

Can Rom-Coms Save America?



Project Proposal

Prepared for: Metis Staff Instructors: Leon Johnson, Kevin Birnbaum

Prepared by: Rahul D. Raju

June 30, 2021

EXECUTIVE SUMMARY

Objective

The world has begun its return to normalcy from a dreary, prolonged lockdown phase. Appetite for a resemblance of life prior to COVID is strong. As such the Hollywood has an opportunity to fill an incredible amount of pent-up demand for America's favorite past-time, going to the movies. As such, it is believed that movies that are light-hearted and seek to serve as a distraction from the memory of the past year would potentially be very welcomed and profitable. The pitfall Hollywood faces however, is to make not only a great movie but avoid the losses that have plagued the industry in the past. Particularly with the stock marked at all time highs, movie industry stocks could be susceptible to anything that diminishes the bottom line. The table below represent the biggest box office losses of all all time, ranked in descending order.

Biggest box office bombs

Title	Year	Net production budget (millions)	Worldwide gross (millions)	Estimated loss (millions)		Ref. [nb 1]
				Nominal	Adjusted for inflation	
<i>John Carter</i>	2012	\$263.7	\$284.1	\$114–200	\$129–225	[# 60]
<i>The Lone Ranger</i>	2013	\$225–250	\$260.5	\$160–190	\$178–211	[# 67]
<i>Mortal Engines</i>	2018	\$110	\$83.7	\$174.8	\$180	[# 73]
<i>King Arthur: Legend of the Sword</i>	2017	\$175	\$148.7	\$114–153.2	\$120–162	[# 63]
<i>Battleship</i>	2012	\$209–220	\$303	\$150	\$169	[# 14]
<i>Tomorrowland</i>	2015	\$180–190	\$209	\$90–150	\$98–164	[# 103]
<i>Pan</i>	2015	\$150	\$128.4	\$85–150	\$93–164	[# 79]
<i>Mulan</i>	2020	\$200	\$66.8	\$147	\$147	[# 74]
<i>Mars Needs Moms</i>	2011	\$150	\$39	\$100–144	\$115–166	[# 69]
<i>Onward</i>	2020	\$175–200	\$142	\$135	\$135	[# 77]
<i>Dark Phoenix</i>	2019	\$200	\$252.4	\$79–133	\$81–137	[# 27]
<i>A Wrinkle in Time</i>	2018	\$125	\$133.4	\$130.6	\$135	[# 111]
<i>Terminator: Dark Fate</i>	2019	\$185–196	\$261.1	\$110–130	\$113–134	[# 101]
<i>The 13th Warrior</i>	1999	\$100–160	\$61.7	\$69–129	\$107–200	[# 1]
<i>Sinbad: Legend of the Seven Seas</i>	2003	\$60	\$80.8	\$125	\$176	[# 92]

Solution

With so much money at risk, a much more scientific approach is needed with respect to idea generation and movie development. One method we hope to incorporate is the use of linear regression as a potential tool for predicting world wide box office grosses. A model can be used to assess the “rom-com hypothesis” by analyzing various features of past box office successes and failures. Movie concepts and pitches could be adjusted to reflect the features of the historically most profitable rom-com movies.

Project Data

The primary source of data will be attained by web-scraping information from boxofficemojo.com. Each row of data will reflect all of the movie's financial, production and scoring information. Each movie will also be supplemented by additional data from Instagram, Rotten Tomatoes, and other information sources. This supplemental data will include information about features within our analysis, such as Facebook likes for directors or number of instagram followers for lead actors.

Project Tools

- 1). Web Scraping and Parsing tools: BeautifulSoup, Selenium
- 2). Data Cleaning & Analysis: Pandas
- 3). Data Visualization: MatPlotLib, Seaborn
- 4). Statistical Analysis: Sci-Kit Learn, StatsModels

Minimum Viable Product Vision

An MVP for this project would consist of the following:

1. A list of all the feature(dependent) variables to be used in this analysis
2. A basic pair plot showing, amongst other things, if a linear relationship exists between our feature variables and our target variable or amongst the features themselves
3. A visual of the distribution of the world wide gross revenues of the movies included in the analysis