**E-COMMERCE SALES DATA ANALYSIS**

**INTRODUCTION**

**Data Analysis** is a process of inspecting, cleansing, transforming, and modeling data with the goal of discovering useful information, informing conclusions, and supporting decision-making. Data analysis has multiple facets and approaches, encompassing diverse techniques under a variety of names while being used in different business, science, and social science domains.

The Analytics team of a Super Store anywhere in the world would want to design a Sales and Performance dashboard to analyze the sales based on various product categories and other factors which have a role to play in the running of the store. The store managing head or the owner wants to add user control for product categories, so users can select a category and can see the trend month-wise and product-wise accordingly.

The Analytics team would also want to analyze various other things like how many days the store takes to ship the product, how many times a customer orders a product, how much time is there between the first and second order of the customer etc.

The Super Store’s database keeps track of the following data fields:

* Row ID- Sequence Number for Each Row
* Order ID- Id of The Order Created by The Customer
* Year- Year Wise Customer to Order the Product
* Order Date & Ship Date- The Date When the Item Was Ordered and The Date When the Item Was Shipped to The Person.
* Ship Mode-Mode of Shipping
* Customer ID-– Unique ID For Each Customer
* Customer Name-Name of The Person Who Ordered the Item.
* Segment- Talks About Customer Type
* Country- Country of Customer
* City- City of Customers
* State-State of Customers
* Postal Code- Postal Code of The City
* Region - Region Where the Sales Were Made.
* Product ID- Unique Id for Each Product
* Category- Category of Product
* Sub-Category -A Sub-Category of The Product
* Product Name- Unique Product Name
* Sales-Quantity of The Items Ordered
* Quantity- Order of Product in Quantity
* Discount-Discount Value on The Product
* Profit-How Much Profit Was Earned on That Product

**SCOPE OF ANALYSIS**

The super store wants to see and analyze the sales trend month-wise and product-wise and work upon the lagging segments and outperforming employees accordingly. The Analytics team also wants to create analyze the database in depth to help the super store grow exponentially. The Analytics team wishes to answer the following objectives: -

1. Sales, Quantity, and Profit of each product category
2. Segment Distribution of each product category
3. Sales and Profit throughout months of sales of each product category
4. Regional Sales of each product category
5. Overall Sales Trend throughout months of a sales year
6. Customer Ordering Trend
7. Comparison of sales and profit product category wise

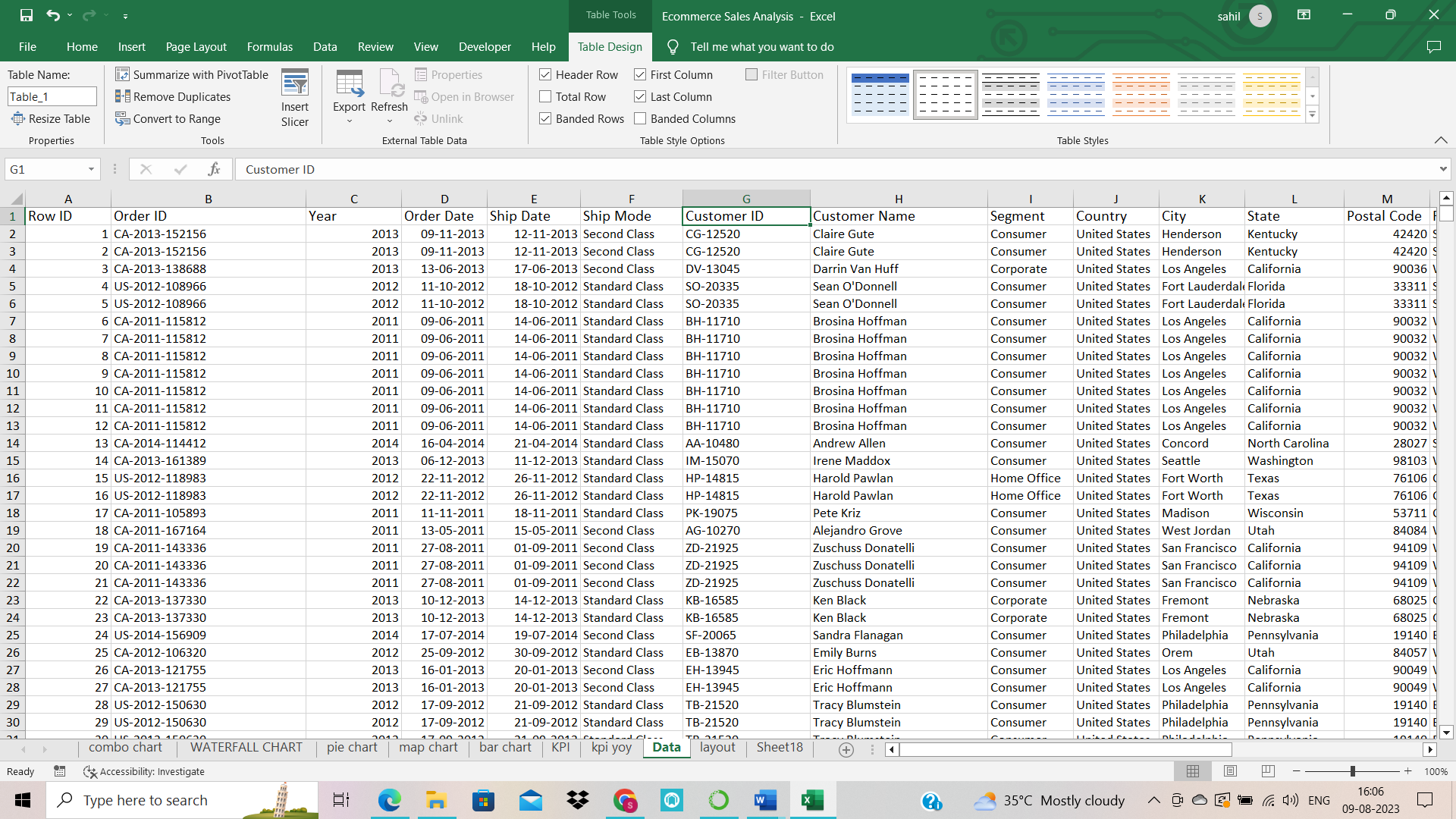
Aim of this project is to answer the above objectives in the form of visualization by creating a dashboard to convey the answers effectively and efficiently.

**ETL PROCESS**

Extract, Transform, Load (ETL) is a procedure used in databases, particularly in data warehousing, to prepare data for analysis. Data transformation processes data by converting them into an appropriate storage format/structure for the purposes of querying and analysis; data loading describes the insertion of data into the final target database, such as an operational data store, a data mart, or a data warehouse. Data extraction involves extracting data from homogeneous or heterogeneous sources. A well-designed ETL system extracts data from the source systems, upholds standards for data quality and consistency, conforms data so that different sources can be combined, and then delivers data in a presentation-ready format so that application developers can create applications and end users can make decisions.

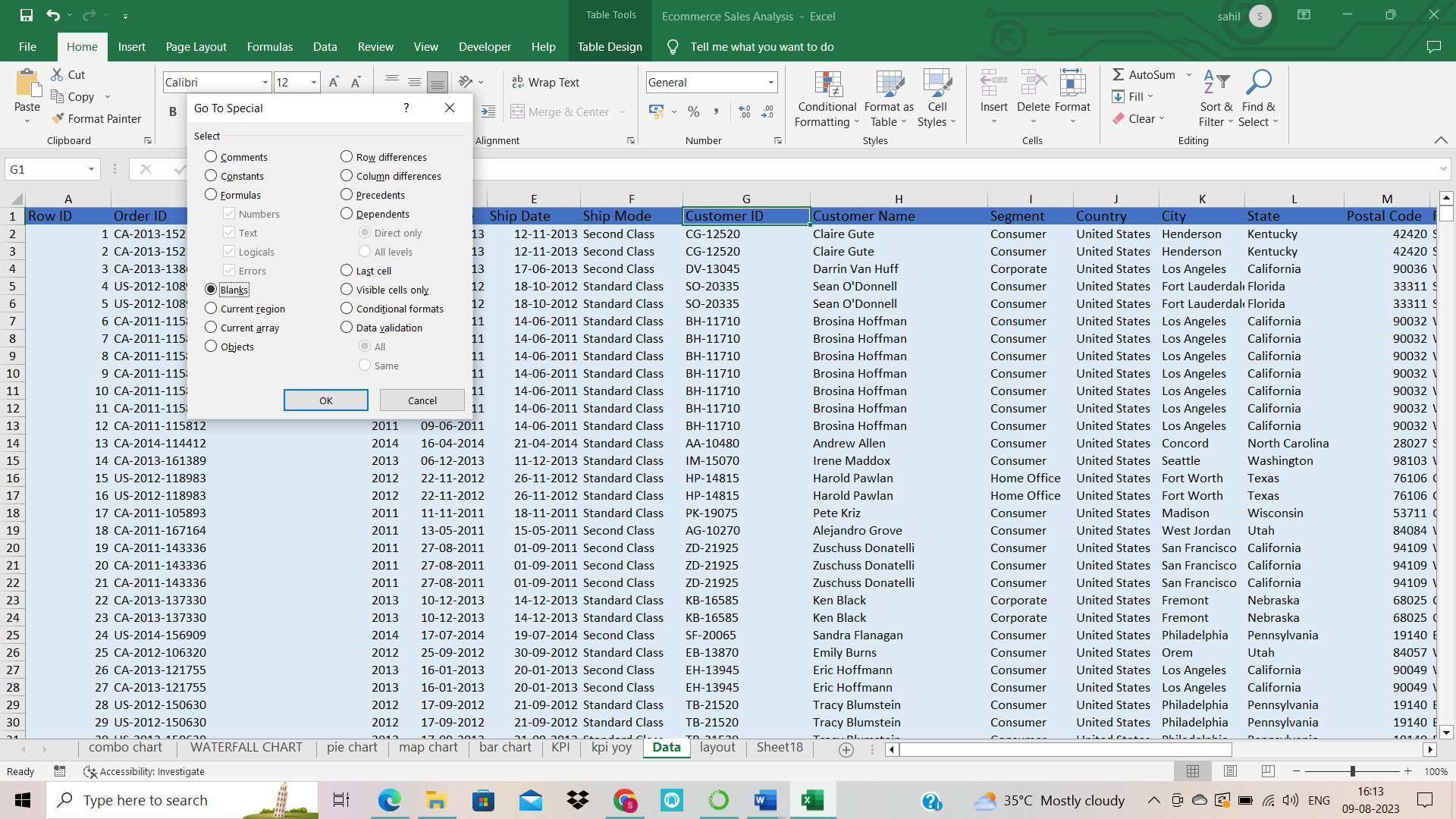
ETL is described specifically as a process that takes data from several RDBMS source systems, transforms the data (by performing computations, concatenating data, etc.), and then loads the data into the Data Warehouse system. Extract, Transform, and Load is referred to as ETL.

The dataset looked like this before ETL. This information came from Kaggle.



**Step 1: Remove the blank cells from the dataset.**

For this, select the whole dataset. Select the whole dataset for this. Select Find and Select from Excel's Home tab. Pick Blank from the drop-down menu after choosing Go to Special. The selection will include every empty cell. Return to the home tab's Delete option and choose Delete Rows from the drop-down menu. Rows containing empty cells will be eliminated as a result.



**Step 2: Removing columns which are not properly defined or not crucial to our analysis.**

We will use redundant columns for this, such as the column that only contains index numbers. To achieve this, we will first choose that specific column, then click the "Delete" button under the "Home" tag and choose "Delete Columns" from the drop-down menu.

i.e. Delete row id column

**Step 3: Give proper and appropriate column names.**

Since the dataset lacks the necessary columns, our next step would be to give the columns the right column names where necessary.

**Step 4: Excluding the NULL values from the data.**

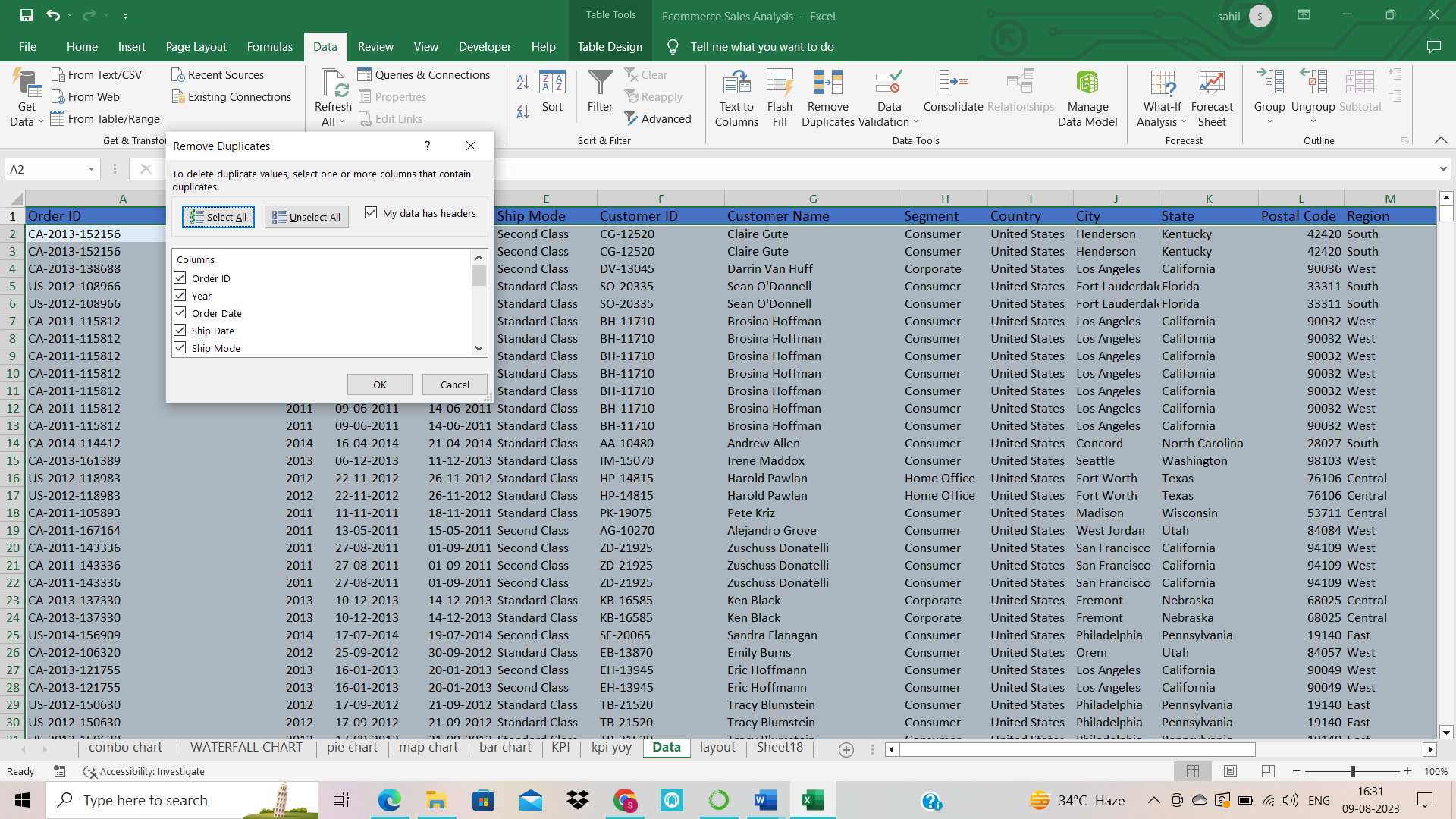
Use the filter for these and remove the null values. Go to sort & filter, then filter select and check the individual column null value.

**Step 5: Improvising Proper Data Formatting**

Analyses won't be done correctly without adequate data formatting. So, we'll style certain columns correctly by bringing them down. For better results, for instance, dates should be in date format, while prices and sales should be in currency format.

**Step 6: Removing Duplicate Values**

It's likely that our data contains duplicate values, which might make accurate analysis difficult. Therefore, deleting duplicate values and preparing our data for analysis will be our last ETL operation.

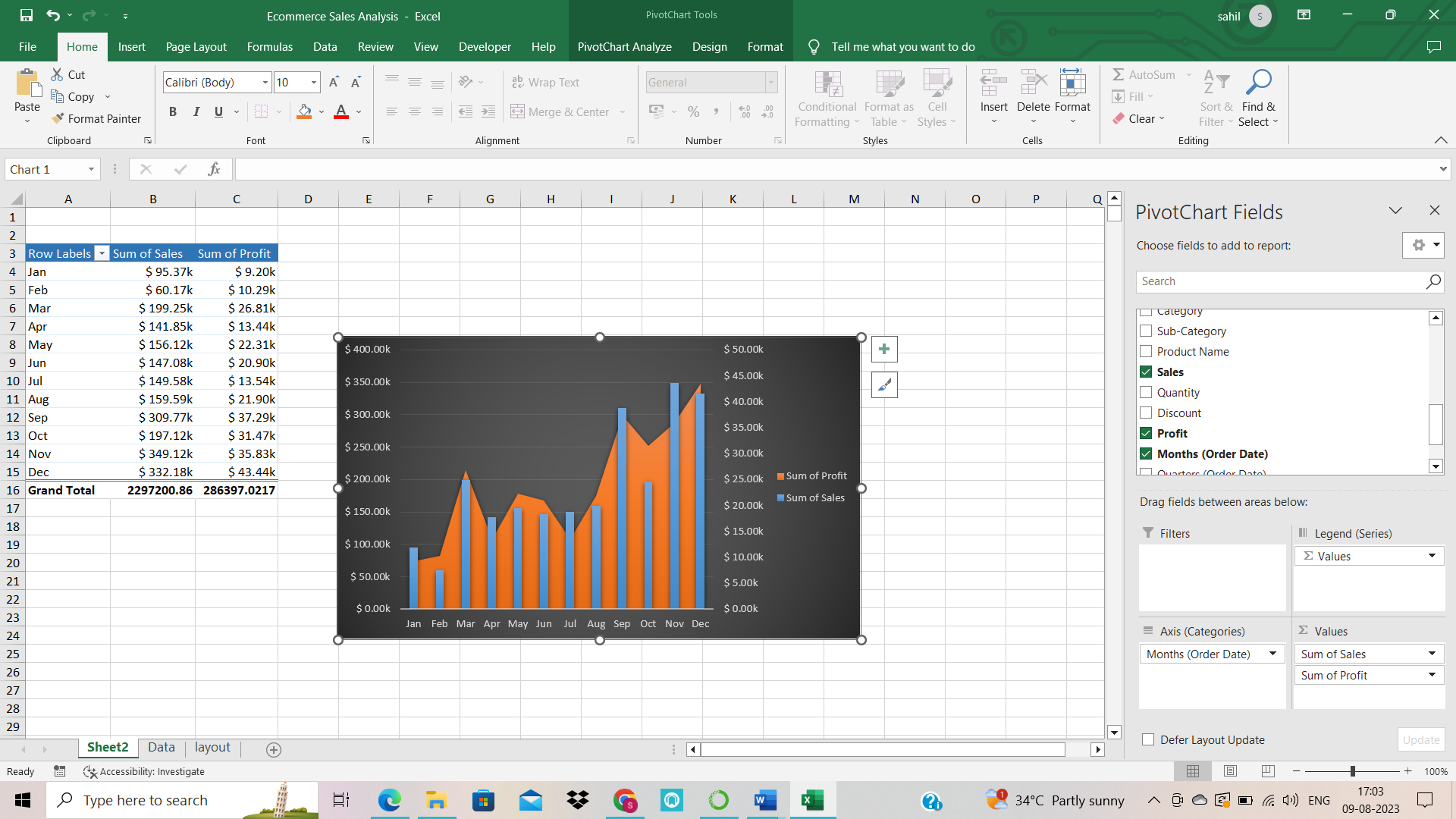


**ANALYSIS OF THE DATASET AND RESULTS**

1. **Monthly Sales and profit of each category** (Combo chart: sum of profit and sum of sales)

**Description:** By knowing about sales and profit over a month, we can know about the months which are more profitable for sales and hence customize our advertisement plan to increase sales even more. After finding out the sales and profit, we visualize the result with the help of a stacked bar graph.

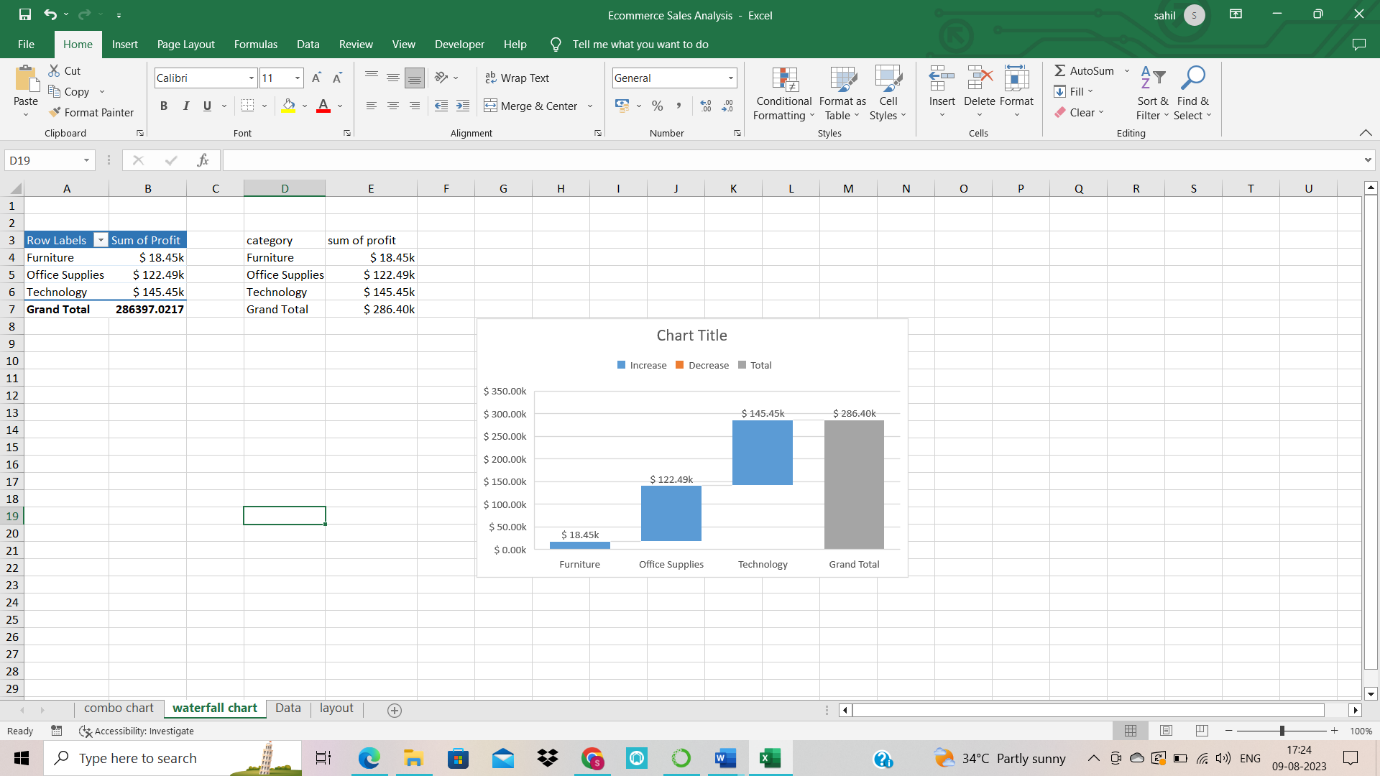
#Convert into dollar currency format and two decimal places.

**Result:**

1. **Category-wise profit**

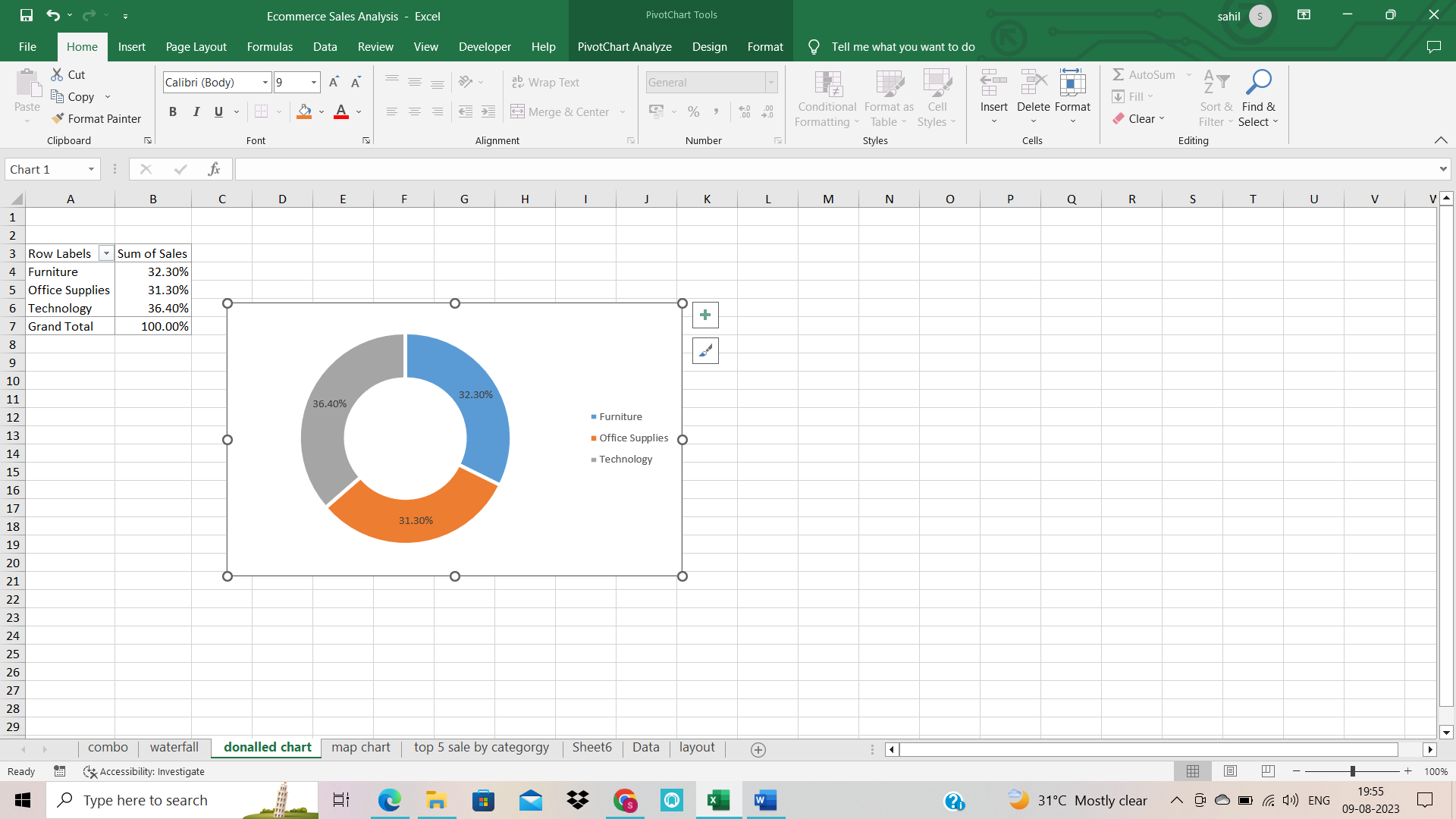
Category-wise profit like furniture, office supplies, and technology with their profit and converted into two decimal currency formats. Then generated the waterfall chart

**Result:**

****

1. **Category-wise sales share %**

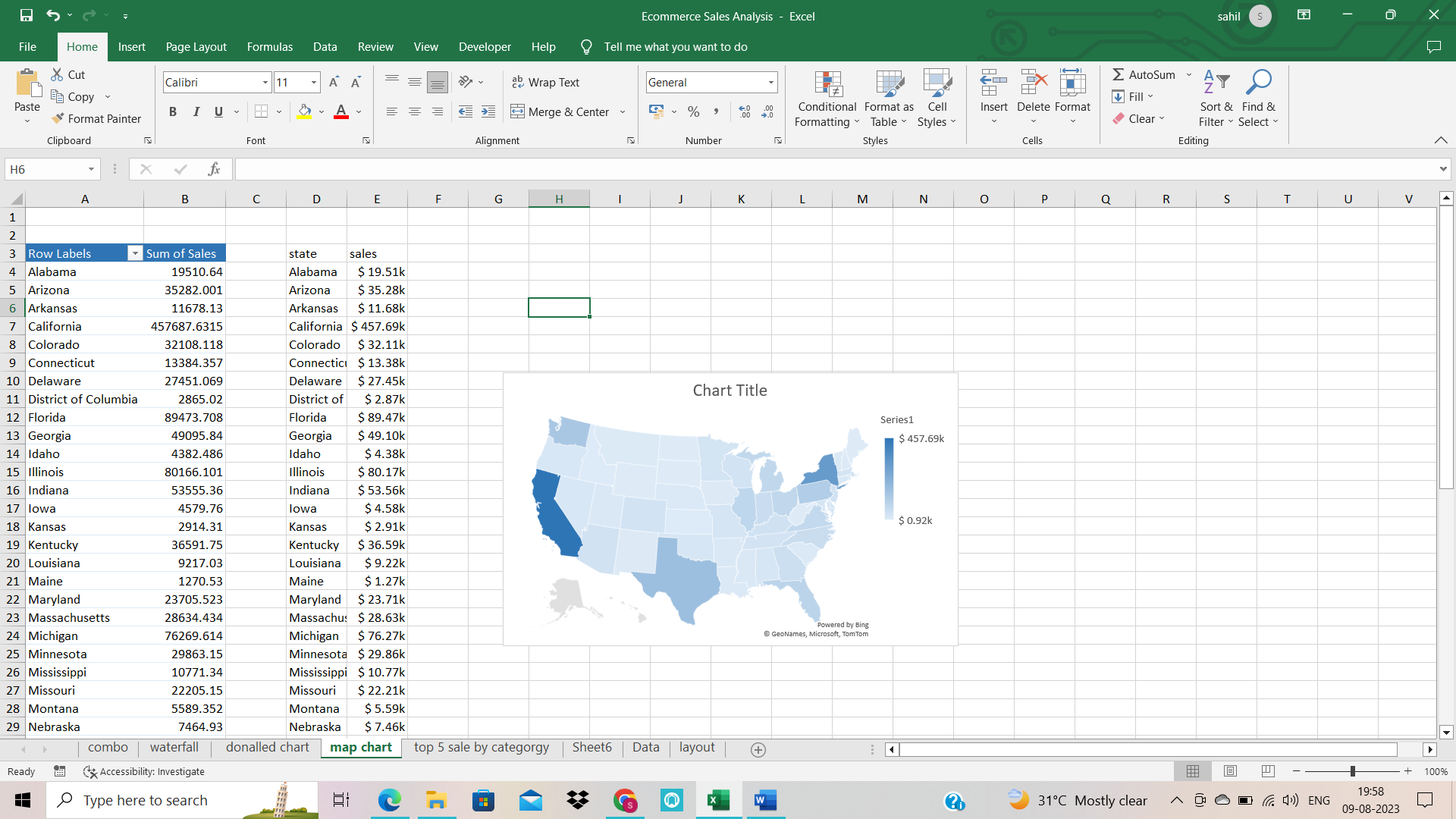
Category-wise share percentage of sales in Donald chart.



1. **Map chart for sate wise sales**

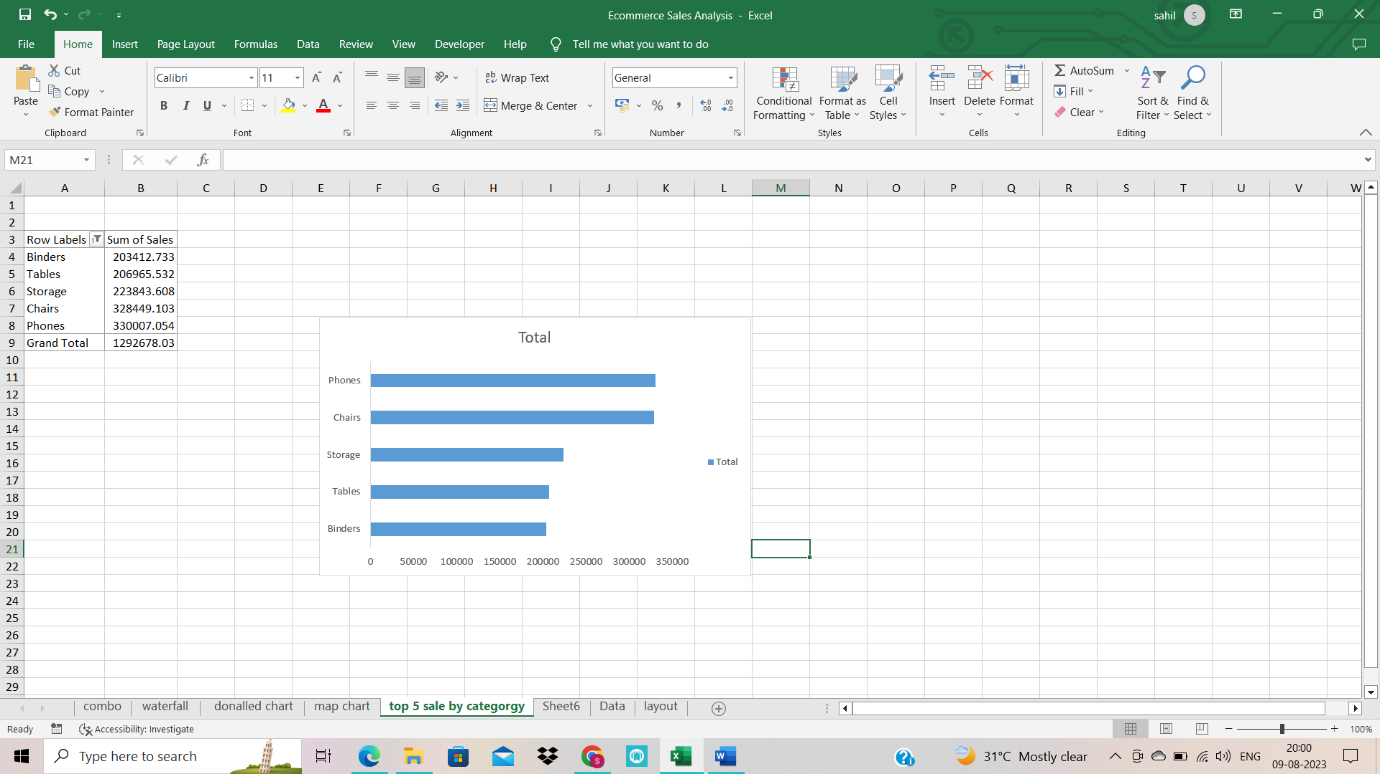
state wise sales in map chart

In this we analyze which particular region is having the most amount of sales and which is least. Furthermore, we can look upon the factors which might be impacting the sales and we can look upon them to increase the sales and invest in the areas of maximum sales.

****

1. **Top 5 sales by categories wise**

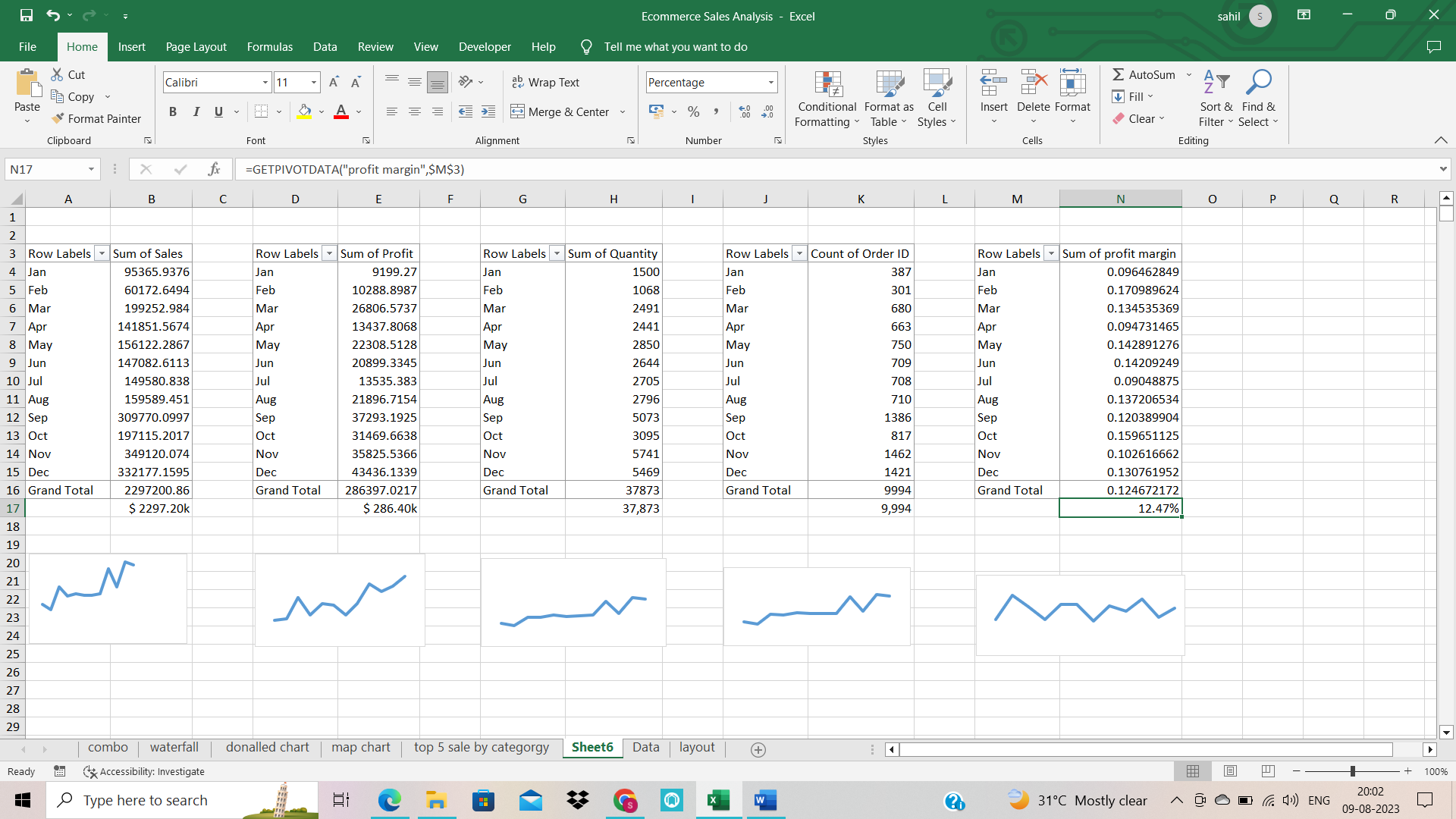
Top 5 products and their sales in bar chart

****

**6. KPI**

Total sales, total profits, total quantity, total order, profit margin

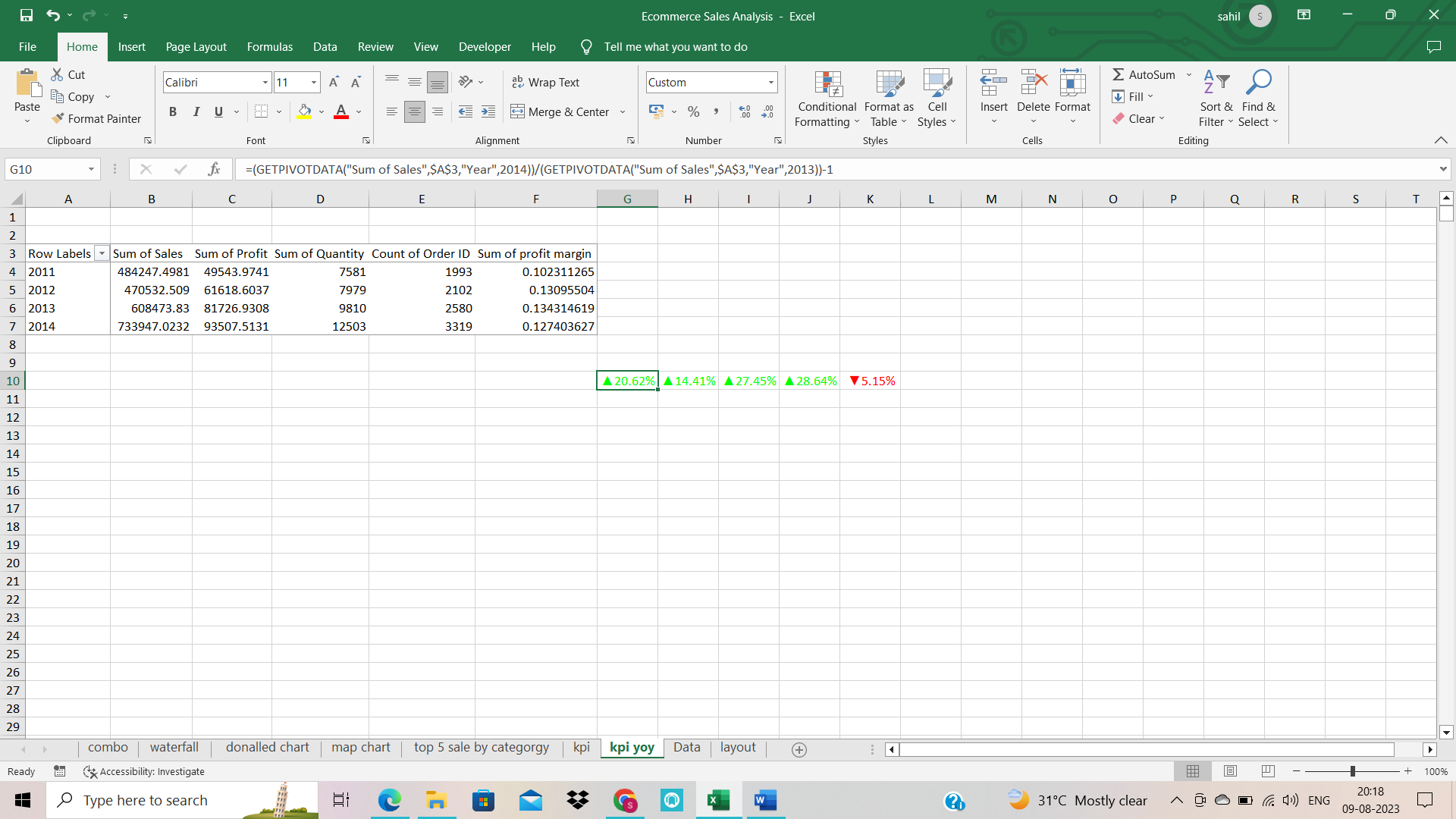
With sparkline



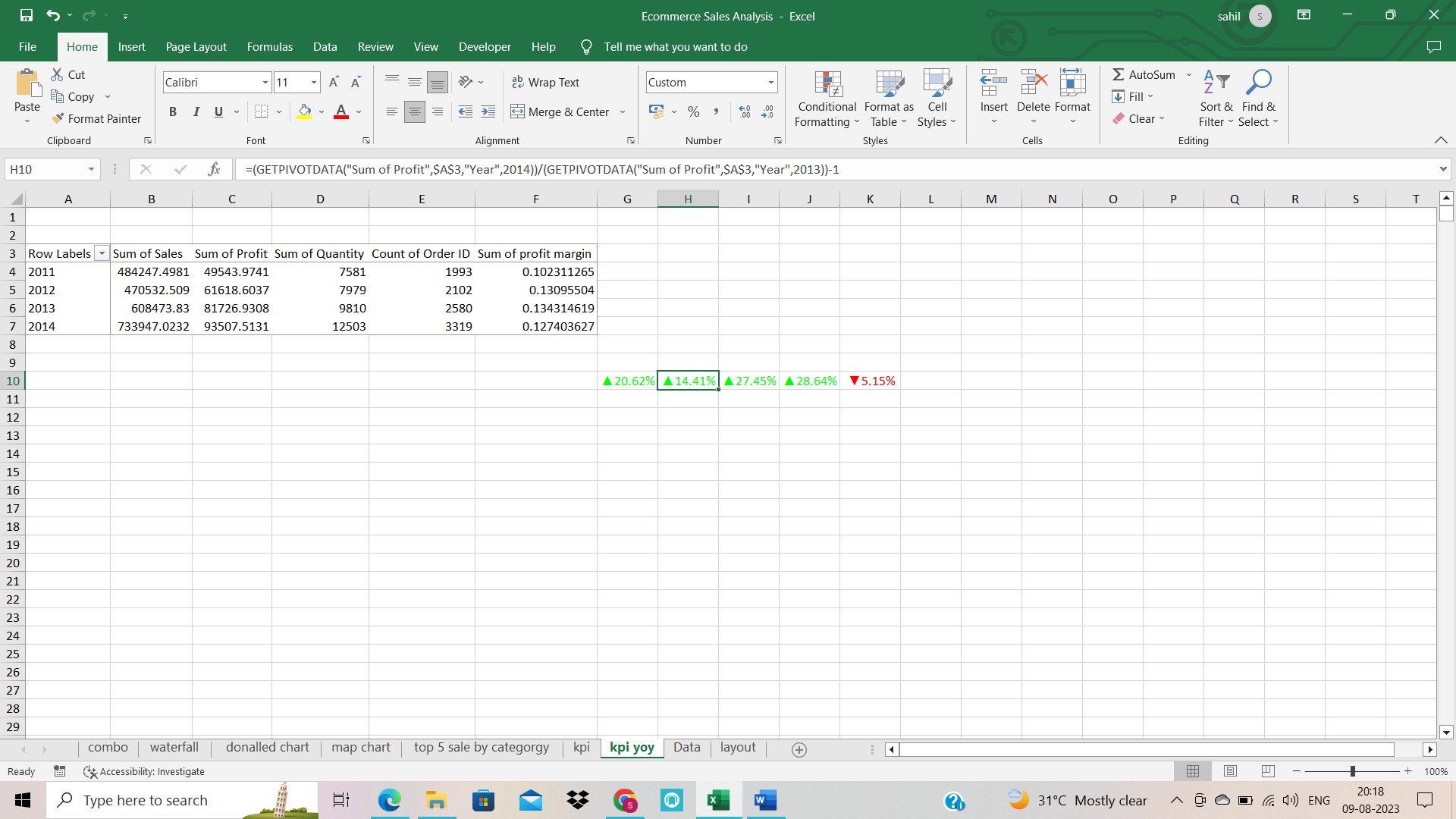
**7. KPI YOY**

**Sales YOY;** formula

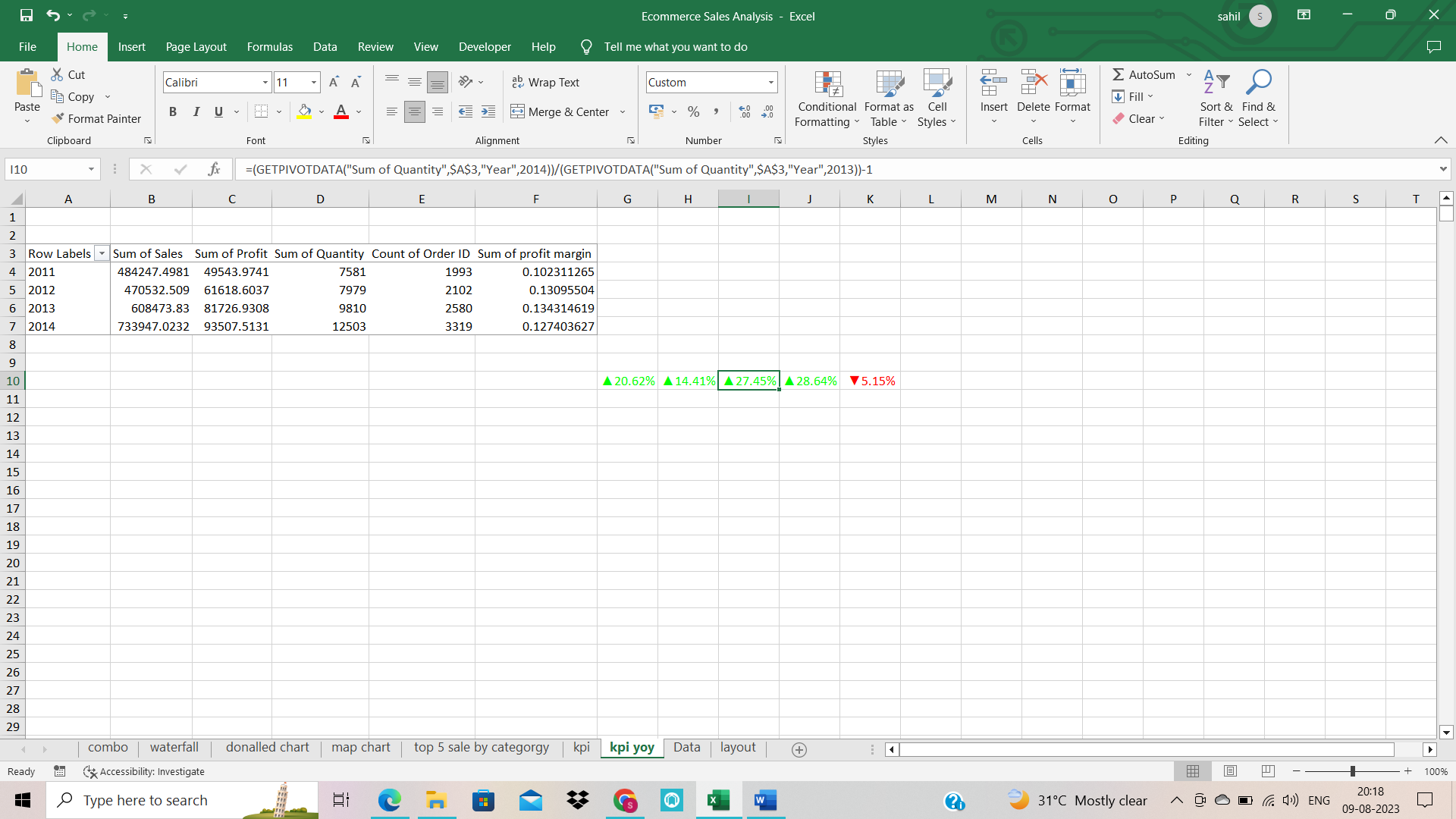
(Current year/pre-year)-1 for all KPI YOY



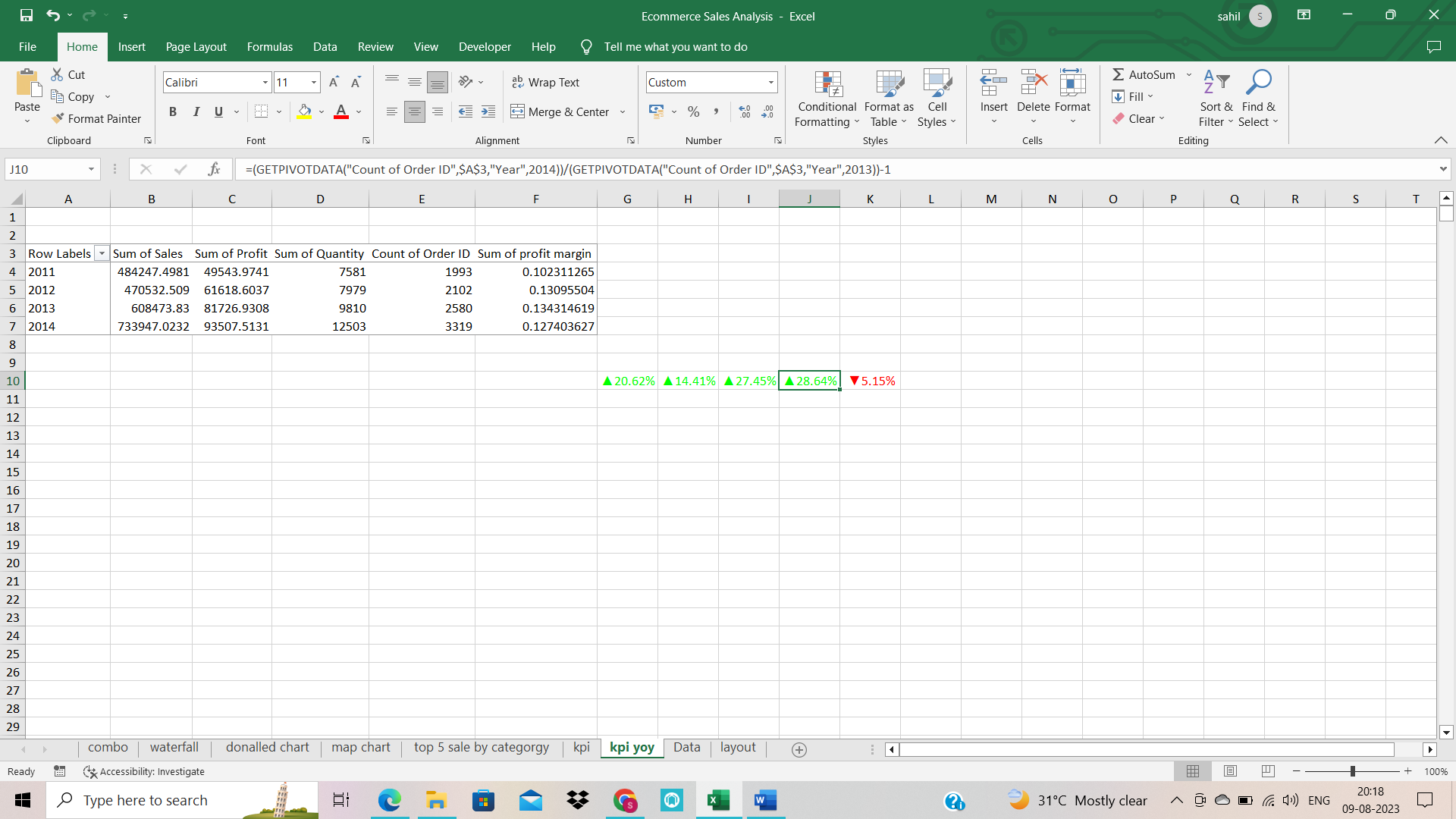
**Profit YOY**



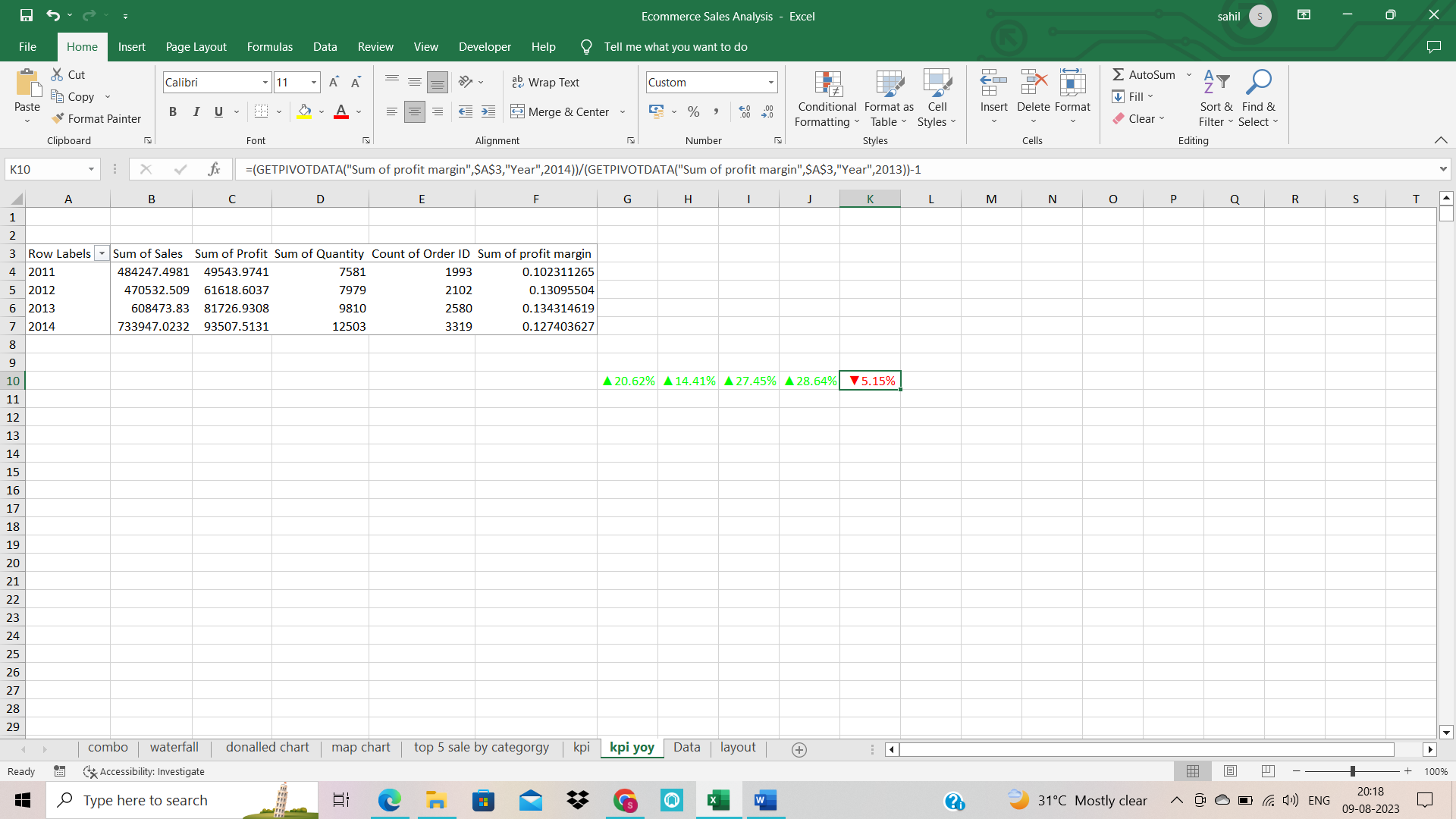
**Quantity YOY**

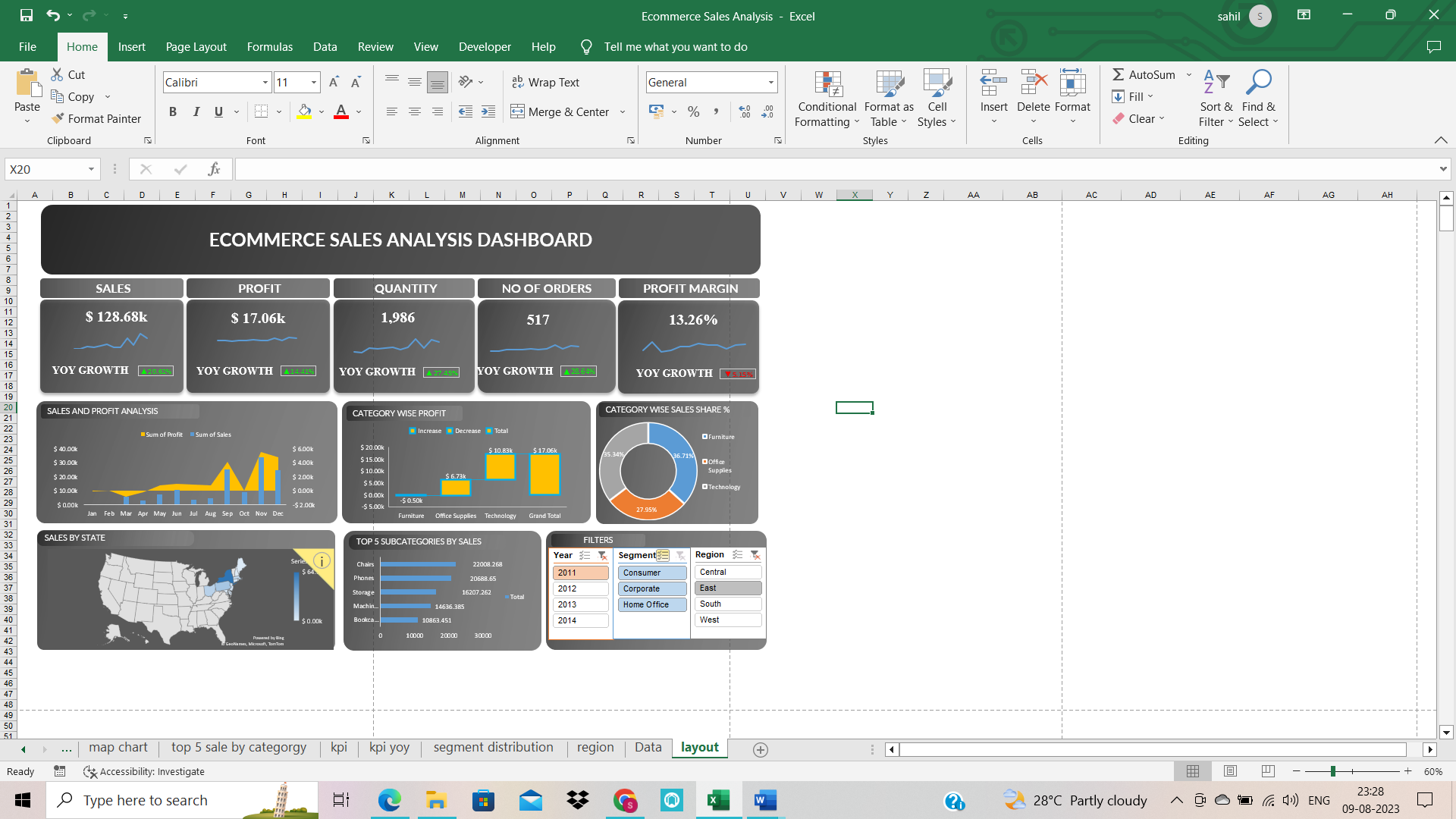


**Order YOY**



**Profit margin YOY**



**FINAL DASHBOARD**