**IMDB MOVIE ANALYSIS**

**Project Description:** The objective of this project is to conduct comprehensive analysis of the IMDB MOVIE dataset. This analysis focus on serval key aspect of the dataset like Genre, language, directors, Duration, Story and Budget etc. our main goal is to find a relation between all these aspect and how movie rating and success define on these aspect so that in future other movie directors can use this analysis for movie making and make a good movies and get their good rating and success.

**Approach:**  To complete the task I follow the structured approach

**Data preparation :** I began by downloading the dataset providing for this project. This dataset contain information about Genre, language, directors, duration, Story, Budget, total gross, imbd rating. First I clean the data then I start working on this project.

**Data Analysis:** I use excel to performed various analysis on the dataset to answer the question outlined in the project. This involve manipulation the data to extract relevant information, calculating statistics and visualization relationship between variables.

**Final report** : Once I complete all the task like find relationship between all the variable then I try to convert the insight into visual representation using graph, chart, report etc. and create the detail report of my finding.

**Tech-Stack Used :** For this project I use excel for extract, transform, clean, statistics approach, data visualization and summarizing the results. And use google drive to save and share the final report.

**Task A:**

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

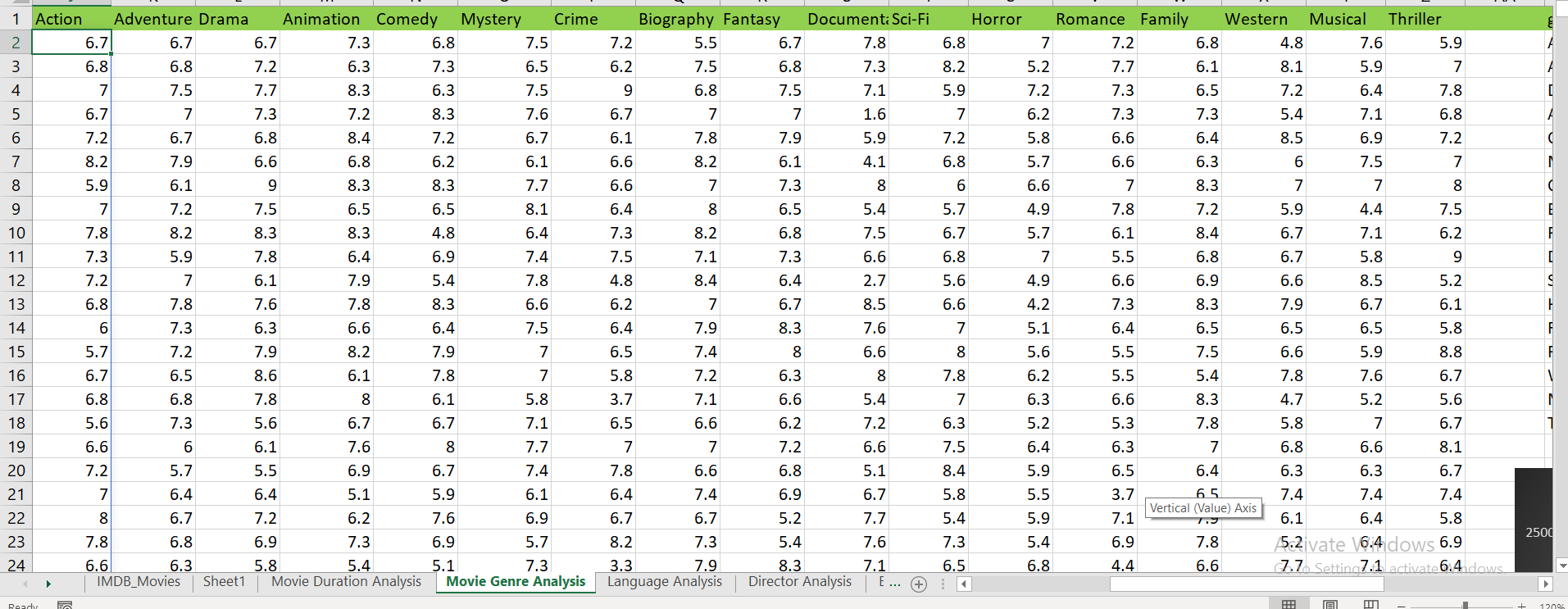
* **Task: Determine the most common genres of movies in the dataset. Then, for each genre, calculate descriptive statistics (mean, median, mode, range, variance, standard deviation) of the IMDB scores.**

Solution: To analyze the distribution the movie genres and their impact on the IMBD Score

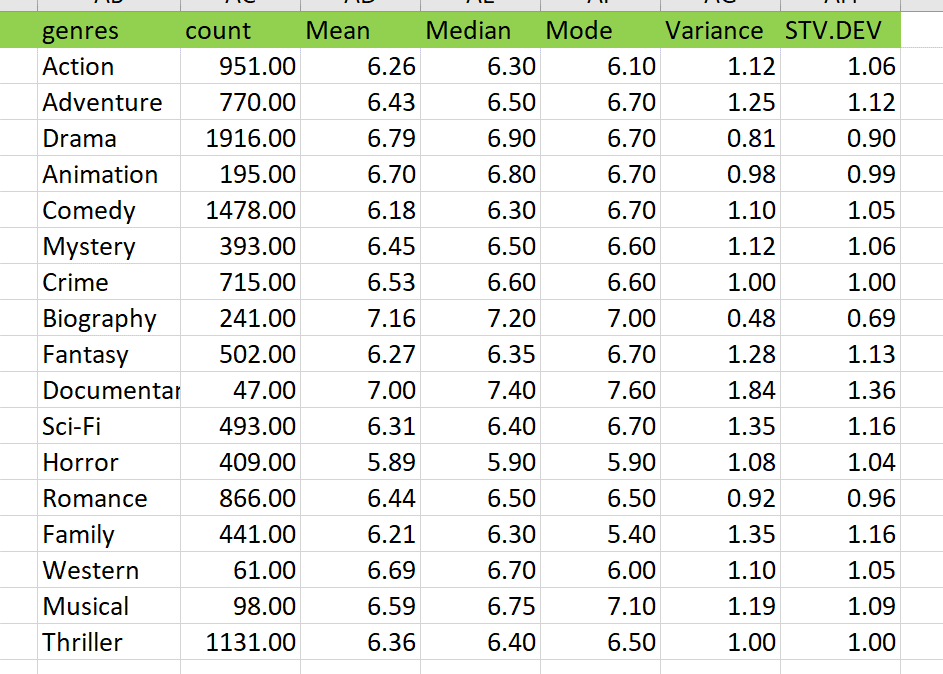
First we bring the Genre, IMBD in the different Sheet because only these two attribute we need for the analyze Movie Genre.

* We use the delimeter to split the column because genre column has multiple genre data in the single row .
* Once we split the column we use remove duplicate to get the total number of genre in the list and after remove the duplicate we find the there are 22 different genre in the list.
* Now we find the IMBD score for each movie using excel formula.

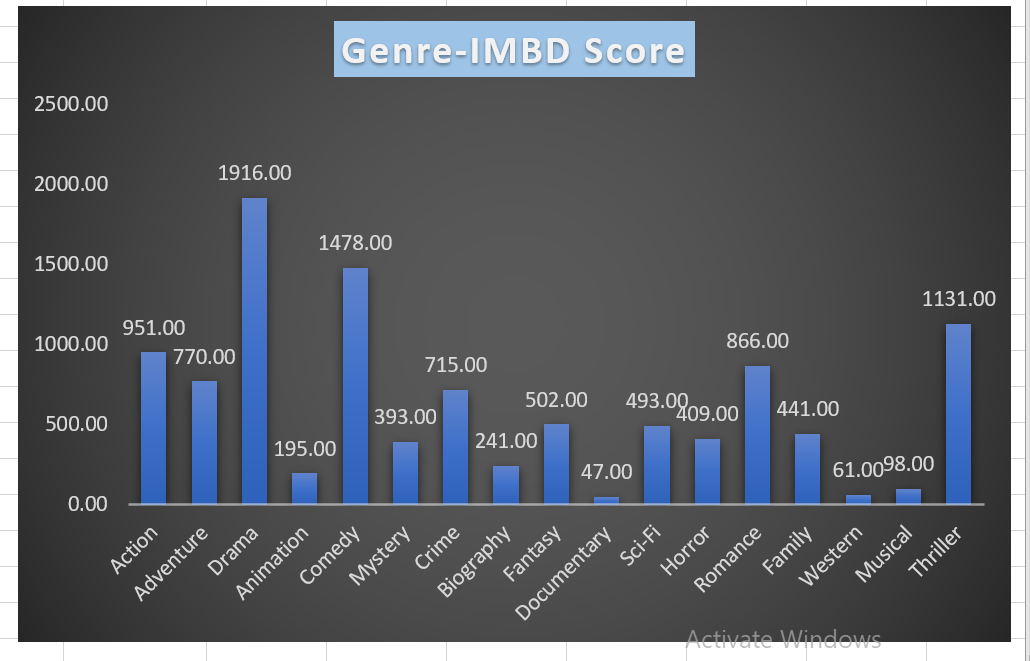
 by using this formula we generate IMBD Score for each Genre and movies.



* After all this we calculate the descriptive statistics for each Genre



* After all this we showed the relation between Movie and Genre count



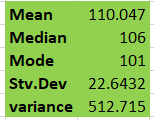
* Through this analysis we found Drama genre has a most number of movie and Comedy, Action, Thriller are second best after the drama.

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

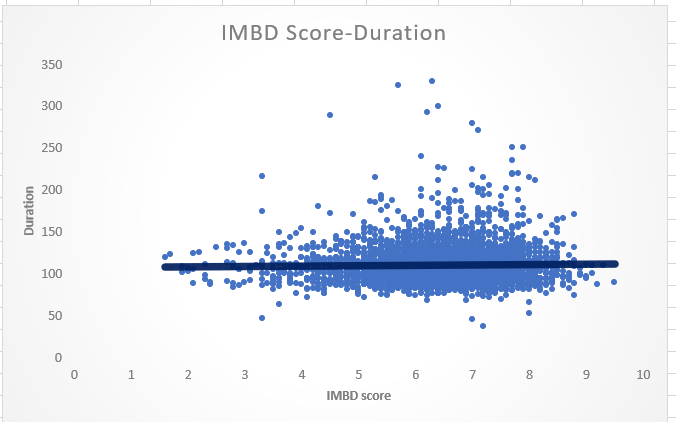
* **Task: Analyze the distribution of movie durations and identify the relationship between movie duration and IMDB score**.

Solution: To find the distribution of movie durations and its impact on the IMBD score first we bring the Duration and IMBD attribute on the different sheet.

* First we calculate the Descriptive Statistics for movie duration



* Now we find the relation between IMBD Score and Movie Duration.



* Through the analysis we find out average movie duration time is 110 min.
* Most of the movie who has average IMBD Score there movie duration is lies between 90 min to 150 min.

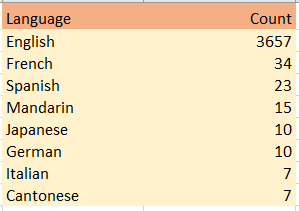
**C. Language Analysis: Situation: Examine the distribution of movies based on their language.**

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

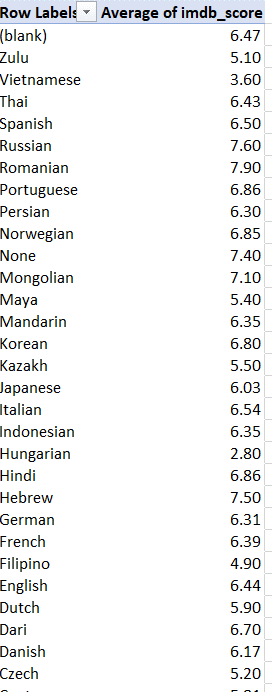
Solution: To Determine the most common language used in movies and analyze their impact on the IMBD first we bring the IMBD Score and Language attribute the spread sheet

Because we have lot of unnecessary column in the main data sheet and we have to work on only this two attribute.

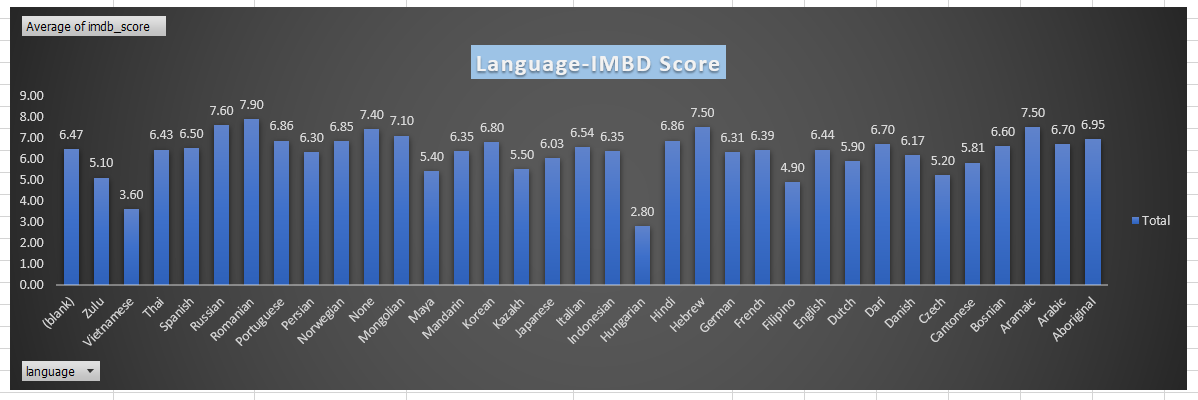
* After that insert pivot table to analyze the relation between these two .
* Then we find out English is the most common language used in movies. Its number is 3657 it is whey more higher than the other language which mean people mostly watching English movie.



* After that we find the average for each language on the basis of IMBD Score.



* Then we find the relation between Language and IMBD Score using column charts.



* After looking at this graph we can say there not any huge impact of language on the IMBD rating.

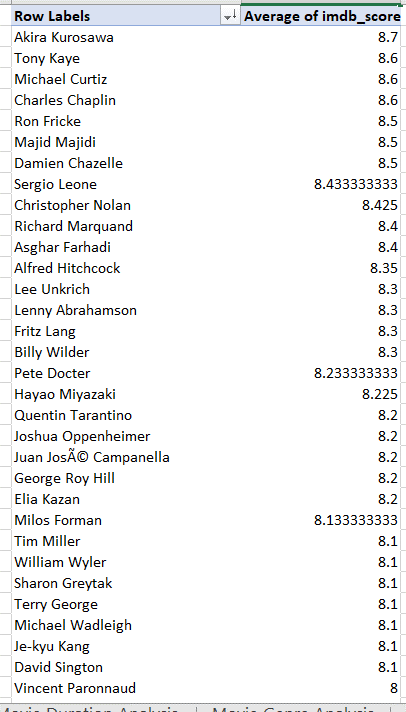
**D. Director Analysis: Influence of directors on movie ratings.**

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

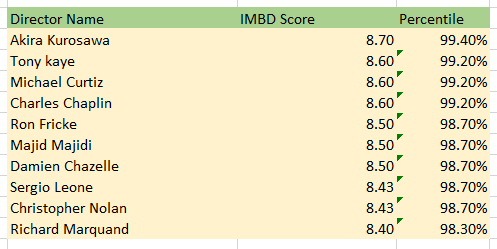
Solution: To find the relation between Director and IMBD score and analyze their contribution to the success first we fetch the director and IMBD score attribute to the different sheet.

* Now, we insert the pivot table on these attributes.
* After insert an pivot table we put Director in row attributes and IMBD score in values

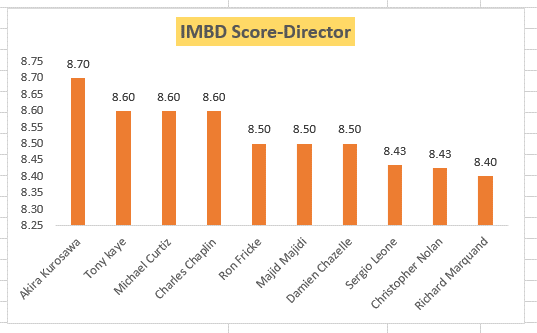
Then we convert it into Average that’s how we find Average of IMBD score for all Director.



* These are the average of all Directors.
* Now, we Separate the top director of all time from this list.



* These are top Director with IMBD score and Percentile.
* Now, we show the relation between IMBD score and Director using column charts.



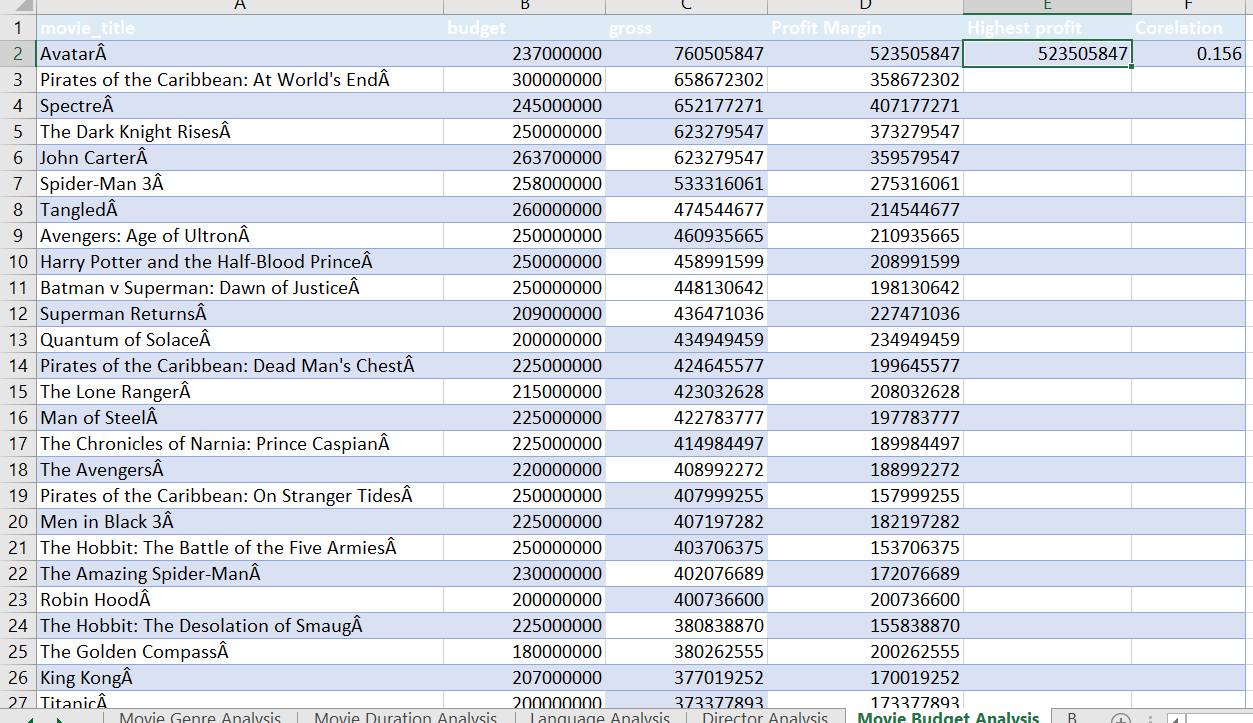
Notice: There is very mild impact on the basis of director.

**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**.

Solution: To Analyze the correlation between movie budgets and gross earning first we bring the Movie Title , Budget, Gross earnings to the different sheet.

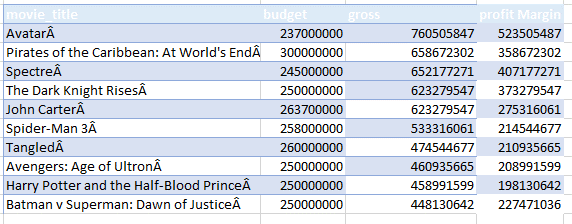
* Now, to find the Total profit we differentiate between Gross earning with Budgets. We get Total profit by doing this.



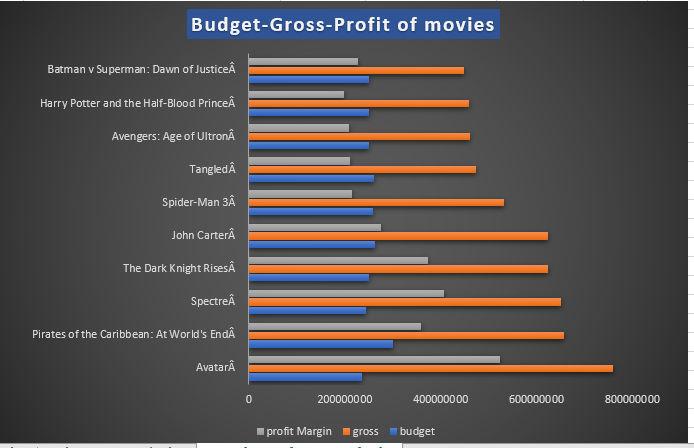
* Then we use Max function to get the highest profit movie from total profit.



* Now we do analysis on the basis of Budget , Gross earning, Total profit.



* Now, we show the relation with using Bar graph.



* Avatar Movie has highest profit among all the movies
* Correlation between Budget and Gross earning is shown using correl function which is 0.156.

**Result:**

* Average Movie Duration is 110 min.
* Drama Genre has most number of movies which is 1916.
* English language has most number of movie no language is near to English which is 3657.
* Avatar movie is the most profitable movie and has big budget.
* Director has barely impact on imbd score.