**Excel - Capstone Project Case Study Assessment Solutions**

**Task 1: Data Preparation and Cleaning**

1. **Import the Dataset:**

* I imported the sales, product, and customer datasets into separate sheets named "Sales\_Data", "Product\_Data", and "Customer\_Data" in an Excel workbook using "Data -> Get Data -> From Workbook." This ensured that the data was organized and ready for analysis.

1. **Data Cleaning:**

* I ensured there were no missing values in key columns and removed duplicate records from the sales data using Excel's "Data -> Go To Special -> Blanks" and "Data -> Remove Duplicates" features. The data was completely clean without any missing values and duplicates.

**Task 2: Data Analysis**

1. **Cell Referencing:**

* Used the cell references to create a summary Table in the new sheet named “Tasks” to display the Total Sales based on the product category.

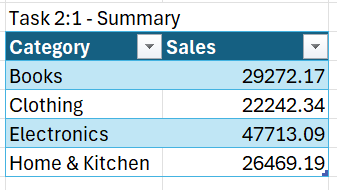
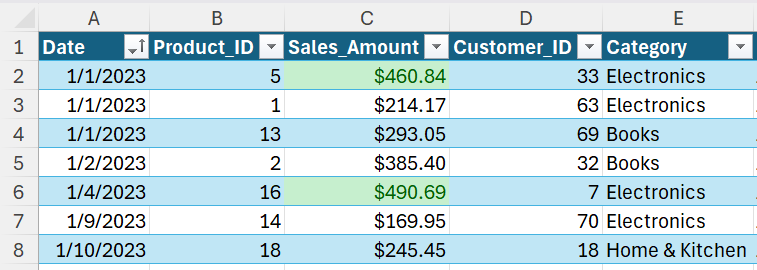
**Formula: =SUMIF(Sales\_Data!$E$2:$E$501, B4, Sales\_Data!$C$2:$C$501)**

Table starts from Column B in excel sheet

Electronics has the highest sales with the value of $47713.09.

1. **Sorting and Filtering:**

* The sales data has been sorted according to the “Date” column in the ascending order.It has been observed that all the sales data has been recorded for the year 2023, with total 501 sales and sold in the categories of Electronics and Books to the customers with id’s 33,63,69.
* To apply the Filters to show only sales data from specific category we use Column level filters which are at the top of the column.

****

1. **Date and Time Functions:**

* Extracted Month and Year from the date column using the formulas.

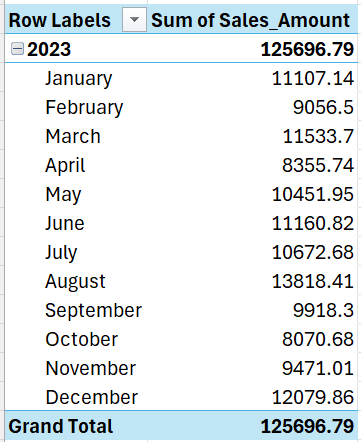
**Month =TEXT(A2,"mmmm")**

**Year =YEAR(A2)**

Created a pivot table by Insert->Pivot Table->Sheet.

A screenshot of a calendar

Description automatically generatedThe pivot table is as follows:



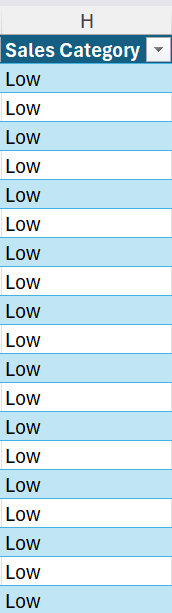
**August** has the highest sales with $13818.41 followed by December, June.

The sales are constantly high in the Middle of the Year.

1. **If-Else Conditions:**

* Created a new column in the Sales\_Data sheet namely “Sales Category” to classify the sales into Low and High.

**Formula =IF(C2 > 500, "High", "Low")**

****

All the sales are categorized in **Low** because the Maximum sale amount done is $499.90.

1. **Descriptive Analysis:**
   * Calculated the Average, median and Total Sales for each product category using Statistical Functions.

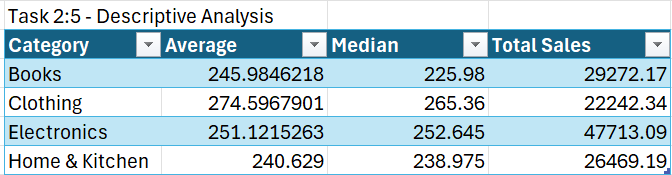
Here are the formulas used:

**=AVERAGEIF(Sales\_Data!$E$2:$E$501, B11, Sales\_Data!$C$2:$C$501)**

**=MEDIAN(IF(Sales\_Data!$E$2:$E$501=B11, Sales\_Data!$C$2:$C$501))**

**=SUMIF(Sales\_Data!$E$2:$E$501, B11, Sales\_Data!$C$2:$C$501)**

Table starts from Column B in excel sheet

****

**Clothing** has highest Average as well as Median but **Electronics** has higher Sales with larger than other categories. Therefore Many Electronic items has been sold under median price compared to products in other categories.

**Task 3: Excel Essentials:**

1. **Excel tables:**
   * Converted the sales data into “Sales\_Table” using the below steps.

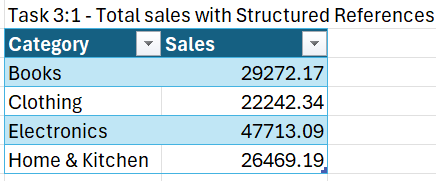
Selected Total data->Insert->Table.

I have used structured reference to calculate the total sales for each product category.

Structured reference is used on a particular column of a table.

**Formula =SUMIF(Sales\_Table[Category], B18, Sales\_Table[Sales\_Amount])**

Table starts from Column B in excel sheet

****

Electronics has the highest total sales followed by Books, Home & Kitchen and Clothing.

1. **Auto-fill, Custom Lists, and Flash-fill**
   * Used Autofill to fill the month. Written “January” in the first cell. Dragged the fill handle from the bottom-right corner of the cell down the column to complete the series of months.

**A screenshot of a calendar

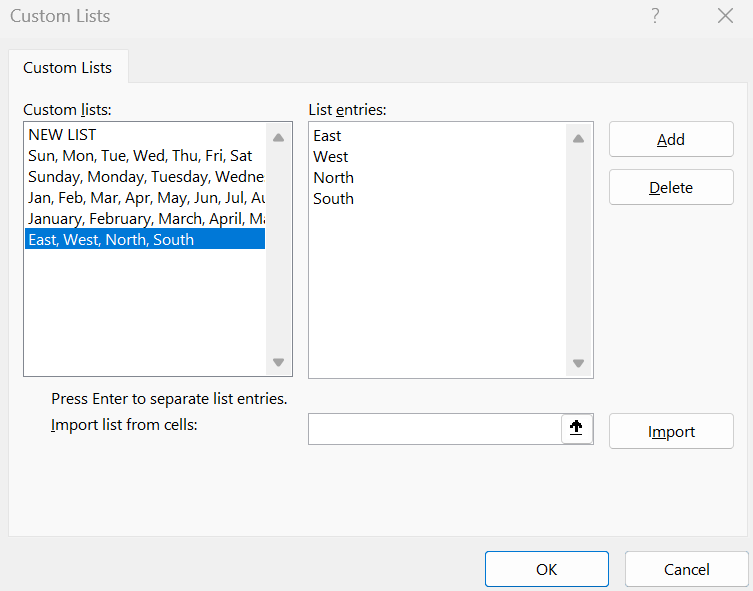
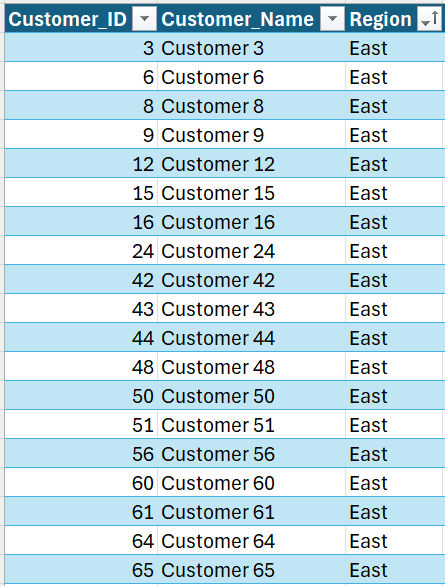
Description automatically generated**

* + Created a custom list for regions and used it to sort the Customer\_Data.

Steps used in creation of the custom list File->Options->Advanced->General->Edit Custom lists->Entered List->Add->Ok.

Custom List – [East, West, North, South]

Sorted the Customer\_Data based on the custom list.

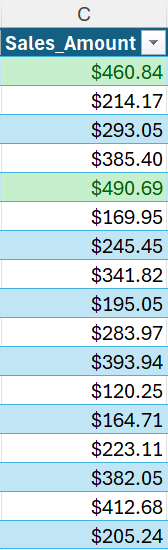
* + Typed First Name and Last name in separate columns i.e “Customer” and “1” and pressed CTRL+E to fill the columns.

A screenshot of a computer

Description automatically generated

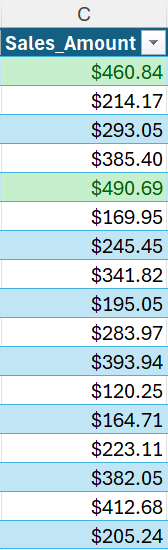
1. **Number Formatting:**
   * Clicked and dragged to highlight the “Sales\_Amount” column.

Home tab -> Number Format-> Currency



* + **Selected the Column:** Highlighted the sales amounts column.

**Applied Conditional Formatting:** Used 'Conditional Formatting' under the 'Home' tab, set a rule for values greater than $1000, and formatted it as “Yellow filled with dark green text”.



Since There are no values greater than $1000, so there is no cell in Yellow colour filled with dark green text.

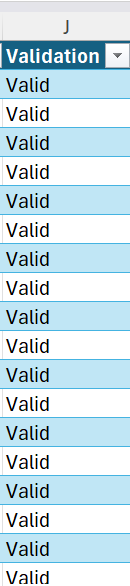
**Task 4: Advanced Excel formulas:**

1. **Logical Formulas:**
   * I have used all the three given logical functions (AND, OR, NOT) and created a column “Validation” which divided the sales amounts into “Valid and Invalid”.

Valid if sales amount between $100 to $1000. Invalid Otherwise.

**Formula:**

**IF(AND(C2 >= 100, C2 <= 1000), "Valid", IF(OR(C2 < 100, C2 > 1000), "Invalid", ""))**

****

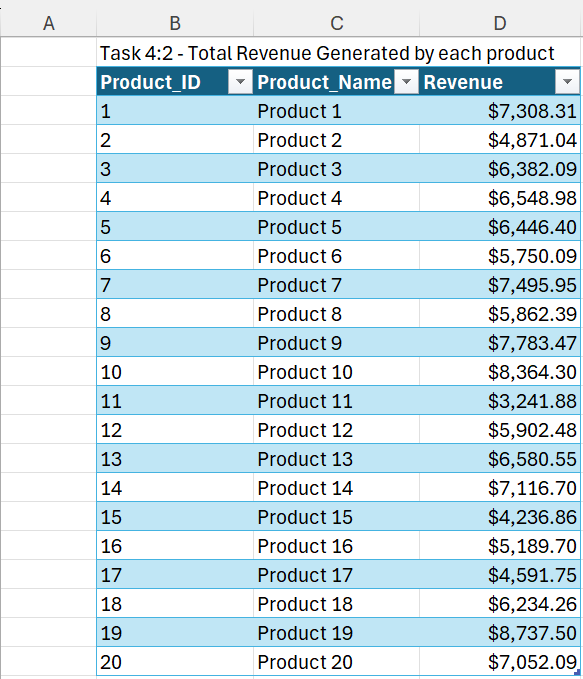
Maximum Proportion of data is “Valid”, i.e the maximum times each sale is done in between $100 to $1000.

1. **Math Formulas:**

* The Total revenue of each product has been calculated using SUMPRODUCT function.

It has been calculated in the sheet **“Tasks”**.

**Formula =SUMPRODUCT((Sales\_Data!$B$2:$B$501=B25) \* (Sales\_Data!$C$2:$C$501))**

****

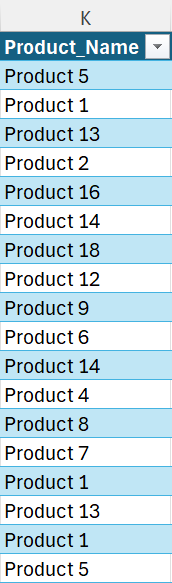
**Product 19** generates highest revenue with **$8,737.50** , followed by Product 10,9.

1. **Lookup and Reference Formulas:**

* Created Column with the name “Product Name” in the sheet “Sales\_Data”.

**VLOOKUP** function is used in the first cell of the column to reference the Product Name which is present in “Product\_Data” sheet and dragged the fill handle down to apply to all.

**Formula =VLOOKUP(B3, Product\_Data!$A$2:$B$100, 2, FALSE)**

****

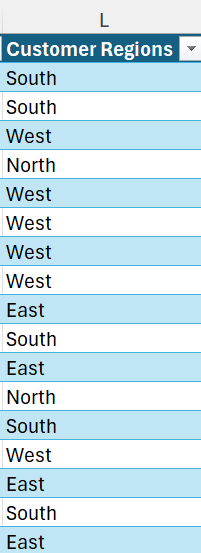
* Uses the INDEX and MATCH functions to retrieve customer regions based on Customer ID.

Added a new column “Customer regions” in the “Sales\_Data” sheet and in the first cell formula with INDEX and MATCH has been used.

Dragged to the end of the column for the function to apply to all the cells.

**Formula**

**=INDEX(Customer\_Data!$C$2:$C$101, MATCH(D2, Customer\_Data!$A$2:$A$101, 0))**

****

**East** is the region where maximum number of Customers ordered with count being 170.

1. **Text Formulas:**

* Created a column with name “Concatenation” that concatenates the Product Name and Category using Text manipulation functions in the sheet “Sales\_Data”.

Entered the below given formula and dragged through all the cells of the column.

**Formula =CONCATENATE(K2, " - ", E2)**

**A screen shot of a list

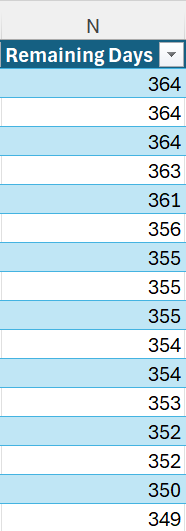
Description automatically generated**

1. **Date and Time Formulas:**

* Calculated the number of days between each sale date and the end of the year.

Created a new column “Remaining Days” in the “Sales\_Data” sheet to showcase the difference between sale date and year end.

The below given formula is used and dragged to all the cells of the column.

**Formula =DATE(YEAR(A2), 12, 31) - A2**

The Highest Difference is 364 days and the least difference is 0 days.

**Task 5: Data Analysis Techniques:**

1. **Named Ranges:**

* Created Named Range for “Sales Amount” column in “Sales\_Data” sheet.

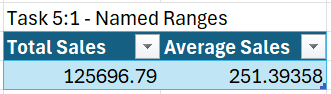
Selected “Sales Amount” column->Formulas->Define Name-> Entered “SalesAmount\_NamedRange”.

This has created a named range for Sales\_Amount.

In the “Tasks” sheet SUM, AVERAGE formulas has been used to the created Name Range.

**Formulas =SUM(SalesAmount\_NamedRange)**

**=AVERAGE(SalesAmount\_NamedRange)**



1. **Data Validation:**

* **Selected the Sales Amounts Column**: Highlighted the range where sales amounts were entered i.e C2.

**Opened Data Validation**: Went to the Data tab and clicked Data Validation in the Data Tools group.

**Set Validation Criteria**: In the Data Validation dialog box, set the criteria to allow Decimal values greater than 0.

**Applied the Rules**: Clicked OK to enforce the data validation rules.

Currently, There is no data in Negative Numbers all are in Positive.

1. **Conditional Formatting:**

* Selected the “Sales Amount” column in the Sales\_Data sheet.

Home tab-> Conditional Formatting->Select Top/Bottom rules->Select Top%...

This selects Top 10% of total sales amounts.

**A screenshot of a computer screen

Description automatically generated**

Top 10% of Sales Amount ranges from $445.14 to $499.90

**Task 6: Excel Charting:**

1. **Intro to Excel Charting:**

* Created a bar chart to display total sales for each product category.

Created a pivot table with “Category” in the rows field and “Sales\_Amount” in the Values field.

Select Insert tab->Insert Column or Bar Chart->Selected 3d column.

A screenshot of a graph

Description automatically generated

From the above chart we can observe **Electronics** has the highest sale in the year 2023.

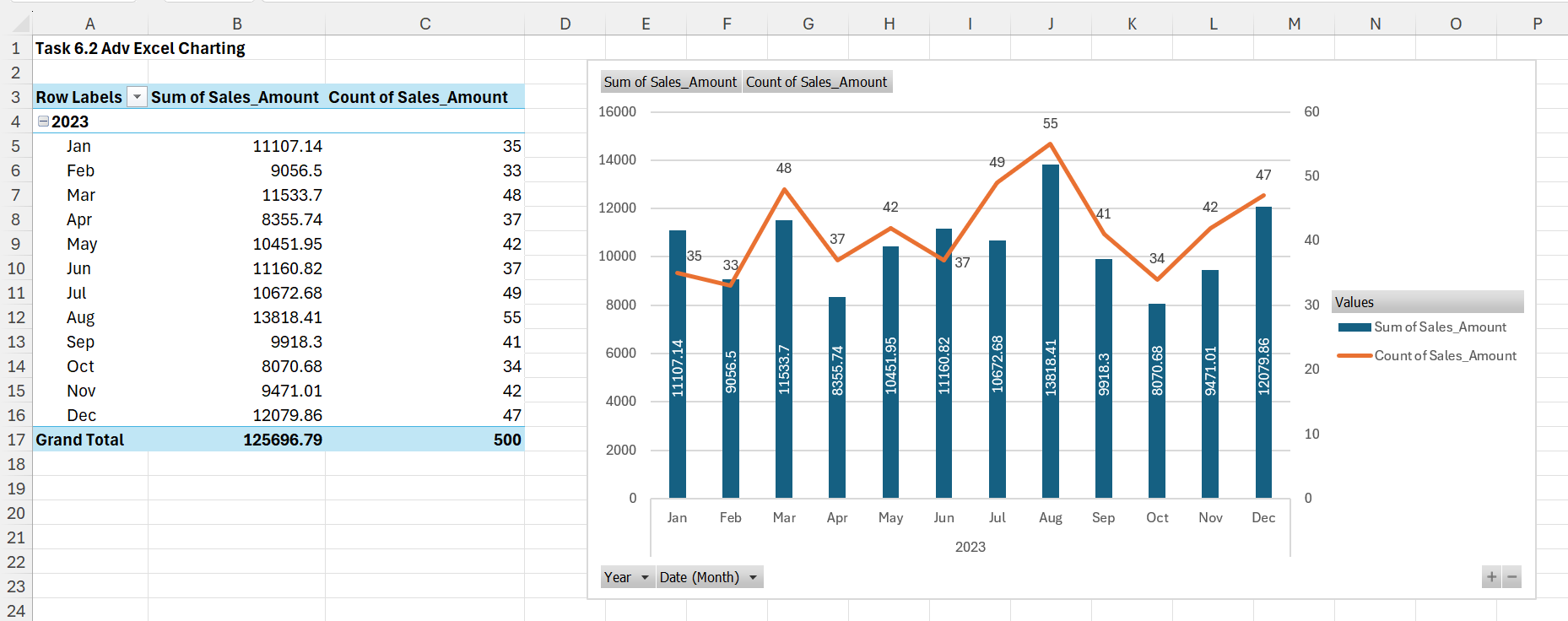
1. **Advanced Excel Charting:**

* Created a combo chart to display total sales and no.of sales for each Month.

Created a pivot table with “Year, Date(Month)” in the rows field ,and “Sum of Sales\_Amount, Count of Sales\_Amount” in the Values field.

Select Insert tab->Insert Combo Chart->Selected Clustured column- Line on Secondary axis.

Sum of Sales on Bars and Count of sales on Line.



**August** month is with both highest sum of sales with **$13818.41** and highest count of Sales with **55** ,i.e, Sales are high at the Middle of the Year.

1. **Dynamic Charts:**

* Created a Dynamic chart for the Total sales month wise which updates according to the category selected on the dropdown menu.

Created a pivot table with “Year, Date(Month)” in the rows field ,and “Sum of Sales\_Amount,” in the Values field.

Created a 2-D Bar chart using the steps followed in Task 6.1

Inserted a slicer with “Categories”.

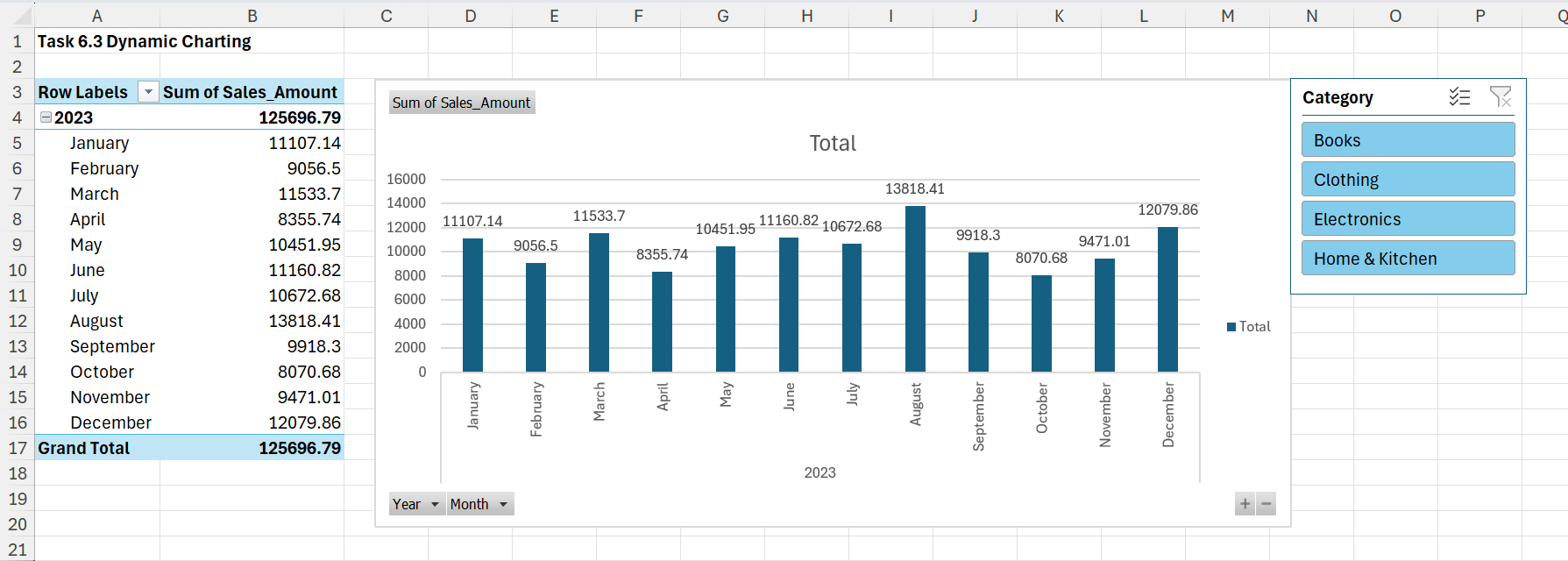
Here are few Observations from the above chart:

**August** has the highest Total sales for all categories combined with **$13818.41**

**July** has the highest sales for **Books** category with **$3508.3**

**August** has the highest sales for **Clothing** category with **$3275.95**

**January** has the highest sales for **Electronics** category with **$5686.69**

**August** has the highest sales for **Home & Kitchen** category with **$****4291.56**

**Task 7: Advanced Topics:**

1. **Pivot Tables:**

* Created Pivot Table that summarizes total sales by customer region and product category.

Insert->Pivot Table->Sheet.

Selected “Customer Regions” in the rows field, “Category” in the column field, “Sum of Sales\_Amount” in the values field.

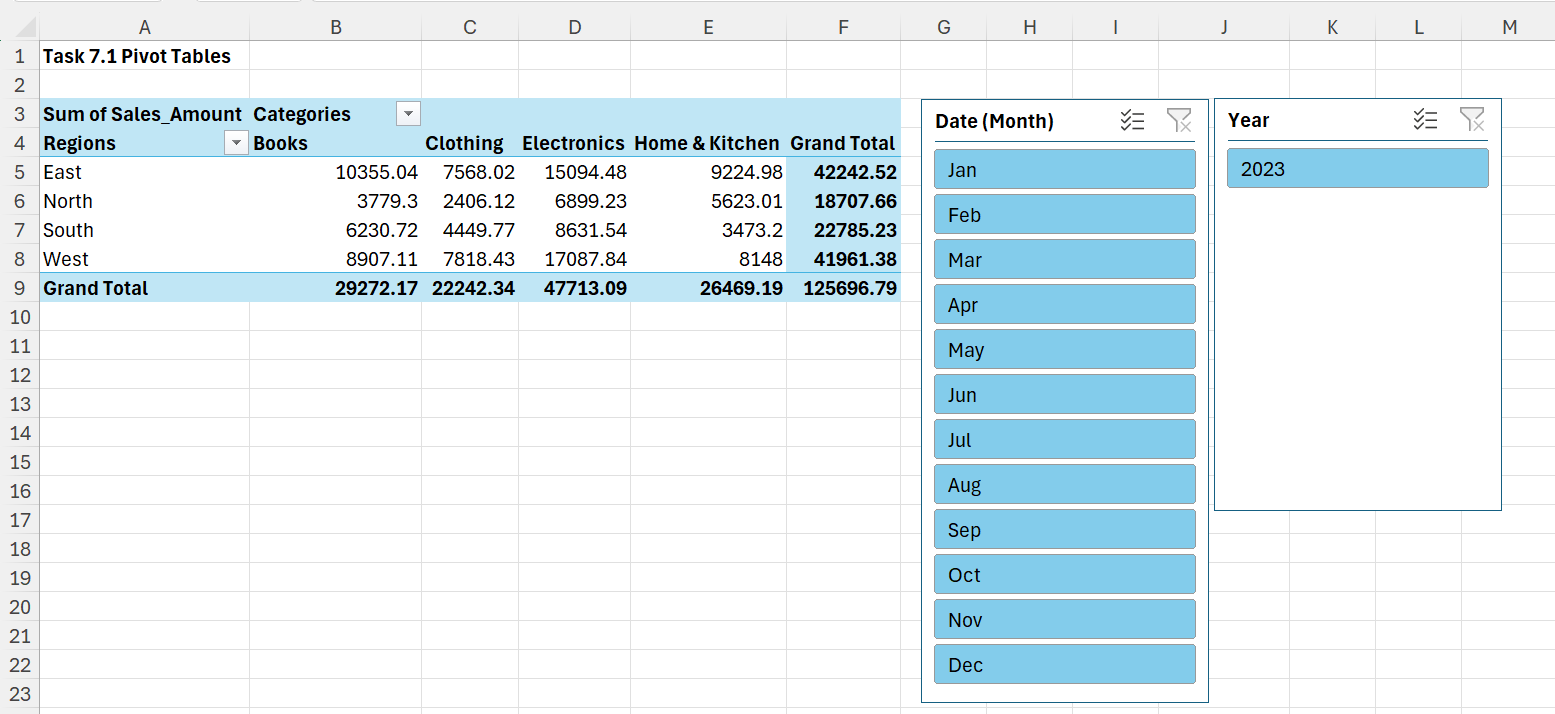
* Inserted 2 Slicers with “Month” and “Year” .

**Observations:**

Total sales are **$125696.79**

**East** region has the highest total sales with **$42242.52**

**Electronics** category has the highest total sales with **$47713.09**

****

1. **Advanced VLOOKUP:**

* Created a LOOKUP table in the “Tasks” sheet which has 0 is low, 200 is medium and 400 is high. The values are considered in this way because 499.90 is the highest sales amount present in the data.

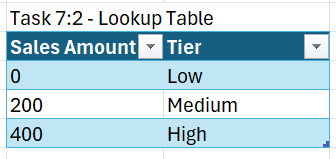
Now a new column named “Tier” is added to the Sales\_Data sheet with reference given to the “Salestier\_Table” created in the tasks sheet.

**Formula =VLOOKUP(C2, Salestier\_Table, 2, TRUE)**

Entered the above formula in 1st cell and dragged to all the cells of the column.

A screenshot of a computer

Description automatically generated

**Medium** tier has the highest count of **212** followed by Low and High Tier.

**Recommendations to the Company:**

1. **Focus on High-Demand Products:**
   * Electronics: Since Electronics lead in sales, prioritize inventory, marketing efforts, and promotional campaigns for this category to capitalize on the strong market demand.
2. **Optimize Inventory Management:**
   * Ensure sufficient stock levels for high-performing categories like Electronics and Books to prevent stockouts and lost sales.
   * Implement demand forecasting techniques to manage inventory levels effectively for all product categories.
3. **Enhance Promotional Strategies:**
   * Develop targeted promotions and discounts for slower-moving categories like Clothing and Home & Kitchen to boost sales.
   * Utilize cross-selling and bundling strategies to increase average order value.
4. **Analyze Profit Margins:**
   * Evaluate profit margins for each product category and identify opportunities to improve profitability, such as renegotiating supplier terms or adjusting pricing strategies.
5. **Customer Segmentation and Personalization:**
   * Segment customers based on purchase behavior and tailor marketing efforts to different segments to improve engagement and conversion rates.
   * Personalize communication and offers based on customer preferences and purchase history.
6. **Expand Product Offerings:**
   * Explore opportunities to introduce new products or expand existing product lines in high-demand categories.
   * Conduct market research to identify emerging trends and customer needs.
7. **Leverage Data Analytics:**
   * Invest in advanced data analytics tools to gain deeper insights into sales trends, customer behavior, and market dynamics.
   * Use these insights to inform strategic decisions and drive continuous improvement.
8. **Improve Online Presence:**
   * Enhance the company's online presence through search engine optimization (SEO), social media marketing, and targeted online advertising to reach a broader audience.

By implementing these recommendations, the company can optimize its operations, improve profitability, and better meet customer needs across all product categories.