

# Document

## eCommerce Transactions Dataset Analysis

---

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

---

#### Dataset Loading & Data Cleaning

- No null values present in Customers, Products & Transactions Dataset
- No duplicate rows present in Customers, Products & Transactions Dataset

I combined the Products, Customers & Transaction Dataframe to form a Single Dataframe. Firstly, I combined Customers & Transactions dataset using Inner join using common Column: CustomerID

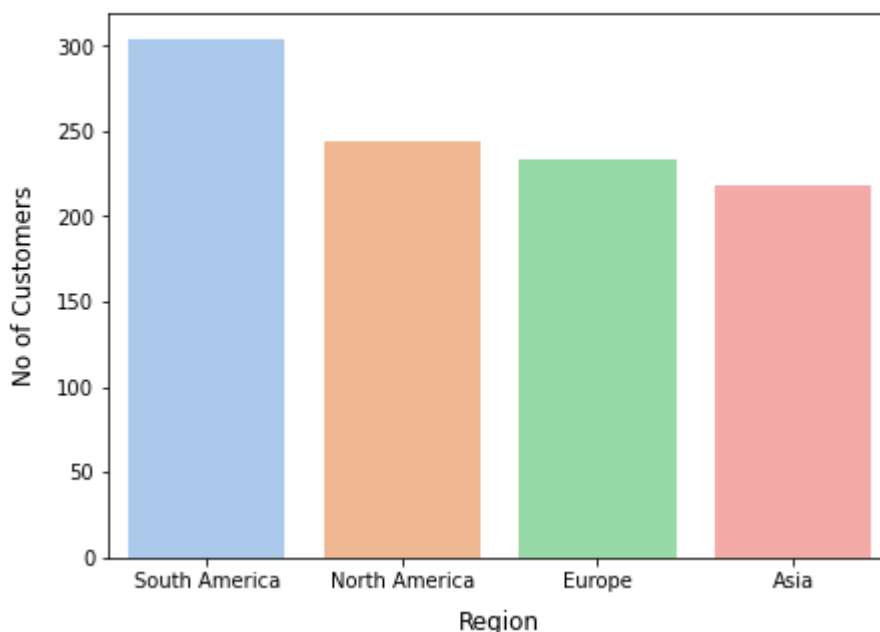
Then combined the above dataset with Products dataset using Inner join using common Column: ProductID

#### EDA

##### Customer Segmentation

- Customer Distribution by Region
  - The Bar plot shows the number of customers belonging to different regions.

Customer Distribution by Region

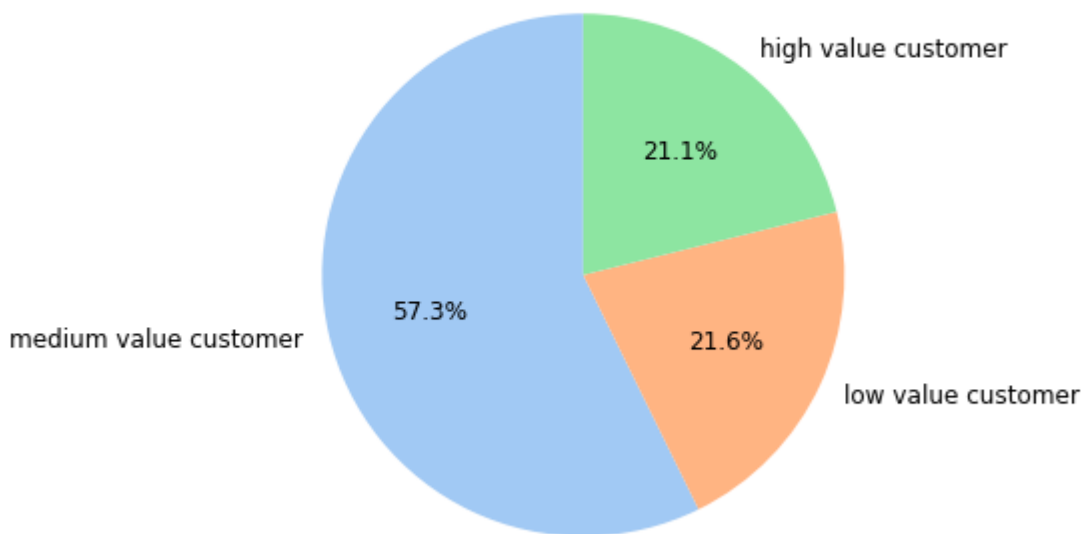


- Customer Net Value

- o Based on the total amount spend, three categories are defined:  
High Value Customer (Total Spent greater 5000)  
Medium Value Customer (Total Spent greater than 2000 but less than or equal to 5000) and  
Low Value Customer (Total Spent less than 2000)

A pie chart is plotted to show this distribution.

### Customers Spending Categories Pie Chart



## Product Performance

- Most Sold Products
  - A word cloud is drawn to show the most sold products

### Most Selling Products Word Cloud



- Most Sold Categories

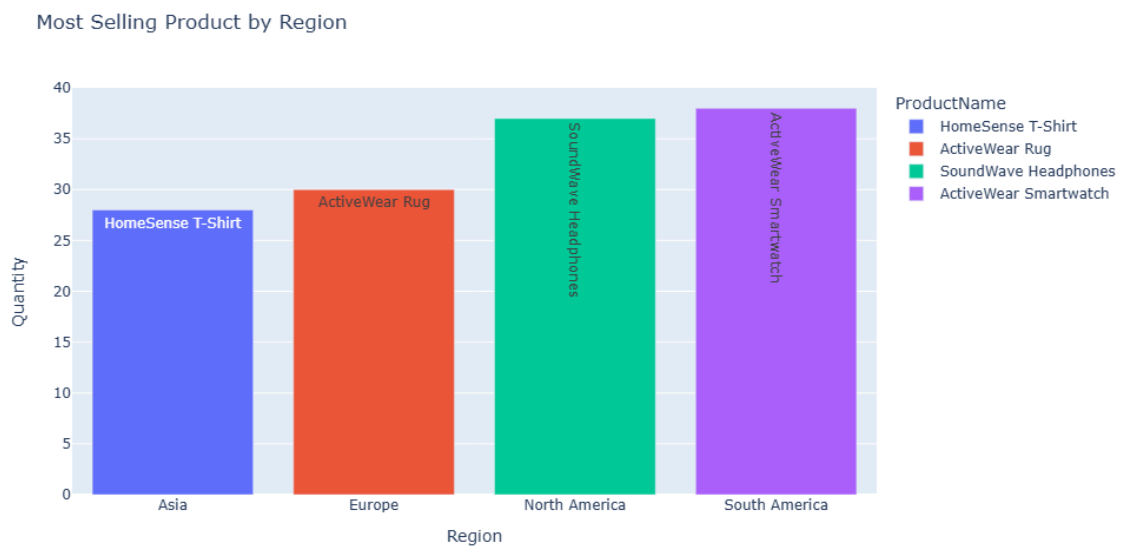
- A word cloud is drawn to show the most sold categories

Most Selling Categories Word Cloud



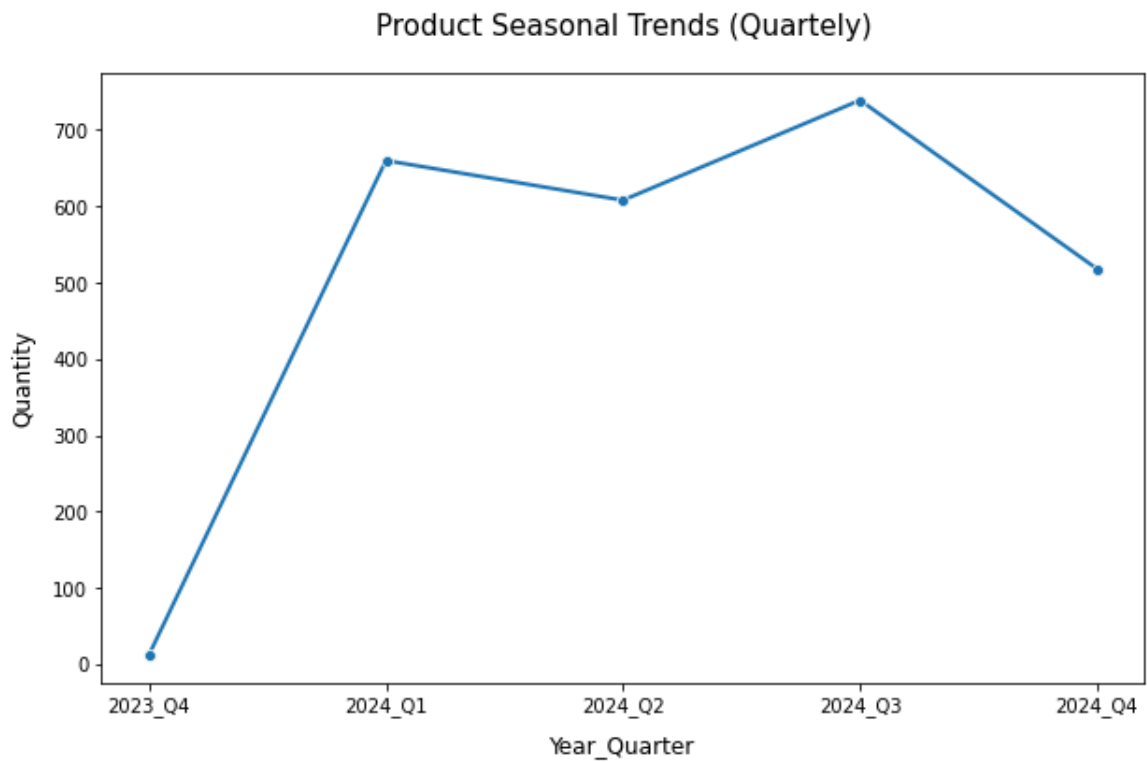
- Most Selling Product by Region

- The bar plot shows the product that was purchased the most times in a particular region.



- Product Seasonal Trends

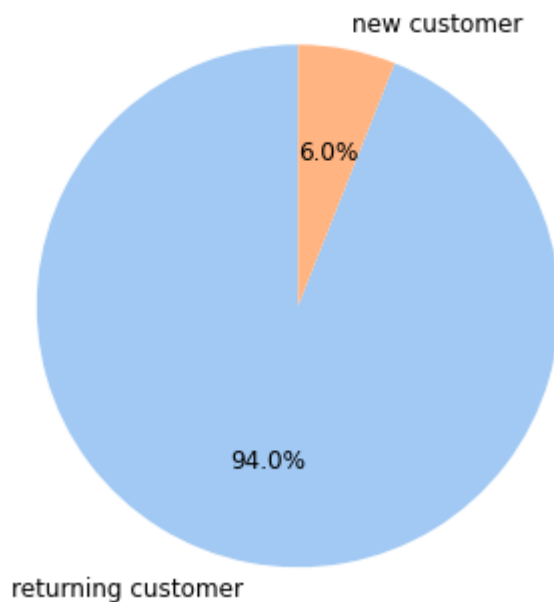
- The TransactionDate is divided into quarters based on the month. It shows the number of products sold in each quarter.



## Customer Retention

- The pie chart shows the number of users who buy frequently and first-time users.

Customers Retention Pie Chart



## Business Insights

- By segmenting users based on geographical location, targeted marketing campaigns can be implemented. (Customer Distribution by Region plot)
- By categorizing users into three groups, targeted high-value products and advertisements can be shown to each segment, improving the relevance of marketing

efforts. (Customer Net Value plot)

- Analyzing data on the most sold products helps make better inventory planning decisions. (Most Sold Products)
- Analyzing data on the most sold categories helps reveal customer preferences and optimize product assortments. (Most Sold Categories)
- Analyzing data on the most sold products by region helps in developing effective expansion and pricing strategies. (Most Selling Product by Region plot)
- By analyzing quarterly product demand and seasonal trends, promotional strategies and sales targets can be set effectively. (Product Seasonal Trends)
- By analyzing the Customer Retention plot, targeted discount vouchers can be offered to first-time buyers, incentivizing repeat purchases and fostering long-term customer loyalty. (Customer Retention plot)