

IMDB MOVIE ANALYSIS

By- Ankita Yadav



Table of contents



01 Project Description

02 Approach

03 Tech-stack Used

04 Insights

05 Results





01

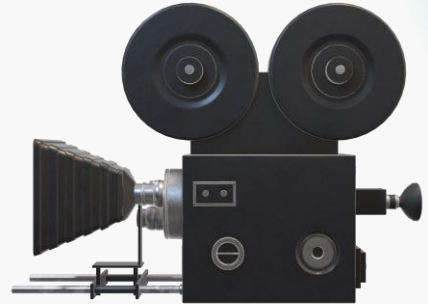
Project Description

Project Description



In this project IMDB movie dataset was provided and we were required to frame the problem. I was expected to define my own problems and expected to shed light on it and then used exploratory data analysis to clean the dataset and derive the required insight.

The [Root Cause Analysis](#) approach or [5 Why's Analysis](#) approach developed by [Sakichi Toyoda](#), founder of Toyota was used to create the report which conveys the data story.





02

Approach

Approach

- Understanding the datasets and its columns.
- Cleaning the dataset to derive the required insights.
- Using 5 why's approach for the deep understanding of the data and root cause problems.
- Analyzing the data for the data for the various problem statements.
- Using statistical and visualization tools to shed lights on results.





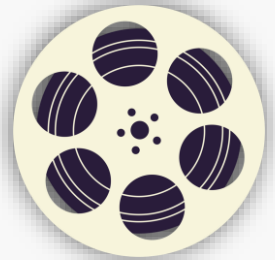
03

Tech-stack Used

Tech-Stack Used



Microsoft Excel
2019





Understanding the dataset



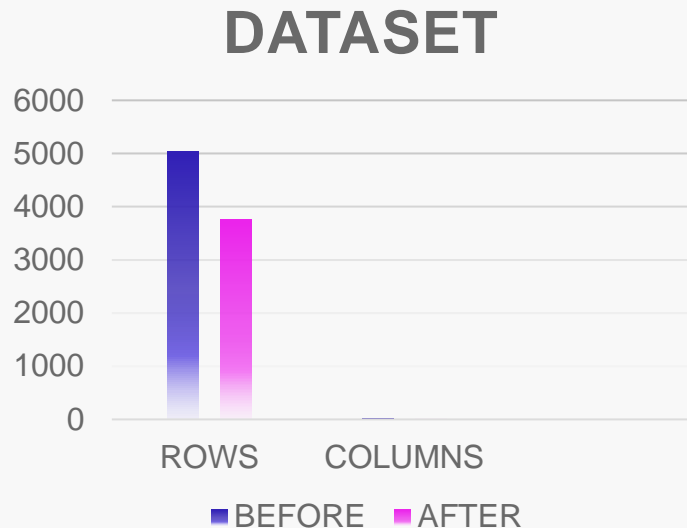
➤ Before Cleaning –

- Total Number of rows- 5044
- Total Number of columns- 28

➤ After Cleaning –

- Total Number of rows- 3724
- Total Number of Columns- 13

In the process of cleaning, 45 duplicates were removed and 3724 rows were obtained after removing the rows having blank cell and the unnecessary columns were dropped.





04

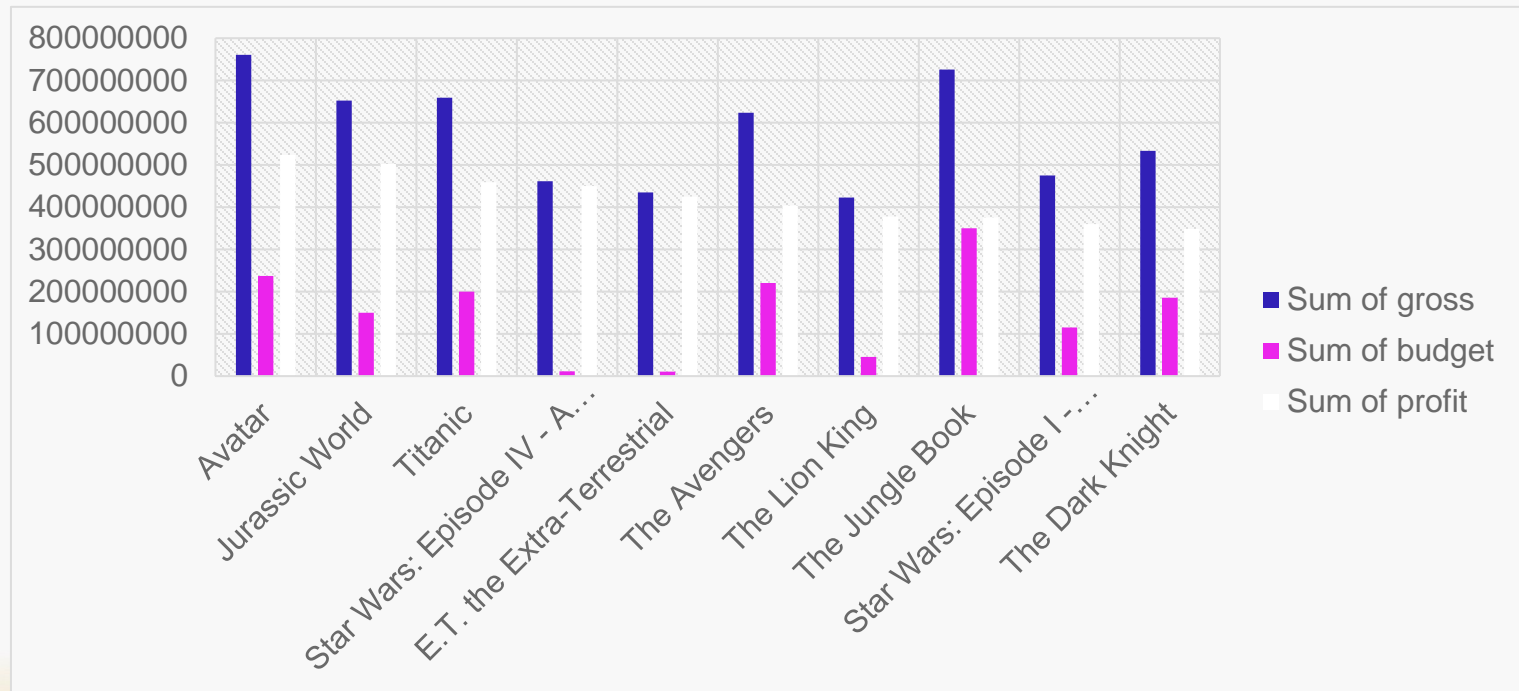
Insights



Most Profitable Movie

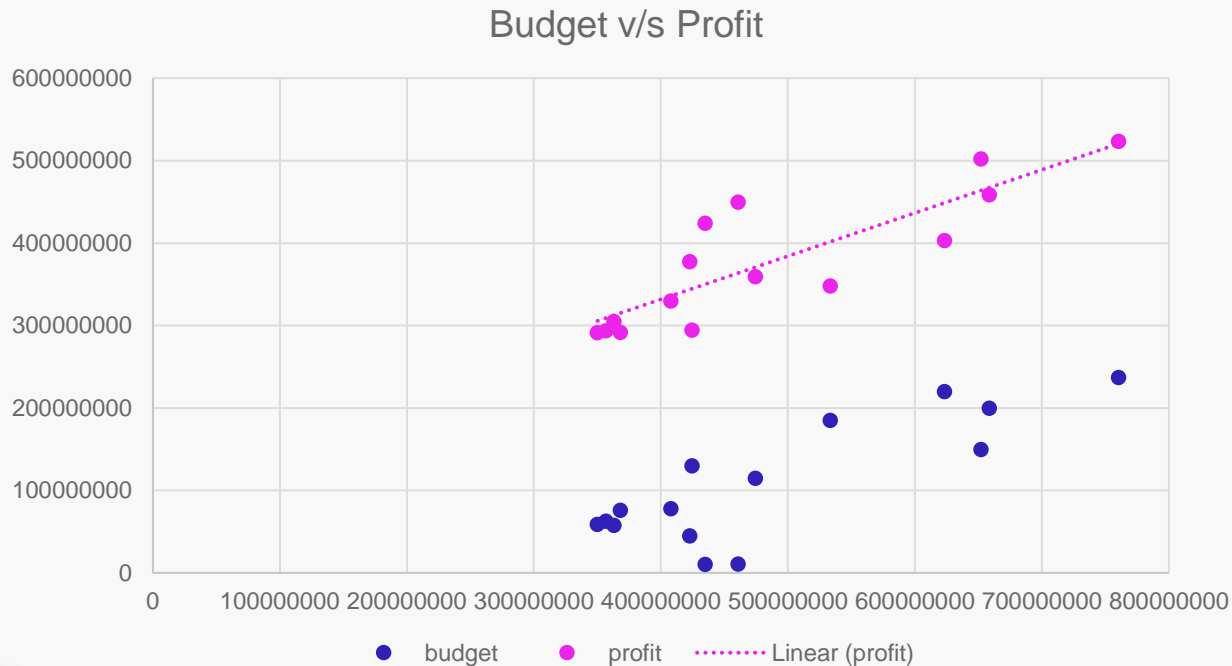
movie_title	Gross	Budget	Profit
Avatar	760505847	237000000	523505847
Jurassic World	652177271	150000000	502177271
Titanic	658672302	200000000	458672302
Star Wars: Episode IV - A New Hope	460935665	11000000	449935665
E.T. the Extra-Terrestrial	434949459	10500000	424449459
The Avengers	623279547	220000000	403279547
The Lion King	422783777	45000000	377783777
Star Wars: Episode I - The Phantom Menace	474544677	115000000	359544677
The Dark Knight	533316061	185000000	348316061
The Hunger Games	407999255	78000000	329999255
Deadpool	363024263	58000000	305024263
The Hunger Games: Catching Fire	424645577	130000000	294645577
Jurassic Park	356784000	63000000	293784000
Despicable Me 2	368049635	76000000	292049635
American Sniper	350123553	58800000	291323553

Based on the analysis of Gross v/s Budget, Avatar was the most profitable movie

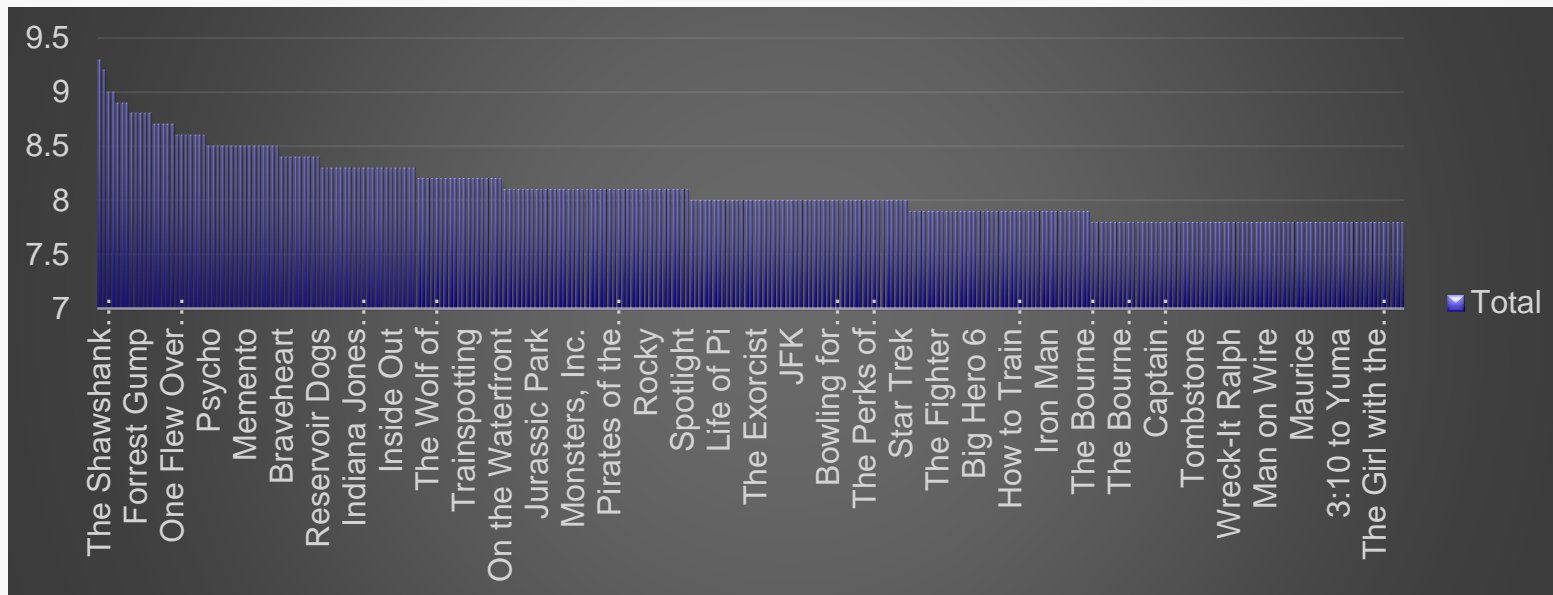




Budget V/s Profit

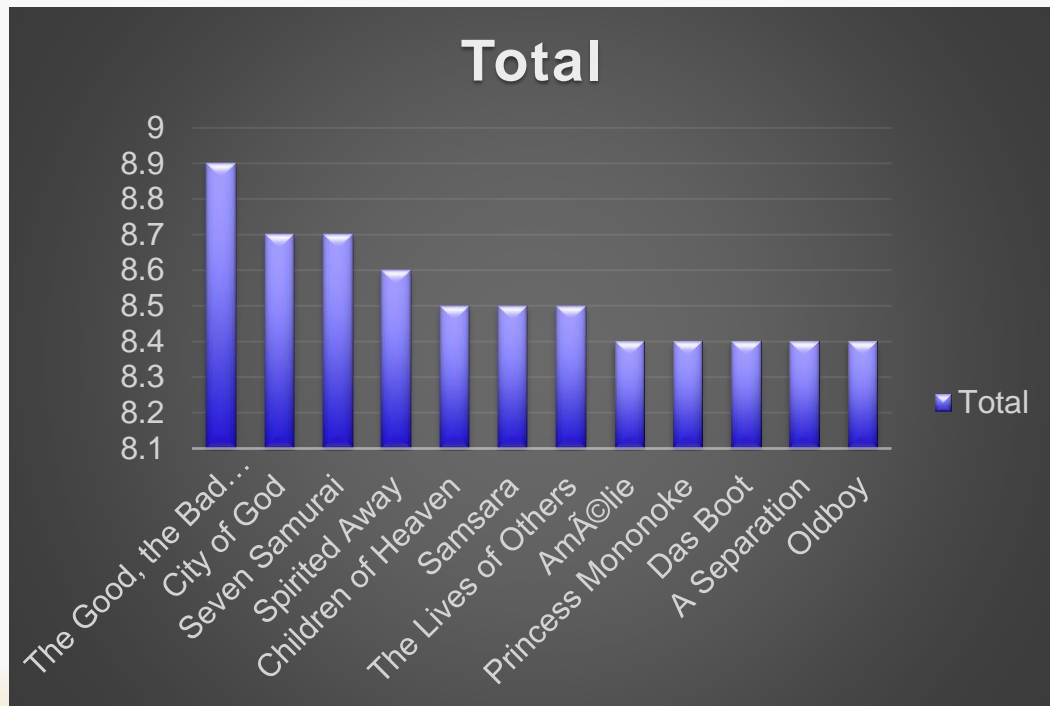


Top 250(English)



The Shawshank Redemption is the best English movie with average IMDB score of 9.3.

Top 250(Foreign)



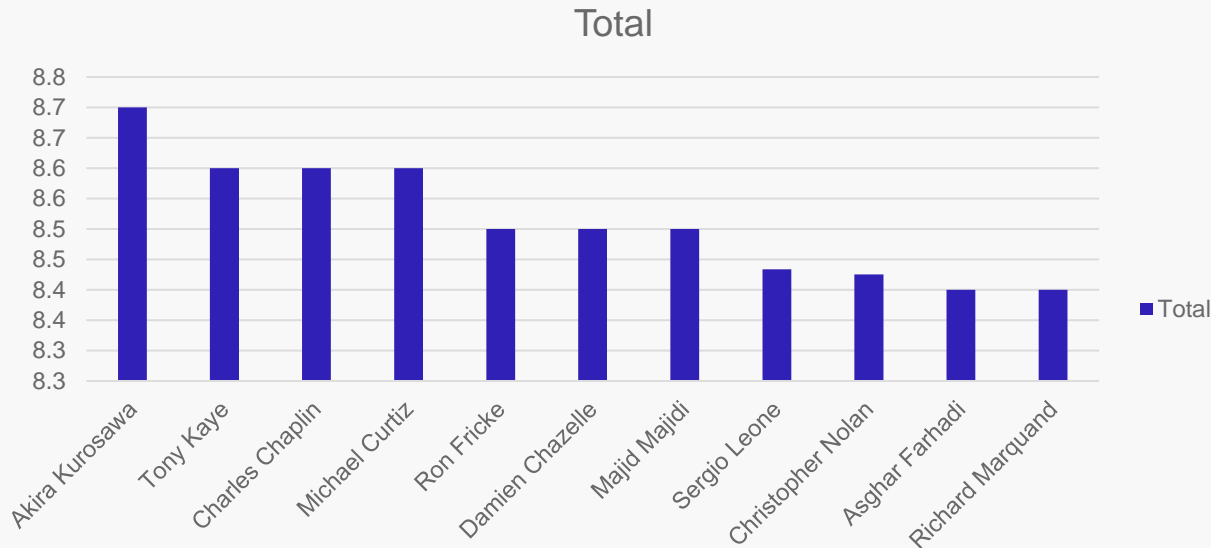
The Good, The Bad and The Ugly is the best Foreign movie(Italian) with highest IMDB ratings. Top 15 represented here.



Best Director

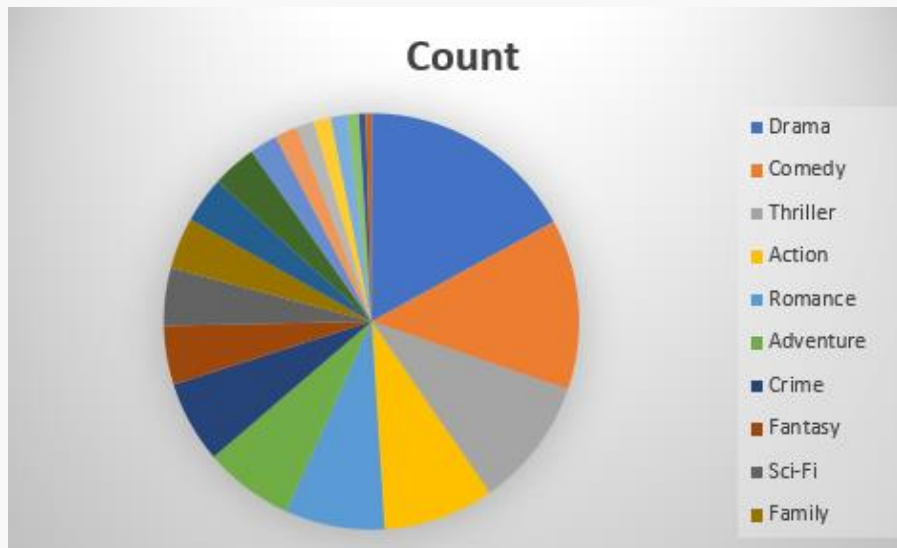
Top Directors	Average of imdb_score
Akira Kurosawa	8.7
Tony Kaye	8.6
Charles Chaplin	8.6
Michael Curtiz	8.6
Ron Fricke	8.5
Damien Chazelle	8.5
Majid Majidi	8.5
Sergio Leone	8.4
Christopher Nolan	8.4
Asghar Farhadi	8.4
Richard Marquand	8.4
Grand Total	8.5

Akira Kurosawa is the Best Director with the Average IMDB rating of 8.7.



Best Genre

genre	count
Drama	1876
Comedy	1455
Thriller	1105
Action	948
Romance	851
Adventure	769
Fantasy	502
Sci-Fi	490
Family	439
Horror	386
Mystery	377
Biography	238
Animation	195
War	150
History	147
Sport	147
Musical	95
Western	56
Documentary	45
Film-Noir	1



Drama, Comedy and Thriller were top three genres based on the movies released by Best Directors.



Actor Analysis

- Meryl Streep – 11
- Leonardo DiCaprio – 21
- Brad Pitt – 17

A helper column was created and VLOOKUP function was used to extract the movie titles of the asked actors.

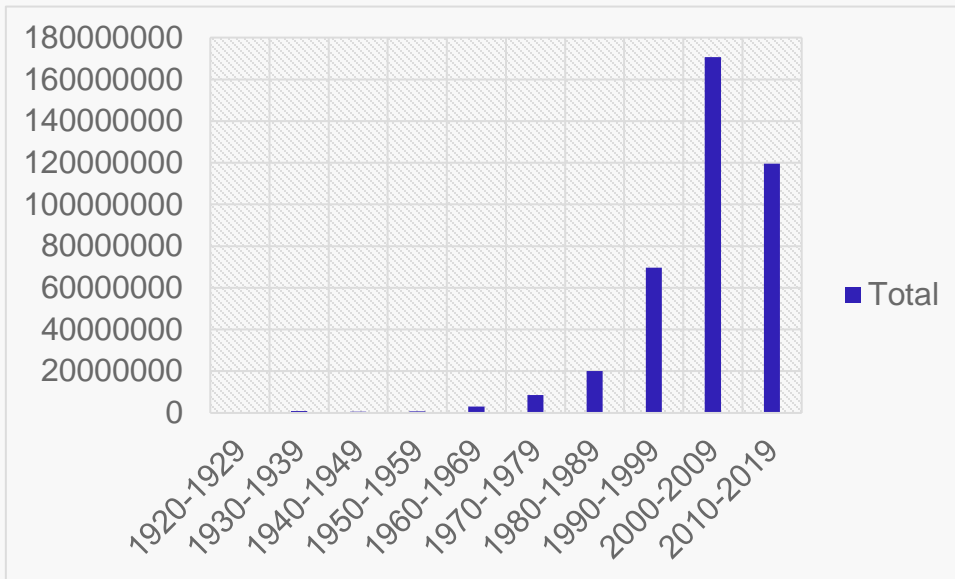
meryl_streep	leo_caprio	brad_pitt
It's Complicated	Titanic	The Curious Case of Benjamin Button
The River Wild	The Great Gatsby	Troy
Julie & Julia	Inception	Ocean's Twelve
The Devil Wears Prada	The Revenant	Mr. & Mrs. Smith
Lions for Lambs	The Aviator	Spy Game
Out of Africa	Django Unchained	Ocean's Eleven
Hope Springs	Blood Diamond	Fury
One True Thing	The Wolf of Wall Street	Seven Years in Tibet
The Hours	Gangs of New York	Fight Club
The Iron Lady	The Departed	Sinbad: Legend of the Seven Seas
A Prairie Home Companion	Shutter Island	Interview with the Vampire: The Vampire Chronicles
	Body of Lies	The Tree of Life
	Catch Me If You Can	The Assassination of Jesse James by the Coward Robert Ford
	The Beach	Babel
	Revolutionary Road	By the Sea
	The Man in the Iron Mask	Killing Them Softly
	J. Edgar	True Romance
	The Quick and the Dead	
	Marvin's Room	
	Romeo + Juliet	
	The Great Gatsby	



Decade Analysis

df_by_decade	Sum of num_voted_users
1920-1929	116387
1930-1939	804839
1940-1949	634449
1950-1959	678336
1960-1969	2982551
1970-1979	8543977
1980-1989	19987476
1990-1999	69581866
2000-2009	170735893
2010-2019	119534391
Grand Total	393600165

Most Number of movies
were released in decade
2000-2009 .

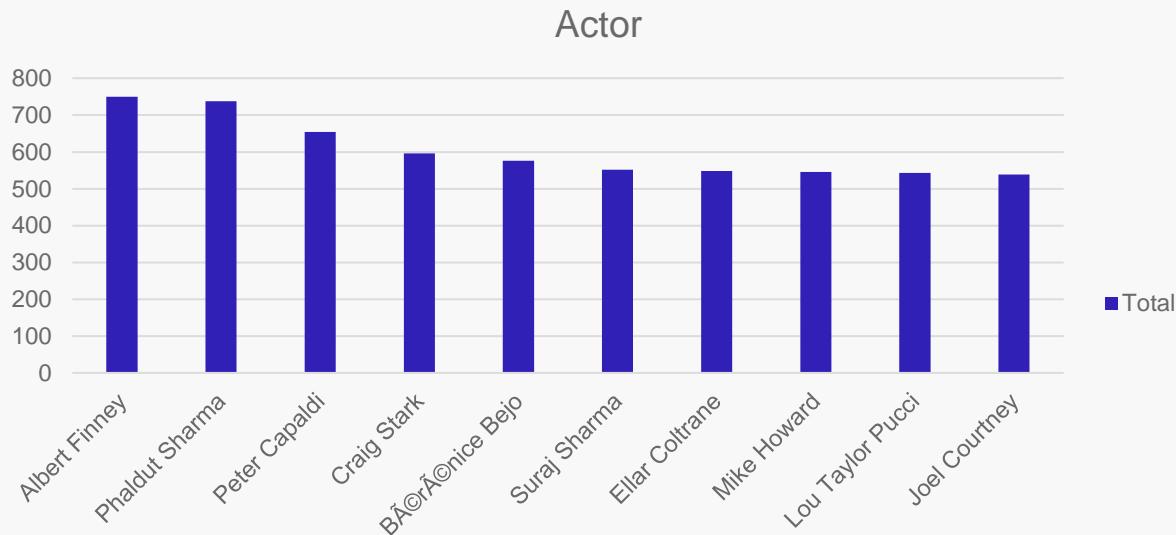




Critic Favorite Actor

Row Labels	Average of num_critic_for_reviews
Albert Finney	750
Phaldut Sharma	738
Peter Capaldi	654
Craig Stark	596
BÃ©rÃ©nice Bejo	576
Suraj Sharma	552
Ellar Coltrane	548
Mike Howard	546
Lou Taylor Pucci	543
Joel Courtney	539
Grand Total	617

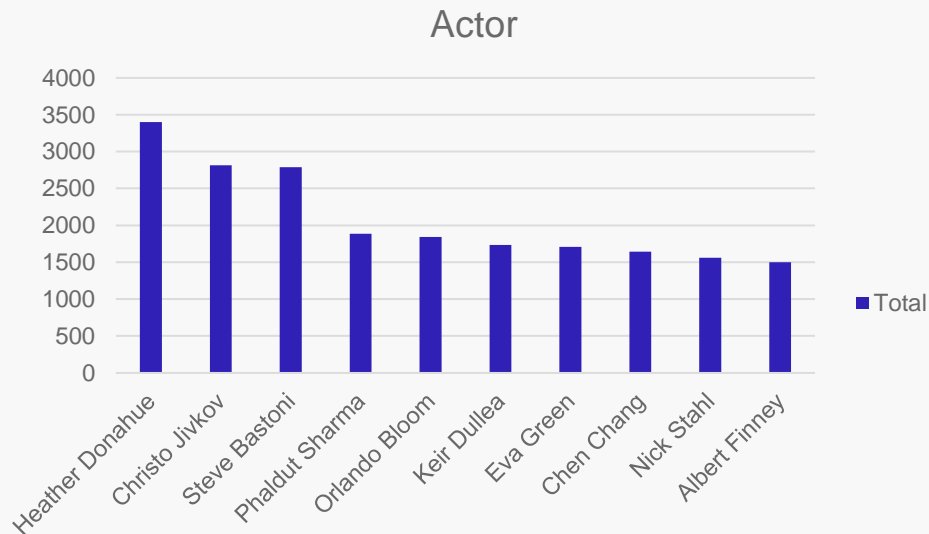
Based on critics reviews,
Albert Finney was the
favorite actor.



Audience Favorite Actor

Row Labels	Average of num_user_for_reviews
Heather Donahue	3400
Christo Jivkov	2814
Steve Bastoni	2789
Phaldut Sharma	1885
Orlando Bloom	1842
Keir Dullea	1736
Eva Green	1708
Chen Chang	1641
Nick Stahl	1562
Albert Finney	1498
Grand Total	1946

Based on the audience votes, Heather Donahue was the favorite actor.





05

Results

Results



Profitable Movie

Avatar was the most
Profitable movie
\$52,35,05,847

Top 250

Shawshank Redemption in
English with average rating
of 9.2 and The Good The
Bad and The Ugly in Italian
with average voting of 8.9

Best Director

Akira Kurosawa with
average rating of 8.7

Genre

Based on the Best
Directors the top 3 genres
were DRAMA, COMEDY
AND THRILLER.

Actor

Critics favorite is Albert
Finney (750 votes)
Audience favorite is
Heather Donahue (3400
votes)

Thanks

ankitaydv07@gmail.com

