Introduction

The Marvel Cinematic Universe films started being made when I joined high school and since then they have become a part of mainstream media. So far, the movies have been popular all over the world. What makes them popular is the superheroes are the stars, using the characters and storylines they created over the last couple of decades. Due to the sizes of both Marvel Studios and Disney Studios, the resources and payoff for these projects are enormous.

Even though I have experienced these films first-hand, I have many questions about how profitable and how successful they are. I want to know if the production of the MCU films has increased over time plus whether they have generated more revenue since the first film. I also want to identify any trends within the scores and reviews the movies receive. Lastly, I am curious if there is any relationship between the ratings for each movie and the amount of money it produces.

Collect and Cleaning the Data from Wikipedia

Data scientists spend 60% of their time cleaning and organizing data. Lucky for us we can scrape the tables from [Wikipedia.org]((https:/en.wikipedia.org/wiki/Marvel_Cinematic_Universe)) with the r package rvest, navigating the HTML code from this URL to extracting the tables is simple. However, some rows are insignificant for our analysis and exist for aesthetic purposes. Also, most of the columns need to be coerced to correct data types so that we can manipulate them later. This includes renaming columns, changing character columns to numeric ones, and removing specific strings from elements.

Let’s answer my questions!

To do this I will be generating plots to visualize the data I cleaned in a way that answers each question. These will be made using the r package ggplot2 and will include scatterplots, histograms, and trendlines. If you want to use the data download these csv files and run the code below.

How profitable are the MCU movies?

To quantify profit for each my movie I calculated the gross profit margin for each movie. To calculate this, you need the revenue and the budget which we have in the data. The gross profit margin is given by the formula

\begin{equation}

\text{Gross Profit Margin} = \frac{Revenue - Budget}{Revenue}.

\end{equation}

Now that I’ve calculated I’m going to make a level variable and group the movies by into 3 intervals of gross profit margin to distinguish groups in the graphs below. The levels are: Greater than 80%, Between 60% and 80%, and Less than 60%.

From this histogram we observe that most of the films lie in the range of between 60% and 80% gross profit margin. We also notice that there is one outlier indicated by the blue bar. This film is The Incredible Hulk, which was the second MCU movie produced in 2008, from this histogram we can see the movie made a lot less profit than the other MCU movies.

Has the amount of profits and budgets for the MCU movies increased since the first?

Now I want to see if the movies have been generating more profits since the first movie in 2008. To do this I plotted the transformed revenues and budgets, distinguishing them by circles and triangles. I transform them by log10 so that we can easily see the trends on the plot. I also used ggplot2 to add a moving average to visualize the average revenue and budget for each year from 2008 to 2019.

From this plot we can see that the revenues for the MCU films has been increasing over the years and the budgets for the movies have stayed relatively constant. From this we can conclude that the movies have been making more money over time without having to increase the budget of the movies. This makes sense because as more movies are produced, more people are exposed to them and hear about them and eventually go watch them.

Have productions for the movies increased?

Next, I wanted to know if they started making more MCU movies as they got more popular. To answer this, I made a histogram to count the number of movies produced each year.

From this histogram we can see that from 2008 to 2016 we have about 2 movies per year being produced. However, from 2017-2019, there have been 3 movies produced each year which shows that the production of films has increased recently. In this histogram we can also see that the movies in this time period have performed well based on their Gross Profit Margin.

Trend in scores and reviews?