

MARGARET'S IMDB PROJECT

A simple project to identify good
movies for dinners



DATA SOURCES

IMDb datasets

Downloaded from <https://datasets.imdbws.com/>

Further data for movies, cast & crew

Unavailable in IMDb datasets online

Collected by Python scripts using IMDbPY library

For additional movie info -

https://github.com/HuynhMargaret/imdb-project/blob/main/get_additional_movie_info.py

For additional cast & crew info -

https://github.com/HuynhMargaret/imdb-project/blob/main/get_additional_castcrew_info.py

DATA ANALYSIS EXERCISES

Some data analysis exercises

Calculate Bayesian average rating of a movie

Find all years that have a movie that received a rating of 9 and above

Find the difference between the average rating of movies released before 1980 and the average rating of movies released after 1980 (Answer: Older movies are rated better)

More to come...

Languages used

Python - https://github.com/HuynhMargaret/imdb-project/blob/main/data_analysis_python.py

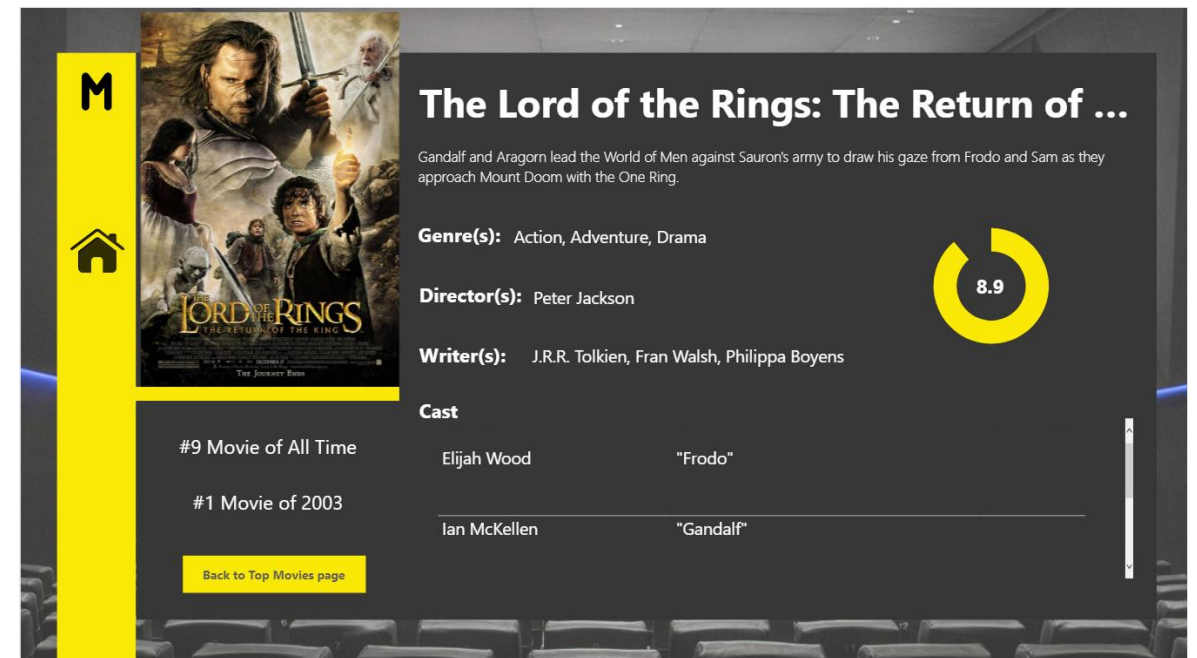
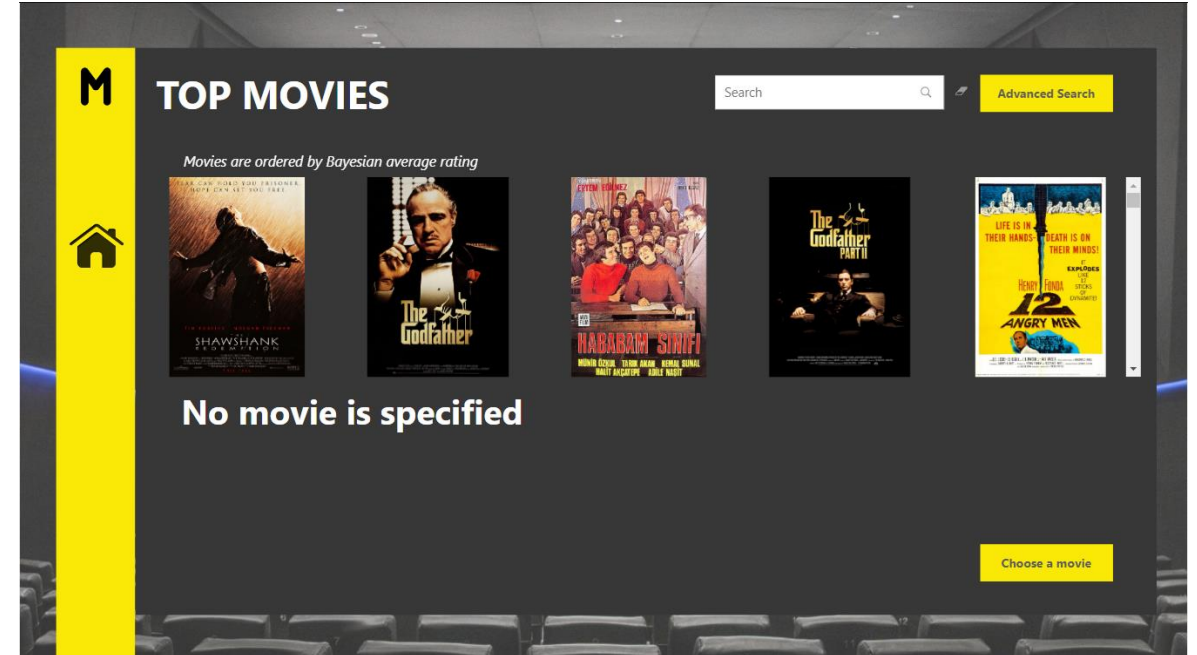
SQL - https://github.com/HuynhMargaret/imdb-project/blob/main/data_analysis_sql.txt

POWER BI DASHBOARD

[https://github.com/HuynhMargaret](https://github.com/HuynhMargaret/imdb-project/blob/main/imdb-top-movies-dashboard.pbix)

[t/imdb-project/blob/main/imdb-](https://github.com/HuynhMargaret/imdb-project/blob/main/imdb-top-movies-dashboard.pbix)

[top-movies-dashboard.pbix](https://github.com/HuynhMargaret/imdb-project/blob/main/imdb-top-movies-dashboard.pbix)



POWER BI DASHBOARD – A CLOSER LOOK (1)

The screenshot shows a movie dashboard with a yellow sidebar on the left containing a house icon and a yellow bar with the letter 'M'. The main area is titled 'TOP MOVIES' and features a search bar with 'tom cruise' and an 'Advanced Search' button. Below the search bar, a row of movie posters is displayed, including 'Rain Man', 'Magnolia', 'Edge of Tomorrow', 'Mission: Impossible - Ghost Protocol', and 'A Few Good Men'. The 'Edge of Tomorrow' poster is highlighted with a yellow box. Below the posters, a detailed view for 'Edge of Tomorrow (2014)' is shown, including a quote, director information, cast list, a 7.9 rating, and a 'Detailed Info' button.

M **TOP MOVIES**

tom cruise [Advanced Search](#)

Movies are ordered by Bayesian average rating

RAIN MAN

magnolia

LIVE DIE REPEAT

EDGE of TOMORROW

MISSION: IMPOSSIBLE GHOST PROTOCOL

A FEW GOOD MEN

Edge of Tomorrow (2014)

'A soldier fighting aliens gets to relive the same day over and over again, the day restarting every time he dies.'

Director(s): Doug Liman

Cast: Emily Blunt, Bill Paxton, Brendan Gleeson, Tom Cruise

7.9

[Detailed Info](#)

Keywords such as actor/actress name, genres, plot detail can be put in search box to filter relevant movies

Once a movie poster is selected, brief info of the movie will be showed, with options for users to check more detailed info

POWER BI DASHBOARD – A CLOSER LOOK (2)

M **TOP MOVIES**

Movies are ordered by Bayesian average rating

Edge of Tomorrow (2014)

'A soldier fighting aliens gets to relive the same day over and over again, the day

Director(s): Doug Liman
Cast: Emily Blunt, Bill Paxton, Brendan Gleeson, Tom Cruise

tom cruise

Apply

United States Genres

Year

1911 2021

Average Bayes Rating

6.50 9.29

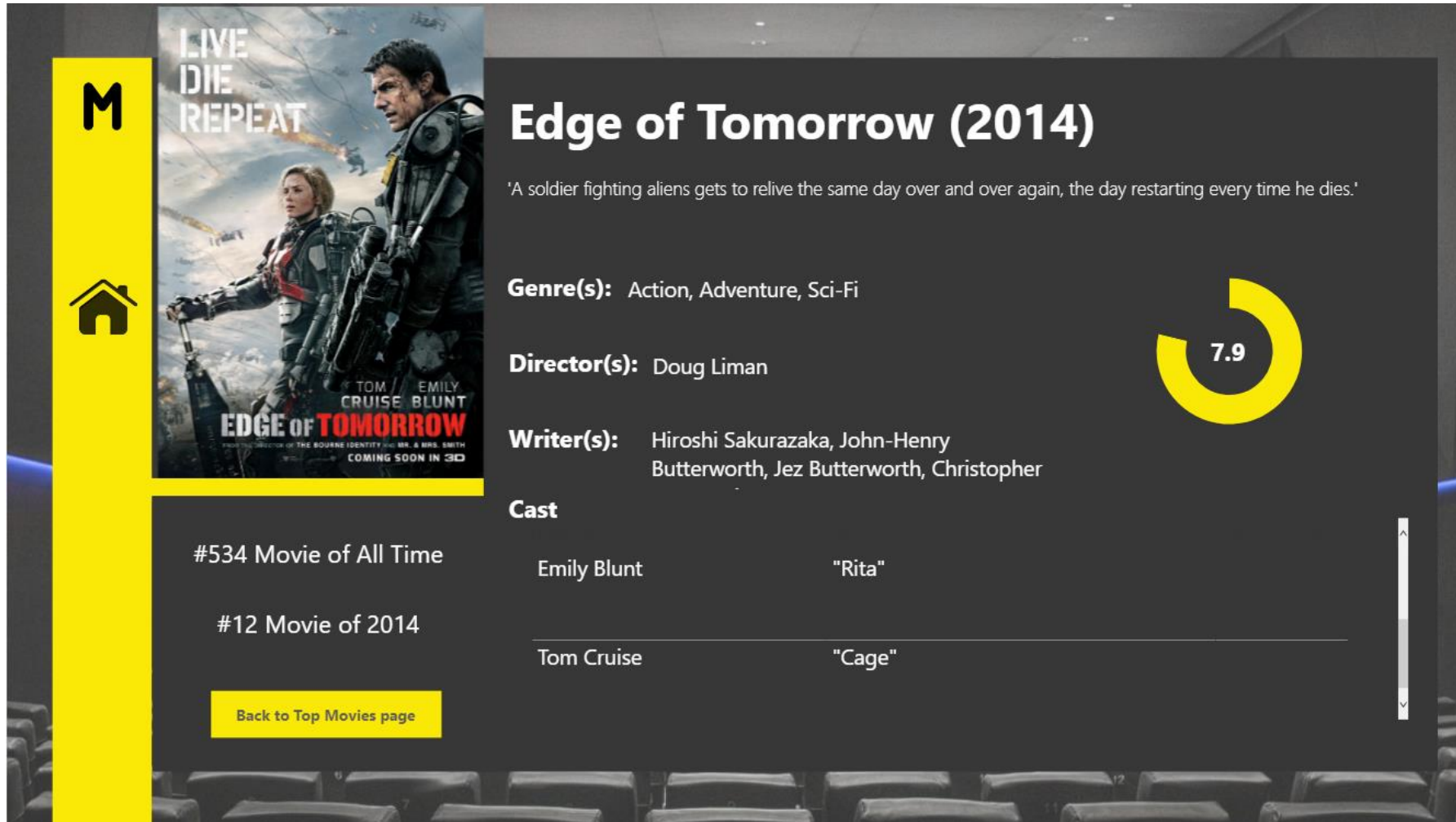
Number of Votes

31 2364351

'Advanced search' button can be used for more refined results

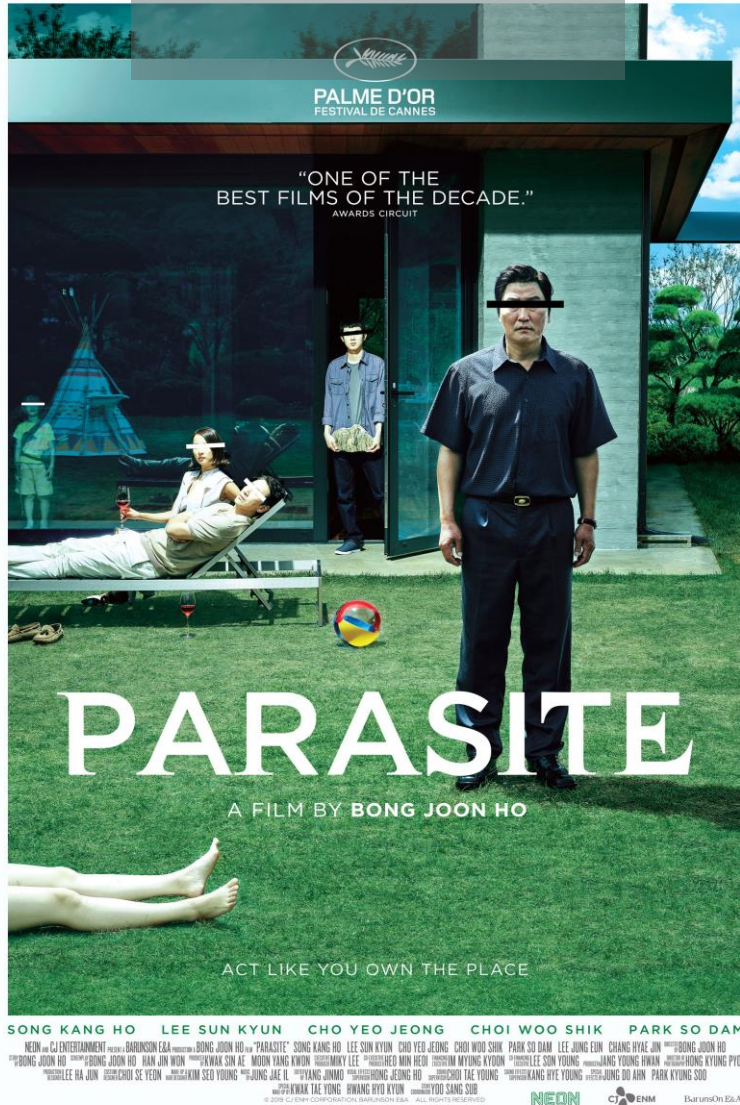
Once finished, users can go back to home page by clicking 'Apply' button

POWER BI DASHBOARD – A CLOSER LOOK (3)



Detailed Info page offers further information about the chosen movie, such as top characters and movie rank within the year or all time.

Once finished, users can go back to relevant movie list by clicking on 'Back to Top Movies page' button



FUTURE PLANNING FOR PROJECT

Further data analysis in Python with matplotlib, seaborn, etc.

Explore data modelling of movie information (plot, genres, cast, etc.) to recommend good movies, with ambition to learn and apply Natural Language Processing & Machine Learning frameworks