FIELDWORK PROJECT 2025:
CRM REACTIVATION

OUR TEAM:

Roberto Vanzetta

Helena Borges Daré

Ignacio Kreis

Jacopo Abete

Jasper von Pachelbel

Master Academic Year (2024/2025)

Reactivation of Inactive Customers can help leverage Economic Potential

OUR CUSTOMER & CHALLENGE

- The customer, a leading European company in manufacturing and sales of tools and for the building a crafting sector
- Significant proportion of customers are **inactive**, with no purchases in the past two years
- Inactive customers are **unrealized economic potential** for the company

EXPECTED OUTCOME

- Development of a **reactivation prediction model** which identifies customers with the highest potential of return
- Providing the company with a starting point to take targeted actions for reactivation and to **maximise conversion**
- CRM reactivation

Understanding CRM

CRM = Customer Relationship Management

includes all measures to build, maintain and **reactivate** customer relationships.

CRM Reactivation

to retarget customers which did not purchase since 2 years.
based on data analysis -> to select the customers with the **highest probability of return** and the **highest economic potential**.

Our Analytical Strategy follows 5 Steps

1 Data Understanding & Exploration
- Understanding of columns
- Analysis and handling of missing values

2 Feature Engineering
- Merging Datasets
- Creating new insightful variables

3 Modeling
- Identification of patterns for active/inactive customers

4 Model Evaluation
- Technical criteria's
- Choice of best model

5 Business Insights
- Translation of findings into valuable business strategies

1. Data Understanding & Exploration

Two datasets consisting of >2 million entries provide the basis for analysis

TOOL_SALES (Sales Data 2017 – 2021)

- Purchase History of customers

→ Allows Identification of purchasing behaviour patterns

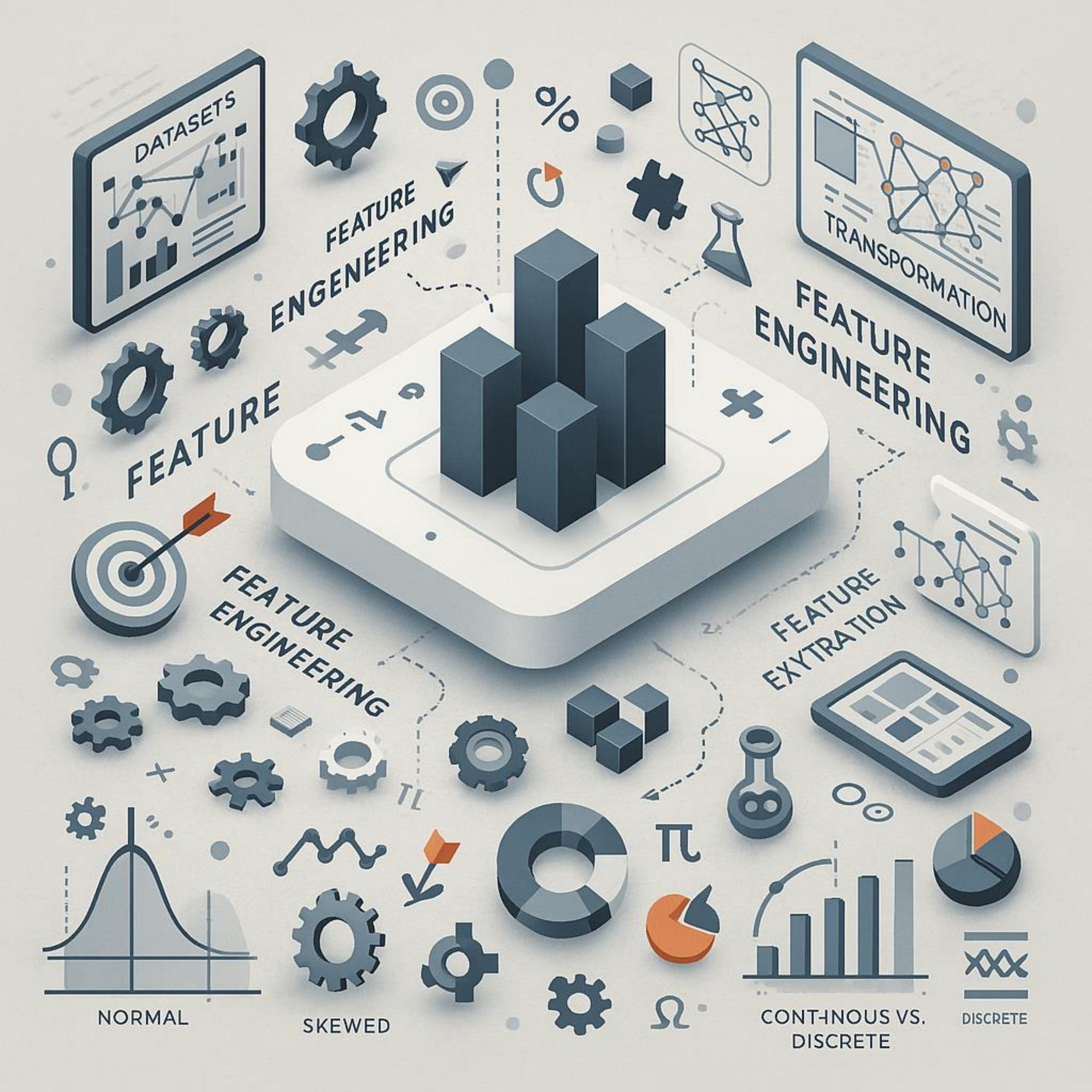
TOOL_CLIENT (Customer Database)

- Client Information

→ Enables Customer Segmentation

! **We merged the customer and sales datasets to create a unified view, and removed columns not relevant for modeling (e.g., canceled orders).**

2. Feature Engineering



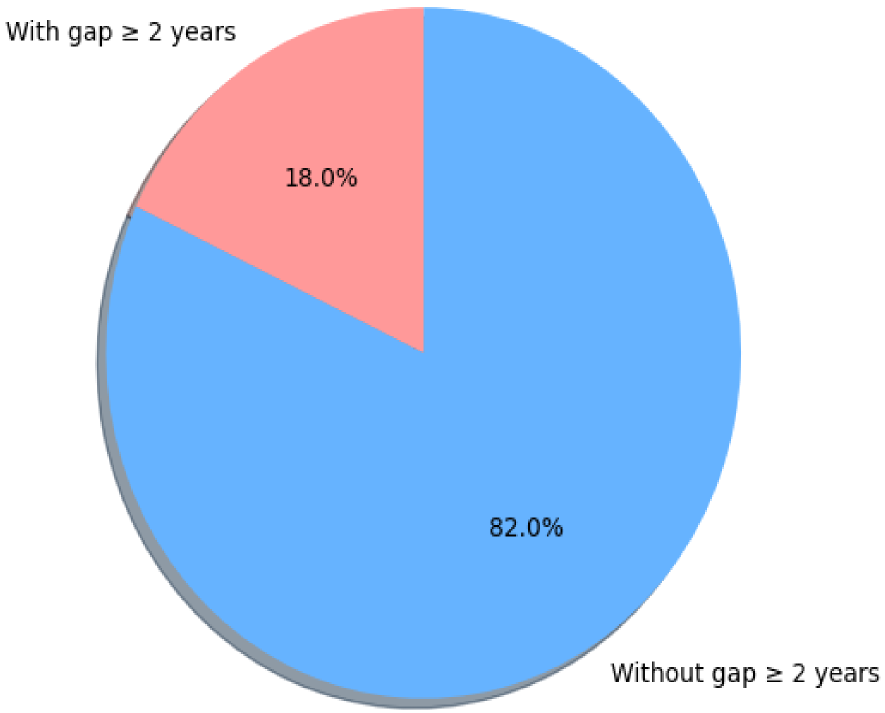
Creation of new features to Spot Reactivation Signals:

- `sales_id`: Created combining `client_id` and `yyyymm`, it represents a single transaction
- `n_purchase`: Created grouping `sales_id` by `client_id`, it represents the total number of purchases per client
- `sales_net`: summarizing the `net` sales by `sales_id`, it shows cumulative sales value per client.
- `time_diff_next`: Obtained extracting time between purchases, it shows the time until the next purchase. (Note: For the missing values, we assume the difference been 0 days)

These new features are expected to have a greater impact on the model, as they aggregate more general variables.

Understanding Customer Purchase Behavior

CUSTOMERS TYPE



KEY INSIGHTS

- The reactivation class is **underrepresented** → Work on dataset or work on models.
- Client purchasing behavior is **heavily skewed** → we considered feature scaling (`RobustScaler`).
- Some engineered features may already show **predictive power** before modeling begins (`n_purchases`, `sales_net`, or `region`).

Creation of new Dataframe on customer base for Modeling

Now each row represents a unique `client_id.`

We created one row per client with aggregated features.

Creation of a `target` column that is 1 if they were reactivated, 0 otherwise.

Why was it necessary?
In order to have a customer based target instead of a transaction based one , since our goal is to predict a particular customer inclination to be reactivated.

3. Modeling

Modeling technique: Logistic Regression

PRO' S

Highly interpretable:
Clear insights into feature impact.

Simple and fast:
Easy to implement and explain.

Good baseline:
Serves as a benchmark for complex models

CON' S

- **Assumes linear relationships:** Can miss complex patterns in behavior.
- **Lower accuracy:** Often underperforms on non-linear problems.
- **Needs careful feature engineering:** Less automatic than tree-based models.

Modeling technique: Random Forest

PRO' S

Handles non-linear relationships between features
→ essential for modeling complex customer behavior.

Works well with mixed data types
(categorical + numerical) without heavy preprocessing.

Robust to missing values and imbalanced classes
like few reactivated vs. many inactive customers.

Provides feature importance insights
helps business understand what drives reactivation.

CON' S

- **Less interpretable** than simple models
- **Slower to train** and predict on large datasets
- Feature **importance** can be **misleading**

Modeling technique: XGBoost

PRO' S

High accuracy:
Often the best-performing model in practice.

Handles rare events well:
Good for imbalanced problems like reactivation.

Flexible and tunable:
You can optimize for precision, recall, etc.

CON' S

- **Sensitive to tuning:** Needs careful parameter adjustment.
- **Longer training time:** Especially with large datasets.
- **Harder to explain**: Less transparent than Logistic Regression.

For modeling the reactivation, we aimed to choose features that:

QUALITIES

Capture **meaningful** behavioral **patterns** (how clients used to buy)

Reflect **business potential** (e.g., economic value of the client)

Available **before reactivation** (predictive, not reactive)

Are consistently **available for most clients**

TO PREDICT THE "TARGET"

Whether the client reactivated or not

→ binary classification

The goal was to model the propensity to return capturing

- profile info
- behavioral patterns

To capture profile info and patterns to model the propensity to return

FEATURES USED TO TRAIN MODELS

-region	→ Client's market segment
-n_employees	→ Size of business
-economic_pot -eco_pot_class	→ Business potential
-sales_channel	→ Channel/ product preferences
-net -n_purchases	→ Purchase behavior
-flg_tool	→ Identifies if purchase was a tool or a part
-risk_cat	→ Client risk

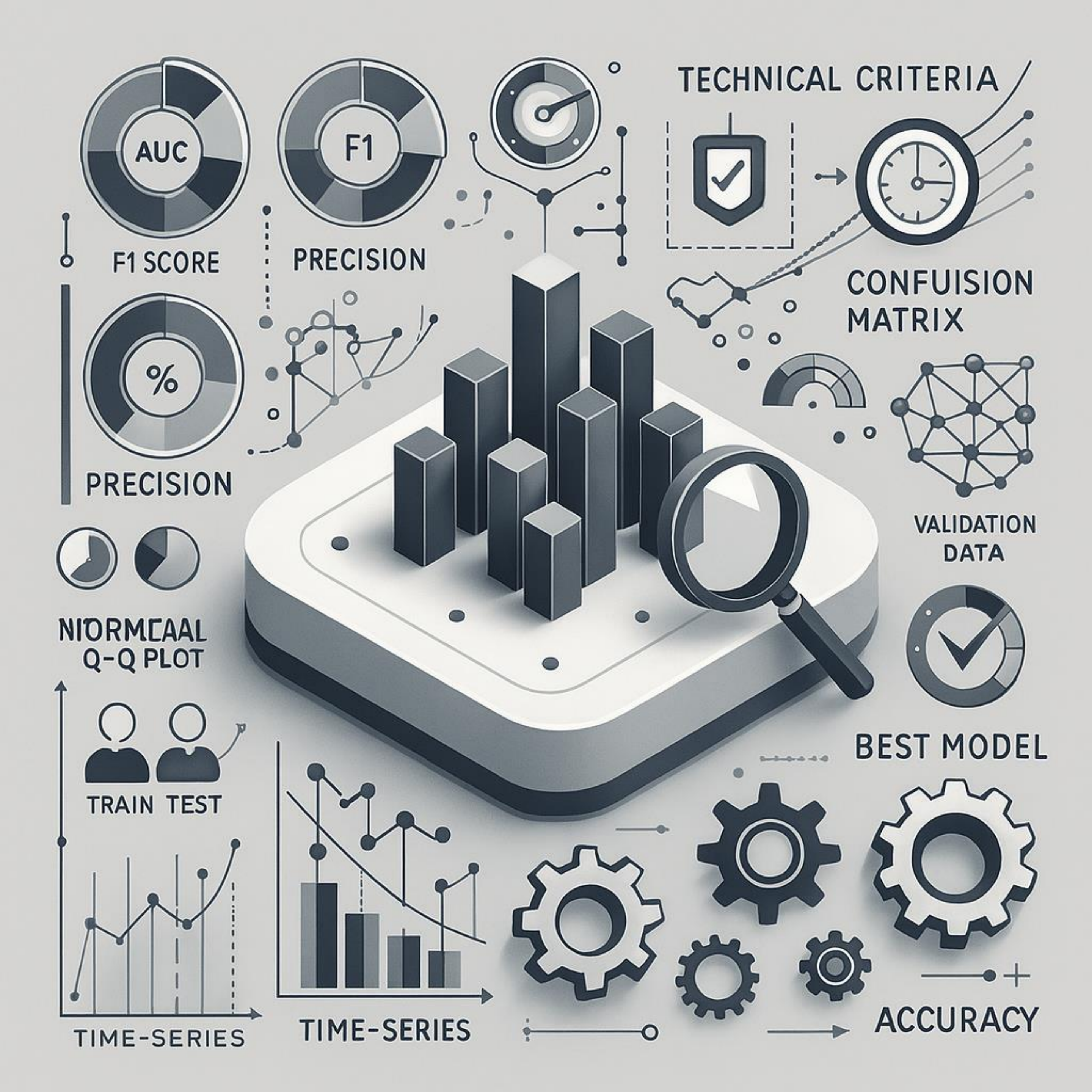
INSIGHTS

- Customers with higher number of purchases → strongest likelihood of reactivation.
- Risk categories help us group customers by behavior
- Sales channel and customer value → decide who to prioritize

For training the model we dropped the client features:

Features dropped	Reason
-item_id	→ Not feasible to aggregate due to complexity
-trade_sector -sales_id	→ Not informative
-family_code	→ grouping already included in **group_code**
-client_create_date -yyyymm	→ Temporal, not predictive

4. Model Evaluation



Evaluation of our 3 Models

- Logistic Regression
- Random Forest
- XGBoost (XGB)

Evaluation Criteria

1 PRECISION

How many of the selected customers actually returned.

High
We don't waste marketing budget on the wrong customers.

Low
We address a lot of "false alarms" - this costs money unnecessarily.

2 RECALL

How many of the customers who actually returned got recognized by the model.

High
We don't miss any potential returnees.

Low
We miss out on many opportunities because we don't address them at all.

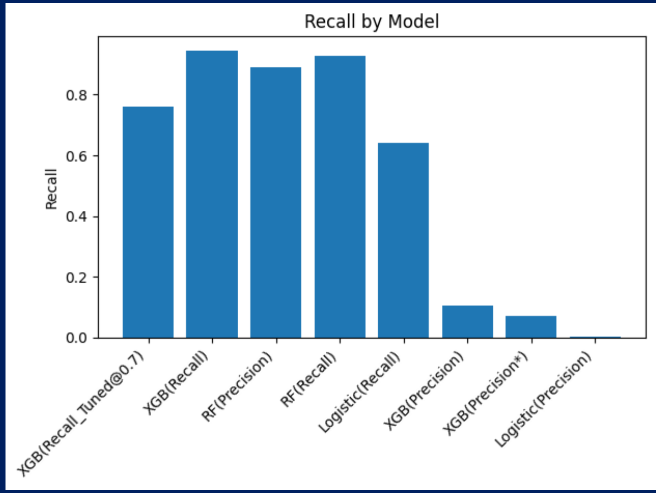
3 F1 SCORE

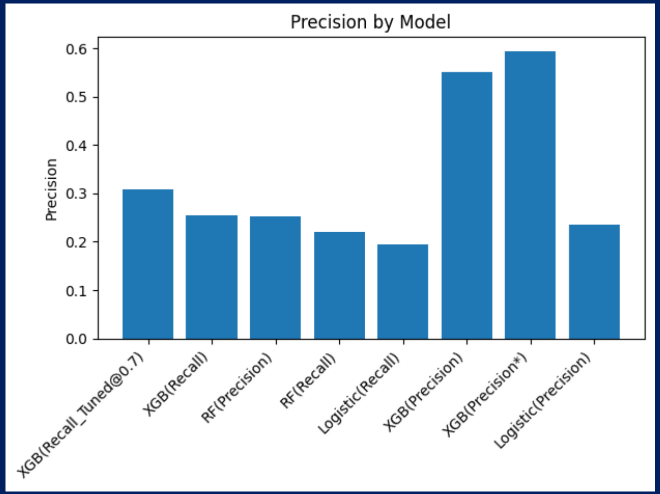
The F1 score is the harmonic mean of precision and recall.
It helps us to find a good **balance**.

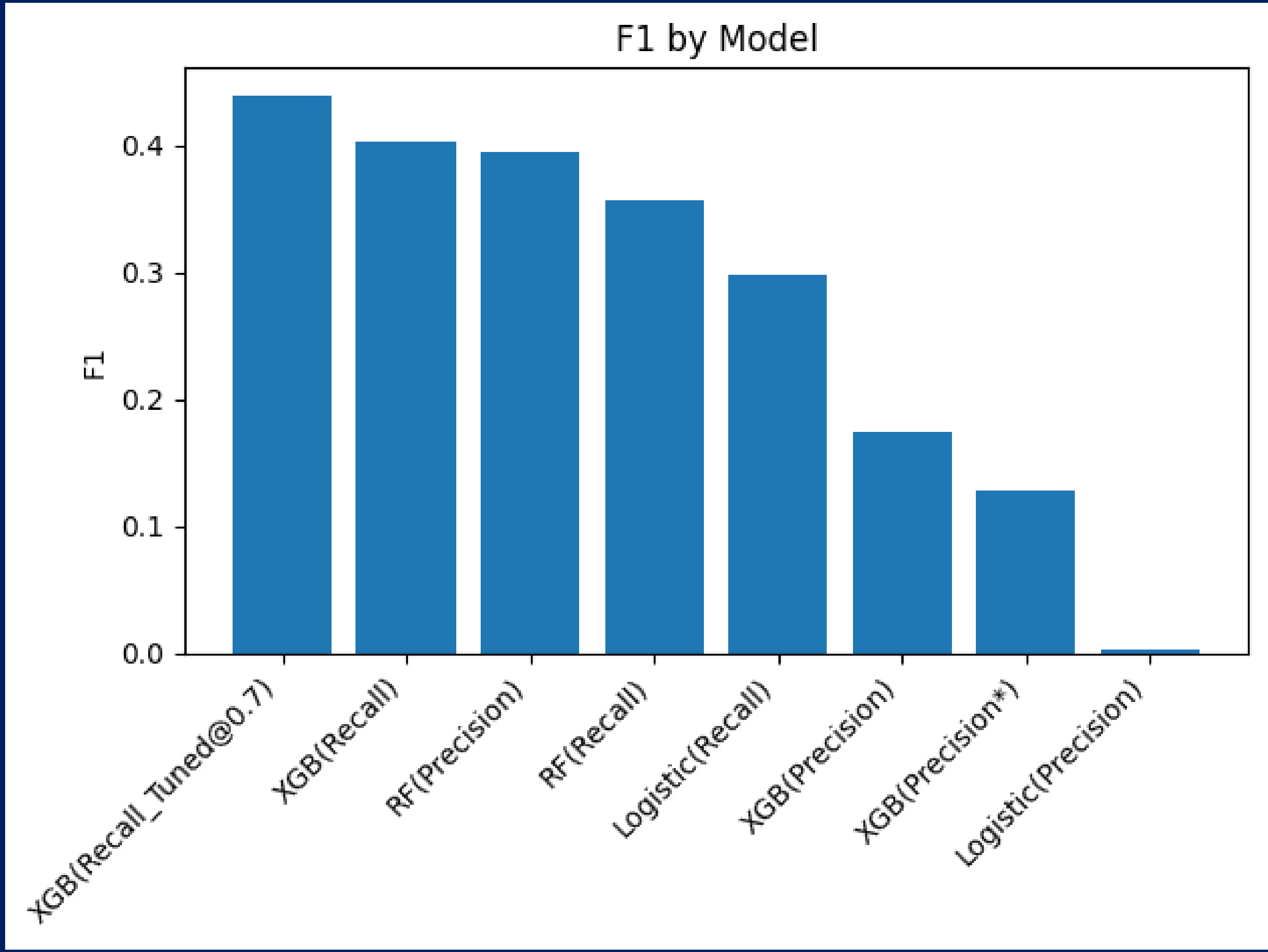
- Not too many unnecessary contacts (= high precision)
- & still as many return customers as possible (= high recall)

Ideal for CRM campaigns with a limited budget and the goal of maximizing impact.

Evaluation Criteria







And the Winner is XGBoost

The best balance between performance & reliability

1. We don't miss any potential returnees.
   Recall = 0.76 -> XGBoost recognizes almost **8 out of 10** reactivatable customers
2. We are concentrating on the right customers and dont annoy wrong ones

XGBoost = Maximum impact with minimum wastage.

Good generalizability
The model behaves consistently on unknown data (no overfitting)

5. Business Insights

Business Insights

REVENUE GROWTH FROM SMART REACTIVATION

- **Higher sales by reactivating the right customers**
  The model captures nearly all returnable customers – unlocking full revenue potential.

- **Recurring revenue through targeted reactivation**
  Reactivated customers often make repeat purchases.→ The model helps unlock sustainable revenue streams without acquiring new customers.

- **Lower acquisition costs**
  Reactivation is cheaper than acquiring new customers – the model focuses on cases where it's most cost-efficient.

Business Insights

MORE EFFICIENT & TARGETED MARKETING

- **Clear targeting, less waste**
  The model defines a precise reactivation group – campaigns focus only on those who matter.
- **Higher marketing ROI**
  Resources are used where the chance of success is highest – with measurable return.
- **Smarter segmentation based on data**
  We identify high-potential customer segments using model insights – improving targeting and outcomes.

Business Insights

CUSTOMER INSIGHTS & BONDING

- **Understanding customer motivation**
  The model shows which factors drive reactivation – enabling more relevant offers and stronger engagement.

- **Data-backed decision-making**
  The model reveals why a customer might return – supporting better strategic decisions.

- **Stronger long-term customer loyalty**
  Reactivated customers can become loyal relationships – increasing their long-term value.

THANK YOU FOR YOUR ATTENTION


Roberto Vanzetta


Helena Borges Daré


Ignacio Kreis


Jacopo Abete


Jasper von Pachelbel