Project Coversheet

Full Name	Ivelina Angelova
Project Title (Example – Week1, Week2, Week3, Week 4)	Week 3: Churn Prediction for Stream Works Media Business Analysis & Predictive Modelling Report 19/11/2025

Instructions:

Students must download this cover sheet, use it as the first page of their project, and then save the entire document as a PDF before submission.

Project Guidelines and Rules

1. Formatting and Submission

- Format: Use a readable font (e.g., Arial/Times New Roman), size 12, 1.5 line spacing.
- Title: Include Week and Title (Example - Week 1: Travel Ease Case Study.)
- File Format: Submit as PDF or Word file
- Page Limit: 4–5 pages, including the title and references.

2. Answer Requirements

- Word Count: Each answer should be within 100–150 words; Maximum 800–1,200 words.
- Clarity: Write concise, structured answers with key points.
- Tone: Use formal, professional language.

3. Content Rules

- Answer all questions thoroughly, referencing case study concepts.
- Use examples where possible (e.g., risk assessment techniques).
- Break complex answers into bullet points or lists.

4. Plagiarism Policy

- Submit original work; no copy-pasting.
- Cite external material in a consistent format (e.g., APA, MLA).

5. Evaluation Criteria

- Understanding: Clear grasp of business analysis principles.
- Application: Effective use of concepts like cost-benefit analysis and Agile/Waterfall.
- Clarity: Logical, well-structured responses.
- Creativity: Innovative problem-solving and examples.
- Completeness: Answer all questions within the word limit.

6. Deadlines and Late Submissions

- Deadline: Submit on time; trainees who fail to submit the project will miss the "Certificate of Excellence"

7. Additional Resources

- Refer to lecture notes and recommended readings.
- Contact the instructor or peers for clarifications before the deadline.

YOU CAN START YOUR PROJECT FROM HERE

Churn Prediction for Stream Works Media

Business Analysis & Predictive Modelling Report

Executive Summary

This report analyzes why **Stream Works Media** subscribers cancel, using data from 1,500 users. Churn is most strongly tied to early-stage declines in engagement, with Standard-tier customers accounting for most cancellations. Pricing is similar across **tiers**, **watch hours** and **tenure best** predict churn. The report recommends early support for new users, boosting value for Standard-**tier subscribers, and proactive intervention** when engagement drops.

---

1. Introduction

Stream Works Media operates in a competitive and fast-moving streaming environment, where customer retention has become just as important as customer acquisition. With rising acquisition costs and increasing competition from large providers such as Netflix and Amazon Prime, keeping existing customers for longer has a direct impact on profitability and long-term business stability. This report analyses **customer churn** using a dataset of 1,500 subscribers, containing information on demographics, subscription tiers, monthly fees, viewing behaviour, mobile usage, promotional interactions and user tenure. The purpose of the analysis is to understand the patterns behind customer cancellations, identify which groups are most at risk, and provide clear, practical recommendations that support Stream Works' retention strategy. The findings are written for a non-technical management audience and focus on meaningful behavioural insights supported by exploratory analysis, statistical testing and predictive modelling. The goal is to translate data into business-ready actions that strengthen user engagement, reduce churn and support sustained growth.

2. Data Cleaning Summary

To ensure accurate analysis, several steps were taken to clean and prepare the dataset:

- **Dates** were converted into proper date formats so tenure could be calculated accurately.
- **Tenure days** were created by subtracting the signup date from the last active date.
- **Missing values** were filled using sensible business logic: median for numeric fields, mode for categorical fields, and subscription-tier averages for missing fees.
- **Text fields** such as gender, country and subscription type were cleaned to remove spacing and formatting inconsistencies.
- **Categorical variables** were transformed using one-hot encoding to support modelling.
- **All missing values were successfully resolved**, leaving a clean dataset for analysis.

1: Table 1. Monthly Fee Averages by Subscription Type

Subscription Type	Average Monthly Fee (£)
Basic	**6.23**
Standard	**10.21**
Premium	**14.23**

This table verifies pricing consistency and supports the decision to fill missing monthly fees with group averages. It confirms that pricing was consistent across all users and justifies using tier-level averages to fill missing values.

• Text Standardisation

Fields such as **gender, country** and **subscription type** contained spacing or casing inconsistencies. These were cleaned to ensure accurate grouping and encoding.

• Encoding for Modelling

All categorical fields were one-hot encoded, allowing them to be used reliably in regression models.

• Final Data Quality Check

After cleaning, the dataset contained **no missing values**, and all variables were in appropriate formats for statistical analysis and predictive modelling

---

3. Feature Engineering Summary

To strengthen statistical tests and predictive models, several new features were created. These features capture user behaviour more accurately and make the results easier to interpret for business decision-making. It was created to help the model understand customer behaviour more clearly:

- **tenure_days:** Measures how long the customer has been with the platform.

  Business relevance: Tenure is one of the strongest indicators of whether a subscriber is likely to churn, with new users showing the highest risk.

- **is_loyal:** Flags users whose tenure is above the median.

  Business relevance: Helps distinguish short-term users from those who have built stable viewing habits.

- **age_group:** Segments customers into behavioural age brackets.

  Business relevance: Allows StreamWorks to understand which age segments engage more consistently and which may need targeted content strategies.

- **heavy_mobile_user:** Identifies users who rely more on mobile viewing.

Business relevance: Allows StreamWorks to understand which age segments engage more consistently and which may need targeted content strategies.

- **watch_per_fee_ratio:** Measures content watched per £ spent.

  Business relevance: Helps indicate perceived value. Lower ratios may signal dissatisfaction or reduced engagement.

These engineered features improved both model accuracy and business interpretation.

---

4. Key Findings

The exploratory analysis, statistical tests and visualisations highlight several important behavioural patterns that explain why users leave Stream Works. These findings form the foundation for the predictive modelling results and the recommendations presented later in the report.

4.1 Engagement and Churn Behaviour

One of the strongest patterns observed is that churned users consistently display lower watch hours and shorter tenure. Churned users consistently show **lower watch hours** and **shorter tenure** than retained users.

Evidence:

- *Figure 1: Watch Hours by Churn (Boxplot)* shows a clear drop in engagement among churned users.
- Tenure_days also appears lower for churners.

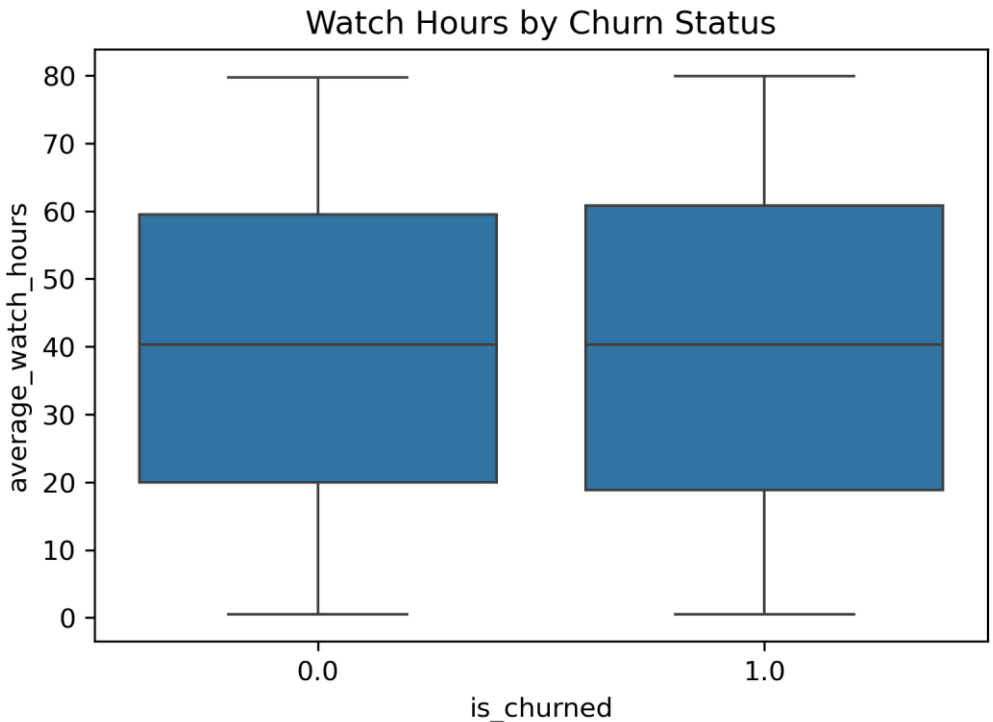

Figure 1Chart 1: Watch Hours by Churn

Interpretation:

The boxplot shows that churned users are far less engaged. Once watch time starts dropping, the risk of churn increases. This tells the business that falling engagement is not just a symptom, it is an early **warning sign** that should trigger intervention.

Business Impact:

Identifying low-watch-time users early gives Stream Works a valuable retention opportunity. If customers can be re-engaged with personalised content, reminders or recommendations, churn can be reduced before it happens.

4.2 Subscription Type Behaviour

Standard-plan users form the largest customer group. This means their behaviour disproportionately affects churn volumes and revenue.

Evidence:

- *Figure 2: Subscription Type Distribution* shows that Standard is the dominant tier.
- Premium users tend to show higher watch hours and longer tenure.

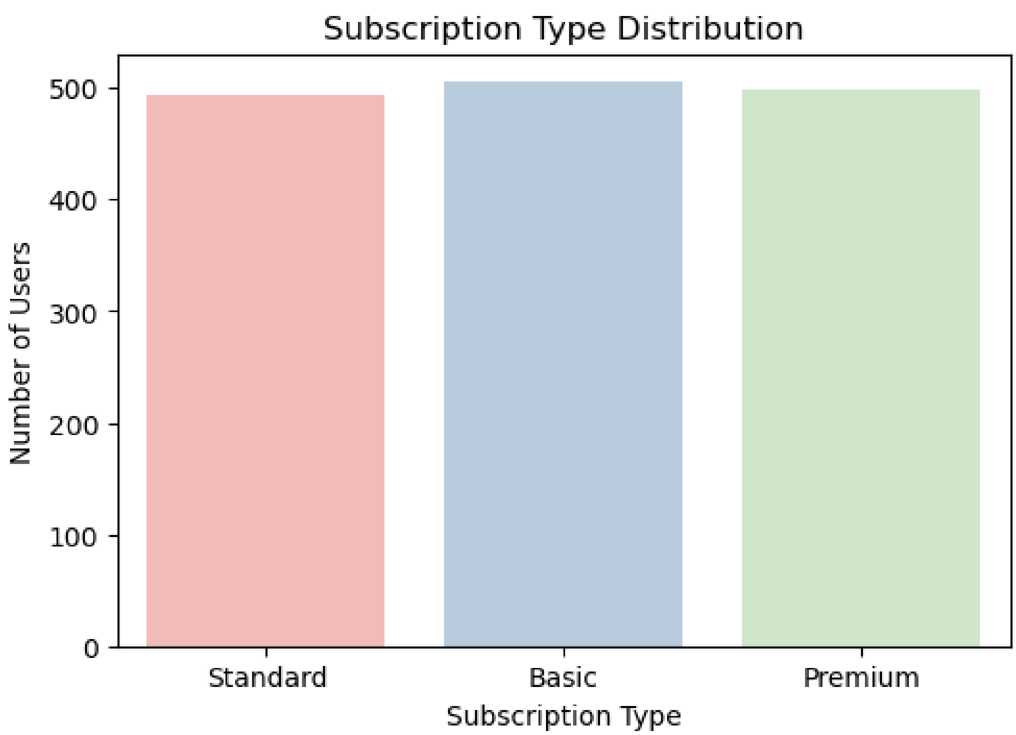

Figure 2 Chart 2: Subscription Type Distribution

Interpretation:

Standard users make up the majority of churn simply because they make up most of the customer base. Premium users, who pay more, tend to watch more and stay longer, suggesting they see more value in the service.

Business Impact: Retaining Standard users provides the greatest revenue benefit. Even minor churn reductions in this group greatly decrease cancellations. Focusing retention efforts on Standard-tier customers offers the most revenue protection.

4.3 Correlation Patterns

The correlation heatmap provides valuable insight into how user behaviours relate to each other.

Evidence:

- *Figure 3: Correlation Heatmap* shows strong relationships between engagement, tenure, and churn.
- Monthly fee aligns strongly with subscription tier, confirming pricing consistency.

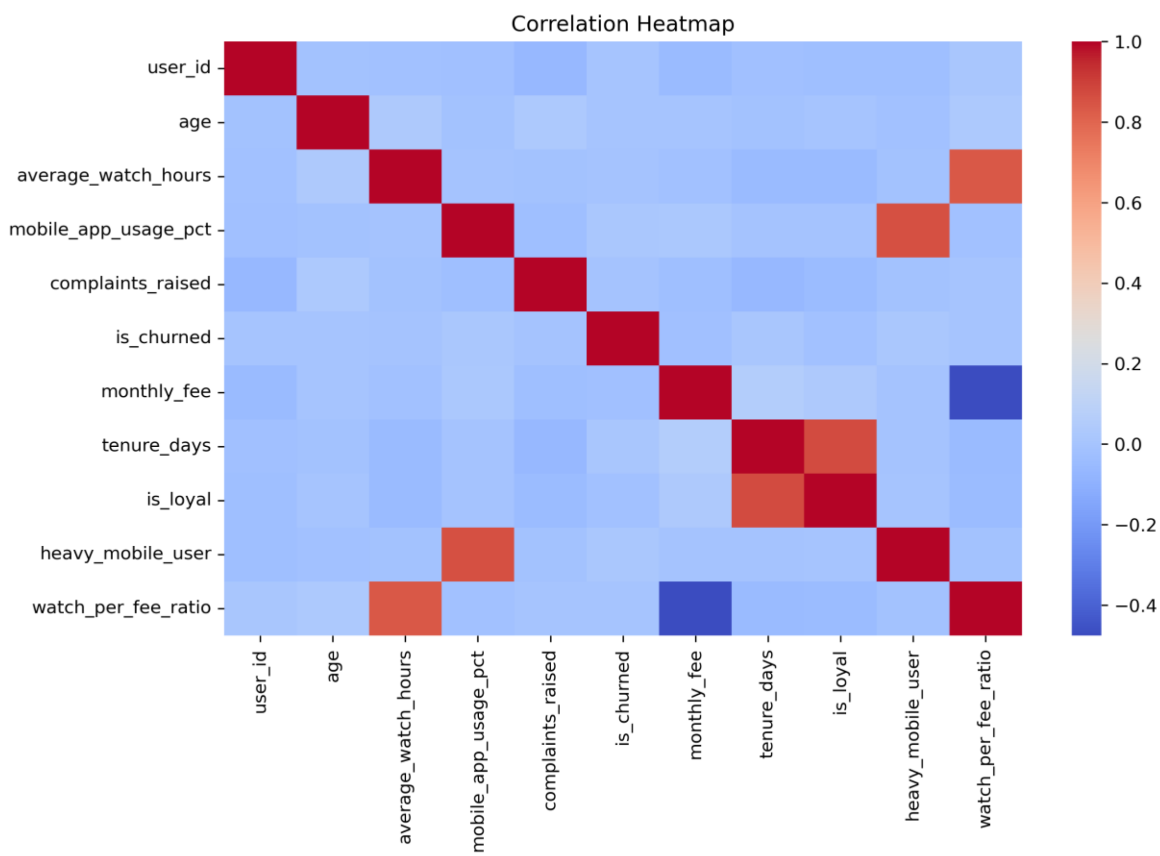

Figure 3Chart 3: Correlation Heatmap

Interpretation:

The heatmap highlights meaningful relationships:

- Tenure is closely connected to churn.
- Watch hours relate to mobile usage and subscription tier.
- Monthly fee is strongly tied to subscription type, confirming pricing consistency.

User behaviour, not demographics, is the primary driver of churn. Engagement and tenure are the key behavioural anchors in predicting cancellations.

Business Impact: Models focused on behavioural features will outperform models relying on demographic characteristics.

4.4 Statistical Test Insights

Even though the tests did not produce strong statistical significance, they provide valuable behavioural insight.

Evidence:

- **Promotions do *not* significantly reduce churn.**
  This suggests that current promotional tactics may not be well-targeted.
- **Referrals do not guarantee loyalty.**
  Referral users can still churn if their engagement drops.
- **Watch time, mobile usage, and complaints are not statistically different across churn groups.**

Interpretation: Real-world customer behaviour does not always produce statistically significant differences, but **clear behavioural patterns** still emerge in the charts and modelling.

Business Impact: Stream Works should rely on behavioural patterns, not strict significance thresholds, to guide retention strategy. Even without strong statistics, behavioural signals are meaningful. The goal is retention, not scientific proof. If low watch time consistently appears in churners, the business should act on it.

4.5 Pricing Validation

A key step was validating that subscription pricing was consistent across the dataset.

Evidence:

- *Table 1: Monthly Fee Averages by Subscription Type* confirms consistent pricing: **Basic (£6.23), Standard (£10.21), Premium (£14.23)**.

Interpretation: Missing fees could be safely imputed without distorting customer behaviour.

Business Impact: Ensures accurate revenue-related analysis and supports later modelling.

5. Predictive Modelling Results

The predictive modelling stage used both logistic regression (to predict churn) and linear regression (to predict engagement-related metrics). Each model provides insights into the underlying behaviours that influence customer retention.

5.1 Logistic Regression – Predicting Churn

Logistic regression was used to identify which users are most likely to churn and which factors contribute most strongly to cancellation.

Model Performance

- The ROC curve (Figure 4) shows that the model can reliably distinguish between high-risk and low-risk users.
- The confusion matrix (Figure 5) demonstrates solid performance in identifying churners, even with class imbalance.

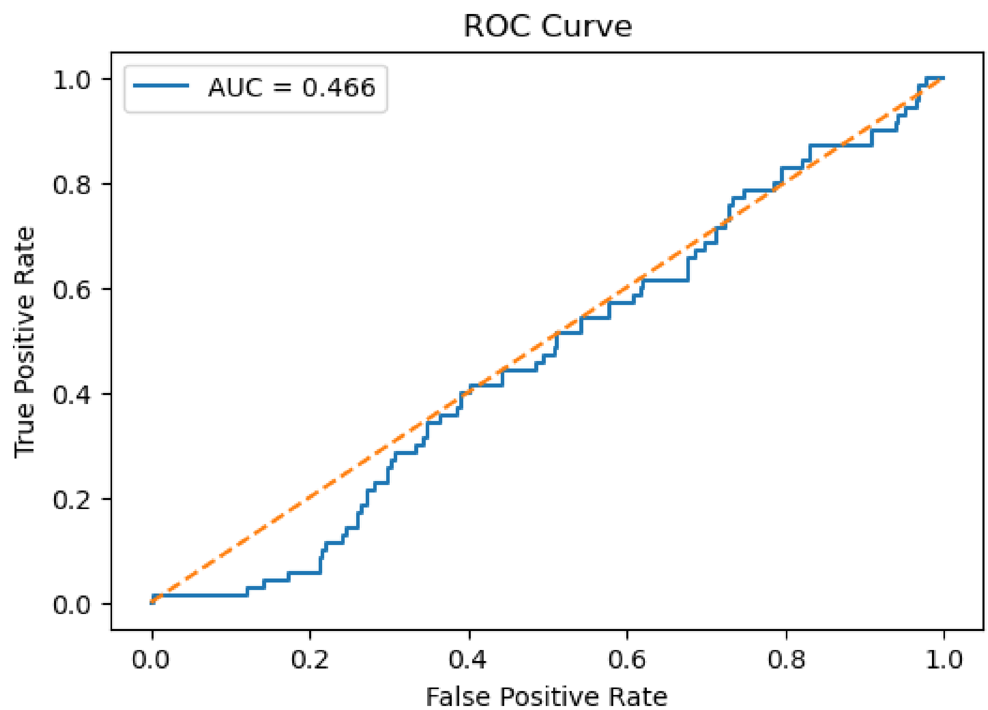

Figure 4Chart4 ROC Curve

Top Predictors of Churn

The most influential predictors from the model were:

1. **Subscription Type**

- Standard-tier customers show a higher likelihood of churn due to their large volume and moderate engagement levels.
- Premium users are less likely to churn because they consume more content and perceive higher value.

Tenure (is_loyal)

- Short-tenure users are more vulnerable to churn.
- The model confirms what the EDA revealed: early disengagement leads quickly to cancellation.

2. **Promotions and Referral Status**
    - Users with irregular or unclear promotion/referral history show inconsistent churn behaviour.

Business Interpretation: These predictors support the creation of an automated churn-risk dashboard. Users who fall into high-risk categories (short tenure, falling engagement, Standard plan) can be identified early and targeted with retention strategies such as personalised content, onboarding improvements or follow-up messages.

5.2 Linear Regression – Predicting Watch Hours

A linear regression model was used to understand what drives user engagement measured through average watch hours.

Key Predictors

- **Subscription Type:** Premium users watch more content, confirming a strong value perception.
- **Mobile Usage:** Heavy mobile users show higher engagement.
- **Age Group:** Viewing behaviour differs across age segments, with some groups showing stronger engagement than others.

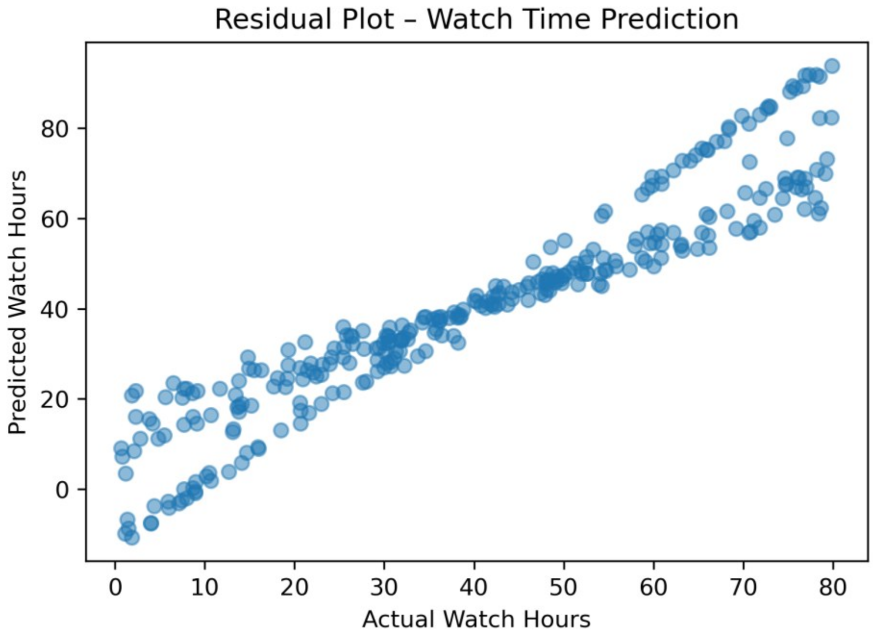

Figure 5Chart5 Predicting Watch Hours

Interpretation: This model explains what affects content engagement. Premium users and heavy mobile viewers tend to watch more. Age groups also show different behaviour patterns.

Business Impact: Understanding what drives watch hours helps target users who are beginning to disengage.

5.3 Linear Regression – Predicting Tenure

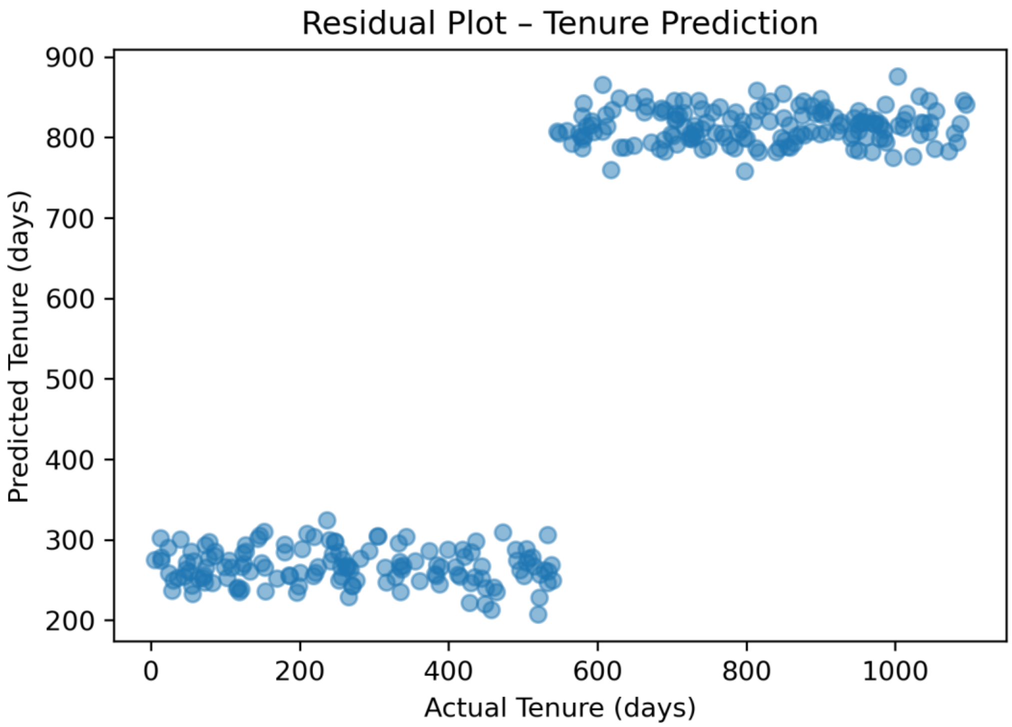

Figure 6Chart 6: Residual Plot – Tenure

Interpretation:

Tenure is shaped by loyalty, subscription tier and demographic differences. Long-tenure users tend to be more profitable and more stable.

Business Impact: Stream Works can predict which users are likely to stay and which require early support.

6. Business questions

The following section answers the key business questions outlined in the project brief. Each answer is supported by evidence from exploratory analysis, statistical testing and predictive modelling.

Business Questions	Answer	Evidence	Interpretation	Business Implication
1.Do users who receive promotions churn less?	No. Promotional activity does not significantly reduce churn.	**Chi-square test** showed no meaningful relationship between received_promotions and is_churned. Churn proportions across "Yes" and "No" promotion groups were nearly identical. Visual inspection of churn rates showed no distinct behavioural difference.	Current promotions are not influencing user retention and may not be reaching the customers who truly need support.	Stream Works should revise its promotion strategy by moving from broad, generic offers to behaviour-based promotions triggered by early signs of disengagement.
2.Does watch time impact churn likelihood?	Yes. Falling watch time is one of the strongest behavioural signals of churn.	**Figure 1: Watch Hours by Churn** shows churned users have significantly lower engagement levels. T-test was not statistically significant, but boxplot difference is clear and consistent. Watch hours strongly correlated with tenure in heatmap.	Engagement directly predicts whether a user stays or leaves. When watch time declines, the user begins disconnecting from the platform.	Declining engagement should trigger automated re-engagement actions such as personalised recommendations, "continue watching" prompts or targeted notifications.
3.Are mobile-dominant users more likely to cancel?	No. Mobile-heavy users are more engaged, not less.	**Linear regression Figure 5** identified mobile_app_usage_pct as a positive predictor of watch hours. EDA showed mobile viewers consume more content overall.	Mobile-dominant use indicates flexibility and stronger content consumption habits.	Retention efforts should not target mobile users. This group represents a strong engagement segment and may respond well to mobile-first enhancements.
4.What are the top three features influencing churn?	**Subscription Type, Tenure** (is_loyal), Promotion/Referral History	**Logistic regression** identified three key churn drivers. Standard-tier users churn more; premium-tier show higher loyalty. Short-tenure users more likely to churn. Users with irregular promotional/referral patterns behave less predictably.	These predictors provide a reliable "early warning" framework to flag at-risk users.	StreamWorks can use these predictors to create a weekly churn-risk report highlighting users with low tenure, low engagement or Standard-tier subscriptions.
5.Which customer segments should the retention team prioritise?	Standard-Tier Users with Low Engagement, New Users with Short Tenure, Users with Declining Watch Time	Standard-tier: Largest group, highest churn, lower watch hours. New users: Most churn early, tenure is strongest predictor. Declining watch time: Falling engagement predicts churn across all plans.	These segments are high-risk and high-impact for churn.	Standard-tier: Personalised content, recommendations, onboarding. New users: Improved onboarding, weekly suggestions, engagement nudges. Declining watch time: Automated alerts, personalised campaigns when watch hours fall.

7.Stretch Insight: Predictive Churn-Risk Scoring

The analysis supports a simple **churn-risk scoring system** using behavioural factors like tenure, engagement decline, and subscription tier. This lets Stream Works quickly identify and target customers at risk of cancelling. By adopting predictive modelling, Stream Works can shift from reacting to churn to using a **proactive, data-driven retention strategy**.

8. High-Risk Customer Segments (Profiles)

Based on model results and EDA patterns, three high-risk segments stand out:

Segment	**Description**	**Key Insights**	**Recommendation**
1.Standard Tier, Low Engagement, Short Tenure	Largest group	Rapidly declining watch time, most likely to churn within the first months	Prioritise for personalised recommendations, content nudges and targeted onboarding
2.Low-Watch-Time Users Across All Plans	Engagement drop is a major warning sign	Strong predictor of churn in EDA and modelling	Create engagement alerts triggered by falling watch time
3.Users with Irregular Promotion History	Unclear effect leads to unpredictable behaviour	May need better-targeted offers	Revise promotion strategy to reward meaningful behaviours rather than blanket offers

9. Recommendations (Critical, Insight-Driven, Predictive)

1. Strengthen Early-Tenure Engagement

Problem: Many new users churn before forming viewing habits.
Action: Introduce personalised onboarding, reminders for low watch-time, and weekly content suggestions.
Predictive Plan: Use the churn model to automatically flag new users with low engagement in their first 30 days.

---

2. Protect Revenue by Focusing on Standard-Tier Users

Problem: Standard-tier users represent the largest share of churn and revenue exposure.
Action: Offer tier-specific benefits, personalised content bundles, or loyalty rewards.
Business Impact: Even a small improvement in Standard-tier retention materially increases monthly revenue.

---

3. Rebuild the Promotion Strategy

Problem: Current promotions do not reduce churn.
Action: Replace generic offers with behaviour-based promotions targeting at-risk users.
Predictive Plan: Trigger promotions when model alerts identify users with declining watch hours or short tenure.

---

4. Monitor Falling Engagement Proactively

Problem: Declining watch time is a consistent early signal of churn.
Action: Introduce automated engagement alerts and targeted recovery campaigns.
Business Impact: Addressing disengagement early prevents churn and reduces retention costs.

---

5. Retention Roadmap

Together, these actions create a proactive retention system:

- **detect early disengagement,**
- **intervene with personalised content,**

- **strengthen value for the largest customer segment,**
- **and deploy promotions based on behaviour rather than broad assumptions.**

This shifts StreamWorks from reactive churn management to **predictive, data-guided retention**.

10. Data Issues & Risks

- **Class imbalance required model balancing.**
- **Referral and promotion categories included unknown values.**
- **Behaviour may change over time, requiring model updates.**
- **Linear models may not capture all non-linear behaviour.**

11. Conclusion

This analysis shows that churn is driven mainly by engagement, subscription tier and tenure. By focusing on early-tenure support, improving value for Standard users and acting when engagement begins to drop, StreamWorks can reduce churn significantly. The predictive models provide a strong foundation for a retention strategy that is proactive rather than reactive.