IMDB MOVIE ANALYSIS

Project Description:

➢ Dataset having 28 columns, 5044 Rows of different IMDB Movies.

➢ Defined a problem, clean the data as necessary remove null and duplicate values.

➢ Explore the data set and derive insights

➢ Use root cause analysis Five 'Whys' approach.

➢ Create chart pivot table and graphs to answer given questions

➢ Give answer to asked questions and create reports for data-driven decision

Approach:

➢ Download and Data open in excel

➢ Find null and duplicate values and clean all data

➢ Process the data and solve as per asked problems.

➢ Use filter, pivot table, sum, average, count and other functions to give answer of asked questions

➢ Create charts to for easy and meaningful data representation

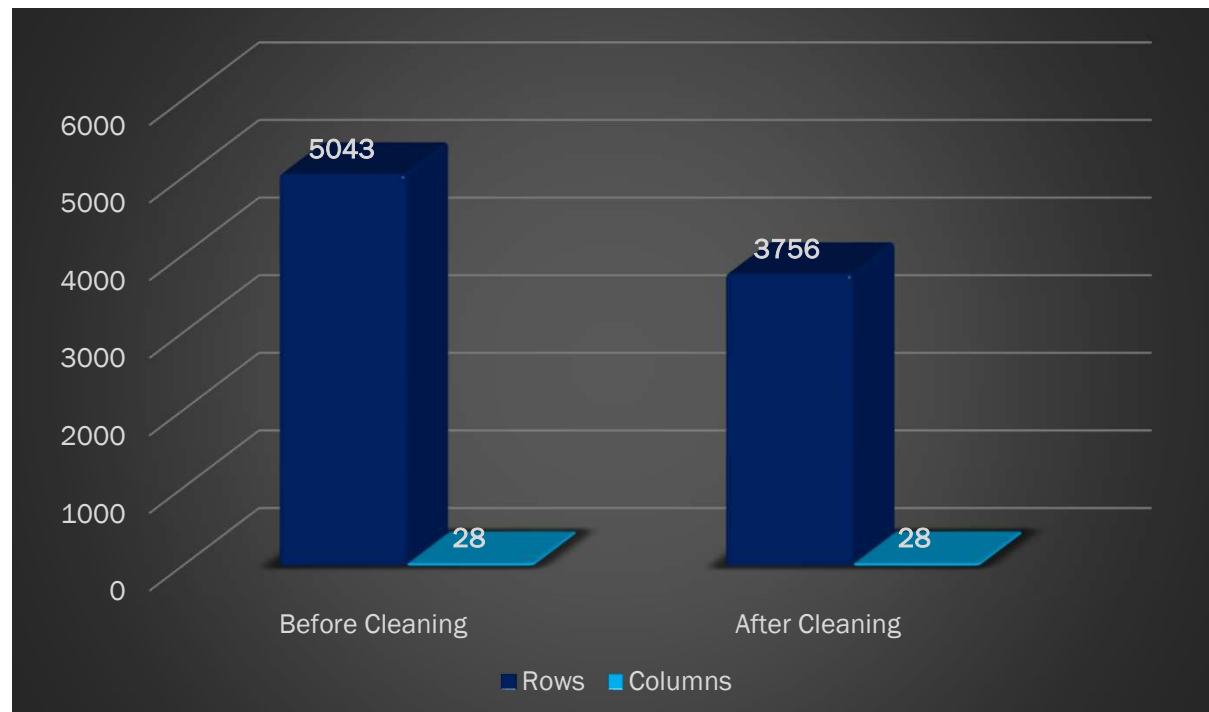➢ Create report in ppt format and submit the project

Tech-Stack Used:

➤ Microsoft Office:

 ➢MS Excel: To analyze the dataset & create charts.

 ➢MS PowerPoint: To Prepare the presentation.

➤ Google Drive: To upload the Presentation

Insights & Results:

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

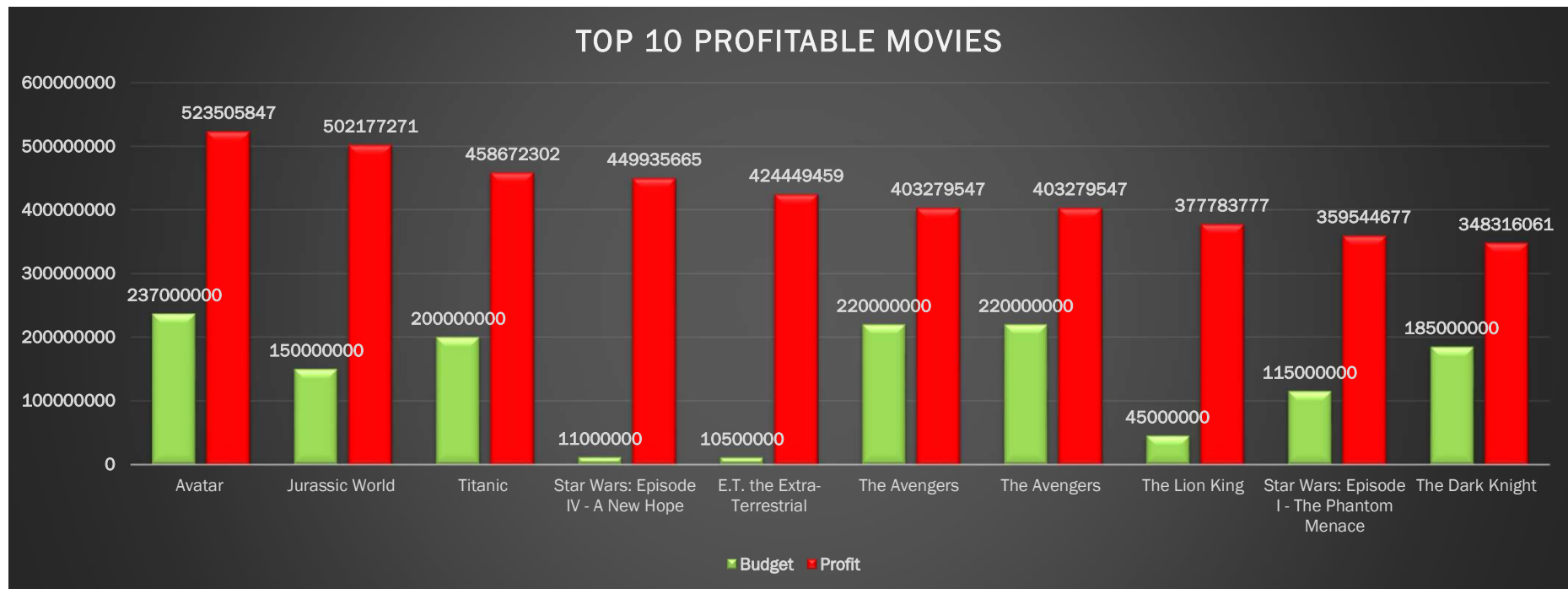|	No_of_Rows	No_of_Columns
Before Cleaning	5043	28
After Cleaning	3756	28

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?

Movie_title	Budget	Profit
Avatar	237000000	523505847
Jurassic World	150000000	502177271
Titanic	200000000	458672302
Star Wars: Episode IV - A New Hope	11000000	449935665
E.T. the Extra-Terrestrial	10500000	424449459
The Avengers	220000000	403279547
The Avengers	220000000	403279547
The Lion King	45000000	377783777
Star Wars: Episode I - The Phantom Menace	115000000	359544677
The Dark Knight	185000000	348316061

Top 10 Movies with highest Profit:

TOP 10 PROFITABLE MOVIES

Movie	Budget	Profit
Avatar	237000000	523505847
Jurassic World	150000000	502177271
Titanic	200000000	458672302
Star Wars: Episode IV - A New Hope	11000000	449935665
E.T. the Extra-Terrestrial	10500000	424449459
The Avengers	220000000	403279547
The Avengers	220000000	403279547
The Lion King	45000000	377783777
Star Wars: Episode I - The Phantom Menace	115000000	359544677
The Dark Knight	185000000	348316061

■ Budget   ■ Profit

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

Top 250 Movies:

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

Top 250 Movies:

Rank	Movie_title	IMDB_score	Rank	Movie_title	IMDB_score	Rank	Movie_title	IMDB_score
76	Some Like It Hot	8.3	101	The Big Lebowski	8.2	126	Groundhog Day	8.1
77	Scarface	8.3	102	Incendies	8.2	127	No Country for Old Men	8.1
78	Eternal Sunshine of the Spotless Mind	8.3	103	Blade Runner	8.2	128	The Revenant	8.1
79	Room	8.3	104	The Thing	8.2	129	Shutter Island	8.1
80	Monty Python and the Holy Grail	8.3	105	Casino	8.2	130	Stand by Me	8.1
81	The Hunt	8.3	106	Warrior	8.2	131	Kill Bill: Vol. 1	8.1
82	Metropolis	8.3	107	Howl's Moving Castle	8.2	132	12 Years a Slave	8.1
83	Downfall	8.3	108	The Avengers	8.1	133	Annie Hall	8.1
84	Raging Bull	8.3	109	The Avengers	8.1	134	Sin City	8.1
85	Finding Nemo	8.2	110	Deadpool	8.1	135	The Grand Budapest Hotel	8.1
86	Gone with the Wind	8.2	111	Jurassic Park	8.1	136	The Terminator	8.1
87	Captain America: Civil War	8.2	112	The Sixth Sense	8.1	137	Spotlight	8.1
88	Gran Torino	8.2	113	Monsters, Inc.	8.1	138	The Best Years of Our Lives	8.1
89	A Beautiful Mind	8.2	114	Pirates of the Caribbean: The Curse of the Black Pearl	8.1	139	The Wizard of Oz	8.1
90	Die Hard	8.2	115	Guardians of the Galaxy	8.1	140	There Will Be Blood	8.1
91	How to Train Your Dragon	8.2	116	The Help	8.1	141	Prisoners	8.1
92	The Bridge on the River Kwai	8.2	117	Platoon	8.1	142	The Princess Bride	8.1
	Pan's Labyrinth	8.2	118	The Martian	8.1	143	Hotel Rwanda	8.1
94	The Secret in Their Eyes	8.2	119	The Bourne Ultimatum	8.1	144	Mad Max: Fury Road	8.1
95	The Wolf of Wall Street	8.2	120	Rocky	8.1	145	Amores Perros	8.1
96	V for Vendetta	8.2	121	Gone Girl	8.1	146	Before Sunrise	8.1
97	Trainspotting	8.2	122	Butch Cassidy and the Sundance Kid	8.1	147	The Celebration	8.1
98	On the Waterfront	8.2	123	The Imitation Game	8.1	148	Donnie Darko	8.1
99	Into the Wild	8.2	124	Million Dollar Baby	8.1	149	Elite Squad	8.1
100	Lock, Stock and Two Smoking Barrels	8.2	125	The Truman Show	8.1	150	The Sea Inside	8.1

Top 250 Movies:

Rank	Movie_title	IMDB_score	Rank	Movie_title	IMDB_score
151	Rush	8.1	176	JFK	8
152	Tae Guk Gi: The Brotherhood of War	8.1	177	The Artist	8
153	Akira	8.1	178	Sling Blade	8
154	Jaws	8	179	Dallas Buyers Club	8
155	The Exorcist	8	180	Boyhood	8
156	Aladdin	8	181	Bowling for Columbine	8
157	The Incredibles	8	182	Casino Royale	8
158	Dances with Wolves	8	183	Casino Royale	8
159	The Sound of Music	8	184	Sicko	8
160	Rain Man	8	185	Shaun of the Dead	8
161	Slumdog Millionaire	8	186	Life of Pi	8
162	The King's Speech	8	187	The Perks of Being a Wallflower	8
163	Catch Me If You Can	8	188	A Fistful of Dollars	8
164	Star Trek	8	189	Before Sunset	8
165	The Pursuit of Happyness	8	190	Central Station	8
166	Doctor Zhivago	8	191	Her	8
167	Black Swan	8	192	Waltz with Bashir	8
168	District 9	8	193	True Romance	8
169	Young Frankenstein	8	194	Persepolis	8
170	Dead Poets Society	8	195	Big Fish	8
171	Mystic River	8	196	The Straight Story	8
172	Ratatouille	8	197	Brazil	8
173	Fiddler on the Roof	8	198	In Bruges	8
174	Kill Bill: Vol. 2	8	199	Mulholland Drive	8
175	X-Men: Days of Future Past	8	200	My Name Is Khan	8

Top 250 Movies:

Rank	Movie_title	IMDB_score	Rank	Movie_title	IMDB_score
201	Dancer in the Dark	8	226	Halloween	7.9
202	Magnolia	8	227	Halloween	7.9
203	Serenity	8	228	Edward Scissorhands	7.9
204	Cinderella Man	8	229	The Hobbit: The Desolation of Smaug	7.9
205	Blood Diamond	8	230	How to Train Your Dragon 2	7.9
206	The Iron Giant	8	231	The Blues Brothers	7.9
207	Avatar	7.9	232	Nightcrawler	7.9
208	E.T. the Extra-Terrestrial	7.9	233	Do the Right Thing	7.9
209	Shrek	7.9	234	The Wrestler	7.9
210	Iron Man	7.9	235	Hot Fuzz	7.9
211	Toy Story 2	7.9	236	The Remains of the Day	7.9
212	Straight Outta Compton	7.9	237	Boogie Nights	7.9
213	Taken	7.9	238	The Hateful Eight	7.9
214	Crouching Tiger, Hidden Dragon	7.9	239	Once	7.9
215	Walk the Line	7.9	240	Glory	7.9
216	The Fighter	7.9	241	Glory	7.9
217	The Bourne Identity	7.9	242	Before Midnight	7.9
218	Big Hero 6	7.9	243	4 Months, 3 Weeks and 2 Days	7.9
219	My Fair Lady	7.9	244	Moon	7.9
220	Captain Phillips	7.9	245	Nine Queens	7.9
221	Little Miss Sunshine	7.9	246	The Chorus	7.9
222	The Untouchables	7.9	247	Letters from Iwo Jima	7.9
223	Crash	7.9	248	The Right Stuff	7.9
224	Crash	7.9	249	Amour	7.9
225	Halloween	7.9	250	Ed Wood	7.9

TOP_FOREIGN_LANG_FILM:

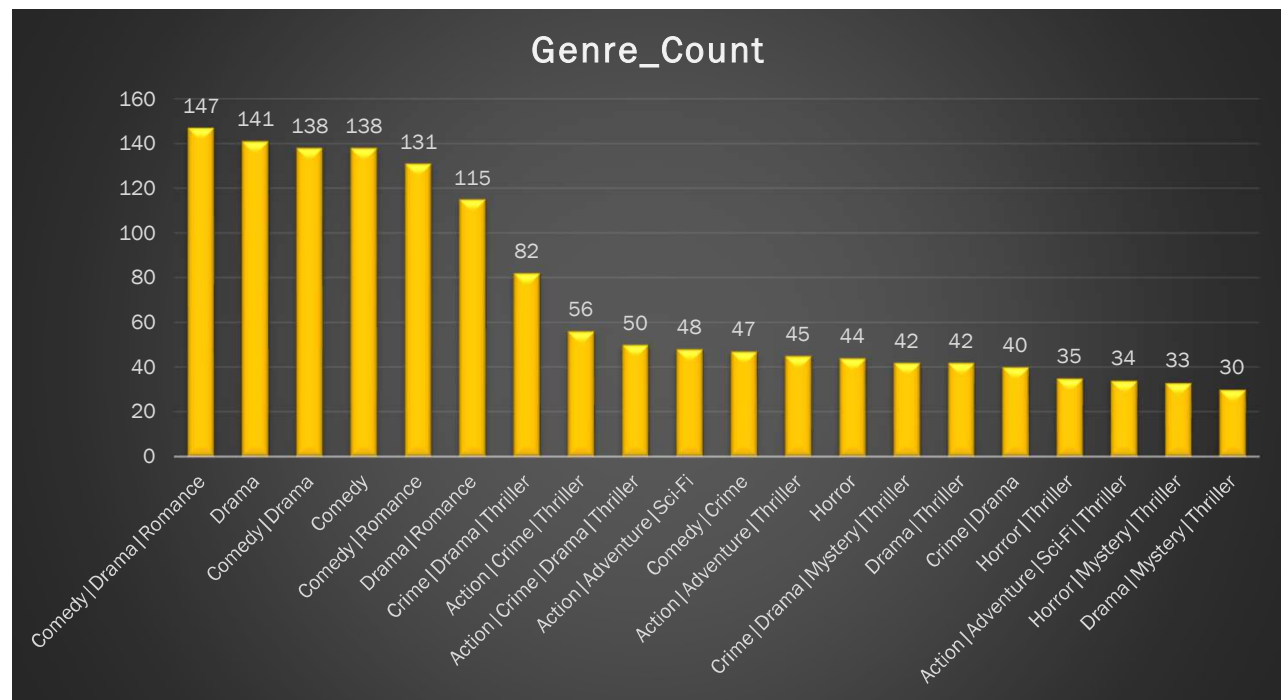| TOP_FOREIGN_LANG_FILM	Language		
The Good, the Bad and the Ugly	Italian	Akira	Japanese
City of God	Portuguese	A Fistful of Dollars	Italian
Seven Samurai	Japanese	Central Station	Portuguese
Spirited Away	Japanese	Waltz with Bashir	Hebrew
The Lives of Others	German		
Children of Heaven	Persian	Persepolis	French
Samsara	None	My Name Is Khan	Hindi
A Separation	Persian		
Oldboy	Korean	Crouching Tiger, Hidden Dragon	Mandarin
Das Boot	German	4 Months, 3 Weeks and 2 Days	Romanian
Amélie	French		
Princess Mononoke	Japanese	Nine Queens	Spanish
The Hunt	Danish	The Chorus	French
Metropolis	German		
Downfall	German	The Second Mother	Portuguese
Pan's Labyrinth	Spanish	Letters from Iwo Jima	Japanese
The Secret in Their Eyes	Spanish		
Incendies	French	Amour	French
The Act of Killing	Indonesian		
Howl's Moving Castle	Japanese	Ernest & Celestine	French
Amores Perros	Spanish	Hero	Mandarin
The Celebration	Danish		
Elite Squad	Portuguese		
The Sea Inside	Spanish		
Tae Guk Gi: The Brotherhood of War	Korean		

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

Director_Name	Mean_IMDB_Score
Akira Kurosawa	8.7
Charles Chaplin	8.6
Tony Kaye	8.6
Alfred Hitchcock	8.5
Damien Chazelle	8.5
Majid Majidi	8.5
Ron Fricke	8.5
Sergio Leone	8.4
Christopher Nolan	8.4
Asghar Farhadi	8.4

E. Popular Genres: Perform this step using the knowledge gained while performing previous steps.
   Your task: Find popular genres

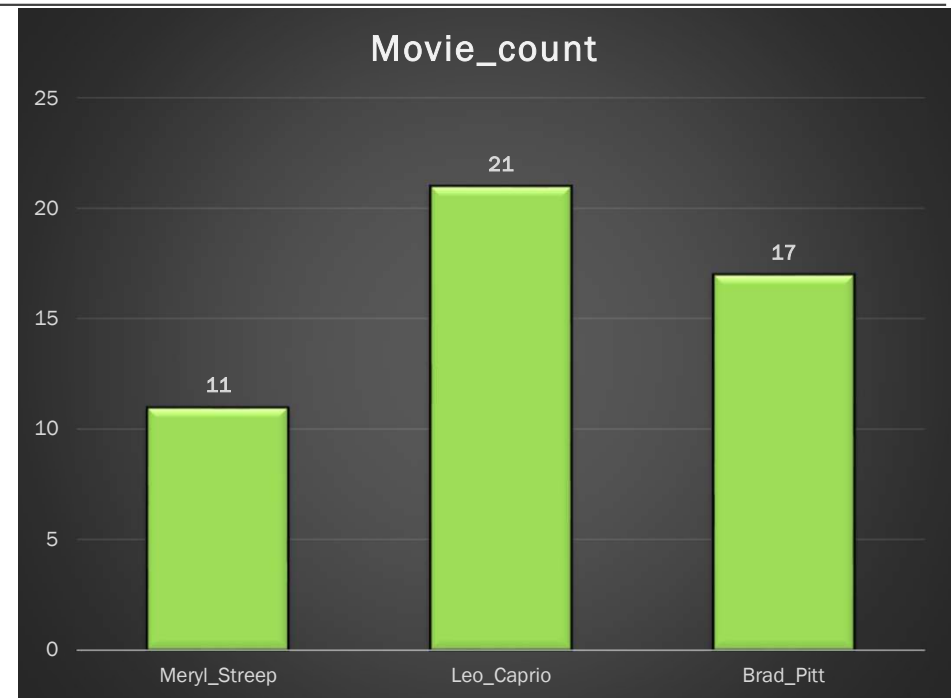| Genre	Count
Comedy|Drama|Romance	147
Drama	141
Comedy|Drama	138
Comedy	138
Comedy|Romance	131
Drama|Romance	115
Crime|Drama|Thriller	82
Action|Crime|Thriller	56
Action|Crime|Drama|Thriller	50
Action|Adventure|Sci-Fi	48
Comedy|Crime	47
Action|Adventure|Thriller	45
Horror	44
Crime|Drama|Mystery|Thriller	42
Drama|Thriller	42
Crime|Drama	40
Horror|Thriller	35
Action|Adventure|Sci-Fi|Thriller	34
Horror|Mystery|Thriller	33
Drama|Mystery|Thriller	30

Movies Filtered based on actor_1_name for Meryl Streep, Leonardo DiCaprio, Brad Pitt
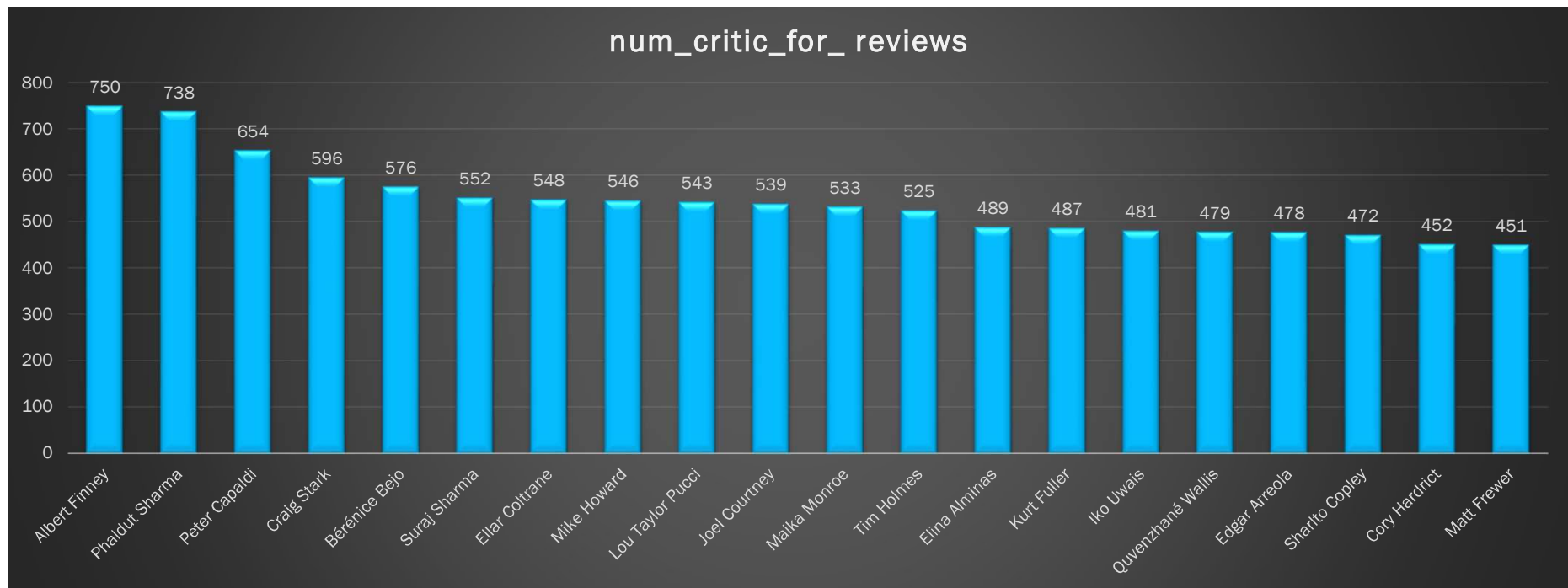
Meryl_Streep	Leo_Caprio	Brad_Pitt
The Hours	Inception	Fight Club
Out of Africa	Django Unchained	True Romance
Julie & Julia	The Departed	Ocean's Eleven
One True Thing	The Wolf of Wall Street	The Curious Case of Benjamin Button
The Devil Wears Prada	The Revenant	Interview with the Vampire: The Vampire Chronicles
A Prairie Home Companion	Shutter Island	Fury
It's Complicated	Catch Me If You Can	Babel
The Iron Lady	Blood Diamond	The Assassination of Jesse James by the Coward Robert Ford
Hope Springs	Titanic	Troy
The River Wild	The Aviator	Seven Years in Tibet
Lions for Lambs	Gangs of New York	Spy Game
	The Great Gatsby	The Tree of Life
	The Great Gatsby	Sinbad: Legend of the Seven Seas
	Revolutionary Road	Mr. & Mrs. Smith
	Body of Lies	Ocean's Twelve
	Romeo + Juliet	Killing Them Softly
	Marvin's Room	By the Sea
	J. Edgar	
	The Beach	
	The Man in the Iron Mask	
	The Quick and the Dead	

Movies Filtered based on actor_1_name for Meryl Streep, Leonardo DiCaprio, Brad Pitt

Actors	Movie_count
Meryl_Streep	11
Leo_Caprio	21
Brad_Pitt	17

Combined:

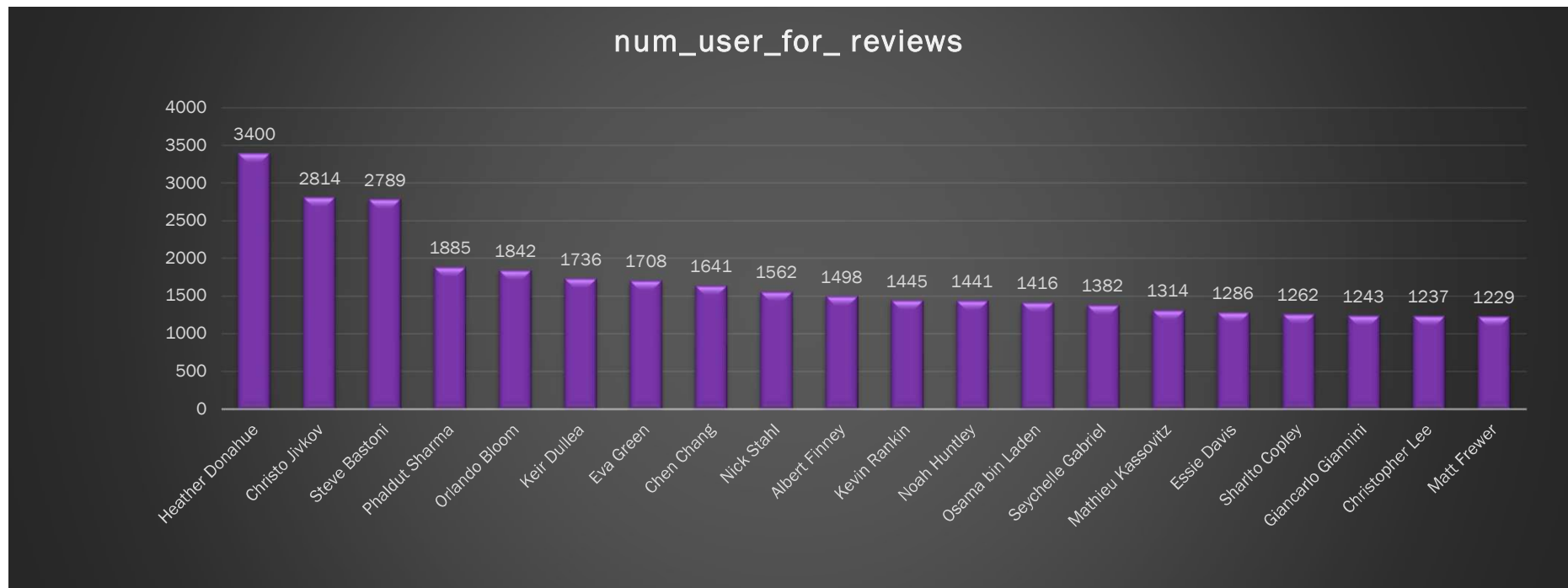| Combined |
| --- |
| A Prairie Home Companion |
| Babel |
| Blood Diamond |
| Body of Lies |
| By the Sea |
| Catch Me If You Can |
| Django Unchained |
| Fight Club |
| Fury |
| Gangs of New York |
| Hope Springs |
| Inception |
| Interview with the Vampire: The Vampire Chronicles |
| It's Complicated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

Mean of num_critic_for_reviews:

num_critic_for_ reviews

Mean of num_users_for_review:

Sum of users voted in each decade:

Decade	num_voted_users
1920s	116387
1930s	804839
1940s	159517
1950s	678336
1960s	2982551
1970s	8681156
1980s	20091781
1990s	69934957
2000s	172800132
2010s	121235552