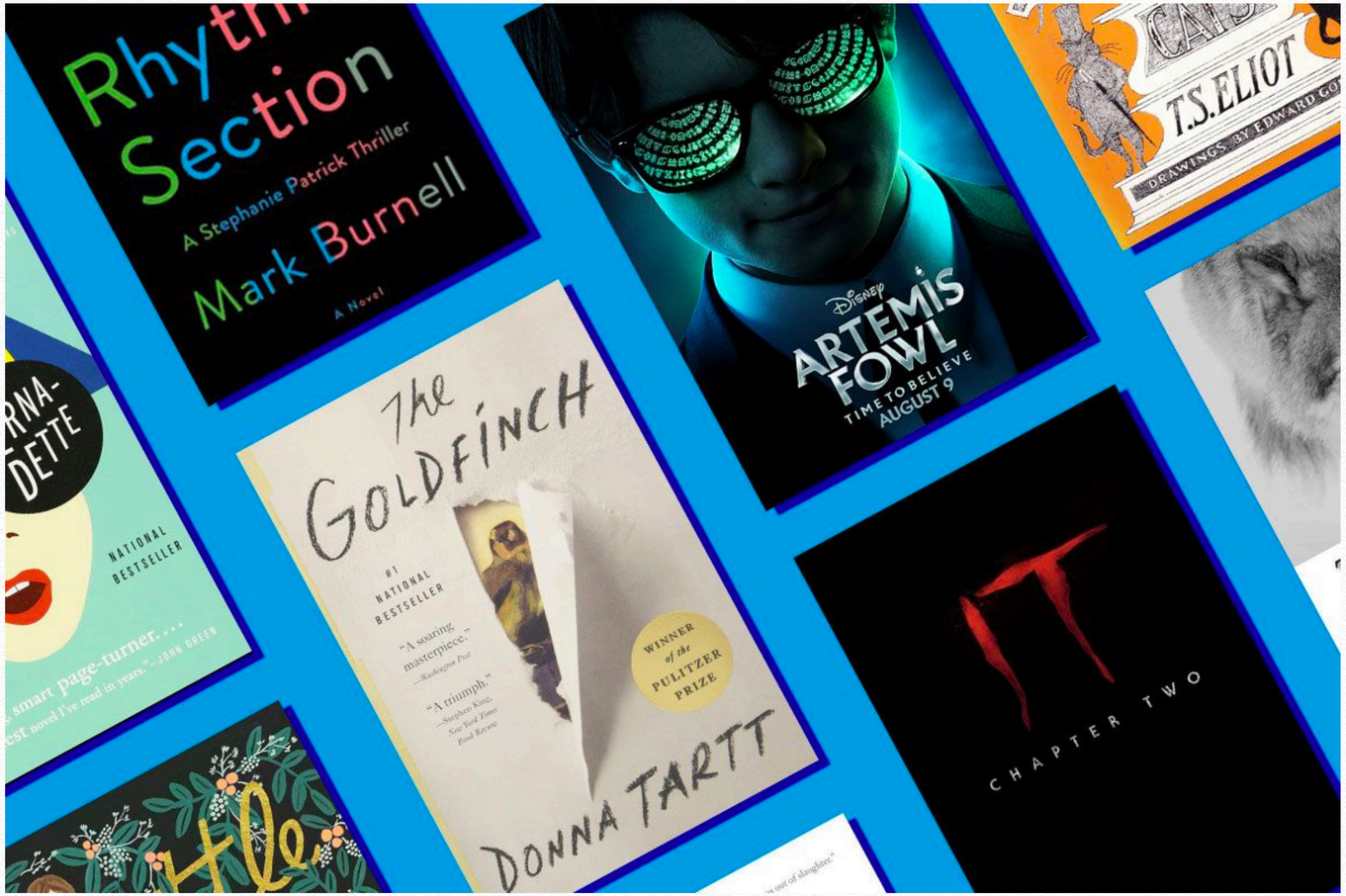
Adapt to succeed?

Revenue prediction of movies based on books.

Katie Huang

January 22nd, 2021

Fulfilled or disappointed?



Are we set to succeed?

Forbes

Jul 11, 2018, 08:44am EDT

Report: Film Adaptations Of Books Earn 53% More At The Worldwide Box Office

Adam Rowe Contributor

Media

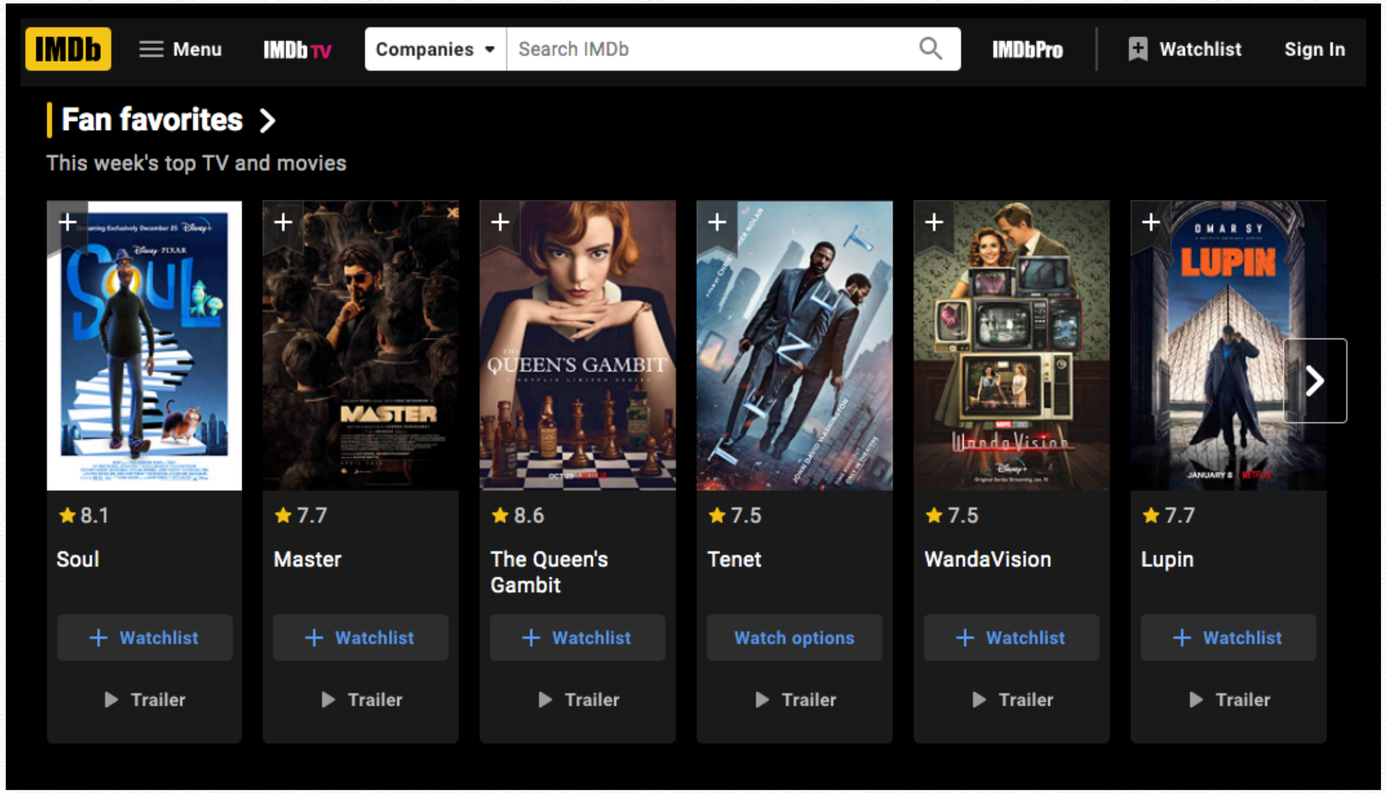
I write about the future of books and the business of storytelling.

This article is more than 2 years old.

J.K. Rowling poses at 'Harry Potter and The Cursed Child parts 1 & 2' on Broadway Opening Night at... [+]

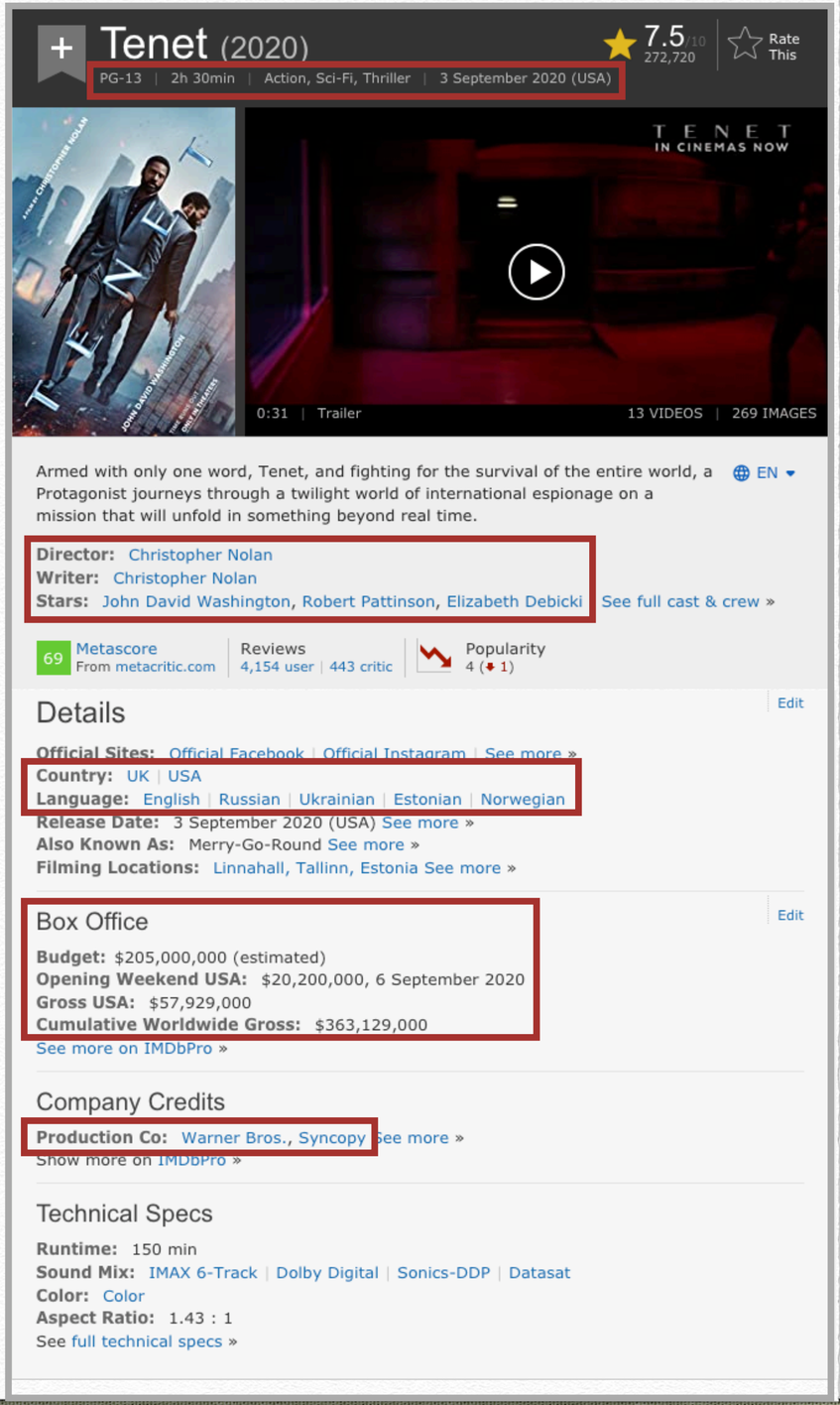
Movie in general

- Predict revenue?
- Data source: IMDb



Movie in general: data

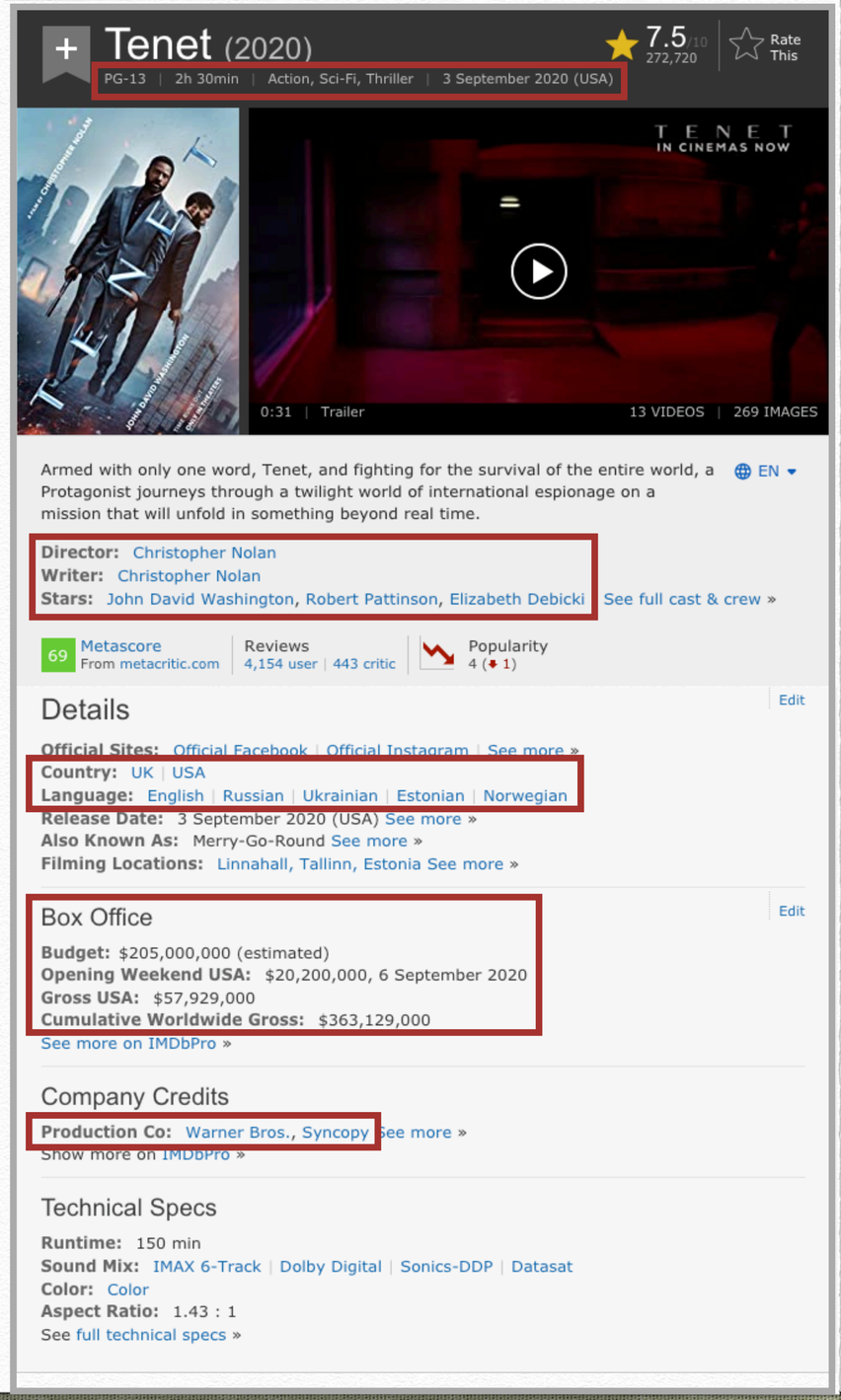
- Predict revenue?
- Data source: IMDb
  - MPAA rating
  - runtime
  - genre
  - release date
  - director, writer, cast
  - country, language
  - box office
  - …



Target & Predictors

- Target
  opening weekend revenue in USA

- Predict revenue?
- Data source: IMDb
  - MPAA rating
  - runtime
  - genre
  - release date
  - director, writer, cast
  - country, language
  - box office

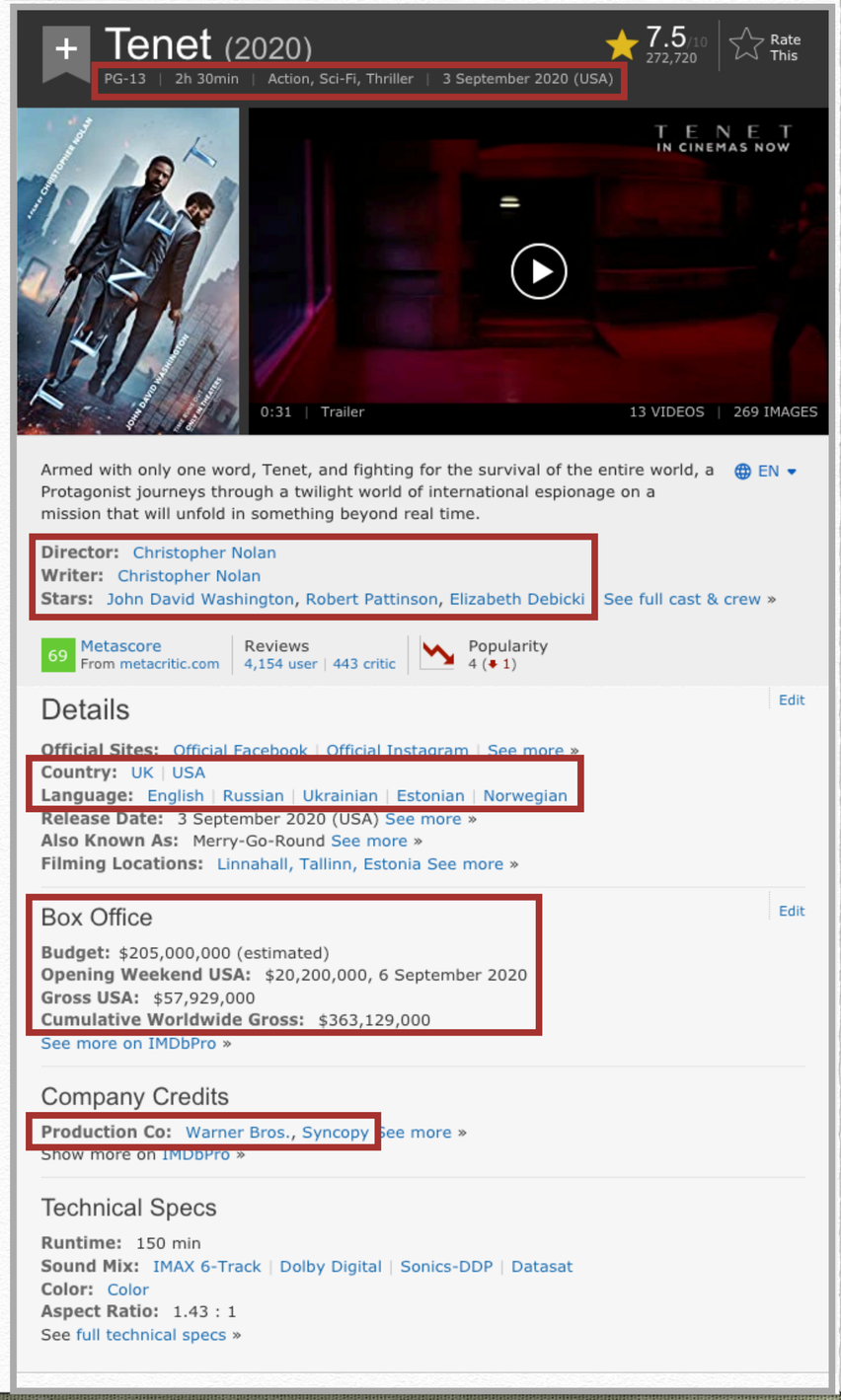


Target & Predictors

- Predict revenue?
- Data source: IMDb
  - MPAA rating
  - runtime
  - genre
  - release date
  - director, writer, cast
  - country, language
  - box office

- **Target**
  opening weekend revenue in USA

- **Continuous**
  - runtime
  - release year
  - budget



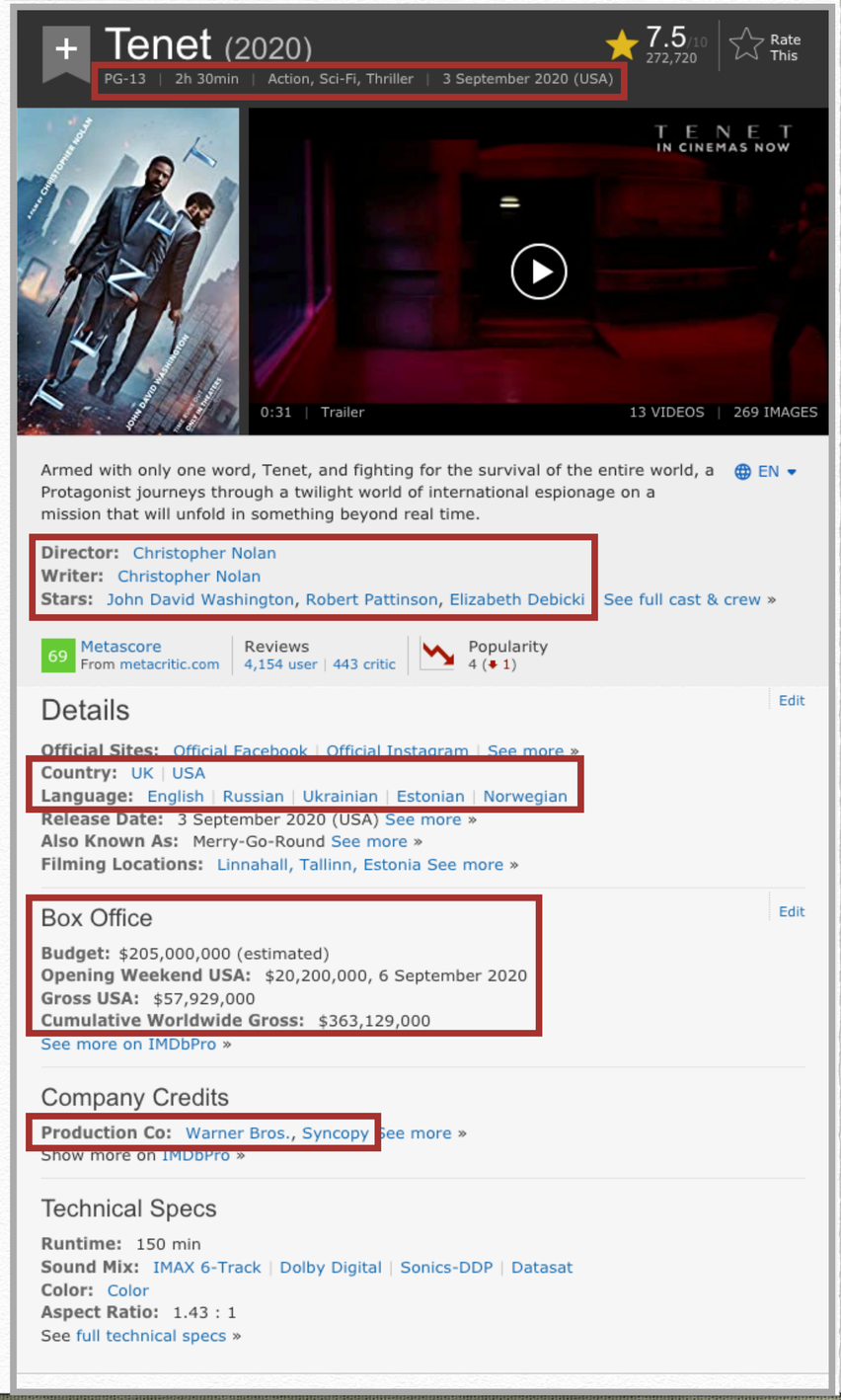
Target & Predictors

- Predict revenue?
- Data source: IMDb
  - MPAA rating
  - runtime
  - genre
  - release date
  - director, writer, cast
  - country, language
  - box office

- **Target**
  opening weekend revenue in USA

- **Continuous**
  - runtime
  - release year
  - budget

- **Categorical**
  - MPAA rating
  - distributor
  - director
  - writer
  - cast
  - genre
  - country
  - language

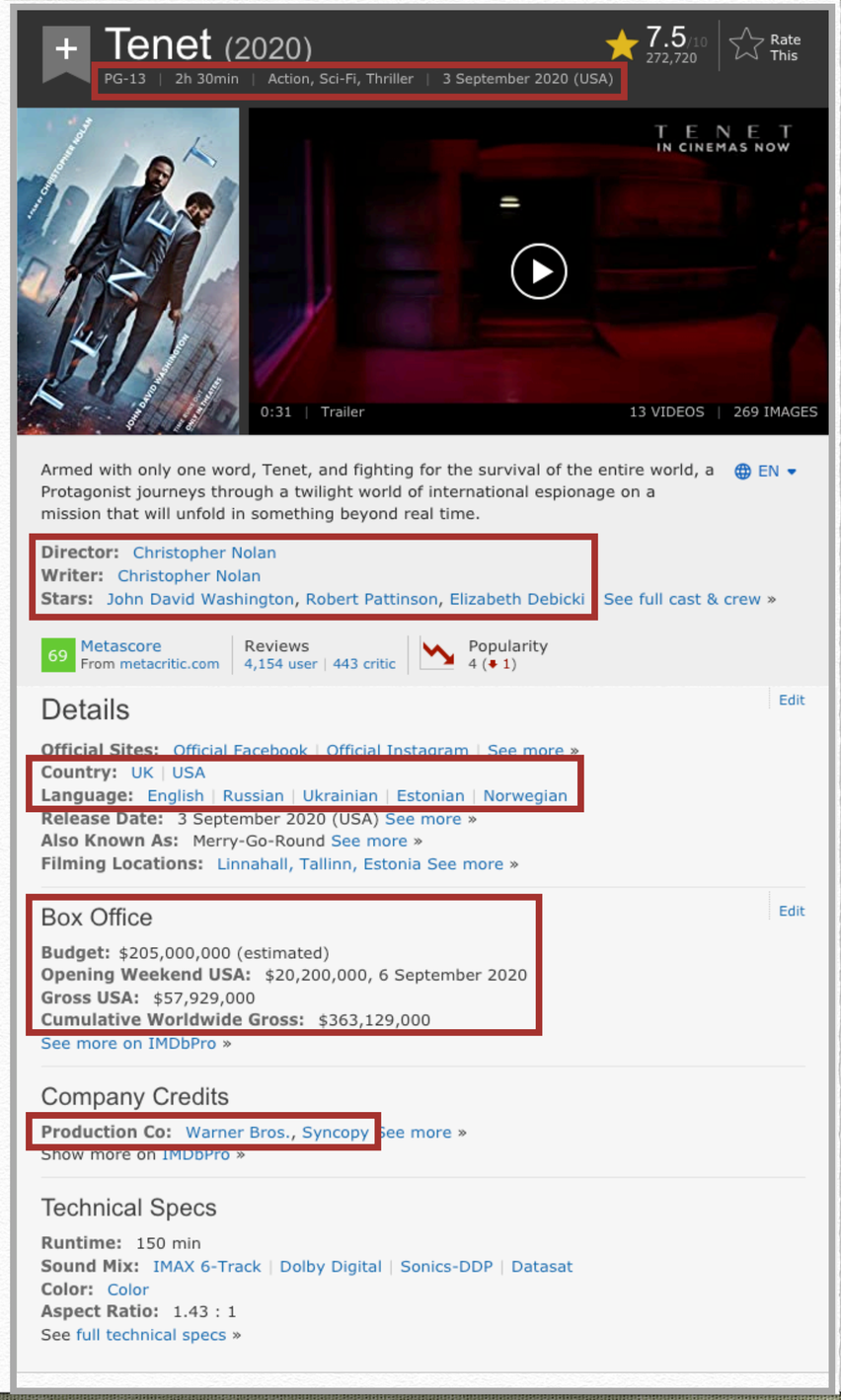


Target & Predictors

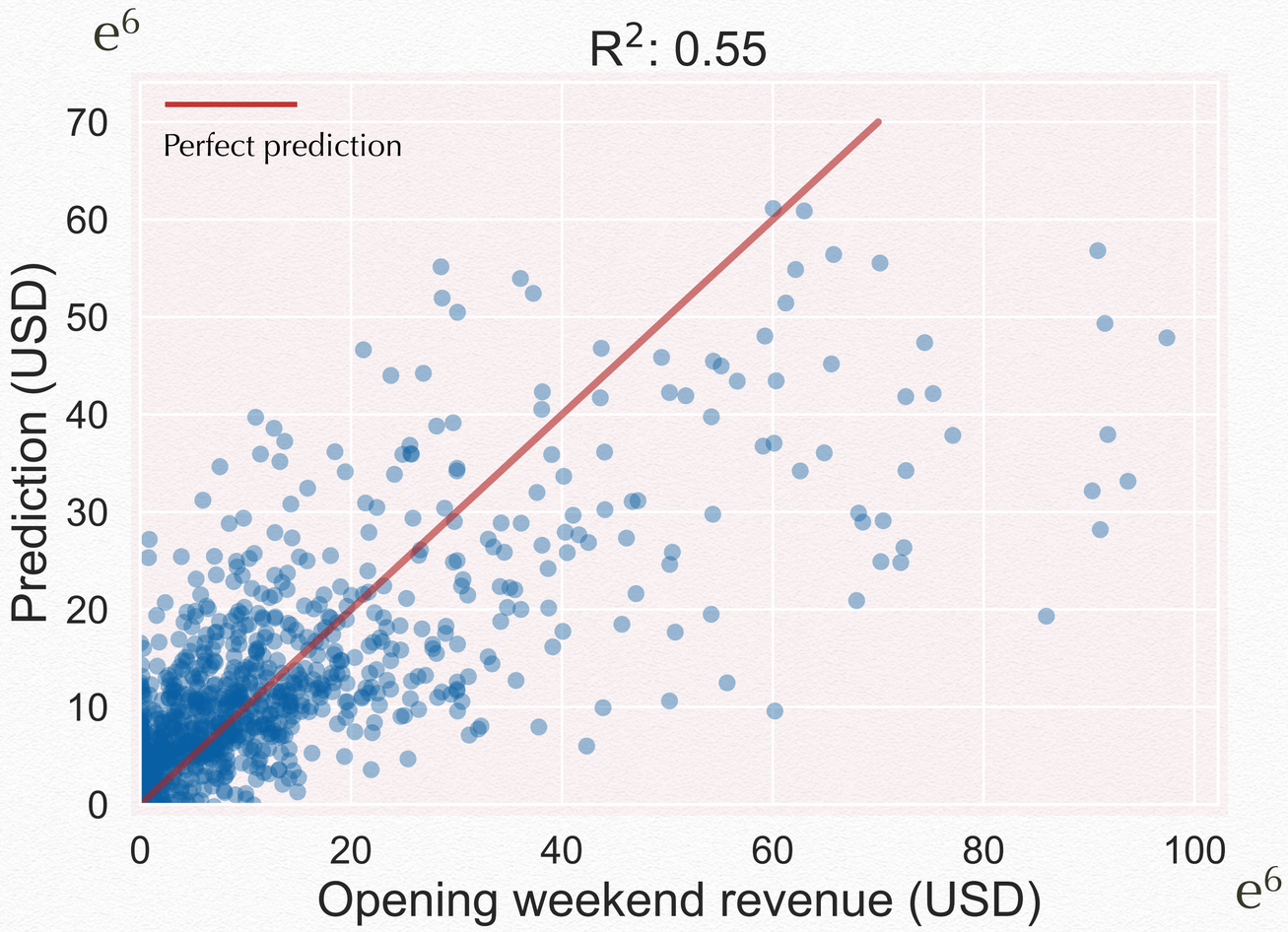
- Predict revenue?
- Data source: IMDb
  - MPAA rating
  - runtime
  - genre
  - release date
  - director, writer, cast
  - country, language
  - box office

6,000+ movies

- **Target**
  opening weekend revenue in USA
- **Continuous**
  - runtime
  - release year
  - budget
- **Categorical**
  - MPAA rating
  - distributor
  - director
  - writer
  - cast
  - genre
  - country
  - language

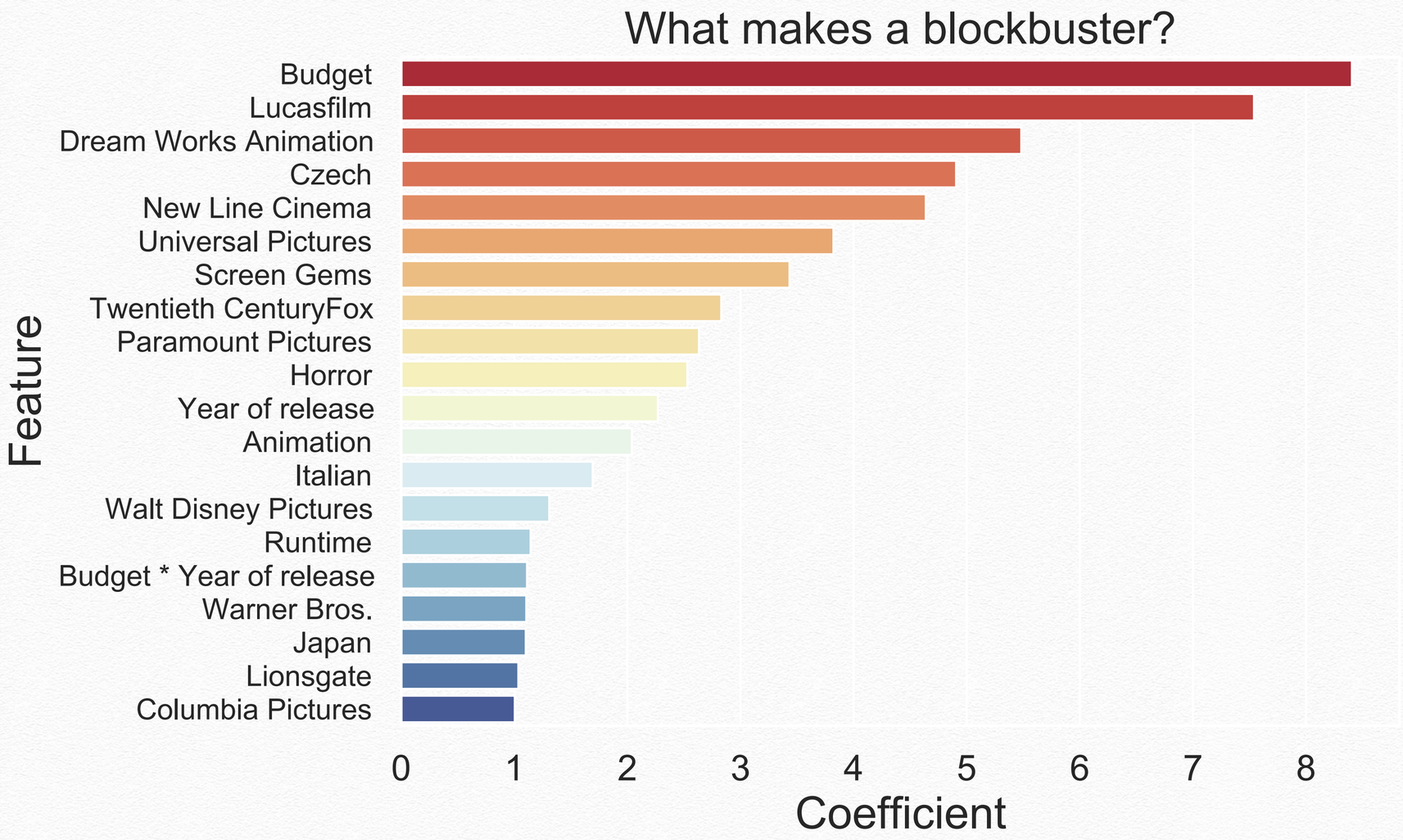


Movie in general: model

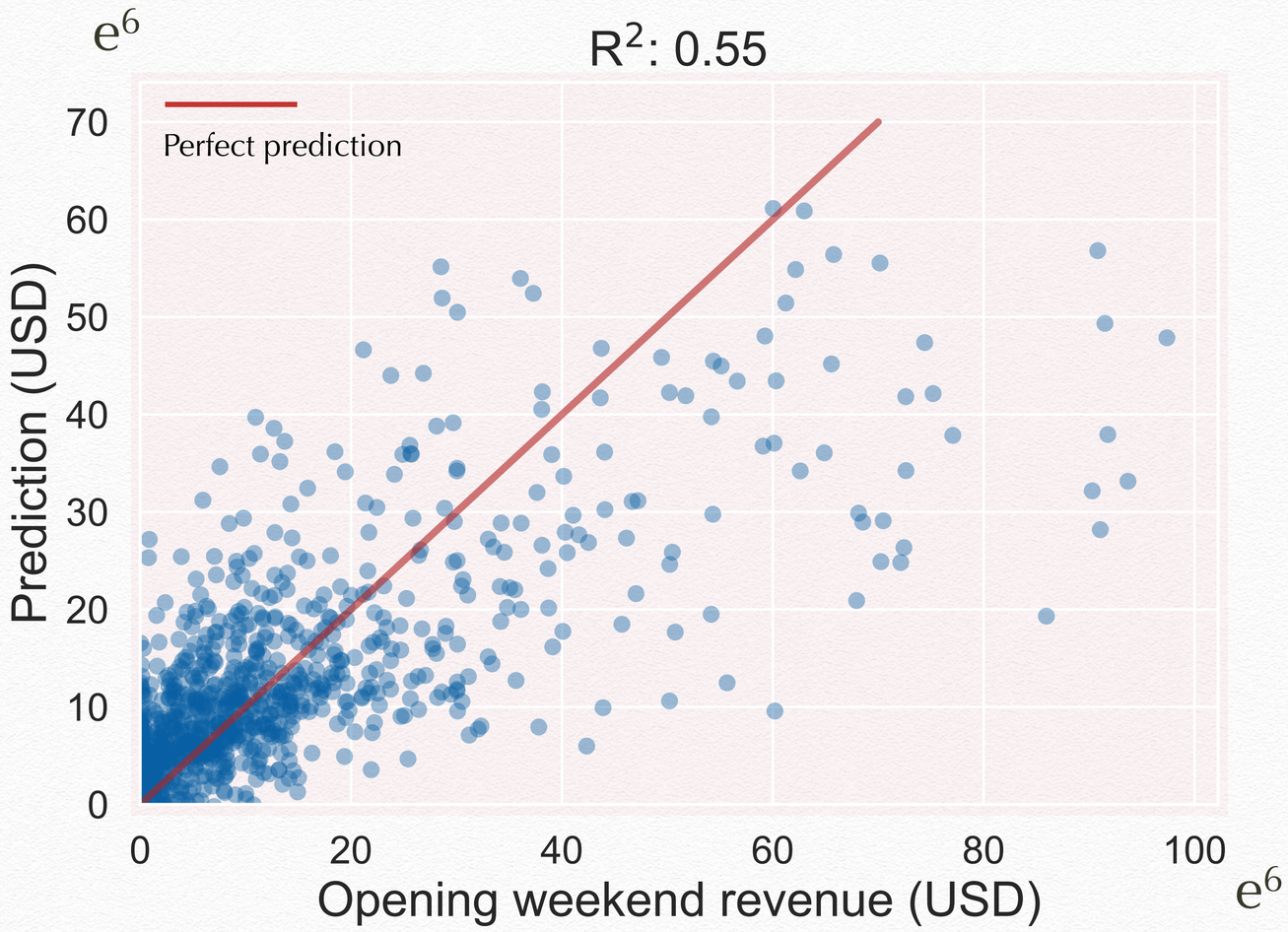


- Continuous
  - runtime
  - release year
  - budget
- Categorical
  - MPAA rating
  - distributor
  - director
  - writer
  - cast
  - genre
  - country
  - language
- polynomial features
- scaling
- Ridge (α=19.2)
- LASSO (α=0.02)

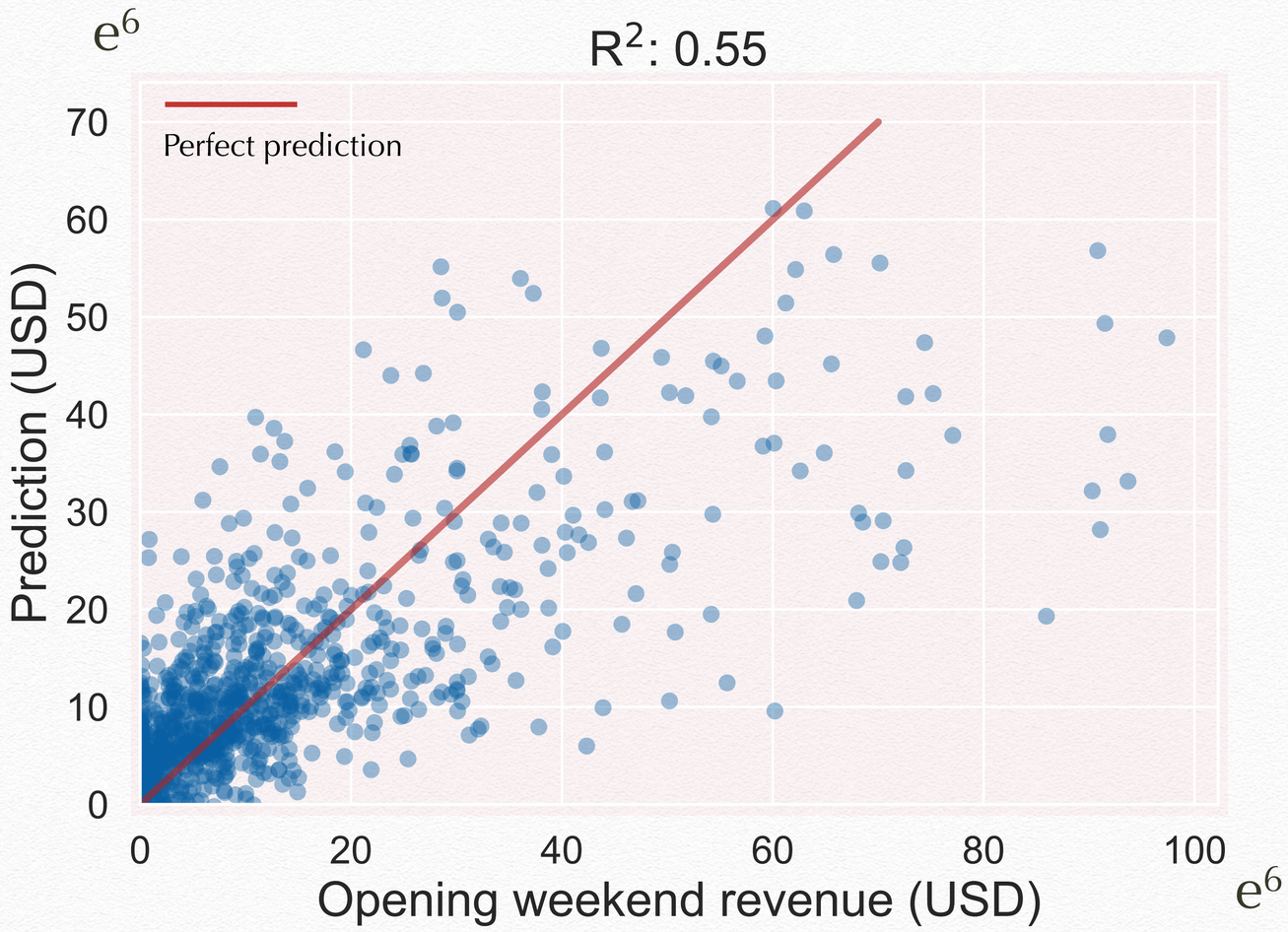
What we learned…



Who are the outliers?



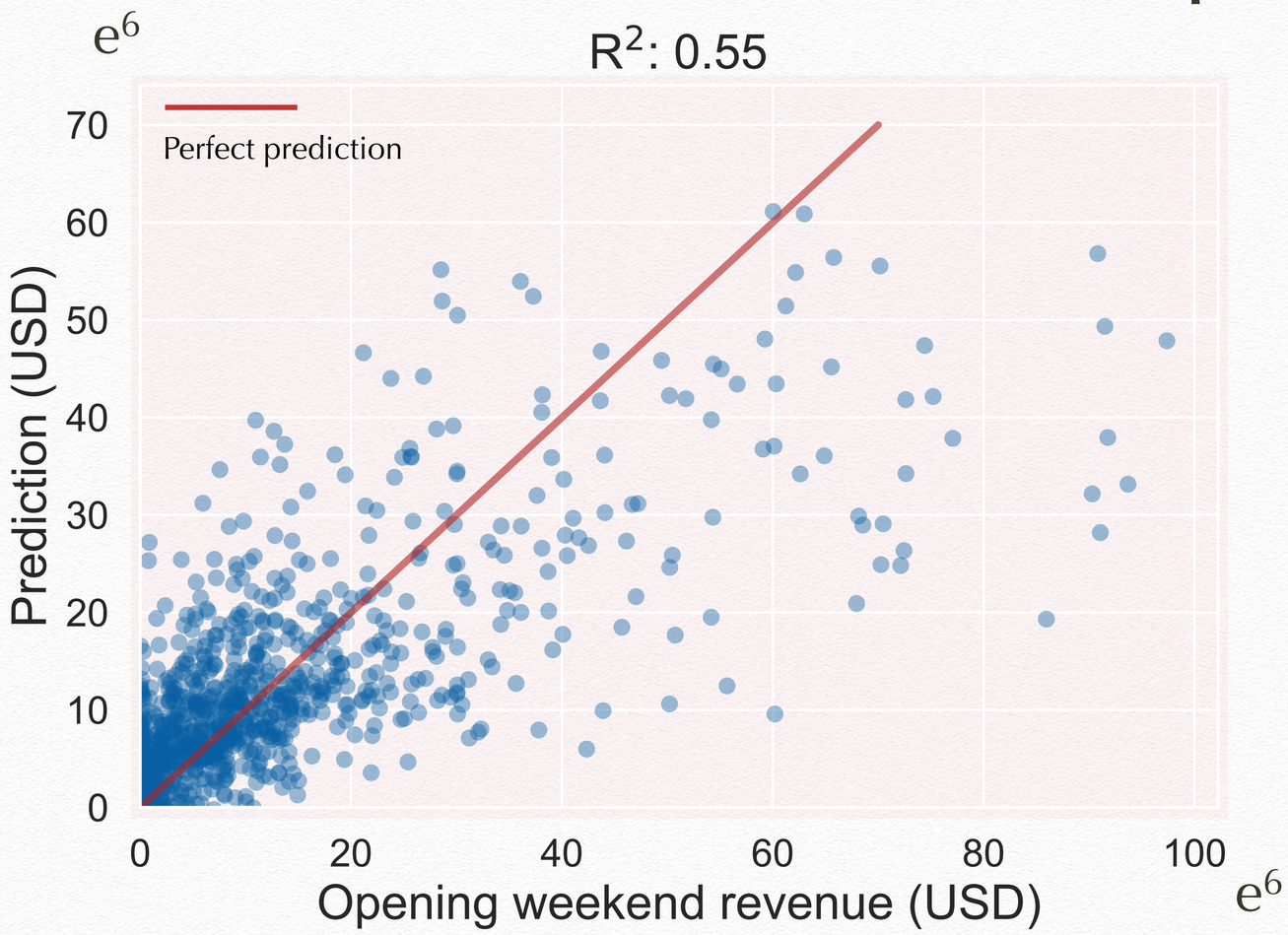
Who are the outliers?



movie_title
Joker
Fifty Shades of Grey
Halloween
It Chapter Two
Us
Harry Potter and the Chamber of Secrets
The Hangover Part II
Twilight
Despicable Me 2
Pitch Perfect 2
Logan
Iron Man
Harry Potter and the Prisoner of Azkaban
Harry Potter and the Sorcerer's Stone
X2: X-Men United

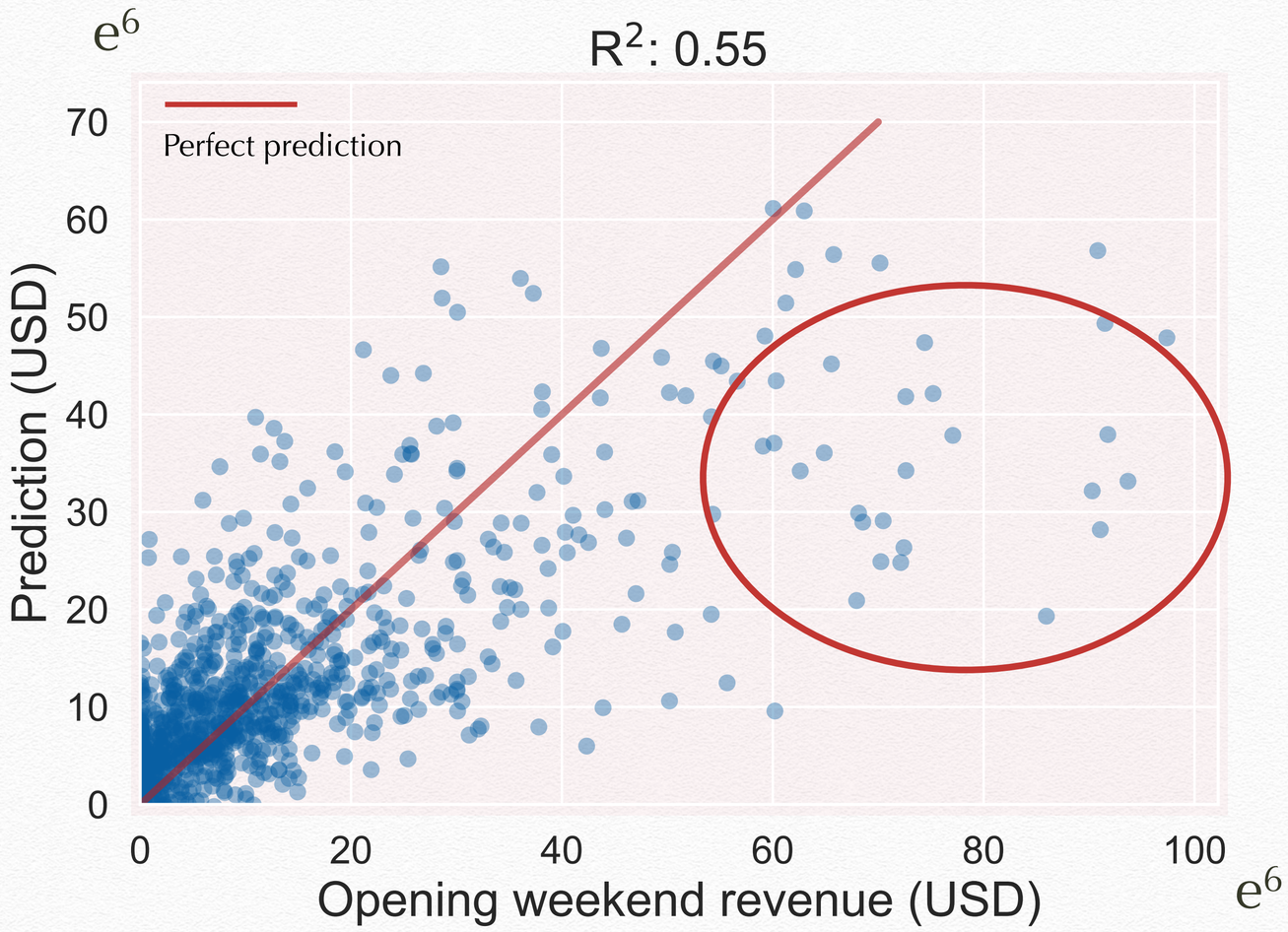
Who are the outliers?

Movie adaptations!



movie_title
Joker
Fifty Shades of Grey
Halloween
It Chapter Two
Us
Harry Potter and the Chamber of Secrets
The Hangover Part II
Twilight
Despicable Me 2
Pitch Perfect 2
Logan
Iron Man
Harry Potter and the Prisoner of Azkaban
Harry Potter and the Sorcerer's Stone
X2: X-Men United

Who are the outliers?



Movie adaptations!

J.K. Rowling poses at 'Harry Potter and The Cursed Child parts 1 & 2' on Broadway Opening Night at... [+]

Movie adaptations: model…not so easy!

~600 movie adaptations

current model: **low R^2**

Movie adaptations: build a new model

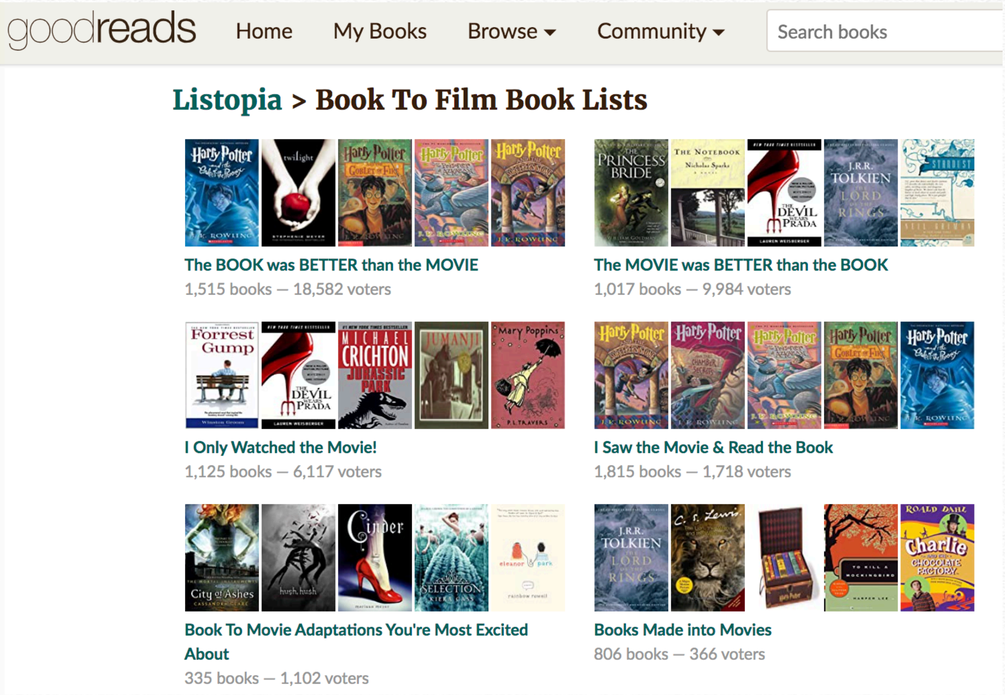
~600 movie adaptations

current model: **low R^2**

new model:**R^2: 0.39**
(trained with ~600 movies)

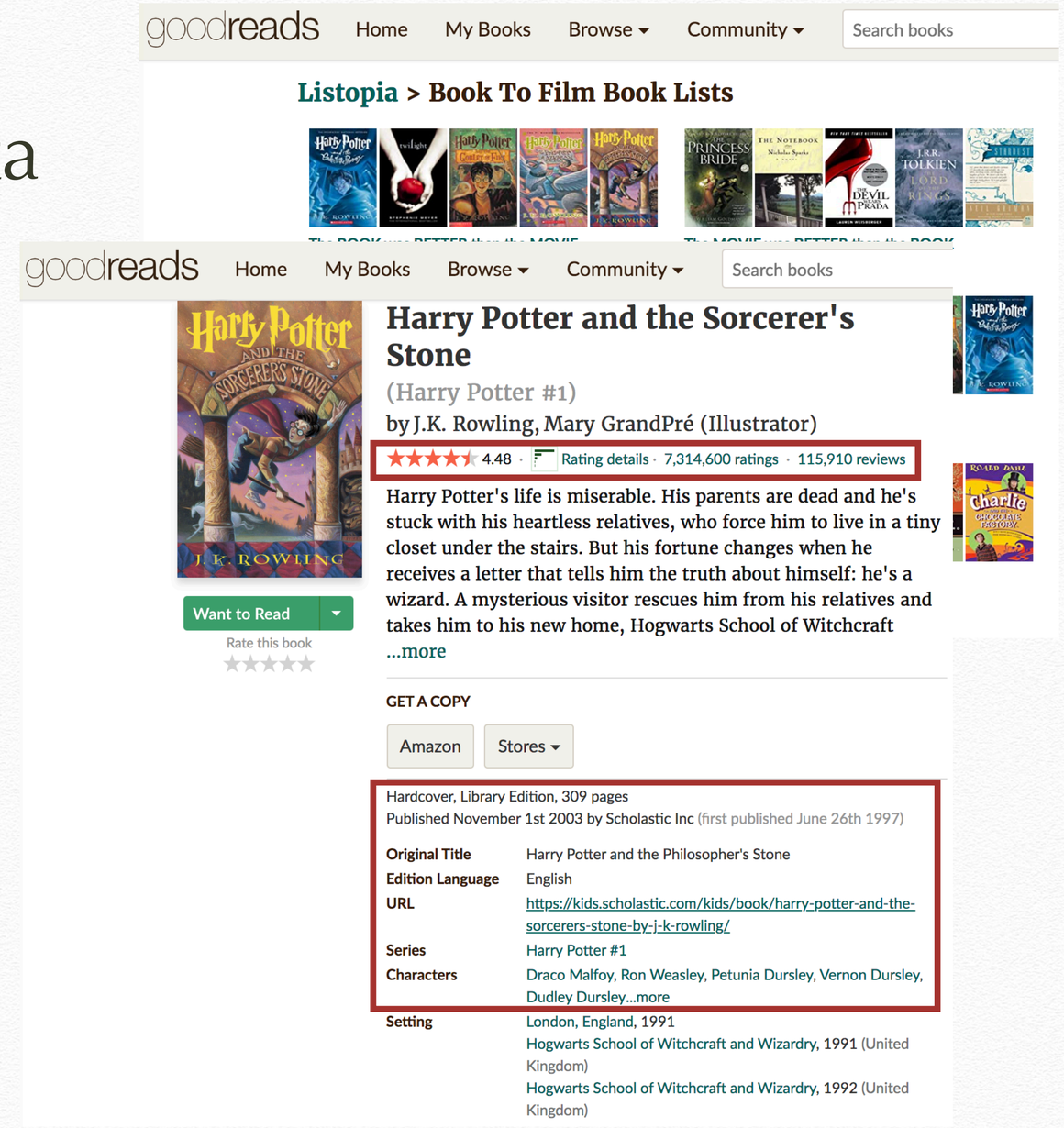
What to do?

- IMDb
- goodreads



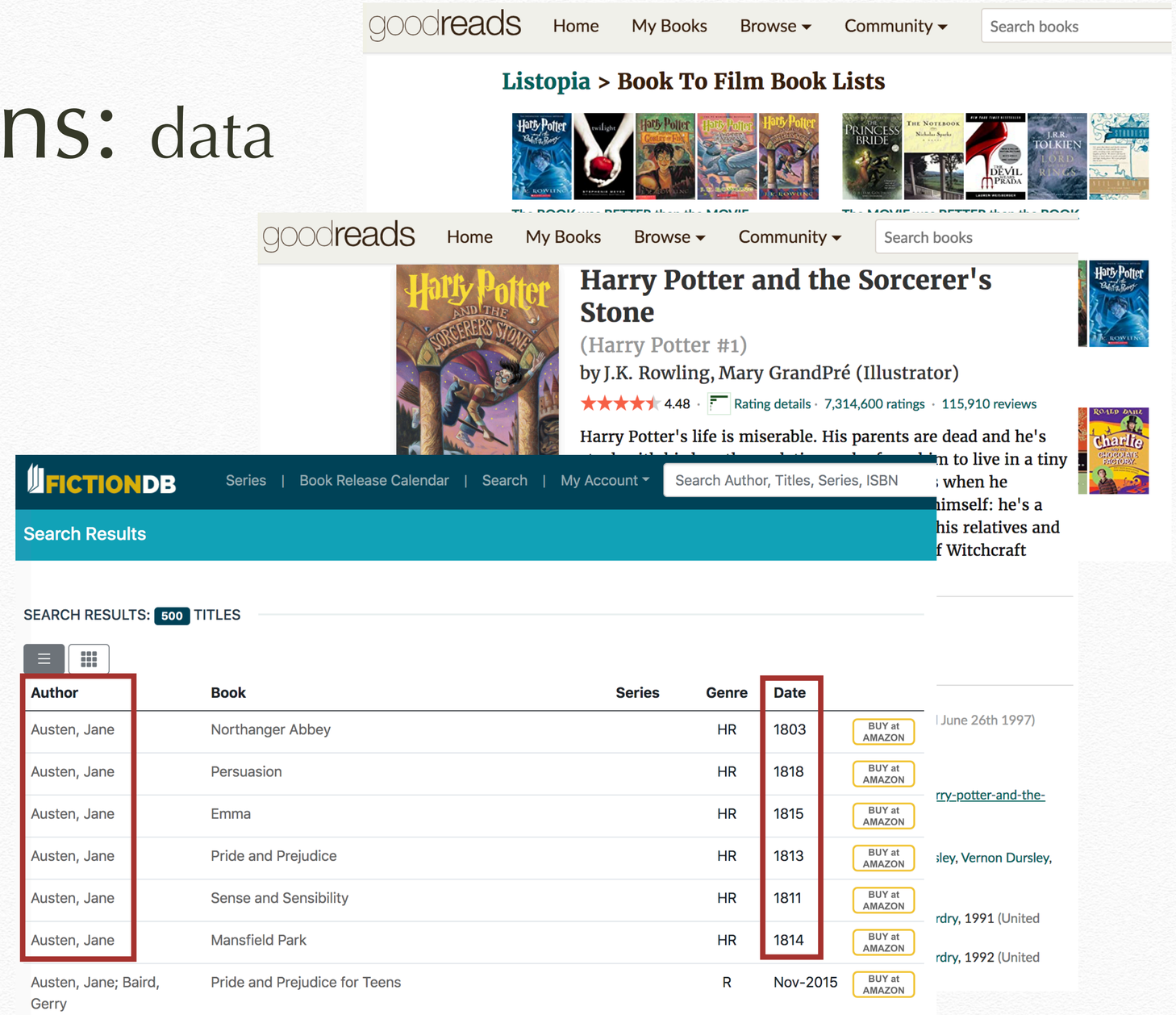
Movie adaptations: data

- IMDb
- goodreads
  - ratings, reviews
  - book info



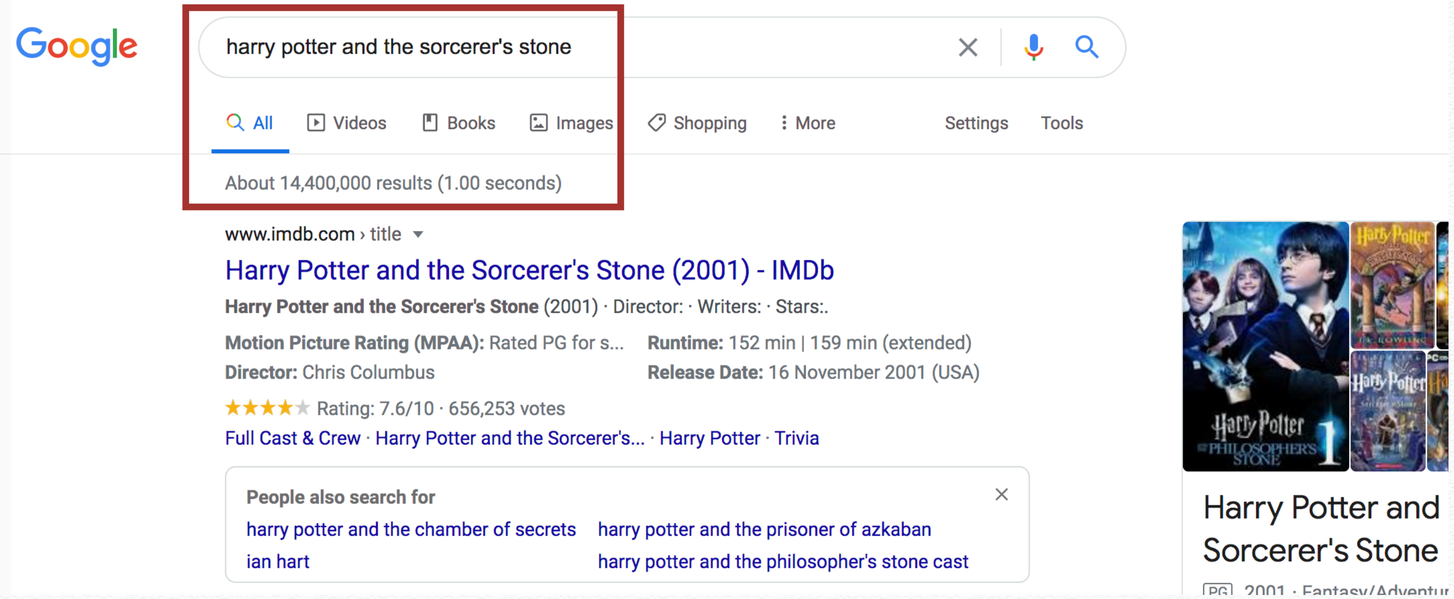
Movie adaptations: data

- IMDb
- goodreads
  - ratings, reviews
  - book info
- fictionDB
  - author publications



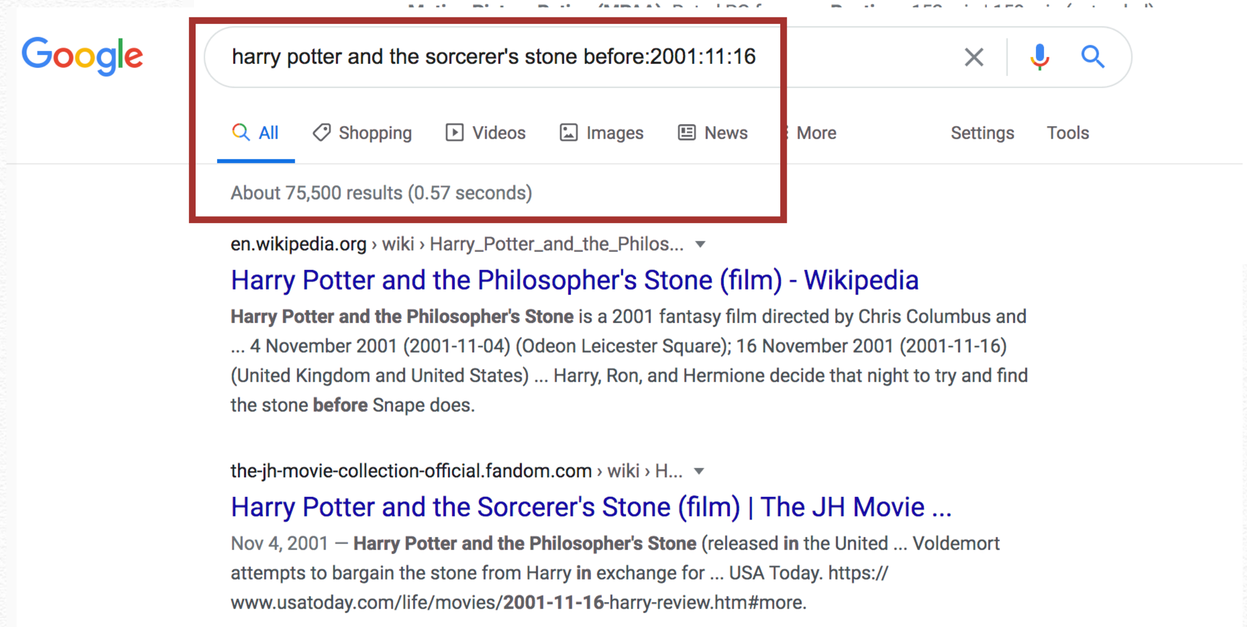
Movie adaptations: data

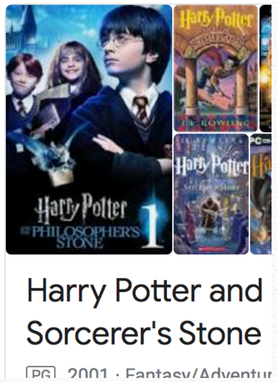
- IMDb
- goodreads
  - ratings, reviews
  - book info
- fictionDB
  - author publications
- Google
  - book popularity
  - author popularity



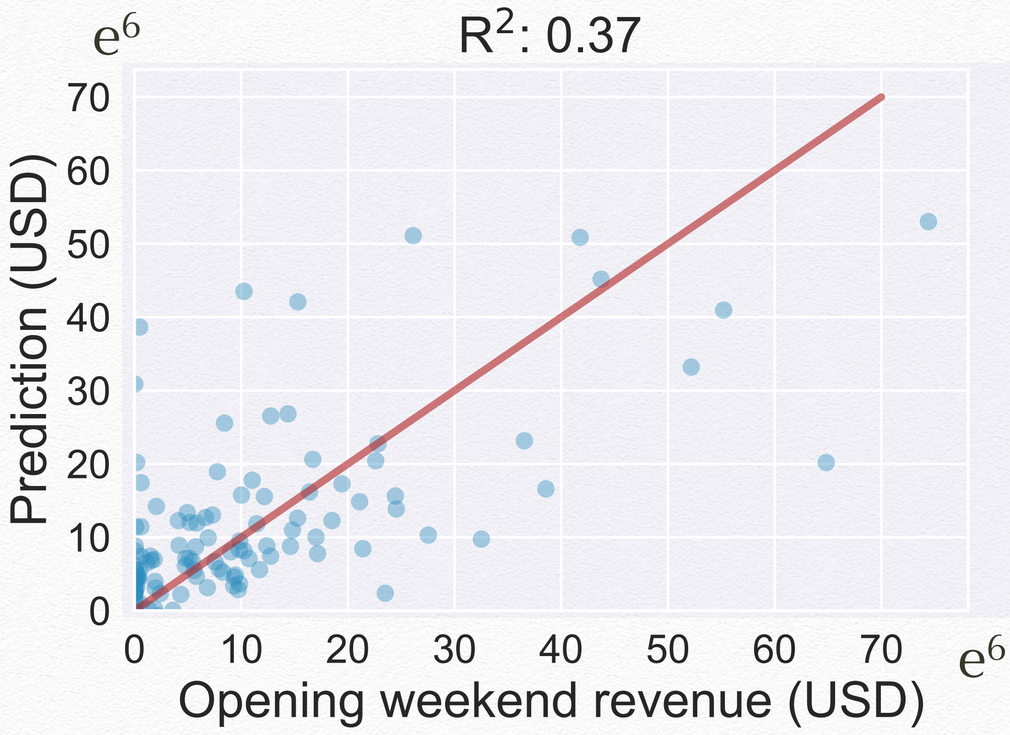
Movie adaptations: data

- IMDb
- goodreads
  - ratings, reviews
  - book info
- fictionDB
  - author publications
- Google
  - book popularity
  - author popularity

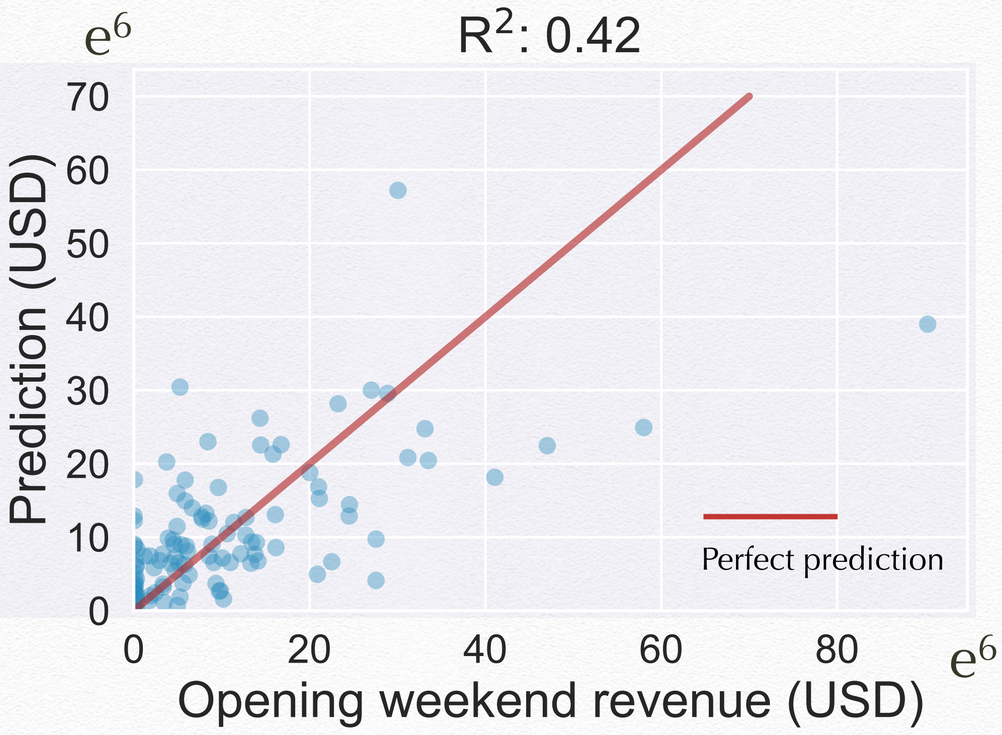




Movie adaptations: model

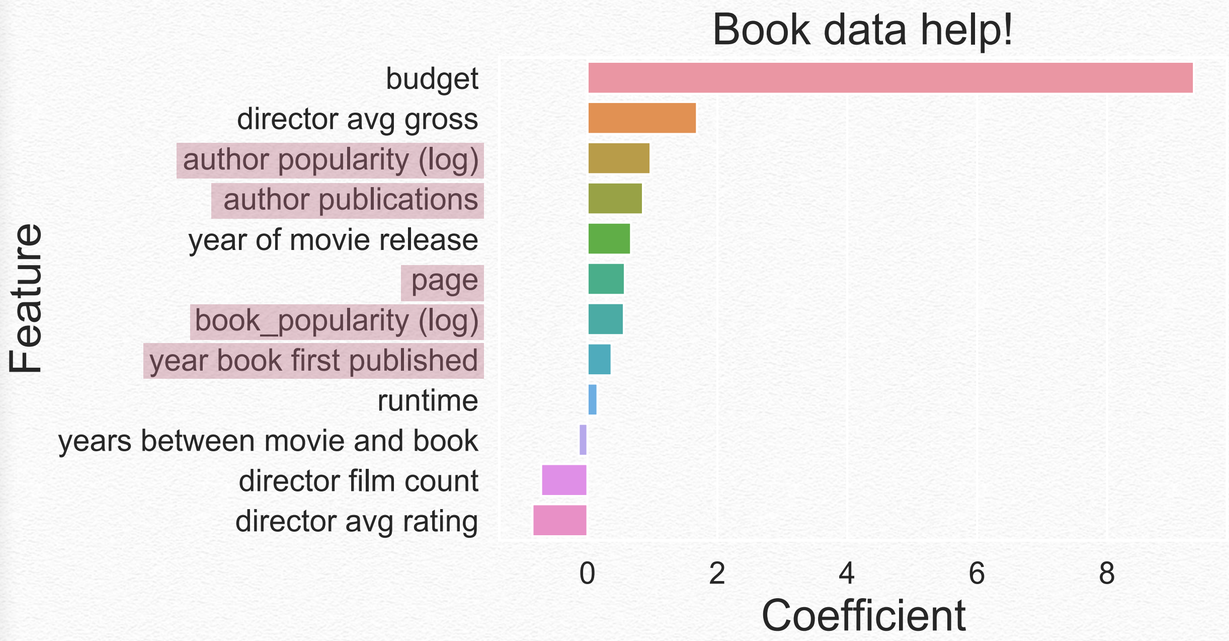


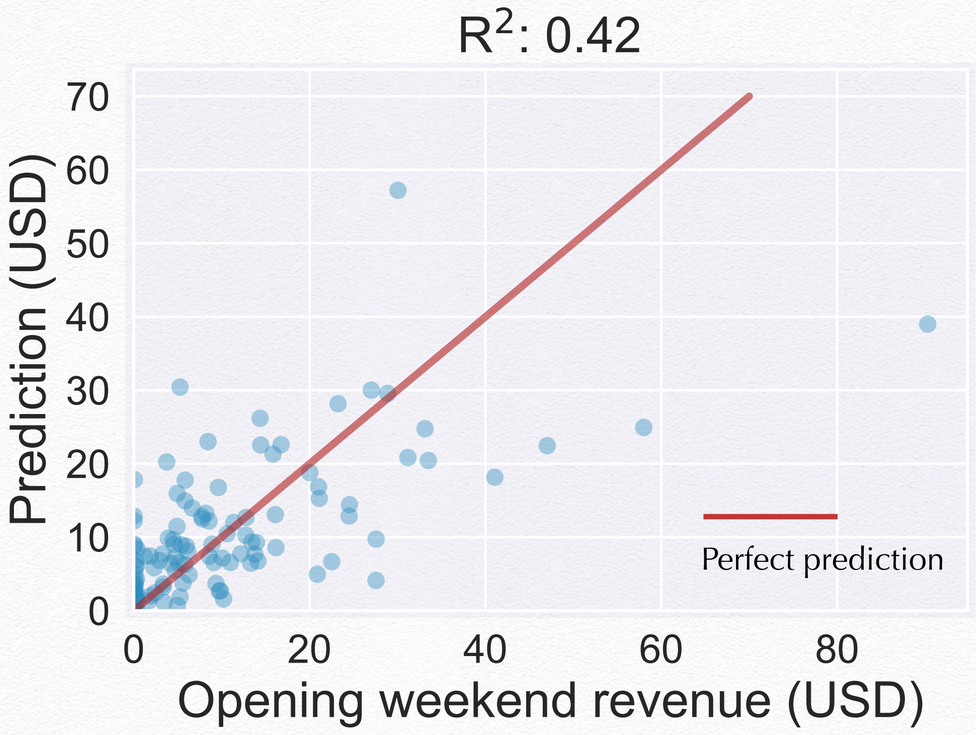
Movie data only



Movie + book data

Movie adaptations: model

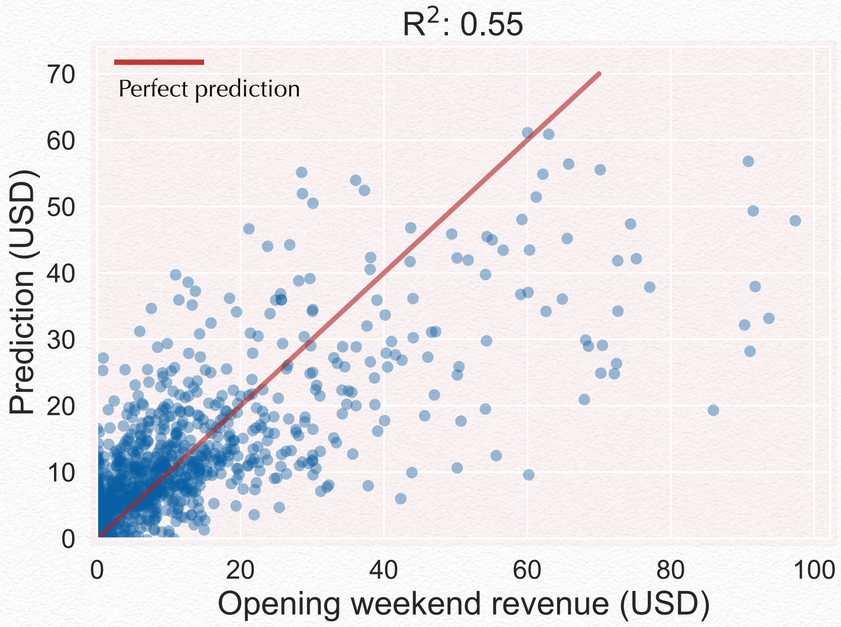




Movie + book data

Summary

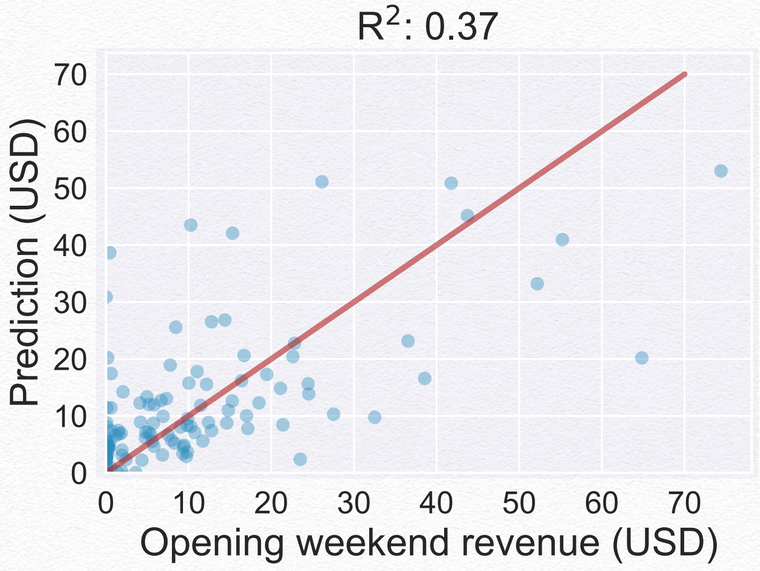
Linear regression model
for **movie in general**



- Adaptation movies among outliers

Summary

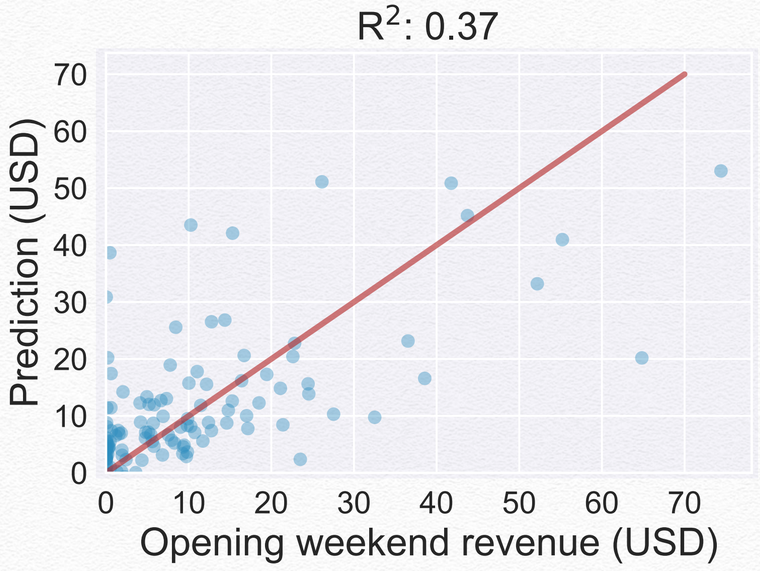
Linear regression model
for **movie adaptations**



Movie data only

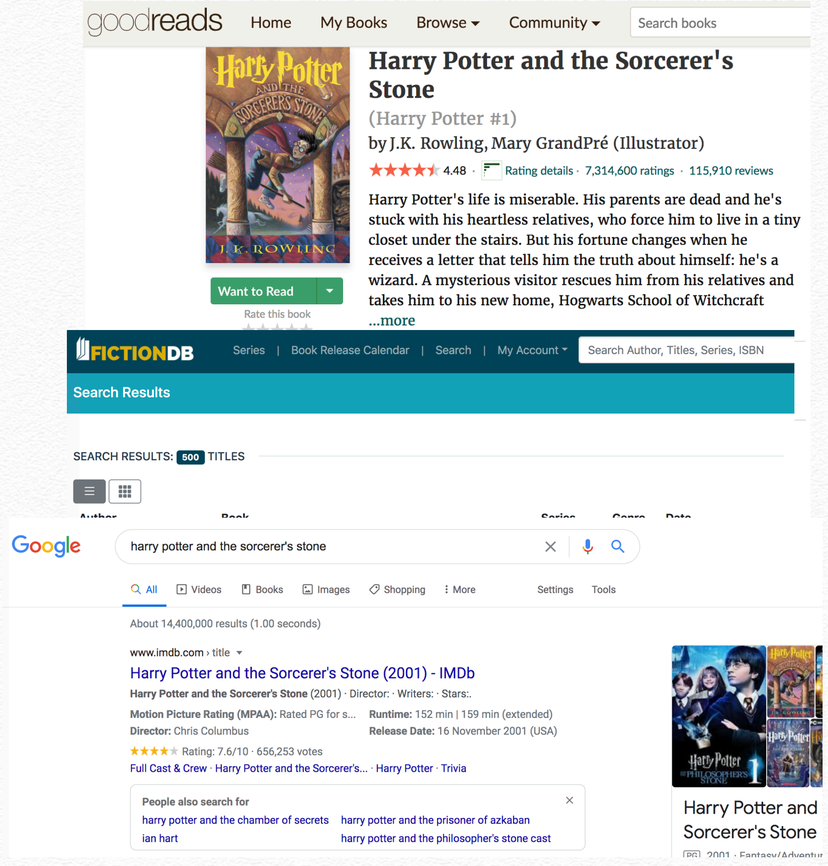
Summary

Linear regression model for **movie adaptations**



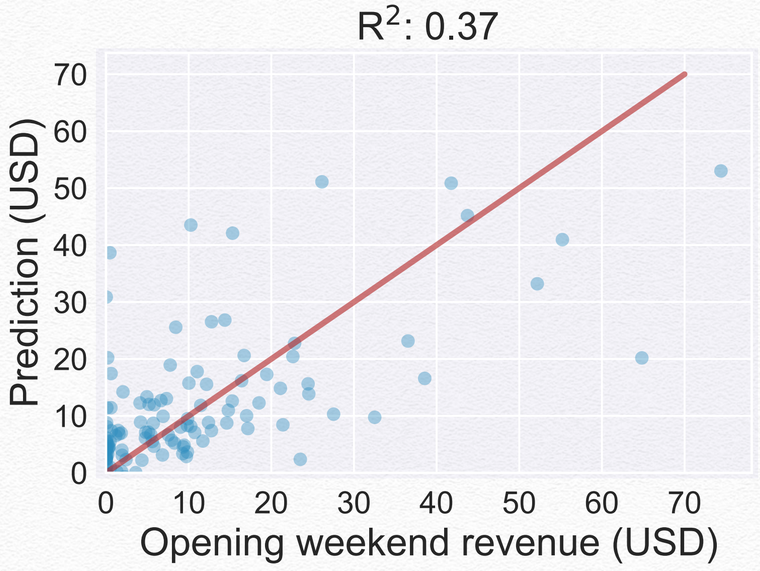
Movie data only

book data



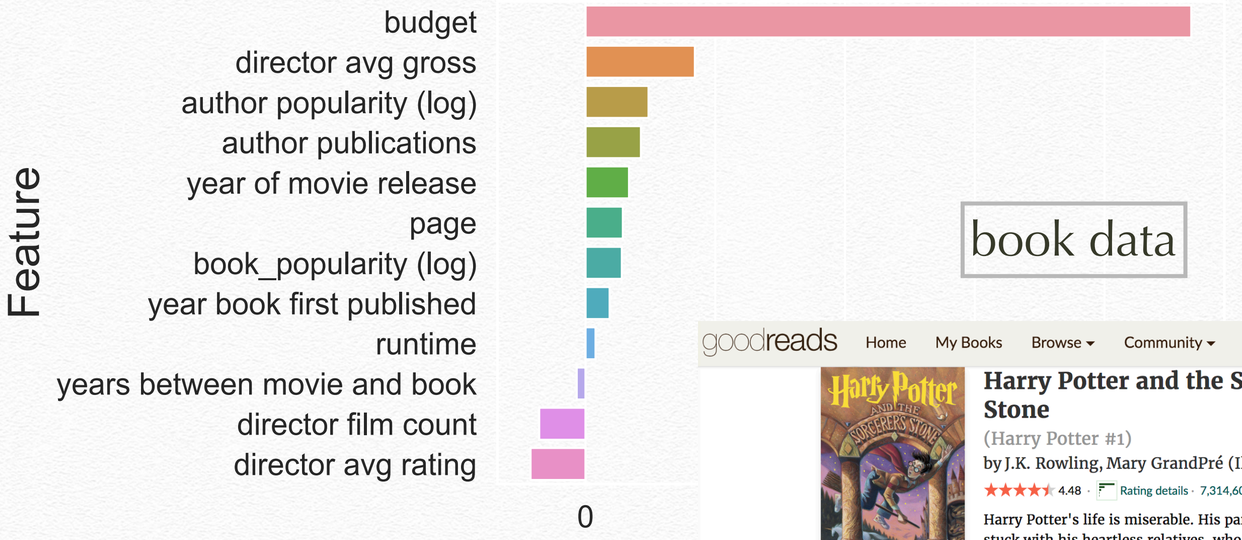
Summary

Linear regression model for **movie adaptations**



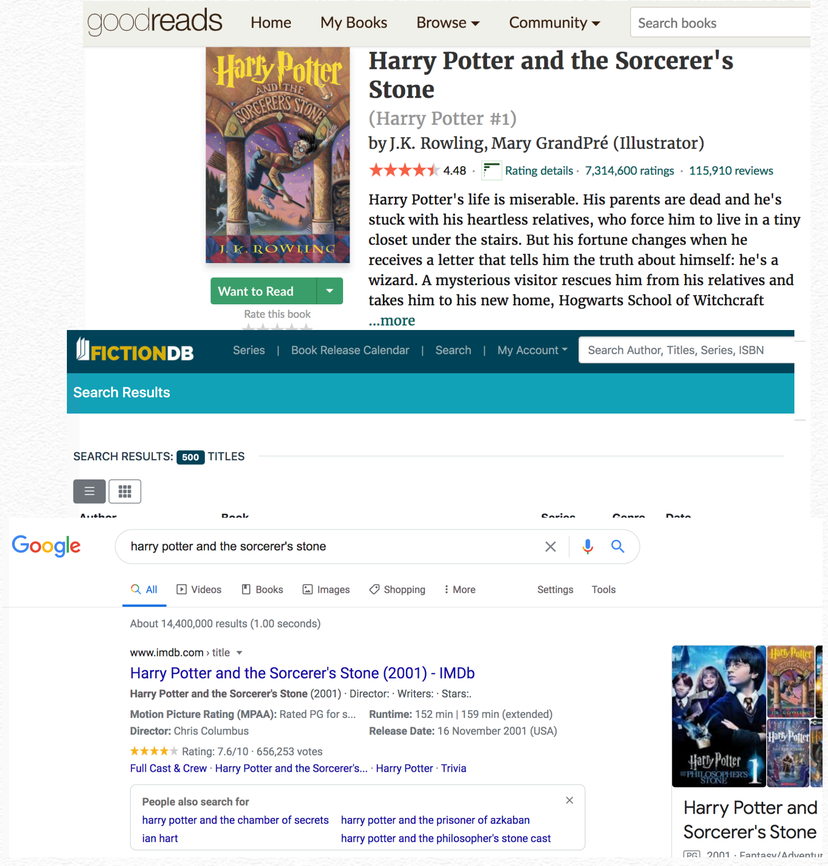
Movie data only

Book data help!



R^2: 0.37 → 0.42

book data



Outlook

- **Improve model for movie adaptations**
  - increase sample size
  - resolve timesline of book ratings

- **Decision making on production of movie adaptation**

Thank you!

Appendix

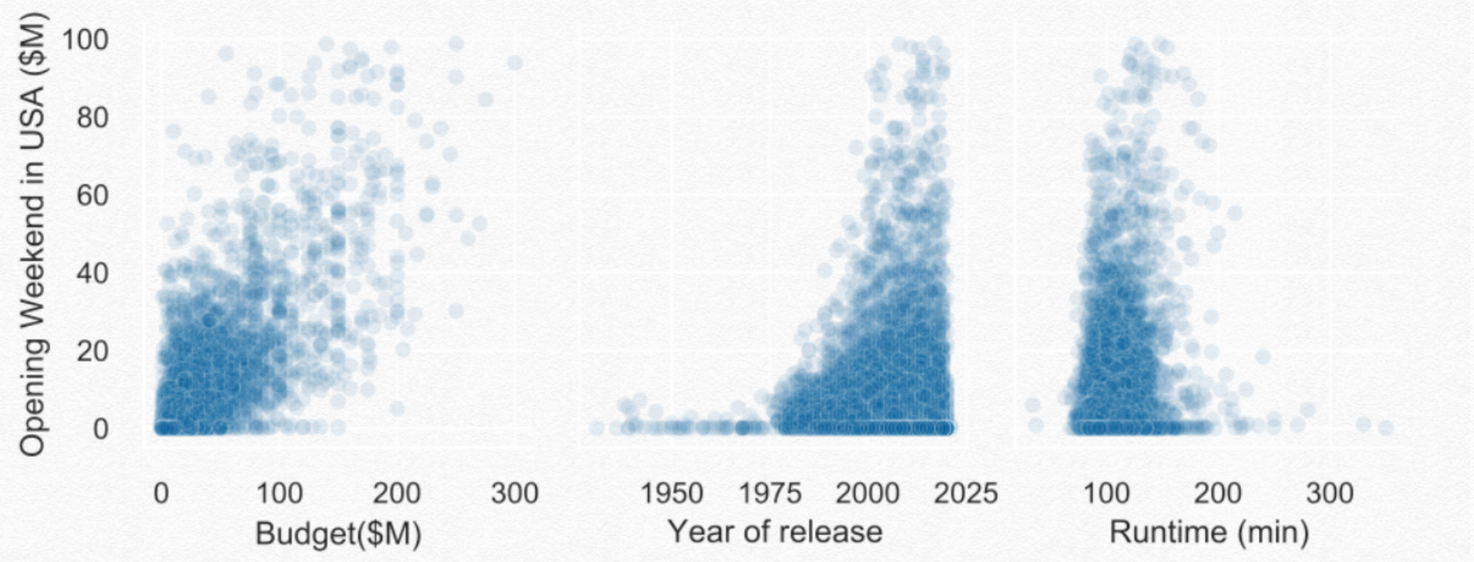
Build a simple model

6,000+ movies

Train	Validation	Test
60%	20%	20%

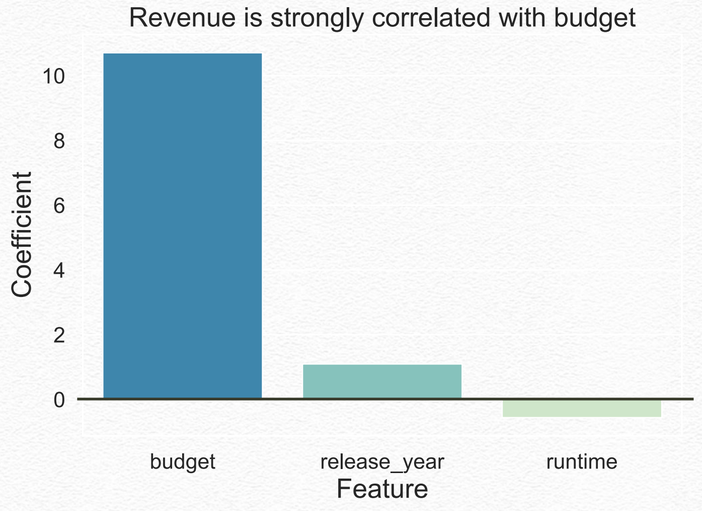
- Target
  opening weekend **revenue** in USA ($M)

- Continuous
  - runtime (min)
  - release year
  - budget ($)



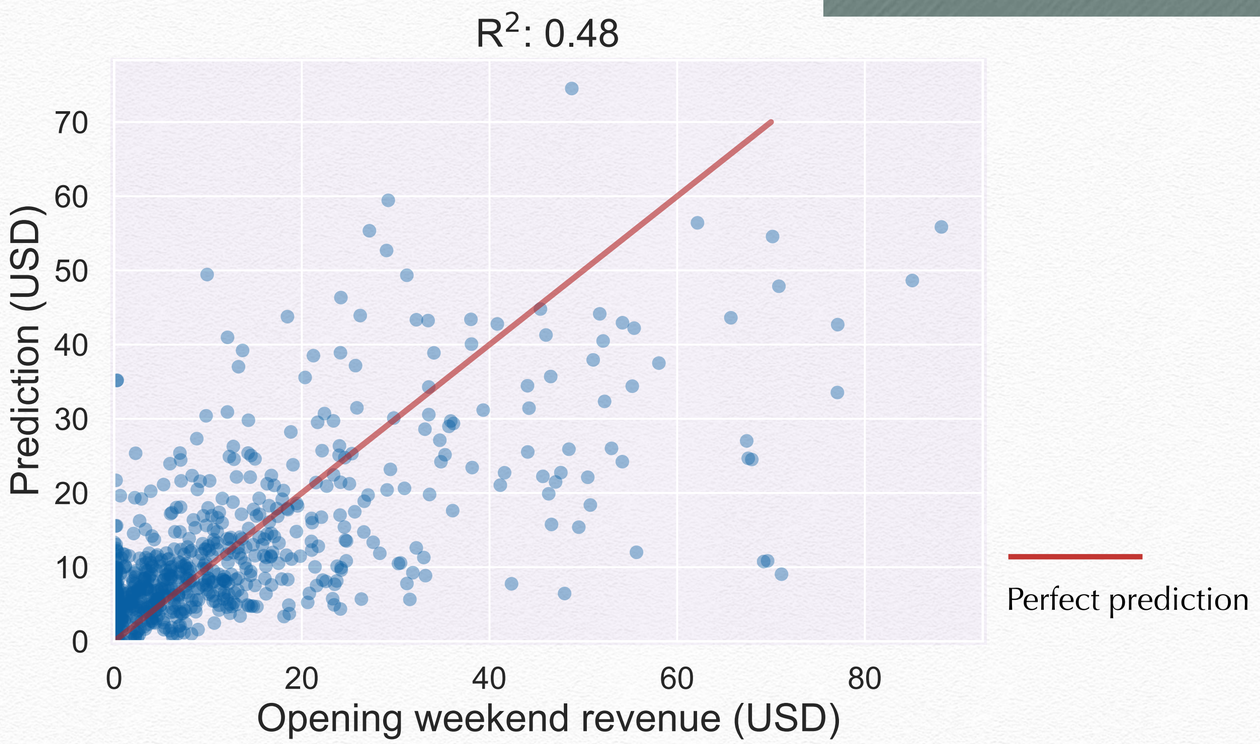
Build a simple model

Cross-Validation (5-fold)	R^2
Linear Regression	0.48
Scaling	0.48
Polynomial Features + Scaling	



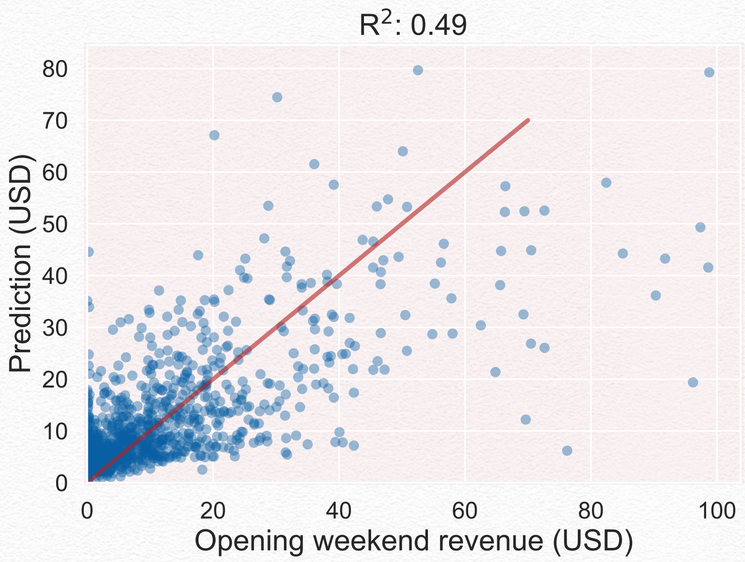
- **Target**
  opening weekend revenue in USA ($M)

- **Continuous**
  - runtime (min)
  - release year
  - budget ($)



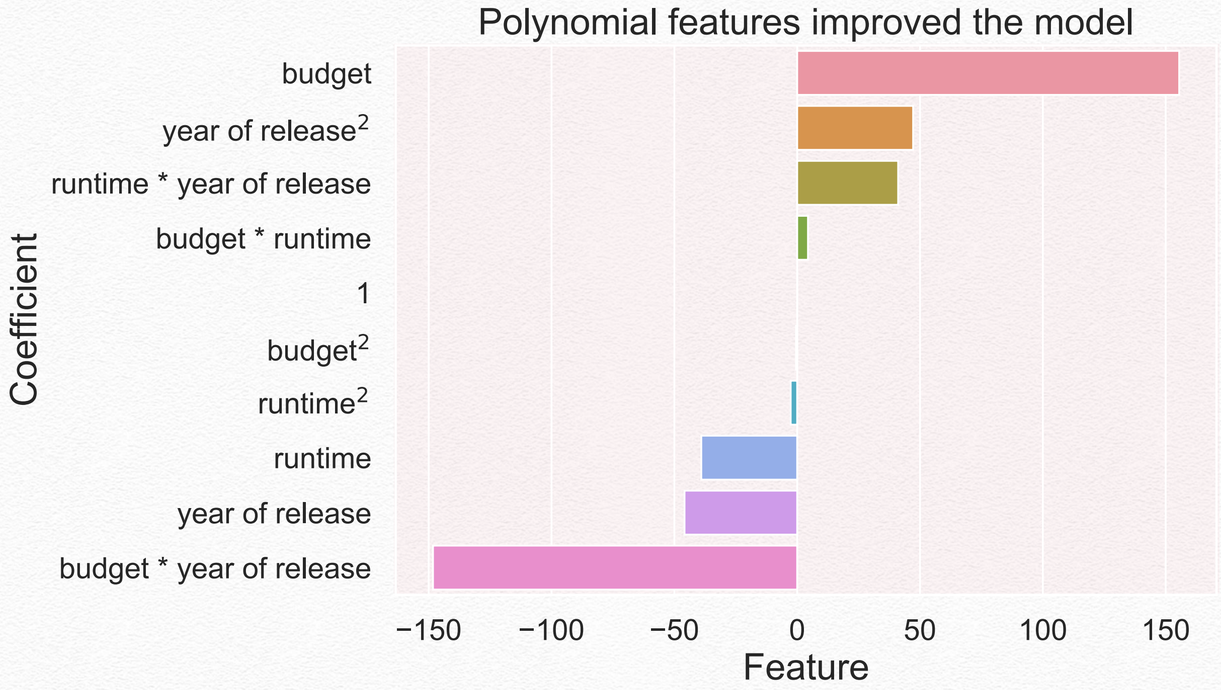
Build a simple model

Cross-Validation (5-fold)	R^2
Linear Regression	0.48
Scaling	0.48
Polynomial Features + Scaling	**0.49**



- **Target**
  opening weekend revenue in USA ($M)

- **Continuous**
  - runtime (min)
  - release year
  - budget ($)



Categorical variables

	Train: R^2	Validation: R^2
genre	0.51	0.52
distributor	0.50	0.51
cast	0.50	0.50
all	0.51	0.51
all+poly	0.60	**negative!**

overfitting!

- Continuous
  - runtime
  - release year
  - budget

- Categorical
  - MPAA rating
  - distributor
  - director
  - writer
  - cast
  - genre
  - country
  - language

Ridge (α=19.2) **R^2: 0.55**
LASSO (α=0.02) **R^2: 0.55**

Evaluation

Test set
R^2: 0.59

- Continuous
  - runtime
  - release year
  - budget
- Categorical
  - MPAA rating
  - distributor
  - director
  - writer
  - cast
  - genre
  - country
  - language
- polynomial features
- scaling
- Ridge (α=19.2)
- LASSO (α=0.02)

Who are the outliers?

movie_title	res
Joker	-76.432654
Fifty Shades of Grey	-74.037323
Halloween	-69.834310
It Chapter Two	-68.496035
Us	-62.068580
Harry Potter and the Chamber of Secrets	-61.293593
The Hangover Part II	-61.176890
Twilight	-58.788479
Despicable Me 2	-58.753424
Pitch Perfect 2	-58.459170
Logan	-57.834601
Iron Man	-57.681644
Harry Potter and the Prisoner of Azkaban	-56.534309
Harry Potter and the Sorcerer's Stone	-55.773166
X2: X-Men United	-54.032581

res = prediction - actual revenue

Movie adaptations!

Movie adaptations: model…not so easy!

	Train: R^2	Validation: R^2
basic movie features	0.46	0.37
genre	0.50	0.50
other categorical	0.43	0.41
actor	0.69	0.20
all+poly	**0.72**	**0.25**

~600 movie adaptations

Ridge (α=10.25) **R^2: 0.26**
LASSO (α=0.1) **R^2: 0.39**

Movie adaptations: model…not so easy!

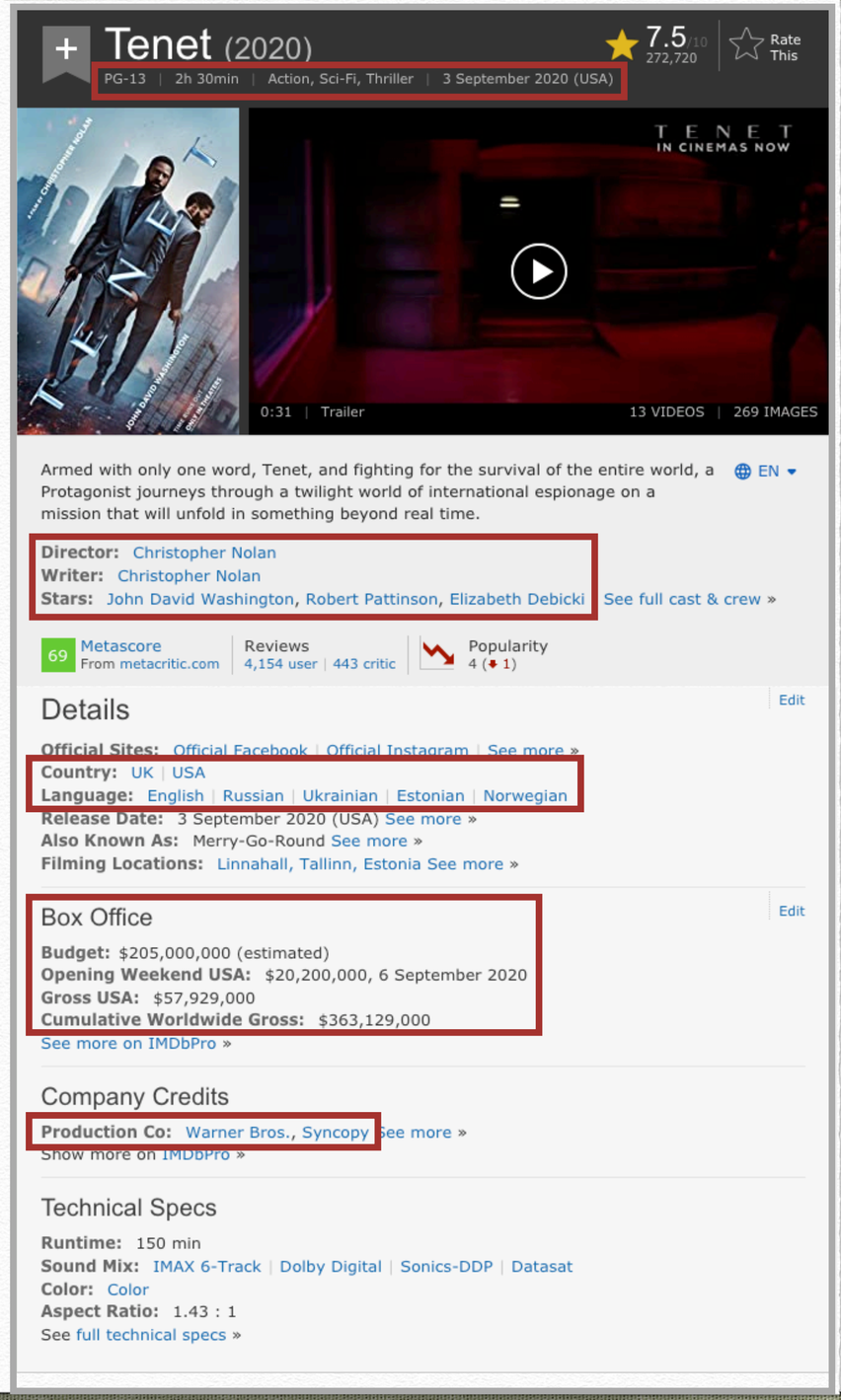
~600 movie adaptations

current model: **low R^2**

new model:
Ridge (α=10.25) **R^2: 0.26**
LASSO (α=0.1) **R^2: 0.39**

Target & Predictors

- Predict revenue?
- Data source: IMDb
  - IMDb rating
  - MPAA rating
  - runtime
  - genre
  - release date
  - director, writer, cast
  - country, language
  - box office
  - …

- Target
  opening weekend revenue in USA

- Continuous
  - ~~IMDb rating~~
  - runtime
  - release year
  - budget