E-commerce Return Rate Reduction Analysis

Introduction :

In today's competitive online retail world, customer product returns can greatly impact profit, logistics, and customer satisfaction. Knowing the main reasons behind returns helps businesses improve their delivery operations, product quality, and marketing strategies. This project focuses on analyzing and predicting product return patterns using Python, SQL, and Power BI. The goal is to pinpoint why customers return products and how return rates differ by category, supplier, geography, and marketing channel.

Abstract:

The project uses data-driven methods to lower return rates in e-commerce. By analyzing historical order and return data, it identifies trends in delivery delays, shipping methods, product types, and customer demographics that lead to higher return chances. A logistic regression model is applied to estimate the likelihood of a return. Power BI dashboards visually display high-risk categories and suppliers. The insights gained can inform inventory management, supplier choice, and customer engagement strategies to cut operational losses and build customer trust.

Tools Used:

Python: For data cleaning, feature engineering, and machine learning modeling

Power BI: For creating interactive dashboards and visual analytics

Libraries: pandas, numpy, matplotlib, seaborn, scikit-learn

Dataset: E-commerce order and return dataset (Train.csv)

Steps Involved in Building the Project

Phase 1: Data Cleaning

Loaded the raw dataset and removed unnecessary columns, such as ID.

Created a new binary column, is_returned, to indicate if a product was returned.

Handled missing values and standardized categorical fields like Warehouse block, Mode of Shipment, Product Importance, and Gender.

Phase 2: Return Rate Analysis

Calculated overall return percentages and those by category.

Visualized return distribution across suppliers, geographies, and shipping methods using Python plots.

Exported summary tables for Power BI analysis.

Phase 3: Predictive Modeling

Built a logistic regression model to predict the chance of a return.

Split the dataset into training and testing subsets.

Evaluated model performance using accuracy, precision, and ROC-AUC metrics.

Exported a list of high-risk products (predicted probability greater than 0.7) as a CSV file for business review.

Phase 4: Power BI Dashboard

Designed an interactive dashboard showing:

- Return percentage by category, supplier, and shipment method
- Risk score visualization by region
- Enabled slicers and filters for managers to explore specific categories or suppliers dynamically.

Conclusion:

The project clearly shows how data analytics and predictive modeling can help e-commerce businesses better manage product returns. By identifying high-risk products and suppliers, companies can enhance their quality control, streamline logistics, and make better marketing decisions. The Power BI dashboard offers an intuitive, real-time view of operational issues, while the logistic regression model provides a scalable way to predict returns. Overall, this prototype serves as a strong basis for a ready-to-use return reduction system that can save costs and boost customer satisfaction.