# IMDB Movie Analysis

**Name : Tanmay Dangat**

**Project 5 : IMDB Movie Analysis**

- **Project Description:** As per the instructions and report, I have assigned a task to perform the data analysis of the IMDB Movie Analysis.

We have to work on the provided dataset and collect the useful insights. The primary goal of this project is to identify the key factors that contribute to a movie's success on IMDB, by high ratings. The project was designed to conduct comprehensive data analytics on the IMDB movie analysis, with objectives of gaining key aspects of the dataset, including movie genres, duration, language, directors, and budget.

Approach :To accomplish the project objectives, a systematic approach was followed. A dataset containing relevant information on movie genres, duration, language, directors, and budget was obtained.

Data Preparation: I began by downloading the dataset provided for the project. The dataset contains information about movie titles, genres, durations, languages, directors, budgets, gross earnings, and IMDB scores. I have done data cleansing activity whenever required and started working on this.

Data analysis: Using excel , we have performed various analyses on the dataset This involved manipulating the data to extract relevant information, calculating descriptive statistics, and visualizing relationships between different variables.

Report Generation: Based on the analysis results, we created a detailed report to present our findings. The report includes insights derived from each analysis task, supported by visualizations and statistical summaries.

# IMDB Movie Analysis

**Tech Stack used : Software: Microsoft Excel 2022**
**Purpose: Excel was chosen for its data analysis functionalities, including pivot tables, charts, and statistical functions. It helps in visualization of the IMDB Movie analysis, enabling a thorough understanding of the underlying trends and patterns**

**Insights: As a beginner, it helped me to understand how the complex problem statement works and how to understand the business and that insights actually works.**

**Results: Following are all the results**

# IMDB Movie Analysis

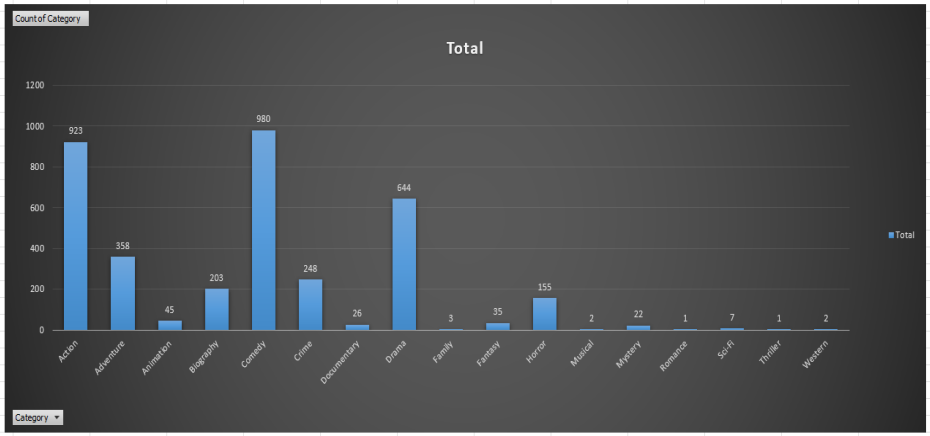
## 1. Data Analysis Task

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

| Category | Count of Category |
|---|---|
| Action | 923 |
| Adventure | 358 |
| Animation | 45 |
| Biography | 203 |
| Comedy | 980 |
| Crime | 248 |
| Documentary | 26 |
| Drama | 644 |
| Family | 3 |
| Fantasy | 35 |
| Horror | 155 |
| Musical | 2 |
| Mystery | 22 |
| Romance | 1 |
| Sci-Fi | 7 |
| Thriller | 1 |
| Western | 2 |
| Grand Total | 3655 |

| Genres Maniplutaion | | |
|---|---|---|
| Median | 35 | |
| Max | 980 | Comedy |
| Min | 1 | |
| Mode | 2 | |
| Var | 100036.4706 | |
| Sd | 316.2854258 | |



- **Genre Distribution**

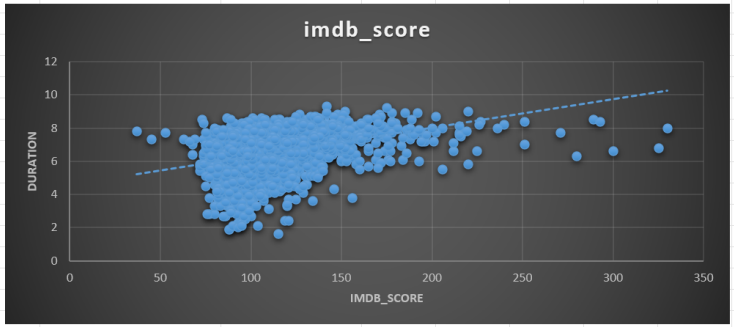
As per the Bar Graph, the most common genres in the dataset that had also impacted on their IMDB scores. Action, Biography, Crime, Comedy and Drama emerged as the most prevalent genres.

# IMDB Movie Analysis

**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.
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

| | A | B | C | D | E | F | G | H | I | J | K |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | duration | imdb_score | | | | | | | | | |
| 2 | 178 | 7.9 | | **Avr of Movie Duration** | **110.2** | | | | | | |
| 3 | 169 | 7.1 | | **Median** | **106.0** | | | | | | |
| 4 | 148 | 6.8 | | **SD** | **22.7** | | | | | | |
| 5 | 164 | 8.5 | | | | | | | | | |
| 6 | 156 | 6.2 | | | | | | | | | |
| 7 | 100 | 7.8 | | | | | | | | | |
| 8 | 141 | 7.5 | | | | | | | | | |
| 9 | 153 | 7.5 | | | | | | | | | |
| 10 | 183 | 6.9 | | | | | | | | | |
| 11 | 169 | 6.1 | | | | | | | | | |
| 12 | 106 | 6.7 | | | | | | | | | |
| 13 | 151 | 7.3 | | | | | | | | | |
| 14 | 150 | 6.5 | | | | | | | | | |
| 15 | 143 | 7.2 | | | | | | | | | |
| 16 | 150 | 6.6 | | | | | | | | | |
| 17 | 173 | 8.1 | | | | | | | | | |
| 18 | 136 | 6.7 | | | | | | | | | |
| 19 | 106 | 6.8 | | | | | | | | | |
| 20 | 164 | 7.5 | | | | | | | | | |



- **Distribution of Movies Duration**

**Insights of the above scatter plot:** The distribution of movie durations varied widely, ranging from short films to lengthy epics. However, we observed a weak correlation between movie duration and IMDB score, suggesting that movie length alone does not significantly influence audience ratings.

From the above scatter plot their is a slight possibility that a movie can score 6 - 8.5 IMDB score if movie duration is more than 250 minutes. With the help of trend line we can also predict IMDB score can increase with respect with the duration of the movie

# IMDB Movie Analysis

**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.
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 | Count of language | imdb_sc |
|---|---|---|
| English | 3498 | 7.9 |
| French | 34 | 7.1 |
| Spanish | 23 | 6.8 |
| Mandarin | 14 | 8.5 |
| Japanese | 10 | 6.2 |
| German | 10 | 7.8 |
| Italian | 7 | 7.5 |
| Cantonese | 7 | 7.5 |
| Portuguese | 5 | 6.9 |
| Korean | 5 | 6.1 |
| Hindi | 5 | 6.7 |
| Norwegian | 4 | 7.3 |
| Persian | 3 | 6.5 |
| Danish | 3 | 7.2 |
| Thai | 3 | 6.6 |
| Dutch | 3 | 8.1 |
| Aboriginal | 2 | 6.7 |
| Dari | 2 | 6.8 |
| Indonesian | 2 | 7.5 |
| Kazakh | 1 | 7 |
| Filipino | 1 | 6.7 |
| Romanian | 1 | 7.9 |

| Language | Mean | Median | Sd |
|---|---|---|---|
| Aboriginal | 6.4 | 6.6 | 1.1 |
| Arabic | 6.5 | 6.6 | 1.1 |
| Aramaic | 6.5 | 6.6 | 1.1 |
| Bosnian | 6.7 | 6.6 | 1.1 |
| Cantonese | 6.7 | 6.6 | 1.1 |
| Czech | 6.8 | 6.6 | 1.1 |
| Danish | 6.5 | 6.6 | 1.1 |
| Dari | 7.0 | 6.6 | 1.1 |
| Dutch | 6.8 | 6.6 | 1.1 |
| English | 6.5 | 6.6 | 1.1 |
| Filipino | 6.9 | 6.6 | 1.1 |
| French | 6.8 | 6.6 | 1.1 |
| German | 6.0 | 6.6 | 1.1 |
| Hebrew | 7.6 | 6.6 | 1.1 |
| Hindi | 7.3 | 6.6 | 1.1 |
| Hungarian | 6.2 | 6.6 | 1.1 |
| Indonesian | 6.2 | 6.6 | 1.1 |
| Italian | 6.0 | 6.6 | 1.1 |
| Japanese | 6.2 | 6.6 | 1.1 |
| Kazakh | 3.3 | 6.6 | 1.1 |
| Korean | 7.9 | 6.6 | 1.1 |
| Mandarin | 6.9 | 6.6 | 1.1 |

- **Insights of the above Pivot table: English was the predominant language in the dataset, followed by French and Spanish. Interestingly, movies in languages other than English tended to have slightly lower IMDB scores on average, indicating potential biases in the audience preferences**

# IMDB Movie Analysis

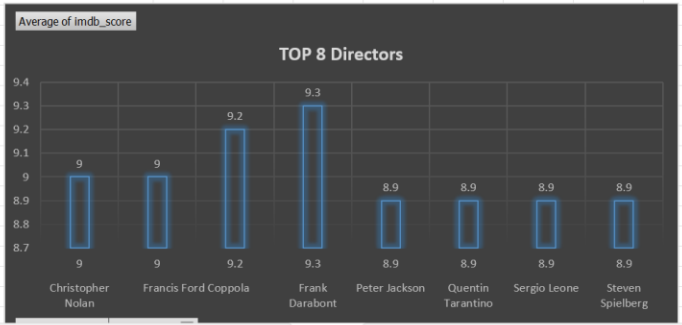
**D. Director Analysis:** Influence of directors on movie ratings.

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

•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.

| director_name | imdb_score | Average of imdb_score |
|---|---|---|
| Christopher Nolan | | 9 |
| | 9 | 9 |
| Francis Ford Coppola | | 9.1 |
| | 9 | 9 |
| | 9.2 | 9.2 |
| Frank Darabont | | 9.3 |
| | 9.3 | 9.3 |
| Peter Jackson | | 8.9 |
| | 8.9 | 8.9 |
| Quentin Tarantino | | 8.9 |
| | 8.9 | 8.9 |
| Sergio Leone | | 8.9 |
| | 8.9 | 8.9 |
| Steven Spielberg | | 8.9 |
| | 8.9 | 8.9 |
| **Grand Total** | | **9.0125** |

Perentile 8.6

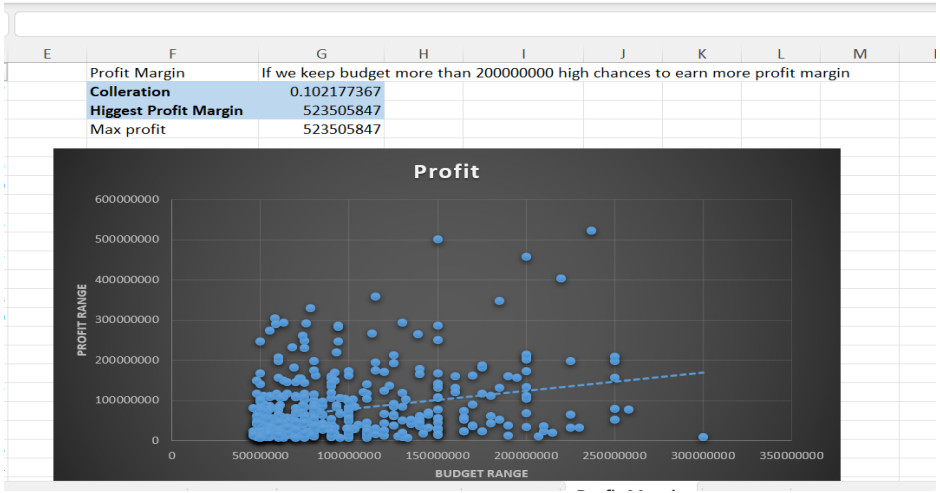


- **Insights of the above Bar Graph: According to the analysis the top directors based on their average IMDB scores and analyzed their contribution to movie success. Directors such as Francis Ford Coppola and Frank Darabont consistently produced movies with high IMDB ratings, reflecting their strong influence on audience perception.**

# IMDB Movie Analysis

**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.
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.



- **Insights of the above scatter plot** : As per the task assigned in IMDB movie analysis,the relationship between movie budgets and financial success by analyzing the correlation between budgets and gross earnings. While there was a positive correlation between the two variables, the strength of the relationship varied depending on other factors such as genre and director.
- If we keep the budget more than 200000000 high chances to earn more profit margin

- **Results :** This project has uncovered valuable insights into what makes movies successful in terms of ratings and financial performance. We looked at different aspects like genres, movie lengths, languages, directors, and budgets. By analyzing these factors, we found trends and patterns that help us understand what audiences like and how the movie industry works. Our discoveries offer practical advice for filmmakers, producers, and studios who want to make better decisions about their movie productions and attract more viewers

- **Drive Link:**
- **Recording Link: https://www.loom.com/share/a6998dd659dc400db67ff2c46d24830b?sid=4b25e218-3a4d-4ae6-8e37-03a4563d41f4**