Georgia Milhem

Thesis Report

Top Movies of 2022

November 28th, 2022

Abstract

Have you ever watched a movie and wondered what makes a movie successful? Well that's what my project is going to dive into. Success for movies is typically determined by the amount it earns at the box office. Three variables that I decided to investigate as influencers of box office success are, time of year that the movie is released, the MPAA movie rating, and the number of movie theaters that the movie is played in. I quickly was able to learn what contributes to a movie's success and what isn't as important. Overall, it was important to investigate the factors of a movie's success for movie producers to know helpful tips to success.

When it came time to decide on my Honors thesis report, I knew that I had to use skills from a class that has significantly enhanced my education. I chose to partner with Jerry Rackley from my marketing analytics and digital marketing classes that I took my junior year. Jerry's classes inspired my passion for data visualization, regression analysis, and Tableau. His classes taught me the importance of looking beyond the data and creating a story.

As I was brainstorming topics to choose from, I wanted to do something that is entertaining to both me and the audience. That's why I choose to do my project on the Top Movies of 2022. I enjoy going to the movie theater and watching the newest movies. Which got me wondering what makes a movie successful? In Hollywood success is rated on what makes the most money at the box office. But what variables contribute to making a successful movie? I was determined to find out.

First, I found my data set on Kaggle.com and began filtering out the data. I chose to only look at the movies that were over $50,000 in worldwide revenue and deleted anything under that. Next, I filtered my data then to only American and Canadian movies. Finally, I used only the US and Canada earnings instead of the global earnings. Then I got to work determining my x-variables. An x-variable is an independent variable, one that we believe "causes" the y or dependent variable. The y is the dependent variable meaning that it is affected by other potential variables. I decided that my x-variables were going to be the time of year that the movie was released, the MPAA movie ratings, and the number of movie theaters that it was shown at. I assumed these would be the best variables to analyze that impacted the success of a movie. I figured that the time of year that the movie was released is important. I mainly took into account that over the spring and summer more people are off due to spring and summer break, therefore, they can watch these movies. I believe that the rating is important because it determines the

frame of appropriateness for who watches it. Finally, I thought that the number of movie theaters is important because it determines how many people are able to go and watch the movie. If there are a lot of movie theaters it might help with the success of the movie.

The first variable that I ran was the time of year. I had to convert the months into numbers to run the regression analysis. It has to be noted that the time of year only includes months 1-11. The regression analysis is a way to determine if there is a relationship between the X and Y variables. The image below shows the results of the regression analysis for time of year.

SUMMARY OUTPUT
Time of Year

Regression Statistics	
Multiple R	0.041444654
R Square	0.001717659
Adjusted R S	-0.023879324
Standard Err	148739853
Observations	41

ANOVA

	df	SS	MS	F	Significance F
Regression	1	1.48458E+15	1.48458E+15	0.067103978	0.796963456
Residual	39	8.62818E+17	2.21235E+16		
Total	40	8.64303E+17			

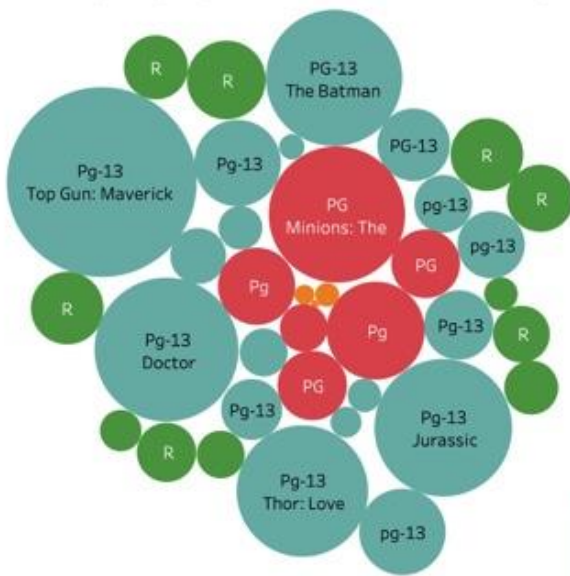|	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	107102378.8	51513018.12	2.07913228	0.044228593	2907464.814	211297293	2907464.81	211297293
X Variable 1	2303423.744	8892005.182	0.259044355	0.796963456	-15682354.4	20289201.9	-15682354	20289201.9

It is important to note that R-Squared refers to the accuracy of the model or relationship. Therefore, the closer R-Squared is to 1, the higher the likelihood that the X variable caused the Y variable. This shows us that the time of year isn't very important for the success of a movie. This is due to the fact that 0.001717659 is very small compared to the relevance of 1.

The second variable I ran was the movie ratings. I had to also convert this into numbers (PG is 1, PG-13 is 2, R is 3, and unrated is 4) to run the regression analysis. The image below represents the regression analysis for the movie ratings.
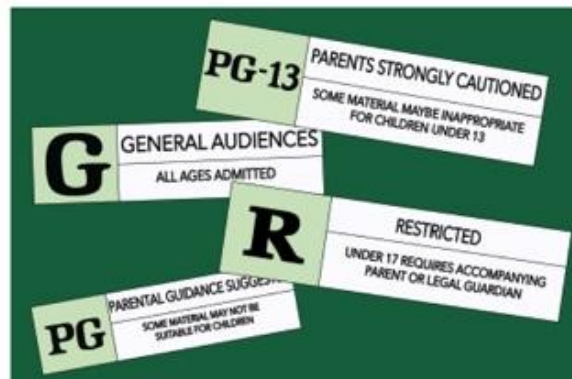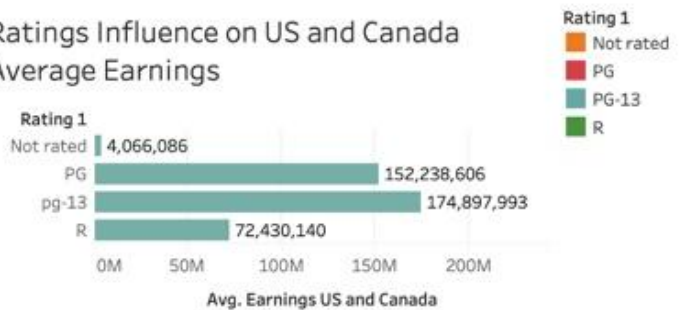
SUMMARY OUTPUT
Movie Rating

Regression Statistics	
Multiple R	0.387209842
R Square	0.149931462
Adjusted R S	0.128134833
Standard Err	137254826.5
Observations	41

ANOVA

	df	SS	MS	F	Significance F
Regression	1	1.29586E+17	1.29586E+17	6.878653608	0.012380828
Residual	39	7.34717E+17	1.88389E+16		
Total	40	8.64303E+17			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	268190151.3	60784007.66	4.41218277	7.83769E-05	145242890.9	391137412	145242891	391137412
X Variable 1	-61780533.71	23555912.68	-2.622718744	0.012380828	-109426864.4	-14134203	-109426864	-14134203

This shows us that the movie ratings also don't play a huge role in determining the earnings and success of the movie. This is shown through the R-Square of 0.149931462, which still is small compared to the importance of 1. The movie rating is however significant due to the P-value of 0.012380828 being less than 0.05, the standard threshold for significance. The visualization below shows the earnings by Top 2022 Movies and Rating and the Ratings Influence on US and Canada Average Earnings. The visualization on the left showcases that Top Gun Maverick had the highest earning with the overall rating of PG 13. The visualization on the right showcases that the PG-13 movies end up making the most of the 2022 movies top earnings.

The third and final variable I chose was the number of movie theaters. I ran this with a regression analysis and the results are shown below.

SUMMARY OUTPUT								
Number of Movie Theaters								
Regression Statistics								
Multiple R	0.475477652							
R Square	0.226078998							
Adjusted R S	0.206234869							
Standard Err	130963107							
Observations	41							
ANOVA								
	df	SS	MS	F	Significance F			
Regression	1	1.95401E+17	1.95401E+17	11.39274019	0.001679448			
Residual	39	6.68902E+17	1.71513E+16					
Total	40	8.64303E+17						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	-35504708.22	50139954.78	-0.708112091	0.483086566	-136922339.5	65912923	-136922339	65912923
X Variable 1	46191.20686	13685.01295	3.375313347	0.001679448	18510.65542	73871.7583	18510.6554	73871.7583

The results show that the number of movie theaters does play a role in determining the success and earnings of a movie. However, it doesn't play a huge role. The R-Square value is 0.226078998 which is the highest variable that I have run, yet it still is far from 1. It does have significance considering that it is below 0.05. All in all, this is the variable that comes closest in helping determine the success of a movie. The visualization below shows that sometimes the
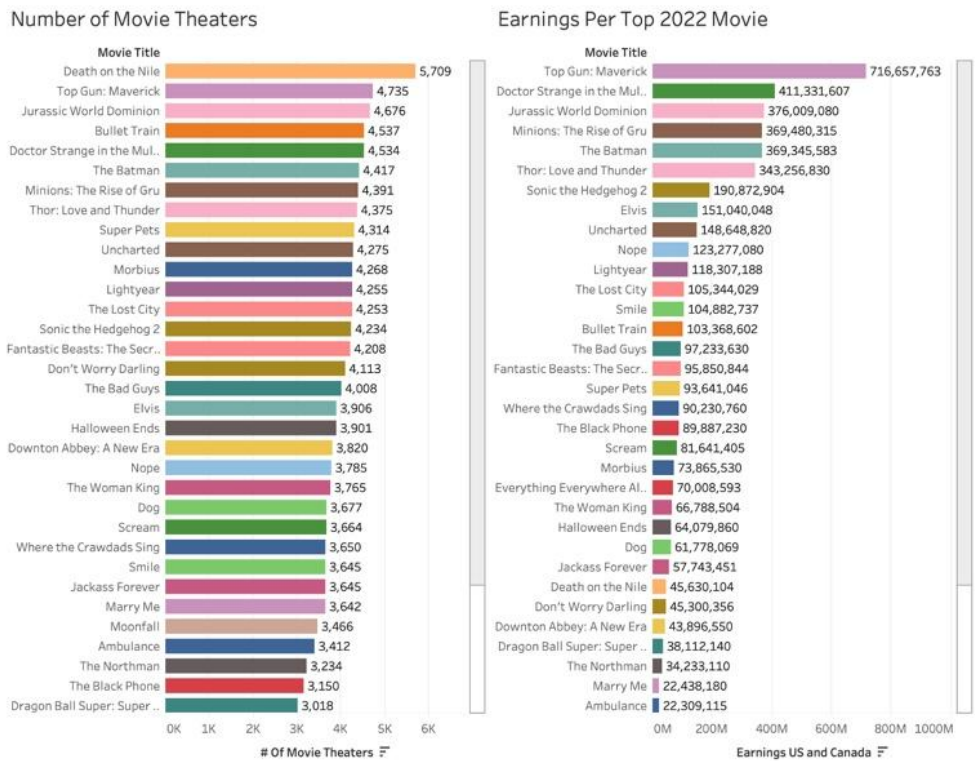
number of movie theaters doesn't determine the highest earnings. Take the Death on the Nile movie for example, it is highest in number of movie theaters, however, lower in the amount of earnings. This shows that the number of movie theaters doesn't always impact the earnings.

To determine if this group of variables is helpful for the success of a movie, I combined all of the variables together and ran a regression analysis. The results are depicted in the image below.

SUMMARY OUTPUT
All variables together

Regression Statistics	
Multiple R	0.47958021
R Square	0.22999718
Adjusted R S	0.16756452
Standard Err	134115278
Observations	41

ANOVA

	df	SS	MS	F	Significance F
Regression	3	1.9879E+17	6.6262E+16	3.68392397	0.02037527
Residual	37	6.6552E+17	1.7987E+16		
Total	40	8.643E+17			

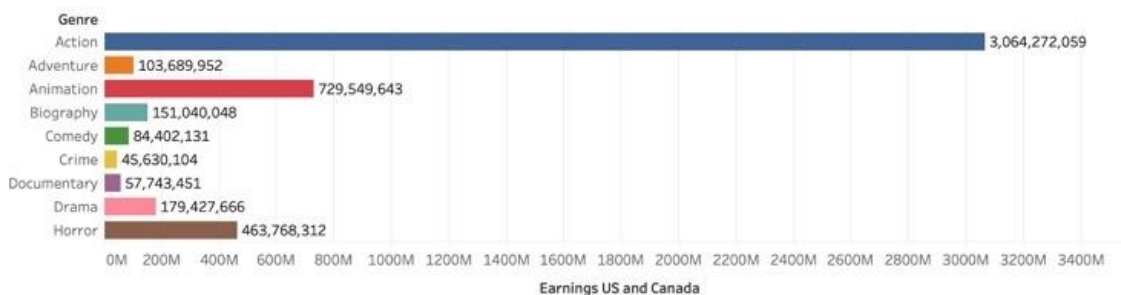|	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	14090936.3	145090664	0.09711815	0.92315675	-279890673	308072546	-279890673	308072546
month	1039512.04	8030962.38	0.12943804	0.89771215	-15232763	17311787.5	-15232763	17311787.5
rating	-14022746	33597545	-0.4173741	0.67881678	-82097839	54052346.3	-82097839	54052346.3
# of movie th	39880.3539	20479.4886	1.94733153	0.05911453	-1615.0316	81375.7394	-1615.0316	81375.7394

In conclusion, these variables together don't give us the answer that we are effectively looking for. They help a little bit, but don't help as much as we would like them to when determining what makes a movie successful. This means that this group of variables isn't what makes a movie successful. Therefore, if I had more time to do further research, I would experiment with more variables. The variables that I would choose would be genre, director, and the length of movie. I think that the genre would be interesting to look at to see if that is what causes a movie to be successful. The visualization below showcases an example of the Top 2022

Movies Genre. Action ended up being the top movie genre with 19 of the 41 movies. Along with being the highest in the influence on US and Canada earnings.

Top 2022 Movies Genre

Top 2022 Genre of Movies Influence on US and Canada Earnings

Genre	Earnings US and Canada
Action	3,064,272,059
Adventure	103,689,952
Animation	729,549,643
Biography	151,040,048
Comedy	84,402,131
Crime	45,630,104
Documentary	57,743,451
Drama	179,427,666
Horror	463,768,312

This shows that the genre would be an interesting variable to investigate with. The director would be good to experiment with because you could see if the movie is more successful based on how many award winning movies the director has filmed. Finally, the length of the movie would be interesting to see if that leads to a successful movie. Does a movie need to be shorter or longer for it to be successful? These would be the three new variables that I would choose to experiment to determine movie success.

Throughout this project, I have learned many new things that will be applicable in my future. I learned the importance of creativity and connection when it comes to picking data to interpret. You want data that connects to the audience and can tell a unique story for them. Professor Rackley's class taught me the usage of the regression analysis and what that means for

your data. I got to understand how important it is to use it to figure out the relationship between two variables. His class also taught me how to use Tableau and I was able to apply that to my project. Finally, I learned that even if you think that data is significant to the outcome, it might not be. I really thought that the number of movie theaters was going to be significantly important for the success of movies. I figured that a movie needed to be played in thousands of movie theaters to be successful. However, after I ran the regression analysis, I found that not to be true. I learned that sometimes things aren't as they seem. Which is an important lesson when it comes to business analytics. While you can assume one thing, what you find could be completely different. These lessons will help me in my future education at grad school. I am planning on attending the University of Arkansas for a Masters of Applied Business Analytics program in the summer of 2023. I am confident that the skills and tools that I have learned in the Spears Marketing Program will provide me a successful advantage in the program.

References

IMBd. (2022). IMBdPro. Retrieved December 7, 2022, from

https://pro.imdb.com/title/tt4998632/boxoffice

Kaggle. (2022). Your machine learning and Data Science Community. Retrieved December 7,

2022, from https://www.kaggle.com/

Author Biography

Georgia is a senior graduating a semester early with a major in marketing and minoring in management. She has enjoyed her time at Oklahoma State and the experiences that she has been a part of. From being a tuba player in the Cowboy Marching Band, the president of Sigma Phi Lambda, member of Student Foundation, and study abroad trips to Ireland and Zimbabwe, Africa, Georgia has expanded her learning outside of the classroom. Recently she was awarded as a Lead Rider (top marketing student of the marketing program), Homecoming Royalty Court, and Senior of Significance. She has a new passion for business analytics and data visualization thanks to her marketing classes. Which has launched her to pursue her masters in business analytics. Georgia has loved her time at OSU and can't wait to further her educational career.