# CEN445 INTRODUCTION TO DATA VISUALIZATION COURSE - ASSIGNMENT 2

**Dataset Name:** Movie Industry

**Data Published by:** DANIEL GRIJALVA

**Source:** Kaggle

**Youtube Link:** https://www.youtube.com/watch?v=VMkj-qgbRLI

**Code Link:**
https://colab.research.google.com/drive/1RG8DFMpz4TS5uk7k6sJG7zmV_FVR_vfO#scrollTo=7SIIEF_oS-V2

**Dataset Kaggle Link:** https://www.kaggle.com/datasets/danielgrijalvas/movies/data

## Description

The dataset used in this exploratory analysis was collected by scraping HTML data from IMDb, utilizing a Python script. The dataset comprises information related to various movies, including details such as "movie name, rating, genre, release year, director, cast, country, budget, gross revenue", and other relevant attributes. The data is diverse, encompassing different genres, years, and production companies, offering a comprehensive view of the film industry.

## Data Fields

The dataset consists of several fields, each providing valuable insights into the movie landscape. Some key attributes include:

**Name**: The title of the movie.

**Rating**: The movie's rating (e.g., G, PG, R).

**Genre**: The genre or category to which the movie belongs (e.g., Drama, Action, Comedy).

**Year**: The year of the movie's release.

**Released**: The release date, including additional details such as the country.

**Score**: The IMDb score assigned to the movie.

**Votes**: The number of votes received by the movie.

**Director, Writer, Star:** People involved in the making of the movie.Country: The country associated with the movie's production.

**Budget, Gross:** Financial details, including the movie's budget and gross revenue.

**Company:** The production company responsible for the movie.

**Runtime**: The duration of the movie.

**Questions:**

1) How is the distribution of average box office revenues across different movie genres?

2) How have the average film budgets and revenues changed over the years, and how did the increasing film budgets impact the revenues?

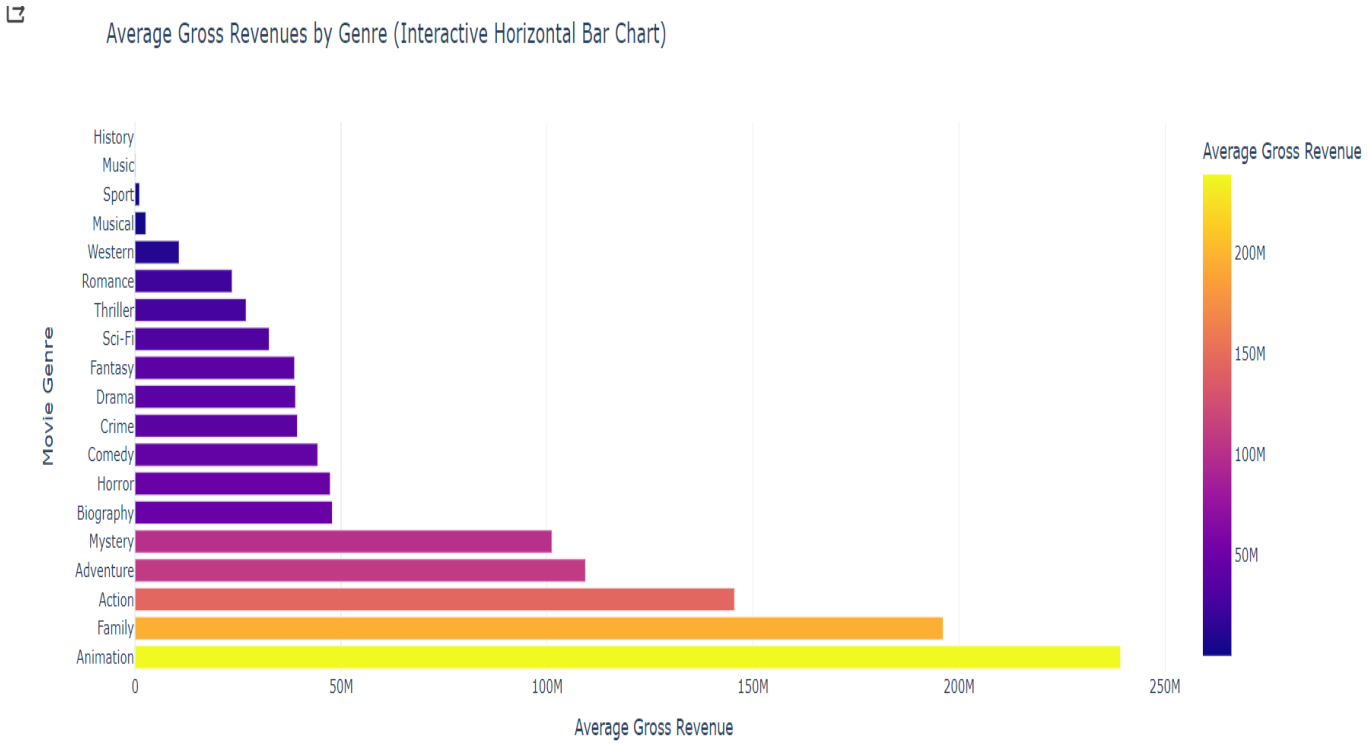
3) Which film production companies have the highest budgets and grossing revenues and how do they compare to each other?

4) Is there a visible relationship between the IMDb scores of movies and the number of users voting?

5) How many films were produced in the top 10 countries, and what is the relationship between the number of films produced in these countries and their IMDb scores?

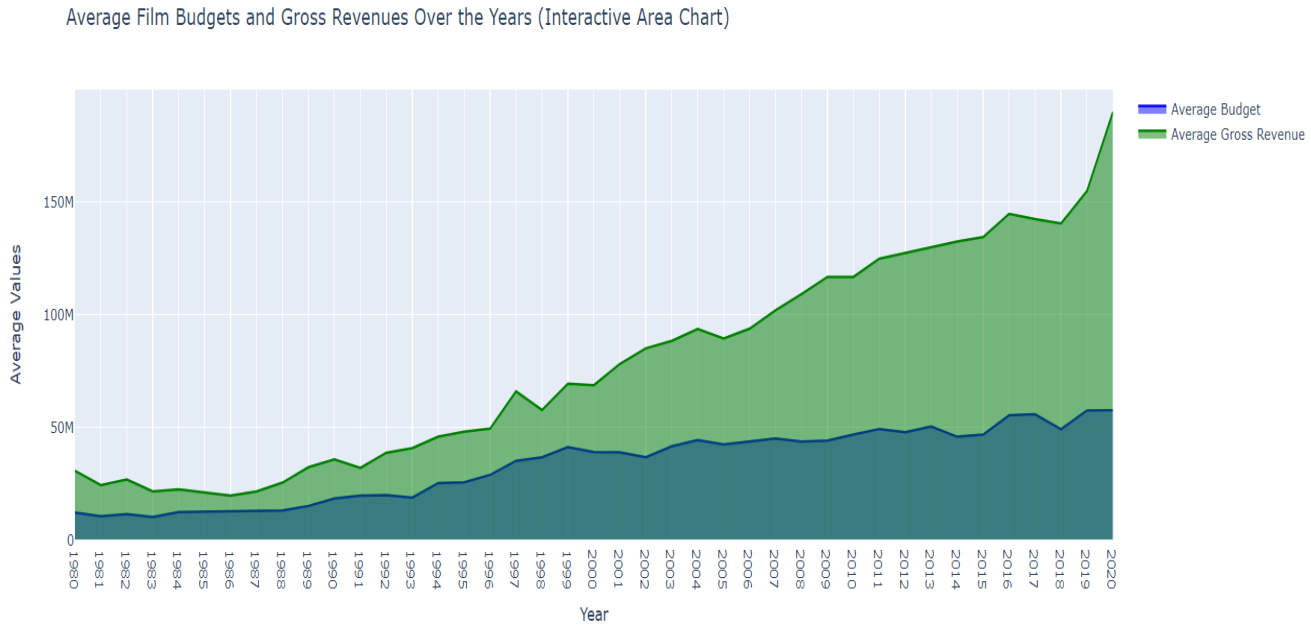
**Answers:**

**1)**




According to the analysis of the dataset, the distribution of average box office revenues varies significantly across different movie genres. The top three genres that consistently generate the highest average gross revenues are Animation, Family, and Action, suggesting a strong box office performance in these categories. On the other hand, genres such as History, Music, and Sports tend to have lower average gross revenues, indicating comparatively less financial success. This insight is derived from the horizontal bar chart, where Animation, Family, and Action genres stand out as the top performers, while History, Music, and Sports genres exhibit lower box office revenues.

**2)**

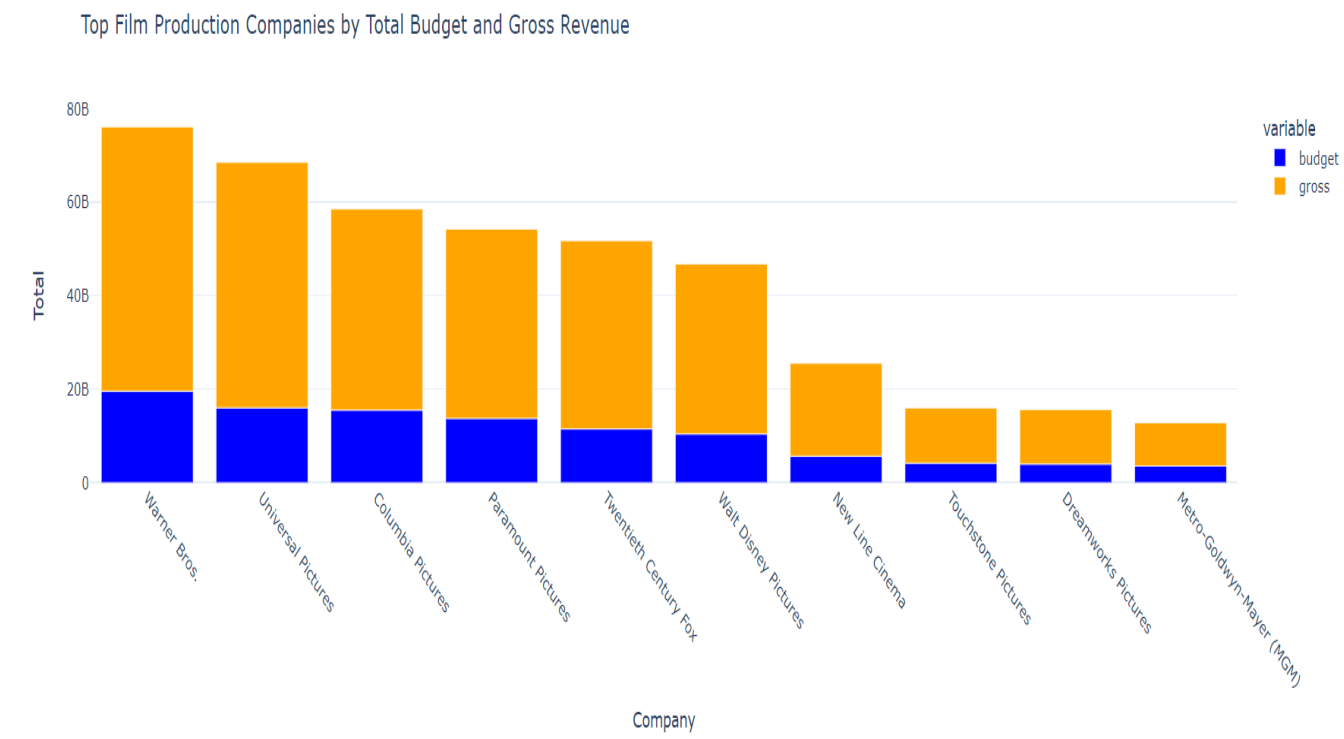



Upon analyzing the average film budgets and gross revenues over the years, a compelling trend emerges. Until the early 2000s, there seems to be a proportional increase in average gross revenues with the rise in film budgets, indicating a positive correlation between investment in movies and box office success.

However, a noteworthy shift occurs in the 2000s, where average budgets have plateaued while average gross revenues continue to experience substantial growth. This divergence suggests that, in recent years, the correlation between film budgets and box office revenues has become less direct. The observed surge in gross revenues from around 2020 could be attributed to factors beyond budget alone, such as a growing population, increased global accessibility, and the influential role of social media in film promotion.

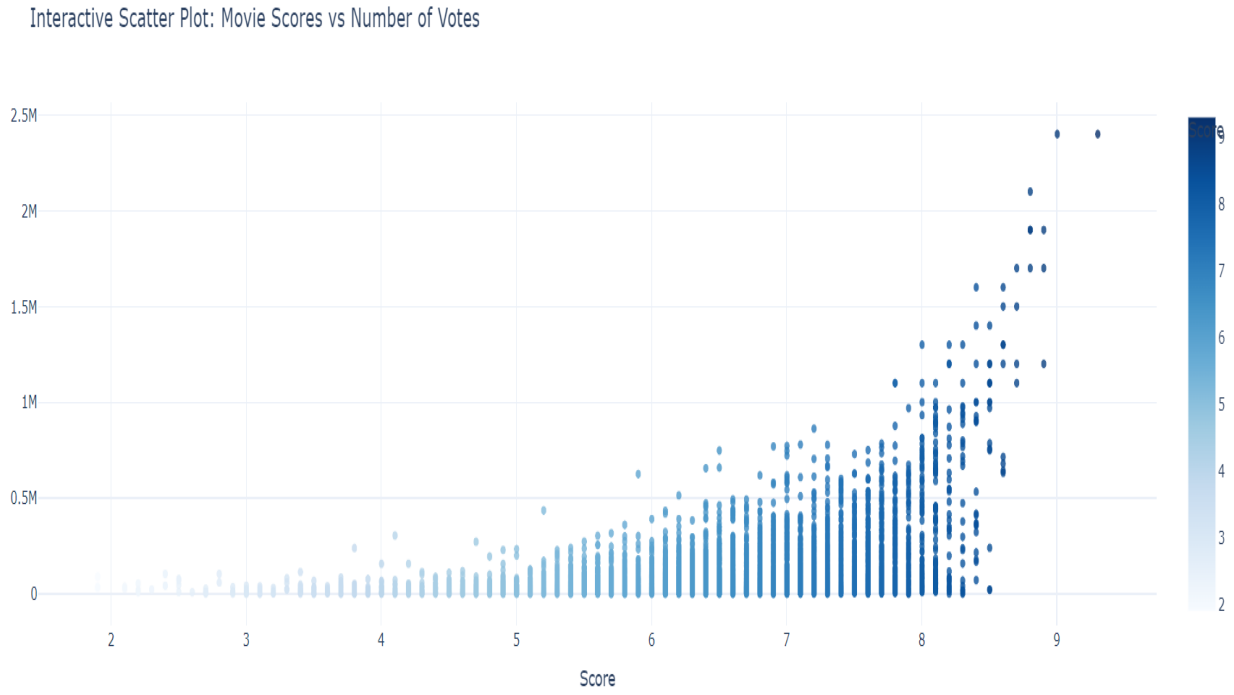
The choice of an area chart for visualization provides a clear representation of the overlapping trends in average film budgets and gross revenues over time.

**3)**




Upon analyzing the output visualization, it is evident that there is a positive correlation between the total budgets and gross revenues of film production companies. This observation suggests that certain film production companies are capable of making significant investments in their projects, resulting in a positive impact on their financial success at the box office. The companies with the highest budgets tend to achieve higher gross revenues, possibly indicating a strong capability to produce successful and financially lucrative films. The bar chart visually emphasizes the financial standing of these top film production companies, providing a clear comparison of their total budgets and gross revenues.

**4)**

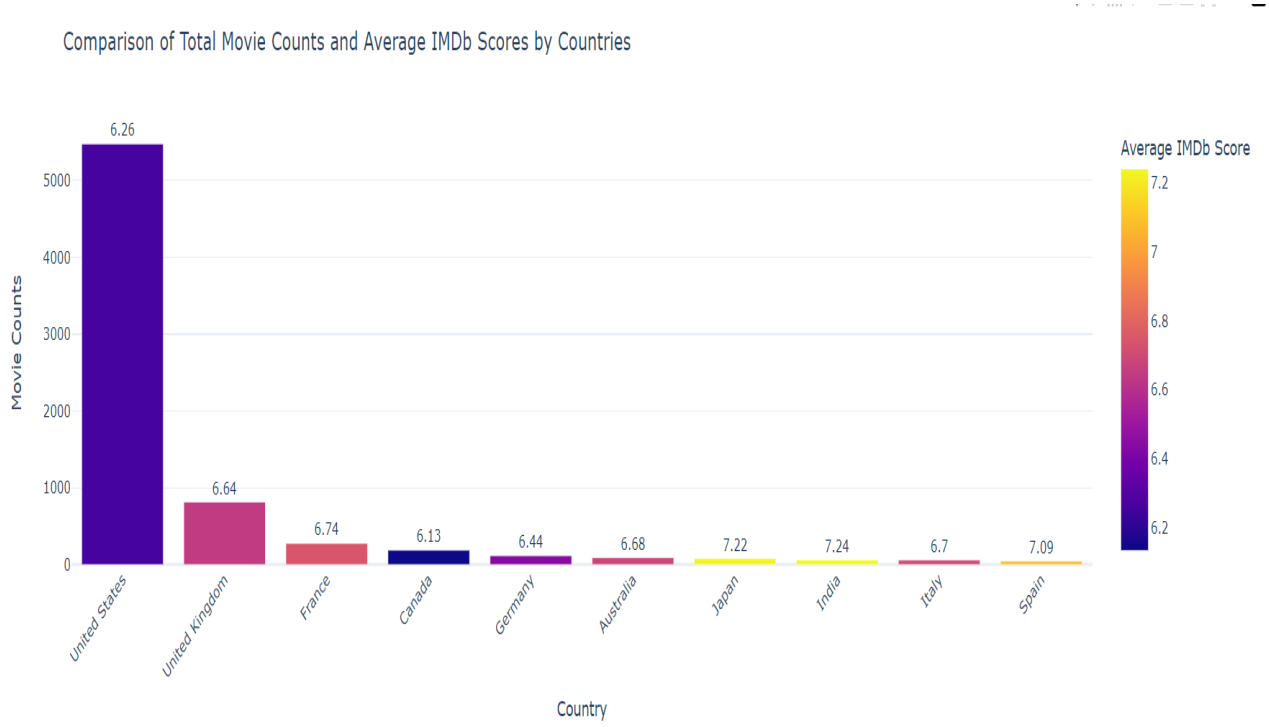



Upon examining the scatter plot, a discernible pattern emerges regarding the relationship between IMDb scores and the number of user votes for movies. Films with higher IMDb scores generally tend to attract a larger number of votes from users. Specifically, it is noticeable that beyond an IMDb score of 5, movies experience a substantial increase in the number of votes. This suggests a positive correlation between the perceived quality of movies (as reflected by IMDb scores) and the level of user engagement (as measured by the number of votes).

In summary, the scatter plot implies that movies receiving higher IMDb scores are likely to generate more interest and participation from the audience, resulting in a higher number of user votes.

A scatter plot was chosen for this analysis as it effectively visualizes the relationship between two numerical variables, IMDb scores, and the number of votes. Scatter plots are ideal for identifying patterns, trends, and correlations in the data.

**5)**




Upon analyzing the output, it is evident that the United States has the highest number of films produced, yet its average IMDb score is around 6. On the contrary, countries with fewer film productions tend to have higher average IMDb scores. Canada stands out as an exception, as it has a relatively low number of films but still maintains a lower IMDb score compared to other countries with similar film counts.

The bar chart illustrates the stark contrast between the total movie counts and the average IMDb scores for each country in the top 10. While the United States dominates in film production, the average IMDb score is not among the highest. This observation suggests that the quantity of films produced in a country may not necessarily correlate with higher IMDb scores, indicating other factors influencing film quality.