

Predicting Lifetime Gross of Box Office Movies

Capstone Project

Stacy Shingleton

The global box office revenue was over 42 billion dollars in 2019, and is projected to continue growing



credit: movienewsnet.com

Production companies invest
hundreds of millions of
dollars on individual movies,
in hope of a box office hit



credit: [imdb.com](https://www.imdb.com)

If production companies could better predict the lifetime gross of their upcoming movies, they could adjust their budget accordingly.

**Box Office Mojo was used to
scrape the top 1000 lifetime
adjusted gross domestic
movies.**

**200 movies were listed on
each page of five pages.**

The screenshot shows the homepage of Box Office Mojo by IMDbPro. The header features the logo 'Box Office Mojo' and 'by IMDbPro' next to a search bar with a magnifying glass icon. Social media links for Facebook and Twitter are also present. Below the header, a navigation bar includes tabs for 'Domestic', 'International', 'Worldwide', 'Calendar', 'All Time' (which is highlighted in yellow), 'Showdowns', and 'Indices'. Underneath the navigation bar, there are links for 'Overall', 'Weekend Records', 'Daily Records', and 'Miscellaneous Records'. The main content area is titled 'Top Lifetime Adjusted Grosses' and includes a sub-section 'By Adjusted Gross'. A dropdown menu for '2020 Ticket Price' is visible. At the bottom, it says 'Data as of Jun 5, 2:40 PDT' and includes navigation links for 'Previous page' and 'Next page', with the current page range '1-200 of 1,000' highlighted in orange.

credit: boxofficemojo.com

From each of the five front pages, the adjusted lifetime gross, lifetime gross, and estimated number of tickets sold was scraped for each of the 200 movies.

Box Office Mojo by IMDbPro

Search for Titles 

IMDbPro |  

Domestic International Worldwide Calendar All Time Showdowns Indices

Overall Weekend Records Daily Records Miscellaneous Records

Top Lifetime Adjusted Grosses

Gross adjustments are explained [here](#).

By Adjusted Gross

2020 Ticket Price ▾

Data as of Jun 5, 2:40 PDT

← Previous page **1-200 of 1,000** Next page →

Rank	Title	Adj. Lifetime Gross	Lifetime Gross	Est. Num Tickets	Year
1	Gone with the Wind	\$1,895,421,694	\$200,852,579	202,286,200	1939
2	Star Wars: Episode IV - A New Hope	\$1,668,979,715	\$460,998,507	178,119,500	1977
3	The Sound of Music	\$1,335,086,324	\$159,287,539	142,485,200	1965
4	E.T. the Extra-Terrestrial	\$1,329,174,791	\$435,110,554	141,854,300	1982
5	Titanic	\$1,270,101,626	\$659,363,944	135,549,800	1997
6	The Ten Commandments	\$1,227,470,000	\$65,500,000	131,000,000	1956
7	Jaws	\$1,200,098,356	\$260,000,000	128,078,800	1975
8	Doctor Zhivago	\$1,163,149,635	\$111,721,910	124,135,500	1965
9	The Exorcist	\$1,036,314,504	\$232,906,145	110,599,200	1973
10	Snow White and the Seven Dwarfs	\$1,021,330,000	\$184,925,486	109,000,000	1937

credit: boxofficemojo.com

From each of the individual movie main pages, the title with year, domestic distributor, earliest release date, MPAA rating, run time, and genres was scraped.

Also the movie ID was scraped from the URL address.

boxofficemojo.com/title/tt0076759/?ref_=bo_cso_table_2

Box Office Mojo by IMDbPro

Search for Titles

Domestic International Worldwide Calendar All Time Showdowns Indices

 Star Wars: Episode IV - A New Hope (1977)

Luke Skywalker joins forces with a Jedi Knight, a cocky pilot, a Wookiee and two droids to save the galaxy from the Empire's world-destroying battle station, while also attempting to rescue Princess Leia from the mysterious Darth Vader.

Title Summary All Releases ▾

All Releases	Domestic Distributor	Twentieth Century Fox See full company information ↗
DOMESTIC (59.4%) \$460,998,507	Domestic Opening	\$1,554,475
INTERNATIONAL (40.6%) \$314,513,557	Budget	\$11,000,000
WORLDWIDE \$775,512,064	Earliest Release Date	May 25, 1977 (Domestic)
	MPAA	PG
	Running Time	2 hr 1 min
	Genres	Action Adventure Fantasy Sci-Fi
	IMDbPro	See more details at IMDbPro ↗

Releases Cast and Crew All-Time Rankings Related Stories Similar Movies

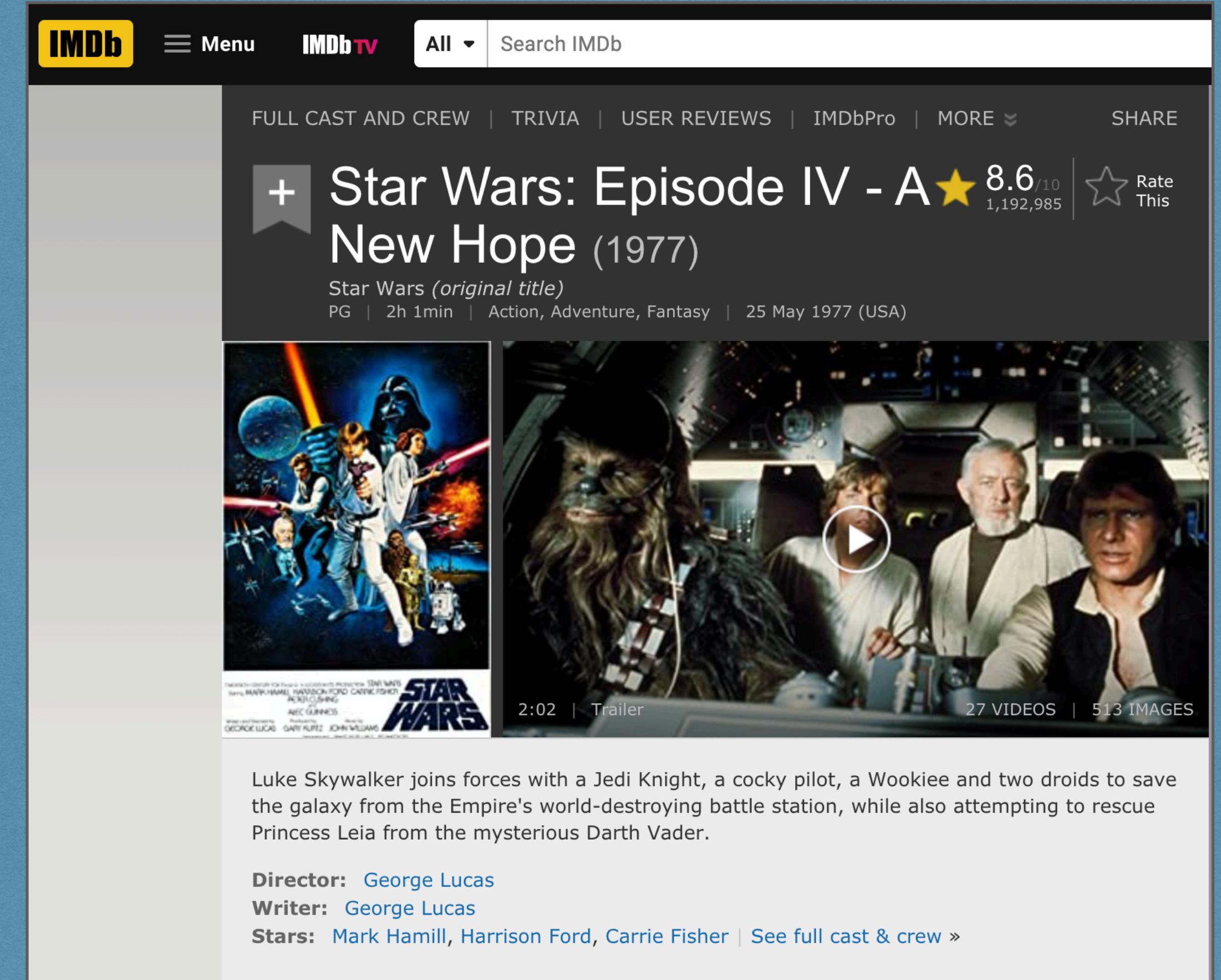
credit: boxofficemojo.com

Within the 'Cast and Crew' tab, all crew members, with their title, and actors, were scraped.

Releases	Cast and Crew	All-Time Rankings	Related Stories	Similar Movies
IMDbPro View contact information for cast and crew ↗				
Crew Member				Role
George Lucas ↗				Director
George Lucas ↗				Writer
Gary Kurtz ↗				Producer
Rick McCallum ↗				Producer
John Williams ↗				Composer
Gilbert Taylor ↗				Cinematographer
Richard Chew ↗				Editor
T.M. Christopher ↗				Editor
Paul Hirsch ↗				Editor
Marcia Lucas ↗				Editor
John Barry ↗				Production Designer
Actor				Character
Mark Hamill ↗				Luke Skywalker
Harrison Ford ↗				Han Solo
Carrie Fisher ↗				Princess Leia Organa
Alec Guinness ↗				Ben Obi-Wan Kenobi

credit: boxofficemojo.com

IMDB was used to scrape individual movie budgets as Box Office Mojo had very limited budget data.



The image shows a screenshot of the IMDb website for the movie "Star Wars: Episode IV - A New Hope". The top navigation bar includes the IMDb logo, a menu icon, "IMDb TV", "All", and a search bar. Below the header, there are links for "FULL CAST AND CREW", "TRIVIA", "USER REVIEWS", "IMDbPro", and "MORE". A "SHARE" button is also present. The main title "Star Wars: Episode IV - A New Hope (1977)" is displayed with a rating of 8.6/10 based on 1,192,985 votes. A "Rate This" button is available. Below the title, it says "Star Wars (original title)", "PG | 2h 1min | Action, Adventure, Fantasy | 25 May 1977 (USA)". To the left is a thumbnail of the movie poster featuring Luke Skywalker, Darth Vader, and other characters. To the right is a video player showing a scene with Chewbacca, Han Solo, and others, with a play button in the center. Below the video player, it says "2:02 | Trailer". At the bottom, there are links for "27 VIDEOS" and "513 IMAGES". A descriptive text block below the trailer says: "Luke Skywalker joins forces with a Jedi Knight, a cocky pilot, a Wookiee and two droids to save the galaxy from the Empire's world-destroying battle station, while also attempting to rescue Princess Leia from the mysterious Darth Vader." Below this, there are sections for "Director: George Lucas", "Writer: George Lucas", and "Stars: Mark Hamill, Harrison Ford, Carrie Fisher | See full cast & crew »".

credit: imdb.com

From each IMDB movie main page, the budget and movie ID was scraped.

This data was joined with the Box Office Mojo data using individual movie IDs.

imdb.com/title/tt0076759/?ref_=fn_al_tt_1

Details

Official Sites: [Official Facebook](#) | [Official site](#)

Country: [USA](#)

Language: [English](#)

Release Date: 25 May 1977 (USA) [See more »](#)

Also Known As: Star Wars: Episode IV - A New Hope [See more »](#)

Filming Locations: [Shed 1, Cardington Airship Hangars, Bedfordshire, England, UK](#) [See more »](#)

Box Office

Budget: \$11,000,000 (estimated)

Opening Weekend USA: \$1,554,475, 30 May 1977

Gross USA: \$460,998,507

Cumulative Worldwide Gross: \$775,512,064

[See more on IMDbPro »](#)

Company Credits

Production Co: [Lucasfilm](#), [Twentieth Century Fox](#) [See more »](#)

[Show more on IMDbPro »](#)

Technical Specs

Runtime: 121 min | 125 min (special edition)

Sound Mix: [70 mm 6-Track](#) (70 mm prints) | [Dolby Stereo](#) (as Dolby System) (35 mm prints) (1977 print) | [DTS-Stereo](#) (as DTS Stereo® in selected theatres) (1997 print) | [Dolby Digital](#) (as Dolby® Digital in selected theatres) (1997 print) | [Dolby Atmos](#) | [SDDS](#) (as Sony Dynamic Digital SoundTM in selected theatres) (1997 print) | [Mono](#) (some 35 mm prints) (other 16 mm prints)

Color: [Color](#) (Technicolor)

Aspect Ratio: 2.39 : 1

[See full technical specs »](#)

credit: imdb.com

**Box Office Mojo listed
franchise data for all domestic
movies, unrelated to the top
1000 gross list.**

Box Office Mojo
by IMDbPro

Search for Titles

IMDbPro | [f](#) [t](#)

Domestic International Worldwide Calendar All Time Showdowns Indices

Brands Franchises Genres

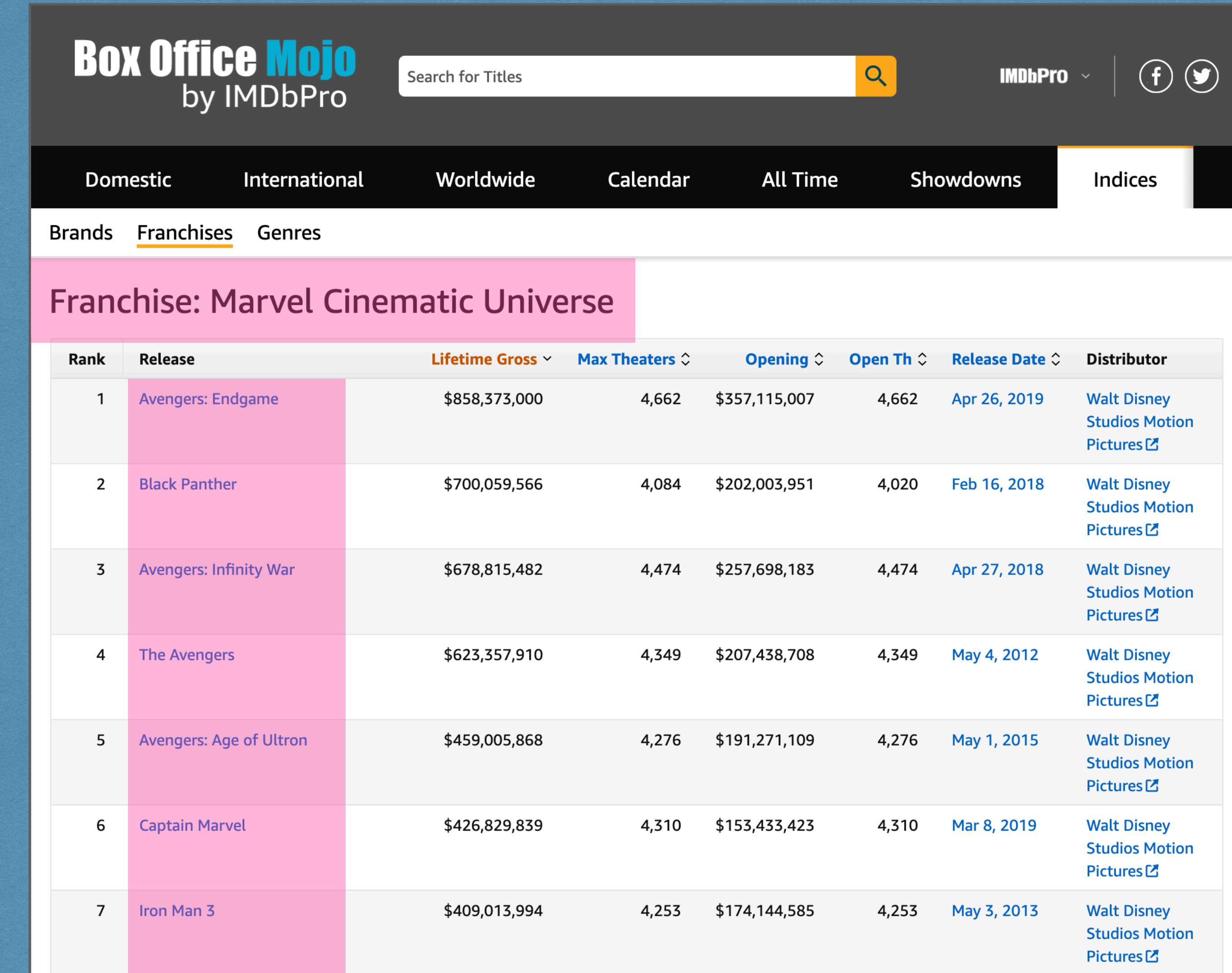
Franchises (US & Canada)

Franchise	Total	Releases	#1 Release	Lifetime Gross
Marvel Cinematic Universe	\$8,545,472,257	23	Avengers: Endgame	\$858,373,000
Star Wars	\$5,079,221,177	20	Star Wars: Episode VII - The Force Awakens	\$936,662,225
Disney Live Action Reimaginings	\$3,407,223,881	15	The Lion King	\$543,638,043
J.K. Rowling's Wizarding World	\$2,786,941,008	26	Harry Potter and the Deathly Hallows: Part 2	\$381,011,219
Avengers	\$2,619,552,260	4	Avengers: Endgame	\$858,373,000
Spider-Man	\$2,496,998,119	9	Spider-Man	\$403,706,375
X-Men	\$2,434,609,697	13	Deadpool	\$363,070,709
Batman	\$2,411,419,947	19	The Dark Knight	\$533,345,358
Harry Potter	\$2,393,347,532	24	Harry Potter and the Deathly Hallows: Part 2	\$381,011,219
DC Extended Universe	\$2,147,685,393	8	Wonder Woman	\$412,563,408
James Bond	\$2,112,905,402	25	Skyfall	\$304,360,277
Jurassic Park	\$1,883,077,059	7	Jurassic World	\$652,270,625
Middle Earth	\$1,852,432,231	11	The Lord of the Rings: The Return of the King	\$377,027,325
The Fast and the Furious	\$1,690,757,371	9	Furious 7	\$353,007,020
Transformers	\$1,582,401,601	7	Transformers: Revenge of the Fallen	\$402,111,870
Pirates of the Caribbean	\$1,451,780,833	5	Pirates of the Caribbean: Dead Man's Chest	\$423,315,812

credit: boxofficemojo.com

The franchise name and all movie titles within that franchise were scraped.

This data was then joined using movie titles.

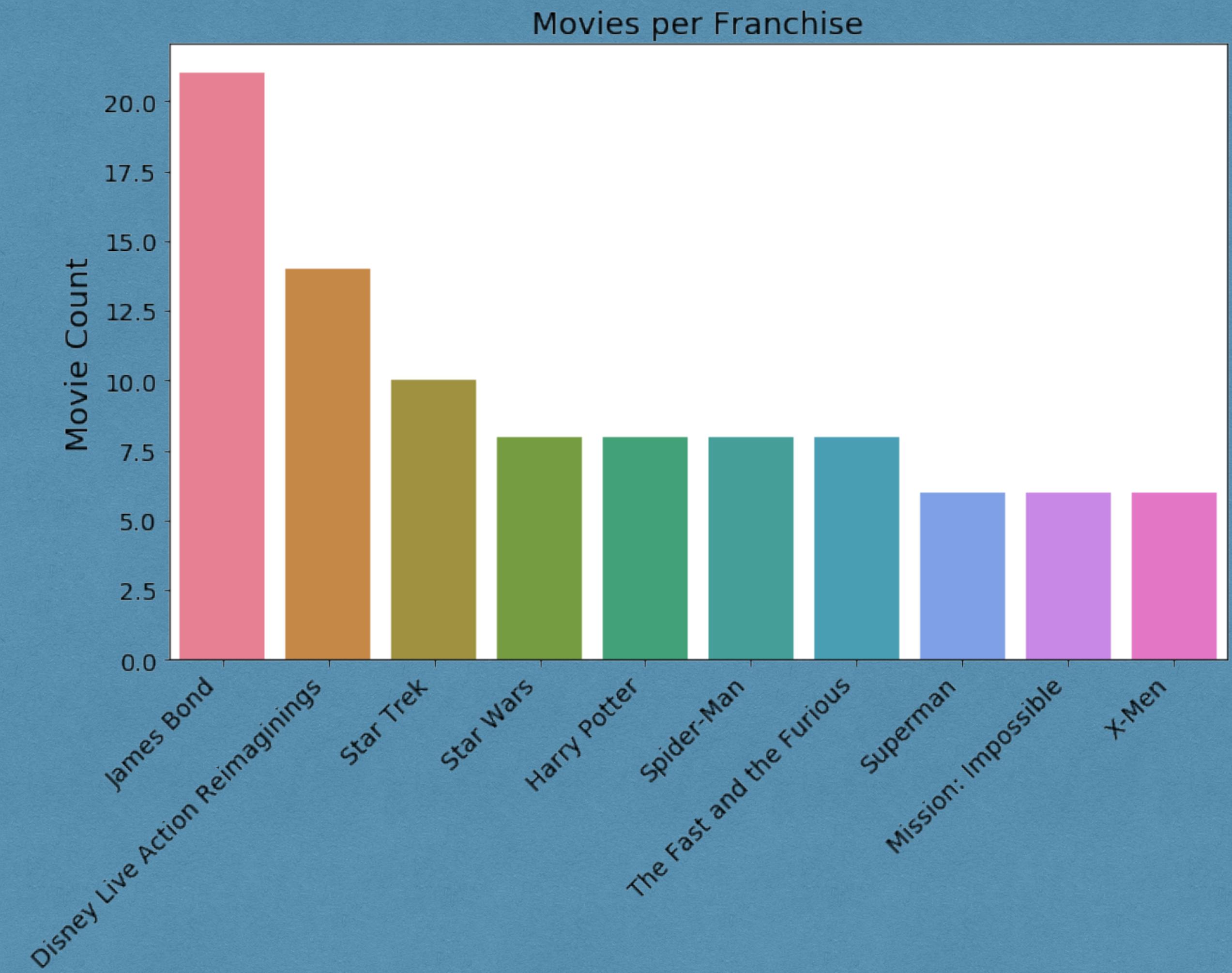


The screenshot shows the Box Office Mojo website interface. At the top, there is a search bar with the placeholder "Search for Titles" and a magnifying glass icon. To the right of the search bar are links for "IMDbPro" and social media icons for Facebook and Twitter. Below the header, there is a navigation menu with tabs: Domestic, International, Worldwide, Calendar, All Time, Showdowns, and Indices. The "Franchises" tab is currently selected, indicated by an orange underline. Below the menu, there are three more tabs: Brands, Franchises (underlined), and Genres. A pink banner across the page reads "Franchise: Marvel Cinematic Universe". The main content area displays a table of movie data for the franchise. The table has columns for Rank, Release, Lifetime Gross, Max Theaters, Opening, Open Th, Release Date, and Distributor. The data shows the top seven movies from the franchise, all released by Walt Disney Studios Motion Pictures. The table is as follows:

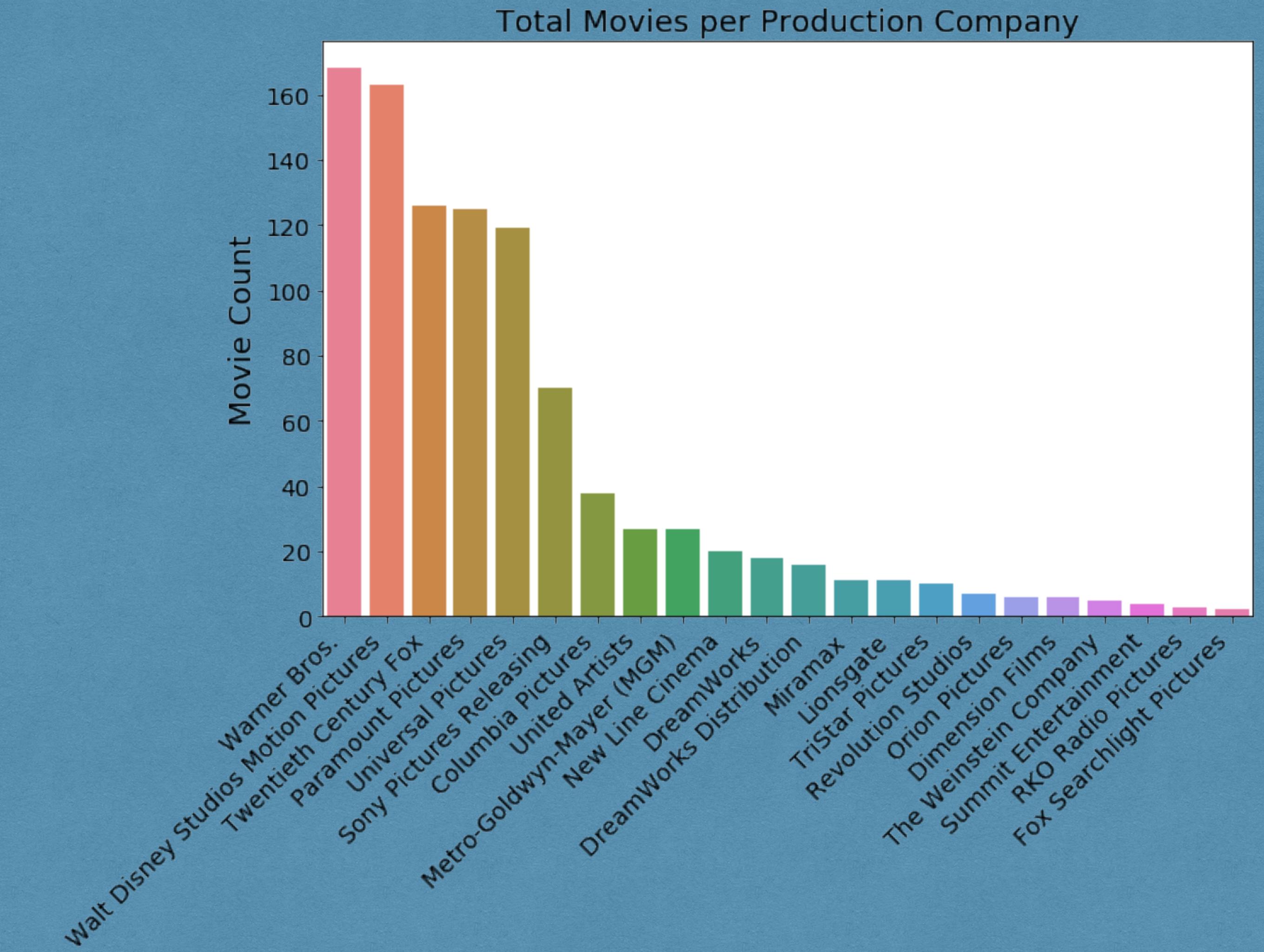
Rank	Release	Lifetime Gross	Max Theaters	Opening	Open Th	Release Date	Distributor
1	Avengers: Endgame	\$858,373,000	4,662	\$357,115,007	4,662	Apr 26, 2019	Walt Disney Studios Motion Pictures
2	Black Panther	\$700,059,566	4,084	\$202,003,951	4,020	Feb 16, 2018	Walt Disney Studios Motion Pictures
3	Avengers: Infinity War	\$678,815,482	4,474	\$257,698,183	4,474	Apr 27, 2018	Walt Disney Studios Motion Pictures
4	The Avengers	\$623,357,910	4,349	\$207,438,708	4,349	May 4, 2012	Walt Disney Studios Motion Pictures
5	Avengers: Age of Ultron	\$459,005,868	4,276	\$191,271,109	4,276	May 1, 2015	Walt Disney Studios Motion Pictures
6	Captain Marvel	\$426,829,839	4,310	\$153,433,423	4,310	Mar 8, 2019	Walt Disney Studios Motion Pictures
7	Iron Man 3	\$409,013,994	4,253	\$174,144,585	4,253	May 3, 2013	Walt Disney Studios Motion Pictures

credit: boxofficemojo.com

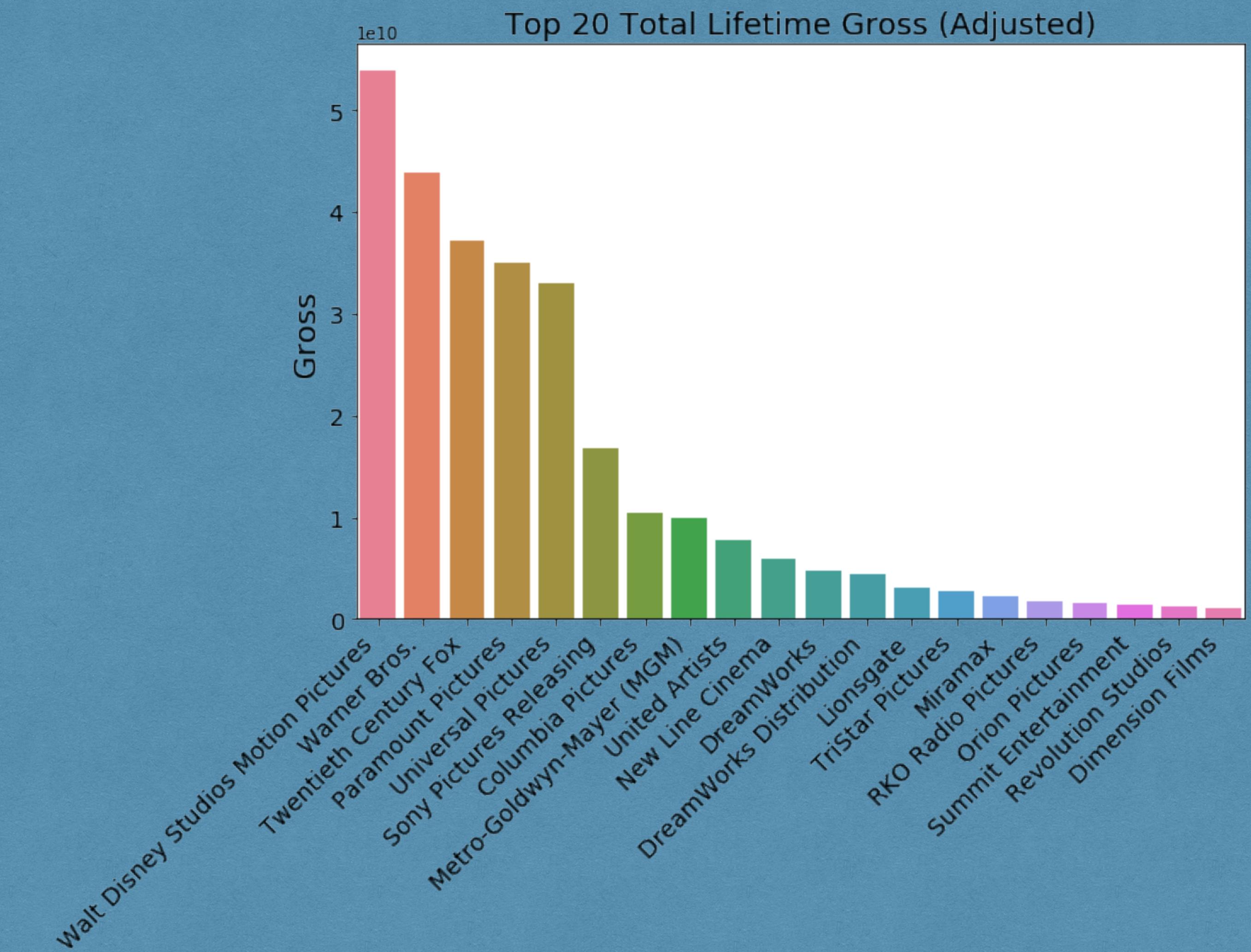
43% of the movies within the data set were listed as part of a franchise



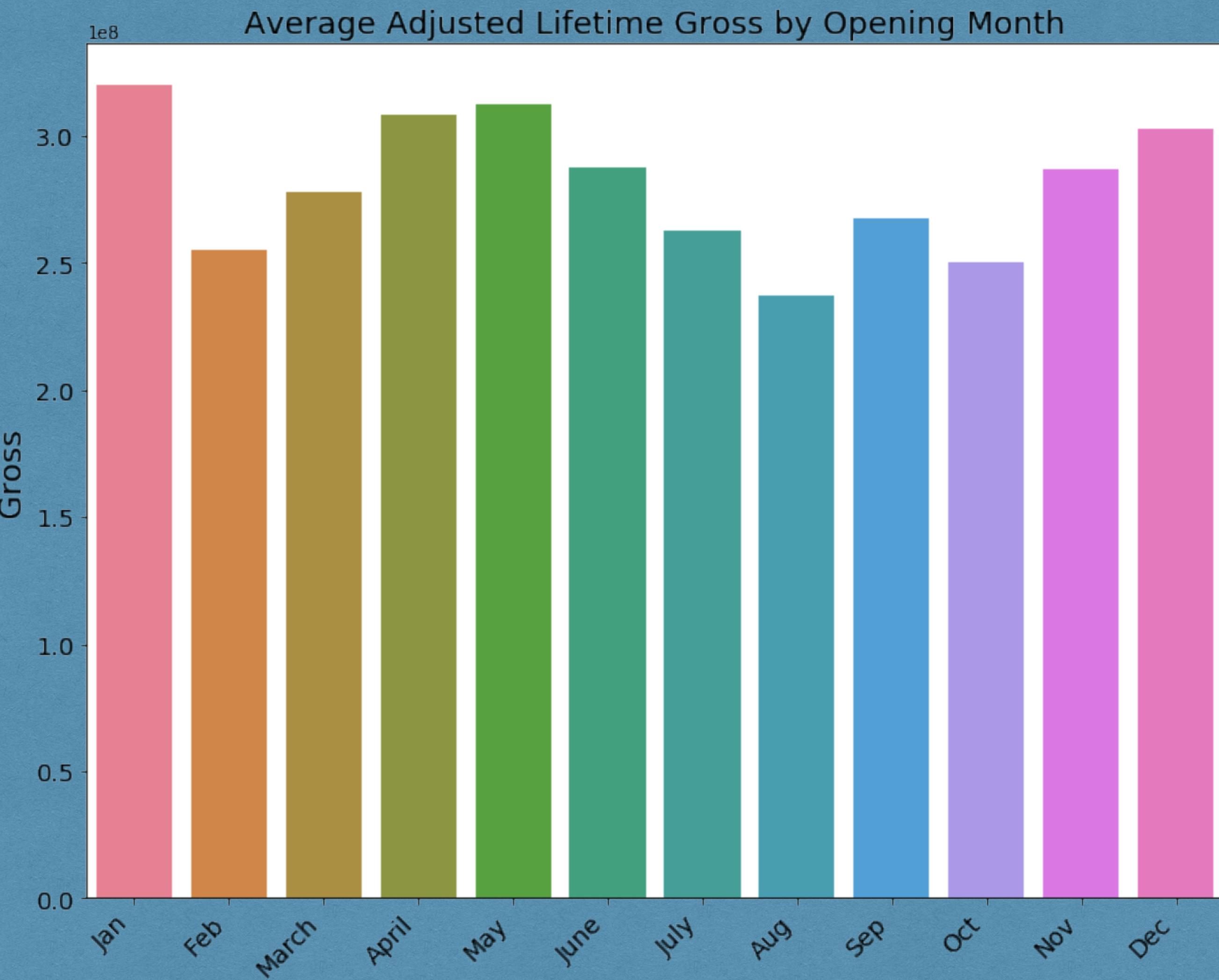
**Only 5 production companies
account for 70% of the movies
in the top 1000 list**



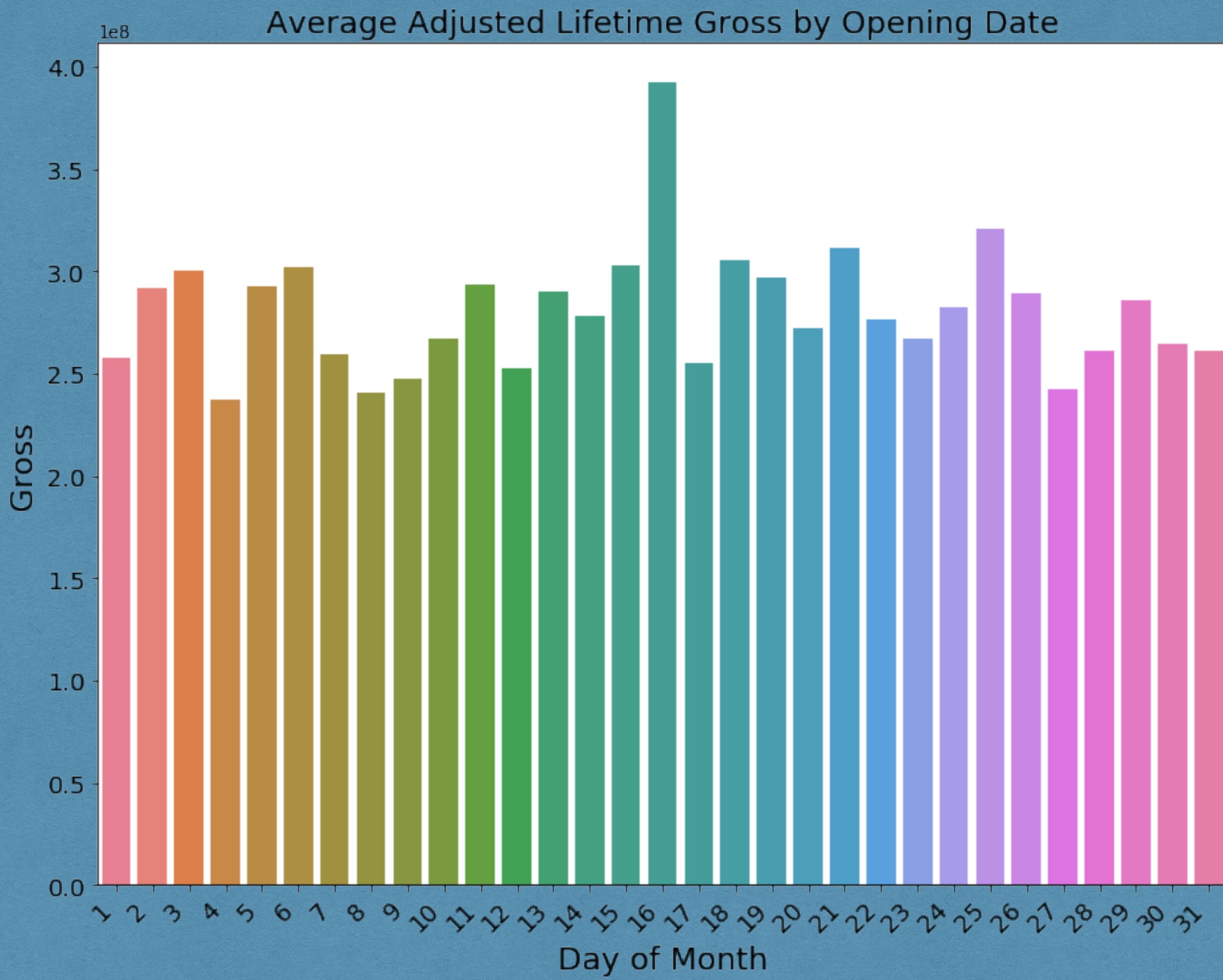
**Some production companies
have since been dissolved,
absorbed, or relabelled**



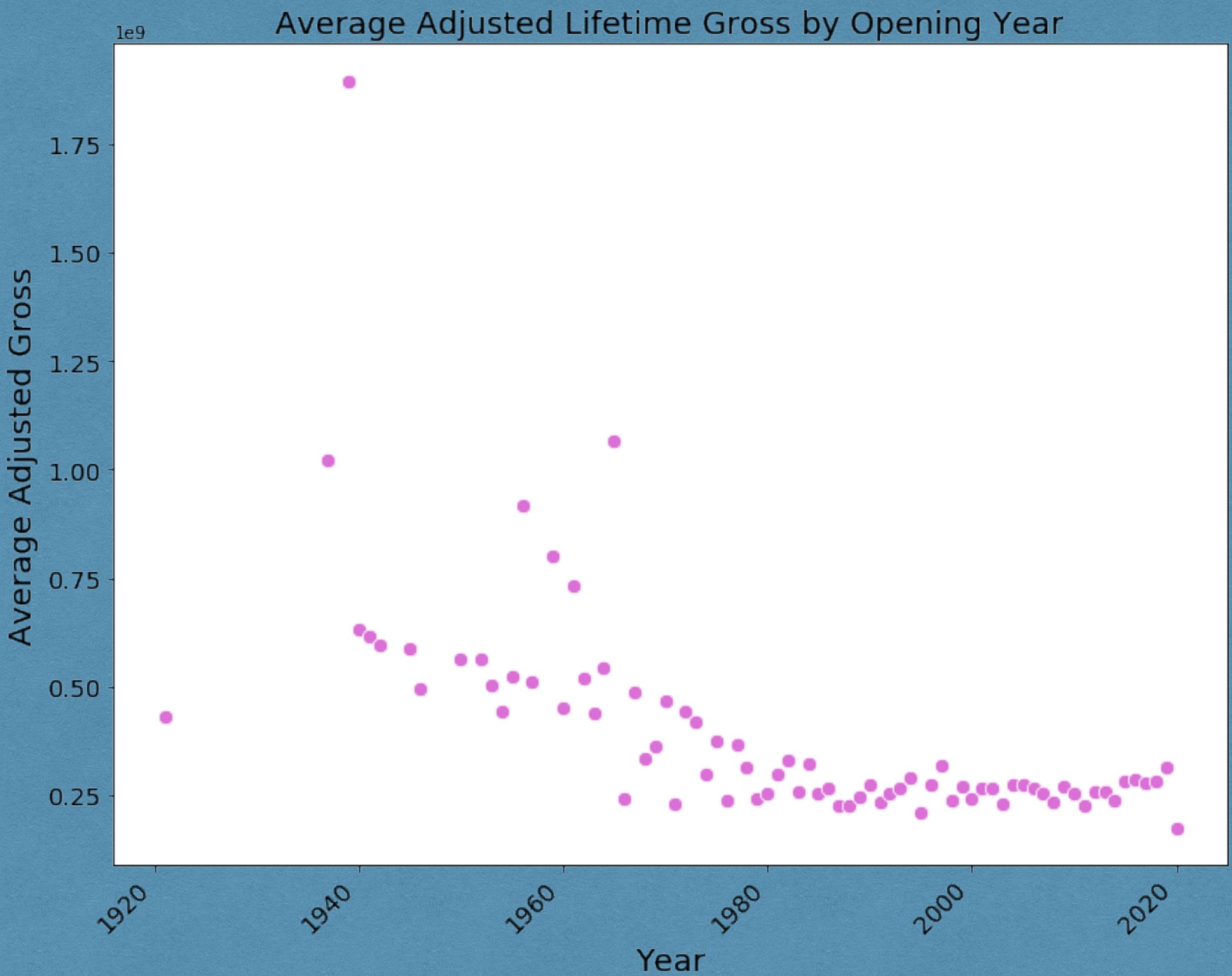
There seems to be a seasonal trend for the month in which movies are released and the average gross



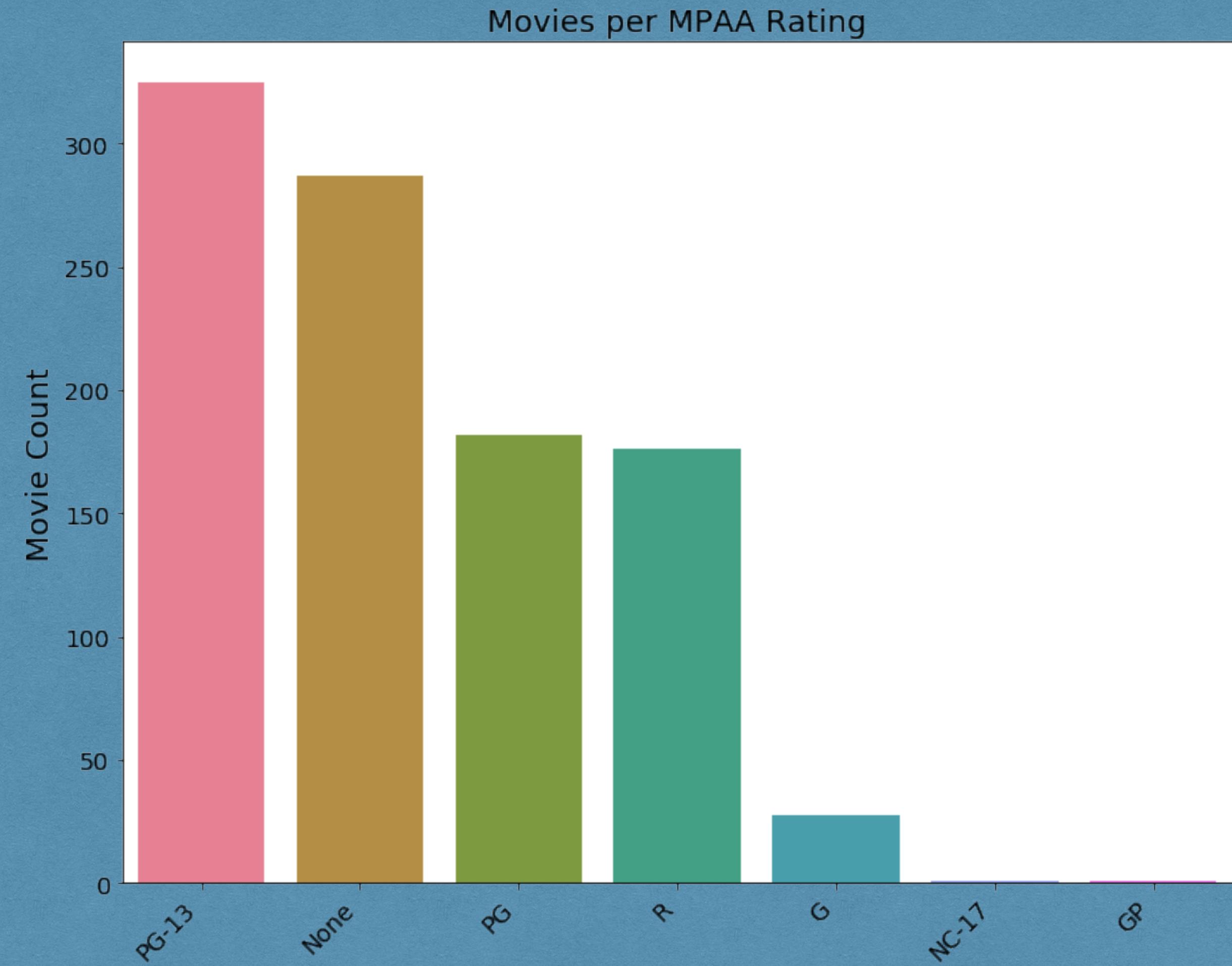
There does not seem to be a correlation between the day of the month the movie is released and the average gross



There does seem to be a downward trend between year released and average gross



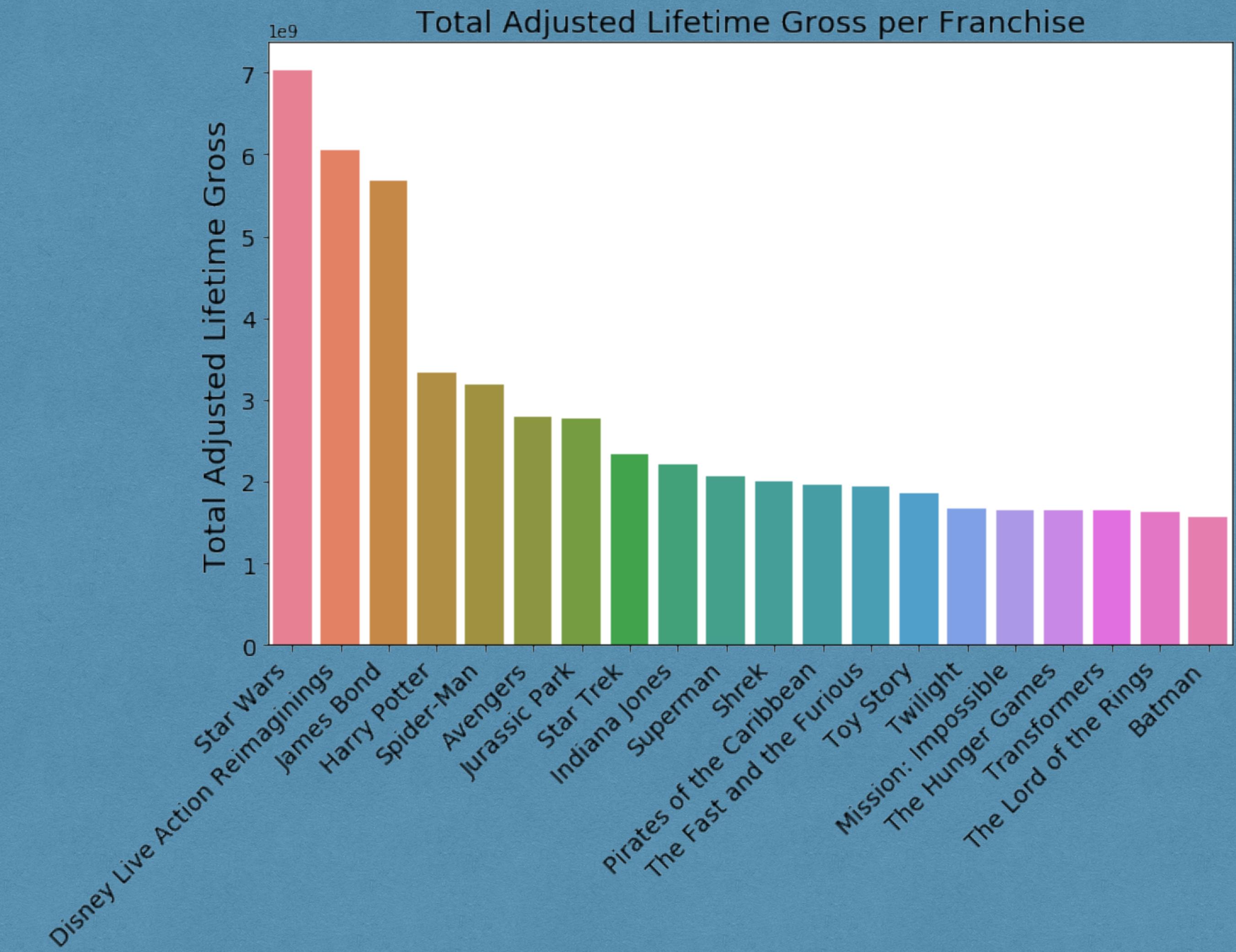
Motion Picture Association of America film rating system which suggests movie suitability for different age groups



A number of regressor models were evaluated and the top performing model predicts adjusted lifetime gross by plus/minus 31%.

An important feature in correctly predicting the gross, was franchise. This was explored further.

The top 20 total gross franchises were added to the model, classifying 136 of the movies within the top 1000.



With 20 additional franchise features added to the model, the model improved to plus/minus 30% in predicting the correct adjusted lifetime gross.

Future work could include building a model on a larger set of movies, possibly improving the model predictions.

Sources:

IMDB.com

boxofficemojo.com

statista.com

Thank You