Predicting Customer Response to a Gold Membership Offer:

A Machine Learning Approach for Smarter Campaign Targeting

Bracha Stein

July 28, 2025

Abstract

This study addresses the challenge of optimizing customer targeting for a retail superstore's Gold Membership marketing campaign. Identifying likely responders is critical for reducing wasted outreach and maximizing return on investment.

The objective was to develop and evaluate a predictive model capable of identifying customers most likely to accept a discounted Gold Membership offer.

The analysis used a dataset of 2,240 customer records containing demographic, transactional, and behavioral features. Three classification models—Logistic Regression, Random Forest, and XGBoost—were trained and compared. Feature engineering and train-test splitting were applied, and SHAP values were used to interpret feature importance.

Among the models, XGBoost achieved the strongest performance, with an AUC of 0.8771, precision of 0.60, and recall of 0.48, outperforming the alternatives in balancing accuracy and interpretability. Key predictors included recency, household composition, and expenditure on wine, meat, and gold products.

The results demonstrate that machine learning can meaningfully improve marketing precision by identifying lapsed or high-spending customers who are most likely to respond. While the model is effective, challenges such as class imbalance and the absence of external behavioral data may limit generalizability beyond the dataset.

In conclusion, this machine learning approach offers a practical and scalable method for improving campaign targeting in retail marketing. Future enhancements could involve A/B testing, real-time prediction pipelines, and integration of customer lifetime value to better prioritize outreach.

Introduction

In today's data-saturated business environment, predictive analytics has become an essential tool for optimizing marketing strategies. By analyzing historical and behavioral data, organizations can anticipate customer behavior, streamline outreach, and enhance return on investment (Siegel, 2013; Delen, 2020). Predictive models enable precise targeting, reducing campaign costs while improving effectiveness (Provost & Fawcett, 2013; Leventhal, 2018). As personalization becomes vital for competitiveness, data-driven decision-making has evolved from a strategic advantage into a fundamental necessity (Sridhar & Mittal, 2021; Davenport & Harris, 2007).

This project applies these principles to a real-world case: developing predictive models to identify which customers are most likely to accept a discounted gold membership offer from a superstore campaign. By leveraging logistic regression, random forest, and XGBoost classifiers, the study aims to optimize campaign performance through data-informed targeting.

A wide range of machine learning and statistical techniques—such as logistic regression, decision trees, and ensemble models—are widely applied in marketing analytics to predict customer behavior (Delen, 2020; Hair et al., 2019). Research supports the effectiveness of decision trees when enhanced with feature selection and class balancing for improving campaign outcomes (Lemke et al., 2009). Clustering and dimensionality reduction methods, including k-medoids and principal component analysis (PCA), also play key roles in customer segmentation and behavioral pattern recognition (Tsiptsis & Chorianopoulos, 2009). Traditional frameworks such as Recency-Frequency-Monetary (RFM) analysis continue to serve as robust foundations for segmenting and targeting customers (Kumar & Reinartz, 2016).

Advancements in analytics platforms have made these predictive methods more accessible. Artificial intelligence–integrated tools now enable marketers to apply complex algorithms without needing deep technical expertise (Chatterjee et al., 2020). Even spreadsheet-based systems can execute segmentation and regression techniques for campaign planning (Winston, 2014). Platforms like R offer user-friendly implementations of supervised learning methods, enhancing analytical efficiency for applied projects (Dinov, 2023). To ensure practical utility, structured forecasting techniques remain essential for validating model accuracy and translating predictions into actionable insights (Armstrong, 2001).

Modern marketing strategy increasingly emphasizes identifying and prioritizing high-value customers to maximize profitability. Customer lifetime value (CLV)–driven targeting has emerged as a cornerstone of sustainable practice (Fader & Toms, 2018), in line with broader shifts toward customer-centric frameworks (Sridhar & Mittal, 2021). While both acquisition and retention are vital (Leventhal, 2018), this study focuses specifically on improving acquisition by predicting campaign responders.

Finally, foundational literature underscores the need to align analytics efforts with business goals to ensure measurable impact (Provost & Fawcett, 2013; Siegel, 2013). This study builds on that foundation by applying proven techniques—feature selection, class balancing, and evaluation metrics—to construct models that improve targeting accuracy, reduce campaign costs, and enhance overall marketing return on investment.

Methods

To predict which existing customers were most likely to respond positively to a discounted Gold Membership offer, a supervised machine learning classification framework was implemented. The primary objective was to identify high-probability responders to enhance marketing efficiency and maximize return on investment. Three classification models were evaluated: XGBoost Classifier, Logistic Regression, and Random Forest Classifier.

The dataset, obtained from a publicly available source on Kaggle, contained detailed demographic, behavioral, and transactional information from a retail superstore. It consisted of 2,240 entries with 22 features, including variables such as Year_Birth (age proxy), education level, marital status, income, and household composition (Kidhome and Teenhome). Additional variables included product category spending (e.g., MntWines, MntMeatProducts, MntGoldProds), digital engagement metrics (e.g., NumWebVisitsMonth), and recency of last purchase (Recency). The binary target variable, *Response*, indicated whether a customer accepted the membership offer (1) or did not (0). The dataset exhibited class imbalance, with only approximately 15% of customers responding positively.

Prior to modeling, comprehensive data cleaning and preprocessing procedures were applied. Missing income values were imputed using the median to preserve distribution characteristics. The Dt_Customer variable was converted to datetime format, and a new feature, Year_Customer_Joined, was derived to represent customer tenure. Categorical variables, including Education and Marital_Status, were one-hot encoded using get_dummies with drop_first=True to mitigate multicollinearity. Additionally, a simplified *MaritalStatus* feature was engineered for visualization purposes, consolidating the most prominent marital status categories.

Feature engineering incorporated both original and derived features. Notably, *Year_Customer_Joined* and *Recency* served as proxies for customer loyalty and engagement. Spending-related features across multiple product categories were retained as continuous variables to capture purchasing behavior. Dimensionality reduction techniques such as PCA were intentionally omitted to preserve model interpretability.

Extensive exploratory data analysis (EDA) was performed to uncover feature relationships and response patterns. Count plots illustrated the degree of class imbalance, while box plots and bar plots were used to explore differences in income, web visits, and product spending across response categories and marital groups. Correlation heatmaps were employed to assess multicollinearity, and line plots depicted response trends by customer join year. Scatter plots compared spending behaviors between responders and non-responders. These insights played a pivotal role in feature selection and model interpretation.

Model selection leveraged the complementary strengths of three algorithms. XGBoost was selected for its effectiveness with structured and imbalanced datasets and was configured with use_label_encoder=False and eval_metric='logloss'. Logistic Regression served as a simple, interpretable baseline and was implemented within a pipeline that included StandardScaler and class_weight='balanced'. The Random Forest Classifier, configured with 100 estimators and balanced class weights, was employed as a robust ensemble method.

All models were trained and evaluated using a stratified train-test split (70% training, 30% testing, random_state=42) to maintain consistent class proportions. Model performance was assessed using metrics suitable for imbalanced classification, with an emphasis on precision, recall, and F1-score for the positive class. ROC-AUC was prioritized as the primary evaluation metric due to its resilience to imbalance. Additional metrics, such as accuracy, confusion matrices, and full classification reports, were also reported to provide a comprehensive evaluation.

Model interpretability was an essential aspect of the analysis. SHAP (SHapley Additive exPlanations) values, generated using the TreeExplainer method, were used to assess feature importance in the XGBoost model. SHAP summary plots revealed both global and local feature contributions. Additionally, XGBoost’s built-in feature importance visualization was used to identify the top 15 predictors. Confusion matrices and ROC curves were plotted for all models to support comparative evaluation.

The analysis was conducted in Google Colab using Python 3.11. Key libraries included pandas and numpy for data manipulation, matplotlib and seaborn for visualization, scikit-learn for modeling and evaluation, xgboost for gradient boosting, and shap for model explainability. To ensure reproducibility, the random seed was consistently set to 42 throughout the entire workflow.

Results

This analysis aimed to identify key predictors of customer response to a Gold Membership offer and evaluate the effectiveness of machine learning models in predicting campaign acceptance. The dataset consisted of 2,240 customer records with demographic, behavioral, and transactional features.

Summary statistics revealed a wide range of customer ages, income levels, and purchasing behaviors. Missing income values were imputed with the median to maintain consistency. A histogram of the target variable showed a class imbalance, with only 15% of customers accepting the membership offer (Response = 1). Visualizations such as box plots, bar charts, and correlation heatmaps helped explore feature distributions and relationships. For example, income tended to be higher among responders (Figure 1), and wine spending was markedly higher in this group (Figure 2). Additionally, heatmaps confirmed moderate correlations among spending variables, prompting the use of tree-based models for robustness to multicollinearity.

The response rate was 15%, indicating a minority of customers accepted the offer. This imbalance was addressed through stratified sampling in training/test splits and by using model evaluation metrics beyond accuracy. Response rates also varied by features such as marital status and year joined, with some groups (e.g., customers who joined in earlier years) showing slightly higher response rates (Figure 3).

Three models were developed: XGBoost, Logistic Regression, and Random Forest. Each was evaluated using accuracy, precision, recall, F1-score, and ROC AUC on a held-out test set. XGBoost achieved the best overall performance, with an accuracy of 88%, precision of 0.60, recall of 0.48, and an ROC AUC of 0.8771 (Table 1). Logistic Regression, despite using class weighting, showed high recall for responders (0.81) but at the cost of precision (0.35), leading to a lower overall accuracy of 75% and ROC AUC of 0.8618. Random Forest reached 87% accuracy and an ROC AUC of 0.8822, but recall for the positive class remained limited (0.23), suggesting difficulty in identifying responders despite high specificity. Figure 4 presents confusion matrices for each model, while Figure 5 displays ROC curves.

Feature importance from the XGBoost model (Figure 6) indicated that spending on gold products, wine, and meat were among the most predictive features. Additionally, recency and number of web visits per month played significant roles. A SHAP summary plot (Figure 7) further confirmed these findings, highlighting that higher gold and wine spending positively influenced campaign response probabilities.

Compared to a naive baseline accuracy (predicting all customers as non-responders, ~85%), all models improved recall and provided actionable predictions for targeting likely responders. XGBoost balanced precision and recall most effectively and outperformed both Logistic Regression and Random Forest on most metrics (Table 2).

ROC curves (Figure 5) illustrate model tradeoffs between sensitivity and specificity. The ROC AUC scores were highest for Random Forest (0.8822), closely followed by XGBoost (0.8771), and Logistic Regression (0.8618), suggesting that all models captured meaningful signals despite the class imbalance.

XGBoost emerged as the best-performing model due to its balance of predictive power and interpretability, particularly when paired with SHAP analysis. While Random Forest performed comparably in terms of accuracy and AUC, it failed to identify many true responders. Logistic Regression, though helpful in identifying responders, suffered from low precision. Feature importance analysis emphasized that campaign success was most strongly tied to prior spending on premium product categories and recency of engagement, informing potential targeting strategies for future campaigns.

Discussion

This study set out to predict customer responsiveness to a discounted Gold Membership campaign using behavioral and demographic data. The final model, an XGBoost classifier, achieved strong performance with 88% accuracy, a precision of 0.60, recall of 0.48, and a ROC AUC of 0.8771. Among the three models tested—XGBoost, Logistic Regression, and Random Forest—XGBoost offered the most balanced trade-off between precision and recall, making it the most suitable for identifying high-potential responders while maintaining reasonable campaign costs. SHAP analysis further supported the model’s credibility by revealing interpretable feature contributions, with key drivers including Recency (days since last purchase), spending on wine, meat, and gold products, as well as customer tenure and household composition.

These findings have significant implications. The model’s ability to flag lapsed yet high-value customers—those who had spent heavily in the past but had not made recent purchases—suggests that the Gold Membership offer can serve as an effective re-engagement tool. This insight enables Superstore to craft more strategic campaigns, particularly those targeting loyal but inactive customers. Moreover, the importance of product-specific spending patterns, such as wine and gold products, highlights an opportunity for targeted messaging aligned with customers’ past preferences. In practical terms, marketing teams can now use these predictions to focus their outreach on segments most likely to convert, enhancing return on investment while minimizing wasted effort.

Nonetheless, the study has several limitations. First, the class imbalance—only 15% of customers responded positively to the offer—posed a significant modeling challenge. Although techniques like stratified sampling and class weighting helped mitigate this issue, it still affects the model's generalizability. Additionally, the dataset lacked external behavioral indicators such as web engagement, campaign exposure history, or seasonality, which could have strengthened predictive accuracy. Computational constraints limited the scope of hyperparameter tuning, and while SHAP values added interpretability, the underlying model remains complex and may be less accessible to business stakeholders without data science expertise.

Despite these constraints, the model provides tangible business value. It can be integrated into existing CRM systems to automate customer scoring and generate targeted campaign lists. Marketing teams could also experiment with different probability thresholds, adjusting them based on campaign goals—whether aiming for broader reach or higher conversion efficiency. Furthermore, by identifying not only likely responders but also high-spend customers who may be at risk of churn, the model supports both acquisition and retention strategies. It thus contributes to a more nuanced and data-informed approach to customer lifecycle management.

When compared to existing approaches, this model demonstrates clear advantages. Logistic Regression, while interpretable, achieved only 35% precision, which would lead to inefficient over-targeting. Random Forest offered strong precision (68%) but failed to identify most actual responders, with a recall of just 23%. In contrast, XGBoost struck a meaningful balance and leveraged nonlinear interactions more effectively. This aligns with broader trends in predictive marketing, where ensemble methods and explainability tools like SHAP are increasingly favored for their performance and transparency.

Future enhancements could further improve the model's utility. More comprehensive hyperparameter optimization could refine performance, and incorporating additional data sources—such as engagement with past campaigns, website activity, or economic indicators—might capture subtler behavioral patterns. The development of a customer lifetime value (CLV) model could be integrated alongside the campaign response model, enabling prioritization not just by likelihood to respond, but by long-term revenue potential. Finally, deploying the model in a real-time scoring environment and evaluating it through A/B testing would allow for direct measurement of its impact on conversion rates and campaign ROI.

Conclusion

This study successfully developed a predictive model to identify customers most likely to respond to a Gold Membership campaign, leveraging demographic and behavioral data from a retail superstore. Among the models tested, XGBoost offered the most effective balance between precision and recall, supported by SHAP analysis for interpretability. The results not only validate the utility of machine learning in marketing strategy but also offer practical recommendations for campaign targeting. While limitations such as class imbalance and missing external data exist, the model provides a strong foundation for future enhancements. Integrating this model into real-world CRM systems has the potential to improve both customer acquisition and retention, making data-driven personalization a scalable reality for modern marketers.

References

Armstrong, J. S. (2001). *Principles of forecasting: A handbook for researchers and practitioners*. Springer.

Chatterjee, S., Nguyen, B., Ghosh, S. K., Bhattacharjee, K. K., & Chaudhuri, R. (2020). Adoption of artificial intelligence-integrated CRM systems in the banking sector: An empirical study. *International Journal of Information Management, 52*, 101934.

Davenport, T. H., & Harris, J. G. (2007). *Competing on analytics: The new science of winning*. Harvard Business Press.

Delen, D. (2020). *Predictive analytics: Data mining, machine learning and data science for practitioners* (2nd ed.). Pearson FT Press.

Dinov, I. D. (2023). *Data science and predictive analytics: Biomedical and health applications using R* (2nd ed.). Springer.

Fader, P., & Toms, S. (2018). Customer lifetime value: Reshaping business thinking for the long term. *Journal of Marketing Analytics, 6*(3), 123–133.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8th ed.). Cengage Learning.

Kumar, V., & Reinartz, W. (2016). Creating enduring customer value. *Journal of Marketing, 80*(6), 36–68.

Lemke, C., Gabrys, B., & Buhmann, J. M. (2009). Automatic selection of decision tree algorithms using evolutionary computation. *Journal of Heuristics, 15*(2), 167–198.

Leventhal, B. (2018). *Predictive analytics for marketers: Using data mining for business advantage*. Kogan Page Publishers.

Provost, F., & Fawcett, T. (2013). *Data science for business: What you need to know about data mining and data-analytic thinking*. O’Reilly Media.

Siegel, E. (2013). *Predictive analytics: The power to predict who will click, buy, lie, or die*. Wiley.

Sridhar, S., & Mittal, V. (2021). Customer-centricity in the age of data: A strategic framework for predictive marketing. *Journal of Business Research, 124*, 347–359.

Tsiptsis, K., & Chorianopoulos, A. (2009). *Data mining techniques in CRM: Inside customer segmentation*. Wiley.

Winston, W. L. (2014). *Marketing analytics: Data-driven techniques with Microsoft Excel*. Wiley.

Appendix

https://colab.research.google.com/drive/1Qty1ELnsWghwsP9sRAqdDRrCPn_c8B03

Figure 1: Income Distribution by Campaign Response

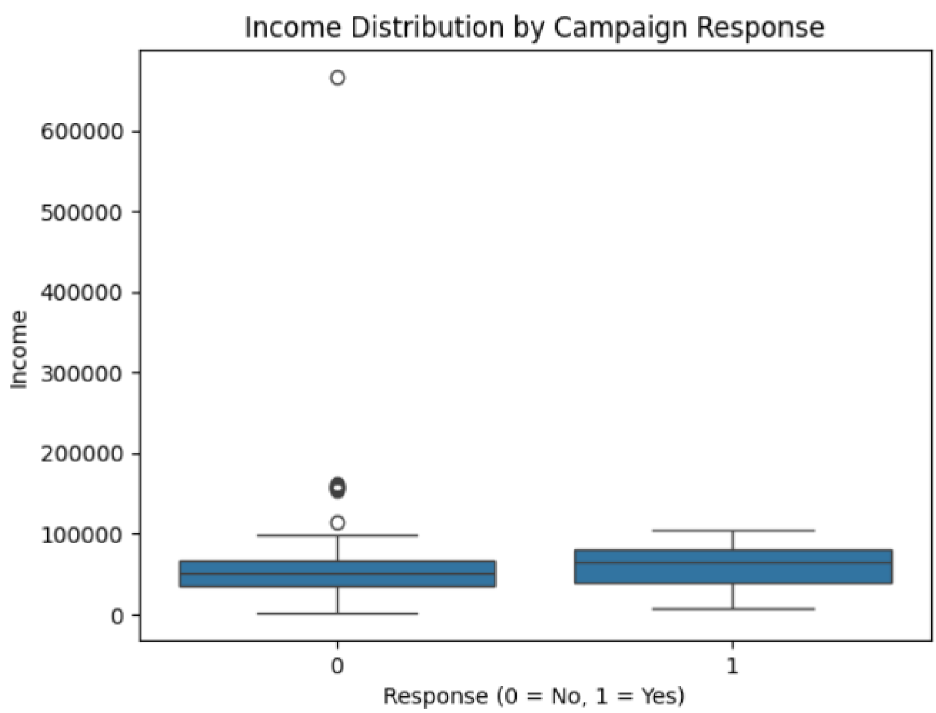


Figure 2: Average Wine Spending by Campaign Response

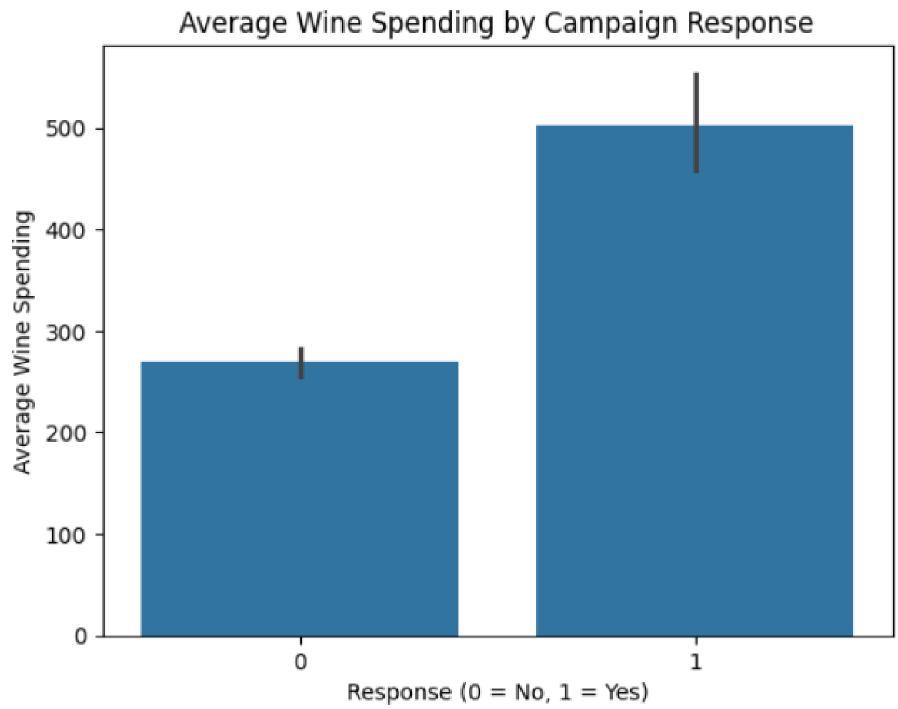


Figure 3: Response Rate by Enrollment Year

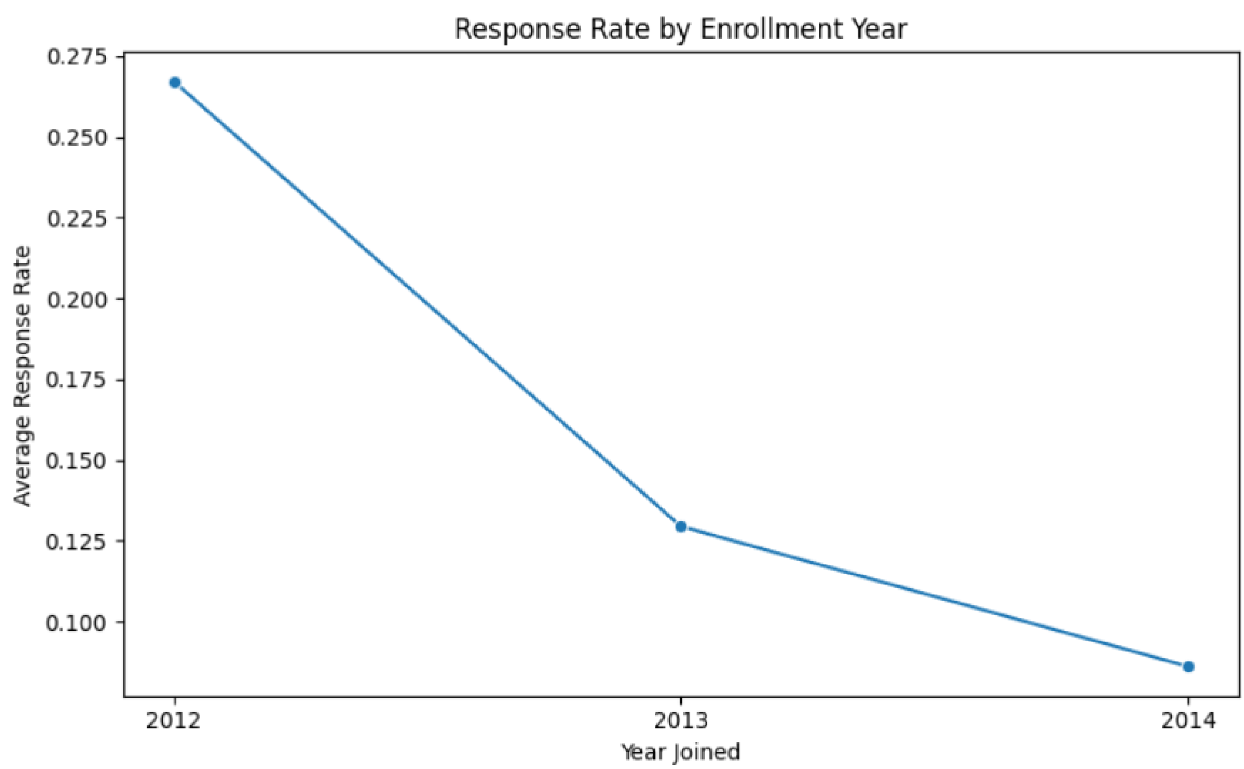
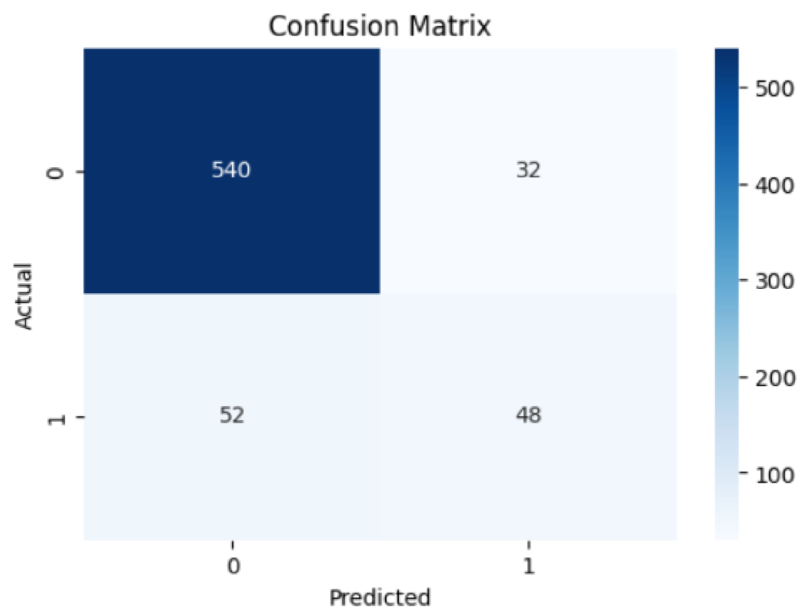
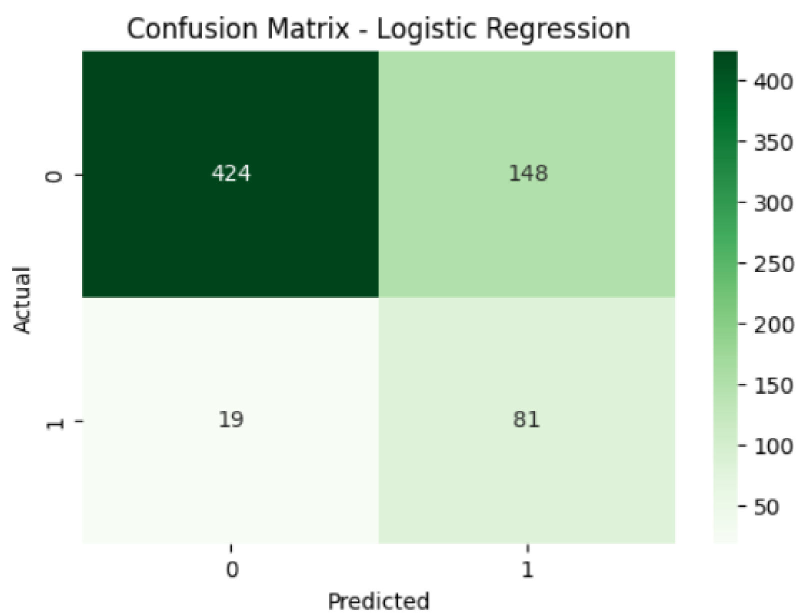


Figure 4: Confusion Matrices for All Models

- Figure 4a: XGBoost Confusion Matrix



- Figure 4b: Logistic Regression Confusion Matrix



- Figure 4c: Random Forest Confusion Matrix

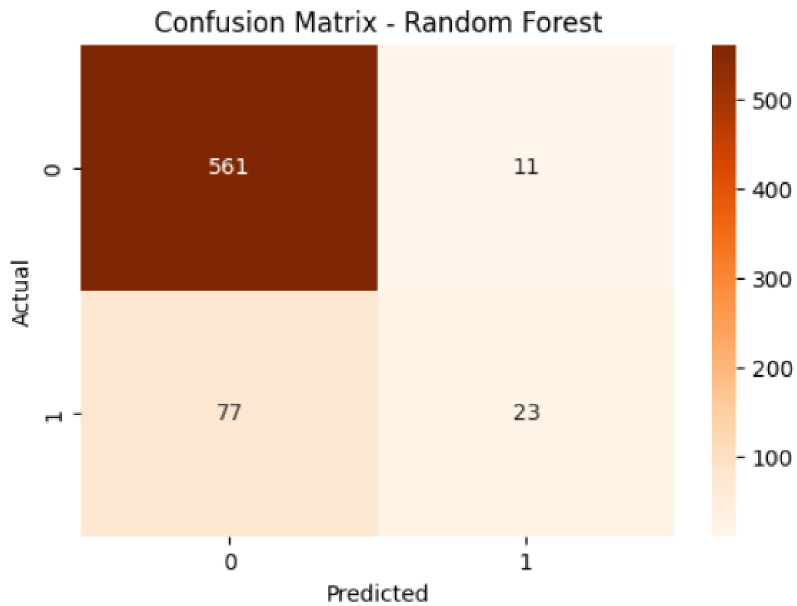


Figure 5: ROC Curves for All Models

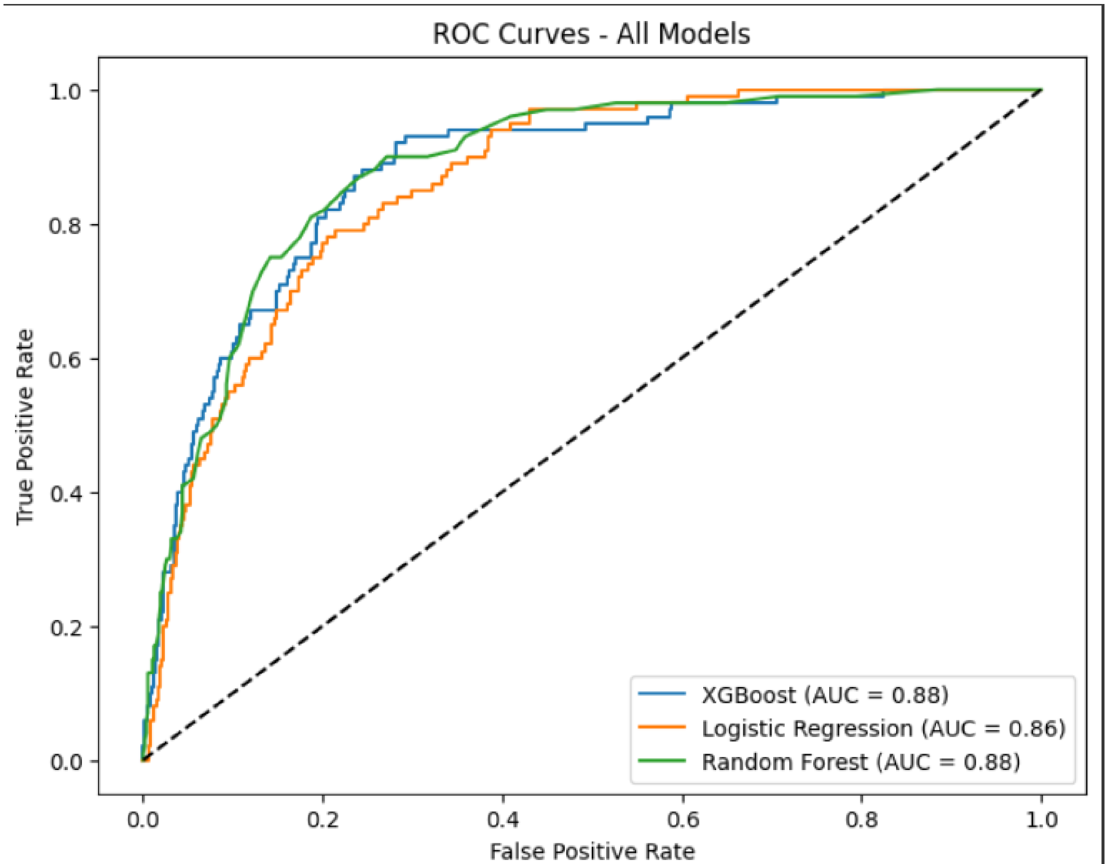


Figure 6: XGBoost Feature Importances

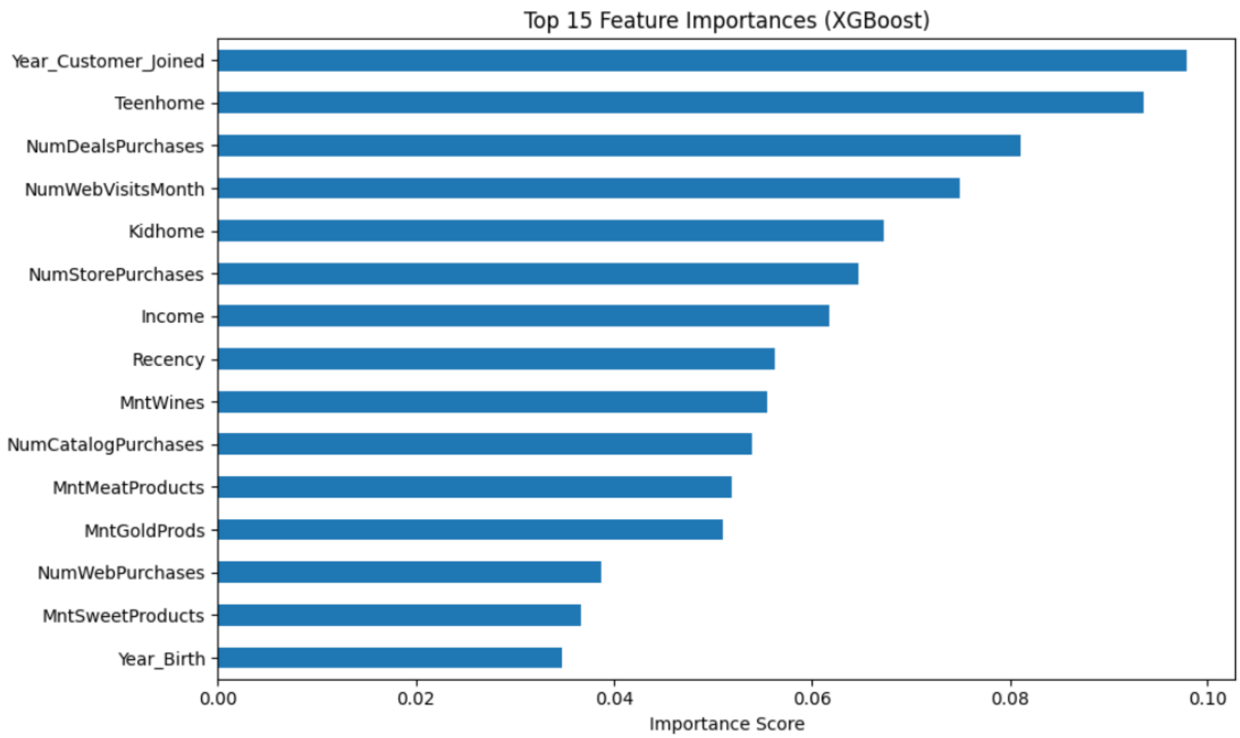


Figure 7: SHAP Summary Plot of Feature Contributions

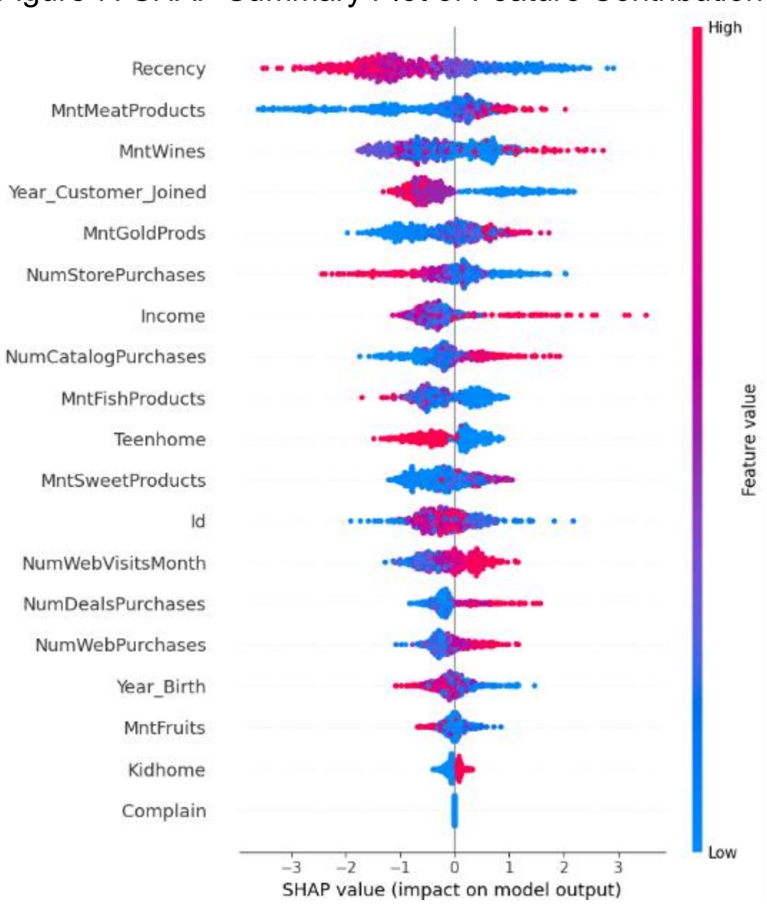


Table 1 – Model Metrics Summary

Model	**Accuracy**	**Precision (1)**	**Recall (1)**	**F1 Score (1)**	**ROC AUC**
XGBoost	0.88	0.60	0.48	0.53	0.8771
Logistic Regression	0.75	0.35	0.81	0.49	0.8618
Random Forest	0.87	0.68	0.23	0.34	0.8822

Table 2 - Model Performance Compared to Baseline

Model	Accuracy	Precision (1)	Recall (1)	F1 Score (1)	ROC AUC	Improvement Over Baseline
Baseline	0.85	0.00	0.00	0.00	0.50	—
XGBoost	0.88	0.60	0.48	0.53	0.8771	Higher precision, recall, and AUC
Logistic Regression	0.75	0.35	0.81	0.49	0.8618	Higher recall, lower precision and accuracy
Random Forest	0.87	0.68	0.23	0.34	0.8822	Higher precision and AUC, lower recall