

# REVOLUTIONIZING BOOK RECOMMENDATIONS

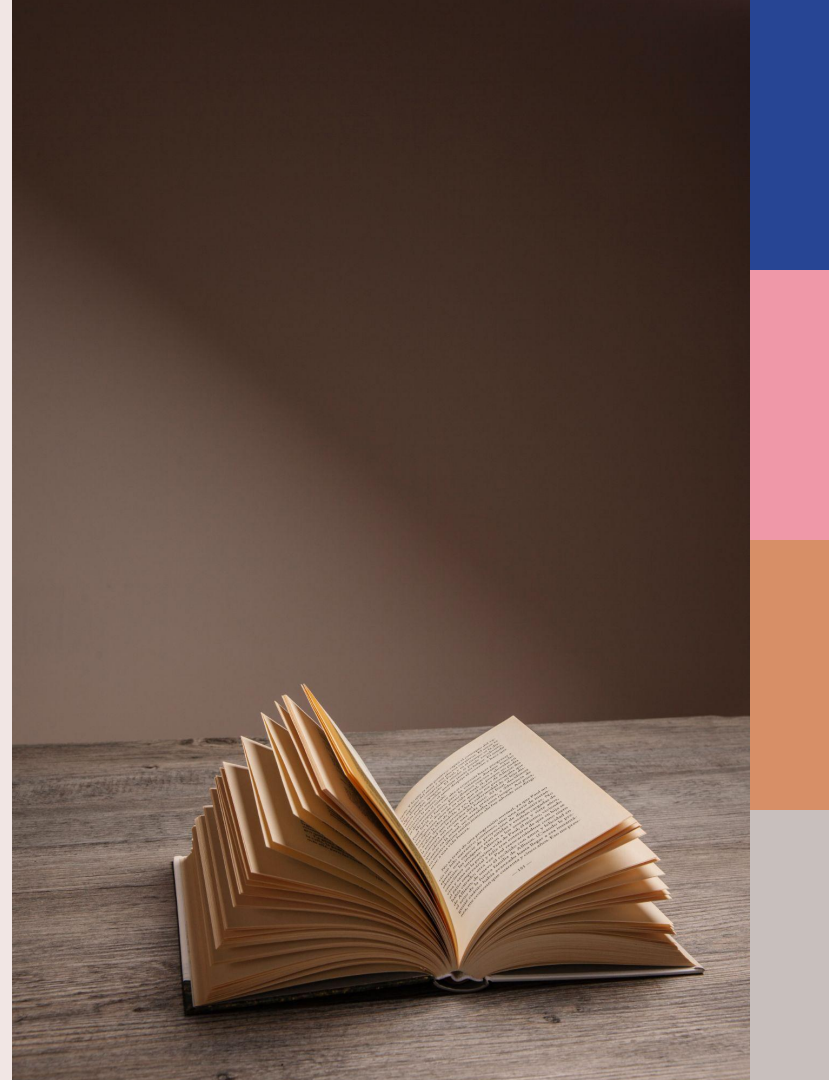
## Team 4

Yash Bhardwaj,  
Nathan Chen,  
William DeForest,  
Daichi Ishikawa,  
Siobhan McDermott



# AGENDA

1. MEET OUR TEAM
2. PROBLEM STATEMENT
3. RFM SEGMENTATION
4. DISCOUNTS & RECOMMENDATIONS
5. COMPARISON TO CURRENT SYSTEM
6. BUSINESS IMPACT



# MEET OUR DIVERSE TEAM OF PROFESSIONALS



Team Lead  
Background: Healthcare, Data  
Management, R, Python

—**Siobhan McDermott**



Analytics Consultant  
Background: Finance, Statistics.  
Machine learning

—**Daichi Ishikawa**



Analytics Consultant  
Background: Machine Learning,  
Python, GCP Cloud Architect

—**Yash Bhardwaj**



Analytics Consultant  
Background : Sport Analytics,  
Bilingual, skilled in R and Tableau

—**Nathan Chen**



Analytics Consultant  
Background: Seattle native, baseball  
player, skilled in R and machine learning

—**William DeForest**

# CURRENT RECOMMENDATION SYSTEM IS LEAVING REVENUE ON THE TABLE

## PROBLEM

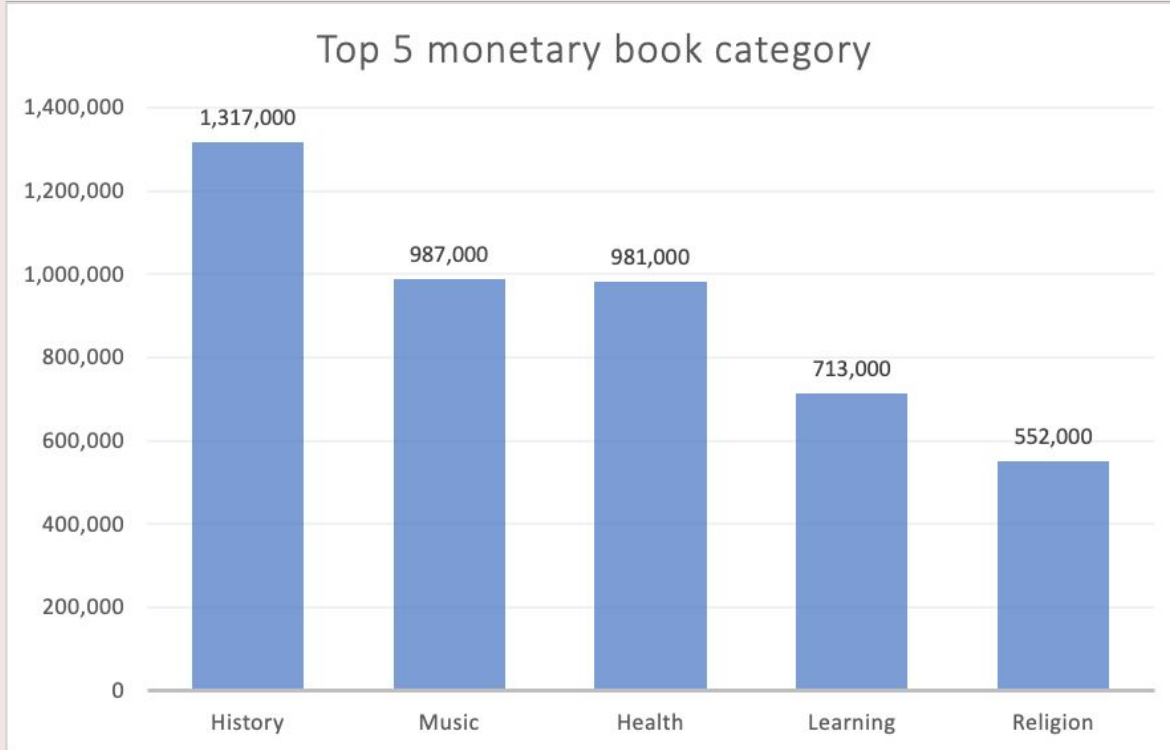
Your e-commerce bookstore's current recommendation system does not tailor recommendations to each customer's specific purchasing habits which is limiting average transaction size.

## SOLUTION

By identifying and understanding the purchasing behavior of various user groups, your company can increase the average size of each transaction by providing more targeted recommendations.

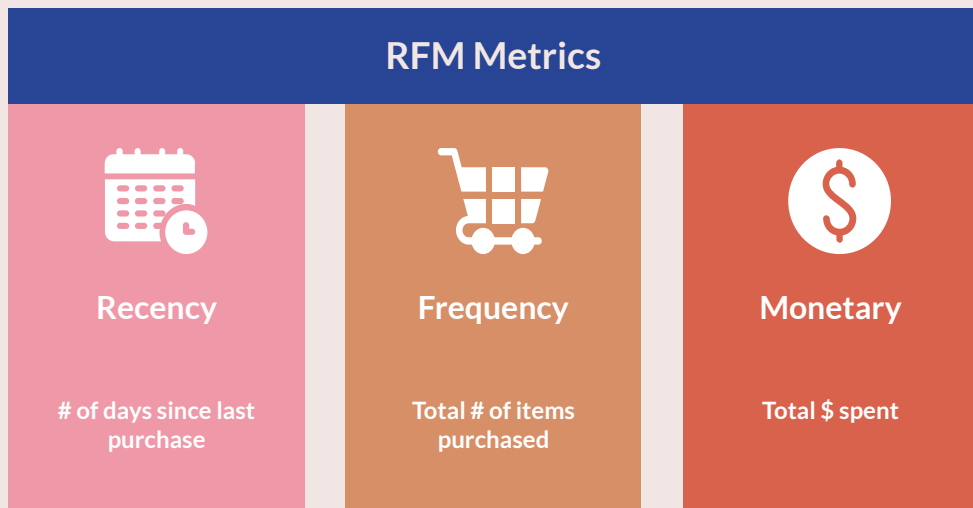


# CURRENT SYSTEM SIMPLY RECOMMENDS “MOST POPULAR” CATEGORIES



- The current system recommends "most popular" items
- Not personalized enough for each customer's needs
- Need more advanced and accurate recommendation solutions

# RFM SEGMENTATION OFFERS A MORE NUANCED, WHILE STILL INTUITIVE, APPROACH TO RECOMMENDATIONS



## RFM FITS YOUR NEEDS BETTER THAN K-MEANS CLUSTERING

- RFM is designed specifically for segmentation based on transactional behavior
- RFM is simpler and more interpretable
- RFM requires fewer assumptions and is less affected by outliers



# RFM SYSTEM SEGMENTS CUSTOMERS BASED ON THEIR STANDARDIZED R-F-M SCORES

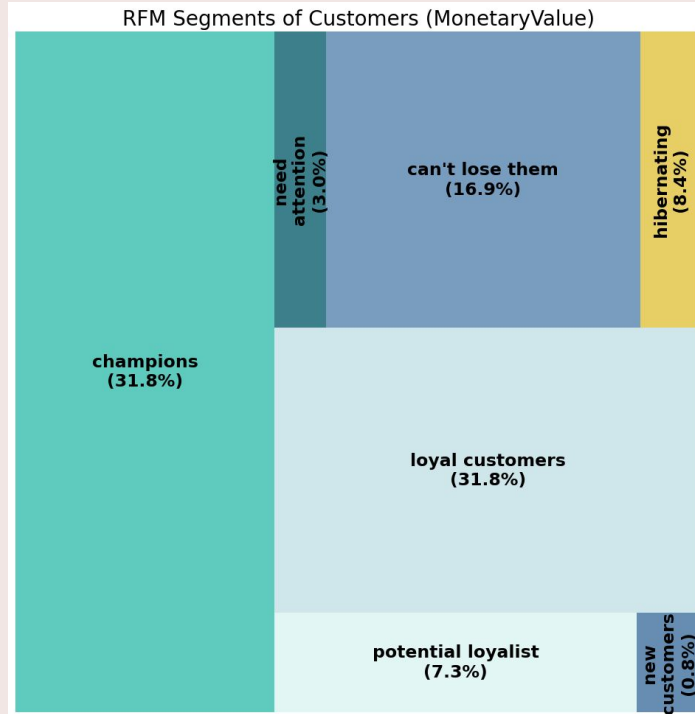
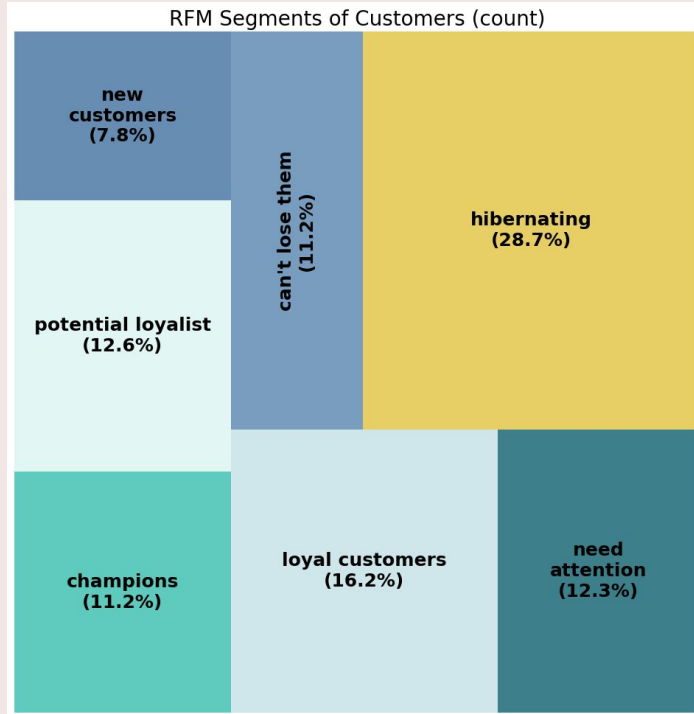
## SEGMENTATION PROCESS

- Each customer's recency, frequency, and monetary values are standardized to a 1-5 scale using quantile-based discretization
- These scores are used to create 7 distinct (MECE) customer groups

Segment	R, F, M Scores	Description
Champions	(5, 4/5, 4/5)	Buy often, spend the most
Loyal Customers	(3/4, 4/5, 4/5)	Buy a lot, pretty active
Potential Loyalist	(4/5, 2/3, 2/3)	Recent, multiple purchases, decent spend
New Customers	(4/5, 1, 1)	Most recent, but not big purchase
Need Attention	(3, 1/3, 1/3)	Need a nudge to keep as customer
Can't Lose Them	(1/2, 4/5, 4/5)	Used to be a champion, but haven't bought recently
Hibernating	(1/2, 1/3, 1/3)	Haven't bought recently, never spent that much



# HIBERNATING CUSTOMERS ARE A LARGE PORTION OF THE CUSTOMER BASE, BUT SMALL PORTION OF REVENUE

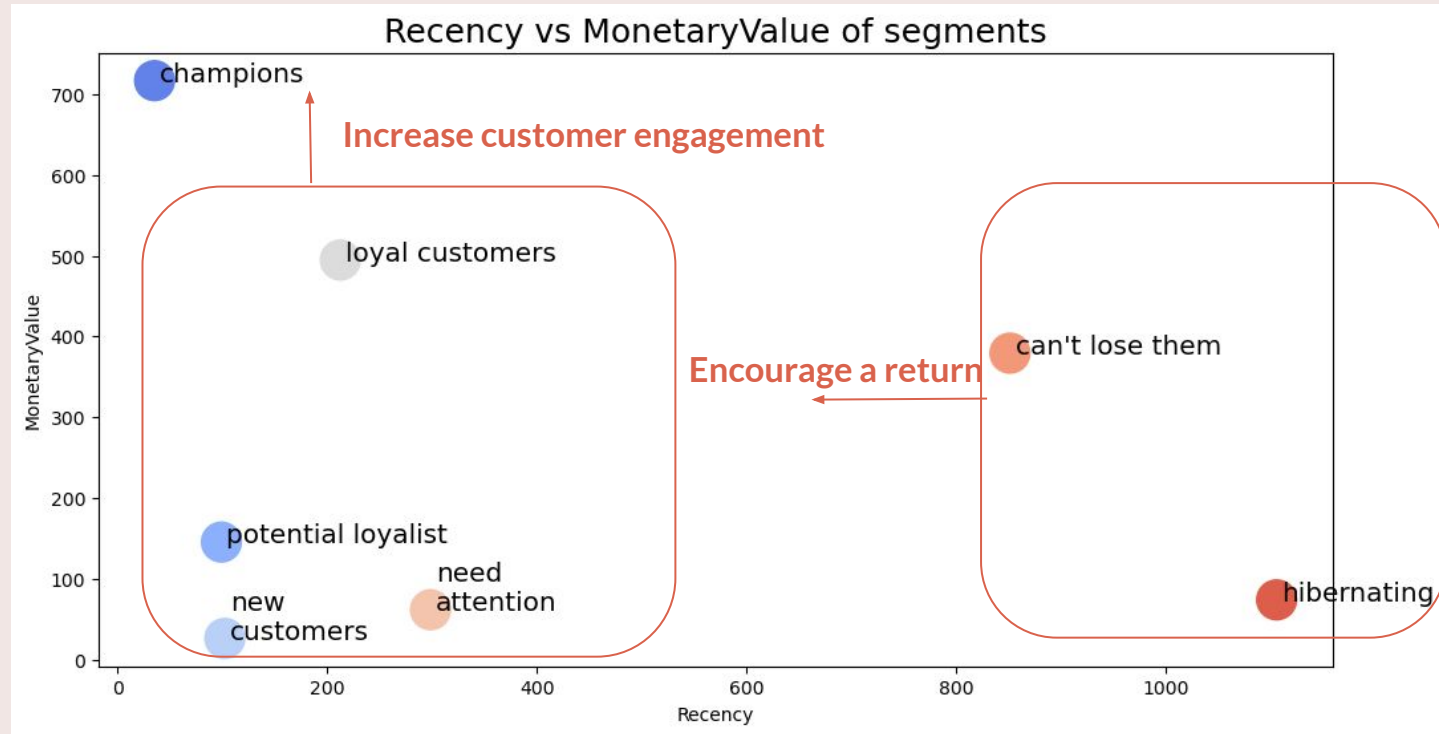


- Over 50% of your customers are at risk of being lost. Action needs to be taken
- Champions and loyal customers make up 23.8% of customers, but 63.6% of revenue
- “20% of customers generate 80% of revenue” is not quite true, but close





# SEGMENTATION HELPS GUIDE STRATEGY TO INCREASE SALES



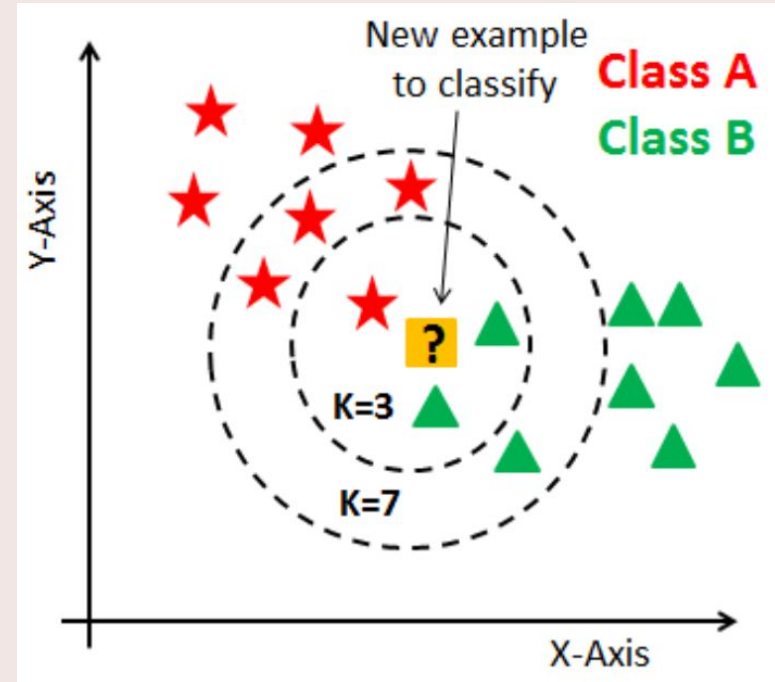
# EACH SEGMENT IS PAIRED WITH A SPECIFIC OFFER

Segment	Offer/Discount	Motivation
Champions	Early access to new releases & book signing	Make them feel exclusive and special and encourage continued high spending
Loyal Customers	10% off orders of \$50-\$100 & 20% off orders of \$100 or more	Tiered discounts to encourage customers to reach higher spending levels for better discounts
Potential Loyalist	“Punch Card” buy 9 books and get the 10th free	Incentivize repeat purchases and build customer loyalty by rewarding customers for consistently choosing your site
New Customers	10% off first-time purchase	Incentivize initial buying decision to reel them in
Need Attention	Time-sensitive 25% off discount	Create urgency with limited time offer to encourage immediate action
Can't Lose Them	30% off discount	Provide a high-value, exclusive discount to win them back
Hibernating	BOGO ‘WelcomeBack’ discount	Provide a substantial discount to encourage a return



# A K-NEAREST NEIGHBOURS MODEL PROVIDES TAILORED CATEGORY RECOMMENDATIONS

- Our K-Nearest Neighbors (KNN) model computes cosine similarity between all categories and uses these similarities to make recommendations that are similar to the categories preferred by similar customers.
- Example: Our model identifies that people who bought sports category items also tend to buy items from the health category. Now we can recommend items from both sports and health and know that they are more likely to be relevant.
- This method is an improvement because it gives personalized recommendations by looking at how other users with similar tastes bought other items.



# OLD SYSTEM DOES NOT SEGMENT CUSTOMERS & SIMPLY RECOMMENDS “MOST POPULAR” CATEGORIES

ID	Most Frequent Purchase Category	R	F	M	Category Rec.
0	History	4	5	5	History, Music, Health, Learning, Religion
1	Music	5	5	5	History, Music, Health, Learning, Religion
2	Health	1	5	5	History, Music, Health, Learning, Religion
3	Travel Guide	4	2	2	History, Music, Health, Learning, Religion
4	Sports	4	1	1	History, Music, Health, Learning, Religion
5	Cooking	1	2	2	History, Music, Health, Learning, Religion
6	Learning	3	2	2	History, Music, Health, Learning, Religion

- Customers 3, 4, 5 do not get relevant category recommendations
- Customer purchasing habits (RFM) are not utilized



# NEW SYSTEM PROVIDES CUSTOMER SEGMENTATION AND MORE ACCURATE CATEGORY RECOMMENDATIONS

ID	Most Frequent Purchase Cat.	R	F	M	Segment	Discount	Category Rec.
0	History	4	5	5	Loyal Customers	10% off orders of \$50-\$100 & 20% off orders of \$100 or more	History, Cont. History, Travel Guides
1	Music	5	5	5	Champions	Early access to new releases	Music, History, Health
2	Health	1	5	5	Can't Lose Them	30% off discount	Health, Learning, History
3	Travel Guide	4	2	2	Potential Loyalist	Buy 9 books and get the 10th free	Travel guide, History, Health
4	Sports	4	1	1	New Customers	10% off first-time purchase	Sports, Health, Hobby
5	Cooking	1	2	2	Hibernating	BOGO 'WelcomeBack' discount	Cooking, Travel guides, Health
6	Learning	3	2	2	Need Attention	25% off if used within 48 hours	Learning, Health, History



# NEW SYSTEM BOOSTS SALES BY \$1M (+12%)

Sales created by our proposal	\$Amount	Change(%)
\$Increase	1,013,624	12%

Projected Sales	\$Amount
\$Total after our proposal	9,506,749

Champions	2,976,947
Loyal Customers	3,028,564
Potential Loyalist	764,990
New Customers	88,730
Needing Attention	297,707
Can't Lose Them	1,503,827
Hibernating	845,984

Projected  
Sales

||

Current  
Sales

**\$8.5M → \$9.5M**  
**(+\$1M/+12%)**

Assumption:  
Lower recency/Monetary value clusters  
might correspond to a higher potential

× ( Recommendation Effect — Cost )

Current sales	\$Amount	Number of Customers	Net Recommendation Effect(%)	Recency contribution(%)	Monetary value contribution(%)	Cost (Discount from sales)
\$Current Total	8,493,125	33,713				
Champions	2,700,179	3,767	10%	5%	5%	0%
Loyal Customers	2,699,255	5,457	12%	20%	10%	15%
Potential Loyalist	618,173	4,252	24%	10%	25%	10%
New Customers	68,943	2,635	29%	10%	30%	10%
Needing Attention	254,450	4,131	17%	20%	30%	25%
Can't Lose Them	1,437,006	3,788	5%	30%	15%	30%
Hibernating	715,118	9,683	18%	30%	30%	30%

- The calculation is based on the entire period, estimated to be around 7 years, as determined by the maximum value in the recency of the dataset
- Cost: Determined according to the promotion description.
- Contribution: As outlined in the chart on page 12. Note that lower recency/monetary value clusters might correspond to a higher potential



# TIMELINE TO YOUR NEW CHAPTER

INITIATIVE	OBJECTIVE	JAN	FEB	MAR	APR	MAY	JUN
Analysis	Validate, examine, re-check data PRN						
Training	Code, optimize, and refine model						
Testing	Check model performance metrics						
Execution	Create, test, and push user-facing interface						
Deployment	Deploy model to subset of users, then roll out for entire customer base						
Monitoring	Conduct post-model review and that check that monitoring is in place						

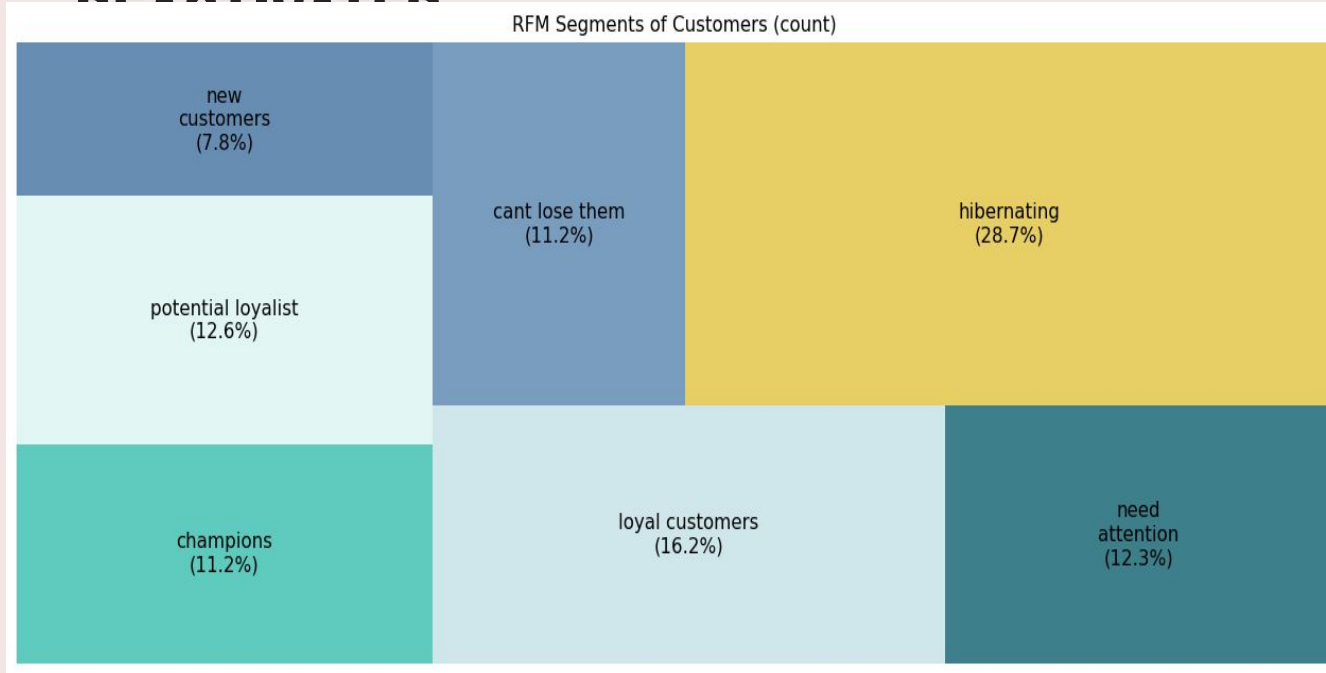
# APPENDIX





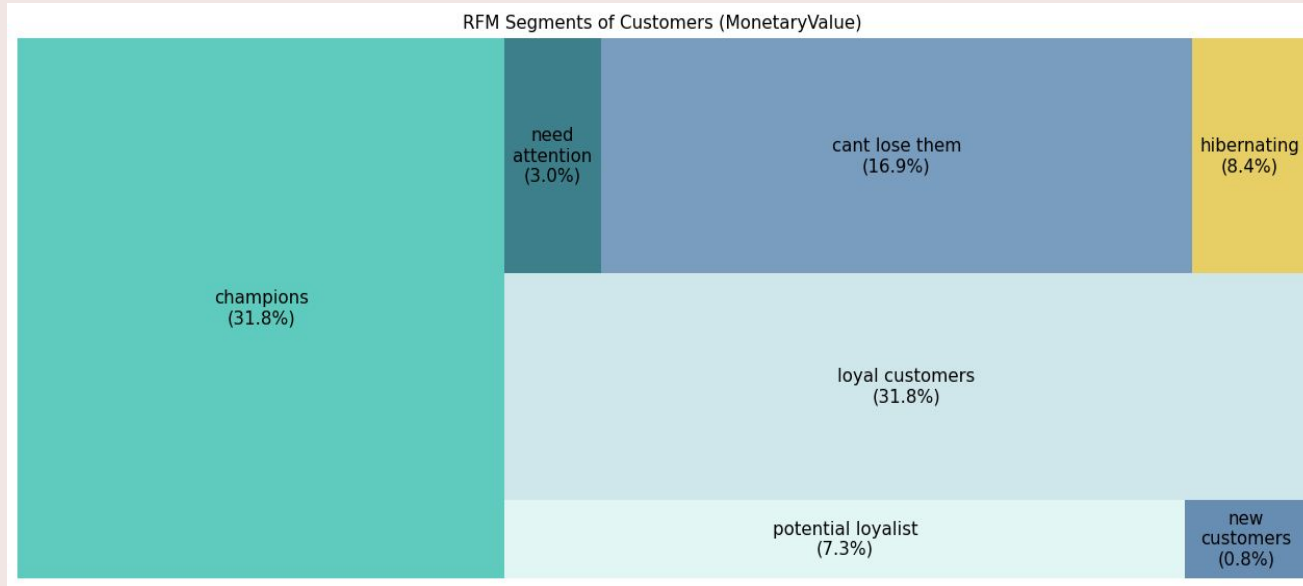
# SEGMENTATION RESULTS SHOW A HIGH NUMBER OF HIBERNATING CUSTOMERS WHO NEED TO BE

REACTIVATED



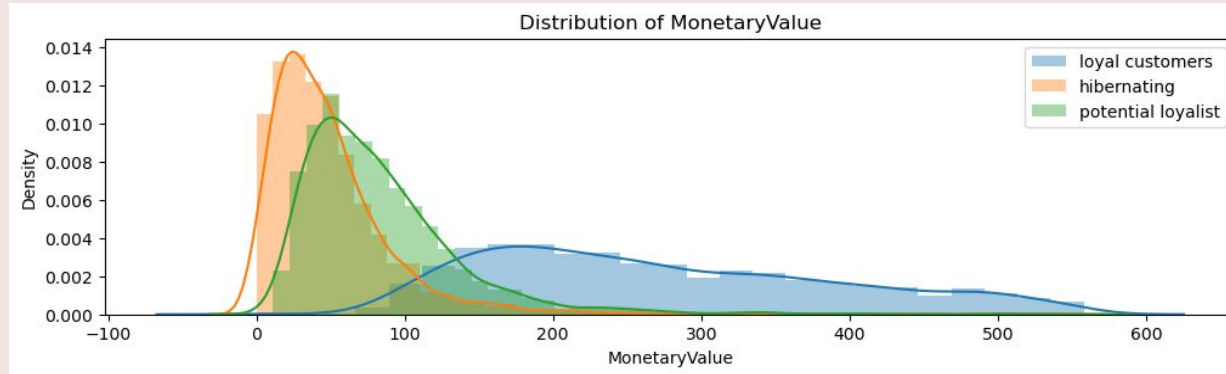
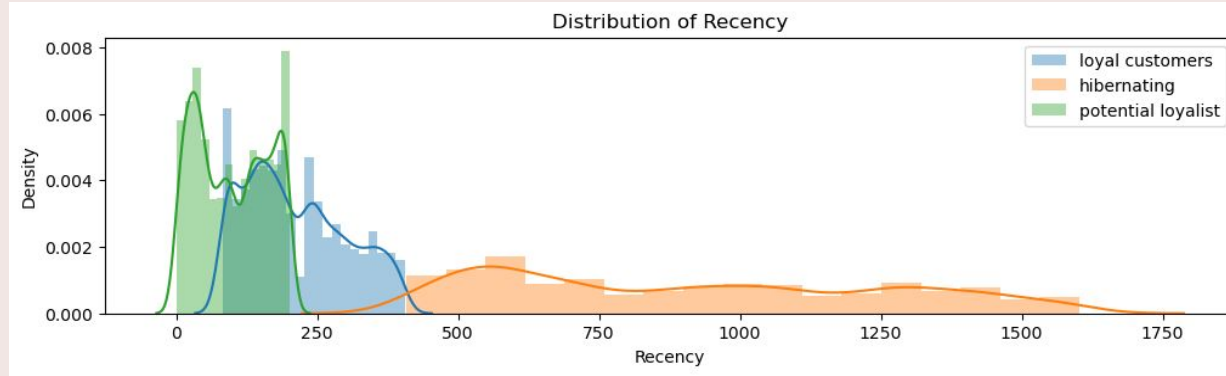
- Hibernating customers make up 28.7% of the customer base
- Adding in “Can’t lose them” and “Needing Attention” segments, over 50% of your customers are at risk of being lost if no action is taken
- Champions and loyal customers are 27.3% of customer base

# HIBERNATING CUSTOMERS ARE A LARGE PORTION OF THE CUSTOMER BASE, BUT SMALL PORTION OF REVENUE



- Hibernating customers make up 8.4% of revenue
- Converting these customers to a more profitable segment should be a high priority
- Champions and loyal customers make up 63.6% of revenue
- “20% of customers generate 80% of revenue” is not quite true, but close

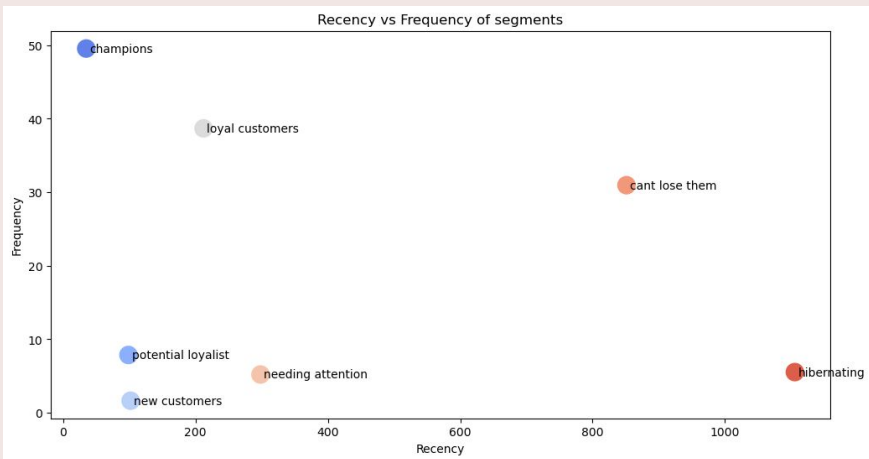
# DISTINCT GROUPS ARE EVIDENT IN RFM DISTRIBUTIONS



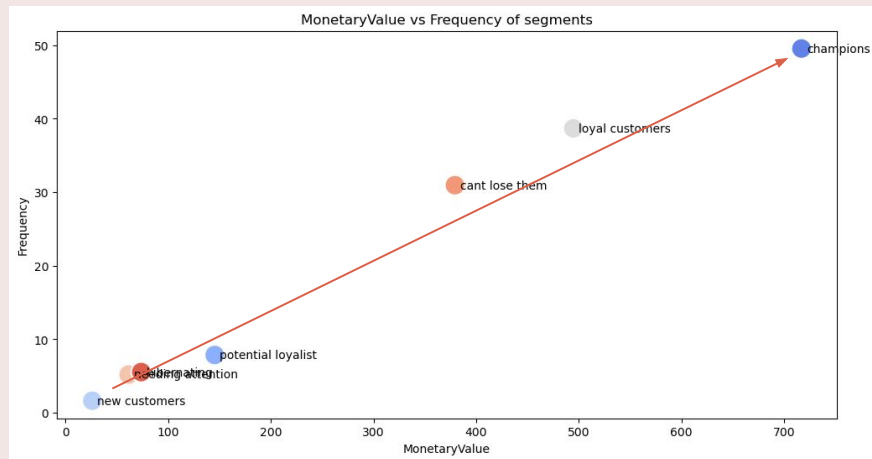
- The three segments with the most customers occupy distinct ranges of the RFM metrics
- Loyal Customers have medium Recency and high Monetary Value
- Hibernating customers have poor Recency and low Monetary Value
- Potential Loyalists have good Recency, but are yet to have spent a lot

# RFM SYSTEM SUCCESSFULLY IDENTIFIES UNIQUE GROUPS

- Almost the same as Recency vs Monetary Value



- Monetary Value and Frequency has a strong correlation



# NEW SYSTEM BOOSTS SALES BY \$2.75M (+32%) (NO COST)

Sales created by our proposal	\$Amount	Change(%)
\$Increase	2,749,232	32%

**\$8.5M → \$11.2M**  
 (+\$2.35M/+32%)

Projected Sales	\$Amount
\$Total after our proposal	11,242,357

Projected  
Sales

||

Current  
Sales

×

Recommendation Effect

Lower recency/monetary value clusters  
might correspond to a higher potential

Current sales	\$Amount	Number of Customers	Net Recommendation Effect(%)	Recency contribution(%)	Monetary value contribution(%)
\$Current Total	8,493,125	33,713			
Champions	2,700,179	3,767	10%	5%	5%
Loyal Customers	2,699,255	5,457	32%	20%	10%
Potential Loyalist	618,173	4,252	38%	10%	25%
New Customers	68,943	2,635	43%	10%	30%
Needing Attention	254,450	4,131	56%	20%	30%
Can't Lose Them	1,437,006	3,788	50%	30%	15%
Hibernating	715,118	9,683	69%	30%	30%

- The calculation is based on the entire period, estimated to be around 7 years, as determined by the maximum value in the recency of the dataset
- Contribution: As outlined in the chart on page 12. Note that lower recency/monetary value clusters might correspond to a higher potential

TEAM

PROBLEM

RFM

DISCOUNTS

COMPARE

IMPACT

# DISCOUNT MESSAGES ARE PERSONALIZED BY SEGMENT

## CHAMPIONS

As a thank you for being a Champion customer, we would like to offer you exclusive early access to the upcoming new release in your favorite category!

## LOYAL CUSTOMERS

As one of our most loyal customers, we would like to extend this exclusive offer of 10% off orders of \$50-\$100 and 20% off orders of \$100 or more!

## POTENTIAL LOYALISTS

As a thank you for being a valued customer, we would like to extend the following exclusive “punch card” offer: When you buy nine books, receive the tenth book for free!

## NEW CUSTOMERS

Welcome to the site! Please accept this introductory offer of 10% off your first purchase.

TEAM

PROBLEM

IMPACT

# DISCOUNT MESSAGES ARE PERSONALIZED BY SEGMENT

**NEED ATTENTION** Act fast! 25% off your next order if made in the next 48 hours!

**CAN'T LOSE THEM**

Thank you for being one of our most valued customers. Please accept this 30% off discount on your next order.

**HIBERNATING**

Welcome back, we've missed you! Pick up reading from your favorite categories and use code 'WelcomeBack' for buy one get one free!"

TEAM

PROBLEM

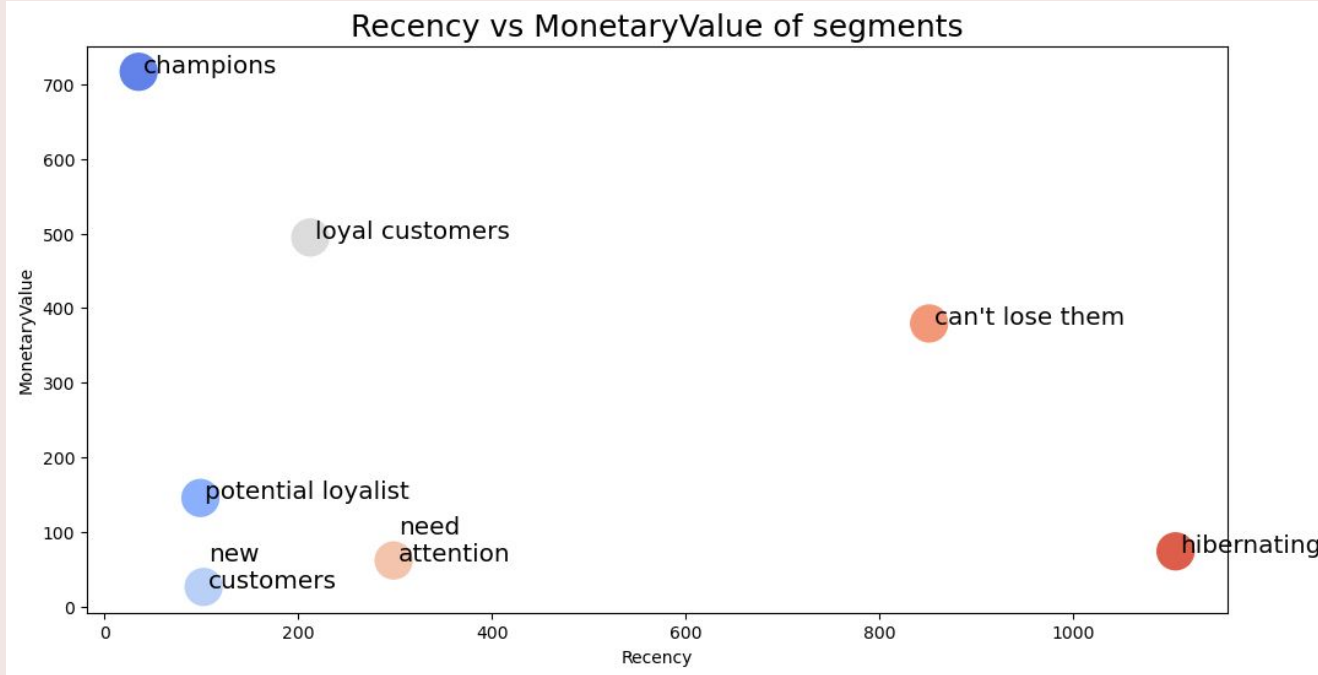
IMPACT

# RESEARCH: EFFECT OF RECOMMENDATION SYSTEMS

Source	Stat
<a href="#">Link</a>	When using recommendation engine, customers had a <b>conversion rate that was 70% higher than normal</b> during that session. Shoppers were <b>20% more likely</b> to return to the site even if they did not complete a purchase.
<a href="#">Link</a>	Amazon attributes <b>35% of its revenue</b> to its recommendation system. <b>~75% of content is discovered</b> on Netflix through its recommendation system.
<a href="#">Link</a>	Product recommendation was responsible for a <b>\$29B increase in sales</b> for Amazon in one year.
<a href="#">Link</a>	These systems add personalization and <b>increase the likelihood of a sale by 75%</b> . Recommendation systems also increase average order value (AOV) and the number of items per order.
<a href="#">Link</a>	Recommendation engines are also <b>good at selling less popular items</b> .



# RFM SYSTEM SUCCESSFULLY IDENTIFIES UNIQUE GROUPS



- This plot shows how the different segments relate along the Recency and Monetary Value dimensions
- Champions purchased very recently and spent a lot, while Hibernating customers purchased a long time ago and spent very little
- Monetary Value and Frequency are highly correlated, so a Recency vs. Frequency plot is redundant