

# E-Commerce Return Rate Reduction Analysis Report

Prepared by D. Bhuvaneswari

June 27, 2025

## Abstract

This report outlines the development of an analytical dashboard to reduce e-commerce return rates by identifying high-risk products and trends. Using synthetic order and return data, the project employs machine learning and SQL to predict return probabilities and aggregate insights, visualized through an interactive Power BI dashboard. The analysis revealed key patterns in return rates across product categories, geographies, and marketing channels, enabling data-driven strategies to minimize returns.

## Introduction

E-commerce businesses face significant challenges due to high return rates, impacting profitability and operational efficiency. This project aims to analyze return patterns using predictive analytics and visualization to identify high-risk products and inform strategic decisions. By leveraging synthetic data, machine learning, and a robust data pipeline, the project delivers actionable insights through an interactive Power BI dashboard, supporting stakeholders in reducing returns effectively.

## Tools Used

- **Python:** For data generation, preprocessing, and predictive modeling using libraries like pandas, NumPy, and scikit-learn.
- **SQLite:** For storing and querying data efficiently using SQL to aggregate return metrics.
- **Power BI Desktop:** For creating interactive visualizations, including treemaps, heatmaps, scatter plots, donut charts, and gauges.
- **Command Prompt (CMD):** For executing Python scripts and managing file operations.
- **ODBC Driver for SQLite:** For connecting Power BI to the SQLite database (optional).

## Steps Involved in Building the Project

1. **Data Generation:** Created synthetic datasets (`orders.csv`, `returns.csv`) with 10,000 orders and 15% returns using Python, simulating product, category, geography, and marketing channel data.
2. **Data Preprocessing:** Merged datasets, handled missing values, removed outliers, and encoded categorical variables to produce `cleaned_data.csv`.
- Predictive Modeling :** *Trained a logistic regression model to predict return probabilities based on product, category, and marketing channel features.*

## **Conclusion**

The e-commerce return rate reduction analysis project successfully delivered a robust pipeline for predicting and visualizing return risks. The Power BI dashboard provides clear insights into high-risk products and trends, with interactive features enabling stakeholders to explore data by category, geography, and marketing channel. Key findings include higher return rates in certain categories (e.g., Clothing) and marketing channels (e.g., Social Media), guiding targeted interventions. Future enhancements could include real-time data integration and advanced predictive models to further optimize return reduction strategies.