Objective

Predict world-wide box office gross using only information available prior to release

Worldwide Box Office Gross
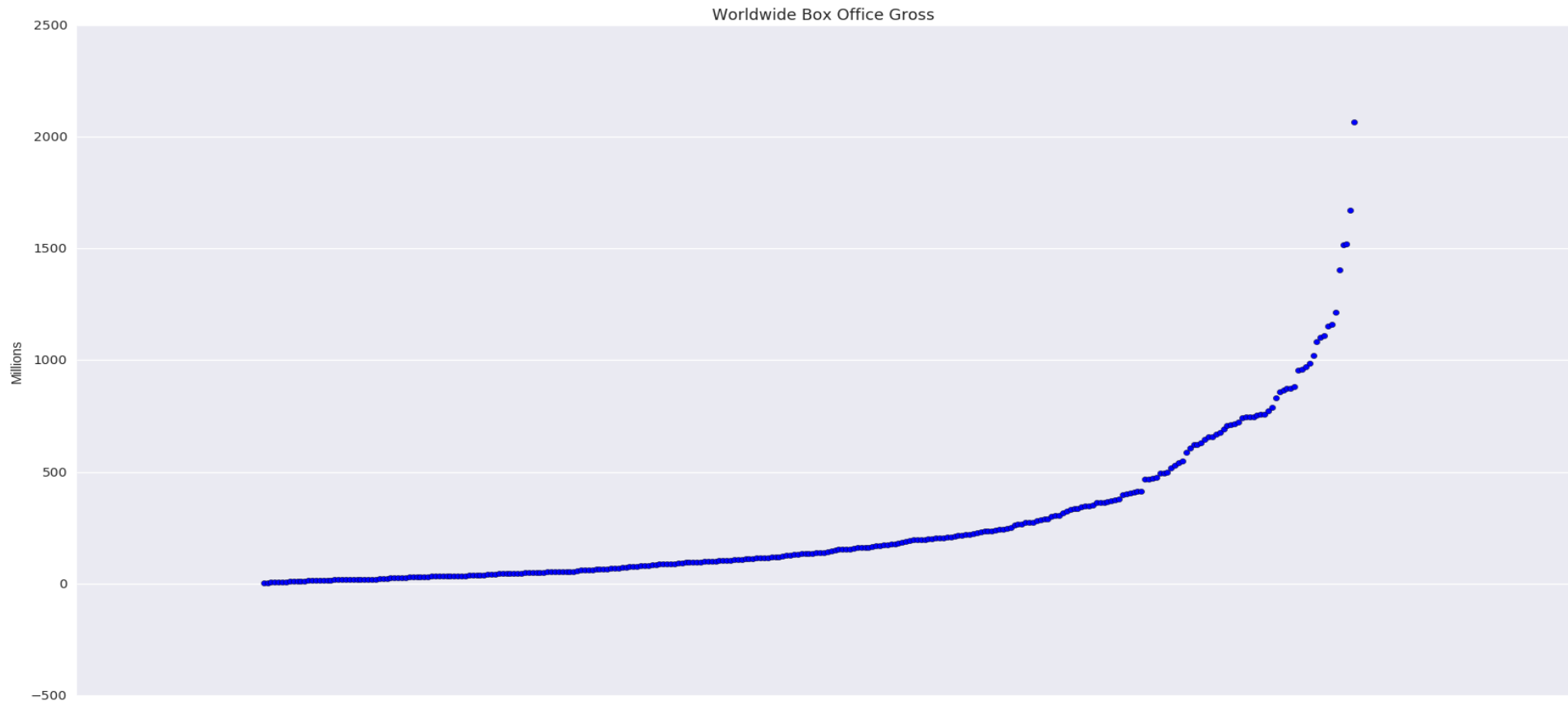
Approach

Linear regression:

- Model has seven features
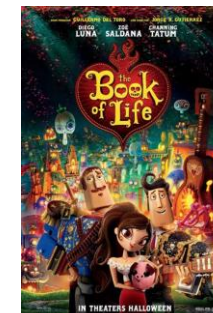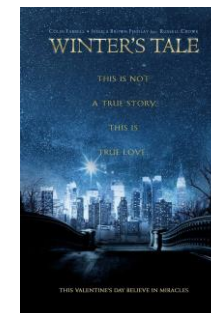- Trained on data from 400+ recent releases (2012 – 2016)
- Sources:

Result

Adjusted R^2 ~ 0.78

Hits and Misses

Movie	Prediction	Actual
Madagascar 3: Europe's Most Wanted	752	747
St. Vincent	55	55
Bridge of Spies	166	165

Movie	Prediction	Actual
Winter's Tale	190	31
The Book of Life (2014)	209	100
Philomena	-10	100

Data

What's known before release?

- Budget
- Production company
- Distributor
- Producers
- Actors
- Directors
- Writers
- Composers
- Genre
- Rating
- Release date?

Data

What's known before release?   What the model uses:
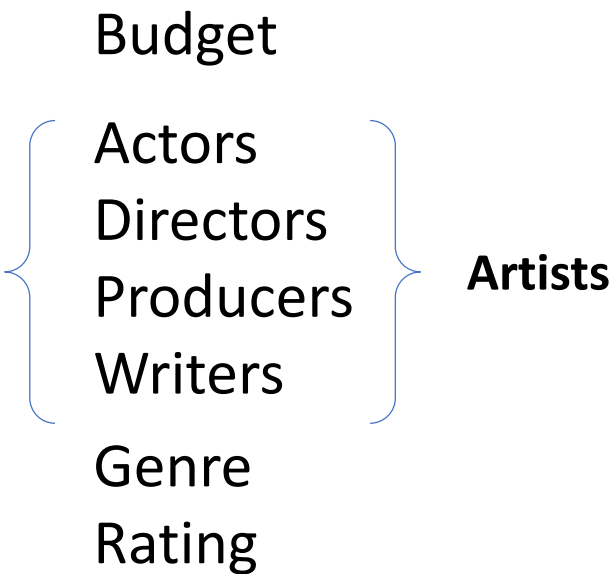
Budget ⟶ Budget

Production company

Distributor

Producers
Actors
Directors ⟶ Actors
Writers   Directors
Composers   Producers
                   Writers

Genre
Rating ⟶ Genre
Release date?   Rating

Data

Budget

Actors
Directors
Producers   } Artists
Writers

Genre
Rating

Artist Scores

Movie	Budget	Actors	Directors	Producers	Writers	Genre	Rating
The Purge	3	42	68	287	43	Drama	R
Sinister	3	17	93	71	221	Horror	R
The Boy Next Door	4	122	53	59	35	Drama	R
The Company You Keep	2	454	31	5	20	Drama	R
Fruitvale Station	1	141	63	16	63	Drama	R
The Last Exorcism Part II	4	28	8	79	40	Horror	PG13

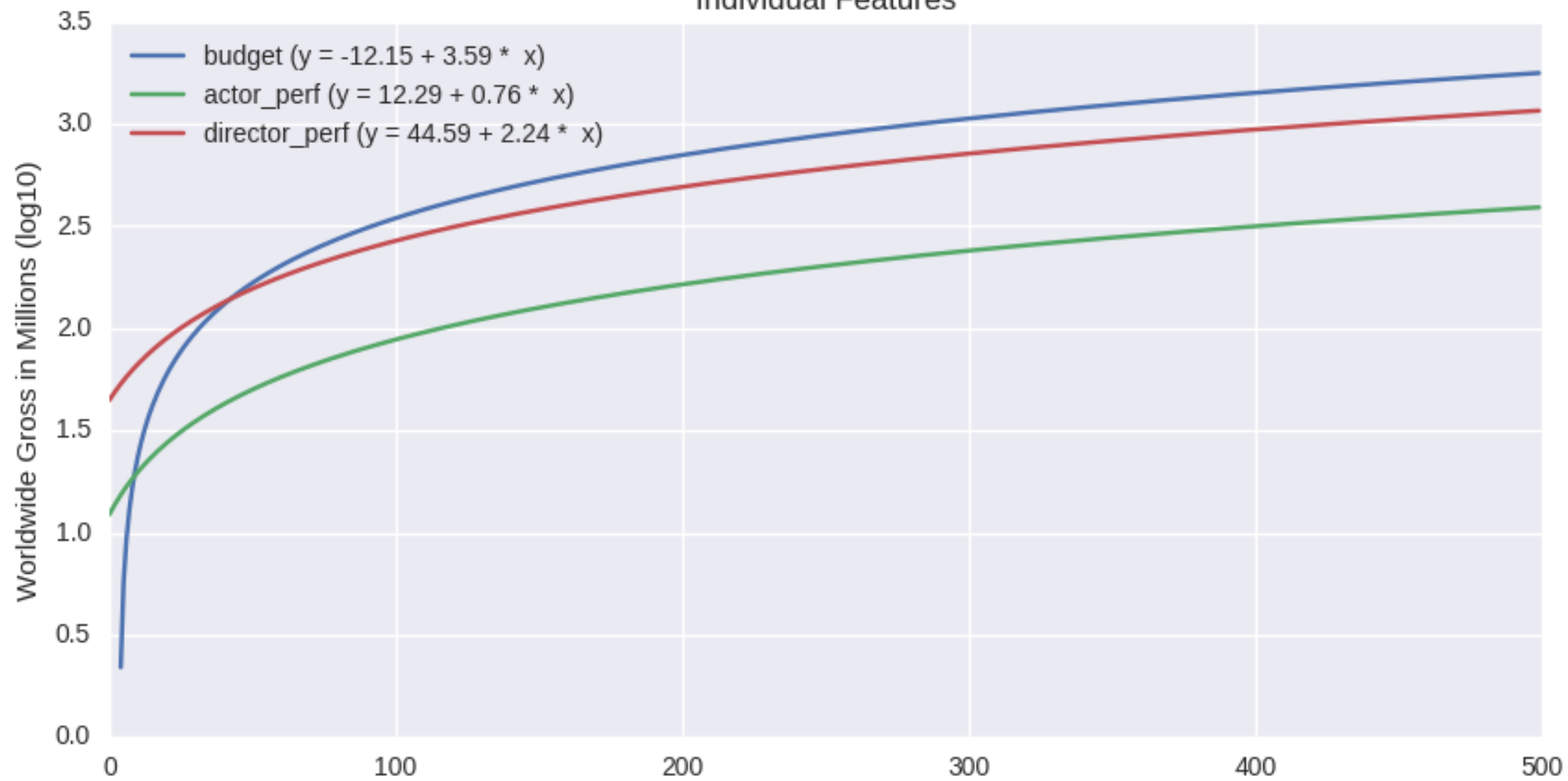| Artists		Average Gross	Total
Director:	Joss Whedon	$278 Million	$278 Million
Writers:	Zak Penn Joss Whedon	$119 Million $151 Million	$270 Million
Actors:	Robert Downey, Jr. Chris Hemsworth Chris Evans Jeremy Renner Mark Ruffalo Scarlett Johansson Clark Gregg Samuel L. Jackson Cobie Smulders Tom Hiddleston Stellan Skarsgard Gwyneth Paltrow Paul Bettany	$75 Million $135 Million $107 Million $118 Million $56 Million $95 Million $101 Million $69 Million - $99 Million $74 Million $56 Million $78 Million	$1.1 Billion
Producers:	Victoria Alonso Louis D'Esposito Jon Favreau Kevin Feige …	$417 Million $297 Million $240 Million $229 Million	$1.8 Billion

Thoughts and Observations

- Franchises with recurring artists are more predictable
- Variables are not independent (e.g., low budget films cannot afford star artists)
- Using a single number for artist score is wrong; it varies over time

Train/Test Split Scores

Individual Features

budget (y = -12.15 + 3.59 * x)
actor_perf (y = 12.29 + 0.76 * x)
director_perf (y = 44.59 + 2.24 * x)

Worldwide Gross in Millions (log10)

Individual Features

budget (y = -12.150 + 3.587 * x)
actor_perf (y = 12.291 + 0.760 * x)
director_perf (y = 44.585 + 2.241 * x)

Worldwide Gross in Millions