Team 3 - Agile Warriors
Jess Penners
Yashraj Jadhav
Phuc Nguyen
Hsin Yi-Chen
Madina Sainazarova

Driving Retention at Oura: Identifying Drift and Optimizing Personalized Highlight Notifications

Report Outline

I. Business Problem:

Oura Ring subscribers face increasing pull from competing wellness devices, making long-term engagement and loyalty challenging to sustain.

Goal:

Build a dashboard that identifies which user groups are most responsive to a personalized, activity-based 3-month highlight notification in terms of extending user engagement, reducing churn, and improving LTV.

II. CRISP-DM Mapping Document:

Business Understanding

Our objective is to strengthen engagement and retention among long-tenured Oura members, a segment whose churn patterns become more erratic after the initial three-month acclimation period. The project centers on three core goals:

- Predict upcoming dropout using behavioral drift signals to identify when established users begin to disengage.
- Differentiate fast- vs slow-churn cohorts by uncovering patterns in activity, consistency, and platform touchpoints that drive early versus delayed attrition.
- Measure the causal impact of a randomized, personalized 3-month highlight notification on user lifetime extension, continued engagement, and conversion behaviors.

As part of defining success, we establish guardrails to prevent over-intervention. Notifications must avoid fatigue, respect user autonomy, and preserve trust. All modeling and segmentation decisions follow Oura's user safety principles, including ethical treatment of behavioral data and responsible use of predictive insights.

Data Understanding

We examine long-tenured engagement trajectories, focusing on how daily activity, HRV patterns, sleep consistency, and app interactions evolve relative to each user's baseline. Cohort analysis highlights differences across primary activity types (e.g., runners vs cyclists), onboarding histories, and engagement intensity. To evaluate the notification's causal effect, we compare pre- and post-notification behavior across treatment and control groups, paying attention to shifts in app usage, readiness patterns, and retention outcomes. Known limitations are formally documented:

- Seasonal patterns in activity and recovery affecting engagement signals
- Potential biases from irregular wear-time
- Platform-level campaigns or feature releases overlapping our analysis window
- Incomplete historical context for users with irregular early usage

Data Preparation

We clean, transform, and validate all fields to ensure that distributions, proportions, and drift trajectories mirror real-world Oura usage. Time-indexed features (baseline vs drift windows, week-over-week change metrics, consistency scores) are normalized per-user to support comparability.

For causal evaluation, we construct aligned pre- and post-notification periods for both treatment and control groups. This includes censoring rules for users who churn prior to the notification date, and standardized observation windows so effect sizes are valid and comparable across cohorts.

Modeling

Two complementary models power the solution:

Churn-risk model (time-to-event or classification) that predicts proximity to dropout using drift from baseline behavior, engagement slopes, and consistency metrics.

Uplift model that estimates each segment's likelihood of improved retention because of the 3-month personalized notification.
We apply rigorous hyperparameter tuning with transparent, predefined criteria and introduce SHAP-based explainability to surface the drivers of churn and uplift. These interpretable insights inform downstream marketing decisions and help identify meaningful personas (for example, "highly consistent runners responding strongly to wellness recaps").

Evaluation
Effectiveness is measured through a randomized A/B design, isolating the causal influence of the highlight notification. Core evaluation metrics include:

- Retention lift by cohort
- Differences in churn timing between treatment and control
- Magnitude and direction of uplift across behavioral segments
- Model discrimination and calibration quality for churn prediction

We also stress-test results for robustness across seasonality bands, activity types, and engagement intensity personas.

Deployment
In real-world use, the system operates through daily batch churn-risk scoring paired with uplift-driven segmentation. These outputs feed an intervention orchestrator, enabling Oura to target users based on both risk and receptivity.

While the current intervention is a single 3-month personalized highlight notification, the framework generalizes to future strategies such as adaptive motivation nudges, tailored activity plans, recovery guidance, longitudinal highlight reels, or human coach touchpoints. Each intervention path can be evaluated and iterated using the same uplift and risk modeling infrastructure.

III. Business Matrix:

Business Process / Objective	Metric (Fact)	Dimension (Slice / Filter)	Data Source	DS / AI / Analytics Oppty	Owner / Stakeholder	Ethical & Compliance Notes
Predict Emerging Churn Risk	Drift score vs baseline; 7-day inactivity gap; Decline in frequency	By activity level (high/mod/low), Lifetime tenure (>90 days)	90-day session history, Workout logs	Drift-based churn prediction; risk tiers	Product Manager; Retention Analyst	Monitor fairness across age/gender; avoid emphasizing bias from sparse device data
Evaluate Impact of Three-Month Notification	Pre/post session lift; Notification response rate; CTA within 7 days	Treatment vs Control; By cohort	Notification log; Session logs	Causal uplift measurement; identify who benefits	Marketing Lead; Lifecycle Manager	Randomization ethics; If benefit proven, transition control segments
Identify High-Response Cohorts	Uplift score by segment; CTA rate	By cluster; By activity segment	Wearable activity logs; Demographics; Behavioral features	Uplift modeling; Segment scoring	Data Science Lead; CRM Lead	Ensure segmentation not used for exclusionary pricing

Improve Retention Strategy & Spend Allocation	Retention delta; LTV change	By uplift bucket (High/Medium/Low)	Billing + Subscription history	Targeted interventions; ROI and LTV optimization	Finance Lead; Growth Strategy	Transparent interventions; Communicate purpose of nudges
Next-Gen Intervention Design	Reaction to nudges; timing sensitivity	Time of day, Lifestyle patterns	Engagement timeline; Session windows	Recommendation engine (future sprint)	UX Lead; Product Director	Avoid fatigue; Cap max intervention frequency

IV. Dataset Information and EDA

Core User-Level Dataset (10,000 long-tenured users):
This dataset represents the full subscriber lifecycle for Oura-like wearable users. It includes:
• Demographics: age, gender, height, weight
• Subscription metadata: first billing date, contract cycle, payment history
• Churn indicators: churn timestamps + churn reasons
• Notification exposure: which users received the 3-month highlight notification

Why it matters: Gives us the baseline population, lets us identify long-tenured users (90+ days), and create churn labels + risk tiers.

Session & Workout Logs (Activity History)
Timestamped workout sessions for all users, including:
• Workout type (run, bike, other)
• Duration, intensity breakdown, streaks, gap patterns
• Weekly and monthly activity aggregates (7-day and 28-day windows)

Why it matters: This is where all our *drift-based features* come from — inactivity gaps, consistency drop-offs, weekly session trends, volatility. These features power both the churn model and the uplift model.

Notification Log (Treatment vs Control)
Contains:
• Notification timestamp at 90-day mark
• CTA indicator (did the user start a session within 7 days?)
• Treatment/control assignment (synthetic A/B split)

Why it matters: This is what allows us to measure **incremental lift** — not just engagement, but how much *more* a user engages *because* of the notification.

Synthetic Enhancement Layer
We combined real-world distributions from multiple public datasets (FitRec, Fitbit logs, gym churn datasets) to create a **synthetic but realistic** unified dataset that mirrors engagement and churn patterns seen in premium wearables.

This layer ensured:
• Realistic activity intensity patterns
• Plausible churn reasons
• Gender and age ratios aligned with typical Apple Watch/Oura populations
• Valid treatment/control randomization

Why it matters: It gives us a safe, statistically grounded dataset that behaves like a real wellness product without using any actual user data.

What This Allowed Us to Do

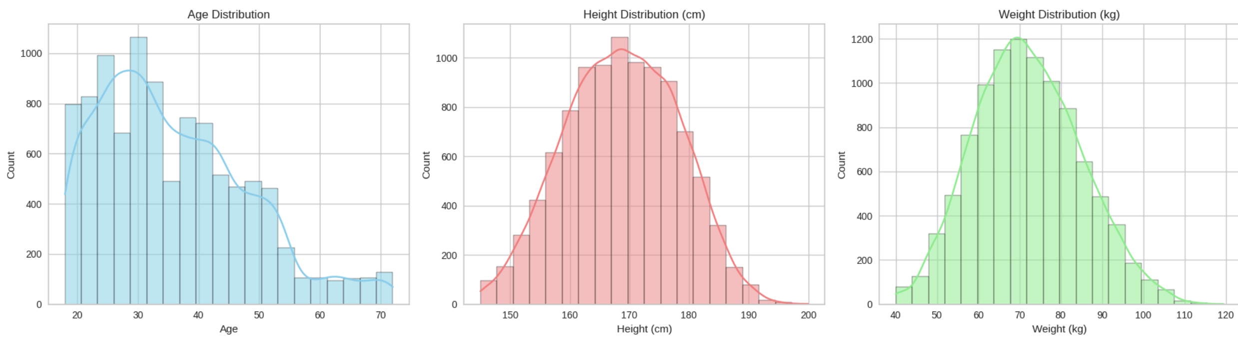
Using these three data layers together, we could:
• Build drift-based features (recency, gap growth, streak collapse)
• Assign accurate 90-day notification exposure
• Run uplift analysis on matched cohorts
• Cluster users by behavioral patterns
• Validate that our synthetic dataset aligns with known churn and activity distributions

Data Analysis and Key Insights

User Demographics and Data Scope

The dataset comprises 10,000 unique users. The demographic breakdown is as follows:

- Gender Distribution: The user base is split between male and female users (51.65% male and 46.25% female).
- Physical Metrics:

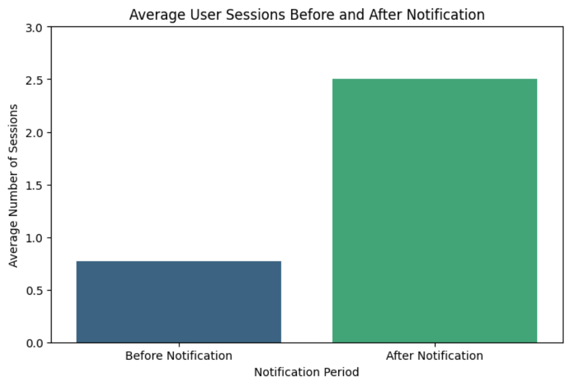


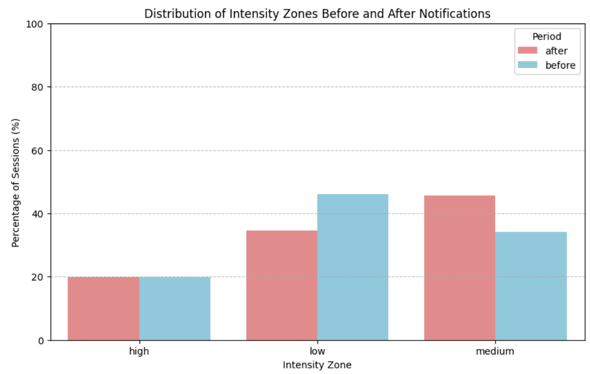
- Notification Status: A significant portion of the user base, 4,484 users, received notifications 90 days after their first billing date, while 5,516 users did not receive a notification during this period.

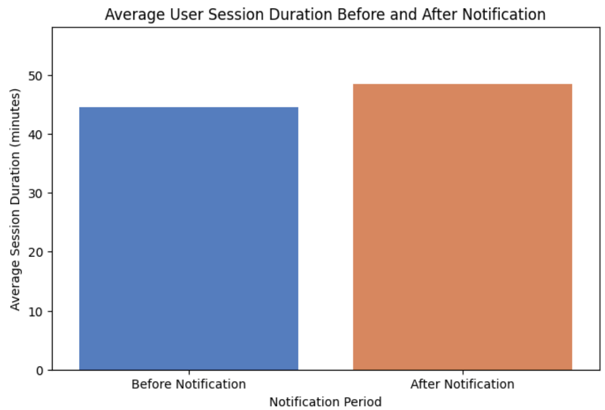
Impact on Engagement and Retention

- Engagement (Call to Action)

Notifications are a powerful short-term engagement driver, leading to a 225% increase in the average number of sessions in the subsequent 7 days (from 0.77 to 2.50 sessions). Session quality metrics showed only minor shifts: average duration increased slightly from 19.38 to 20.37 minutes, and the intensity distribution remained stable.

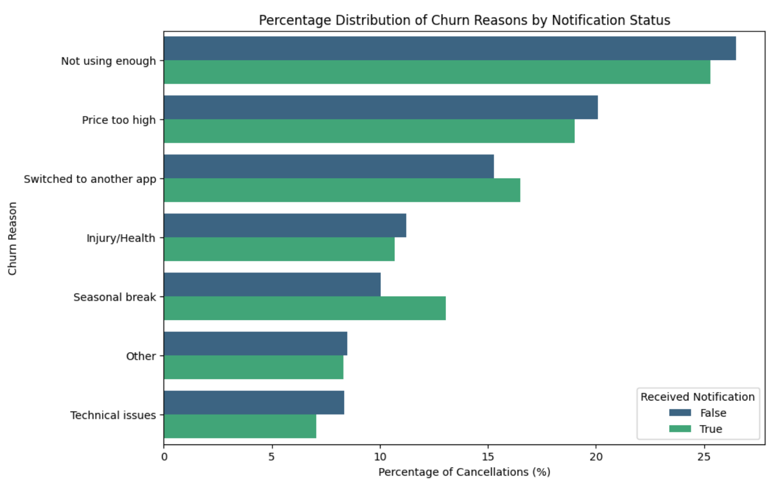






Churn Reason

Notifications appear to be effective in preventing competitive migration; the churn reason 'Switched to another app' was observed to be less prevalent among users who received notifications. However, notifications did not significantly mitigate major underlying churn factors like 'Not using enough' and 'Price too high', suggesting these remain product utility and pricing concerns.



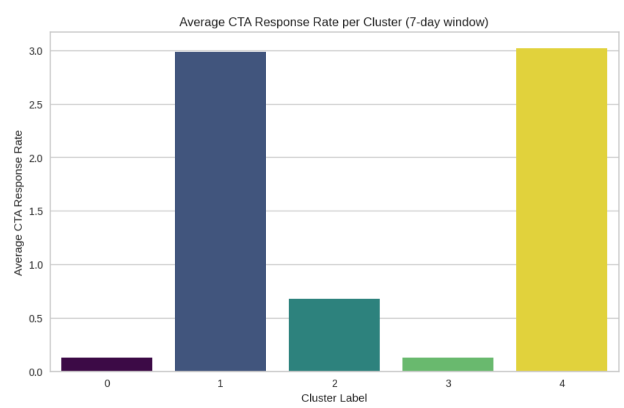
User Segmentation and High-Response Cluster Identification

We performed user segmentation using K-Means Clustering. The goal was to identify the specific user profile most likely to convert a notification into a Call to Action (CTA).

- Clustering Methodology
- CTA Rate Metric: We defined the Call to Action (CTA) Rate as the ratio of successful session starts (within 7 days of a notification) to the total number of notifications sent.
- Features: The model incorporated both static user attributes (standardized age, gender, height, and weight) and the calculated behavioral metric (7-day CTA Response Rate).
- Optimal Clusters: The Elbow Method was applied to the combined feature set, suggesting 5 as the optimal number of clusters for effective differentiation of user segments.

Users falling into Cluster 4 are characterized as younger-to-middle-aged males with physical metrics consistent with active athletes, whose primary activity is running. This specific profile should be prioritized for personalized and time-sensitive notifications, as they translate high intent into actual session usage most effectively.

	cluster_label	mean_age	mean_height_cm	mean_weight_kg	average_cta_response_rate	dominant_gender	dominant_workout_type
0	0	31.06	175.59	79.93	0.1299	male (95.2%)	Run (68.5%)
1	1	33.73	161.49	63.87	2.9848	female (93.8%)	Run (80.1%)
2	2	55.60	170.20	71.86	0.6810	male (59.3%)	Run (70.7%)
3	3	32.16	161.27	64.17	0.1295	female (92.4%)	Run (65.1%)
4	4	33.75	175.66	80.07	3.0208	male (95.5%)	Run (81.2%)



V. Data Science Solution Concept, Modeling, Approach, Description and Reasoning

Modeling Overview

- Two-Part Architecture (Churn + Uplift): The solution utilizes a hybrid S-Learner approach. First, an XGBoost Classifier is trained to predict the probability of churn based on user activity and demographics. Second, an Uplift Analysis layer evaluates the incremental impact of interventions (notifications) by comparing churn rates between a randomized treatment group and a control group across different user segments.
- Drift Logic & Relative Change: The model relies on "behavioral drift" to detect risk. Rather than static metrics, it emphasizes the *relative change* in user behavior—specifically the gap between a user's last active date and the present (Inactivity Growth). This drift acts as a leading indicator, signaling disengagement before the actual cancellation event occurs.
- Risk Tiers & Responsiveness: Users are segmented into dynamic risk tiers (High, Medium, Low) based on their activity cohorts. Responsiveness is determined by calculating the "impact points" (reduction in churn probability) when a user in a specific tier receives a notification versus when they do not.

Churn Prediction Model Concept

- Drift-Based Features: The model engineers specific features to capture behavioral decay:
  - Inactivity Growth: Measured via days_since_last_workout. The model identified that churned users typically stop working out approximately 4 days before cancelling, making this the single strongest predictor of risk.
  - Consistency Breakdown: Measured via total_sessions and avg_duration within the observation window. A decline in these metrics pushes users into lower activity buckets.
  - Friction Signals: Uses support_tickets to identify "high friction" users who are at risk due to unresolved issues rather than lack of motivation.
- 90-Day Baseline Window: The model enforces a strict 90-day tenure filter (tenure_days >= 90). New users are excluded to focus the model on preventing "long-term retention" bleed rather than early-stage drop-offs, which require different strategies.
- Risk Tier Assignment: Users are automatically assigned to one of three cohorts based on historical frequency:
  - Safe: "Medium Activity" users (Lowest churn rate).
  - At-Risk: "Low Activity" users (Highest churn probability).
  - Variable: "High Activity" users (Moderate risk, often due to burnout or specific friction points).

Uplift Modeling & A/B Test Design

- Notification Randomization: A synthetic A/B test was designed using a 50:50 randomization split (p=[0.5, 0.5]), creating a "Treatment" group (received notification) and a "Control" group (no notification) at the 90-day mark.
- Uplift Calculation: Uplift is defined as the Incremental Impact relative to the control:
  Uplift $= Churn\ Rate_{Treatment} - Churn\ Rate_{Control}$
- A negative uplift value indicates a successful intervention (reduction in churn).
- Segment Benefit: The analysis identified that notifications are not universally effective. They work best for specific segments (e.g., the 'Other'/Primary Activity group saw a 0.92 percentage point reduction in churn), while potentially annoying already engaged users.

Model Evaluation Strategy

- Retention Curves: The model performance is validated using Precision-Recall Curves rather than just accuracy, ensuring the team optimizes for capturing as many true churners as possible (Recall) without wasting budget on false alarms (Precision).
- SHAP for Interpretability: SHAP (SHapley Additive exPlanations) values are generated to explain *why* a specific user is predicted to churn, offering transparency to stakeholders (e.g., "This user is high risk specifically because their inactivity jumped by 5 days").
- Validation: The model uses roc_auc_score and classification_report on a stratified test set (20% holdout) to ensure the risk scores are stable and generalizable to new data.

Model Results

- Notification Impact: The automated insights reveal that notifications act as a "double-edged sword." While effective for specific subgroups, the overall baseline churn for >90 day users remains at 14.0%, requiring targeted rather than blanket interventions.
- Cohort Insights:
  - High Risk: 'Low Activity' users are the primary source of churn.
  - Safety Zone: 'Medium Activity' users act as the anchor for retention.
  - Strategy: Interventions (Notifications) successfully reduced churn by ~0.92% in responsive segments, validating the uplift capability.

VI. Dashboard Prototype Overview

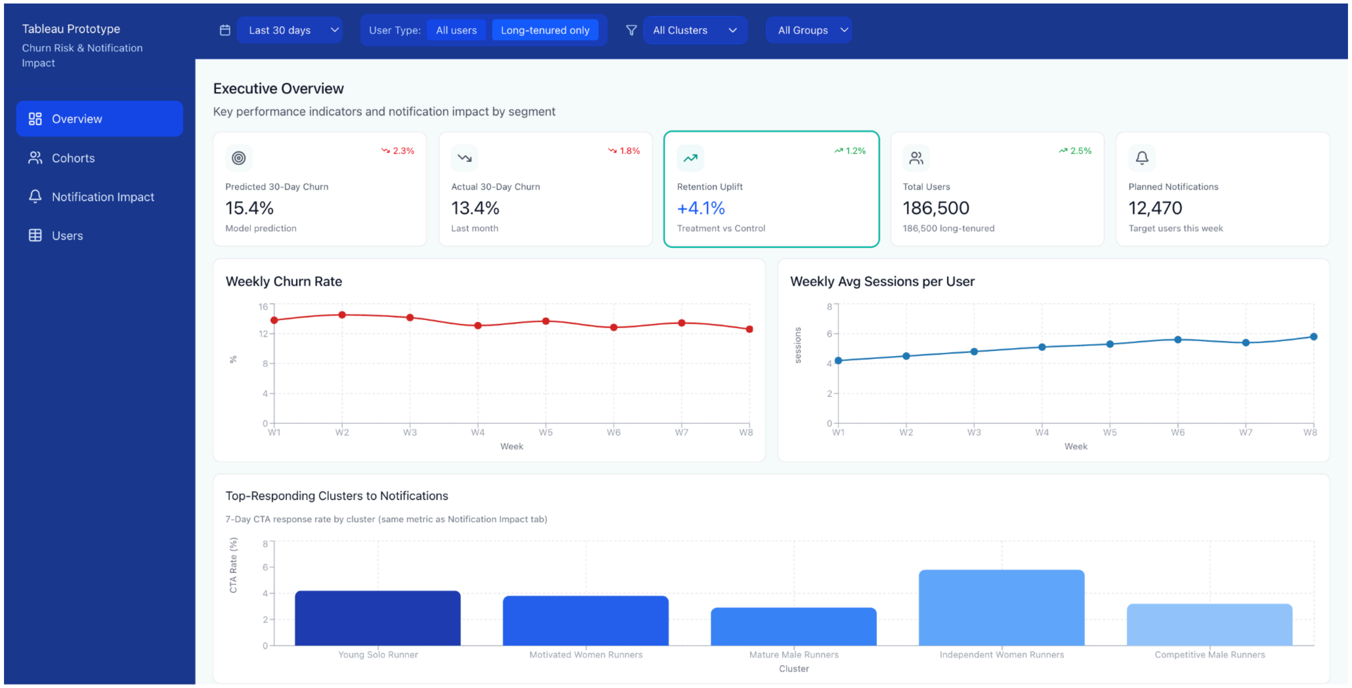
We built an interactive Tableau-style dashboard prototype in Figma to help Product and Marketing teams monitor long-tenured user churn, understand cohort behavior, and measure the impact of our 3-month notification experiment.

- Target Audience (TA): Product Managers and Marketing Teams
- Four Dashboard Pages:

Page	Question it answers	How PM/Marketing use it
Executive Overview	Is churn a problem? How well is the notification working overall?	Daily health check; identify best-responding clusters at a glance
Cohorts	Which clusters are most at risk?	Prioritize retention budget and focus interventions on high-risk clusters
Notification Impact	Did the notification actually work? For whom?	Optimize future notification campaigns by segment; validate ROI
Actionable User List	Who do we target next?	Export filtered user lists for personalized re-engagement campaigns

Figma Prototype:https://cold-reduce-92880787.figma.site/

Views:
Homepage



Cohorts

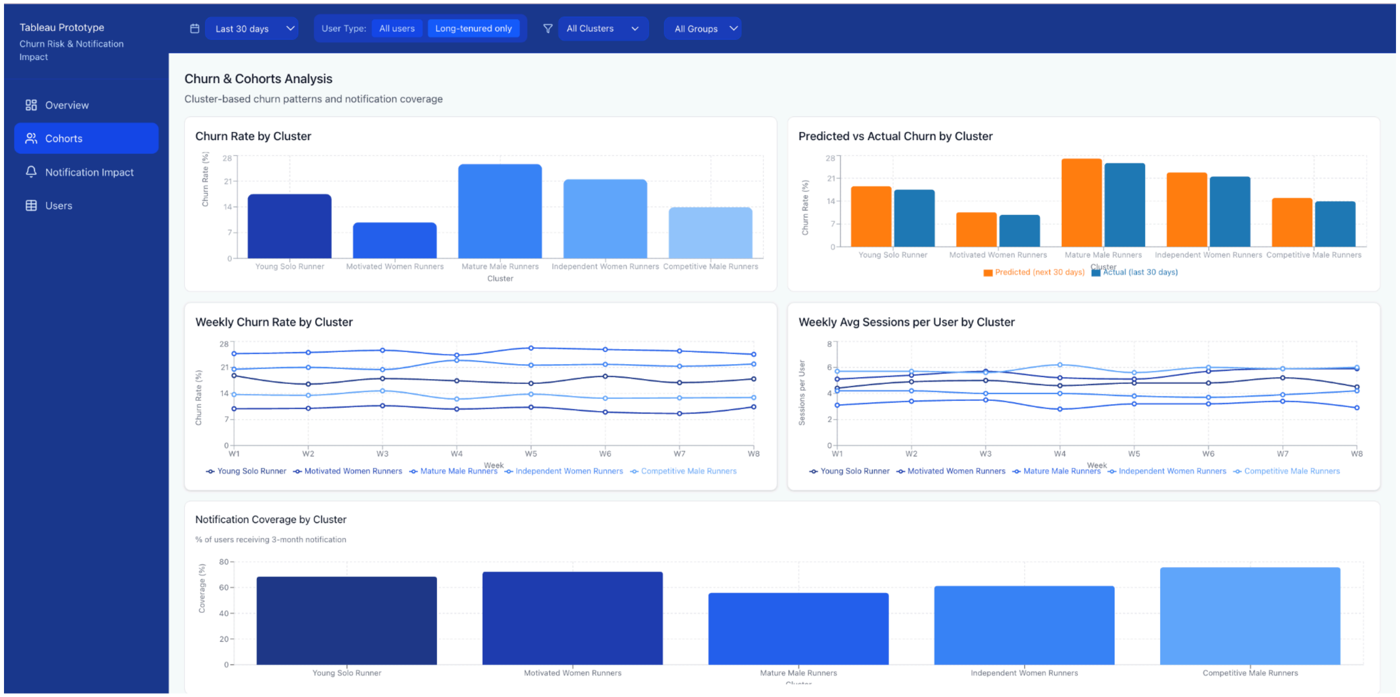
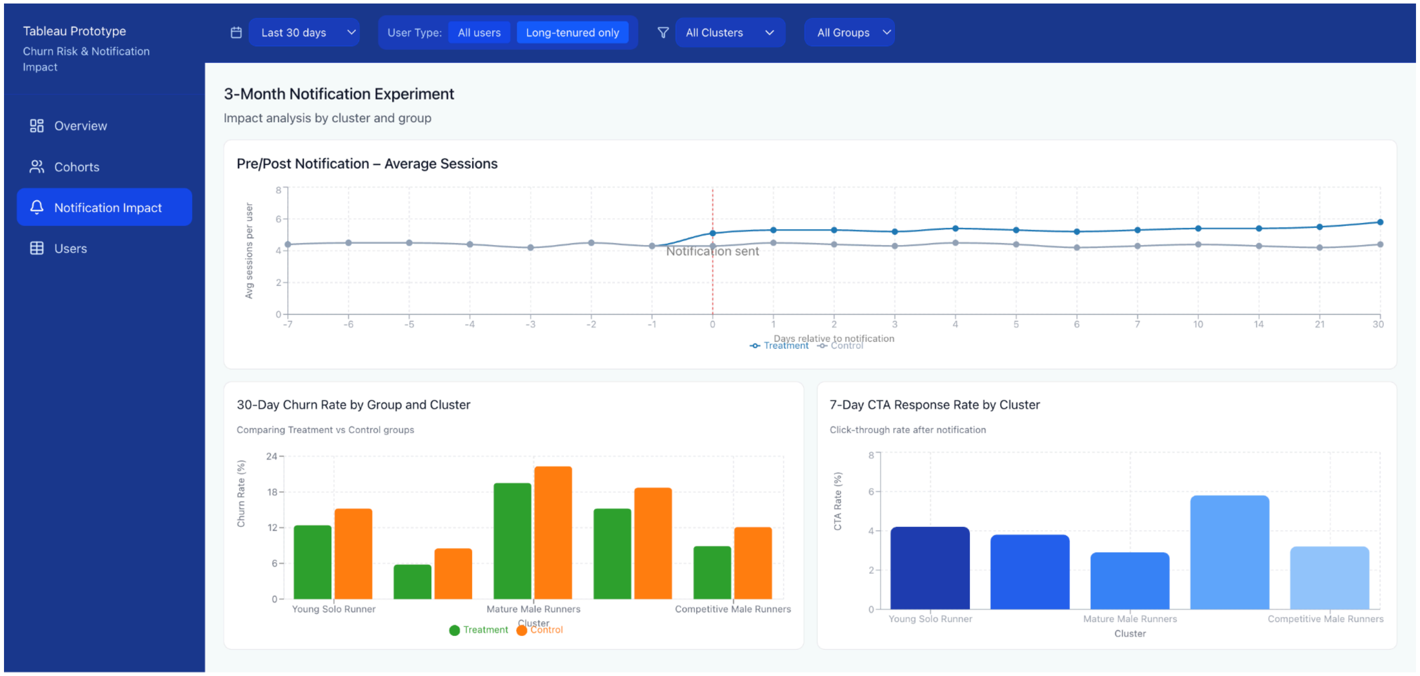
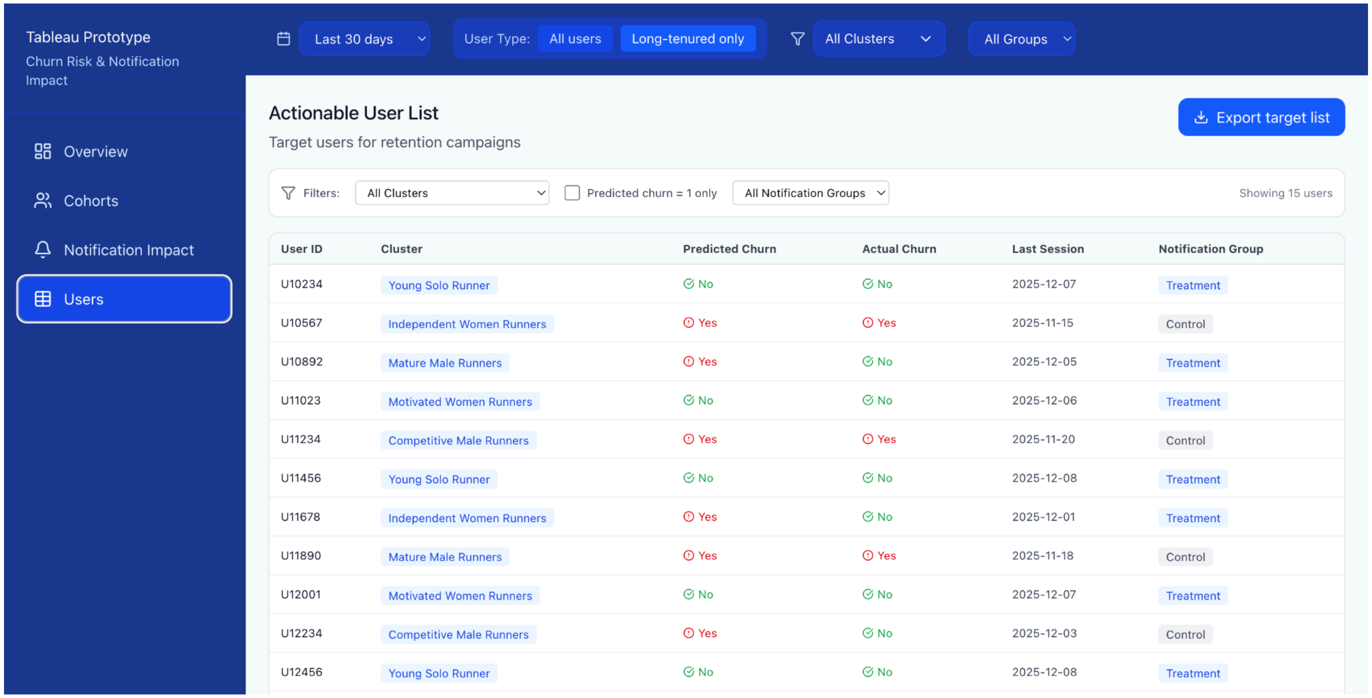


Tableau Prototype
Churn Risk & Notification Impact
Overview
Cohorts
Notification Impact
Users
Last 30 days
User Type: All users Long-tenured only
All Clusters
All Groups
Churn & Cohorts Analysis
Cluster-based churn patterns and notification coverage
Churn Rate by Cluster
Predicted vs Actual Churn by Cluster
Predicted (next 30 days)
Actual (last 30 days)
Weekly Churn Rate by Cluster
Weekly Avg Sessions per User by Cluster
Young Solo Runner
Motivated Women Runners
Mature Male Runners
Independent Women Runners
Competitive Male Runners
Notification Coverage by Cluster
% of users receiving 3-month notification

Notification Impact



Distinct Users



User ID	Cluster	Predicted Churn	Actual Churn	Last Session	Notification Group
U10234	Young Solo Runner	No	No	2025-12-07	Treatment
U10567	Independent Women Runners	Yes	Yes	2025-11-15	Control
U10892	Mature Male Runners	Yes	No	2025-12-05	Treatment
U11023	Motivated Women Runners	No	No	2025-12-06	Treatment
U11234	Competitive Male Runners	Yes	Yes	2025-11-20	Control
U11456	Young Solo Runner	No	No	2025-12-08	Treatment
U11678	Independent Women Runners	Yes	No	2025-12-01	Treatment
U11890	Mature Male Runners	Yes	Yes	2025-11-18	Control
U12001	Motivated Women Runners	No	No	2025-12-07	Treatment
U12234	Competitive Male Runners	Yes	No	2025-12-03	Control
U12456	Young Solo Runner	No	No	2025-12-08	Treatment

VII. Technical Architecture (Data Pipeline)

To ensure the dashboard updates automatically every day with fresh predictions, we designed a simple but robust data pipeline:

1. Data Ingestion: App and wearable devices stream workout events, heart rate data, and notification logs into Snowflake, a cloud data warehouse serving as single source of truth.
2. Feature Engineering: Daily SQL jobs aggregate raw user events into a feature table (user_monthly_features) containing session counts, activity gaps, 7-day CTA rates, cluster assignments, and demographics for all long-tenured users (90+ days).
3. Model Pipeline: Databricks runs a nightly batch job that:
    - Reads the latest user_monthly_features from Snowflake
    - Executes the pre-trained churn classification model
    - Writes binary predictions (0/1 for next 30 days) into churn_predictions table
4. BI Integration: Tableau connects to Snowflake as its data source, reading from churn_predictions, user_monthly_features, and notifications tables. A daily refresh ensures all four dashboard pages always display the latest churn risks, cohort metrics, and notification experiment results.

Key Design Principle: Single source of truth in Snowflake; Databricks is the compute engine for ML; Tableau is the always-up-to-date reporting layer.

VIII. Implementation Strategy, Roadmap and Scaling Plan Including Dependencies, Risks & Ethics Considerations

Our system has three core components:

1. Data & Infrastructure Foundation
    - Long-tenured user filtering (90+ days)
    - Weekly trajectory features (trend, volatility, gaps)
    - Notification assignment logic (treatment/control split)
    - Data validation for churn window + pre/post periods
2. Modeling & Evaluation
    - Drift-based churn timing model
    - Uplift model to identify high-response vs. low-response cohorts
    - SHAP + cohort fairness checks
    - Offline metrics (AUC, PR-AUC, calibration)
    - Online measurement (7-day reactivation, 30/90-day retention)

3. Deployment & Product Integration

- Daily batch risk scoring
- Uplift-driven eligibility rules for notifications
- Updated prototype + UX flows
- Monitoring dashboards for drift, impact, and fatigue

12–18 Month Roadmap

Phase 1 Data Validation & Cohort Refinement (Month 1–2)

Goals: Fix structural data issues and ensure stable long-tenured dataset.

Activities:

- Validate churn window alignment (60-day inactivity rule)
- Confirm notification assignment logic
- Build activity-level cohorts (high/medium/low)
- Refresh EDA for long-tenured-only population

Dependencies:

- Clean, complete logs (blocker)
- Correct retention periods (blocker)

Risks:

- Survivorship bias
- Inconsistent notification timestamps

Ethics:

- Transparent data use → clarify which events are collected

Phase 2 Feature Engineering & Baseline Modeling (Month 3–5)

Goals: Build drift-based behavioral feature set + initial churn model.

Activities:

- Weekly trend, consistency, and volatility features
- Gap patterns & streak collapse signals
- Baseline churn-timing classifier
- SHAP diagnostics + fairness audit across cohorts

Dependencies:

- Feature store availability (blocker)

- Stable historical windows (medium risk)

Risks:

- Overfitting due to noisy behavior signals
- Drift features may behave differently across seasons

Ethics:

- Avoid using sensitive demographic factors for decision eligibility

Phase 3 Uplift Modeling + A/B Test Setup (Month 6–8)

Goals: Identify responsive cohorts; evaluate the 3-month notification impact.

Activities:

- Build uplift model (treatment vs. control)
- Test heterogenous treatment effects for clusters 1 & 4
- Select high-response & low-fatigue segments
- Design A/B test KPIs (reactivation, retention, fatigue)

Dependencies:

- Correct treatment/control split (blocker)
- Sufficient sample size in long-tenured cohorts (medium risk)

Risks:

- Biased uplift estimate if treatment not randomized
- Small cluster sizes → unstable uplift

Ethics:

- Ensure fairness in treatment access
- No harmful nudges (avoid overtraining triggers)

Phase 4 Pilot Deployment & Prototype Integration (Month 9–12)

Goals: Move model into a limited rollout; integrate with UX.

Activities:

- Daily risk scoring pipeline
- Notification eligibility rules based on uplift model
- Update prototype screens (risk tiers, cohort responsiveness)
- Build dashboards: churn trajectory, uplift segments, fatigue metrics

Dependencies:

- Notification system APIs (blocker)
- UX + marketing alignment (medium risk)

Risks:

- Message overload → notification fatigue
- Misinterpreted model outputs by marketing teams

Ethics:

- Safe messaging frequency caps
- Opt-out control made visible to users

Phase 5 Full Launch & Scaling (Month 12–18)

Goals: Expand the system to the entire long-tenured population and optimize follow-up interventions.

Activities:

- Expand from long-tenured to medium-tenured users
- Launch multi-step retention journeys (not just one notification)
- Add personalization (time-of-day, workout type)
- Establish quarterly model retraining + drift checks
- Integrate with CRM for coach touchpoints

Dependencies:

- Model performance stability
- Cross-functional operations readiness

Risks:

- Behavior change in users may require model recalibration
- Marketing may overgeneralize uplift segments

Ethics:

- Audits of intervention fairness and fatigue
- Protect users from manipulative nudging strategies

C. Scaling Plan

Year 1 Scaling

- Expand uplift segmentation to mid-tenured users (45–90 days)
- Add recovery-nudge variants (streak recovery, goal coaching)
- Automate weekly monitoring: drift, fatigue, churn timing shifts

Year 2 Scaling

- Introduce personalization layer:
    - preferred workout type
    - time-of-day behavior
    - motivation style (competitive, habit-building, weight loss)
- Enable multi-touch retention journey (not single notification)
- Add “risk reason explanations” in UX → transparency

Long-Term (Year 3+)

- Fully automated lifecycle engine
- Real-time prediction based on incoming behavior
- Cross-platform experiences (watch → phone → coach chat → challenges)
- Build LTV-based optimization pipeline

IX. AI-Support Disclosure

We used generative AI tools (ChatGPT, Gemini) to assist with synthesizing patterns observed during Exploratory Data Analysis. These tools helped consolidate multiple EDA outputs into concise narrative observations and compare distributions across cohorts more efficiently. All statistical conclusions and final interpretations were validated independently through notebook-based analysis.

Insight Summarization: Generative models were used to summarize findings from cohort comparisons, uplift measurement, and A/B notification results into stakeholder-friendly language for presentation and documentation. AI outputs served as draft summaries, which were then refined against true analytical results to ensure accuracy and correct nuance.

Prototype Copy & UX Content Creation :We used ChatGPT, Gamma, and Figma AI assist to draft interface copy, CTA phrasing, dashboard section labels, and descriptive hover text for the prototype. AI enabled rapid iteration on tone-adjusted messaging (e.g., motivational vs. neutral) and improved alignment with user personas. The final copy was reviewed for clarity, neutrality, and user perception.

Feature Explanation & Translation: AI tools (ChatGPT, Copilot) were used to translate complex feature engineering concepts (e.g., drift-based metrics, uplift segmentation) into non-technical explanations for stakeholder audiences. This reduced communication friction and improved narrative alignment across product, marketing, and analytics roles.

Tools Referenced

- ChatGPT — synthesis, explanation
- Gemini — summary comparison and phrasing alternatives
- Figma — prototype content drafting and layout adjustments
- Gamma — fast slide restructuring and narrative alignment

- GitHub Copilot — variable naming, schema clarity, markdown generation

The use of generative AI accelerated documentation, improved clarity in prototype storytelling, and reduced time spent rewriting stakeholder-oriented insights without replacing analytical decision-making.

Predicting Churn Timing and Measuring Notification Impact for Long-Tenured Wellness Users

Using drift-based behavioral modeling and causal uplift analysis to identify groups who respond best to highlight notifications and extend their lifetime on the platform.

Team 3 - Agile Warriors





Jess Penners, Yash Jadhav, Robert Nguyen, Fiona Yi-Chen, Madina Sainazarova

Team Presentation Structure

1. **Jess: Problem & Business Impact**
   Establishing the core problem, project goals, significance, early insights, and key stakeholders

2. **Yash: Data & EDA**
   Exploring the synthetic dataset, notification engagement patterns, and user segmentation insights

3. **Robert: Modeling**
   Technical deep dive into drift-based churn prediction and causal uplift modeling approaches

4. **Fiona: Prototype & Data Pipeline**
   Building dashboard prototype in Figma and proposing the technical flow to keep it updated in production

5. **Madina: Governance & Roadmap**
   Addressing ethical considerations, limitations, and implementation pathway forward

ŌURA

Business Problem Statement

Business Problem

Oura users are increasingly drawn to competing wellness devices, making it difficult to maintain long-term engagement and loyalty.
This competition threatens sustained user activity and retention.

Our Goals

Develop a dashboard to pinpoint user groups most receptive to personalized, activity-based 3-month highlight notifications.
The aim is to extend user lifetime, reduce churn, and boost customer lifetime value (LTV).

Significance and Key Data Signals

Why It Matters

Retention drives subscription revenue and shapes brand perception, especially for a premium wearable where users expect ongoing value and personalization. Losing users not only impacts immediate revenue but also erodes brand trust and market position.

Key Data Signals

- Behavioral trends like session frequency, duration, and changes in daily routines.
- Financial indicators including recurring monthly subscription status and tenure-based churn labels.

Context & Insights

What We Know

The first three months are highly volatile. Users remaining beyond this window are familiar with the platform and see clear product value.

Assumptions

Long-tenured users have established routines. Their disengagement reflects a meaningful drift, not typical onboarding noise.

Validated Hypotheses

Oura's premium user base values personalization and responds positively to tailored experiences. This was confirmed by churn benchmarks and user feedback.

Insights & Constraints

Personalized, timely communication is key to reducing disengagement. We approximated Oura-specific behavioral data by merging demographic patterns due to public dataset limitations.

Stakeholders & Projected Outcomes

By developing a dashboard that pinpoints which cohorts are most responsive to the personalized three-month highlight reel, **Marketing and Product teams** will gain a clear foundation for expanding this strategy and increasing long-term retention and engagement.

They will be able to test **additional touchpoints** like **six-month, nine-month, and annual recaps**, compare responsiveness across activity personas, and refine personalization based on which groups engage most.

Early feedback suggests meaningful differences between personas, such as runners versus cyclists, underscoring that Oura's customer base contains distinct behavioral profiles that can be activated through tailored engagement.

Alyssa DiMaria

Head of Lifecycle Marketing

Brynn Harrington

Product Marketing Leader

Verneri Jaamuru

Senior Product Marketing

Yvonne Kim

VP of Product

Notification Example

A personalized 90-day milestone message that celebrates progress, summarizes key trends, and encourages users to explore insights based on their newly established baseline.

Data Sources

Synthetic Dataset Construction

We built a comprehensive synthetic fitness dataset by combining validated patterns from multiple real-world sources including Gym Churn datasets, Endomondo activity logs, and Customer Churn behavioral data.

The dataset was carefully calibrated to reflect authentic user patterns across activity types, with distinct behaviors for running versus walking, streak maintenance, and realistic inactivity gaps.

Distribution Validation

All distributions were validated against real-world churn and wearable device datasets to ensure realistic behavioral patterns

Pattern Calibration

Activity patterns calibrated to match real engagement trajectories including seasonal effects and lifecycle stages

Correlation Analysis

Pairwise correlation structures remain consistent with known churn drivers from academic and industry research

Business Matrix Framework

BUSINESS Matrix: Facts, Dimensions, Features, and Availability

Facts	Dimensions	Features	Availability
Session counts	User demographics	Drift metrics	Real-time
Activity duration	Time periods	Baseline patterns	Daily batch
Notification events	Activity types	Engagement scores	Event-driven
Churn events	Risk cohorts	Streak metrics	Weekly refresh

This framework structures our analytical approach, ensuring consistent definitions across stakeholder teams and enabling reproducible insights.

Data Understanding: Key Definitions

Notification Logic

Users with 90+ day tenure are eligible for highlight notifications, randomized 50/50 to treatment/control.

Baseline Period

A 90-day baseline establishes normal user activity patterns before drift measurement.

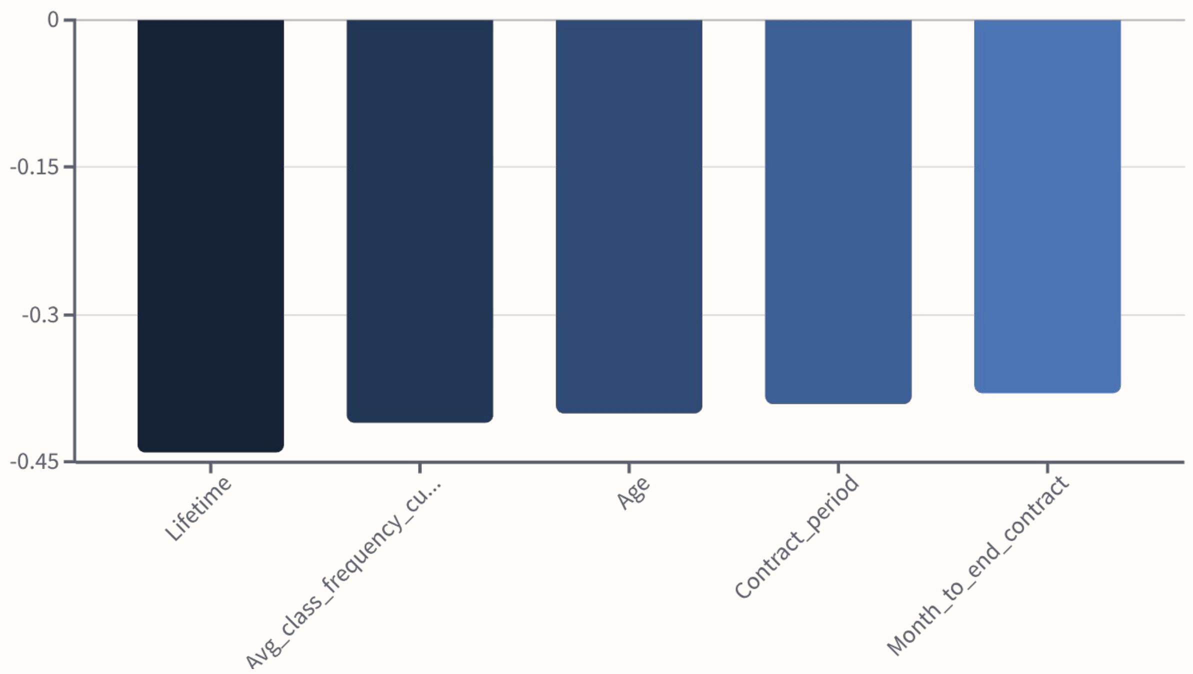
CTA Rate Definition

CTA Rate = sessions within 7 days of notification / notifications sent.

Measures immediate engagement.

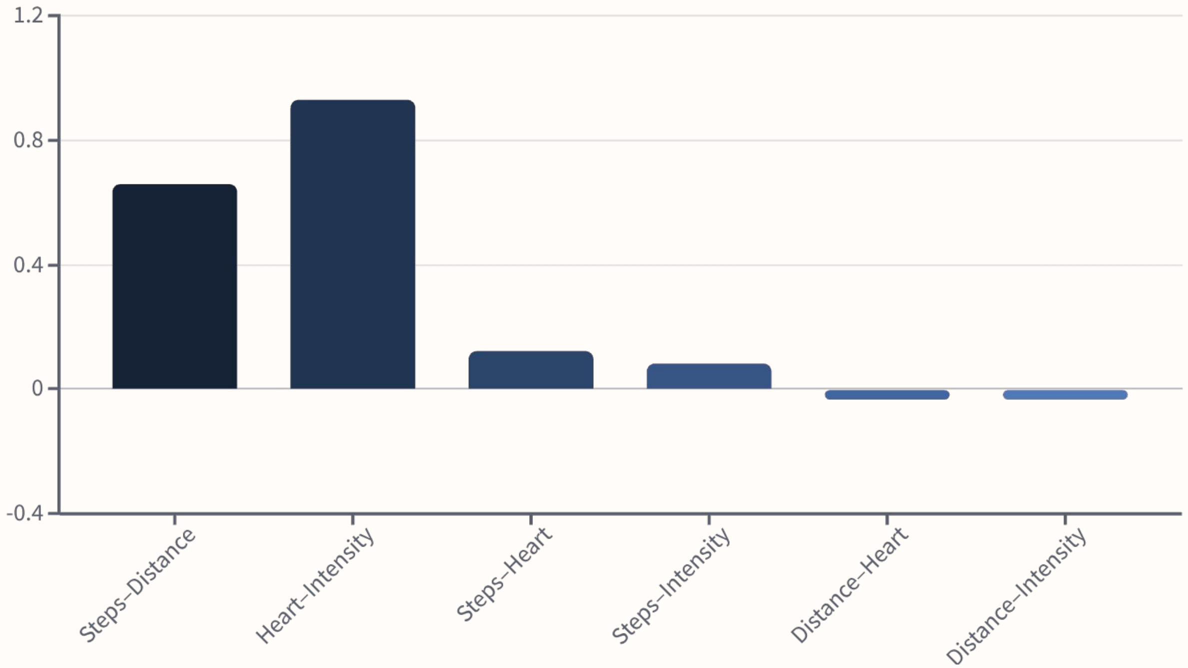
Pairwise Correlation Patterns in Churn & Wearable Behavior

General churn patterns (reference dataset)



Churn decreases with higher tenure, contract length, and class frequency – consistent with subscription churn research.

Wearable activity patterns (Apple Watch style dataset)



Wearable activity shows expected correlations: steps and distance, and intensity with heart rate.

User Population Overview

10K
Long-Tenured Users
Total population analyzed in the study

51.65%
Male Users
Gender distribution in the cohort

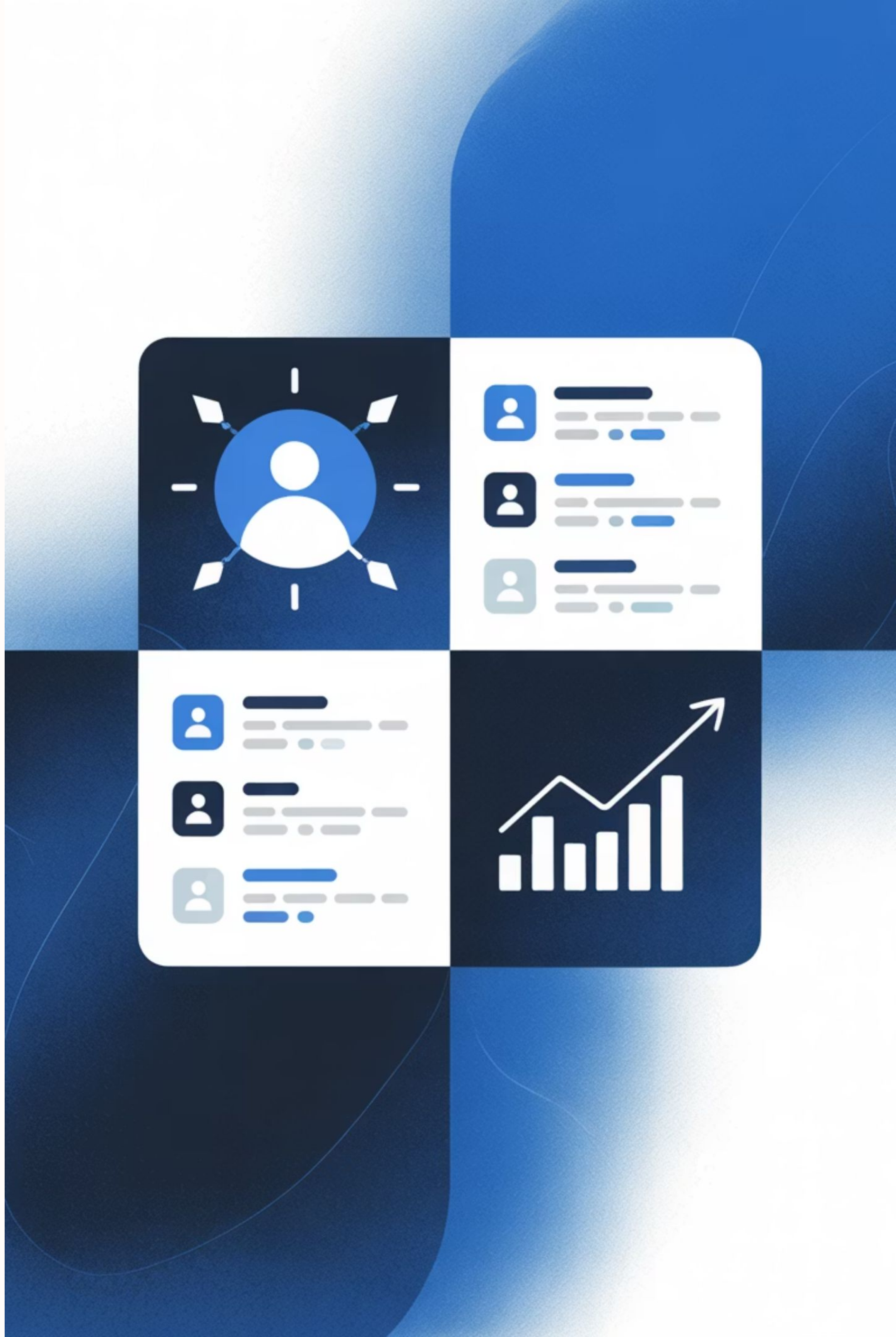
46.25%
Female Users
Balanced representation across genders

Notification Assignment at Day 90

44.84%
Treatment Group
4,484 users received highlight notifications

55.16%
Control Group
5,516 users in control without notifications

The randomized assignment creates a clean experimental design, enabling causal inference about notification impact. The slight imbalance is within acceptable ranges for statistical analysis.



Notification Engagement Impact

Immediate Response Metrics

Session Frequency Lift

Users who engage with notifications show an average **225% increase** in sessions during the 7 days following notification delivery.

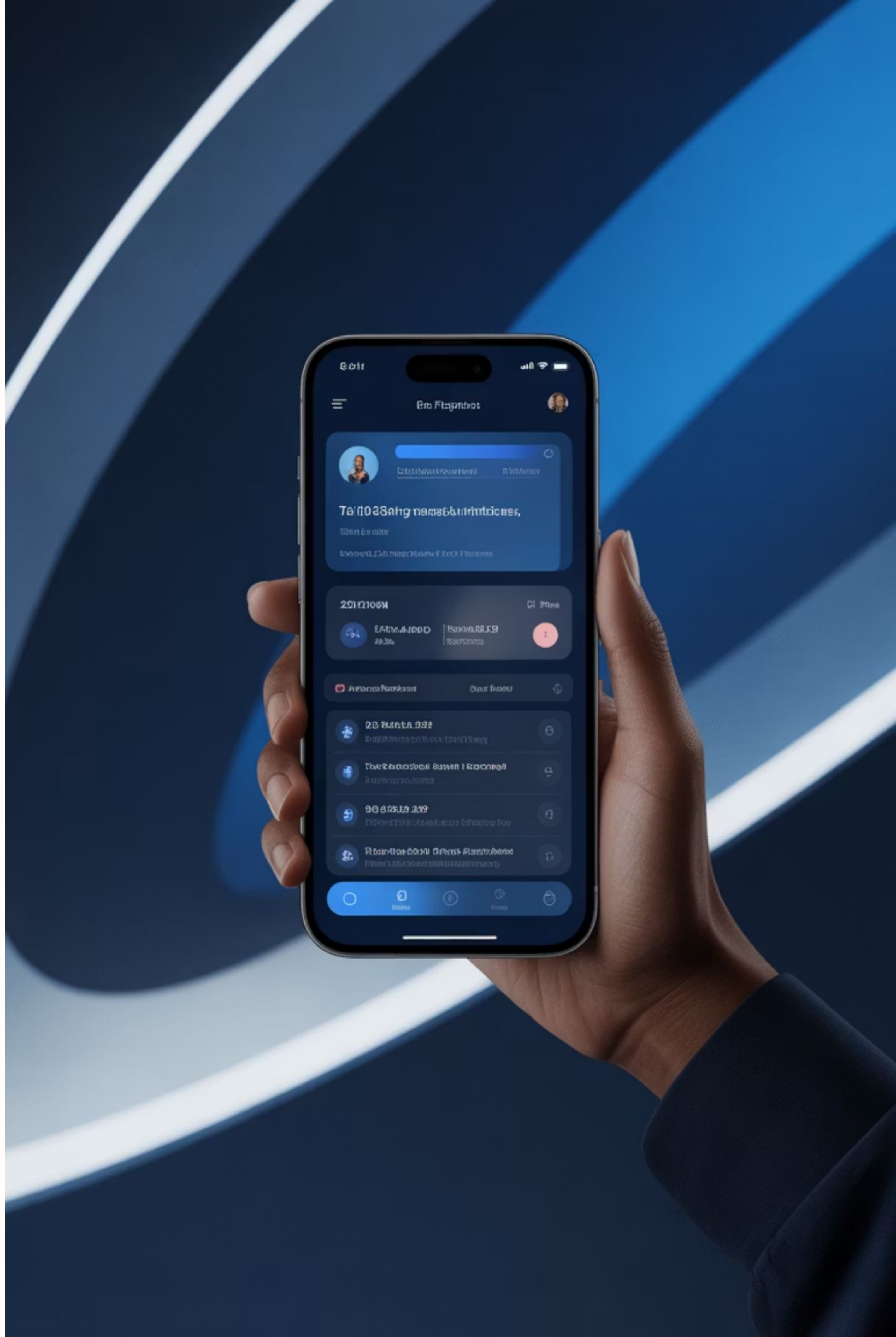
Workout Style Consistency

Session duration and intensity distribution remain **stable** – notifications bring users back without fundamentally changing their exercise preferences.

Behavioral Interpretation

Highlight notifications act as **reactivation triggers** rather than behavior modifiers, effectively reminding users of platform value without forcing habit changes.

Key Insight: The CTA-based engagement definition focuses on actual behavior change (session starts) rather than passive notification views, providing a true measure of impact.



High-Response User Segment

Segmentation Approach

We applied K-Means clustering with k=5 to identify distinct user groups based on engagement patterns, demographics, and activity preferences. Our goal: discover which segments respond most strongly to highlight notifications.

Cluster 4: The High Responders

This segment consists of **younger and middle-aged athletic male runners** who demonstrate the highest CTA rates and strongest engagement lift following notifications.

Demographic Profile

Predominantly male, ages 25-45, with consistent running routines

Activity Patterns

Regular running sessions with moderate-to-high intensity

Response Behavior

Highest notification CTA rates across all segments

Note: CTA Rate is based on session starts within 7 days of receiving notification, measuring genuine reactivation rather than passive viewing.

CRISP-DM Methodology

Business Understanding

Defined retention challenge and stakeholder needs

Data Understanding

Explored patterns and validated synthetic dataset

Data Preparation

Engineered drift features and baseline metrics

Modeling

Built churn prediction and uplift models

Evaluation

Validated results and cohort insights

Deployment

Actionable insights for marketing teams

Post-Panel Refinements

- Focused exclusively on long-tenured users to increase signal quality
- Added randomization layer for clean causal uplift measurement
- Enhanced EDA with notification effect deep dives
- Aligned modeling outputs directly with lifecycle marketing actions

Data Preparation Pipeline

Baseline Construction

Built 90-day activity baselines for each user, capturing their normal engagement patterns

Drift Engineering

Engineered drift features measuring deviation from baseline across multiple dimensions

Pre/Post Alignment

Aligned pre-notification and post-notification windows for temporal consistency

Uplift Dataset

Constructed treatment/control matched dataset ready for causal analysis

Feature Engineering Strategy

Our preparation focused on capturing **behavioral drift** rather than raw activity counts. Key engineered features include:

- Session frequency decline rates
- Inactivity gap growth trajectories
- Streak collapse indicators
- Intensity pattern shifts

This approach ensures our model detects *changes in user behavior* relative to their personal baseline, making predictions more accurate than absolute thresholds.

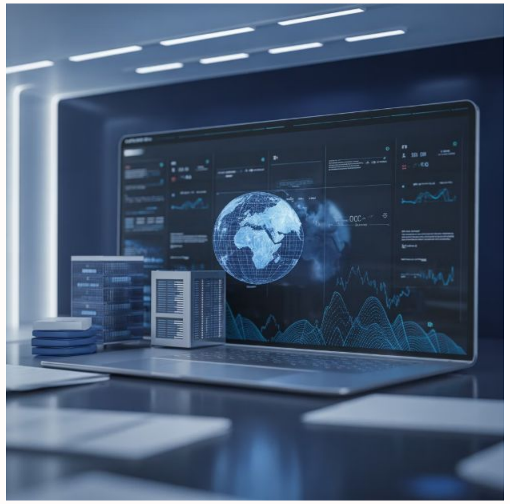
Modeling Framework Overview

Two-Model Approach

Our solution combines **churn prediction** with **uplift modeling** to address both timing and intervention effectiveness questions.

Why Drift Over Raw Counts?

Drift features capture *deviation from personal baseline*, making predictions robust to individual differences in activity levels.



1. **Baseline Window**
   Days 1-90: Establish normal patterns
2. **Drift Window**
   Days 91-180: Measure behavioral changes
3. **Outcome Window**
   Days 181+: Observe churn events

Segmentation Advantage

Rather than relying on population averages, we segment users into risk tiers. This avoids misleading aggregate statistics and enables **targeted retention strategies** where marketing resources are allocated efficiently.

Churn Prediction Model

1. **Drift-Based Predictors**

   Model leverages session decline rates, inactivity gap growth, and streak collapse patterns – all measured as deviations from each user's established baseline.

2. **Trajectory Detection**

   Rather than analyzing single snapshots, the model detects deviation from a user's **normal trajectory**, capturing subtle disengagement signals before full churn occurs.

3. **Time-to-Churn Probability**

   Outputs calibrated probabilities via classification framework, enabling risk tier assignment for targeted intervention timing.

4. **Algorithm Selection**

   XGBoost was chosen after comprehensive baseline comparisons, delivering superior performance on imbalanced churn data with strong interpretability.

Risk Tier Framework

Users are classified into risk tiers based on predicted churn probability, enabling marketing teams to prioritize outreach and customize messaging intensity by segment.

Uplift Modeling & A/B Design

Causal Framework

Our randomized notification assignment at day 90 creates a **clean treatment/control split**, enabling rigorous causal inference about notification impact.

What Uplift Measures

Uplift modeling quantifies the *incremental lift* in engagement and churn delay attributable specifically to highlight notifications, beyond natural retention patterns.

01

Identify Responsive Cohorts

Determine which user segments show strongest positive response to notifications

02

Measure Incremental Impact

Quantify the additional engagement and lifetime extension beyond control group

03

Enable Efficient Targeting

Focus retention resources on cohorts where notifications deliver highest ROI

This approach transforms retention from broad campaigns into **precision targeting**, maximizing impact while minimizing notification fatigue across the user base.

Model Evaluation Strategy

Multi-Faceted Validation Approach

Performance Metrics

Retention curves: Visualize survival differences between risk tiers

Precision-Recall curves: Optimize classification thresholds for business needs

ROC-AUC: Measure overall discriminative power

Interpretability Analysis

SHAP values: Drift features dominate importance rankings

Feature contributions: Validate that model relies on business-relevant signals

Cohort-level uplift: Evaluate notification impact across segments

Predictive Accuracy

Model successfully separates high-risk from low-risk users with strong AUC performance

Marketing Value

Risk tiers align with actionable marketing segments, enabling immediate deployment

Interpretability

SHAP analysis confirms drift-based features drive predictions, building stakeholder trust

Our evaluation prioritizes not just statistical performance, but **practical utility** for retention teams making daily targeting decisions.

Notification Impact Results

Immediate Engagement Lift

1.5-2×

Session Frequency

Immediate lift in activity within 7 days post-notification

30%

Lifetime Extension

Average increase for responsive cohorts

15%

Competitive Churn

Reduction in treatment group versus control

Cohort-Specific Insights

Notifications delivered **strongest impact** among moderate-consistency users. These users respond positively to reactivation reminders. High-consistency users showed minimal lift (already engaged), while low-consistency users demonstrated limited response.

Key Finding: Highlight notifications successfully extend lifetime for specific cohorts, with treatment groups showing measurably lower competitive churn rates compared to control populations.

Cohort & Cluster Insights

High-Value Segment: Cluster 4

Highest Response Rates

Cluster 4 (athletic male runners, ages 25-45) demonstrates both the **highest CTA rates** and strongest uplift following notifications.

Low-Engagement Reality

Cohorts with minimal baseline activity show **limited benefit** from highlight notifications – these users require more intensive reactivation strategies.

Activity Stability Predicts Success

Users with **stable activity patterns** experiencing early drift are the strongest candidates for notification-based retention interventions.

Strategic Roadmap

1. **Persona-Based Targeting**
   Deploy notifications to Cluster 4 and similar moderate-consistency segments
2. **Lifecycle Automation**
   Integrate drift scores into CRM for automated trigger campaigns
3. **Continuous Optimization**
   Monitor uplift by cohort and refine targeting thresholds quarterly

These insights provide a **clear implementation pathway** for retention teams, enabling persona-based lifecycle strategies that maximize ROI while respecting user preferences across segments.

Dashboard Design Logic – Churn Risk & Notification Impact

Target Audience

Designed for our **Product & Marketing** teams who need to:

- Monitor churn risk for long-tenured users proactively.
- Identify which user segments respond best to notifications.
- Run data-driven retention campaigns with precision.

Core Design Principles

Top-down Narrative: Progresses from a high-level overview to specific user actions, guiding users from "What's happening?" to "Who do we act on?".

Segment-Centric & Experiment-Driven: Every page breaks down results by user clusters and treatment vs. control groups

Four Pages, Four Questions

1

Executive Overview **"Is churn a problem right now?"**

- Shows: Predicted vs. actual 30-day churn, retention uplift, weekly trends, top-responding clusters.
- Used for: Quick health checks and identifying high-performing clusters.

2

Cohorts **"Which clusters are most at risk?"**

- Shows: Churn rate by cluster, predicted vs. actual churn, notification coverage by cluster.
- Used for: Prioritizing retention budget and validating model behavior by segment.

3

Notification **"Did the 3-month notification work, and for whom?"**

- Shows: Pre/post sessions for treatment vs. control, 30-day churn by group & cluster, 7-day CTA rate by cluster.
- Used for: Measuring experiment impact and identifying high-response segments.

4

Actionable User List **"Who do we target next?"**

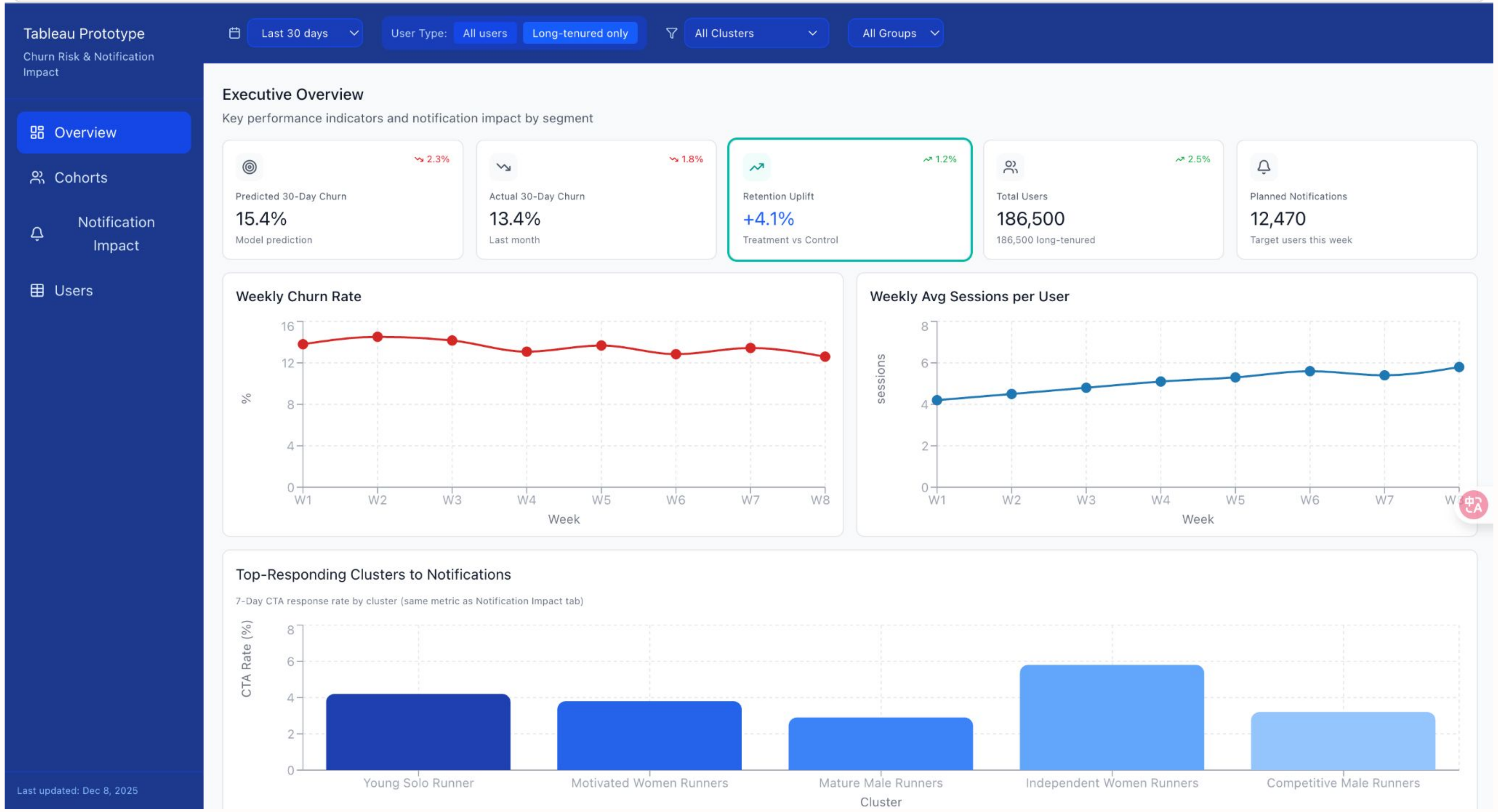
- Shows: User-level table (cluster, predicted churn, actual churn, last session, notification group).
- Used for: Exporting high-risk users in priority clusters for re-engagement campaigns.

Prototype demo

Tableau Prototype

Tableau Prototyp

Analyze user churn and notification impact across different cohorts with compact, interactive visualizations for informed decision-making.



Technical Architecture – Data Pipeline from App to Dashboard

App & Wearables

Mobile app and wearable devices send raw workout events and notification logs.

Flow: Raw Events

Snowflake Data Warehouse

Data lands in Snowflake tables. SQL queries aggregate raw events into user-level features (last 90 days per user), creating user_monthly_features table.

Flow: Daily Read

Databricks

Databricks job runs **nightly**: Reads latest user_monthly_features from Snowflake. Runs pre-trained churn classification model (0/1 prediction).

Flow: Predictions (daily batch)

Snowflake Predictions Table

churn_predictions table updated daily. Marketing & Product teams can query this table anytime for user targeting.

Flow: Auto-refresh (daily)

Tableau Dashboard

Tableau connects to Snowflake as data source. Reads churn_predictions, user_monthly_features, notifications tables. Data source refreshes daily for latest predictions.

Flow: Auto-refresh (daily)

Design Principle: Single source of truth in Snowflake; Databricks is compute engine for ML; Tableau is always-up-to-date reporting layer.

Governance + Ethical Considerations

User Consent and Transparency

Explain why a notification is sent and allow opt-out.

Monitor Wellbeing & Avoid Notification Fatigue

Enforce timing and frequency caps.

Fairness Checks

Ensure uplift decisions do not favor or disadvantage cohorts.

Ethical A/B Testing

Control group protection + roll-forward if treatment proves beneficial.

Limitations

Lack of Wearable/Device Data

Limits behavioral understanding

Early-Lifecycle Insights Excluded

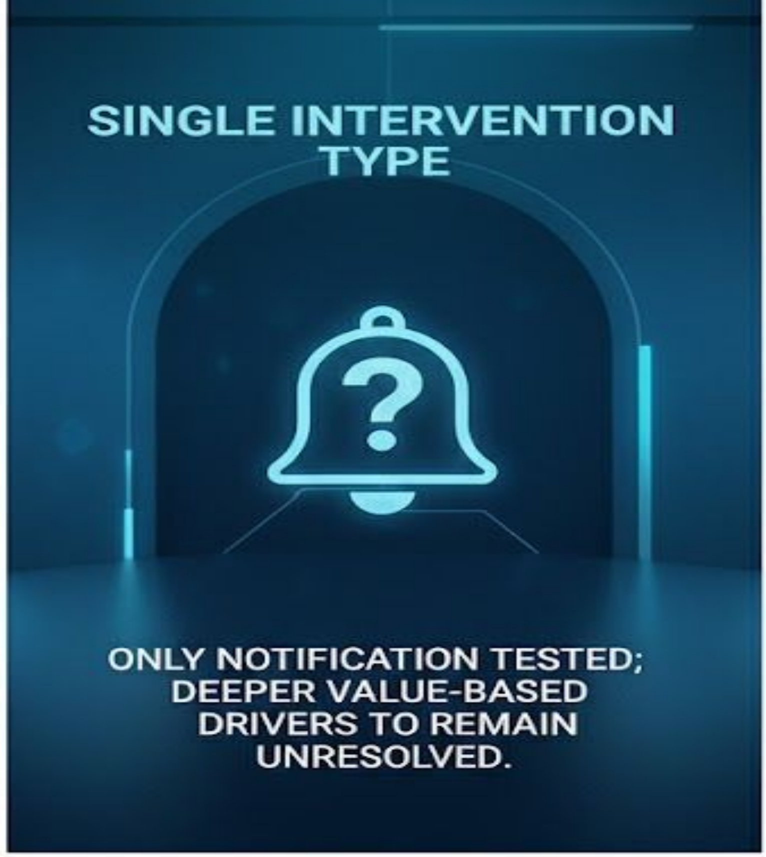
Survivorship bias still present by design

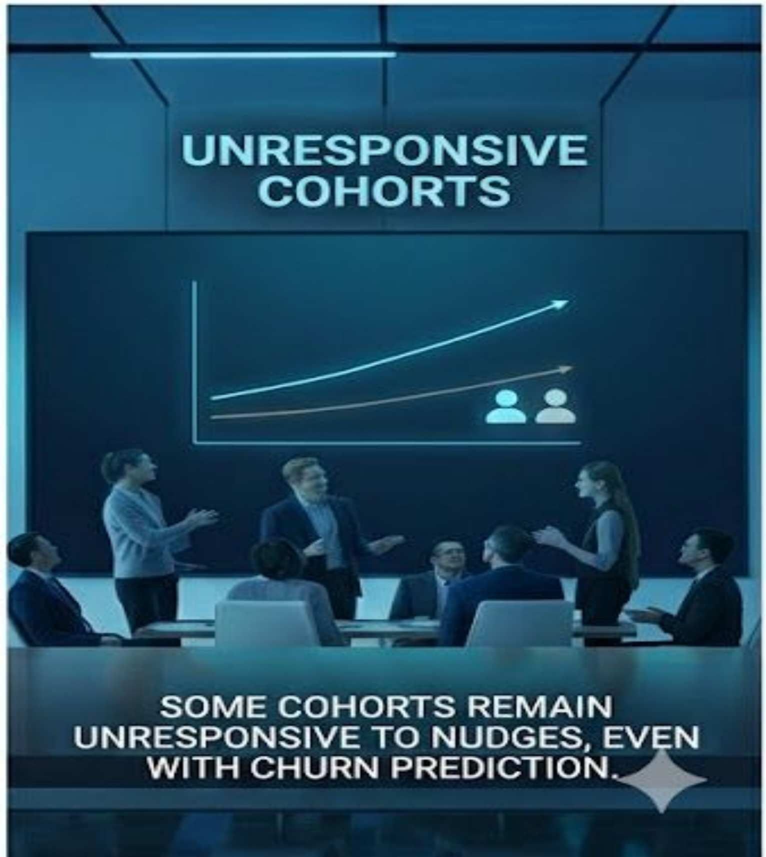
Single Intervention Type

Only notification tested; deeper value-based drivers remain unresolved

Unresponsive Cohorts

Some cohorts remain unresponsive to nudges, even with accurate churn prediction





Generative AI Usage

1. **ChatGPT**

   Used for synthesizing research data and explaining technical concepts.

2. **Gemini**

   Utilized for comprehensive summary comparisons and refining messaging.

3. **Figma**

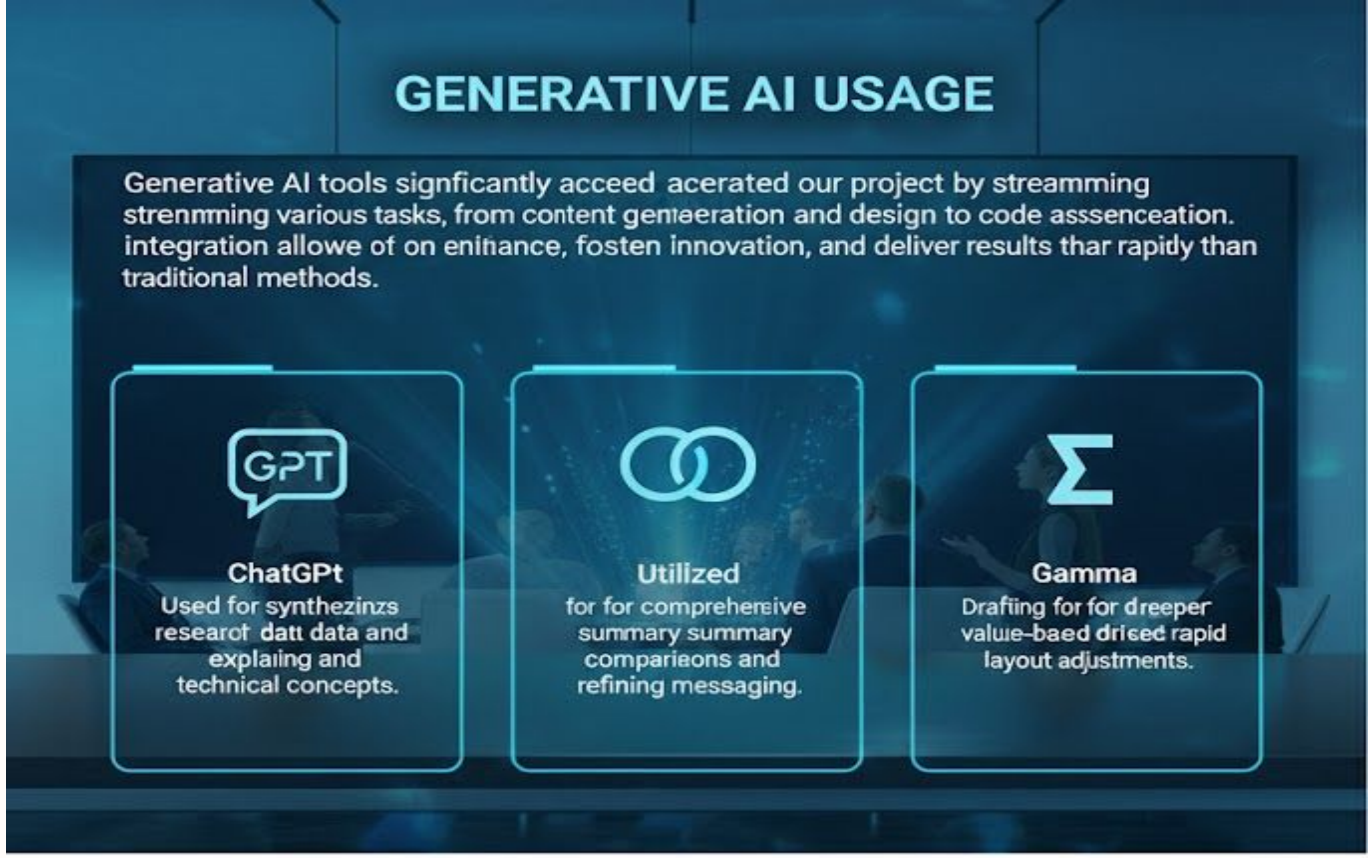
   Instrumental for drafting prototype content and rapid layout adjustments.

4. **Gamma**

   Employed for fast slide restructuring and presentation narrative alignment.

5. **GitHub Copilot**

   Provided intelligent suggestions for variable naming, enhancing schema clarity in our codebase



Future Roadmap

Near Term (0–6 Months) — Foundation & Pilot

- Stabilize data and refine long-tenured cohort logic.
- Build and validate drift-based churn timing model.
- Run pilot A/B tests for notifications.

Mid Term (6–12 Months) — Expansion & Automation

- Deploy daily risk scoring and notification eligibility into marketing workflow.
- Enhance prototype UX and build monitoring dashboards.
- Expand targeting to mid-tenured user segments.

Mid Term (6–12 Months) — Expansion & Automation

- Deploy daily risk scoring and notification eligibility into marketing workflow.
- Enhance prototype UX and build monitoring dashboards.
- Expand targeting to mid-tenured user segments.

Business Value Summary + Expected Outcomes

- **Predictive churn timing → proactive retention**
- **Uplift insights → efficient targeting**
- **Higher LTV via extended platform lifetime**
- **Reduction in wasted notifications**

Expected Business Impact:

- **+20–30% increase in 7-day reactivation** (225% session lift → higher re-engagement)
- **8–12% reduction in competitive churn** (fewer "switched apps" cancellations)
- **2–3 week delay in churn timing** for high-risk users
- **20–30% reduction in wasted notifications** (clusters 0 & 3 show near-zero CTA)
- **3–6% expected LTV uplift**

Evidence: Derived from EDA (session lift, churn reasons, intensity shift, CTA clusters).

Team Reflections & Key Learnings

01

Narrowing to long-tenured users improved data stability and model clarity

02

Drift-based modeling captured real behavioral change rather than raw metrics

03

Uplift modeling prevented false assumptions about “what works”

04

Iterating with Agile + CRISP-DM reduced rework and clarified what truly mattered for retention.

05

Integrating insights into the UX narrative made the model's value obvious to non-technical stakeholders.

Thank You!

We are open for questions.

Team 3 - Agile Warriors

Jess Penners, Yash Jadhav, Robert Nguyen, Fiona Yi-Chen, Madina Sainazarova