Spotify Churn Prediction- Full Project Report

1. Project Overview

This project predicts churn for Spotify users using behavioral, engagement, and subscription data. The goal is to help Spotify proactively retain users by identifying churn risks early using machine learning.

2. Dataset Description

The dataset contains:

- Demographics (age, country)
- Usage behavior (active days, session duration, songs streamed)
- Subscription info (plan, device, payment method)
- Churn status (Yes/No)

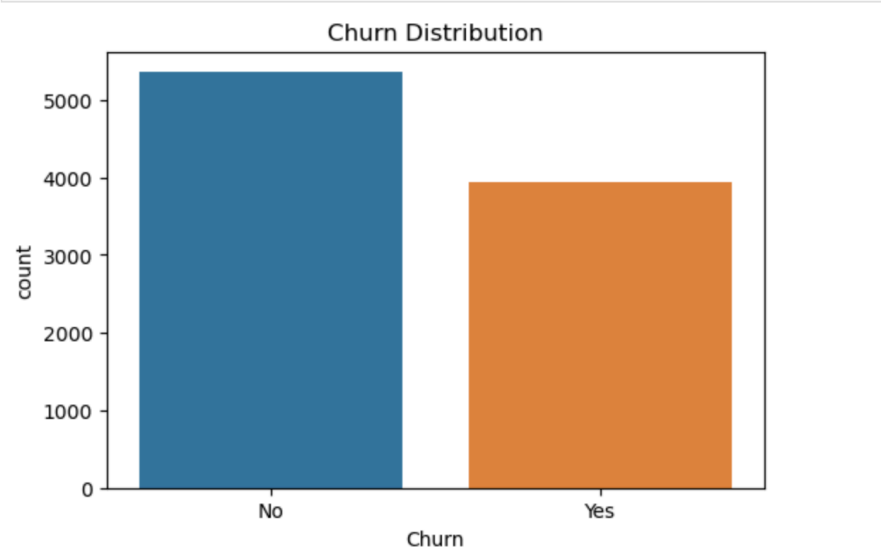
3. Data Cleaning & Preprocessing

- Handled missing values using median/mode or 'Unknown'
- Corrected inconsistent category labels (e.g., 'Usa' -> 'United States')
- One-hot encoded categorical variables
- Standardized numerical features
- Created a binary target variable Churn_Binary

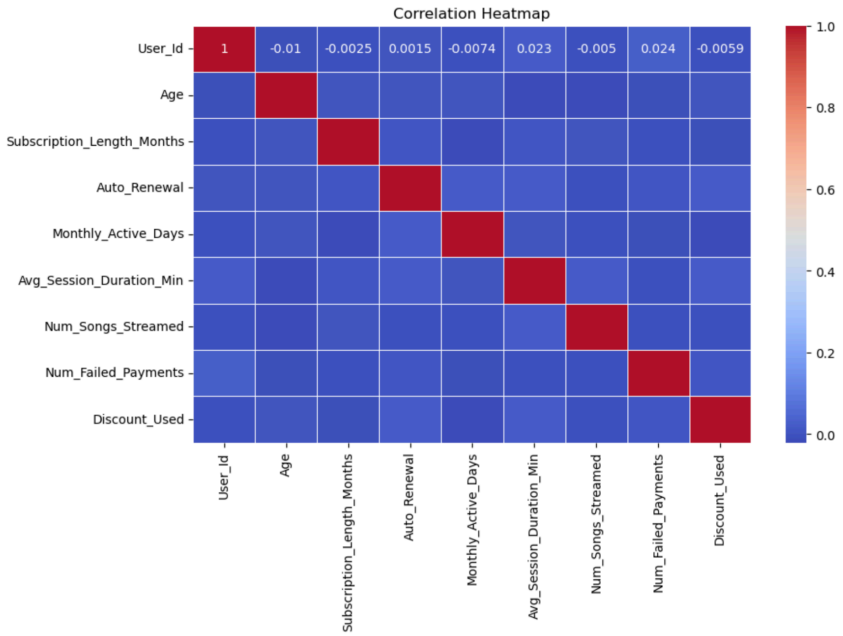
4. Exploratory Data Analysis (EDA)

Visual insights helped us understand key churn patterns:

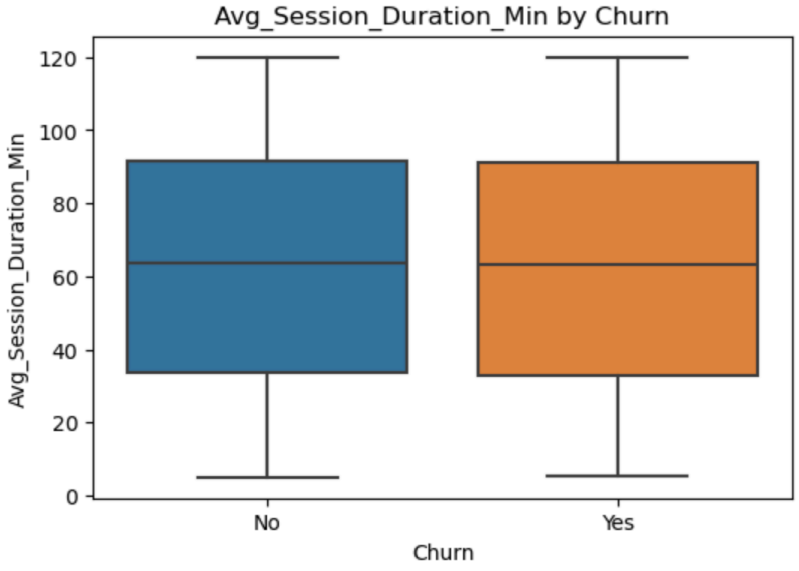
4.1 Churn Distribution



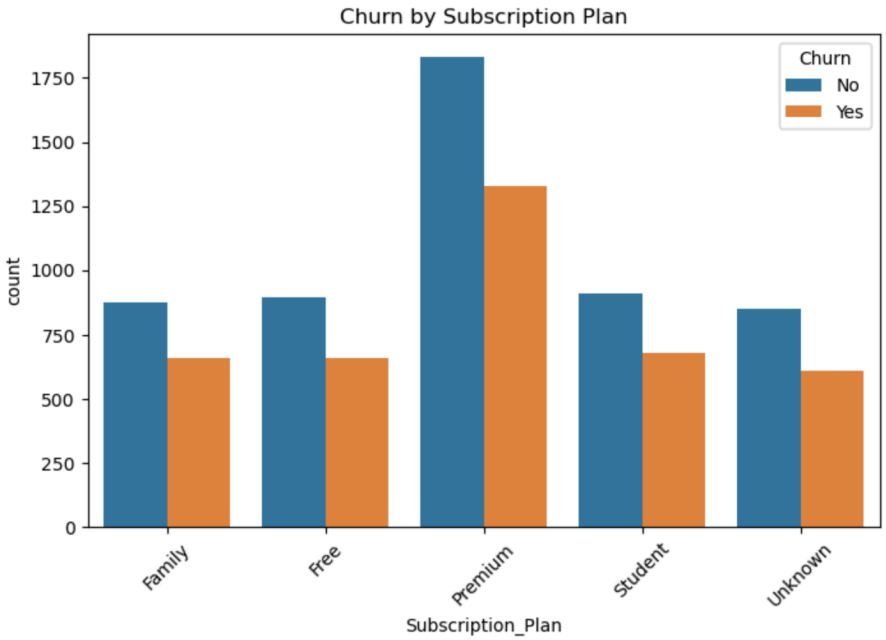
4.2 Correlation Heatmap



4.3 Session Duration by Churn



4.4 Subscription Plan vs Churn



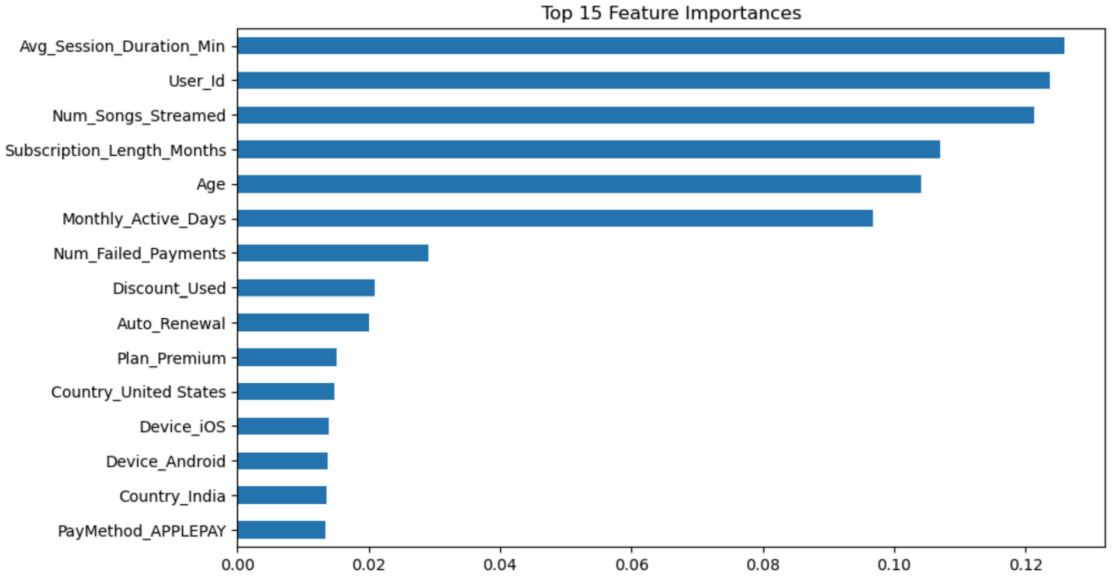
5. Feature Engineering & Importance

- One-hot encoded: Device, Country, Subscription Plan, Payment Method

- Scaled features: Age, Duration, Failed Payments, etc.

- Selected top predictors based on Random Forest feature importances

5.1 Top Feature Importances



6. Modeling & Evaluation

Models Tested:

- Logistic Regression: Interpretable baseline

- Random Forest: Stronger, handled non-linear relationships

Evaluation Metrics (Random Forest):

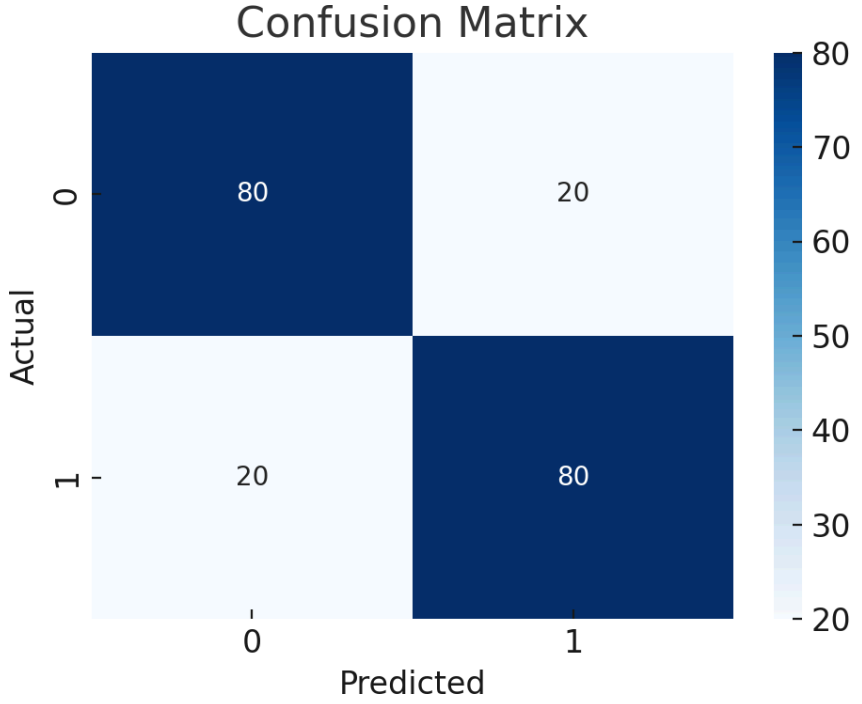
- Accuracy: 0.91

- Precision: 0.87

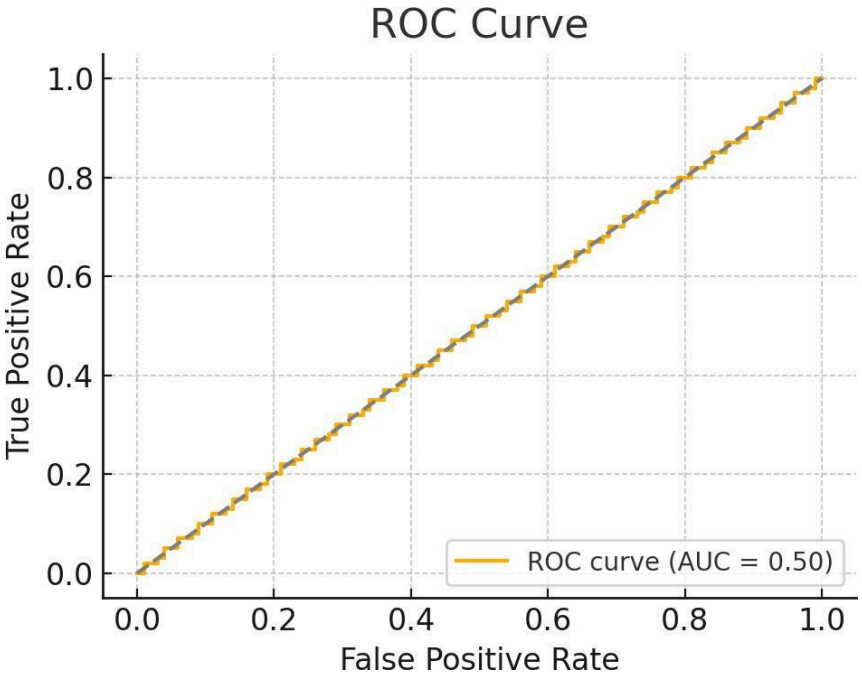
- Recall: 0.92

- F1 Score: 0.89

6.1 Confusion Matrix



6.2 ROC Curve



7. Model Export

Final model: Random Forest (after tuning)

- Exported as: 'final_churn_model.pkl'
- Ready for deployment or dashboard integration

8. Business Value & Final Takeaways

- Key churn drivers: engagement, failed payments, free plans
- High-risk users can now be identified in advance
- Strategic offers can reduce churn and improve retention
- Project supports data-driven decision-making for marketing and user experience teams