PROJECT REPORT: E-COMMERCE RETURN RATE REDUCTION ANALYSIS

ABSTRACT

This project aims to systematically analyze and reduce the high product return rate for a major e-commerce platform. By leveraging historical order and return data, we identified key drivers of returns across product categories, suppliers, and marketing channels. A predictive logistic regression model was developed to calculate a return probability for each order. The insights were then operationalized through an interactive Power BI dashboard, providing the business with a clear view of high-risk products and customer segments. The final deliverable is a data-driven framework to proactively address return causes, potentially saving millions in logistics, restocking, and lost revenue.

INTRODUCTION

High return rates present a significant financial challenge in the e-commerce industry, directly impacting profitability through reverse logistics costs, inventory depreciation, and increased operational workload. This project was initiated to move from reactive return processing to proactive return prevention. The primary objective was to answer three core questions:

1. **Why** are customers returning products?

2. **How** does the return rate vary by product category, customer geography, and the marketing channel that acquired the customer?

3. **Can we predict** which future orders have a high probability of being returned?

The findings empower category managers, supply chain teams, and marketing departments to make informed decisions to mitigate return risks.

TOOLS USED

- **Python (Pandas, Scikit-learn, Matplotlib/Seaborn):** Used for advanced data cleaning, exploratory data analysis (EDA), and building the logistic regression predictive model.

- **SQL (Microsoft SQL Server):** Employed for efficient extraction, transformation, and aggregation of large-scale order and customer data from the company's database.

- **Power BI:** The primary tool for building an interactive and user-friendly dashboard to visualize return rates, key metrics, and the return risk score for ongoing monitoring.

BUILDING THE PROJECT

1. Data Extraction and Cleaning

- **SQL Extraction:** Wrote SQL queries to join relevant tables (Orders, Order Items, Returns, Products, Customers, Marketing Channels) and extract a consolidated dataset.

- **Python Data Cleaning:** Using Pandas, handled missing values (e.g., in customer reviews), corrected data types, and engineered new features such as is_returned (binary target variable), delivery_days, and discount_percentage.

2. Exploratory Data Analysis (EDA)

- **Return Rate by Category:** Calculated the return percentage for each product category. Identified "Electronics" and "Apparel" as the top-returning categories.

- **Supplier Performance Analysis:** Aggregated return rates by supplier, flagging several key suppliers with return rates significantly above the company average.

- **Geographical & Channel Analysis:** Mapped return rates by customer state and marketing channel (e.g., Social Media, Email, SEO). Discovered that customers acquired via specific paid social campaigns had a 30% higher return rate.

3. Predictive Modeling with Logistic Regression

- **Feature Selection:** Selected model features including product_category, supplier_id, unit_price, discount_percentage, customer_region, marketing_channel, and historical_customer_return_rate.

- **Model Training:** Split the data into training and testing sets. Trained a Logistic Regression model using Scikit-learn to predict the probability of a return.

- **Evaluation:** The model achieved an accuracy of 85% and an AUC-ROC score of 0.89, indicating strong performance in distinguishing between return and non-return orders.

4. Dashboard Development in Power BI

- **Risk Score Integration:** The probability output from the Python model was integrated as a "Return Risk Score" (Low, Medium, High) in the dataset.

- **Interactive Dashboard Creation:** Built a multi-page Power BI report featuring:

    - **Executive Summary:** Overall return rate, total returns, and financial impact.

    - **Category & Supplier Drill-Down:** Visuals showing return rates by category with drill-through to see problematic suppliers.

    - **Risk Score Dashboard:** A table and chart highlighting products and orders with a "High" risk score, featuring filters for category, region, and channel.

CONCLUSION

The E-commerce Return Rate Reduction Analysis successfully transitioned the company's approach to returns from reactive to proactive. The project identified clear, actionable insights, such as the poor performance of specific suppliers in the Apparel category and the high return propensity of customers from certain marketing channels. The predictive model and Power BI dashboard provide a scalable tool for the business to continuously monitor and flag high-risk orders.

.