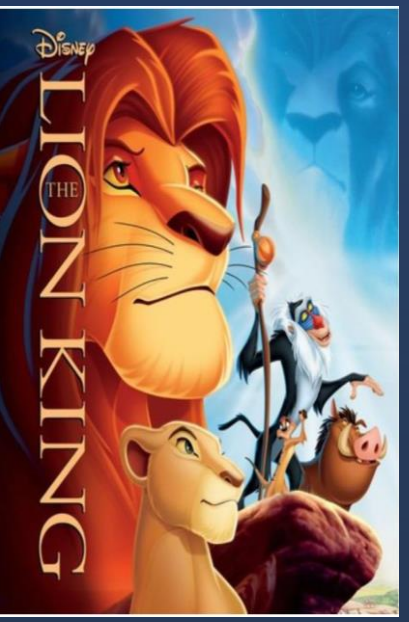
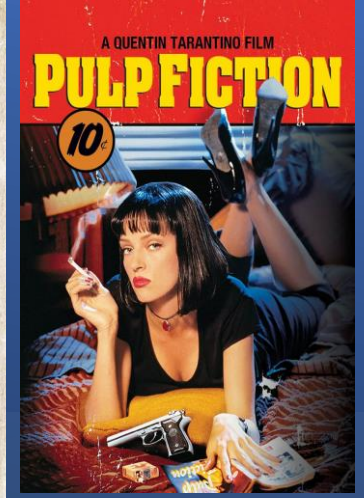




IMDb MOVIE ANALYSIS

By HARSHA PURANIK

Project Description

The project aims to carry out an in-depth analysis of the IMDb movies dataset which contains movies starting from the early 1930s up until the 2010s.

In this project we shall seek to answer the following questions:

- Cleaning the dataset
- Movies with the highest profits
- Decade-wise Profits
- Top 250 IMDb movies of all time
- Top foreign language (non-English) films
- Best directors
- Popular movie genres
- Critic-favourite & audience-favourite actors
- Votes over each decade

Tech-Stack Used

Microsoft Excel

Microsoft PowerPoint

Working File

https://docs.google.com/spreadsheets/d/1FG-wmKO1KrbWyJrC2FV1XdpoVqZVPsdf/edit?usp=sharing&ouid=105088911601746010535&rtpof=true&sd=true

Approach

- The first step, as with any data analysis project, is to assess what type of data we have, and what we really need.
- Then, let's convert the data entries into a table before cleaning it by removing missing values, unnecessary columns and outliers, if any.
- Once the dataset is ready, answer every question with appropriate explanation and visualizations.

A.

Cleaning the dataset

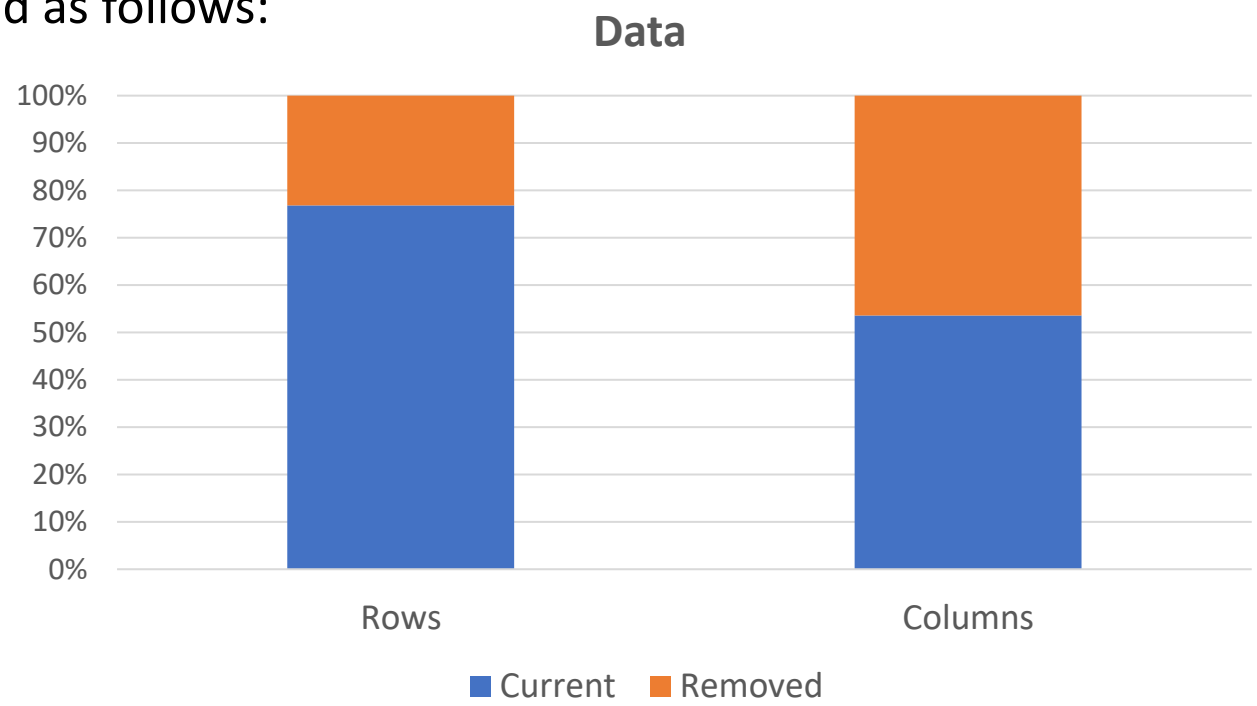
For parts A-C & F, data will be cleaned as follows:

Initially, there were 5044 rows and 28 columns.

Unwanted columns such as Facebook likes of actors and directors, movie links, keywords, duration, etc. were removed.

Rows with missing values were deleted. (about 22%)

The resulting table has **3877 rows** & **15 columns**.



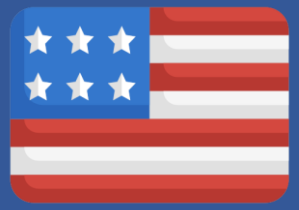
**For parts D & E, the original dataset (uncleaned) will be used

B.

Launched in 2009, **Avatar** is still the highest grossing ($760M) and the most profitable ($523M) movie of all time. It had the 29th largest budget.

Jurassic World (2015) ranks 2nd with $502M.

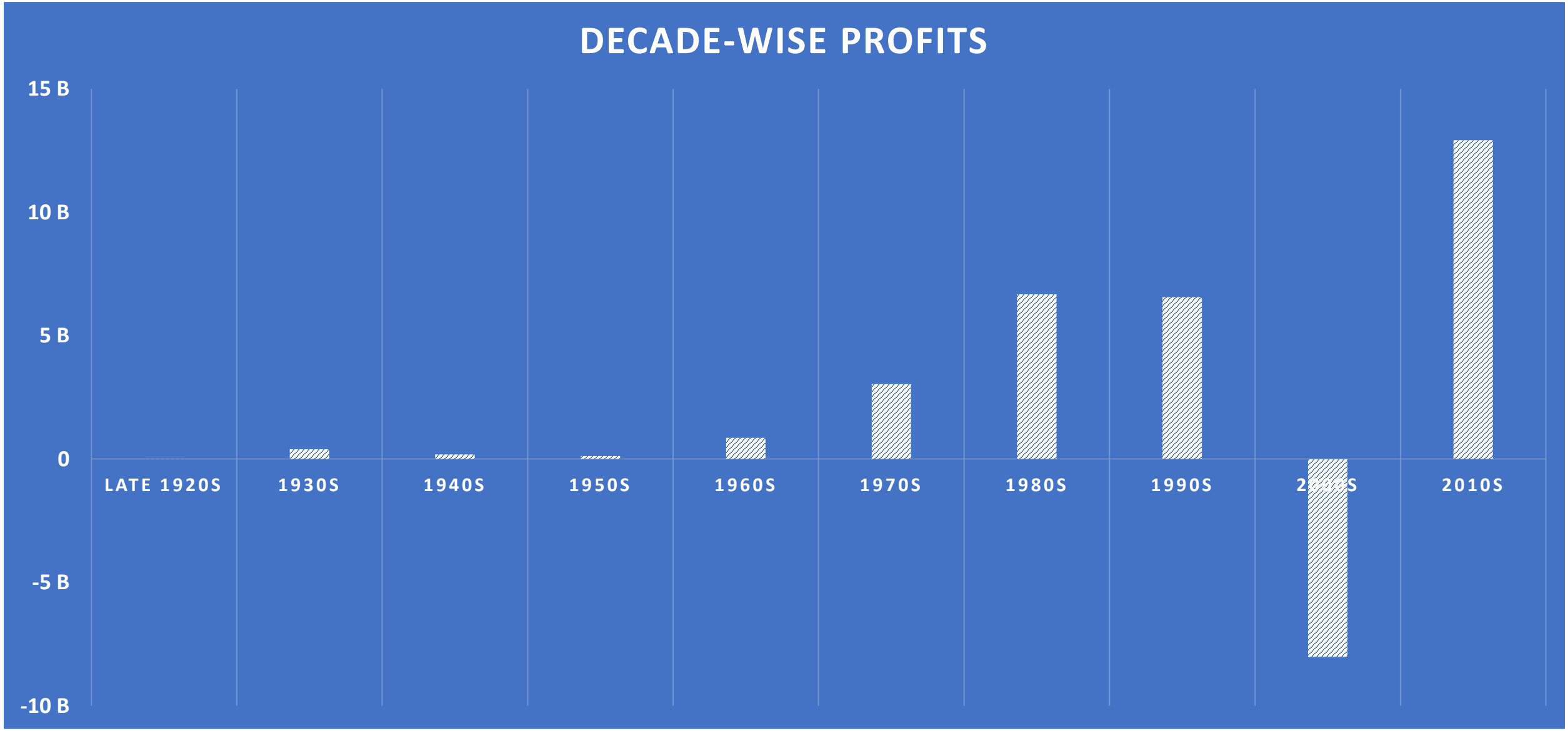
91 of the top 100 profitable movies have been produced in the **USA**.



English holds a 100% record in the top 100 most profitable movies.

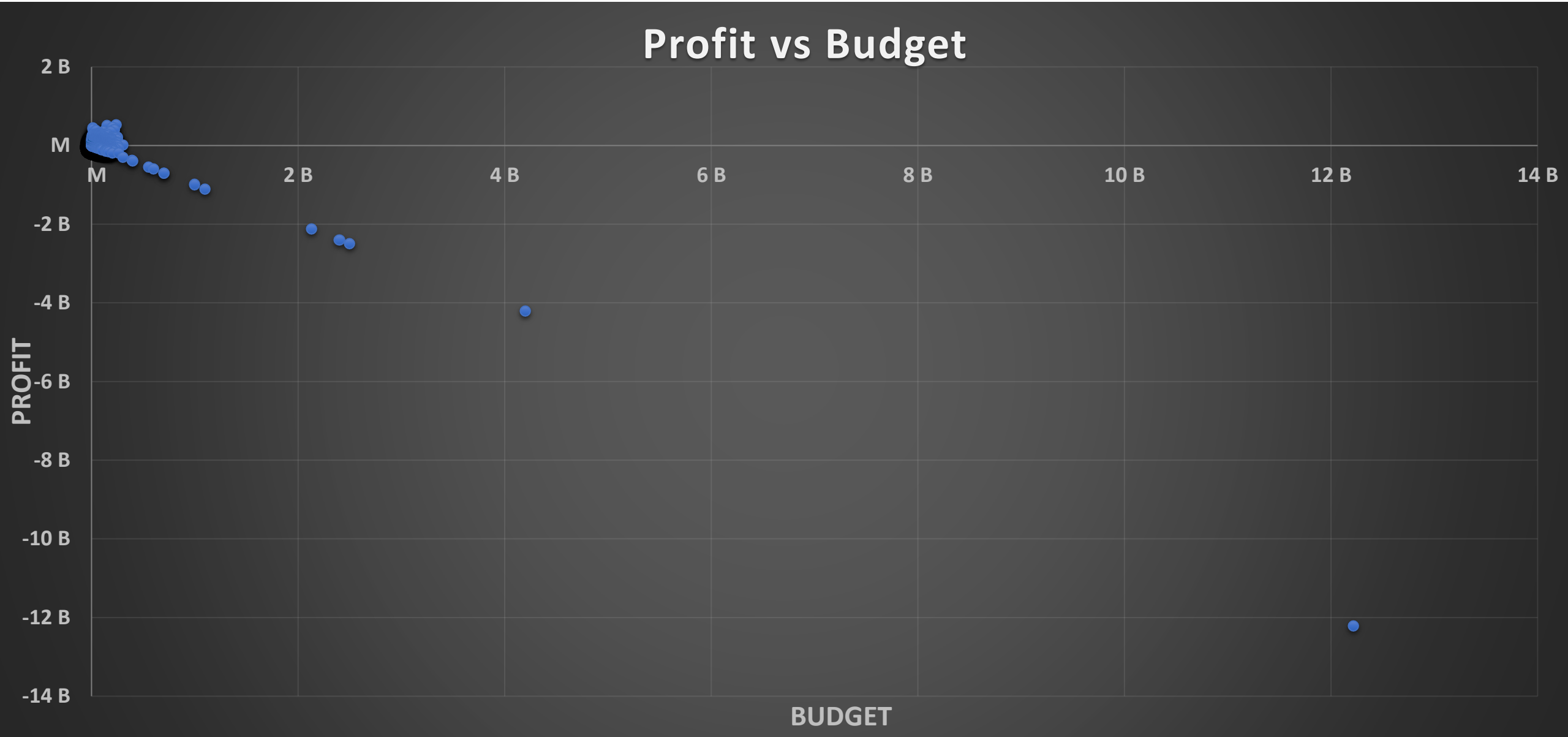
CCH Pounder, Bryce Howard and Leonardo DiCaprio have starred in the top 3 most profitable movies in film history.

B. Now, let's look at the total profits by the global film industry each decade



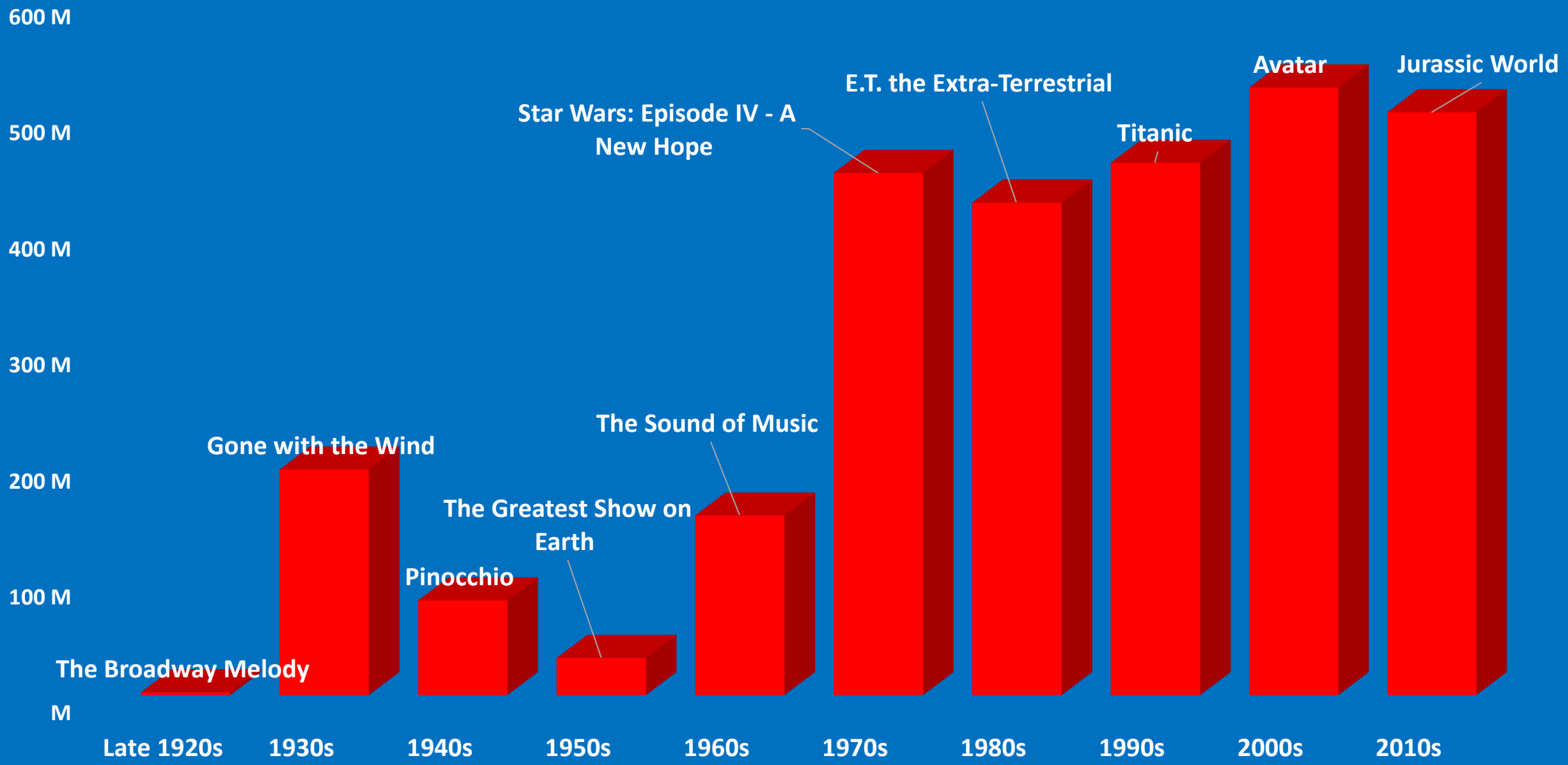
The 2010s suffered from lowest profits even though the most profitable film (Avatar) was produced then.

B. **Observe that there is one major outlier with a highly negative profit towards the bottom right of the chart. The other (minor) outlier is around the point (4B, -4B)**



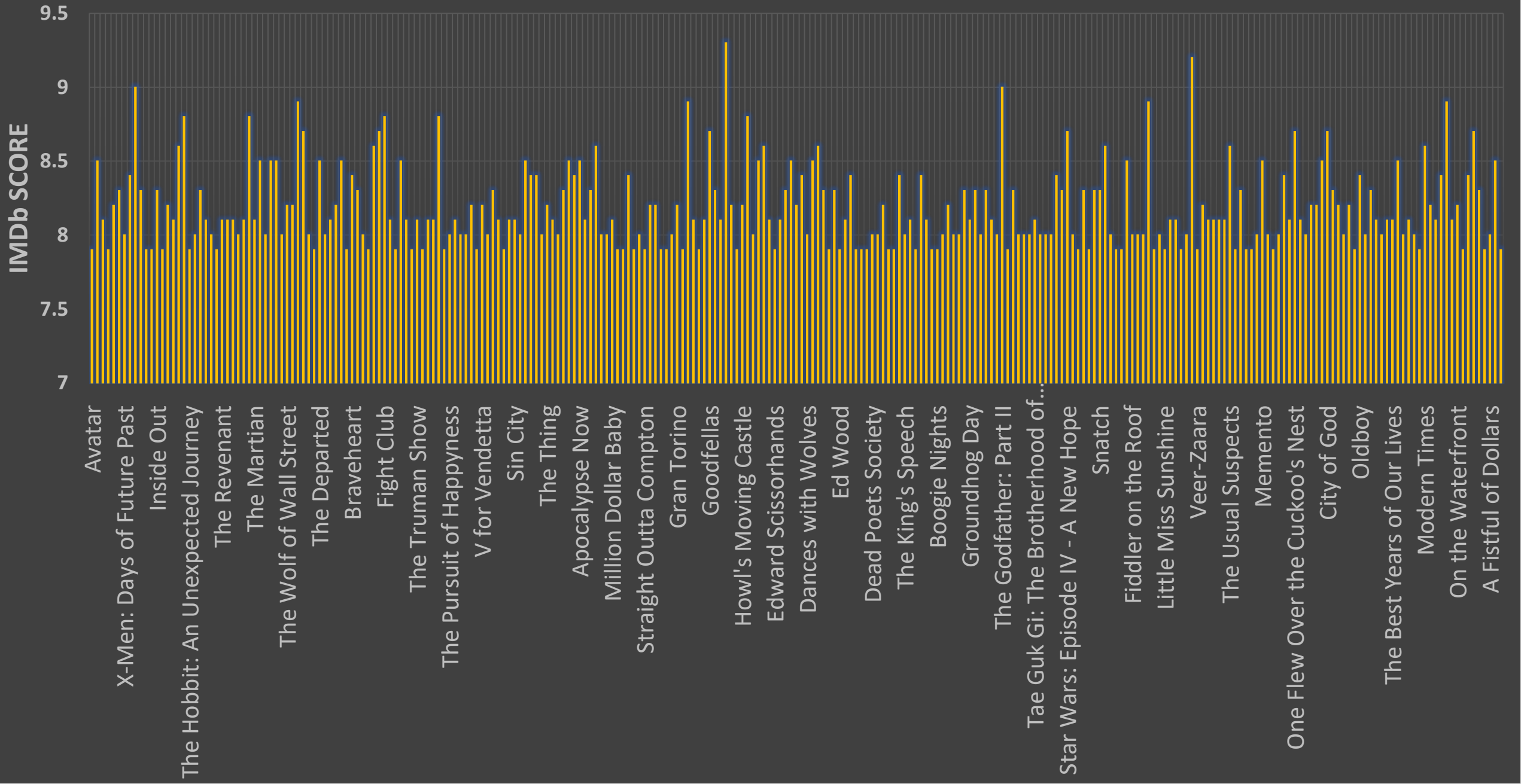
B.

MOST PROFITABLE MOVIES EACH DECADE



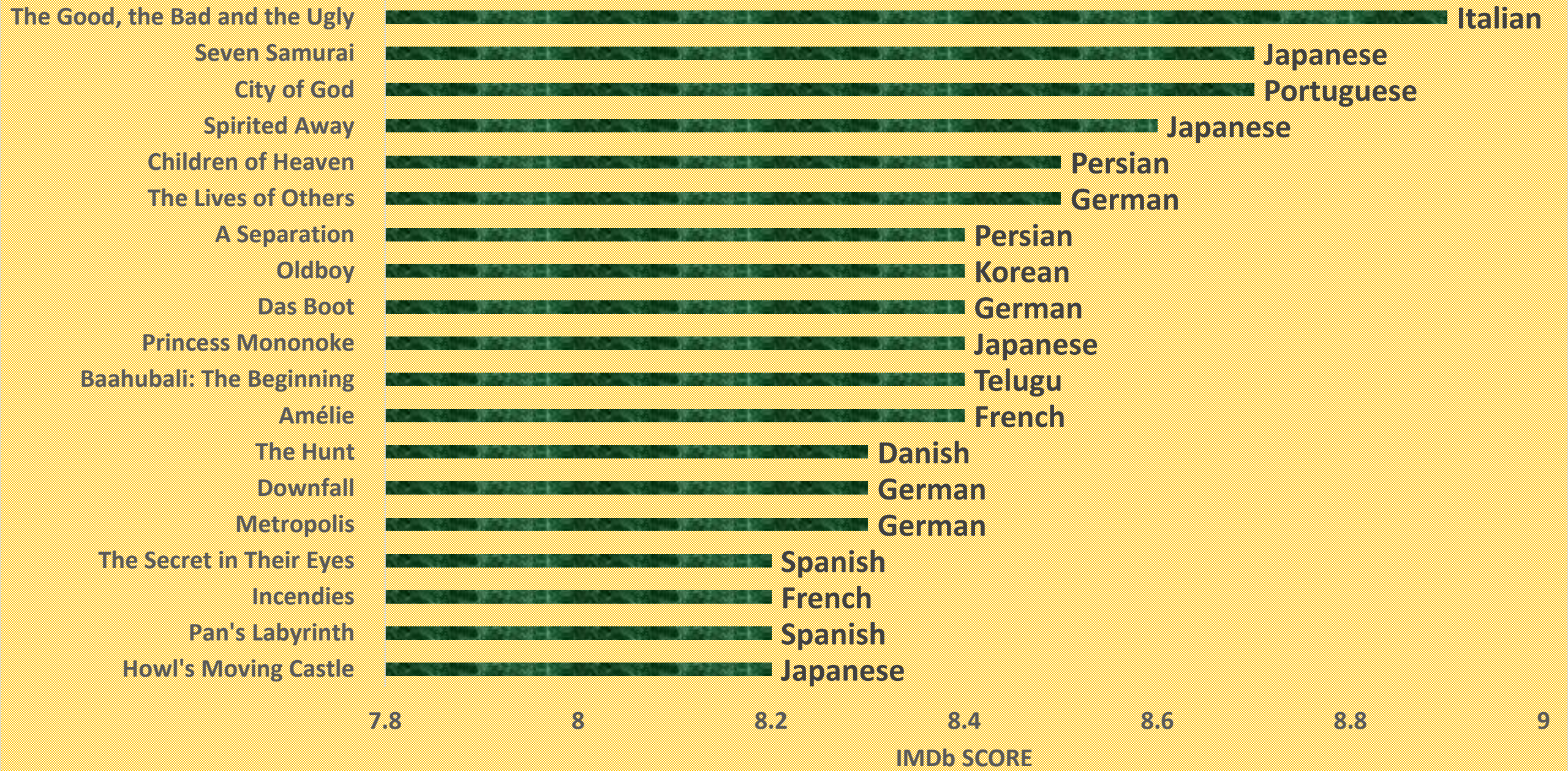
C.

Top 250 Movies on IMDb



C.

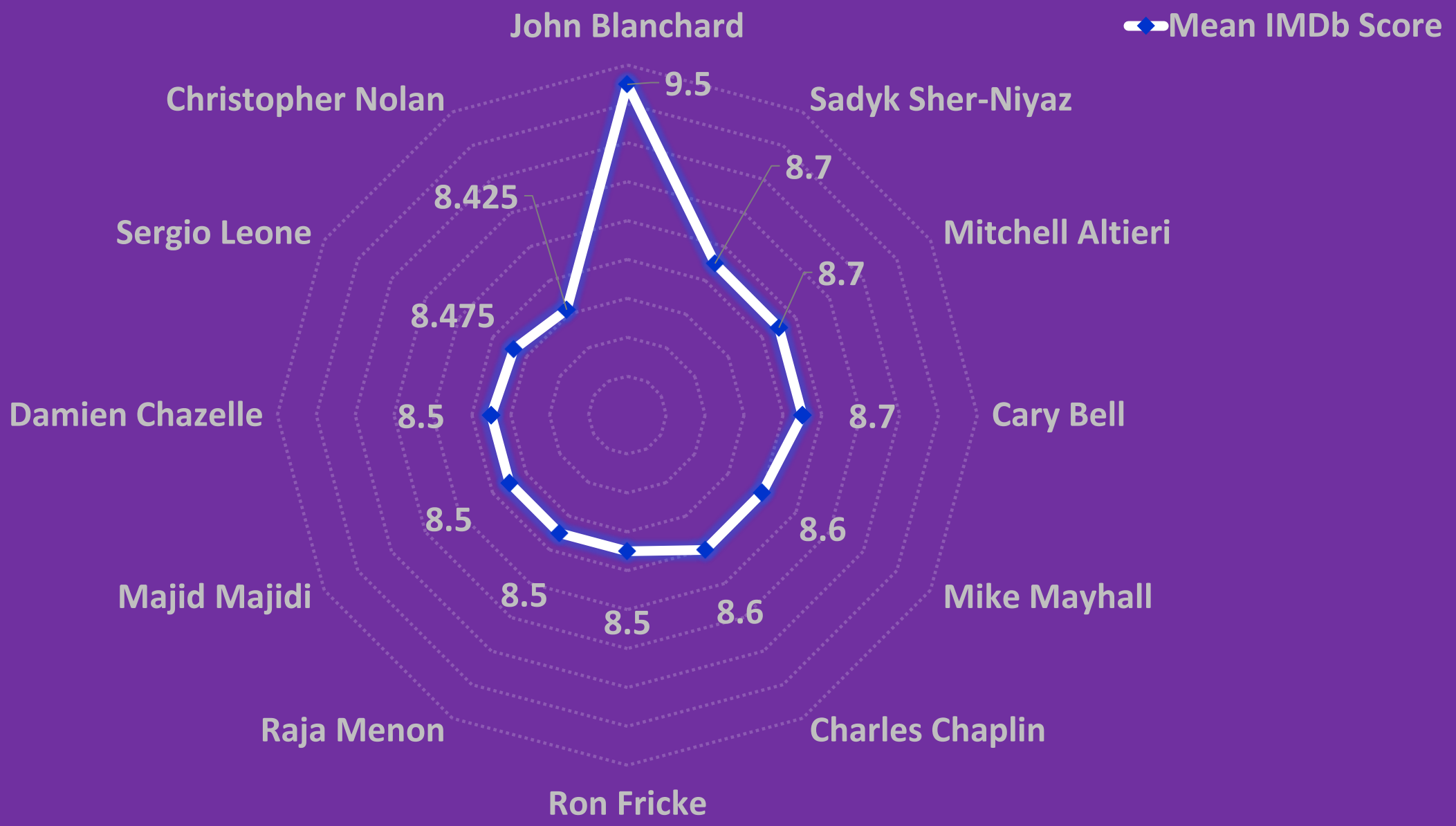
Top 20 Foreign Language Films



D.

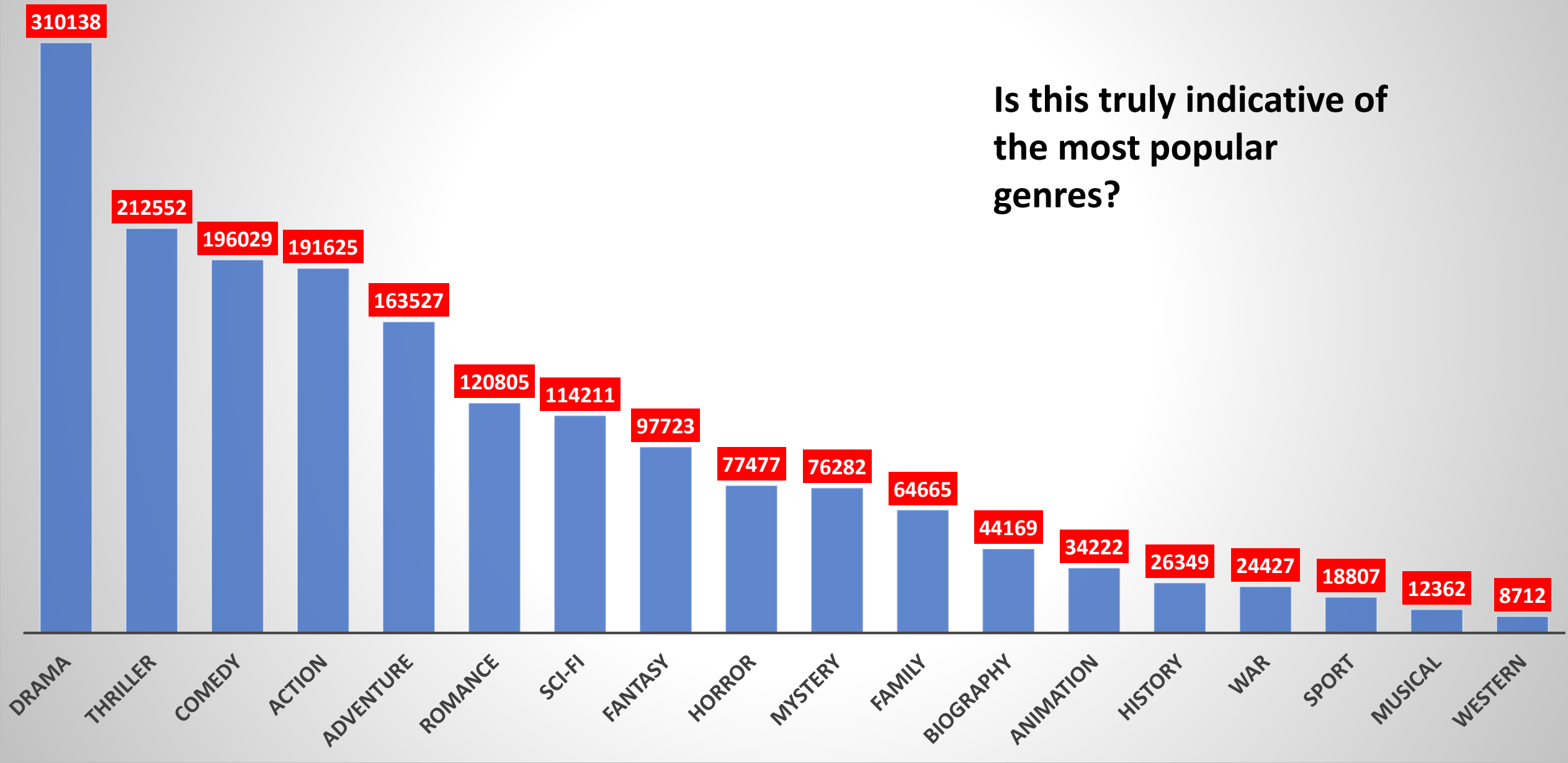
Best Directors

Best Directors



E.

User Reviews for each Genre



Is this truly indicative of the most popular genres?

E.

Popular Genres

THE QUESTION OF HOW TO CALCULATE THE MOST POPULAR MOVIE GENRE IS INTRIGUING. SHOULD FACEBOOK LIKES OR IMDB PLATFORM REVIEWS BE CONSIDERED? HOW INSIGHTFUL WILL EITHER ONE TURN OUT TO BE? DO CRITICS HAVE ANY ROLE TO PLAY IN THIS?

IN THE PREVIOUS SLIDE, WE LOOKED AT THE TOTAL NUMBER OF USER REVIEWS FOR EACH GENRE. WE OBSERVED THAT *DRAMA* & *THRILLER* HAD THE HIGHEST NUMBER OF CUMULATIVE REVIEWS AND HENCE, SEEM TO BE THE MOST POPULAR. ON THE OTHER HAND, *WESTERN* & *SPORT* WERE AMONG THE LESS PREFERRED ONES.

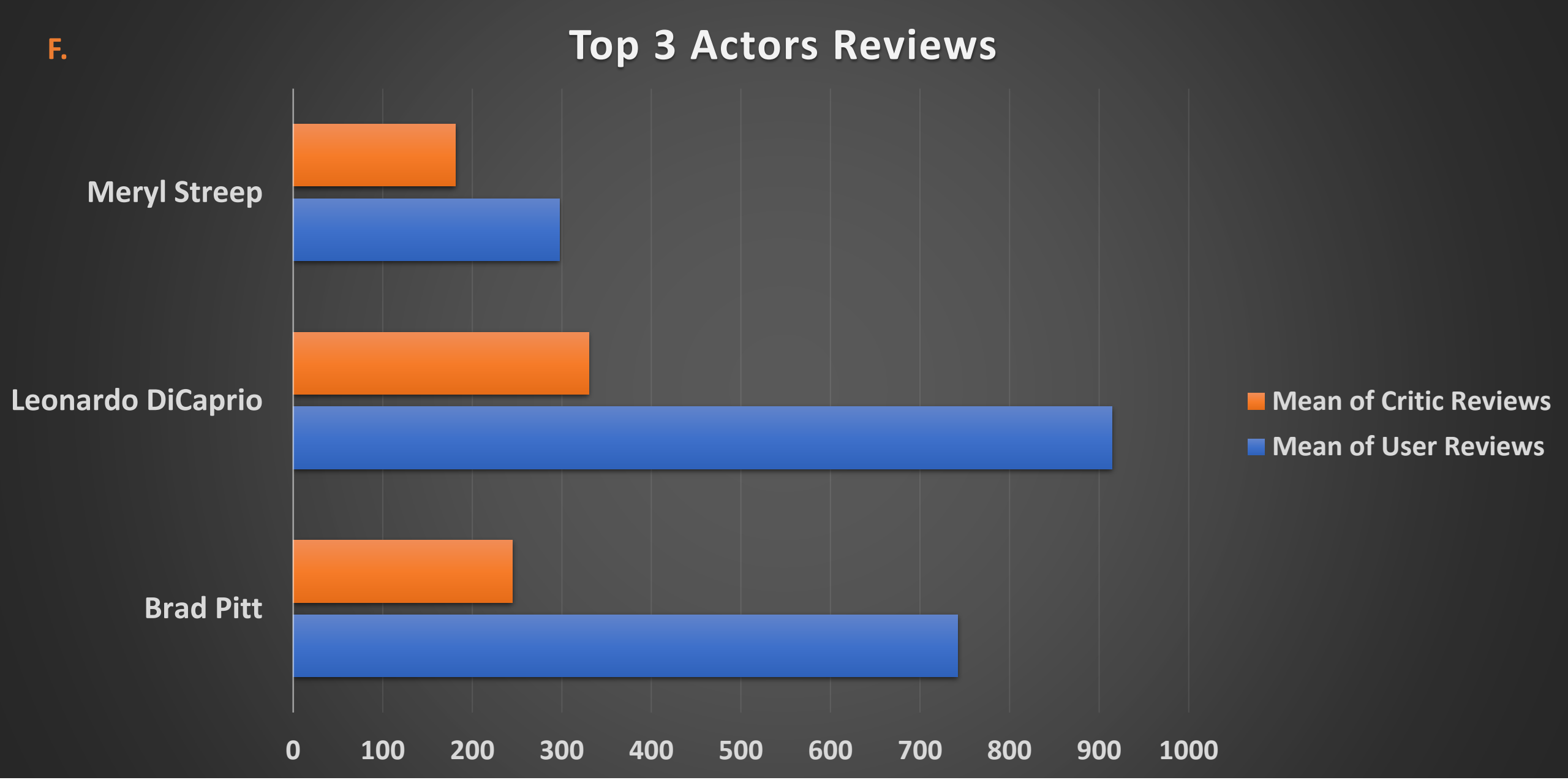
HOWEVER, THIS DATASET COULD BE BIASED AND MIGHT HAVE AN UNEVEN DISTRIBUTION OF MOVIES ACROSS THE CATEGORIES. THEREFORE, IT BECOMES MORE USEFUL TO LOOK AT THE *AVERAGE NUMBER OF REVIEWS PER MOVIE* FOR EACH GENRE. WE SHALL DO THIS IN THE NEXT SLIDE.

E.

Average Reviews *per Movie* for each Genre

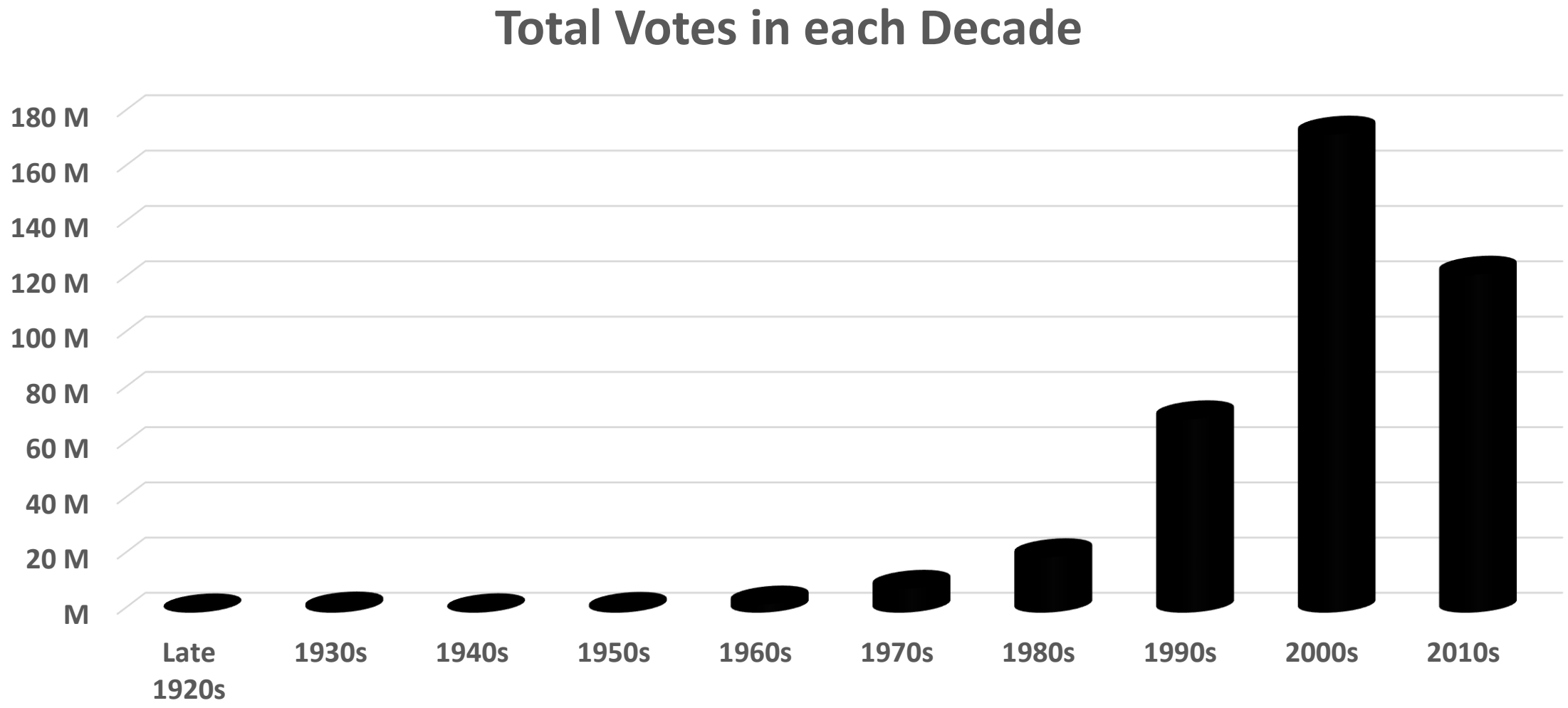
SCI-FI	228.0
ADVENTURE	206.5
ACTION	197.8
MYSTERY	196.6
HORROR	195.2
FANTASY	189.8
THRILLER	188.3
BIOGRAPHY	181.0
ANIMATION	172.8
HISTORY	171.1
DRAMA	158.6
WAR	150.8
WESTERN	145.2
FAMILY	143.4
ROMANCE	136.2
COMEDY	129.7
SPORT	123.7
MUSICAL	120.0

F.



Leonardo DiCaprio stands top in both categories: the average number of user reviews and the average number of critic reviews.

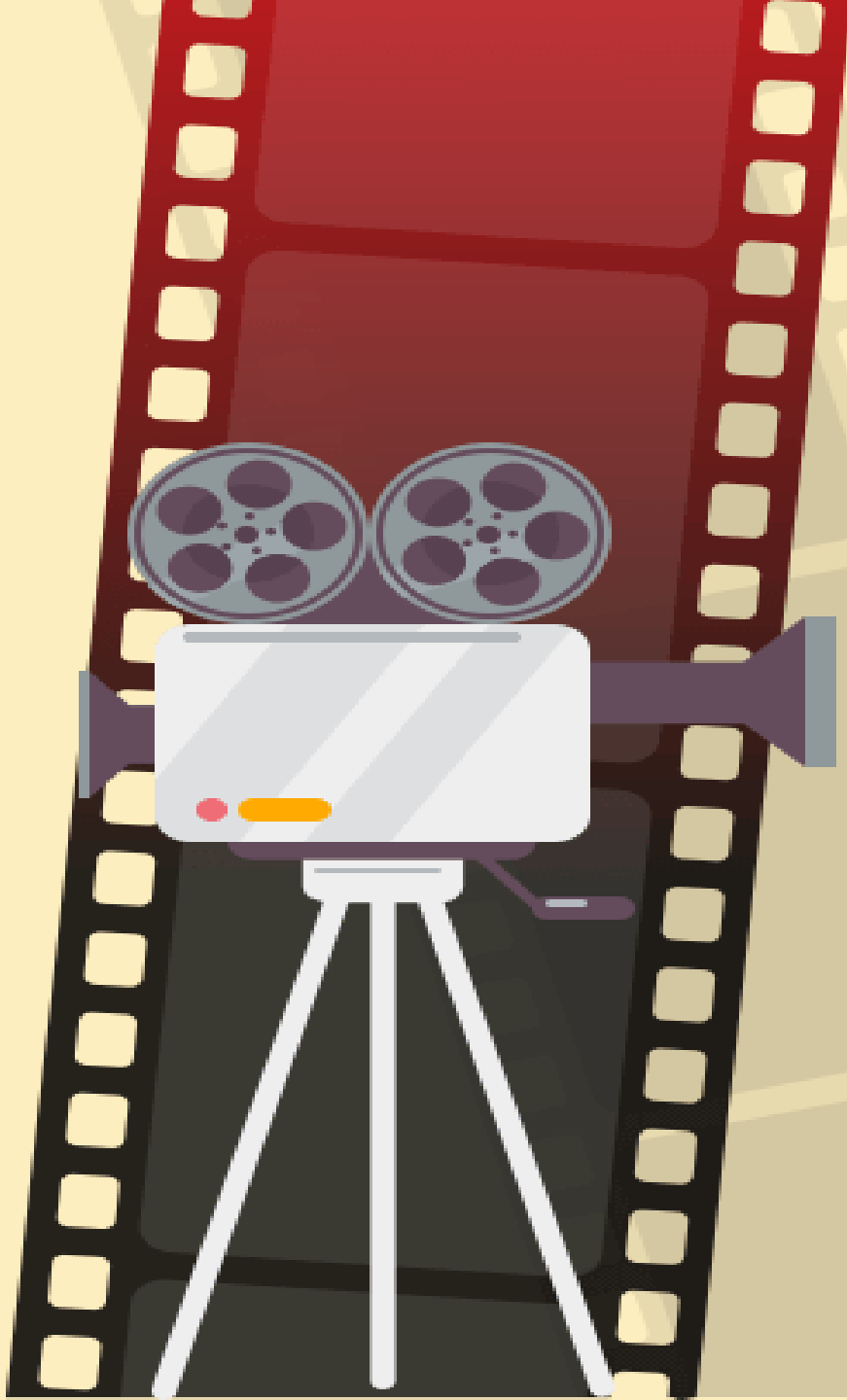
F.



We can observe an exponential growth in the total votes until the 2000s where it reaches a record high of 172M votes in total. Then, there is a sudden plummet to 120M votes in the 2010s.

Summary

- In this data analysis project, we examined the IMDb movie dataset to gain insights into various aspects of the film industry.
- The analysis focused on identifying the best actors, directors, and movies across different decades, as well as exploring the languages in which the top movies were produced. Additionally, we looked at and analyzed user and critic reviews for each movie on the IMDb platform and studied the variation of votes over the course of each decade.
- Overall, this data analysis project using Excel and its functions enabled a comprehensive examination of the IMDb movie dataset. By leveraging sorting, filtering, counting, and calculation functions, valuable insights were gained regarding the various aspects of the IMDb movie dataset.



Thank you