* Link of excel data:

<https://docs.google.com/spreadsheets/d/1uDMxBIwXHJTJpgJtnTwKhKYY9RBIBphn/edit?usp=sharing&ouid=117198255086466412144&rtpof=true&sd=true>

**IMDB Movie Analysis**

When it comes to watching a movie or series, most of us first go to google and search about IMDB ratings of the interested movies or series. And this project is about IMDB movie Analysis. Before starting the project, it is important to know a few things about IMDB, like what it is and how it works.

IMDB stands for Internet Movie Database, an online database that stores information about films, television series, podcasts, video games and online streaming content. It also includes cast, production crew, plot summaries, ratings and fan and critical reviews. Since 1998, it has been owned and operated by one of the subsidiaries of Amazon, IMDb.com Inc.

In this project, we will analyze the IMDb movie rating and find useful insights.

**PROJECT DESCRIPTION**

For this project, we are provided with a dataset of movies and all the related information like actor\_1\_name, director\_name, duration, movie\_title, genre, budget, gross, imdb\_score and so on. There are some extra information in the dataset that is not required, like actor2\_name, director\_facebook\_like, color, etc.

We need to answer the questions and derive insights using this dataset.

**Tools used:**

* Microsoft Excel 2016 for cleaning data and visualization of data.

**Problem:**

*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.

***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.

***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.

***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.

***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.

**Approach:**

**We need to first clean our data by using these steps.**

1. Removing all irrelevant columns from the dataset.

(Color , director\_facebook\_likes , actor\_3\_facebook\_likes, actor\_2\_name , actor\_1\_facebook\_likes , cast\_total\_facebook\_likes, actor\_3\_name , facenumber\_in\_posts , plot\_keywords, movie\_imdb\_link , content\_rating , actor\_2\_facebook\_likes, aspect\_ratio , movie\_facebook\_likes)

1. Delete the rows having null or missing values
2. Delete rows having duplicate values.

We will then make use of our cleaned data for analyzing and finding insights from it.

**Cleaned data excel file for reference:**

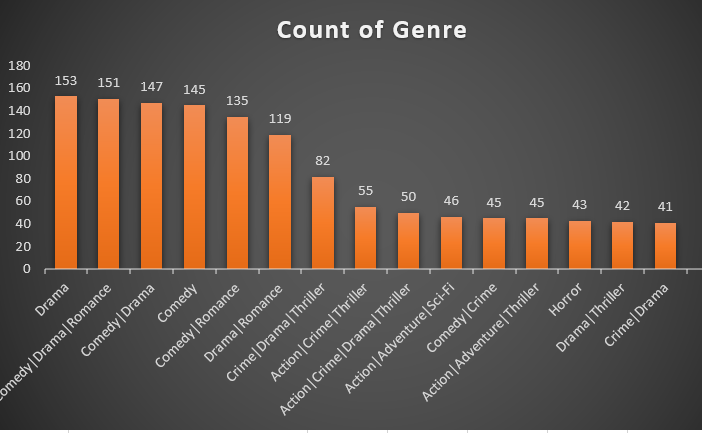
<https://docs.google.com/spreadsheets/d/1MLNTcazOLzVm9UGOjpA0ngjEiBwwuZfk/edit?usp=sharing&ouid=117198255086466412144&rtpof=true&sd=true>

**Findings**

1. **Problem-1**

To find the most common genres of the movies in the dataset,

* We first created a pivot table.
* Then chose the genres in the rows option and count of genres in the values option
* Next, sorted the count of genres in descending order and this gives us the idea of most common genres of movies.
* For better visual representation, we chose the top 15 common genres.

****

From the above diagram, it is clear that the genre ‘Drama’ is most common genre with count of 153.

To analyze the relation between the most common genres and IMDB Score,

* we find out the descriptive statistics of imdb\_score (i.e. Min, Max, Average, Standard deviation, Variance) for each genre.

It can be inferred that most people prefer to watch movies of the genre Drama and Comedy-Drama-Romance.

The highest count of genres is drama.

Thus, it can be concluded that nowadays, most viewers watch movies for entertainment to relieve stress in their busy lives.

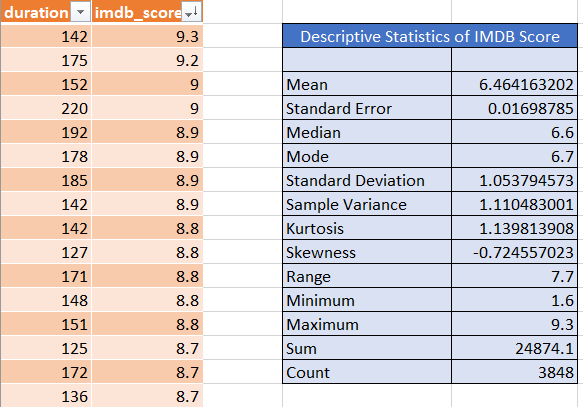
Our choice of genre greatly impacts our emotions as most people who choose drama are emotionally moved or sad; those choosing comedy want to experience joy.

So, people mostly prefer drama and comedy over horror and thriller, which shows fear or any other genre.

1. **Problem-2**

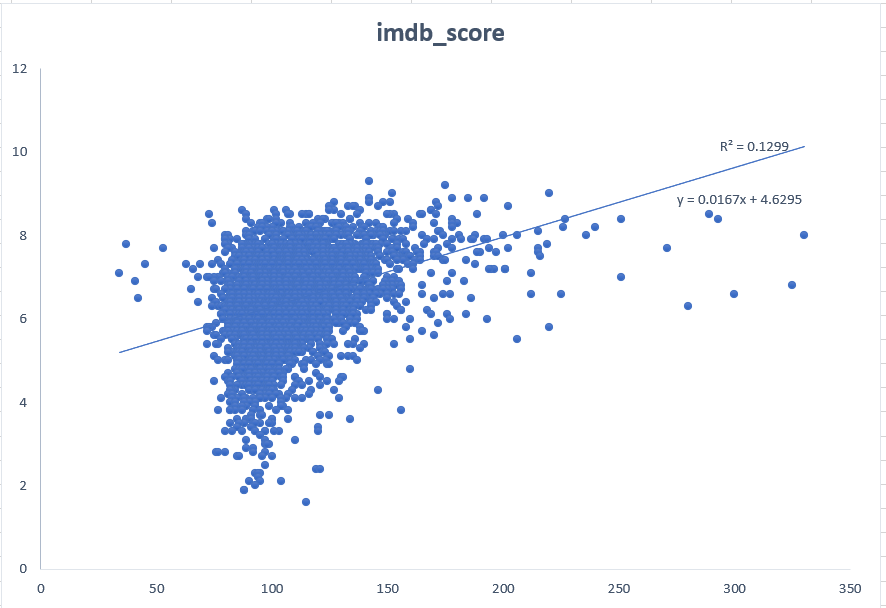
To identify the relationship between the movie duration and IMDB Score,

* We have created an another excel sheet with data containing the columns of only duration and imdb\_score.
* Convert this data into table and sort in accord to decreasing order of imdb\_score.

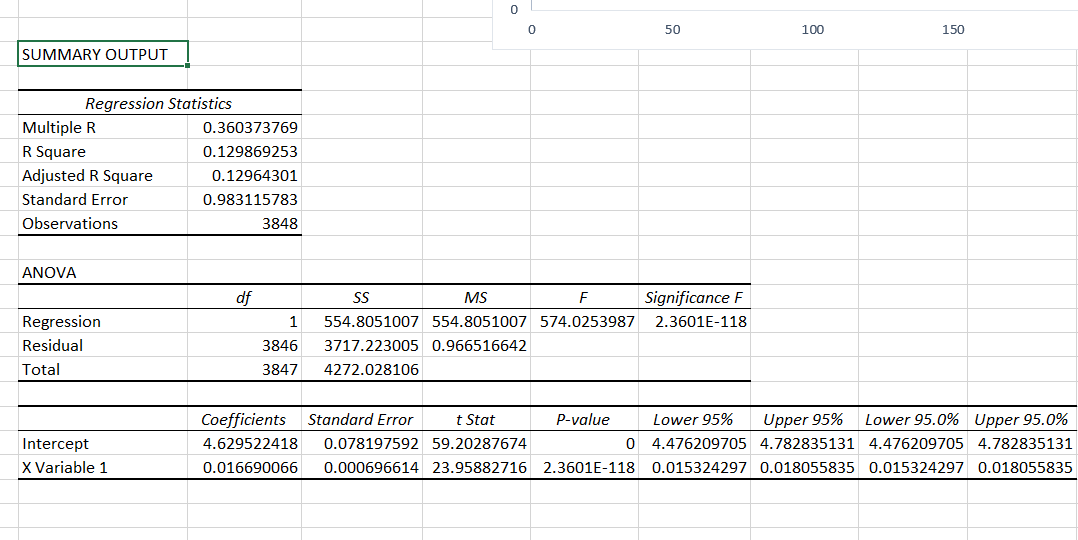


The IMDB score is highest for movie duration 142.

To visualize the relationship between the duration and imdb\_score we created a scattered plot.



The scattered plot shows a trend line with R-square=0.1299 which is too small and indicates a correlation between them but the correlation is weak.



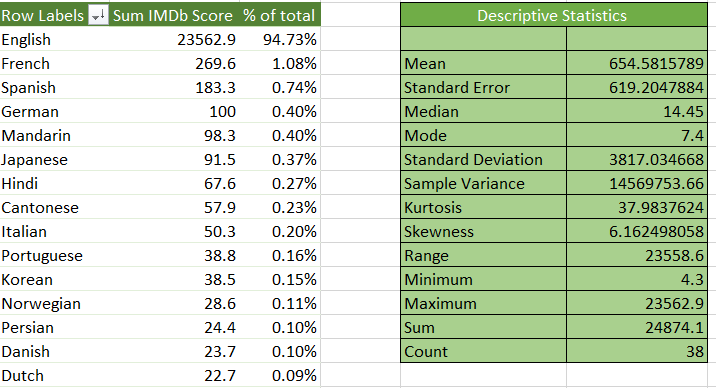
Also, if we analyze regression Statistics from the Data tab in excel,

The higher t-stat value and smaller p value indicates that there is relationship between the two variables.

1. **Problem-3:**

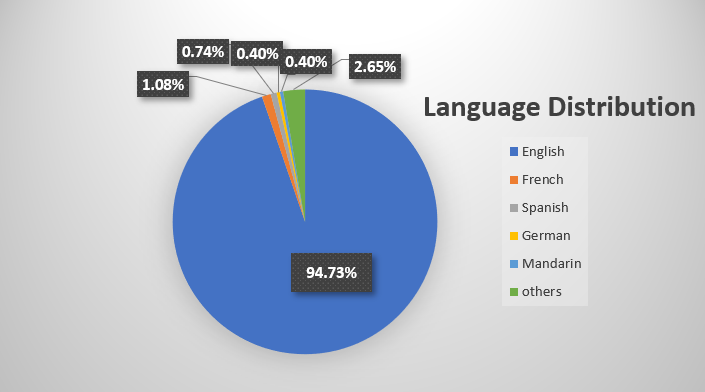
To determine the most common languages used in movies and analyze their impact on the IMDB score -

* Firstly, we select row as language and value as sum of imdb\_score
* Sort the table in decreasing order of sum of imdb\_score.
* For better clarity we can also find the % of sum of imdb\_score for each language.



From above, it is clear that English language movies comprise around 95% of total language distribution.

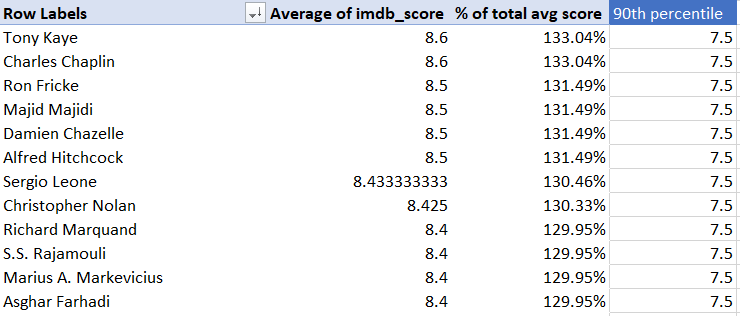
Thus, it can be concluded that movies in the English language have the highest imdb\_score, and people prefer to watch movies in English over other languages.

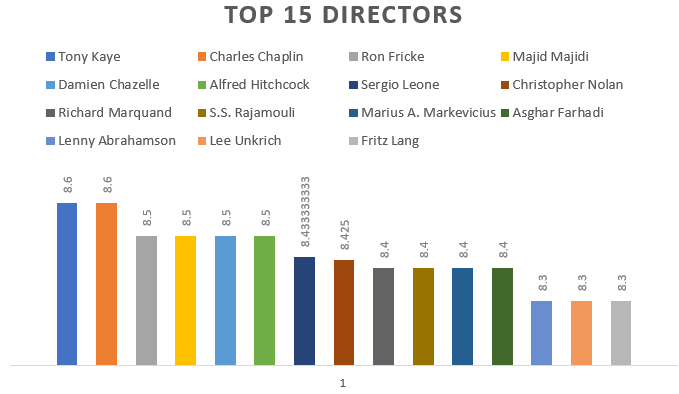


1. **Problem-4**

To find out whether there is any influence of directors on imdb\_score of a movie-

* We select director\_name in row label and Average of imdb\_score in the values label.
* Sort the data in decreasing order of average of imdb\_score.
* For better visualization, we will take records of top 15 directors and plot the data into a column diagram.
* Also, we have calculated 90the Percentile to compare the score of each director.





It is evident from above column chart that the director Tony Kaye has the highest average imdb\_score of 8.60.

1. **Problem-5:**

To analyze the correlation between movie budgets and gross earnings and identify the movies with the highest profit margin,

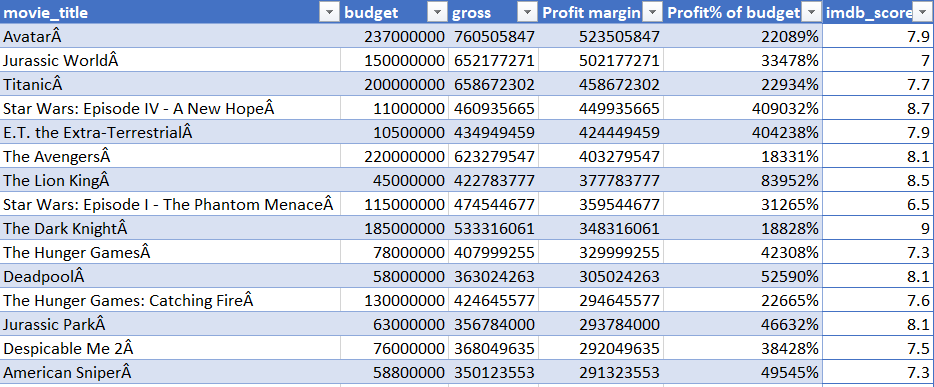
* Firstly, we calculate the profit margin for each movie and add a column named profit in our cleaned Excel data file.

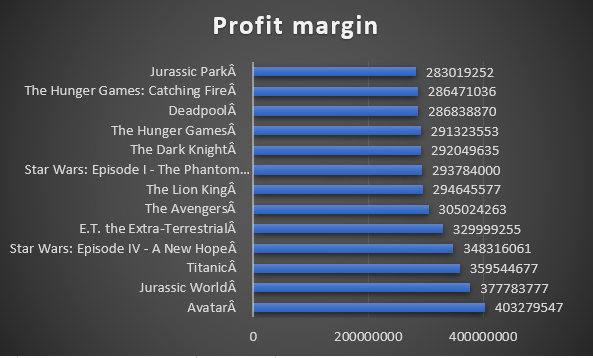
Where *Profit margin = Gross earnings-budget*

* Then, we calculate the correlation coefficient of the movie budget and gross earnings by using the Excel CORREL function as

=CORREL(Table3[gross], Table3[budget])

* Next, we take columns movie\_title, budget, gross and profit on another sheet
* Sort the data in decreasing order of profit.





It shows that

* Avatar movie has the highest profit margin.
* The highest-budget movie was ‘The Host’, but it made a massive loss.
* And Tarnation, which was the most low-budget movie but made a profit of 271466% of its budget.

Also, we calculated the correlation coefficient of budget and gross earnings; its value was 0.10085, representing a very weak relationship between the two variables, budget and gross earnings.

This means one gross earnings increase with the increase in the movie's budget, but not in a consistent or predictable way.

**Conclusion**

Based on our analysis, it can be concluded that its not all about the movie's budget, but a good direction and genre of public interest can have a good IMDB score.

Also, a good IMDB score does not guarantee a good profit margin.

But as many stakeholders, investors keep an eye on IMDB rating before investing, so this analysis plays an important role in the pre-production and post-production of the movies.

1. (Color, director\_facebook\_likes, actor\_3\_facebook\_likes,
2. actor\_2\_name, actor\_1\_facebook\_likes, cast\_total\_facebook\_likes,
3. actor\_3\_name, facenumber\_in\_posts, plot\_keywords,
4. movie\_imdb\_link, content\_rating, actor\_2\_facebook\_likes,
5. aspect\_ratio, movie\_facebook\_likes
6. Color, director\_facebook\_likes, actor\_3\_facebook\_likes,
7. actor\_2\_name, actor\_1\_facebook\_likes, cast\_total\_facebook\_likes,
8. actor\_3\_name, facenumber\_in\_posts, plot\_keywords,
9. movie\_imdb\_link, content\_rating, actor\_2\_facebook\_likes,
10. aspect\_ratio, mov