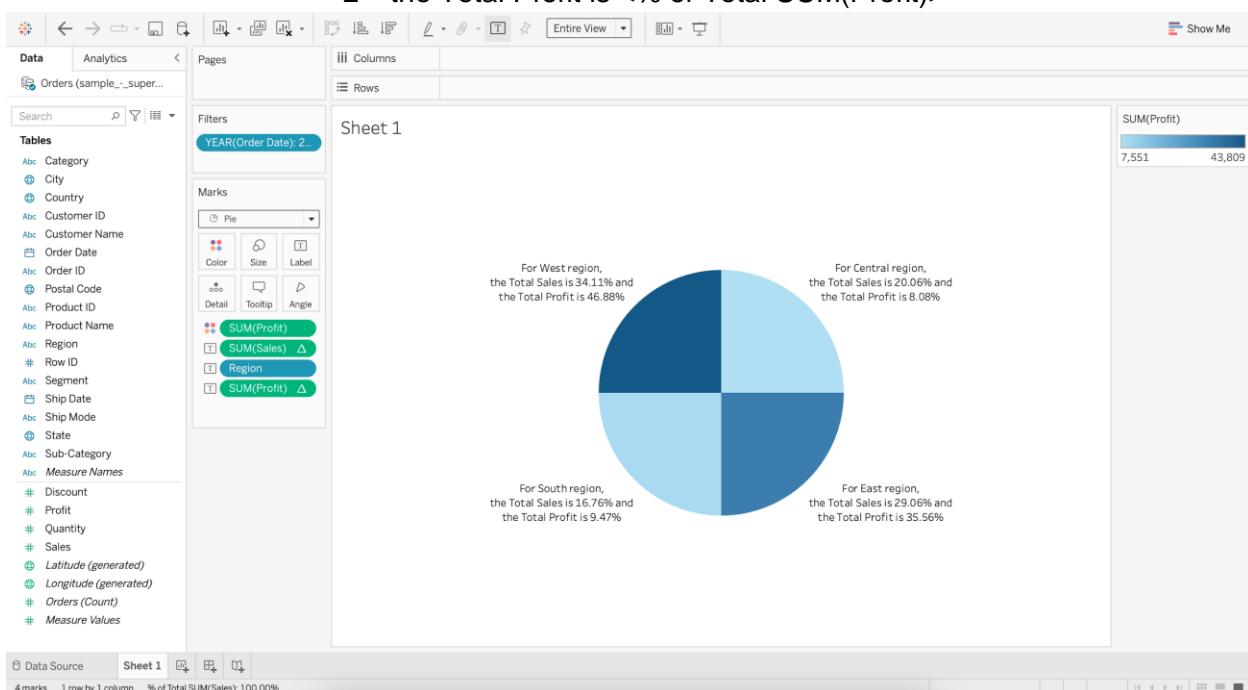


- Put Order Date into Filters
- Select Years and press next
- Select 2017 and press OK
- Add Profit to Color
- Add Profit to Label
- For SUM(Profit), go to Quick Table Calculation
- Select Percent of Total



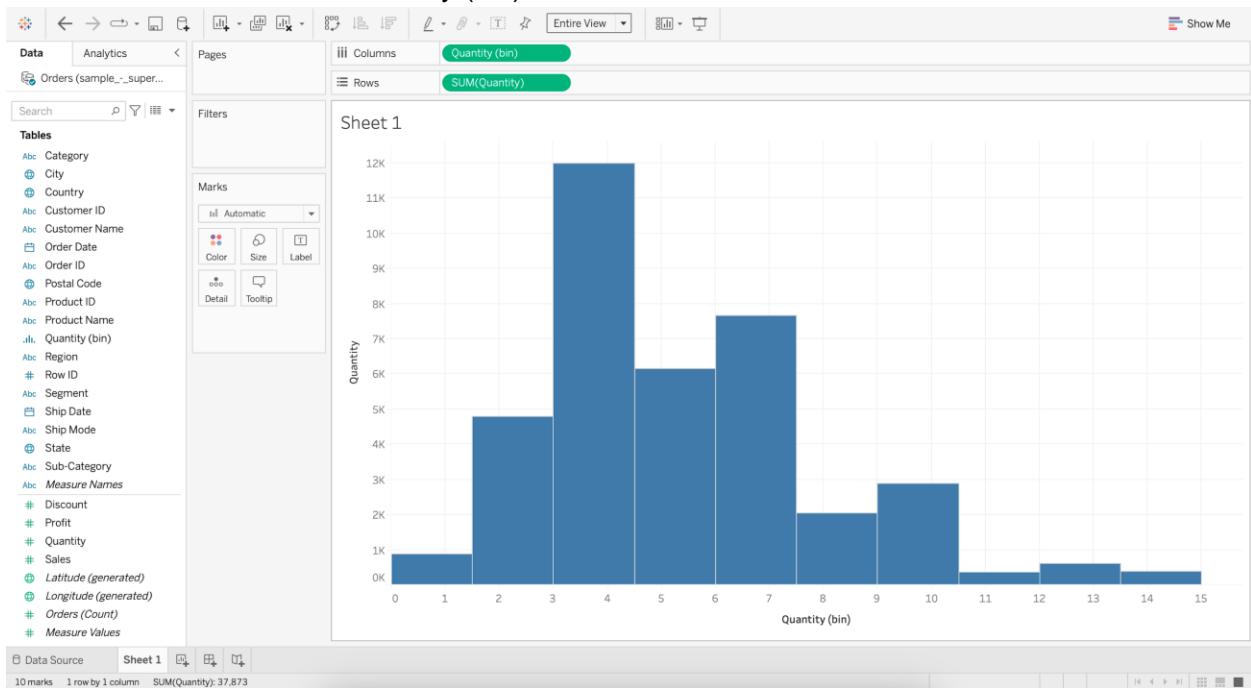
○ Annotating the Pie chart

- Go to Label and edit the Text
- Enter
 - For <Region> region,
 - the Total Sales is <% of Total SUM(Sales)> and
 - the Total Profit is <% of Total SUM(Profit)>

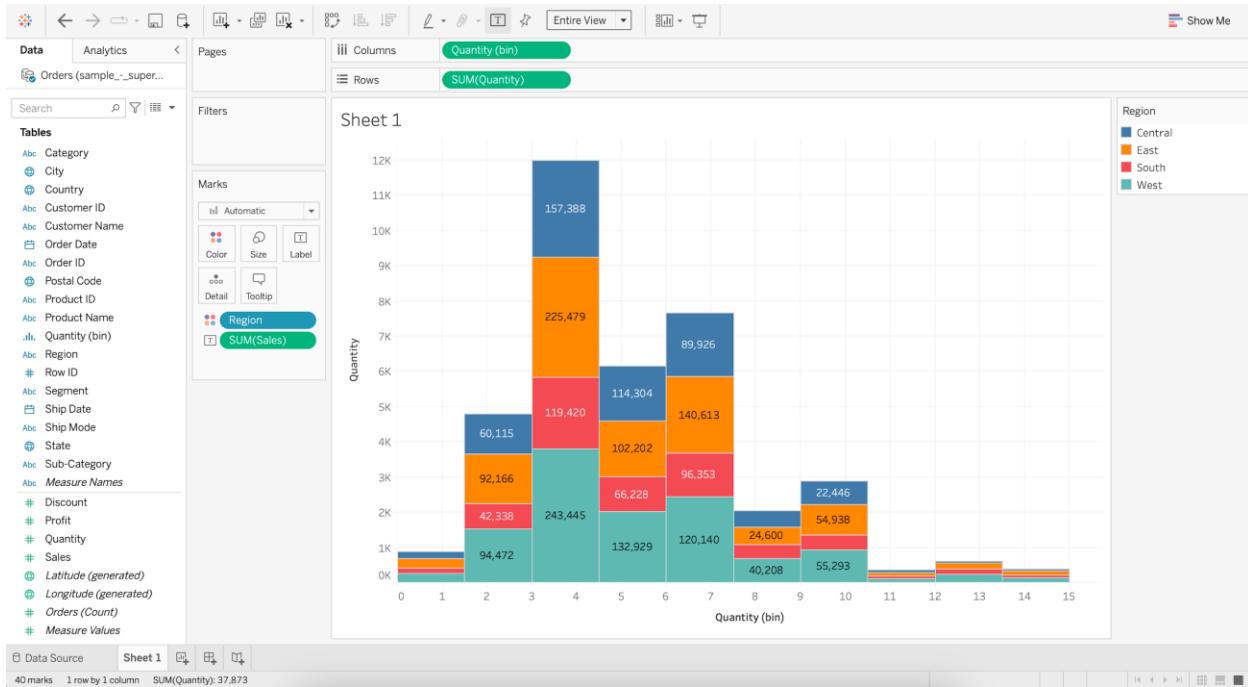


Histogram

- It is a chart that displays the shape of a distribution. A histogram looks like a bar chart but groups values for a continuous measure into ranges, or bins.
 - **Business problem 7:** Find sales by region based on the number of quantity sold (Bins).
 - For Quantity, go to Create and choose Bins
 - Change the Size of bins to 1.5
 - Drag Quantity (bin) to Columns
 - Drag Quantity to Rows
 - Set Quantity (bin) to Continuous



- Add Region to Color
- Add Sales to Label



Cleaning and Formatting data for analysis

Dataset : World_Bank_CO2.xlsx [Use CO2 (kt) RAW DATA Table]

Dataset url : [link](#)

- **Business problem 8:**

- Find the CO2 emissions per country per year for those countries which have CO2 emissions above 2000 kt.
 - Load the World_Bank_CO2.xlsx dataset
 - Drag the CO2 (kt) RAW DATA table
 - Check the Use Data Interpreter option

■ Select the columns 1960 to 2011 and Pivot

The screenshot shows the Power BI Data Editor interface. On the left, the 'Connections' pane is open, showing a connection to 'World_Bank_CO2 Microsoft Excel'. The 'Sheets' pane lists various data transformations: 'Cleaned with Data Interpreter' (checked), 'CO2 (kt) for Split', 'CO2 (kt) Pivoted', 'CO2 (kt) RAW DATA', 'CO2 Data Cleaned', 'CO2 Per Capita (Pivoted)', 'CO2 Per Capita RAW DATA', 'Metadata - Countries', 'New CO2(KT) Pivoted', and 'Sheet1'. A 'New Union' and 'New Table Extension' option are also present.

The main area displays the 'CO2 (kt) RAW DATA' table with 56 fields and 248 rows. The table has columns: Name, Fields, Type, Field Name, Physical Table, and Remote Field N... . The data includes rows for Aruba, Andorra, Afghanistan, Angola, Albania, and Arab World, with various CO2 emissions values.

On the right, there are filter controls for 'Filters' (0 | Add) and a preview of the pivoted data with columns 1960 to 1961.

■ Rename the PIVOT FIELD NAMES column as Year

■ Rename the PIVOT FIELD VALUES column as CO2 (kt)

This screenshot shows the same Power BI Data Editor interface after the pivot operation. The 'Sheets' pane remains the same. The main area now shows the pivoted 'CO2 (kt) RAW DATA' table with columns: Name, Fields, Type, Field Name, Physical Table, and Remote Field N... . The data is grouped by year (2006, 2007, 2008, 2009) and contains specific CO2 values for each country.

On the right, the 'Rename' context menu is open over the first column, showing options like 'Rename', 'Copy Values', and 'Hide'. The 'Rename' option is highlighted.

The screenshot shows the Power BI Data Interpreter interface. On the left, the 'Connections' pane lists 'World_Bank_CO2 Microsoft Excel'. The 'Sheets' pane shows various tables: 'CO2 (kt) for Split', 'CO2 (kt) Pivoted', 'CO2 (kt) RAW DATA', 'CO2 Data Cleaned', 'CO2 Per Capita (Pivoted)', 'CO2 Per Capita RAW DATA', 'Metadata - Countries', 'New CO2(KT) Pivoted', 'Sheet1', 'New Union', and 'New Table Extension'. The main area displays the 'CO2 (kt) RAW DATA' table with 6 fields and 12896 rows. The table has columns: Country Code, Indicator Name, Indicator Code, Year, and CO2 (kt). The data shows CO2 emissions for ABW from 1960 to 1965, with values like 'null', 'null', 'null', 'null', and 'null'. A 'Fields' section below the table lists columns: Type, Field Name, Physical Table, and Remote Field N... with entries for Country Name, Country Code, and Indicator Name.

- Go the Filters in the upper right corner and click Add
- Again click Add and Select a field as CO2 (kt)
- Select the minimum Range of values as 2000

This screenshot shows the same Power BI Data Interpreter interface as the previous one, but with filters applied. In the top right corner, there is a 'Filters' section with a count of 1 and a 'Edit' button. The 'CO2 (kt) RAW DATA' table now shows data for the year 2000 only, with values: 2,321.21, 2,357.88, 2,372.55, 2,416.55, 2,420.22, and 2,497.23.

Data Interpreter

- When you track data in Excel spreadsheets, you create them with the human interface in mind.

- To make your spreadsheets easy to read, you might include things like titles, stacked headers, notes, maybe empty rows and columns to add white space, and you probably have multiple tabs of data too.
- When you want to analyze this data in Tableau, these aesthetically pleasing attributes make it very difficult for Tableau to interpret your data.

That's where a Data Interpreter can help.

What does a Data Interpreter do?

- It can give you a head start when cleaning your data by detecting things like titles, notes, footers, empty cells, etc.
- And bypassing them to identify the actual fields and values in your data set.

Note:

- Data Interpreter is only available for Microsoft Excel, Text (.csv) files, PDF files and Google Sheets.
- For Excel, your data must be in the .xls or .xlsx format.

For more information: https://help.tableau.com/current/pro/desktop/en-us/data_interpreter.htm

Pivoting Data

- Sometimes analyzing data from a spreadsheet or crosstab format can be difficult in Tableau.
- Tableau prefers data to be “tall” instead of “wide”, which means that you often have to pivot your data from columns to rows so that Tableau can evaluate it properly.

Note: Pivoting from rows to columns is not available in Tableau Public.

For more information: https://help.tableau.com/current/prep/en-us/prep_pivot.htm

Data Source Filter

- The data source filter is used to filter the data in data source proportion.
- It restricts the files present in the dataset.
- A data source filter works on both Live connection and Extract connection.

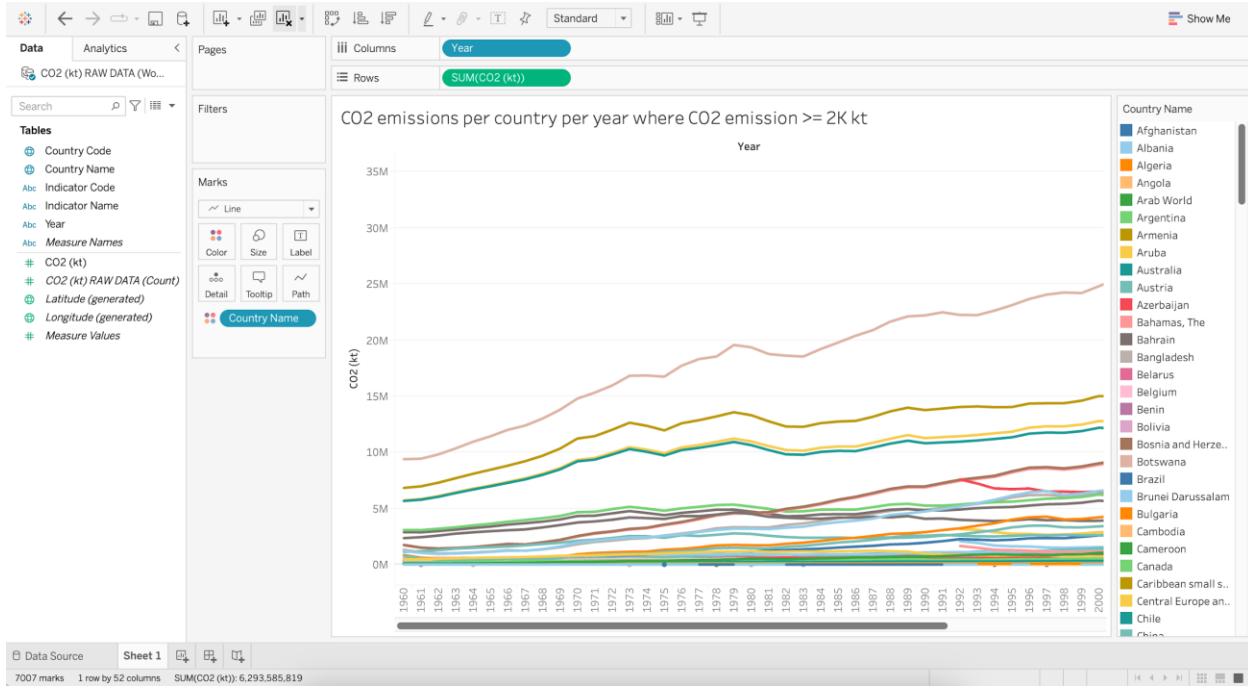
Filter shelf

We can use a filter shelf to filter data in the view.

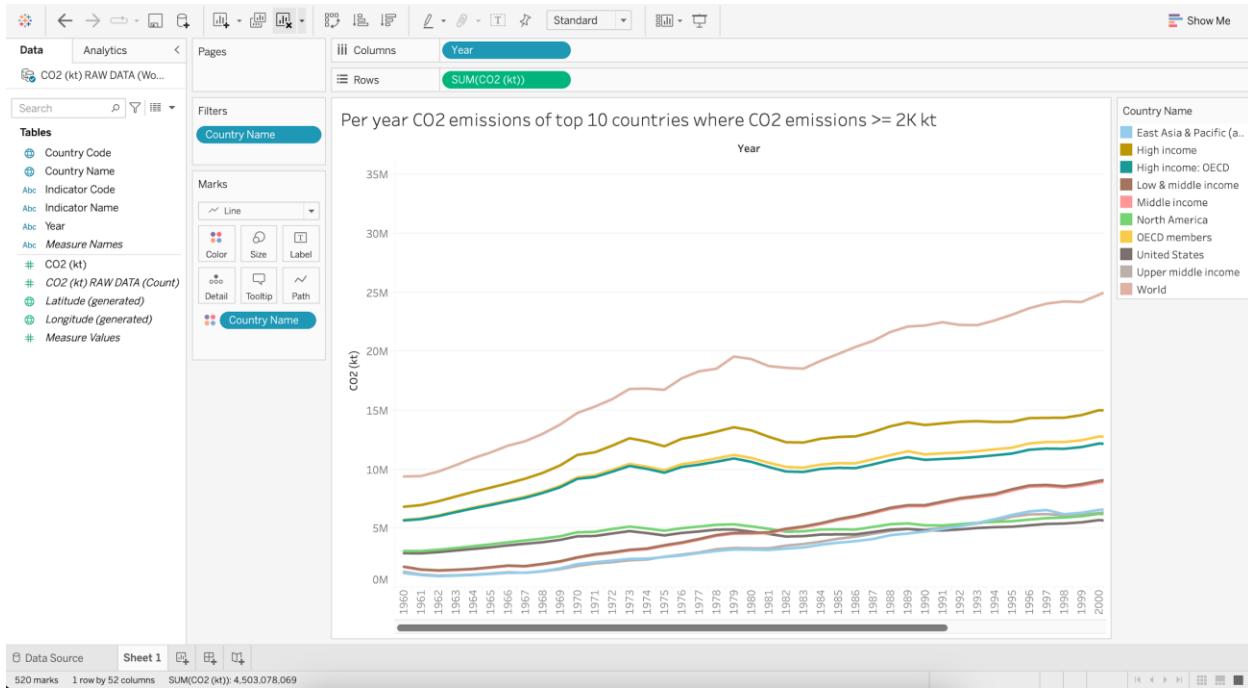
Dataset : World_Bank_CO2.xlsx [Use CO2 (kt) RAW DATA Table]

- **Business problem 9:**
 - Find the CO2 emissions per country per year where CO2 emission $\geq 2K\ kt$
 - Drag Year to Columns
 - Drag CO2 (kt) to Rows
 - Change Marks from Automatic to Line

- Add Country Name to Color
- Edit Title as “CO2 emissions per country per year where CO2 emission $\geq 2K\text{ kt}$ ”



- Find the per year CO2 emissions of top 10 countries where CO2 emissions $\geq 2K\text{ kt}$.
 - Put Country Name into Filters
 - Go to Top and Select Top 10 Fields by Sum of CO2 (kt)
 - Edit Title as “Per year CO2 emissions of top 10 countries where CO2 emissions $\geq 2K\text{ kt}$ ”



Creating & Uploading Dashboard

- Consider this as an extra activity.
 - Create and Upload a basic dashboard based on the worksheets created above (Business problem 1 to 7).
-

Tableau Lecture 3: Data Structuring options

Agenda

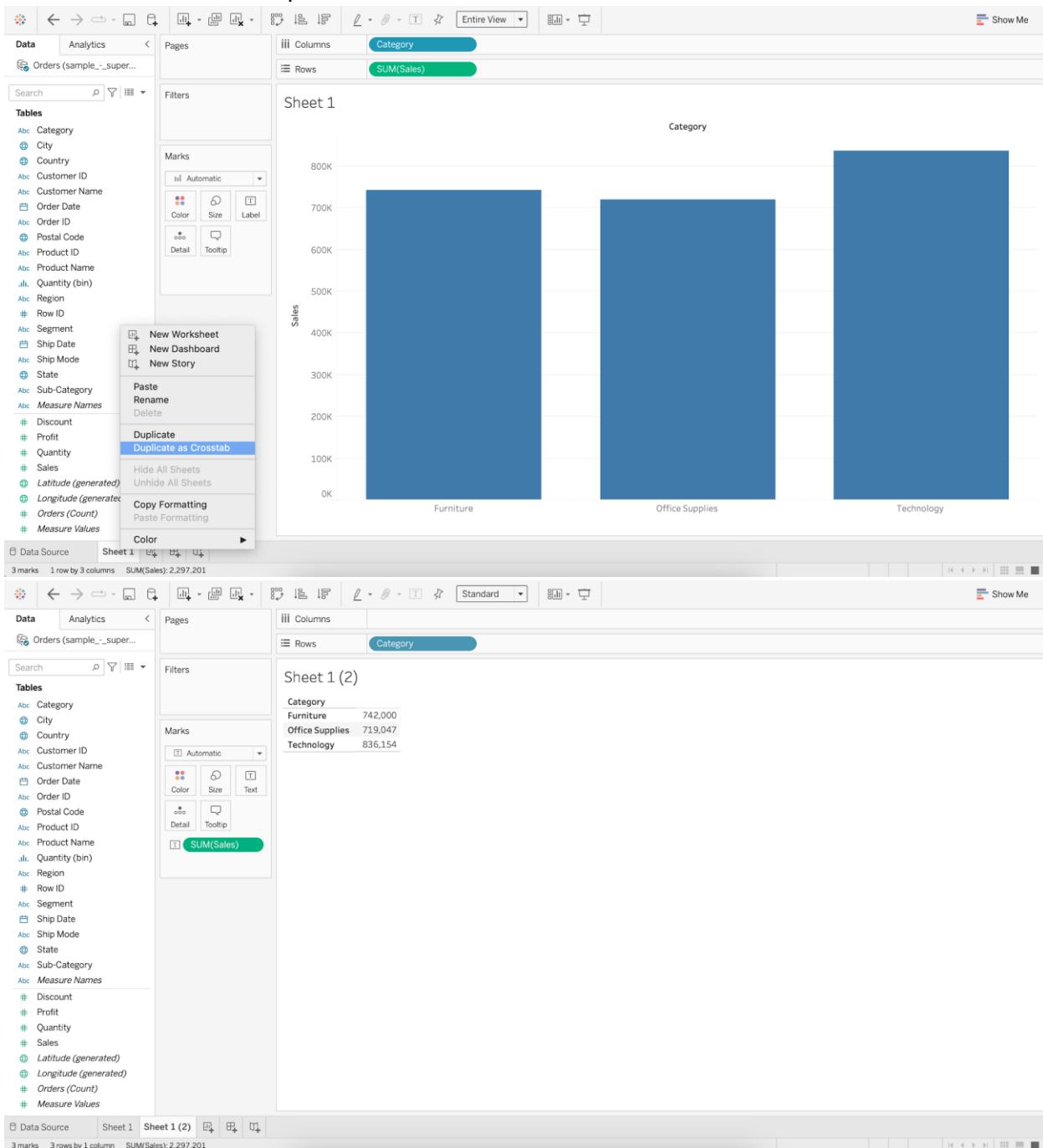
- Basic Charts
 - Text Table
 - Highlight Table
 - Heat Map
- Data Structuring
 - Groups
 - Sets
 - Group vs Set
 - Parameters

Dataset : [sample superstore](#)

Text Table

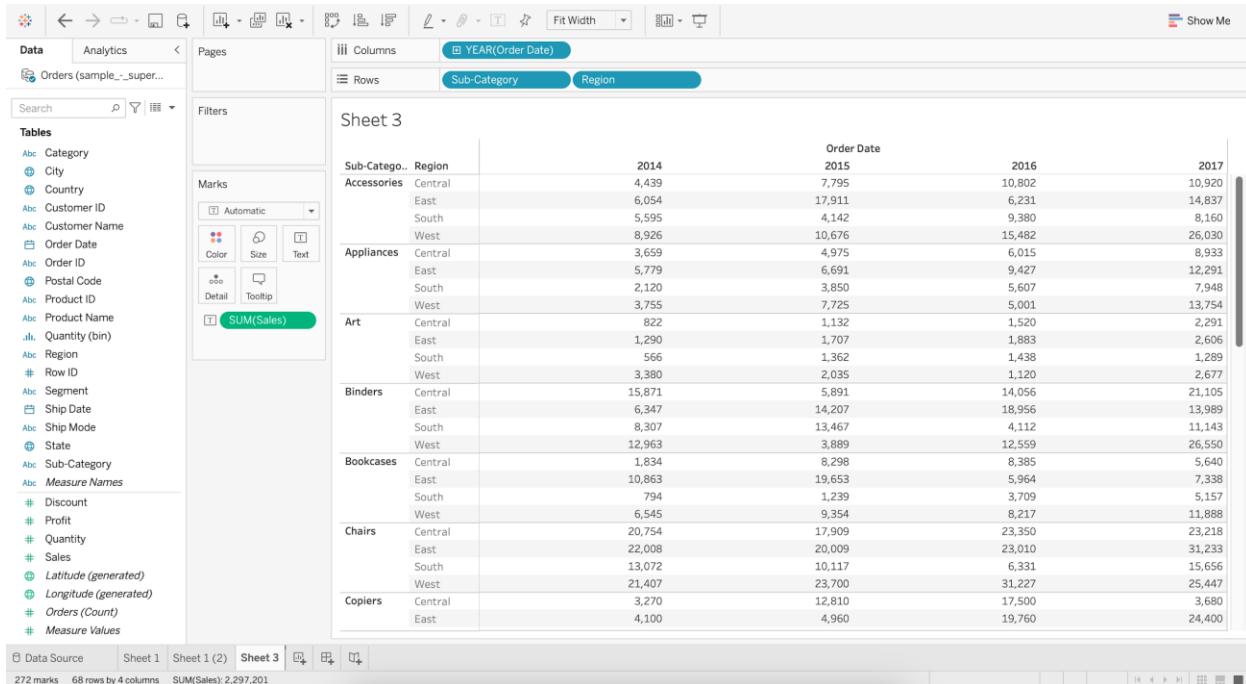
- Also known as **CrossTab** or **Pivot Table**.

- They are useful in below cases :
 - When the end user is more comfortable in reading data in tabular form.
 - To export Raw Data - any view in a worksheet can be duplicated as text table.
 - To create a filter action in dashboards.
- Duplicating a view as Text Table
 - Drag Category to Columns
 - Drag Sales to Rows
 - Right Click on the Sheet
 - Choose Duplicate as Crosstab



Business problem 1 :

- Display sales of for each subcategory by region for all the years in a table format.
 - Drag Order Date to Columns
 - Drag Sub-Category to Rows
 - Drag Region to Rows
 - Add Sales to Label
 - Switch to Fit Width

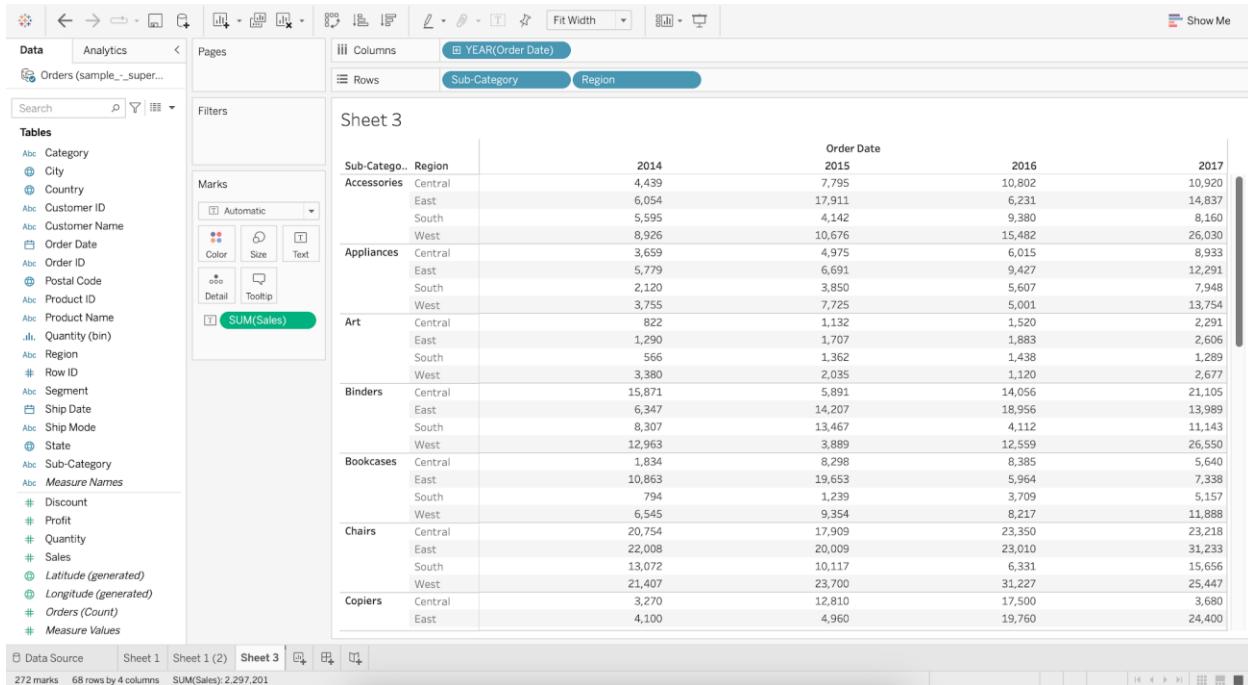


Highlight Table

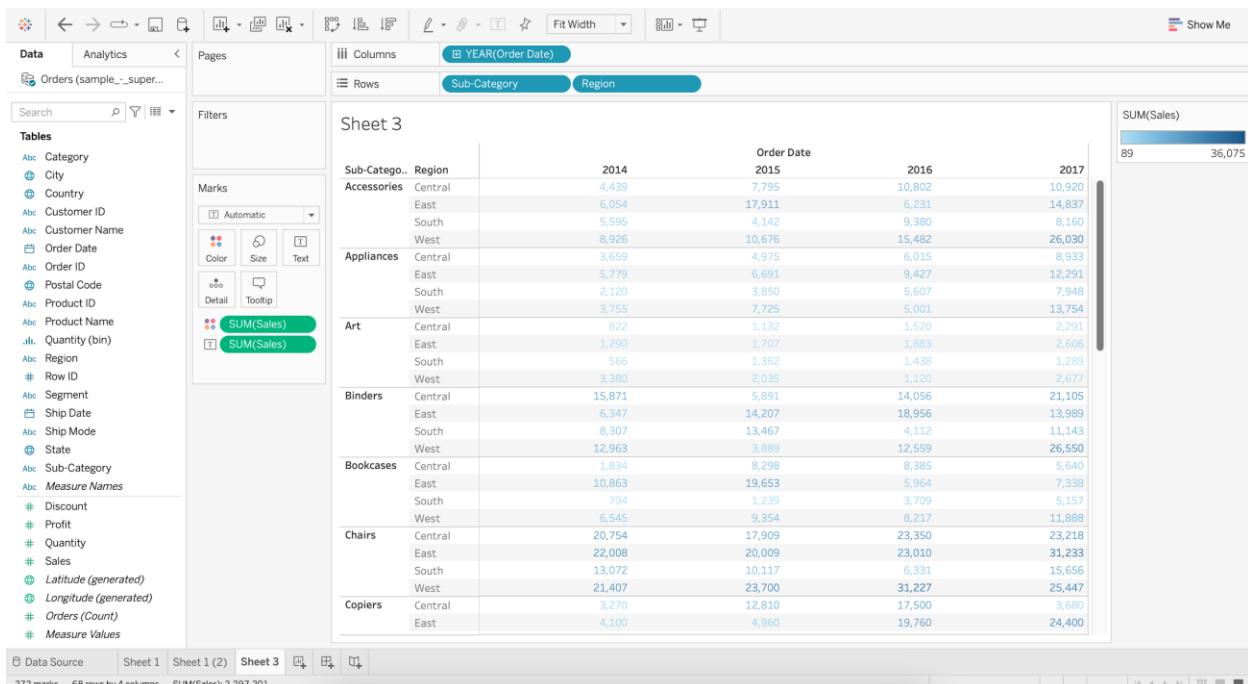
- It is used to enhance the text table and provide better insights in a glance.
- It requires 1 or more dimension on the Column shelf, 1 or more dimension on the Row shelf and 1 measure in the Color marks shelf.

Business problem 2 :

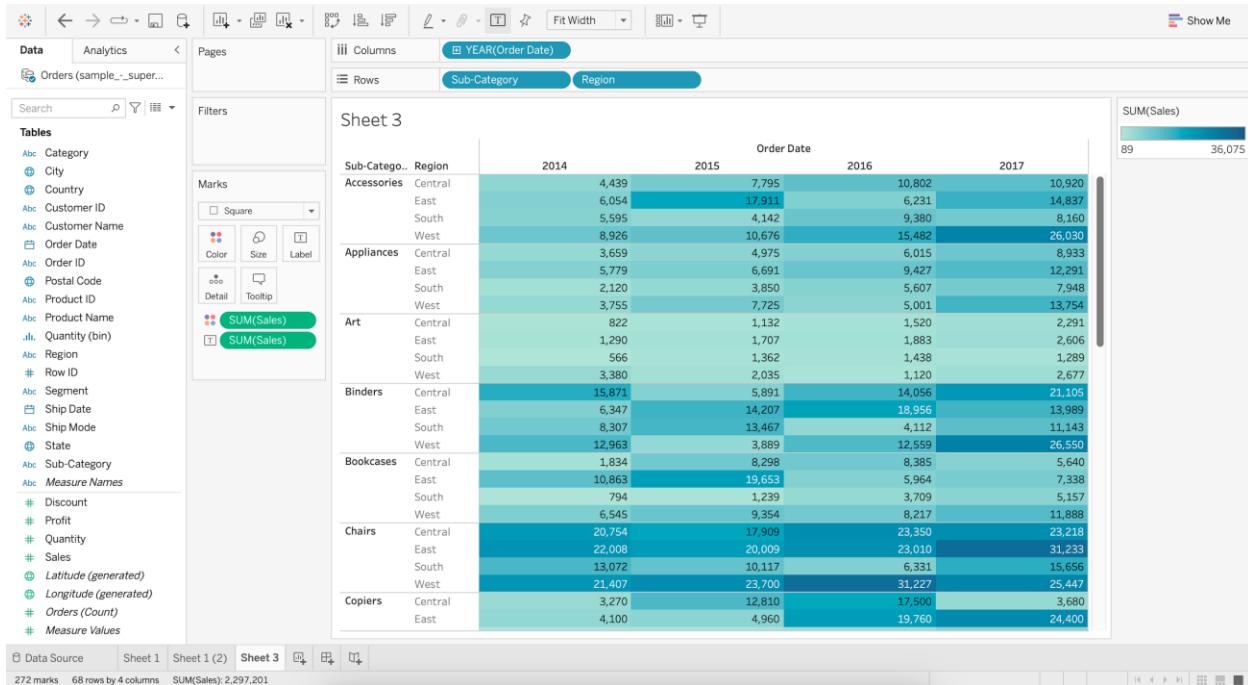
- Display sales of for each subcategory by region for all the years in a table format and highlight values that have the highest and lowest sales.
 - Create the Text Table
 - Drag Order Date to Columns
 - Drag Sub-Category to Rows
 - Drag Region to Rows
 - Add Sales to Label
 - Switch to Fit Width



Add Sales to Color



Change Marks from Automatic to Square



Heat Map

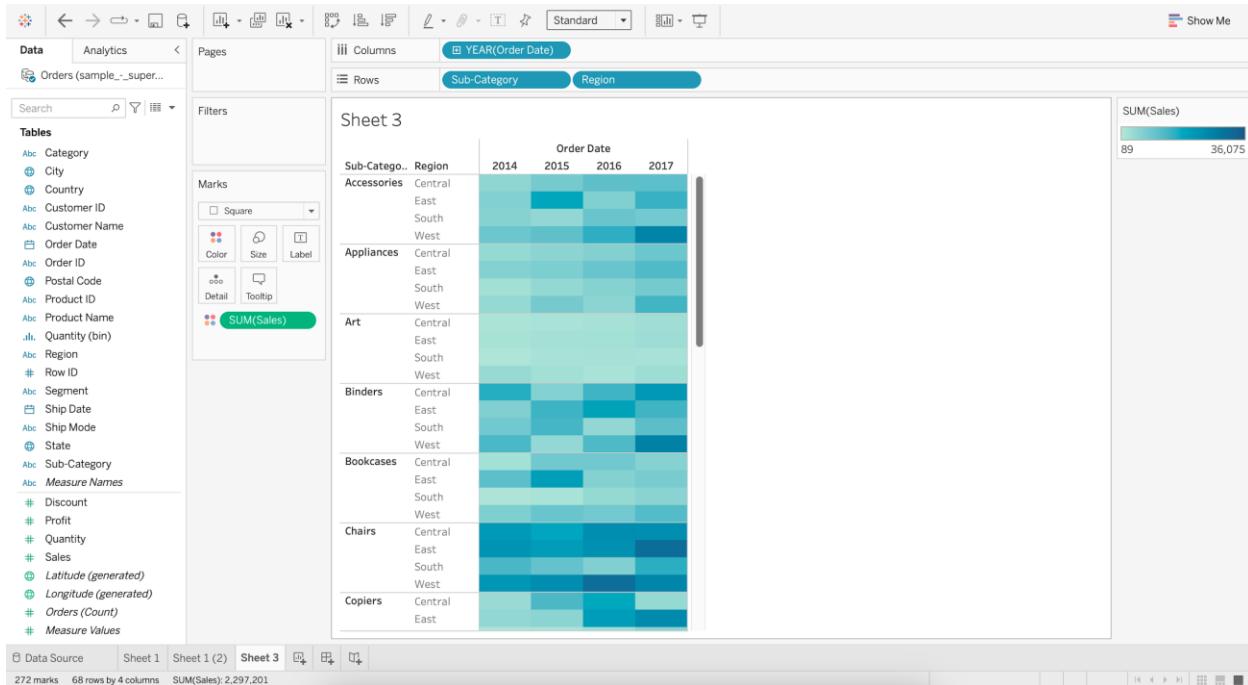
- It requires 1 or more dimensions and 1 or 2 measures.

Difference between Heat Map and Highlight Table

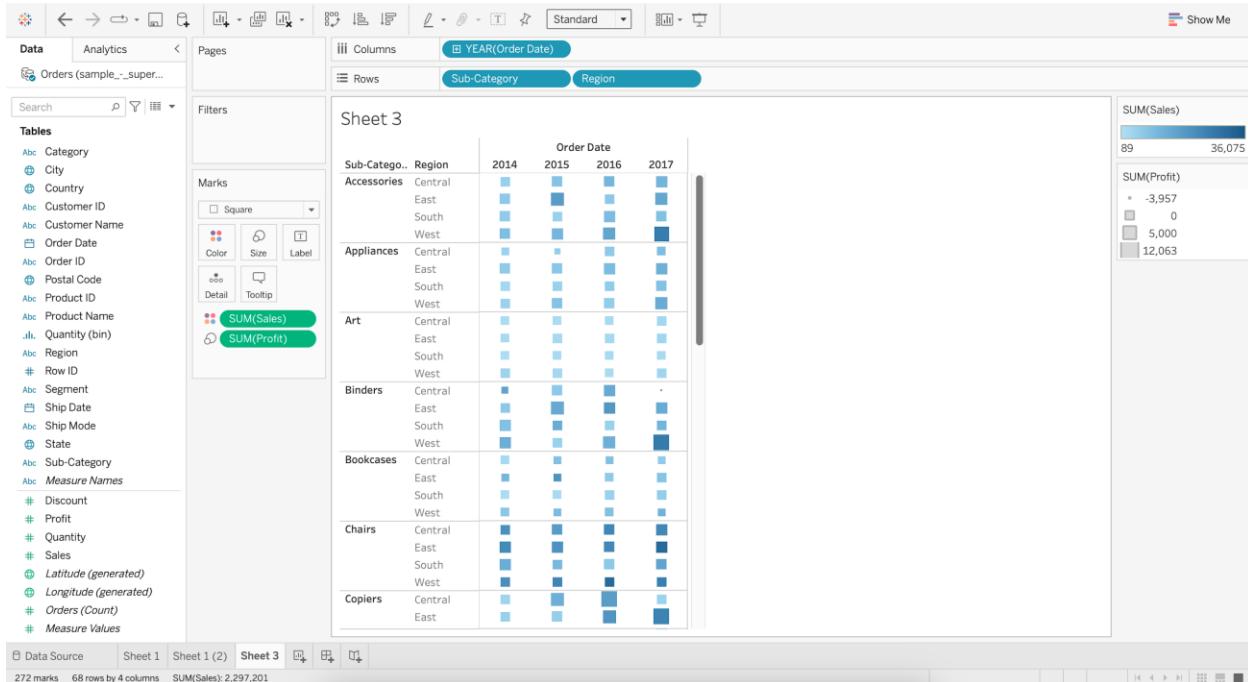
- In the highlight table our only option is to Color the marks using the measures.
- In the heat map in addition to this we can also Size the marks by 1 measure.

Business problem 3 :

- Display the sales and profit for each subcategory by region for all the years in a table format.
 - Create the Highlight Table
 - Drag Order Date to Columns
 - Drag Sub-Category to Rows
 - Drag Region to Rows
 - Add Sales to Color



- Add Profit to Size
- Increase the Size



We can easily see that the highest profit for accessories was for the west region in the year 2017.

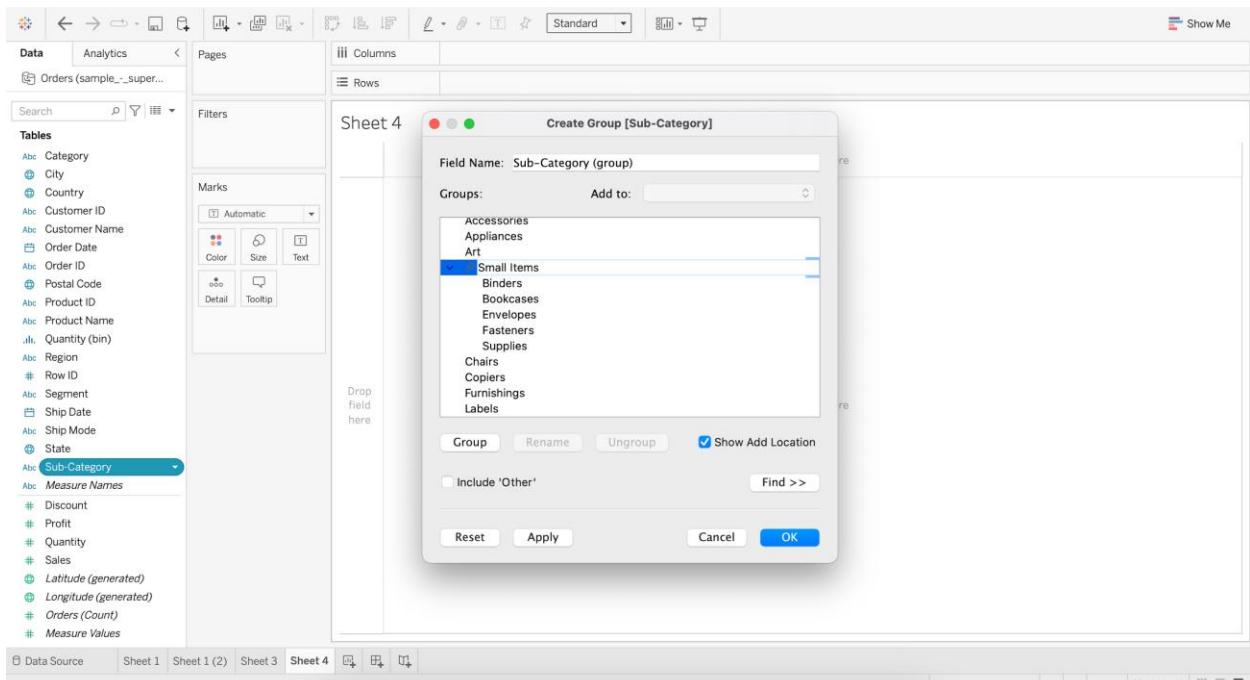
Group

- Tableau allows 1-dimensional members to be grouped together and automatically creates a new dimension with Tableau Group at the end of the name.

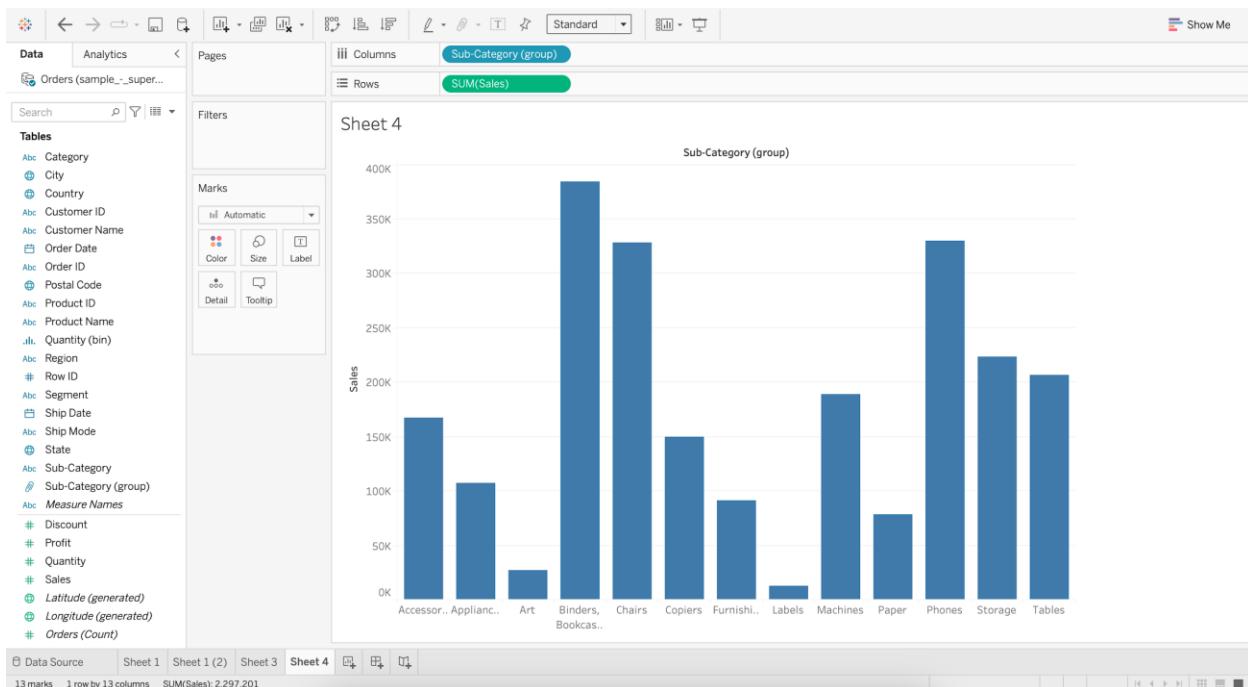
- Tableau makes no adjustments to the members' original dimensions.

Business problem 4 :

- Create a group called “small items” which includes Binder, Bookcases, Envelope, Fastener, Supplies and compare its sales with the rest of the sub-categories.
 - For Sub-Category, go to Create and choose Group
 - Select Binder, Bookcases, Envelope, Fastener, and Supplies
 - Click on Group
 - Rename the Group as “Small Items”
 - Click OK



- Drag Sub-Category (group) to Column
- Drag Sales to Rows



Set

- Sets are custom fields that are used to keep a subset of data based on a condition.
- When you drag a Set to the visualization, Tableau displays the Set using the In/Out mode.
- This is a binary categorization of Sets :
 - In-Members that are inside the Set
 - Out-Members that aren't part of the Set

Static Set

- Members of a static set do not change if the underlying data changes.

Business problem 5 :

- Create a set of customers with the highest profit.
 - Create a table which displays profit by each customer.
 - Drag Customer Name to Rows
 - Add Profit to Label
 - Sort in descending order

The screenshot shows the Tableau Data Source view for a dataset named "Orders (sample_...super...)". The interface includes a top navigation bar with icons for back, forward, search, and refresh, followed by a "Standard" dropdown. Below the navigation is a "Show Me" button. The main area is divided into several sections: "Data" (selected), "Analytics" (disabled), "Tables" (listing various dimensions like Category, City, Customer ID, etc.), "Pages" (showing "Sheet 6"), "Filters" (with a Marks card showing "Automatic" and "SUM(Profit)"), and a large "Sheet 6" pane displaying a list of customers sorted by profit. The data table has columns for "Customer Name" and "Profit". The first 11 rows are highlighted in blue, representing the selected set.

Customer Name	Profit
Tamara Chand	8,981
Raymond Buch	6,976
Sanjit Chand	5,757
Hunter Lopez	5,622
Adrian Barton	5,445
Tom Ashbrook	4,704
Christopher Martinez	3,900
Keith Dawkins	3,039
Andy Reiter	2,885
Daniel Raglin	2,869
Tom Boekenhauer	2,798
Nathan Mautz	2,752
Sanjit Engle	2,651
Bill Shonely	2,616
Harry Marie	2,438
Todd Sumrall	2,372
Brian Moss	2,199
Christopher Conant	2,177
Jane Waco	2,174
Helen Wasserman	2,164
Greg Tran	2,163
Laura Armstrong	2,059
Adam Bellavance	2,055
Fred Hopkins	2,050
Pete Kriz	2,038
Steven Roelle	1,990
Shirley Daniels	1,985

Sheet 6

Customer Name Profit

793 marks 793 rows by 1 column SUM(Profit): 286,397

- Select the first 11 records
- Click on Create Set

The screenshot shows the Tableau interface with the same dataset and view as the previous screenshot. The "Marks" card in the sidebar now has a green "SUM(Profit)" button highlighted. A context menu is open over the first 11 selected rows, with the "Create Set..." option highlighted. A tooltip indicates "11 items selected · SUM(Profit): 52,976". The main data table below shows the full list of 793 customers.

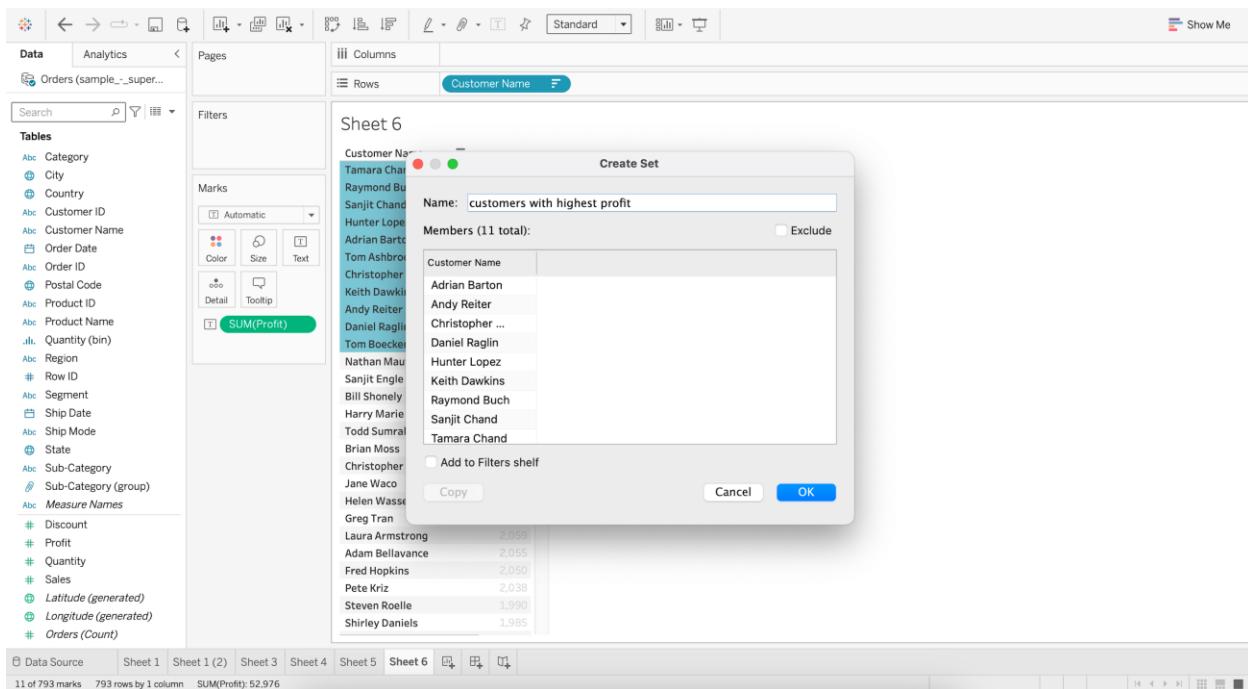
Customer Name	Profit
Tamara Chand	8,981
Raymond Buch	6,976
Sanjit Chand	5,757
Hunter Lopez	5,622
Adrian Barton	5,445
Tom Ashbrook	4,704
Christopher Martinez	3,900
Keith Dawkins	3,039
Andy Reiter	2,885
Daniel Raglin	2,869
Tom Boekenhauer	2,798
Nathan Mautz	2,752
Sanjit Engle	2,651
Bill Shonely	2,616
Harry Marie	2,438
Todd Sumrall	2,372
Brian Moss	2,199
Christopher Conant	2,177
Jane Waco	2,174
Helen Wasserman	2,164
Greg Tran	2,163
Laura Armstrong	2,059
Adam Bellavance	2,055
Fred Hopkins	2,050
Pete Kriz	2,038
Steven Roelle	1,990
Shirley Daniels	1,985

Sheet 6

Customer Name Profit

11 of 793 marks 793 rows by 1 column SUM(Profit): 52,976

- Rename the Set as “customers with highest profit”

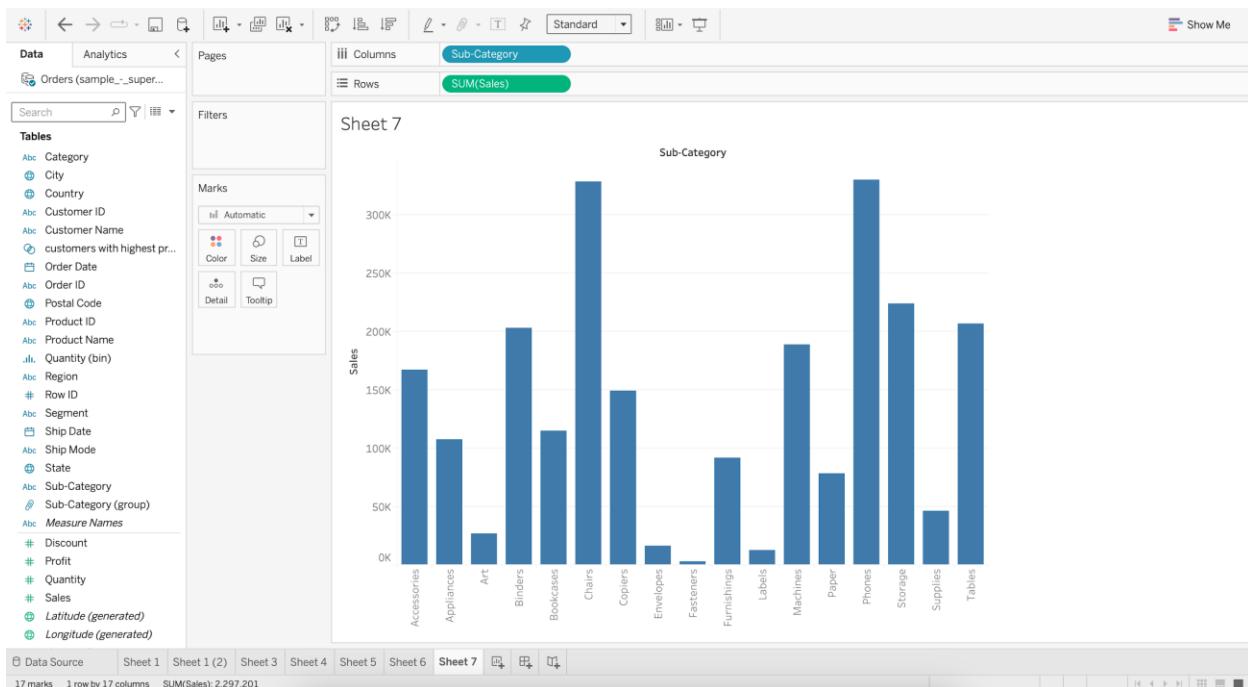


Dynamic Set

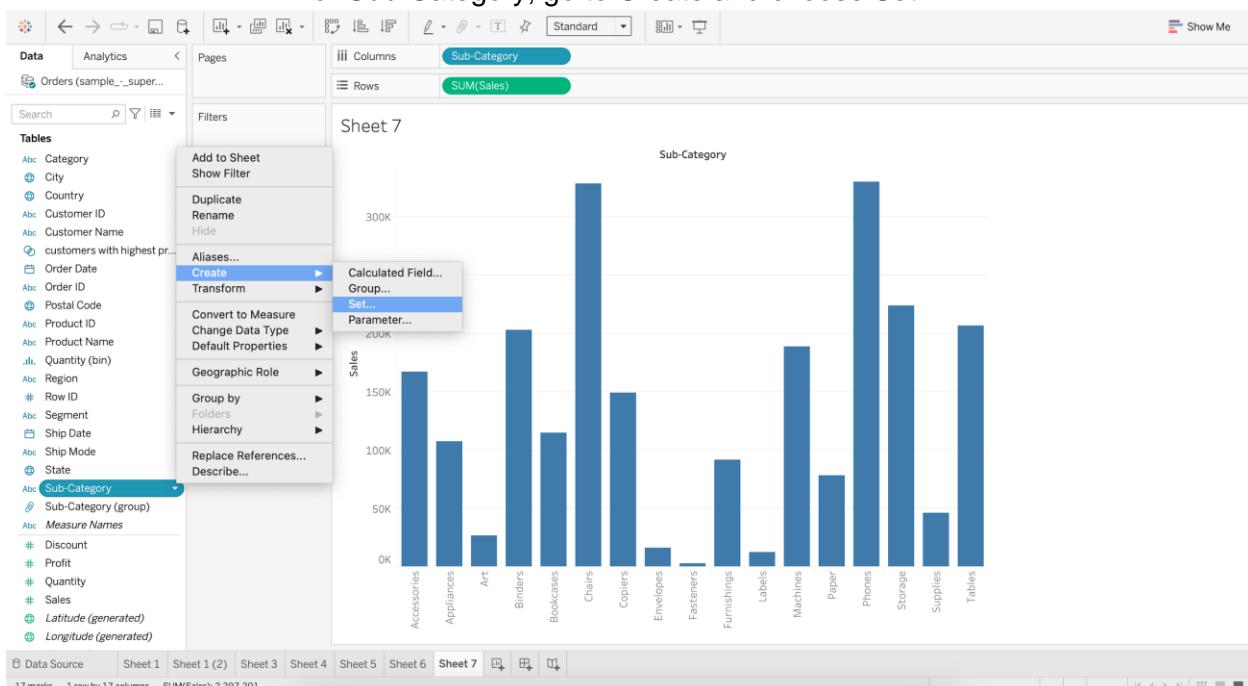
- Members of the dynamic set change when the underlying data changes.

Business problem 6 :

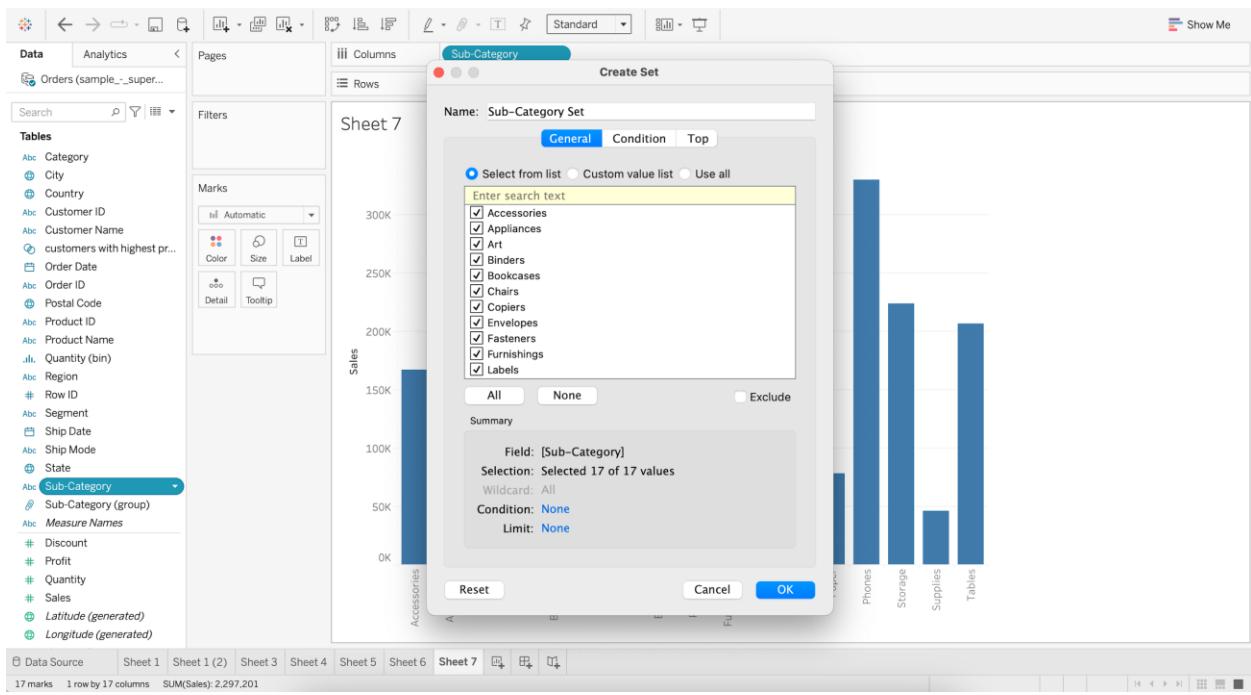
- Find the top 5 products subcategories by sales.
 - Create a bar chart to visualize sales by each sub-category.
 - Drag Sub-Category to Columns
 - Drag Sales to Rows



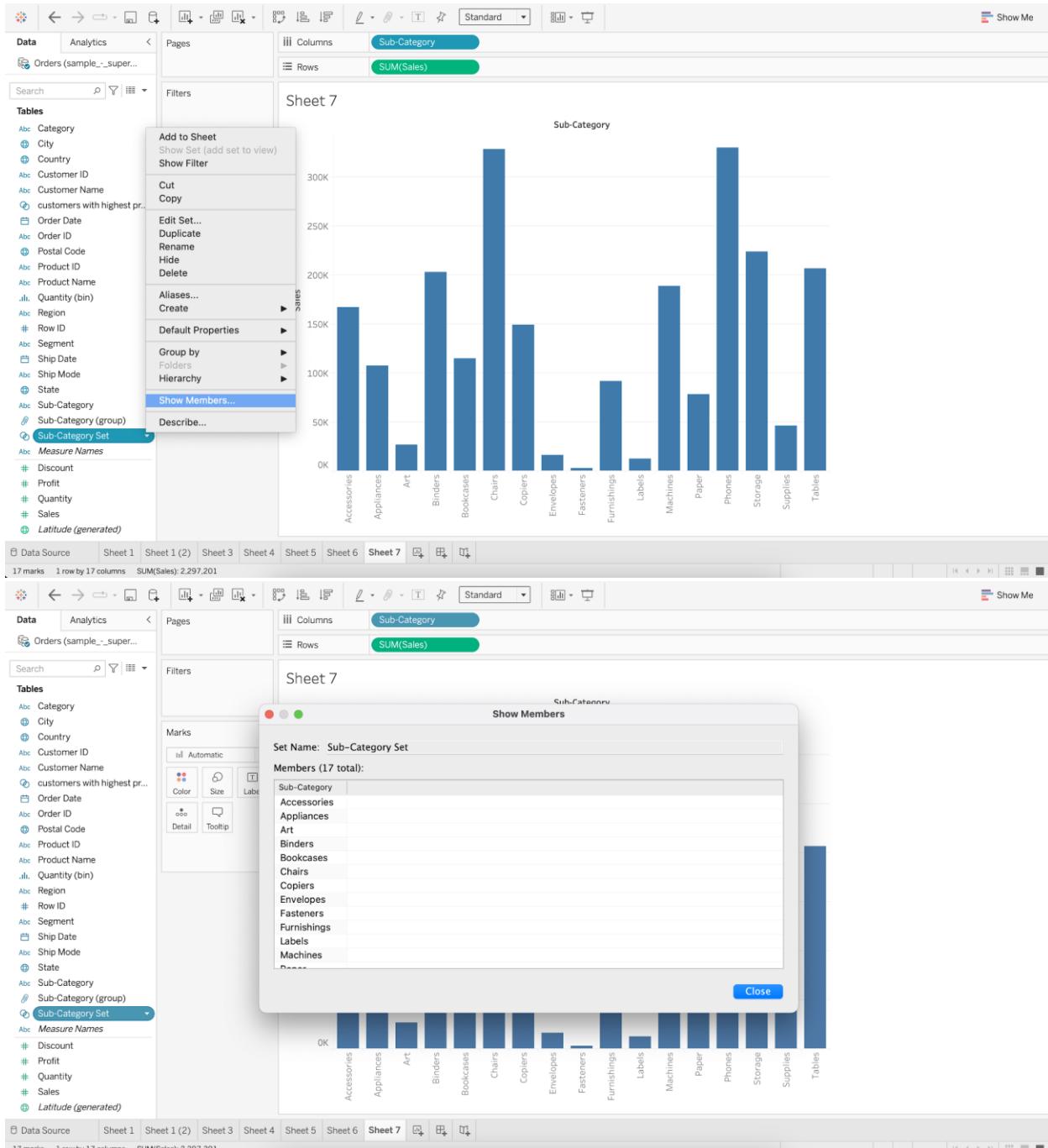
■ For Sub-Category, go to Create and choose Set



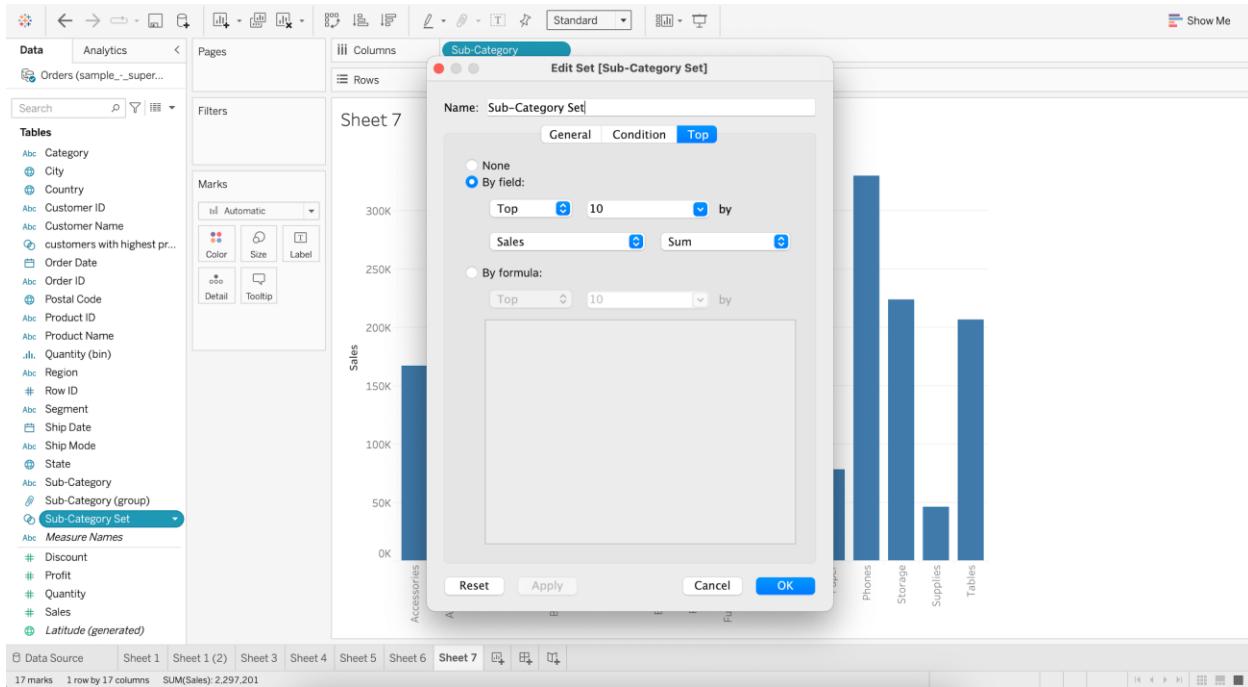
■ Select ALL from list and click OK



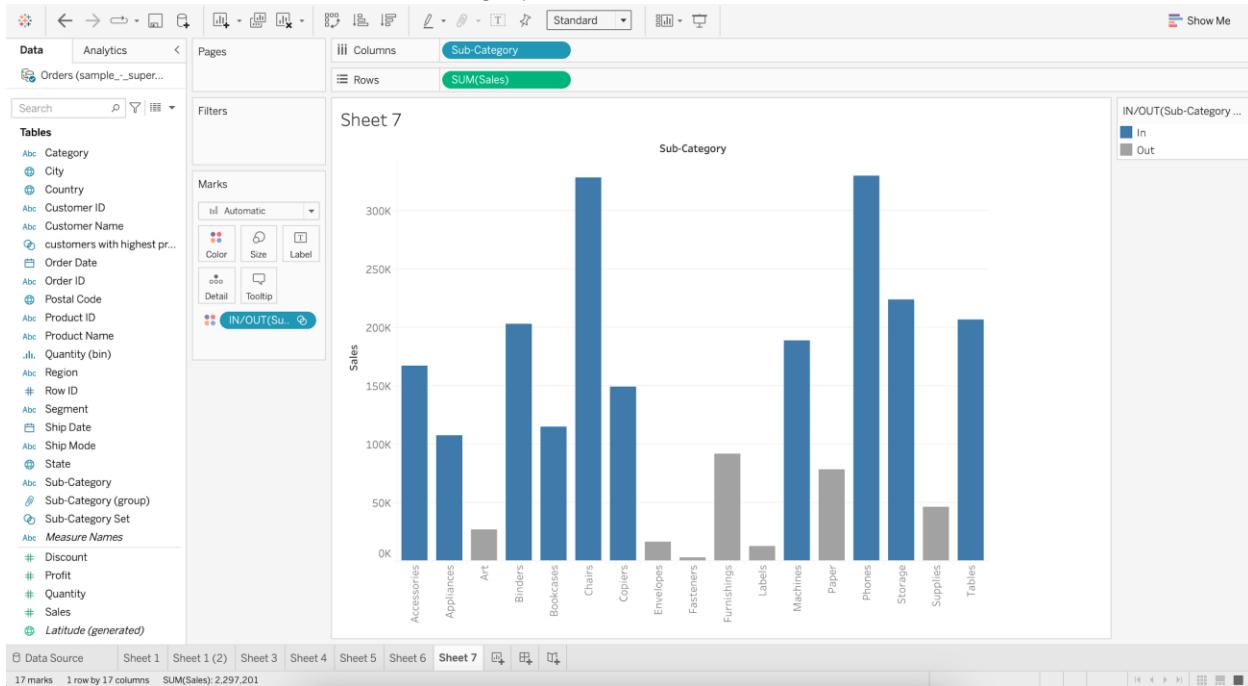
■ Show Members of Sub-Category Set



- For Sub-Category Set, go to Edit Set
- Go to Top and Select Top 5 Fields by Sum of Sales



■ Add Sub_Category Set to Color



The ones in Blue are within the Set.
The ones in Grey are outside the Set.

Combined Set

- Allows us to compare members from 2 different sets.

Business problem 7 :

- Find the top customers by sales but with low profitability.
- Create Set 1, which is “Top 20 customers by Sales”

The screenshot shows the Tableau Data Source interface with the 'Create Set' dialog open. The dialog is titled 'Create Set' and has a 'Name:' field containing 'Top 20 customers by Sales'. The 'Condition' tab is selected. Under 'By field:', 'Top' is chosen, and '20' is entered. The 'Sales' field is selected, and 'Sum' is chosen as the aggregation method. The 'OK' button is highlighted in blue.

- Create Set 2, which is “Bottom 20 customers by Profit”

The screenshot shows the Tableau Data Source interface with the 'Create Set' dialog open. The dialog is titled 'Create Set' and has a 'Name:' field containing 'Bottom 20 customers by Profit'. The 'Condition' tab is selected. Under 'By field:', 'Bottom' is chosen, and '10' is entered. The 'Profit' field is selected, and 'Sum' is chosen as the aggregation method. The 'OK' button is highlighted in blue.

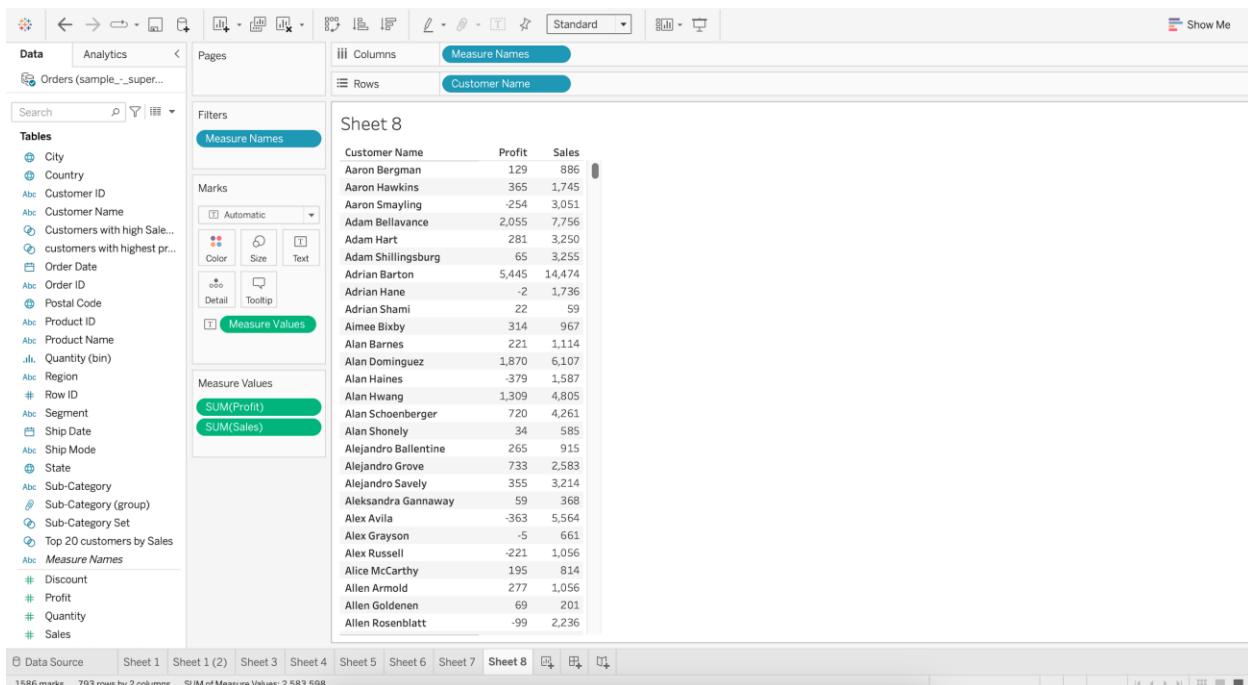
- Create a Combined Set as “Customers with high Sales and low Profit”

The screenshot shows the Tableau Data Source interface. A context menu is open over a set named "Top 20 customers by Sales". The menu includes options like "Add to Sheet", "Cut", "Copy", "Edit Set...", "Create Combined Set...", "Duplicate", "Rename", "Hide", "Delete", "Aliases...", "Create", "Default Properties", "Group by", "Folders", "Hierarchy", "Show Members...", and "Describe...". The "Create Combined Set..." option is highlighted.

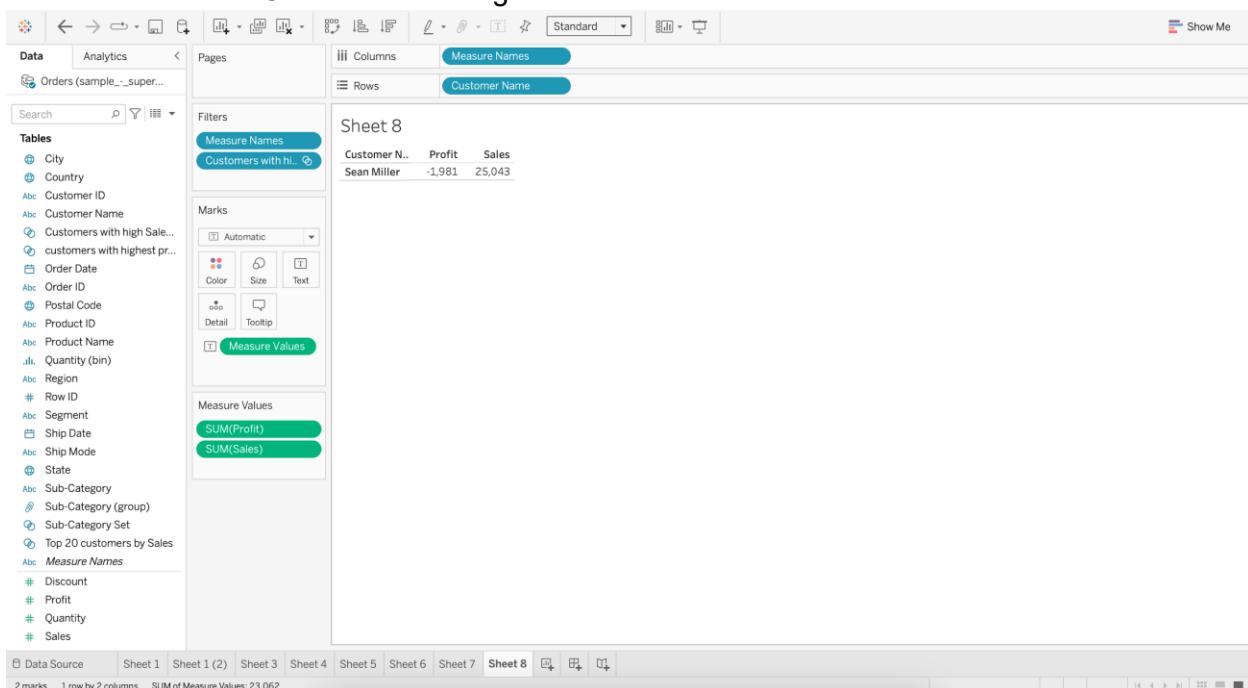
- Combine sets “Top 20 customers by Sales” and “Bottom 20 customers by Profit”
- Choose Shared members in both sets and click OK

The screenshot shows the Tableau Data Source interface with the "Create Set [Set 1]" dialog box open. The dialog box has a "Name:" field containing "Customers with high Sales and low Profit". Below it, a question "How would you like to combine the two sets?" is followed by a "Sets:" dropdown containing "Top 20 customers by Sales" and "Bottom 20 customers by Pr...". Underneath, there are four radio button options: "All members in both sets" (unchecked), "Shared members in both sets" (checked), "'Top 20 customers by Sales'" except shared members (unchecked), and "'Bottom 20 customers by Profit'" except shared members (unchecked). At the bottom, there is a "Separate members by:" dropdown set to "East, Green Tea, 2012" and "OK" and "Cancel" buttons.

- Drag Customer Name to Rows
- Add Sales & Profit as Measure Values



- Put “Customers with high Sales and low Profit” into Filter



Groups vs Sets

Group	Set
Static (must be manually changed)	(Can be) Dynamic

No limit to number of groups	Binary (In/Out)
Can be created on both Dimensions and Measures	Can be created only on Dimension

Parameter

- It is like a workbook variable that is used to replace constant value in a filter, calculation or reference lines.
- They are user generated values that are not necessarily attached to the dataset.

Business problem 8 :

- Allow users to filter out top N subcategories of products by average discount.
 - Create a Parameter as “Filter top N products”

The screenshot shows the Tableau Data Source pane on the left side of the interface. The pane lists various fields categorized under 'Orders (sample_..._super...)'. The categories include:

- Data
- Analytics
- Pages
- Filters
- Marks

Under 'Data', the following fields are listed:

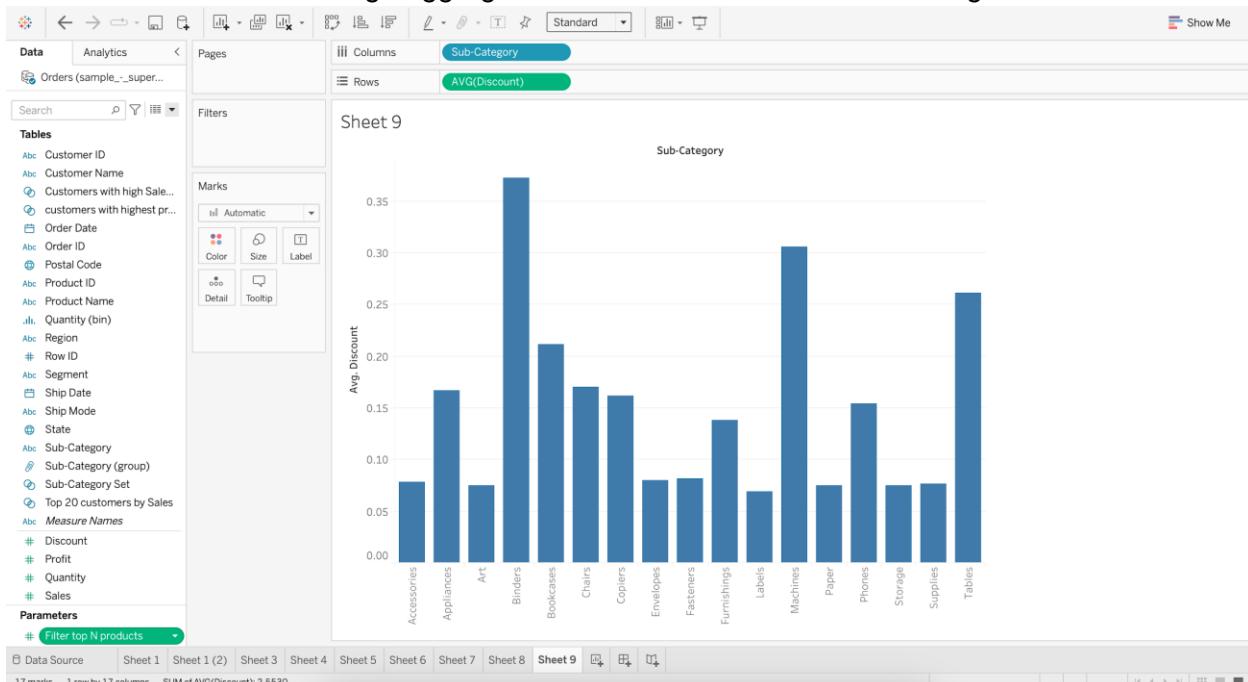
- Create Calculated Field...
- Create Parameter...** (highlighted in blue)
- Group by Folder
- ✓ Group by Data Source Table
- ✓ Sort by Name
- Sort by Data Source Order
- Hide All Unused Fields
- Show Hidden Fields
- Expand All
- Collapse All
- Abc: Product Name
- .Jlh.: Quantity (bin)
- Abc: Region
- # Row ID
- Abc: Segment
- Ship Date
- Abc: Ship Mode
- State
- Abc: Sub-Category
- Sub-Category (group)
- Sub-Category Set
- Top 20 customers by Sales
- Abc: Measure Names
- # Discount
- # Profit
- # Quantity
- # Sales

- Set :
 - Data type as Integer
 - Current value as 1
 - Allowed values as Range
 - Minimum value as 1
 - Maximum value as 10
 - Step size as 1

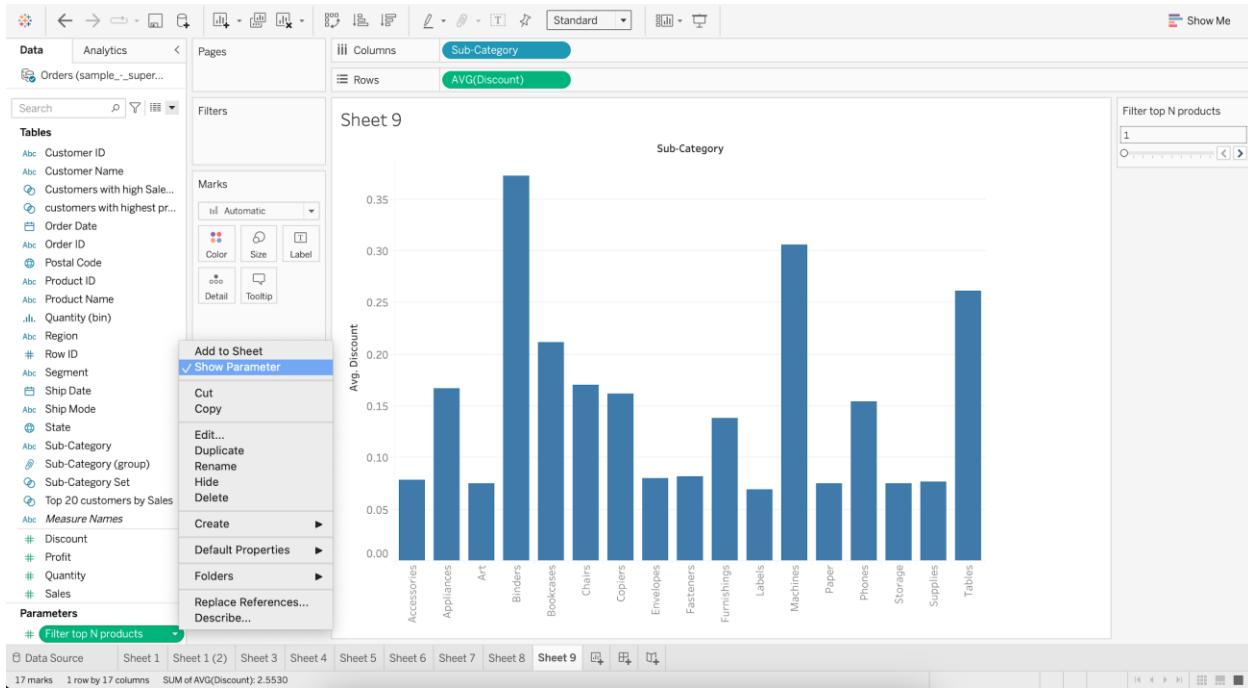
The screenshot shows the Tableau Data Source interface with the 'Create Parameter' dialog box open. The dialog box is titled 'Create Parameter' and contains the following fields:

- Name:** Filter top N products
- Properties:** Data type: Integer, Display format: 1
- Current value:** 1
- Allowable values:** Range selected, Minimum: 1, Maximum: 10, Step size: 1
- Range of values:** Fixed selected

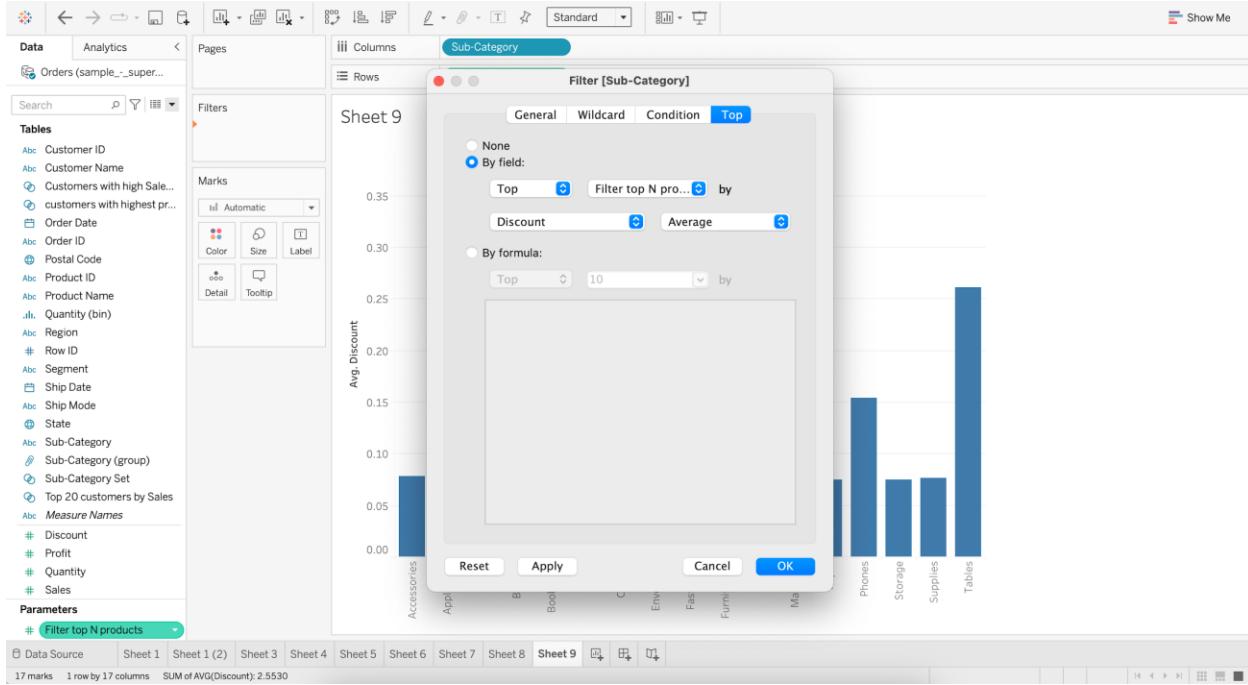
- Display average discount for each subcategory.
 - Drag Sub-Category to Columns
 - Drag Discount to Rows
 - Change aggregation of Discount from Sum to Average



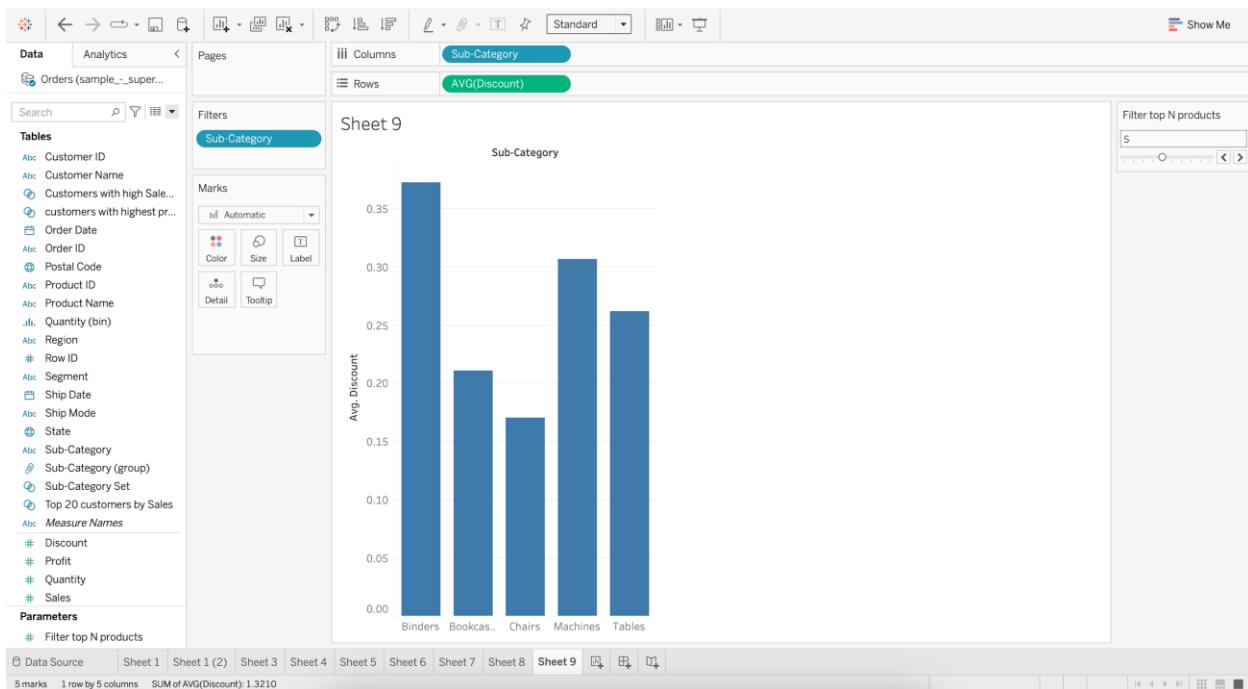
- Go to “Filter top N products” and Show Parameter



- Put Sub-Category into Filter
- Go to Top and Select Filter top N products by Average of Discount



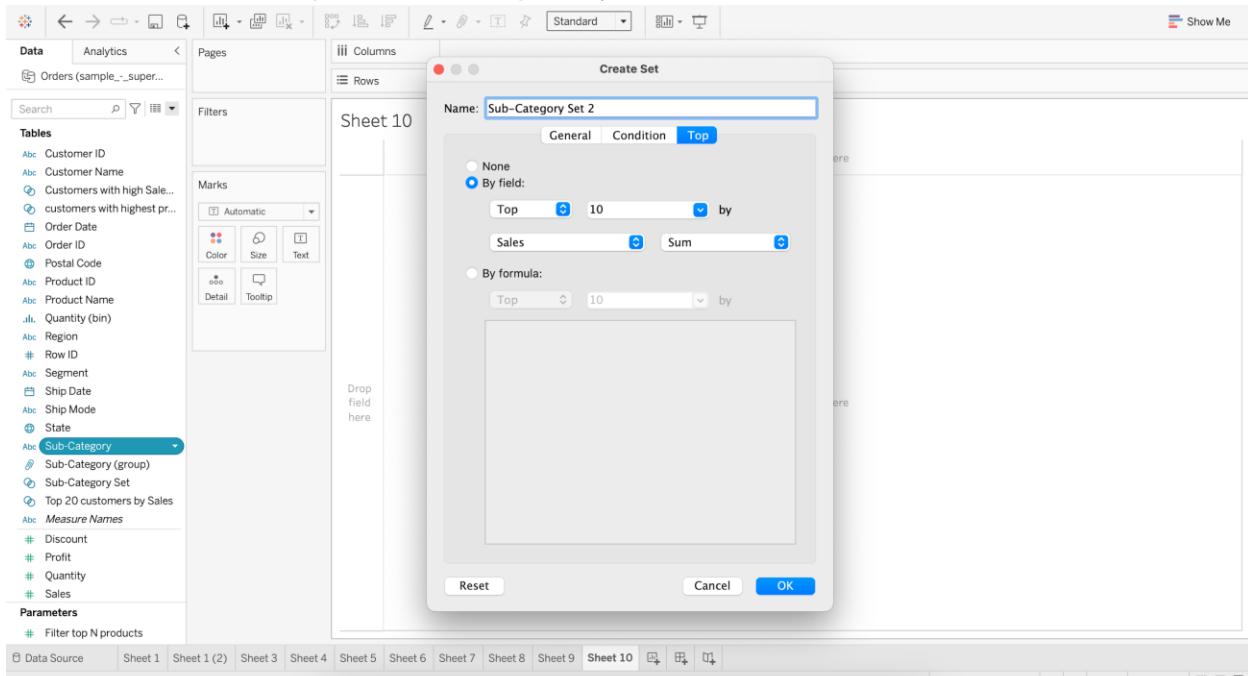
- Increase value of "Filter top N products" to 5



Control Set with Parameter

Business problem 9 :

- Have a parameter where the user can choose top N product subcategories by sales.
 - Create a null Set of Sub-Category
 - Go to Top and Select Top 10 by Sum of Sales



- Create a Parameter as “Sub-Category Selector”

- Set :

- Data type as Integer
- Current value as 1
- Allowed values as Range
- Minimum value as 1
- Maximum value as 10
- Step size as 1

The screenshot shows the Tableau interface with the 'Create Parameter' dialog open. The dialog is titled 'Create Parameter' and contains the following settings:

- Name:** Sub-Category Selector
- Properties:**
 - Data type: Integer
 - Display format: 1
 - Current value: 1
 - Value when workbook opens: Current value
- Allowable values:**
 - Range of values:
 - Minimum: 1
 - Maximum: 10
 - Step size: 1

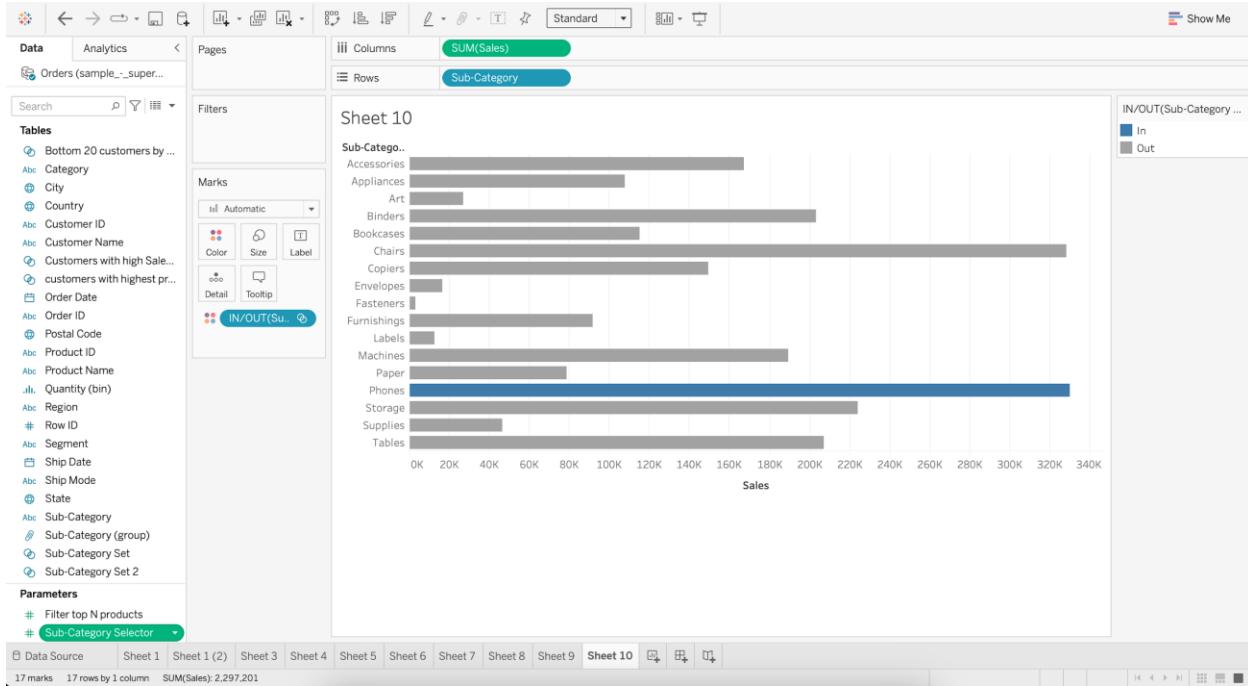
- Edit Sub-Category Set 2

- Go to Top and Select Sub-Category Selector

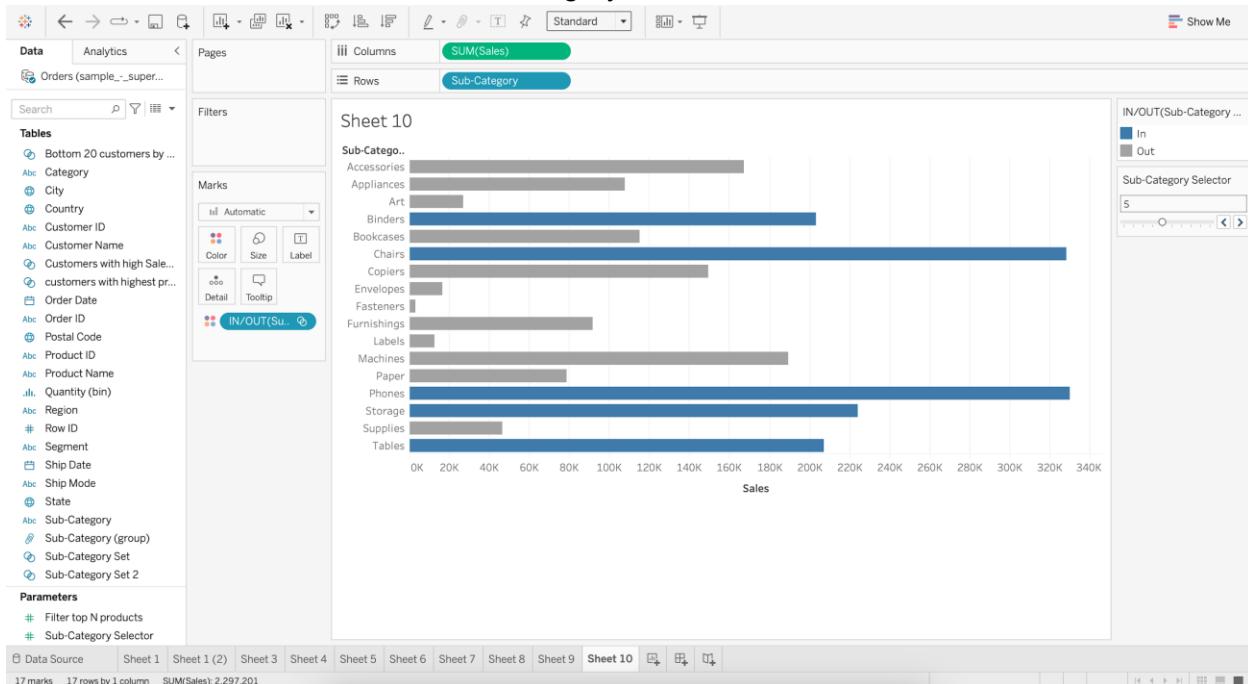
The screenshot shows the Tableau interface with the 'Edit Set [Sub-Category Set 2]' dialog open. The dialog is titled 'Edit Set [Sub-Category Set 2]' and has the 'Top' tab selected. The configuration is as follows:

- Name:** Sub-Category Set 2
- General:** None
- Condition:** By field: Top 10 Sub-Category ... by Sales Sum
- Top:** Top 10
- By formula:** Top 10

- Drag Sub-Category to Rows
- Drag Sales to Columns
- Add Sub-Category Set 2 to Color



- Go to “Sub-Category Selector” and Show Parameter
- Increase value of “Sub-Category Selector” to 5



Post-read: [Hierarchy & Sorting](#)

Tableau Lecture 4: Filters and Calculations

Agenda

- Charts
 - Stacked Bar chart
 - Scatter plot
 - Treemap
 - Combined axis chart
 - Dual-axis chart
- Filters and Calculations
 - Types of Filter
 - Order of Operations

Dataset : [sample superstore](#)

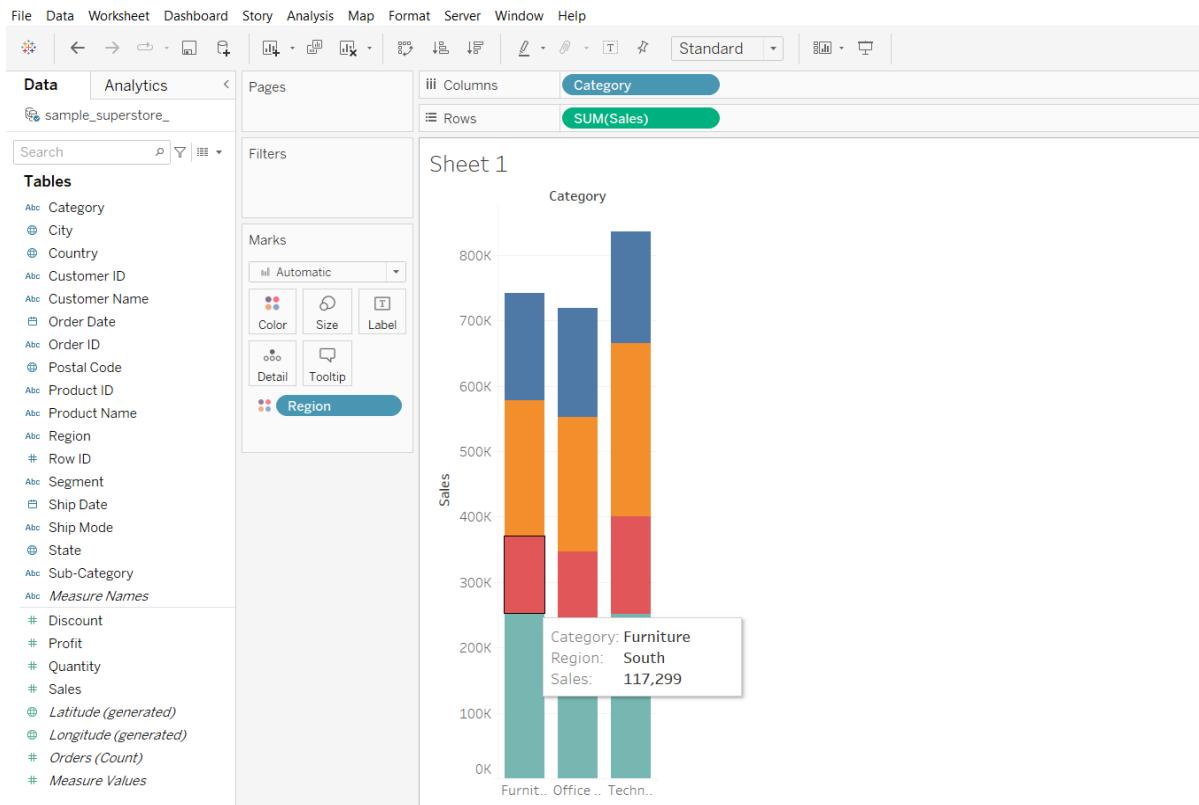
Stacked Bar Chart

- It is a simple bar chart with segmented bars.
- It requires
 - 1 or more dimension and
 - 1 or more measures.
- When you select dimension into the color marks chart what you get is a stacked bar chart.

Business problem 1:

Display total sales for each category by region.

- Drag Category to Columns
- Drag Sales to Rows
- Drag Region to Color in Marks card



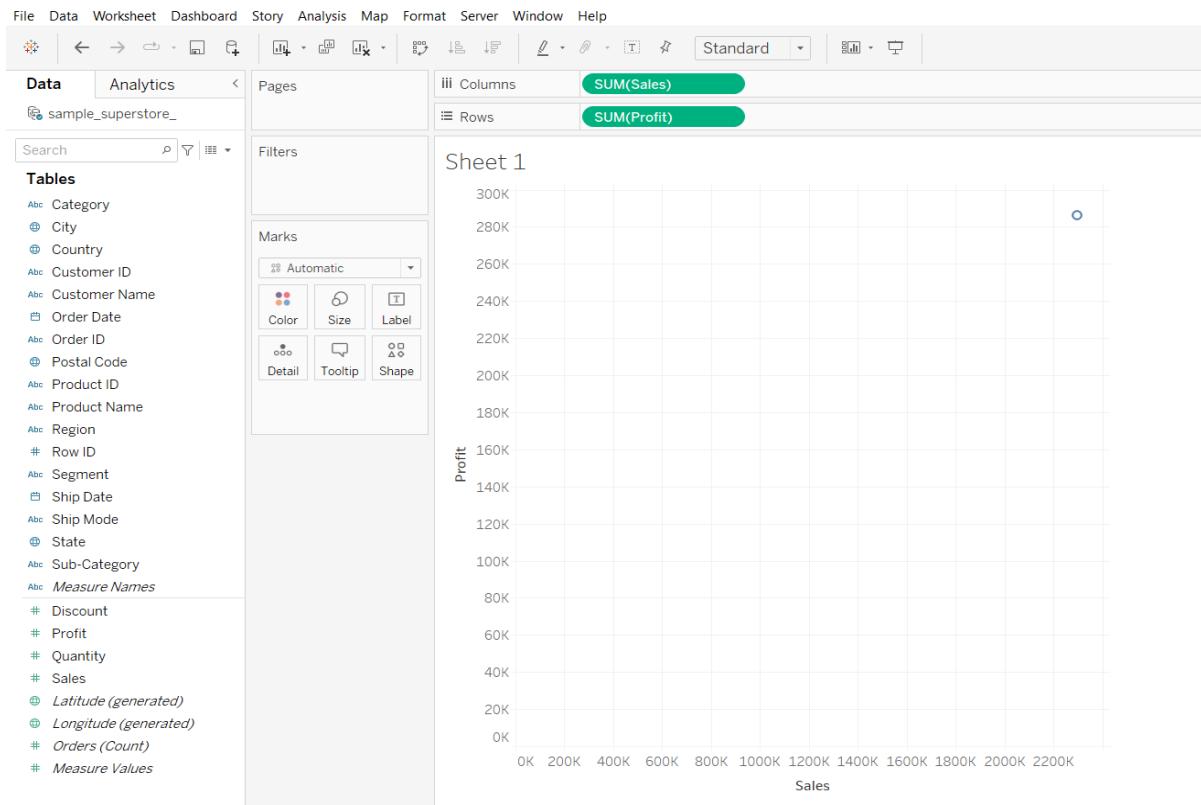
Scatter Plot

- It displays a data point at their respective intersection of both the measures
- We can create a scatter plot by adding at least one measure on row shelf and one measure on column shelf

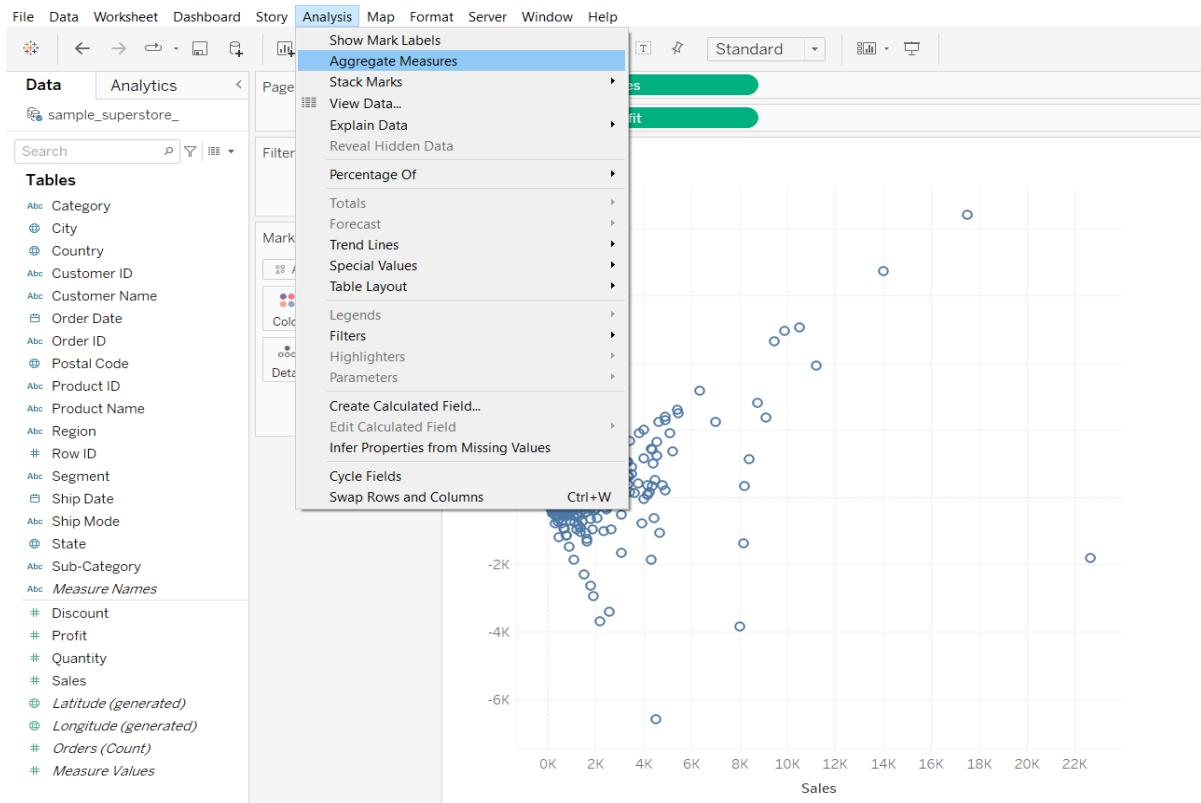
Business problem 2:

Find the order id that has the highest sales and highest profit values.

- Creating a scatter chart what we see below is the total sales and total profit.
 - Drag Profit to Rows
 - Drag Sales to Columns



- Disaggregating the measures to plot every row values in the plot and solving the business problem.
 - Click on Analysis in Menu bar
 - Uncheck the Aggregate Measures option



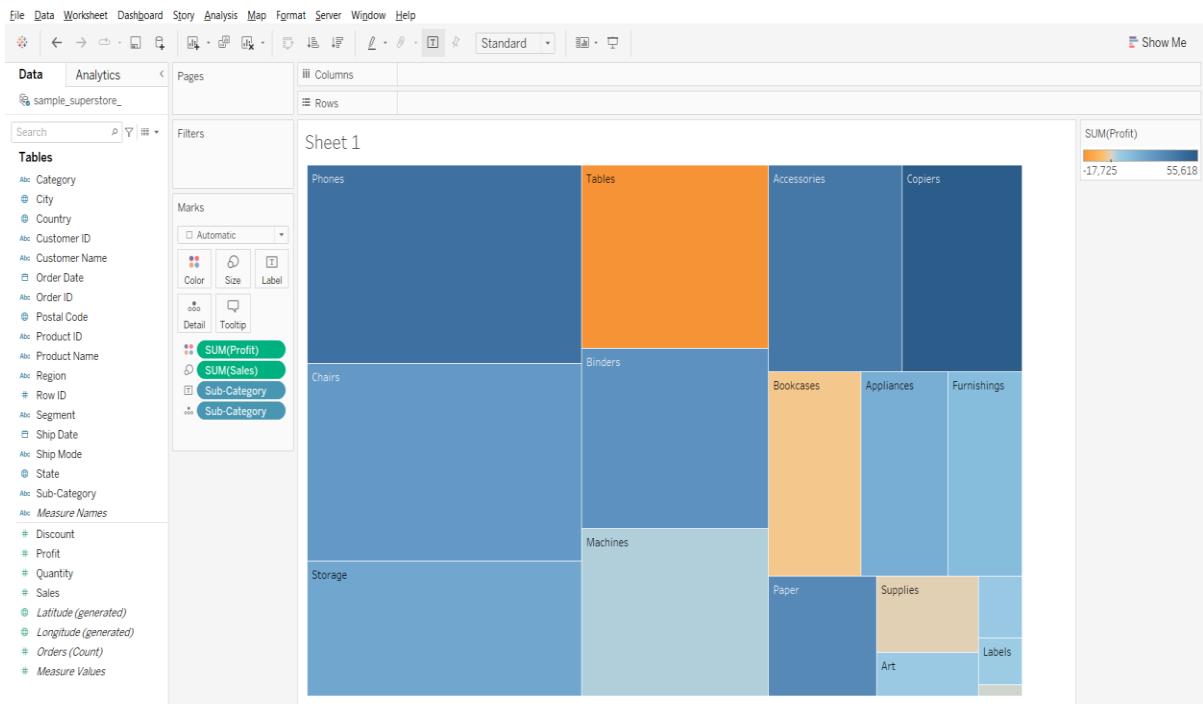
Tree Map

- It is a visualization that nests rectangles in hierarchies
- It uses dimension to define the structure of the tree map and measure to define the size or color of each of the rectangle
- It requires one or more dimensions and one or 2 measures

Business problem 3:

Find the subcategory that has the highest sales and highest profit.

- Drag sub-category to Detail in Marks Card
- Drag sub-category to Label in Marks card
- Drag Sales to Size in Marks card
- Drag Profit to Color in Marks Card



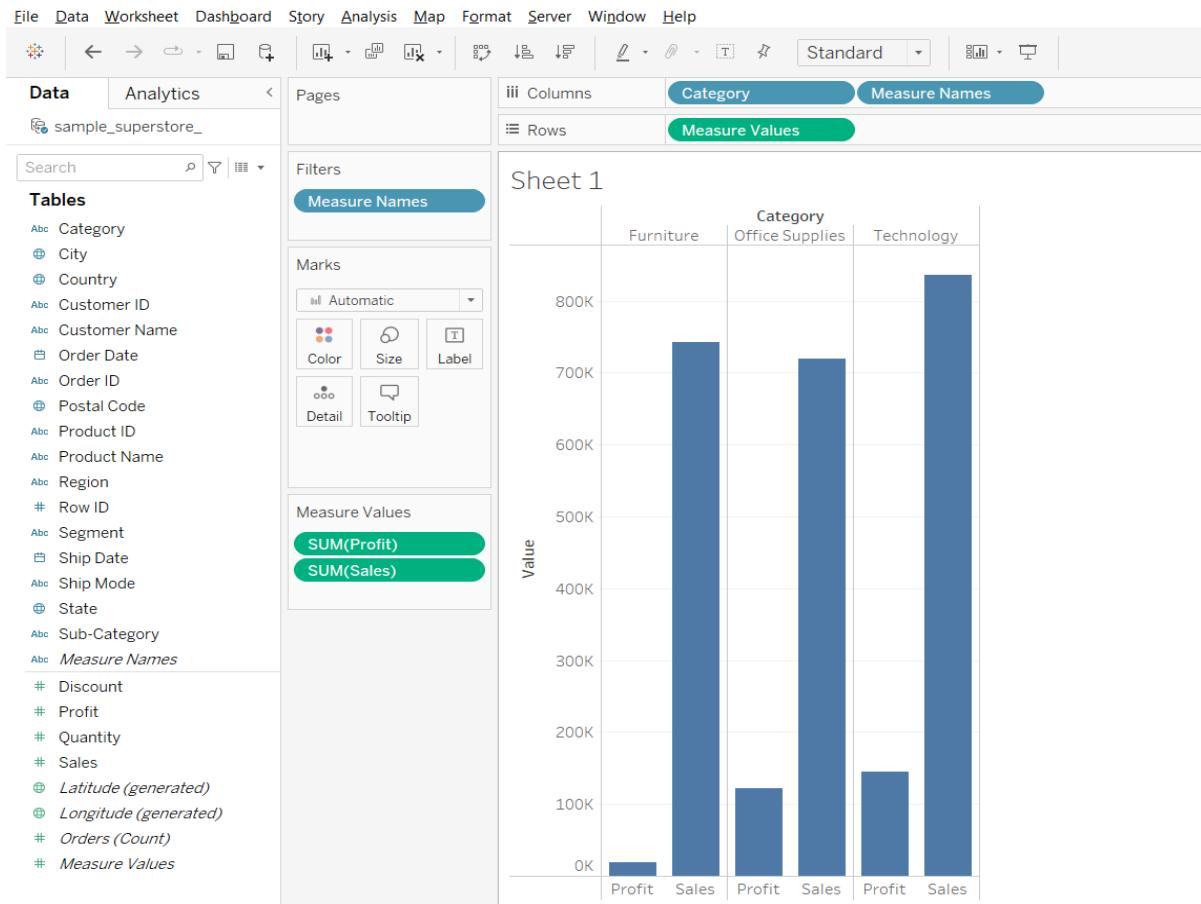
Combined Axis Chart

- Also called as Blended axis or shared axis chart
- Measures share a single axis so that all the marks are shown in a single pane
- They are used to compare measures of same data type and same scale
- Only single Mark card is present and we can compare more than 2 measure
- To create a combined axis chart drag the measure on the left side of the already existing axis
- It requires one or more dimension and 2 or more measures

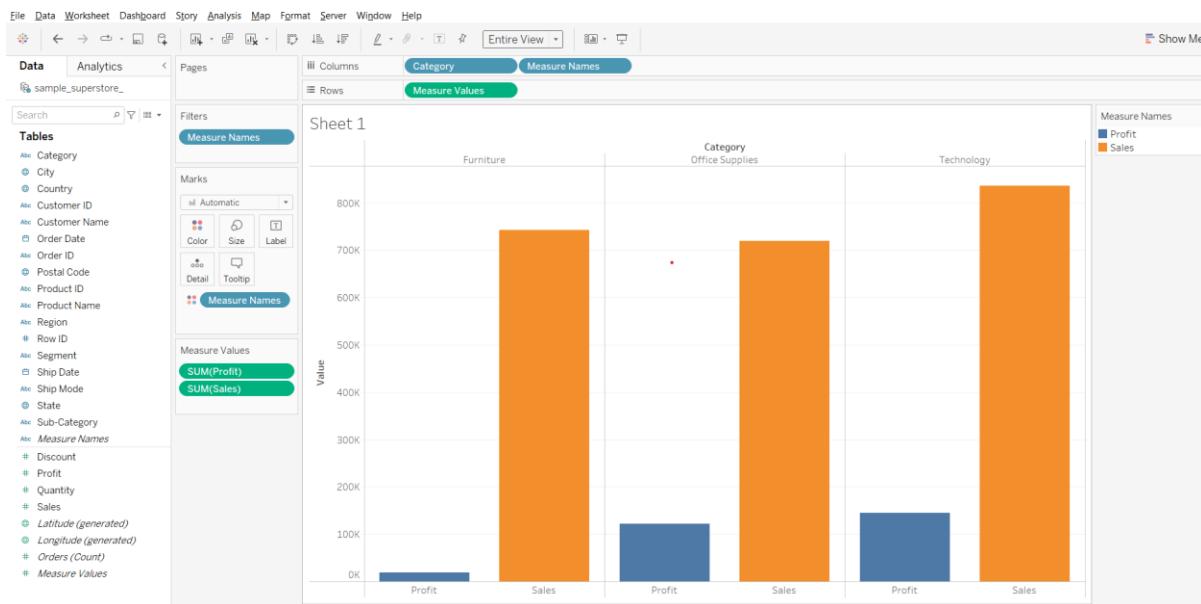
Business problem 4:

Compare sales and profit for each category in a single plot.

- Drag Category to Columns
- Drag Sales to Rows
- Drag Profit to vertical axis of the chart



- Drag Measure Names to Color in Marks Card
- Change Fit from Standard to Entire View in toolbar



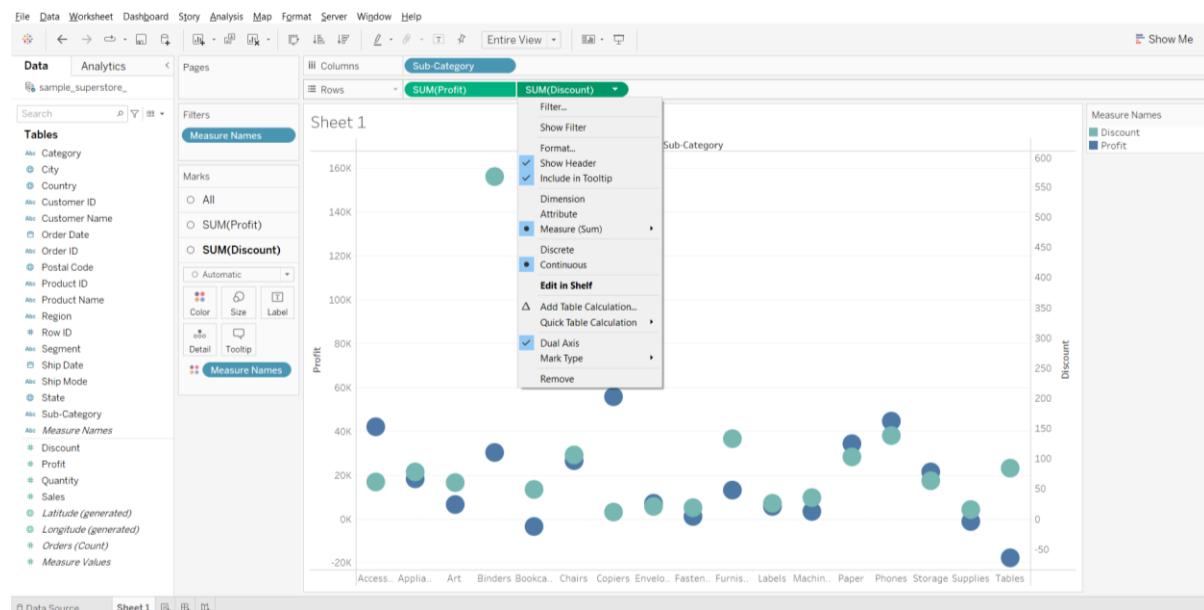
Dual Axes Charts

- You can compare multiple measures using dual axes, which are two independent axes that are layered on top of each other.
- Dual axes are useful for analyzing two measures with different scales.
- Creates multiple marks
- Can't compare more than 2 measures
- To create a dual axis chart we can drag 2nd measure to right side of the plot to create another axis or right click on the 2nd measure and select dual axis
- It requires 2 measures and at least one dimension

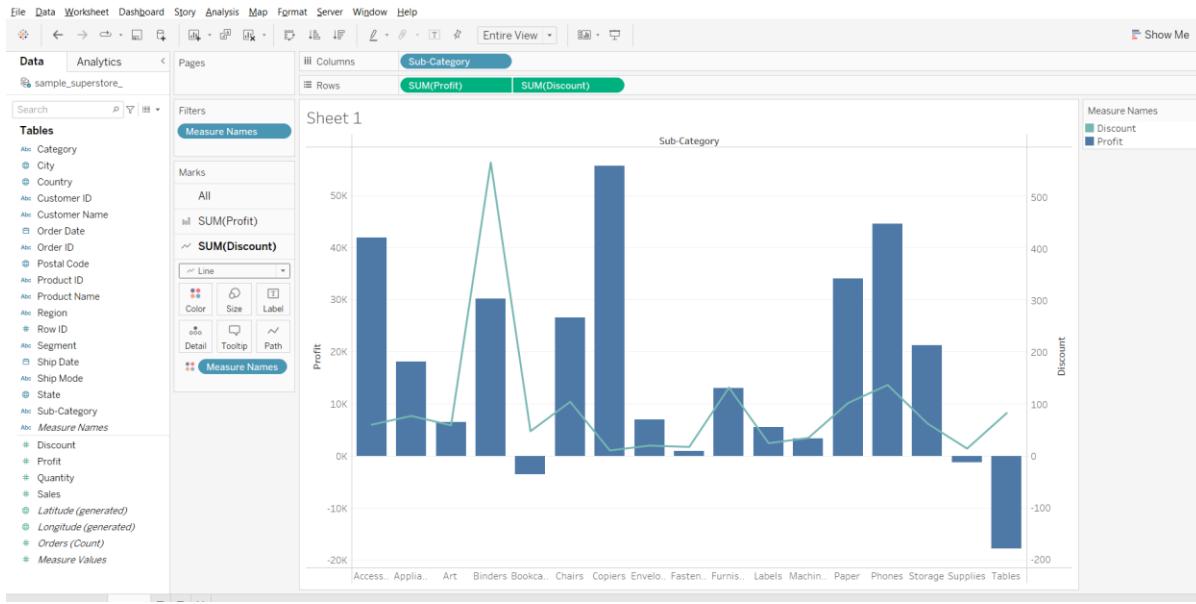
Business problem 5:

Find the sub-category that has the lowest discount but highest profit.

- Drag Sub-Category to Columns
- Drag Profit to rows
- Drag Discount to Rows
- Open the Discount field dropdown and choose Dual Axis



- In Marks card, change plot type of Profit to Bar
- In Marks card, change plot type of Discount to Line



Filters

- Filter removes some scope of data from a data set.
- Filters are very helpful to create dashboards in Tableau.
- Filters can help to minimize the size of data sets for efficient use, eliminate irrelevant dimension elements, clean up underlying data, set date ranges and measures as required, simplify and organize data, etc.

Types of Filters

1. Extract filters

- Extract filters in Tableau are used to extract a small subset of data from the original data source.
- Tableau then creates a local copy of the data set that is to be stored in the repository.
- These methods reduce Tableau queries.
- The data size can be further reduced by applying the measure or dimension filter to the extract as required.
- Unlike Tableau Desktop(paid) which supports both extract and live connections, Tableau public, by default, only supports extract.

Important Note:

Tableau Public always works with extracts, and therefore does not show the options 'Live' and 'Extract'.

Here is a screenshot of Tableau Desktop Professional that supports both extract and live connections.

The screenshot shows the Tableau interface with the following details:

- Connections:** Sample - Superstore
- Sheets:** Orders, People, Returns, Orders, People, Returns, New Union
- Data View:** A hierarchical diagram showing 'Orders' connected to 'People' and 'Returns'. The 'People' node has two children: 'Claire Gute' and 'Darrin Van Huff'.
- Connection Options:** Connection dropdown set to 'Extract' (highlighted with a red circle). Note: 'Extract will include subset of data'.
- Data Grid:** A table with 10 rows of order data. The columns are: Row ID, Order ID, Order Date, Ship Date, Ship Mode, Customer ID, Customer Name, Segment, Country/Region, City, State, Postal Code, Region, Profit.

Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country/Region	City	State	Postal Code	Region	Profit
1	CA-2020-152156	11/8/2020	11/11/2020	Second Class	00-12520	Claire Gute	Consumer	United States	Henderson	Kentucky	42420	South	1
2	CA-2020-152156	11/8/2020	11/11/2020	Second Class	00-12520	Claire Gute	Consumer	United States	Henderson	Kentucky	42420	South	1
3	CA-2020-188688	6/12/2020	6/16/2020	Second Class	00-13045	Darrin Van Huff	Corporate	United States	Los Angeles	California	90036	West	1
4	US-2019-109966	10/11/2019	10/18/2019	Standard Class	00-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	Florida	33311	South	1
5	US-2019-109966	10/11/2019	10/18/2019	Standard Class	00-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	Florida	33311	South	1
6	CA-2018-115812	6/9/2018	6/14/2018	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	1
7	CA-2018-115812	6/9/2018	6/14/2018	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	1
8	CA-2018-115812	6/9/2018	6/14/2018	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	1
9	CA-2018-115812	6/9/2018	6/14/2018	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	1
10	CA-2018-115812	6/9/2018	6/14/2018	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	1

2. Data Source filters

- Data source filters in Tableau are mainly used to restrict sensitive data from viewers and reduce data feeds.
- Viewers can, however, have certain access rights to view the underlying data.
- Data source filters allow the direct application to source data.
- One important thing to mention is that the extract filter and the data source filter are not linked, and if you happen to go back to a live connection, the data source filter will remain intact.

Business Problem 6 :

Show orders which made profit $\geq 1K$.

- Go to the Data Source tab
- Click on add filter on the top right.
- Click on Add...
- Select Profit from the list of fields.
- Set the minimum value to 1000.

The screenshot shows the Tableau Data Source interface for a 'sample_superstore_'. The 'Orders' sheet is selected. A context filter dialog box is open, titled 'Add Filter', with 'Profit' selected from a list of fields.

Date	Ship Date	Ship Mode	Customer ID	Customer Name
08-11-2016	11-11-2016	Second Class	CG-12520	Claire Gute
08-11-2016	11-11-2016	Second Class	CG-12520	Claire Gute
12-06-2016	16-06-2016	Second Class	DV-13045	Darrin Van Huff
			SO-20335	Sean O'Donnell
			SO-20335	Sean O'Donnell
			BH-11710	Brosina Hoffman
			BH-11710	Brosina Hoffman
			BH-11710	Brosina Hoffman
			BH-11710	Brosina Hoffman

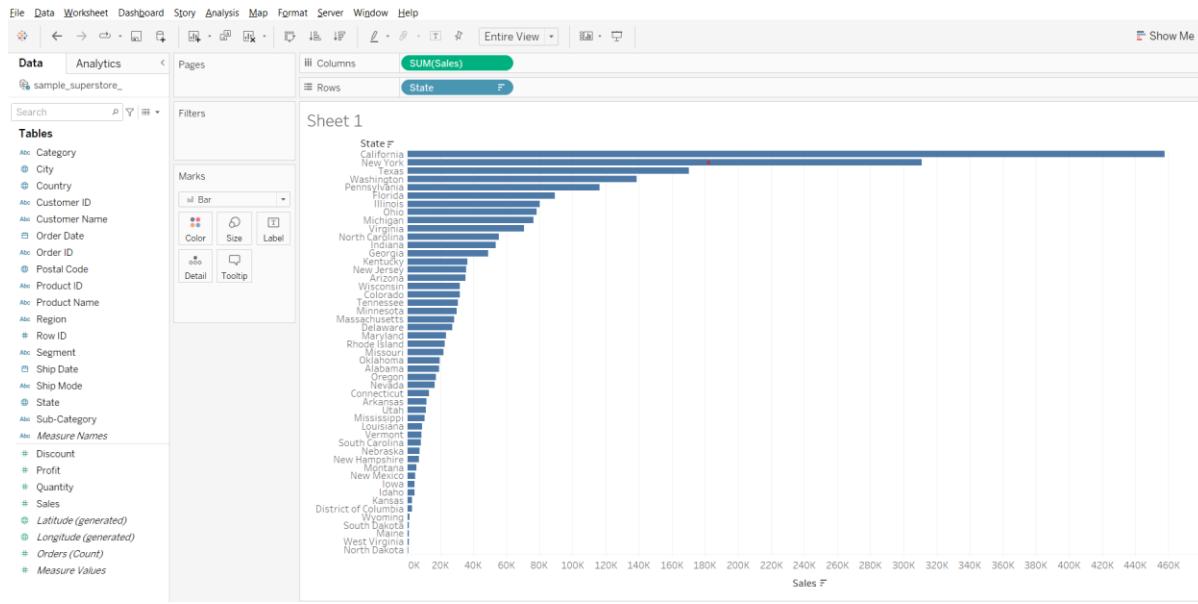
3. Context filters

- By default, all filters that you set in Tableau are computed independently. That is, each filter accesses all rows in your data source without regard to other filters. However, you can set one or more categorical filters as context filters for the view.
- You can think of a context filter as being an independent filter. Any other filters that you set are defined as dependent filters because they process only the data that passes through the context filter.
- The context filter adds an actionable context to data analysis, but if the data is not reduced enough, the cost of computing can be very high.

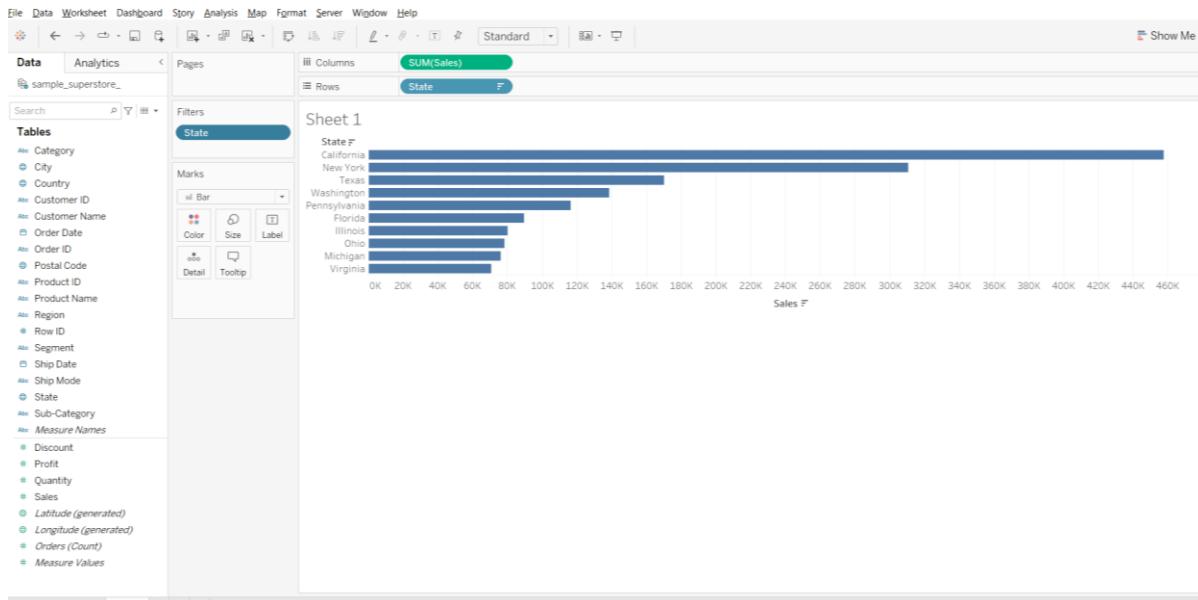
Business Problem 7 :

Find the top ten states in the West region with the highest sales.

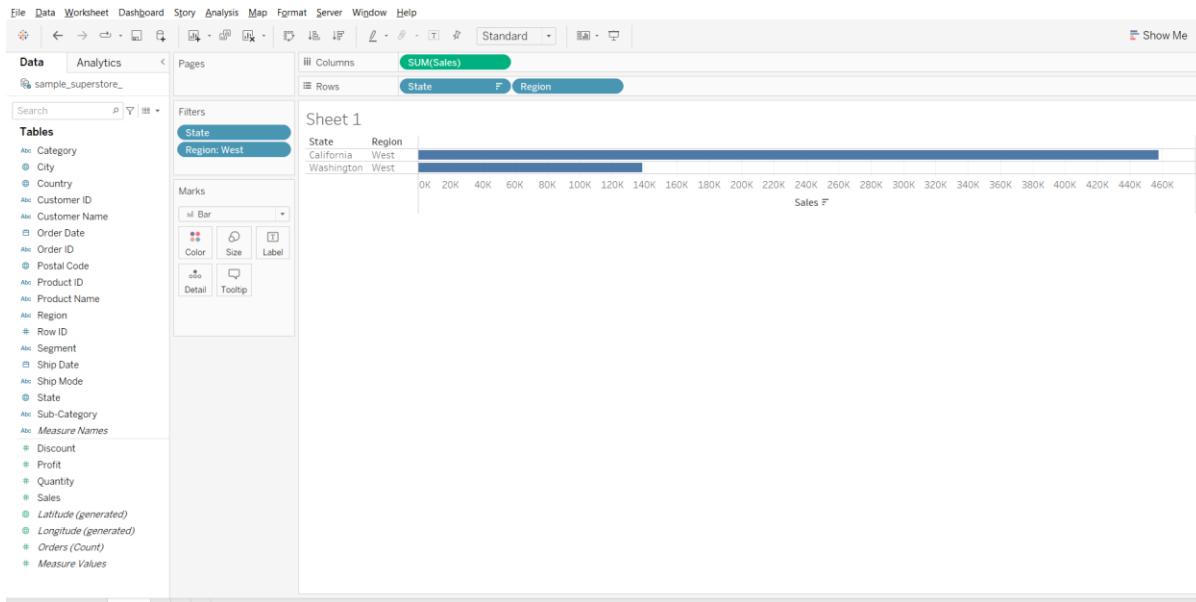
- Drag State to Rows
- Drag Sales to Columns
- Click on the sort-descending button in toolbar



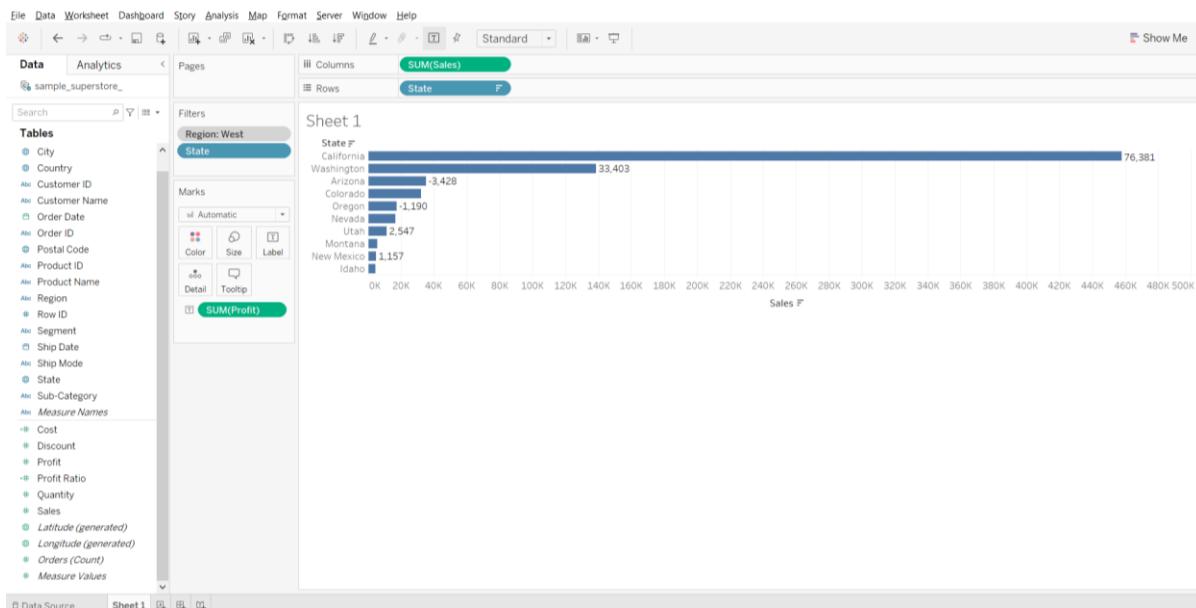
- Drag State to Filters card
- Click on the Top tab of Filter dialog box
- Choose Top 10 by Sales field



- Drag Region to Rows
- Drag Region to Filters card
- Deselect all regions, then select West



- Right click on Region in Filters card, select Add to Context.



Explanation:

- First we filter out top 10 states by sales
- Next when we apply filter for region=West
- We only see 2 entries that instead of 10 that is because top 10 filter gets executed first and then the region filter is executed if you see from the 10 states only 2 states belong to west region
- Now in order to change the order of operation we add region filter to context filter this ensures that first our region filter gets executed and then the top 10 by sales is executed.

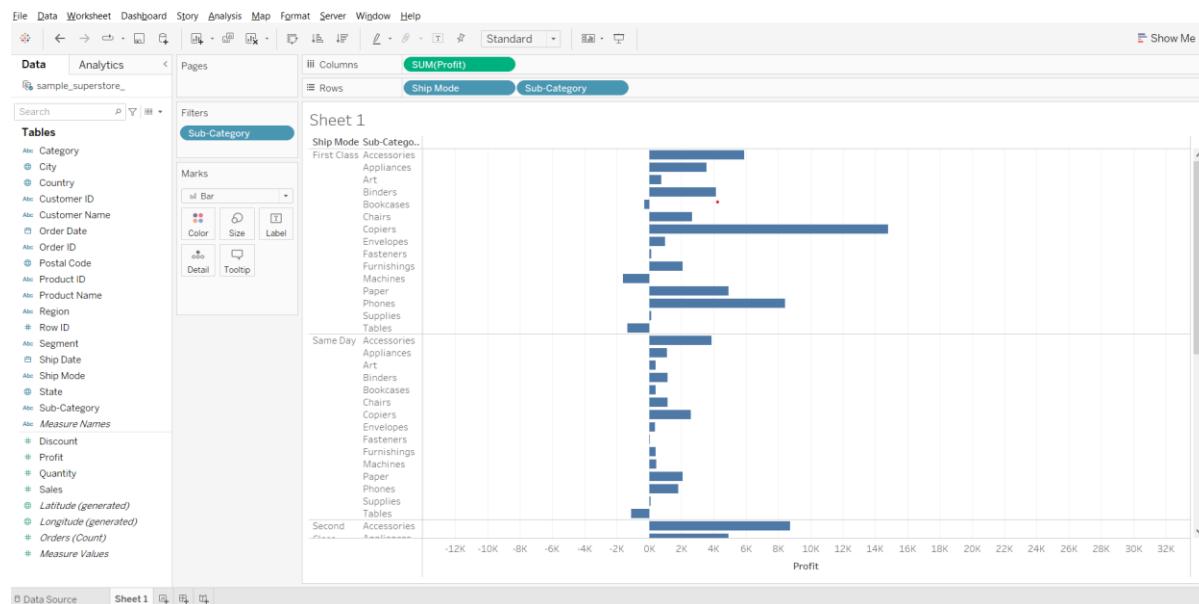
4. Filters on dimensions (whether on the Filters shelf or in filter cards in the view)

- Dimension filters in Tableau are non-aggregated filters. The dimensions that are used are mostly blue pills. Blue pills correspond to discrete data.
- If there are many dimensions, one can search for them. Dimension filter provides four options, General, Wildcard, Condition, and Top/Bottom. You can pick up any of the four options to select the right data or remove the unwanted data.
- One can create their own formula as well and then use it in the Condition filter and the Top/Bottom filter for data selection. They provide a channel to measure to get the required data.

Business Problem 8:

Show ship mode and subcategories wrt profit where subcategories Labels and Storage are excluded.

- Drag Ship Mode to Rows
- Drag Sub-Category to Rows
- Drag Profit to Columns
- Drag Sub-Categories to Filters card
- Deselect all boxes and select Labels and Storage
- Check the exclude box and click OK.



5. Filters on measures (whether on the Filters shelf or in filter cards in the view)

- Using a Measure filter in Tableau allows for various operations and aggregate functions such as sum, median, avg, standard deviation, etc.
- Aggregated filters are always applied after non-aggregated filters, no matter what the order is on the Filters pane.

- The filters are applied to Measure fields consisting of quantitative data.

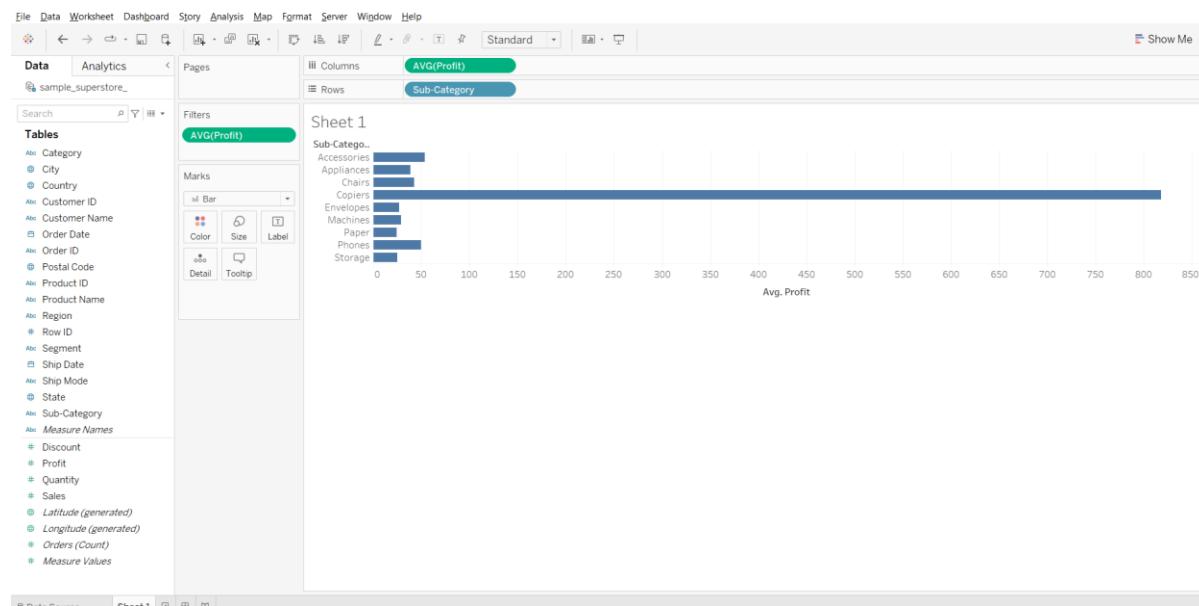
In a subsequent dialog box, you will get four types of filters:

- Range: Select the range of values to include in the result
- At least: Select the minimum value of a measure
- At most: Select the maximum value of a measure
- Special: Select null or non-null values

Business Problem 9 :

Show only the subcategories whose average profit is greater than 20.

- Drag Profit to Columns
- From the Profit drop down change measure to AVG.
- Drag Sub-Category to Rows
- From the Profit dropdown open Filter dialog box
- Change the minimum value to 20 and click OK.



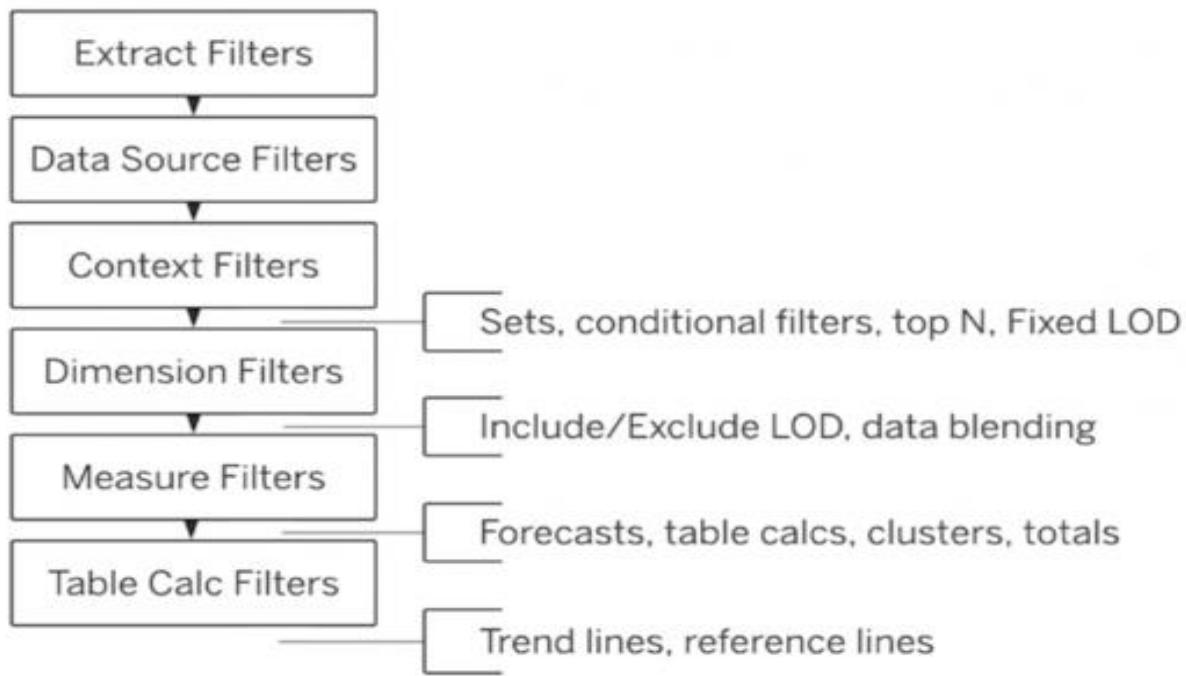
Filtering Order of Operations

Tableau performs actions on our view in a very specific order; this is called the Order of Operations. It helps you avoid filter conflicts and achieve efficiency with your dashboard.

Filters are executed in the following order:

- Extract filters
- Data source filters
- Context filters
- Filters on dimensions (whether on the Filters shelf or in filter cards in the view)

5. Filters on measures (whether on the Filters shelf or in filter cards in the view)
6. The Table Calculation filter



Calculated Fields

- Calculated fields allow you to create new data from data that already exists in your data source.
- When you create a calculated field, you are essentially creating a new field (or column) in your data source, the values or members of which are determined by a calculation that you control.
- This new calculated field is saved to your data source in Tableau, and can be used to create more robust visualizations.
- But don't worry: your original data remains untouched.

You can use calculated fields for many, many reasons. Some examples might include:

- To segment data
- To convert the data type of a field, such as converting a string to a date.
- To aggregate data
- To filter results
- To calculate ratios

Formatting Calculations in Tableau

There are four basic components to calculations in Tableau:

- Functions - Statements used to transform the values or members in a field.
- Fields - Dimensions or measures (columns) from your data source.
- Operators - Symbols that denote an operation.
- Literal expressions - Constant values that are represented “as is”, such as “Profitable” and “Unprofitable”

Note: The syntax (how to format them to work in Tableau) of each of these components will be discussed in class

Operators in Tableau

We have mainly 3 types of operators in Tableau.

- Arithmetic operator
 - Few types of arithmetic operator are addition, subtraction, division, multiplication, modulo
 - Example-profit+sales, profit/sales,
- Comparison operator
 - Few types of comparison operator are “==,>,<,>=,<=,!=”
 - Example-sales>100k, profit<=1k
- Logical operator
 - Few types of logical operator are “And, OR, Not”
 - Example-profit=100 AND sales=1000, Not(sales=profit)

Order of precedence of Operators : [Link](#)

Tableau Functions

Tableau supports many functions for use in Tableau calculations.

They have been organized category-wise below :

- [Number functions](#) : Number functions allow you to perform computations on the data values in your fields. Number functions can only be used with fields that contain numerical values.
- [String functions](#) : String functions allow you to manipulate string data (i.e. data made of text).
- [Date functions](#) : Date functions allow you to manipulate dates in your data source.
- [Type conversion](#) : Type conversion functions allow you to convert fields from one data type to another.
- [Logical functions](#) : Logical calculations allow you to determine if a certain condition is true or false (boolean logic).
- [Aggregate functions](#) : Aggregate functions allow you to summarize or change the granularity of your data.

- [User functions](#) : User functions can be used to create user filters or row-level security(RLS) filters that affect visualizations published to Tableau Server or Tableau Online, so that only certain people can see your visualization. Not available in Tableau Public.
- [Table calculations functions](#): Table calculation functions allow you to perform computations on values in a table.
- [Pass-Through functions \(RAWSQL\)](#) : These RAWSQL pass-through functions can be used to send SQL expressions directly to the database, without first being interpreted by Tableau. If you have custom database functions that Tableau doesn't know about, you can use these pass-through functions to call these custom functions. Not available in Tableau Public. [Link1](#) [Link2](#)
- [Spatial functions](#) : Spatial functions allow you to perform advanced spatial analysis and combine spatial files with data in other formats like text files or spreadsheets.
- [Additional functions](#) :
 - Regular Expressions
 - Hadoop Hive Specific Functions
 - Google BigQuery Specific Functions

Types of Calculations

There are three main types of calculations you can use to create calculated fields in Tableau:

- Basic expressions : They allow you to transform values or members at the :
 - Data source level of detail (a row-level calculation) or
 - At the visualization level of detail (an aggregate calculation).
- Level of Detail (LOD) expressions :
 - They too allow you to compute values at the data source and the visualization levels.
 - LOD expressions give you even more control (compared to basic expressions) on the level of granularity you want to compute.
 - They can be performed at a more granular level (INCLUDE), a less granular level (EXCLUDE), or an entirely independent level (FIXED).
- Table calculations : They allow you to transform values at the level of detail of the visualization only and are added to measures.

Basic expressions

Row level calculations:

Calculations are performed for every row of underlying data.

Business problem 10:

Find average cost of each product sub-categories (cost=Sales-Profit)

- Click on the drop down at top of left pane

- Choose Create Calculated Field...
- Enter the formula “[Sales] - [Profit]”
- Rename the field as Cost and click ok.

The screenshot shows the Tableau Data Editor interface. On the left, the 'Tables' pane lists various dimensions and measures. In the center, the 'Marks' card is selected, showing options for Bar, Color, Size, and Label. A tooltip for 'Label' is open, showing the formula '[Sales] - [Profit]' and a preview of the calculated field 'Cost'. A callout box highlights the 'Cost' field in the preview. On the right, the 'Sheet 1' canvas has a placeholder 'Drop field here'.

- We can verify that it is done for each row by clicking on view data (near the search bar in left pane) we can see a new column added “cost”

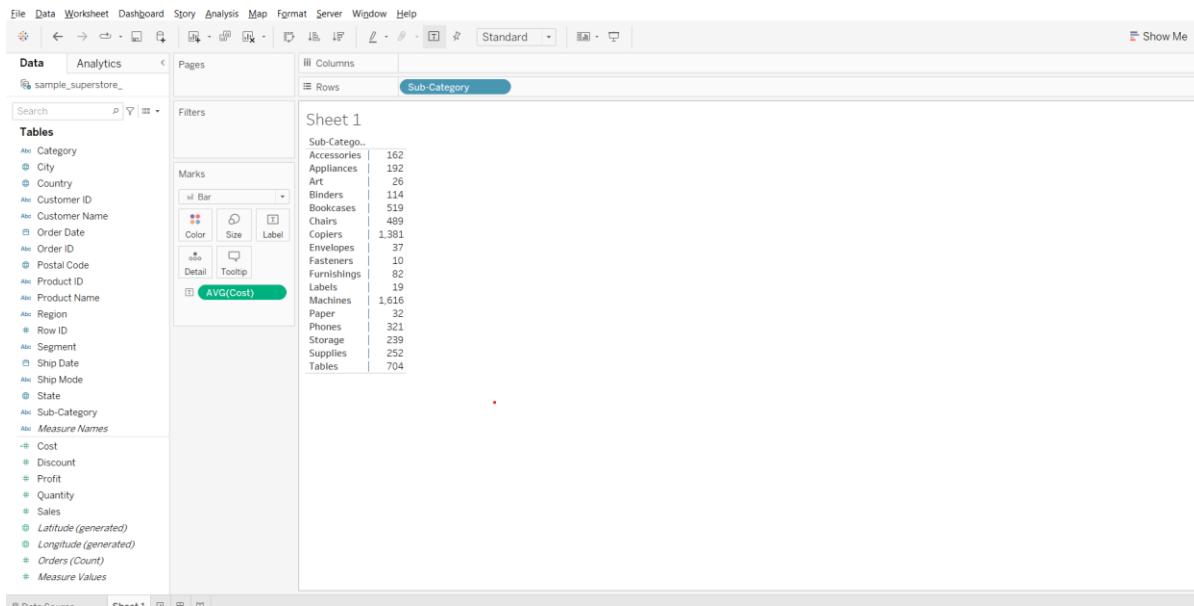
The screenshot shows the 'View Data' window for the 'sample_superstore' data source. The 'Tables' tab is selected, showing the 'Orders' table with columns: Category, Orders, Cost, Discount, Profit, Orders, Quantity, and Sales. The 'Cost' column is highlighted with a green border. The data grid displays several rows of order details. At the bottom, there are buttons for '10,000 rows' and 'Show Fields'.

Note: Subtraction arithmetic operator is used in this calculation.

Solving the business problem:

- Drag Sub-Category to Rows
- Drag Cost to Label in Marks card

- Change the measure of Cost to AVG using its dropdown menu



Aggregation calculations:

Calculations are performed at an aggregate level which is defined by the dimension used in the view.

Business problem 11

Find profit ratio of each product sub-category. Profit ratio = profit / sales

- Open the calculated field dialog box
- Enter the formula - $\text{SUM}([\text{Profit}])/\text{SUM}([\text{Sales}])$
- Name the new field as Profit Ratio and click ok.

The screenshot shows the Tableau Data Editor interface. On the left, the 'Tables' pane lists various dimensions and measures. In the center, the 'Sheet 1' canvas has a single row labeled 'Drop field here'. A tooltip window is open over the row, titled 'Profit Ratio', containing the formula `SUM({[Profit]}) / SUM({[Sales]})`. To the right of the formula is a help card for the `SUM(expression)` function, which states: 'Returns the sum of all the values in the expression. SUM can be used with numeric fields only. Null values are ignored.' Below the help card are examples like `SUM([Profit])`.

Note: Division arithmetic operator is used in this calculation.

Solving the business problem:

- Drag Sub-Category to Rows
- Drag Profit to the labels column next to Sub-Category in the sheet
- Drag Sales to the Profit column in the sheet
- Drag Profit Ratio to the Sales column
- From the Measure Values card change the format of Profit Ratio to Percentage

The screenshot shows the Tableau Data Editor with the 'Format AGG(Profit Ratio)' card open. The 'Measure Names' tab is selected, showing the field 'Sub-Category'. The 'Rows' section is set to 'Sub-Category'. The 'Measure Values' section contains three items: 'SUM(Profit)', 'SUM(Sales)', and 'AGG(Profit Ratio)'. The 'AGG(Profit Ratio)' item is highlighted. The main canvas displays a table with columns: Sub-Cat., Profit, Sales, and Profit Ra.. The data includes rows for Accessories, Appliances, Art, Blinders, Bookcases, Chairs, Copiers, Envelopes, Fasteners, Furnishings, Labels, Machines, Paper, Phones, Storage, Supplies, and Tables. The 'Profit Ra..' column shows values like 25.05%, 16.87%, 24.07%, etc. At the bottom of the screen, there are tabs for 'Data Source', 'Sheet 1', and other sheet icons.

Note:

- Notice AGG in pill profit ratio refer here the SUM is performed 1st based on the dimension sub category (i.e. sum of profit and sum of sales for each sub category) and then the division occurs.
 - Now if we change the dimension from sub category to category the calculation will be done based on it.
-

Tableau Lecture 5: Level of Detail (LoD) calculation

Agenda

- Using parameters to make visualizations more interactive
- Data aggregation and granularity
- Level of detail calculation

Dataset: [sample superstore](#)

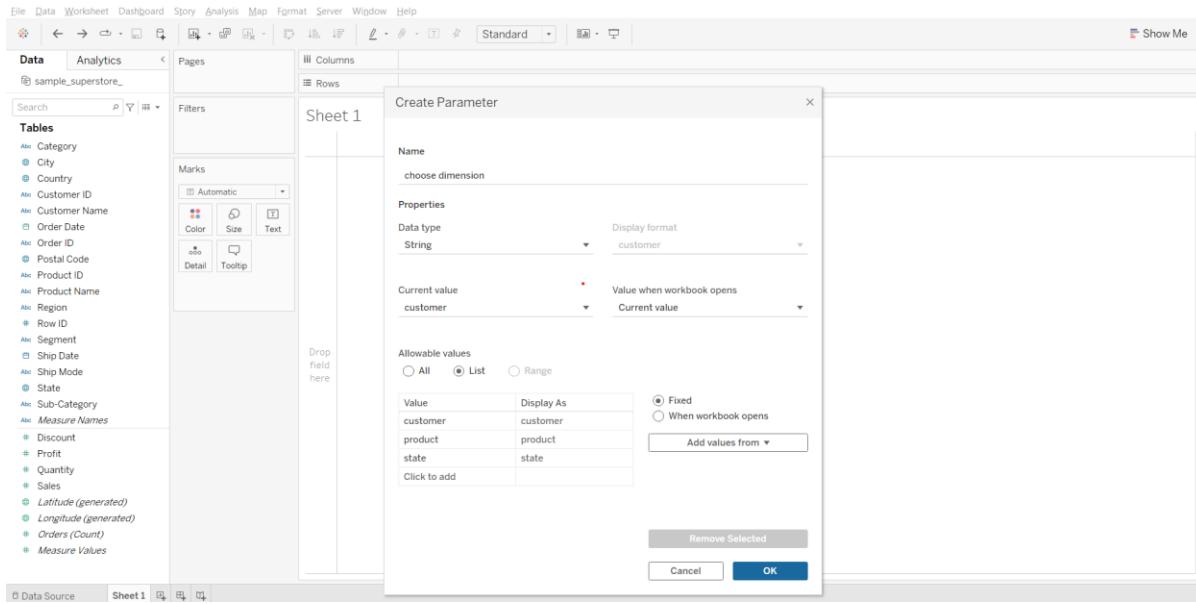
Using parameters to make visualizations more interactive

- Parameters are useful when you want to add interactivity and flexibility to a report, or to experiment with what-if scenarios.
- Suppose you are unsure which fields to include in your view or which layout would work best for your viewers.
- You can incorporate parameters into your view to let viewers choose how they want to look at the data.

Business problem 1:

Allow users to select the field against which they want to compare the sales value.

1. Create parameters where we display the fields that they can have on the y axis
 - Open the dropdown menu at the top of the left pane, click on Create Parameter...
 - Enter parameter name as 'choose parameter'
 - Select data type as string, Allowable values as list
 - Enter customer, product and state in the list of values section. Click OK.



- Creating a calculated field that will help us select the correct dimension based on the parameter value selected

Formula:

```
CASE [choose dimension]
WHEN 'customer'
THEN [Customer Name]
WHEN 'product'
THEN [Product Name]
WHEN 'state'
THEN [State]
END
```

Note: The strings in the above formula have to match the string that we had entered when creating the parameter otherwise it won't work

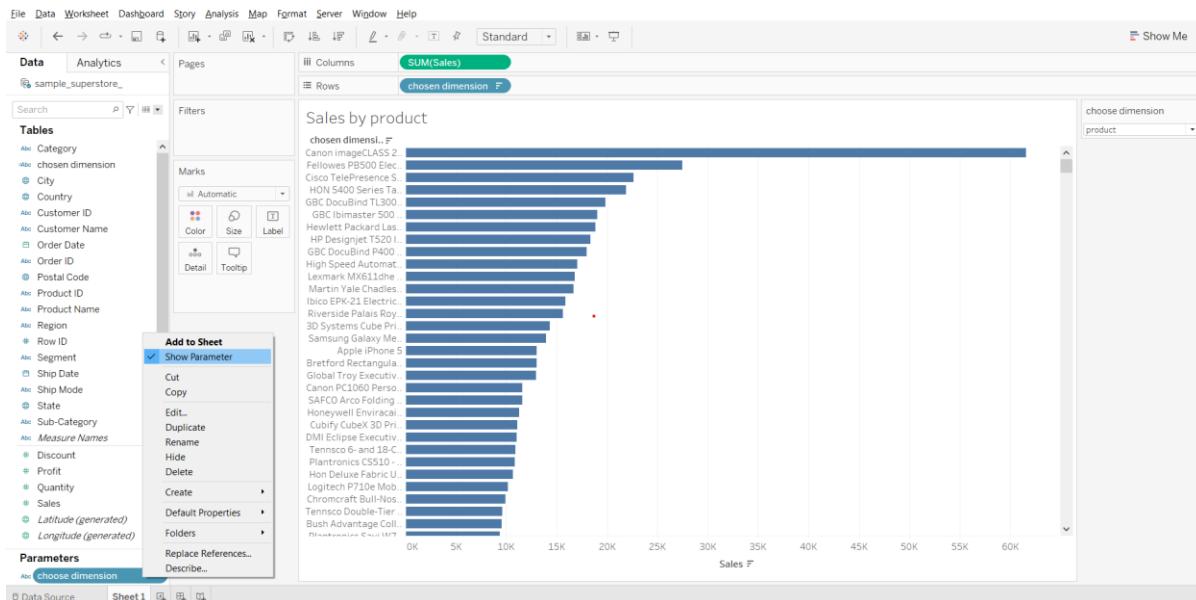
The screenshot shows the Tableau Data Editor interface. On the left, the 'Tables' and 'Parameters' panes are visible. The 'Parameters' pane contains a single parameter named 'choose dimension'. In the center, a sheet titled 'Sheet 1' is displayed with a tooltip over the 'chosen dimension' field. The tooltip shows the following code:

```
CASE [choose dimension]
WHEN 'customer'
THEN [Customer Name]
WHEN 'product'
THEN [Product Name]
WHEN 'state'
THEN [State]
END
```

The tooltip also includes a note: 'The calculation is valid.' and buttons for 'Apply' and 'OK'. To the right of the tooltip, a dropdown menu is open, showing various mathematical functions like ABS, ACOS, AND, AREA, ASCII, ASIN, ATAN, ATAN2, and ATTR. The 'ABS (number)' function is highlighted.

3. Solving the business problem:

- Drag 'chosen dimension' to Rows
- Drag Sales to Columns
- Right click choose dimension parameter at the bottom of left pane and click show parameter.
- Change sheet name to “Sales by <Parameters.choose dimension>” by using insert parameter option.
- Dimension can now be chosen from the right pane.

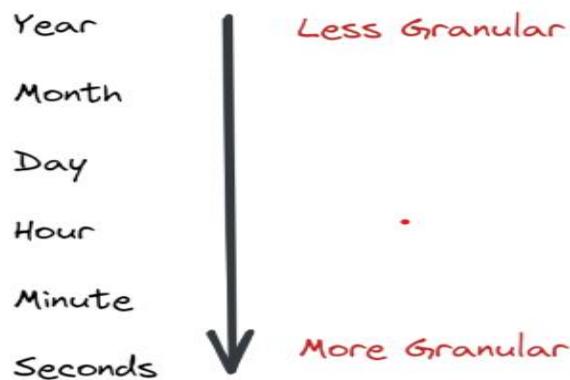


What is Data Granularity?

- Data granularity is a measure of the level of detail in a data structure.
- Granularity in simple terms is how detailed the data is.
 - The more detailed, the higher the granularity. (Hence the more rows of data you will have)
 - The less detailed, the lower the granularity. (Vice versa, the less number of rows of data you will have)

Example:

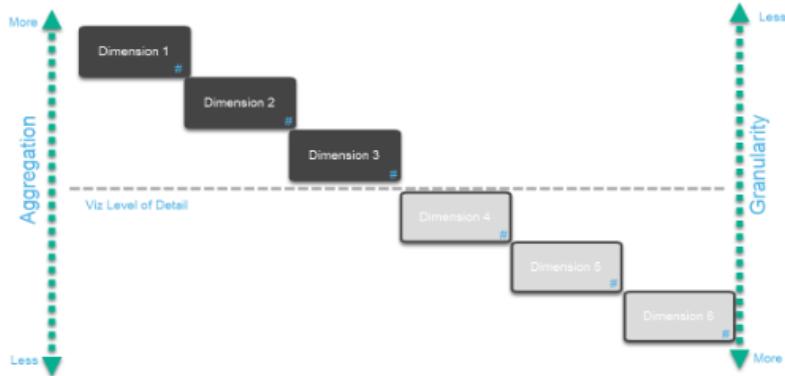
- In time-series data, for example, the granularity of measurement might be based on intervals of years, months, weeks, days, or hours.
- The name field could represent the full name or have separate entries for first name, middle name, and last name.



YouTube video for better understanding: [link](#)

Data Granularity vs Aggregation

- Granularity and aggregation work opposite of each other.
- In Tableau, when you bring dimensions and measure to the view, the measures are aggregated by whatever dimensions are on the view.
- As you remove and add more dimensions you are decreasing and increasing the granularity. The more dimensions/details on the view, the more granular it is.



In this example, we have 3 dimensions out on the view so your viz level of detail is that line shown in the photo above. Every measure on your viz will be aggregated at that level which means it will be aggregated by the combination of those 3 dimensions.

Why is Data Granularity Important?

The level of data granularity determines what analysis can be performed on the data, and whether results from that analysis lead to appropriate conclusions.

The more granularity, the more information is available for analysis, but at the cost of increased storage, memory, and required computing resources.

Some analysis may require information to be analyzed at a higher level, which would require aggregating the underlying detail into the higher level of granularity; this is where Level of details works magic.

Fixed Level of Detail

It computes a value using the specified dimension without reference to the dimension in the view.

The FIXED LOD can specify the granularity of the calculation independently of what is in the view.

This is what the syntax looks like:

```
{ keyword [dimension 1], [dimension 2]: aggregate expression }
```

Business Problem 2:

Find the percentage sales contribution of each sub-category to the total sales of each product category.

1. Getting the sum of sales for each product subcategory
 - o Drag sub-category to Rows

- Drag sales to Text in Marks card

The screenshot shows the Tableau Data Editor interface. On the left, there's a sidebar with 'Tables' and 'Parameters' sections. In the main area, a 'Marks' card is open, showing a list of sub-categories and their sales values. A tooltip for the 'fixed sum of sales' calculation is visible.

Sub-Catego...	Sales
Accessories	167,380
Appliances	107,532
Art	27,119
Binders	203,413
Bookcases	114,880
Chains	328,449
Copiers	149,539
Envelopes	16,476
Fasteners	3,024
Furnishings	91,705
Labels	12,486
Machines	189,239
Paper	78,479
Phones	330,007
Storage	223,844
Supplies	46,674
Tables	206,966

2. Creating fixed LoD calculation

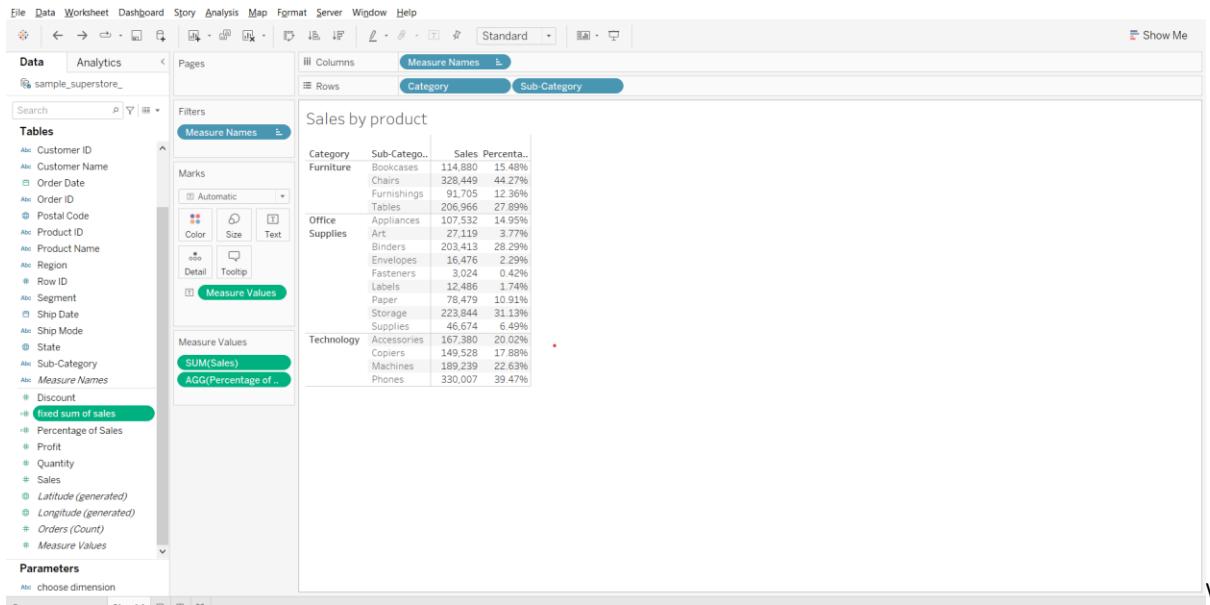
Calculation formula- { FIXED [Category]:SUM([Sales])}

The screenshot shows the Tableau Data Editor interface. A tooltip for the 'fixed sum of sales' calculation is visible, stating 'The calculation is valid.' A dropdown menu for 'Category' is open, showing various string functions like ABS, ACOS, AND, AREA, ASCII, ASIN, ATAN, ATAN2, and ATTR.

3. Solving the business problem:

Calculation formula- SUM([Sales])/SUM([fixed sum of sales])

- Create a calculated field 'Percentage of Sales' using the above formula.
- Drag 'percentage of sales' to the labels column in the sheet.
- Change format of 'Percentage of Sales' to percentage, from the Measure Values card.
- Drag Category to Rows shelf keeping it to the left of sub-category.



Include Level of Detail

It computes the value using the specified dimension in addition to whatever dimension are in the view.

You can translate INCLUDE LOD as 'For every dimension in the view AND every listed dimension, calculate the aggregate expression':

This is what the syntax looks like:

```
{ include [Product] : sum([Sales])}
```

For every dimension in
the view AND Product Compute
sum of sales

Business problem 3:

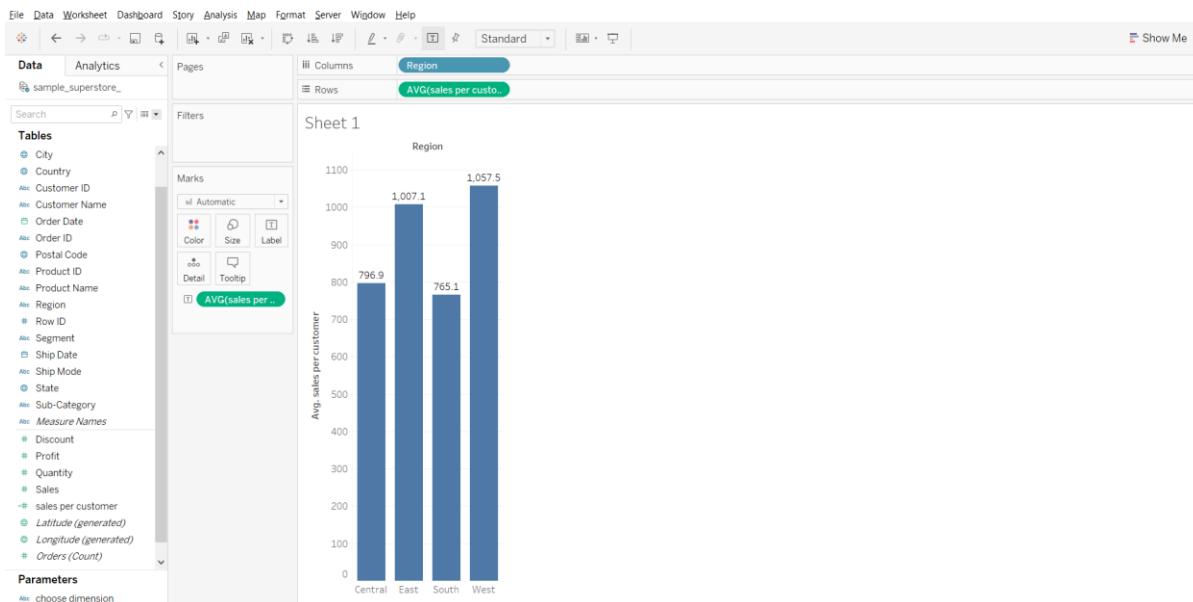
Find the difference between the total sales for each region and average sales per customer for each region.

1. Creating a include LoD calculation by customer name
Calculation formula- { INCLUDE [Customer Name]: SUM([Sales])}

The screenshot shows the Tableau Data Editor interface. On the left, the 'Tables' shelf lists various dimensions and measures. In the center, the 'Marks' card is set to 'Automatic'. A tooltip for the 'Text' icon is open, showing a calculated field definition: { INCLUDE [Customer Name]:SUM([Sales]) }. To the right, a context menu for the 'Text' icon is displayed, with 'ABS(number)' selected. The menu includes a search bar, a list of functions (ABS, ACOS, AND, AREA, ASCII, ASIN, ATAN, ATAN2, ATTR), and an example: ABS(-7) = 7. Below the menu, a message says 'The calculation is valid.'

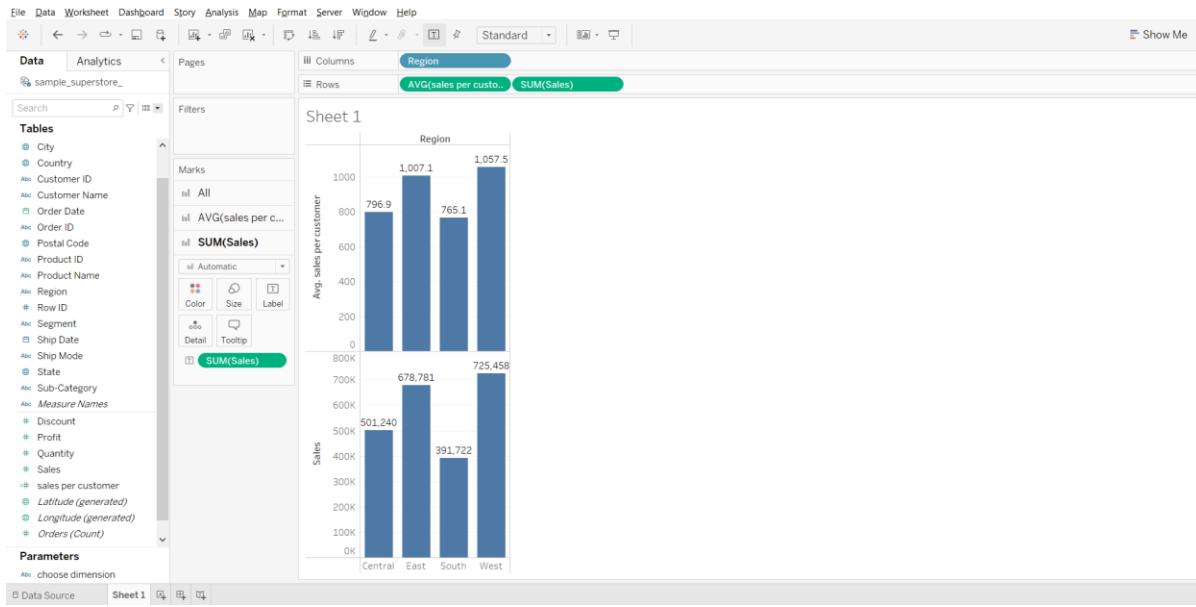
2. Plotting average sales per customer by region

- Drag Region to Columns shelf
- Drag sales per customer to Rows shelf
- Change measure type of sales per customer to AVG in Rows shelf
- Drag sales per customer to text in Marks card
- Change measure of sales per customer to AVG in Marks card.



3. Solving the business problem:

- Drag sales to Rows shelf
- Click on sales in marks card
- Remove AVG(sales per customer) labels
- Drag sales to labels in sales marks card



Exclude Level of Detail

It is pretty much the opposite of INCLUDE. Instead of adding more dimensions, you're getting rid of them, i.e. we declare dimension to omit from the view level of detail.

You can translate EXCLUDE LoD as 'For every dimension in the view EXCEPT the listed dimension(s), calculate the aggregate expression':

This is what syntax looks like:

```
{ exclude [Sub-Category] : sum([Sales])}
  For every dimension in the
  view EXCEPT Sub-Category      Compute
                                sum of sales
```

Business problem 4:

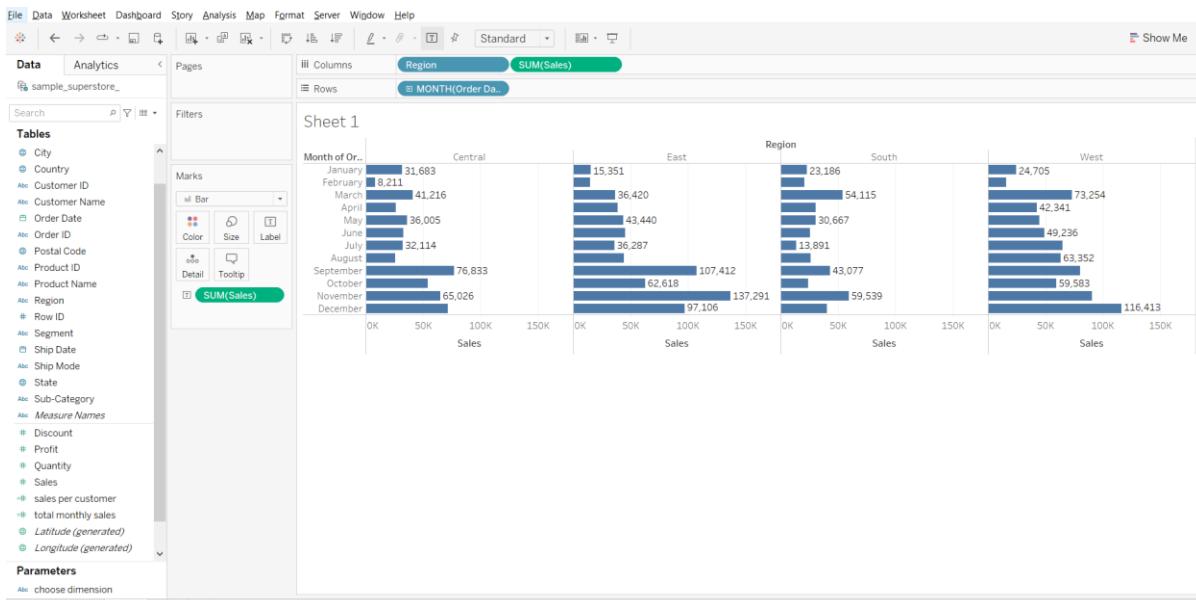
Find the difference between sales for each region by month and total sales for each month.

1. Creating exclude calculation
 - o Calculation formula - { EXCLUDE [Region]:SUM([Sales])}

The screenshot shows the Tableau Data Editor interface. On the left, the 'Tables' pane lists various dimensions and measures from the 'sample_superstore' data source. In the center, a tooltip for the 'ABS' function is displayed, showing its definition: 'Returns the absolute value of the given number.' Below the tooltip, a message says 'The calculation is valid.' At the bottom right of the tooltip are 'Apply' and 'OK' buttons.

2. Plotting sales by month for each region

- Drag order date to rows shelf
- From the order date drop down in the rows shelf, choose Month.
- Drag Region to Columns shelf
- Drag sales to Columns shelf
- Change graph type to bar
- Drag sales to label in marks card



3. Using exclude calculation in color marks card to get total sales done for each month

- Drag total monthly sales to color in Marks card

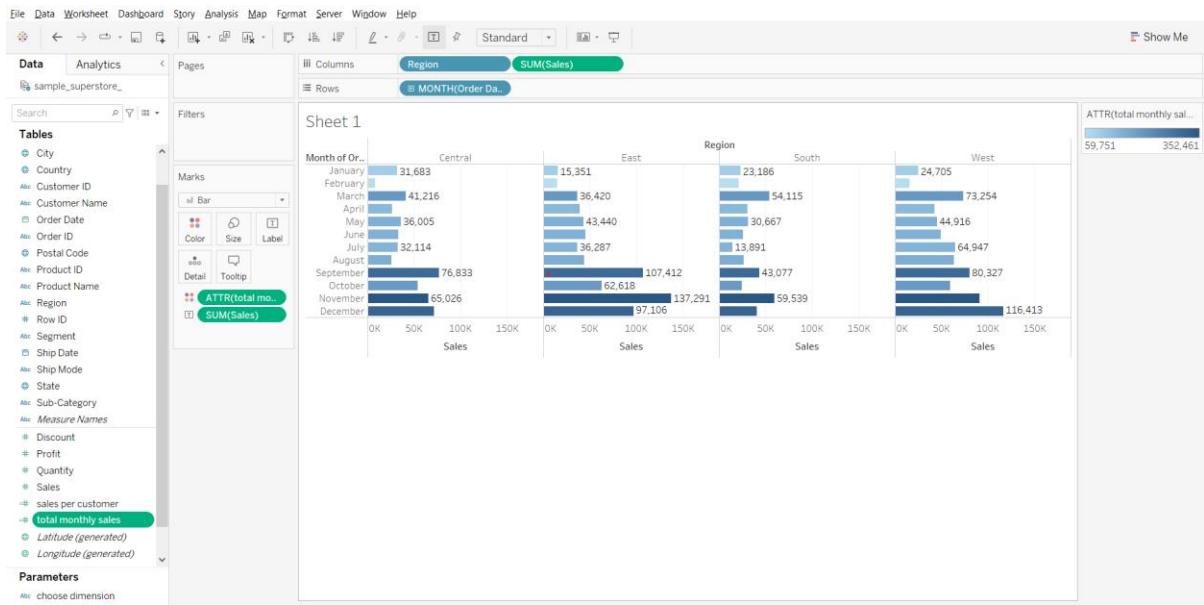


Tableau Lecture 6: Table Calculation and Analytics

Agenda

- Table Calculations
- Tooltip
- Reference Lines
- Trend Lines

Dataset: [sample superstore](#)

Table Calculations

- Table calculations are a special type of calculated field within Tableau that apply transformations (i.e. additional math) on values within a visualization.
- When you add a table calculation, you must use all dimensions in the level of detail either for partitioning (scoping) or for addressing (direction).
- Scope- it defines the boundaries within which a given table calculations can reference other values
 - Scope options-Table, pane, cell
- Direction- defines how the table calculations moves within the scope
 - Direction options-down, across, down then across, across then down
- Index-Returns the index of the current row in the partition

Index Problem: Table calculation using Index to explain the concept.

1. Create a calculated field index as shown below

The screenshot shows the Tableau Data Editor interface. On the left, there's a sidebar with 'Tables' and 'Measures' sections. The main area is 'Sheet 1'. In the 'Marks' card, the 'index' field is selected. A tooltip for 'INDEX()' is displayed, stating: 'Returns the index of the current row in the partition.' Below the tooltip, it says 'Example (for the first row in the partition): INDEX() = 1'. At the bottom right of the Marks card, there are 'Default Table Calculation' buttons for 'Apply' and 'OK'.

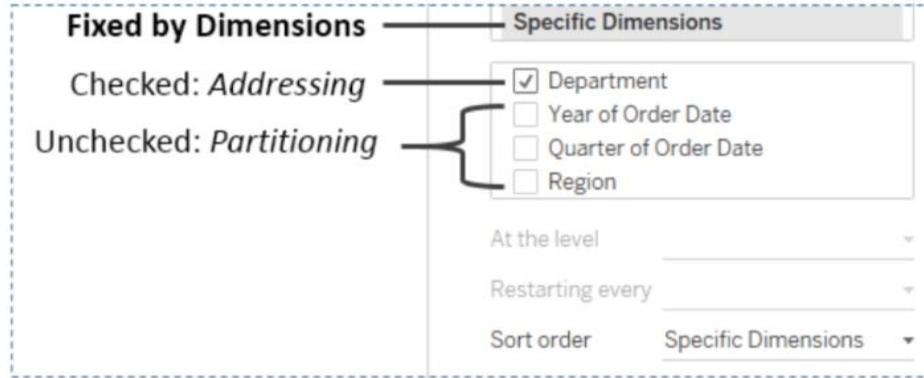
2. Explain types of “compute using” options using index calculation

- Drag Order Date to Rows shelf
- Click on the small plus sign to get the quarters for each year
- Drag Region and Category to Columns shelf
- Drag index to label in Marks card
- Open the index drop down in the Marks card and try different ‘compute using’ options.

The screenshot shows the Tableau Data Editor interface. The 'Marks' card has 'index' selected. A context menu is open over the 'Compute Using' dropdown, listing options: 'Table (across)', 'Table (down)', 'Table (across then down)', 'Table (down then across)', 'Pane (across)', 'Pane (down)', 'Pane (across then down)', 'Pane (down then across)', 'Cell', 'Category', 'Order Date', and 'Region'. 'Table (across)' is currently selected.

Direction and Scope

- Here we define which dimension in the view define direction(addressing) and all others define the scope(partitioning).



Demo steps:

- After performing the steps for explaining ‘compute using’, open the index dropdown in the menu card and select ‘Edit Table Calculation...’

We can notice that tableau is computing the index in the direction of the checked dimension category thus each new category increments the index.

Table calculations “compute using”

- Table across
- Table down
- Table across then down
- Table down then across
- Pane across
- Pane down
- Pane across then down
- Pane down then across

- Cell-computation is done within a single cell

Table Calculation Types

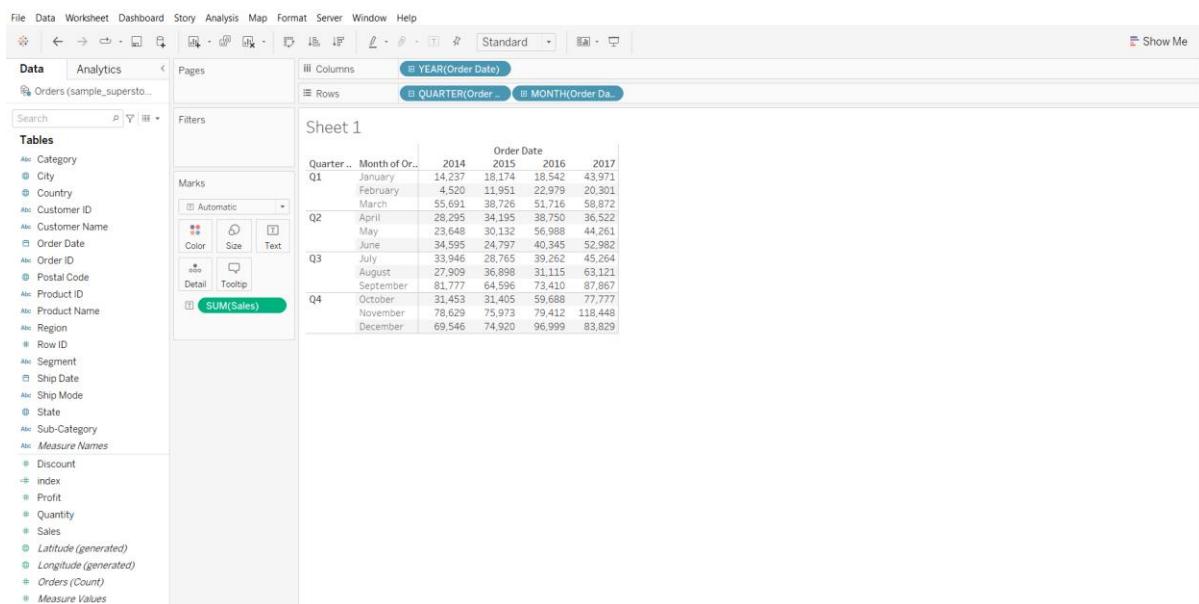
- Difference from - computes difference between current value and another value in the table
- Percentage difference from - computes difference between current value and another value in the table as a percentage
- Percentage from - computes a value as a percentage of some other value
- Percent of total - computes a value as a percentage of all values in the current partition
- Rank - computes a ranking for each value in a partition
- Percentile - computes a percentile rank for each value in a partition
- Running total - it aggregates values cumulative in a partition
- Moving calculation - it determines the value by performing aggregation across specified number of values before or after current value

Percent of total

Business problem 1:

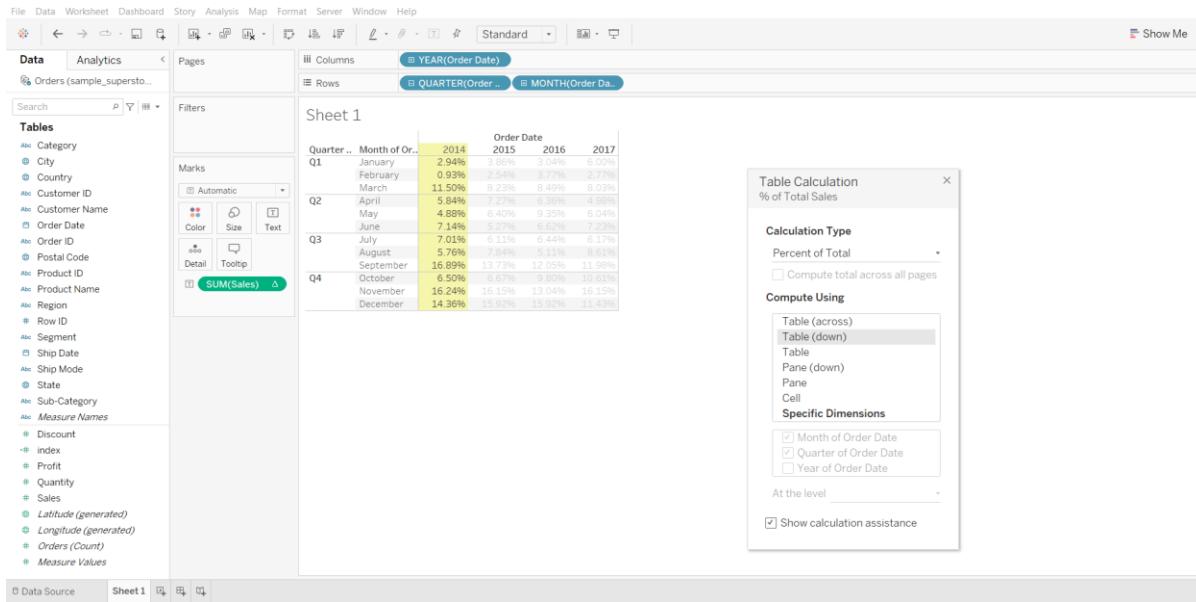
Find the percentage of total sales each month makes for each year.

1. Creating total sales per month for the year 2014 to 2017
 - Drag Order Date to Rows shelf
 - From dropdown of Order Date in Rows shelf, select Quarter
 - Click on the small plus on Quarter(Order Date) to get months
 - Drag Order Date to Columns shelf
 - Drag sales to labels in Marks card



2. Creating table calculation and solving business problem

- From the sales dropdown in the Marks card, click 'Add Table Calculation...'
- In calculation type choose 'Percent of total'
- In 'compute using', choose Table(down)



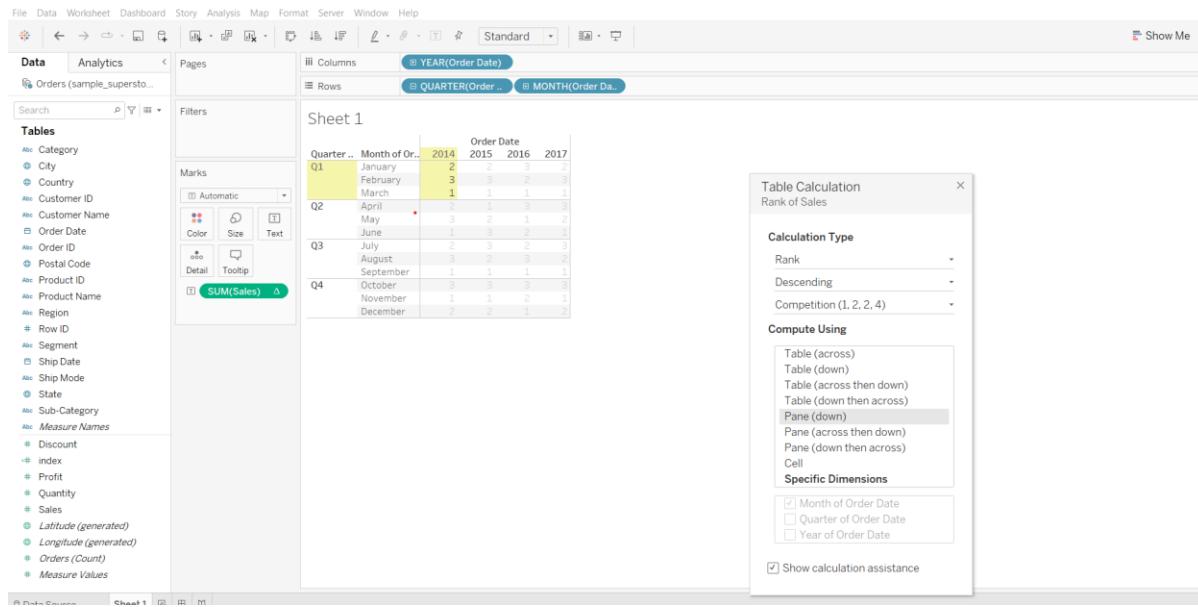
- We can see for the year 2017 highest sales was done in the month of november

Rank

For each year and for each quarter we are finding a **rank based on sales** so we can see below that for all the years in quarter1 March month had highest sales and hence it has rank=1

Demo steps:

- After performing the steps of the previous business problem, click 'Edit Table Calculation...' in the sales dropdown in the Marks card.
- Change calculation type to Rank
- Change 'compute using' to Pane(down)



Reference for [Table Calculation Types](#)

Table Calculations functions: Index vs Rank

RANK and INDEX are part of Table calculations in Tableau.

INDEX mainly deals with physical position of the record, incremental numbers are assigned based on the physical order of the records.

Example:

SUM(SALES)	INDEX
10000	1
40000	2
20000	3
30000	4

RANK mainly deals with the value of the record, highest value gets highest rank, lowest gets the lowest rank.

Example:

SUM(SALES)	RANK
10000	4
40000	1
20000	3
30000	2

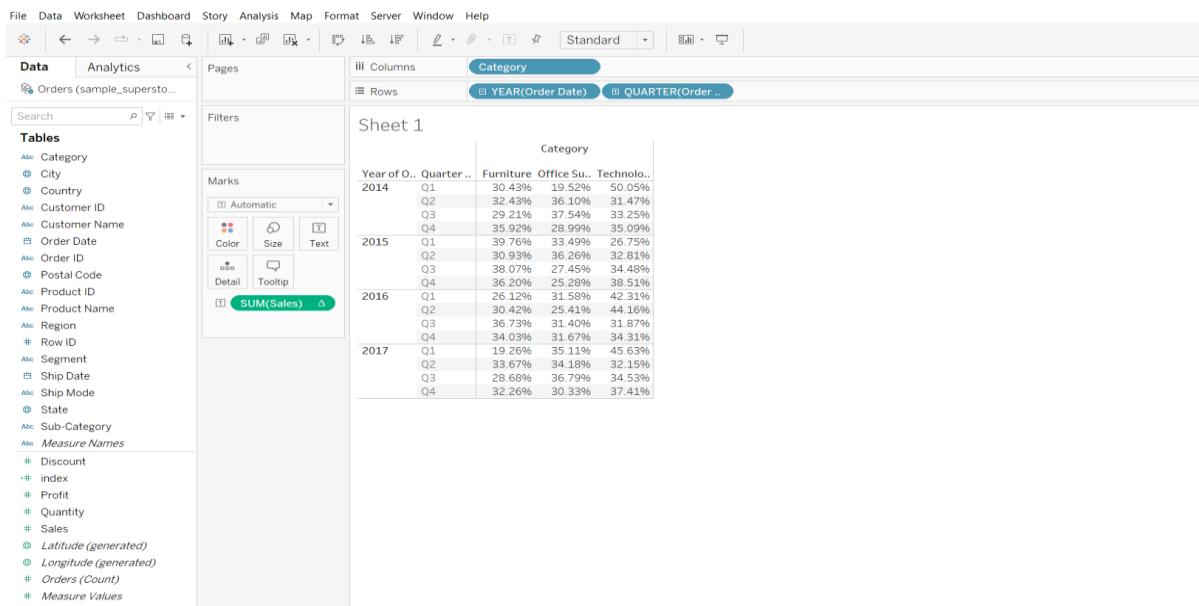
Quick table calculation

It allows us to quickly apply a common table calculation to your visualization

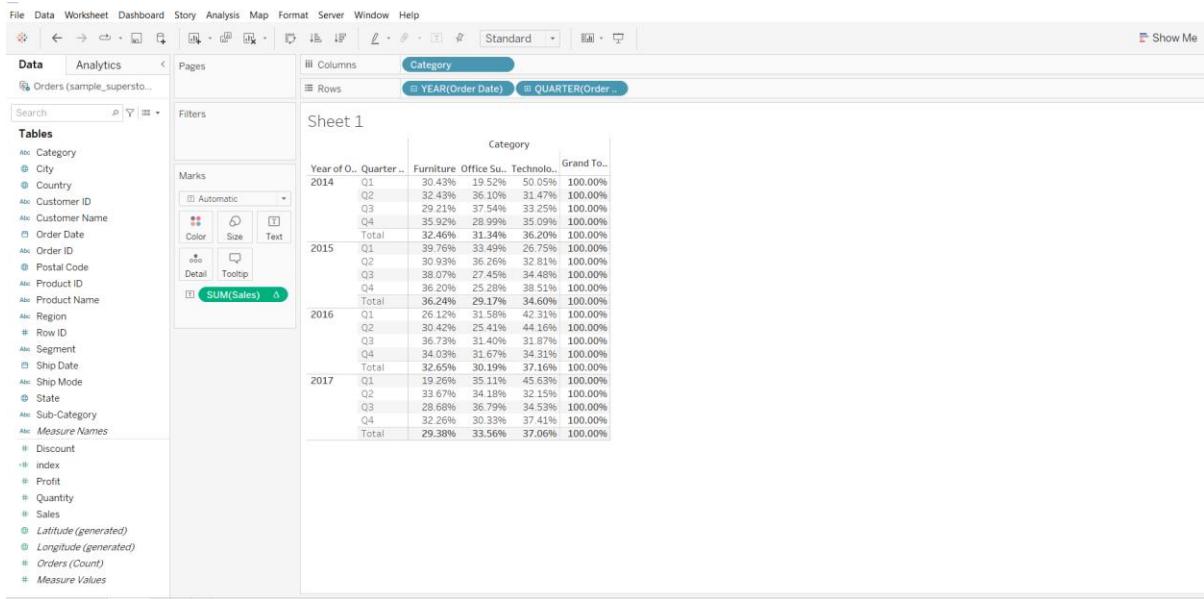
Business problem 2:

Find the percentage of total sales done across different categories for each quarter of the given year.

- Drag Order Date to Rows shelf
- Click on the small plus sign on Order Date to get Quarters
- Drag Category to Columns shelf
- Drag sales to label in Marks card
- From sales dropdown in Marks card, go to Quick Table Calculation and select ‘Percent of Total’



- Go to Analysis in menu bar, go to totals and select ‘Show Row Grand Totals’
- Again from Analysis menu, go to Total and select ‘Add All Subtotals’



- We can see that for the year 2016 for the 1st quarter 42.31% of total sales was done in technology category and for the year 2016 also the highest sales was done by technology category.

Difference between calculated field and table calculation

- Table calculations are created in a Tableau view and stay there locally in our worksheet
- Calculated fields are created on a data level and appear as a separate column in the data source

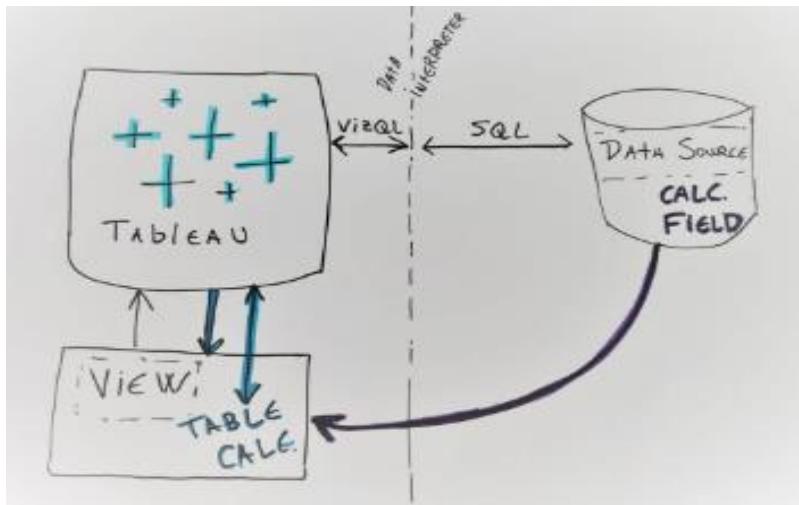


Tableau Functions post-read: [link](#)

Tooltip

Tooltips are additional data details that display when you hover over one or more marks in the view.

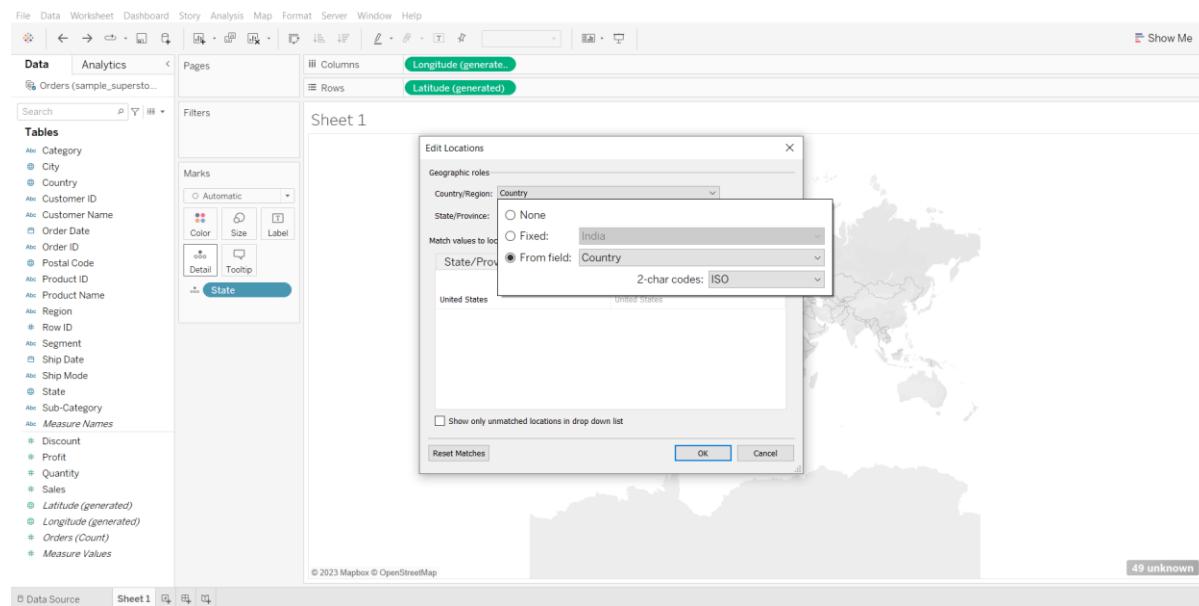
Tooltip command button for exploring data -

- Keep only- creates a filter that removes all the other data
- Exclude- creates a filter that removes the selected data
- Group members - creates a group based on the selection
- Create set- creating a new set containing the selected members
- View data- opens a window displaying the data

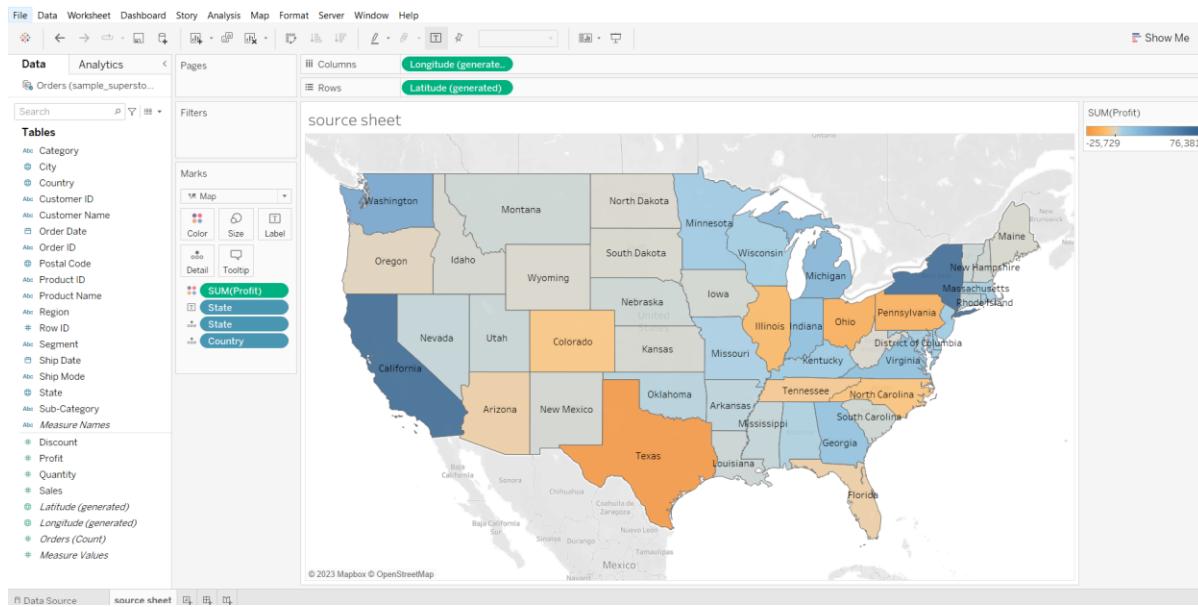
Business problem 3:

Create a map visualization that displays profit for each state and allows the user to view profit by subcategory for that state in the tooltip when they hover over a state.

1. Creating a source sheet which plots profit by each state in a map visualization
 - Drag Latitude to rows shelf
 - Drag Longitude to columns shelf
 - Drag State to Details in marks card
 - Click on the '49 unknown' box on the bottom right corner
 - Click on Edit Locations and change Country from India to 'From Field' Country.

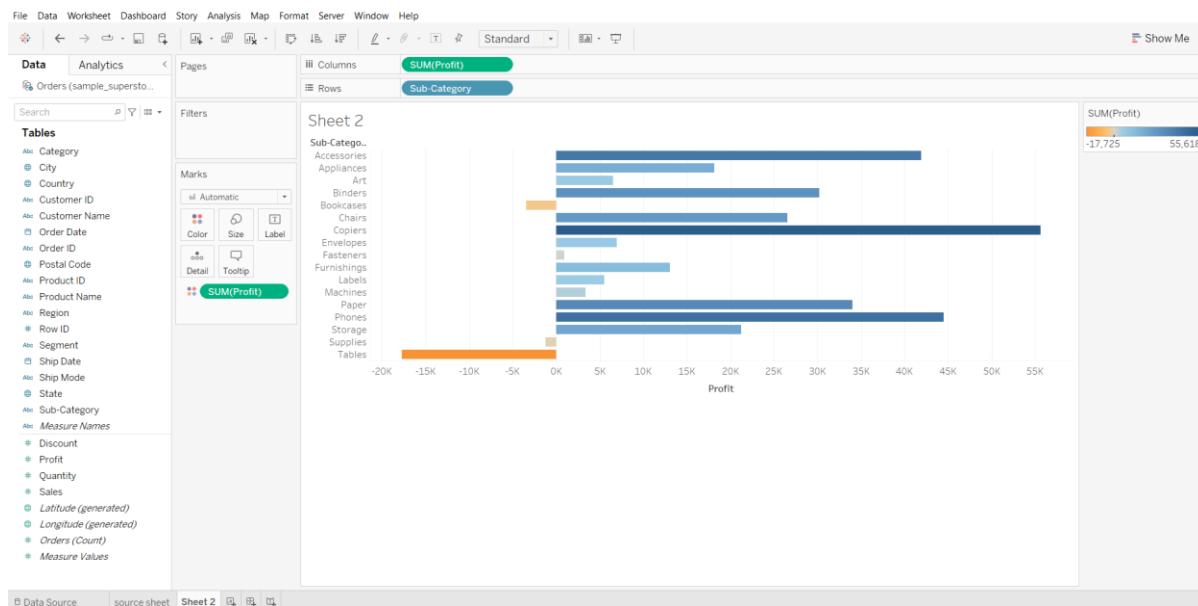


- Change graph type to map
- Drag Profit to color in marks card
- Drag State to label in marks card
- Change the sheet name to source sheet



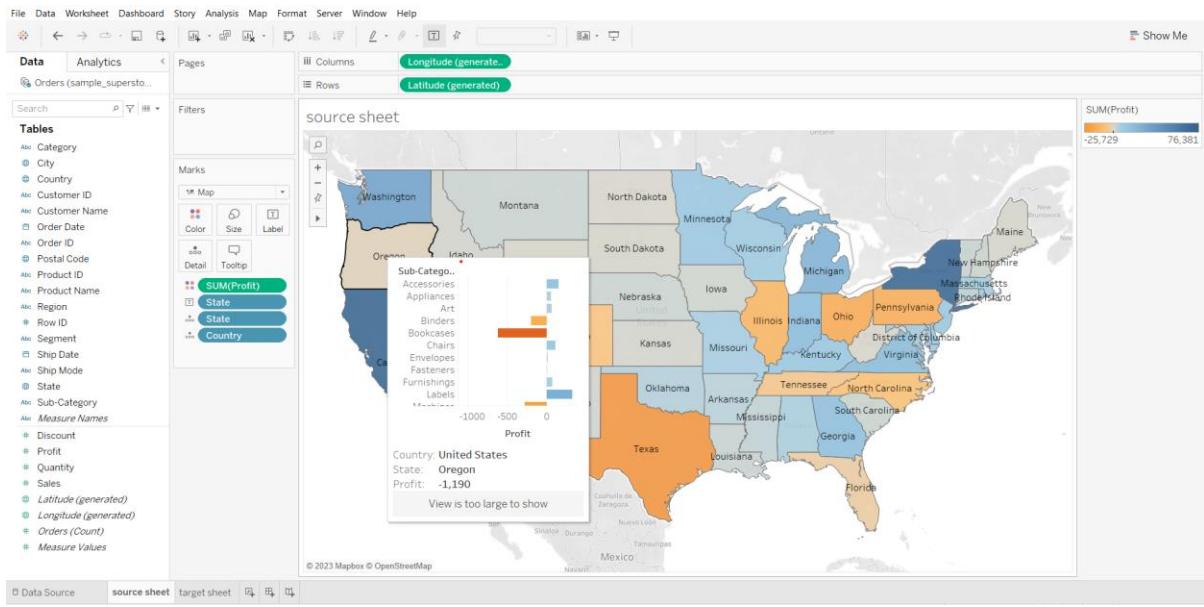
2. Creating a target sheet which plots subcategory wise profit

- Drag sub-category to rows shelf
- Drag Profit to Columns shelf
- Drag profit to color in marks card
- Change the sheet name to target sheet

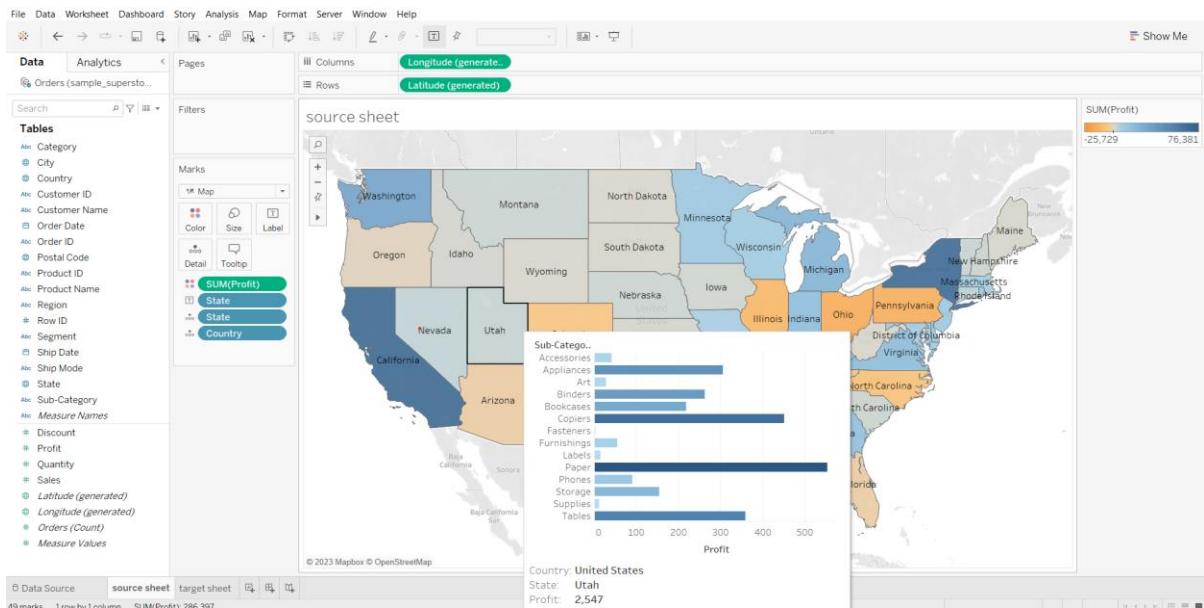


3. Adding target sheet visualization in the tooltip

- In the source sheet, click on tooltip in marks card
- Click on Insert > Sheet > target sheet
- Hover over different states to see the result



- We can see that most of the product subcategories are losing money in Texas
- #### 4. Editing the size of the tooltip visualization
- Click on tooltip in marks card
 - Change max-width and max-height to 500.



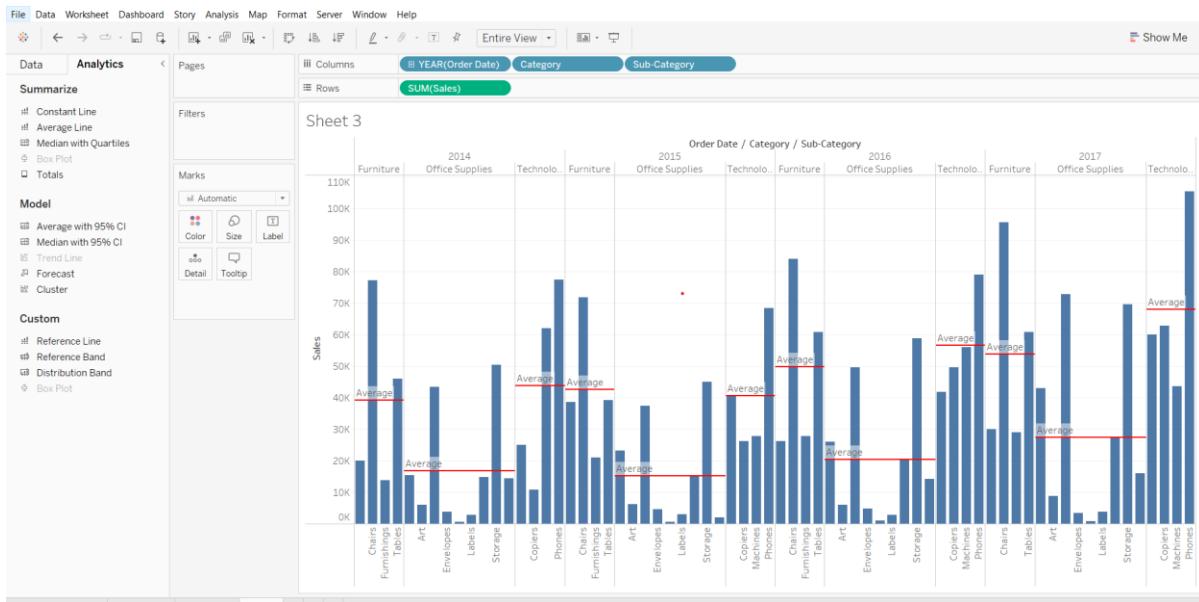
Reference Line

It is simply a line that gets drawn on a chart that represents another measure or point of reference

Business problem 4:

Find subcategories with above-average sales for each category and year.

- Drag Order Date, Category and sub-category to columns shelf
- Drag sales to rows shelf
- Go to the Analytics tab in the left pane
- Drag reference line to the pane option
- Choose color and opacity of line and click ok.

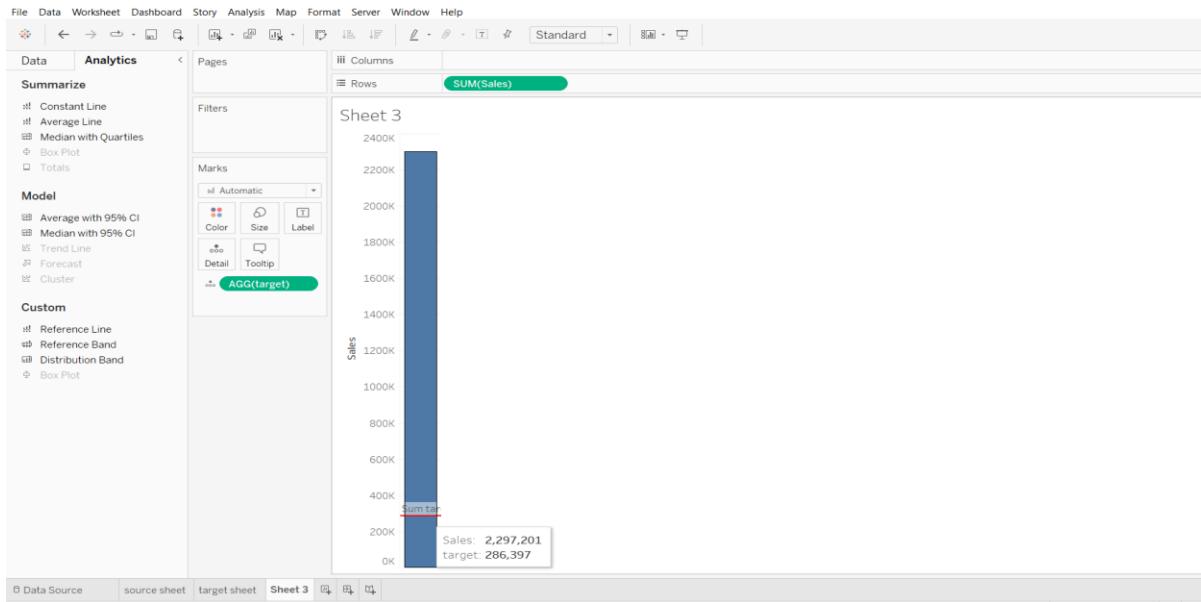


Controlling Reference Line with Parameters

Business problem 5:

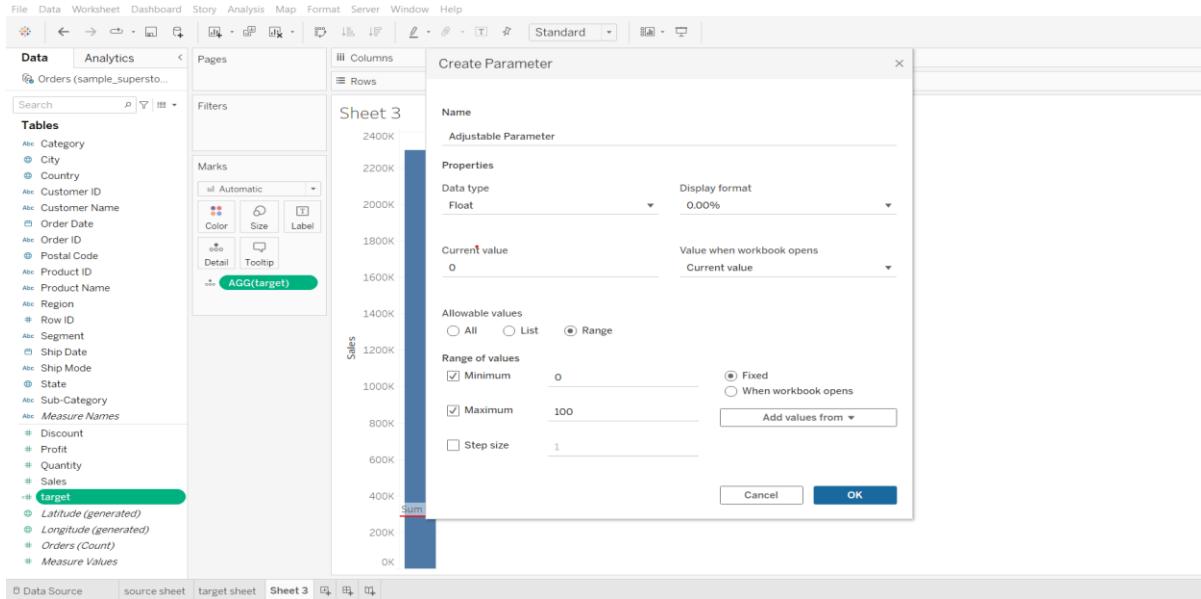
You want to decide the sales target for next year based on this year profit*target percentage with the flexibility to decide the target percentage.

1. Creating a basic view that displays total sales and a target field that will be displayed using a reference line in the view
 - Drag sales to rows shelf
 - Create a calculated field 'target' using calculation sum([Profit])
 - Drop target on detail in marks card
 - Go to the Analytics tab and create a 'table' reference line.
 - Choose value as AGG(target), measure as SUM, click OK.



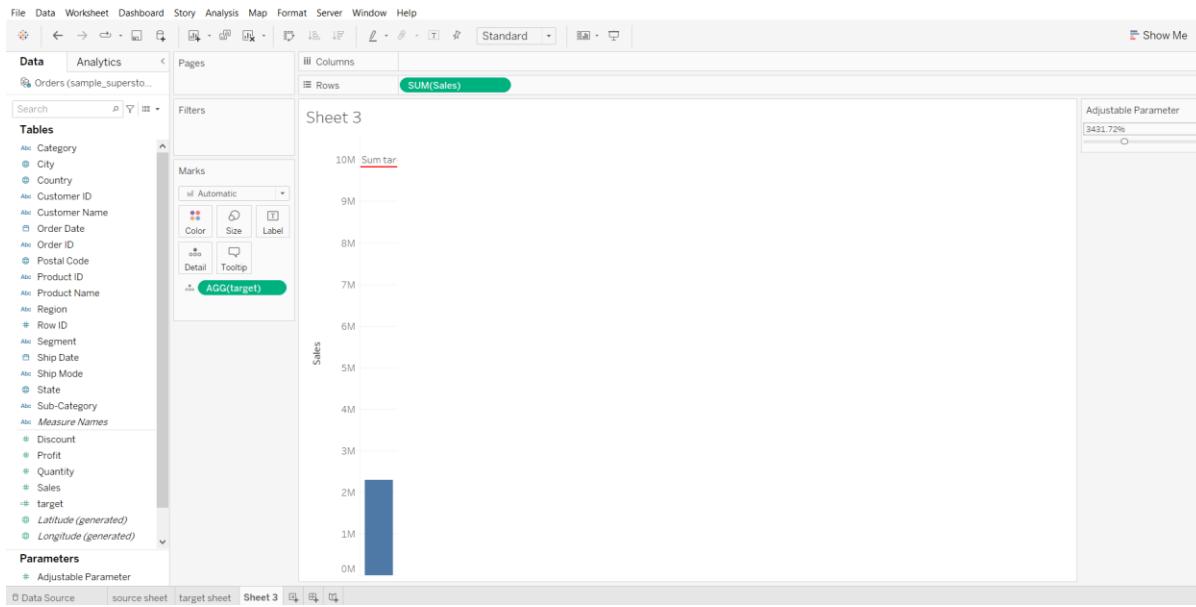
2. Creating a parameter

- From the target dropdown in the Data panel, choose Create > Parameter...
- Name the parameter 'Adjustable Parameter', set its minimum value to 0 and maximum value to 100, set current value to 0, display format to percentage. Click OK.



3. Tying up the parameter with the target field and solving the problem

- Change the formula of target to -> $\text{SUM}([\text{Sales}]) * [\text{Adjustable Parameter}]$
- From the dropdown of Adjustable Parameter click 'Show Parameter'
- Change the value Adjustable Parameter in the right pane to change the target.



Trend Lines

- Trend lines are used to predict the continuation of a certain trend of a variable.
- It also helps to identify the correlation between two variables by observing the trend in both of them simultaneously.
- There are many mathematical models for establishing trend lines. Tableau provides four options. They are Linear, Logarithmic, Exponential, and Polynomial
- Tableau takes a time dimension and a measure field to create a Trend Line.

Trends are computed by Tableau after the query of the data source and are based on various elements in the view:

- The two fields that define X and Y coordinates: The fields on Rows and Columns that define the x and y axes describe coordinates, allowing Tableau to calculate various trend models. In order to show trend lines, you must use a continuous (green) field on both Rows and Columns. The only exception to this rule is that you may use a discrete (blue) date field. If you use a discrete date field to define headers, the other field must be a continuous field.
- Additional fields that create multiple, distinct trend lines: Discrete (blue) fields on the Rows, Columns, or Color shelves can be used as factors to split a single trend line into multiple, distinct trend lines.

Intuition on different trendline models:

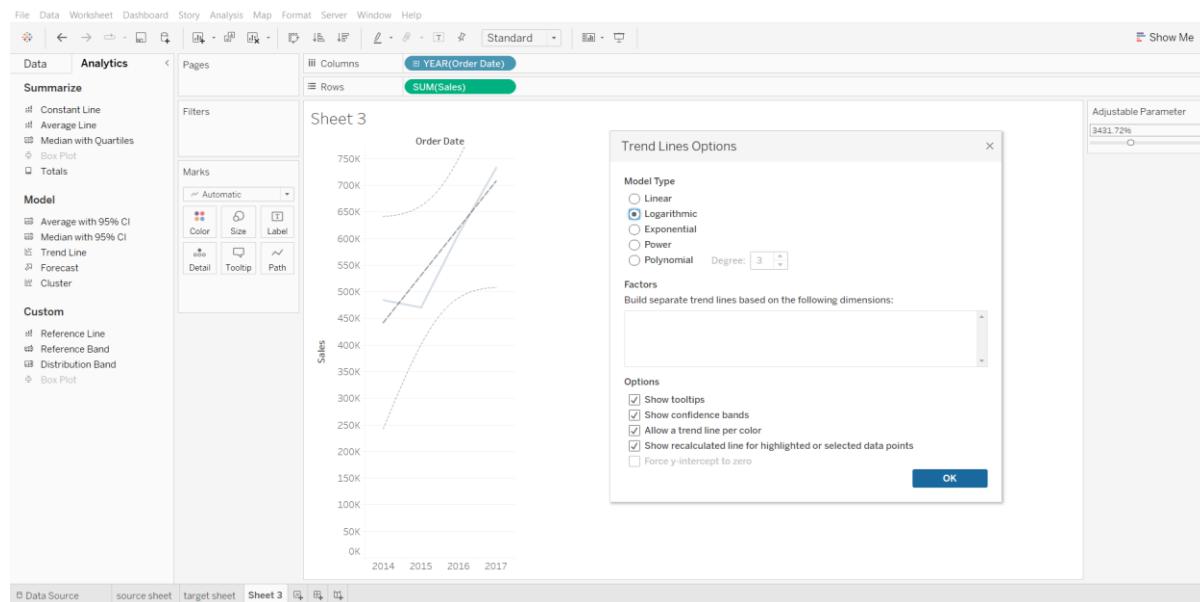
Consider example where x axis has size of a house in square feet and price on y axis

- **Linear:** We'd use a linear model if we assumed that, as size increases, other price will increase at a constant rate
- **Logarithmic:** We'd employ a logarithmic model if we expected the law of diminishing returns in effect—that is size can increase only so much before buyers will stop paying much more
- **Exponential:** We'd use an exponential model to test the idea that each additional increase in size results in a dramatic (exponential!) increase in price
- **Polynomial:** We'd employ a power trend model if we felt relation would indicate that the price was a function of the size to a certain power.
- **Power:** We'd use this model if we felt the relationship between Size and Price was complex and followed more of an S-shaped curve

Business problem 6:

Determine the year-wise trend in sales.

- Drag Order Date to columns
- Drag Sales to rows
- Go to the Analytics tab, drag the trend line to the linear option.
- Click on the trend line created, choose edit.
- Tick the show confidence bands option
- Check how the different trend lines fit the data.



To get more detailed information like MSE, SSE, R-squared, p-value etc. go to Analytics > Trend Lines > Describe Trend Model.

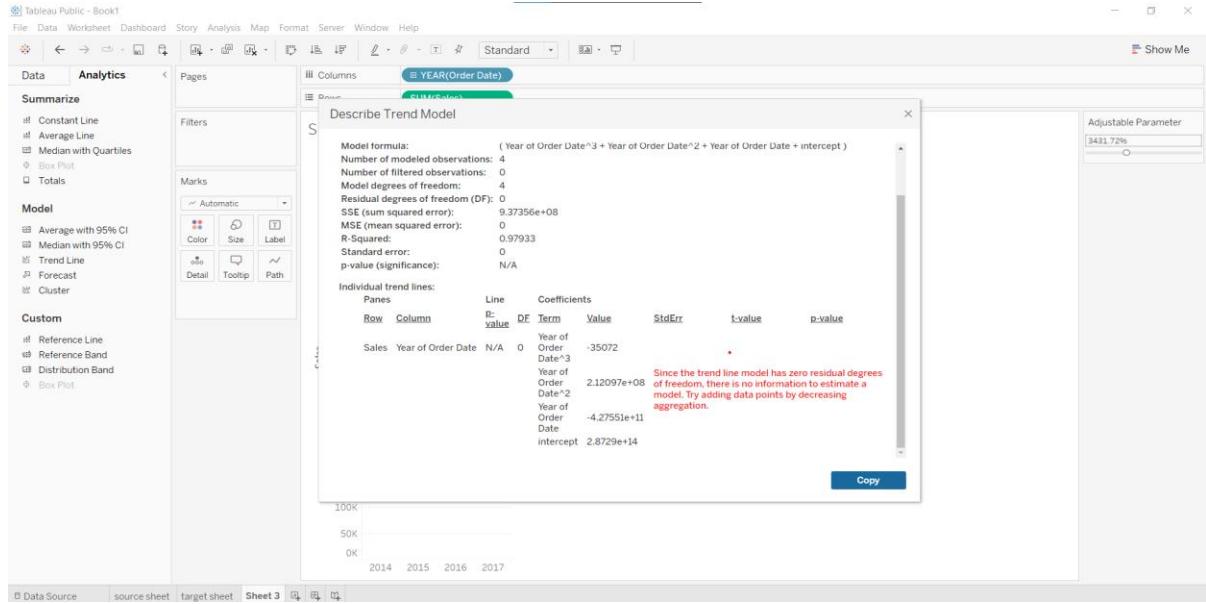


Tableau Lecture 7: Operations on Dataset

Agenda

- Joins
 - Building a map visualization
 - Cross-database joins
- Relationships
- Unions
- Blends

Joins

The data that you want to analyze is often made up of a collection of tables that are related by specific fields.

Joining is a method for combining the related data on those common fields. The result of combining data using a join is a table that's typically extended horizontally by adding fields of data.

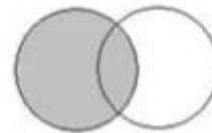
Name	# of Siblings
Taylor	2
Alex	3
Shannon	0
Tracy	1



Name	Eye Color
Taylor	Blue
Alex	Brown
Morgan	Brown

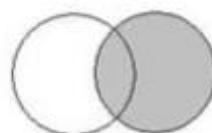
Left Join

Name	# of Siblings	Eye Color
Taylor	2	Blue
Alex	3	Brown
Shannon	0	null
Tracy	1	null



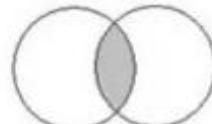
Right Join

Name	Eye Color	# of Siblings
Taylor	Blue	2
Alex	Brown	3
Morgan	Brown	null



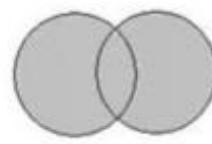
Inner Join

Name	# of Siblings	Eye Color
Taylor	2	Blue
Alex	3	Brown



Outer Join

Name	# of Siblings	Eye Color
Taylor	2	Blue
Alex	3	Brown
Shannon	0	null
Tracy	1	null
Morgan	null	Brown



Business problem 1:

Find per country CO2 KT and per capita emission and filter top 10 countries by CO2 KT.

Dataset : [World_bank_CO2.xlsx](#)

1. Creating Join

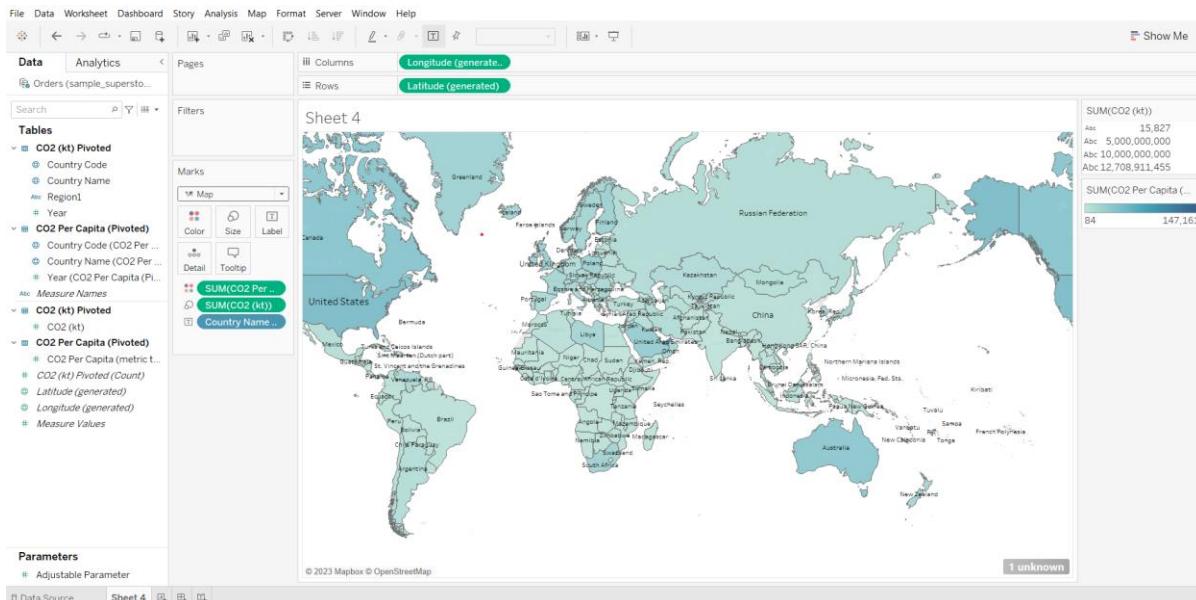
- Drag the 'CO2 (kt) Pivoted' table to the Canvas
- Double click on the 'CO2 (kt) Pivoted' box to get the physical layer.
- Drag 'CO2 per capita (pivoted)' to the canvas.
- Click on the join symbol
- Change the key field for both tables to the country name.

B- Orders (sample_superstore_)

The screenshot shows the Tableau Data Source interface. On the left, there's a sidebar with 'Connections' (World_Bank_CO2 (I) Microsoft Excel), 'Sheets' (including CO2 (kt) for Split, CO2 (kt) Pivoted, CO2 (kt) RAW DATA, CO2 Data Cleaned, CO2 Per Capita (Pivoted), CO2 Per Capita RAW DATA, Metadata - Countries, New CO2(KT) Pivoted, Sheet1, New Union, and New Table Extension), and 'About'. The main area displays a join diagram between 'CO2 (kt) Pivoted' and 'CO2 Per Capita (Pivoted)' tables. A tooltip says 'CO2 (kt) Pivoted is made of 2 tables.' Below the diagram is a preview of the data with 578604 rows.

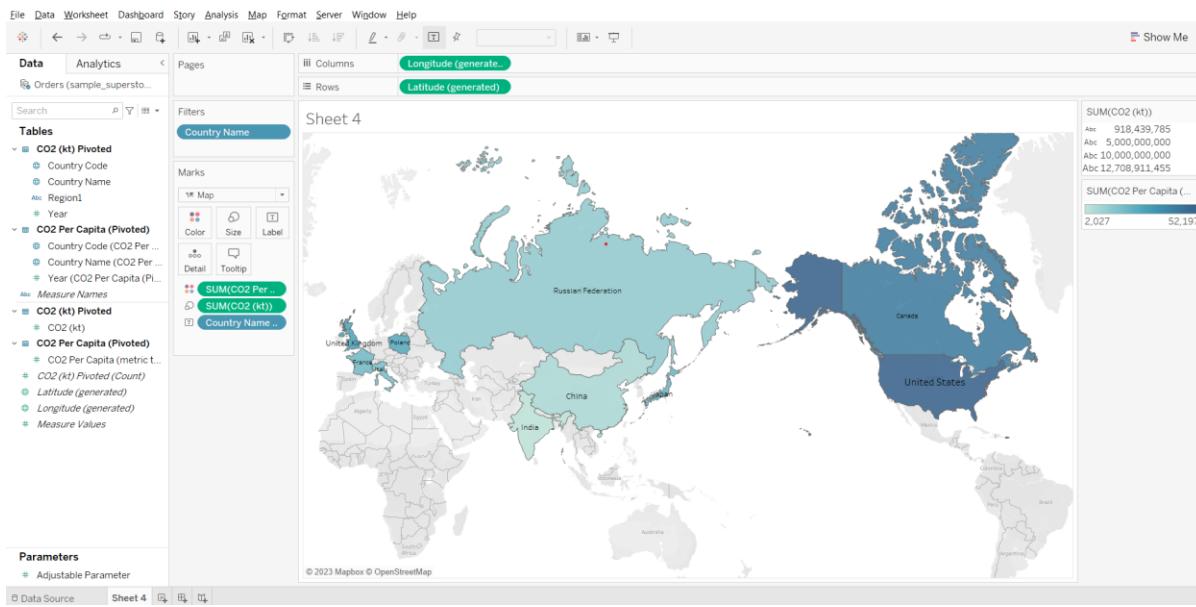
2. Visualization

- Drag Latitude to rows shelf
- Drag longitude to columns shelf
- Drag country name to label in marks card
- Drag CO2 (kt) to size in marks card
- Drag CO2 Per Capita to color in marks card
- In the marks card, change the visualization type to map.



3. Applying filter

- Drag country name to filters card
- Go to the 'Top' tab, choose the CO2 (kt) field, click OK.



Geographic data type:

Format geographic data in Tableau

- Depending on the type of map you want to create, you must assign certain data types, data roles, and geographic roles to your fields (or columns).
- For example, in most cases, your latitude and longitude fields should have a data type of number (decimal), a data role of measure, and be assigned the Latitude and Longitude geographic roles.
- All other geographic fields should have a data type of string, a data role of dimension, and be assigned the appropriate geographic roles.

Change the data type of a column

- Tableau might incorrectly assign a Postal Code column a data type of Number (whole). To create map views, your Postal Code data must have a data type of String.

Assign geographic roles to your geographic data:

How does tableau know where to plot on a map?

When you assign the correct geographic role to a field, Tableau assigns latitude and longitude values to each location in that field by finding a match that is already built into the installed geocoding database. This is how Tableau knows where to plot your locations on the map.

Geographic roles are sometimes **automatically assigned** to your data. You can tell a geographic role has been assigned to your data because the column includes a **globe icon**. If a **geographic role** is not automatically assigned, you can **manually assign one to your field**.

To assign or edit a geographic role:

1. On the Data Source page, click the globe icon.
2. Select **Geographic Role**, and then select a role that best matches your data.

For example, in this case, the Country column does not have a geographic role assigned to it, so the **Country/Region** geographic role is assigned.

The screenshot shows the Tableau Data Source pane. On the left, there's a list of data types: Number (decimal), Number (whole), Date & Time, Date, ✓ String, Boolean, ✓ Default. On the right, there's a preview of a data source with columns: Orders, State, Postal Code. The 'State' column has three rows: Kentucky, 42420; Kentucky, 42420; California, 90036. Below the preview, there's a 'Geographic Role' dropdown menu with options: ✓ None, ✓ Airport, ✓ Area Code (U.S.), ✓ CBSA/MSA (U.S.), ✓ City, ✓ Congressional District (U.S.), ✓ County/Region (highlighted in blue), ✓ County, ✓ NUTS Europe, ✓ Census Division, ✓ Census Place, ✓ ZIP Code Tabulation Area. The 'County/Region' option is selected.

Importance of maps:

- You have some location data in your data source
- You think that it would enhance getting better insight from data compared to other visualization
- It makes representing data in much more beautiful manner
- To answer spatial questions like
 - What is state wise sales?
 - Which states have the highest profit?
 - In which state we had the highest sales for a particular product?

Few things to keep in mind when creating a map view in tableau

- If your data source has **latitude and longitude fields** then they should have a **data type of numeric(decimal)** and a **data role of measure** should be assigned to them.
- And other **geographical fields like country, city, state** should be assigned **string data type** and a **data role of dimension**.
- Tableau will automatically assign geographic roles based on the field name and values present in the data.
- **If our data source does not contain latitude and longitude values we can still create maps using the geographical fields available in the data source.**

In the upper-left corner of a view, the view toolbar lets you zoom in and out, pan the view, and select data marks.

1

In the view toolbar click on the + to zoom in and - to zoom out

In the view toolbar click on the to zoom in to specific area and

In the view toolbar click on to pan the view

to reset zoom or pan click on

In the view toolbar click on to select marks within a rectangular area

In the view toolbar click on to select marks within a circular area

In the view toolbar click on to select multiple marks by drawing a freehand shape around them

Cross-database Joins

- In Tableau, a cross-database join is when you join two or more tables from different databases using a common field. This creates a single table in Tableau that can be used to create visualizations.
- Cross-database joins require a multi-connection data source - that is, you create a new connection to each database before you join the tables.
- We'll try to join tables from a CSV file and an Excel file.
 - Dataset:
 - [Retail_store_order_details.xlsx](#)
 - [Retail_store_product_details.csv](#)

Sample problem:

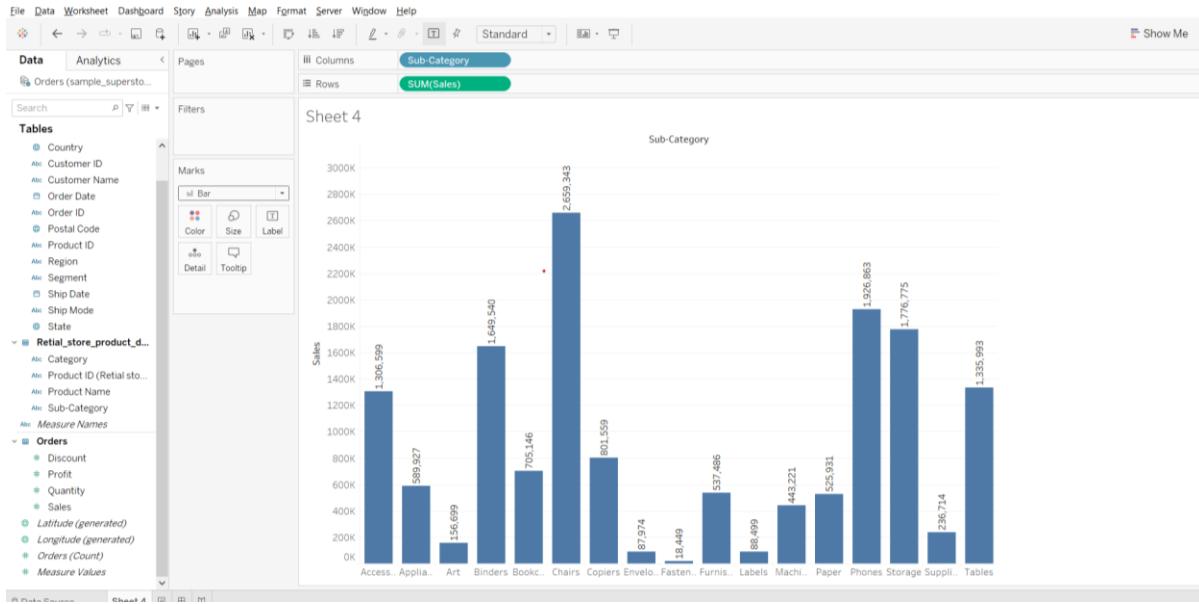
Identify subcategories that made the highest and lowest sales.

1. Loading the two different data sources and adding an inner join between the two tables
Note: one is the orders table and the other is the products table.

The screenshot shows the Tableau Data Source editor. On the left, under 'Connections', 'Retail_store_order_details' is selected. Under 'Files', several CSV files are listed. In the center, the 'Orders' connection is expanded, showing it's made of two tables: 'Orders' and 'Retail_store_product_details.csv'. A 'Join' dialog is open between these two tables, with 'Inner' selected. The 'Data Source' dropdown shows 'Retail_store_pro...', and the join condition 'Product ID = Product ID (Retai...' is set. Below this, a preview of the 'Orders' table is shown with 21 fields and 66870 rows. The preview includes columns like Order ID, Order Date, Ship Date, Ship Mode, Customer ID, Customer Name, and Segment.

2. Solving the business problem:

- Go to a worksheet
- Drag Sub-Category to Columns
- Drag Sales to Rows



Relationships

Relationships are a dynamic, flexible way to combine data from multiple tables for analysis.

A relationship describes how two tables relate to each other, based on common fields, but does not merge the tables together.

When a relationship is created between tables, the tables remain separate, maintaining their individual level of detail and domains.

Think of a relationship as a contract between two tables. When you are building a viz with fields from these tables, Tableau brings in data from these tables using that contract to build a query with the appropriate joins.

[Advantages of Relationships](#)

[Relationships vs Joins](#)

Note: We'll be using the same example as we saw for Joins but instead of Joins here we'll be building relationships for understanding the topic.

- Add the World Bank CO2 dataset
- Drag CO2 (kt) Pivoted on the canvas
- Drag CO2 Per Capita Pivoted on the canvas
- In the clause, set the field for both tables to the country name.

The screenshot shows the Microsoft Power BI Data Window interface. On the left, the 'Connections' pane lists 'World_Bank_CO2 (1)' as a Microsoft Excel connection. The 'Sheets' pane contains various items like 'About', 'CO2 (kt) for Split', 'CO2 (kt) Pivoted', 'CO2 (kt) RAW DATA', 'CO2 Data Cleaned', 'CO2 Per Capita (Pivoted)', 'CO2 Per Capita RAW DATA', 'Metadata - Countries', 'New CO2(KT) Pivoted', and 'Sheet1'. The main canvas area shows two tables connected by a relationship: 'CO2 (kt) Pivoted' and 'CO2 Per Capita (Pivoted)'. A preview window displays the resulting data, which is a union of the two tables. The preview table has columns: 'Country Name (CO2 Per...)', 'CO2 Per Capita (Pivoted)', 'Country Code (CO2 Per ...)', 'Year (CO2 Per Capita (Pl...)', and 'CO2 Per Capita (metric ...)'. The data shows multiple entries for Afghanistan from 1974 to 1981.

Country Name (CO2 Per...)	CO2 Per Capita (Pivoted)	Country Code (CO2 Per ...)	Year (CO2 Per Capita (Pl...)	CO2 Per Capita (metric ...)
Afghanistan	AFG		1974	0.1557
Afghanistan	AFG		1975	0.1690
Afghanistan	AFG		1976	0.1549
Afghanistan	AFG		1977	0.1831
Afghanistan	AFG		1978	0.1633
Afghanistan	AFG		1979	0.1687
Afghanistan	AFG		1980	0.1332
Afghanistan	AFG		1981	0.1526

Union

Though union is not a type of join, union is another method for combining two or more tables by appending rows of data from one table to another.

Make sure the tables you union have the same number of fields, the same field names, and the fields are the same data type.

Data: [World_bank_CO2.xlsx](#)

Perform Union on table CO2 (kt) Pivoted and New CO2 (kt) Pivoted

- Drag CO2 (kt) Pivoted to the canvas
- Double click 'CO2 (kt) Pivoted' to open the physical layer.
- Drag New CO2 (kt) Pivoted just below 'CO2 (kt) Pivoted' on the canvas.
- Close the physical layer.

Blending

Data blending is a method for combining data from multiple sources. Data blending brings in additional information from a secondary data source and displays it with data from the primary data source directly in the view.

Example: Consider the Sales data is present in a relational database and Sales Target data in an Excel spreadsheet. Now, to compare actual sales to target sales, you can blend the data based on common dimensions to get access to the Sales Target measure. The two sources involved in data blending are referred to as primary and secondary data sources.

Note: A left join is created between the primary data source and the secondary data source.

Business problem 2 :

For each country, find the total number of olympic medals won and CO2 per capita (metric tons) emission till 2008.

Dataset : [Modified_Summer_Olympic_medallists_1896-2008.xlsx](#) & [World_bank_CO2.xlsx](#)

Use table Team event fixed from Modified_Summer_Olympic_medallists_1896-2008

Use table CO2 per capita (Pivoted) from World_bank_CO2 dataset

1. Loading 2 Data sources in a Single workbook

- Since world bank CO2 dataset is already connected, drag CO2 (kt) Pivoted to the canvas

- Go to worksheet
- In the Data menu, click new data source and select the modified summer olympics excel file.
- Go to Data Source and from modified summer olympics file, drag the team events fixed table to the canvas.

The screenshot shows the Tableau Data pane. The 'Data' tab is selected. Under 'Pages', there is a list of pages: 'Pages', 'iii Columns', and 'Rows'. Below 'Pages' is a 'Filters' section. To the right is a large workspace titled 'Sheet 5' with a placeholder 'Drop field here'. On the left, under 'Tables', several fields are listed: 'Country Code', 'Country Name', 'Year', 'CO2 Per Capita (metric tons)', 'CO2 Per Capita (Pivoted) (...)', 'Latitude (generated)', 'Longitude (generated)', and 'Measure Values'. The 'Marks' section on the right includes options for 'Automatic' and buttons for 'Color', 'Size', 'Text', 'Detail', and 'Tooltip'.

2. Edit Blend Relationship

- Click on CO2 Per Capita in Data panel
- Go to the Data menu, click Edit Blend Relationships...
- Select Custom and click on Add...
- Choose the country name from the Primary Data Source field, click ok on both dialog boxes.

The screenshot shows the Tableau interface with the 'Data' tab selected. The 'Modified_Summer_Olym...' data source is visible in the Data pane. A 'Blend' icon is present in the Data pane. A 'Show Me' button is in the top right. A 'Add/Edit Field Mapping' dialog box is open in the center, showing a mapping between 'Country Name' in the primary data source and 'Country' in the secondary data source. The dialog has 'OK' and 'Cancel' buttons at the bottom.

3. Edit Alias

- If you drag the country name from CO2 data to the rows shelf, an orange link appears next to 'country' in Olympics data, saying - 'Stop using country as a linking field'.

This is because the name of some countries (like the USA) is written differently in the two files. We can set an alias of a country if needed using the following steps -

- In CO2 data, from the country name dropdown in the data panel, click Aliases...
- Set the alias for the United States as the USA.
- Drag country name from CO2 data and country from Olympics data to rows shelf.
- Observe that the entries for 'USA' are next to each other.

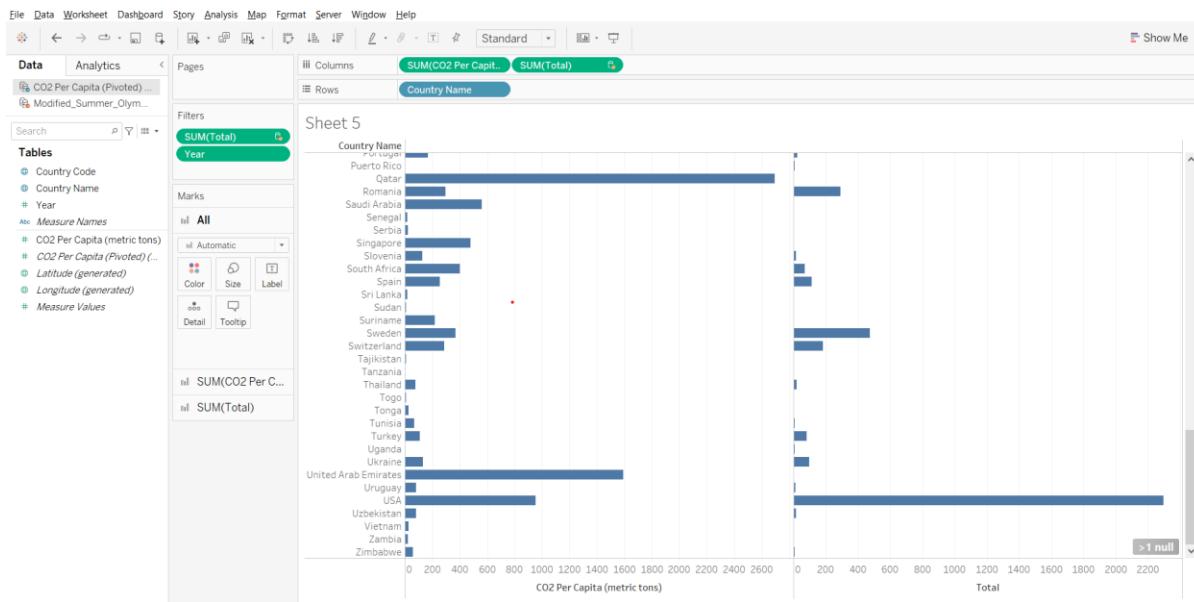
The screenshot shows the Tableau Data Editor interface. On the left, the 'Tables' pane lists 'CO2 Per Capita (Pivoted)' and 'Modified_Summer_Olympics'. The main area displays two tables side-by-side:

Country Name	Country	Type
Sabah	Sabah	Abs
Suriname	Suriname	Abs
Swaziland	Null	Abs
Sweden	Sweden	Abs
Switzerland	Switzerland	Abs
Syrian Arab Republic	Null	Abs
Tajikistan	Tajikistan	Abs
Taiwan	Tanzania	Abs
Thailand	Thailand	Abs
Timor-Leste	Null	Abs
Togo	Togo	Abs
Tonga	Tonga	Abs
Trinidad and Tobago	Null	Abs
Tunisia	Tunisia	Abs
Turkey	Turkey	Abs
Turkmenistan	Null	Abs
Turks and Caicos Isl.	Null	Abs
Tuvalu	Null	Abs
Uganda	Uganda	Abs
Ukraine	Ukraine	Abs
United Arab Emirates	United Arab Emirates	Abs
United Kingdom	Null	Abs
Upper middle income	Null	Abs
Uruguay	Uruguay	Abs
USA	USA	Abs
Uzbekistan	Uzbekistan	Abs
Vanuatu	Null	Abs
Venezuela, RB	Null	Abs
Vietnam	Vietnam	Abs
Virgin Islands (U.S.)	Null	Abs
West Bank and Gaza	Null	Abs
World	Null	Abs
Yemen, Rep.	Null	Abs
Zambia	Zambia	Abs
Zimbabwe	Zimbabwe	Abs

Note: If the same error pops up when bringing country name from one table just click on the orange link that appears

4. Solving the business problem:

- Drop country name to rows
- Drop CO2 Per Capita to columns
- Drop Total from Olympics data to columns
- Click on '>149 nulls' in bottom right corner of the sheet and select Filter data
- Drag Year in CO2 data to filters and set maximum value to 2008.



Joins vs Blend

Data Joining

Used when the data set is from the same source

Ability to use different types of join (left join, right join, inner join and full outer join)

Data has to be maintained at the same level of granularity

Joins data at a row-level

Data Blending

Used when the dataset is from a different data source

Uses only left join.

Data can be available in different level of granularity.

Sends separate query to each dataset, aggregates and then performs blending

Cross-database Join vs Blending

Cross database Join	Blending
It is used when you want to join two or more tables from different databases using a common field.	It is a way to combine data from multiple sources within a single visualization.
This creates a single table in Tableau that can be used to create visualizations.	It does not create a single table in Tableau. Instead, data from different sources is blended on the fly as needed to create a visualization.
It require that the databases should have a common field that can be used for the join.	It is not require that the databases should have a common field that can be used for Blending
Not recommended to use when Data sources are at different levels of detail	Recommended to use when Data sources are at different levels of detail.

Tableau Lecture 8: Advanced charts, Dashboard & Tableau Capstone Intro

Agenda

- Creating a Dashboard
- Dashboard Objects
 - Horizontal and vertical containers
 - URL object
- Actions in dashboard
 - Parameter and Set action
- Stories
- Increase the efficiency of the dashboard
- Advanced charts
 - Box and Whiskers plot
 - Bullet chart

Dataset : [sample superstore](#)

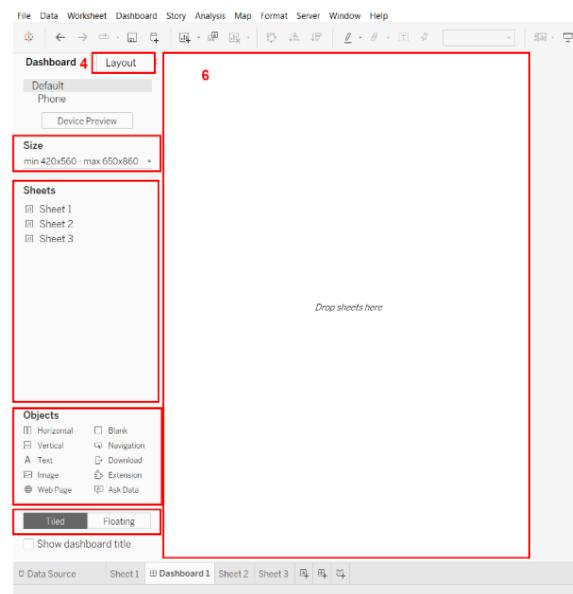
Introduction to Dashboards

- A dashboard is a collection of several views, letting you compare a variety of data simultaneously.
- For example, if you have a set of views that you review every day, you can create a dashboard that displays all the views at once, rather than navigate to separate worksheets.
- Like worksheets, you access dashboards from tabs at the bottom of a workbook.
- Data in sheets and dashboards is connected; when you modify a sheet, any dashboards containing it change, and vice versa.

The Dashboard Interface

The dashboard interface is similar to the worksheet workspace in many ways.

1. A list of worksheets that have not been hidden and which are sorted according to the order in which they are arranged in your workbook.
2. Containers and objects that can be dragged onto your dashboard to increase the integrity of the design.
3. A controller for new objects.
4. A hierarchy of dashboard objects that can be used for navigation.
5. A controller for the dashboard size.
6. The canvas where you will compose the dashboard.



Building a Dashboard

Demo : <https://www.tableau.com/learn/tutorials/on-demand/getting-started-part9?playlist=391099>

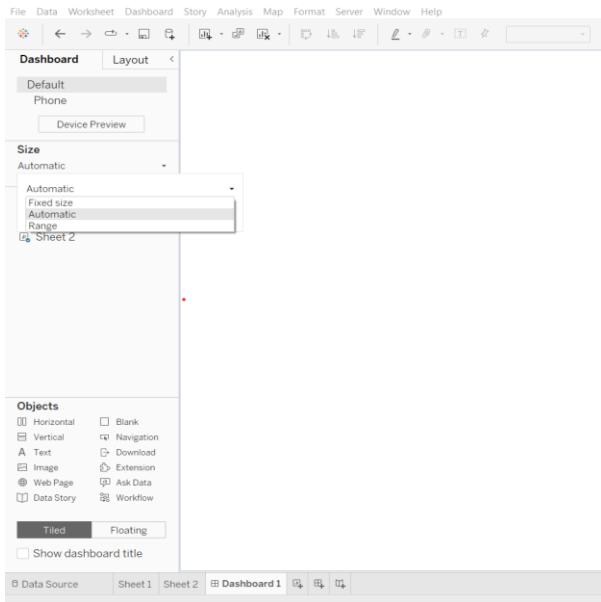
Dashboard Objects

- Tableau Public has several different layout objects that you can add to your dashboard to control the composition.
- Containers allow you to essentially group dashboard objects together
- Show/Hide container buttons within dashboards help to declutter the view and make it easier to understand the data.(demo added)

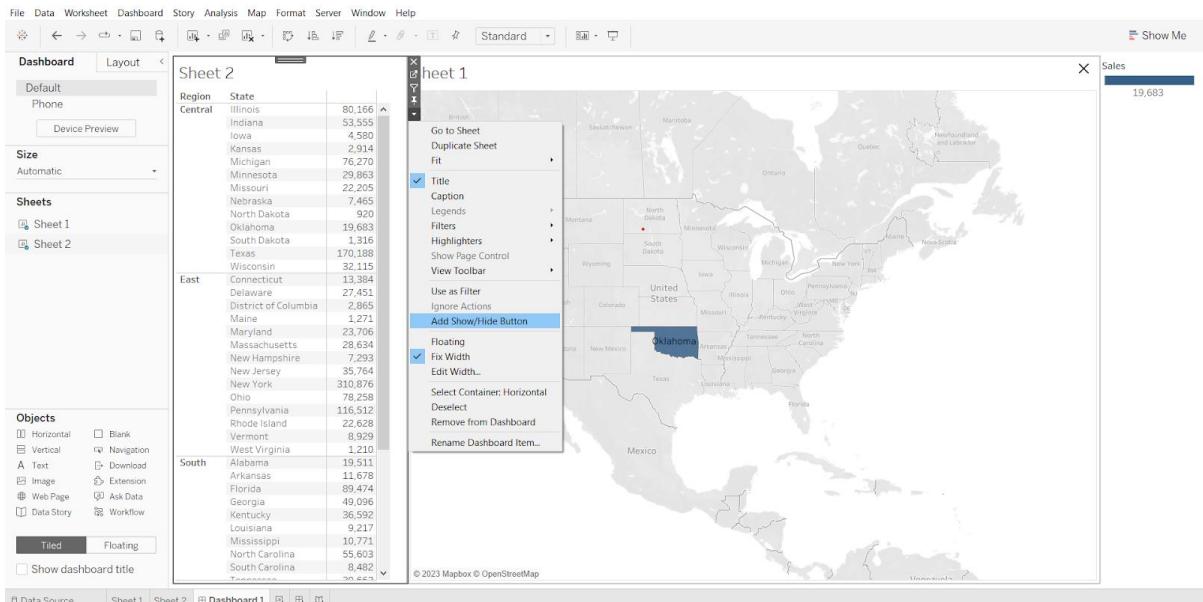
Demo :

- Using the sample superstore data create a visualization (map type) showing state wise sales.

- In another worksheet, drop Region and state to the Rows shelf and drop sales to labels in the Marks card.
- Create a new dashboard. Set size to automatic.
- Select Tiled



- Drop both Sheet 1 and Sheet 2 on the canvas.
- Two containers have been created. Open the menu of a container from its top-right and select 'Add show/hide button.'
- Right click on the button and choose the sheet.
- Alt + Click on the button will show or hide the container



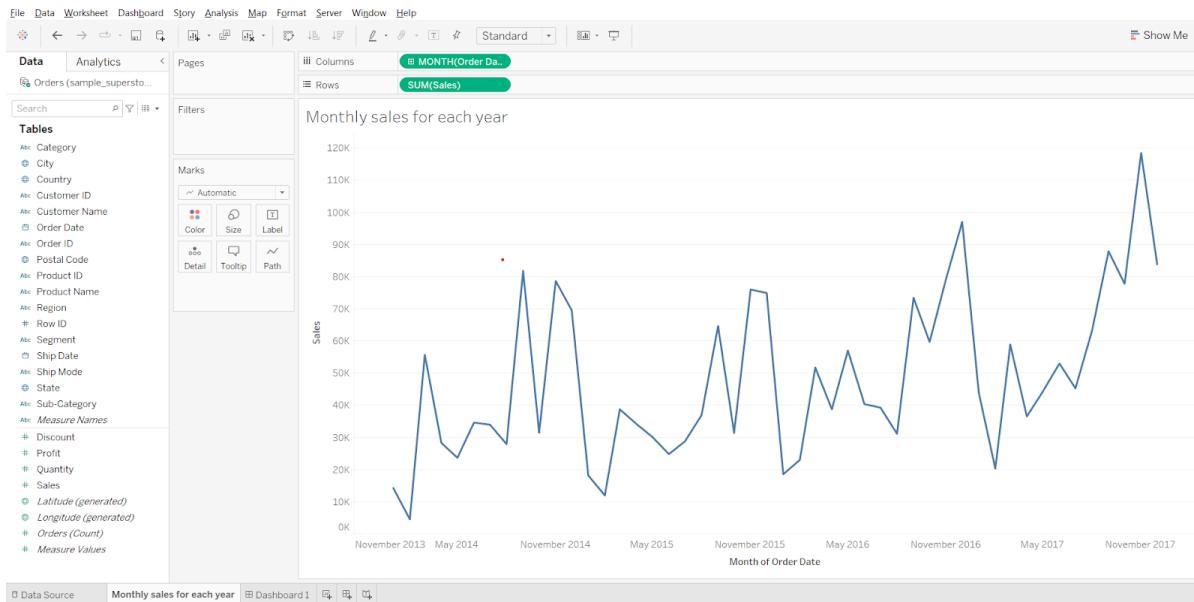
The following objects can be added to your dashboard: (demo of each has been added below)

- Horizontal layout containers: You can drag worksheets that you want next to each other into these containers. It has a Dark Blue border.
- Vertical layout containers: You can drag other objects and worksheets that you want to stack from top to bottom into these containers. It has a Dark blue border
- Text objects: These can be used to add titles and calls to action. Though you can't add field tokens to text objects, you can add parameter tokens, which will be discussed in depth in the next chapter.
- Images: You can use these to browse logos or branding elements that add richness to your dashboard. You can also use images as links to web pages, since each image can have a URL attribute.
- Web page objects: These can be used to add content from the Internet. You can even create dynamic content.
- Blank objects: These objects can be used to control space. Blank containers are transparent. So, the background colors that you are using will show through them.
- Extension: Extensions let you add unique features to dashboards or directly integrate them with applications outside Tableau. Adding extensions is easy; you incorporate them into dashboard layouts just like other dashboard objects. Extensions expand dashboard functionality with the help of web applications created by third-party developers.
- Navigation Object: Lets your audience navigate from one dashboard to another, or to other sheets or stories.

Dashboard Objects Demonstrations

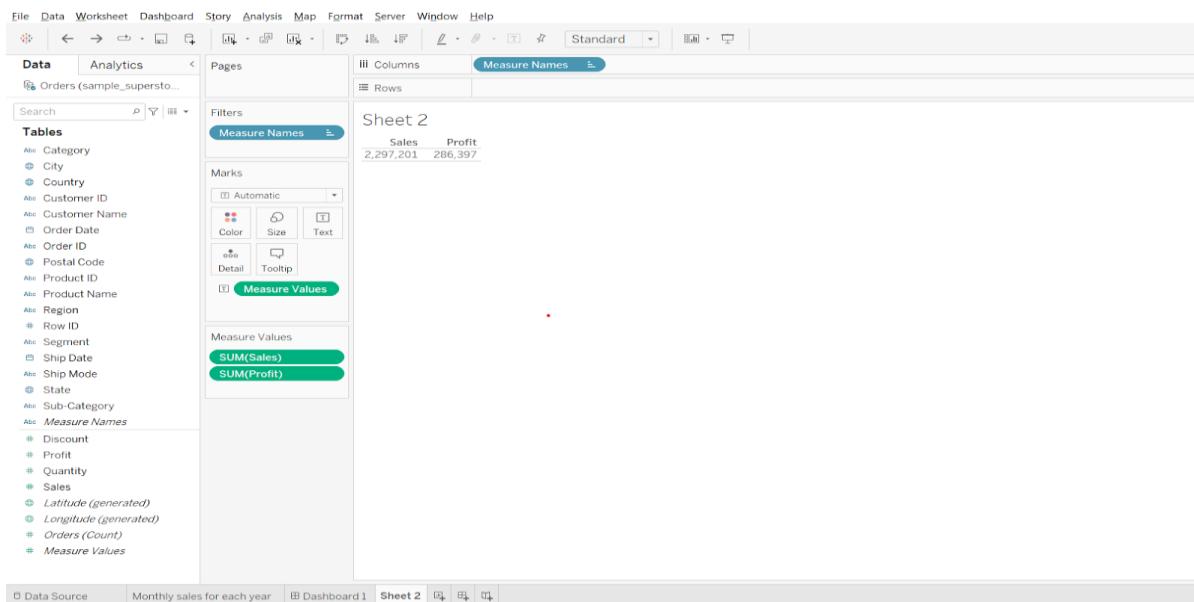
Creating 3 sheets that will be used to explain the all concept.

1. Sheet 1: Monthly sales for each years
 - Drag order date to columns shelf
 - From its dropdown select 'month with year'
 - Drop sales on rows shelf
 - Rename sheet to Monthly sales for each year



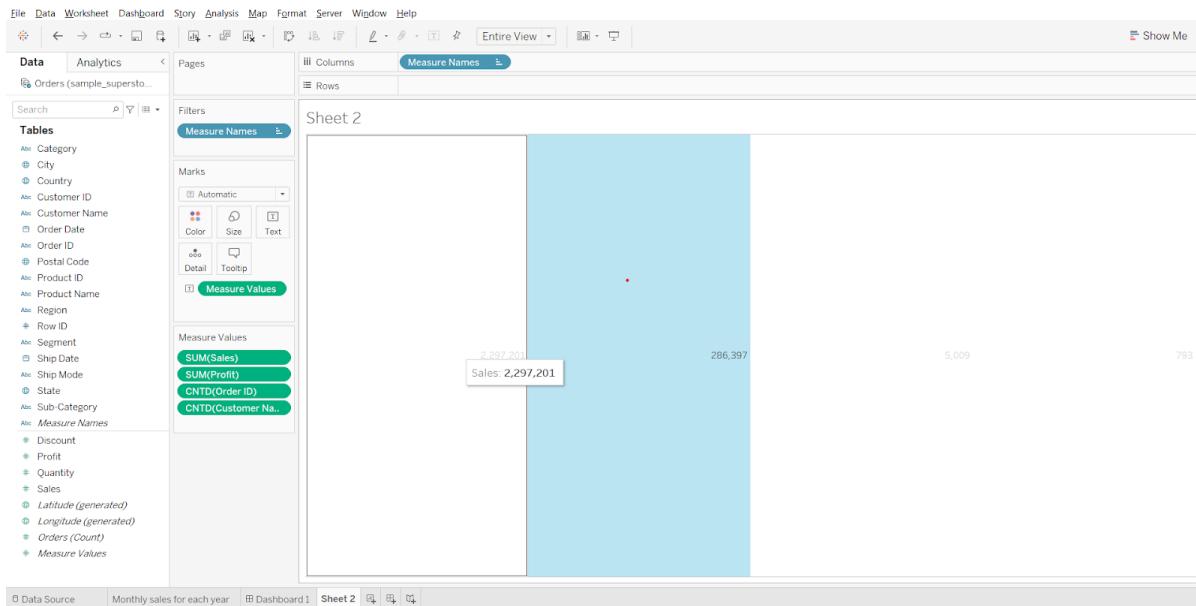
2. Sheet 2: Summary Table

- Drag measure names to columns shelf
- Drag measure values to text on marks card
- From measure values card, remove CNT(Orders), SUM(Discount) and SUM(Quantity)
- In measure values card, reorder sales to above profit.

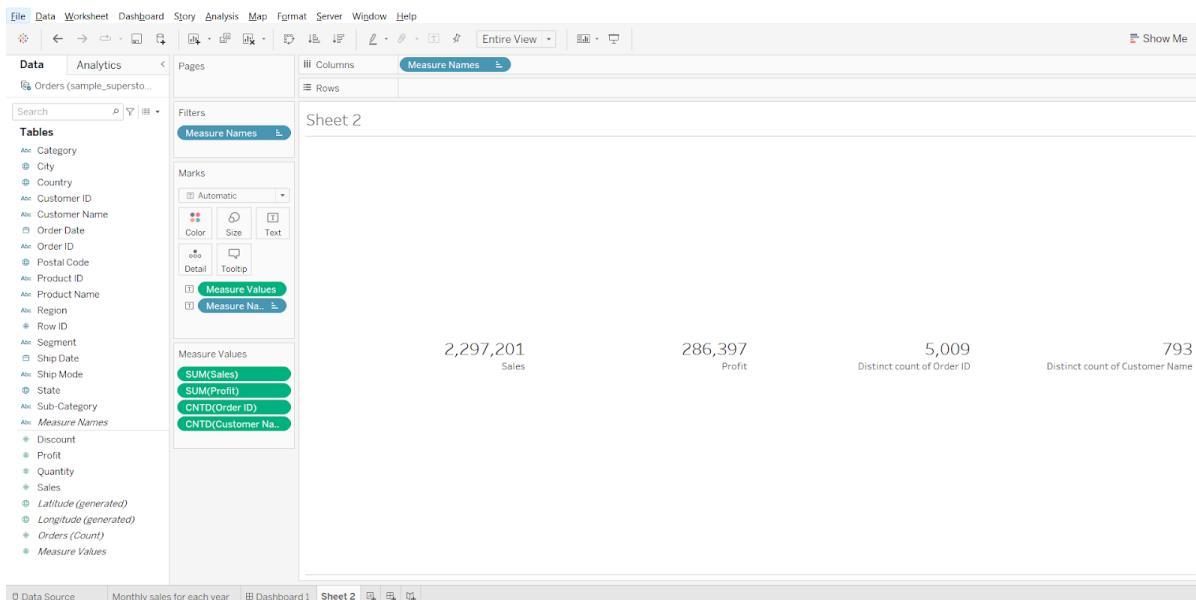


- Drop Order Id to details in marks card, change its measure to count distinct.
- Drag CNTD(Order ID) to measure values card
- Drop Customer Name of text in the Marks card and change its measure to count distinct.
- Drag CNTD(Customer Name) to measure values card

- Change Fit to ‘Entire View’
- Right click on the header in the sheet and uncheck ‘Show Header’ option.

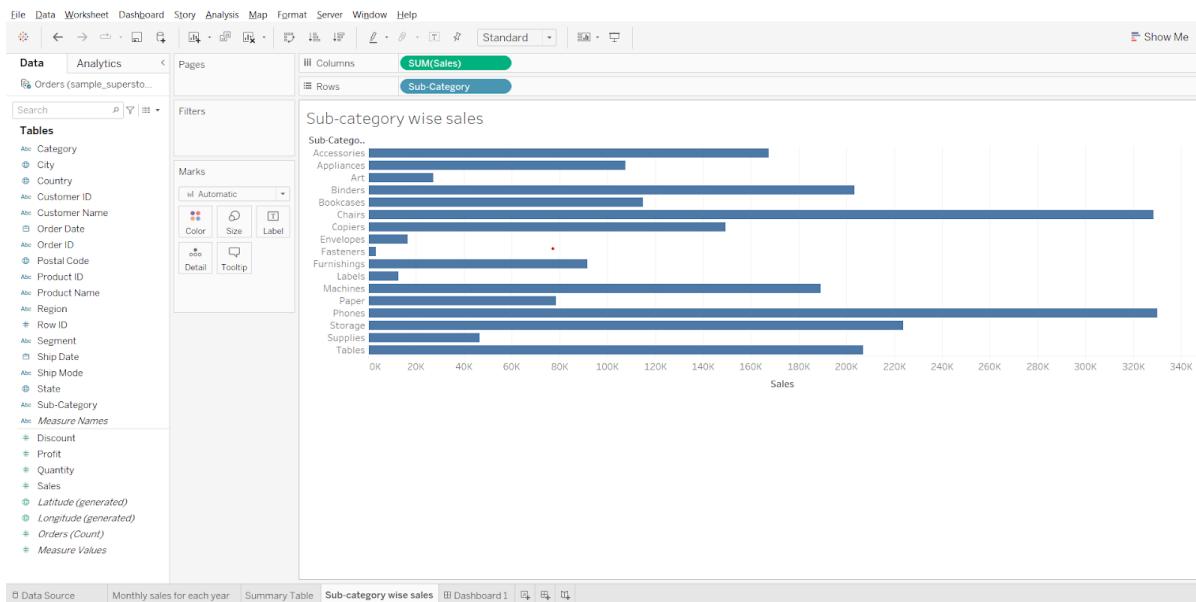


- Drag measure names to text in marks card
- Double click on text in the Marks card, open the dialog box.
- Select Measure Values, change size to 16. Click ok.
- Change sheet name to ‘summary table’



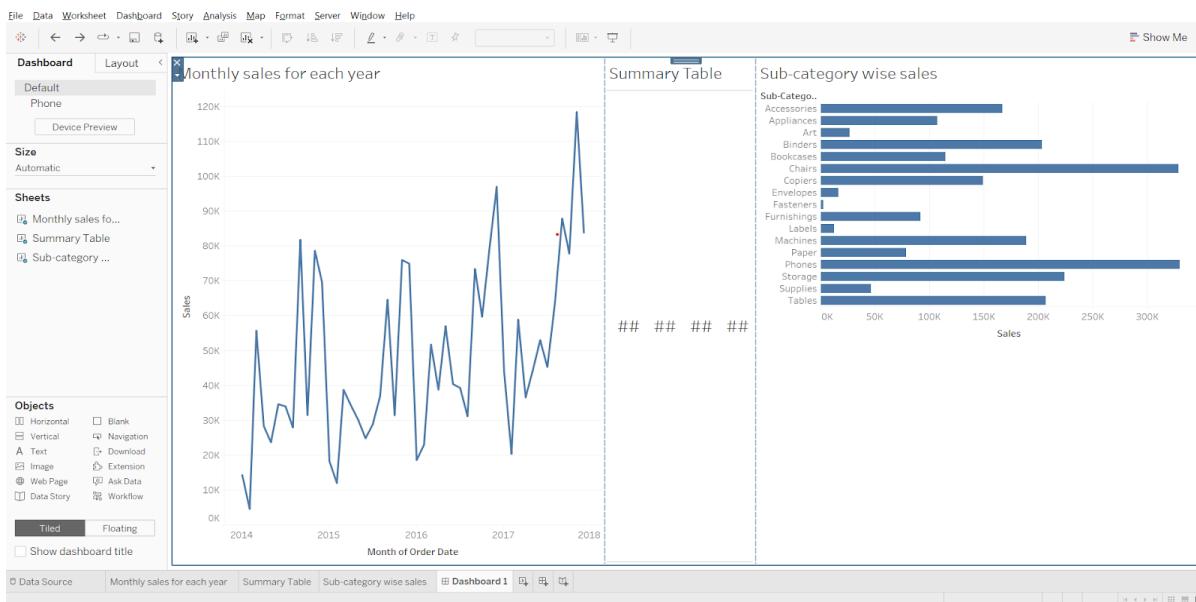
3. Sheet 3: Sub-category wise sales

- Drag sub-category to rows
- Drag sales to columns
- Rename sheet to sub-category wise sales.

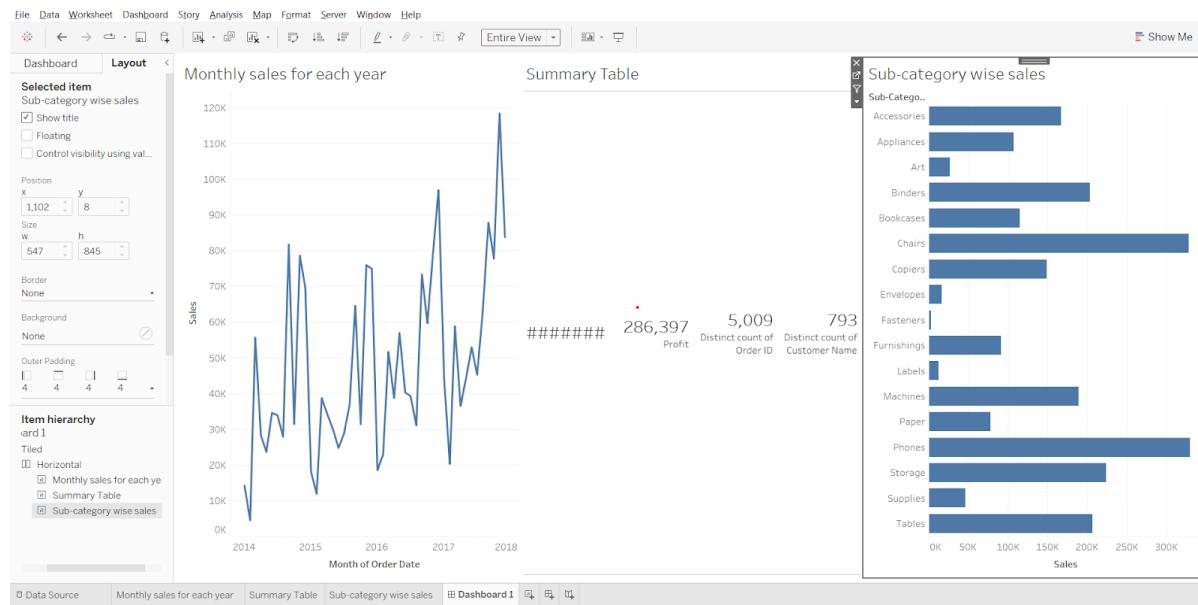


Creating a Dashboard using the Horizontal container

- Create a new dashboard tab, select tiled.
- From objects, drag horizontal to the canvas.
- Drag the 1st sheet to roughly the centre of the canvas.
- Drag the 2nd sheet to the right end, then drag the 3rd sheet to the right end.



- Go to the layout tab in the left pane.
- In Item Hierarchy, click on Horizontal.
- From the top left corner of the blue box (horizontal container) that shows, open the menu and select distribute contents evenly.
- Select the third sheet and change fit to the entire view.



Creating a Dashboard using the Vertical container

- Create a new dashboard and drag ‘vertical’ from ‘objects’ to the canvas
- Set size to automatic
- Drag the 1st sheet to roughly the centre of the canvas.
- Drag 2nd and 3rd sheets to the bottom, be careful not to drop them outside the vertical container.
- Go to layout, click on ‘vertical’ in the item hierarchy section.
- Open the menu from top left corner of the vertical container, select Distribute Content Evenly

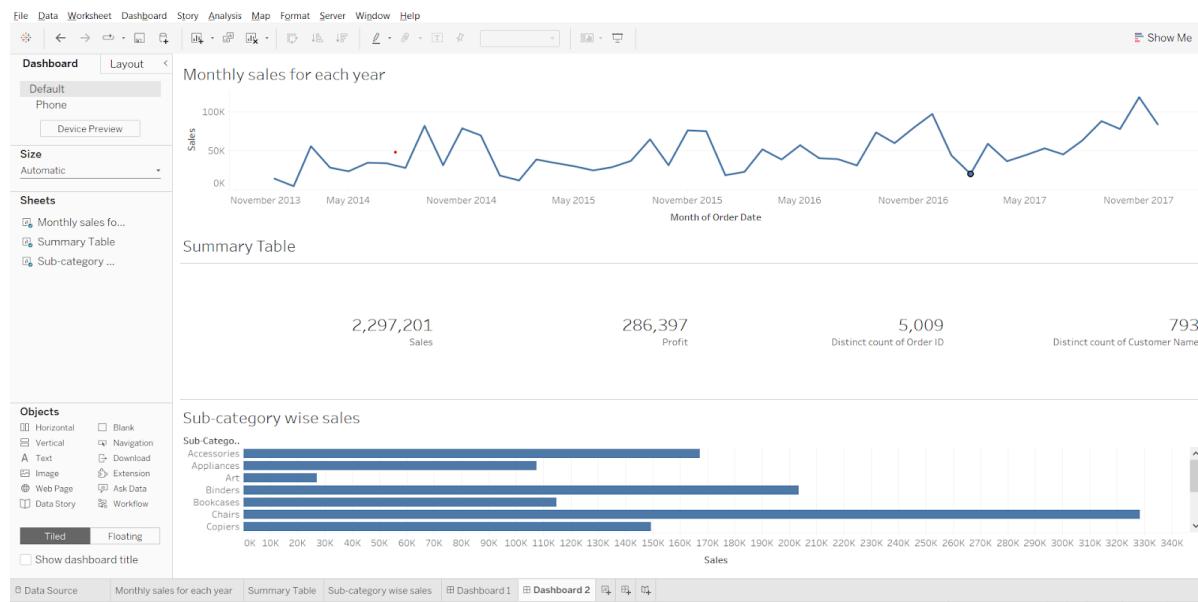
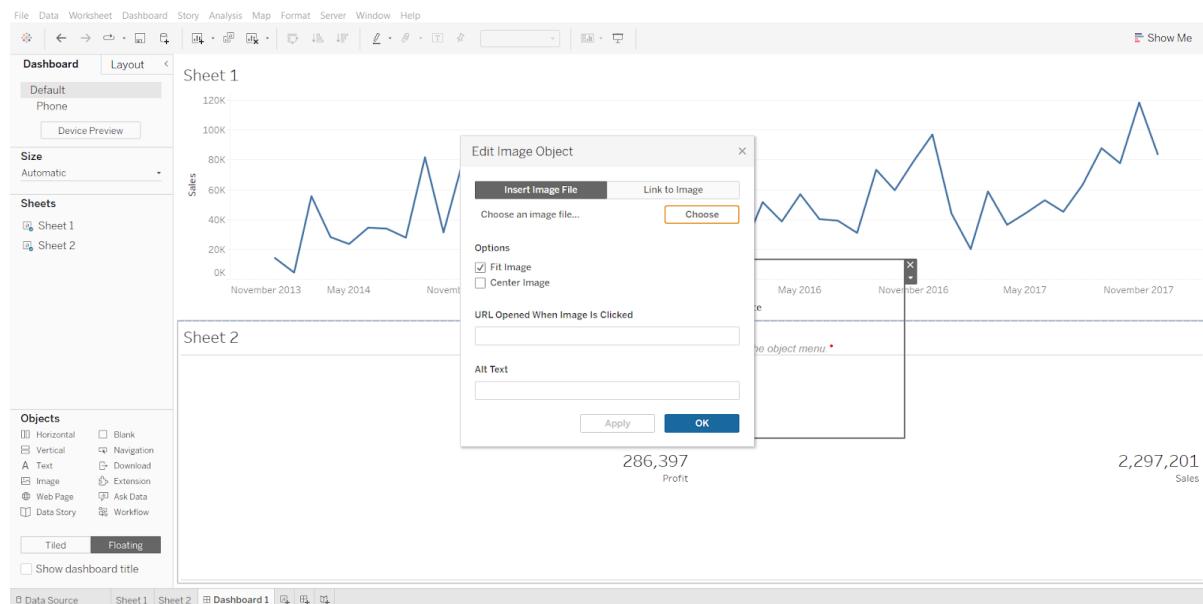


Image object

Using an image object we are using a floating container.

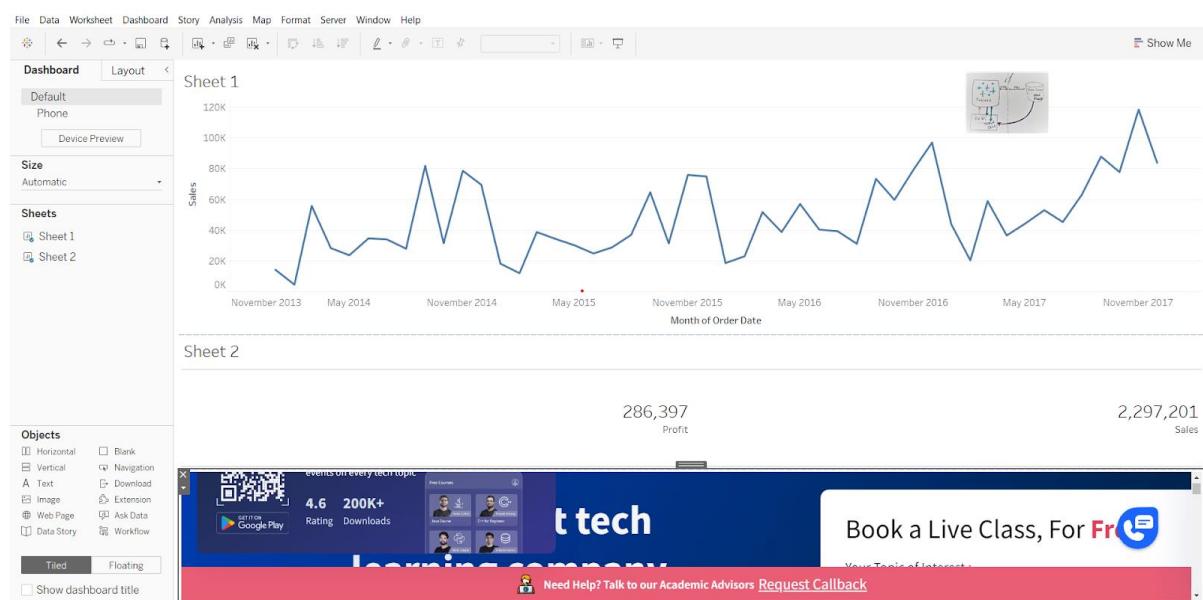
- Click on floating, drag the image from objects to the canvas.
 - In the box, choose an image, click OK.



Position the image as needed.

Web Page object

Click on tiled, drag ‘web page’ from objects to the bottom of the canvas. Enter the url, say ‘www.scaler.com’.



Interactive Dashboards with actions

To make your visuals interactive in Tableau is to use a series of actions :

- **Filter action** - Use the data from one view to filter data in another to help guide analysis
- **Highlight action** - Call attention to marks of interest by coloring specific marks and dimming all others.
- **Go to URL action** - Create hyperlinks to external resources, such as a web page, file, or another Tableau worksheet.
- **Go to sheet action** - Simplify navigation to other worksheets, dashboards, or stories.
- **Change Parameter action** - Let users change parameter values by directly interacting with marks on a viz.
- **Change set values** - Let users change the values in a set by directly interacting with marks on a viz.

The screenshot shows the 'Actions' dialog box in Tableau. At the top, it says 'Actions' and has a close button 'X'. Below that, a descriptive text reads: 'Actions let you create interactive relationships between data, dashboard objects, other worksheets, and the web.' Underneath, there's a section titled 'Show actions for' with two radio buttons: 'This workbook' (selected) and 'This sheet'. A table lists actions for a row named 'Set1':

Name	Run On	Source	Fields
Set1	Select	Dashboard 1 (sales by...)	Region Set

A context menu is open over the 'Set1' row, listing the following options:

- Filter...
- Highlight...
- Go to URL...
- Go to Sheet...
- Change Parameter...
- Change Set Values...

At the bottom of the dialog, there are buttons for 'Add Action ▾', 'Edit', 'Remove', 'Cancel', and a prominent blue 'OK' button.

Change Parameter action demonstration:

Using parameter action to choose chart type - [video link](#)

To use custom icons download the icons that you want and save it in this location -
\\Documents\\My Tableau Repository\\Shapes

Icons

Bar chart icon



Line chart icon-



Map icon

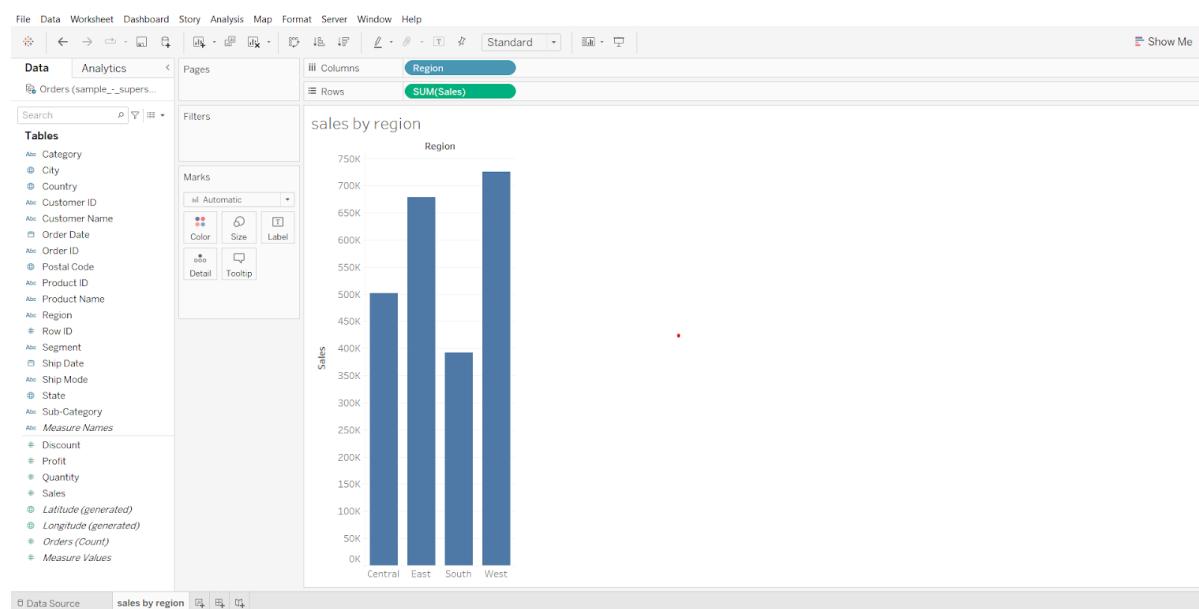


Change set values:

Demo:

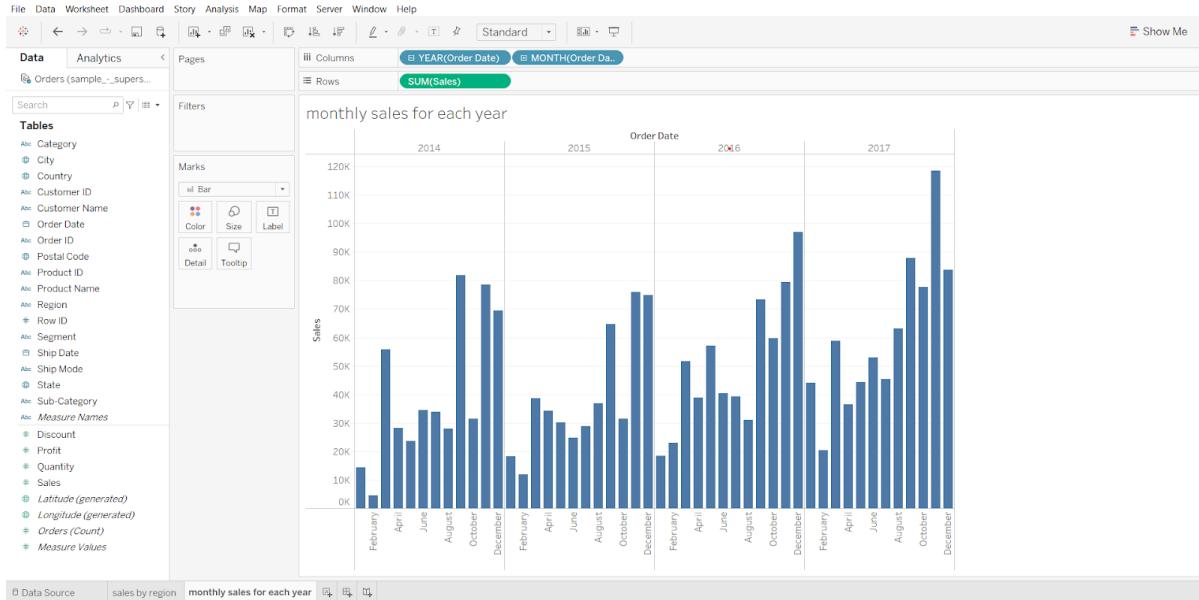
Display what proportion each region contributed to total monthly sales for each year using set action.

1. Displaying total sales for each region
 - o Drop sales on rows shelf
 - o Drop region on columns shelf
 - o Rename sheet to sales by region



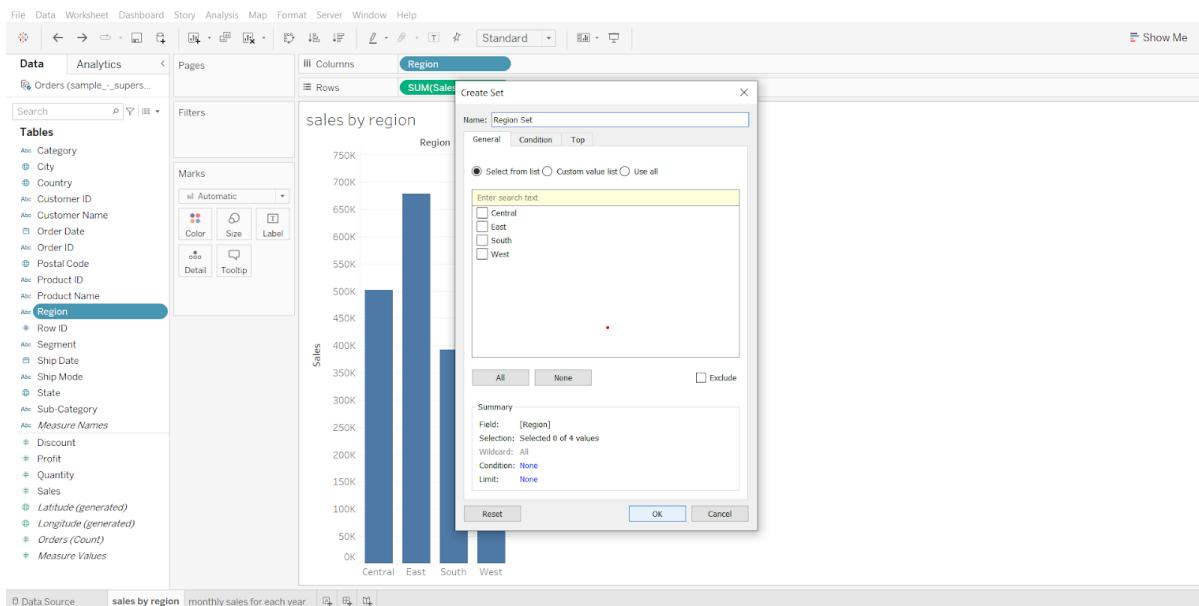
2. Displaying monthly sales data for each year

- Drag Order Date to columns shelf
- Click on the small plus to get quarters
- Open Quarter(Order Date) dropdown and select month (without year)
- Drag sales to rows
- Rename the sheet to 'monthly sales for each year'



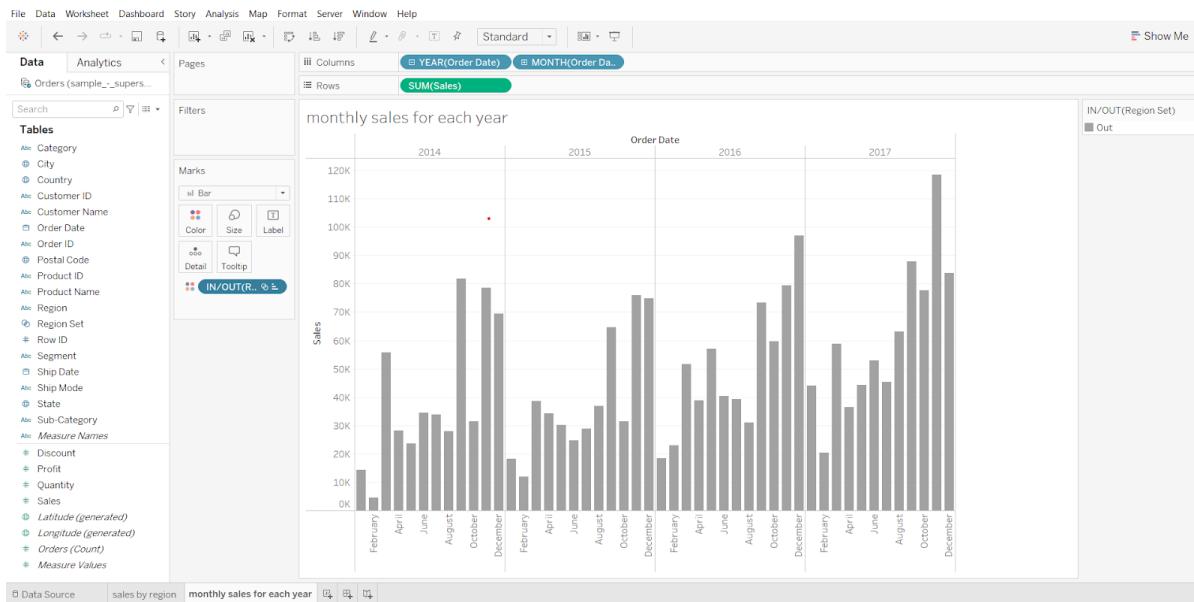
3. Creating an empty set for region field

- Open the sales by region worksheet
- Go to region in data panel, open its dropdown, go to Create > Set
- In the dialog box, click ok.



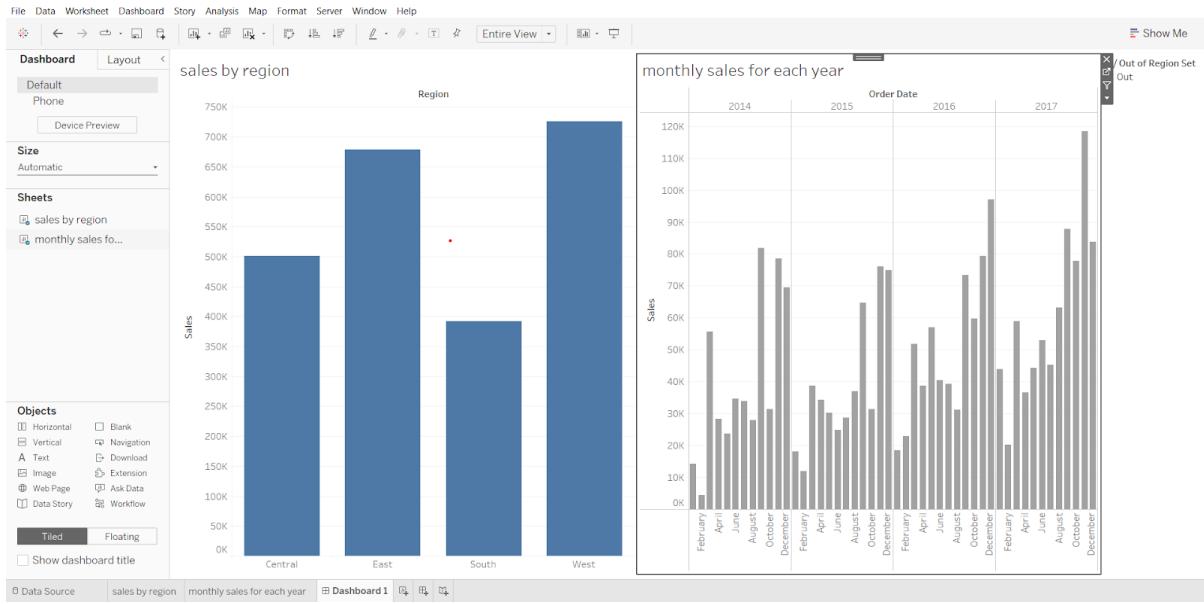
4. Adding newly created set to the time series graph and nothing is present in this set currently hence everything is in gray color
- Go to 'monthly sales for each year' sheet, drop region set on color in marks card
 - Open the In/Out(Region) dropdown from the marks card, click on Sort
 - Click on Descending, then click on ascending, close the dialog box.

Here we are sorting the result of it will be visible in the final dashboard such that the selected region values will be displayed at the bottom rather than on top.



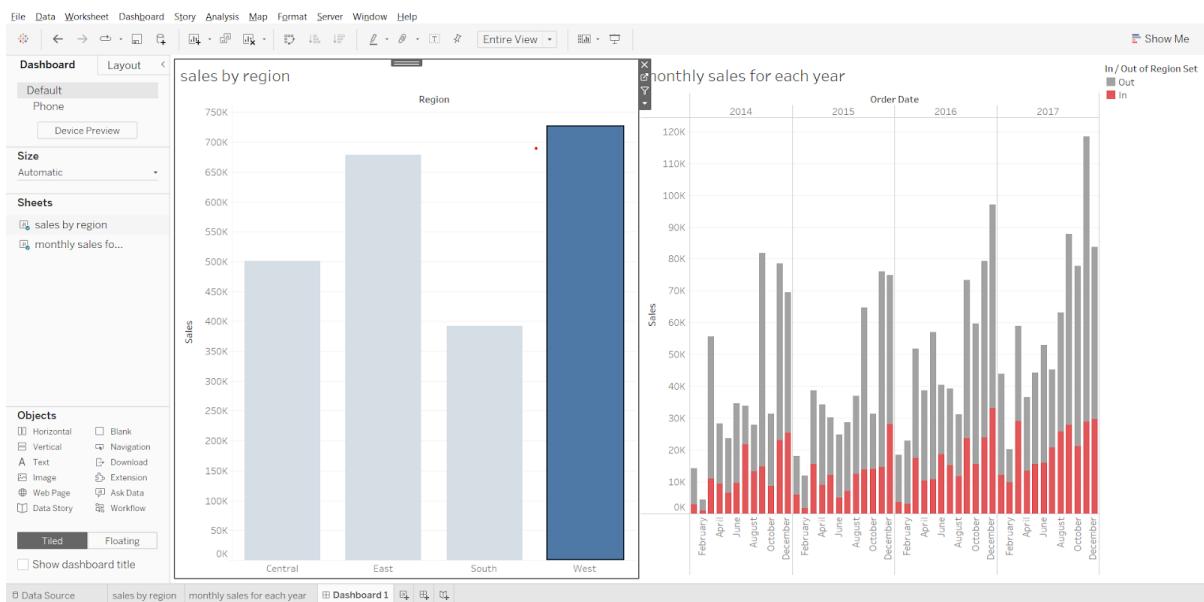
5. Creating a Dashboard

- Create a new dashboard
- Drag sales by region to the canvas
- Drag monthly sales for each year to the right of the canvas
- Set size to automatic



6. Creating Set action in Dashboard and solving the business problem

- In the Dashboard menu (from menu bar) click on Actions...
- Click on Add Action and select 'Change Set Values'
- In the new dialog box, deselect 'monthly sales for each year' in source sheet
- In target set, select orders(sample superstore) > Region set.
- Change 'clearing the selection will' to 'remove all values from set'
- Click ok in both boxes.
- You can change the colors used in 'monthly sales for each year' by going to the sheet, clicking on color in the marks card and choosing Edit colors...



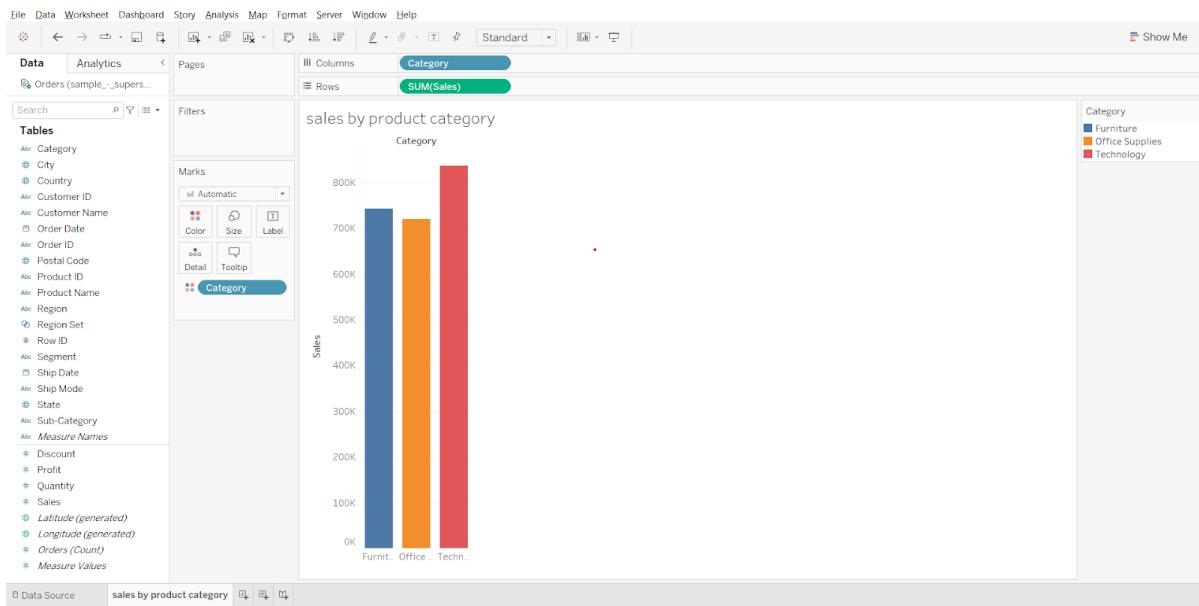
Filters Action

Demo:

Filter Out sales of products based on categories and subcategories

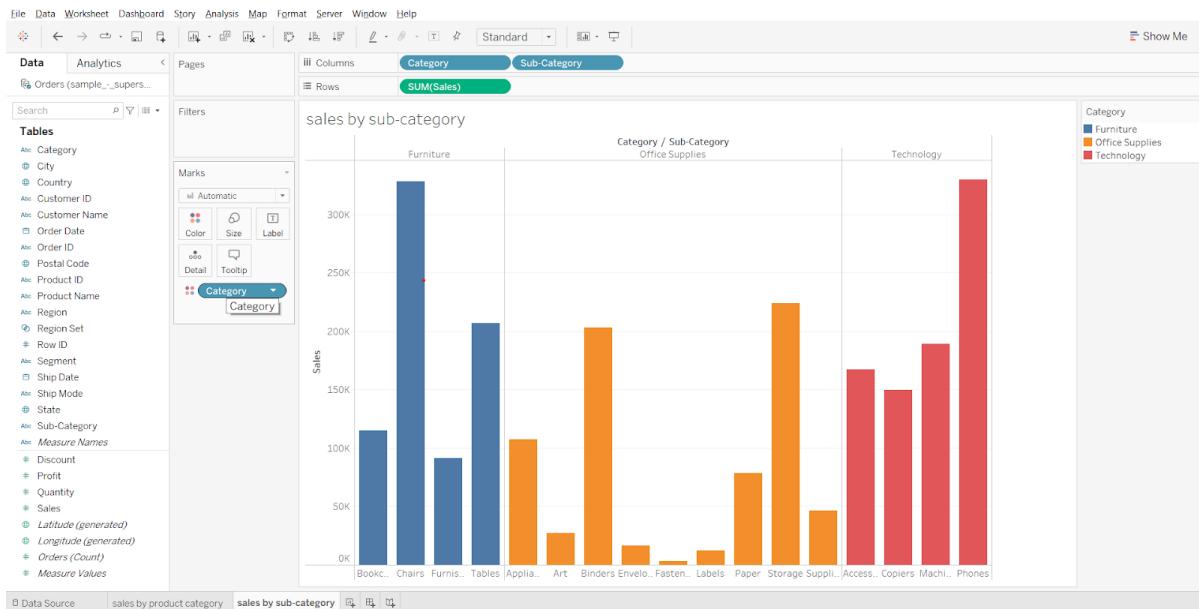
1. Creating Bar chart to visualize sales for each category

- Drag sales to rows shelf
- Drag category to columns shelf
- Drag category to color in marks card
- Rename sheet to sales by product category



2. Creating Bar chart to visualize sales for each sub category

- Create a new worksheet, name it 'sales by sub-category'
- Drag sales to rows shelf
- Drag category to columns shelf
- Drag sub-category to columns shelf
- Drag category to color in marks card



3. Creating Filter action and solving business problem in a worksheet

- Go to the worksheet in the menu bar, click on Actions...
- From Add Action, select Filter
- In the box that appears, select 'sales by product category' as source sheet
- Select 'sales by subcategory' as target sheet
- In 'run action on' section choose 'select' and check 'single-select only'
- In 'clearing the selection' section choose 'show all values'. Click ok in both boxes.
- Go to the 'sales by product category' sheet and click on any of the 3 bars in the chart. Observe the change in target sheet 'sales by sub-category'.

Edit Filter Action

Name: Filter1

Source Sheets: Orders (sample_-_superstore (1))

Run action on: Hover (radio button)

Marks: Automatic

Target Sheets: sales by sub-category

Clearing the selection will: Show all values (radio button)

Filter: All fields (radio button)

Source Field: Click to add

Target Data Source: Click to add

Target Field: Click to add

OK

Data Source: sales by product category

File Data Worksheet Dashboard Story Analysis Map Format Server Window Help

Category: Furniture, Office Supplies, Technology

Tables: Category, City, Country, Customer ID, Customer Name, Order Date, Order ID, Postal Code, Product ID, Product Name, Region, Row ID, Segment, Ship Date, Ship Mode, State, Sub-Category, Measure Names, Discount, Profit, Quantity, Sales, Latitude (generated), Longitude (generated), Orders (Count), Measure Values

Marks: Automatic, Color, Size, Label, Detail, Tooltip, Category

Rows: SUM(Sales)

Columns: Category

Filters:

Marks: Automatic

Color: Category

Size: SUM(Sales)

Label:

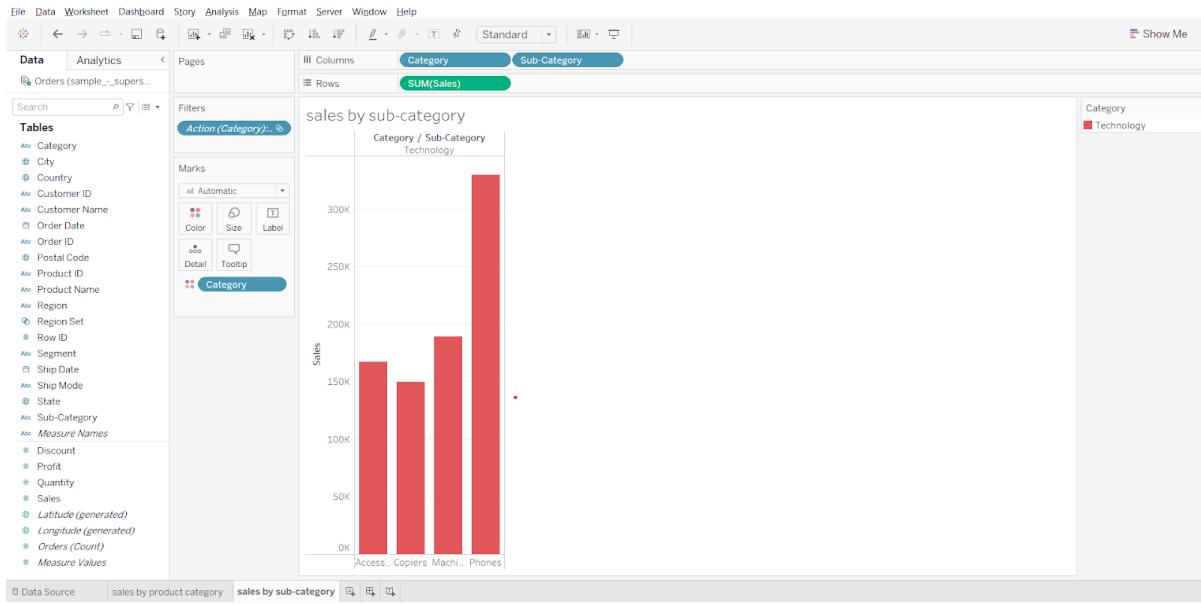
Detail:

Tooltip:

Category: Furniture, Office Supplies, Technology

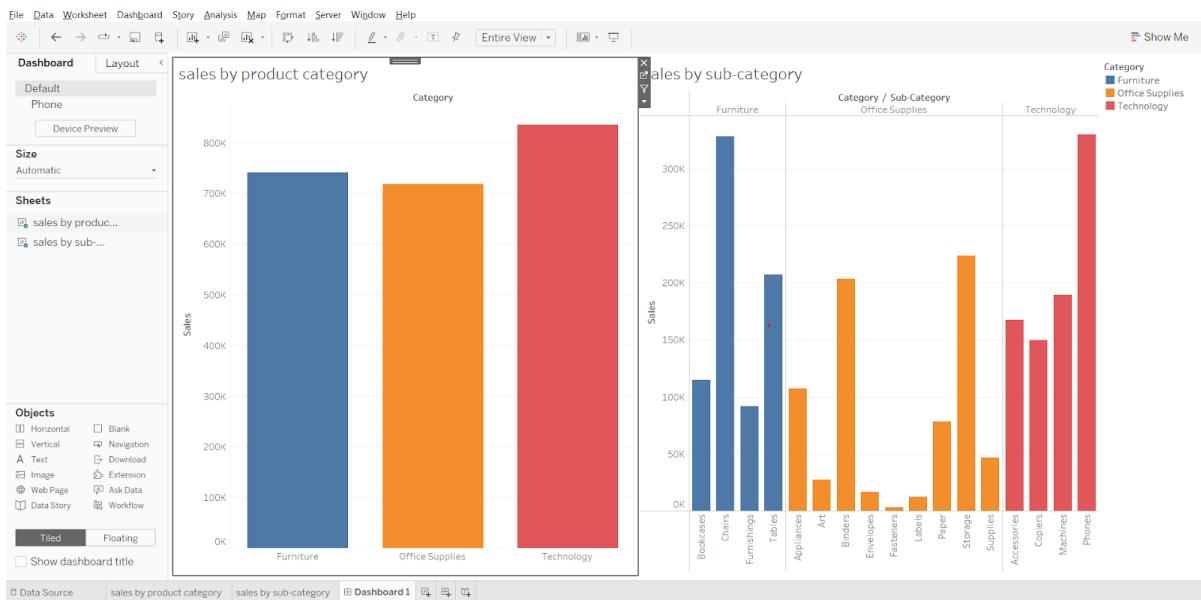
sales by product category

Category	Sales
Furniture	~750K
Office Supplies	~720K
Technology	~850K



4. Creating a Dashboard

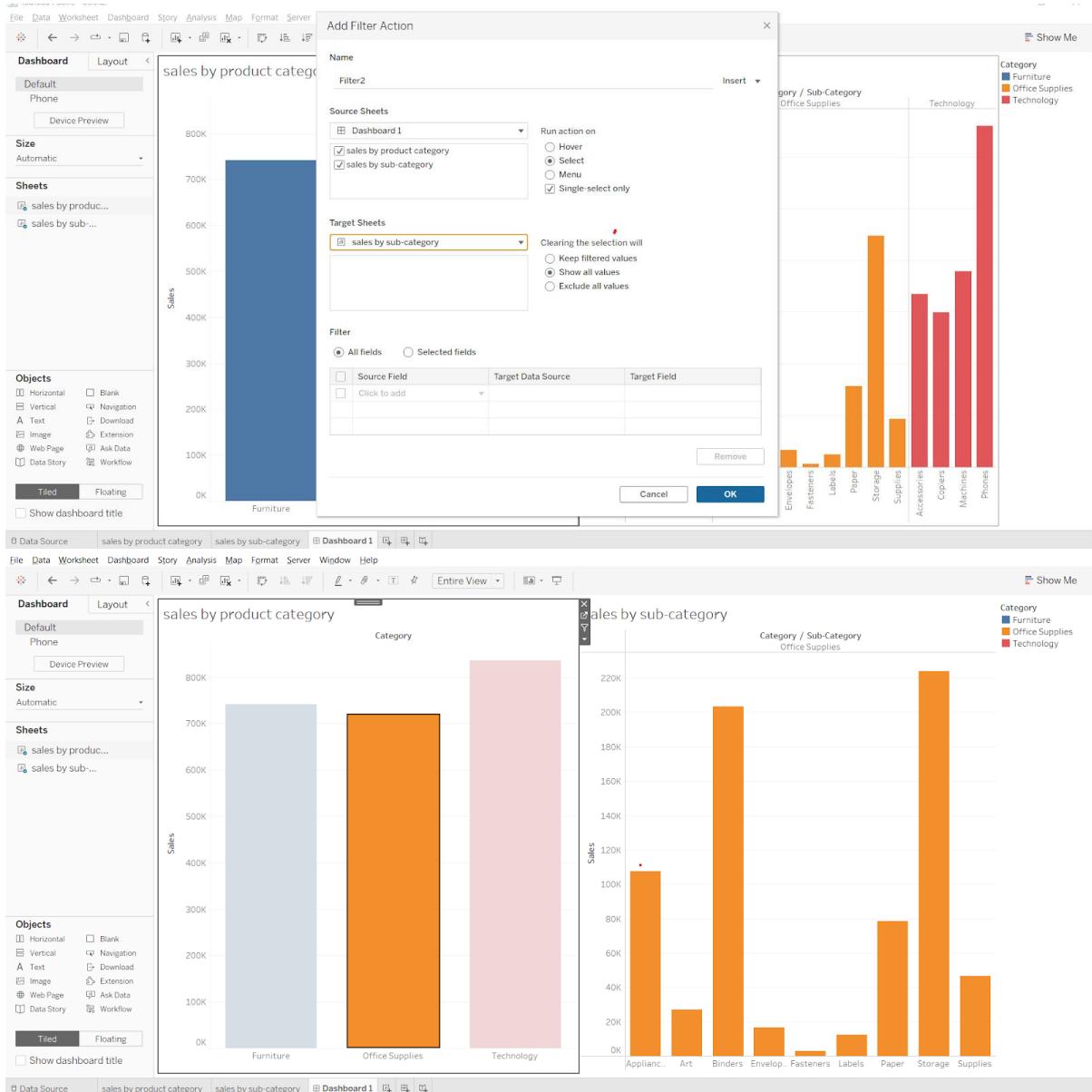
- Create a new dashboard
- Drag sales by product category to the canvas
- Drag sales by subcategory to the right side of canvas, but left of the legend (because both charts have the same legend).
- Set size to automatic.



Demo in Dashboard :

- In the dashboard created above, go to the dashboard in the menu bar, click on Actions...
- In Add Action, select filter
- In source sheets, select Dashboard 1, then select both the sheets.

- Rest of the steps are the same as the previous filter. (See screenshot below)
- Click on different bars in the 'sales by product category' container to see the results.



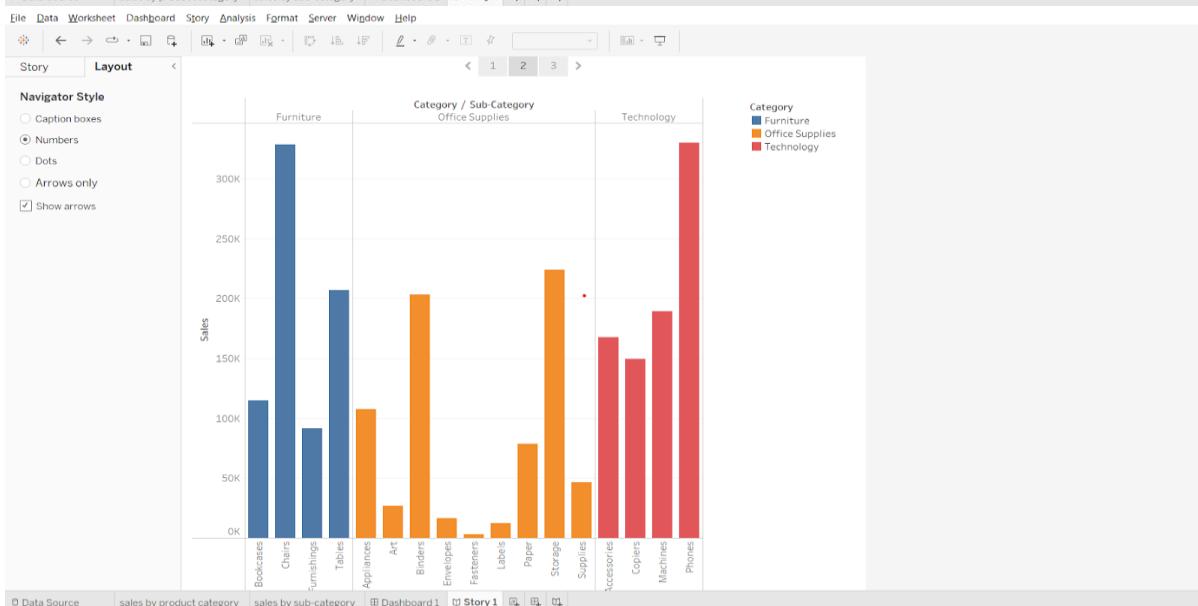
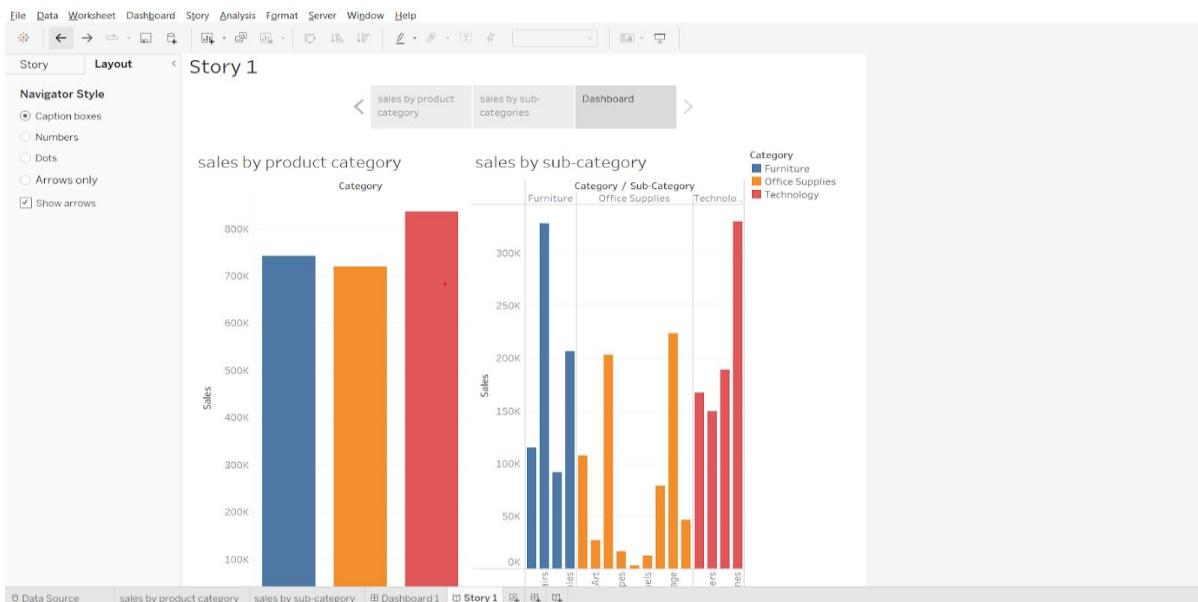
Story Points

- Use stories to make your case more compelling by showing how facts are connected, and how decisions relate to outcomes. You can then publish your story to the web, or present it to an audience.
- Each story point can be based on a different view or dashboard, or the entire story can be based on the same visualization seen at different stages, with different filters and annotations

Demo :

Open a new story tab.

- The sheets and dashboards created earlier are shown in the left pane.
- Drag a sheet on the canvas
- Click on 'Blank'
- Drag another sheet on the new canvas.
- Write a caption for the 2 story points.
- Go to layout and pick a desired navigation style.



Dashboards and Stories (Business Problem)

- Dashboard 1 :
 - Find out which state is losing money and get an overall picture/overall summary.
- Dashboard 2:
 - Find the City and the subcategory where it is losing money.
- Dashboard 3 :
 - For the selected city and subcategory, find why it is losing money.

Story :

- Summary - Find out which state is losing money and get an overall picture/overall summary
 - Answer : Texas
- What went wrong in Texas - Find the City and the subcategory where it is losing money
 - Answer : Houston & Binder
- Product Analysis - Analysis of the products under the category Binder if they are all losing money
 - Answer : Yes
- Why so? – For the selected city and subcategory, why is it losing money?
 - Answer : All the products that are there in subcategory binder and for city houston it is giving an average discount of 0.8 or 80%

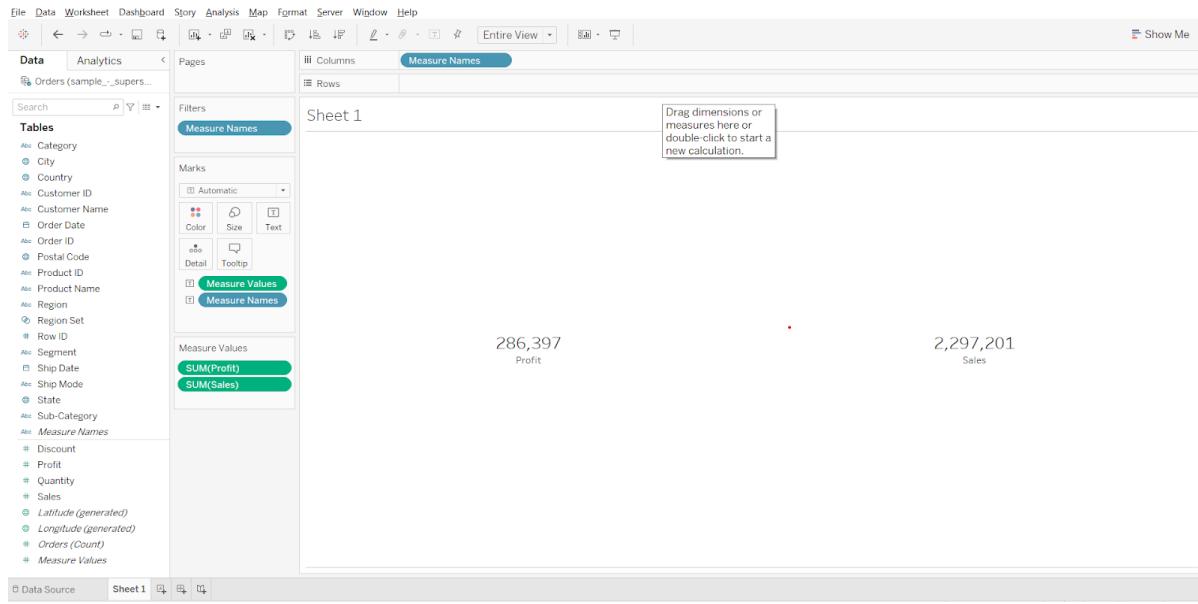
Dashboard 1 :

- Find out which state is losing money and get an overall picture/overall summary
 - Show overall summary for the profit and sales value as text
 - Show state wise profit and sales in single map (<https://youtu.be/x-k6PYernug>)
 - Display sales and profit for each subcategory using heat map
 - Have a Order date filter through which we can select year and the visualization will be updated based on the year selected

Sheet1 -

Overall summary of Profit and sales for US.

- Drag measure names to columns
- Drag measure values to text in marks card
- From measure values card remove all measures except sales and profit
- Change fit to entire view
- Drag measure names to text in marks card
- Click on text in the Marks card, open the dialog box, change size of 'measure values' to 16.
- Click on text, change horizontal alignment to center.
- Right click on header, uncheck 'show header' option.

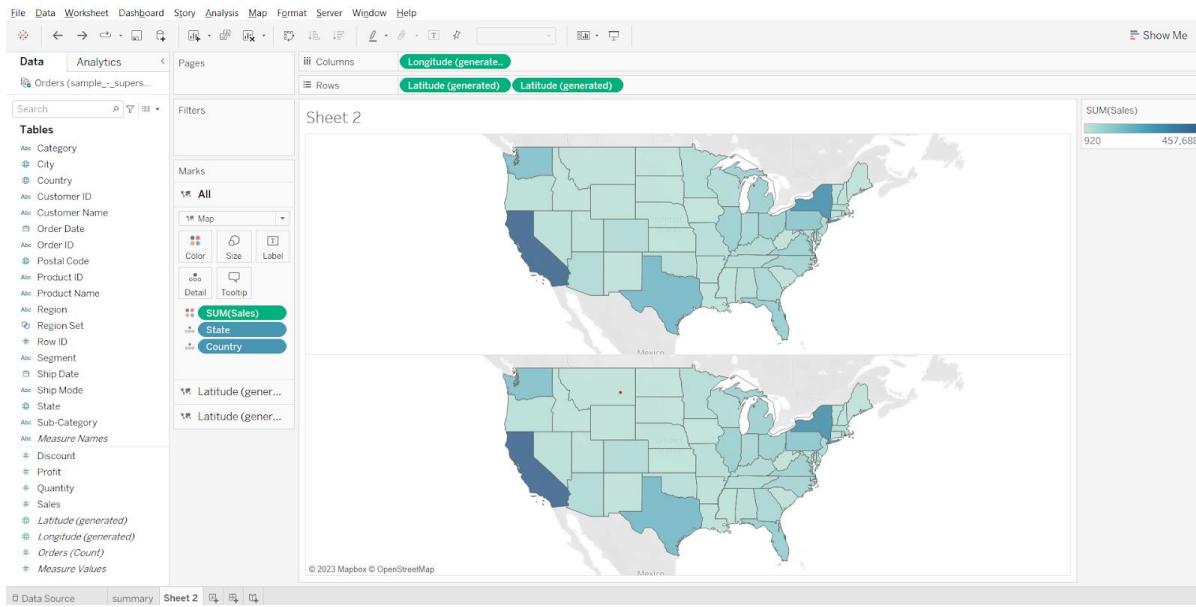


- Rename the sheet as summary

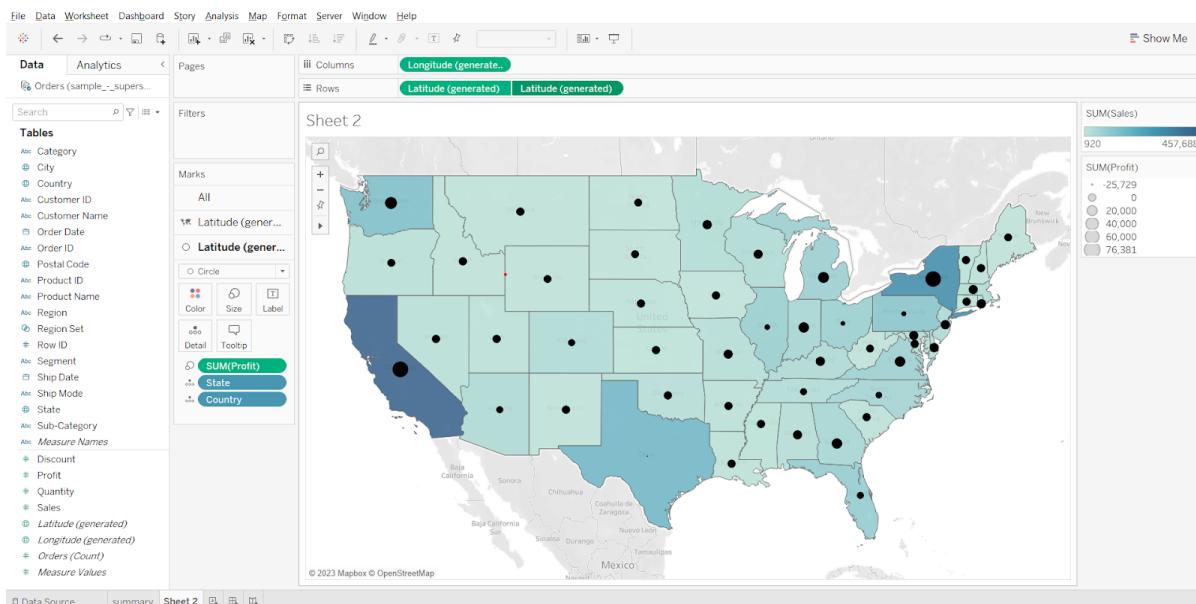
Sheet 2 -

State wise profit and sales using map visualization.

- Drag longitude to columns
- Drag latitude to rows
- Drag state to detail in marks card
- Click on the '49 unknown' box in the bottom right corner.
- Click edit locations, change country/region to from field > country. Click OK.
- Change chart type to map
- Drag Sales to color in marks card
- Drag latitude to rows shelf to get 2 graphs



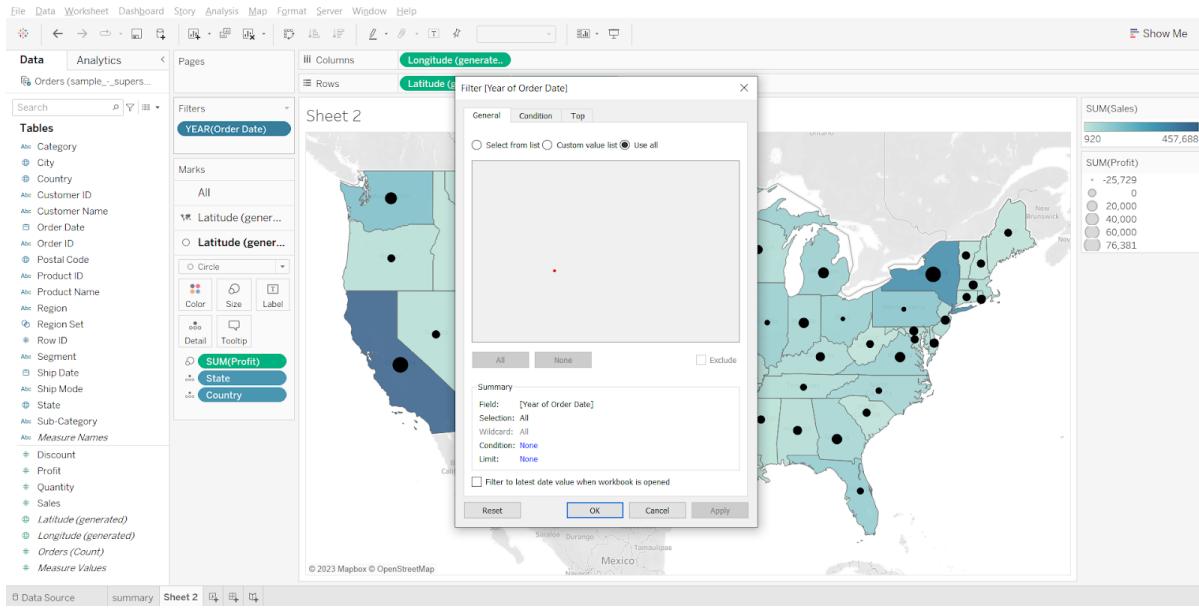
- Click on the second Latitude chart in the marks card.
- Remove sales, drop profit on size.
- Change mark type to circles.
- Click on color and choose black.
- Click on size and adjust the size of the circles.
- Open the latitude dropdown in the rows shelf and choose 'Dual Axis'.



From the plot we can see that Texas has the lowest profit compared to the rest of the states.

Adding a filter for order date :

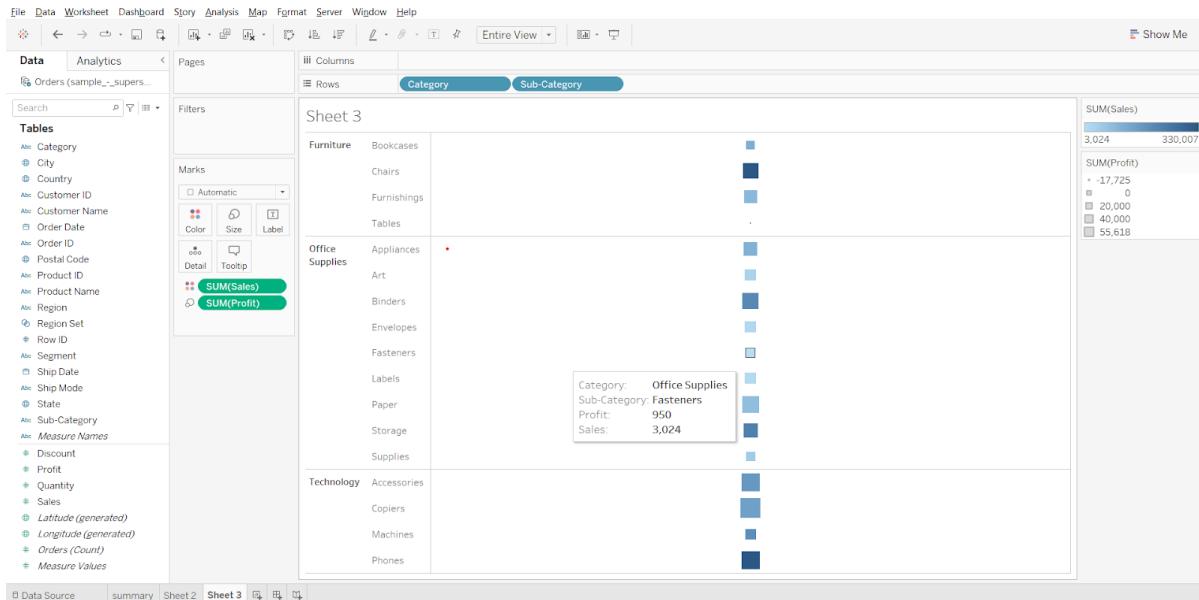
- Drag order date to filter card, select year, click next, select ‘use all’, click OK.



Sheet 3 -

Subcategory wise profit and sales.

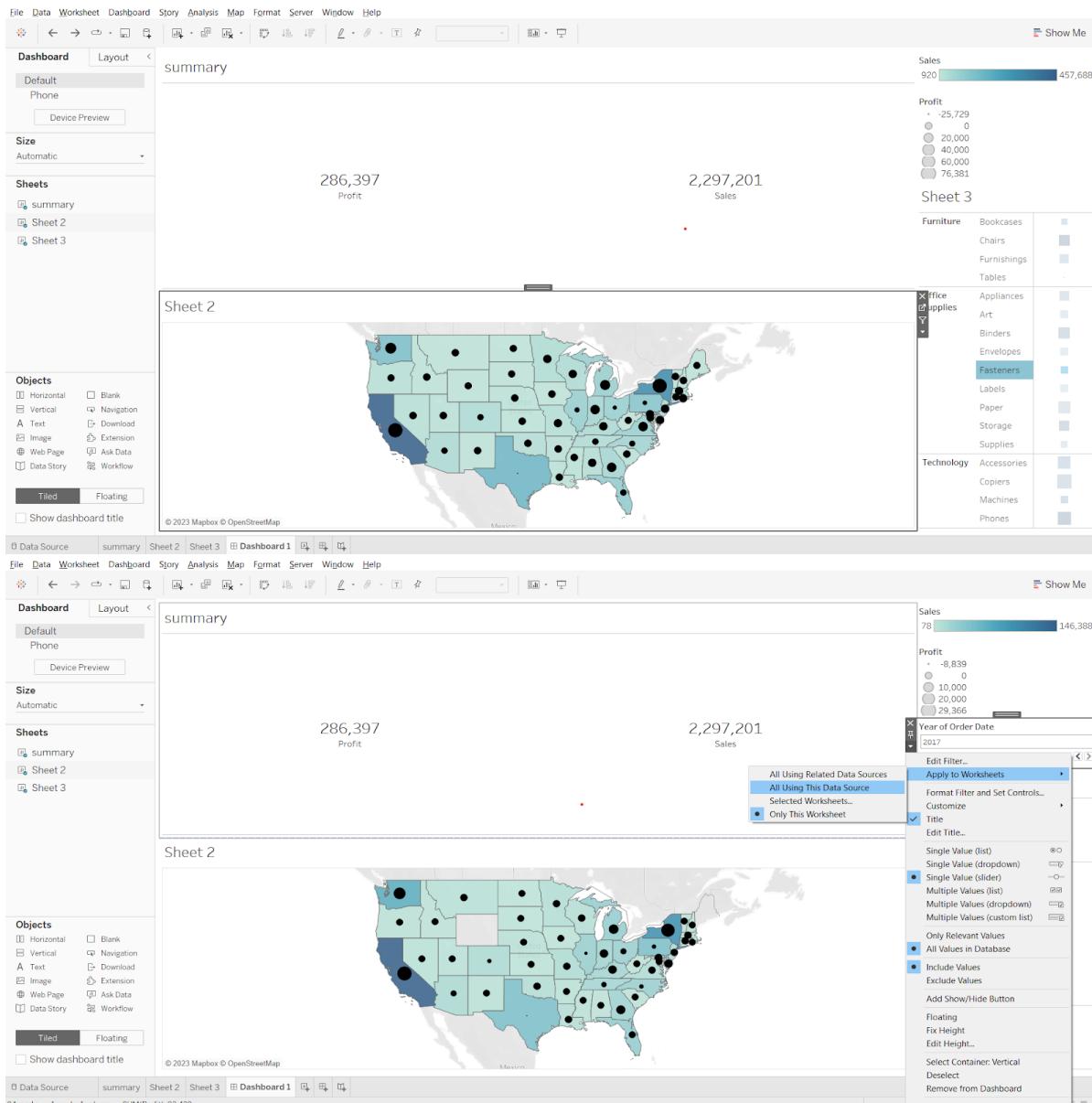
- Drag category and sub-category to rows, drop sales on color, profit on size, adjust size as per need.
- Change fit to entire view, right click on header in sheet, remove header labels.



From the above dashboard we can see that for the state of Texas we have negative profit and for subcategory we have the negative profit so let's analyze further why this is so.

Adding order date filter for the dashboard :

- Create a new dashboard, set size to automatic, drop all 3 sheets on the canvas as shown in the screenshot below, open sheet 2 menu (from top right corner of sheet 2 container), select filters > year of order date.
- Right click on the filter and choose a single value slider. Open the menu from the filter container, select Apply to worksheets > All using this data source.



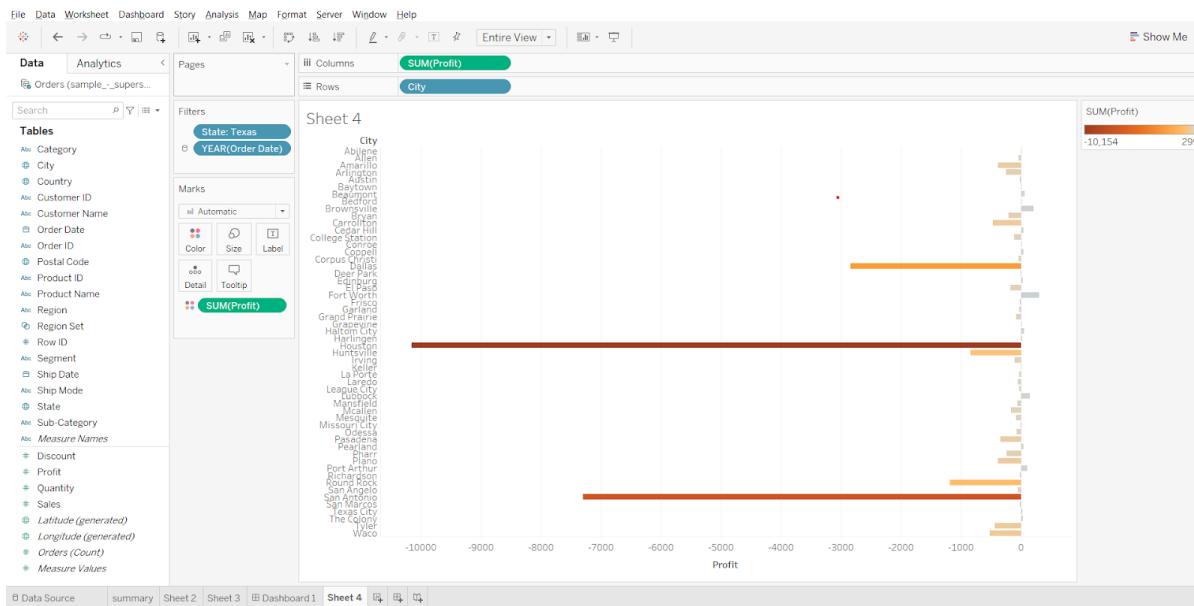
Note: Keep the Order date filter field to All for further analysis.

- Selecting State-Texas and Subcategory-Table and profit field for further analysis

Sheet 4 -

Plotting city wise profit for the state of Texas.

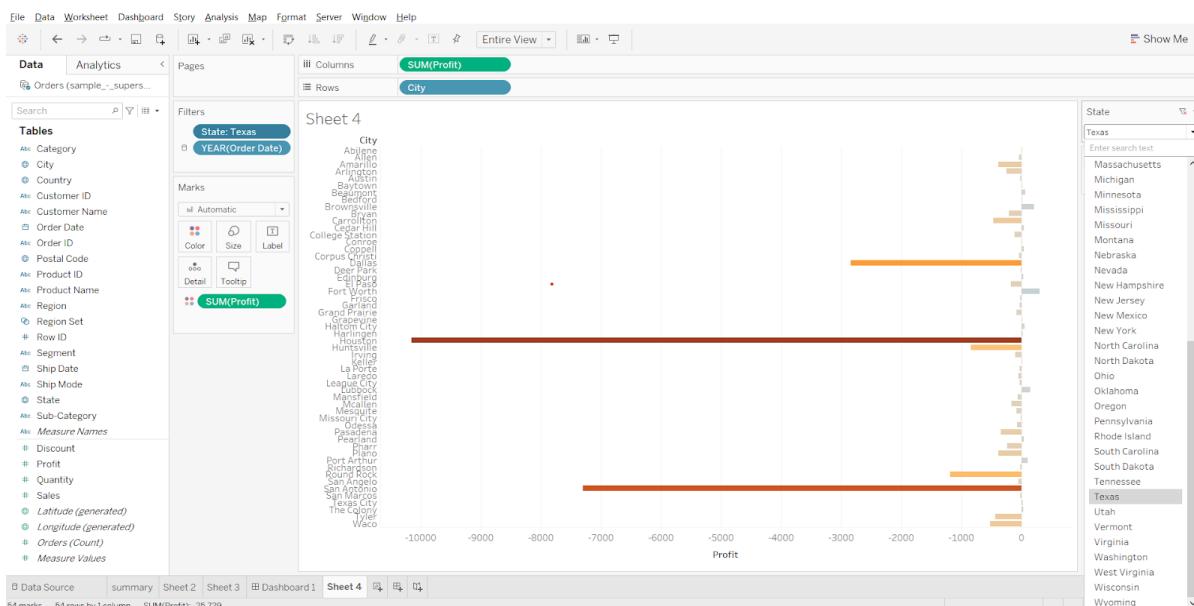
- Drag state to filter, select Texas, click OK. Drag city to rows, profit to columns, profit to color. Change fit to the entire view.



We can see that the city Houston is where we are losing more money.

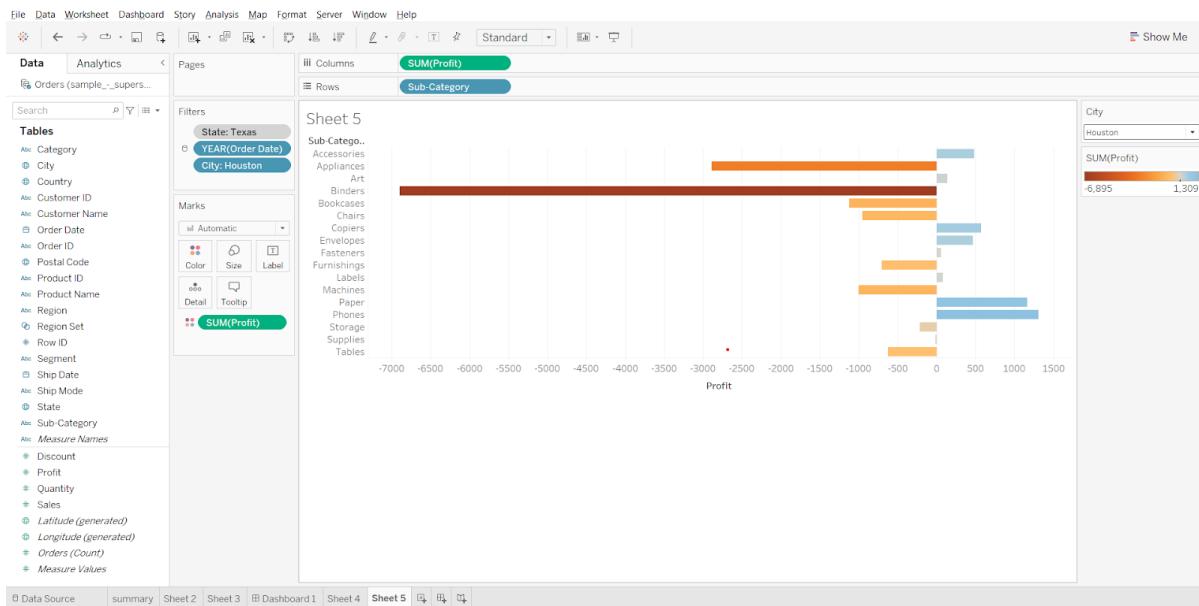
Showing city interactive filter :

- Open the State menu in the filter card, choose show filter.
- Convert the checklist into a single value dropdown.



Sheet 5 -

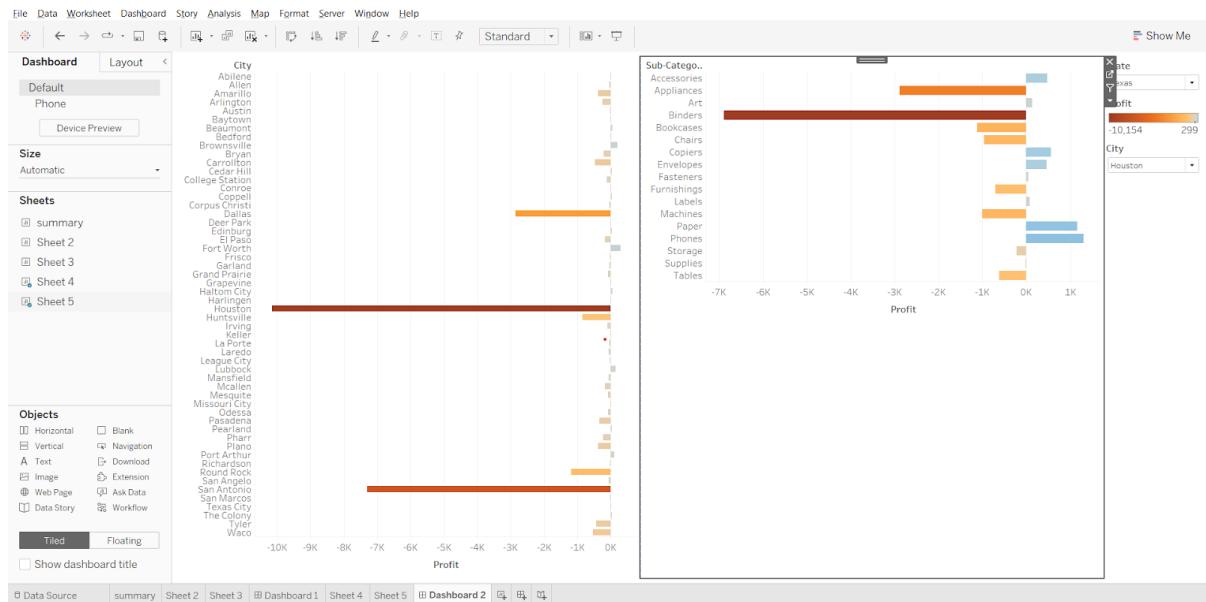
- Drag subcategory to rows, profit to columns. Drag state to filters, select Texas, click ok.
- From the State dropdown in filters, add it to context. Drag city to filters, from dropdown select show filter, convert the checklist to a single value dropdown. Drag Profit to color.
- From city dropdown select Houston.



We can see that for city houston we are losing more money in subcategory binders.

Creating a dashboard :

- Create a new dashboard, set size to automatic, drop sheet 4 and sheet 5 on the canvas. Hide title for both sheets.

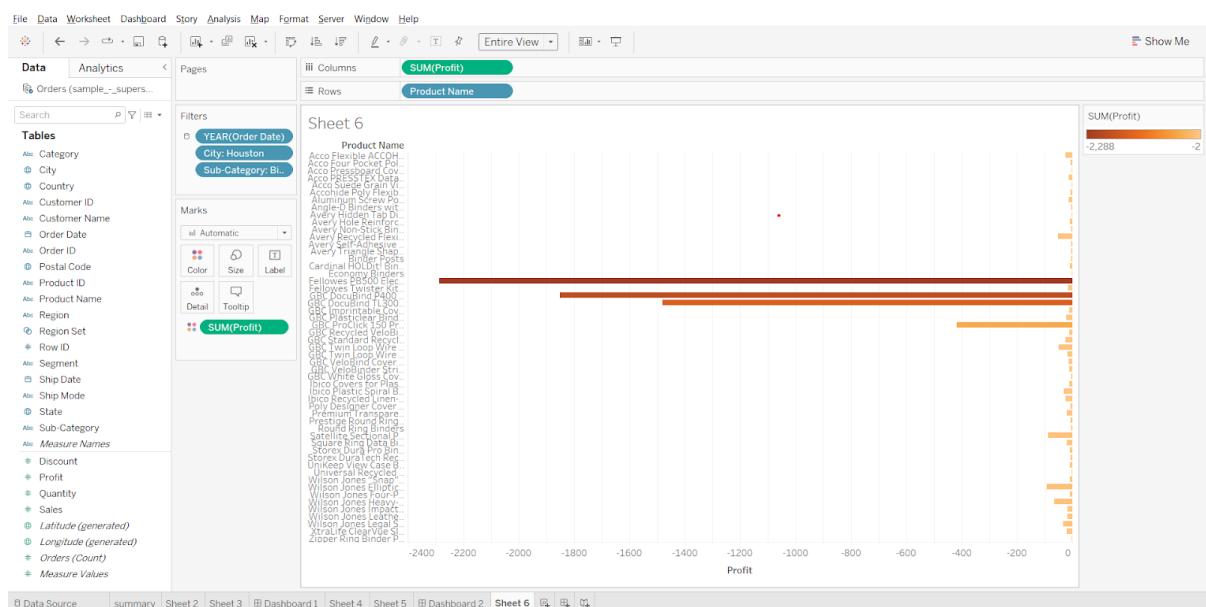


So we can see that for the state of Texas in the city of Houston we are losing money in subcategory binder so let's further analyze exactly which product we are losing money in.

Sheet 6 -

Creating products wise profit for city houston and subcategory binders.

- Drag product name to rows, add all members. Drag profit to columns, drag city to filters, select houston, click OK.
- Drag subcategory to filters, select binders, click OK. Drag profit to color, change fit to Entire View.

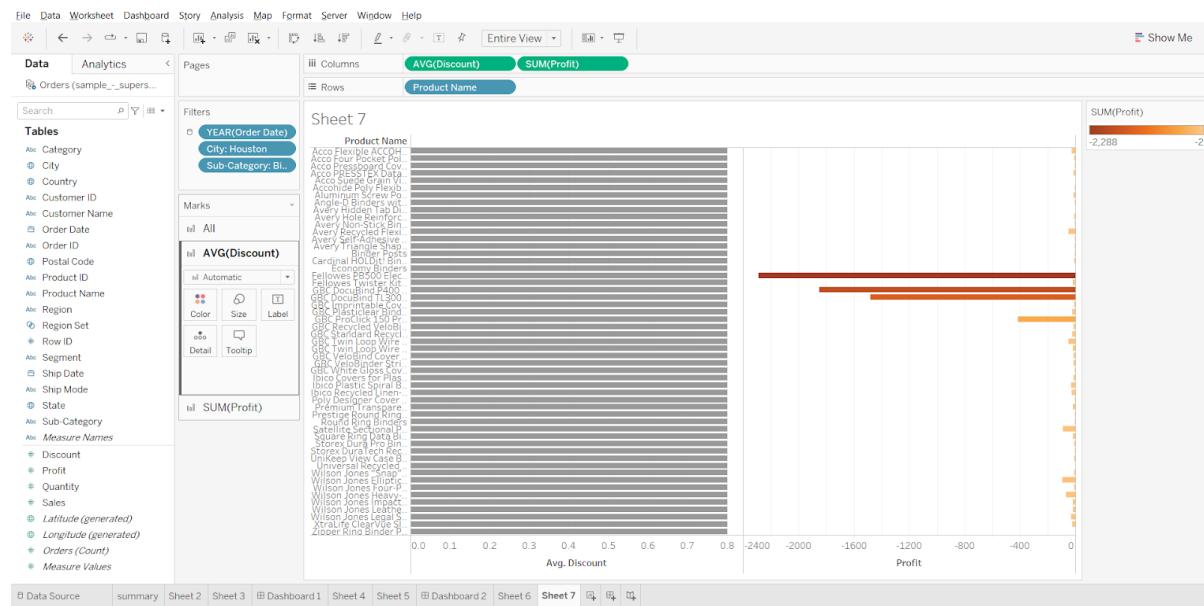


So we can see that within binders there are 3 products where we are losing more money, mainly there are binding machines. Let's analyze why that is so.

Sheet 7 -

Plotting average discount vs profit for each product within subcategory binder and for city houston.

- Duplicate sheet 6, rename it as Sheet 7, drag discount to columns before profit. Change its measure to average.
- From discount marks cards, remove profit from color.

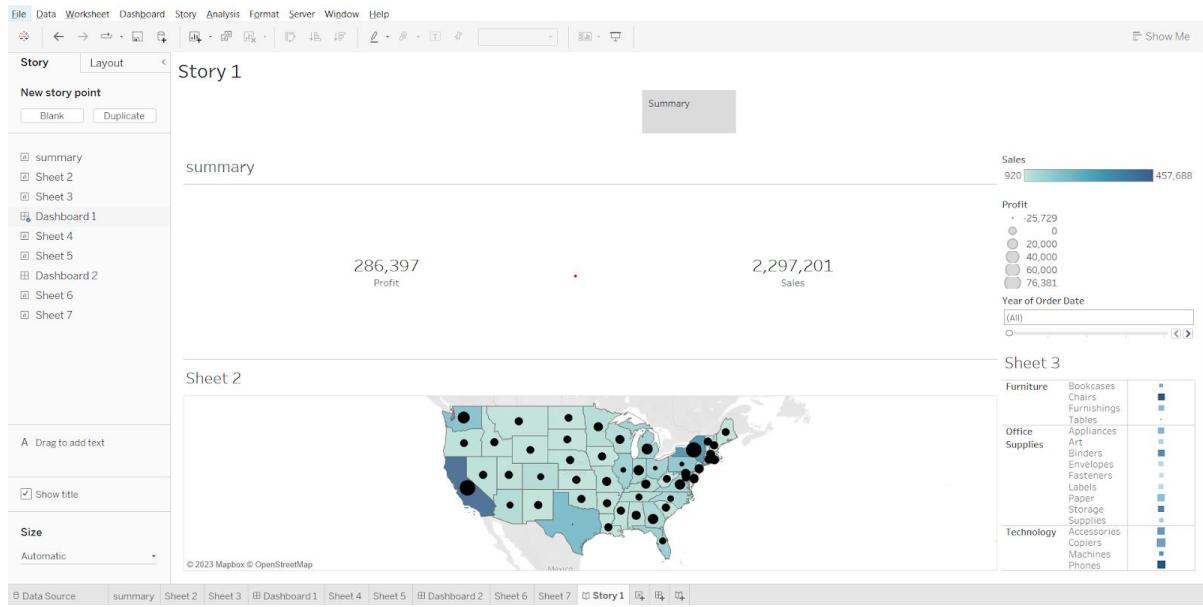


So we can see that for city houston and for subcategory binders all the products have an average discount of 0.8 or 80% on them which is why the store in houston is losing more money in houston city.

Creating a story for above analysis:

The 1st tab is the sales and profit summary for the US.

- Create a new story, add dashboard 1 to the canvas, set size to automatic, write 'Summary' in caption



The 2nd tab is analyzing what went wrong in Texas.

So we can see that for city Houston and for the binder subcategory we are losing money.

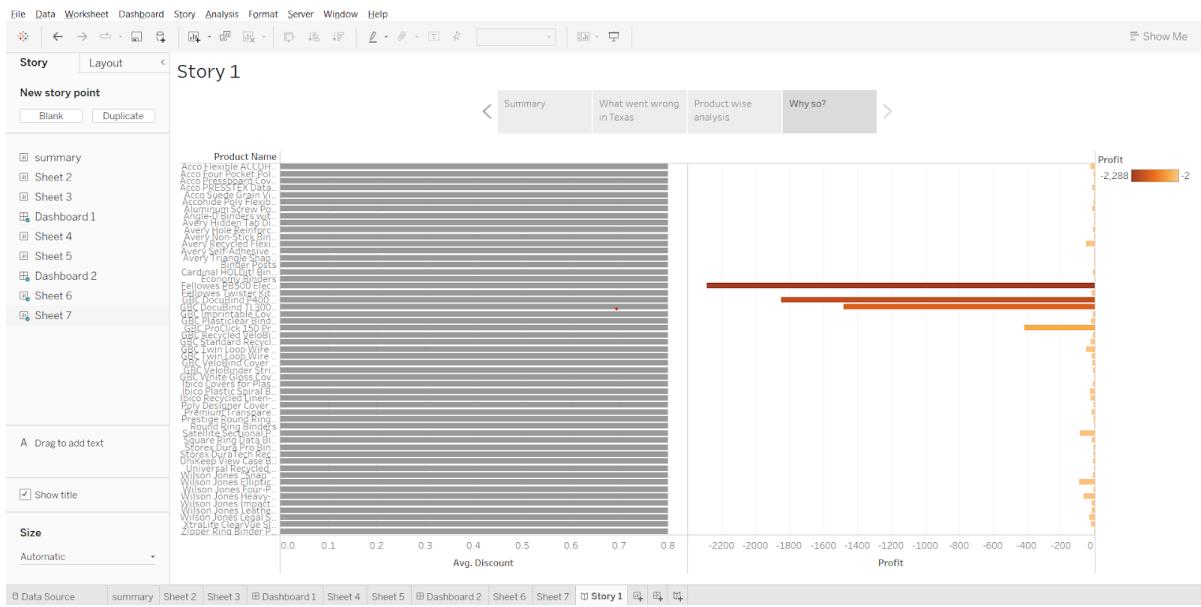
- Create a new story point, write 'What went wrong in Texas' as caption, drag dashboard 2 to the canvas.

The 3rd tab we are diving deeper to find for which products are we losing money for Houston city and Binders category.

- Create a new story point, type 'Product wise analysis' in caption, drag sheet 6 to the canvas.

We can see that almost all the products that come under the Binder category are losing money. So let see why is that so?

- Create a new story point, type 'Why so?' in the caption and drag sheet 7 to the canvas.



So we can see that for all the products that are there in subcategory binder and for city Houston it is giving an average discount of 0.8 or 80% which is why it is losing more money compared to other subcategories.

To publish the workbook, press Ctrl + S, sign in, name the workbook.

How to increase efficiency of Tableau Dashboards

- Your data strategy drives performance
 - Minimize the number of fields
 - Minimize the number of records
 - Optimize extracts
- Reduce the marks (data points) in your view
 - Practice guided analytics.
 - Remove unneeded dimensions
 - Explore.
- Limit your filters by number and type
 - Reduce the number of filters in use.
 - Use an include filter
 - Use a continuous date filter.
 - Use Boolean or numeric filters.
 - Use parameters and action filters.
- Optimize and materialize your calculations
 - Perform calculations in the database.
 - Reduce the number of nested calculations.

- Reduce the granularity of LoD or table calculations in the view. The more granular the calculation, the longer it takes.
 - LoDs
 - Table Calculations
 - Where possible, use MIN or MAX instead of AVG. Make groups with calculations.
 - Use Booleans or numeric calculations instead of string calculations.
- Take advantage of Tableau's query optimization
 - Blend on low-granularity dimensions.
 - Minimize joined tables.
 - Assume referential integrity
 - Remove custom SQL.
- Clean up your workbooks
 - Reduce dashboard scope.
 - Delete or consolidate unused worksheets and data sources.

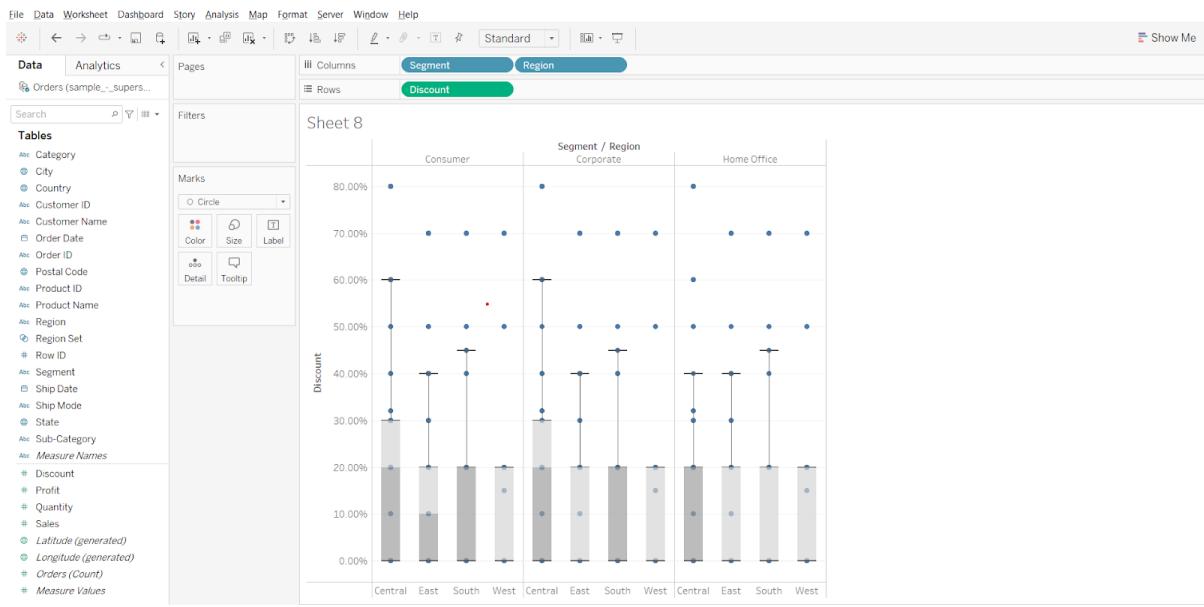
Box and Whiskers plot

- Also called as Boxplot
- Used to plot a distribution of values for a selected measure
- The Box contains the lower quartile, upper quartile and the median in the center
- You can configure the Upper and lower lines called as whiskers to be to be within 1.5 times the Interquartile range or all the points at the maximum extent of the data
- [Building blocks for Box and Whisker plot](#)

Business problem 1:

Compare the average discount given to each segment across various regions.

- Creating a bar chart for each region and segment
 - Drop segment and region on columns, drop discount on rows. Change the measure of discount to average.
- Creating a Box and whiskers plot
 - Go to 'show me', click on box and whiskers plot
- Solving the business problem
 - Change format of vertical axis values to percentage, drag region from marks card to columns.
 - Go to Analysis in the menu bar, uncheck aggregate measures.



Insight: We can see that the median discount value for west region for all the segments is 0 we can say that west region has lowest discount values compared to other region for each segment

You can edit the properties of the box and whiskers plot by right clicking on one of the boxes in the plot and clicking on Edit...

Bullet chart

- A bullet graph is a bar marked with extra encodings to show progress towards a goal or performance against a reference line.
- A bullet graph is useful for comparing the performance of a primary measure to one or more other measures.

Business problem 2:

Compare 2017 and 2016 sales for each product subcategory.

- First we will create two calculated fields to get 2016 sales and 2017 sales.

Sales 2016= IF YEAR([Order Date])=2016 THEN [Sales] END

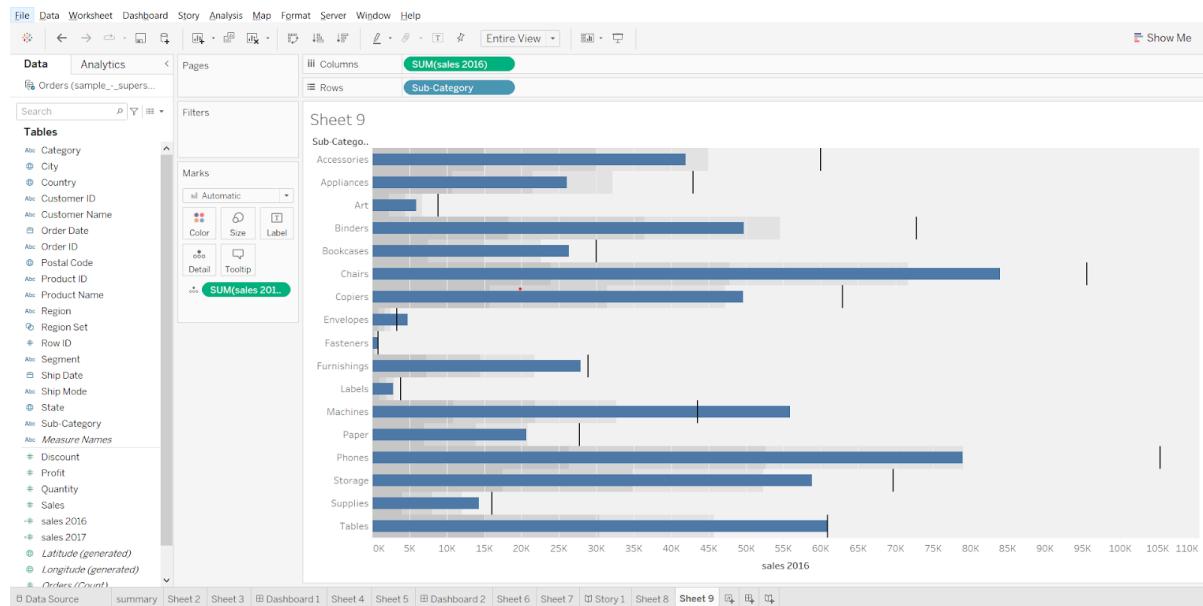
Sales 2017= IF YEAR([Order Date])=2017 THEN [Sales] END

- Create a bullet chart using the show me card
 - Drop sales 2016, sales 2017 on columns, subcategory on rows
 - Go to show me and select the bullet chart, change fit to the entire view.

- Editing the properties of bullet chart
Changing the vertical reference bands to display total sales in 2017
- Right click on the horizontal axis, choose Edit reference line > Average sales 2017
- Change measure to total, click OK.

Change the background bars from 60, 80 to 25, 50, 75 % of sales in 2017, here the blue bars denote the total sales in 2016.

- Right click on the horizontal axis, go to Edit reference line > 60%, 80% of Average Sales 2017
- Change percentage values to 25, 50, 75. Change measure to total. Click ok.



We can swap the reference lines so that the blue bars represent sales in 2017 and the vertical reference line and the background will represent 25,50 and 75% of total sales in 2016.

To do so, right click on the horizontal axis, choose 'swap reference lone fields'.

Tableau Capstone Project introduction

Doc: [link](#)

1. You'll get assignment questions on this case study under the homework section from the first lecture of Excel.
2. You do not have to submit this project and there is no specific evaluation that will be done for the case study as of now.
3. You are not allowed to share this data publicly as it contains sensitive information about the company.

4. If you have created a dashboard and want to save it then you can upload the same on tableau public then download it on to your local system and delete the dashboard from tableau public website.
5. You are not allowed to showcase this project in your resume. What you can do is take a dataset that is publicly available and showcase that as a project in your resume.