

# IMDB Movie Analysis

## Project Description

The 'IMDB Movie Analysis' project is based on finding out different factors that affect to the movie success. In this project first of all the 'IMDB Movies' dataset is provided to us. This dataset contains essential information about movies, including director names, movie titles, duration, genre, budget, gross earnings, IMDB ratings and more.

By the use of Excel's skills, statistics techniques and data visualization, we perform different analysis on provided data and extract valuable insights that effect the imdb ratings.

In this project, the following analysis have been performed:-

- **Movie genre analysis** : Analyzing the different genres, the no of movies based on genres and and performing different statistics functions.
- **Movie duration analysis** : Analyzing the distribution of movie durations and its impact on the IMDB score.
- **Language analysis** : Analyzing the different language and performing statistics functions.
- **Director analysis** : Finding out the top directors and their excellence on movies success.
- **Budget analysis** : Calculating the profit margin and finding out the correlation with movies budget and also analyzing the top movie which have highest profit margin.

## Approach

For this project I apply the following approach:-

1. First of all I analyze the "IMDB\_Movies" dataset. This dataset contains different information like director name, genres, language, actor name, facebook likes, duration, gross earnings, country and so on. I assume that this dataset included

unwanted columns, null values, duplicated value , empty rows. So before performing data analysis we need to clean the data.

2. Then I perform data cleaning activity. It includes handling missing values, removing duplicates and deleting unwanted columns etc.
3. After that using excel i perform various analysis on the dataset to answers the questions mentioned in the project. It involves, genre analysis, duration analysis, language analysis, director and budget analysis. During the data analysis I find out the correlation between movie ratings and other factors like genre, director, budget etc., calculating descriptive statistics like mean, median, mode, range, variance, standard deviation. To visualize the data we use visualization tools like scatter plot . I also use different formulas and functions like unique function, max and min functions, index etc.

## **Tech Stack Used**

**Excel | Microsoft 356:** Since the project is totally based on excels and statistics, therefore I use ms excel. I use excel's different tools , different tabs, editing tools, formulas, sort and filter option , table design options and many more. They are easy to access and comfortable.

## **Insights**

The insights and knowledge gained during the project:

This project help me to expand my statistics knowledge for eg. exploratory data analysis like Descriptive statistics, different formulas and functions like mean, median,stdev, variance, range, max and min functions, index and unique functions etc. data cleaning activity helps me to preprocess the data to make it valuable for data analysis

Also it expands my excel skills like how to use tables, sort and filtering options, data visualization tools help me how to convert the data into scatter, and all other tools and techniques like to design tables , editing options and use of different tabs and many more that are available on ms excel.

In this project the following observations and meaningful trends are covered:-

**Data Cleaning:** This step involves preprocessing the data to make it suitable for analysis. The cleaned IMDB Movie dataset :-

	director_name	duration	gross	genres	movie_title	language	country	budget	title_year	imdb_score	aspect_ratio
1	James Cameron	178	760505847	Action Adventure Fantasy Sci-Fi	Avatar	English	USA	237000000	2009	7.9	
2	Gore Verbinski	169	309404152	Action Adventure Fantasy	Pirates of the Caribbean: At World's End	English	USA	300000000	2007	7.1	
3	Sam Mendes	148	200074175	Action Adventure Thriller	Spectre	English	UK	245000000	2015	6.8	
4	Christopher Nolan	164	448130642	Action Thriller	The Dark Knight Rises	English	USA	250000000	2012	8.5	
5	Sam Raimi	156	336530303	Action Adventure Romance	Spider-Man 3	English	USA	258000000	2007	6.2	
6	Nathan Greno	100	200807262	Adventure Animation Comedy Family	Tangled	English	USA	260000000	2010	7.8	
7	Joss Whedon	141	458991599	Action Adventure Sci-Fi	Avengers: Age of Ultron	English	USA	250000000	2015	7.5	
8	David Yates	153	301956980	Adventure Family Fantasy Mystery	Harry Potter and the Half-Blood Prince	English	UK	250000000	2009	7.5	
9	Zack Snyder	183	330249062	Action Adventure Sci-Fi	Batman v Superman: Dawn of Justice	English	USA	250000000	2016	6.9	
10	Bryan Singer	169	200069408	Action Adventure Sci-Fi	Superman Returns	English	USA	209000000	2006	6.1	
11	Marc Forster	106	168368427	Action Adventure	Quantum of Solace	English	UK	200000000	2008	6.7	
12	Gore Verbinski	151	423032628	Action Adventure Fantasy	Pirates of the Caribbean: Dead Man's Chest	English	USA	225000000	2006	7.3	
13	Gore Verbinski	150	89289910	Action Adventure Western	The Lone Ranger	English	USA	215000000	2013	6.5	
14	Zack Snyder	143	291021565	Action Adventure Fantasy Sci-Fi	Man of Steel	English	USA	225000000	2013	7.2	
15	Andrew Adamson	150	141614023	Action Adventure Family Fantasy	The Chronicles of Narnia: Prince Caspian	English	USA	225000000	2008	6.6	
16	Joss Whedon	173	623279547	Action Adventure Sci-Fi	The Avengers	English	USA	220000000	2012	8.1	
17	Rob Marshall	136	241063875	Action Adventure Fantasy	Pirates of the Caribbean: On Stranger Tides	English	USA	250000000	2011	6.7	
18	Barry Sonnenfeld	106	179020854	Action Adventure Comedy Family	Men in Black 3	English	USA	225000000	2012	6.8	
19	Peter Jackson	164	255108370	Adventure Fantasy	The Hobbit: The Battle of the Five Armies	English	New Zealand	250000000	2014	7.5	
20	Marc Webb	153	262030663	Action Adventure Fantasy	The Amazing Spider-Man	English	USA	230000000	2012	7	
21	Ridley Scott	156	105219735	Action Adventure Drama History	Robin Hood	English	USA	200000000	2010	6.7	
22	Peter Jackson	186	258355354	Adventure Fantasy	The Hobbit: The Desolation of Smaug	English	USA	225000000	2013	7.9	
23	Chris Weitz	113	70083519	Adventure Family Fantasy	The Golden Compass	English	USA	180000000	2007	6.1	
24	Peter Jackson	201	218051260	Action Adventure Drama Romance	King Kong	English	New Zealand	207000000	2005	7.2	
25	James Cameron	194	658672302	Drama Romance	Titanic	English	USA	200000000	1997	7.7	

**Tasks:-**

## A. Movie Genre Analysis:

**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.

**Result:-**

The common genres are:-

GENRES	NO. OF MOVIES
Action	923
Adventure	358
Drama	644
Animation	45
Comedy	980
Mystery	22
Crime	248
Biography	203
Fantasy	35
Documentary	26
Sci-Fi	7
Horror	155
Romance	1
Family	3
Western	2
Musical	2
Thriller	1

The median, max, min, mode, variance and STDEV for each genre:-

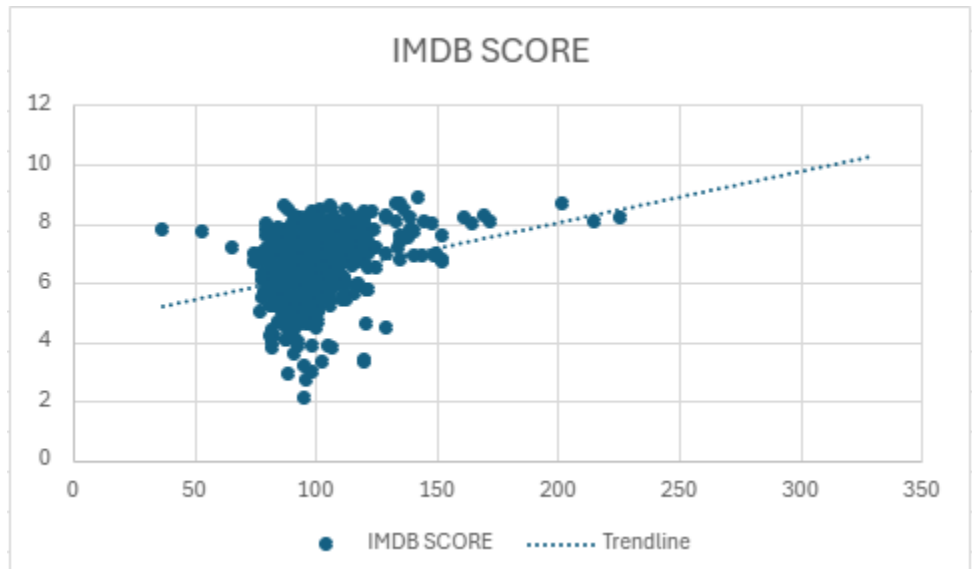
GENRES	NO. OF MOVIES	MEAN	MEDIAN	MODE	MAX	MIN	RANGE	VARIANCE	STDEV
Action	923	6.28	6.30	6.10	9.00	2.10	6.90	1.08	1.04
Adventure	358	6.56	6.70	7.30	8.60	2.30	6.30	1.28	1.13
Drama	644	6.84	6.95	6.70	8.80	2.10	6.70	0.82	0.91
Animation	45	6.74	7.00	7.10	8.00	4.50	3.50	0.94	0.97
Comedy	980	6.17	6.30	6.40	8.80	1.90	6.90	1.06	1.03
Mystery	22	6.63	6.70	6.60	8.50	3.30	5.20	1.24	1.11
Crime	248	6.94	7.00	7.30	9.30	3.30	6.00	0.76	0.87
Biography	203	7.16	7.20	7.00	8.90	4.50	4.40	0.49	0.70
Fantasy	35	6.23	6.50	6.80	7.90	4.30	3.60	0.80	0.89
Documentary	26	6.80	7.45	7.50	8.50	1.60	6.90	2.95	1.72
Sci-Fi	7	6.63	6.40	#N/A	8.20	5.00	3.20	1.23	1.11
Horror	155	5.81	5.90	5.90	8.50	2.30	6.20	1.02	1.01
Romance	1	7.10	7.10	#N/A	7.10	7.10	0.00	#DIV/0!	#DIV/0!
Family	3	6.50	5.90	#N/A	7.90	5.70	2.20	1.48	1.22
Western	2	8.10	8.10	#N/A	8.90	7.30	1.60	1.28	1.13
Musical	2	6.75	6.75	#N/A	7.20	6.30	0.90	0.41	0.64
Thriller	1	4.80	4.80	#N/A	4.80	4.80	0.00	#DIV/0!	#DIV/0!

## B. Movie Duration Analysis:

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

**Result:-**

MEAN for movie duration	110.16
MEDIAN	106.00
STDEV	22.70



### C. Language Analysis:

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

#### Result:-

Common Languages are mentioned below:

D	E	
LANGUAGE	NO OF MOVIES	
English	3498	
Mandarin	14	
Aboriginal	2	
Spanish	23	
French	34	
Filipino	1	
Maya	1	
Kazakh	1	
Cantonese	7	
Japanese	10	
Aramaic	1	
Italian	7	
Dutch	3	
Dari	2	
German	10	
Mongolian	1	
Thai	3	
Bosnian	1	
Korean	5	
Hungarian	1	
Hindi	5	
Danish	3	

Portuguese	5	
Norwegian	4	
Czech	1	
Russian	1	
None	1	
Zulu	1	
Hebrew	1	
Arabic	1	
Vietnamese	1	
Indonesian	2	
Romanian	1	
Persian	3	

The mean, median, and standard deviation of the IMDB scores for each language.

LANGUAGE	NO OF MOVIES	MEAN	MEDIAN	STDEV
English	3498	6.42	6.50	1.05
Mandarin	14	7.02	7.25	0.77
Aboriginal	2	6.95	6.95	0.78
Spanish	23	7.08	7.20	0.86
French	34	7.36	7.30	0.52
Filipino	1	6.70	6.70	#DIV/0!
Maya	1	7.80	7.80	#DIV/0!
Kazakh	1	6.00	6.00	#DIV/0!
Cantonese	7	7.34	7.30	0.35
Japanese	10	7.66	8.00	0.99
Aramaic	1	7.10	7.10	#DIV/0!
Italian	7	7.19	7.00	1.16
Dutch	3	7.57	7.80	0.40
Dari	2	7.50	7.40	0.73
German	10	7.77	7.80	0.71
Mongolian	1	7.30	7.30	#DIV/0!
Thai	3	6.63	6.60	0.45
Bosnian	1	4.30	4.30	#DIV/0!
Korean	5	7.70	7.70	0.57
Hungarian	1	7.10	7.10	#DIV/0!
Hindi	5	7.22	7.40	0.80
Danish	3	7.90	8.10	0.53

Danish	3	7.90	8.10	0.53
Portuguese	5	7.76	8.00	0.98
Norwegian	4	7.15	7.30	0.57
Czech	1	7.40	7.40	#DIV/0!
Russian	1	6.50	6.50	#DIV/0!
None	1	8.50	8.50	#DIV/0!
Zulu	1	7.30	7.30	#DIV/0!
Hebrew	1	8.00	8.00	#DIV/0!
Arabic	1	7.20	7.20	#DIV/0!
Vietnamese	1	7.40	7.40	#DIV/0!
Indonesian	2	7.90	7.90	0.42
Romanian	1	7.90	7.90	#DIV/0!
Persian	3	8.13	8.40	0.55

#### D. Director Analysis:

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

**Result:-**

B	C	D	E
IMDB SCORE	DIRECTORS	AVERAGE	PERCENTILE
8.5	Christopher Nolan	8.43	0.99
6.6	Richard Marquand	8.40	0.99
6.7	Sergio Leone	8.43	1.00
7.1	Alfred Hitchcock	8.50	1.00
7.2	Tony Kaye	8.60	1.00
5.5	Ron Fricke	8.50	1.00
7.2	Damien Chazelle	8.50	1.00
7	Charles Chaplin	8.60	1.00
6.5	Asghar Farhadi	8.40	0.99
7.9	Akira Kurosawa	8.70	1.00
6.9	Majid Majidi	8.50	1.00

## E. Budget Analysis:

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

**Result:-**



	A	B	C	D
1	MOVIE TITLE	GROSS	BUDGET	PROFIT MARGIN
2	Avatar	760505847	237000000	523505847
3	Pirates of the Caribbean: At World's End	309404152	300000000	9404152
4	Spectre	200074175	245000000	-44925825
5	The Dark Knight Rises	448130642	250000000	198130642
6	Spider-Man 3	336530303	258000000	78530303
7	Tangled	200807262	260000000	-59192738
8	Avengers: Age of Ultron	458991599	250000000	208991599
9	Harry Potter and the Half-Blood Prince	301956980	250000000	51956980
10	Batman v Superman: Dawn of Justice	330249062	250000000	80249062
11	Superman Returns	200069408	209000000	-8930592
12	Quantum of Solace	168368427	200000000	-31631573
13	Pirates of the Caribbean: Dead Man's Chest	423032628	225000000	198032628
14	The Lone Ranger	89289910	215000000	-125710090
15	Man of Steel	291021565	225000000	66021565
16	The Chronicles of Narnia: Prince Caspian	141614023	225000000	-83385977
17	The Avengers	652177271	150000000	502177271
18	Pirates of the Caribbean: On Stranger Tides	241063875	250000000	-8936125
19	Men in Black 3	179020854	225000000	-45979146
20	The Hobbit: The Battle of the Five Armies	255108370	250000000	5108370
21	The Amazing Spider-Man	262030663	230000000	32030663
22	Robin Hood	105219735	200000000	-94780265
23	The Hobbit: The Desolation of Smaug	258355354	225000000	33355354

	A	B	C	D
24	The Golden Compass	70083519	180000000	-109916481
25	King Kong	218051260	207000000	11051260
26	Titanic	658672302	200000000	458672302
27	Captain America: Civil War	407197282	250000000	157197282
28	Battleship	65173160	209000000	-143826840
29	Jurassic World	460935665	11000000	449935665
30	Skyfall	304360277	200000000	104360277
31	Spider-Man 2	373377893	200000000	173377893
32	Iron Man 3	408992272	200000000	208992272
33	Alice in Wonderland	334185206	200000000	134185206
34	X-Men: The Last Stand	234360014	210000000	24360014
35	Monsters University	268488329	200000000	68488329
36	Transformers: Revenge of the Fallen	402076689	200000000	202076689
37	Transformers: Age of Extinction	245428137	210000000	35428137
38	Oz the Great and Powerful	234903076	215000000	19903076
39	The Amazing Spider-Man 2	202853933	200000000	2853933
40	TRON: Legacy	172051787	170000000	2051787
41	Cars 2	191450875	200000000	-8549125
42	Green Lantern	116593191	200000000	-83406809
43	Toy Story 3	414984497	200000000	214984497
44	Terminator Salvation	125320003	200000000	-74679997
45	Furious 7	350034110	190000000	160034110
46	World War Z	202351611	100000000	12351611

COFFIECIENT CORREALTION r =
-0.95514642
HIGHEST PROFIT MARGIN
523505847 MOVIE AVATAR

## Result

This analysis clears my queries like how to improve imdb score, what are the profitable and harmful factors for movies etc. This project provides important insights which helps in making Decisions like top genres, top languages, top directors, profit margin and their impact of imdb score. I am able to understand the importance of Data Analytics in Movies success.

I have also gained experience for the data analysis using statistical knowledge and excel's tools and techniques. Through this I have learnt to apply my data analytics skills in solving real life problems.

## Hyperlink the Excel Sheet to report:

Excel sheet link:

[Click Here](#)

## The Video Presentation Link:

Presentation link:

[Click Here](#)