
# School of Advanced Technology

# Project 1 Report

Project Title: Assignment 1: Web Scraping & Data Analysis

Student Name: Shen Wang

Student ID: 1824260

Project field: INT303

Supervisor: Jia Wang

Co-supervisor (if applicable):

# Introduction

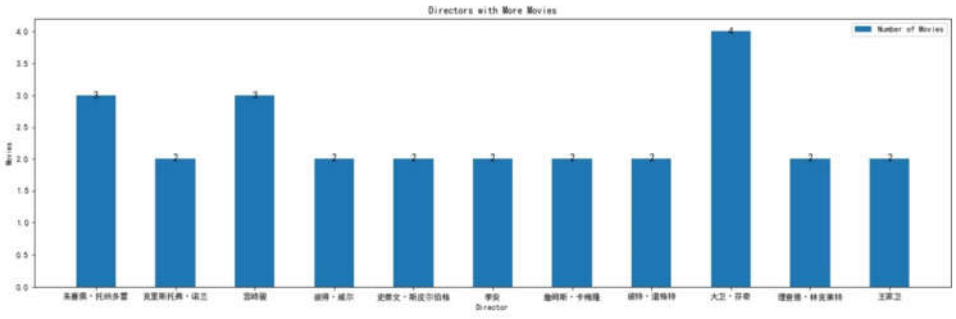
The report will complete data analysis according to the data which is scraped from the Maoyan Movies website about the top 100 movies. The ranking list from Maoyan Movies is ordered by the feedback and the number of people giving feedback. In general, the ranking is published with a complex comparison.

The data set provides information about the ranking, movie title, director, actors, rating, movie box office revenue, published country, movie duration, movie type and published date. The data analysis will try to find the relationship among some of these features.

# Visualization and Data Analysis

### 1. Some directors have more than one film in the top 100

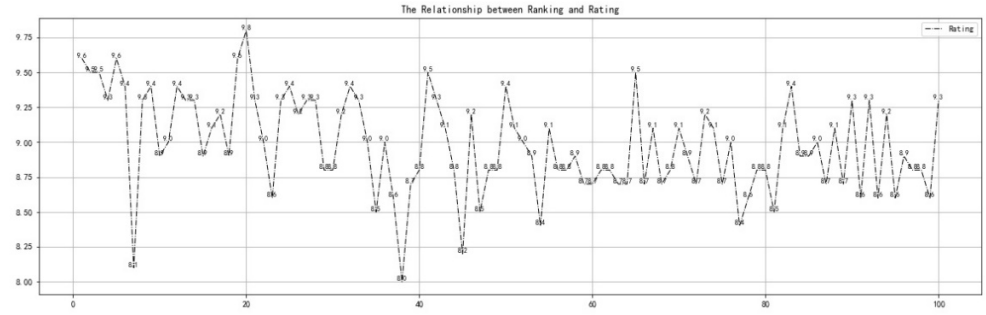
It is amazing that some directors have more than one work in the top 100 ranking list. As the following diagram, David Fincher even has 4 films in the list. A little less than him, Giuseppe Tornatore and Miyazaki Hayao have 3 most popular films. Moreover, there are other 8 directors have 2 movies in the top 100.




In general, this shows that these directors are good at making high quality films that are widely loved and have a high artistic aesthetic.

## 2. The relation between ranking and rating

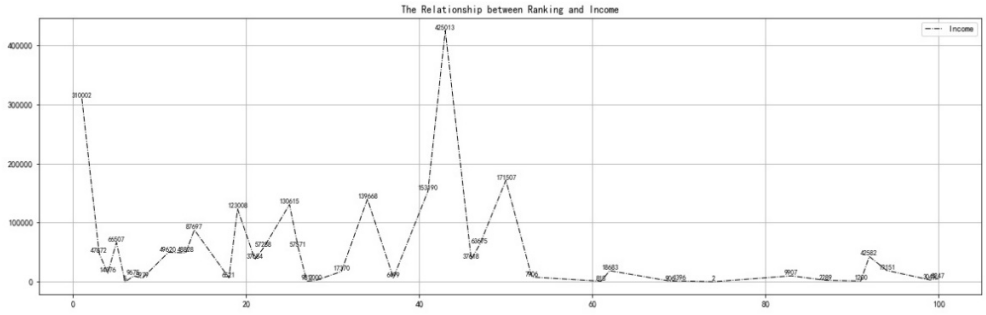
The ratings are all higher than 8.0, which means these 100 movies do not have any bad ones. However, movies at the top of the list may have relatively low ratings, and movies at the bottom may have high ratings.




The ranking is considered from a number of aspects. Overall, the ratings tend to go down as the rankings go down.

## 3. The relation between ranking and income

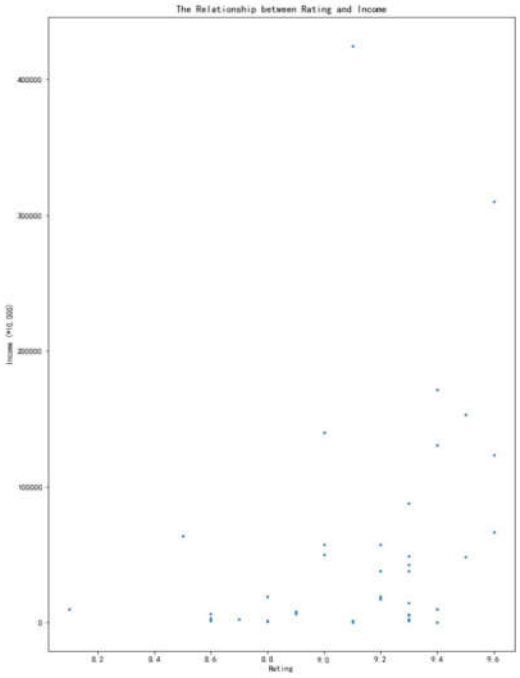
Due to the missing of income data for many movies, the visualization can only be use to show the possible trend.




Similarly, the ranking and income are not linear. Movies in the first half of the ranking list have large fluctuation changes. Movies in the second half have more smooth changes. In general, movies have better income with better ranking.

## 4. The relation between rating and income

Here is a scatter diagram to show the relation between rating and income.

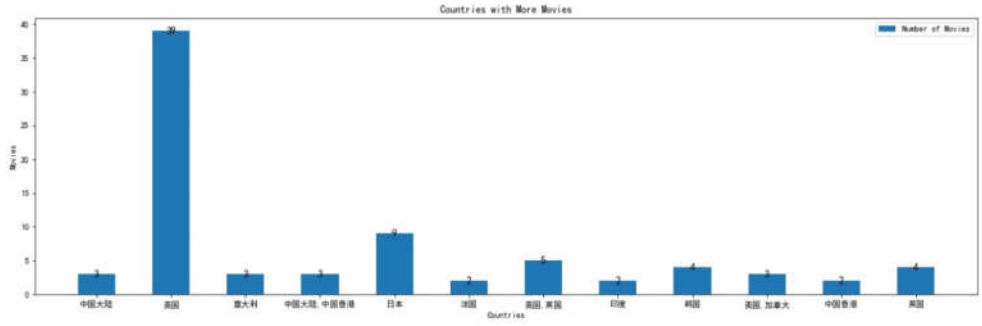



Many movies have the same rating, however, with really different incomes. There are many possible reasons. Firstly, the types of movies may cause the difference of incomes. Furthermore, the publish date may be influential. If a movie had been released in the last century, the number of people paying to see it would have been much lower.

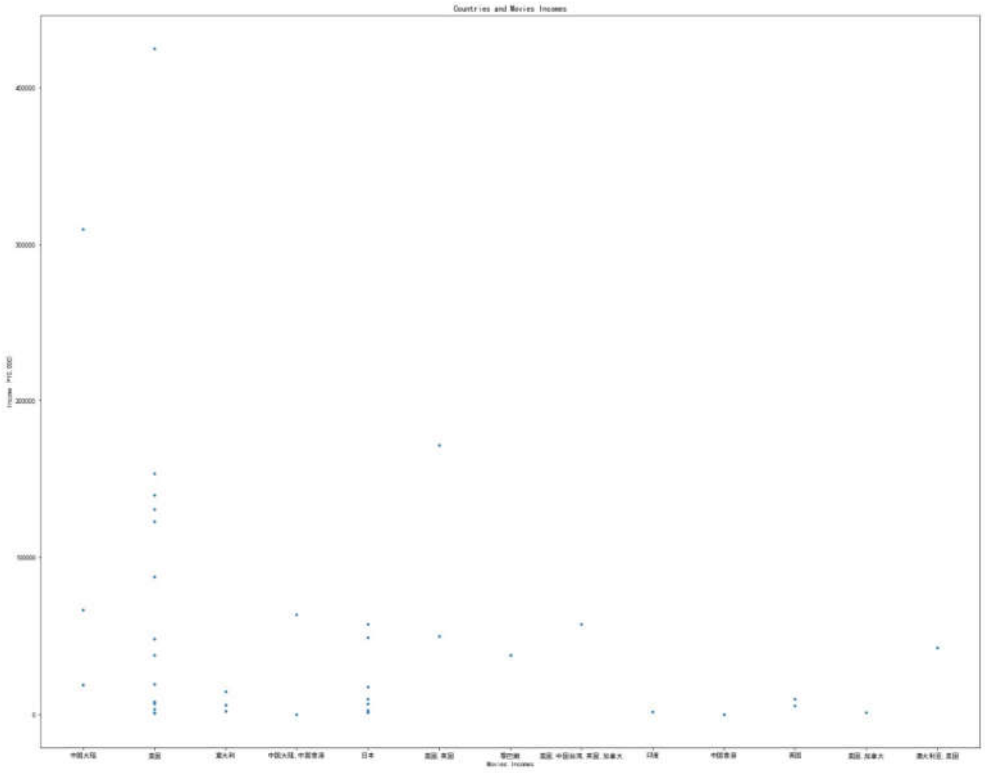
In general, the trend is that movies with higher rating may have better incomes.

## 5. The relation between country and movies

The diagram shows the number of movies form different countries. It only shows the countries with more than one movie in the top 100 list. Amazingly, 39 movies are from America. Japan is the second with 9 movies. There are also some movies are made by cooperation.



## 6. The relation between country and income




Similarly, the movies from America have both higher incomes and numbers of movies. Perhaps the technology and art of American film are indeed excellent.