

# E-Commerce Return Rate Reduction Analysis

---

## Abstract

Product returns are a major challenge in e-commerce, leading to increased operational costs and reduced profitability. This project focuses on analysing historical order data to identify patterns behind product returns and predict high-risk orders. Using Python for data analysis and machine learning, SQL for structured querying, and Power BI for visualization, the project provides actionable insights to help businesses reduce return rates and improve decision-making. A predictive model was built to estimate return probability, and an interactive dashboard was developed for business stakeholders.

---

## Introduction

E-commerce platforms frequently face high product return rates due to factors such as customer dissatisfaction, incorrect product selection, and logistical issues. High return rates negatively impact revenue, inventory management, and customer experience.

The objective of this project is to analyze return behavior, identify high-risk categories and locations, and provide data-driven recommendations to reduce return rates. By combining analytical techniques and visualization tools, the project aims to support strategic planning and operational improvements.

---

## Tools Used

- **Python:** Pandas, NumPy, Matplotlib, Scikit-learn
  - **SQL Server:** Data querying and aggregation
  - **Power BI:** Dashboard design and data visualization
  - **Excel / CSV:** Data storage and export
- 

## Steps Involved in Building the Project

1. **Data Collection & Cleaning (Python)**
  - Imported the e-commerce returns dataset
  - Handled missing values and corrected data types
  - Created derived fields such as return indicators

## 2. Exploratory Data Analysis (Python & SQL)

- Calculated overall return rate
- Analysed returns by product category and customer location
- Identified trends and patterns in returned orders

## 3. SQL Analysis

- Computed return rates and high-risk order counts
- Grouped data by category and location
- Validated analytical results against Python outputs

## 4. Predictive Modelling (Python)

- Built a logistic regression model to predict return probability
- Handled class imbalance using balanced class weights
- Classified orders into risk levels (High / Medium / Low)

## 5. Power BI Dashboard Development

- Created KPI cards for total orders, return rate, and high-risk orders
- Designed bar and column charts for category-wise and location-wise analysis
- Enabled drill-through for detailed high-risk order inspection

---

## Conclusion

The analysis revealed that return rates vary significantly across product categories and customer locations. Certain categories and regions consistently showed higher return risk. The predictive model successfully identified high-risk orders, enabling proactive intervention.

By integrating Python, SQL, and Power BI, the project delivers a complete analytical workflow—from raw data to actionable insights. These findings can help e-commerce businesses reduce return rates, optimize inventory, and improve customer satisfaction through targeted strategies.