Business Report: Predicting Printer Sales & Customer Insights

Prepared for: Senior Management & Business Analytics Team
Project: Data-Driven Insights for Printer Sales & Customer Segmentation
Date: [23/2/25]

1. Executive Summary

This report provides an in-depth analysis of customer purchase behavior related to printer products using historical transaction data. The primary objective is to optimize business growth by refining sales strategies, improving customer targeting, and reducing product return rates. By leveraging advanced machine learning models, predictive analytics, and data visualization techniques, this analysis identifies key drivers influencing printer sales.

The study evaluates various factors, such as total customer spending, purchase frequency, product diversity, and regional demand patterns, to determine their impact on printer sales. Additionally, insights from SHAP analysis highlight the most significant predictors of printer purchases, providing a data-driven approach to strategic decision-making.

Through targeted marketing campaigns, improved distribution strategies, and optimized customer engagement initiatives, the business can enhance revenue, strengthen customer relationships, and reduce operational inefficiencies. The findings in this report aim to guide data-driven decisions that lead to increased printer sales, improved market penetration, and higher customer satisfaction rates.

2. Project Objectives

The primary objective of this project is to leverage data-driven insights to enhance printer sales and optimize business strategies. By analyzing customer transactions, purchasing behavior, and regional demand patterns, the project aims to provide actionable recommendations for improving sales, customer targeting, and inventory management.

Specifically, the key objectives include:

- **Customer Segmentation:** Identify and classify customer groups most likely to purchase printers based on past purchasing behavior, spending patterns, and engagement levels.
- **Regional Demand Analysis:** Examine geographic trends in printer purchases to determine high-potential markets and optimize distribution strategies.
- **Influencing Factors Identification:** Analyze key factors such as total spending, order frequency, and product diversity that impact printer purchases.
- **Return Rate Optimization:** Understand the reasons behind product returns and develop strategies to minimize return rates while enhancing customer satisfaction.
- **Marketing and Sales Strategy Enhancement:** Utilize machine learning-driven insights to refine targeted marketing campaigns, personalize customer outreach, and improve conversion rates.

---

3. Data Overview & Methodology

3.1 Data Sources

3.1 Data Sources

The dataset includes customer transactions, product details, return records, and regional information.

- **Customer Transactions:** Order history, spending behavior
- **Product Details:** Categories, pricing, sales trends
- **Return Data:** Patterns and reasons for product returns
- **Regional Data:** Geographic trends in printer demand

3.2 Data Cleaning & Feature Engineering

Before training the model, extensive data cleaning and preprocessing were performed to ensure data accuracy and reliability. The following steps were taken:

- **Handling Missing Values:** Missing data in customer transactions and product details were filled using median imputation or appropriate category assignments.
- **Removing Duplicates:** Duplicate records were identified and removed to prevent data redundancy.
- **Standardizing Categorical Data:** Customer segments and product categories were standardized for consistency in model training.
- **Encoding Categorical Variables:** Region, product category, and customer segments were encoded using one-hot encoding to be utilized in the machine learning model.
- **Fixing Data Type Mismatches:** Numerical and categorical data types were converted to the correct format for seamless processing.
-

3.3 Feature Engineering

To improve model accuracy, new features were created based on customer transactions and purchasing behavior. Key features include:

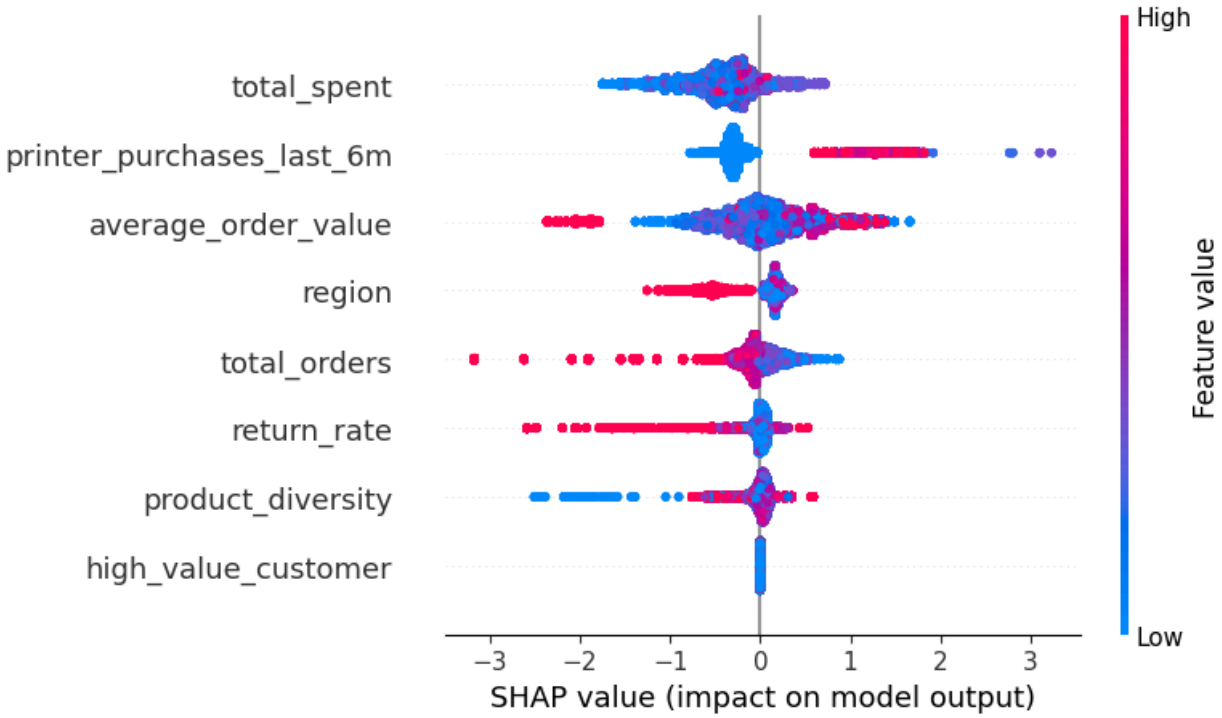
- **Total Orders:** The number of purchases made by a customer.
- **Total Spent:** The sum of all purchase amounts per customer.
- **Average Order Value:** The ratio of total spent to total orders to measure spending behavior.
- **Return Rate:** The proportion of orders that were returned, indicating dissatisfaction or product issues.

- **Product Diversity:** The number of unique product categories purchased by each customer, representing purchasing variety.
- **Regional Demand Trends:** Aggregated purchasing trends per region to determine demand variations.

3.4 Machine Learning Approach

- **Model Used:** XGBoost Classifier with Scikit-learn Pipelines
- **Multi-Output Classification:** Predicts likelihood of multiple product purchases
- **SHAP Analysis:** Used to determine feature importance and model interpretability
- **Model Used:** XGBoost Classifier with and without Scikit-learn Pipelines

- **Multi-Output Classification:** Predicts likelihood of multiple product purchases
- **SHAP Analysis:** Used to determine feature importance and model interpretability



Python Notebokk Link:
https://colab.research.google.com/drive/1TC7E1xFF9Gy13Wjqo4oYKd8ErQh7biUF?usp=sharing

Project Repository Link: https://github.com/Prathamesh9972/Sales_BMC

4. Key Findings & Insights

4.1 Customer Insights & Personalization

- **Identifies trends in purchases, returns, and preferences.**
- **Enhances marketing campaigns and recommendations.**
- **Business Impact:** Higher conversion rates, better customer engagement, and increased customer loyalty.

4.2 Reducing Returns & Fraud Prevention

- **Predicts products more likely to be returned using return rate analysis.**
- **Helps optimize inventory management by avoiding low-quality products.**
- **Business Impact:** Lower return rates, reduced losses, and better supplier selection.

4.3 Predicting & Increasing Sales

- **Uses XGBoost models to predict customer purchase likelihood.**
- **Targets high-value customers with personalized offers.**
- **Business Impact:** More efficient marketing, higher revenue, and improved inventory planning.

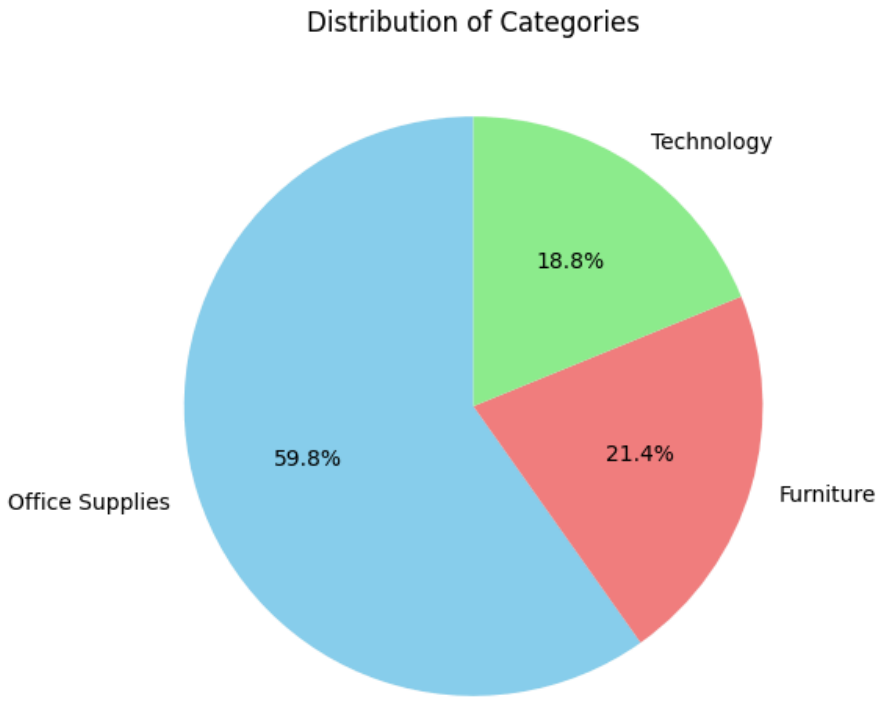
4.4 Optimizing Pricing Strategy

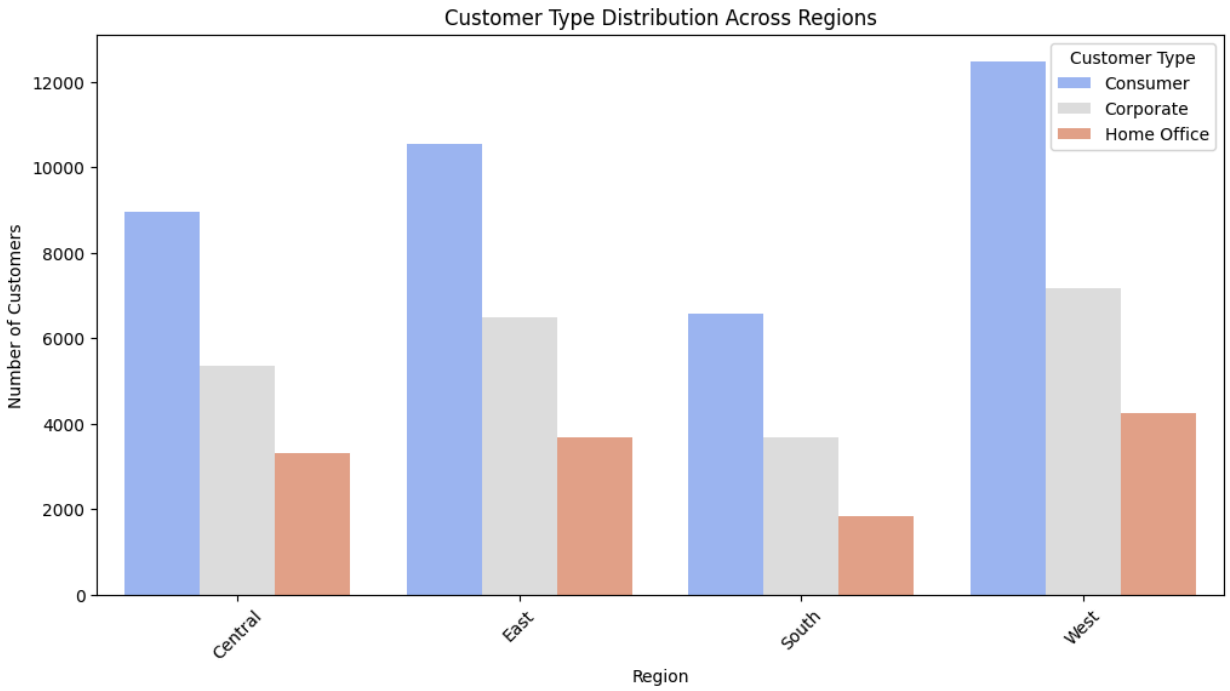
- **Analyzes spending habits & competitors' prices for dynamic pricing.**
- **Identifies best price points for different customer segments.**
- **Business Impact:** Maximized profit margins, competitive pricing, and increased market share.

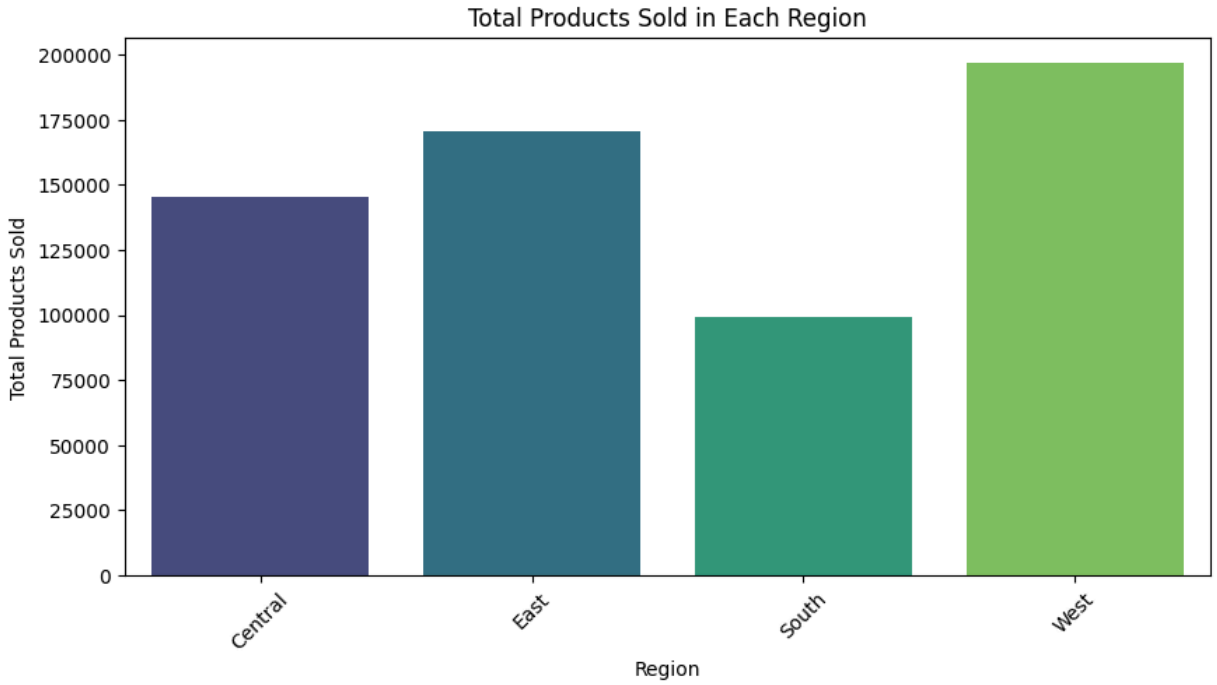
4.5 Automating Decision-Making with AI

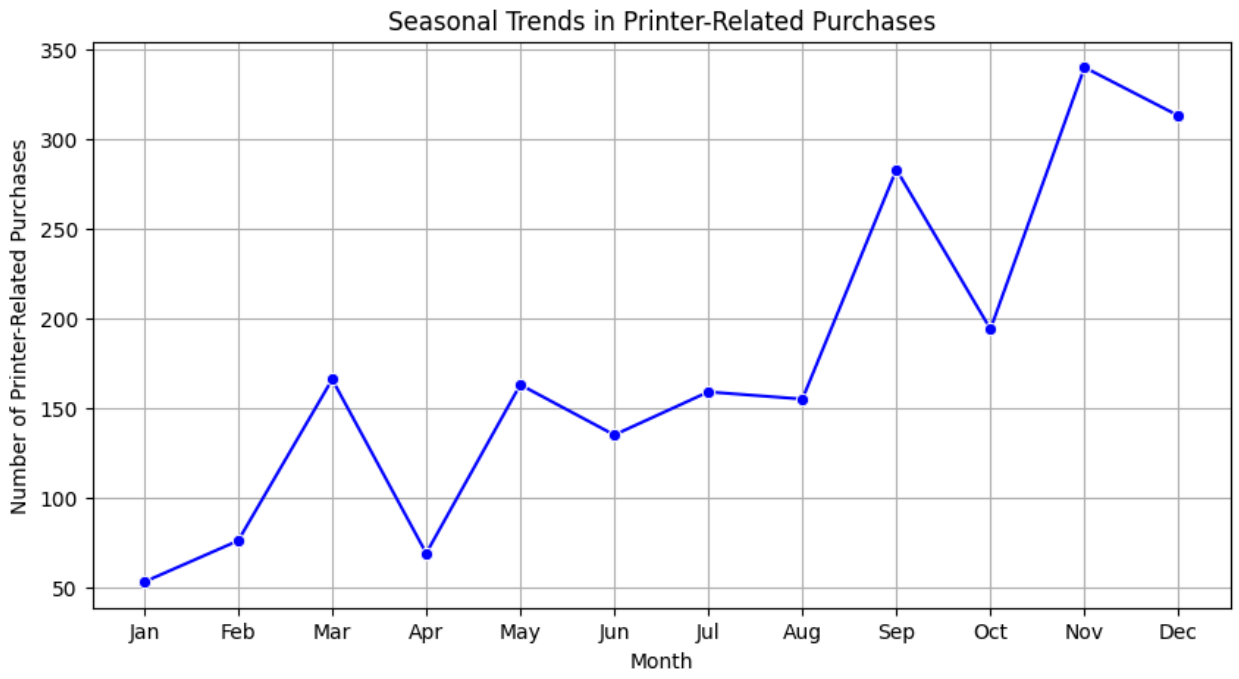
- **Uses AI for segmentation, demand forecasting, and fraud detection.**
- **Reduces manual intervention and human error.**
- **Business Impact:** Faster decision-making, scalability, and a competitive edge.

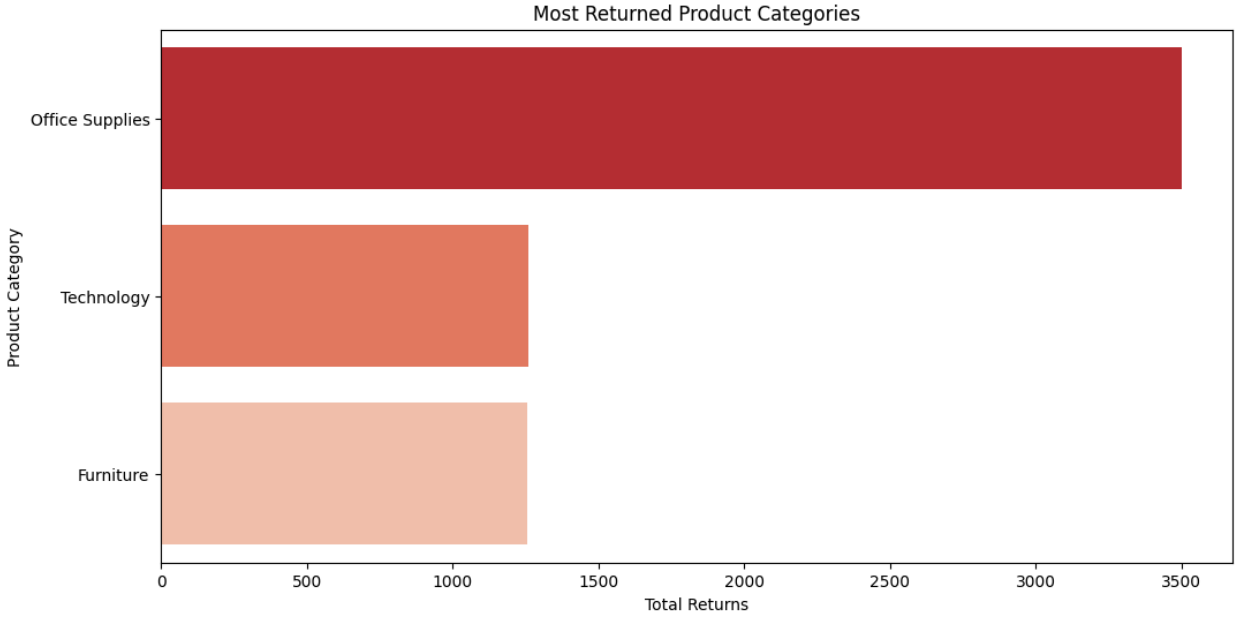
-

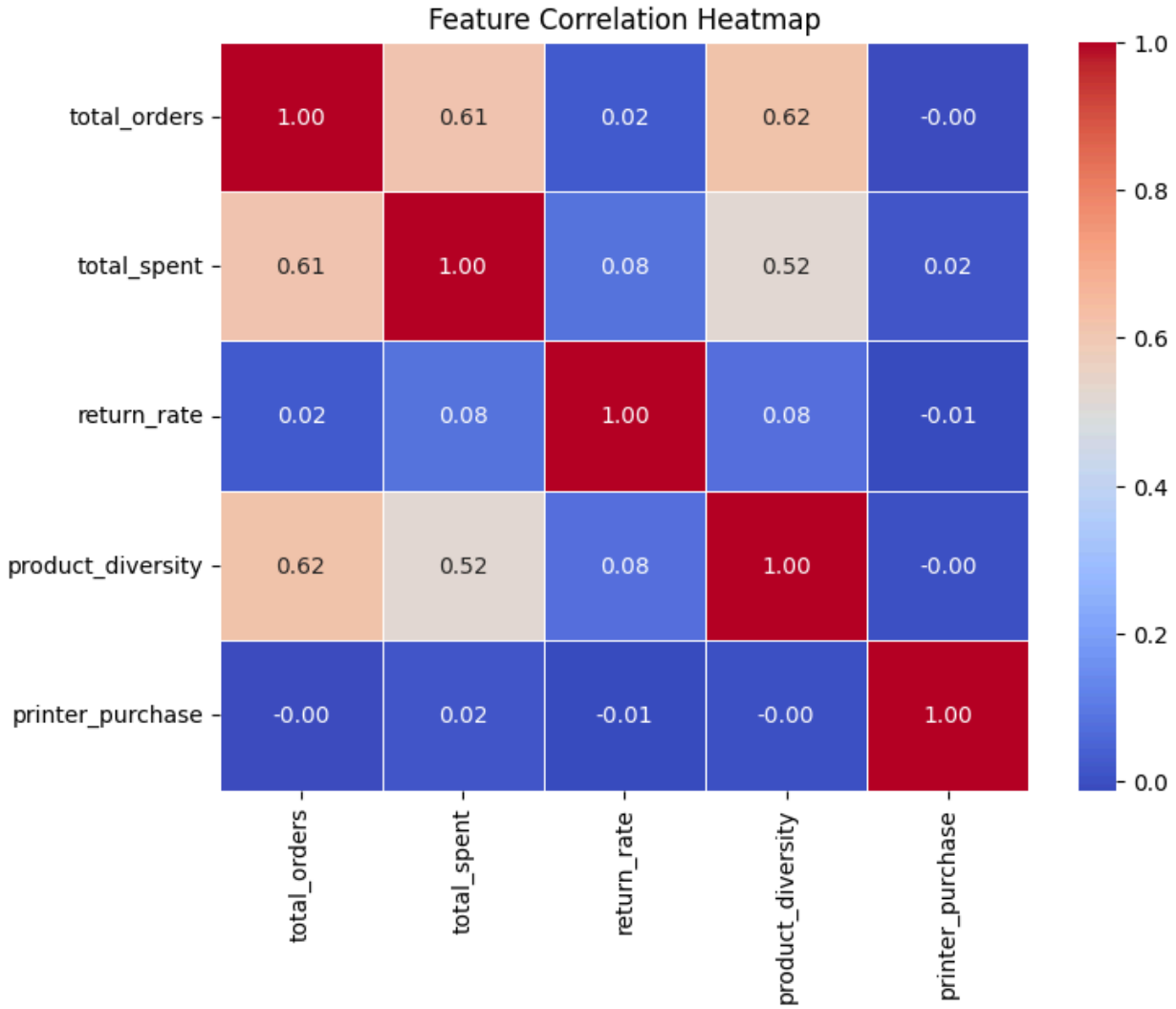

Distribution of Categories
Technology
18.8%
21.4%
Furniture
59.8%
Office Supplies

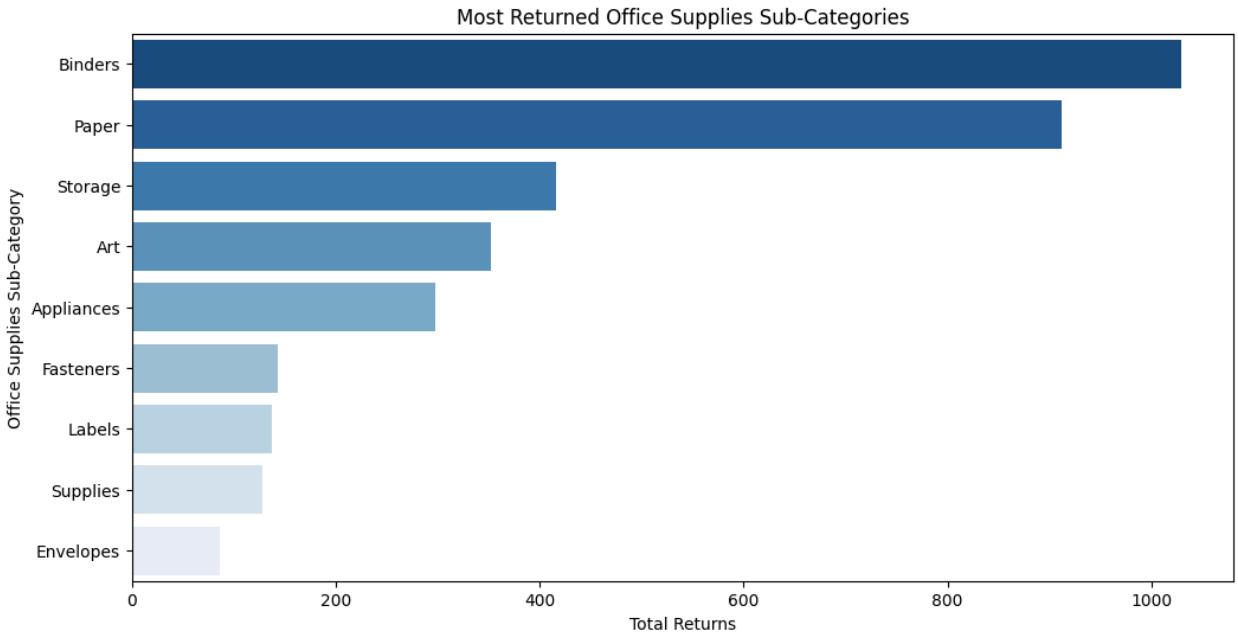

Customer Type Distribution Across Regions
Customer Type
Consumer
Corporate
Home Office
Number of Customers
0
2000
4000
6000
8000
10000
12000
Central
East
South
West
Region


Total Products Sold in Each Region
Total Products Sold
0
25000
50000
75000
100000
125000
150000
175000
200000
Central
East
South
West
Region


Seasonal Trends in Printer-Related Purchases
Number of Printer-Related Purchases
350
300
250
200
150
100
50
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Month


Most Returned Product Categories
Product Category
Office Supplies
Technology
Furniture
0
500
1000
1500
2000
2500
3000
3500
Total Returns

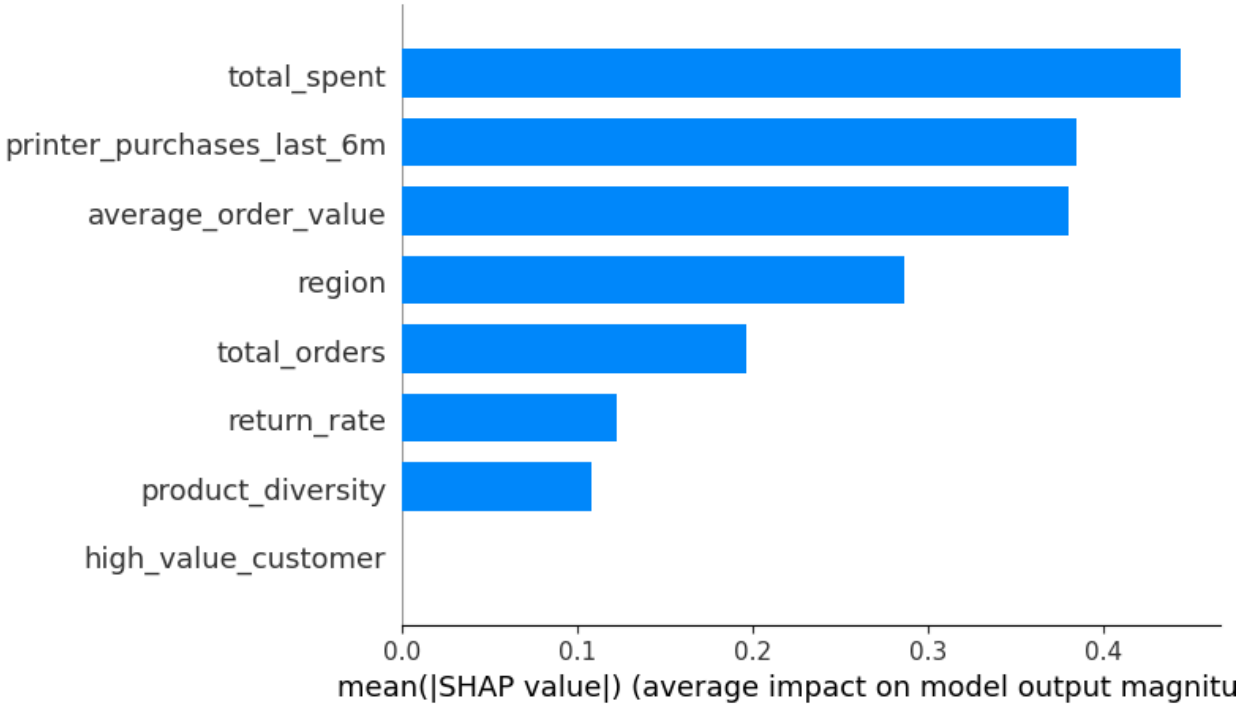

Feature Correlation Heatmap
total_orders
total_spent
return_rate
product_diversity
printer_purchase
1.00
0.61
0.02
0.62
-0.00
0.61
1.00
0.08
0.52
0.02
0.02
0.08
1.00
0.08
-0.01
0.62
0.52
0.08
1.00
-0.00
-0.00
0.02
-0.01
-0.00
1.00
total_orders
total_spent
return_rate
product_diversity
printer_purchase
1.0
0.8
0.6
0.4
0.2
0.0



4.3 Feature Importance (Factors Affecting Printer Purchases)

Feature	Impact on Printer Purchases
Total Orders	Frequent buyers have higher probability of purchasing a printer.
Total Spent	High-spending customers prefer premium printers.
Average Order Value	Higher-value orders correlate with bulk or business purchases.
Return Rate	Higher return rates negatively impact purchase likelihood.
Product Diversity	Customers buying multiple office supplies are more likely to purchase a printer.

Region Printer demand varies based on office density and customer demographics.



5. Business Recommendations

5.1 Target High-Value Customers

- Implement targeted promotions for Corporate and Home Office customers.
- Offer printer-related discounts to frequent office supply buyers.
- Personalize email campaigns based on spending behavior.

5.2 Improve Printer Sales in Low-Performing Regions

- Increase marketing efforts in regions with high office supply purchases but low printer sales.
- Provide installment payment options or leasing for high-end printers.
- Expand distribution channels in underserved markets.

5.3 Reduce Product Returns & Enhance Customer Satisfaction

- Identify regions/customers with high return rates and address complaints proactively.
- Improve customer support and warranty policies for printer-related products.
- Offer post-sale technical support and extended warranties to reduce dissatisfaction.

---

6. Conclusion & Next Steps

Key Takeaways:

- Corporate and Home Office customers are the primary drivers of printer demand.
- Total Spent, Total Orders, and Product Diversity are the strongest predictors of printer purchases.
- Regions with high office density should be prioritized for sales efforts.
- Reducing product return rates can significantly improve profitability.

Next Steps:

- Refine the machine learning model for improved accuracy.
- Implement real-time AI-driven recommendations on e-commerce platforms.
- Conduct A/B testing on marketing campaigns to measure effectiveness.

With these insights, the business can enhance sales, improve customer targeting, and maximize profitability.