

# What Drives A Movie's Success?

Laith Waqas Mohammed Ahmed

School Of Computing, National College Of Ireland, Ireland

X23103493@student.ncirl.ie

---

## I. INTRODUCTION

Over the last four decades, the industry has experienced significant transformations, driven by advancements in technology, changes in audience preferences, and the drastic rise of streaming platforms like Netflix and Amazon Prime. These transformations have shaped how movies are produced, spread, and watched around the world. This project aims to analyze such transformations by digging deep into a dataset covering decades of cinema and helping the viewer understand the transformations by presenting various distinct visualizations.

The dataset contains over 6820 pieces of cinema released between 1986 and 2016 with detailed information such as budget, gross revenue, production company, country of origin, genre, IMDb rating, runtime, and many more such information will allow us to determine what factors influence a movie's success; such determination can assist industry professionals, researchers, and movie lovers understand the industry of cinema and might possibly help make a new path for the cinema industry helping us figure out which form of movies has the highest chance of success.

### A. Data Source Background

The dataset was published on Kaggle by Daniel Grijalva [1] and the method of extracting the data was through web scraping from the IMDb website, as mentioned previously, the data provides detailed information on almost 7000 movies released between 1986 and 2016 with several attributes such as Genre, categorizing the movies into several genres with action being the major one, Gross Revenue and Budget, allowing us to make analysis comparing both, and IMDb score rating which can be used to more or less determine the quality of the movie, the variety in the dataset's attributes allows us to perform in-depth analysis of the industry and to make valuable regarding about this competitive field.

### B. Data Characteristics

The dataset is mainly quantitative, with numerical attributes like budget, gross revenue, IMDb rating, and runtime. Such quantitative attributes facilitate the making of useful statistical analysis and make patterns regarding a movie's success; the dataset also has qualitative data that is as impressive as the quantitative, which information such as genre, production company, country of origin, director, star, and more.

This dataset holds around 15 attributes and a whopping amount of 6820 instances making this dataset a ginormous one that is very relevant and intriguing when it comes to the topic of movie success

## II. OBJECTIVES

### C. Target Audience

The main target audiences for this project split into two groups, the first being the group that seeks this project for professionalism, such as industry professionals like film producers, directors, and studio executives who can utilize this project to understand market trends, audience preferences, and the main factors affecting a movie's success, another group in the professional side of the spectrum are researchers and data analysis who have an interest in the world of cinema and want to analyze the relationships between budget, gross revenue, genre popularity share throughout time, and finally, the second group, my personal favourite, is the community of casual viewers and cinephiles, more or less known in the movie community as "Filmbros," who are interested in understanding the trends in the movie industry purely for personal enjoyment and passion

### D. Information Aim

This project aims to determine the factors influencing the success of movies by examining datasets and providing visualizations. Another aim is to demonstrate how the demand and the audience's preferences influence how successful certain genres were in terms of popularity, budget, and revenue over the four decades, highlighting the rise and fall of genres, change in audience preferences, and more, and finally, the project focuses on demonstrating the relationships between attributes, seeing how attributes like IMDb score, runtime,

and star influence the movie's overall success, helping the viewer draw connections between the causes of success.

#### *E. Narrative Direction*

The concept is for the narrative to start with familiarizing the viewer with the dataset and its contents, which was initiated within the text given earlier, explaining to the viewer the transformation of the cinema industry throughout the last four decades and providing the viewer with insights on the questions that shall be answered within this journey, such as which factors influence a movie's success the most, how the audience's demand has changed over the decades and which particular attributes when combined together can make a successful blockbuster movie.

The path taken to answer such questions is through visualizations, with each visualization covering a unique aspect of the dataset, providing insights that can be used for movie making and analysis, the viewers will be deliberately guided through this path, providing explanations for each and every visualization and explaining the insight extracted from it.

### **III. DESIGN PROCESS**

#### *F. Ideation and conceptualization*

Initially, while looking for datasets and brainstorming for ideas, We sought after a topic that I would personally enjoy working on and have passion for, a topic more oriented towards the younger audience, at first, the idea was to look for something sport related, and the biggest option with the highest amount of datasets was football related datasets and ideas. However, that did not end up being as successful as We imagined due to a lack of consistency in the datasets, and then WE moved to the next topic, which was the music industry, and it was going well. However, it just did not "click", and finally, We settled with the concept of movie making due to my intriguing interest in movies and movie making and due to the huge datasets provided for that topic, thanks to IMDb.

When it came to sketching we wanted to make something that would catch the viewer's attention, something relatable with names that the viewer might recognize from the directors and stars lists, and We sought to cover insights on genres since every viewer has their own personal favourite genre and it would be interesting to know more about how that certain genre is doing when it comes to cinematic success.

#### *G. Data Processing*

Thanks to the person who published the dataset on Kaggle, we did not have to go through a lot of data processing, the first issue we faced was the fact that the dataset was in the form of CSV, which was an issue because Tableau could not run CSV so we had to convert it to XLSX, the second issue is that when it comes to budget and gross revenue, some movies did not contain that information simply due to it being unavailable to the public so we had to disregard them when it came to insights regarding budget and revenue, the third and final issue that was processed is that there was no attribute counting the number of movies produced per year or country so I had to make a new one using the "Count" function provided in the tableau base version.

#### *H. Prototyping and Iterations*

The pilot design was initially made to test which software is the most suitable to be used for this task, testing software like Excel, Tableau, and Power BI, in this case, we went with Tableau as it was the easiest to use and most practical, the first prototype made was fairly basic, it was seven visualizations 4 of which being static, and the rest were interactive, and I displayed them to my peers, and they initially liked the content of the visualizations however they did feel like it was missing something that would have made it special, after looking into it thoughtfully we realized that the issue was that the information given was interesting however the coloring and theme simply did not look appealing to the target audience, so we drastically changed the theme and color palette with something more suitable for the topic, we will get into more details in the following design choices section

### **IV. DESIGN CHOICES**

#### *I. Techniques And Layout*

Two visualizations take advantage of the classic bar chart form, the first comparing the top ten countries based on movie releases and the second comparing the top 10 directors based on average IMDb score, the reason such a visualization type was chosen was due to its ease of understanding when it comes to comparing certain elements in certain fields, making it the most practical one for both of the concepts.

A line chart was also utilized in the visualizations as it was used on a visualization comparing how many movies were released for each genre per quarter of a year from 1986 to 2016, you can easily determine that this visualization uses more than 3 variables and one of them is date, when it comes to comparisons through dates a line chart is the most suitable as it is the clearest form of visualizations when it comes to showing progress over time.

A heat map was used in this project, where the topic was top genres by average IMDb score, and the reason this form of visualization was used was due to its clarity, showing the higher values with a fierce shade of orange and the lower ones with a blue-ish grey allowing the viewer to easily distinguish the highs and lows

The three forms of visualizations mentioned above were merely for static visualizations; now, we're going to get into the interactive ones. For the first one, we used a map chart that demonstrates the number of movies each country made per year, with fierce orange showing higher amounts and blue-ish gray showing the lower amounts, the reason such a visualization form was chosen was the fact that this concept compares countries throughout years, so it's easier to comprehend when shuffling between the pages showing the movies produced during each year.

A scatterplot was utilized in this project, comparing the gross revenue to budget ratio between genres throughout the years, the reason it was used was to allow the viewer to easily distinguish the drastic difference in gross and budget between each genre.

Finally, a horizontal bar chart was used for our favorite topic that compares how successful each of the top 10 actors is in each genre, with every genre having its page, the success was measured by summing the total scores of every movie the star starred in in that genre.

### *J. Styling And Annotations*

Initially, we were aiming to use the iconic Star Wars font due to its extreme relevance to the topic, however, due to copyright issues and sharing issues being the fact that if another user downloads the visualizations he would be unable to download the font as well, the font was regretfully avoided, so we stuck with something basic and undaring, being Tableau Bold for headers and Tableau Light for everything else, when it comes to annotations, every visualization has text annotations demonstrating what every part of the visualization represents, and if the annotations were not in the visualization itself due to the visualization being too cramped, we made sure to add a legend to the left of the visualization demonstrating what each symbol means helping the viewers comprehend and understand the visualizations.

### *K. Interactivity Features*

There are three interactive visualizations displayed in this project, and each of them has its own unique concept, starting strong with the star power analysis, it is interactive as every page shows a different genre, and every page also shows who the most successful actor in that genre is, allowing the viewer to observe and interact with the visualization shuffling between the pages and understanding which stars have the highest chance of success in each genre, for the other two visualizations, the user could shuffle between the years with every year having its own page, one of the visualizations shows off the number of movies each country produced during the years and the other shows off the gross revenue to the budget ratio for each country, the user can interact with the visualizations shuffling between the years allowing the viewer to understand trends in movies throughout the years and understanding the ups and downs of the movie industry.

### *L. Color choices*

The coloring of the set of visualizations for this project was derived mainly from the colors of cinema, when you envision cinema, you envision dark colors and backgrounds bombarded with fierce colors like red and gold and that's what we went for, we made the background of visualizations dark navy in color, resembling the darkness you're in when you're watching a movie in a cinema hall and then for the content in the visualizations we used a palette of youthful fierce colors like Deep Gold (#FFD700), Crimson Red (#DC143C) and more colors of that sort like midnight blue, charcoal grey, emerald green and more, the choice of these colors was derived from a research published by Jon Hellerman on the website NoFilmSchool [2].

## V. DEVELOPMENT OF DESIGN SOLUTIONS

### M. Static Visualizations

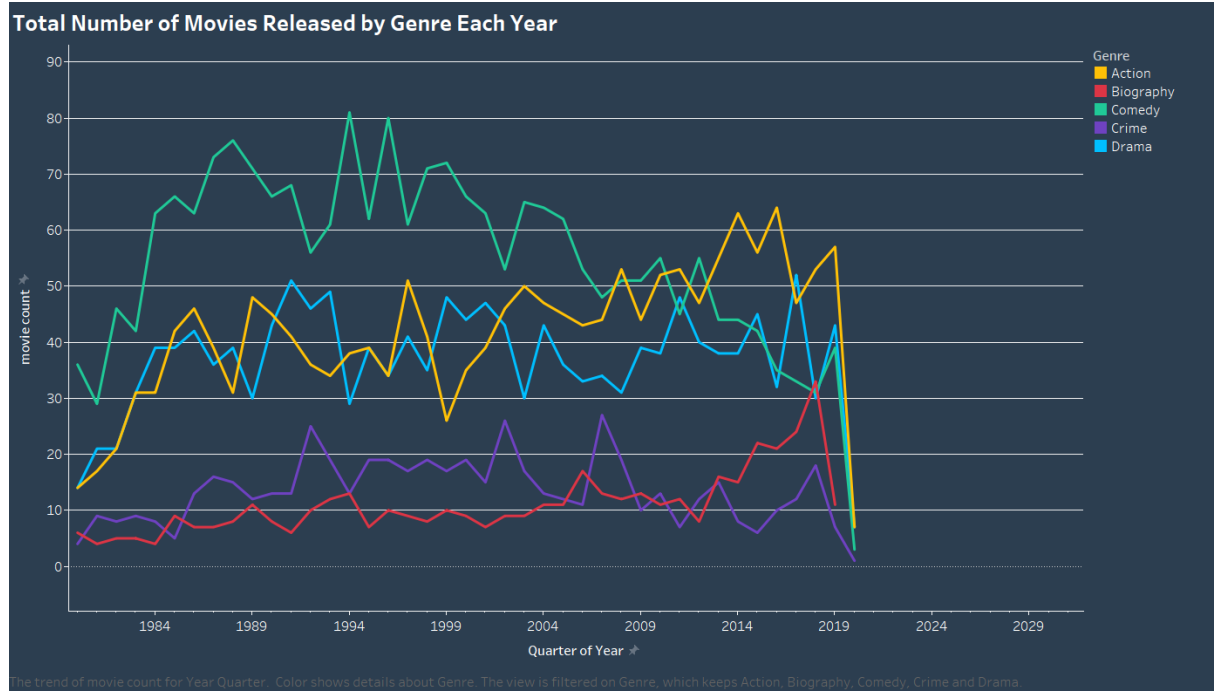


Fig. 1 A line graph demonstrating the total number of movies released by the top 5 genres each year

The visualization displayed on top displays a comparison of the number of movies released by genre every year, by observing the visualization, you can easily derive the fact that comedic movies dominated the movie industry until action movies took over. In the 2010s, with around 63 action movies releasing on the first quarter of 2014, comedic movies have been on a decline ever since and drama movies took over the 2nd place of the amount of movies produced, however as you can see in the visualizations, all of the genres faced a devastating drastic decline in 2020 due to the Covid-19 Pandemic [3] which, according to Wikipedia, delayed and even canceled a lot of the movie's production at that time.

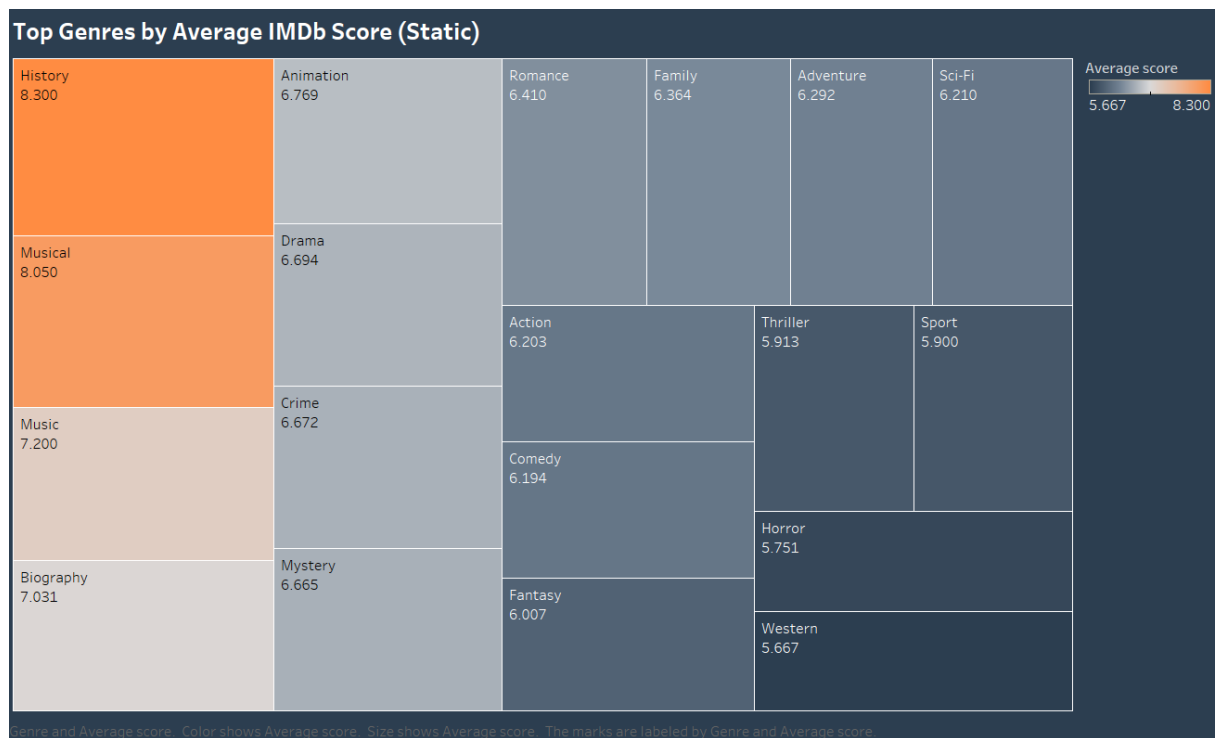


Fig. 2 A heat map showing the highest-rated genres based on the IMDb score

The visualization displayed above is a head map putting the average scores of each genre on comparison, history the most orange colored square dominates the score field with an average of 8.3, with movies like Kingdom Of Heaven and Dundirk making it the genre with the highest-quality movies even though it is not the most produced, as shown in the last visualization, following history is musical movies with movies like LaLa Land and Babylon with an average score of 8.050 and in the last place, devastatingly are western movies displayed with the navy colored square with an average of 5.6 making it the worst movie genre when it comes to quality.

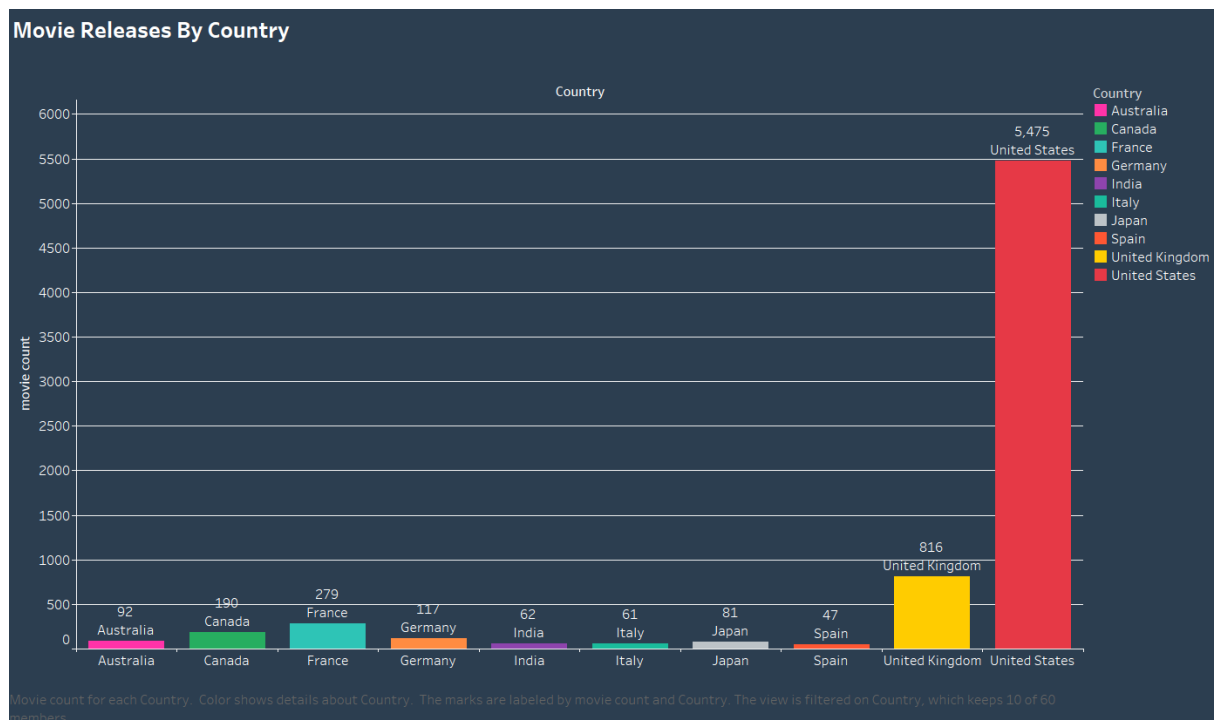


Fig. 3 A bar chart comparing the top 10 movie-producing countries

In this visualization it compares the top 10 movie-producing countries and its obvious to tell that the United States Of America clearly dominates this field by levels with almost 5500 movies screened between 1986 and 2016 , following it is the United Kingdom with 816 , which is fairly impressive however it still does not compare to the United Of America.

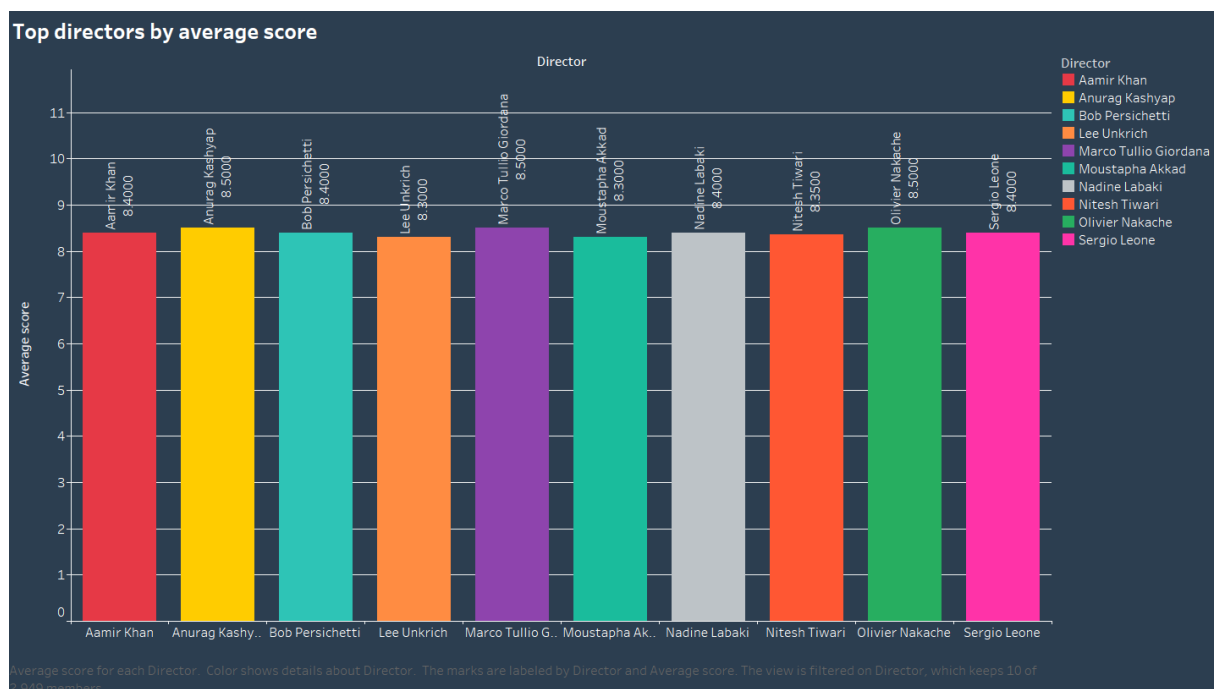


Fig. 4 A bar chart comparing the top 10 movie directors based on average IMDb score

In this visualization it compares the top 10 movie directors based on average IMDb score, in this case, three directors are tied for first place, with a score of 8.5 making it a very competitive field and also making it hard to figure out which director is exactly the best director, but also it offers us a selection of prime directors to choose from when looking for a director to direct a movie.

## N. Interactive Visualizations

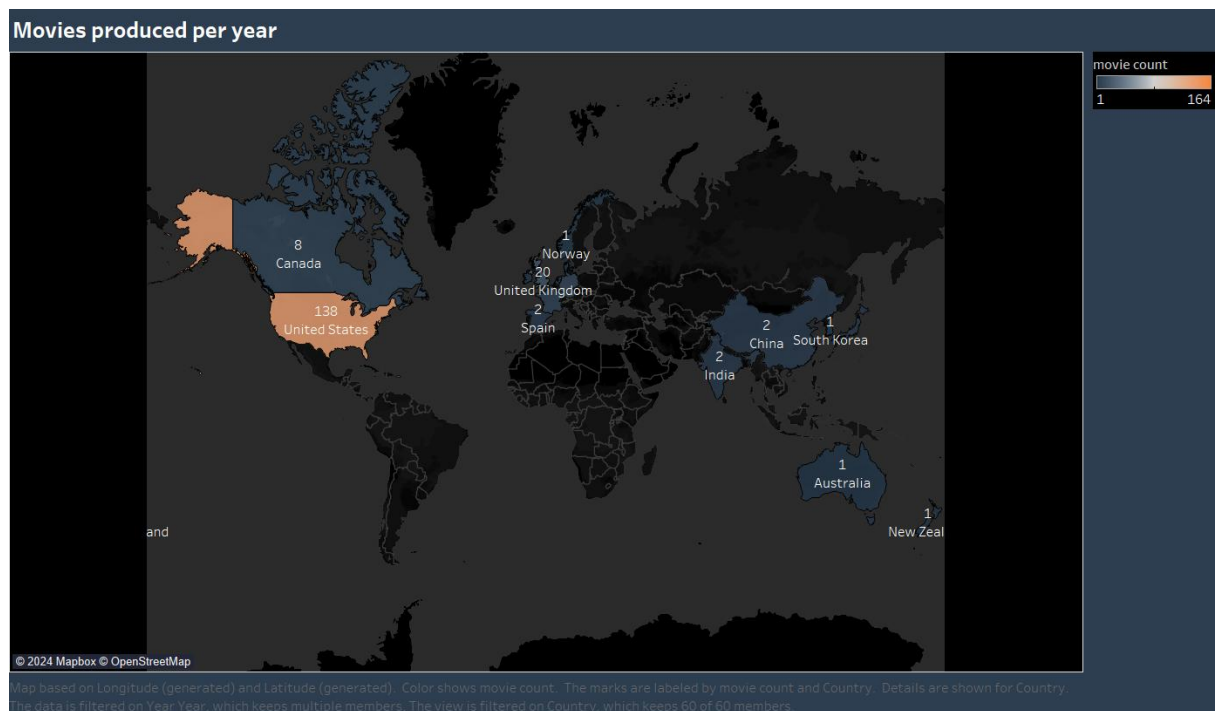


Fig. 5 A map chart showcasing the top 10 movie directors based on their average IMDb scores across various countries over the years.

\*You can find all of the interactive visualizations on this link: <https://public.tableau.com/app/profile/laith.ahmed/viz/WhatdrivesAMoviesSuccessTheDashboard/Dashboard1> \*

In this visualization, it is an interactive one the viewer gets to see how many movies were produced each year between 1986 and 2016, having the ability to shuffle between the pages. Clearly, the United States Of America produced the most movies, and the second rank shuffled between several countries, this visualization allows the user to understand which countries influenced movie production the most.

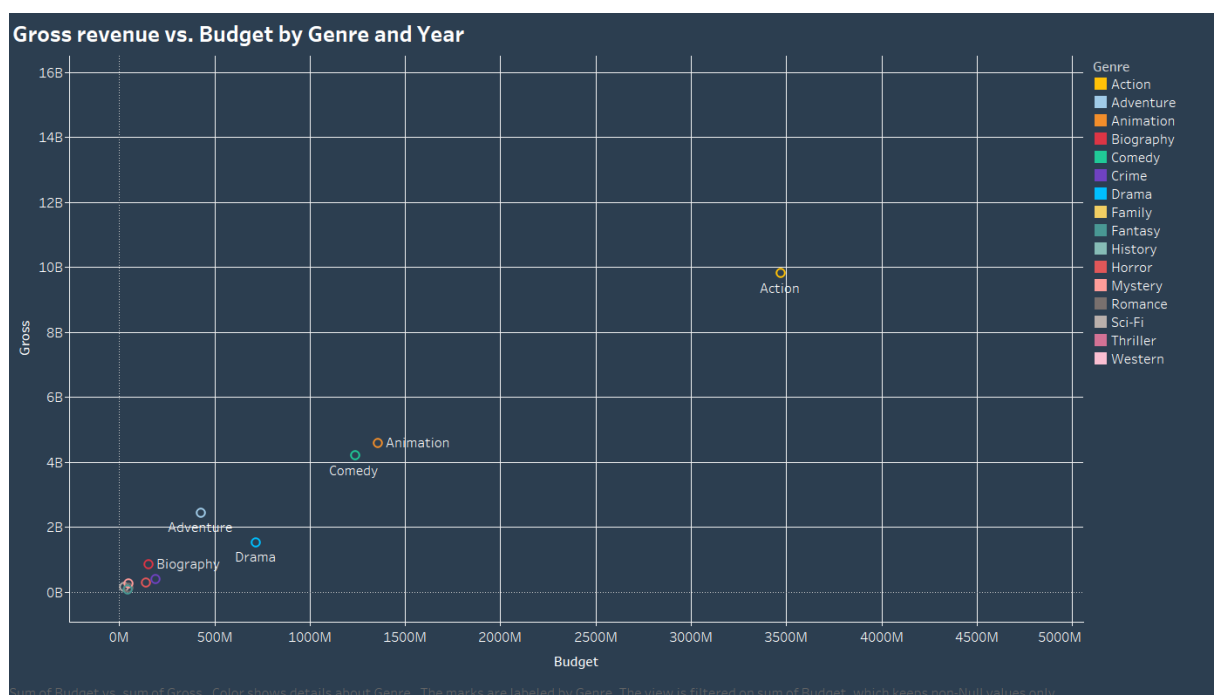


Fig. 6 A scatter plot chart visualizing the trends in gross revenue and budget for movies over the years.

\*You can find all of the interactive visualizations on this link: <https://public.tableau.com/app/profile/laith.ahmed/viz/WhatdrivesAMoviesSuccessTheDashboard/Dashboard1> \*

The visualization displayed above is a scatter plot showing the budget and gross revenue ratio across several genres, it appears the action genre dominates the chart with almost 10 billion in revenue! And that's only during the year 2011, when the user can interact with the visualization and shuffle between the years, in most years action is still the biggest genre followed by comedy, however in the last 10 years animation surpassed comedy's gross revenue crowning it with the second place.

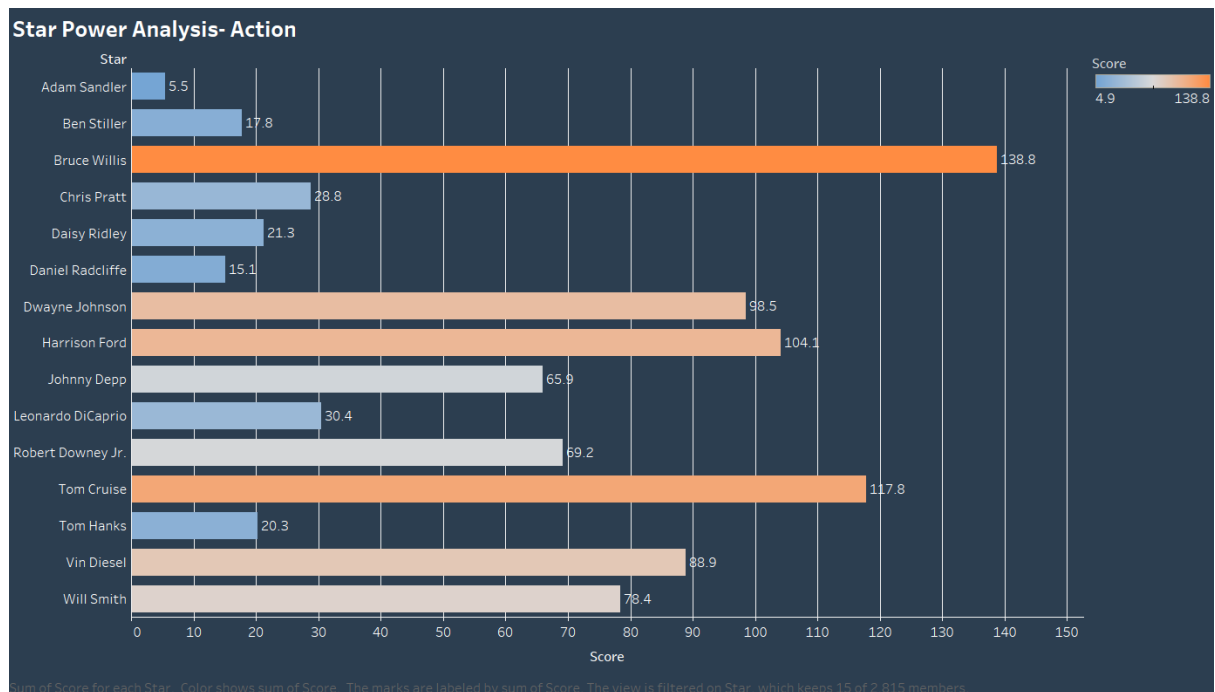


Fig. 7 A horizontal bar chart comparing Star Success across genres.

\*You can find all of the interactive visualizations on this link: <https://public.tableau.com/app/profile/laith.ahmed/viz/WhatdrivesAMoviesSuccessTheDashboard/Dashboard1> \*

This visualization is a horizontal bar chart comparing the success of the top 10 stars among different genres; the user gets to interact with the visualization by shuffling between the genres, in this case, the genre demonstrated is action, the most successful genre, and the actor with the highest sum of movie scores is Bruce Willis, making him a perfect choice for an action movie, followed by the infamous Tom Cruise.

## VI. CONCLUSION

After looking at the visualizations and recognizing patterns, you could tell that the path with the highest chance of a successful movie in the modern is for it to be an American action movie with Bruce Willis as its star combining the 3 most influential factors in making a successful movie. However, that was not always the case because comedy movies were in the highest demand during the 80s. This project was enjoyable to make, I did not face many issues making it, If I had to change something in it, I would improve my Tableau visualization-making skills to provide more appealing visualizations.

- [1] D. Grijalva, "Movie Industry," Kaggle. Accessed: Dec. 12, 2024. [Online]. Available: <https://www.kaggle.com/datasets/danielgrijalvas/movies>
- [2] J. Hellerman, "The Psychology of Color in Film (with examples) | No Film School." Accessed: Dec. 13, 2024. [Online]. Available: <https://nofilmschool.com/color-psychology-in-film>



- [3] "Impact of the COVID-19 pandemic on cinema - Wikipedia." Accessed: Dec. 13, 2024. [Online]. Available: [https://en.wikipedia.org/wiki/Impact\\_of\\_the\\_COVID-19\\_pandemic\\_on\\_cinema](https://en.wikipedia.org/wiki/Impact_of_the_COVID-19_pandemic_on_cinema)