

IMDB Movie Analysis

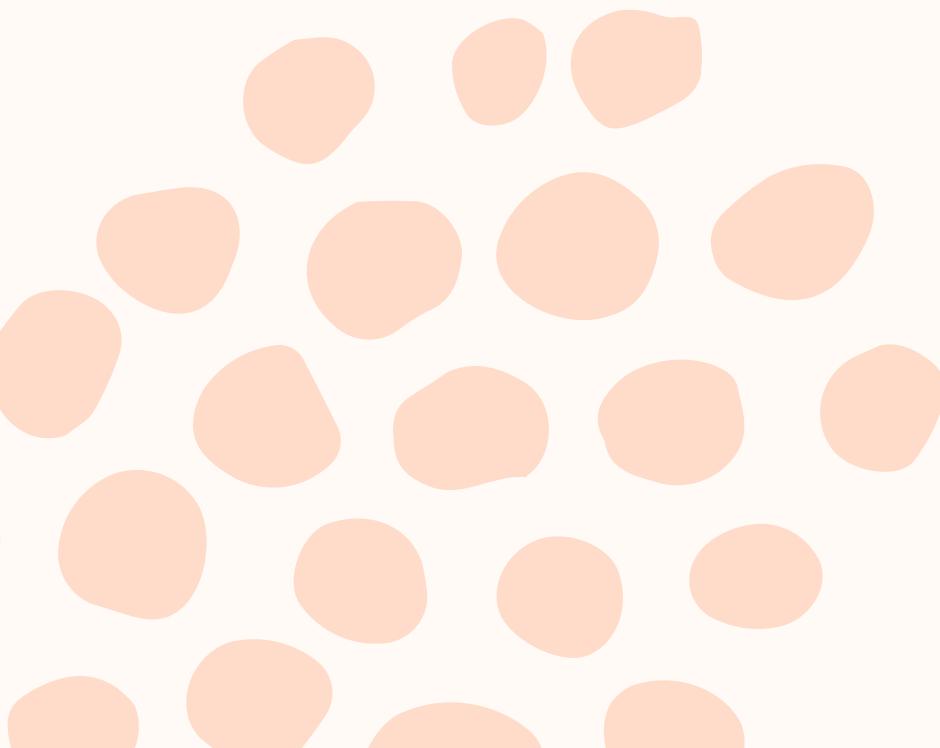
PREETI KUMARI

Contents

1. Project Description
2. Approach
3. Tech- Stack analysis
4. Insights
5. Results



PROJECT DESCRIPTION



Introduction

This project involves analysis of dataset having various columns of different IMDB Movies.

1. **color**
2. **director_name**
3. **num_critic_for_reviews**
4. **duration**
5. **director_facebook_likes**
6. **actor_3_facebook_likes**
7. **actor_2_name**
8. **actor_1_facebook_likes**
9. **gross**
10. **genres**
11. **actor_1_name**
12. **movie_title**
13. **num_voted_users**
14. **cast_total_facebook_likes**
15. **actor_3_name**
16. **facenumber_in_poster**
17. **plot_keywords**
18. **movie_imdb_link**
19. **num_user_for_reviews**
20. **language**
21. **country**
22. **content_rating**
23. **budget**
24. **title_year**
25. **actor_2_facebook_likes**
26. **imdb_score**
27. **aspect_ratio,**
28. **movie_facebook_likes**

Objective

OUR OBJECTIVE

- First objective is to perform exploratory data analysis which involves identifying the redundant data which is not required for analysis, identifying the missing or the null values in the dataset, removing the duplicates and identification of outliers.
- Find out the movies with the highest profit.
- Find out the IMDB Top 250.
- Find out the best directors
- Find out the popular genres.
- Find out the critic-favorite and audience-favorite actors.
- Observe the change in number of voted users over decades using a bar chart.

APPROACH



1. Cleaning the data:: This is one of the most important step to perform before moving forward with the analysis

- Removing the unwanted columns.
- Identifying and remove the missing values or null values from the datasets
- Identify and remove duplicates.

2. Movies with highest profit: Creating a new column called profit which contains the difference of the two columns: gross and budget then Sorting the column using the profit column as reference. Plot profit (y-axis) vs budget (x- axis) and observing the outliers using Scatter Plot and movies with higher profit using Bar chart/Column Chart.

3. Top 250:

- Creating a new column IMDb_Top_250 and store the top 250 movies with the highest IMDb Rating (corresponding to the column: imdb_score). As per given criteria on all of these movies, the num_voted_users is greater than 25,000.
- Addition of a Rank column containing the values 1 to 250 indicating the ranks of the corresponding films.
- Extracting 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.

4. Best Directors: Grouping 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 top 10 director.

5. Popular Genres: Find out the popular top 10 genres

6. Charts: Using charts Observe the change in number of voted users over decades and the critic-favorite and audience-favorite actors

TECH STACK



MICROSOFT EXCEL



The background features several abstract, rounded organic shapes in various colors: a large dark teal shape on the left, a smaller orange shape below it, and a cluster of light green, orange, and pink shapes at the bottom left. A large, semi-transparent orange circle is positioned behind the word 'INSIGHTS'.

INSIGHTS

DATA CLEANING

DATA PRE-PROCESSING

This is one of the most important step to perform before moving forward with the analysis. It involves

- Identifying the missing/Null values.
- Identifying the duplicate values.
- Dropping irrelevant Columns.

COLUMN DROPPED

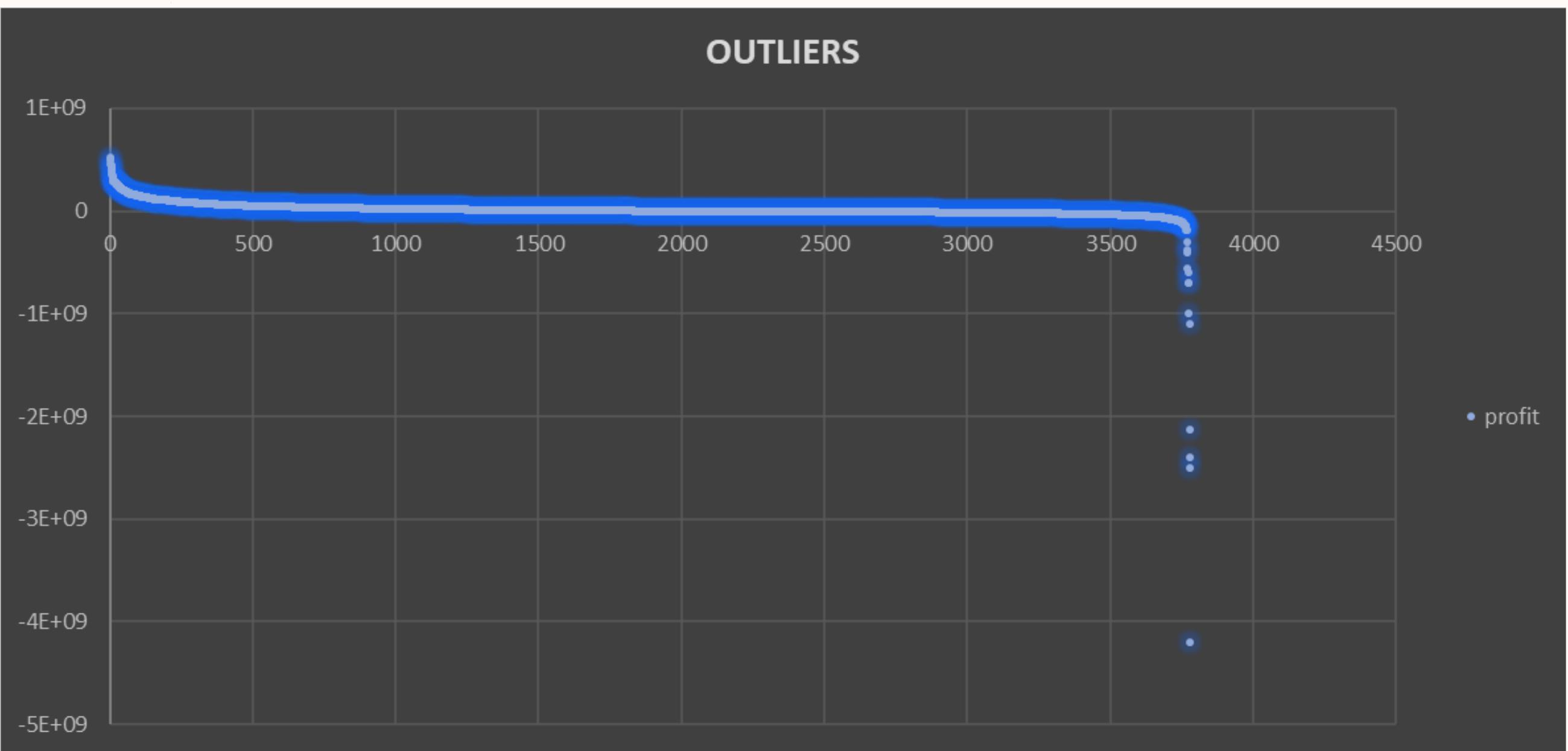
1. 'Color',
2. 'director_facebook_likes'
3. 'actor_3_facebook_likes',
4. 'actor_2_name',
5. 'actor_1_facebook_likes',
6. 'cast_total_facebook_likes',
7. 'actor_3_name',
8. 'facenumber_in_posts',
9. 'plot_keywords',
10. 'movie_imdb_link',
11. 'content_rating'
12. 'actor_2_facebook_likes',
13. 'aspect_ratio',
14. 'movie_facebook_likes'

DATA CLEANING

Total Number of rows	5044
Total no. of missing values	1160
No. of Rows after removing missing values	3884
Duplicate Values(Movie_Title)	102
No. of rows after removing duplicates	3782

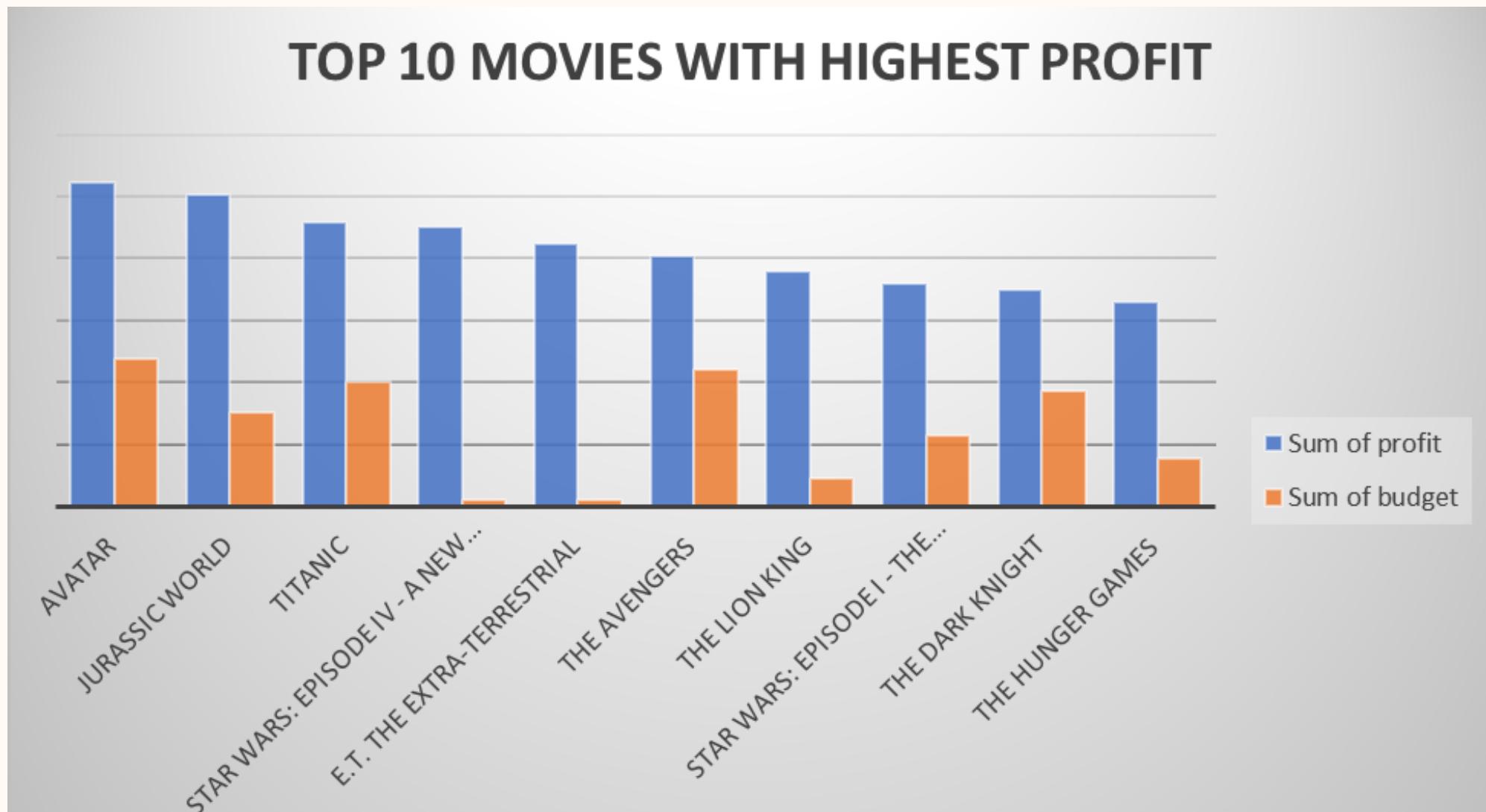
1.(a).Identifying the outliers using Scatter Plot

OUTLIERS



MIN	-4199788333
MAX	523505847
QUARTILE 1	-10217492
QUARTILE 2	794341
QUARTILE 3	24335698
Q3-Q1	34553190
UPPER BOUND	76165483
LOWER BOUND	-62047277

1.(b.) Find the movies with the highest profit



director_name	num_critic_for_reviews	duration	gross	genres	actor_1_name	movie_title	num_voted_users	num_user_for_reviews	language	country	budget	title_year	profit	imdb_score	outliers
James Cameron	723	178	760505847	Action Adventure Fantasy Sci-Fi	CCH Pounder	Avatar	886204	3054	English	USA	237000000	2009	523505847	7.9	TRUE
Colin Trevorrow	644	124	652177271	Action Adventure Sci-Fi Thriller	Bryce Dallas Howard	Jurassic World	418214	1290	English	USA	150000000	2015	502177271	7	TRUE
James Cameron	315	194	658672302	Drama Romance	Leonardo DiCaprio	Titanic	793059	2528	English	USA	200000000	1997	458672302	7.7	TRUE
George Lucas	282	125	460935665	Action Adventure Fantasy Sci-Fi	Harrison Ford	Star Wars: Episode IV - A New Hope	911097	1470	English	USA	11000000	1977	449935665	8.7	TRUE
Steven Spielberg	215	120	434949459	Family Sci-Fi	Henry Thomas	E.T. the Extra-Terrestrial	281842	515	English	USA	10500000	1982	424449459	7.9	TRUE
Joss Whedon	703	173	623279547	Action Adventure Sci-Fi	Chris Hemsworth	The Avengers	995415	1722	English	USA	220000000	2012	403279547	8.1	TRUE
Roger Allers	186	73	422783777	Adventure Animation Drama Family Musical	Matthew Broderick	The Lion King	644348	656	English	USA	45000000	1994	377783777	8.5	TRUE
George Lucas	320	136	474544677	Action Adventure Fantasy Sci-Fi	Natalie Portman	Star Wars: Episode I - The Phantom Menace	534658	3597	English	USA	115000000	1999	359544677	6.5	TRUE
Christopher Nolan	645	152	533316061	Action Crime Drama Thriller	Christian Bale	The Dark Knight	1676169	4667	English	USA	185000000	2008	348316061	9	TRUE
Gary Ross	673	142	407999255	Adventure Drama Sci-Fi Thriller	Jennifer Lawrence	The Hunger Games	701607	1959	English	USA	78000000	2012	329999255	7.3	TRUE

2. IMDB Top 250 movies

Top 10 movies in different languages (including English)

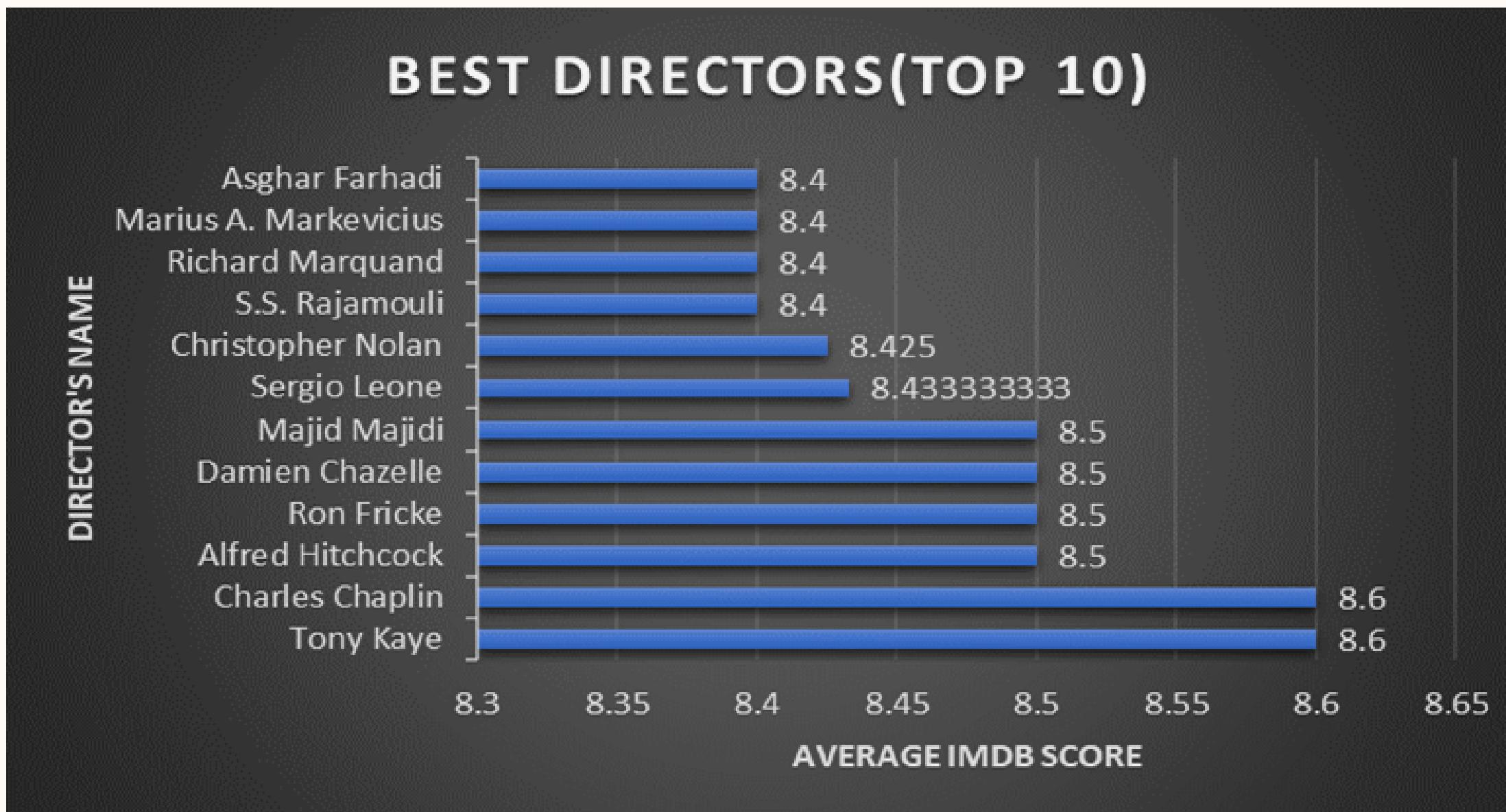
director_name	num_critic_for_reviews	duration	gross	genres	actor_1_name	movie_title	num_voted_users	num_user_for_review	language	country	budget	title_year	RANK	imdb_score
Frank Darabont	199	142	28341469	Crime Drama	Morgan Freeman	The Shawshank Redemption	1689764	4144	English	USA	25000000	1994	1	9.3
Francis Ford Coppola	208	175	134821952	Crime Drama	Al Pacino	The Godfather	1155770	2238	English	USA	6000000	1972	2	9.2
Christopher Nolan	645	152	533316061	Action Crime Drama Thriller	Christian Bale	The Dark Knight	1676169	4667	English	USA	185000000	2008	3	9
Francis Ford Coppola	149	220	57300000	Crime Drama	Robert De Niro	The Godfather: Part II	790926	650	English	USA	13000000	1974	4	9
Peter Jackson	328	192	377019252	Action Adventure Drama Fantasy	Orlando Bloom	The Lord of the Rings: The Return of the King	1215718	3189	English	USA	94000000	2003	5	8.9
Steven Spielberg	174	185	96067179	Biography Drama History	Liam Neeson	Schindler's List	865020	1273	English	USA	22000000	1993	6	8.9
Quentin Tarantino	215	178	107930000	Crime Drama	Bruce Willis	Pulp Fiction	1324680	2195	English	USA	8000000	1994	7	8.9
Sergio Leone	181	142	6100000	Western	Clint Eastwood	The Good, the Bad and the Ugly	503509	780	Italian	Italy	1200000	1966	8	8.9
Christopher Nolan	642	148	292568851	Action Adventure Sci-Fi	Leonardo DiCaprio	Inception	1468200	2803	English	USA	160000000	2010	9	8.8
Peter Jackson	297	171	313837577	Action Adventure Drama Fantasy	Christopher Lee	The Lord of the Rings: The Fellowship of the Ring	1238746	5060	English	New Zealand	93000000	2001	10	8.8

Top 10 movies in different languages (Except English)

director_name	num_critic_for_reviews	duration	gross	genres	actor_1_name	movie_title	num_voted_users	num_user_for_review	language	country	budget	title_year	RANK	imdb_score
Sergio Leone	181	142	6100000	Western	Clint Eastwood	The Good, the Bad and the Ugly	503509	780	Italian	Italy	1200000	1966	8	8.9
Fernando Meirelles	214	135	7563397	Crime Drama	Alice Braga	City of God	533200	749	Portuguese	Brazil	3300000	2002	19	8.7
Akira Kurosawa	153	202	269061	Action Adventure Drama	Takashi Shimura	Seven Samurai	229012	596	Japanese	Japan	2000000	1954	20	8.7
Hayao Miyazaki	246	125	10049886	Fantasy	Bunta Sugawara	Spirited Away	417971	902	Japanese	Japan	19000000	2001	25	8.6
Florian Henckel von Donnersmark	215	137	11284657	Drama Thriller	Sebastian Koch	The Lives of Others	259379	407	German	Germany	2000000	2006	45	8.5
Majid Majidi	46	89	925402	Drama Family	Bahare Seddiqi	Children of Heaven	27882	130	Persian	Iran	180000	1997	46	8.5
Jean-Pierre Jeunet	242	122	33201661	Comedy Romance	Mathieu Kassovitz	Amélie	534262	1314	French	France	77000000	2001	49	8.4
S.S. Rajamouli	44	159	6498000	Action Adventure Drama Fantasy	Tamannaah Bhatia	Baahubali: The Beginning	62756	410	Telugu	India	18026148	2015	50	8.4
Hayao Miyazaki	174	134	2298191	Adventure Animation Fantasy	Minnie Driver	Princess Mononoke	221552	570	Japanese	Japan	2400000000	1997	53	8.4
Wolfgang Petersen	96	293	11433134	Adventure Drama Thriller War	Jürgen Prochnow	Das Boot	168203	426	German	West Germany	14000000	1981	57	8.4

3. Find Best Directors

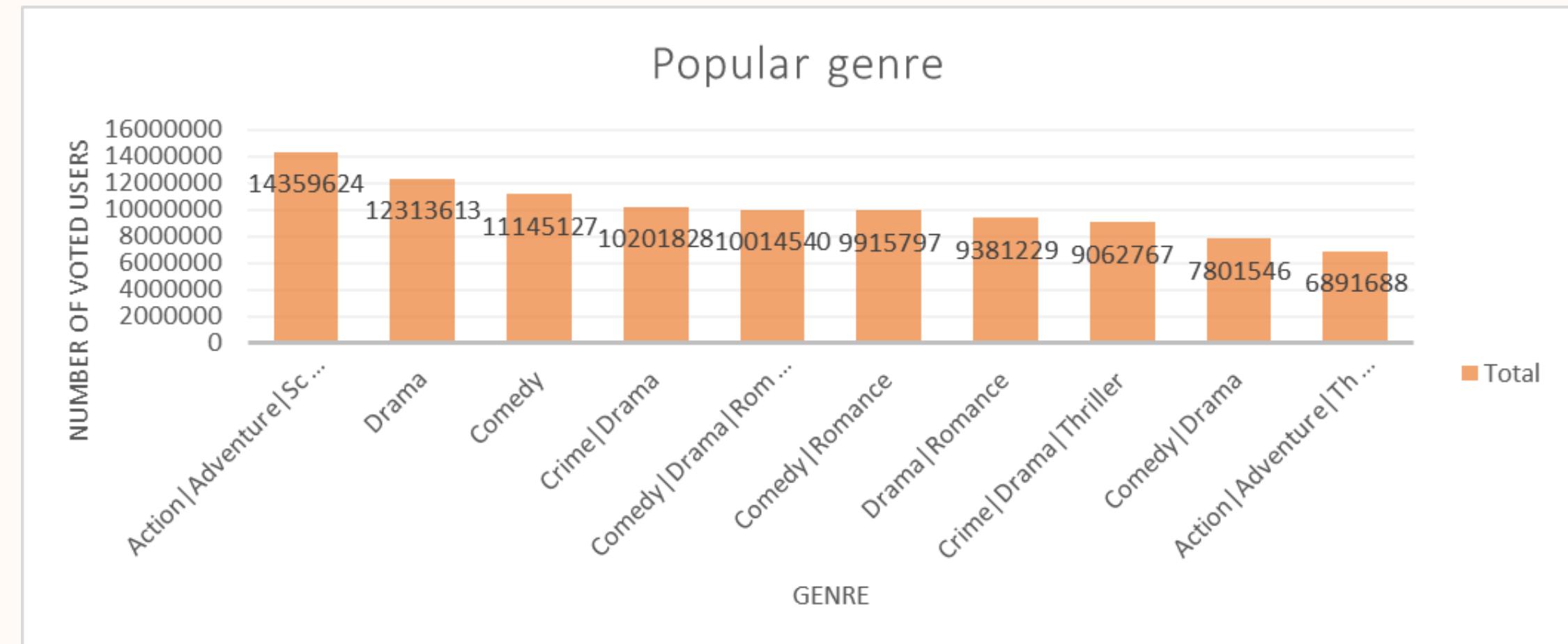
Best Director	Average of imdb_score
Tony Kaye	8.6
Charles Chaplin	8.6
Alfred Hitchcock	8.5
Ron Fricke	8.5
Damien Chazelle	8.5
Majid Majidi	8.5
Sergio Leone	8.433333333
Christopher Nolan	8.425
S.S. Rajamouli	8.4
Richard Marquand	8.4
Marius A. Markevicius	8.4
Asghar Farhadi	8.4



4. Popular genre

Top 10 movies in different languages (including English)

Popular genre	Sum of num_voted_users
Action Adventure Sci-Fi	14359624
Drama	12313613
Comedy	11145127
Crime Drama	10201828
Comedy Drama Romance	10014540
Comedy Romance	9915797
Drama Romance	9381229
Crime Drama Thriller	9062767
Comedy Drama	7801546
Action Adventure Thriller	6891688



5. Find the critic favorite and audience favourite actors

Brad pitt

director_name	num_critic_for_reviews	duration	gross	genres	actor_1_name	movie_title	num_voted_users	num_user_for_review	language	country	budget	title_year	RANK	imdb_score
David Fincher	315	151	37023395	Drama	Brad Pitt	Fight Club	1347461	2968	English	USA	63000000	1999	11	8.8
Tony Scott	122	121	12281500	Action Crime Drama Romance Thriller	Brad Pitt	True Romance	163492	460	English	USA	13000000	1993	184	8
David Fincher	362	166	127490802	Drama Fantasy Romance	Brad Pitt	The Curious Case of Benjamin Button	459346	822	English	USA	150000000	2008	266	7.8
Steven Soderbergh	186	116	183405771	Crime Thriller	Brad Pitt	Ocean's Eleven	402645	845	English	USA	85000000	2001	266	7.8
David Ayer	406	134	85707116	Action Drama War	Brad Pitt	Fury	303185	701	English	USA	68000000	2014	438	7.6
Neil Jordan	120	123	105264608	Drama Fantasy Horror	Brad Pitt	Interview with the Vampire: The Vampire L	239752	406	English	USA	60000000	1994	438	7.6
Andrew Dominik	273	160	3904982	Biography Crime Drama History Western	Brad Pitt	The Assassination of Jesse James by the C	136104	415	English	USA	30000000	2007	539	7.5
Alejandro G. Iñárritu	285	143	34300771	Drama	Brad Pitt	Babel	243799	908	English	France	25000000	2006	539	7.5
Wolfgang Petersen	220	196	133228348	Adventure	Brad Pitt	Troy	381672	1694	English	USA	175000000	2004	869	7.2
Tony Scott	142	114	26871	Action Crime Thriller	Brad Pitt	Spy Game	121259	361	English	Germany	92000000	2001	1149	7
Jean-Jacques Annaud	76	136	37901509	Adventure Biography Drama History War	Brad Pitt	Seven Years in Tibet	96385	119	English	USA	70000000	1997	1149	7
Patrick Gilmore	98	85	26288320	Science Fiction Drama Family Fantasy Roma	Brad Pitt	Sinbad: Legend of the Seven Seas	36144	91	English	USA	60000000	2003	1571	6.7
Terrence Malick	584	139	13303319	Drama Fantasy	Brad Pitt	The Tree of Life	136367	975	English	USA	32000000	2011	1571	6.7
Doug Liman	233	126	186336103	Action Comedy Crime Romance Thriller	Brad Pitt	Mr. & Mrs. Smith	348861	798	English	USA	120000000	2005	1910	6.5
Steven Soderbergh	198	125	125531634	Crime Thriller	Brad Pitt	Ocean's Twelve	284852	627	English	USA	110000000	2004	2066	6.4
Andrew Dominik	414	97	14938570	Crime Thriller	Brad Pitt	Killing Them Softly	111625	369	English	USA	15000000	2012	2355	6.2

Find the critic favorite and audience favourite actors

Meryl Streep

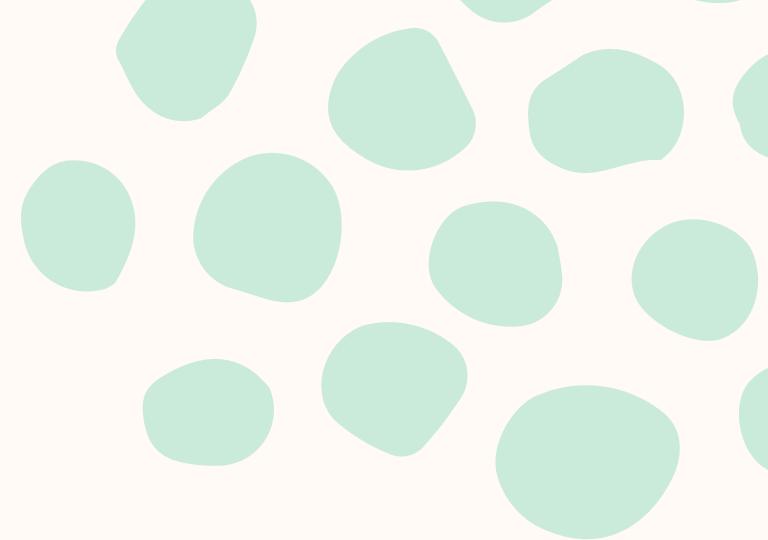
director_name	num_critic_for_reviews	duration	gross	genres	actor_1_name	movie_title	num_voted_users	num_user_for_review	language	country	budget	title_year	RANK	imdb_score
Stephen Daldry	174	114	41597830	Drama Romance	Meryl Streep	The Hours	102123	660	English	USA	25000000	2002	438	7.6
Sydney Pollack	66	161	87100000	Biography Drama Romance	Meryl Streep	Out of Africa	52339	200	English	USA	31000000	1985	869	7.2
Nora Ephron	252	123	94125426	Biography Drama Romance	Meryl Streep	Julie & Julia	79264	277	English	USA	40000000	2009	1149	7
David Frankel	208	109	124732962	Comedy Drama Romance	Meryl Streep	The Devil Wears Prada	286178	631	English	USA	35000000	2006	1427	6.8
Nancy Meyers	187	120	112703470	Comedy Drama Romance	Meryl Streep	It's Complicated	69860	214	English	USA	85000000	2009	1745	6.6
Phyllida Lloyd	331	105	29959436	Biography Drama History	Meryl Streep	The Iron Lady	82327	350	English	UK	13000000	2011	2066	6.4
Curtis Hanson	42	111	46815748	Action Adventure Crime Thrill	Meryl Streep	The River Wild	32544	69	English	USA	45000000	1994	2217	6.3
David Frankel	234	100	63536011	Comedy Drama Romance	Meryl Streep	Hope Springs	34258	178	English	USA	30000000	2012	2217	6.3
Robert Redford	227	92	14998070	Drama Thriller War	Meryl Streep	Lions for Lambs	41170	298	English	USA	35000000	2007	2355	6.2

Find the critic favorite and audience favourite actor

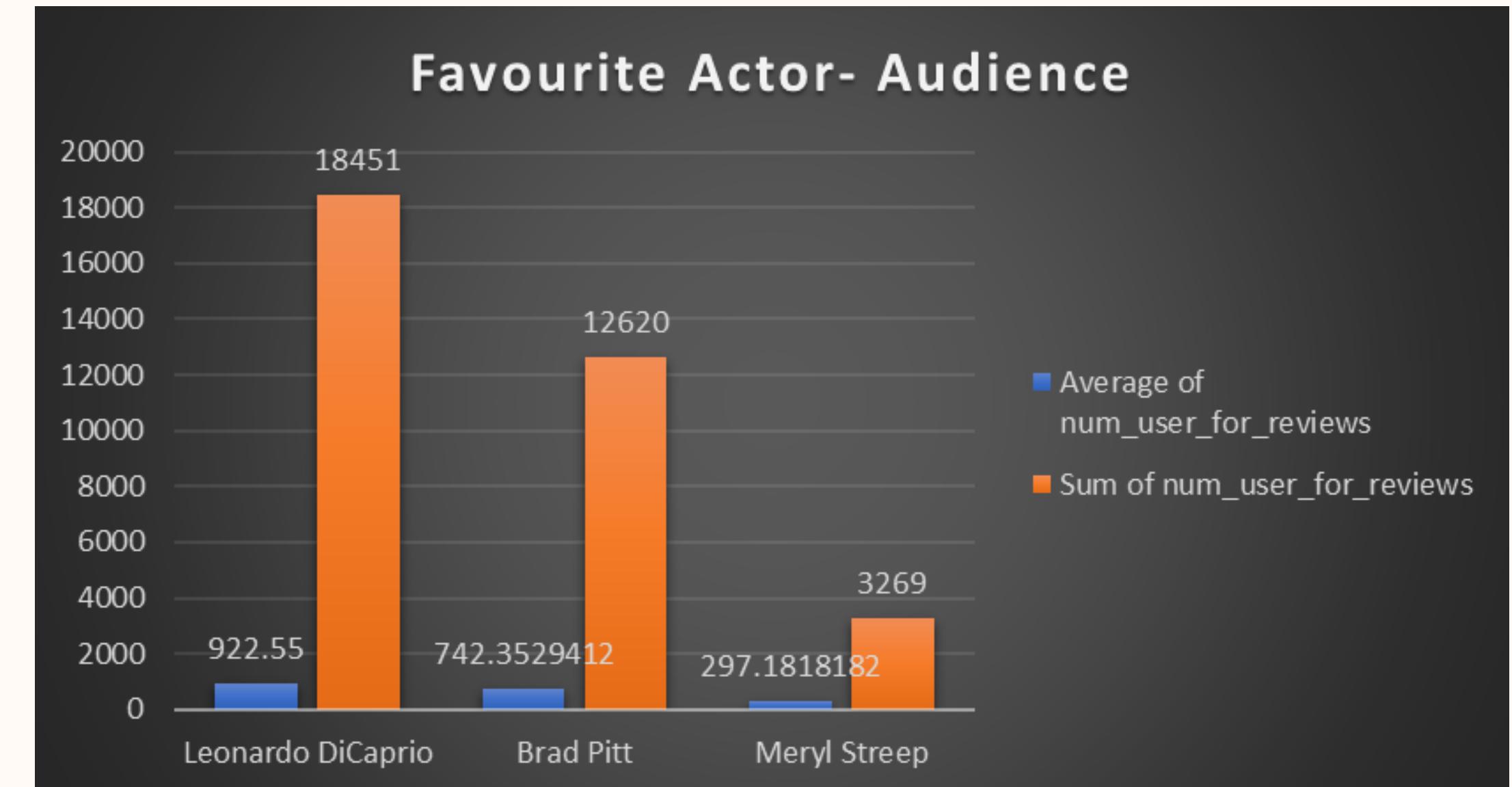
Leonardo DiCaprio

director_name	num_critic_for_reviews	duration	gross	genres	actor_1_name	movie_title	num_voted_users	num_user_for_review	language	country	budget	title_year	RANK	imdb_score
Christopher Nolan	642	148	292568851	Action Adventure Sci-	Leonardo DiCaprio	Inception	1468200	2803	English	USA	160000000	2010	9	8.8
Quentin Tarantino	765	165	162804648	Drama Western	Leonardo DiCaprio	Django Unchained	955174	1193	English	USA	100000000	2012	32	8.5
Martin Scorsese	352	151	132373442	Crime Drama Thriller	Leonardo DiCaprio	The Departed	873649	2054	English	USA	90000000	2006	33	8.5
Martin Scorsese	606	240	116866727	Biography Comedy Crime Dra	Leonardo DiCaprio	The Wolf of Wall Street	780588	1138	English	USA	100000000	2013	87	8.2
Alejandro G. Iñárritu	556	156	183635922	Adventure Drama Thriller We	Leonardo DiCaprio	The Revenant	406020	1188	English	USA	135000000	2015	111	8.1
Martin Scorsese	490	138	127968405	Mystery Thriller	Leonardo DiCaprio	Shutter Island	786092	964	English	USA	80000000	2010	116	8.1
Edward Zwick	166	143	57366262	Adventure Drama Thriller	Leonardo DiCaprio	Blood Diamond	400292	657	English	Germany	100000000	2006	158	8
Steven Spielberg	194	141	164435221	Biography Crime Drama	Leonardo DiCaprio	Catch Me If You Can	525801	667	English	USA	52000000	2002	164	8
James Cameron	315	194	658672302	Drama Romance	Leonardo DiCaprio	Titanic	793059	2528	English	USA	200000000	1997	347	7.7
Martin Scorsese	267	170	102608827	Biography Drama	Leonardo DiCaprio	The Aviator	264318	799	English	USA	110000000	2004	539	7.5
Martin Scorsese	233	216	77679638	Crime Drama	Leonardo DiCaprio	Gangs of New York	314033	1166	English	USA	100000000	2002	539	7.5
Baz Luhrmann	490	143	144812796	Drama Romance	Leonardo DiCaprio	The Great Gatsby	362912	753	English	Australia	105000000	2013	729	7.3
Sam Mendes	323	119	22877808	Drama Romance	Leonardo DiCaprio	Revolutionary Road	152591	414	English	USA	35000000	2008	729	7.3
Ridley Scott	238	128	39380442	Action Drama Thriller	Leonardo DiCaprio	Body of Lies	174248	263	English	USA	70000000	2008	1008	7.1
Baz Luhrmann	106	120	46338728	Drama Romance	Leonardo DiCaprio	Romeo + Juliet	167750	506	English	USA	14500000	1996	1427	6.8
Danny Boyle	118	119	39778599	Adventure Drama Thriller	Leonardo DiCaprio	The Beach	176169	548	English	USA	50000000	2000	1745	6.6
Clint Eastwood	392	137	37304950	Biography Crime Drama	Leonardo DiCaprio	J. Edgar	102728	279	English	USA	35000000	2011	1745	6.6
Randall Wallace	83	132	56876365	Action Adventure	Leonardo DiCaprio	The Man in the Iron Mask	125219	244	English	USA	35000000	1998	2066	6.4
Sam Raimi	63	107	18636537	Action Thriller Western	Leonardo DiCaprio	The Quick and the Dead	69197	216	English	Japan	32000000	1995	2066	6.4

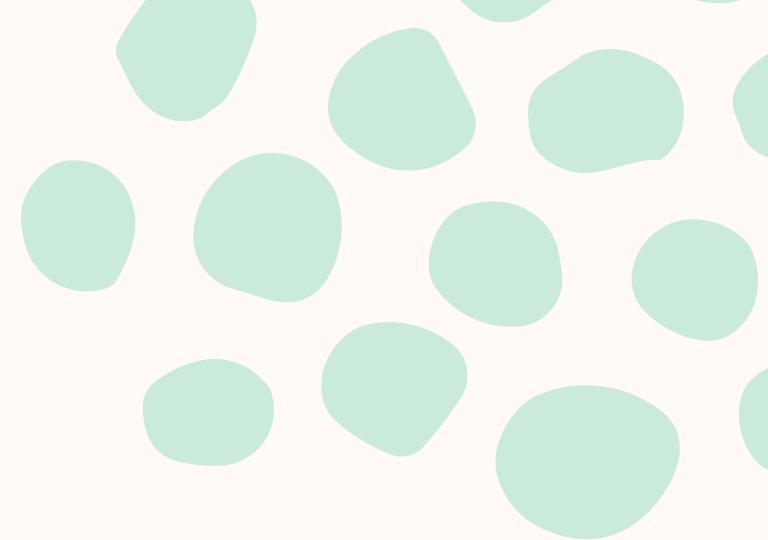
Favourite Actor – Audience



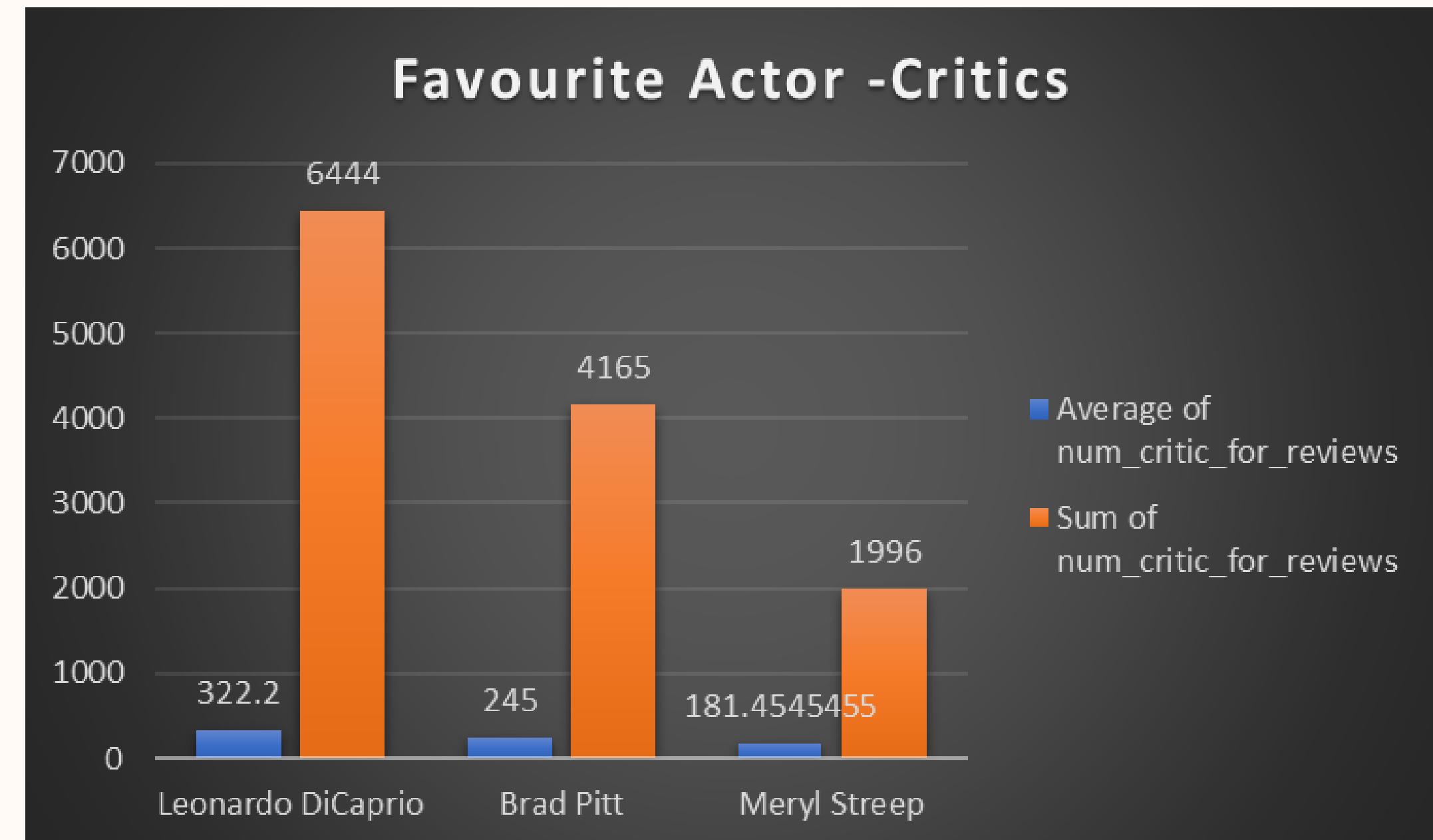
Actor Name	Average of num_user_for_reviews	Sum of num_user_for_reviews
Leonardo DiCaprio	922.55	18451
Brad Pitt	742.3529412	12620
Meryl Streep	297.1818182	3269



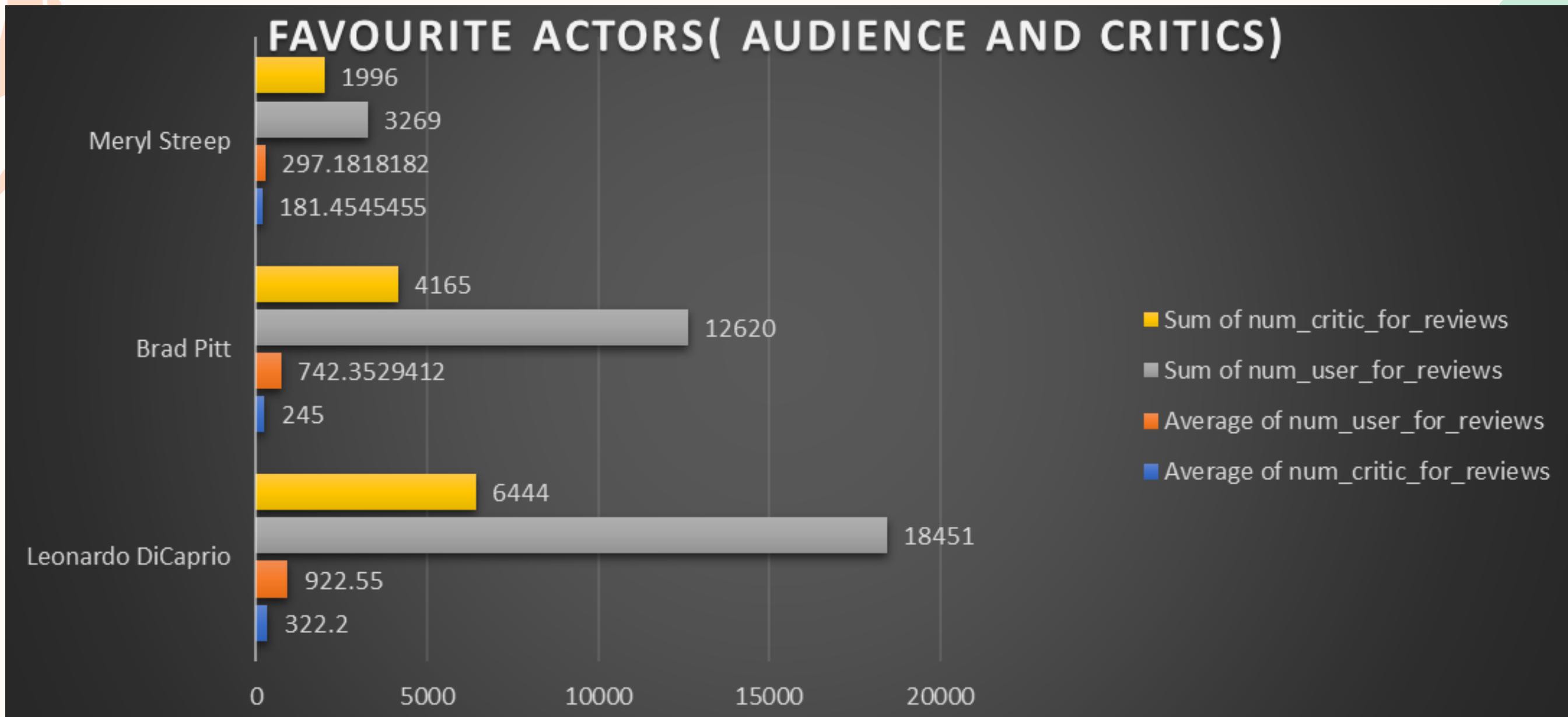
Favourite Actor – Critics



Actors Name	Average of num_critic_for_reviews	Sum of num_critic_for_reviews
Leonardo DiCaprio	322.2	6444
Brad Pitt	245	4165
Meryl Streep	181.4545455	1996



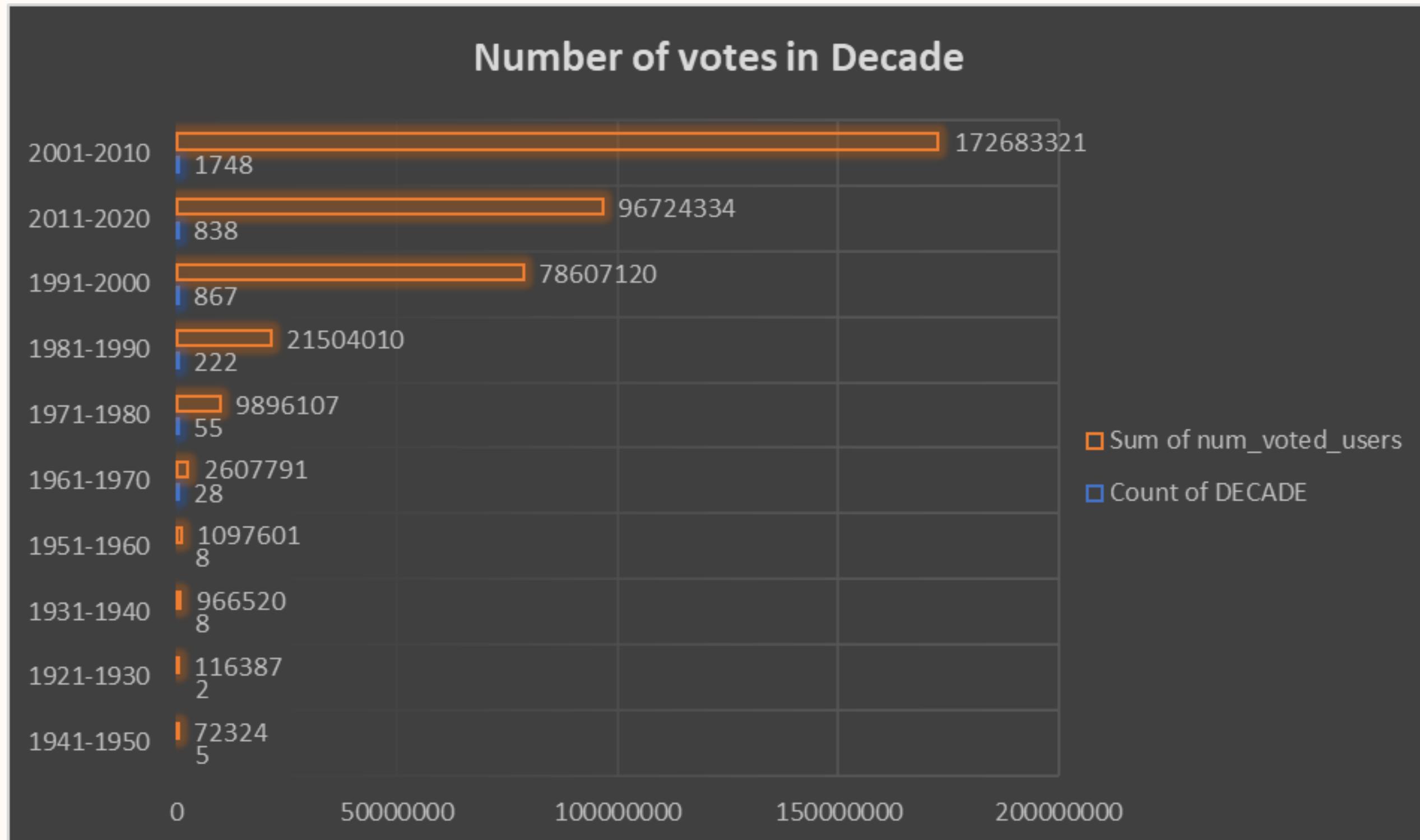
Favourite actor(Audience and critics)



actor name	Average of num_critic_for_reviews	Average of num_user_for_reviews	Sum of num_user_for_reviews	Sum of num_critic_for_reviews
Leonardo DiCaprio	322.2	922.55	18451	6444
Brad Pitt	245	742.3529412	12620	4165
Meryl Streep	181.4545455	297.1818182	3269	1996

Change in number of voted users over decades

Decade	Count of DECADE	Sum of num_voted_users
1941-1950	5	72324
1921-1930	2	116387
1931-1940	8	966520
1951-1960	8	1097601
1961-1970	28	2607791
1971-1980	55	9896107
1981-1990	222	21504010
1991-2000	867	78607120
2011-2020	838	96724334
2001-2010	1748	172683321





Results

- Data analysis plays a crucial role in understanding customer behaviour, preferences, and needs. By analyzing customer data, organizations can gain insights into their target audience and tailor their products, services, and marketing strategies to meet customer expectations more effectively.
- Enables organizations to personalize their offerings and enhance customer satisfaction.
- Organizations can understand individual preferences, anticipate needs, and deliver personalized experiences. This can lead to increased customer loyalty, higher satisfaction levels, and improved customer retention rates.



thank
you