

## A large grid of movie posters and actor portraits, with the IMDb logo prominently displayed in the center. The grid is composed of many small squares, each containing a different image, likely representing a large collection of movies or actors. The IMDb logo is a yellow rectangle with the text "IMDb" in black, positioned in the lower-left quadrant of the grid.

# Pritesh Kumar Bag



C-97, C-97, Ahinsa Cir, Panch Batti, C Scheme,  
Ashok Nagar, Jaipur, Rajasthan 302001

**AUGUST 29, 2023**

## **PROJECT DESCRIPTION**

IMDb, short for the Internet Movie Database, is a widely recognized online platform that provides a comprehensive database of information related to films, television shows, actors, directors, and other entertainment industry professionals. Users can access details such as cast and crew lists, plot summaries, user reviews, ratings, and trivia for a vast array of movies and TV shows from around the world. IMDb serves as a valuable resource for film enthusiasts, industry professionals, and casual viewers seeking information about their favourite media. It has become a go-to platform for discovering, exploring, and learning about the world of cinema and television.

## **APPROACH**

- Downloaded the given datafile
- Understanding the datafile
- Checking the blanks and outliers
- Removing unwanted data
- Drawing summary from the data
- Used formulas, pivot table and charts

## **TECH STACK USED**

- Microsoft Excel 2016

## INSIGHTS

- 1. Movie Genre Analysis: Analyse 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. Hint: Use Excel's COUNTIF function to count the number of movies for each genre. You might need to manipulate the 'genres' column to separate multiple genres for a single movie. Use Excel's functions like AVERAGE, MEDIAN, MODE, MAX, MIN, VAR, and STDEV to calculate descriptive statistics. Compare the statistics to understand the impact of genre on movie ratings.

Descriptive Statistics from IMDB Scores						
Genre	Count of Movies	Mean Of IMDB score		Genre	Count of Movies	Mean Of IMDB score
Action	962	6.290748441		Horror	391	5.926086957
Adventure	787	6.458322745		Romance	877	6.429190422
Drama	1941	6.786192684		Family	450	6.210444444
Animation	199	6.700502513		Western	58	6.765517241
Comedy	1503	6.184497671		Musical	103	6.559223301
Mystery	383	6.475195822		Thriller	1117	6.378066249
Crime	714	6.544817927		History	153	7.136601307
Biography	243	7.141563786		Music	248	6.460483871
Fantasy	514	6.290077821		War	160	7.05375
Documentary	64	6.99375		Sport	151	6.603311258
Sci-Fi	497	6.322736419		Short	2	6.8
				Film-Noir	1	7.7

	Action	Adventure	Drama	Animation	Comedy	Mystery	Crime	Biography	Fantasy	Documentary	Sci-Fi	
MEDIAN	6.3	6.6	6.9	6.8	6.3	6.5	6.6	7.2	6.4	7.2	6.4	
MODE	6.1	6.7	6.7	6.7	6.7	6.6	6.6	7	6.7	6.6	6.7	
VARIANCE	1.0638	1.227980898	0.79406	0.97732298	1.07741	1.01192	0.96070943	0.5026868	1.27661	1.499642857	1.337	
SD	1.0314	1.108142995	0.8911	0.98859647	1.03798	1.00594	0.98015786	0.70900409	1.12987	1.22459906	1.156	
MAX	9	8.9	9.3	8.6	8.8	8.6	9.3	8.9	8.9	8.5	8.8	
MIN	2.1	2.3	2.1	2.8	1.9	3.1	2.4	4.5	2.2	1.6	1.9	
RANGE	6.9	6.6	7.2	5.8	6.9	5.5	6.9	4.4	6.7	6.9	6.9	
	Horror	Romance	Family	Western	Musical	Thriller	History	Music	War	Sport	Short	Film-Noir
MEDIAN	6	6.5	6.3	6.8	6.7	6.4	7.2	6.7	7.1	6.8	6.8	7.7
MODE	5.9	6.5	6.7	6.8	7.1	6.5	7.7	6.2	7.1	7.2	#N/A	#N/A
VARIANCE	0.9942	0.935608134	1.35158	0.99703569	1.30185	0.93347	0.45246732	1.41122568	0.65282	1.088588962	0.18	#DIV/0!
SD	0.9971	0.967268388	1.16258	0.99851675	1.14099	0.96616	0.67265691	1.1879502	0.80797	1.043354668	0.424	#DIV/0!
MAX	8.6	8.5	8.6	8.9	8.5	9	8.9	8.5	8.6	8.4	7.1	7.7
MIN	2.3	2.1	1.9	4.1	2.1	2.7	5.5	1.6	4.3	2	6.5	7.7
RANGE	6.3	6.4	6.7	4.8	6.4	6.3	3.4	6.9	4.3	6.4	0.6	0

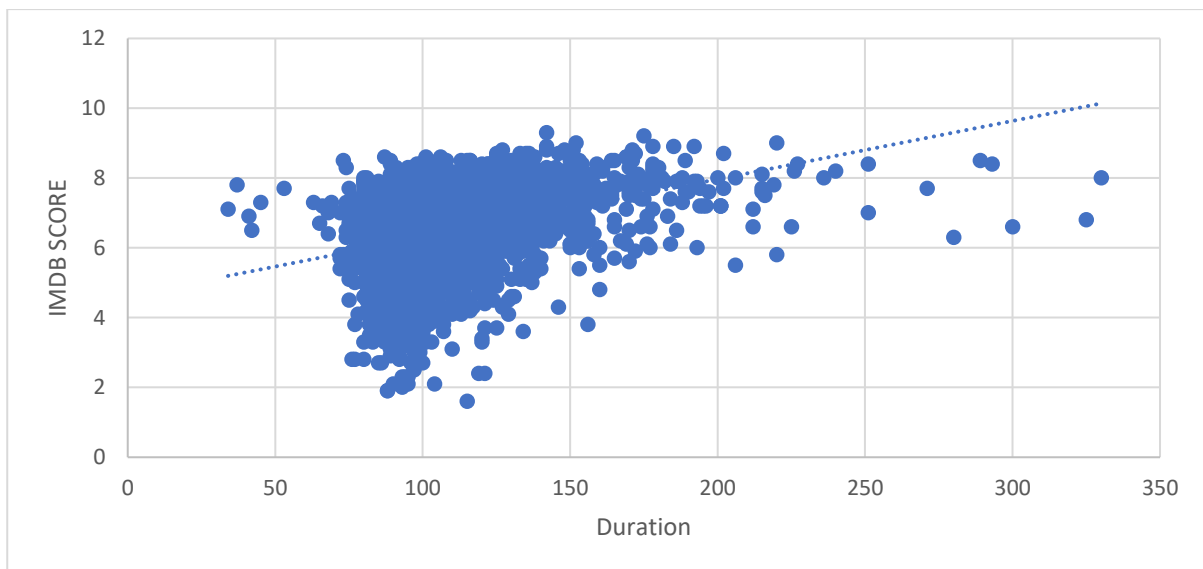
**INFERENCE**-Maximum count of movies is Drama followed by Comedy genre. Lowest is film Noir Genre. War, Documentary, Biography and History genre though have less count in movies but their ratings are more than 7 on average. It means people like movies based on facts and history.

**2. Movie Duration Analysis: Analyse the distribution of movie durations and its impact on the IMDB score.** Task: Analyse the distribution of movie durations and identify the relationship between movie duration and IMDB score.

Hint: Calculate descriptive statistics such as mean, median, and standard deviation for movie durations. Use Excel's functions like AVERAGE, MEDIAN, and STDEV. 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.

**DESCRIPTIVE STATISTICS FOR MOVIE DURATION**

duration		
Mean	109.9241	
Standard Error	0.366804	
Median	106	
Mode	101	
Standard Deviation	22.75365	
Sample Variance	517.7286	
Kurtosis	12.28119	
Skewness	2.338397	
Range	296	
Minimum	34	
Maximum	330	
Sum	422988	
Count	3848	



**INFERENCE**-On average duration of movie is 109 minutes. Median is 106 and mode 101. Maximum time duration of a movie is 330 minutes and the minimum duration is 34. There is little positive correlation between IMDB score and duration of the movie.

### 3. 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.

Hint: Use Excel's COUNTIF function to count the number of movies for each language.

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

Compare the statistics to understand the impact of language on movie ratings.

Language	No. of Movies	Mean	Median	Mode	SD	Variance
English	3668	6.423909487	6.5	6.7	1.048751	1.099878
Mandarin	14	7.021428571	7.25	7.6	0.765786	0.586429
Aboriginal	2	6.95	6.95	#N/A	0.777817	0.605
Spanish	26	7.05	7.25	7.2	0.826196	0.6826
French	37	7.286486486	7.25	7.2	0.561329	0.31509
Filipino	1	6.7	7.25	#N/A	#DIV/0!	#DIV/0!
Maya	1	7.8	7.25	#N/A	#DIV/0!	#DIV/0!
Kazakh	1	6	7.25	#N/A	#DIV/0!	#DIV/0!
Telugu	1	8.4	7.25	#N/A	#DIV/0!	#DIV/0!
Cantonese	8	7.2375	7.25	7.3	0.440576	0.194107
Japanese	12	7.625	7.25	#N/A	0.899621	0.809318
Aramaic	1	7.1	7.25	#N/A	#DIV/0!	#DIV/0!
Italian	7	7.185714286	7.25	#N/A	1.155319	1.334762
Dutch	3	7.566666667	7.25	7.8	0.404145	0.163333
Dari	2	7.5	7.25	#N/A	0.141421	0.02
German	13	7.692307692	7.25	7.4	0.640913	0.410769
Mongolian	1	7.3	7.25	#N/A	#DIV/0!	#DIV/0!
Thai	3	6.633333333	7.25	#N/A	0.450925	0.203333
Bosnian	1	4.3	7.25	#N/A	#DIV/0!	#DIV/0!
Korean	5	7.7	7.25	#N/A	0.570088	0.325
Hungarian	1	7.1	7.25	#N/A	#DIV/0!	#DIV/0!
Hindi	10	6.76	7.25	#N/A	1.111755	1.236
Icelandic	1	6.9	7.25	#N/A	#DIV/0!	#DIV/0!
Danish	3	7.9	7.25	#N/A	0.52915	0.28
Portuguese	5	7.76	7.25	#N/A	0.978775	0.958
Norwegian	4	7.15	7.25	7.6	0.574456	0.33
Czech	1	7.4	7.25	#N/A	#DIV/0!	#DIV/0!
Russian	1	6.5	7.25	#N/A	#DIV/0!	#DIV/0!
None	1	8.5	7.25	#N/A	#DIV/0!	#DIV/0!
Zulu	1	7.3	7.25	#N/A	#DIV/0!	#DIV/0!
Hebrew	3	7.5	7.25	#N/A	0.43589	0.19
Dzongkha	1	7.5	7.25	#N/A	#DIV/0!	#DIV/0!
Arabic	1	7.2	7.25	#N/A	#DIV/0!	#DIV/0!
Vietnamese	1	7.4	7.25	#N/A	#DIV/0!	#DIV/0!
Indonesian	2	7.9	7.25	#N/A	0.424264	0.18
Romanian	1	7.9	7.25	#N/A	#DIV/0!	#DIV/0!
Persian	3	8.133333333	7.25	#N/A	0.550757	0.303333
Swedish	1	7.6	7.25	#N/A	#DIV/0!	#DIV/0!

**INFERENCE-** More number of movies are made in English language. A non-verbal movie and some movies by Persian Language and Indonesian language movies though produce less movies have IMDB score near and above 8. So, the stories of such language's night have attracted the audience and reviewers' attention. More movies and remakes can be made from them.

- 4. Director Analysis: Influence of directors on movie ratings.** Task: Identify the top directors based on their average IMDB score and analyse their contribution to the success of movies using percentile calculations. Hint: Calculate the average IMDB score for each director. Use Excel's PERCENTILE function to identify the directors with the highest scores. Compare the scores of these directors to the overall distribution of scores.

Row Labels	Average of imdb_score
⊕ Alfred Hitchcock	8.5
⊕ Asghar Farhadi	8.4
⊕ Charles Chaplin	8.6
⊕ Christopher Nolan	8.425
⊕ Damien Chazelle	8.5
⊕ Majid Majidi	8.5
⊕ Marius A. Markevicius	8.4
⊕ Richard Marquand	8.4
⊕ Ron Fricke	8.5
⊕ S.S. Rajamouli	8.4
⊕ Sergio Leone	8.433333333
⊕ Tony Kaye	8.6
<b>Grand Total</b>	<b>8.452380952</b>
LARGE	8.6
PERCENT RANK	0.909
PERCENTILE	8.5999

**INTERPRETATION-**The top ten directors with highest average IMDB scores were taken. Based on IMDB score the success rate of a movie is 90 percent from the director Tony Kaye and Charles Chaplin whose average IMDB score is 8.6.

- 5. Budget Analysis: Explore the relationship between movie budgets and their financial success.** Task: Analyse the correlation between movie budgets and gross earnings, and identify the movies with the highest profit margin.

Hint: 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 and identify the movies with the highest profit margin using Excel's MAX function.

<b>CORRELATION</b>	0.100850218
<b>Director name</b>	James Cameron
<b>HIGHEST PROFIT MARGIN</b>	523505847

**INFERENCE-** There is relatively weak positive correlation between budget and gross income. The highest profit margin is around 52 crore.

## **RESULTS**

- Maximum count of movies is Drama followed by Comedy genre. Lowest is film Noir Genre. War and History genre though have less count in movies but their ratings are more than 7 on average. It means people like movies based on facts and history.
- On average duration of movie is 110 minutes. Median is 106 and mode 101. Maximum time duration of a movie is 330 minutes and the minimum duration is 37. There is little positive correlation between IMDB score and duration of the movie.
- More number of movies are made in English language. A non-verbal movie and some movies by Persian Language and Indonesian language movies though produce less movies have IMDB score near and above 8. So, the stories of such language's might have attracted the audience and reviewers' attention. More movies and remakes can be made from them.
- The top ten directors with highest average IMDB scores were taken. Based on IMDB score the success rate of a movie is 90 percent from the director Tony Kaye and Charles Chaplin whose average IMDB score is 8.6.
- There is a relatively weak positive correlation between budget and gross income. The highest profit margin is around 52 crores.