E-commerce Return Rate Reduction Analysis

1. Introduction

In the e-commerce industry, product returns significantly impact profitability, customer satisfaction, and logistics. Returns often result in extra costs like reverse logistics, damaged goods, and lost customer trust. Understanding why customers return products and who is more likely to return them can help businesses reduce return rates and improve operations.

This project aims to analyze product return patterns, predict high-risk orders, and build a dashboard to guide business decisions on reducing return rates.

2. Abstract

The E-commerce Return Rate Reduction Analysis project analyzes return data to uncover key drivers behind product returns. We used a synthesized dataset of 200 records that includes orders, customers, products, returns, and marketing channel data. The project involved data cleaning, exploratory data analysis (EDA), logistic regression modeling, and interactive dashboard creation using Power BI.
The final outcomes include:
- A prediction model to assign return risk scores.
- Insights into return trends by category, region, and marketing channel
- A Power BI dashboard for business users to drill down into high-risk segments

3. Tools Use

Python (Pandas, Seaborn, Scikit-learn)
Data cleaning, EDA, and predictive modeling

Power BI –
Dashboard creation and visualization

Google Colab –
Python coding environment

CSV Files
Input data

4. Steps Involved in Building the Project

Step1:Data Collection -
Five CSV datasets were prepared: orders, returns, products, customers, and marketing channels

Step2:Data Cleaning and Merging (Python) -
- Converted date fields into proper datetime format
- Merged all datasets using keys like order_id, product_id, and customer_id
- Created new features such as is_returned, discount_pct, delivery_days

Step 3: Exploratory Data Analysis -
- Visualized return rates by category, region, supplier, and marketing channel
- Investigated the impact of discounts and delivery delays on return behavior

Step 4: Predictive Modeling (Logistic Regression)
- Used logistic regression to predict return probability based on order attributes
- Outputted a return_risk_score between 0 and 1

Step 5: Dashboard Creation in Power BI
- Included visuals like return trends, return rates by category and region, and high-risk product

5. Conclusion

This project provides a comprehensive approach to minimizing product return rates by combining data science for predictive modeling and business intelligence for visualization. By identifying return-prone categories, risky regions, and customer behavior patterns, e-commerce companies can:
- Reduce unnecessary returns
- Adjust marketing campaigns
- Optimize product descriptions or supplier quality

The dashboard allows continuous monitoring and helps prioritize interventions in real time.