

# Technical Report

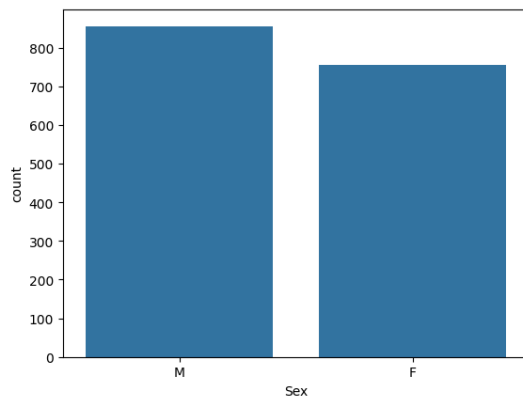
## Problem statement

- To optimize the direct marketing strategy for Mutual Funds, Credit Cards and Consumer Loan.
- Given consumers data which consumer would be highly likely to buy one of these products and what would be the expected revenue with the strategy.

## EDA

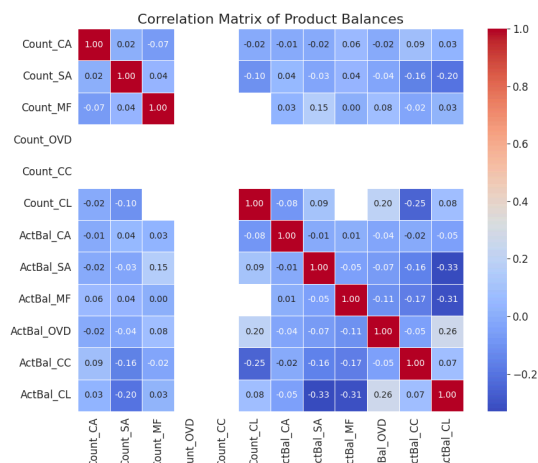
### Demographic Data

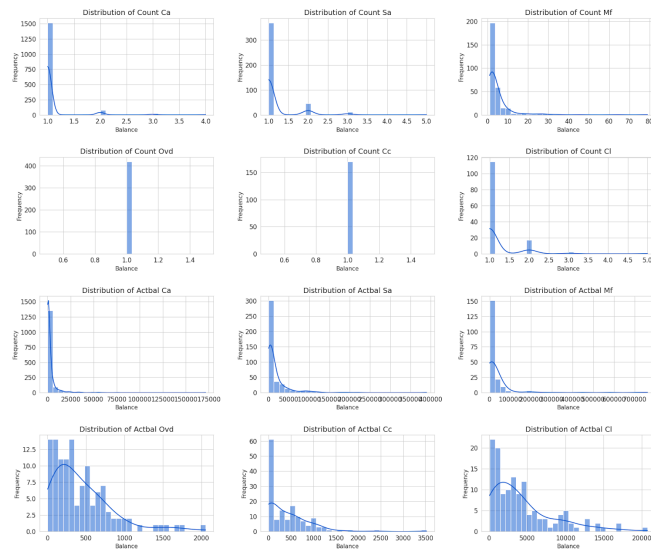
1. The demographic data consists of almost equal split of Male and Female



2. Tenure for some of the accounts is higher than the age. Possible explanation could be hereditary accounts.

### Account Balance Data





## Distribution of Account Balances

The balances for most products are heavily right-skewed, meaning a majority of clients have modest balances, while a few have exceptionally high balances.

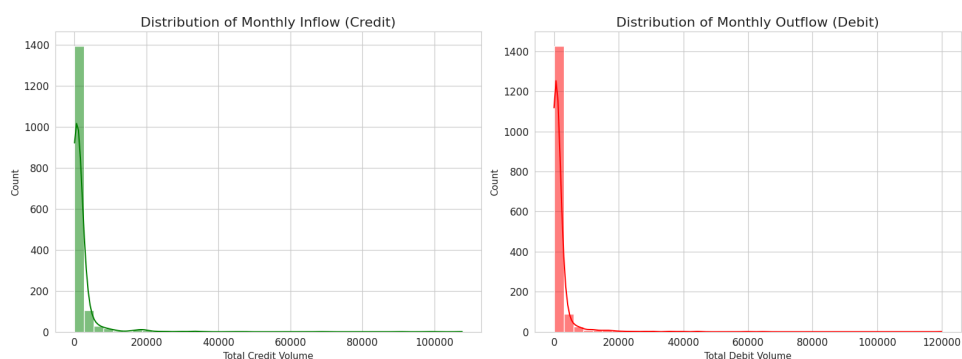
- **Mutual Funds and Consumer Loans** show the widest range and highest values, indicating they are high-value products.
- **Credit Card** and **Overdraft** balances are concentrated at lower values.
- **Thoughts:** The presence of high-balance outliers in products like Mutual Funds and Consumer Loans suggests that segmenting customers into "high-value" and "standard" tiers could be a very effective strategy.

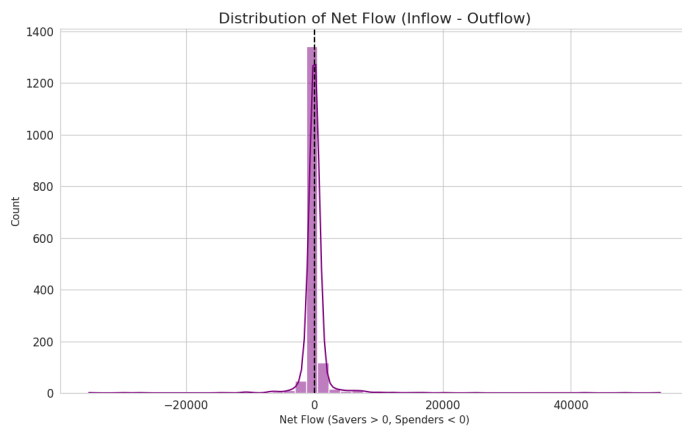
## Correlation of Balances

The correlation heatmap reveals how different product balances relate to each other.

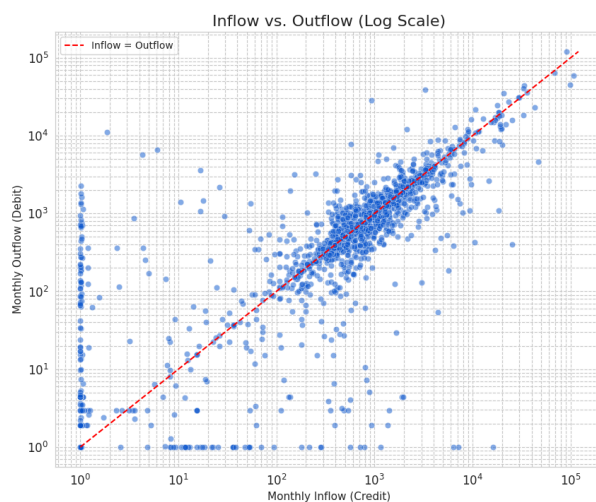
- There is a notable **positive correlation between Saving Account and Mutual Fund values.**
- There is a notable **negative correlation between Account balance of Over Draft and Consumer Loan values.**

## Inflow Outflow Data





- Both the total monthly inflow (credit) and outflow (debit) are heavily skewed to the right. This indicates that most clients have relatively modest monthly transaction volumes, while a smaller group of clients transact with significantly larger sums of money.



## Feature Engineering

Various features were created based on

- Interaction Features:
  - Which were mainly based on interactions between two columns example Account balance for MF X count of MF accounts
- Ratio Based Features:
  - Ratio based features like volume credited / volume debited, transaction credit ratio
- Grouping Features
  - Binning of various features like age was binned to age group

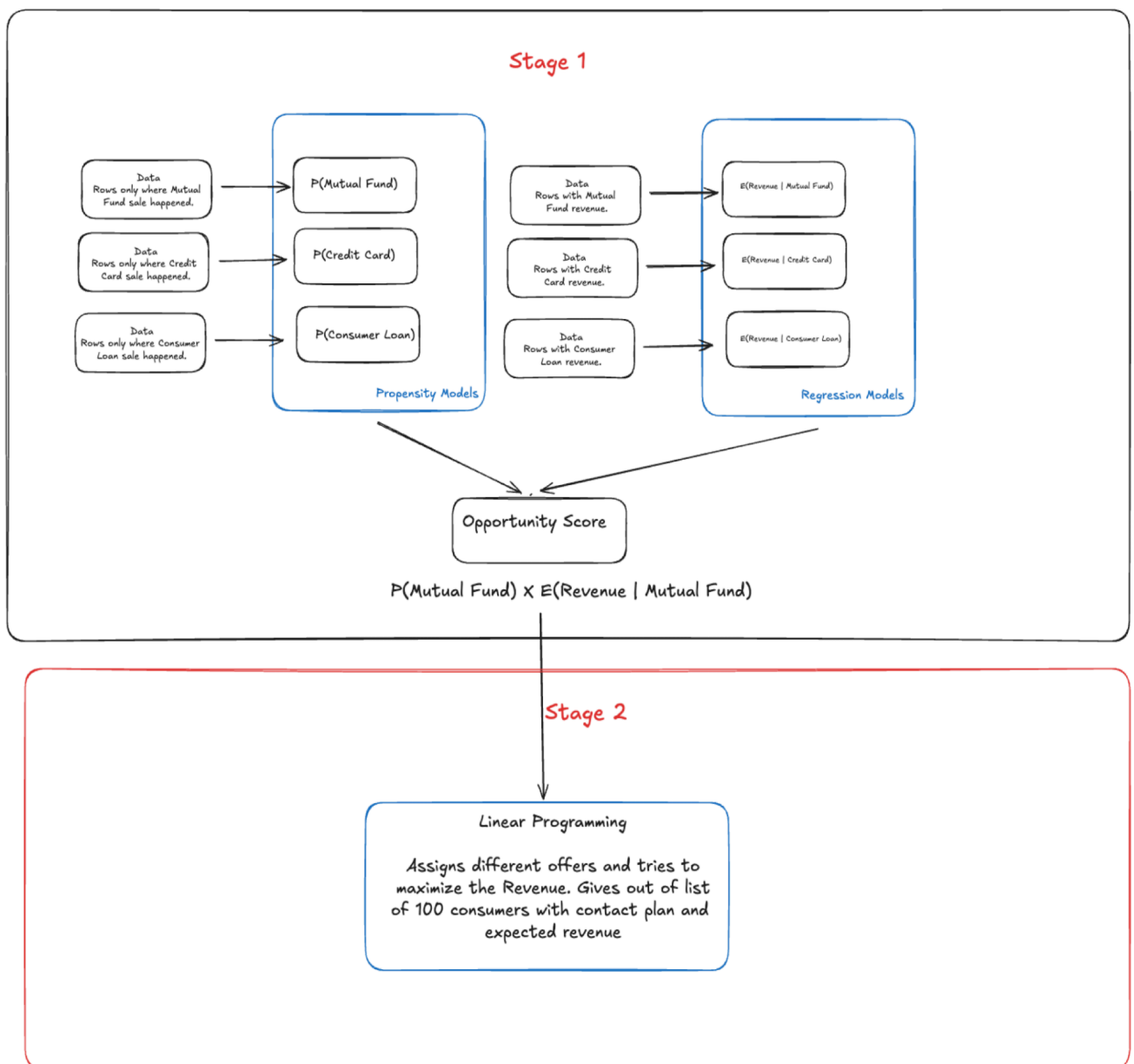
- b. Features like spender vs saver were created
- c. The Beginning of tenure was done too.

After all, the feature engineering model's performance did not improve significantly, the reason being low samples and high cardinality. At any given instance only 150-200 samples were present for each product. Resulting in no visible improvement.

Hence to make analysis simple I carried out with original features.

## Model Architecture

### Two Stage Approach



## **Stage One (Prediction Stage)**

### **Prediction Stage (Propensity Models)**

- Predicts Probability of a consumer buying a product (Credit Card, Mutual Fund or Consumer Loan)
- Method : Train 3 separate classification models which predict probability to buy.

### **Revenue Modelling Stage (Regression Models)**

- Predicts revenue from a user once they buy a product (Credit Card, Mutual Fund or Consumer Loan)
- Method : Train 3 separate regression models with forecasts of the user revenue.

## **Stage Two (Linear Programming Stage)**

Incorporates linear programming to optimise the consumer contact plan, keeping in all the constraints.

## **Model Performance**

### **Propensity Models Performance**

Xgboost	Roc-AUC	F1 Score	Recall	Precision
Mutual Fund	0.5024	0.3014	0.5641	0.2056
Credit Card	0.5304	0.3030	0.3125	0.2941
Consumer Loan	0.6202	0.4928	0.5862	0.4250

Model performance is based on a threshold of 0.5. This can be further improved based on the fine tuning of threshold. However my architecture used probability values and it wouldn't

affect the performance much. Fine Tuning threshold will improve results instilling confidence amongst the stakeholders.

### Regression Models Performance

Xgboost	MAE	RMSE	R2 Score
Mutual Fund	10.2576	18.6216	-0.3041
Credit Card	5.5560	10.5454	-0.1755
Consumer Loan	7.6468	17.9013	-0.0612

Regression models did not perform well. Since the dataset was very small XGBoost was not able to capture the complex relationships. Moreover the target variable had extreme outliers and was heavily skewed.

At this current state using the mean values of revenue would lead to better results But i strongly believe that with more data this could easily outperform the mean value strategy.

## Contact Plan

### Contact Plan

Product	Client Count	Revenue
Mutual Fund	15	40.970592
Consumer Loan	65	642.699463
Credit Card	20	83.309937

**Expected Revenue - € 767**

I have added additional constraints like 15% of the total contacted population should be Mutual Funds and 20% should be Credit Cards.

Currently the revenue of Consumer Loans are very high compared to other products. Due to which a bias was introduced in the architecture. Leading the LPP to always allocate the consumer loans.

## **Improvements**

1. Major improvement would be training regression models on more data, this would lead to better and robust performance.
2. With the increase in data, I would be able to incorporate better feature engineering.
3. Currently I was only able to explore XGBoost based models, better results can be obtained with various ensemble techniques.