

# Data Science – Report Task 3

## Project Overview:

This project aims to create an interactive and insightful dashboard that explores various factors influencing the success of movies based on IMDb ratings, gross earnings, genre, certification, and more. By utilizing the IMDb top 1000 movie dataset, we created multiple visualizations to better understand the relationships between these attributes. The project uses the Dash framework for web development and Plotly for generating interactive visualizations.

## Key Features and Visualizations:

1. Bar Chart: Top 10 Movies by Gross Earnings
  - Description: This bar chart showcases the top 10 highest-grossing movies in the dataset, allowing users to see which films have earned the most.
  - Key Insight: It highlights blockbuster hits that dominate the box office and provides insights into which films, based on earnings, are likely to have left a significant cultural impact.
2. Scatter Plot: IMDb Rating vs Gross Earnings
  - Description: This visualization explores the relationship between a movie's IMDb rating and its gross earnings. The scatter plot plots the IMDb ratings on the y-axis against the gross earnings on the x-axis. The size of the points is proportional to the number of votes, and the color corresponds to the movie title.
  - Key Insight: It highlights how higher-rated movies often correlate with greater financial success, providing a glimpse into how audience reception can impact a film's box office performance.
3. Pie Chart: Gross Earnings Based on Certificates
  - Description: This pie chart visualizes the gross earnings distribution across different movie certification categories (PG, R, PG-13, etc.). The chart shows which certification categories contribute most significantly to the gross earnings in the dataset.
  - Key Insight: It reveals how the classification of movies into different certificate categories (family-friendly, mature, etc.) influences their earnings, providing valuable insights for understanding the financial impact of movie ratings.

#### 4. Bar Chart: Average IMDb Rating by Certificate

- Description: This bar chart compares the average IMDb ratings of movies across different certification categories. It allows for easy comparison of the reception of films in various rating groups.
- Key Insight: It uncovers whether family-friendly movies (e.g., PG) tend to get higher ratings compared to more intense films (e.g., R-rated), providing valuable information for movie enthusiasts and filmmakers alike.

#### 5. Bubble Chart: Gross Earnings by Released Year and Runtime

- Description: This interactive bubble chart shows how movie runtime and release year influence gross earnings. The chart displays each movie as a bubble, with the x-axis representing the year of release, the y-axis representing the gross earnings, and the bubble size representing the movie's runtime.
- Key Insight: This chart reveals trends over time in terms of movie length and their corresponding financial success, showcasing how the movie industry's financial performance has changed based on these two attributes.

#### Data Preprocessing and Analysis:

- Dataset Cleaning:
  - Removed null values from the dataset to ensure the accuracy of the visualizations.
  - Converted the 'Released\_Year' column to numeric format and handled any errors in conversion.
  - Cleaned the 'Gross' column by removing commas and extracting only numeric values, making it suitable for analysis.
  - The 'Runtime' column was cleaned by removing the 'min' text and converted to a numeric format to allow for meaningful analysis.
- Exploratory Data Analysis (EDA):
  - Investigated the dataset for missing values and corrected any data inconsistencies.
  - Used summary statistics to gain an understanding of the dataset's overall structure and key trends, such as the most common genres, average ratings, and gross earnings.

## User Interaction and Features:

The dashboard is built with Dash, making it interactive for users. Each visualization offers hover-over interactions, where users can see additional details such as movie titles, ratings, and other related attributes.

- The scatter plot allows users to explore the relationship between gross earnings and IMDb ratings visually.
- The pie chart provides insights into how different certifications (e.g., PG, PG-13) are associated with movie earnings.
- The bar charts let users compare the average IMDb ratings of different certification categories and view the top 10 movies by gross earnings.
- The bubble chart offers a deeper analysis of how movie length and release year affect box office performance.

## Insights:

- **Certification and Earnings:** The gross earnings by certification chart shows that certain movie certifications (like PG-13) dominate the box office, making it clear that audience age group and content restrictions play a major role in determining financial success.
- **IMDb Rating vs Gross Earnings:** The scatter plot reveals that, generally, higher-rated movies tend to have higher earnings. This highlights the influence of audience and critic approval on a movie's success.
- **Runtime and Box Office Trends:** The bubble chart shows that longer movies, especially from the 2000s and beyond, tend to perform better in terms of gross earnings. This might suggest that longer movies are becoming more common in the blockbuster space.
- **Audience Preferences:** The bar chart on average ratings by certificate reveals that films rated PG or PG-13 tend to have higher average ratings, possibly indicating that lighter films are more favourably received by audiences.

## Future Improvements:

- **More Interactivity:** Future iterations of the dashboard could include filter options for viewers to drill down into specific genres, release years, or gross earnings ranges to explore the data in more detail.
- **Additional Visualizations:** Adding trend lines, heat maps, or detailed genre breakdowns could further enhance the analysis, helping to uncover more nuanced insights.

- **Data Updates:** The dataset could be regularly updated with the latest movie data to keep the dashboard fresh and relevant.

## **Conclusion:**

This project offers a comprehensive and interactive way to explore how various factors, such as movie certification, runtime, and IMDb ratings, impact the box office performance of films. By using the power of Plotly visualizations and Dash interactivity, the project delivers a rich, engaging experience for users to analyze and understand movie trends and success factors in a data-driven manner.