

# DECODING IMDB REVIEWS

UNCOVERING INSIGHTS THROUGH DATA ANALYTICS



# CONTENT

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

# PROJECT DESCRIPTION

- **IMBD IS A WELL-KNOWN WEBSITE THAT ENABLES USERS AND REVIEWERS TO RATE FILMS AND TELEVISION SHOWS. WE MAY DISCOVER REVIEWS OF FILMS AND TELEVISION SHOWS ON IMBD, ALONG WITH BIOS OF THE CAST MEMBERS AND DIRECTORS. HERE, THE IMBD DATASET FOR MOTION PICTURES FROM 1927 TO 2016 IS MADE AVAILABLE. THIS INCLUDES DETAILS ON THE PERFORMERS, DIRECTORS, BUDGET, AND OTHER ASPECTS OF THE FILM. WE'RE GOING TO USE OFFICE 365 EXCEL'S APPROACH TO CLEAN THE DATASET AND OBTAIN THE REQUESTED INFORMATION.**

# APPROACH

- **IN ORDER TO COMPLETE THIS ASSIGNMENT, WE MUST FIRST COMPREHEND THE PROVIDED DATA.**
- **THEN, WE WILL CLEAN THE DATA TO MEET OUR NEEDS BY ELIMINATING NULL VALUES, EXTRANEIOUS COLUMNS, ETC.**
- **FOLLOWING THE CLEANING, WE WILL USE PIVOT TABLES, SEVERAL FUNCTIONS, AND CHARTS TO OBTAIN THE NECESSARY ANSWERS TO THE QUERIES.**
- **WE'LL KEEP ASKING WHY QUESTIONS OF THE DATA TO GO FURTHER INTO THE ISSUE'S CAUSE.**
- **AT THE CONCLUSION, WE'LL STRUCTURE OUR RESPONSES PROPERLY IN A TABLE AND SOME GRAPHS.**

# TECH-STACK USED

- **MICROSOFT EXCEL ( 2021 )**
- **GOOGLE SHEETS**

# INSIGHTS

- **TASK 1 : CLEAN THE DATA**
- **DROP UNNECESSARY COLUMNS( COLOR, DIRECTOR\_FACEBOOK\_LIKES, ACTOR\_1\_FACEBOOK\_LIKES, ACTOR\_2\_FACEBOOK\_LIKES, ACTOR\_3\_FACEBOOK\_LIKES, ACTOR\_2\_NAME, CAST\_TOTAL\_FACEBOOK\_LIKES, ACTOR\_3\_NAME, DURATION, FACENUMBER\_IN\_POSTER, CONTENT\_RATING, MOVIE\_IMDB\_LINK, ASPECT\_RATIO, PLOT\_KEYWORDS )**
- **REMOVE NULL VALUE.**
- **REMOVE DUPLICATES.**
- **LINK :** [https://bit.ly/task\\_5\\_1](https://bit.ly/task_5_1)



- **TASK 2 :FIND THE MOVIES WITH THE HIGHEST PROFIT**
- **TO CALCULATE THE PROFIT, WE MUST FIRST SUBTRACT THE BUDGETED AMOUNT FROM THE GROSS SUM.**
- **NEXT, WE'LL PLOT VALUES BY UTILISING THE SCATTER PLOT OPTION.**
- **THEN, WE WILL VISUALISE THE VALUES OF PROFIT (Y\_AXIS) AND BUDGET (X\_AXIS) USING THE SCATTER PLOT OPTION.**
- **THEN, USING THE GRAPH, WE WILL IDENTIFY ANY OUTLIERS. THEN, USING THE GRAPH, WE WILL IDENTIFY ANY OUTLIERS.**

TOP 5 HIGHEST PROFITED MOVIES	
Movie Title	Profit
Avatar	523505847
Jurassic World	502177271
Titanic	458672302
Star Wars: Episode IV - A New Hope	449935665
E.T. the Extra-Terrestrial	424449459

**LINK:** [https://bit.ly/task\\_5\\_2](https://bit.ly/task_5_2)



- **TASK 3:- FIND IMDB TOP 250**
- **USE THE FILTER TO EXCLUDE ANY DATA WHEN NUM\_VOTED\_USERS EXCEEDS 25,000.**
- **THE IMBD\_SCORE COLUMN IS USED TO DESCENDINGLY RANK THE DATA. WE WILL ONLY BE ANALYSING THE TOP 250 ENTRIES.**
- **WE'LL USE THE FORMULA “=RANK(M2,\$M\$2:\$M\$251,0)+COUNTIFS(\$M\$2:M2,M2)-1” TO GET RANK.**
- **REMOVE THE LANGUAGE FILTER BY DESELECTING ENGLISH. IT RESULTS IN FOREIGN FILMS BEING ON OUR TOP 250 LIST.**

Rank	IMDB Top 250
1	The Shawshank Redemption
2	The Godfather
3	The Dark Knight
4	The Godfather: Part II
5	The Lord of the Rings: The Return of the King
6	Schindler's List
7	Pulp Fiction
8	The Good, the Bad and the Ugly
9	Inception
10	The Lord of the Rings: The Fellowship of the Ring
11	Fight Club
12	Forrest Gump
13	Star Wars: Episode V - The Empire Strikes Back
14	The Lord of the Rings: The Two Towers
15	The Matrix
16	Goodfellas
17	Star Wars: Episode IV - A New Hope
18	One Flew Over the Cuckoo's Nest
19	City of God
20	Seven Samurai
21	Interstellar
22	Saving Private Ryan
23	Se7en
24	The Silence of the Lambs
25	Spirited Away
26	American History X

27	The Usual Suspects
28	Modern Times
29	The Dark Knight Rises
30	Gladiator
31	Terminator 2: Judgment Day
32	Django Unchained
33	The Departed
34	The Lion King
35	The Green Mile
36	The Prestige
37	The Pianist
38	Apocalypse Now
39	Raiders of the Lost Ark
40	Psycho
41	Back to the Future
42	Alien
43	Memento
44	Whiplash
45	The Lives of Others
46	Children of Heaven
47	WALL·E
48	Braveheart
49	Amélie
50	Baahubali: The Beginning
51	Star Wars: Episode VI - Return of the Jedi
52	Once Upon a Time in America
53	Princess Mononoke

54	Aliens
55	American Beauty
56	Lawrence of Arabia
57	Das Boot
58	Requiem for a Dream
59	Oldboy
60	Reservoir Dogs
61	A Separation
62	Toy Story 3
63	Up
64	Inside Out
65	Batman Begins
66	Inglourious Basterds
67	Indiana Jones and the Last Crusade
68	L.A. Confidential
69	Toy Story
70	Scarface
71	Eternal Sunshine of the Spotless Mind
72	Amadeus
73	Raging Bull
74	Metropolis
75	Unforgiven
76	Downfall
77	Room
78	Snatch
79	2001: A Space Odyssey
80	Good Will Hunting

81	The Sting
82	The Hunt
83	Some Like It Hot
84	Monty Python and the Holy Grail
85	Captain America: Civil War
86	How to Train Your Dragon
87	The Wolf of Wall Street
88	Finding Nemo
89	A Beautiful Mind
90	Casino
91	V for Vendetta
92	The Thing
93	Die Hard
94	Blade Runner
95	Gran Torino
96	Warrior
97	Howl's Moving Castle
98	Into the Wild
99	Pan's Labyrinth
100	The Big Lebowski
101	Incendies
102	Gone with the Wind
103	The Secret in Their Eyes
104	Trainspotting
105	The Bridge on the River Kwai
106	Lock, Stock and Two Smoking Barrels
107	On the Waterfront

108	The Avengers
109	Guardians of the Galaxy
110	Mad Max: Fury Road
111	The Revenant
112	The Bourne Ultimatum
113	Pirates of the Caribbean: The Curse of the Black
114	Monsters, Inc.
115	The Martian
116	Shutter Island
117	Jurassic Park
118	Gone Girl
119	The Truman Show
120	Deadpool
121	Kill Bill: Vol. 1
122	Prisoners
123	The Sixth Sense
124	Sin City
125	Rush
126	The Grand Budapest Hotel
127	Million Dollar Baby
128	The Help
129	No Country for Old Men
130	There Will Be Blood
131	12 Years a Slave
132	Spotlight
133	Hotel Rwanda
134	The Imitation Game

135	The Princess Bride
136	Groundhog Day
137	The Sea Inside
138	Tae Guk Gi: The Brotherhood of War
139	Stand by Me
140	Akira
141	Elite Squad
142	The Terminator
143	Platoon
144	Butch Cassidy and the Sundance Kid
145	Donnie Darko
146	Annie Hall
147	The Wizard of Oz
148	Before Sunrise
149	The Best Years of Our Lives
150	Amores Perros
151	The Celebration
152	Rocky
153	X-Men: Days of Future Past
154	Ratatouille
155	Star Trek
156	Life of Pi
157	Casino Royale
158	Blood Diamond
159	The Incredibles
160	Cinderella Man
161	Big Fish

162	The Pursuit of Happyness
163	Kill Bill: Vol. 2
164	Catch Me If You Can
165	The Iron Giant
166	JFK
167	Serenity
168	Magnolia
169	District 9
170	Mystic River
171	Aladdin
172	Rain Man
173	Her
174	Dances with Wolves
175	Dead Poets Society
176	The Artist
177	The King's Speech
178	Brazil
179	In Bruges
180	Mulholland Drive
181	Slumdog Millionaire
182	Black Swan
183	The Perks of Being a Wallflower
184	True Romance
185	Dancer in the Dark
186	The Exorcist
187	Jaws
188	Casino Royale

189	Doctor Zhivago
190	The Straight Story
191	Fiddler on the Roof
192	Sicko
193	My Name Is Khan
194	The Sound of Music
195	Persepolis
196	Dallas Buyers Club
197	Shaun of the Dead
198	Sling Blade
199	Boyhood
200	Bowling for Columbine
201	Central Station
202	Young Frankenstein
203	Before Sunset
204	Waltz with Bashir
205	A Fistful of Dollars
206	Avatar
207	The Hobbit: The Desolation of Smaug
208	Iron Man
209	Edge of Tomorrow
210	Big Hero 6
211	The Hobbit: An Unexpected Journey
212	How to Train Your Dragon 2
213	Toy Story 2
214	Children of Men
215	The Insider

216	The Hateful Eight
217	The Bourne Identity
218	Almost Famous
219	Captain Phillips
220	Shrek
221	Hero
222	The Notebook
223	Glory
224	Walk the Line
225	Straight Outta Compton
226	The Blues Brothers
227	The Right Stuff
228	Taken
229	The Untouchables
230	The World's Fastest Indian
231	Edward Scissorhands
232	Glory
233	Ed Wood
234	My Fair Lady
235	Halloween
236	Hot Fuzz
237	Crouching Tiger, Hidden Dragon
238	The Remains of the Day
239	Boogie Nights
240	Letters from Iwo Jima
241	The Fighter
242	E.T. the Extra-Terrestrial

243	Crash
244	Amour
245	Nightcrawler
246	Little Miss Sunshine
247	Veer-Zaara
248	The Wrestler
249	The Chorus
250	Do the Right Thing

**LINK:** [https://bit.ly/task\\_5\\_3\\_top250](https://bit.ly/task_5_3_top250)

**LINK:** [https://bit.ly/task\\_5\\_3\\_Top\\_Foreign\\_Lang\\_Film](https://bit.ly/task_5_3_Top_Foreign_Lang_Film)

- **TASK 4: FIND THE BEST DIRECTORS**
- **FIRST, CHOOSE THE IMDB\_SCORE COLUMN FROM THE CLEANED DATASET.**
- **THE PIVOT TABLE WILL THEN BE SELECTED.**
- **DIRECTOR\_NAME WILL BE ADDED TO THE PIVOT TABLE'S SERIES SECTION.**
- **THE PIVOT TABLE'S VALUES SECTION WILL THEN RECEIVE THE ADDITION OF THE AVERAGE IMDB\_SCORE.**
- **THE DATA WILL THEN BE SORTED BASED ON THE AVERAGE IMDB\_SCORE AFTER THAT IMDB\_SCORE IS FIRST IN DESCENDING ORDER, THEN ALPHABETICALLY BY DIRECTOR NAME.**



TOP_10_DIRECTORS	MEAN_IMDB_SCORE
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	8.433333333
Christopher Nolan	8.425
Asghar Farhadi	8.4
Marius A. Markevicius	8.4

**LINK:** [https://bit.ly/tasl\\_5\\_4\\_Top\\_10\\_Directors](https://bit.ly/tasl_5_4_Top_10_Directors)



- **TASK 5: FIND POPULAR GENRES**
- **CHOOSE THE GENRE COLUMN IN THE CLEANED DATASET FIRST.**
- **NEXT, WE'LL CHOOSE THE PIVOT TABLE OPTION.**
- **NEXT, WE'LL CHOOSE THE GENRE NAMES FOR THE ROW LABELS.**
- **THE VALUES ARE THEN CONVERTED TO A COUNT OF THE GENRES, AND THE GENRE COUNTS ARE USED TO SORT THE DATA IN DECREASING ORDER.**

Top 10 Popular genres are:-

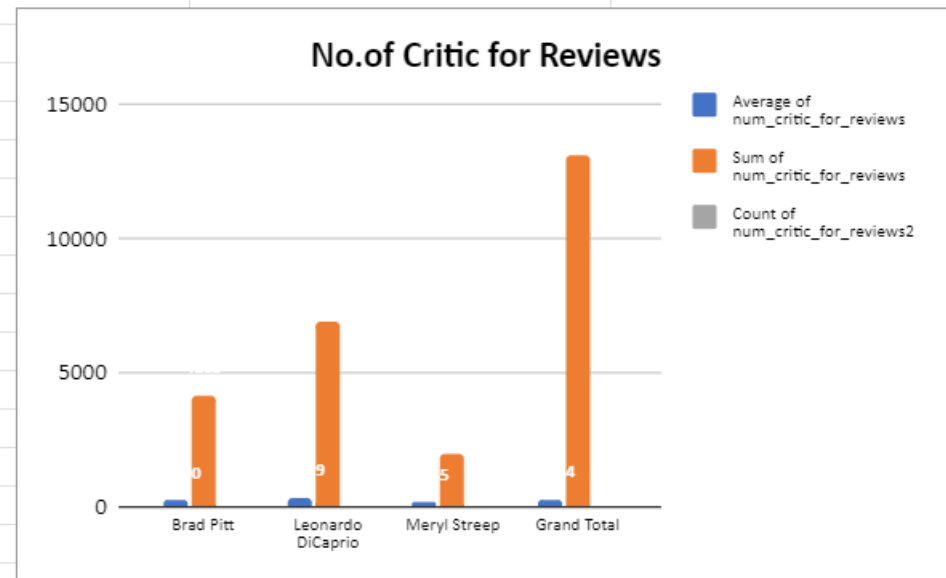
Genre name	count
Drama	153
Comedy   Drama   Romance	151
Comedy   Drama	147
Comedy	145
Comedy   Romance	135
Drama   Romance	119
Crime   Drama   Thriller	82
Action   Crime   Thriller	55
Action   Crime   Drama   Thriller	50
Action   Adventure   Sci-Fi	46

**LINK :** <https://bit.ly/3reXFHH>

- **TASK 6 : FIND THE CRITIC-FAVORITE AND AUDIENCE-FAVORITE ACTORS**
- **ADDED THREE ADDITIONAL DATA FRAMES WITH THE NAMES MERYL STREEP, LEONARDO DICAPRIO, AND BRAD PITT.**
- **THESE FRAMES FEATURED THE FILMS IN WHICH THE AFOREMENTIONED ACTORS PLAYED THE PRINCIPAL ROLES.**
- **THE ROWS OF EACH OF THESE DATA FRAMES WERE THEN ADDED TO AND SAVED IN A NEW DATA FRAME CALLED "COMBINED."**
- **NEXT, USE THE ACTOR\_1\_NAME COLUMN TO GROUP THE 'COMBINED' , CALCULATED THE MEAN OF NUM\_CRITIC\_FOR\_REVIEWS AND NUM\_USERS\_FOR\_REVIEW.**

actor_1_name	Average of num_user_for_reviews	Sum of num_user_for_reviews	Count of num_user_for_reviews2
Brad Pitt	742.35	12620	17
Leonardo DiCaprio	914.48	19204	21
Meryl Streep	297.18	3269	11
Grand Total	716.18	35093	49

actor_1_name	Average of num_critic_for_reviews	Sum of num_critic_for_reviews	Count of num_critic_for_reviews2
Brad Pitt	245.00	4165	17
Leonardo DiCaprio	330.19	6934	21
Meryl Streep	181.45	1996	11
<b>Grand Total</b>	<b>267.24</b>	<b>13095</b>	<b>49</b>



**LINK :** <https://bit.ly/3rpreZ>

Decade	Total No of voted users
1920	116387
1930	804839
1940	230838
1950	678336
1960	2983442
1970	8524102
1980	19987476
1990	69735679
2000	170904334
2010	120640346

**LINK :** <https://bit.ly/3pNxxmV>



# RESULT

- **IN ORDER TO EXTRACT THE REQUIRED FINDINGS FROM THE DATASET FOR THIS IMBD MOVIE ANALYSIS PROJECT, I HAVE DEVELOPED A RANGE OF LOGICAL, STATISTICAL, AND TECHNOLOGICAL ABILITIES. I CAN BETTER UNDERSTAND THE DATA AND EVALUATE IT WITH THE USE OF STATISTICS CONCEPTS LIKE AVERAGE, FREQUENCY TABLES, AND IDENTIFYING OUTLIERS. MY ABILITY TO USE STATISTICS AND MICROSOFT EXCEL'S TECHNICAL SKILLS TO ASSESS DATA PROCESSES USING DATA ANALYTICS ARE CONSIDERABLY EXPEDITED. SIMPLIFIES COMPLICATED CALCULATIONS THAT TAKE A LONG TIME. THROUGH ITS DATA VISUALISATION FEATURES, I ALSO GET A SENSE OF HOW EASILY UNDERSTANDABLE DATA IS WHEN IT IS PRESENTED VISUALLY. BASED ON THE FACTS AND THE DESIRED RESULTS, I LEARN WHEN TO UTILISE EACH GRAPH OR CHART.**