

# ANUSHKA GHARAGE

## PROJECT 5: IMDB Movie Analysis

### Project Overview:

The project focused on analyzing IMDB movie data to understand the factors influencing a movie's success, with success defined by high IMDB ratings. The primary objectives were to identify patterns, correlations, and trends within the dataset, providing insights for filmmakers, producers, and investors to make informed decisions. The analysis covered aspects such as genres, directors, budgets, and their impact on IMDB ratings.

### Approach:

**Data Collection:** Utilized a dataset containing information about IMDB movies, including features like genres, directors, budgets, and IMDB ratings.

**Data Cleaning:** Cleaned the dataset by handling missing data.

**Feature Selection:** Identified influential features for further analysis, such as genres, directors, and budgets.

**Statistical Analysis:** Conducted Statistical Analysis to understand the distribution of IMDB ratings, explore correlations, and identify trends in different features.

**Visualization and Reporting:** Created visualizations to present key findings to enhance understanding.

### Tech-Stack Used:

**Software:** Microsoft Excel 2016

**Purpose:** Excel was used for data cleaning and statistical analysis. Excel's functions, including COUNTIF, AVERAGE, MEDIAN, STDEV, and others, were crucial for various calculations and analyses.

### Insights:

#### **Movie Genre Analysis:**

- Discovered that certain genres tend to have higher IMDB ratings, influencing the overall success of movies.
- Most common genres in the dataset were Drama, Comedy, Thriller, Action and Romance.

#### **Movie Duration Analysis:**

- On an average Movie durations tend to be around **107 to 120 minutes (around 2 hours approximately)**.
- There isn't much relation between movie duration and IMDB score.

#### **Language Analysis:**

- The most common languages used in movies are **English, French and Spanish**.

- And these movies have an IMDB score between 6-7 which is more than average score and thus determines more success than movies in other languages.

#### **Director Analysis:**

- Identified directors with consistently high average IMDb scores, indicating the impact of directors on a movie's success.

#### **Budget Analysis:**

- Explored the correlation between movie budgets and gross earnings, highlighting movies with the highest profit margins.
- Top 5 highest grossing movies are **Avatar, Jurassic World, Titanic, Star Wars and E.T. the Extra.**

#### **Results:**

- The project provided valuable insights for decision-makers in the film industry, including recommendations for genres, directors, and budget considerations.
- Achieved a better understanding of the factors contributing to a movie's success on IMDb, enabling stakeholders to make more informed decisions in their future projects.
- Strengthened data analysis skills, particularly in using tools like Excel for exploring and interpreting complex datasets.

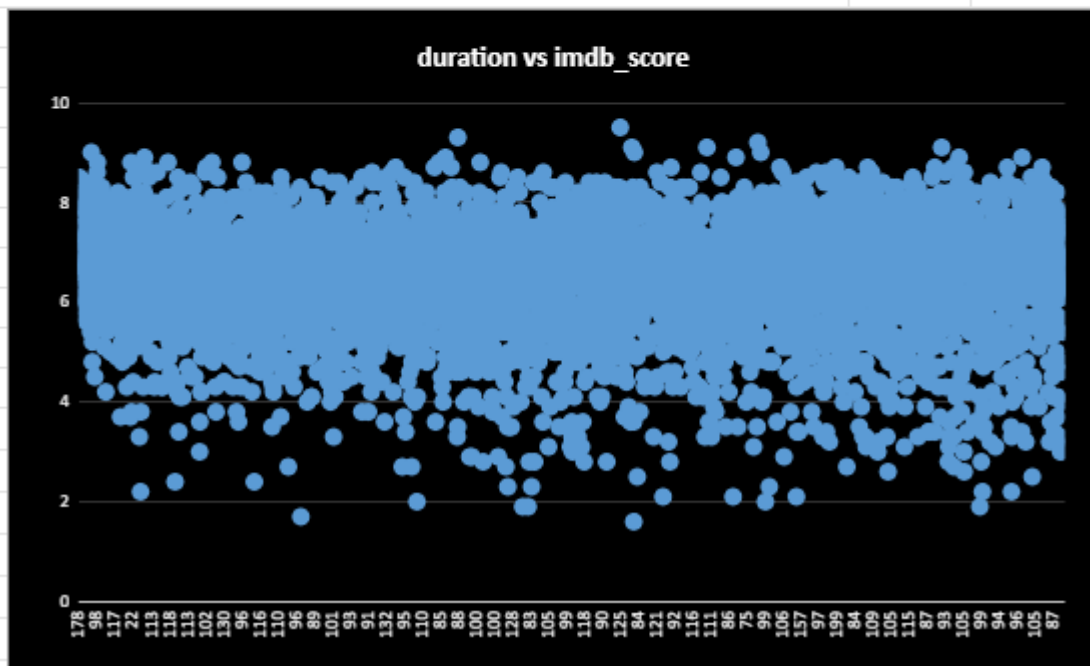
**Task-1:** Determine the most common genres of movies in the dataset. Then, for each genre, calculate descriptive statistics (mean, median, mode, range, variance, standard deviation) of the IMDB scores.

AX	AY	BI	BJ	BK	BL	BM	BN
Distinct genres	No. of movies	Mean	Median	Mode	Range	Variance	Std. Var.
Drama	2594	6.846480	6.9	7.2	7.3	0.916173	0.957169
Comedy	1872	6.287782	6.3	6.7	7.8	1.189021	1.090422
Thriller	1411	6.122844	6.4	6.1	6.8	1.110831	1.053960
Action	1153	6.239895	6.3	6.1	7.4	1.250706	1.118349
Romance	1107	6.476837	6.5	6.5	6.5	0.991189	0.995585
Adventure	923	6.442689	6.6	6.7	7	1.278218	1.130583
Crime	889	6.393854	6.6	6.6	6.9	1.052427	1.025878
Sci-Fi	616	6.250483	6.4	6.7	6.9	1.463695	1.209832
Fantasy	610	6.349329	6.4	6.7	7.2	1.344983	1.159734
Horror	565	5.879629	5.9	6.2	6.5	1.275697	1.129467
Family	546	6.329901	6.4	6.7	7	1.441193	1.200497
Mystery	500	6.456039	6.6	6.6	6.4	1.187375	1.089667
Biography	293	7.095609	7.2	7	4.4	0.520247	0.721281
Animation	242	6.612516	6.7	6.7	6.9	1.293309	1.137237
Music	214	6.257149	6.6	6.5	6.9	1.383165	1.176080
War	213	6.811018	7.1	7.1	5.9	0.761519	0.872650
History	207	7.175356	7.2	7.5	6.9	0.784561	0.885754
Sport	182	6.207907	6.8	7.2	6.7	1.207600	1.098908
Musical	132	6.493948	6.7	7	6.4	1.491003	1.221066
Documentary	121	7.204802	7.4	7.5	7.1	1.107044	1.052161
Western	97	6.710129	6.8	6.5	5.1	1.075563	1.037093
Film-Noir	6	7.605555	7.65	#N/A	1.1	0.155555	0.394405
Short	5	6.5	6.5	#N/A	1.9	0.4456	0.667532

- The most common genres are Drama, Comedy, Thriller, Action and Romance.
- The green highlighted table shows descriptive statistics of IMDb score.

**Task-2:** Analyze the distribution of movie durations and identify the relationship between movie duration and IMDB score.

AD	AE	AF
Mean duration		
107.201074		
Median duration		
103		
Standard deviation of durations		
25.19493497		



- On an average movies tend to be 107 minutes(1hr 45mins) long.
- The dataset contained movies ranging from 22minutes to more than 3hours duration but all were evenly distributed, mostly concentrated between 6-8 IMdb score. There is no specific relation between movie duration and IMDB score.

**Task-3:** Determine the most common languages used in movies and analyze their impact on the IMDB score using descriptive statistics.

AD	AE	AF	AG	AH	AI
language	No. of movies		Mean	Median	Std. Var.
English	4704		6.398426871	6.6	1.121948655
Japanese	18		7.394444444	7.6	0.962907762
French	73		7.038356164	7.2	0.721989287
Mandarin	26		6.788461538	7.05	1.021810958
Aboriginal	2		6.95	6.95	0.55
Spanish	40		6.9375	7.15	0.844300746
Filipino	1		6.7	6.7	0
Hindi	28		6.632142857	6.95	1.37374711
Russian	11		6.363636364	6.5	1.319278541
Maya	1		7.8	7.8	0
Kazakh	1		6	6	0
Telugu	1		8.4	8.4	0
Cantonese	11		6.954545455	7.2	0.671989767
Icelandic	2		7.55	7.55	0.65
German	19		7.342105263	7.6	0.928675224
Aramaic	1		7.1	7.1	0
Italian	11		7.227272727	7.3	1.186354929
Dutch	4		7.425	7.45	0.376662979
Dari	2		7.5	7.5	0.1
Hebrew	5		7.58	7.6	0.299332590
Chinese	3		5.666666667	5.7	0.449691252
Mongolian	1		7.3	7.3	0
Swedish	5		7.44	7.6	0.677052435
Korean	8		7.3875	7.5	0.772071078
Thai	3		6.633333333	6.6	0.368178700
Polish	4		8.25	8.25	0.85

- The top 3 most languages used in movies are English, French and Spanish.
- The movies in these top languages had average IMDb score between 6-7 which indicates that most people will like movies in these languages.

**Task-4:** Identify the top directors based on their average IMDB score and analyze their contribution to the success of movies using percentile calculations.

AD	AE	AF	AG	AH	AI	AJ
director_name	No. of movies	Mean			98th percentile	Directors with IMDB score above 98th percentile
John Blanchard	1	9.5			8.3	John Blanchard
Mitchell Altieri	1	8.7				Mitchell Altieri
Sadyk Sher-Niyaz	1	8.7				Sadyk Sher-Niyaz
Cary Bell	1	8.7				Cary Bell
Mike Mayhall	1	8.6				Mike Mayhall
Charles Chaplin	1	8.6				Charles Chaplin
Raja Menon	1	8.5				Raja Menon
Ron Fricke	1	8.5				Ron Fricke
Damien Chazelle	1	8.5				Damien Chazelle
Majid Majidi	1	8.5				Majid Majidi
Sergio Leone	4	8.475				Sergio Leone
Christopher Nolan	8	8.425				Christopher Nolan
S.S. Rajamouli	1	8.4				S.S. Rajamouli
Moustapha Akkad	1	8.4				Moustapha Akkad
Richard Marquand	1	8.4				Richard Marquand
Catherine Owens	1	8.4				Catherine Owens
Rakeysh Omprakash Mehra	1	8.4				Rakeysh Omprakash Mehra
Jay Oliva	1	8.4				Jay Oliva
Robert Mulligan	1	8.4				Robert Mulligan
Asghar Farhadi	1	8.4				Asghar Farhadi
Marius A. Markevicius	1	8.4				Marius A. Markevicius
Bill Melendez	1	8.4				Bill Melendez
Lee Unkrich	1	8.3				Lee Unkrich

- Top directors based on their average IMDb score are John Blanchard, Mitchell Altieri, Sadyk Sher-Niyaz, Cary Bell and Mike Mayhall.
- The output provides a list of directors whose movies consistently achieve IMDb scores higher than a significant portion of other movies in the dataset.
- Directors with higher average IMDb scores contribute significantly to the success of their movies, as reflected by their positions in the upper percentiles.

**Task-5:** Analyze the correlation between movie budgets and gross earnings, and identify the movies with the highest profit margin.

AG	AH	AI	AJ	AK
Correlation coefficient	Profit Margin		TOP 5	Movie Name
0.1021794535	523505847		523505847	Avatar
	9404152		502177271	Jurassic World
	-44925825		458672302	Titanic
	198130642		449935665	Star Wars: Episode IV - A New Hope
	-190641321		424449459	E.T. the Extra-Terrestrial
	78530303			
	-59192738			
	208991599			
	51956980			
	80249062			
	-8930592			
	-31631573			
	198032628			
	-125710090			
	66021565			
	-83385977			
	403279547			
	-8936125			
	-45979146			
	5108370			
	32030663			
	-94780265			
	33355354			
	-109916481			
	11051260			
	458672302			

- The correlation coefficient between movie budgets and gross earnings is approximate 0.102, which means success or gross earning of a movie does not depend on the movie budget.
- Movies with highest profit margin are Avatar, Jurassic world, Titanic, Star Wars and E.T. the Extra.

Excel Sheet:

[IMDb Movies](#) (download pdf for hyperlink)