

Creating movies is no easy feat. Directors, actors, writers, and producers have to come together to develop a coherent and exciting story that will attract audiences. However, the fundamental question for film studios is not if a movie will become a critical success, but rather this: *Will this film be profitable for our company?* To help movie studios answer this question, we aim to analyze trends within the film industry. Specifically, we presented ourselves as consultants for Universal Pictures to help guide their executives and producers with developing their next film. Our goal is not to instruct Universal Pictures on how to make an upcoming movie. Instead, our final project's objective is to provide information about the contemporary cinematic environment and identify key factors that affect a movie's profitability in order to help Universal Pictures' executives and producers make informed decisions about developing future films.

To start our project, we needed to find a dataset that could provide clear data about the film industry. We eventually found the "Movies.csv" file from Kaggle's *Movie Industry: Four Decades of Movies (Movie Industry)* dataset. This dataset holds information on films released from 1980 to 2020, such as a movie's runtime, director, budget, gross revenue, etc. (Grijalva 2021). We decided to use this dataset for its wide breadth of clear, detailed information about films. Later, we chose Tableau as our business intelligence platform. We then produced the following dashboards and data visualizations for Universal Pictures' executives and producers.

As a means of organizing our data visualizations, we divided them into two dashboards. Our first dashboard, Box Office Analysis, contains visualizations that give Universal Pictures a glimpse into the contemporary cinematic landscape (Figure 1). For example, we created a visualization, "Profit Change Across Years," to compare Universal Pictures' profits, based on film genre and measured in United States Dollars (USD), with competitors such as Columbia Pictures, Twentieth Century Fox, and Warner Bros (Figure 2). We calculate profit as the "gross" (i.e., gross revenue) values minus "budget" values from the "Movies.csv" file. Each line in "Profit

Change Across Years,” represents a different movie genre, and depicts changes in a particular genre’s profitability by company for the years 1980 to 2020. This visualization can change depending on the genre filters used, which can either isolate a genre or compare it with other genres. The genre filter also affects the other visualizations in Box Office Analysis, as well as the visualizations in our second dashboard. The ability to filter genres allows Universal Pictures’ executives and producers to create fresh insights into how a movie genre affects a film studio’s profits, or how the profitability of their own films made in a certain genre compare to other companies. In regards to the latter consideration, understanding how well a competitor performs is important. However, Universal Pictures cannot only compare themselves with their competitors. They must also make informed decisions about which movies to produce and which movies will be the most profitable for their company. Thus, we created a second dashboard that can help Universal Pictures make those decisions about a future film.

Key Profit Factors, our second dashboard, identifies variables that Universal Pictures’ executives and producers should consider when predicting an upcoming movies’ profitability (Figure 3). For example, one visualization, “Director vs. Avg. Profit,” shows a list of directors in descending order of average profit, measured in USD (Figure 4). This list includes a search bar, which can help Universal Pictures’ executives and producers specify a director that they are interested in collaborating with, e.g., Martin Scorsese. This search feature does more than allow users to focus on a specific director or a set of directors. Selecting a director or a set of directors affects the visualizations across *both* dashboards. For instance, the visualization, “Top Movies by Company” in the Box Office Analysis dashboard (Figure 5a), changes when the filter, “Steven Spielberg,” is applied, as it shows only films directed by Steven Spielberg, e.g., *Jurassic Park* (Figure 5b). Thus, the search bar for “Director vs Avg. Profit” allows Universal Pictures’ executives and producers to have a concrete understanding of how a potential director can impact a film’s profitability. Moreover, the Key Profit Factors dashboard can spark reflection and

considerations about how Universal Pictures' executives and producers want to assess profitability for their upcoming movies.

Our dashboards can help Universal Pictures understand the contemporary film industry environment and specific factors that can affect a movie's profitability. However, in the process of creating these dashboards, we encountered the following challenges. First, we needed to change our selected dataset. We had decided initially that our dataset for this project would be Kaggle's *Cinema Tickets: Time Series Analysis and Forecasting (Cinema Tickets)* dataset. This dataset focused on movie theaters' ticket sales. However, upon closer inspection, we learned that this dataset did not provide the titles of films, despite providing IDs for them within it (Möbius 2020). Without film titles, we realized that the *Cinema Tickets* dataset is essentially useless. Thus, we needed to find a new dataset. Fortunately, we were able to find and use Kaggle's *Movie Industry* dataset instead, which proved to be vastly more insightful for our project.

Second, as we analyzed the data within the *Movie Industry* dataset, we realized that we could not simply make arbitrary data visualizations about the film industry. We needed to have a cohesive, compelling data story that could organize our data visualizations together. Thus, we decided that we would frame our dashboards and data visualizations through the perspectives of consultants advising Universal Pictures about the contemporary movie industry and about the factors that affect the potential profitability of their next film. Having this framework was foundational to our project. Without it, our project would have been much more difficult to assemble cohesively. Ultimately, implementing our data story as Universal Pictures' consultants gave us a usable narrative that guided how we developed and grouped our data visualizations.

Third, with our data story in place, we then needed to decide what our data visualizations were going to be and what factors affecting a film's profitability we wanted to highlight with Universal Pictures. However, as we tested different data visualizations and potential factors for our project, we learned that certain combinations of variables produced

unusable data visualizations. For example, although a scatterplot of comparing a movie's profit and score, filtered by director, could be made, it ultimately provided no meaningful insights with which we could share with Universal Pictures (Figure 6). This meant that any visualizations that did not provide meaningful insights needed to be discarded from our dashboards. Similarly, we also encountered the challenge of missing data from the *Movie Industry* dataset. This resulted in null values appearing in our data visualizations. This meant that we had to ensure that our data visualizations did not incorporate null values into their calculations. Likewise, we also understood that Universal Pictures is not interested in making a critically-acclaimed movie. They are more interested in a film's profitability. Because of this, we disregarded any attempts of using score as a variable in our data visualizations. Ultimately, any variable used in our dashboards had to affect a movie's profitability. This led us to focus our data visualizations on factors such as genre and director, rather than score, runtime, or the country that released a certain movie.

Finally, our main challenge was deciding which platform would be able to create the dashboards and data visualizations necessary for our project. We began using Power BI at the onset to create our bar charts and graphs for our project. However, we later realized that Power BI is unable to create calculated columns. This proved to be a setback, as we needed to have the ability to create calculated columns for our project. In addition, we had already created data visualization drafts using Power BI. Despite these drafts, we decided that we needed to transition onto a new platform for our project. This situation led us to switching over to using Tableau instead. This challenge ultimately turned into a benefit, as Tableau could create all the calculated columns and data visualizations necessary for our project. Thus, although the shift from using Power BI to Tableau was an initial hindrance to our project, we would not have been able to complete this project without it.

The movie industry is fast-paced, fast-changing, and expensive. This means that producers and executives need tangible, trustworthy data to make decisions that can keep up

with artistic, cultural, and financial trends. With our dashboards and data visualizations, we believe that we can confidently provide our clients at Universal Pictures with an understanding of the contemporary film industry, as well as key factors to consider when assessing a future movie's profitability. We hope that our business intelligence can help Universal Pictures produce a profitable film that entertains audiences.

Figures

Figure 1: Box Office Analysis Dashboard

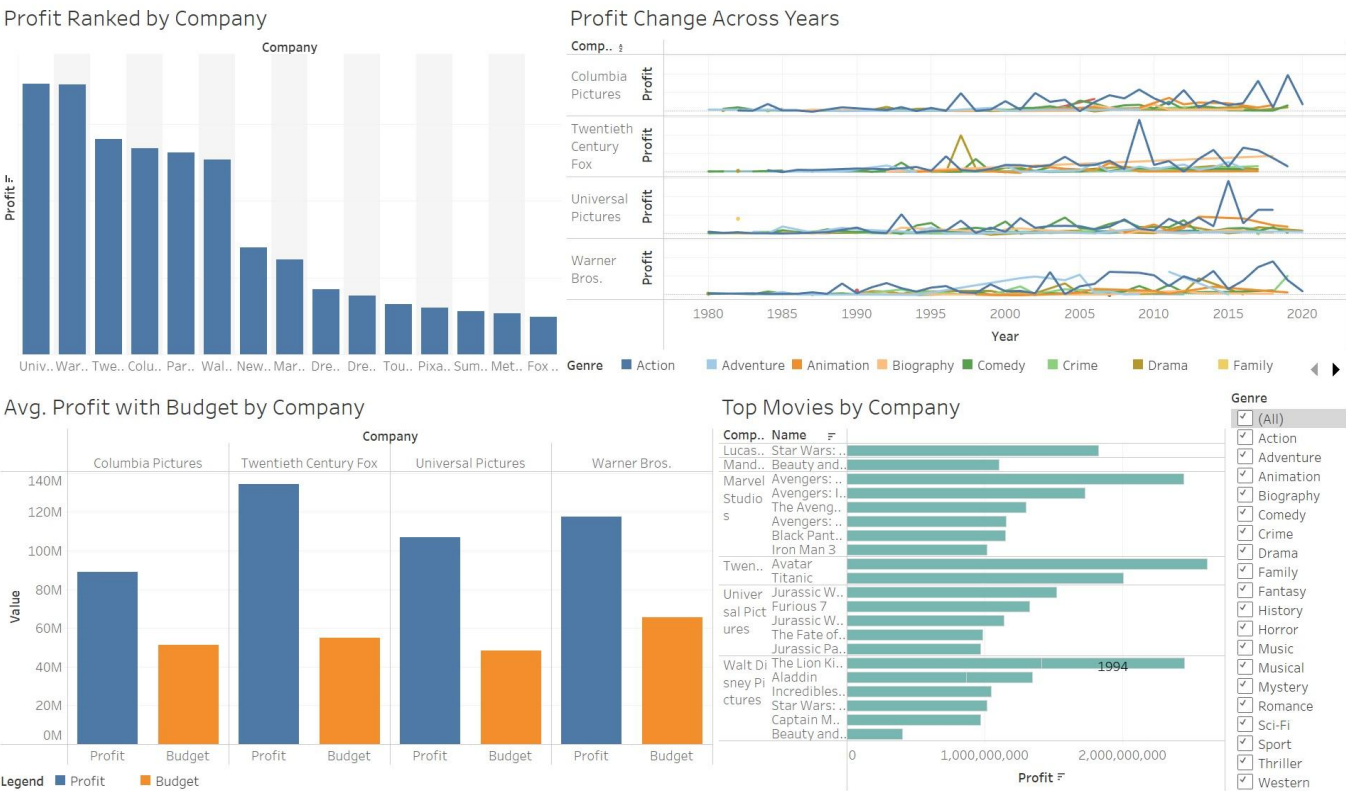


Figure 2: “Profit Change Across Years” Data Visualization

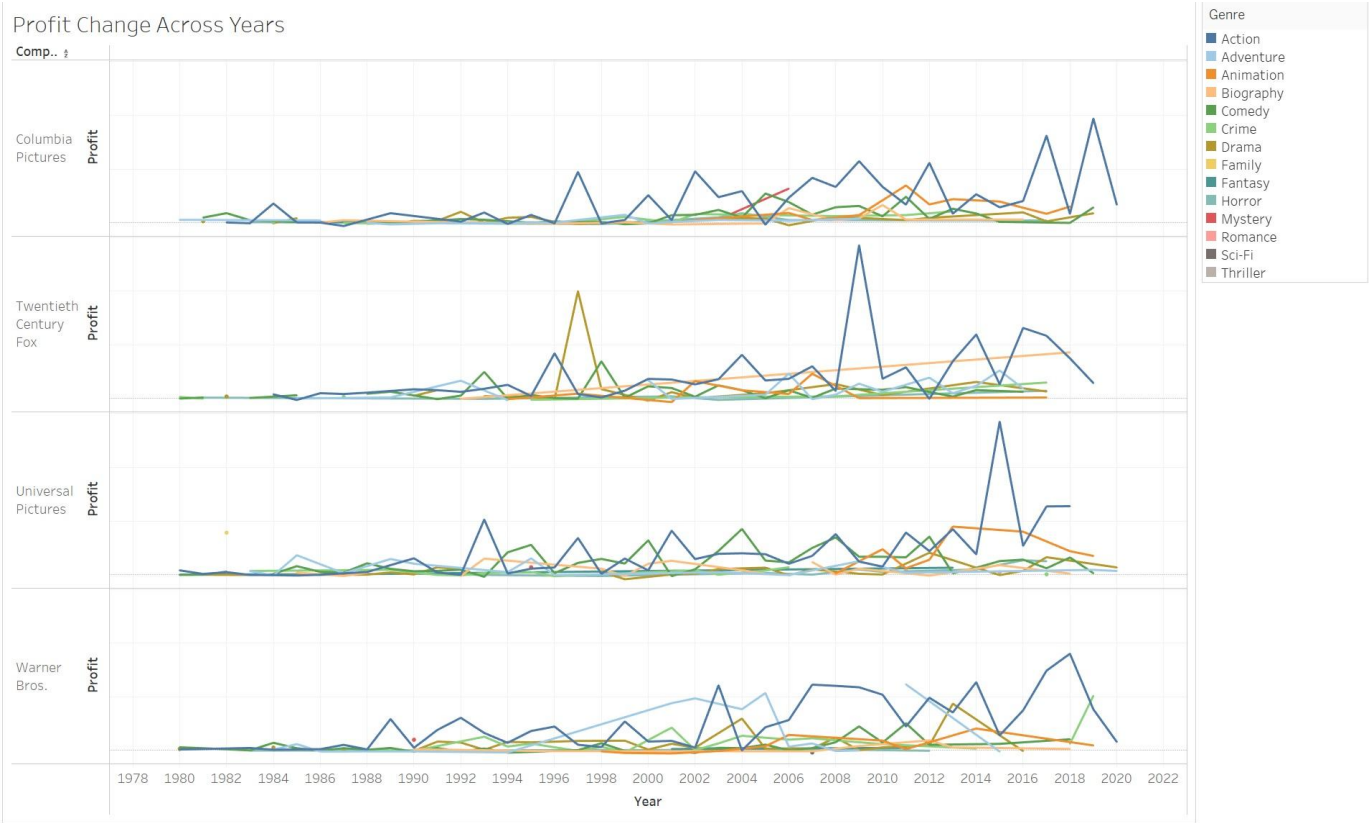
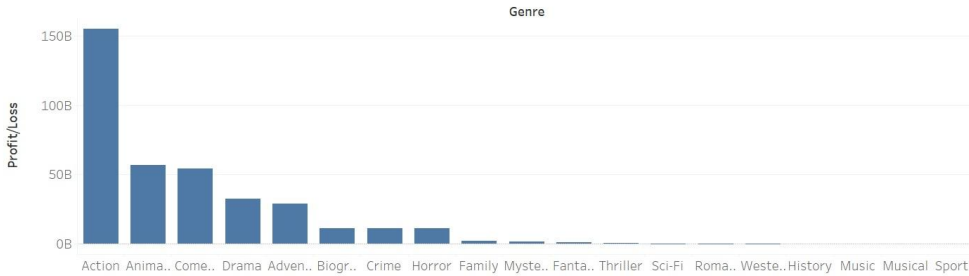


Figure 3: Key Profit Factors Dashboard

Genre vs Profit



Director vs Avg. Profit

Director	Avg. Profit
Anthony Russo	1,138,649,797
Kyle Balda	1,020,122,397
Chris Buck	916,575,617
Josh Cooley	873,394,593
Joss Whedon	817,690,061
James Cameron	785,598,368
Lee Unkrich	749,894,350
George Lucas	735,417,746
Pierre Coffin	684,439,995

Actor Analysis by Genre



Avg. Profit by Rating

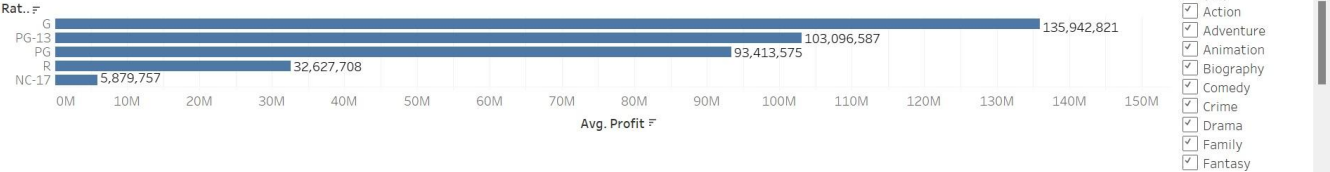


Figure 4: "Director vs Avg. Profit" Data Visualization

Director vs Avg. Profit

Director	AVG(Profit)
Anthony Russo	1,138,649,797
Kyle Balda	1,020,122,397
Chris Buck	916,575,617
Josh Cooley	873,394,593
Joss Whedon	817,690,061
James Cameron	785,598,368
Lee Unkrich	749,894,350
George Lucas	735,417,746
Pierre Coffin	684,439,995
David Yates	659,924,736
Frant Gwo	651,992,512
J.J. Abrams	599,895,911
Robert Stromberg	578,411,779
Roger Allers	577,266,283
Pete Docter	569,551,620
Dan Scanlon	543,559,645
Chris Miller	539,177,429
Sam Taylor-John..	529,651,467
Byron Howard	517,050,549
Andrew Stanton	512,593,644
Colin Trevorrow	506,262,616
Don Hall	492,869,686
Andrew Adamson	491,573,193
Peter Jackson	488,367,799
Chris Renaud	476,450,080
Carlos Saldanha	471,699,631
Steve Martino	464,738,948
David Silverman	461,414,293
Kirk DeMicco	452,205,319
Yarrow Cheney	437,492,755
David Leitch	435,180,481
Ryan Coogler	434,217,128
J.A. Bayona	432,620,940
Brad Bird	429,983,633
Jon Watts	419,830,045
Gareth Edwards	408,531,409
Christopher Miller	408,060,692
Tim Miller	400,478,042
Bradley Cooper	400,188,866
Andy Muschietti	397,439,284
Jordan Vogt-Rob..	381,652,812
Hu Guan	381,421,559



Figure 5a: “Top Movies by Company” Data Visualization without Steven Spielberg filter

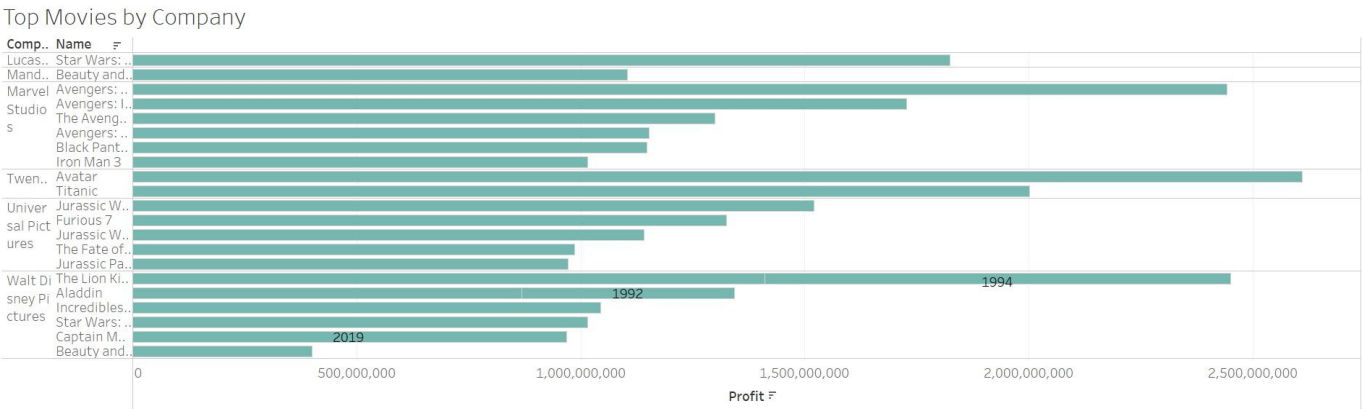


Figure 5b: “Top Movies by Company” Data Visualization with Steven Spielberg filter

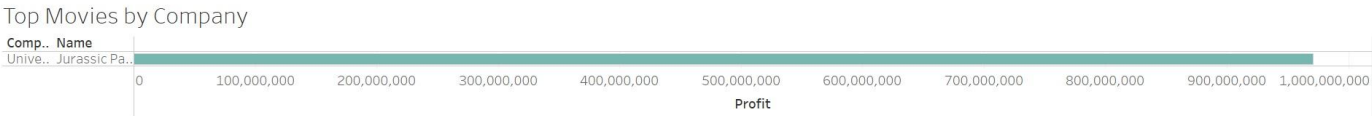
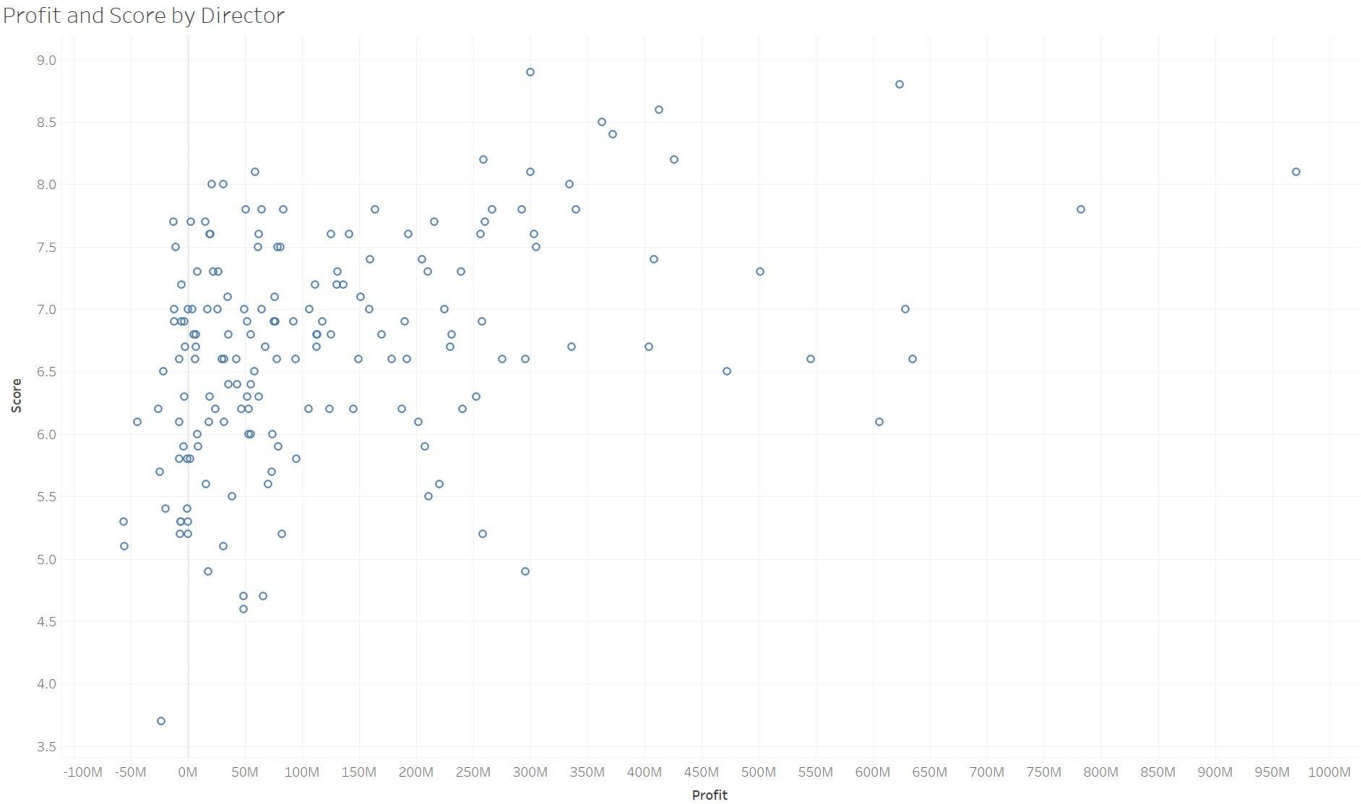


Figure 6: Unusable Scatterplot for Our Project



Works Cited

Grijalva, D. (2021). *Movie industry: Four decades of movies* (Version 2) [Data set]. Kaggle. <https://www.kaggle.com/danielgrijalvas/movies>

Möbius (2020). *Cinema tickets: Time series analysis and forecasting* (Version 2) [Data set]. Kaggle. <https://www.kaggle.com/datasets/arashnic/cinema-ticket>