What makes a movie earn more money?

Dataset:
https://www.kaggle.com/datasets/harshitshankhdhar/imdb-dataset-of-top-1000-movies-and-tv-shows

Project Description:

For our DATA-151 project, we plan to research what factors are most important when determining how much money a movie will earn. There are many seemingly important factors such as release year, runtime, genre, rating, and director that could have significant impacts on a film's earnings. We plan to investigate and quantify the various effects that these factors have and discover which one of them has the largest impact. We believe that this topic is important because movies are a major part of modern culture. By looking into what leads to a movie earning more money, we can get a better understanding of what modern culture values, what stories appeal to people, and what people want out of entertainment. It is also possible that the results of this research may lead us to finding the most desirable movie to produce next, hoping it to be a massive moneymaker (assuming all of the factors could reasonably be assembled).

Our dataset contains the top 1000 ranked movies on IMDB. Since our dataset only accounts for movies that appeared in the top 1000, it is possible that our data does not entirely account for all movies. In addition, our dataset doesn't have any movies that were created after 2020, so our data may be slightly outdated and thereby somewhat less accurate to modern culture. Despite these shortcomings, we believe that it provides a good depiction of the broader film industry.

Our hypothesis is that a movie's release year is the most important factor for determining how much money it will earn. We plan to isolate, measure, and compare the various effects that the different factors have, by comparing correlation coefficients and average earnings. By doing this, we can determine what factors are more important than others.

What factor is most important for a movie to earn more money?
-Year
-Runtime
-Genre
-IMDB Rating
-Metascore
-Director
-Stars
-Num of votes

Hypothesis: A movie's release year is the most important factor for determining how much money it will earn.

Why is our hypothesis important?
- Movies are a big part of modern culture, and determining what most draws viewers to watch movies is a way to discern cultural values and what people want out of entertainment. The results of this data would allow us to better understand what makes a movie successful and perceived as successful.

How will you be able to tell/measure the success of the research question and hypothesis?
- We will be able to tell the success of the research question and hypothesis if we can conclusively measure and compare the effects of the various factors on movie earnings. Specifically, if we can determine correlation strength and coefficients as well as average earnings.

How will we model our data?
- We plan to compare our categorical correlation coefficients against each other, and attempt to display them in a graphical way, most likely making use of excel, and probably comparing average earnings of the quantitative variables in graphs (also using excel). Until we have calculated our variables and can try creating the graphical representations for them, we are unable to say specifically how we are choosing, but we do plan to use visual displays as they are more easily able to communicate the information in a readable way. Upon a small amount of research; pie charts, frequency tables, and bar graphs appear to be the best way to visually display this categorical data as the many other methods of graphing do not work with variables that aren't in number form.

Schedule:
-Meet at 10:30 on Fridays to work on project. (Work for however long we deem necessary)
-Other times as needed.
Part 1 (Due 2/10)
-Proposal done by 1/21
-Presentation ready by 2/4
Part 2 (Due 3/24)

-Report done by 2/25
-Short presentation ready by 3/4

Part 3 (Due 4/19)

-Demonstration ready by 4/1
-Paper done by 4/15

Part 4 (Due 5/8)

-Miscellaneous done by 4/29
-Final presentation done by 5/6