B0sco technical training society

Don bosco technical school

Okhla(Adjacent to Sukhdev Vihar Metro Station Gate No. 4),New Delhi-110025



Guru gobind singh Indraprastha university

Sector-16C, Dwarka, new Delhi-110075



Project on: Superstore Data Analysis

Submitted by: Rishab Khurana

Nitin Kumar

Kashish Kumar

Certificate

This is to certify that the project is entitled **“Supermarket Data Analysis”** is the work carried out by Rishab Khurana, Nitin kumar, kashish kumar student of BCA Course, **“BOSCO TECHNICAL TRAINING SOCIETY,** **affiliated to Guru Gobind Singh Indraprastha University, Dwarka, New Delhi”.** This project was an authentic work done by him under my supervision and guidance.  
This project has not been submitted to any other institution for the award of a diploma.

Date:  
(name of project guide)  
**BOSCO TECHNICAL TRAINING SOCIETY,** **affiliated to Guru Gobind Singh Indraprastha University, Dwarka, New Delhi**

acknowledgement

we have taken efforts in the project. However, it would not have been possible without the kind support and help of many individuals and organizations. We would like to extend our sincere thanks to all of them.

We are highly indebted to “Mr. Rakesh kumar” for his guidance and constant supervision as well as for providing necessary information regarding the project and also for his support in completing the project.

We would like to express our gratitude towards our professors for their kind co-operation and encouragement which helped us in completing of this project.

We would like to express our special gratitude and thanks to industry person for giving us such attention and time.

Our thanks and appreciations also go to our colleague in developing the project and people who have willingly helped us out with their abilities.

Submitted by:

Rishab Khurana

Nitin Kumar

Kashish Kumar

Content

|  |  |  |
| --- | --- | --- |
| **S.No.** | **Topic** | **Pg no.** |
|  | **Introduction of the project** | **5** |
|  | **Objective** | **6** |
|  | **Introduction about the Technology** | **7** |
|  | **Database Link** | **8** |
|  | **SRS**   * **Hardware And Software Requirement** | **9-12** |
|  | **Code And Screenshot** | **13-31** |
|  | **Testing** | **32-33** |
|  | **Future Scope** | **34** |
|  | **Bibliography/ References** | **35** |

Introduction about the PROJECT.

Superstores industry comprises of companies that operate by having large size spaces which store and supply large amounts of goods. The superstore industry is comprised of extensive stores that sell a typical product line of grocery items and merchandise products, such as food, pharmaceuticals, apparel, games and toys, hobby items, furniture and appliances. The analysis of such industry is of great importance as it gives insights for the sales and profits of various products. Our analysis is based on a superstore dataset for US country where the products are ordered between 2016 and 2019.

OBJECTIVE.

To find various supermarket statistics such as-

* Trend in profit/sales over time (years/months/quarters).
* Trend in profit/sales over region (years/months/quarters).
* Product (Segment/Category) with highest and lowest sales.
* Forecasting future sales according to shipping date.

With this analysis, the Superstore can identify various aspects of the shopping pattern and take measures if required.

Introduction about the TECHNOLOGY.

Tableau is data visualization software designed to make the process of presenting data more manageable. Tableau provides many different versions of their software, such as Tableau Desktop, Reader, Public, Server, and Online.  For this project, we will be using Tableau Public. Tableau Public is a free version of the software.

Tableau Desktop is considered to be the gold standard in contemporary data visualization tools, as evidenced by winning Gartner’s Magic Quadrant award for Analytics and BI Platforms nine years in a row. Its power and ease of use encourage vast data exploration. Interactive workbooks and dashboards help users uncover insights as queries or hunches arise, and to make discovery easy via visual patterns.

Tableau's Applications

Tableau bridges the gap between a workforce with little or no technical skills and tasks such as:

Data visualization

Business intelligence

Data blending

Data mining

Data collaboration

Query translation into analytics

Real-time data analysis

No-code data queries

Importing large data sets

Managing large metadata

**Database Link**

[**https://github.com/Rishabkhurana/Tableau-Project**](https://github.com/Rishabkhurana/Tableau-Project)

**Software Requirement Specification**

**Introduction**

In this present project we are working on superstore dataset. This dataset is taken from [https://www.kaggle.com/datasets/bravehart101 /sample-supermarket-dataset](https://www.kaggle.com/datasets/bravehart101%20/sample-supermarket-dataset) . We firstly fetch the file form Kaggle in Tableau. Then we used graphs maps and other features of Tableau to mold information as we intended.

* **Purpose**

The purpose of our project is to represent the performance of the superstores around United State of America and forecasting future sales according to shipping date.

* **Definition**

Our project aids companies and government organizations to collect data and identify patterns in the data. This helps to generate better insights for the organization with regards to decision-making, and sometimes, automates the decision-making process itself. Our project can help to transform business and reduce the cost and time to make quick decisions. The data analytics industry has made a significant impact on Indian companies, especially in their day-to-day operations. Our project has enabled the industry to completely automate some aspects of business. Data is considered to be the new oil, and industries.

* **Overview**

Data aggregation and analysis

Our project can assist in the collection of big data and solve major issues faced by big companies.

Discovering potential of untapped markets

India has numerous large untapped markets that can be identified, analysed, and tapped into through data analytics.

Efficient allocation of resources

Our project can help a government, company or even an individual to allocate its resources efficiently.

Increased efficiency

Forecasting helps companies boost efficiency.

**Specific Requirement**

Once the components of the framework have been established can review the data that will need to be collected in order to measure the program indicators. This should also involve listing where data to be collected is being recorded and the questions which they hope the project will respond to. As each question is reviewed, consider what data will address it. Questions may be listed in a table format alongside the specific variables being monitored to ensure there are no gaps in the data. Once all questions and indicators have been linked in this way.

* **Functions**

Employment analysis is a process of inspecting, cleansing, transforming, and modelling data with the goal of discovering useful information, suggesting conclusions, and supporting decision-making. This allows us to make informed decisions and to stop guessing

* **Logical Database Requirement**

A logical database is the collected information stored on multiple physical disk files and hard drives within a computer. This database provides a structure to house all the accumulated information within the device and determines the relationships between different types of files and programs.

A logical database can stretch over multiple physical hard disks information files. The data storage unit is still a single database for information retrieval purpose. To have a logical database, all given hard disks and information files must be accessible from a single source.

* **Software requirement**

Microsoft Excel, Tableau Desktop

* **Software System Attribute**

Reliability

Identifies the degree to which different raters agree in the components of a target work role or job. It is typically assessed using Tableau

Availability

This describes the process of identifying the number of members belonging to a minority or a disadvantage groped who are qualified for any available position within the organization. Those are often candidates who suffer discrimination within a culture.

Security

It is an approach to cybersecurity that uses data collection, data aggregation and analysis tools for threat detection and security monitoring. An organization that deploys security analytics tools can analyse security events to detect potential threats before they can negatively affect the company’s infrastructure and bottom line.

This combines big data capabilities with threat intelligence to help detect, analyse and mitigate threats, as well as persistent cyberthreats and targeted attacks from external bad actors.

Maintainability

It is used to predict maintainability of systems and equipment of any type, including avionics, ground and shipboard electronics, mechanical equipment, etc.

Our project will receive a uniform checking for bug and problem, throughout the time period it will maintain and be ready to use all the time.

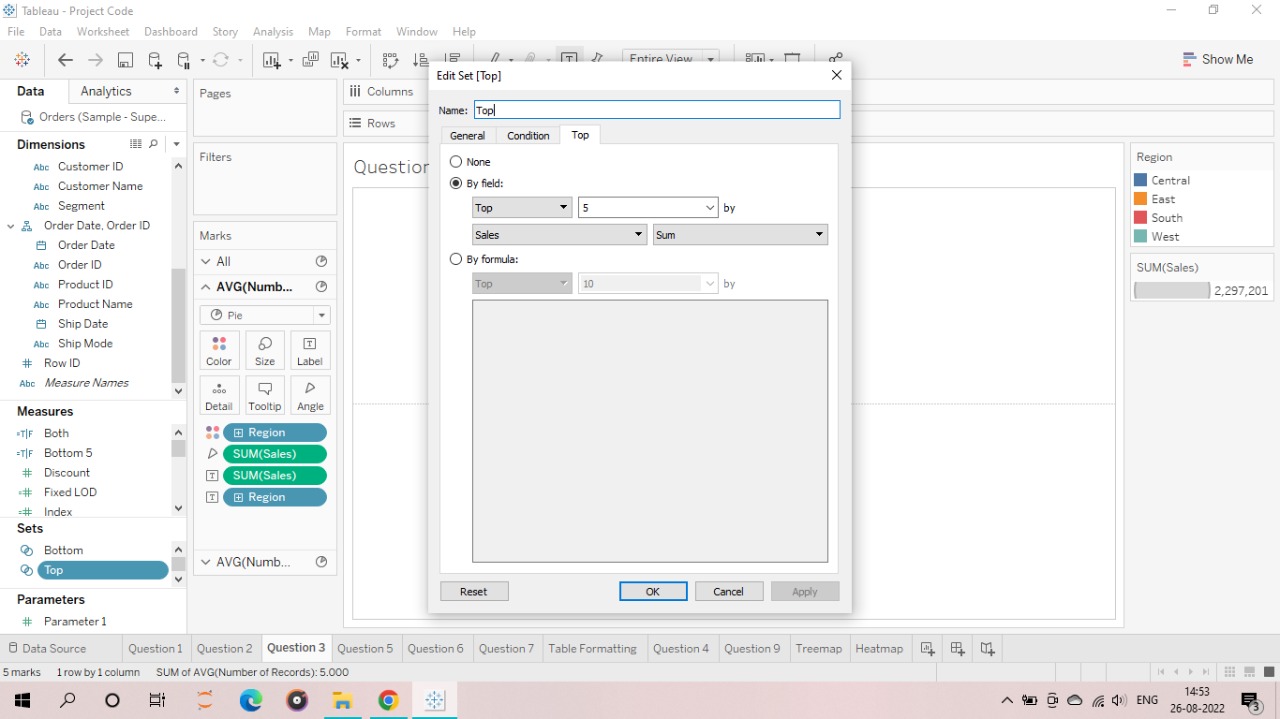
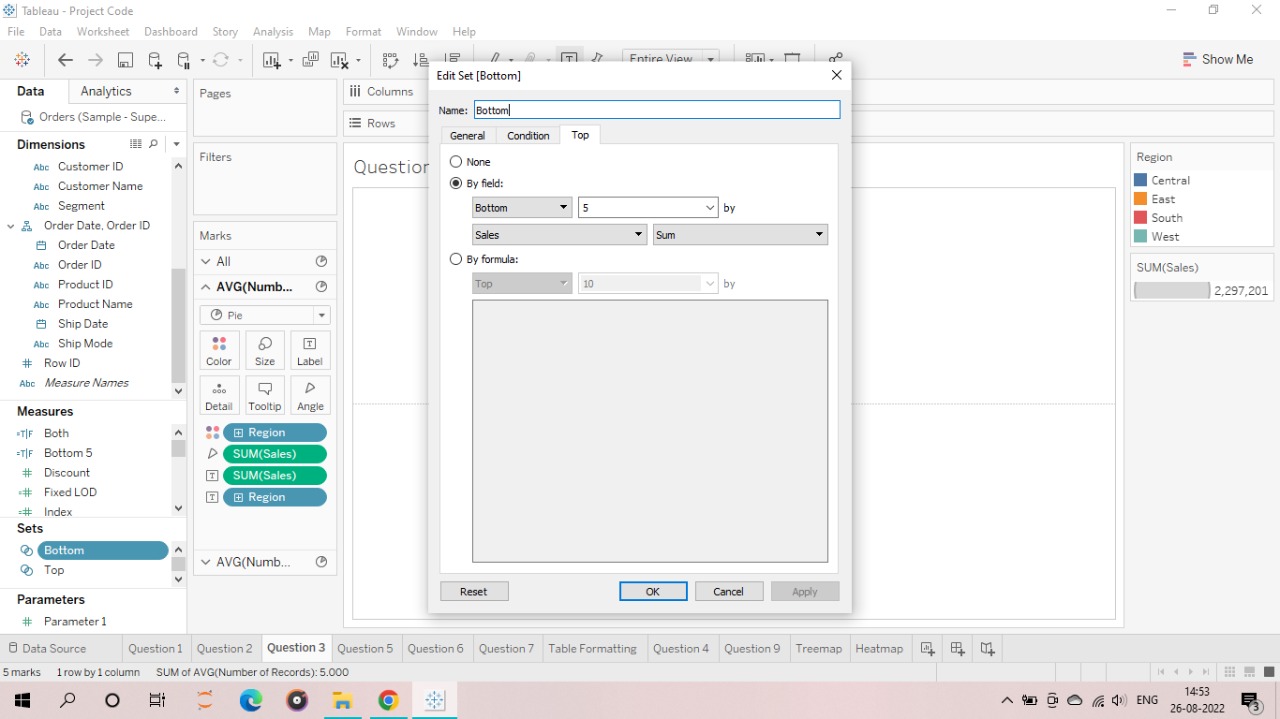
Portability

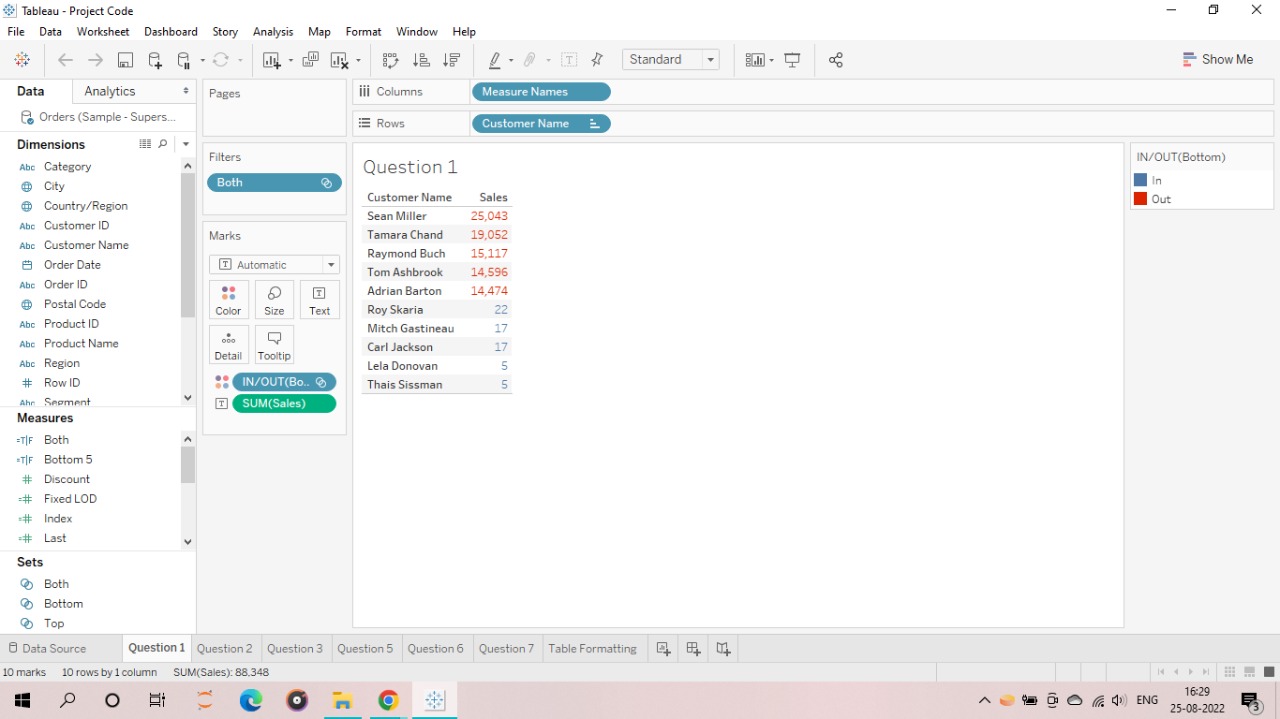
Portability is a characteristic attributed to a computer program if it can be used in an operating system other than the one in which it was created without requiring major rework. Porting is the task of doing any work necessary to make the computer program run in the new environment

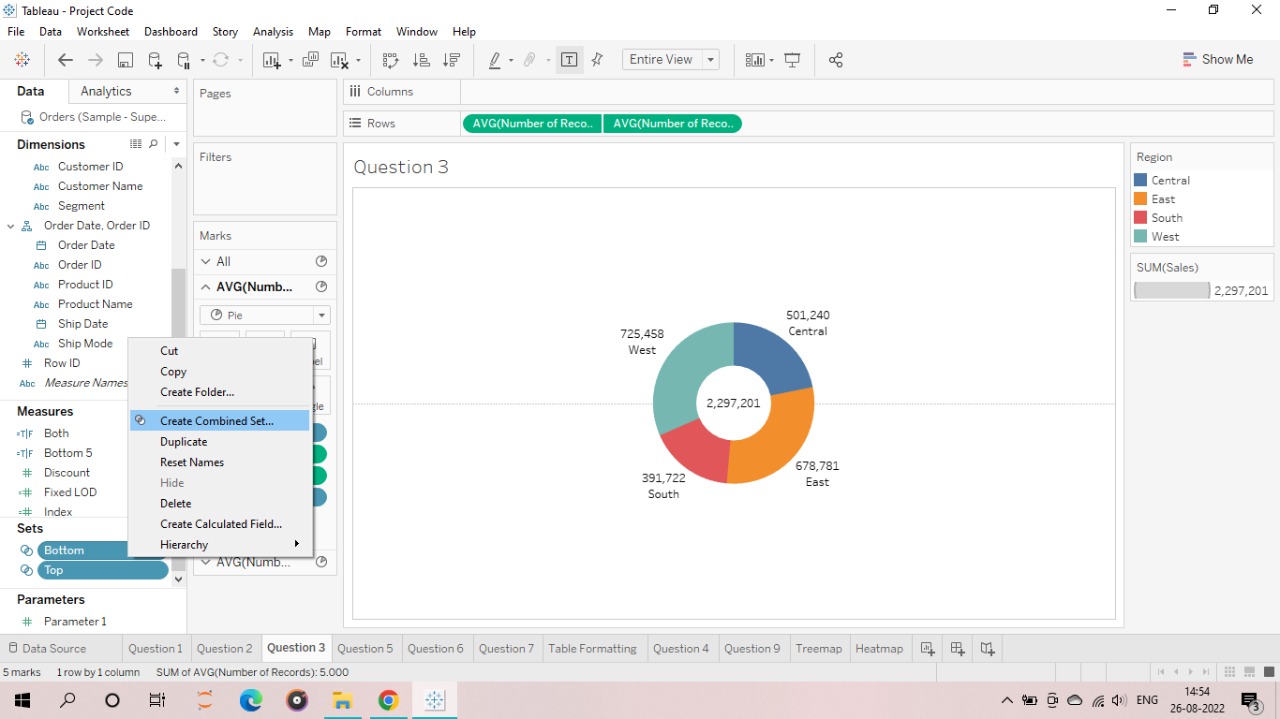
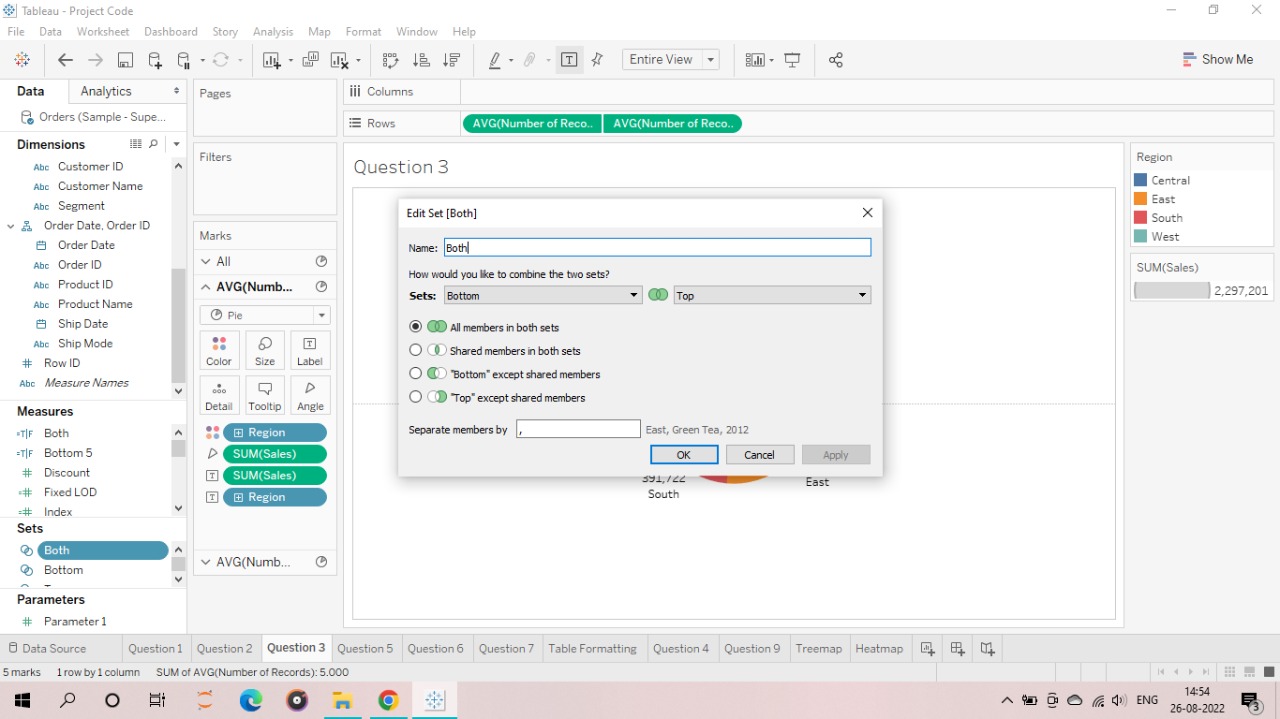
Code & Screenshot

**1. Showing Top 5 and Bottom 5 Customer in a single sheet.**

**CODE:**

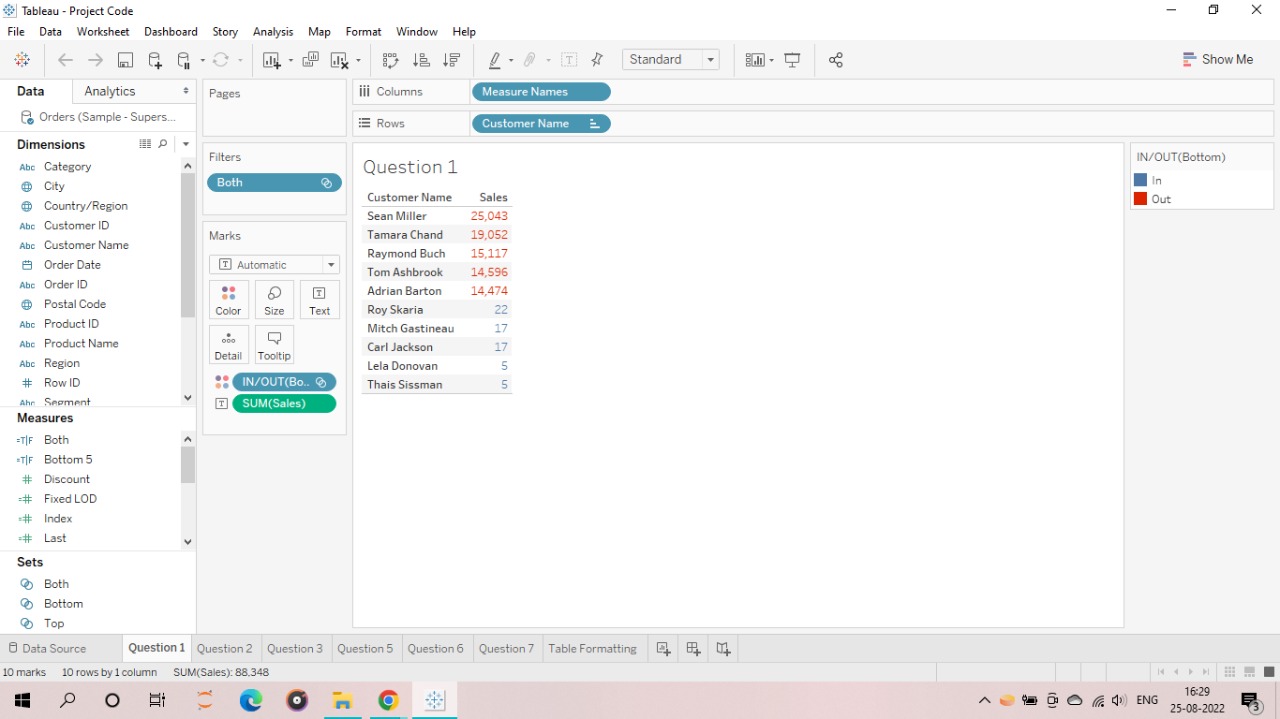






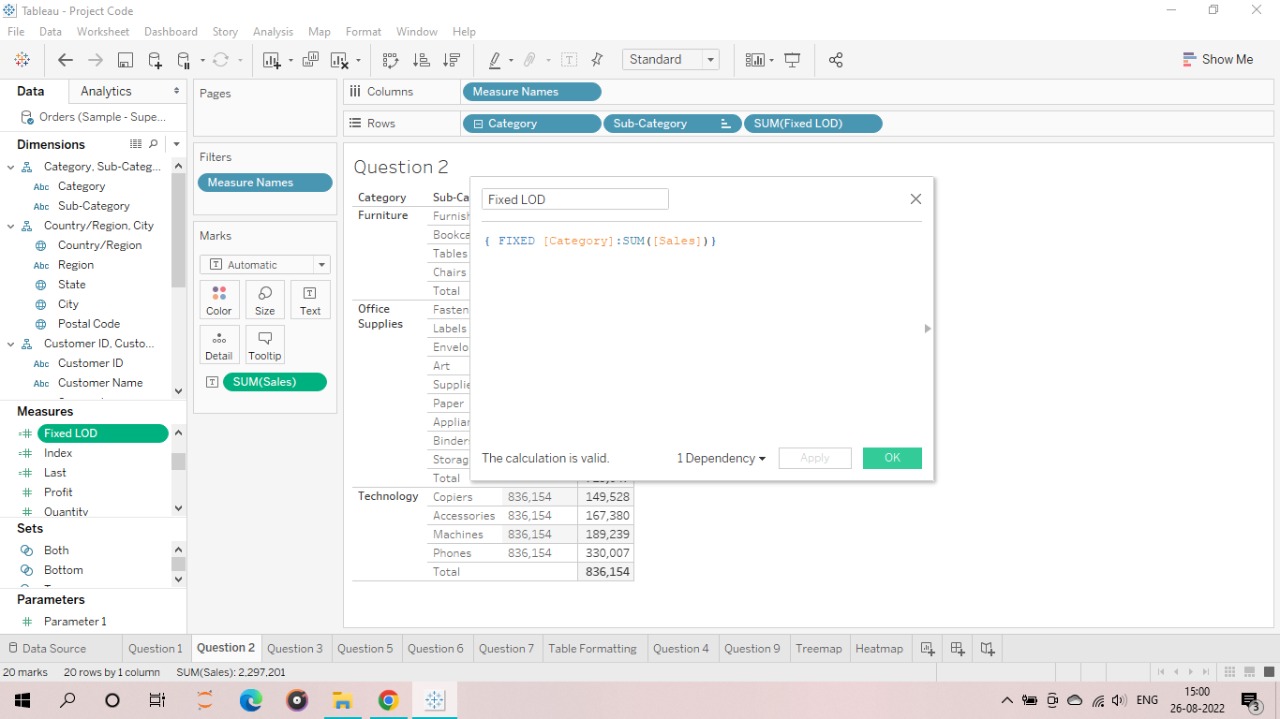
**OUTPUT:**

Top 5 Customers are shown in red and Bottom 5 customers are shown in blue

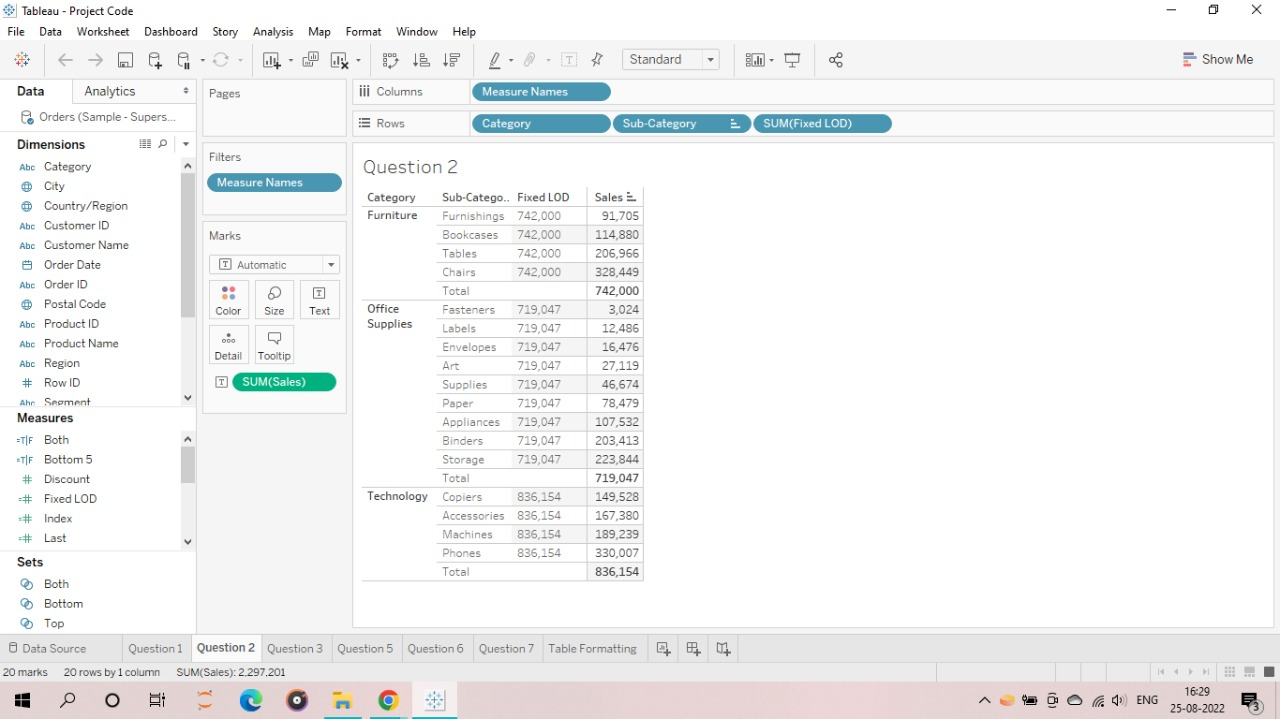


**2. Using Fixed LOD to show category wise sales.**

**CODE:**

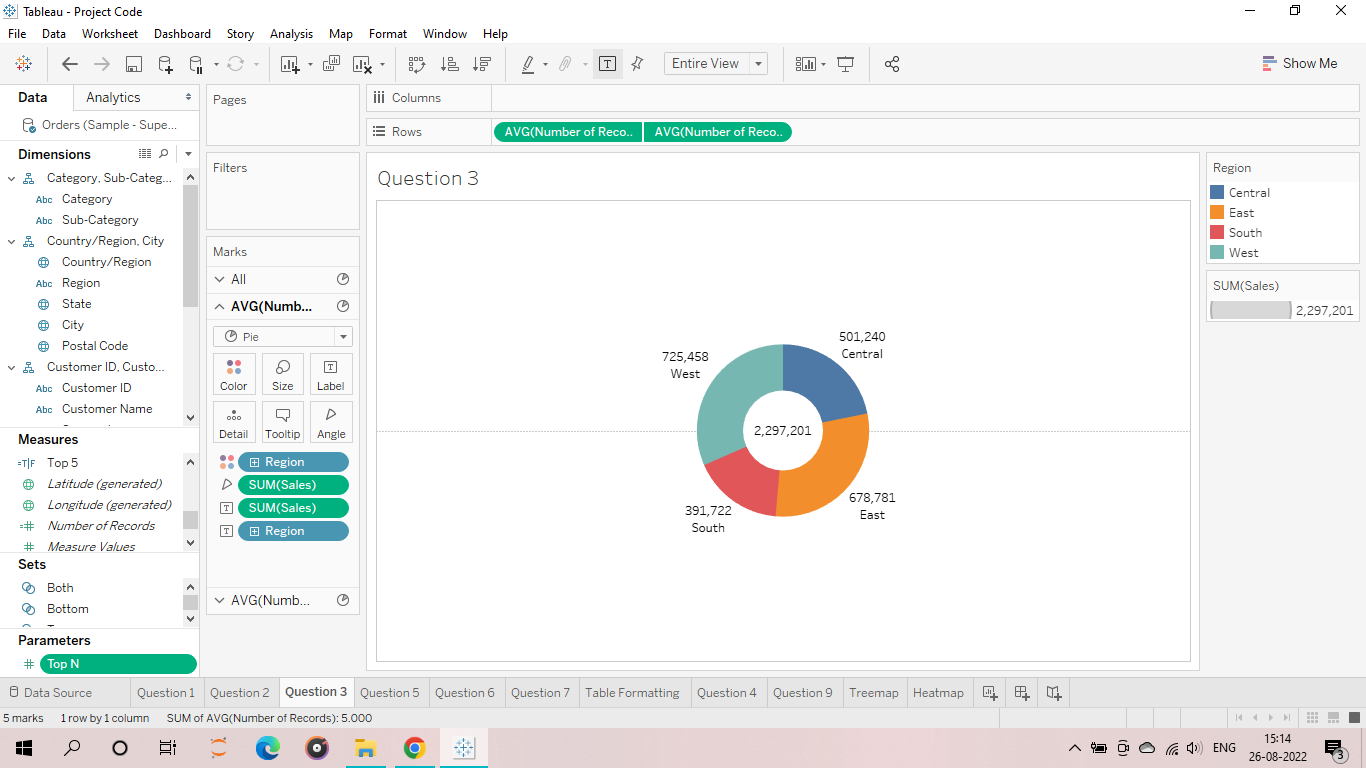


**OUTPUT:**

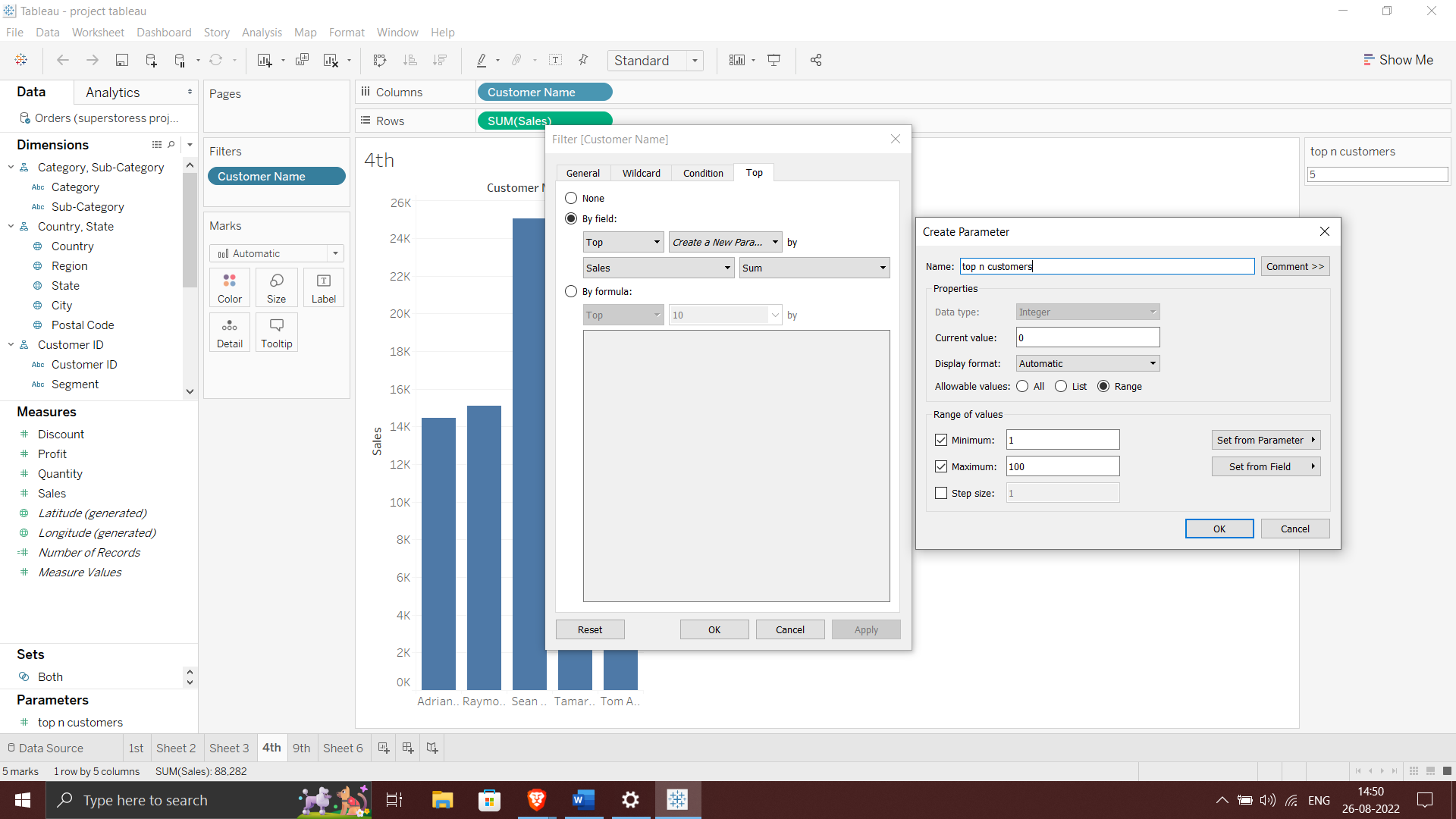


**3. Designing donut chart.**

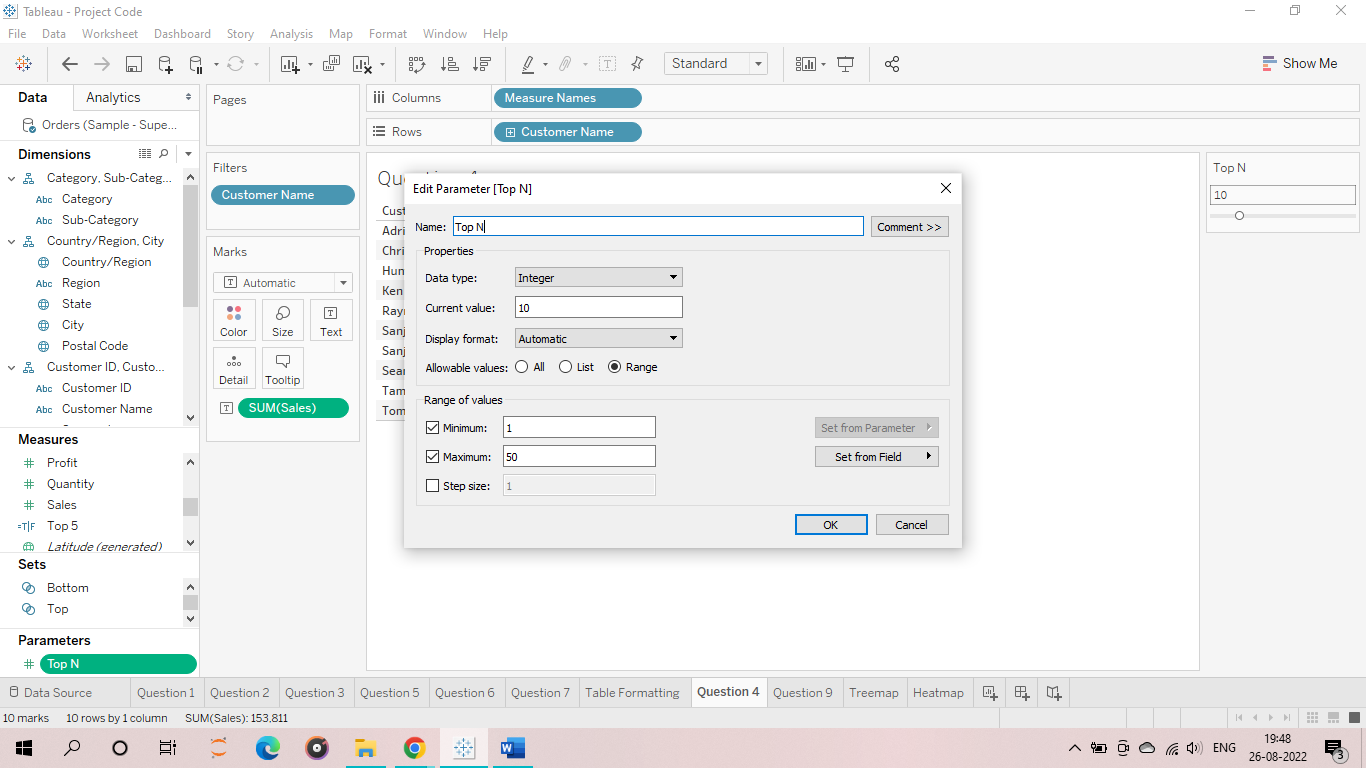
Creating a donut chart that shows Category-wise sales in different region and displaying the total sales as well in this donut chart.

****

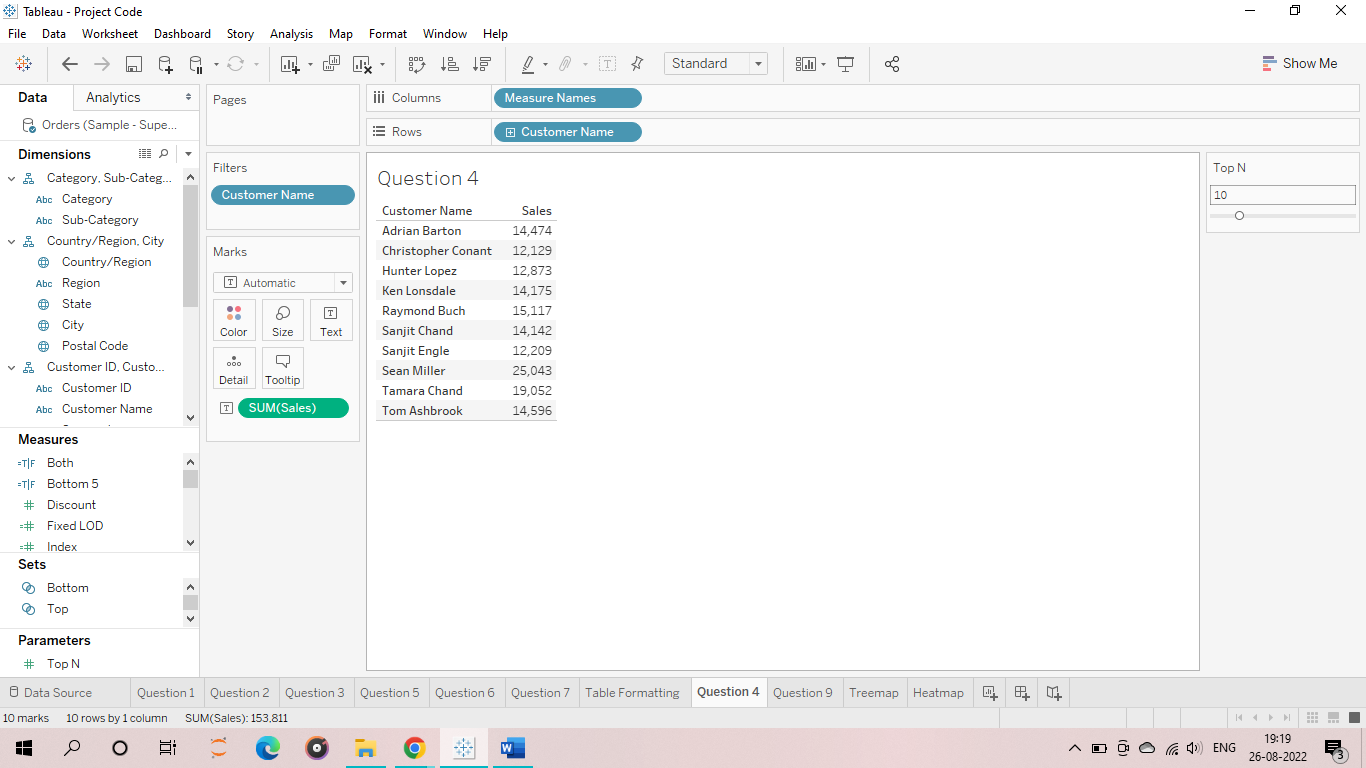
**4. Show Top N Customers by sales using parameter.**

**CODE:**

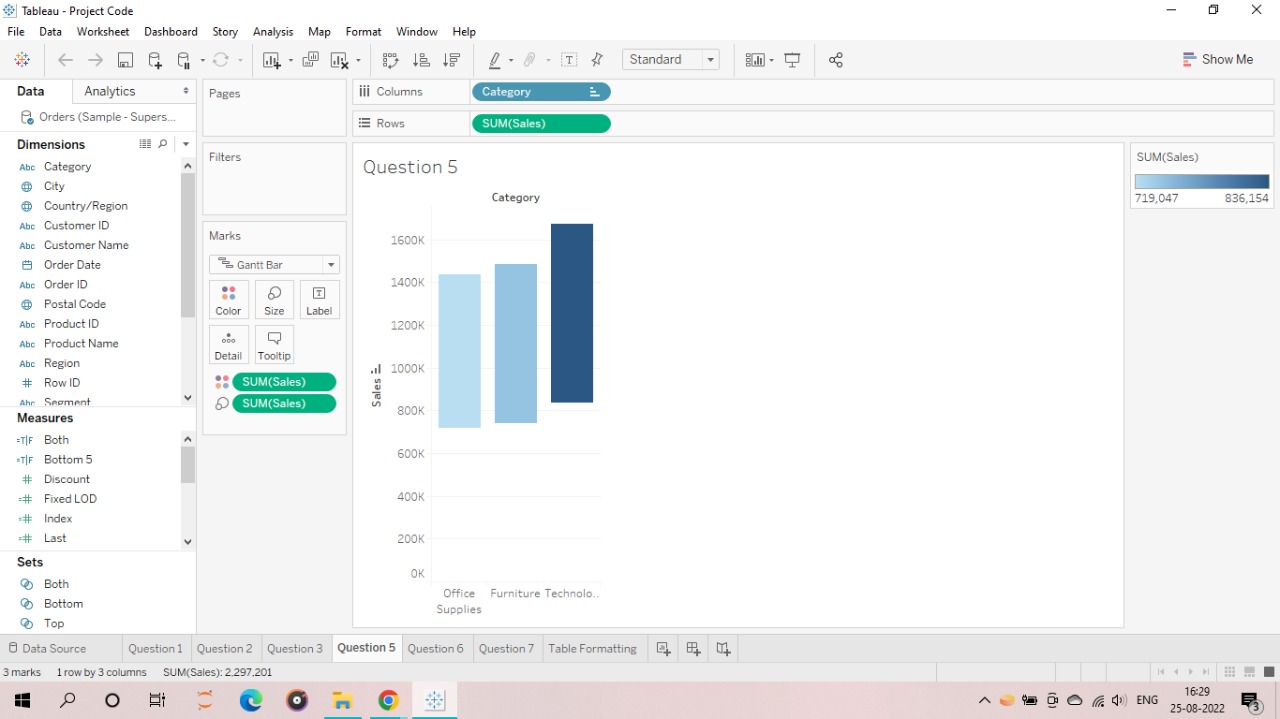
Parameter



**OUTPUT:**

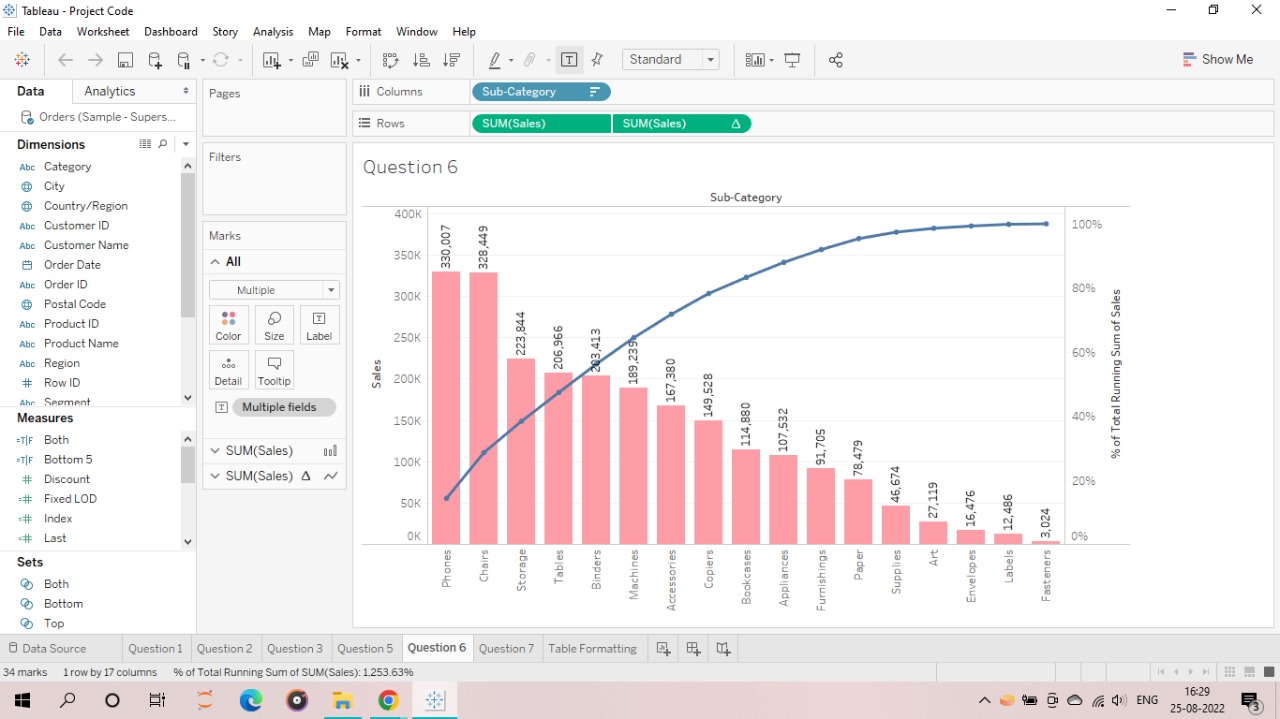
****

**5. Show water fall chart in tableau by using sales and category.**



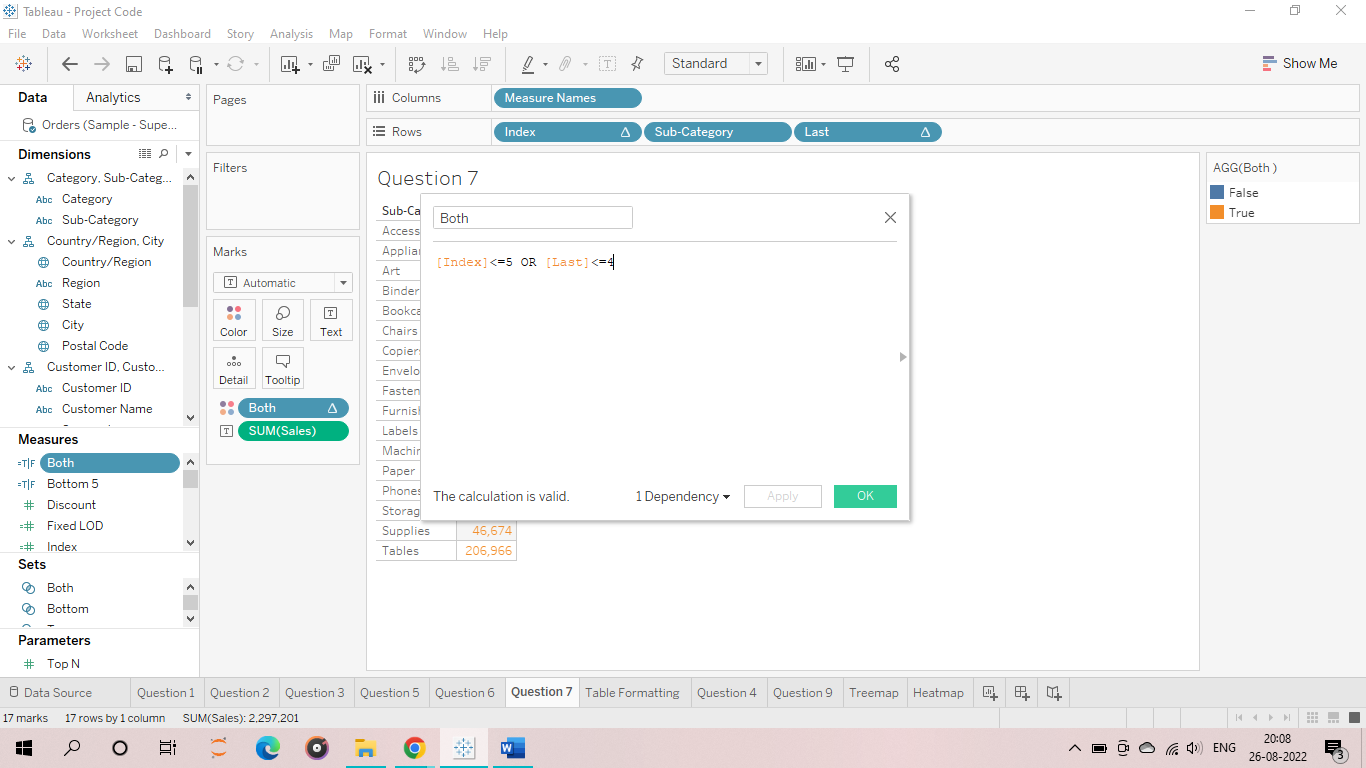
**6. Using Pareto chart in tableau.**

Creating a Pareto chart using sales by sub-category analysis.



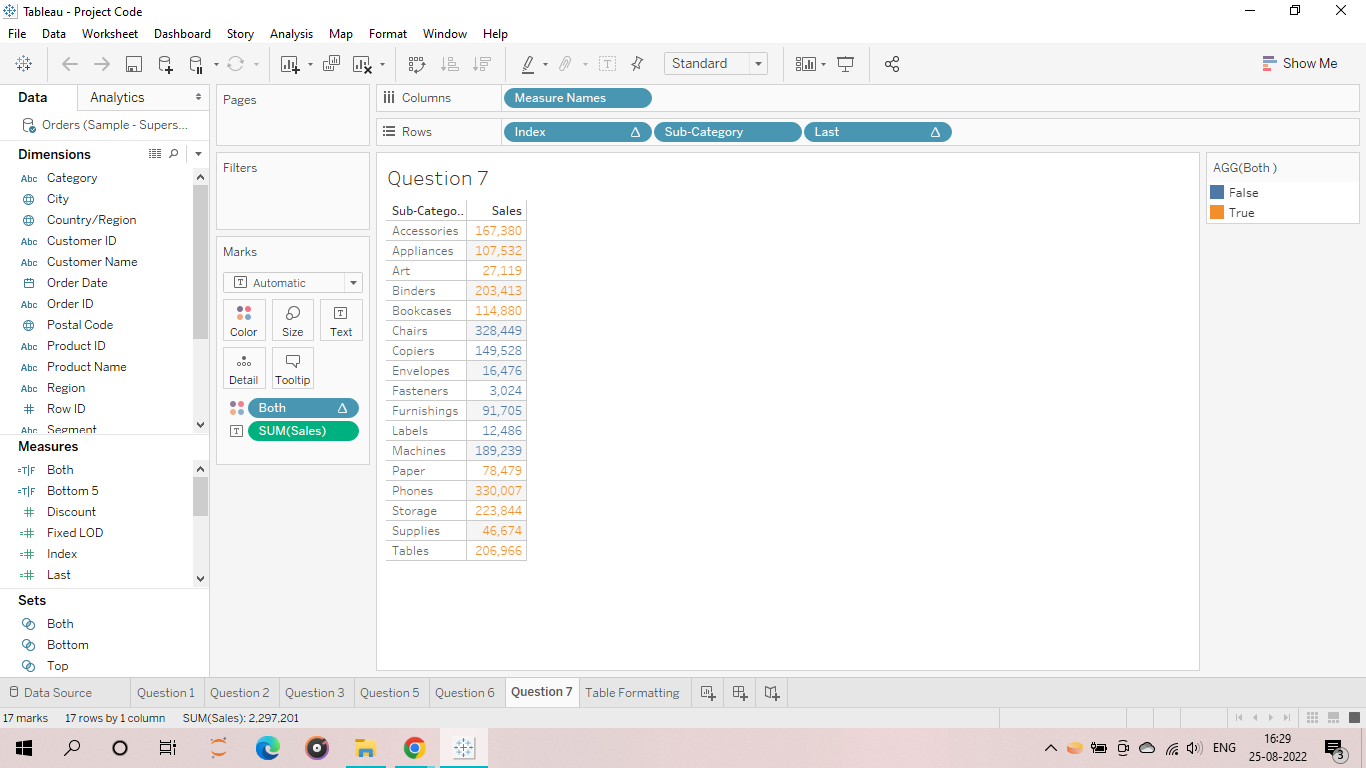
**7. Using conditional formatting in tableau.**

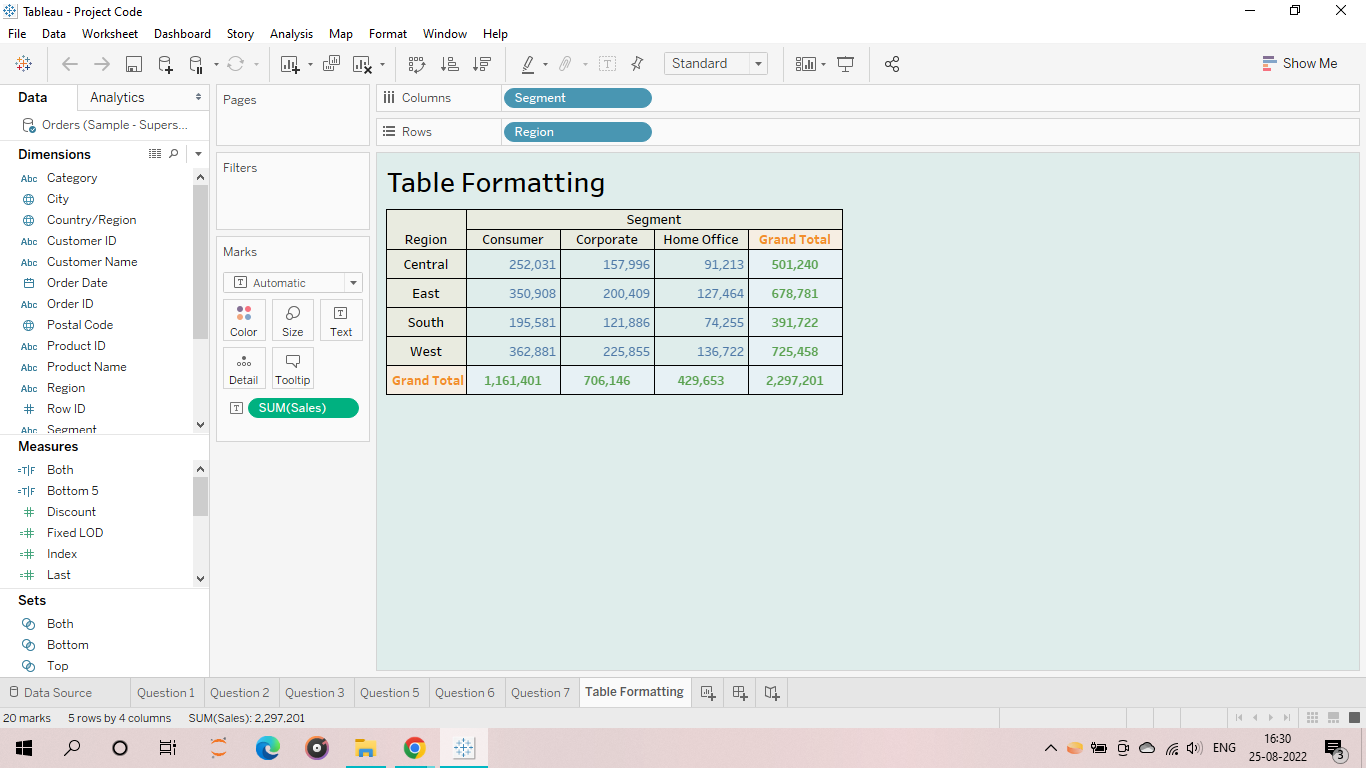
**CODE:**

****

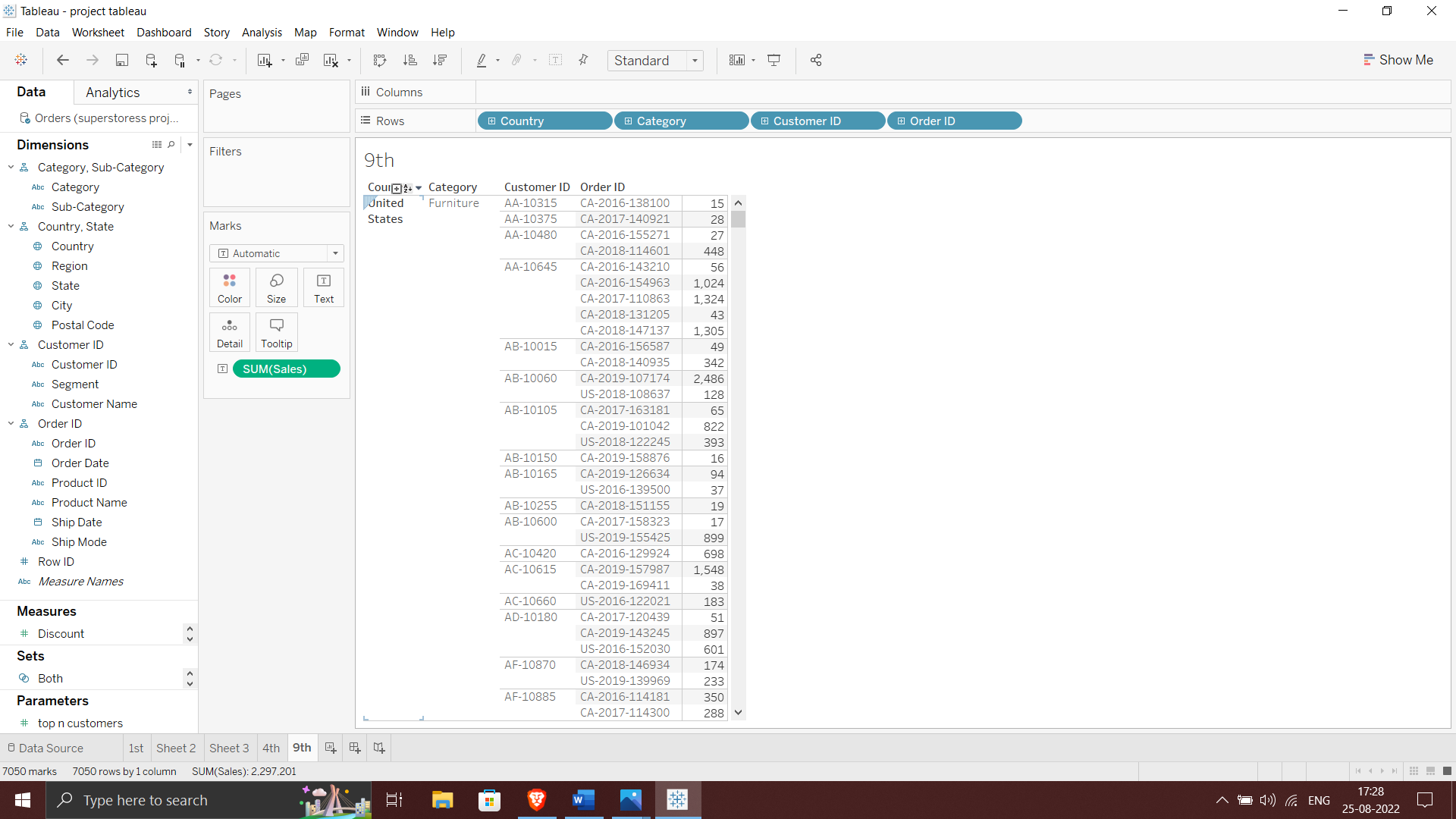
**OUTPUT:**

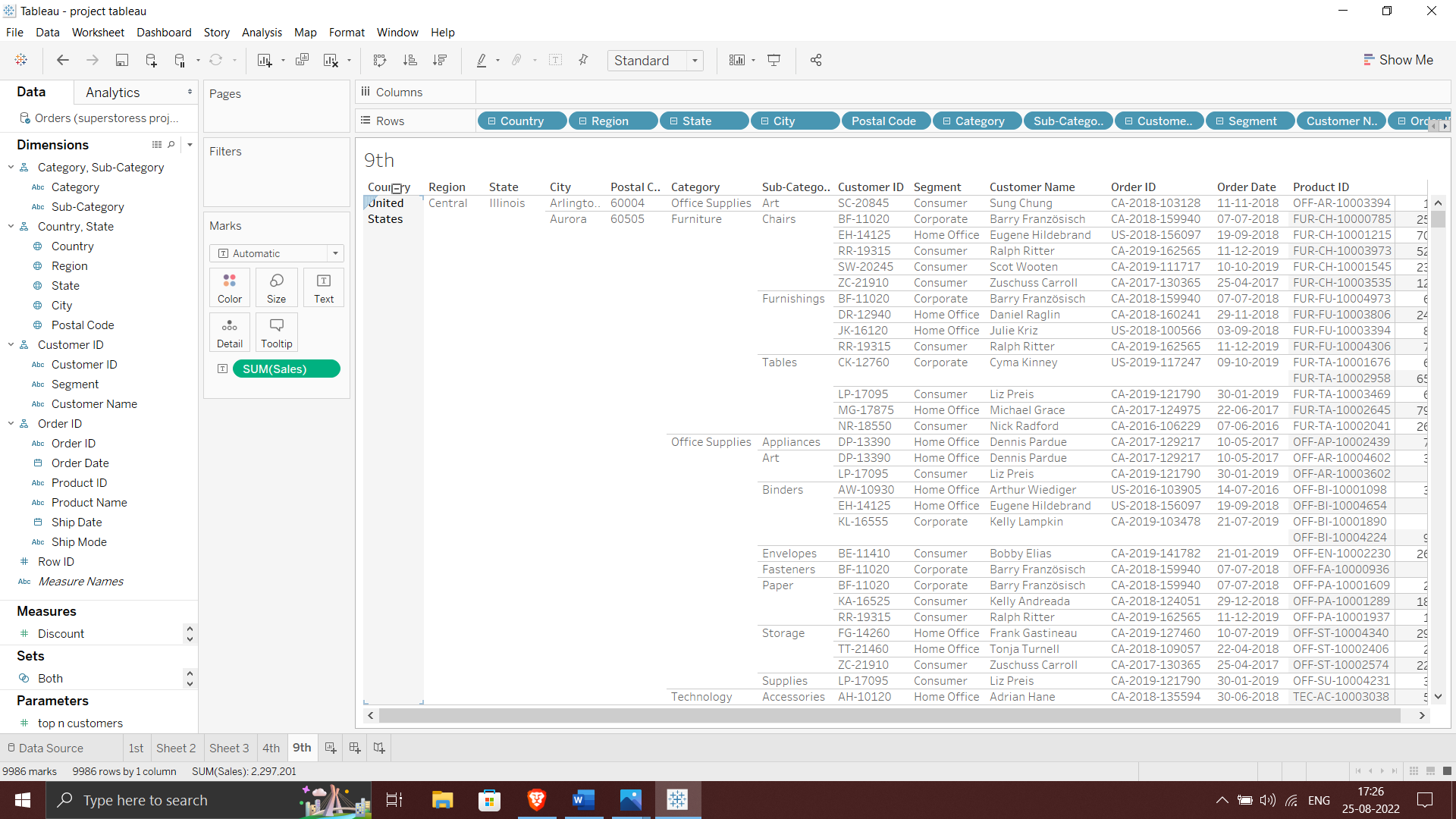
Highlighting Top 5 and Bottom 5 Sales by Categories using Conditional Formatting

****

**8. Using formatting in table in tableau.**

**9. Create hierarchy in tableau.**





**10.Difference between tree map and heat map in tableau.**

**Tree Map:**

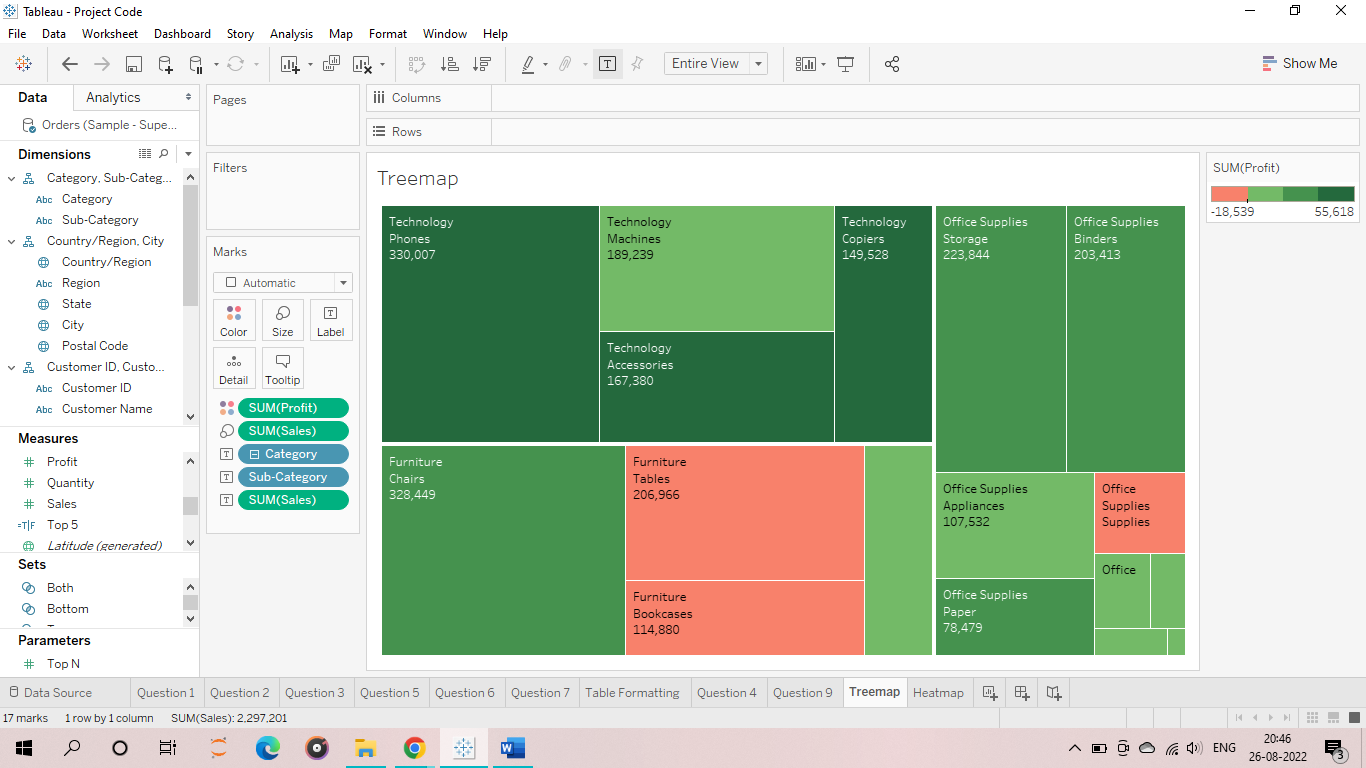
Tree maps are a relatively new feature in Tableau, first appearing in version 8.0. The ‘tree map’ is a chart type that displays hierarchical or part-to-whole relationships via rectangles. In case of hierarchical (tree-structured) data these rectangles are nested. The space in the view is divided into rectangles that are sized and ordered by a measure. Nested rectangles mean that hierarchy levels in the data are expressed by larger rectangles (above in the hierarchy) containing smaller ones (below in the hierarchy). The rectangles in the tree map range in size from the top left corner of the chart to the bottom right corner, with the largest rectangle positioned in the top left corner and the smallest rectangle in the bottom right corner.

In a tree map 1 or more dimensions & up to 2 measures are used to create such a map.

**Example:**

Below is a tree map that show the sales and profit in all regions for different products category and Sub-Category.

**Note:** The profit (color) and sales (size) of products are given at a Category and Sub- Category level. Bigger the size of the node, greater is the sales in that state. Similarly, the greener the node, more is the profit in that state.



**Heat Map:**

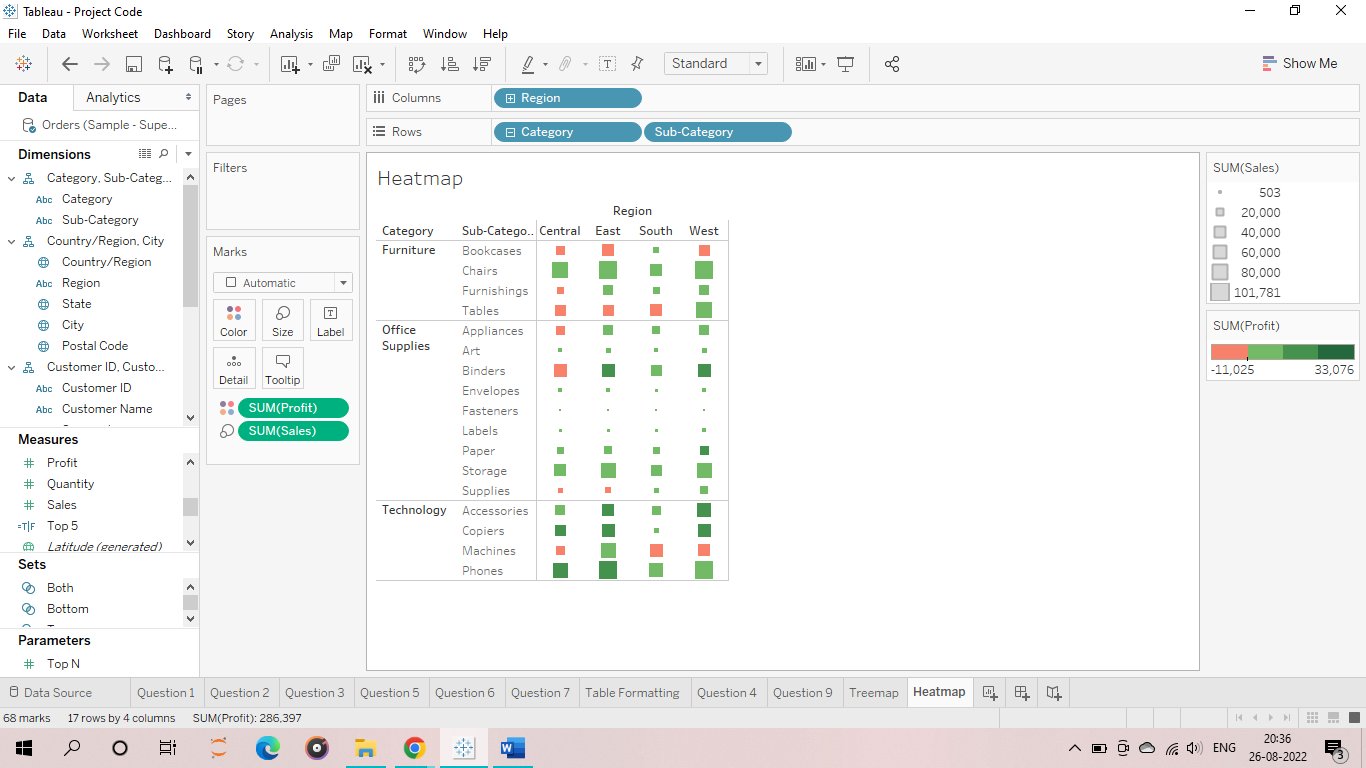
Heat map is a type of visualization tool that is very apt to compare different categories. It helps to visualize measures against dimensions with the help of colors and size to compare one or more dimensions & up to two measures. The layout is similar to a text table with variations in values encoded as colors. In heat map, you can quickly see a wide array of information.

In a heat map, one measure can be assigned to the color and another measure can be assigned to the size.

**Example:**

Below is a heatmap that shows the sales and profit in all regions for different products category and sub-Category.

**Note:** The profit is represented by the color and ranges from red for loss to green for profit. The total sales are represented by the size.

****

**Disadvantage:** A slight disadvantage of using tree maps in Tableau is that, as the number of items increase, the amount of space allocated for each item decreases. Hence, the area available to print the labels, become small.  As a result, usually, in a tree map almost all the squares or nodes will appear blank. This defect can be overcome by providing appropriate tooltips for each node. Like in heat maps, measures can be assigned to give different colors and sizes to the nodes in the tree map.

**11.Difference between blending and joining in tableau.**

**Data Blending:**

Data Blending sounds similar to Join but it is not the same. Let us look at its usage:

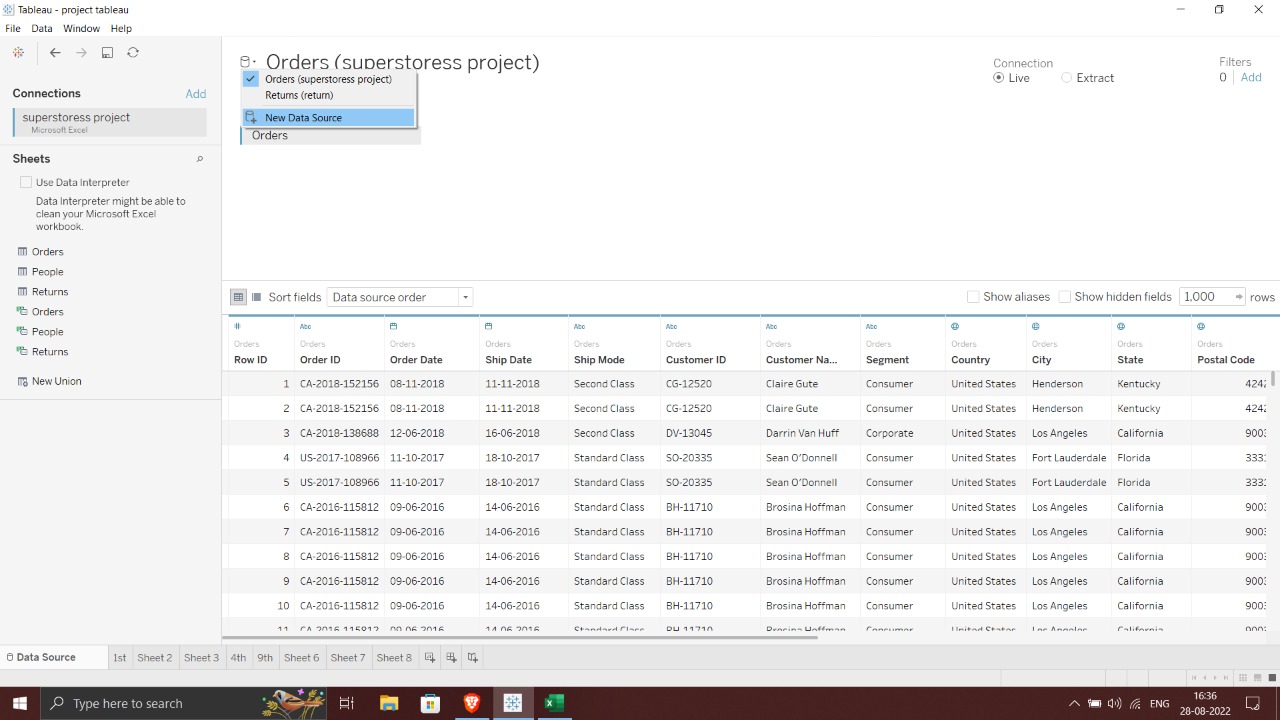
1.  It should be used when you want to analyse data available in different sources.

* For example, Actual Sales for geography like Region Name can come from one source (relational) and Month details from a different source (Excel File).
* In case you wish to visualize monthly sales for a region, you will be using Data Blending.

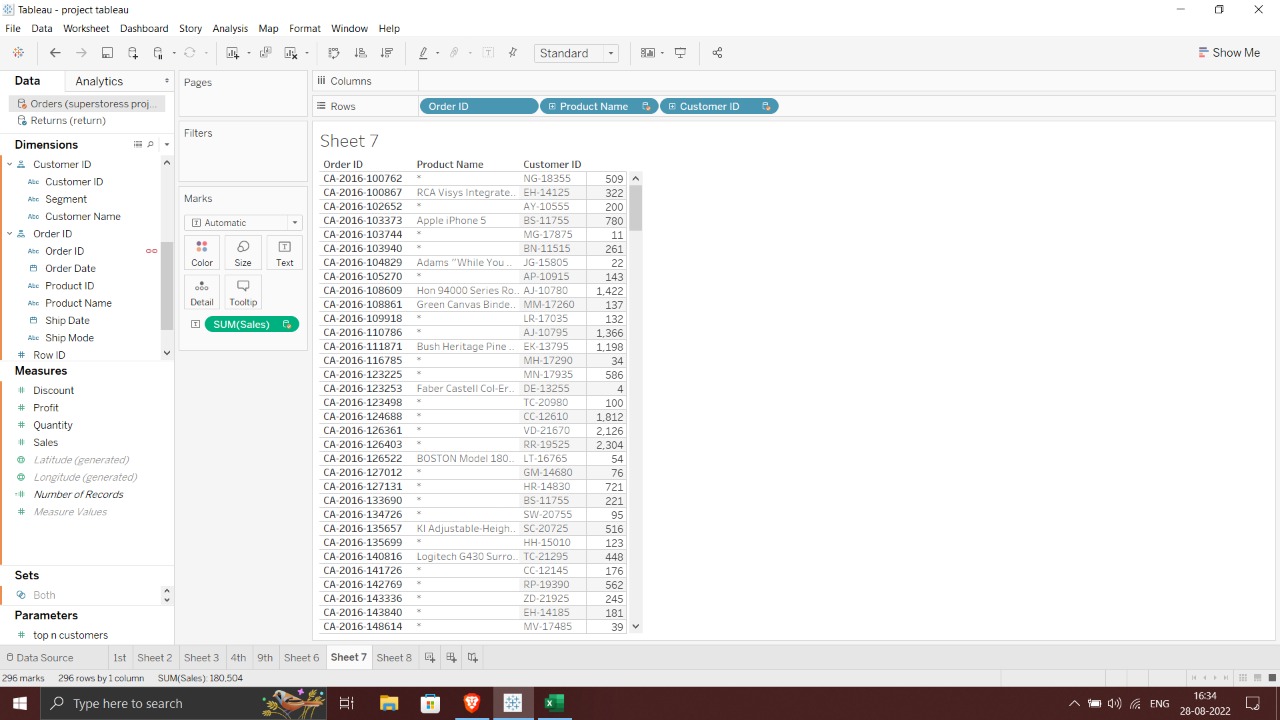
2. Ensure that there is a common field called “Linking field” while performing Data Blending.

3. Data Blending performs like a left join operation and does not accept or perform any other type of join.

4. Since data is coming from different sources, it is not mandatory that both data sets have the same level of detail. Data blending operation is independent of granularity.

****

****

****

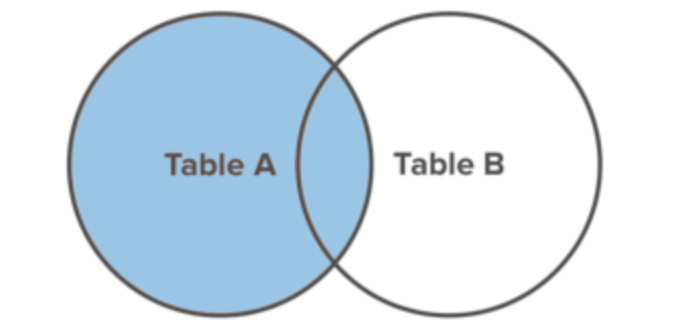
**Data Joining:**

It is primarily used when you have to merge data sets from the same source. For example, you want to look at Student details where his grades are coming from one source and personal information from another source. Here, the key will be the student roll number which will be unique.

There are different types of joins:

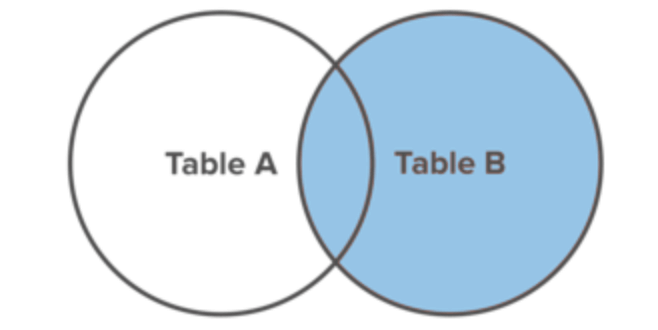
**1. Left join**

Merging contents between 2 tables and also retaining left table details.



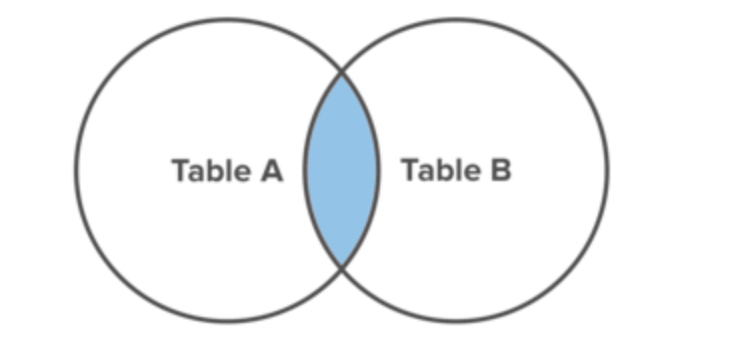
**2. Right Join**

Merging contents between 2 tables and also retaining right table details.



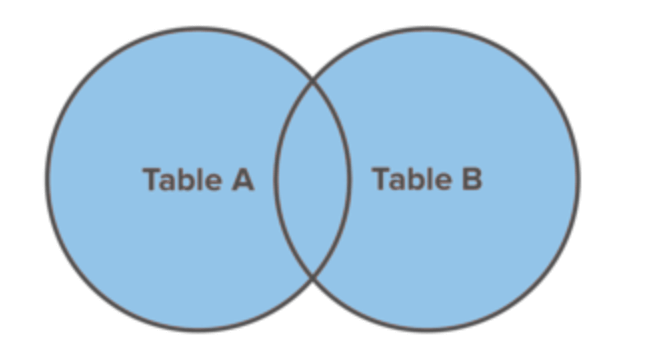
**3. Inner join**

Only the common matching data between 2 tables are displayed.



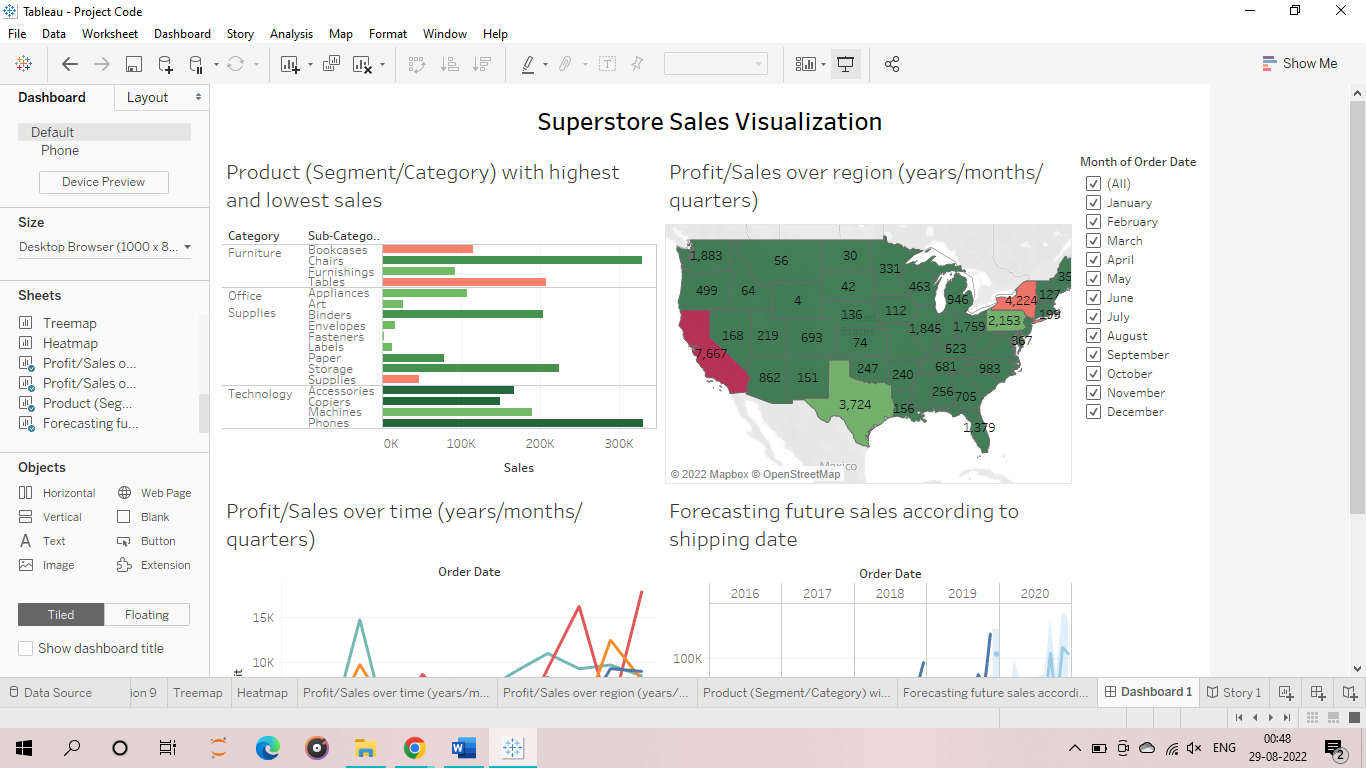
**4. Full Outer Join**

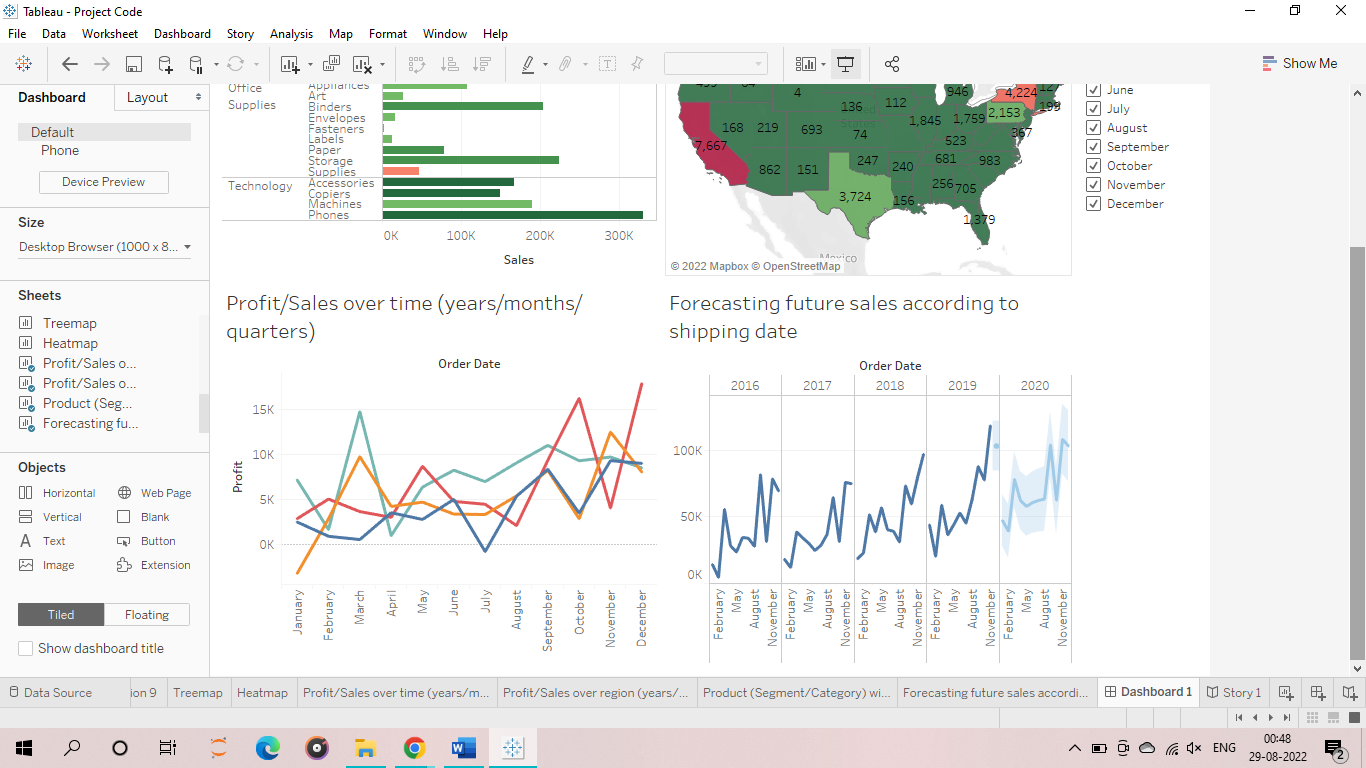
Result set from both the tables are merged and displayed.



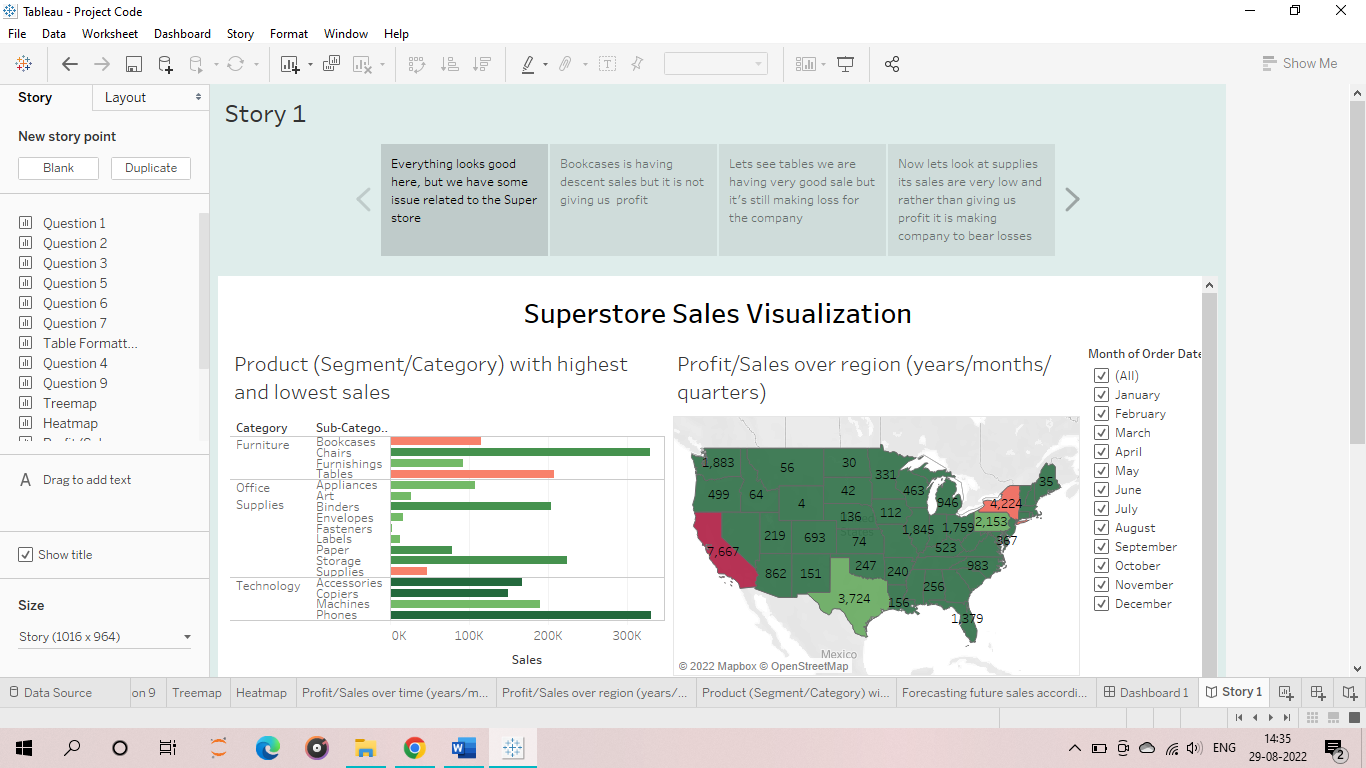
**12.Create a Dashboard and Story Point for all possible dataset**

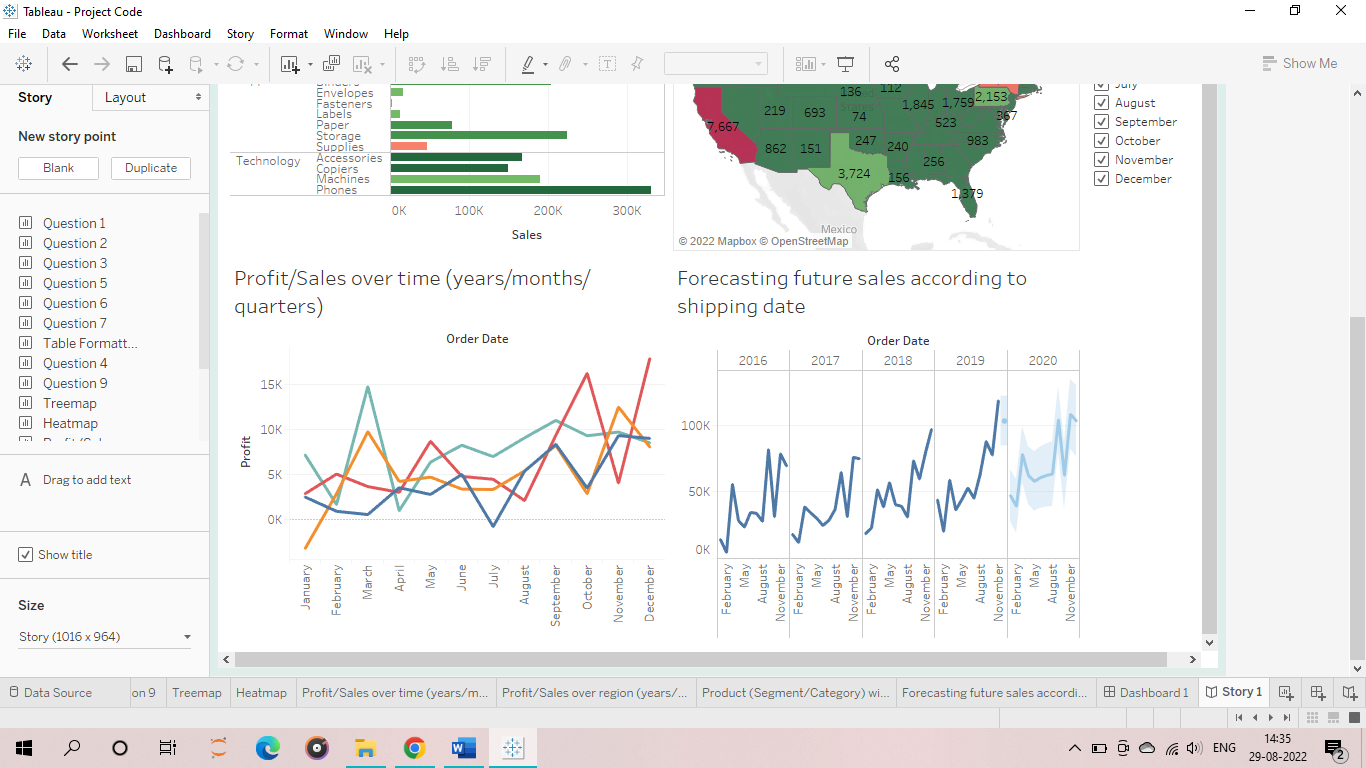
**Dashboard**

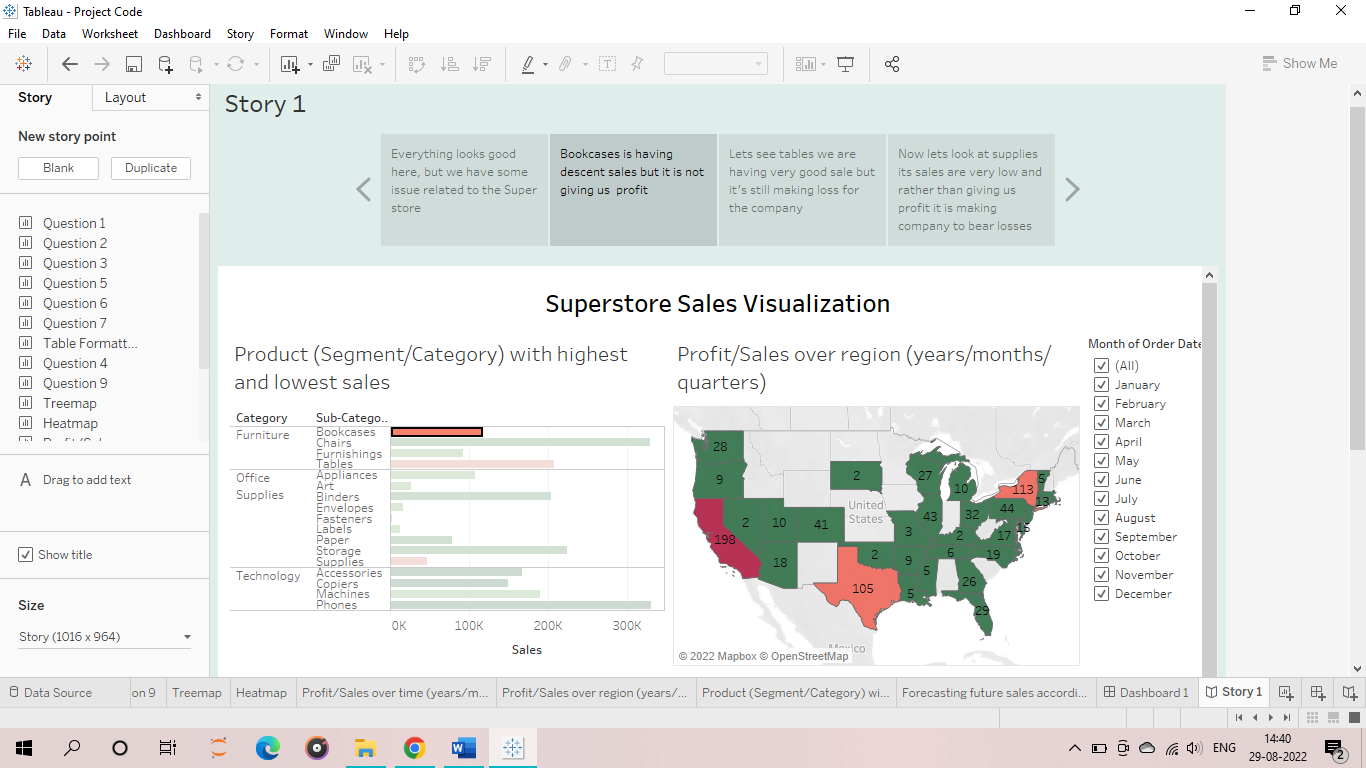
****

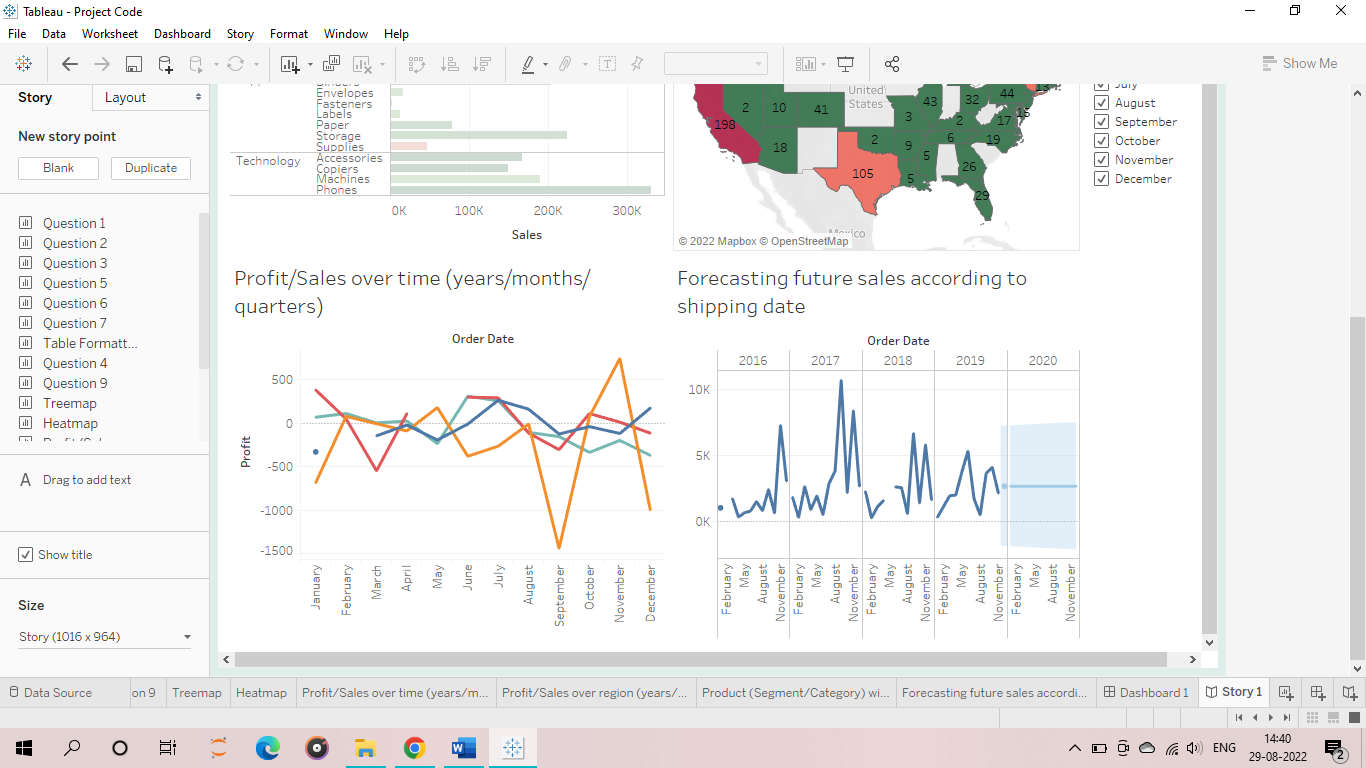
****

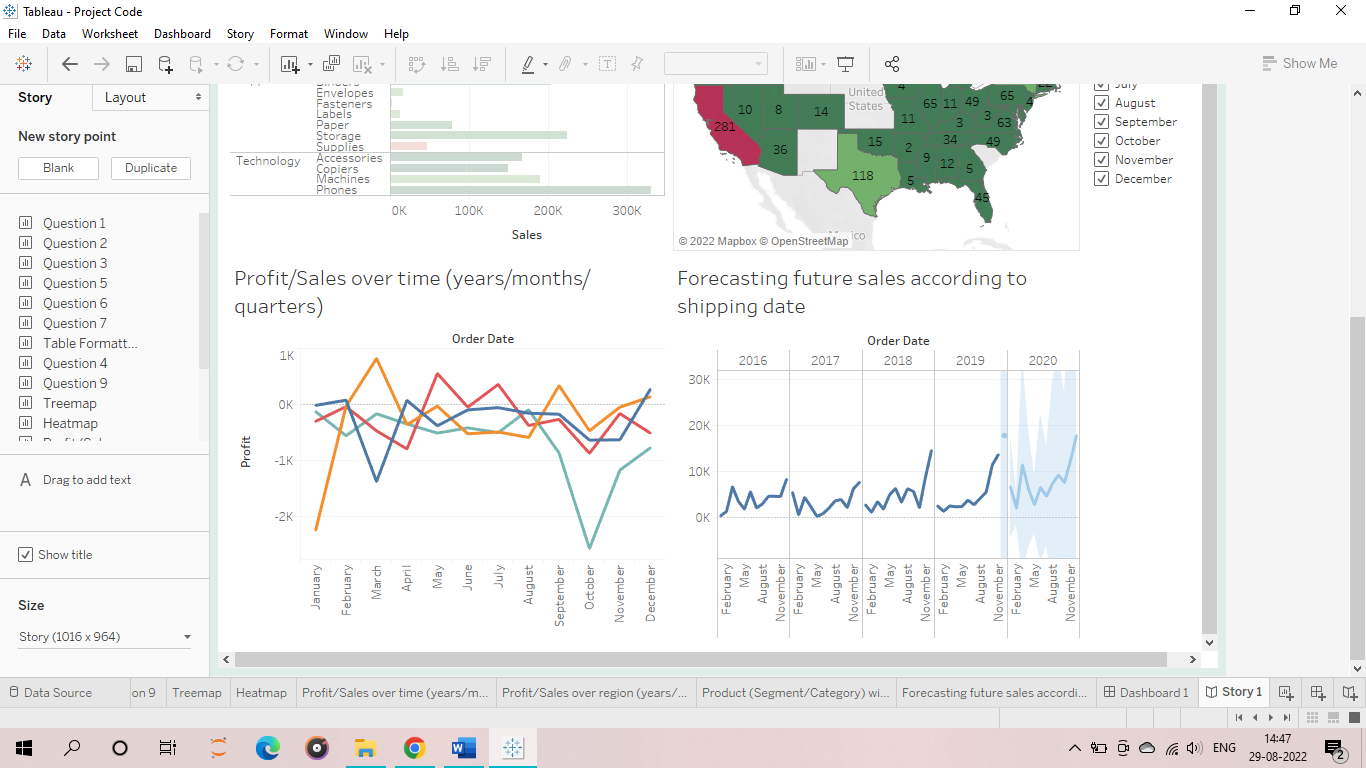
**Story Point**

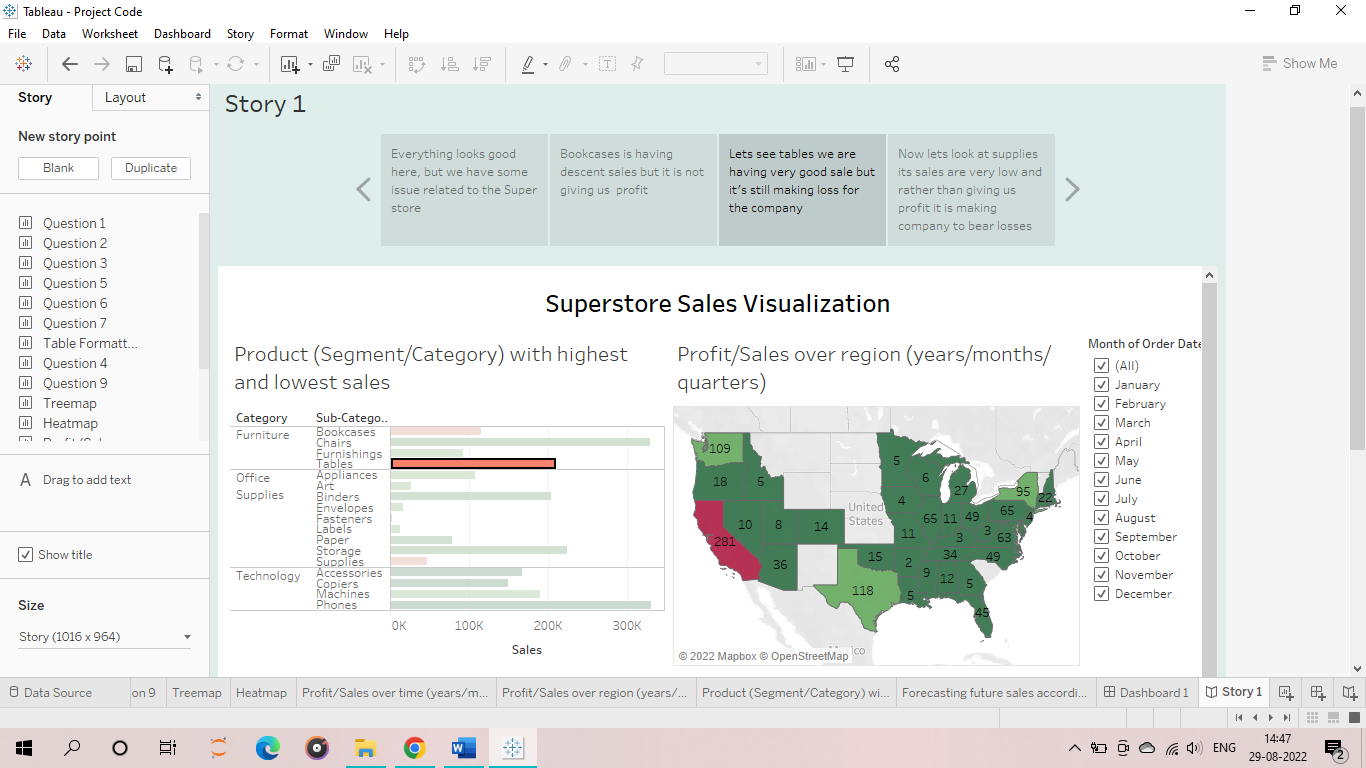


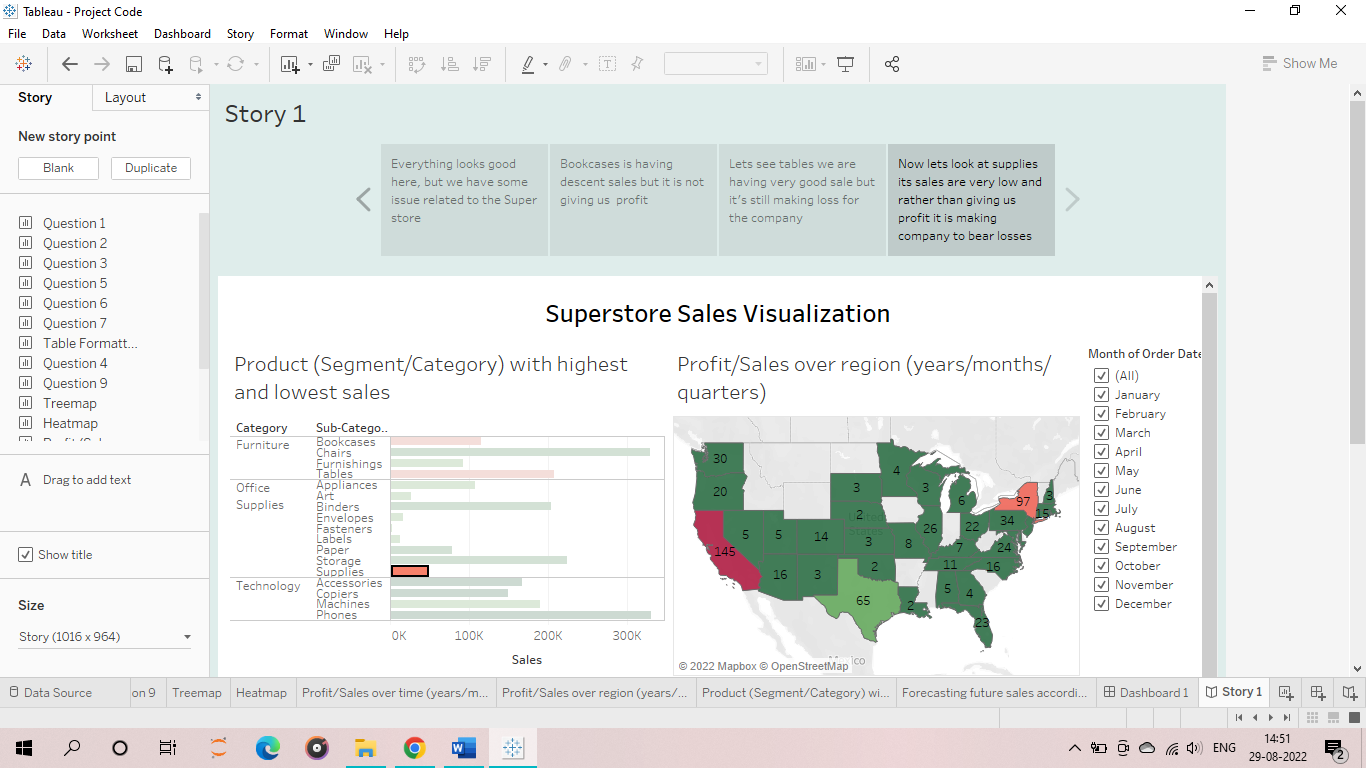


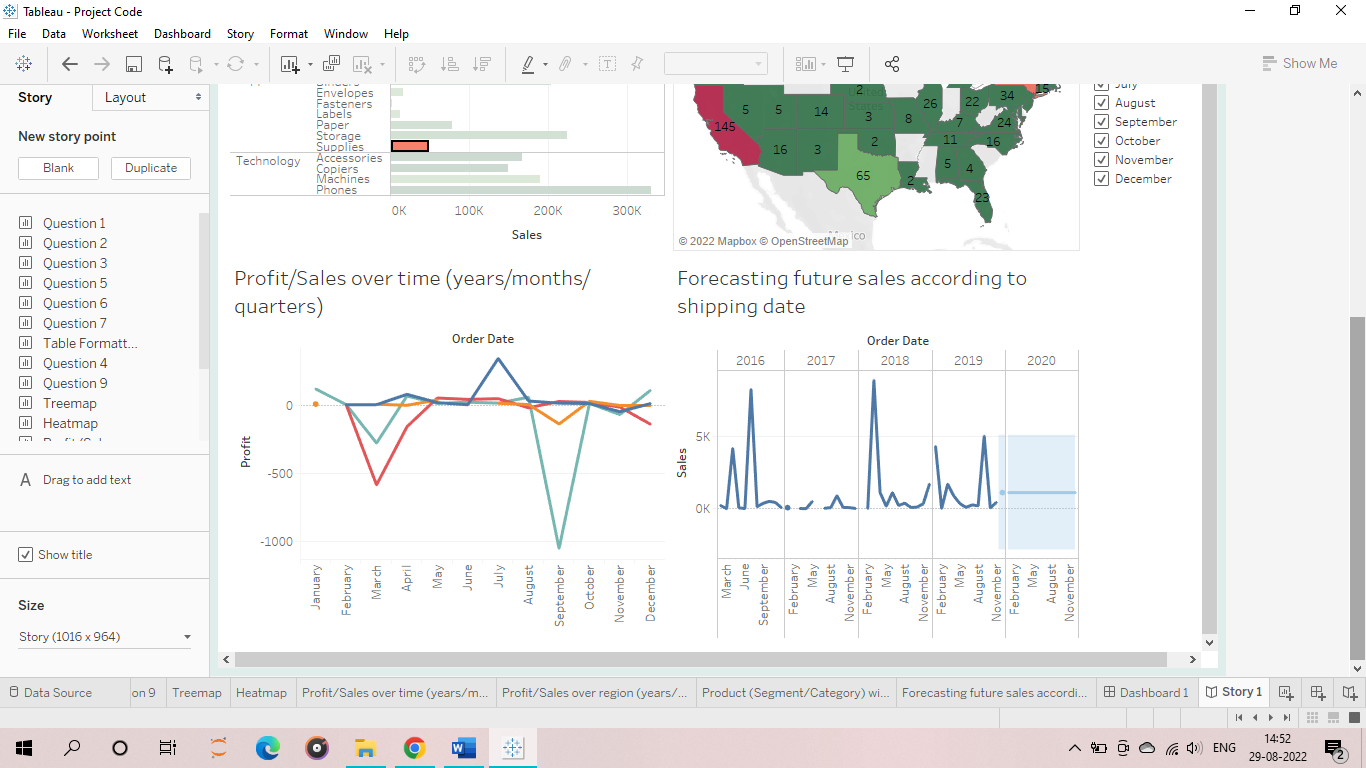












**Black Box Testing**

Black Box Testing is a software testing method in which the functionalities of software applications are tested without having knowledge of internal code structure, implementation details and internal paths. Black Box Testing mainly focuses on input and output of software applications and it is entirely based on software requirements and specifications. It is also known as Behavioural Testing.



The above Black-Box can be any software system you want to test. For Example, an operating system like Windows, a website like Google, a database like Oracle or even your own custom application. Under Black Box Testing, you can test these applications by just focusing on the inputs and outputs without knowing their internal code implementation.

## How to do Black Box Testing

Here are the generic steps followed to carry out any type of Black Box Testing.

* Initially, the requirements and specifications of the system are examined.
* Tester chooses valid inputs (positive test scenario) to check whether SUT processes them correctly. Also, some invalid inputs (negative test scenario) are chosen to verify that the SUT is able to detect them.
* Tester determines expected outputs for all those inputs.
* Software tester constructs test cases with the selected inputs.
* The test cases are executed.
* Software tester compares the actual outputs with the expected outputs.
* Defects if any are fixed and re-tested.

## Types of Black Box Testing

There are many types of Black Box Testing but the following are the prominent ones –

* **Functional testing** – This black box testing type is related to the functional requirements of a system; it is done by software testers.
* **Non-functional testing**– This type of black box testing is not related to testing of specific functionality, but non-functional requirements such as performance, scalability, usability.
* **Regression testing**– Regression testing is done after code fixes, upgrades or any other system maintenance to check the new code has not affected the existing code.

## Tools used for Black Box Testing:

Tools used for Black box testing largely depends on the type of black box testing you are doing.

* For Functional/ Regression Tests you can use – QTP, Selenium
* For Non-Functional Tests, you can use – LoadRunner, JMeter

## Black Box Testing Techniques

Following are the prominent Test Strategy amongst the many used in Black box Testing

* **Equivalence Class Testing:** It is used to minimize the number of possible test cases to an optimum level while maintains reasonable test coverage.
* **Boundary Value Testing:** Boundary value testing is focused on the values at boundaries. This technique determines whether a certain range of values are acceptable by the system or not. It is very useful in reducing the number of test cases. It is most suitable for the systems where an input is within certain ranges.

**Decision Table Testing**: A decision table puts causes and their effects in a matrix. There is a unique combination in each column.

Future Scope

The future scope for our project is to analyse big data around the world and representing it in simpler form and forecasting so that the company, organization and individual can take best decision in their Interest. Our project can radically change the present business situation by capturing a large volume of data, expand business model, energize the imaginative procedures and overall growth and development of a company.

Bibliography/ References

[**https://www.youtube.com/**](https://www.youtube.com/)

[**https://www.tutorialsduniya.com/**](https://www.tutorialsduniya.com/)

[**https://www.slideshare.net/**](https://www.slideshare.net/)

[**https://github.com/**](https://github.com/)

[**https://www.simplilearn.com/**](https://www.simplilearn.com/)