THE UNIVERSITY
OF TEXAS AT DALLAS

The Correlation Between Google Trends Drop Rate and IMDb and Rotten Tomatoes Rating

A Data-Driven Analysis of Movie Popularity and Viewer Reception

Presented by: Oliver Myers, Calvin Hanebeck, Allen Hernandez & Lynn Kuhlwein

EPPS 6302

Overview

- **Research Question:**

  Is there a correlation between the drop rate of Google Trends searches (Day 1 to Day 21) and their IMDb or Rotten Tomatoes rating?

- **Purpose:**

  To understand the relationship between digital search interest and audience reception (IMDb and Rotten Tomatoes ratings).

Literature Review - Google Trends in Scientific Research

- **Benefits:**
  - Easily accessible
  - Massive dataset (since 2004)
  - No usage cost
  - Daily data updates/real-time insights
  - “Digital truth serum”
- **Limitations:**
  - Requires high internet penetration and reliance on Google as primary search instrument
  - No absolute search volume
  - Use of sampled data
  - Search term ambiguity (Apple & *apple*)

Literature Review - Google Trends in Scientific Research

- Google Trends is often used due to its predictive power to enhance **forecasting** models
- Applies in a variety of disciplines:
  - Forecasting the emergence of financial crises (Petropoulos et al., 2022)
  - Emergence of epidemics through infectious disease spread (Carneiro and Mylonakis, 2009)
  - Employment growth and labor market conditions (Borup and Schütte, 2022)
  - Consumption behavior (Woo and Owen, 2018)
- Only one study examining GT-IMDb rating nexus (Demir et al., 2012)
  - Tested prediction accuracy of search terms for movies + classifying movies into high/low rating
  - Overall, an unreliable study (short, no journal publication, only two sources)
  - Other studies: movie ticket sales, cinema admissions

Methodology (Data Collection and Production)

- **Data Collection:**
  - Movies list scrapped (rvest) from Box Office Mojo
    i. Titles, Theaters
  - Movie metadata from OMDb API
    i. IMDb_Votes, Release_date, IMDb_Rating, RottenTomatoes
  - Daily earnings from Box Office Mojo, using rvest package in R
    i. Day, Daily_Earnings
  - Google Trends data for Day 1 and Day 21 post-release.
    i. Hits
- **Tools Used:**
  - R packages: rvest, gtrendsR, and httr (OMDb API).

Methodology (Data Collection and Production)

Methodology (Data Collection and Production)

- **R-Studio Code:** https://oliverjackmyers.github.io/EPPS6302_finalproject_code.pdf

BoxOfficeMojo Google Trends OMDb

	Title	date	DOW	Day	Daily_Earning	gt_hits	gt_geo	gt_keyword	imdbID	releaseID	Max_Theaters	BoxOfficeTotal	imdbVotes	imdbRating	Metascore	RottenTomatoes_Rating
1	Bones and All	2022-11-23	Wednesday	1	$921,168	97	US	Bones and All	tt10168670	rl734495489	3754	$7,834,907	63370	6.8	74	82%
2	Bones and All	2022-11-24	Thursday	2	$436,609	98	US	Bones and All	tt10168670	rl734495489	3754	$7,834,907	63370	6.8	74	82%
3	Bones and All	2022-11-25	Friday	3	$834,421	97	US	Bones and All	tt10168670	rl734495489	3754	$7,834,907	63370	6.8	74	82%
4	Bones and All	2022-11-26	Saturday	4	$856,894	100	US	Bones and All	tt10168670	rl734495489	3754	$7,834,907	63370	6.8	74	82%
5	Bones and All	2022-11-27	Sunday	5	$567,247	81	US	Bones and All	tt10168670	rl734495489	3754	$7,834,907	63370	6.8	74	82%
6	Bones and All	2022-11-28	Monday	6	$257,304	50	US	Bones and All	tt10168670	rl734495489	3754	$7,834,907	63370	6.8	74	82%
7	Bones and All	2022-11-29	Tuesday	7	$323,704	37	US	Bones and All	tt10168670	rl734495489	3754	$7,834,907	63370	6.8	74	82%
8	Bones and All	2022-11-30	Wednesday	8	$259,241	35	US	Bones and All	tt10168670	rl734495489	3754	$7,834,907	63370	6.8	74	82%
9	Bones and All	2022-12-01	Thursday	9	$246,386	34	US	Bones and All	tt10168670	rl734495489	3754	$7,834,907	63370	6.8	74	82%
10	Bones and All	2022-12-02	Friday	10	$383,099	34	US	Bones and All	tt10168670	rl734495489	3754	$7,834,907	63370	6.8	74	82%
11	Bones and All	2022-12-03	Saturday	11	$488,889	48	US	Bones and All	tt10168670	rl734495489	3754	$7,834,907	63370	6.8	74	82%
12	Bones and All	2022-12-04	Sunday	12	$316,710	46	US	Bones and All	tt10168670	rl734495489	3754	$7,834,907	63370	6.8	74	82%
13	Bones and All	2022-12-05	Monday	13	$155,194	28	US	Bones and All	tt10168670	rl734495489	3754	$7,834,907	63370	6.8	74	82%
14	Bones and All	2022-12-06	Tuesday	14	$193,258	25	US	Bones and All	tt10168670	rl734495489	3754	$7,834,907	63370	6.8	74	82%
15	Bones and All	2022-12-07	Wednesday	15	$161,254	27	US	Bones and All	tt10168670	rl734495489	3754	$7,834,907	63370	6.8	74	82%
16	Bones and All	2022-12-08	Thursday	16	$155,367	21	US	Bones and All	tt10168670	rl734495489	3754	$7,834,907	63370	6.8	74	82%
17	Bones and All	2022-12-09	Friday	17	$188,304	25	US	Bones and All	tt10168670	rl734495489	3754	$7,834,907	63370	6.8	74	82%
18	Bones and All	2022-12-10	Saturday	18	$228,428	31	US	Bones and All	tt10168670	rl734495489	3754	$7,834,907	63370	6.8	74	82%
19	Bones and All	2022-12-11	Sunday	19	$166,147	29	US	Bones and All	tt10168670	rl734495489	3754	$7,834,907	63370	6.8	74	82%
20	Bones and All	2022-12-12	Monday	20	$95,167	22	US	Bones and All	tt10168670	rl734495489	3754	$7,834,907	63370	6.8	74	82%

Key Findings (Data Collection and Production)

- **Insight:**
  - Unique identifiers (e.g., imdbID, releaseID) are crucial for linking records and avoiding data loss caused by inconsistent movie titles.
- **Challenges:**
  - Scraping IMDb directly with rvest caused issues, so we switched to BoxOfficeMojo (a part of IMDb Pro).
  - The pandemic skewed data as many 2022 movies were released on streaming and excluded from the dataset.
  - Missing rating values led to the deletion of numerous movies.

Methodology (Data Analysis)

- **Tools Used:**
  - Data analysis: Stata for regression modeling
- **Key Variables:**
  - Dependent variables:
    i. Rotten Tomatoes Rating
    ii. IMDb ratings
  - Key explanatory variable:
    i. Drop rate (% decline in search interest between release date and 21 days later)
    drop_rate = (hits_day1 - hits_day21) / hits_day1 * 100

Methodology (Data Analysis)

- **Key Variables:**
  - Control variables:
    i. Runtime in mins
    ii. Box office revenue in $
    iii. Number of votes on IMDb
- **2 Models:**

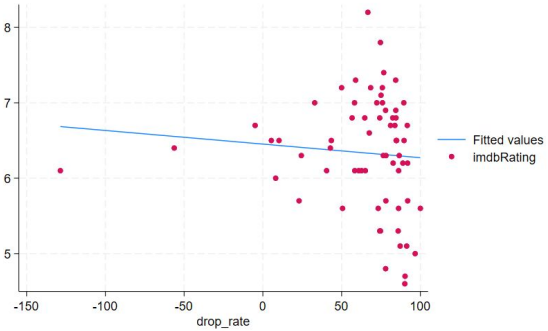
$$imdbrating_i = \beta_0 + \beta_1 * drop\_rate_i + \beta_2 * runtime_i + \beta_3 * boxoffice_i + \beta_4 * imdbvotes_i + \mu_i$$

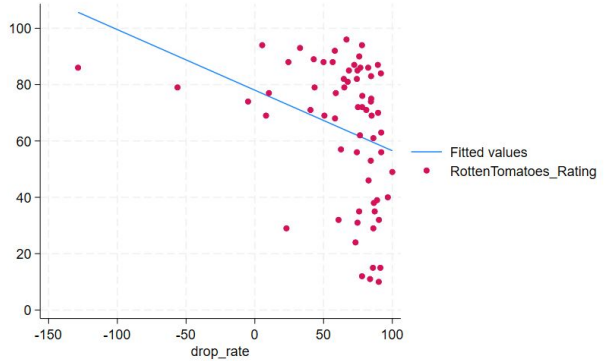
$$rottentomatoes\_rating_i = \beta_0 + \beta_1 * drop\_rate_i + \beta_2 * runtime_i + \beta_3 * boxoffice_i + \mu_i$$

- **Final Data:** 65 movies

Key Findings

- **Results:**





  - Effect of the drop rate on the Rotten Tomatoes ranking is statistically significant: -0.2481
  - Lower drop rate → associated with higher Rotten Tomatoes rating
  - No significant effect on the IMDb rating
- **Challenges:**
  - Outliers and data limitations (e.g., incomplete data)

Discussion

- **Limitations:**
  - Dependence on Google Trends data (relative, not absolute measures).
  - Sample restricted to a single year.
  - Combining diverse datasets (Google Trends, IMDb, Box Office Mojo) required extensive cleaning and introduced constraints due to missing or incomplete data.

- **Future Directions:**
  - Incorporate machine learning for better predictions.
  - Analyze additional factors: social media sentiment, reviews, demographics
  - Expanding the scope to include global search trends and data from multiple years would improve generalizability.

Conclusion

Key Observations and Platform Differences:

The drop rate in Google Trends search interest showed no significant correlation with IMDb ratings but was significantly correlated with Rotten Tomatoes ratings.

- IMDb is more influenced by the number of votes, which skews its ratings.
- Rotten Tomatoes incorporates a broader mix of audience and critic opinions, making it more reflective of trends in search interest.

Summary:

- We explored the drop rate of Google Trends searches and movie ratings, finding a weak IMDb correlation but a **significant negative link with Rotten Tomatoes**, highlighting platform-specific audience dynamics.
- Yes, there is a correlation, but it **varies depending on the platform**.

References

Wickham H (2024). *rvest: Easily Harvest (Scrape) Web Pages*. R package version 1.0.4, https://github.com/tidyverse/rvest, https://rvest.tidyverse.org/.

OMDb API. (n.d.). *The Open Movie Database*. Retrieved [November 29], from https://www.omdbapi.com

OpenAI. (2024). *ChatGPT (November 30 version)*. Retrieved from https://openai.com/chatgpt

Box Office Mojo. (n.d.). *Yearly box office results*. Retrieved [date], from https://www.boxofficemojo.com

Google Trends API Massicotte P, Eddelbuettel D (2022). _gtrendsR: Perform and Display Google Trends Queries_. R package version 1.5.1, https://CRAN.R-project.org/package=gtrendsR