The Gross Decline:
Predicting First to Second Week Box Office Decline
...

Aaron Beyer

Goal

- Predict the decline in box office between week 1 and 2
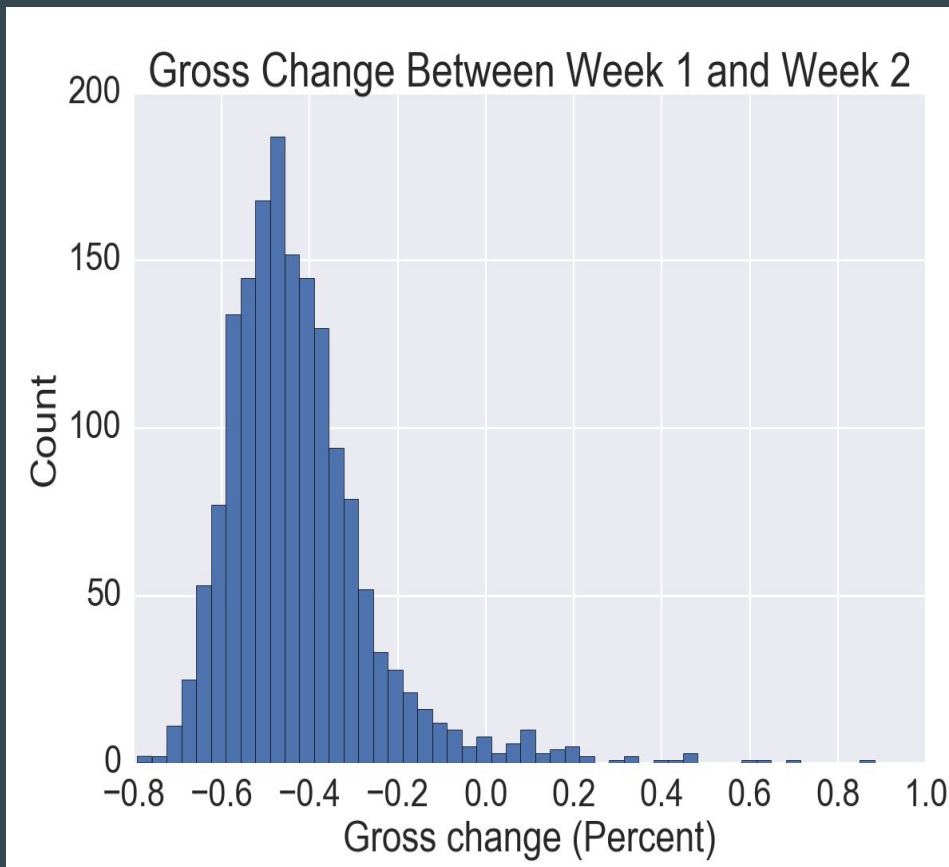- Help movie studios advertise more effectively with limited budgets

Data

- Used Beautiful Soup to obtain 5500 links of top grossing movies on Box Office Mojo
- Scraped basic information about movies and box office information with Scrapy
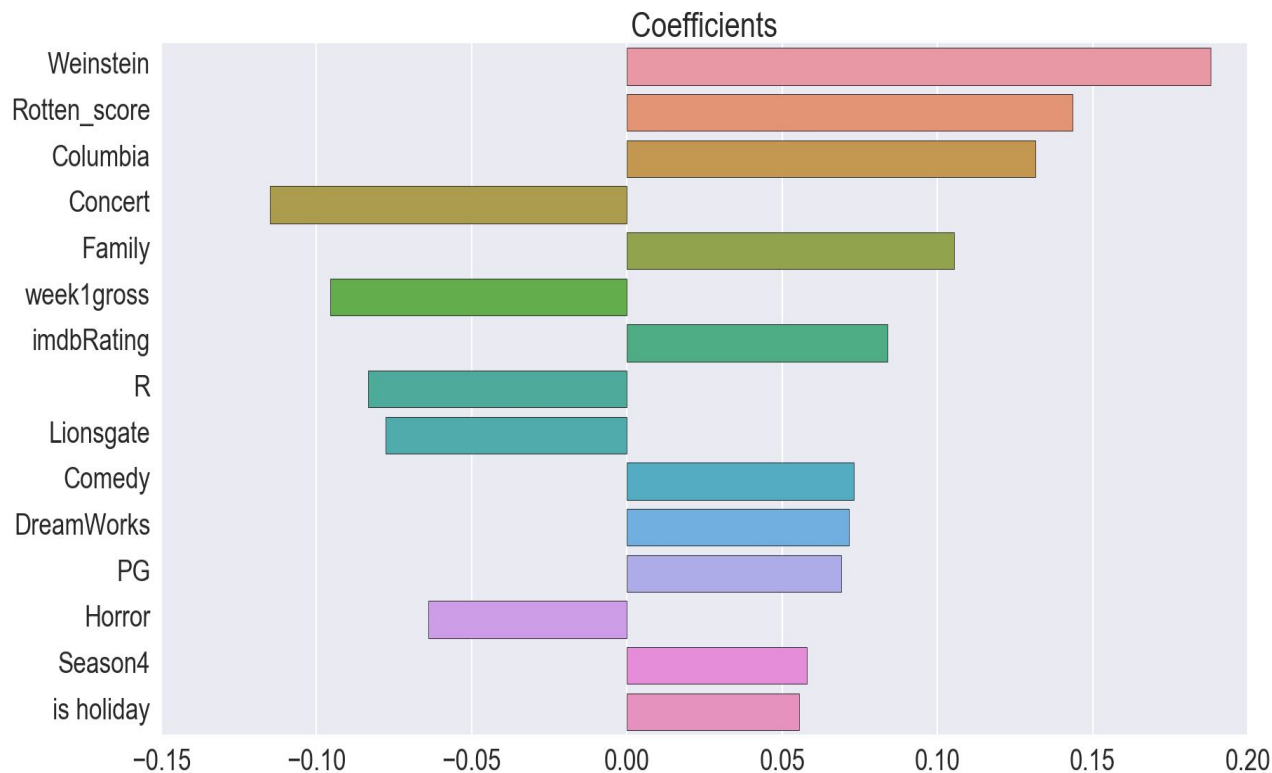- Obtained review info with OMDb API

Data

- Engineered two features:
  - Seasons
  - Is Holiday?
- LOTS of cleaning needed
  - Missing/Misleading data
- Target Variable:
  - Log of Gross Change (percent)

Results

Coefficients

- Used Ridge Regression CV,
  - R2: 0.33, MSE: 0.068

- Also tried Gradient Boosted Regressor
  - R2: 0.40, MSE: 0.071

Fun stuff

Worst
Performing
Movies

(Relative
to week 1)

	Title	grosschange	Rotten_score	imdbRating	Genre	Release Date
496	Friday the 13th (2009)	-0.792	0.59	6.5	Horror	2009-02-13
931	One Direction: This is Us	-0.769	0.63	4.1	Concert	2013-08-30
1333	The Bye Bye Man	-0.750	0.23	3.8	Horror Thriller	2017-01-13
1365	The Devil Inside	-0.732	0.06	4.2	Horror	2012-01-06
1630	The Twilight Saga: New Moon	-0.724	0.28	4.6	Romance	2009-11-20
464	Fifty Shades of Grey	-0.720	0.25	4.1	Romance	2015-02-13
1559	The Purge	-0.717	0.37	5.7	Thriller	2013-06-07
1628	The Twilight Saga: Breaking Dawn Part 2	-0.709	0.49	5.5	Romance	2012-11-16
1731	Valentine's Day	-0.707	0.18	5.7	Romantic Comedy	2010-02-12
379	Doom	-0.701	0.19	5.2	Sci-Fi Horror	2005-10-21
289	Cloverfield	-0.700	0.77	7.0	Action Horror	2008-01-18
430	Endless Love (2014)	-0.696	0.15	6.3	Romance	2014-02-14
43	A Nightmare on Elm Street (2010)	-0.694	0.15	7.5	Horror	2010-04-30
114	Apollo 18	-0.694	0.24	5.2	Sci-Fi Thriller	2011-09-02
157	Batman v Superman: Dawn of Justice	-0.693	0.27	6.7	Action / Adventure	2016-03-25
53	About Last Night (2014)	-0.688	0.69	6.1	Romantic Comedy	2014-02-14
548	Grind	-0.688	0.08	6.0	Sports Comedy	2003-08-15
1000	Predators	-0.687	0.64	6.4	Sci-Fi Horror	2010-07-09
1649	The Wolfman	-0.687	0.34	5.8	Horror	2010-02-12
584	Hellboy II: The Golden Army	-0.686	0.85	7.0	Action Horror	2008-07-11

Next steps

- Bring in more features (hype)
- Try non-linear algorithms /explore non-linear relationships
- Look more at holiday/non-holiday movies
- Add more data (deal with missing variables)

Thank You !