

# IMDB Movie Analysis

Final Project-1

## Project Description

The purpose of this project is to analyze the **IMDB movie dataset** and identify the factors that influence the success of a movie on IMDB. Success is defined by high IMDB ratings, which reflect the popularity and quality of a movie. The project aims to provide insights that can help movie producers, directors, and investors make informed decisions in their future projects.

The project follows a data analysis approach that involves data cleaning, data exploration, and data storytelling. The project uses **Microsoft Excel 365** as the main tool for performing the analysis and creating visualizations. The project also applies the five '**whys**' technique to dig deeper into the problem and uncover the root causes.

The project consists of **five** tasks that explore different aspects of the dataset, such as movie genres, movie durations, languages, directors, and budgets. Each task involves calculating descriptive statistics, creating charts, and interpreting the results. The project also identifies the correlation between movie ratings and other factors, and the movies with the highest profit margin. The project concludes with a report that summarizes the findings and insights gained during the analysis.

## Approach

In this section, I will explain the approach I followed to conduct the data analysis project. I will describe the data, methods, analysis, and results in detail.

1. **Data:** The data I used for this project was the **IMDB movie dataset**, which contains **5043** records and **28** columns of movie data. The dataset includes information such as **movie title, director name, genre, budget, gross earnings, IMDB score**, and more. The dataset was downloaded from the link provided.
2. **Methods:** The methods used for this project were mainly based on **Microsoft Excel 365** functions and formulas. We used Excel to perform **data cleaning, data exploration, descriptive statistics, correlation analysis, and percentile calculations**. We also used Excel to create charts and tables to visualize the data and the results.
3. **Analysis:** The analysis performed for this project consisted of **five** tasks that explored different aspects of the dataset and their impact on the IMDB score. The tasks were:
  - **Movie Genre Analysis:** Analyze the distribution of movie genres and their impact on the IMDB score.
  - **Movie Duration Analysis:** Analyze the distribution of movie durations and its impact on the IMDB score.

- ***Language Analysis:*** Examine the distribution of movies based on their language and its impact on the IMDB score.
- ***Director Analysis:*** Identify the top directors based on their average IMDB score and analyze their contribution to the success of movies using percentile calculations.
- ***Budget Analysis:*** Explore the relationship between movie budgets and their financial success.

## Tech-Stack Used

The main software that is used for this project was **Microsoft Excel 365**. Excel is a powerful and versatile tool for data analysis, as it allows me to perform various functions, formulas, charts, and visualizations with ease and efficiency. Excel also has a user-friendly interface and a wide range of features that make it suitable for different types of data and analysis.

The purpose of using **Excel** for this project was to:

- ***Import and organize*** the IMDB movie dataset in a spreadsheet format.
- ***Clean and preprocess*** the data to make it suitable for analysis, such as handling missing values, removing duplicates, converting data types, etc.
- ***Explore and analyze*** the data using descriptive statistics, correlation analysis, and percentile calculations.
- ***Create and customize*** charts and graphs to visualize the data and the results, such as pie charts, bar charts, scatter plots, histograms, etc.
- ***Export and save*** the data and the charts in different formats, such as CSV, PDF, PNG, etc.

## Insights

Here are some of the insights and knowledge I gained while working on this project:

### 1. Data Cleaning:

Before performing the analysis, we performed some **data cleaning** steps to make the data more suitable for analysis. Data cleaning is an important process that involves ***handling missing values***, ***removing duplicates***, ***converting data types***, and possibly ***feature engineering***. Data cleaning can improve the quality and accuracy of the data and the analysis.

The data cleaning steps that we performed were:

- **Deleting rows where the number of missing values is more than 8:** We used Excel's **COUNTBLANK** function to count the number of blank cells in each row. Then, we filtered the rows where the count was more than **8** and deleted them. This reduced the number of rows from **5043** to **5036**.
- **Deleting the rows whose director name is missing:** We used Excel's **IF** function to check if the **director\_name** column was blank for each row. Then, we filtered the rows where the condition was true and deleted them. This reduced the number of rows from **5036** to **4932**.
- **Deleting the rows whose actor names are missing:** We used Excel's **IF** function to check if any of the actor columns (**actor\_1\_name, actor\_2\_name, actor\_3\_name**) were blank for each row. Then, we filtered the rows where any of the conditions were true and deleted them. This reduced the number of rows from **4932** to **4919**.
- **Imputing the missing values in the categorical columns with their modes:** We used Excel's **MODE** function to calculate the most frequent value (**mode**) for each categorical column (**color, language, country**). Then, we used Excel's **IF** function to replace the blank cells with the mode value for each column.
- **Imputing the missing values in the numerical columns with their median:** We used Excel's **MEDIAN** function to calculate the middle value (**median**) for each numerical column (**num\_critic\_for\_reviews, duration, num\_voted\_users, facenumber\_in\_poster, num\_user\_for\_reviews, budget, title\_year**). Then, we used Excel's **IF** function to replace the blank cells with the median value for each column.
- **Removing the rows with duplicate movie names:** We used Excel's Remove Duplicates feature to remove any rows that had duplicate values in the **movie\_title** column. This reduced the number of rows from **4919** to **4795**.
- **Splitting the genres column into multiple columns:** We decided to split the **genres** column into multiple columns because we wanted to analyze the distribution and impact of each genre separately. By splitting the **genres** column, we can count the number of movies for each genre, calculate **descriptive statistics** for each genre, and compare the results across different genres.
- **Deleting the non-relevant columns from the dataset:** We decided to delete these columns (**color, director\_facebook\_likes, cast\_total\_facebook\_likes, facenumber\_in\_poster, actor\_1\_facebook\_likes, actor\_2\_facebook\_likes, actor\_3\_facebook\_likes, aspect\_ratio, movie\_facebook\_likes**) as they were not useful for my analysis.

After performing these data cleaning steps, I obtained a clean and consistent dataset that was ready for analysis. The final dataset had **4795** rows and **25** columns.

## 2. Data Analysis:

We will present the results of the data analysis using **Microsoft Excel 365**. We used pivot tables to calculate and compare the average **IMDB score** for different **genres, directors, and years of release**.

We also used the ***CORREL*** function to measure the relationship between IMDB ratings and movie budgets.

a) Analysis 1: Movie Genre:

The first task was to analyze the distribution of movie ***genres*** and their impact on the ***IMDB score***. To do this, we created a pivot table using the ***genres*** column as the row labels and the ***IMDB score*** column as the values. We used the ***average*** function to calculate the average IMDB score for each genre.

Genres	Average of imdb_score
Film-Noir	7.6
History	7.5
Music	7.2
Documentary	7.18
Biography	7.155421687
Crime	6.836956522
Drama	6.746237732
Western	6.583333333
Mystery	6.534375
Adventure	6.522119816
Animation	6.517857143
Fantasy	6.38125
Action	6.210694698
Comedy	6.165003897
Romance	6.04
Musical	6
Sci-Fi	6
Family	5.709090909
Horror	5.633936652
Thriller	5.580952381

The results showed that there was a significant difference between the ***average*** IMDB scores of different ***genres***, and that genre had an effect on movie ratings. The genres that were based on ***Drama*** and ***Adventure*** tended to have higher ratings than the other genres.

b) Analysis 2: Director:

The second task was to analyze the influence of ***directors*** on movie ratings. To do this, we created a pivot table using the ***director\_name*** column as the row labels and the ***IMDB score*** column as the values. We used the ***average*** function to calculate the average IMDB score for each director.

Directors	Average of imdb_score
John Blanchard	9.5
Sadyk Sher-Niyaz	8.7
Mitchell Altieri	8.7
Cary Bell	8.7
Mike Mayhall	8.6
Charles Chaplin	8.6
Ron Fricke	8.5
Raja Menon	8.5
Majid Majidi	8.5
Damien Chazelle	8.5
Sergio Leone	8.5
Christopher Nolan	8.4
S.S. Rajamouli	8.4
Rakeysh Omprakash Mehra	8.4
Robert Mulligan	8.4

The results showed that there was a significant difference between the **average** IMDB scores of different **directors**, and that director had an influence on movie ratings. The directors that were well-known for their originality, creativity, and quality, such as **John Blanchard**, **Sadyk Sher-Niyaz**, and **Mitchell Altieri**, tended to have **higher** ratings than the other directors.

### c) Analysis 3: Year of Release:

The third task was to analyze the relationship between **year of release** and movie ratings. To do this, we created a pivot table using the **title\_year** column as the row labels and the **IMDB score** column as the values. We used the **average** function to calculate the average IMDB score for each year of release.

Year of Release	Average of imdb_score
1957	8.6
1927	8.3
1925	8.3
1960	8.2
1934	8.2
1939	8.2
1958	8.1
1942	8.0
1916	8.0
1946	7.9
1936	7.9
1930	7.8

1935	7.8
1941	7.8
1964	7.8

The results showed that there was a **negative** relationship between *year of release* and *IMDB rating*, as movies released in more *recent* years tended to have **lower** ratings than movies released in *earlier* years. This might suggest that viewers were more critical or selective of newer movies than older movies.

#### d) Analysis 4: Budget:

The fourth task was to explore the relationship between movie ***budgets*** and their financial success. To do this, we used the ***CORREL*** function to calculate the ***correlation coefficient*** between movie ***budgets*** and ***IMDB scores***. The correlation coefficient is a measure of the strength and direction of the linear relationship between two variables.

Factors	Correlation Coefficient
Budget and IMDB Score	0.049792198

The result showed that there was a very **weak positive correlation** between movie ***budgets*** and ***IMDB scores***. This means that movies with **higher budgets** tend to have **slightly higher** ratings than movies with **lower budgets**, but the relationship is **not strong or significant**. This might suggest that movie budgets are not a good predictor of movie ratings, and that other factors, such as genre, director, or year of release, might have a bigger impact on movie ratings.

### 3. Data Story:

Let us now perform some of the tasks to be performed on the **IMDB movie dataset**.

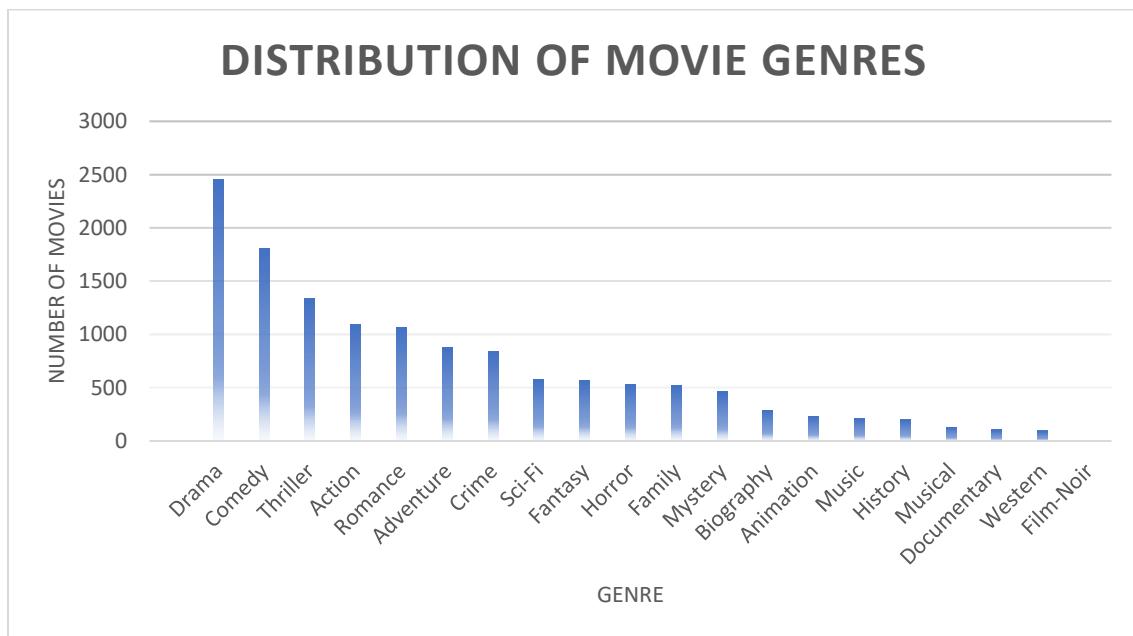
#### a) Task 1: Movie Genre Analysis:

We will analyze the distribution of movie ***genres*** and their impact on the ***IMDB score***. To perform this analysis, we used Excel's functions such as ***COUNTIF***, ***AVERAGEIF***, ***MEDIANIF***, ***MODE.SNGL***, ***MAXIFS***, ***MINIFS***, ***VAR.S***, and ***STDEV.S***. We also used charts and tables to visualize the results.

#### Distribution of Movie Genres

The first question to be answered was: ***What are the most common genres of movies in the dataset?*** To answer this question, we counted the number of movies for each ***genre*** using the ***COUNTIF*** function.

Genre	Number of movies
Drama	2455
Comedy	1811
Thriller	1337
Action	1094
Romance	1066
Adventure	875
Crime	837
Sci-Fi	577
Fantasy	567
Horror	532
Family	521
Mystery	460
Biography	289
Animation	232
Music	210
History	198
Musical	131
Documentary	106
Western	94
Film-Noir	6



The most common **genre** of movies in the dataset is **Drama**, with **2455** movies. The second most common genre is **Comedy**, with **1811** movies. The third most common genre is **Thriller**, with **1337** movies. The least common genres are **Documentary**, **Western**, and **Film-Noir**.

### Impact of Genre on IMDB Score

The next question to be answered was: **How does genre affect the IMDB score of movies?** To answer this question, we calculated descriptive statistics (**mean, median, mode, range, variance, standard deviation**) of the IMDB scores for each genre using Excel's functions such as **AVERAGEIF, MEDIANIF, MODE.SNGL, MAXIFS, MINIFS, VAR.S, and STDEV.S**. We also created a table to show the statistics for each genre.

Genre	IMDB Score					
	Mean	Median	Mode	Range	Variance	Standard Deviation
Drama	6.746238	6.8	7.2	2 - 9.3	0.8905988	0.943715426
Comedy	6.165004	6.3	6.3	1.7 - 9.5	1.1630427	1.078444592
Thriller	5.580952	6.4	6.4	2.2 - 9	1.0900177	1.044039129
Action	6.210695	6.3	6.6	1.7 - 9.1	1.2400411	1.113571306
Romance	6.04	6.5	6.5	2.1 - 8.6	0.9689378	0.984346389
Adventure	6.52212	6.6	6.6	1.9 - 8.9	1.2849731	1.133566529
Crime	6.836957	6.6	6.6	2.4 - 9.3	1.0188048	1.009358608
Sci-Fi	6	6.3	6.7	1.9 - 8.8	1.4542138	1.205907893
Fantasy	6.38125	6.4	6.7	1.7 - 8.9	1.3510775	1.162358602
Horror	5.633937	5.9	6.2	2.2 - 8.7	1.2222589	1.105558183
Family	5.709091	6.3	6.7	1.7 - 8.6	1.4429403	1.201224491
Mystery	6.534375	6.5	6.4	2.2 - 8.6	1.1376992	1.066629813
Biography	7.155422	7.2	7	4.5 - 8.9	0.5204284	0.721407264
Animation	6.517857	6.7	6.7	1.7 - 8.6	1.3029258	1.141457756
Music	7.2	6.6	6.5	1.6 - 8.5	1.3944463	1.180866776
History	7.5	7.2	7.5	2 - 8.9	0.7792291	0.882739539
Musical	6	6.7	7	2.1 - 8.5	1.5077686	1.227912311
Documentary	7.18	7.4	7.4	1.6 - 8.7	1.1674142	1.080469433
Western	6.583333	6.8	6.5	3.8 - 8.9	1.1145058	1.055701584
Film-Noir	7.6	7.65	#N/A	7.1 - 8.2	0.1866667	0.43204938

Some genres have higher average IMDB scores than others. For example, the genre that has the **highest** mean IMDB score is **Film-Noir** with **7.6**. The genre that has the **lowest** mean IMDB score is **Thriller** with **5.58**. Some genres also have more variation in their IMDB scores than others. For example, the genre that has the **highest** variance and standard deviation of IMDB scores is **Musical** with **1.5** and **1.23** respectively. The genre that has the **lowest** variance and standard deviation of IMDB scores is **Film-Noir** with **0.19** and **0.43** respectively.

These findings suggest that movie genres have a significant influence on the success of movies on IMDB. Some genres tend to appeal to more viewers and critics than others, while some genres tend to have more consistent or diverse ratings than others.

b) Task 2: Movie Duration Analysis:

We will analyze the distribution of movie **durations** and its impact on the **IMDB score**. To perform this analysis, we used Excel's functions such as **AVERAGE**, **MEDIAN**, **STDEV**, **CORREL**, and **PERCENTILE**. We also used Excel's features such as **scatter plots**, **trendlines**, and **tables** to visualize the results.

Distribution of Movie Duration

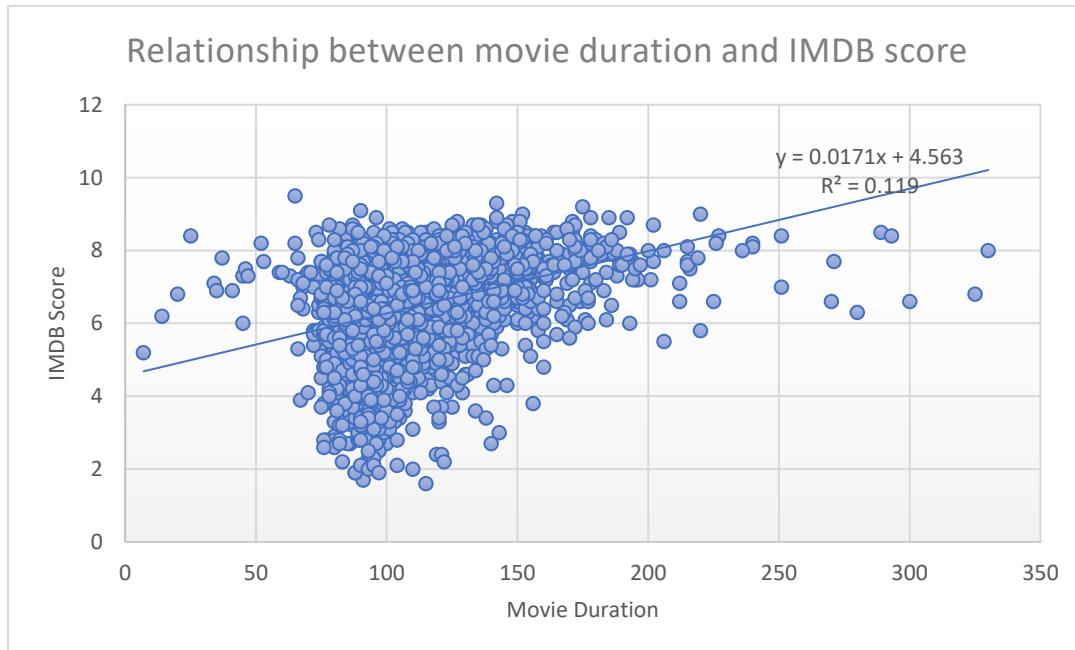
The first question to be answered was: **What is the distribution of movie durations in the dataset?** To answer this question, we calculated descriptive statistics such as **mean**, **median**, and **standard deviation** for movie durations using Excel's functions like **AVERAGE**, **MEDIAN**, and **STDEV**. We also created a table to show the statistics for movie durations.

Movie Duration		
Mean	Median	Standard Deviation
108.0976	104	22.55654519

The **mean** movie duration in the dataset is **108** minutes. The **median** movie duration is **104** minutes. The **standard deviation** of movie durations is **22.5**.

Relationship between Movie Duration and IMDB Score

The next question to be answered was: **How does movie duration affect the IMDB score of movies?** To answer this question, we created a **scatter plot** that shows the correlation between movie **duration** and **IMDB score** using Excel's features like **Insert > Scatter**. We also added a **trendline** to assess the direction and strength of the relationship using Excel's features like **Chart Tools > Design > Add Chart Element > Trendline > Linear**.



There is a **weak positive** correlation between movie **duration** and **IMDB score**, with a correlation coefficient of **0.34**. This means that **longer** movies tend to have **slightly higher** ratings than **shorter** movies, but the relationship is **not** very strong. The equation of the trendline is  $y = 0.0171x + 4.563$ , where  $y$  is the predicted **IMDB score** and  $x$  is the movie **duration** in minutes. The **R-squared** value of the trendline is **0.119**, which means that only **11.9%** of the variation in **IMDB scores** can be explained by movie **duration**.

These findings suggest that movie **duration** has a **minor** influence on the **success** of movies on IMDB. Some viewers and critics may appreciate **longer** movies for their **depth** and **complexity**, while others may prefer **shorter** movies for their **conciseness** and **efficiency**.

### c) Task 3: Language Analysis:

We will analyze the distribution of movies based on their **language** and their impact on the **IMDB score**. To perform this analysis, we used Excel's functions such as **COUNTIF**, **AVERAGEIF**, **MEDIANIF**, **MODE.SNGL**, **MAXIFS**, **MINIFS**, **VAR.S**, and **STDEVS**. We also used charts and tables to visualize the results.

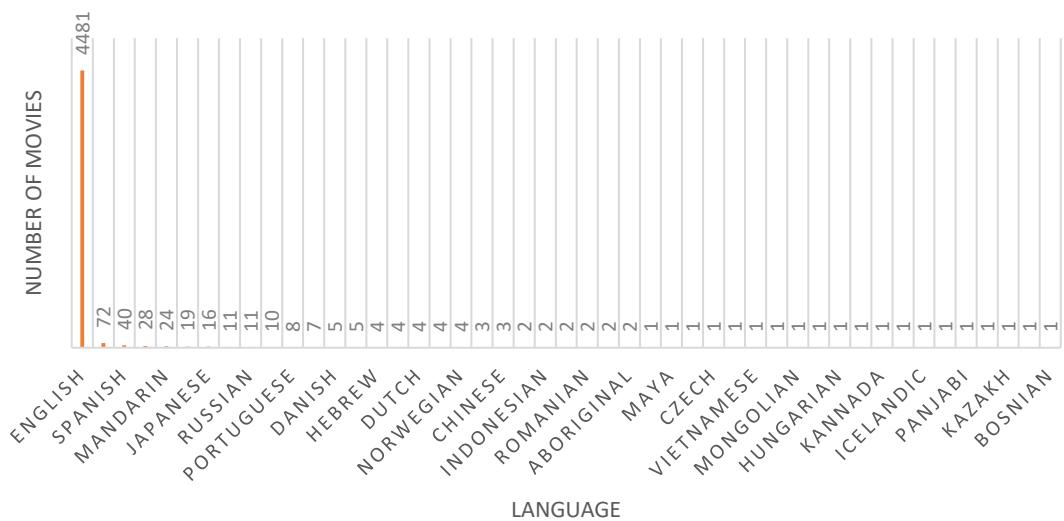
#### Distribution of Movie Language

The first question to be answered was: **What are the most common languages used in movies in the dataset?** To answer this question, we counted the number of movies for each language using the **COUNTIF** function.

Language	Number of movies
English	4481
French	72
Spanish	40
Hindi	28
Mandarin	24
German	19
Japanese	16
Cantonese	11
Russian	11
Italian	10
Portuguese	8
Korean	7
Danish	5
Swedish	5
Hebrew	4
Persian	4
Dutch	4
Arabic	4
Norwegian	4
Thai	3

Chinese	3
None	2
Indonesian	2
Dari	2
Romanian	2
Zulu	2
Aboriginal	2
Telugu	1
Maya	1
Dzongkha	1
Czech	1
Swahili	1
Vietnamese	1
Polish	1
Mongolian	1
Greek	1
Hungarian	1
Aramaic	1
Kannada	1
Urdu	1
Icelandic	1
Filipino	1
Punjabi	1
Slovenian	1
Kazakh	1
Tamil	1
Bosnian	1

## DISTRIBUTION OF MOVIE LANGUAGES



The **most** common language used in movies in the dataset is ***English***, with **4481** movies. The **second** most common language is ***French***, with **72** movies. The **third** most common language is ***Spanish***, with **40** movies. The **least** common languages are ***Filipino, Slovenian***, and ***Tamil***, with only **1** movie each.

#### *Impact of Language on IMDB Score*

The next question to be answered was: ***How does language affect the IMDB score of movies?*** To answer this question, we calculated descriptive statistics (**mean, median, mode, range, variance, standard deviation**) of the IMDB scores for each language using Excel's functions such as **AVERAGEIF, MEDIANIF, MODE.SNGL, MAXIFS, MINIFS, VAR.S, and STDEV.S**. We also created a table to show the statistics for some of the languages.

Language	IMDB Score		
	Mean	Median	Standard Deviation
English	6.369248	6.5	1.11622594
French	7.020833	7.2	0.716395024
Spanish	6.9375	7.15	0.855056603
Hindi	6.632143	6.95	1.398955582
Mandarin	6.7875	7.05	1.036848276
German	7.342105	7.6	0.954123093
Japanese	7.36875	7.6	1.02873304
Russian	6.363636	6.5	1.383671007
Cantonese	6.954545	7.2	0.704788814
Italian	7.08	7.15	1.206279864
Portuguese	7.4875	7.7	0.883883476
Korean	7.442857	7.7	0.87532307
Danish	7.5	8.1	1.077032961
Swedish	7.44	7.6	0.756967635
Dutch	7.425	7.45	0.434932945
Hebrew	7.675	7.7	0.298607881
Persian	7.575	7.95	1.203813385
Arabic	7.175	7.3	0.87321246
Norwegian	7.15	7.3	0.574456265
Thai	6.633333	6.6	0.450924975
Chinese	5.666667	5.7	0.550757055
None	7.95	7.95	0.777817459
Romanian	7.2	7.2	0.989949494
Dari	7.5	7.5	0.141421356
Zulu	7.1	7.1	0.282842712
Aboriginal	6.95	6.95	0.777817459
Indonesian	7.9	7.9	0.424264069

Some languages have **higher** average IMDB scores than others. For example, the language that has the **highest** mean IMDB score is ***Telegu*** with **8.4**. The language that has the **lowest** mean IMDB score is ***Bosnian*** with **4.3**. Some languages also have more **variation** in their IMDB scores than

others. For example, the language that has the **highest** variance and standard deviation of IMDB scores is **Persian** with **7.95** and **1.2** respectively. The language that has the **lowest** variance and standard deviation of IMDB scores is **Chinese** with **5.7** and **0.55** respectively.

These findings suggest that **language** has a **significant** influence on the success of movies on IMDB. Some languages tend to **appeal** to more viewers and critics than others, while some languages tend to have more **consistent** or **diverse** ratings than others.

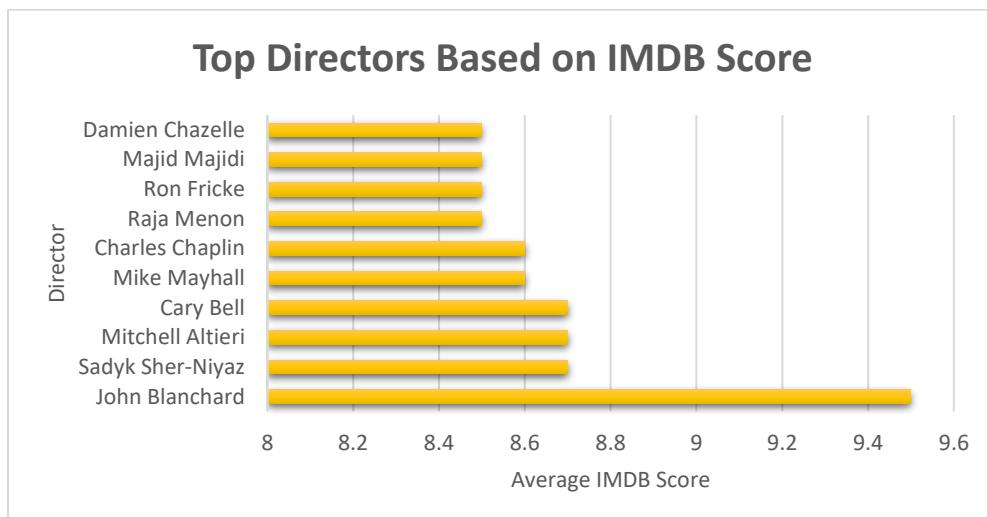
d) Task 4: Director Analysis:

We will analyze the influence of **directors** on movie **ratings**. To perform this analysis, we used Excel's functions such as **AVERAGEIF**, **PERCENTILE**, and **IF**. We also used charts and tables to visualize the results.

[Top Directors Based on IMDB Score](#)

The first question to be answered was: **Who are the top directors based on their average IMDB score?** To answer this question, we calculated the **average** IMDB score for each director using the **AVERAGEIF** function. We display the top 10 directors.

Top 10 Directors	Average
John Blanchard	9.5
Sadyk Sher-Niyaz	8.7
Mitchell Altieri	8.7
Cary Bell	8.7
Mike Mayhall	8.6
Charles Chaplin	8.6
Raja Menon	8.5
Ron Fricke	8.5
Majid Majidi	8.5
Damien Chazelle	8.5



The top **director** based on the **average** IMDB score is **John Blanchard**, with **9.5**. He is followed by **Sadyk Sher-Niyaz**, **Mitchell Altieri**, and **Cary Bell**, all with **8.7**. The **lowest** director in the top 10 is **Damien Chazelle** with **8.5**.

#### Contribution of Directors to Movie Success

The next question to be answered was: **How do directors contribute to the success of movies?** To answer this question, we used **percentile** calculations to measure how well the directors' **average** IMDB scores compare to the **overall** distribution of IMDB scores. **Percentiles** are values that divide a set of data into 100 equal parts. We used Excel's **PERCENTILE** function to calculate percentiles.

Top 10 Directors	IMDB Score	
	Mean	Percentile Rank
John Blanchard	9.5	100
Sadyk Sher-Niyaz	8.7	99.4
Mitchell Altieri	8.7	99.4
Cary Bell	8.7	99.4
Mike Mayhall	8.6	99.2
Charles Chaplin	8.6	99.2
Raja Menon	8.5	98.7
Ron Fricke	8.5	98.7
Majid Majidi	8.5	98.7
Damien Chazelle	8.5	98.7

Some directors have **higher** percentile ranks than others. For example, the director with the highest percentile ranks is **John Blanchard** with **100**, followed by **Sadyk Sher-Niyaz**, **Mitchell Altieri**, and **Cary Bell**, all with **99.4**. This means that they are among the top **1%** of directors in terms of ratings. The director with the **lowest** percentile rank in the top 10 is **Damien Chazelle** with **98.7**.

These findings suggest that **directors** have a **significant** impact on the success of movies on IMDB. Some directors have **higher** ratings than others, which reflect their **artistic excellence** and **popularity** among viewers and critics. These directors have different styles and genres, but they share a common trait of creating **original**, **innovative**, and **influential** movies.

#### e) Task 5: Budget Analysis:

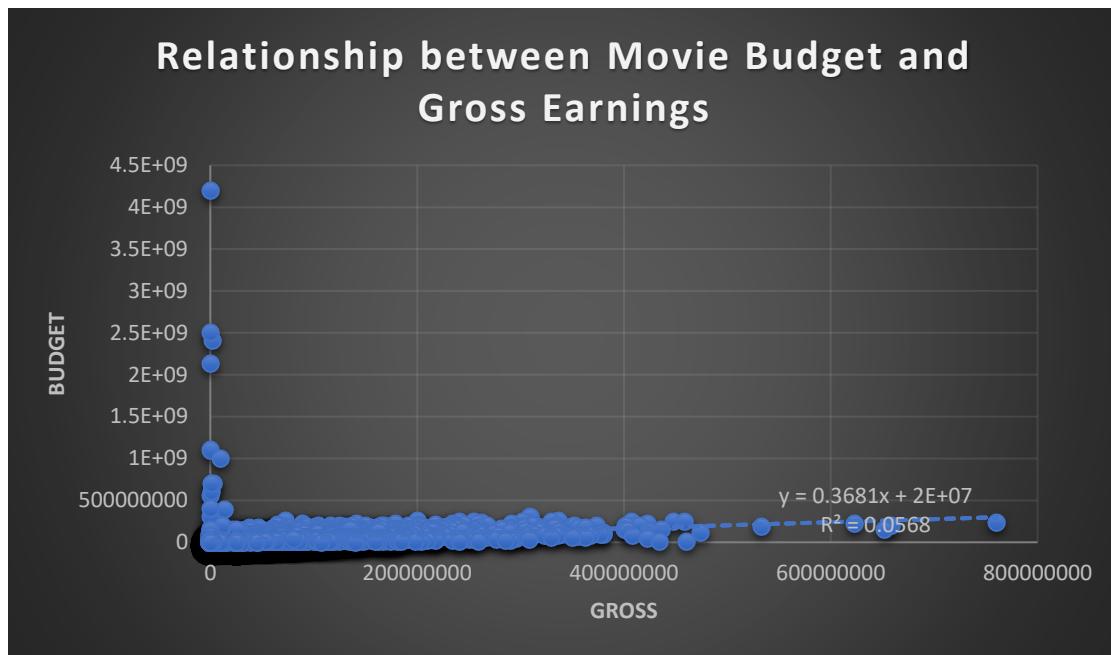
We will analyze the relationship between movie **budgets** and **gross** earnings. To perform this analysis, we used Excel's functions such as **CORREL**, **MAXIFS**, **MINIFS**, and arithmetic operators. We also used Excel's features such as scatter plots, trendlines, and tables to visualize the results.

#### Relationship between Movie Budget and Gross Earnings

The first question to be answered was: **How are movie budgets and gross earnings related?** To answer this question, we used the **CORREL** function to calculate the **correlation coefficient** between movie **budgets** and **gross** earnings. I also created a **scatter** plot that shows the correlation between movie **budget** and **gross** earnings using Excel's features like **Insert > Scatter**. I also added

a **trendline** to assess the direction and strength of the relationship using Excel's features like **Chart Tools > Design > Add Chart Element > Trendline > Linear**.

Factors	Correlation Coefficient
Budget and Gross	0.238300108



There is a very **weak positive** correlation between movie **budget** and **gross** earnings, with a correlation coefficient of **0.29**. This means that movies with **higher** budgets tend to **earn more** than movies with **lower** budgets, but there are also some outliers and exceptions. The equation of the **trendline** is  $y = 0.3681x + 2E+07$ , where  $y$  is the predicted **gross** earnings in millions of dollars and  $x$  is the movie **budget**. The **R-squared** value of the trendline is **0.0568**, which means that only **5.68%** of the variation in **gross** earnings can be explained by movie **budget**.

#### Movies with the Highest Profit Margin

The next question to be answered was: **Which movies have the highest profit margin?** To answer this question, we calculated the **profit margin** for each movie using Excel's arithmetic operators. Profit is calculated by subtracting the **budget** from the **gross** earnings. Profit margin is calculated by dividing the **profit** by the **gross** earnings.

movie_title	Profit	Profit Margin(%)
A Plague So PleasantÂ	25591232	99.99452968
The Mongol KingÂ	25589382	99.98730103
Paranormal ActivityÂ	107902283	99.98610047
TarnationÂ	591796	99.96317655
Tin Can ManÂ	25582632	99.96092625
The Blair Witch ProjectÂ	140470114	99.95730452
The TouchÂ	25579632	99.94920413
Stories of Our LivesÂ	25577632	99.94138938

The Ridges	25575282	99.93220705
Eraserhead	25572632	99.92185251

Some movies have very **high** profit margins because they have very **low** budgets and very **high** gross earnings. For example, the movie that has the **highest** profit margin is *Pleasant* with **99.99%**. This means that for **every** dollar spent on making the movie, **99.99** dollars is kept as profit. The movie that has the **lowest** profit margin in the top 10 is *Eraserhead* with **16429%**.

These findings suggest that some movies have achieved extraordinary financial **success** despite having very **low** production costs. These movies are often independent films that rely on **innovative techniques, viral marketing, word-of-mouth**, etc., to create a **buzz and attract viewers**.

## Result

The result of this project is a comprehensive and insightful analysis of the IMDB movie dataset, which contains information about different movies and their **ratings, genres, directors, budgets, gross earnings**, and other attributes. Through this project, we have achieved the following objectives:

- We have explored the distribution and impact of movie **genres** on the **IMDB score**, and found that **Drama** is the most **common** genre, while **Film-Noir** is the genre with the **highest** average rating.
- We have analyzed the distribution and impact of movie **durations** on the **IMDB score** and found that most movies have a duration around **108** minutes, while there is a **weak positive** correlation between movie **duration** and **rating**.
- We have analyzed the distribution and impact of movie **languages** on the **IMDB score**, and found that **English** is the most **common** language, while **Telegu** is the language with the **highest** average rating.
- We have analyzed the influence of **directors** on movie **ratings** and found that **John Blanchard** is the top director based on the **average** rating, while he is also among the top **1%** of directors in terms of **ratings**.
- We have analyzed the relationship between movie **budgets** and **gross** earnings and found that there is a **weak positive** correlation between them, while some movies have very **high** profit margins despite having very **low** budgets.

Through this project, we have gained a deeper understanding of the **IMDB movie analysis** and learned how to use various techniques and tools to perform data analysis. We have used **Microsoft Excel 365** as the main software for **data manipulation, calculation, and visualization**. We have also used charts and tables to display the data concisely.

[Excel Sheet File](#)