



IMDb

The background of the slide features a large grid of movie posters. A large yellow rectangle with the IMDb logo is positioned on the left side. The right side of the slide is white, containing the title and a stylized graphic of a movie camera in the bottom right corner.

IMDb Movie Analysis

1985 - 1995 (0 < rating ≤ 10)

Project Description:

- Dataset having 28 columns and 5044 rows of different IMDB Movies.
- Framing a problem and finding solutions for the problem defined.
- Clean the data as necessary remove null and duplicate values.
- Explore the cleaned dataset and derive insights.
- Use root cause analysis Five 'Whys' approach.
- Create chart pivot table and filter the column to get required data and graphs to answer given questions.
- Give answer to asked questions for taking data-driven decisions.

Approach:

1. Download and Data open in excel
2. Find null and duplicate values and clean all data
3. Understand and process the data to solve as per asked problems.
4. Use pivot table and other functions to give answer of asked questions
5. Create charts for easy and meaningful insights and data interpretation.
6. Create report in ppt format.

Tech-Stack Used:

1. MS Excel 2019: To analyze the dataset & create charts.
2. Google Slides: To prepare the presentation slides.

Results & Insights:

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.

Task: Clean the Data

Result: After Cleaning the data, removing duplicates and nulls using Power Query Editor, the number of rows has been reduced to **3722** rows from **5044** rows.

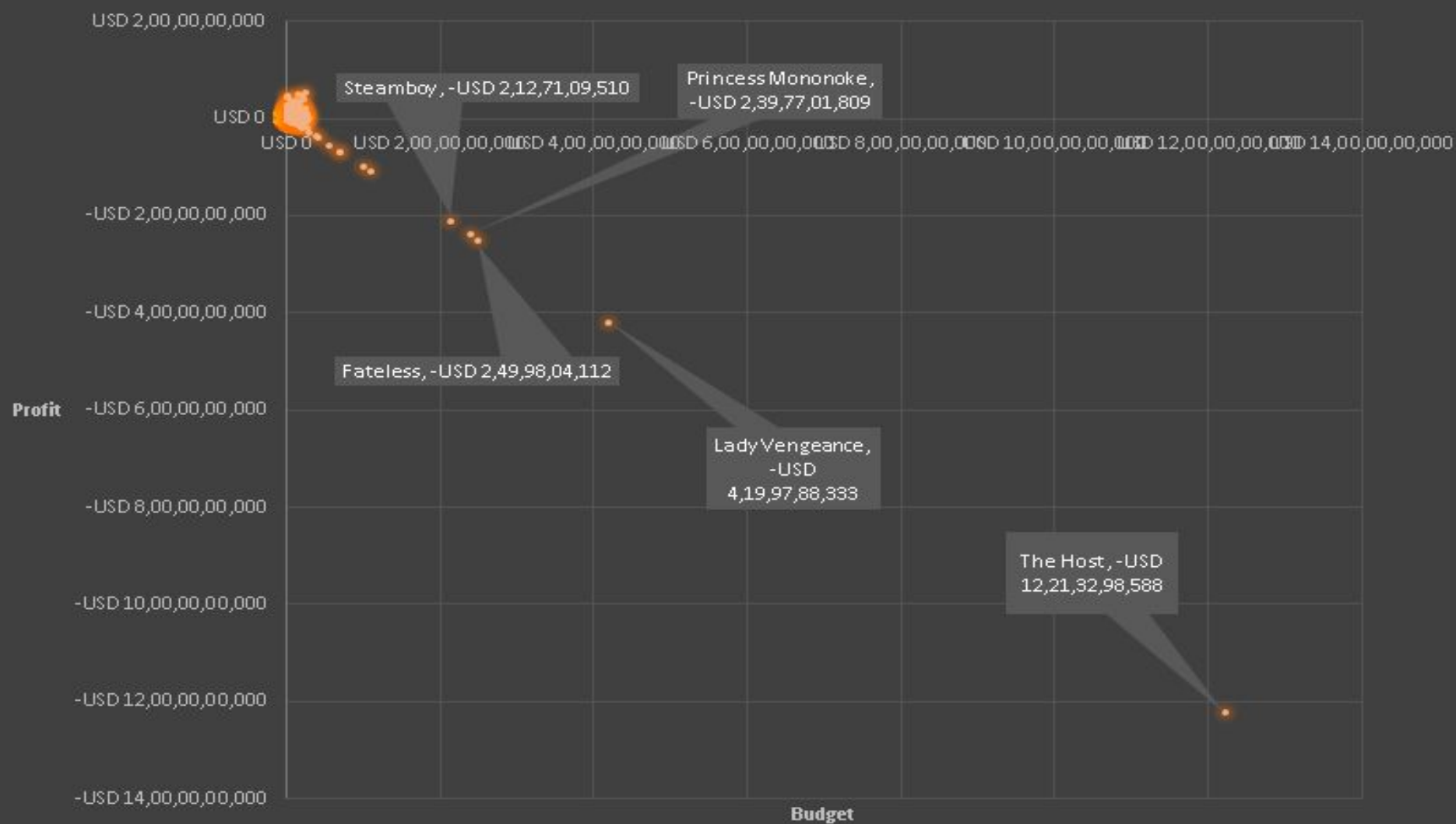
Movies with highest profit:

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.

Task: Find the movies with the highest profit?

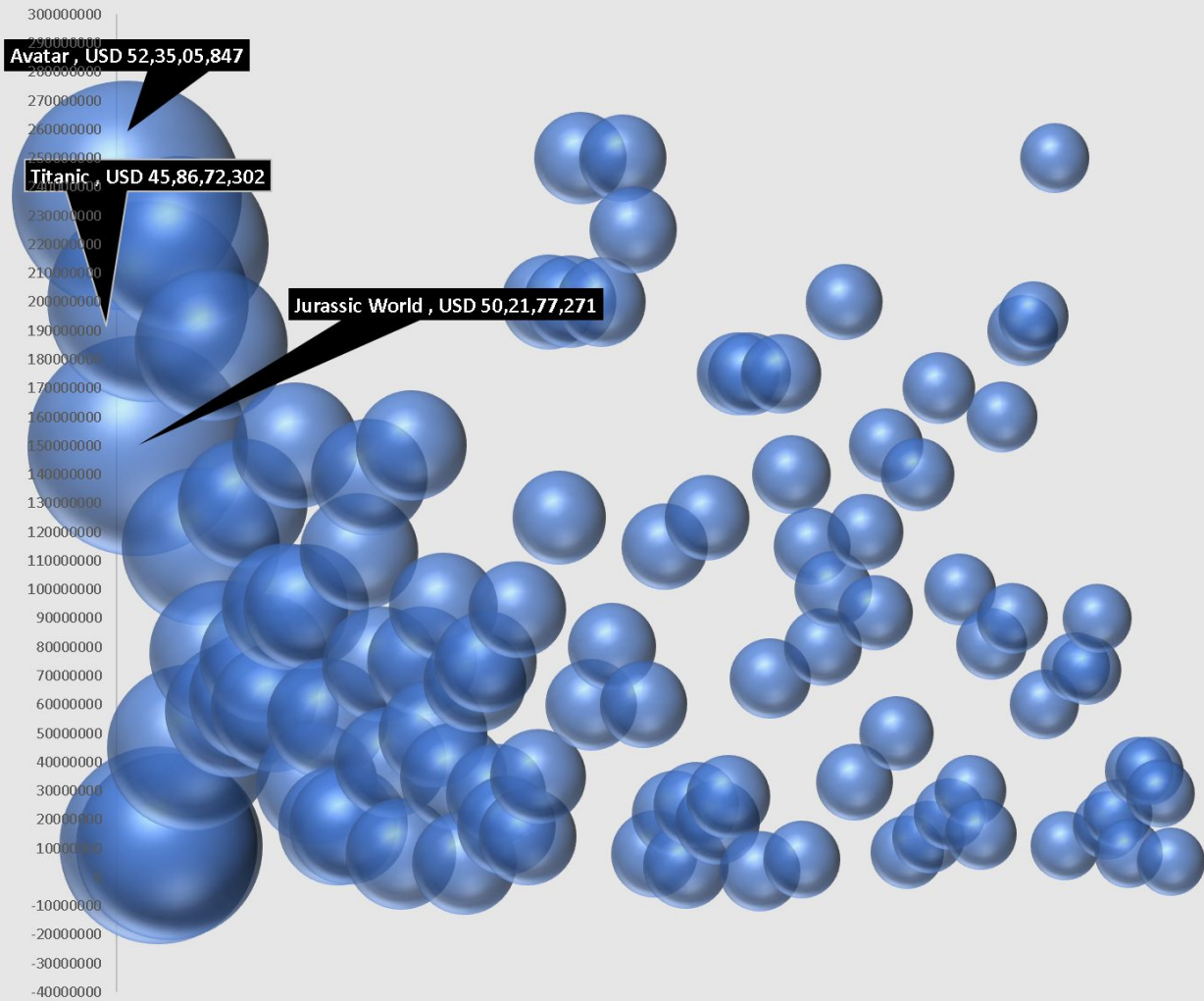
Results:

1. In the following first chart, the labelled ones are the possible outliers in the dataset deviating from other data-points.
2. In the second chart, the movie title is labelled along with profit with bubble size interpreting total profit earned by top 100 movies.



Highest profitable Movies

Budget



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`.

Task: Find IMDB Top 250.

IMDb_rank	movie_title	imdb_score			
1	The Shawshank Redemption	9.3	31	The Lives of Others	8.5
2	The Godfather	9.2	32	Whiplash	8.5
3	The Godfather: Part II	9	33	Memento	8.5
4	The Dark Knight	9	34	Alien	8.5
5	The Good, the Bad and the Ugly	8.9	35	Raiders of the Lost Ark	8.5
6	Pulp Fiction	8.9	36	Back to the Future	8.5
7	Schindler's List	8.9	37	Apocalypse Now	8.5
8	The Lord of the Rings: The Return of the King	8.9	38	The Pianist	8.5
9	Star Wars: Episode V - The Empire Strikes Back	8.8	39	The Prestige	8.5
10	Forrest Gump	8.8	40	The Lion King	8.5
11	Fight Club	8.8	41	The Green Mile	8.5
12	The Lord of the Rings: The Fellowship of the Ring	8.8	42	The Departed	8.5
13	Inception	8.8	43	Django Unchained	8.5
14	Seven Samurai	8.7	44	Terminator 2: Judgment Day	8.5
15	City of God	8.7	45	Gladiator	8.5
16	One Flew Over the Cuckoo's Nest	8.7	46	The Dark Knight Rises	8.5
17	Star Wars: Episode IV - A New Hope	8.7	47	A Separation	8.4
18	Goodfellas	8.7	48	Reservoir Dogs	8.4
19	The Matrix	8.7	49	Oldboy	8.4
20	The Lord of the Rings: The Two Towers	8.7	50	Requiem for a Dream	8.4
21	Modern Times	8.6	51	Das Boot	8.4
22	The Usual Suspects	8.6	52	American Beauty	8.4
23	American History X	8.6	53	Lawrence of Arabia	8.4
24	The Silence of the Lambs	8.6	54	Aliens	8.4
25	Spirited Away	8.6	55	Once Upon a Time in America	8.4
26	Se7en	8.6	56	Star Wars: Episode VI - Return of the Jedi	8.4
27	Saving Private Ryan	8.6	57	Braveheart	8.4
28	Interstellar	8.6	58	Amélie	8.4
29	Children of Heaven	8.5	59	WALL·E	8.4
30	Psycho	8.5	60	Princess Mononoke	8.4
			61	Monty Python and the Holy Grail	8.3
			62	Some Like It Hot	8.3
			63	The Hunt	8.3
			64	The Sting	8.3
			65	Metropolis	8.3
			66	Snatch	8.3
			67	Good Will Hunting	8.3
			68	2001: A Space Odyssey	8.3
			69	Room	8.3
			70	Downfall	8.3
			71	Unforgiven	8.3
			72	Amadeus	8.3
			73	Raging Bull	8.3
			74	Eternal Sunshine of the Spotless Mind	8.3
			75	Scarface	8.3
			76	Toy Story	8.3
			77	L.A. Confidential	8.3
			78	Indiana Jones and the Last Crusade	8.3
			79	Inglourious Basterds	8.3
			80	Batman Begins	8.3
			81	Up	8.3
			82	Inside Out	8.3
			83	Toy Story 3	8.3
			84	On the Waterfront	8.2
			85	Lock, Stock and Two Smoking Barrels	8.2
			86	The Secret in Their Eyes	8.2
			87	The Bridge on the River Kwai	8.2
			88	Trainspotting	8.2
			89	Gone with the Wind	8.2
			90	Incendies	8.2

91	Pan's Labyrinth	8.2	121	Tae Guk Gi: The Brotherhood of War	8.1	151	Akira	8.1
92	The Thing	8.2	122	The Imitation Game	8.1	152	A Fistful of Dollars	8
93	Into the Wild	8.2	123	Groundhog Day	8.1	153	Sling Blade	8
94	The Big Lebowski	8.2	124	The Princess Bride	8.1	154	Waltz with Bashir	8
95	Howl's Moving Castle	8.2	125	Hotel Rwanda	8.1	155	Before Sunset	8
96	Warrior	8.2	126	12 Years a Slave	8.1	156	Young Frankenstein	8
97	Die Hard	8.2	127	Spotlight	8.1	157	Central Station	8
98	Blade Runner	8.2	128	The Grand Budapest Hotel	8.1	158	Shaun of the Dead	8
99	Gran Torino	8.2	129	The Help	8.1	159	Boyhood	8
100	Casino	8.2	130	No Country for Old Men	8.1	160	Bowling for Columbine	8
101	V for Vendetta	8.2	131	There Will Be Blood	8.1	161	Dallas Buyers Club	8
102	A Beautiful Mind	8.2	132	Kill Bill: Vol. 1	8.1	162	Persepolis	8
103	Finding Nemo	8.2	133	Million Dollar Baby	8.1	163	The Exorcist	8
104	The Wolf of Wall Street	8.2	134	Rush	8.1	164	Jaws	8
105	How to Train Your Dragon	8.2	135	The Sixth Sense	8.1	165	The Sound of Music	8
106	Captain America: Civil War	8.2	136	Sin City	8.1	166	Fiddler on the Roof	8
107	Rocky	8.1	137	Prisoners	8.1	167	Sicko	8
108	The Celebration	8.1	138	Deadpool	8.1	168	The Straight Story	8
109	Amores Perros	8.1	139	The Truman Show	8.1	169	Doctor Zhivago	8
110	The Best Years of Our Lives	8.1	140	Gone Girl	8.1	170	My Name Is Khan	8
111	Before Sunrise	8.1	141	Jurassic Park	8.1	171	Dancer in the Dark	8
112	The Wizard of Oz	8.1	142	Shutter Island	8.1	172	Black Swan	8
113	Elite Squad	8.1	143	The Martian	8.1	173	The Perks of Being a Wallflower	8
114	Annie Hall	8.1	144	The Bourne Ultimatum	8.1	174	True Romance	8
115	Donnie Darko	8.1	145	Monsters, Inc.	8.1	175	The Artist	8
116	Platoon	8.1	146	The Revenant	8.1	176	The King's Speech	8
117	Butch Cassidy and the Sundance Kid	8.1	147	Pirates of the Caribbean: The Curse of the Black Pearl	8.1	177	Brazil	8
118	The Terminator	8.1	148	Mad Max: Fury Road	8.1	178	In Bruges	8
119	Stand by Me	8.1	149	Guardians of the Galaxy	8.1	179	Mulholland Drive	8
120	The Sea Inside	8.1	150	The Avengers	8.1	180	Slumdog Millionaire	8

181	Dead Poets Society	8			
182	Dances with Wolves	8			
183	Her	8			
184	Mystic River	8	211	The Chorus	7.9
185	Rain Man	8	212	The Wrestler	7.9
186	Aladdin	8	213	Do the Right Thing	7.9
187	Kill Bill: Vol. 2	8	214	Hot Fuzz	7.9
188	District 9	8	215	Little Miss Sunshine	7.9
189	Magnolia	8	216	Nightcrawler	7.9
190	JFK	8	217	Amour	7.9
191	Serenity	8	218	E.T. the Extra-Terrestrial	7.9
192	Catch Me If You Can	8	219	The Remains of the Day	7.9
193	The Pursuit of Happyness	8	220	Crouching Tiger, Hidden Dragon	7.9
194	Big Fish	8	221	Boogie Nights	7.9
195	The Iron Giant	8	222	My Fair Lady	7.9
196	Cinderella Man	8	223	Glory	7.9
197	The Incredibles	8	224	Glory	7.9
198	Blood Diamond	8	225	Ed Wood	7.9
199	Life of Pi	8	226	Letters from Iwo Jima	7.9
200	Ratatouille	8	227	Edward Scissorhands	7.9
201	Star Trek	8	228	Taken	7.9
202	Casino Royale	8	229	The Untouchables	7.9
203	Casino Royale	8	230	The World's Fastest Indian	7.9
204	X-Men: Days of Future Past	8			
205	Once	7.9			
206	Halloween	7.9			
207	4 Months, 3 Weeks and 2 Days	7.9			
208	Nine Queens	7.9			
209	Before Midnight	7.9			
210	Moon	7.9			
			231	The Fighter	7.9
			232	The Blues Brothers	7.9
			233	The Right Stuff	7.9
			234	Walk the Line	7.9
			235	Straight Outta Compton	7.9
			236	The Notebook	7.9
			237	The Hateful Eight	7.9
			238	Captain Phillips	7.9
			239	The Bourne Identity	7.9
			240	Almost Famous	7.9
			241	Shrek	7.9
			242	The Insider	7.9
			243	Children of Men	7.9
			244	Toy Story 2	7.9
			245	Iron Man	7.9
			246	How to Train Your Dragon 2	7.9
			247	Big Hero 6	7.9
			248	Edge of Tomorrow	7.9
			249	The Hobbit: The Desolation of Smaug	7.9
			250	Avatar	7.9

Top Foreign Language Film:

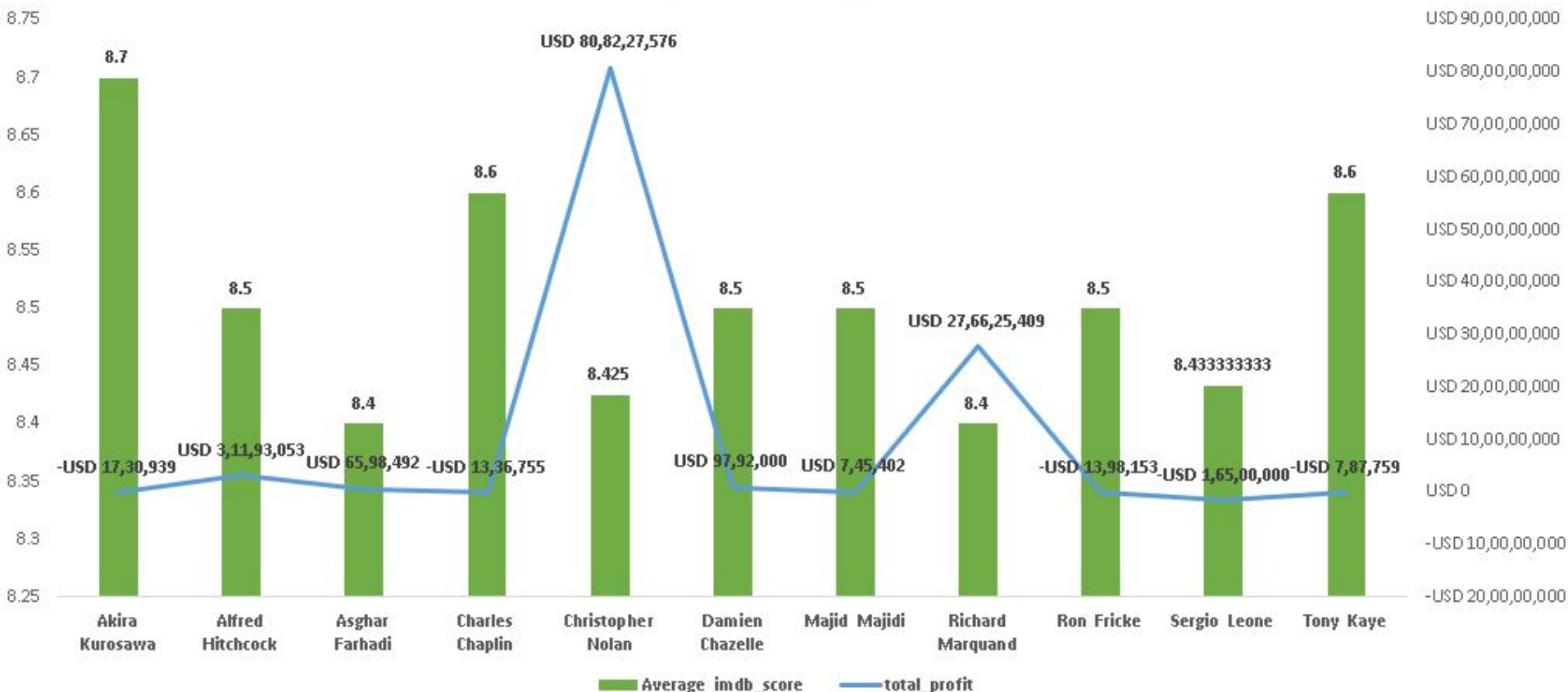
IMDb_rank_250	Top_Foreign_Lang_Film	language			
5	The Good, the Bad and the Ugly	Italian	95	Howl's Moving Castle	Japanese
14	Seven Samurai	Japanese	108	The Celebration	Danish
15	City of God	Portuguese	109	Amores Perros	Spanish
25	Spirited Away	Japanese	113	Elite Squad	Portuguese
29	Children of Heaven	Persian	120	The Sea Inside	Spanish
31	The Lives of Others	German	121	Tae Guk Gi: The Brotherhood of War	Korean
47	A Separation	Persian	151	Akira	Japanese
49	Oldboy	Korean	152	A Fistful of Dollars	Italian
51	Das Boot	German	154	Waltz with Bashir	Hebrew
58	Amélie	French	157	Central Station	Portuguese
60	Princess Mononoke	Japanese	162	Persepolis	French
63	The Hunt	Danish	170	My Name Is Khan	Hindi
65	Metropolis	German	207	4 Months, 3 Weeks and 2 Days	Romanian
70	Downfall	German	208	Nine Queens	Spanish
86	The Secret in Their Eyes	Spanish	211	The Chorus	French
90	Incendies	French	217	Amour	French
91	Pan's Labyrinth	Spanish	220	Crouching Tiger, Hidden Dragon	Mandarin
			226	Letters from Iwo Jima	Japanese

Best Directors:

Task: Find out the top 10 directors for whom the mean of imdb_score is the highest and store them in a new column top10director.

director_name	Average imdb_score
Akira Kurosawa	8.7
Alfred Hitchcock	8.5
Asghar Farhadi	8.4
Charles Chaplin	8.6
Christopher Nolan	8.425
Damien Chazelle	8.5
Majid Majidi	8.5
Richard Marquand	8.4
Ron Fricke	8.5
Sergio Leone	8.4333333333
Tony Kaye	8.6

Top 10 director's profit

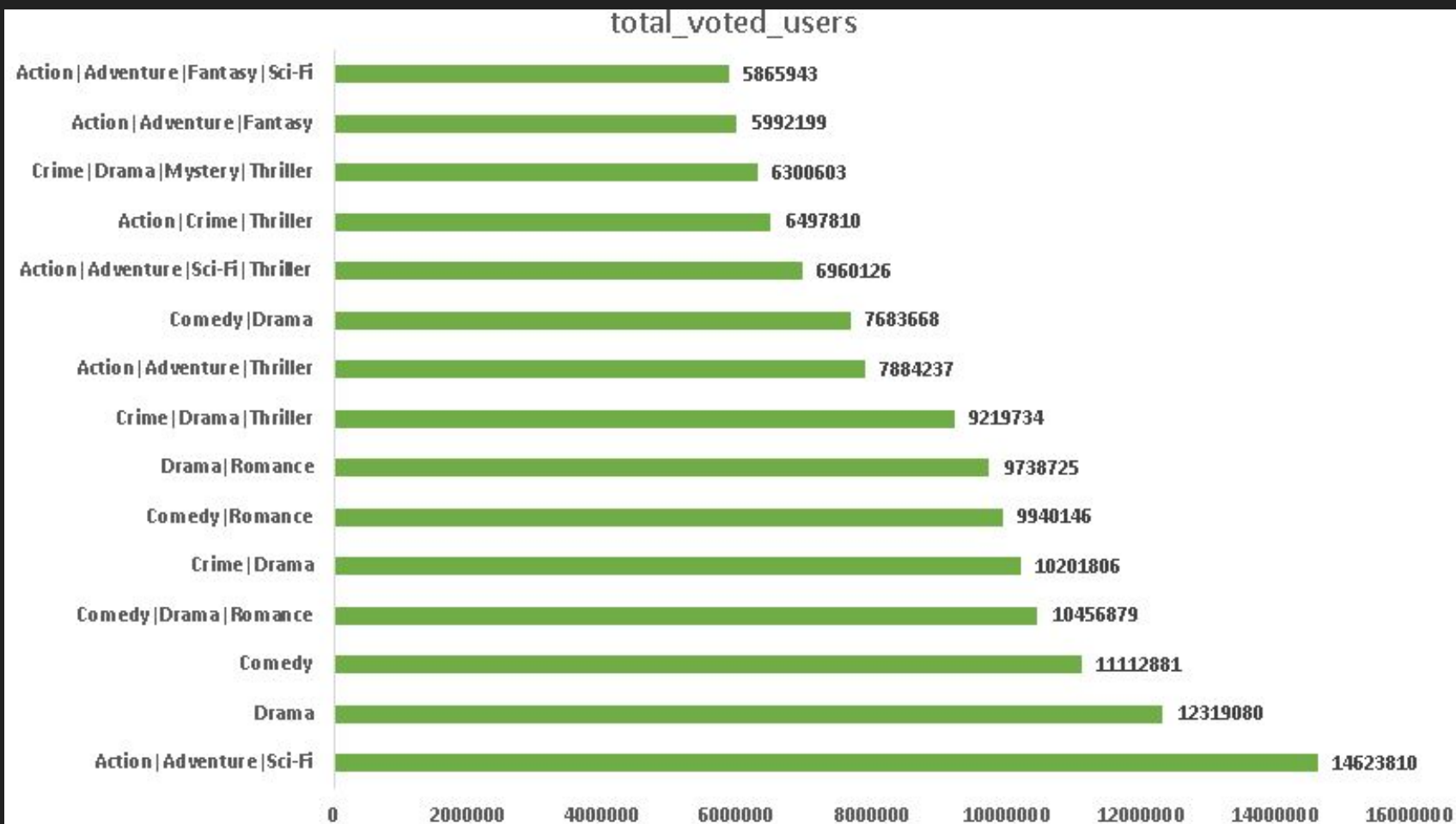


Popular Genres:

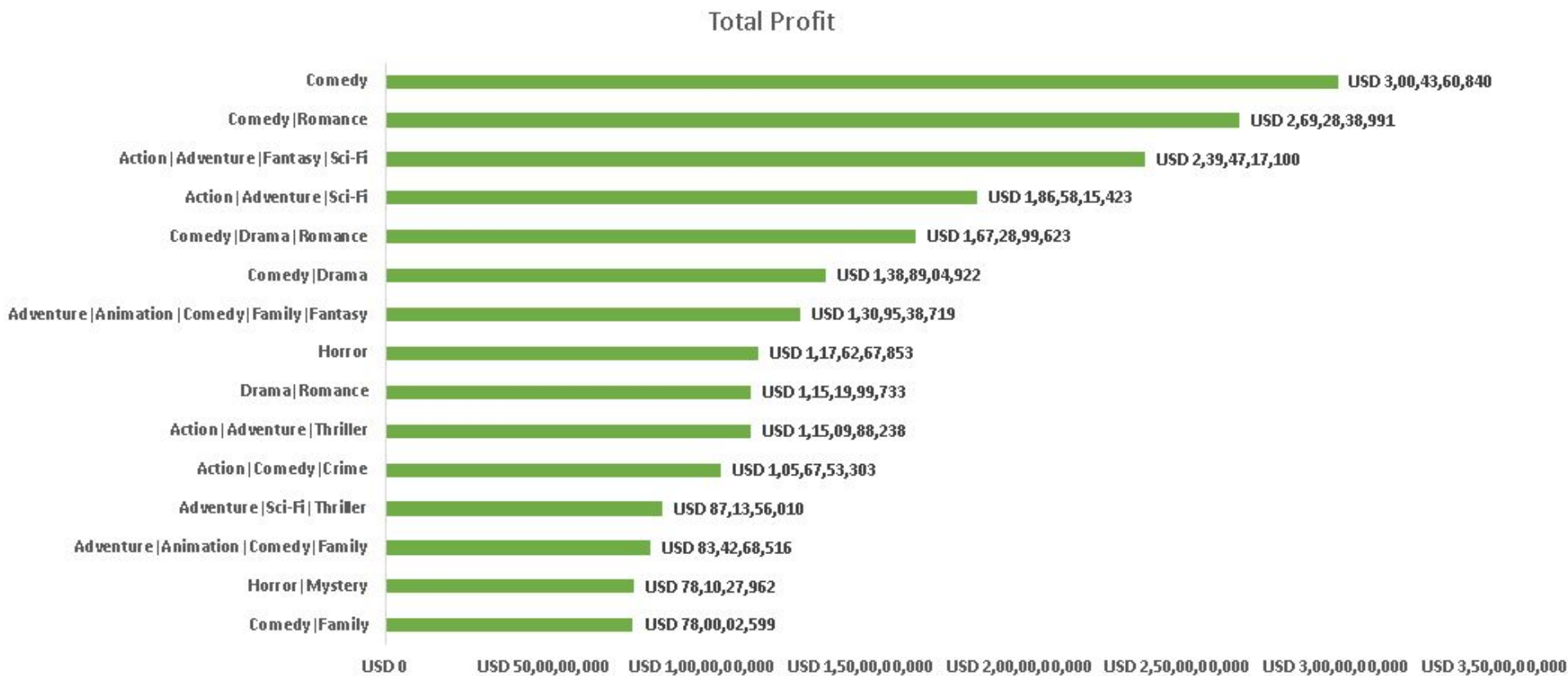
Perform this step using the knowledge gained while performing previous steps.

Task: Find popular genres.

Popular Genres based on total voted users:



Popular Genres based on total Profits:



Popular genres based on average IMDb scores					
genres	▼ total_voted_users	▼ Total profit	▼ Average imdb_score	▼	
Crime Drama Fantasy Mystery	782610	USD 7,68,01,374	⬆️		8.5
Adventure Animation Drama Family Musical	644348	USD 37,77,83,777	⬆️		8.5
Adventure Animation Fantasy	221552	USD -2,39,77,01,809	↔️		8.4
Adventure Drama Thriller War	168203	USD -25,66,866	↔️		8.4
Adventure Animation Comedy Drama Family Fantasy	345198	USD 18,14,54,367	➡️		8.3
Documentary Drama Sport	18980	USD 71,30,611	➡️		8.3
Biography Drama History Music	270790	USD 3,36,00,000	➡️		8.3
Adventure Drama War	1034619	USD 6,97,23,073	⬇️		8.25
Drama Fantasy War	467234	USD 2,41,23,143	⬇️		8.2
Drama Mystery War	80429	USD 57,096	⬇️		8.2
Biography Crime Documentary History	23836	USD -5,15,779	⬇️		8.2
Adventure Drama Thriller Western	406020	USD 4,86,35,922	⬇️		8.1
Action Animation Sci-Fi	106160	USD -1,09,95,60,838	⬇️		8.1
Documentary History Music	12631	USD 1,27,00,000	⬇️		8.1
Adventure Comedy Crime Drama	475518	USD 3,40,73,773	⬇️		8.1

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.

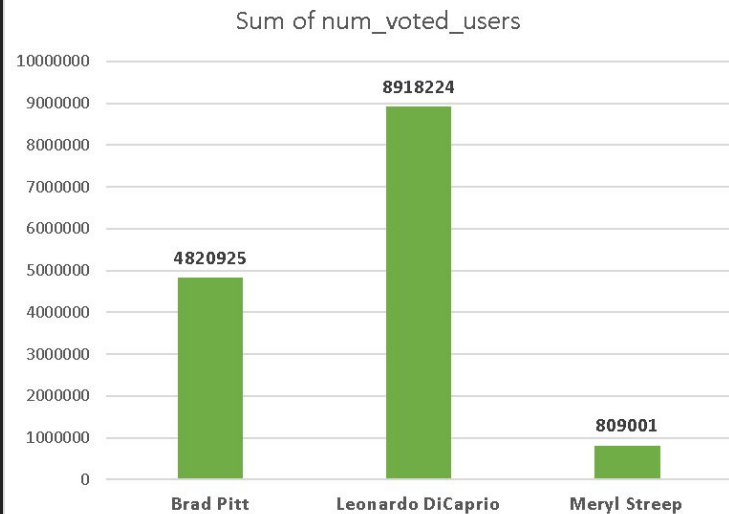
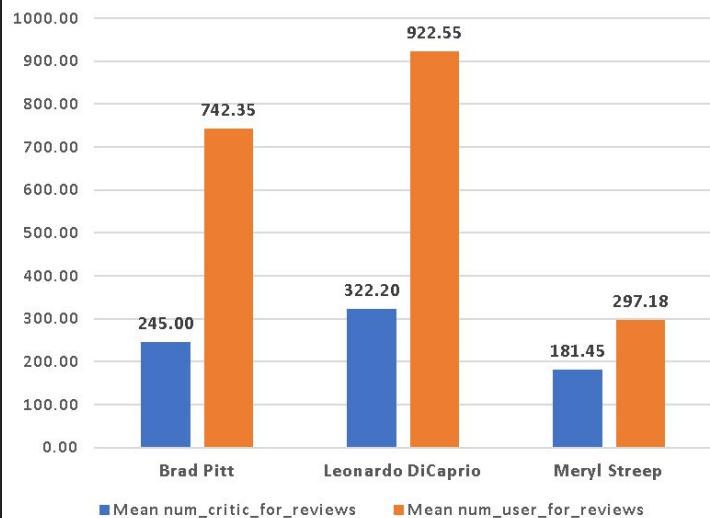
Append the rows of all these columns and store them in a new column named `Combined`.

Find the mean of the `num_critic_for_reviews` and `num_users_for_review` and identify the actors which have the highest mean.

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

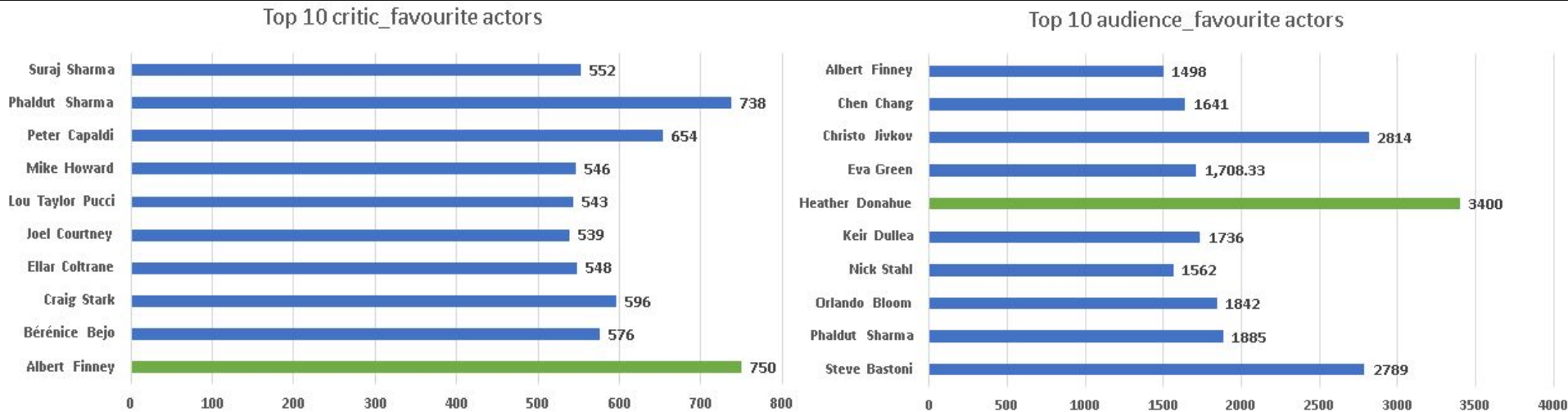
actor_1_name	total_profit	Mean imdb_score	Mean num_critic_for_reviews	Mean num_user_for_reviews	Sum of num_voted_user
Brad Pitt	-50535372	7.15	245.00	742.35	4820925
Leonardo DiCaprio	869268978	7.51	322.20	922.55	8918224
Meryl Streep	280117002	6.75	181.45	297.18	809001

Leonardo DiCaprio has gained the highest rating in terms of all insights shown here.



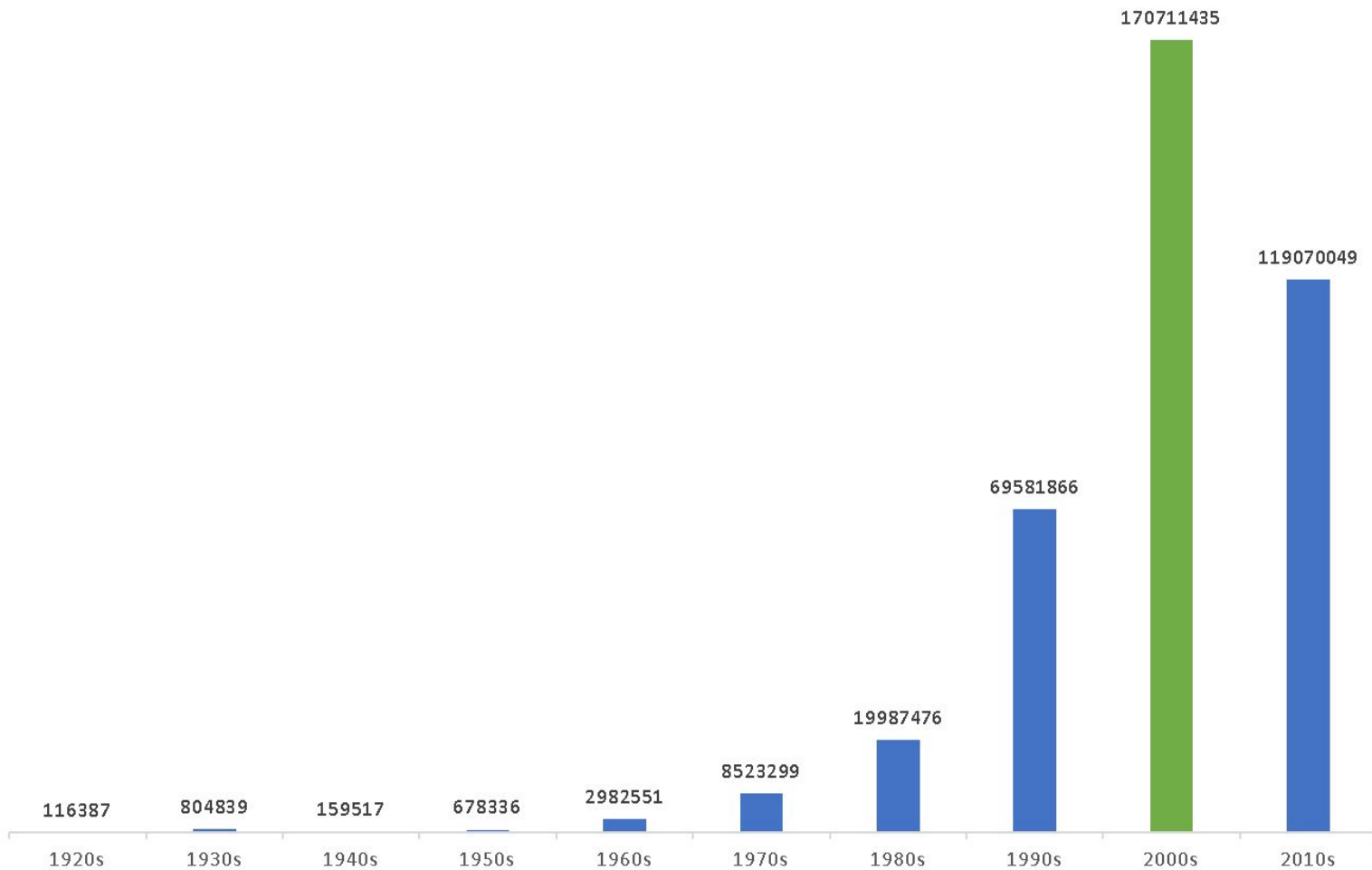
Observe the change in number of voted users over decades using a bar chart. Create a column called decade which represents the decade to which every movie belongs to. For example, the title_year year 1923, 1925 should be stored as 1920s. Sort the column based on the column decade, group it by decade and find the sum of users voted in each decade.

Task: Find the critic-favorite and audience-favorite actors



From above chart, it shows that Albert Finney has bagged the most votes for critic-favorite actor while Heather Donahue is the most audience-favorite actor.

Sum of num_voted_users decade-wise



In the 2000s decade,
There was highest
number of voters.

Conclusion:

The project successfully brings out the required insights. All the questions were answered by carrying out analysis using Excel. The insights drawn were important and it answered the required questions.

For much deep-dive understanding and other findings, you can refer to this [project link](#).

It is recommended that you open this file in MS Excel 2019 and above versions as some of the features may not be available in google spreadsheets and other lower versions of Excel.

