IMDB Movie Analysis

Project Description

- IMDB is a well-known movie and series rating site for users and critics worldwide. In IMDB we can find movies or series' ratings as well as its director's and actors compiled profiles as well as financials of it. Here we're provided with IMDB's dataset for movies from 1920-2010, which contains information about the movie, its actors, directors, budget, collection, etc.

- In this project we are going to perform analysis on the IMDB data to derive insights to the business problems.

Problem Statements

- ❑ Clean the data
- ❑ Find the movies with the highest profit
- ❑ Find IMDB Top 250 (on basis of IMDB ratings)
- ❑ Find the best directors
- ❑ Find popular genres
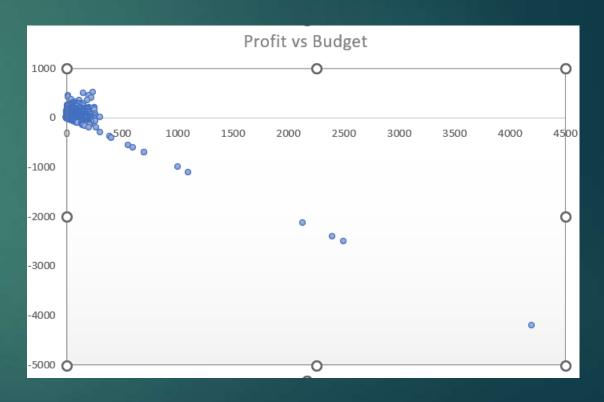- ❑ Find the critic-favourite and audience-favourite actors

Approach

➤ Firstly we have to understand the dataset provided

➤ Then clean the data as per the requirement by removing blanks, redundant values and drop columns which have no means in our project

➤ Perform analysis to derive insights as per the business problem

► 1. Clean the data

- It is the most important step to perform before doing the analysis

- As originally the datasets contains 28 columns and 5044 rows

- Dropping unnecessary columns (color, duration, director_facebook_likes, actor_3_facebook_likes, actor_2_name, actor_1_facebook_likes, cast_total_facebook_likes, actor_3_name, facenumber_in_poster, plot_keywords, movie_imdb_link, content_rating, actor_2_facebook_likes, aspect_ratio, movie_facebook_likes)

- Then removing blanks and redundant values

- Then 13 columns and 3782 rows remains

► 2. Find the movies with the highest profit

List of top 20 profitable movies

Movies	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 Lion King	45000000	377783777
Star Wars: Episode I - The Phantom Menace	115000000	359544677
The Dark Knight	185000000	348316061
The Hunger Games	78000000	329999255
Deadpool	58000000	305024263
The Hunger Games: Catching Fire	130000000	294645577
Jurassic Park	63000000	293784000
Despicable Me 2	76000000	292049635
American Sniper	58800000	291323553
Finding Nemo	94000000	286838870
Shrek 2	150000000	286471036
The Lord of the Rings: The Return of the King	94000000	283019252
Star Wars: Episode VI - Return of the Jedi	32500000	276625409
Forrest Gump	55000000	274691196

Profit vs Budget

Find IMDB Top 250

movie_title	language	imdb_score	Rank
The Shawshank Redemption	English	9.3	1
The Godfather	English	9.2	2
The Dark Knight	English	9	3
The Godfather: Part II	English	9	4
The Lord of the Rings: The Return of the King	English	8.9	5
Schindler's List	English	8.9	6
Pulp Fiction	English	8.9	7
The Good, the Bad and the Ugly	Italian	8.9	8
Inception	English	8.8	9
The Lord of the Rings: The Fellowship of the Ring	English	8.8	10
Fight Club	English	8.8	11
Forrest Gump	English	8.8	12
Star Wars: Episode V - The Empire Strikes Back	English	8.8	13
The Lord of the Rings: The Two Towers	English	8.7	14
The Matrix	English	8.7	15
Goodfellas	English	8.7	16
Star Wars: Episode IV - A New Hope	English	8.7	17
One Flew Over the Cuckoo's Nest	English	8.7	18
City of God	Portuguese	8.7	19
Seven Samurai	Japanese	8.7	20
Interstellar	English	8.6	21
Saving Private Ryan	English	8.6	22
Se7en	English	8.6	23
The Silence of the Lambs	English	8.6	24
Spirited Away	Japanese	8.6	25
American History X	English	8.6	26
The Usual Suspects	English	8.6	27
Modern Times	English	8.6	28
The Dark Knight Rises	English	8.5	29
Gladiator	English	8.5	30
Terminator 2: Judgment Day	English	8.5	31
Django Unchained	English	8.5	32
The Departed	English	8.5	33
The Lion King	English	8.5	34

movie_title	language	imdb_score	Rank
The Green Mile	English	8.5	35
The Prestige	English	8.5	36
The Pianist	English	8.5	37
Apocalypse Now	English	8.5	38
Raiders of the Lost Ark	English	8.5	39
Psycho	English	8.5	40
Back to the Future	English	8.5	41
Alien	English	8.5	42
Memento	English	8.5	43
Whiplash	English	8.5	44
The Lives of Others	German	8.5	45
Children of Heaven	Persian	8.5	46
WALLÂ·E	English	8.4	47
Braveheart	English	8.4	48
Amélie	French	8.4	49
Baahubali: The Beginning	Telugu	8.4	50
Star Wars: Episode VI - Return of the Jedi	English	8.4	51
Once Upon a Time in America	English	8.4	52
Princess Mononoke	Japanese	8.4	53
Aliens	English	8.4	54
American Beauty	English	8.4	55
Lawrence of Arabia	English	8.4	56
Das Boot	German	8.4	57
Requiem for a Dream	English	8.4	58
Oldboy	Korean	8.4	59
Reservoir Dogs	English	8.4	60
A Separation	Persian	8.4	61
Toy Story 3	English	8.3	62
Up	English	8.3	63
Inside Out	English	8.3	64
Batman Begins	English	8.3	65
Inglourious Basterds	English	8.3	66
Indiana Jones and the Last Crusade	English	8.3	67
L.A. Confidential	English	8.3	68

movie_title	language	imdb_score	Rank
Toy Story	English	8.3	69
Scarface	English	8.3	70
Eternal Sunshine of the Spotless Mind	English	8.3	71
Amadeus	English	8.3	72
Raging Bull	English	8.3	73
Metropolis	German	8.3	74
Unforgiven	English	8.3	75
Downfall	German	8.3	76
Room	English	8.3	77
Snatch	English	8.3	78
2001: A Space Odyssey	English	8.3	79
Good Will Hunting	English	8.3	80
The Sting	English	8.3	81
The Hunt	Danish	8.3	82
Some Like It Hot	English	8.3	83
Monty Python and the Holy Grail	English	8.3	84
Captain America: Civil War	English	8.2	85
How to Train Your Dragon	English	8.2	86
The Wolf of Wall Street	English	8.2	87
Finding Nemo	English	8.2	88
A Beautiful Mind	English	8.2	89
Casino	English	8.2	90
V for Vendetta	English	8.2	91
The Thing	English	8.2	92
Die Hard	English	8.2	93
Blade Runner	English	8.2	94
Gran Torino	English	8.2	95
Warrior	English	8.2	96
Howl's Moving Castle	Japanese	8.2	97
Into the Wild	English	8.2	98
Pan's Labyrinth	Spanish	8.2	99
The Big Lebowski	English	8.2	100
Incendies	French	8.2	101
Gone with the Wind	English	8.2	102

movie_title	language	imdb_score	Rank
The Secret in Their Eyes	Spanish	8.2	103
Trainspotting	English	8.2	104
The Bridge on the River Kwai	English	8.2	105
Lock, Stock and Two Smoking Barrels	English	8.2	106
On the Waterfront	English	8.2	107
The Avengers	English	8.1	108
Guardians of the Galaxy	English	8.1	109
Mad Max: Fury Road	English	8.1	110
The Revenant	English	8.1	111
The Bourne Ultimatum	English	8.1	112
Pirates of the Caribbean: The Curse of the Black Pear	English	8.1	113
Monsters, Inc.	English	8.1	114
The Martian	English	8.1	115
Shutter Island	English	8.1	116
Jurassic Park	English	8.1	117
Gone Girl	English	8.1	118
The Truman Show	English	8.1	119
Deadpool	English	8.1	120
Kill Bill: Vol. 1	English	8.1	121
Prisoners	English	8.1	122
The Sixth Sense	English	8.1	123
Sin City	English	8.1	124
Rush	English	8.1	125
The Grand Budapest Hotel	English	8.1	126
Million Dollar Baby	English	8.1	127
The Help	English	8.1	128
No Country for Old Men	English	8.1	129
There Will Be Blood	English	8.1	130
12 Years a Slave	English	8.1	131
Spotlight	English	8.1	132
Hotel Rwanda	English	8.1	133
The Imitation Game	English	8.1	134
The Princess Bride	English	8.1	135
Groundhog Day	English	8.1	136

movie_title	language	imdb_score	Rank
The Sea Inside	Spanish	8.1	137
Tae Guk Gi: The Brotherhood of War	Korean	8.1	138
Stand by Me	English	8.1	139
Akira	Japanese	8.1	140
Elite Squad	Portuguese	8.1	141
The Terminator	English	8.1	142
Platoon	English	8.1	143
Butch Cassidy and the Sundance Kid	English	8.1	144
Donnie Darko	English	8.1	145
Annie Hall	English	8.1	146
The Wizard of Oz	English	8.1	147
Before Sunrise	English	8.1	148
The Best Years of Our Lives	English	8.1	149
Amores Perros	Spanish	8.1	150
The Celebration	Danish	8.1	151
Rocky	English	8.1	152
X-Men: Days of Future Past	English	8	153
Ratatouille	English	8	154
Star Trek	English	8	155
Life of Pi	English	8	156
Casino Royale	English	8	157
Blood Diamond	English	8	158
The Incredibles	English	8	159
Cinderella Man	English	8	160
Big Fish	English	8	161
The Pursuit of Happyness	English	8	162
Kill Bill: Vol. 2	English	8	163
Catch Me If You Can	English	8	164
The Iron Giant	English	8	165
JFK	English	8	166
Serenity	English	8	167
Magnolia	English	8	168
District 9	English	8	169
Mystic River	English	8	170

movie_title	language	imdb_score	Rank
Aladdin	English	8	171
Rain Man	English	8	172
Her	English	8	173
Dances with Wolves	English	8	174
Dead Poets Society	English	8	175
The Artist	English	8	176
The King's Speech	English	8	177
Brazil	English	8	178
In Bruges	English	8	179
Mulholland Drive	English	8	180
Slumdog Millionaire	English	8	181
Black Swan	English	8	182
The Perks of Being a Wallflower	English	8	183
True Romance	English	8	184
Dancer in the Dark	English	8	185
The Exorcist	English	8	186
Jaws	English	8	187
Doctor Zhivago	English	8	188
The Straight Story	English	8	189
Fiddler on the Roof	English	8	190
Sicko	English	8	191
My Name Is Khan	Hindi	8	192
The Sound of Music	English	8	193
Persepolis	French	8	194
Dallas Buyers Club	English	8	195
Shaun of the Dead	English	8	196
Sling Blade	English	8	197
Boyhood	English	8	198
Bowling for Columbine	English	8	199
Central Station	Portuguese	8	200
Young Frankenstein	English	8	201
Before Sunset	English	8	202
Waltz with Bashir	Hebrew	8	203
A Fistful of Dollars	Italian	8	204

movie_title	language	imdb_score	Rank
Avatar	English	7.9	205
The Hobbit: The Desolation of Smaug	English	7.9	206
Iron Man	English	7.9	207
Edge of Tomorrow	English	7.9	208
Big Hero 6	English	7.9	209
The Hobbit: An Unexpected Journey	English	7.9	210
How to Train Your Dragon 2	English	7.9	211
Toy Story 2	English	7.9	212
Children of Men	English	7.9	213
The Insider	English	7.9	214
The Hateful Eight	English	7.9	215
The Bourne Identity	English	7.9	216
Almost Famous	English	7.9	217
Captain Phillips	English	7.9	218
Shrek	English	7.9	219
Hero	Mandarin	7.9	220
The Notebook	English	7.9	221
Glory	English	7.9	222
Walk the Line	English	7.9	223
Straight Outta Compton	English	7.9	224
The Blues Brothers	English	7.9	225
The Right Stuff	English	7.9	226
Taken	English	7.9	227
The Untouchables	English	7.9	228
The World's Fastest Indian	English	7.9	229
Edward Scissorhands	English	7.9	230
Ed Wood	English	7.9	231
My Fair Lady	English	7.9	232
Halloween	English	7.9	233
Hot Fuzz	English	7.9	234
Crouching Tiger, Hidden Dragon	Mandarin	7.9	235
The Remains of the Day	English	7.9	236
Boogie Nights	English	7.9	237
Letters from Iwo Jima	Japanese	7.9	238

movie_title	language	imdb_score	Rank
The Fighter	English	7.9	239
E.T. the Extra-Terrestrial	English	7.9	240
Crash	English	7.9	241
Amour	French	7.9	242
Nightcrawler	English	7.9	243
Little Miss Sunshine	English	7.9	244
Veer-Zaara	Hindi	7.9	245
The Wrestler	English	7.9	246
The Chorus	French	7.9	247
Do the Right Thing	English	7.9	248
Moon	English	7.9	249
Before Midnight	English	7.9	250

► IMDB Top Non English Movies from the list of Top 250

movie_title	num_voted_users	language	imdb_score	Rank
The Good, the Bad and the Ugly	503509	Italian	8.9	8
City of God	533200	Portuguese	8.7	19
Seven Samurai	229012	Japanese	8.7	20
Spirited Away	417971	Japanese	8.6	25
The Lives of Others	259379	German	8.5	45
Children of Heaven	27882	Persian	8.5	46
Amélie	534262	French	8.4	49
Baahubali: The Beginning	62756	Telugu	8.4	50
Princess Mononoke	221552	Japanese	8.4	53
Das Boot	168203	German	8.4	57
Oldboy	356181	Korean	8.4	59
A Separation	151812	Persian	8.4	61
Metropolis	111841	German	8.3	74
Downfall	248354	German	8.3	76
The Hunt	170155	Danish	8.3	82
Howl's Moving Castle	214091	Japanese	8.2	97
Pan's Labyrinth	467234	Spanish	8.2	99
Incendies	80429	French	8.2	101
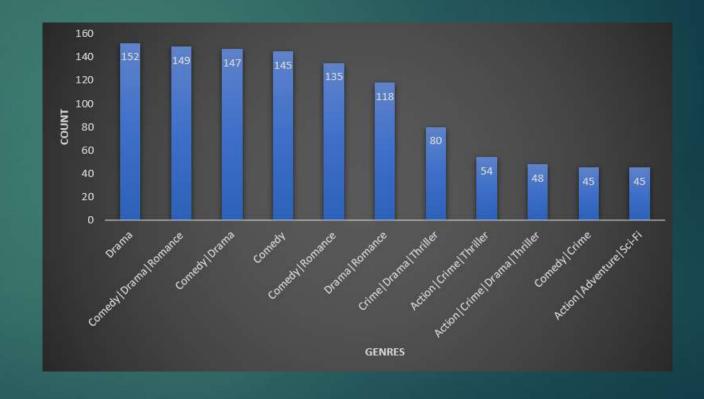
movie_title	num_voted_users	language	imdb_score	Rank
The Secret in Their Eyes	131831	Spanish	8.2	103
The Sea Inside	64556	Spanish	8.1	137
Tae Guk Gi: The Brotherhood of War	31943	Korean	8.1	138
Akira	106160	Japanese	8.1	140
Elite Squad	81644	Portuguese	8.1	141
Amores Perros	173551	Spanish	8.1	150
The Celebration	65951	Danish	8.1	151
My Name Is Khan	69759	Hindi	8	192
Persepolis	70194	French	8	194
Central Station	28951	Portuguese	8	200
Waltz with Bashir	46107	Hebrew	8	203
A Fistful of Dollars	147566	Italian	8	204
Hero	149414	Mandarin	7.9	220
Crouching Tiger, Hidden Dragon	217740	Mandarin	7.9	235
Letters from Iwo Jima	132149	Japanese	7.9	238
Amour	70382	French	7.9	242
Veer-Zaara	34449	Hindi	7.9	245
The Chorus	44151	French	7.9	247

▶ Find the best directors

Top 10 best directors as per their mean of IMDB ratings

director_name	imdb_score_mean
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
Marius A. Markevicius	8.4

Find popular genres

Genres	Count of genres
Drama	152
Comedy|Drama|Romance	149
Comedy|Drama	147
Comedy	145
Comedy|Romance	135
Drama|Romance	118
Crime|Drama|Thriller	80
Action|Crime|Thriller	54
Action|Crime|Drama|Thriller	48
Comedy|Crime	45
Action|Adventure|Sci-Fi	45

▶ Find the critic-favourite and audience-favourite actors

Actor_name	mean of num_critic_for_reviews
Albert Finney	750
Phaldut Sharma	738
Peter Capaldi	654
Craig Stark	596
Bérénice Bejo	576
Suraj Sharma	552
Ellar Coltrane	548
Mike Howard	546
Lou Taylor Pucci	543
Joel Courtney	539
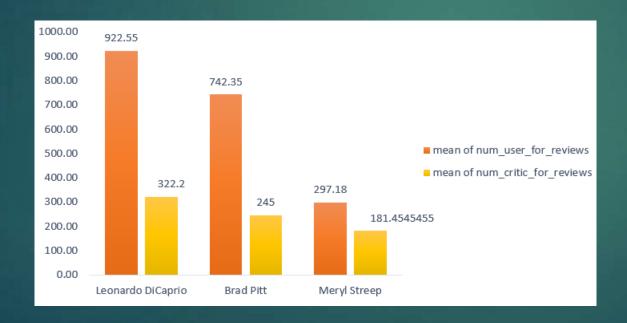
critic favourite

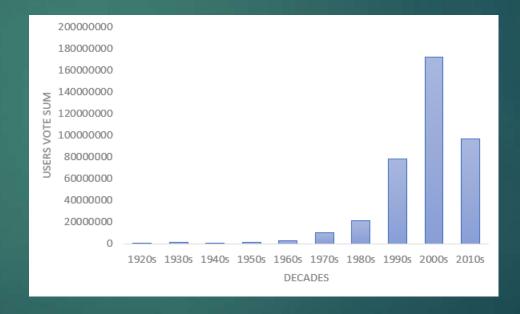| Actor_name	mean of num_user_for_reviews
Heather Donahue	3400
Christo Jivkov	2814
Steve Bastoni	2789
Phaldut Sharma	1885
Orlando Bloom	1842
Keir Dullea	1736
Chen Chang	1641
Nick Stahl	1562
Albert Finney	1498
Kevin Rankin	1445

audience favourite

► Critic and Audience favourite actor between the three Meryl Streep, Leonardo DiCaprio and Brad Pitt

actor_name	mean of num_user_for_reviews	mean of num_critic_for_reviews
Leonardo DiCaprio	922.55	322.20
Brad Pitt	742.35	245.0
Meryl Streep	297.18	181.45

▶ Numbers of voted users over decade

title_year	df_by_decade	Sum of num_voted_users
1920s	1921-1930	116387
1930s	1931-1940	966520
1940s	1941-1950	72324
1950s	1951-1960	1097601
1960s	1961-1970	2607791
1970s	1971-1980	9896107
1980s	1981-1990	21504010
1990s	1991-2000	78607120
2000s	2001-2010	172474182
2010s	2011-2020	96724334

Tech Stack Used

➤ Microsoft Excel :

For performing analysis because it is user-friendly, allows for data organization and manipulation, offers a wide range of functions and formulas, enables data visualization, has powerful pivot tables, integrates with other tools, and is widely accessible and available.

➤ Microsoft PowerPoint :

For creating a detailed report on the project. Used PowerPoint because it offers visually appealing designs, easy sharing options and presenter friendly tools.

Insights

- The most profitable movie is Avatar having a profit of 52,35,05,847

- Outlier present in the dataset movie named Lady Vengeance has the biggest loss value of -4,19,97,88,333

- Shawshank Redemption has the highest IMDB rating 9.3 among all movies

- In non english movies, The Good, the Bad and the Ugly has the highest IMDB rating of 8.9

- The director Charles Chaplin and Tony Kaye has the highest mean of IMDB score 8.6

- Most popular genre is drama followed by comedy|drama|romance

- Albert Finney is most critic favourite and Heather Donahue is audience favourite actor

- Between the three Meryl Streep, Leonardo DiCaprio and Brad Pitt. Leonardo DiCaprio is audience favourite and critic favourite

- The most numbers of users voted in 2000s followed by 2010s

Result

▶ By completing this project I gained deep understanding of excel functions, pivot table and formulas

▶ It helped in enhancing my data analysis skills

▶ It helped a lot in to gain mastery in advanced excel

▶ Improved my visualisation skills by using variety of excel charts

❑ Google Drive link for excel project file