

# Ecommerce Customer Behavior Analysis

## Project Overview

This project analyzes ecommerce customer behavior to identify patterns of churn, optimize decision-making, and enhance customer retention strategies. Using Python and machine learning, we explored behavior trends, built predictive models, and evaluated churn-related factors.

## Exploratory Data Analysis (EDA)

- Univariate Analysis: Explored distributions of Age Group, Product Category, Gender, Payment Method, and Purchase Weekday.
- Bivariate Analysis: Compared churn rates across categories using grouped visualizations.
- Returns: Visualized top categories with high return rates.

## Feature Engineering

- Calculated Average Order Value (AOV) per customer.
- Defined 'Customer Type': One-time vs Repeat.
- Encoded Weekday and Weekend purchases.
- Created interaction features (e.g., Payment × Product Category).

## Handling Imbalance with SMOTE

Used SMOTE (Synthetic Minority Over-sampling Technique) to balance the training dataset and improve the classifier's ability to detect minority churn cases.

## Models Built & Evaluated

- Logistic Regression: Moderate performance with low recall on churners.
- Decision Tree: Accuracy = 80%, but very low churn recall.
- Random Forest: Improved recall for churners after SMOTE, but accuracy dropped to ~39%.
- ROC AUC Score (Random Forest): 0.49

## Key Takeaways

- 'Average Order Value', 'Purchase Weekday', and 'Customer Type' were strong churn indicators.
- Handling imbalance improved model sensitivity but affected precision.
- Additional tuning or advanced models (e.g., Gradient Boosting) may enhance performance.