Project MCU

Group Members:

Ana Morales, Eric Dowd, Marcus Guerrier, Nick Cresta

Motivation:

For our project, we wanted to focus on something fun and light-hearted. Since all of our group members are fans of Marvel, we wanted to use this opportunity to learn as much as we could about the Marvel Cinematic Universe (MCU) using what we've learned in class so far.

Marvel is the highest grossing movie franchise of all time - grossing over $22.5B world-wide. There is no question that the MCU is a successful franchise, but how could we explore this further using data analysis? After much deliberation, our main question that we asked throughout our project work was this: **"How has the cultural footprint for MCU grown since the release of the first movie (Iron Man) in 2008?"**

Summary:

- Based on our analysis and observations from our data (outlined in detail in below sections), we can conclude that the answer to our main question is yes. The MCU footprint has grown significantly over time as it relates to revenues and theater releases.
- With respect to ratings, however, there is less of an obvious positive trend. Critics do seem to 'like' movies in Phases 2 and 3 more than 1, but rise/fall of audience ratings did not seem to be affected by time.
- The Avengers series was definitely the most successful boasting the highest revenues, theater releases, opening box office revenues, and profits. To sum up - Avengers are strongest when they're a team.
- The Incredible Hulk, Ant-Man series, and Thor series appear to be the least successful and popular overall - with lower revenues, profits, and ratings.
- International footprint as it relates to gross revenues was significantly higher than domestic footprint.

Data Cleanup & Exploration:

To answer our main question, we wanted to focus on several key data sets including critic and audience ratings, box office revenues, international and domestic revenues, budget, and opening weekend revenues. We gathered this data using 2 main sources - an API call from OMDB coupled with several datasets from Kaggle[1].

After several rounds of reading in CSVs, merging dataframes, cleaning data, formatting data, and renaming columns, we arrived at a final dataframe - housing all the relevant data needed for analysis for all MCU movies (this is the 'final_df' in our code).

Some limitations to the data gathering process was definitely the availability of the data and when it was gathered. Marvel does keep a tight lid on their data, so gaining access to data that was up-to-date and accurate was a little bit of a

---

[1] Data sets used can be found at the following URLs:

(1) https://www.kaggle.com/zeegerman/hollywood-stock-exchange-box-office-data
(2) https://www.kaggle.com/narmelan/opening-weekend-box-office-performance-50mm
(3) https://www.kaggle.com/abilionb/umc-movies2008-2019
(4) http://www.omdbapi.com/

challenge. However, after several hours of searching Kaggle and other sources, we were able to obtain what was needed.

Data Analysis:

Using our final_df created in previous cleanup efforts, we created several graphs/charts to help us answer our main question, looking at several different data points in the process.

Important fact: MCU movies are broken up into 3 phases with a fourth incoming in early 2021 and storylines planned into 2028.

- Phase 1: Iron Man (2008) to The Avengers (2012)
- Phase 2: Iron Man 3 (2013) to Ant-Man (2015)
- Phase 3: Civil War (2016) to Spider-Man: Far From Home (2019)

*Total Box Office Revenue by Release Date*

- Line chart showing Worldwide Gross revenue graphed against each movie's release date.
- Dates are in chronological order starting with the first release (Iron Man 2008) to the last release (Spider-Man: Far From Home 2019).
- There seems to be a general upward trend as time goes on, even though peaks and valleys do exist.
- The most successful movies with respect to total box office revenue are the 4 Avengers movies: Avengers, Avengers: Age of Ultron, Avengers: Infinity War, and Avengers: Endgame.
- An interesting point about the valleys is that as time went on, the valleys also had an upward trend; becoming more "shallow" with time.
- Some exceptions to these observations are the Ant-Man series; they returned significantly less revenues than the other movies.
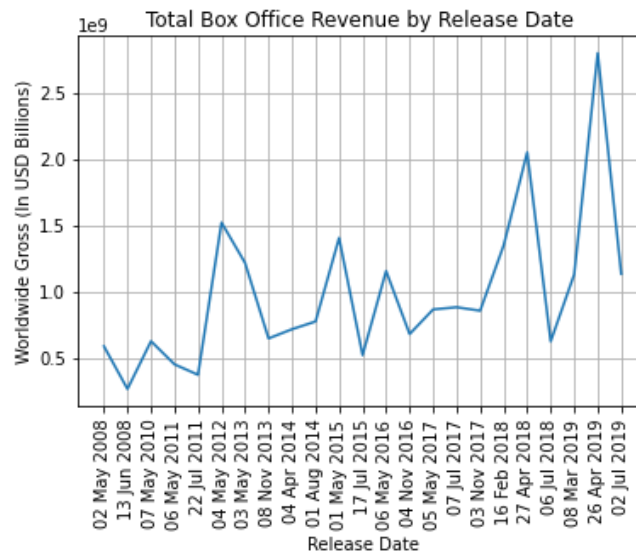

Critic Ratings

- Bar chart showing critic ratings by MCU movie. Critic ratings are sourced from Rotten Tomatoes and Metacritic.
- Movies are still shown in chronological order, except x-axis data labels are movie names instead of release dates.
- None of the MCU movies went below a 50 rating.
- Although there is no obvious trend as it relates to release date; however, Phase 3 has consistently higher ratings than the previous 2 phases.

- Blank Panther has the highest scores for both Metascore and Tomatometer for MCU with 88 and 96 respectively.
- Third movies in each franchise have the best Tomatometer average (Thor: Ragnarok, Captain America: Civil War, Iron Man 3, Avengers: Infinity War).
- Worst Tomatometer run happened between 2008-2012, including Incredible Hulk (67%), Iron Man 2 (73%) and Thor (77%); none of these movies being able to muster more than an 80% rating.


Critic Ratings (Source: Rotten Tomatoes and Metacritic)
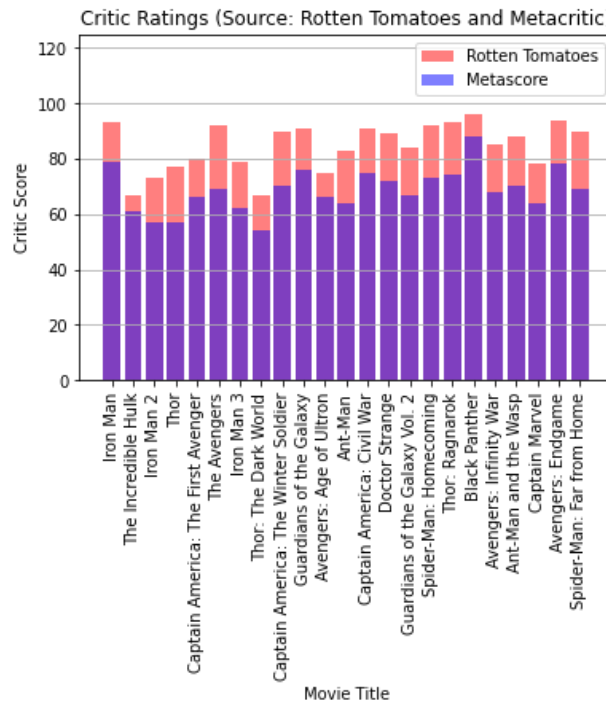
IMDb Scores for each Film

- Bar chart showing IMDb user ratings by MCU movie.
- Movies are still shown in chronological order, except x-axis data labels are movie names instead of release dates.
- Lowest scoring movie is Incredible Hulk with 6.7; highest score is tied between the last 2 Avengers movies at an 8.4 rating (Infinity War and Endgame).
- Similar to the previous bar chart, there is no obvious trend as it relates to time.

IMDb Scores for Each Film
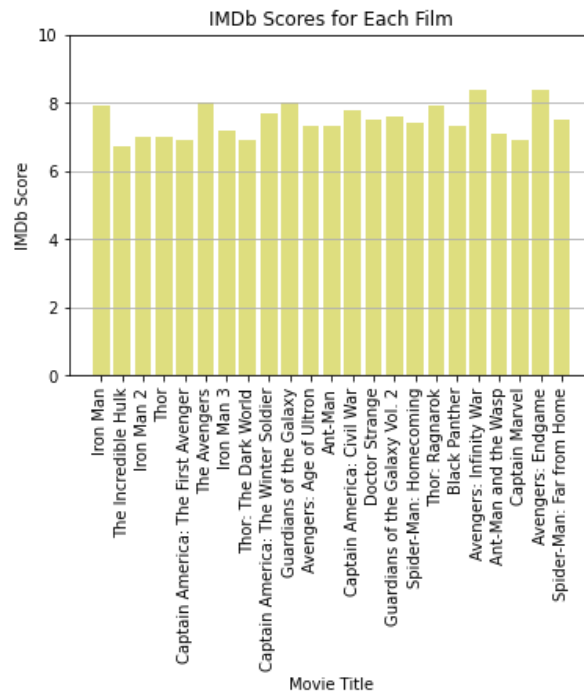
Worldwide Gross vs. Budget

- Bar chart showing worldwide gross revenues vs. budget per MCU movie.
- Movies are still shown in chronological order, except x-axis data labels are movie names instead of release dates.
- Immediately you can see the revenue for almost every MCU movie surpassed the budget by more than 100%, significantly higher in some cases. Incredible Hulk was the only one to not have as big of a return and the lowest worldwide gross revenue.
- Highest worldwide gross revenue is Avengers: Endgame with the closest to it being Avengers: Infinity War.
- The highest revenues were seen in Phase 3 with a caveat being that Phase 3 has significantly more movies in it than either Phase 1 or 2.

Worldwide Gross vs. Budget
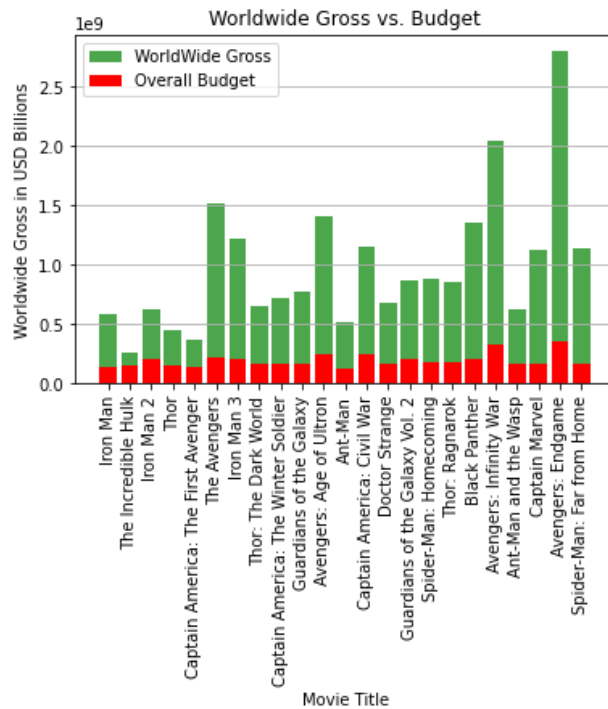
Opening Box Office Gross revenue

- Line chart showing Opening box office gross revenue graphed against each movie.
- Movies are still shown in chronological order, except x-axis data labels are movie names instead of release dates.
- The distribution seen in this graph follows an extremely similar pattern to the first graph of Total Box Office revenues.
- There is a slight positive trend, with the biggest spikes being the Avengers movies.
- Surprisingly, Thor (2011) has the lowest opening box office gross revenue, even though it was not the lowest rated (i.e. The Incredible Hulk).

Opening Box Office Gross in Chronological Order
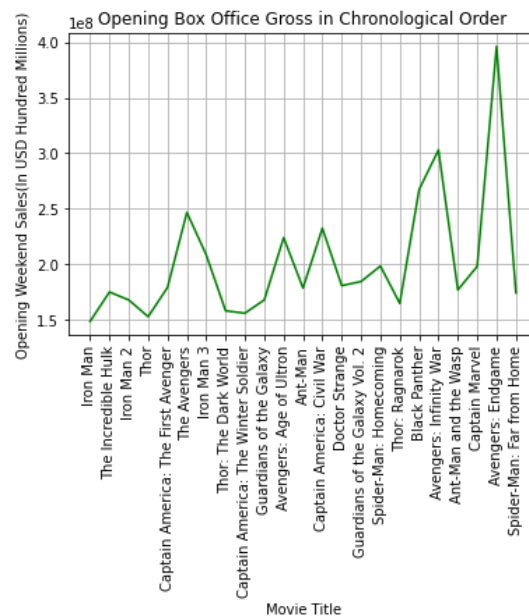
Domestic Gross vs. International Gross

- Line chart showing Domestic vs. International gross revenues graphed against each movie.
- Movies are still shown in chronological order, except x-axis data labels are movie names instead of release dates.
- An interesting observation is that for most of Phase 1, domestic and international gross revenues were almost the same. As time goes on and Phase 2 and 3 are created and released, the international footprint grows.
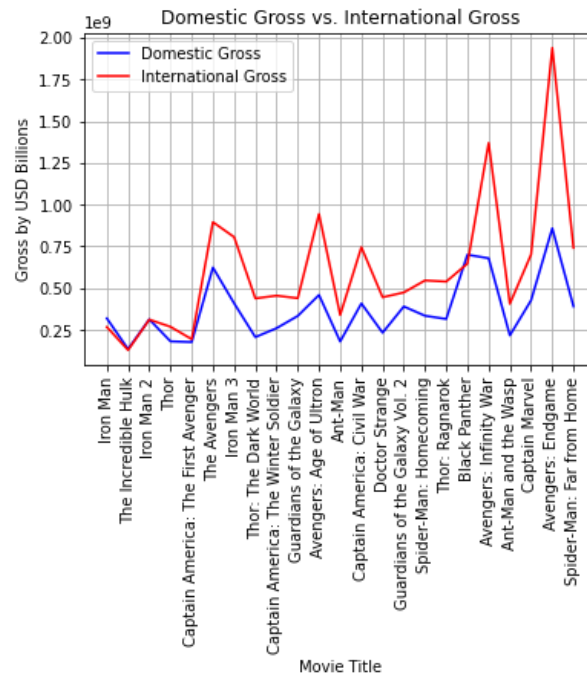
_Critic reception vs. Worldwide Box Office revenue_

- Scatterplot showing relationship between critic reception (Tomatometer) to worldwide box office revenues.
- Linear regression and correlation coefficient calculated/added to the graph as well.
- Based on correlation coefficient, there is a positive correlation (not strong, but existent) between Tomatometer and Worldwide Box Office revenues. Meaning, as Tomatometer rating increases, Worldwide Box Office revenues tend to increase as well.
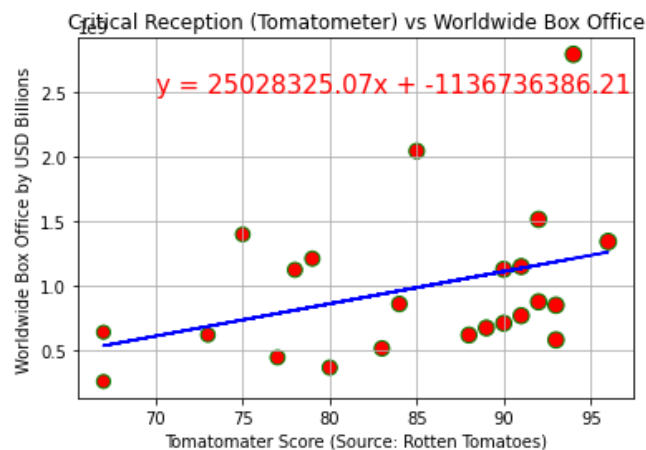
Widest Theater release

- Line chart showing widest theater release per movie.
- Movies are still shown in chronological order, except x-axis data labels are movie names instead of release dates.

- There is a much stronger positive trend with this data versus the revenues. Widest theater releases for each movie increased dramatically throughout the life of the MCU franchise.
- Additionally, it seems there was a tendency to release the group films (i.e. Avengers, Iron Man) to a greater number of theaters than solo films with the exception of Spider-Man: Far From Home - the last movie of Phase 3 has the highest theater release even though it's a solo film.

Widest Theater Release for Each Film
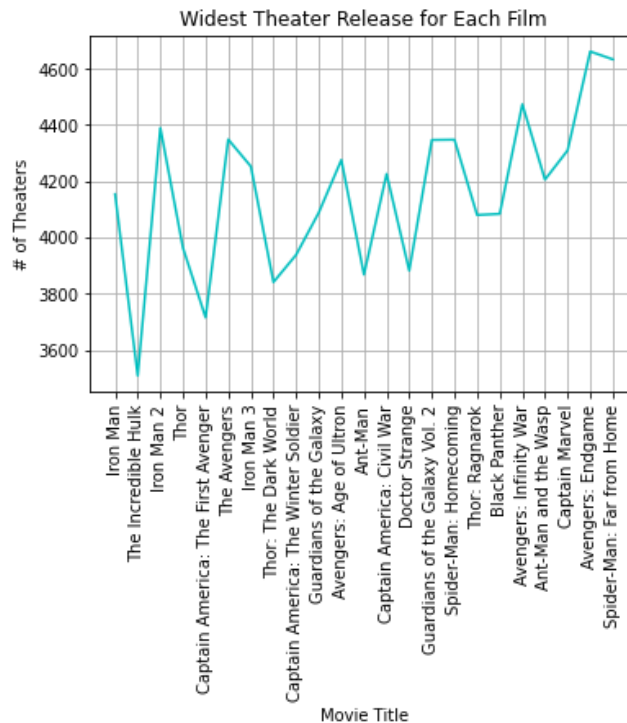
Discussion & Post-Mortem

Reminding ourselves of our findings:

- Based on our analysis and observations from our data (outlined in detail in below sections), we can conclude that the answer to our main question is yes. The MCU footprint has grown significantly over time as it relates to revenues and theater releases.
- With respect to ratings, however, there is less of an obvious positive trend. Critics do seem to 'like' movies in Phases 2 and 3 more than 1, but rise/fall of audience ratings did not seem to be affected by time.
- The Avengers series was definitely the most successful boasting the highest revenues, theater releases, opening box office revenues, and profits. To sum up - Avengers are strongest when they're a team.
- The Incredible Hulk, Ant-Man series, and Thor series appear to be the least successful and popular overall - with lower revenues, profits, and ratings.
- International footprint as it relates to gross revenues was significantly higher than domestic footprint.

One point in particular that was surprising analyzing this data was the fact that both critic and audience ratings did not follow a positive trend similar to the revenues and theater releases. Otherwise, our findings were in line with our expectations - revenues (opening weekend, domestic, international, worldwide, and total overall) followed an upward trend as time went on, signifying a rise in interest as each movie was released.

Apart from the difficulties mentioned above regarding sourcing up-to-date and accurate data, we did not experience any difficulties that we were not able to overcome. There were some instances in which we had to ensure our data was the correct type in order to be plotted/graphed, in addition to some minor errors along the well, but for the most part, we worked together to solve our problems and finish our analysis without any major problems.

If we had two more weeks, some other topics around MCU that we would've loved to explore would have included analysis around merchandise per movie, box office revenues by US city that would've allowed for some heatmap exploration, perhaps some comparisons with other movie franchises other than MCU (i.e. Star Wars or DC), social media impact of each movie (such as YouTube searches, Instagram or Facebook posts, etc.), or even a look into how the US/world economy affected (if at all) the interest in MCU. The main reason for us not exploring these further, other than time and holidays, would be access to this data. Most of these sites have extreme restrictions on access to this data accompanied by hefty fees - maybe next time!