****

**School of Advanced Technology**

**Project 1 Report**

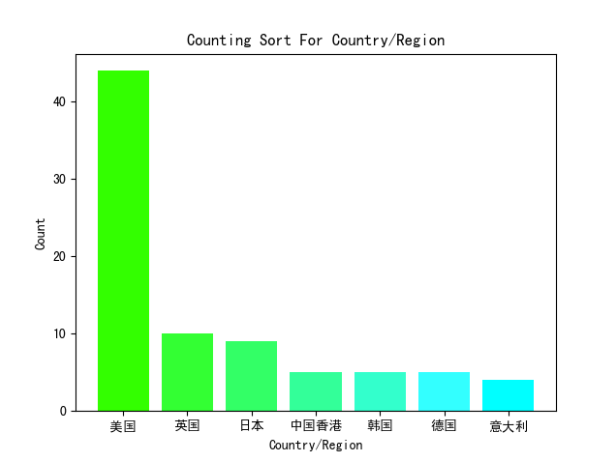
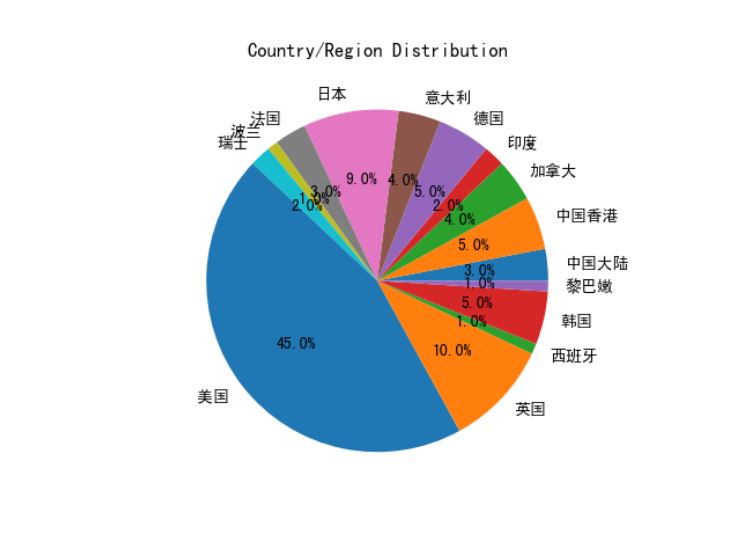
Project Title: Assignment 1: Web Scraping & Data Analysis

Student Name: Xingbo Wei

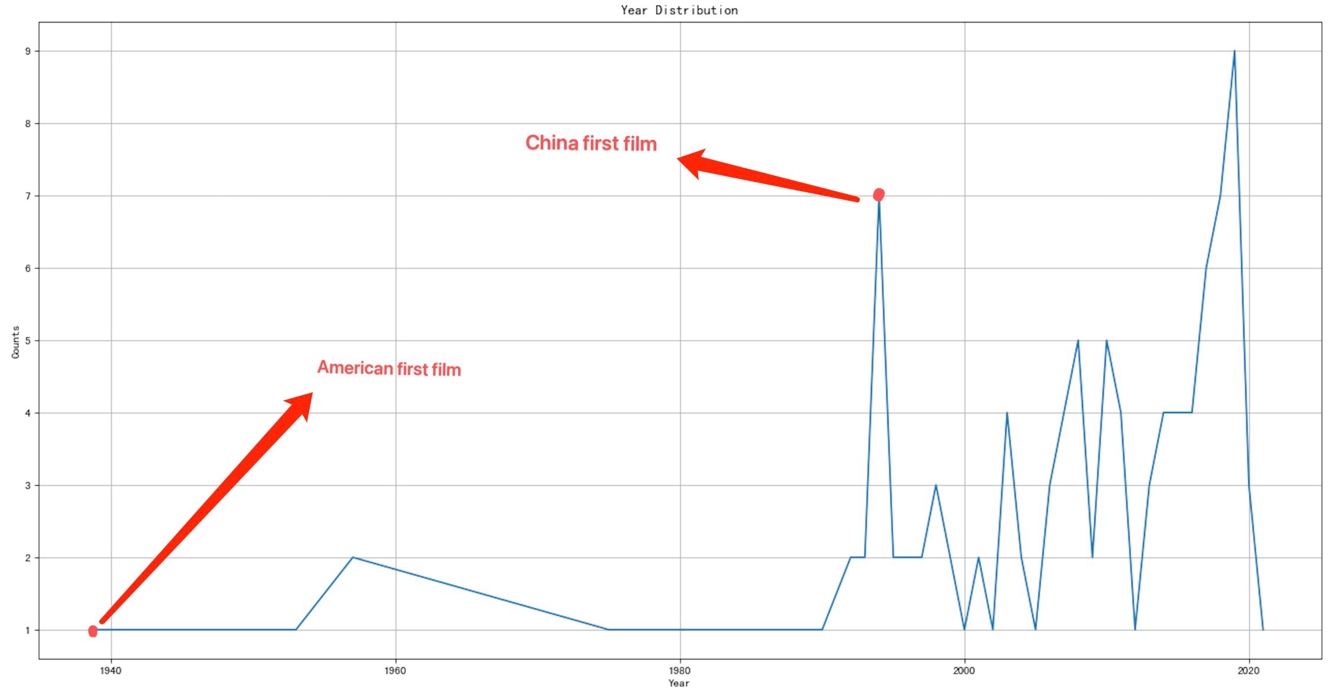
Student ID: 1824150

In this project, I made Maoyan web crawlers using python. I got the Maoyan top 100 movies data after crawling. I use pandas python package to operate and conduct the data. Finally, I use matpltlib, collections and wordcloud python package to visualize those data after conduct.

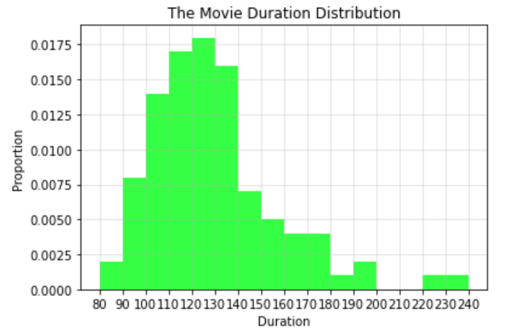
Maoyan top 100 means those 100 movies are fascinating in generally. I an interest in where are they from so, based on movies data I draw the char which is the region/country for the top 100 movies, as shown in the finger, 45% of movies are from American, which means that American movie field is very strong and have a mature system and environment in this field, Hollywood is a good example.

****

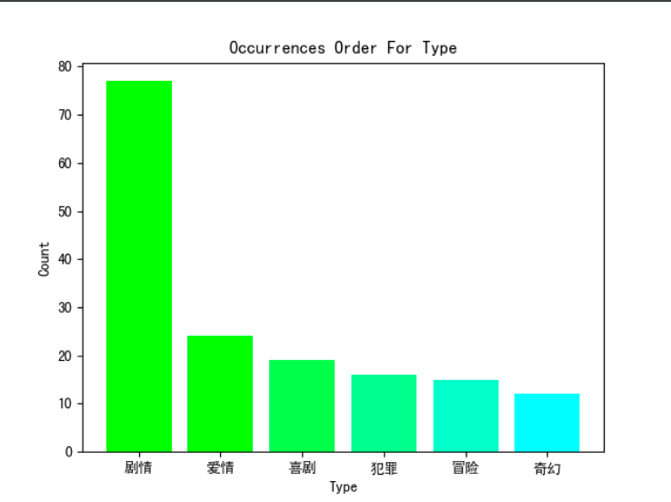
famous in the world as the movie industry. But there is an interesting phenomenon is that the earliest movie in Maoyan top 100 from American in 1939 named *Gone with the Wind* and the earliest movie from China in 1993 named *Farewell My Concubine* 55, years later than American. Based on the image, in 1993-2021 there are eight movies to be Maoyan top 100 and average 3.37 movies per year, 81 American movies in the Maoyan top 100 from 1939-2021, average 1.76 movies per year lower than China. It proves that China movie industry has enormous potential I think with the development of China, Maoyan top 100 will have more and more China movies. Next, I will give some suggestions based on the movie type and movie runtime for the fact of adopted in Maoyan top 100.

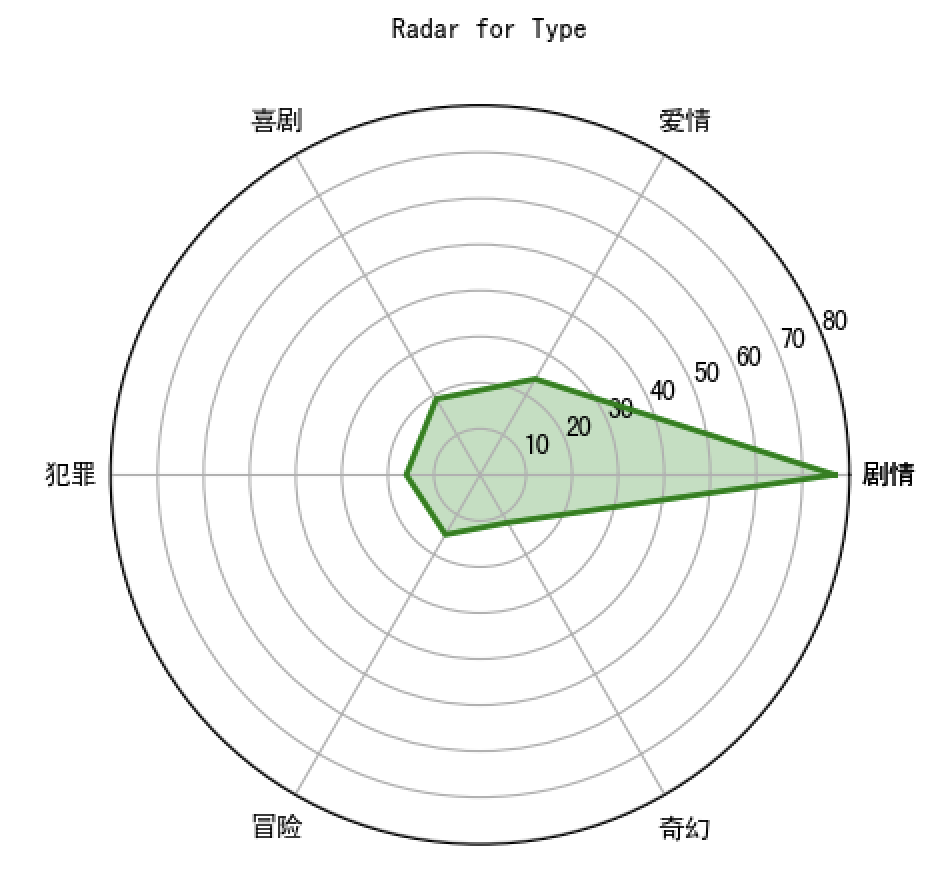


As the image shows, over fifty per cent movies duration in the 110 minutes to 140minutes range, this is an appropriate range for movies. too long running time will let the audience feel tired even the film is fascinating, to the contrary, too short runtime is also unreasonable, because it is hard to tell a whole story in a limit time.

****

A good film has other factors, for example the genre is also an important factor for a good movie. I draw the picture which rank the type of the top 100 movies. 78 films contain the drama element, nearly 80% of movies contain it. Those data imply director can choose the appropriate runtime and genre to shoot the movie that easily for the population, but it does not demonstrate a good film have to choose the specific runtime and genre, it just means the factor I mentioned is an easy way to create a popularity movie, and the movie not adopted in Maoyan top 100 it doesn’t mean is not a good movie.

****



A fascinating movie have many factors not just only about I mentioned above. The acting of actor, shooting costs, background and so on… I believe there are various of type and country in the future.