

# Project Description

---

- **Motivation**

I myself quite enjoy watching movies. Douban movie website is a great website that I usually check when I look for good movies to watch or simply want to see movie comments. IMDB which is international movie database is also a quite famous movie website. I am curious about the differences between two databases and would like to find out some fun trend about worldwide top movies.

- **Data Source Description**

1. Top 250 douban movies data (scraped)

e.g. <https://movie.douban.com/top250?start=0&filter=>

Douban top 250 movies metadata and other information like comments are stored in these websites.

2. Top 250 IMDB movies data (API)

API is used to extract IMDB top 250 movies data.

e.g. [https://imdb-api.com/en/API/Top250Movies/k\\_w7cj8b1c](https://imdb-api.com/en/API/Top250Movies/k_w7cj8b1c)

3. movies box office (scraped)

I used "box office mojo by imdb pro" website to scrape box office data which includes domestic box office, international box office and worldwide box office.

e.g. [https://www.boxofficemojo.com/title/{id}/?ref\\_=bo\\_se\\_r\\_1](https://www.boxofficemojo.com/title/{id}/?ref_=bo_se_r_1)

- **Analysis and Conclusion**

- douban movie type analysis

This analysis is mainly focused on the types of Top 250 movies in douban websites. We can see what types top movies belong to. Through the bar chart, we can notice that drama movie is the most popular movie type in douban top 250 movies and romance, comedy, adventure and other kinds of movies are less popular. In order to see this trend much more clearly, a wordcloud diagram is created, which shows most popular movie types in large font size and less popular types in smaller size.

- douban movie country/region analysis

In this analysis, I create a bar chart and a pie chart to see the country/region trend of movies. Results present to us that in douban top 250 movie list, American movie occupies a quite large proportion. What next to it are movies made from Japan, England, Hongkong, China mainland.

- shared movies analysis

I scraped the top 250 movies data separately from douban website and IMDB website. Douban website is a famous movie website in China, and IMDB is the international movie database. The topic I want to analyze most is how different are those movies on the lists. Through above tentative analysis, we find out that in those top 250 movie lists, 95 movies are overlapped. Let's find out more information about these movies. We can see that movies belong to both top lists are mostly drama, crime genre as the word cloud diagram shows. Also, most of them come from American and England as the pie chart shows.

- movie rating analysis

We are able to see that the ratings in douban website are between 8.3 and 9.7 while those in IMDB are between 8.0 and 9.2. The rating mean is higher in douban movie rating. With the help of the visualization, we can see the distribution of the ratings in two datasets. The rating distribution in imdb dataset is more to the left.

- worldwide box office analysis

In this analysis, I grabbed top 10 movies with highest worldwide box office in both datasets. It turns out that the ranks of movies in the top 250 lists have little to do with the box office. Few of those movies in top 10 box office lists belong to the top 10 movies lists in general.