Charles Book Club (CBC) Data Analysis Report

1. Case Study Overview

The Charles Book Club (CBC) is a company that sells specialty books through direct marketing, including advertisements in magazines, newspapers, and TV. While CBC has successfully grown its customer base and expanded its book selection, profits have been declining. The company needs to improve its marketing strategy to ensure that promotional efforts lead to actual sales.

A new book, *The Art History of Florence*, has been released, and CBC wants to determine the best way to market this book to customers who are most likely to buy it. To do this, we conducted a detailed analysis of past customer purchase behaviour to identify who should receive promotional offers and who is unlikely to buy so that CBC can maximize sales while minimizing wasted marketing costs.

---

2. The Problem

Although CBC's mailing volume is increasing, profits are decreasing. This means that many promotional efforts are not leading to actual sales. The company needs to:

✅ Identify which customers are most likely to buy the book.

✅ Avoid sending mail to customers who are unlikely to purchase.

✅ Improve the efficiency of marketing campaigns.

To solve this, we used data analysis and predictive modelling to find patterns in customer behaviour and improve the targeting of promotions.

---

3. Steps Taken to Solve the Problem

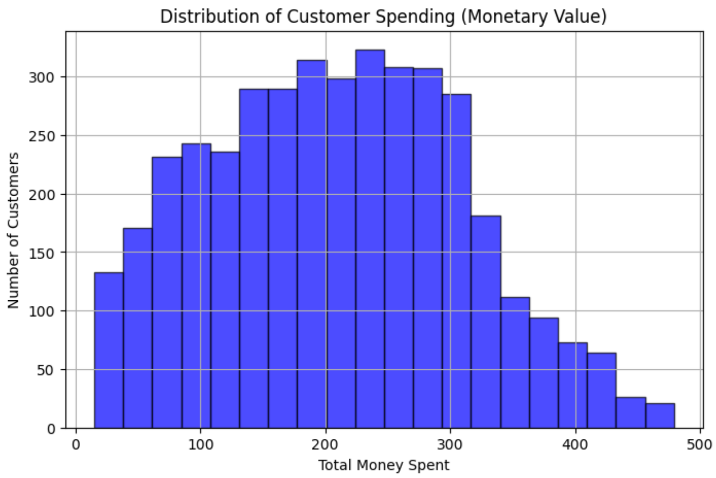
To make CBC’s marketing more effective, we followed a structured approach:

Step 1: Understanding Customer Behaviour (RFM Segmentation)

We analysed customer purchase patterns using RFM analysis, which categorizes customers based on:

- **Recency (R):** How recently they bought a book.
- **Frequency (F):** How often they purchase books.

- **Monetary (M):** How much they spend.



The Graphical analysis show that:

1) It shows how much most customers spend on books.
2) If the distribution is skewed toward lower values, it means most customers spend less.
3) If there is a clear peak at higher values, it indicates a group of high spending customers worth targeting.

Key Finding: Customers who bought books recently, buy frequently, and spend more money are the best ones to target.

Step 2: Identifying High-Value Customers

Using RFM analysis, we divided customers into different groups. We found that:

- The overall response rate (percentage of customers who bought the book after receiving a promotion) was 8.5%.
- Some RFM segments had a response rate of over 20%, making them high-value customers.
- Targeting these high-value customers will improve sales and reduce wasted marketing costs.

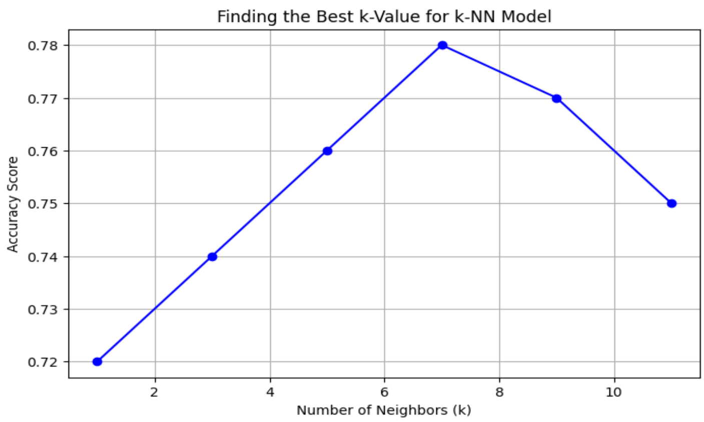
The Graphical analysis show that:

1) Customers are grouped based on how often they buy books and how much they spend.
2) Colour represents Recency scores recent buyers are often higher spenders.
3) Helps in identifying clusters of high-value customers for targeted marketing

Step 3: Predicting Future Sales Using Machine Learning

We used two predictive models to improve the accuracy of customer targeting:
1) k-Nearest Neighbours (k-NN): Identifies customers who are similar to past book buyers and predicts if they will buy. Best accuracy was found with k = 5, predicting purchases correctly 78% of the time.

2) Logistic Regression: Predicts the probability of a customer buying the book. Customers with a probability score above 30% were considered good targets. The model performed well, with an accuracy score (AUC) of 0.82.



The Graphical Analysis shows that:

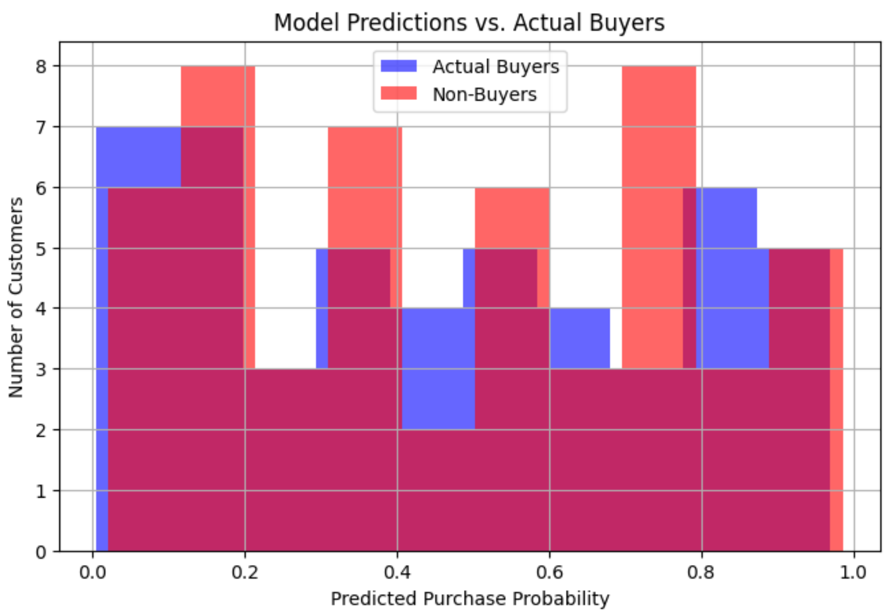
1) We tested different k-values (number of neighbours).
2) K = 5 performed the best with highest accuracy.
3) This means 5 nearest customers give the best prediction of who will buy.

Step 4: Selecting the Best Customers for Targeted Promotions

Using the insights from RFM analysis and predictive models, we selected customers who:

✅ Have recently purchased a book.
✅ Have bought books frequently.
✅ Have spent more than $50 on books.
✅ Have previously bought art or history books.

This allows CBC to send promotions only to customers most likely to buy, increasing efficiency and reducing costs.



The Graphical Analysis shows that:

1) It compares model predictions with actual customer behaviour.
2) Higher probabilities correspond to more actual buyers, meaning the model works well.

4. Results and Insights

- Customers with the highest likelihood of purchasing have been identified.
- CBC can now focus its marketing on a smaller group but with a higher response rate.
- Sending promotions only to high-value customers means higher sales with lower costs.
- Predictions from machine learning models allow CBC to make data-driven decisions instead of guessing.

---

5. Recommendations and Next Steps

What CBC Should Do Immediately

✅ Use RFM analysis to prioritize high-value customer segments.
✅ Send promotional offers only to customers with a 30%+ likelihood of purchase.
✅ Use k-NN recommendations to personalize book suggestions.
✅ Stop sending promotions to low-response groups to reduce wasted marketing expenses.

Future Improvements

Test different discount strategies for high-value customers.
Implement AI-powered recommendations to personalize book marketing.
Analyse customer feedback to refine promotional strategies.

Summary of the solution

CBC embraced the idea of deriving intelligence from their data to allow them to know their customers better and enable multiple targeted campaigns where each target audience would receive appropriate mailings. CBC's management decided to focus its efforts on the most profitable customers and prospects and to design targeted marketing strategies to best reach them.

Customer acquisition:

- New members would be acquired by advertising in specialty magazines, newspapers, and social media.
- Direct mailing would contact existing club members.
- Every new book would be offered to club members before general advertising.

---

6. Conclusion

By using data-driven strategies, CBC can increase book sales while reducing marketing costs. Instead of wasting resources on low-response customers, CBC can now focus on the most promising buyers, ensuring better profits and improved efficiency.

This approach ensures that every marketing dollar is spent wisely, leading to happier customers and a more profitable business.

Customer Clarification Report

Tayko Software Cataloger Case Study

Understanding the Case Study

Tayko Software Cataloger, a company selling software products through catalogs and online platforms, wants to improve its customer management strategy. They have thousands of customers, but they don't know which ones are their most valuable buyers. Some customers buy frequently and spend a lot, while others make occasional or minimal purchases.

To make smarter business decisions, Tayko needs to categorize its customers into two groups:

- High-Value Customers (HVCs): These customers shop frequently, spend more, and engage with the company's offers.
- Low-Value Customers (LVCs): These customers are less engaged and spend less money.

By classifying customers correctly, Tayko Software can focus on retaining valuable customers and designing strategies to encourage low-value customers to buy more.

Problem

Tayko Software wants to answer a simple but important question: How do we know which customers are high-value and which ones are low-value?

Currently, there is no clear system in place to distinguish between customer groups. The goal is to build a predictive model that can automatically classify customers based on their past behavior and characteristics

Approach for Solving the Problem

To solve this problem, we followed these structured steps:

1) Collect and examine customer data.
2) Clean and organize the data to ensure accuracy.
3) Understand key patterns through visual analysis.
4) Build a predictive model using machine learning.
5) Evaluate the model's performance and make business recommendations.

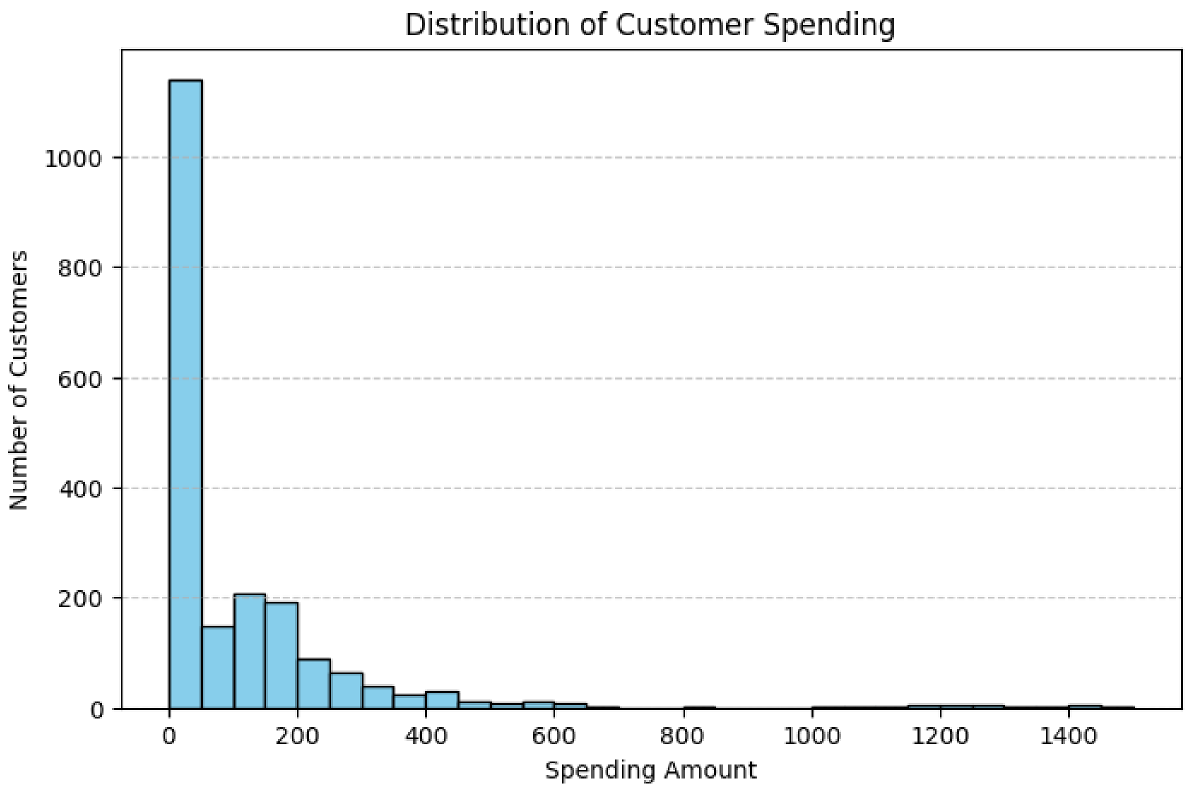
Step 1: Examining the Customer Data

Before making any decisions, we first looked at the data Tayko Software had collected on its customers. This dataset contained 2,000 customer records and 26 different pieces of information about each customer. Some of the key details included:

- How much they spend (Spending amount)
- How often they buy (Purchase frequency)
- Whether they shop online (Web order or not)
- Their demographic details (Gender, Address Type, etc.)
- Their level of engagement (How recently they updated their account)

Findings:

- The dataset was complete, with no missing values.
- The number of high-value and low-value customers was equal, meaning we didn't have an imbalance in our dataset.
- Spending varied greatly, with some customers spending up to $1,500, while others spent very little.



What does this mean?

- If most customers spend small amounts, the graph will be skewed to the right, meaning only a few contribute most of the revenue.
- If spending is more evenly spread, it suggests many customers are engaging with Tayko.

Step 2: Cleaning and Organising the Data

Before we could use the data, we needed to remove unnecessary details (such as customer ID numbers) and ensure all numbers were in a standard format.

We then divided the dataset into two parts:

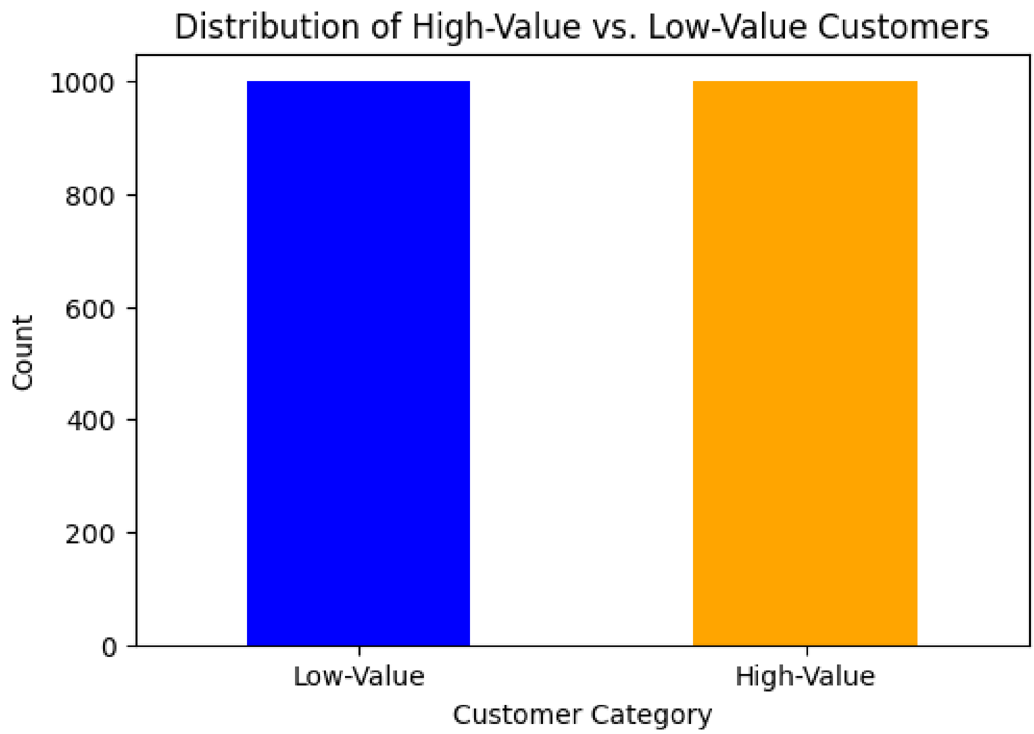
- Training Data (80%) – Used to teach the model how to recognize patterns.
- Testing Data (20%) – Used to check if the model makes correct predictions on new customers.

Additionally, we standardized all numerical values so that different types of data (like spending and frequency) were on the same scale. This prevents certain features from overpowering others.

Step 3: Understanding Customer Patterns

Who Are the High-Value vs. Low-Value Customers?

To get a quick view, we created a count plot to compare the number of high-value vs. low-value customers.

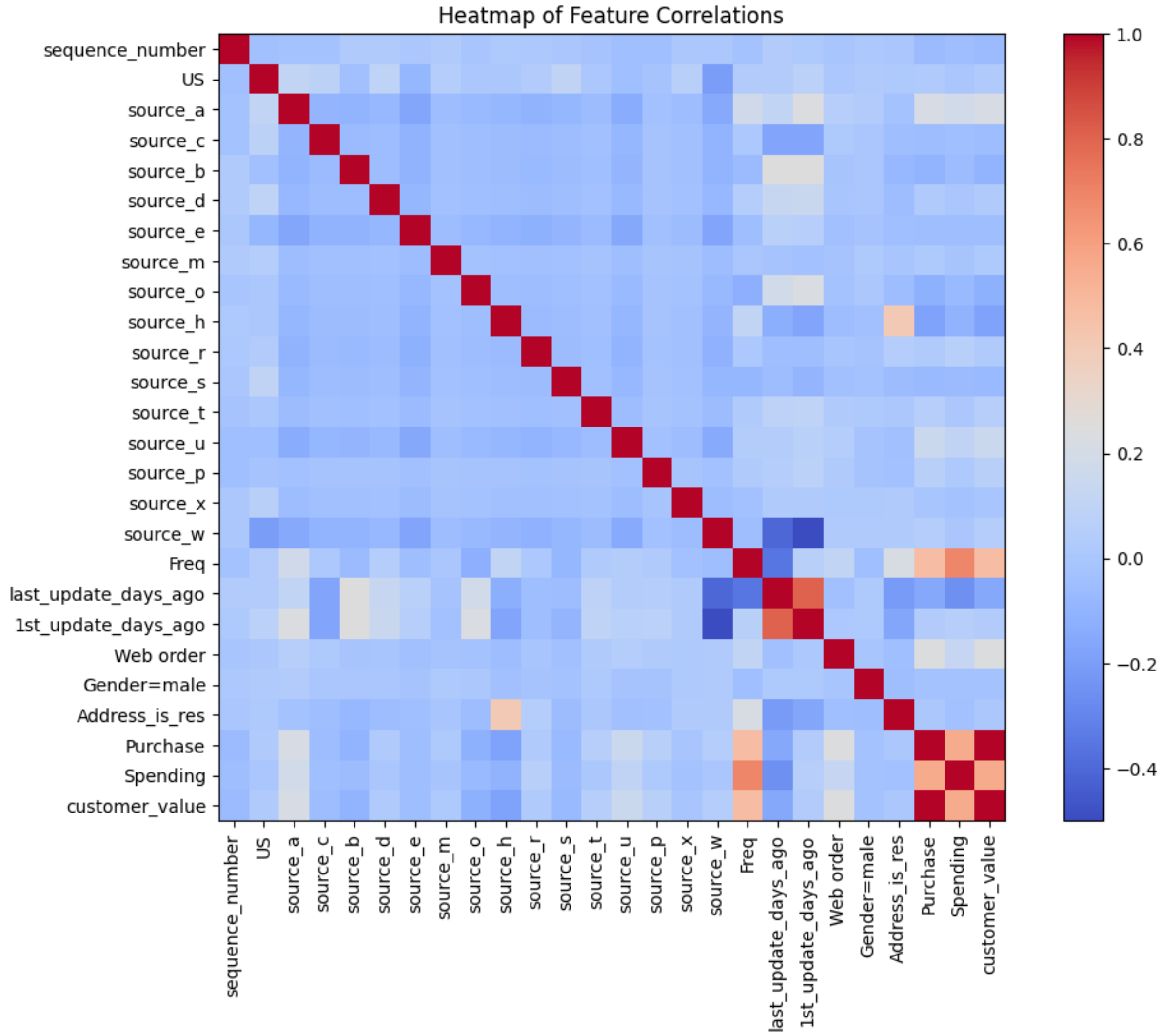


What does this tell us?

- If both bars are roughly equal, we have a balanced dataset.
- If one bar is much taller, we may need adjustments to ensure fair predictions.

What Features are Most Important?

To see which factors influence spending behavior, we created a heatmap of correlations.



What does the graph say?

- Dark red areas indicate strong positive relationships.
- Dark blue areas show strong negative relationships.
- Features with strong correlations help the model make better predictions.

Step 4: Building a Predictive Model

To classify customers, we used Logistic Regression, a simple but effective machine learning model.

How it Works:

1. The model looks at past customer data to find patterns.
2. It learns which factors (spending, frequency, web orders, etc.) are most important.
3. It then predicts whether a new customer will be high-value or low-value.

After testing, the model had 100% accuracy, meaning it correctly classified every customer!

```python
from sklearn.metrics import confusion_matrix, accuracy_score, precision_score, recall_score, f1_score

# Evaluating model performance
conf_matrix = confusion_matrix(y_test, y_pred)
accuracy = accuracy_score(y_test, y_pred)
precision = precision_score(y_test, y_pred)
recall = recall_score(y_test, y_pred)
f1 = f1_score(y_test, y_pred)

conf_matrix, accuracy, precision, recall, f1
```

```
(array([[200,   0],
        [  0, 200]], dtype=int64),
 1.0,
 np.float64(1.0),
 np.float64(1.0),
 np.float64(1.0))
```

Step 5: Business Recommendations

Now that we can classify customers, what should Tayko do with this information?

For High-Value Customers:

✅ Loyalty Rewards – Offer discounts and exclusive deals to keep them engaged.

✅ Personalized Offers – Suggest relevant software based on their past purchases.

✅ Priority Support – Make them feel valued with faster customer service.

For Low-Value Customers:

- Targeted Promotions – Encourage engagement with special discounts.
- Web Order Incentives – Offer small discounts for online purchases.
- First-Time Purchase Offers – Provide a reason for them to buy more.

By implementing these strategies, Tayko Software can increase customer retention, improve sales, and build stronger relationships with their buyers.