E-commerce Return Rate Reduction Analysis

1. Introduction

This project focuses on analyzing customer return behavior using e-commerce order data. The main objective is to identify patterns in returned products and predict future return probabilities based on features like product category, supplier, price, region, and marketing channel.

2. Abstract

We performed data cleaning and exploratory data analysis (EDA) to identify which product categories and suppliers experience the highest return rates. Logistic Regression was used as a classification model to predict the probability of a product being returned. The model was evaluated using confusion matrix, accuracy score, and ROC AUC score.

3. Tools Used

- Python (Pandas, Seaborn, Scikit-learn, Matplotlib)
- Jupyter Notebook / Google Colab
- CSV Dataset of 500 orders

4. Steps Involved

1. Loaded and cleaned the dataset (500 rows).
2. Conducted EDA to find category-wise and supplier-wise return rates.
3. Encoded categorical variables for modeling.
4. Trained Logistic Regression model to classify return likelihood.
5. Evaluated model using confusion matrix, classification report, and ROC AUC.
6. Identified high-risk categories and suppliers.
7. Suggested using Power BI dashboard for visual storytelling.

5. Conclusion

The analysis revealed that certain product categories and suppliers consistently showed higher return percentages. The logistic regression model provided reasonable accuracy in predicting returns. This model can help businesses focus on reducing return rates by improving product quality or refining marketing strategies.