E-Commerce Return Rate Reduction Analysis

By Harish Kumar

Introduction

This project focuses on analyzing and reducing product return rates in an e-commerce business using real transactional data. The goal is to understand which products and regions have the highest return probability and identify actionable steps to reduce returns. Using data analytics and visualization tools, the project transforms raw data into insights that help optimize inventory, improve customer satisfaction, and increase profit margins.

Abstract

Product returns are a major challenge for online retailers, directly impacting profitability and operations. This project combines Python-based machine learning and Power BI visual analytics to predict and analyze product return rates. A logistic regression model was developed to calculate the probability of returns for each product, and the results were visualized in Power BI using a clean, light professional dashboard theme. The dashboard highlights total sales, return risk percentages, and high-risk products and countries. These insights allow businesses to identify problem areas and take preventive actions such as improving product descriptions, quality control, and packaging.

Tools Used

• Python (Pandas, Scikit-learn): Data cleaning and logistic regression modeling
• Power BI: Dashboard creation and interactive visualization
• Excel / CSV: Data inspection and manual validation
• ChatGPT (OpenAI): Project guidance, code assistance, and report writing support

Steps Involved in Building the Project

1. Data Cleaning & Preparation (Python): Removed duplicates, missing values, and incorrect data types. Created key fields such as IsReturn (to mark returned items) and ReturnProbability (predicted likelihood of return).
2. Feature Engineering: Added total sales and month of purchase columns. Prepared data for logistic regression modeling.
3. Model Development: Used Logistic Regression to estimate return probability for each product. Exported high-risk product predictions to High_Risk_Products.csv.
4. Dashboard Creation (Power BI): KPI Cards, Bar Chart, Map, Scatter Plot, Table Visual, and Slicers were added to analyze sales and return risk distribution.

Conclusion

This project demonstrates how combining Python analytics and Power BI visualization can provide deep insights into e-commerce operations. The dashboard effectively identifies high-return products and regions, helping the business reduce unnecessary losses and improve decision-making. Products with lower unit prices and higher order quantities showed greater return probabilities. By focusing on these areas, companies can improve product quality, descriptions, and policies to reduce returns. Overall, this project successfully integrates data science, visualization, and AI-powered guidance (ChatGPT) to create a complete, insight-driven business intelligence solution.