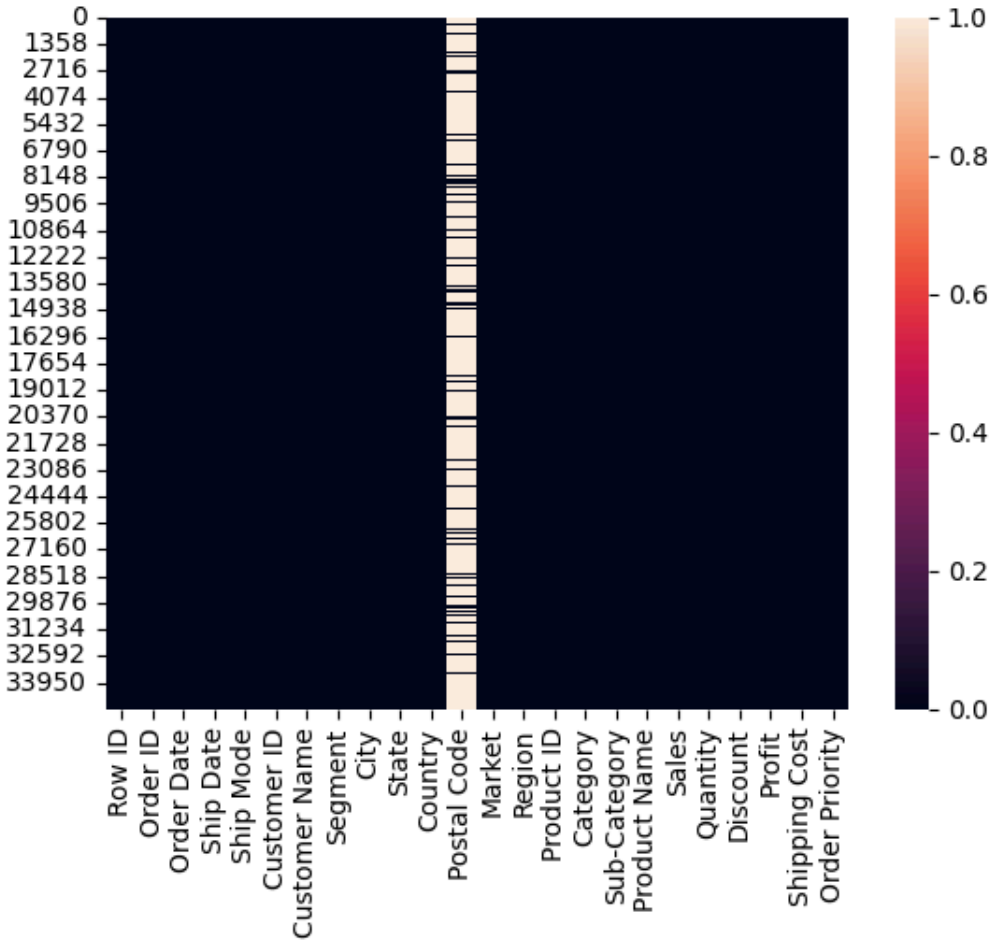
EDA REPORT

Overview of Dataset:

The dataset provided is an Excel file containing data on various transactions made by customers. The data includes information such as the date of the transaction, the customer's location, the class of the transaction, the product purchased, the price of the product, the quantity purchased, order criticality etc.

Preprocessing for Analysis:

- Cleaning Data (Removing Null Values): By heatmap of df.isnull() it is observed that Postal Code is the only column with null values and the count of non null values is only 5723. So it is dropped for better analysis and predictions



- Filtering columns for Sales and Profit analysis: A separate data frame is sliced from the original one for sales and profit analysis based on Market, Region, Segment, Category and subcategory, and later Sales and Profits are plotted based on this. The columns like Customer ID, Product ID, Product Name, Ship Date, Order Date are dropped.

- Financial Ratio Analysis: Four financial ratios which include GPR, NPR, PPQ, SCR are calculated for each entry and added as a column in the orders dataset. These are then used to get insights of a company's position in different markets, segments, categories and Sub-Category.

- Consumer Base Analysis: The number of customers in subsequent years is analyzed in different markets, segments, regions, Category and Sub-Category. The new column of year is added, and then the dataset is grouped based on year for analysis.

Structure of Data

- Rows and Columns:

  The dataset contains a total of 51290 rows and 24 columns. The columns present in dataset are:

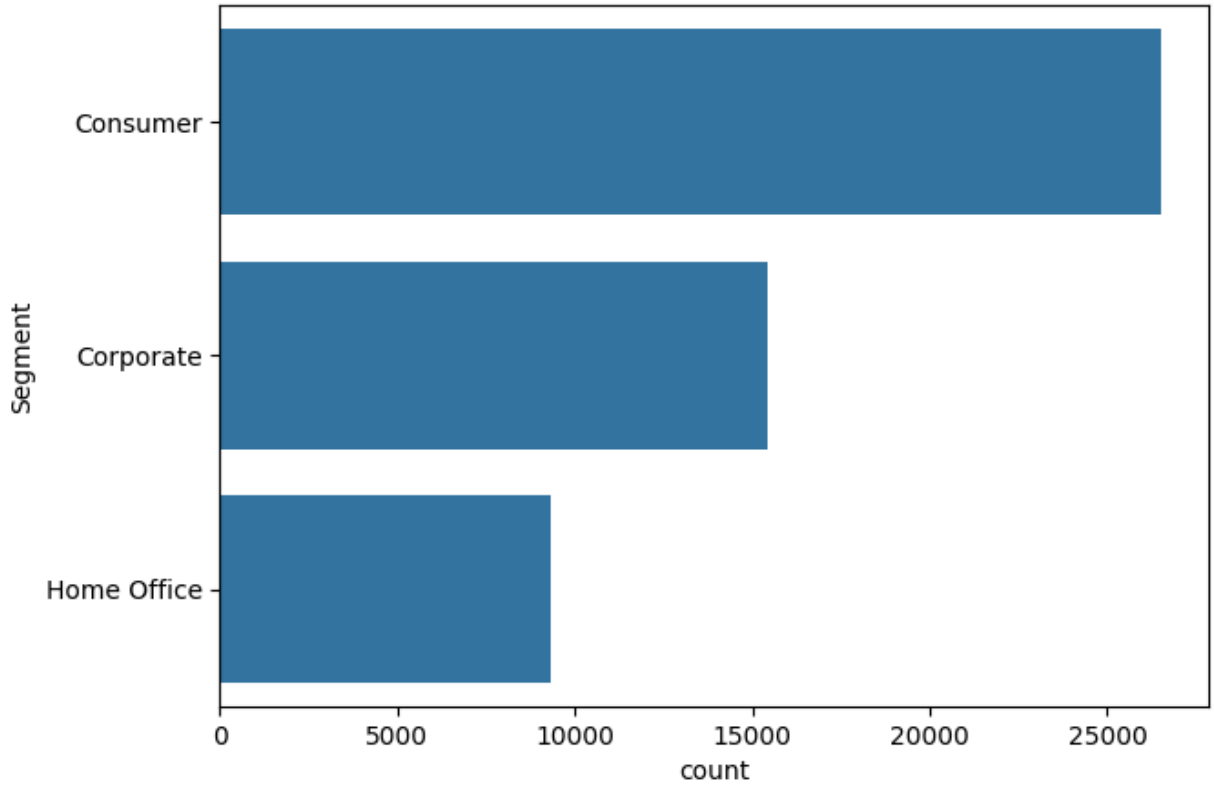
  1. Row ID
  2. Order ID
  3. Order Date
  4. Ship Date
  5. Ship Mode
  6. Customer ID
  7. Customer Name
  8. Segment
  9. City
  10. State
  11. Country
  12. Postal Code
  13. Market
  14. Region
  15. Product ID
  16. Category
  17. Sub-Category
  18. Product Name
  19. Sales
  20. Quantity
  21. Discount
  22. Profit
  23. Shipping Cost
  24. Order Priority

- Broad Classification of Columns:

  The columns belong one of the 4 main categories: Identity Columns, Numerical Columns, Categorical Columns, Date Time Columns and Name columns:

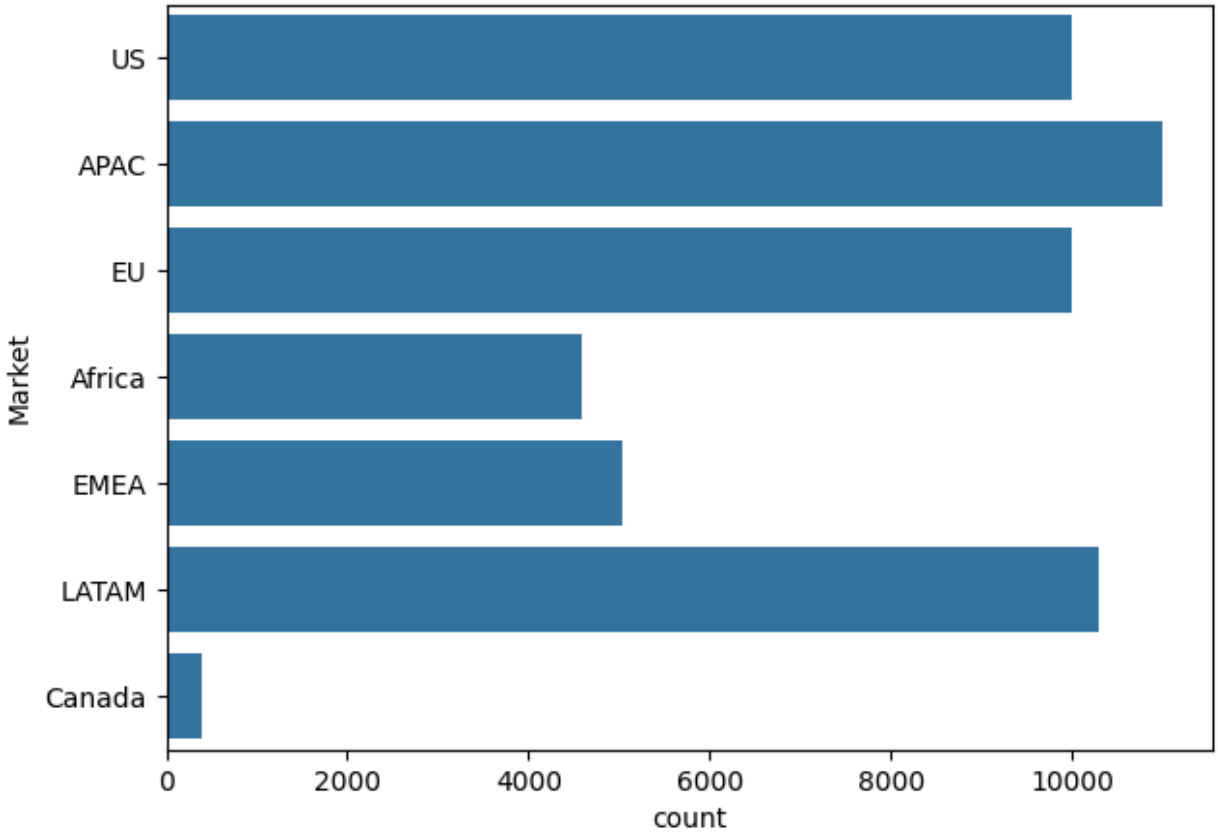
  1. Identity Columns: Row ID, Order ID, Customer ID, Product ID
  2. Numerical Columns: Sales, Quantity, Discount, Profit, Shipping Cost
  3. Categorical columns: Ship Mode, Segment, City, State, Country, Postal Code, Market, Region, Category, Sub-Category, Order Priority
  4. Date Time Columns: Order Date, Ship Date
  5. Name Columns: Product Name, Customer Name

- Categorical Columns:
    - Segment: There are three segments: Consumer, Corporate and Home Office. The number of orders in each segment:



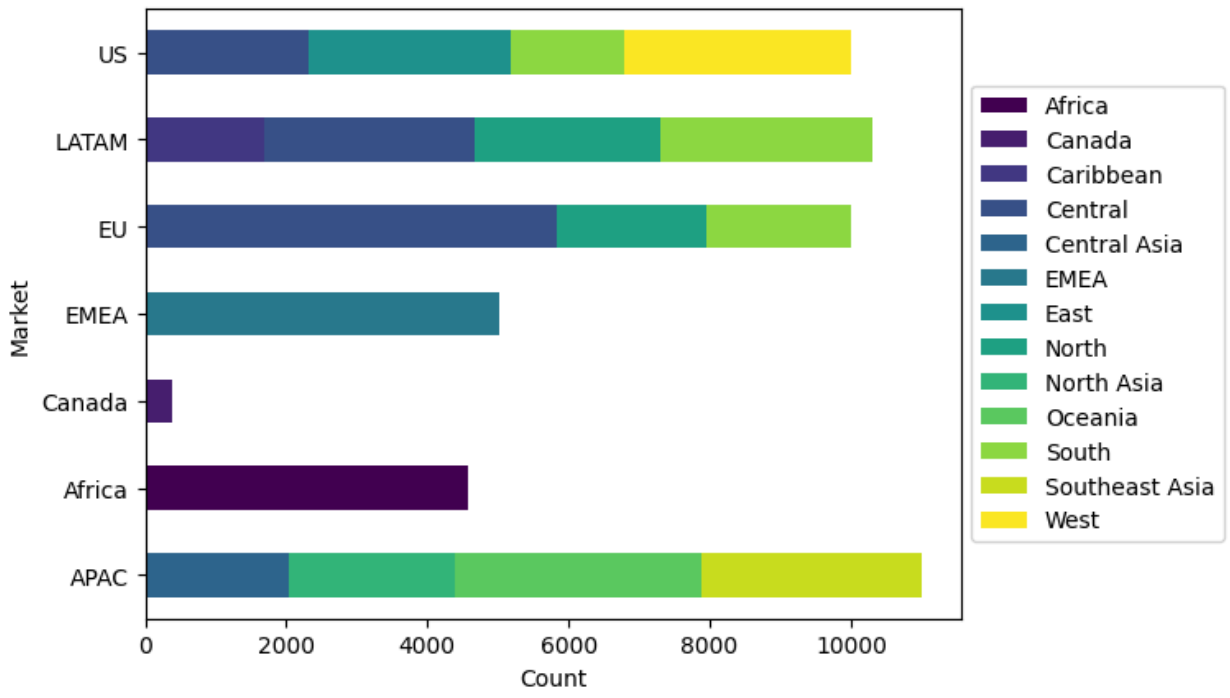
    - Market: There are 7 markets
        - US
        - APAC (Asia Pacific)
        - EU (European Union)
        - Africa
        - EMEA (Europe, Middle East, and Africa)
        - LATAM (Latin America)
        - Canada

The number of product ordered in each market:

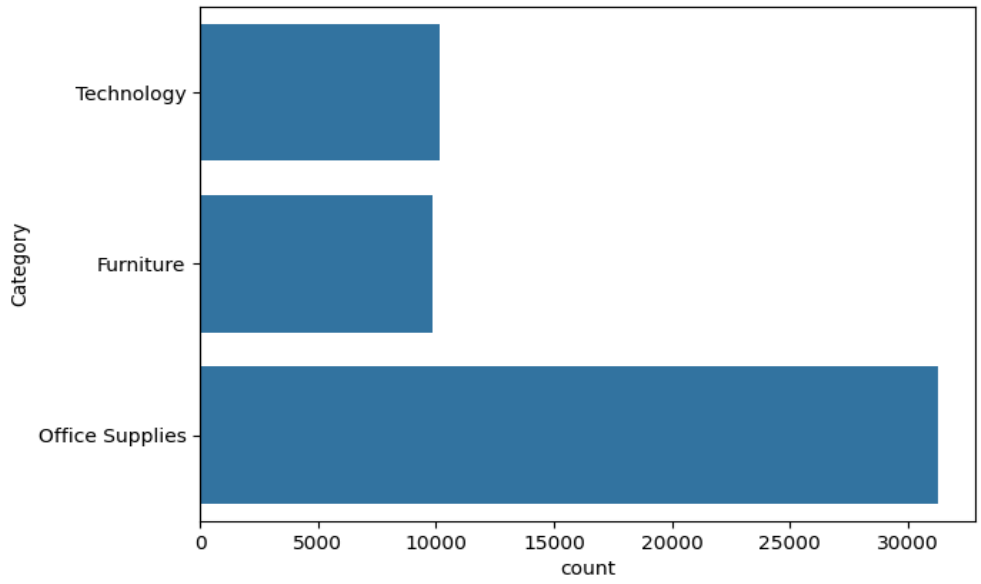


- Region: There are certain regions in each market:
  - Africa
  - Canada
  - Caribbean
  - Central
  - Central Asia
  - EMEA
  - East
  - North
  - North Asia
  - Oceania
  - South
  - Southeast Asia
  - West

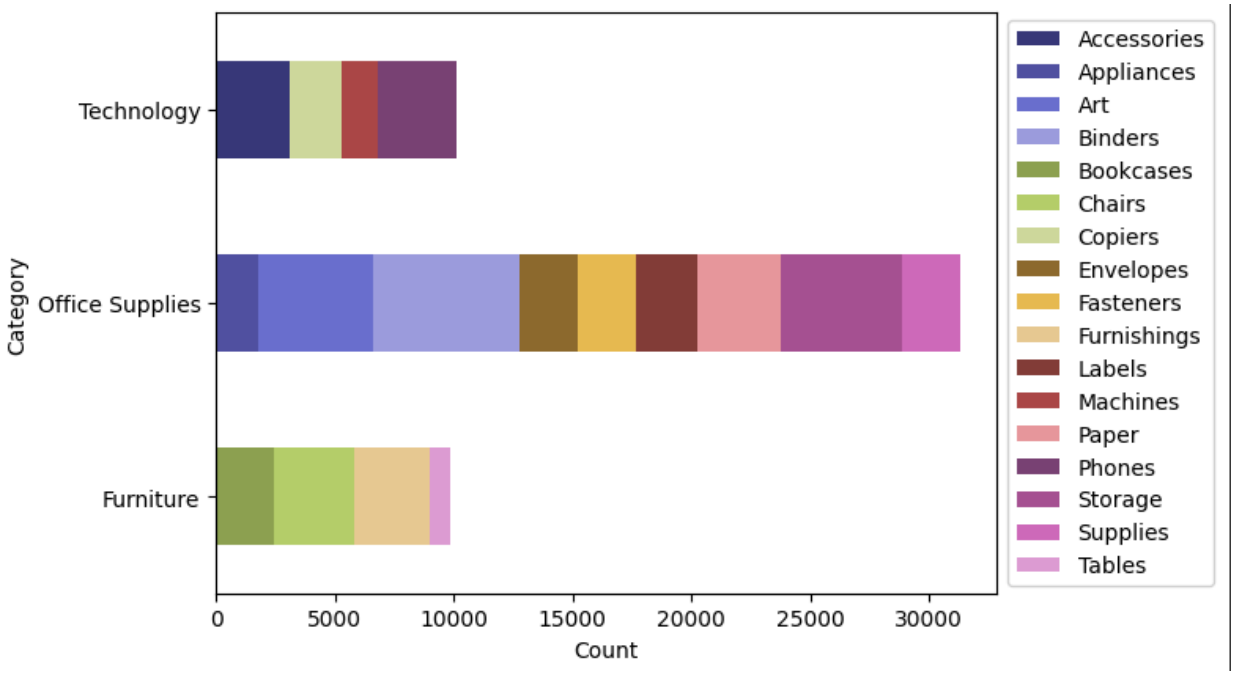
The number of orders in different regions of market:

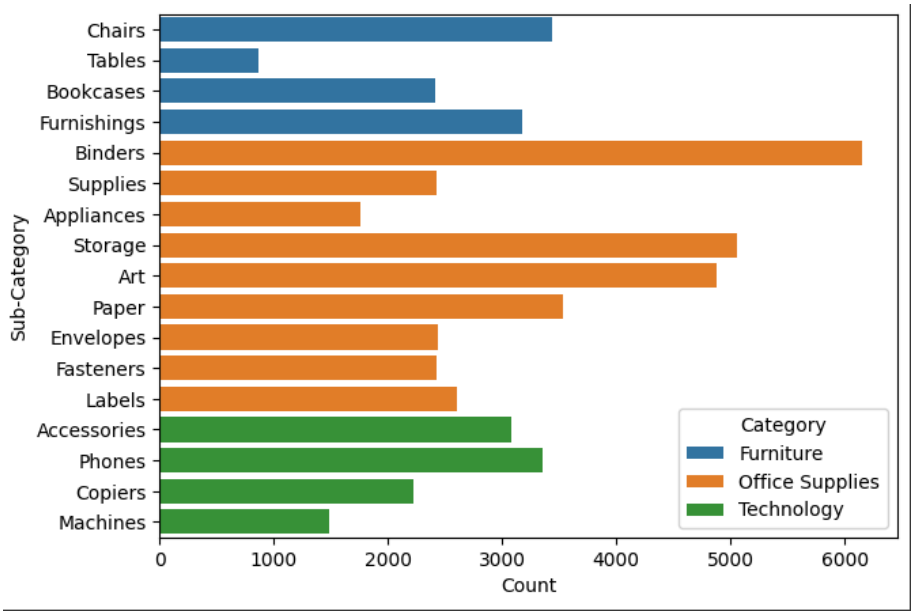


- Category: There are three categories: Technology, Furniture and Office Supplies. The number of orders in each category:

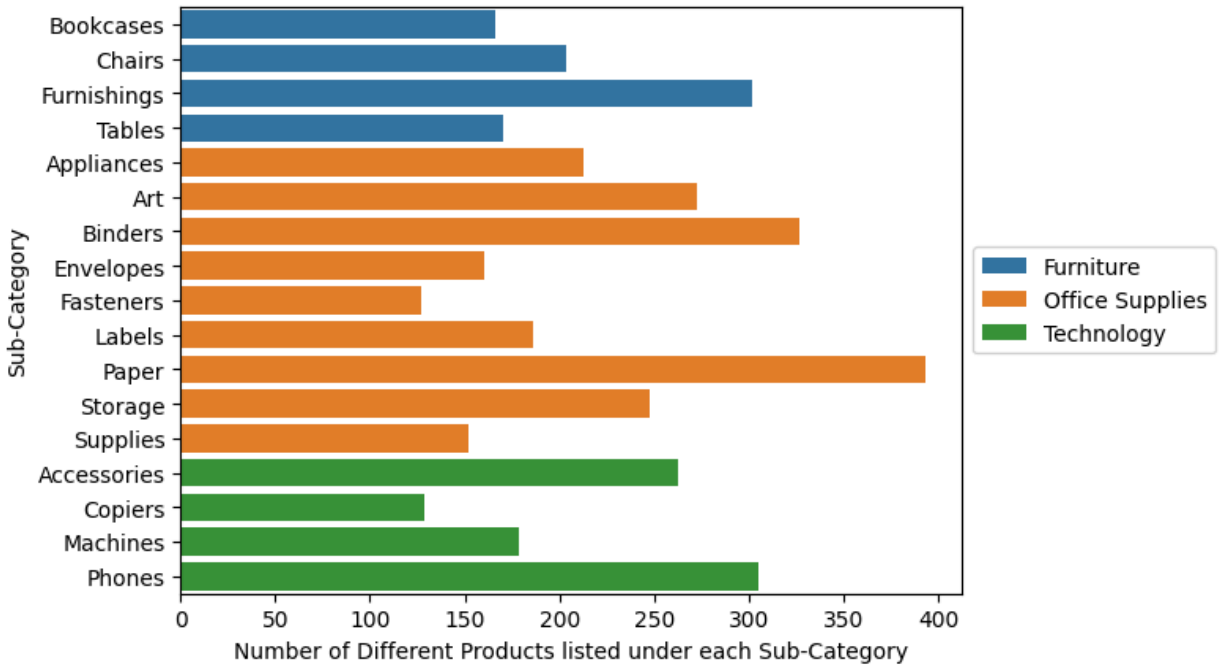


- Sub-Category: There are certain sub-categories in each category. It includes
  - Technology: Accessories, Copiers, Machines and Phones
  - Office Supplies: Appliances, Art, Binders, Envelopes, Fasteners, Labels, Paper, Storage and Supplies
  - Furniture: Book Cases, Chairs, Furnishings and Tables

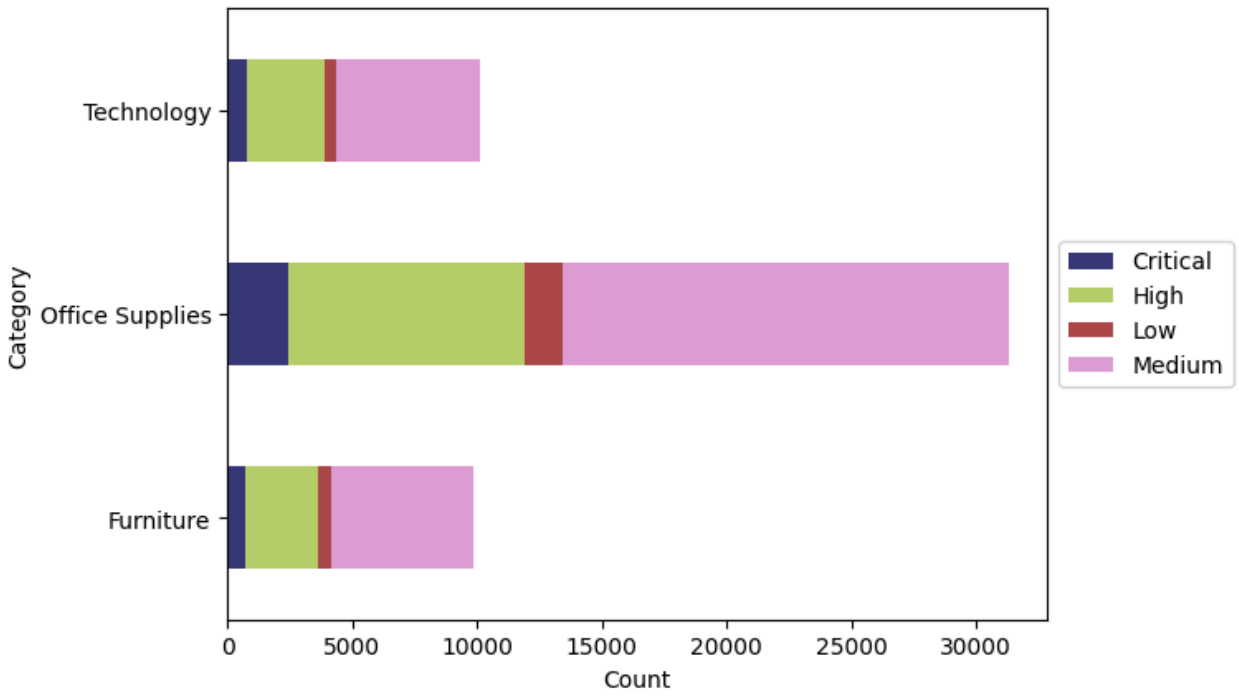




- Product Name: There are a total of 3788 different products listed in all categories and Sub-categories. The number of products listed for each Sub-Category are:

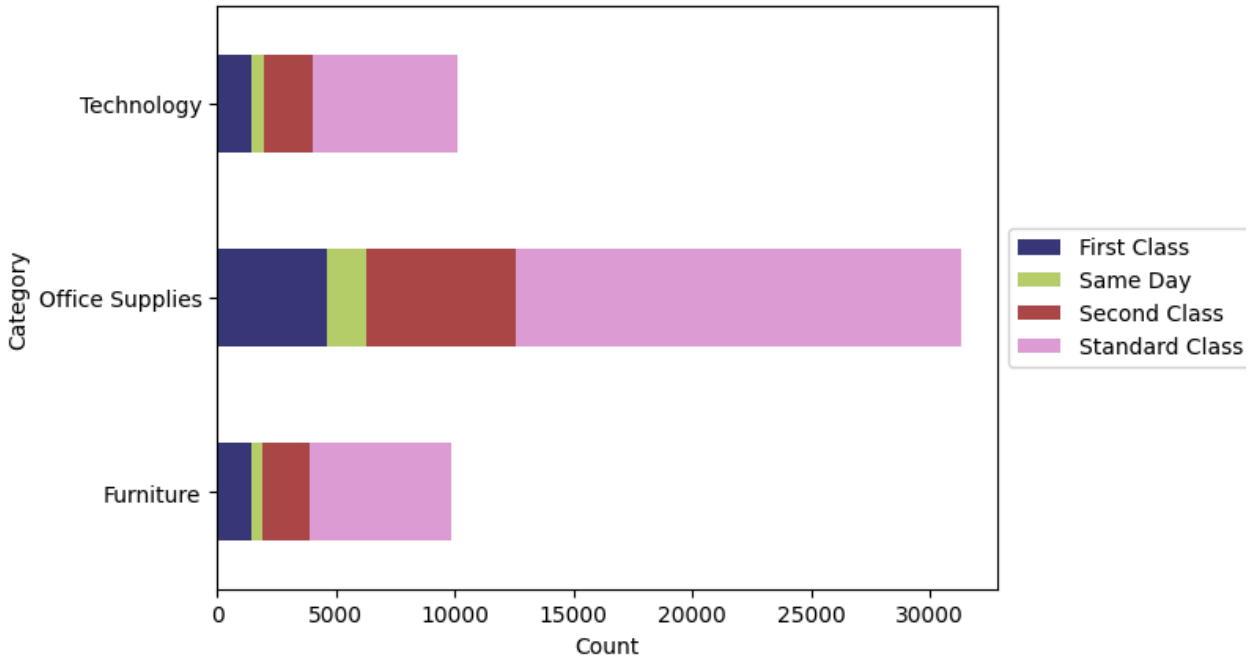


- Orders Priority: There are 4 categories for order criticality: Critical, High, Medium and Low. Medium is the highest in number followed by High, Critical and then Low.



- Ship Mode: There are four Ship Modes Company is offering
  - Same Day
  - First Class
  - Second Class
  - Standard Class

Here, the Standard Class is most chosen by customers, followed by Second Class, First Class and Same Day



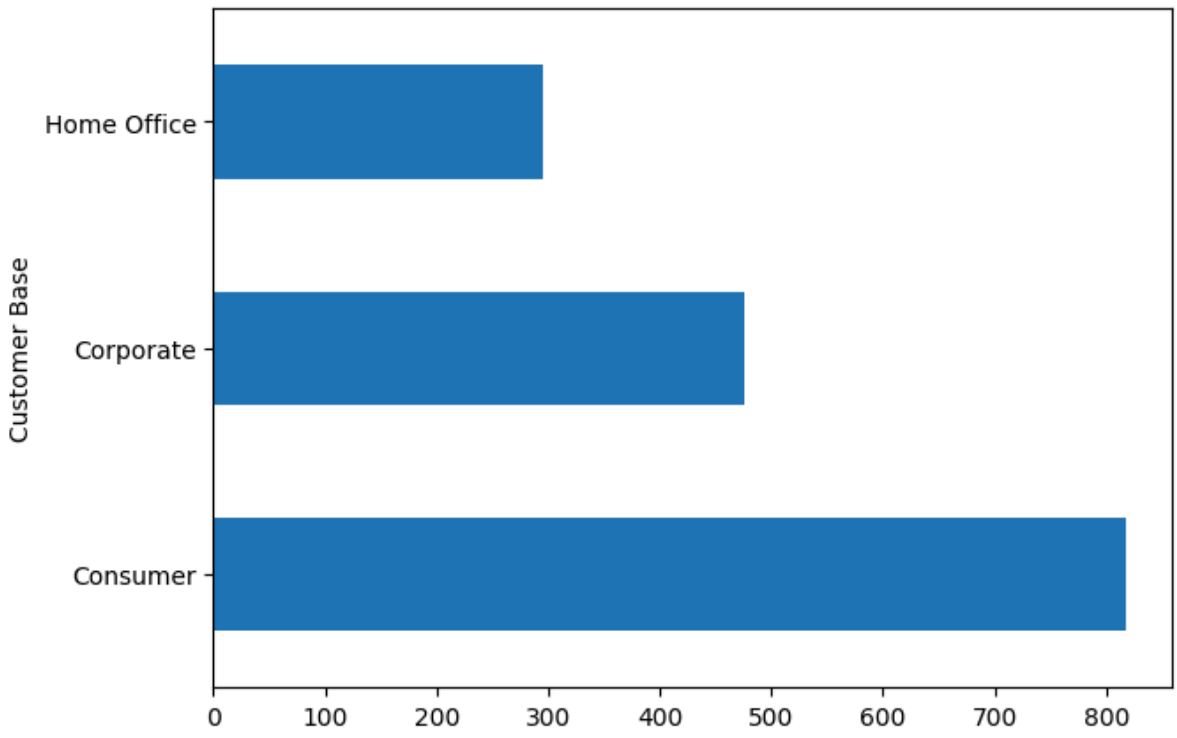
Correlation Matrix

A correlation matrix of sales quantity, discount, and shipping cost displays the pairwise correlation coefficients between these variables, indicating the strength and direction of their linear relationships, which is valuable for understanding how changes in one variable may be associated with changes in another.

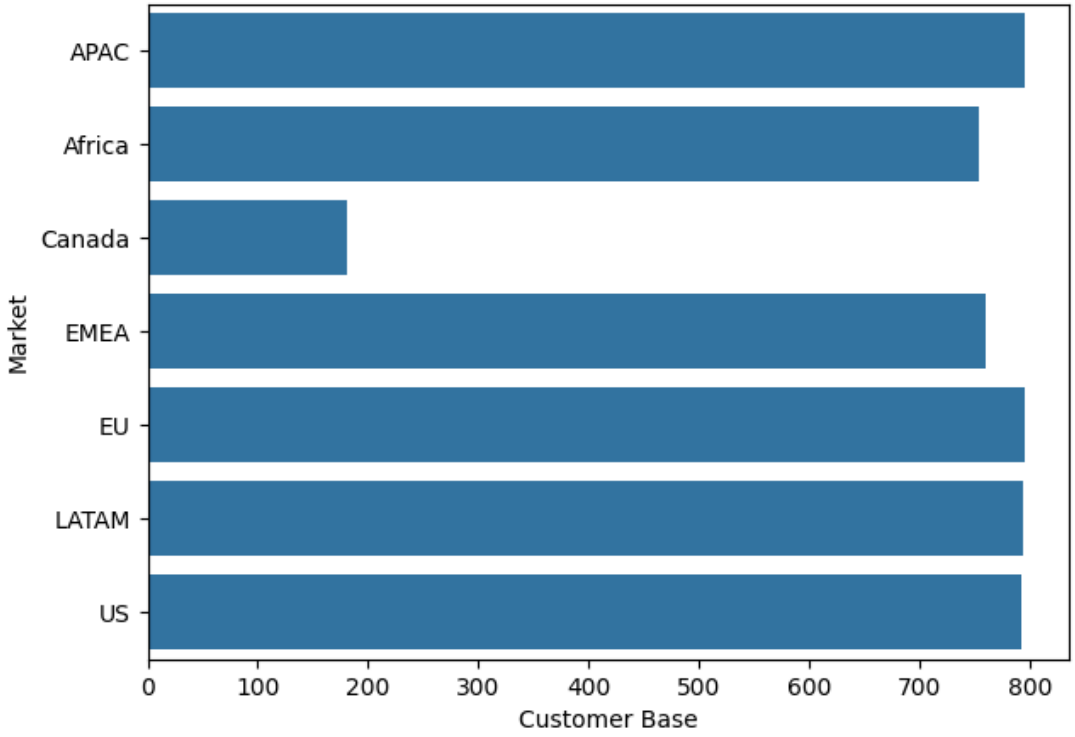
Customer Base Analysis

The company has a customer base of 1590 customers over the time of three years: 2011-2014.

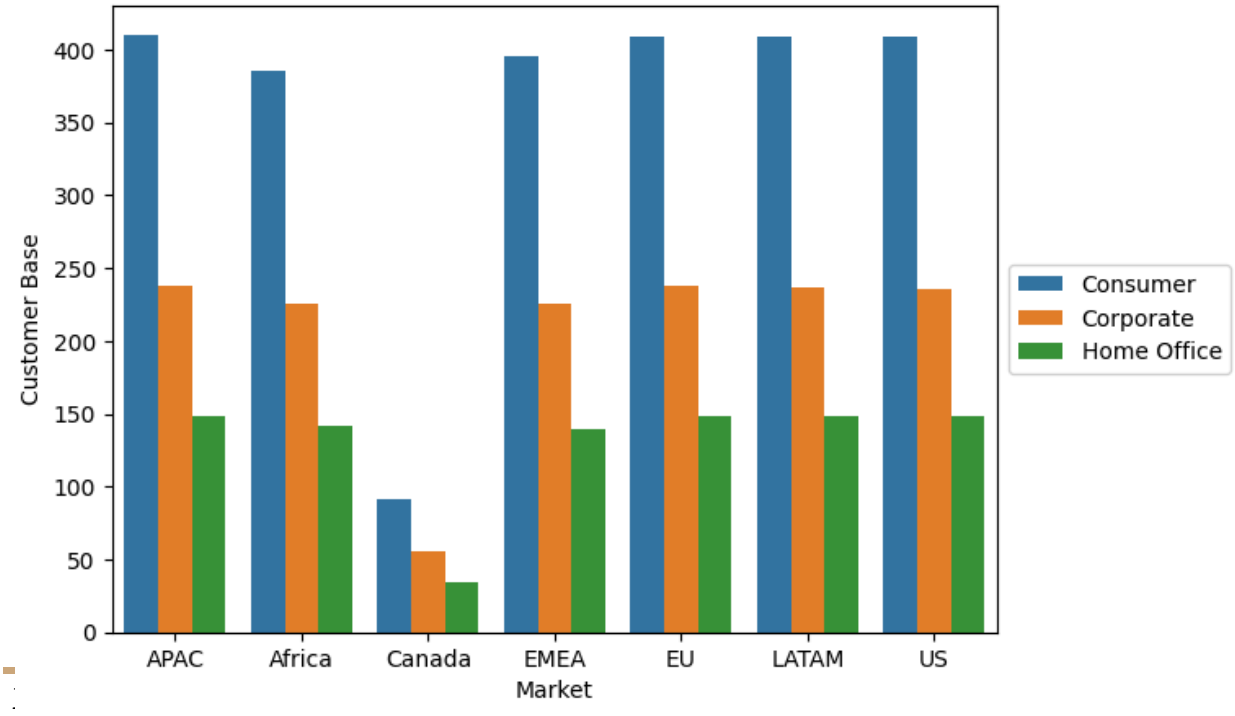
- The Customer Base in the 3 segments:



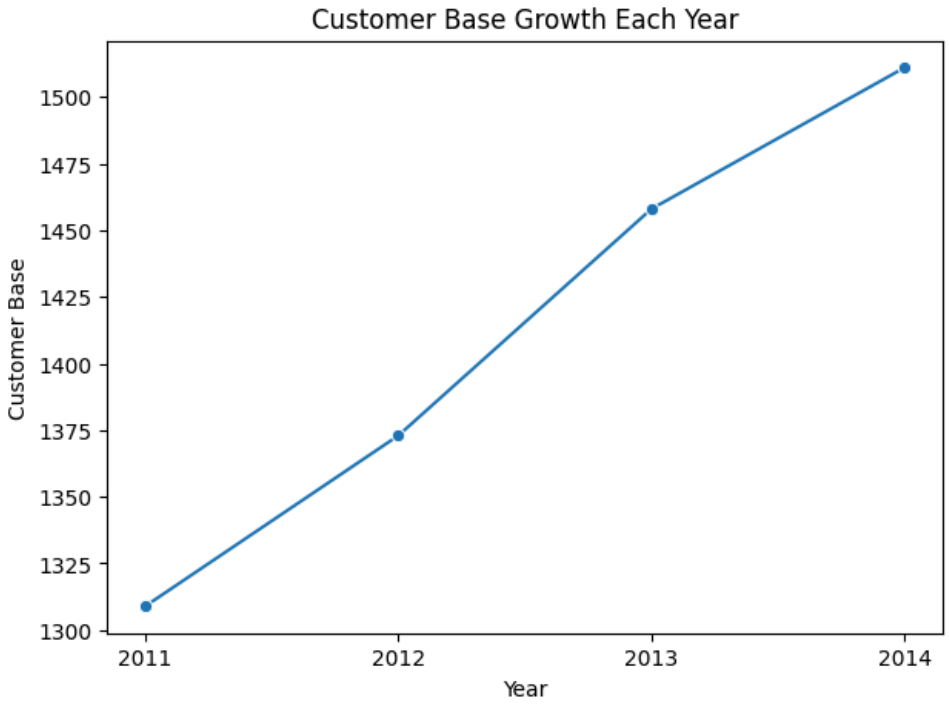
- The Customer Base in the different Markets:



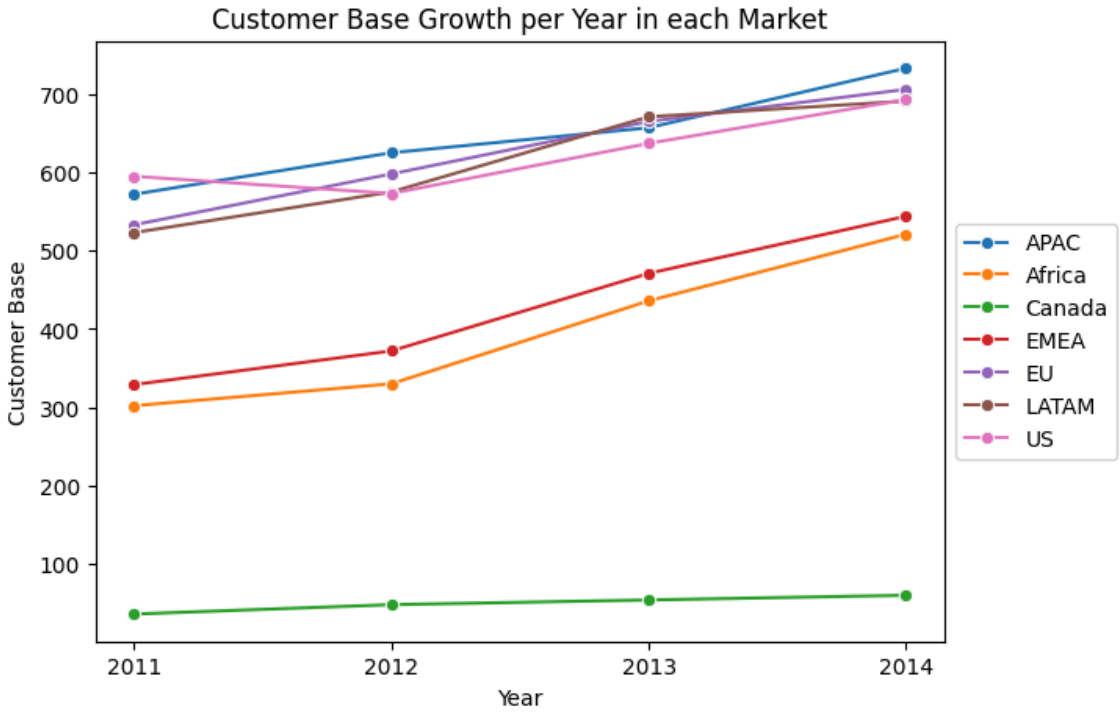
- The Customer Base of different Segments in different Markets:



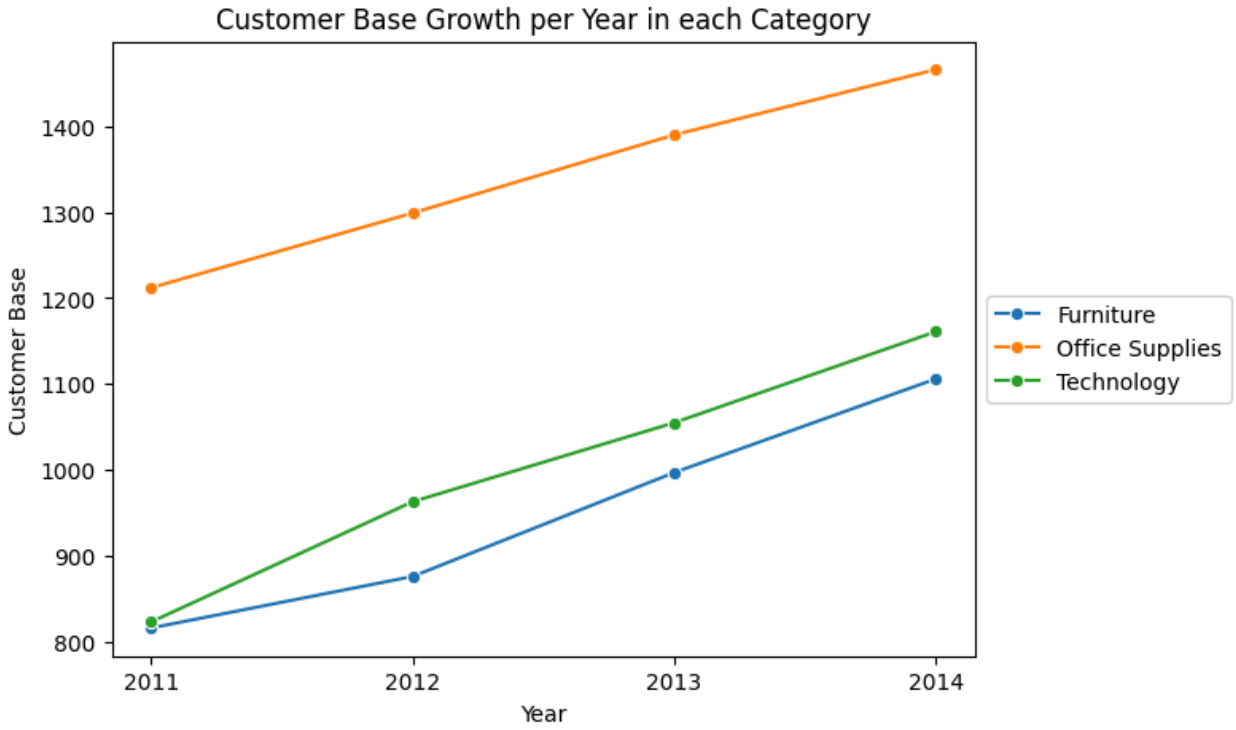
- The Customer Base Growth each year:



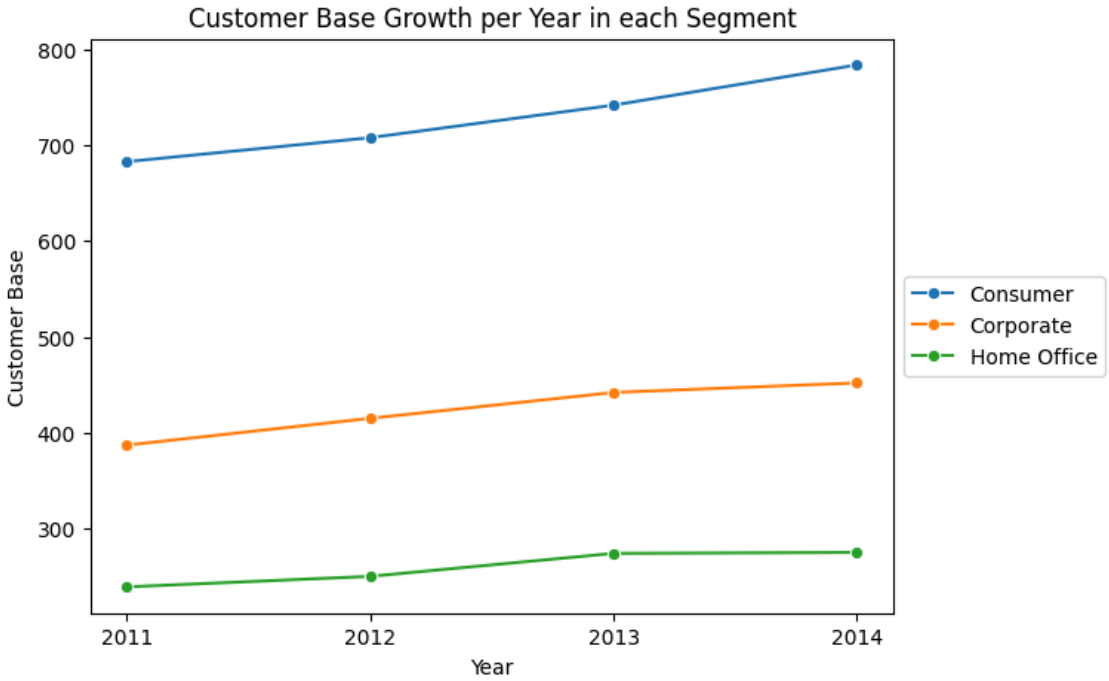
- The Customer Base Growth per year in each Market:



- The Customer Base Growth per year in each Category:

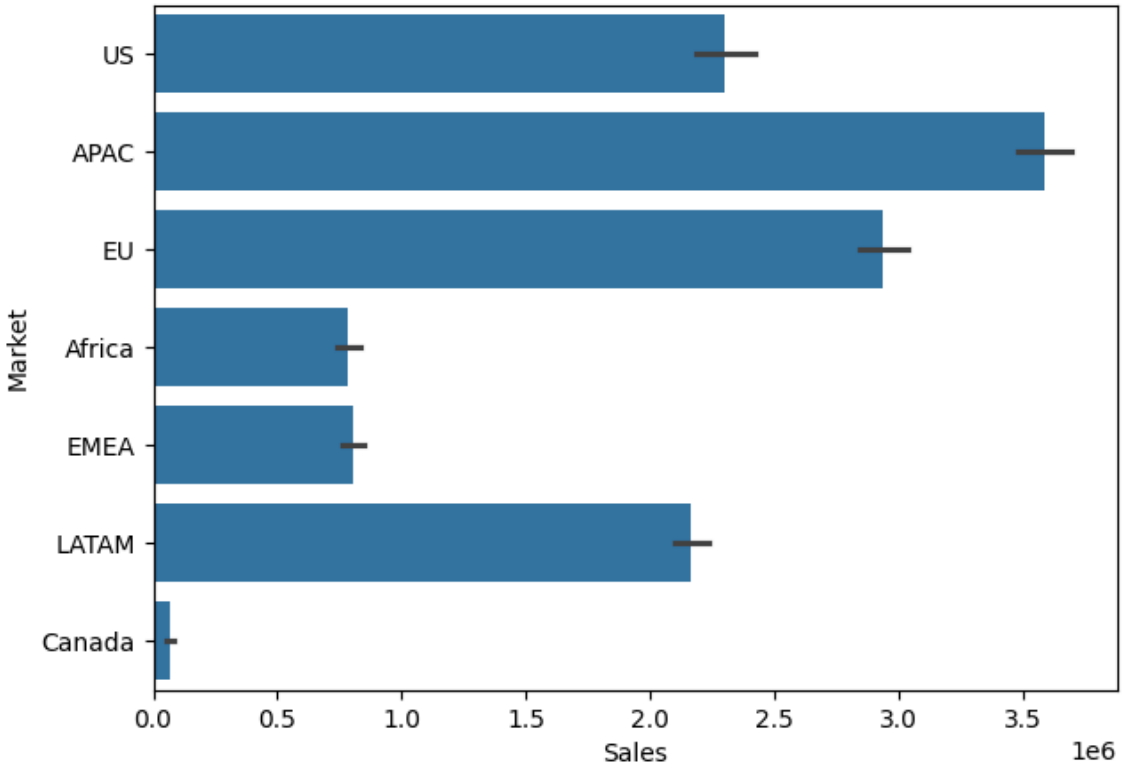


- The Customer Base Growth per year in each Segment:

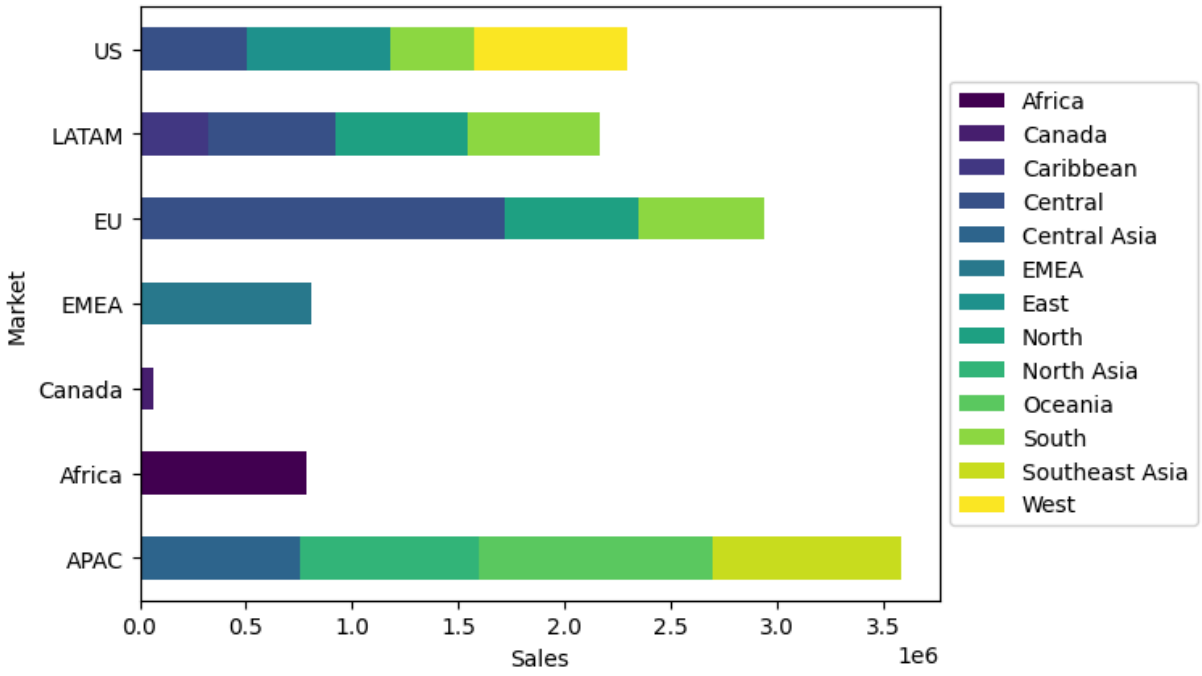


Sales and Profit Analysis

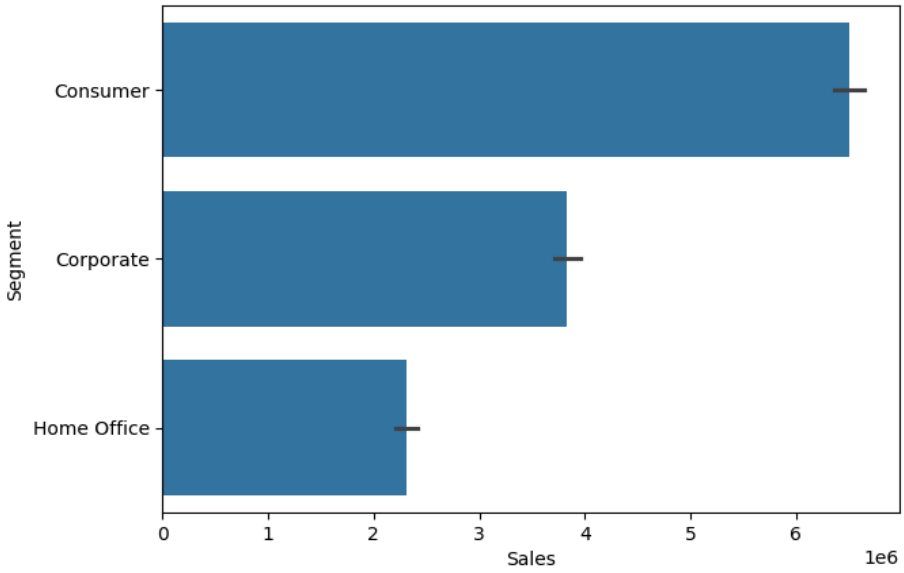
- Sales Analysis
    - Based on Market: The total sales in millions for each market is:



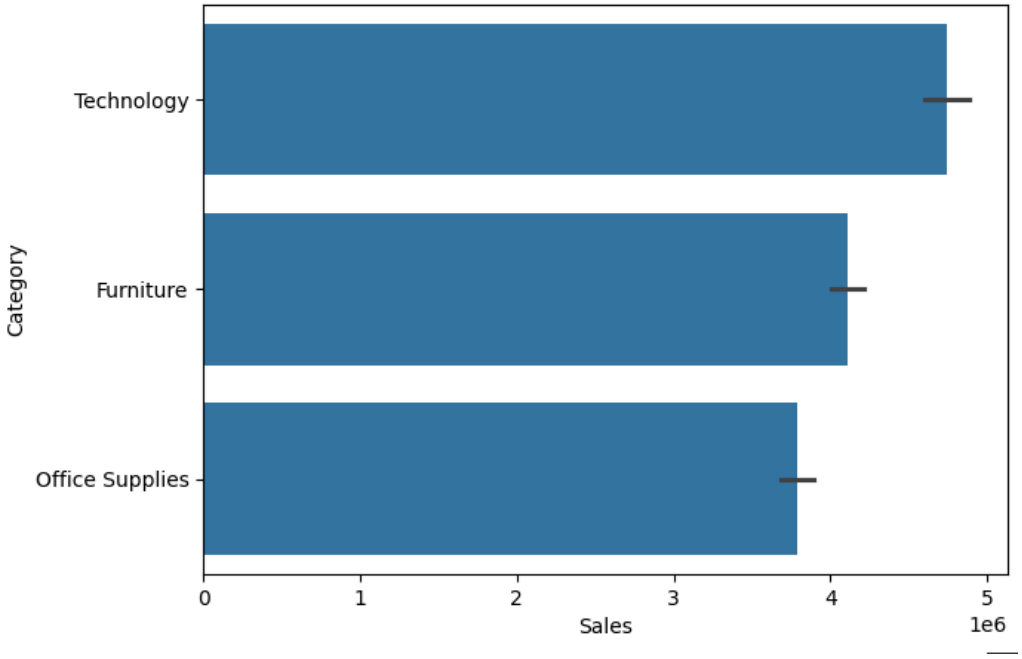
- Based on Different Regions of markets: The total sales in millions for each region in different market is:



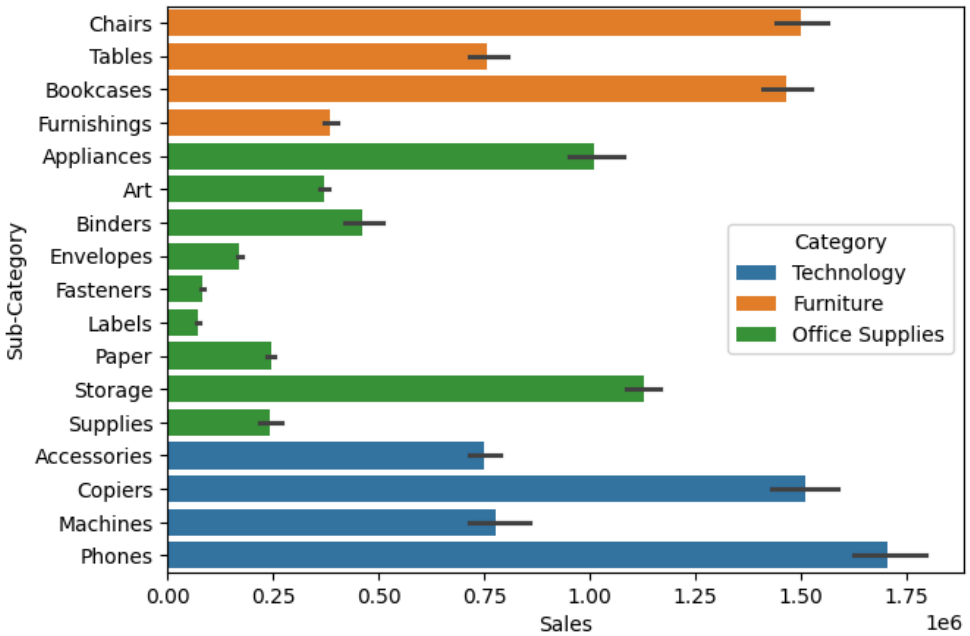
- Based on Segment: The total sales in millions for each segment is:



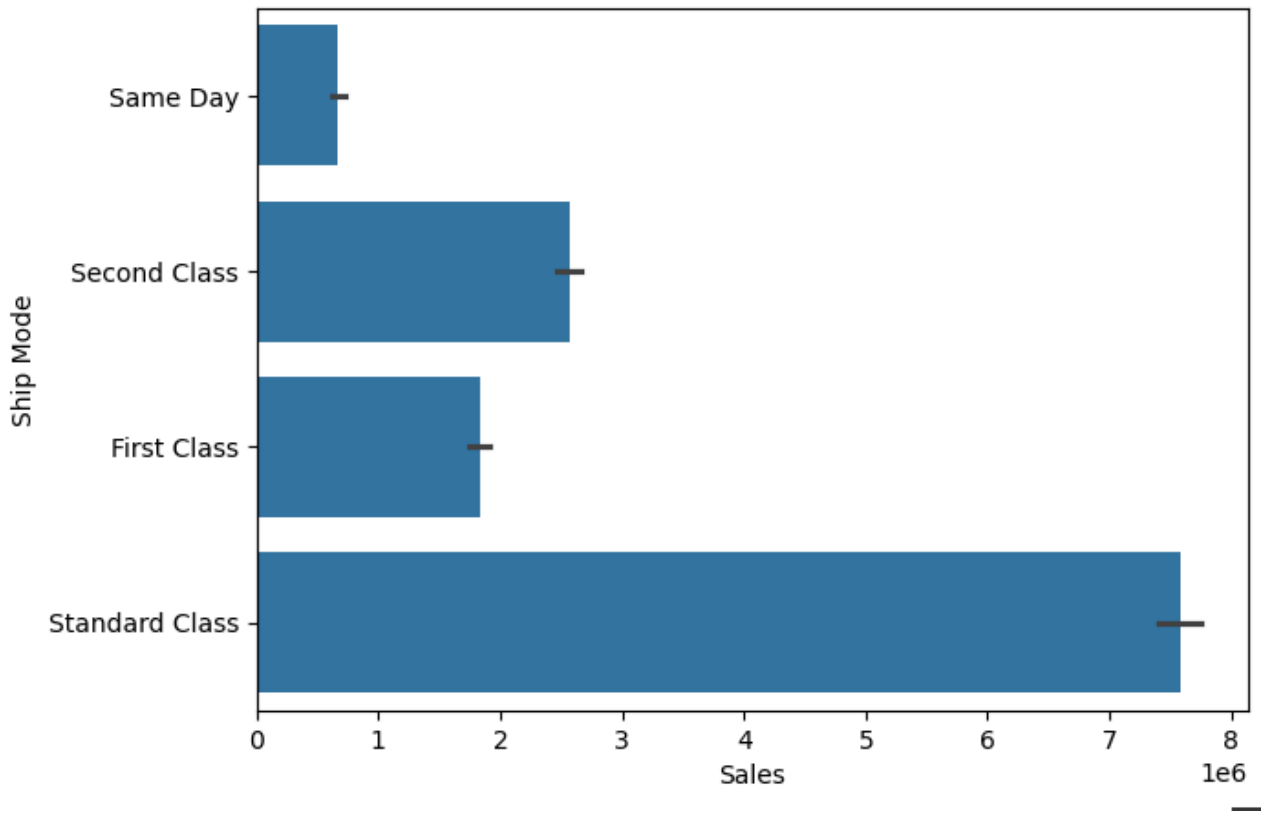
- Based on Category: The total sales in millions for each category is:



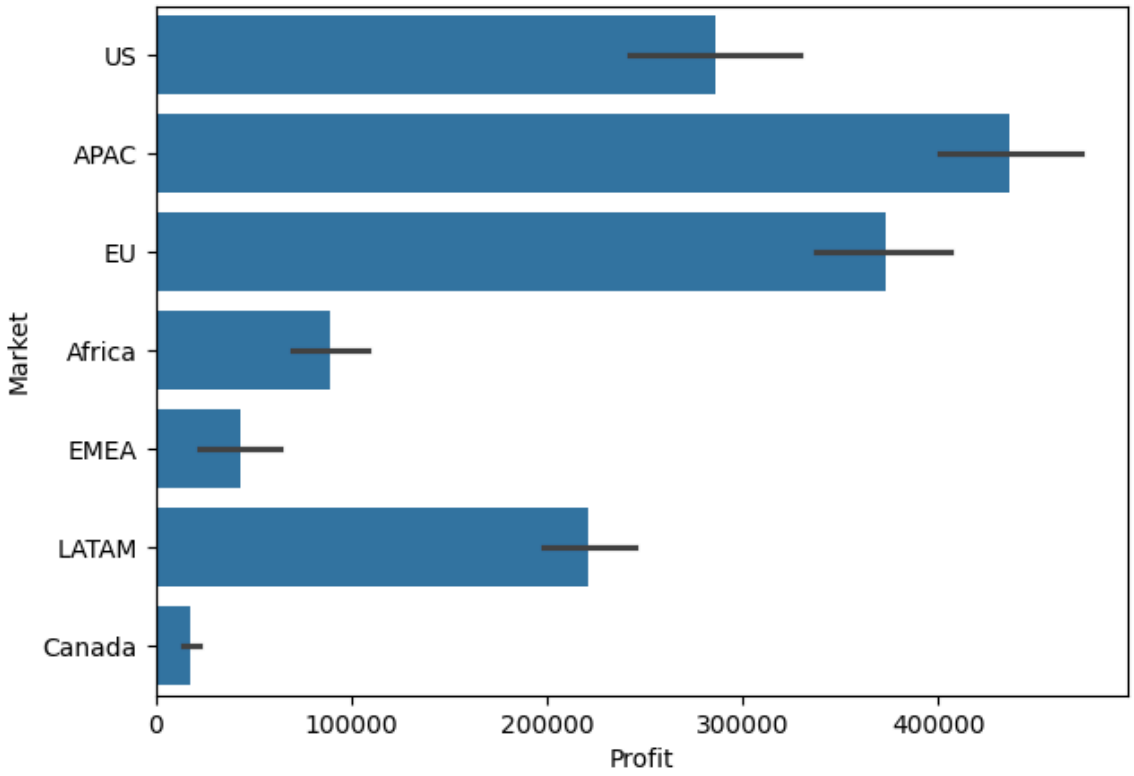
- Based on Sub-Category: The total sales in millions for each Sub-Category is:



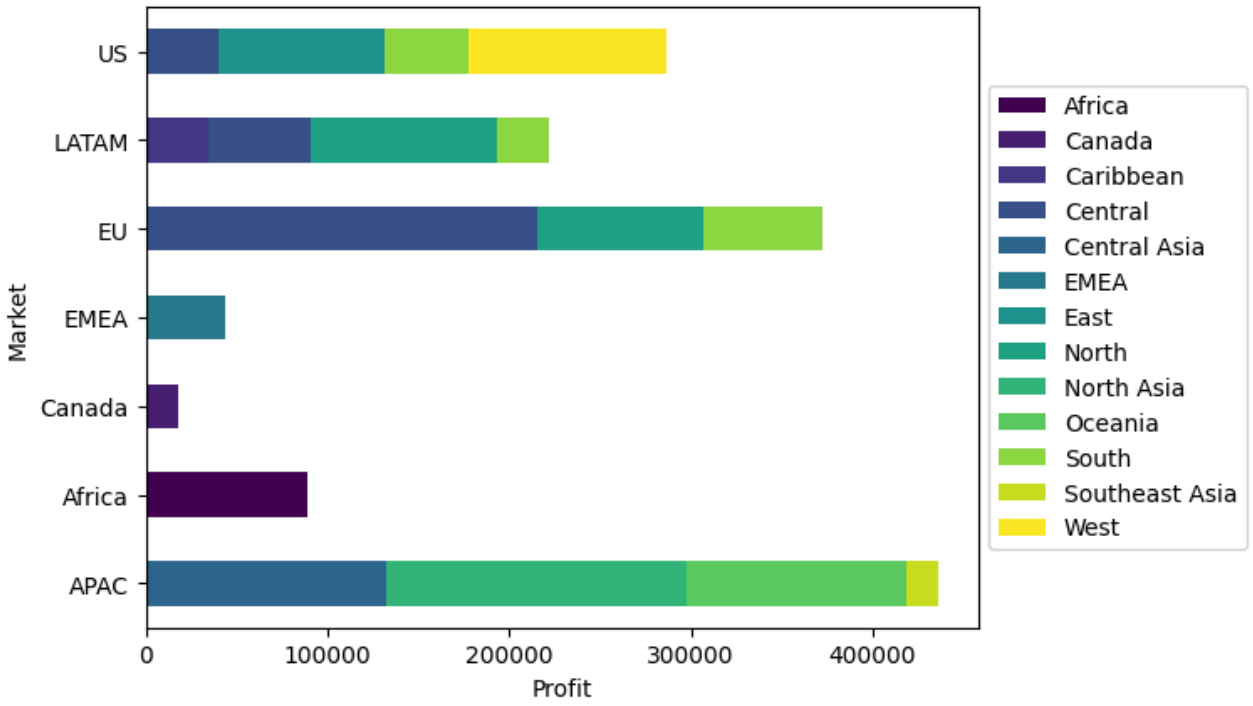
- Based on Ship Mode: The total sales in millions by each Ship Mode is:



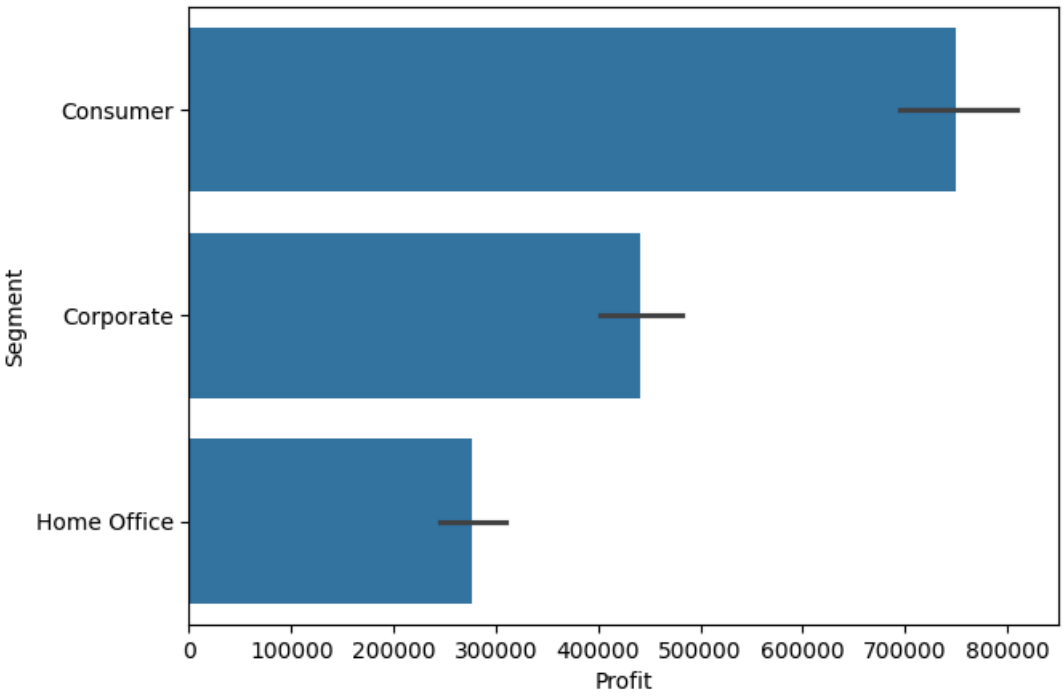
- Profit Analysis:
    - Based on Market: The total profit by each Market is:



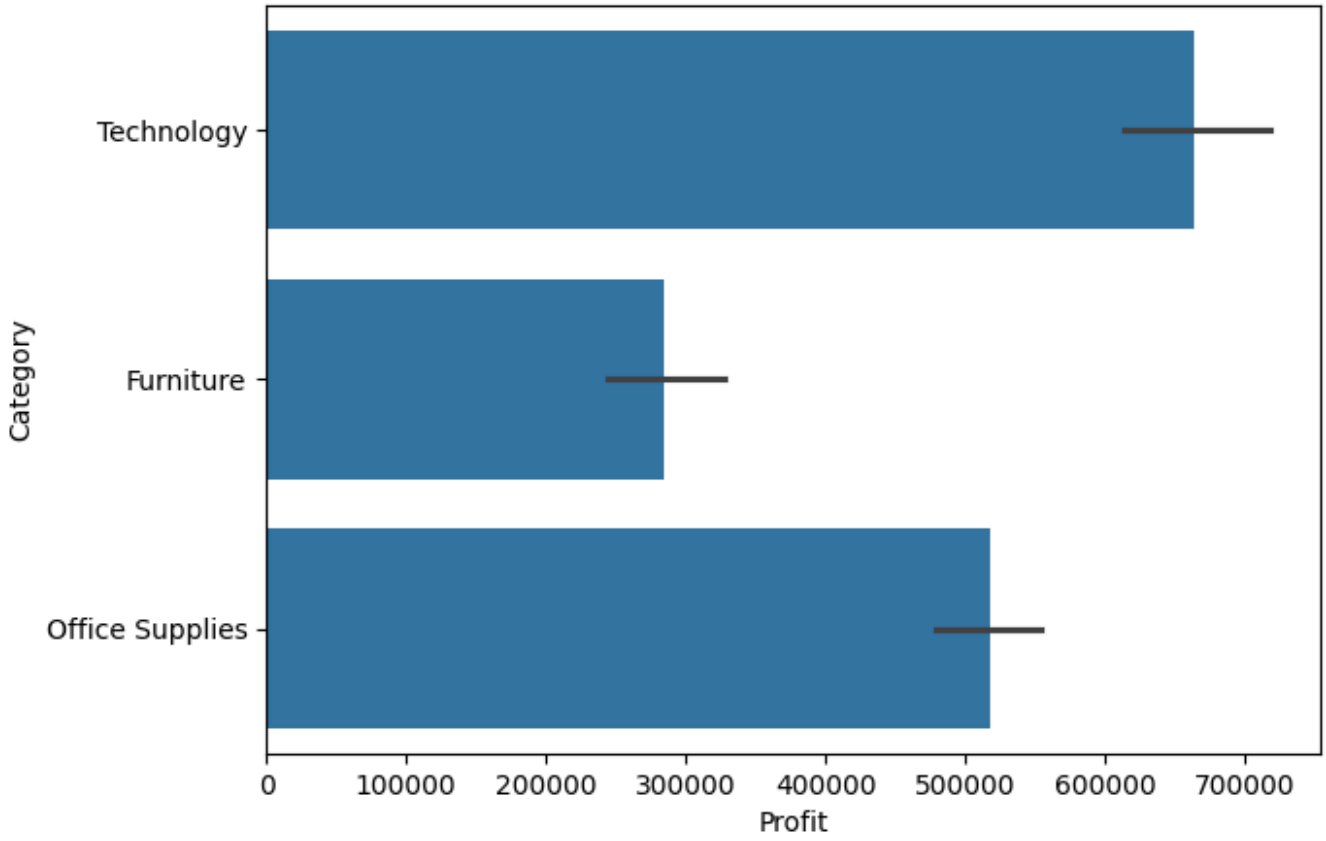
- Based on Different Regions in Market: The total profit of each region of different Market is:



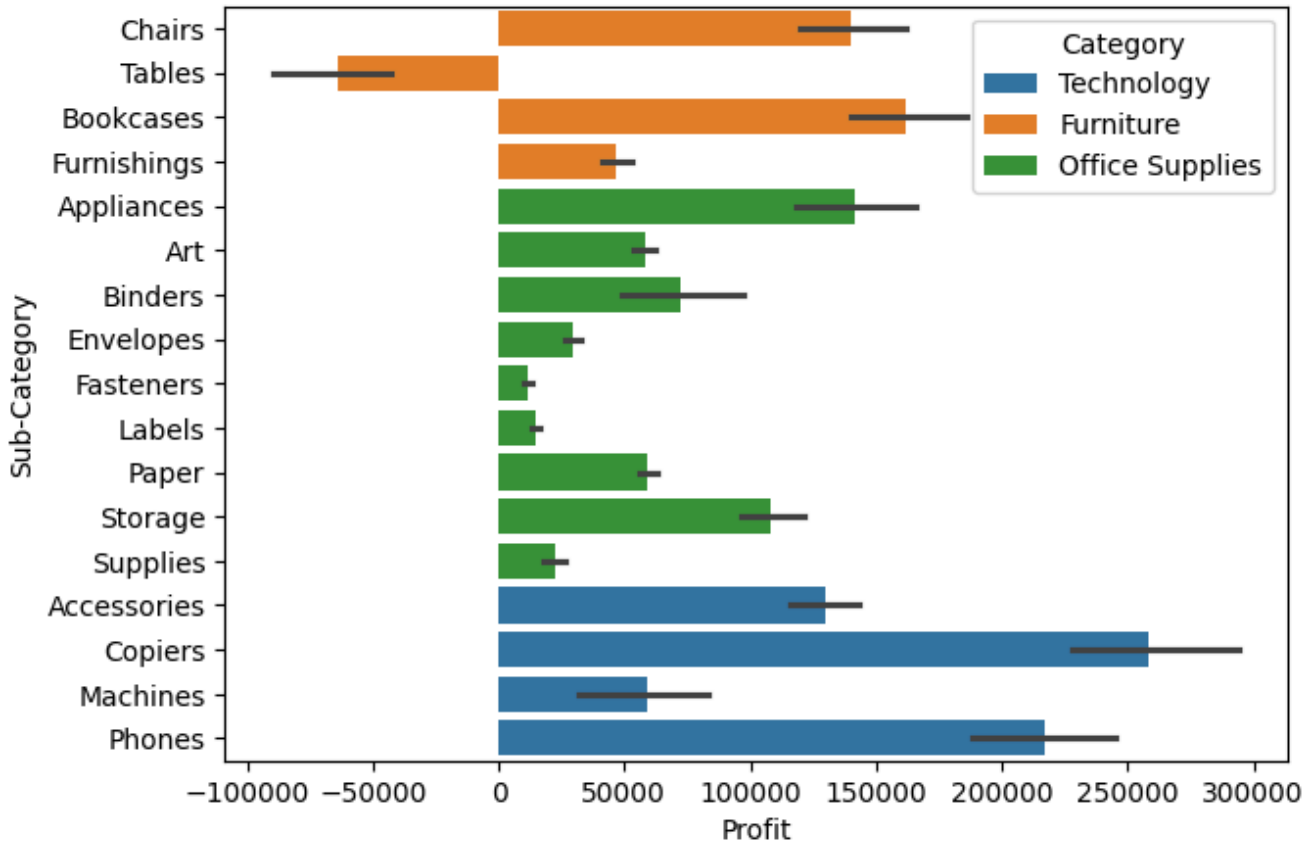
- Based on Segment: The total profit of each Segment is:



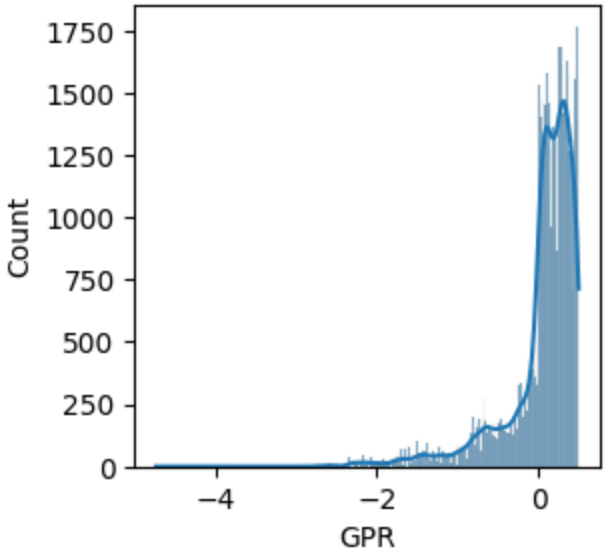
- Based on Category: The total profit of each Category is:

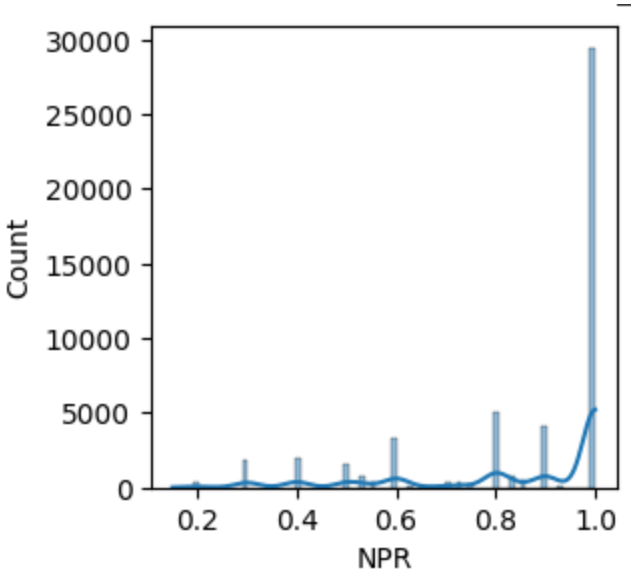


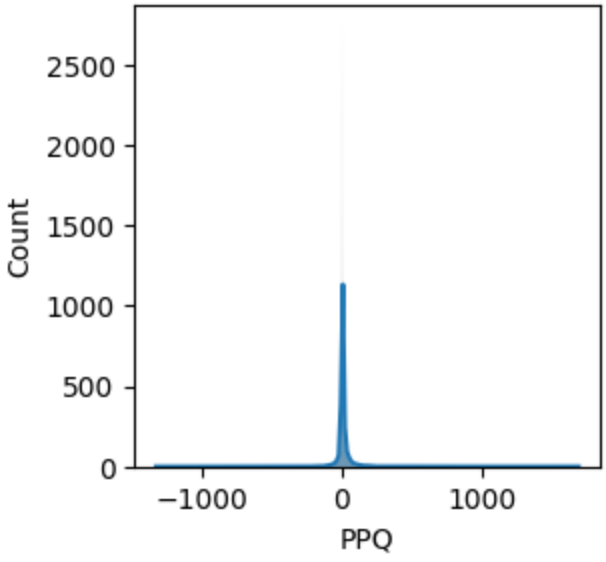
- Based on Sub-Category: The total profit of each Sub-Category is:

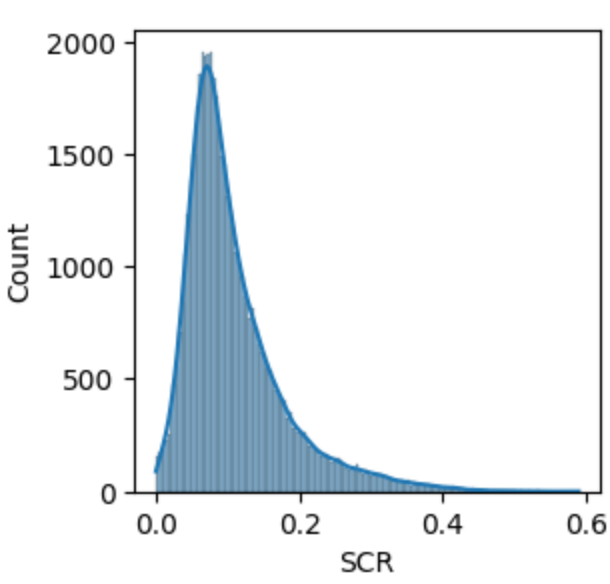


Financial Ratios of full dataset:









Implementation of Financial Ratios:

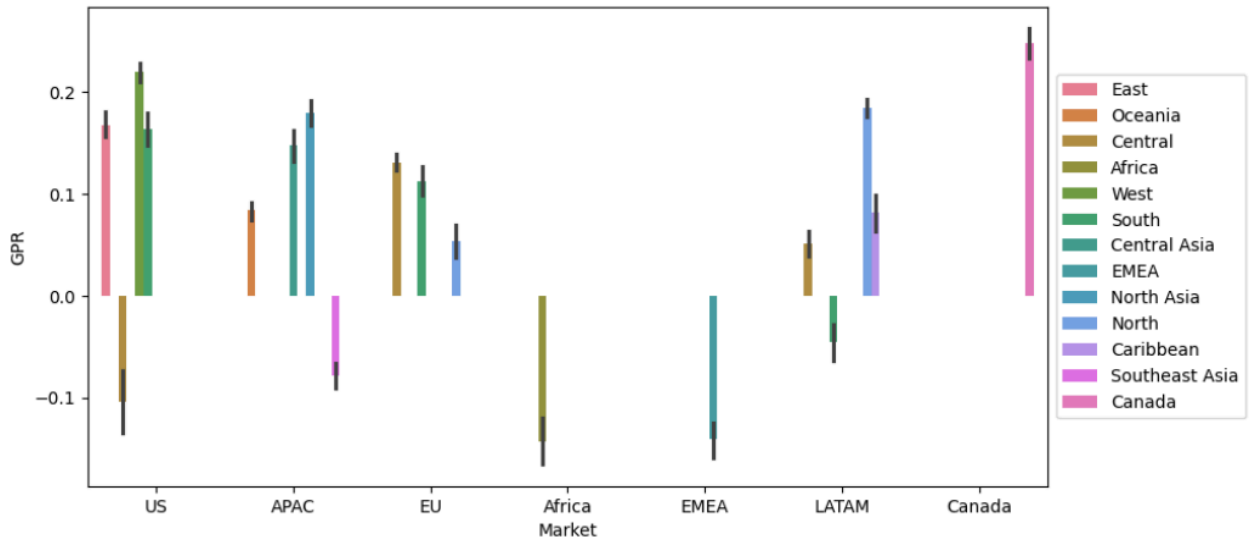
Gross Profit Margin (Profit/Sales) (GPR):

Gross Profit Margin (GPR) is a financial metric that expresses the profitability of a company's core business activities by measuring the percentage of revenue that exceeds the cost of goods sold (COGS). It is a key indicator of a company's operational efficiency and profitability at the gross profit level.

$$GPR = (\text{Gross Profit}/\text{Sales}) \times 100$$

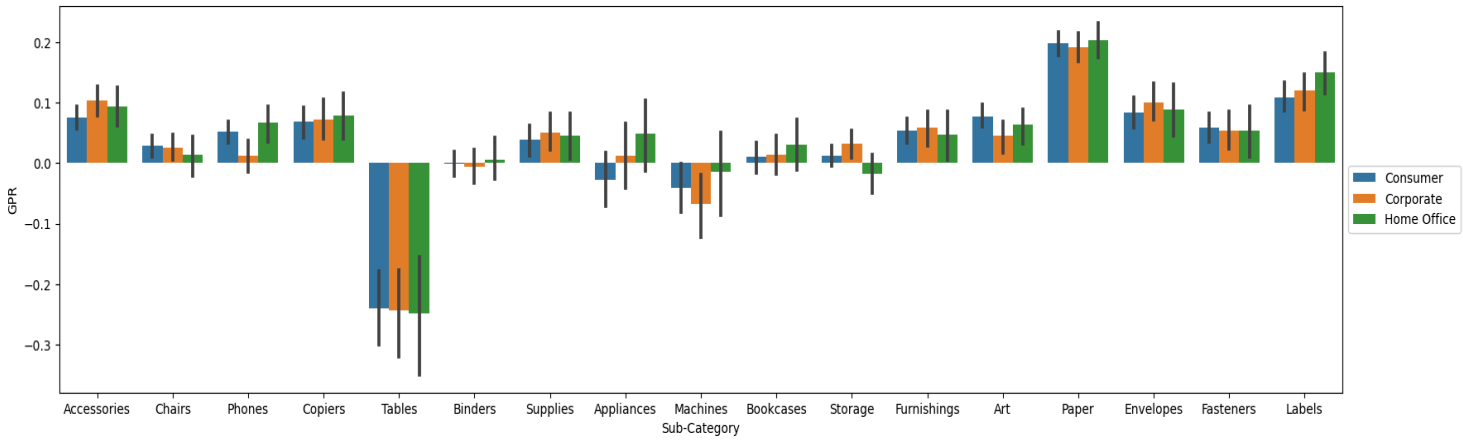
- **GPR(Market):**

  The plot shows the comparison of GPR of different markets based on regions.



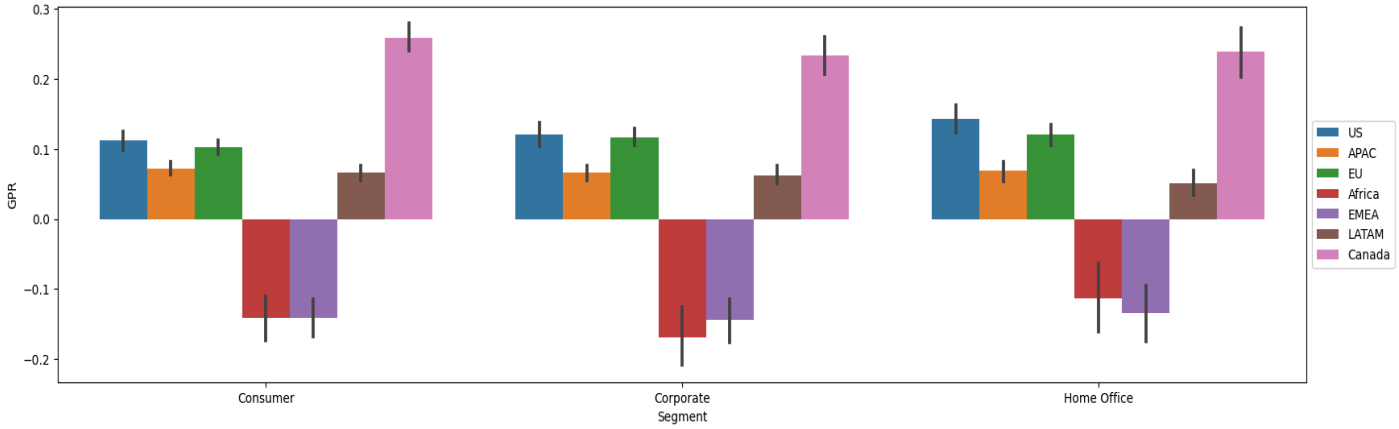
- **GPR(Sub-Category):**

  This plot shows the comparison of GPR of different segments of consumers, based on different sub-categories.



- **GPR(Segment- Market Wise):**

  This plot shows the comparison of GPR for different markets, based on different segments.



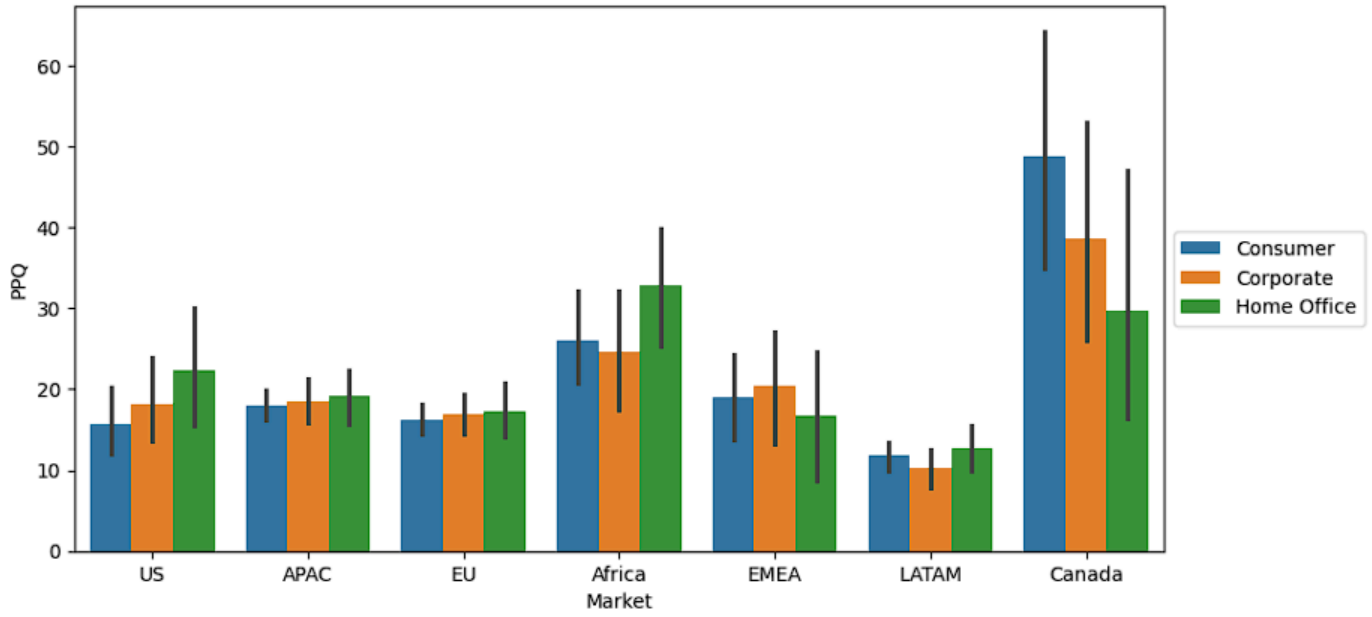
Profit Per Quantity (PPQ) :

Profit Per Quantity (PPQ) is a financial metric that measures the profit generated by a business for each unit of quantity sold. It is particularly relevant in industries where the unit of quantity can vary, such as retail, manufacturing, or services. The formula for calculating Profit Per Quantity is:

PPQ=Profit/Quantity

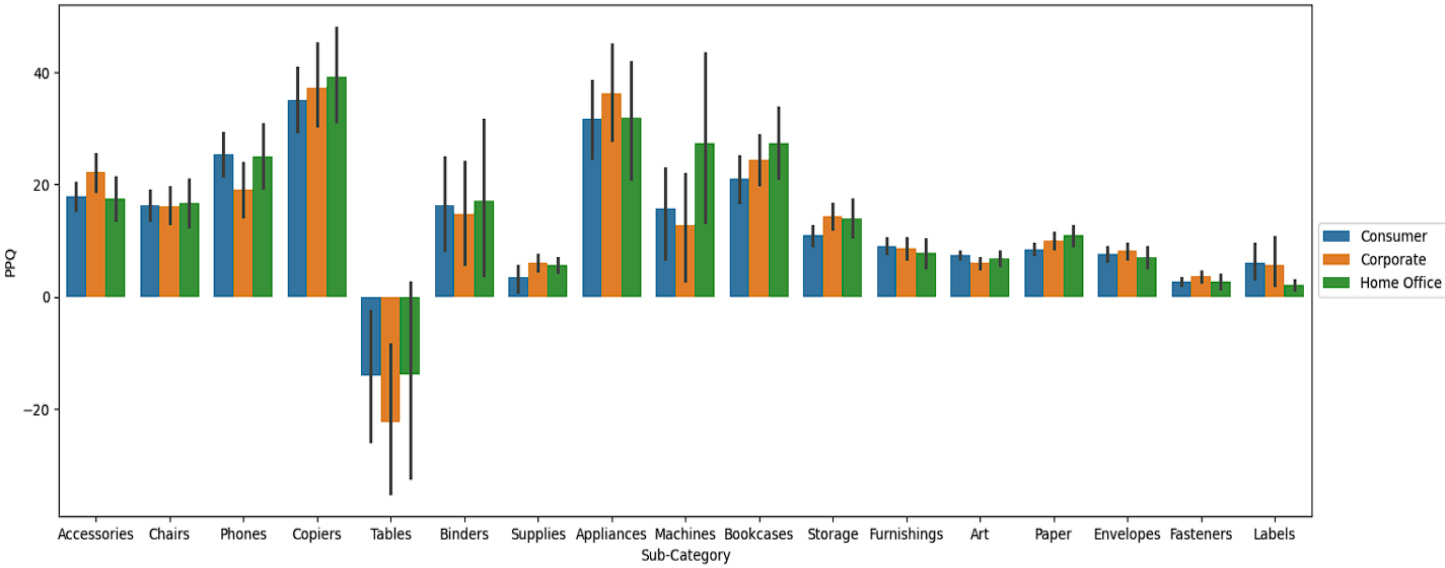
- **Market:**

  The plot shows the comparison of GPR of different markets based on segments.



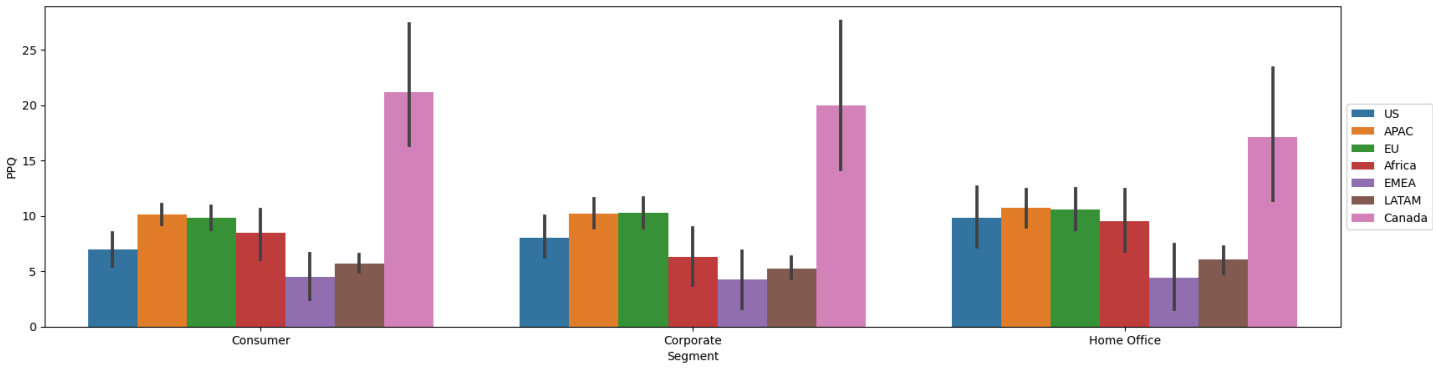
- **Sub-Category:**

  This plot shows the comparison of PPQ of different subcategories of goods, based on different segments.



- **PPQ(Segment- Market Wise):**

  This plot shows the comparison of PPQ for different markets, based on different segments.



NPR - Net Profit Ratio:

Net Profit Ratio (NPR) is a financial ratio that measures the profitability of a business by expressing the net profit as a percentage of total revenue.

NPR=(Total Revenue/Net Profit)×100

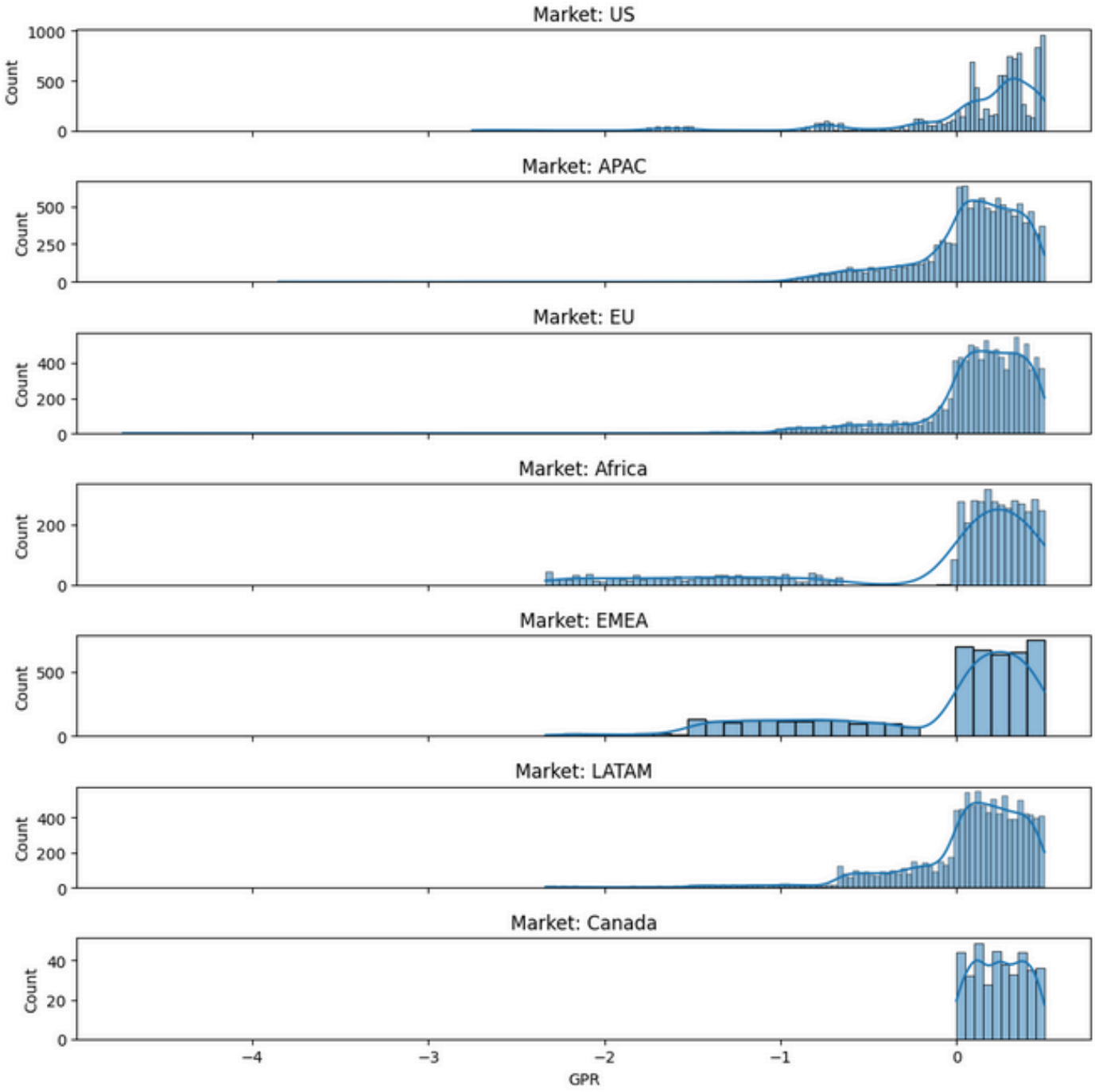
SCR - Shipment Cost Ratio:

SCR Shipment Cost Ratio could potentially refer to a ratio that assesses the relationship between shipment costs and Supply Chain Resilience (SCR).

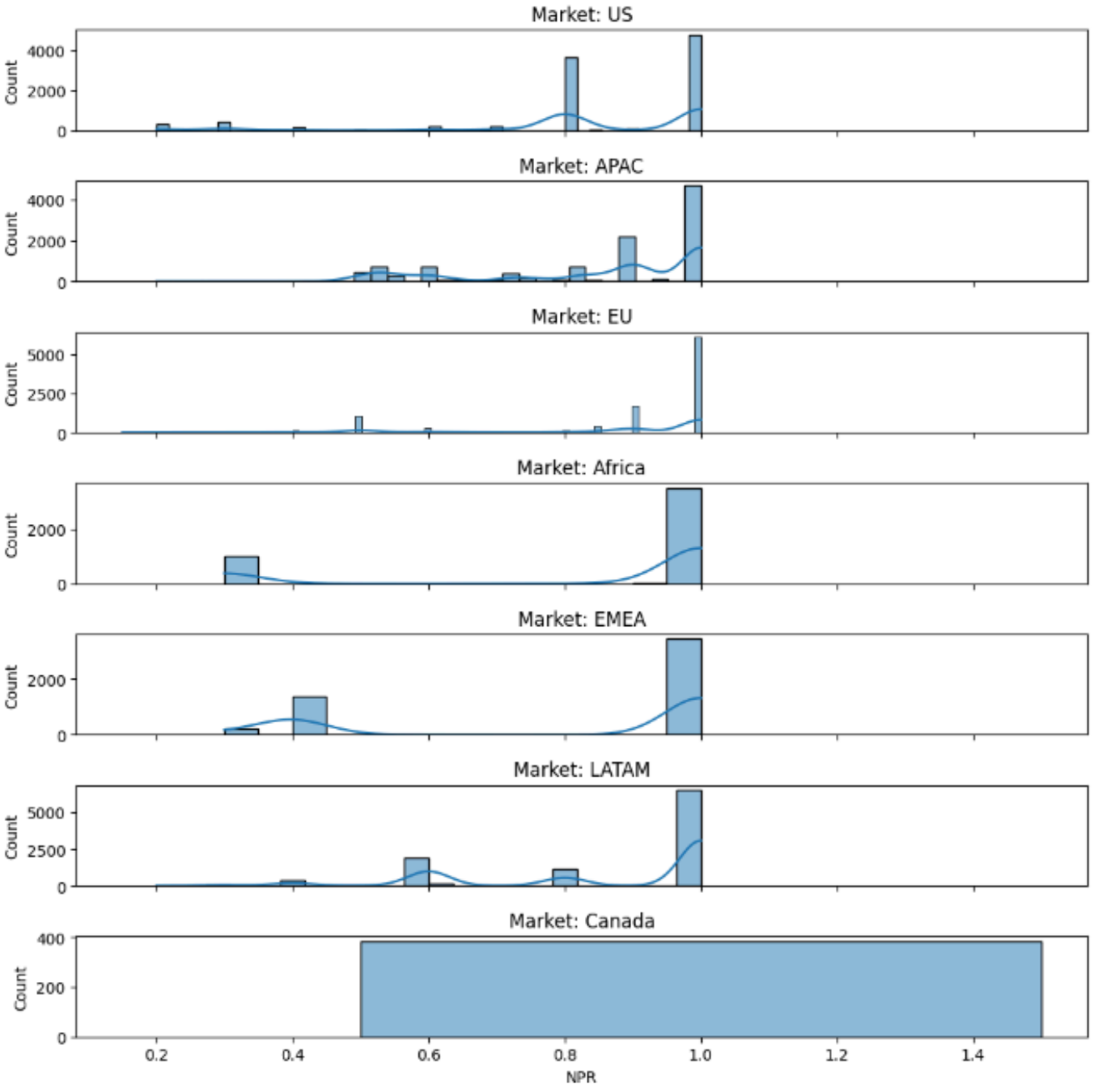
SCR Shipment Cost Ratio=(Total Shipment CostsSupply /Chain Resilience Metrics)×100

Market Analysis using financial ratios:

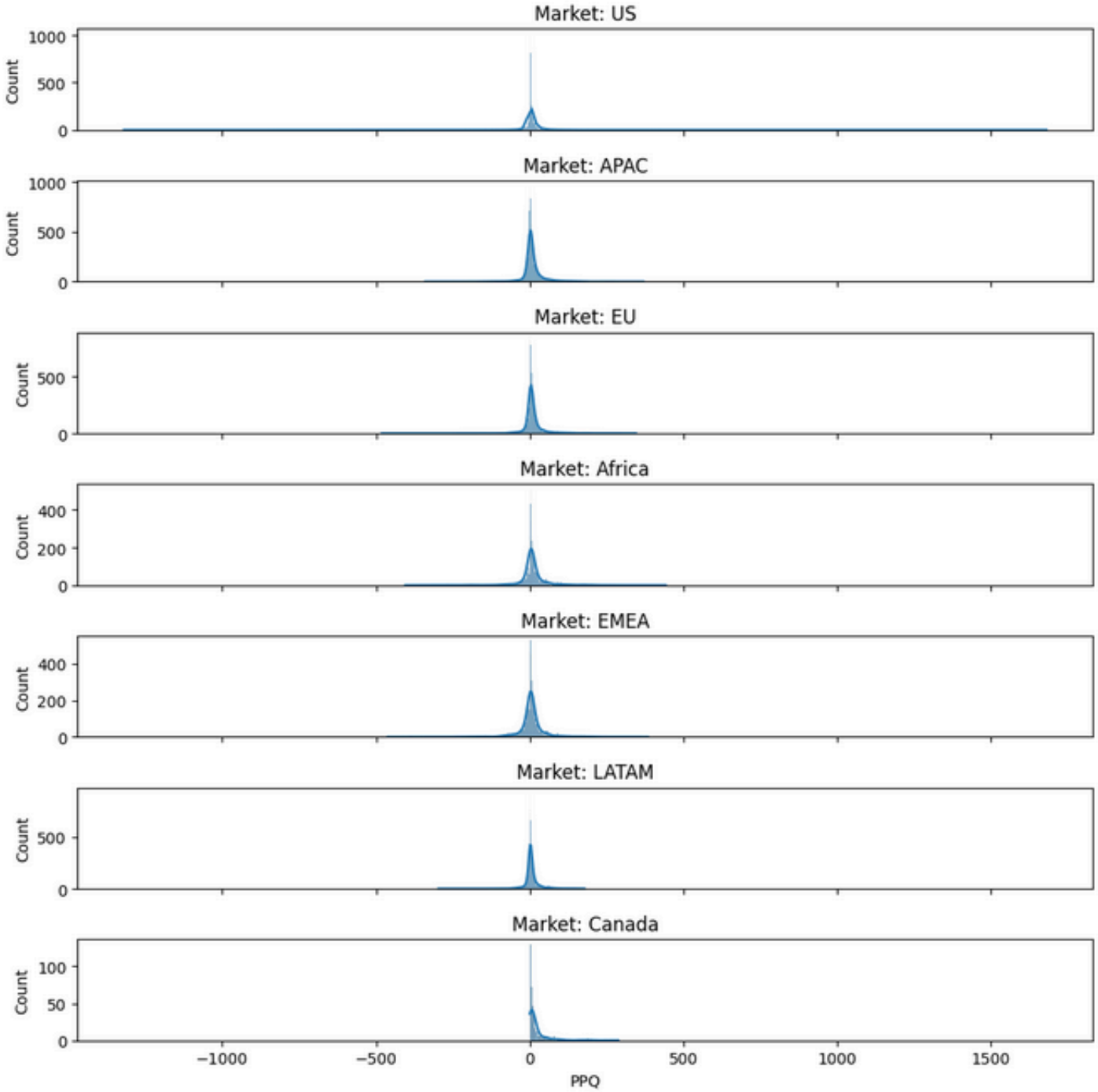
Distribution of GPR in different Markets



Distribution of NPR in different Markets:

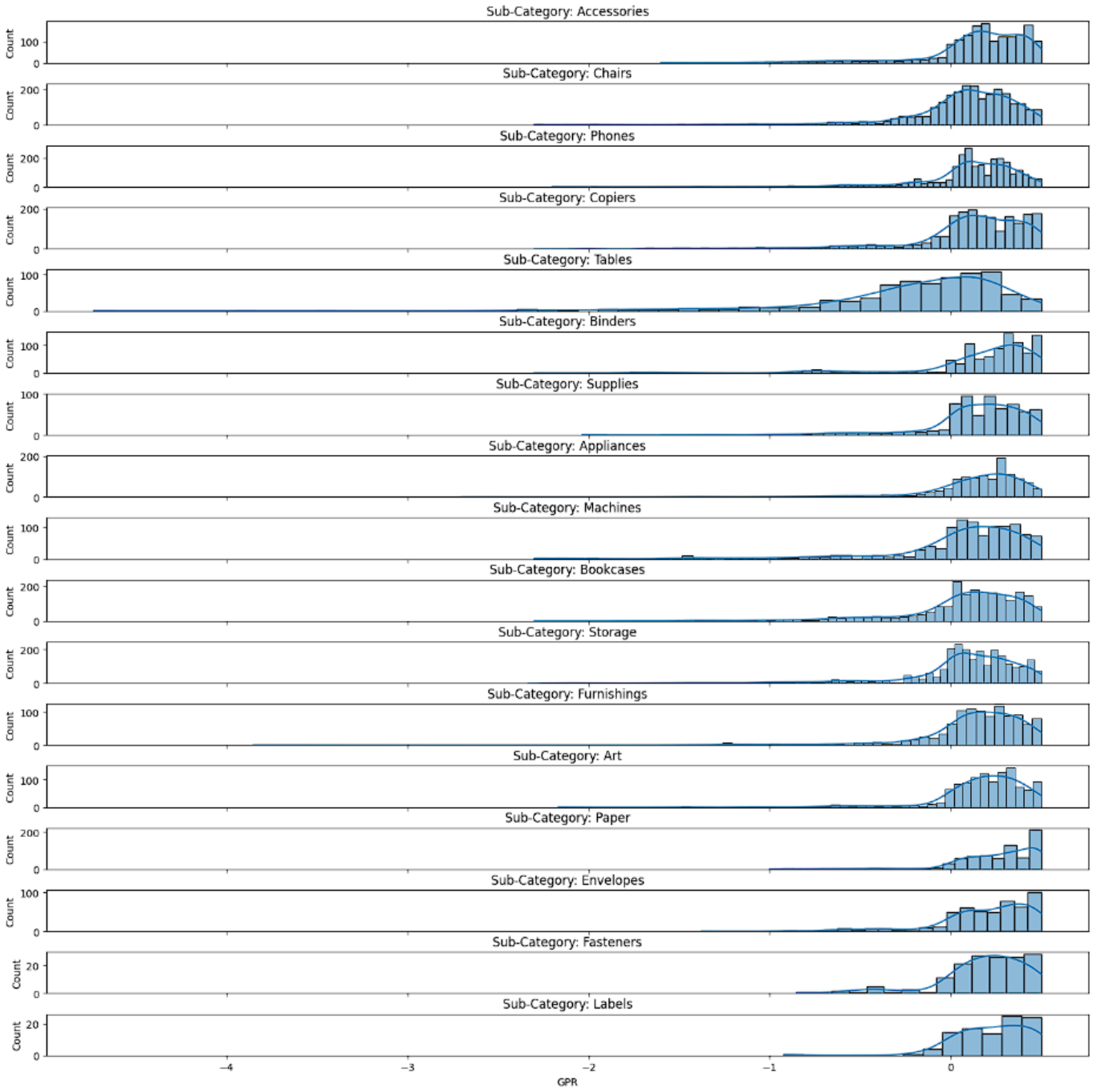


Distribution of PPQ in different Markets:

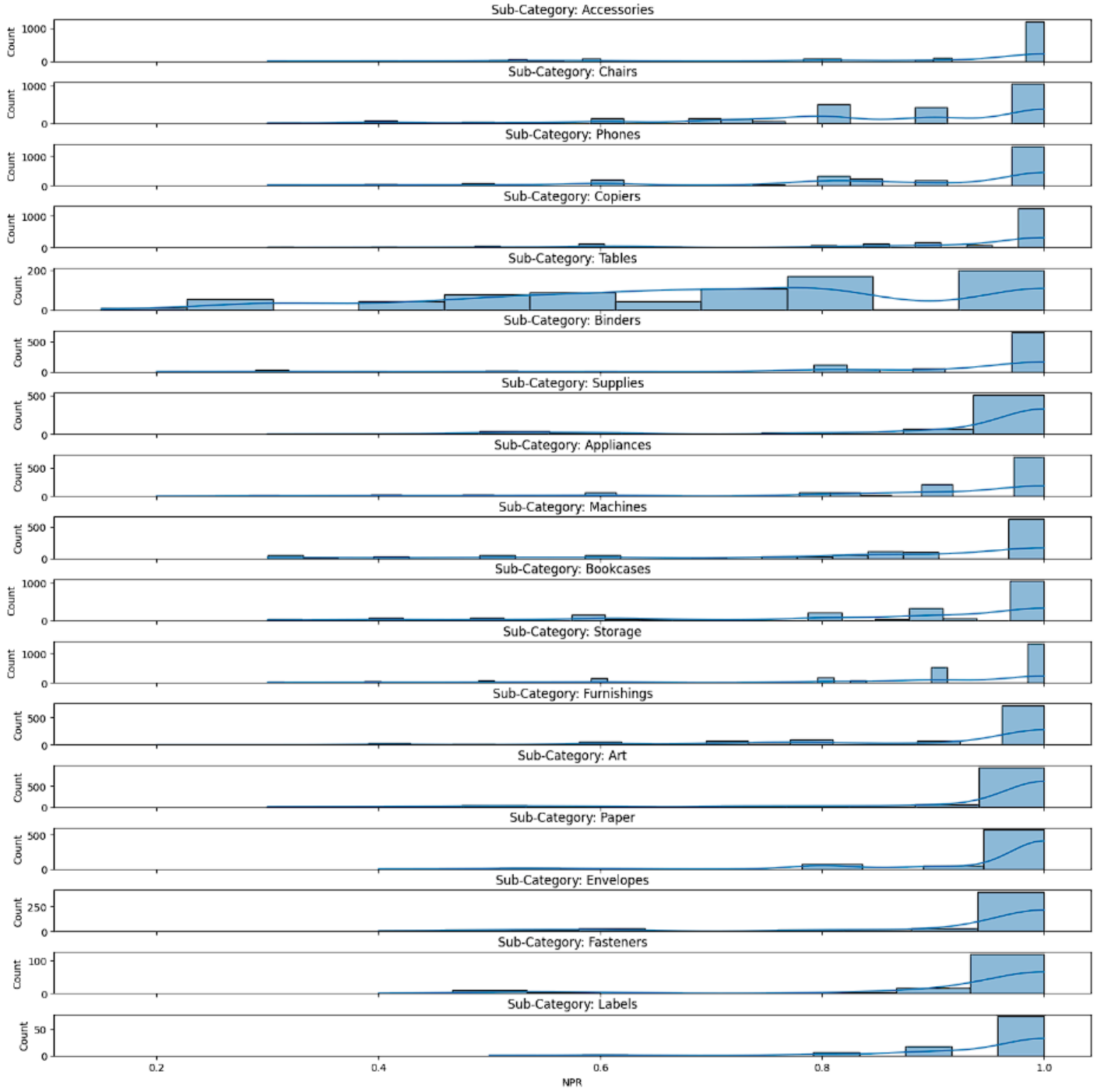


Sub-Categories Analysis using financial ratios:

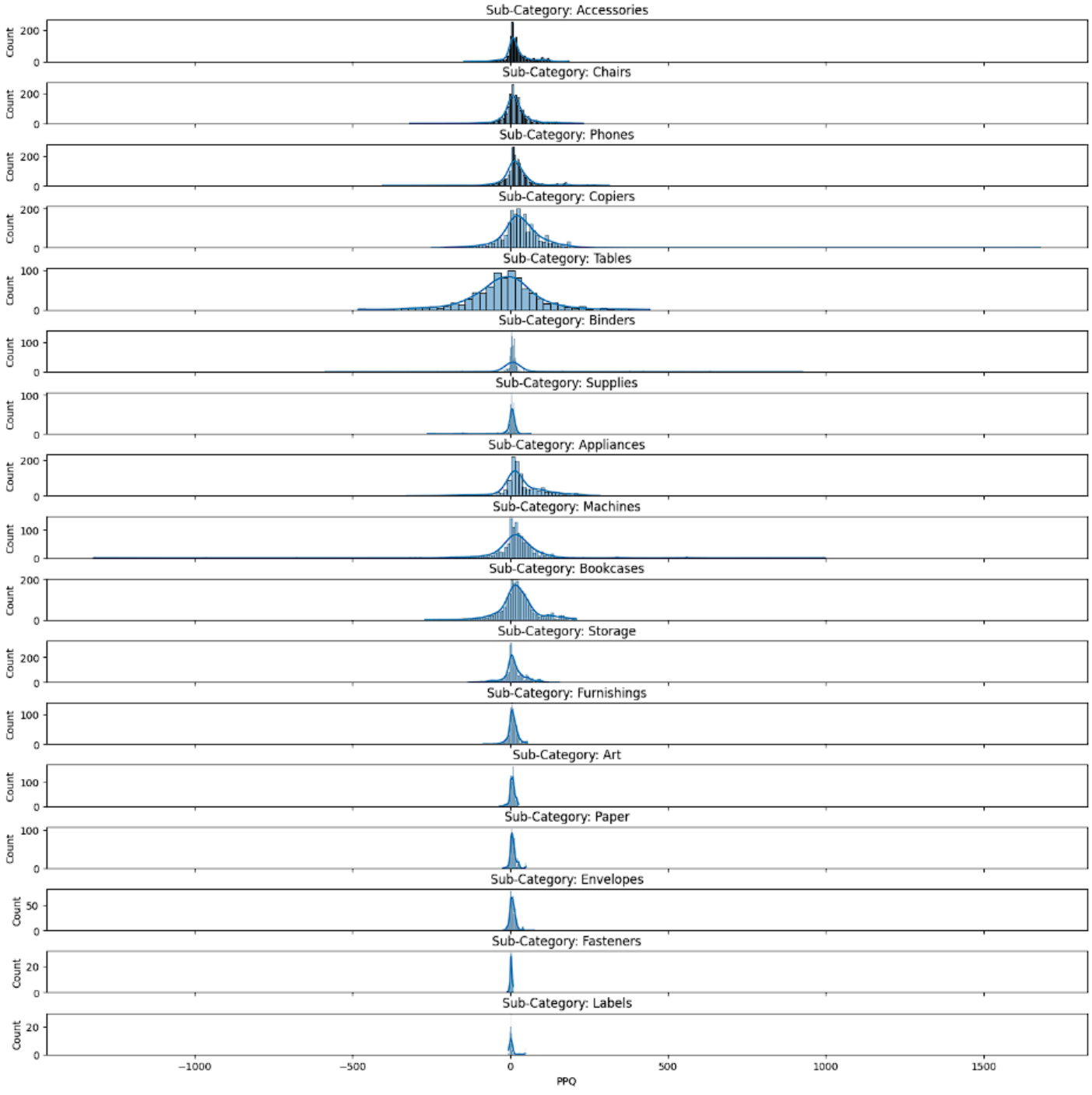
GPR of Sub-categories:



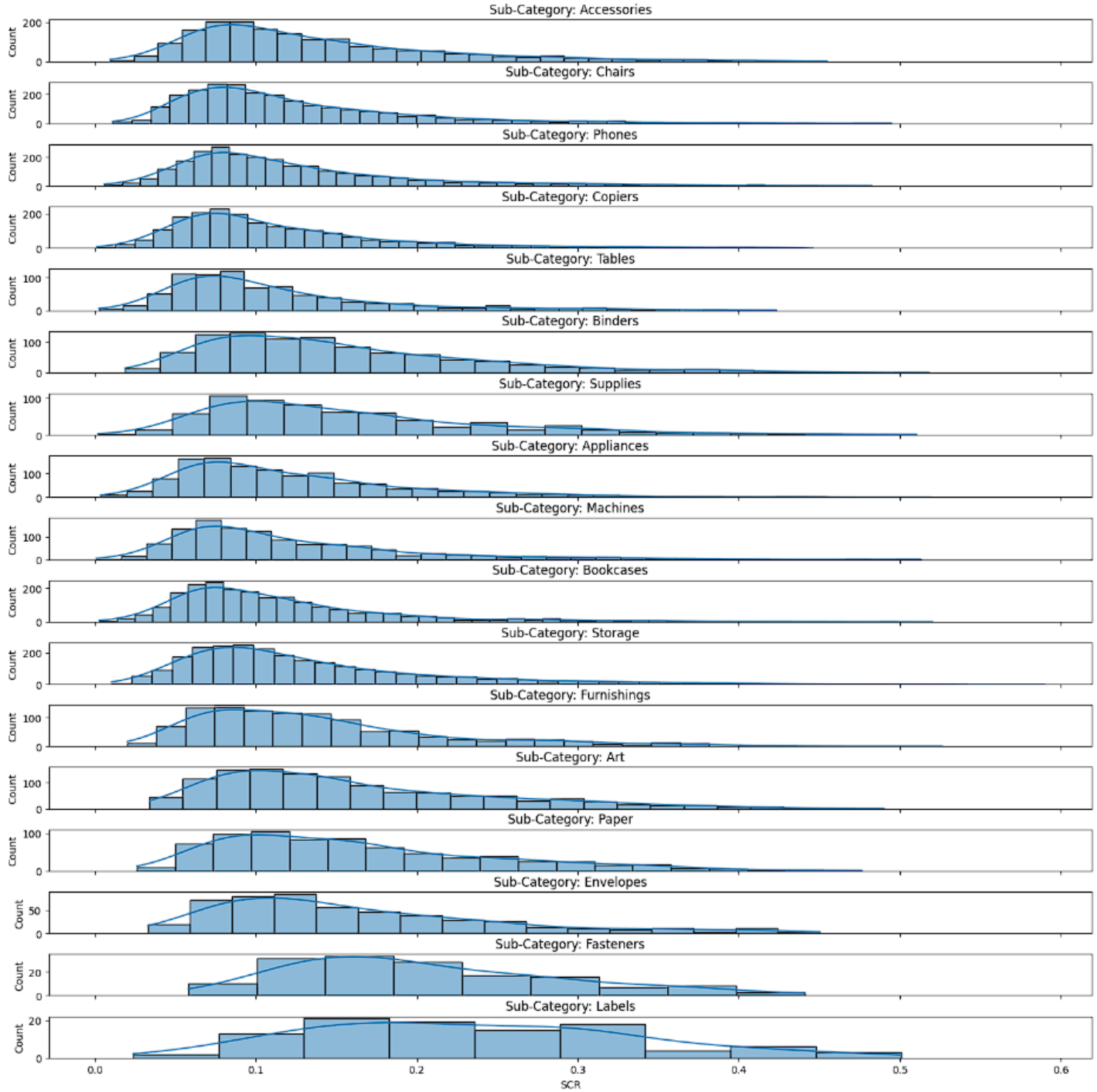
NPR of Sub-categories:



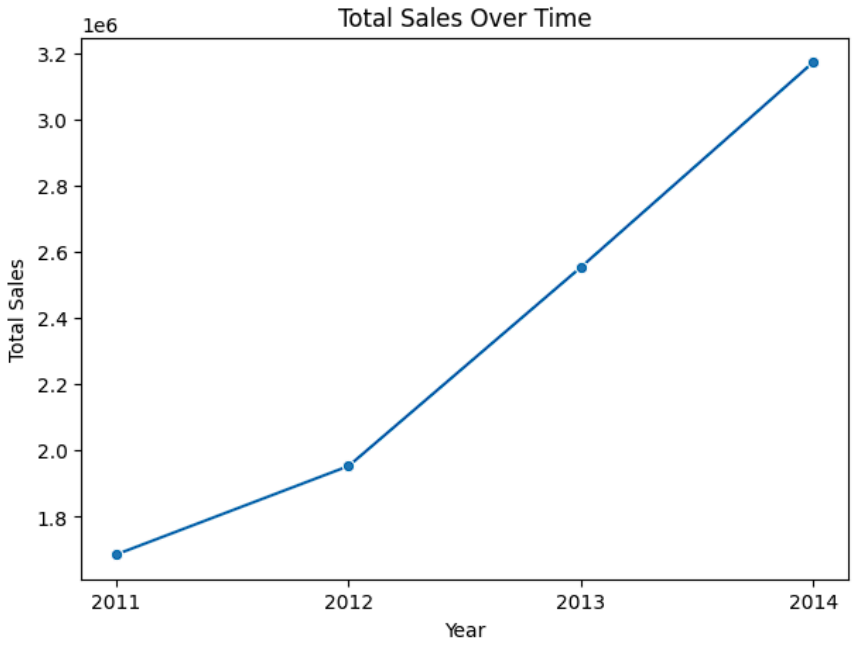
PPQ of Sub-categories:

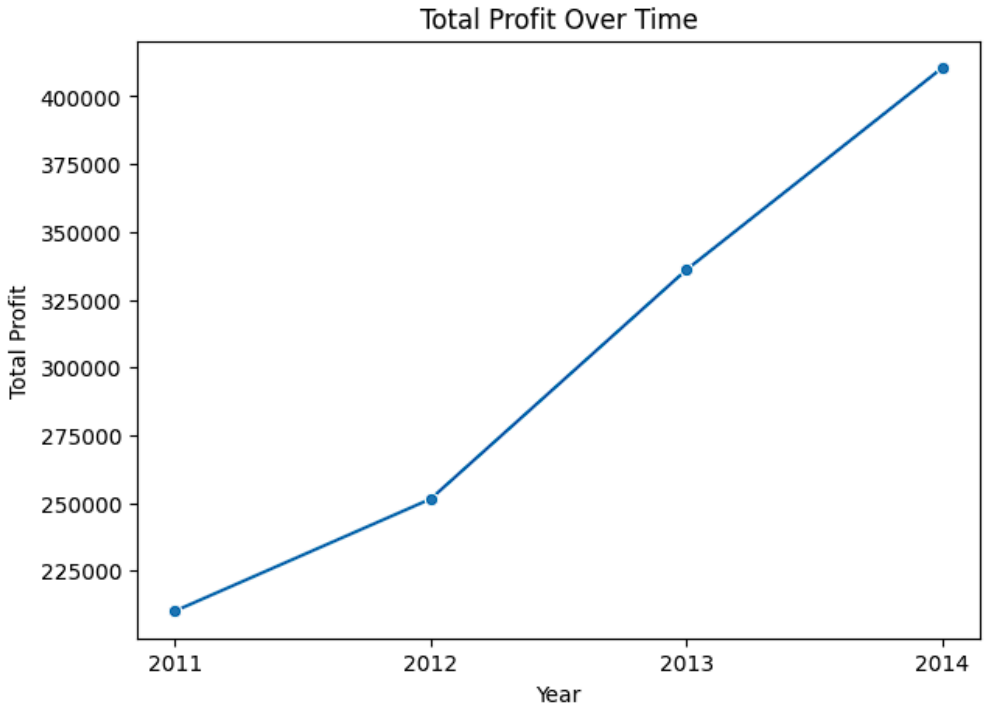


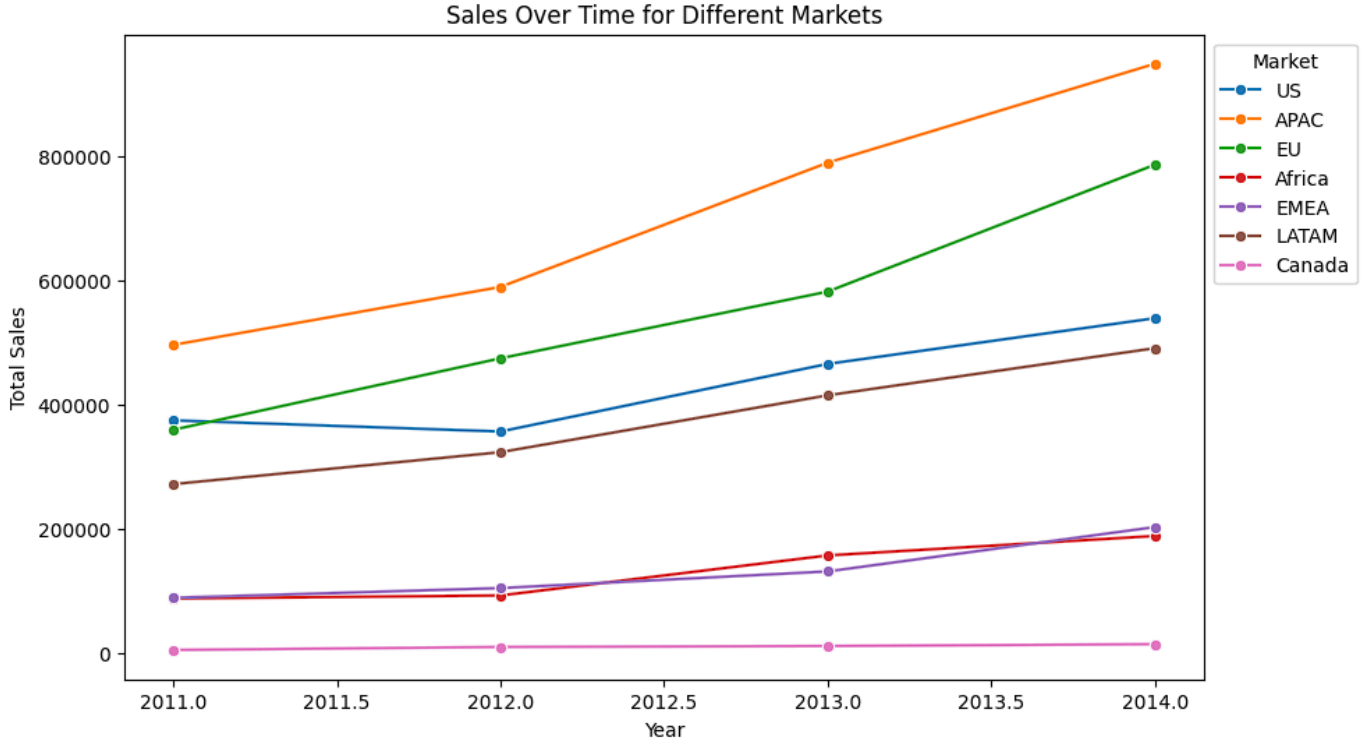
SCR of Sub-categories:

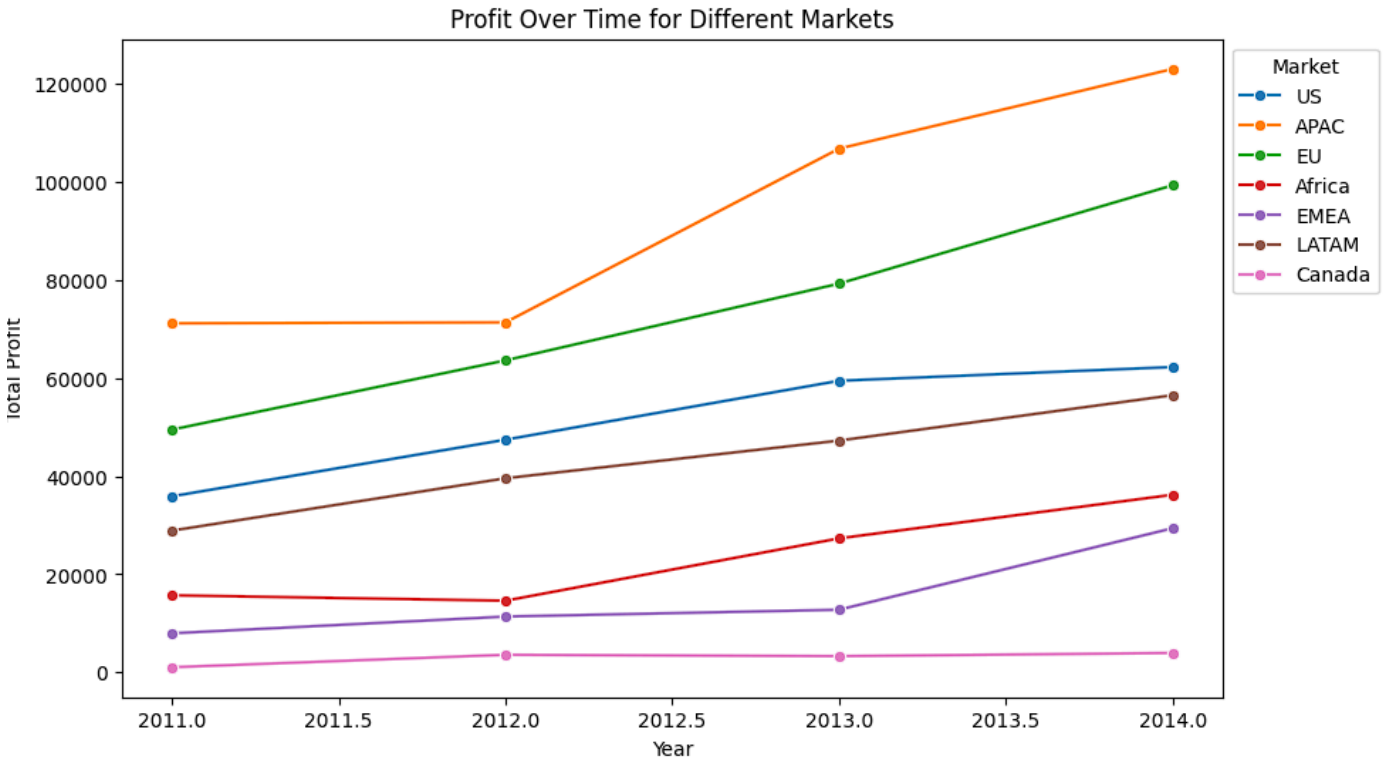


Yearly Sales and Profit Analysis






Sales Over Time for Different Markets
Total Sales
Year
Market
US
APAC
EU
Africa
EMEA
LATAM
Canada
0
200000
400000
600000
800000
2011.0
2011.5
2012.0
2012.5
2013.0
2013.5
2014.0


Profit Over Time for Different Markets
Total Profit
Year
Market
US
APAC
EU
Africa
EMEA
LATAM
Canada
0
20000
40000
60000
80000
100000
120000
2011.0
2011.5
2012.0
2012.5
2013.0
2013.5
2014.0