Data-Driven Customer Targeting and Segmentation Strategy

JPMorganChase

Yoon Nam | Feb 22, 2026

Executive Summary

Objective — Identify high-income customers and optimize marketing targeting efficiency

Predictive Modeling

- The predictive model achieved strong discrimination power (ROC-AUC 0.94).
- Income probability is strongly driven by age, education, and weeks worked.

Segmentation Analysis

- Two segments (Premium Target & High-Value Mature) generate 4–5x higher lift compared to the overall population.
- One segment (Avoid) shows near-zero conversion probability, indicating significant cost inefficiency.

Expected Effect

- **Reallocating budget toward top-performing segments improves ROI and reduces waste.**
- **Potential for a 3–4x increase in targeting efficiency.**

Data Overview

The dataset contains 199,523 observations with demographic, employment, household, and financial variables.
The target variable is binary: **High Income (>$50K) vs. Low Income (≤$50K).**

Features

Demographic

Education & Household

Employment

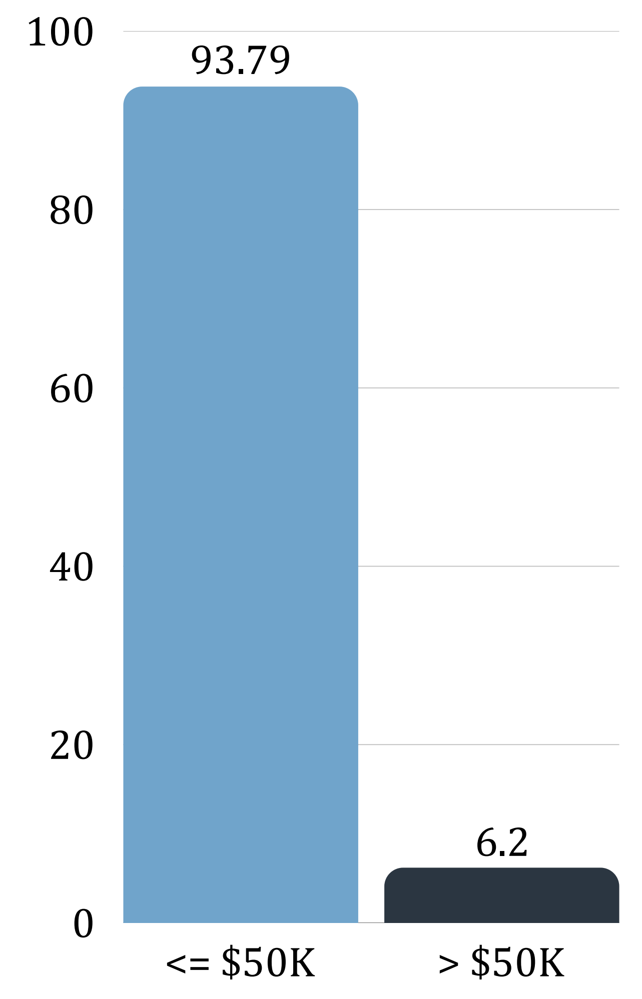
Financial

Geographic & Migration

*** Missing Value Handling :**
'country of birth father', 'country of birth mother' features showed age gaps exceeding 10 years, indicating structured (non-random) missingness. These features were encoded as "Unknown" to preserve signal while minimizing bias.

Severe Class Imbalance

93.8% Low Income, 6.2% High Income.
Evaluation focused on metrics robust to class imbalance rather than accuracy.

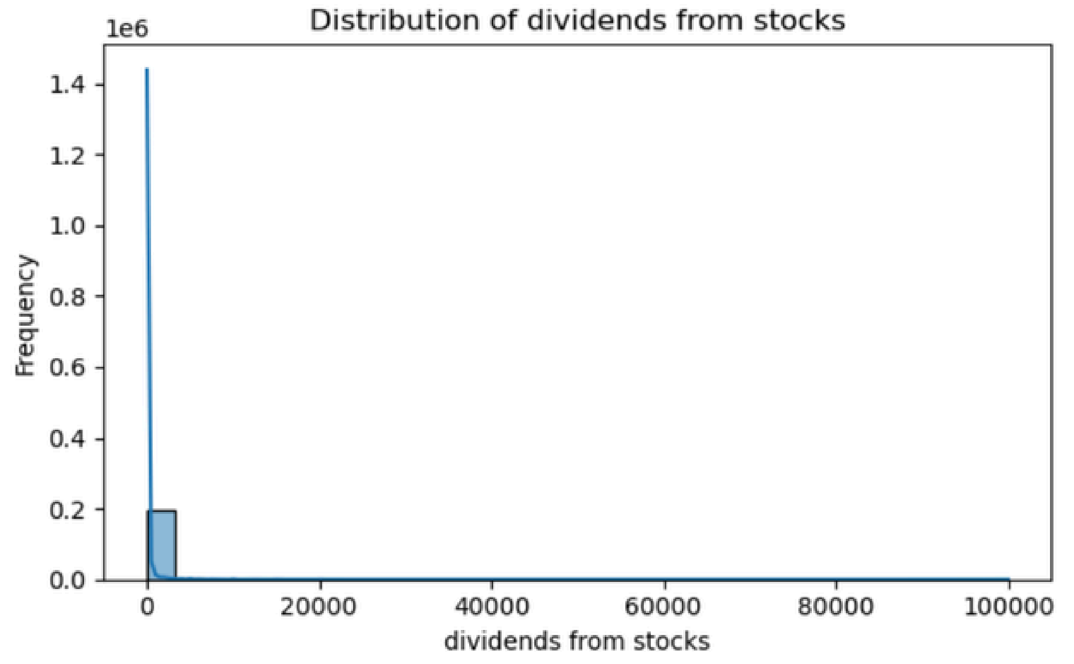


Heavy-tailed Financial Variables

Capital Gains, Dividends from Stocks, and Wage per Hour variables have a long-tail pattern:

Most individuals have zero values
A small subset has extremely high values

To stabilize variance and reduce model bias,
Log transformation was applied prior to modeling.



Predictive Modeling : Optimized Model for Precision Marketing

1 Evaluation Metric Selection

F1 Score

F1 Score balances Precision and Recall.
We prioritize F1-score to balance coverage and cost efficiency.

Recall

“how many actual high-income individuals we correctly identify”

if Recall is high but Precision is low:
Many low-income customers will also be targeted.

Marketing budget may be wasted.

Precision

“how many predicted high-income individuals are actually high-income”

if Precision is high but Recall is low:
We miss many true high-income customers.

Opportunity loss.

2 Model Benchmarking

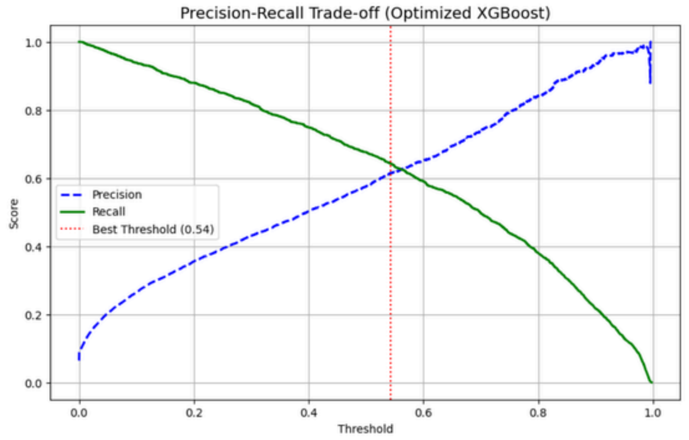
XGBoost achieved the highest F1-score (0.53) while maintaining strong ROC-AUC (0.94).

Random Forest showed high precision but weak recall.
LightGBM achieved strong recall but lower precision.

	Recall	Precision	F1-Score	ROC-AUC
LightGBM	0.877246	0.318861	0.467713	0.950935
Logistic Regression	0.872905	0.244918	0.382512	0.925974
XGBoost	0.783263	0.402471	0.531711	0.942210
Random Forest	0.361296	0.727397	0.482782	0.934624

3 Threshold Tuning

Optimal threshold selected to maximize F1-score.



4 Final Model Selection

XGBoost Final Validation

	Precision	Recall	F1 Score
High Income	0.98	0.97	0.97
Low Income	0.61	0.64	0.63

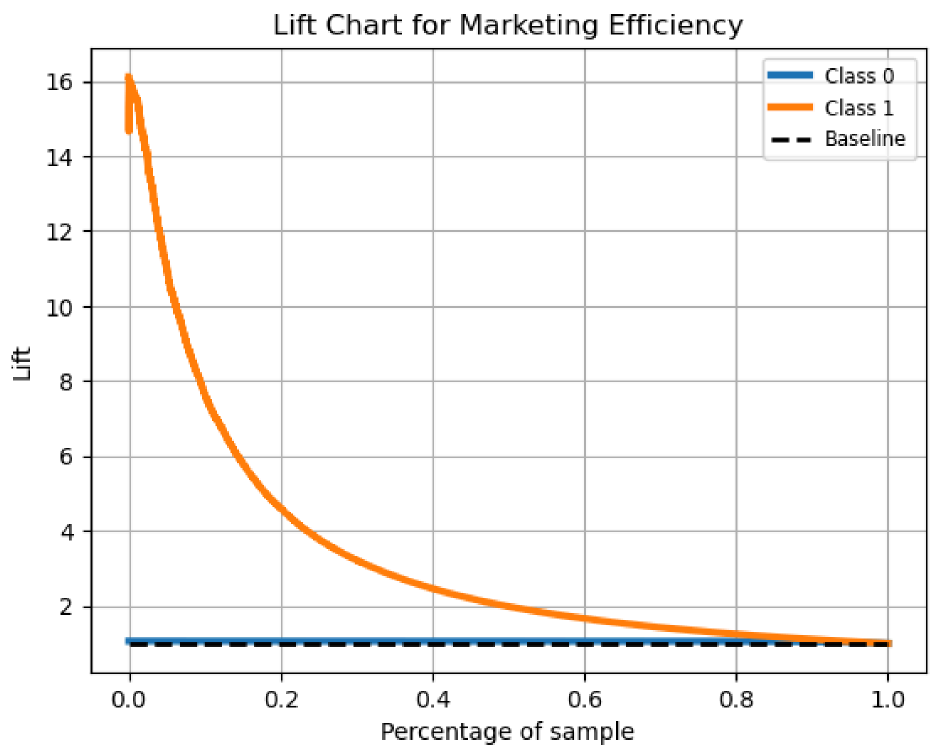
XGBoost captures 97% of actual high-income individuals. This means that among individuals predicted as high-income, 97% are truly high-income.

Given the class imbalance, **a recall of 64%** for high-income individuals is a strong outcome, meaning the **model correctly identifies 64 out of 100 actual high-income individuals.**

A **precision of 61%** indicates that **61% of predicted high-income individuals truly belong to that segment,** making the model suitable for targeted marketing applications.

Lift Performance & Key Drivers

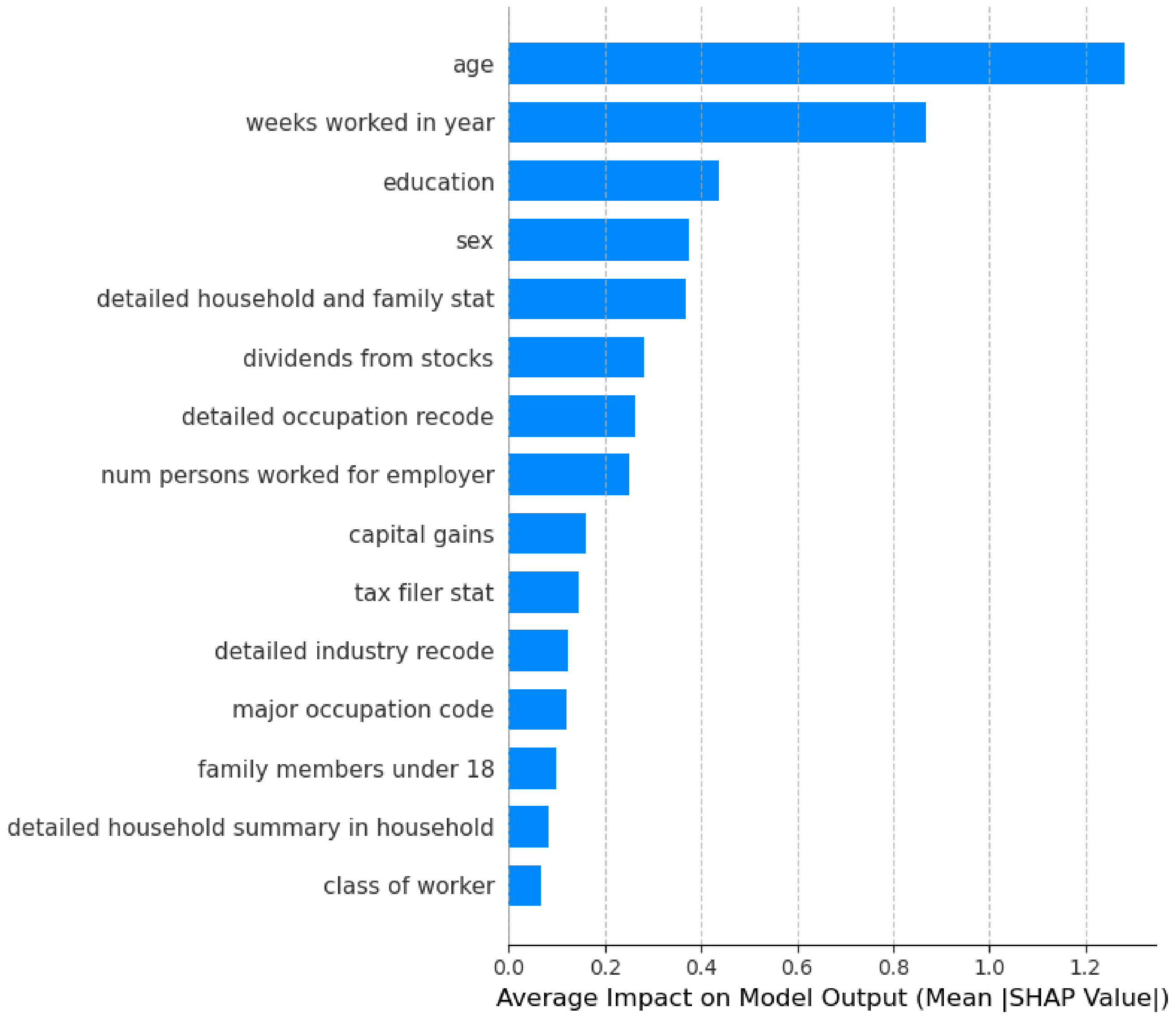
Lift Curve



The Lift chart demonstrates marketing targeting efficiency. At the top 5% of ranked customers, the lift reaches approximately 16x, meaning that **targeting the highest-scoring individuals yields 16 times more high-income individuals** compared to random selection.

Even within the top 20%, lift remains around 4–5x.
This indicates that the model effectively ranks high-income individuals at the top, providing **strong ROI potential for targeted campaigns.**

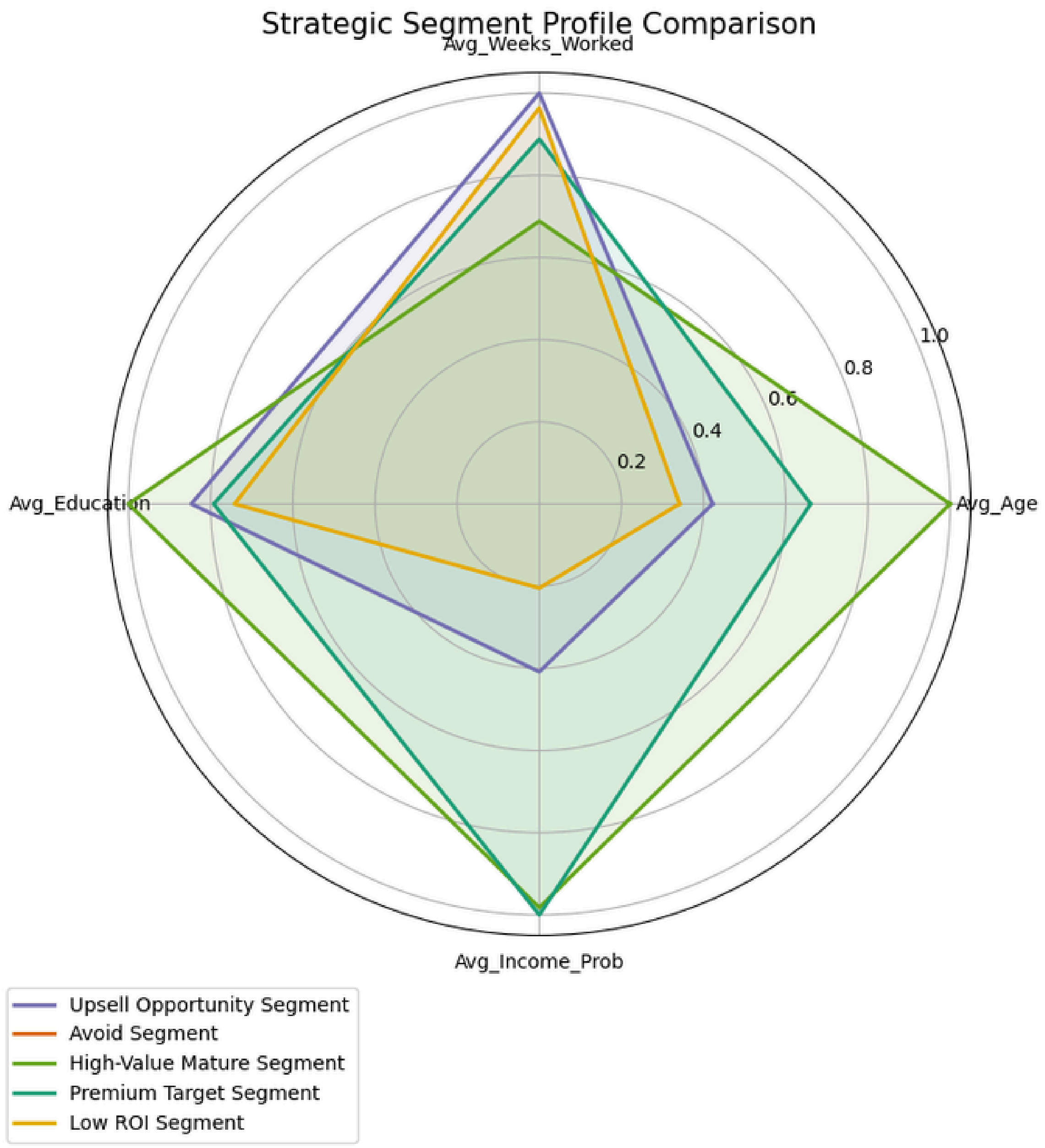
Feature Importance



Age, employment intensity (weeks worked), and education level are the strongest structural predictors of high income.

Strategic Customer Segmentation

- Premium and High-Value segments exhibit 3–5x higher high-income concentration.
- Upsell segment works more weeks but lower realized income
- Avoid segment shows structurally weak employment patterns
- Targeting focus should prioritize Premium and High-Value segments while minimizing spend on Avoid.



Segment	High Income Rate	Age	Weeks Worked	Education
★ **Premium Target Segment**	30.2%	43.78	41.53	10.91
★ **High-Value Mature Segment**	26.2%	51.83	32.46	11.41
Upsell Opportunity Segment	8.9%	38.13	46.64	11.04
Low ROI Segment	4.5%	36.21	44.96	10.79
Avoid Segment	0.3%	28.12	1.23	9.01

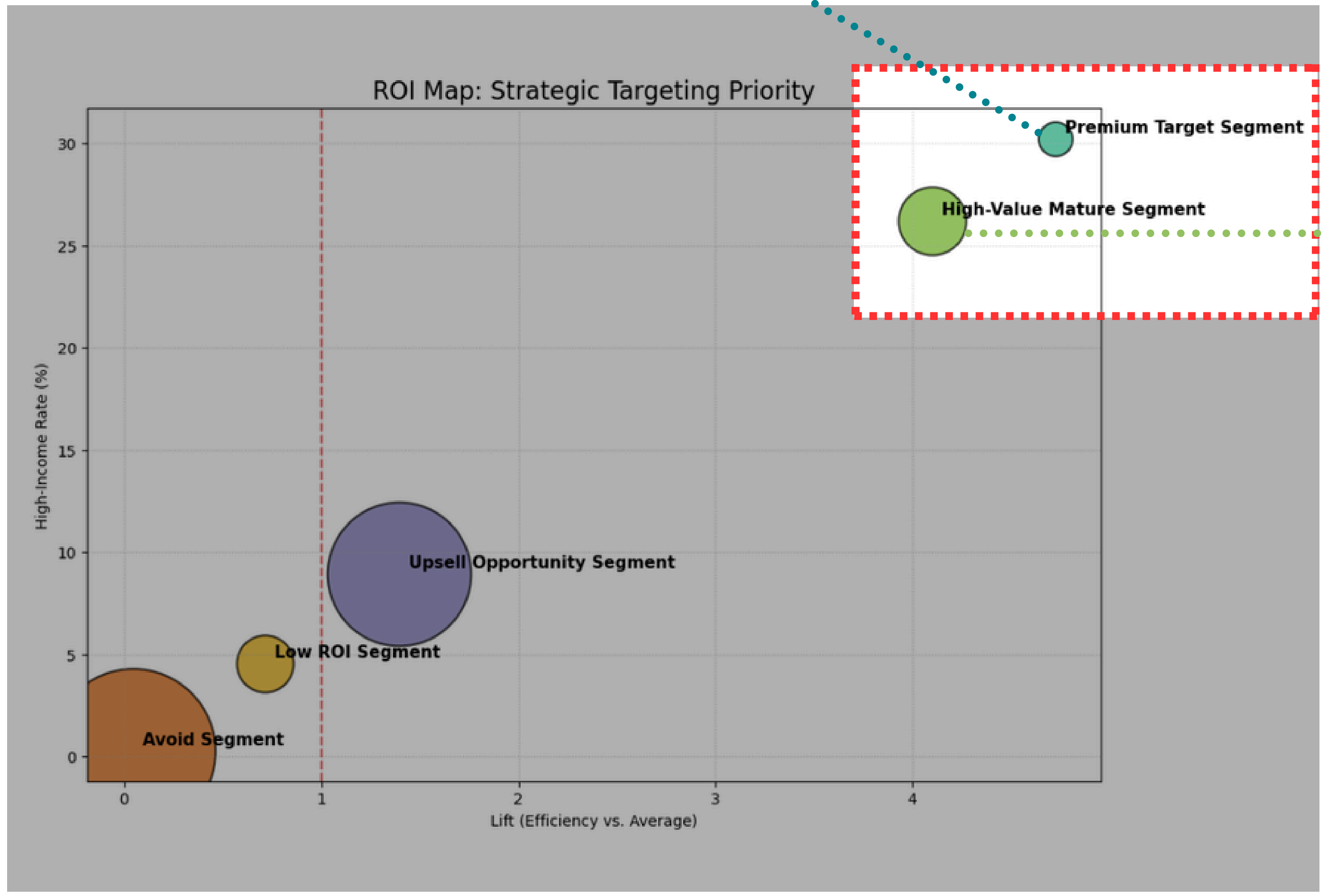
Concentrate Spend Where Income Probability Peaks

Premium Target Segment ★

- High-Income Rate: 30.2% (4.7x population average)
- Profile: Age 44, full-time workers, highly educated

Strategy: Focus on premium product categories and loyalty expansion

- Promote high-margin premium products (luxury apparel, premium electronics, designer goods).
- Offer exclusive membership tiers with early access and VIP discounts.
- Personalize recommendations and subscription-based premium services for high AOV bundles.



High-Value Mature Segment ★

High-Income Rate: 26.2% (4.1x population average)
Profile: Age 52, with shorter working weeks (32.5 weeks) but high capital gains.
Established career professionals, Wealthy Seniors

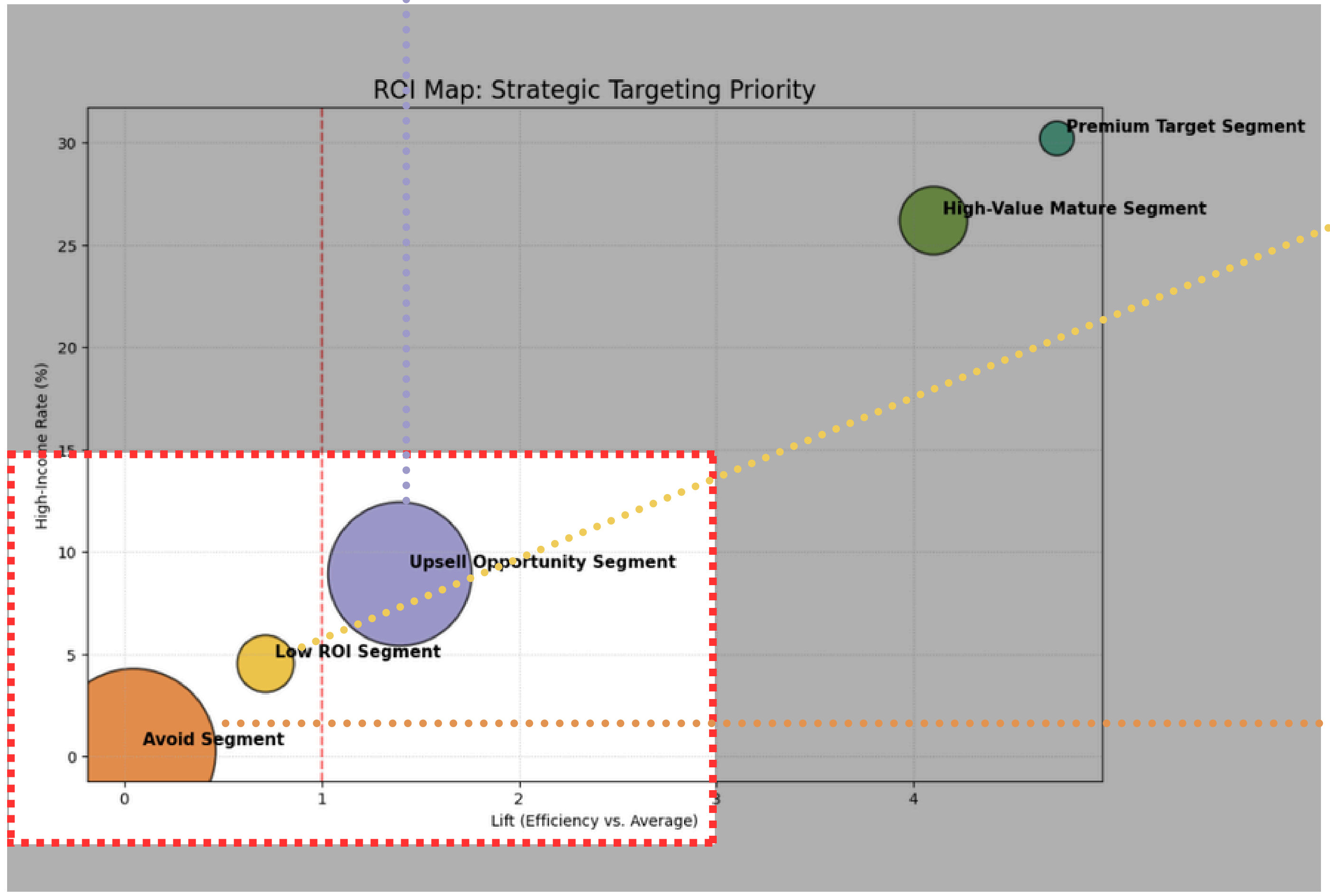
Strategy: Focus on premium lifestyle and repeat purchase programs

- Promote home, wellness, travel, and lifestyle products.
- Offer curated bundles (home improvement, seasonal packages).
- Target with email campaigns emphasizing quality and reliability.
- Implement cross-sell programs for complementary premium products.

Minimize Low-Return Targeting

Upsell Opportunity Segment
High-Income Rate: 8.9% (near population average)
Profile: Age 38, growth potential, career development phase

Strategy:
- Promote installment payment options.
- Offer mid-tier product upgrades.
- Target entry-level premium categories.
- Use targeted promotions with limited discounts.



Low ROI Segment
High-Income Rate: 4.6% (below average)
Age mid-30s, leading to low ROI for aggressive marketing.

Strategy
- Focus on essential, price-sensitive product categories
- Use automated email and app push campaigns
- Offer clearance and promotional inventory

Avoid Segment
High-Income Rate: 0.3% (extremely low)

Strategy:
Minimize marketing spend and avoid high-CAC targeting

Execution Risk & Mitigation Plan

Risk	Mitigation & Next Steps
Underperforming Marketing ROI Actual results below target (+50% ROI improvement)	• Real-time A/B testing for performance validation • Phased rollout to minimize exposure • Threshold optimization for precision improvement
Model Performance Degradation Model accuracy decline over time due to data drift	• Quarterly model retraining schedule • Real-time performance monitoring dashboard • Automated alerts when precision drops below 35%
Regulatory Compliance Financial regulation violations (fairness, privacy)	• Regular bias auditing for gender/race discrimination • GDPR/CCPA compliance framework • Legal team quarterly review process

End of Document