Churn Analytics Retention - BRD

> Business Requirements Document

Project Name: Streaming Subscription Churn Analytics
Company: Drillinsight
Version: 1.0
Date: 2025-12-04
Author: Drillinsight Data & Analytics Team

---

1. Document Control

Item	Detail
Document Title	Streaming Subscription Churn Analytics BRD
Version	1
Status	Draft for Review / Approval Ready
Owner	Head of Data Analytics, Drillinsight
Author	Data Analytics Lead
Date Created	12/4/2025
Last Updated	12/4/2025

0.1 Revision History

Version	Date	Author	Description
1	12/4/2025	Data Analytics Lead	Initial BRD for stakeholder review

0.2 Approvals (To Be Signed)

Role	Name	Signature	Date
VP, Product			
Director, Growth / Marketing			
Head of Data & Analytics			
Engineering Manager (Data)			

---

2. Purpose

The purpose of this document is to define the **business requirements** for the **Streaming Subscription Churn Analytics** initiative at Drillinsight.

This BRD:

- Describes the **business problem** and **context** for churn analysis in a subscription streaming business.
- Defines the **scope**, **objectives**, and **success metrics** for the project.
- Specifies **functional** and **non-functional** requirements for the data, analytics, and reporting solution.
- Identifies **assumptions**, **constraints**, **dependencies**, and **risks**.
- Serves as the basis for **design, implementation, and stakeholder approval**.

This document is intended to be sufficiently detailed so that data engineering, analytics, and BI teams can implement the solution without ambiguity.

---

3. Background & Business Context

3.1 Company & Product Context

Drillinsight is supporting a **subscription-based streaming service** offering on-demand video content (series, movies, live TV add-ons) via monthly recurring subscriptions.

The streaming product offers:

- **Subscription plans:** Basic, Standard, Premium.
- **Feature add-ons:**
  - Multi-profile / multi-device access.
  - Parental controls.

    - Cloud watchlist & progress synchronization.
    - Live TV add-on.
    - Premium Movies add-on.
    - Premium support / phone support.
- **Access channels:** Mobile, desktop, tablet, TV devices.

3.2 Business Problem

Over the last **three billing quarters**, the business observed a **material increase in churn**:

- Historical average monthly churn: **~4.8%**
- Recent three-month average monthly churn: **~5.7%**
- Relative increase: **~18–20%**

This has significant impact on:

- **Monthly Recurring Revenue (MRR)** and annual revenue forecasts.
- **Customer Lifetime Value (CLV)**.
- **Unit economics** of paid acquisition (CAC vs LTV).
- Investor and executive confidence in the subscription model.

At the same time:

- Customer acquisition costs (CAC) have increased year-over-year.
- Market competition from other streaming providers has intensified.
- Content and infrastructure costs have risen.

As a result, **improving retention and reducing churn** is now a top-priority initiative.

---

4. Business Objectives

The project will deliver a **data-driven churn analytics capability** to answer:

1. **Who is churning?**
    - By plan, cohort, demographics, device usage, and add-on subscriptions.
2. **When do users churn?**
    - At which lifecycle stages (first month, 3–6 months, >12 months).
    - Which signup cohorts (e.g., March 2024) exhibit unusually high churn.
3. **Why are they churning?**
    - Behavioral drivers (drop in engagement, silent users).

    - Pricing and plan mix.
    - Feature adoption / add-on combinations.
4. **What can we do to reduce churn?**
    - Which users should be prioritized for retention campaigns.
    - Which product and UX changes can improve retention.

4.1 Measurable Objectives

Within **3 months** after the solution is adopted:

- Reduce the **monthly churn rate** by **0.5–1.0 percentage points** (e.g., from 5.7% to ⩽5.0%).
- Improve **Month-1 retention** (M1 retention) by **3–5 percentage points** for new cohorts targeted by interventions.
- Increase **contribution of "high engagement" users** (as % of active subscriber base) by **5+ percentage points**.
- Provide a **churn prediction model** achieving:
    - **Recall ⩾ 0.75** on churners at a reasonable precision (⩾ 0.3) for top risk deciles.
- Enable Growth/CRM team to generate **monthly prioritized retention lists** (top 10–20% at-risk users) based on model and rules.

---

5. Scope

5.1 In Scope

The following areas are **in scope** for this initiative:

1. **Data Modeling & Integration**
    - Use three core analytical tables:
        - `users` (demographics / profile).
        - `subscriptions` (subscription lifecycle & add-ons).
        - `user_events` (behavior logs).
    - Build analytical views / tables for:
        - Monthly churn and retention.
        - Signup cohorts.
        - Behavior-aggregated features (logins, sessions, events).
2. **Analytics & Modeling**

- Descriptive analytics:
  - Retention curves (cohort-based).
  - Churn breakdown by plan, add-on, segment.
  - Behavioral drop-off analysis prior to churn.
- Predictive modeling:
  - Baseline churn risk model (Logistic Regression).
  - Enhanced model (XGBoost) including behavior & add-on features.
- Explainability:
  - Feature importance analysis (e.g., SHAP values) for top drivers.

3. **BI & Reporting**
   - KPI dashboards (Tableau/Power BI), including:
     - Monthly churn & retention by cohort.
     - Plan / add-on segmentation.
     - Engagement & activity trends.
   - Ad-hoc query support for Growth, Product, and CS teams.
4. **Data Documentation & Governance**
   - Data dictionary for `users`, `subscriptions`, `user_events`.
   - Clear churn and retention definitions.
   - Data quality rules and monitoring for critical metrics.
5. **Operationalization (Light-weight)**
   - Scheduled data refresh (daily or weekly).
   - Regular churn reporting to stakeholders.
   - Monthly export of high-risk user lists for campaigns.

5.2 Out of Scope

The following are **explicitly out of scope** for this phase:

- Real-time model scoring and in-app personalization.
- Deployment of models into production microservices.
- End-to-end marketing attribution and CAC optimization.
- Detailed content-level analysis (e.g., recommendations by title/genre).
- Experimentation platform / A/B test infrastructure implementation.

These may be addressed in subsequent phases.

---

6. Stakeholders & Roles

Role	Responsibility
VP, Product	Sponsor, approves scope and prioritization
Director, Growth / Marketing	Uses outputs for retention and campaign strategy
Head of Data & Analytics	Owns analytical methodology and project delivery
Data Analytics Lead	Leads analysis, modeling, and KPI design
Data Engineer	Implements data pipelines, tables, and performance tuning
BI Developer	Builds dashboards and reporting views
Customer Success Lead	Consumes churn insights and high-risk lists for outreach
Finance Representative	Validates revenue / CLV impact assumptions

---

7. Current State

7.1 Data & Reporting

- Basic churn reporting may exist in spreadsheets or ad-hoc queries.
- There is **no unified, trusted churn dashboard** with:
  - Cohort analysis.
  - Behavior-driven segmentation.
  - Model-based risk scoring.
- Data is available from:
  - Subscription billing system.
  - Authentication / streaming event logs.
  - Account profile database.

However, it is not yet organized into a consistent **analytical data model**.

7.2 Pain Points

- Conflicting churn numbers across teams (Finance vs Product).
- No clear understanding of:
  - Which segments drive the recent 18–20% churn increase.
  - Early warning signals before churn.
- CRM and Growth teams manually pull lists using static rules (e.g., “no login in 30 days”), which:
  - Miss many high-risk users.
  - Target some low-risk users unnecessarily.

---

8. Future State Overview

In the target state, Drillinsight will have:

1. A **single source of truth** for churn, retention, and engagement KPIs.
2. A **standardized cohort and retention framework** for all new campaigns and features.
3. A **churn risk scoring capability** that generates:
   - Ranked list of at-risk subscribers.
   - Explanations of top drivers (e.g., “low recent watch activity”, “no add-ons”, “Short tenure plan”).
4. **Self-serve dashboards** for Product, Growth, and CS teams.

This enables:

- Data-driven prioritization of retention campaigns.
- Product experiments to improve onboarding, engagement, and add-on adoption.
- Revenue impact estimation for churn-reduction initiatives.

---

9. High-Level Use Cases

UC-1: Monitor Churn & Retention

Growth and Product teams can:

- View monthly churn rate by cohort, plan, and region.
- Identify months where churn spiked (e.g., specific cohorts like 2024-03).
- Compare churn rate of subscribers with vs without specific add-ons.

UC-2: Segment High-Churn Cohorts

Analyst can:

- Identify customer segments with >8% monthly churn (e.g., Basic plan, no add-ons, mobile-only devices).
- Quantify revenue impact of these segments.

UC-3: Detect Behavior-Driven Risk Patterns

Analyst can:

- Compare behavior metrics (logins, watch events) of churned vs retained users.
- Quantify behavior drop in the 30 days and 7 days before churn.

UC-4: Generate Retention Target Lists

Churn model outputs a ranked list of users with:

- Churn probability ⩾ defined threshold.
- Top behavioral / profile drivers attached.

Growth team exports top deciles (e.g., top 10–20% highest risk) as a **retention campaign audience**.

UC-5: Evaluate Intervention Impact

Over time, teams can:

- Track churn and retention trends for targeted cohorts vs control cohorts.
- Measure uplift in retention following targeted campaigns or feature rollouts.

---

10. Detailed Requirements

10.1 Data Model & Sources

The **grain** and **core tables** are:

1. `users` (User-level dimension)
    - Grain: 1 row per user.
    - Example fields:
        - `user_id` (PK)
        - `gender`

        - `senior_citizen`
        - `has_partner`
        - `has_dependents`
    - Business meaning: stable demographics & basic profile attributes.
2. `subscriptions` (Subscription-level fact table)
    - Grain: 1 row per active/historical subscription per user.
    - Example fields:
        - `user_id` (FK → users.user_id)
        - `signup_date`
        - `cancel_date` (nullable)
        - `plan` (Basic / Standard / Premium)
        - `price` (monthly fee)
        - `status` (active / cancelled)
        - Add-on indicators:
            - Multi-profile / device (mapped from `multiple_lines`)
            - Parental control (`online_security`)
            - Cloud sync (`online_backup`)
            - Device / account control (`device_protection`)
            - Premium support (`tech_support`)
            - Live TV add-on (`streaming_tv`)
            - Premium Movies add-on (`streaming_movies`)
        - `paperless_billing`
        - `total_charges` (with known deviations simulating discounts/promotions).
3. `user_events` (Event-level behavior log; stored as CSV / analytical table)
    - Grain: 1 row per user event.
    - Example fields:
        - `user_id`
        - `event_type` (login / watch / click / cancel)
        - `event_time` (timestamp)
        - `device_type` (mobile / desktop / tablet / TV)

        - `page_url`
    - Used for:
        - Daily / weekly activity metrics.
        - Behavior-based features (engagement scores).
        - Pre-churn trajectory analysis.

Data retention:

- Subscription data: at least **24 months** historical.
- Event logs: at least **12 months**, preferably 18–24 months.

10.2 Functional Requirements

FR-1: Churn & Retention Calculation

- FR-1.1: Implement standardized **churn user** definition: a user with non-NULL `cancel_date`
- FR-1.2: Implement **monthly churn rate** calculation:
    - Numerator: users whose `cancel_date` falls within the month.
    - Denominator: users with active subscriptions at the start of the month.
- FR-1.3: Implement **monthly retention**:
    - Users active in month N+1 / users active in month N.
- FR-1.4: Implement **cohort-based retention tables** (by signup month).

FR-2: Segmentation & Cohort Analysis

- FR-2.1: Enable churn and retention breakdown by:
    - Plan.
    - Add-on combination.
    - Device type (primary).
    - Demographic segments (age proxy via senior_citizen, partner/dependent flags).
- FR-2.2: Provide tables / views supporting comparison of **high-churn vs low-churn** segments.

FR-3: Behavioral Aggregations

- FR-3.1: Compute daily / weekly aggregation per user:
    - Number of logins.
    - Number of watch events.
    - Number of clicks.

- FR-3.2: Compute **behavior in 30 days and 7 days prior to cancel_date** for churned users.
- FR-3.3: Identify **silent users** (e.g., no events in last 60 days but still active).
- FR-3.4: Flag **abnormal users** with extremely high event counts.

FR-4: Churn Risk Modeling

- FR-4.1: Construct feature set including:
  - Tenure (days since signup).
  - Recent engagement metrics (7/30 day windows).
  - Plan & price.
  - Add-on indicators.
  - Billing behavior (if available).
- FR-4.2: Train baseline **Logistic Regression** model to classify churn vs non-churn within a defined horizon (e.g., next 30 days).
- FR-4.3: Train enhanced **XGBoost** model with the same target.
- FR-4.4: Evaluate model performance using:
  - AUC.
  - Recall & precision at top risk deciles.
- FR-4.5: Generate **scored user tables** with:
  - Churn probability.
  - Risk segment (e.g., High / Medium / Low).
  - Top contributing features (via SHAP or equivalent).

FR-5: BI Dashboards & Reporting

- FR-5.1: Provide interactive dashboard with:
  - Global monthly churn and retention trends (12–24 months).
  - Cohort retention heatmaps.
  - Churn by plan, add-on, and demographics.
- FR-5.2: Provide engagement analytics:
  - DAU/WAU/MAU trends.
  - Engagement distribution across segments.
- FR-5.3: Provide **churn driver** view summarizing key variables impacting churn.
- FR-5.4: Provide exportable views or CSV extracts for:

- High-risk user lists (for CRM tools).
- Summarized segment metrics (for presentations).

10.3 Non-Functional Requirements

- **Performance:**
  - Monthly churn & retention dashboards should refresh within **5–10 minutes** on scheduled runs.
  - Ad-hoc queries on aggregated tables should return within **30 seconds** for typical filters.
- **Data Quality:**
  - No orphaned subscriptions (`subscriptions.user_id` must exist in `users`).
  - Churn and retention metrics must reconcile within **±1%** of Finance' s verified numbers.
  - Data validation checks for:
    - Impossible dates (`cancel_date` < `signup_date`).
    - Violated dependency rules (e.g., internet_service = 'No' but streaming add-ons enabled).
- **Security & Privacy:**
  - No PII beyond necessary user identifiers used internally (e.g., hashed IDs).
  - Access to detailed user-level data restricted to authorized analysts.
- **Availability:**
  - Dashboards should be available during business hours with **99%+** uptime (excluding maintenance windows).

---

11. Assumptions & Constraints

11.1 Key Assumptions

- Behavior prior to churn follows realistic patterns:
  - ~50% event reduction in the 30 days before churn.
  - ~80% event reduction in the 7 days before churn.
- 5–10% of active users are "silent" in recent 60 days.
- <1% of users have abnormally high event volumes (heavy or bot-like usage).
- Add-on subscriptions influence churn probability (Premium feature users tend to churn less).

- Market and product conditions remain broadly similar during the first analysis window (no massive rebranding or pricing overhaul).

These assumptions are consistent with industry benchmarks and are necessary to:

- Simulate realistic churn behavior patterns.
- Provide enough signal for churn modeling and behavioral analysis.

11.2 Constraints

- Event data is available at least at **user-event granularity**, but not necessarily tied to specific content titles (content-level analysis is out of scope).
- Real-time or sub-minute latency is not required; daily/weekly batch processing is acceptable.
- Infrastructure choices (MySQL vs cloud warehouse etc.) may be defined by existing Drillinsight or client stack and are not fully controlled by the analytics team.

---

12. KPIs & Success Metrics (Business-Facing)

This BRD references a separate **KPI Definition Document**, but key KPIs are:

- **Monthly Churn Rate** (primary KPI).
- **Monthly Retention Rate** (cohort and global).
- **M1 / M3 / M6 Retention** for new cohorts.
- **Engagement Metrics:**
  - % of users with activity in last 7 / 30 days.
  - Average events per active user.
- **Model Performance:**
  - AUC.
  - Recall @ top X% at-risk.
- **Business Impact:**
  - Reduction in monthly churn (% points).
  - Incremental retained revenue.
  - CLV uplift estimates for targeted cohorts.

Targets:

- Reduce monthly churn by **0.5–1.0 percentage points**.
- Improve M1 retention by **3–5 percentage points** in targeted segments.

- Achieve **Recall ⩾ 0.75** for churn prediction on top-risk cohorts.

---

13. Dependencies

- Access to subscription, user, and events data in the source systems.
- Data Engineering capacity to:
  - Build ingest and transformation pipelines.
  - Optimize storage and indexing for analytical workloads.
- BI licenses (Tableau / Power BI) for dashboard consumption.
- Alignment with Finance on baseline churn and retention calculations.
- Collaboration with Growth / CS teams for:
  - Validating high-risk user lists.
  - Implementing retention campaigns.

---

14. Risks & Mitigation

Risk ID	Description	Impact	Likelihood	Mitigation Strategy
R1	**Data quality issues** in event logs (missing or delayed events).	High	Medium	Implement data validation checks; flag days with incomplete data; document gaps.
R2	Misalignment on **churn definition** between teams (Product vs Finance).	High	Medium	Lock definitions in KPI document; obtain explicit Finance sign-off before rollout.
R3	Model performance lower than expected due to weak signal or noise.	Medium	Medium	Start with interpretable baseline models; iterate on features; use behavior windows.
	Over-targeting			Use thresholds

R4	Over-targeting customers based on noisy model outputs (customer fatigue).	Medium	Medium	+ business rules; cap outreach frequency; monitor opt-out rates.
R5	Underestimation of engineering effort for data pipelines & optimization.	High	Low–Medium	Phase delivery; start with limited history; optimize progressively.
R6	Changes in pricing or packaging during analysis period distort metrics.	Medium	Low	Track major product/ pricing changes; annotate dashboards; analyze separately by era.
R7	Stakeholders misinterpret model as deterministic decision engine.	Medium	Medium	Provide clear documentation and explainability; position model as decision support.

---

15. Implementation Approach & Timeline (High Level)

Assuming **project kickoff = T0**, target timeline:

1. **Phase 1 – Data Foundation (~2–3 days max)**
   - Add missing data quality checks
   - Implement base retention/churn SQL views using existing schema
   - Validate event-time ranges & subscription consistency
2. **Phase 2 – Analytics & Insights (Weeks 1-2)**
   - Retention and cohort analysis.
   - Segmentation by plan, add-on, and behavior.
   - Behavior drop-off and silent user analysis.
   - Draft initial “Top Insights” report.
3. **Phase 3 – Modeling & Scoring (Weeks 3-4)**

- Feature engineering (behavior, tenure, add-ons).
- Train baseline Logistic Regression model.
- Train XGBoost model; evaluate performance and interpretability.
- Produce initial scored user list and driver analysis.

4. **Phase 4 – BI & Operationalization (Weeks 5-6)**
   - Build production BI dashboards.
   - Define retention campaign workflows using high-risk lists.
   - Handoff, documentation, and stakeholder training.

---

16. Acceptance Criteria

The project will be considered successfully implemented when:

1. A **single, agreed-upon churn and retention metric** is available in the dashboard and reconciled with Finance.
2. Stakeholders can self-serve:
   - Cohort retention analyses.
   - Churn breakdown by key segments.
3. A churn model is in place with:
   - Documented feature set and performance.
   - Regularly updated risk scores.
4. Growth / CS teams actively use:
   - Monthly high-risk user lists.
   - Dashboards for decision support.
5. At least **one full cycle** of retention campaign has been run using model-driven lists, and pre/post metrics are captured.