## **Tableau Lecture 2: Basic Charts and Operations**

#### **Agenda**

* Visual Analytics
* Data Connections in Tableau
  + Live connection
  + Extract connection
* Aggregation on Dimensions and Measures
* Basic Charts
  + Line chart
  + Bar chart
  + Pie chart
  + Histogram
* Cleaning and Formatting data for analysis
  + Data Interpreter
  + Pivoting Data
  + Data Source Filter
* Introduction to Summer Olympics dataset

**Dataset**: [**sample superstore**](https://docs.google.com/spreadsheets/d/15W0gVPEHCDDnrKHpgj9XlUTIpNJsS9qX/edit?usp=share_link&ouid=116847520093376081608&rtpof=true&sd=true)

### **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 dimensions and measures in Tableau.
* [Types of aggregation on measures](https://help.tableau.com/current/pro/desktop/en-us/calculations_aggregation.htm#AggFuncs)
* We can convert a dimension to a measure and perform aggregation on it.
* [Types of aggregation on dimension](https://help.tableau.com/current/pro/desktop/en-us/calculations_aggregation.htm#:~:text=its%20context%20menu%3A-,Aggregating%20Dimensions,-You%20can%20aggregate)
* If you want to remove the default aggregation for fields in the view, go to **Analysis** and uncheck the **Aggregate Measures** option.

### **Business problem 1:**

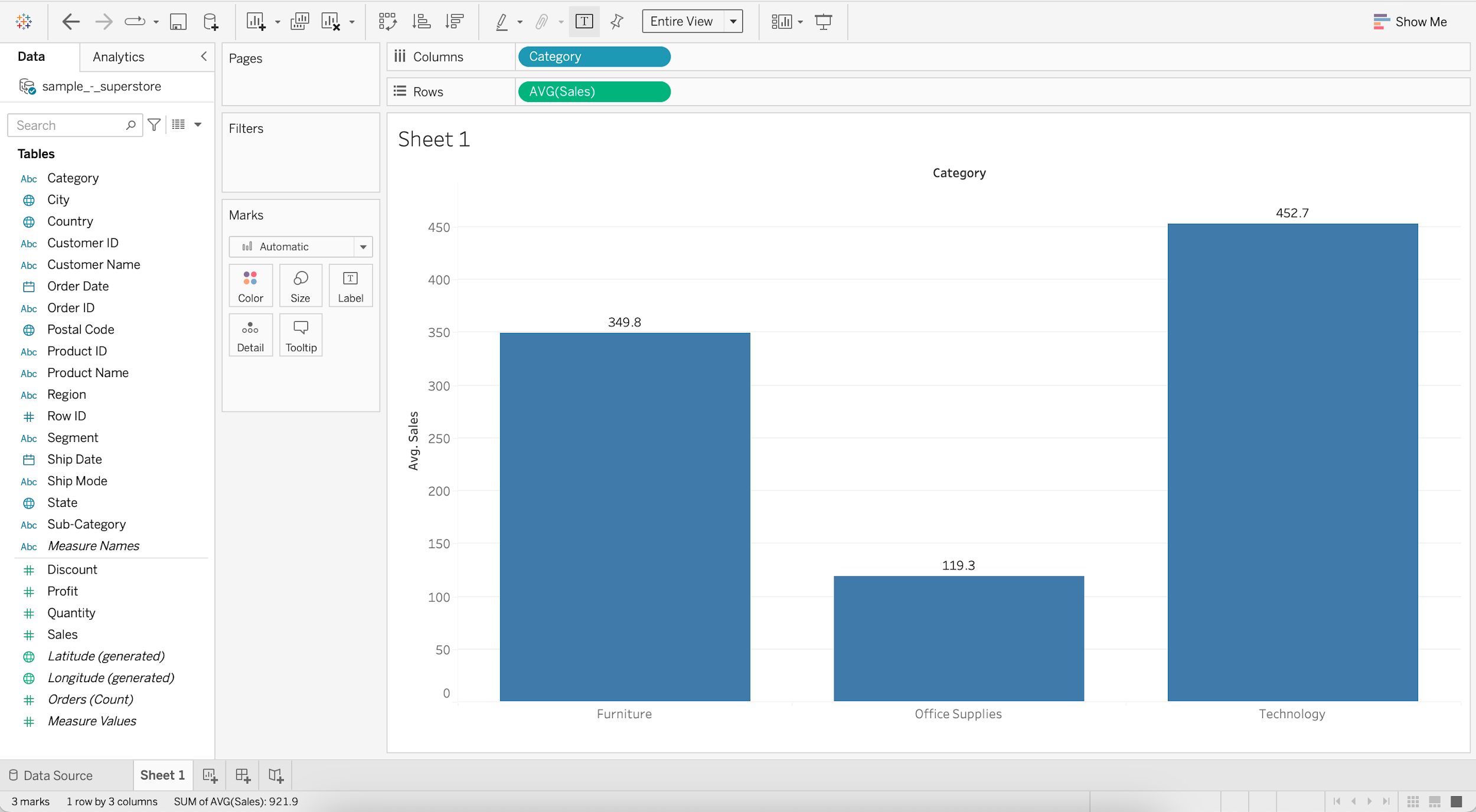
Determine the average sales value of each category.

**How to choose what goes to Rows shelf and what goes to Columns shelf?**

* **Category** being a **dimension** by which we wish to group the data, goes to **Columns** shelf (X-axis).
* **Sales** being a **measure** that we want to aggregate (AVG), goes to **Rows** shelf (Y-axis).

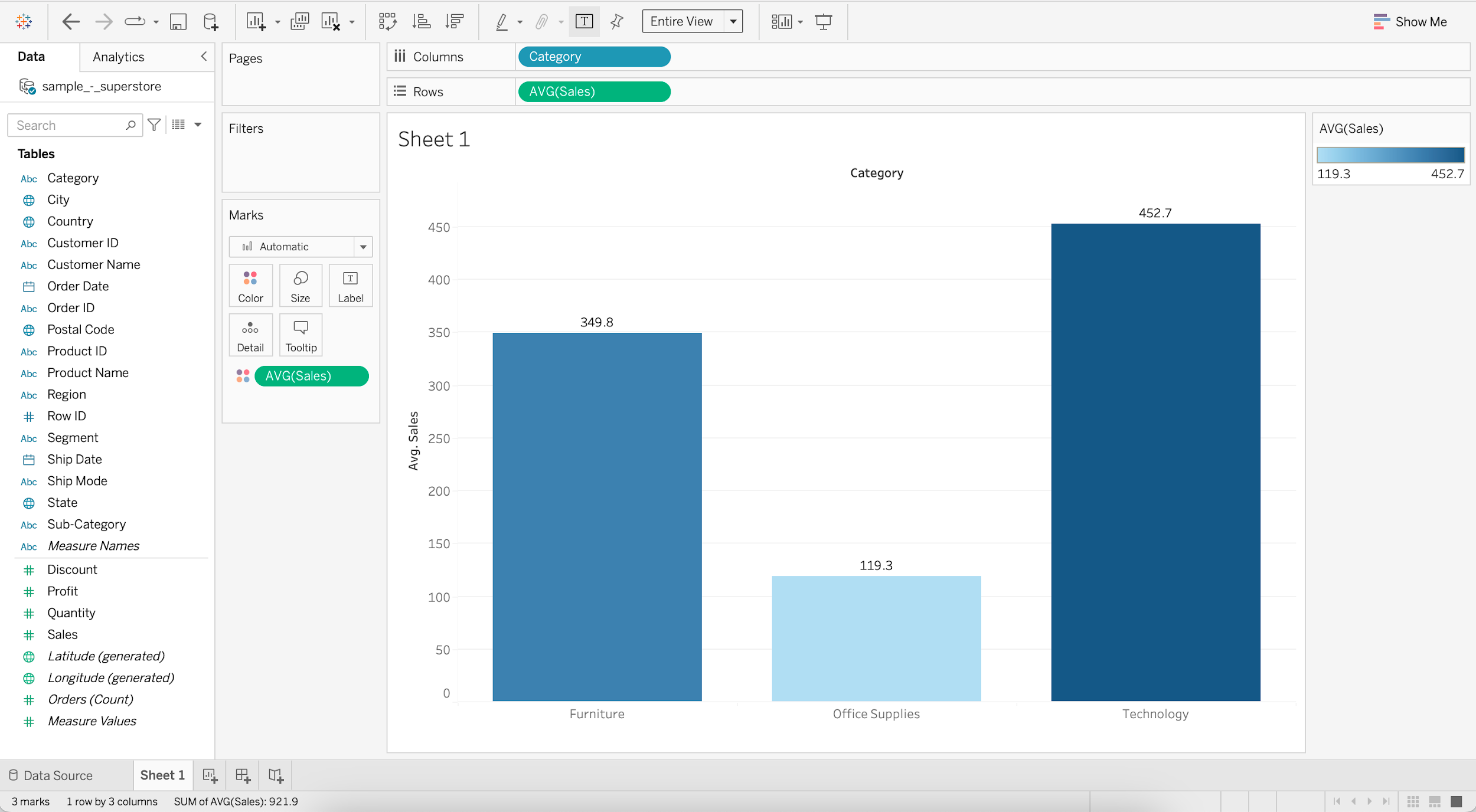
**Steps:**

* Drag **Category** to Columns
* Drag **Sales** to Rows
* Change aggregation of **Sales** from Sum to Average

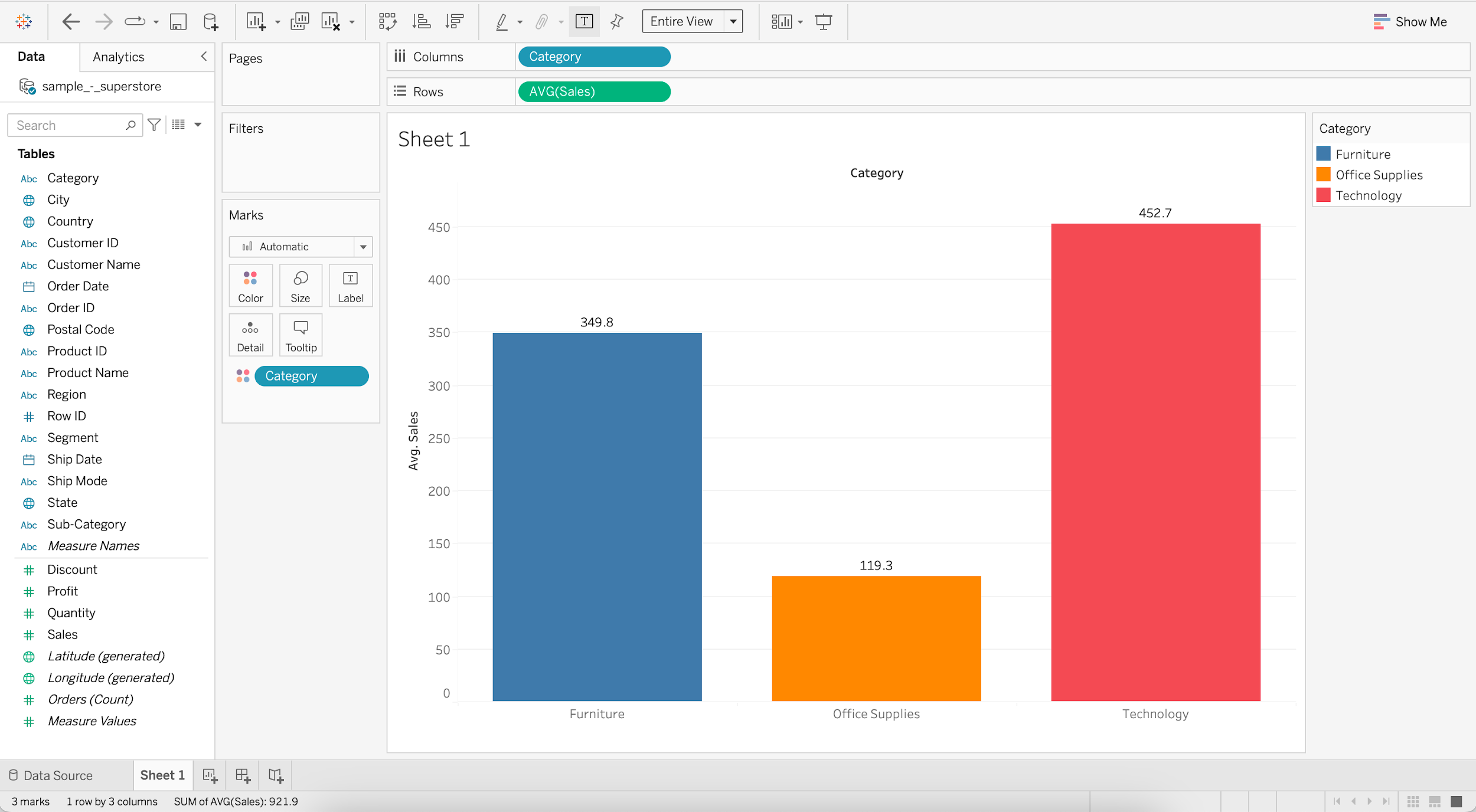


#### **We can demonstrate Color marks shelf -**

* If we drag **Sales** (Measure) onto the **Color** marks shelf, it’ll assign different shades of a single color to the **AVG of Sales** values.



* If we drag the **Category** (Dimension) to the **Color** marks shelf, we’ll get a different color for each category.



### **Business problem 2:**

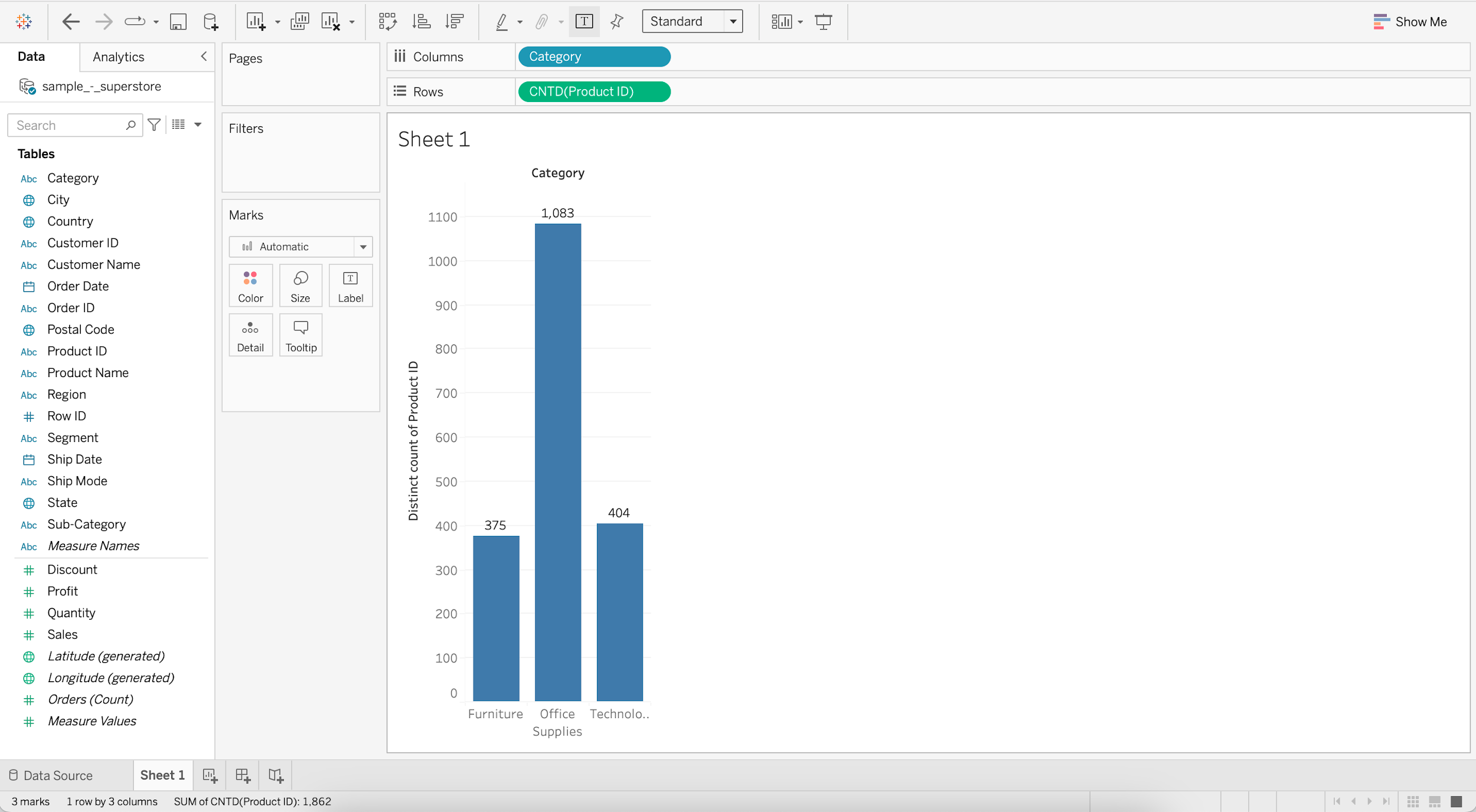
Determine total number of products within each category.

**How to choose what goes to Rows shelf and what goes to Columns shelf?**

* **Category** being a **dimension** by which we wish to group the data, goes to **Columns** shelf (X-axis).
* **Product ID** is a **dimension** but it still goes to **Rows** shelf (Y-axis) since we plan to use it as a **measure** and count (CNTD) the unique Product IDs.

**Steps:**

* 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**

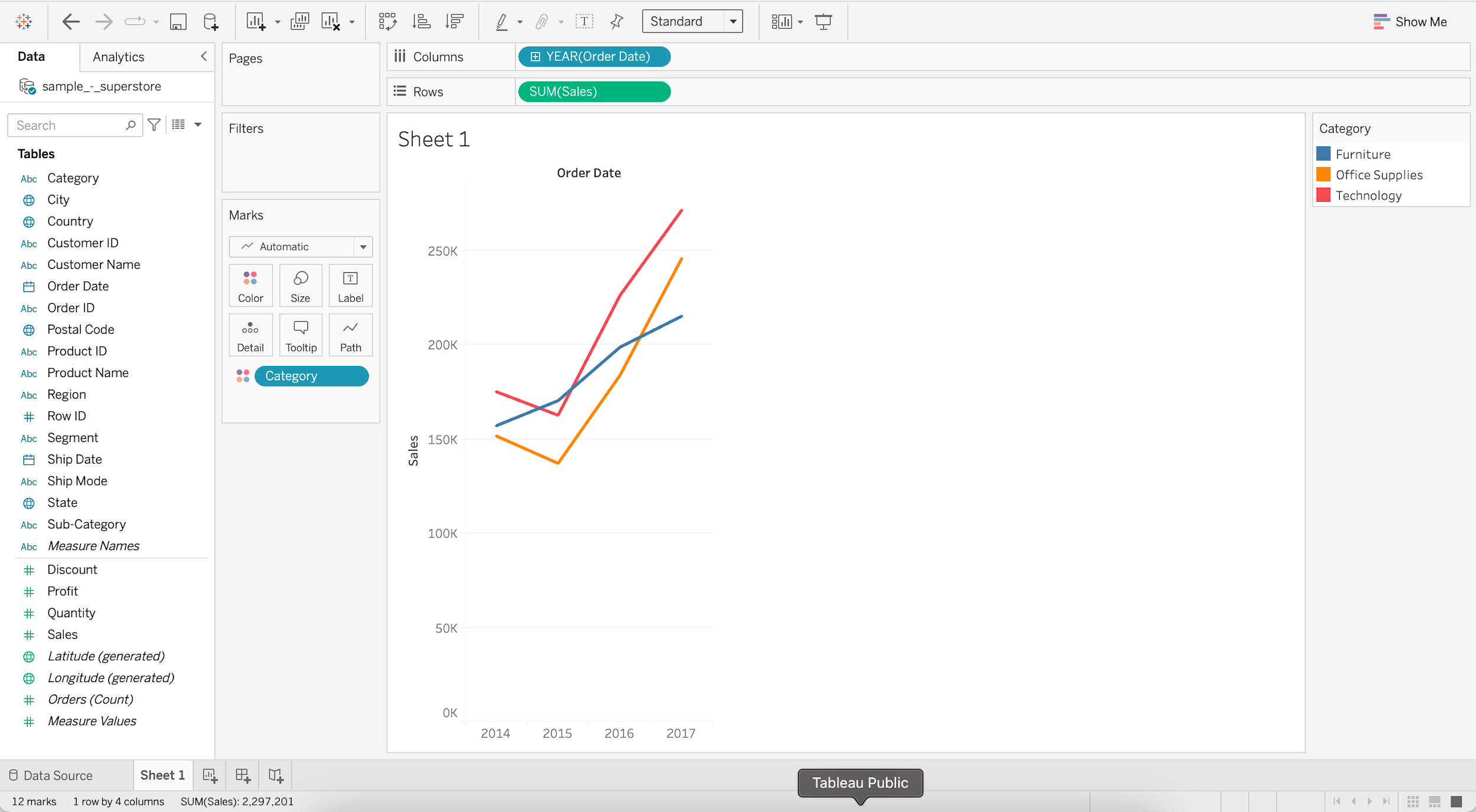
**Business problem 3 :** Find the product category that has the highest sales throughout most of the years.

**How to choose what goes to Rows shelf and what goes to Columns shelf?**

* Year of **Order Date** being a **dimension** by which we wish to group the data, goes to **Columns** shelf (X-axis).
* **Sales** being a **measure** that we want to aggregate (SUM), goes to **Rows** shelf (Y-axis).
* **Category** being a **dimension**, if put into the **Color** marks shelf, we’ll get a different color for each category.

**Steps:**

* Drag **Order Date** to Columns
* Drag **Sales** to Rows
* Add **Category** to Color



### **Continuous line chart**

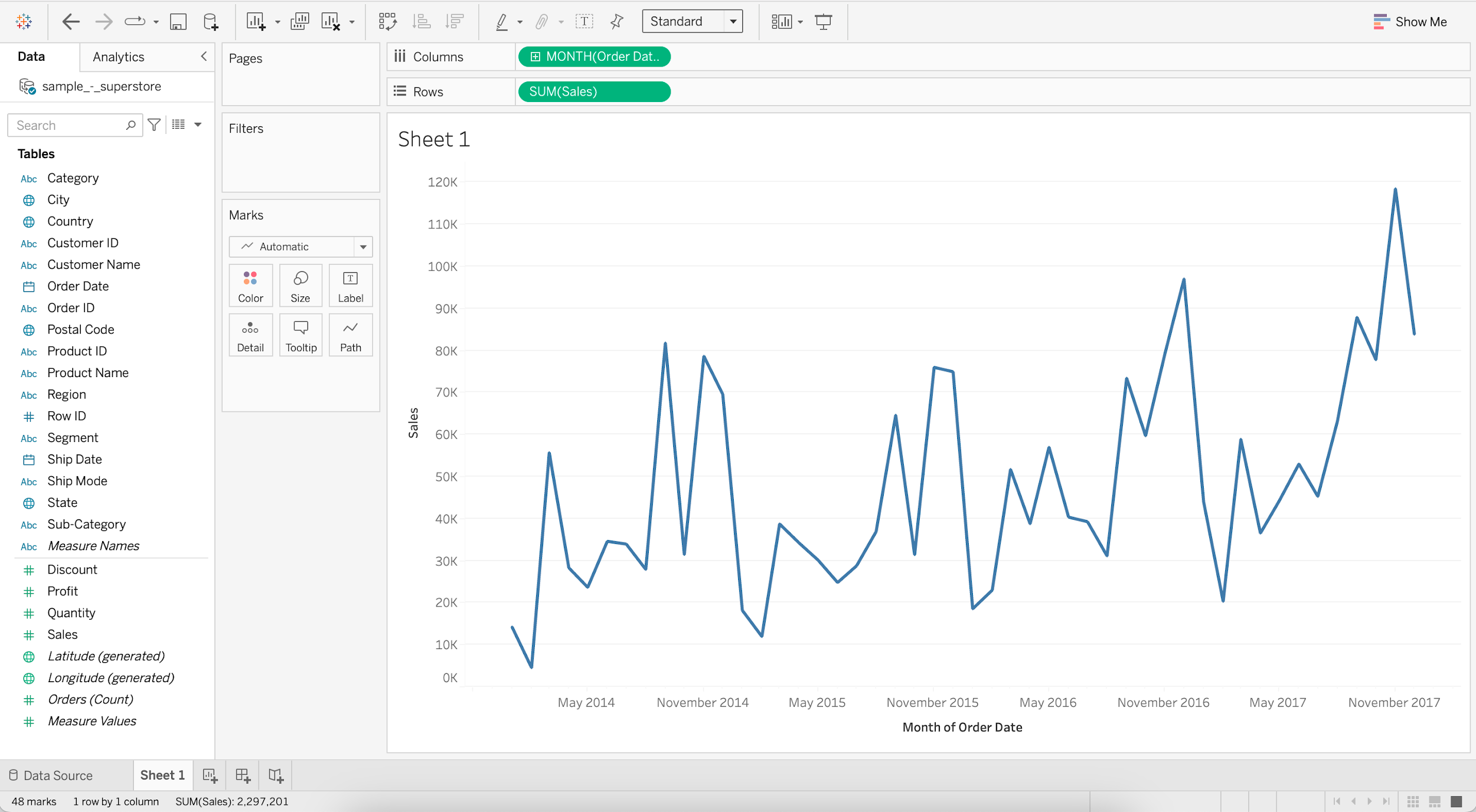
**Business problem 4 :** Find the year and month that had the highest and lowest sales.

**How to choose what goes to Rows shelf and what goes to Columns shelf?**

* Month of **Order Date** being a **dimension** by which we wish to group the data, goes to **Columns** shelf (X-axis).
* **Sales** being a **measure** that we want to aggregate (SUM), goes to **Rows** shelf (Y-axis).

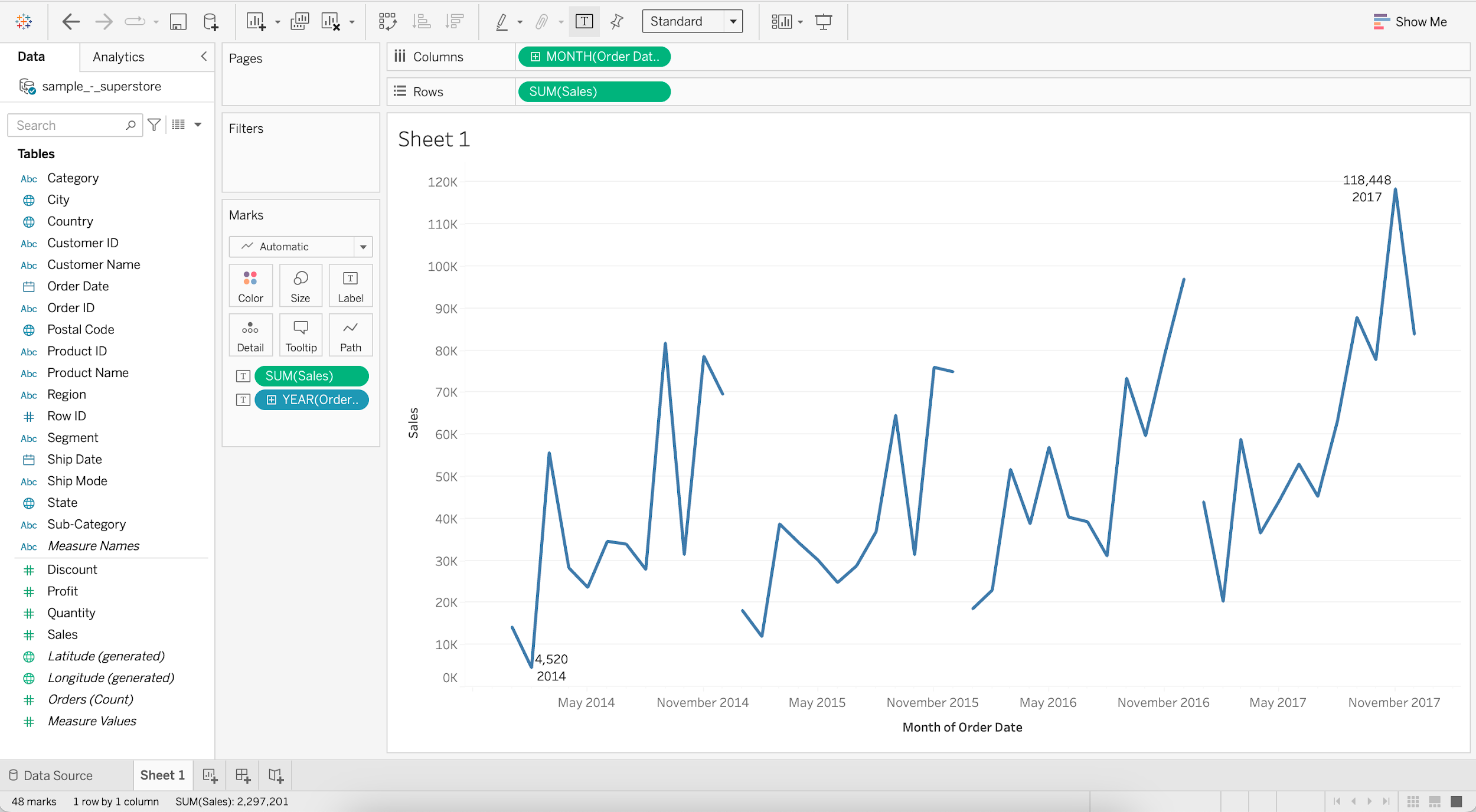
**Steps:**

* 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



* Putting **Sales** (Measure) on the **Label** marks shelf will add written information about **SUM of Sales** to the chart.
* Putting **Order Date** (Dimension) on the **Label** marks shelf will add written information about **YEAR of Order Date** to the chart.

## **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.

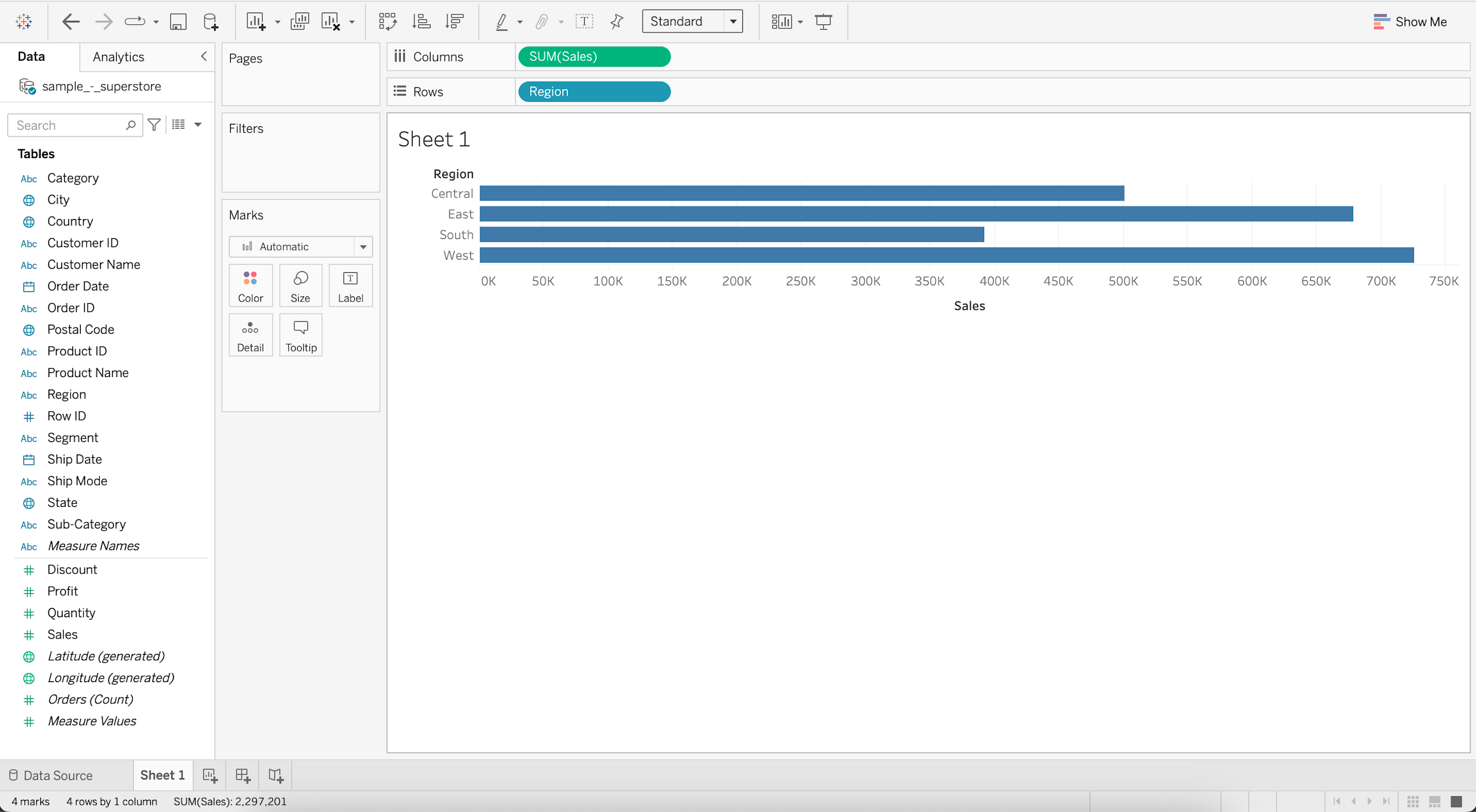
**Business problem 5 :** Find sales and profit by region for the year 2017.

**How to choose what goes to Rows shelf and what goes to Columns shelf?**

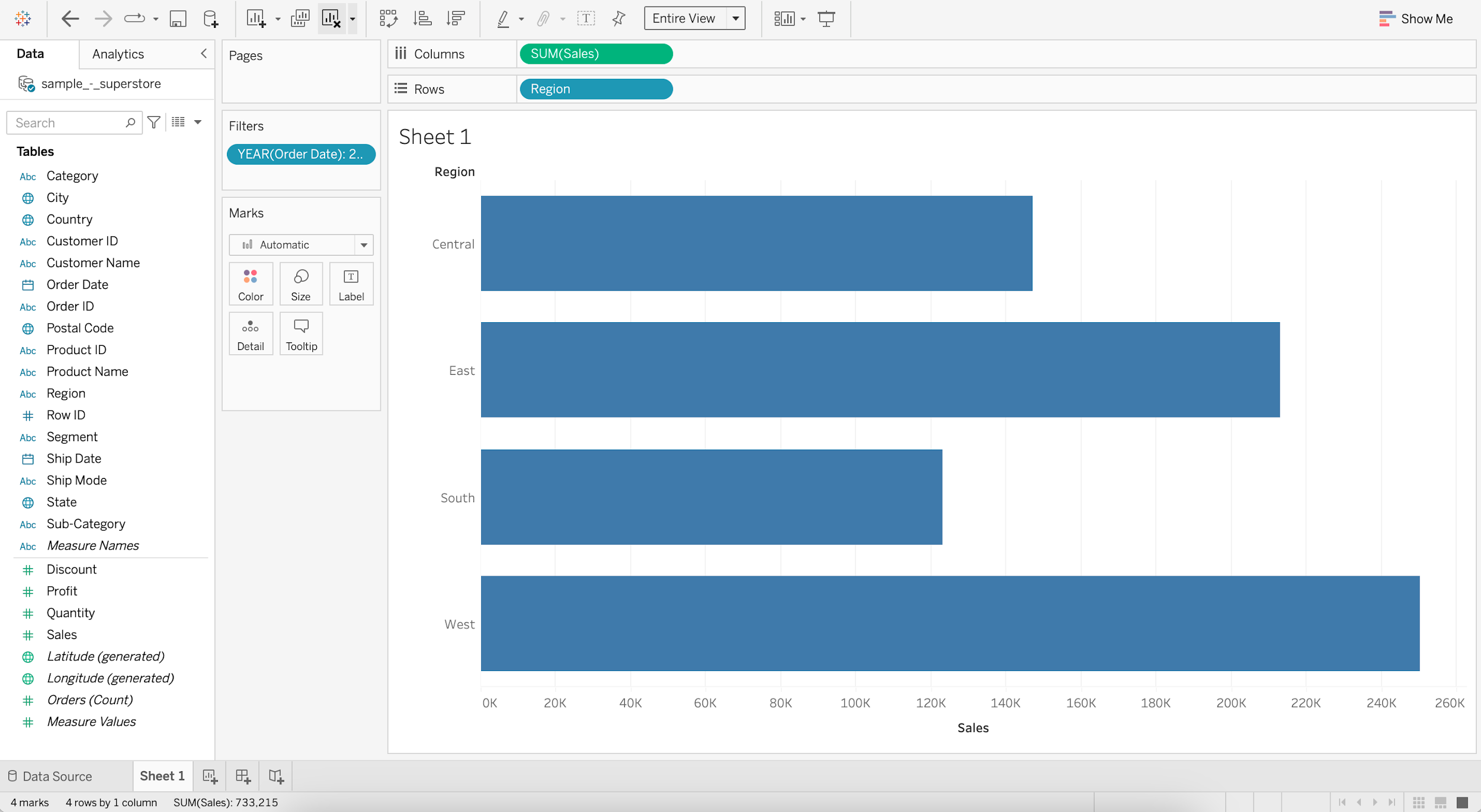
* Usually, **Sales** (Measure) would go to Rows shelf and **Region** (Dimension) would go to Columns shelf.
* But since we're creating a **Horizontal bar chart** (instead of a Vertical bar chart),
  + **Sales** (Measure) goes to **Columns** shelf.
  + **Region** (Dimension) goes to **Rows** shelf.

**Steps:**

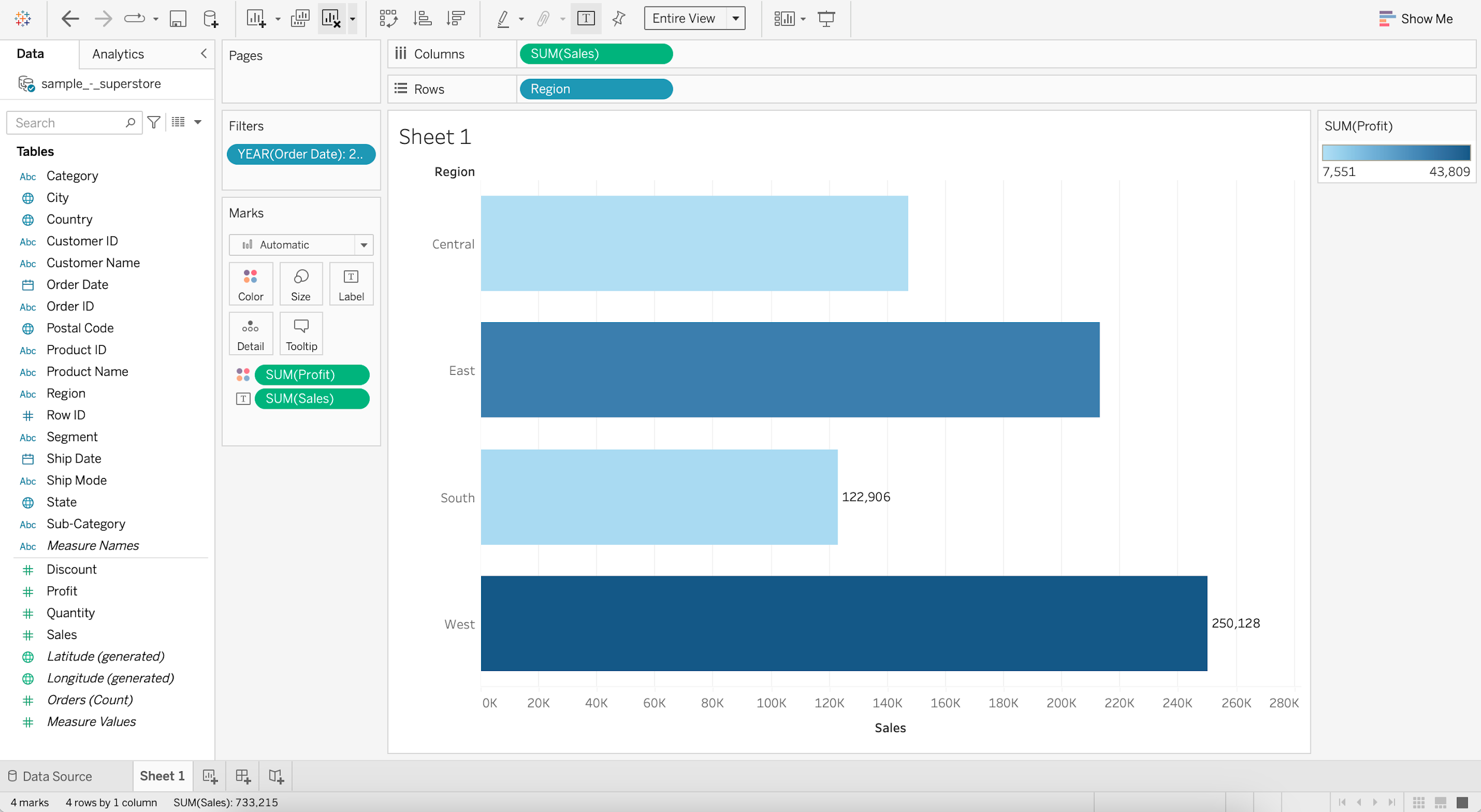
* Drag **Sales** to Columns
* Drag **Region** to Rows



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



* Add **Profit** to Color
* Add **Sales** to Label



* Adding **Profit** (Measure) to the **Color** marks shelf assigns different shades of a single color to the **SUM of Profit** values.
* Adding **Sales** (Measure) to the **Label** marks shelf will add written information to the chart about the **SUM of Sales** values.

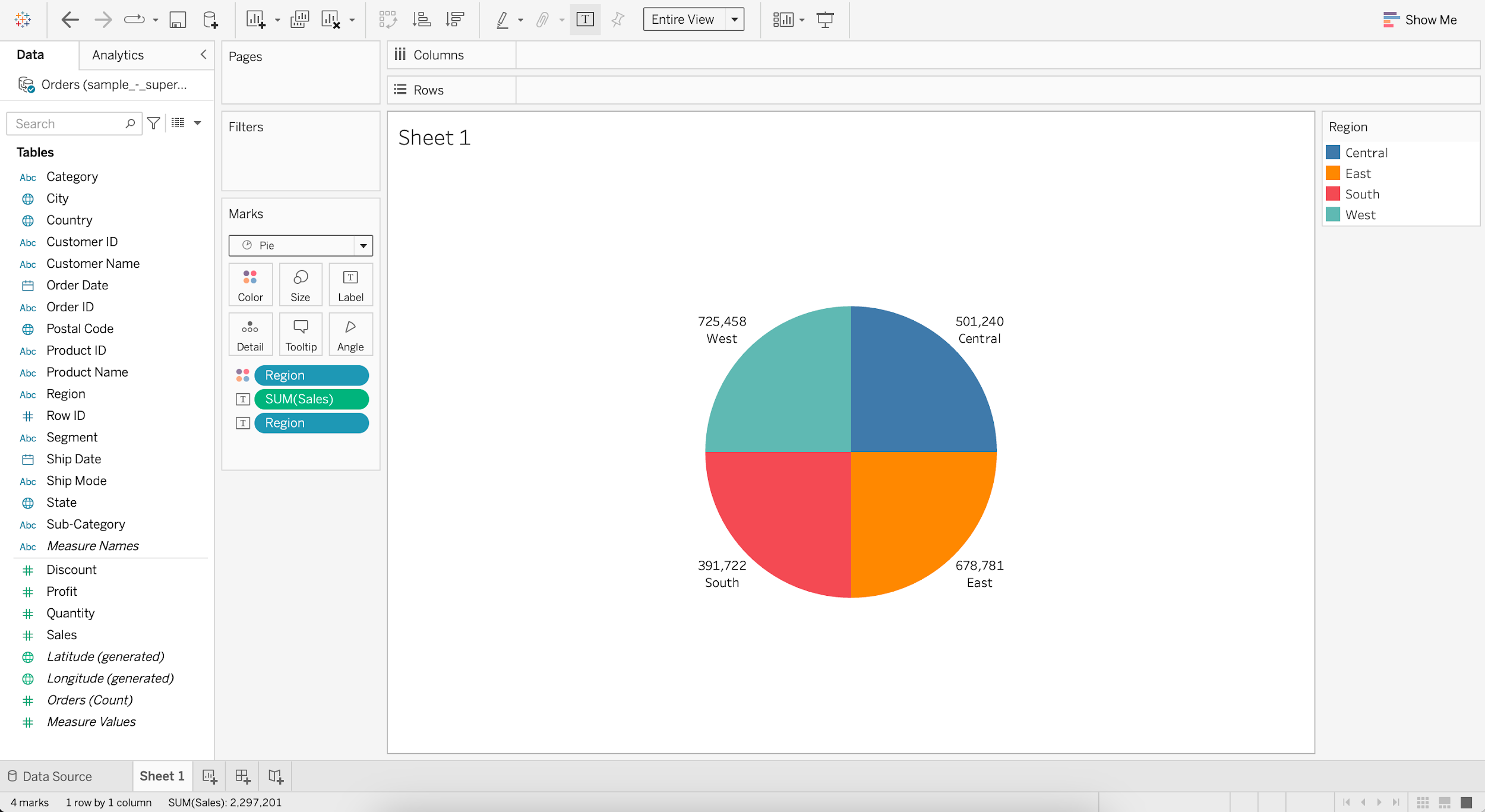
## **Pie chart**

It represents data as a slice of a circle with different sizes and colors.

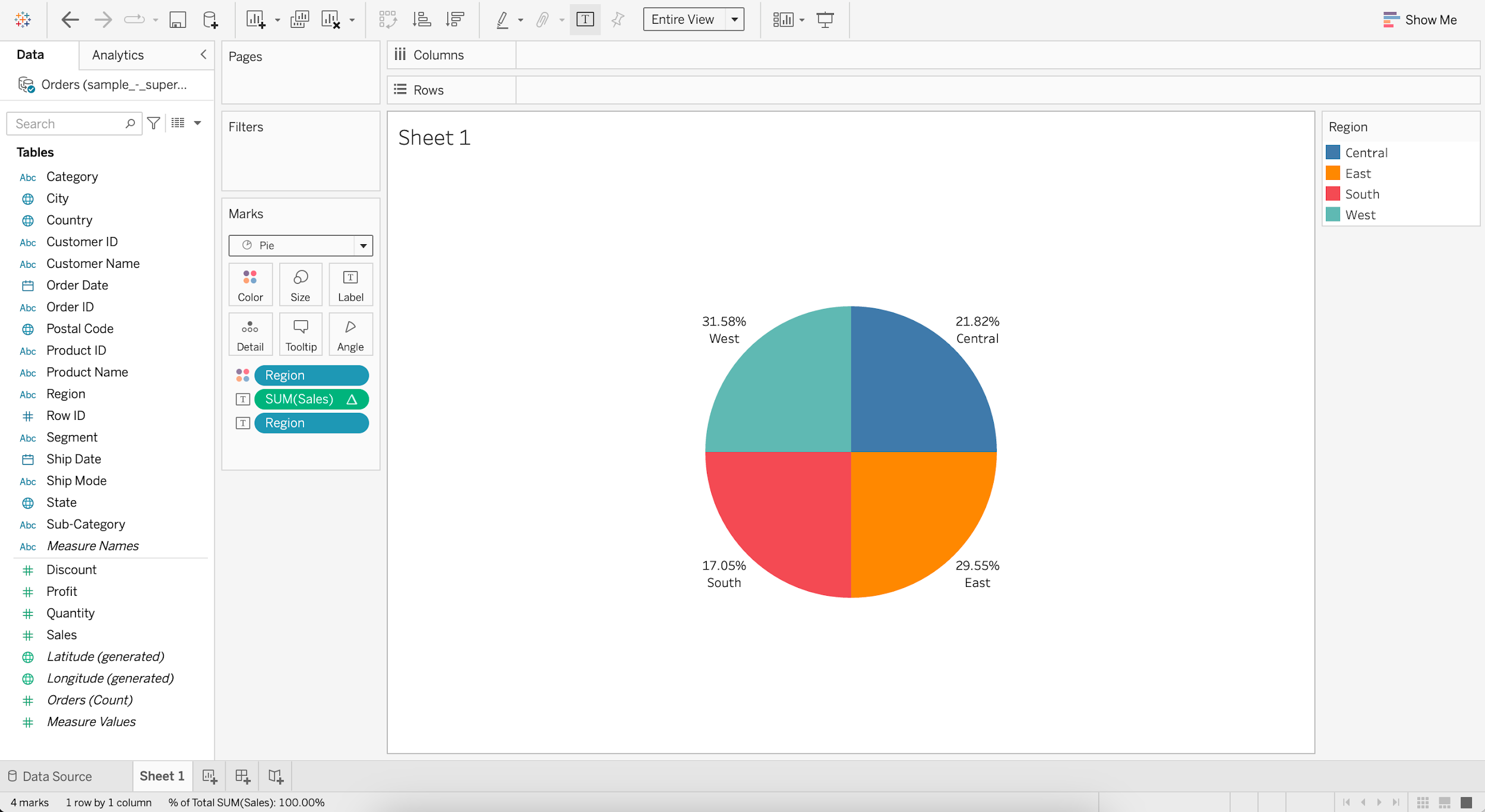
**Business problem 6:** Show relative percentage of sales and profit by region for the year 2017.

**Steps:**

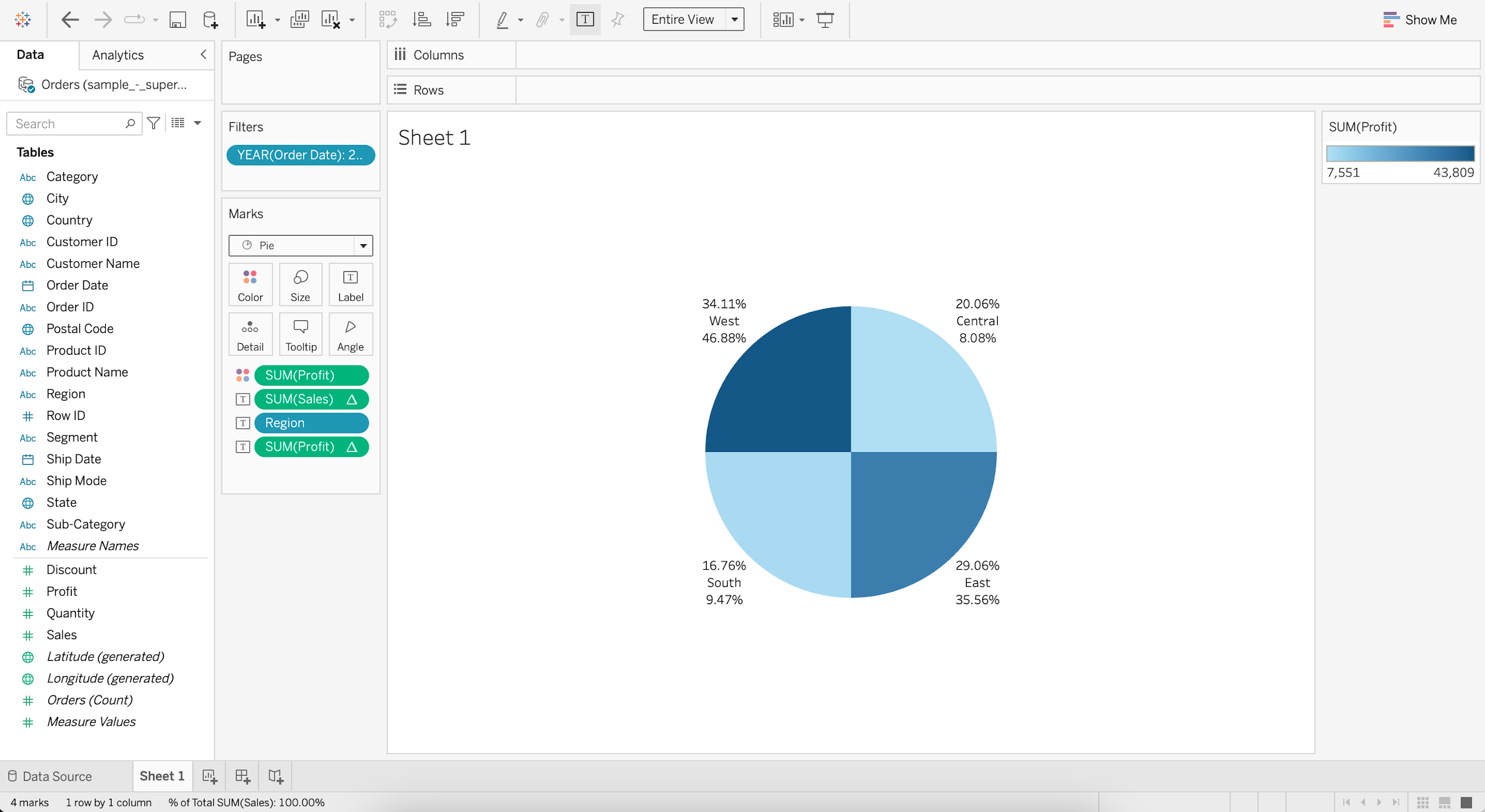
* Drag **Sales** to the Sheet
* Change Marks from Automatic to Pie
* Add **Region** to Color
* Add **Region** to Label
* Switch to Entire View



* If we drag **Region** (Dimension) to the **Color** marks shelf, we’ll get a different color for each region.
* Putting **Region** (Dimension) on the **Label** marks shelf will add written information about region to the chart.
* Adding **Sales** (Measure) to the **Label** marks shelf will add written information to the chart about the **SUM of Sales** values.
* For SUM(Sales), go to Quick Table Calculation
* Select Percent of Total

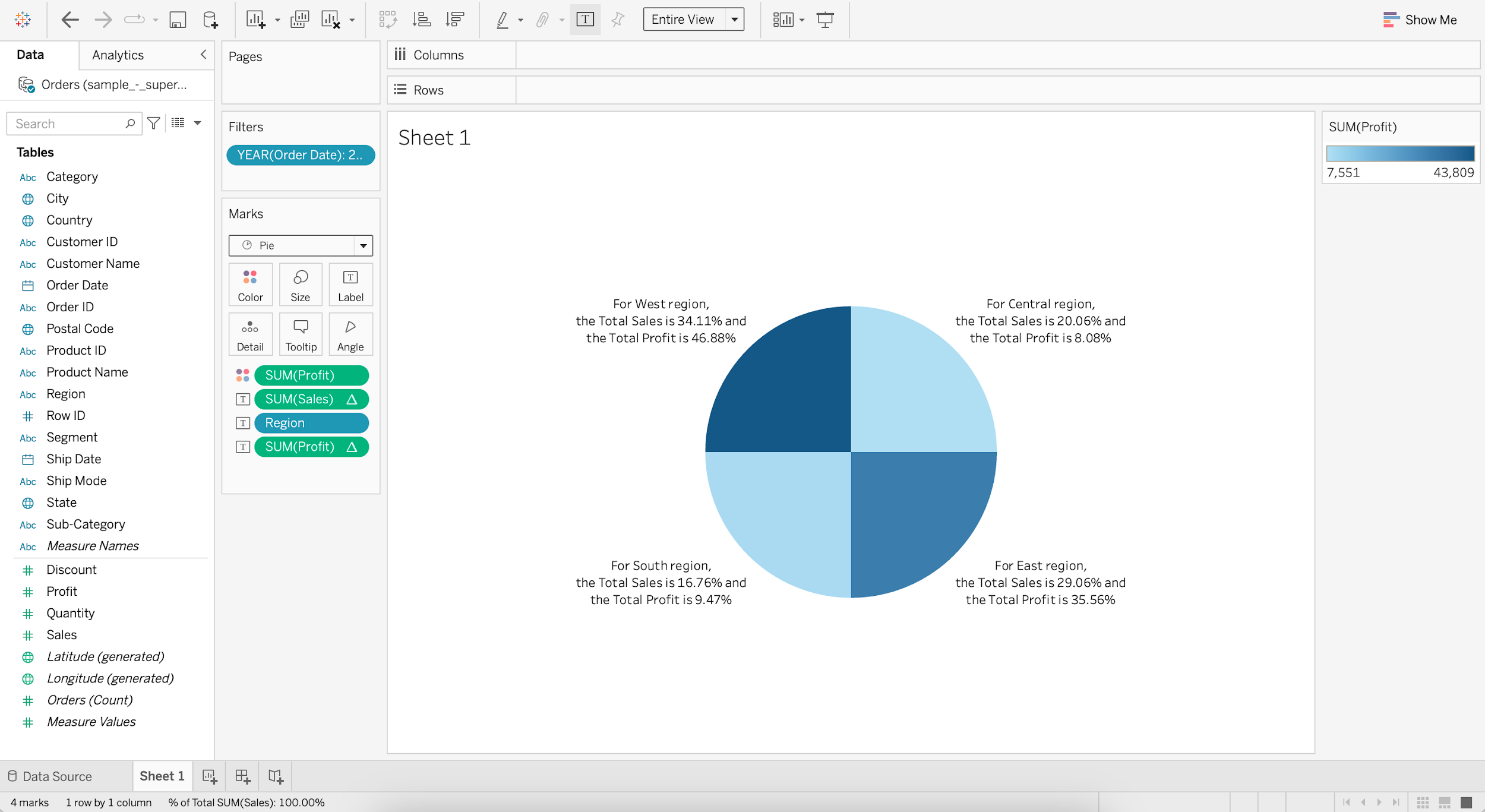


* Put **Order Date** into Filters
* Select Years and press next
* Select 2017 and press OK
* Add **Profit** to Color
* Add **Profit** to Label
* For SUM(Profit), go to Quick Table Calculation
* Select Percent of Total



#### **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.

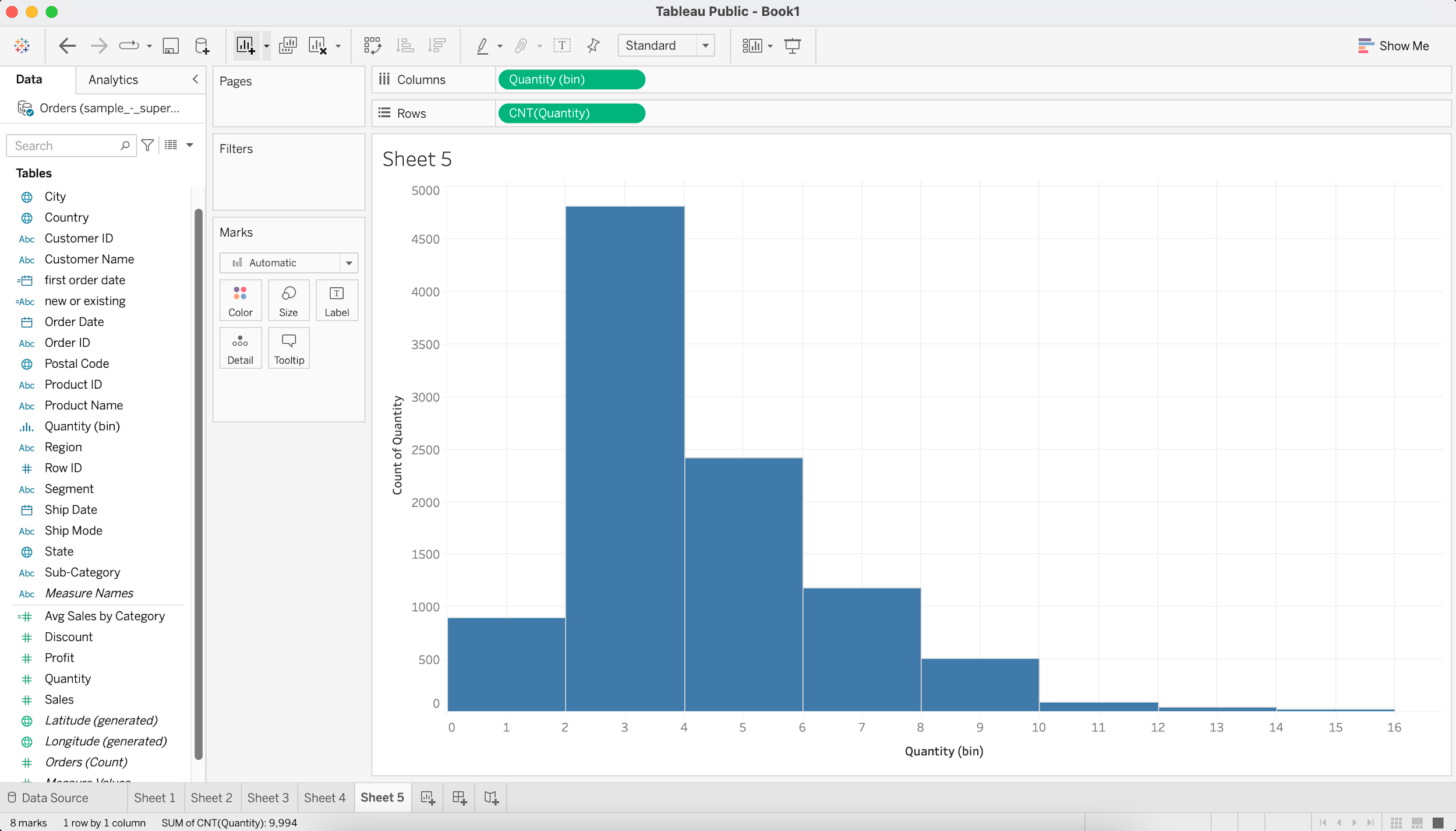
**Business problem 7:** Find the distribution of the number of quantities sold based on the quantity bins.

**How to choose what goes to Rows shelf and what goes to Columns shelf?**

* **Quantity (bin)** being a **dimension** by which we wish to group the data, goes to **Columns** shelf (X-axis).
* **Quantity** being a **measure** that we want to aggregate (CNT), goes to **Rows** shelf (Y-axis).

**Steps:**

* For **Quantity**, go to Create and choose Bins
* Change the Size of bins to 2
* Drag **Quantity (bin)** to Columns
* Drag **Quantity** to Rows
* Change aggregation of **Quantity** to Count
* Set **Quantity (bin)** to Continuous



## **Cleaning and Formatting data for analysis**

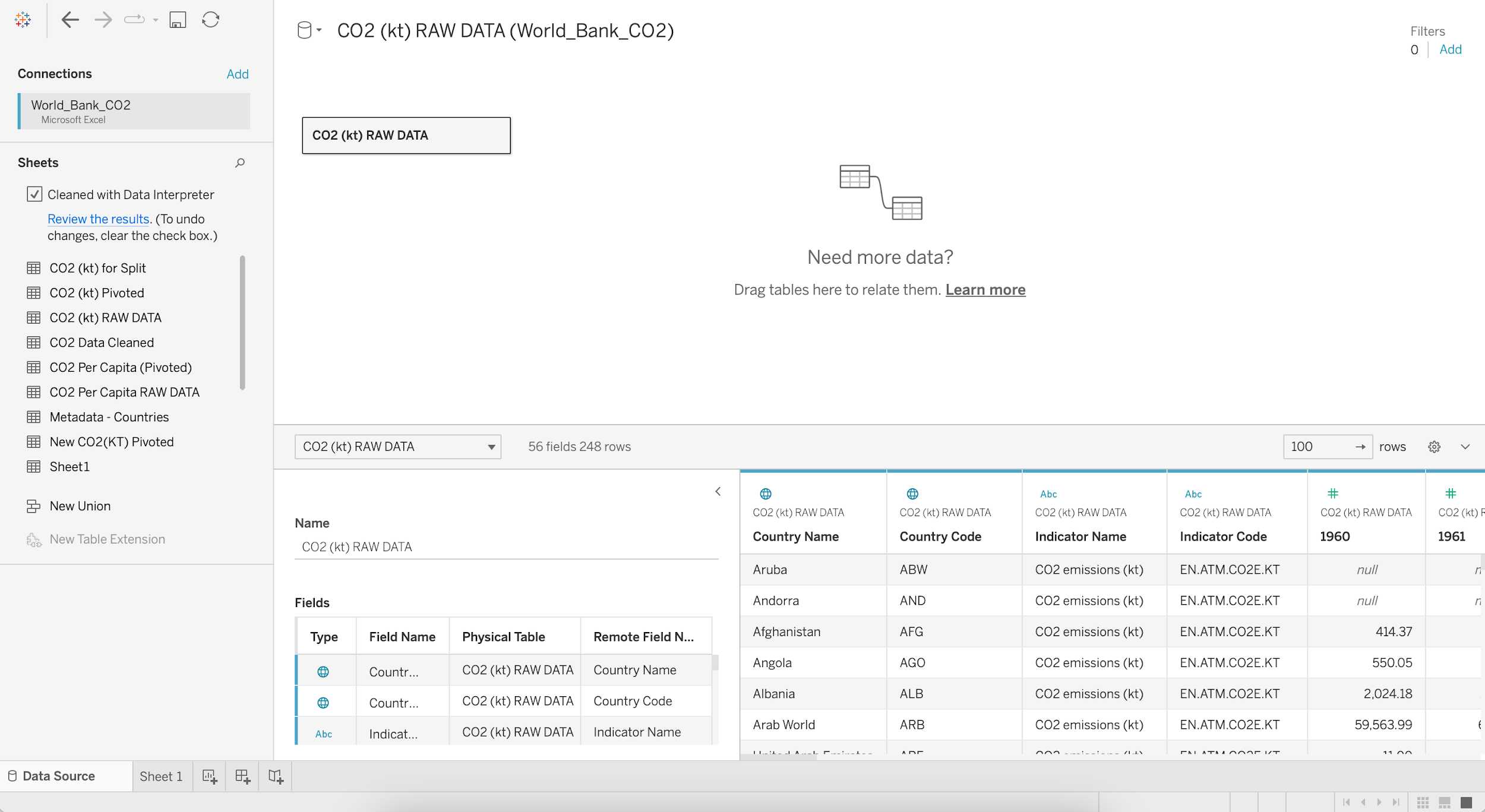
#### **Dataset :** [**World\_Bank\_CO2.xlsx**](https://docs.google.com/spreadsheets/d/1PUpy6uhtn15n18f7oS9Y1Iy-VoWtEzba/edit?usp=share_link&ouid=100659516601446935794&rtpof=true&sd=true) **[Use CO2 (kt) RAW DATA table]**

### **Business problem 8:**

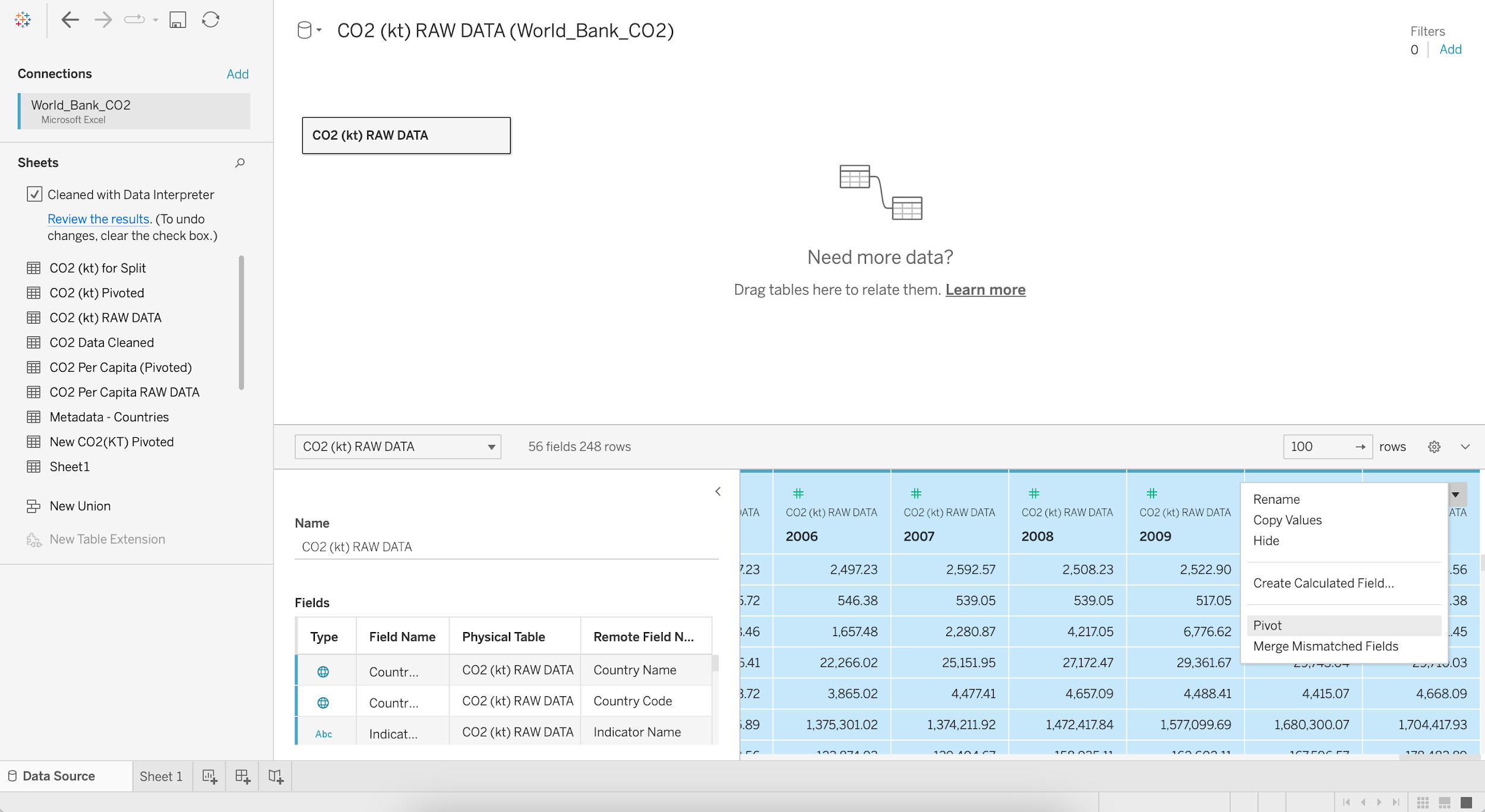
Find the CO2 emissions per country per year for those countries which have CO2 emissions above 2000 kt.

**Steps:**

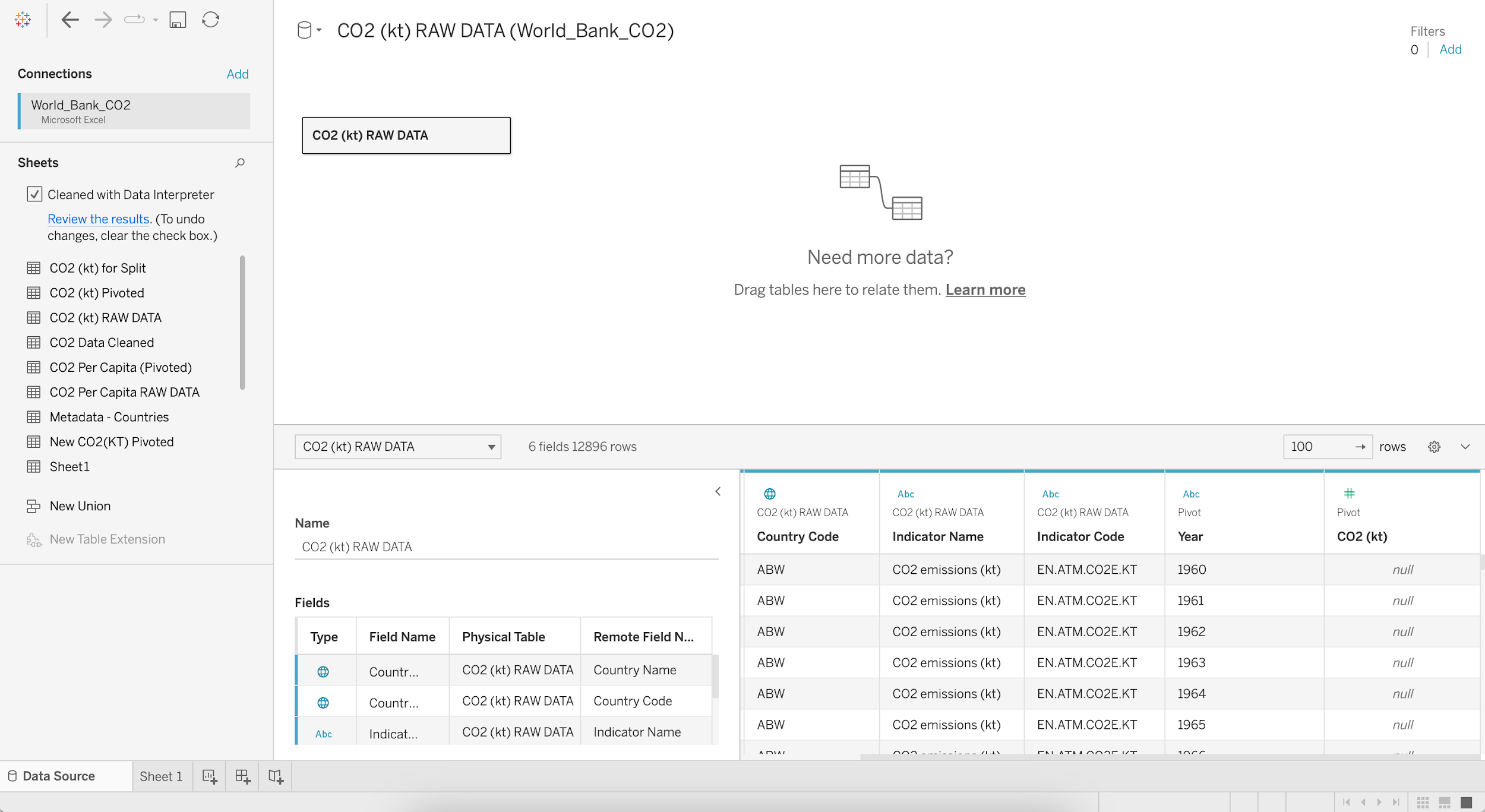
* Load the **World\_Bank\_CO2.xlsx** dataset
* Drag the **CO2 (kt) RAW DATA** table to the Canvas
* Check the **Use Data Interpreter** option



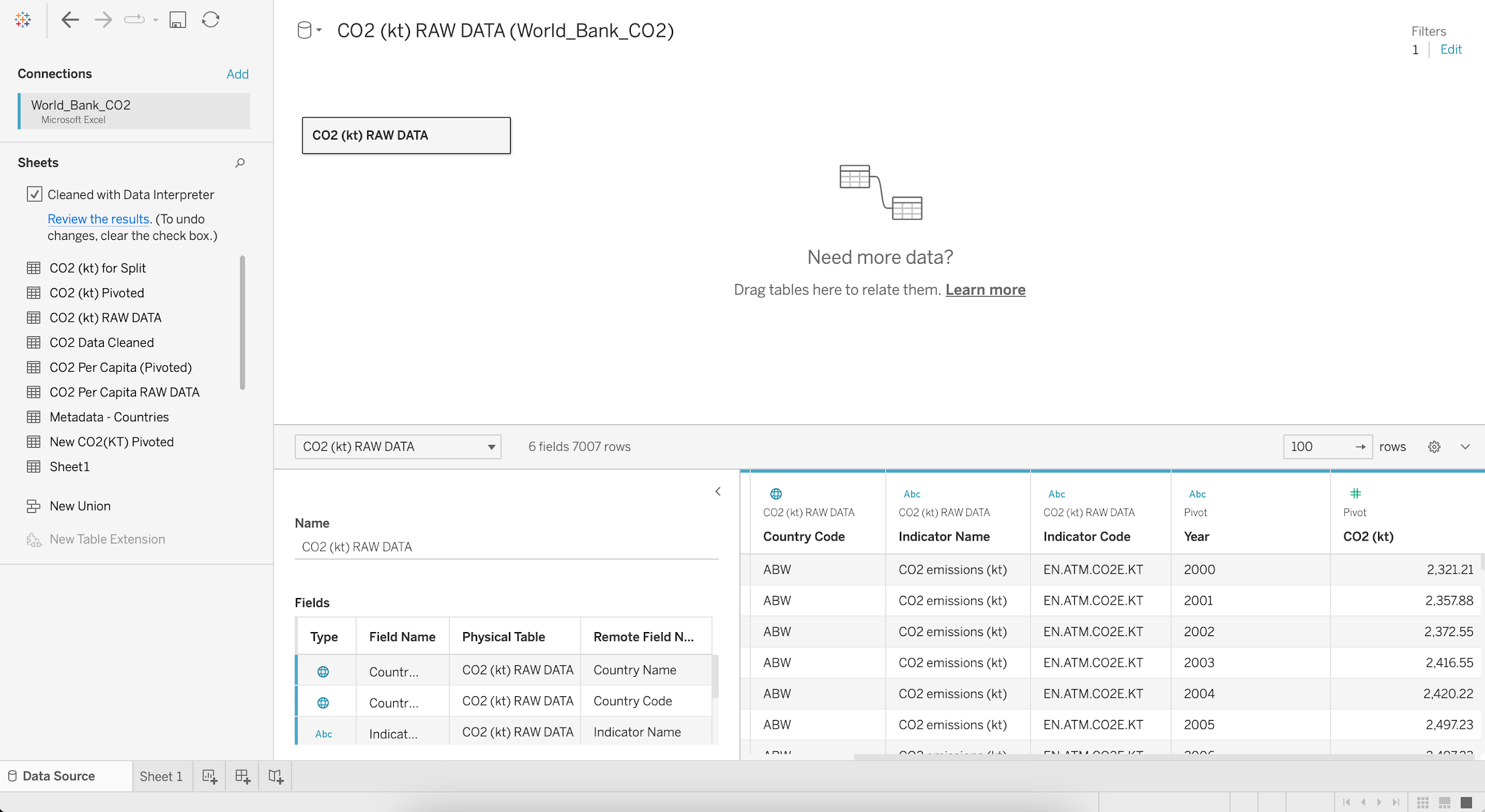
* Select the columns **1960** to **2011** and **Pivot**

****

* Rename the PIVOT FIELD NAMES column as **Year**
* Rename the PIVOT FIELD VALUES column as **CO2 (kt)**

****

* 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**

****

### **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.

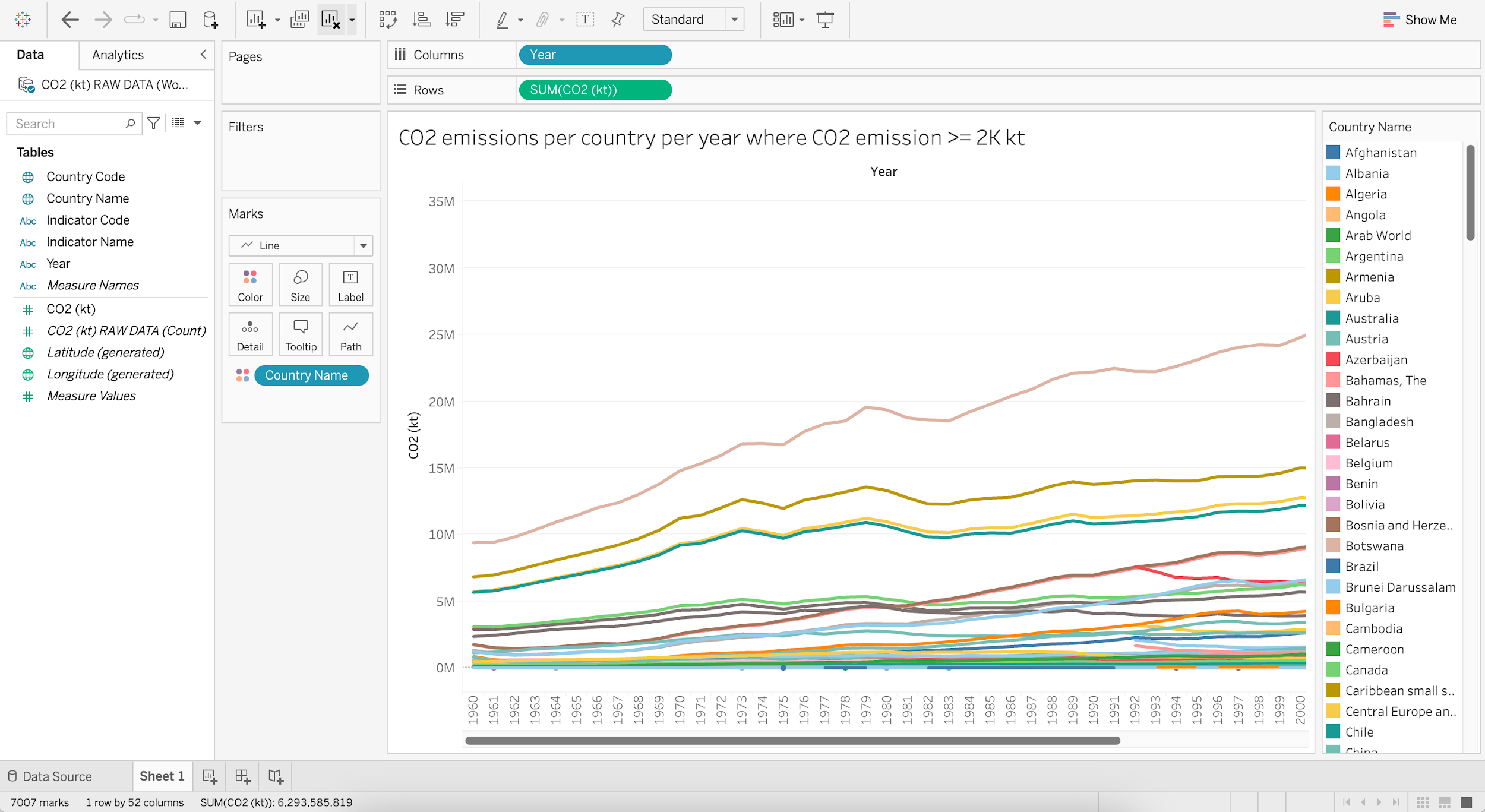
We can use the **Filter shelf** to filter data in the view.

### **Business problem 9:**

Find the CO2 emissions per country per year where CO2 emission >= 2K kt.

**Steps:**

* 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 >= 2K kt"



* **Year** being a **dimension** by which we wish to group the data, goes to **Columns** shelf (X-axis).
* **CO2 (kt)** being a **measure** that we want to aggregate (SUM), goes to **Rows** shelf (Y-axis).
* **Country Name** being a **dimension**, if put into the **Color** marks shelf, we’ll get a different color for each country.

### **Business problem 10:**

Find the per year CO2 emissions of top 10 countries where CO2 emissions >= 2K kt.

**Steps:**

* 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 >= 2K kt"

