Amazon Return Rate Reduction Analysis

Introduction

This project aims to analyze Amazon order data to identify return trends and predict the likelihood of product returns using Machine Learning (Logistic Regression). Additionally, a Power BI dashboard was created to visualize returns, high-risk products, and payment method distributions.

Abstract

A complete end-to-end workflow was executed:

- SQL was used for initial data extraction and cleaning.
- Python was used for modelling return probability using Logistic Regression.
- Power BI was used for creating an interactive dashboard to track key metrics like return rate by State, City, Payment Method.

Tools Used

- SQL — for querying
- Python(Jupyter Notebook) — for Logistic Regression Model
- Power BI — for dashboard creation
- Libraries: Pandas, matplotlib, seaborn, sklearn

Steps Involved

1. Data Cleaning (Python)

- Imported Amazon order dataset using `pandas`
- Cleaned data: Removed duplicates, Handled missing/null values
- Formatted columns like `order_date`, `payment_mode`, `is_returned`
- Performed Exploratory Data Analysis (EDA) with `matplotlib` and `seaborn`

2. Data Preparation (SQL)

- Imported Amazon Cleaned dataset.
- Ran some SQL queries to understand the return rates based on some key metrics.

3. Predictive Modeling (Python)

- Built Logistic Regression model

- Evaluated with accuracy score, confusion matrix
- Identified High- Risk Products

Result: Achieved ~78% accuracy in predicting whether an order will be returned.

4. Power BI Dashboard

- Imported cleaned dataset into Power BI.
- Created DAX Measures
- Built Visuals: KPI Cards, Bar chart, Line Chart, Pie Chart, Slicers etc.
- Drill-Through Feature: Added “State Details” page
- Design: Used Amazon branding colors

Conclusion

This project covers the complete workflow from data extraction to predictive modeling and dashboard visualization. Using SQL, Python, and Power BI, I analyzed Amazon's return trends, predicted return probabilities, and provided actionable insights to support better inventory and customer management.