E-commerce Return Rate Reduction Analysis Project Report

A Data Science Project using Python, SQL, and Power BI

# 1. Objective

This project focuses on identifying and mitigating the factors contributing to high product return rates in an e-commerce environment. High return rates negatively impact profit margins, customer satisfaction, and inventory management. By analyzing historical product return data, the aim is to uncover trends across product categories, pricing, seller performance, and customer reviews. Using statistical modeling and business intelligence, we provide a framework to predict return risks and recommend practical steps for reducing them.

# 2. Tools & Technologies

- Python: Utilized for data preprocessing, feature engineering, and implementing logistic regression for classification tasks.  
- SQL: Used for querying structured data and summarizing product-level and seller-level insights.  
- Power BI: Developed an interactive dashboard to visualize product returns, high-risk segments, and supplier performance for stakeholders.

# 3. Dataset Overview

The dataset contains approximately 500 rows, each representing a product listing. The key features include:  
- Product\_ID, Product\_Name: Unique identifiers and titles for each product.  
- Category, Seller\_Name: Categorical information used to group return trends.  
- Units\_Sold, Revenue, Price: Sales and pricing metrics used to evaluate financial impact.  
- Rating: Customer review average used to correlate satisfaction with returns.  
- Return\_Rate: The percentage of sold units returned — the primary metric for modeling.

# 4. Methodology

1. Conducted initial data exploration to assess null values, data types, and distribution of return rates.  
2. Created a binary classification target, marking entries with Return\_Rate > 0.2 as high-risk.  
3. Applied one-hot encoding on categorical variables (Category, Seller\_Name) and normalized numeric features.  
4. Built a logistic regression model using Scikit-learn to estimate the probability of a product being returned.  
5. Evaluated model performance with accuracy, precision, and F1-score; refined thresholds to maximize high-risk detection.  
6. Generated CSV outputs for high-risk products and summary insights for visualization.

# 5. Power BI Dashboard Design

- Page 1: Executive Summary  
 • KPIs showing total products analyzed, average return rate, and count of high-risk products.  
 • Bar chart comparing return rates across product categories.  
 • Pie chart displaying product distribution across sellers.  
 • Filters for interactive analysis by category and seller.  
  
- Page 2: High-Risk Products  
 • Tabular view with filtering options on predicted return risk, category, and seller.  
 • Bar charts to highlight high-risk product clusters.  
  
- Page 3: Seller Performance  
 • Visualization of average return rates per seller.  
 • Drill-through features for deeper inspection of product-level risk by seller.

# 6. Key Insights

- Products in categories like Electronics and Toys exhibited consistently higher return rates.  
- Items with low average ratings (below 3.5 stars) and high price points were more prone to being returned.  
- A few sellers had a disproportionate number of high-risk products, suggesting operational or quality issues.  
- Despite low overall returns, a small segment of products contributed to a majority of reverse logistics costs.

# 7. Strategic Recommendations

- Enforce stricter quality control and compliance checks for sellers flagged with high return ratios.  
- Improve transparency with better product descriptions, clearer images, and more accurate specifications.  
- Use predictive scoring to limit aggressive marketing on high-risk products.  
- Regularly retrain and update the prediction model with recent data for adaptive monitoring.  
- Leverage customer feedback loops to understand dissatisfaction and implement actionable improvements.

# 8. Conclusion

By leveraging machine learning, data visualization, and business intelligence, this project provides a robust and scalable framework for return risk mitigation in e-commerce. From predicting return likelihood to visualizing insights across suppliers and categories, the solution enables proactive decision-making that can lead to improved operational efficiency and customer experience.

## Prepared By

Ankit Kumar Upadhyay

Date: June 2025