E-commerce Return Rate Reduction Analysis

Introduction

The rapid expansion of the e-commerce industry has brought convenience to customers and growth to businesses. However, one of the most pressing operational challenges is the high rate of product returns. These returns negatively impact profitability, logistics, and customer satisfaction. This project focuses on analyzing the patterns and predictors of product returns to enable proactive measures that reduce return rates and operational losses.

Abstract

This analysis was conducted on a structured dataset comprising e-commerce order information, return status, payment methods, product categories, customer locations, and delivery metrics. The goal was to understand which factors influence the likelihood of a return and to develop a predictive model for flagging high-risk products. Using Python, logistic regression modeling, and Power BI, the project uncovered actionable insights. These include category-wise return trends, the influence of payment methods (especially Cash on Delivery), and the impact of delivery time on return probability. A return risk dashboard was also built to help stakeholders visualize patterns and take timely action.

Tools Used

- Python (Pandas, Seaborn, Scikit-learn, Jupyter Notebook)
- Power BI (for dashboarding)
- Excel (for initial inspection and basic preprocessing)

Steps Involved in Building the Project

1. Data Cleaning:
   - Removed missing values, duplicates, and irrelevant identifiers such as Order ID, Product ID, and Return Reason.

2. Feature Engineering:
   - One-hot encoded categorical variables like Product Category, Region, Payment Method, etc.

3. Exploratory Data Analysis:
   - Plotted return percentages by category, geography, and payment method.
   - Identified that electronics and fashion had the highest return rates.

4. Predictive Modeling:

   - Used Logistic Regression to train a binary classification model.

   - Achieved approximately 80% accuracy on test data.

5. Evaluation:

   - Confusion matrix, precision, recall, and F1-score were used for performance measurement.

6. Dashboard Development:

   - Created an interactive Power BI dashboard with filters for category, geography, and channel.

   - Enabled product-level drilldowns to isolate high-return segments.

7. Output Generation:

   - Exported a list of high-risk products (with >50% return probability) for business intervention.

Conclusion

The return rate analysis highlighted the critical importance of focusing on category-wise return trends and logistics performance. High return rates were notably concentrated in Electronics and Apparel, with COD orders and longer delivery timelines emerging as contributing factors. By leveraging logistic regression, the project successfully identified products at risk of return. The interactive dashboard empowers internal teams to visualize risk segments and take immediate action. Going forward, this model can be retrained on live data periodically, and its predictions integrated into inventory and marketing strategies. This solution not only reduces operational burden but also enhances customer satisfaction and trust.