**Tableau Assignment;**

* **1.What is the difference between discreate and continuous data?**

**Definition of Discrete Data;**

The term discrete implies distinct or separate. So, discrete data refers to the type of quantitative data that relies on counts. It contains only finite values, whose subdivision is not possible. It includes only those values that can only be counted in whole numbers or integers and are separate which means the data cannot be broken down into fraction or decimal.

**For example,** Number of students in the school, the number of cars in the parking lot, the number of computers in a computer lab, the number of animals in a zoo, etc.

**continuous data;** contains data that can be measured, that includes fractions and decimals. Take a read of the article to know the difference between discrete and continuous data can be broken down into fractions and decimal, i.e. it can be meaningfully subdivided into smaller parts according to the measurement precision.

[](https://keydifferences.com/wp-content/uploads/2016/05/discrete-vscontinuous-data.jpg)

Figure;1

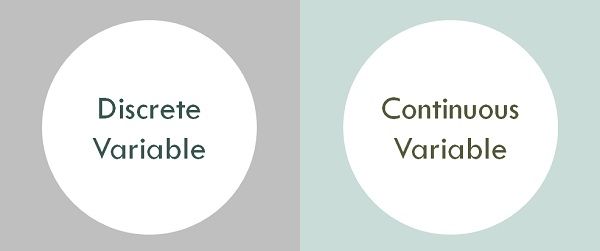
**Comparison Chart;**

| **BASIS FOR COMPARISON** |  | **DISCRETE DATA** | **CONTINUOUS DATA** |
| --- | --- | --- | --- |
| Meaning |  | Discrete data is one that has clear spaces between values. | Continuous data is one that falls on a continuous sequence. |
| Nature |  | Countable | Measurable |
| Values |  | It can take only distinct or separate values. | It can take any value in some interval. |
| Graphical Representation |  | Bar Graph | Histogram |
| Tabulation is known as |  | Ungrouped frequency distribution. | Grouped frequency distribution. |
| Classification |  | Mutually Inclusive | Mutually Exclusive |
| Function graph |  | Shows isolated points | Shows connected points |
| Example |  | Days of the week | Market price of a product |

**Key Differences Between Discrete and Continuous Data**

The difference between discrete and continuous data can be drawn clearly on the following grounds:

1. Discrete data is the type of data that has clear spaces between values. Continuous data is data that falls in a continuous sequence.
2. Discrete data is countable while continuous data is measurable.
3. Discrete data contains distinct or separate values. On the other hand, continuous data includes any value within range.
4. Discrete data is graphically represented by bar graph whereas a histogram is used to represent continuous data graphically.
5. Tabulation of discrete data, done against a single value, is called as an ungrouped frequency distribution. On the contrary, tabulation for continuous data, done against a group of value, called as grouped frequency distribution.
6. Overlapping or mutually exclusive classification, such as 10-20, 20-30, etc. s done for continuous data. As opposed to, non-overlapping or mutually inclusive classification like 10-19,20-29…., etc. is done for discrete data.
7. In a graph of the discrete function, it shows distinct point which remains unconnected. Unlike, continuous function graph, the points are connected with an unbroken line.

. 

**2.what is the criteria for data to land into dimensions and measures?**

**Ans;**

**Meaning**

* **dimensions;** A dimension is a field that can be considered an**independent variable**. Dimensions in Tableau produce headers when added to the Rows or Columns shelves in the view. By default, Tableau treats any field containing qualitative, categorical information as a dimension.

1. Slowly changing dimension: If the data in the dimension are changing over the period of time, then such a kind of dimension is called as SCD. Example: – In employ dimension data will change over a period of time.
2. Rapidly Changing dimension: If the data is changing rapidly then such kind of dimension is called a rapidly changing dimension.

Example: - Age change from time to time.

3. Unchanged dimension: If the data is constant and it won’t change then it is called as an unchanged dimension or static dimension.

Example: - Traffic signals, surname etc.

4. Conformed dimension: It is a dimension which is shared by the multiple business areas. Example: Java, DWH, Oracle etc.

5. Shrunken dimension: Subset of one dimension or subpart of one dimension is called as a shrunken dimension. Example: Quarter is the shrunken dimension of the year.

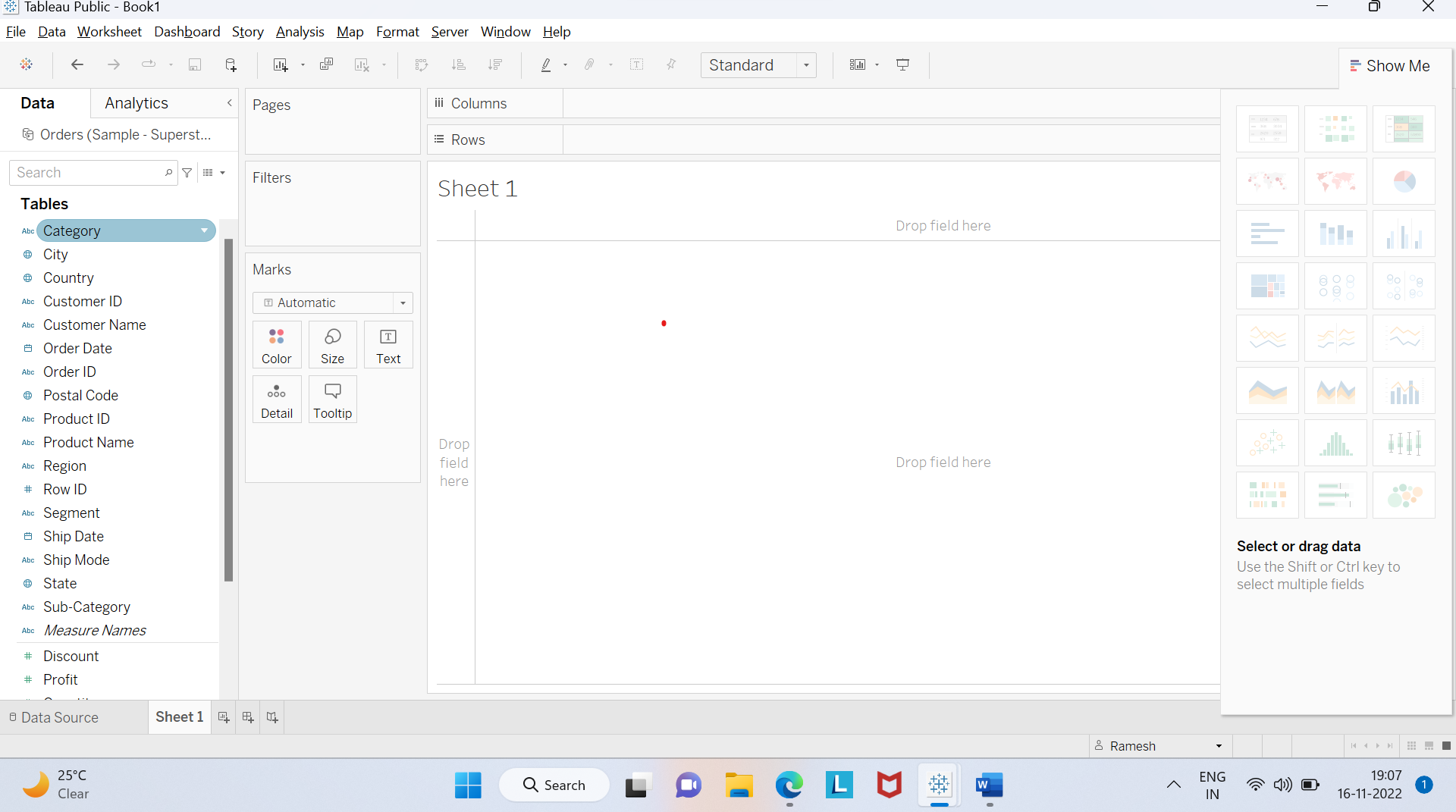
6. Role-playing dimension: Once dimension is playing multiple roles in the fact table is called as a role-playing dimension. Example: Take data dimension, the date dimension contains data of order, date of delivery, date of invoice etc.

7. Degenerated dimension: It is a dimension, where all the values get stored in a fact table, not in a separate dimension table. Degenerated dimension always contains the dimension keys, it won’t contain any other values

8. Junk dimension: Junk dimension is a dimension where we store Junk data. It is a single table with a combination of different related and unrelated data values like flags, indicators, some other unwanted data. A Junk dimension is mainly used to avoid the large no of foreign keys in the fact table.

9. Informed dimensions: Here all the dimensions are attached to fact. If we want to load employ table to the fact, it would be done by surrogate key. Measures in Tableau Depending on the measure values the Facts or measures in Tableau are three types semi additive facts non-additive. 10.Additive fact: It is a type of fact where it supports all the group functions (Sum, min, max, Avg etc.

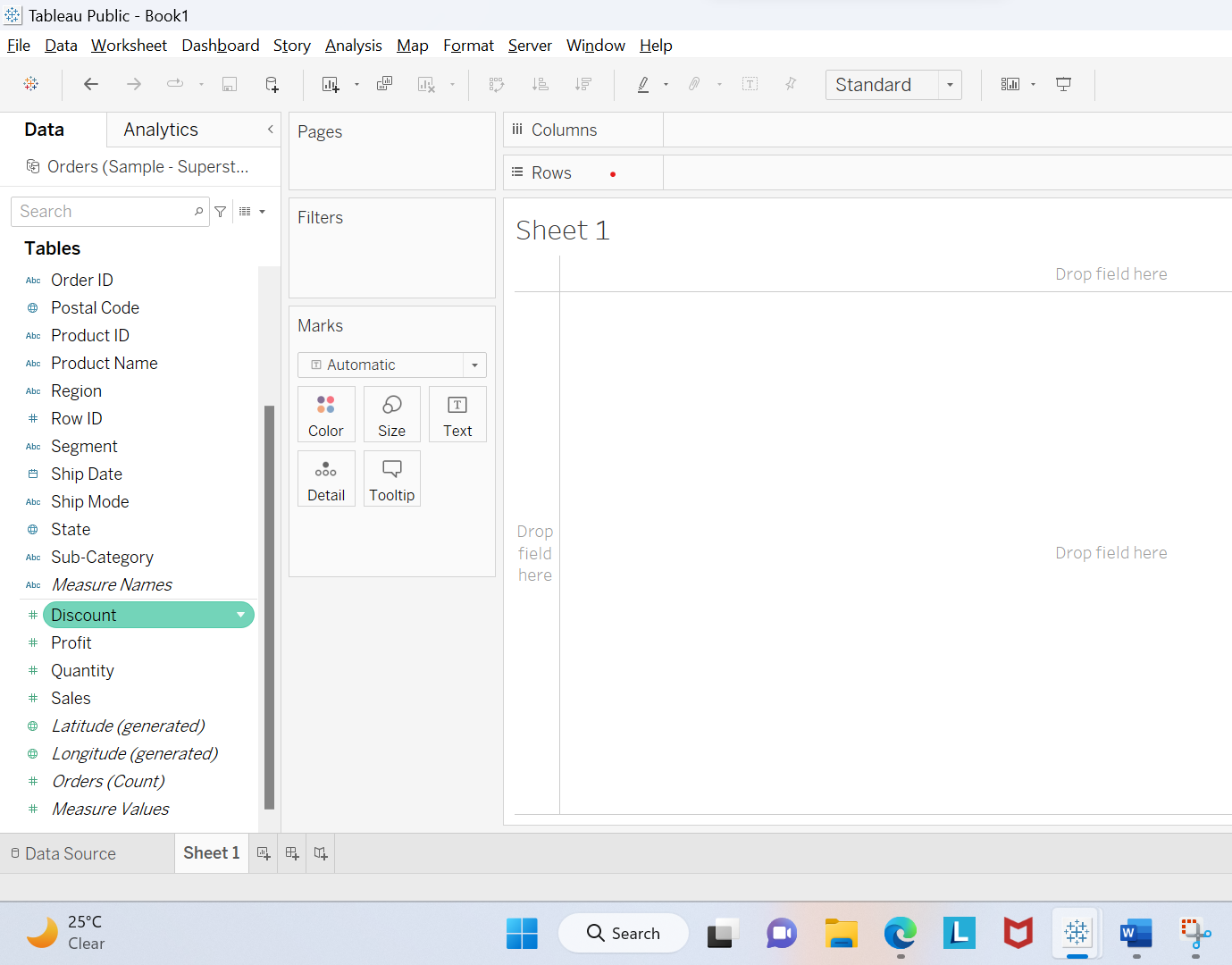
Example visualization of Dimensions;



* **Measures;** Depending on the measure values the Facts or measures in Tableau are three types, Additive fact, semiadditive fact, non-additive fact.
* Additive fact: It is a type of fact where it supports all the group functions (Sum, min, max, Avg etc.)
* Semi-additive fact: It is a fact which supports only a few of the group functions. Example: For account balance, we can apply only Max and Min which gives the meaningful output.
* Non-additive fact: It is a type of measure where we can’t apply any type of aggregations.

Example: we will take rations, percentages we can’t apply any group functions.

Example visualization of measures;



**3.what is meta data where is it present in the work book?**

**Ans;**

* What is Metadata?

Metadata is simply defined as data about data. The data that is used to represent other data is known as metadata. For example, the index of a book serves as a metadata for the contents in the book. In other words, we can say that metadata is the summarized data that leads us to detailed data. In terms of data warehouse, we can define metadata as follows.

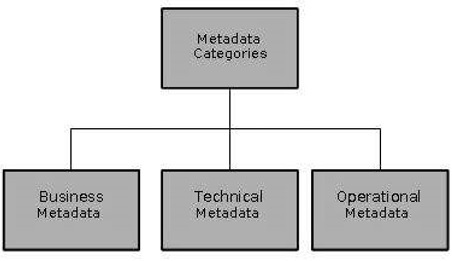
* Metadata is the road-map to a data warehouse.
* Metadata in a data warehouse defines the warehouse objects.
* Metadata acts as a directory. This directory helps the decision support system to locate the contents of a data warehouse.

**Note** − In a data warehouse, we create metadata for the data names and definitions of a given data warehouse. Along with this metadata, additional metadata is also created for time-stamping any extracted data, the source of extracted data.

* Categories of Metadata

Metadata can be broadly categorized into three categories –

* **Business Metadata** − It has the data ownership information, business definition, and changing policies.
* **Technical Metadata** − It includes database system names, table and column names and sizes, data types and allowed values. Technical metadata also includes structural information such as primary and foreign key attributes and indices.
* **Operational Metadata** − It includes currency of data and data lineage. Currency of data means whether the data is active, archived, or purged. Lineage of data means the history of data migrated and transformation applied on it.

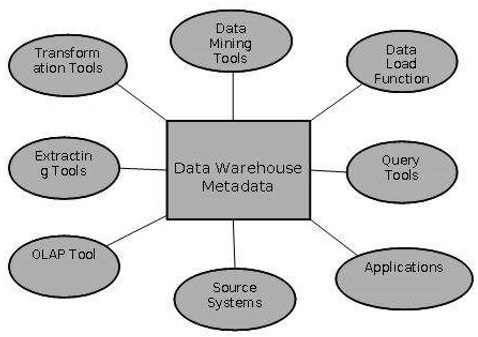


* Role of Metadata

Metadata has a very important role in a data warehouse. The role of metadata in a warehouse is different from the warehouse data, yet it plays an important role. The various roles of metadata are explained below.

* Metadata acts as a directory.
* This directory helps the decision support system to locate the contents of the data warehouse.
* Metadata helps in decision support system for mapping of data when data is transformed from operational environment to data warehouse environment.
* Metadata helps in summarization between current detailed data and highly summarized data.
* Metadata also helps in summarization between lightly detailed data and highly summarized data.
* Metadata is used for query tools.
* Metadata is used in extraction and cleansing tools.
* Metadata is used in reporting tools.
* Metadata is used in transformation tools.
* Metadata plays an important role in loading functions.

The following diagram shows the roles of metadata.



## Metadata Repository

Metadata repository is an integral part of a data warehouse system. It has the following metadata −

* **Definition of data warehouse** − It includes the description of structure of data warehouse. The description is defined by schema, view, hierarchies, derived data definitions, and data mart locations and contents.
* **Business metadata** − It contains has the data ownership information, business definition, and changing policies.
* **Operational Metadata** − It includes currency of data and data lineage. Currency of data means whether the data is active, archived, or purged. Lineage of data means the history of data migrated and transformation applied on it.
* **Data for mapping from operational environment to data warehouse** − It includes the source databases and their contents, data extraction, data partition cleaning, transformation rules, data refresh and purging rules.
* **Algorithms for summarization** − It includes dimension algorithms, data on granularity, aggregation, summarizing, etc.

## Challenges for Metadata Management

The importance of metadata cannot be overstated. Metadata helps in driving the accuracy of reports, validates data transformation, and ensures the accuracy of calculations. Metadata also enforces the definition of business terms to business end-users. With all these uses of metadata, it also has its challenges. Some of the challenges are discussed below.

* Metadata in a big organization is scattered across the organization. This metadata is spread in spreadsheets, databases, and applications.
* Metadata could be present in text files or multimedia files. To use this data for information management solutions, it has to be correctly defined.
* There are no industry-wide accepted standards. Data management solution vendors have narrow focus.
* There are no easy and accepted methods of passing metadata.

**4.what happens when you aggregate or disaggregate the data?**

**Ans;**

**Aggregation**.?

The process of viewing numeric values or measures at higher and more summarized levels of the data is called **aggregation**.

When you place a measure on a shelf, Tableau automatically aggregates the data, usually by summing it.

You can easily determine the aggregation applied to a field because the function always appears in front of the field ‘s name when it is placed on a shelf.

For example, Sales becomes SUM(Sales).

You can aggregate measures using Tableau only for relational data sources. Multidimensional data sources contain aggregated data only. In Tableau, multidimensional data sources are supported only in Windows.

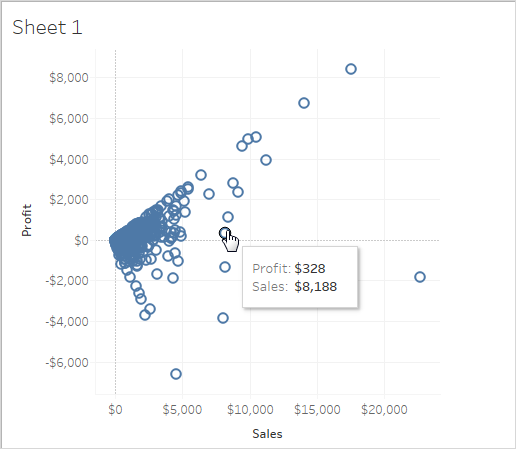
**Disaggregation;**

According to Tableau, **disaggregating** your data allows you to view every row of the data source which can be useful when you are analysing measures that you may want to use both independently and dependently in the view.

For example, you may be analysing the results from a product satisfaction survey with the Age of participants along one axis.

You can aggregate the Age field to determine the average age of participants or disaggregate the data to determine at what age participants were most satisfied with the pro

**Aggregate data combines and summarizes information**, whereas disaggregate data separate aggregated data into separate points or pieces of information. Disaggregating data might help gain a deeper understanding of various subsets within a larger dataset



**5.you are working on a dataset the client adds in more data to the dataset. what happens to the visualization that you had created? Give the explanation for both live and extracted data?**

**Live & extract connection**

## Data extracts vs. live connections

“Extract” is a word you’re going to hear a lot in Tableau. Extracts are one of the most powerful but overlooked tools in Tableau’s arsenal. Tableau Data Extracts are snapshots of data optimized for aggregation and loaded into system memory to be quickly recalled for visualization. Extracts tend to be much faster than live connections, especially in more complex visualizations with large data sets, filters, calculations, etc. For a deep dive into how Tableau extracts are created, check out  on the subject. When you create an extract from a local file (such as a .csv or an Excel workbook) or an on-premise database, you’re speeding up the workbook through optimization. As a result, Tableau doesn’t need the database to build the visualization. Instead, Tableau’s in-memory data engine queries the extract directly. However, because an extract is a snapshot of the data, the extract will need to be refreshed to receive updates from the original data source, whether it is a local file or an on-premise database. Live connections offer the convenience of real-time updates, with any changes in the data source reflected in Tableau. But live connections also rely on the database for all queries. And unlike extracts, databases are not always optimized for fast performance. With live connections, your data queries are only as fast as the database itself. There are also more variables at play when using a live connection. Workbook speeds are affected by a variety of factors, including your network speed, traffic on that network, and any custom SQL.

## An extract or a live connection—which to use?

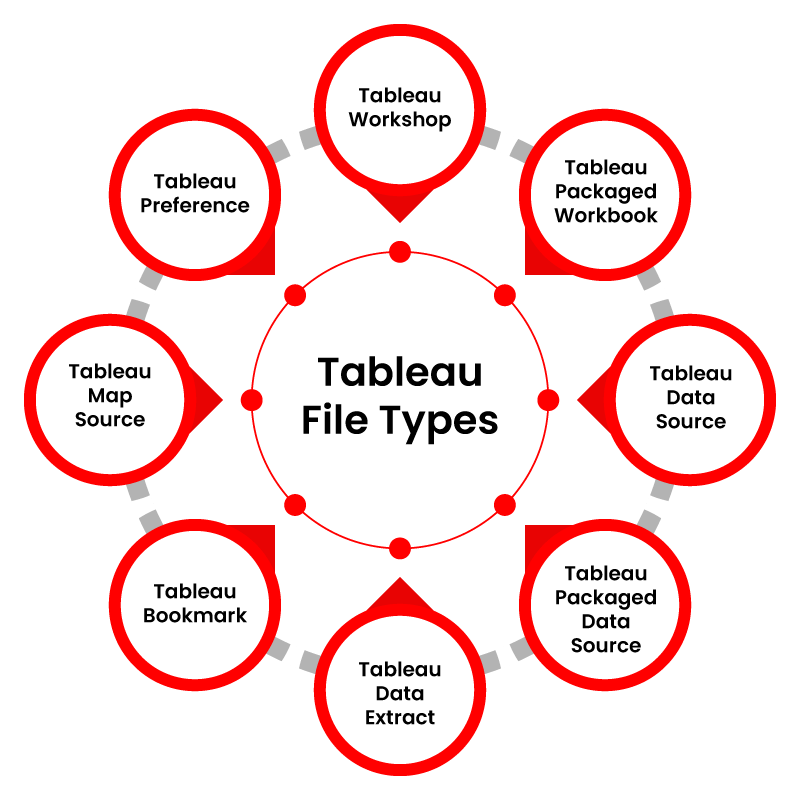
Both types of connections have their place. Hospitals that monitor incoming patient data need to make real-time decisions. These situations necessitate a live database connection. But in the same hospital, there may also be visualizations that monitor daily or weekly trends. For these analytics, using an extract of the data source helps build a faster workbook.

**6.What are the file extensions in Tableau and how each one is different?**

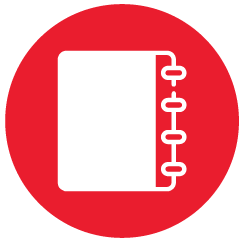
**Ans;**

* **What is file extensions;**

Tableau allows you to save your work in roughly seven to eight different file extensions and directories. This blog examines all of the tableau file extensions that you can use to store your work, as well as what they include, why you'd use them, and how they're created.

**Tableau file types;**

### 1. Tableau Workbooks:



The Tableau Workbook file format is the one you'll utilize the most. The extension of this file format is. twb, and it is the default for users. A workbook in Tableau is a file that contains sheets, dashboards, and other components. As a result, this Tableau file type includes data about the worksheets and dashboards in a workbook. These files contain all of the information on the fields, aggregate kinds, styles, formatting, filters, etc. To create a.twb file, go to the active data connection's data source control panel, then to the File option (in the toolbar), and select Save As. Then, from the Save As Type drop-down list, choose Tableau Workbook as the file type.

### 2. Tableau Bookmarks

Tableau Bookmark files are those with the. Tbm extension. These Tableau file types can store and share worksheets with others to utilize them in their workbooks without starting from scratch.

Go to the Windows option on the toolbar to create a.tbm file. Select Bookmark and then Create Bookmark from the drop-down menu. The bookmark will save the active worksheet as an a.tbm file.The use of Bookmark files has decreased since the release of a newer version of Tableau. We can immediately copy and paste worksheets from one workbook to another in Tableau versions 8.1 and beyond without creating an a.tbm file.

### 3. Tableau Packaged Workbooks

The Tableau Packaged Workbook file type contains metadata about a workbook's constituents and data derived from the data source. A.tde file includes the data extracted from the start. The extension. twbx utilized for Tableau Packaged Workbooks. When sharing a workbook with a user who does not have access to the live data connection, you can use a.twbx file type instead of a.twb (Tableau Workbook) file. As a result, you'll require a file containing the data taken from the source and other information about the workbook in this scenario.

Tableau Packaged Workbook files can also include information about associated images or geocoding that gets altered. To make a twbx file, go to File, then Save As, and then choose. twbx from the drop-down list.

### 4. Tableau Extract (.hyper or .tde)

The extension. tde is used for Tableau Data Extract files. Only a local copy of the complete or a subset of data from the source gets stored in these Tableau file types. It's worth noting that. tde files don't include a file The extension. tde is used for Tableau Data Extract files. Only a local copy of the complete or a subset of data from the source gets stored in these Tableau file types. It's worth noting that. tde files don't include a file location or information .

The data in such Tableau file formats cannot be automatically renewed when refreshes at the source, a known shortcoming. On the other hand, Tableau features a two-step approach for refreshing data stored as an extract in your.

### 5. Tableau Data Source (.tds)

Tableau Data Source files are files that contain all the information required for a Tableau data connection. When we create a new link to a data source, we make several changes to it to meet our needs, such as data types, aggregations, custom fields, and so on.

The Tableau Data Source files provide all necessary information for setting up a data connection and metadata for any additional user customizations. The. tds file can save data on data connections, including custom fields and table joins.

This Tableau file type, on the other hand, simply keeps the information required to connect to a data source, not the data itself.

Go to the Data tab on the toolbar to create an a.tds file. Then click Add To Saved Data Source from the drop-down menu and choose a data source to connect. Save the file as a Tableau Data Source file after that.

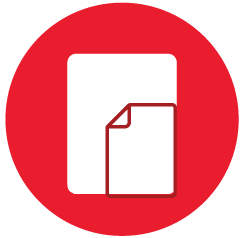
### 6. Tableau Packaged Data Source (.tdsx)



A Tableau Packaged Data Source file contains information about a data source connection and its data. The extracted data is saved as an a.tde file, while the source information is kept as an a.tds file. Data can be removed from any local file, including text files, extract files (. hyper or. tde), Excel files, Access files, and so on.A Tableau Packaged Data Source file, on the other hand, has the extension. tdsx. When we wish to share data and additional relevant information about a data source with a user who does not have access to the data source or its data, we utilize Tableau Packaged Data Source files.

Go to the Data tab on the toolbar to create an a.tdsx file. Then click Add to Saved Data Source from the drop-down menu and choose a data source to connect. Save the file as a Tableau Packaged Data Source file after that.

**7.Tableau preference file.**



Preference A Tableau Preference file stores all of the data associated with a custom color palette. You can design a custom colour palette or theme and save it as an a.tps file to utilize it consistently across the worksheet.

Tableau Preference files are in XML format and contain the extension. tps.My Tableau Repository contains these Tableau files.

**8.** **Tableau Map Source (.tms)**

For usage in Tableau, a Tableau Map Source file contains information on maps and their elements. .tms is the file extension for such files. By default, tableau will acquire map data such as backdrop and other layers from a specific map server or provider.

You can add map details from a WMS server of your choosing or a custom map from Map box to Tableau. Tableau will retrieve map details from that file instead of the default one and load map images and information accordingly whenever you generate a map file (.tms) of your choice. You can also share these. tms users with others in your organization.

**Conclusion:**

Tableau is a robust business intelligence and visualization tool. Because users do not need prior programming or coding abilities, this BI solution is straightforward to learn and use. Tableau provides a variety of visualization assets. Tableau File Types and Tableau Data Types are two of them that you've seen in this blog. The purpose of essential extension files and Tableau data types and how to generate them is discussed in this article.

**Practical questions**

**2.0. Text Table, Highlight Tables, Heat Maps, Tree Map:**

1. Create a text table for the Avg (Sales) for each subcategory using Sample Superstore? List which Sub -Category is got Avg (Sale) more than $1000? - **Sample Superstore**

2.Create a Heat Table for the order date and Region against the Sub Category based in Count of Sales with two colours diverging that is distinguished by Sum of Profit - **Sample Superstore**

3. Create a Highlight table for the States for the Order Date Year whose highlighting is done based on Sum of profits - **Sample Superstore**

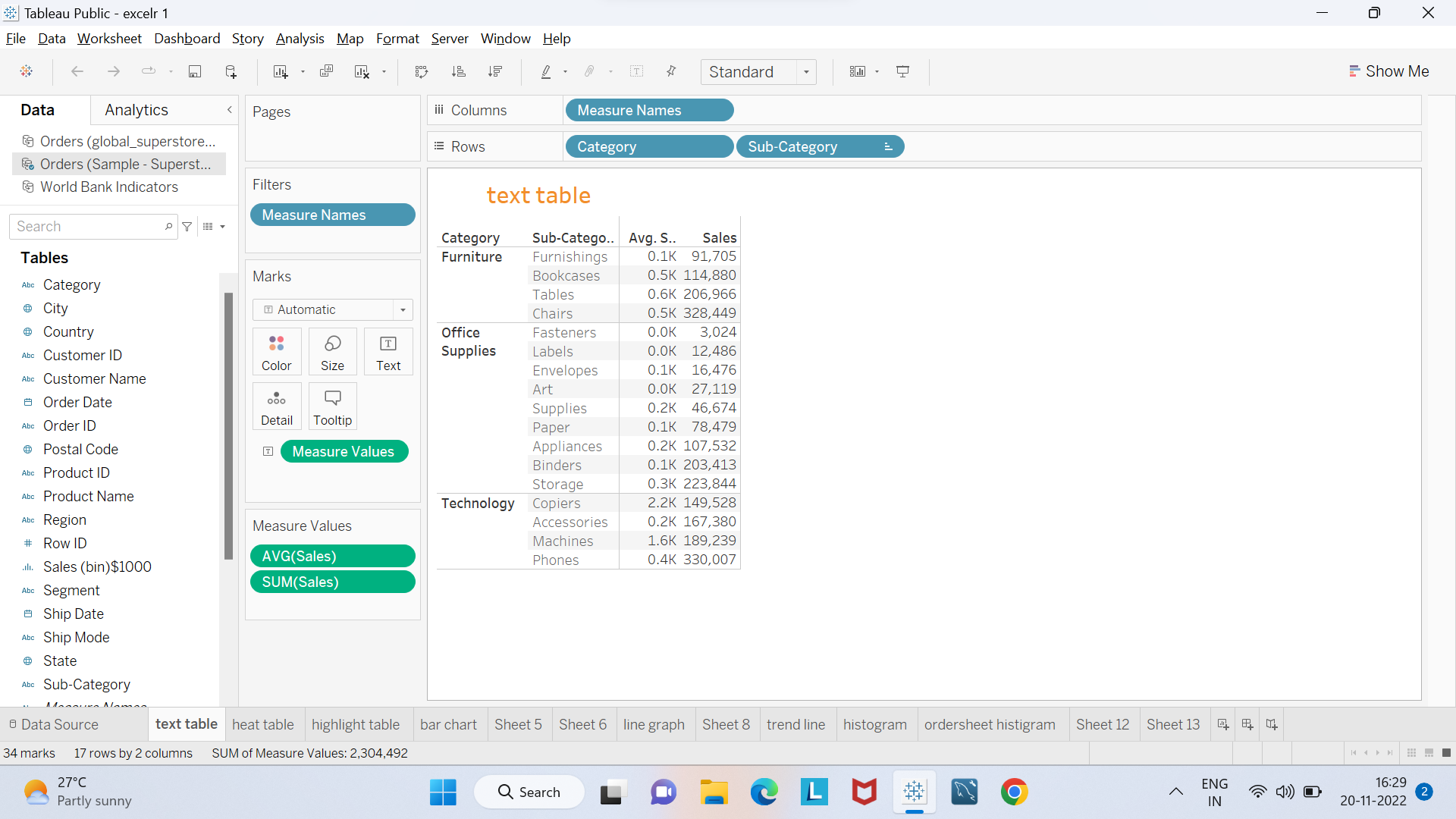
4.Which customer is having maximum of sales in the year 2012? - **Global Superstore**.

5.How much is profit share less in Pennsylvania when compared to New York? - **Sample Superstore**

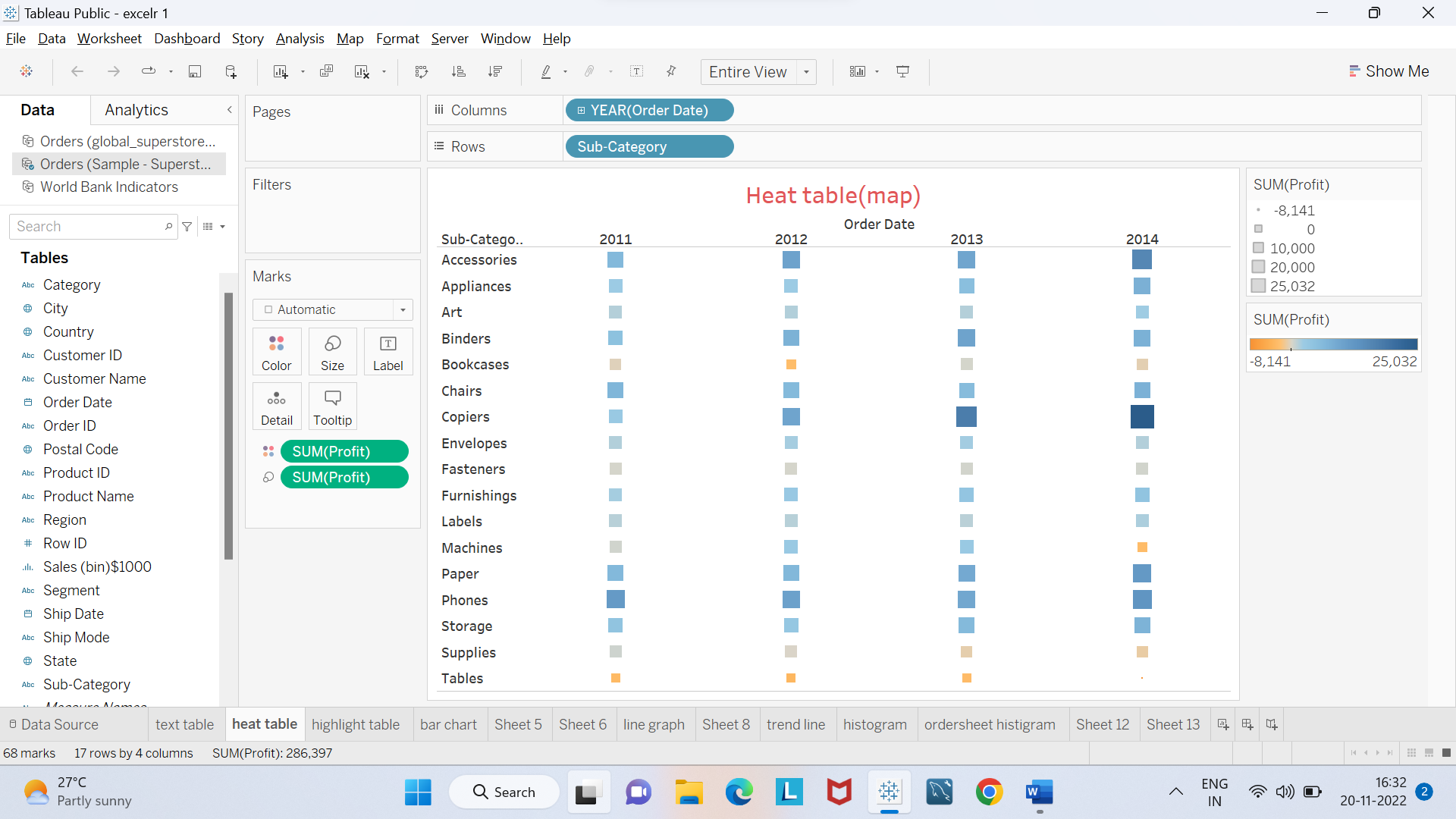
6.Check for the pane wise percentages of sales with Category, Sub- Category and quarter wise order date, also check for the Row wise grand totals and Column wise grand totals. - **Sample Superstore**

Solutions:

2.1 . Create a text table for the Avg (Sales) for each subcategory using Sample Superstore? List which Sub -Category is got Avg (Sale) more than $1000? - **Sample Superstore**



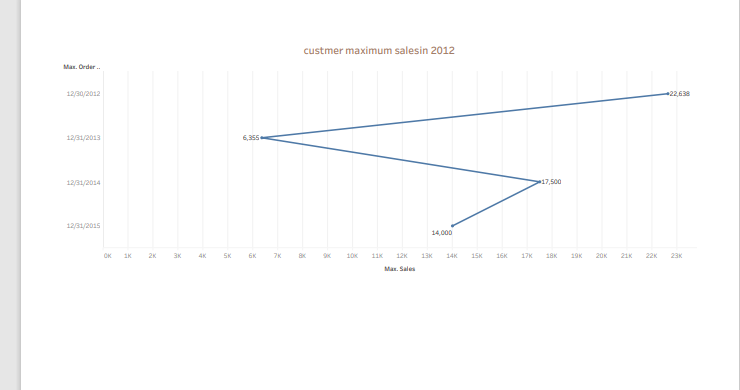
2.2: Create a Heat Table for the order date and Region against the Sub Category based in Count of Sales with two colours diverging that is distinguished by Sum of Profit - **Sample Superstore**



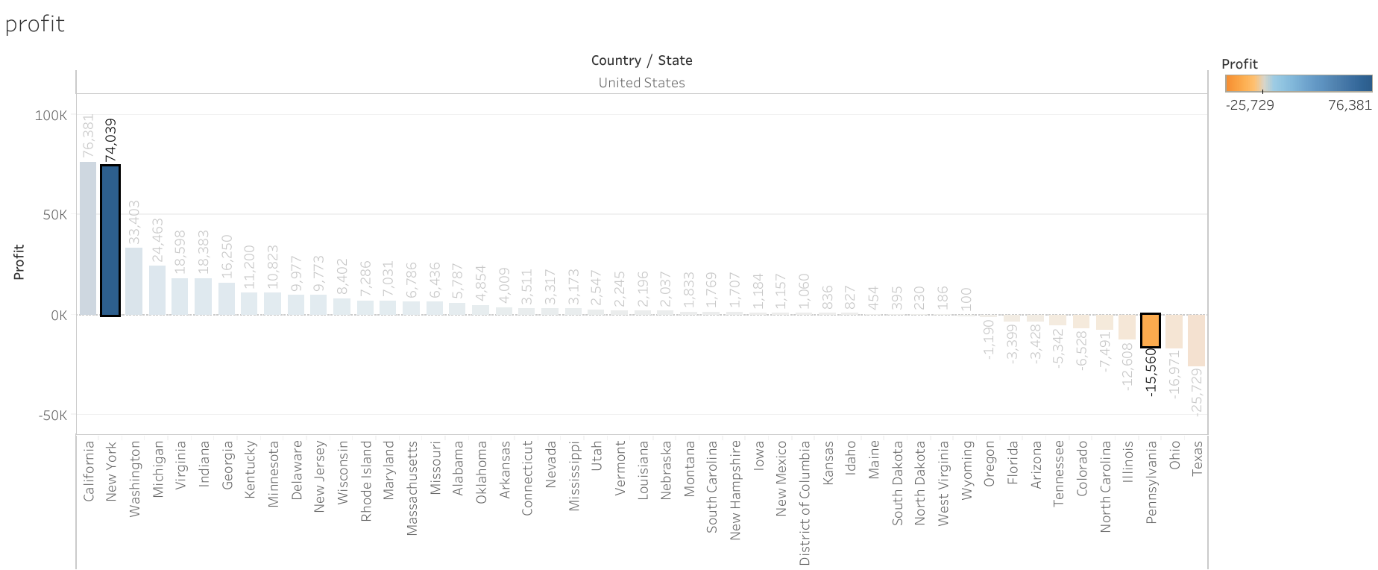
2.3: Create a Highlight table for the States for the Order Date Year whose highlighting is done based on Sum of profits - **Sample Superstore**



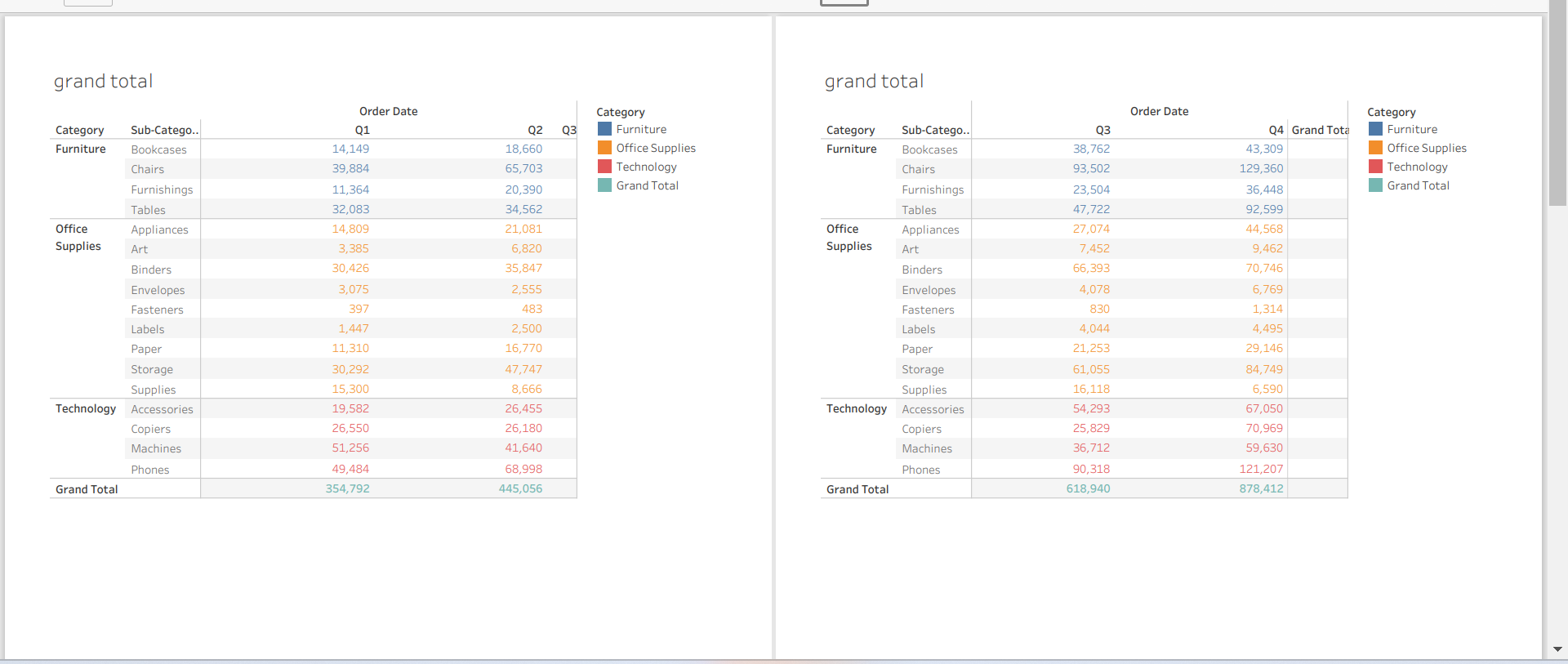
2.4 : Which customer is having maximum of sales in the year 2012? - **Global Superstore**



2.5: How much is profit share less in Pennsylvania when compared to New York? - **Sample Superstore**



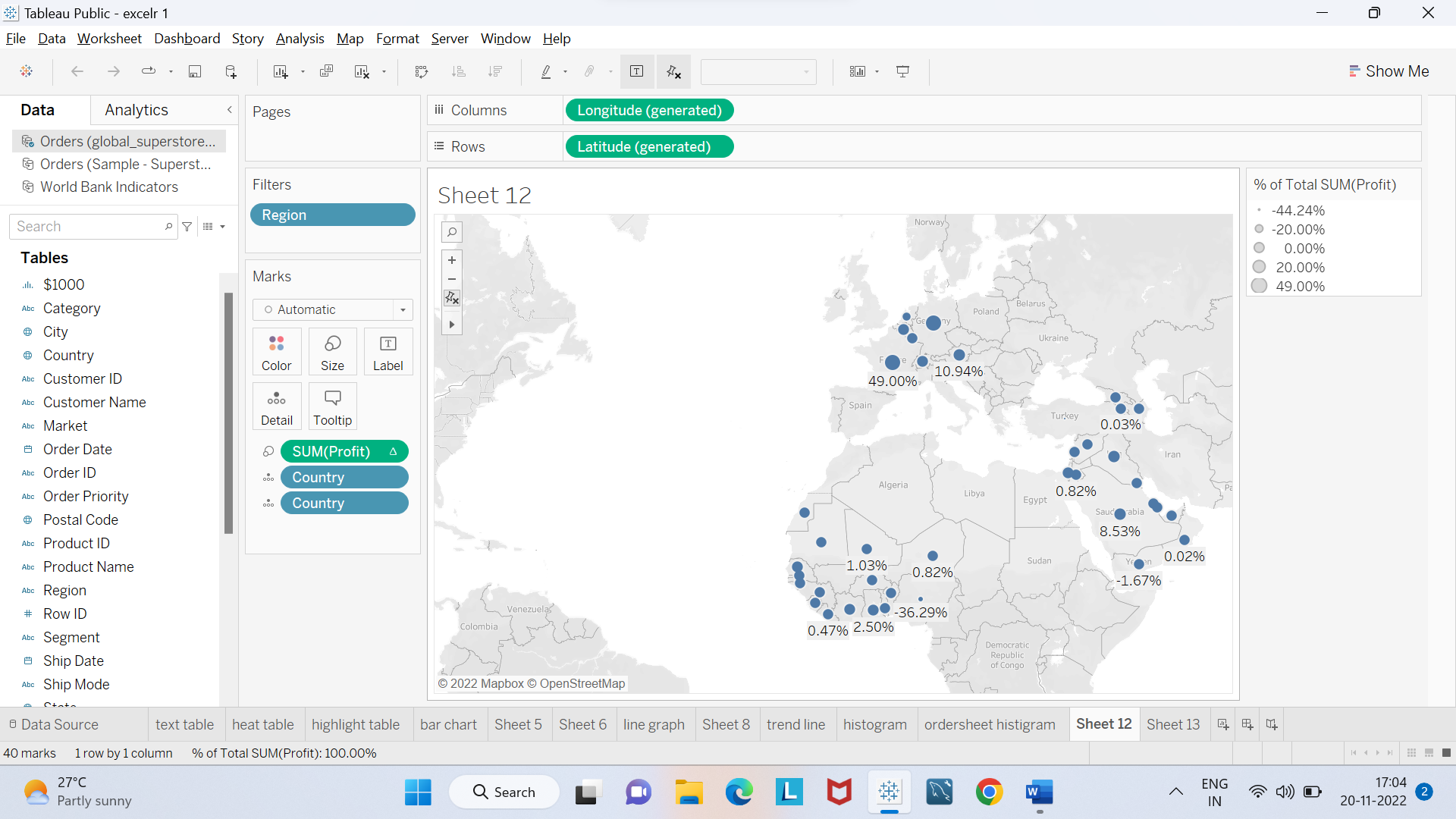
2.6: Check for the pane wise percentages of sales with Category, Sub- Category and quarter wise order date, also check for the Row wise grand totals and Column wise grand totals. - **Sample Superstore**



**3. Filled Maps, Symbol Maps:**

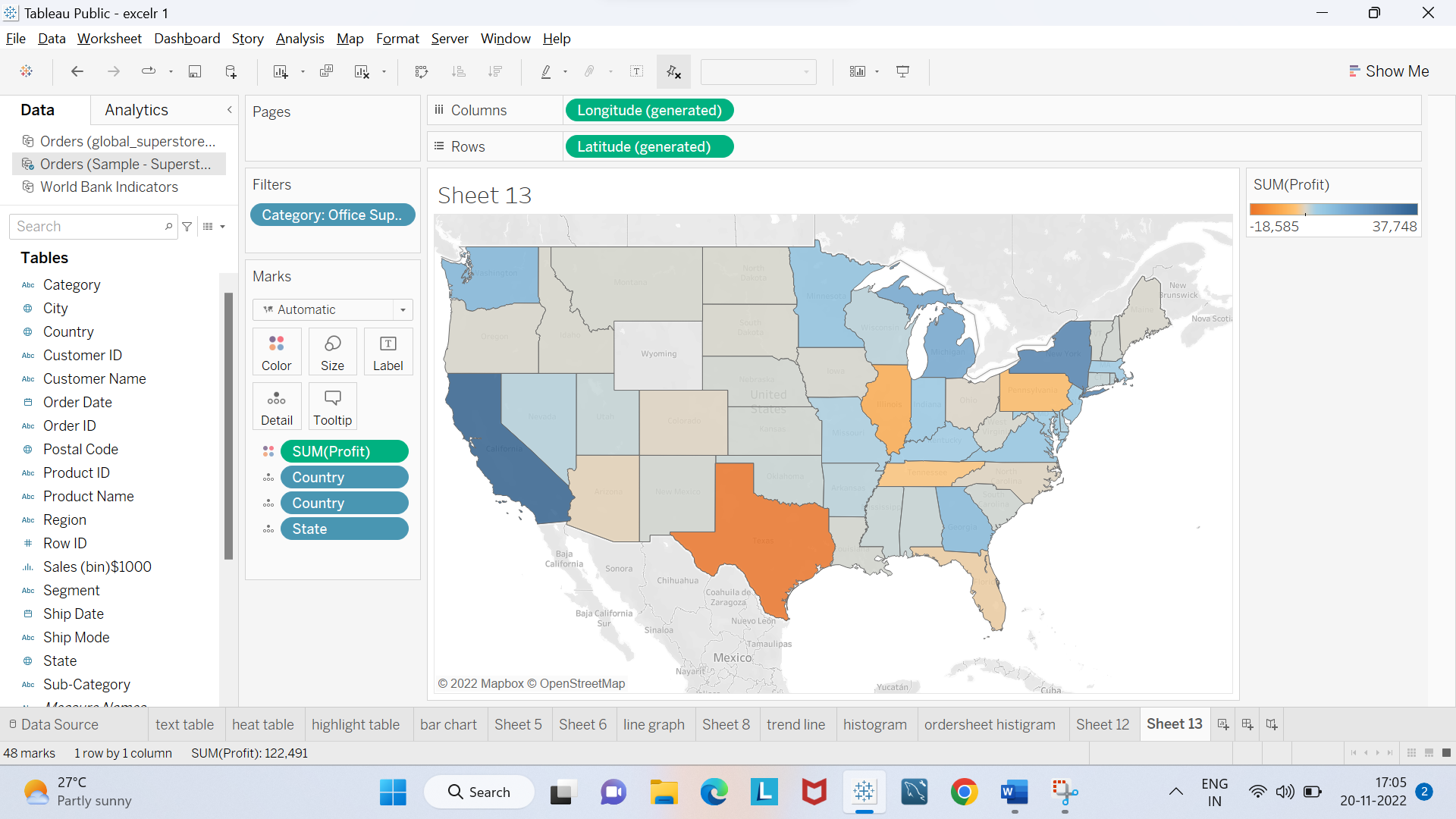
1. Use Global Superstore. Check Which Western Country in EMEA region has least profit percentage.
2. Use **“Sample Superstore. Xls”,** which state shares boarders only profit for tables
3. Use **“Sample Superstore. Xls”,** which state has no data for Profits for Office Supplies

3.1: Use Global Superstore. Check Which Western Country in EMEA region has least profit percentage.



3.2: Use **“Sample Superstore. Xls”,** which state shares boarders only profit for tables

**3.3**: Use **“Sample Superstore. Xls”,** which state has no data for Profits for Office Supplies

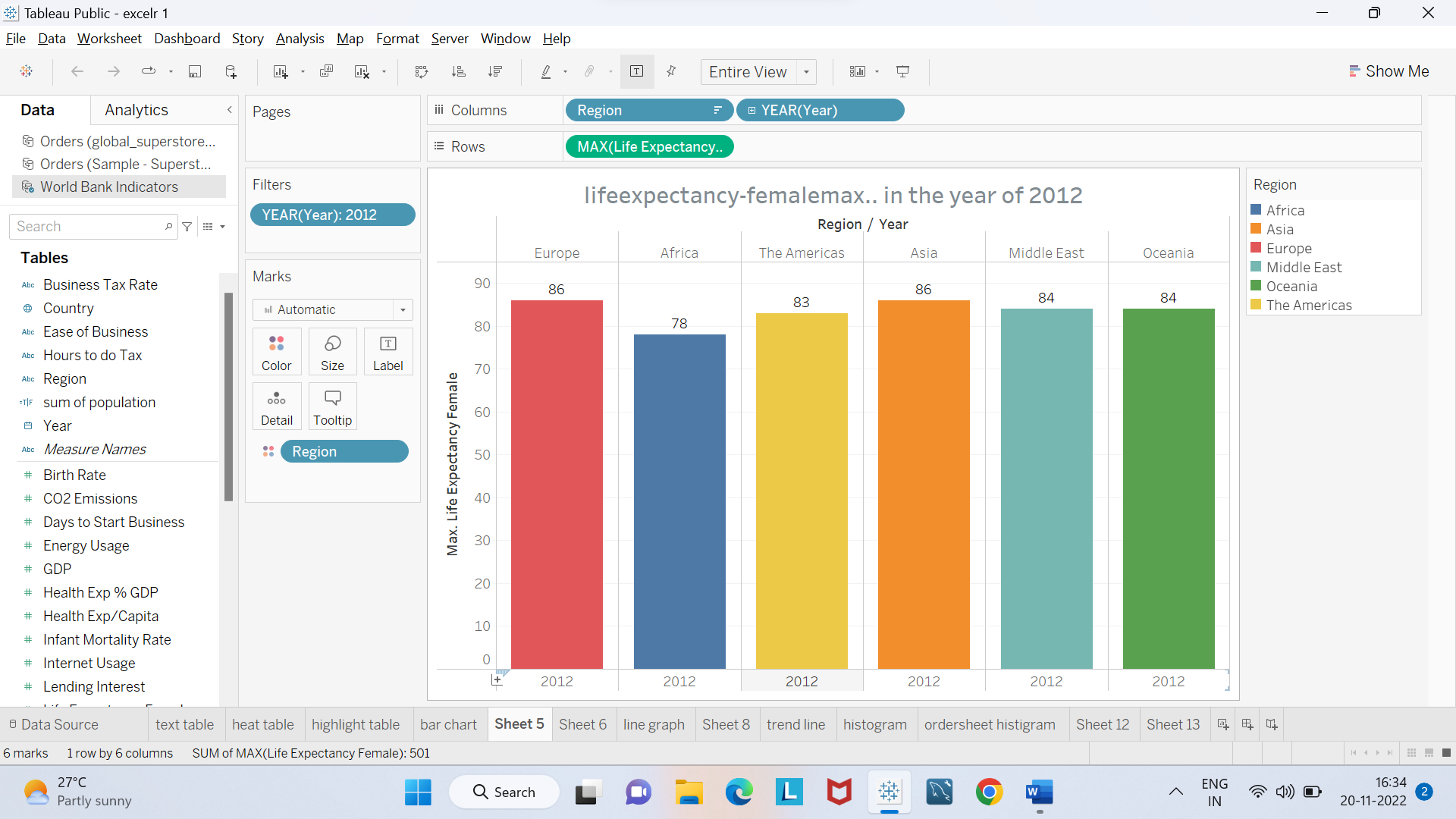


**4. Bar Charts, Stacked, Side by Side:**

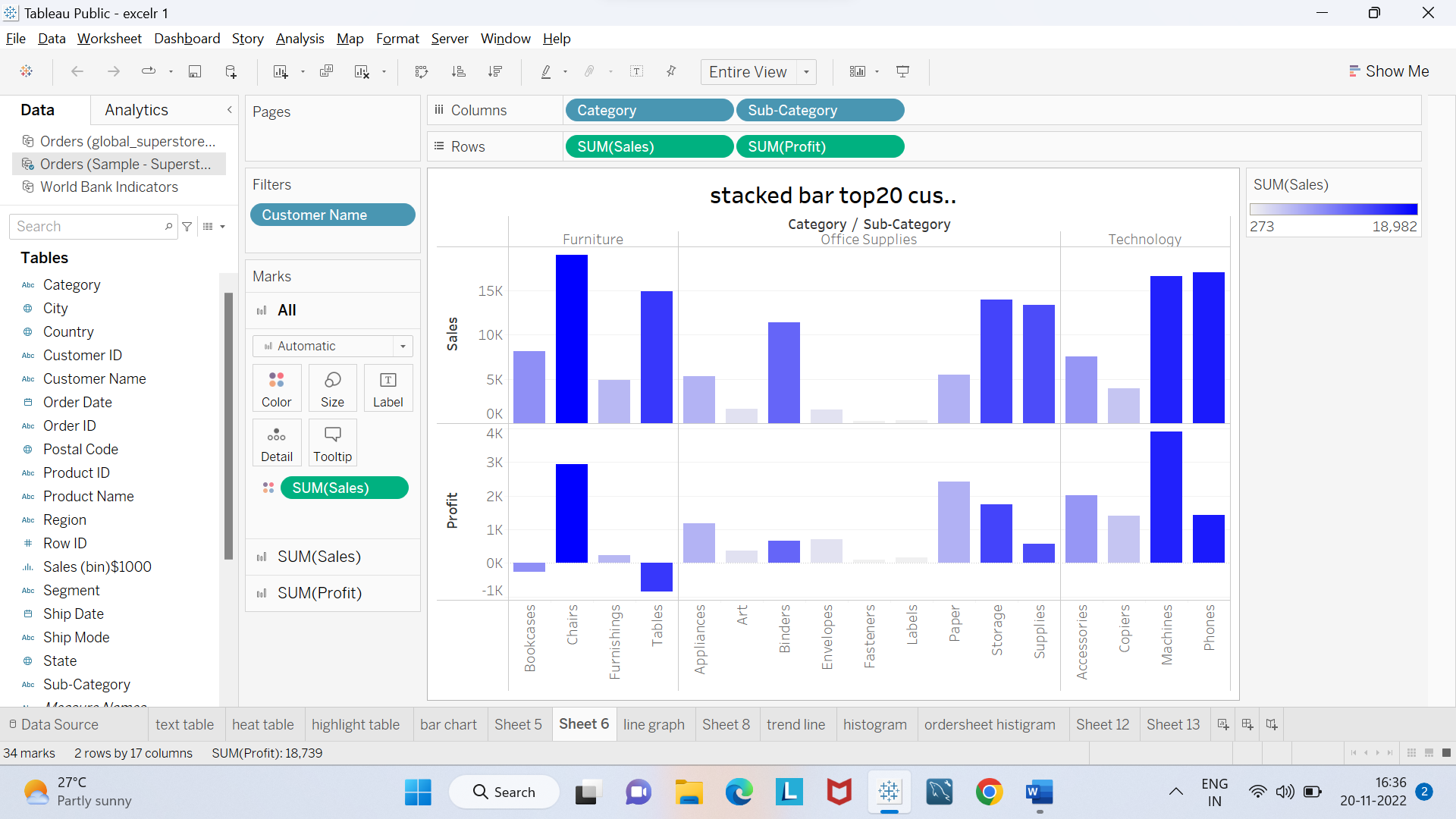
1. Which Customer name & Year is having all the Product Categories sum of profit less than over-all Average profit? - **Sample Superstore**
2. What is the Maximum of Life Expectancy Female for the region Africa & year 2012? - **World Indicators**
3. What is the share of the top 20 customers based on the sales amount compared to the customers based on profit amounts - **Sample Superstore**

4.1: Which Customer name & Year is having all the Product Categories sum of profit less than over-all Average profit? - **Sample Superstore**

4.2: What is the Maximum of Life Expectancy Female for the region Africa & year 2012? - **World Indicators**



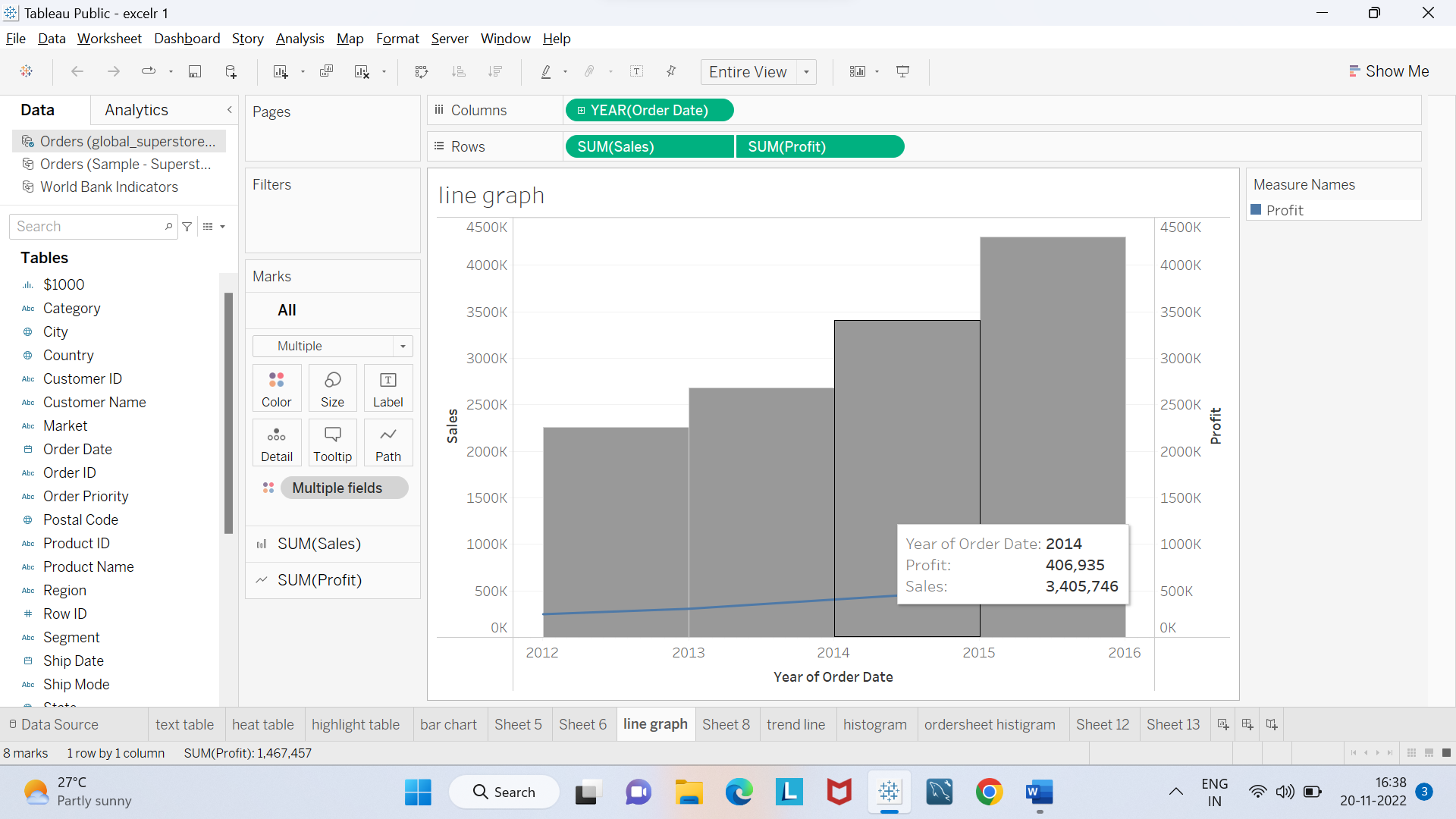
4.3: What is the share of the top 20 customers based on the sales amount compared to the customers based on profit amounts - **Sample Superstore**



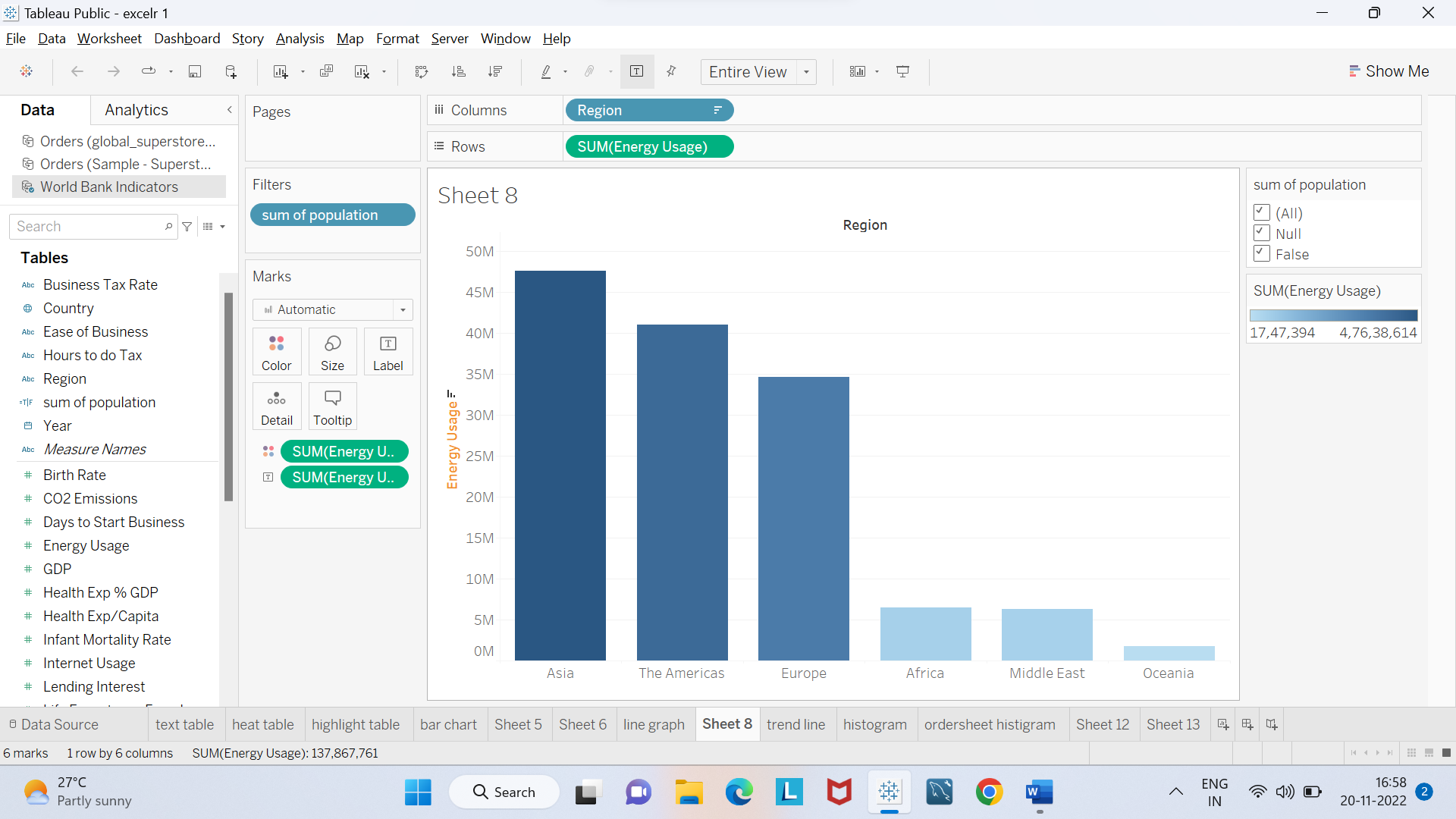
**5. Line Graphs, Dual Line, dual axis:**

1. How can you show two different graphs in one view? - **Global Superstore**
2. Which Region is having Sum of Energy Usage>1000000 and sum of Population 65+>10? - **World Indicators**

5.1: How can you show two different graphs in one view? - **Global Superstore**



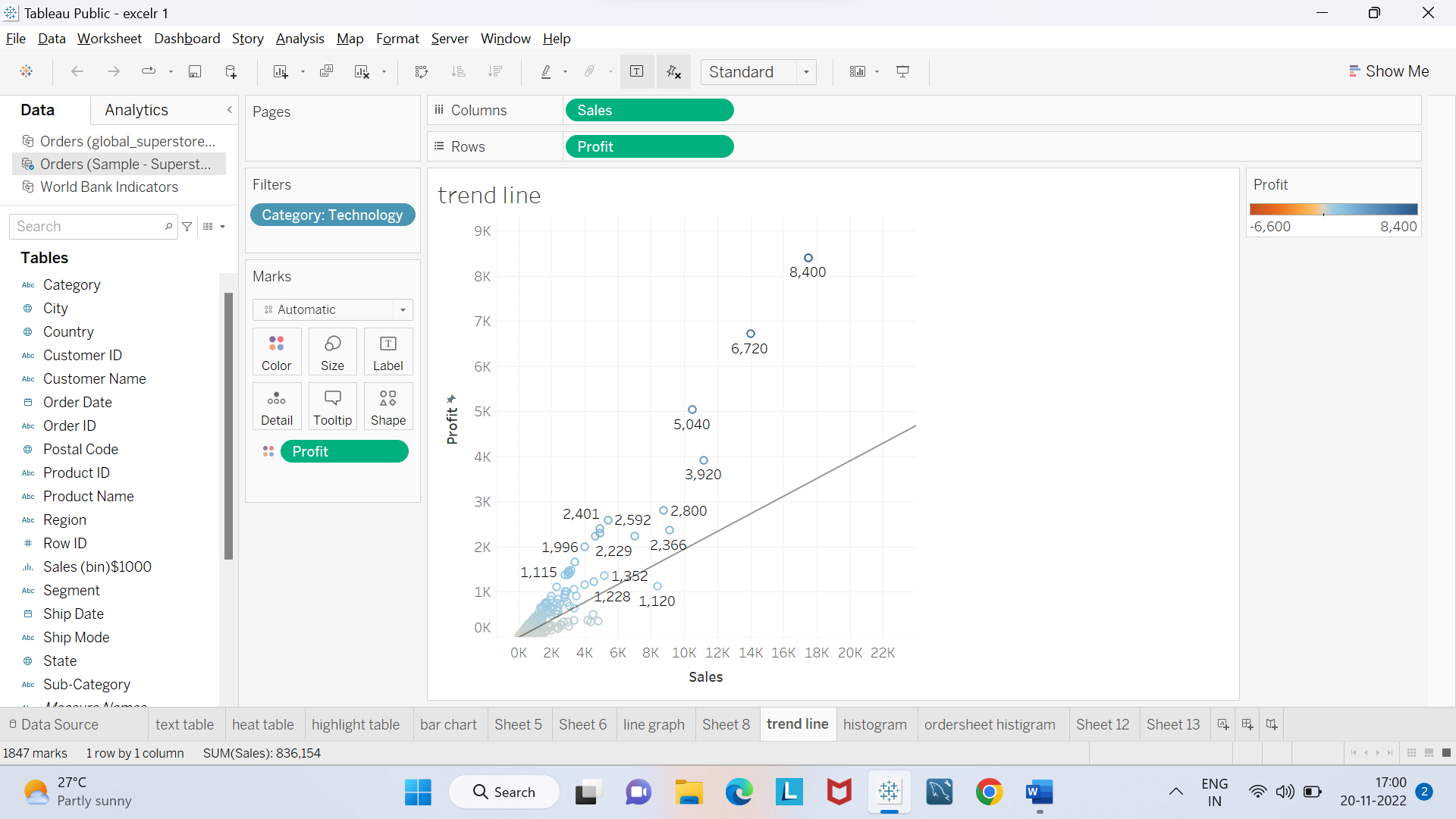
5.2: Which Region is having Sum of Energy Usage>1000000 and sum of Population 65+>10? - **World Indicators**



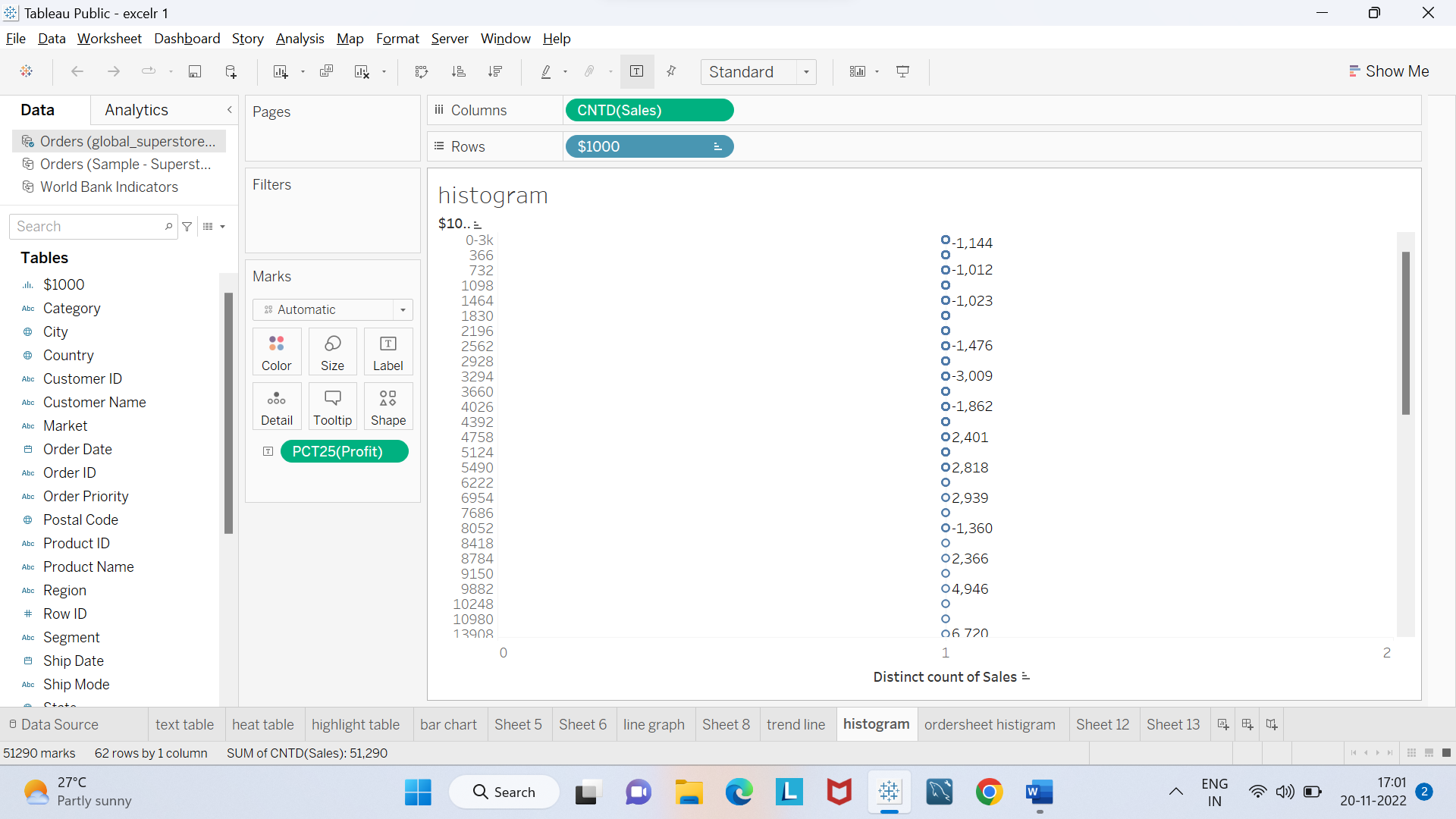
**6. Trendlines, Cluster, scatter Plot, boxplot, Word Cloud (Packed Bubbles), Histogram:**

1. Draw a trend line for profit as a linear function of sales only for product technology? - **Sample Superstore**
2. Create a histogram showing the number of Sales using Sales Bins of $1000. Which bins have profit ratios of more than 25%? - **Global Superstore**
3. Using “**Sample Superstore”**, use order sheet create a histogram showing the number of orders using sales bins of $1000.
4. equation for linear regression for products in Technology? Using **“Global Superstore**”, use the orders sheet, build a scatter plot showing the sum of sales on the x-axis and sum of profits on the y axis for all products (Product name). What is the
5. Use **“World Indicators”.**  Take Health Exp% GDP, Health Exp/Capita, Life Expectancy Male, Female. What are the variables that are considered to create the clusters by default?

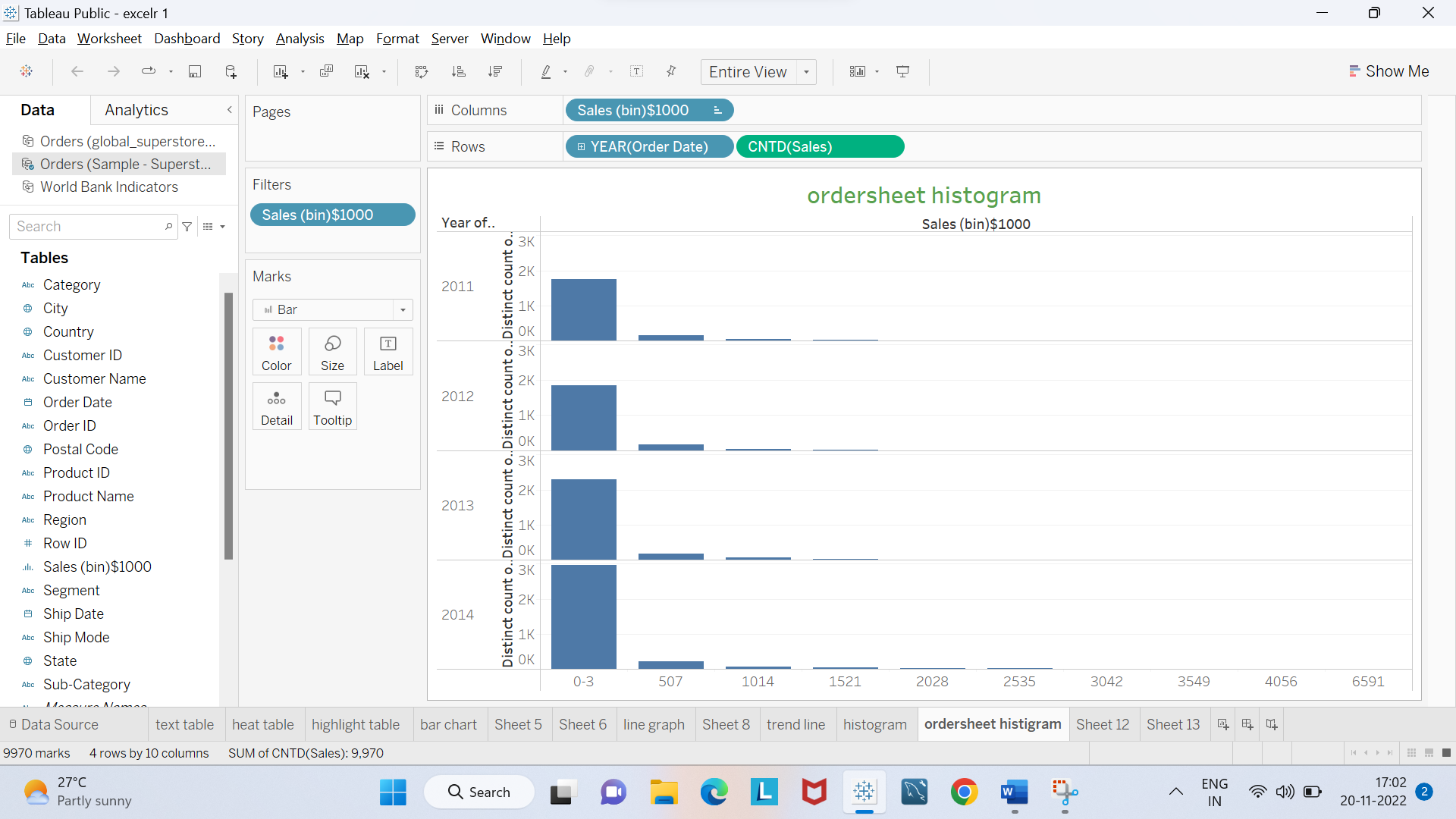
**6.1:** Draw a trend line for profit as a linear function of sales only for product technology? - **Sample Superstore**



**6.2**: Create a histogram showing the number of Sales using Sales Bins of $1000. Which bins have profit ratios of more than 25%? - **Global Superstore**



6.3: Using “**Sample Superstore”**, use order sheet create a histogram showing the number of orders using sales bins of $1000.



6.4: equation for linear regression for products in Technology? Using **“Global Superstore**”, use the orders sheet, build a scatter plot showing the sum of sales on the x-axis and sum of profits on the y axis for all products (Product name). What is the

6.5: Use **“World Indicators”.**  Take Health Exp% GDP, Health Exp/Capita, Life Expectancy Male, Female. What are the variables that are considered to create the clusters by default?

**7. Calculate Fields, Quick table calculations, LOD:**

1. How do you create a profit ratio using the Calculated fields?
2. Global Superstore data set; Region wise year wise sales are ranked. What is the rank of some country when compared to last year?
3. What percent of total profits do the top 10 customers by Sales represent? - **Sample Superstore**
4. Find the customer with the lowest overall profit. What is his/her profit ratio? - **Sample Superstore**
5. Ranking States based on Sales what is the rank of state which has sales crossed $20000. - **Sample Superstore**
6. What is the percent of orders which took more than 7 days on an average to deliver.
7. Use **“World Indicators”.** Without using table calculations what is the proper syntax to build a calculated field which will display overall total GDP on this view?