E-commerce Return Rate Reduction Analysis

Introduction

In today's fast-paced e-commerce world, product returns are a growing challenge for businesses. They not only increase costs but can also impact customer trust and satisfaction. This project was all about understanding why customers return products and how those return rates differ across product categories, suppliers, locations, and marketing channels. By using data analysis and machine learning, we aimed to uncover insights that could help reduce return rates and improve customer experience.

Abstract

We started by exploring and cleaning datasets containing order and return details. Then, we looked closely at return percentage breaking them down by category, supplier, region, and more. To take it a step further, we built a logistic regression model that predicts the chances of a product being returned. Using Power BI, we created an interactive dashboard that visualizes key patterns and lets users drill down into the data by different filters. Finally, we generated a list of high-risk products those most likely to be returned to help businesses take proactive steps.

Tools Used

- **Python** – for data cleaning, analysis, and building the prediction model.
- **Power BI** – to create an engaging dashboard for visualizing return patterns and risks.

Steps Involved in Building the Project

1. **Data Cleaning**

   - We cleaned both the order and return datasets by handling missing values, removing duplicates, and standardizing inconsistent entries.
   - Special attention was given to fields like product categories, regions, and supplier names to ensure uniformity.

2. **Exploratory Data Analysis**

   - Using Python, we examined return trends across different segments.
   - Key insights included which categories had the highest return rates and whether certain suppliers or regions had more issues.
   - Visualizations helped us spot patterns and anomalies more easily.

3. **Predictive Modelling**

   o A logistic regression model was developed to predict whether a product would be returned.

   o We used features like product price, category, supplier rating, shipping duration, and region.

4. **Dashboard Development**

- The Power BI dashboard was built to make the findings accessible and interactive.
- It includes visual breakdowns of return rates by different filters and a return risk score for each product.
- Users can drill down by supplier, product type, or location to explore data in more detail.

5. **Identifying High-Risk Products**

- o We filtered products with a high predicted return probability and exported them into a CSV file.

- o This list can help businesses make informed decisions like reviewing product descriptions, improving packaging, or communicating better expectations.

Conclusion

This project showed how data and machine learning can be powerful tools in solving real-world business problems. By identifying the key drivers behind product returns and predicting which products are most at risk, businesses can take meaningful action to reduce return rates. The interactive dashboard adds value by allowing users to explore and interpret the data with ease. Overall, this project helps improve not just profits but also customer satisfaction making it a win-win for both the business and its users.