**Topic:**

What are the differences between Chinese and American audience’s taste for movies?

**Background:**

Douban is the Chinese version of IMDB, where users can rate movies and leave comments. Similar to IMDB TOP250, Douban has its own version of TOP250 based on Chinese audience’s taste. It’s common that same movie ranks high on IMDB but ranks much lower on Douban, vice versa.

**Data Source:**

I will begin scrapy at the Douban TOP250 site (<https://movie.douban.com/top250>) and go further, crawling each movie’s basic infomaiton, key words, quote, rate, IMDB ID and etc.

Then I will use the IMDB API (<http://www.omdbapi.com/>) to get the information of corresponding movie

Web API you haven’t used before that requires API key or HTTP Basic authorization

Crawling [and scraping] multiple pages in a site you haven’t used before

**Overall challenge score should be 4 + 8 = 12**

**Overall records should be more than 250 + 250 = 500**

**Data Presentation:**

Users can select different methods to compare rating, comments and etc. for same movie.

* Concise method: simple and clear, only movie name, rank in Douban and IMDB would be presented
* Thorough method: complete information including key words, rank, actor&actress, comments
* Comparative method: focus on the differences between Douban and IMDB based on frequency of different key words, key words cluster and etc.

Users can select different methods to compare the overall situation of movies in Douban Top250 and IMDB.

* Users can see the distribution of top250 movies in Douban and their ranks in IMDB and what’s the differences
* Users can see the key words cluster that top250 movies in Douban has and in IMDB has
* More to come like distribution based on release date, actor/actress and etc.

**Data Presentation Tools:**

I will try to create a **Flask App** using HTML links/form elements to prompt for the user to choose data/visualization options. Display selected data using HTML tables in different methods like concise method, thorough method, comparative methods.

**Basic Deliverable:**

Python program where users can input number to check movie in Douban TOP 250 and corresponding rank in IMDB TOP250, how many people rated, rates and other information

**Advanced Deliverable:**

1. A flask web app where users can select from movie’s poster to see the information above.

2. Add key words cluster function. So users can find what Chinese audiences like and what American audiences like.

3\*. Improve UI design of the web app

4\*. Add export function. So users can generate an information card display the differences

5\*\*. Add search function. So users can search and compare any movie between Douban and IMDB