# Problem Statement

Analyze data on movie reviews in order to provide movie recommendations

# Client

Movie recommendation engines can be used in a number of different ways:

1. Streaming sites like Netflix of Flixter can use the reviews given by their users to suggest other movies that they may like.
2. Television networks can boost viewership by advertising recommended shows during commercial breaks of shows with an audience likely to enjoy them.
3. Cinema theatres can provide recommendations based on your past bookings when you log on to their website.
4. Websites like IMDB and Rotten Tomatoes can provide recommendations based on your past reviews.

By analyzing past reviews to recommend movies or TV shows, advertising can be more targeted based on the specific tastes of a user/audience. Viewership/customer satisfaction and engagement may increase.

# Data

For the purpose of this analysis, I will be using the data collected and made available by the GroupLens Research group. The anonymized data has been collected from the MovieLens website and was created by users of the website between 1995 and 2016. It consists of 4 different files:

* links.csv
* movies.csv
* ratings.csv
* tags.csv

# Approach

Most of the data I will be using for my analysis are present in movies.csv and ratings.csv. The movies file contains details of the movie i.e. movie ID, title and genre. The ratings file contains the user ID, movie ID, rating and timestamp. Ratings are made on a 5 star scale with half-star increments. Each user has reviewed at least 20 movies and each movie has at least one rating or tag.

I will first create a smaller suitable dataset using the data of about 100k movie ratings. By analyzing the ratings previously created by the user as well as the genres of the movies that are popular with the user, new movie recommendations may be generated.

# Deliverables

Following are my expected deliverables at the end of the project

* Code
* Documentation including visualizations and analysis
* Presentation slides summarizing my analysis