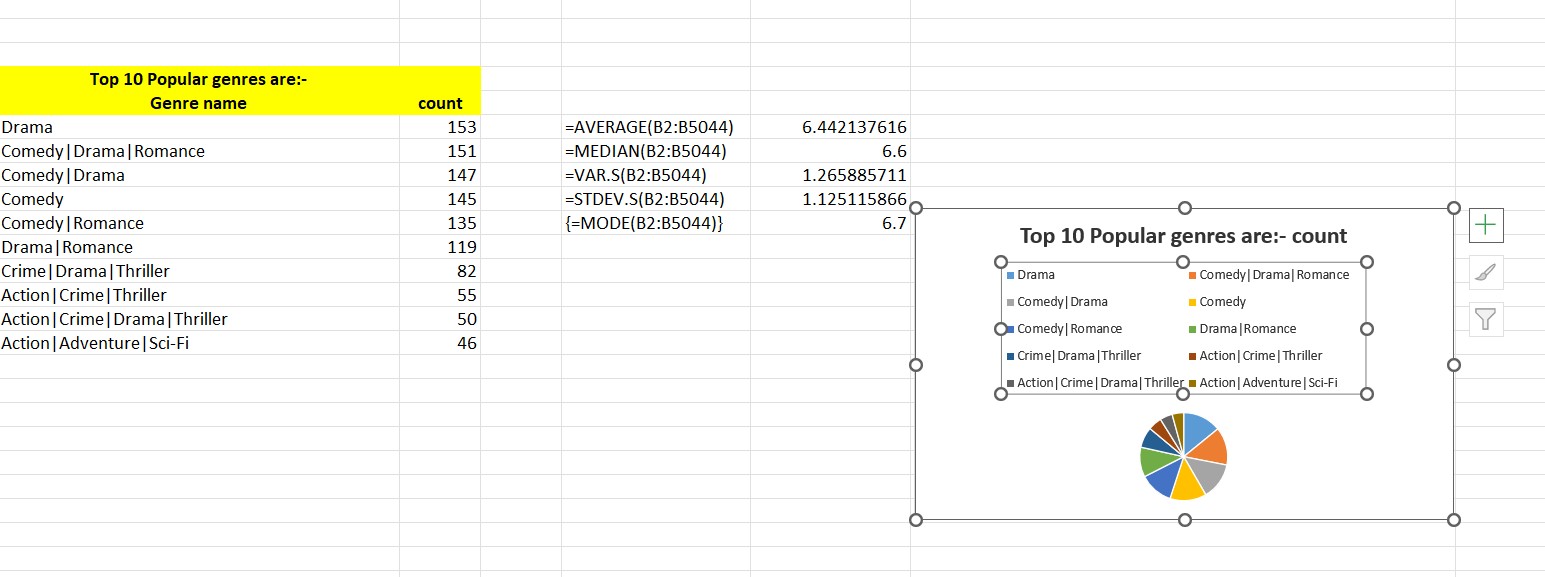
**DESCRIPTION:**

1. **Movie Genre Analysis:**
   * To determine the most common genres in the dataset.
   * To calculate descriptive statistics (mean, median, mode, range, variance, standard deviation) for IMDB scores within each movie genre.
   * Understand how genre impacts movie ratings.
2. **Movie Duration Analysis:**
   * To analyze the distribution of movie durations.
   * Explore the relationship between movie duration and IMDB score.
   * Create a scatter plot with a trendline to assess this relationship.
3. **Language Analysis:**
   * Identify the most common languages used in movies.
   * Calculate descriptive statistics (mean, median, standard deviation) of IMDB scores for each language.
   * Evaluate language’s impact on movie ratings.
4. **Director Analysis:**
   * Find top directors based on their average IMDB scores.
   * Assess their contribution to movie success using percentile calculations.
   * Compare director scores to the overall distribution.
5. **Budget Analysis:**
   * Explore the correlation between movie budgets and gross earnings.
   * Calculate profit margin ie; gross earnings – budget for each movie.
   * Identify movies with the highest profit margin.

**APPROACH:**

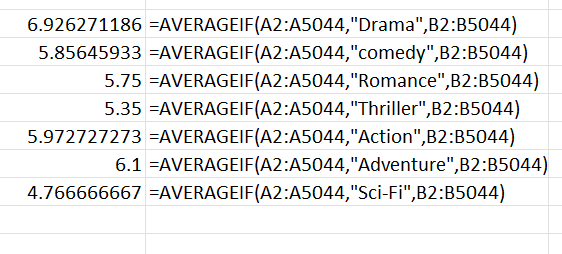
**A. Movie Genre Analysis:** Analyze the distribution of movie genres and their impact on the IMDB score.

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.

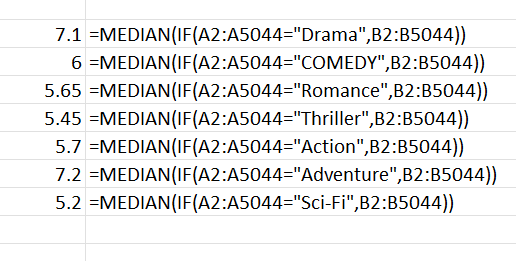
The below is the list of top 10 popular genres.

Below are the average, median, mode, variance and standard for the above top 10 popular genres.

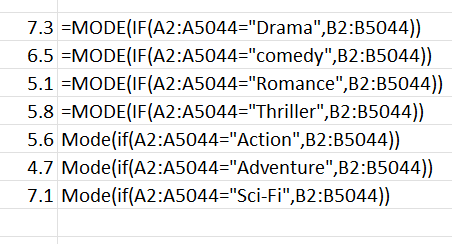
**AVERAGE**: using =averageif(cells,”genre”,cells).



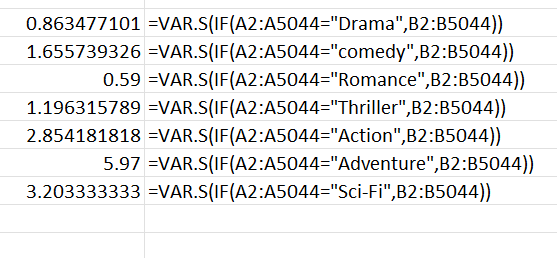
**MEDIAN:** using =median(if(cells=”genre”,cells)).



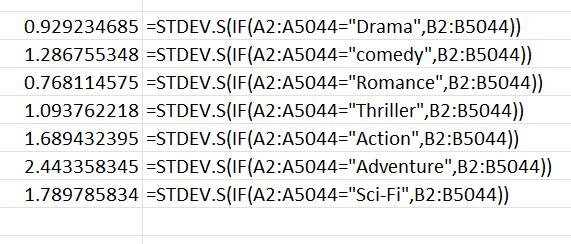
**MODE:** using mode(if(cells=”genre”,cells))



VARIANCE: using var.s(if(cells=”genre”,cells))



**STANDARD DEVIATION:** using =stedv.s(if(cells=”genre”,cells))



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

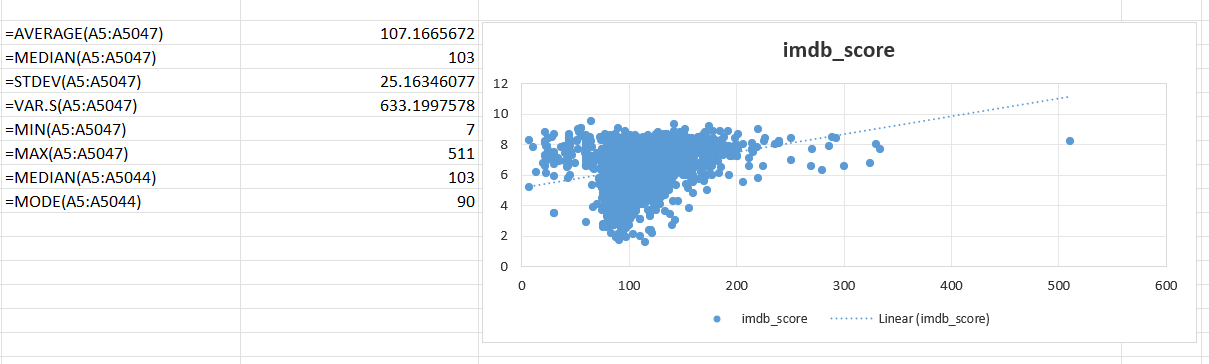
* Task: Analyze the distribution of movie durations and identify the relationship between movie duration and IMDB score.

**Relationship Between Duration and IMDB Score:**

Create a scatter plot:

* X-axis: Movie duration.
* Y-axis: IMDB score

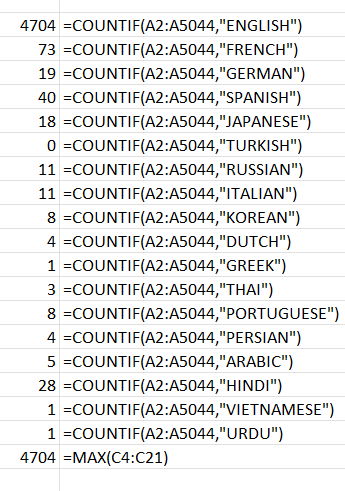
The below are the standard deviation, average and median of the duration of the movies.



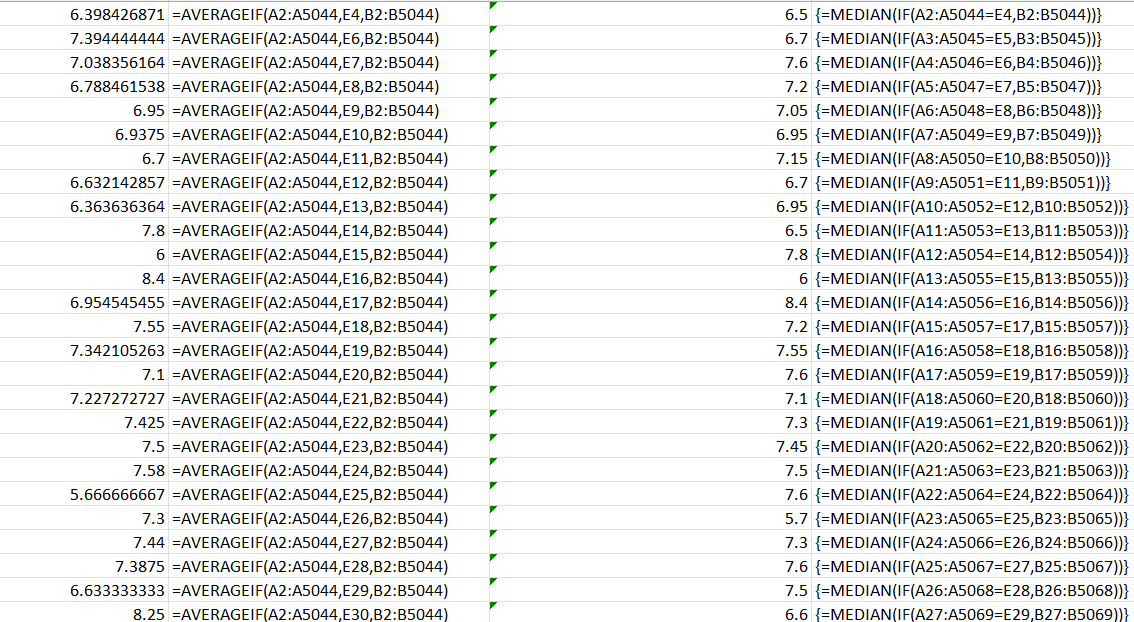
**C. Language Analysis:**Situation: Examine the distribution of movies based on their language.

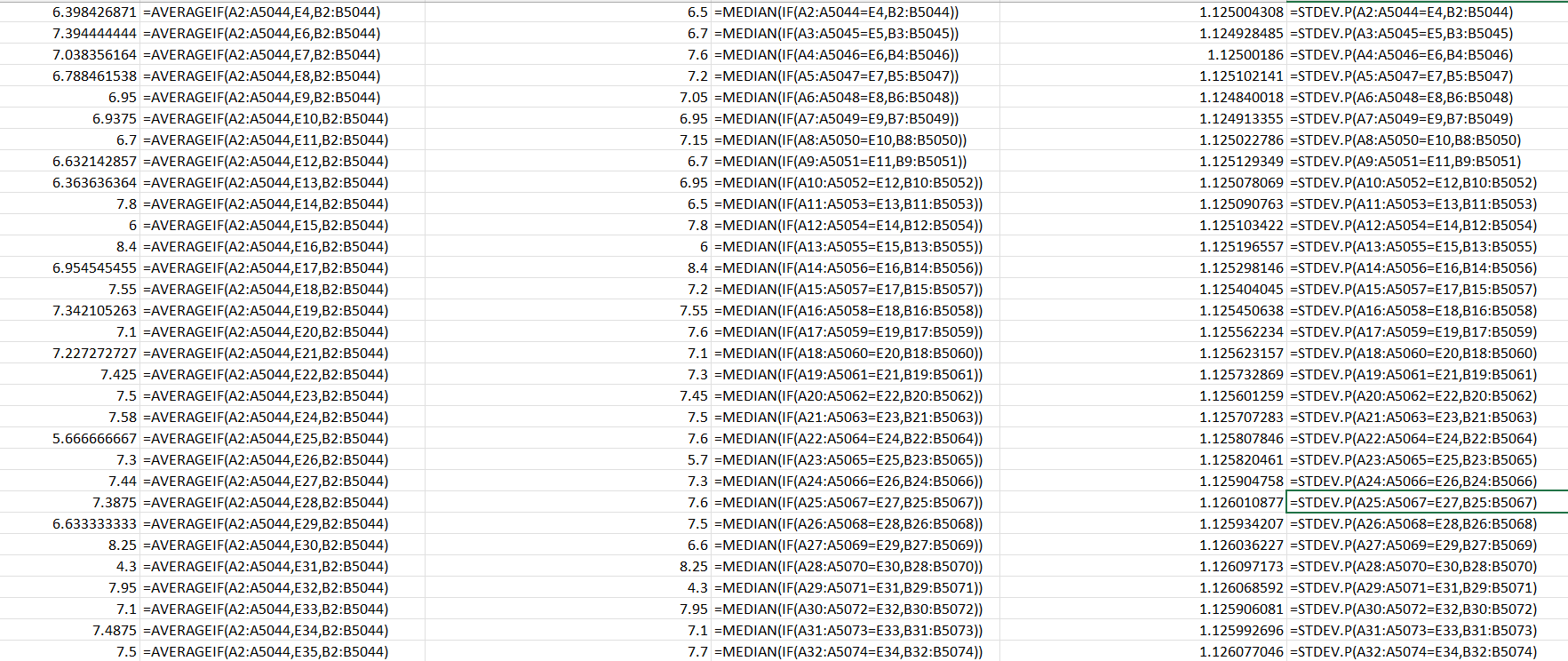
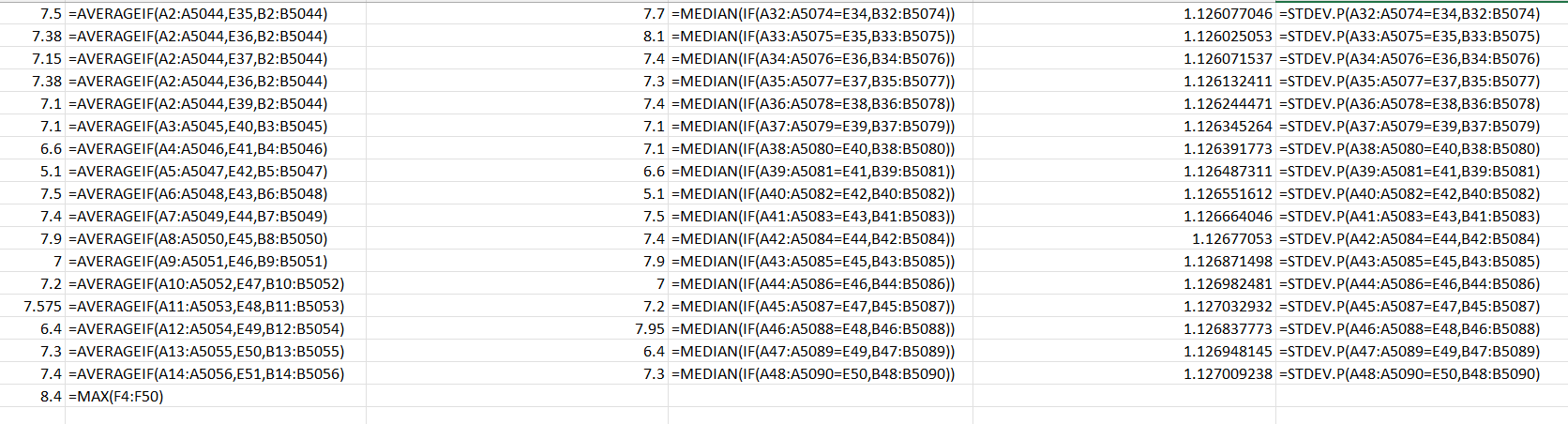
* **Task:** Determine the most common languages used in movies and analyze their impact on the IMDB score using descriptive statistics.

The below are the number of movies from specific languages. I used countif function to find the number of movies in each language.

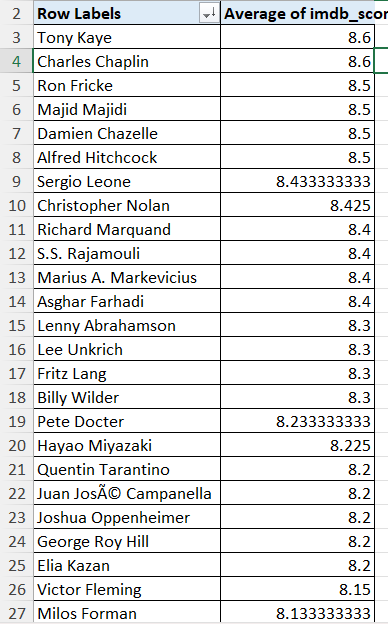


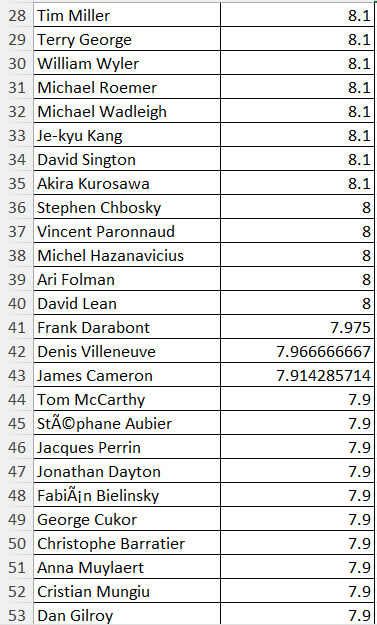
The below are the mean, median and standard deviation for the above mentioned languages.

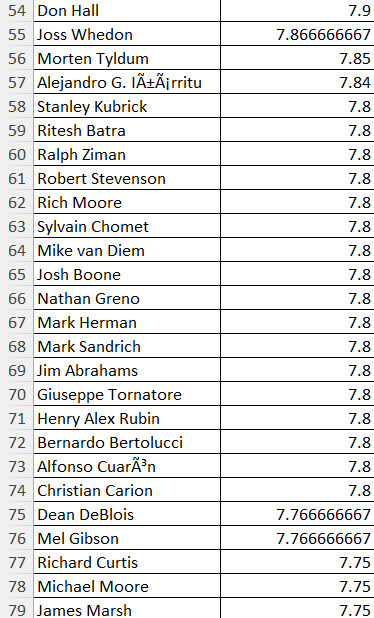




**D. Director Analysis:**Influence of directors on movie ratings.

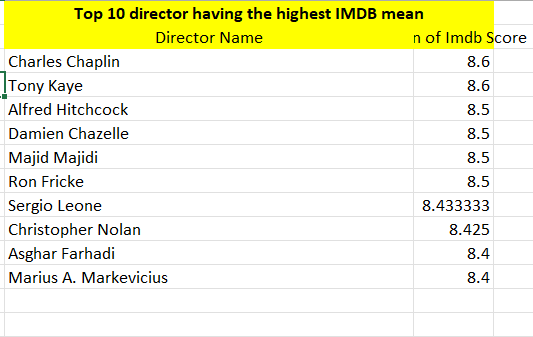
* Task: Identify the top directors based on their average IMDB score and analyze their contribution to the success of movies using percentile calculations.
* 



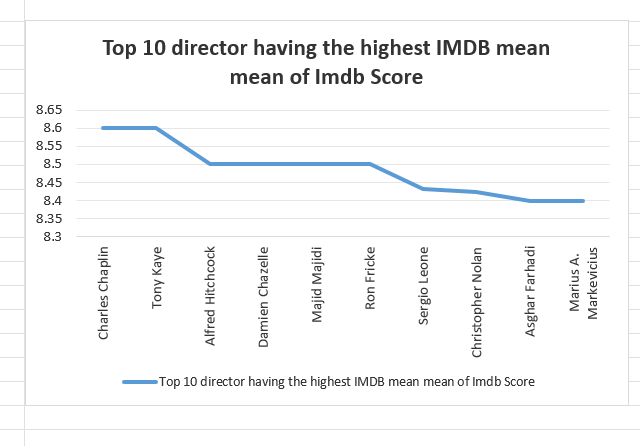


The above are the averages of IMDB score of each director.

The below are the averages of IMDB score of top 10 directors.



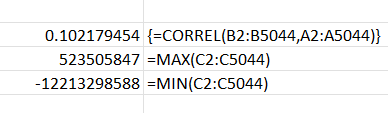
Here is the graphical representation of the IMDB score of top 10 directors.



**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.

These are the gross earnings and movie budget using max, min and correlation function.



**TECH- STACK USED:** I used MS Excel to apply the function keys and get the needed output. In MS Excel I could make pivot tables, graphs, relations from the given dataset.

**INSIGHTS: Certainly! Let’s dive into each task and provide the needed insights:**

**A. Movie Genre Analysis:**

1. **Most Common Genres:**
   * **Identify the genres with the highest frequency (using COUNTIF).**
   * **These genres are likely popular among filmmakers and audiences.**
2. **Descriptive Statistics for IMDB Scores by Genre:**
   * **Calculate the following statistics for each genre:** 
     + **Mean (Average): Indicates the central tendency of scores within a genre.**
     + **Median: Represents the middle value, minimizing the impact of outliers.**
     + **Mode: Identifies the most frequent score.**
     + **Range: Shows the spread from the lowest to the highest score.**
     + **Variance: Measures the dispersion of scores.**
     + **Standard Deviation: Indicates how scores deviate from the mean.**
3. **Impact of Genre on Movie Ratings:**
   * **Compare the statistics across genres.**
   * **Genres with higher mean and median scores may indicate greater audience appreciation.**

**B. Movie Duration Analysis:**

1. **Distribution of Movie Durations:**
   * Calculate descriptive statistics for movie durations:
     + Mean: Average duration.
     + Median: Middle value, less affected by outliers.
     + Standard Deviation: Measure of dispersion.
   * Visualize the distribution using graphs.
2. **Relationship Between Duration and IMDB Score:**
   * Create a scatter plot:
     + X-axis: Movie duration.
     + Y-axis: IMDB score.
   * Add a trendline to assess the direction and strength of the relationship.
   * Interpret:
     + Positive slope: Longer movies tend to have higher IMDB scores.
     + Negative slope: Shorter movies may be favoured by audiences.

**C. Language Analysis:**

1. **Commonly Used Languages:**
   * Count the number of movies for each language.
   * Identify the languages with the highest occurrence.
2. **Impact of Language on Movie Ratings:**
   * Calculate descriptive statistics (mean, median, standard deviation) of IMDB scores for each language.
   * Compare the statistics across languages.
   * Languages with higher average scores could point to cultural preferences or production quality.

**D. Director Analysis:**

1. **Top Directors by IMDB Score:**
   * Calculate the average IMDB score for each director.
   * Use Excel’s PERCENTILE function to identify directors with the highest scores.
   * Directors in higher percentiles likely have a consistent track record of successful movies.
2. **Contribution to Movie Success:**
   * Compare the scores of top directors to the overall distribution.
   * Assess how influential these directors are in shaping movie ratings.

**E. Budget Analysis:**

1. **Correlation Between Budget and Gross Earnings:**
   * Use Excel’s CORREL function to find the correlation coefficient.
   * A strong positive correlation may indicate that higher budgets contribute to financial success.
2. **Profit Margin Calculation:**
   * Calculate profit margin for each movie: (Gross Earnings - Budget).
   * Identify movies with the highest profit margin using Excel’s MAX function.

**RESULT:** This is the most interesting task I’ve done so far. I’ve learned how to combine theoretical knowledge with practical knowledge. I’ve learned to translate complex data insights into understandable and actionable information.