**Project Name: - IMDB Movie Analysis**

**Done by: - Arpan Paul**

**GitHub Link: -**

**Project Description**

This project focuses on Analyzing various attributes of movies, such as genre, duration, language, director influence, and budget, to understand their impact on IMDB scores and financial success. The primary objectives are to identify which genres, durations, languages, and directors are most associated with high IMDB ratings and to examine the correlation between movie budgets and gross earnings. The approach involves using descriptive statistics and visualizations in Microsoft Excel 2019 to gain a comprehensive understanding of the factors influencing movie ratings and profitability.

**Approach**

The project is divided into five primary analytical tasks:

1. **Movie-Genre-Analysis**  
   I analyze the distribution of movie genres and calculated the average IMDB score for each genre. This analysis helped identify the most popular genres and assess whether certain genres tend to have higher IMDB scores. I used Excel’s COUNTIF function to tally the number of movies per genre. Then, I used Excel’s statistical functions—AVERAGE, MEDIAN, MODE, VAR, STDEV, and MAX/MIN—to gather descriptive statistics for each genre's IMDB scores. These statistics allowed for a comparative analysis to understand the general rating trends across genres.
2. **Movie-Duration-Analysis**  
   For this task, I calculated descriptive statistics (mean, median, and standard deviation) for movie durations and examined the correlation between duration and IMDB score. A scatter plot was created to visualize this relationship, with a trendline added to assess the strength and direction of any correlation. Excel’s AVERAGE, MEDIAN, and STDEV functions provided insights into typical movie durations, and the trendline on the scatter plot helped indicate any relationship between duration and rating.
3. **Language-Analysis**  
   The distribution of languages in movies was examined to see if language impacts IMDB scores. Using the COUNTIF function, I counted the number of movies in each language. For each language, I calculated the mean, median, and standard deviation of IMDB scores. This analysis provided an understanding of the rating variation across languages and identified the languages associated with the highest-rated movies.
4. **Director-Analysis**  
   Here, I calculated the average IMDB score for each director and identified the top-performing directors based on their scores. Using Excel's PERCENTILE function, I determined directors with IMDB scores above the 90th percentile to highlight those who consistently produced high-rated movies. This analysis shed light on the directors’ contributions to movie success and revealed how top directors compare to the overall distribution.
5. **Budget-Analysis**  
   I explored the relationship between movie budgets and gross earnings, calculating the correlation coefficient between the two using Excel's CORREL function. Additionally, I calculated the profit margin (gross earnings - budget) and used the MAX function to identify movies with the highest profit margins. This task highlighted any trends in budget allocation versus revenue generation, showing how financial investment impacts profitability.

**Tech-Stack Used**

* **Microsoft Excel 2019**: Chosen for its powerful data analysis and visualization capabilities. Excel was used to apply statistical functions (such as AVERAGE, MEDIAN, STDEV, VAR, COUNTIF, CORREL, and PERCENTILE), create scatter plots, and add trendlines. Excel’s ease of use and extensive analytical tools made it ideal for extracting meaningful insights from the dataset.

**Insights**

The project offered valuable insights into movie trends and patterns:

* **Genres**: Certain genres consistently achieved higher average IMDB scores, suggesting genre-specific audience preferences.
* **Duration**: The analysis indicated a moderate positive relationship between movie duration and IMDB scores, implying that longer movies may have a slight advantage in ratings.
* **Languages**: English was the most common language; however, certain foreign-language movies showed higher average ratings, likely due to niche audiences and storytelling styles.
* **Directors**: Top-rated directors consistently produced highly rated movies, emphasizing the influence of direction on movie quality.
* **Budget and Profitability**: A strong correlation between budget and gross earnings was observed, along with the insight that higher budgets often correspond with high grossing, though profit margins varied significantly.

**Result**

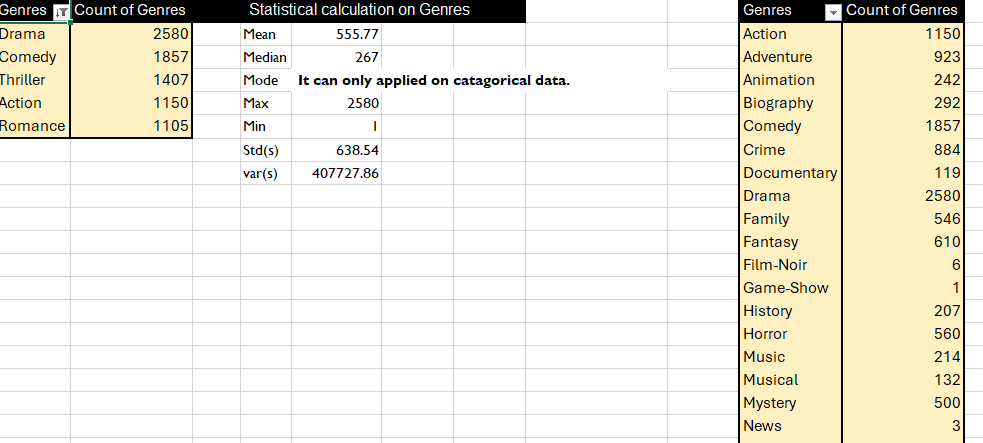
The project successfully highlighted the factors that contribute most to high IMDB scores and profitability in the movie industry. By understanding the genre, duration, language, director influence, and budget, I gained a more detailed view of what contributes to a movie’s success. This analysis deepened my understanding of how these factors impact movie performance, providing valuable insights that could inform future movie production and marketing decisions.

**IMDB Movie Analysis Tasks:**

You are required to provide a detailed report for the below data record mentioning the answers of the questions that follows:

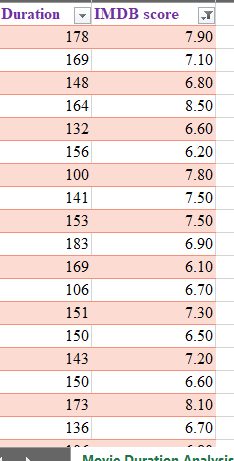
A. **Movie Genre Analysis:** Analyze the distribution of movie genres and their impact on the IMDB score.

* **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.
* **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.



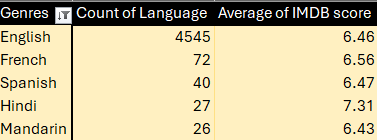
**B. Movie Duration Analysis:**Analyze the distribution of movie durations and its impact on the IMDB score.

* 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.

****

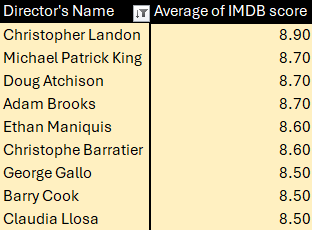
**C. Language Analysis:**Situation: Examine the distribution of movies based on their language.

* **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.



**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.



**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.

