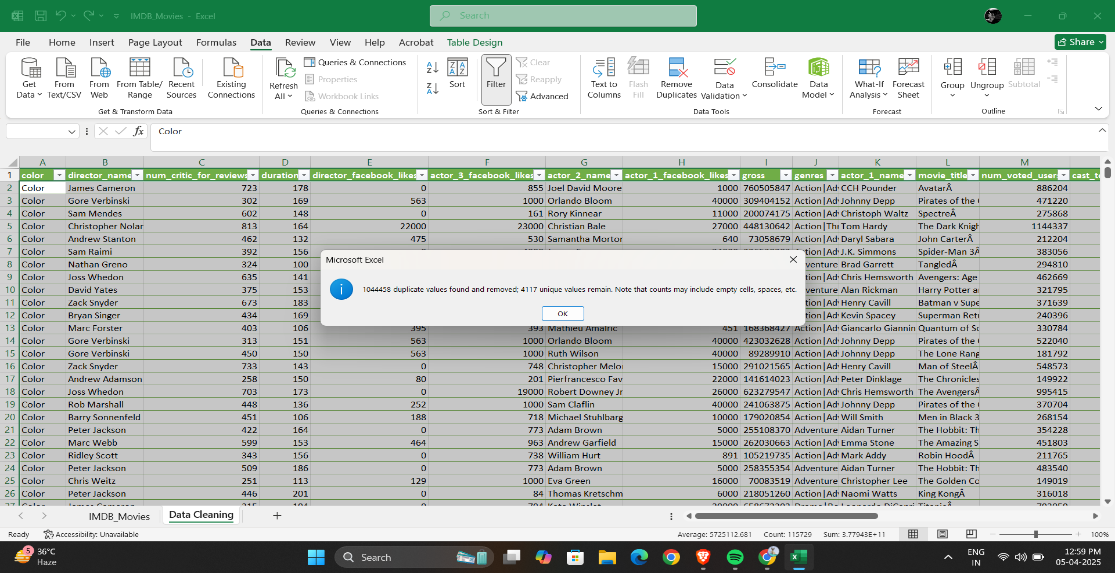
**5. IMDB MOVIES ANALYSIS**

**Data Cleaning:** In the IMDB Movies dataset, cleaning involved:

* Removing duplicates
* Handling missing values in columns like budget, gross and IMDB score.
* Converting data types, such as ensuring numerical columns like budget, gross and duration are in proper number format
* Splitting multi-value columns (e.g., genres like "Action,Adventure") for better analysis
* Removing or replacing symbols/invalid entries



1. **Movie Genre Analysis:**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.

**Output 1:**

|  |  |
| --- | --- |
| **Genres** | **Count of genres** |
| Action | 977 |
| Adventure | 385 |
| Animation | 50 |
| Biography | 223 |
| Comedy | 1126 |
| Crime | 280 |
| Documentary | 58 |
| Drama | 762 |
| Family | 3 |
| Fantasy | 40 |
| Horror | 163 |
| Music | 1 |
| Musical | 3 |
| Mystery | 24 |
| Romance | 2 |
| Sci-Fi | 8 |
| Thriller | 6 |
| Western | 4 |
| **Grand Total** | **4115** |

**Output 2:** Descriptive Statistics:

|  |  |
| --- | --- |
| Average | 228.6111 |
| Median | 45 |
| Mode | 3 |
| Max | 1126 |
| Min | 1 |
| Var | 127848 |
| Sd | 357.5584 |

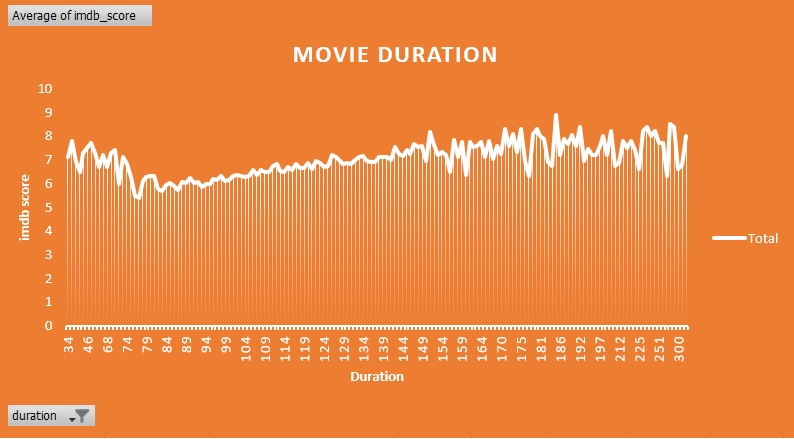
**Insights:** The analysis showed that genres like Drama, Action, and Comedy are most frequent, but genres like Documentary and Thriller often scored higher on IMDB.

1. **Movie Duration Analysis:** Analyze the distribution of movie durations and identify the relationship between movie duration and IMDB score.

**Output 1:**

****

**Output 2:** Visualize the relationship between movie duration and IMDB score.

****

**Insights:** Movies that are too short or too long tend to have lower ratings, whereas mid-length movies generally receive higher viewer satisfaction.

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

**Output 1:**

****

**Output 2:** Descriptive Statistics

|  |  |  |  |
| --- | --- | --- | --- |
| **Languages** | **Mean** | **Median** | **Stdev** |
| **Aboriginal** | 6.469121 | 6.6 | 1.04984 |
| **Arabic** | 6.469121 | 6.6 | 1.04984 |
| **Aramaic** | 6.468878 | 6.6 | 1.05099 |
| **Bosnian** | 6.468878 | 6.6 | 1.05099 |
| **Cantonese** | 6.471064 | 6.6 | 1.048969 |
| **Chinese** | 6.468878 | 6.6 | 1.05099 |
| **Czech** | 6.468878 | 6.6 | 1.05099 |
| **Danish** | 6.469363 | 6.6 | 1.048921 |
| **Dari** | 6.469121 | 6.6 | 1.04984 |
| **Dutch** | 6.469363 | 6.6 | 1.048921 |
| **Dzongkha** | 6.468878 | 6.6 | 1.05099 |
| **English** | 7.41397 | 6.6 | 60.5572 |
| **Filipino** | 6.468878 | 6.6 | 1.05099 |
| **French** | 6.480296 | 6.6 | 1.231397 |
| **German** | 6.472279 | 6.6 | 1.055929 |
| **Greek** | 6.468878 | 6.6 | 1.05099 |
| **Hebrew** | 6.469849 | 6.6 | 1.047776 |
| **Hindi** | 6.473737 | 6.6 | 1.071728 |
| **Hungarian** | 6.468878 | 6.6 | 1.05099 |
| **Icelandic** | 6.468878 | 6.6 | 1.05099 |
| **Indonesian** | 6.469121 | 6.6 | 1.04984 |
| **Italian** | 6.470335 | 6.6 | 1.047558 |
| **Japanese** | 6.47155 | 6.6 | 1.051065 |
| **Kazakh** | 6.468878 | 6.6 | 1.05099 |
| **Korean** | 6.470092 | 6.6 | 1.047551 |
| **Mandarin** | 6.472522 | 6.6 | 1.058005 |
| **Maya** | 6.468878 | 6.6 | 1.05099 |
| **Mongolian** | 6.468878 | 6.6 | 1.05099 |
| **None** | 6.468878 | 6.6 | 1.05099 |
| **Norwegian** | 6.469606 | 6.6 | 1.048233 |
| **Persian** | 6.469606 | 6.6 | 1.048233 |
| **Polish** | 6.468878 | 6.6 | 1.05099 |
| **Portuguese** | 6.470092 | 6.6 | 1.047551 |
| **Romanian** | 6.468878 | 6.6 | 1.05099 |
| **Russian** | 6.469363 | 6.6 | 1.048921 |
| **Spanish** | 6.475923 | 6.6 | 1.109874 |
| **Swedish** | 6.469121 | 6.6 | 1.04984 |
| **Telugu** | 6.468878 | 6.6 | 1.05099 |
| **Thai** | 6.469363 | 6.6 | 1.048921 |
| **Vietnamese** | 6.468878 | 6.6 | 1.05099 |
| **Zulu** | 6.468878 | 6.6 | 1.05099 |

**Insights:** English dominates the dataset in terms of number of movies, but other languages like French and Spanish also have high average scores. This can suggests that while English movies are most produced, quality storytelling transcends language barriers.

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

**Output 1:**

|  |  |
| --- | --- |
| **Top 10 Director\_name** | **Average of imdb\_score** |
| Alfred Hitchcock | 8.5 |
| Asghar Farhadi | 8.4 |
| Catherine Owens | 8.4 |
| Charles Chaplin | 8.6 |
| Christopher Nolan | 8.425 |
| Damien Chazelle | 8.5 |
| Majid Majidi | 8.5 |
| Marius A. Markevicius | 8.4 |
| Rakeysh Omprakash Mehra | 8.4 |
| Richard Marquand | 8.4 |
| Ron Fricke | 8.5 |
| S.S. Rajamouli | 8.4 |
| Sergio Leone | 8.433333333 |
| Tony Kaye | 8.45 |
| **Grand Total** | **8.441666667** |

**Output 2:** Percentile calculation:

|  |  |
| --- | --- |
| Large | 8.6 |
| Percent Rank | 0.692 |
| Percentile | 8.4998 |

**Insights:** The director’s vision, style, and storytelling clearly play a strong role in a movie's success, influencing its reception among viewers and critics alike.

1. **Budget Analysis:** Analyze the correlation between movie budgets and gross earnings, and identify the movies with the highest profit margin.

**Output:**

|  |  |
| --- | --- |
| **Correlation:** | **0.1009825** |
|  |  |
| **Profit Margin:** | **523505847** |

**Insights:** There is a moderate positive correlation between budget and gross earnings, indicating that bigger budgets often result in higher earnings.

**Project Description**

This project focuses on analyzing a dataset of IMDB movies to uncover factors that influence the success of a movie based on its IMDB score. Success, in this context, is primarily measured by higher ratings. The goal was to identify patterns and trends related to genre, duration, language, director performance, and budget-financial performance.

By evaluating these key variables, we aim to offer useful insights for movie producers, directors, and stakeholders to help them make better decisions in future film projects. The dataset used consists of multiple records across various attributes such as budget, gross, director, genre, language, duration, and more.

**Approach**

The analysis began with a thorough data cleaning phase to ensure consistency and accuracy in the dataset. We handled missing values, removed duplicate entries, standardized numerical formats (e.g., budget and gross), and derived useful new columns such as profit. The project was then broken down into five main analytical tasks—genre analysis, duration analysis, language influence, director analysis, and budget-profit exploration.

For each task, statistical functions in Excel such as AVERAGE, MEDIAN, STDEV, COUNTIF, CORREL, and PERCENTILE are used.

**Tech-Stack Used**

The primary software tool used in this project was Microsoft Excel 2022. It was selected for its powerful spreadsheet capabilities and built-in statistical functions. Excel's Pivot Tables were used to summarize categorical data like genres and languages. Scatter or Line plots and trendlines helped visualize relationships such as between duration and IMDB scores.

**Insights**

1. **Movies Genre :** Genres like Drama and Thriller consistently showed higher IMDB ratings, despite not being the most frequent in the dataset.
2. **Movie Duration:** Average-length movies (mid-minutes) performed better in ratings than very short or very long movies.
3. **Language Influence:** English was the dominant language in the dataset, but a few non-English movies showed very high average ratings, indicating quality over quantity.
4. **Director Analysis:** Directors with consistently high IMDB scores fell into the top percentile, demonstrating that creative leadership plays a major role in movie success.
5. **Budget and Profit:** Although high-budget movies earned higher gross revenues, some low-budget movies achieved the highest profit margins, showing that financial success isn’t solely investment-dependent.

**Result**

Through this project, I gained a strong understanding of how different factors contribute to a movie’s success on IMDB. I learned how to extract meaningful insights from raw data by applying statistical techniques and visualization tools. This project not only deepened my technical skills in Excel but also helped me improve my data storytelling and reporting abilities.

**Drive Link:** Here is the full Excel dataset and analysis can be accessed through the following link.

[**IMDB Movie Analysis**](file:///C:\Users\HP\OneDrive\Documents\Trainity\Trainity%20Projects\IMDB_Analysis.xlsx)