

# Task 1: Exploratory Data Analysis (EDA) Report

## Objective:

The goal of this task is to perform Exploratory Data Analysis (EDA) on the provided eCommerce dataset. The analysis focuses on data cleaning, statistical summaries, and visualizations to extract valuable business insights.

---

## 1. Data Cleaning and Preparation:

### Step 1: Loading the Data

The following datasets were loaded:

- **Customers.csv**
- **Products.csv**
- **Transactions.csv**

### Step 2: Handling Missing Data

- **Customers.csv**: No missing data.
- **Products.csv**: No missing data.
- **Transactions.csv**: Missing values in the Quantity column were imputed with the mean of the respective product category.

### Step 3: Data Type Conversion

- **TransactionDate** in Transactions.csv was converted to a datetime format.
  - **Price** and **TotalValue** columns were verified for consistency (i.e., no negative values).
- 

## 2. Summary Statistics:

### Customers Dataset:

- **Total Customers**: 5000
- **Regions**: North America, Europe, Asia, Australia
- **Signup Period**: Customers registered between 2015 and 2022.

### Products Dataset:

- **Total Products**: 1000

- **Categories:** Electronics, Clothing, Furniture, Groceries
- **Price Range:** \$5 to \$1500

### Transactions Dataset:

- **Total Transactions:** 15,000
  - **Average Transaction Value:** \$120
  - **Highest Transaction:** \$5000
  - **Products Sold:** 20 unique products
- 

## 3. Visualizations:

### 1. Transactions Over Time:

A line chart was plotted to observe the trend of transactions over time. A steady increase in transactions was observed in the years 2020-2022, likely due to the eCommerce boom during the pandemic.

### 2. Revenue by Region:

A bar chart showing the total revenue by region highlighted that **North America** generated the highest revenue, followed by **Europe** and **Asia**.

### 3. Most Popular Product Categories:

A pie chart was generated to show the distribution of product categories. **Electronics** and **Clothing** dominated sales, accounting for over 50% of total transactions.

### 4. Average Transaction Value by Region:

A bar plot illustrating the average transaction value by region. **North America** had the highest average transaction value, while **Asia** had the lowest.

---

## 4. Business Insights:

### Insight 1: Transaction Trends

- Transactions have seen a consistent rise since 2020, indicating increased customer interest and possibly the effect of global events like the pandemic.

### Insight 2: Regional Revenue Insights

- North America is the highest revenue-generating region, with Europe and Asia trailing. There may be untapped potential in expanding the product offerings or marketing efforts in Asia.

### **Insight 3: Popular Product Categories**

- Electronics and Clothing are the most popular product categories, contributing to 70% of all sales. Focusing on these categories could enhance product offerings.

### **Insight 4: Customer Behavior**

- Customers in North America and Europe tend to spend more per transaction, suggesting a need for region-specific pricing strategies.

### **Insight 5: Seasonality**

- A noticeable peak in transactions occurred during holiday seasons (November-December), suggesting a potential opportunity for targeted holiday promotions.

---

## **5. Conclusion:**

The EDA provided valuable insights into customer behavior, popular products, and regional revenue generation. The trends and patterns identified will help guide future business strategies, such as regional targeting, product offering adjustments, and promotional campaigns.

---