E-commerce Return Rate Reduction Analysis

Introduction

In the rapidly evolving world of online retail, high return rates significantly impact profitability and customer experience. This project aimed to investigate the key reasons behind product returns in an e-commerce platform and to identify high-risk categories, locations, and customer segments. The goal was to provide actionable insights through predictive modelling and interactive dashboards to help reduce unnecessary returns and improve operational efficiency.

🎯 Abstract

The analysis focused on a dataset containing 30,000 orders and their return-related details, including customer demographics, payment methods, product categories, and return reasons. Through data cleaning, exploratory data analysis, logistic regression modelling, and Power BI visualization, we identified patterns in return behaviour. We calculated return probabilities, highlighted high-risk products, and built dashboards to explore return trends by category, geography, age, gender, pricing, and discounts. Our model and visual insights aim to help stakeholders optimize sales strategies, product listings, and logistics operations.

🛠 Tools Used

- **Python**: Data preprocessing, logistic regression model
- **SQL**: Data extraction, filtering, grouping
- **Power BI**: Visual analytics, dashboard creation
- **Pandas, Scikit-learn**: For data manipulation and predictive modelling

---

🔍 Steps Involved in Building the Project

1. **Data Cleaning & Preprocessing**
    - Handled missing values in Days_to_Return, standardized column types
    - Encoded categorical fields for model compatibility

2. **Exploratory Data Analysis**
    - Identified top return reasons: *Defective*, *Wrong item*, *Changed mind*
    - Analysed return rates by location, price, category, and discounts
3. **Feature Engineering**
    - Created new columns: Return_Probability, Is_High_Risk, age groups
    - Categorized price brackets and order quantity segments
4. **Logistic Regression Modelling**
    - Built a model to predict return probability using features like Category, Discount, Price, and Shipping_Method
    - Evaluated the model and exported high-risk product list
5. **Power BI Dashboard**
    - Designed 3-page dashboard:
        - Page 1: Return overview (top reasons, category & location)
        - Page 2: Price, discount, quantity-based risk
        - Page 3: Customer behaviour by age, gender, payment type
    - Enabled drill-through filters and KPI cards

✅ Conclusion

This project successfully identified core factors driving high return rates in an e-commerce setting. Clothing, Electronics, and Books emerged as the most return-prone categories, with certain locations and customer age groups (above 50) contributing disproportionately to returns. The logistic regression model helped predict return risk, while Power BI dashboards presented insights interactively. These findings can guide product listing optimization, quality assurance, and customer targeting strategies to minimize returns and enhance user satisfaction.