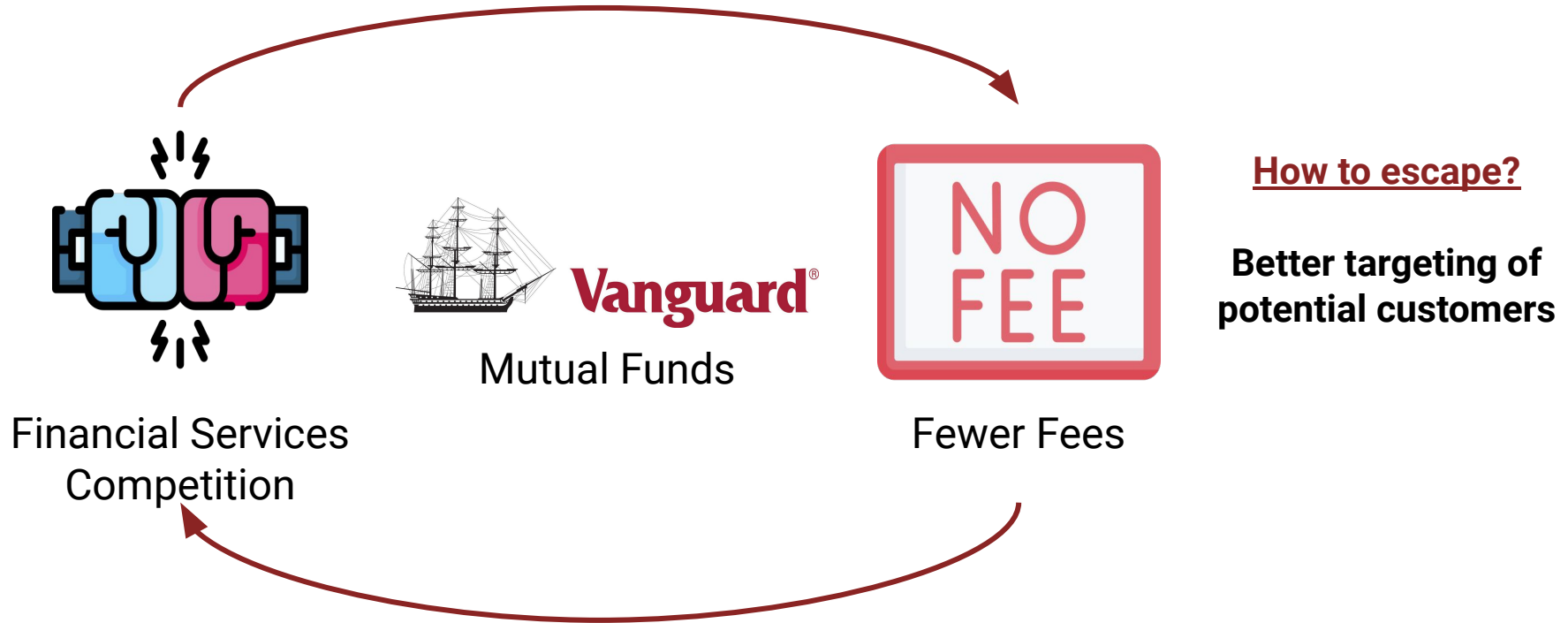


Vanguard's \$1,000 Question

Analyzing U.S. Households' Liquid Assets

Dan Valenzuela

Business Background

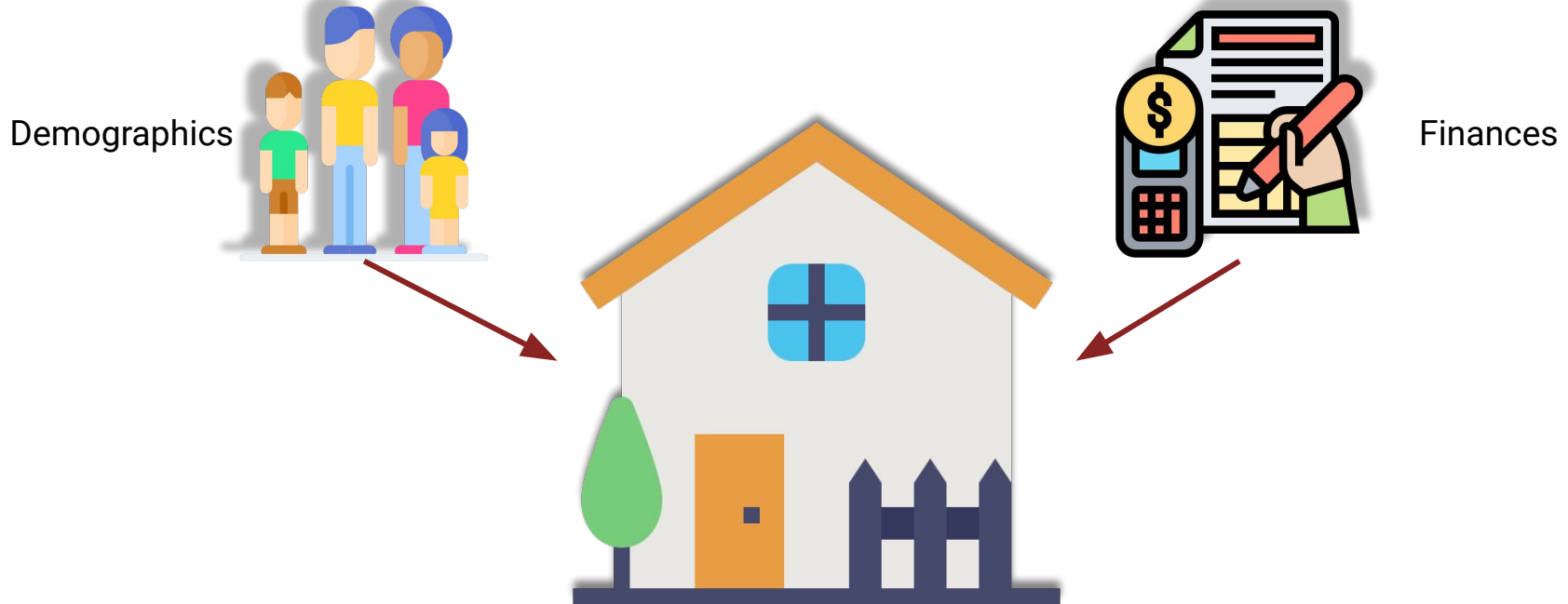


Method

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

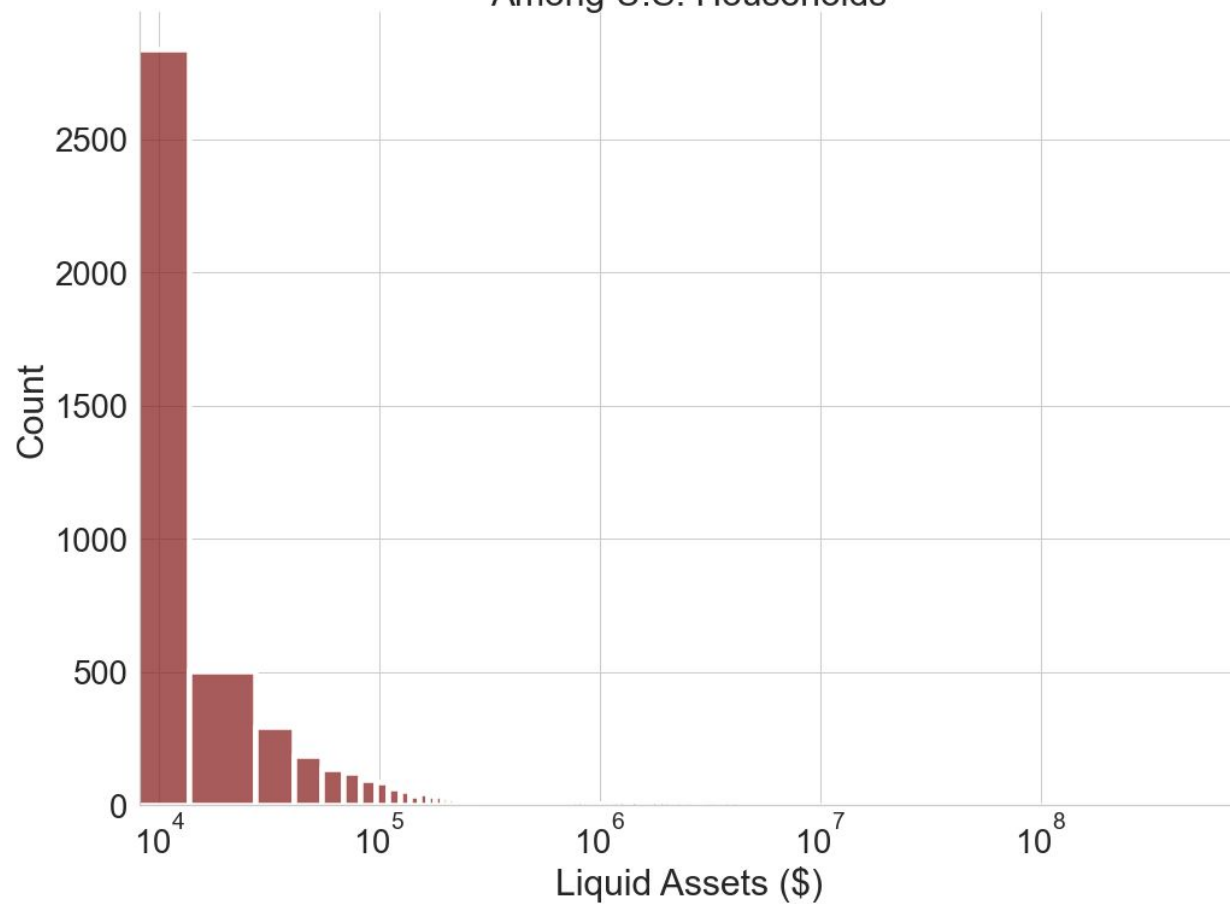
Data - Potential Customers



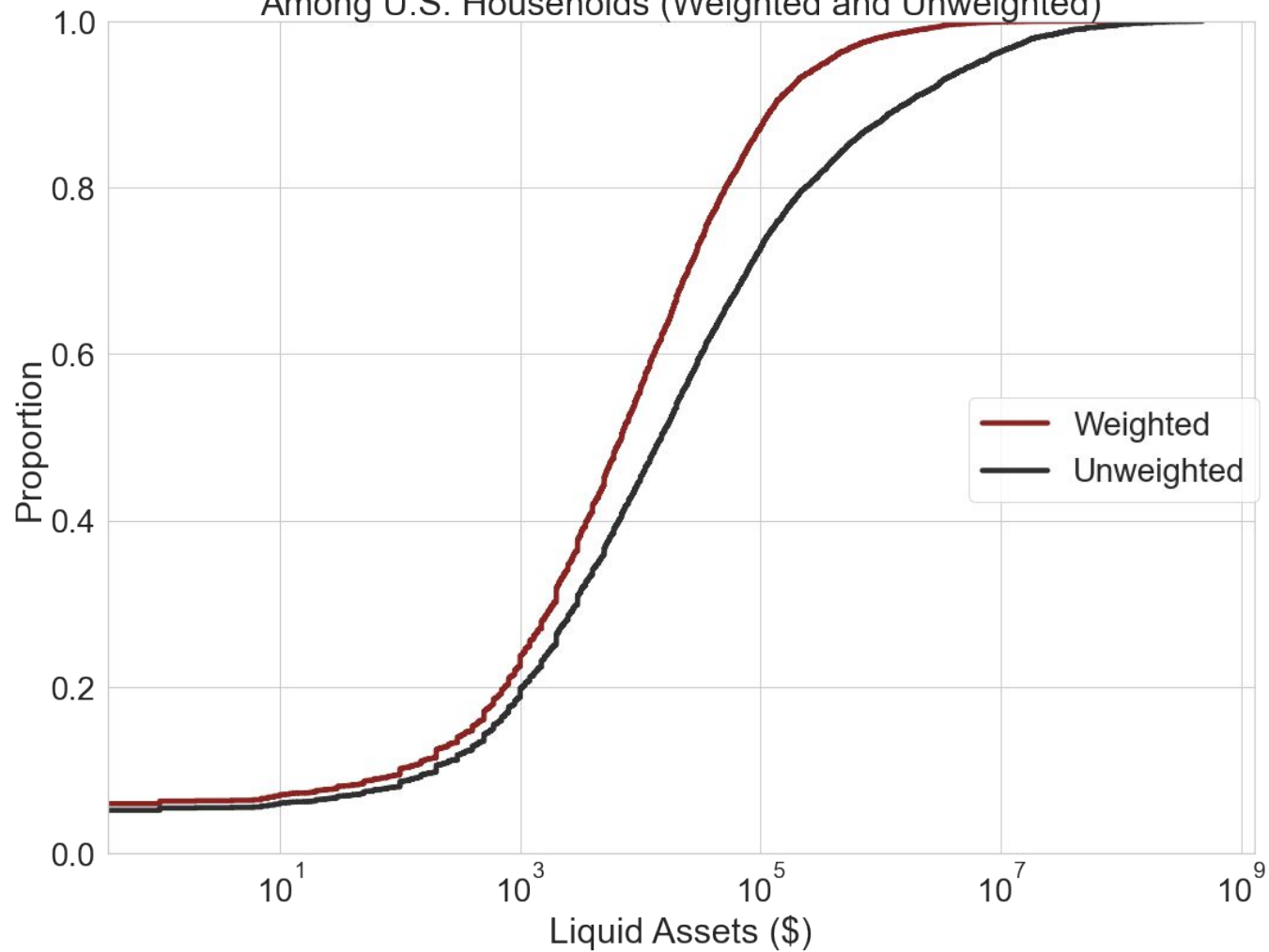
Data Hurdles

1. Skewed distribution of wealth
2. Overrepresentation of wealthy households

Distribution of Liquid Assets
Among U.S. Households



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



Modeling

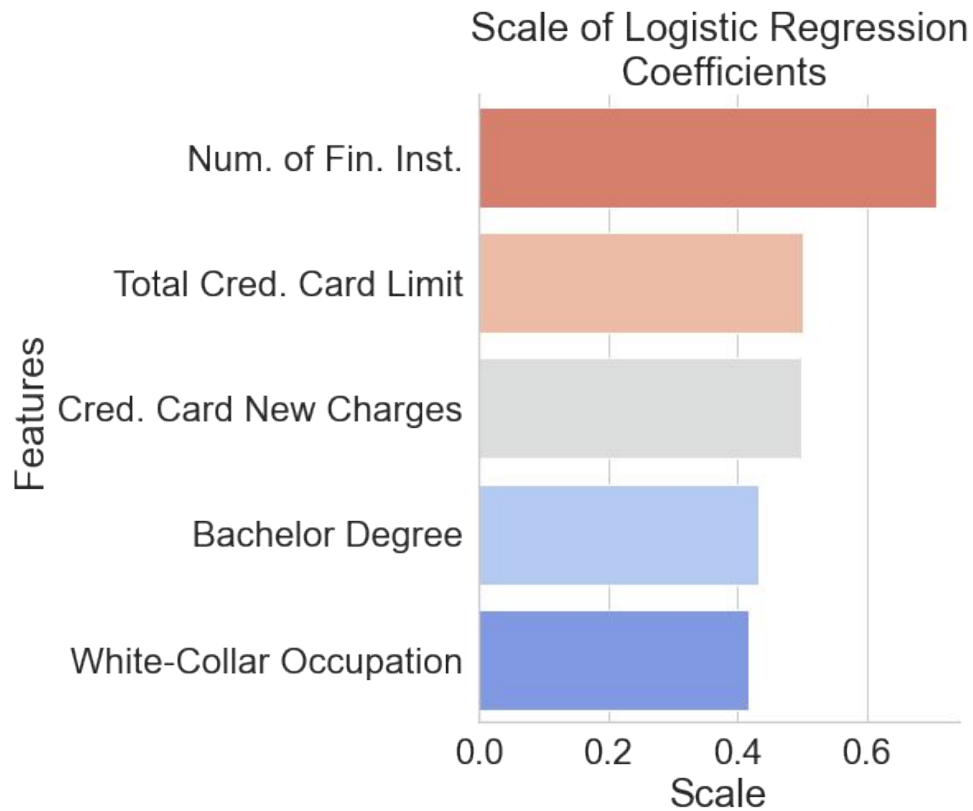
Overcoming Hurdles

- Sample weights
- Scaling

Evaluating final model

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

Modeling



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

Contact Info & Credits

Dan Valenzuela

dan@danvalenzuela.com

[GitHub.com/danvalen1](https://github.com/danvalen1)

Credits

Wikimedia

[Eucalyp](#)

[Freepik](#)

[Becris](#)