

THANK YOU SO MUCH TEAM TRINITY , FOR GIVING THIS PROJECT TO ME.

THIS PROJECT IS BASED ON THE DATASET OF IMDB\_MOVIES. DATA SET CONSISTED DATA ON MOVIE DETAILS LIKE :- COLOR, DIRECTION,NUMBER OF REVIEWS, ACTOR 1, ACTOR2, ACTOR3, FACEBOOK LIKES, TIME DURATION , YEAR THE MOVIE WAS MADE, ETC.

IN THIS PROJECT , I OBSERVED MULTIPLE LIKINGS OF PEOPLE , WHICH CAN LEAD TO THE SUCCESS OF MOVIE IF MADE PROPER OBSERVATIONS AND PRACTICED THEM IN REAL LIFE SCENARIO.

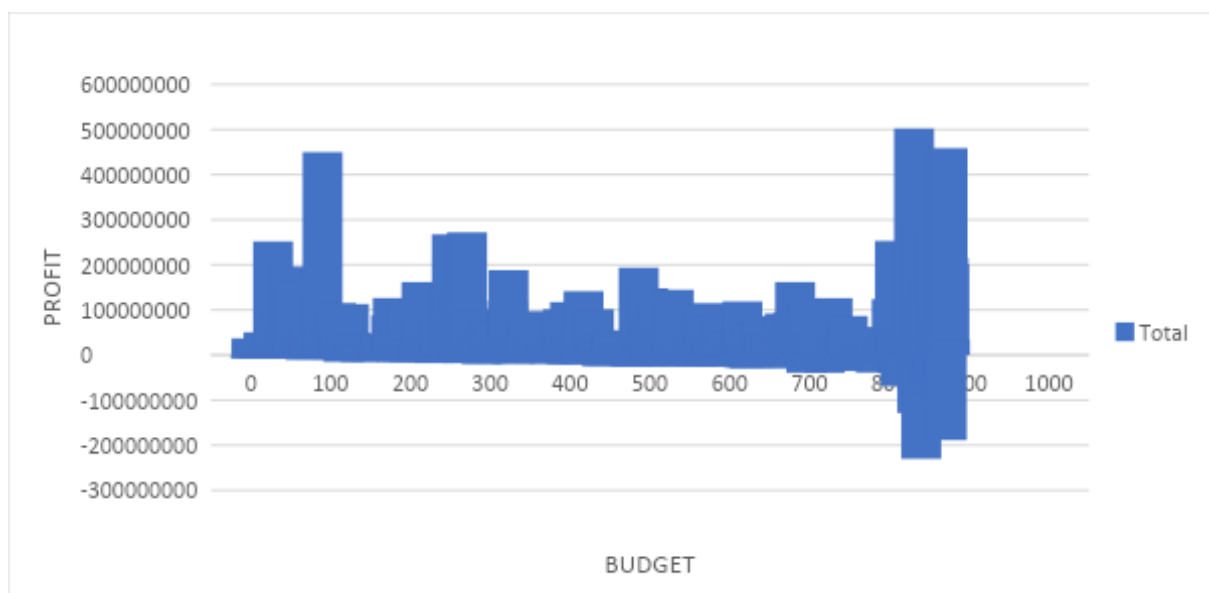
I OBSERVED THE GENRE WHICH WAS MOST LIKED BY USERS i.e. CRIME/DRAMA/FANTASY/MYSTERY. USERS LIKE WATCHING MOVIES IN THESE PARTICULAR GENRE .

I OBSERVED THE MOST LIKED LEAD ACTORS.(can see and compare during the document)

I OBSERVED THE MOST LIKED DIRECTION OF MOVIES.(can see and compare during the document)

I OBSERVED THAT THE NUMBER OF MOVIES RELEASED WAS SIMILAR DURING 1970'S TO 2000'S.

### IN THE FOLLOWING TABLE



## WE CAN SEE THE PROFIT OF MOVIE AND THE BUDGET USED IN MAKING OF IT.

You can follow up this report for further details and query focused to solve in this data.

I HAVE USED MS-EXCEL FOR DATA CLEANING AND DATA ANALYSIS.

### **THE QUERY ANSWERING:-**

You are required to provide a detailed report for the below data record mentioning the answers of the questions that follows:

- A. **Cleaning the data::** This is one of the most important step to perform before moving forward with the analysis. Use your knowledge learned till now to do this. (Dropping columns, removing null values, etc.)

**Your task:** Clean the data

HOW I CLEANED THE DATA:-

STEP 1 – REMOVING DUPLICATED VALUES

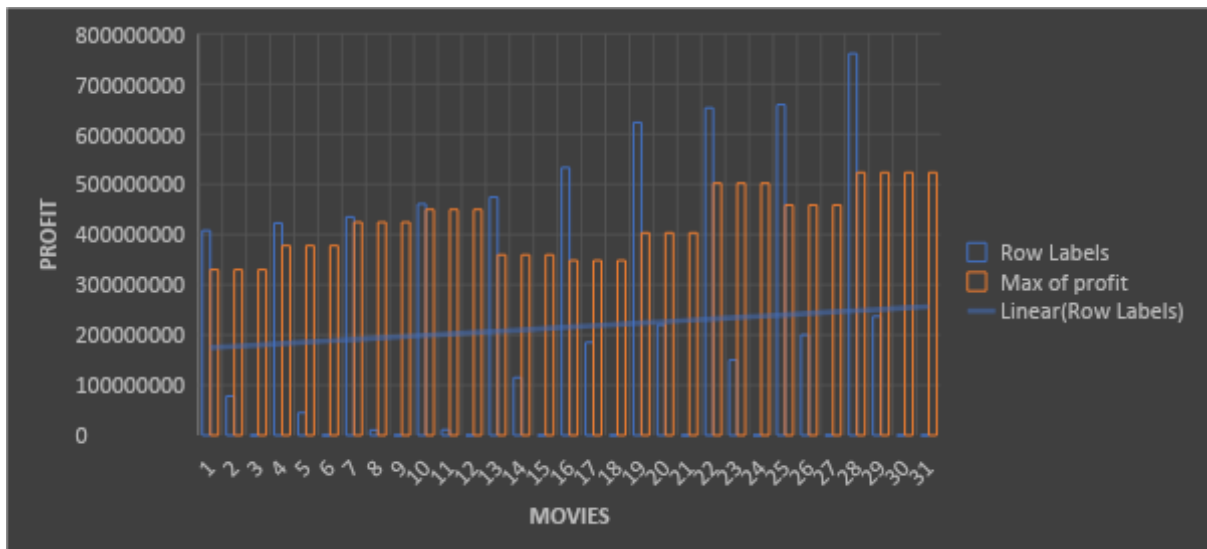
STEP 2- REMOVING NULL VALUES

- B. **Movies with highest profit:** Create a new column called profit which contains the difference of the two columns: gross and budget. Sort the column using the profit column as reference. Plot profit (y-axis) vs budget (x-axis) and observe the outliers using the appropriate chart type.

**Your task:** Find the movies with the highest profit?

Row Labels	Max of profit
407999255	329999255
78000000	329999255
The Hunger Games	329999255
422783777	377783777
45000000	377783777
The Lion King	377783777
434949459	424449459
10500000	424449459
E.T. the Extra-Terrestrial	424449459
460935665	449935665

<b>11000000</b>	<b>449935665</b>
Star Wars: Episode IV - A New Hope	449935665
<b>474544677</b>	<b>359544677</b>
<b>115000000</b>	<b>359544677</b>
Star Wars: Episode I - The Phantom Menace	359544677
<b>533316061</b>	<b>348316061</b>
<b>185000000</b>	<b>348316061</b>
The Dark Knight	348316061
<b>623279547</b>	<b>403279547</b>
<b>220000000</b>	<b>403279547</b>
The Avengers	403279547
<b>652177271</b>	<b>502177271</b>
<b>150000000</b>	<b>502177271</b>
Jurassic World	502177271
<b>658672302</b>	<b>458672302</b>
<b>200000000</b>	<b>458672302</b>
Titanic	458672302
<b>760505847</b>	<b>523505847</b>
<b>237000000</b>	<b>523505847</b>
Avatar	523505847
<b>Grand Total</b>	<b>523505847</b>



**THIS CHART SHOWS THE PROFIT GAINED BY TOP 10 MOVIES.**

**WE CAN SEE THAT AVATAR HAS GAINED THE MOST PROFIT ALL THESE YEARS - FROM 1920'S TO 2010'S**

C. **Top 250:** Create a new column `IMDb_Top_250` and store the top 250 movies with the highest IMDb Rating (corresponding to the column: `imdb_score`). Also make sure that for all of these movies, the `num_voted_users` is greater than 25,000. Also add a `Rank` column containing the values 1 to 250 indicating the ranks of the corresponding films.

Extract all the movies in the `IMDb_Top_250` column which are not in the English language and store them in a new column named `Top_Foreign_Lang_Film`. You can use your own imagination also!

**Your task:** Find IMDB Top 250

IMDB_TOP_250
Row Labels
<b>2001: A Space Odyssey</b>
427357
8.3
<b>A Separation</b>
151812
8.4
<b>Alien</b>
563827
8.5
<b>Aliens</b>
488537
8.4
<b>Amélie</b>
534262
8.4
<b>Amadeus</b>
270790
8.3
<b>American Beauty</b>
822500
8.4
<b>American History X</b>
782437
8.6
<b>Apocalypse Now</b>

<b>450676</b>
8.5
<b>Back to the FutureÂ</b>
<b>732212</b>
8.5
<b>Batman BeginsÂ</b>
<b>980946</b>
8.3
<b>BraveheartÂ</b>
<b>736638</b>
8.4
<b>CasablancaÂ</b>
<b>387508</b>
8.6
<b>Children of HeavenÂ</b>
<b>27882</b>
8.5
<b>City of GodÂ</b>
<b>533200</b>
8.7
<b>Das BootÂ</b>
<b>168203</b>
8.4
<b>Django UnchainedÂ</b>
<b>955174</b>
8.5
<b>DownfallÂ</b>
<b>248354</b>
8.3
<b>Eternal Sunshine of the Spotless MindÂ</b>
<b>666937</b>
8.3
<b>Fight ClubÂ</b>
<b>1347461</b>
8.8
<b>Forrest GumpÂ</b>
<b>1251222</b>
8.8
<b>GladiatorÂ</b>
<b>982637</b>
8.5
<b>Good Will HuntingÂ</b>
<b>604904</b>
8.3
<b>GoodfellasÂ</b>
<b>728685</b>

8.7
<b>Inception</b>
<b>1468200</b>
8.8
<b>Indiana Jones and the Last Crusade</b>
<b>515306</b>
8.3
<b>Inglourious Basterds</b>
<b>885175</b>
8.3
<b>Inside Out</b>
<b>345198</b>
8.3
<b>Interstellar</b>
<b>928227</b>
8.6
<b>L.A. Confidential</b>
<b>414219</b>
8.3
<b>Lawrence of Arabia</b>
<b>192775</b>
8.4
<b>Memento</b>
<b>845580</b>
8.5
<b>Metropolis</b>
<b>111841</b>
8.3
<b>Modern Times</b>
<b>143086</b>
8.6
<b>Monty Python and the Holy Grail</b>
<b>382240</b>
8.3
<b>Oldboy</b>
<b>356181</b>
8.4
<b>Once Upon a Time in America</b>
<b>221000</b>
8.4
<b>One Flew Over the Cuckoo's Nest</b>
<b>680041</b>
8.7
<b>Princess Mononoke</b>
<b>221552</b>
8.4

<b>Psycho</b>
<b>422432</b>
8.5
<b>Pulp Fiction</b>
<b>1324680</b>
8.9
<b>Raging Bull</b>
<b>235133</b>
8.3
<b>Raiders of the Lost Ark</b>
<b>661017</b>
8.5
<b>Requiem for a Dream</b>
<b>573541</b>
8.4
<b>Reservoir Dogs</b>
<b>664719</b>
8.4
<b>Room</b>
<b>161288</b>
8.3
<b>Saving Private Ryan</b>
<b>881236</b>
8.6
<b>Scarface</b>
<b>537442</b>
8.3
<b>Schindler's List</b>
<b>865020</b>
8.9
<b>Se7en</b>
<b>1023511</b>
8.6
<b>Seven Samurai</b>
<b>229012</b>
8.7
<b>Snatch</b>
<b>600996</b>
8.3
<b>Some Like It Hot</b>
<b>175196</b>
8.3
<b>Spirited Away</b>
<b>417971</b>
8.6
<b>Star Wars: Episode IV - A New Hope</b>

<b>911097</b>
8.7
<b>Star Wars: Episode V - The Empire Strikes BackÂ</b>
<b>837759</b>
8.8
<b>Star Wars: Episode VI - Return of the JediÂ</b>
<b>681857</b>
8.4
<b>Terminator 2: Judgment DayÂ</b>
<b>744891</b>
8.5
<b>The Dark Knight RisesÂ</b>
<b>1144337</b>
8.5
<b>The Dark KnightÂ</b>
<b>1676169</b>
9
<b>The DepartedÂ</b>
<b>873649</b>
8.5
<b>The Godfather: Part IIÂ</b>
<b>790926</b>
9
<b>The GodfatherÂ</b>
<b>1155770</b>
9.2
<b>The Good, the Bad and the UglyÂ</b>
<b>503509</b>
8.9
<b>The Green MileÂ</b>
<b>782610</b>
8.5
<b>The HuntÂ</b>
<b>170155</b>
8.3
<b>The Lion KingÂ</b>
<b>644348</b>
8.5
<b>The Lives of OthersÂ</b>
<b>259379</b>
8.5
<b>The Lord of the Rings: The Fellowship of the RingÂ</b>
<b>1238746</b>
8.8
<b>The Lord of the Rings: The Return of the KingÂ</b>
<b>1215718</b>

8.9
<b>The Lord of the Rings: The Two Towers</b>
<b>1100446</b>
8.7
<b>The Matrix</b>
<b>1217752</b>
8.7
<b>The Pianist</b>
<b>497946</b>
8.5
<b>The Prestige</b>
<b>844052</b>
8.5
<b>The Shawshank Redemption</b>
<b>1689764</b>
9.3
<b>The Silence of the Lambs</b>
<b>887467</b>
8.6
<b>The Sting</b>
<b>175607</b>
8.3
<b>The Usual Suspects</b>
<b>740918</b>
8.6
<b>Toy Story 3</b>
<b>544884</b>
8.3
<b>Toy Story</b>
<b>623757</b>
8.3
<b>Unforgiven</b>
<b>277505</b>
8.3
<b>Up</b>
<b>665575</b>
8.3
<b>WALL·E</b>
<b>718837</b>
8.4
<b>Whiplash</b>
<b>399138</b>
8.5
<b>Grand Total</b>

TOP_FOREIGN_LANG_MOVIES
Row Labels
A Separation
151812
8.4
Persian
Amélie
534262
8.4
French
Children of Heaven
27882
8.5
Persian
City of God
533200
8.7
Portuguese
Das Boot
168203
8.4
German
Downfall
248354
8.3
German
Metropolis
111841
8.3
German
Oldboy
356181
8.4
Korean
Princess Mononoke
221552
8.4
Japanese
Seven Samurai
229012
8.7
Japanese
Spirited Away
417971
8.6
Japanese

The Good, the Bad and the Ugly
503509
8.9
Italian
The Hunt
170155
8.3
Danish
The Lives of Others
259379
8.5
German
Grand Total

**THE ABOVE GIVEN LIST IS OF TOP 250 IMDB SCORED MOVIES**

**WITH DIFFERENT LANGUAGE MOVIES ALSO. BUT ENGLISH IS ALWAYS IS THE MOST WATCHED AMONG THEM**

D. **Best Directors:** Group the column using the director\_name column.

Find out the top 10 directors for whom the mean of imdb\_score is the highest and store them in a new column top10director. In case of a tie in IMDb score between two directors, sort them alphabetically.

**Your task:** Find the best directors

	AVG_IMDB_SCORE	DIRECTORS
1	8.7	AKIRA KUROSAWA
2	8.6	CHARLES CHAPLIN
3	8.6	MICHAEL CURTIZ
4	8.6	TONY KAYE
5	8.5	DAMIEN CHAZELLE
6	8.5	MAJID MAJIDI
7	8.5	RON FRICKE
8	8.433333	SERGIO LEONE
9	8.425	CHRISTOPHER NOLAN
10	8.4	ASGHAR FARHADI

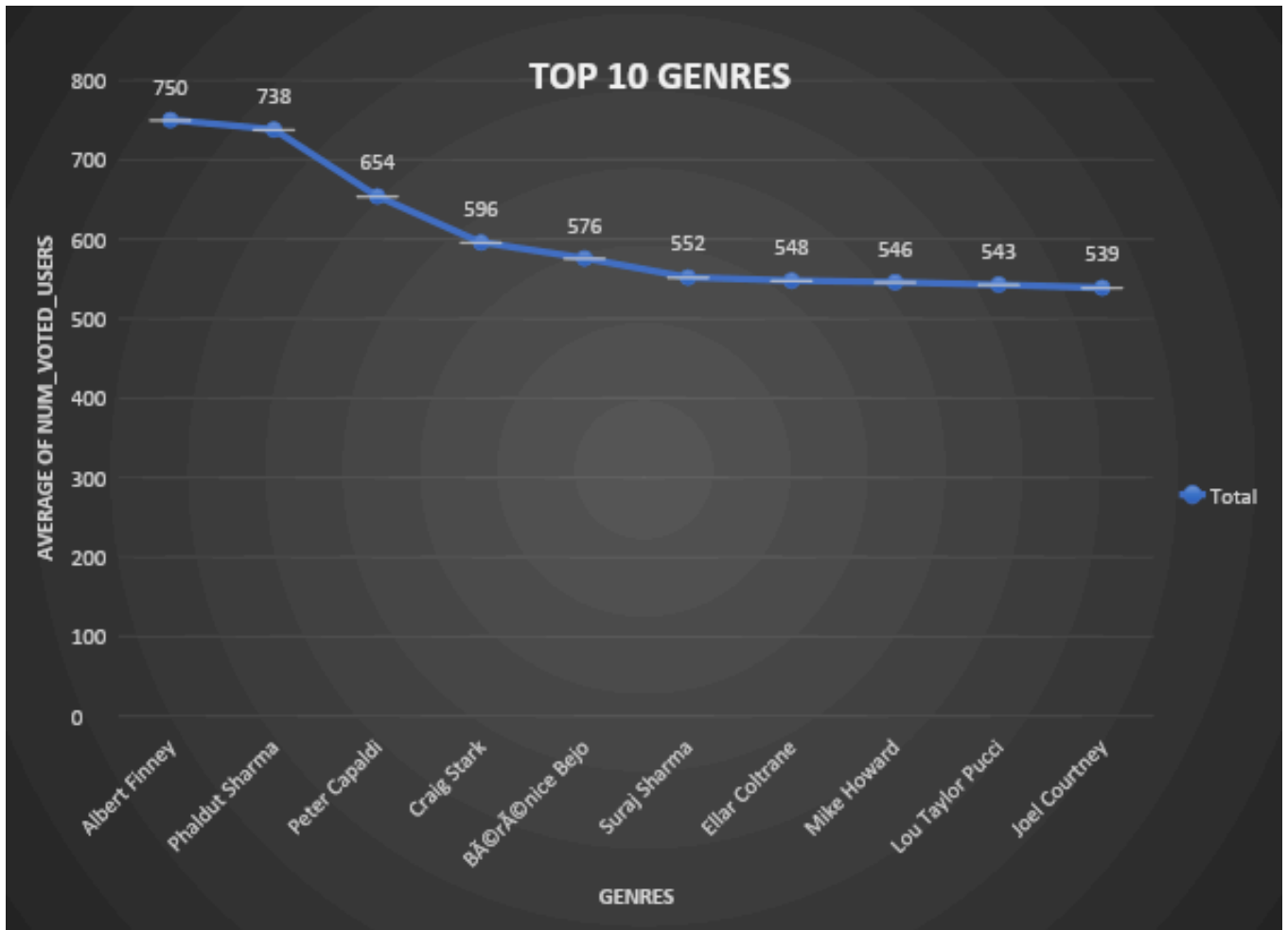
**THIS IS THE LIST OF MOST LIKED DIRECTORS.**

## WE CAN SEE THAT USERS LIKE THE DIRECTION OF AKIRA KUROSAWA THE MOST

E. **Popular Genres:** Perform this step using the knowledge gained while performing previous steps.

**Your task:** Find popular genres

Row Labels	Average of num_voted_users	RANK
Crime   Drama   Fantasy   Mystery	782610	1
Action   Adventure   Drama   Fantasy	656368	2
Adventure   Animation   Drama   Family   Musical	644348	3
Adventure   Drama   War	517309.5	4
Action   Adventure   Fantasy   Sci-Fi	488828.5833	5
Action   Adventure   Animation   Family	479166	6
Adventure   Comedy   Crime   Drama	475518	7
Action   Drama   Romance	473182.6667	8
Drama   Fantasy   War	467234	9
Action   Drama   Fantasy   War	416254	10
<b>Grand Total</b>	<b>531323.4</b>	



**THE MOST ENJOYED GENRE BY THE USERS IS -CRIME / DRAMA/ FANTASY/MYSTERTY**

F. **Charts:** Create three new columns namely, Meryl\_Streep, Leo\_Caprio, and Brad\_Pitt which contain the movies in which the actors: 'Meryl Streep', 'Leonardo DiCaprio', and 'Brad Pitt' are the lead actors. Use only the actor\_1\_name column for extraction. Also, make sure that you use the names 'Meryl Streep', 'Leonardo DiCaprio', and 'Brad Pitt' for the said extraction.

Append the rows of all these columns and store them in a new column named Combined.

Group the combined column using the actor\_1\_name column.

Find the mean of the num\_critic\_for\_reviews and num\_users\_for\_review and

identify the actors which have the highest mean.

Observe the change in number of voted users over decades using a bar chart. Create a column called decade which represents the decade to which every movie belongs to. For example, the title\_year year 1923, 1925 should be stored as 1920s. Sort the column based on the column decade, group it by decade and find the sum of users voted in each decade. Store this in a new data frame called df\_by\_decade.

**Your task:** Find the critic-favorite and audience-favorite actors

#### COMBINED

##### A Prairie Home Companion

Meryl Streep

##### Babel

Brad Pitt

##### Blood Diamond

Leonardo DiCaprio

##### Body of Lies

Leonardo DiCaprio

##### By the Sea

Brad Pitt

##### Catch Me If You Can

Leonardo DiCaprio

##### Django Unchained

Leonardo DiCaprio

##### Fight Club

Brad Pitt

##### Fury

Brad Pitt

##### Gangs of New York

Leonardo DiCaprio

##### Hope Springs

Meryl Streep

##### Inception

Leonardo DiCaprio

##### Interview with the Vampire: The Vampire Chronicles

Brad Pitt

##### It's Complicated

Meryl Streep

##### J. Edgar

Leonardo DiCaprio

##### Julie & Julia

Meryl Streep

##### Killing Them Softly

---

Brad Pitt

**Lions for Lambs**

---

Meryl Streep

**Marvin's Room**

---

Leonardo DiCaprio

**Mr. & Mrs. Smith**

---

Brad Pitt

**Ocean's Eleven**

---

Brad Pitt

**Ocean's Twelve**

---

Brad Pitt

**One True Thing**

---

Meryl Streep

**Out of Africa**

---

Meryl Streep

**Revolutionary Road**

---

Leonardo DiCaprio

**Romeo + Juliet**

---

Leonardo DiCaprio

**Seven Years in Tibet**

---

Brad Pitt

**Shutter Island**

---

Leonardo DiCaprio

**Sinbad: Legend of the Seven Seas**

---

Brad Pitt

**Spy Game**

---

Brad Pitt

**The Assassination of Jesse James by the Coward Robert Ford**

---

Brad Pitt

**The Aviator**

---

Leonardo DiCaprio

**The Beach**

---

Leonardo DiCaprio

**The Curious Case of Benjamin Button**

---

Brad Pitt

**The Departed**

---

Leonardo DiCaprio

**The Devil Wears Prada**

---

Meryl Streep

**The Great Gatsby**

---

Leonardo DiCaprio

**The Hours**

---

Meryl Streep

**The Iron Lady**

---

Meryl Streep

**The Man in the Iron Mask**

---

Leonardo DiCaprio

**The Quick and the Dead**

Leonardo DiCaprio

**The Revenant**

Leonardo DiCaprio

**The River Wild**

Meryl Streep

**The Tree of Life**

Brad Pitt

**The Wolf of Wall Street**

Leonardo DiCaprio

**Titanic**

Leonardo DiCaprio

**Troy**

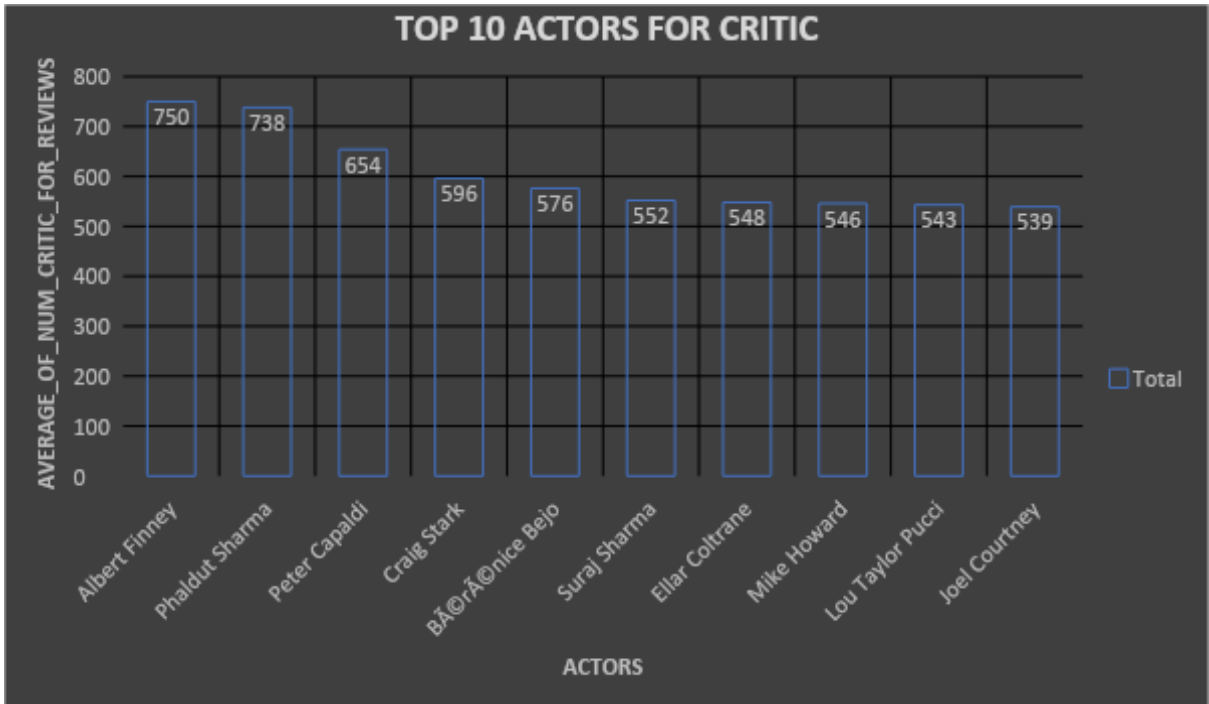
Brad Pitt

**True Romance**

Brad Pitt

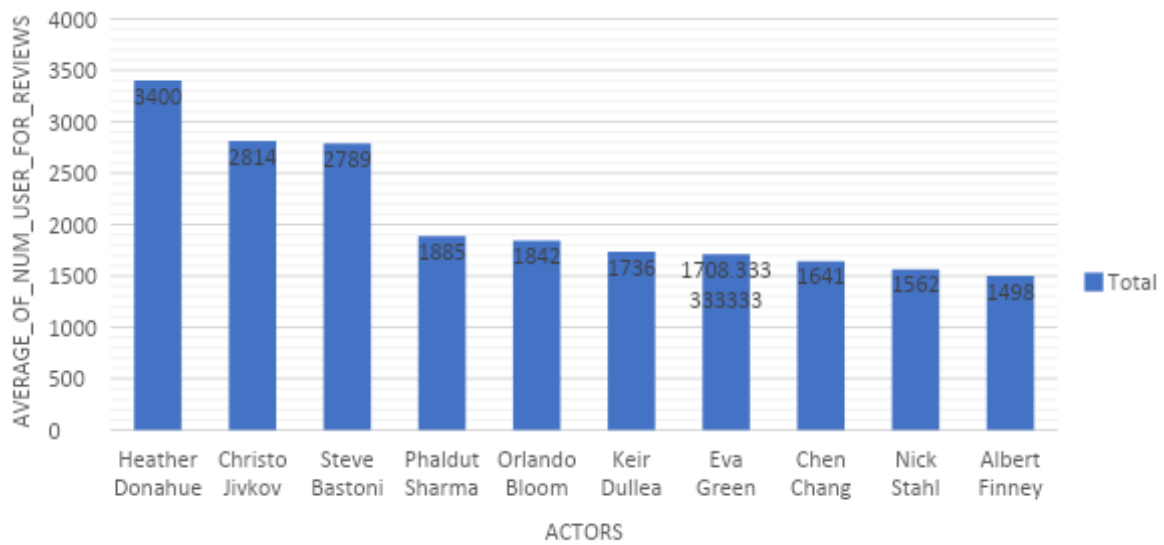
**Grand Total**

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
<b>Grand Total</b>	<b>617.4545455</b>



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	1708.333333
Chen Chang	1641
Nick Stahl	1562
Albert Finney	1498
<b>Grand Total</b>	<b>1946.266667</b>

## TOP 10 ACTORS FOR USER REVIEWS

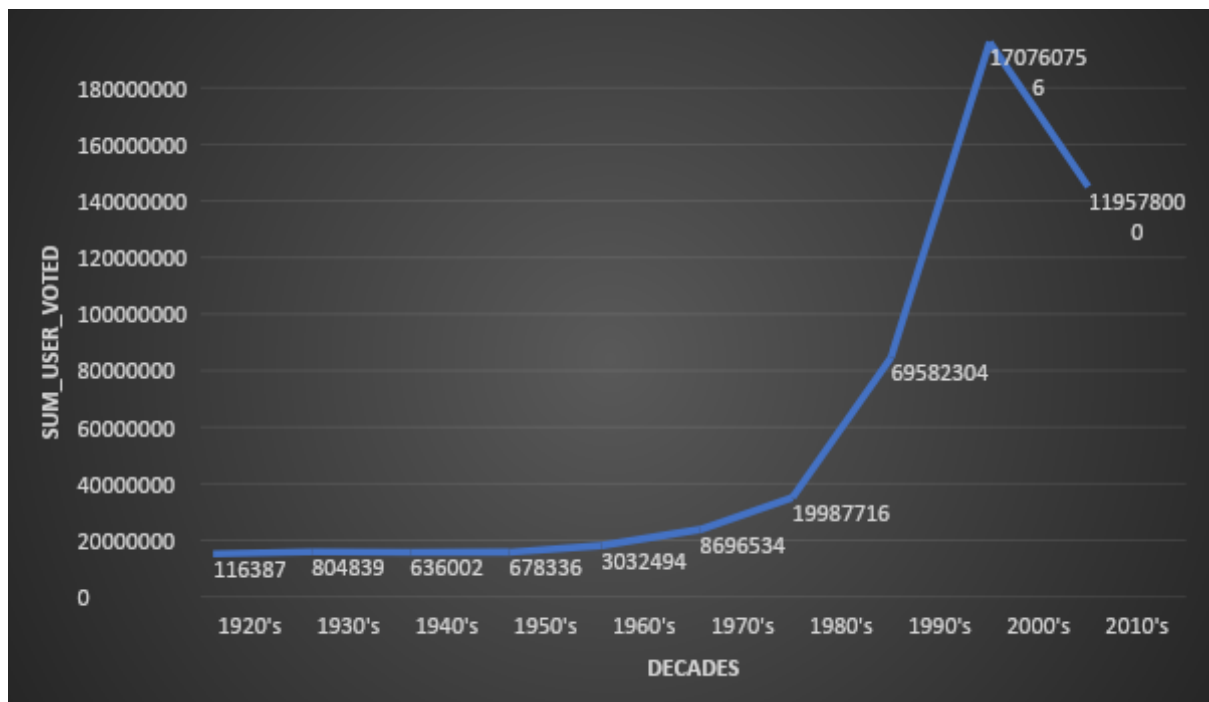


**CRITIC FAVOURITE ACTOR – ALBERT FINNEY**

**USERS FAVOURITE ACTOR- HEATHER DONAHUE**

decades	sum_user_voted
1920's	116387
1930's	804839
1940's	636002
1950's	678336
1960's	3032494
1970's	8696534
1980's	19987716

1990's	69582304
2000's	170760756
2010's	119578000



( WE CAN SEE THE RISE OF USERS VOTING FROM 1960'S , WITH IT'S PEAK IN 2000'S)

THANK YOU