Haley Hoang
Elsie Uwera
Keerthana Katragadda

Customer Analytics for Grocery Retail

by Group 3

Transforming marketing efficiency through data-driven customer insights. Analysis of 2,240 customers across demographics, purchasing behavior, and campaign responses to optimize targeting and profitability.

Project Overview

Transforming marketing efficiency through data-driven customer insights

2,240

Customers Analyzed

Comprehensive dataset across demographics and purchasing behavior

1M

Annual Consumers

Served through physical stores, catalogs, and website

The Challenge

Despite healthy revenue, profit growth is slowing. Many campaigns waste resources due to ineffective targeting in today's competitive grocery market.

Our Objective

Use analytics to target the right customers, increase campaign response rates, and improve profitability through data-driven insights.

Key Performance Metrics

$51K
Median Income

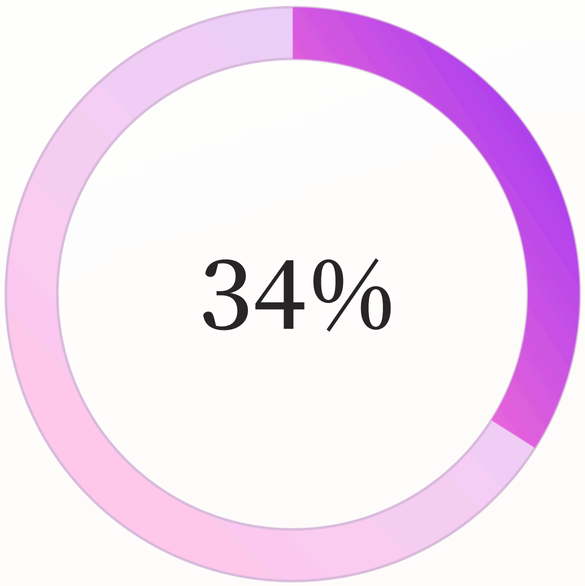
Mid-to-uppersegmentcustomerbasewith significant purchasing power

$562
Average Spend

Total spending per customer across all product categories

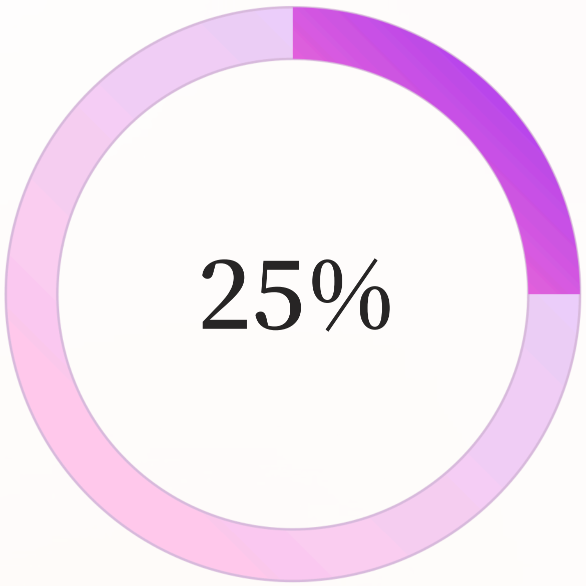
14.9%
Response Rate

Baseline campaign acceptance rate across all customer segments



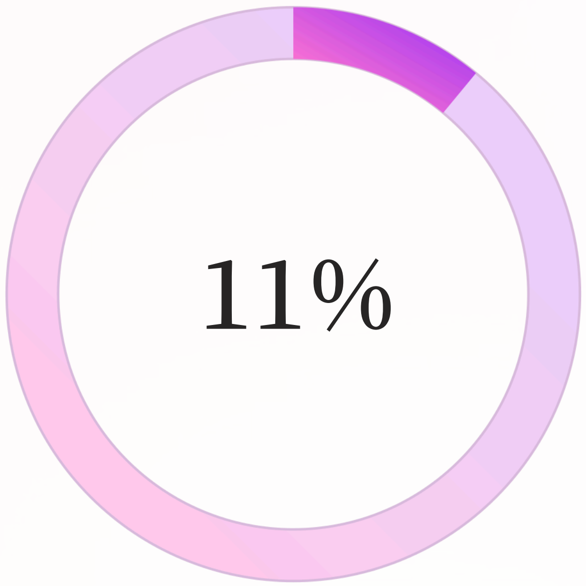
Catalog Response

Highest response rate despite being only 6% of customer base



Web Response

Strong digital engagement with 6.8 visits per month average



Store Response

Largest segment at 79% but lowest campaign response rate

Key Insights from EDA

Insight	Marketing Implication
Most customers shop in stores (around 80%) but spend the least	Focus on in-store upselling — use small checkout coupons or product-pair deals to increase basket size without heavy discounts.
Catalog shoppers are few but very valuable	They spend 2–3× more and respond to campaigns the most (34% response rate). Keep them loyal with exclusive offers, premium bundles, and thank-you rewards.
Families with fewer kids respond more	Target child-free or older households
Store channel still dominant	Strengthening in-store promotions

High-value customers who already buy premium items are significantly more responsive to marketing campaign

Strategic Recommendations

Prioritize Catalog Customers

Target with premium loyalty offers and personalized reordering campaigns. Highest spend and 34% response rate.

Convert Web Browsers

Use personalized recommendations and cart-reminder emails. Apply remarketing ads to frequent visitors.

Boost In-Store Spending

Use in-aisle promotions and checkout coupons. Promote cross- category deals to increase basket size.

Segment by Family Size

Smaller households more responsive. Target singles/couples with premium quality, families with value bundles.

Market Segmentation

Variables Selection

These five variables capture the key behavioral and demographic drivers identified in Phase 1. Collectively, they represent customer value, recent purchasing activity, household structure, preferred shopping channels, and online engagement—all of which are highly relevant to predicting likelihood of response in grocery marketing campaigns.

Total Spend

Captures customer value and purchase engagement

Recency

Reflects how recently customer interacted with store

Children

Represents household size and consumption needs

Income

Indicates financial capacity

Web Visits

Measures online engagement and purchase intent

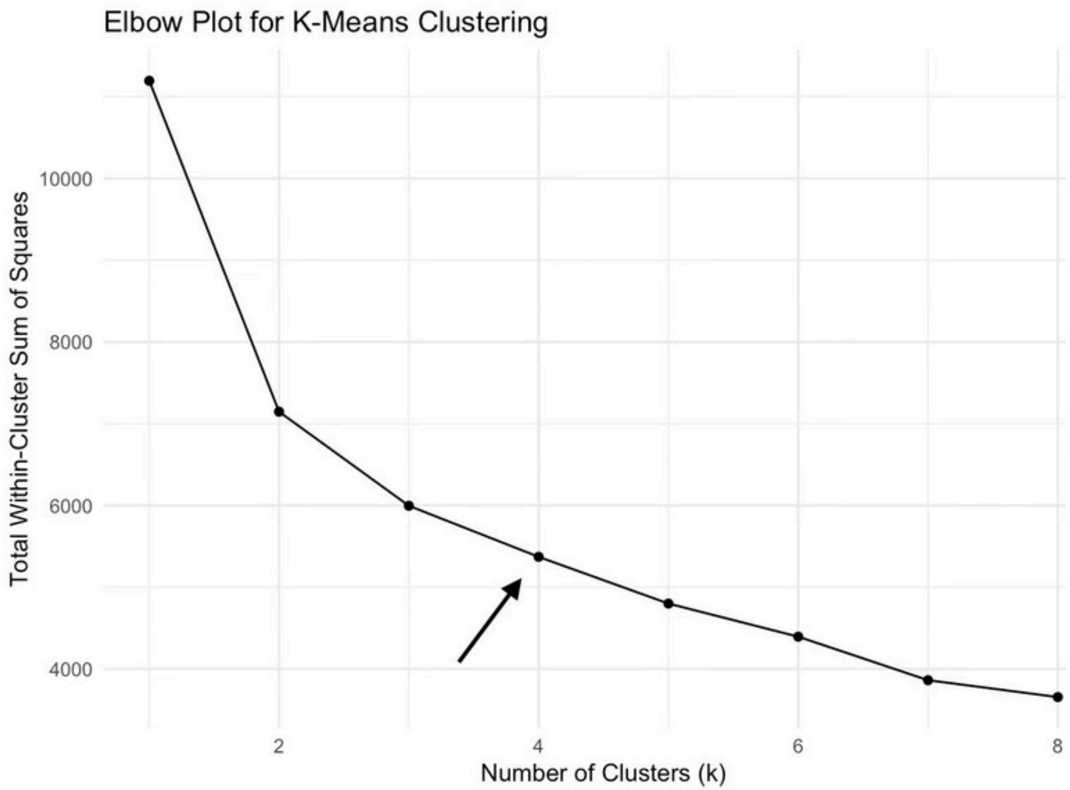
Determining Optimal Clusters

Elbow Method Analysis

The elbow curve revealed diminishing returns after approximately 4 clusters. This balance provided sufficient differentiation between customer behaviors without unnecessary complexity.

Selected: k = 4 clusters This configuration delivers actionable

insights while maintaining interpretability for marketing strategy.



Overview of Customer Segments

Segment	Income	Spend	Recency	Web Visits	Response Rate	Label
1	High	Very High	Moderate	Very Low	30%	Affluent traditional shoppers
2	Mid	Mid	Moderate	Mid	13%	Mixed channel spenders
3	Low	Very Low	Best	High	17%	Digital but low spend
4	Low	Low	Worst	High	5%	Browsers but disengaged

Segment 1:
High-Income, High-Spend Customers

Financial Profile

- **Average income: $80,000**
- **Total spend: $1,300+** Highest economic value of all segments

Shopping Behavior

- Moderaterecency4regular shoppers
- Prefer in-store experiences
- Strong campaign responsiveness

Response Rate

- **~30%**4 highestacross all segments
- Affluent, loyal customers driving premium revenue

Segment 2: Mid-Income, Moderate-Spend Customers

Economic Profile

- Average income: **$65,000**
- Total spend: **$880** (second highest)
- Valuable but not premium-focused

Engagement Patterns

- Recency:~50 days
- Web visits: **5 per month**
- Split between digital and in-store

Marketing Potential

Response rate:**~13%** (moderate)

Segment 3: Low-Income, Low-Spend but Recently Active

$37K

Average Income

Lower financial capacity

$150

Average Spend

Currently low spenders

22

Days Recency

Best across all segments

7

Monthly Web Visits

Highly digitally engaged

17%

Response Rate

Surprisinglystrong

Segment 4: High Web Engagement but High Risk of Churn

Low Conversion

Income: ~$38,000

Spend: $160 average

Largest segment size

High Browsing

6+ monthly web visits

Highest digital engagement

Frequent site interaction

Poor Recency

~70days since purchase

Haven't bought recently

Drifting from brand

Low Response

Only**~5%** responserate

Browsing but not buying

Price sensitive or disengaged

Logistic Regression

Building a predictive model to identify which customers are most likely to respond to future campaigns, allowing our grocery store to target high-probability, high-value shoppers more efficiently and improve marketing ROI

Coefficient & Odds Ratios Table

The coefficient & odds ratios results based on previous variable selection:

```
Call:
glm(formula = Response ~ total_spend + Recency + children + channel_pref +
    NumWebVisitsMonth, family = binomial, data = train)

Coefficients:
                   Estimate Std. Error z value Pr(>|z|)
(Intercept)       -1.6627214  0.3841180  -4.329 1.50e-05 ***
total_spend        0.0014556  0.0001631   8.924  < 2e-16 ***
Recency           -0.0262455  0.0029732  -8.827  < 2e-16 ***
children          -0.4802645  0.1397763  -3.436 0.000590 ***
channel_prefStore -0.8742348  0.2629025  -3.325 0.000883 ***
channel_prefWeb   -0.0907907  0.3092132  -0.294 0.769049
NumWebVisitsMonth  0.2056474  0.0374655   5.489 4.04e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1349.0  on 1567  degrees of freedom
Residual deviance: 1072.3  on 1561  degrees of freedom
AIC: 1086.3

Number of Fisher Scoring iterations: 5
```

```
      (Intercept)       total_spend           Recency
        0.1896222         1.0014567         0.9740960
         children channel_prefStore   channel_prefWeb
        0.6186198         0.4171811         0.9132088
NumWebVisitsMonth
        1.2283200
```

Coefficients Interpretation

Total Spend: +0.001456

Every additional $1,000 spent increases odds of accepting offer by ~15%. **High-value customers are the #1 driver of success.**

Web Visits: +0.2056

Strongest predictor: each extra monthly visit boosts response odds by 22.8%. **Frequent browsers are extremely valuable.**

Recency: -0.0262

Every additional day of inactivity reduces response odds by ~2.6%. **Recent buyers are hot leads - timing is critical.**

Children: -0.480

Customers with children are ~40% less likely to respond. **Family obligations significantly suppress campaign uptake.**

Store Channel: -0.874

Pure in-store shoppers are 58% less likely to accept offer. **They rarely respond to direct campaigns.**

Web channel: -0.091

Slightly lower than Catalog shoppers (baseline), but still much better than Store-only customers.

Recommended Target Segment: High-Value Digital Engagers

Ideal customer profile for maximum ROI:

High Historical Spend

High historical spend (top 30–40% of spenders)

Recent Activity

Last purchase within 3-6 months

HighWeb Engagement

Frequent website visitors (6+ visits/month)

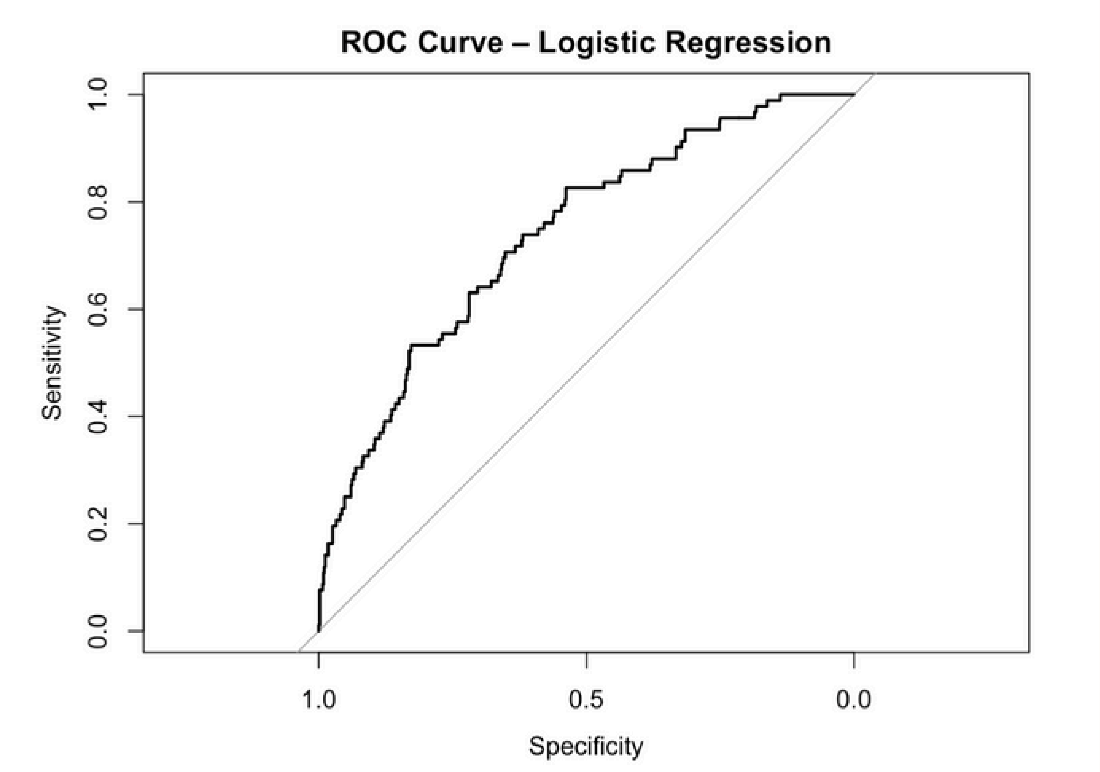
Few or No Children

Households without family obligations

Web/Catalog Preference

Pure in-store shoppers with no web/catalog activity

Is this model better than random selection?



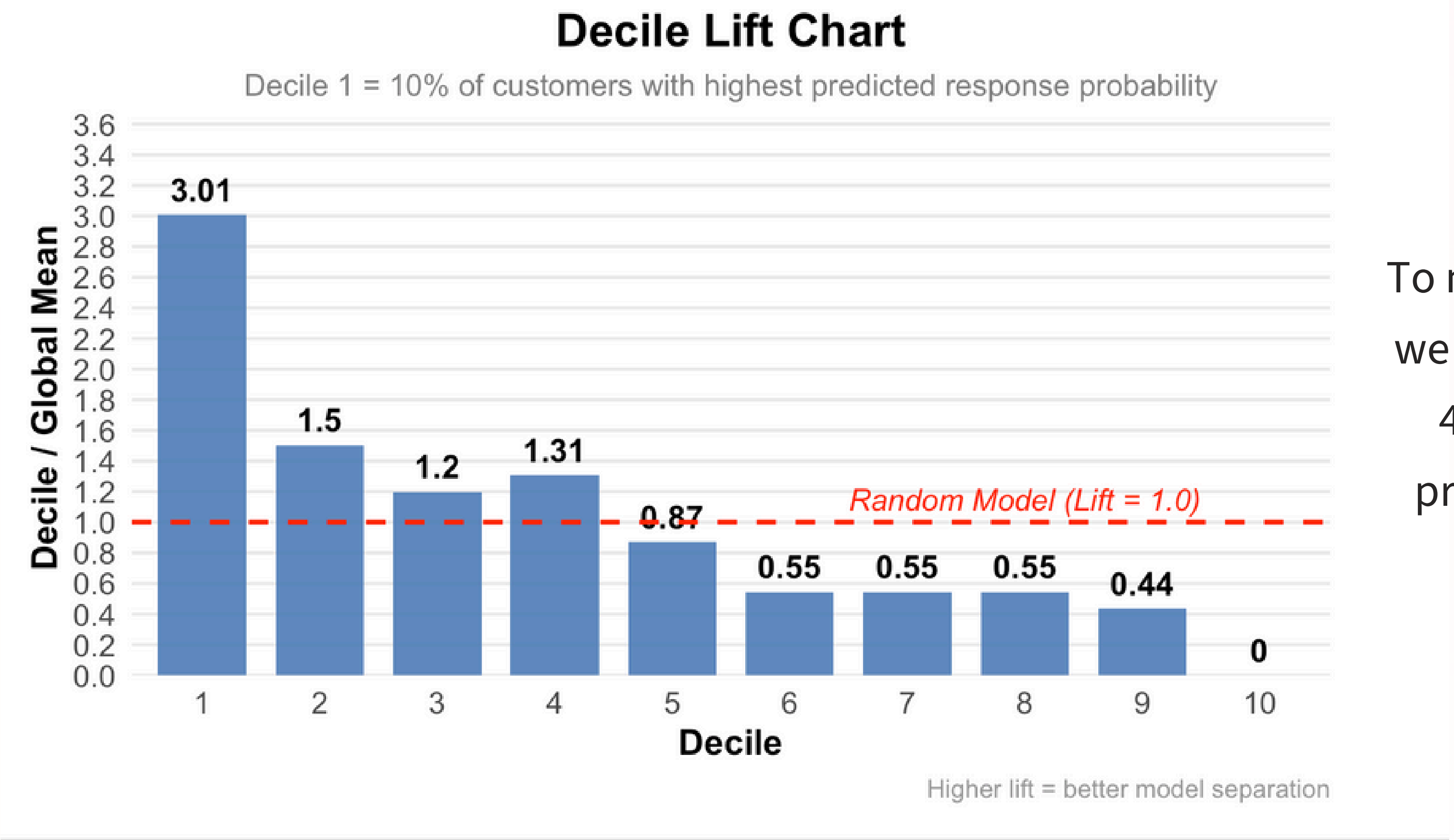
0.73

AUC Score

Model works well, reasonably separates responders from non-responders

The ROC AUC of 0.73 is threshold-independent, evaluating the model's overall ranking ability across all possible cutoffs.

Decile Lift Analysis



To maximize campaign efficiency, we recommend targeting the top 40% of customers ranked by predicted response probability (Deciles 1–4).

Choose Best Cutoff

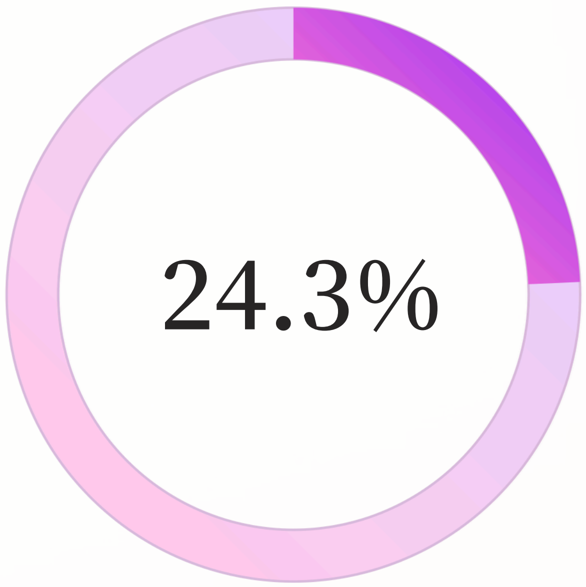
The optimal cutoff identifies customers most likely to respond while minimizing marketing waste, maximizing ROI, and aligning with business constraints.

target_group <chr>	Customers <int>	Responders <int>	Response_Rate <chr>	%_of_Total_Customers <dbl>	%_of_Total_Responders <dbl>
Not Targeted	405	27	6.67%	60.3	29.3
Targeted (Top 40%)	267	65	24.34%	39.7	70.7

```
=== RECOMMENDED CAMPAIGN CUTOFF FOR TOP 40% ===
Target: Top 40% of customers (Deciles 1-4)
Number of customers to contact: 268 out of 672 (40.0%)
Probability cutoff (include if pred_prob >= 0.1340)
Customers with pred_prob >= 0.1340 will be targeted
Baseline response rate on test set: 13.69%
```

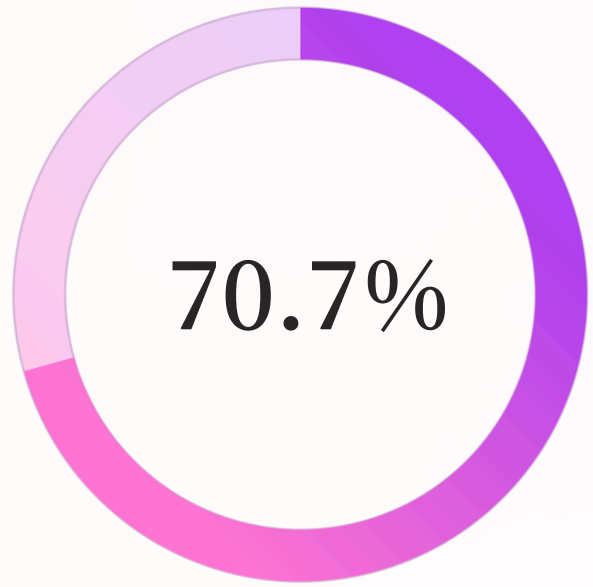
Optimal Targeting Strategy with Optimal Cutoff

Target the top 40% of customers ranked by predicted response probability (New cutoff = 0.1340)



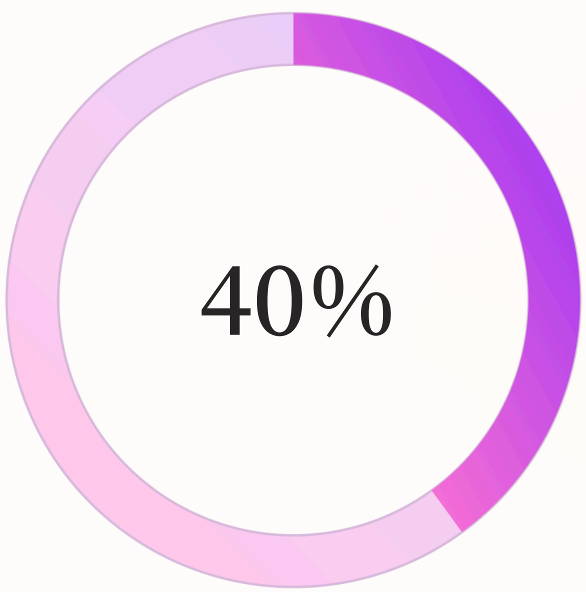
Response Rate

1.78× higher than baseline of 13.69%



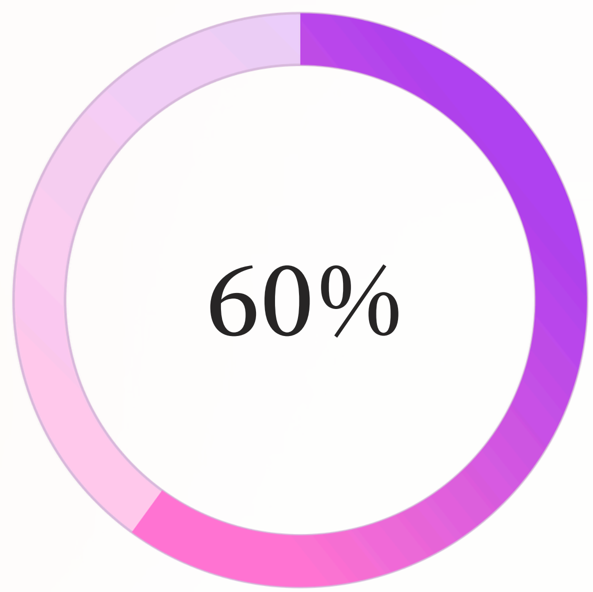
Responder Capture

Of all potential responders reached



Database Contact

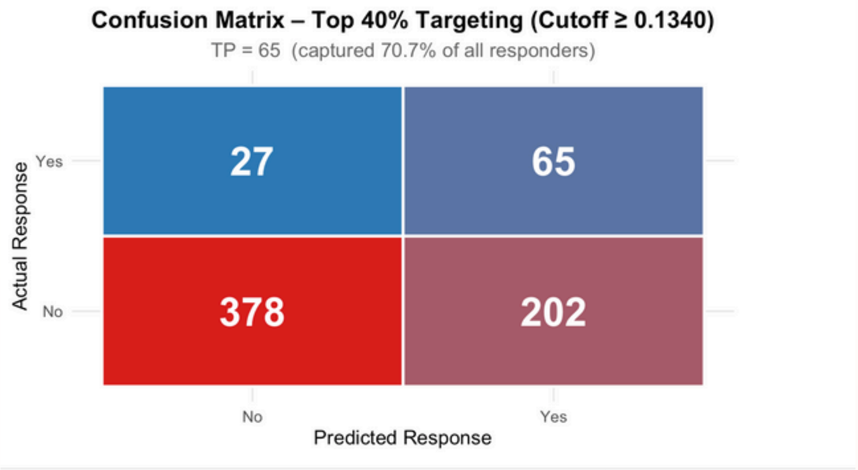
Only 40% of customers contacted



Cost Reduction

Compared to mass campaign

Confusion Matrix with Optimal Cutoff



Metric <chr>	Value <chr>
Accuracy	65.9%
Sensitivity (Recall)	70.7%
Specificity	65.2%
Precision	24.3%
F1-Score	36.2%

5 rows

Accuracy: 65.9%

Model makes correct prediction for two-thirds of customers

Sensitivity: 70.7%

Correctly identify 70.7% of all customers who would have responded to the campaign

Precision: 24.3%

Of the 267 customers we decide to contact, 24.3% actually respond — 1.78× higher than the baseline 13.69%. This is the response rate the marketing team will see in the targeted campaign

The trade-off is acceptable

we accept a moderate precision (24.3%) in exchange for significantly higher reach among true responders — exactly the right balance when the cost of missing a responder is higher than the cost of contacting a few non-responders

Market Segmentation + Predictive Modeling: A Strategic Powerhouse

Combining k-means clustering (4 customer segments) with logistic regression (individual response probability) transforms grocery retail from volume player to high-margin, customer-centric leader—maximizing ROI and driving measurable growth.

Segmentation: The "Who"

Clusters 2,240 customers into 4 actionable groups (e.g., "Affluent Traditional Shoppers" at 30% response vs. "Browsers but Disengaged" at 5%). Enables tailored strategies at scale—premium bundles for high-spenders, value packs for families.

Predictive Modeling: The "How Likely"

Ranks individuals by conversion probability—contact only high-probability leads (recent buyers, 6+ web visits). Cuts waste by 60%, boosts ROI through targeted emails and in-aisle upsells, achieving 70.7% responder capture.

60%
Cost Reduction

Lower CAC by avoiding blanket campaigns and targeting high-probability customers

2-3×
Basket Size Increase

Personalized offers drive higher spending per transaction

30-40%
LTV Boost

Tailored engagement increases customer lifetime value

70.7%
Responder Capture

Precision targeting without budget overextension

> **The Power of Combination:** Segmentation provides the strategic map for planning; predictive modeling delivers the GPS for execution. Together, they enable precision personalization at scale - transforming data into profitable, customer-centric growth.

Business Recommendations

Recommendation 1: Prioritize High-Value Segments

Segmentation analysis reveals stark performance differences. Segment 1 customers spend over $1,200 on average with response rates near 30% - dramatically exceeding baseline performance across all customer groups.

Retention Initiatives:

- Launch VIP loyalty rewards program
- Offer early access to new products and sales
- Create personalized shopping experiences

Recommendation 2: Tailor Channels to Behavior

Digital-First Approach

Segments3 & 4

- Personalized email product recommendations
- Real-time app push notifications for flash sales
- Web-exclusive deals and early-bird promotions
- Dynamic retargeting ads based on browsing history

Store-Centric Strategy

Segments 1 &2

- Point-of-sale triggered coupons at register
- High-quality physical mailers with premium offers
- In-store displays featuring high-margin wine selections
- Personal shopping appointments and consultations

Recommendation 3: Deploy Predictive Scoring

Model Performance: AUC = 0.73

Our logistic regression model demonstrates **strong discriminatory power**, making it reliable for campaign targeting and customer prioritization decisions.

Top Positive Predictors

- **Total spend:** Higher spending signals greater response likelihood
- **Purchase recency:** Recent buyers show elevated engagement
- **Web visits:** Digital activity correlates with conversion

Score All Customers: Generate predicted response probability for entire database using validated model

Create Ranked Lists: Sort customers by probability score to identify highest-value targets first

Recommendation 4: Win Back At-Risk Customers

The Segment 4 Challenge

Despite leading in digital engagement, Segment 4 shows 70-day average recency and sub-5% response rates. These customers are browsing extensively but not converting4indicating price sensitivity or disengagement.

Reactivation Campaign Triggers

- Automatic deployment when recency exceeds 45-60 days
- Personalized "We miss you" messaging with customer name
- Time-limited offers creating urgency (72-hour windows)

Win-Back Incentives

15-20% comeback discount codes Free shipping with no minimum Personalized product reminders

Exclusive early access to sales

Key Takeaways & Next Steps

Concentrate Resources

Direct 60-70%of marketinginvestment toward Segments 1 & 3 for maximum ROI

Match Channels to Behavior

Deploy digitalstrategies forSegments 3 & 4, store-centric tactics for Segments 1 & 2

Leverage Predictive Models

UseAUC0.73modeltorankandprioritize all campaign audiences systematically

Prevent Churn Proactively

Launch automated win-backcampaigns when customer recency exceeds 45 days

---

These four evidence-based recommendations provide a clear roadmap to optimize targeting, increase campaign efficiency, and drive measurable revenue growth through sophisticated customer segmentation and predictive analytics.