



IMDB Movie Analysis



by Kiran Kumar



Project Description

As a data analyst we have to analyze what factors influence the success of a movie on IMDB Score or Ratings. The impact of this problem is significant for movie producers, directors, and investors who want to understand what makes a movie successful to make informed decisions in their future projects.



Approach

In the given dataset there were so many duplicates, null cells and movie names were having special characters, there was no proper alignment. So first we cleaned the data for getting right result. The following steps were followed:

- There were 45 duplicates which were removed
- The cells with null values were removed
- Cleaned movie title names which were having special characters
- Removed unwanted columns which were unnecessary before it was 27, after cleaning we have 10 columns
- Before cleaning the rows were 5044 and after cleaning the rows were 3724

So after cleaning the data we moved forward for our analysis

Before Cleaning

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC
	color	director_r_num_criti	duration	director_f	actor_3_f	actor_2_n	actor_1_f	gross	genres	actor_1_n	movie_tit	num_vote	total	actor_3_n	facenum	plot_keyw	movie_im	num_user	language	country	content_r	budget	title_year	actor_2_f	imdb_sco	aspect_ra	movie_face	book_like	
1	Color	James Car	723	178	0	855	Joel David	1000	7.61E+08	Action	Ac CCH Pouni	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	
2	Color	Gore Ver	302	169	563	1000	Orlando B	40000	3.09E+08	Action	Ac Johnny De	Pirates of	471220	48350	Jack Dave	0	goddess	http://www	1238	English	USA	PG-13	3E+08	2007	5000	7.1	2.35	0	
3	Color	Sam Menc	602	148	0	161	Rory Kinni	11000	2E+08	Action	Ac Christoph	Spectre	275868	11700	Stephanie	1	bomb es	http://www	994	English	UK	PG-13	2.45E+08	2015	393	6.8	2.35	85000	
4	Color	Christoph	813	164	22000	23000	Christian I	27000	4.48E+08	Action	Th Tom Hard	The Dark k	1144337	106759	Joseph Gc	0	deception	http://www	2701	English	USA	PG-13	2.5E+08	2012	23000	8.5	2.35	164000	
5		Doug Walker			131		Rob Walk	131		Document	Doug Wall	Star Wars	8	143		0	http://www						12	7.1		0			
6	Color	Andrew S	462	132	475	530	Samantha	640	73058679	Action	Ac Daryl Sab	John Carte	212204	1873	Polly Wall	1	alien ame	http://www	738	English	USA	PG-13	2.64E+08	2012	632	6.6	2.35	24000	
7	Color	Sam Raim	392	156	0	4000	James Fra	24000	3.37E+08	Action	Ac J.K. Simm	Spider-Me	383056	46055	Kirsten Du	0	sandman	http://www	1902	English	USA	PG-13	2.58E+08	2007	11000	6.2	2.35	0	
8	Color	Nathan Gr	324	100	15	284	Donna Mu	799	2.01E+08	Adventuri	Brad Garre	Tangled	294810	2036	M.C. Gain	1	17th cent	http://www	387	English	USA	PG	2.6E+08	2010	553	7.8	1.85	29000	
9	Color	Joss Whec	635	141	0	19000	Robert Do	26000	4.59E+08	Action	Ac Chris Hem	Avengers	462669	92000	Scarlett Jc	4	artificial i	http://www	1117	English	USA	PG-13	2.5E+08	2015	21000	7.5	2.35	118000	
10	Color	David Yate	375	153	282	10000	Daniel Ra	25000	3.02E+08	Adventuri	Alan Rickr	Harry Pott	321795	58753	Rupert Gri	3	blood bo	http://www	973	English	UK	PG	2.5E+08	2009	11000	7.5	2.35	10000	
11	Color	Zack Snyder	673	183	0	2000	Lauren Co	15000	3.3E+08	Action	Ac Henry Cav	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	
12	Color	Bryan Sing	434	169	0	903	Marlon Br	18000	2E+08	Action	Ac Kevin Spa	Superman	240396	29991	Frank Lang	0	crystal ep	http://www	2367	English	USA	PG-13	2.09E+08	2006	10000	6.1	2.35	0	
13	Color	Marc Fors	403	106	395	393	Mathieu A	451	1.68E+08	Action	Ac Giancarlo	Quantum	330784	2023	Rory Kinn	1	action her	http://www	1243	English	UK	PG-13	2E+08	2008	412	6.7	2.35	0	
14	Color	Gore Ver	313	151	563	1000	Orlando B	40000	4.23E+08	Action	Ac Johnny De	Pirates of	522040	48486	Jack Dave	2	box office	http://www	1832	English	USA	PG-13	2.25E+08	2006	5000	7.3	2.35	5000	
15	Color	Gore Ver	450	150	563	1000	Ruth Wils	40000	89289910	Action	Ac Johnny De	The Lone l	181792	45757	Tom Wilki	1	horse ou	http://www	711	English	USA	PG-13	2.15E+08	2013	2000	6.5	2.35	48000	
16	Color	Zack Snyder	733	143	0	748	Christoph	15000	2.91E+08	Action	Ac Henry Cav	Man of Ste	548573	20495	Harry Lenr	0	based on	http://www	2536	English	USA	PG-13	2.25E+08	2013	3000	7.2	2.35	118000	
17	Color	Andrew A	258	150	80	201	Pierfrance	22000	1.42E+08	Action	Ac Peter Dinl	The Chron	149922	22697	Dami	4	brother br	http://www	438	English	USA	PG	2.25E+08	2008	216	6.6	2.35	0	
18	Color	Joss Whec	703	173	0	19000	Robert Do	26000	6.23E+08	Action	Ac Chris Hem	The Aven	995415	87697	Scarlett Jc	3	alien inva	http://www	1722	English	USA	PG-13	2.2E+08	2012	21000	8.1	1.85	123000	
19	Color	Rob Marsh	448	136	252	1000	Sam Clafl	40000	2.41E+08	Action	Ac Johnny De	Pirates of	370704	54083	Stephen G	4	blackbear	http://www	484	English	USA	PG-13	2.5E+08	2011	11000	6.7	2.35	58000	
20	Color	Barry Soni	451	106	188	718	Michael Si	10000	1.79E+08	Action	Ac Will Smith	Men in Bl	268154	12572	Nicole Sc	1	alien crin	http://www	341	English	USA	PG-13	2.25E+08	2012	816	6.8	1.85	40000	
21	Color	Peter Jack	422	164	0	773	Adam Bro	5000	2.55E+08	Adventuri	Aidan Turi	The Hobbi	354228	9152	James Ne	0	army elf	http://www	802	English	New Zeal	PG-13	2.5E+08	2014	972	7.5	2.35	65000	
22	Color	Marc Webb	599	153	464	963	Andrew G	15000	2.62E+08	Action	Ac Emma Sto	The Amaz	451803	28489	Chris Zylk	0	lizard out	http://www	1225	English	USA	PG-13	2.3E+08	2012	10000	7	2.35	56000	
23	Color	Ridley Scc	343	156	0	738	William H	891	1.05E+08	Action	Ac Mark Add	Robin Hoc	211765	3244	Scott Grim	0	1190s ar	http://www	546	English	USA	PG-13	2E+08	2010	882	6.7	2.35	17000	
24	Color	Peter Jack	509	186	0	773	Adam Bro	5000	2.58E+08	Adventuri	Aidan Turi	The Hobbi	483540	9152	James Ne	6	dwarf elf	http://www	951	English	USA	PG-13	2.25E+08	2013	972	7.9	2.35	83000	
25	Color	Chris Wei	251	113	129	1000	Eva Green	16000	70083519	Adventuri	Christoph	The Golde	149019	24106	Kristin Sc	2	children	http://www	666	English	USA	PG-13	1.8E+08	2007	6000	6.1	2.35	0	
26	Color	Peter Jack	446	201	0	84	Thomas Ki	6000	2.18E+08	Action	Ac Naomi Wa	King Kong	316018	7123	Evan Park	0	animal na	http://www	2618	English	New Zeal	PG-13	2.07E+08	2005	919	7.2	2.35	0	
27	Color	James Car	315	194	0	794	Kate Wins	29000	6.59E+08	Drama	Rc Leonardo	Titanic	793059	45223	Gloria Stu	0	artist lov	http://www	2528	English	USA	PG-13	2E+08	1997	14000	7.7	2.35	26000	
28	Color	Anthony F	516	147	94	11000	Scarlett Jc	21000	4.07E+08	Action	Ac Robert Do	Captain Ai	272670	64798	Chris Evan	0	based on	http://www	1022	English	USA	PG-13	2.5E+08	2016	19000	8.2	2.35	72000	
29	Color	Peter Berg	377	131	532	627	Alexande	14000	65173160	Action	Ac Liam Nee	Battleship	202382	26679	Tadanobu	0	box office	http://www	751	English	USA	PG-13	2.09E+08	2012	10000	5.9	2.35	44000	
30	Color	Colin Trev	644	124	365	1000	Judy Gree	3000	6.52E+08	Action	Ac Bryce Dall	Jurassic W	418214	8458	Omar Sy	0	dinosaur	http://www	1290	English	USA	PG-13	1.5E+08	2015	2000	7	2	150000	
31	Color	Sam Menc	750	143	0	393	Helen Mcr	883	3.04E+08	Action	Ac Albert Fin	Skyfall	522030	2039	Rory Kinni	0	brawl chi	http://www	1498	English	UK	PG-13	2E+08	2012	563	7.8	2.35	80000	
32	Color	Sam Raim	300	135	0	4000	James Fra	24000	3.73E+08	Action	Ac J.K. Simm	Spider-Me	411164	43388	Kirsten Du	1	death doi	http://www	1303	English	USA	PG-13	2E+08	2004	11000	7.3	2.35	0	
33	Color	Shane Bla	608	195	1000	3000	Jon Favre	21000	4.09E+08	Action	Ac Robert Do	Iron Man	557489	30426	Don Chea	3	armor ex	http://www	1187	English	USA	PG-13	2E+08	2013	4000	7.2	2.35	95000	
34	Color	Tim Burto	451	108	13000	11000	Alan Rickr	40000	3.34E+08	Adventuri	Johnny De	Alice in W	306320	79957	Anne Hat	0	alice in w	http://www	736	English	USA	PG	2E+08	2010	25000	6.5	1.85	24000	
35	Color	Brett Ratn	334	104	420	560	Kelsey Gri	20000	2.34E+08	Action	Ac Hugh Jack	X-Men: Th	383427	21714	Daniel Cu	0	battle mu	http://www	1912	English	Canada	PG-13	2.1E+08	2006	808	6.8	2.35	0	
36	Color	Dan Scanl	376	104	37	760	Tyler Labi	12000	2.68E+08	Adventuri	Steve Bus	Monsters	235025	14863	Sean Haye	0	cheating	http://www	265	English	USA	G	2E+08	2013	779	7.3	1.85	44000	
37	Color	Michael B	366	150	0	464	Kevin Dun	894	4.02E+08	Action	Ac Glenn Mo	Transform	323207	3218	Ramon Ro	0	autobot	http://www	1439	English	USA	PG-13	2E+08	2009	581	6	2.35	0	
38	Color	Michael B	378	165	0	808	Sophia My	974	2.45E+08	Action	Ac Bingbing L	Transform	242420	3988	Kelsey Gri	2	blockbust	http://www	918	English	USA	PG-13	2.1E+08	2014	956	5.7	2.35	56000	
39	Color	Sam Raim	535	130	0	11000	Mila Kuv	14000	2.35E+08	Adventuri	Tim Hol	On the Gr	175400	72441	James Fra	4	circus ma	http://www	511	English	USA	PG	2.15E+08	2013	15000	6.4	2.35	50000	

After Cleaning

	A	B	C	D	E	F	G	H	I	J
1	director_name	duration	gross	genres	movie_title	language	country	budget	title_year	imdb_score
2	James Cameron	178	760505847	Action	Avatar	English	USA	237000000	2009	7.9
3	Gore Verbinski	169	309404152	Action	Pirates of the Caribbean: At World's End	English	USA	300000000	2007	7.1
4	Sam Mendes	148	200074175	Action	Spectre	English	UK	245000000	2015	6.8
5	Christopher Nolan	164	448130642	Action	The Dark Knight Rises	English	USA	250000000	2012	8.5
6	Andrew Stanton	132	73058679	Action	John Carter	English	USA	263700000	2012	6.6
7	Sam Raimi	156	336530303	Action	Spider-Man 3	English	USA	258000000	2007	6.2
8	Nathan Greno	100	200807262	Adventure	Tangled	English	USA	260000000	2010	7.8
9	Joss Whedon	141	458991599	Action	Avengers: Age of Ultron	English	USA	250000000	2015	7.5
0	David Yates	153	301956980	Adventure	Harry Potter and the Half-Blood Prince	English	UK	250000000	2009	7.5
1	Zack Snyder	183	330249062	Action	Batman v Superman: Dawn of Justice	English	USA	250000000	2016	6.9
2	Bryan Singer	169	200069408	Action	Superman Returns	English	USA	209000000	2006	6.1
3	Marc Forster	106	168368427	Action	Quantum of Solace	English	UK	200000000	2008	6.7
4	Gore Verbinski	151	423032628	Action	Pirates of the Caribbean: Dead Man's Chest	English	USA	225000000	2006	7.3
5	Gore Verbinski	150	89289910	Action	The Lone Ranger	English	USA	215000000	2013	6.5
6	Zack Snyder	143	291021565	Action	Man of Steel	English	USA	225000000	2013	7.2
7	Andrew Adamson	150	141614023	Action	The Chronicles of Narnia: Prince Caspian	English	USA	225000000	2008	6.6
8	Joss Whedon	173	623279547	Action	The Avengers	English	USA	220000000	2012	8.1
9	Rob Marshall	136	241063875	Action	Pirates of the Caribbean: On Stranger Tides	English	USA	250000000	2011	6.7
0	Barry Sonnenfeld	106	179020854	Action	Men in Black 3	English	USA	225000000	2012	6.8
1	Peter Jackson	164	255108370	Adventure	The Hobbit: The Battle of the Five Armies	English	New Zealand	250000000	2014	7.5
2	Marc Webb	153	262030663	Action	The Amazing Spider-Man	English	USA	230000000	2012	7
3	Ridley Scott	156	105219735	Action	Robin Hood	English	USA	200000000	2010	6.7
4	Peter Jackson	186	258355354	Adventure	The Hobbit: The Desolation of Smaug	English	USA	225000000	2013	7.9

Tech Stack Used

In this analytics process we used MS- Excel 2021 for analyzing the dataset and for presentation we used MS-Powerpoint 2021



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.

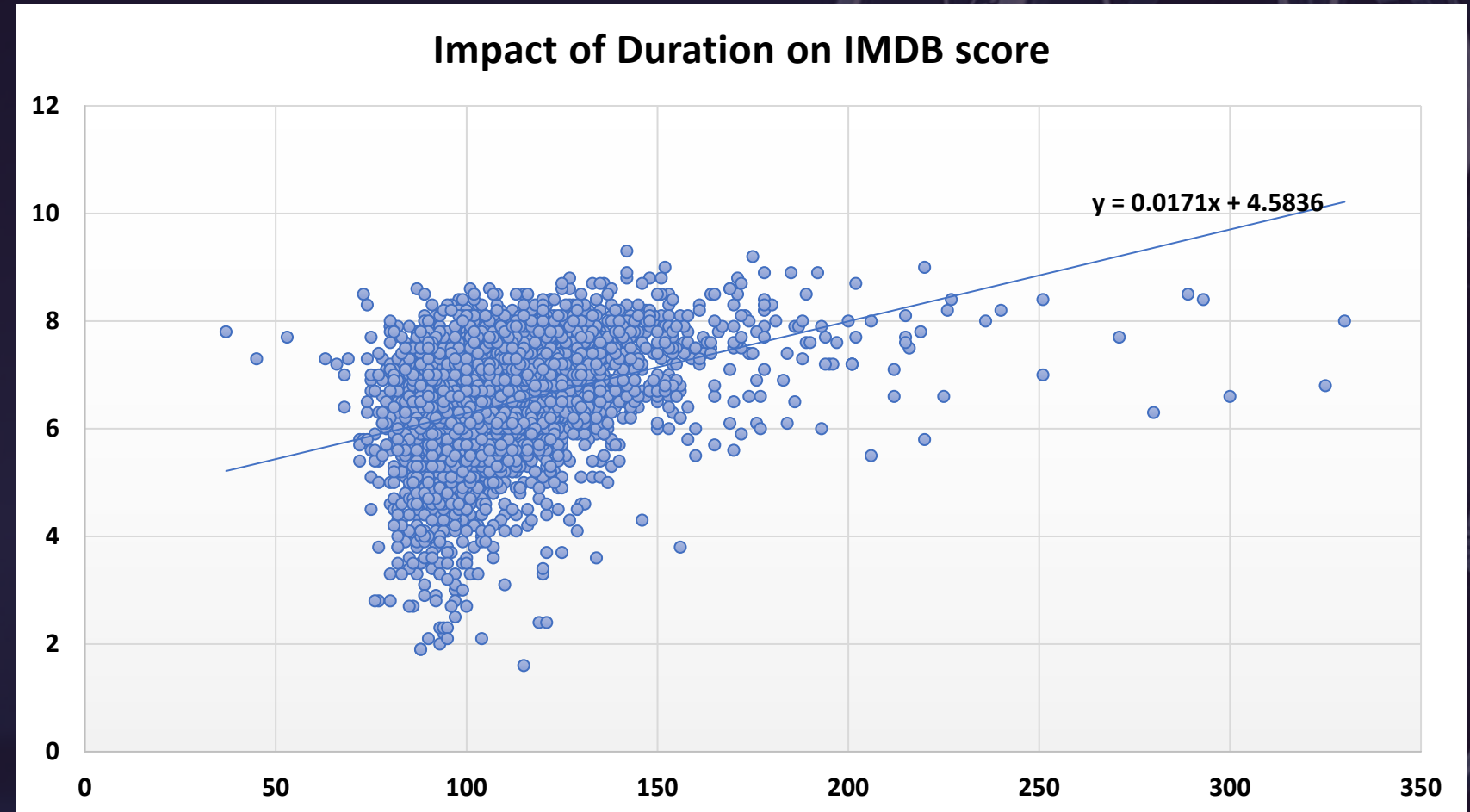
Then, for each genre, calculate descriptive statistics (mean, median, mode, range, variance, standard deviation) of the IMDB scores.

Unique Genres	Total Genres	MEAN(Average)	MEDIAN	MODE	MAX	MIN	RANGE	VARIANCE	STD DEV.
Comedy	984	6.17	6.30	6.40	8.80	1.90	6.90	1.05	1.03
Action	951	6.29	6.30	6.10	9.00	2.10	6.90	1.06	1.03
Drama	659	6.83	6.90	6.70	8.80	2.10	6.70	0.82	0.91
Adventure	366	6.56	6.70	7.30	8.60	2.30	6.30	1.26	1.12
Crime	253	6.94	7.00	7.30	9.30	3.30	6.00	0.75	0.87
Biography	204	7.16	7.20	7.40	8.90	4.50	4.40	0.48	0.70
Horror	159	5.84	5.90	6.70	8.50	2.30	6.20	1.05	1.03
Animation	45	6.74	7.00	7.10	8.00	4.50	3.50	0.92	0.96
Fantasy	37	6.28	6.50	6.80	7.90	4.30	3.60	0.78	0.88
Documentary	26	6.80	7.45	7.50	8.50	1.60	6.90	2.83	1.68
Mystery	23	6.65	6.70	7.10	8.50	3.30	5.20	1.14	1.07
Sci-Fi	7	6.63	6.40	6.70	8.20	5.00	3.20	1.05	1.02
Family	3	6.50	5.90	6.70	7.90	5.70	2.20	0.99	0.99
Western	2	8.10	8.10	6.70	8.90	7.30	1.60	0.64	0.80
Musical	2	6.75	6.75	6.70	7.20	6.30	0.90	0.20	0.45
Romance	1	7.10	7.10	6.70	7.10	7.10	0.00	0.00	0.00
Thriller	1	4.80	4.80	6.70	4.80	4.80	0.00	0.00	0.00

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

Duration of Movie	
Average	110.26
Median	106.00
Stdev	22.68



- Here the Trendline is going upwards means, longer length movies are getting good IMDB Score, so there is a positive relation for movie duration and IMDB score.

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.

There are total 34 languages movies in the given Dataset, among them we have taken top 10 languages movies which are very high in number. These are:

Languages	Total Movies	Mean	Median	Stdev
English	3566	6.43	6.50	1.05
French	34	7.36	7.30	0.51
Spanish	23	7.08	7.20	0.84
Mandarin	14	7.02	7.25	0.74
Japanese	10	7.66	8.00	0.94
German	10	7.77	7.80	0.68
Cantonese	7	7.34	7.30	0.32
Italian	7	7.19	7.00	1.07
Korean	5	7.70	7.70	0.51
Hindi	5	7.22	7.40	0.72

- Among all the languages “English” is the most common language used in movies
- The average IMDB score of English is 6.43 and all other languages are mentioned in the table
- The Highest average IMDB score is language is “German”

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.

There are total 1659 Directors in the given dataset after, among them we have taken top 10 directors whose IMDB score is very high and movies are above 10. They are:

Director	Total Movies	Avg. IMDB
Martin Scorsese	15	7.68
Steven Spielberg	24	7.54
Richard Linklater	11	7.33
Clint Eastwood	19	7.21
Rob Reiner	11	7.02
Woody Allen	19	7.00
Oliver Stone	12	6.91
John Carpenter	12	6.83
Spike Lee	15	6.73
Steven Soderbergh	13	6.71
Barry Levinson	11	6.58
Wes Craven	11	6.11

The contribution of Directors to the success of movies is 75% (Done using percentile functions)

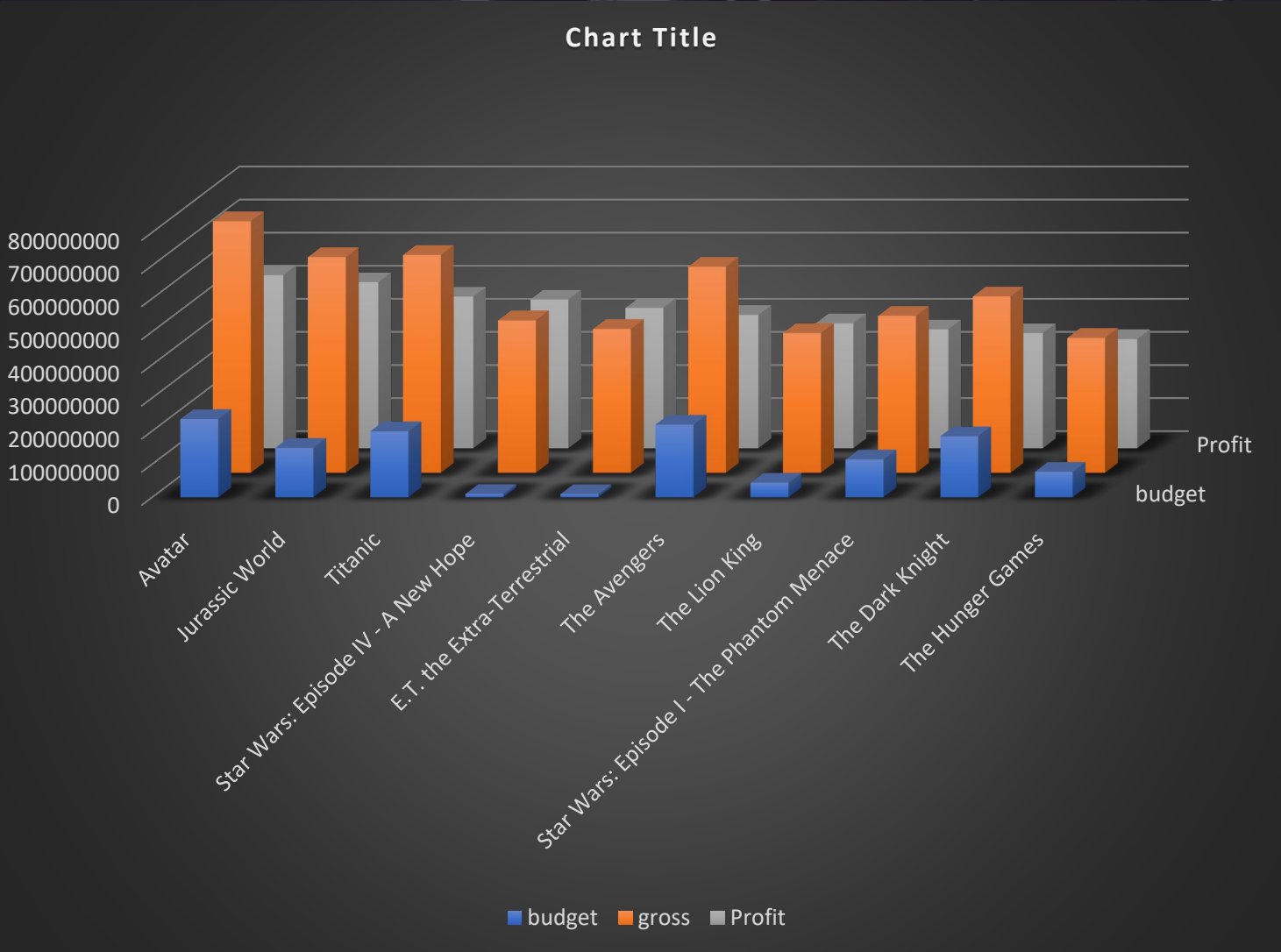
E. Budget Analysis: Explore the relationship between movie budgets and their financial success.

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

The Correlation between Movie budget and gross earnings is 0.0983(so it is mostly on the positive side), with this we can see that as budget increases it will bring profits to the movie.

The top 10 with highest profit margin are:

movie_title	budget	gross	Profit	Highest Profit	Highest profit Movie
Avatar	237000000	760505847	523505847	523505847	Avatar
Jurassic World	150000000	652177271	502177271		
Titanic	200000000	658672302	458672302		
Star Wars: Episode IV - A New Hope	11000000	460935665	449935665		
E.T. the Extra-Terrestrial	10500000	434949459	424449459		
The Avengers	220000000	623279547	403279547		
The Lion King	45000000	422783777	377783777		
Star Wars: Episode I - The Phantom Menace	115000000	474544677	359544677		
The Dark Knight	185000000	533316061	348316061		
The Hunger Games	78000000	407999255	329999255		





Result

The purpose of this presentation is to analyze IMDB movie ratings and identify what factors influence them. By the end of this presentation, you will have a better understanding of what makes a good movie. This project made me think as an analyst and do the work on it to get the desired solutions. Thank you for giving me this opportunity to explore myself to understand the Data Analytics.

Note: You can checkout my working file here

<https://docs.google.com/spreadsheets/d/18cPMhpffAc86F3kcAXK91hnSNoOtxMoKgs4fSe8XAcg/edit?usp=sharing>

THANK YOU