

IMDB Movie Analysis

Shruti Jha

Description

The IMDb logo is displayed in a yellow, tilted rectangular box in the top right corner of the slide. The text 'IMDb' is in a bold, black, sans-serif font.

The objective of this project is to conduct a comprehensive analysis of the IMDB movie dataset, aiming to provide valuable insights. The analysis will address key questions, including:

1. **Data Cleaning**: Ensuring the dataset is accurate and free from inconsistencies.
2. **Movies with Highest Profit**: Identifying the films that have generated the highest profits.
3. **Top Movies with the Best IMDB Rating Score**: Determining the movies with the highest ratings according to IMDB users.
4. **Best Directors**: Recognizing the most accomplished directors based on various criteria.
5. **Movie Releases Over Decades**: Examining the trends and patterns of movie releases across different decades.
6. **User Rating Over Decades**: Analyzing how user ratings have changed over time.
7. **Critic Favorite Actors**: Identifying the actors who have received significant praise from critics.
8. **Audience Favorite Actors**: Determining the actors who are most beloved by the audience.

Through this project, we aim to provide meaningful insights into the IMDB movie dataset.

Approach

Before moving forward with the analysis, we performed three important steps:

- Firstly, we dropped the columns that have no use for the analysis.
- Secondly, we removed the rows that were blank or null.
- Thirdly, we eliminated duplicate row values.

These steps were taken as a crucial preliminary procedure to ensure the dataset is streamlined and ready for analysis.



Tech Stack Used



Microsoft Excel

View the excel sheet here:

https://1drv.ms/x/s!Aqaf_DqYTupjkSWXyL-z897nnjD0?e=HTP0Zc

Task 1: Cleaning the data



Before moving forward with the analysis, we performed three important steps:

- Firstly, we dropped the columns that have no use for the analysis.
- Secondly, we removed the rows that were blank or null.
- Thirdly, we eliminated duplicate row values.

These steps were taken as a crucial preliminary procedure to ensure the dataset is streamlined and ready for analysis.

Out of the 27 variables(columns) that were initially there, we required only 14 of them for processing the required answers from our data. So those columns were deleted.

Next, the duplicate rows were deleted, followed by rows having null values for necessary variables like Gross, Budget, etc were also removed.

Look for any abnormal values using the filter option

Task 2: Movies with highest profits

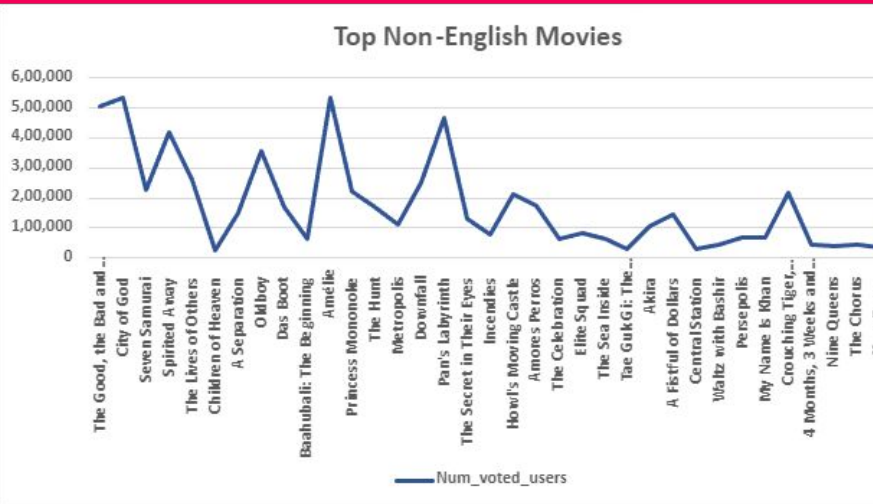
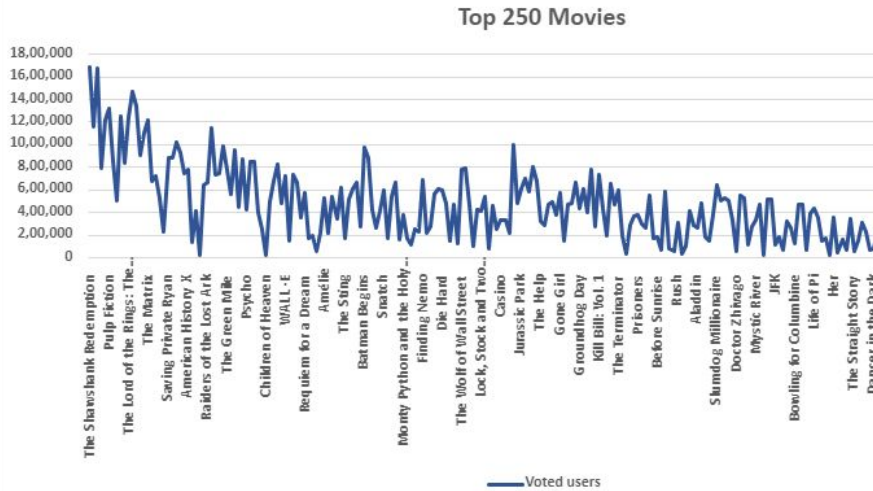
IMDb

Movie_title	Profit
Avatar	52,35,05,847
Jurassic World	50,21,77,271
Titanic	45,86,72,302
Star Wars: Episode IV - A New Hope	44,99,35,665
E.T. the Extra-Terrestrial	42,44,49,459

- ★ Avatar is the highest profit of 523 million dollars
- ★ Star wars: A new hope has the least budget among all the top 5 movies with a scoring profit of 449 million dollars

Task 3: Top 250 movies

The Shawshank Redemption is the highest rated movie on the IMDB platform with a 9.3 rating



The Good, the Bad and the Ugly (Italian) is the top most foreign language movie with an IMDB score of 8.9, followed by City of God (Portuguese)

Task 4: Top 10 Best Directors



Director_name	Average IMDB score
Charles Chaplin	8.6
Tony Kaye	8.6
Alfred Hitchcock	8.5
Damien Chazelle	8.5
Majid Majidi	8.5
Ron Fricke	8.5
Sergio Leone	8.4
Christopher Nolan	8.4
Asghar Farhadi	8.4
Marius A. Markevicius	8.4

Charles Chaplin is the best director with an average IMDB score of 8.6 followed by Tony Kaye and Alfred Hitchcock

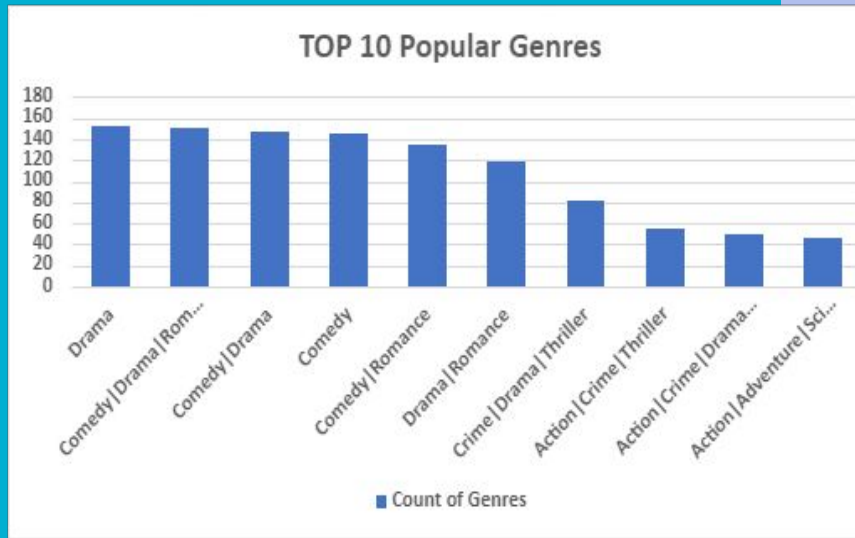
The top 4 directors direct English movies whereas, Majid and Asghar direct Persian movies, Sergio directs Italian movies and the rest direct English movies

Task 5: Most popular genres

Genres	Count
Drama	153
Comedy Drama Romance	151
Comedy Drama	147
Comedy	145
Comedy Romance	135
Drama Romance	119
Crime Drama Thriller	82
Action Crime Thriller	55
Action Crime Drama Thriller	50
Action Adventure Sci-Fi	46

The most popular genre among users is Drama, followed by Comedy|Drama|Romance and Comedy|Drama

IMDb

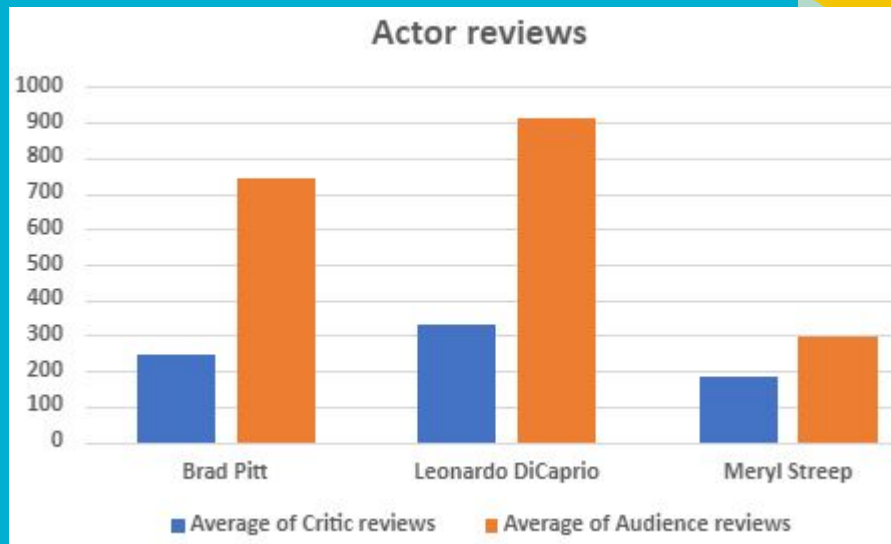


Task 6: Critic & Audience Choice

IMDb

Actor Name	Average Critic reviews	Average Audience reviews
Brad Pitt	245	742.4
Leonardo DiCaprio	330.2	914.5
Meryl Streep	181.5	297.2

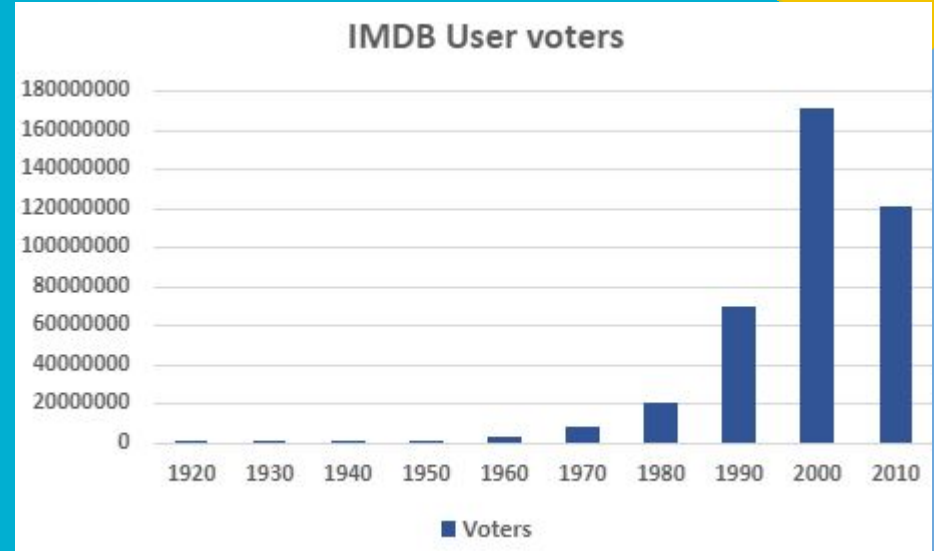
Leonardo DiCaprio is both the critic and audience's most voted and favourite actor.



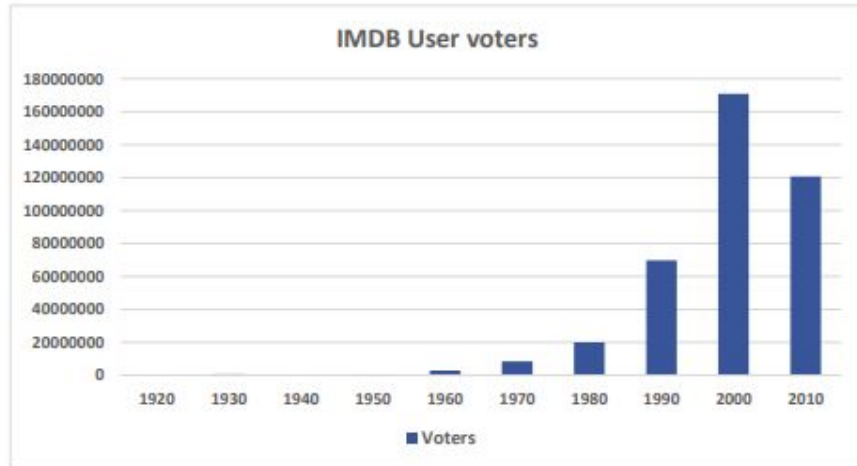
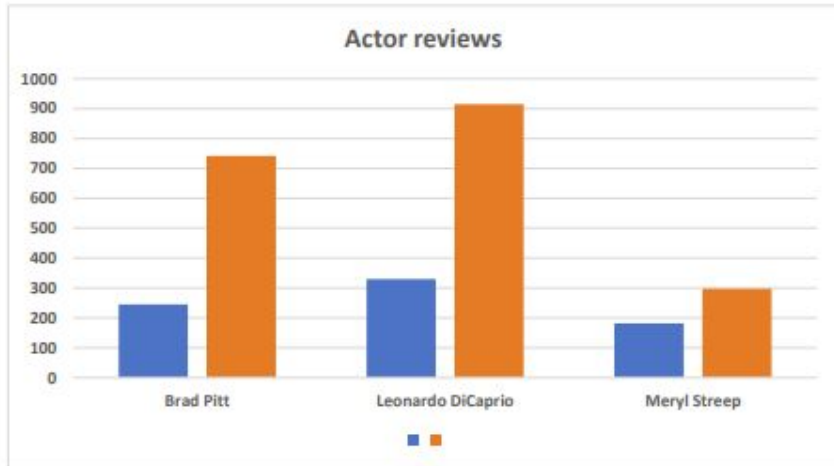
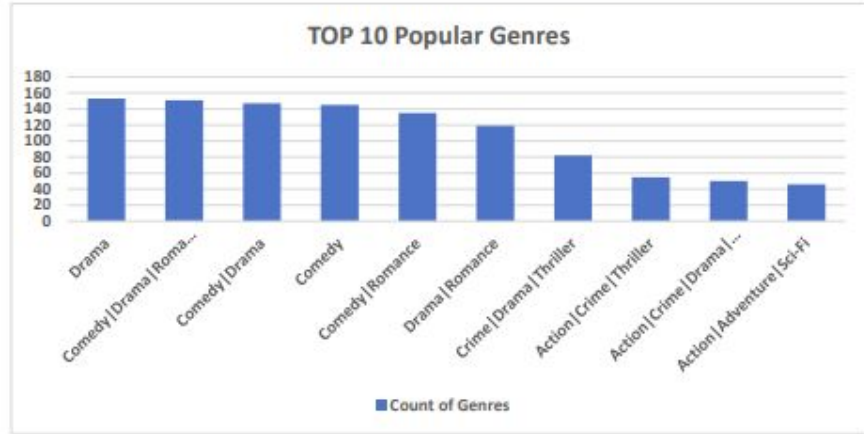
Task 7: Voter growth per decade

There is a very slow growth of users from 1920-1960.

Between 1960-1980 there is a linear growth of users that voted and soon after that, we can see the exponential spike in user count hitting its peak in the year 2000 of 170 million.



DASHBOARD



CONCLUSION

This IMDB movie analysis project has provided valuable insights into various aspects of the dataset, including financial success, ratings, directors, release patterns, user preferences, and actor popularity.

The cleaned and analyzed dataset provides a foundation for further exploration and decision-making in the field of movie analysis.

These findings contribute to a deeper understanding of the movie industry and can be utilized for further research, decision-making, and exploration within the field.

View the excel sheet here:

[IMDB Movies Analysis](#)

https://1drv.ms/x/s!Aqaf_DqYTupjkSWXyL-z897nnjD0?e=HTP0Zc



THANK YOU!

Prepared by: Shruti Jha