Date: 24 March, 2025

Objective: Maximize revenue from direct marketing campaigns using provided data.

Proposed Solution:

1. Code and Results: Notebook Marketing\_Optimization.ipynb contains code for train data preparation, building propensity models for each product category and code for maximization of revenue under given constraints along with result expected revenue. List of selected clients for targeting is saved in a separate file targeted\_clients\_lp.csv. Code file contains explanation of logic along with necessary plots for visualization.

2. Brief Explanation of Solution:

Training Data preparation: In provided dataset, there are four major categories of data: social demographics, product holdings, sales/revenue and financial transactions. Data per client was merged before creating training and test sets. Product holdings and Sales Revenue had missing values for certain columns which were handled as 0. Since problem statement specifically mentioned that only 60% of client samples contain sales and revenue data and meant for training set, the fact was used for train/test split.

Segmentation: To better understand client characteristics and provide meaningful insights after modeling and optimization, clients were segmented with k-means. This also helped identify few high-revenue groups.

Propensity models: To estimate the likelihood of a client purchasing consumer loan, credit cards or mutual funds, three logistic regression models were trained. Inspected coefficients for every feature to explain and inspect importance while determining propensity score for particular product i.e. why a certain client was likely to adopt that product. For each of these products, based on output probabilities of models, expected revenue for per product was calculated for use during optimization step. [Note: From alternative propensity models, due to time constraints, need for further fine tuning and explainability reasons did only preliminary explorations using xgboost and attention-mechanism based neural networks which are currently not included in notebook]

Revenue maximization: For marketing campaigns under given constraints of contact limits (approx 100 clients) and only one offer per customer, prepared linear programming formulation as follows:

Objective function: maximize Total expected revenue

Decision Variable: Whether a client is targeted for particular offer

Constraints:

* Contact limit: only specific number of clients (I.e. 100) can be contacted
* Single offer per client: each client receives exactly one offer
* Binary variables: Decision variable can only be 0(not selected) or 1(selected).

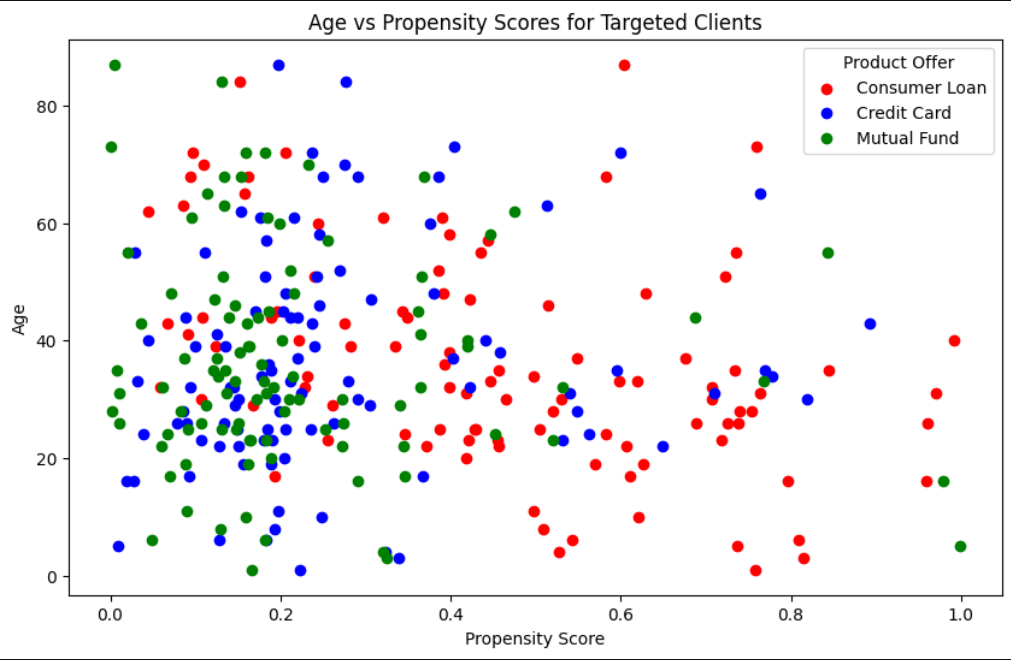
After of running linear programming optimization using linprog from scipy.optimize, as a result got total revenue post maximization under given constrains which is sum of revenues for chosen clients for targeting and selected offers. The clients with non-zero decision variables are the ones targeted in the campaign. Here, due to single offer constraints included in formulation, each client is targeted with exactly one offer. For solver, used highs method which is designed to efficiently handle large-scale problems with sparse structures.

3. Analysis and Insights:

Users with High-propensity for credit cards are users with mostly moderate balances with high transaction volumes. For example, Users targeted in segment 5 is aged between 30-35, with around relatively longer tenure and high number of credit/debit transactions, were chosen for credit card offer.

High-propensity clients for mutual funds generally exhibit financial stability with higher savings balances and existing investments. For example, among targeted users belonging to segment 1 are mid to old age customers, with moderate to high balances and high volumes for debit transaction along with good existing adoption of offers. So they were primarily chosen for Mutual fund investments.

High-propensity clients are for consumer loans are typically those with liquidity needs, lower account balances, and younger tenure profiles. Users targeted in segment 3 are mostly young clients with moderate to low balances with higher monetary needs, who were primarily chose for consumer loans. Additionally, upon inspection of Age vs. Propensity scores plot, we can see that Clients targeted for consumer loans are mostly situated in younger age buckets.



Note: Current approach is bit limited due to missing revenue numbers for some existing sales, which will impact expected revenue calculation. Given more time, models can be developed to impute/predict revenue for such cases which will drive final maximized revenue during optimization step to be more precise and bit higher than current estimate.