

Movies worldwide box-office gross

Can we predict it



Roberto Linares

The problem:

How to choose what
movies rights to
acquire?

Solution:

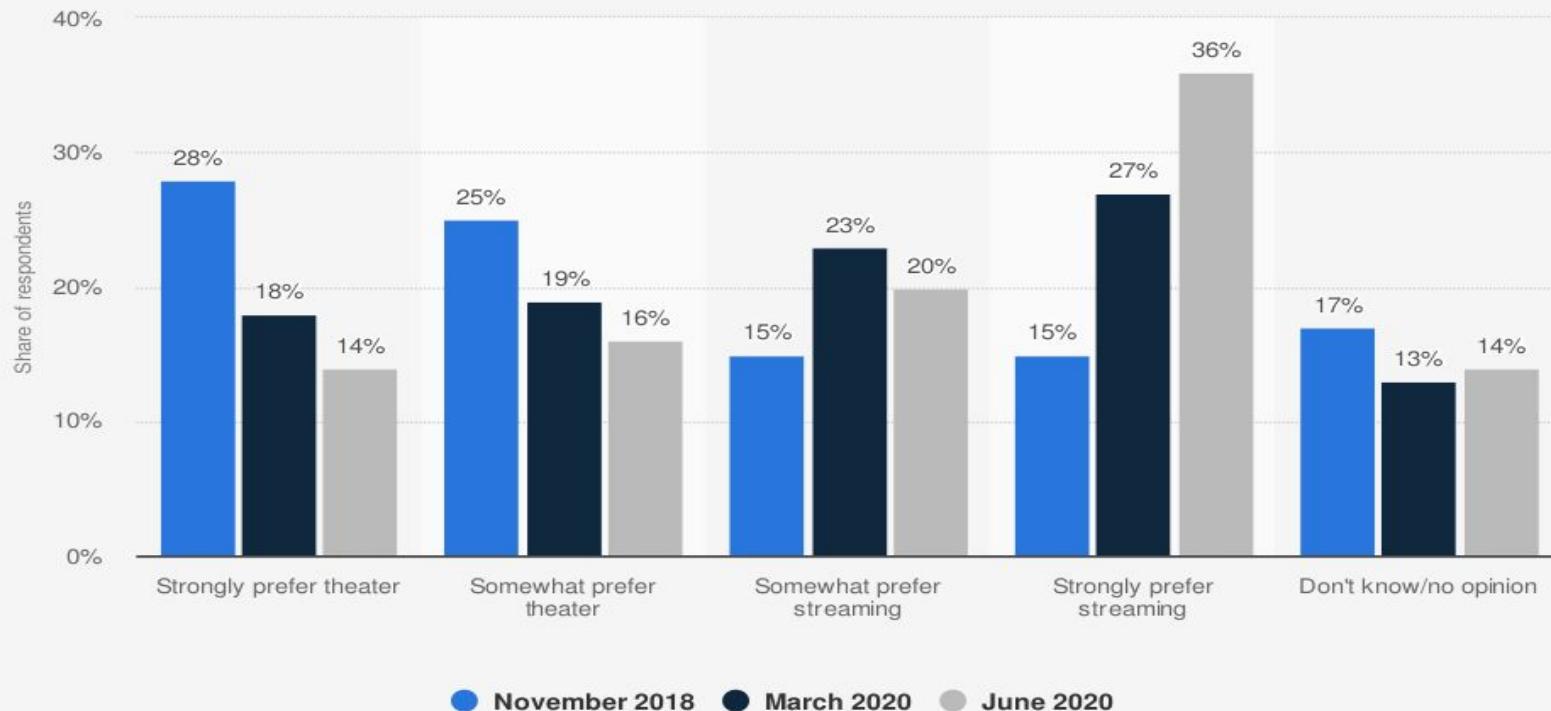
Let's unleash the
power of data to help
us find out.

NETFLIX

VS



Preference for watching a movie for the first time at a theater instead of via a streaming service in the United States from November 2018 to June 2020



The data

Box Office Mojo
by IMDbPro

Search for Titles

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

Domestic International Worldwide Calendar All Time Showdowns Indices

Star Wars: Episode VII - The Force Awakens (2015)
As a new threat to the galaxy rises, Rey, a desert scavenger, and Finn, an ex-stormtrooper, must join Han Solo and Chewbacca to search for the one hope of restoring peace.

[Title Summary](#) [All Releases](#)

All Releases	Domestic Distributor	Walt Disney Studios Motion Pictures See full company information
DOMESTIC (45.3%) \$936,662,225	Domestic Opening	\$247,966,675
INTERNATIONAL (54.7%) \$1,132,859,475	Budget	\$245,000,000
WORLDWIDE \$2,069,521,700	Earliest Release Date	December 16, 2015 (EMEA, APAC)
	MPAA	PG-13
	Running Time	2 hr 18 min
	Genres	Action Adventure Sci-Fi
	IMDbPro	See more details at IMDbPro



Star Wars: The Force Awakens

Theatrical release poster

Directed by J. J. Abrams
Written by Lawrence Kasdan
J. J. Abrams
Michael Arndt

Cruella

Theatrical release poster

Directed by Craig Gillespie
Screenplay by Dana Fox
Tony McNamara

How it looks like

Title	Domestic opening	Domestic total gross	Worldwide Gross	Budget	Runtime	Rating	Release date	Distributor	Genre
Avatar	77025481	760507625	2847246203.00	237000000.00	162.00	PG-13	December 16, 2009	Twentieth Century Fox	Action Adventure Fantasy...
Avengers: Endgame	357115007	858373000	2797501328.00	356000000.00	181.00	PG-13	April 24, 2019	Walt Disney Studios Motion Pictures	Action Adventure Drama ...

Models

- Linear Regression
- Ridge

Metrics

R^2

Comparison

MAE

For results

LASSO



Results

$R^2 = 0.44$

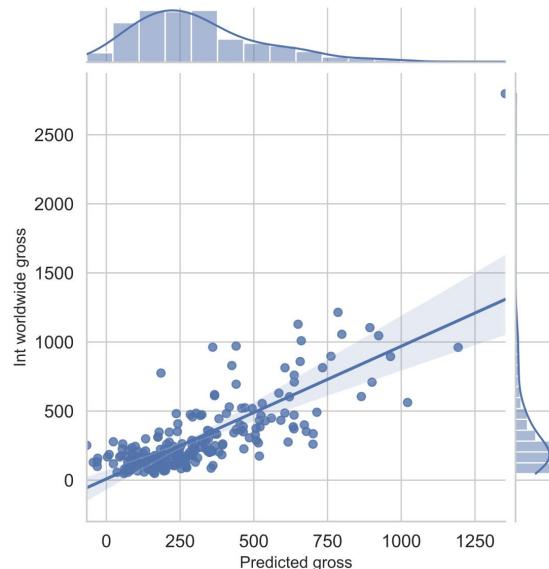
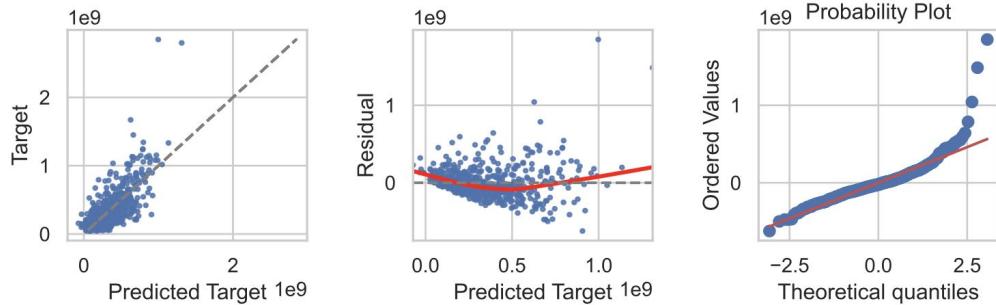
MAE = \$156 million

Lasso

$R^2 = 0.54$

MAE = \$130 million

LinearRegression approach



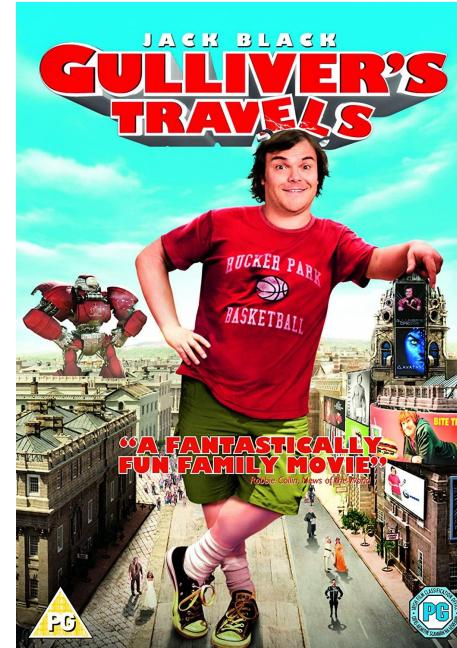
Model predictions



Pred = \$762 million

Real = \$896 million

Diff = \$134 million



Pred = \$318 million

Real = \$237 million

Diff = \$81 million

Key takeaways

- Budget is the feature most related with the predicted goal.
- PG-13 rated films will outperform the rest.
- Most profitable genres are:
 - Animation, adventure, family, music/musical and last but not least, horror 😱.

Future Work

- Increase the amount of data and features (consider director and lead cast).
- Include the factor of the seasonality of movies.
- Consider the trend of streaming/theaters releases.