

# **TASK 1: Exploratory Data Analysis (EDA) and Business Insights**

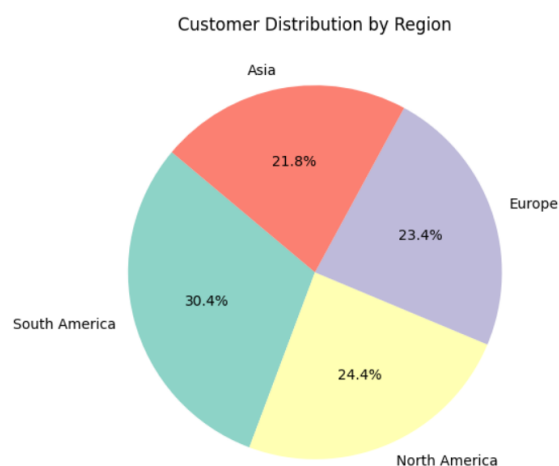
Three dataset files—Customers.csv, Products.csv, and Transactions.csv—are used in the provided report. Initially, exploratory data analysis is carried out, which involves examining each dataset for any missing values. The dataset then includes columns that have been transformed to desired formats, such as dates in a particular datetime format and price, quantity, and total value in a numerical format. After that, we combine the datasets and apply some preprocessing. We use visualisations to extract business insights following EDA. This examination of sales and customer data yields important information on product performance, consumer behaviour, and regional distributions. The following important conclusions have been drawn:

## 1. Sales by Category



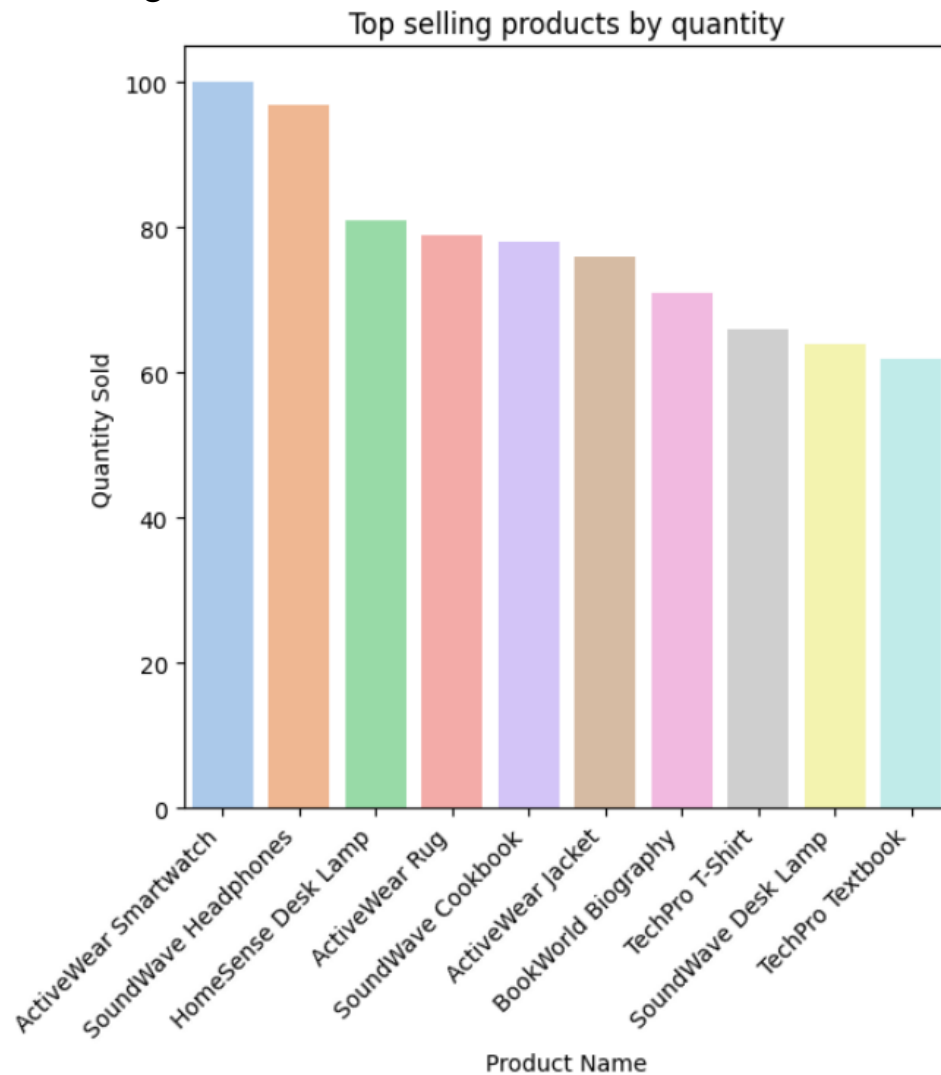
The Sales by Category insight shows that certain product categories dominate the sales. Categories with higher total sales may indicate customer preferences or trends toward specific types of products. The treemap visualization highlights the proportional contribution of each category to overall sales, helping identify the most profitable categories.

## 2. Customer Distribution by Region



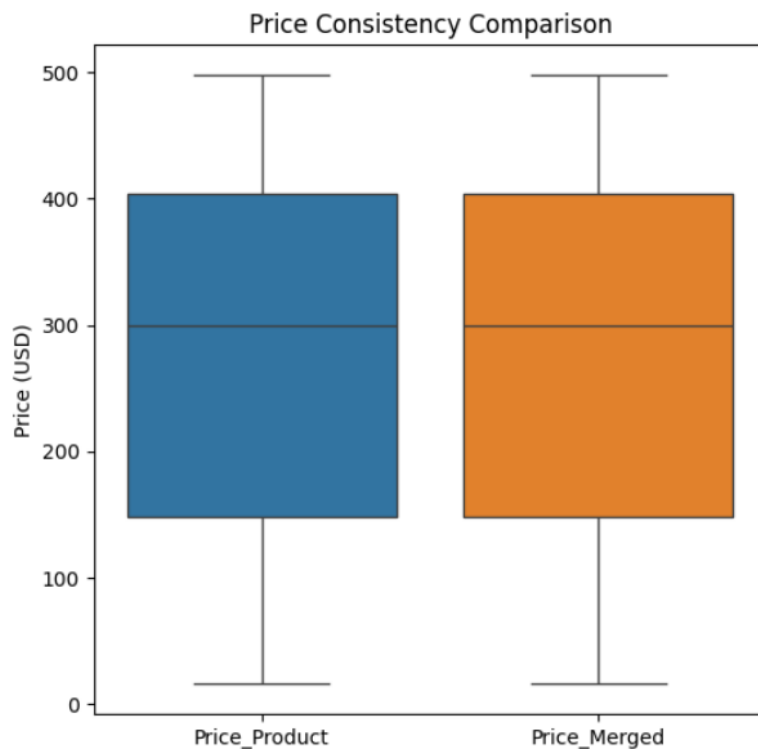
Geographical patterns in the customer base can be found by analysing the consumer distribution by region. South America accounts for 30.4% of the total consumer base, followed by North America (24.4%), Europe (23.4%), and Asia (21.8%), according to the pie chart. Higher client concentration areas might need more inventory, while other areas might benefit from focused marketing campaigns to increase interaction.

### 3. Top 10 Selling Products



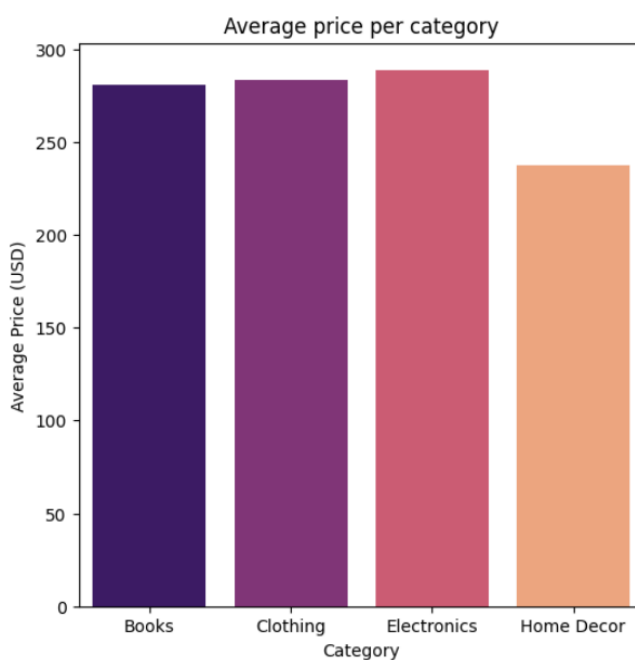
Which products are generating the most sales in terms of quantity are shown by the Top-Selling Products insight. The top ten products by quantity sold are displayed in a bar plot. The ActiveWear SmartWatch has the highest sales. By determining which products sell the best, the company may concentrate on keeping or growing its inventory of these items and take cross-selling or bundling tactics into account.

#### 4. Price Consistency Comparison



A key component of client pleasure and trust is price constancy. The distribution of product prices between the company-set price (Price\_Product) and the actual price charged to customers (Price\_Merged) is displayed in the Price Consistency Comparison box plot. A significant price range could indicate problems like price inconsistencies or possible mistakes in the sales process. While disparities may result in unhappy consumers or lost sales opportunities, a consistent pricing strategy guarantees that customers see value and fairness.

#### 5. Average Price per Category



Price trends for several product categories can be seen in the Average Price per Category bar plot. Electronics have the highest average price, followed by apparel, books, and home décor. We can determine which categories have higher or lower pricing strategies by figuring out the average price for each category. While lower-priced categories may appeal to more budget-sensitive consumers, higher-average price categories may be positioned as premium. This knowledge can be useful when setting prices.