

E-commerce Return Rate Reduction Analysis

Introduction

Product returns are a critical concern for e-commerce businesses. They affect profitability, customer satisfaction, and inventory flow. This project focuses on analyzing return patterns from an e-commerce dataset to identify which customer, product, and transactional factors contribute to high return rates. The aim is to develop data-driven strategies and predict high-risk returns to minimize losses.

Abstract

This project analyzes an e-commerce dataset to identify the key drivers behind product returns and predict return likelihood using a logistic regression model. By conducting exploratory data analysis (EDA) and visualizing patterns through Power BI, actionable insights were derived on how product category, customer-rating, region, and discounts influence returns. A machine learning model was trained to predict return risk, and high-probability transactions were exported for review. The final deliverables include an interactive dashboard, model code, and a high-risk transaction CSV.

Tools Used

- Python: Data cleaning, EDA, logistic regression modeling
- Pandas, Scikit-learn, Seaborn: Python libraries used
- Power BI: Interactive dashboards and visual insights
- GitHub: Project storage and deliverable submission

Steps Involved in Building the Project

1. Dataset Selection & Loading

- Used E_Commerce_Customer_Analytics.csv with 18 features related to orders, returns, ratings, and Discounts.

2. Data Cleaning & Feature Engineering

- Cleaned Return_Status and created a binary return_flag.
- Converted dates, handled missing values, and selected relevant features.

3. Exploratory Data Analysis (EDA)

- Analyzed return rates by product category, region, discount applied, and customer rating.
- Found strong correlation between low ratings/high discounts and return rates.

E-commerce Return Rate Reduction Analysis

4. Modeling with Logistic Regression

- Selected key features: Customer_Rating, Discount_Applied, Quantity, Order_Amount
- Built and evaluated a logistic regression model to predict return probability.

5. High-Risk Transaction Export

- Predicted return probabilities on test data
- Exported high-risk transactions to high_risk_products.csv (probability > 0.5)

6. Power BI Dashboard

- Created clean, interactive visuals:
Return rate by category, region, rating, and discount
KPI card and filters for user exploration
Slicers for Region, Product_Category, Customer_Rating

Conclusion

Through this project, a deeper understanding of return behavior was developed. The logistic regression model enabled the identification of potentially returnable transactions before they occurred. Visualizations helped isolate high-risk categories and regions, and the final CSV allows the business to take action on those. These insights can help e-commerce platforms reduce losses and improve operational efficiency.