

Project Luther

...

Analysis of Data from boxofficemojo.com
Prashant Tatineni

Project Summary

Can we predict the success of a movie sequel based on the performance of previous films in the franchise?

This project applies a Linear Regression model to movies from [boxofficemojo.com](https://www.boxofficemojo.com) franchise data to find expected gross domestic earnings of a movie sequel.

Summary of Steps

1. Retrieve Data from boxofficemojo.com.
2. Clean up data and reduce to features and target.
3. Run Linear Regression model.
4. Review model performance.

Retrieving data from boxofficemojo.com

Franchise Index					
Series	Brands				
Franchise (click to view chart)	Total Gross	# Movies / Average	# 1 Picture		Gross
James Bond	\$2,112.9	25	\$84.5	Skyfall	\$304.4
Marvel Cinematic Universe	\$4,234.0	14	\$302.4	The Avengers	\$623.4
Star Trek	\$1,401.0	13	\$107.8	Star Trek	\$257.7
Friday the 13th	\$380.6	12	\$31.7	Freddy Vs. Jason	\$82.6
Batman	\$2,232.0	10	\$223.2	The Dark Knight	\$533.3
Halloween	\$308.5	10	\$30.9	Halloween (2007)	\$58.3
J.K. Rowling's Wizarding World	\$2,624.0	10	\$262.4	Harry Potter / Deathly Hallows (P2)	\$381.0
The Pink Panther	\$282.6	10	\$28.3	The Pink Panther (2006)	\$82.2
Harry Potter	\$2,391.8	9	\$265.8	Harry Potter / Deathly Hallows (P2)	\$381.0
Nightmare on Elm Street	\$370.5	9	\$41.2	Freddy Vs. Jason	\$82.6
Star Wars	\$3,376.0	9	\$375.1	Star Wars: The Force Awakens	\$936.7

Key Data

- Adjusted Gross
- Release Date
- Number of Theaters

Drop re-issued films

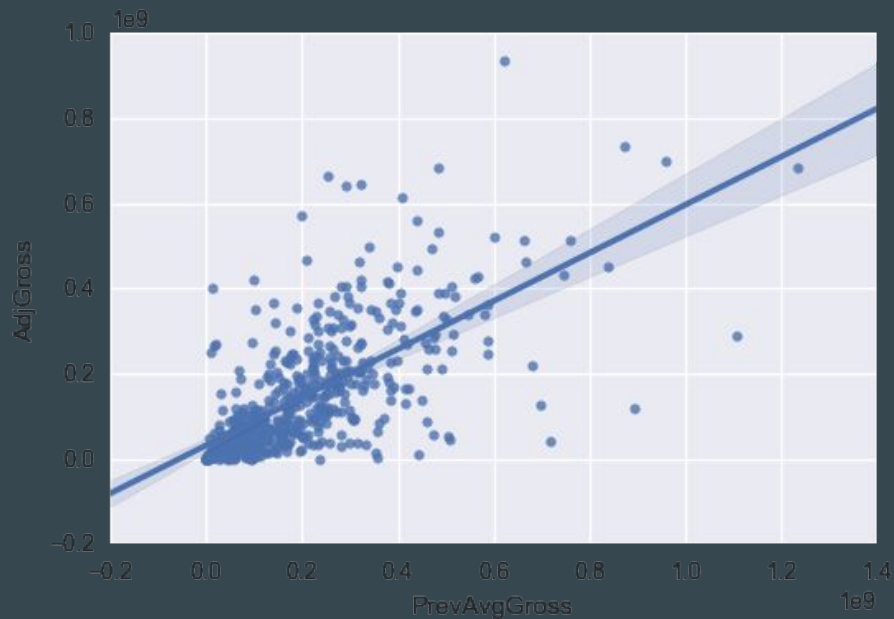
Adjusted for Ticket Price Inflation					
Rank	Title (click to view)	Studio	Adjusted Gross	Unadjusted Gross	Release
1	Star Wars	Fox	\$1,234,649,200	\$307,263,857	5/25/77
2	Star Wars: The Force Awakens	BV	\$935,195,600	\$936,662,225	12/18/15
3	Star Wars: Episode I - The Phantom Menace	Fox	\$733,743,200	\$431,088,295	5/19/99
4	Return of the Jedi	Fox	\$700,506,300	\$252,583,617	5/25/83
5	The Empire Strikes Back	Fox	\$682,158,400	\$209,398,025	5/21/80
6	Rogue One: A Star Wars Story	BV	\$515,381,200	\$521,359,922	12/16/16
7	Star Wars: Episode III - Revenge of the Sith	Fox	\$513,157,600	\$380,270,577	5/19/05
8	Star Wars: Episode II - Attack of the Clones	Fox	\$449,906,100	\$302,191,252	5/16/02
9	Star Wars (Special Edition)	Fox	\$260,551,300	\$138,257,865	1/31/97
10	The Empire Strikes Back (Special Edition)	Fox	\$127,390,000	\$67,597,694	2/21/97
11	Return of the Jedi (Special Edition)	Fox	\$85,690,500	\$45,470,437	3/14/97
12	Star Wars: Episode I - The Phantom Menace (in 3D)	Fox	\$47,456,200	\$43,456,382	2/10/12
13	Star Wars (Re-issue)	Fox	\$45,534,000	\$15,476,285	8/13/82
14	Star Wars: The Clone Wars	WB	\$42,360,400	\$35,161,554	8/15/08
15	The Empire Strikes Back (Re-issue)	Fox	\$39,061,000	\$13,276,241	11/19/82
16	Return of the Jedi (Re-issue)	Fox	\$27,417,100	\$11,252,123	3/29/85
17	Attack of the Clones: The IMAX Experience (IMAX)	Fox	\$12,596,200	\$8,485,488	11/1/02

Target and Feature variables

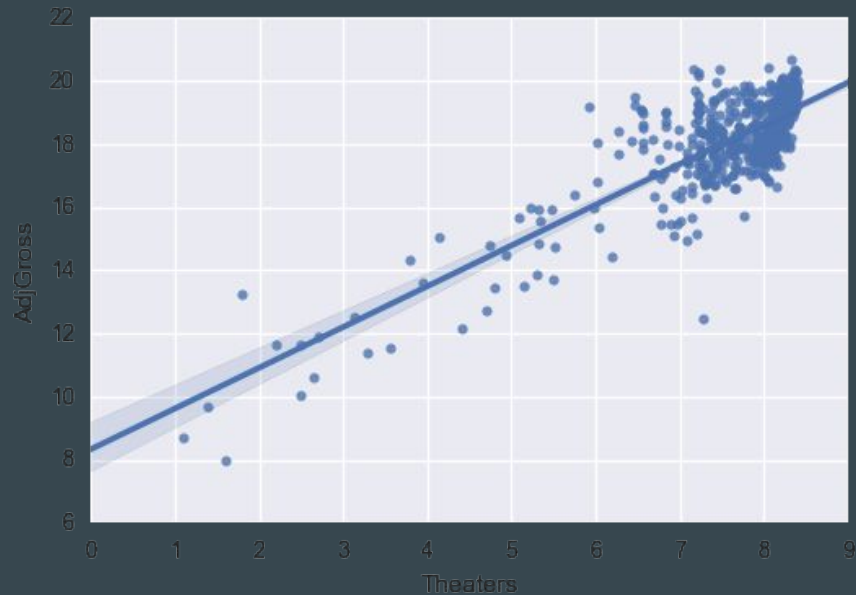
	Title	Franchise	AdjGross	Theaters	SeriesNum	PrevAvgGross	FirstGross	DaysSinceFirstFilm	DaysSincePrevFilm
1	3 Ninjas Kick Back	./chart/?id=3ninjas.htm	24416300	2043.0	2.0	6.044640e+07	60446400	637	637
2	3 Ninjas Knuckle Up	./chart/?id=3ninjas.htm	822200	52.0	3.0	4.243135e+07	60446400	945	308
3	3 Ninjas: High Noon at Mega Mountain	./chart/?id=3ninjas.htm	693100	120.0	4.0	2.856163e+07	60446400	2072	1127
5	Along Came a Spider	./chart/?id=alexcross.htm	113211300	2573.0	2.0	1.140667e+08	114066700	1281	1281
6	Alex Cross	./chart/?id=alexcross.htm	27818900	2541.0	3.0	1.136390e+08	114066700	5495	4214
8	Aliens	./chart/?id=alien.htm	198554200	1454.0	2.0	2.720611e+08	272061100	2611	2611
9	Alien 3	./chart/?id=alien.htm	115625600	2227.0	3.0	2.353076e+08	272061100	4746	2135
10	Alien Resurrection	./chart/?id=alien.htm	90072400	2449.0	4.0	1.954136e+08	272061100	6760	2014
11	Alien Vs. Predator	./chart/?id=alien.htm	111826300	3401.0	5.0	1.690783e+08	272061100	9212	2452
12	Aliens Vs. Predator - Requiem	./chart/?id=alien.htm	51238500	2617.0	6.0	1.576279e+08	272061100	10441	1229
13	Prometheus	./chart/?id=alien.htm	135344100	3442.0	7.0	1.398964e+08	272061100	12068	1627

Potential Correlations

Linear Plot: AdjGross vs. PrevAvgGross

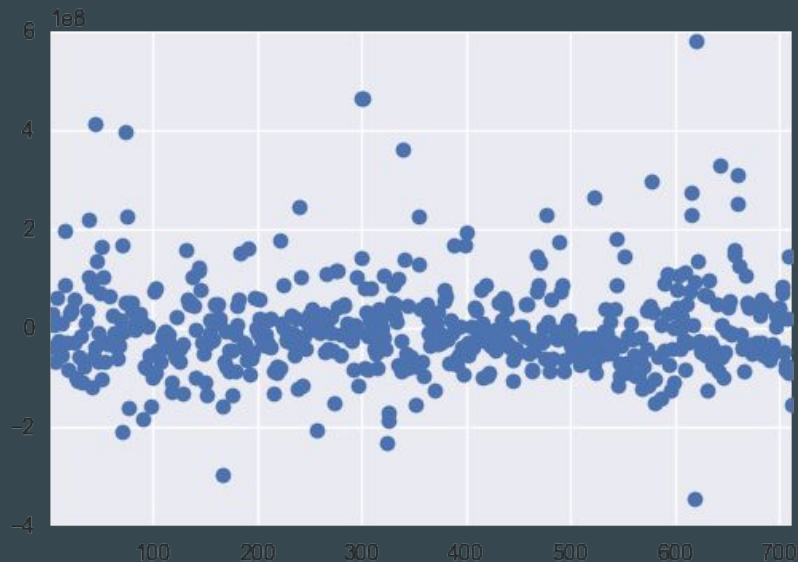


Log-Log Plot: AdjGross vs. Theaters

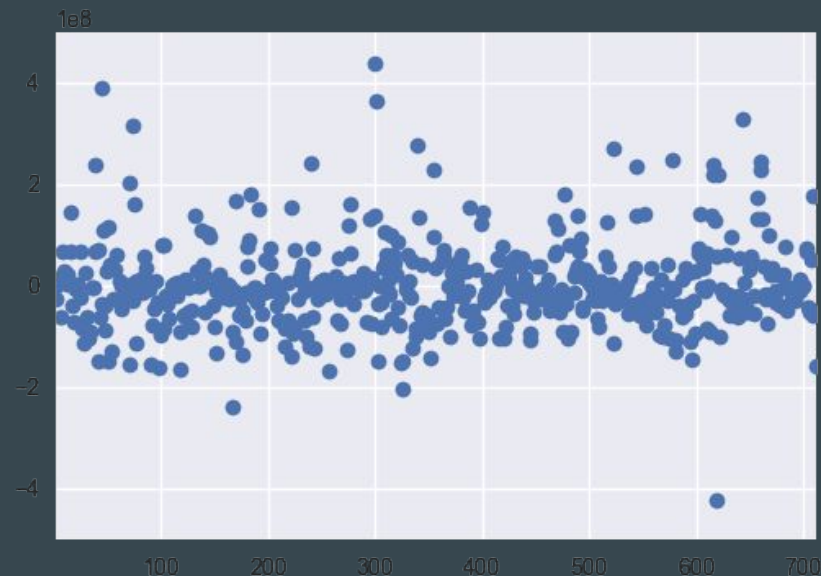


Initial Regression models

Linear, Adj. R-squared = 0.548

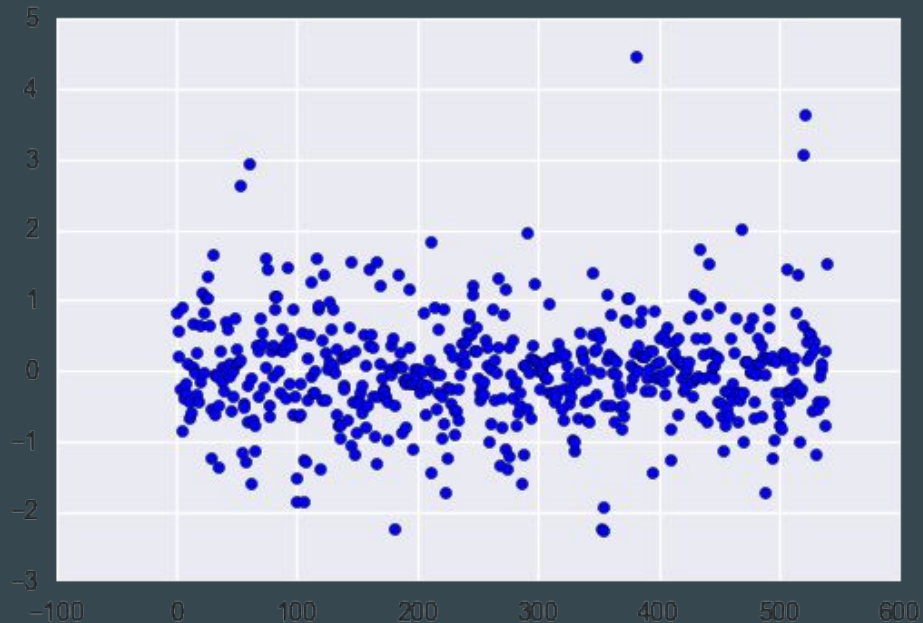


Polynomial, Adj. R-squared = 0.626



Tests and Transforms

- Tested for Heteroskedasticity
- Applied Box-Cox Transformation to X data and scaled
- Applied log to Y
- Used ElasticNetCV to regularize.



Applying this model to the newest sequels...



Release: 1/27/17

Predicted Gross: \$3,550,700

Resident Evil: Apocalypse	SGem	\$71,319,300
Resident Evil: Afterlife	SGem	\$65,422,900
Resident Evil: Extinction	SGem	\$63,678,900
Resident Evil	SGem	\$59,730,700
Resident Evil: Retribution	SGem	\$47,080,800
Resident Evil: The Final Chapter	SGem	\$16,634,600



Release: 1/20/17

Predicted Gross: \$752,268,200

XXX	SonR	\$211,558,500
XXX: State of the Union	SonR	\$36,265,100
xXx: The Return of Xander Cage	Par.	\$35,814,600



Release: 1/6/17

Predicted Gross: \$149,625,700

Underworld: Evolution	SGem	\$82,299,000
Underworld	SGem	\$74,535,000
Underworld Awakening	SGem	\$68,847,600
Underworld: Rise of the Lycans	SGem	\$53,108,600
Underworld: Blood Wars	SGem	\$29,756,100

If time permitted...

- > More in depth review of data and outliers
- > Further consideration of what affects sequel success
- > Additional work on improving model fit