E-commerce Return Rate Reduction Analysis

Introduction

This project aims to analyze and reduce the return rate in an e-commerce platform by understanding patterns behind customer returns. High return rates often indicate issues such as poor product quality, mismatched customer expectations, or ineffective marketing. By identifying these patterns and understanding return behavior, the project provides actionable insights to enhance product offerings, optimize logistics, and improve customer satisfaction.

A significant part of the project is focused on building an interactive Power BI dashboard and training a predictive model to flag high-risk products likely to be returned.

Abstract

The E-commerce Return Rate Reduction Analysis focuses on identifying key factors responsible for high product returns across various product categories, regions, and marketing channels. Using a synthetically generated dataset, we explore customer return behavior by conducting exploratory data analysis (EDA) and statistical modeling. A logistic regression model is developed to predict the likelihood of return for each product, helping to isolate risk factors.

These insights are visualized using Power BI to create an executive dashboard that business stakeholders can use to monitor return rates in real time and drill down into categories and suppliers with the highest return risk. This combined approach supports proactive decision-making and operational optimization.

Tools Used

- Python (pandas, scikit-learn, seaborn, matplotlib) for data analysis and machine learning
- Power BI for dashboard creation and interactive data visualizations
- GitHub for source code and version control
- Visual Studio Code as the development environment
- Microsoft Excel for dataset inspection and minor data formatting

Steps Involved in Building the Project

1. Created a synthetic dataset representing product orders and return behavior.
2. Cleaned and preprocessed the dataset for analysis, handling nulls and inconsistent types.
3. Conducted exploratory data analysis (EDA) using Python to identify key return patterns based on product category, location, supplier, and marketing channel.
4. Trained a logistic regression model to classify orders as likely to be returned or not.
5. Evaluated model accuracy using classification metrics and confusion matrix.
6. Filtered high-risk products (those with predicted return probability > 0.5) and exported them to a CSV.
7. Built a Power BI dashboard showing return percentages, high-risk products, supplier-wise risk, and return rate trends over time.
8. Shared GitHub repository containing all code files and visual outputs.
9. Created this project report to summarize the methodology, tools, and findings.

Conclusion

This project demonstrated how data analytics, machine learning, and visual dashboards can be combined to address real-world business challenges. The return prediction model helped identify which products are at higher risk of being returned, enabling businesses to take preventive measures such as revising product descriptions, improving delivery processes, or tightening quality control.

The Power BI dashboard enables stakeholders to monitor return activity and drill into granular insights by product category or customer segment. This end-to-end project, from dataset creation to actionable dashboarding, highlights the power of data-driven strategies to reduce return rates and boost operational efficiency.