

# IMDB Movie Analysis

---

AUGUST 11

---

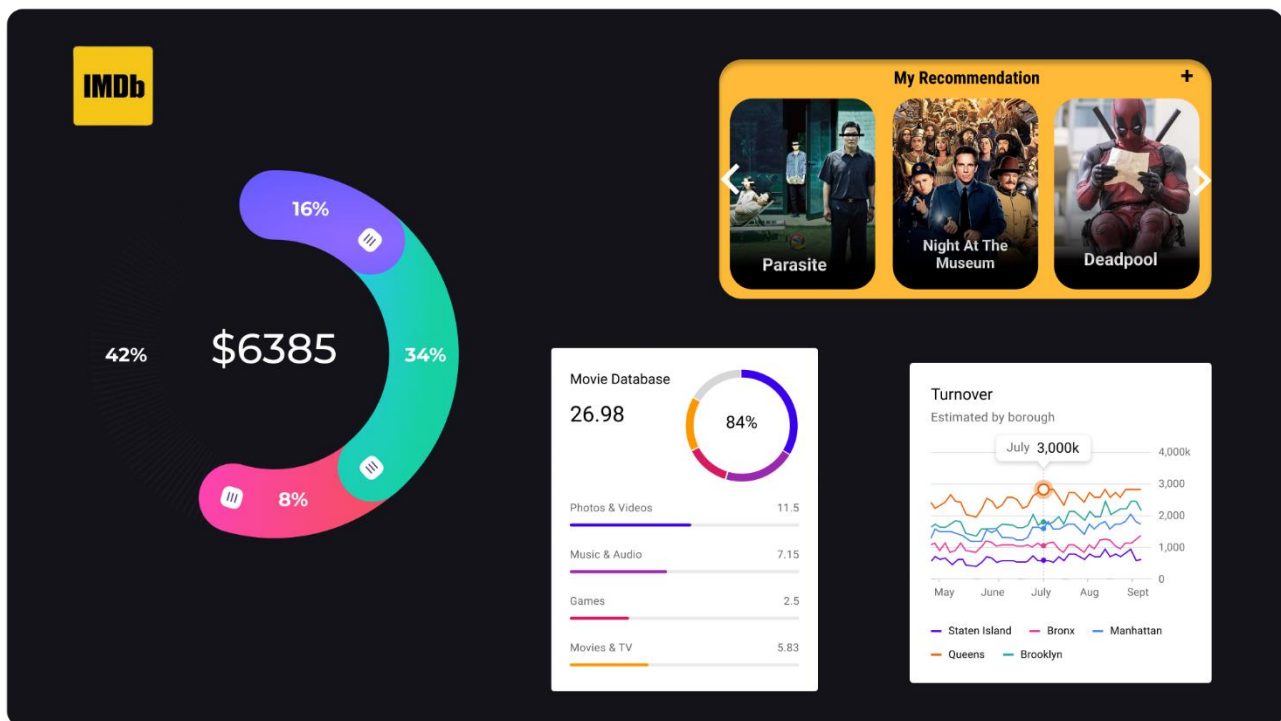
By Aastha Kumar



# Description

I have been provided with a dataset having various columns of different IMDB Movies and am required to Frame the problem. For this task, I will need to define a problem I want to shed some light on.

I can do this by asking 'What?' This is where I frame the problem i.e. What is the problem? Once I have defined a problem I will use my Data Analysis skills to explore the data set and derive insights.



# Approach

First, I downloaded the dataset from Google Sheets onto my personal device for making modifications. I then used my knowledge in statistics and used different formulas in excel to draw necessary conclusions about the company.

# Tech-Stack Used

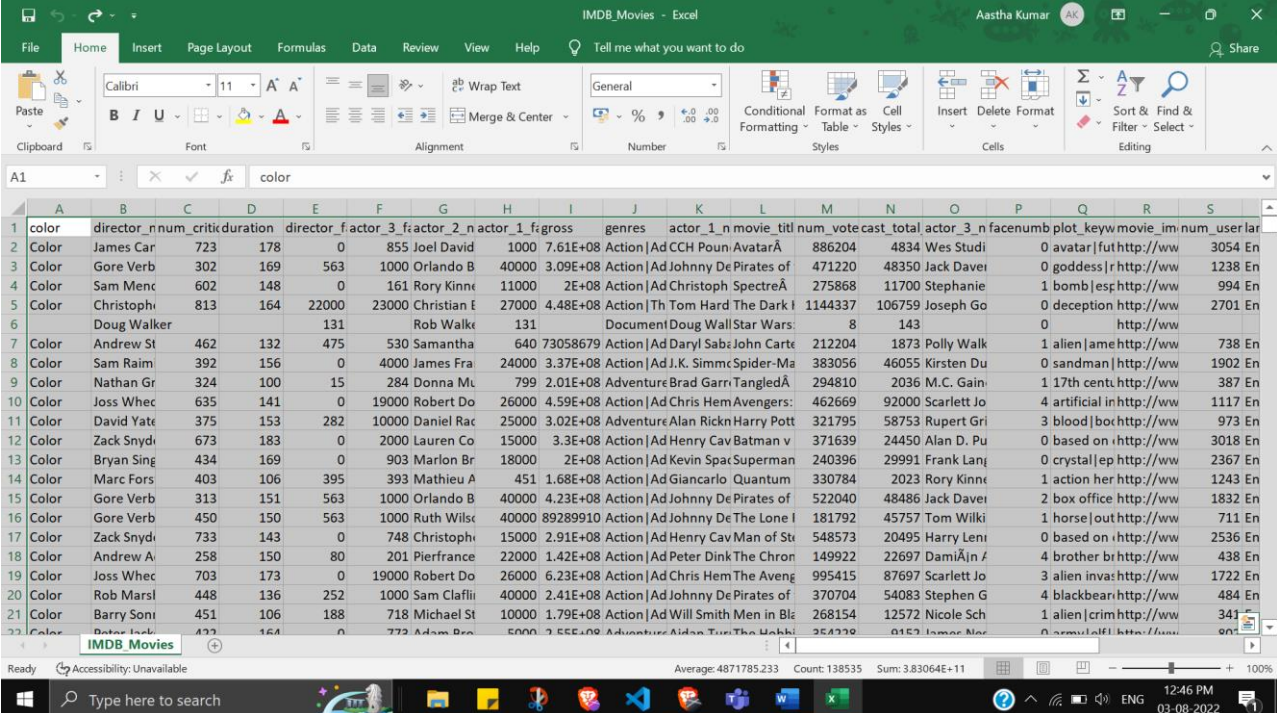
- Google Sheet

## Insights

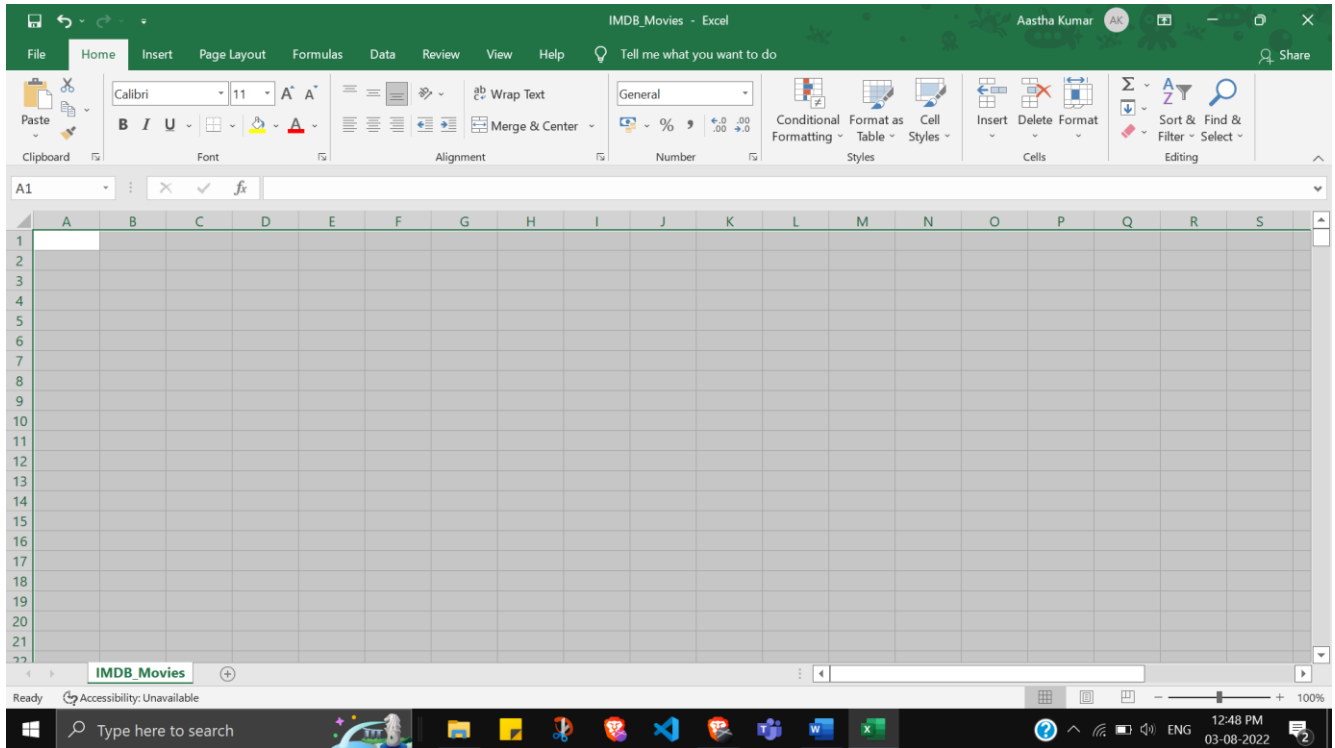
- A. Cleaning the data: This is one of the most important step to perform before moving forward with the analysis. Use your knowledge learned till now to do this. (Dropping columns, removing null values, etc.)

**My Task:** Clean the data

To delete all rows and columns of the spreadsheet, I used the keyboard shortcut ctrl+A to select all the values at once and then hit the delete button.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	color	director_n	critic	duration	director_f	actor_3_f	actor_2_n	actor_1_f	gross	genres	actor_1_n	movie_titl	num_vote	cast_total	actor_3_n	facenumb	plot_keyw	movie_im	num_user lar
2	Color	James Car	723	178	0	855	Joel David	1000	7.61E+08	Action	Ad CCH Poun	Avatar	886204	4834	Wes Studi	0	avatar fut	http://www	3054 En
3	Color	Gore Verb	302	169	563	1000	Orlando B	40000	3.09E+08	Action	Ad Johnny De	Pirates of	471220	48350	Jack Daver	0	goddess r	http://www	1238 En
4	Color	Sam Menc	602	148	0	161	Rory Kinn	11000	2E+08	Action	Ad Christoph	Spectre	275868	11700	Stephanie	1	bomb esq	http://www	994 En
5	Color	Christoph	813	164	22000	23000	Christian E	27000	4.48E+08	Action	Th Tom Hard	The Dark	1144337	106759	Joseph Go	0	deception	http://www	2701 En
6		Doug Walker			131		Rob Walk	131		Document	Doug Wall	Star Wars	8	143		0		http://www	
7	Color	Andrew St	462	132	475	530	Samantha	640	73058679	Action	Ad Daryl Sab	John Carte	212204	1873	Polly Walk	1	alien ame	http://www	738 En
8	Color	Sam Raim	392	156	0	4000	James Fra	24000	3.37E+08	Action	Ad J.K. Simm	Spider-Ma	383056	46055	Kirsten Du	0	sandman	http://www	1902 En
9	Color	Nathan Gr	324	100	15	284	Donna Mi	799	2.01E+08	Adventure	Brad Garr	Tangled	294810	2036	M.C. Gain	1	17th cent	http://www	387 En
10	Color	Joss Whec	635	141	0	19000	Robert Do	26000	4.59E+08	Action	Ad Chris Hem	Avengers:	462669	92000	Scarlett Jo	4	artificial	http://www	1117 En
11	Color	David Yate	375	153	282	10000	Daniel Rac	25000	3.02E+08	Adventure	Alan Rickn	Harry Pott	321795	58753	Rupert Gri	3	blood bo	http://www	973 En
12	Color	Zack Snyder	673	183	0	2000	Lauren Co	15000	3.3E+08	Action	Ad Henry Cav	Batman v	371639	24450	Alan D. Pu	0	based on	http://www	3018 En
13	Color	Bryan Sing	434	169	0	903	Marlon Br	18000	2E+08	Action	Ad Kevin Spa	Superman	240396	29991	Frank Lan	0	crystal	http://www	2367 En
14	Color	Marc Fors	403	106	395	393	Mathieu A	451	1.68E+08	Action	Ad Giancarlo	Quantum	330784	2023	Rory Kinn	1	action her	http://www	1243 En
15	Color	Gore Verb	313	151	563	1000	Orlando B	40000	4.23E+08	Action	Ad Johnny De	Pirates of	522040	48486	Jack Daver	2	box office	http://www	1832 En
16	Color	Gore Verb	450	150	563	1000	Ruth Wils	40000	89289910	Action	Ad Johnny De	The Lone I	181792	45757	Tom Wilki	1	horse out	http://www	711 En
17	Color	Zack Snyder	733	143	0	748	Christoph	15000	2.91E+08	Action	Ad Henry Cav	Man of Sti	548573	20495	Harry Leni	0	based on	http://www	2536 En
18	Color	Andrew A	258	150	80	201	Pierfrance	22000	1.42E+08	Action	Ad Peter Dink	The Chron	149922	22697	Dami	4	brother b	http://www	438 En
19	Color	Joss Whec	703	173	0	19000	Robert Do	26000	6.23E+08	Action	Ad Chris Hem	The Aveng	995415	87697	Scarlett Jo	3	alien inva	http://www	1722 En
20	Color	Rob Marsl	448	136	252	1000	Sam Claflin	40000	2.41E+08	Action	Ad Johnny De	Pirates of	370704	54083	Stephen G	4	blackbear	http://www	484 En
21	Color	Barry Soni	451	106	188	718	Michael St	10000	1.79E+08	Action	Ad Will Smith	Men in Bla	268154	12572	Nicole Sch	1	alien crim	http://www	341 En
22	Color	Adam B	423	164	0	773	Adam B	5000	2.55E+08	Adventure	Adam B	The Hobbit	254730	0153	James Mac	0	comul	http://www	801 En



- 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.

My Task: Find the movies with the highest profit.

	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC
1	actor_1_n	movie_title	num_votes	cast_total	actor_3_n	facenumb	plot_keyw	movie_im	num_user	language	country	content_r	budget	title_year	actor_2_f	imdb_score	aspect_ra	movie_fac	profit
2	CCH Poun	Avatar	886204	4834	Wes Studi	0	avatar fut	http://ww	3054	English	USA	PG-13	2.37E+08	2009	936	7.9	1.78	33000	=I2-W2
3	Johnny De	Director of	471220	48350	Jack Daver	0	goddess r	http://ww	1238	English	USA	PG-13	3E+08	2007	5000	7.1	2.35	0	
4	275868		11700	Stephanie	1	bomb esp	http://ww	994	English	UK	PG-13	2.45E+08	2015	393	6.8	2.35	85000	-44925825	
5	1144337		106759	Joseph Go	0	deception	http://ww	2701	English	USA	PG-13	2.5E+08	2012	23000	8.5	2.35	164000	198130642	
6	8	143			0		http://ww							12	7.1		0	0	
7	212204		1873	Polly Walk	1	alien ame	http://ww	738	English	USA	PG-13	2.64E+08	2012	632	6.6	2.35	24000	-1.91E+08	
8	383056		46055	Kirsten Du	0	sandman	http://ww	1902	English	USA	PG-13	2.58E+08	2007	11000	6.2	2.35	0	78530303	
9	294810		2036	M.C. Gain	1	17th cent	http://ww	387	English	USA	PG	2.6E+08	2010	553	7.8	1.85	29000	-59192738	
10	462669		92000	Scarlett Jo	4	artificial	http://ww	1117	English	USA	PG-13	2.5E+08	2015	21000	7.5	2.35	118000	208991599	
11	321795		58753	Rupert Gri	3	blood bo	http://ww	973	English	UK	PG	2.5E+08	2009	11000	7.5	2.35	10000	51956980	
12	371639		24450	Alan D. Pu	0	based on	http://ww	3018	English	USA	PG-13	2.5E+08	2016	4000	6.9	2.35	197000	80249062	
13	240396		29991	Frank Lan	0	crystal	ep http://ww	2367	English	USA	PG-13	2.09E+08	2006	10000	6.1	2.35	0	-8930592	
14	330784		2023	Rory Kinn	1	action her	http://ww	1243	English	UK	PG-13	2E+08	2008	412	6.7	2.35	0	-31631573	
15	522040		48486	Jack Daver	2	box office	http://ww	1832	English	USA	PG-13	2.25E+08	2006	5000	7.3	2.35	5000	198032628	
16	181792		45757	Tom Wilki	1	horse out	http://ww	711	English	USA	PG-13	2.15E+08	2013	2000	6.5	2.35	48000	-1.26E+08	
17	548573		20495	Harry Leni	0	based on	http://ww	2536	English	USA	PG-13	2.25E+08	2013	3000	7.2	2.35	118000	66021565	
18	149922		22697	Damián	4	brother br	http://ww	438	English	USA	PG	2.25E+08	2008	216	6.6	2.35	0	-83385977	
19	995415		87697	Scarlett Jo	3	alien invas	http://ww	1722	English	USA	PG-13	2.2E+08	2012	21000	8.1	1.85	123000	403279547	
20	370704		54083	Stephen G	4	blackbear	http://ww	484	English	USA	PG-13	2.5E+08	2011	11000	6.7	2.35	58000	-8936125	
21	268154		12572	Nicole Sch	1	alien crim	http://ww	341	English	USA	PG-13	2.25E+08	2012	816	6.8	1.85	40000	-45979146	
22	264279		8157	James Mac	0	army elf	http://ww	902	English	New Zeala	PG-13	2.5E+08	2014	873	7.5	2.25	65000	5108270	

- 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.

### My Task: Find IMDB Top 250

Select the entire spreadsheet by hitting ctrl+A

	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD
1	movie_title	num_vote	cast_total	actor_3_n	facenum	plot_keyw	movie_im	num_user	language	country	content_r	budget	title_year	actor_2_f	imdb_score	aspect_ra	movie_fac	profit	IMDb_Top_2
2	Avatar	886204	4834	Wes Studi	0	avatar fut	http://www	3054	English	USA	PG-13	2.37E+08	2009	936	7.9	1.78	33000	5.24E+08	
3	Pirates of the Caribbean: The Curse of the Black Pearl	471220	48350	Jack Davenport	0	goddess r	http://www	1238	English	USA	PG-13	3E+08	2007	5000	7.1	2.35	0	9404152	
4	Spectre	275868	11700	Stephanie	1	bomb esq	http://www	994	English	UK	PG-13	2.45E+08	2015	393	6.8	2.35	85000	-4.5E+07	
5	The Dark Knight	1144337	106759	Joseph Gordon-Levitt	0	deception	http://www	2701	English	USA	PG-13	2.5E+08	2012	23000	8.5	2.35	164000	1.98E+08	
6	Star Wars: The Force Awakens	8	143		0		http://www							12	7.1		0	0	
7	John Carter	212204	1873	Polly Walker	1	alien ame	http://www	738	English	USA	PG-13	2.64E+08	2012	632	6.6	2.35	24000	-1.9E+08	
8	Spider-Man	383056	46055	Kirsten Dunst	0	sandman	http://www	1902	English	USA	PG-13	2.58E+08	2007	11000	6.2	2.35	0	78530303	
9	Tangled	294810	2036	M.C. Gainey	1	17th centu	http://www	387	English	USA	PG	2.6E+08	2010	553	7.8	1.85	29000	-5.9E+07	
10	Avengers: Endgame	462669	92000	Scarlett Johansson	4	artificial	in http://www	1117	English	USA	PG-13	2.5E+08	2015	21000	7.5	2.35	118000	2.09E+08	
11	Harry Potter and the Chamber of Secrets	321795	58753	Rupert Gri	3	blood bo	http://www	973	English	UK	PG	2.5E+08	2009	11000	7.5	2.35	10000	51956980	
12	Batman v	371639	24450	Alan D. Pu	0	based on	(http://www	3018	English	USA	PG-13	2.5E+08	2016	4000	6.9	2.35	197000	80249062	
13	Superman	240396	29991	Frank Lan	0	crystal ep	http://www	2367	English	USA	PG-13	2.09E+08	2006	10000	6.1	2.35	0	-8930592	
14	Quantum of Solace	330784	2023	Rory Kinne	1	action her	http://www	1243	English	UK	PG-13	2E+08	2008	412	6.7	2.35	0	-3.2E+07	
15	Pirates of the Caribbean: On Stranger Tides	522040	48486	Jack Davenport	2	box office	http://www	1832	English	USA	PG-13	2.25E+08	2006	5000	7.3	2.35	5000	1.98E+08	
16	The Lone I	181792	45757	Tom Wilki	1	horse out	http://www	711	English	USA	PG-13	2.15E+08	2013	2000	6.5	2.35	48000	-1.3E+08	
17	Man of Ste	548573	20495	Harry Leni	0	based on	(http://www	2536	English	USA	PG-13	2.25E+08	2013	3000	7.2	2.35	118000	66021565	
18	The Chron	149922	22697	Dami	4	brother br	http://www	438	English	USA	PG	2.25E+08	2008	216	6.6	2.35	0	-8.3E+07	
19	The Aveng	995415	87697	Scarlett Jo	3	alien inva	http://www	1722	English	USA	PG-13	2.2E+08	2012	21000	8.1	1.85	123000	4.03E+08	
20	Pirates of the Caribbean: The Curse of the Black Pearl	370704	54083	Stephen G	4	blackbear	http://www	484	English	USA	PG-13	2.5E+08	2011	11000	6.7	2.35	58000	-8936125	
21	Men in Bla	268154	12572	Nicole Sch	1	alien crim	http://www	341	English	USA	PG-13	2.25E+08	2012	816	6.8	1.85	40000	-4.6E+07	
22	The Backli	354779	8152	James Nes	0	sexual self	https://www	803	English	New Zeal	PG-13	2.5E+08	2014	873	7.5	2.35	65000	5108270	

Under Sort and Filter menu, click the custom sort option

Sort

Column: imdb\_score, Sort On: Cell Values, Order: Largest to Smallest

OK Cancel

Click OK. All the contents of the spreadsheet will be sort according to the highest to lowest imdb\_score. Select the top 250 movies corresponding to the highest imdb\_scores and paste the values in the new column created (IMDb\_Top\_250)

plot_key	movie_im	num_user	language	country	content_r	budget	title_year	actor_2	fi	imdb_sco	aspect	ra	movie_fac	profit	IMDb_Top_250
1	escape fr	http://www.imdb.co	English	Canada	R	25000000	1994	179	9.5	1.33	0	0	0	0	Towering Inferno
2	crime fam	http://www	English	USA	R	6000000	1972	745	9.3	1.85	108000	3341469	1.29E+08	0	The Shawshank Redemption
3	meaning ch	http://www	Polish	Poland	TV-MA			10000	9.2	1.85	43000	1.29E+08	0	0	The Godfather
4	meaning ch	http://www	Polish	Poland	TV-MA			3	9.1	1.33	0	447093	0	0	Dekalog
5	meaning ch	http://www	Polish	Poland	TV-MA			3	9.1	1.33	0	447093	0	0	Dekalog
6	based on	http://www	English	USA	PG-13	1.85E+08	2008	454	9.1	1.33	0	-1.7E+07	0	0	Kickboxer: Vengeance
7	1950s	cor	http://www	English	USA	R	13000000	2016	13000	9	2.35	37000	3.48E+08	0	The Dark Knight
8	anthology	http://www	English	USA	TV-MA			14000	9	1.85	14000	44300000	0	0	The Godfather: Part II
9	anthology	http://www	English	USA	TV-MA			1000	9	1.78	61000	0	0	0	Fargo
10	battle epi	http://www	English	USA	PG-13	94000000	2003	857	8.9	2.35	16000	2.83E+08	0	0	The Lord of the Rings: The Return of the King
11	german g	http://www	English	USA	R	22000000	1993	795	8.9	1.85	41000	74067179	0	0	Schindler's List
12	black com	http://www	English	USA	R	8000000	1994	902	8.9	2.35	45000	99930000	0	0	Pulp Fiction
13	civil war	http://www	Italian	Italy	Approved	1200000	1966	34	8.9	2.35	20000	4900000	0	0	The Good, the Bad and the Ugly
14	courtroom	http://www	English	USA	Not Rated	350000	1957	259	8.9	1.66	40000	-350000	0	0	12 Angry Men
15	ambiguou	http://www	English	USA	PG-13	1.6E+08	2010	27000	8.8	2.35	175000	1.33E+08	0	0	Inception
16	elf hobbit	http://www	English	New Zeala	PG-13	93000000	2001	5000	8.8	2.35	21000	2.21E+08	0	0	The Lord of the Rings: The Fellowship of the Ring
17	corruption	http://www	English	USA	TV-MA			4	8.8	1.6	55000	0	0	0	Daredevil
18	anti estab	http://www	English	USA	R	63000000	1999	783	8.8	2.35	48000	-2.6E+07	0	0	Fight Club
19	amputee	http://www	English	USA	PG-13	55000000	1994	294	8.8	2.35	59000	2.75E+08	0	0	Forrest Gump
20	alcohol b	http://www	English	USA	TV-MA			547	8.8	1.33	0	0	0	0	It's Always
21	duell fam	http://www	English	USA	PG	18000000	1980	504	8.8	2.35	17000	2.75E+08	0	0	Star Wars

---

D. Best Directors: TGroup 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.

My Task: Find the best directors

E. Popular Genres: Perform this step using the knowledge gained while performing previous steps.

My Task: Find popular genres

F. Charts: Create 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. Use only the actor\_1\_name column for extraction. Also, make sure that you use the names 'Meryl Streep', 'Leonardo DiCaprio', and 'Brad Pitt' for the said extraction.

Append the rows of all these columns and store them in a new column named Combined. Group the combined column using the actor\_1\_name column. Find the mean of the num\_critic\_for\_reviews and num\_users\_for\_review and identify the actors which have the highest mean. Observe the change in number of voted users over decades using a bar chart. Create a column called decade which represents the decade to which every movie belongs to. For example, the title\_year year 1923, 1925 should be stored as 1920s. Sort the column based on



---

the column decade, group it by decade and find the sum of users voted in each decade. Store this in a new data frame called df\_by\_decade.

My Task: Find the critic-favorite and audience-favorite actors

## Resources

---



[IMDB Movies](#)

## Result

---

It would have been impossible or very time consuming for an ordinary human to be able to process such large bits of information. A computer can do it within seconds with the right commands. That's why companies like Instagram hire data analysts to control the waves of data they collect every day, makes sense of it, and then draw conclusions or make predictions. This is the process of turning data into insights, and it's how analysts help businesses put all their data to good use.

The more detailed definition you learned earlier is that data analysis is the collection, transformation, and organization of data in order to draw conclusions, make predictions, and drive informed decision-making.

---

Data analytics can help organizations completely rethink something they do or point them in a totally new direction. For example, maybe data leads them to a new product or unique service, or maybe it helps them find a new way to deliver an incredible customer experience.