## Introduction

Customer churn, or the percentage of customers who discontinue a service within a given time frame, is a critical business metric for subscription-based services such as Netflix. This study aims to identify the key drivers of customer churn and uncover patterns that can help in reducing churn rates. Through a comprehensive exploration of the Netflix customer churn dataset, various statistical techniques and exploratory data analysis (EDA) methods were employed to gain insights into customer behaviour.

## Objectives

The primary objectives of this study are as follows:

- To analyse customer churn behaviour using statistical methods.
- To identify significant variables contributing to churn.
- To develop a basic predictive model to estimate the likelihood of churn.
- To suggest actionable strategies for retention based on findings.

## Data Understanding

### Dataset Description

The dataset used for this analysis consists of customer data from Netflix, including variables such as:

- customer_id: Unique identifier for each customer.
- gender: Gender of the customer.
- age: Age in years.
- monthly_fee: Subscription fee paid by the customer.
- watch_time: Average monthly watch time (in hours).
- plan_type: Type of subscription plan.
- signup_date: Date of registration.
- last_active_date: Most recent date of activity.
- churned: Binary variable indicating whether the customer has churned (1 = Yes, 0 = No).

The data spans over multiple months and includes both numerical and categorical attributes relevant for churn analysis.

### Data Preprocessing

Prior to analysis, the dataset was cleaned and processed using the following steps:

- Categorical variables were encoded using label encoding or one-hot encoding where required.
- Outliers were detected using boxplots and treated accordingly.
- Derived variables such as customer tenure were created to enhance feature relevance.

## Exploratory Data Analysis (EDA)

The EDA phase focused on understanding the distribution of data and relationships between variables. Key observations include:

- **Churn Rate**: Approximately 50.3% of customers were identified as churned.
- **Gender Distribution**: Balanced representation across genders.
- **Plan Type vs Churn**: Basic plans had a higher churn rate compared to premium plans.
- **Watch Time and Churn**: Customers with lower average watch times were more likely to churn.
- **Monthly Fee**: Churned customers generally subscribed to lower-priced plans.
- Correlation analysis showed significant relationships between watch_time, monthly_fee, and churn.

Various visualization techniques such as histograms, boxplots, and bar graphs were used to support these findings.

## Statistical Testing

The following statistical tests were applied to validate assumptions:

- **Independent t-test**: Applied to determine if average monthly_fee and watch_time differ significantly between churned and non-churned customers.
- **Chi-square test**: Used to test the independence between categorical variables such as plan_type and churned.
- **ANOVA**: Applied to compare monthly_fee across multiple churn categories or plan types.
- **Correlation Analysis:** To check how numeric variables (e.g., monthly_fee, watch_hours) relate to each other.

- **Logistic Regression**: A binary classification model was built to identify variables that significantly predict churn. Key predictors included monthly_fee, plan_type, and watch_time.

**Findings and Interpretation**

- Customers on lower-tier plans and with less engagement (low watch time) are more prone to churn.
- Plan type and usage patterns can be used to segment at-risk customers.
- Pricing strategies and personalized engagement can play a key role in retention.

## Conclusion

This study provided valuable insights into the factors contributing to Netflix customer churn. By leveraging statistical analysis and machine learning, patterns and trends were uncovered that can support data-driven decision-making. Future work can incorporate time-series analysis and real-time churn prediction for ongoing business optimization.