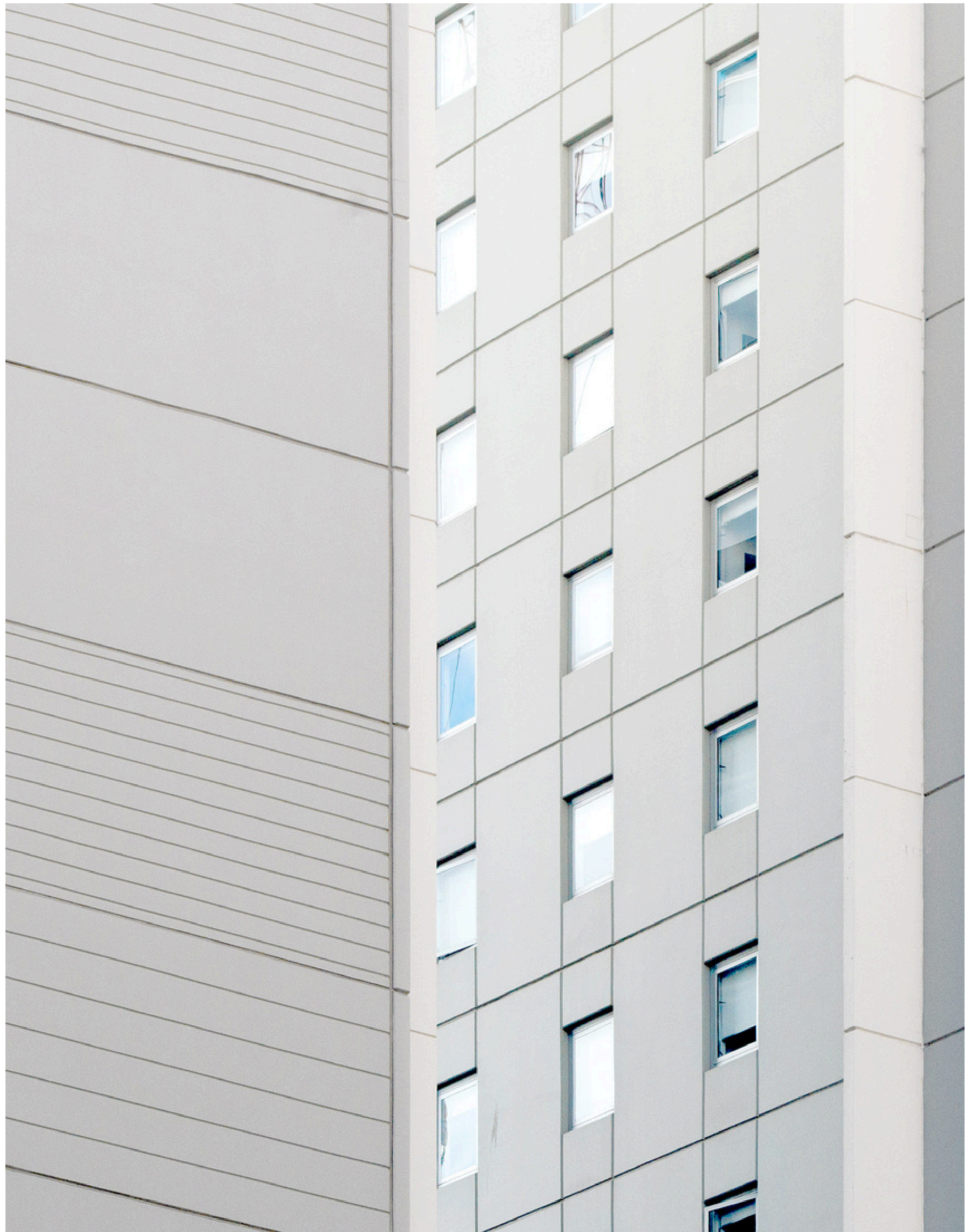
July 25, 2025

Data-driven marketing campaign

Jed Jerrel K. Escaran
BS Applied Physics - Instrumentation
National Institute of Physics, UP Diliman
Class of 2025

+639266552985 Quezon City, Philippines www.linkedin.com/in/jed-escaran



Scenario:

You are a **data analyst** of a bank and have the budget to contact only 100 people as part of a new marketing campaign.

You can offer these customers 1 of 3 different products:

1. Mutual Fund (MF)
2. Credit Card (CC)
3. Consumer Loan (CL)

Each of these 100 people can only receive 1 offer for 1 product and they should not have been part of the previous marketing campaign.

Objective:
select the **best 100 people** and **which product to offer them** based on maximum potential revenue

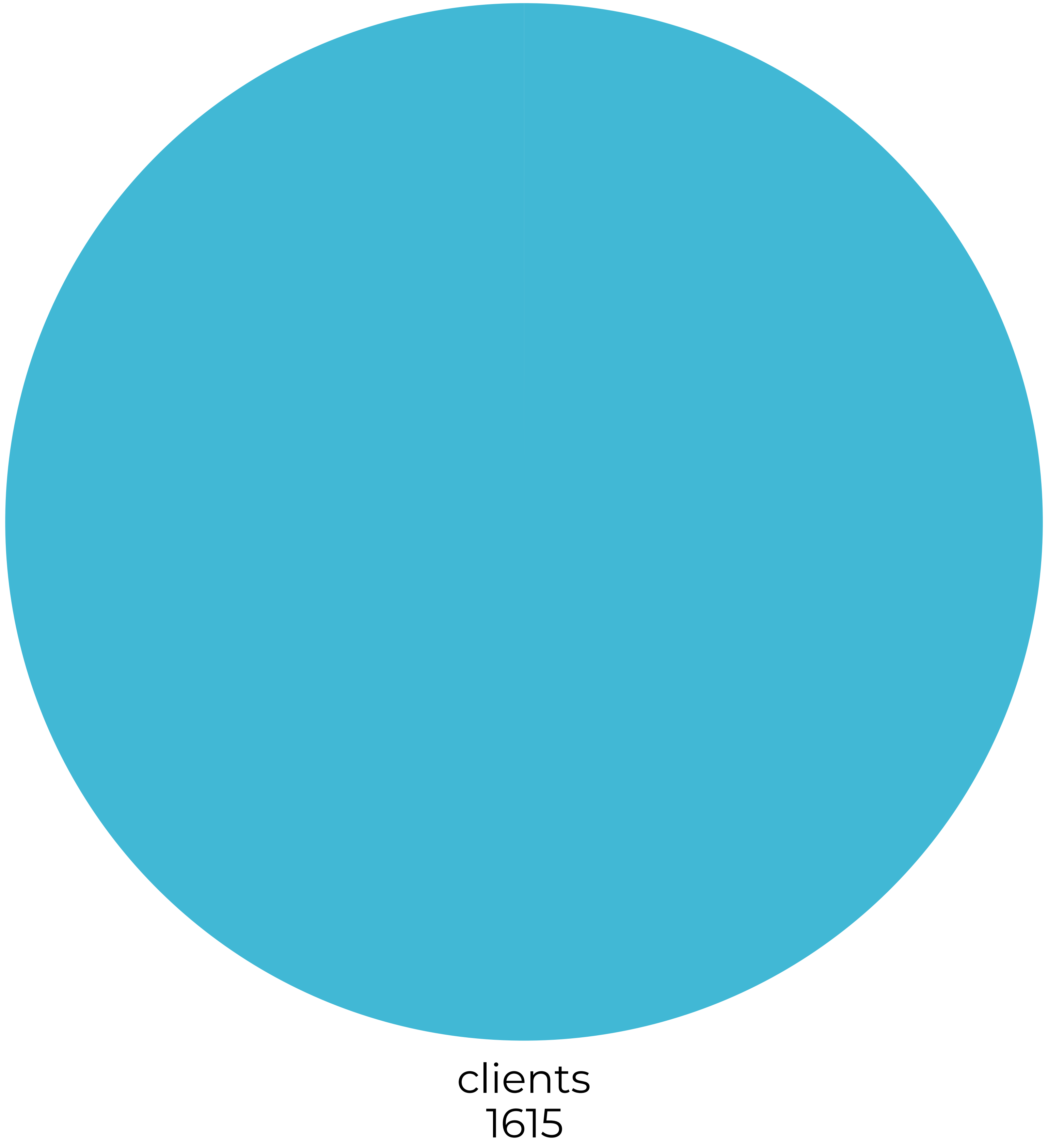
Data

Social-demographical data

Products owned and balances

Behavioral data

Sales and revenue



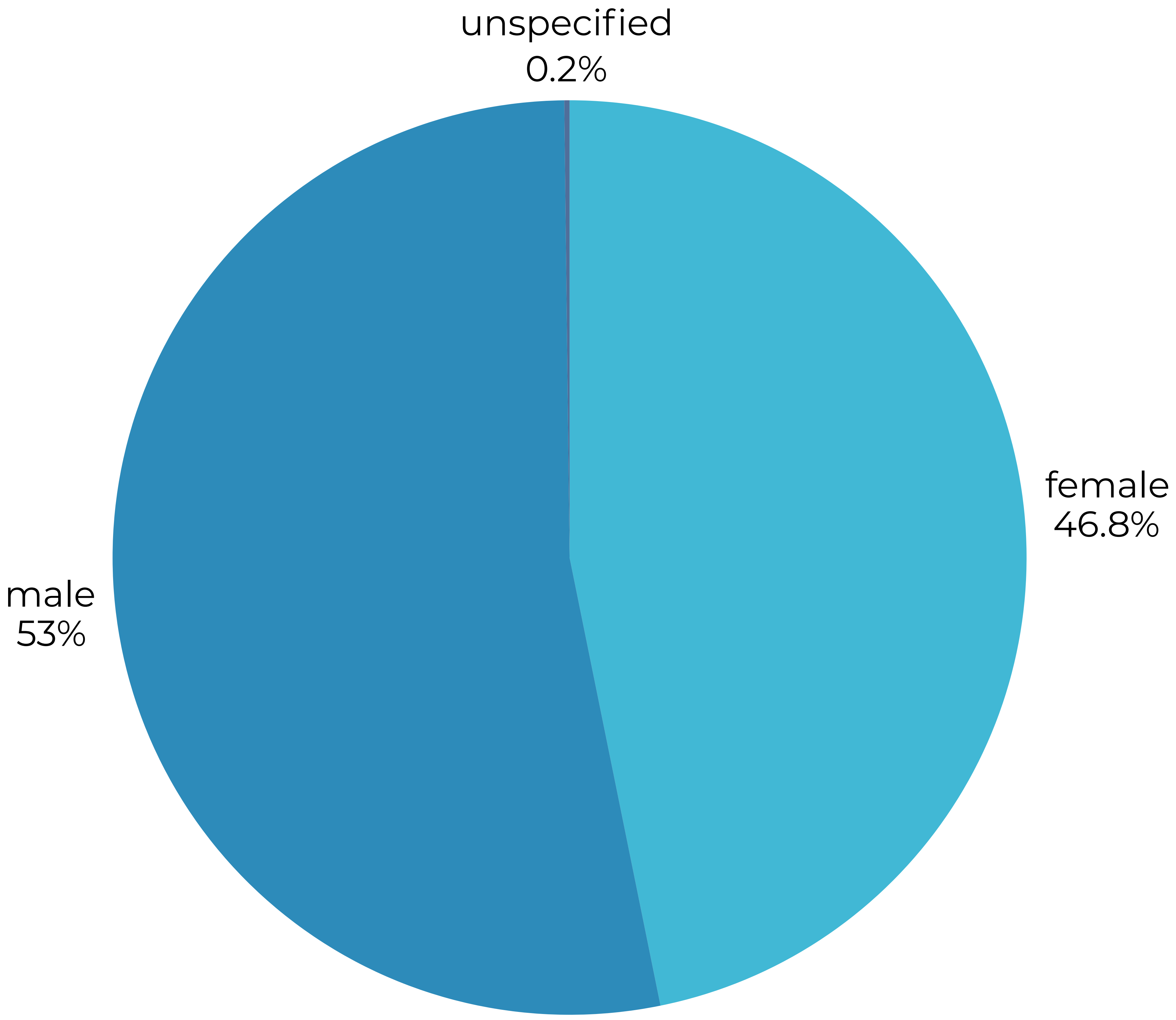
Data

Social-demographical data

Products owned and balances

Behavioral data

Sales and revenue



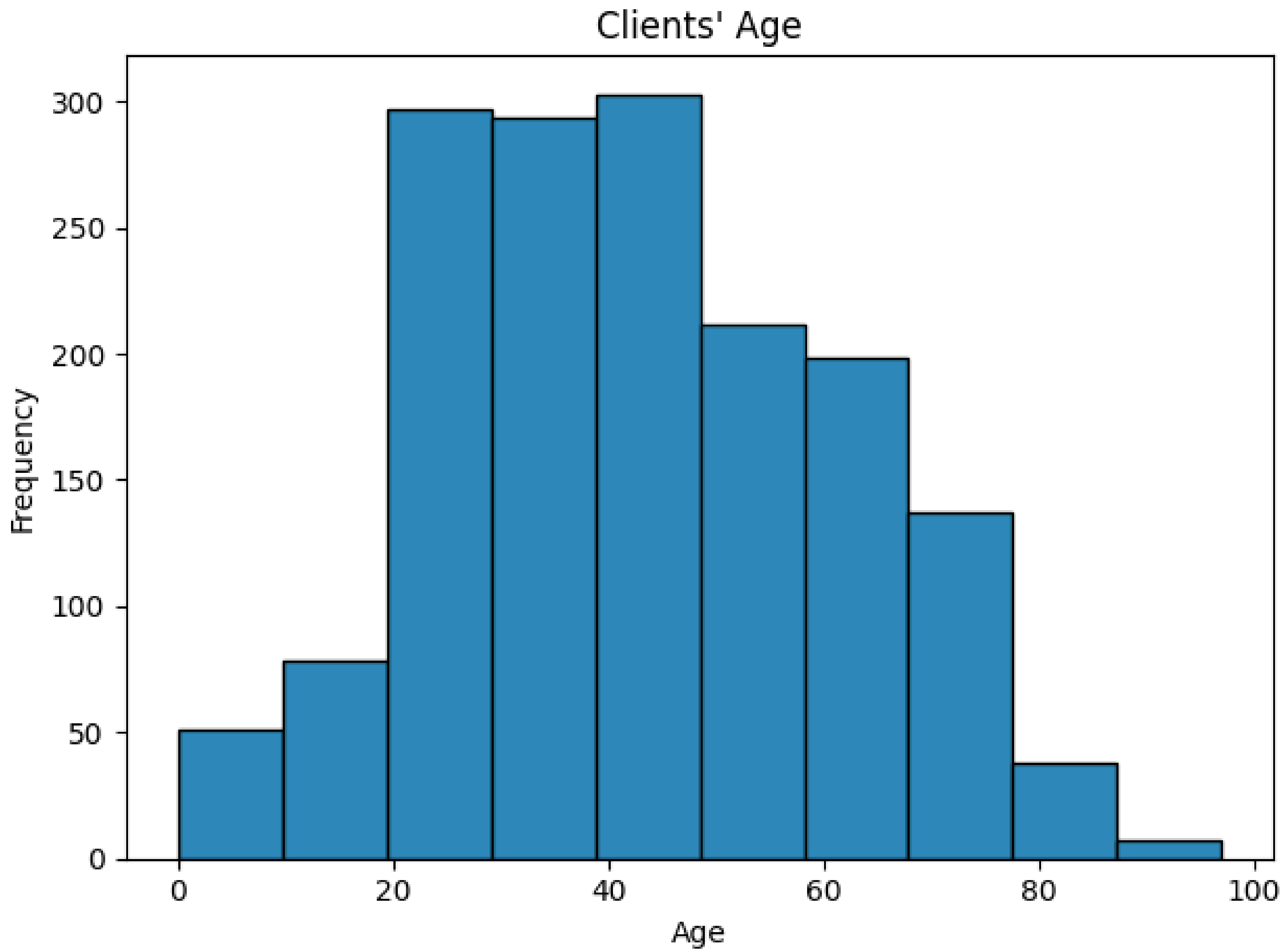
Data

Social-demographical data

Products owned and balances

Behavioral data

Sales and revenue



Data

Social-demographical data

Products owned and balances

Behavioral data

Sales and revenue

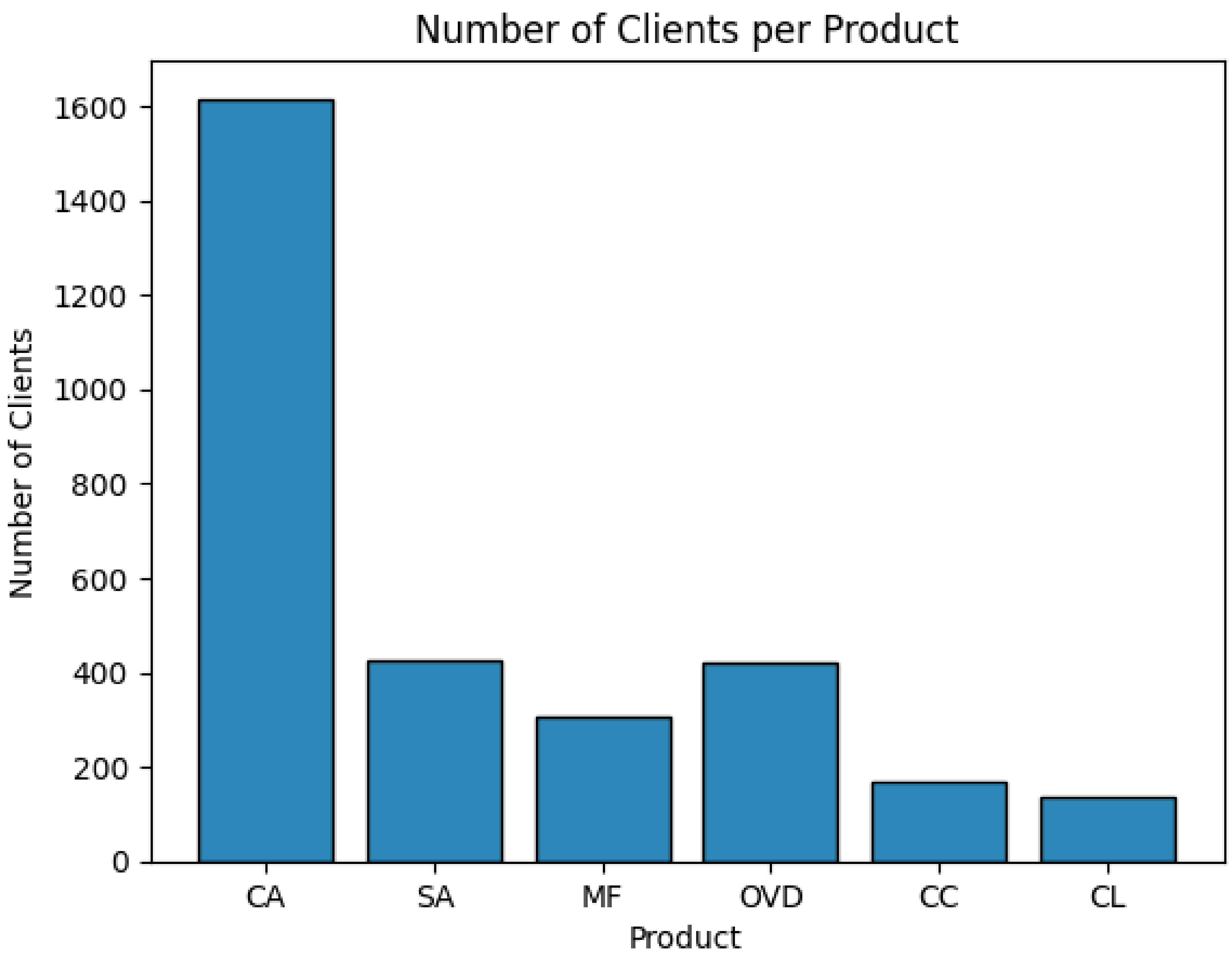
Data

Social-demographical data

Products owned and balances

Behavioral data

Sales and revenue



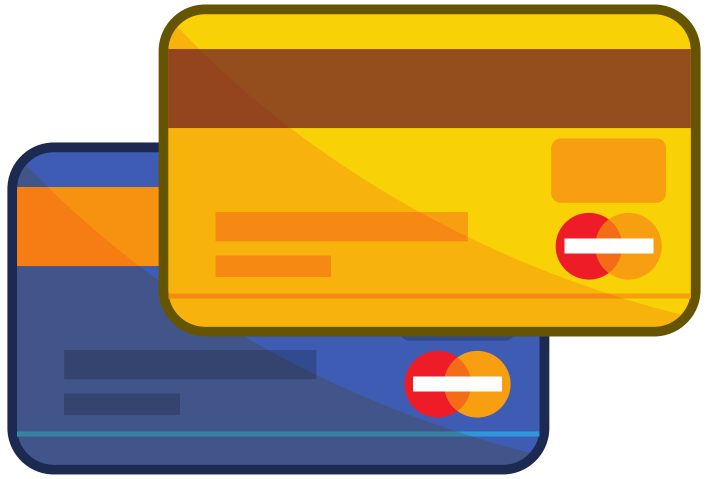
Data

Social-demographical data

Products owned and balances

Behavioral data

Sales and revenue



Inflow/outflow on current account, aggregated card turnover (monthly average over past 3 months)

contains per-client information on:

- credit and debit turnover
- number of transactions
- volume of transactions

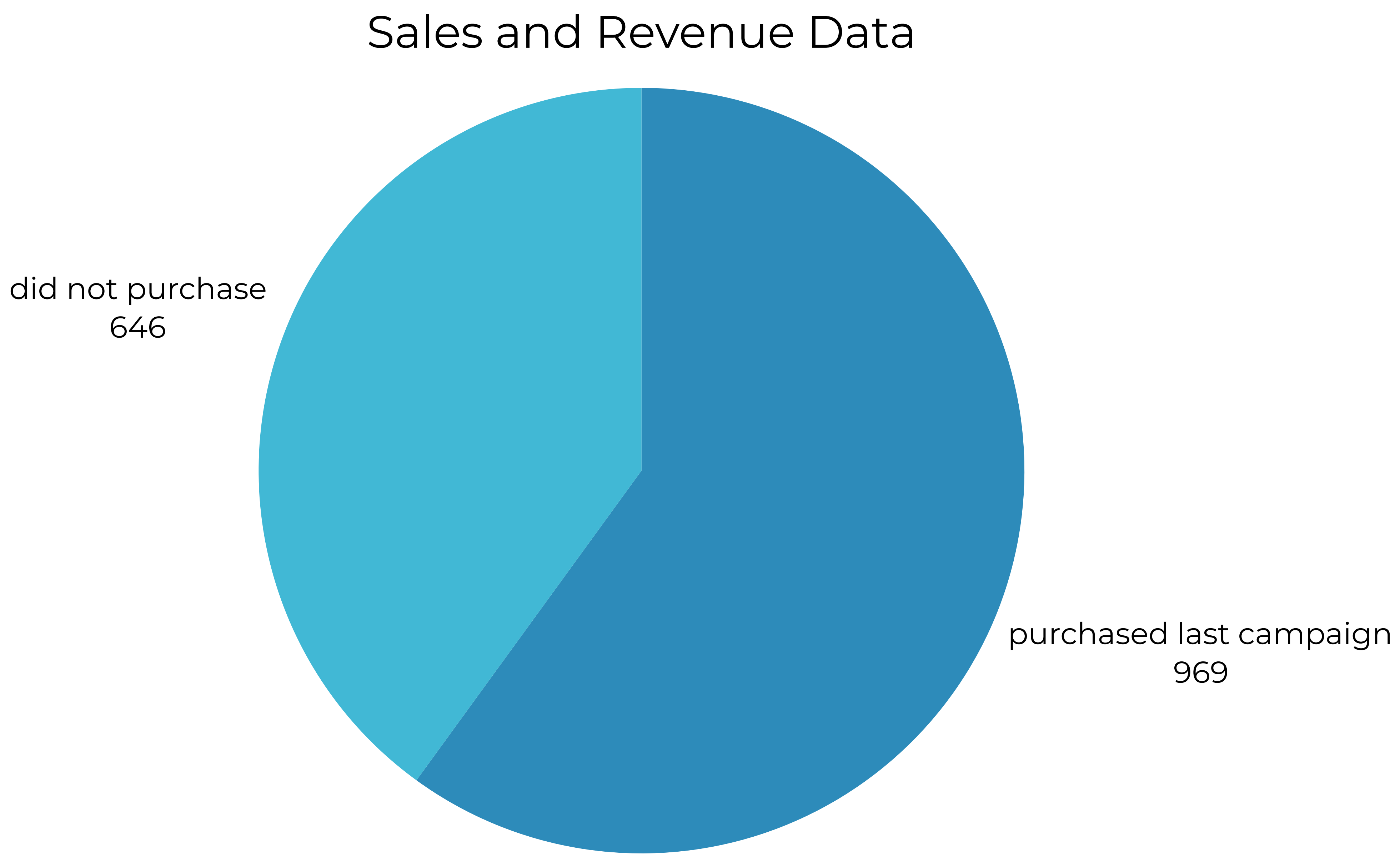
Data

Social-demographical data

Products owned and balances

Behavioral data

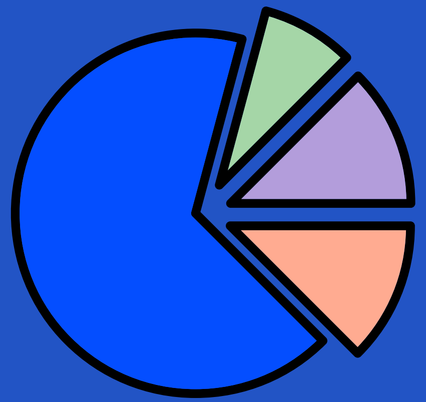
Sales and revenue



Approach

Data preprocessing

- Data reorganization, cleaning, and scaling
- Feature importance and selection



Data splitting and allocation

Dataset with sales and revenue information was divided into three parts:

- 60% for training
- 20% for testing
- 20% for validation

Model training and validation

Supervised machine learning models tested were:

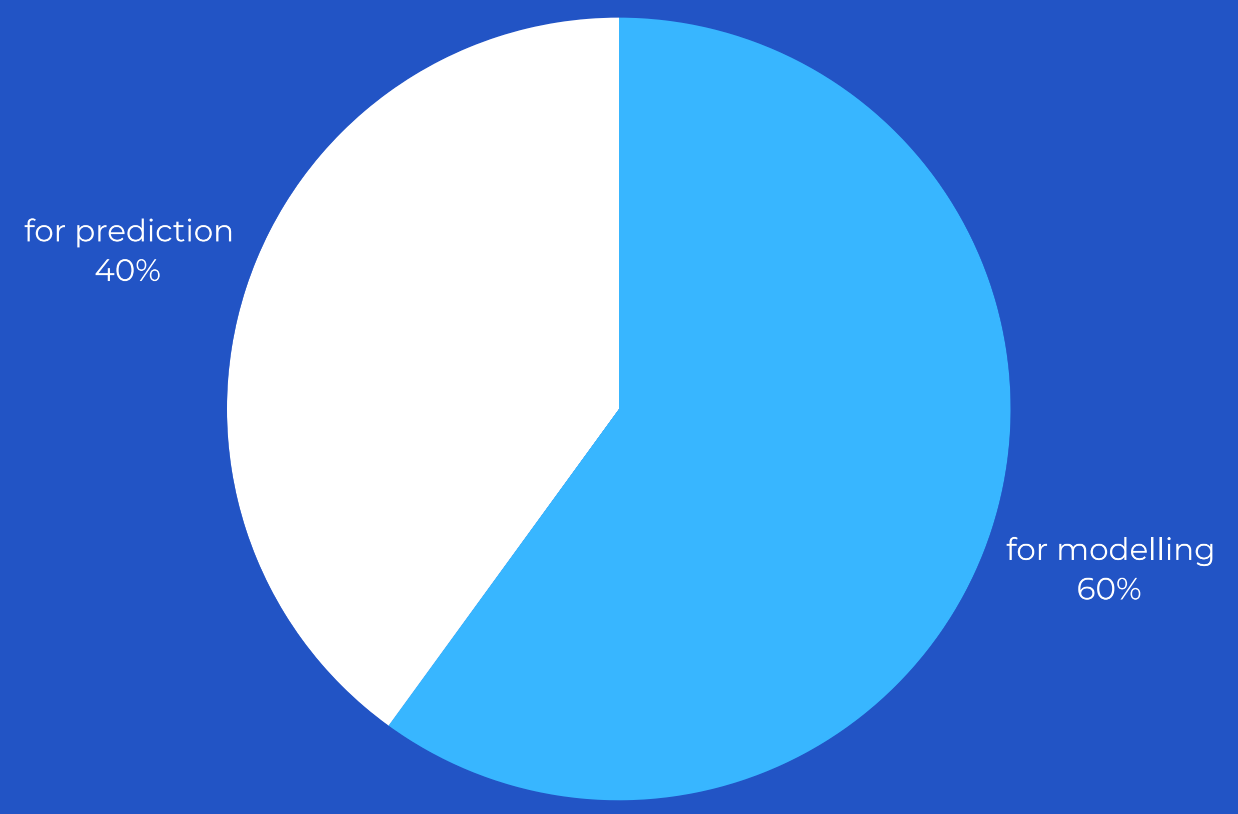
- Logistic Regression
- Random Forest Classifier
- Linear Regression
- Random Forest Regressor

Application of predictive model

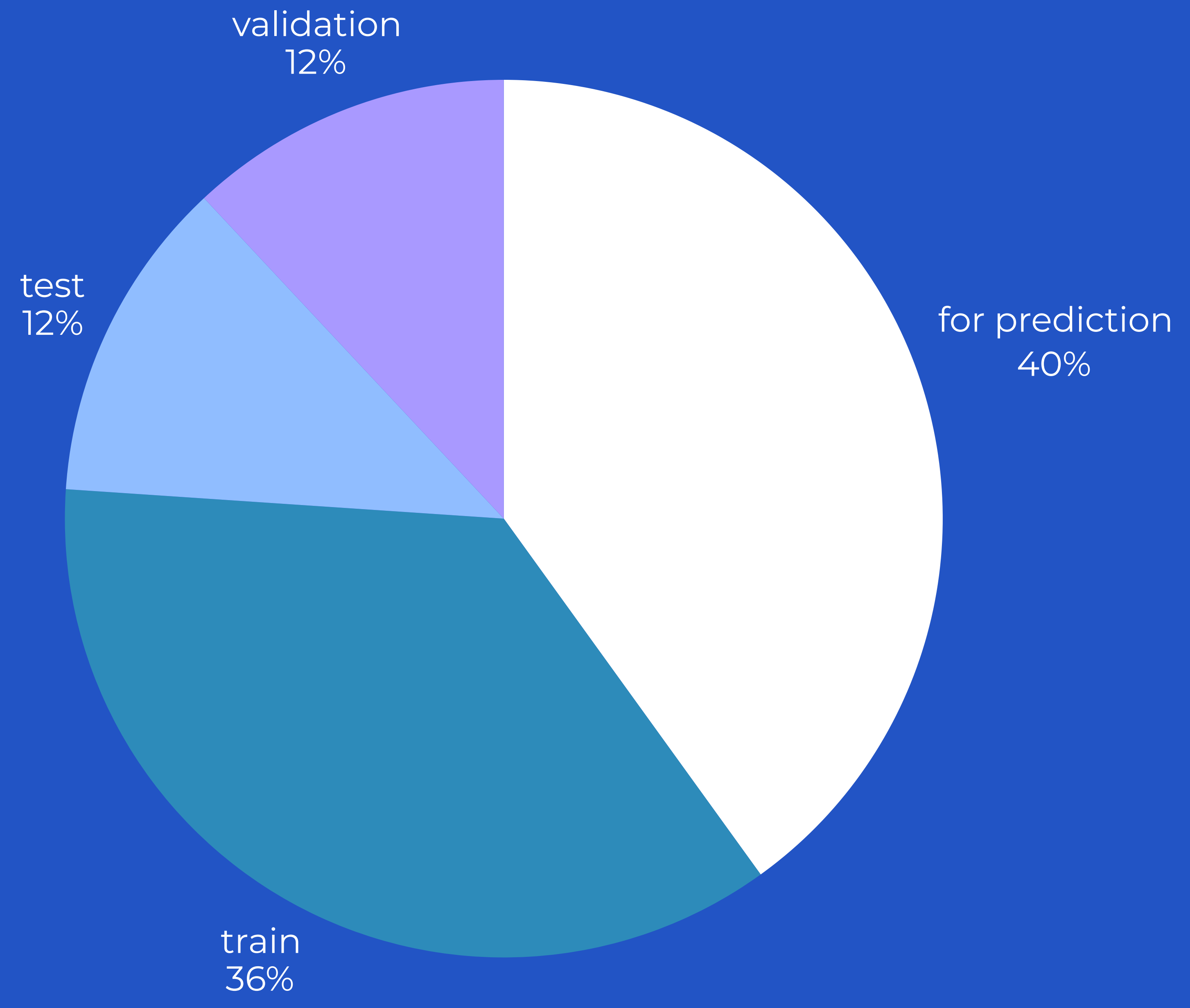
Probability for each client to accept an offer were obtained, and the potential revenue was calculated.

Approach

Sales and Revenue Data



Train, Test, and Validation split



Model Training

Features

Targets

	Sex	Age	Tenure	Count_CA	Count_SA	Count_MF	Count_OVD	Count_CC	Count_CL	ActBal_CA	...	Sale_MF	Sale_CC	Sale_CL	Revenue_MF	Revenue_CC	Revenue_CL
Client																	
1	0.0	51	7	1	0.0	0.0	1.0	0.0	0.0	1333.802857	...	1.0	0.0	0.0	26.972679	0.0	0.0
2	1.0	43	152	1	1.0	0.0	0.0	0.0	0.0	110.768571	...	0.0	0.0	0.0	0.000000	0.0	0.0
3	1.0	17	140	1	0.0	1.0	0.0	0.0	0.0	482.654643	...	NaN	NaN	NaN	NaN	NaN	NaN
4	1.0	24	153	1	1.0	0.0	0.0	1.0	0.0	1599.840714	...	NaN	NaN	NaN	NaN	NaN	NaN
5	0.0	58	200	1	1.0	0.0	0.0	0.0	0.0	5353.483929	...	NaN	NaN	NaN	NaN	NaN	NaN
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1611	0.0	41	181	1	0.0	0.0	1.0	0.0	0.0	0.000000	...	NaN	NaN	NaN	NaN	NaN	NaN
1612	1.0	63	114	1	0.0	0.0	1.0	0.0	1.0	170.593214	...	0.0	0.0	0.0	0.000000	0.0	0.0
1613	1.0	46	45	1	0.0	0.0	0.0	0.0	0.0	676.008571	...	1.0	0.0	0.0	2.088571	0.0	0.0
1614	1.0	48	65	1	0.0	0.0	0.0	0.0	0.0	0.010357	...	NaN	NaN	NaN	NaN	NaN	NaN
1615	1.0	7	180	2	1.0	0.0	1.0	0.0	0.0	118.938929	...	0.0	0.0	0.0	0.000000	0.0	0.0

The Sale_* columns were used as the target variables for **Logistic Regression**. Three models were trained, one for each product, using all 29 features.

Model Training

Features | **Targets**

Client	Sex	Age	Tenure	Count_CA	Count_SA	Count_MF	Count_OVD	Count_CC	Count_CL	ActBal_CA	...	Sale_MF	Sale_CC	Sale_CL	Revenue_MF	Revenue_CC	Revenue_CL
1	0.0	51	7	1	0.0	0.0	1.0	0.0	0.0	1333.802857	...	1.0	0.0	0.0	26.972679	0.0	0.0
2	1.0	43	152	1	1.0	0.0	0.0	0.0	0.0	110.768571	...	0.0	0.0	0.0	0.000000	0.0	0.0
3	1.0	17	140	1	0.0	1.0	0.0	0.0	0.0	482.654643	...	NaN	NaN	NaN	NaN	NaN	NaN
4	1.0	24	153	1	1.0	0.0	0.0	1.0	0.0	1599.840714	...	NaN	NaN	NaN	NaN	NaN	NaN
5	0.0	58	200	1	1.0	0.0	0.0	0.0	0.0	5353.483929	...	NaN	NaN	NaN	NaN	NaN	NaN
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1611	0.0	41	181	1	0.0	0.0	1.0	0.0	0.0	0.000000	...	NaN	NaN	NaN	NaN	NaN	NaN
1612	1.0	63	114	1	0.0	0.0	1.0	0.0	1.0	170.593214	...	0.0	0.0	0.0	0.000000	0.0	0.0
1613	1.0	46	45	1	0.0	0.0	0.0	0.0	0.0	676.008571	...	1.0	0.0	0.0	2.088571	0.0	0.0
1614	1.0	48	65	1	0.0	0.0	0.0	0.0	0.0	0.010357	...	NaN	NaN	NaN	NaN	NaN	NaN
1615	1.0	7	180	2	1.0	0.0	1.0	0.0	0.0	118.938929	...	0.0	0.0	0.0	0.000000	0.0	0.0

The Revenue_* columns were used as the target variables for **Random Forest Regressor**. Three models were trained, one for each product revenue, using all 29 features.

Model Tuning

Since we can only contact the top 100 clients out of the 646 potential buyers, we prioritized **Precision**.

Using RandomizedSearchCV, candidate models were evaluated based on their precision score during cross-validation by tuning the hyperparameter C.

mutual fund (MF)

	precision	recall	f1-score	support
0	0.8128	0.9806	0.8889	155
1	0.5714	0.1026	0.1739	39
accuracy			0.8041	194
macro avg	0.6921	0.5416	0.5314	194
weighted avg	0.7643	0.8041	0.7452	194

credit card (CC)

	precision	recall	f1-score	support
0	0.7772	0.9862	0.8693	145
1	0.8000	0.1633	0.2712	49
accuracy			0.7784	194
macro avg	0.7886	0.5747	0.5702	194
weighted avg	0.7829	0.7784	0.7182	194

consumer loan (CL)

	precision	recall	f1-score	support
0	0.7308	0.9779	0.8365	136
1	0.7500	0.1552	0.2571	58
accuracy			0.7320	194
macro avg	0.7404	0.5666	0.5468	194
weighted avg	0.7365	0.7320	0.6633	194

Predicting buyers

Models for MF, CC, and CL were all used to assess each client. The clients are then to be offered the product they are most likely to avail. The top 100 most likely buyers of any of the three products were selected.

Of all the 646 clients:

Mutual Fund	Credit Card	Consumer Loan
141	203	302

Of the top 100 clients:

Mutual Fund	Credit Card	Consumer Loan
33	24	43

Predicting revenue

Revenue-predicting models for MF, CC, and CL were all used on each client. The predicted revenues for each product per client were used to obtain the expected revenues.

Expected Value (EV):

$$EV = p(x) \cdot v(x)$$

where *p(x)* represent a probability of occurrence of *x* and *v(x)* represents the value if outcome *x* happens

Top 100 likely buyers:

Client	prob_1	product	predicted revenue	expected revenue
506	0.999747	MF	15.538361	15.534423
1414	0.999690	CC	13.633134	13.628904
1077	0.997296	CC	17.729156	17.681217
1455	0.994288	CC	14.495135	14.412339
766	0.982818	MF	18.326443	18.011552
...	...	...	...	...
940	0.524211	MF	10.646401	5.580964
1093	0.517859	MF	20.273788	10.498959
1119	0.517334	CL	12.727506	6.584373
710	0.515998	CL	14.038576	7.243879
1148	0.511020	MF	11.560733	5.907771

	Top 100 likely buyers	Top 100 expected revenue
Total expected revenue	€796	€930
Mutual Fund	33	16
Credit Card	24	45
Consumer Loan	43	39

Top 100 expected revenue:

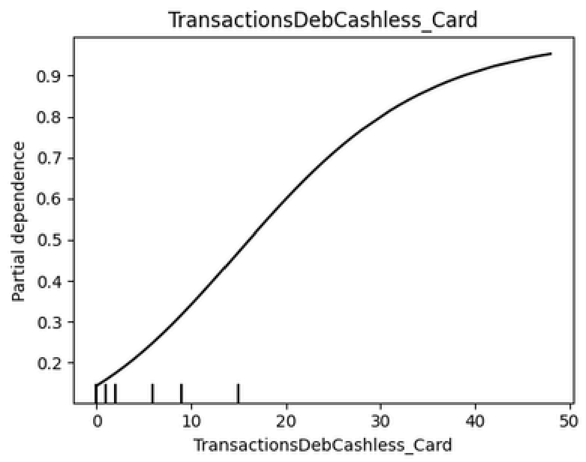
Client	proba_1	product	predicted revenue	expected revenue
766	0.982818	MF	18.326443	18.011552
1077	0.997296	CC	17.729156	17.681217
1278	0.794550	CC	21.814629	17.332813
506	0.999747	MF	15.538361	15.534423
1289	0.879492	CC	17.424485	15.324700
...	...	...	...	...
1343	0.533100	CL	12.781885	6.814025
1242	0.283602	CC	23.888789	6.774899
1026	0.259167	CC	26.077884	6.758515
706	0.497064	CL	13.592100	6.756141
1240	0.542945	CL	12.441903	6.755269

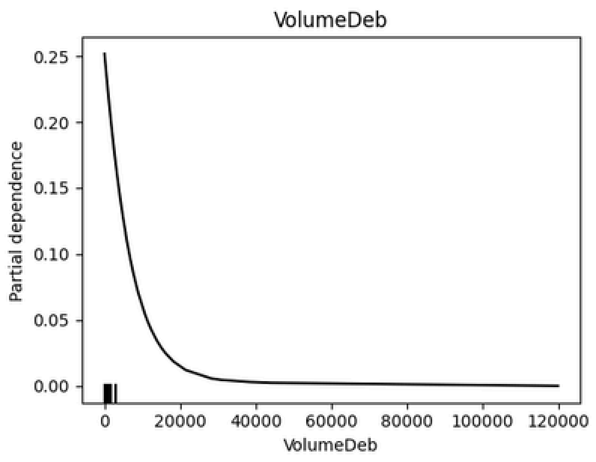
Identifying Clients

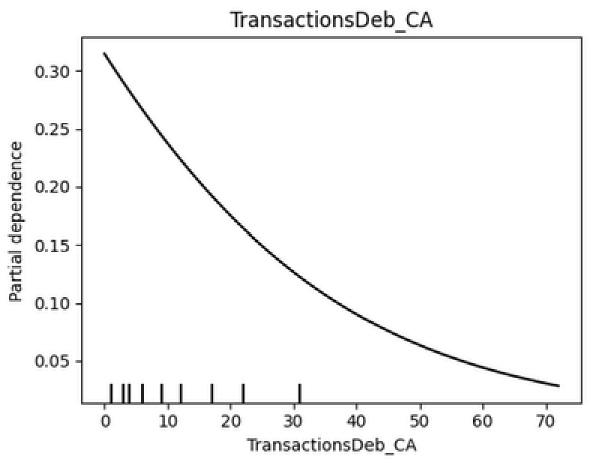
High likelihood clients for product **MF**

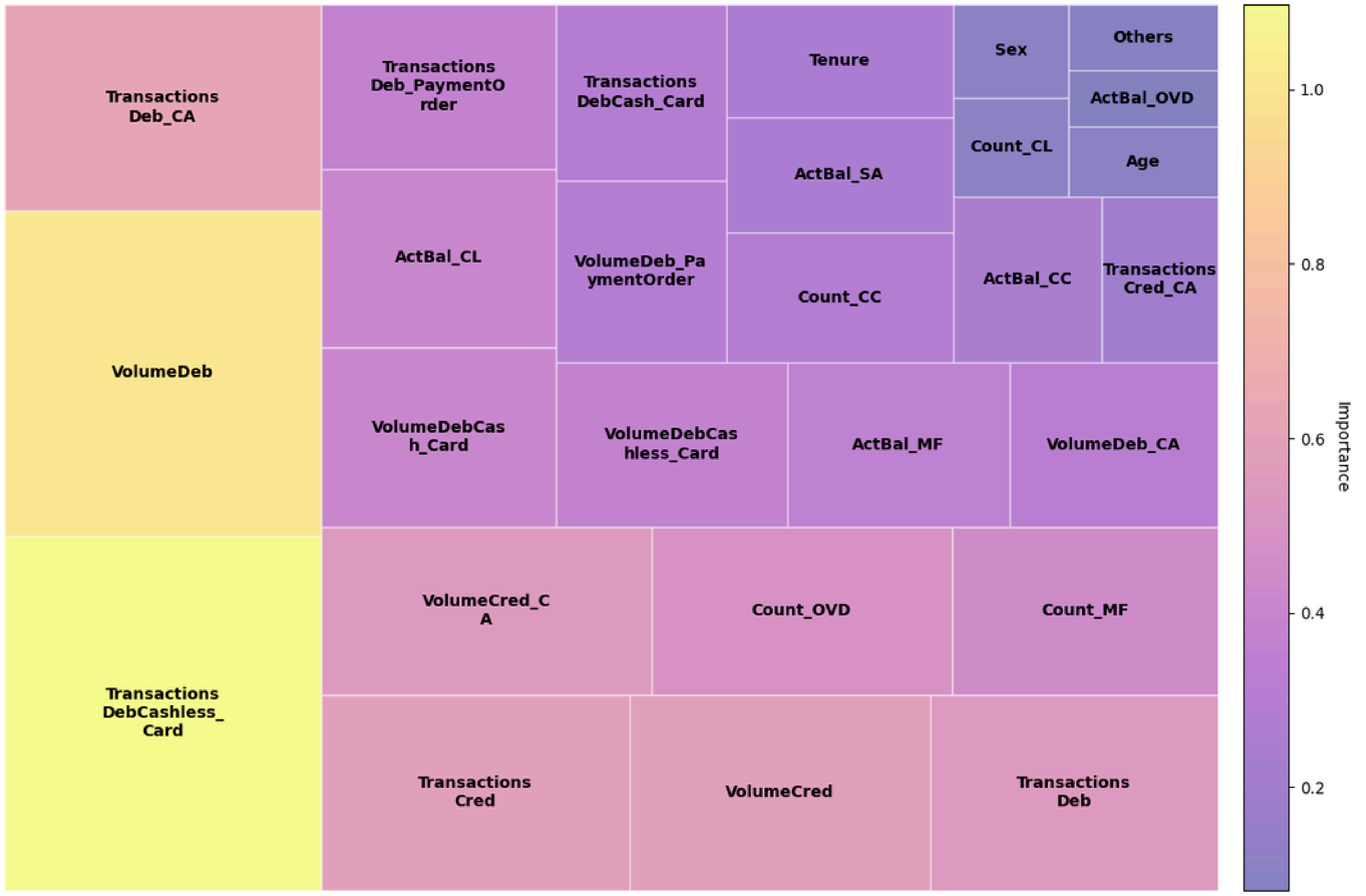
From the three dominant features, we can infer that clients who are highly likely to avail MF:

- prefer digital cashless transactions,
- disciplined or not huge spenders, which leads to
- lower total monthly debit volumes.







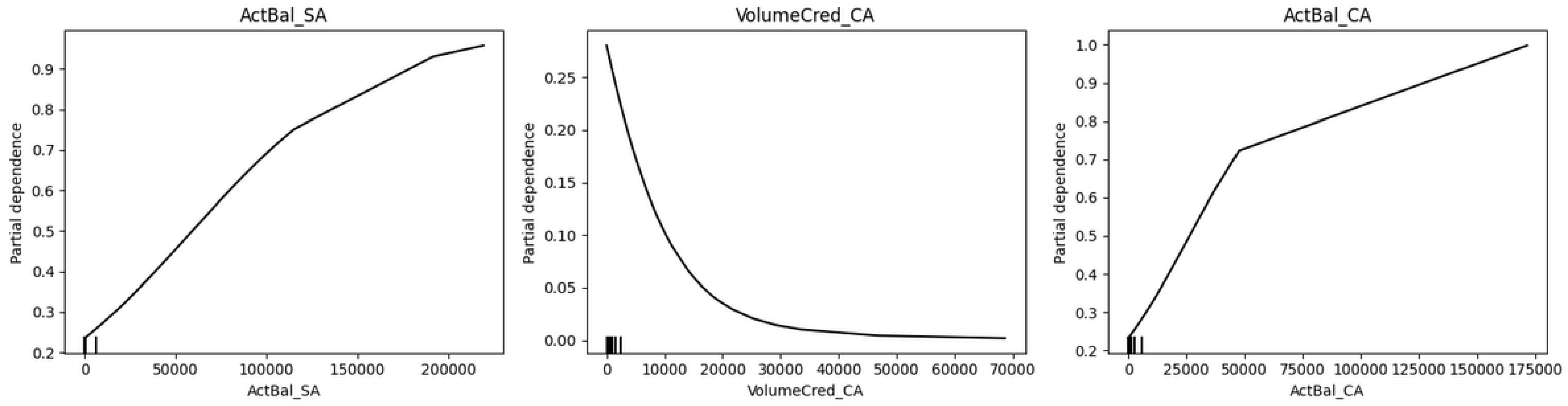


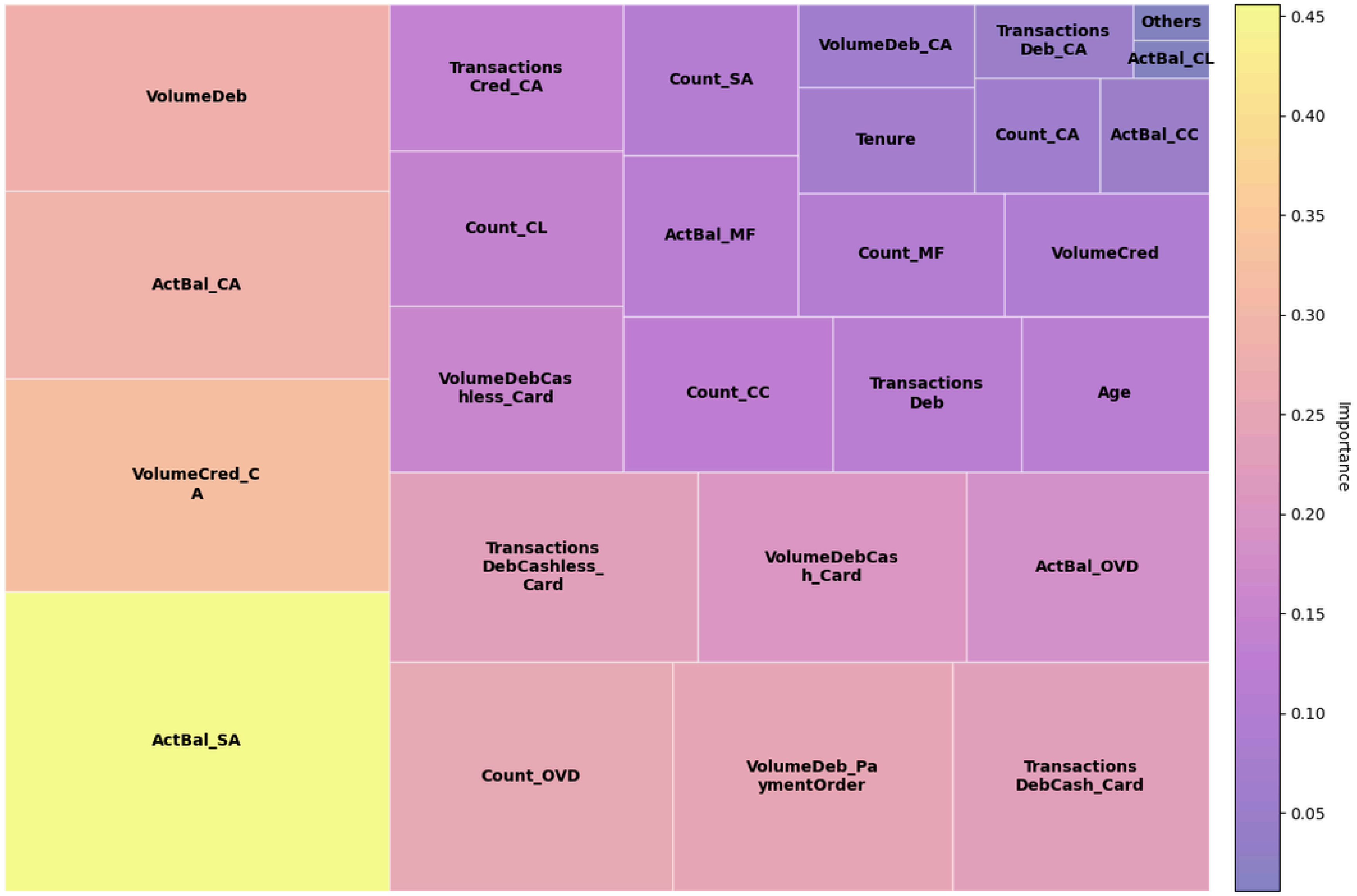
Identifying Clients

High likelihood clients for product **CC**

From the three dominant features, we can infer that clients who are highly likely to avail CC are those with:

- substantial savings account balances,
- lower monthly credit turnover on their CA, and
- healthy current account balances.



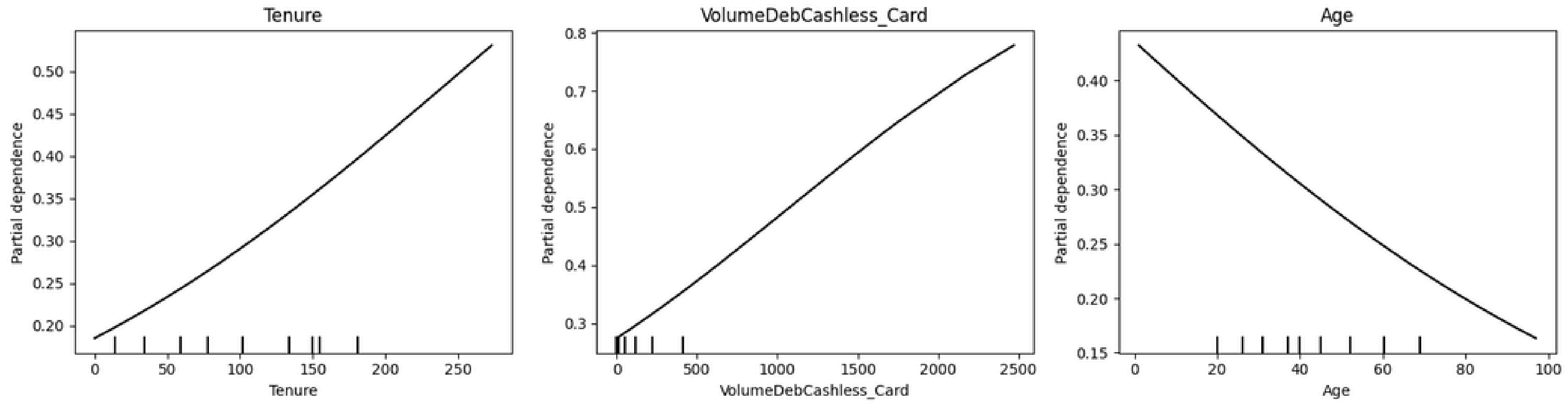


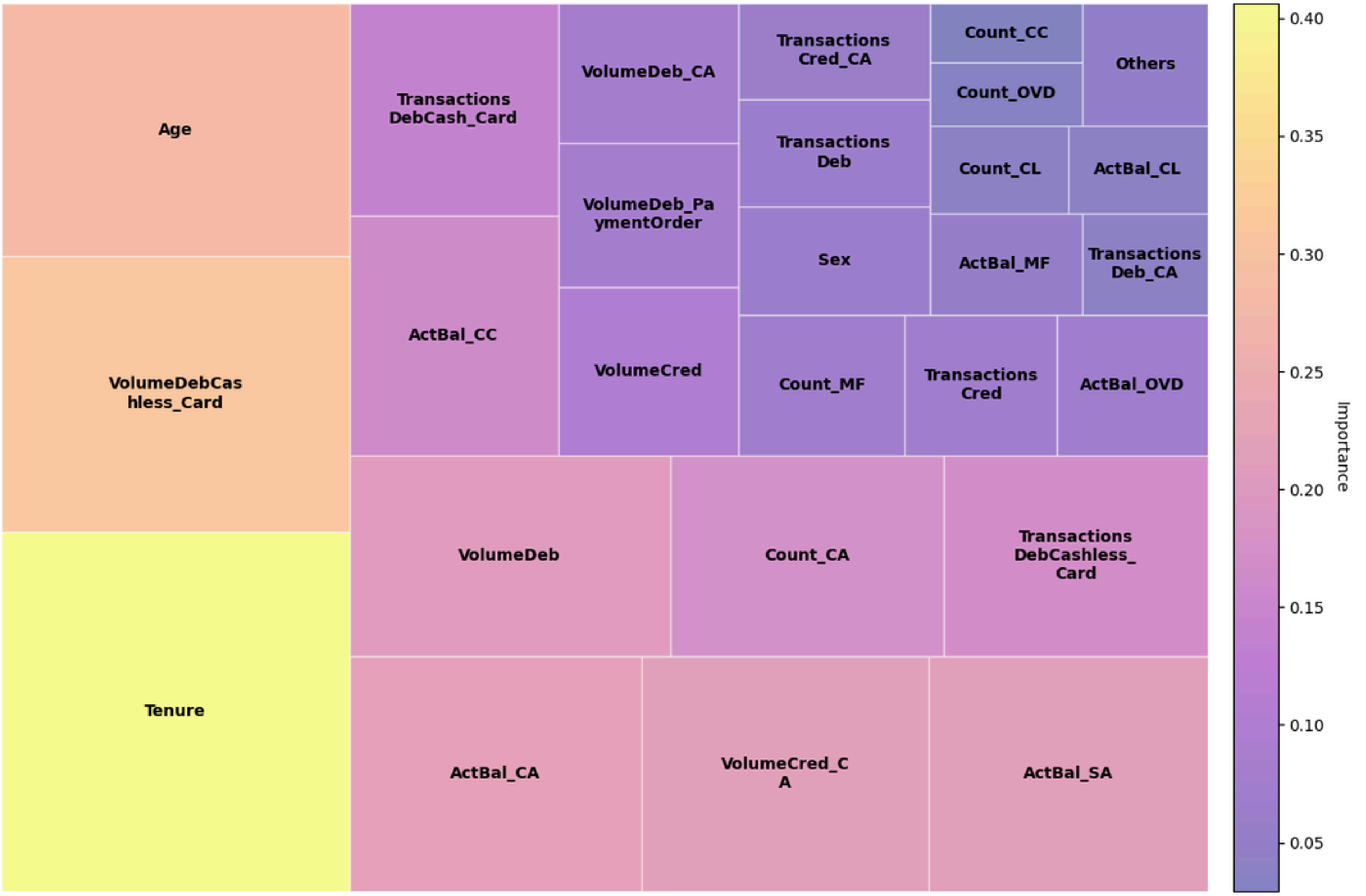
Identifying Clients

High likelihood clients for product **CL**

From the three dominant features, we can say that clients that are highly likely to avail CL are:

- mostly young clients,
- trusted and tenured customers, and
- have a high monthly volume of debit cashless transactions via card.



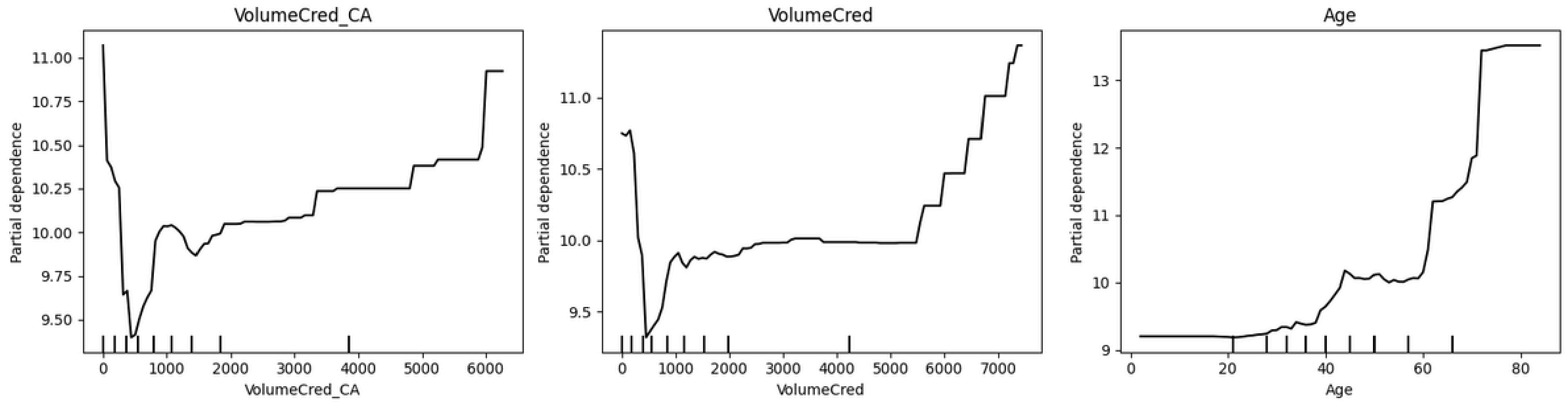


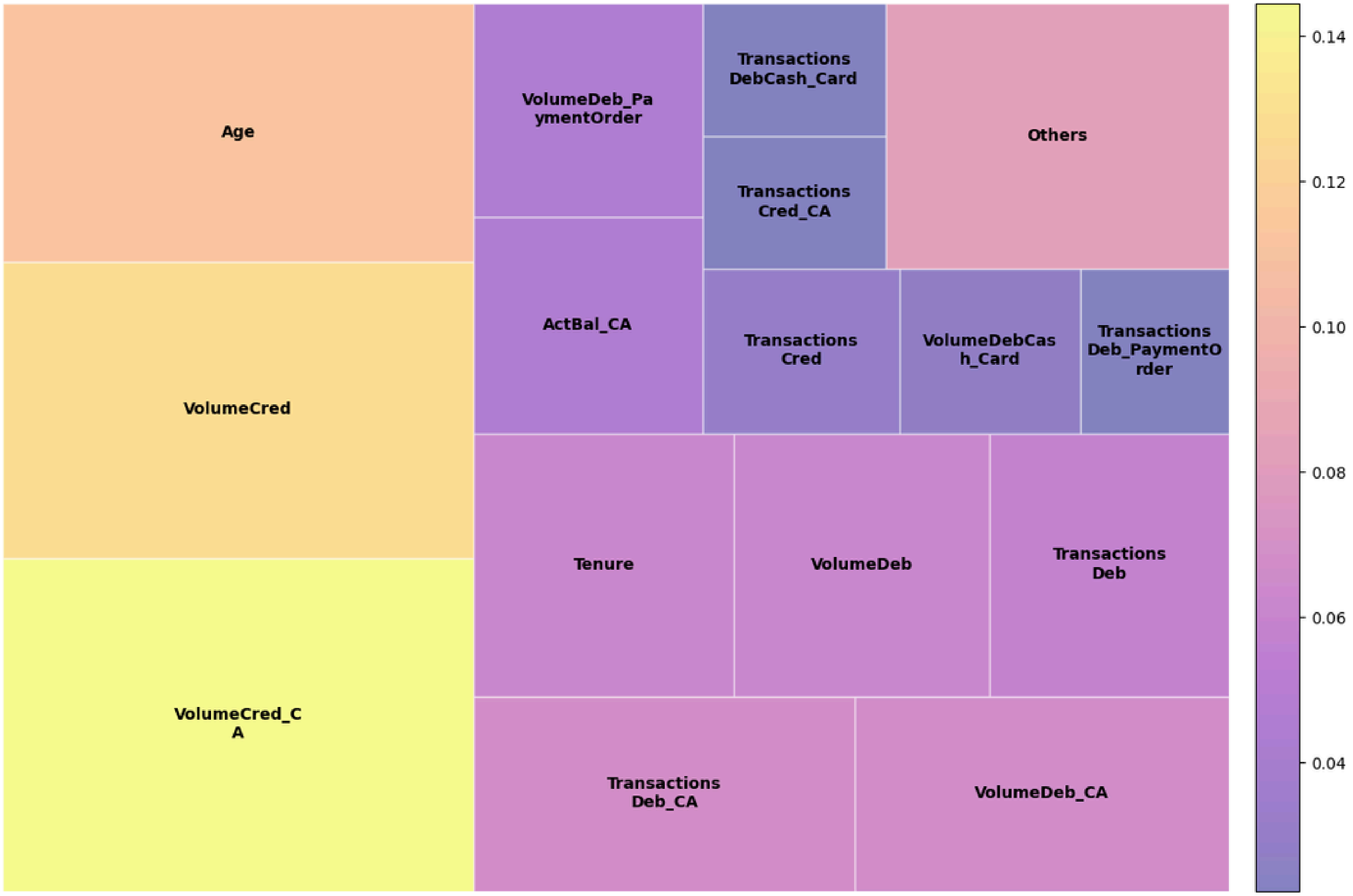
Key Drivers for Revenue

High revenue clients for product **MF**

From the three dominant features, we can infer that clients who yield high revenue for MF are:

- mostly 60 year old clients or older, and
- have high monthly credit turnovers.



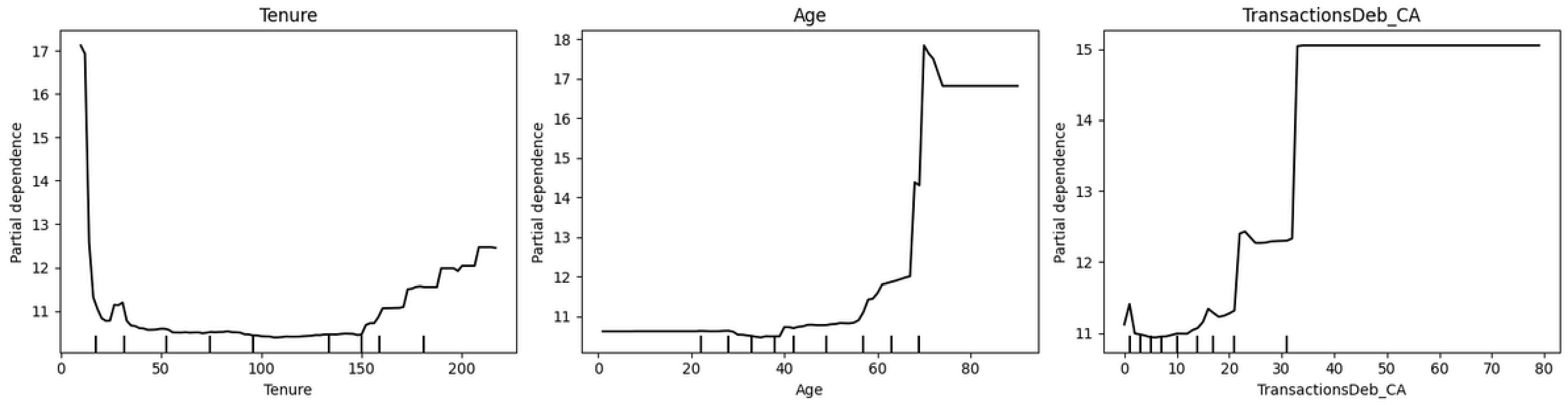


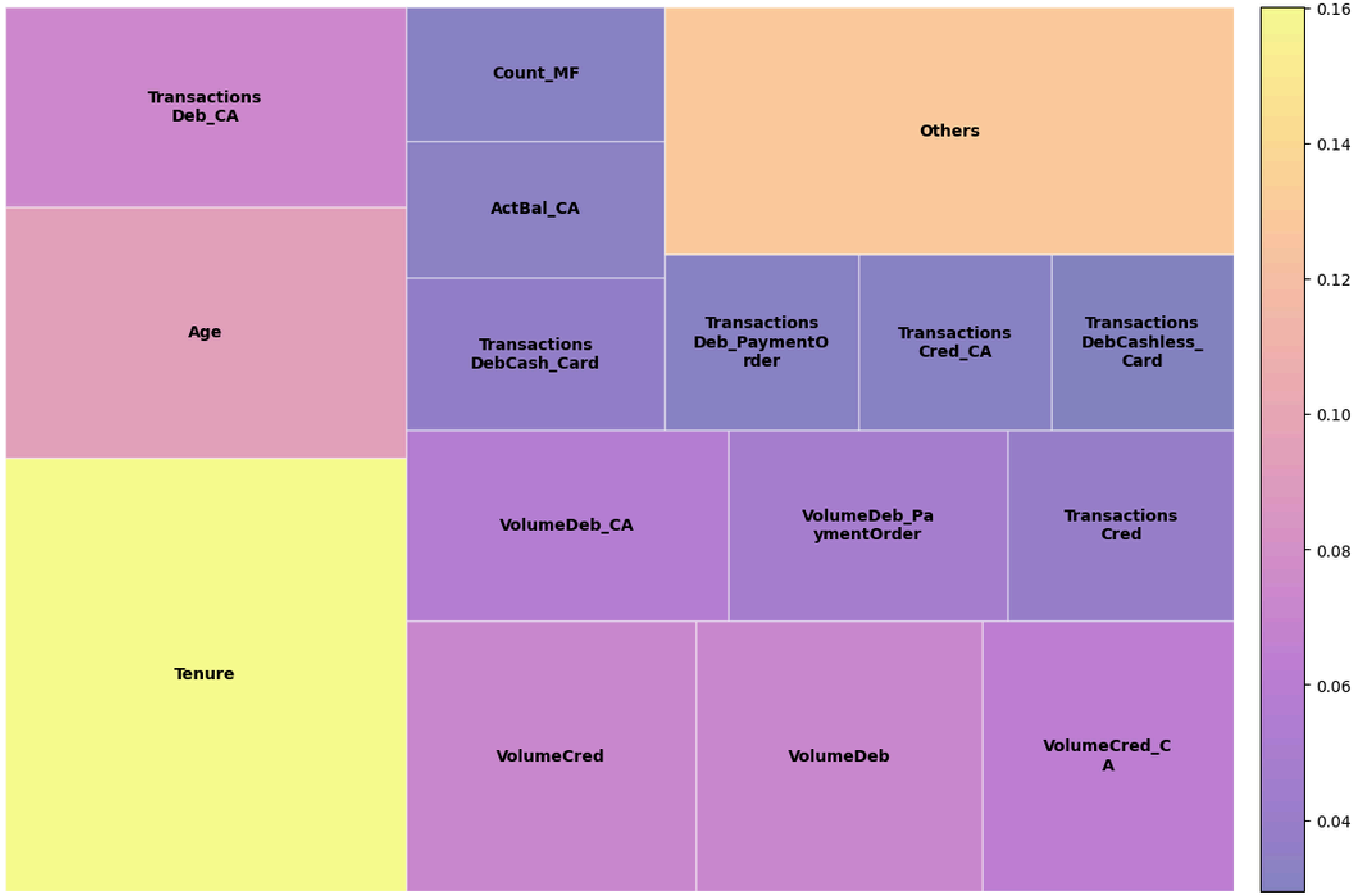
Key Drivers for Revenue

High revenue clients for product **CC**

From the three dominant features, we can infer that clients who yield high revenue for CC are:

- either new clients or tenured (~12 years) clients,
- of retirement age, and
- have more than 30 current account debit transactions.



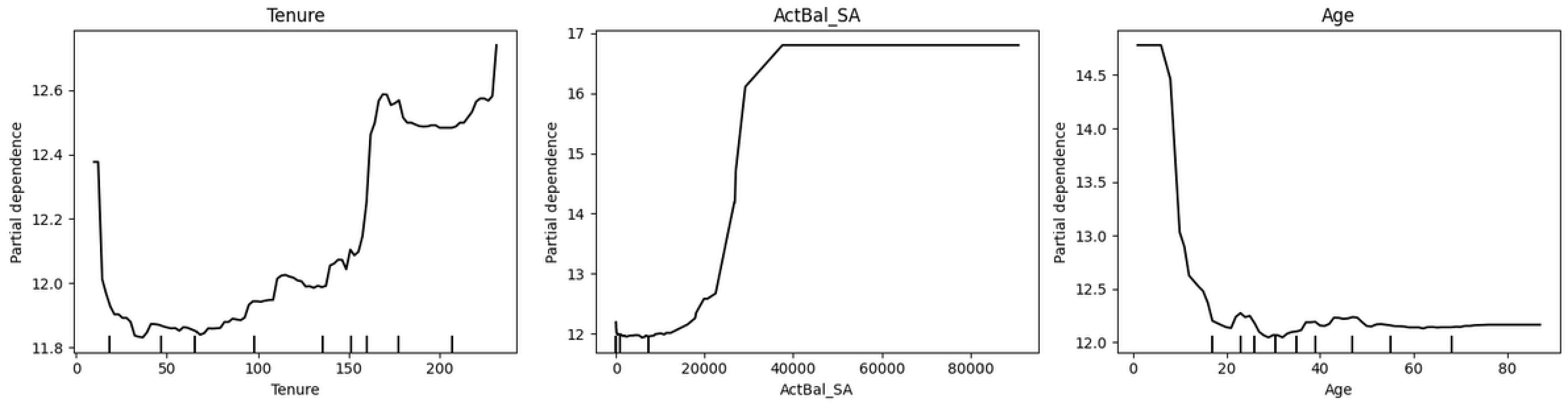


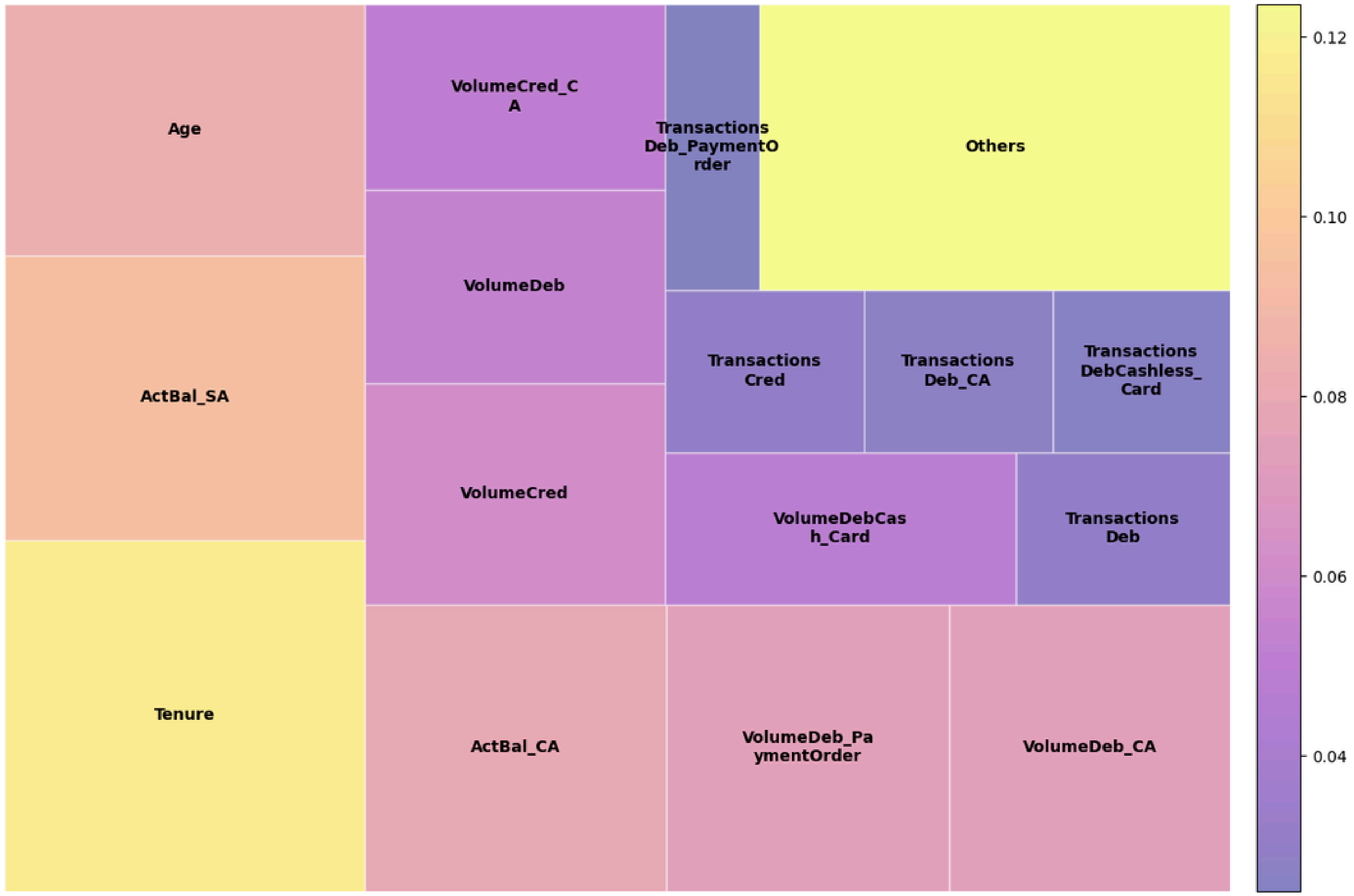
Key Drivers for Revenue

High revenue clients for product **CL**

From the three dominant features, we can infer that clients who yield high revenue for CL are:

- either new clients or tenured (8-20 years) clients,
- mostly 20-50 year old clients, and
- have large savings account balances (+€30K)





Summary

- Potential MF buyers prefer cashless transactions and are disciplined spenders
- Potential CC buyers have substantial savings and current account balances
- Potential CL buyers are young and loyal customers that are big spenders

- High revenue MF clients are retired customers with high monthly credit turnovers
- High revenue CC clients are either new or tenured, retirement age, and have a number of current account debit transactions
- High revenue CL clients are either new or tenured, 20-50 years of age, and have large savings account balances

Top 100 Likely Buyers: (0.99 - 0.51) €796 (€1192)

Top 100 Max Expected Revenue: (0.99 - 0.20) €930 (€1864)