IMDB MOVIE ANALYSIS

PROJECT 05 TRAINITY [EXCEL FILE](https://docs.google.com/spreadsheets/d/1L2ry1N_qZF9-jNvnYEYb6rR5MwREEPRj/edit?usp=sharing&ouid=100859955328054714520&rtpof=true&sd=true)

**PROJECTDESCRIPTION:** In this project, I have used Excel as my tool to analyse the imdb movie datagiven in the project. My goal in this project is to find details insight from the imdb data .I found significant for movie producers, directors, and investors who want to understand what makes a movie successful to make informed decisions in their future projects.

**APPROACH:** I have looked into the data given in the project. Then I have downloaded the IMDB data given in the project. Then cleaned the data properly ,filled the missing value after that I have performed operation.

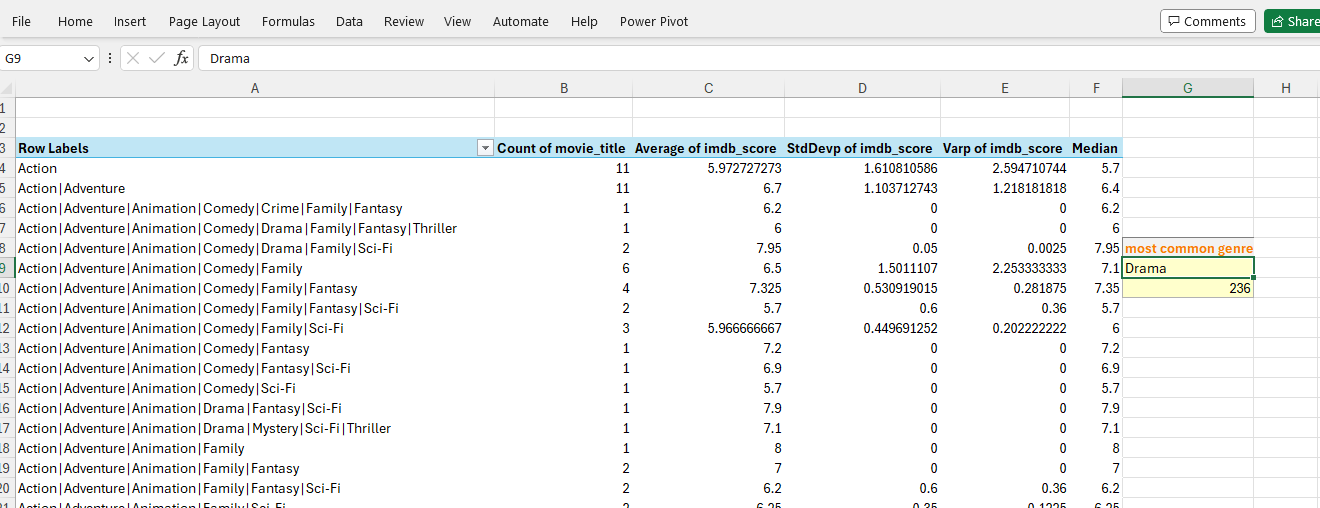
**DATA CLEANING:** At first I observe in the data. Then after looking the task I dropped the unnecessary columns. Then in the columns. If there is numbers I filled the blank cells with mean value. If there is string I filled the data with most used string used in the entire column .

**TECH-STACK USED:**Excel

**PROJECT DETAILS AND DATA ANALYSIS**

**Movie Genre Analysis:**

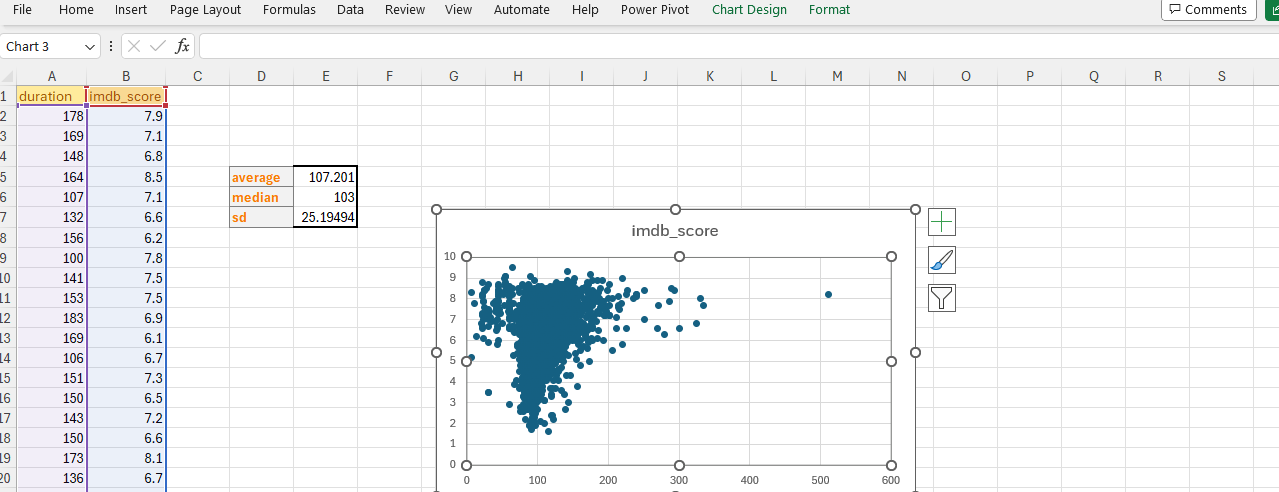
Here my task is to determine the most common genres of movies in the dataset. Then, for each genre, calculate descriptive statistics (mean, median, mode, range, variance, standard deviation) of the IMDB scores.. Here is the snapshot of the result on Excel.



Report: The most common genre is drama which is repeated 236 times in the data. Here I made a simple pivotable from original table the just calculated mean of imdb\_score ,counted movies, stdDev of imdb\_score,varp of imdb etc.With the help of power pivot I calculated the median also. With help of the data one can see which types of movie get more rating.

**Movie Duration Analysis:**

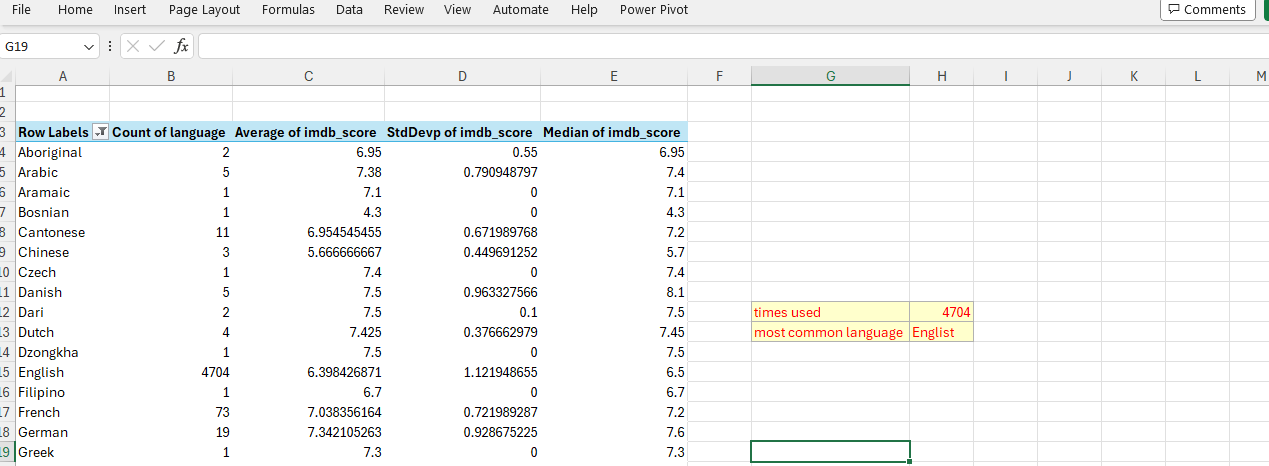
Here is my task to analyse the distribution of movie durations and identify the relationship between movie duration and IMDB score.. Here is the snapshot of the operation on Excel.



Report :Here we can clearly see the relationship between movie duration and imdb\_score by a scatter plot. I have also calculated average duration with top 10 imdb .That is 128 minutes .Movies with These length get highest rating.

**Language Analysis:**

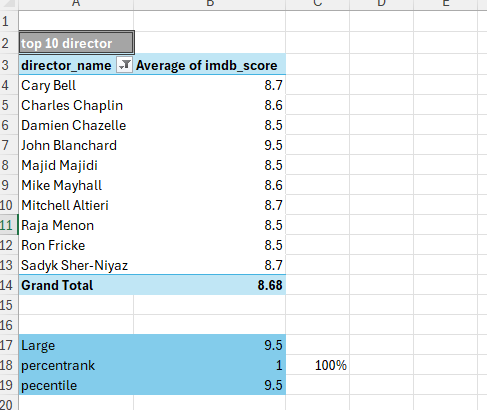
Here I have to determine the most common languages used in movies and analyse their impact on the IMDB score using descriptive statistics.



Report : Most common language is English which is used 4704 times. I made a pivot table with language and counting movie title and imdb\_score. I found that It is all about quality the impact the rating not the language .Tamil has highest average rating it has only one movie in the data. So we can not judge based on that.

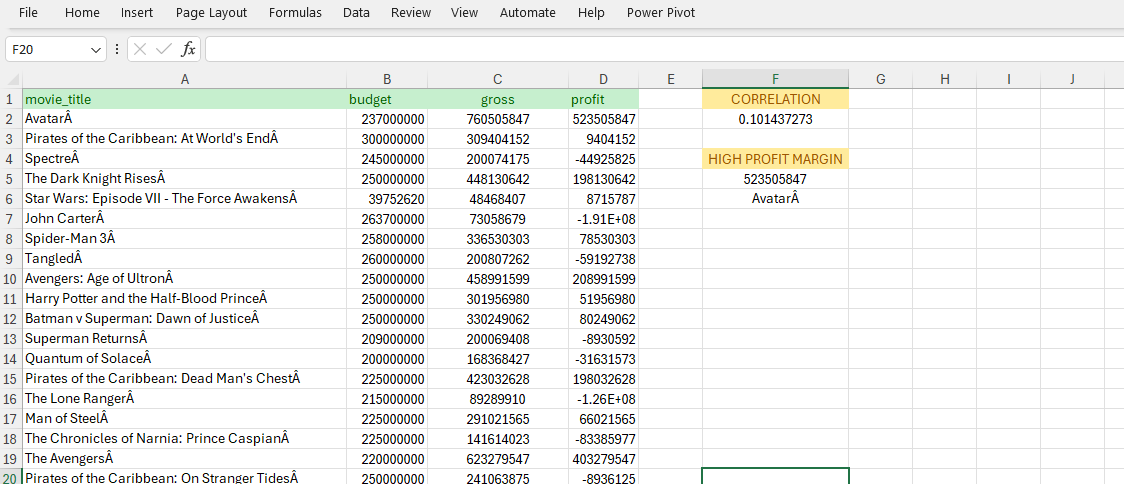
**Director Analysis:**   
My task is to Identify the top directors based on their average IMDB score and analyze their contribution to the success of movies using percentile calculations.

Here is the snapshot of to ten director with highest imdb ratings.



Report : Carry bell has been most successful director based on data. As per the list impact of a director for success of a movie is 100%.

**Budget Analysis:**   
Here I have to analyse the correlation between movie budgets and gross earnings, and identify the movies with the highest profit margin.



Report: Movie with highest profit is AvatarA .

Correlation between movie budget and profit is 0.101437273 which is significantly low. Profit depends on budget but not that much . Some time low budget movie earns a lot.

Result: After all the operation on excel I got that .Drama movies are highest in number, Movie with 2 hours and some minutes gets highest rating, a good director is must for a movie success, good budget is necessary but not always good budget movie do good on market.