PROJECT MCU

Ana Morales, Eric Dowd, Marcus Guerrier, Nick Cresta
Rutgers Data Analysis Bootcamp 2020/2021

Motivation & Data

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?"**

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 (resources listed on last page in presentation).

Summary

➔ Based on our analysis and observations from our data, 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**.

Exploration & Cleanup

Data cleanup process was documented in detail in Jupyter notebook.

Some limitations that we encountered during this process was ensuring our data was the right type in order to be plotted correctly.

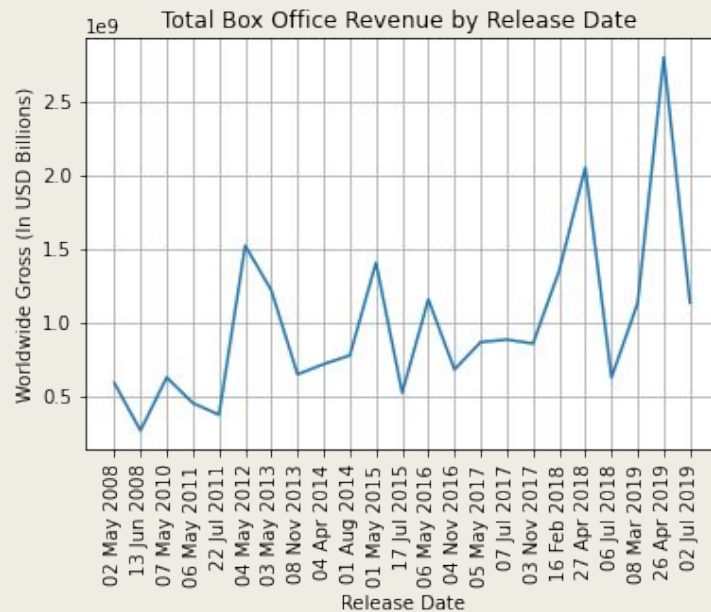**Let's get to the demo!**

Data Analysis - Intro Information

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)
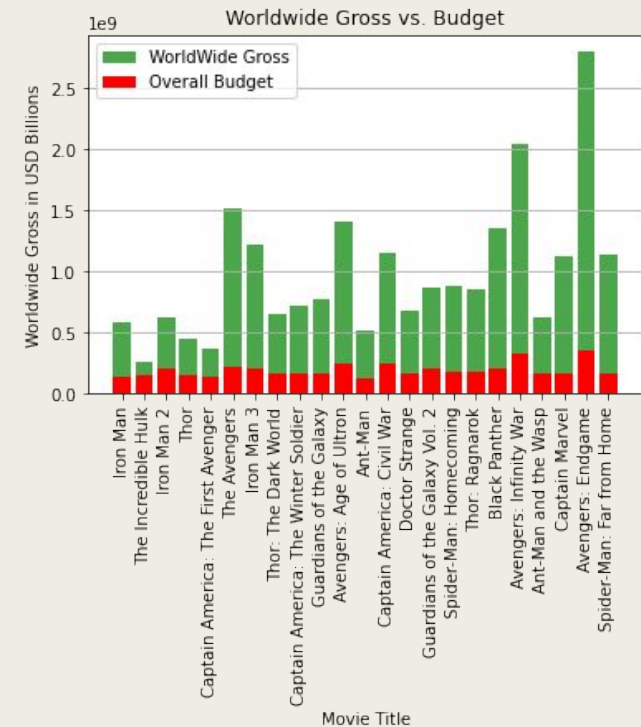- Phase 3: Civil War (2016) to Spider-Man: Far From Home (2019)

Data Analysis

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.

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.
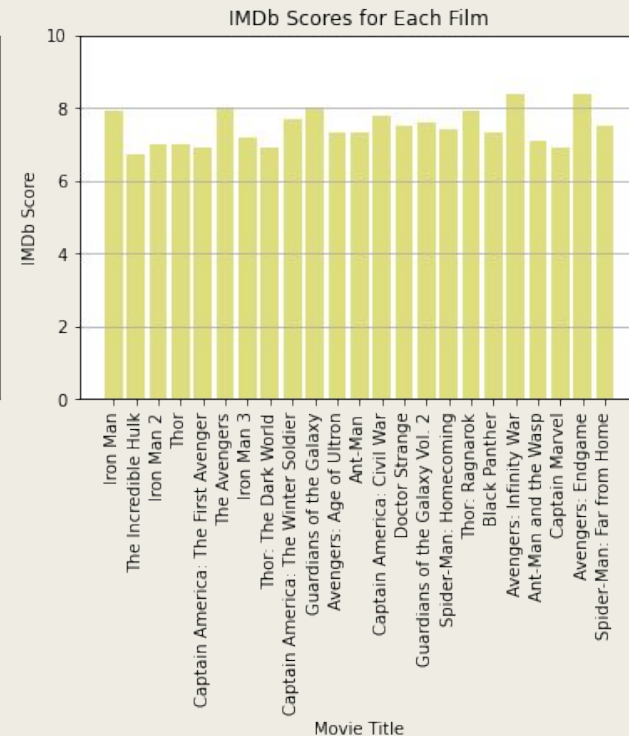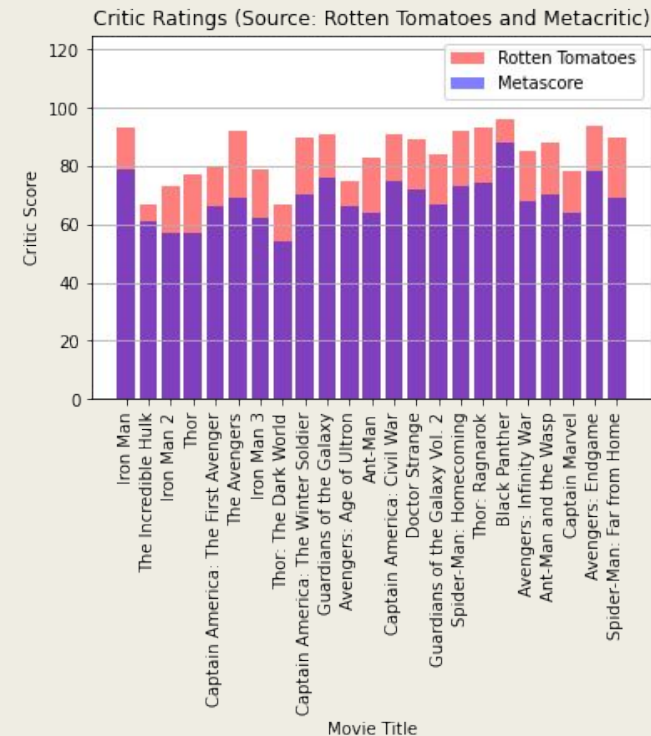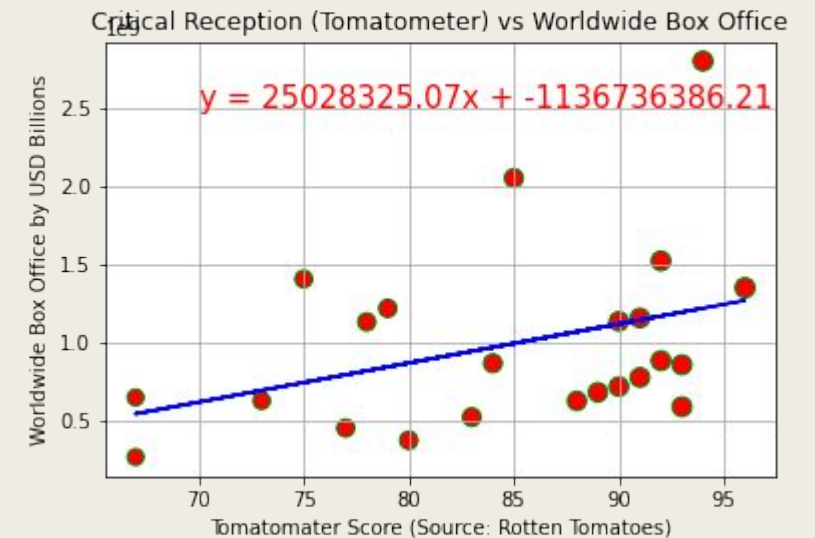
Data Analysis cont'd

Although there is no obvious trend for critic ratings as it relates to release date, critics did consistently **give Phase 3 higher ratings** than Phases 1+2. None of the MCU movies **went below a 50** rating.
Third movies in each franchise have the best Tomatometer average (Thor: Ragnarok, Captain America: Civil War, Iron Man 3, Avengers: Infinity War).
For user ratings, lowest scoring movie is Incredible Hulk with 6.7 and highest score is tied between the last 2 Avengers movies at an 8.4 rating (Infinity War and Endgame); **there were no other obvious trends seen.**
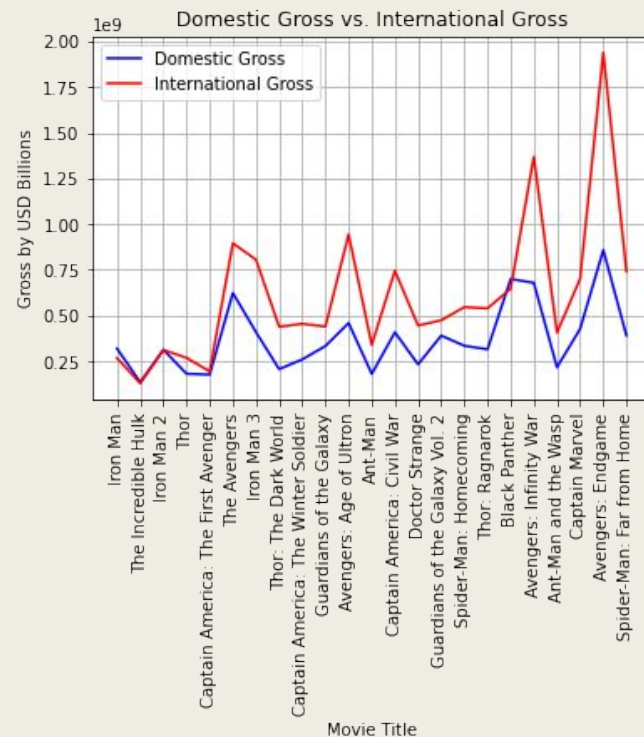
Correlation coefficient calculated was 0.37. Based on this, there is a **slight positive correlation** between Tomatometer and Worldwide Box Office revenues. Meaning, as Tomatometer rating increases, Worldwide Box Office revenues tend to increase as well.

Critic Ratings (Source: Rotten Tomatoes and Metacritic)

IMDb Scores for Each Film

Critical Reception (Tomatometer) vs Worldwide Box Office

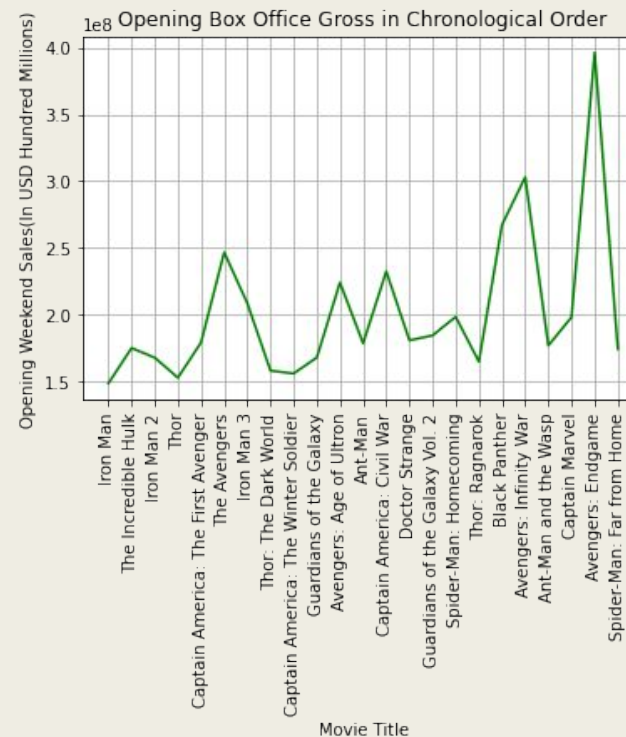$$y = 25028325.07x + -1136736386.21$$
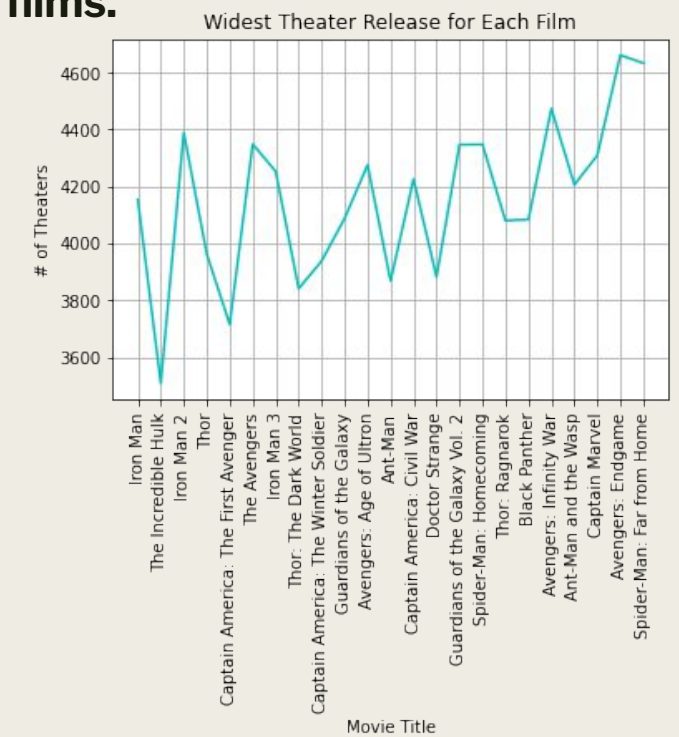
Data Analysis cont'd

An interesting observation for the below 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**.

The distribution seen in this graph follows an extremely **similar pattern** to the first graph of Total Box Office revenues. Surprisingly, Thor (2011) has the lowest opening box office gross revenue, even though it was **not the lowest rated** (i.e. The Incredible Hulk).

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. 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.**

Discussion & Post-mortem

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) and number of theaters released to followed an upward trend as time went on, signifying a rise in interest as each movie was released.

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
- 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

Resources

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/

Questions?

THANK YOU!