Introduction:

This analysis focuses on reducing return rates in e-commerce by identifying key factors influencing customer returns. Using data-driven insights, it aims to enhance customer satisfaction, optimize inventory management, and improve profitability. The study leverages statistical techniques and visualizations to recommend strategies for minimizing product returns and improving operational efficiency.

Abstract:

This project analyzes e-commerce return rates to identify the primary causes behind product returns and propose effective reduction strategies. By examining customer behavior, product attributes, and return patterns, the analysis provides insights into return trends. Using data visualization and statistical methods, the study aims to improve customer satisfaction and operational efficiency. The findings help e-commerce businesses reduce return-related losses, enhance product quality, and optimize marketing and logistics strategies for better decision-making and increased profitability.

Tools:

Jupyter Notebook:

Jupyter Notebook was used for data cleaning, exploration, and modeling. Its interactive environment enabled step-by-step data analysis with Python, including visualization using libraries like seaborn and matplotlib. It facilitated efficient debugging, annotation, and execution of code blocks, making the analytical workflow transparent, reproducible, and easy to share.

MySQL:

MySQL served as the primary database system for storing and querying the e-commerce dataset. It helped perform structured queries to extract relevant features, filter conditions, and aggregate metrics. MySQL ensured efficient data management, allowing for scalable and organized data retrieval essential for accurate analysis and further processing.

Power BI:

Power BI was used to create interactive dashboards and visualizations for the return rate analysis. It enabled a clear understanding of trends, customer behavior, and return patterns through charts and graphs. The tool allowed stakeholders to explore insights dynamically and make data-driven decisions to reduce return rates.

Description:

1. **Data Import and Cleaning:**
   Loaded the e-commerce dataset using Pandas and handled missing values by filling or replacing them with appropriate defaults (e.g., median, mode).

2. **Data Preprocessing:**
   Converted date columns to proper datetime format and created new indicators like Is_Returned to flag returned orders.

3. **Feature Engineering:**
   Extracted new insights such as supplier information from product IDs and calculated return percentages by product category and supplier.
4. **Exploratory Data Analysis (EDA):**
   Visualized return patterns using bar charts to identify high-return categories and suppliers.
5. **Label Encoding:**
   Encoded categorical features into numeric values using LabelEncoder to prepare for machine learning.
6. **Train-Test Split:**
   Split the dataset into training and test sets for model evaluation using train_test_split.
7. **Random Forest Modeling:**
   Built a Random Forest classifier to predict returns, then evaluated it using a classification report, confusion matrix, and feature importance plot.
8. **Logistic Regression Modeling:**
   Created a machine learning pipeline using LogisticRegression with preprocessing (standardization and one-hot encoding), and evaluated the model with ROC AUC and confusion matrix.
9. **Feature Interpretation:**
   Interpreted logistic regression coefficients to understand which features most significantly influence return probability.
10. **Order Probability Prediction:**
    Predicted the return probability for a sample order and displayed the likelihood and influencing factors.

The SQL file performs comprehensive analysis on the **ecommerce** dataset. It includes queries to calculate total and average revenue, return rates by product category and discount range, and customer segmentation by age and gender. It also analyzes product pricing, payment methods, and shipping methods to identify trends and patterns influencing e-commerce performance and returns.