##### **IMDB Movie Analysis**

**Description:**

**Problem Statement**: The dataset provided is related to IMDB Movies. A potential problem to investigate could be: "What factors influence the success of a movie on IMDB?" Here, success can be defined by high IMDB ratings. The impact of this problem is significant for movie producers, directors, and investors who want to understand what makes a movie successful to make informed decisions in their future projects.

**Data Cleaning**: This step involves preprocessing the data to make it suitable for analysis. It includes handling missing values, removing duplicates, converting data types if necessary, and possibly feature engineering.

**Data Analysis**: Here, you'll explore the data to understand the relationships between different variables. You might look at the correlation between movie ratings and other factors like genre, director, budget, etc. You might also want to consider the year of release, the actors involved, and other relevant factors.

**Five 'Whys' Approach**: This technique will help you dig deeper into the problem. For instance, if you find that movies with higher budgets tend to have higher ratings, you can ask "Why?" repeatedly to uncover the root cause. Here's an example:

* Q: "Why do movies with higher budgets tend to have higher ratings?"
* A: They can afford better production quality.
* Q: "Why does better production quality lead to higher ratings?"
* A: It enhances the viewer's experience.
* Q: "Why does an enhanced viewer experience lead to higher ratings?"
* A: Viewers are more likely to rate a movie highly if they enjoyed watching it.
* Q: "Why are viewers more likely to rate a movie highly if they enjoyed watching it?"
* A: Positive experiences lead to positive reviews.
* Q: "Why do positive reviews matter?"
* A: They influence other viewers' decisions to watch the movie, increasing its popularity and success.

**Report and Data Story**: After your analysis, you'll create a report that tells a story with your data. This should include your initial problem, your findings, and the insights you've gained. Use visualizations to help tell your story and make your findings more understandable.

Remember, as a data analyst, your goal is not just to answer questions but to provide insights that can drive decision-making. Your analysis should aim to provide actionable insights that can help stakeholders make informed decisions.

**Data Analytics 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:** 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.

**B. 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.

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

* **Task:** Determine the most common languages used in movies and analyse 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 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.

**E. 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.