

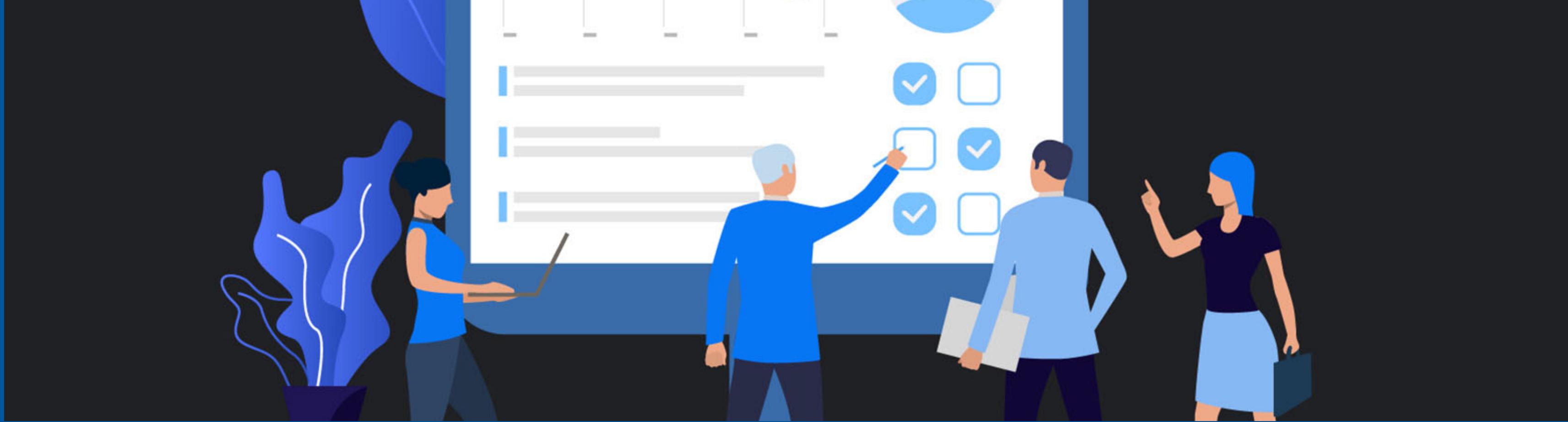
IMDB MOVIE ANALYSIS

Presented by: Shivam Choudhary

OVERVIEW

- PROJECT DESCRIPTION
- APPROACH
- TECH-STACK USED
- DATA CLEANING
- INSIGHTS
- RESULT





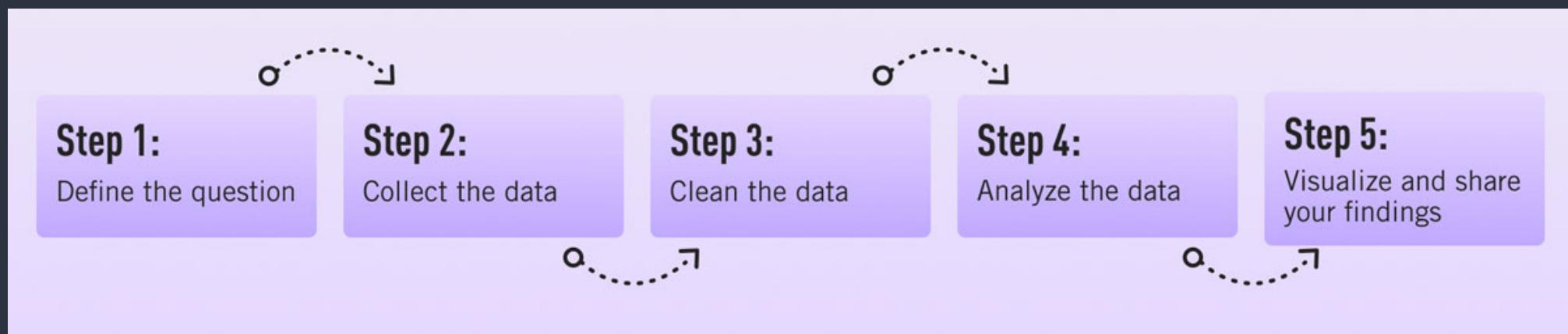
PROJECT DESCRIPTION

- This project aims to make you think like an analyst, how to approach a business problem analytically, what is the business logic behind it, etc. Imdb Data Consist of several columns and rows. Since it is a big data :

For the Project we First need to Frame the problem. For this task, we will need to define a problem we want to shed some light on.

- We can do this by asking 'What?' This is where we frame the problem i.e. What is the problem?
- By using these questions to guide our thinking:
 - What do you see happening?
 - What is your hypothesis for the cause of the problem? (this will be broadly based on intuition initially)
 - What is the impact of the problem on stakeholders?
 - What is the impact of the problem not being solved?

APPROACH



- Cleaning the data:: This is one of the most important step to perform before moving forward with the analysis
- Analysing the data:This involves Finding trends, patterns,etc.
- Insights:Finding insights through visualization tools.



TECH-STACK USED



Microsoft excel

The distinct features of Excel include analyzing data and making complex calculations. Excel has been used because it is easy to store and recover data and provides neat and secure visibility of information

In Excel, charts are used to make a graphical representation of any set of data. A chart is a visual representation of the data, in which the data is represented by symbols such as bars in a Bar Chart or lines in a Line Chart

DATA CLEANING

CLEANING THE DATA

This is one of the most important step to perform before moving forward with the analysis.(Dropping columns, removing null values, etc.)



First we will remove unwanted Columns:

- **Removed column-**
color
director_facebook_likes,
actor_3_facebook_likes,
actor_2_facebook_likes,
actor_1_facebook_likes,
cast_total_facebook_likes,
facenumber_in_poster,
plot_keywords,
movie_imdb_link
,content_rating,
aspect_ratio,
movie_facebook_likes,
actor_2_name,
actor_3_name

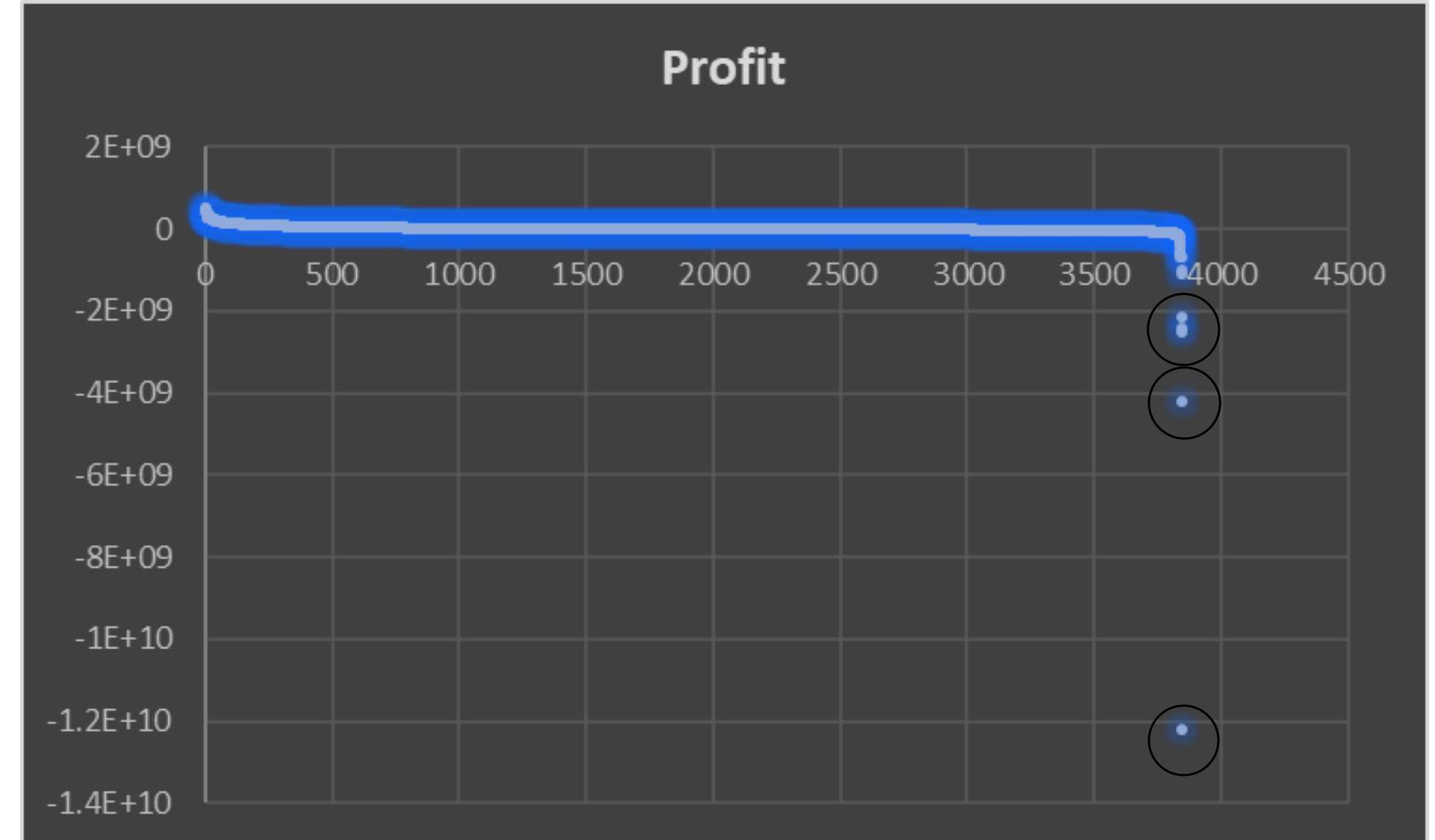
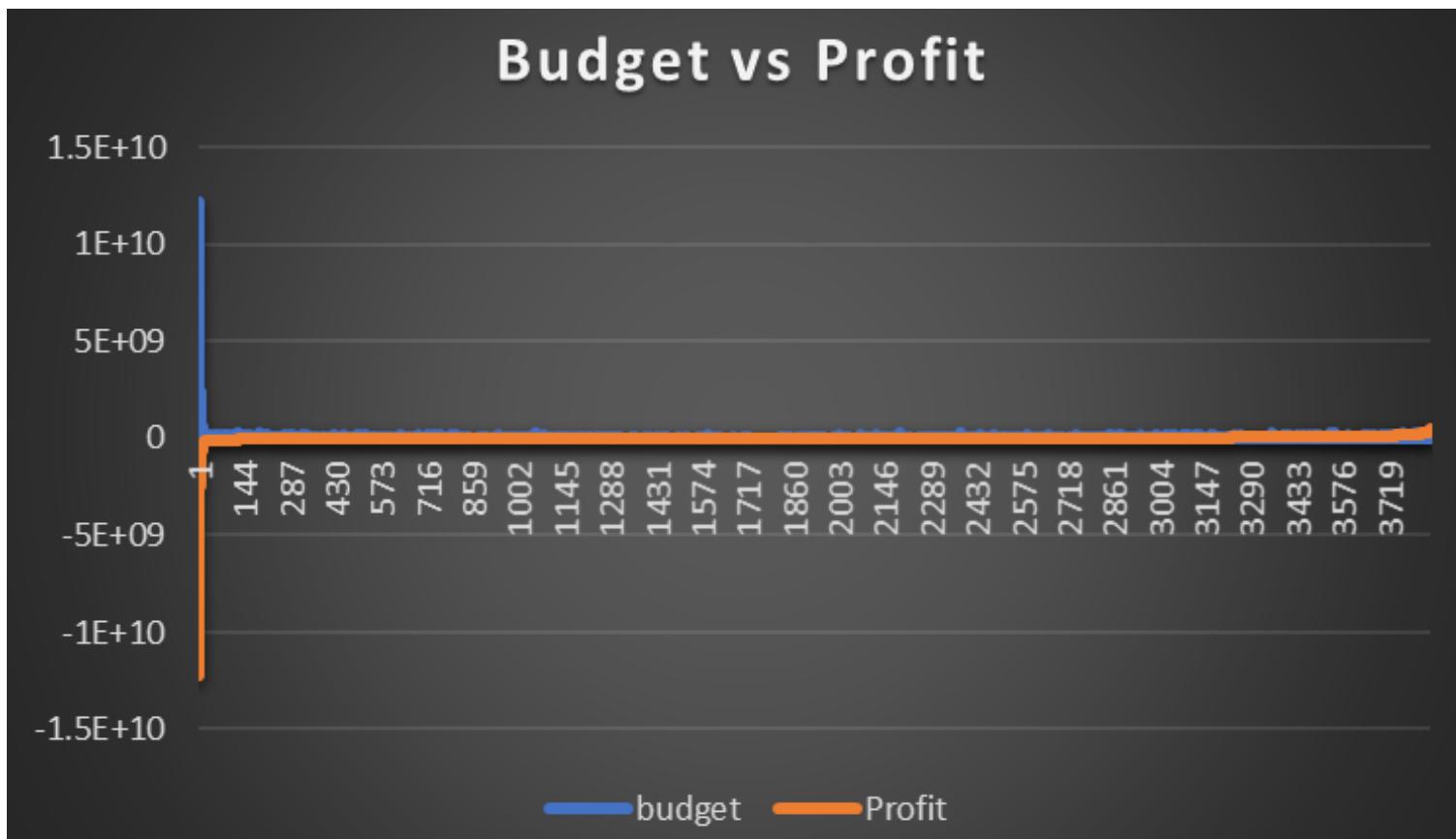
- 47 Duplicate Cells were Found AND Removed
- Á was replaced with Blank Value and Ã was replaced by A

INSIGHTS



MOVIES WITH HIGHEST PROFIT

a new column called profit which contains the difference of the two columns: gross and budget. was Created



Outliers Found from the Graph are:

-12213298588

-4199788333

-2397701809

-2499804112

-2127109510

MOVIES WITH HIGHEST PROFIT

a new column called profit which contains the difference of the two columns: gross and budget. was Created

TOP 25 HIGHEST PROFIT MOVIES:

director_name	actor_1_name	movie_title	language	country	gross	budget	Profit	title_year	imdb_score
James Cameron	CCH Pounder	Avatar	English	USA	7.61E+08	237000000	523505847	2009	7.9
Colin Trevorrow	Bryce Dallas Howard	Jurassic World	English	USA	6.52E+08	150000000	502177271	2015	7
James Cameron	Leonardo DiCaprio	Titanic	English	USA	6.59E+08	200000000	458672302	1997	7.7
George Lucas	Harrison Ford	Star Wars: Episode IV - A New Hope	English	USA	4.61E+08	11000000	449935665	1977	8.7
Steven Spielberg	Henry Thomas	E.T. the Extra-Terrestrial	English	USA	4.35E+08	10500000	424449459	1982	7.9
Joss Whedon	Chris Hemsworth	The Avengers	English	USA	6.23E+08	220000000	403279547	2012	8.1
Roger Allers	Matthew Broderick	The Lion King	English	USA	4.23E+08	45000000	377783777	1994	8.5
George Lucas	Natalie Portman	Star Wars: Episode I - The Phantom Menace	English	USA	4.75E+08	115000000	359544677	1999	6.5
Christopher Nolan	Christian Bale	The Dark Knight	English	USA	5.33E+08	185000000	348316061	2008	9
Gary Ross	Jennifer Lawrence	The Hunger Games	English	USA	4.08E+08	78000000	329999255	2012	7.3
Tim Miller	Ryan Reynolds	Deadpool	English	USA	3.63E+08	58000000	305024263	2016	8.1
Francis Lawrence	Jennifer Lawrence	The Hunger Games: Catching Fire	English	USA	4.25E+08	130000000	294645577	2013	7.6
Steven Spielberg	Wayne Knight	Jurassic Park	English	USA	3.57E+08	63000000	293784000	1993	8.1
Pierre Coffin	Steve Carell	Despicable Me 2	English	USA	3.68E+08	76000000	292049635	2013	7.5
Clint Eastwood	Bradley Cooper	American Sniper	English	USA	3.5E+08	58800000	291323553	2014	7.3
Andrew Stanton	Alexander Gould	Finding Nemo	English	USA	3.81E+08	94000000	286838870	2003	8.2
Andrew Adamson	Rupert Everett	Shrek 2	English	USA	4.36E+08	150000000	286471036	2004	7.2
Peter Jackson	Orlando Bloom	The Lord of the Rings: The Return of the King	English	USA	3.77E+08	94000000	283019252	2003	8.9
Richard Marquand	Harrison Ford	Star Wars: Episode VI - Return of the Jedi	English	USA	3.09E+08	32500000	276625409	1983	8.4
Robert Zemeckis	Tom Hanks	Forrest Gump	English	USA	3.3E+08	55000000	274691196	1994	8.8
Irvin Kershner	Harrison Ford	Star Wars: Episode V - The Empire Strikes Back	English	USA	2.9E+08	18000000	272158751	1980	8.8
Chris Columbus	Macaulay Culkin	Home Alone	English	USA	2.86E+08	18000000	267761243	1990	7.5
George Lucas	Natalie Portman	Star Wars: Episode III - Revenge of the Sith	English	USA	3.8E+08	113000000	267262555	2005	7.6
Sam Raimi	J.K. Simmons	Spider-Man	English	USA	4.04E+08	139000000	264706375	2002	7.3

TOP 250 IMDB:

a new column IMDb_Top_250 was constructed that stored the top 250 movies with the highest IMDb Rating (corresponding to the column: imdb_score). having the num_voted_users greater than 25,000

TOP 10 IMDB MOVIES:

director_name	genres	movie_title	num_voted_users	language	title_year	TOP 250 Imdb	imdb_score
Frank Darabont	Crime Drama	The Shawshank Redemption	1689764	English	1994	1	9.3
Francis Ford Coppola	Crime Drama	The Godfather	1155770	English	1972	2	9.2
Francis Ford Coppola	Crime Drama	The Godfather: Part II	790926	English	1974	3	9
Christopher Nolan	Action Crime Drama Thriller	The Dark Knight	1676169	English	2008	4	9
Sergio Leone	Western	The Good, the Bad and the Ugly	503509	Italian	1966	5	8.9
Steven Spielberg	Biography Drama History	Schindler's List	865020	English	1993	6	8.9
Quentin Tarantino	Crime Drama	Pulp Fiction	1324680	English	1994	7	8.9
Peter Jackson	Action Adventure Drama Fantasy	The Lord of the Rings: The Return of the King	1215718	English	2003	8	8.9
Irvin Kershner	Action Adventure Fantasy Sci-Fi	Star Wars: Episode V - The Empire Strikes Back	837759	English	1980	9	8.8
Robert Zemeckis	Comedy Drama	Forrest Gump	1251222	English	1994	10	8.8

TOP 250 IMDB:

a new column IMDb_Top_250 was constructed that stored the top 250 movies with the highest IMDb Rating (corresponding to the column: imdb_score). having the num_voted_users greater than 25,000

TOP 10 FOREIGN LANGUAGE IMDB MOVIES:

director_name	genres	movie_title	num_voted_users	language	title_year	TOP 250 Imdb	imdb_score
Sergio Leone	Western	The Good, the Bad and the Ugly	503509	Italian	1966	5	8.9
Akira Kurosawa	Action Adventure Drama	Seven Samurai	229012	Japanese	1954	14	8.7
Fernando Meirelles	Crime Drama	City of God	533200	Portuguese	2002	20	8.7
Hayao Miyazaki	Adventure Animation Family Fantasy	Spirited Away	417971	Japanese	2001	27	8.6
Majid Majidi	Drama Family	Children of Heaven	27882	Persian	1997	36	8.5
Florian Henckel von Donnersmarc	Drama Thriller	The Lives of Others	259379	German	2006	43	8.5
Wolfgang Petersen	Adventure Drama Thriller War	Das Boot	168203	German	1981	48	8.4
Hayao Miyazaki	Adventure Animation Fantasy	Princess Mononoke	221552	Japanese	1997	54	8.4
Jean-Pierre Jeunet	Comedy Romance	Amélie	534262	French	2001	57	8.4
Chang-wook Park	Drama Mystery Thriller	Oldboy	356181	Korean	2003	58	8.4

BEST DIRECTORS:

Grouping the column using the director_name column.

Finding out the top 10 directors for whom the mean of imdb_score is the highest

TOP 10 DIRECTORS::

Top 10 Directors having highest IMDB Mean:

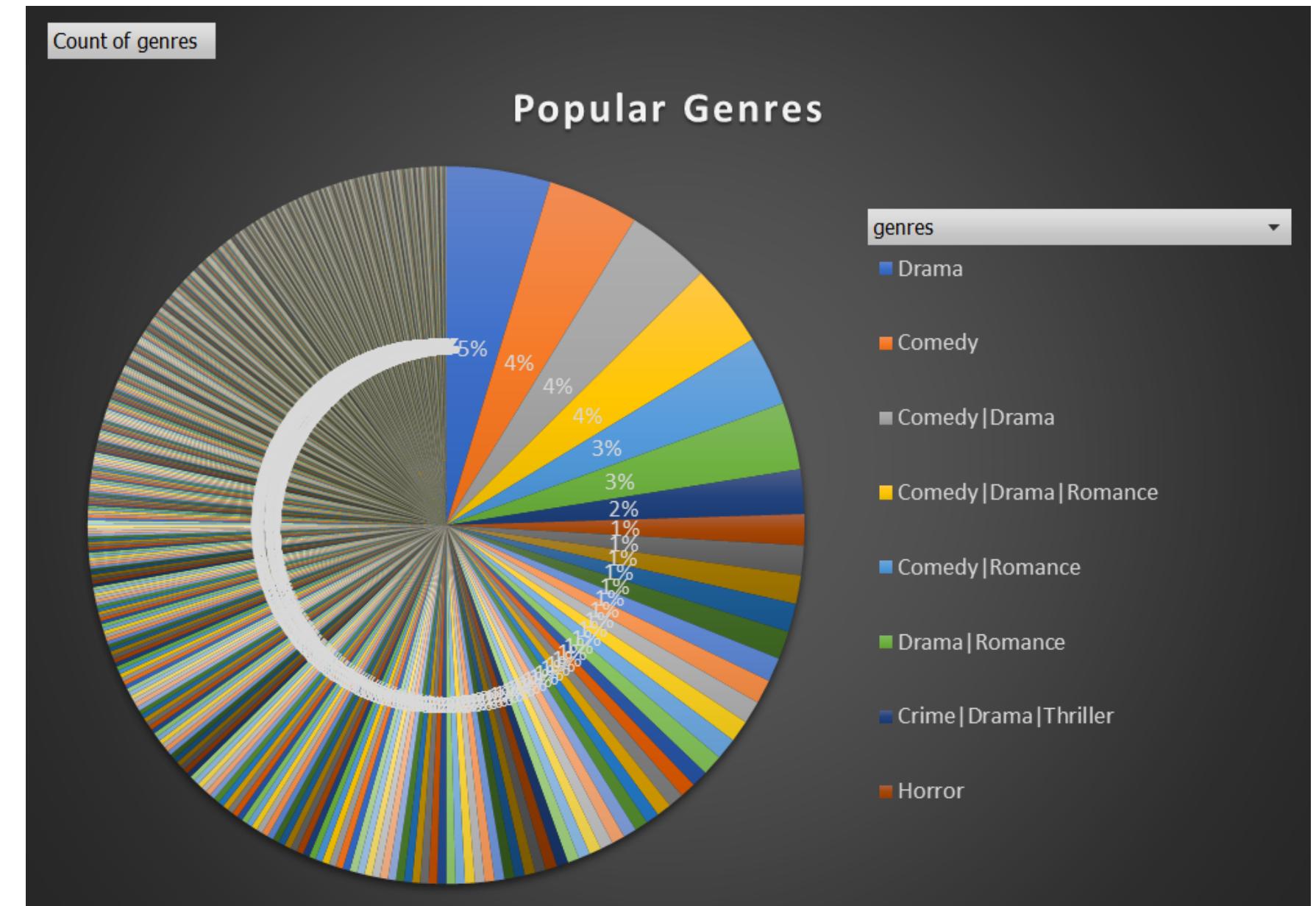
Director Name	Mean Of Imdb
1 John Blanchard	9.5
2 Sadyk Sher-Niyaz	8.7
3 Mitchell Altieri	8.7
4 Charles Chaplin	8.6
5 Mike Mayhall	8.6
6 Damien Chazelle	8.5
7 Majid Majidi	8.5
8 Raja Menon	8.5
9 Ron Fricke	8.5
10 Sergio Leone	8.475

POPULAR GENRES:

Performing analysis on genres column
to identify most popular genre using pivot table

TOP 10 Genres:

Rank	Genres	Count Of Genre
1	Drama	235
2	Comedy	205
3	Comedy Drama Romance	187
4	Comedy Romance	157
5	Drama Romance	151
6	Crime Drama Thriller	100
7	Horror	70
8	Action Crime Drama Thriller	67
9	Action Crime Thriller	64
10	Drama Thriller	64



CHARTS:

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 were created.

Tables of Actor:

- Leonardo Dicaprio
- Meryl Streep
- Brad Pitt

director_name	actor_1_name	movie_title	num_voted_us	num_user_for_review	num_critic_for_review	language	country	gross	budget	title_y	imdb_sc
James Cameron	Leonardo DiCaprio	Titanic	793059	2528	315	English	USA	6.59E+08	20000000	1997	7.7
Baz Luhrmann	Leonardo DiCaprio	The Great Gatsby	362912	753	490	English	Australia	1.45E+08	105000000	2013	7.3
Christopher Nolan	Leonardo DiCaprio	Inception	1468200	2803	642	English	USA	2.93E+08	160000000	2010	8.8
Alejandro G. Iñárritu	Leonardo DiCaprio	The Revenant	406020	1188	556	English	USA	1.84E+08	135000000	2015	8.1
Martin Scorsese	Leonardo DiCaprio	The Aviator	264318	799	267	English	USA	1.03E+08	110000000	2004	7.5
Quentin Tarantino	Leonardo DiCaprio	Django Unchained	955174	1193	765	English	USA	1.63E+08	100000000	2012	8.5
Edward Zwick	Leonardo DiCaprio	Blood Diamond	400292	657	166	English	Germany	57366262	100000000	2006	8
Martin Scorsese	Leonardo DiCaprio	The Wolf of Wall Street	780588	1138	606	English	USA	1.17E+08	100000000	2013	8.2
Martin Scorsese	Leonardo DiCaprio	Gangs of New York	314033	1166	233	English	USA	77679638	100000000	2002	7.5
Martin Scorsese	Leonardo DiCaprio	The Departed	873649	2054	352	English	USA	1.32E+08	90000000	2006	8.5
Martin Scorsese	Leonardo DiCaprio	Shutter Island	786092	964	490	English	USA	1.28E+08	80000000	2010	8.1
Ridley Scott	Leonardo DiCaprio	Body of Lies	174248	263	238	English	USA	39380442	70000000	2008	7.1
Steven Spielberg	Leonardo DiCaprio	Catch Me If You Can	525801	667	194	English	USA	1.64E+08	52000000	2002	8
Danny Boyle	Leonardo DiCaprio	The Beach	176169	548	118	English	USA	39778599	50000000	2000	6.6
Sam Mendes	Leonardo DiCaprio	Revolutionary Road	152591	414	323	English	USA	22877808	35000000	2008	7.3
Randall Wallace	Leonardo DiCaprio	The Man in the Iron Mask	125219	244	83	English	USA	56876365	35000000	1998	6.4
Clint Eastwood	Leonardo DiCaprio	J. Edgar	102728	279	392	English	USA	37304950	35000000	2011	6.6
Sam Raimi	Leonardo DiCaprio	The Quick and the Dead	69197	216	63	English	Japan	18636537	32000000	1995	6.4
Jerry Zaks	Leonardo DiCaprio	Marvin's Room	20163	71	45	English	USA	12782508	23000000	1996	6.7
Baz Luhrmann	Leonardo DiCaprio	Romeo + Juliet	167750	506	106	English	USA	46338728	14500000	1996	6.8
Baz Luhrmann	Leonardo DiCaprio	The Great Gatsby	362933	753	490	English	Australia	1.45E+08	105000000	2013	7.3

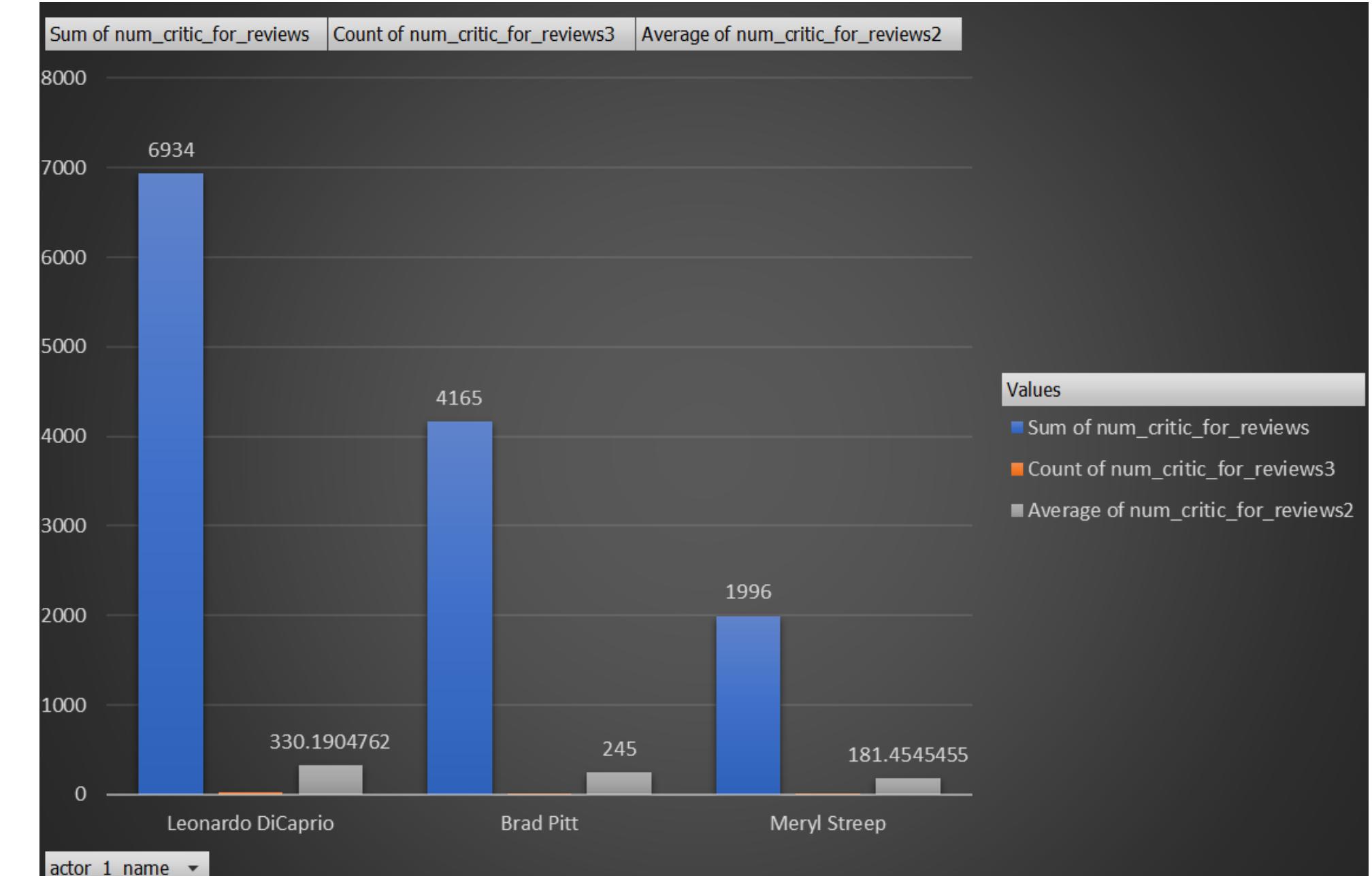
director_name	actor_1_name	movie_title	num_voted_us	num_user_for_review	num_critic_for_review	language	country	gross	budget	title_y	imdb_sc
Nancy Meyers	Meryl Streep	It's Complicated	69860	214	187	English	USA	1.13E+08	85000000	2009	6.6
Curtis Hanson	Meryl Streep	The River Wild	32544	69	42	English	USA	46815748	45000000	1994	6.3
Nora Ephron	Meryl Streep	Julie & Julia	79264	277	252	English	USA	94125426	40000000	2009	7
David Frankel	Meryl Streep	The Devil Wears Prada	286178	631	208	English	USA	1.25E+08	35000000	2006	6.8
Robert Redford	Meryl Streep	Lions for Lambs	41170	298	227	English	USA	14998070	35000000	2007	6.2
Sydney Pollack	Meryl Streep	Out of Africa	52339	200	66	English	USA	87100000	31000000	1985	7.2
David Frankel	Meryl Streep	Hope Springs	34258	178	234	English	USA	63536011	30000000	2012	6.3
Carl Franklin	Meryl Streep	One True Thing	9283	112	64	English	USA	23209440	30000000	1998	7
Stephen Daldry	Meryl Streep	The Hours	102123	660	174	English	USA	41597830	25000000	2002	7.6
Phyllida Lloyd	Meryl Streep	The Iron Lady	82327	350	331	English	UK	29959436	13000000	2011	6.4
Robert Altman	Meryl Streep	A Prairie Home Companion	19655	280	211	English	USA	20338609	10000000	2006	6.8

director_name	actor_1_name	movie_title	num_voted_us	num_user_for_review	num_critic_for_review	language	country	gross	budget	title_y	imdb_sc
David Fincher	Brad Pitt	The Curious Case of Benjamin Button	459346	822	362	English	USA	1.27E+08	15000000	2008	7.8
Wolfgang Petersen	Brad Pitt	Troy	381672	1694	220	English	USA	1.33E+08	17500000	2004	7.2
Steven Soderbergh	Brad Pitt	Ocean's Twelve	284852	627	198	English	USA	1.26E+08	11000000	2004	6.4
Doug Liman	Brad Pitt	Mr. & Mrs. Smith	348861	798	233	English	USA	1.86E+08	12000000	2005	6.5
Tony Scott	Brad Pitt	Spy Game	121259	361	142	English	Germany	26871	9200000	2001	7
Steven Soderbergh	Brad Pitt	Ocean's Eleven	402645	845	186	English	USA	1.83E+08	85000000	2001	7.8
David Ayer	Brad Pitt	Fury	303185	701	406	English	USA	85707116	68000000	2014	7.6
Jean-Jacques Annaud	Brad Pitt	Seven Years in Tibet	96385	119	76	English	USA	37901509	70000000	1997	7
David Fincher	Brad Pitt	Fight Club	1347461	2968	315	English	USA	37023395	63000000	1999	8.8
Patrick Gilmore	Brad Pitt	Sinbad: Legend of the Seven Seas	36144	91	98	English	USA	26288320	60000000	2003	6.7
Neil Jordan	Brad Pitt	Interview with the Vampire: The Vampire Ch	239752	406	120	English	USA	1.05E+08	60000000	1994	7.6
Terrence Malick	Brad Pitt	The Tree of Life	136367	975	584	English	USA	13303319	32000000	2011	6.7
Andrew Dominik	Brad Pitt	The Assassination of Jesse James by the Co	136104	415	273	English	USA	3904982	30000000	2007	7.5
Alejandro G. Iñárritu	Brad Pitt	Babel	243799	908	285	English	France	34300771	25000000	2006	7.5
Angelina Jolie Pitt	Brad Pitt	By the Sea	7976	61	131	English	USA	531009	10000000	2015	5.3
Andrew Dominik	Brad Pitt	Killing Them Softly	111625	369	414	English	USA	14938570	15000000	2012	6.2
Tony Scott	Brad Pitt	True Romance	163492	460	122	English	USA	12281500	13000000	1993	8

CHARTS:

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 were created.

num critic for reviews:



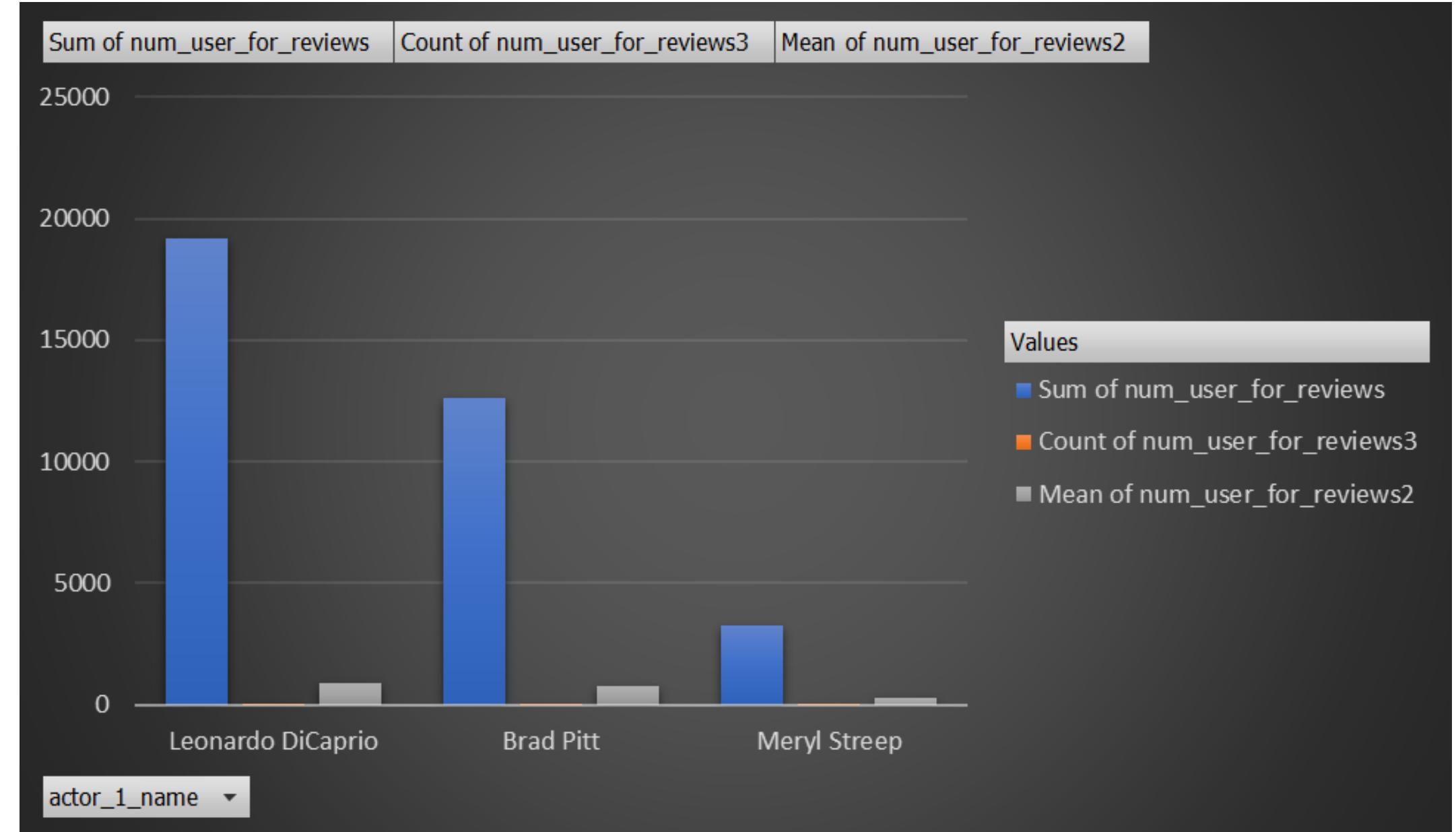
Actor Name	Sum of num_critic_for_reviews	Count of num_critic_for_reviews3	Average of num_critic_for_reviews2
Leonardo DiCaprio	6934	330.1904762	21
Brad Pitt	4165	245	17
Meryl Streep	1996	181.4545455	11

CHARTS:

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 were created.

From the above two graphs we can infer that 'Leonardo DiCaprio' is both critic favourite actor and audience favourite actor

num_users_for_review



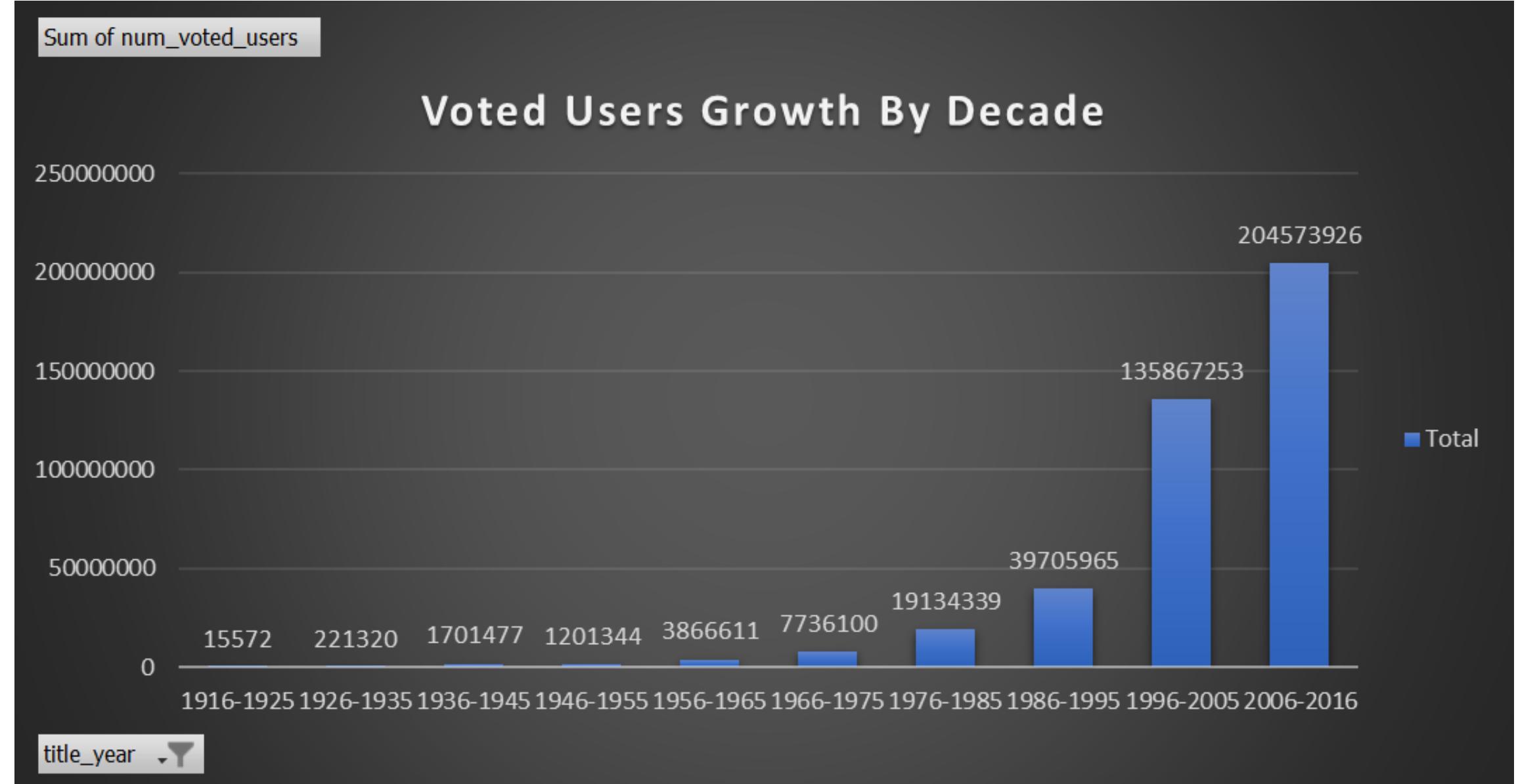
Row Labels	Sum of num_user_for_reviews	Count of num_user_for_reviews3	Mean of num_user_for_reviews2
Leonardo DiCaprio	19204	21	914.4761905
Brad Pitt	12620	17	742.3529412
Meryl Streep	3269	11	297.1818182

VOTED USERS PER DECADE:

there is change in number of voted users over decades . A column called decade which represents the decade to which every movie belongs to is created with class intervals of 10 years.

Year Vs num Voted users

Year	Sum of num_voted_users
1916-1925	15572
1926-1935	221320
1936-1945	1701477
1946-1955	1201344
1956-1965	3866611
1966-1975	7736100
1976-1985	19134339
1986-1995	39705965
1996-2005	135867253
2006-2016	204573926



RESULT:

Complete Analysis and visualization was carried out of the dataset IMDB with the use of excel



Top Highest Grossing movies was filtered out with excel formulaes and sorting process



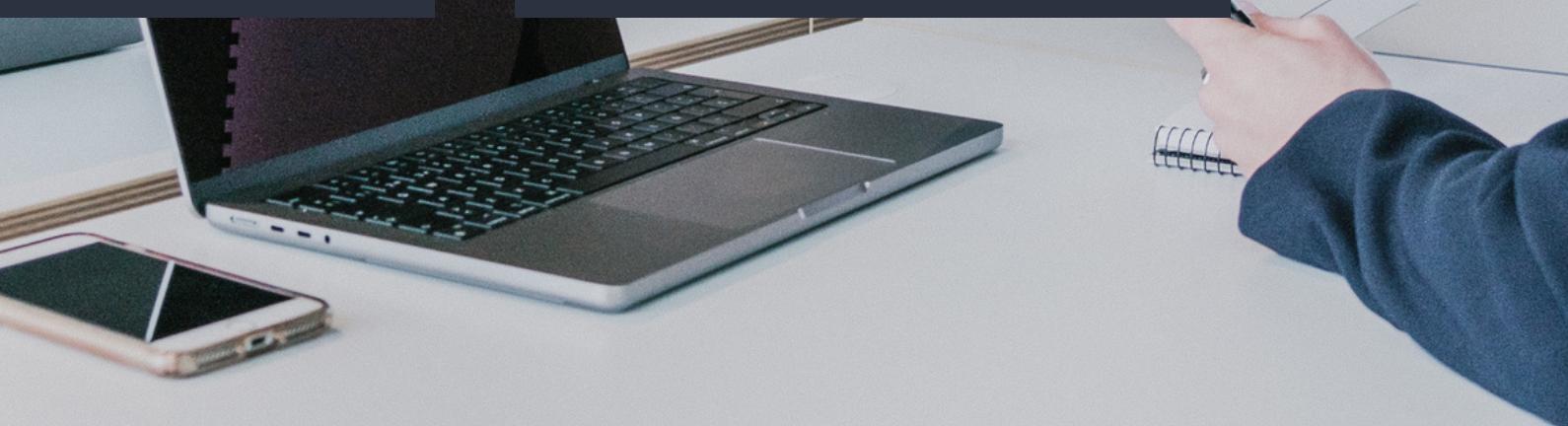
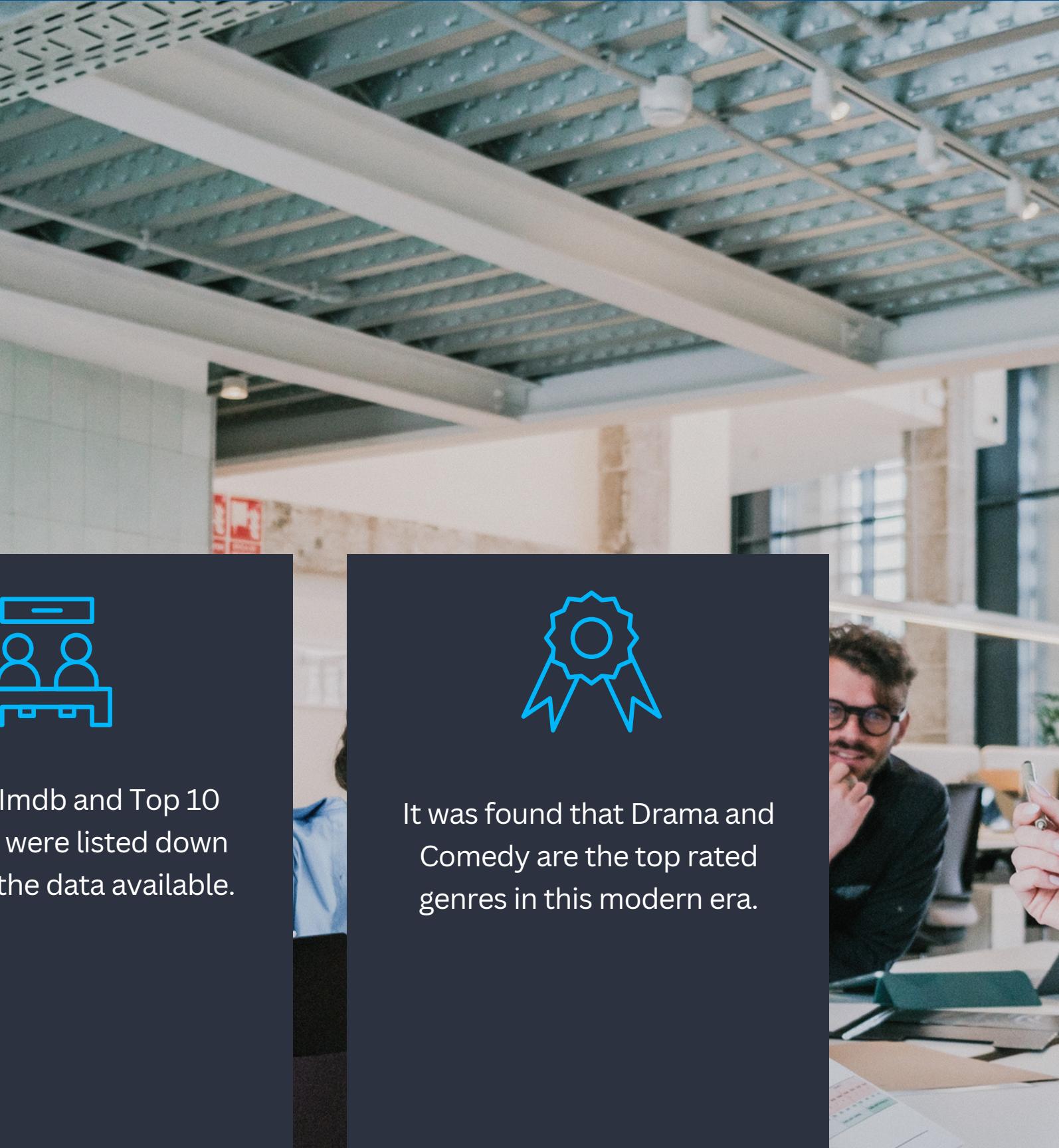
Most critic and most favourite actor was figured out from the dataset. Surprisingly it was found that single actor belonged to both category.



Top 250 Imdb and Top 10 Directors were listed down based on the data available.



It was found that Drama and Comedy are the top rated genres in this modern era.



Thank's For Watching

