

IDC409 Project report

Yogendra Kumar MS19113

October 13, 2022

1 Aim:

Scrap a movie database and analyse the gender bias as a function of time. [scrapping + analysis]

2 Introduction:

Based on facts, women's engagement in the workforce has significantly altered during the previous 100 years. Women only made up 20% of the labour force in 1920, but as laws and women's rights improved, more women entered and stayed in the workforce. Women's participation in the work force reached a high of 60% in 1999, after which it fell to 47%.

Even while women's involvement in the workforce has generally increased throughout time, certain industries and professions still have far lower gender representation than others. For instance, just 25.5% of architects and only 33% of workers in the software industry are women. In the same manner, women are notably underrepresented in the film business, where men predominate in all professions, including acting, directing, and writing.

3 Data From Web-scrapping:

We use [IMDB](#) website to scrap the deta. It has all U.S. Released Movies from 1964 to 2027. It contains total 9935 movies, each listed with their general information. I scraped some important information like movie name, year of release, director name, stars name and their gender. After scrapping, we save it into a csv file to further analysis the data.

4 Data Analysis:

First we plot the gender proportions in the hollywood film industry over all time and we got that 61% occupy the male actors and remaining 39% has female.

Gender Proportions in the Hollywood film industry

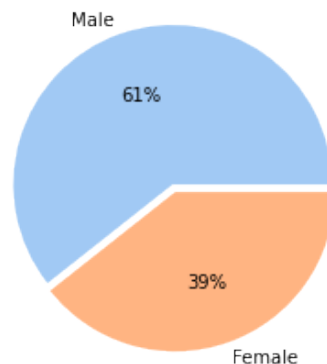


Figure 1: Gender Proportions

After this, we plot the bar chart of male/female stars percent in a year over all time in the hollywood film industry.

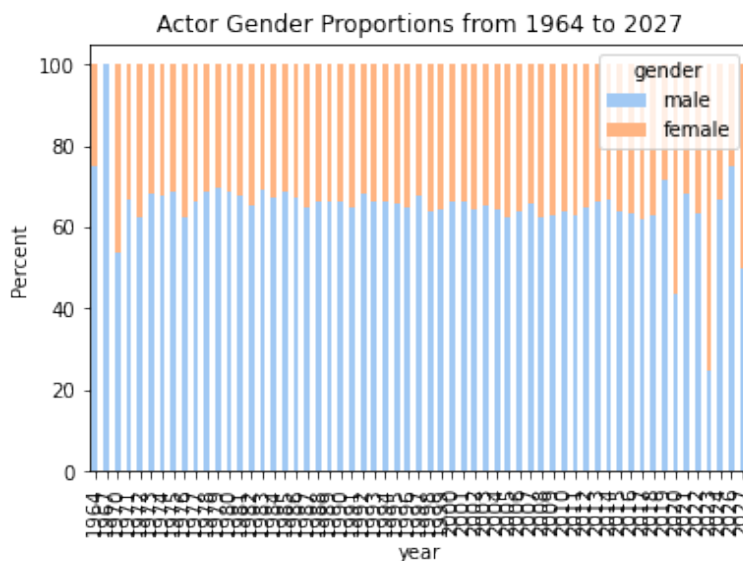


Figure 2: Gender Proportions over time

5 conclusion

We scrap the data from the IMDB website and then analysis it and found that in the hollywood film industry there are 61% males and 39% females stars. Also we see the gender proportions over all time. All the scraping part done on my local computer and analysis part done on the google colab and file is attached with this report [Here](#).

6 References

- <https://www.imdb.com/list/ls057823854/>
- https://en.wikipedia.org/wiki/Cinema_of_the_United_States