

# Report - Movies Recommendation in Flixster

Pedro Unas e Solange Sampaio

08/05/2020

## Introdução

O objetivo deste trabalho prático consiste em comparar diferentes métodos de recomendações para um dado conjunto de dados, de modo a fazer recomendações a um conjunto de utilizadores. Foi-nos disponibilizado os 3 seguintes ficheiros de texto pertencentes a um site de filmes, chamado Flixster:

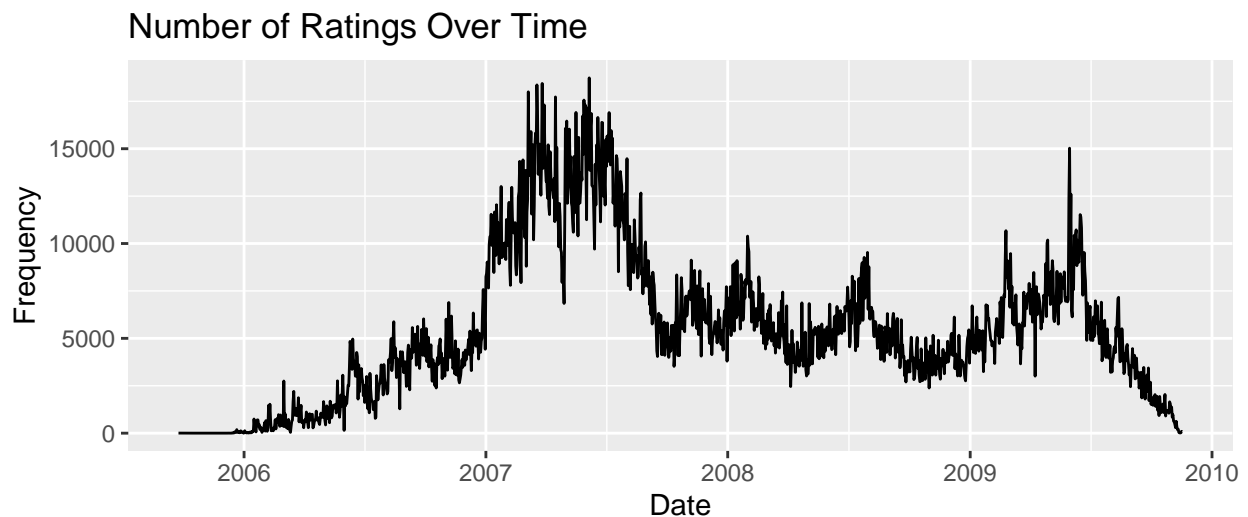
- **Ratings.timed.txt:** Contêm os ratings que cada utilizador deu aos filmes numa respetiva data.
- **movie.txt:** Contêm o nome dos filmes e o respetivo id.
- **profile.txt:** Contêm a informação de cada utilizador (id, género, idade, localização e o tempo que é membro)

## Pré-processamento dos dados

Ao analisar os ficheiros, reparámos que não estavam bem formatados, por isso começamos por fazer um script em python para os tratar. Para o ficheiro movie.txt, criámos duas colunas, moviename e movieid, contendo os nomes dos filmes e os id's e guardámos num ficheiro csv. O ficheiro Ratings.timed.txt também foi convertido para um ficheiro csv, contendo 4 colunas (userid, movieid, rating, date). Por fim, no ficheiro profile.txt eliminámos os valores em falta utilizando a função dropna() e convertimos este ficheiro para um ficheiro csv com 7 colunas (userid, gender, memberfor, lastlogin, profileview, age).

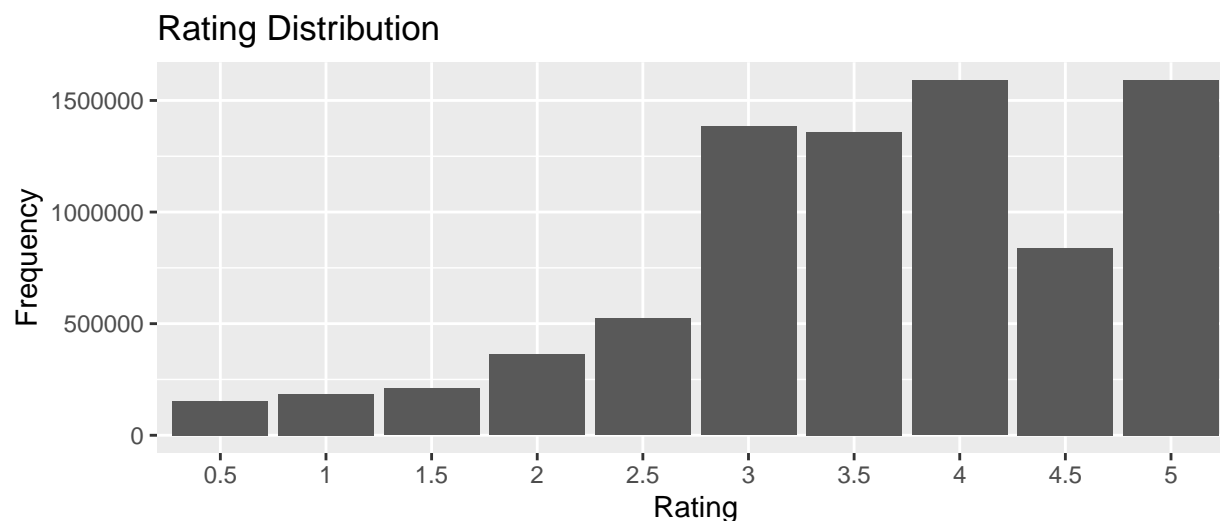
## Exploração dos dados

### Número de Ratings ao longo do tempo



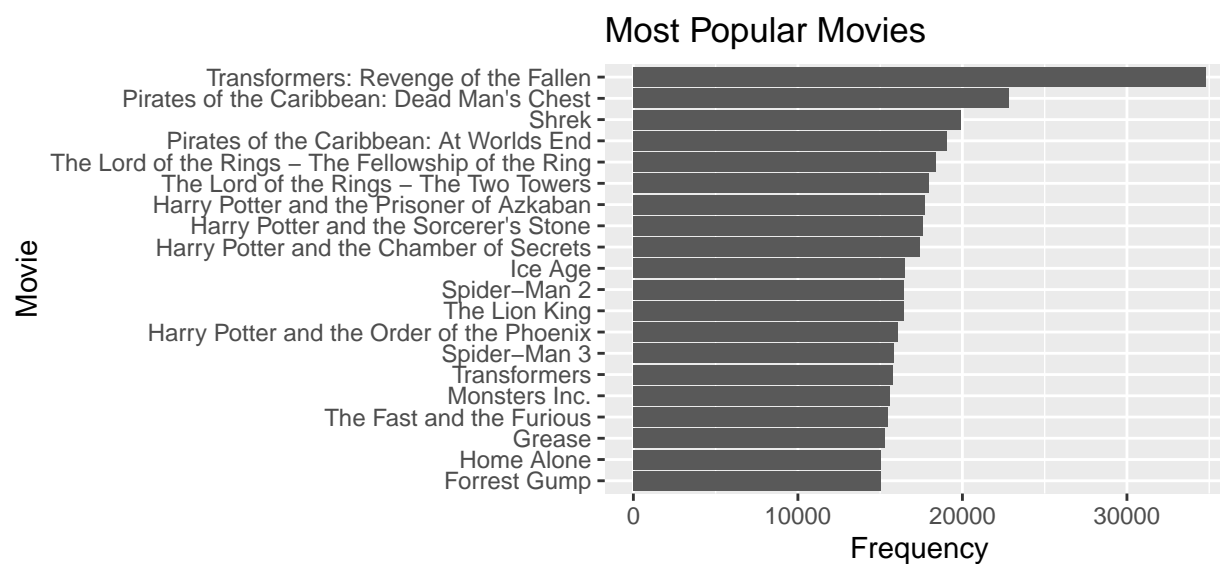
Ao analisar o gráfico é possível ver que houve um pico entre abril e maio do ano de 2007, o que significa que havia uma grande quantidade de utilizadores ativos, havendo um decréscimo nos anos seguintes com algumas subidas não muito significativas. No ano 2009 houve uma subida mais acentuada no mês de maio e a partir daí houve um decréscimo muito acentuado.

### Distribuição do rating



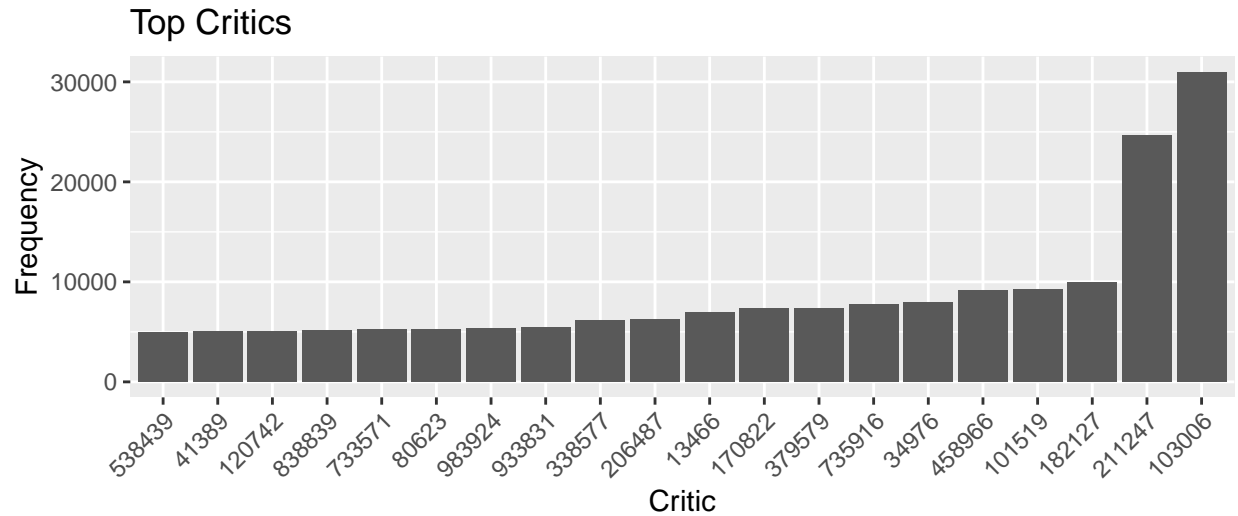
É possível ver que os ratings atribuídos pelos utilizadores aos filmes são na sua maioria positivos, ou seja, maior que 2.5, o que podemos concluir que os utilizadores têm uma maior tendência para avaliar os filmes que mais gostaram.

### Filmes mais populares



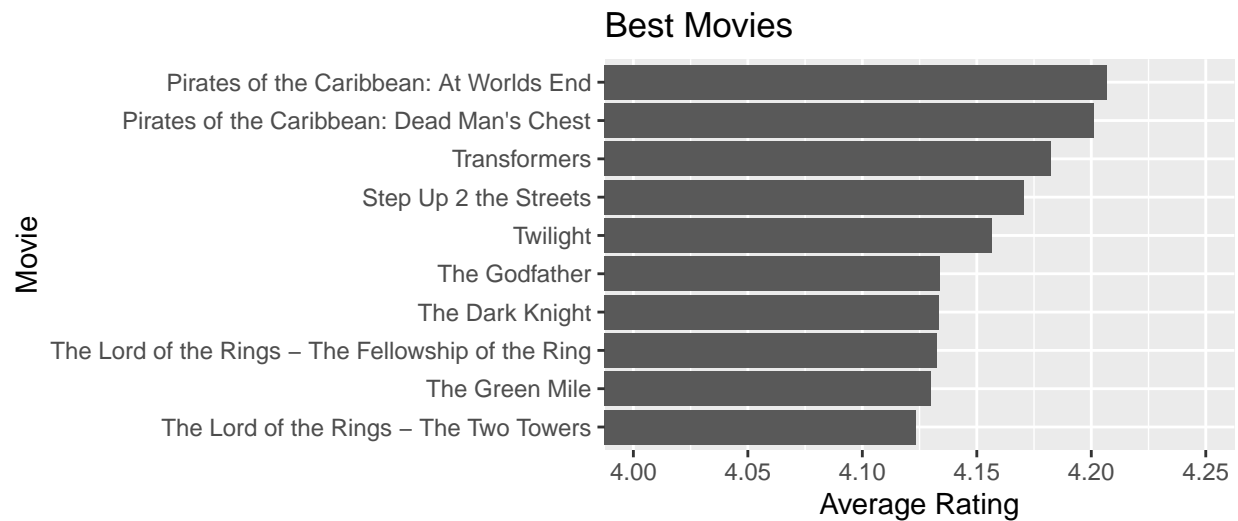
Como podemos observar o filme com mais avaliações é o *Transformers: Revenge of the Fallen* com cerca de 35,000 avaliações. Já o segundo filme com mais popular, possui cerca de 22,500 avaliações, havendo assim uma grande diferença comparativamente com o primeiro.

## Top críticos



Neste gráfico é apresentado os 20 utilizadores que avaliaram o maior número de filmes, sendo o utilizador com o id 103006 com mais de 30000 avaliações atribuídas aos filmes.

## Top filmes



Os 10 filmes com uma classificação maior estão apresentados neste gráfico, sendo *Pirates of the Caribbean-. At Worlds End* o filme com melhor classificação.

## Sistemas de recomendação

Começamos por fazer uma amostra dos dados, que contém os id's dos utilizadores, os nomes dos filmes e o rating que deram a um determinado filme. Essa amostra contém 3745 observações.

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.