Proactive Churn Reduction Initiative

Presenter: Omri Shahar **Date:** 08/12/2026

Made with GAMMA

Business Challenge and **Why It Matters**

Topic	Current Situation	Impact
Business Challenge	**High customer attrition:** The current churn rate is approximately **27%**.	This represents a significant baseline loss of about **8,224,160.70$**.
Current Strategy	**Reactive Churn Chasing:** Employees contact clients *only after* they leave, offering high discounts (20–40% off standard rate - an average of **300$**).	This strategy is highly inefficient, with success rates for recruiting customers back ranging only between **15–30%**. in addition to the discounts and the men power required.
Financial Stakes	The average CLV (Customer Lifetime Value) is **$4,400.30**. The estimated loss based on 1869 churners in the dataset is approximately **8.2 million dollars**.	We are currently losing 70–85% of clients who leave. Implementing a predictive model is essential to **save costly and time-consuming client chasing**.

Resources Needed & Success Criteria

- **Data Science Expertise:** Model development, optimization, and pipeline integration.

- **Data Access:** Continued access to customer feature data for scoring.

- **Integration:** Collaboration with IT/ML Ops to feed model output into the existing ML discount system.

- The model must run automatically at fixed intervals on all clients. The output must be an **easy-to-process format** (JSON, yaml, etc.) containing only the client IDs of positive predictions (predicted churners).

- **Financial Optimization:** Success is defined by the model's ability to maximize net financial value. The target value function is designed to maximize: **(Recall × (avg CLV - avg discount) - (FPR × avg discount)**. → **(Recall × 4,100.30$) - (FPR × 300$)**.

- Preserving a client yields a net value of **4,100.30$** (CLV of 4,400.30$ minus average retention discount cost of 300$).

Problem Statement & Proposed Solution

Problem

The company needs to shift <u>from</u> a reactive, high-cost strategy for client retention <u>to</u> a **proactive, automated predictive system**. The current 27% churn rate is financially unsustainable.

Proposed Approach

1. Develop a predictive **churn classification model** using current customer behavior data.

2. Optimize the model not purely on accuracy, but on maximizing the **financial value score** (balancing True Positives and False Positives).

3. Automate the model to feed client IDs identified as high-risk into a secondary ML system that generates custom, temporary, and smaller discounts.

Data & Resource Plan

- **Data Source:** Internal telecom customer data (prototyped using the Telco-Customer-Churn dataset).

- **Data Volume:** 7043 entries.

- **Key Data Challenge:** The target variable (Churn) is unbalanced, requiring techniques like **Stratified K-Fold cross-validation** to ensure robust performance measurement.

- **Key Features:** The model relies heavily on customer tenure, contract type, and monthly charges.

- **Resource Requirement:** Development phase concluded; focus now shifts to integration and monitoring of the optimal solution.

Key Retention Insights from Data Analysis

Analyzing customer behavior revealed critical factors for retention:

Contract Type Matters Most

Long-term contracts (negative correlation of -0.396) significantly reduce churn. *Recommendation: Incentivize long-term contracts.*

Critical Early Phase

Low tenure strongly predicts churn (negative correlation of -0.352).

Recommendation: Focus on onboarding and early experience during the critical first 6–12 months.

Services Retain

Add-on services, particularly **Tech Support** (negative correlation of -0.282) and **Online Security** (negative correlation of -0.289), substantially reduce the likelihood of leaving. *Recommendation: Bundle these services.*

Expected Outcomes (Target Financial Performance)

The model is optimized for the threshold that yields the maximum Value Score.

- **Baseline Loss (No Model):** 8,224,160.70$.

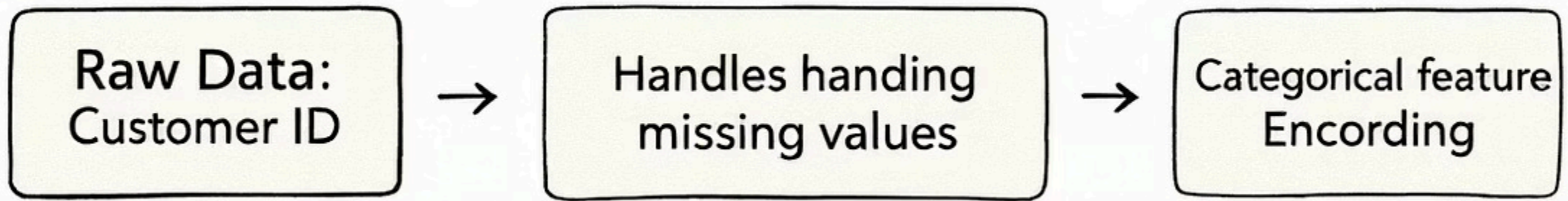- **Expected Model Performance (Logistic Regression at TH=0.30):**

$3,023.439

Optimal Value Score

per customer.

75.6%

Recall (True Positive Rate)

We correctly identify over three-quarters of actual churners.

25.5%

FPR (False Positive Rate)

The cost of unnecessary discounts is managed.

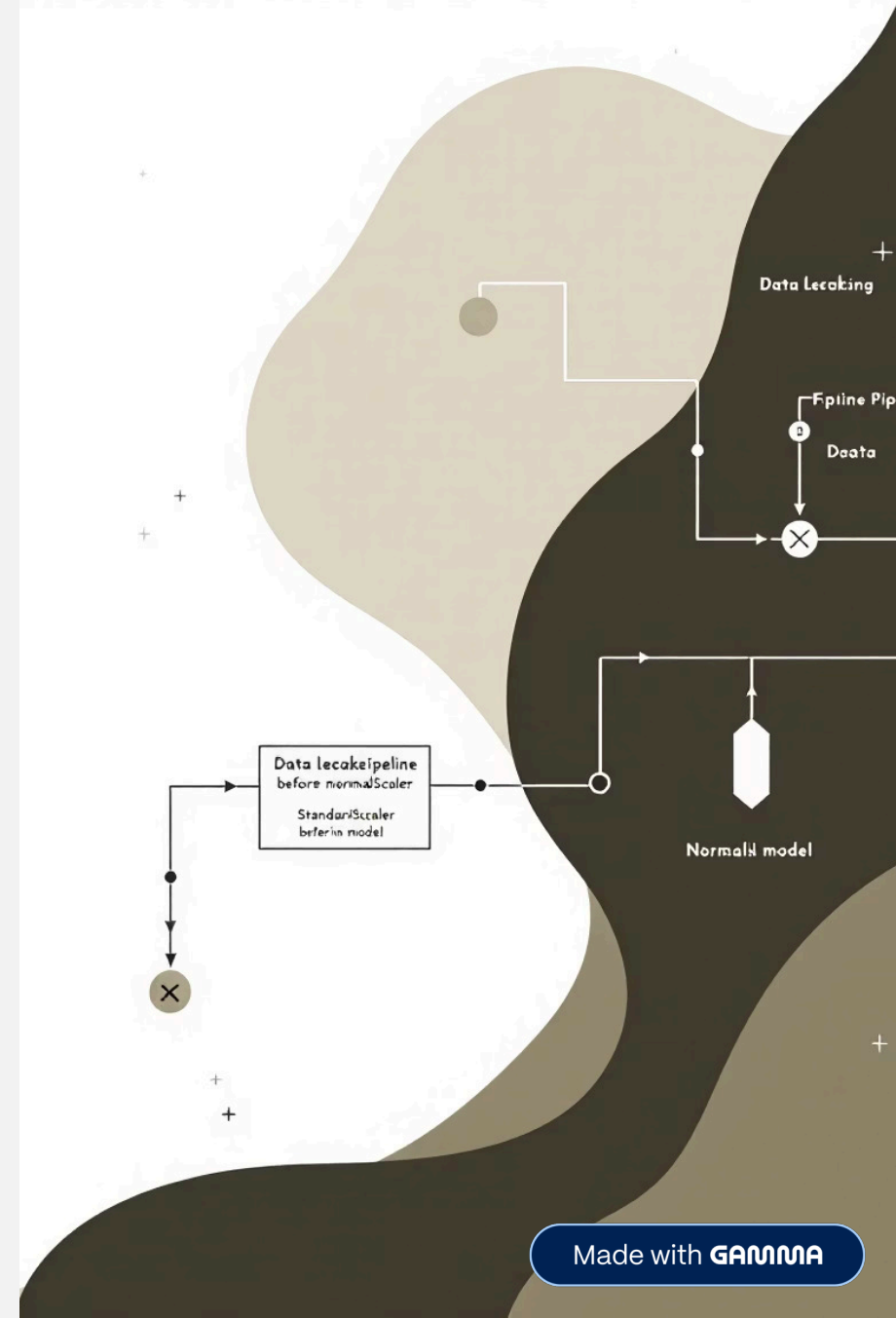| Raw Data: Customer ID	→	Handles handing missing values	→	Categorical feature Encording

Data Preparation Process: Initial Cleaning & Encoding

- **Initial Dataset:** 7043 rows.

- **ID Removal:** The customerID column was dropped.

- **Handling Missing Values:** Rows where tenure was zero (11 rows) were dropped, as they represent new customers with no measurable tenure. (and critical for the feacture engineered)

- **Feature Conversion:** The majority of features (17 columns) were of object dtype. A LabelEncoder was applied to convert all these categorical columns into numerical representations.

- *Example:* gender, Partner, Dependents, Contract, and Churn were converted to integers.

Data Preparation Process: Scaling and Pipeline

- **Normalization Rationale:** Numerical features need to be on the same scale (mean=0, std=1) for many ML algorithms, particularly Logistic Regression (large value will produce larger weight updates, not because it's more important — just because its numbers are bigger)

- **Scaling Method:** StandardScaler was used for normalization.

- **Data Leakage Prevention:** The scaling process was encapsulated within a Pipeline using ColumnTransformer to ensure the scaling parameters are fitted *only* on the training sets during cross-validation, thereby avoiding data leak.

EDA Key Findings: Imbalance and Correlation

Target Imbalance

The dataset is imbalanced: 5174 Non-Churners (0) versus **1869 Churners (1)**. This corresponds to a **27% churn rate**. This imbalance necessitated using Stratified K-Fold for robust validation - keeping the save ratio in the train\test splits.

Contract type matters most

Contract-type feature has -0.397 correlation value with churn, according to the following diagram we can learn that the vast majority of churned clients had a month-to-month contract.

Add-on services work

Online backup, tech support and online security have correlation values of -0.195525, -0.282492, -0.289309.

Top 5 Positive Correlates with Churn:

1. MonthlyCharges (0.193).
2. PaperlessBilling (0.191).
3. SeniorCitizen (0.150).
4. PaymentMethod (0.107).

Top 5 Negative Correlates with Churn (Retention Drivers):

1. Contract (-0.396).
2. tenure (-0.352).
3. OnlineSecurity (-0.289).
4. TechSupport (-0.282).