E-commerce Return Rate Reduction Analysis

Introduction

In the dynamic world of e-commerce, product returns are a significant operational and financial concern. This project focuses on analyzing return patterns and identifying key drivers of returns across categories, regions, suppliers, and marketing channels. The goal is to predict the likelihood of product returns and help businesses take proactive steps to reduce avoidable return rates.

Abstract

This analysis involved cleaning and preprocessing a synthetic dataset of 3000 e-commerce orders. We explored return trends by various dimensions (category, supplier, region, etc.) and built a logistic regression model to predict return probability. Based on the model output, high-risk products were identified and exported for targeted attention. The entire analysis was visualized through an interactive Power BI dashboard, enabling drill-down by supplier and product type to support strategic decisions.

Tools Used

- **Python (Pandas, scikit-learn):** Data cleaning, feature engineering, modeling

- **Power BI:** Dashboard creation and interactive reporting

- **Excel/CSV:** Dataset manipulation and export

Steps Involved in Building the Project

1. **Data Import & Cleaning**

   The dataset was loaded from a CSV file. Null values were handled, date formats were corrected, and a binary column `is_returned` was created to mark returned orders.

2. **Exploratory Data Analysis (EDA)**
   Analyzed return % by product category, region, supplier, and marketing channel. Identified categories and regions with high return rates.

3. **Model Building**
   Applied logistic regression using scikit-learn, encoded categorical variables using OneHotEncoder, evaluated model accuracy, and generated return probability scores.

4. **Risk Scoring & Export**
   Appended `return_probability` to the dataset. Filtered high-risk products (probability > 0.5) and exported to CSV.

5. **Dashboard Creation**
   Built an interactive Power BI dashboard showing return % by various dimensions, KPI summary cards, return risk tables, and supplier drill-through views.

Conclusion

This project successfully highlights how data-driven insights can reduce operational inefficiencies in e-commerce. By modeling the probability of returns, businesses can proactively adjust product offerings, vendor contracts, and customer policies. The Power BI dashboard makes these insights easily accessible for stakeholders to drive informed actions. This approach not only minimizes return-related losses but also improves customer satisfaction and supply chain resilience.