Vanguard's $1,000 Question

Analyzing U.S. Households' Liquid Assets

Dan Valenzuela

Business Background

Financial Services
Competition

Vanguard®
Mutual Funds

NO
FEE

Fewer Fees

How to escape?
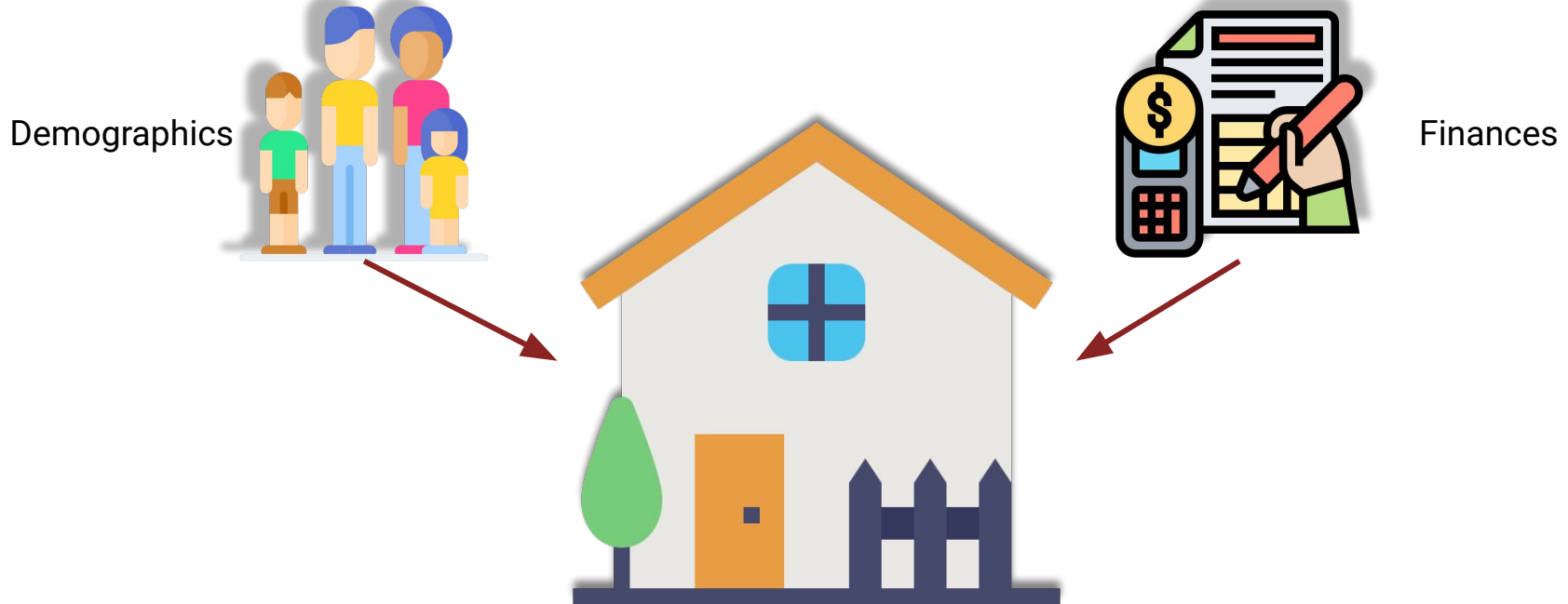
**Better targeting of
potential customers**

Vanguard®
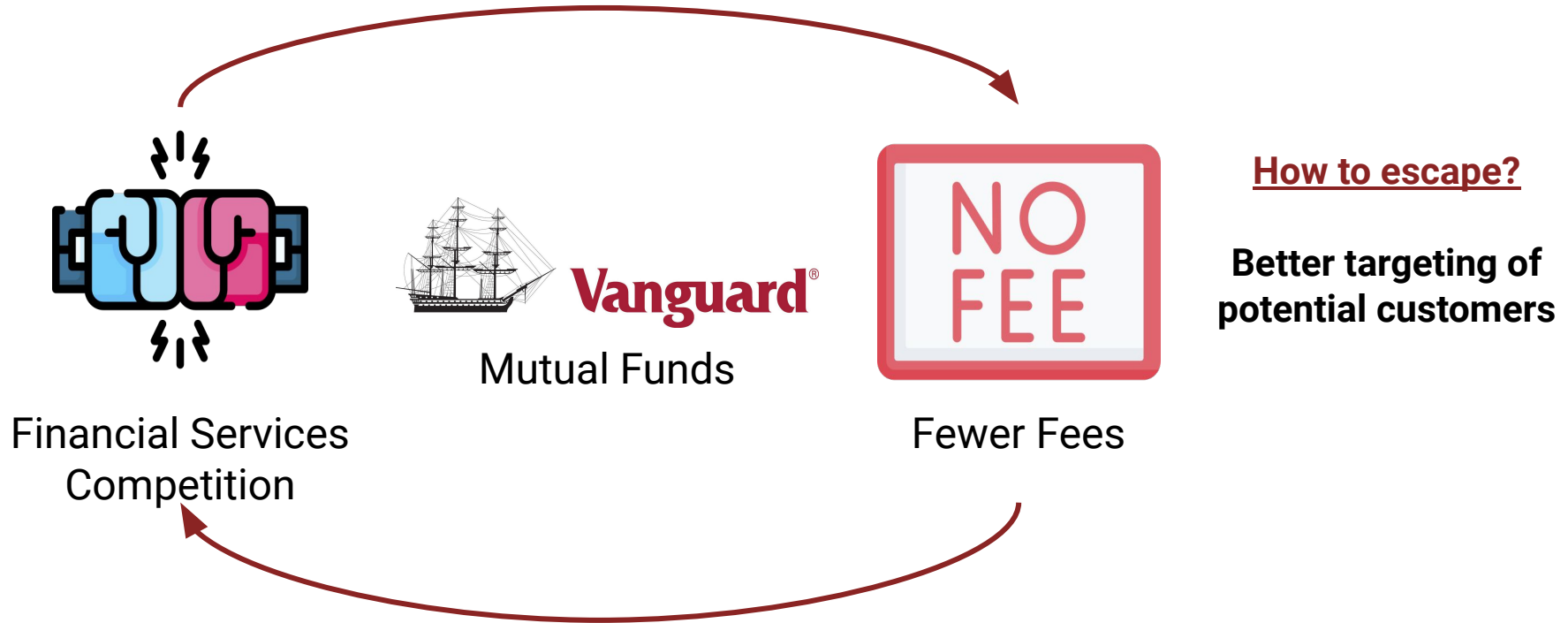
Method

Overall Goal: Target households with >$1,000 in assets

1. Collect data on household finances

2. Construct and retrieve relevant household features

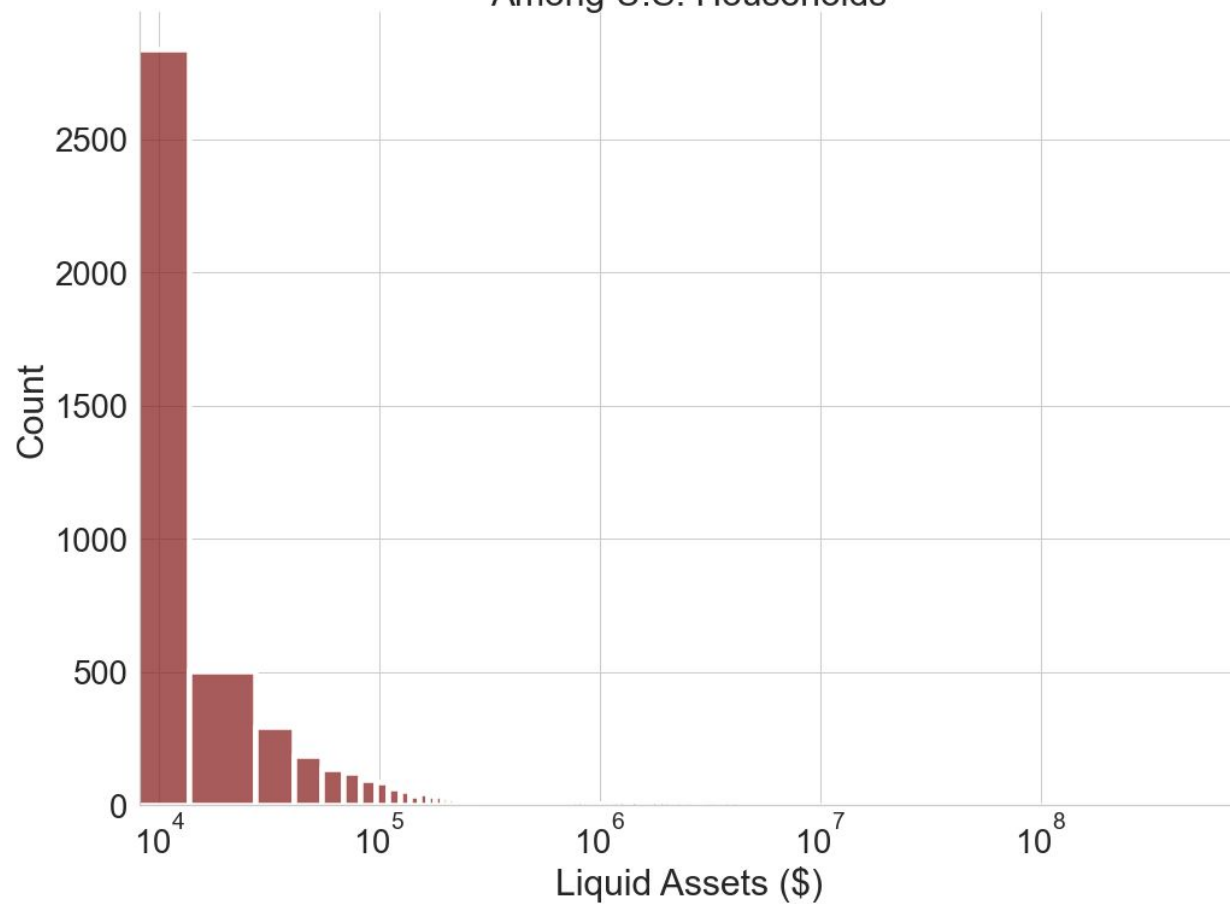3. Classification model to target households and understand features

Vanguard®

Data - Potential Customers

Demographics

Finances

Liquid assets of 6,000 households

Data source: Federal Reserve

Vanguard®

Data Hurdles

1. Skewed distribution of wealth

2. Overrepresentation of wealthy households

Vanguard®

Distribution of Liquid Assets
Among U.S. Households

Cumulative Distribution of Liquid Assets Among U.S. Households (Weighted and Unweighted)

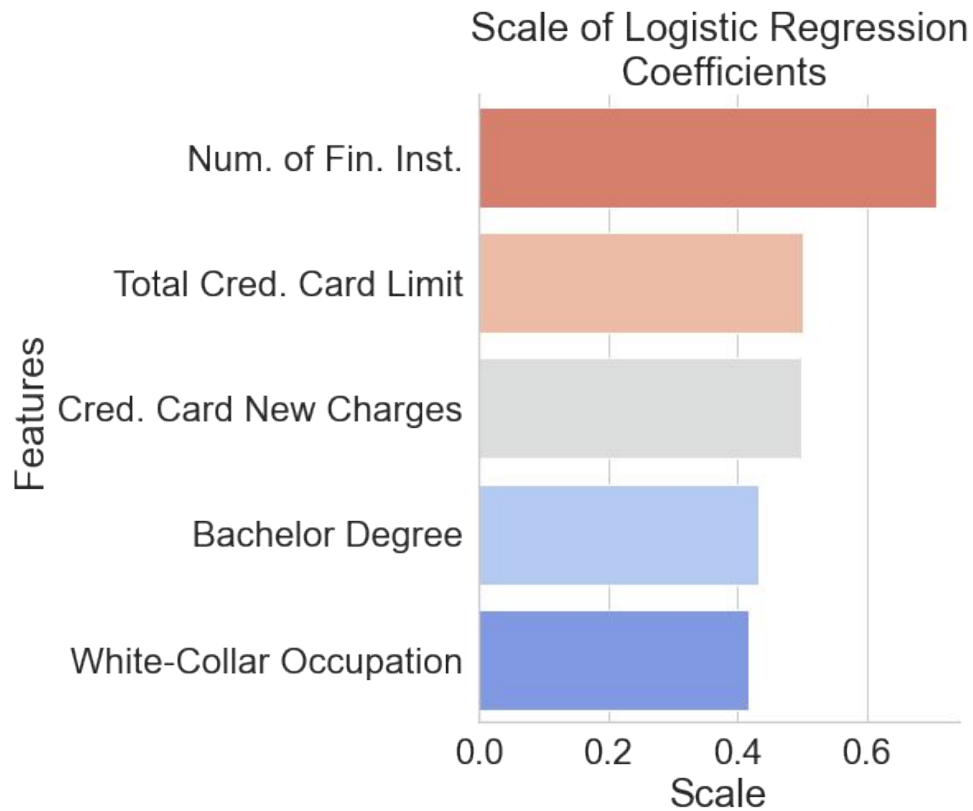Vanguard®
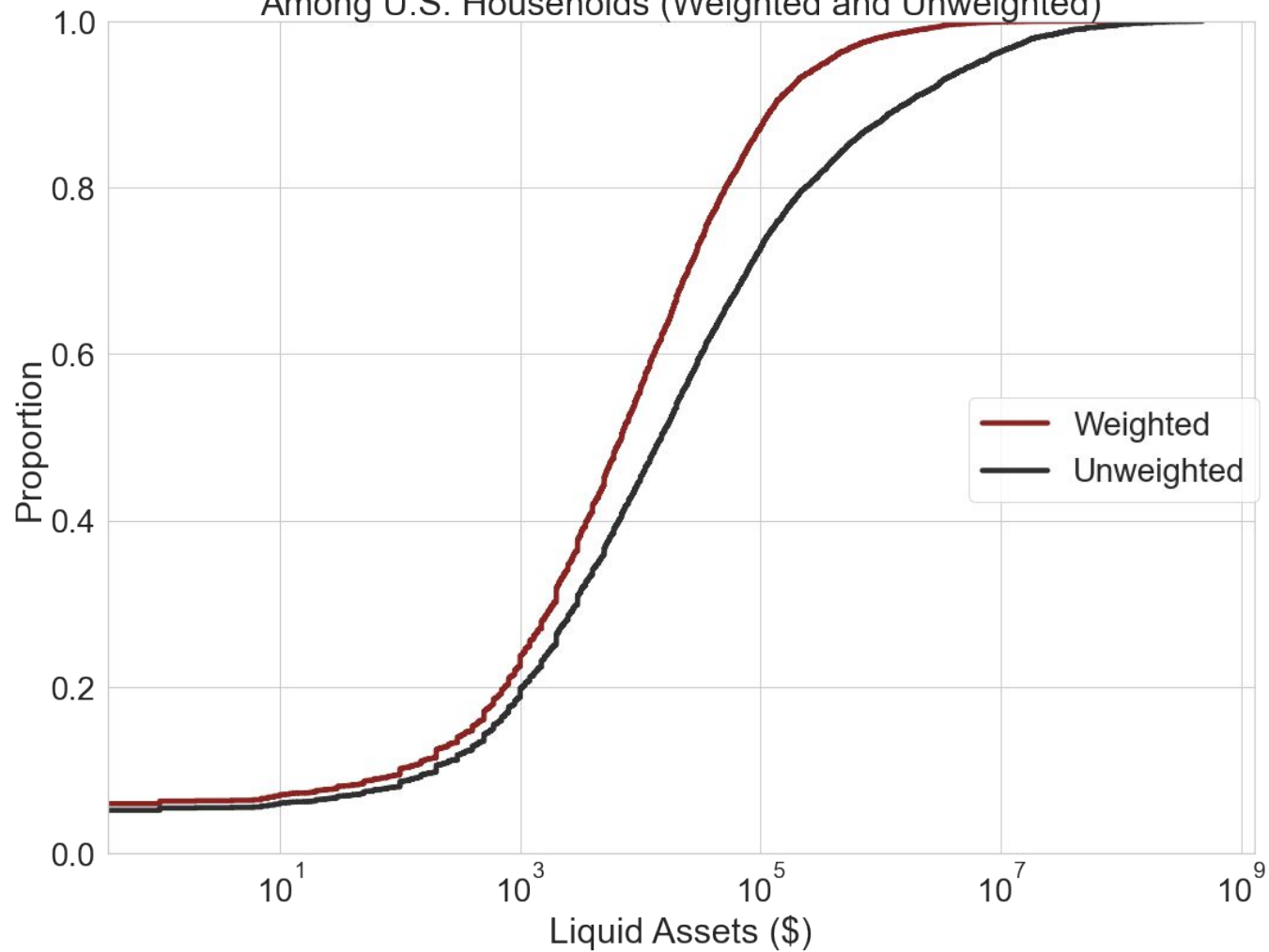
Modeling

Overcoming Hurdles

- Sample weights
- Scaling

Evaluating final model

- Optimizing F1 Score: decreasing both false positives and false negatives
- .916 compared to .baseline of 907

Vanguard®

Modeling

Scale of Logistic Regression Coefficients

Conclusions & Next Steps

Conclusions

- Credit reports and banking integration will yield useful information

Next Steps

- Improve feature engineering from abundance of data
- Consider removing outliers

Vanguard®

Contact Info & Credits

Dan Valenzuela

dan@danvalenzuela.com

GitHub.com/danvalen1

Credits

Wikimedia   Eucalyp

Freepik   Becris

Vanguard®