# E-commerce Return Rate Reduction Analysis

## Introduction

Product returns are a major challenge in e-commerce, impacting logistics, inventory management, and customer satisfaction. This project aims to build a machine learning model that predicts whether a product will be returned based on various factors such as delivery time, product category, vendor, and customer behaviour. Accurate predictions can help optimize operations and reduce return-related losses.

## Abstract

The goal of this project is to classify whether a product will be returned using historical order data. The dataset includes features like delivery days, vendor region, order timing, product category, and delivery status. After cleaning the data and handling leakage issues, a Logistic Regression model was trained and evaluated using ROC AUC score and classification metrics. The final model achieved an accuracy of 87% with a ROC AUC score of approximately 0.79, indicating a reasonably effective performance.

## Tools

- **Language**: Python
- **Tools**: Power BI (Dashboard)
- **Libraries**: pandas, (scikit) sk-learn,
- **IDE**: Jupyter Notebook / VS Code
- **Model**: Logistic Regression(also used  Random Forest Classifier)
- **Evaluation Metrics**: ROC AUC Score, Precision, Recall, F1-Score

## Steps Involved in Building the Project

1. **Data Collection and Cleaning**: Imported a structured dataset containing order and delivery details.
2. **Data Preprocessing**: Removed target leakage columns like Return Reason

Handled missing values and converted categorical columns using encoding

3. **Feature Engineering**:
    - Created new features like Delivery time (Days)
    - Used get_dummies() for one-hot encoding categorical variables
4. **Model Training**:
    - Applied train_test_split() and used Logistic Regression Model
    - Trained the model on real features, avoiding leakage
5. **Model Evaluation**:
    - Achieved ROC AUC Score of **0.7935**
    - Accuracy of **87%**
    - Reasonable performance confirmed through classification_report and confusion_matrix
6. **Dashboard Creation**
   To complement the predictive model, an interactive dashboard was developed using Power BI. The dashboard includes

    - Impactful visuals to easily understand the data Visually
    - Comparing the data based on different attributes
    - Understanding the Return Risk of the products and analysing to reduce the return rate

## Conclusion

The project successfully demonstrates how machine learning can be applied to predict product returns in e-commerce. After resolving target leakage issues and performing effective feature engineering, the model achieved good predictive accuracy and reliability. The approach can be further improved by integrating customer-level data, seasonality patterns, and optimizing the model using hyperparameter tuning. Finally a interactive Dashboard to easily understand and visualize the data.