

IMDB Movie Analysis

- By Sandeep Tomar



Contents

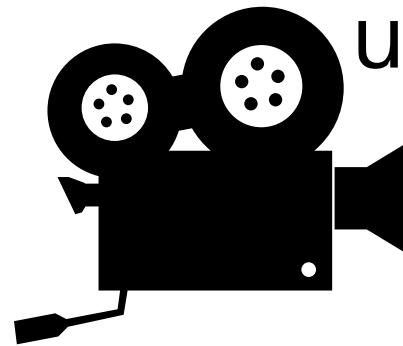
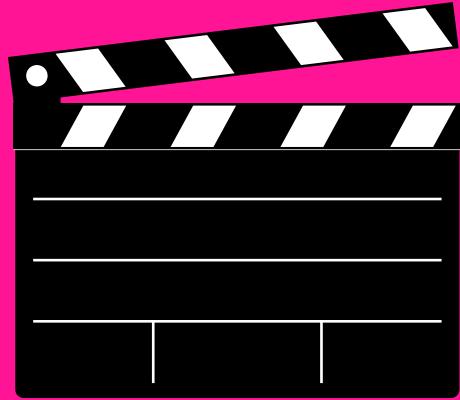


- **Project Description**
- **Approach**
- **Tech-Stack Used**
- **IMDB Movie Analysis**
- **Insights**
- **Result**

Project Description

This project aims to analyze the IMDB Movies dataset to learn more about the movie business. The dataset contains a number of columns relating to numerous films, including their budget, gross, director, actors, genres, and ratings.

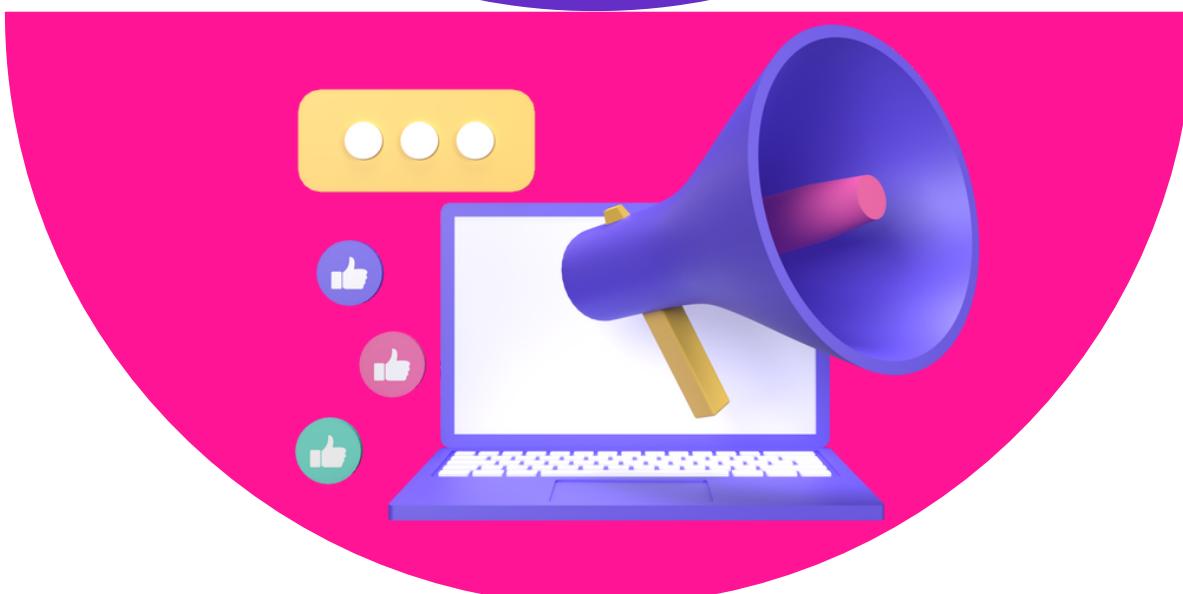
The project wants to answer several dataset-related queries, including determining the highest-grossing films, the top 250 films based on IMDb ratings, the top directors based on IMDb ratings, popular genres, and more. Excel Spreadsheet will be used for the analysis.



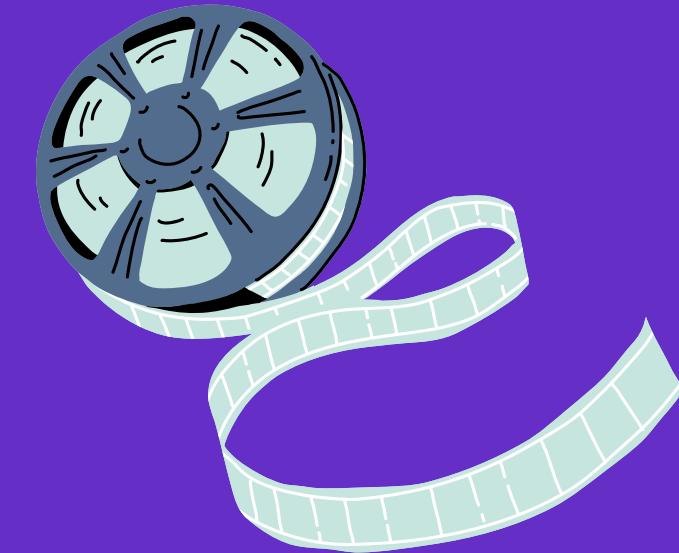
Approach

The approach towards this project is to first download all the data provided and then after understanding the data. I clean the dataset and remove null values or drop the unnecessary columns and duplicate values.

Then, I do a step-by-step analysis, answering the questions mentioned in the project description. I use various Excel functions and tools such as sorting, filtering, grouping, and plotting to find the required insights



Tech-Stack Used



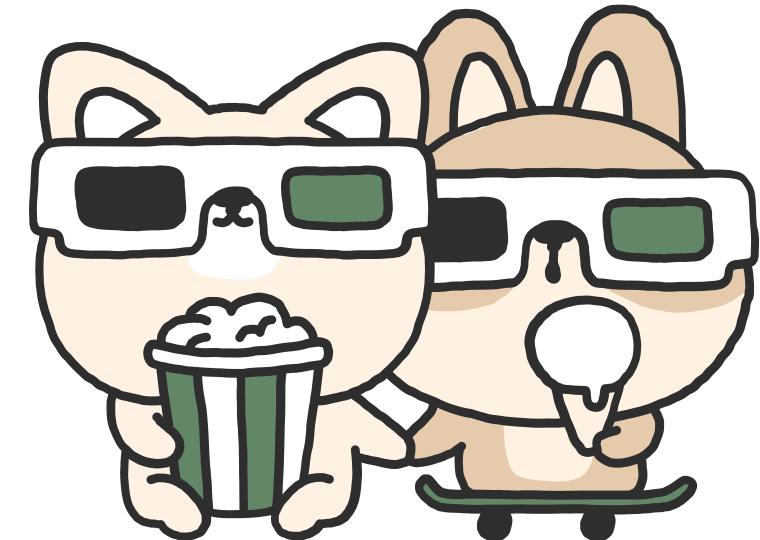
For this project, I used Microsoft Excel (2021).

This software program offers a variety of tools and features that will be useful in carrying out the necessary analysis, making them perfect for doing data analysis and modification.

IMDB Movie Analysis



- 01. Cleaning the data.
- 02. Movies with highest profit.
- 03. IMDB Top 250.
- 04. Best Directors.
- 05. Popular Genres.
- 06. Charts.

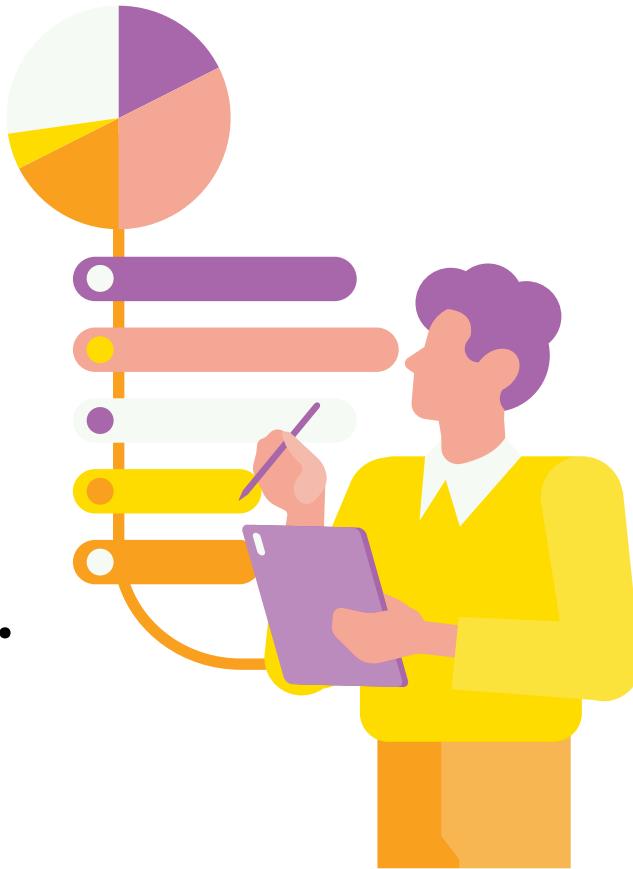


Insights

Excel may be used to provide answers to all of the questions asked. I used various tools and functions that will be helpful in performing the required analysis.

01. Cleaning the data !

- Count the number of empty cells in all columns and rows using the COUNTBLANK() method.
- I am analyzing the movies in terms of gross collection, ratings, popularity, and so on, and many columns in the dataset are unnecessary. This is the reason I have removed the following columns. `color, director_facebook_likes, actor_1_facebook_likes, actor_2_facebook_likes, actor_3_facebook_likes, actor_2_name, cast_total_facebook_likes, actor_3_name, duration, facenumber_in_poster, content_rating, movie_imdb_link, aspect_ratio, plot_keywords, movie_facebook_likes.`
- Drop duplicate values.
- I can also observe that some columns have a high percentage of null values, which means that such rows would be dropped.
- I had 5044 rows and 28 columns before cleaning (including the title in the first column). Now, I have 3783 rows and 13 columns (including the title in the first column) after cleaning.



Cleaning the data

6000

5000

4000

3000

2000

1000

0

Columns

28

13

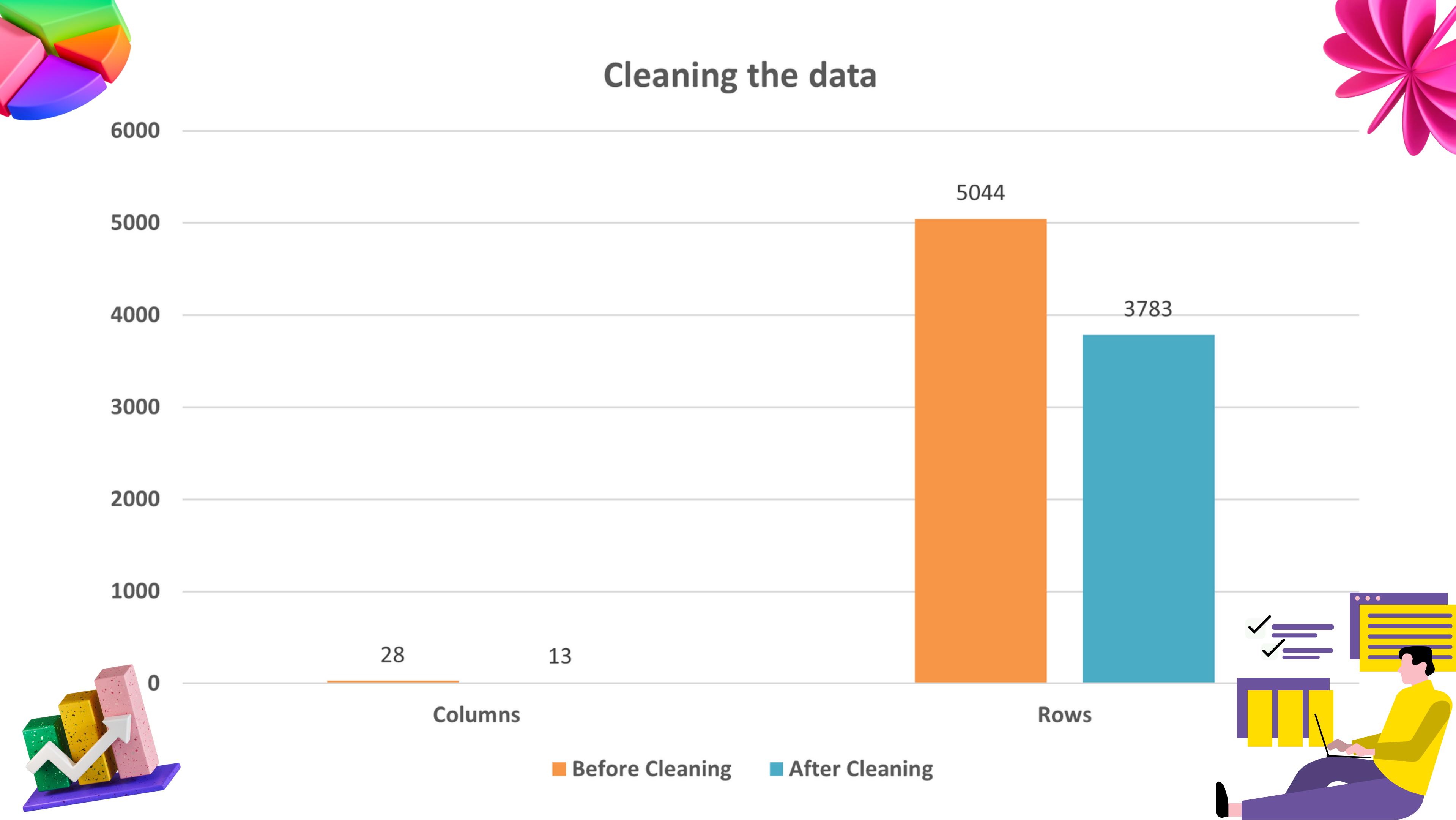
5044

3783

Rows

Before Cleaning

After Cleaning





02. Movies with Highest Profit !

Movies Title	Profit
Avatar	\$523.50
Jurassic World	\$502.20
Titanic	\$458.70
Star Wars: Episode IV - A New Hope	\$449.90
E.T. the Extra-Terrestrial	\$424.40
The Avengers	\$403.30

The Lion King

\$377.80

Star Wars: Episode I - The Phantom Menace

\$359.50

The Dark Knight

\$348.30

The Hunger Games

\$330.00

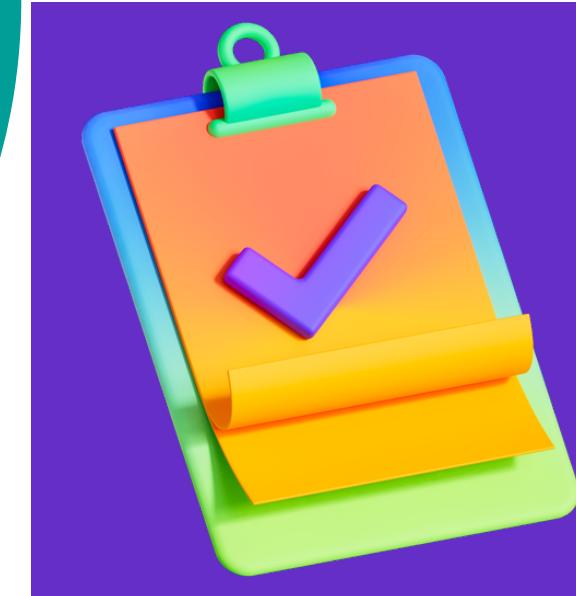


Movies with highest profit





03. IMDb Top 250 !



- I create a new column called IMDb_Top_250 and fill it with the top 250 films having the highest IMDb Rating (equivalent to the column: imdb_score). Also, verify that the num_voted_users for all of these films is greater than 25,000. Add a Rank column with the numbers 1 to 250 showing the positions of the related films.
- Extract all the non-English-language films from the IMDb_Top_250 column and store them in the Top_Foreign_Lang_Film column.

Click on the links below to view the CSV files.

- [IMDB_Top_250.csv](#)
- [Top_Foreign_Language_Film.csv](#)





04. Best Directors !



director_name

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

8.43

Christopher Nolan

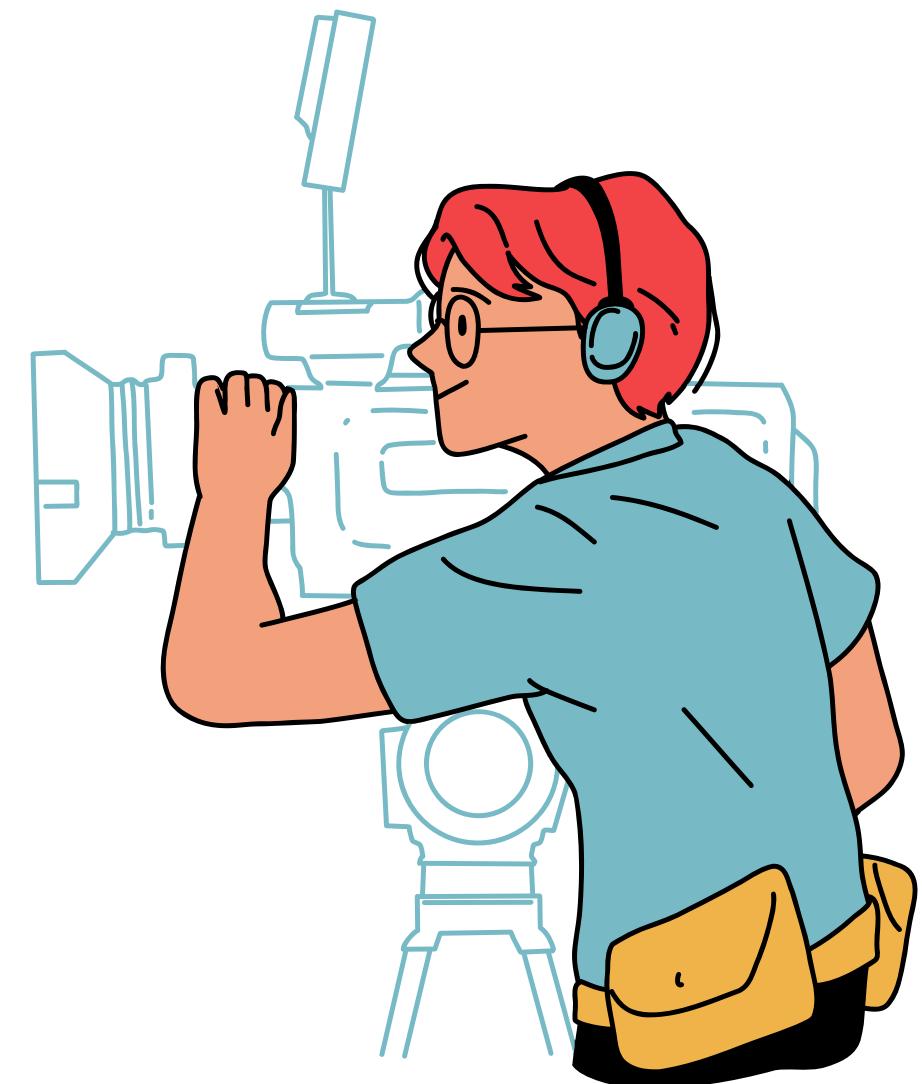
8.42

Asghar Farhadi

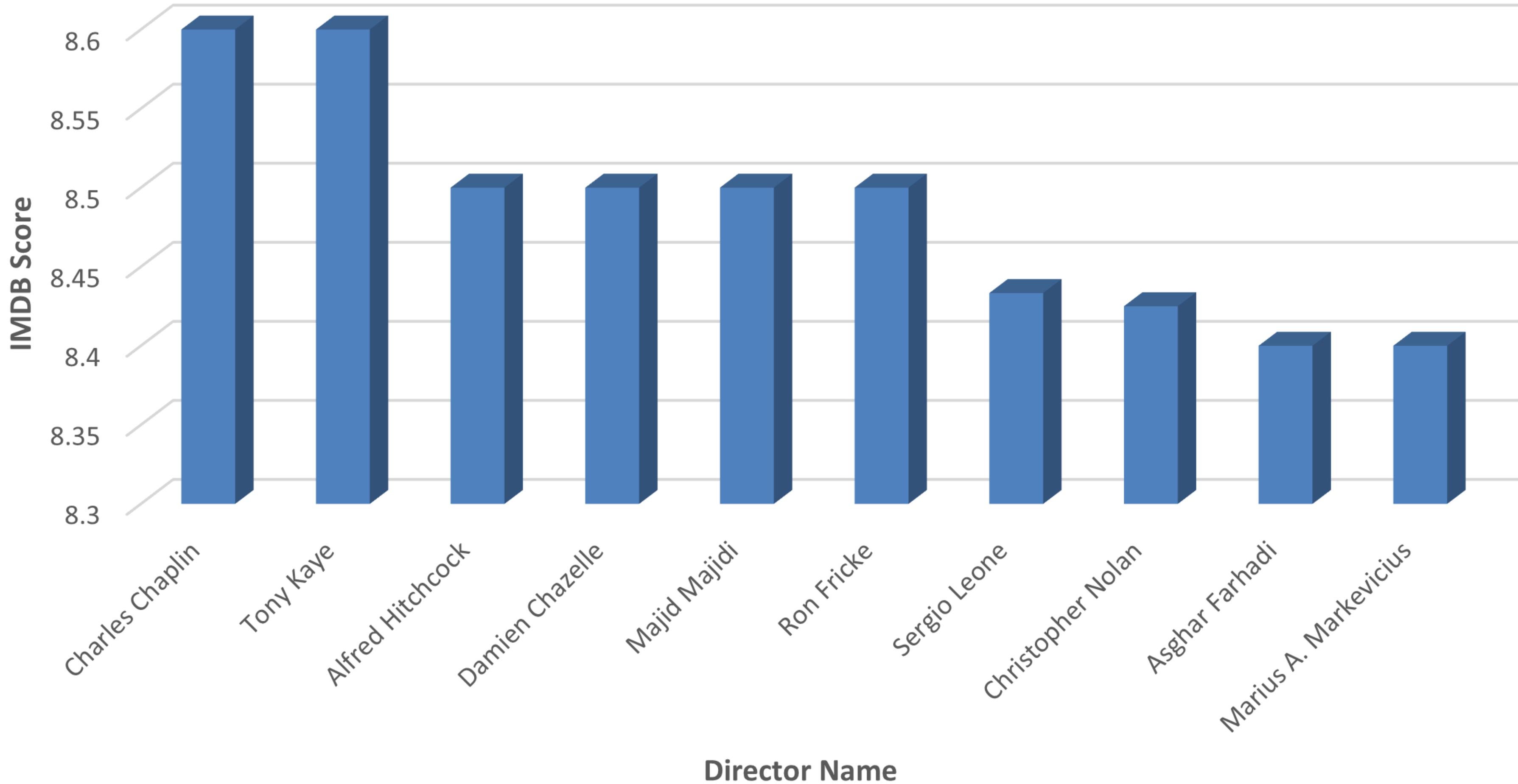
8.4

Marius A. Markevicius

8.4

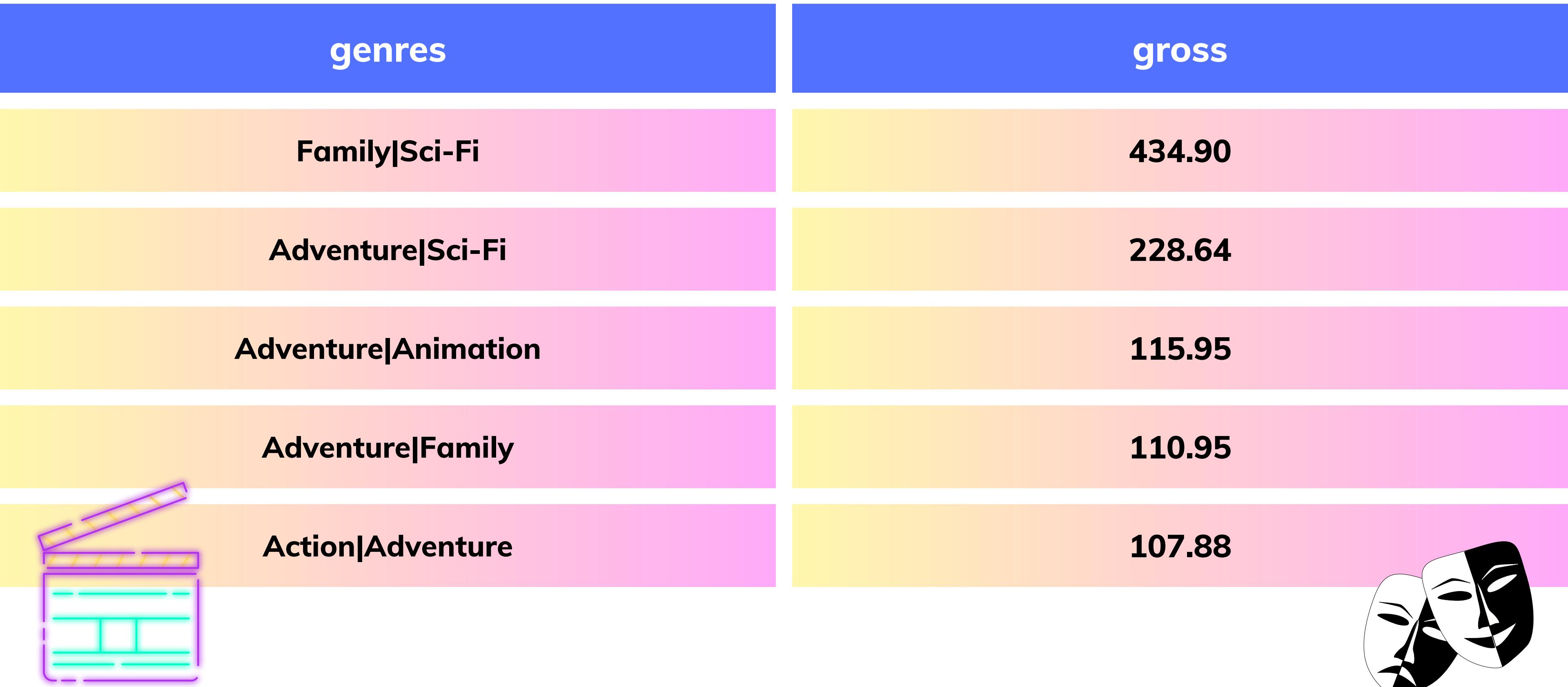
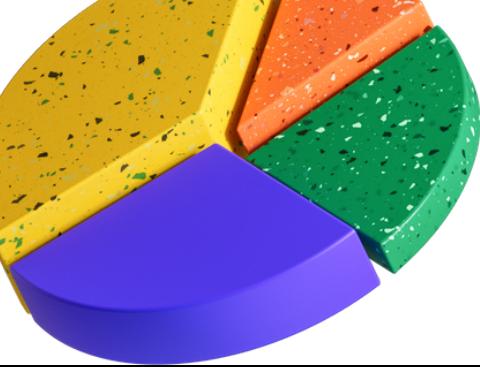


Best Directors

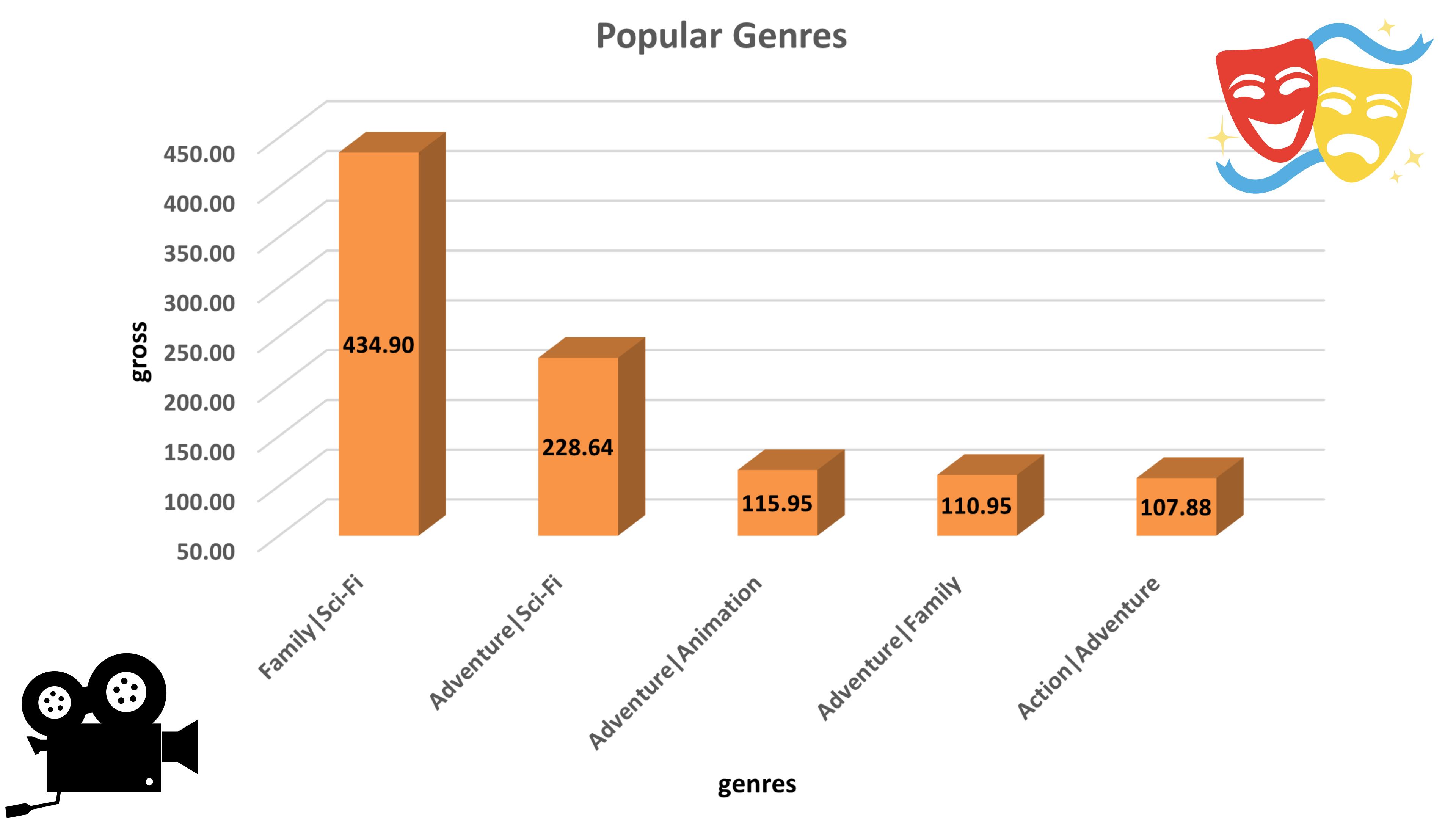




05. Popular Genres !

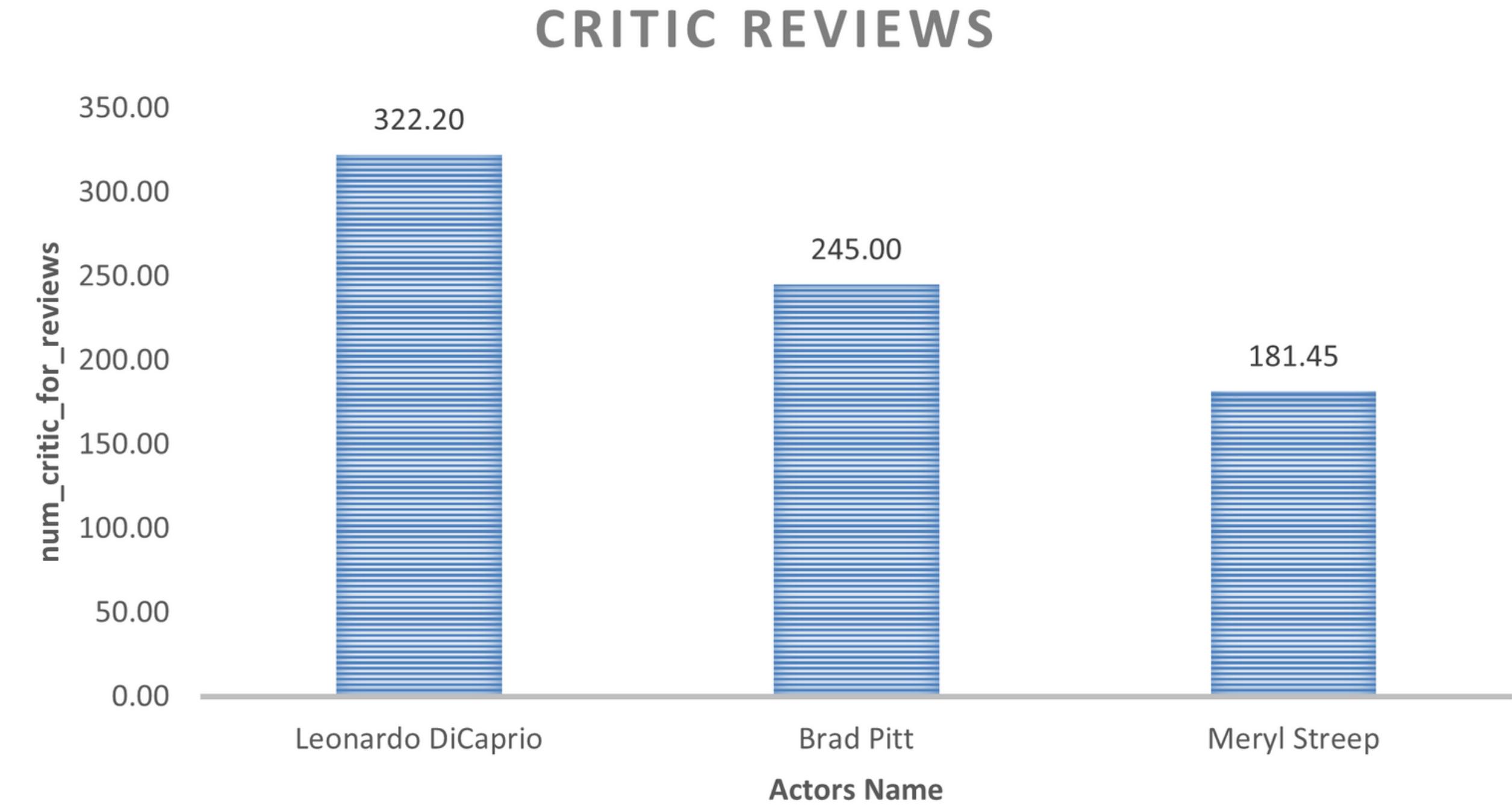
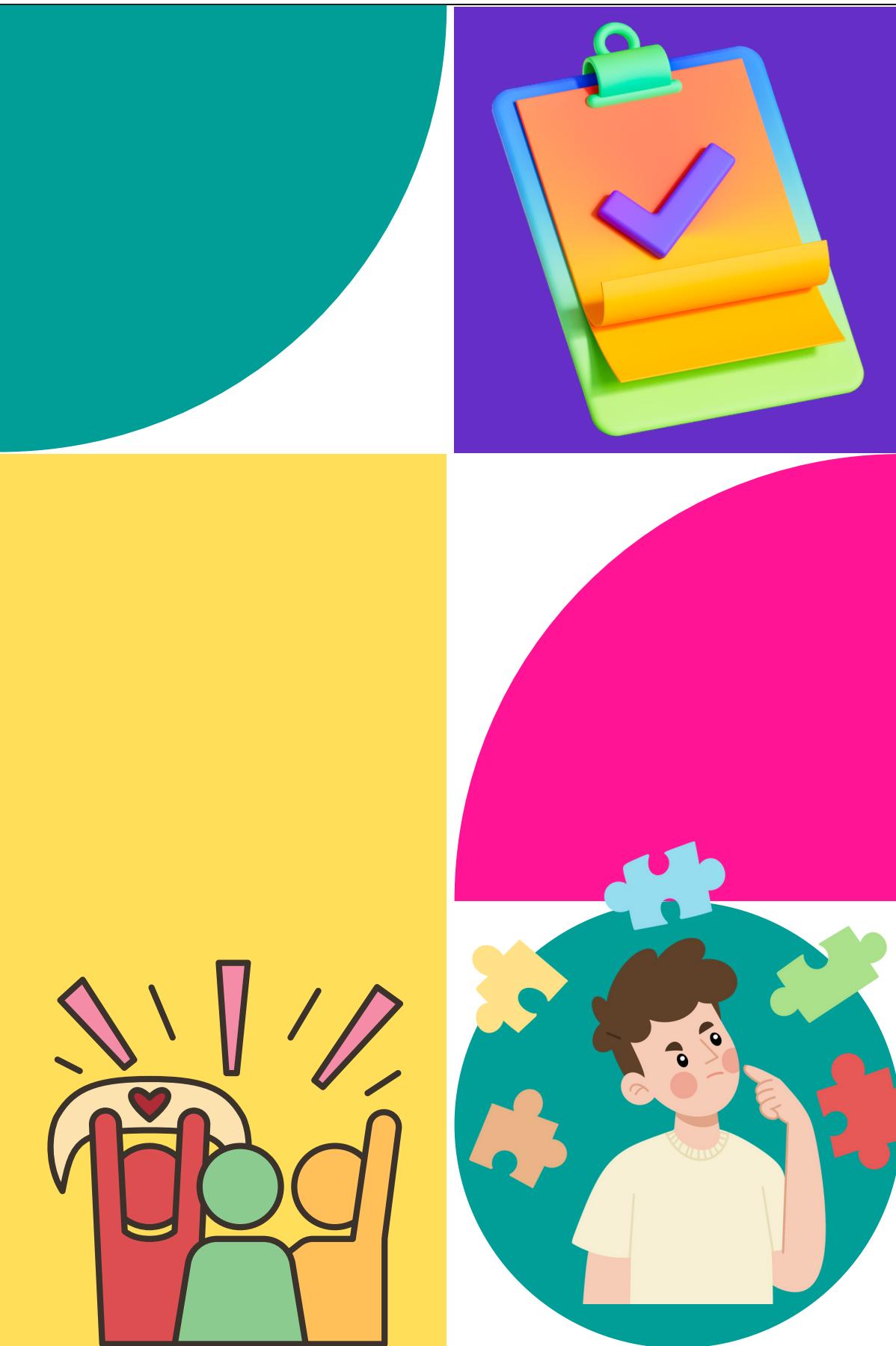


Popular Genres

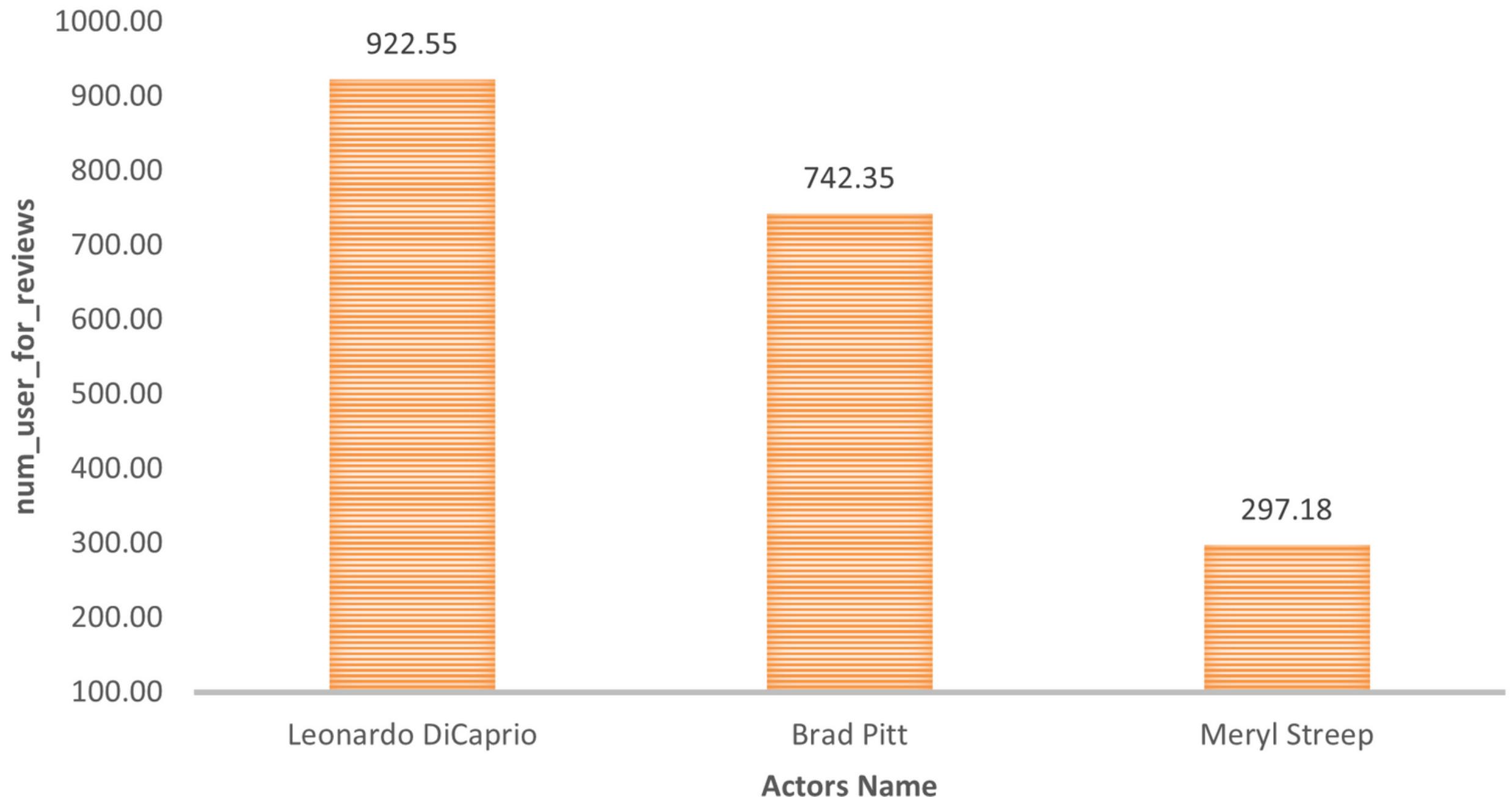




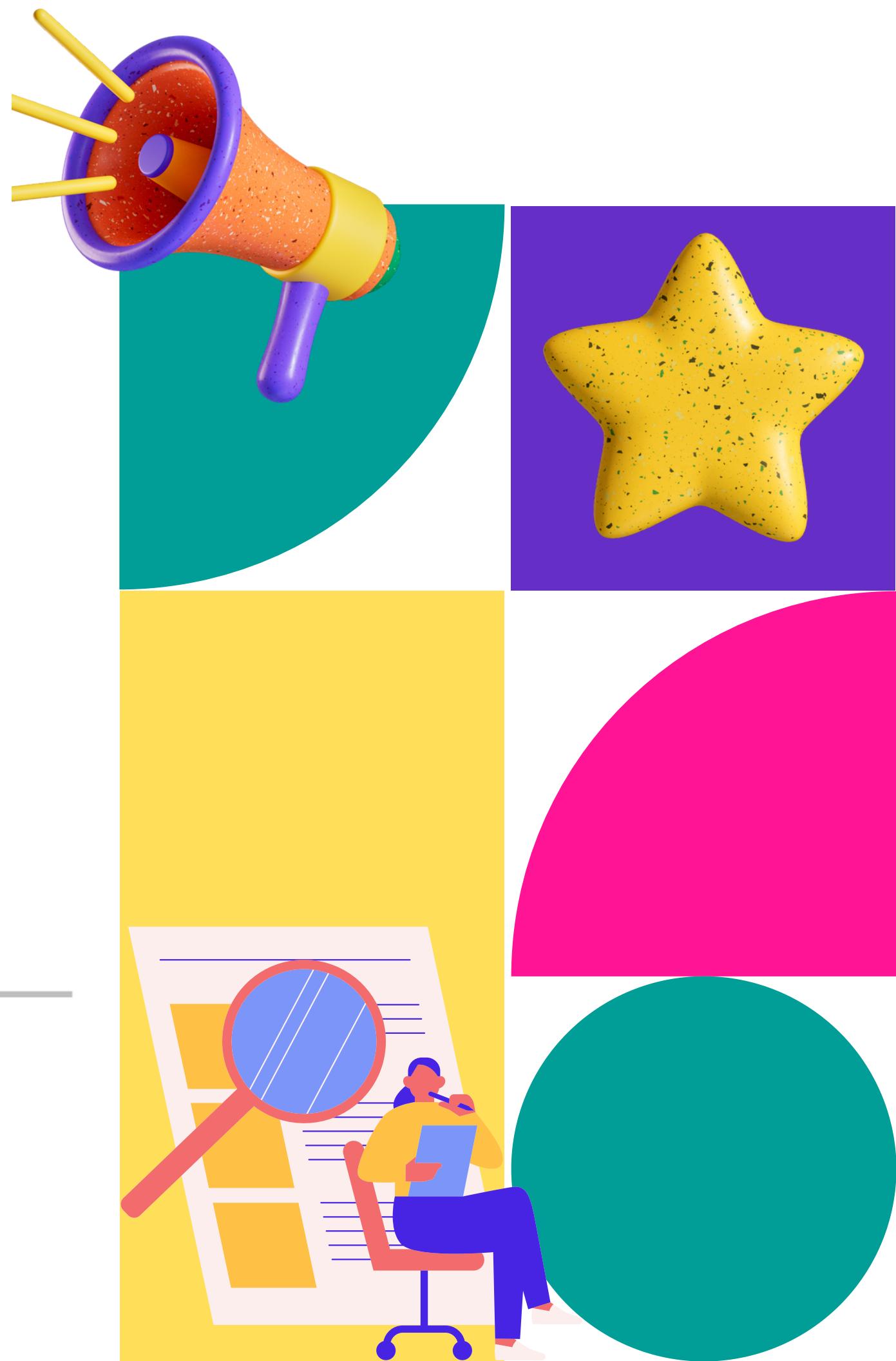
06. Critic-favourite and Audience-favourite Actors !



AUDIENCE_REVIEWS

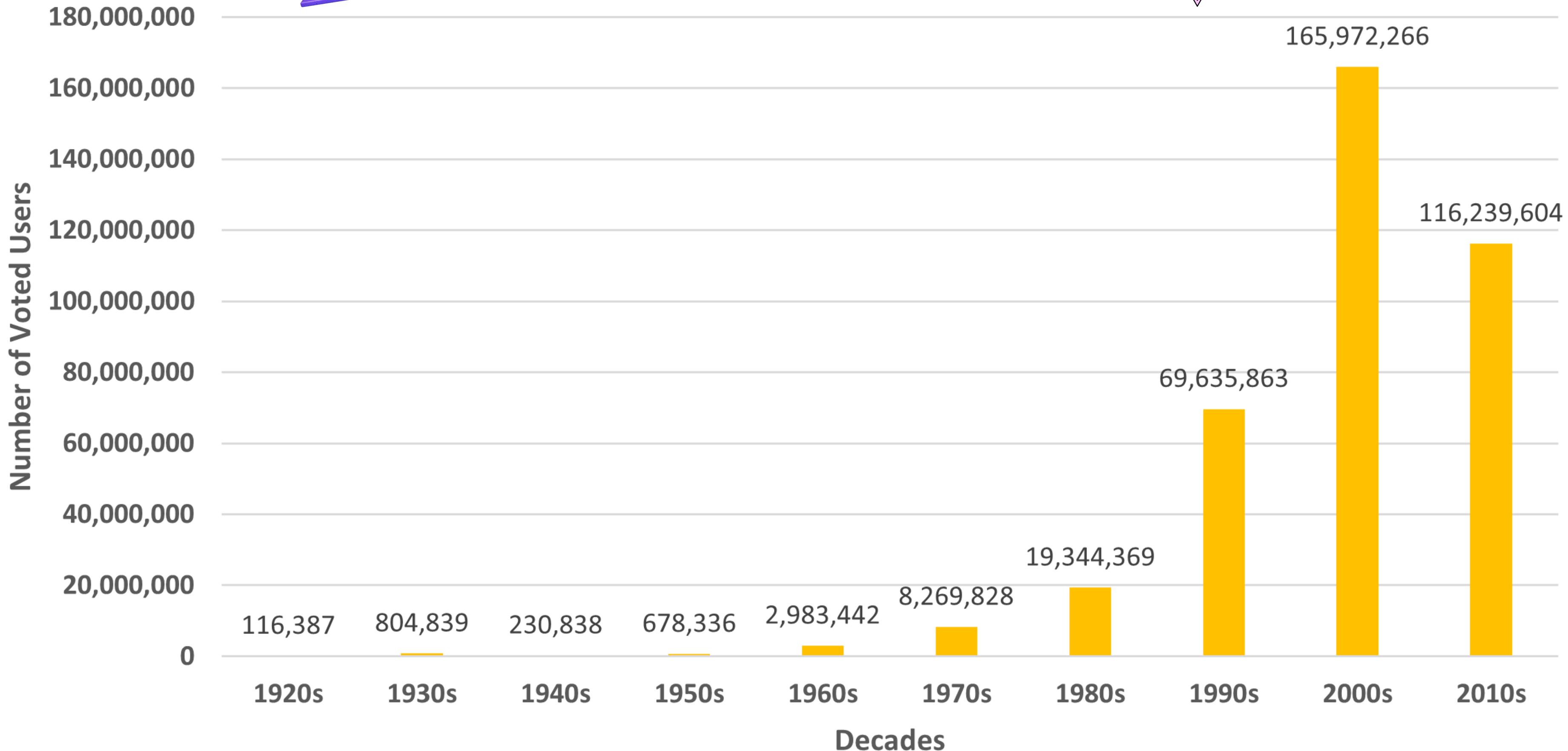


- Source File :- [Combined.csv](#)





Number of Voted Users Over Decades





Result



The result of this project is an in-depth analysis that includes the findings and insights. The report will be provided to the leadership team and will help them to make informed choices concerning the movie industry.

To make the insights more clear and visually appealing, the report will feature a variety of charts, graphs, and tables. Overall, this research will help us better understand the film business as well as the trends and patterns in the data.



Thank You

