



# E-commerce Return Rate Reduction Analysis



## Project Overview

This project analyses e-commerce order return patterns and predicts **return risk levels** using data analysis, machine learning, SQL, and Power BI.

The goal is to help businesses **identify high-risk return orders** and take preventive actions to reduce return rates.



---

## Objectives

- Analyse return trends across **categories, suppliers, locations, and marketing channels**
- Build a **Logistic Regression model** to predict return probability
- Classify orders into **Low / Medium / High return risk**
- Visualize insights using **Power BI dashboards**



---

## Tools & Technologies Used

- **Python** (Pandas, NumPy, Matplotlib, Scikit-learn)
- **SQL** (SQLite / SQL Server)
- **Power BI**
- **Jupyter Notebook**
- **CSV Dataset**



---

## Steps Involved

1. **Data Cleaning & Exploration**
  - Checked missing values and data types
  - Analysed return rate by category, supplier, location, and channel
2. **Feature Engineering**
  - Encoded categorical variables
  - Selected key features: price, quantity, category, supplier, marketing channel
3. **Machine Learning Model**
  - Built **Logistic Regression** model
  - Evaluated using Precision, Recall, F1-score, ROC-AUC
  - Generated **return probability**
4. **Risk Classification**

- Converted probabilities into risk levels:
  - Low (0–30%)
  - Medium (30–60%)
  - High (60–100%)

#### 5. Power BI Dashboard

- KPI cards: Total Orders, Returned Orders, High-Risk Orders
- Bar & column charts by category, supplier, channel
- Table view with return probability and risk level



## Key Insights

- **Fashion category** has the highest return rate
- **Supplier C & D** contribute more to returns
- **Paid Ads** channel shows higher return risk
- Logistic Regression effectively predicts return probability (ROC-AUC  $\approx 0.84$ )



## Final Outcome

- Successfully built an **end-to-end data analytics project**
- Combined **Python + SQL + Power BI**
- Delivered actionable insights to reduce return rates



## Dataset

The dataset contains 1,000 e-commerce orders with:

- Product, price, quantity
- Customer location
- Marketing channel
- Return flag, return probability, return risk level



## Author

**Indramma M**  
*Data Analyst*