THANK YOU SO MUCH TEAM TRAINITY , 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.

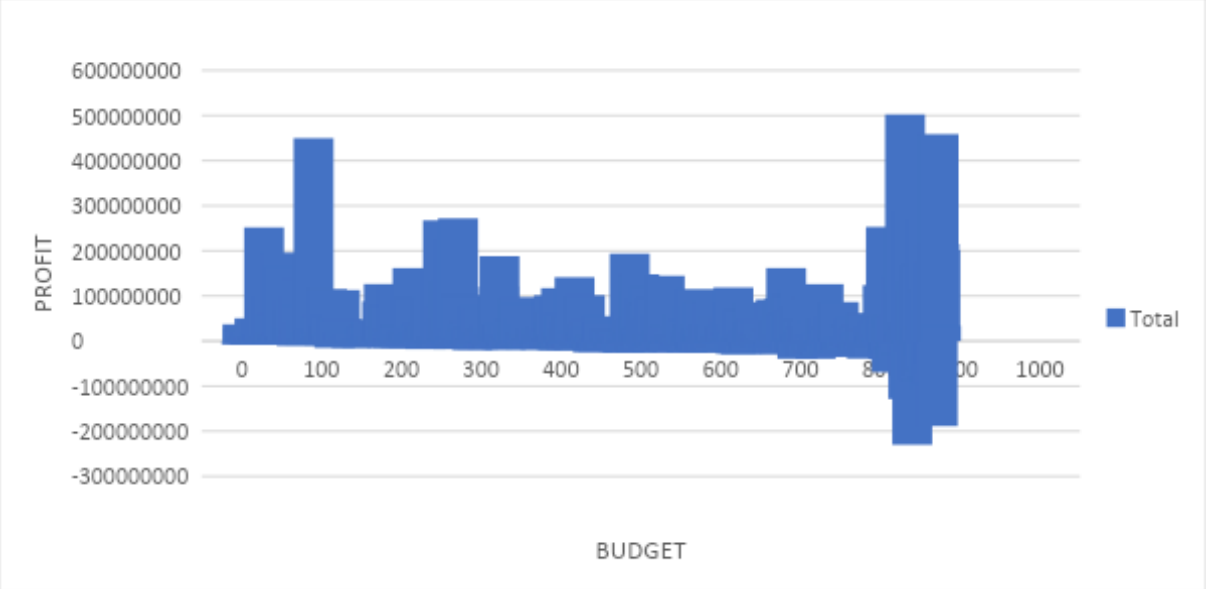
I OBSERVED THE GENRE WHICH WAS MOST LIKED BY USERS i.e. CRIME/DRAMA/FANTASY/MYSTERY. USERS LIKE WATCHING MOVIES IN THESE PARTICULAR GENRE .

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:-

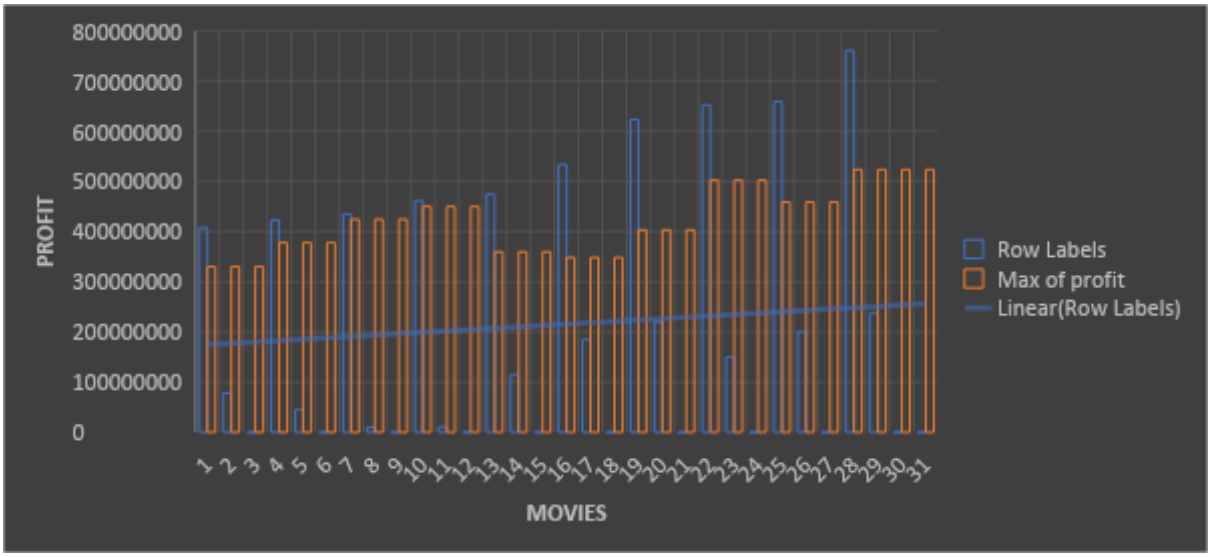
STEP 1 – REMOVING DUPILCATE VALUES

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 GamesÂ	329999255
422783777	**377783777**
45000000	**377783777**
The Lion KingÂ	377783777
434949459	**424449459**
10500000	**424449459**
E.T. the Extra-TerrestrialÂ	424449459
460935665	**449935665**

11000000	**449935665**
Star Wars: Episode IV - A New HopeÂ	449935665
474544677	**359544677**
115000000	**359544677**
Star Wars: Episode I - The Phantom MenaceÂ	359544677
533316061	**348316061**
185000000	**348316061**
The Dark KnightÂ	348316061
623279547	**403279547**
220000000	**403279547**
The AvengersÂ	403279547
652177271	**502177271**
150000000	**502177271**
Jurassic WorldÂ	502177271
658672302	**458672302**
200000000	**458672302**
TitanicÂ	458672302
760505847	**523505847**
237000000	**523505847**
AvatarÂ	523505847
Grand Total	**523505847**



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
2001: A Space OdysseyÂ
427357
8.3
A SeparationÂ
151812
8.4
AlienÂ
563827
8.5
AliensÂ
488537
8.4
AmÃ©lieÂ
534262
8.4
AmadeusÂ
270790
8.3
American BeautyÂ
822500
8.4
American History XÂ
782437
8.6
Apocalypse NowÂ

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

8.7
InceptionÂ
1468200
8.8
Indiana Jones and the Last CrusadeÂ
515306
8.3
Inglourious BasterdsÂ
885175
8.3
Inside OutÂ
345198
8.3
InterstellarÂ
928227
8.6
L.A. ConfidentialÂ
414219
8.3
Lawrence of ArabiaÂ
192775
8.4
MementoÂ
845580
8.5
MetropolisÂ
111841
8.3
Modern TimesÂ
143086
8.6
Monty Python and the Holy GrailÂ
382240
8.3
OldboyÂ
356181
8.4
Once Upon a Time in AmericaÂ
221000
8.4
One Flew Over the Cuckoo's NestÂ
680041
8.7
Princess MononokeÂ
221552
8.4

PsychoÂ
422432
8.5
Pulp FictionÂ
1324680
8.9
Raging BullÂ
235133
8.3
Raiders of the Lost ArkÂ
661017
8.5
Requiem for a DreamÂ
573541
8.4
Reservoir DogsÂ
664719
8.4
RoomÂ
161288
8.3
Saving Private RyanÂ
881236
8.6
ScarfaceÂ
537442
8.3
Schindler's ListÂ
865020
8.9
Se7enÂ
1023511
8.6
Seven SamuraiÂ
229012
8.7
SnatchÂ
600996
8.3
Some Like It HotÂ
175196
8.3
Spirited AwayÂ
417971
8.6
Star Wars: Episode IV - A New HopeÂ

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

8.9
The Lord of the Rings: The Two TowersÂ
1100446
8.7
The MatrixÂ
1217752
8.7
The PianistÂ
497946
8.5
The PrestigeÂ
844052
8.5
The Shawshank RedemptionÂ
1689764
9.3
The Silence of the LambsÂ
887467
8.6
The StingÂ
175607
8.3
The Usual SuspectsÂ
740918
8.6
Toy Story 3Â
544884
8.3
Toy StoryÂ
623757
8.3
UnforgivenÂ
277505
8.3
UpÂ
665575
8.3
WALLÂ·EÂ
718837
8.4
WhiplashÂ
399138
8.5
Grand Total

TOP_FOREIGN_LANG_MOVIES
Row Labels
A SeparationÂ
151812
8.4
Persian
AmÃ©lieÂ
534262
8.4
French
Children of HeavenÂ
27882
8.5
Persian
City of GodÂ
533200
8.7
Portuguese
Das BootÂ
168203
8.4
German
DownfallÂ
248354
8.3
German
MetropolisÂ
111841
8.3
German
OldboyÂ
356181
8.4
Korean
Princess MononokeÂ
221552
8.4
Japanese
Seven SamuraiÂ
229012
8.7
Japanese
Spirited AwayÂ
417971
8.6
Japanese

The Good, the Bad and the UglyÂ
503509
8.9
Italian
The HuntÂ
170155
8.3
Danish
The Lives of OthersÂ
259379
8.5
German
Grand Total

THE ABOVE GIVEN LIST IS OF TOP 250 IMDB SCORED MOVIES

WITH DIFFERENT LANGUAGE MOVIES ALSO. BUT ENGLISH AS 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

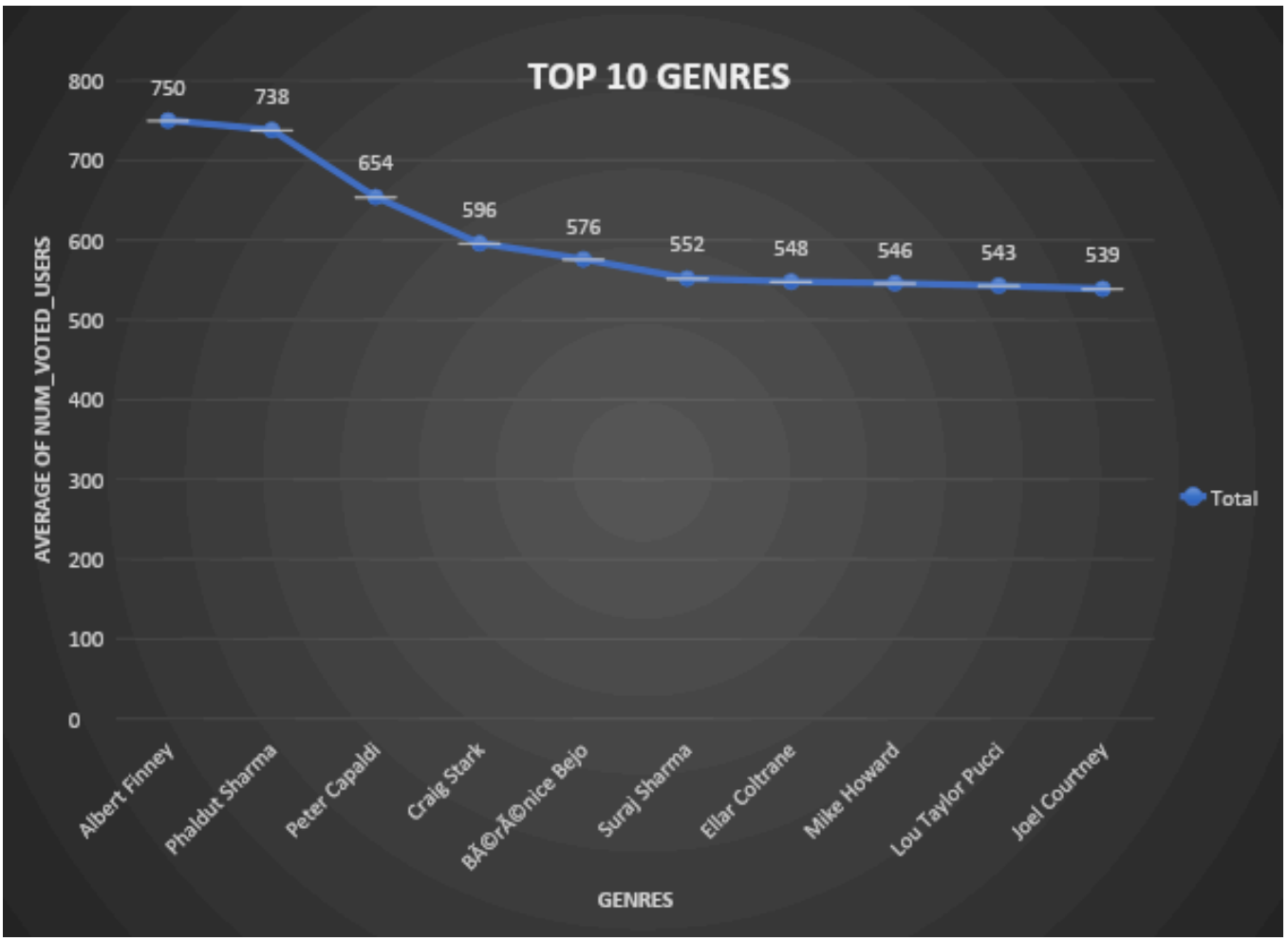
	AVG_IMDB_SCORE	DIRECTORS
1	8.7	AKIRA KUROSAWA
2	8.6	CHARLES CHAPLIN
3	8.6	MICHAEL CURTIZ
4	8.6	TONY KAYE
5	8.5	DAMIEN CHAZELLE
6	8.5	MAJID MAJIDI
7	8.5	RON FRICKE
8	8.433333	SERGIO LEONE
9	8.425	CHRISTOPHER NOLAN
10	8.4	ASGHAR FARHADI

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
Grand Total	**531323.4**	



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 CompanionÂ
Meryl Streep
BabelÂ
Brad Pitt
Blood DiamondÂ
Leonardo DiCaprio
Body of LiesÂ
Leonardo DiCaprio
By the SeaÂ
Brad Pitt
Catch Me If You CanÂ
Leonardo DiCaprio
Django UnchainedÂ
Leonardo DiCaprio
Fight ClubÂ
Brad Pitt
FuryÂ
Brad Pitt
Gangs of New YorkÂ
Leonardo DiCaprio
Hope SpringsÂ
Meryl Streep
InceptionÂ
Leonardo DiCaprio
Interview with the Vampire: The Vampire ChroniclesÂ
Brad Pitt
It's ComplicatedÂ
Meryl Streep
J. EdgarÂ
Leonardo DiCaprio
Julie & JuliaÂ
Meryl Streep
Killing Them SoftlyÂ

Brad Pitt

Lions for LambsÂ

Meryl Streep

Marvin's RoomÂ

Leonardo DiCaprio

Mr. & Mrs. SmithÂ

Brad Pitt

Ocean's ElevenÂ

Brad Pitt

Ocean's TwelveÂ

Brad Pitt

One True ThingÂ

Meryl Streep

Out of AfricaÂ

Meryl Streep

Revolutionary RoadÂ

Leonardo DiCaprio

Romeo + JulietÂ

Leonardo DiCaprio

Seven Years in TibetÂ

Brad Pitt

Shutter IslandÂ

Leonardo DiCaprio

Sinbad: Legend of the Seven SeasÂ

Brad Pitt

Spy GameÂ

Brad Pitt

The Assassination of Jesse James by the Coward Robert FordÂ

Brad Pitt

The AviatorÂ

Leonardo DiCaprio

The BeachÂ

Leonardo DiCaprio

The Curious Case of Benjamin ButtonÂ

Brad Pitt

The DepartedÂ

Leonardo DiCaprio

The Devil Wears PradaÂ

Meryl Streep

The Great GatsbyÂ

Leonardo DiCaprio

The HoursÂ

Meryl Streep

The Iron LadyÂ

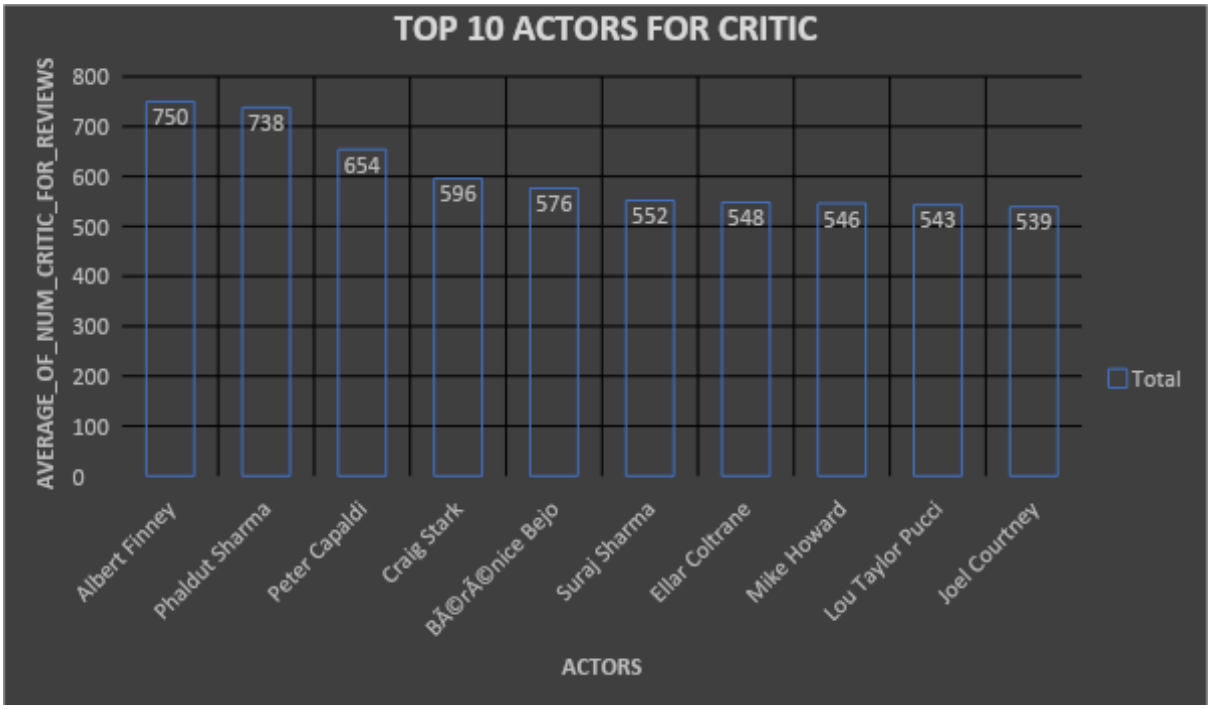
Meryl Streep

The Man in the Iron MaskÂ

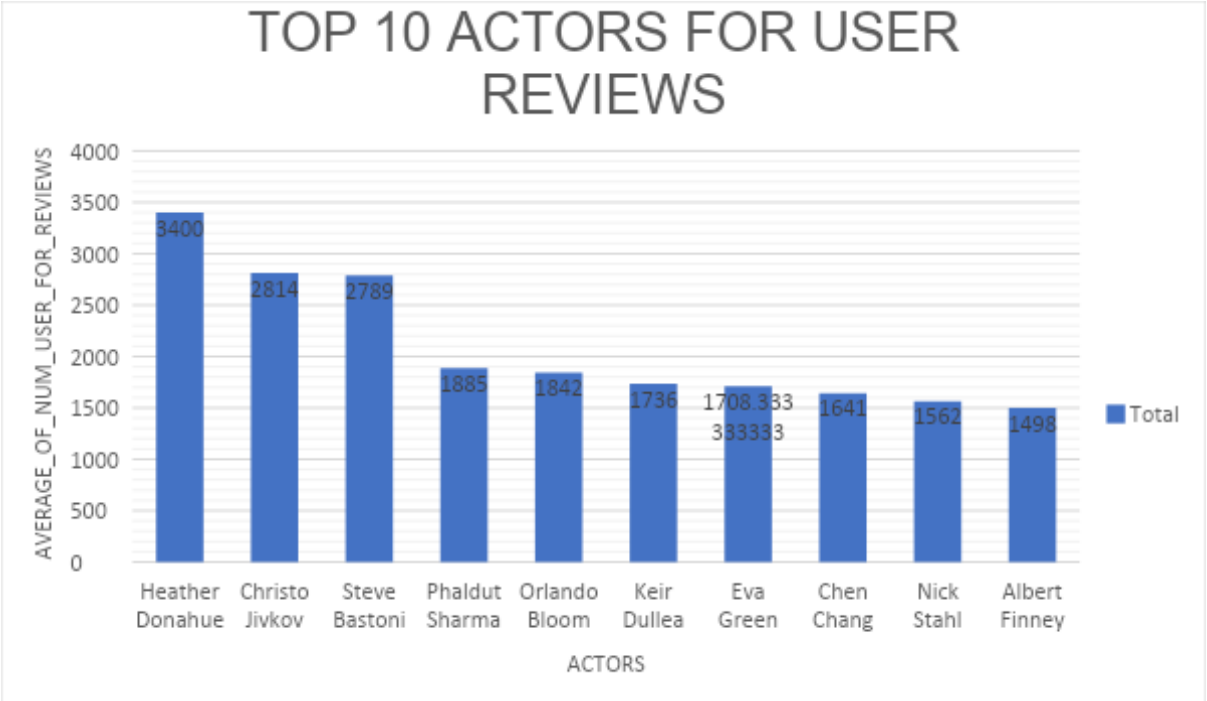
Leonardo DiCaprio

The Quick and the DeadÂ	
Leonardo DiCaprio	
The RevenantÂ	
Leonardo DiCaprio	
The River WildÂ	
Meryl Streep	
The Tree of LifeÂ	
Brad Pitt	
The Wolf of Wall StreetÂ	
Leonardo DiCaprio	
TitanicÂ	
Leonardo DiCaprio	
TroyÂ	
Brad Pitt	
True RomanceÂ	
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
Grand Total	**617.4545455**



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
Grand Total	**1946.266667**

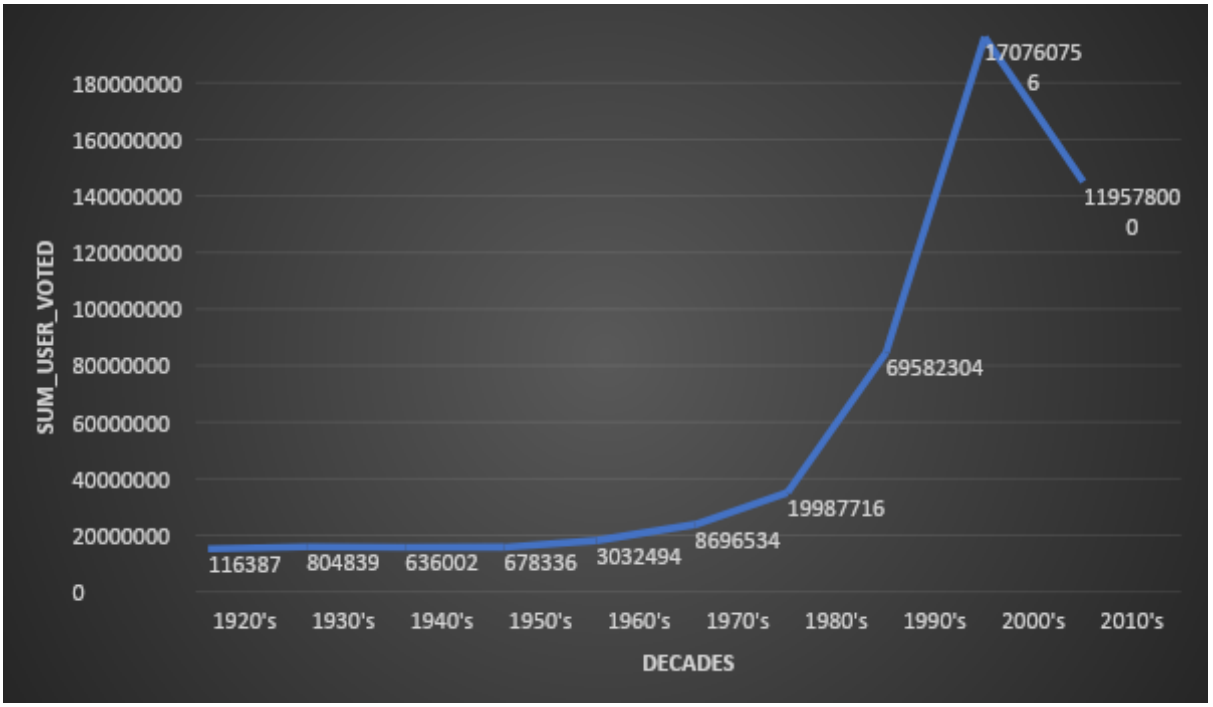


CRITIC FAVOURITE ACTOR – ALBERT FINNEY

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