IMDb Movie Analysis

The **Internet Movie Database** (IMDb) is an online database containing information and statistics about movies, TV shows and video games as well as actors, directors and other film industry professionals. Every year hundreds of movies get released but not all of them are successful. Movie analysis can be a long process. This is the series of actions. A movie is analyzed according to different aspects: cinematography, direction, editing, and sound. There are different ways of how a person can analyze a film. For example, short scenes and clips usually undergo the shot-by-shot analysis. The aim of the project is to analyze the IMDb movie dataset which has details about the movies, their production cost and revenue generated along with rating information. We want to come up with insights by analyzing the dataset. We are trying to analyze the dataset to find answers to the questions posed in the problem statement.

The success of a movie could be measured in terms of the following metrics:

- **Return on Investment**: Which will be a measure of the Revenue generated
- **Rating by Public**: Which will also take into account the number of votes posted.

We will look over the budget on our project.

Why do we go over the budget on our project?

A film's budget plays a dominant role throughout the film's life cycle. The budget includes all costs relating to the development, production, post-production of a film. Every filmmaker wants his movies to get high returns and good rating. But not every movie is successful because of many reasons. The reason is that sometimes high budget movies also get flopped or do not get enough return to protect their investment.

Why do they not get enough returns to protect their investment?

There can be many reasons for which the movies are not successful.
- The primary reason is that one cannot successfully plan to make a blockbuster film.
- The movie does not have a great script. You can make a terrible film with a great script, but you cannot make a great film with a terrible script. So the script must be good to entertain the audience.

Why these movies does not have great scripts or stories?

- Terribly written movies with dreadful dialogue are a huge reason for movie failure.
- The story begins too late in the script: It's not uncommon for writers to slowly ramp up to the main conflicts in their first drafts. They need time to settle in and explore the world and the characters, warming themselves up before diving head-first into the teeth of the story.
- The story is too thin.
- The characters are underwritten: Some writers make these mistakes because they are forcing their characters to do or say things to adhere to certain pre-ordained plot points. But for other writers, particularly more conceptual writers, the reasons go deeper, having to do with the way that they write.
- The story may be dubbed or may be out of public interest.
- Wrong casting

These are some mistakes done other than wrong casting by the writer in writing the script.

Why does wrong casting happen?

People expect a huge from their favorite stars and if they do not perform up to the mark they get disappointed. Take Bruce Willis. In the 1980s and 90s, with the Die Hard films, he was one of the biggest action stars in Hollywood. But when he was casted in literary adaptation The Bonfire of the Vanities (1990), his fans stayed away. They were used to see him as a wisecracking action hero, not a shady, drunken journalist. But even a huge celebrity will fails to draw in an audience, or can actually drive them away, if they're casted in a film as a kind of character that people don't expect from them. In case if the research team does not analyze the data correctly and do not use much analytical skills

the wrong casting can take place because of this they can make incorrect assumptions about what users wanted.

Why is the data analyzed wrongly?

The problem with big budget films is that there is too much risk on the line, so to mitigate that risk and protect their investment, studios use analysts and data and algorithms to try and find "the formula" that will make a movie that the largest audience will like. Analyzing characters, their roles and archetypes, and the meaning they provide to the story is a great place to start. Reasons for wrong analyzing the data are:

- The user experience research team doesn't ask effective questions.

- Concentrating too much or not Enough on Outliers.

- Focusing on the Wrong Metrics.

- Failure to Factor in Seasons.

- You Rely On Inaccurate Algorithms

- Your Models Are Bad.

PROJECT APPROACH

I will provide a detailed report for the below data record mentioning the answers of the questions that follows:

1. Cleaning the data: This is one of the most important steps to perform before moving forward with the analysis.

2. 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.

3. Top 250: Create a new column IMDb_Top_250 and store the top 250 movies with the highest IMDb Rating. 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.

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.

4. 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 top 10 director. In case of a tie in IMDb score between two directors, sort them alphabetically.

5. Popular Genres: Find popular genres.
6. Charts: Find the critic-favorite and audience-favorite actors.

HOW TO DO THIS PROJECT?

1. **Download the dataset:** I downloaded the dataset and performed the analysis using Excel sheets.

2. **Perform Analysis:** I used Excel Sheets to perform entire analysis answering the questions asked above.

3. **Submit a Report:** I made a report (PDF) to be presented to the leadership team.

THE TOOLS AND TECHNOLOGIES

The tools and technologies that I am going to use in this project would be the **IMDB_MOVIES** dataset provided by Trainity and MS Excel. Excel is a spreadsheet software application used to store, organize, and analyze data. In business, any function in any industry can benefit from those with strong Excel knowledge. Excel is a powerful tool that has become entrenched in business processes worldwide—whether for analyzing stocks or issuers, budgeting, or organizing client sales lists.

WHY I AM USING THE MS EXCEL?

- Microsoft Excel is a spreadsheet application used to manipulate stored data.

- Finance and accounting professionals choose Excel for its complex analytical and computing features.

- Microsoft Excel enables users to identify trends and organize and sort data into meaningful categories.

- Pivot tables summarize data from a worksheet into a chart.

A. Clean the data.

When I go through this data I found that there are:

Total number of columns : 28

Total number of rows : 5043

Duplicate values : 126

Missing values : 2654

Firstly, I found that there are many duplicate values in the data. So I removed these values then I found that there are many missing values so I tried to replace them. In case of categorical column (color, num_critic_for_reviews, plot_keywords, num_user_for_reviews, language, country, content_rating, aspect_ratio) by replacing these values by their respective mode.
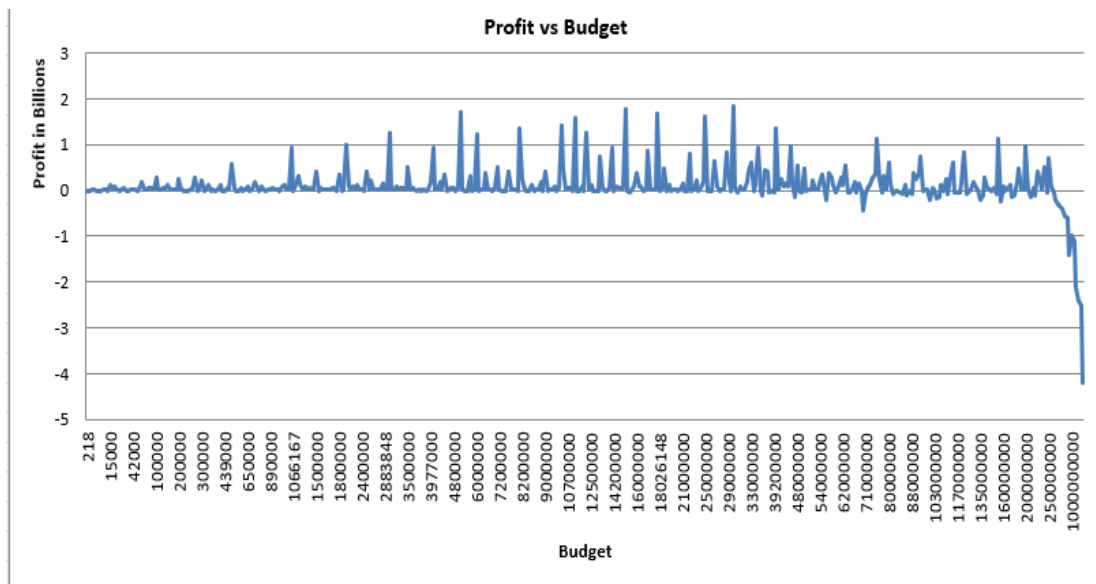
And in the case of a numerical column (duration, gross, facenumber_in_poster, budget) I filled these missing values with the median.

Now no missing values are left .Hence the data is cleaned.

B. Find the movies with the highest profit?

I created a new column called profit which contains the difference between the two columns: gross and budget. Then sort the column using the profit column as reference and here is my output for the top 20 movies with highest profit.

	movie_title	gross	budget	profit
1				
2	AvatarÂ	760505847	237000000	523505847
3	Jurassic WorldÂ	652177271	150000000	502177271
4	TitanicÂ	658672302	200000000	458672302
5	Star Wars: Episode IV - A New HopeÂ	460935665	11000000	449935665
6	E.T. the Extra-TerrestrialÂ	434949459	10500000	424449459
7	The AvengersÂ	623279547	220000000	403279547
8	The Lion KingÂ	422783777	45000000	377783777
9	Star Wars: Episode I - The Phantom MenaceÂ	474544677	115000000	359544677
10	The Dark KnightÂ	533316061	185000000	348316061
11	The Hunger GamesÂ	407999255	78000000	329999255
12	DeadpoolÂ	363024263	58000000	305024263
13	The Hunger Games: Catching FireÂ	424645577	130000000	294645577
14	Jurassic ParkÂ	356784000	63000000	293784000
15	Despicable Me 2Â	368049635	76000000	292049635
16	American SniperÂ	350123553	58800000	291323553
17	Finding NemoÂ	380838870	94000000	286838870
18	Shrek 2Â	436471036	150000000	286471036
19	The Lord of the Rings: The Return of the KingÂ	377019252	94000000	283019252
20	Star Wars: Episode VI - Return of the JediÂ	309125409	32500000	276625409

Profit vs Budget

C. Find IMDb Top 250.

I created a new sheet IMDb_Top_250 stored the top 250 movies with the highest IMDb Rating (corresponding to column: imdb_score).

Also num_voted_users is greater than 25000. And here is my output for top IMDB 250 movies.

rank	movie_title	num_voted_users	language	imdb_score
1	The Shawshank RedemptionÂ	1689764	English	9.3
2	The GodfatherÂ	1155770	English	9.2
3	The Dark KnightÂ	1676169	English	9
4	The Godfather: Part IIÂ	790926	English	9
5	12 Angry MenÂ	447785	English	8.9
6	Pulp FictionÂ	1324680	English	8.9
7	Schindler's ListÂ	865020	English	8.9
8	The Good, the Bad and the UglyÂ	503509	Italian	8.9
9	The Lord of the Rings: The Return of the KingÂ	1215718	English	8.9
10	Fight ClubÂ	1347461	English	8.8
11	Forrest GumpÂ	1251222	English	8.8
12	InceptionÂ	1468200	English	8.8
13	Star Wars: Episode V - The Empire Strikes BackÂ	837759	English	8.8
14	The Lord of the Rings: The Fellowship of the RingÂ	1238746	English	8.8
15	City of GodÂ	533200	Portuguese	8.7
16	GoodfellasÂ	728685	English	8.7
17	One Flew Over the Cuckoo's NestÂ	680041	English	8.7
18	Seven SamuraiÂ	229012	Japanese	8.7
19	Star Wars: Episode IV - A New HopeÂ	911097	English	8.7
20	The Lord of the Rings: The Two TowersÂ	1100446	English	8.7
21	The MatrixÂ	1217752	English	8.7

22	American History XÂ	782437	English	8.6
23	CasablancaÂ	387508	English	8.6
24	InterstellarÂ	928227	English	8.6
25	It's a Wonderful LifeÂ	275720	English	8.6
26	Modern TimesÂ	143086	English	8.6
27	Once Upon a Time in the WestÂ	216005	English	8.6
28	Saving Private RyanÂ	881236	English	8.6
29	Se7enÂ	1023511	English	8.6
30	Spirited AwayÂ	417971	Japanese	8.6
31	The Silence of the LambsÂ	887467	English	8.6
32	The Usual SuspectsÂ	740918	English	8.6
33	AirliftÂ	30977	Hindi	8.5
34	AlienÂ	563827	English	8.5
35	Apocalypse NowÂ	450676	English	8.5
36	Back to the FutureÂ	732212	English	8.5
37	Children of HeavenÂ	27882	Persian	8.5
38	Django UnchainedÂ	955174	English	8.5
39	Dr. Strangelove or: How I Learned to Stop Worrying and Love the BombÂ	342585	English	8.5
40	GladiatorÂ	982637	English	8.5
41	MementoÂ	845580	English	8.5
42	PsychoÂ	422432	English	8.5
43	Raiders of the Lost ArkÂ	661017	English	8.5
44	Terminator 2: Judgment DayÂ	744891	English	8.5
45	The Dark Knight RisesÂ	1144337	English	8.5
46	The DepartedÂ	873649	English	8.5
47	The Green MileÂ	782610	English	8.5
48	The Lion KingÂ	644348	English	8.5
49	The Lives of OthersÂ	259379	German	8.5
50	The PianistÂ	497946	English	8.5
51	The PrestigeÂ	844052	English	8.5
52	WhiplashÂ	399138	English	8.5
53	A SeparationÂ	151812	Persian	8.4
54	AliensÂ	488537	English	8.4
55	AmÃ©lieÂ	534262	French	8.4
56	American BeautyÂ	822500	English	8.4
57	Baahubali: The BeginningÂ	62756	Telugu	8.4
58	Batman: The Dark Knight Returns, Part 2Â	30434	English	8.4
59	BraveheartÂ	736638	English	8.4
60	Das BootÂ	168203	German	8.4
61	Lawrence of ArabiaÂ	192775	English	8.4
62	OldboyÂ	356181	Korean	8.4
63	Once Upon a Time in AmericaÂ	221000	English	8.4
64	Princess MononokeÂ	221552	Japanese	8.4
65	Rang De BasantiÂ	70233	Hindi	8.4
66	Requiem for a DreamÂ	573541	English	8.4
67	Reservoir DogsÂ	664719	English	8.4
68	Star Wars: Episode VI - Return of the JediÂ	681857	English	8.4
69	The ShiningÂ	610333	English	8.4
70	To Kill a MockingbirdÂ	215086	English	8.4
71	WALLÂ·EÂ	718837	English	8.4
72	2001: A Space OdysseyÂ	427357	English	8.3
73	AmadeusÂ	270790	English	8.3
74	Batman BeginsÂ	980946	English	8.3
75	DownfallÂ	248354	German	8.3
76	Eternal Sunshine of the Spotless MindÂ	666937	English	8.3
77	Good Will HuntingÂ	604904	English	8.3
78	Indiana Jones and the Last CrusadeÂ	515306	English	8.3
79	Inglourious BasterdsÂ	885175	English	8.3
80	Inside JobÂ	55382	English	8.3
81	Inside OutÂ	345198	English	8.3
82	Judgment at NurembergÂ	44457	English	8.3
83	L.A. ConfidentialÂ	414219	English	8.3
84	MetropolisÂ	111841	German	8.3
85	Monty Python and the Holy GrailÂ	382240	English	8.3
86	Raging BullÂ	235133	English	8.3

87	Room	161288	English	8.3
88	Scarface	537442	English	8.3
89	Singin' in the Rain	150020	English	8.3
90	Snatch	600996	English	8.3
91	Some Like It Hot	175196	English	8.3
92	Taxi Driver	507063	English	8.3
93	The Apartment	109335	English	8.3
94	The Great Escape	165638	English	8.3
95	The Hunt	170155	Danish	8.3
96	The Sting	175607	English	8.3
97	Toy Story 3	544884	English	8.3
98	Toy Story	623757	English	8.3
99	Unforgiven	277505	English	8.3
100	Up	665575	English	8.3
101	A Beautiful Mind	610568	English	8.2
102	Blade Runner	461609	English	8.2
103	Captain America: Civil War	272670	English	8.2
104	Casino	333542	English	8.2
105	Die Hard	592582	English	8.2
106	Finding Nemo	692482	English	8.2
107	Gone with the Wind	215340	English	8.2
108	Gran Torino	561773	English	8.2
109	How to Train Your Dragon	485430	English	8.2
110	Howl's Moving Castle	214091	Japanese	8.2
111	Incendies	80429	French	8.2
112	Into the Wild	426359	English	8.2
113	It Happened One Night	64888	English	8.2
114	Lage Raho Munna Bhai	27569	Hindi	8.2
115	Lock, Stock and Two Smoking Barrels	414976	English	8.2
116	Mr. Smith Goes to Washington	77392	English	8.2
117	On the Waterfront	100890	English	8.2
118	Pan's Labyrinth	467234	Spanish	8.2
119	Rebecca	87424	English	8.2
120	The Big Lebowski	537419	English	8.2
121	The Bridge on the River Kwai	149444	English	8.2
122	The Deer Hunter	232577	English	8.2
123	The Elephant Man	161972	English	8.2
124	The Secret in Their Eyes	131831	Spanish	8.2
125	The Thing	258078	English	8.2
126	The Wolf of Wall Street	780588	English	8.2
127	Trainspotting	469561	English	8.2
128	V for Vendetta	791783	English	8.2
129	Warrior	332276	English	8.2
130	12 Years a Slave	439176	English	8.1
131	A Christmas Story	104908	English	8.1
132	Akira	106160	Japanese	8.1
133	Amores Perros	173551	Spanish	8.1
134	Annie Hall	192940	English	8.1
135	Barry Lyndon	101627	English	8.1
136	Before Sunrise	183288	English	8.1
137	Butch Cassidy and the Sundance Kid	152089	English	8.1
138	Cat on a Hot Tin Roof	33741	English	8.1
139	Deadpool	479047	English	8.1
140	Donnie Darko	580999	English	8.1
141	Elite Squad	81644	Portuguese	8.1
142	Gandhi	171726	English	8.1
143	Gone Girl	569841	English	8.1
144	Groundhog Day	437418	English	8.1
145	Guardians of the Galaxy	682155	English	8.1
146	Hachi: A Dog's Tale	155249	English	8.1
147	High Noon	80193	English	8.1
148	Hotel Rwanda	264533	English	8.1
149	Jurassic Park	613473	English	8.1
150	Kill Bill: Vol. 1	735784	English	8.1
151	Mad Max: Fury Road	552503	English	8.1

152	Million Dollar BabyÂ	482064	English	8.1
153	Monsters, Inc.Â	585659	English	8.1
154	NetworkÂ	103493	English	8.1
155	No Country for Old MenÂ	612060	English	8.1
156	Pirates of the Caribbean: The Curse of the Black PearlÂ	809474	English	8.1
157	PlatoonÂ	291603	English	8.1
158	PrisonersÂ	383591	English	8.1
159	RockyÂ	375240	English	8.1
160	RushÂ	312629	English	8.1
161	Shutter IslandÂ	786092	English	8.1
162	Sin CityÂ	656640	English	8.1
163	SolarisÂ	54057	Russian	8.1
164	SpotlightÂ	195333	English	8.1
165	Stand by MeÂ	271794	English	8.1
166	Tae Guk Gi: The Brotherhood of WarÂ	31943	Korean	8.1
167	The AvengersÂ	995415	English	8.1
168	The Best Years of Our LivesÂ	40359	English	8.1
169	The Bourne UltimatumÂ	491077	English	8.1
170	The CelebrationÂ	65951	Danish	8.1
171	The Grand Budapest HotelÂ	475518	English	8.1
172	The HelpÂ	318955	English	8.1
173	The Imitation GameÂ	467613	English	8.1
174	The Man Who Shot Liberty ValanceÂ	53741	English	8.1
175	The MartianÂ	472488	English	8.1
176	The Princess BrideÂ	294163	English	8.1
177	The RevenantÂ	406020	English	8.1
178	The Sea InsideÂ	64556	Spanish	8.1
179	The Sixth SenseÂ	704766	English	8.1
180	The TerminatorÂ	600266	English	8.1
181	The Truman ShowÂ	667983	English	8.1
182	The Wizard of OzÂ	291875	English	8.1
183	There Will Be BloodÂ	372990	English	8.1
184	Touching the VoidÂ	26926	English	8.1
185	A Fistful of DollarsÂ	147566	Italian	8
186	A Streetcar Named DesireÂ	78454	English	8
187	AladdinÂ	260939	English	8
188	Before SunsetÂ	168398	English	8
189	Big FishÂ	350698	English	8
190	Black SwanÂ	551363	English	8
191	Blood DiamondÂ	400292	English	8
192	Bowling for ColumbineÂ	123090	English	8
193	BoyhoodÂ	266020	English	8
194	BrazilÂ	152306	English	8
195	Casino RoyaleÂ	470483	English	8
196	Catch Me If You CanÂ	525801	English	8
197	Central StationÂ	28951	Portuguese	8
198	Cinderella ManÂ	148238	English	8
199	Dallas Buyers ClubÂ	326494	English	8
200	Dancer in the DarkÂ	79330	English	8
201	Dances with WolvesÂ	186485	English	8
202	Days of HeavenÂ	37594	English	8
203	Dead Poets SocietyÂ	277451	English	8
204	District 9Â	531737	English	8
205	Doctor ZhivagoÂ	55816	English	8
206	Fiddler on the RoofÂ	29839	English	8
207	HerÂ	355126	English	8
208	In BrugesÂ	307639	English	8
209	JawsÂ	412454	English	8
210	JFKÂ	113472	English	8
211	Kill Bill: Vol. 2Â	512749	English	8
212	Life of PiÂ	440084	English	8
213	MagnoliaÂ	241030	English	8
214	Mulholland DriveÂ	235992	English	8
215	My Name Is KhanÂ	69759	Hindi	8
216	Mystic RiverÂ	338415	English	8
217	Night of the Living DeadÂ	87978	English	8

218	PattonÂ	76398	English	8
219	PersepolisÂ	70194	French	8
220	Rain ManÂ	383784	English	8
221	RatatouilleÂ	473887	English	8
222	Rosemary's BabyÂ	140527	English	8
223	SerenityÂ	242599	English	8
224	Shaun of the DeadÂ	395921	English	8
225	SickoÂ	66610	English	8
226	Sling BladeÂ	72443	English	8
227	Slumdog MillionaireÂ	641997	English	8
228	Star TrekÂ	504419	English	8
229	The ArtistÂ	190030	English	8
230	The Diving Bell and the ButterflyÂ	89906	French	8
231	The ExorcistÂ	284252	English	8
232	The HustlerÂ	62860	English	8
233	The IncrediblesÂ	479166	English	8
234	The Iron GiantÂ	128455	English	8
235	The King's SpeechÂ	503631	English	8
236	The Man from EarthÂ	129799	English	8
237	The Perks of Being a WallflowerÂ	351274	English	8
238	The Pursuit of HappynessÂ	338383	English	8
239	The ReturnÂ	31589	Russian	8
240	The Sound of MusicÂ	148172	English	8
241	The Straight StoryÂ	63733	English	8
242	The Wild BunchÂ	63192	English	8
243	True RomanceÂ	163492	English	8
244	Waltz with BashirÂ	46107	Hebrew	8
245	X-Men: Days of Future PastÂ	514125	English	8
246	Young FrankensteinÂ	112671	English	8
247	4 Months, 3 Weeks and 2 DaysÂ	44763	Romanian	7.9
248	Almost FamousÂ	207287	English	7.9
249	AmourÂ	70382	French	7.9
250	AvatarÂ	886204	English	7.9

Then I 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 and here is my output.

rank	movie_title	num_voted_users	language	imdb_score
8	The Good, the Bad and the UglyÂ	503509	Italian	8.9
15	City of GodÂ	533200	Portuguese	8.7
18	Seven SamuraiÂ	229012	Japanese	8.7
30	Spirited AwayÂ	417971	Japanese	8.6
33	AirliftÂ	30977	Hindi	8.5
37	Children of HeavenÂ	27882	Persian	8.5
49	The Lives of OthersÂ	259379	German	8.5
53	A SeparationÂ	151812	Persian	8.4
55	AmÃ©lieÂ	534262	French	8.4
57	Baahubali: The BeginningÂ	62756	Telugu	8.4
60	Das BootÂ	168203	German	8.4
62	OldboyÂ	356181	Korean	8.4
64	Princess MononokeÂ	221552	Japanese	8.4
65	Rang De BasantiÂ	70233	Hindi	8.4
75	DownfallÂ	248354	German	8.3
84	MetropolisÂ	111841	German	8.3
95	The HuntÂ	170155	Danish	8.3
110	Howl's Moving CastleÂ	214091	Japanese	8.2
111	IncendiesÂ	80429	French	8.2
114	Lage Raho Munna BhaiÂ	27569	Hindi	8.2
118	Pan's LabyrinthÂ	467234	Spanish	8.2

124	The Secret in Their EyesÂ		131831	Spanish	8.2
132	AkiraÂ		106160	Japanese	8.1
133	Amores PerrosÂ		173551	Spanish	8.1
141	Elite SquadÂ		81644	Portuguese	8.1
163	SolarisÂ		54057	Russian	8.1
166	Tae Guk Gi: The Brotherhood of WarÂ		31943	Korean	8.1
170	The CelebrationÂ		65951	Danish	8.1
178	The Sea InsideÂ		64556	Spanish	8.1
185	A Fistful of DollarsÂ		147566	Italian	8
197	Central StationÂ		28951	Portuguese	8
215	My Name Is KhanÂ		69759	Hindi	8
219	PersepolisÂ		70194	French	8
230	The Diving Bell and the ButterflyÂ		89906	French	8
239	The ReturnÂ		31589	Russian	8
244	Waltz with BashirÂ		46107	Hebrew	8
247	4 Months, 3 Weeks and 2 DaysÂ		44763	Romanian	7.9
249	AmourÂ		70382	French	7.9

D. Find the best directors.

The top 10 directors for whom the mean of imdb_score is the highest are here and in case of tie if imdb_score is sorted alphabetically.

Row Labels	Average of imdb_score
Cary Bell	8.7
Mitchell Altieri	8.7
Sadyk Sher-Niya	8.7
Charles Chaplin	8.6
Mike Mayhall	8.6
Damien Chazelle	8.5
Majid Majidi	8.5
Raja Menon	8.5
Ron Fricke	8.5
Sergio Leone	8.5
Grand Total	8.6

E. Find popular genres.

Movie genres are stylistic categories where a particular movie can be placed based on the setting, characters, mood, tone, and theme. A Film's main genre category will be based on where the majority of the content land. In the given dataset the most popular genres are:

Row Labels	Values Average of imdb_score	Average of num_voted_users	Average of profit
Adventure|Animation|Drama|Family|Musical	8.5	644348	377783777
Action|Biography|Drama|History|Thriller|War	7.3	325264	291323553
Family|Sci-Fi	5.7	141026	224610375
Adventure|Drama|Fantasy|Romance	5.0	196930	217148557
Drama|Fantasy|Romance|Thriller	7.0	154199	195631306
Action|Adventure|Fantasy|Sci-Fi	7.0	451464	184602953
Adventure|Animation|Comedy|Drama|Family|Fantasy	8.3	345198	181454367
Action|Adventure|Animation|Family	8.0	479166	169437578
Adventure|Comedy|Family|Mystery|Sci-Fi	7.3	403014	160147615
Animation|Comedy|Family|Sci-Fi	7.3	219739	158459955
Grand Total	**6.8**	**366636**	**199738574**

F. Find the critic-favorite and audience-favorite actors.

I created 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 by using a pivot table along with their average of num_critic_for_reviews and average of num_users_for_review and then I combined these in new column name combined.
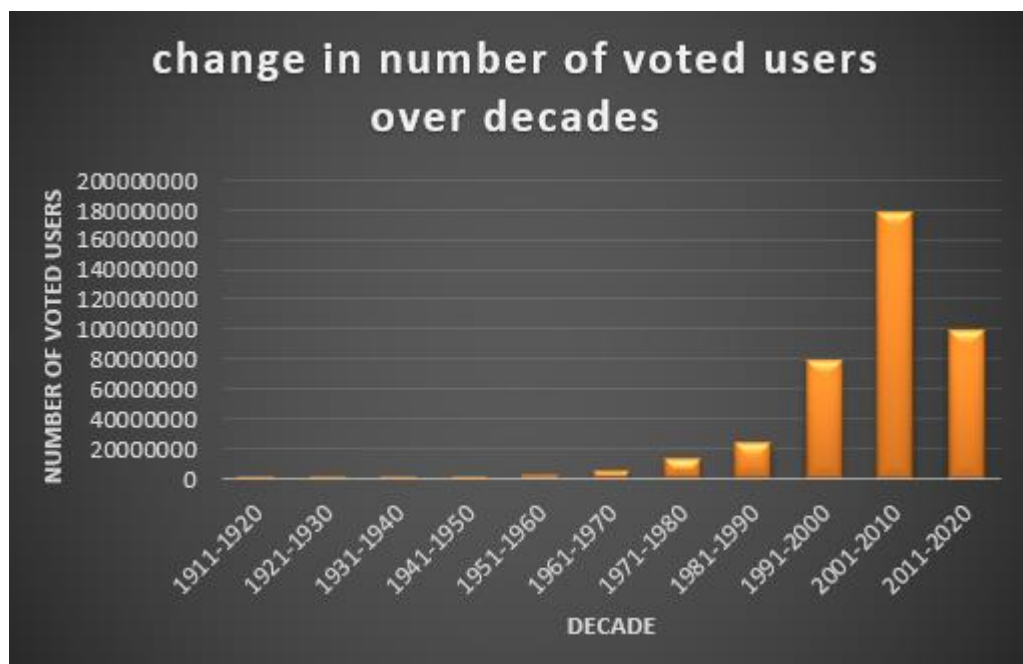
combined	num_critic_for_reviews	num_user_for_reviews
Meryl Streep		
A Prairie Home CompanionÂ	211	280
Florence Foster JenkinsÂ	87	32
Hope SpringsÂ	234	178
It's ComplicatedÂ	187	214
JuliaÂ	38	44
Julie & JuliaÂ	252	277
Lions for LambsÂ	227	298
One True ThingÂ	64	112
Out of AfricaÂ	66	200
The Devil Wears PradaÂ	208	631
The HoursÂ	174	660
The Iron LadyÂ	331	350
The River WildÂ	42	69
Mean	**163.2**	**257.3**
Leonardo DiCaprio		
Blood DiamondÂ	166	657
Body of LiesÂ	238	263

Catch Me If You CanÂ	194	667
Django UnchainedÂ	765	1193
Gangs of New YorkÂ	233	1166
InceptionÂ	642	2803
J. EdgarÂ	392	279
Marvin's RoomÂ	45	71
Revolutionary RoadÂ	323	414
Romeo + JulietÂ	106	506
Shutter IslandÂ	490	964
The AviatorÂ	267	799
The BeachÂ	118	548
The DepartedÂ	352	2054
The Great GatsbyÂ	490	753
The Man in the Iron MaskÂ	83	244
The Quick and the DeadÂ	63	216
The RevenantÂ	556	1188
The Wolf of Wall StreetÂ	606	1138
TitanicÂ	315	2528
Mean	**322.2**	**922.6**

Brad Pitt		
BabelÂ	285	908
By the SeaÂ	131	61
Fight ClubÂ	315	2968
FuryÂ	406	701
Interview with the Vampire: The Vampire ChroniclesÂ	120	406
Johnny SuedeÂ	10	24
Killing Them SoftlyÂ	414	369
Mr. & Mrs. SmithÂ	233	798
Ocean's ElevenÂ	186	845
Ocean's TwelveÂ	198	627
Seven Years in TibetÂ	76	119
Sinbad: Legend of the Seven SeasÂ	98	91
Spy GameÂ	142	361
The Assassination of Jesse James by the Coward Robert FordÂ	273	415
The Curious Case of Benjamin ButtonÂ	362	822
The Tree of LifeÂ	584	975
TroyÂ	220	1694
True RomanceÂ	122	460
Mean	**231.9**	**702.4**

Then I observed the change in number of voted users over decades using a bar chart.

Row Labels	Sum of num_voted_users
1911-1920	10723
1921-1930	132420
1931-1940	1161744
1941-1950	962634
1951-1960	2152978
1961-1970	4664167
1971-1980	13117909
1981-1990	23332522
1991-2000	79208555
2001-2010	178783140
2011-2020	99458013
Grand Total	402984805

change in number of voted users over decades

Here we can clearly see that in 2001-2010 there are a maximum number of voted users.

RESULT

How this project helps me: This project helps me to understand the importance of data analysis. Through this project I am able to understand IMDb data analysis and how the companies use this data as a secret weapon. With an informed and proactive approach, they can leverage insights to make data-backed decisions that optimize their strategy and boost ROI.

Challenges that I faced in this project: Personally I feel that it was easy to solve the questions but the IMDb dataset is relatively huge and has lots of missing values. To find the missing values using Excel and then replacing the null with the right values either median or mode is a challenge.

Conclusion: Living in a world where there are sophisticated platforms to make everyday work simpler, we need to take full advantage and ms excel is the perfect example of this by knowing how to interpret in real time can help us to find out what performs best. Whenever utilized correctly data analytics achieved a significant positive effect on our general public.