BADM590 Consumer Analytics Final Project

Case: Bookbinders Book Club

Group 10: Devina Chhajer, Divya Gupta, Jamie Liu, Panshul Saraswat, Vishwas Rao, Zoe Chen

1. Executive Summary

The Bookbinders Book Club (BBBC), established in 1986 and specializing in direct marketing of specialty books, aims to enhance customer engagement through predictive modeling despite intense competition and changing consumer habits. BBBC's case study evaluates the effectiveness of RFM (Recency, Frequency, and Monetary Value), Ordinary Linear Regression, and Binary Logit Model in improving response rates to direct mail campaigns. An experimental campaign mailing "The Art History of Florence" to 20,000 customers achieved a 9.03% response rate, using data from 1,600 customers to develop these models. These were validated against a holdout sample of 2,300 customers, determining the Binary Logit model as the most effective with an 82.34% accuracy.

BBBC is advised to invest $50,000 in developing logistic regression expertise for better direct mail campaign evaluation and an additional $30,000 to enhance the Binary Logit model by incorporating factors like customer sentiment and demographic data for more targeted marketing.

2. Complete Recommendations

a) Linear Regression Model

Based on the results from the linear regression model, we developed some recommendations for Bookbinders Book Club (BBBC). Firstly, BBBC should focus its marketing strategies on segments that have shown interest in art books, as these customers are more likely to respond to similar offerings. Tailored promotions and targeted mailings for art-themed books could significantly enhance conversion rates among this demographic. Secondly, considering that females are more likely to purchase than males, marketing efforts should be customized to appeal more to female customers, potentially highlighting themes that resonate with them. Additionally, for frequent buyers, who show decreased likelihood of purchase, implementing a loyalty program or offering special incentives like exclusive previews and member-only discounts could help maintain engagement and mitigate purchase fatigue. By integrating these strategies, BBBC can optimize its resource allocation, boost customer engagement, and increase the efficacy of its direct mail campaigns.

Given the limitations observed with the linear regression model, particularly its inadequacy in handling binary outcomes effectively, it would be advisable for Bookbinders Book Club not to invest heavily in developing in-house expertise specifically for linear regression as it relates to direct mail campaign evaluations. The binary nature of the outcome variable 'Choice' suggests that logistic regression, which is specifically designed for binary outcomes, might provide more accurate and relevant insights for such marketing decisions.

b) Binary Logit Model

Targeting Based on Purchase Likelihood: The Binary Logit model allows BBBC to forecast customer purchase probabilities effectively. By targeting individuals most likely to buy, BBBC can optimize resource allocation and enhance the success rates of its marketing campaigns.

Understanding Customer Behavior:

Demographic Influence: BBBC should tailor its marketing and services to appeal more to female consumers, who are more likely to purchase, according to the model's insights on gender-based buying habits. Additionally, the lower purchase likelihood among men suggests a need for BBBC to review and possibly develop offerings that better cater to this demographic.

Purchase Drivers: Variables such as purchase amount, frequency, and recency highlight customer engagement; those with higher or recent spending are ideal for new offers, whereas strategies should be devised to boost purchase frequency among infrequent buyers.

Segmentation and Personalization: With the help of the model's insights, BBBC can divide up its customer base into behavioral categories like product preferences in addition to more traditional demographic groups like P_Child, P_Youth, P_Cook, P_DIY, and P_Art. By speaking directly to the interests of each segment, this segmentation enables highly personalized marketing that can boost customer satisfaction and loyalty.

For instance, to capitalize on their expressed interest, art book enthusiasts could be sent invitations to art events, special edition releases, or exclusive content about art.

c) RFM Model

Bookbinders Book Club (BBBC) can leverage RFM analysis results to optimize its marketing strategies by focusing on high-value customer segments. These segments, identified by their recent, frequent, and substantial purchases, are likely to generate repeat business and significant revenue. By classifying customers into "High Value," "Medium Value," and "Low Value" segments, BBBC can tailor its marketing approaches to meet the distinct needs and preferences of each group. For example, high-value customers might benefit from exclusive offers and loyalty programs designed to enhance long-term relationships and increase customer lifetime value, while strategies for low-value customers could focus on re-engagement and incentivizing additional purchases. Strategically allocating resources to these targeted segments enables BBBC to maximize marketing efficiency, boost customer engagement, and drive revenue growth more effectively.

3. Analysis Details

a) Linear Regression Model

In the analysis using an ordinary linear regression model, we utilized the training dataset from the Bookbinders Book Club containing variables such as gender, amount purchased, purchase frequency, months since last/ first purchases, and types of books purchased (children's, youth, cookbooks, DIY, art). The dependent variable, Choice, indicates whether the customer purchased "The Art History of Florence". The model aimed to identify the key factors influencing this purchasing decision. We developed the model by regressing Choice on all provided independent variables and evaluated its performance using standard statistical outputs such as coefficients, R-squared values, and significance levels.

The results from the linear regression analysis reveal several insights. Significant predictors include the type of books previously purchased, with art books (coefficient = 0.07931, $p < 0.00001$) increasing the likelihood of purchasing "The Art History of Florence", while purchases of children's, youth, DIY, and cookbooks decrease it. The negative coefficients for gender (coefficient = -0.07016, $p < 0.00001$) suggest that males are less likely to purchase the book compared to females. Other influential factors include the recency of last purchase (coefficient = 0.03763, $p < 0.0001$) and purchase frequency (coefficient = -0.00650, $p < 0.00001$), where recent purchases correlate with a higher likelihood of buying the book, but increased frequency correlates with a lower likelihood.

The linear regression model shows a Multiple R-squared value of 0.1208 and an Adjusted R-squared of 0.1169, indicating it explains about 12% of the variance in the decision to purchase. This suggests the model captures some, but not all, factors influencing purchase decisions. When tested on holdout data, the model achieved an accuracy of 75.56%, which points to room for improvement in predicting customer behavior more accurately. This performance supports using the model for some predictive insights while also highlighting the need for additional data or more complex modeling techniques to enhance predictive accuracy.

The advantage of the linear regression model in this context is its simplicity and interpretability, which allows for an easy understanding of how each factor influences the likelihood of a purchase. However, a significant limitation of using linear regression here is the binary nature of the outcome variable, 'Choice', which indicates whether or not a customer made a purchase. Linear regression assumes a continuous outcome, not a categorical one like in this case, which can lead to issues such as predictions falling outside the meaningful range of 0 and 1. This misalignment can affect the accuracy and appropriateness of the model for binary decision-making processes, suggesting the potential need for a logistic regression model that inherently accommodates binary outcomes.

b) Binary Logit Model

Gender Influence: According to the model, gender is a significant predictor, with men being less likely than women to make a purchase. A highly significant negative coefficient for gender (-0.8632, $p < 0.0001$) supports this.

Spending and Purchase Behavior: The amount purchased and the probability of making a purchase had a positive relationship, though with a tiny effect size (coefficient = 0.0018641, p =

0.0186). On the other hand, the probability of buying this particular book decreases with purchase frequency, suggesting that the market possibly reached saturation (coefficient = -0.0755142, $p < 0.00001$).

Recency and Customer Engagement: The likelihood of a purchase is significantly increased by recent interactions, as indicated by the most recent purchase (coefficient = 0.6117713, $p < 0.00001$). On the other hand, a longer interval between purchases reduces the likelihood of making a purchase, though this effect is not statistically significant (coefficient = -0.0147792, $p = 0.2483$).

Product Categories: Interest in specific book categories has varied impacts. Notably, purchases of art books are positively associated with buying "The Art History of Florence" (coefficient = 0.6861124, $p < 0.00001$), suggesting a targeted interest. Other book categories, such as children's and DIY, exhibit a negative correlation, highlighting unique consumer preferences.

The Binary Logit model, useful for predicting purchasing decisions based on various factors, assumes all observations are independent and follows a linear pattern, necessitating regular updates and checks with new data to maintain its accuracy and relevance.

c) RFM Model

In conducting the RFM analysis, we utilized the RFM framework to segment customers based on their recency, frequency, and monetary values. The dataset provided by Bookbinders Book Club (BBBC) included variables such as last purchase date, purchase frequency, amount spent, and customer ID. By calculating RFM scores for each customer, we were able to categorize them into distinct segments representing their purchasing behavior and value to the business.

The RFM analysis revealed several key findings about BBBC's customer base. Firstly, we observed a spectrum of RFM scores among customers, ranging from high-value segments characterized by recent, frequent, and high-value purchases, to low-value segments with less frequent and lower-value transactions. This segmentation allowed us to identify segments with the highest potential for revenue generation and prioritize marketing efforts accordingly.

Analyzing the distribution of customers across RFM segments, we found that the majority of customers fell into the medium-value segment, comprising 1196 customers in the training dataset and 1748 customers in the testing dataset. This indicates that while there is a significant portion of customers with moderate purchasing behavior, there are also substantial opportunities to target high-value segments for more lucrative returns. Additionally, we identified 266 high-value customers in the training dataset and 382 in the testing dataset, suggesting a valuable cohort of customers with recent, frequent, and high-value purchases. Targeting these high-value segments with personalized offers, exclusive discounts, or premium services could significantly enhance customer satisfaction and loyalty, ultimately driving long-term profitability for BBBC.

Moreover, we observed 138 low-value customers in the training dataset and 170 in the testing dataset. While these customers may contribute less revenue individually, implementing targeted. Retention strategies or reactivation campaigns could help prevent attrition and maximize their lifetime value to the business.

Furthermore, we calculated the accuracy score of the RFM model on the testing dataset to evaluate its predictive performance. The accuracy score helps assess the model's effectiveness in correctly classifying customers into their respective RFM segments and predicting their purchasing behavior. The RFM model achieved an accuracy score of 76.09% on the testing dataset, indicating its ability to accurately classify customers into their RFM segments and predict their purchasing behavior.

While the RFM model provides valuable insights into customer segmentation and purchasing behavior, it also comes with certain advantages and limitations. One advantage is its simplicity and ease of interpretation, making it accessible for businesses to implement and understand. Additionally, RFM segmentation allows for targeted marketing strategies, leading to improved customer engagement and higher conversion rates.

However, a limitation of the RFM model is its reliance on historical transaction data, which may not capture changes in customer behavior or preferences over time. Moreover, RFM analysis does not consider external factors such as market trends, competitive landscape, or socio-economic factors, which could influence customer behavior.

4. Profit Exercise

Mailing Costs per Address: $0.65 per addressee

Book Purchase and Mailing Cost = $15 per book

Overhead Cost = 45% of Book Cost = 45% x $15 = $6.75 per book

Total Cost = Mailing Cost + Book Purchase and Mailing Cost + Overhead Cost= $0.65 + $15 + $6.75 = $22.40 per targeted customer

Selling Price = $31.95 per book

Profit = Selling Price - Total Cost = $31.95 - $22.40 = $9.55 per book

Model Insights

a) Linear Regression Model: This model identifies significant factors like gender and the type of books purchased influencing the purchase decision. It suggests that females and customers interested in art books are more likely to purchase, while frequent buyers and those buying children's, youth, DIY, and cookbooks are less likely.

b) Binary Logit Model: This model also highlights the influence of gender, with females more likely to make a purchase. It further emphasizes that recent purchase activity increases the likelihood of buying again. It points out that focusing on customers who have shown an interest in art books could be beneficial, as this positively correlates with the likelihood of purchasing "The Art History of Florence."

c) RFM Model: This segmentation model shows that targeting customers based on recent, frequent, and high monetary transactions (high RFM scores) could yield better responses. This is particularly effective for identifying which customer segments are likely to generate the most revenue.

Targeting Strategy Based on Models to maximize the campaign's effectiveness and profitability, Bookbinders should:

- Target Females who have shown interest in art books as identified by the Linear and Logit models
- Focus on Recent Buyers with significant spending on BBBC books, especially those who have recently purchased art books
- Prioritize High-Value RFM Segments, which include customers with the highest recency, frequency, and monetary values

Assuming a response rate enhancement from the typical 5% to 20% by utilizing model insights :

- Without Targeting (5% response): 2,500 sales x $9.55 = $23,875 profit
- With Targeting (20% response): 10,000 sales x $9.55 = $95,500 profit
- Additional Profit with Targeted Approach: $95,500 - $23,875 = $71,625

5. Next Steps

To enhance its analysis, Bookbinders Book Club (BBBC) could employ dynamic modeling techniques like time-series analysis to track customer behavior changes over time and adapt marketing strategies accordingly. This approach would also benefit from incorporating external factors such as economic trends and competitor activities to better predict market dynamics and improve strategic decision-making. Additionally, exploring the lifetime value of customer segments identified through RFM analysis could guide more effective resource allocation and retention strategies by revealing the profitability of different groups.

Investigating cross-selling opportunities by analyzing product associations could help BBBC identify complementary book categories, potentially increasing revenue per customer. Attribution modeling could also refine marketing channel effectiveness, optimizing budget distribution to enhance return on investment.

Further refinement in customer segmentation could be achieved by applying clustering techniques to behavioral data from RFM analysis, enabling more targeted marketing approaches. Combining logistic regression, RFM, and machine learning in a predictive modeling ensemble could also heighten predictive accuracy and robustness. Additionally, models predicting customer lifetime value from historical transactions and RFM scores would allow BBBC to focus on high-value customers, optimizing long-term revenue potential.