

# E-commerce Return Rate Reduction Analysis

---

## Abstract

The E-commerce Return Rate Reduction project aims to identify the key factors contributing to product returns and reduce unnecessary returns by analyzing historical order and return data. By leveraging Python for data cleaning and prediction, SQL for querying datasets, and Power BI for data visualization, the project provides deep insights into return trends across regions, categories, and marketing channels. The outcome helps business teams understand high-risk return areas and make informed decisions to reduce return rates.

## Introduction

Product returns are a critical issue in e-commerce, impacting profitability and customer satisfaction. This project investigates why products are returned and identifies patterns based on region, category, and marketing channels. A logistic regression model is used to predict high-risk return products. Visual dashboards support data-driven decisions to optimize operations and reduce returns.

## Tools & Technologies Used

- Python – Data cleaning, preprocessing, and logistic regression modeling (using pandas and scikit-learn)
- SQL (PostgreSQL) – Data extraction, aggregation, joins, subqueries, and views
- Power BI – Interactive dashboards, drill-through pages, slicers, KPIs, and data storytelling
- Excel – Initial dataset review and formatting

## Steps Involved in Building the Project

### 1. Data Collection:

- Combined e-commerce order and return data from .csv files

### 2. Data Cleaning (Python):

- Removed duplicates and null values

- Converted data types (dates, numerical)
- Created return rate columns by category, supplier, region, and marketing channel

### 3. Data Analysis (SQL):

- Analyzed return rate per category, supplier, and region
- Grouped return counts by marketing channel and month
- Created views and joins to structure the dataset

### 4. Predictive Modeling (Python):

- Applied logistic regression to predict return probability
- Identified high-risk products and exported to CSV for business use

### 5. Data Visualization (Power BI):

- Built dashboards showing return rates by category, region, and channel
- Displayed total orders vs returns
- Added drillthrough and filters for interactive analysis

## Conclusion

The project successfully identified key areas where return rates were high, such as specific categories (e.g., Electronics) or marketing channels (e.g., Email). The predictive model provided early insights into potentially high-return products, helping reduce future return-related losses. Power BI dashboards allowed stakeholders to visually explore the data and take action based on the insights.