Enhancing Ecommerce Recommendation

Utilizing RFM Analysis for Tailored Book Shopping Experience

Group 2
Dec 1, 2023
Presenters: Khushi, Jiayi, Yixuan, Andy, Jason
ADSP 31016 IP04 – Leadership & Consulting for Data Science

Alpha Analytics

Alpha Analytics

Agenda

- Introduction
- Exploratory Data Analysis
- Model Selection, Construction and Assessment
- Tailored Customer Incentives
- Business Impact
- Conclusion

Alpha Analytics

Introduction

Business Problem:

The current ecommerce platform relies on a 'most popular' item recommendation strategy to boost cart additions and increase average purchase sizes. However, this approach falls short in addressing individual customer preferences.

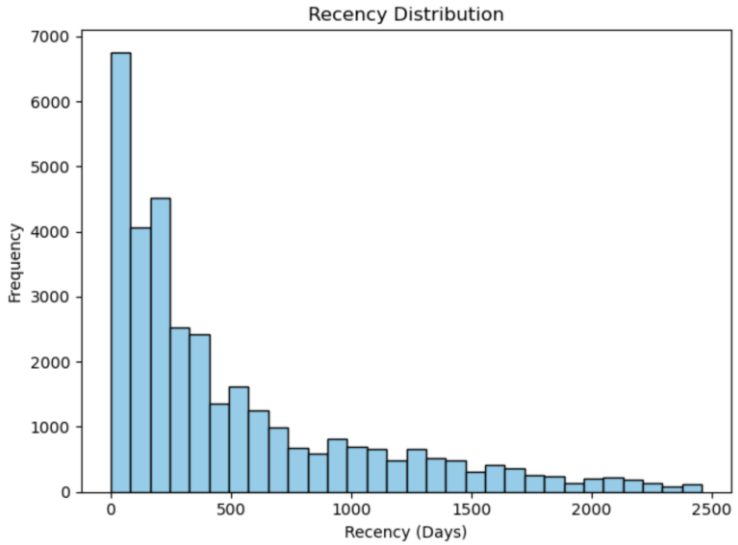
Analysis Goal:

Our objective involves utilizing customer transaction data to generate personalized recommendations, with the aim of enhancing engagement and sales beyond the constraints of the current approach.

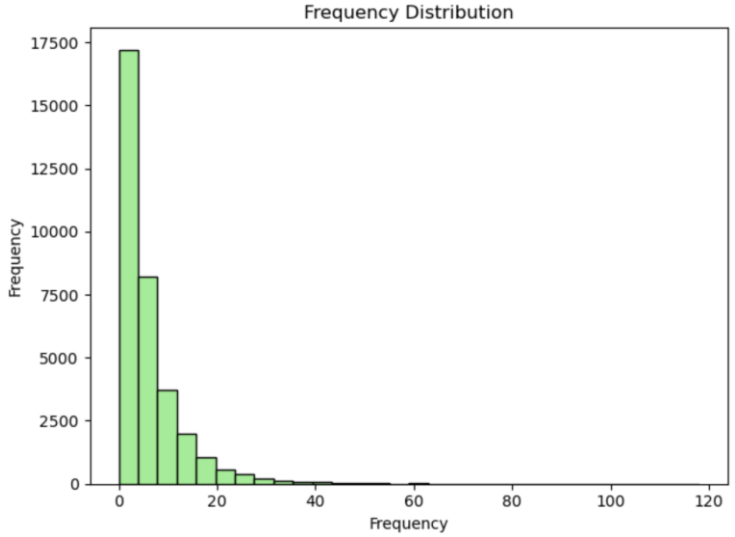
Distributions of *Recency*, *Frequency* & *Monetary Value* are right-skewed

EDA → Method Comparison → Modelling and Assessment → Incentives → Business Impact

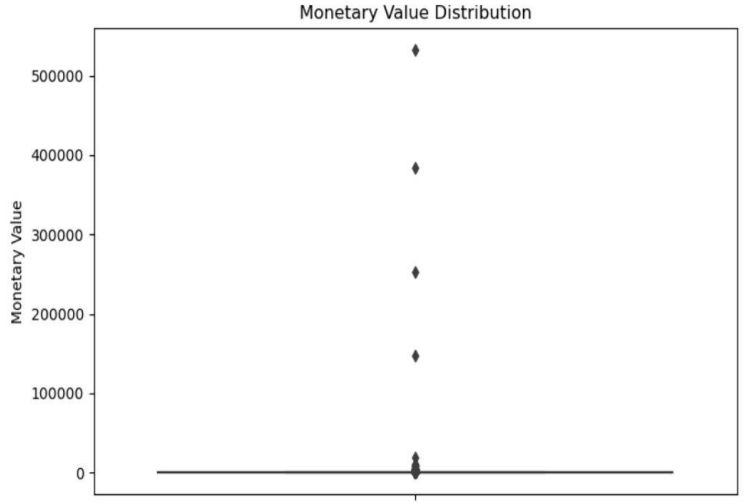
- 33,713 entries
- 25,402 Missing Values in *logtarg*
- Otherwise no null values in the data
- Distributions of *Recency*, *Frequency* and *Monetary Value* are right-skewed (see graph 1, 2)
- Existence of outliers in *Monetary Value* => Several customers making very high purchase amounts (graph 3)



Graph 1



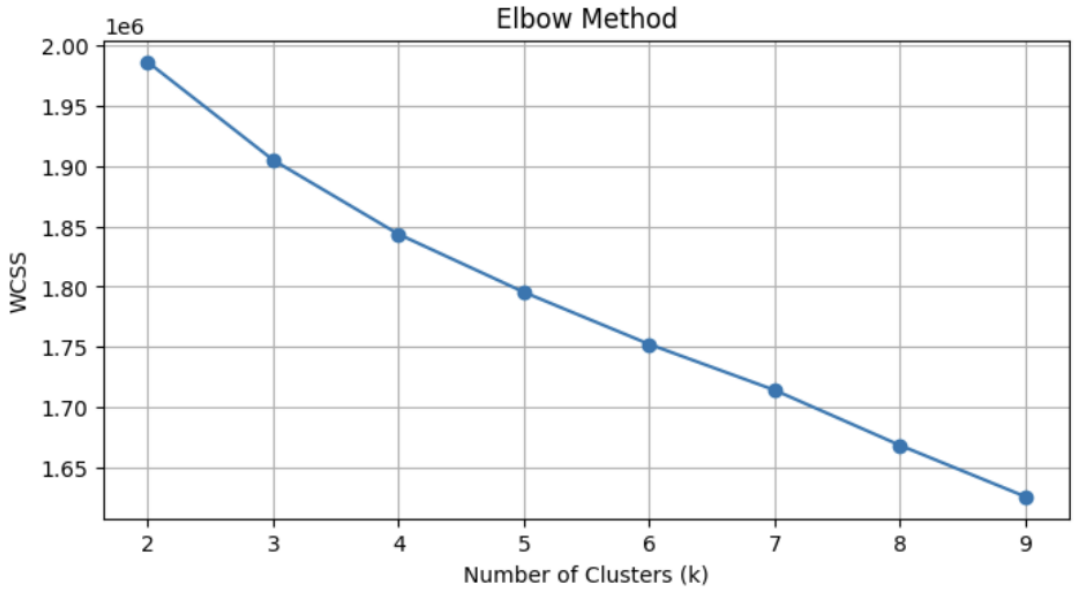
Graph 2



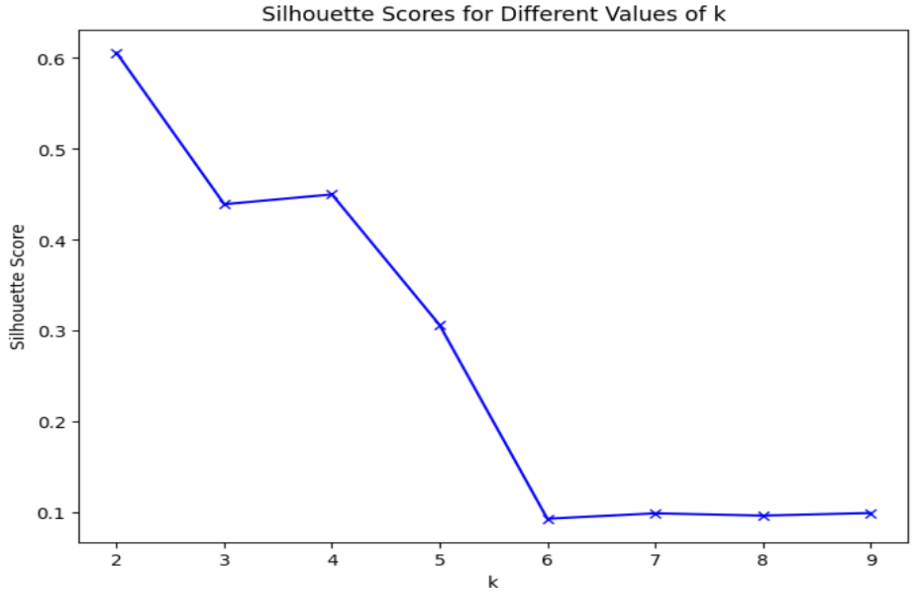
Graph 3

K-means is not perfect for this dataset

EDA → **Method Comparison** → Modelling and Assessment → Incentives → Business Impact



We applied **Elbow Method** to find optimal k value for k-means model. But there is no inflection point (turning point) in this graph.

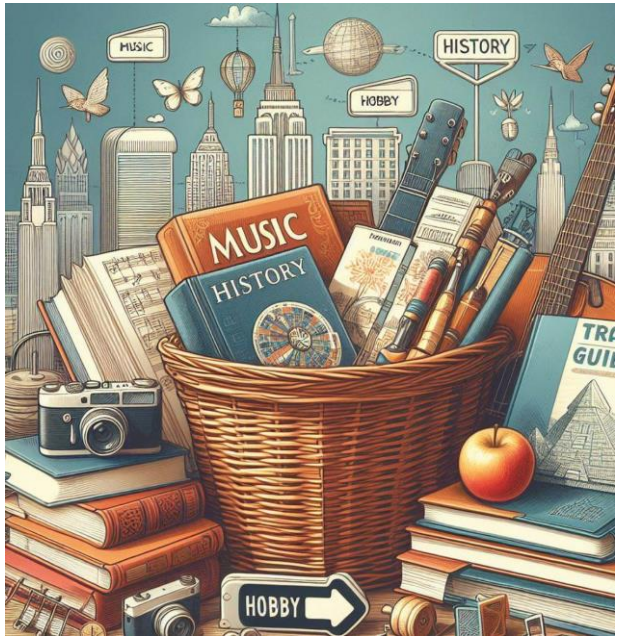


Silhouette Scores are really low (<0.5) while k is beyond 3. K-Means might not be the ideal choice for customer segmentation..

Alpha Analytics

Association Mining is not suitable for segmentation

EDA → **Method Comparison** → Modelling and Assessment → Incentives → Business Impact



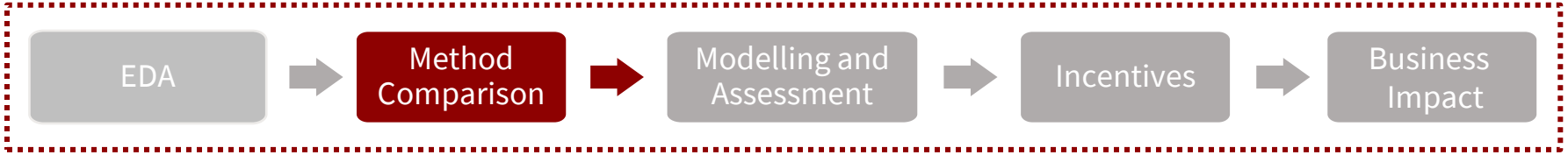
Association mining identifies patterns within data, revealing relationships or co-occurrences between items in a dataset

Reasons why it's not an appropriate method:

1. More focused on relationships between items **rather than on segmenting customers.**
2. Limited to Binary Data: Traditional association mining methods work with binary data, where items are either present or absent in a transaction. This limitation makes it less suitable for datasets with continuous or numerical attributes without proper discretization.

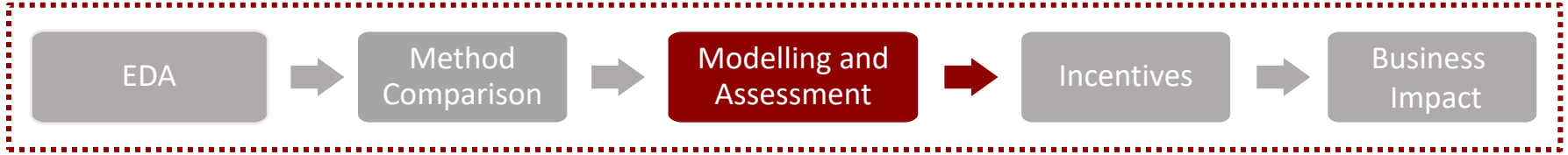
Alpha Analytics

Why RFM is the best way to segment



1. **Clear threshold for segmentation**: RFM thresholds are set by analysts based on specific business objectives and domain knowledge.
2. **Flexible**: RFM analysis can be applied effectively to both small and large customer datasets.
3. **MECE(Mutually Exclusive & Comprehensively Exhaustive)**: it provides a structured and systematic approach to customer segmentation that ensures all customers are appropriately categorized into segments without overlapping, resulting in a comprehensive and well-organized understanding of the customer base.

Assigning RFM scores based on quantiles



RFM Modeling Process:

1. **Segmenting by Quantiles**: Segment *Recency*, *Frequency* and *Monetary Value* respectively using quartiles (25%, 50%, 75%), with scores from 1 to 4 (integer).
   - Lower *Recency* gets higher R_Score
   - Higher *Frequency* gets higher F_Score
   - Higher *Monetary Value* gets higher M_Score

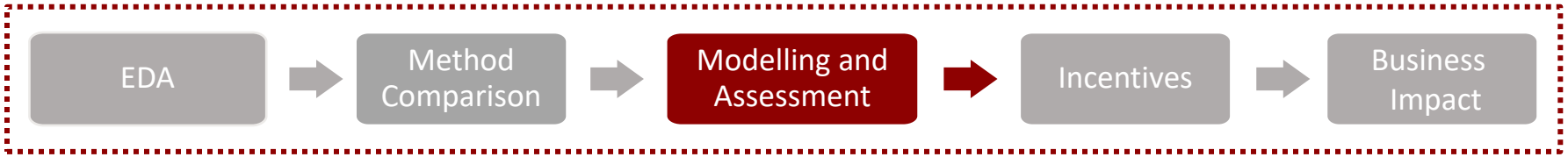
Table 1 Boundaries of *r*, *f* & *m*

	r	f	m
25%	117	1	39.90
50%	293	3	102.63
75%	712	7	247.29

2. **Handling Outliers**: For *Monetary Value*, consider a fixed monetary threshold of $100,000 for an additional category. Customers with a *Monetary Value* exceeding $100,000 will receive an M_Score of '5'.
3. **Assigning Scores**: Combine individual R, F and M Score into a unified three-digit RFM Score for each customer.

*See Appendix C for sample RFM scores

All combinations of RFM Scores are categorized into 9 customer segments without overlapping or omission



1)High-Value Segments	**VIP Customers** M:5	**Big Spenders** R:2-4, M:4	**Loyal Customers** F:4, M:1-3
2)Potential Growth Segments	**Recent Incoming Customers** R:4, F:1-3, M:1-3	**Promising Customers** R:2-3, F:2-3, M:2-3	**At Risk** R:2-3, F:2, M:1-3/ R:1, M:4
3)Retention and Engagement Segments	**Hibernating** R:2-3, F:1, M:1-3	**Low-Value Customers** R:1-3, F:2-3, M:1	**Almost Lost** R:1, F:1, M:1-3

Verifying RFM model using synthetic new customers

EDA → Method Comparison → **Modelling and Assessment** → Incentives → Business Impact

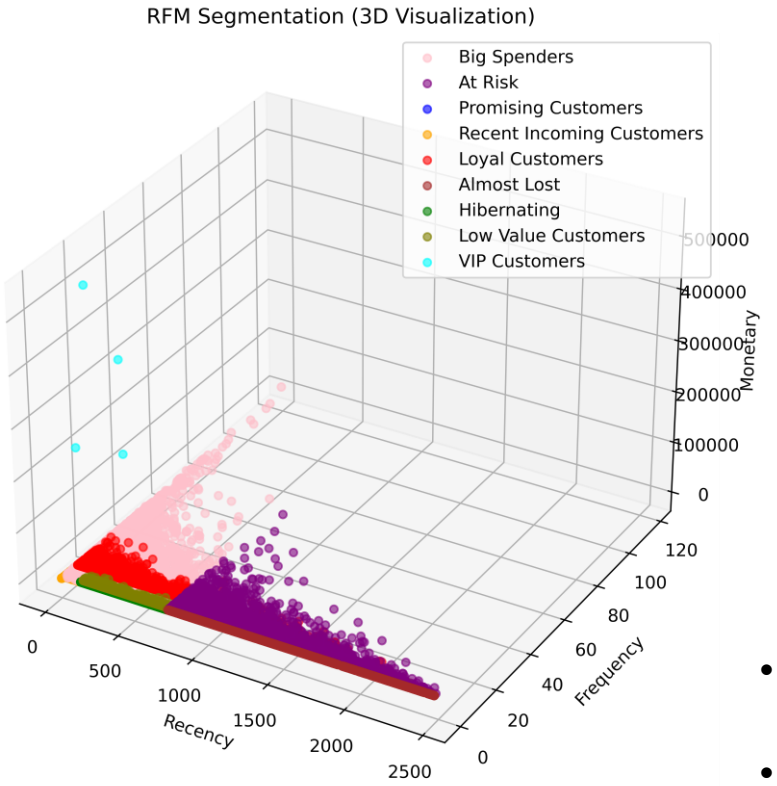
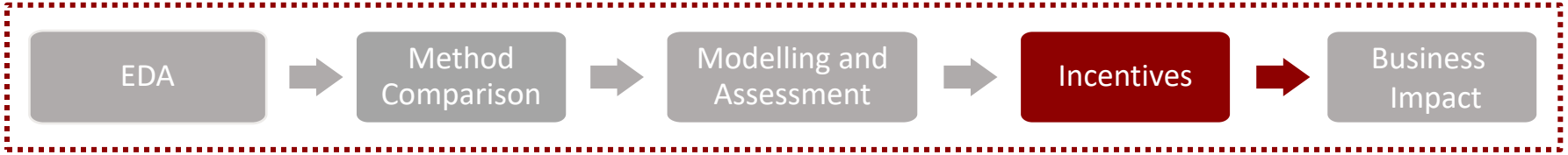


Table 2 Synthetic New Customers

id	r	f	m	Segment
1	238	1	237.54	Promising Customers
2	668	9	86.72	Loyal Customers
3	463	2	114.53	Promising Customers
4	86	1	386.02	Big Spenders
5	30	2	55.91	Recent Incoming Customers
6	1021	4	181.23	At Risk
7	466	3	225.27	Promising Customers
8	625	2	11.94	Low Value Customers

- Synthetic new customer data are randomly generated from exponential distributions based on the shape of original distributions of *r*, *f* & *m*.
- The segments of new customers as well as the visualization are aligned with the set logic and MECE principle.

Incentives for High-Value Segments



Objective: Foster Strong Loyalty and Drive Incremental Revenue from Top Spenders.

VIP Customers

- Examples – Schools

Incentive Plan:
Aiming to **maintain satisfaction and retain** loyalty among VIP customers, we offer an exclusive annual discount of 10% on all purchases.

Assumptions:
Based on this, schools buy more books, 20% more in monetary value

Big Spenders

Incentive Plan:
Yearly $50 Voucher for spending over $500 in a calendar year

Assumptions:
Expect a 15% rise in average spending from this segment to reach the $500 threshold, but only 80% of customers eligible redeem this reward

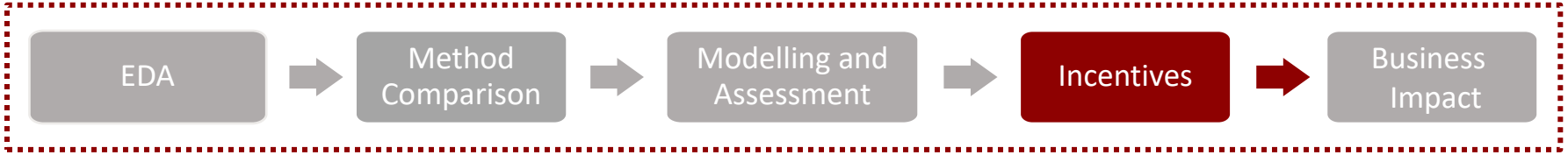
Loyal Customer

Incentive Plan:
Loyalty Points Program Earn 1 point/$10 spent, redeemable at 100 points for a $10 voucher.

Assumptions:
Because of this, there is an average 15% rise in spending for each customer

Alpha Analytics

Incentives for Potential Growth Segments



Objective: Encourage Repeat Purchases and Prevent Churn.

Recent Incoming Customers

Incentive Plan:
Second Purchase Discount with 15% off on second purchase.

Assumptions:
Overall 5% increase in customer sales across all buyers in this segment

Promising Customers

Incentive Plan:
Focus on nurturing and converting them into loyal customers. Second Purchase Discount with 10% off on second purchase.

Assumptions:
Overall 3% increase in customer sales across all buyers in this segment

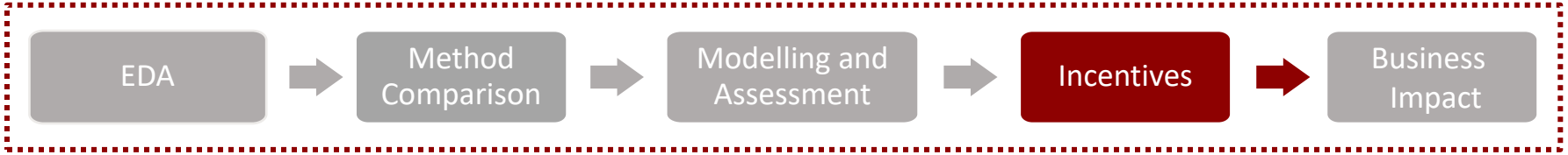
At Risk

Incentive Plan:
Implement strategies to re-engage and retain these customers: Re-Engagement Voucher worth $10. Redeemable in 30 days.

Assumptions:
Only 50% will redeem, but they will be grouped into 'promising customer' the year after.

Alpha Analytics

Incentives for Retention and Engagement Segments



Objective: Revitalize Engagement and Enhance Lifetime Value of Customers

Hibernating

Incentive Plan:
Rekindle their interest through - $5 coupon. No limit.

Assumptions:
Only 30% will redeem, but they will be grouped into 'promising customer' the year after.

Low-Value Customers

Incentive Plan:
Promotional campaign on $3 books. This will increase overall traction and move some customers to the 'promising customer' segment

Assumptions:
Set the price at the average cost of the book so revenue stays even.

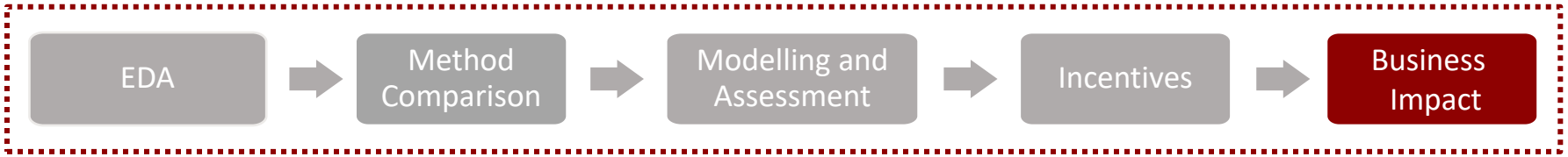
Almost Lost

Incentive Plan:
These customers probably buy books once a year, for their friends/family as gifts. So, set a 'buy any book get a $5 off redeemable next year!'

Assumptions:
Overall 6% increase in customer sales across all buyers.

Alpha Analytics

Final Outcome: a revenue surge of $670,000



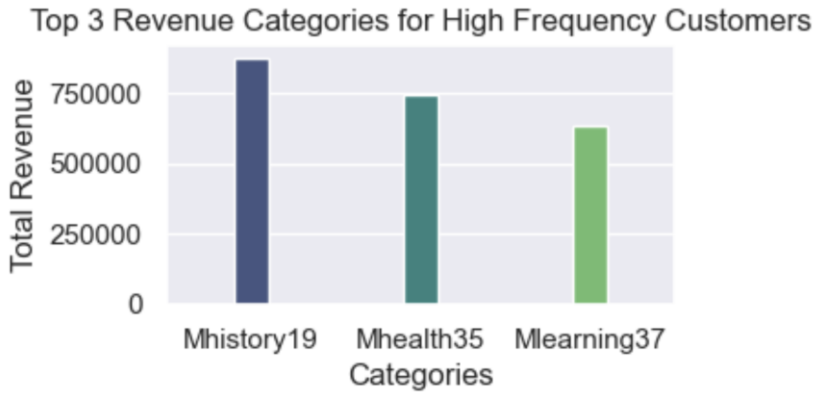
Customer Segment	# of Customers	Total Revenue Before	Projected Next Year's Revenue*	Change in Revenue (Year over Year)
VIP Customers	4	$ 1,316,669.50	$ 1,422,003.06	$ 105,333.56
Big Spenders	7,209	$ 4,461,907.04	$ 4,980,633.10	$ 518,726.06
Loyal Customers	1,958	$ 356,682.34	$ 390,924.69	$ 34,242.35
Recent Incoming Customers	4,603	$ 366,995.63	$ 385,345.41	$ 18,349.78
Promising Customers	5,976	$ 674,064.67	$ 694,286.61	$ 20,221.94
At Risk	5,039	$ 1,028,365.65	$ 1,003,170.65	$ (25,195.00)
Hibernating	4,673	$ 143,619.42	$ 136,609.92	$ (7,009.50)
Low-Value Customers	1,687	$ 44,958.92	$ 44,958.92	$ 0
Almost Lost	2,564	$ 99,861.60	$ 105,853.30	$ 5,991.70

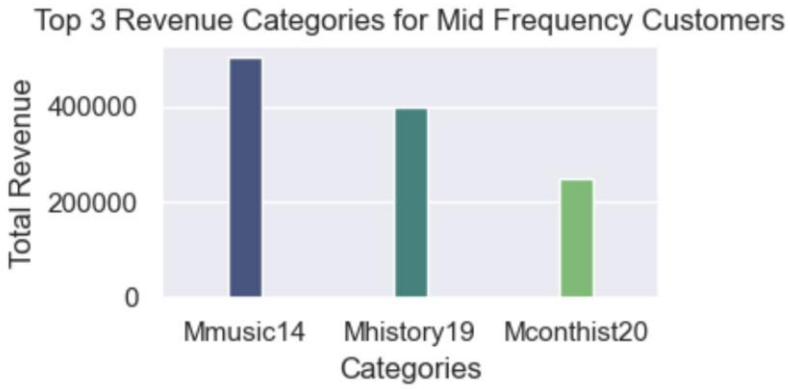
*See Appendix F for full calculation

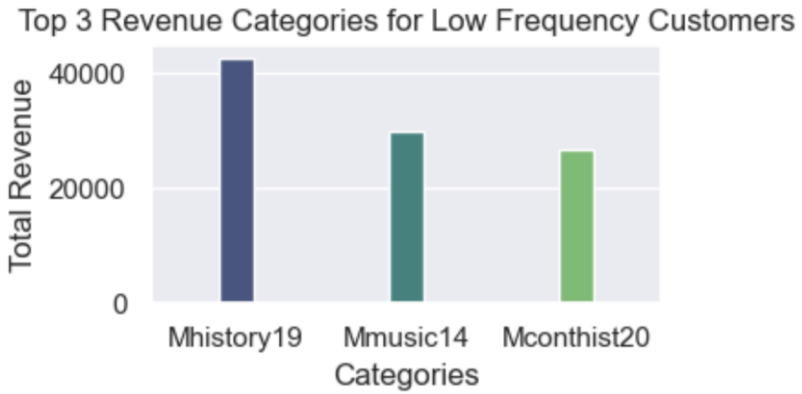
Conclusion

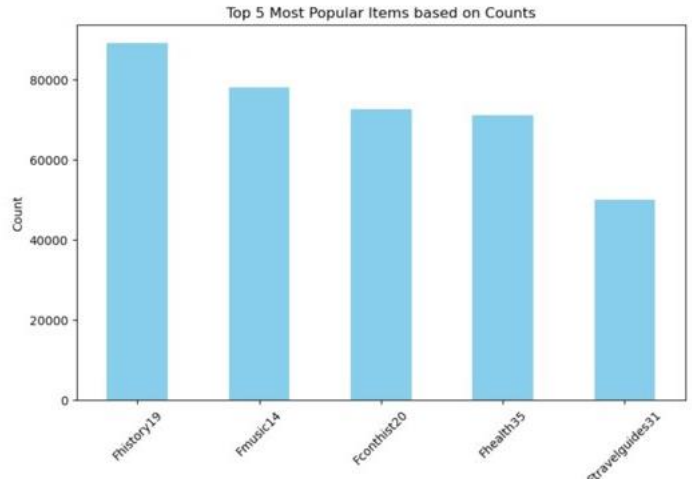
- Insights from EDA:
  - Varied customer engagement level from right-skewed distribution of *r*, *f* and *m*
  - Particular requirement for top-tier customer segment due to outliers in *Monetary(m)*
- Methodological Exploration and Decision:
  - Evaluated K-means clustering and Association Mining but found them less suited for our nuanced customer segmentation needs
  - Adopted RFM analysis due to its robustness in identifying customer behavior patterns
- RFM Segmentation Implementation:
  - Created 9 customer segments, each aligned with tailored incentives, elevating our strategy beyond generic "most popular" item recommendations
  - Our RFM model adheres to MECE principle, ensuring clear and actionable segmentation
- Business Impact:
  - The solution demonstrates the potential business impact of a significant revenue surge of $670,000 with reasonable assumptions

Appendix A – EDA of Book Categories





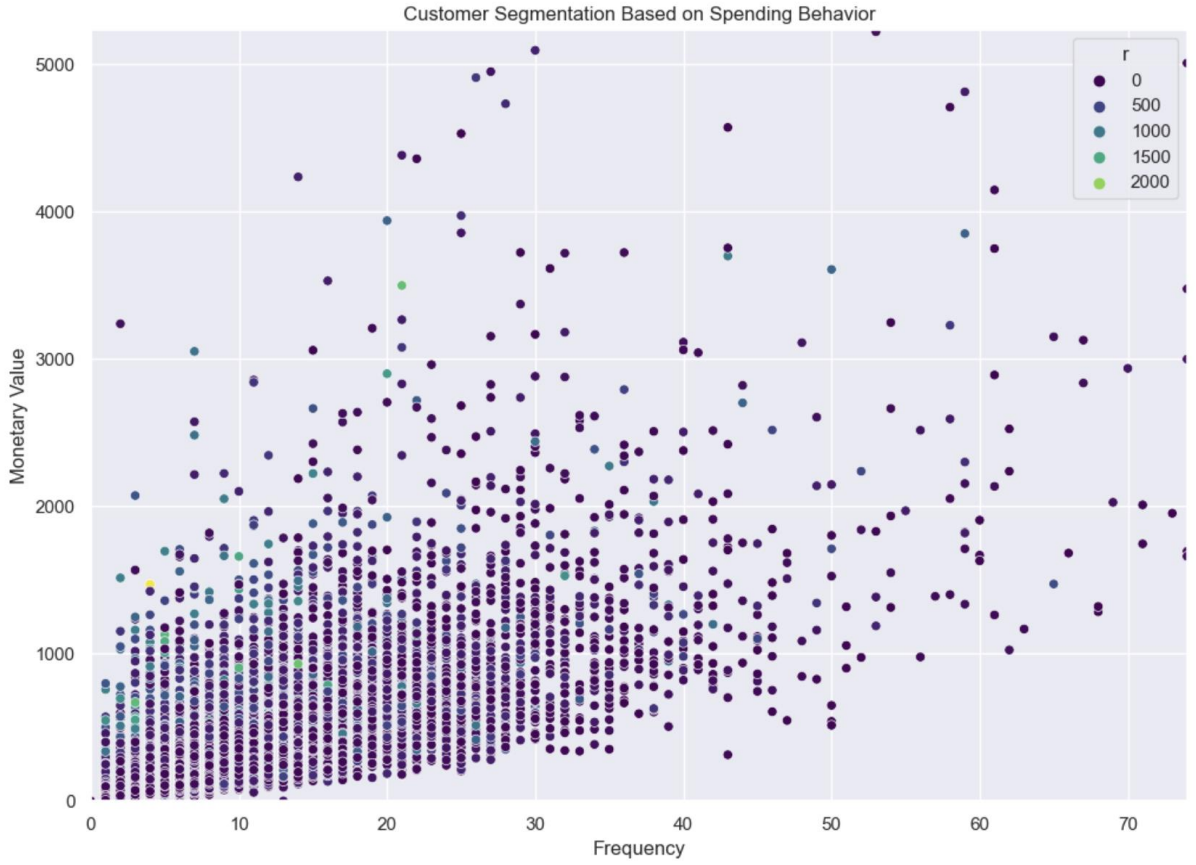




History Music Cont. History Health Travel

Alpha Analytics

Appendix B – EDA Customer Spending Pattern



Appendix C – Sample RFM Score

id	r	f	m	R_Score	F_Score	M_Score	RFM_Score
914	194	7	318.89	3	3	4	334
957	3	14	368.05	4	4	4	444
1406	1489	15	423.30	1	4	4	144
1414	155	4	71.22	3	3	2	332
1546	194	6	422.64	3	3	4	334
1651	1797	2	47.93	1	2	2	122

Appendix D – Top Association Mining Outputs

Top 5 Association Rules with Highest Lift:

	antecedents	consequents	lift
129	(Fmusic14, Fhobby40, Fhistory19)	(Ftravelguides31)	1.9785
139	(Fconthist20, Fhobby40, Fhistory19)	(Ftravelguides31)	1.9504
145	(Fhobby40, Fhealth35, Fhistory19)	(Ftravelguides31)	1.9448
111	(Ffiction1, Fhealth35, Fhistory19)	(Ftravelguides31)	1.9251
98	(Ffiction1, Fmusic14, Fhistory19)	(Ftravelguides31)	1.9037

Appendix E - Logic of Segmentation

Segment	RFM Score Like	Explanation(Talking Points)
VIP Customers	__5	Top-tier customers with the highest monetary contribution
Big spenders	[2,3,4]_4	Customers with recent large amounts of spending, though not necessarily spend as much as VIPs
Loyal Customers	_4[1,2,3]	Customers who shop frequently and have a strong relationship with the bookstore
Recent Incoming	4[1,2,3][1,2,3]	They have started shopping with you recently
Promising Customers	[2,3][2,3][2,3]	Recent customers with average frequency, showing potential to become more valuable
At Risk	1[2,3][2,3]/1_4	These were once regular customers but haven't made purchases recently
Hibernating	[2,3]1[1,2,3]	Customers who do not purchase very frequently
Low-Value Customers	[1,2,3][2,3]1	Customers who spend small amounts across history
Almost Lost	11[1,2,3]	Customers who may have churned or chosen a competitor

Appendix F - Calculation of all incentive schemes

VIP (like schools):
Assumption: 10% discount of every purchase, leading to 20% more in monetary value.
- ($ 1,316,669.50) x 90% x 120% = **$1,422,003.06**

Big Spenders:
Assumption: Yearly $50 Voucher for spending over $500 in a calendar year. Expect a 15% rise in average spending from this segment.
- After applying 15% increase to each customer in this segment, 3445 is still below $500, 3764 is above $500 and qualifies for the voucher.
- We only assumed 80% would redeem the free $50 (rest might have forgotten, etc.)
- So, ($4,461,907.04 x 1.15) - 3764*$50*80% = **$4,980,633.10**

Loyal Customers:
Assumption: Loyalty Points Program Earn 1 point/$10 spent, redeemable at 100 points for a $10 voucher. Expect a 15% rise in average spending.
- After applying 15% increase to each customer in this segment, 32 still don't spend $100. Since the rest are loyal, they will all redeem the $10.
- So, ($356,682.34 x 1.15) - 1926*$10 = **$390,924.69**

Recent Incoming/Promising/Almost Lost Customers:
Calculation: This is just a (1 + __%) * Previous Revenue

ANY VOUCHER – for At Risk/Hibernating:
Calculation: x% that will redeem x Coupon Value x How many Customers