Final Report

Netflix Customer Churn:

Identifying the Behavioral Drivers of User Retention

Executive summary

This project analyzed Netflix customer behavior to identify drivers of churn and develop predictive models. While baseline models struggled, XGBoost slightly improved over random classification (PR AUC = 0.153 vs 0.148 baseline). Engagement, discovery, satisfaction and early tenure emerged as the strongest behavioral drivers of churn. Based on these insights, we recommend strengthening onboarding, monitoring satisfaction proactively and focusing retention spend on high-value at-risk users.

1. Introduction

Netflix faces a persistent challenge with customer churn that directly erodes recurring revenue and limits upselling opportunities. Nowadays, the streaming industry has become increasingly competitive, driving up acquisition costs while consumers experience growing subscription fatigue. In this environment, retaining existing subscribers (particularly those with high lifetime value) proves far more cost-effective than constantly acquiring new customers.

Without clear insights into which behaviors distinguish loyal users from those likely to cancel, retention campaigns risk being overly broad and ineffective. Fortunately, Netflix's comprehensive behavioral dataset, encompassing watch history, search patterns, reviews, and recommendation interactions, presents a valuable opportunity to conduct an in-depth churn analysis.

This project pursues three primary objectives: developing a churn prediction model able to identify at-risk users despite significant class imbalance, uncovering the key behavioral drivers that differentiate retention from departure and delivering actionable recommendations to implement targeted strategies.

2. Data wrangling

Our analysis leveraged the Netflix 2025 User Behavior Dataset, containing approximately 210 000 records distributed across six interconnected tables: users, watch_history, search_logs, reviews, movies and recommendation_logs.

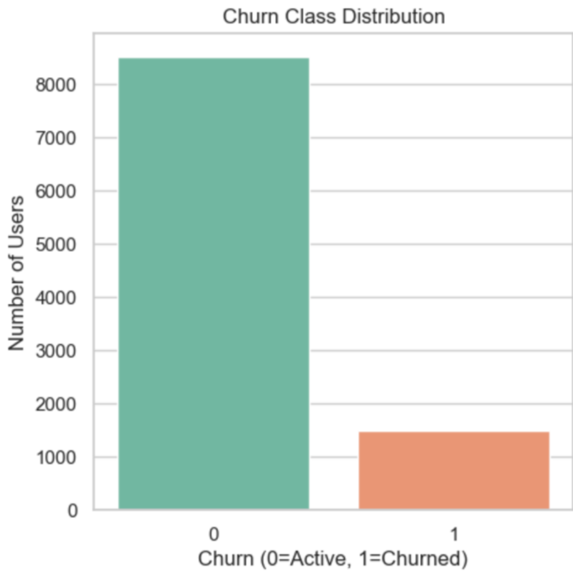
First, we began by preparing and comprehensively cleaning the data, including removing duplicates across all fields and systematically handling missing values through median imputation for numeric features and “Unknown” categorization for sparse fields. Highly incomplete metrics were filled with zeros to maintain data integrity.

We then engineered sophisticated user-level features, including total and average watch time, completion rates, search activity metrics, review sentiment scores and recommendation click-through rates. A binary churn label was created where inactive users (is_active=False) were coded as churned (churn=1) and user tenure was calculated in days from subscription start.

These engineered features were consolidated into a clean dataset (user_features.csv) containing 10 000 users across 33 carefully crafted variables, forming the foundation for our modeling efforts.

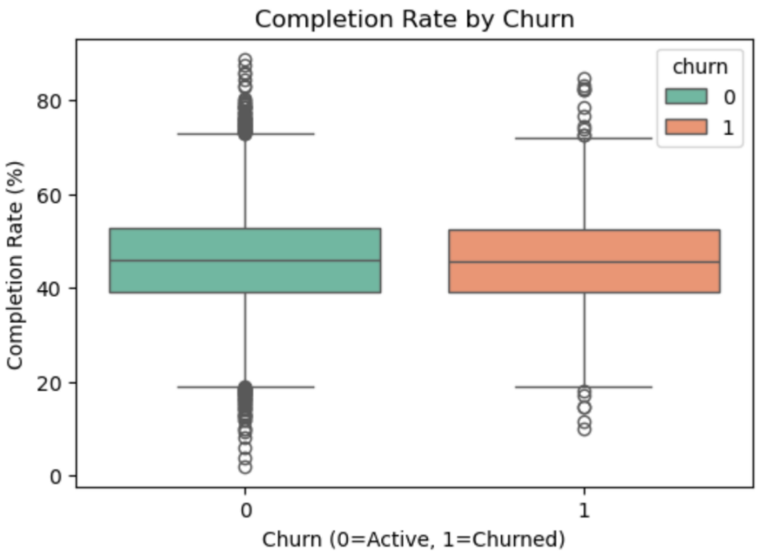
3. Exploratory Data Analysis

The initial analysis revealed a significant class imbalance, with 85% of users remaining active while only 15% had churned (a pattern that would heavily influence our modeling approach).



The exploratory phase uncovered behavioral patterns that clearly distinguished churned from retained users. Churn risk peaked dramatically among new subscribers, with the majority of cancellations occurring within the first 90 days of subscription. This finding indicates the importance of the onboarding period.

Engagement emerged as the most powerful predictor of retention: users who have higher watch times, more frequent viewing sessions and greater content completion rates demonstrate much lower churn rates.



Similarly, the active use of discovery features (including search activity and recommendation interactions) strongly correlates with user loyalty. On the other hand, churned users largely ignore recommendations and show much lower search frequencies.

User satisfaction provided another critical dimension: subscribers who consistently left lower ratings or wrote reviews with negative sentiment showed higher churn probabilities. On the contrary, demographic variables and subscription preferences had minimal predictive power: churn rates only vary by 14-15% across different plans.

These patterns demonstrate that behavioral indicators prove to be effective in predicting churn risk.

4. Modeling

Our modeling framework treated churn as the target variable while carefully removing potential data leakage sources such as user identifiers and raw timestamps. The preprocessing pipeline included one-hot encoding for categorical variables, median imputation for missing values and standardization for numerical features. We applied an 80/20 training-testing split while preserving the original churn distribution.

Due to the significant class imbalance, we prioritized Precision-Recall AUC (PR AUC) as our primary evaluation metric, alongside ROC AUC, recall, precision and F1 scores.

Baseline performance

The Dummy Classifier, which consistently predicted the majority class, established our baseline with a PR AUC of 0.148.

Logistic Regression with balanced class weights achieved 41% recall but suffered from low precision (12%), resulting in a PR AUC of 0.133 (actually underperforming the baseline).

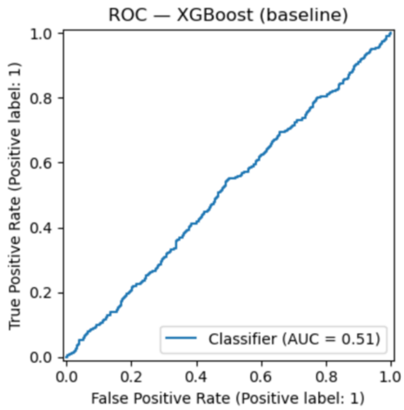
The Random Forest model proved even less effective, completely collapsing to predict only the majority class with zero recall for churners.

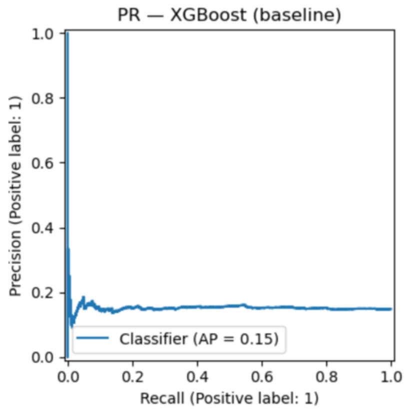
	ROC AUC	PR AUC	F1	Recall	Precision
XGBoost (baseline)	0.510436	0.152613	0.135531	0.125000	0.148000
Dummy (majority)	0.500000	0.148000	0.000000	0.000000	0.000000
Random Forest (balanced)	0.491695	0.139825	0.000000	0.000000	0.000000
Logistic Regression (balanced)	0.439774	0.133487	0.190325	0.405405	0.124352

XGBoost results

XGBoost performed best among all models tested and achieved a PR AUC of 0.153, outperforming the baseline.

On the test set, XGBoost identified 12.5% of actual churners with 14.8% precision, yielding an F1 score of 0.136. Though the improvement was small, it showed that the model could detect patterns that simpler approaches missed entirely. This suggests that more advanced methods could achieve even better results with additional feature development.





5. Key findings

Our analysis revealed five important patterns in Netflix churn behavior:

- More engaged users stay longer.

Users with higher watch times and completion rates are much less likely to churn.

- Active users explore more.
  Users who search for content and click on recommendations tend to keep their subscriptions.

- Satisfaction matters.
  Users who give low ratings or write negative reviews often cancel soon after, providing early warning signals for intervention.

- New users are at highest risk.
  The first 90 days represent a critical window, since the majority of churn occurs during this onboarding period.

- Personal details don't predict churn.
  Demographics and subscription tiers have little impact compared to behavioral variables.

6. Recommendations

Based on these insights, we recommend Netflix implement the three following initiatives:

- Strengthen the onboarding experience.
  Implement personalized welcome content, completion reminders and early exposure to recommendation features during the critical first 90 days to reduce new-user churn.

- Monitor satisfaction proactively.
  Create automated systems to flag users leaving low ratings or negative reviews, then trigger personalized outreach with tailored content suggestions or retention offers.

- Focus on valuable customers.

Calculate Customer Lifetime Value using subscription spend and tenure data, then deploy retention resources strategically toward high-value and at-risk users.

7. Limitations

Several factors limited our analysis. Indeed, some churn labels may reflect temporary inactivity rather than permanent cancellations, which introduces noise into the dataset. The absence of detailed revenue data required us to approximate customer value using basic subscription metrics. Additionally, the class imbalance fundamentally constrained our models' ability to achieve both high precision and recall.

8. Future work

Future improvements should focus on developing richer behavioral features, including engagement recency patterns and decay analysis. Most importantly, any model-driven interventions should be validated through A/B testing to measure their actual impact on retention and revenue.

9. Conclusion

While Netflix churn prediction presents significant challenges due to weak behavioral signals and class imbalance, this project successfully identified engagement, content discovery, user satisfaction and early tenure dynamics as the primary drivers of retention.

Although baseline models struggled with the prediction task, XGBoost's improvement over random classification demonstrates that advanced machine learning methods can detect meaningful patterns in complex user behavior data.

For Netflix leadership, the strategic implications are clear: investing in enhanced onboarding experiences, implementing proactive satisfaction monitoring and targeting high-value users can measurably reduce churn and protect subscription revenue growth.