# Recommendation Engine Proposal for eCommerce Bookstore

*Business Presentation for Chief Marketing Officer & Webmaster*

SCIENTIA CONSULTING

# Our Team

**Anusha Bhat**

***Data Engineer***

**Sara Chaker**

***Team Lead***

**Louis Gao**

***Data Analyst***

**Hritik Jhaveri**

***Business Analyst***

**Q Li**

***Data Scientist***

**John Melel**

***Associate Partner, SME***

**Olivia Rumere**

***Data Scientist***

SCIENTIA CONSULTING

# Why Scientia Consulting?

## Domain Expertise

Our consulting group brings **specialized knowledge in the ecommerce industry** which enables us to understand the unique challenges and nuances that clients face. We develop data-driven solutions grounded in industry insights, ensuring relevance and impact for our clients.

## Proven Track Record

We pride ourselves on **over two decades of successful engagements**, demonstrated through measurable outcomes and **over 100 satisfied clients**. Our team's past projects highlight our ability to drive results, from increasing operational efficiency to boosting profitability.



SCIENTIA CONSULTING

# Agenda

## 01 Overview of Business Problem and Achieved Solution

*Summary of the eCommerce Bookstore's primary challenge and our achieved solution*

## 02 Model Choice & Result

*Overview of the selected model and model performance*

## 03 Customer Segmentation

*Demonstration of customer segmentation from model*

## 04 Business Impact & Implementation Process

*Credible estimate of the model's financial impact and a suggested implementation plan*

# Personalizing Book Recommendations for Increased Sales

- Your Ecommerce Bookstore promotes the same books to all visitors, **limiting potential sales** by not accounting for individual customer preferences.

- The challenge is to develop more **advanced recommendation system** that **segments users** by behaviors and purchase patterns to allow for **personalized product suggestions**.



SCIENTIA CONSULTING

# Solution Driven Analytics for Personalized Sales

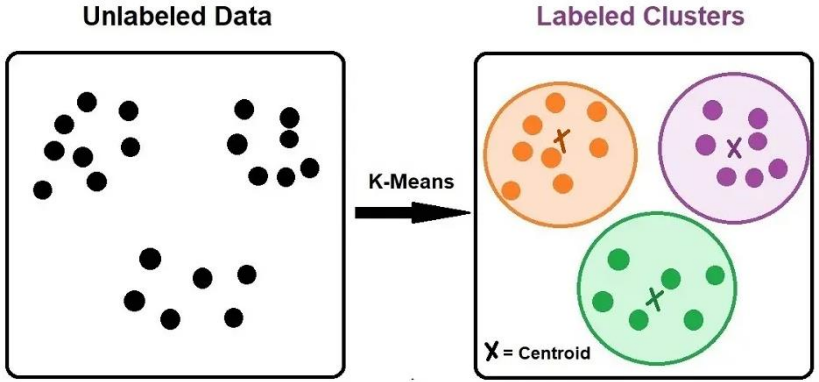
- We have used **unsupervised machine learning techniques** to create this scalable recommendation engine.
- We have created an intelligent customer segmentation system using **K-means clustering** that has generated **5 distinct customer profiles**
- The overall **sales** is forecasted to improve by **11%** and **revenue** is expected to grow by **$930,000**.



SCIENTIA CONSULTING

# Utilizing K-Means to Effectively Segment Customers for Personalized Recommendations

K-Means Clustering is an Unsupervised Machine Learning algorithm, which groups an unlabeled dataset into different clusters.




By grouping customers into clusters, we can identify patterns and similarities that allow us to make personalized and targeted recommendations.



SCIENTIA CONSULTING

# Actionable Customer Segments for Growth and Retention

The clustering is done based on key customer behavior and value –

- Purchase Recency
- Purchase Frequency
- Purchase Amount
- Membership Duration

| Segment Name | % of Customers | Key Characteristics | Advertisement Strategies |
|---|---|---|---|
| New & Frugal | 44% | Newly-acquired, low spend | Cost-effective offers |
| Potential Loyal | 30% | Moderate spend, high potential | Personalized loyalty programs, conversion focus |
| Long Lost | 21% | Infrequent buyers, non-recent | Re-engagement campaigns |
| Loyal | 6% | Frequent and recent buyers | Exclusive rewards, retention focus |
| VIP | <1% | High spend, highly engaged | Personalized, high-value retention plans |

# The "New and Frugal Customer"

## Segment Pairing - Cluster 1

*Low recency values but low frequency and monetary value, indicating recent but minimal engagement.*

| Metric | Value |
|---|---|
| Total Customers | 14,676 |
| Avg Recency | 255 |
| Avg Frequency | 2.5 |
| Avg Monetary Value | $84 |
| Avg Time on File | 476 |
| Avg Visits per Year | 3.2 |

***Personalized Incentive***

Send a personalized **welcome email** thanking them for their first purchase and introducing your brand's value propositions.

"Thank you for making your first purchase with us! To show our appreciation, we'd love to offer you **5% off your next purchase**! Just use the coupon code **WELCOME5** at checkout."



SCIENTIA CONSULTING

# The "Potential Loyal Customer"

## Segment Pairing - Cluster 2

*Moderate recency, frequency, and monetary value.*

| Metric | Value |
|---|---|
| Total Customers | 10,135 |
| Avg Recency | 313 |
| Avg Frequency | 8.5 |
| Avg Monetary Value | $310 |
| Avg Time on File | 1,940 |
| Avg Visits per Year | 1.4 |

*Personalized Incentive*

Provide them with **bundled offers** on complementary items to increase average order value. Based on their purchase history, also **recommend similar items.**

"Pair your [recent purchase] with [complementary product] for a perfect match – **save 10%** when you buy them together!

Based on your purchase history, we think you might love: **[Recommended Product #1]**."



SCIENTIA CONSULTING

# The "Long Lost Customer"

## Segment Pairing - Cluster 3

*High recency (haven't interacted recently), low frequency, and low monetary value.*

| Metric | Value |
|---|---|
| Total Customers | 6,980 |
| Avg Recency | 1,422 |
| Avg Frequency | 3 |
| Avg Monetary Value | $119 |
| Avg Time on File | 1,846 |
| Avg Visits per Year | 0.6 |

*Personalized Incentive*

**Special discount or coupon** to encourage re-engagement by sending email/message.

**"We miss you! Enjoy 5% off your next purchase using coupon GETBACK5 "**

# The "Loyal Customer"

## Segment Pairing - Cluster 4

*Very loyal to the brand, making frequent and recent purchases.*

| Metric | Value |
|---|---|
| Total Customers | 1,919 |
| Avg Recency | 154 |
| Avg Frequency | 27.1 |
| Avg Monetary Value | $1,104 |
| Avg Time on File | 2,165 |
| Avg Visits per Year | 4.8 |

*Personalized Incentive*

Offer **exclusive rewards** such as **early access** to products, **loyalty perks** like reward points programs, and **personalized incentives.**

"Celebrate your birthday with an **exclusive 5% discount** – because you deserve it!"



SCIENTIA CONSULTING

# The "VIP Customer"

## Segment Pairing - Cluster 5

*Extreme spenders, very recent*

| Metric | Value |
|---|---|
| Total Customers | 3 |
| Avg Recency | 117 |
| Avg Frequency | 10.7 |
| Avg Monetary Value | $389,662 |
| Avg Time on File | 1,636 |
| Avg Visits per Year | 2.9 |

*Personalized Incentive*

**Early access** to **limited edition products** or collections before they are available to the public. Provide also a **personalized shopping concierge service.**

"Be the first to explore our **limited-edition collections** and **new arrivals** before they go live to the public!"

"Need help finding the perfect product? Our dedicated concierge team is here for you!"



SCIENTIA CONSULTING

# The Personalized Recommendation Model Can Lead to ~1 Million Dollars Increase in Revenue

***Projected Growth with Personalized Recommendation***

| Customer Segment | Growth Rate | Estimated Revenue | Revenue Increase | Profit Increase |
|---|---|---|---|---|
| VIP | 20% | $1,402,782 | $233,797 | $70,139 |
| Loyal | 15% | $2,435,686 | $317,698 | $95,309 |
| Potential Loyal | 9% | $3,427,271 | $282,986 | $84,896 |
| New and Frugal | 5% | $1,289,303 | $61,395 | $18,419 |
| Long Lost | 4% | $867,317 | $33,358 | $10,008 |
| **Total** | **11%** | **$9,422,359** | **$929,234** | **$278,770** |

Estimated Revenue = Current Revenue + Revenue Increase
Profit Increase = Current Revenue x Growth rate x Profit Margin

**Current Revenue: $8,492,816**

Growth Assumptions:

- Higher rates for VIP and loyal customers (20-15%), medium rates for potential loyal (9%), and lower rates for new frugal and long-lost customers (5-4%).
- Purchase Frequency Increase:10%
- Spending Per Purchase Increase: 20%
- Profit Margin: 30%



SCIENTIA CONSULTING

# Scaling Personalized Recommendations for Maximum Impact

Projected Cost: $30,000

| Task/ Deliverable* | | Q1 | | | Q2 | | |
|---|---|---|---|---|---|---|---|
| | | Dec 2024 | Jan 2025 | Feb 2025 | Mar 2025 | Apr 2025 | May 2025 |
| 1 | *Refine Customer Segmentation* ***$5,000*** | ■ | | | | | |
| 2 | *Implement quarterly personalization* ***$6,000*** | | ■ | | | | |
| 3 | *A/B Testing* ***$4,000*** | | | ■ | ■ | | |
| 4 | *Scale the Infrastructure for Recommendations* ***$15,000*** | | | | | ■ | ■ |

*details of tasks in appendix



SCIENTIA CONSULTING

# Key Takeaways

## Intelligent Segmentation

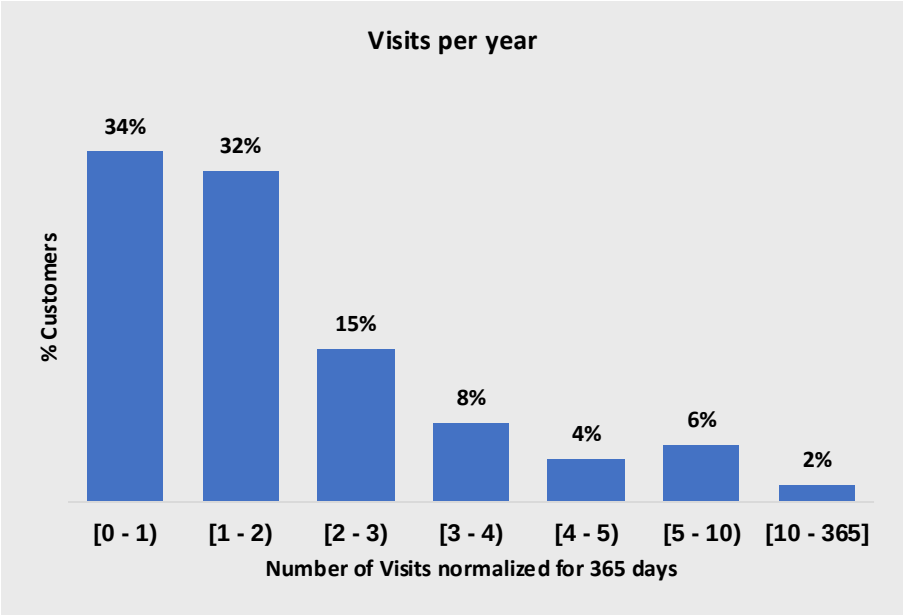
5 distinct clusters enabling personalized recommendations by K-Means clustering.

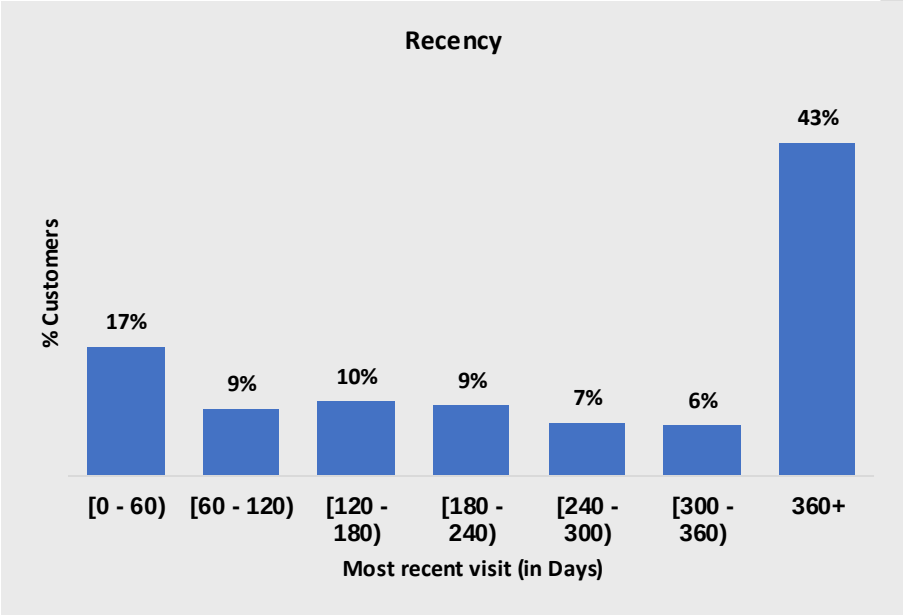
## Revenue Impact

Increase overall revenue by 11% and an additional profit of $280,000.

# Appendix

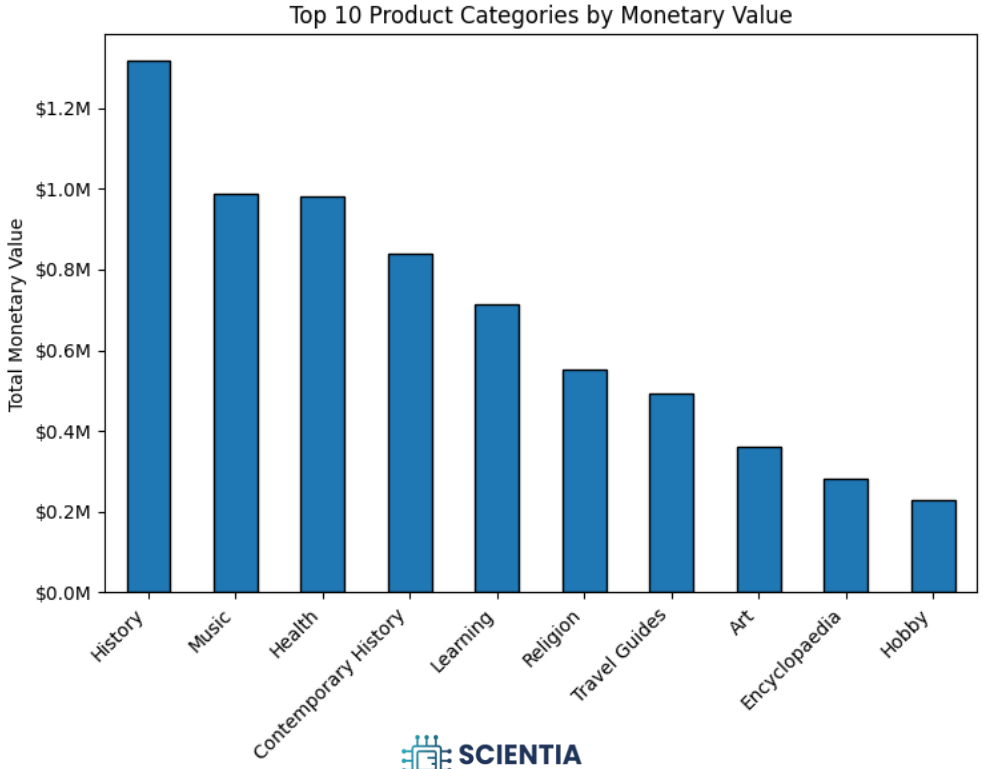
# A majority of customers visit the site infrequently and have less recent purchases.



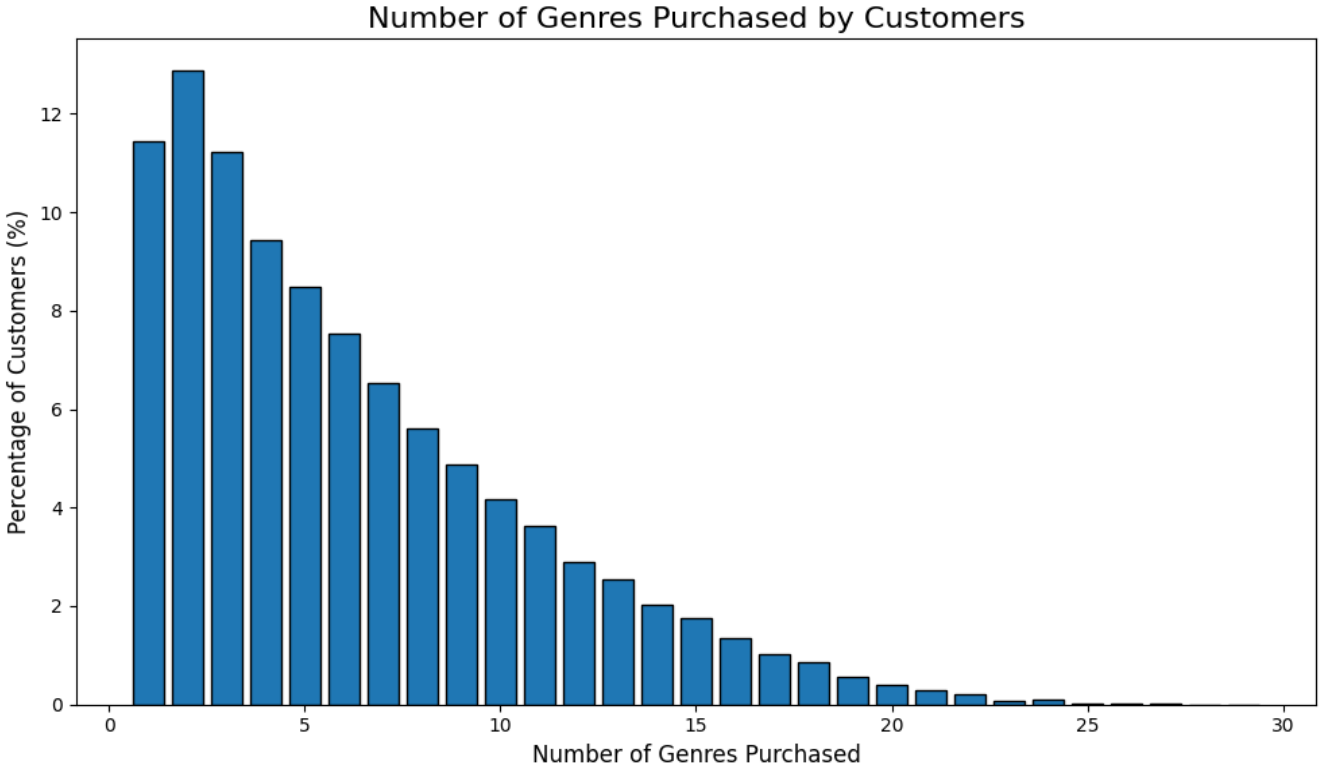

SCIENTIA CONSULTING

# History, Music, and Health books contribute the most to sales.



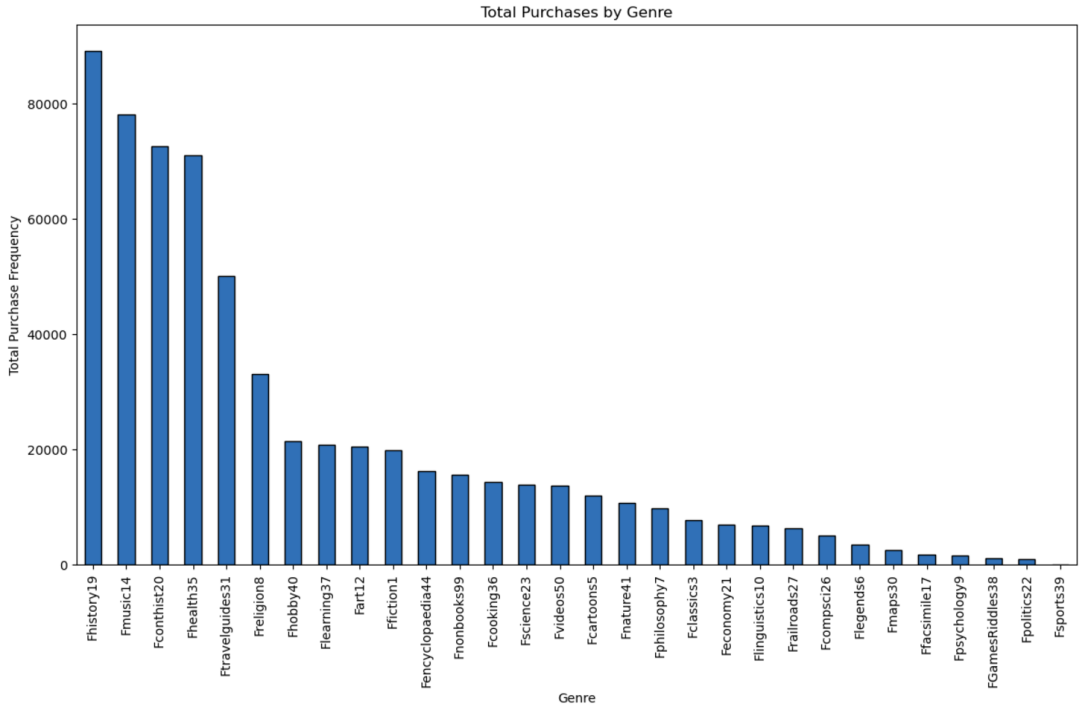

SCIENTIA CONSULTING

# Most customers purchase fewer than 5 genres.

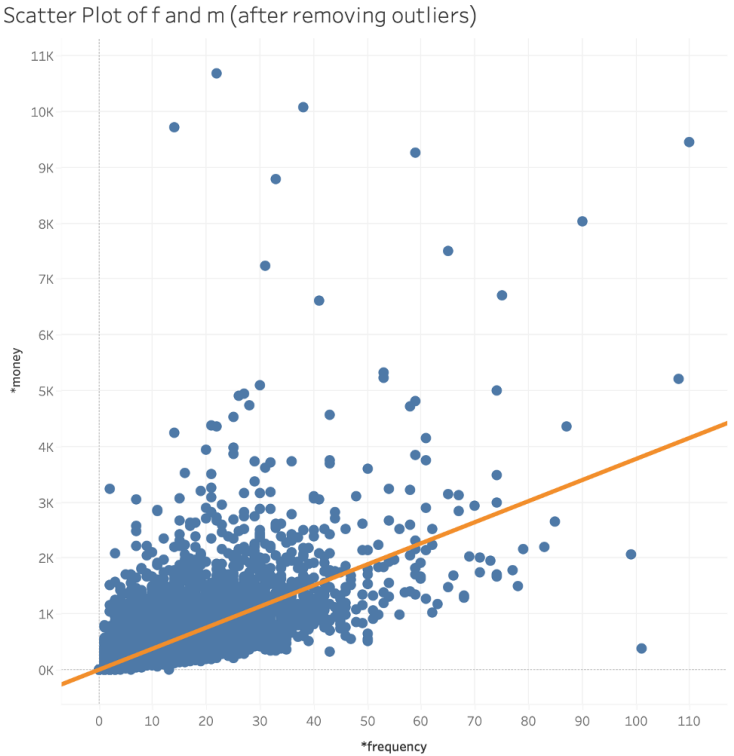



SCIENTIA CONSULTING

# History is the most frequently bought genre meanwhile sports books are the least frequently bought.

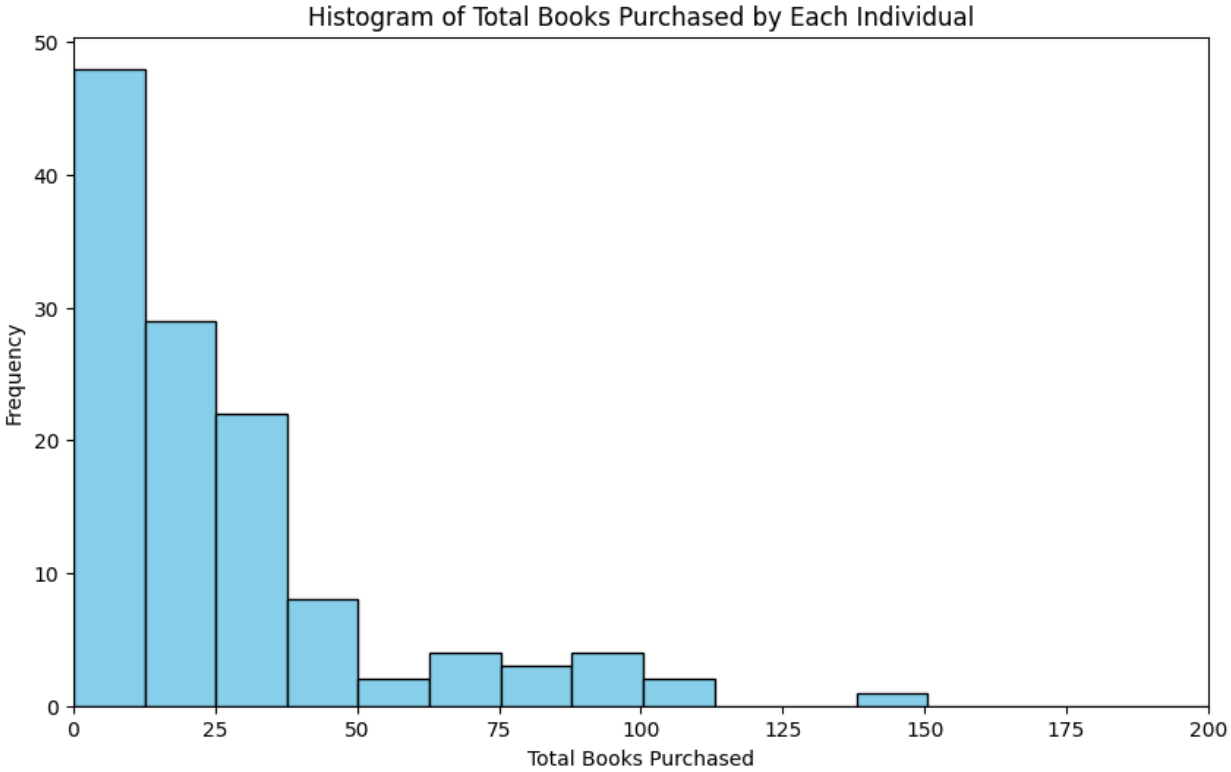


# There is an increase in money spent as frequency increases.




Sum of *frequency vs. sum of *money. Details are shown for Id. The view is filtered on Id, which excludes 5405602, 5900190, 8192553, 13729756 and 14158205.

**Most customers buy less than 25 books.**



# Benefits of K-Means as a Practical Solution for Effective Segmentation

**Key Advantages of K-Means**

- ***Simplicity and Efficiency:*** K-Means is computationally efficient for large datasets, making it ideal for creating a recommendation engine that needs to handle a significant volume of data in a timely manner.

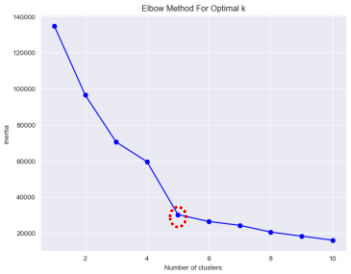
- ***Interpretability:*** The clusters formed by K-Means are easy to interpret and visualize, which helps in clearly understanding the distinct segments of book categories.

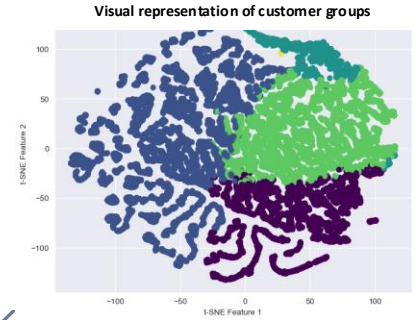
- ***Scalability:*** As our dataset grows, K-Means remains a scalable option, adapting well to higher dimensional data.

- ***Objective Alignment:*** The algorithm's goal of minimizing intra-cluster variance aligns perfectly with the need to group similar customers together to make accurate recommendations.

# Customers Can Be Segmented Into 5 Clusters

- We selected **k=5** based on within-cluster sum-of-squares elbow plot for our K-means model.
- The clusters are fairly differentiated as seen in the table.




| | Clusters | | | | |
|---|---|---|---|---|---|
| | [1] **Long Lost** | [2] **New Frugal** | [3] **Loyal** | [4] **Potential Loyal** | [5] **VIP** |
| #customers | 6,980 | 14,676 | 1,919 | 10,135 | 3 |
| %customers | 21% | 44% | 6% | 30% | 0% |
| avg **Recency** | 1422 | 255 | 154 | 313 | 117 |
| avg **Frequency** | 3.0 | 2.5 | 27.1 | 8.5 | 10.7 |
| avg **Amount** | $ 119 | $ 84 | $ 1,104 | $ 310 | $ 389,662 |
| avg **Time on File** | 1846 | 476 | 2165 | 1940 | 1636 |
| avg **Visits per Year** | 0.6 | 3.2 | 4.8 | 1.7 | 2.9 |

SCIENTIA CONSULTING

# Advertisement Strategies From Clustering

| Cluster ID | Segment | Total Customer | Recommendation |
|---|---|---|---|
| 1 | Long lost customers | 6,980 (21%) | Offer reactivation incentives such as "Welcome Back" discounts or special campaigns to entice them to re-engage. |
| 2 | New and Frugal Customers | 14,676 (44%) | Offer upselling opportunities, such as discounts on higher-value items or loyalty program enrollment. |
| 3 | Loyal customers | 1,919 (6%) | Provide exclusive rewards, loyalty perks, or early access to products to strengthen their loyalty. |
| 4 | Potential loyal customers | 10,135 (30%) | Encourage higher spending through bundle offers. |
| 5 | VIP Customers | 3 (<1%) | Early access to new product before launching or sell on public for same product or product in same genre |

# Business Impact Estimate

**Current Baseline Revenue**

| Segment | Customers | Avg. Spending | Current Revenue | % of Revenue |
|---|---|---|---|---|
| Long Lost | 6,980 | $119 | $830,620 | 9.8% |
| New and Frugal | 14,676 | $84 | $1,232,784 | 14.5% |
| Loyal | 1,919 | $1,104 | $2,118,576 | 24.9% |
| Potential Loyal | 10,135 | $310 | $3,141,850 | 37.0% |
| VIP | 3 | $389,662 | $1168,986 | 13.8% |
| **Total** | **33,713** | **$251.92** | **$8,492,816** | 100% |

Current Revenue = Number of Customers x Average Spending

# Growth rate calculation for clusters

***General Growth Rate Projected by eCommerce Bookstore Based on Engagement***

| Customer Segment | Growth Rate | Current Revenue | Revenue Increase |
|---|---|---|---|
| Top 15% Engagement | 20% | $3,344,065 | $668,813 |
| Bottom 85% Engagement | 5% | $5,149,060 | $257,453 |
| **Total** | **11%** | **$8,493,125** | **$929,266** |

***General Growth Rate Projected by Scientia Based on Clusters***

| Customer Segment | Growth Rate | Current Revenue | Revenue Increase |
|---|---|---|---|
| VIP | 20% | $1,168,985 | $233,797 |
| Loyal | 15% | $2,117,988 | $317,698 |
| Potential Loyal | 9% | $3,144,285 | $282,986 |
| New and Frugal | 5% | $1,227,908 | $61,395 |
| Long Lost | 4% | $833,959 | $33,358 |
| **Total** | **11%** | **$8,493,125** | **$929,234** |

Individual clusters' growth rates extrapolated while keeping the projected increase in revenue same

# Details of Each Implementation Deliverable

We will complete the following steps to implement and deploy our recommender system:

1. **Refine Customer Segmentation**
   - Collect more data (e.g. browsing history, book ratings, and demographics) to create more personalized recommendations and implement the model.
2. **Implement quarterly personalization**
   - Use website session data to personalize recommendations periodically as a customer is using the website
3. **A/B Testing**
   - Compare the performance of the new model on the most engaged customers versus the performance of the current model
   - Use results to refine the new model
4. **Scale the Infrastructure for Recommendations**
   - Use batch processing for daily updates to the data
   - Utilize cloud infrastructure to handle large volumes of user data and real-time session-based recommendations