

Project

E-commerce Return Rate Reduction Analysis

Introduction

In the fast-growing e-commerce sector, **product returns** impact profitability, customer experience, and logistics costs. This project aims to identify **key drivers of product returns**, analyze **return patterns by product, geography, and time**, and build a model to predict **return risk** before shipment. This allows businesses to act proactively and reduce costly returns.

Abstract

The project leverages the **UCI Online Retail dataset** to conduct a complete **data-driven return analysis**. It combines **data cleaning, exploratory data analysis (EDA), predictive modeling using logistic regression, and Power BI dashboarding**. A return probability (**risk_score**) was calculated using logistic regression, and high-risk orders were identified. These insights were visualized in an interactive dashboard, enabling decision-makers to filter return behavior by country, month, and product.

Tools Used

Tool	Purpose
Python (Pandas, Matplotlib, Seaborn, scikit-learn)	Data cleaning, EDA, modeling
SQLite	Lightweight SQL analysis

Power BI	Interactive visual dashboard
Jupyter Notebook	Development and code execution

Steps Involved in Building the Project

1. Data Cleaning

- Loaded the *Online Retail.xlsx* dataset
- Filtered out canceled orders, nulls, and invalid quantities
- Created new features: order_value, return_flag, Month, Hour

2. Exploratory Data Analysis (EDA)

- Visualized return trends by country, product, and time
- Identified top-returning items and seasonal patterns
- Grouped orders by value tiers and return rates

3. SQL Insights

- Used SQLite to calculate return rates by country, product, and time
- Extracted summary tables for modeling and Power BI

4. Logistic Regression Modeling

- Built a logistic regression model to predict return_flag
- Used features: Quantity, UnitPrice, Country, Month, etc.
- Predicted return probability (risk_score) for each order
- Exported high-risk products to a CSV for action

5. Power BI Dashboard

- Created KPIs: total returns, total orders, avg. risk score
- Charts: return risk by product, country, and month
- Interactive table with filters for high-risk orders

Conclusion

This end-to-end project successfully combined **Data science and BI** to tackle a real-world e-commerce problem: product returns. The predictive model and Power BI dashboard allow decision-makers to:

- Identify risky orders before shipment
- Target specific products or regions for improvement
- Visualize return behavior across time and categories

The approach can be extended to marketing, fraud, or logistics optimization.