Final Project Summary

1. Introduction

In the competitive world of e-commerce, product returns are a significant challenge. High return rates impact profits, supply chains, and customer satisfaction. This project focuses on analyzing return patterns and predicting the likelihood of a product being returned based on order attributes, geography, and customer behaviour.

2. Abstract

This project aims to:

- Analyze product return patterns across various categories and regions.
- Identify key factors influencing returns such as category, delivery time, and marketing channels.
- Build a logistic regression model to predict return probability for each order.
- Create an interactive Power BI dashboard with drill-through filters to visualize return risks.
- Highlight high-risk products for proactive interventions.

The dataset used contains historical order details, including product category, geography, price, quantity, delivery time, and whether the product was returned along with the reason.

3. Tools Used

- **Python (Pandas, Scikit-learn)**: For data cleaning, feature engineering, and logistic regression modelling.
- **Power BI**: For building interactive dashboards with slicers, KPIs, and drill-through pages.
- **Excel**: Initial data formatting and file integration.
- **SQL**: Used for querying return rates by dimension for faster aggregation.

4. Steps Involved in Building the Project

1. **Data Preparation**
    - Load and clean the data
    - Handle missing values and convert dates
    - Feature engineering using quantity, price, delivery time, and encoded categories
2. **Exploratory Data Analysis (EDA)**
    - Return % by Category, Geography, and Marketing Channel
    - Top return reasons
    - Visualization of overall return trends
3. **Predictive Modeling**
    - Logistic Regression trained using encoded features
    - Model evaluation using classification report
    - Risk score added to each order as Return_Probability
4. **Dashboard Creation in Power BI**
    - Import scored data
    - Visuals: Category-wise risk, geography maps, channel performance
    - Drill-through setup to explore order-level insights
5. **High-Risk Product Export**
    - Orders with risk score > 0.7 exported to CSV for proactive action

5. Conclusion

This project delivers both analytical and predictive capabilities to reduce return-related losses in e-commerce. By identifying high-risk orders and visualizing key trends through Power BI, businesses can take preventive steps—such as modifying delivery policies, improving sizing information, or targeting certain regions with quality improvements.

The combination of **data science and business intelligence** empowers stakeholders to make informed, data-driven decisions that optimize returns management.