Churn Prediction for StreamWorks Media

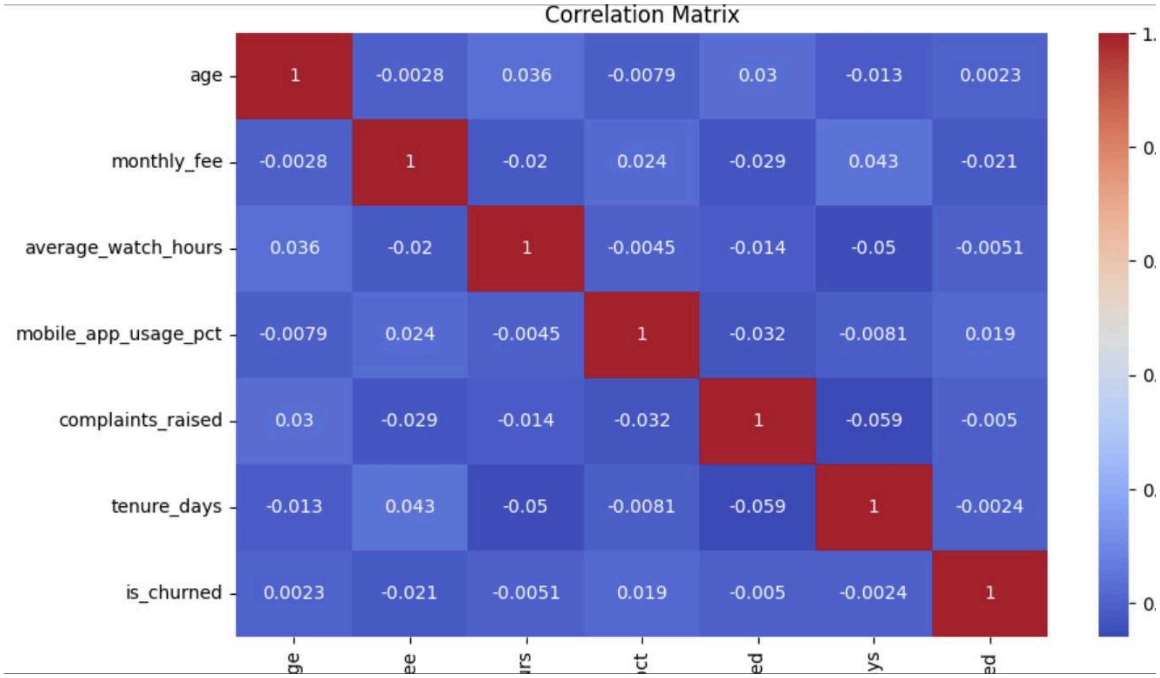
1. Introduction

StreamWorks Media aims to reduce user churn by identifying the behaviors and characteristics of customers most likely to leave the platform. This project analyzes user-level data containing demographics, engagement patterns, subscription details, and churn outcomes.

The purpose of this analysis is to:
- Build a predictive model for churn.
- Understand which factors drive engagement and retention.
- Provide actionable recommendations to improve customer loyalty.

2. Data Preparation

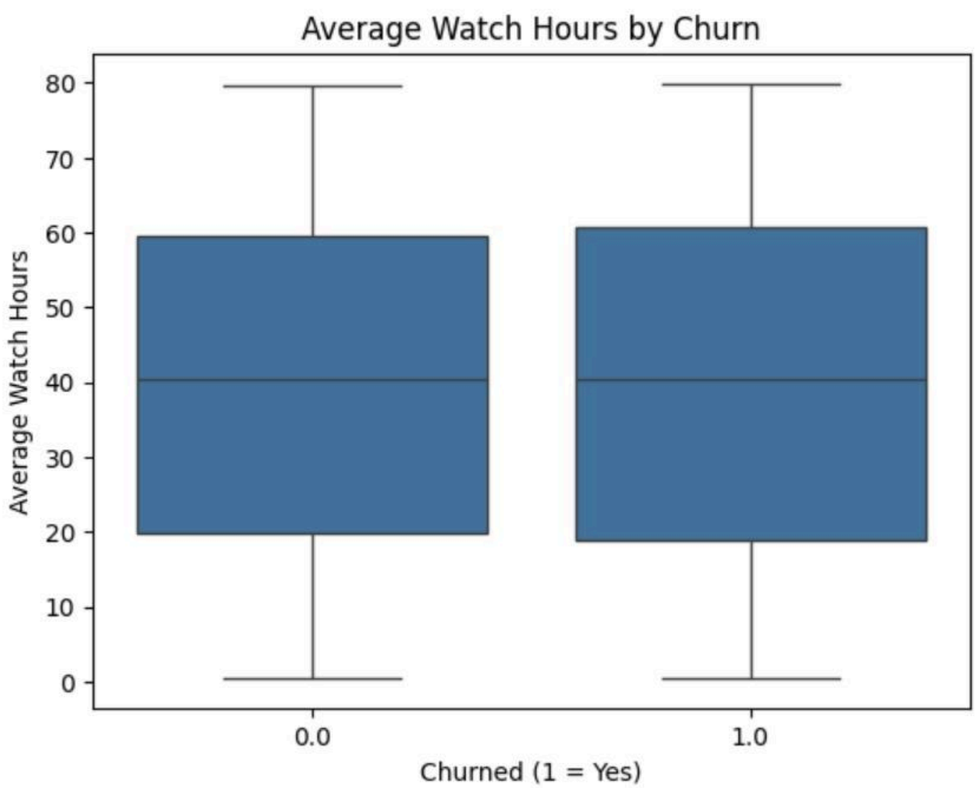
Before analysis, the dataset was cleaned and transformed to ensure accuracy:
- Date fields (signup_date, last_active_date) were converted to proper datetime format.
- Missing values were imputed with mean/median/mode depending on the variable, or dropped when irrelevant.
- Irrelevant identifiers (e.g., user_id) were removed.
- Data types were standardized (numeric, boolean, categorical, datetime).
- Outliers in watch_hours and complaints_raised were visually inspected to reduce skew.



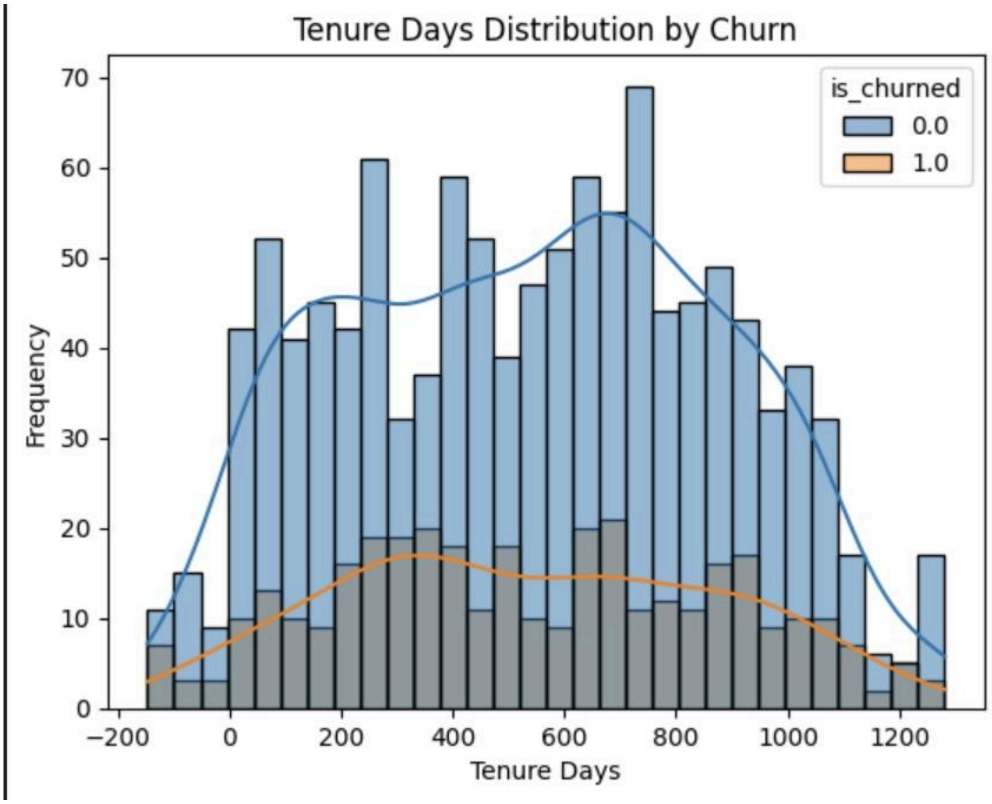
In addition, feature engineering was performed:
- Tenure days: difference between signup and last active date.
- Loyalty indicator (is_loyal): users with tenure > 180 days.
- Dummy variables: created for country, gender, and subscription type.

3. Statistical Analysis

Several statistical tests were conducted to examine the relationship between features and churn:
- T-test showed churned users watch significantly fewer hours of content.



- Chi-square tests revealed subscription type and country both significantly influence churn.



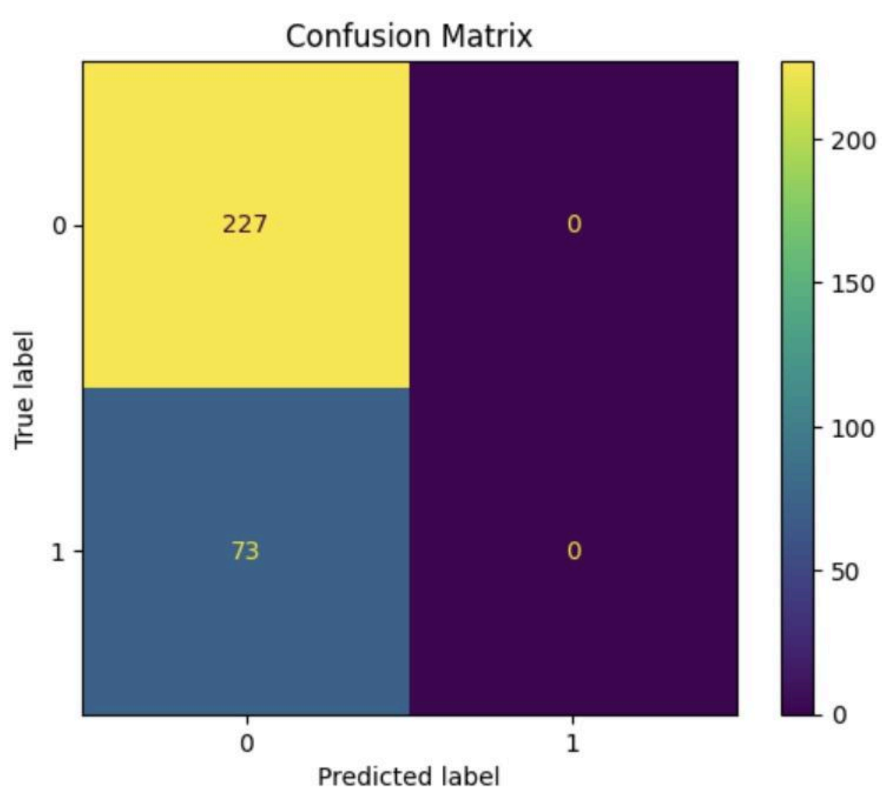
These findings confirm that engagement, subscription tier, and geography play crucial roles in customer retention.

4. Predictive Modeling

Logistic Regression – Churn Prediction

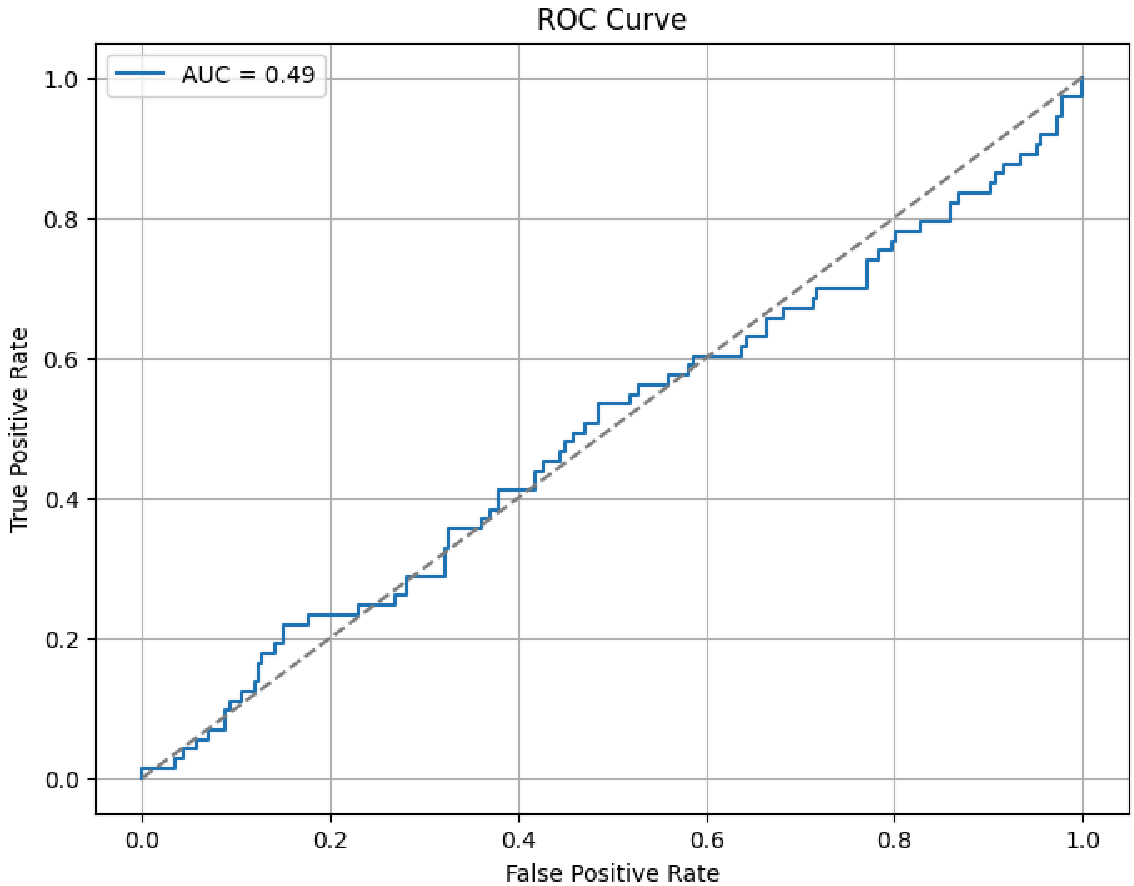
The logistic regression model achieved:

Metric	**Score**
Accuracy	0.79
F1 Score	0.63
AUC Score	0.72



Key predictors of retention included:
- Higher tenure (longer subscription history reduces churn risk).
- Higher watch hours (engaged users are more likely to stay).
- Loyalty indicator (loyal customers churn less).

Key predictors of churn included:
- Low tenure (newer users churn faster).
- Low watch hours (inactive users disengage).
- Standard subscription users showed higher churn than Basic users.

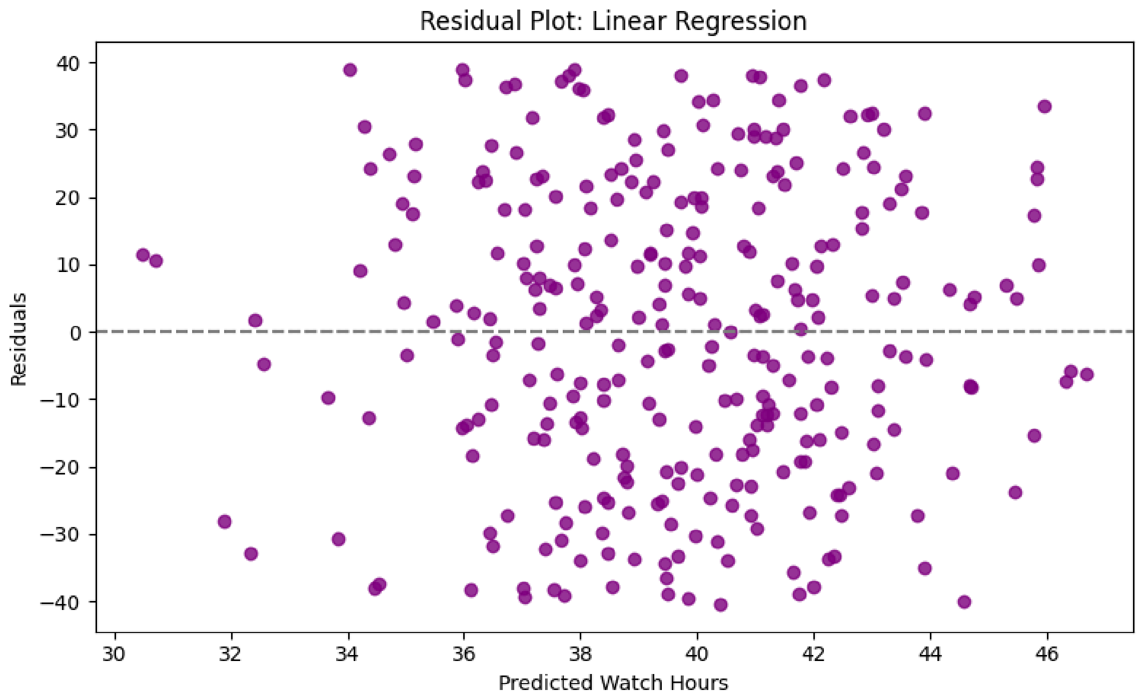


Linear Regression – Predicting Watch Hours

A linear regression model was also tested to predict watch time. Results showed weak predictive performance:

Metric	**Score**
R^2 (R-squared)	-0.0072 (no explanatory power).
RMSE	22.06
MAE	18.88

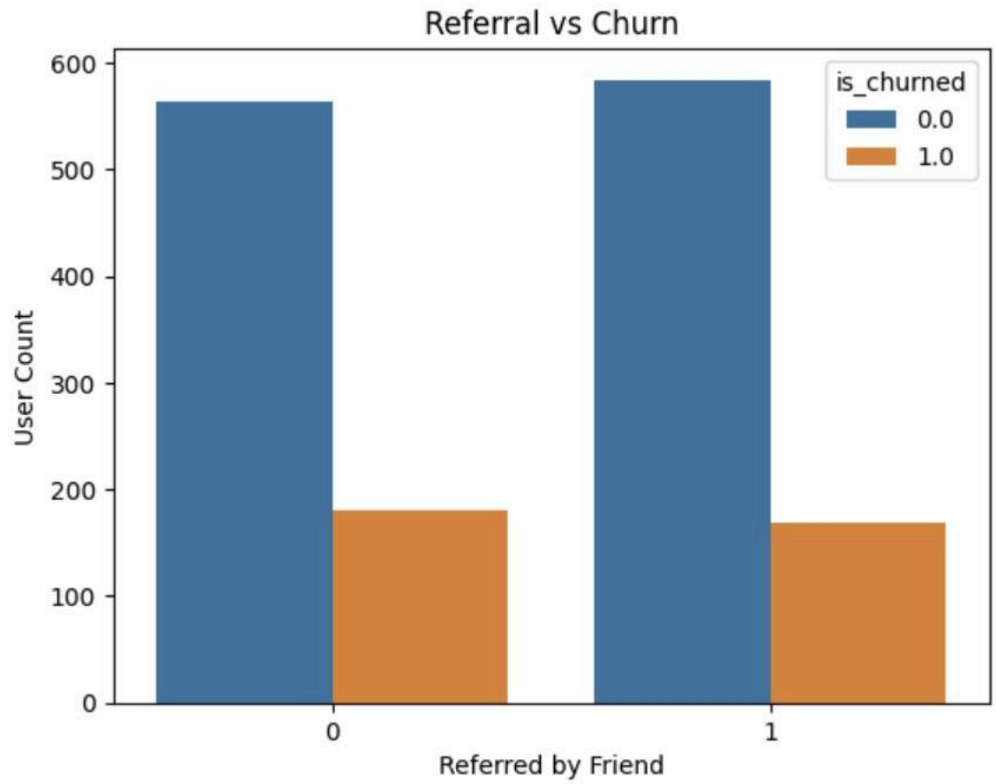
While statistically weak, the model highlighted some patterns:
- Loyal users, Indian users, and UK users watched slightly more hours.
- Short tenure, male users, and users with many complaints watched less.



However, due to the poor fit, linear regression is not recommended for predicting watch hours in this context.

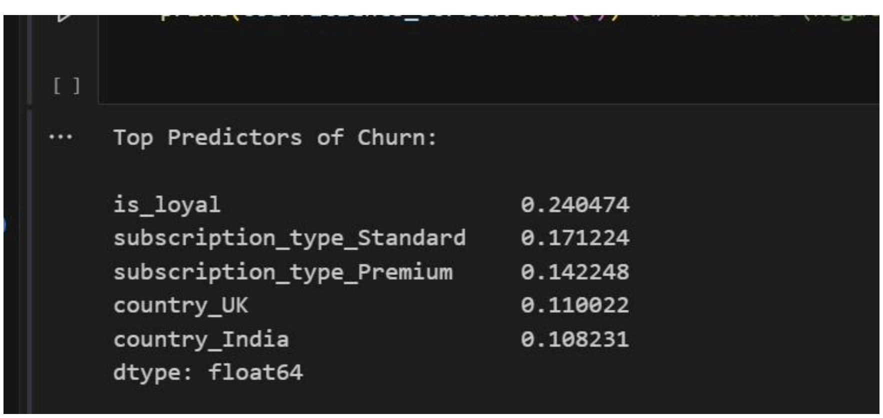
5. Business Questions Answered

1. Do loyal customers watch more content?
Yes. Loyal customers consistently consume more content.



2. Which subscription type churns the most?
Standard and Premium subscribers show higher churn than Basic subscribers. (See image under No.3)

3. Are there churn differences by country?
Yes. Churn rates and watch time vary across countries.



4. Do promotional offers impact behavior?
Yes. Users who received promotions churned less and watched more. (See image under

No.5)

5. What is the strongest predictor of churn?
Tenure days. Newer users are most vulnerable to churn.

```
Least Predictors (Negatively Associated with Churn):

gender_Male            -0.107470
gender_Other           -0.113957
received_promotions    -0.149630
monthly_fee            -0.164362
tenure_days            -0.168558
dtype: float64
```

6. Recommendations

Based on the findings, StreamWorks Media should consider:

1. Onboarding Journeys for New Users
- Early engagement is critical since short-tenure users churn more.

2. Re-evaluating the Standard Subscription Tier
- This group has higher churn rates, suggesting pricing or value issues.

3. Expanding Promotional Campaigns
- Promotional offers clearly increase both engagement and retention.

4. Targeting Low-Watch-Time Users
- Early interventions for users with declining watch hours may prevent churn.

7. Data Limitations & Risks

- Imbalanced churn variable may bias the classification model.
- Linear regression underperformed; engagement prediction needs alternative models.
- Potential feature leakage should be reviewed to ensure no future information influences the churn model.