Netflix customer churn: identifying the behavioral drivers of user retention and departure

Problem statement

Netflix faces significant churn that reduces revenue and limits upsell opportunities. Without knowing which behaviors drive loyalty versus cancellations, retention campaigns risk being ineffective. Building a churn prediction model will enable the company to identify at-risk users and target them more effectively.

Context

The streaming industry is highly competitive, with rising acquisition costs and subscription fatigue. Retaining existing users, especially high-value ones, is more profitable than constant acquisition. Netflix's behavioral dataset (watch history, searches, reviews, recommendations) enables end-to-end churn analysis.

Criteria for success

- Develop a churn prediction model that accurately identifies users at risk of leaving, even though churn cases are less frequent.
- Identify high-risk and high-value segments and key behavioral drivers.
- Provide actionable recommendations for retention campaigns and ROI.

Scope of solution space

- Define churn as "is_active=False".
- Use classification for churn prediction and clustering for segmentation.
- Use Customer Lifetime Value (CLV) estimates to focus retention strategies on the most valuable users.

Constraints

- The dataset may include missing values, duplicates, and/or anomalies.
- Because revenue per show or film isn't available, customer value will be estimated using their monthly subscription fee ("monthly_spend" in the "users" table) and the time they remain subscribed.
- Some churn labels may reflect temporary inactivity rather than true cancellations.

Stakeholders

- **Marketing**: Run focused campaigns to target existing users and re-engage those who have left.
- **Product**: improve recommendations, search, and other features to keep users engaged.
- **Data Science/Analytics**: develop churn prediction models and track KPIs.
- **Finance/Strategy**: estimate CLV and evaluate retention ROI.

Dataset

Netflix 2025:User Behavior Dataset (210K+ Records) (6 linked tables):
https://www.kaggle.com/datasets/sayeeduddin/netflix-2025user-behavior-dataset-210k-records

- *watch_history.csv*: sessions, duration, completion.
- *users.csv*: demographics, plans, spend, activity, subscription length.
- *search_logs.csv:* queries, interactions.
- *reviews.csv*: ratings, text (sentiment).
- *recommendation_logs.csv*: exposures, clicks.
- *movies.csv*: catalog metadata.

Approach

- **Data cleaning**: handle missing values, duplicates and outliers.
- **Feature engineering**: focus on variables related to engagement, content discovery, recommendation usage, user satisfaction and spending behavior.
- **EDA**: compare loyal vs churned cohorts.
- **Modeling**: supervised classification for churn risk.
- **Insights and recommendations** : design targeted interventions and estimate retention ROI.

Deliverables

- Clean dataset with documented transformations.
- EDA and modeling notebooks showing churn definition, key insights and model results.
- Segmentation and churn profiles highlighting high-risk and/or high-value profiles.
- Actionable recommendations with estimated ROI of retention strategies.
- Executive summary and slide deck for business stakeholders.