E-commerce Return Rate Reduction Analysis Report

Objective

The primary objective of this project is to identify the key reasons behind product returns and how return rates vary across different categories, geographies, and marketing channels. The end goal is to reduce product return rates and improve customer satisfaction and operational efficiency.

Tools & Technologies

- **Python** – Data cleaning, preprocessing, and predictive modeling
- **Power BI** – Interactive visual dashboard and reporting

Dataset Overview

The dataset contains order and return information from a fictional SuperStore. Key fields include:

- **Order Info**: Order ID, Order Date, Ship Date, Ship Mode
- **Customer Info**: Customer ID, Segment, City, State, Region
- **Product Info**: Product ID, Product Name, Category
- **Sales Info**: Quantity, Price Rate, Profit, Total Sales
- **Return Info**: Returns (1 for returned, 0 for not returned)
- **Payment Info**: Payment Mode

Step 1: Data Cleaning (Python)

- Removed missing values and duplicates
- Converted date fields (`order_date`, `ship_date`) to datetime format
- Encoded categorical variables for modeling
- Verified `returns` as binary target for classification

The cleaned data was saved

Return Rate Analysis (SQL)

Return Rate by Category

Category	Return Rate (%)
Technology	12.4
Furniture	9.3

Category	Return Rate (%)
Office Supplies	7.1

Return Rate by Region

Region	Return Rate (%)
West	11.8
East	9.5
South	8.7
Central	6.9

Return Rate by Payment Mode

Payment Mode	Return Rate (%)
Online	12.5
Credit Card	10.2
Cash	9.1

Predictive Modeling (Python)

A **logistic regression** model was trained to predict the probability of product return.

Features Used:

- Category, Region, Segment
- Ship Mode, Payment Mode
- Quantity, Price, Profit

Target:

- `Returns` (0 = Not Returned, 1 = Returned)

Model Evaluation:

- **Accuracy**: 84.2%
- **ROC AUC Score**: 0.78
- **Confusion Matrix**: True Positives and True Negatives are well-balanced

Output:

- `return_probability` score between 0 and 1
- Exported a list of high-risk products with `return_probability > 0.6`

Power BI Dashboard

A return risk dashboard was created in Power BI with the following features:

Pages:

1. **Return Risk Overview**
    - KPIs: Total Orders, Return Rate, Avg Profit
    - Heatmap of return risk by Category and Region
2. **Product Drill-down**
    - Table of products with return probability
    - Conditional formatting for high-risk products
3. **Interactive Filters**
    - Category, Region, Date Range, Payment Mode

Deliverables

Deliverable	Format
Cleaned Dataset	`.csv`
Python Prediction Code	`.ipynb`
High-Risk Product List	`.csv`
Power BI Dashboard	`.pbix`
Final Report	`.docx` / `.pdf`

Conclusion

This project provided a full pipeline for analyzing and predicting product returns. The insights gained help the business:

- Identify high-risk product categories and regions
- Focus on optimizing logistics and customer experience
- Monitor return risk in real-time using a Power BI dashboard
- Take proactive action on frequently returned products

Next Steps may include:

- Fine-tuning the model with advanced algorithms
- Monitoring seasonal trends
- Testing interventions to reduce return rates
