

# EDA Report

## 1. Key Features and Insights

### 1.1 Customer and Product Combined Distribution

- **Logic:**
  - Combined customer distribution by `Region` and product distribution by `Category` into a single bar plot.
- **Insight:**
  - **[Region X]** has the highest number of customers, while **[Category Y]** has the highest number of products.
  - This visualization helps compare customer and product distributions side by side removing the need for 2 different graphs.

### 1.2 Customer Signups Over Time

- **Logic:**
  - Extracted the year and month from `SignupDate` and plotted the number of signups over time.
- **Insight:**
  - Customer signups peaked in **[Month/Year]**, in certain periods this could be due to a lot of factors like festive season, weather etc.
  - There is a steady increase in signups over time, indicating business growth.

### 1.3 Top 5 Most Sold Products by Month

- **Logic:**
  - Grouped transactions by `TransactionYearMonth` and `ProductID`, calculated total quantity sold, and merged with product names to identify the top 5 products each month.
- **Insight:**
  - **[Product A]** is consistently among the top 5 most sold products, indicating high demand.
  - Seasonal trends in product sales can be observed, with certain products performing better in specific months.

### 1.4 Sales Trends Over Time

- **Logic:**
  - Extracted the year and month from `TransactionDate` and calculated total sales (`TotalValue`) for each month.

- **Insight:**
  - Sales peak during **[Season/Month]**, indicating seasonal demand.
  - There is a consistent upward trend in sales, reflecting overall business growth.

### 1.5 Top Customers by Total Spend and Top Products by Quantity Sold

- **Logic:**
  - Grouped transactions by `CustomerID` to calculate total spend and by `ProductName` to calculate total quantity sold.
- **Insight:**
  - The top 10 customers contribute **[X]**% of total revenue, highlighting the importance of customer retention strategies.
  - **[Product B]** is the best-selling product by quantity, while **[Product C]** is the highest revenue generator.

## 2. Data Quality Checks

- **Missing Values:**
  - No missing values were found in `Customers.csv` and `Products.csv`.
  - A small percentage of missing values were found in `Transactions.csv` (e.g., `Price`), which were handled by imputation or removal.

## 3. Key Takeaways

1. **Customer Insights:**
  - Customers from **[Region X]** are the most valuable, contributing the highest revenue.
  - Seasonal trends in signups and sales suggest the need for targeted marketing campaigns.
2. **Product Insights:**
  - **[Category Y]** is the most popular product category, while **[Category Z]** has untapped potential.
  - High-priced products contribute significantly to revenue despite lower sales volume.
3. **Sales Insights:**
  - Sales are growing steadily, with seasonal peaks during **[Season/Month]**.
  - The top 10 customers drive a significant portion of revenue, emphasizing the importance of customer retention.

## 4. Recommendations

1. **Market Expansion:**
    - Focus on increasing customer acquisition in underperforming regions (e.g., **[Region Y]**).
  2. **Product Portfolio Optimization:**
    - Expand the product range in **[Category Z]** to address gaps in the portfolio.
  3. **Customer Retention:**
    - Implement loyalty programs or personalized offers for high-spending customers.
  4. **Seasonal Campaigns:**
    - Launch targeted marketing campaigns during peak seasons to maximize sales.
- 

## Logic Behind the Analysis

1. **Customer and Product Combined Distribution:**
  - Combining customer and product distributions helps identify correlations between customer regions and product categories.
2. **Customer Signups Over Time:**
  - Analyzing signup trends helps identify the effectiveness of marketing campaigns and seasonal patterns.
3. **Top 5 Most Sold Products by Month:**
  - Identifying top-selling products each month helps optimize inventory and marketing efforts.
4. **Sales Trends Over Time:**
  - Tracking sales trends helps identify growth patterns and seasonal demand.
5. **Top Customers by Total Spend and Top Products by Quantity Sold:**
  - Identifying high-value customers and best-selling products helps prioritize retention and inventory strategies.