📄 E-commerce Return Rate Reduction Analysis — Project Report

---

1. Introduction

Product returns in e-commerce not only increase operational costs but also affect customer satisfaction and profitability. This project aims to analyze return trends in an e-commerce dataset and identify high-risk products and customer segments responsible for frequent returns. A return risk prediction model is built using logistic regression, and an interactive dashboard is developed in Power BI to visualize and monitor return behavior.

---

2. Abstract

This project leverages Python, SQL, and Power BI to analyze customer return behavior in an e-commerce dataset. It focuses on identifying key contributors to returns such as product category, supplier, geography, and marketing channel. A logistic regression model is trained to predict the probability of a return, and a Return Risk Score is generated for each transaction. The results are visualized in a Power BI dashboard, enabling decision-makers to drill down into high-risk areas and take preventive measures.

---

3. Tools Used

- **Python**: Data cleaning, feature engineering, logistic regression modeling, CSV generation
- **SQL**: Initial data exploration and summarization (if needed)
- **Power BI**: Return risk dashboard with filters and drill-through functionality
- **Jupyter Notebook / VS Code**: Python code execution and model building

---

4. Steps Involved in Building the Project

1. **Data Cleaning**:
   Removed null values, handled duplicates, and formatted columns for consistency.
2. **Exploratory Data Analysis (EDA)**:
   - Analyzed return rates by Category, Geography, Payment Method
   - Identified high return rates in specific product categories and regions
3. **Feature Engineering**:
   - Encoded categorical variables
   - Extracted customer age groups, simplified product names, and aggregated past returns
4. **Model Building (Python)**:
   - Used logistic regression to predict return probability

- Evaluated model using accuracy, precision, recall, and F1-score
- Exported final dataset with Return_Probability as return_risk_score_data.csv

5. **Dashboard Development (Power BI)**:
    - Imported prediction CSV
    - Built bar charts for Return Risk by Category & Geography
    - Table with conditional formatting for high-risk products
    - Added slicers (Category, Geography, Payment Method) and drill-through to product detail

---

5. Conclusion

The analysis revealed that certain categories and regions have significantly higher return rates. The logistic regression model effectively predicts return probability based on order features. The Power BI dashboard enables quick identification of risky products and customer segments, helping businesses take corrective actions such as adjusting return policies or improving product descriptions.

This integrated approach of using Python for modeling and Power BI for visualization offers a scalable and data-driven solution for e-commerce return reduction.
