PROJECT- IMDB MOVIE ANALYSIS

-Final Project-1

Project Description:

This project is about IMDB movie analysis. A dataset containing IMDB movie information is provided along with this project. It contains data such as movies with its corresponding imdb ratings, directors, actors, genres, language, ratings and reviews by critics and audience, releasing year, budget, gross collection, etc.
Through this project, from the dataset provided, I will be finding out the answers for the questions in tasks asked. I will be cleaning the data, doing analysis and finding out the factors that influence the success of a movie on IMDB, movies with high IMDB ratings, etc which will be helpful for the movie makers to understand what makes a movie successful and thus will be useful in making informed decisions in their future projects.

Approach:

I tried to understand the purpose of this project. I read each and every tasks provided by the team and understood what data they needed from it. Microsoft Excel is used to perform the analysis. A data record containing IMDB movie information is provided to perform the analysis and I was asked to answer certain questions to collect insights from the results got. I watched and understood tutorial videos to gain knowledge about statistics, excel, shortcuts, pivot tables, various functions, formulas and filters. I could also learn advanced excel. It helped me to practically apply my knowledge to complete the given tasks. I also used Google and YouTube videos to clear my doubts.

I downloaded the IMDB movie dataset provided by the team and performed the analysis of the dataset using Excel, after cleaning the data such as making the data into readable form, removing the duplicates, unnecessary columns, blank rows and handling the missing data, converting the data types, etc. From the dataset given, I uniquely selected the columns required for solving each tasks. I solved and specified each tasks in separate sheets. I used my knowledge in statistics and different formulas and functions, pivot tables in excel and found answers to the questions asked to draw necessary conclusions. The screenshots of data used to find answers and output data are attached in the output section of each tasks.

Tech-Stack Used:

I used Microsoft Excel to perform the analysis and solving the tasks given. It helped to do the analysis effectively and easier.

Insights:

While doing the IMDB movie analysis project using Microsoft excel, I performed some statistical functions, several formulas, shortcuts, filters, functions, pivot tables, charts, etc. to solve the tasks provided and gained many insights and knowledge, and provided a detailed report. I tried to understand the relationships between different variables, correlation between movie ratings and other factors like genre, language, duration, director, budget, etc., and other relevant factors.

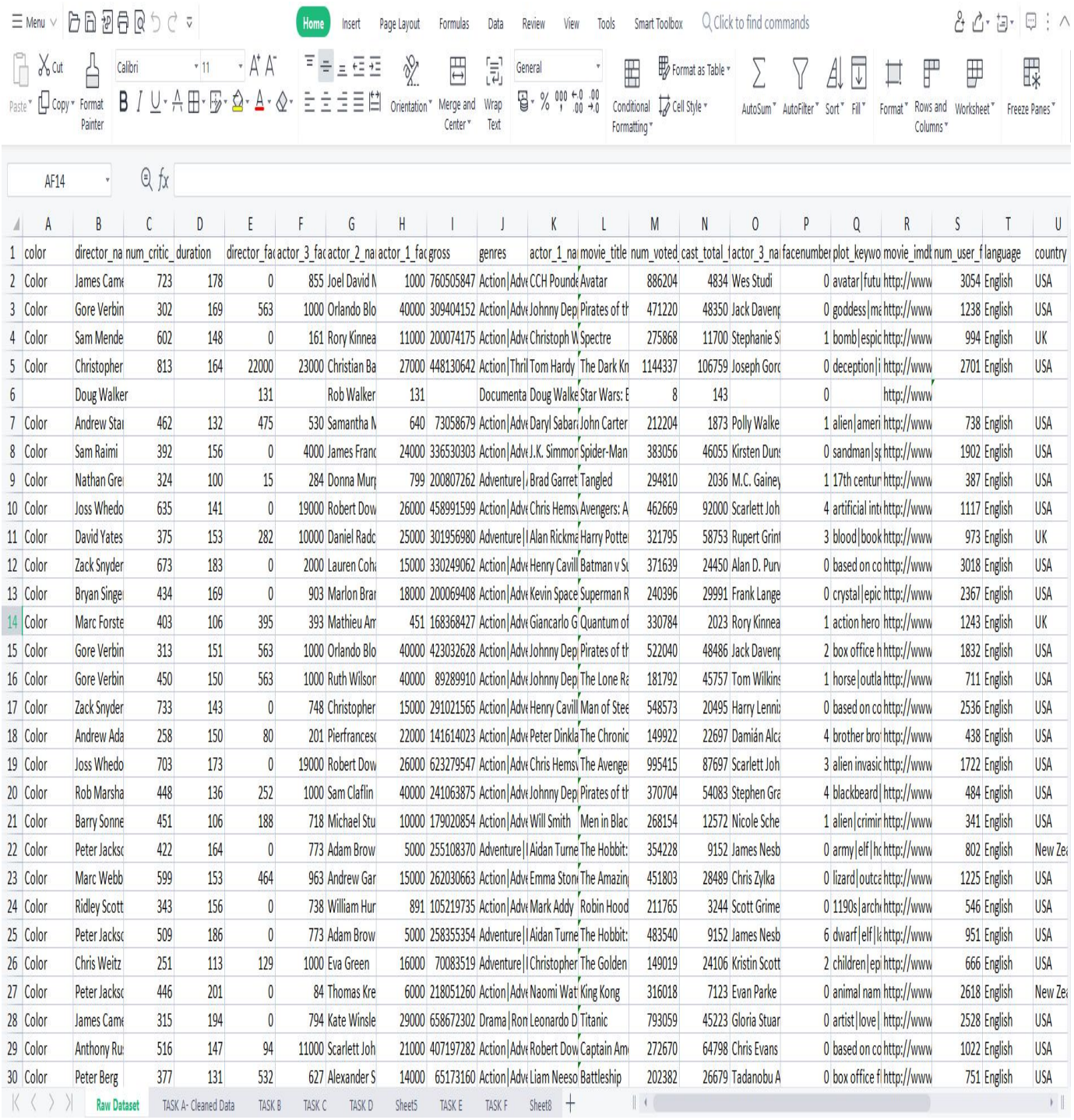
After solving the tasks by analyzing the data and getting the desired output, I got insights about the importance of cleaning and analyzing the data, movies with highest profit and IMDB ratings. Performed movie genre analysis, movie duration analysis, language analysis, director analysis and budget analysis. IMDB ratings of movies can influence people's viewing decisions. IMDB movie analysis helps online streaming platforms or movie makers to understand people's interests, so that they can provide content with value and also gain good profit from it.

Result:

While making this project, I achieved the confidence to practically apply my knowledge to analyze the dataset after cleaning the data, carry out the tasks provided by the team and got the required output. It has helped me to derive useful insights for the team, from the output data. This project has helped me to gain a clear understanding about statistics, advanced excel understanding such as functions, filters, shortcuts, conditional formatting, sorting, formulas, grouping, creating pivot tables, creating appropriate charts, finding percentile, etc. I learned to visually represent the output data using charts/graphs. I learned to use Microsoft excel as a data analyst.

Here is the screenshot of the raw dataset provided.

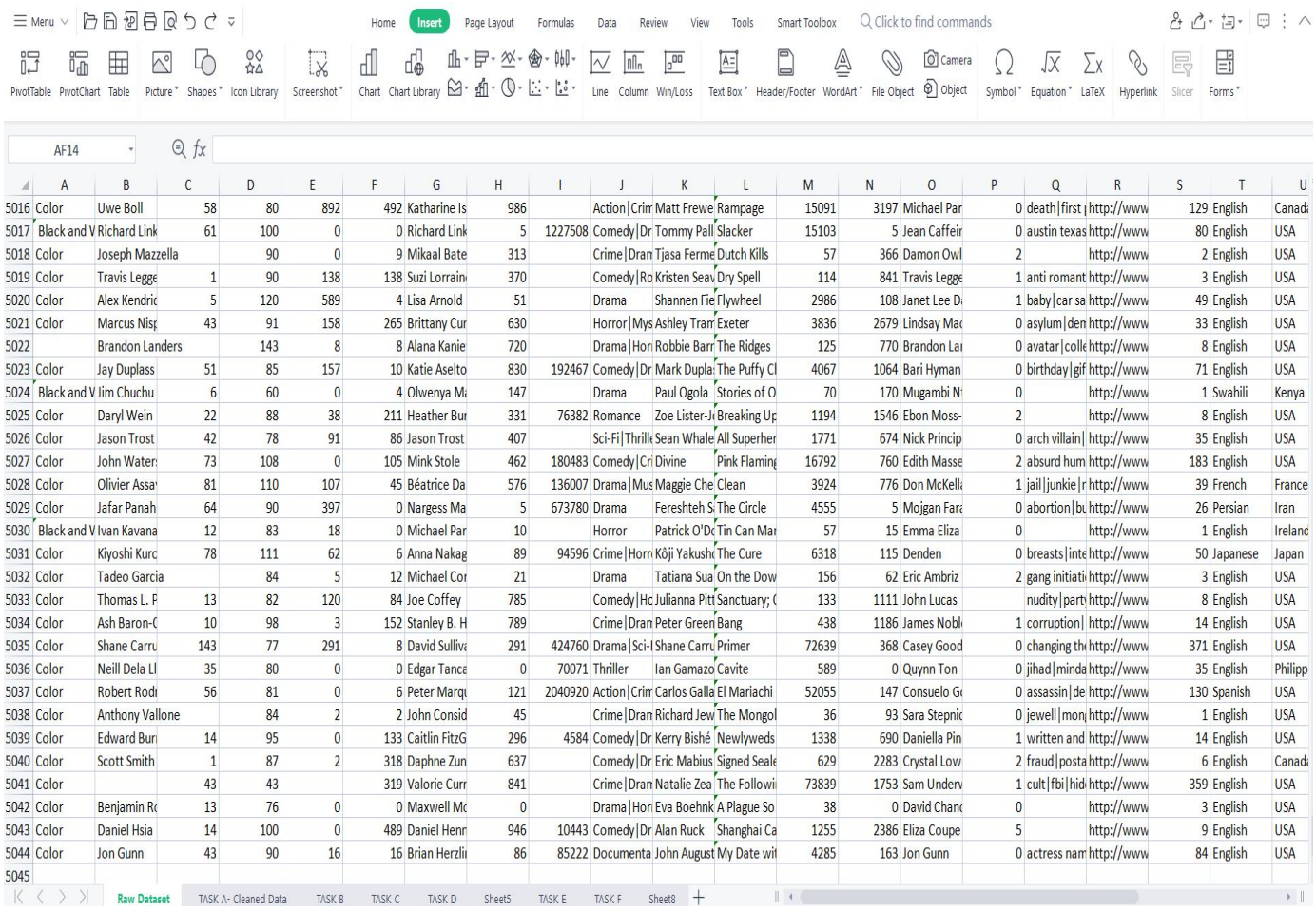
The raw dataset given contains 28 columns and 5044 rows.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	color	director_na	num_critic_	duration	director_fa	actor_3_fa	actor_2_na	actor_1_fa	gross	genres	actor_1_na	movie_title	num_voted	cast_total_	actor_3_na	facenumber	plot_keywo	movie_imdb	num_user_f	language	country
2	Color	James Came	723	178	0	855	Joel David M	1000	760505847	Action|Adve	CCH Pounde	Avatar	886204	4834	Wes Studi	0	avatar|futu	http://www	3054	English	USA
3	Color	Gore Verbin	302	169	563	1000	Orlando Blo	40000	309404152	Action|Adve	Johnny Dep	Pirates of th	471220	48350	Jack Davenp	0	goddess|ma	http://www	1238	English	USA
4	Color	Sam Mende	602	148	0	161	Rory Kinnea	11000	200074175	Action|Adve	Christoph W	Spectre	275868	11700	Stephanie Si	1	bomb|espic	http://www	994	English	UK
5	Color	Christopher	813	164	22000	23000	Christian Ba	27000	448130642	Action|Thril	Tom Hardy	The Dark Kn	1144337	106759	Joseph Gord	0	deception|i	http://www	2701	English	USA
6		Doug Walker			131		Rob Walker	131		Documenta	Doug Walke	Star Wars: E	8	143		0		http://www			
7	Color	Andrew Stai	462	132	475	530	Samantha M	640	73058679	Action|Adve	Daryl Sabar	John Carter	212204	1873	Polly Walke	1	alien|ameri	http://www	738	English	USA
8	Color	Sam Raimi	392	156	0	4000	James Franc	24000	336530303	Action|Adve	J.K. Simmor	Spider-Man	383056	46055	Kirsten Dun	0	sandman|sp	http://www	1902	English	USA
9	Color	Nathan Gre	324	100	15	284	Donna Murp	799	200807262	Adventure|A	Brad Garret	Tangled	294810	2036	M.C. Gainey	1	17th centur	http://www	387	English	USA
10	Color	Joss Whedo	635	141	0	19000	Robert Dow	26000	458991599	Action|Adve	Chris Hemsv	Avengers: A	462669	92000	Scarlett Joh	4	artificial int	http://www	1117	English	USA
11	Color	David Yates	375	153	282	10000	Daniel Radc	25000	301956980	Adventure|I	Alan Rickma	Harry Potte	321795	58753	Rupert Grint	3	blood|book	http://www	973	English	UK
12	Color	Zack Snyder	673	183	0	2000	Lauren Coha	15000	330249062	Action|Adve	Henry Cavill	Batman v Su	371639	24450	Alan D. Purv	0	based on co	http://www	3018	English	USA
13	Color	Bryan Singe	434	169	0	903	Marlon Brar	18000	200069408	Action|Adve	Kevin Space	Superman R	240396	29991	Frank Lange	0	crystal|epic	http://www	2367	English	USA
14	Color	Marc Forste	403	106	395	393	Mathieu Am	451	168368427	Action|Adve	Giancarlo G	Quantum of	330784	2023	Rory Kinnea	1	action hero	http://www	1243	English	UK
15	Color	Gore Verbin	313	151	563	1000	Orlando Blo	40000	423032628	Action|Adve	Johnny Dep	Pirates of th	522040	48486	Jack Davenp	2	box office h	http://www	1832	English	USA
16	Color	Gore Verbin	450	150	563	1000	Ruth Wilson	40000	89289910	Action|Adve	Johnny Dep	The Lone Ra	181792	45757	Tom Wilkins	1	horse|outla	http://www	711	English	USA
17	Color	Zack Snyder	733	143	0	748	Christopher	15000	291021565	Action|Adve	Henry Cavill	Man of Stee	548573	20495	Harry Lennix	0	based on co	http://www	2536	English	USA
18	Color	Andrew Ada	258	150	80	201	Pierfrancesc	22000	141614023	Action|Adve	Peter Dinkla	The Chronic	149922	22697	Damián Alca	4	brother bro	http://www	438	English	USA
19	Color	Joss Whedo	703	173	0	19000	Robert Dow	26000	623279547	Action|Adve	Chris Hemsv	The Avenge	995415	87697	Scarlett Joh	3	alien invasic	http://www	1722	English	USA
20	Color	Rob Marsha	448	136	252	1000	Sam Claflin	40000	241063875	Action|Adve	Johnny Dep	Pirates of th	370704	54083	Stephen Gra	4	blackbeard|	http://www	484	English	USA
21	Color	Barry Sonne	451	106	188	718	Michael Stu	10000	179020854	Action|Adve	Will Smith	Men in Blac	268154	12572	Nicole Sche	1	alien|crimir	http://www	341	English	USA
22	Color	Peter Jackso	422	164	0	773	Adam Brow	5000	255108370	Adventure|I	Aidan Turne	The Hobbit:	354228	9152	James Nesb	0	army|elf|hc	http://www	802	English	New Zea
23	Color	Marc Webb	599	153	464	963	Andrew Gar	15000	262030663	Action|Adve	Emma Ston	The Amazin	451803	28489	Chris Zylka	0	lizard|outca	http://www	1225	English	USA
24	Color	Ridley Scott	343	156	0	738	William Hur	891	105219735	Action|Adve	Mark Addy	Robin Hood	211765	3244	Scott Grime	0	1190s|arch	http://www	546	English	USA
25	Color	Peter Jackso	509	186	0	773	Adam Brow	5000	258355354	Adventure|I	Aidan Turne	The Hobbit:	483540	9152	James Nesb	6	dwarf|elf|k	http://www	951	English	USA
26	Color	Chris Weitz	251	113	129	1000	Eva Green	16000	70083519	Adventure|I	Christopher	The Golden	149019	24106	Kristin Scott	2	children|epi	http://www	666	English	USA
27	Color	Peter Jackso	446	201	0	84	Thomas Kre	6000	218051260	Action|Adve	Naomi Wat	King Kong	316018	7123	Evan Parke	0	animal nam	http://www	2618	English	New Zea
28	Color	James Came	315	194	0	794	Kate Winsle	29000	658672302	Drama|Ron	Leonardo D	Titanic	793059	45223	Gloria Stuar	0	artist|love|	http://www	2528	English	USA
29	Color	Anthony Ru	516	147	94	11000	Scarlett Joh	21000	407197282	Action|Adve	Robert Dow	Captain Am	272670	64798	Chris Evans	0	based on co	http://www	1022	English	USA
30	Color	Peter Berg	377	131	532	627	Alexander S	14000	65173160	Action|Adve	Liam Neeso	Battleship	202382	26679	Tadanobu A	0	box office fl	http://www	751	English	USA

Raw Dataset | TASK A- Cleaned Data | TASK B | TASK C | TASK D | Sheet5 | TASK E | TASK F | Sheet8

.....



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
5016	Color	Uwe Boll	58	80	892	492	Katharine Is	986		Action|Crim	Matt Frewe	Rampage	15091	3197	Michael Par	0	death|first|	http://www	129	English	Canada
5017	Black and V	Richard Link	61	100	0	0	Richard Link	5	1227508	Comedy|Dr	Tommy Pall	Slacker	15103	5	Jean Caffeir	0	austin texas	http://www	80	English	USA
5018	Color	Joseph Mazzella		90	0	9	Mikaal Bate	313		Crime|Dran	Tjasa Ferme	Dutch Kills	57	366	Damon Owl	2		http://www	2	English	USA
5019	Color	Travis Legge	1	90	138	138	Suzi Lorrain	370		Comedy|Ro	Kristen Seav	Dry Spell	114	841	Travis Legge	1	anti romant	http://www	3	English	USA
5020	Color	Alex Kendric	5	120	589	4	Lisa Arnold	51		Drama	Shannen Fie	Flywheel	2986	108	Janet Lee Di	1	baby|car sa	http://www	49	English	USA
5021	Color	Marcus Nisp	43	91	158	265	Brittany Cur	630		Horror|Mys	Ashley Tram	Exeter	3836	2679	Lindsay Mac	0	asylum|den	http://www	33	English	USA
5022		Brandon Landers		143	8	8	Alana Kanie	720		Drama|Hor	Robbie Barr	The Ridges	125	770	Brandon Lai	0	avatar|colle	http://www	8	English	USA
5023	Color	Jay Duplass	51	85	157	10	Katie Aselto	830	192467	Comedy|Dr	Mark Dupla	The Puffy Cl	4067	1064	Bari Hyman	0	birthday|gif	http://www	71	English	USA
5024	Black and V	Jim Chuchu	6	60	0	4	Olwenya Ma	147		Drama	Paul Ogola	Stories of O	70	170	Mugambi N	0		http://www	1	Swahili	Kenya
5025	Color	Daryl Wein	22	88	38	211	Heather Bur	331	76382	Romance	Zoe Lister-J	Breaking Up	1194	1546	Ebon Moss-	2		http://www	8	English	USA
5026	Color	Jason Trost	42	78	91	86	Jason Trost	407		Sci-Fi|Thrille	Sean Whale	All Superher	1771	674	Nick Princip	0	arch villain|	http://www	35	English	USA
5027	Color	John Waters	73	108	0	105	Mink Stole	462	180483	Comedy|Cri	Divine	Pink Flaming	16792	760	Edith Masse	2	absurd hum	http://www	183	English	USA
5028	Color	Olivier Assay	81	110	107	45	Béatrice Da	576	136007	Drama|Mus	Maggie Che	Clean	3924	776	Don McKella	1	jail|junkie|r	http://www	39	French	France
5029	Color	Jafar Panah	64	90	397	0	Nargess Ma	5	673780	Drama	Fereshteh S	The Circle	4555	5	Mojgan Fara	0	abortion|bu	http://www	26	Persian	Iran
5030	Black and V	Ivan Kavana	12	83	18	0	Michael Par	10		Horror	Patrick O'D	Tin Can Mar	57	15	Emma Eliza	0		http://www	1	English	Ireland
5031	Color	Kiyoshi Kurc	78	111	62	6	Anna Nakag	89	94596	Crime|Horr	Kôji Yakush	The Cure	6318	115	Denden	0	breasts|inte	http://www	50	Japanese	Japan
5032	Color	Tadeo Garcia		84	5	12	Michael Cor	21		Drama	Tatiana Sua	On the Dow	156	62	Eric Ambriz	2	gang initiati	http://www	3	English	USA
5033	Color	Thomas L. P	13	82	120	84	Joe Coffey	785		Comedy|Hc	Julianna Pitt	Sanctuary; (	133	1111	John Lucas		nudity|part	http://www	8	English	USA
5034	Color	Ash Baron-C	10	98	3	152	Stanley B. H	789		Crime|Dran	Peter Green	Bang	438	1186	James Nobl	1	corruption|	http://www	14	English	USA
5035	Color	Shane Carru	143	77	291	8	David Sulliva	291	424760	Drama|Sci-I	Shane Carru	Primer	72639	368	Casey Good	0	changing th	http://www	371	English	USA
5036	Color	Neill Dela Ll	35	80	0	0	Edgar Tanca	0	70071	Thriller	Ian Gamazo	Cavite	589	0	Quynn Ton	0	jihad|minda	http://www	35	English	Philipp
5037	Color	Robert Rodr	56	81	0	6	Peter Marqu	121	2040920	Action|Crim	Carlos Galla	El Mariachi	52055	147	Consuelo Gu	0	assassin|de	http://www	130	Spanish	USA
5038	Color	Anthony Vallone		84	2	2	John Consid	45		Crime|Dran	Richard Jew	The Mongol	36	93	Sara Stepnic	0	jewell|mon	http://www	1	English	USA
5039	Color	Edward Burr	14	95	0	133	Caitlin FitzG	296	4584	Comedy|Dr	Kerry Bishé	Newlyweds	1338	690	Daniella Pin	1	written and	http://www	14	English	USA
5040	Color	Scott Smith	1	87	2	318	Daphne Zun	637		Comedy|Dr	Eric Mabius	Signed Seale	629	2283	Crystal Low	2	fraud|posta	http://www	6	English	Canada
5041	Color		43	43		319	Valorie Curr	841		Crime|Dran	Natalie Zea	The Followi	73839	1753	Sam Underv	1	cult|fbi|hid	http://www	359	English	USA
5042	Color	Benjamin Ro	13	76	0	0	Maxwell Mc	0		Drama|Hor	Eva Boehnk	A Plague So	38	0	David Chanc	0		http://www	3	English	USA
5043	Color	Daniel Hsia	14	100	0	489	Daniel Henr	946	10443	Comedy|Dr	Alan Ruck	Shanghai Ca	1255	2386	Eliza Coupe	5		http://www	9	English	USA
5044	Color	Jon Gunn	43	90	16	16	Brian Herzli	86	85222	Documenta	John August	My Date wit	4285	163	Jon Gunn	0	actress nam	http://www	84	English	USA
5045																					

Output:

As a part of cleaning the data, first of all, I converted the dataset into a readable form. For that I used Ctrl A and double clicked on boundaries. After that I used Ctrl shift arrow keys to select the dataset. I removed the duplicates. 45 duplicates were removed.

Data--> Highlight duplicates--> Remove duplicates

I dropped unnecessary and not useful columns. I also deleted the rows containing null values.

Home--> Find and Replace--> Go To--> Blanks--> Go to

Selected altogether the rows containing blanks/null values and deleted.

After cleaning the data, I got 14 columns and 3850 rows. After cleaning the dataset, the next step is data analysis.

……

	director_name	num_critic_for_reviews	duration	gross	genres	actor_1_name
2	James Cameron	723	178	760505847	Action|Adventure|Fantasy|Sci-Fi	CCH Pounder
3	Gore Verbinski	302	169	309404152	Action|Adventure|Fantasy	Johnny Depp
4	Sam Mendes	602	148	200074175	Action|Adventure|Thriller	Christoph Waltz
5	Christopher Nolan	813	164	448130642	Action|Thriller	Tom Hardy
6	Andrew Stanton	462	132	73058679	Action|Adventure|Sci-Fi	Daryl Sabara
7	Sam Raimi	392	156	336530303	Action|Adventure|Romance	J.K. Simmons
8	Nathan Greno	324	100	200807262	Adventure|Animation|Comedy|Family|Fantasy|Musical|Romance	Brad Garrett
9	Joss Whedon	635	141	458991599	Action|Adventure|Sci-Fi	Chris Hemsworth
10	David Yates	375	153	301956980	Adventure|Family|Fantasy|Mystery	Alan Rickman
11	Zack Snyder	673	183	330249062	Action|Adventure|Sci-Fi	Henry Cavill
12	Bryan Singer	434	169	200069408	Action|Adventure|Sci-Fi	Kevin Spacey
13	Marc Forster	403	106	168368427	Action|Adventure	Giancarlo Giannini
14	Gore Verbinski	313	151	423032628	Action|Adventure|Fantasy	Johnny Depp
15	Gore Verbinski	450	150	89289910	Action|Adventure|Western	Johnny Depp
16	Zack Snyder	733	143	291021565	Action|Adventure|Fantasy|Sci-Fi	Henry Cavill
17	Andrew Adamson	258	150	141614023	Action|Adventure|Family|Fantasy	Peter Dinklage
18	Joss Whedon	703	173	623279547	Action|Adventure|Sci-Fi	Chris Hemsworth
19	Rob Marshall	448	136	241063875	Action|Adventure|Fantasy	Johnny Depp
20	Barry Sonnenfeld	451	106	179020854	Action|Adventure|Comedy|Family|Fantasy|Sci-Fi	Will Smith
21	Peter Jackson	422	164	255108370	Adventure|Fantasy	Aidan Turner
22	Marc Webb	599	153	262030663	Action|Adventure|Fantasy	Emma Stone
23	Ridley Scott	343	156	105219735	Action|Adventure|Drama|History	Mark Addy
24	Peter Jackson	509	186	258355354	Adventure|Fantasy	Aidan Turner
25	Chris Weitz	251	113	70083519	Adventure|Family|Fantasy	Christopher Lee
26	Peter Jackson	446	201	218051260	Action|Adventure|Drama|Romance	Naomi Watts
27	James Cameron	315	194	658672302	Drama|Romance	Leonardo DiCaprio
28	Anthony Russo	516	147	407197282	Action|Adventure|Sci-Fi	Robert Downey Jr.

	director_name	num_critic_for_reviews	duratio	gross	genres	actor_1_name	movie_title
3824	Dennis Iliadis	241	114	32721635	Crime|Horror|Thriller	Tony Goldwyn	The Last House on the Left
3825	Darren Aronofsky	138	84	3216970	Drama|Mystery|Thriller	Mark Margolis	Pi
3826	Myles Berkowitz	32	87	536767	Biography|Comedy|Romance	Tia Carrere	20 Dates
3827	Morgan Spurlock	193	100	11529368	Comedy|Documentary|Drama	Chemeeka Walker	Super Size Me
3828	Brandon Trost	66	82	40557	Comedy	Clifton Collins Jr.	The FP
3829	Joe Swanberg	65	82	30084	Comedy|Drama	Anna Kendrick	Happy Christmas
3830	Edward Burns	36	98	10246600	Comedy|Drama|Romance	Shari Albert	The Brothers McMullen
3831	Lena Dunham	113	98	389804	Comedy|Drama|Romance	Lena Dunham	Tiny Furniture
3832	David Gordon Green	75	90	241816	Drama	Paul Schneider	George Washington
3833	Kevin Jordan	21	90	277233	Comedy|Romance	Derick Martini	Smiling Fish & Goat on Fire
3834	Mike Bruce	3	78	243768	Western	Joseph Campanella	The Legend of God's Gun
3835	Kevin Smith	136	102	3151130	Comedy	Jason Mewes	Clerks
3836	James Bidgood	8	65	8231	Drama|Fantasy	Don Brooks	Pink Narcissus
3837	Neil LaBute	80	97	2856622	Comedy|Drama	Stacy Edwards	In the Company of Men
3838	David Ayer	233	109	10499968	Action|Crime|Drama|Thriller	Mireille Enos	Sabotage
3839	Richard Linklater	61	100	1227508	Comedy|Drama	Tommy Pallotta	Slacker
3840	Jay Duplass	51	85	192467	Comedy|Drama|Romance	Mark Duplass	The Puffy Chair
3841	Daryl Wein	22	88	76382	Romance	Zoe Lister-Jones	Breaking Upwards
3842	John Waters	73	108	180483	Comedy|Crime|Horror	Divine	Pink Flamingos
3843	Olivier Assayas	81	110	136007	Drama|Music|Romance	Maggie Cheung	Clean
3844	Jafar Panahi	64	90	673780	Drama	Fereshteh Sadre Orafaiy	The Circle
3845	Kiyoshi Kurosawa	78	111	94596	Crime|Horror|Mystery|Thriller	Kôji Yakusho	The Cure
3846	Shane Carruth	143	77	424760	Drama|Sci-Fi|Thriller	Shane Carruth	Primer
3847	Neill Dela Llana	35	80	70071	Thriller	Ian Gamazon	Cavite
3848	Robert Rodriguez	56	81	2040920	Action|Crime|Drama|Romance|Thriller	Carlos Gallardo	El Mariachi
3849	Edward Burns	14	95	4584	Comedy|Drama	Kerry Bishé	Newlyweds
3850	Jon Gunn	43	90	85222	Documentary	John August	My Date with Drew

Below are each of the Data Analytics tasks given by the team and their respective outputs.

A.) Movie Genre Analysis: Analyze the distribution of movie genres and their impact on the IMDB score.

Task: Determine the most common genres of movies in the dataset. Then, for each genre, calculate descriptive statistics (mean, median, mode, range, variance, standard deviation) of the IMDB scores.

Output:

For getting the unique genres of movies, I followed the following steps.

Select Column-->Data-->Text to Columns-->Smart Split Columns

And after that, I used filters to find the unique genres and used COUNTIF(eg; =COUNTIF(C2:J3850,"Drama")) function to find the count of movies made in each genre. After that sorted the count in descending order to find the genres with the most movie count.

Next is to calculate descriptive statistics (mean, median, mode, range, variance, standard deviation) of the IMDB scores.

For that, I created a pivot table to find out the average, max, min, var, stdev of the IMDB scores corresponding to each unique movie genre.

I used excel functions such as MEDIAN and MODE to find out the median and mode of the IMDB scores of the corresponding movie genre.

For finding the range, I used the following formula: Range = Max - Min

movie_title	imdb_score	genres						
AvatarÂ	7.9	Action	Adventure	Fantasy	Sci-Fi			
Pirates of the Caribbean: At World's EndÂ	7.1	Action	Adventure	Fantasy				
SpectreÂ	6.8	Action	Adventure	Thriller				
The Dark Knight RisesÂ	8.5	Action	Thriller					
John CarterÂ	6.6	Action	Adventure	Sci-Fi				
Spider-Man 3Â	6.2	Action	Adventure	Romance				
TangledÂ	7.8	Adventure	Animation	Comedy	Family	Fantasy	Musical	Romance
Avengers: Age of UltronÂ	7.5	Action	Adventure	Sci-Fi				
Harry Potter and the Half-Blood PrinceÂ	7.5	Adventure	Family	Fantasy	Mystery			
Batman v Superman: Dawn of JusticeÂ	6.9	Action	Adventure	Sci-Fi				
Superman ReturnsÂ	6.1	Action	Adventure	Sci-Fi				
Quantum of SolaceÂ	6.7	Action	Adventure					
Pirates of the Caribbean: Dead Man's ChestÂ	7.3	Action	Adventure	Fantasy				
The Lone RangerÂ	6.5	Action	Adventure	Western				
Man of SteelÂ	7.2	Action	Adventure	Fantasy	Sci-Fi			
The Chronicles of Narnia: Prince CaspianÂ	6.6	Action	Adventure	Family	Fantasy			
The AvengersÂ	8.1	Action	Adventure	Sci-Fi				
Pirates of the Caribbean: On Stranger TidesÂ	6.7	Action	Adventure	Fantasy				
Men in Black 3Â	6.8	Action	Adventure	Comedy	Family	Fantasy	Sci-Fi	
The Hobbit: The Battle of the Five ArmiesÂ	7.5	Adventure	Fantasy					
The Amazing Spider-ManÂ	7	Action	Adventure	Fantasy				
Robin HoodÂ	6.7	Action	Adventure	Drama	History			
The Hobbit: The Desolation of SmaugÂ	7.9	Adventure	Fantasy					
The Golden CompassÂ	6.1	Adventure	Family	Fantasy				

The following are the unique movie genres and the number of movies made in the corresponding genre.

	A	B
1	**Movie genre**	**Movie count**
2	Drama	1941
3	Comedy	1504
4	Thriller	1117
5	Action	962
6	Romance	878
7	Adventure	787
8	Crime	714
9	Fantasy	514
10	Sci-Fi	497
11	Family	450
12	Horror	391
13	Mystery	383
14	Biography	243
15	Animation	199
16	War	160
17	Music	159
18	History	153
19	Sport	151
20	Musical	103
21	Documentary	64
22	Western	58
23	Short	2
24	Film-Noir	1

Below is the Descriptive Statistics of the most common movie genres(top 10)

Genre	movie Count	mean_imdb	median_imdb	mode_imdb	max_imdb	min_imdb	Range_imdb	var_imdb	stdev_imdb
Drama	1941	6.786192684	6.9	6.7	9.3	2.1	7.2	0.794056682	0.891098582
Comedy	1504	6.184441489	6.3	6.7	8.8	1.9	6.9	1.076697226	1.03764022
Thriller	1117	6.378066249	6.4	6.5	9	2.7	6.3	0.933470092	0.96616256
Action	962	6.290748441	6.3	6.1	9	2.1	6.9	1.06375407	1.03138454
Romance	878	6.42881549	6.5	6.5	8.5	2.1	6.4	0.93466473	0.966780601
Adventure	787	6.458322745	6.6	6.7	8.9	2.3	6.6	1.227980898	1.108142995
Crime	714	6.544817927	6.6	6.6	9.3	2.4	6.9	0.960709434	0.980157862
Fantasy	514	6.290077821	6.4	6.7	8.9	2.2	6.7	1.276607011	1.129870352
Sci-Fi	497	6.322736419	6.4	6.7	8.8	1.9	6.9	1.336558626	1.156096287
Family	450	6.210444444	6.3	6.7	8.6	1.9	6.7	1.351583321	1.162576157

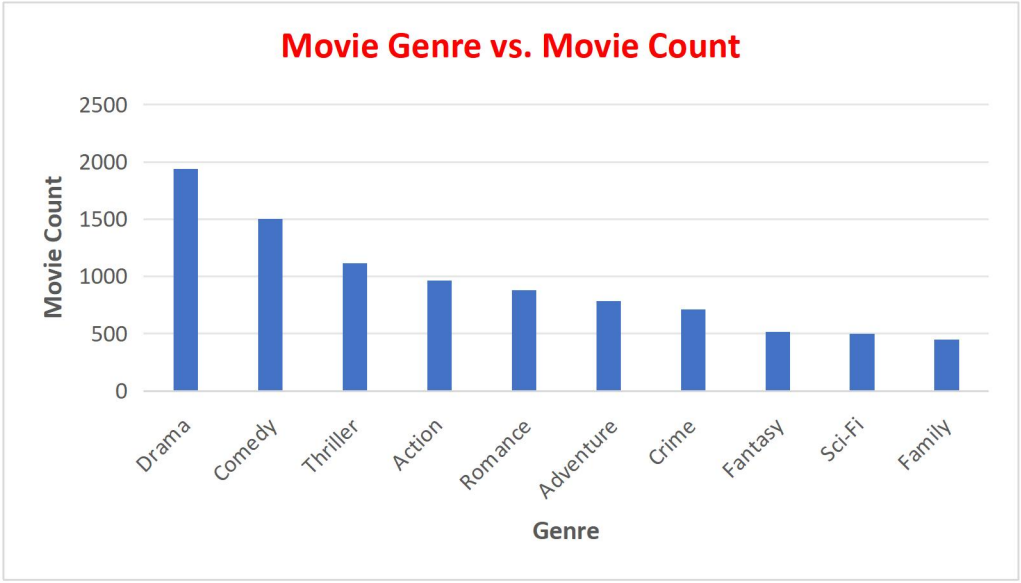
The table below shows the most common genres sorted in terms of their Mean IMDB scores.

Genre	mean_imdb	median_imdb	mode_imdb
Drama	6.786192684	6.9	6.7
Crime	6.544817927	6.6	6.6
Adventure	6.458322745	6.6	6.7
Romance	6.42881549	6.5	6.5
Thriller	6.378066249	6.4	6.5
Sci-Fi	6.322736419	6.4	6.7
Action	6.290748441	6.3	6.1
Fantasy	6.290077821	6.4	6.7
Family	6.210444444	6.3	6.7
Comedy	6.184441489	6.3	6.7

The table below shows the most common genres sorted in terms of their variability in IMDB scores.

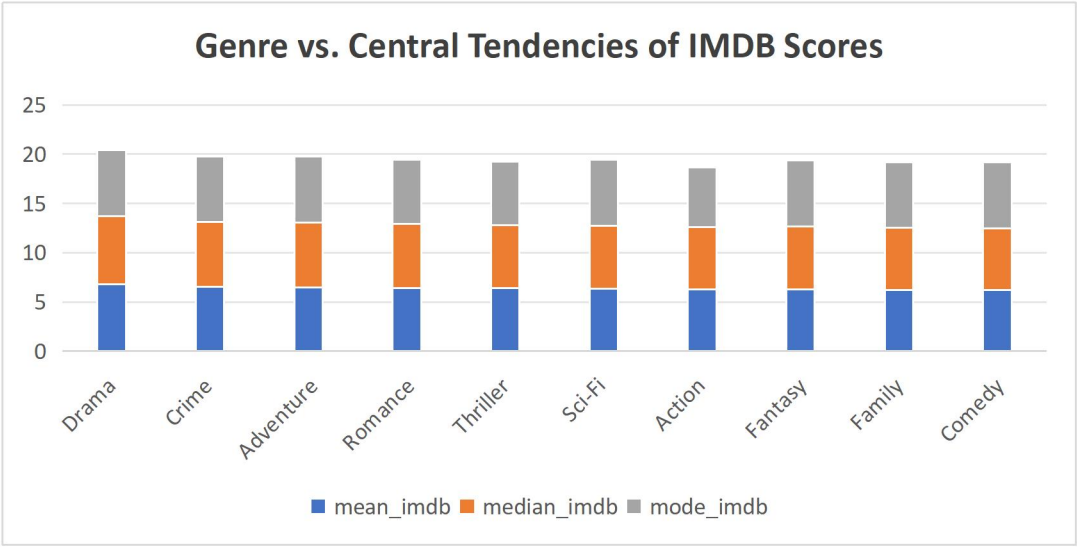
Genre	Range_imdb	var_imdb	stdev_imdb
Family	6.7	1.351583321	1.162576157
Sci-Fi	6.9	1.336558626	1.156096287
Fantasy	6.7	1.276607011	1.129870352
Adventure	6.6	1.227980898	1.108142995
Comedy	6.9	1.076697226	1.03764022
Action	6.9	1.06375407	1.03138454
Crime	6.9	0.960709434	0.980157862
Romance	6.4	0.93466473	0.966780601
Thriller	6.3	0.933470092	0.96616256
Drama	7.2	0.794056682	0.891098582

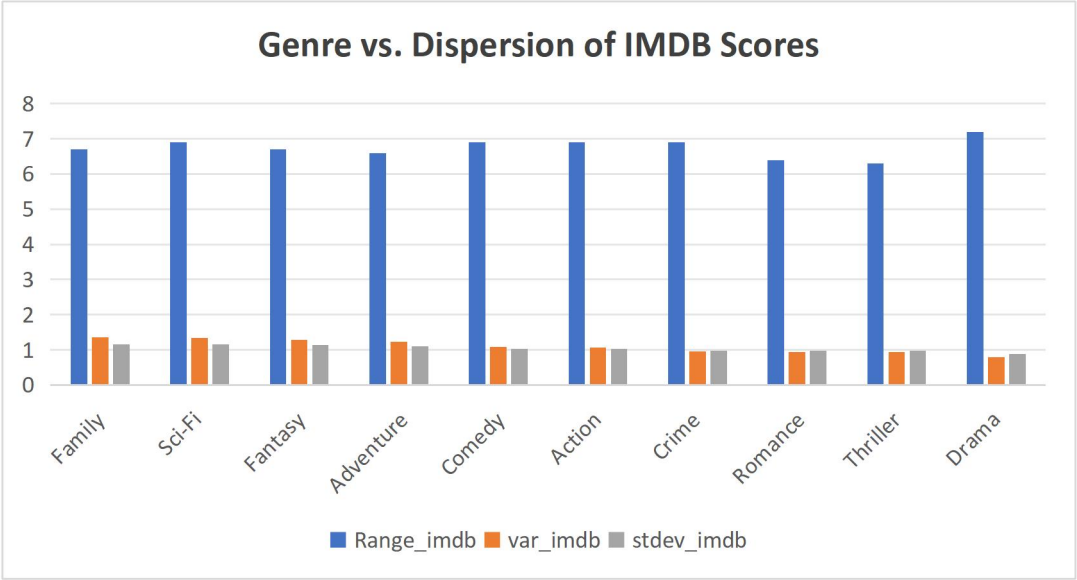
Below is the data visualization showing the number of movies made in the popular genres.



In the descriptive statistics, the measures of central tendencies are mean, median, and mode. And the measures of dispersion or variability are range, variance, and standard deviation.

Below are the data visualizations showing the descriptive statistics of the IMDB scores of the most common movie genres.





Insights:

1) The most common and the genres having the highest movie count are Drama, Comedy, Thriller, Action, Romance, Adventure, Crime, Fantasy, Sci-fi and Family.

2) Genres like Drama and Crime have highest average IMDB ratings, while Comedy and Family have least average IMDB ratings.

3) Family and Sci-Fi movies have greater variability(variance and standard deviation) in IMDB ratings, while Drama, Thriller and Romance movies have lower variability in IMDB ratings.

4) Genre of movies with high average IMDB scores and less variability in the IMDB scores are tend to be successful as they have consistent high IMDB ratings

5) Genre of movies with low average IMDB scores and less variability in the IMDB scores tend to perform less compared to others

6) High variation in the IMDB scores means the ratings are inconsistent. There is no constant success or failure of these movies. These movies can be well-received or poorly-received by the audience.

7) Drama has the highest average IMDB score and lowest variability rates. It tends to have stable IMDB ratings. It is most the popular and most consistently rated genre. This means that it is well-received by the audience generally. Drama movies usually have great stories.

8) Many movies are made in Comedy genre, but it has the least average IMDB score. It maybe because audience have different taste in humor. The moderate variability in IMDB scores shows it tends to have low or mixed IMDB ratings, making it unpredictable.

9) Movies in genres like Crime, Adventure, Romance, Thriller have fairly good average IMDB scores and also having only low and moderate variability in IMDB scores, implying fairly consistent ratings. This maybe because younger audience prefer these types of movies. These movies tend to perform high or average.

10) Family, Sci-fi, Fantasy movies high variance in IMDB scores implying inconsistency in ratings and also have low average IMDB ratings. These movies tend to perform less compared to others

11) Sci-fi, Fantasy, Action, Adventure movies can be very successful or flop due to the technicalities in the making of the film. It may require high budget. It can be at high risk if it is not well executed.

12) Action movies tend to have moderate average IMDB scores and moderate variance in IMDB scores which means it have mixed reception from audience

13) The success of the movies of any particular genre is depending on the preference of the audience. Different age groups of audience have diverse taste. Some genres might be liked by the people of all age groups. The movies which are not family-friendly will be limited to some age groups.

14) Each movie belongs to a main genre with a bit of mix of multiple genres. A compelling story-telling well executed with great production quality tend to be successful and can get high IMDB ratings.

For Example; The Movie Titanic is a successful film with great IMDB score and it is very well-received by the audience. The IMDB ratings are given by the audience and critics. It belongs to genre Drama with Romance. It has a unique and interesting story and characters. It is directed by a great filmmaker James Cameron. The main actor is Leonardo DiCaprio who is a favourite actor of many. It has high production quality and it is a well-executed movie. So the viewing experience of the movie was very good. So it attracted world-wide audience and became commercially successful. So the audience give higher IMDB ratings. There are many factors involved.

So it means Genres do have a great impact on IMDB ratings but it alone doesn't determine a movies success.

Five 'Whys' Approach:

Drama, crime, romance, thriller, and adventure movies tend to have higher IMDB ratings, while comedy and family movies tend to have lower ratings.

Based on High IMDB rated Genres

1) Why drama crime, romance, thriller, and adventure movies tend to have higher IMDB ratings?

A. These genres usually focus on great story-telling, character development, and emotional impact which resonates strongly with large audiences.

2) Why does great story-telling and emotional impact lead to higher IMDB ratings?

A. Audience like meaningful stories that stir strong feelings which keep them engaged and interested in the movie.

3) Why does audience engagement lead to higher IMDB ratings?

A. Audience are inclined to give positive reviews and suggest it to others when they sense an emotional connection to it.

4) Why does positive word-of-mouth lead to higher IMDB ratings?

A. Movies that are well-received by the audience attract more viewers, increasing their popularity and ratings.

5) Why does increased popularity leads to higher IMDB ratings?

A. A positive feedback cycle is created when more people view and appreciate the movie, resulting in higher ratings

Based on Low IMDB rated Genres

1) Why does comedy and family movies tend to have lower IMDB ratings?

A. Comedy is subjective as the audience have varied humor preferences. So some people may not find it funny. Family movies cater to children which limits their appeal to adults.

2) Why does subjectivity affect IMDB ratings?

A. The IMDB ratings can be pulled down compared to the broadly appealing genre like Drama, due to the bigger variations in IMDB ratings caused by mixed reactions

3) Why does broader appeal lead to higher IMDB ratings?

A. Broadly appealing genres focus on all human emotions which makes it relatable to a wider audience, leading to stable high IMDB ratings

4) Why does some comedy and family movies still succeed?

A. Some movies have great story and emotional depth which maybe relatable and could connect to a large audience, so it can transcend limitations and score high IMDB ratings

5) Why does storytelling matters for all genres?

A. No matter what genre is, audience connect and respond positively to movies with a compelling story with emotional depth, leading to higher IMDB ratings

Recommendations for Filmmakers:

1) Drama and Romance movies tend to have stable ratings which makes them a safe option for filmmakers

2) Crime and Thriller movies are generally well-received making them a good choice

3) Comedy movies can perform well or flop due to diverse humor preferences

4) Sci-fi, Adventure, Fantasy, and Action movies are risky and at the same time can be blockbusters if it is well executed. They also have the potential for low IMDB ratings

5) Family movies tend to perform less but can be successful depending on how good the story resonate with the audience

6) No matter what the genre is, a movie receives success and high IMDB ratings when it has factors like great story-telling, emotional impact and connection creation, good marketing, high production quality and well execution

B.) Movie Duration Analysis: Analyze the distribution of movie durations and its impact on the IMDB score.

Task: Analyze the distribution of movie durations and identify the relationship between movie duration and IMDB score.

Output:

	movie_title	duratic	imdb_scor
2	Blood In, Blood Out	330	8
3	Heaven's Gate	325	6.8
4	The Legend of Suriyothai	300	6.6
5	Das Boot	293	8.4
6	Apocalypse Now	289	8.5
7	Gods and Generals	280	6.3
8	Gettysburg	271	7.7
9	Cleopatra	251	7
10	Once Upon a Time in America	251	8.4
11	The Wolf of Wall Street	240	8.2
12	Dances with Wolves	236	8
13	Lawrence of Arabia	227	8.4
14	Gone with the Wind	226	8.2
15	The Greatest Story Ever Told	225	6.6
16	All the Pretty Horses	220	5.8
17	The Godfather: Part II	220	9
18	The Last Emperor	219	7.8
19	Gangs of New York	216	7.5
20	Watchmen	215	7.7
21	The Thin Red Line	215	7.6
22	Woodstock	215	8.1
23	Wyatt Earp	212	6.6
24	Nixon	212	7.1
25	Alexander	206	5.5
26	JFK	206	8
27	Malcolm X	202	7.7
28	Seven Samurai	202	8.7

Using filters, the longest and shortest movie durations are found.

The longest movie is 330 minutes long, i.e., approximately 5.5 hours.

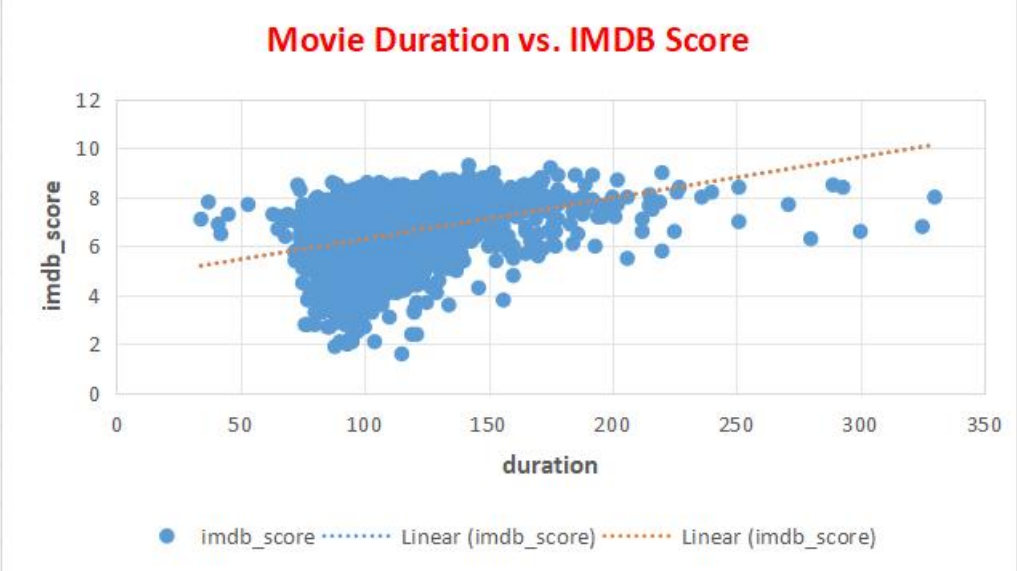
The shortest movie is 34 minutes long.

The next step is to find the descriptive statistics for movie durations. I found the mean, median and standard deviation of movie durations using the excel functions such as AVERAGE,MEDIAN,STDEV, etc.

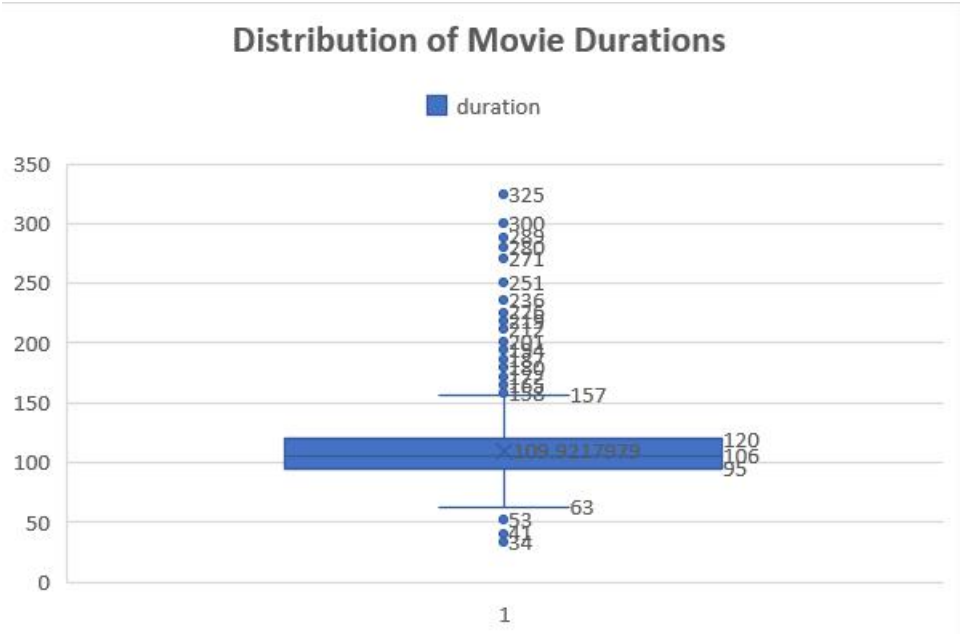
The mean,median and standard deviation of movie durations are 109.92 minutes(approx. 1 hour 50 minutes), 106 minutes and 22.75 minutes respectively.

I created a Scatter plot to visualize the relationship between movie duration and IMDB score. A trendline is added to assess the direction and strength of the relationship.

The correlation coefficient between the movie durations and IMDB scores is found using CORREL function. Correlation coefficient = 0.3603



Below is a data visualization(box plot) showing the distribution of movie durations.



The scatter plot shows a slight positive correlation which means the movies with longer duration tend to have slightly higher IMDB ratings. The IMDB score increases with respect to duration in a slightly less manner. The strength of the relationship between duration and IMDB score is slightly weak. Most of the movies have duration in between 80 and 140 minutes. Some outliers exist for very short or very long movies with low/high ratings.

Insights:

1) Most of the movies have duration in between 80 and 140 minutes. But their IMDB ratings ranges from 4 to 8 which means duration alone does not determine the success of a movie

2) The trendline shows positive correlation which means that movies with longer duration tend to have slightly higher IMDB ratings. It shows positive trend and the strength of the relationship between movie duration and IMDB score is slightly weak and not so strong.

3) The movies with very long duration and very short duration have fairly great IMDB ratings

4) The movies with very long duration that is greater than 200 minutes tend to have higher ratings. There are not many such films. They seem as outliers. Because they are epic classic movies and are critically acclaimed.

5) The average length of a movie is approximately 110 minutes

6) Movies with duration below 60 minutes tend to have lower IMDB ratings but some have good ratings

7) Outliers exist for both long and short movies which means IMDB ratings are varying regardless of the movie duration.

Five 'Whys' Approach:

1) Why do longer movies tend to have higher IMDB ratings?

A. Longer movies has more time for intricate storytelling, world-building character development, which makes audience impressed and more deeply engaged in it.

2) Why does more time for storytelling lead to higher IMDB ratings?

A. Longer movies usually has good stories with well developed characters and complex narratives leads to stronger emotional impact that connects with the audience better, giving them great viewing experience and satisfaction. This can lead to increased IMDB ratings.

3) Why does significant emotional impact leads to higher IMDB ratings?

A. Audience are more inclined to give higher IMDB ratings to a movie and recommend it to others, when they feel connected and deeply engaged to it.

4) Why does audience engagement important for higher IMDB ratings?

A. A movie that keeps the audience engaged in it, leaves a lasting impact which can lead to favourable word-of-mouth and higher IMDB ratings.

5) Why do shorter movies have a hard time achieving higher IMDB ratings?

A. These movies often lack depth which makes it harder to build emotional connections with the audience, leading to lower engagement and inconsistent IMDB ratings.

Recommendations for Filmmakers:

1) If targeting the mainstream audience, it is better to make movies with duration in the range 90-150 minutes, as these tend to have stable and higher IMDB ratings. These tend to be commercially successful
2) Movies with duration less than 60 minutes are risky because it tend to have low IMDB ratings. It has chance of creating less emotional impact or an unfinished feeling among audience. It is not suitable for theatrical releases which aims larger audience. A short film can get higher ratings if it has a great quality story progression.

3) Movies with duration greater than 150 minutes tend to receive good ratings but it can also vary and can be poorly received sometimes.
4) Movies which are very lengthy greater than 200 minutes can get very high ratings. So it should ensure high quality otherwise it can perform poorly.
5) Duration alone does not determine the success of a movie
6) Make sure the movies have high quality story and engaging content, character development, elements which connects the audience, well edited and executed, etc., to justify its length of the duration.

C.) Language Analysis: Situation: Examine the distribution of movies based on their language.

Task: Determine the most common languages used in movies and analyze their impact on the IMDB score using descriptive statistics.

Output:

	A	B	C
1	movie_title	language	imdb_score
2	Avatar	English	7.9
3	Pirates of the Caribbean: At World's End	English	7.1
4	Spectre	English	6.8
5	The Dark Knight Rises	English	8.5
6	John Carter	English	6.6
7	Spider-Man 3	English	6.2
8	Tangled	English	7.8
9	Avengers: Age of Ultron	English	7.5
10	Harry Potter and the Half-Blood Prince	English	7.5
11	Batman v Superman: Dawn of Justice	English	6.9
12	Superman Returns	English	6.1
13	Quantum of Solace	English	6.7
14	Pirates of the Caribbean: Dead Man's Chest	English	7.3
15	The Lone Ranger	English	6.5
16	Man of Steel	English	7.2
17	The Chronicles of Narnia: Prince Caspian	English	6.6
18	The Avengers	English	8.1
19	Pirates of the Caribbean: On Stranger Tides	English	6.7
20	Men in Black 3	English	6.8
21	The Hobbit: The Battle of the Five Armies	English	7.5
22	The Amazing Spider-Man	English	7
23	Robin Hood	English	6.7
24	The Hobbit: The Desolation of Smaug	English	7.9
25	The Golden Compass	English	6.1
26	King Kong	English	7.2
27	Titanic	English	7.7
28	Captain America: Civil War	English	8.2

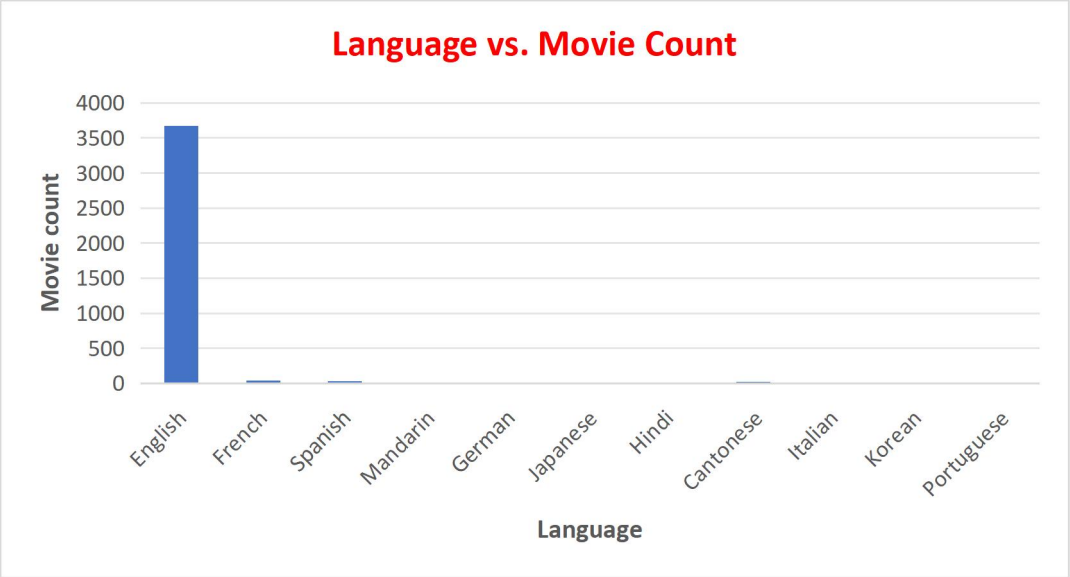
Using the pivot table, I found the Unique languages and the number of movies made in each language, and also the mean and standard deviation of IMDB scores corresponding to each language.

I found the median of IMDB scores of movies corresponding to each language using MEDIAN function.

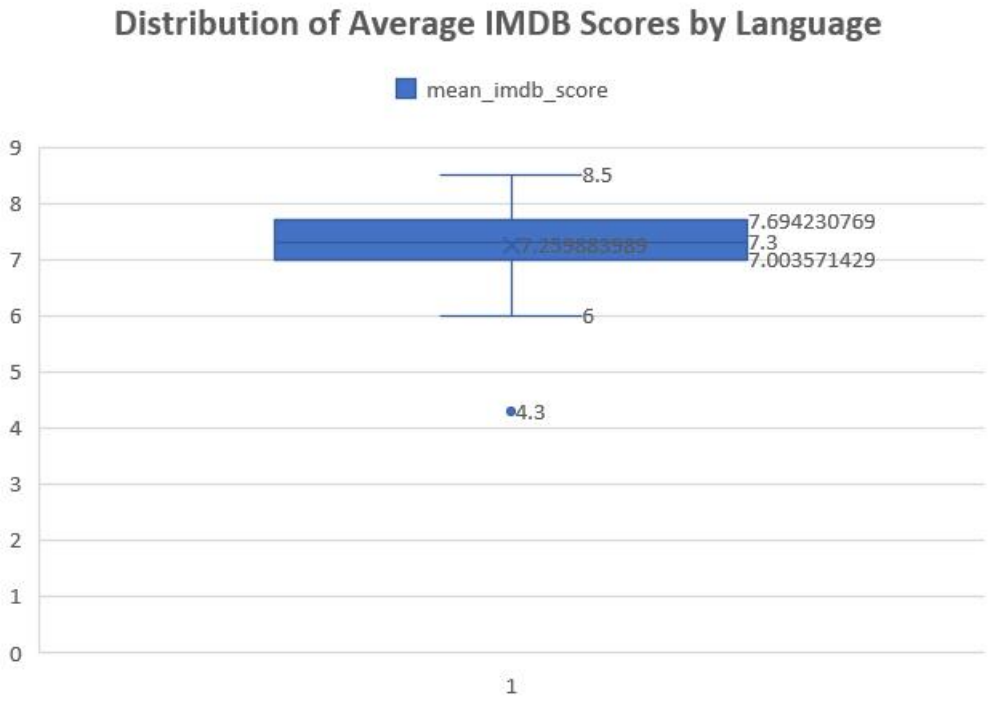
The most common languages used in movies are highlighted in yellow.

Language	Movie Count	mean_imdb_score	median_imdb	StdDev of imdb_score
English	3669	6.423821205	6.6	1.048621418
French	37	7.286486486	6.5	0.561328861
Spanish	26	7.05	6.5	0.826196103
Mandarin	14	7.021428571	6.5	0.765786244
German	13	7.692307692	6.7	0.640912811
Japanese	12	7.625	6.6	0.899621132
Hindi	10	6.76	6.7	1.111755369
Cantonese	8	7.2375	6.6	0.440575922
Italian	7	7.185714286	6.6	1.155318962
Korean	5	7.7	6.7	0.570087713
Portuguese	5	7.76	6.8	0.978774744
Norwegian	4	7.15	6.8	0.574456265
Danish	3	7.9	6.8	0.529150262
Dutch	3	7.566666667	7.45	0.404145188
Hebrew	3	7.5	7.3	0.435889894
Persian	3	8.133333333	8.4	0.550757055
Thai	3	6.633333333	6.6	0.450924975
Aboriginal	2	6.95	6.95	0.777817459
Dari	2	7.5	7.5	0.141421356
Indonesian	2	7.9	7.9	0.424264069
Arabic	1	7.2	7.2	#DIV/0!
Aramaic	1	7.1	7.1	#DIV/0!
Bosnian	1	4.3	4.3	#DIV/0!
Czech	1	7.4	7.4	#DIV/0!
Dzongkha	1	7.5	7.5	#DIV/0!
Filipino	1	6.7	6.7	#DIV/0!
Hungarian	1	7.1	7.1	#DIV/0!
Icelandic	1	6.9	6.9	#DIV/0!
Kazakh	1	6	6	#DIV/0!
Maya	1	7.8	7.8	#DIV/0!
Mongolian	1	7.3	7.3	#DIV/0!
None	1	8.5	8.5	#DIV/0!
Romanian	1	7.9	7.9	#DIV/0!
Russian	1	6.5	6.5	#DIV/0!
Swedish	1	7.6	7.6	#DIV/0!
Telugu	1	8.4	8.4	#DIV/0!
Vietnamese	1	7.4	7.4	#DIV/0!
Zulu	1	7.3	7.3	#DIV/0!

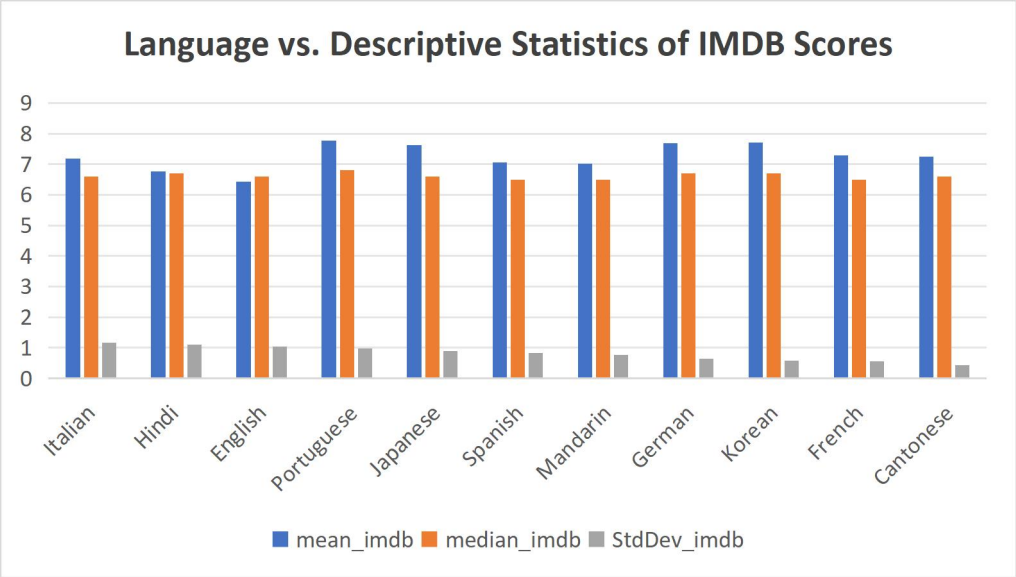
Below is the data visualization showing the most common languages used in the movies.



The data visualization below shows the distribution of average IMDB scores by language.



Below is the data visualization showing the descriptive statistics of IMDB scores by most common languages.



Insights:

1) The top 11 languages in which most number of movies are made are English, French, Spanish, Mandarin, German, Japanese, Hindi, Cantonese, Italian, Korean, and Portuguese.
2) More than 95% of the movies are made in English language and other languages have comparatively very less movies made in it.
3) English movies have higher standard deviation means inconsistent ratings. It has low average IMDB score. It has the biggest count of movies made in it. So it has a wider range of IMDB scores indicating a mix of highly, moderately, and poorly rated movies. It is a globally accepted language for communication so people from all over the world tend to see it. These movies can be successful or flops.
4) Languages with less number of movies tend to have higher average IMDB ratings maybe due to the small size of sample data or small count of the movies.
5) Hindi movies have high standard deviation and low average IMDB scores. This maybe because Hindi movies are mostly watched in India, a country which has the largest population in the world. And additionally, some people from other countries also have watched it. So many people might have rated it differently.
6) Movies made in languages such as Portuguese, Korean, German, etc. have high average IMDB scores and low standard deviation which means consistent good ratings.
7) Movies made in languages such as Japanese, Spanish, Mandarin, etc., has good average IMDB scores and moderate standard deviation. They perform well but their consistency can be improved.

8) Movies made in languages such as Italian and Hindi have higher standard deviation which means that some movies can be great and some others can be poor ratings. It shows inconsistency.
9) Cantonese and French movies have fairly good average IMDB ratings and low standard deviation which means they tend to perform well and consistent in good IMDB ratings
10) Language alone doesn't determine high IMDB ratings and the success of a movie.

Five 'Whys' Approach:

1) Why do English movies tend to have higher IMDB ratings?

A. These movies get international reach making them more accessible to a wider audience leading to higher IMDB ratings.

2) Why does wider international reach leads to higher IMDB ratings?

A. It attracts viewers from all over the world making a way for more discussion, exposure, and recognition of the movie. It can lead to a larger and more engaged voter base on the IMDB platform resulting in higher IMDB ratings.

3) Why does larger voter base helps to maintain higher IMDB ratings?

A. It means more votes. More votes can create a balanced rating distribution minimizing the effects of extreme ratings and boosts reliability.

4) Why do some non-English movies tend to have varied IMDB ratings?

A. Due to limited distribution and language barriers,they attract niche audiences which leads to polarized ratings.

5) Why do some non-English films still achieve higher IMDB ratings?

A. They transcend language barriers and receive higher IMDB ratings if they have great story telling, universal themes, and positive reviews.

Recommendations for Filmmakers:

1) English is a commercially viable language because people from all over the world tend to see it as it is a globally accepted communicating language. Filmmakers should focus on the

quality over quantity of the English movies to avoid too many low IMDB ratings. Add elements which the top IMDB rated English movies have.

2) Other languages filmmakers should exhibit effective marketing strategies to gather global attention for their movies. They should make sure their movies have larger international visibility. Make the movies available on the streaming platforms so that people from all over the world can see it.

3) Even though the movies in other popular languages such as French, Spanish, Mandarin, German, Japanese, Hindi, Cantonese, Italian, Portuguese, Korean, etc., has fairly good IMDB ratings, some do get consistently good ratings and some get lower or moderate ratings. It can be inconsistent sometimes. So to improve or sustain the consistency in IMDB ratings, despite the language, make sure that the movies have great story-telling, cultural uniqueness and depth, good script, better production quality, global availability, well-execution, etc. This will help to make more people watch it and create good impression among the audience which leads to higher IMDB ratings.

D.) Director Analysis: Influence of directors on movie ratings.

Task: Identify the top directors based on their average IMDB score and analyze their contribution to the success of movies using percentile calculations.

Output:

movie_title	director_name	imdb_score
Avatar	James Cameron	7.9
Pirates of the Caribbean: At World's End	Gore Verbinski	7.1
Spectre	Sam Mendes	6.8
The Dark Knight Rises	Christopher Nolan	8.5
John Carter	Andrew Stanton	6.6
Spider-Man 3	Sam Raimi	6.2
Tangled	Nathan Greno	7.8
Avengers: Age of Ultron	Joss Whedon	7.5
Harry Potter and the Half-Blood Prince	David Yates	7.5
Batman v Superman: Dawn of Justice	Zack Snyder	6.9
Superman Returns	Bryan Singer	6.1
Quantum of Solace	Marc Forster	6.7
Pirates of the Caribbean: Dead Man's Chest	Gore Verbinski	7.3
The Lone Ranger	Gore Verbinski	6.5
Man of Steel	Zack Snyder	7.2
The Chronicles of Narnia: Prince Caspian	Andrew Adamson	6.6
The Avengers	Joss Whedon	8.1
Pirates of the Caribbean: On Stranger Tides	Rob Marshall	6.7
Men in Black 3	Barry Sonnenfeld	6.8
The Hobbit: The Battle of the Five Armies	Peter Jackson	7.5
The Amazing Spider-Man	Marc Webb	7
Robin Hood	Ridley Scott	6.7
The Hobbit: The Desolation of Smaug	Peter Jackson	7.9
The Golden Compass	Chris Weitz	6.1
King Kong	Peter Jackson	7.2
Titanic	James Cameron	7.7

I used Pivot table to find the unique directors, the number of movies they made, and the average IMDB scores of their movies.

director_name	Count of movie_title	Average of imdb_score
Charles Chaplin	1	8.6
Tony Kaye	1	8.6
Alfred Hitchcock	1	8.5
Damien Chazelle	1	8.5
Majid Majidi	1	8.5
Ron Fricke	1	8.5
Sergio Leone	3	8.433333333
Christopher Nolan	8	8.425
Asghar Farhadi	1	8.4
Marius A. Markevicius	1	8.4
Richard Marquand	1	8.4
S.S. Rajamouli	1	8.4
Billy Wilder	1	8.3
Fritz Lang	1	8.3
Lee Unkrich	1	8.3
Lenny Abrahamson	1	8.3
Pete Docter	3	8.233333333
Hayao Miyazaki	4	8.225
Elia Kazan	1	8.2
George Roy Hill	2	8.2
Joshua Oppenheimer	1	8.2
Juan José Campanella	1	8.2
Quentin Tarantino	8	8.2
Victor Fleming	2	8.15
Milos Forman	3	8.133333333
Akira Kurosawa	2	8.1

To find the Percentile, firstly, I found the highest average IMDB score using LARGE function. Then using PERCENTRANK function, I found the top directors are in the 99% percentile. Using that, I found the percentile using the PERCENTILE function.

Highest Average_imdb_score =LARGE(H2:H1748,1) =8.6

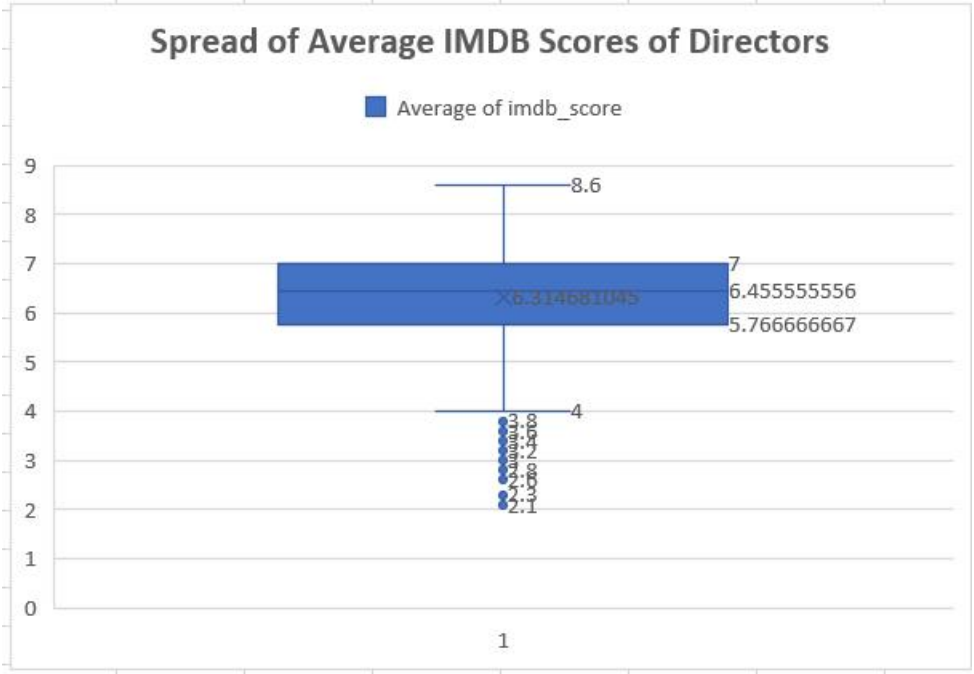
Percentrank =PERCENTRANK(H2:H1748,8.6) =0.999

Percentile =PERCENTILE(H2:H1748,K4) =8.524

Below are the Top 10 directors according to the Average IMDB scores of their movies. The top directors which belong to the 99% percentile are highlighted in green.

director_name	Count of movie_title	Average of imdb_score
Charles Chaplin	1	8.6
Tony Kaye	1	8.6
Alfred Hitchcock	1	8.5
Damien Chazelle	1	8.5
Majid Majidi	1	8.5
Ron Fricke	1	8.5
Sergio Leone	3	8.433333333
Christopher Nolan	8	8.425
Asghar Farhadi	1	8.4
Marius A. Markevicius	1	8.4

Below is the data visualization showing the Distribution of the Average IMDB scores of Directors.



Insights:

1) The directors Charles Chaplin and Tony Kaye belongs to the top 1%. They have the highest average IMDB scores.

2) The best and highly-rated directors consistently make high IMDB rated movies. They make many movies with higher IMDB ratings. They rarely make bad movies.

3) Directors like Christopher Nolan and Sergio Leone have high average IMDB scores and considering the number of movies they create, they often make high-rated movies.

4) The distribution of average IMDB scores of directors shows that most of the director's average IMDB scores fall between 5.7 and 7 which makes them the middle performers.

5) The directors who have average IMDB scores below 4 are very poor performers.

6) The directors having average IMDB scores greater than 8 are good performers. They belong to 95% percentile and above.

7) Directors from various countries have greater average IMDB scores including them in the top directors and showing international impact

8) Bad direction leads to low IMDB ratings

9) Poor direction, lack of vision, weak story-telling and poor execution, etc., of a movie lead to negative audience reception and thereby leads to lower IMDB ratings

Five 'Whys' Approach:

1) Why do movies by highly-rated directors tend to have higher IMDB ratings?

A. With their years of experience and mastery of excellent film-making, these directors make visually stunning movies with great quality and story telling, and emotionally engaging content. This leads to increase audience satisfaction and higher IMDB ratings.

2) Why do experience and great story telling skills lead to higher IMDB ratings?

A. Experienced directors know effective film-making techniques like how to create compelling narratives, control pacing, how to emotionally engage the audience, how to maintain the flow, etc., which leads to more positive reviews and IMDB ratings.

3) Why do positive reviews important for higher IMDB ratings?

A. A movie's overall IMDB score is greatly influenced by audience feedback. Higher IMDB ratings are the result of positive reviews from the audience who appreciated the film's direction, story telling and character development.

4) Why does audience feedback influence IMDB ratings?

A. IMDB ratings are a direct reflection of audience engagement and satisfaction. Viewers are more likely to leave positive ratings when they feel the movie was well-directed and captivated their attention and interest.

5) Why does poor direction of the movie leads to lower IMDB ratings?

A. Poor direction leads to lower IMDB ratings due to the problems in story telling, pacing and execution, making it frustrating for the people to watch. These can lead to negative reviews and thereby lead to lower IMDB ratings.

Recommendations for Filmmakers:

1) Directors alone don't determine the success of a movie and they have a great role in contributing to the success of a movie and IMDB ratings

2) Strong direction skills shows the great vision of the director that leads to great film-making process. It can help to make a movie successful and gain high IMDB ratings

3) A strong directorial debut with a unique directing style can help a filmmaker to create a good first impression, build strong reputation and make his/her mark in the history

4) Building a solid reputation as a director leads to long-term success and can influence audience expectations and reception.

For example; When the films of legendary director like Christopher Nolan releases, people expects the film to be excellent because they think it was Christopher Nolan who directed the film so the film can't be bad and they have an excitement and trust on Nolan that they believe the film would be great. So they go and watch the film without any second guess and rate the movie high.

5) Legendary directors are the ones who consistently make multiple movies with high quality and high IMDB ratings.

6) Modern directors can achieve legendary status through consistent quality and innovative film-making

7) A director should make sure that he elevates the movie's quality and maintain continuous excellence to become a highly-rated director.

8) New directors should learn from great and best directors and can improve their direction skills

9) International directors can compete with great Hollywood directors by focusing on great storytelling and quality

10) Even if a movie has good good story-telling, poor direction can lead the movie to failure

11) To make a movie highly successful and to get higher IMDB ratings, filmmakers should focus on strong direction, vision, quality script, great storytelling, production-quality, audience engagement and well execution

E.) Budget Analysis: Explore the relationship between movie budgets and their financial success.

Task: Analyze the correlation between movie budgets and gross earnings, and identify the movies with the highest profit margin.

Output:

gross	movie_title	budget	profit margi
760505847	Avatar	237000000	523505847
652177271	Jurassic World	150000000	502177271
658672302	Titanic	200000000	458672302
460935665	Star Wars: Episode IV - A New Hope	11000000	449935665
434949459	E.T. the Extra-Terrestrial	10500000	424449459
623279547	The Avengers	220000000	403279547
422783777	The Lion King	45000000	377783777
474544677	Star Wars: Episode I - The Phantom Menace	115000000	359544677
533316061	The Dark Knight	185000000	348316061
407999255	The Hunger Games	78000000	329999255
363024263	Deadpool	58000000	305024263
424645577	The Hunger Games: Catching Fire	130000000	294645577
356784000	Jurassic Park	63000000	293784000
368049635	Despicable Me 2	76000000	292049635
350123553	American Sniper	58800000	291323553
380838870	Finding Nemo	94000000	286838870
436471036	Shrek 2	150000000	286471036
377019252	The Lord of the Rings: The Return of the King	94000000	283019252
309125409	Star Wars: Episode VI - Return of the Jedi	32500000	276625409
329691196	Forrest Gump	55000000	274691196
290158751	Star Wars: Episode V - The Empire Strikes Back	18000000	272158751
285761243	Home Alone	18000000	267761243
380262555	Star Wars: Episode III - Revenge of the Sith	113000000	267262555
403706375	Spider-Man	139000000	264706375
336029560	Minions	74000000	262029560
293501675	The Sixth Sense	40000000	253501675

I found out the Profit Margin using the formula, PROFIT MARGIN = GROSS EARNINGS - BUDGET

The correlation between the movie budget and gross earnings is found out using the function CORREL.

Correlation Coefficient =0.1008

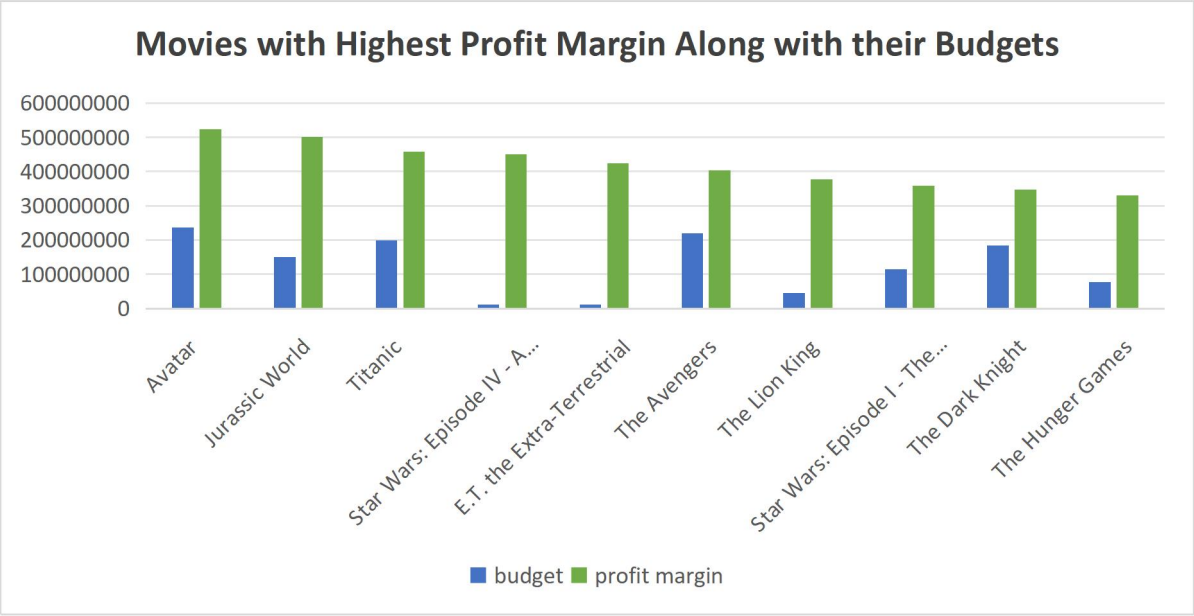
Avatar is the movie with the highest profit margin. It is found out using the MAX function.

Highest profit Margin =MAX(D2:D3850) =523505847 ==> Avatar

Below are the Top 10 movies according to their Profit Margin.

movie_title	profit margin
Avatar	523505847
Jurassic World	502177271
Titanic	458672302
Star Wars: Episode IV - A New Hope	449935665
E.T. the Extra-Terrestrial	424449459
The Avengers	403279547
The Lion King	377783777
Star Wars: Episode I - The Phantom Menace	359544677
The Dark Knight	348316061
The Hunger Games	329999255

Below is the data visualization showing the movies with highest profit margin along with their budgets.



Insights:

1) The correlation between the budget and gross earnings shows weak positive correlation. Movies with higher budgets tend to earn more and financially successful. But the relationship is not strong. This means spending high on budget does not guarantee a movie's box office success. Others factors can be influential. Budget can contribute to a movie's success but it alone doesn't determine a movie's commercial success.

2) The visualization given above shows the movies with highest profit margin along with their budgets. A movie is financially successful when the profit earned exceeds the overall budget spent. Even though Avatar is the movie with highest profit margin, it's budget was big compared to the other movies. And the profit it gained is double the budget expenses. Meanwhile movies like Star Wars and ET only required small budget but the profit gained was 40 times its budget. The Lion King and The Hunger Games also gained profit multiples times their budget. It means it received huge financial success.

3) Iconic and classic blockbuster movies like Titanic, ET, Star Wars and The Lion King tend to perform well over time. Decades after their release people still like to watch it.

4) Low-budget movies can outperform high-budget movies in profit margins.

5) Franchise movies like Star Wars, Avengers, Jurassic World, The Dark Knight, etc., gained huge profit every time the movie was released. Some franchise movies require heavy budget for movie production, completion, distribution and marketing. But they tend to receive huge box office success and profit.

6) If the first movie of a franchise is superhit blockbuster, people will have an anticipated excitement to watch the upcoming movies of the franchise. They tend to be successful. So brand power has a crucial role.

7) A well- executed low-budget movie can gain huge profit.

8) A low-budget well executed movie can be more profitable than a high-budget flop movie.

9) Smart budgeting is important as it can influence both IMDB ratings and financial success.

Five 'Whys' Approach:

1) Why do high-budget movies tend to be more financially successful?

A. Because it enables better production quality, visual effects, star power and great marketing, which helps in attracting larger audiences.

2) Why does better production quality and marketing attract larger audiences?

A. High production quality makes the movies aesthetically and visually pleasing, while well-known stars and effective marketing efforts generate hype, attracting more people to theatres.

3) Why does a larger audience leads to financial success?

A. More viewers translates into more ticket sales, more positive word-of-mouth, and better global reach, leading to higher box office earnings.

4) Why do some high-budget movies still flop?

A. No amount of budget can save a movie if it lacks a compelling story, good direction and audience appeal. Even movies with high budget flop if they fail to resonate with the audience.

5) Why do some low-budget movies gain massive financial success?

A. Because, instead of expensive special effects or A-list celebrities, they often rely on strong story telling, unique concepts, well-execution and effective marketing. This can capture audience attention and attract many people to watch the movies, leading to greater box office performances.

Recommendations for Filmmakers:

1) It doesn’t matter how much you spend for budgeting. It does not guarantee financial success.

2) If you are aiming to spend low or medium budget for the movie, make sure that you focus on great story telling, quality content, strong execution and audience engagement and connection. So that it becomes massively profitable.

3) No matter the budget is low or high, smart budgeting is important. Utilizing the budget efficiently can enhance the movie quality and viewing experience. A good budget combined with strong execution can maximizes the profit.

4) High budget can be beneficial as it helps in enhancing production quality, makes use of technological advancements, includes stunning visuals, better viewing or theatrical experience, completion, worldwide distribution and strong marketing. This can attract worldwide audience and can help in gaining huge financial success. But it cannot be always successful. High budget movies can flop too and lead to financial loss.

5) Despite the budget, the others factors that contribute to a movie's financial success are star power, great performances, right timing of the release, effective marketing, strong direction, audience appeal and interest, quality script and concept, great story telling, well-execution, etc.

6) Classy iconic films generate long-term profits. For example, If Titanic is released now worldwide, there will be many people going to theatres to relive that wonderful unique experience again. People also watch it on streaming platforms. So these type of films continue to earn more profit.

7) Franchise movies tend to dominate the box office. They have established and loyal fan bases. So it consistently earn higher profits making It a financially safer option for filmmakers.

8) Make the movie interesting to watch that can impress a large audience. It should have wonderful takeaways for the audience whether it is an enjoyable viewing experience, a feel of connection or any good message. This can increase audience satisfaction and can lead to success of the movie.

Final Conclusion:

Doing the project IMDB Movie Analysis has helped me gain a better understanding through a detailed analysis of different factors that influence the IMDB ratings and success of a movie. In this project, I did genre analysis, duration analysis, language analysis, director analysis, and budget analysis. From all these analyses, I came to know that one factor doesn't alone determine the IMDB ratings and success of a movie. The combination of these factors can lead to better results. The ratings can vary due to many factors and reasons. By these analyses, I could explore the patterns and trends and use appropriate data visualizations. I gained valuable insights from it. Genre analysis helped me to understand how genres impact IMDB ratings, which genre movies perform better and likely receiving frequently stable IMDB ratings and which genre movies do not. Duration analysis helped me to understand how duration impact IMDB ratings. Long movies tend to get higher ratings and short movies can perform well with an innovative approach. Language analysis helped me to understand how languages impact the IMDB ratings, which language movies perform better and how can both English and non-English movies gain high IMDB ratings. Director analysis made me understand the qualities, experience, expertise, tactics and characteristics of the top high-rated directors. They make high-rated movies consistently that build strong

reputation. I also understood how poor direction impact IMDB ratings. Budget analysis helped me to understand how budget influences a movie's financial success and understood that no matter the size of the budget, it can gain financial profit or loss based on how the audience respond to it. There are many other important factors that lead to higher IMDB ratings and financial success of a movie. A movie can be highly-rated or successful if it has great story telling, better script, outstanding performances, strong direction, audience connection, engagement and interest, audience appeal, good quality, great reach, marketing, better viewing experience, well-execution, right release timing, etc. The aim of these analyses is to improve the quality of the film-making, adding essential elements, providing valuable and preferable content to attract larger audience to see the movies which can lead to higher IMDB ratings and financial success. This project has helped me to gain a deeper understanding and insights about the relationship between various movie elements and their IMDB ratings which allows me to make informed decisions or recommendations to help filmmakers or streaming platforms regarding the predicting factors of a movie's success. This project helped me to understand how data analysis is useful for film or entertainment industry.

Project Report made by: Sweena Tony

Link to Excel Analysis