



EXPLORATORY DATA ANALYSIS OF TMBD MOVIE DATASET

BY
OBIORAH CHINENYE JULIET
Data Analyst

Table of Content

Introduction	page 1
Data Source	page 2
Data Cleaning.....	page 2
Exploratory Data Analysis	page 3
Conclusion	page 10

INTRODUCTION

This project aims at using excel to conduct an Exploratory Data Analysis (EDA) on the TMDb Movies Dataset to discover patterns and trends and extract meaningful insights. The dataset contains 24 columns and 1287 rows, it includes a range of attributes, such as movie titles, release dates, genres, cast and crew details, ratings, and financial information like budgets, profits, and revenues. By analyzing these elements, we can gain valuable insights into what influences a movie's success, genre popularity over time, and the impact of directors on ratings. This analysis will provide a deeper understanding of the dynamics of the film industry.

The dataset also includes adjusted budget and revenue values in terms of 2010 dollars, accounting for inflation over time. These columns are called 'budget_adj' and 'revenue_adj' respectively.

DATA SOURCE

The data used in this analysis was sourced from Kaggle, a popular platform for data science and machine learning.

DATA CLEANING

- Identified and filled missing data:
 - Replaced a blank entry in the "popularity level" field, replaced with "NA".
 - Replaced incorrectly entered movie names with "NA" when the correct title could not be determined.
- Researched and fixed misspelled director names, ensuring consistency for repeated directors.
- Validated and adjusted financial data by Cross-checking revenue and budget data, correcting inaccuracies through external research.
- Conducted comprehensive checks for other inconsistencies across data fields to ensure accuracy.
- Processed multi-genre entries:
 - Identified movies with multiple genres listed together, separated by a pipe delimiter (e.g., "Adventure|Action|Drama").
 - Used the **"Text.Split"** function in Power Query to separate genres into a list, then expanded these lists, example Text.Split([genres], "|")
 - Ensured that each movie was counted for each genre it represents, enabling more accurate genre analysis.
- Excluded columns such as "s/n," "id," "cast," "homepage," "tagline," "keywords," "overview," and "production companies" that were not essential for the analysis.

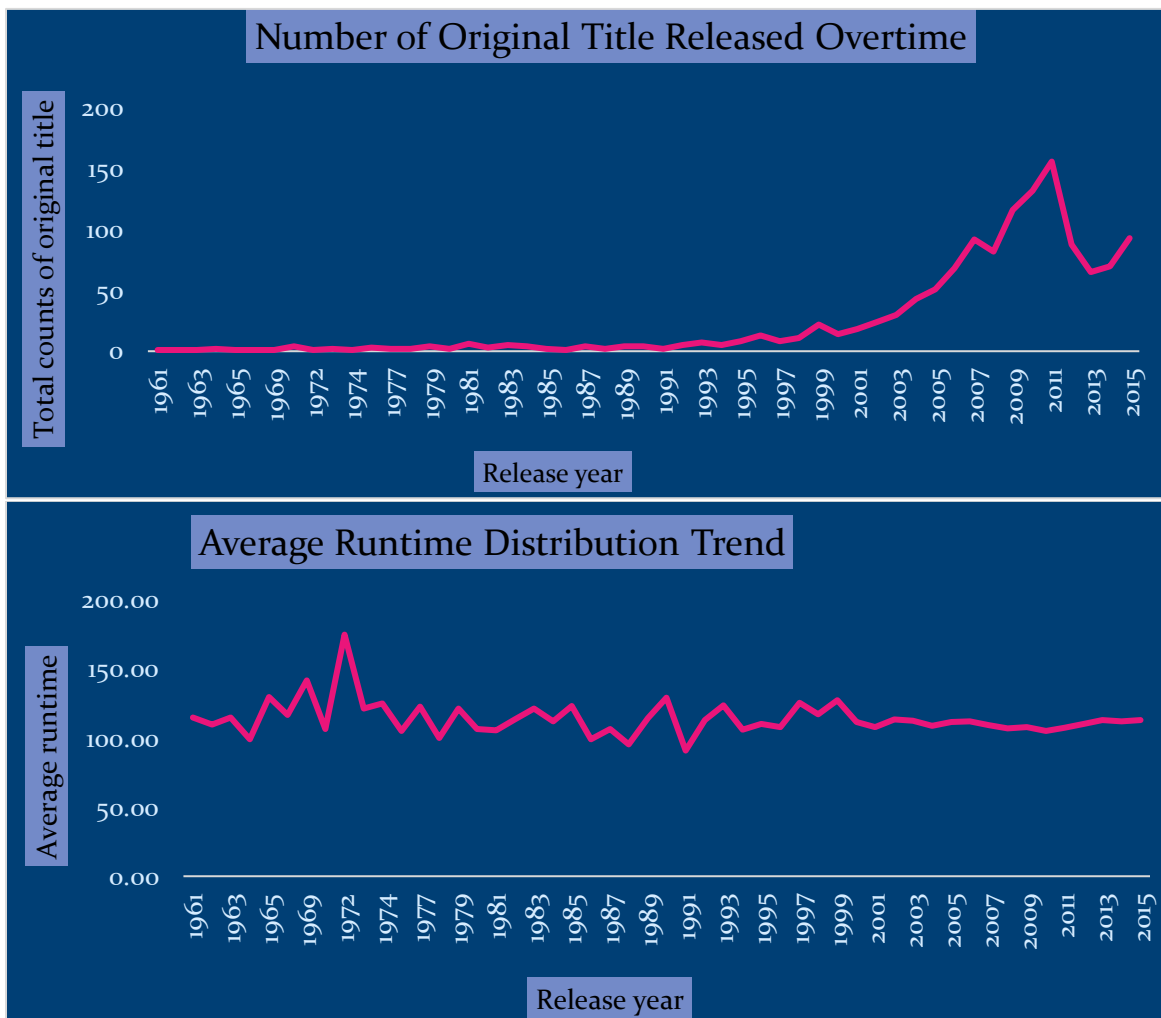
EXPLORATORY DATA ANALYSIS

Basic summary statistics for numerical columns (mean, median, min, max)

	vote_average	Popularity	Revenue	Budget	Revenue_adj	Budget_adj	Profit
mean	6.28	1.79	\$ 176,608,099.56	\$ 52,099,834.02	\$ 199,677,531.86	\$ 54,752,614.95	\$ 124,400,603.36
median	6.3	1.152354	\$ 82,347,656.00	\$ 32,000,000.00	\$ 88,260,560.37	\$ 35,574,015.19	\$ 45,412,898.00
max	8.3	32.98576	\$ 2,781,505,847.00	\$ 425,000,000.00	\$ 2,827,123,750.41	\$ 425,000,000.00	\$ 2,544,505,847.00
min	2.2	0.010335	\$ 7,306.00	\$ 7,000.00	\$ 7,425.82	\$ 8,081.12	\$ (413,912,431.00)

RESEARCH QUESTIONS AND FINDINGS:

Research Question 1: What are trends in the volume of releases across different years? Compare with distribution film's duration overtime.

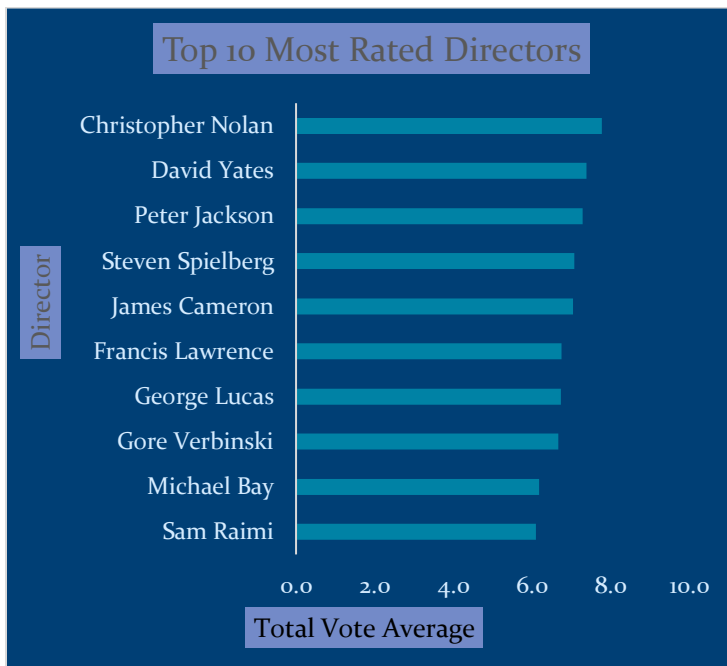


The chart above shows an upward trend, especially from the early 2000s, suggests a significant increase in movie production over time, peaking around 2010-2012 before a slight decline.

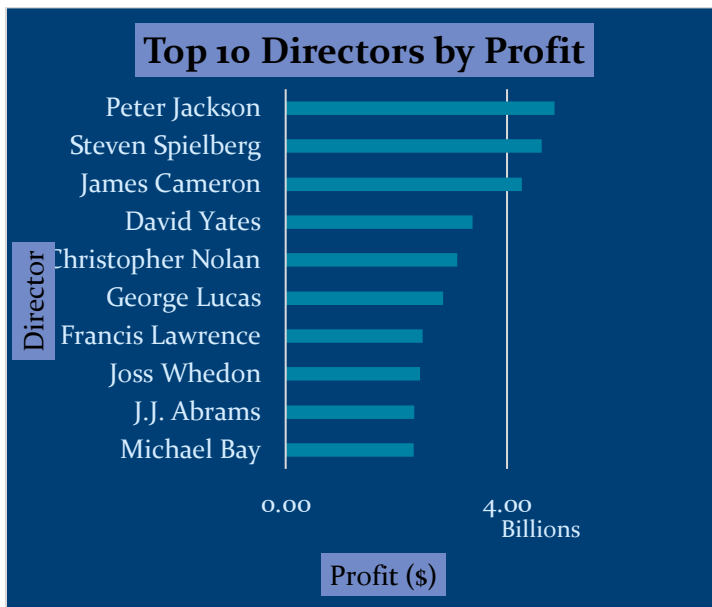
Comparing with movie duration over time:

- **Overall Trend:** Both charts show a similar upward trend in the number of movies released over time, with a notable increase starting from the 2000s.
- **Peak Periods:** Both charts display a peak in movie releases around 2010-2012, followed by a slight decline.

Research Question 2: Who are the top 10 most rated directors? Who are the top 10 directors whose movies generated the revenue and profit? Compare with least 10 generated revenues and profit.



The chart shows that Christopher Nolan, David Yates, Peter Jackson, and Steven Spielberg are among the most highly rated directors, based on total vote averages. These high ratings suggest strong viewer engagement and critical acclaim for their films.



Here are some analysis based on directors rating and the box office performance:

1. Consistency Between Ratings and Revenue:

Directors Peter Jackson, Steven Spielberg, James Cameron, and David Yates appear prominently in both the "Most Rated Directors" chart and the "Top 5" Directors by Box Office Revenue." This indicates not only positive rating but also commercial success, as these directors have generated substantial box office profits.

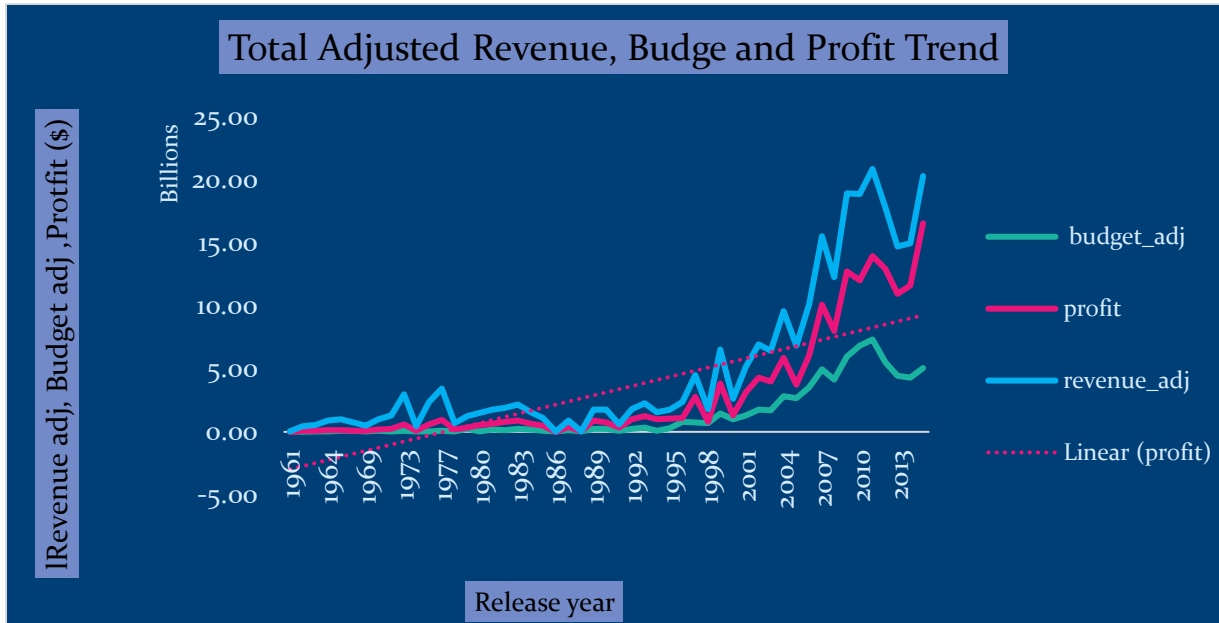
2. Leading Directors by Profit:

- Peter Jackson leads the profit chart with a total of **\$4.8 billion** in box office profit, showcasing his films' strong financial impact alongside high ratings.
- Steven Spielberg follows closely, with **\$4.6 billion** in profit, reinforcing his dual standing as a highly-rated and financially successful director.

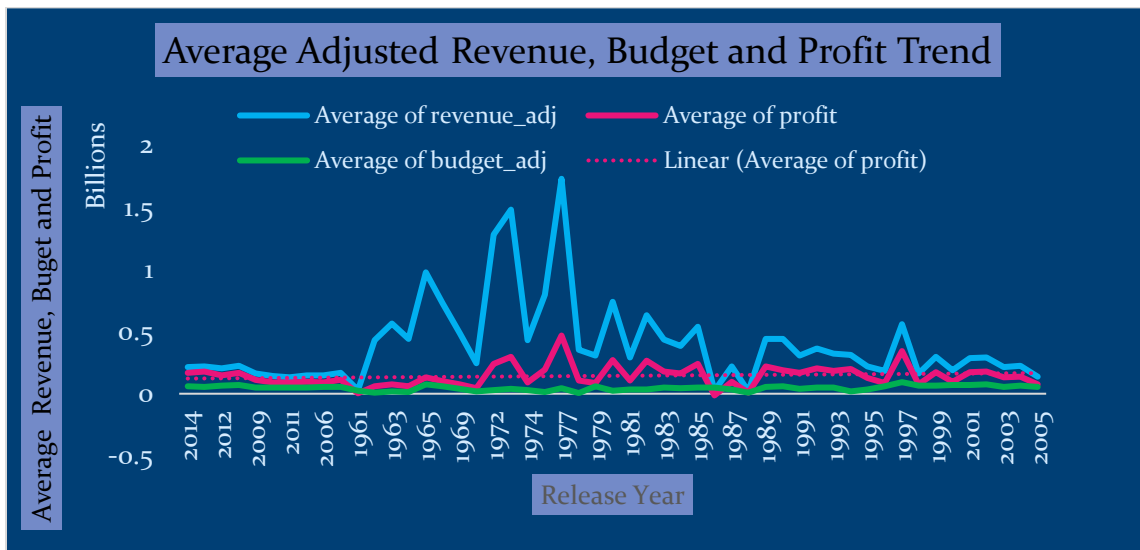
Other Top Earners:

- James Cameron and David Yates also rank highly in profit, solidifying their positions as directors with both widespread audience approval and financial success.
- Michael Bay, though ranked lower in the top 10 ratings, secures the 10th spot on the profit chart with a substantial **\$2.3 billion** in box office profit.

Research Question 3. How do budget and revenue influence the profitability of movies, and what trends can be identified in box office profits across different budget ranges?



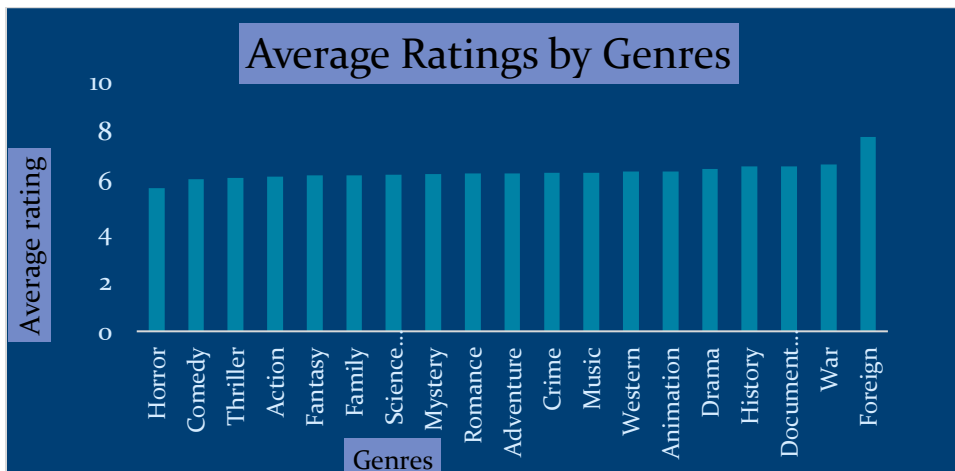
- The revenue adjustment line (blue line) shows a general increase, with significant growth starting around the 1980s, and a sharp rise around the early 2000s.
- The green line shows the adjusted budget over the years. It generally increases with some fluctuations, especially after the 1990s.
- The orange line tracks the profit, which is the difference between the adjusted revenue and the adjusted budget. Initially close to zero or negative, profit begins to rise significantly around the late 1990s, peaking in the 2000s. The dotted line indicates a linear trend in profit, showing an upward trend over time.



Comparing with the “total adjusted revenue, budget and profit chart” with the “average adjusted revenue, budget and profit chart” here are some observations made:

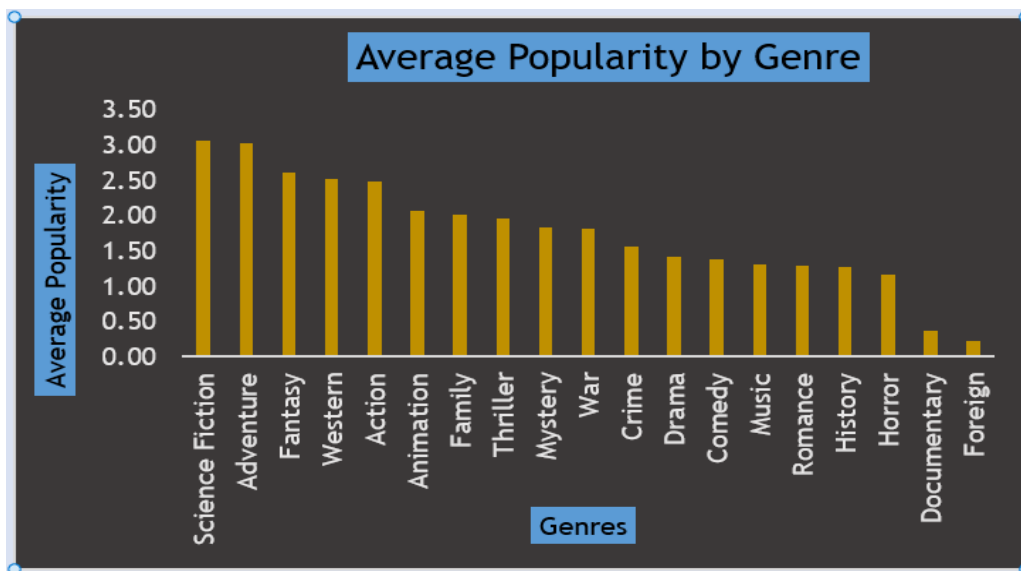
- In the average adjusted revenue, budget and profit chart, revenue peaks sharply in the 1970s and early 1980s, but then levels off. Profit remains close to zero in most years, indicating that, on average, profit levels are much lower than in the first chart.
- In the average adjusted revenue, budget and profit chart, peak periods are concentrated in the 1970s, with high average revenue and profit for that decade. Afterward, both values flatten out, showing fewer peaks and a more stable, low-level trend in the 1980s and beyond.

Research Question 4: Which movie genres tend to receive the highest ratings?



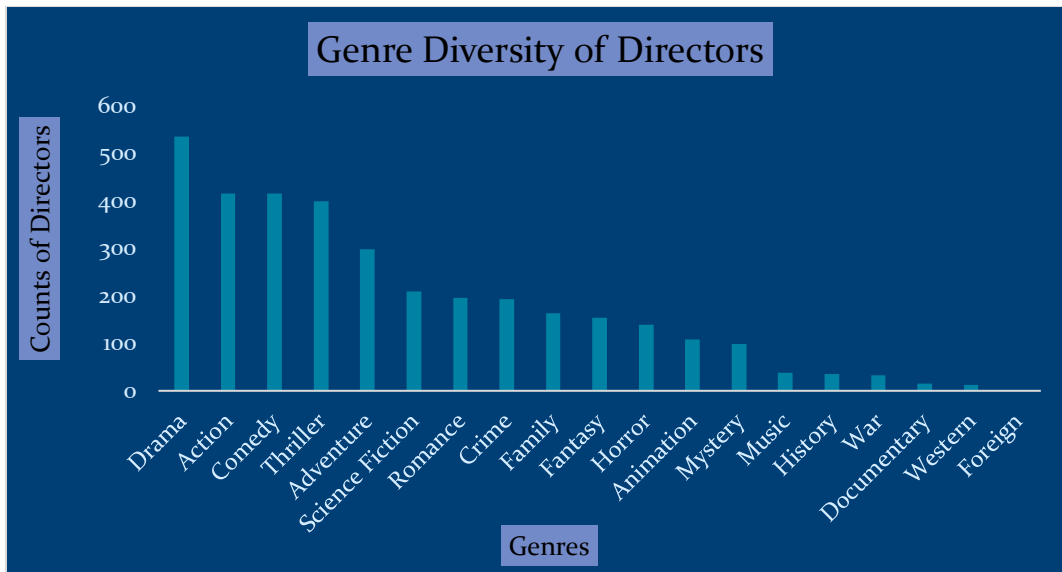
- The average ratings across genres are relatively close, with only slight differences between most genres. This suggests that, on average, no genre is rated significantly lower than others, indicating a balanced distribution in perceived quality.
- Most genres have an average rating between 5 and 7. This range includes genres such as Horror, Comedy, Thriller, Action, Fantasy, and more.
- The "Foreign" genre has the highest average rating, close to 9 while genres like Horror, Comedy, and Thriller have the lowest average ratings, though they still hover around 5 to 6.

Research Question 5: Identify the most popular genre and compare it with the ratings of the genre.



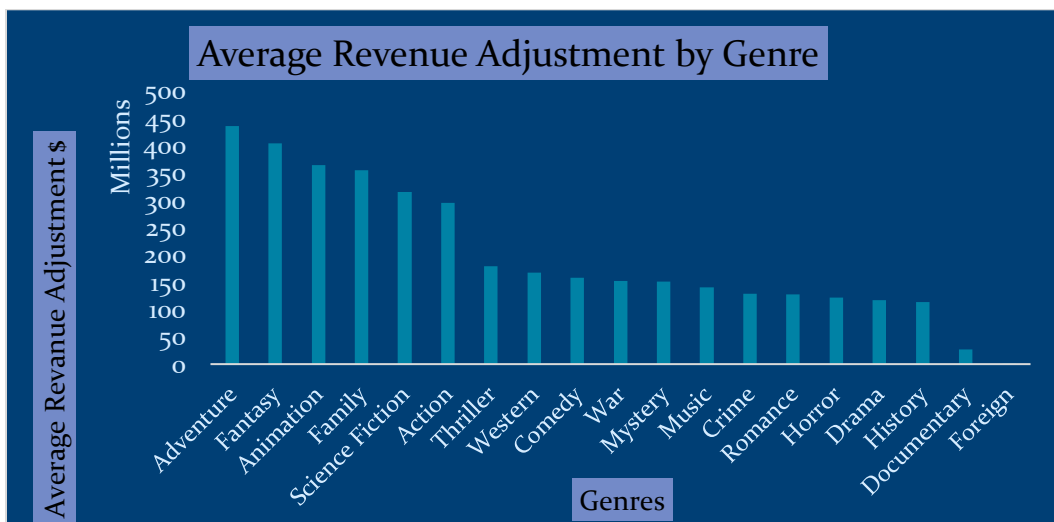
- Science Fiction and Adventure have the highest average popularity, with values around 3.5.
- "Fantasy", "Western", "Action", "Animation", and "Family" have moderately high popularity while "Documentary", "Foreign" and "Horror" rank the lowest in popularity.
- Comparing the rating and popularity:
 - Genres such as "Comedy", "Horror" and "Thriller" tend to have lower ratings and popularity.
 - "Documentary" and "Foreign" films, while rated highly, have lower popularity.

Research Question 6: Illustrate the diversity of directors across various movie genres.



- “Drama” has the highest count of directors, with 535 counts.
- “Action”, “Comedy”, and “Thriller” also have high director diversity, each with around 415 to 400 directors.
- “Adventure”, “Science Fiction”, “Romance”, and “Crime” have moderate director diversity, with counts ranging between 193 and 297 directors.
- Genres such as “Documentary”, “Western”, “War” and “Foreign” have the lowest director diversity, with fewer than 50 directors

Research Question 7: Analyze the revenue adjustment generated from each genre of the movies.



- “Adventure”, “Fantasy”, and “Animation” have the highest average revenue adjustments, each around or exceeding \$350 million.
- “Family” and “Science Fiction” also show high adjusted revenue, while Genres such as “Action”, “Thriller”, “Western”, and “Comedy” fall in the mid-range for adjusted revenue.
- “Documentary” and “Foreign” films have the lowest adjusted revenue, with values well below \$100 million.

CONCLUSIONS

- Directors with high audience ratings tend to also perform well financially, especially in the case of blockbuster directors like Peter Jackson and Steven Spielberg.
- This consistency suggests that films by these directors not only attract large audiences but also resonate well, leading to both high ratings and significant revenue generation.
- The gap between revenue and budget widens over time, indicating increasing profitability, particularly from the early 2000s onward. This might suggest improvements in revenue generation or cost efficiencies over the years.
- The upward slope of the trend line suggests an overall increase in profit over the years. This implies that, on average, profits have been growing steadily across the timeline.
- There appears to be a correlation between the adjusted revenue and profit, as both lines generally move in the same direction. This suggests that higher revenues often lead to higher profits.
- Some of the highest-rated genres appear as the least popular which suggests that these genres may be critically acclaimed but have limited appeal to a general audience.
- This genre popularity analysis suggests that Drama is a popular genre among directors, possibly because it allows for a wide range of storytelling and artistic expression.
- “Action” and “Adventure” have both high popularity (as measured by director count) and a high count of directors. This indicates that these genres are both commercially appealing and offer directors ample creative opportunities.

