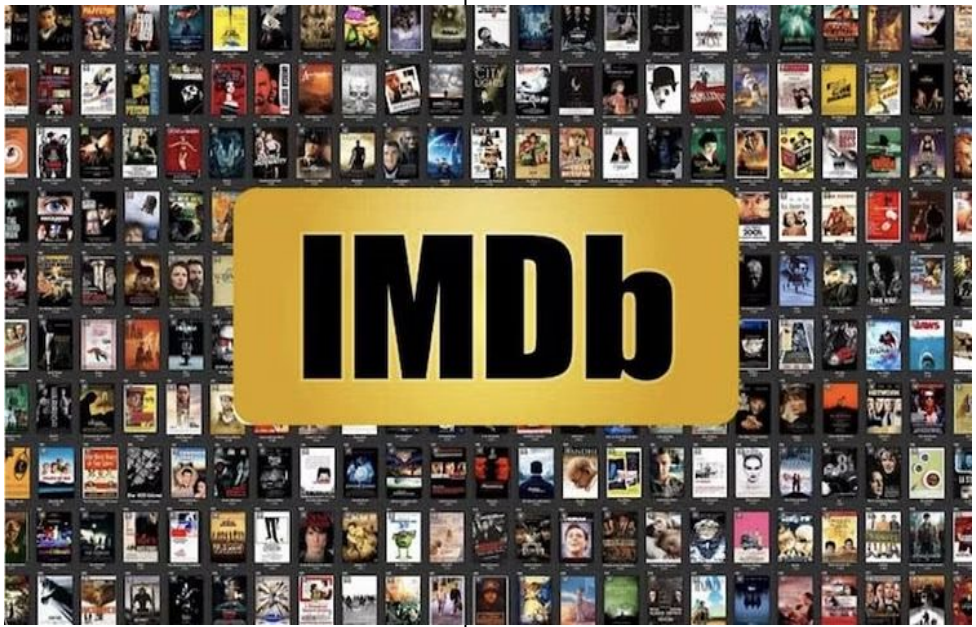
IMDB Movie Analysis

Divyasri Jegan



Project Description :

- A potential question for filmmakers 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 question 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.

- This Project is about giving insights about success of a Movie based on IMDB data provided which will be helpful for filmmakers and other stakeholders during production of a movie.

Approach :

1. Tech Stack Used
2. Dataset Overview
3. Data Pre-Processing
4. Insights
5. Conclusion
6. Links

Tech Stack Used :

- Extract, manipulate and process the data to make decisions.
- Make interactive charts for better visualization.

Prepare a presentation to tell the data story.

Dataset Overview :

- Source of Data: https://drive.google.com/file/d/1XpGThHzLnXxL_7aQo2sCpYL3SeB18MMB/view?usp=sharing
- The dataset provided is related to IMDB Movies and contains records of movies from a number of years and geographical locations.
- The Dataset details are:

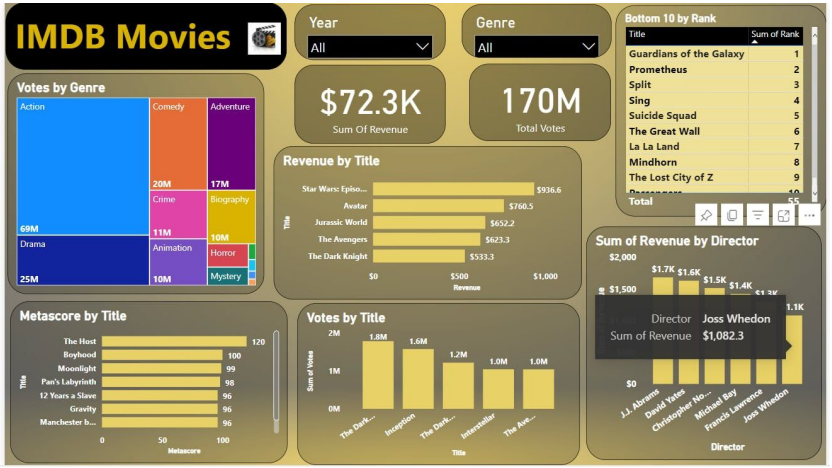
  - Number of Data-Points: 5,043
  - Number of Features: 28
  - Column Details:

    Color, director_name, num_critic_for_reviews, duration, director_facebook_likes, Actor_3_facebook_likes, actor_2_name, actor_1_facebook_likes, gross, genres, Actor_1_name, movie_title, num_voted_users, cast_total_facebook_likes, actor_3_name, Facenumber_in_poster, plot_keywords, movie_imdb_link, num_user_for_reviews, Language, Country, content_rating, budge,t title_year, actor_2_facebook_likes, imdb_score, Aspect_ratio, movie_facebook_likes …

Insights :

Data Analytics Tasks:

1. Movie Genre Analysis
2. Movie Duration Analysis
3. Language Analysis
4. Director Analysis
5. Budget Analysis

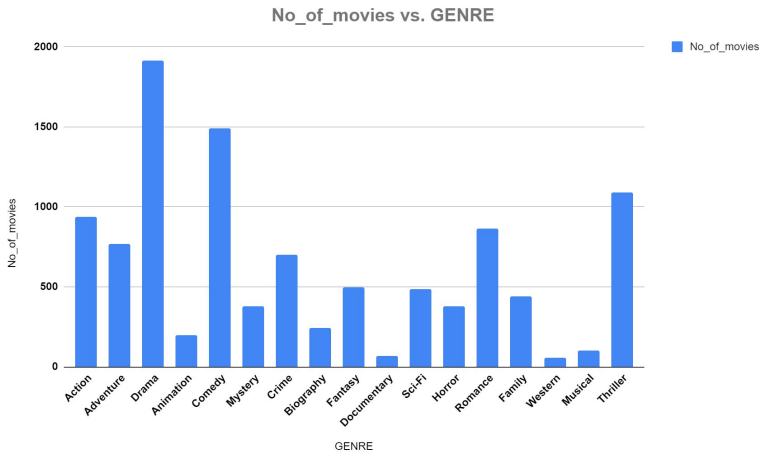


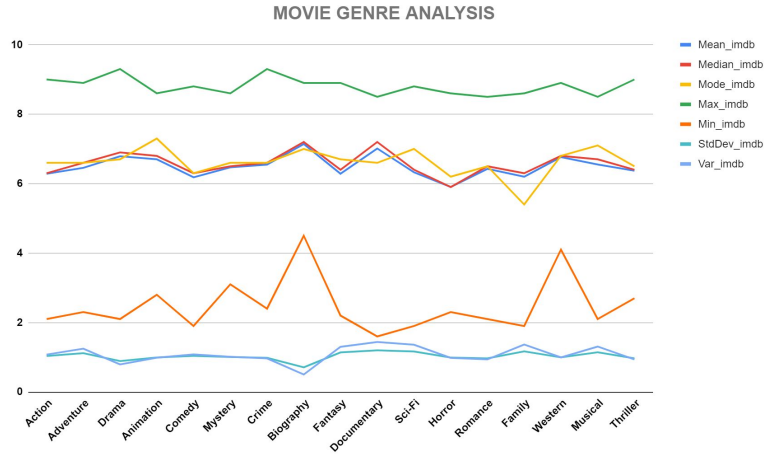
01

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.
- **Result:** The top 7 most common genres are **Drama**, **Comedy**, **Thriller**, **Action**, **Romance**, **Adventure** and **Crime.** Also all the top 7 genres' descriptive statistics are almost at same level.



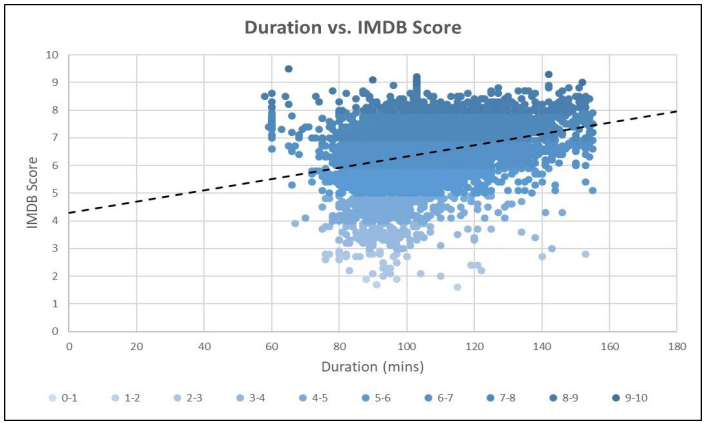


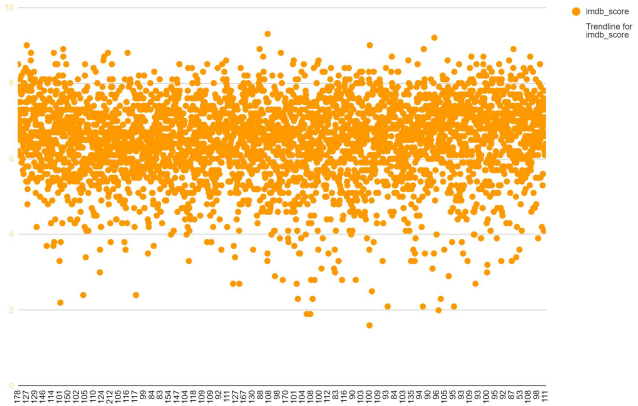
02

Movie Duration Analysis

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.
- **Result:** The distribution of **Movie Durations** shows that it closely follows a **Normal Distribution**. Also the scatter plot shows that **duration** and **imdb_scores** have a positive relationship.



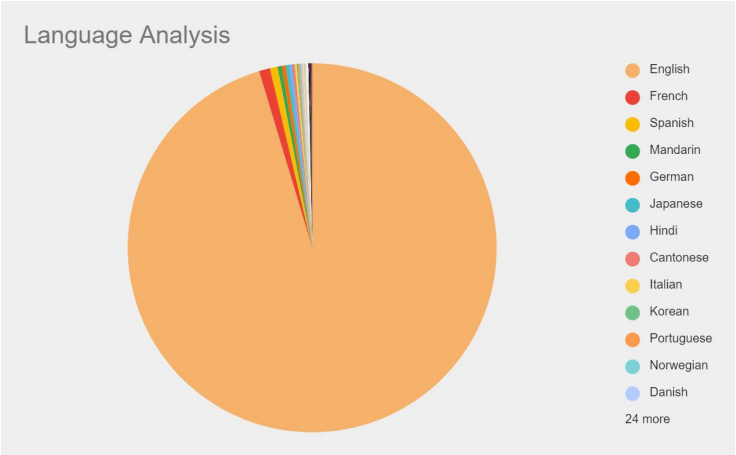


03

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.
- **Result:** The left plot show that **English** is the most common language used in movies followed by **French**, **Spanish**, **Hindi** and **Mandarin**. The right plot shows that **French** language has comparatively higher **mean** and **median** but lower **variance** and **standard deviation** implying that most of the French language movies have their **imdb** score on the higher side.

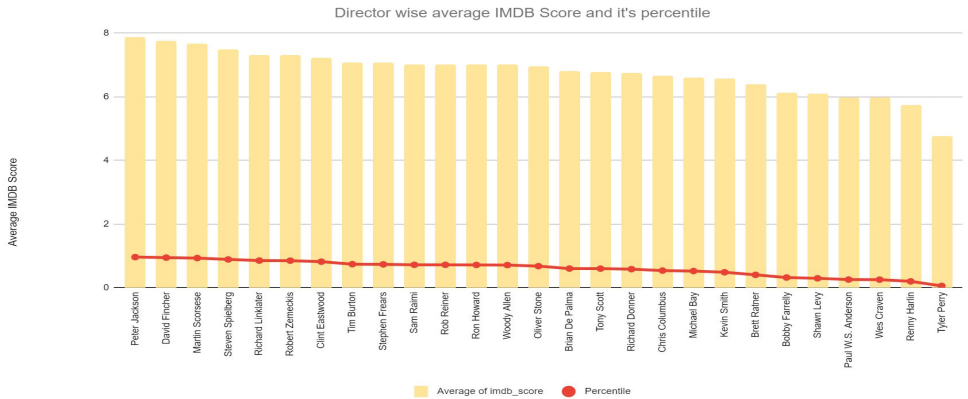


04

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.
- **Result:** The plot considers only those directors whose **movie counts are more than 9** and **the range of IMDB Scores is less than equal to 3** as otherwise it would be unfair for those who has maintained consistently high scores for large number of movies to be compared for top directors to those who has performed well in few movies.

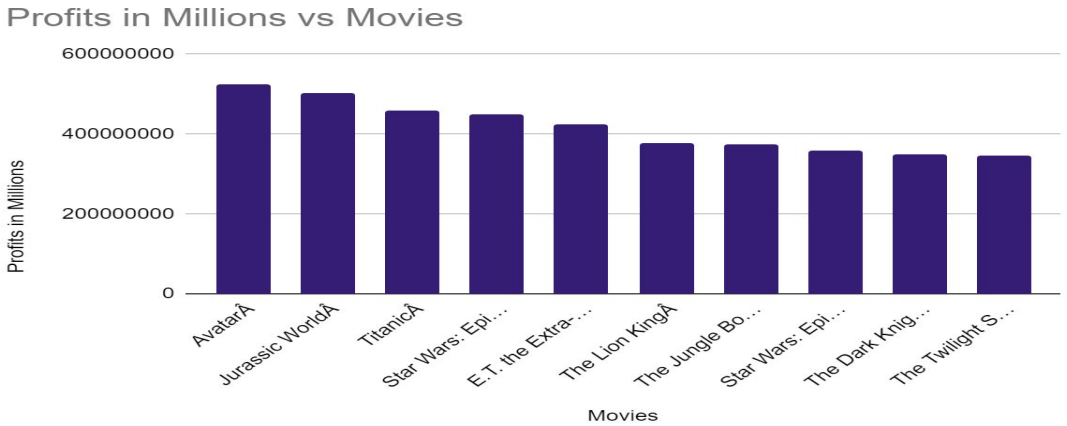


05

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**.**
- **Result:** The table shows that the correlation between **Gross** and **Budget** is **positive** and more than **0.5**. That is, the relationship shows that as budget of movies increase, there is a very high probability that the gross collection of the movie will also increase. The plot shows the relationship between **Gross** and **Budget**. The overall trendline has a slope close to **1**.



Links:

Source dataset :
https://drive.google.com/file/d/1XpGThHzLnXxL_7aQo2sCpYL3SeB18MMB/view?usp=sharing

Google sheet :
https://docs.google.com/spreadsheets/d/1X1tjxrw9C2sRbrcvVmY9-h1aIm-zvkI4iXwdHdFL8G4/edit?usp=sharing

Conclusion:

- Through this project, I was able to understand the importance of **Data Analytics** in **Movies analysis** as it provides valuable insights such as director's relationship with IMDB Score, genre's relationship with IMDB Score, budget's relationship with IMDB Score etc. which helps in making **Data-Driven Decisions**.

- This helps in better usage and handling of data in excel and it provide a strong foundation on it too..