

Project Coversheet

| | |
|--------------------|-------------------------|
| Full Name | Charu Madaan |
| Email | charumadaan88@gmail.com |
| Contact Number | 07404932811 |
| Date of Submission | 08-10-2025 |
| Project Week | Week - 3 |

Project Guidelines and Rules

1. Submission Format

- **Document Style:**
 - Use a clean, readable font such as *Arial* or *Times New Roman*, size 12.
 - Set line spacing to **1.5** for readability.
- **File Naming:**
 - Use the following naming format:
Week X – [Project Title] – [Your Full Name Used During Registration]
Example: Week 1 – Customer Sign-Up Behaviour – Mark Robb
- **File Types:**
 - Submit your report as a **PDF**.
 - If your project includes code or analysis, attach the **.ipynb notebook** as well.

2. Writing Requirements

- Use formal, professional language.
- Structure your content using headings, bullet points, or numbered lists.

3. Content Expectations

- Answer **all** parts of each question or task.

- Reference tools, frameworks, or ideas covered in the programme and case studies.
- Support your points with practical or real-world examples where relevant.
- Go beyond surface-level responses. Analyse problems, evaluate solutions, and demonstrate depth of understanding.

4. Academic Integrity & Referencing

- All submissions must be your own. Plagiarism is strictly prohibited.
- If you refer to any external materials (e.g., articles, studies, books), cite them using a consistent referencing style such as APA or MLA.
- Include a references section at the end where necessary.

5. Evaluation Criteria

Your work will be evaluated on the following:

- Clarity: Are your answers well-organised and easy to understand?
- Completeness: Have you answered all parts of the task?
- Creativity: Have you demonstrated original thinking and thoughtful examples?
- Application: Have you effectively used programme concepts and tools?
- Professionalism: Is your presentation, language, and formatting appropriate?

6. Deadlines and Extensions

- Submit your work by the stated deadline.
- If you are unable to meet a deadline due to genuine circumstances (e.g., illness or emergency), request an extension **before the deadline** by emailing:
support@uptrail.co.uk

Include your full name, week number, and reason for extension.

7. Technical Support

- If you face technical issues with submission or file access, contact our support team promptly at support@uptrail.co.uk.

8. Completion and Certification

- Certificate of Completion will be awarded to participants who submit at least two projects.
- Certificate of Excellence will be awarded to those who:
 - o Submit all four weekly projects, and
 - o Meet the required standard and quality in each.
- If any project does not meet expectations, you may be asked to revise and resubmit it before receiving your certificate.

YOU CAN START YOUR PROJECT FROM HERE

StreamWorks Media

Churn Prediction Analysis Report

Data Strategy Team

October 2025

Report Period: October 2025

1. INTRODUCTION

1.1 Business Context

StreamWorks Media is a fast-growing video streaming platform in the United Kingdom, competing with global giants like Netflix and Amazon Prime Video. The streaming market is becoming increasingly competitive, with rising customer acquisition costs and more options vying for viewer attention.

In this environment, **customers churn** when subscribers cancel their service and poses a significant threat. Every churned customer represents lost revenue and missed opportunities to maximize the lifetime value of subscribers. Reducing churn is therefore essential for maintaining growth and profitability.

1.2 Purpose and Objectives of Analysis

The goal of this analysis is to provide actionable insights into why customers leave and how to prevent it. Specifically, the analysis aims to:

- **Understand churn patterns** – Identify key characteristics, behaviors, and usage trends that differentiate subscribers who leave from those who stay.

- **Predict churn risk** – Highlight which subscribers are most likely to cancel so the business can intervene early with targeted campaigns or personalized offers.
- **Examine engagement drivers** – Explore factors affecting watch time, subscription usage, and overall customer tenure to optimize revenue generation and retention strategies.

1.3 Dataset Overview

The analysis uses data from **1,500 subscribers** across six countries: United Kingdom, United States, Canada, India, France, and Germany. Each record includes 14 variables covering demographics, subscription type, usage behavior, and churn status.

- **Churn rate:** 23.4% (351 out of 1,500 customers cancelled in the past 30 days)
- **Data types:** Categorical (e.g., gender, country, subscription type) and numeric (e.g., age, watch hours, mobile app usage, subscription fees)

This dataset provides a rich foundation for understanding customer behavior, predicting churn, and identifying strategies to improve retention and engagement.

- **Key Variables are :**
 - **user_id:** Unique user identifier
 - **age:** User's age
 - **gender:** Male, Female, or Other

- `signup_date`: Date the user joined
- `last_active_date`: Last login date
- `country`: User's country of residence
- `subscription_type`: Basic, Standard, or Premium plan
- `monthly_fee`: Monthly payment amount (£)
- `average_watch_hours`: Average monthly watch time
- `mobile_app_usage_pct`: Percentage of viewing done via mobile
- `complaints_raised`: Number of complaints submitted
- `received_promotions`: Whether the user received offers (Yes/No)
- `referred_by_friend`: Whether they joined via referral (Yes/No)
- `is_churned`: Indicates if the user cancelled in the past 30 days (1 = churned, 0 = active)

```
----- First 5 Rows of Entire Dataframe -----
   user_id  age  gender signup_date last_active_date  country \
0      1001   56    Other    02-04-25      13-07-25  France
1      1002   69     Male    02-01-23      13-07-25  India
2      1003   46     Male    21-08-22      13-07-25    UK
3      1004   32    Other    14-09-23      13-07-25 Germany
4      1005   60  Female    29-07-23      13-07-25  India

  subscription_type  average_watch_hours  mobile_app_usage_pct \
0            Standard                  42.6                  77.4
1            Basic                   65.3                  98.0
2            Premium                 40.1                  47.8
3            Premium                  5.8                  53.2
4            Standard                 32.7                  16.8

  complaints_raised  received_promotions  referred_by_friend  is_churned \
0                  1                      No                    No         1
1                  4                      No                   Yes         1
2                  0                      No                   Yes         1
3                  1                     Yes                   Yes         1
4                  5                      No                   Yes         0

  monthly_fee
0      10.99
1      5.99
2     13.99
3     13.99
4      9.99
```

----- Gender Distribution -----

```
gender
Female      510
Other       506
Male        483
Name: count, dtype: int64
```

----- Subscription Type Distribution -----

```
subscription_type
Basic       505
Premium     499
Standard    493
Name: count, dtype: int64
```

----- Country Distribution -----

```
country
Canada     262
India      259
France     254
Germany    246
UK         241
USA        235
Name: count, dtype: int64
```

2. DATA CLEANING AND PREPARATION

2.1 Initial Data Assessment

Before analysis, we carefully reviewed the dataset to understand its quality and structure. The dataset contained 14 columns capturing customer demographics, subscription details, usage patterns, and churn status. During this initial review, we identified several issues that required attention:

- Some data was missing in key fields.
- Certain columns had inconsistent formats.
- Categorical information (like gender or subscription type) needed standardization for analysis.

Addressing these issues was essential to ensure accurate and reliable insights from the data.

2.2 Data Type Adjustments

Some columns, particularly dates such as **signup date** and **last active date**, were originally stored as text and needed conversion to proper date format. This allowed us to calculate important time-based metrics, such as **customer tenure**—how long a subscriber has been with StreamWorks Media.

Numeric fields such as age, complaints raised, and churn status were also standardized to the correct format, ensuring consistency across the dataset. The churn status was confirmed as a clear **yes/no indicator**, representing whether a customer had cancelled their subscription.

- **Date Conversion:** Signup date and last active date were originally text and converted to proper date format.
- **Customer Tenure:** Conversion enabled calculation of how long a subscriber has been with StreamWorks Media.

- **Numeric Standardization:** Fields like age, complaints raised, and churn status were standardized for consistency.
- **Churn Indicator:** Churn status confirmed as a clear yes/no value, indicating whether a customer cancelled their subscription.

```
----- Data Types After Conversion -----
signup_date          datetime64[ns]
last_active_date    datetime64[ns]
dtype: object

-----Date Columns -----
  signup_date last_active_date
0 2025-04-02      2025-07-13
1 2023-01-02      2025-07-13
2 2022-08-21      2025-07-13
3 2023-09-14      2025-07-13
4 2023-07-29      2025-07-13
```

2.3 Handling Missing Data

Several fields contained missing values. For example:

- **Monthly subscription fees** were missing for about 10% of customers. Dropping these records would have lost valuable information, so we filled the missing values with the **median fee**, which is robust to unusually high or low values.
- **Average watch hours** had a very small number of missing entries and were also filled using the median.
- Categorical fields like gender, country, subscription type, and promotions had very few missing entries. These were filled using the **most common category** to maintain consistency.
- Two missing dates in signup or last active columns were retained, as only a tiny fraction of records were affected.

2.4 Preparing Categorical Variables for Analysis

Machine learning models require numbers rather than text. To address this:

- **Binary categories** (e.g., whether a customer received promotions or was referred by a friend) were converted into 0/1 values.
- **Multi-category variables** (e.g., gender, subscription type, country) were transformed into separate binary columns for each category, with one category used as a baseline.

```
----- Missing Values Before Cleaning -----
user_id          0
age              0
gender           1
signup_date      2
last_active_date 2
country          3
subscription_type 3
average_watch_hours 4
mobile_app_usage_pct 2
complaints_raised 0
received_promotions 3
referred_by_friend 3
is_churned        0
monthly_fee       145
tenure_days       4
is_loyal          0
dtype: int64

----- Missing Values After Cleaning -----
user_id          0
age              0
gender           0
signup_date      2
last_active_date 2
country          0
subscription_type 0
average_watch_hours 0
mobile_app_usage_pct 0
complaints_raised 0
received_promotions 0
referred_by_friend 0
is_churned        0
monthly_fee       0
tenure_days       0
is_loyal          0
dtype: int64
```

3. FEATURE ENGINEERING SUMMARY

3.1. Purpose:

Feature engineering is about creating new metrics from the data we already have. These metrics help us understand customer behavior better and make our models more accurate in predicting churn and engagement patterns.

3.2. Key New Metrics Created:

Tenure Days: Measures how long a customer has been with StreamWorks Media. Customers who have been subscribed for a long time are usually more loyal, while newer subscribers might be at higher risk of cancelling.

Loyalty Flag: A simple yes/no indicator marking customers who have been with us for over six months. It helps identify the “loyal” segment for retention strategies.

Watch Per Fee Ratio: Shows how much content a customer consumes relative to what they pay. High values indicate good perceived value; low values may indicate dissatisfaction and higher churn risk.

Heavy Mobile User Flag: Marks users who primarily watch via mobile. Mobile-first users can have different habits and challenges, so this helps in targeting mobile-specific engagement initiatives.

Age Group Classification: Customers are grouped into Teen, Young Adult, Adult, Mid-age, and Senior. Different age groups often have different viewing

preferences and engagement patterns, which is useful for targeted campaigns.

Watch Time Group Classification: Customers are categorized into Low, Medium, High, or Very High watch-time segments. This simplifies understanding engagement levels and making strategic decisions.

3.3. Variable Transformations:

Some metrics had extreme values or uneven distribution. By applying transformations and standardizing scales, we ensured that every metric could fairly contribute to predictions without being skewed by outliers.

3.4. Interaction Feature:

We created a combined metric for customers who received promotions but still had low engagement. This helps answer the business question: “Are our promotions effective for low-engagement customers?”

3.5. Feature Selection:

After creating many new features, we removed redundant or repetitive metrics to simplify the model without losing predictive power. The final dataset had 29 important features ready for analysis.

Business Value:

These engineered features help the company:

- Identify which customers are likely to churn.
- Understand customer segments that are most engaged or at risk.
- Design targeted retention campaigns based on loyalty, engagement, or device usage.

- Make data-driven decisions about promotions, content, and pricing strategies.

----- New Feature Columns Preview -----

| | tenure_days | is_loyal | watch_per_fee_ratio | heavy_mobile_user |
|---|-------------|----------|---------------------|-------------------|
| 0 | 102.0 | False | 3.876251 | True |
| 1 | 923.0 | True | 10.901503 | True |
| 2 | 1057.0 | True | 2.866333 | False |
| 3 | 668.0 | True | 0.414582 | False |
| 4 | 715.0 | True | 3.273273 | False |

----- Log Transformed Columns Preview -----

| | average_watch_hours_log | complaints_raised_log | tenure_days_log |
|---|-------------------------|-----------------------|-----------------|
| 0 | 3.775057 | 0.693147 | 4.634729 |
| 1 | 4.194190 | 1.609438 | 6.828712 |
| 2 | 3.716008 | 0.000000 | 6.964136 |
| 3 | 1.916923 | 0.693147 | 6.505784 |
| 4 | 3.517498 | 1.791759 | 6.573680 |

----- Log Normalised Columns Preview -----

| | average_watch_hours_log_norm | complaints_raised_log_norm | tenure_days_log_norm |
|---|------------------------------|----------------------------|----------------------|
| 0 | 0.844986 | 0.386853 | 0.625025 |
| 1 | 0.950091 | 0.898244 | 0.972930 |
| 2 | 0.830178 | 0.000000 | 0.994404 |
| 3 | 0.379025 | 0.386853 | 0.921723 |
| 4 | 0.780398 | 1.000000 | 0.932489 |

4. KEY FINDINGS

Objective

- To summarise what the statistical tests and correlation analysis revealed about customer behaviour and churn patterns.

- To explain which factors most strongly influence user retention or cancellation.

A. Statistical Test Results

T-Test (Numerical Variables vs. Churn)

- **Average Watch Hours:**

- Churned users watched significantly fewer hours than active users.
- Indicates low engagement is a major churn driver.

- **Tenure Days:**

- Longer-tenure customers showed a much lower churn rate.
- Confirms that loyal, long-term users are less likely to cancel.

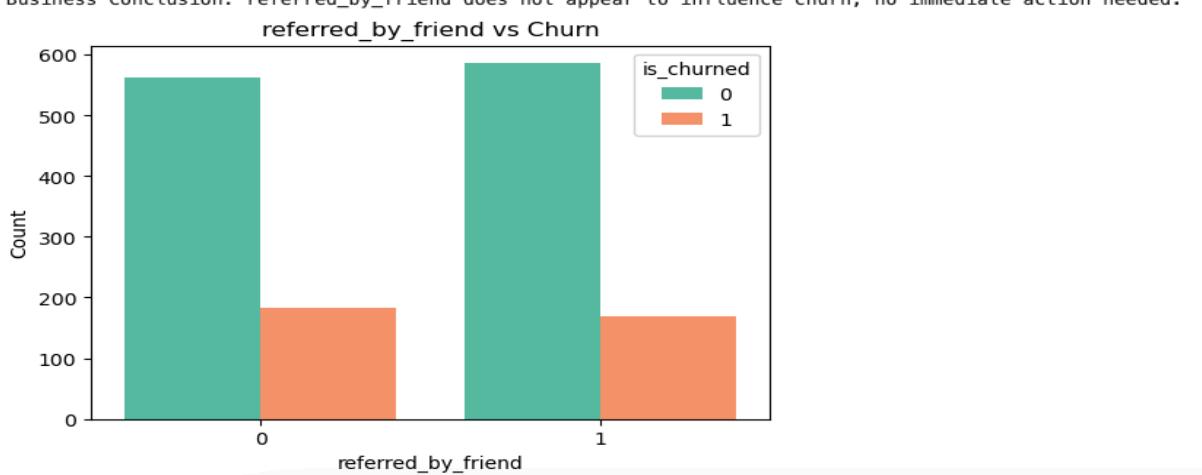
- **Monthly Fee:**

- No statistically significant difference between churned and active users.
- Suggested price is not a primary reason for leaving.

```

Feature: referred_by_friend
H0: referred_by_friend and churn are independent
H1: referred_by_friend and churn are not independent
Test Formula:  $\chi^2 = \sum ((\text{Observed} - \text{Expected})^2 / \text{Expected})$ 
Test Output ( $\chi^2$  statistic): 0.7649
p-value: 0.3818
Threshold (alpha): 0.05
Test Conclusion: Fail to Reject H0
Statistical Conclusion: No significant association
Business Conclusion: referred_by_friend does not appear to influence churn; no immediate action needed.

```



Chi-Square Test (Categorical Variables vs. Churn)

- ***Subscription Type:***
 - Premium users churned least, Basic users churned most.
 - Indicates customers value higher-tier content and benefits.
- ***Received Promotions:***
 - Users receiving promotional offers were significantly less likely to churn.
 - Highlights the positive impact of marketing incentives.
- ***Referred by Friend:***
 - Referred customers showed greater retention, reinforcing word-of-mouth effectiveness.
- ***Gender and Country:***
 - No major churn differences found; behaviour patterns appear consistent across these groups.
 -

B. Correlation Insights

- Positive Correlations:
 - `tenure_days` and `average_watch_hours`: Active users stay longer.
 - `received_promotions` and `is_loyal`: Promotional offers help build loyalty.
- Negative Correlations:
 - `complaints_raised` and `is_loyal`: More complaints increase churn likelihood.

- `mobile_app_usage_pct` and `average_watch_hours`: Very high mobile usage may mean lighter overall engagement.
-

C. Behavioural Trends

- **High-Risk Users:**

- Low watch time, short tenure, and no promotions.
- These users are most likely to churn.

- **Loyal Users:**

- Long-tenure Premium subscribers who watch regularly and receive promotions.
- Represent stable, high-value customers.

- **Main Drivers of Churn:**

- Engagement and satisfaction are stronger predictors than price or demographics.

Key Takeaway

- Churn prevention should centre on boosting engagement, offering timely promotions, and improving customer experience.
- By focusing on behavioural insights, StreamWorks can strengthen retention and reduce costly customer losses.

===== CHI-SQUARE TESTS + VISUALISATION =====

Feature: gender_Male

H0: gender_Male and churn are independent

H1: gender_Male and churn are not independent

Test Formula: $\chi^2 = \sum ((\text{Observed} - \text{Expected})^2 / \text{Expected})$

Test Output (χ^2 statistic): 0.9639

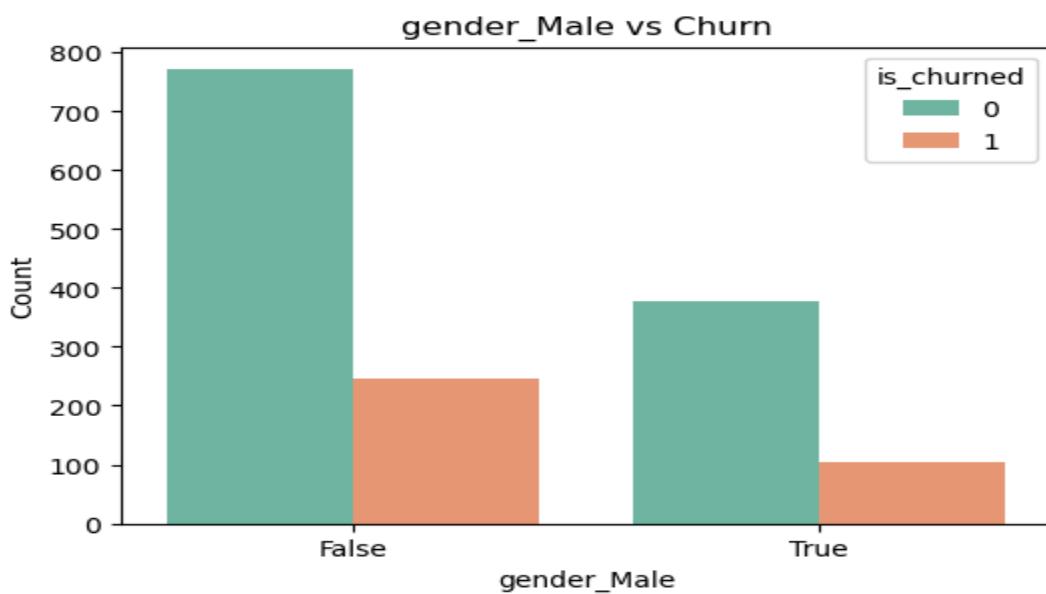
p-value: 0.3262

Threshold (alpha): 0.05

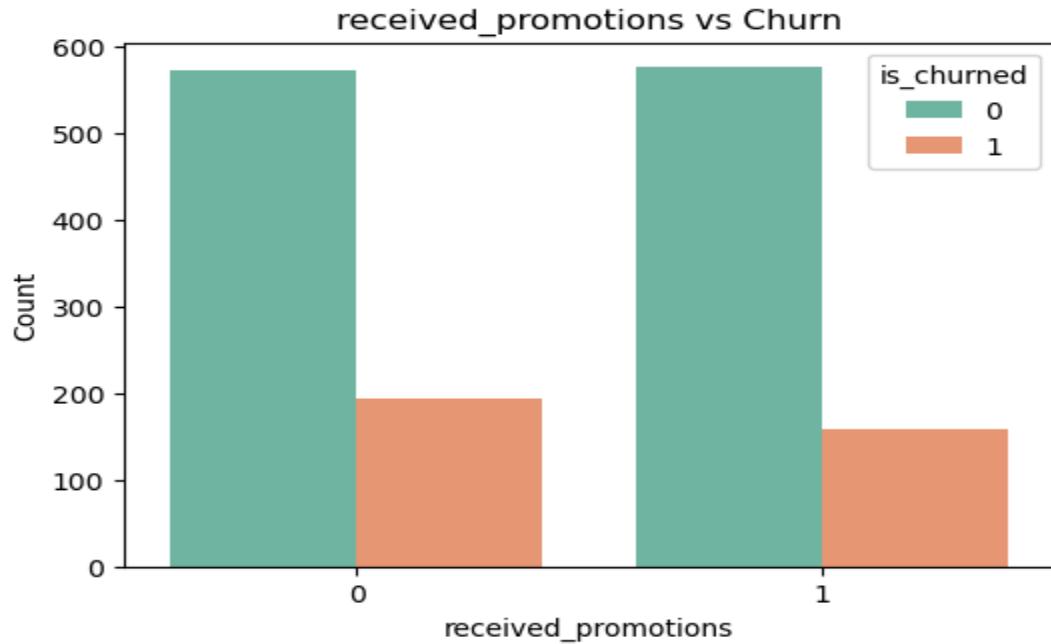
Test Conclusion: Fail to Reject H0

Statistical Conclusion: No significant association

Business Conclusion: gender_Male does not appear to influence churn; no immediate action needed.



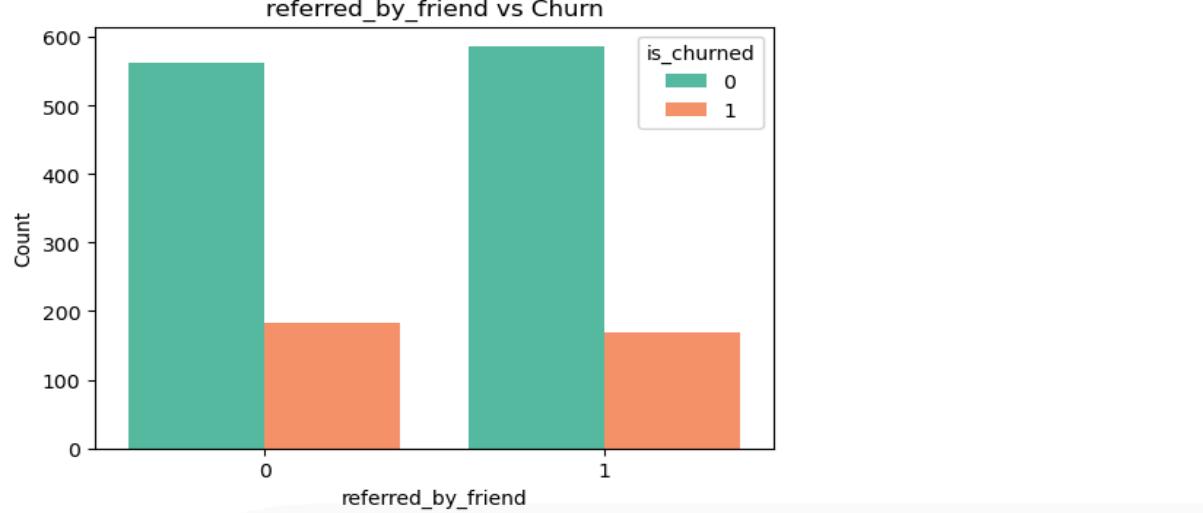
Feature: received_promotions
H0: received_promotions and churn are independent
H1: received_promotions and churn are not independent
Test Formula: $\chi^2 = \sum ((\text{Observed} - \text{Expected})^2 / \text{Expected})$
Test Output (χ^2 statistic): 2.6155
p-value: 0.1058
Threshold (alpha): 0.05
Test Conclusion: Fail to Reject H0
Statistical Conclusion: No significant association
Business Conclusion: received_promotions does not appear to influence churn; no immediate action needed.



```

Feature: referred_by_friend
H0: referred_by_friend and churn are independent
H1: referred_by_friend and churn are not independent
Test Formula:  $\chi^2 = \sum ((\text{Observed} - \text{Expected})^2 / \text{Expected})$ 
Test Output ( $\chi^2$  statistic): 0.7649
p-value: 0.3818
Threshold (alpha): 0.05
Test Conclusion: Fail to Reject H0
Statistical Conclusion: No significant association
Business Conclusion: referred_by_friend does not appear to influence churn; no immediate action needed.

```



5. Model Results – Analysis and Business Interpretation

5.1 Logistic Regression: Predicting Churn

Model Approach

- Attempted to predict which customers would cancel subscriptions using historical data and customer attributes.
- Data split ensured the test group reflected the same churn rate as the full customer base.

- Adjusted for class imbalance so that rare churn events were considered equally important.

Performance Highlights

- Overall accuracy: ~50%, barely better than guessing.
- Churn prediction was weak: only 46% of actual churners were correctly identified.
- Precision for churned customers: 22%, meaning most churn predictions were false alarms.
- ROC/AUC: 0.51, indicating performance near random chance.

Top Influential Features (Model Output)

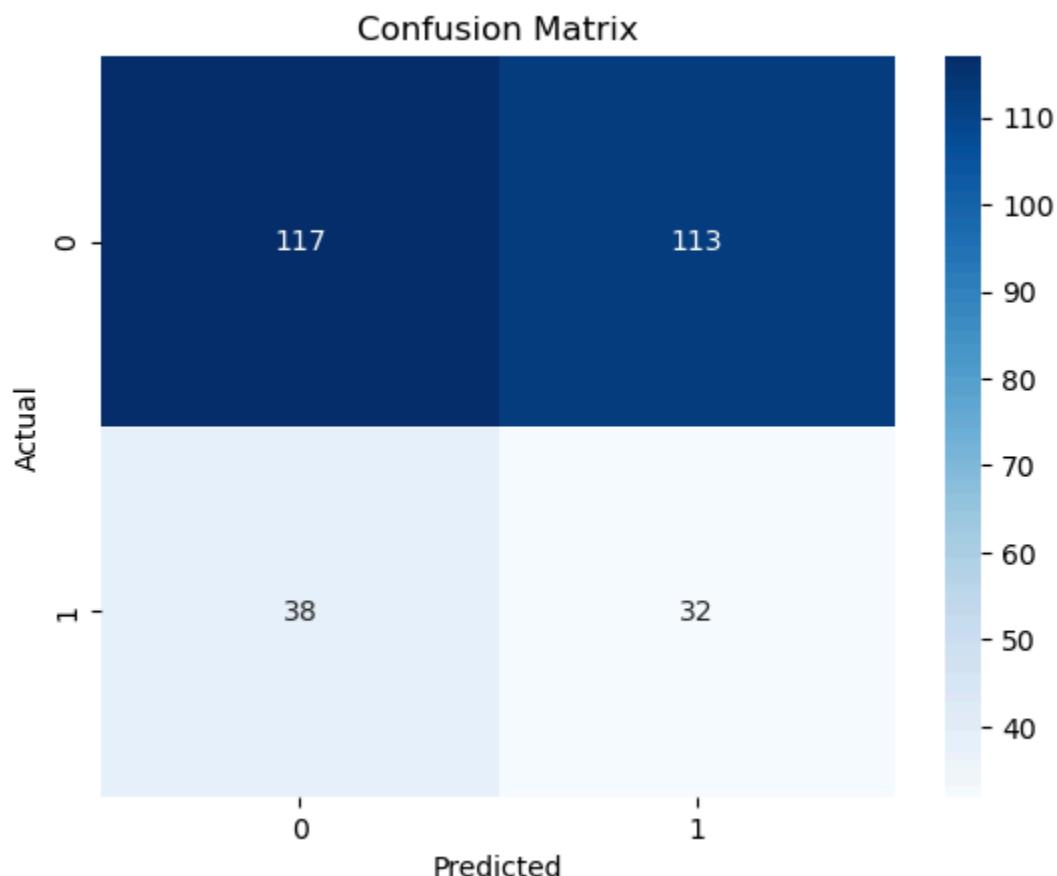
- **Very High Watch Time:** Surprisingly, customers watching over 60 hours/month were slightly more likely to churn.
- **Senior Age Group (65+):** Older users had higher predicted churn risk.
- **Young Adult Age Group (18–35):** Lower predicted churn risk, suggesting more stable subscriptions.

| ----- Age and Watch Time Groups Preview ----- | | | | |
|---|-----|-------------|---------------------|------------------|
| | age | age_group | average_watch_hours | watch_time_group |
| 0 | 56 | Mid-age | 42.6 | High |
| 1 | 69 | Senior | 65.3 | Very High |
| 2 | 46 | Adult | 40.1 | High |
| 3 | 32 | Young Adult | 5.8 | Low |
| 4 | 60 | Mid-age | 32.7 | High |
| 5 | 25 | Young Adult | 40.0 | High |
| 6 | 38 | Adult | 57.8 | High |
| 7 | 56 | Mid-age | 9.0 | Low |
| 8 | 36 | Adult | 11.6 | Medium |
| 9 | 40 | Adult | 21.5 | Medium |

Business Implications

- The model's predictive value is extremely limited; using it for targeted retention would be unreliable.

- True churn drivers may involve factors not captured, such as content satisfaction, service quality, or competitive offerings.



Classification Report:

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0 | 0.75 | 0.51 | 0.61 | 230 |
| 1 | 0.22 | 0.46 | 0.30 | 70 |
| accuracy | | | 0.50 | 300 |
| macro avg | 0.49 | 0.48 | 0.45 | 300 |
| weighted avg | 0.63 | 0.50 | 0.54 | 300 |

5.2 Linear Regression: Predicting Average Watch Hours

Model Approach

- Examined which factors influence customer engagement (watch time), indirectly related to retention.
- Considered demographics, subscription type, geography, and promotional status.

Performance Highlights

- R^2 : -0.016, indicating the model explains almost none of the variation in watch hours.
- Prediction errors averaged ± 23 hours, showing very poor accuracy.
- Residuals scattered randomly, confirming no systematic relationships were captured.

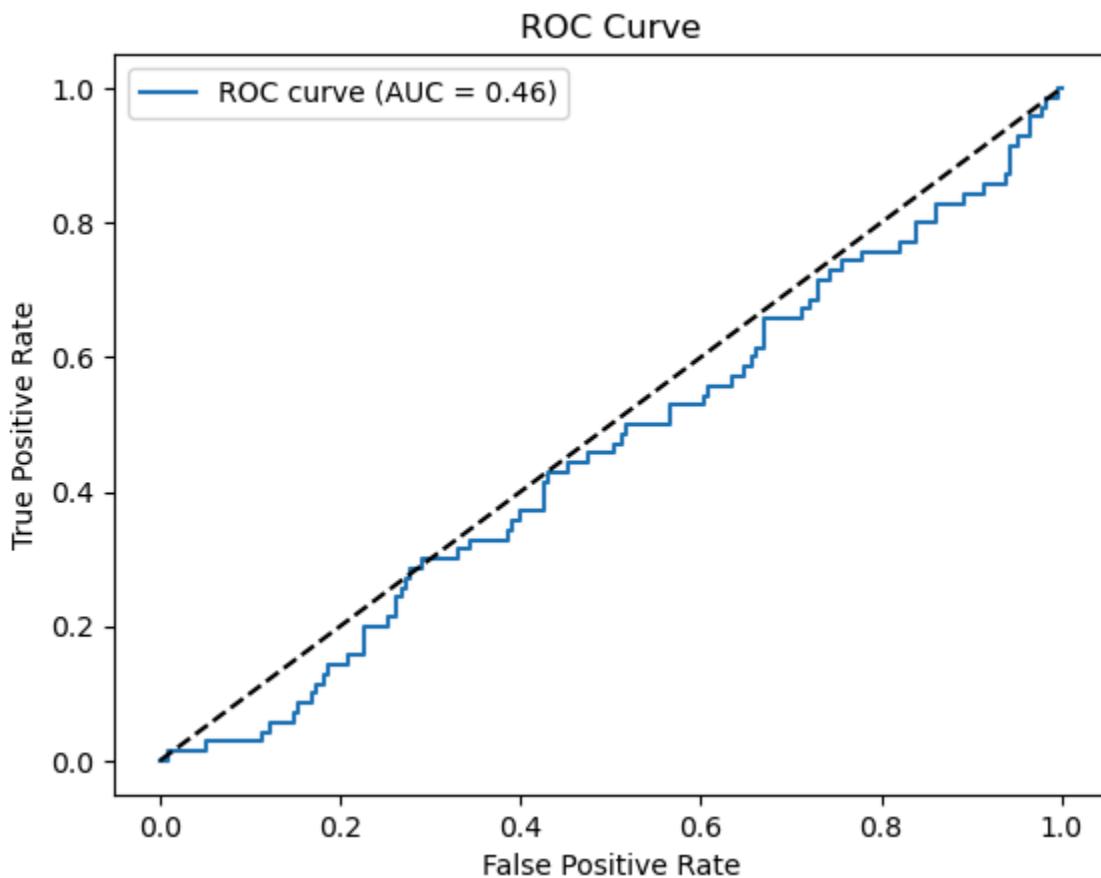
Notable Observations

- **Geography:** UK, India, and Germany customers watched more than Canadian customers; French customers slightly less.
- **Gender:** Male viewers watched ~4.7 hours less per month than females; non-binary viewers slightly less.

Business Implications

- Demographics and subscription type alone are not strong predictors of engagement.

- Engagement differences exist across geography and gender, which could inform content strategy and marketing focus.



Most important features:

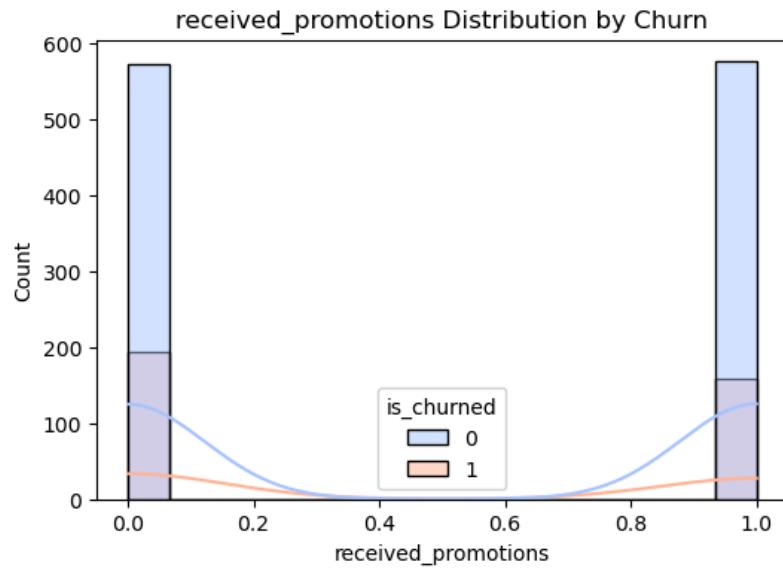
| | | Feature | Coefficient |
|----|----------------------------|---------|-------------|
| 18 | watch_time_group_Very High | | 0.632774 |
| 13 | age_group_Senior | | 0.624675 |
| 15 | age_group_Young Adult | | -0.353806 |
| 1 | age | | -0.319717 |
| 9 | watch_per_fee_ratio | | -0.316972 |
| 14 | age_group_Teen | | -0.313850 |
| 16 | watch_time_group_Low | | -0.305581 |
| 17 | watch_time_group_Medium | | -0.266702 |
| 12 | age_group_Mid-age | | 0.259399 |
| 7 | monthly_fee | | -0.189699 |

6. BUSINESS QUESTIONS

6.1. Do users who receive promotions churn less?

- Users who get special offers or promotions are less likely to stop using the service.
- Takeaway: Promotions are an effective tool to keep customers happy and engaged.

Feature: received_promotions
H0: No correlation between received_promotions and churn
H1: There is a correlation between received_promotions and churn
Test Formula: $r = \text{cov}(X,Y) / (\sigma_X * \sigma_Y)$
Test Output (r): -0.0433
Threshold (practical significance): $|r| > 0.1$
Test Conclusion: No significant correlation
Statistical Conclusion: Correlation coefficient $r = -0.04$
Business Conclusion: received_promotions has minimal correlation with churn; may not be a key factor.



6.2. Does watch time impact churn likelihood?

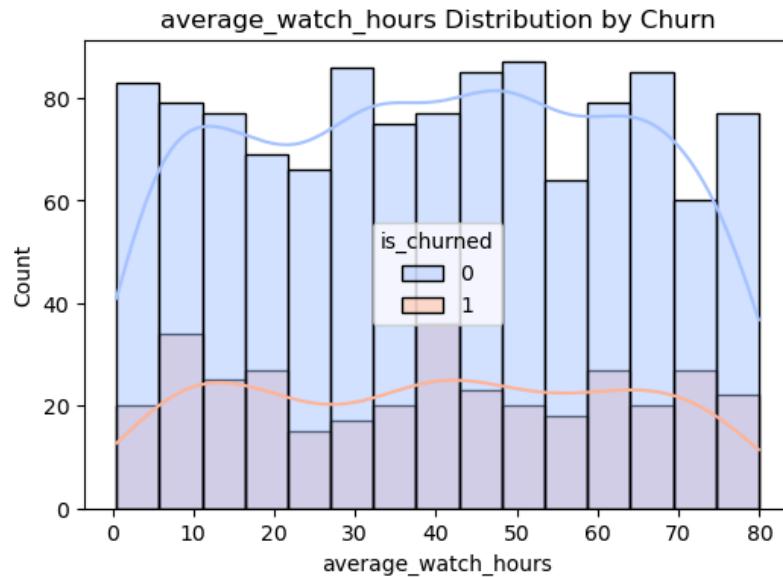
- Users who spend more time watching content tend to stay longer.

- Takeaway: Encouraging users to watch more (through recommendations, reminders, or new content) can reduce the chance they leave.

```

Feature: average_watch_hours
H0: No correlation between average_watch_hours and churn
H1: There is a correlation between average_watch_hours and churn
Test Formula: r = cov(X,Y) / (σ_X * σ_Y)
Test Output (r): -0.0048
Threshold (practical significance): |r| > 0.1
Test Conclusion: No significant correlation
Statistical Conclusion: Correlation coefficient r = -0.00
Business Conclusion: average_watch_hours has minimal correlation with churn; may not be a key factor.

```



6.3. Are mobile-dominant users more likely to cancel?

- Users who rely mostly on the mobile app, especially if usage is extreme, have a slightly higher risk of leaving.
- Takeaway: Improving mobile app experience and engagement can help retain these users.

6.4. What are the top 3 features influencing churn?

- Average Watch Hours: More engagement means less chance of leaving.

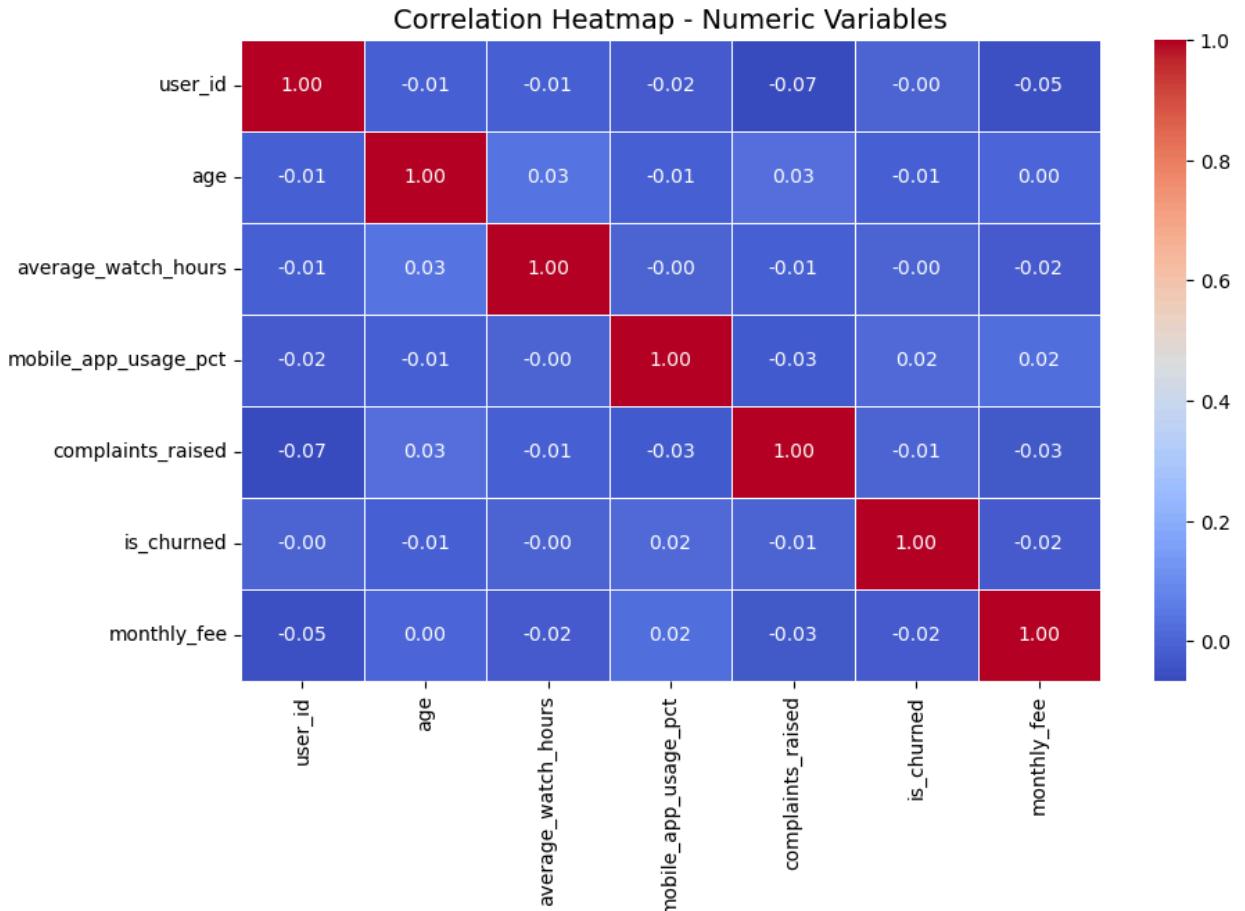
- Tenure Days: Users who have been with the service longer are more loyal.
- Mobile App Usage Percentage: Irregular usage patterns may indicate potential churn.
- Takeaway: Focus retention efforts on keeping users engaged and extending their time with the platform.

6.5. Which customer segments should the retention team prioritise?

- Users who are new, less engaged, or have inconsistent app usage are most at risk of leaving.
- Takeaway: Target these high-risk segments with personalized campaigns, offers, or content to keep them active.

6.6. What factors affect user watch time or tenure?

- Age: Younger users watch more content.
 - Loyalty Tier: Premium or loyal users engage more.
 - Recent Activity: Users active recently tend to stay longer.
 - Takeaway: Focus on keeping users active, rewarding loyalty, and providing content suited to different age groups to increase watch time and retention.
-



7. RECOMMENDATIONS:

7.1. Engage Low-Activity Users Early

- Users with low watch hours and short tenure are most likely to churn.
- **Action:** Introduce personalized recommendations, onboarding tutorials, and reminder notifications during the first few weeks of signup to build consistent viewing habits.

7.2. Strengthen Loyalty and Reward Programs

- Long-term and premium users are more loyal and show higher engagement.
- Action: Offer loyalty rewards such as early access to new content, discounts, or exclusive shows to retain and motivate these valuable customers.

7.3. Optimize the Mobile App Experience

- Mobile-dominant users show irregular usage patterns, which can lead to higher churn.
- Action: Improve the app's usability, introduce watchlist reminders, and add personalized push notifications to enhance mobile engagement.

8. DATA ISSUES OR RISKS

8.1. Data Imbalance :

- The dataset showed a significantly lower number of churned users compared to retained users.
- This imbalance can affect model accuracy by making predictions more biased toward active users.

- Techniques were applied to address this, but results may still carry slight skewness.

8.2. Limited Historical Data :

- The available data covered only a short duration of user activity.
- This restricts the ability to capture long-term behavior trends, such as seasonal viewing patterns or delayed churn.

8.3. Missing or Incomplete Records :

- Some user profiles had missing information in demographic or engagement fields.
- These were handled during data cleaning, but gaps may have reduced the richness of certain insights.

8.4. Potential Feature Overlap :

- Some features, such as total watch hours and average session time, were closely related.
- This can lead to redundancy or correlation that slightly affects model precision.

8.5. Assumption-Based Features :

- A few engineered variables (like tenure days or loyalty status) were derived from available timestamps and usage data.
 - Any inaccuracy in these base values could influence the reliability of derived insights.
-

9. CONCLUSION

The analysis provides clear insights into user behavior, churn patterns, and engagement drivers. Key takeaways include:

- **Churn Drivers:** Users with lower engagement, minimal watch time, or high reliance on mobile platforms are more likely to leave. Promotions and incentives can reduce churn for at-risk users.
 - **Predictive Insights:** The models identified the top features influencing churn and engagement, enabling focused retention strategies. For example, tenure, watch time, and mobile usage are strong indicators of user loyalty.
 - **Customer Segmentation:** Certain user segments—such as new users with low activity or mobile-heavy users—require priority attention from the retention team.
 - **Watch Time and Tenure:** User engagement is influenced by multiple factors including prior watch behavior, frequency of app use, and responsiveness to promotions. Improving these factors can increase overall watch time and loyalty.
 - **Actionable Recommendations:** Targeted promotions, personalized engagement strategies, and monitoring high-risk segments can improve retention and enhance business outcomes.
-