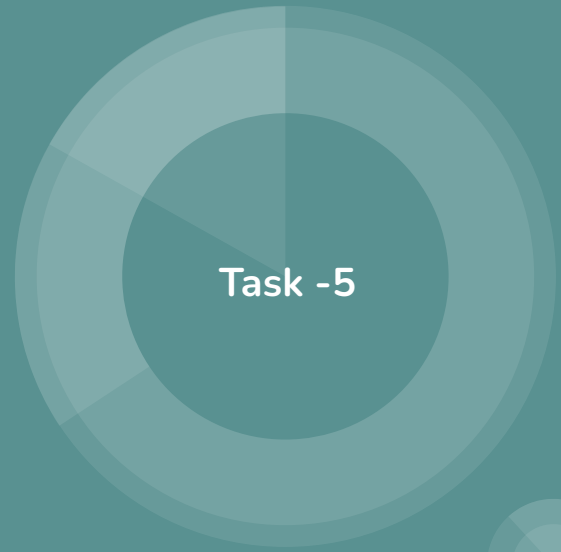


IMDB Movie Analysis



Email Id - dhritimlayadutta@gmail.com

Data Analytics Trainee





Introduction

The dataset provided is related to IMDB Movies. A potential problem to investigate could be: "What factors influence the success of a movie on IMDB?" Here, success can be defined by high IMDB ratings. The impact of this problem is significant for movie producers, directors, and investors who want to understand what makes a movie successful to make informed decisions in their future projects.

My goal is not just to answer questions but to provide insights that can drive decision-making. my analysis aim to provide actionable insights that can help stakeholders make informed decisions.



Methodology

Data Cleaning: This step involves preprocessing the data to make it suitable for analysis. It includes handling missing values, removing duplicates, converting data types if necessary, and possibly feature engineering.

Data Analysis: Here, we explore the data to understand the relationships between different variables. We might look at the correlation between movie ratings and other factors like genre, director, budget, etc. we might also want to consider the year of release, the actors involved, and other relevant factors.

After analysis, we will create a report that tells a story with data. This includes initial problem, findings, and the insights gained.

Here we use visualizations to help tell story and make findings more understandable.



Data Cleaning

This is one of the most important steps of Data Analysis.

- First Dropping some of the irrelevant columns like - 'Color', 'director_facebook_likes', 'actor_3_facebook_likes', 'actor_2_name', 'actor_1_facebook_likes' etc.
- After dropping the irrelevant columns now we need to remove the rows from the dataset having any one of its column value as blank/NULL.
- Then we need to get rid off the duplicate values in the dataset which can be achieved by using the 'Remove Duplicate Values/Cells' available in the 'Data' tab

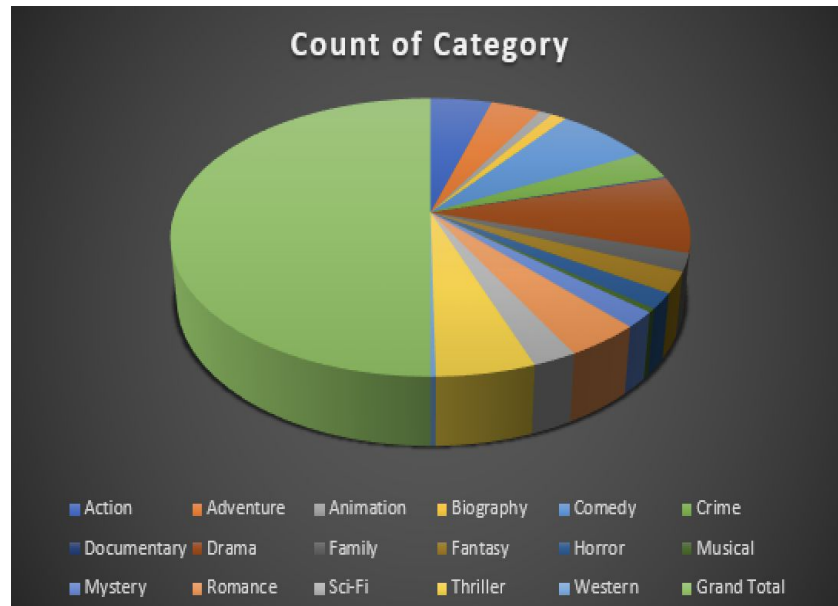


Project Description

- A. Movie Genre Analysis:** Analyze the distribution of movie genres and their impact on the IMDB score.
- B. Movie Duration Analysis:** Analyze the distribution of movie durations and its impact on the IMDB score.
- C. Language Analysis: Situation:** Examine the distribution of movies based on their language.
- D. Director Analysis:** Influence of directors on movie ratings.
- E. Budget Analysis:** Explore the relationship between movie budgets and their financial success.

Software Used : Microsoft Excel

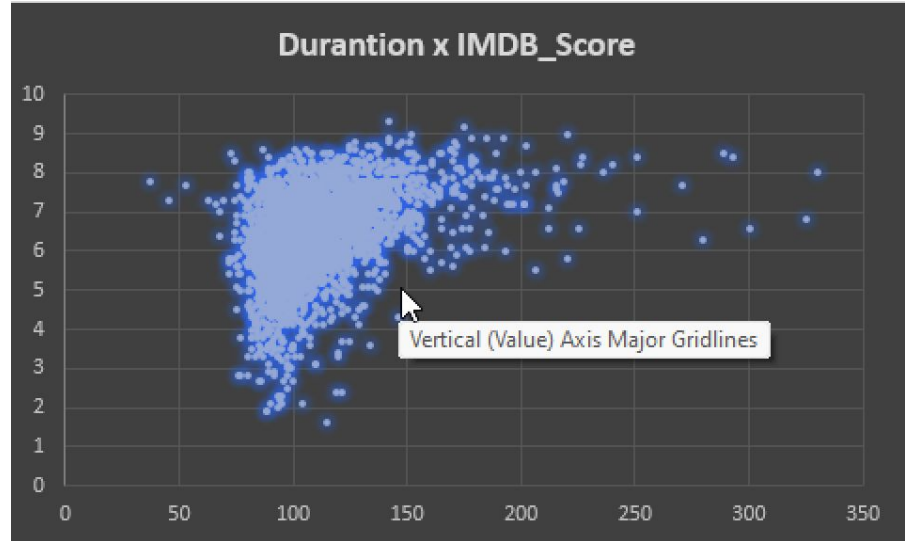
A. Movie Genre Analysis: Determine the most common genres of movies in the dataset. Then, for each genre, calculate descriptive statistics (mean, median, mode, range, variance, standard deviation) of the IMDB scores.



| Genres Maniplutaioir | |
|----------------------|----------|
| Median | 493 |
| Max | 1877 |
| Min | 42 |
| Mode | 380 |
| Var | 242990.6 |
| Sd | 492.9408 |

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

The distribution of movie durations varied widely, ranging from short films to lengthy epics. However, we observed a correlation between movie duration and IMDB score, suggesting that movie length alone does not significantly influence audience ratings.



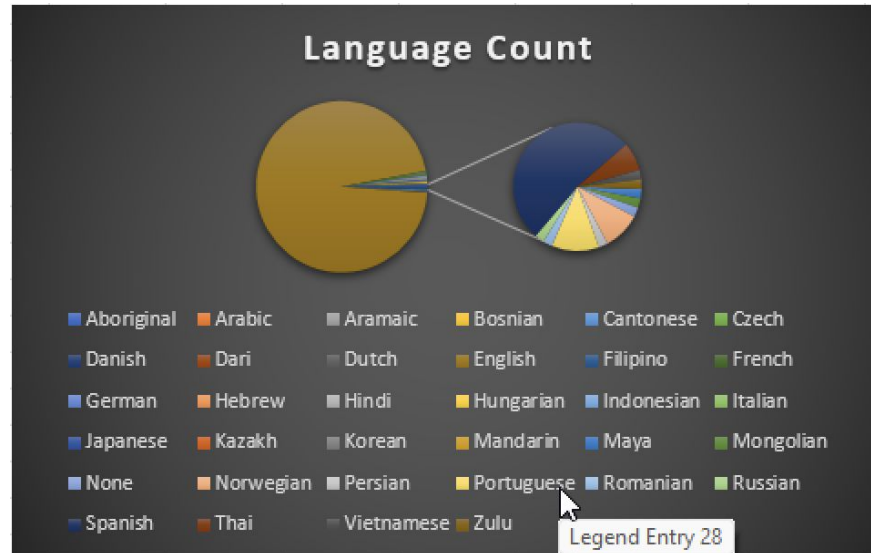


C. Language Analysis: Determine the most common languages used in movies and analyze their impact on the IMDB score using descriptive statistics.

| Languages | Count | Mean | Median | SD |
|------------|-------|-----------|--------|----------|
| Aboriginal | 2 | 5.95 | 6.5 | 1.059507 |
| Arabic | 1 | 6.9 | 6.5 | 1.059507 |
| Aramaic | 1 | 5.2 | 6.5 | 1.059507 |
| Bosnian | 1 | 7.1 | 6.5 | 1.059507 |
| Cantonese | 7 | 7.3428571 | 6.5 | 1.059507 |
| Czech | 1 | 7.3 | 6.5 | 1.059507 |
| Danish | 3 | 5.9333333 | 6.5 | 1.059507 |
| Dari | 2 | 7.35 | 6.5 | 1.059507 |
| Dutch | 3 | 5.9666667 | 6.5 | 1.059507 |
| English | 3581 | 6.4577149 | 6.5 | 1.059507 |
| Filipino | 1 | 4.8 | 6.5 | 1.059507 |
| French | 32 | 6.678125 | 6.5 | 1.059507 |
| German | 11 | 6.2727273 | 6.5 | 1.059507 |
| Hebrew | 2 | 7.3 | 6.5 | 1.059507 |
| Hindi | 5 | 6.54 | 6.5 | 1.059507 |
| Hungarian | 1 | 7 | 6.5 | 1.059507 |
| Indonesian | 2 | 7.55 | 6.5 | 1.059507 |
| Italian | 5 | 7.4 | 6.5 | 1.059507 |
| Japanese | 10 | 6.23 | 6.5 | 1.059507 |

| | | | | |
|------------|----|-----------|-----|----------|
| Kazakh | 1 | 3.3 | 6.5 | 1.059507 |
| Korean | 5 | 6.74 | 6.5 | 1.059507 |
| Mandarin | 15 | 5.8866667 | 6.5 | 1.059507 |
| Maya | 1 | 7.7 | 6.5 | 1.059507 |
| Mongolian | 1 | 5.6 | 6.5 | 1.059507 |
| None | 1 | 7.6 | 6.5 | 1.059507 |
| Norwegian | 4 | 6.5 | 6.5 | 1.059507 |
| Persian | 1 | 7.7 | 6.5 | 1.059507 |
| Portuguese | 5 | 6.18 | 6.5 | 1.059507 |
| Romanian | 1 | 6.4 | 6.5 | 1.059507 |
| Russian | 1 | 7.5 | 6.5 | 1.059507 |
| Spanish | 22 | 6.5363636 | 6.5 | 1.059507 |
| Thai | 3 | 6.2333333 | 6.5 | 1.059507 |
| Vietnamese | 1 | 7.3 | 6.5 | 1.059507 |
| Zulu | 1 | 7.5 | 6.5 | 1.059507 |

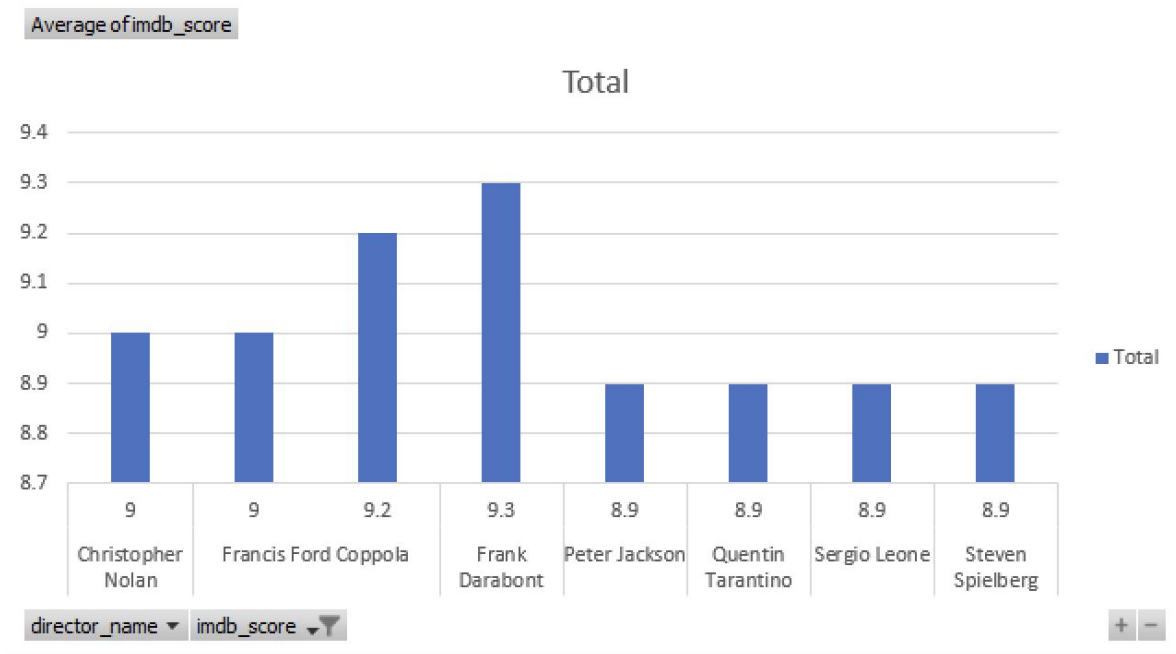
C. Language Analysis: Determine the most common languages used in movies and analyze their impact on the IMDB score using descriptive statistics.



Language English was the predominant language in the dataset, Followed by Spanish and French



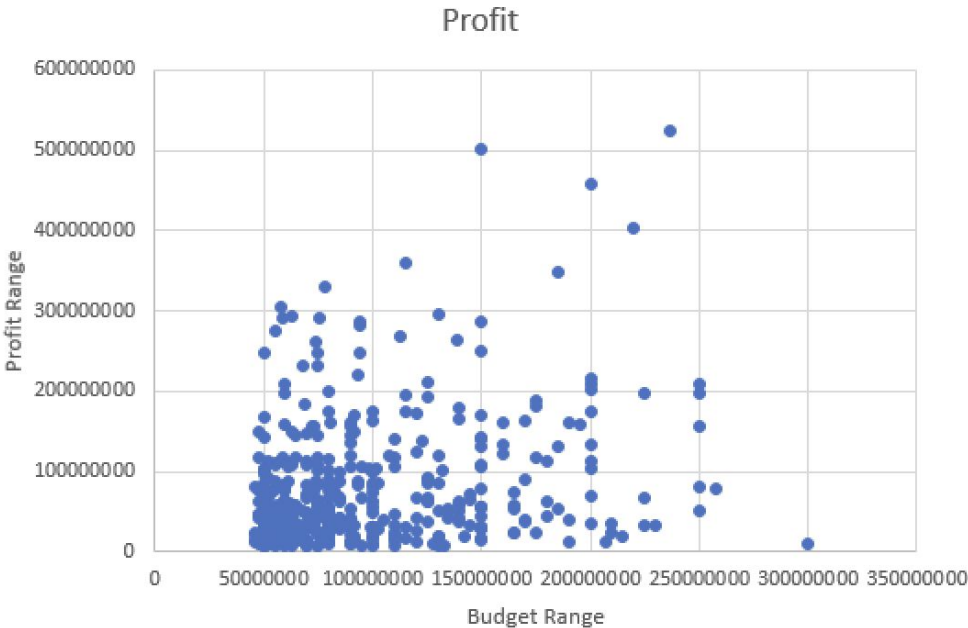
D. Director Analysis: identify the top directors based on their average IMDB score and analyze their contribution to the success of movies using percentile calculations.





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

| MAX | MIN |
|-----------|--------------|
| 523505847 | -12213298588 |
| Avatar | Tomcats |





Conclusion & Link

Through this project, we gained valuable insights into the factors influencing movie ratings and financial performance. By analyzing genres, durations, languages, directors and budgets, we were able to identify trends and patterns that shed light on audience preferences and industry dynamics. Our findings provide actionable insights for filmmakers, producers, and studios seeking to optimize their movie production strategies and maximize audience appeal.

Final Dataset with Analysis -

<https://docs.google.com/spreadsheets/d/1e8MWlSkIBBdSvTAgHLl71mIRWJ7OQwPz/edit?usp=sharing&oid=118268620399913105117&rtpof=true&sd=true>

Loom Video Presentation -

<https://www.loom.com/share/cf444364c9fe438d9b8cd7b40d8424db?sid=180dfef8-678e-4da9-9cb6-0c4db56251b8>