DATA ANALYSIS Canvas

™ Why Data Wins

PROJECT TITLE *state the goal in one sentence*

Can we identify churners 90 days earlier than our retention team?

END USER *who are we helping? what friction do we remove?*

Retention agent: less searching, more focus on highest urgency customers.

TEAM

Data & CRM — TelecomCo

DATE

June 2026

VERSION

v1.0

USE INSIGHT (how?)

1 Information gain

Our project will yield new information. Will it be better than what we have today?

How will the new information be generated?

What is our current baseline? Can we fundamentally improve it?

Baseline today: agent decides on gut feeling and manual CRM searches.

New: churn score per customer, 90 days before cancellation.

Improvement: fundamental — agent sees what they cannot see today.

✓ Go/no-go: accuracy ≥ 78% on holdout set.

2 Information use

How do we act on the new information? Who makes which decisions?

How do we use information to create added value?

What does our decision process look like and who is responsible?

Who: retention agent, not the manager.

Form: score + top-3 reasons in CRM screen. When: at the start of every customer call.

Decision: agent chooses approach (discount, upgrade, visit). System advises, human decides.

Parameter: threshold ≥ 0.72 → set bi-annually by PM + data scientist.

3 Added value end user

Central to our project is the end user. What concrete friction do we remove?

Who is the end user of our solution?

What added value will the solution provide?

What does the end user do differently afterwards?

End user: retention agent. Not the manager, not the data scientist.

Friction today: calling 18,000 customers/week on gut feeling — no direction.

After project: focused list of 2,400 customers with highest urgency.

Different: 'Who should I call?' disappears. Agent calls with focus and context.

■ End user ≠ decision-maker in information use.

4 Added value measuring

Value must be measured in end user behaviour — not in model accuracy.

When will our project deliver added value?

How do we measure the added value for end users?

Retention conversion: 14% → ≥ 25%. Time per file: – 40%. Adoption: ≥ 80% use tool after 60 days.

No AUC or F1 as success. Value = behavioural change.

Measurement: interviews (week 4, 8, 12) + CRM conversion ratio data.

5 Data collection

Where does data come from? Costs, quality, dependencies and risks over time.

What data sources do we use and who provides them?

What is the cost and quality of our data?

Sources: CRM, billing, network platform, service desk.

Quality: CRM has 12% missing values on phone number — resolve before go-live.

Dependency: billing connection requires IT ticket, lead time 6 weeks.

Risk: data owner billing = different dept — governance must be agreed.

9 Offline evaluation

Before going live: how do we test impact? When do we give the green light?

How do we evaluate our application before it goes live?

What is the impact on our people, processes and systems?

Go/no-go: accuracy ≥ 78%, precision ≥ 70% on 6-month holdout set.

Cost of error: false positive = €15 (unnecessary discount). False negative = €380 (missed retention).

Go/no-go decision: PM + data lead. No experiment to production without formal approval.

7 Development

Elaborating a method is one thing. Developing it into a working application is another step.

How will we develop our data application?

What technical requirements must it meet, today and in the future?

Phase 1: weekly batch scoring, Excel export to CRM.

Phase 2: Azure ML pipeline, extendable to real-time scoring.

Scalability: containerized model, automated retraining.

Responsibility: data scientist is model owner, IT manages pipeline.

6 Data methods

Which methods do we apply? And how do we improve them over time?

Which data methods do we apply in our analysis?

How do we test, validate and improve our methods?

Method: XGBoost classification. Baseline: logistic regression.

Iteration: monthly evaluation of precision & recall on new production data.

Improvement: feature importance review every 90 days at retraining.

Iteration with decision-makers: quarterly threshold review with PM.

8 Input data

The key to information gain is selecting the right data. Who decides on this?

How do we decide which input data we need?

What should our input data look like and how do we document it?

Variables: 6-month usage data, complaint history, contract type, tenure, NPS.

Format: monthly snapshot. Aggregations: avg. data, 3-month trend, # complaints per quarter.

Decision: domain expert + data scientist together — never solo.

Documentation: data dictionary in Confluence.

CREATE INSIGHT (how?)

EVALUATE (how well?)

10 Monitoring

No application works forever. Monitor data, methods, usage and value.

How do we monitor performance, consistency and error margins over time?

What procedures do we follow when performance changes?

Data: weekly check on missing values and input variable drift. Methods: monthly accuracy measurement. Usage: adoption dashboard for team leads. Value: quarterly conversion ratio review.

Who pulls the plug? → PM decides. Escalation if drift > 5pp accuracy loss.

Credits: Based on the Machine Learning Canvas by Louis Dorard (www.ownml.co), adapted by Maarten Vanhoof (Craftzing). Reworked by Joachim De Vos — Why Data Wins.