**CDT 502**

**Principles of Data Analytics and Decision Making**

**Research on the User Churn Problem of Netflix**

**Group Name:** Teletubbies (Group 9)

**Group Members:**

|  |  |  |
| --- | --- | --- |
| Student ID | Student name | Contribution |
| 3223952 | CHEN Xin | Coding support |
| 3229839 | JIANG Yiyi | Report draft |
| 3224229 | MA Yutao | PPT draft |
| 3229724 | YAO Xinyu | Coding support |

25 April, 2025

**Abstract**

Netflix, the leading platform in streaming media, had faced the challenge of declining subscriber renewals back to 2021 and 2022 during the pandemic. The purpose of this report is to explore and discuss the reasons for this decline through an in-depth analysis with the data disclosed from Netflix annual report. The analysis reveals the diverse competitions in the market, including the gap between content quality and subscriber expectations, pricing strategy, and changing economic environment. Meanwhile, in order to accurately predict the risk of the churn, Markov Chain model, data decision-making and machine learning methods were applied during the process. With the subscriber behavioral data, viewing history and interaction, we also used logistic regression, Random Forest, and XGBoost for modelling and evaluation.

The results show that XGBoost has excellent performance in terms of AUC and recall rate in our research. Further analysis reveals that users with low activities at early stage, simple content needs or low interaction are more likely to churn.

In summary, this report provides a data foundation and strategy support for Netflix to identify and intervene potential churn users via data and behavioral analysis, together with machine learning. The relevant results help the platform to increase the LTV (Life Time Value) and upgrade the ability of content recommendation so that the users can be more interactive with the platform.

# **1 Introduction**

In the digital age, data has become a key asset that is changing the decision-making process in all fields. Data-driven decision refers to the use of collected big data from different sources to make decisions. This approach is becoming critical for enterprises, governments and organizations who are aiming to thrive in a competitive and dynamic environment. For business, data-driven decisions enhance the brand advantage. It enables companies to gain deep insights into customer behaviors, market trends and inefficiencies in daily operations. Companies are able to adjust their product design, services and marketing strategies more effectively to meet customer needs by analyzing the customers' purchase history, online browsing patterns and feedbacks. This not only improves customer satisfaction, but also boosts sales and profits.

As the global leader in the streaming media, Netflix has been playing the role of data-driven decision proponent from the start. The company uses data to recommend content for its subscribers, which has been a major factor for its success. However, Netflix had faced a challenge back to 2022 during the global pandemic - a decline in subscriber renewal, which directly impacts Netflix's revenue. As a subscription-based platform, the number of active subscribers is a key to its financial health. The decline means lower recurring revenue, which could limit the company ability to invest in content creation, technology development and global expansion. In addition, subscriber loss can also damage Netflix's brand image and effect the share price since word-of-mouth and reputation are crucial in the industry. If a large number of subscribers abandon the platform, this could be a signal for those potential subscribers that there are problems with the service, which might be harder for Netflix to attract new subscribers and maintain its market share. Netflix can only identify the root causes of subscriber dissatisfaction by analyzing the data in depth and take targeted measures to address them. In the following sections, we will analyse data from Netflix's annual report to reveal the reasons to the decline in subscriber renewals and use data-driven machine learning methods to predict subscriber churn risk.

**2 Exploring the Causes**

# **2.1 Analysis of User Churn Research**

According to the data disclosed in the annual report from Netflix in 2022, we can find that global user subscriptions declined in the first half year of 2022. Some users chose to stop their subscriptions after the first three-month subscription.

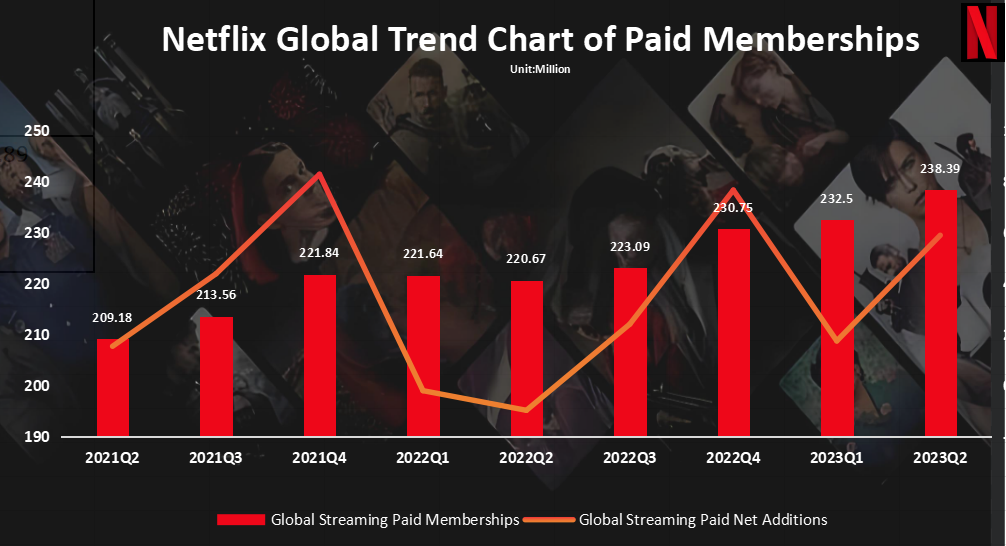


Figure 1: Netflix Global Trend Chart of Paid Memberships  
Source: Netflix   
By plotting the data of the Global Streaming Paid Memberships and the Global Streaming Paid Net Additions into a bar chart, it is obvious that the decline in subscribers is concentrated on Q1 and Q2 in 2022. Further reasons will be addressed for the decline based on the behaviors and user profiles from Netflix during this period.

**2.2 Reasons for Subscriber Churn**

**2.2.1 Subscription Price Increase**

It was reported that from 2018 to 2021, Netflix had raised the subscription fee for three times. In January 2022, the cost of basic, standard and premium accounts in the United States increased by $1, $1.5 and $2 respectively. To a certain extent, this has affected the subscription decision from users.

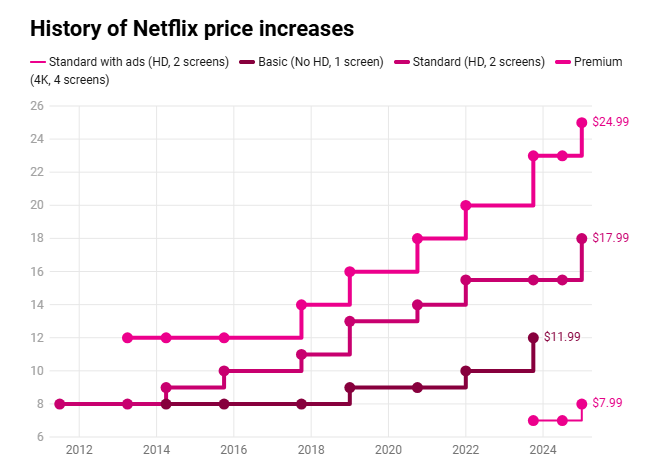


Figure 2: History of Netflix Subscription Fee

Source: Netflix & The Verge

**2.2.2 Market Competition**

With the fall of traditional media, the media industry has been rushing to transform. Discovery and Warner Media merge, online streaming launched by CNN, in addition to older competitors such as YouTube or Disney+, above of which cause some user lost for Netflix. The chart below shows the trend of US TV's share by platform from May 2021 to February 2022.

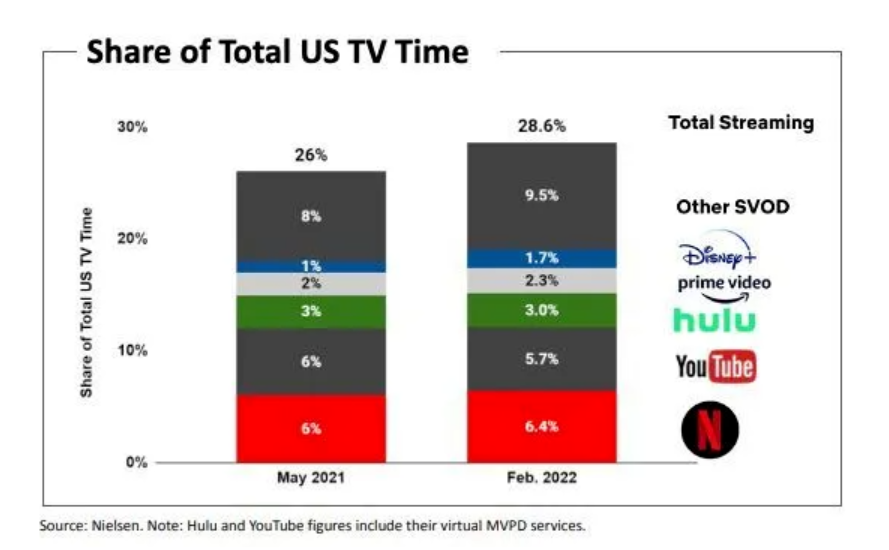
****

Figure 3: Share of Total US TV Time

Source: Nielsen Note: Hulu and YouTube figures include their virtual MVPD services

**2.2.3 Political Impact**

The fast gained subscription users from 2020 to 2021 caused by the global pandemic dropped significantly for the open-up in the United States in 2022. People tended to travel and experience the real life when they finally had the opportunities.

Due to the Russian-Ukrainian conflict on 24 April 2022, Netflix pulled out of the Russian market for the political reason, which lost 700,000 subscribers as a result.

**2.2.4 Large Base of Shared Accounts**Netflix is known to have more than 100 million households accessing its paid subscription content through account sharing. And Netflix single account can create up to 5 IDs, which means that perhaps more than 500 million users watch videos through account sharing. Just as we don't pay for a separate membership after borrowing a video membership account from a friend, account sharing in households makes it more difficult to add subscribers.

**2.2.5 Original Content**

In 2021, *The Squid Game* became a phenomenal breakout series on the Netflix platform globally. However, the following original series Netflix released struggled to reach the level of heat and quality of *The Squid Game* due to the lower quality. Most of the new dramas lack in depth and charisma, which makes it difficult for the audience to share emotional resonance. In terms of production level, the quality of graphics and special effects in some series is also poor. Three of Netflix's main series in the first quarter, *Bridgeton 2, Inventing Anna,* and *All of Us Are Dead*, were all below expectation. As a result, the number of Netflix subscribers renewing their subscriptions in the first and second quarters of 2022 has dropped significantly.

3 **User Churn Risk Analysis**

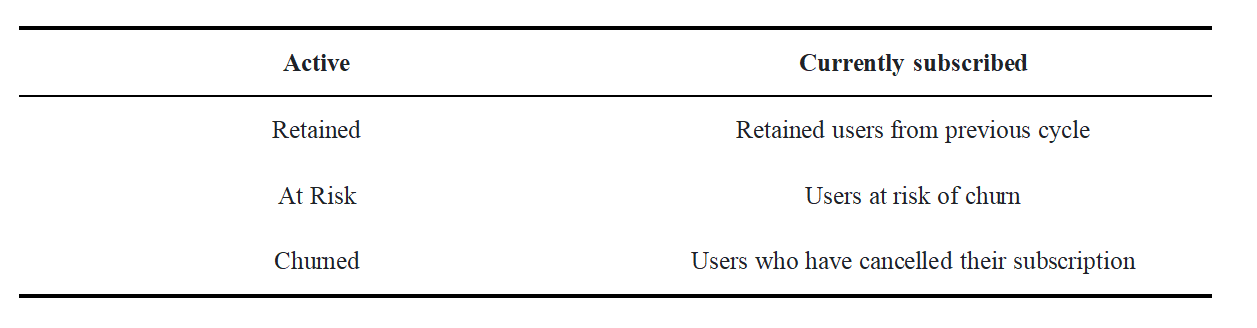
Based on the above analysis, we can find that the subscription renewals for Netflix had reduced in 2022, which needed to be highly valued for the company. In this chapter, we will discuss the potentially lost user profiles by means of the Markov Chain model. The model can evaluate the valuation of incremental account lifecycle as well as predict the risk. Relevant user data will be collected at the end to generate the churn risk through the logistic regression model.

**3.1 Markov Chain Models**

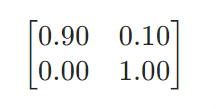
The Markov Chain Model is a model of stochastic processes with memoryless properties. It is based on "memorylessness", which means that the transfer of the current state depends only on the previous state and is independent of more previous states. It was first mentioned in 1906 by the Russian mathematician Andrey Markov in his Extension of the *Laws of Large Numbers for Dependent Random Variables with Bounded Covariance*. To prove that independence between random variables is not necessary for the weak law of large numbers and the central limit theorem to hold, Markov constructed a random process that is interdependent with conditional probability, known as a Markov Chain.

**3.1.1 Matrix construction**

We define the Netflix subscriber states, based on previous data assuming a subscriber sample size of 10,000 and an observation period of 24 months, into three states as “Retained”, “At Risk” and “Churned”.

Table 1: Netflix subscriber states

It is assumed that 90% of the users will continue to subscribe and 10% will be churned. Once churned, subscribers do not return and the Churned state is irreversible. Below is the matrix construction.



Also, based on the matrix, several assumptions are made as below.

|  |  |
| --- | --- |
| **Type** | **Explanation** |
| **Basic** | 85% Retention rate (lowest stickiness) |
| **Standard** | 90% Retention rate (industry benchmark) |
| **Premium** | 93% Retention rate |

Table 2: Netflix Retention Rate

#### ****3.1.2 Results And Graphs Analysis****

#### (1) ****Monthly Retention Rate****

#### As shown below among the 3 graphs, the monthly retention rate (left) illustrates the actual retention rate of Netflix subscribers versus the industry benchmark. The retention rate is slightly lower than the industry benchmark (90%) in the initial phase, but gradually climbs over time. Fluctuations may be related to the randomness of user behavior or the distribution of different packages.

#### (2) Comparison of Retention Rates for Different Packages

##### The graph in the middle shows the retention rates among three different packages. the Premium package has a consistently higher retention rate and less fluctuation than the Basic and the Standard, which indicates that the high-end subscribers are more adhesive to the subscription. While the Basic package has the lowest retention rate and more fluctuation, which indicates price-sensitive subscribers are less reliant on the service.

#### (3) User Activity Rate

#### The graph on the right shows the trend of user activity rate over time. Initially, all users are active (activity rate of 1.0). As time passes, the proportion of active users gradually decreases and eventually becomes stable, which indicates the cumulative effect of churn.

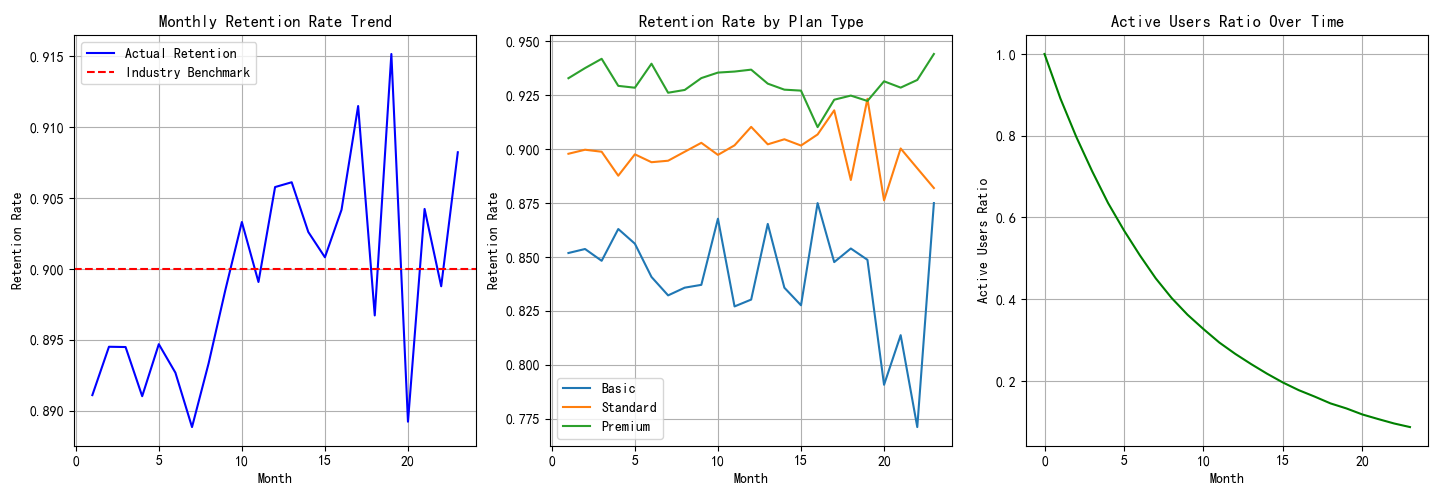
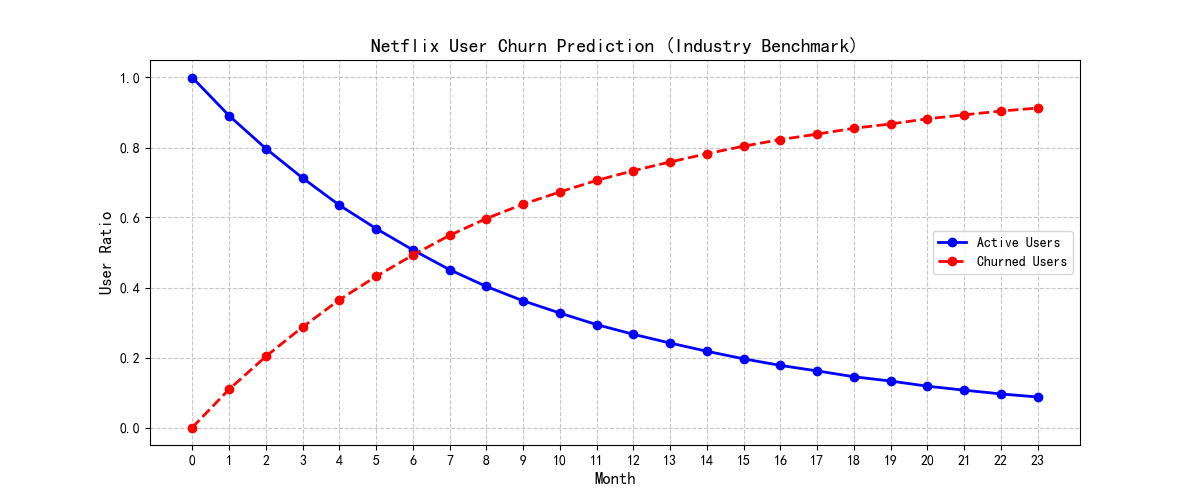


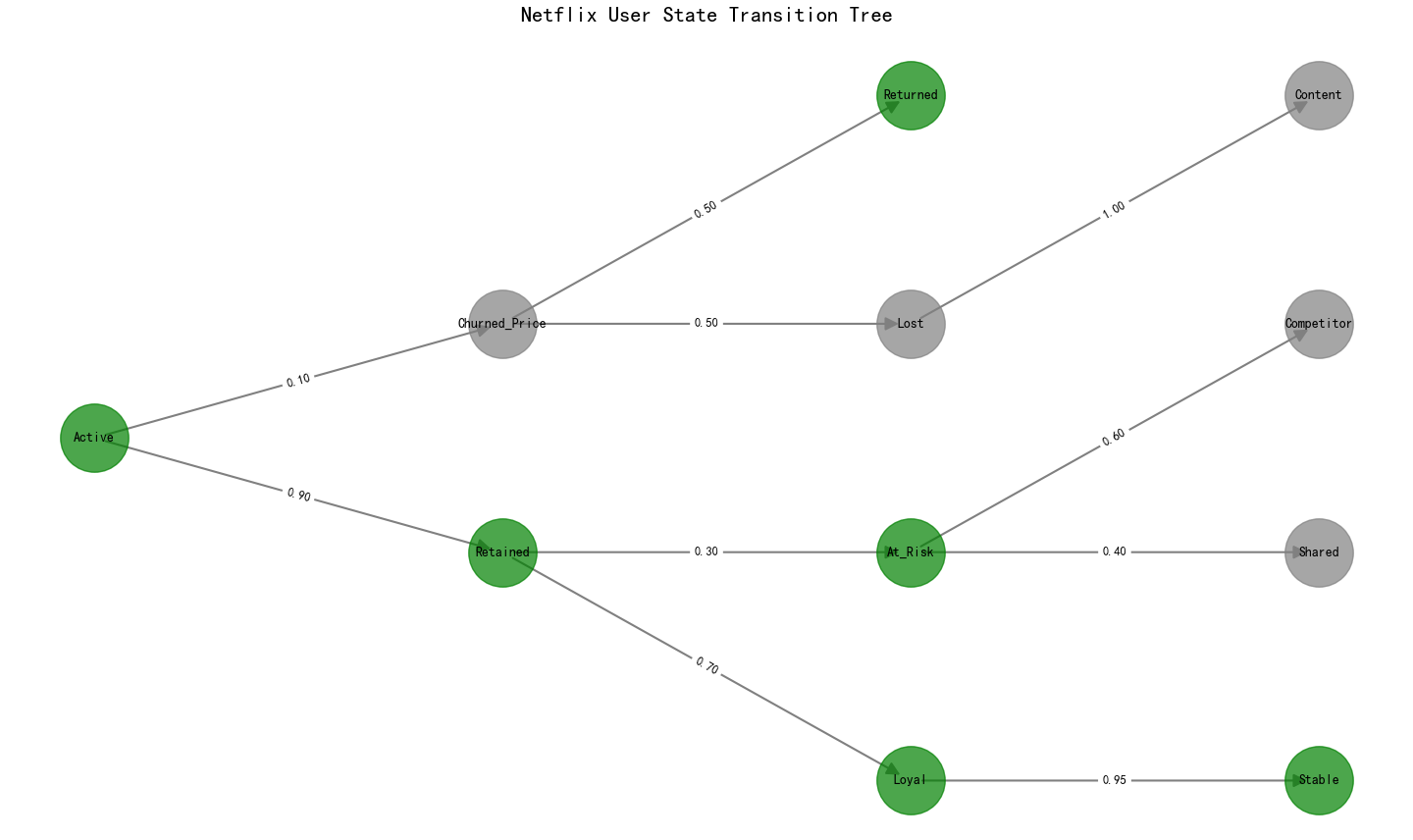
Figure 4-6: Monthly Retention Trend & Retention Rate by Plan Type & Active Users Ratio Over Time

#### ****(4) Forecast of User Churn****

Figure 7 below shows how the proportion of active and churned users changes over time. The blue curve (Active Users) is gradually decreasing over time. The red dotted line (Churned Users) is gradually increasing. Both curves turn stable at 24 months, indicating that the ratio between active and churned users is close to 1:9. This chart highlights the dynamics of the shift from active to churn over the Netflix user lifecycle, and provides a key reference for the company strategy making.Figure 7: Netflix User Churn Prediction

#### ****(5) State Transition Tree****

A Transition Tree Structure to represent the user states is shown to reflect the dynamics of the user among different states. **The first level stands for Active and shows** the initial state of the user, with a 90% probability of staying in the active state and 10% going to the churned state. The **Level 2 refers to the Retained / Churned Price. That means that** active users may be further segmented into retained users or churned due to price sensitivity. **The Level 3 is about the Loyal / At Risk / Lost / Returned.** Retained users may be converted to Loyal (70%) or Churned Risk (30%). Churned users may be still attracted by the platform (50%) or permanently churned. **The last level includes Stable / Shared / Competitor / Content. So** loyal users end up in a stable state (95%), while At-risk users may drift to shared accounts, competitors or return temporarily due to content attraction. The diagram visualizes the potential path and likelihood of users moving from active to churn in a hierarchical way, which may provide a multi-dimensional perspective for Netflix to optimize users’ interactions.

Figure 8: Netflix User State Transition Tree

**3.1.3 Conclusion**

From the diagrams and analysis above, the percentage of Netflix active subscribers declines and the churn rises, while eventually tend to stable. The industry baseline - 90% retention, provides a benchmark that Netflix has a 10% churn rate based on the subscriber performance in Q1 and Q2 in 2022.

**3.2 Logistic Regression Models**

**3.2.1 Data Sources And Modelling Logic**

As Netflix did not publish any subscriber-behavior-related data, this study builds a dataset to demonstrate the behaviors for the typical subscribers in the early stages of subscription, according to the public news and industry reports (de Zilwa, 2023). The simulated data contains a total of 10,000 subscriber samples with different variables as below table.

|  |  |  |
| --- | --- | --- |
| **Variable Name** | **Type** | **Meaning Explanation** |
| watch\_hours\_weekly | Numerical type | Average number of hours users watch per week |
| plan\_type | Categorical type | Subscription package type: Basic / Standard / Premium |
| use\_mobile\_only | Boolean type | Whether to watch only on mobile devices (True/False) |
| shared\_account | Boolean type | Whether it is a shared account (simulating multi-device login) |
| watched\_netflix\_original | Boolean type | Whether to watch Netflix original content recently |
| paused\_account | Numerical type | Number of times the account has been suspended within the last 90 days |
| churned | Label variable | Whether to cancel the subscription within 90 days (0 = not churned, 1 = churned) |

Table 3: Logistic Regression Models Variables

These variables are designed under the issues in where Netflix has exposed to the public for its churn in 2022, such as less appeal of original content, account sharing controls or increased mobile usage.

**3.2.2 Analysis of Methodology And Modelling Process**

In this study, logistic regression model was used for classifying prediction, with Python programming language. Data reading and processing are through pandas. Visualization to show the different variable relationships are through seaborn and matplotlib. Model training and evaluation are through scikit-learn. Jupyter Notebook was used to write and present the analysis process.

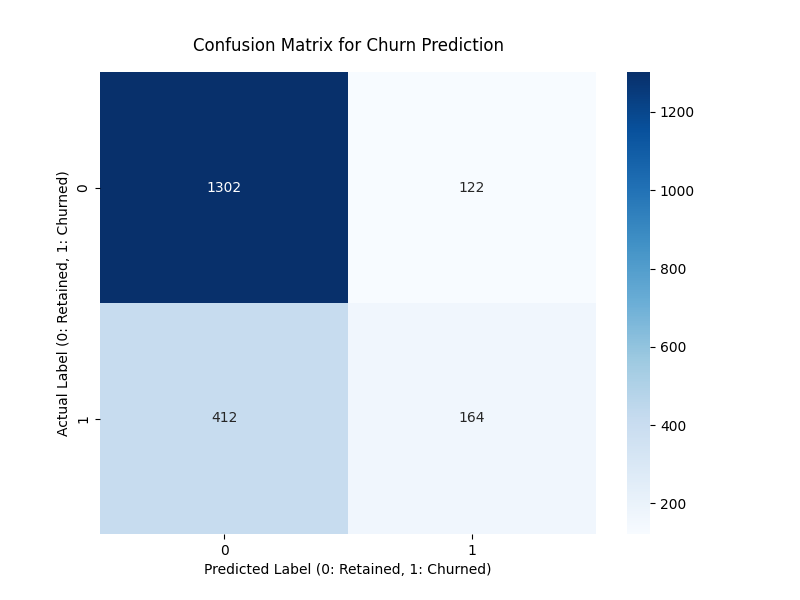
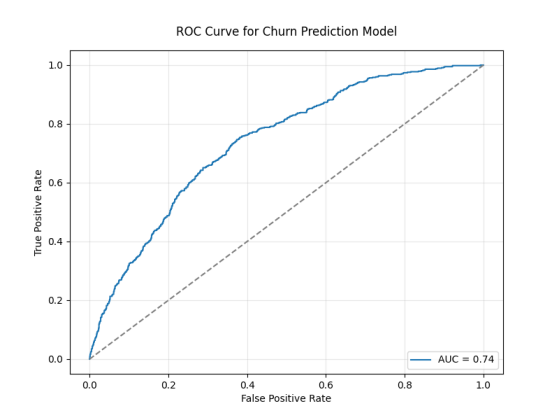


Chart 9-10: ROC Curve & Confusion Matrix

**3.2.3 Regression Model Results and Analysis**

This study uses a logistic regression model to classify and predict whether a user is going to churn or not. We divided the data into training and test sets in an 8:2 ratio and trained the model using default parameters. The following is the performance of the model on the test set.

|  |  |
| --- | --- |
| **Indicator** | **Value** |
| Accuracy | 0.73 |
| AUC | 0.74 |

Table 4: Performance of Model

“Accuracy” indicates the level of the overall prediction. “Recall” indicates the ability to identify churned users; “**AUC”** describes the ability of the model to distinguish whether churned or not. The AUC value 0.74 belongs to a high level, which shows that the model has a good discriminative performance.

To understand the impact on the probability of churn from each variable, we output the standardized coefficients from the logistic regression model. Below are some of the variables. Positive values mean that they prompt churn, and vice versa.

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Coefficient (Standardized)** | | **Explanation** |
| watch\_hours\_weekly | 0.12 | The more hours a user watches, the lower the likelihood of churn. | |
| plan\_type\_Standard | 1.00 | Compared with users of the Basic package, those of the Standard package are more stable. | |
| plan\_type\_Premium | 0.92 | Users of the Premium package have a slightly lower churn rate. | |
| use\_mobile\_only | 0.98 | Users who only watch on mobile devices are more likely to churn. | |
| shared\_account | 0.26 | Shared accounts slightly increase the risk of churn. | |
| watched\_netflix\_original | 0.89 | Watching original content significantly reduces the probability of churn. | |
| paused\_account | 0.44 | Users who have suspended their accounts are more likely to churn. | |

Table 5: Variables and Corresponding Direction

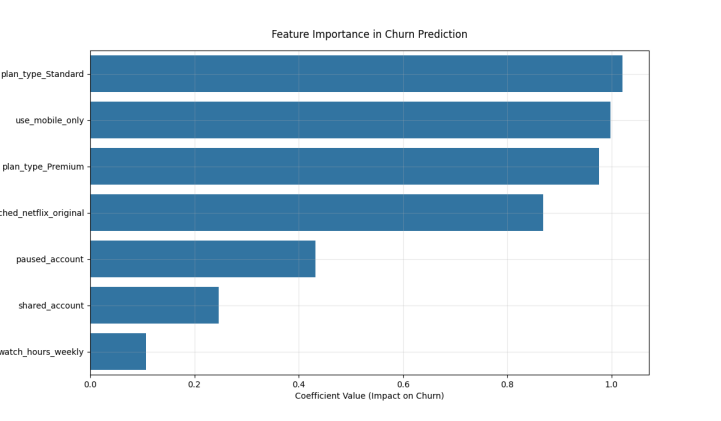


Figure 11: Feature Importance in Prediction

Netflix produced content quality, viewing duration and the subscription package play key roles in user retention according to the diagrams.

**3.3 Visualized Analysis Review**

We further validate the consistency of the model inferences with the visualized trends.

The box plot shows the average viewing hours of the churned users is much lower.

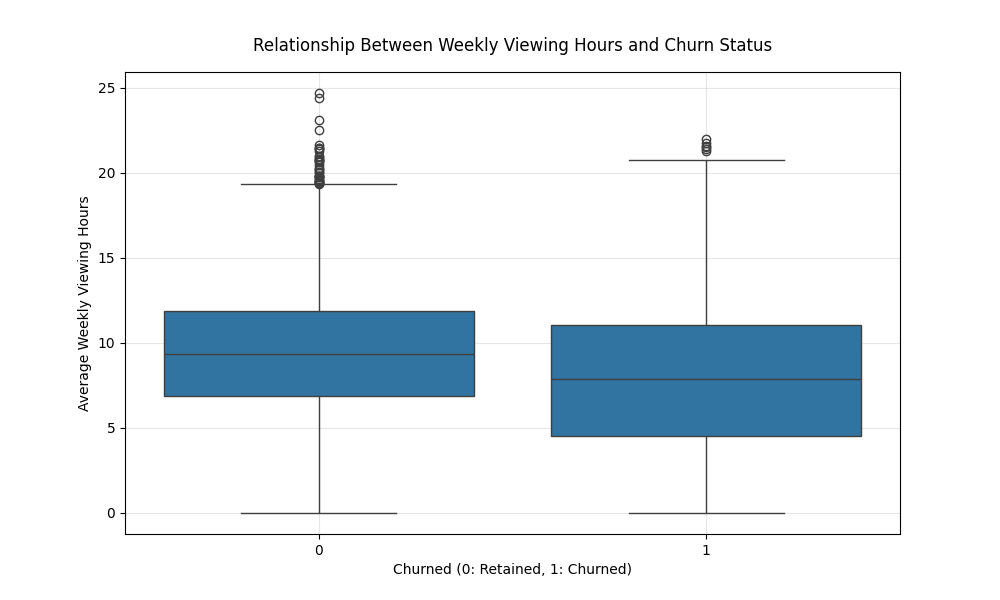
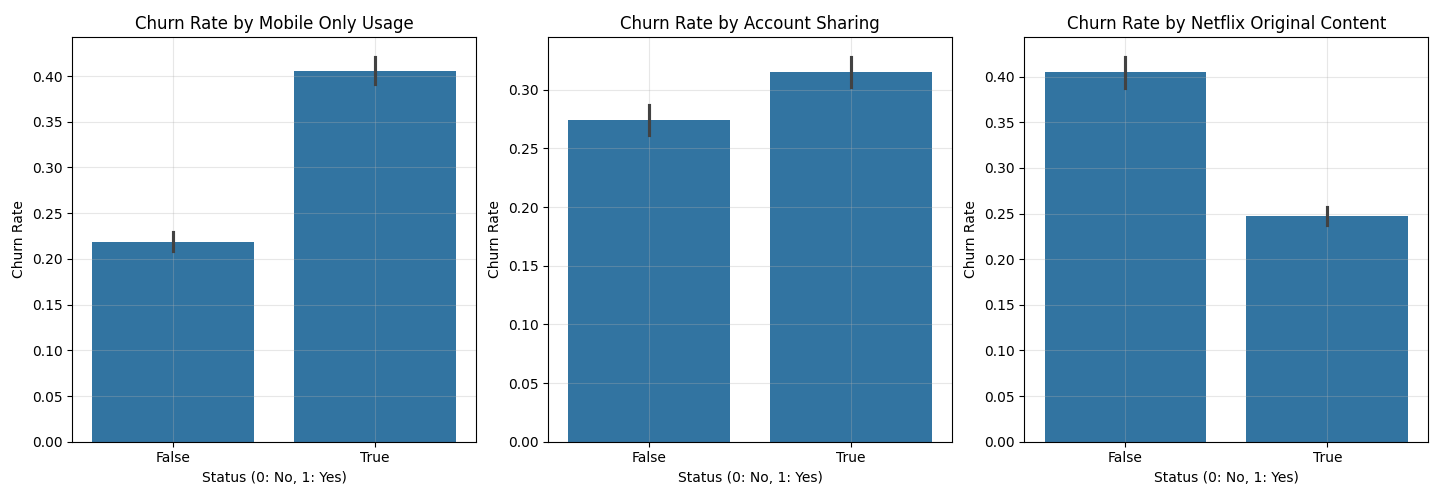


Figure 12: Relationship between Weekly Viewing Hours and Churn Status

The bar charts indicate that the churn rate for subscribers who failed watching Netflix original content was about 15% higher.



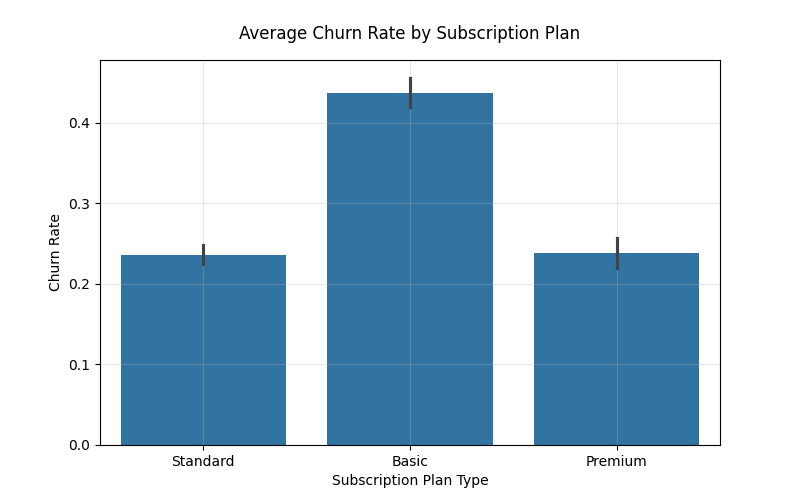


Figure 13-16: 3 Bar Charts of User Status & Average Churn Rate by Subscription Plan

The correlation heatmap shows that churn is negatively correlated with viewing hours and original content, but positively with suspended accounts and mobile usage.

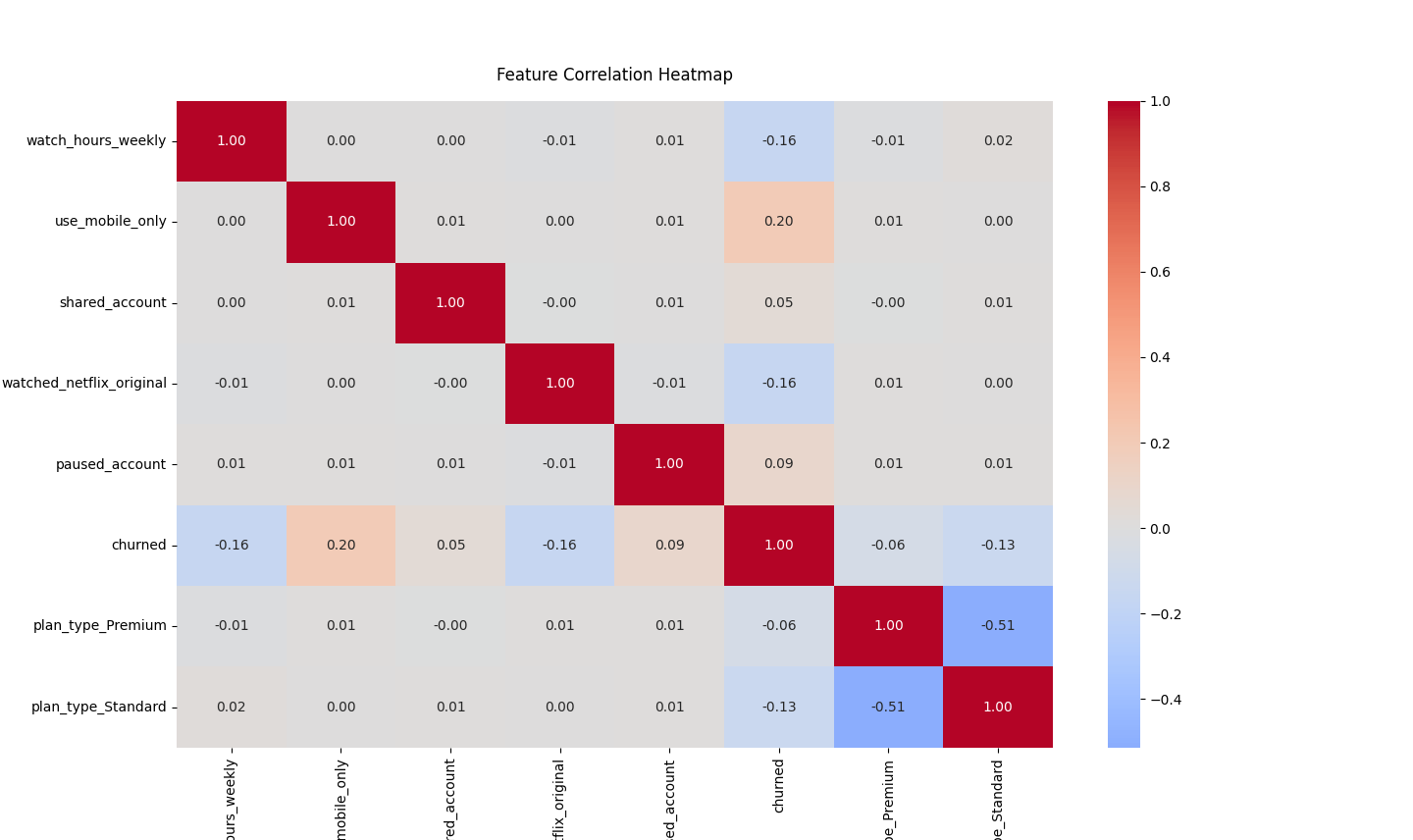


Figure 17: The Correlation Heatmap

These graphical evidences are highly consistent with the model results and enhance the credibility of the research.

**3.4 Final Result Analysis**

Based on the results mentioned above, 3 major factors are concluded from this study – user viewing behaviors, Netflix original content interaction and AUC value.

User viewing behavior shows that the shorter the weekly viewing time is or the less attractive the content is, the users are more likely to cancel their subscription.

Frequency of interaction with Netflix original content is a key variable in predicting subscriber retention. The users who watch Netflix original series frequently shows much lower churn rate, which can also be an indication that the Netflix exclusive series are still one of the most competitive advantages for the platform in the market.

With an AUC value of 0.74, the logistic regression model indicates that even with basic machine learning methods, high-risk users can be accurately identified with a reasonable data structure.

Based on the analysis, we can summarize the following typical churn user profiles as people who watch less than 5 hours per week, people who watch only on mobile devices, people who have never watched any original content, people who subscribe Basic Package and people who have suspended their account once or more times.

**4 Summary**

With the open data source from Netflix's annual report and Markov Chain modelling and logistic regression, the major reasons for Netflix subscription renewal decline in 2022 are the subscription pricing increase and original content quality. At the same time, we conclude that Basic level subscribers spend less time per week and the users who are dissatisfied with the recent original content are the vulnerable groups to churn. Netflix wants to change the current situation by targeting these groups accordingly to give suggestions to reduce churn. Netflix can try to work on these groups of people and meet their needs first to twist the declining trend.  
The data analysis and model prediction in this paper can make us better understand what we have learnt from the course, and feel the charm of data analysis and decision-making better.

**Reference**

Andre. Markov. (1906). An Extension of the Law of Large Numbers with Respect to Dependent Variables.

Catie Keck.(2022).Yes, Netflix just got even more expensive.The Verge.

<https://www.theverge.com/2022/3/24/22993562/netflix-price-increase-us-plans-2022>

Netflix.(2022).Final-Q1-22-Shareholder-Letter.

<https://ir.netflix.net/financials/quarterly-earnings/default.aspx>

Netflix.(2022).Final-Q2-22-Shareholder-Letter.

<https://ir.netflix.net/financials/quarterly-earnings/default.aspx>

Netflix.(2022).Final-Q3-22-Shareholder-Letter.

<https://ir.netflix.net/financials/quarterly-earnings/default.aspx>

Netflix.(2022).Final-Q4-22-Shareholder-Letter.

<https://ir.netflix.net/financials/quarterly-earnings/default.aspx>

Netflix.(2023).Final-Q1-23-Shareholder-Letter.

<https://ir.netflix.net/financials/quarterly-earnings/default.aspx>

Nielsen Note.(2022).*Hulu and YouTube figures include their virtual MVPD services*

Weiqi Cheng, Ximin Huang, Xi Chen, et al. (2015). *Markov Chains. Tsinghua University Press*.