Executive Summary: Direct Marketing Optimization Project

Project Objectives

The primary goal was to develop a data-driven approach to determine the most appropriate offer for each client and estimate the expected revenue from the targeted marketing strategy

Methodology Overview

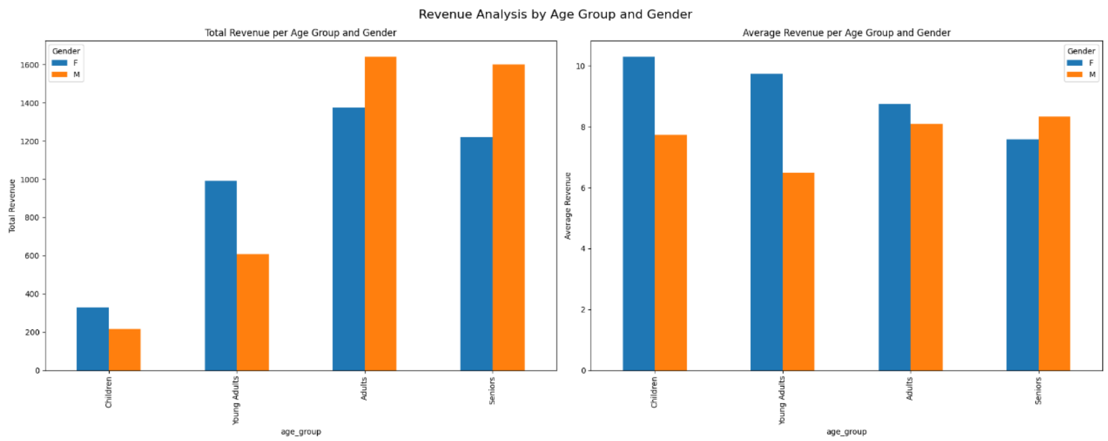
A two-stage predictive modelling approach was used:

- **Stage 1: Propensity Modelling -** Predict the likelihood of product subscription
- **Stage 2: Revenue Prediction -** Estimate potential revenue for each product

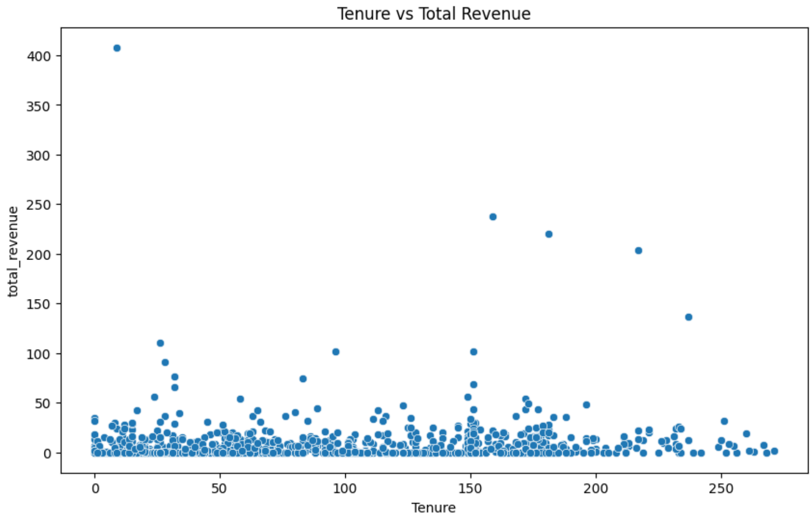
By combining the outputs of both stages—propensity scores and expected revenue—the clients were ranked based on their potential value to the company. This integrated approach ensures that marketing efforts are directed towards clients who not only have a high likelihood of purchasing but also represent significant revenue potential. This dual focus on both likelihood and potential revenue enables more strategic targeting, ultimately maximizing overall revenue generation for the company.

Insights from exploratory analysis

1. **Data Quality and Consistency**:
    - There were no duplicated client IDs found in the dataset.
    - Checks were done to find inconsistencies in the data. All product counts were consistent with their respective account balances; if a product existed, the corresponding balance was also present.
    - Assumption that df_inflow_outflow was generated from the dates qualified as "active" period was also tested. Meaning if there were no transactions for a client it would mean that the client was not active during the period and hence marketing should not target these clients. But this assumption was invalidated as some “inactive” clients contributed to revenue
2. Gender distribution was almost balanced. Interesting insight uncovered was that Young Females (<30 years of age) are generating more total revenue as compared to males. Average revenue per client is higher for only Senior males compared to females in other age groups



- Total revenue from clients had no correlation with their tenure with the company



- There was right skew in almost every numeric column except

<u>Modelling Methodology</u>

- Handling missing values
  - The ‘Sex’ column had 3 null values, and they were imputed using the mode
  - Product_ActBalance had many missing values. Flags were created for situations where product exists but balance is 0 and then the remaining missing values were imputed with 0
  - Not all clients that had a MF, CC or CL were present in the Soc_Dem client list. This resulted in more null values after merging all tables into one
- Feature creation
  - Many new features were created like ratios, volume per transaction, one hot encoding of gender
- Correlation plots revealed highly correlated variables which were removed from the feature list. The threshold was set to 0.85
- Feature Transformation
  - Since the numeric columns were highly skewed, sqrt and yeo-johnson transformation was applied but both methods did not significantly improve the skew
- Pre-processing
  - Standard scaler was applied to numerical variables and one-hot-encoding was applied to categorical variables. The only categorical variable available was the gender.
- Feature selection
  - Top 10 features were selected for the propensity and revenue prediction model using Randomforest
- Model training
  - Propensity modelling and Revenue prediction modelling was done separately for each of the 3 products.
  - 80% train data was chosen and 20% test. The gender ratio was neutral so stratified split wasn’t necessary.

  - Grid search with cross validation was conducted and model with best params was selected. The model used was Randomforest classifier for propensity and Randomforest regressor for revenue prediction
- Combining both models
  - A combined score was calculated for every client, which was essentially propensity_scores * expected_revenues
- Selecting best product for each client
  - The best product was selected for each client based on the max expected revenue. For e.g, if expected revenue on credit card was the highest for the client, then credit card was chosen as the product
- Selecting top 15% clients
  - After selecting the best product, all clients were ranked in descending order based on the expected revenue they would bring if they opted for the product. Top 15% clients were chosen based on the expected revenue. This method was chosen to maximize the total revenue for the company. There could be other methods too, which need to be aligned with the product manager
  - A total of 242 clients were selected with one offer, and they are expected to bring $681.35 in revenue for the company.
  - Among these 242 clients, 65% should be targeted with consumer loan, 21.5% with credit card and the rest with mutual fund