

Customer Segmentation And Predicting Behavior

1. Problem Statement

This project focuses on predicting customer behavior by classifying customers into pre-defined segments based on demographic and behavioral features. These segments help in targeted marketing and loyalty strategy development.

2. Dataset Overview

The dataset consists of household-level demographic and transactional features such as SEC, Affluence Index, Age, Gender, Education, Promo Usage, and Purchase Volume. The goal is to assign each customer to a meaningful segment like Loyalists, Variety Seekers, or Promo Shoppers.

3. Tools and Techniques Used

- Languages: Python
- Libraries: pandas, NumPy, scikit-learn, XGBoost, matplotlib, seaborn, joblib
- Environment: Jupyter Notebook
- Deployment: Flask API for real-time customer input and prediction

4. Data Preprocessing

- Categorical variables encoded using one-hot and label encoding
- Numerical features standardized using StandardScaler
- Train-test split performed with stratification to preserve segment distribution

5. Exploratory Data Analysis

- Cluster profiling showed distinct differences in Promo Usage, Purchase Volume, and Loyalty
- Demographic insights: Loyalists skewed older and more affluent
- Visualizations helped validate cluster labeling

6. Model Building

- Supervised classification using Logistic Regression, Random Forest, and XGBoost
- Evaluated using accuracy, precision, recall, and confusion matrix
- Feature importance and SHAP values used for explainability

7. Evaluation Metrics

Best model: XGBoost Classifier

Accuracy: 0.89

Precision by Segment:

- Loyalists: 0.90
- Variety Seekers: 0.87
- Promo Shoppers: 0.86

8. Key Takeaways

- XGBoost outperformed other models in overall accuracy and segment-specific precision
- Feature importance revealed that Affluence Index, Promo Usage %, and Total Value were key drivers
- Flask app successfully deployed for real-time scoring of individual customers

9. Resources

[Kaggle Notebook](#)

[GitHub Repo](#)

[Flask App Demo](#)