Project Coversheet

Full Name	Sai Kiran Dhandi
Email	Saikirandhandi25@gmail.com
Contact Number	+353 892402305
Date of Submission	27-07-2025
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:
  - 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

Intoduction

The business goal is to understand and predict customer churn for StreamWorks, a subscription-based streaming platform. The dataset streamworks_user_data.csv contains 1,500 users and 14 attributes, covering demographics, subscription details, engagement metrics, and churn status.

The purpose of this analysis is to:

- Audit and clean the customer dataset.
- Identify behavioural patterns linked to churn.
- Build a predictive model to flag at-risk customers.
- Provide data-driven recommendations to reduce churn.

Data Cleaning Summary

- Handled Missing Values: Filled missing monthly_fee values using subscription-type averages. Filled categorical nulls with mode and numeric nulls with mean. Dropped rows with missing critical dates.
- User IDs**:** Imputed missing user_id for two rows with unique values and cast the column to categorical type.
- Date Conversion: Converted signup_date and last_active_date to datetime objects to allow tenure calculations.
- Target Variable Fix: Corrected an anomalous is_churned value (0.234156) to 0.0 after inspection.
- Encoding: Prepared categorical features for modeling using label encoding (binary) and one-hot encoding (multi-class).

Feature Engineering Summary

- tenure_days: Number of days between signup and last active date, used to measure user longevity.
- is_loyal: Binary flag for users active > 180 days. Initially hypothesized as a strong churn predictor.

- Dummy Variables: Gender, country, and subscription_type were one-hot encoded for model input.
- High Mobile Usage Flag: Segmented users into high/low mobile app usage groups for churn pattern analysis.

Key Findings

- T-tests: No significant difference in average watch hours between churned and retained users (p=0.846).

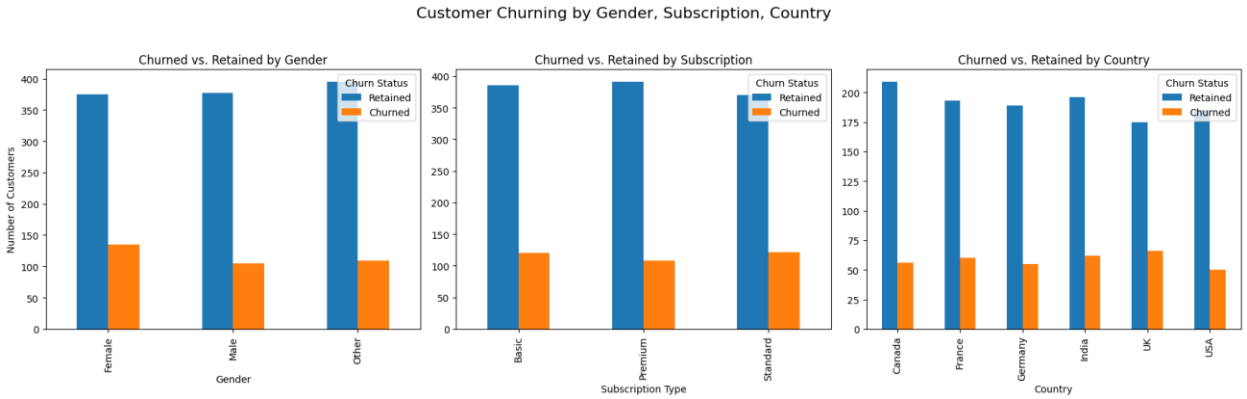
```
T-statistic: -0.194, p-value: 0.8460
No significant difference in watch time between churned and retained users
```

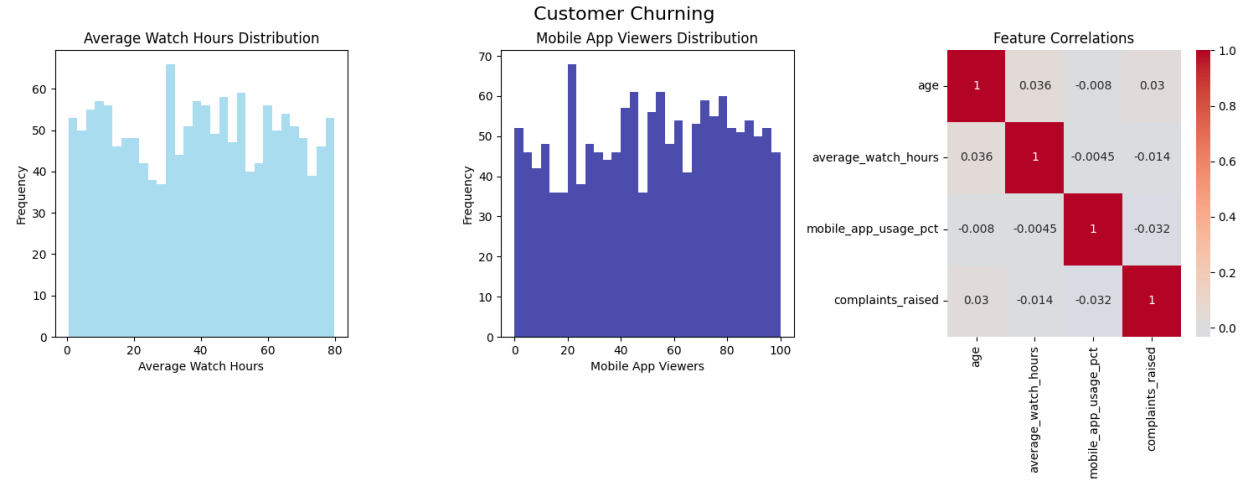
- Chi-square Tests: No statistically significant relationship between churn and receiving promotions, gender, or friend referrals (p-values > 0.05).

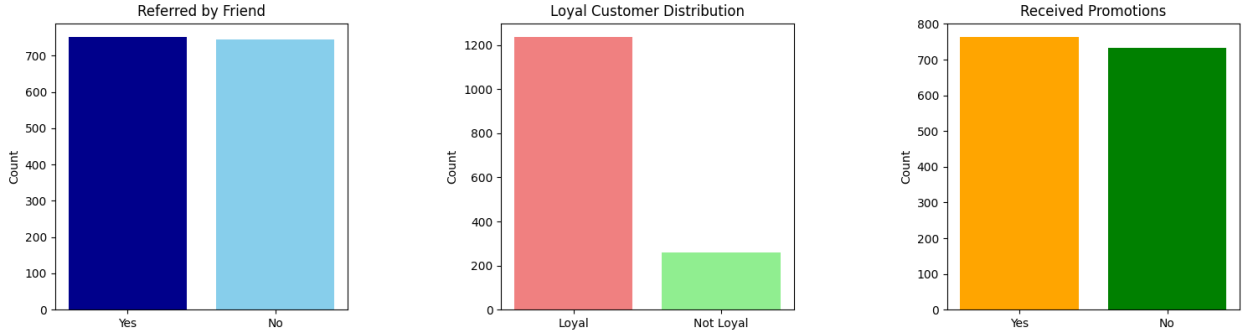
```
gender:
Chi2 = 4.273, p-value = 0.1180
No relationship with churn

received_promotions:
Chi2 = 2.633, p-value = 0.1047
No relationship with churn

referred_by_friend:
Chi2 = 0.719, p-value = 0.3966
No relationship with churn
```

- Correlations: Age, watch hours, mobile usage, and complaints showed near-zero Pearson correlations with churn, suggesting no strong linear effects.

Visualizations







Model Results

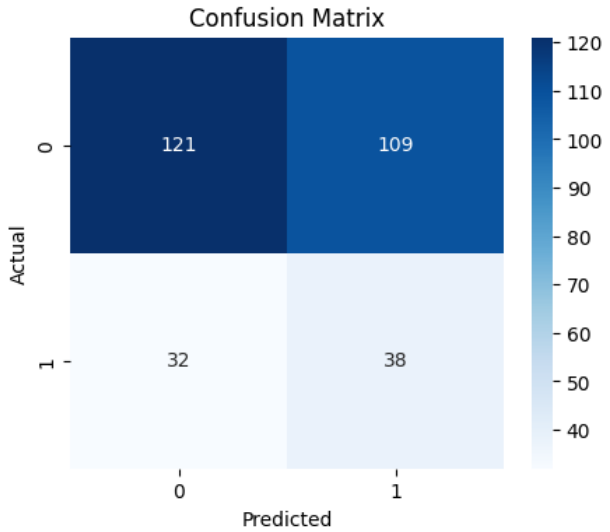
- Classification Report

	precision	recall	f1-score	support
Retained	0.79	0.53	0.63	230
Churned	0.26	0.54	0.35	70
accuracy			0.53	300
macro avg	0.52	0.53	0.49	300
weighted avg	0.67	0.53	0.57	300

- Top predictors for Churn

```
Top predictors of churn:
gender_Other                -0.184420
monthly_fee                 -0.132569
is_loyal                     0.106049
received_promotions         -0.102088
subscription_type_Premium    0.101420
gender_Male                 -0.088413
country_France               0.072362
country_India                0.071815
country_UK                   0.069946
subscription_type_Standard   0.066793
dtype: float64
```

- Confusion Matrix of the Model



Business Questions Answered

- Does receiving promotions reduce churn

  No significant effect found (Chi-square p > 0.05). Promotions alone aren't driving retention.

		count
received_promotions	**is_churned**	
No	**0.0**	572
	1.0	192
Yes	**0.0**	575
	1.0	157

- Are heavy mobile users less likely to churn

  No strong pattern. Both high and low mobile usage groups had similar churn rates.

		count
mobile_app_usage_pct	**is_churned**	
0	**0.0**	545
	1.0	161
1	**0.0**	602
	1.0	188

- Do watch hours correlate with churn

  No significant difference in average watch hours between churned and retained users. Engagement alone may not explain churn.

is_churned	average_watch_hours
0.0	39.988413
1.0	39.713190

- Do loyal customers (>180 days) churn less
  Not necessarily. The model even found a positive coefficient for is_loyal, suggesting tenure isn't a direct protective factor.
- Are certain subscription types or countries more prone to churn
  Yes. Some countries show higher churn rates.

country	is_churned	count
Canada	0.0	209
	1.0	56
France	0.0	193
	1.0	60
Germany	0.0	189
	1.0	55
India	0.0	196
	1.0	62
UK	0.0	175
	1.0	66
USA	0.0	185
	1.0	50

Data Issues or Risks

- Data Imbalance: Churned users are ~23% of the dataset, which limits model performance despite class weighting.
- Feature Limitations: Lack of granular engagement data (e.g., content preferences) may restrict the ability to capture true churn drivers.