## Introduction

The movie recommendations have undergone a great transformation with the improvement of machine learning techniques. In this report, we embark on a comprehensive exploration of user preferences in movies, employing data-driven methodologies to find patterns and hidden insights. Our primary objective is to develop a predictive model that understands user's wishes toward movies based on demographic attributes and genre preferences.

## Data analysis

I started by looking through three datasets: one on users, another on movies, and the third on ratings. These datasets hold the key to understanding what people like and dislike. Then I looked through each dataset and found out the following information: Firstly, there is no missing data in columns that are important for future evaluation. Secondly, movies dataset has a lot of different genres that are mixed, so we can use them in prediction. Thirdly, ratings are distributed pretty well, so predictions will be good.

## Model Implementation

In this phase, I used a logistic regression model to predict user preferences for movies, leveraging demographic information and movie info. This model aimed to discern whether a user would like a movie or not.

## Model advantages and disadvantages

Advantages:

- Interpretability: Logistic regression provides transparent insights into how different features influence the likelihood of liking a movie.
- Efficiency: It's computationally efficient, making it suitable for large datasets.

Disadvantages:

- Linearity Assumption: Logistic regression assumes a linear relationship between features, which might not capture complex interactions

## Training process

Model was trained using a logistic regression algorithm with a maximum of 10,000 iterations. I split the dataset into training and testing sets and decided not to do the validation, because the dataset is pretty simple, so the model should give good results even without it.

## Evaluation

To analyze model's performance, I evaluated it using various metrics:

- Accuracy: The proportion of correctly classified instances.
- F1 Score: A balance between precision and recall, crucial for imbalanced datasets.
- Precision: The accuracy of positive predictions.
- Recall: The proportion of actual positives correctly predicted.

## Results

I achieved the following results:

- Accuracy: 1.0
- F1: 1.0
- Precision: 1.0
- Recall: 1.0
- Benchmark evaluation: 0.999

These results show that the model did a great, almost perfect job, predicting exactly the movies that the user will like and discarding the movies that the user will dislike.