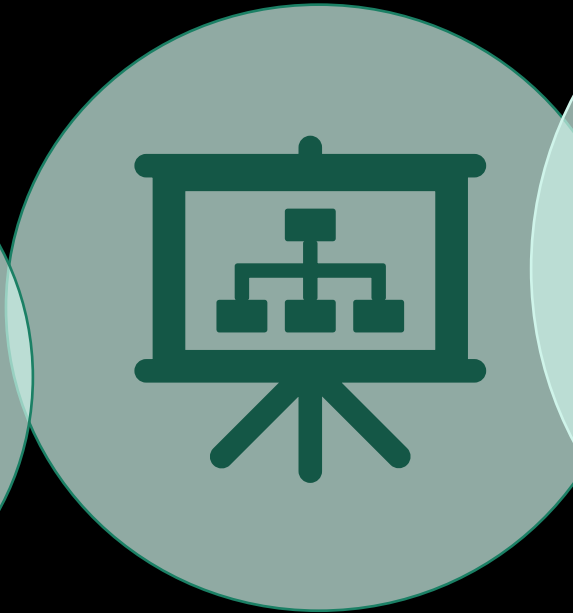
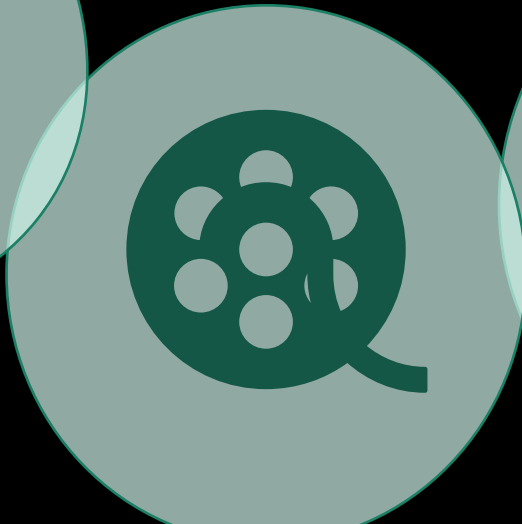
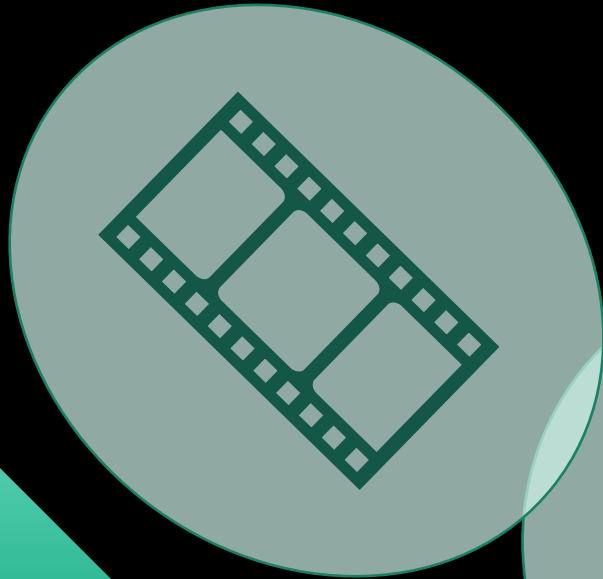


# IMDB MOVIE ANALYSIS

-By Pilli Bhavana



- This task requires defining a problem to be solved using a dataset of IMDb Movies.
- The most important idea is to frame the problem by asking 'What?' and using Data Analysis skills to explore the data set and derive insights.
- After defining the problem, using 5 Whys Analysis to create a report which conveys a data story.

## PROJECT DESCRIPTION

- A. Cleaning the data
- B. Find the movies with the highest profit?
- C. Find IMDB Top 250
- D. Find the best directors
- E. Find popular genres
- F. Find the critic-favorite and audience-favorite actors

# Task

- Data Collection: The project started by downloading the data from the provided source in .xlsx format. The project description was carefully studied to understand the project requirements and key factors to focus on.
- Data Cleaning: The initial step in the data analysis process was data cleaning. This involved removing unnecessary columns, handling outliers, arranging columns, and performing other necessary data cleaning operations to ensure the data was suitable for analysis.
- Data Transformation: A new column called "Profit" was added, which represented the difference between gross and budget. This additional column provided valuable insights into the profitability of each movie.

# APPROACH

- Data Analysis:
  - a) Top 10 Movies by Profit: A pivot table was created using the “Movie Name” and “Profit” columns to identify the top 10 movies with the highest profits.
  - b) Top 250 English and Other Language Movies: Two separate tables were created to showcase the top 250 English movies and top movies in other languages based on their IMDb scores.
  - c) Top 10 Directors: A table was created to display the top 10 directors based on the average IMDb scores of their movies.
  - d) Popular Genres: A table was generated to identify the top 10 popular genres using the data from the top 250 English movies and top movies in other languages.
  - e) Movies of Specific Actors: A table was created to list the movies featuring actors such as “Meryl Streep”, “Leonardo DiCaprio”, and “Brad Pitt”.
  - f) Critic and Audience Favorites: Two separate pivot tables were used to determine the favorite actors based on critic reviews and audience ratings.
  - g) Change in Number of Voted Users Over Decades: A pivot table was generated to analyze the change in the number of voted users over different decades.
  - h) Visualizations: Graphs and charts were created for each table to provide visual representations of the data and facilitate better understanding.
- Project Report: After completing the data analysis, project report was prepared to present the findings, insights, and visualizations obtained from the analysis.



MICROSOFT EXCEL (VERSION:  
MICROSOFT 360): FOR CLEANING DATA,  
ANALYZING DATA, etc.



MICROSOFT POWERPOINT (VERSION:  
MICROSOFT 360): FOR CREATING THE  
REPORT

# TECH- STACK USED

- **Movie Profitability:**
  - “Avatar” is the most profitable movie with a profit of “\$52,350,5847”, followed by “Jurassic World” with a profit of “\$50,217,7271”. These movies demonstrate exceptional financial success in the film industry.
- **Top IMDb Scores:**
  - “The Shawshank Redemption” holds the highest IMDb score among English movies, with a rating of “9.3”. This indicates the film's outstanding quality and its strong appeal to viewers.
  - “The Good, the Bad and the Ugly” stands out as the highest-rated movie in languages other than English, with a score of “8.9”. This suggests its universal appreciation beyond language barriers.

# INSIGHTS

- Best Directors:

- Directors “Tony Kaye” and “Charles Chaplin” have the highest average IMDb scores of 8.6. Their consistent high ratings reflect their talent and influence in the film industry.

- Popular Genres:

- The genres “Drama”, “Crime, Drama” and “Drama, Romance” are popular choices among audiences based on the analysis of the top-rated movies. This highlights the impact of these genres in captivating viewers.

- Lead Actors:

- “Meryl Streep” has acted as the lead in “11 movies”, “Leonardo DiCaprio” has acted as the lead in “21 movies” and “Brad Pitt” has acted as the lead in “17 movies”. Their continued participation in important roles strengthens their status in the film industry.



- Critic and Audience Favorites:

- “Heather Donahue” is identified as the audience's favorite actor, indicating how well-liked and popular she is with viewers.
- “Albert Finney” is recognized as the critic's favorite actor, reflecting his remarkable talent.

- User Voting Trends Over Decades:

- The analysis reveals that the highest number of user votes occurred in the “2000s” decade, with a total of “1,709,025,320” votes. This indicates a significant engagement and participation of users during this period.
- However, the “1920s” decade had the lowest number of votes, with only “116,392” votes. This can be defined as limited availability of movies and the lower user population during that era.

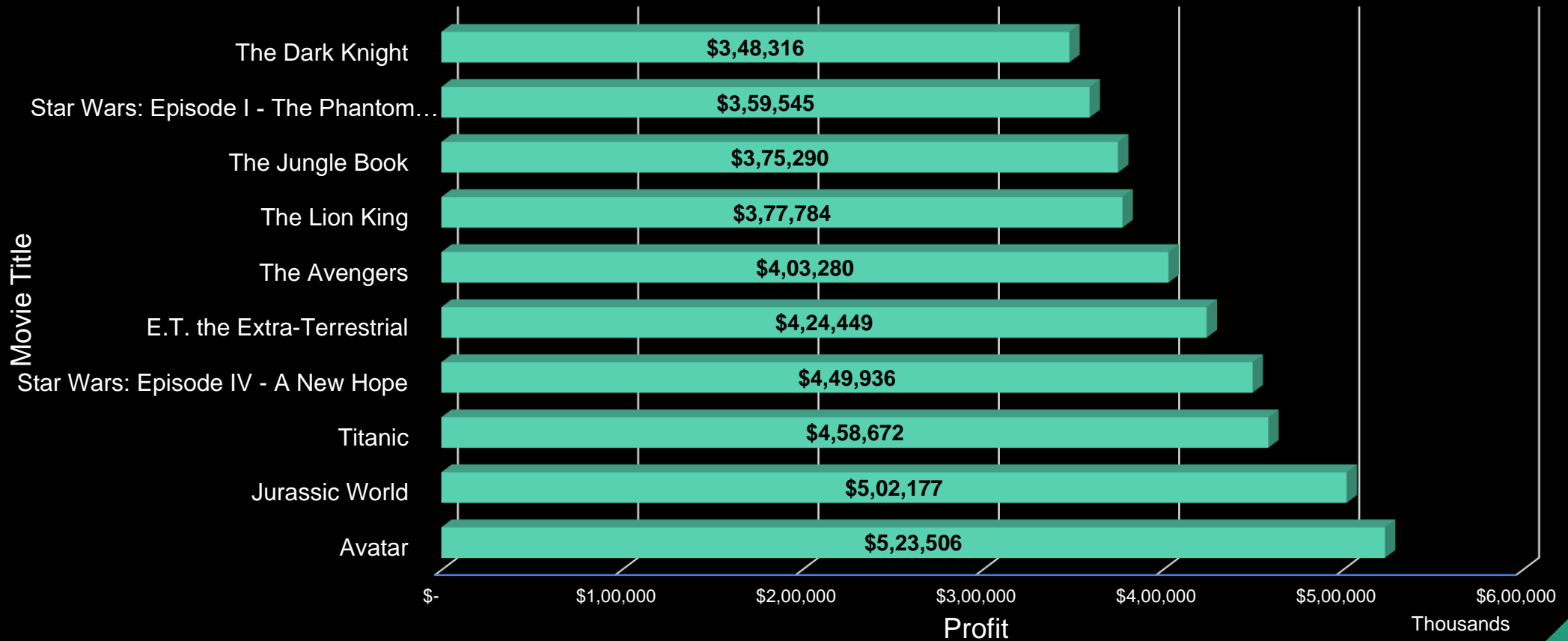
# RESULT

# A. Cleaning The Data

- The data was cleaned by removing:
  - Outliers
  - Duplicates
  - Null Values
- After completely cleaning the dataset, the final dataset with cleaned data was obtained and it was used to do the rest of the analysis.
- The link for the cleaned dataset is in the slide “Drive Link”.

## B. Find the movies with the highest profit?

### Top 10 Movies With Highest Profit



C. Find IMDB Top 250

# C- 1: Top 250 English Movies

Rank	Top_250_English_Movie_Title	IMDb_Score
1	The Shawshank Redemption	9.3
2	The Godfather	9.2
3	The Dark Knight	9
4	The Godfather: Part II	9
5	Pulp Fiction	8.9
6	Schindler's List	8.9
7	The Lord of the Rings: The Return of the King	8.9
8	Fight Club	8.8
9	Forrest Gump	8.8
10	Inception	8.8

**NOTE:** Here only 10 records are shown the remaining records are in the excel workbook and the link for the excel workbook is provided in the slide “**DRIVE LINK**”

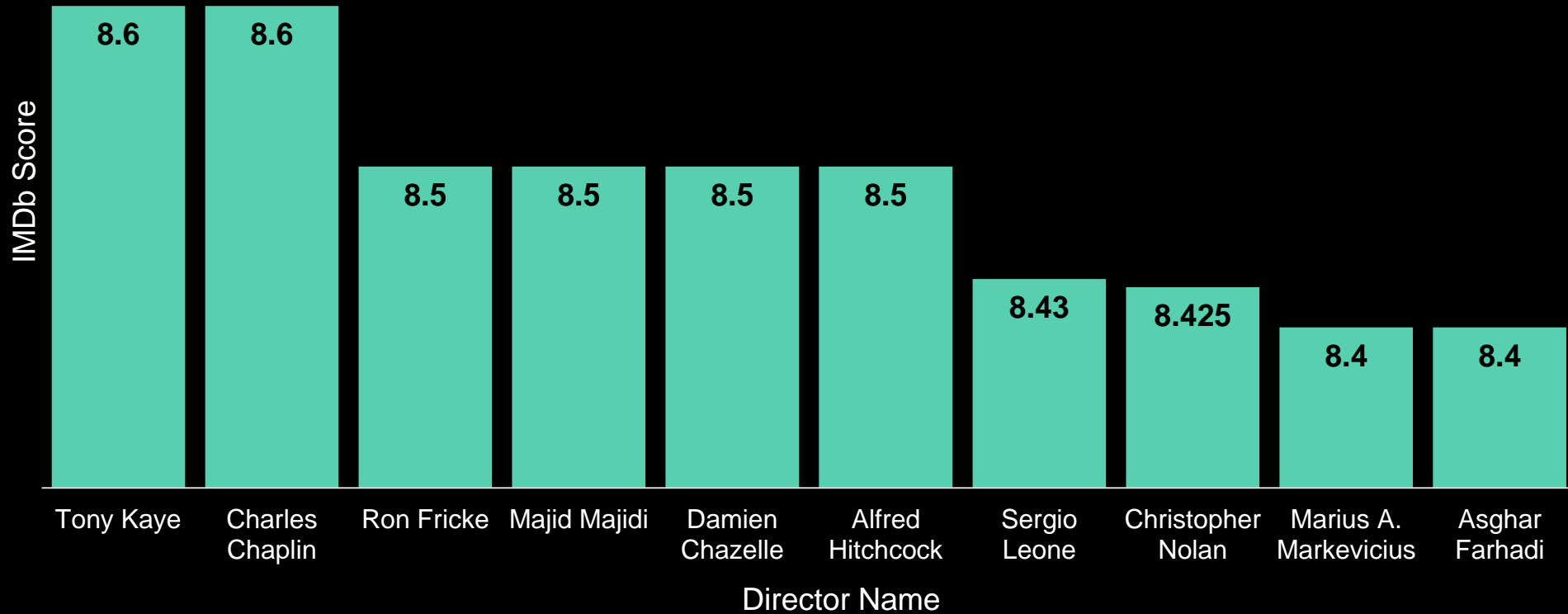
# C- 2: Top Other Language Movies

Rank	Top_Other_Language_Movie_Title	IMDb_Score	Language
1	The Good, the Bad and the Ugly	8.9	Italian
2	City of God	8.7	Portuguese
3	Seven Samurai	8.7	Japanese
4	Spirited Away	8.6	Japanese
5	Children of Heaven	8.5	Persian
6	The Lives of Others	8.5	German
7	A Separation	8.4	Persian
8	Amélie	8.4	French
9	Baahubali: The Beginning	8.4	Telugu
10	Das Boot	8.4	German

**NOTE:** Here only 10 records are shown the remaining records are in the excel workbook and the link for the excel workbook is provided in the slide “**DRIVE LINK**”

# D. Find the best directors

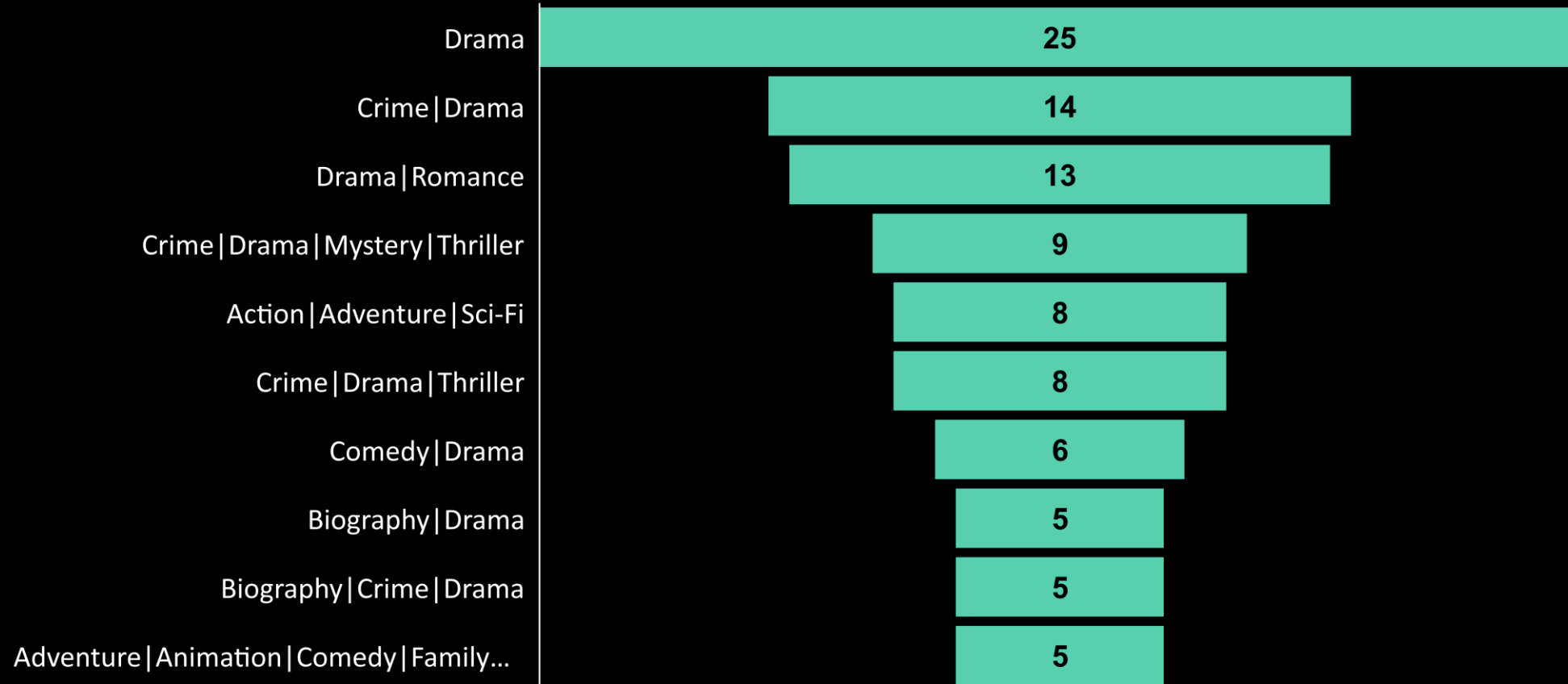
## Top 10 Best Directors





# E. Find popular genres

Top 10 Popular Genres



F. Find the critic-favorite and audience-favorite actors

# F– 1.1: Movies Of Meryl Streep

Movie Title
A Prairie Home Companion
Hope Springs
It's Complicated
Julie & Julia
Lions for Lambs
One True Thing
Out of Africa
The Devil Wears Prada
The Hours
The Iron Lady
The River Wild

# F- 1.2: Movies of Leonardo DiCaprio

Movie Title
Blood Diamond
Body of Lies
Catch Me If You Can
Django Unchained
Gangs of New York
Inception
J. Edgar
Marvin's Room

Movie Title
Revolutionary Road
Romeo + Juliet
Shutter Island
The Aviator
The Beach
The Departed
The Great Gatsby
The Great Gatsby

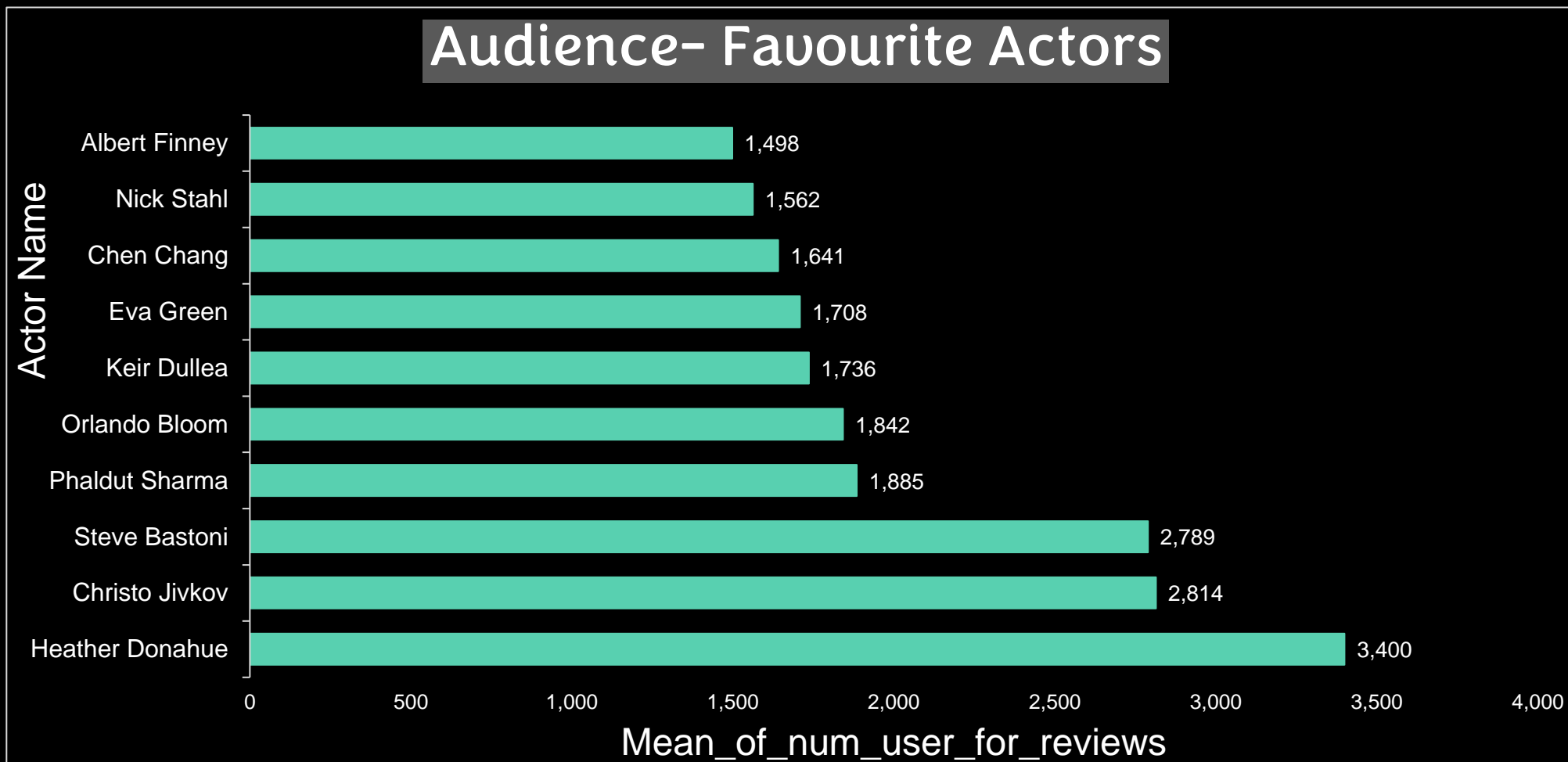
Movie Title
The Man in the Iron Mask
The Quick and the Dead
The Revenant
The Wolf of Wall Street
Titanic

# F- 1.3: Movies Of Brad Pitt

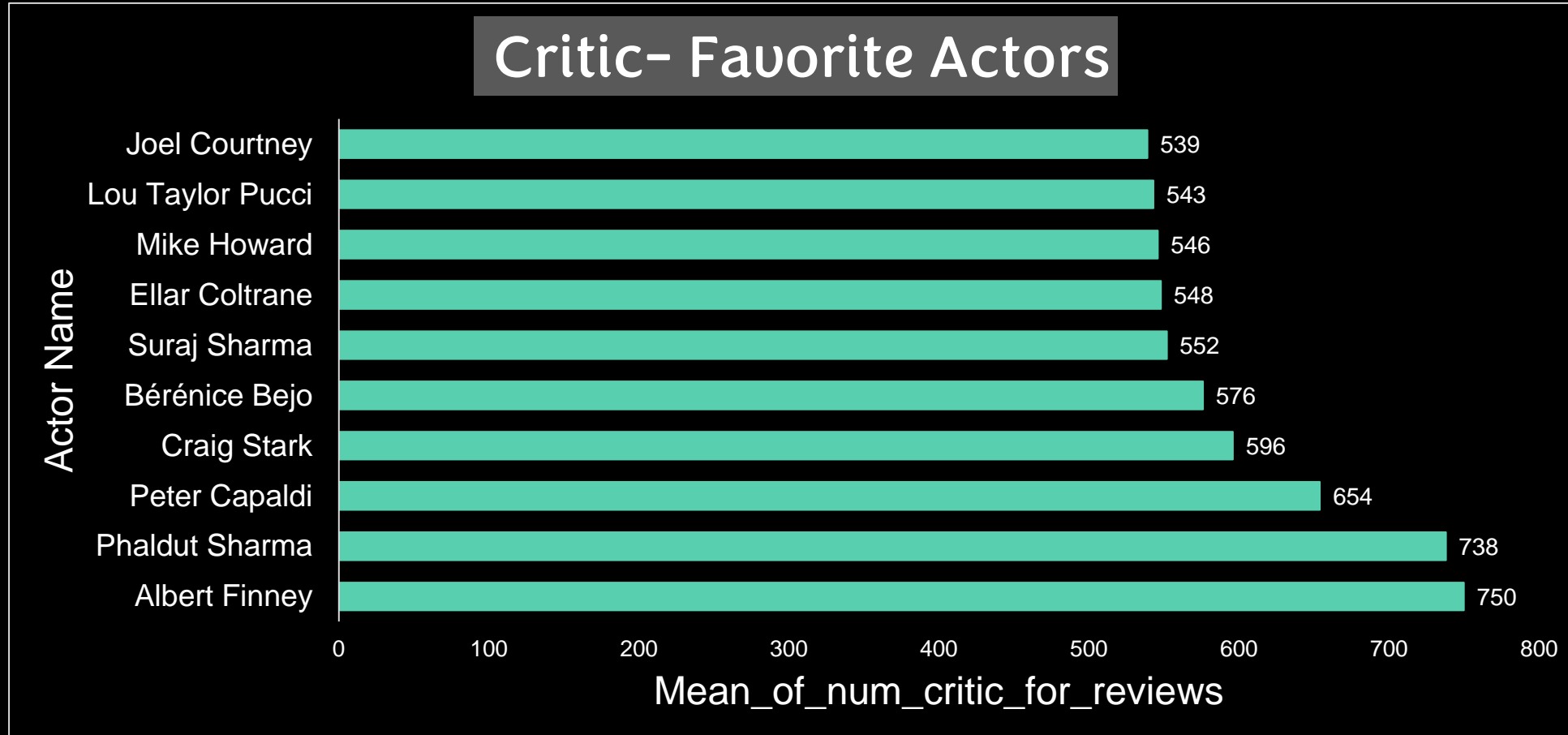
Movie Title
Babel
By the Sea
Fight Club
Fury
Interview with the Vampire: The Vampire Chronicles
Killing Them Softly
Mr. & Mrs. Smith
Ocean's Eleven
Ocean's Twelve

Movie Title
Seven Years in Tibet
Sinbad: Legend of the Seven Seas
Spy Game
The Assassination of Jesse James by the Coward Robert Ford
The Curious Case of Benjamin Button
The Tree of Life
Troy
True Romance

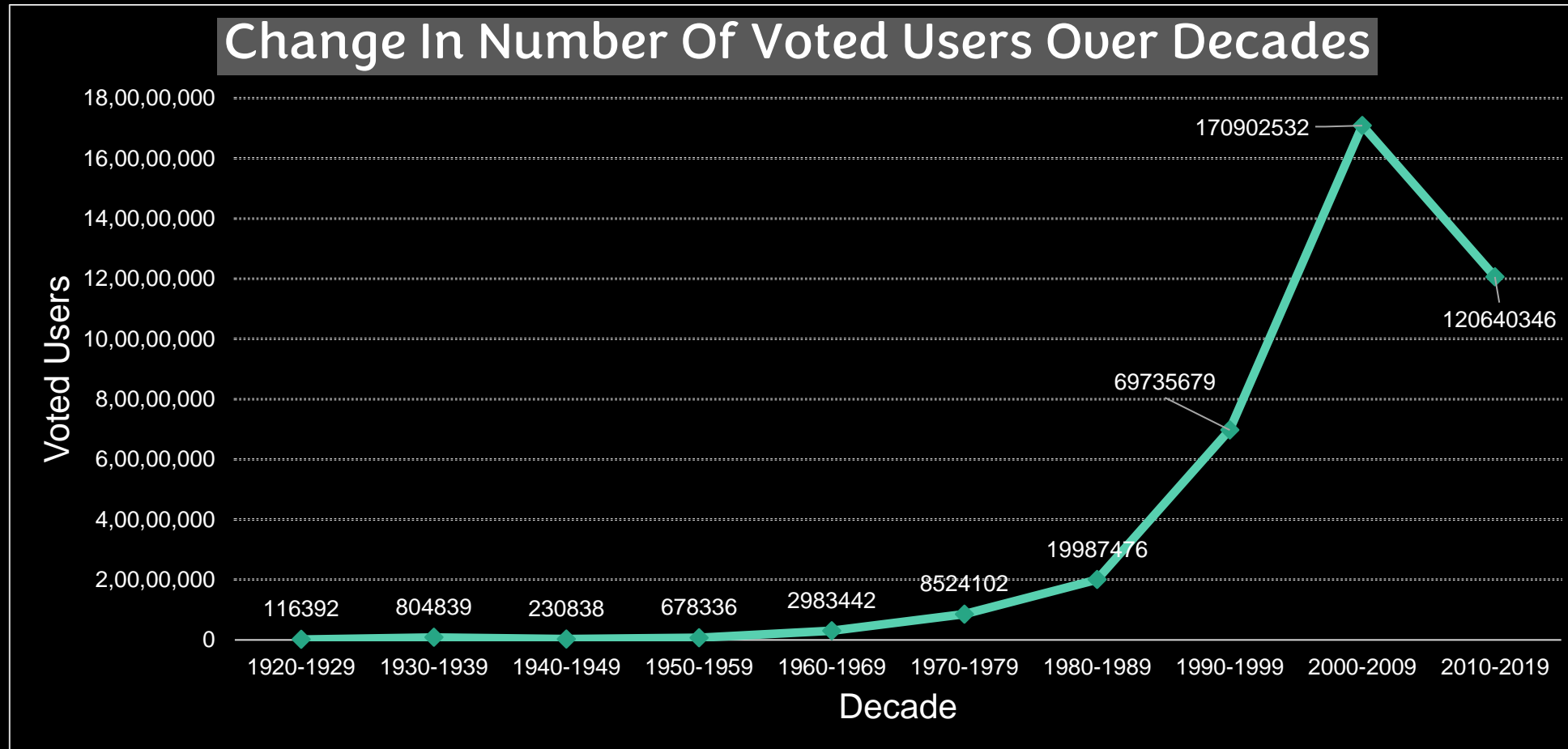
# F- 2.1: Audience- Favorite Actors



## F- 2.2: Critic- Favorite Actors



# F- 3: Change In Number Of Voted Users Over Decades





- Drive link of the excel file I created for the analysis.

<https://docs.google.com/spreadsheets/d/1DU0IGsvEKO-EBjHomDdJJFzdzF46VMpm/edit?usp=sharing&ouid=116897921059711507240&rtpof=true&sd=true>

DRIVE  
LINK

THANK  
YOU

