****

**School of Advanced Technology**

**Project 1 Report**

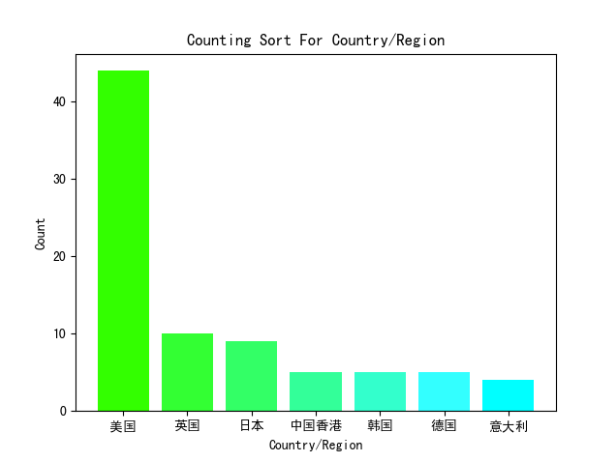
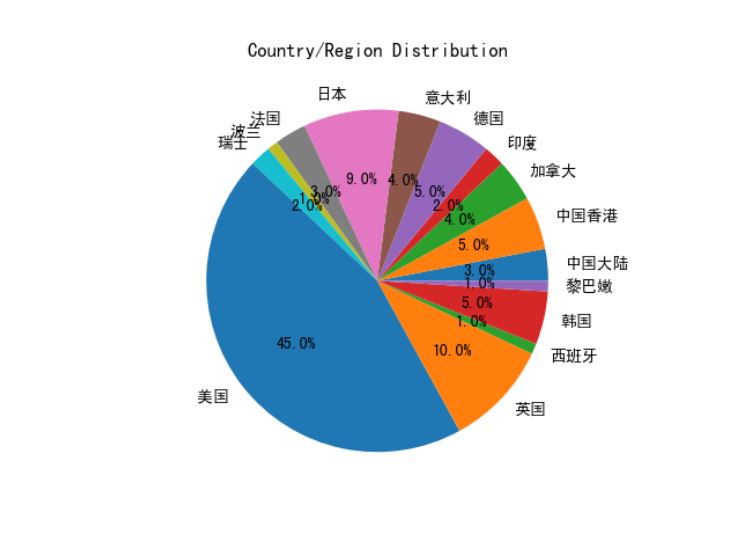
Project Title: Assignment 1: Web Scraping & Data Analysis

Student Name: Xingbo Wei

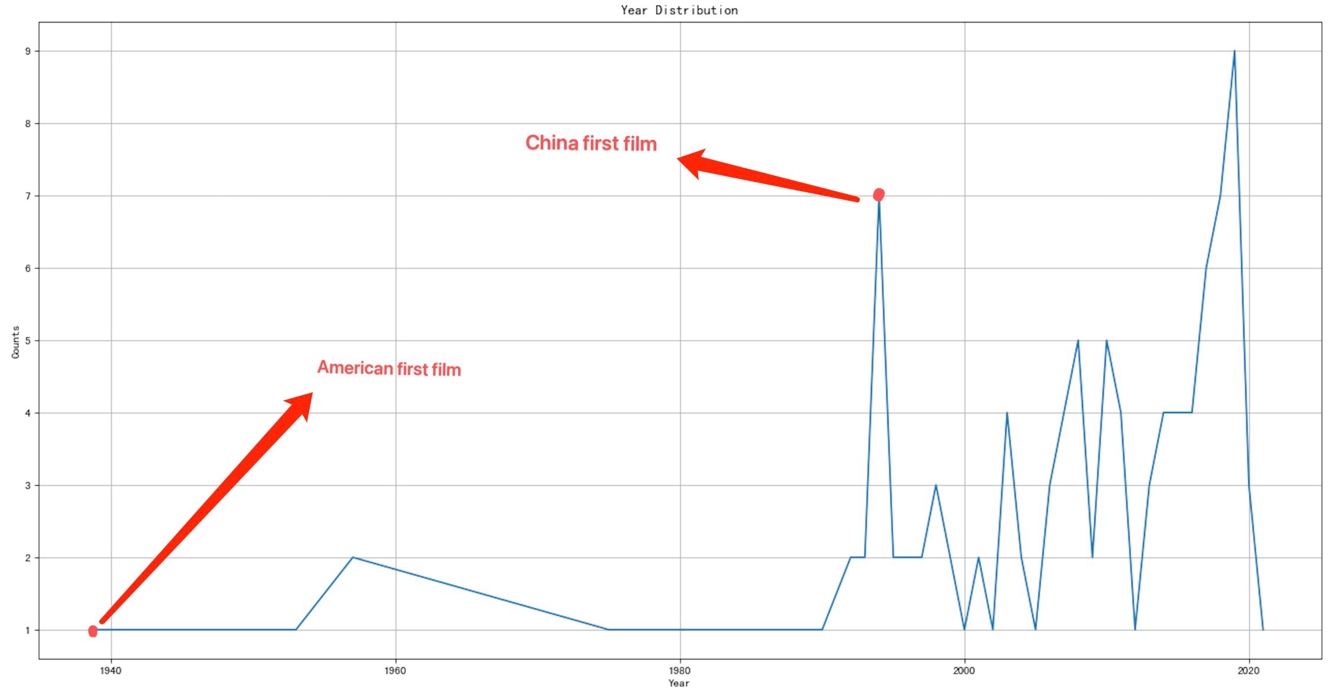
Student ID: 1824150

In this project, I making Maoyan web crawlers using python. I got the Maoyan top 100 movies data after crawling. The analysis and visualization of data are as follows.

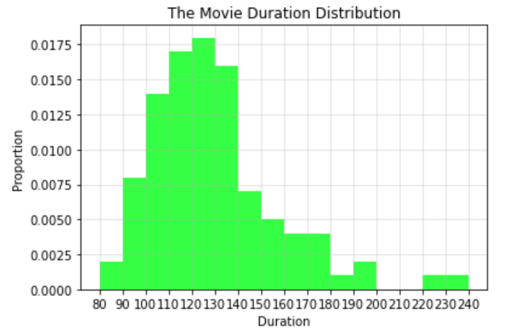
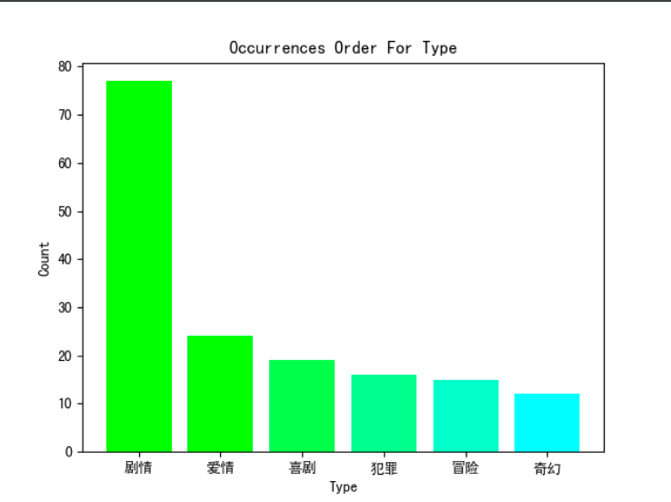
Maoyan top 100 mains those 100 movies are fascinating in generally. Based on movies data I draw the char which is the region/country for the top 100 movies, 45% of movies are from American, which means that American have mature system and environment in the movie field, Hollywood is also

****

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. 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 has more and more China movies. Next, I will give some ideas based on movie type and movie runtime for the fact of adopted in Maoyan top 100.



As the image shows, over fifty percent movies duration in the 110 minutes to 140minutes range, too long will let the audience feel boring and it hard to catch the audience eyes throughout the file, to the contrary, actors are hard to show the characters features and the audience hard gets emotion from the film if the film is too short. The genre is also an important factor for a good movie. 78 films contain the drama element, nearly 80% of movies contain it. Those data imply director can choose the property runtime and genre to create the movie that easily for 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 easily way to create a population movie, and the move not adopted in Maoyan top 100 it doesn’t mean which is not a good movie.

****