



SCHOOL OF
SCIENCE &
TECHNOLOGY



RFM, Churn and Forecasting

Group K - MBD Oct 21

Meet The Team



Julian Krauth



Julius Prestin



Diego Scheibling



Quirijn Bolhuis



Diego Mata



Thomas Truyts



Thomas Van den Borg

Agenda

■ Business overview

Get an understanding of ClientCo and the Scope of Work

■ Hypothesis

Identify the main challenges and their potential sources

■ EDA

Explore transactional dataset and the distribution of the features

■ Analytics

RFM, Retention, CLV analysis
Churn prediction

■ Recommendation

Find key customer groups, their KPIs and how to target them

■ Implementation

Timeline for next steps of the transformation project

Challenges: Weak online store, liquidity shortages and decreasing market share in Western Europe as well as South East Asia make rise for a need of transformation



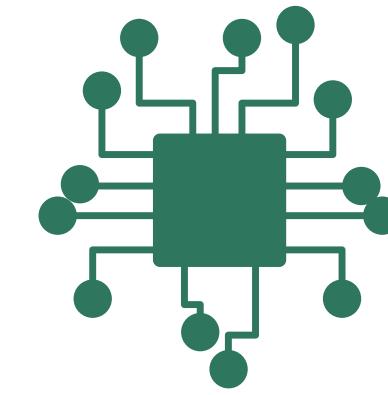
Liquidity shortage



Regulations



Currency



Digitalisation

Hypotheses

Portfolio is
too large

Channels are
wrongly
handled

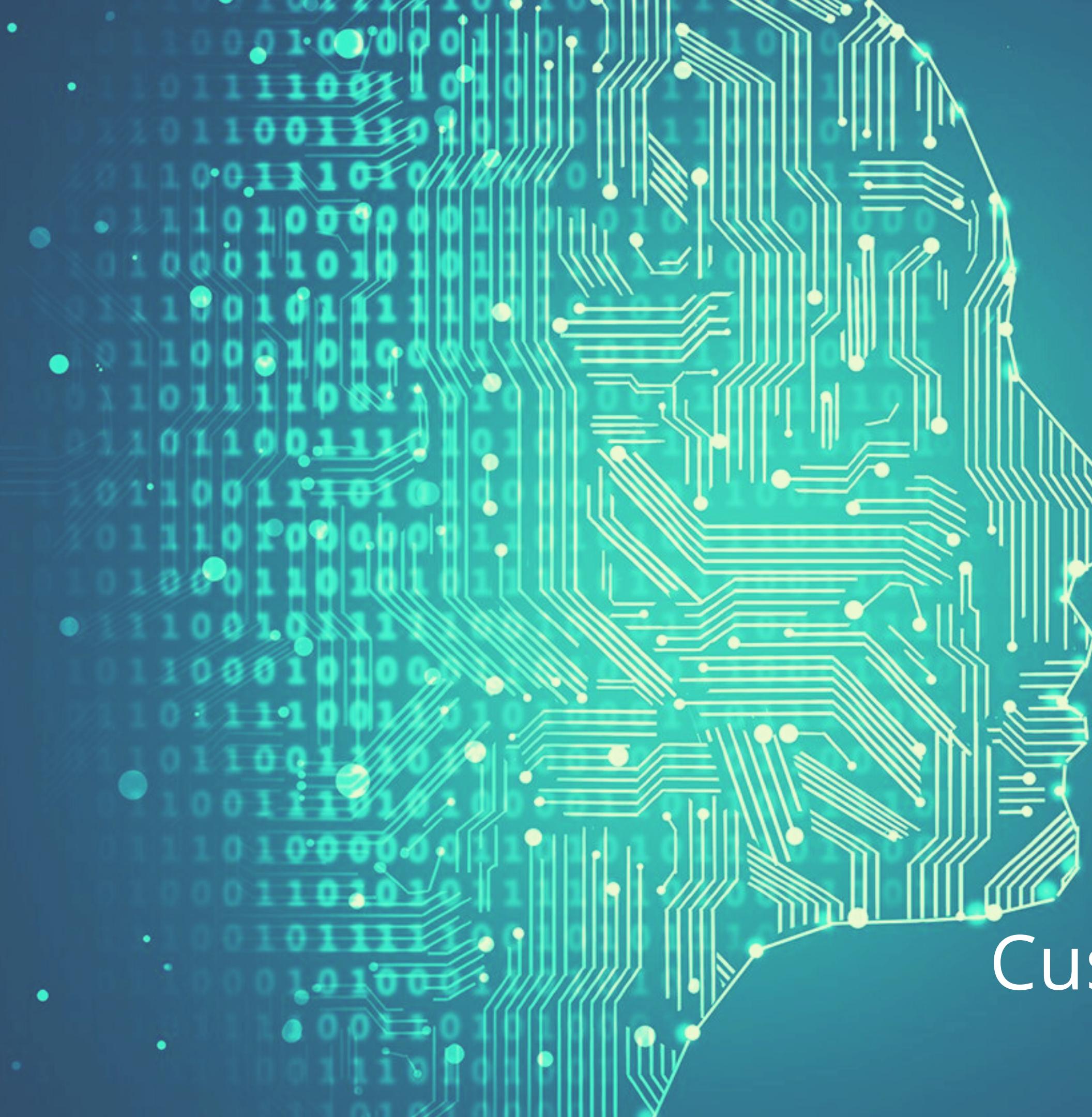
Branches are
operating
inefficient

Dataset: Size of the dataset with anonymised and time selective data is going to be the main challenge of this project





THE MODELS
RFM
Retention Curves
Propensity model
Customer Lifetime Value



RFM for segmentation: Calculating a R, F, M and combined metrics in order to put customers into segments that can be targeted differently

Scoring

Recency Frequency Monetary



Champions

High Value

Medium Value

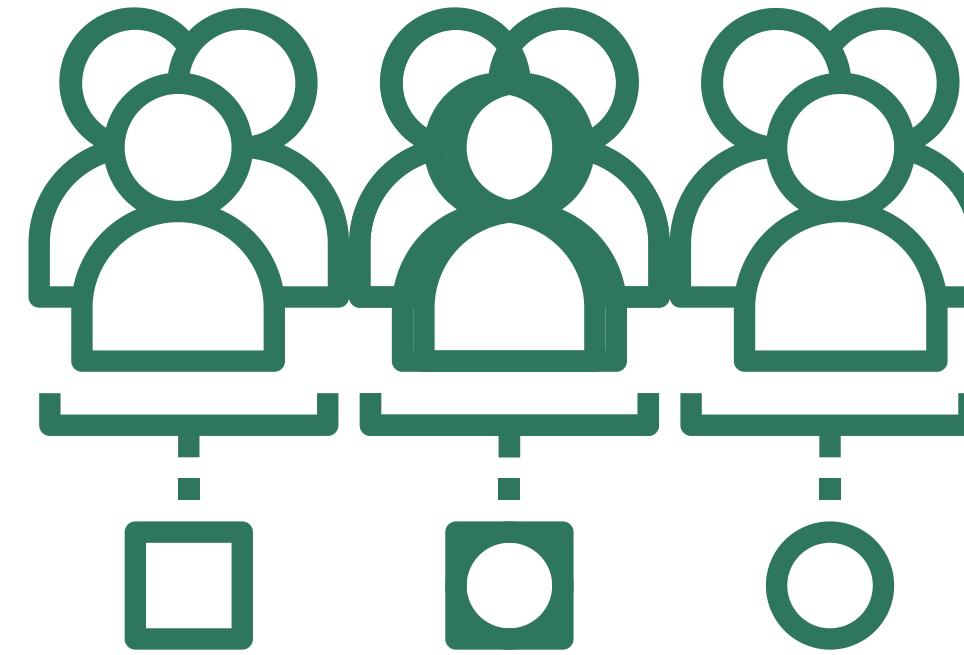
Low Value

Lost

Churn ratio per customer segment

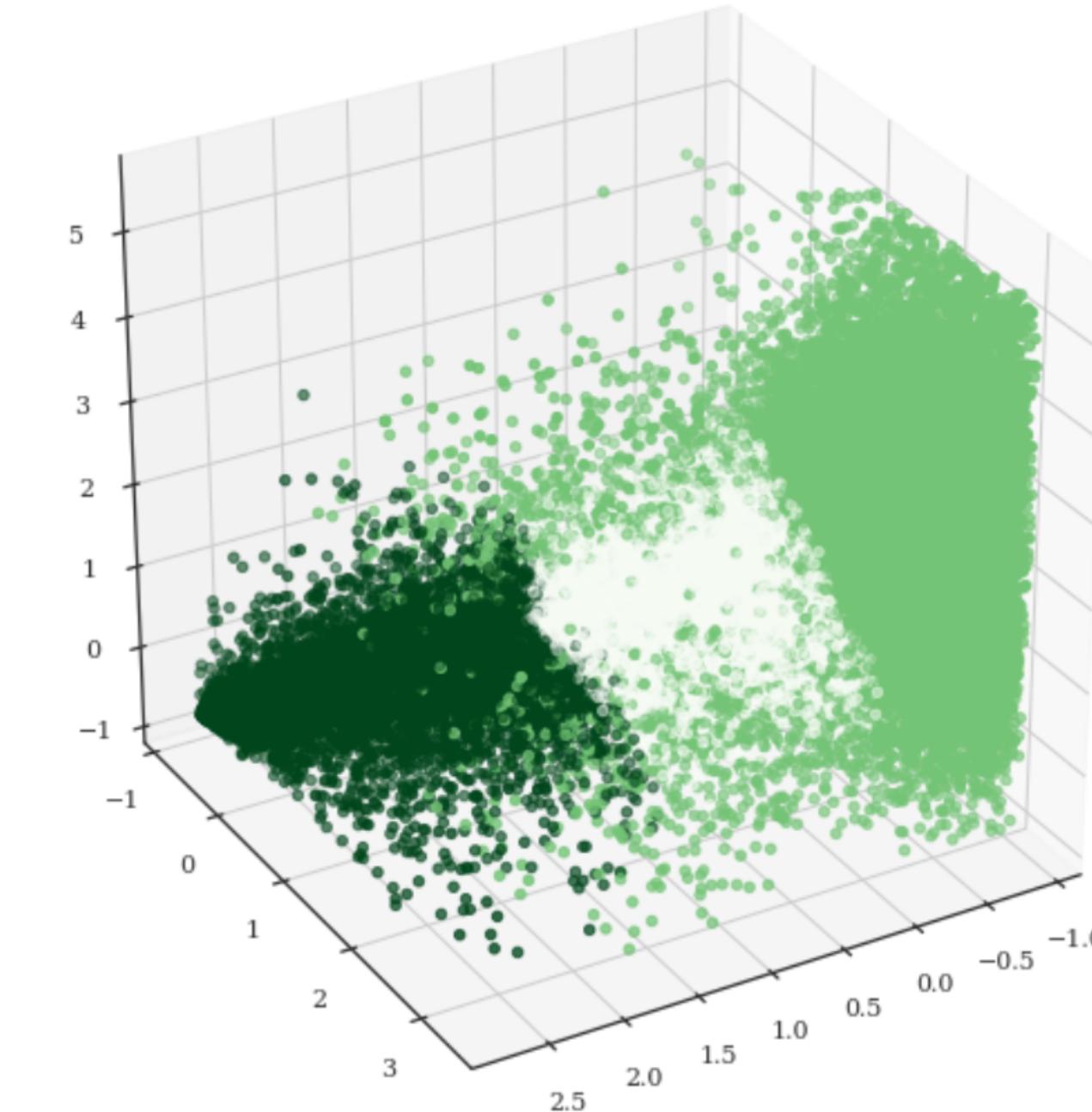


RFM for Clustering: Applying KMeans clustering to the clients and their RFM scores results in 3 different client groups



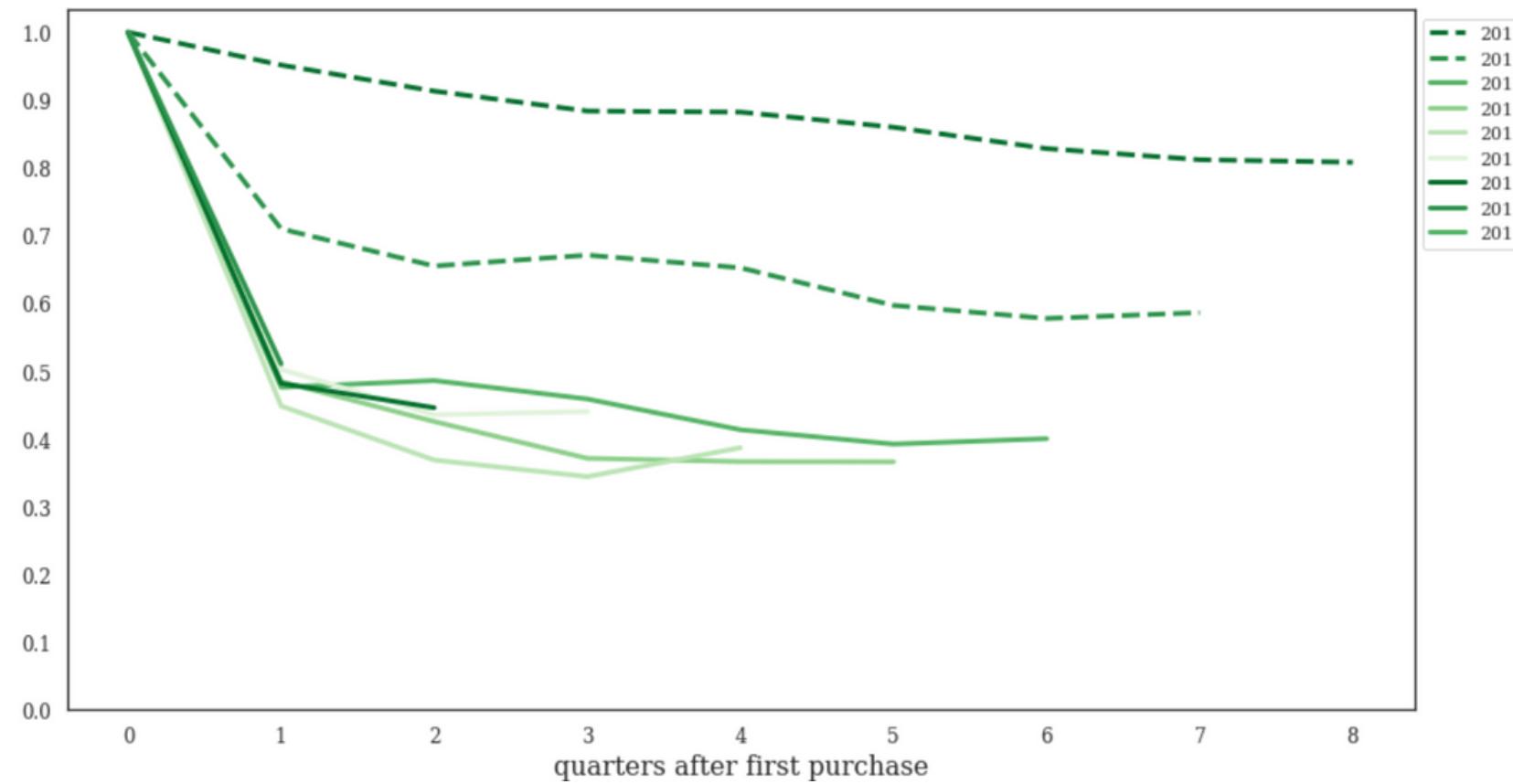
KMeans

- ✖ Remove outliers
- ↗ Scale values
- ❸ Optimal number of clusters
- ÷ Silhouette = 0.5

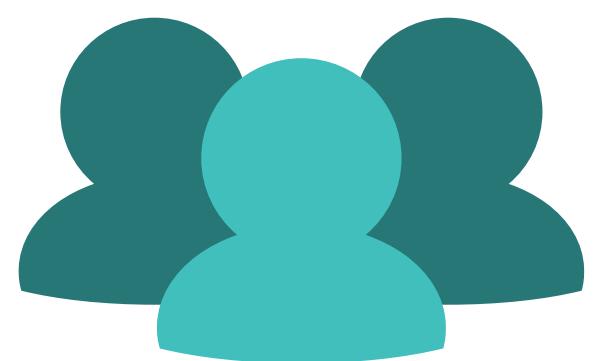
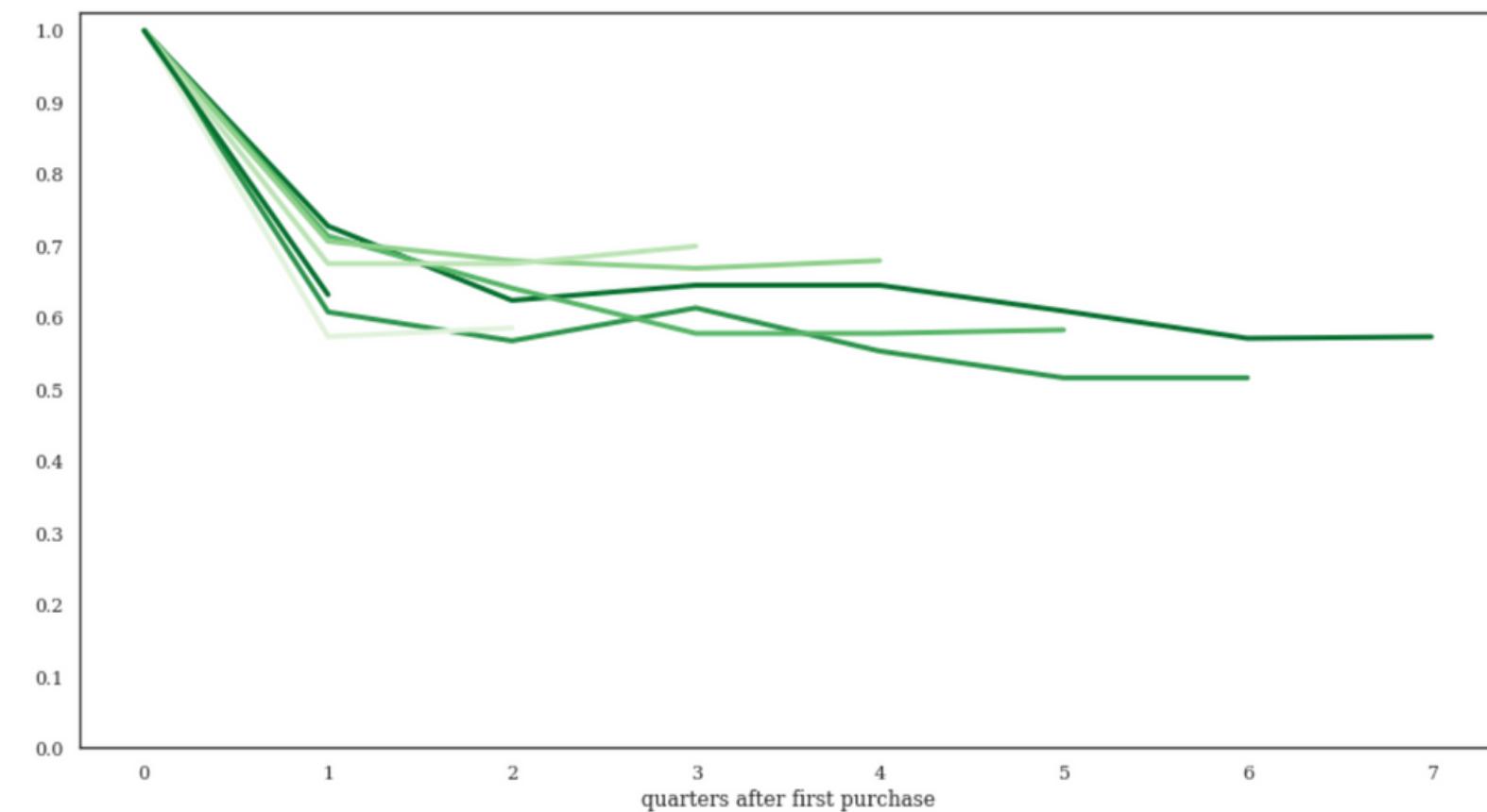


Retention Curves: Overall the clients show an average retention rate after one year is 40% - customers who buy online have a significantly higher rate of 60%

All clients



Mostly online



CLV: Customer Lifetime Value has been calculated for three different models using different assumptions about the profit margins and the acquisition costs



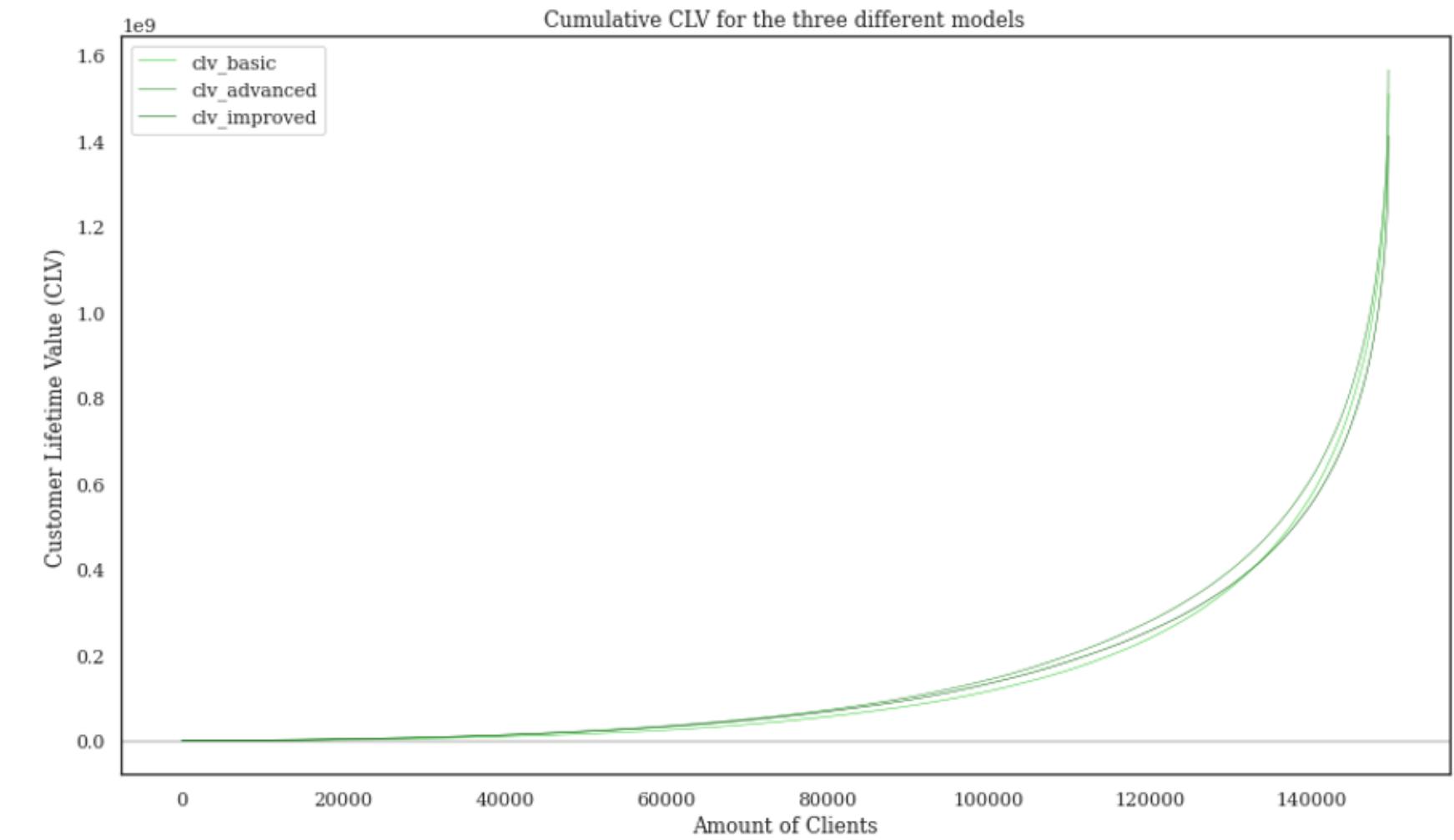
Basic, advanced and improved



Advanced: adapted separate profit margins

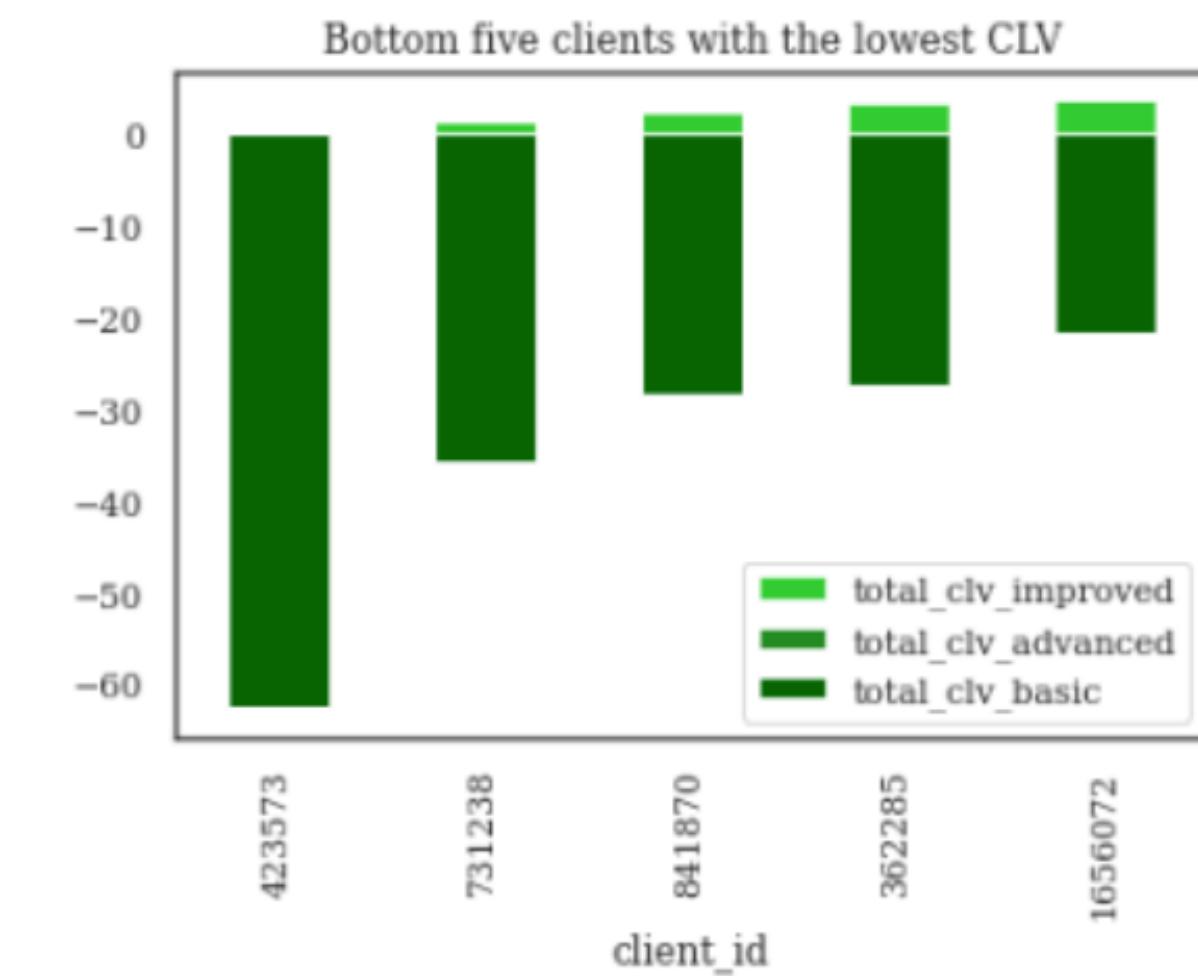
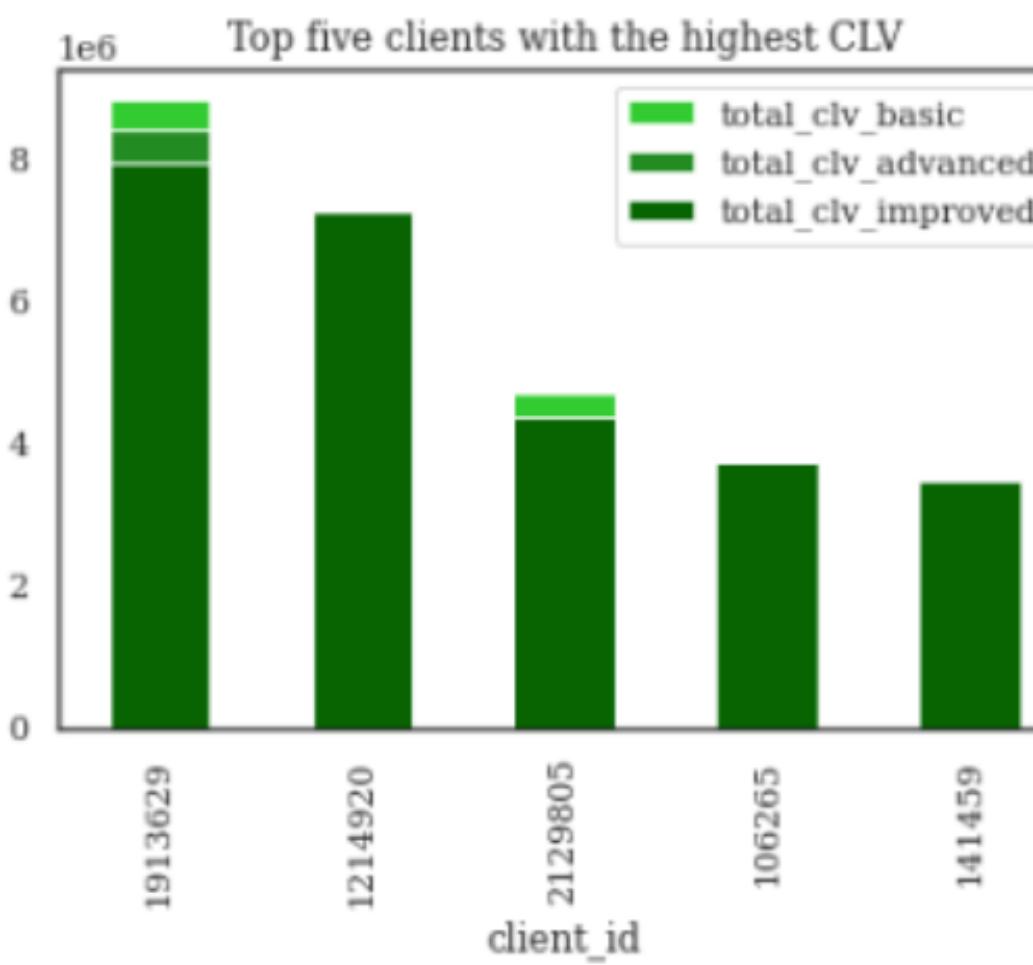
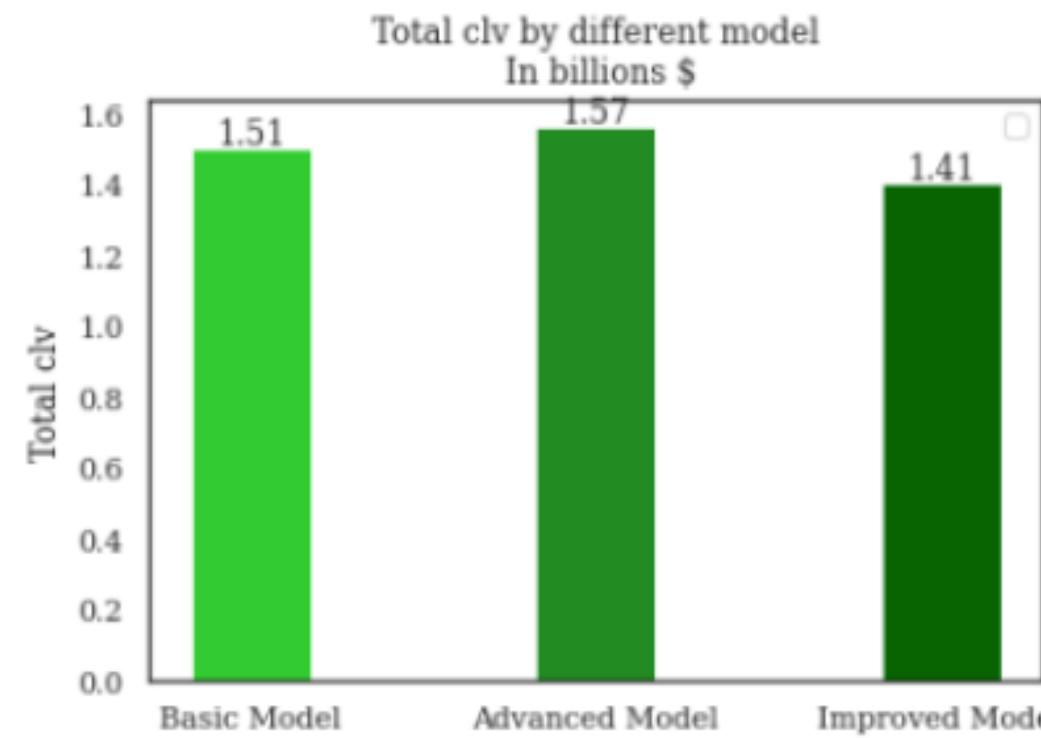


Improved: adapted separate acquisition costs



$$CLV = \frac{\text{profit margin} * \text{average lifetime}}{(1 + d)^{\text{average lifetime}}} - CAC$$

CLV per client: The negative values that are found for the bottom five clients can be explained by product returns



Propensity Model (Labelling): The thresholds for the propensity model are decided based on domain knowledge and analyses of ClientCo's product catalogue



Product categories



Negative prices (3,77%) dropped



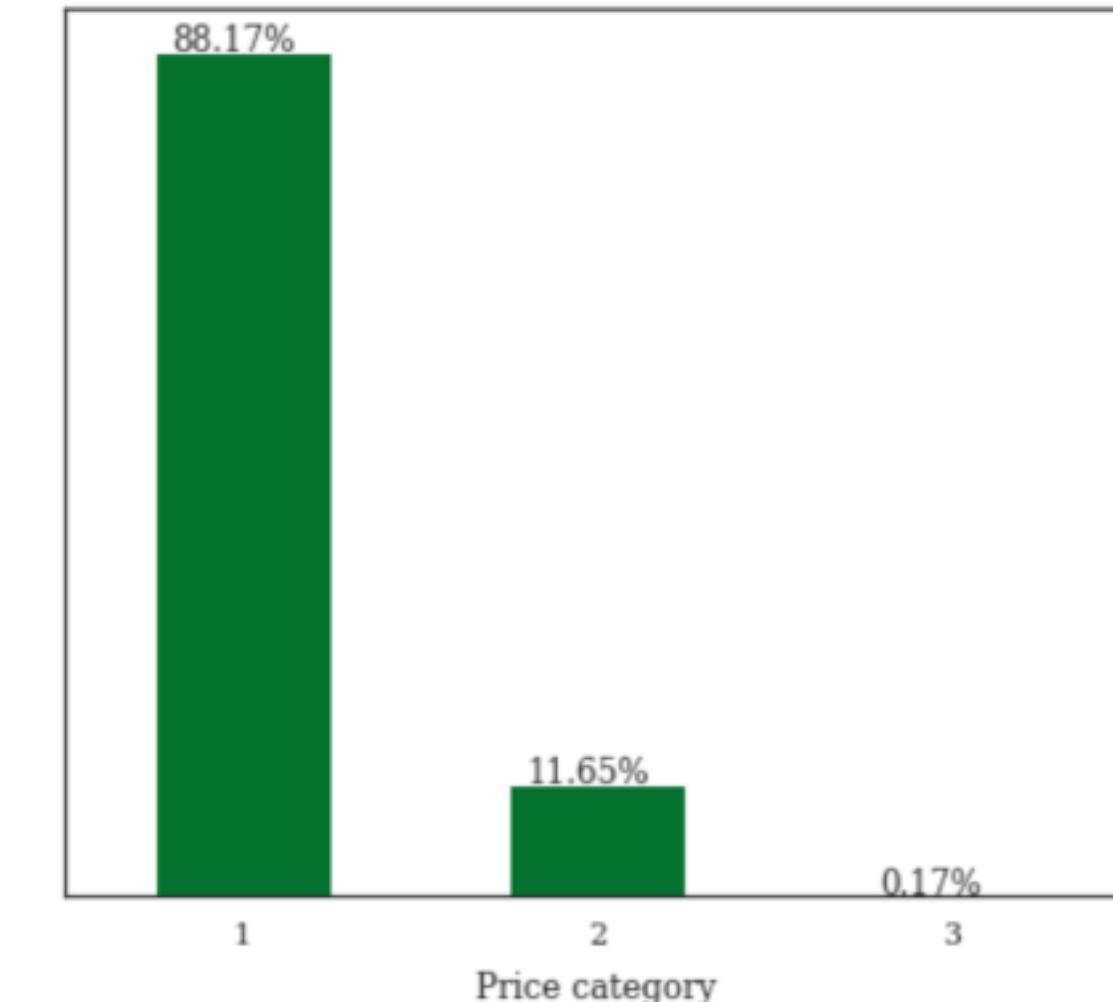
Average timespan per client per product category

$$AT_{ci} = \frac{LP_{ci} - FP_{ci}}{purchases_{ci} - 1}$$

- 97,5th percentile per product category
 - Threshold-product1 = 99 days
 - Threshold-product2 = 270 days
 - Threshold-product3 = 450 days

Product categories	Price ranges
Product category 1	0 < x <= 50
Product category 2	50 < x <= 1000
Product category 3	1000 < x <= 500000

Distribution of price categories
(by number of transactions)



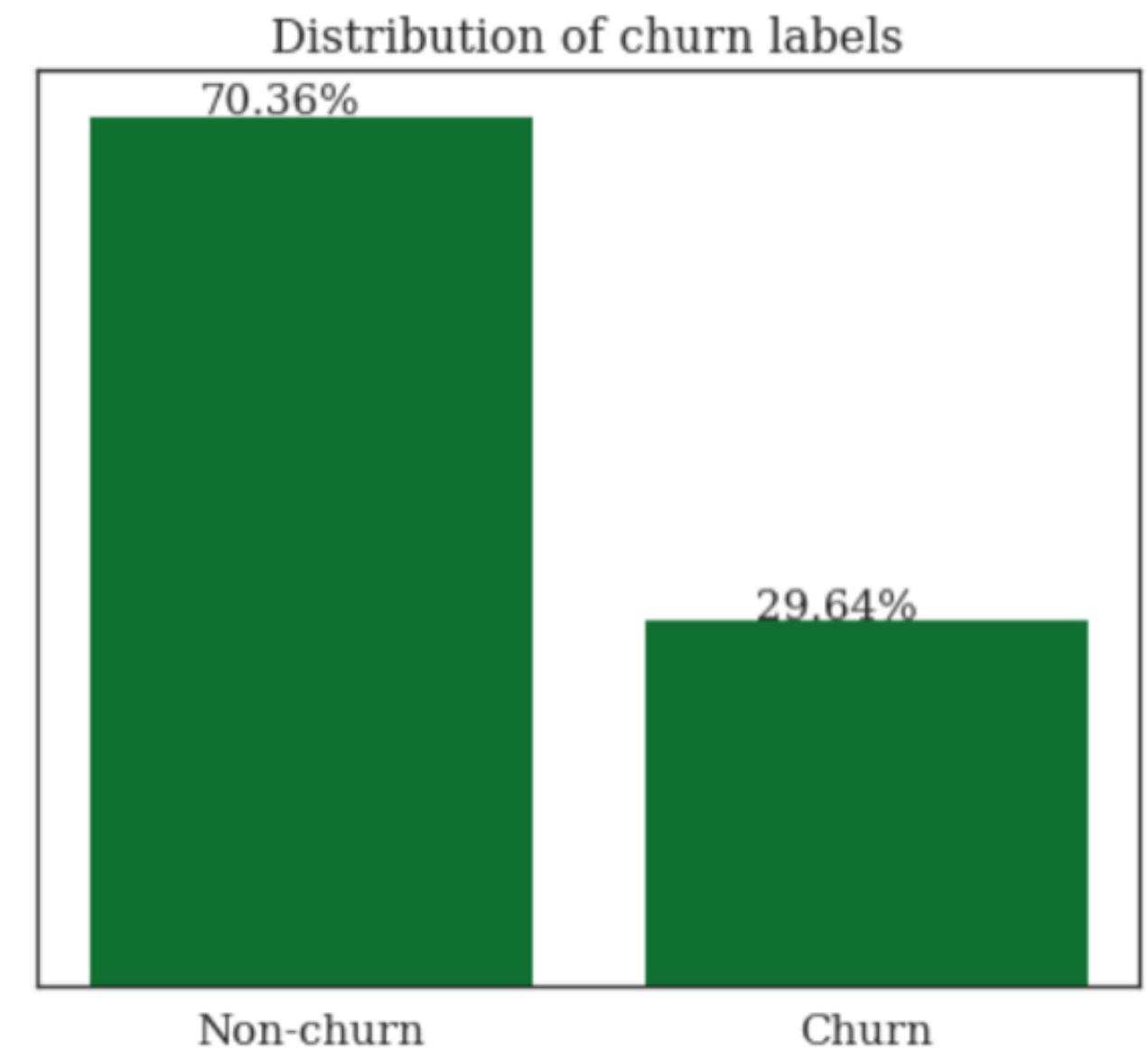
Propensity Model (Labelling): The extra variable *past_churn* is calculated as the number of times a client exceeded the threshold in every product category before his last purchase



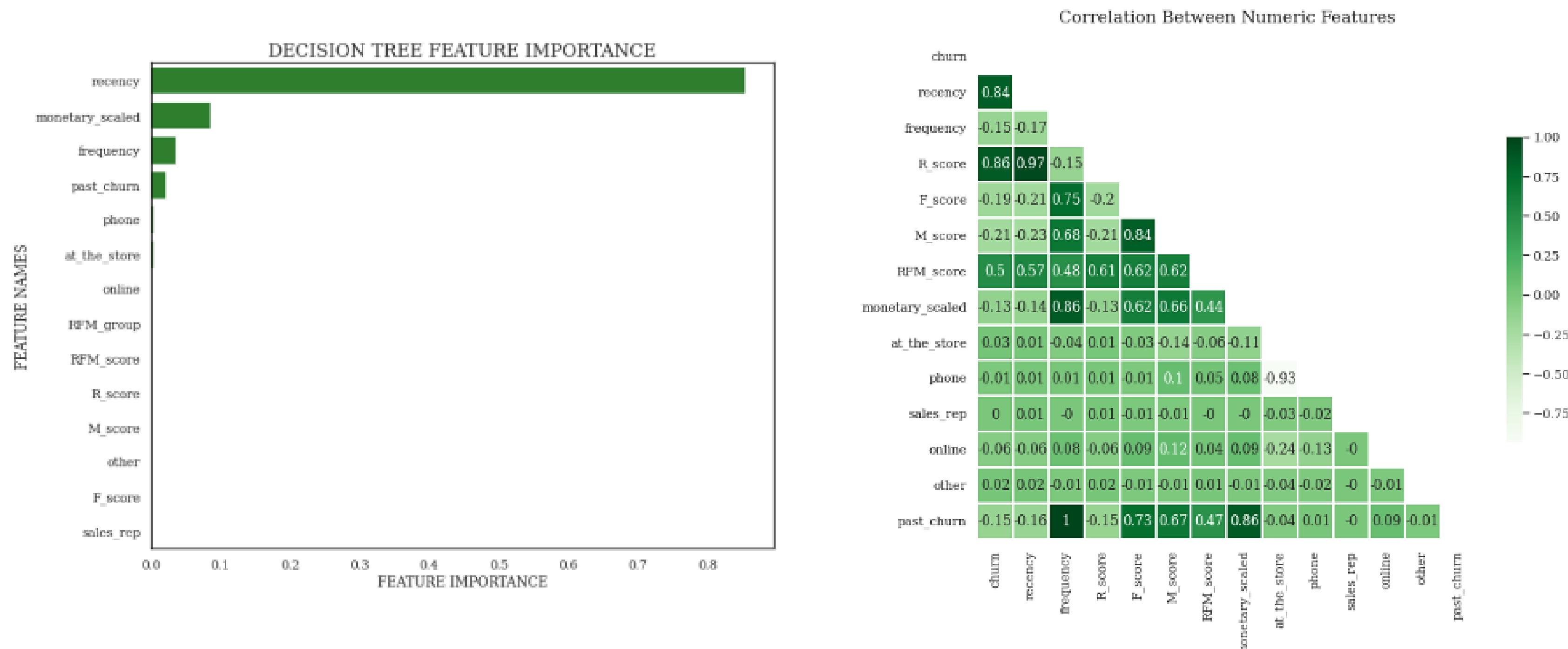
Churn for client per product category
(22/09/2019 - last purchase) > threshold



Preventing information leakage with the extra variable *past_churn* looking only into the past



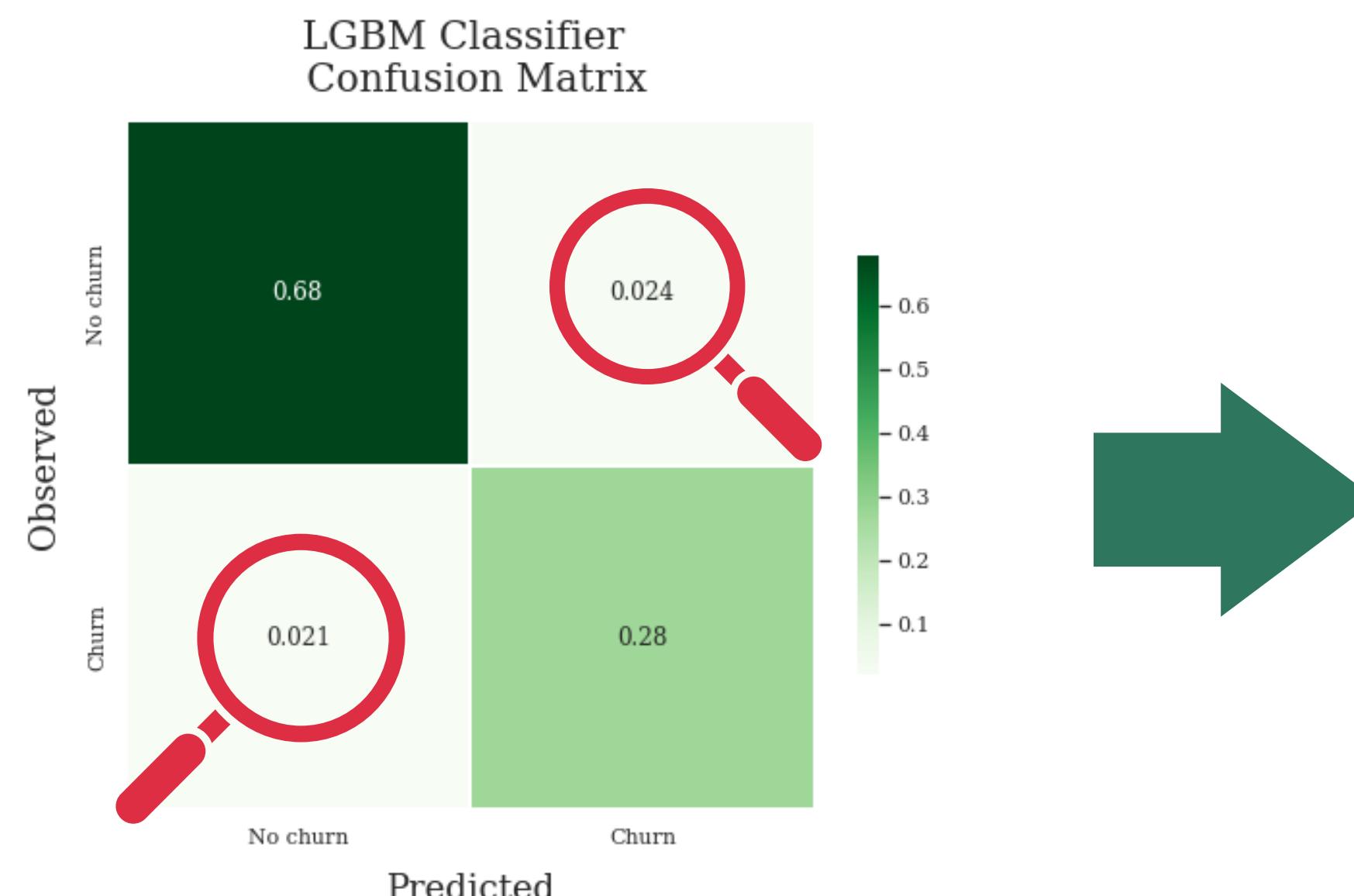
Propensity Model (Features): After analyzing the fourteen different features before applying the machine learning algorithms, no missing values were found



Propensity Model (Selection): After applying different classification models LGBM is selected with an Accuracy of 95.5% on the randomly selected test split (20%)

Model	Accuracy	Precision	Recall	F1-Score
Decision Tree	94%	93%	93%	93%
Decision Tree + PT	94%	93%	93%	93%
Random Forest	95%	94%	94%	94%
Extra Trees	95%	94%	94%	94%
Gradient Boosting	95.4%	94.2%	95.0%	94.6%
Gradient Boosting Scaled	70%	35%	50%	41%
LGBM	95.5%	94.5%	94.8%	94.7%
LGBM with SMOTE	95.5%	93.7%	95.4%	94.5%

Propensity Model (Prediction): Looking at the misprediction the most features that deviate the most seem to be frequency and past_churn



	mean_wrong	mean_correct
recency	183.402994	144.788215
frequency	63.343731	407.511482
R_score	2.023082	1.594347
F_score	1.011853	1.157845
M_score	1.019339	1.183837
RFM_score	4.054273	3.936029
monetary_scaled	0.000380	0.002199
at_the_store	0.645041	0.628762
phone	0.333125	0.337077
sales_rep	0.001248	0.000599
online	0.019963	0.032553
other	0.000624	0.001009
past_churn	12.489083	122.039745

A faint, semi-transparent background image shows a construction site with several large tower cranes and industrial buildings under construction. One building has the word 'SWECO' visible on its facade. The overall color palette is muted, with various shades of grey and teal.

RECOMMENDATIONS
KEY CHALLENGES
NICHES
PROJECT TIMELINE

Recommendations: Two-fold strategy of restructuring phase and a strategy refreshment phase to overcome current internal and external key challenges

Key Challenges



Liquidity



Digital platform



Regional presence



Product assortment

ClientCo 2.0



1 Restructuring

A Product Portfolio Optimization

Strengthen the core body of the business and make it more resilient and agile

B Branch Network Streamlining

Optimize the supply chain based on its ability to serve main customers and future channels

2 Refreshment

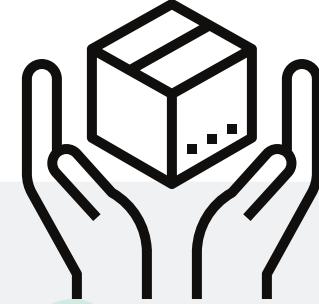
A Conservative Core Segment

Consider the characteristics and preferences of the actual customer base

B Digital Future Segment

Consider the emerging trend of digitization with customers predominantly ordering digitally

Restructuring phase: Strengthen the core body of the business and make it more resilient and agile for the upcoming changes inside and outside the organization



Product Portfolio

Downsize non-performing units in order to encounter the **liquidity problem** and the upcoming **assortment reviews**

Branch Network



Filters

- **Sales < USD 500 & quantity < 10 units**
- **Churned sales > 33%**
- **Days Sales Outstanding > 90 days**

- **Sales < USD 5 m & quantity < 100k units & online channel coverage < 50%**
- **Sales < USD 10 m & churned sales > 20% & online channel coverage < 50%**

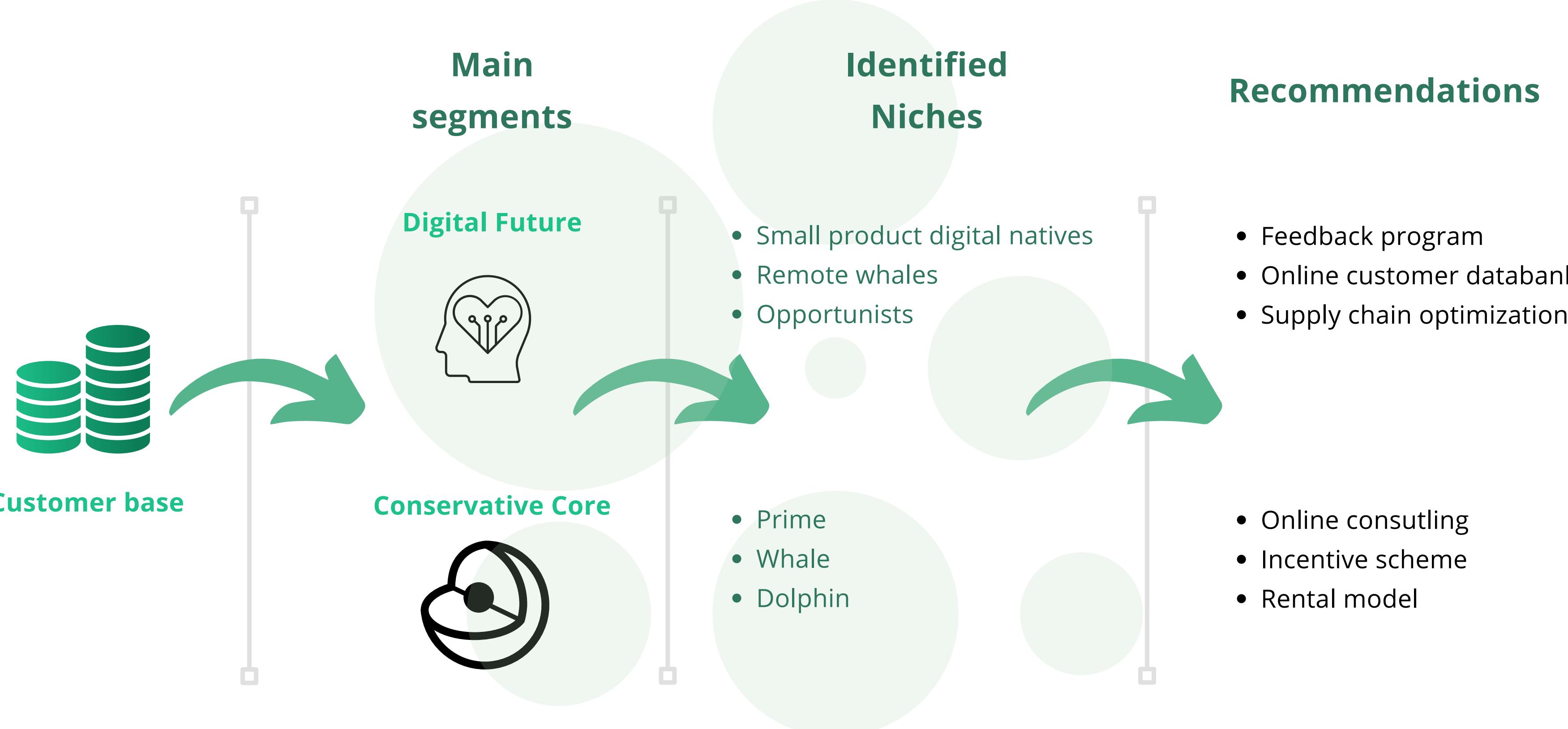


Outcome

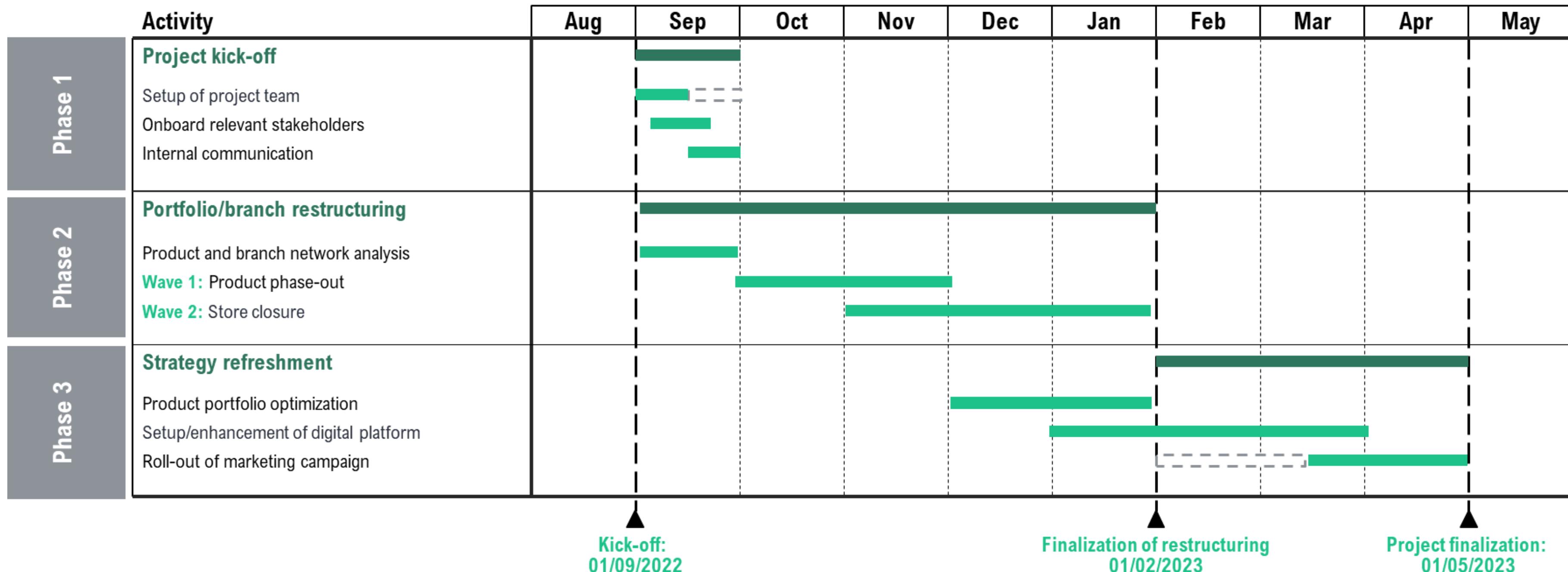
167k products identified for removal that contributes **USD 124 m sales** and **38 m units sold**, thereof **USD 38 m churned sales**

58 branches identified for closure that contributes **USD 38 m sales** and **26 k units sold**, thereof **USD 16 m churned sales**

Strategy refreshment: Two-step approach to identify the most important customer segments that need to be prioritized when reallocating the strategic product focus



Implementation timeline: Three main stages project kick-off, restructuring phase, and strategy refreshment to be implemented within 9 months



**Thank you for your attention! Please ask any
question you want.**





APPENDIX

EDA - Clients: Exploratory Data Analysis (EDA) of the used dataset to further understand the features behaviours and insights over time (1/3)

- No cluster risk
- Top 10% generate 75% revenue and 77% quantity
- Many clients with small turnovers
- Phone and store most used sales channel
- Phone the most important sales channel



EDA - Products: Exploratory Data Analysis (EDA) of the used dataset to further understand the features behaviours and insights over time

- Price range \$0 - \$435k
- Top 10% net sales accountable for 90% total revenue
- Top 10% quantity accountable for 99% total items dispatched
- Majority of turnover generated by product category 1
- Average YtY price increase 2.99%
- Net sales increase of 4.94%



EDA - Branches: Exploratory Data Analysis (EDA) of the used dataset to further understand the features behaviours and insights over time

- Top 10% generate 36% revenue stream and 41% quantity
- Top 10% generate 36% revenue stream and 41% quantity
- 103 branches generate 50% revenue stream
- Lower valued products sold through these branches
- 58% of top clients are customer in top branches

