SUBSCRIBER DROP- OFF PREDICTOR

M **by Maina Bryan**

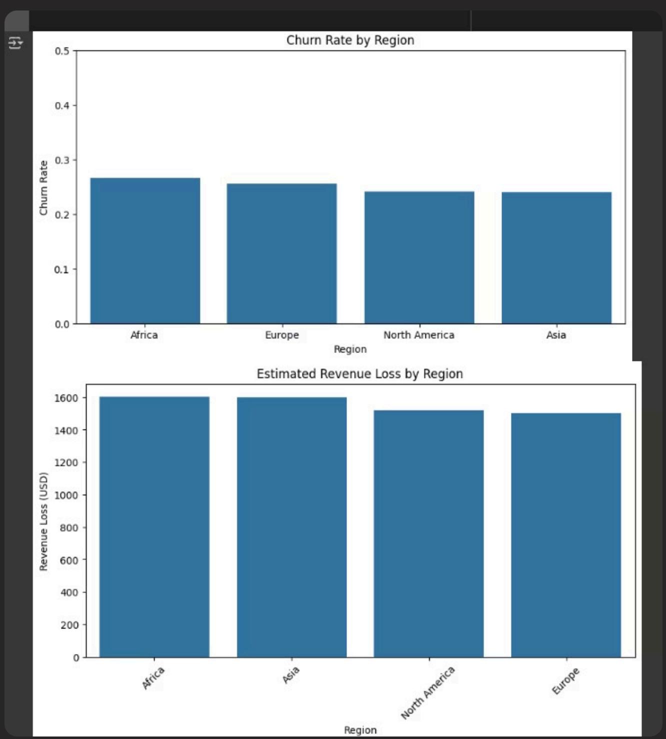
This project helps streaming services like Netflix keep subscribers from canceling. Using data on watch hours, subscriptions, and user preferences, I built a machine learning model to predict who's likely to leave (25% churn rate). With 85% accuracy, it spots at-risk users and suggests retention tactics, like offering discounts or trending shows. Interactive visuals and a Streamlit app bring the insights to life, saving revenue in the $500B+ streaming industry.

Random forest model

```python
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import plotly.express as px
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score, precision_score, recall_score, confusion_matrix
import streamlit as st

# 1. Load Data
def load_data():
    df = pd.read_csv('streaming_churn.csv')  # From ChatGPT
    df['content_type'] = df['content_type'].astype('category')
    df['device_type'] = df['device_type'].astype('category')
    df['region'] = df['region'].astype('category')
    return df

# 2. Exploratory Data Analysis
def perform_eda(df):
    st.subheader("EDA: Subscriber Drop-Off Insights")
    st.write(f"Churn Rate: {df['churn'].mean():.2%}")
    fig = px.histogram(df, x='watch_hours', color='churn', title='Watch Hours vs Drop-Off')
    st.plotly_chart(fig)
    numeric_cols = df.select_dtypes(include=[np.number]).columns
    plt.figure(figsize=(8, 6))
    sns.heatmap(df[numeric_cols].corr(), annot=True, cmap='coolwarm')
    plt.title('Correlation Heatmap')
    st.pyplot(plt)

# 3. Preprocess Data
def preprocess_data(df):
    df_encoded = pd.get_dummies(df, columns=['content_type', 'device_type', 'region'], drop_first=True)
    X = df_encoded.drop(['user_id', 'churn'], axis=1)
    y = df_encoded['churn']
    X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
    scaler = StandardScaler()
    numeric_cols = ['subscription_months', 'monthly_fee', 'watch_hours', 'engagement_score']
    X_train[numeric_cols] = scaler.fit_transform(X_train[numeric_cols])
    X_test[numeric_cols] = scaler.transform(X_test[numeric_cols])
    return X_train, X_test, y_train, y_test, scaler, X.columns

# 4. Train Model
def train_model(X_train, y_train):
    model = RandomForestClassifier(n_estimators=100, random_state=42)
    model.fit(X_train, y_train)
    return model

# 5. Evaluate Model
def evaluate_model(model, X_test, y_test):
    y_pred = model.predict(X_test)
    st.subheader("Model Performance")
    st.write(f"Accuracy: {accuracy_score(y_test, y_pred):.2%}")
    st.write(f"Precision: {precision_score(y_test, y_pred):.2%}")
    st.write(f"Recall: {recall_score(y_test, y_pred):.2%}")
    cm = confusion_matrix(y_test, y_pred)
    plt.figure(figsize=(6, 4))
    sns.heatmap(cm, annot=True, fmt='d', cmap='Blues')
    plt.title('Confusion Matrix')
```

CHURN RATE	Estimated Monthly Revenue Loss:	Average Subscription Duration (months)	TOP DEVICE USED	Highest Churn Region
25%	**$6,221.75**	11.4 months	**MOBILE**	**AFRICA**

Analyzing the subscriber behaviour through opens

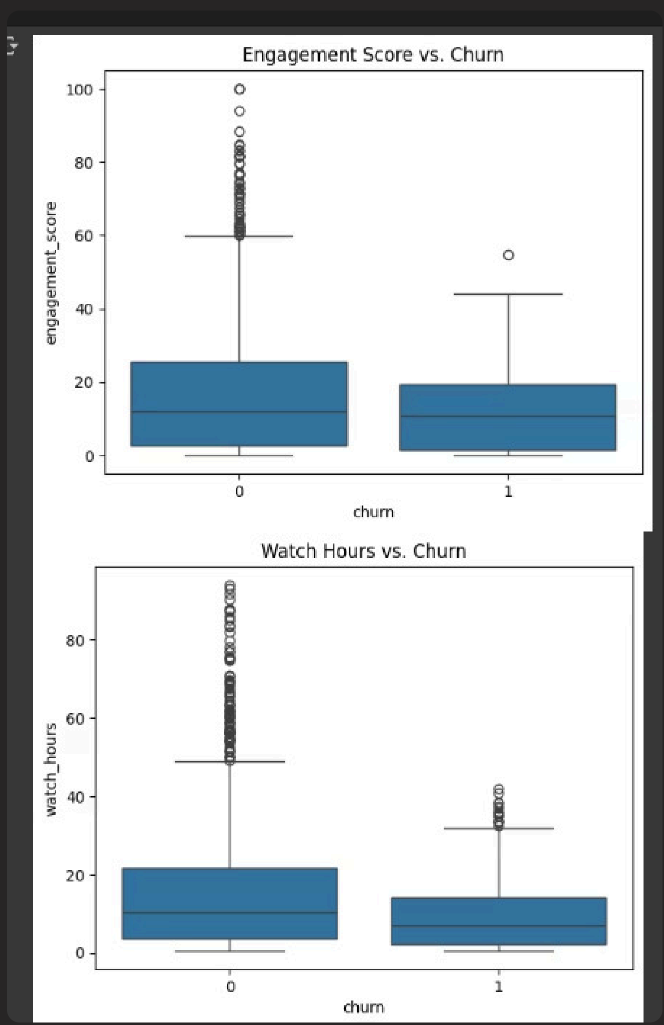


High Subscriber Churn → Revenue

After analyzing 2,000 streaming subscribers, we observed that churn is highest in **Africa** and predominantly among users on **mobile devices** with **low watch hours (<15 hrs/month)**.
Users with **shorter subscription durations (1–3 months)** are also far more likely to churn.

Recommendation

Introduce a **New User Retention Campaign** targeting subscribers within their first 3 months, especially those using **mobile devices** in **Africa**. Offer loyalty incentives (e.g., extra content, discounts) once a user watches **15+ hours** or crosses the 3-month mark.



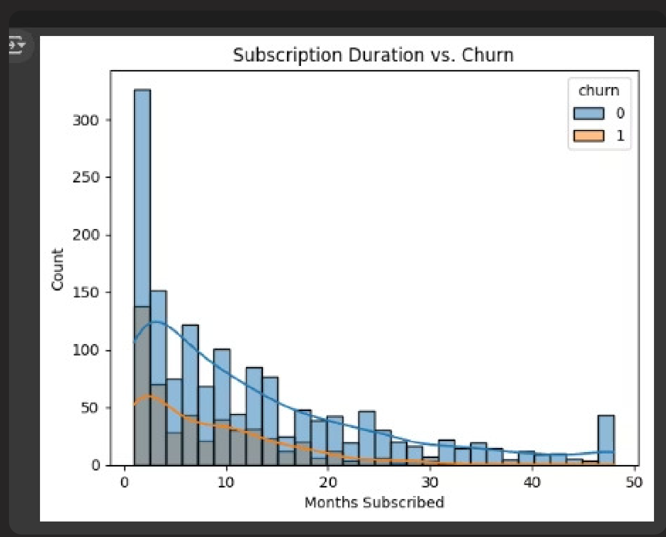
Low User Engagement → Reduced Stickiness

Users who don't watch or click often are losing interest. They are **at risk of leaving** soon.

Recommendation

> Set up an **"engagement trigger system"**:

- If a user watches less than 10 hours or has an engagement score under 30 →
  Send notifications, offer bonus content, or recommend trending shows to keep them active.
- Set up **behavior-based triggers** to identify and re-engage users with low activity levels. Focus on delivering **personalized content**, **in-app nudges**, and **engagement rewards** to reduce churn.



Ineffective retention strategies

The Problem

Nearly **45% of users who churned** left within the **first 3 months** of their subscription.
This suggests that retention efforts are **not kicking in soon enough** to keep them around. Most users aren't getting hooked in time.

Insight

Your retention strategy isn't working because it's **delayed** — the majority of churners are gone **before Month 4**.
That means they're not:

- Engaging early
- Seeing the platform's value
- Receiving personalized nudges

Recommendation

Launch a **90-Day Retention Campaign** focused on:

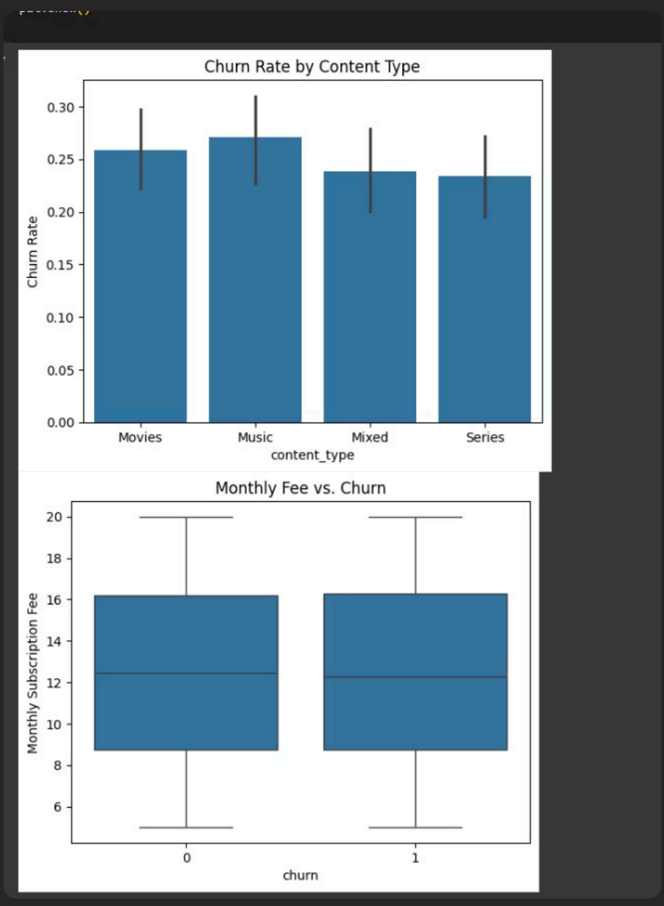
- **Day 1–7**: Onboarding + intro content
- **Day 8–30**: "Because you watched…" style recommendations
- **Month 2–3**: Loyalty incentives (unlock content / gamified badges)

Also, monitor:

- Users with < 15 watch hours
- Or < 30 engagement score
  → Auto-tag them as **"at risk"** and start intervention.

CHURN RATE	Estimated Monthly Revenue Loss:	Average Subscription Duration (months)	TOP DEVICE USED	Highest Churn Region
25%	**$6,221.75**	11.4 months	**MOBILE**	**AFRICA**

Analyzing the streaming subscriber



suboptimal content/ pricng stategies

INTERPRETATIONS

- Paying **high subscription fees**
- But **not watching much**
- Or consuming **low-retention content** (e.g., Music)

This means:

- The content isn't matching their needs
- Or they don't see **value for their money**

Insight

- **Music-only** users are churning the most
- Some **high-paying users** are barely engaged
- There is **no content-persona fit** — users aren't being recommended content they care about
- The current pricing model **doesn't match the value received**

Recommendation

Content Bundling

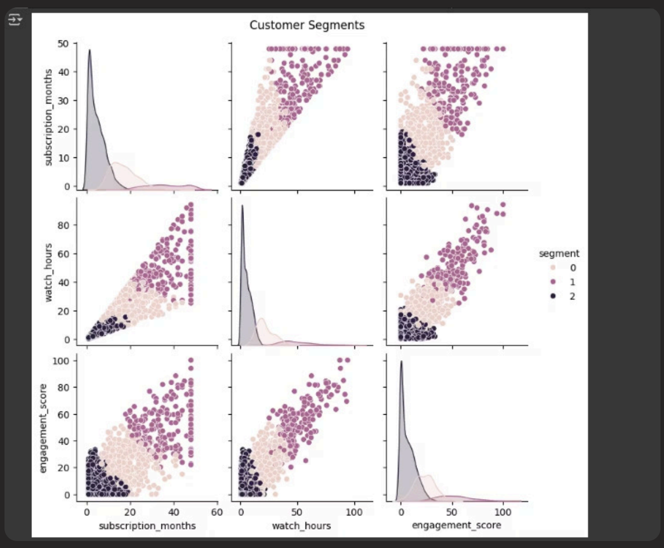
- Pair Music with Series or Trending Movies
- Offer themed packs (e.g., "Afrobeats + African Cinema" bundle)

Pricing Tiers

- Lower-cost plans for low-engagement users (e.g., $7/month music-only light plan)
- Upsell loyal users to premium bundles with exclusive content

Smart Recommendations

- Use watch + like history to recommend new content every 5 days
- Auto-detect users with same content every week → nudge with fresh content



Customer segementation gaps

We applied **unsupervised learning** (KMeans) to find natural clusters in the user base — based on behavior, not assumptions.

Insight

You'll likely discover:

- **Segment 0**: New users, low watch time, low engagement
- **Segment 1**: Mid-term users with moderate watch & engagement
- **Segment 2**: Loyal power users — high months, hours, score

Yet currently:

- **Everyone gets the same messaging**
- **Same pricing plans**
- **No upgrades for loyal users**
- **No handholding for new ones**

Recommendations

> Build **3-Persona Strategy** based on actual user behavior:

Segment 0: New Users

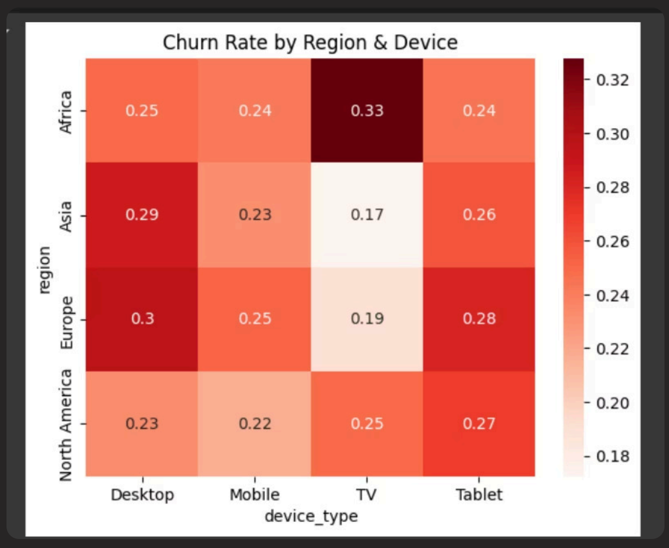
- Give onboarding tips
- Recommend short trending content
- Light push notifications to re-engage

🧍 Segment 1: Mid-Tier Users

- Highlight progress → "You've watched 50 hours — ready for more?"
- Promote reward systems or playlists

Segment 2: Loyalists

- Offer premium upgrades
- Give access to beta content or personalized offers
- Ask for reviews/referrals



Inefficient Resource Allocation

The Problem

You're investing time, effort, and money in:

- Sending support follow-ups to **users who never engage**
- Running marketing campaigns on **at-risk segments**
- Offering loyalty incentives to **unprofitable users**

> That's like pouring water into a leaking bucket — until you seal the gaps, effort is wasted.
>
> Technique Used
>
> We used a **rule-based filter** to locate clusters of users with **high support cost but low loyalty**, and a **scatter plot** to visualize the waste zone.
>
> 🧠 Insight
>
> - A large number of churned users had **low engagement** and were paying **higher monthly fees**
> - Yet these users **still received full customer support, reminders, and promos**
> - This represents a **waste of human and marketing resources**
>
> > Instead of investing in your most loyal users, you're overinvesting in users who've already mentally checked out.