

---

# Retail Return Risk & Profitability Optimization

## Final Project Report

Asmit P. Patil

Data Analyst Intern | Skilled in Python, SQL, Power BI

Project Duration: 15 Days

Dataset: Orders, Returns, People (Superstore)

---

### Project Objective

To empower retail business stakeholders with **strategic visibility** into **return risks, customer behavior, and profitability** by analyzing operational data and building a **robust 3-page Power BI dashboard**. The goal is to combine **real-world business insight (50%)** with **technical data execution (50%)**, offering a solution that drives action, not just analysis.

---

### Business Problems

Retailers often struggle with:

- High product return rates eating into profits
  - Low visibility into customer lifetime value
  - Ineffective segmentation of high-risk customers or low-performing products
  - Lack of data-backed strategy for inventory, pricing, and marketing
- 

### Methodology

#### 1. Data Preparation

- **Datasets Used:**
  - orders.csv: Order-level transactional data
  - returned.csv: Return flags
  - peoples.csv: Regional sales rep data
- **Merging & Cleaning:**

- Created Is\_Returned flag
  - Merged return and people data
  - Calculated Days\_to\_Ship, Profit\_Margin
  - Removed invalid/missing records
- 

## 2. Exploratory Data Analysis (Python)

- **Visualizations:**
    - Heatmap (Sales, Profit, Discount, Margin)
    - Boxplots for Profit by Category
    - Region-based Return Patterns
  - **Insights:**
    - High discounts often correlate with negative profit
    - Certain sub-categories (e.g., binders, tables) show unusually high return rates
    - High-return regions and slow shipping impact profit
- 

## 3. RFM Analysis (Customer Segmentation)

- Used **Recency**, **Frequency**, and **Monetary** value to segment customers into:
    - Champions, Loyal, At Risk, Low Value
  - Derived RFM Score and grouped customers by engagement
- 

## 4. (Planned but Not Executed) Return Risk Prediction – Logistic Regression

- Attempted to train a logistic regression model to predict Is\_Returned based on:
    - Sales, Discount, Profit, Days to Ship
  - **Result:** Dropped from implementation due to imbalanced classes (only one return class present)
- 

# Power BI Dashboard Overview

## Page 1: Executive Overview

- **KPIs:** Total Sales, Total Returns, Return Rate, Net Profit %
- **Map:** Sales & Returns by Region

- **Trend:** Sales vs Returns Over Time
- **Impact:** Understand overall business health & loss due to returns

## Page 2: Product & Category Intelligence

- **Charts:**
  - Return Rate by Sub-Category
  - Profitability vs Discount
  - Top & Bottom Performing Products
- **Impact:** Spot high-risk products, optimize pricing/stocking

## Page 3: Customer & Risk Intelligence

- **Visuals:**
    - RFM Segments
    - High-Risk Customers (Returned + At Risk)
    - Sales Rep Performance
  - **Impact:** Targeted customer engagement, manage churn risk
- 

## Tools & Tech Stack

- **Python:** Data Cleaning, EDA, Clustering, RFM
  - **Pandas, Seaborn, Scikit-learn**
  - **Power BI:** DAX, Visual Storytelling, Dashboard
  - **SQL (used optionally):** for slicing data logic
  - **Git/GitHub (optional):** Version Control
- 

## Business Value Delivered

Identified **loss hotspots** and return-heavy regions

Created **data-driven customer segments** for marketing

Highlighted product categories needing price/profit optimization

Offered **risk intelligence** to reduce future returns

---