

# IMDb Movie Analysis by Excel

## Project description :

Here we are having IMDb dataset, contains movie names, director names, actors, movie ratings, year of movie released, number of votes ,critic reviews. On the basis of this data we have to analyze top 250 movies, popular genres, top directors, distribution over year.

## Approach:

- First understand the data as well as project description carefully.
- Familiarized with dataset ,what columns actually important , columns are more highlighted like movies .
- Cleaned the data , removed unwanted columns, blank rows.
- Example, color of the movie does not give more insight about movie we can get this information by year too, so no requirement of such columns.

## Tech- Stack:

Microsoft Excel 365 have used.

## Insights:

- a) **Cleaning the data :** I have cleaned the data and removed duplicates also. I have used formattings for understanding each row and filtering too.

	A	B	C	D	E	F	G	H	I
1	movie_title	director_name	genres	actor_1_name	num_critic_for_rev	num_voted_users	num_user_for_reviews	gross	budget
2	Avatar	James Cameron	Action Adv CCH Pounder		723	886204	3054	760505847	23
3	Jurassic World	Colin Trevorrow	Action Adv Bryce Dallas Howard		644	418214	1290	652177271	15
4	Titanic	James Cameron	Drama Ror Leonardo DiCaprio		315	793059	2528	658672302	20
5	Star Wars: Episode IV - A New Hope	George Lucas	Action Adv Harrison Ford		282	911097	1470	460935665	1
6	E.T. the Extra-Terrestrial	Steven Spielberg	Family Sci-Fi Henry Thomas		215	281842	515	434949459	1
7	The Lion King	Roger Allers	Adventure Matthew Broderick		186	644348	656	422783777	4
8	Star Wars: Episode I - The Phantom Menace	George Lucas	Action Adv Natalie Portman		320	534658	3597	474544677	11
9	The Dark Knight	Christopher Nolan	Action Crim Christian Bale		645	1676169	4667	533316061	18
10	The Hunger Games	Gary Ross	Adventure Jennifer Lawrence		673	701607	1959	407999255	7
11	Deadpool	Tim Miller	Action Adv Ryan Reynolds		579	479047	1058	363024263	5
12	The Hunger Games: Catching Fire	Francis Lawrence	Adventure Jennifer Lawrence		502	498397	706	424645577	13
13	Jurassic Park	Steven Spielberg	Adventure Wayne Knight		308	613473	895	356784000	6
14	Despicable Me 2	Pierre Coffin	Animation Steve Carell		306	286877	284	368049635	7
15	American Sniper	Clint Eastwood	Action Biog Bradley Cooper		490	325264	916	350123553	5
16	Finding Nemo	Andrew Stanton	Adventure Alexander Gould		301	692482	866	380838870	6
17	Shrek 2	Andrew Adamson	Adventure Rupert Everett		205	314630	483	436471036	15
18	The Lord of the Rings: The Return of the King	Peter Jackson	Action Adv Orlando Bloom		328	1215718	3189	377019252	9
19	Star Wars: Episode VI - Return of the Jedi	Richard Marquand	Action Adv Harrison Ford		197	681857	647	309125409	3
20	Forrest Gump	Robert Zemeckis	Comedy Dr Tom Hanks		149	1251222	1398	329691196	5
21	Star Wars: Episode V - The Empire Strikes Back	Irvin Kershner	Action Adv Harrison Ford		223	837759	900	290158751	1

	actor_1_name	num_critic_for_revie	num_voted_users	num_user_for_reviews	gross	budget	Profit	language	country	content
2	CCH Pounder	723	886204	3054	760505847	237000000	523505847	English	USA	PG-13
3	Bryce Dallas Howard	644	418214	1290	652177271	150000000	502177271	English	USA	PG-13
4	Leonardo DiCaprio	315	793059	2528	658672302	200000000	458672302	English	USA	PG-13
5	Harrison Ford	282	911097	1470	460935665	110000000	449935665	English	USA	PG
6	Henry Thomas	215	281842	515	434949459	105000000	424449459	English	USA	PG
7	Matthew Broderick	186	644348	656	422783777	450000000	377783777	English	USA	G
8	Natalie Portman	320	534658	3597	474544677	115000000	359544677	English	USA	PG
9	Christian Bale	645	1676169	4667	533316061	185000000	348316061	English	USA	PG-13
10	Jennifer Lawrence	673	701607	1959	407999255	780000000	329999255	English	USA	PG-13
11	Ryan Reynolds	579	479047	1058	363024263	580000000	305024263	English	USA	R
12	Jennifer Lawrence	502	498397	706	424645577	130000000	294645577	English	USA	PG-13
13	Wayne Knight	308	613473	895	356784000	630000000	293784000	English	USA	PG-13
14	Steve Carell	306	286877	284	368049635	760000000	292049635	English	USA	PG
15	Bradley Cooper	490	325264	916	350123553	588000000	291323553	English	USA	R
16	Alexander Gould	301	692482	866	380838870	940000000	286838870	English	USA	G
17	Rupert Everett	205	314630	483	436471036	150000000	286471036	English	USA	PG
18	Orlando Bloom	328	1215718	3189	377019252	940000000	283019252	English	USA	PG-13
19	Harrison Ford	197	681857	647	309125409	325000000	276625409	English	USA	PG
20	Tom Hanks	149	1251222	1398	329691196	550000000	274691196	English	USA	PG-13
21	Harrison Ford	223	837759	900	290158751	180000000	272158751	English	USA	PG
22	Macaulay Culkin	102	311075	305	285761243	180000000	267761243	English	USA	PG

## b) Movies with highest profit:

so here profit column have added to table by subtracting budget from gross. Negative profit shows loss.

Book1 - Excel										
File Home Insert Page Layout Formulas Data Review View Help										
Clipboard Font Alignment Number Styles Cells Editing										
Q4 =RANK(F4,\$F\$3:\$F\$316,0)										
	director_name	genres	actor_1_name	num_critic_for_rev	num_voted_users	num_user_for_reviews	gross	budget	Profit	lang
1	Frank Darabont	Crime Drama	Morgan Freeman	199	1689764	4144	28341469	25000000	3341469	
4	Francis Ford Coppola	Crime Drama	Al Pacino	208	1155770	2238	134821952	60000000	128821952	
6	Christopher Nolan	Crime Drama	Christian Bale	645	1676169	4667	533316061	185000000	348316061	
7	Francis Ford Coppola	Crime Drama	Robert De Niro	149	790926	650	57300000	13000000	44300000	
8		Crime Drama Thriller	Kirsten Dunst	54	170055	173			0	
9	Peter Jackson	Adventure Drama	Orlando Bloom	328	1215718	3189	377019252	94000000	283019252	
10	Quentin Tarantino	Crime Drama	Bruce Willis	215	1324680	2195	107930000	80000000	99930000	
11	Steven Spielberg	Adventure Drama	Liam Neeson	174	865020	1273	96067179	22000000	74067179	
12	Sergio Leone	Western	Clint Eastwood	181	503509	780	6100000	1200000	4900000	
13	Sidney Lumet	Crime Drama	Jack Warden	177	447785	888		350000	-350000	
14	Robert Zemeckis	Comedy Drama	Tom Hanks	149	1251222	1398	329691196	55000000	274691196	
15	Irvin Kershner	Adventure Fantasy	Harrison Ford	223	837759	900	290158751	180000000	272158751	
16	Peter Jackson	Adventure Drama	Christopher Lee	297	1238746	5060	313837577	93000000	220837577	
17	Christopher Nolan	Adventure Sci-Fi	Leonardo DiCaprio	642	1468200	2803	292568851	160000000	132568851	
18		Action Adventure Crime Drama	Elden Henson	95	213483	394			0	
19		Comedy	Rob McElhenney	79	133415	151			0	
20	David Fincher	Drama	Brad Pitt	315	1347461	2968	37023395	63000000	-25976605	
21	George Lucas	Adventure Fantasy	Harrison Ford	282	911097	1470	460935665	110000000	449935665	
22	Peter Jackson	Adventure Drama	Christopher Lee	294	1100446	2417	340478898	94000000	246478898	
23	Lana Wachowski	Action Sci-Fi	Keanu Reeves	313	1217752	3646	171383253	63000000	108383253	

movie_title	Profit
Star Wars: Episode IV - A New Hope	449935665
The Lion King	377783777
The Dark Knight	348316061
Deadpool	305024263
Jurassic Park	293784000
Finding Nemo	286838870
The Lord of the Rings: The Return of the King	283019252
Star Wars: Episode VI - Return of the Jedi	276625409
Forrest Gump	274691196
Star Wars: Episode V - The Empire Strikes Back	272158751
The Sixth Sense	253501675
Jaws	252000000

- c) **Top 250 Movies:** using IMDb score I have analyzed top 250 movies having high score or ratings. After I found rank according to top 250 movies. And the 1<sup>st</sup> rank is of The Shawshank Redemption.

The screenshot shows an Excel spreadsheet with the following data columns: budget, Profit, language, country, content\_rating, title\_year, imdb\_score, IMDb\_Top\_50, and RANK. The formula bar shows the formula for the RANK column: `=RANK(F4,$F$3:$F$316,0)`. The data is sorted by IMDb score in descending order, with The Shawshank Redemption at the top (rank 1).

	budget	Profit	language	country	content_rating	title_year	imdb_score	IMDb_Top_50	RANK
3	25000000	3341469	English	USA	R	1994	9.3	The Shawshank Redemption	1
4	6000000	128821952	English	USA	R	1972	9.2	The Godfather	10
6	185000000	348316061	English	USA	PG-13	2008	9	The Dark Knight	2
7	13000000	44300000	English	USA	R	1974	9	The Godfather: Part II	30
8	0	0	English	USA	TV-MA		9	Fargo	176
9	94000000	283019252	English	USA	PG-13	2003	8.9	The Lord of the Rings: The Return of the King	9
10	8000000	99930000	English	USA	R	1994	8.9	Pulp Fiction	5
11	22000000	74067179	English	USA	R	1993	8.9	Schindler's List	23
12	1200000	4900000	Italian	Italy	Approved	1966	8.9	The Good, the Bad and the Ugly	84
13	3500000	-3500000	English	USA	Not Rated	1957	8.9	12 Angry Men	100
14	55000000	274691196	English	USA	PG-13	1994	8.8	Forrest Gump	6
15	18000000	272158751	English	USA	PG	1980	8.8	Star Wars: Episode V - The Empire Strikes Back	26
16	93000000	220837577	English	New Zealand	PG-13	2001	8.8	The Lord of the Rings: The Fellowship of the Ring	7
17	160000000	132568851	English	USA	PG-13	2010	8.8	Inception	3
18	0	0	English	USA	TV-MA		8.8	Daredevil	163
19	0	0	English	USA	TV-MA		8.8	It's Always Sunny in Philadelphia	196
20	63000000	-25976605	English	USA	R	1999	8.8	Fight Club	4
21	11000000	449935665	English	USA	PG	1977	8.7	Star Wars: Episode IV - A New Hope	18
22	94000000	246478898	English	USA	PG-13	2002	8.7	The Lord of the Rings: The Two Towers	12
23	63000000	108383253	English	USA	R	1999	8.7	The Matrix	8

Again I analyzed foreign movies ( no English) as follows.

	language	country	content	title	year	imdb_score	IMDb_Top_250	RANK	Top_Foreign_Language_Film
12	Italian	Italy	Approved		1966	8.9	The Good, the Bad and the Ugly	84	Italian
26	Portuguese	Brazil	R		2002	8.7	City of God	75	Portuguese
33	Japanese	Japan	Unrated		1954	8.7	Seven Samurai	156	Japanese
48	Japanese	Japan	PG		2001	8.6	Spirited Away	107	Japanese
64	German	Germany	R		2006	8.5	The Lives of Others	148	German
65	Persian	Iran	PG		1997	8.5	Children of Heaven	256	Iran
72	Hindi	India			2016	8.5	Airliift	251	India
78	Persian	Iran	PG-13		2011	8.4	A Separation	187	Iran
80	Hindi	India	Not Rated		2006	8.4	Rang De Basanti	222	India
89	Korean	South Korea	R		2003	8.4	Oldboy	121	South Korea
93	German	West Germany	R		1981	8.4	Das Boot	178	West Germany
96	Telugu	India			2015	8.4	Baahubali: The Beginning	234	India
100	French	France	R		2001	8.4	Amélie	74	France
101	Japanese	Japan	PG-13		1997	8.4	Princess Mononoke	157	Japan
135	Danish	Denmark	R		2012	8.3	The Hunt	175	Denmark
137	German	Germany	Not Rated		1927	8.3	Metropolis	203	Germany
138	German	Germany	R		2004	8.3	Downfall	150	Germany
148	Spanish	Spain	R		2006	8.2	Pan's Labyrinth	97	Spain
149	Spanish	Argentina	R		2009	8.2	The Secret in Their Eyes	197	Argentina
158	French	Canada	R		2010	8.2	Incendies	214	Canada

#### d) Best directors

Here is the list of top 10 directors and their movies for which they got high ratings.

According to average ratings here I got the top 10 directors leading Carry Bell

	Top_10_Best_Directors	Average of imdb_score
4	Cary Bell	8.7
5	Charles Chaplin	8.6
6	Damien Chazelle	8.5
7	John Blanchard	9.5
8	Majid Majidi	8.5
9	Mike Mayhall	8.6
10	Mitchell Altieri	8.7
11	Raja Menon	8.5
12	Ron Fricke	8.5
13	Sadyk Sher-Niyaz	8.7
14	Grand Total	8.68

#### e) Popular genres:

Popular genres by using movies list and imdb ratings

	A	B	C
1	imdb_score	(Multiple Items)	
2			
3	popular_genres	Average of imdb_score	
4	Action Adventure Drama Fantasy	8.8	
5	Western	8.75	
6	Action Biography Drama History	8.7	
7	Comedy Horror Thriller	8.7	
8	Family	8.6	
9	Drama Family Fantasy Romance	8.6	
10	Action	8.6	
11	Crime Drama Horror Thriller	8.6	
12	Action Drama History Thriller War	8.5	
13	Horror Sci-Fi	8.5	
14	Crime Drama Fantasy Mystery	8.5	
15	Adventure Animation Drama Family Musical	8.5	
16			
17			
18			

**f) Charts: critic favorite actors and audience favorite:**

I grouped three actors by their movies in which they were lead actors.

Leonardo DiCaprio has maximum number of movies.

Trinity\_project5\_IMDB - Excel

File Home Insert Page Layout Formulas Data Review View Help Tell me what you want to do

Clipboard Font Alignment Number Styles Cells Editing

Calibri 11 A A General Conditional Formatting Format as Table Cell Styles Insert Delete Format Sort & Filter Find & Select

H363

	C	D	E	F	G	H	I	J	K
1		Brad Pitt	Leonardo DiCaprio	Meryl Streep					
28		The Curious Case of Benjamin Button	Titanic	It's Complicated					
52		Troy	The Great Gatsby	The River Wild					
99		Ocean's Twelve	Inception	Julie & Julia					
103		Mr. & Mrs. Smith	The Revenant	The Devil Wears Prada					
149		Spy Game	The Aviator	Lions for Lambs					
81		Ocean's Eleven	Django Unchained	Out of Africa					
256		Fury	Blood Diamond	Hope Springs					
257		Seven Years in Tibet	The Wolf of Wall Street	One True Thing					
259		Fight Club	Gangs of New York	Florence Foster Jenkins					
298		Sinbad: Legend of the Seven Seas	The Departed	The Hours					
309		Interview with the Vampire: The Vampire Chronicles	Shutter Island	The Iron Lady					
310		The Tree of Life	Body of Lies	A Prairie Home Companion					
328		The Assassination of Jesse James by the Coward Robert Ford	Catch Me If You Can	Julia					
363		Babel	The Beach						
384		By the Sea	Revolutionary Road						
402		Killing Them Softly	The Man in the Iron Mask						
412		True Romance	J. Edgar						
454		Johnny Suede	The Quick and the Dead						
472			Marvin's Room						
613			Romeo + Juliet						
643			The Great Gatsby						
685									

Sheet1 Sheet2 Sheet5 Sheet6 Sheet3 Sheet4

Average number of critic for review of Phaldut Sharma actor along with that, average number of users for review of Heather Donahue.

The image displays two screenshots of an Excel spreadsheet titled 'Trainity\_project5\_IMDB - Excel'. The spreadsheet is organized into columns: D (Row Labels), E (Average of num\_user\_for\_reviews), and F (Average of num\_critic\_for\_reviews). The data is presented in a table format with rows numbered 1 through 21.

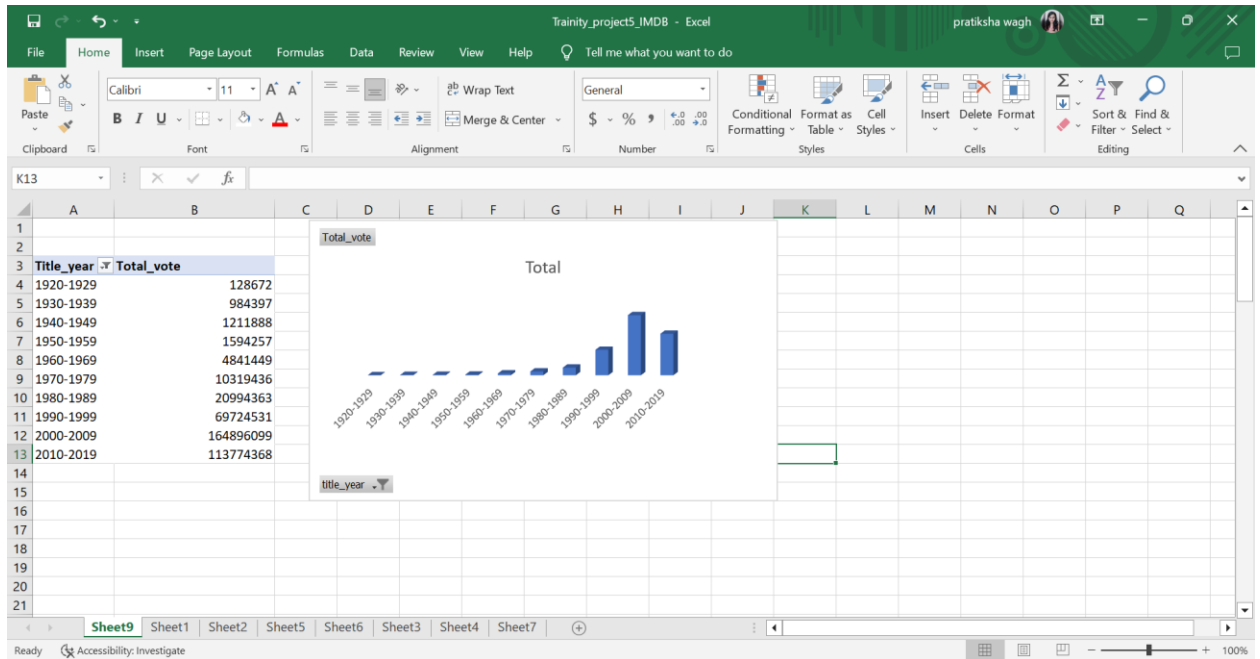
**Top Screenshot Data:**

Row Labels	Average of num_user_for_reviews	Average of num_critic_for_reviews
Phaldut Sharma	1885	738
Peter Capaldi	995	654
Craig Stark	1018	596
BÃ©rÃ©nice Bejo	583	576
Suraj Sharma	755	552
Ellar Coltrane	836	548
Mike Howard	405	546
Lou Taylor Pucci	789	543
Maika Monroe	631	533
Tim Holmes	511	525
Albert Finney	1017.333333	510
Elina Alminas	611	489
Kurt Fuller	509	487
Iko Uwais	316	481
QuvenzhanÃ© Wallis	392.333333	478.666667
Edgar Arreola	461	478
Sharlto Copley	1262	472
Cory Hardict	326	452
Matt Frewer	1229	451
Elizabeth McGovern	801	447

**Bottom Screenshot Data:**

Row Labels	Average of num_user_for_reviews	Average of num_critic_for_reviews
Heather Donahue	3400	360
Christo Jivkov	2814	406
Steve Bastoni	2789	275
Phaldut Sharma	1885	738
Keir Dullea	1736	285
Chen Chang	1641	287
Nick Stahl	1562	218
Kevin Rankin	1445	267
Noah Huntley	1441	224
Osama bin Laden	1416	288
Seychelle Gabriel	1382	280
Mathieu Kassovitz	1314	242
Eva Green	1290	313.5
Essie Davis	1285.5	245
Sharlto Copley	1262	472
Giancarlo Giannini	1243	403
Orlando Bloom	1242.333333	195.666667
Matt Frewer	1229	451
Luenell	1198	343
Micah Sloat	1189	409

I have finally created one chart using pivot table, pivot charts as follows:  
Which shows sum of numbers of votes according to year in which movie has released.



## Results:

I have analyzed the data by using excel, pivoting. Lot of factors affecting the popularity of movies. This project help me to build my analytical skills and gave me more insights in business or real time projects.