By: Michael Kennedy

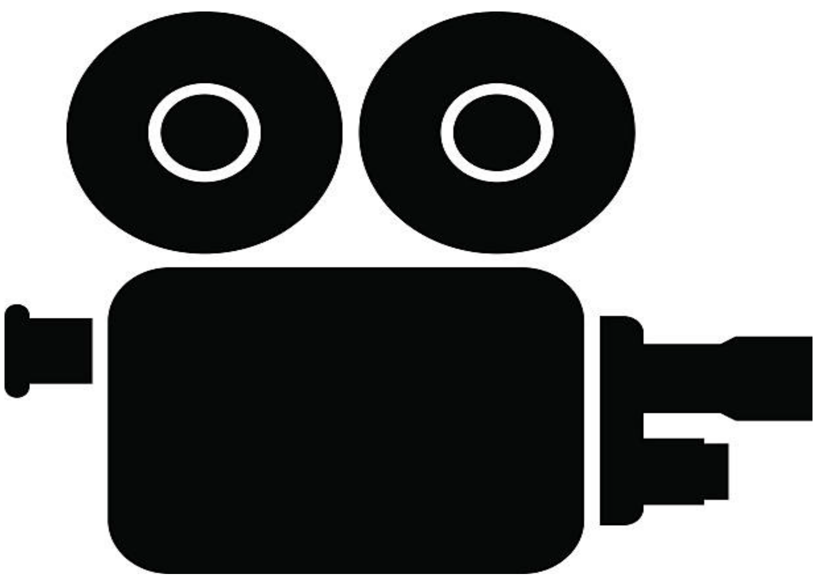
Peacock Streaming Services Movie Production Report

A Collection of Data on the Top 50 Movies of All time

Table of Contents

The New Environment

Our research team at Peacock Streaming Services is gathering data on the best 50 movies of all time to gather information on the movie production industry so our production team can vertically integrate our service to start making our own films like Netflix, HBO, and Paramount. The environment we chose specifical the environment we chose specifically was what was qualified as the best movies of all time according to accumulation of reviews, audience scores, revenue, genre, and other factors. This was to determine what types of movies audiences enjoy at the level of film production we would like to achieve. Other streaming industries have already been able to tap into creating their own content rather than relying on other studios to create content that an audience might not even enjoy. By gathering this data, we can accurately determine what an audience will like and make something similar they will enjoy. The data was gathered from the data-sharing platform Kaggle.



Variable Information

Variable Name	Description	Variable Type
Movie Title	This is the unique identifier of the data. This splits the Data into 50 different rows.	Unique Identifier
Released Before or After 2000 (Binary variable)	The binary categorical data point splits our data up into a 17/33 split (or a 34/66). Determines if a movie was released before or after the year 2000.	Categorical (Binary)
Director	Variable contains the name of every director of each movie even if one director has directed multiple movies on the list.	Categorical (35 categories)
Lead Actor	Variable contains the name of every lead in each movie even if an actor has been in multiple movies on the list.	Categorical (43 categories)
Genre	Variable lists the genre of each movie.	Categorical (10 categories)
Country of Origin	Name of the country each movie originated from	Categorical (6 categories)
Runtime (Hours)	The runtime of each movie in hours rounded to two decimal places.	Interval/Ratio (Continuous)
Votes Received (IMDB)	Numbers of Voted the movie received on the online film database IMDB	Interval/Ratio
Rating on IMDB (Out of 10)	Rating of the movie out of 10 by film critics	Interval/ Ratio
Metascore (Out of 10)	Rating of the movie as an accumulation of scores by revered critics	Interval/ Ratio
Gross Earning in Mill (USD)	Total gross earnings of the movie in Millions (USD).	Interval/Ratio
Budget in Mill (USD)	Budget of Movie in millions (USD)	Interval/ Ratio
Height of Lead Actor	The height of the lead actor performing in each movie.	Interval/Ratio (Natural World)

Variable Statistics

Variable Name	Mean	Standard Deviation
Runtime (Hours)	2.3156	0.507479405
Votes Received on IMDB	900488.16	435633.6321
IMDB Rating (Out of 10)	8.666	0.193369689
Metascore (Out of 100)	83.31111111	85
Gross Earning in Mill (USD)	144.0477273	147.2184405
Budgets in Mill (USD)	51.17877551	80.63331691
Height of Lead actor	5.8782	0.292244586

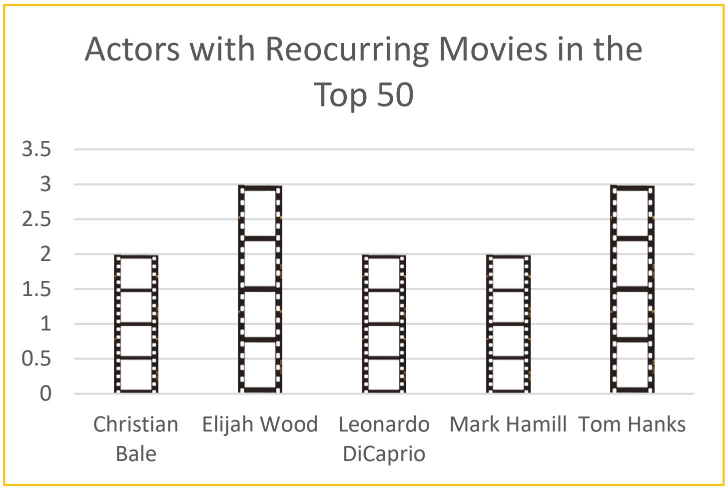
Variable Name	Mode	Range
Released Before or After the Year 2000	Before	18
Director	Christopher Nolan	4
Lead Actor	Tom Hanks	2
Genre	crime	12
Country of Origin	43	42

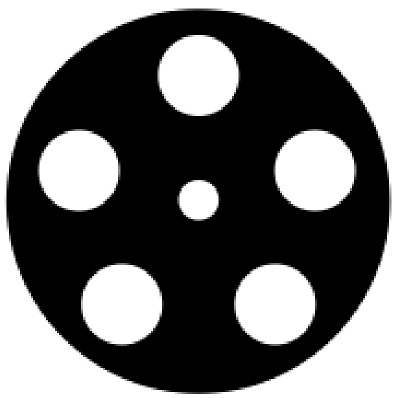
Our range of data if we were to be 90%, 95%, and 99% confident in our data for each variable

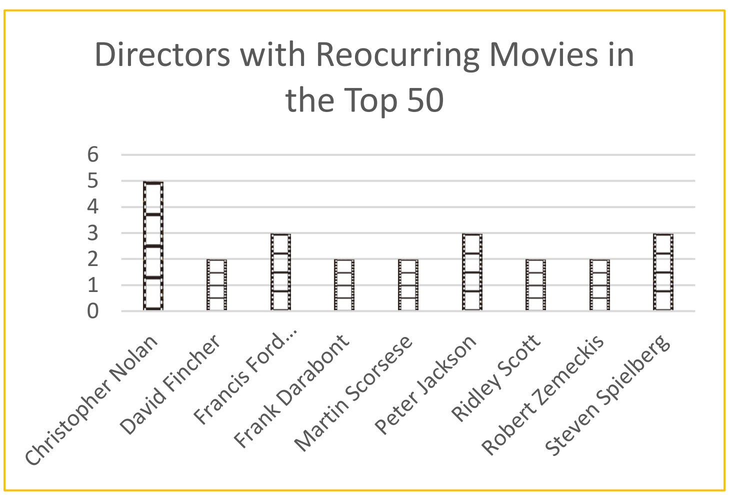
Variable	Degree of Confidence	Lower Level	Upper Level
Budget In Mill (USD)	90%	35.067	67.29
	95%	31.87	70.49
	99%	25.42	76.93
Gross in Mill (USD)	90%	99.47	160.16
	95%	93.44	166.18
	99%	81.31	178.32
Metascore (Out of 10)	90%	80.7900	85.28
	95%	80.29	86.33
	99%	79.29	87.33

*Visuals for Confidence Intervals are continued on the Confidence Intervals Section

Data

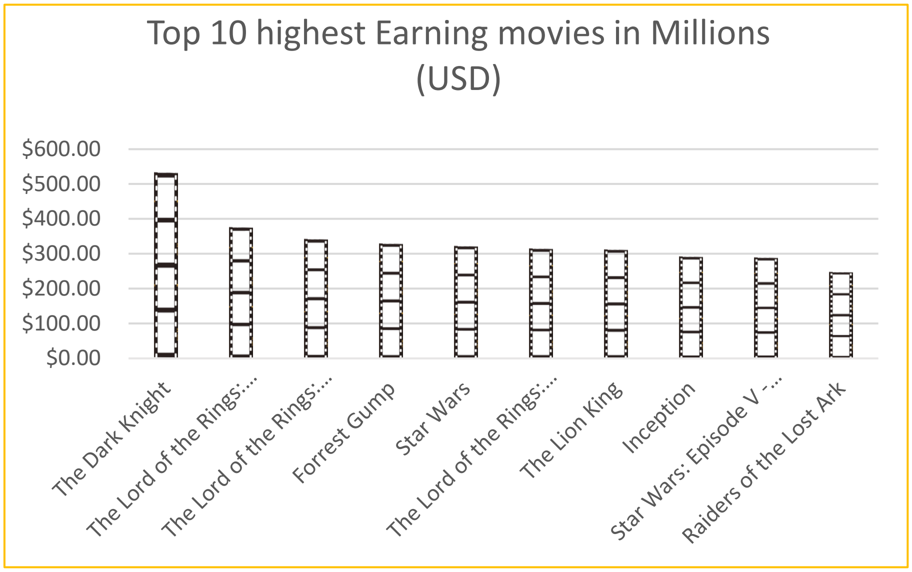






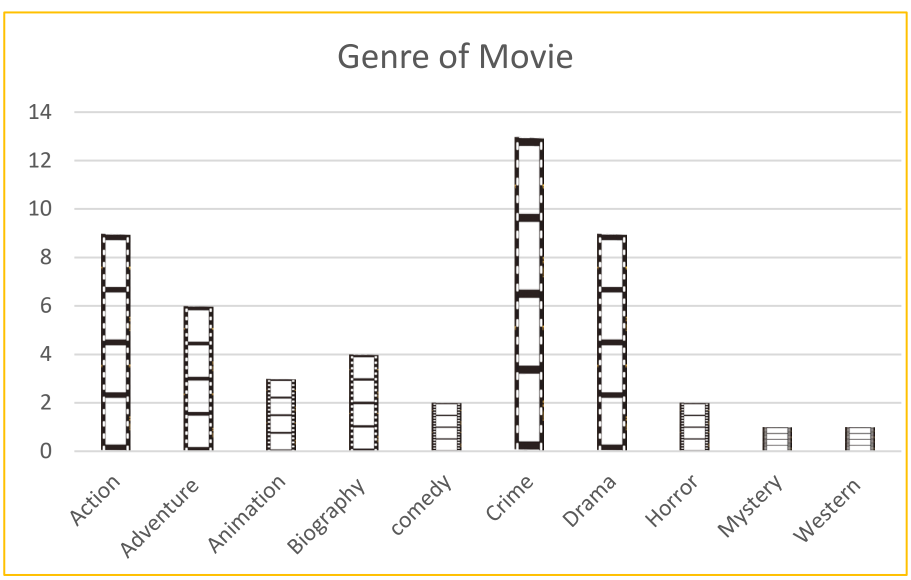
ACTORS & DIRECTORS

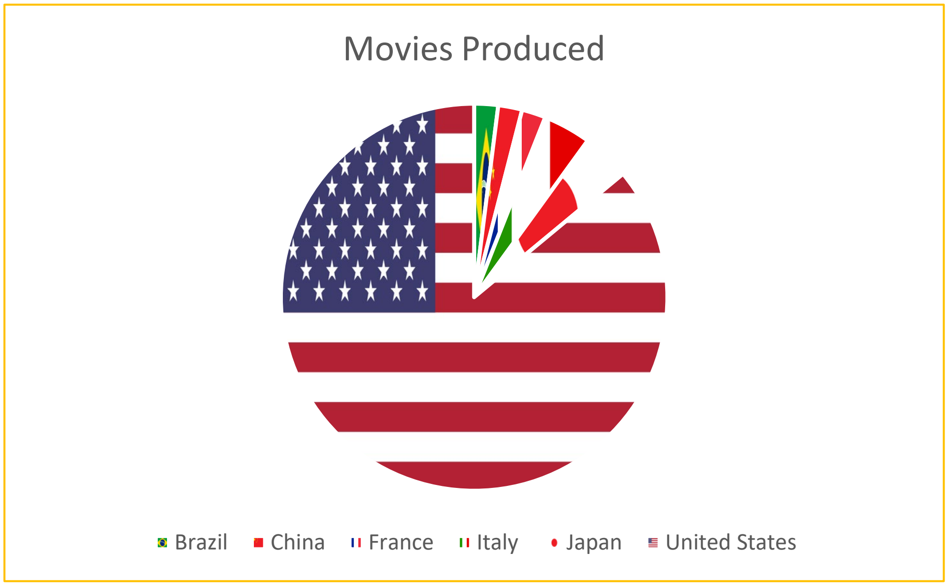
In our research of the top 50 movies of all time we found reoccurring names in both the director and lead actor category. Those who were in multiple movies in the top 50 list are represented. Tom Hanks and Elijah Wood are tied for first both with 3 movies on the list. Christopher Nolan has directed 5 of the 50 movies on our list.

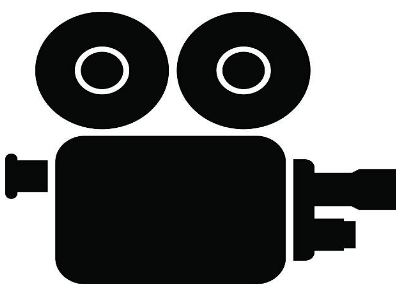


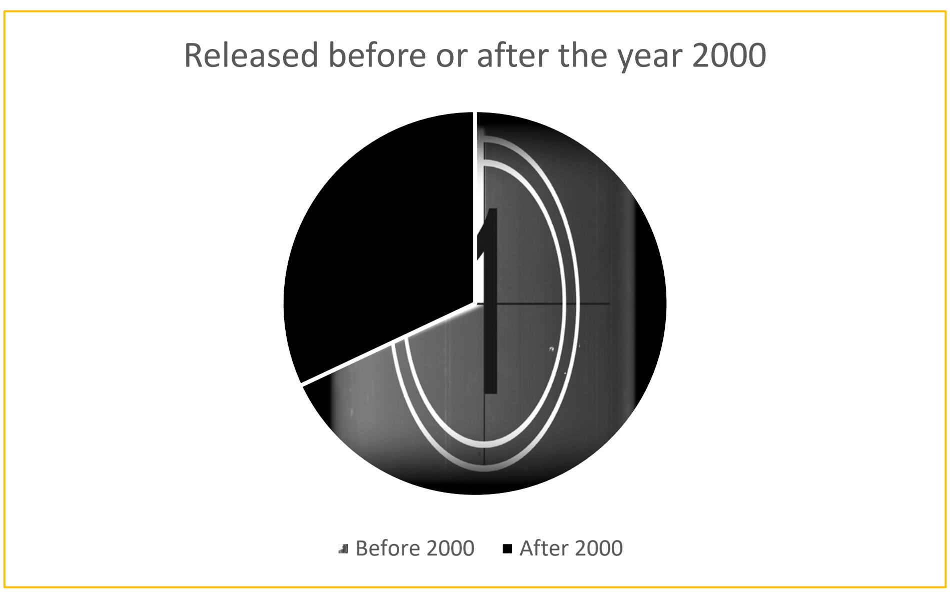
The genre that produced the most films on the top 50 list was the crime genre with 13 movies on the list. The second most frequent genre was both Action and Drama tied with 9 movies each.

The top 10 highest grossing movies on our list were all produced by big name directors from big studios. Although indie films have shown success, it is more probable to create a hit movie with a revered director and a big studio behind it.







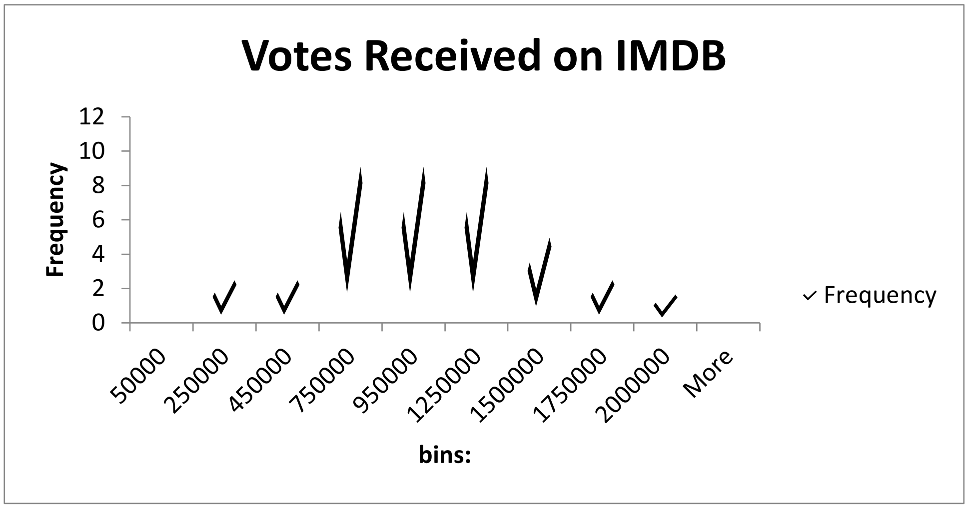


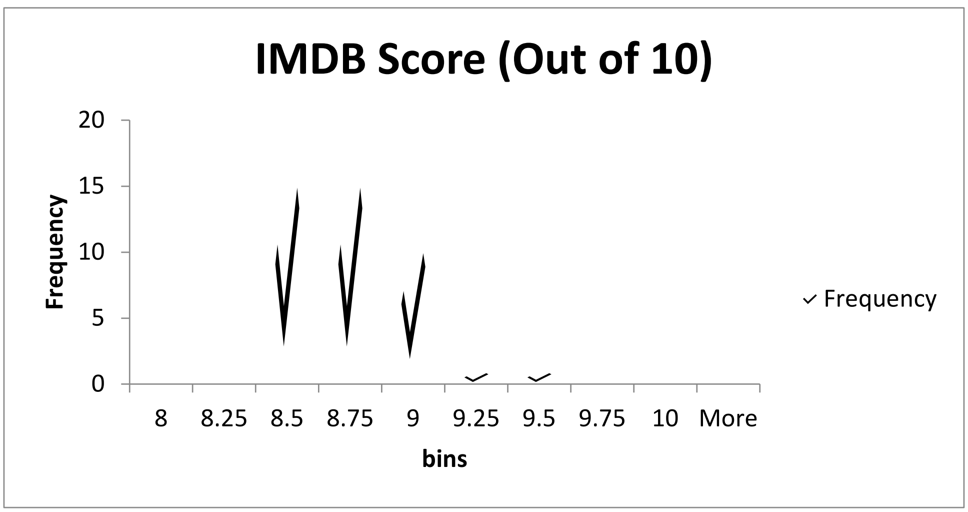
Patterns In Production

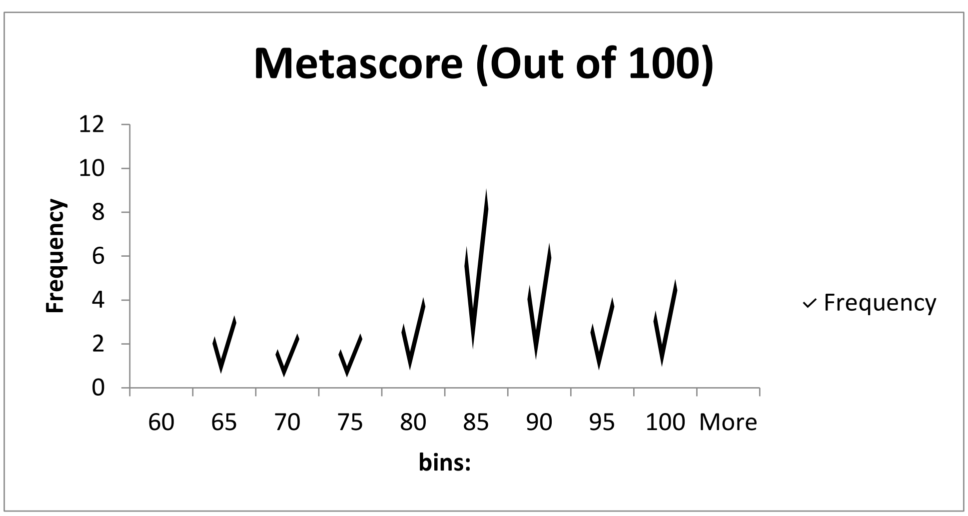
Patterns such as most frequent genre, Actors, directors, runtime, votes, etc., allow for us to identify the elements that all these films share and what aspects benefitted them when it came to producing the project. The country that produced the most movies on the list was the United States producing 43 movies. The majority of the list was also released before the year 2000.

peacock

Overall Ratings



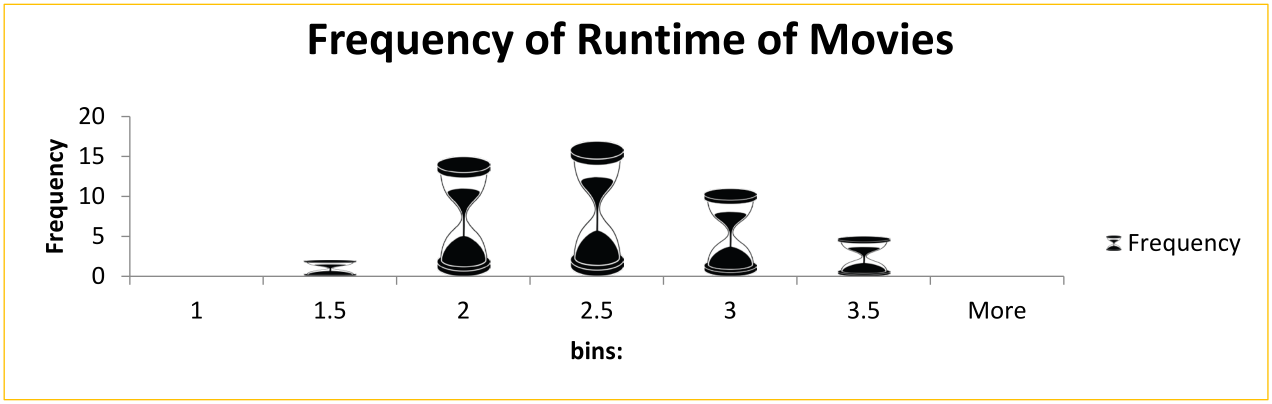


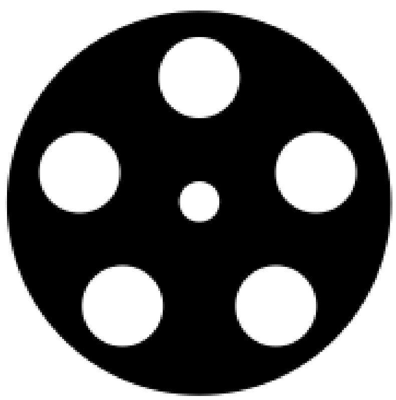


Datapoints:

Each graph shows the most common number of votes for that movie on IMDB and both their Metascore and their IMDB score. Most movies received between 750,000 votes and 1,250,000 votes. The IMDB score and Metascore are different grading scales created by different film review platforms and ae used widely amongst judging films. Most of the top 50 movies scored between an 8.5-8.75 out of a max score of ten. The Metascore has a wider range of data than that of the IMDB score. Although, most of the data still resides in the 85 out of 100 range and higher.

Runtime





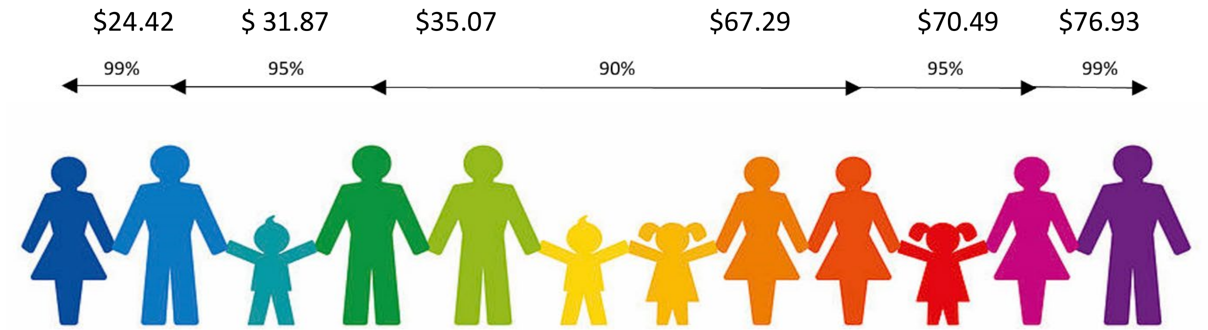
RUNTIME

Another important factor that was investigated was the runtime of a movie. The runtime of a movie determines how long a director thinks he can keep an audience member in their seat without them deciding the film wasn't worth their time. From this data, we concluded that the perfect runtime for a movie at the caliber we are going for to be 2.5 hours.

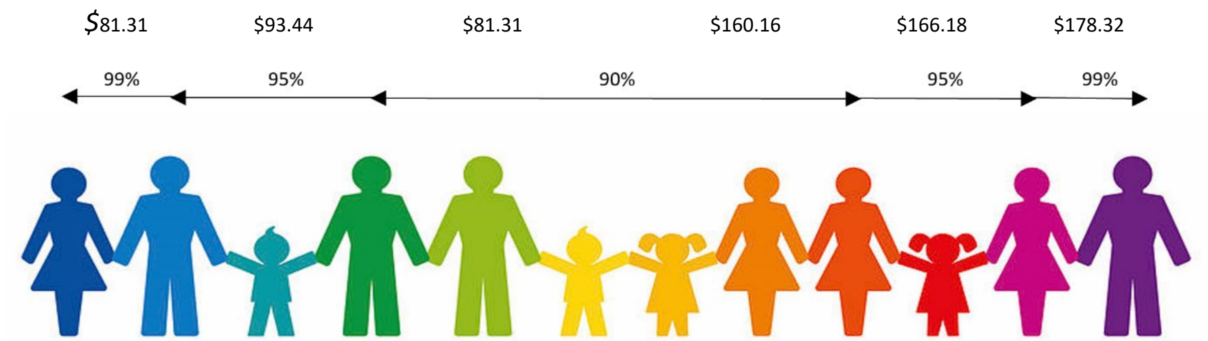
peacock

Confidence Intervals

Budget in Mill (USD):



Gross Earnings in Mill (USD)



Metascore (Out of 10)

peacock

Data Tests

Single Sample

The Level of Confidence for all tests is 95%

Test 1

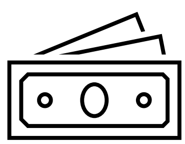
We were given an average for gross earnings amongst the top 50 movies in millions and the value was $162 million we believe that the population means is greater than or equal to that average (Ho). The other possibility is that my original hypothesis could be incorrect, and the population means is less than the average we were given. If our P-Value is greater than the percentage, we are willing to be wrong (5%) then we will accept the hypothesis. However, if the P-Value is lower than our Alpha (5%) then our data will support the contrary argument (Ha).

Ho: population average gross earnings in millions > or = 162

Ha: population average gross earnings in millions < 162

P-value: 0.04

Impact: Based on the sample data we can support the claim that the population average for gross earnings of movies is less than the calculated $162 million. This is the claim we can make based on the sample data we have



Test 2

The average amount of votes received on IMDB amongst our top 50 movies of all time sample is 950,000 votes. We are not sure whether the population average will be higher or lower than this value We performed a two-tailed hypothesis test using the information below.

Ho: The averages vote for a movie on IMDB >= 950,000

Ha: The average amount of votes for a movie on IMDB < 950,000

P-Value: 0.07

Impact: Our data support the claim that the average amount of IMDB votes for a movie is greater than or equal to 950,000. We will want our future projects that be of this caliber to achieve at least 950,000 votes by focusing exposure on IMDB.

Test 3

The average budget for the top 50 movies of all time is $56 million. However, we think that number is not entirely accurate. We believe that the population means is higher than the value we are given. We used an upper tail hypothesis test to find budget values for production.

Ho: The population average budget for a movie is >= $56 million

Ha: The population average budget for a movie is < $56 million

P-Value: 0.69

Impact: Our data support the claim that the population average budget for a movie is great than or equal to $56 million. This means our budgets for upcoming projects could look like these numbers. Understanding the budgeting necessary for full-scale movie production allows us to work with more information while also financing a separate production branch.

Two Sample

Test 1

To dissect the movie industry's profitability, we wanted to see if the average gross earnings of movies have gone up. We did this by splitting the gross earnings variable between two categories. One category is the movies released before the year 2000 and the movies released after the year 2000. Our hypothesis was that the movies released after the year 2000 on average made more than movies released before the year 2000.

Ho: The population average of gross earnings of movies released after the year 2000 is >= The population average of gross earnings of movies released before the year 2000 is

Ha: The population average of gross earnings of movies released after the year 2000 is < The population average of gross earnings of movies released before the year 2000 is

P-Value: 0.077

Impact: Our data has sufficient evidence that the population average gross earnings for movies released after the year 2000 are greater than or equal to that of movies released before the year 2000.

Test 2

Our team wanted to determine if the demographic for a lead actor has changed at all since after the year 2000. By splitting the actors in movies released before and after the year 2000 their populations could be compared, and we could make an educational deduction about our lead actors going forward. We believe that due to increases in medicine, health, and hygiene the lead actors in movies released after the year 2000 will be naturally taller.

Ho: The average height for the lead actor of the population of top 50 movies released after the year 2000 is greater than or equal to the average heights of lead actors in the top 50 movies released before the year 2000

Ha: The average height for lead actors of movies after 2000 Is less than the average height of lead actors in movies released before the year 2000

P-Value: 0.63

Impact: Our data support the claim that the average height for lead actors in movies released after the year 2000 is greater than or equal to the average height of lead actors in movies released before the year 2000. Knowing this data, we can infer that the modern audience prefers a taller lead actor in their films.

<u>Test 3</u>

Considering the advancements made in film and the plethora of ideas that are continuously explored we decided to compare the average IMDB Ratings (out of 10) of the movies on the list released before the year 2000 and after the year 2000. Our hypothesis preceding the upper tail two-sample test was that the average IMDB rating for movies released after the year 2000 is more than the average IMDB rating for movies released before the year 2000.

Ho: The average IMDB rating for movies released after the year 2000 is >= the average IMDB rating for movies released before the year 2000

Ha: The average IMDB rating for movies released after the year 2000 is< the average IMDB rating for movies released before the year 2000

P-Value: 0.864

Impact: There is sufficient data to support the claim that the average IMDB rating for movies released after the year 2000 is greater than or equal to the average IMDB rating for movies released before the year 2000. Based on the modern film rating systems movies that explore newer themes and take advantage of the growing modern film era. Movies are being rated on a higher scale than they have been in the past.

ANOVA Tests

<u>Test 1</u>

Our team wanted to understand if every genre had the sum average runtime per movie to determine if longer (less cost-effective) movies were doing just as well in their respective genres as others. By understanding, if one genre has a difference in averages than the rest, we can make an educated response moving forward to the types of genres we use. Our hypothesis was that all genres have the same average movie runtime.

Ho: All ten genres have the same average movie runtime per genre

Ha: At least one of the genres does not share the same average movie runtime as the others

P-Value: 0.014

Impact: Now that we understand that there is a difference in trend length when transitioning between genres in the production industry. Depending on gross earnings and budgets we can determine if we can cost-effectively have a longer runtime movie that can still keep the audience at the edge of their seat.

Test 2:

The second ANOVA test we ran was to see if all countries received the same average Metascore (out of 200) for the movies they release to help us understand if some countries have a higher average than others, our hypothesis was that All countries receive the same average Metascore out of 100.

Ho: All countries receive the same average Metascore out of 100.

Ha: At least one country does not have the same average Metascore as the others

P-Value: 0.12

Impact: Our data show sufficient evidence to support the claim that all countries do indeed receive the same average Metascore for the films they release. This means that no country has an advantage over others when being compared to the Metascore film review platform.

Test 3

To accurately develop data on genres, we ran a test to see if all ten genres had the same average gross earnings for their movies. Our hypothesis was that based on our last test relating to Metascore averages being the same amongst the genres, we believe that all ten genres will have the same gross earnings for movies.

Ho: All 10 genres have the same average gross profit in millions (USD)

Ha: At least one genre does not share the same average gross profit as the other genres

P-Value: 0.004

Impact: Now moving forward knowing that some genres amongst the 10 genres have higher average gross earnings for those movies compared to the others we can produce movies with the higher average earnings out of those genres. Knowing they all aren't equal in average earnings we can take that into consideration for the project moving forward.

Testing for Independence

Test 1

Our team wanted to understand if directors liked to stick to their genre or sway towards others for whatever reason they so choose. We wanted to know that our genre doesn't restrict the list of great directors we can choose from. Our hypothesis was that the director of a movie is independent of genre.

Ho: The Director of a movie is independent of genre

Ha: The director of a movie is not independent of genre

P-Value: 0.621105368

peacock

Impact: Moving forward we now have sufficient data to support our earlier claim that the director of a movie is independent of genre. This creates flexibility for our new production department as we can have in-house directors that can work with us directly and produce our projects.

Test 2

Our team wanted to know if some lead actors stick with their type-cast role or move on to other things. We performed an independent test between lead actors on the top 50 list and genre. Our hypothesis was that actors are indeed independent of genre

Ho: Actors are independent of genre

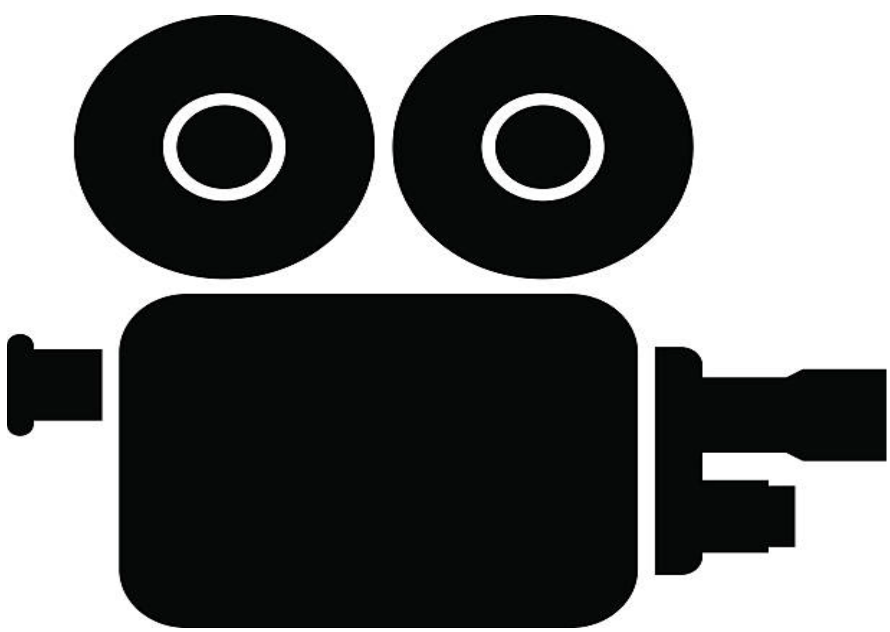
Ha: Actors or not independent of genre

P-Value: 0.43076759

Impact: There is sufficient evidence to support the claim that actors are independent of genre. Moving forward our production team now knows that neither the director nor the actor is dependent on genre. This information gives us more flexibility when deciding who we are going to hire for upcoming projects.

Summary

For Peacock streaming services to take the company in a direction of more production-oriented business rather than securing contracts on other movies Peacock needs to understand the similarities in the data that can lead them to create a successful movie. For example, it would be in Peacock's benefit to produce crime and drama-oriented movies because of their likelihood to succeed. The movies we would need to produce to produce movies at this high of caliber is a sizeable budget, a big named director (American), a big named lead actor. The goal being for our films would be to reach the 8.5 out of 10 or above range on the IMDB score (Where most of the top movies reside) and in the 85-100 range for our average Metascore ratings. We would also want to produce movies that do not exceed 2.5 hours as to not run the risk of losing the audience's attention thus ruining the credibility of our films.



References:

IMDb. (2008). *IMDb - Movies, TV and Celebrities*. IMDb. https://www.imdb.com/

IMDB top 50 movies. (n.d.). Kaggle.com. Retrieved January 29, 2022, from https://www.kaggle.com/aditya1303/imdb-top-50-movies