How Listening Behavior Predicts Retention

Spotify®

Using Data Science to Predict and Prevent Spotify User Churn

The Business Challenge

- User churn **reduces revenue** and increases marketing costs
- Replacing users who churn is **expensive** in a crowded music streaming market
- Retaining existing subscribers maximizes lifetime value and loyalty

GOAL : Identify users at high risk of leaving before they churn!

The Data Science Solution

- Use **unsupervised learning** (clustering, networks) to segment users by behavior
- Use **supervised learning** tools (logistic regression, random forest, LASSO, etc.) to predict churn probability
- Identify key churn drivers (listening time, ad exposure, subscription type)

GOAL : Deliver targeted retention strategies → reduce churn, optimize marketing, increase loyalty

Understanding Our User Data

Overview:

- 2025 Spotify analysis dataset from **Kaggle**
- Each record = one user

Limitations:

- **Synthetic** dataset
- Use results to identify **general churn trends**, not exact real-world predictions

Key Variables

Demographics	age, gender, country
Subscription	subscription_type, device_type, ads_listened_per_week
Engagement	listening_time, skip_rate, songs_player_per_day

Preparing Clean, Reliable Data For Churn Prediction

Data Quality and Consistency Check

- No missing values
- No duplicates
- Reasonable ages
- Skip rate between 0-1
- Listening time / songs per day are not negative

Outliers & Correlated Variables

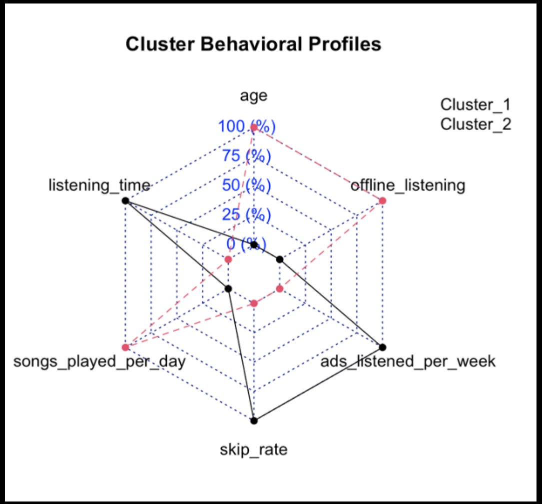
- Ad exposure differs by subscription type
- Offline listening is categorical
- Strong correlations between subscription_type, ads_listened_per_week, and offline_listening

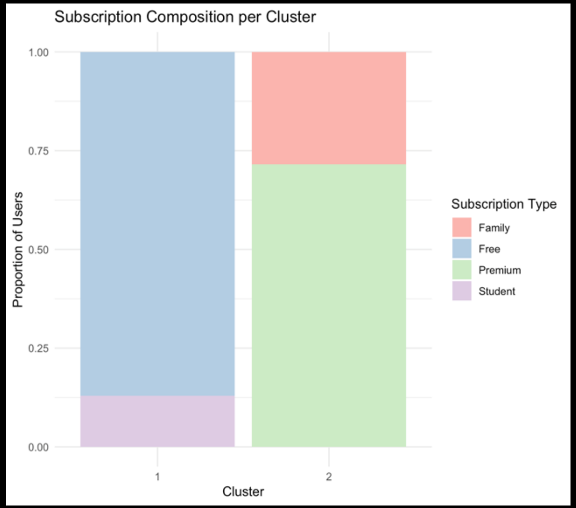
Summary Stats

- Overall churn rate = 53.88%
- Avg listening time = 148.8 hrs/week
- Avg songs per day = 32.1
- Avg skip rate = 28.7%
- Churn by subscription: Free (72.6%), Student (57.8%), Family (34.1%), Premium (29.6%)

Spotify User Segmentation via K-Means Clustering

Goal: Identify distinct user segments based on listening behavior, engagement, and subscription type using **unsupervised learning (k-means clustering)**





Key Findings (For k=2)

- **Cluster 1:** Free / Student : high skip, low engagement, higher churn (~65%)
- **Cluster 2:** Premium / Family : steady engagement, lower churn (~30%)

Insight Discovery
Identified two distinct user personas:
- Explorers (Free/Student)
- Committed Subscribers (Premium/Family)

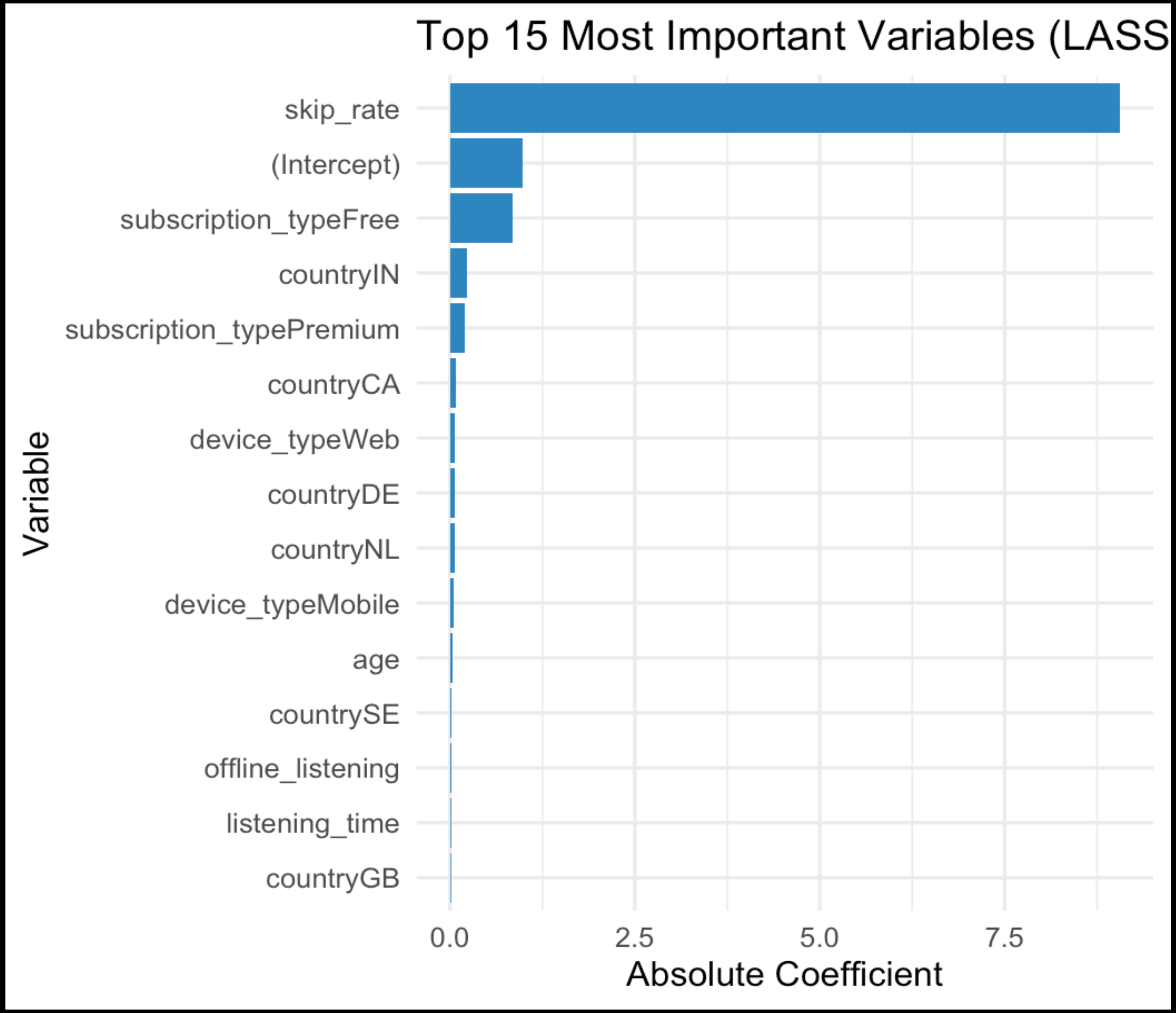
Behavioral Understanding
Observed differences in engagement, skip behavior, and churn patterns (behavioral vs. situational churn)

Strategic Application
- Targeted retention campaigns for Explorers
- Improve recommendation algorithms
- Loyalty programs for Subscribers

Modeling Approach: Finding the Best Predictor of User Churn

Model	Why?	Key Findings
Logistic Regression	Establish baseline relationships	Found key churn predictors (age, listening time, skip rate, offline listening)
Logistic Regression + Interactions	Capture cross-variable effects	Improved deviance but overfit
LASSO Logistic Regression	Feature selection and high accuracy	AUC = 0.889, Accuracy 0.80
Classification Tree	Easy interpretation	Top splits: offline listening, skip rate
Random Forest	Strong performance	Confirmed our top predictors
Neural Network	Capture nonlinear patterns	Low interpretability

LASSO Model: Key Insights Driving Churn Prediction



Top Predictors:

- High skip rate → Increases churn likelihood
- Low listening time → Increases churn likelihood
- Offline listening enabled → Reduces churn
- Premium subscription → Strongly reduces churn
- Country (Germany, India) → Higher churn

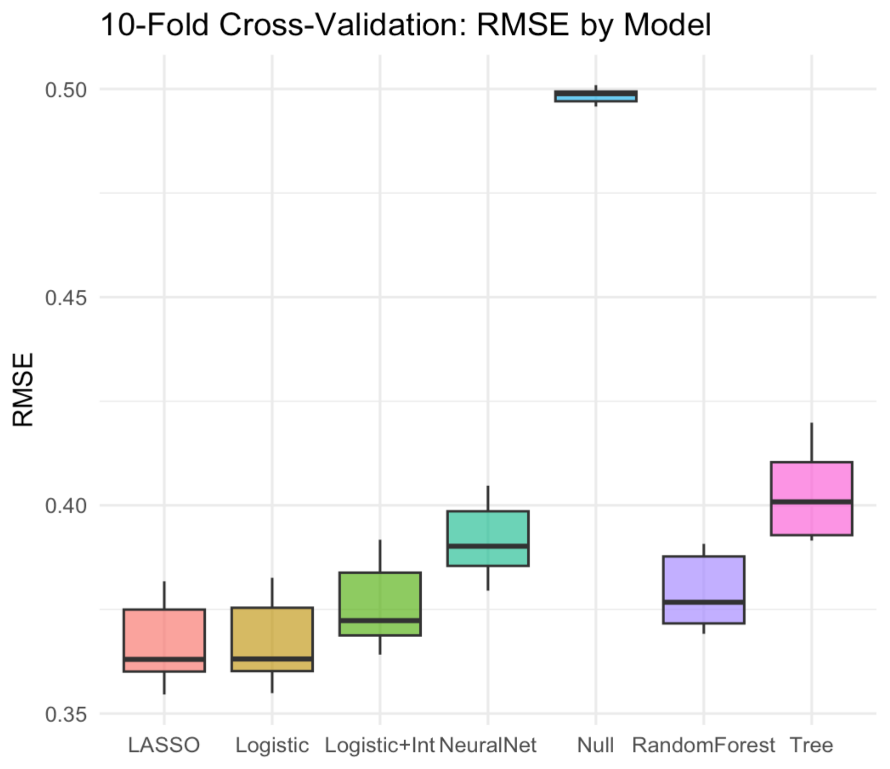
Business Implications:

- Encourage **offline listening and longer engagement** sessions
- Retain **Free and high skip-rate** users with discounts or personalized playlists
- Target churn-prone regions (Germany, India) with **localized campaigns**

Evaluation of Churn Prediction Models

10-Fold Cross-Validation with four metrics

- **AUC** (ranking churners)
- **RMSE** (prediction precision)
- **Log-loss** (penalizes overconfidence)
- **Accuracy** (overall correctness)



Key Result:

- **LASSO Logistic Regression** performed best
- AUC = 0.889, RMSE = 0.37, Accuracy = 0.80

Why LASSO Wins:

- Strong prediction **and** interpretability
- Automatically selects key predictors: **skip rate**, **free-tier use**, **low listening time**
- Helps Spotify understand ***why*** users churn, not just ***who***

Business Impact

- Pinpoint top 30% high-risk users → target retention offers, playlists, or discounts
- Reduces wasted marketing spend through data-driven personalization

Deployment : Free & Student Listeners

Expose at-risk free users to Premium experiences convert satisfaction into commitment

Rationale

- Price-sensitive and less committed users
- High skip rates, low total listening time
- Churn driven by *music relevance and satisfaction*, not value perception
- Free-tier recommendation system limits personalization improvements
- Opportunity: convert high-churn users by *exposing Premium experience*

Deployment

- **Churn Prediction Integration:** Flag users with churn probability > 0.7 for proactive engagement
- **Targeted Offers**: Send short Premium trial offers or curated playlist refreshes to re-engage
- **Experience Sampling:** Use refreshed playlists to showcase Premium algorithm quality
- **Algorithmic Focus:** Improve specific recommendation areas (e.g., *Daily Mix*, *Discover*) for Free users

Deployment : Premium and Family Listeners

Keep loyal users engaged by reinforcing value and not discounts

Rationale

- High engagement but still show measurable churn
- Churn linked to *value perception* and *pricing fatigue*, not content
- Small churn % large revenue impact
- Requires richer behavioral & lifecycle data for early detection
- Retention priority group for sustaining long-term value

Deployment

- **Enhanced Features:** Integrate behavioral & lifecycle data e.g. tenure, renewal history, device diversity, payment stability
- **Early Warning System**
  - P(churn) > **0.7** → automatic loyalty campaign
  - P(churn) > **0.4** → engagement reinforcement track
- **Retention Tactics:** Loyalty incentives: “3 months at a discounted rate”, Pause options: “Take a break, your playlists will wait”. Usage reminders: “You listened to 1,200 minutes this month”
- **Regional Targeting:** Promote mobile adoption or flexible billing in high-churn regions (e.g., Sweden, India)