

# Final report

\*Final report

Jirong Yi  
ECE Department  
Stevens Institute of Technology  
Hangzhou, China  
10468494

Bingcong Zhai  
ECE Department  
Stevens Institute of Technology  
Beijing, China  
10467423

## I. Introduction

We often watch movies in our daily life. Movies are closely related to our life. However, there are thousands kinds of films, and the scores of movies are also different. The reason is related to the quality of the film, and also closely related to the people who have different age, gender, personality and growth background, so they enjoy different genres film. Based on the above reasons, we do a data analysis deeply explore how do different ages, genders impact movie's score. Based on the above analysis, we have also made a simple movie information management system. In this system, you can as a user to create an account, which includes your name, your age, the movie you want to evaluate and the score you want give to the movie. You can also view this account, change the information in this account, and delete this account.

## II. Movie data set analysis

In this project, we use jupyter notebook as a compiler for analysis. We got three data sets from the Internet. One data set is the information of users, including user ID, age and gender. The second data set is the information of movies, including movie ratings. The third data set is also the information of movies, including the movie names and type of the movie and so on. We first merge the three data sets into one data set, and then watch a basic analysis of the data, such as its count, mean, standard deviation and so on. As shown in Figure 1

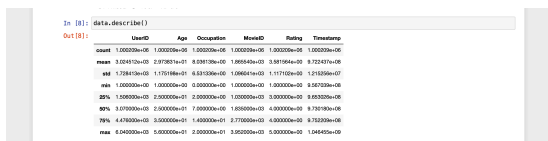


Fig. 1.

Secondly, we choose gender as a key, explore the influence of different genders on film scoring. We find out the difference between the scores of different genders for movies, and then sort out the top ten movies with the biggest differences between men and women, and what kind of movies they belong to, as shown in Figure 2

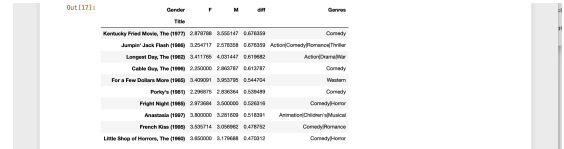


Fig. 2.

Finally, we find out the frequency of each type of movie and use the histogram to reflect it find out the age distribution and use the histogram to reflect it. We also rank the movie scores and get the names of the top 20 movies.

## III. Movie information management system

In this system, we first divide the project into four modules. The first module is the interface view, which is responsible for processing interface logic, such as input and output. The second module is the core processing module, which is mainly used to write add, delete, modify and query code. The third module is the film information model, which is used to encapsulate the information of the first two modules. The fourth module is the main module, use to run this program in this module. For the final demonstration, please see video. In this system, you can create an account as a user, which includes your name, your age, the movie you want to evaluate and your rating of the movie. You can also view this account, change the information in this account, and delete this account.

## IV. Future and Prospect

This data analysis and movie information system is a relatively basic programming, fully show our learning results. In our follow-up study, we will further explore the reasons for different movie scores, increase the amount of data, and increase the direction of research. For the movie information management system, we will increase the function of the film information system and so on. In the future, we will keep moving and we hope to benefit mankind through our own little power