IBI Internship Project – Customer & Sales Data Analysis

By: Jagadeeswar R.

# Executive Summary

This project analyzes a retail dataset containing customer, product, and transaction data. The goal is to generate business insights, forecast future sales, segment customers, and predict churn. Advanced techniques such as Gaussian Mixture Models, Prophet forecasting, Random Forests, and Association Rule Mining were applied.

# Dataset Overview

The dataset includes fields like Customer ID, Age, Gender, Product Category, Price, Discount, Quantity, Date of Purchase, Payment Method, and Invoice Number. This dataset enables segmentation, lifetime value estimation, and behavior modeling.

# Data Preprocessing & Feature Engineering

- Missing values were handled using Iterative Imputer.  
- Outliers were clipped using Tukey’s method.  
- Features were scaled using Min-Max Scaler.  
- New features included: TotalPrice, Recency, Customer Lifetime Value (CLV), and Loyalty Score.

# Exploratory Data Analysis

Time-based trends showed seasonal spikes in monthly sales. Recency and Frequency metrics highlighted customer engagement patterns.

# Customer Segmentation

Using Gaussian Mixture Models, customers were grouped into four distinct segments. These segments differ by frequency, monetary value, and recency scores. This allows targeted marketing strategies to be created for each group.

# Sales Forecasting

Prophet model was used to forecast future sales. The forecast includes daily trends and a 3-month outlook. Business can prepare for sales dips and peaks based on this analysis.

# Churn Prediction

A Random Forest classifier predicted whether a customer is likely to churn. The model achieved good accuracy. This insight helps prioritize retention campaigns for high-risk customers.

# Market Basket Analysis

Apriori algorithm discovered product associations for cross-selling. Rules with high lift and confidence were found, suggesting optimal product bundles.

# Strategic Recommendations

- Focus on 'loyal' customer segments with high CLV.  
- Target potential churners with exclusive discounts.  
- Prepare for forecasted demand spikes with inventory alignment.  
- Create product bundles using association rules to boost cart size.