

*IMDb Movie Analysis  
Final Project-1*

**IMDb**

**SURAJ OD**

# PROJECT DESCRIPTION

- *The goal of this project is to analyze a dataset containing various columns of different IMDB movies.*
- *The dataset needs to be cleaned and prepared for analysis. Once the data cleaning is done, the following tasks need to be performed.*
- *This project involves cleaning a dataset of IMDB movies and performing various analyses on it.*
- *Data Analysis is the data will be thoroughly explored to discern relationships between various variables.*
- *An emphasis will be placed on examining correlations between movie ratings and other factors such as genre, director, budget, actors, year of release, and more.*
- *Understanding these connections can provide valuable insights into what drives high ratings.*



# APPROACH

- ❖ **Data Cleaning:-** Drop unnecessary columns that are not relevant to the analysis.
  - Remove rows with missing or null values to ensure data integrity.
- ❖ **Data Preprocessing:-** Load the movie dataset into Excel.
  - Ensure that the columns for genres and IMDB scores are in the appropriate format and do any necessary data cleaning.
- ❖ **Determine Most Common Genres:** - Use Excel's COUNTIF function to count the occurrences of each genre in the dataset.
  - Sort or filter the genre count to identify the most common genres.
- ❖ **Movie Duration Analysis :-** Calculate descriptive statistics (mean, median, standard deviation) for movie durations using Excel's functions.
  - Create a scatter plot to visualize the relationship between movie duration and IMDB score.
  - Add a trendline to assess the direction and strength of the relationship.
- ❖ **Language Analysis (Task C):** - Use Excel's COUNTIF function to count occurrences of each language in the dataset.
  - Determine the most common languages based on the count.
  - Calculate descriptive statistics (mean, median, standard deviation) of IMDB scores for each language.
- ❖ **Director Analysis (Task D):-** Calculate the average IMDB score for each director.
  - Use Excel's PERCENTILE function to identify directors with the highest scores.
  - Compare the scores of top directors to the overall distribution of scores.
- **Budget Analysis (Task E):-** Calculate the correlation coefficient between movie budgets and gross earnings using Excel's CORREL function.
- Calculate the profit margin (gross earnings - budget) for each movie.

## TECH-STACK USED

- **Microsoft Excel** :- In this tech stack I used for data cleaning, analysing and data visualization.
- **Microsoft PowerPoint** :- In this tech stack I used for presenting the data and reporting my Project. **Tableau** as well used.

# INSIGHTS

- ☐ Cleaning the data.
- ☐ Data Analysis.
- ☐ Genre Impact on Ratings.
- ☐ Movie Duration and Ratings Relationship.
- ☐ Language and Ratings.
- ☐ Director Influence.
- ☐ Budget and Financial Success.

# ➤ CLEANING THE DATA

## ❖ Steps for clean data :-

1. Identify the data to be cleaned.
2. Make a backup of the original data.
3. Remove unnecessary spaces using the TRIM function or Find and Replace.
4. Remove duplicate values using the built-in "Remove Duplicates" feature.
5. Correct inconsistent data formatting or naming conventions.
6. Handle missing data by deleting rows.
7. Format data appropriately based on its type (numbers, dates, text).
8. Save the cleaned data.

A	B	C	D	E	F	G	H	I
color	director_name	num_critic_for_review	duration	director_facebook_likes	actor_3_facebook_likes	actor_2_name	actor_1_facebook_likes	gross
Color	James Cameron	723	178	0	855	Joel David Moore	1000	76050584
Color	Colin Trevorrow	644	124	365	1000	Judy Greer	3000	65217727
Color	James Cameron	315	194	0	794	Kate Winslet	29000	65867230
Color	George Lucas	282	125	0	504	Peter Cushing	11000	46093566
Color	Steven Spielberg	215	120	14000	548	Dee Wallace	861	43494945
Color	Joss Whedon	703	173	0	19000	Robert Downey Jr.	26000	62327954
Color	Roger Allers	186	73	28	847	Nathan Lane	2000	42278377
Color	George Lucas	320	136	0	1000	Liam Neeson	20000	47454467
Color	Christopher Nolan	645	152	22000	11000	Heath Ledger	23000	53331606
Color	Gary Ross	673	142	378	575	Josh Hutcherson	34000	40799925
Color	Tim Miller	579	108	84	361	Ed Skrein	16000	36302426
Color	Francis Lawrence	502	146	508	523	Josh Hutcherson	34000	42464557
Color	Steven Spielberg	308	127	14000	191	Ariana Richards	967	35678400
Color	Pierre Coffin	306	98	275	1000	Miranda Cosgrove	7000	36804963
Color	Clint Eastwood	490	133	16000	318	Leonard Roberts	14000	35012355
Color	Andrew Stanton	301	100	475	799	Stephen Root	1000	38083887
Color	Andrew Adamson	205	93	80	48	Jennifer Saunders	692	43647103
Color	Peter Jackson	328	192	0	416	Billy Boyd	5000	37701925
Color	Richard Marquand	197	134	37	504	Ian McDiarmid	11000	30912540
Black and	Robert Zemeckis	149	142	0	194	Siobhan Fallon Hogan	15000	32969119



S	T	U	V	W	X	Y	Z	AA	AB
num_user_for_reviews	language	country	content_rating	budget	title_year	actor_2_facebook_likes	imdb_score	aspect_ratio	movie_facebook_likes
3054	English	USA	PG-13	237000000	2009	936	7.9	1.78	33000
1290	English	USA	PG-13	150000000	2015	2000	7	2	150000
2528	English	USA	PG-13	200000000	1997	14000	7.7	2.35	26000
1470	English	USA	PG	11000000	1977	1000	8.7	2.35	33000
515	English	USA	PG	10500000	1982	725	7.9	1.85	34000
1722	English	USA	PG-13	220000000	2012	21000	8.1	1.85	123000
656	English	USA	G	45000000	1994	886	8.5	1.66	17000
3597	English	USA	PG	115000000	1999	14000	6.5	2.35	13000
4667	English	USA	PG-13	185000000	2008	13000	9	2.35	37000
1959	English	USA	PG-13	78000000	2012	14000	7.3	2.35	140000
1058	English	USA	R	58000000	2016	805	8.1	2.35	117000
706	English	USA	PG-13	130000000	2013	14000	7.6	2.35	82000
895	English	USA	PG-13	63000000	1993	610	8.1	1.85	19000
284	English	USA	PG	76000000	2013	2000	7.5	1.85	56000
916	English	USA	R	58800000	2014	962	7.3	2.35	112000
866	English	USA	G	94000000	2003	939	8.2	1.85	11000
483	English	USA	PG	150000000	2004	309	7.2	1.85	0
3189	English	USA	PG-13	94000000	2003	857	8.9	2.35	16000
647	English	USA	PG	32500000	1983	1000	8.4	2.35	14000
1398	English	USA	PG-13	55000000	1994	294	8.8	2.35	59000

SQL IMDB									
Accessibility: Unavailable									
Wes Studi	0	avatar future marine native paraplegic	http://www.imdb.com/title/tt0499549/?ref_=fn_tt_tt_1	Action Adventure Fantasy Sci-Fi	CCH Pounder	Avatar	886204	4834	
Omar Sy	0	dinosaur disaster film experiment gone wrong jurassic park velociraptor	http://www.imdb.com/title/tt0369610/?ref_=fn_tt_tt_1	Action Adventure Sci-Fi Thriller	Bryce Dallas Howard	Jurassic World	418214	8458	
Gloria Stuart	0	artist love ship titanic wet	http://www.imdb.com/title/tt0120338/?ref_=fn_tt_tt_1	Drama Romance	Leonardo DiCaprio	Titanic	793059	45223	
Kenny Baker	1	death star empire galactic war princess rebellion	http://www.imdb.com/title/tt0076759/?ref_=fn_tt_tt_1	Action Adventure Fantasy Sci-Fi	Harrison Ford	Star Wars: Episode IV - A New Hope	911097	13485	
Peter Coyote	0	bicyclist boy doll star wars reference voice impersonation	http://www.imdb.com/title/tt0083866/?ref_=fn_tt_tt_1	Family Sci-Fi	Henry Thomas	E.T. the Extra-Terrestrial	281842	2811	
Scarlett Johansson	3	alien invasion assassin battle iron man soldier	http://www.imdb.com/title/tt0848228/?ref_=fn_tt_tt_1	Action Adventure Sci-Fi	Chris Hemsworth	The Avengers	995415	87697	
Niketa Calame	0	king prince scar uncle unnecessary guilt	http://www.imdb.com/title/tt0110357/?ref_=fn_tt_tt_1	Adventure Animation Drama Family Musical	Matthew Broderick	The Lion King	644348	6458	
Ian McDiarmid	1	alien character says i have a bad feeling about this hand to hand combat seven word title wilhelm scream	http://www.imdb.com/title/tt0120915/?ref_=fn_tt_tt_1	Action Adventure Fantasy Sci-Fi	Natalie Portman	Star Wars: Episode I - The Phantom Menace	534658	37723	
Morgan Freeman	0	based on comic book dc comics psychopath star died before release urban setting	http://www.imdb.com/title/tt0468569/?ref_=fn_tt_tt_1	Action Crime Drama Thriller	Christian Bale	The Dark Knight	1676169	57802	
Anthony Reynolds	0	fight to the death game massacre self survival suicide attempt	http://www.imdb.com/title/tt1392170/?ref_=fn_tt_tt_1	Adventure Drama Sci-Fi Thriller	Jennifer Lawrence	The Hunger Games	701607	49942	
Stefan Kapicic	0	based on comic book breaking the fourth wall mercenary mutant self healing	http://www.imdb.com/title/tt1431045/?ref_=fn_tt_tt_1	Action Adventure Comedy Romance Sci-Fi	Ryan Reynolds	Deadpool	479047	18239	
Sandra Ellis Lafferty	1	president snow television tour uprising	http://www.imdb.com/title/tt1951264/?ref_=fn_tt_tt_1	Adventure Sci-Fi Thriller	Jennifer Lawrence	The Hunger Games: Catching Fire	498397	49355	
Bob Peck	0	dinosaur jurassic park survival theme park tyrannosaurus rex	http://www.imdb.com/title/tt0107290/?ref_=fn_tt_tt_1	Adventure Sci-Fi Thriller	Wayne Knight	Jurassic Park	613473	2129	
Steve Coogan	0	dating minion overprotective father undercover spy woman agent	http://www.imdb.com/title/tt1690953/?ref_=fn_tt_tt_1	Animation Comedy Family Sci-Fi	Steve Carell	Despicable Me 2	286877	11905	
Keir O'Donnell	0	assassin death of child iraq sniper sniper rifle	http://www.imdb.com/title/tt2179136/?ref_=fn_tt_tt_1	Action Biography Drama History Thriller War	Bradley Cooper	American Sniper	325264	16277	
Brad Garrett	0	great barrier reef protective father separation from family shark short term memory loss	http://www.imdb.com/title/tt0266543/?ref_=fn_tt_tt_1	Adventure Animation Comedy Family	Alexander Gould	Finding Nemo	692482	5641	
Conrad Vernon	1	drag queen physical appearance princess queen sequel	http://www.imdb.com/title/tt0298148/?ref_=fn_tt_tt_1	Adventure Animation Comedy Family Fantasy Romance	Rupert Everett	Shrek 2	314630	1148	
Bernard Hill	2	battle epic king orc ring	http://www.imdb.com/title/tt0167260/?ref_=fn_tt_tt_1	Action Adventure Drama Fantasy	Orlando Bloom	The Lord of the Rings: The Return of the King	1215718	6434	
Kenny Baker	0	death star emperor empire princess rebel	http://www.imdb.com/title/tt0086190/?ref_=fn_tt_tt_1	Action Adventure Fantasy Sci-Fi	Harrison Ford	Star Wars: Episode VI - Return of the Jedi	681857	13649	
Sam Anderson	0	amputee love vietnam vietnam war war hero	http://www.imdb.com/title/tt0109830/?ref_=fn_tt_tt_1	Comedy Drama	Tom Hanks	Forrest Gump	1251222	15700	

# ❖ MOVIE GENRE ANALYSIS:-

**your 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.

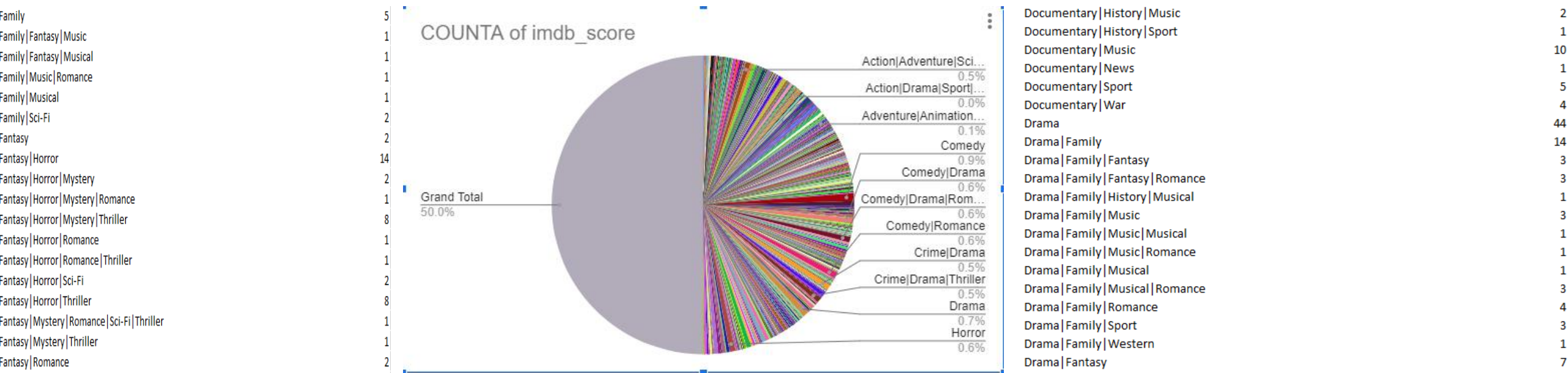
The most common genres of movie in the dataset is **Comedy** with count of IMDB Score is 56.

A	B	C	D	E	F	G	H
genres	imdb_score	Average	Median	mode	Range	variance	std.deviation
Action Adventure Fantasy Sci-Fi	7.9	7.008333333	7.4	6.7	5	2.571742	1.603665
Action Adventure Fantasy	7.1	6.238888889	6.3	6.7	4.1	1.360163	1.16626
Action Adventure Thriller	6.8	6.636363636	6.75	6.7	3.6	0.864329	0.929693
Action Thriller	8.5	5.8	5.75	6.7	5.3	1.991429	1.41118
Documentary	7.1	6.961904762	7.2	6.7	6	1.815476	1.347396
Action Adventure Sci-Fi	6.6	6.440625	6.55	6.7	6	1.821845	1.349757
Action Adventure Romance	6.2	5.9	6	6.7	2.2	0.846667	0.920145
Adventure Animation Comedy Family Fantasy Musical Romance	7.8	7.9	7.9	6.7	0.2	0.02	0.141421
Action Adventure Sci-Fi	7.5	6.440625	6.55	6.7	6	1.821845	1.349757
Adventure Family Fantasy Mystery	7.5	7.575	7.55	6.7	0.4	0.029167	0.170783
Action Adventure Sci-Fi	6.9	6.440625	6.55	6.7	6	1.821845	1.349757
Action Adventure Sci-Fi	6.1	6.440625	6.55	6.7	6	1.821845	1.349757
Action Adventure	6.7	6.7625	6.55	6.7	3.7	1.654107	1.286121
Action Adventure Fantasy	7.3	6.238888889	6.3	6.7	4.1	1.360163	1.16626
Action Adventure Western	6.5	6.45	6.55	6.7	3	0.979	0.989444
Action Adventure Fantasy Sci-Fi	7.2	7.008333333	7.4	6.7	5	2.571742	1.603665
Action Adventure Family Fantasy	6.6	5.1	5.1	6.7	3.1	1.655	1.286468
Action Adventure Sci-Fi	8.1	6.440625	6.55	6.7	6	1.821845	1.349757
Action Adventure Fantasy	6.7	6.238888889	6.3	6.7	4.1	1.360163	1.16626
Action Adventure Comedy Family Fantasy Sci-Fi	6.8	5.875	5.95	6.7	2	0.689167	0.830161

A	B	C	D	E	F	G	H
Adventure Fantasy	7.5	6.68	6.9	6.7	3.1	1.472	1.21326
Action Adventure Fantasy	7	6.238888889	6.3	6.7	4.1	1.360163	1.16626
Action Adventure Drama History	6.7	6.4	6.4	6.7	0.6	0.18	0.424264
Adventure Fantasy	7.9	6.68	6.9	6.7	3.1	1.472	1.21326
Adventure Family Fantasy	6.1	6.330769231	6.4	6.7	2.7	0.557308	0.74653
Action Adventure Drama Romance	7.2	7.075	7.05	6.7	1	0.1825	0.4272
Drama Romance	7.7	6.544827586	6.7	6.7	4.6	1.291847	1.136595
Action Adventure Sci-Fi	8.2	6.440625	6.55	6.7	6	1.821845	1.349757
Action Adventure Sci-Fi Thriller	5.9	6.580952381	6.4	6.7	3.8	1.053619	1.026459
Action Adventure Sci-Fi Thriller	7	6.580952381	6.4	6.7	3.8	1.053619	1.026459
Action Adventure Thriller	7.8	6.636363636	6.75	6.7	3.6	0.864329	0.929693
Action Adventure Fantasy Romance	7.3	6.95	6.95	6.7	0.7	0.245	0.494975
Action Adventure Sci-Fi	7.2	6.440625	6.55	6.7	6	1.821845	1.349757
Adventure Family Fantasy	6.5	6.330769231	6.4	6.7	2.7	0.557308	0.74653
Action Adventure Fantasy Sci-Fi Thriller	6.8	5.6	6.2	6.7	5.3	3.21	1.791647
Adventure Animation Comedy Family Fantasy	7.3	6.205263158	6.3	6.7	5.5	2.006082	1.416362
Action Adventure Sci-Fi	6	6.440625	6.55	6.7	6	1.821845	1.349757
Action Adventure Sci-Fi	5.7	6.440625	6.55	6.7	6	1.821845	1.349757
Adventure Family Fantasy	6.4	6.330769231	6.4	6.7	2.7	0.557308	0.74653
Action Adventure Fantasy Sci-Fi	6.7	7.008333333	7.4	6.7	5	2.571742	1.603665
Action Adventure Sci-Fi	6.8	6.440625	6.55	6.7	6	1.821845	1.349757



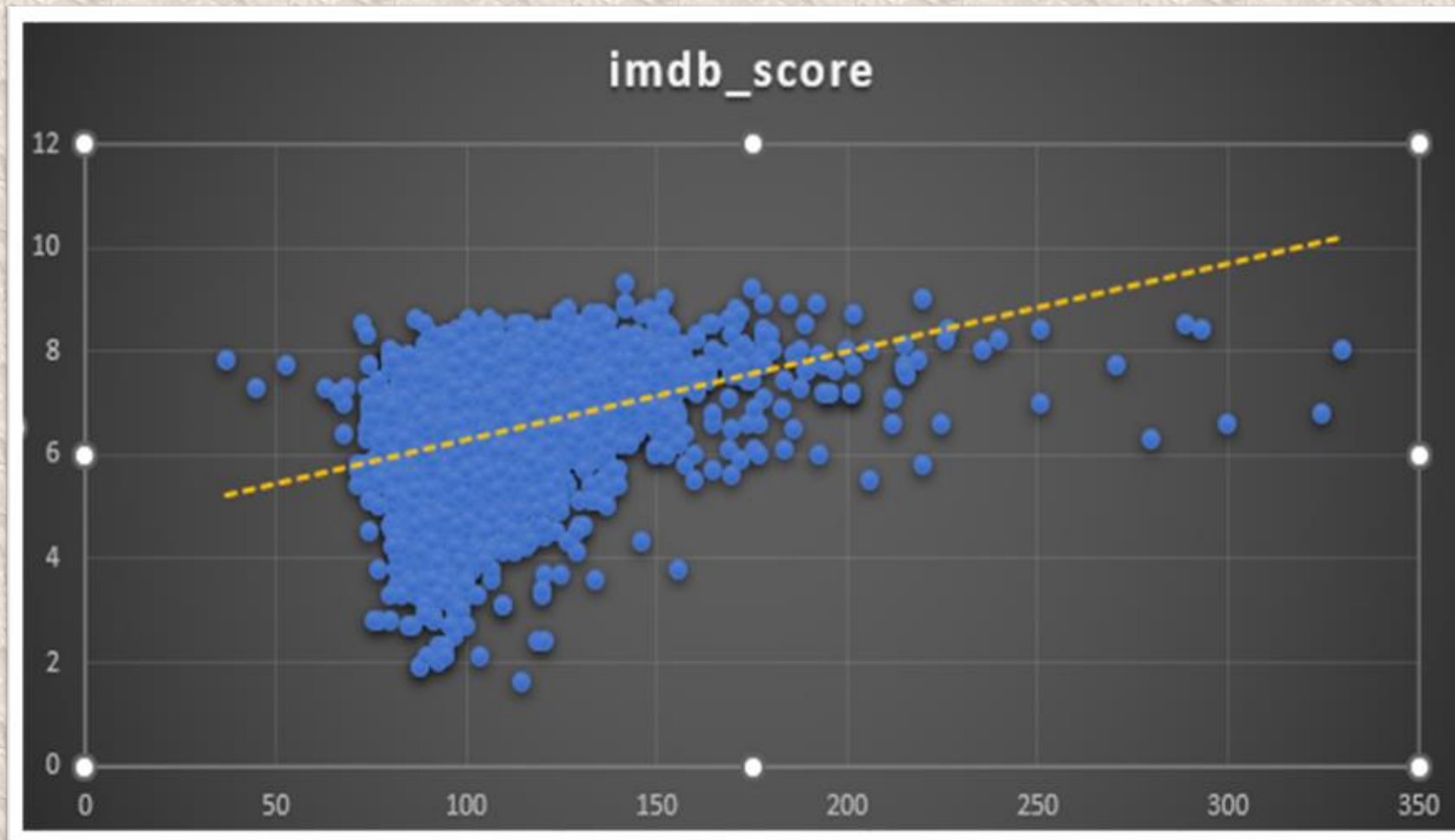
Row Labels	Count of imdb_score		
Action	11	Biography   Drama   Music   Romance	4
Action   Adventure	8	Biography   Drama   Romance	9
Action   Adventure   Animation   Comedy   Crime   Family   Fantasy	1	Biography   Drama   Romance   Sport	3
Action   Adventure   Animation   Comedy   Drama   Family   Fantasy   Thriller	1	Biography   Drama   Romance   War	2
Action   Adventure   Animation   Comedy   Drama   Family   Sci-Fi	2	Biography   Drama   Romance   Western	1
Action   Adventure   Animation   Comedy   Family	6	Biography   Drama   Sport	16
Action   Adventure   Animation   Comedy   Family   Fantasy	4	Biography   Drama   Sport   War	1
Action   Adventure   Animation   Comedy   Family   Fantasy   Sci-Fi	2	Biography   Drama   Thriller	5
Action   Adventure   Animation   Comedy   Family   Sci-Fi	3	Biography   Drama   Thriller   War	2
Action   Adventure   Animation   Comedy   Fantasy	1	Biography   Drama   War	4
Action   Adventure   Animation   Comedy   Fantasy   Sci-Fi	1	Comedy	56
Action   Adventure   Animation   Comedy   Sci-Fi	1	Comedy   Crime	30
Action   Adventure   Animation   Drama   Fantasy   Sci-Fi	1	Comedy   Crime   Drama	16
Action   Adventure   Animation   Drama   Mystery   Sci-Fi   Thriller	1	Comedy   Crime   Drama   Horror   Mystery   Thriller	1
Action   Adventure   Animation   Family	1	Comedy   Crime   Drama   Horror   Thriller	1
Action   Adventure   Animation   Family   Fantasy	1	Comedy   Crime   Drama   Music   Romance	1
Action   Adventure   Animation   Family   Fantasy   Sci-Fi	2	Comedy   Crime   Drama   Mystery	1
Action   Adventure   Animation   Family   Sci-Fi	2	Comedy   Crime   Drama   Mystery   Romance	3
Action   Adventure   Animation   Family   Sci-Fi   Thriller	1	Comedy   Crime   Drama   Mystery   Thriller	1
Action   Adventure   Animation   Fantasy	1	Comedy   Crime   Drama   Romance	2
		Comedy   Crime   Drama   Romance   Thriller	3



## ❖ MOVIE DURATION ANALYSIS:-

*your task :-* Analyze the distribution of movie durations and identify the relationship between movie duration and IMDB score.

duration	imdb_score	average	median	std.deviation
178	7.9	110.25799	106	22.64671656
169	7.1	110.23995	106	22.62272402
148	6.8	110.22429	106	22.60539247
164	8.5	110.21423	106	22.59998962
132	6.6	110.19989	106	22.58593046
156	6.2	110.19408	106	22.58613562
100	7.8	110.18187	106	22.5767531
141	7.5	110.18458	106	22.57915207
153	7.5	110.17636	106	22.57655142
183	6.9	110.16493	106	22.56871853
169	6.1	110.14549	106	22.54032262
106	6.7	110.12977	106	22.52279787
151	7.3	110.13088	106	22.52570515
150	6.5	110.11996	106	22.51880347
143	7.2	110.1093	106	22.51237027
150	6.6	110.10051	106	22.50895371
173	8.1	110.08984	106	22.5025024
136	6.7	110.07301	106	22.48197106
106	6.8	110.06608	106	22.48097747
164	7.5	110.06717	106	22.48388753

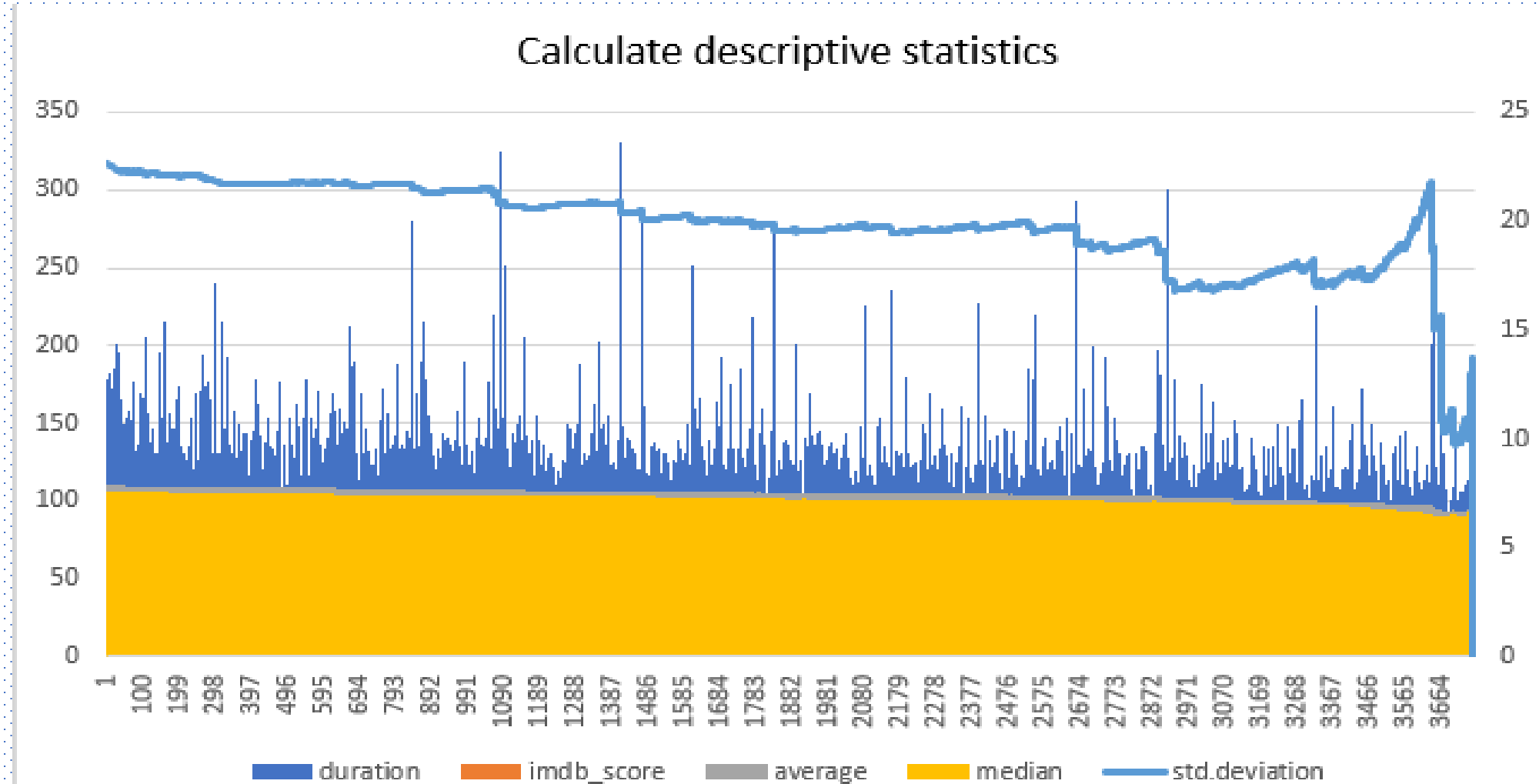


- ❖ Calculating the descriptive statistics such as average, median, and standard deviation for movie durations in chart.

**AVERAGE Function formula:-**  
= AVERAGE(A2:A3757).

**MEDIAN Function formula:-**  
= MEDIAN(A2:A3757).

**STDEV Function formula:-**  
= STDEV(A2:A3757)





# ❖ Language Analysis:-

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

movie_title	language	imdb_score	count	Average	Median	Std.Dev
Avatar	English	7.9	3624	6.415700883	6.6	1.0681016
Pirates of the Caribbean: At World's End	English	7.1	3624	6.415700883	6.6	1.0681016
Spectre	English	6.8	3624	6.415700883	6.6	1.0681016
The Dark Knight Rises	English	8.5	3624	6.415700883	6.6	1.0681016
John Carter	English	6.6	3624	6.415700883	6.6	1.0681016
Spider-Man 3	English	6.2	3624	6.415700883	6.6	1.0681016
Tangled	English	7.8	3624	6.415700883	6.6	1.0681016
Avengers: Age of Ultron	English	7.5	3624	6.415700883	6.6	1.0681016
Harry Potter and the Half-Blood Prince	English	7.5	3624	6.415700883	6.6	1.0681016
Batman v Superman: Dawn of Justice	English	6.9	3624	6.415700883	6.6	1.0681016
Superman Returns	English	6.1	3624	6.415700883	6.6	1.0681016
Quantum of Solace	English	6.7	3624	6.415700883	6.6	1.0681016
Pirates of the Caribbean: Dead Man's Chest	English	7.3	3624	6.415700883	6.6	1.0681016
The Lone Ranger	English	6.5	3624	6.415700883	6.6	1.0681016
Man of Steel	English	7.2	3624	6.415700883	6.6	1.0681016
The Chronicles of Narnia: Prince Caspian	English	6.6	3624	6.415700883	6.6	1.0681016
The Avengers	English	8.1	3624	6.415700883	6.6	1.0681016
Pirates of the Caribbean: On Stranger Tides	English	6.7	3624	6.415700883	6.6	1.0681016
Men in Black 3	English	6.8	3624	6.415700883	6.6	1.0681016
The Hobbit: The Battle of the Five Armies	English	7.5	3624	6.415700883	6.6	1.0681016
The Amazing Spider-Man	English	7	3624	6.415700883	6.6	1.0681016
Robin Hood	English	6.7	3624	6.415700883	6.6	1.0681016

Zombie Hunter	English	7.1	3624	6.415700883	6.6	1.0681016
Banshee Chapter	English	6.1	3624	6.415700883	6.6	1.0681016
Ask Me Anything	English	6.4	3624	6.415700883	6.6	1.0681016
And Then Came Love	English	7	3624	6.415700883	6.6	1.0681016
Food Chains	English	6.6	3624	6.415700883	6.6	1.0681016
On the Waterfront	French	7.8	34	7.355882353	6.6	1.0681016
Llife Happens	French	7.7	34	7.355882353	6.6	1.0681016
4 Months, 3 Weeks and 2 Days	French	7.2	34	7.355882353	6.6	1.0681016
Hard Candy	French	8.4	34	7.355882353	6.6	1.0681016
The Quiet	French	7.1	34	7.355882353	6.6	1.0681016
Fruitvale Station	French	6.7	34	7.355882353	6.6	1.0681016
The Brass Teapot	French	7.8	34	7.355882353	6.6	1.0681016
Snitch	French	7.3	34	7.355882353	6.6	1.0681016
Latter Days	French	6.9	34	7.355882353	6.6	1.0681016
For a Good Time, Call...	French	6.6	34	7.355882353	6.6	1.0681016
Time Changer	French	7.2	34	7.355882353	6.6	1.0681016
A Separation	French	7.5	34	7.355882353	6.6	1.0681016
Welcome to the Dollhouse	French	7.3	34	7.355882353	6.6	1.0681016
Ruby in Paradise	French	7.3	34	7.355882353	6.6	1.0681016
Raising Victor Vargas	French	7.9	34	7.355882353	6.6	1.0681016
Deterrence	French	7.7	34	7.355882353	6.6	1.0681016
Dead Snow	French	7.9	34	7.355882353	6.6	1.0681016

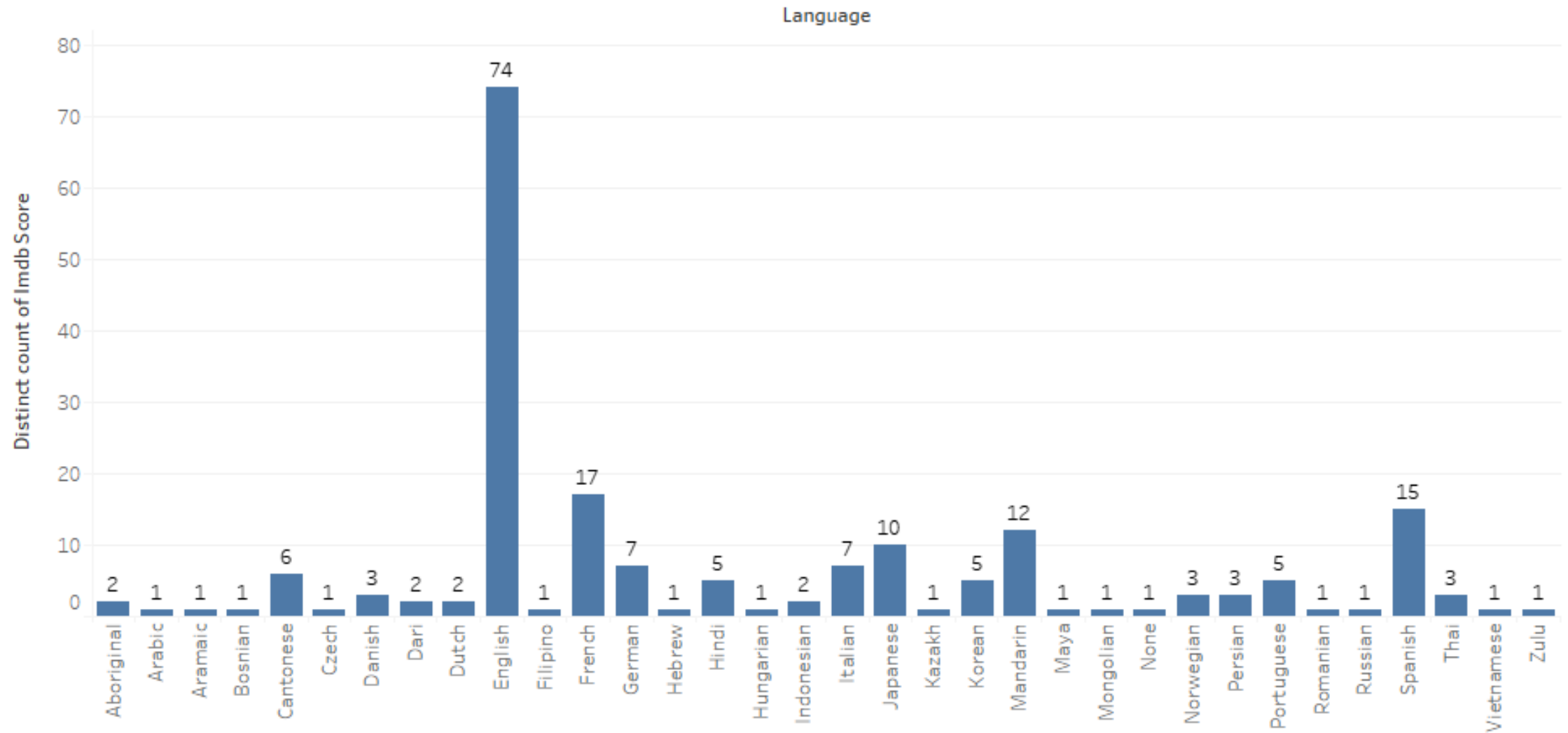
The most common Language used in movies is English.

Super Size Me	Persian	7.5	3	8.133333333	6.6	1.0681016
The FP	Aboriginal	6.4	2	6.95	6.6	1.0681016
Happy Christmas	Dari	7.6	2	7.5	6.6	1.0681016
Absentia	Aboriginal	7.5	2	6.95	6.6	1.0681016
The Brothers McMullen	Indonesian	7.6	2	7.9	6.6	1.0681016
Tiny Furniture	Indonesian	8.2	2	7.9	6.6	1.0681016
The Call of Cthulhu	Dari	7.4	2	7.5	6.6	1.0681016
The Signal	Filipino	6.7	1	6.7	6.6	1.0681016
This Is Martin Bonner	Maya	7.8	1	7.8	6.6	1.0681016
George Washington	Kazakh	6	1	6	6.6	1.0681016
Smiling Fish & Goat on Fire	Aramaic	7.1	1	7.1	6.6	1.0681016
Clerks	Mongolian	7.3	1	7.3	6.6	1.0681016
In the Company of Men	Bosnian	4.3	1	4.3	6.6	1.0681016
Sabotage	Hungarian	7.1	1	7.1	6.6	1.0681016

George Washington	Kazakh	6	1	6	6.6	1.0681016
Smiling Fish & Goat on Fire	Aramaic	7.1	1	7.1	6.6	1.0681016
Clerks	Mongolian	7.3	1	7.3	6.6	1.0681016
In the Company of Men	Bosnian	4.3	1	4.3	6.6	1.0681016
Sabotage	Hungarian	7.1	1	7.1	6.6	1.0681016
Slacker	Czech	7.4	1	7.4	6.6	1.0681016
Pink Flamingos	Russian	6.5	1	6.5	6.6	1.0681016
Clean	None	8.5	1	8.5	6.6	1.0681016
The Circle	Zulu	7.3	1	7.3	6.6	1.0681016
Bang	Hebrew	8	1	8	6.6	1.0681016
Primer	Arabic	7.2	1	7.2	6.6	1.0681016
El Mariachi	Vietnamese	7.4	1	7.4	6.6	1.0681016
My Date with Drew	Romanian	7.9	1	7.9	6.6	1.0681016

# Language Analysis by Bar Chart:-

Sheet 1



# ❖ DIRECTOR ANALYSIS:-

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

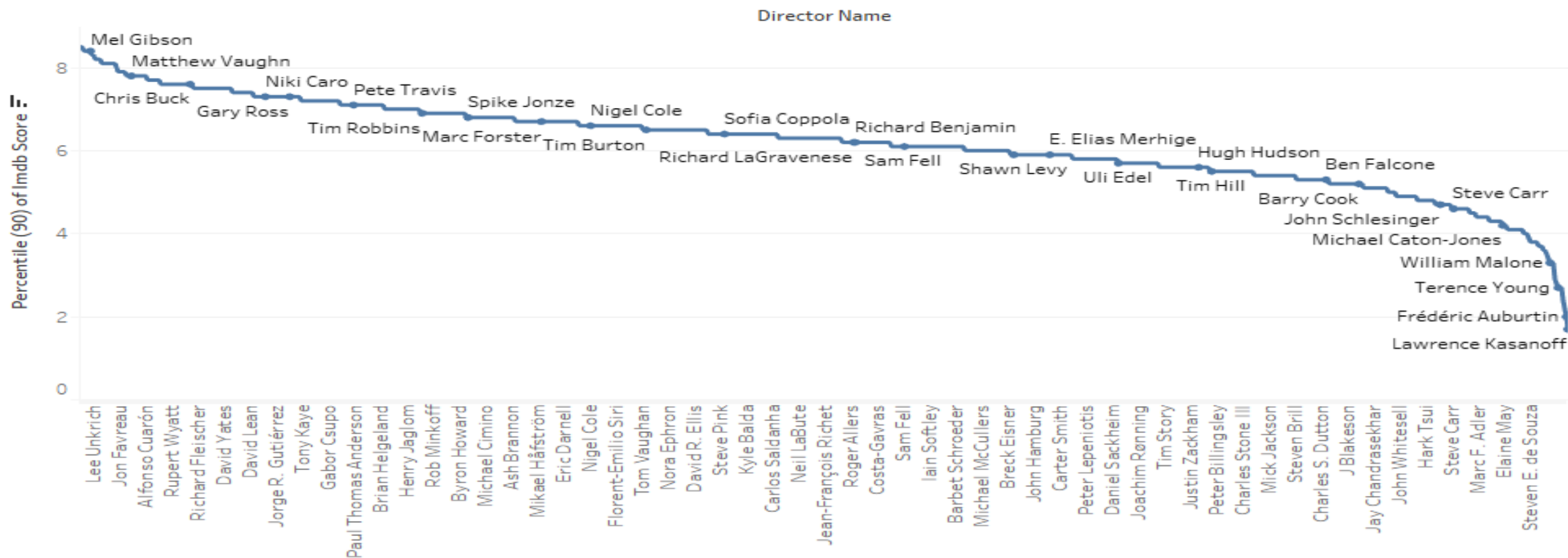
director_name	imdb_score	average score	percentile	comparision	Fernando Meirelles	7.5	7.5	7.5	Yes
Christopher Nolan	8.5	8.5	7.5	Yes	Kirsten Sheridan	7.5	7.5	7.5	Yes
Quentin Tarantino	8.5	8.5	7.5	Yes	Richard Fleischer	7.5	7.5	7.5	Yes
Mel Gibson	8.4	8.4	7.5	Yes	Dexter Fletcher	7.5	7.5	7.5	Yes
S.S. Rajamouli	8.4	8.4	7.5	Yes	Bill Paxton	7.5	7.5	7.5	Yes
Moustapha Akkad	8.4	8.4	7.5	Yes	Ava DuVernay	7.5	7.5	7.5	Yes
Richard Marquand	8.4	8.4	7.5	Yes	Christopher McQuarrie	7.4	7.4	7.5	No
Sergio Leone	8.4	8.4	7.5	Yes	Peter Weir	7.4	7.4	7.5	No
Lee Unkrich	8.3	8.3	7.5	Yes	Peyton Reed	7.4	7.4	7.5	No
Pete Docter	8.3	8.3	7.5	Yes	David Silverman	7.4	7.4	7.5	No
Anthony Russo	8.2	8.2	7.5	Yes	Milos Forman	7.4	7.4	7.5	No
Dean DeBlois	8.2	8.2	7.5	Yes	Richard Curtis	7.4	7.4	7.5	No
Hideaki Anno	8.2	8.2	7.5	Yes	Julie Taymor	7.4	7.4	7.5	No
James McTeigue	8.2	8.2	7.5	Yes	Tobe Hooper	7.4	7.4	7.5	No
James Gunn	8.1	8.1	7.5	Yes	Jason Reitman	7.4	7.4	7.5	No
George Miller	8.1	8.1	7.5	Yes	Rian Johnson	7.4	7.4	7.5	No
Alejandro G. Iñárritu	8.1	8.1	7.5	Yes	Paul Haggis	7.4	7.4	7.5	No
Paul Greengrass	8.1	8.1	7.5	Yes	David Lean	7.4	7.4	7.5	No
Tim Miller	8.1	8.1	7.5	Yes	Richard Eyre	7.4	7.4	7.5	No
Andrei Tarkovsky	8.1	8.1	7.5	Yes	Dan Scanlon	7.3	7.3	7.5	No
Denis Villeneuve	8.1	8.1	7.5	Yes	Baz Luhrmann	7.3	7.3	7.5	No



# Line graph :- Director Analysis.

The top directors based on their average IMDB score is **Christopher Nolan**.

Director Analysis



# BUDGET ANALYSIS :-

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

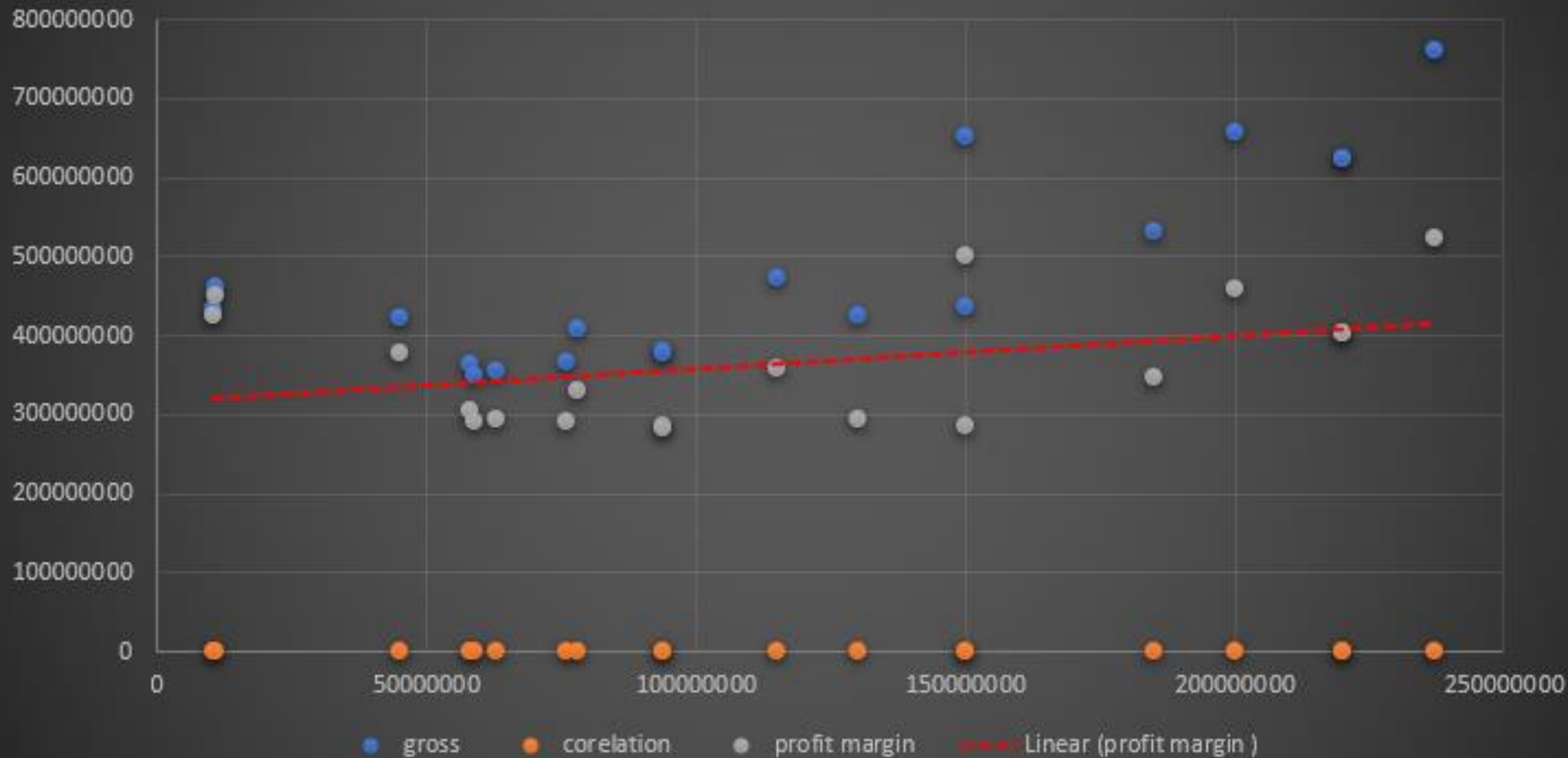
The movies with the highest profit margin is **Avatar**

*The correlation coefficient formula is :-  
= CORREL(Budget, Correlation).*

*The Profit margin formula is :- = Gross - Budget*

movie_title	budget	gross	corelation	profit margin
Avatar	237000000	760505847	0.099496423	523505847
Jurassic World	150000000	652177271	0.09858198	502177271
Titanic	200000000	658672302	0.098512556	458672302
Star Wars: Episode IV - A New Hope	11000000	460935665	0.097939736	449935665
E.T. the Extra-Terrestrial	10500000	434949459	0.098665989	424449459
The Avengers	220000000	623279547	0.099329364	403279547
The Avengers	220000000	623279547	0.098547553	403279547
The Lion King	45000000	422783777	0.097759116	377783777
Star Wars: Episode I - The Phantom Menace	115000000	474544677	0.098180188	359544677
The Dark Knight	185000000	533316061	0.098201276	348316061
The Hunger Games	78000000	407999255	0.097707991	329999255
Deadpool	58000000	363024263	0.097895447	305024263
The Hunger Games: Catching Fire	130000000	424645577	0.098131747	294645577
Jurassic Park	63000000	356784000	0.097998807	293784000
Despicable Me 2	76000000	368049635	0.098200681	292049635
American Sniper	58800000	350123553	0.098347791	291323553
Finding Nemo	94000000	380838870	0.098566438	286838870
Shrek 2	150000000	436471036	0.098627631	286471036
The Lord of the Rings: The Return of the King	94000000	377019252	0.098370228	283019252
Star Wars: Episode VI - Return of the Jedi	32500000	309125409	0.098429422	276625409

# Budget Analysis





## ❖ RESULT :-

- Overall, the project provided valuable insights into movie data analysis and the factors that can influence movie ratings and success. It allowed me to gain a deeper understanding of the relationships between variables such as genre, duration, language, director, and budget with IMDB scores.
- This understanding can be valuable for filmmakers, studios, and other industry professionals to make informed decisions when creating and promoting movies.
- Additionally, the project enhanced our proficiency in using various Excel functions for data analysis, including COUNTIF, AVERAGE, MEDIAN, MODE, MAX, MIN, VAR, STDEV, and PERCENTILE.
- These skills can be applied in other data analysis tasks across various domains beyond movie analysis.
- Overall, the IMDB Movie Analysis project has contributed me to a better understanding of movie data analysis, the importance of data-driven decision-making in the film industry, and the utilization of Excel as a powerful tool for statistical analysis and visualization.