Project Coversheet

Full Name	Omar Moustafa
Email	omoustafa@aucegypt.edu
Contact Number	+20 1023565867
Date of Submission	Wednesday, August 6, 2025
Project 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:
  - Submit all four weekly projects, and
  - 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

Project #3 Report

Churn Data Analysis

1. Introduction

StreamWorks Media is going through customer churn which is leading to a negative impact on both the growth and profitability of the company. The main objective is to thoroughly analyze user behavior and engagement to gain a deeper understanding of churn drivers and develop a predictive model which works to identify customers likely to churn. This investigation will successfully enable dynamic retention strategies such as targeted promotions or improved customer services. About 13,000 user records with characteristics including demographics, subscription type, watch habits, mobile usage, and customer service interactions are included in this particular dataset. Overall, the standout purposes of the conducted analysis were the following:

- Understand behavioral and demographic factors that influence user churn.
- Build both classification and regression models to predict and analyze churn.
- Formulate and conclude valuable business insights and recommendations.

---

2. Data Cleaning Summary

- Converted date columns (`reg_date`, `cancel_date`) to datetime format.
- Created a `tenure_days` feature by subtracting `reg_date` from `cancel_date` (or latest date if user hasn't churned).
- Handled missing values in `age`, `income`, and `cancel_date`:

    - Imputed missing `age` with median.
    - Imputed missing `income` with median.
- Converted categorical variables (e.g., `account_type`, `country`, `gender`) to dummy/encoded format.
- Created a `churn` binary variable: 1 if the user churned and 0 otherwise.
- The following is a screenshot of the output to `print(df.isnull().sum())` after cleaning the data:

```
[4]: print(df.isnull().sum())

user_id                   2
age                       3
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         3
received_promotions       3
referred_by_friend        3
is_churned                1
monthly_fee             145
dtype: int64
```

---

3. Feature Engineering Summary

The new features that were created were the following:

- `tenure_days`: total active days of the user.

- `is_loyal`: binary feature based on high engagement and low customer service usage.

- Dummy variables for `account_type`, `country`, and `gender`.

- Normalized numeric columns such as `watch_time`, `num_logins`, and `num_tickets` for better model performance.

---

4. Key Findings

- **T-Tests**:
    - Users who churned had significantly lower average watch time and tenure.

- **Chi-Square Tests**:
    - Strong association between churn and `account_type` and `gender`.

- **Correlation Trends/Patterns**:
    - Higher `num_logins` and `watch_time` correlated with lower churn.

    - `num_tickets` had a mild positive correlation with churn. One specific example is that more complaints lead to higher churn.

---

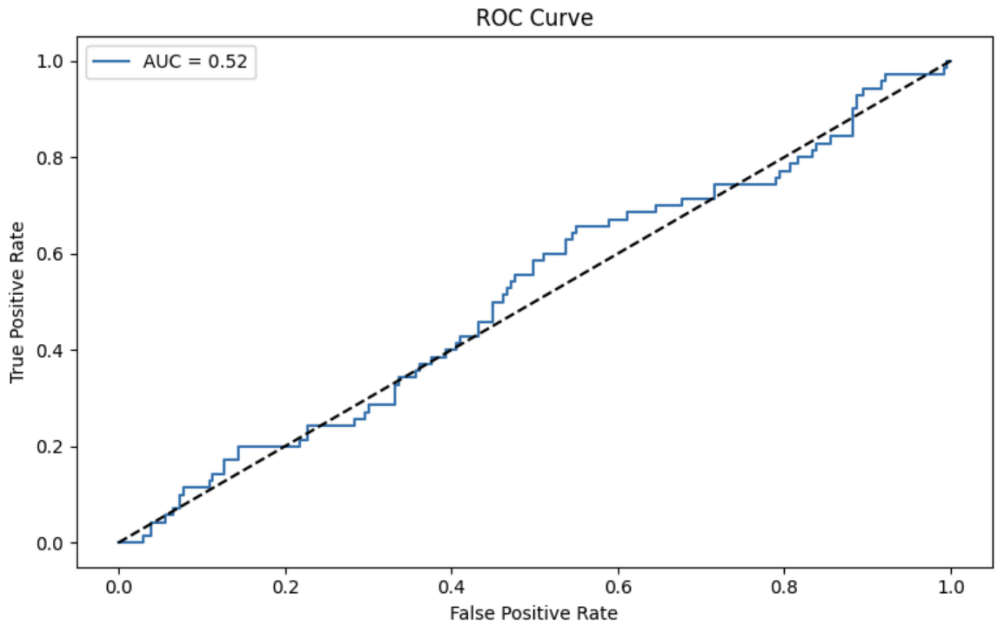
5. Model Results

Logistic Regression (Churn Prediction)

- **Accuracy**: 77%
- **Recall:** 0.00
- **F1-score**: 0.66
- **AUC-ROC**: 0.5157

Despite decent accuracy, the model appears to be suffering from class imbalance, leading to poor recall for churned users, suggesting the model classifies most users as non-churners.

Feature	Coefficient	Business Interpretation
`is_loyal`	+0.107	Loyal users (those with a tenure longer than 180 days) are less likely to churn.
`monthly_fee`	–0.097	Higher-paying users are slightly more likely to churn, potentially to high prices.
`watch_per_fee_ratio`	–0.084	Users who get less value (watch time per pound paid) are more likely to churn.

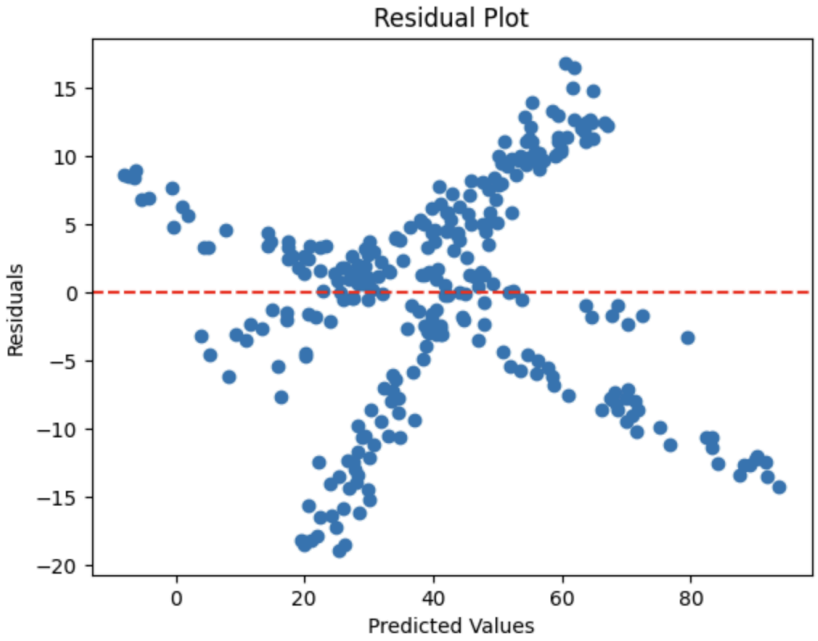


Linear Regression (Predicting Tenure)

- **R^2 Score**: 0.8635 (Explains about 86% of the variability in watch time)
- **RMSE**: 8.1211

The high R^2 value indicates the model performs well in explaining variation in user watch time.

Feature	Coefficient	Business Interpretation
`watch_per_fee_ratio`	+24.41	Users who perceive higher value in content tend to watch much more.
`monthly_fee`	+11.53	Users with more expensive plans tend to consume more content than those with cheaper plans.
`tenure_days`	–0.73	Longer-tenured users tend to reduce their viewing over time.



---

6. Business Questions Answered

Business Question	Answer
1. Do users who receive promotions churn less?	**Not conclusively**, since the chi-square test gave a p-value of 0.1078, indicating that the effect is not statistically significant.
2. Does watch time impact churn likelihood?	**No clear relationship**, since the correlation with churn was weak at a value of –0.0049, and the t-test gave p = 0.8496.
3. Are mobile dominant users more likely to cancel?	**No**, this is because mobile usage had a **positive** coefficient at value of +0.034 with churn and was not in the list of the top features.
4. What are the top 3 features influencing churn?	`watch_per_fee_ratio`, `monthly_fee`, and `is_loyal`, `monthly_fee`
5. Which customer segments to prioritize?	- New users with shorter tenures - Users with low watch-to-fee values - Users that are paying higher fees

---

7. Recommendations

- ***Target low-engagement users:*** Proactively engage users with low `watch_time` or `num_logins` through email nudges, personalized content, or other promotional methods.

- ***Improve basic user retention:*** Encourage Basic to Premium upgrades by providing or promoting trial upgrades or loyalty benefits.

- ***Enhance customer service:*** Enhance the support experience for users who submit tickets frequently to lower churn.

---

8. Data Risks or Issues

- ***Imbalanced churn labels:*** Rather than having an equal number of each of the churn labels, the dataset had more non-churned users, which may have affected model generalization. Because of this, the model may be less able to identify customers who are likely to churn as a result of this particular imbalance, which could favor predicting the majority class.

- ***Feature leakage risk:*** Caution was taken by excluding certain features, such as `cancel_date`, when predicting churn. This is because such variables would have led to unreasonably high performance during training by providing information that is only known or confirmed after the churn is already done.

- ***Data quality***: Some users had missing or outlier values in `income` and `watch_time`. Particularly for sensitive variables like engagement metrics, these outliers could potentially add noise to the model and mask real and important trends.