

# Predicting Movie Lifetime Domestic Gross

...

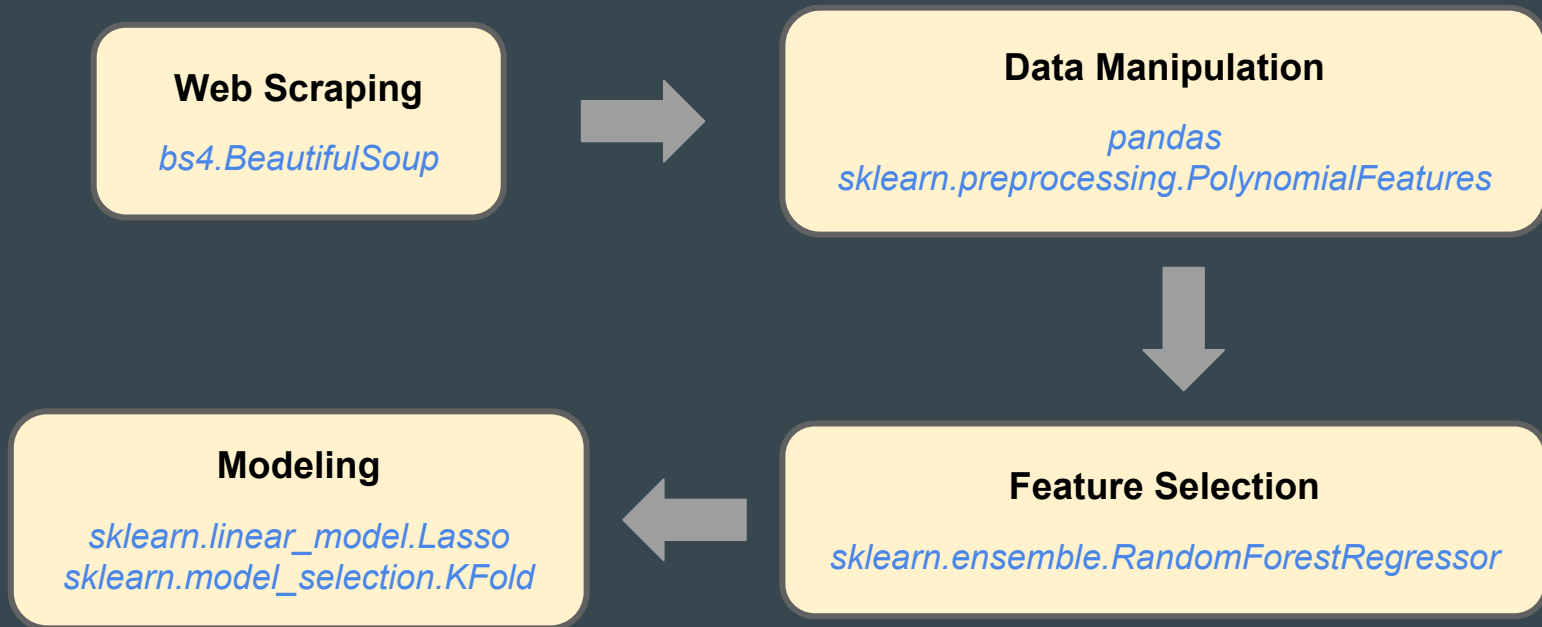
Katharina Huang

2017/04/28

# Aim


- Predict movie lifetime domestic gross with movie and player metadata
- Learn web scraping and linear regression

# Method



# Web Scrapping

Source:  
[boxofficemojo.com](http://boxofficemojo.com)



## The Conjuring

Domestic Total Gross: <b>\$137,400,141</b>	
Distributor: <b>Warner Bros. (New Line)</b>	Release Date: <b>July 19, 2013</b>
Genre: <b>Horror</b>	Runtime: <b>1 hrs. 52 min.</b>
MPAA Rating: <b>R</b>	Production Budget: <b>\$20 million</b>

Summary

Daily

Weekend

Weekly

Foreign

Similar Movies

### Total Lifetime Grosses

<b>Domestic:</b>	<b>\$137,400,141</b>	<b>43.2%</b>
+ Foreign:	\$180,600,000	56.8%

= **Worldwide: \$318,000,141**

### Domestic Summary

Opening Weekend:	\$41,855,326
(#1 rank, 2,903 theaters, \$14,418 average)	
% of Total Gross:	30.5%
<a href="#">View All 15 Weekends</a>	
Widest Release:	3,115 theaters
Close Date:	October 31, 2013
In Release:	105 days / 15 weeks

### The Players

Director:	James Wan
Writers:	Carey Hayes Chad Hayes
Actors:	Vera Farmiga Patrick Wilson Lili Taylor Ron Livingston Joey King
Producers:	Rob Cowan Tony DeRosa-Grund Peter Safran
Composer:	Joseph Bishara

### Related Stories

- 9/6/13 Grading Mojo's Summer Forecast
- 9/6/13 Grading Mojo's Summer Forecast (cont.)
- 8/11/13 Weekend Report: Audiences Travel to 'Elysium' Weekend
- 8/4/13 Weekend Report: '2 Guns' Takes Top Spot, 'Smile' Holds
- 8/2/13 'Despicable' Drives July to Second-Highest Month
- 7/28/13 Weekend Report: 'Wolverine' Bleeds, But Still Holds
- 7/21/13 Weekend Report: 'Conjuring' Haunts First With
- 7/18/13 Forecast: 'Conjuring' Should Scare Off 'Turbo,'

# Web Scrapping

Source:  
[boxofficemojo.com](http://boxofficemojo.com)

## James Wan

Director

Producer

Writer

Date	Title (click to view)	Studio	Lifetime Gross / Theaters		Opening / Theaters		Rank
6/10/16	<b>The Conjuring 2</b>	WB (NL)	\$102,470,008	3,356	\$40,406,314	3,343	3
4/3/15	<b>Furious 7</b>	Uni.	\$353,007,020	4,022	\$147,187,040	4,004	1
9/13/13	<b>Insidious Chapter 2</b>	FD	\$83,586,447	3,155	\$40,272,103	3,049	4
7/19/13	<b>The Conjuring</b>	WB (NL)	\$137,400,141	3,115	\$41,855,326	2,903	2
4/1/11	<b>Insidious</b>	FD	\$54,009,150	2,419	\$13,271,464	2,408	6
8/31/07	<b>Death Sentence</b>	Fox	\$9,534,258	1,823	\$4,231,321	1,822	8
3/16/07	<b>Dead Silence</b>	Uni.	\$16,809,076	1,806	\$7,842,725	1,805	7
10/29/04	<b>Saw</b>	Lions	\$56,815,693	2,467	\$18,276,468	2,315	5

Lifetime Gross Total (8): \$813,631,793

**Average: \$101,703,974**

Opening Gross Average (8): \$39,167,845 (Wide Releases Only)

# Data Manipulation

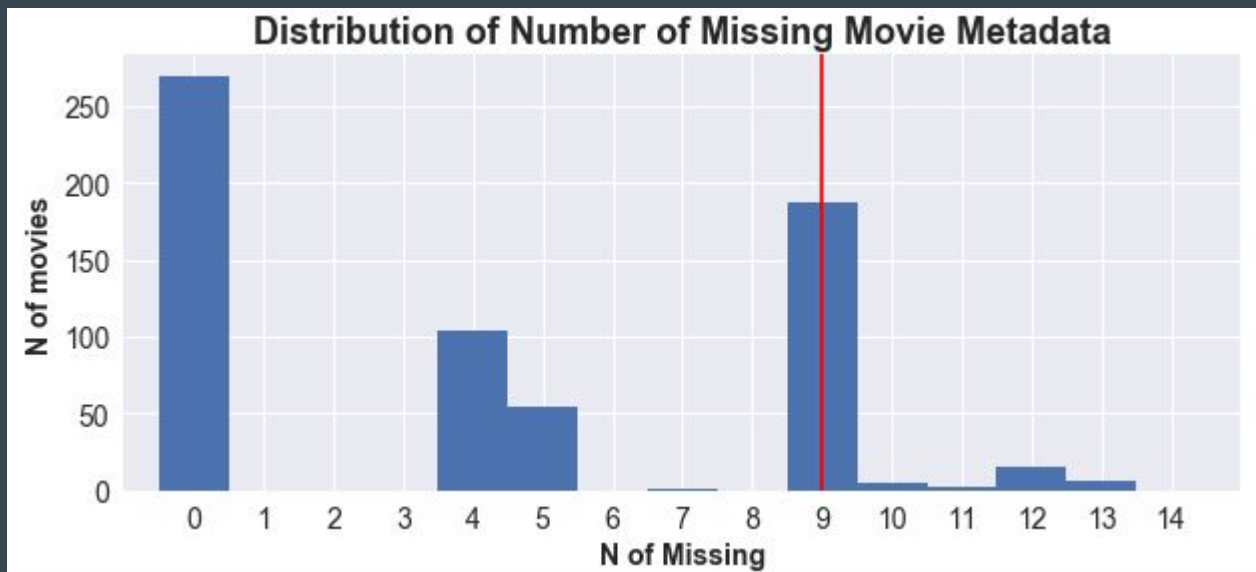
- Filter in 5 genres
  - Total 635 movies

**N of Movies and No. 1 Picture by Genre**

Genre	N of Movies	No. 1 Picture	
		Title	Gross (USD in millions)
3D	296	Star Wars: The Force Awakens	936.7
..	..	..	..
Mouse/Rat - Featured	25	Alice in Wonderland	334.2
Drama - Summer	145	Forrest Gump	330.3
Sports - Football (Featured)	14	Forrest Gump	330.3
..	..	..	..
Werewolf	28	The Twilight Saga: Eclipse	300.5
Horror - Supernatural	147	Sixth Sense	293.5
Crime Time	123	Inception	292.6
Heist / Caper	105	Inception	292.6
Christmas	67	Home Alone	285.8
..	..	..	..
Pregnancy	34	Breaking Dawn Part 1	281.3
Comedy - Sequel (Live Action)	112	Meet the Fockers	279.3
Comedy - R-Rated Youth	115	The Hangover	277.3
Travelogue - Las Vegas	23	The Hangover	277.3
..	..	..	..
Sundance 2012	28	Beasts of the Southern Wild	12.8

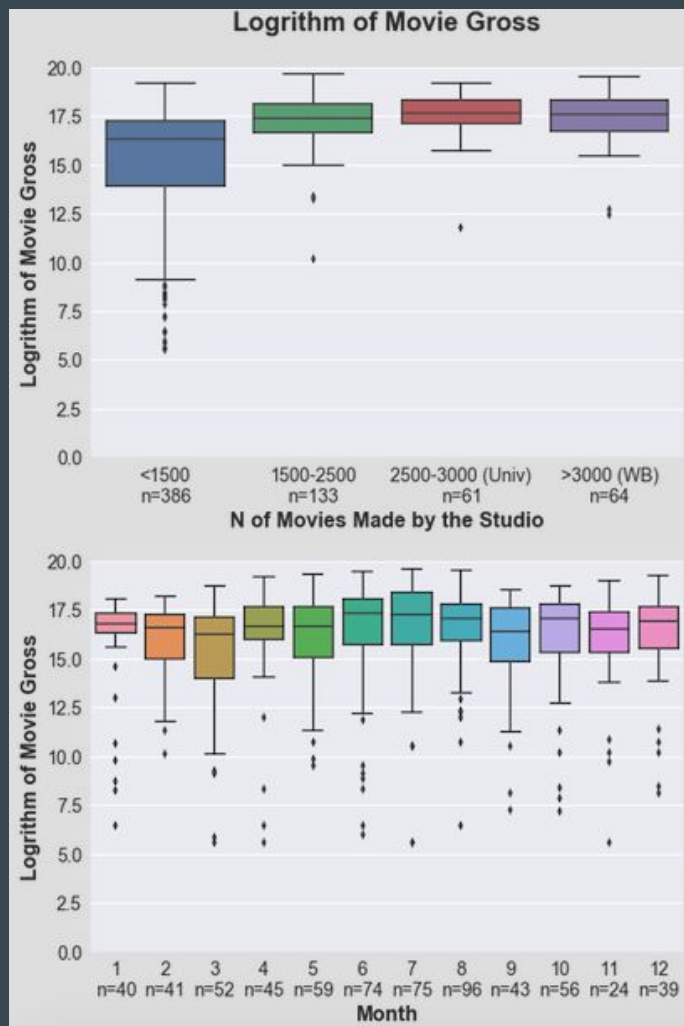
# Data Manipulation

- Systematically more missing in less popular movies
  - Exclude movies missing at least 9 features (N of movies 635 => 429)
  - Impute missing revenue and movie ranking-related features with minimum values



# Data Manipulation

- Exclude "future" variables
  - Lifetime foreign and worldwide gross
  - Number of total theaters
  - Percentage of opening weekend gross
  - Movie ranking
- Take log on revenues and budget
- Extract new features from release date
  - Year
  - Month
  - Day of week
- Add polynomial terms (*see plot*)





# Data Manipulation

- Aggregate player historical information
  - 1 year
  - 3 years
  - 5 years
  - All years
- Filter in 4 roles
  - Director
  - Producer
  - Writer
  - Actor

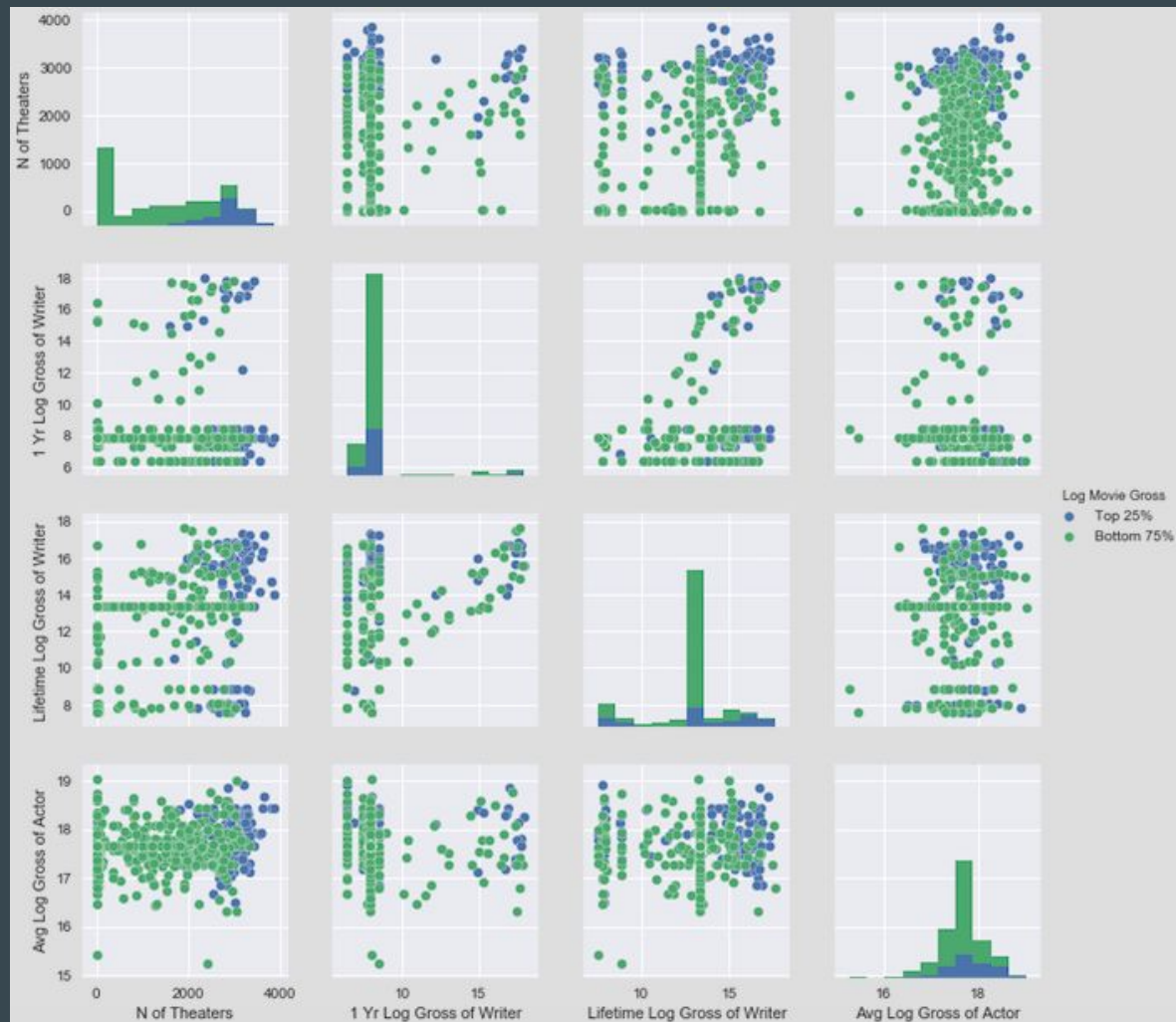
James Wan							
Director		Producer		Writer			
Date	Title (click to view)	Studio	Lifetime Gross / Theaters		Opening / Theaters		Rank
6/10/16	The Conjuring	Univ.	\$2,470,008	3,356	\$40,406,314	3,343	3
4/3/15	Furious 7	Uni.	\$353,007,0				1
9/13/13	Insidious Chapter 2	FD	\$83,586,447	3,158	\$40,272,103	3,049	4
7/19/13	The Conjuring	WB (NL)	\$137,400,141	1,715	\$41,855,326	2,903	2
4/1/11	Insidious	FD	\$54,009,150	2,419	\$13,271,464	2,408	6
8/31/07	Death Sentence	Fox	\$9,534,258	1,823	\$4,231,321	1,822	8
3/16/07	Dead Silence	Uni.	\$16,809,076	1,806	\$7,842,725	1,805	7
10/29/04	Saw	Lions	\$56,815,693	2,467	\$18,276,468	2,315	5

Note: Titles in grey are cameo or bit parts and not counted in totals and averages.

Lifetime Gross Total (8): \$813,631,793  
**Average: \$101,703,974**  
Opening Gross Average (8): \$39,167,845 (Wide Releases Only)

# Data Manipulation

- Add interaction



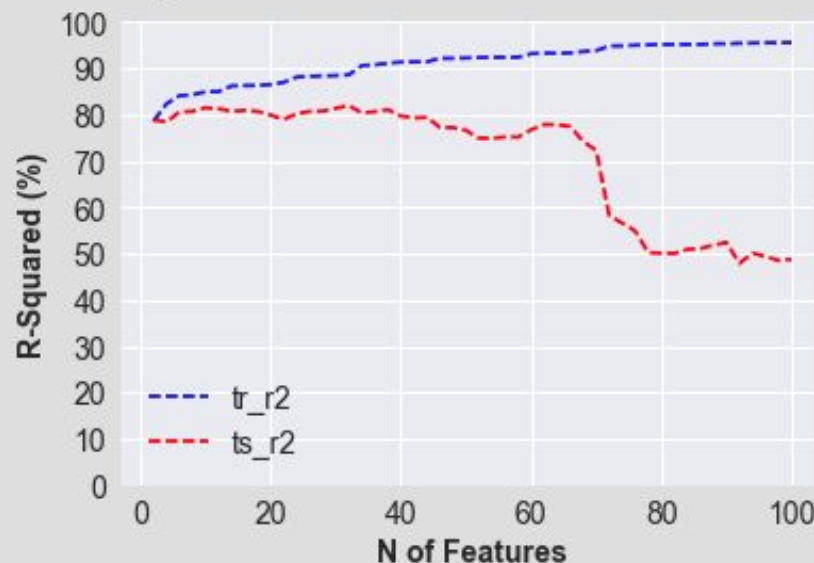
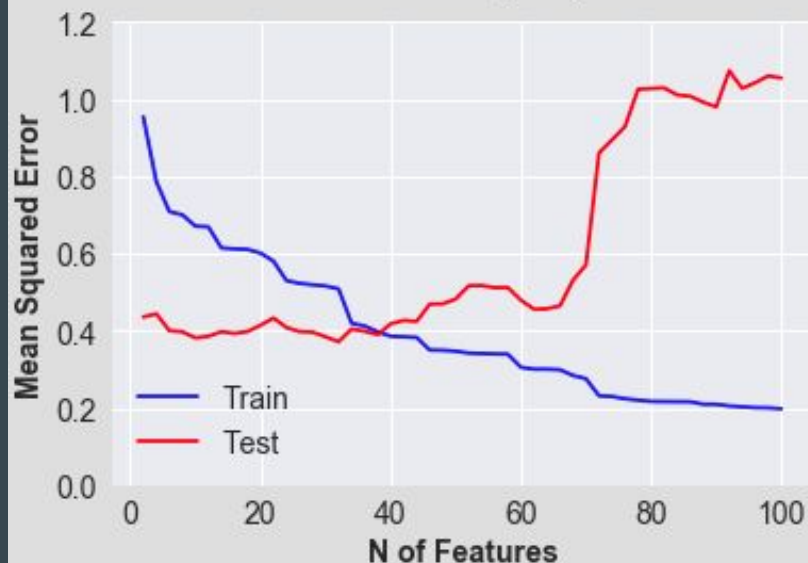
# Feature Selection

14,195 features

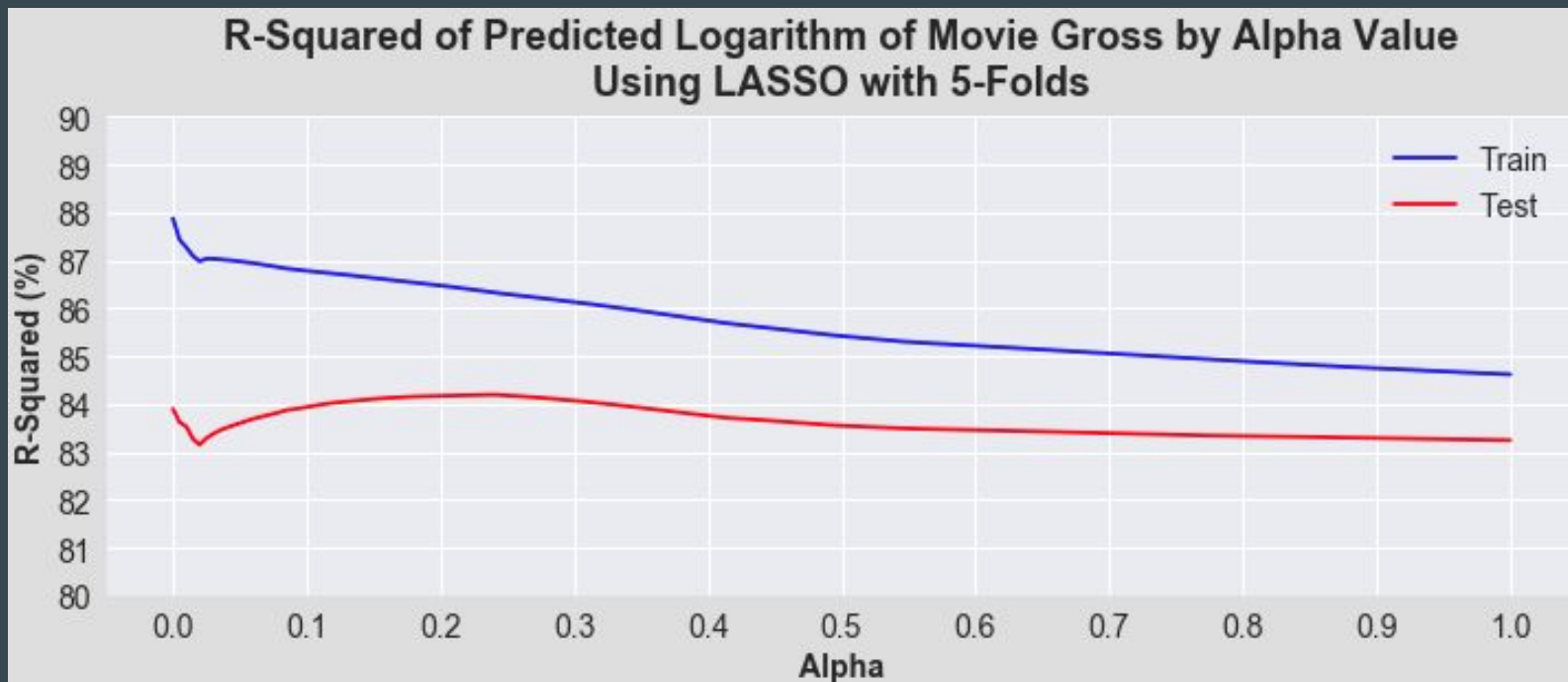
32 features



Errors of Predicted Logarithm of Movie Gross by Number of Features  
Using Importance Determined by Random Forest



# Regularization Tuning



# Results

	Training	Holdout
Final Model	Final Model: LASSO (alpha=0.24)	
Label	Log of lifetime movie gross	
N of Features	32	
N of Movies	343	86
MSE	0.5189	0.5143
R-Squared	86.1033%	85.6635%

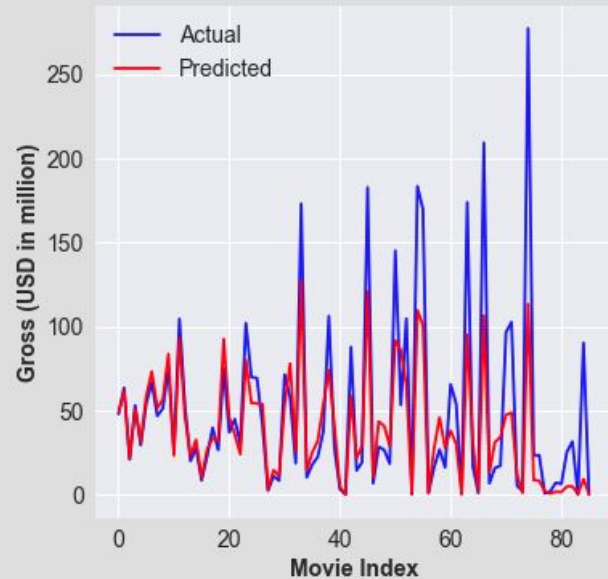
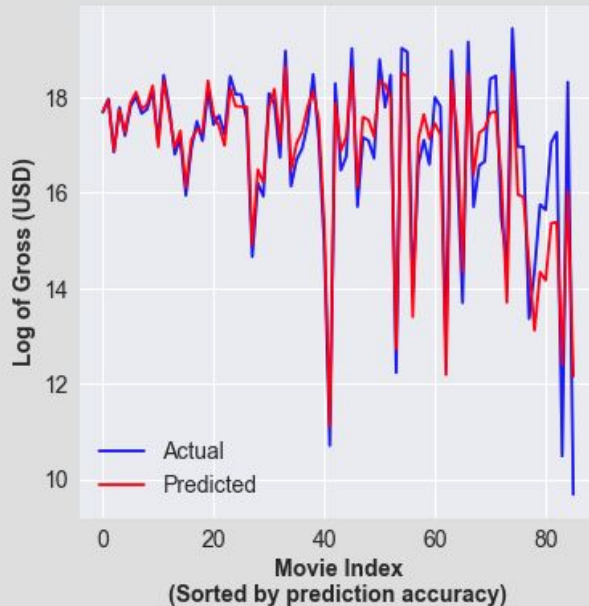
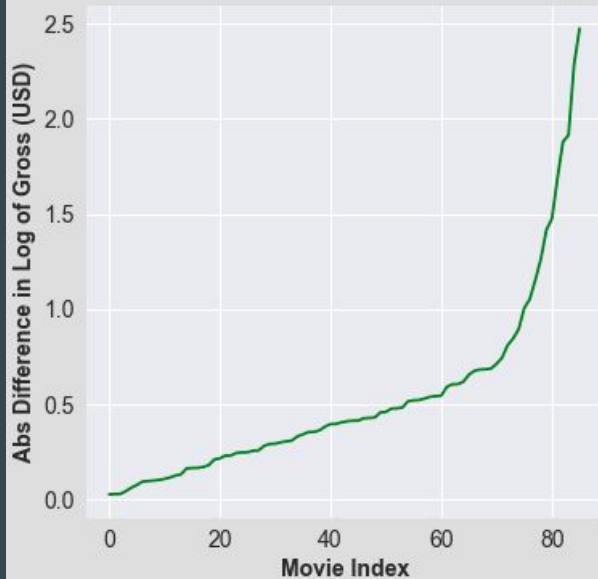
## Top 7 Features with the Strongest Impact

Rank	Feature	Est. Beta Coefficient
1	Interaction between log of opening weekend gross and studio ranking	0.001434901
2	Interaction between log of opening weekend gross and release year	0.000362029
3	Interaction between log of opening weekend gross and runtime (mins)	0.000338021
4	Interaction between log of mean opening weekend gross per theater and release year	0.000117422
5	Interaction between log of opening weekend gross and N of opening weekend theaters the director was seen at in the past year	0.000000223
6	Interaction between N of opening weekend theaters and N of movies by the studio	0.000000093
7	Interaction between N of opening weekend theaters and studio ranking	-0.000000001



# Results

Compare Actual and Predicted Lifetime Movie Gross



# Future Work

- Predict opening weekend domestic gross with historical data
- Scale up number of movies

Q & A



# Appendix