IMDB Movie Analysis

22.11.2024

Vinay Rana

vr.ranavinay@gmail.com

Project Summary:

The goal of this project is to perform an in-depth analysis of the IMDB movie dataset. The study will examine crucial elements such as movie genres, runtime, languages, directors, and production budgets. By investigating these aspects, the project seeks to uncover their influence on IMDB ratings and the financial performance of movies.

Approach

We will start by understanding the dataset and its structure. Then, we will clean the data by fixing missing values, removing unnecessary columns, and preparing it for analysis. After that, we'll use tools like pivot tables, charts, and functions to answer key questions. By digging deeper into the data, we aim to uncover important insights. Finally, the results will be clearly presented using tables and graphs.

Tools and Software

Software: Microsoft Excel Version 16.91

Reason: Microsoft Excel 2024 offers robust features like pivot tables, charts, and statistical functions, making it an excellent choice for analyzing, visualizing, and interpreting trends in the hiring data with precision and efficiency.

Data Cleaning Instructions

Before proceeding with the analysis, apply your existing knowledge to clean the dataset effectively. Follow these steps:

Drop Unnecessary Columns

Remove irrelevant columns to streamline the dataset. Columns to drop include:

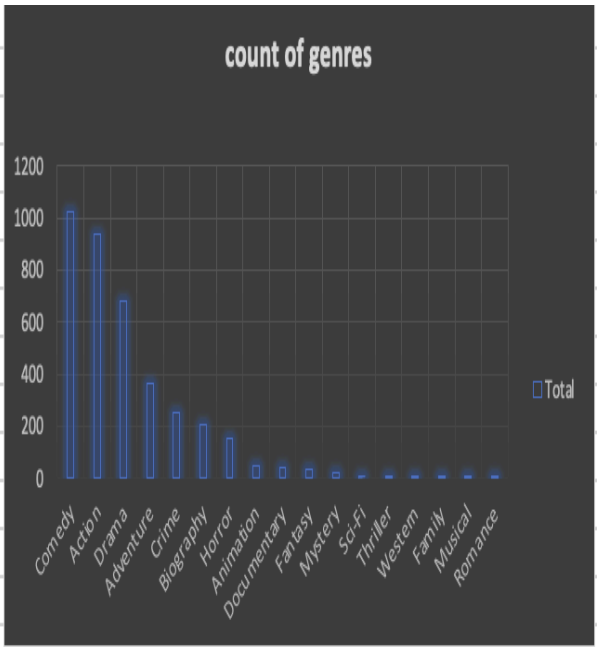
color	actor_1_facebook_likes	actor_3_name	movie_imdb_link	aspect_ratio	
director_facebook_likes	cast_total_facebook_likes	facenumber_in_posts	content_rating	movie_facebook_likes	
actor_3_facebook_likes	actor_2_name	plot_keywords	actor_2_facebook_likes		

Movie Genre Analysis:

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.

I examined the dataset to determine the most common genres and evaluated their influence on IMDb scores. The findings revealed that Action, Comedy, and Drama were the most frequently occurring genres.

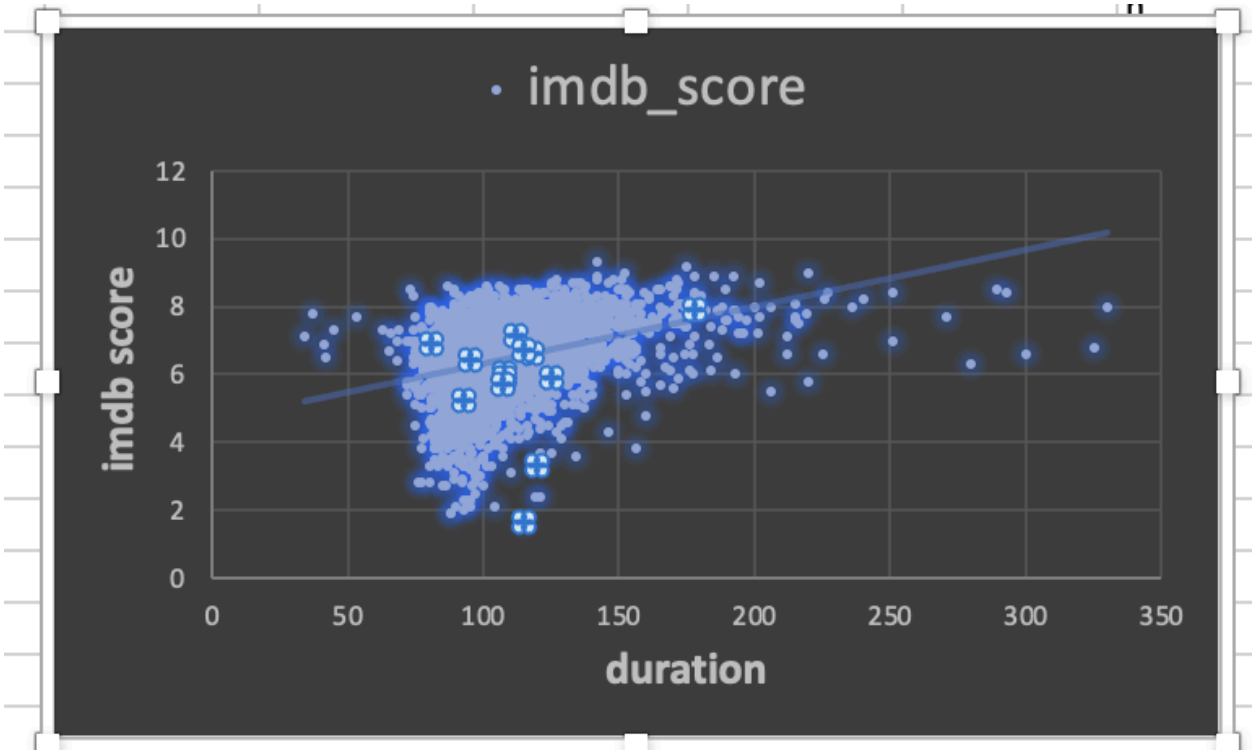
Genres	Count of genres
Comedy	1025
Action	935
Adventure	367
Biography	206
Animation	46
Grand Total	2579



Movie Duration Analysis:

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

The dataset revealed a wide variation in movie durations, ranging from short films to lengthy epics. However, the analysis showed a weak correlation between movie duration and IMDb scores, suggesting that duration alone has minimal impact on audience ratings. The scatter plot indicates a slight tendency for movies longer than 250 minutes to achieve IMDb scores between 6 and 8.5. Additionally, the trend line suggests a potential increase in IMDb scores as movie duration increases, though the effect is not strong.



Language Analysis:

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

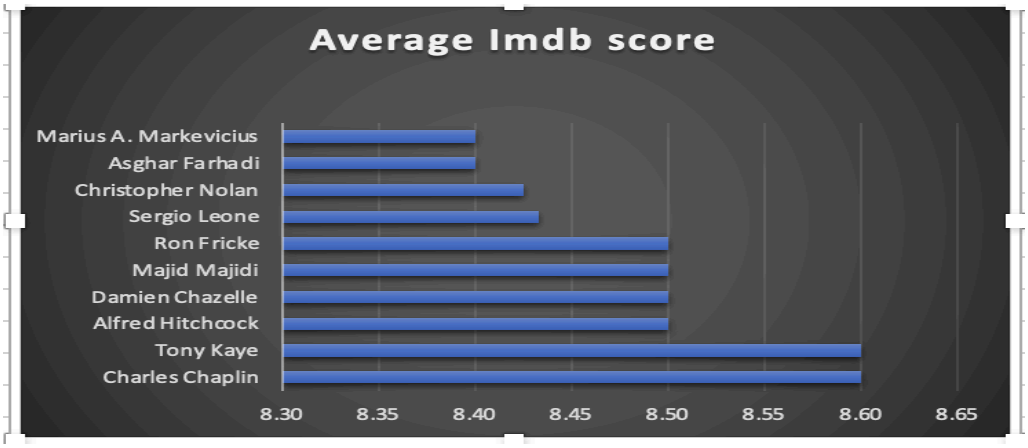
English emerged as the most prevalent language in the dataset, followed by French and German. Interestingly, films in non-English languages tended to have slightly lower average IMDb scores, possibly reflecting biases in audience preferences. Below is a list of the top ten common languages observed in the dataset.

unique language	Count of language
English	3603
French	37
Spanish	26
Mandarin	14
German	13
Japanese	12
Hindi	10
Cantonese	8
Italian	7
Portuguese	5

Director Analysis:

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

The analysis identified top directors based on their average IMDb scores, highlighting their contributions to movie success. Renowned filmmakers such as Charles Chaplin, Tony Kaye, and Alfred Hitchcock consistently delivered movies with high ratings, underscoring their significant impact on audience reception and overall cinematic excellence.

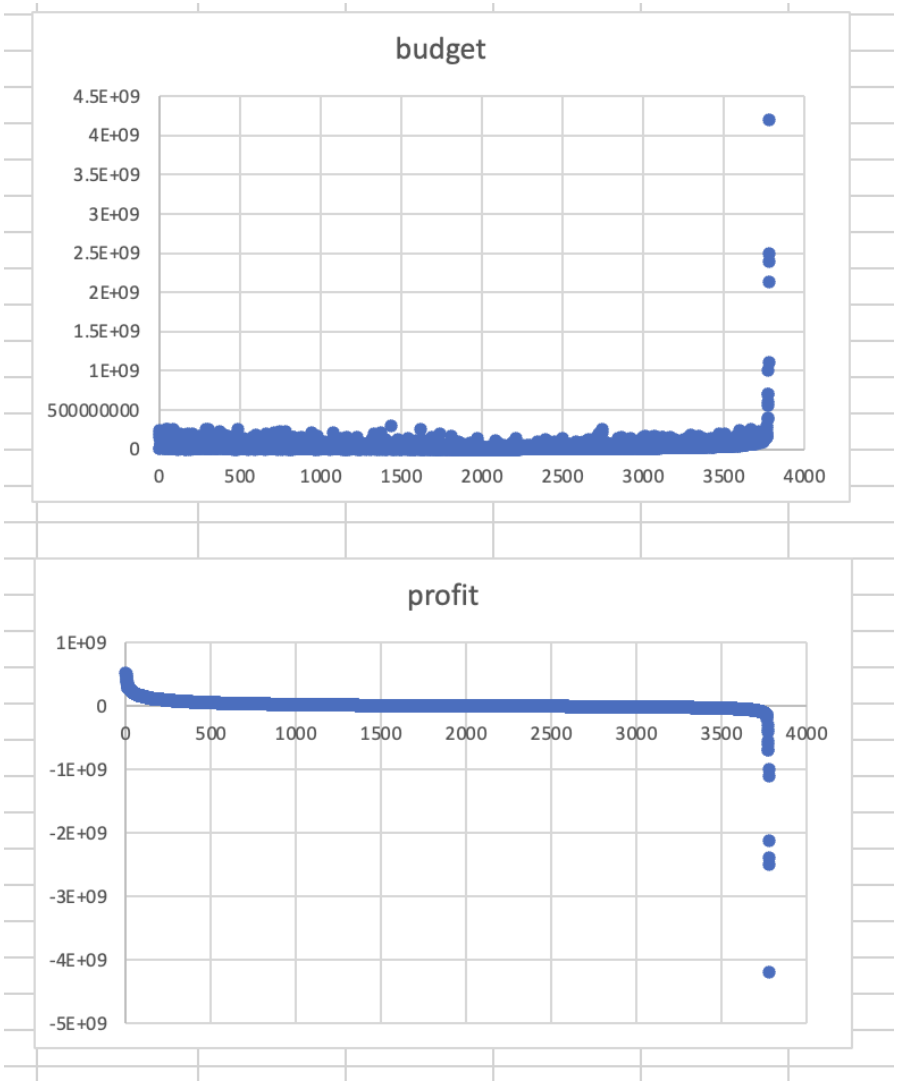


Top 10 Director name	Average of imdb_score
Charles Chaplin	8.60
Tony Kaye	8.60
Alfred Hitchcock	8.50
Damien Chazelle	8.50
Majid Majidi	8.50
Ron Fricke	8.50
Sergio Leone	8.43
Christopher Nolan	8.43
Asghar Farhadi	8.40
Marius A. Markevicius	8.40

Budget Analysis:

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

I analyzed the relationship between movie budgets and financial performance by examining the correlation between production budgets and gross earnings. The results revealed a positive correlation, indicating that higher budgets often lead to greater financial success. However, the strength of this relationship varied based on factors such as genre and the director's influence.



Result:

Summary of Analysis:

Task 1: The top three most popular genres identified are Comedy, Action, and Drama.

Task 2: Movies with a duration of around 200 minutes have a higher likelihood of achieving an IMDb score greater than 6.

Task 3: The top three languages used in movies are English, French, and Spanish.

Task 4: The top three directors known for consistently delivering highly rated movies are Charles Chaplin, Tony Kaye, and Alfred Hitchcock.

Task 5: The top three movies with the highest profit margins are Avatar, Jurassic World, and Titanic.

Through this project, we uncovered valuable insights into the factors affecting movie ratings and financial performance. By examining genres, durations, languages, directors, and budgets, we identified key trends and patterns that highlight audience preferences and industry dynamics. These findings offer actionable recommendations for filmmakers, producers, and studios aiming to enhance their production strategies and boost audience engagement.

Excel