COMCAST

NBCUniversal Analytics Challenge

Jordan Dunn, Tee Gildemeister, Andrew Hughes

Understanding the seemingly unexpected behavior of the movie market can be difficult. However, in-depth analyses reveal that some of the complex patterns in consumer trends are in fact expected. The following report details three key investigations, each of which elucidate real patterns in market behavior and the attributes that lead to success.

Actor & Director Power Pairings

One of the most impactful attributes in a movie's U.S. box office return is if it has a strong actor and director pairing. These pairings are measured using a relationship matrix that identifies actors and directors who have worked together on 3 or more films, then finding the average return on investment for the movies these pairings have made together.

Between 2000 and 2017, movies with strong actor-director pairings made an average return on investment of $23,656,200. In comparison, the average return for all movies in our sample was $6,220,779. This indicates that strong actor-director pairings bring in 3.8 times higher returns than the average movie. Where this is most notable is in the comedy and adventure genres. Of movies with strong pairings that made a positive monetary return, 50 percent fell under comedy or adventure. Based on this, it is accurate to assume that a movie in either comedy or adventure with a strong actor-director combination will be certain to yield higher-than-average returns.

Multiple Regression Analysis & Predictive Model

To tease out the most prevalent explanatory factors for a movie's success in the U.S. box office, a series of multiple regressions were devised and executed with the aim of creating predictive models. This series included small budget movies, large budget movies, and all movies in the provided data sample released between 2000 and 2016. Each of these, in their original form, contained a total of 44 categorical and continuous variables. These included, for each movie: run-time; MPAA rating, genre (including genre combinations); the sum of the number of movies for all billed actors; the average revenue per movie of each billed actor; various economic and weather metrics; the month of release; and budget. Testing revealed that few of these factors had statistical significance. The models were refined to the features listed in the data visuals section. Each model revealed that the largest determinant of box office revenue, for small and large budget movies, as well as the overall sample of films, is a movie's budget. The surprising result of these analyses is that few factors other than a movie's budget correlated significantly with box office success – regardless of the budget category the movie falls into.

The regression model with the greatest accuracy is the total sample model. With an adjusted R-squared of 0.61, it is able to predict any movie's box office revenue with a residual standard error less than 1 standard deviation of the mean. As a result, the authors of this study believe it has the potential to be a powerful tool in predicting box office success.

Sequels, Remakes, and Series

Between 2010 and 2017, about 58 percent of the top three grossing movies in each quarter were either sequels, remakes, or part of a series. While these three categories account for only 8 percent of the total number of movies made, they capture 43 percent of total quarterly revenue. It is clear from this data that sequels, remakes, and series garner a disproportionately high amount of market share.

As a result, it is necessary to remain cautious when releasing an original film in the same quarter as any of these three types of movies. However, anything after the second sequel typically experiences a diminishing return. For example, third sequels make an average of 17 percent less revenue than second sequels. Based on this, one can assume that this 17 percent of revenue is potentially available market share. Therefore, the likelihood of gaining greater market share exists when releasing in the same quarter as third sequels.

Conclusion

Though box office results are often described as within the realm of the unexpected, the investigations undertaken and described have revealed meaningful patterns in the factors that can lead to box office success. In understanding the highlighted relationships between U.S. box office revenue, strong actor-director combinations; sequels, remakes and series; and budget, it is possible to make sense of the order that exists within this market. By turning the unexpected into the expected, one can assuredly realize greater success in the box office.

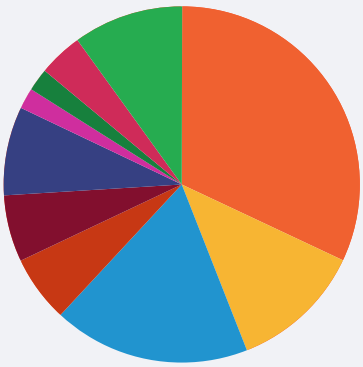
DATA VISUALS

AVERAGE ROI
2000-2017

$23 million
Strong Actor-Director Pairing

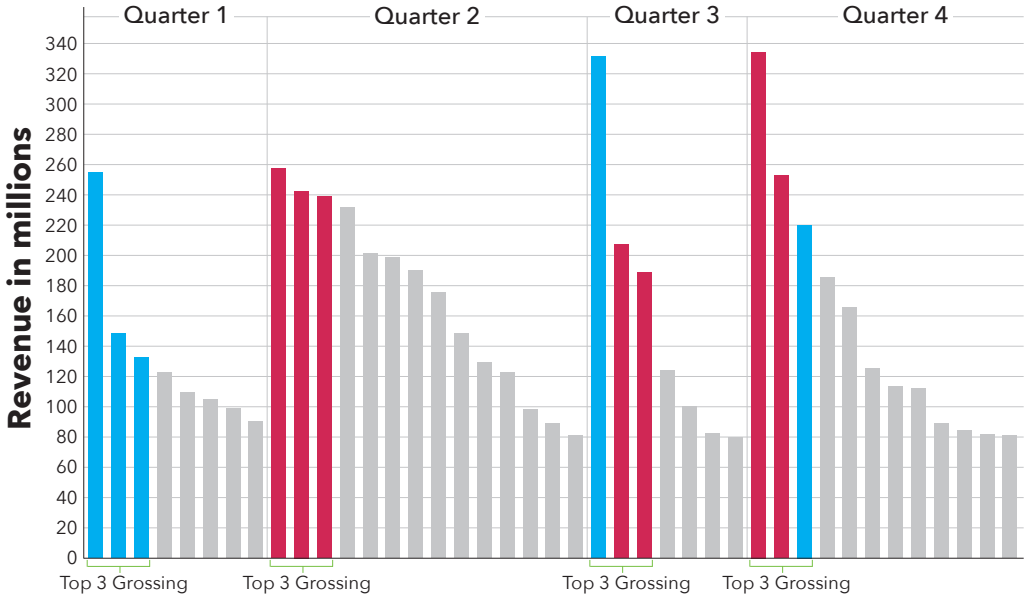
$6 million
Average Movie

GENRES WITH STRONG ACTOR-DIRECTOR PAIRINGS



- **Comedy** 32%
- **Action** 12%
- **Adventure** 18%
- **Crime** 6%
- **Drama** 6%
- **Fantasy** 8%
- **Horror** 2%
- **Mystery** 2%
- **Romance** 4%
- **Thriller** 10%

EXAMPLE: TOP GROSSING MOVIES BY QUARTER 2014
Above 80m in Revenue

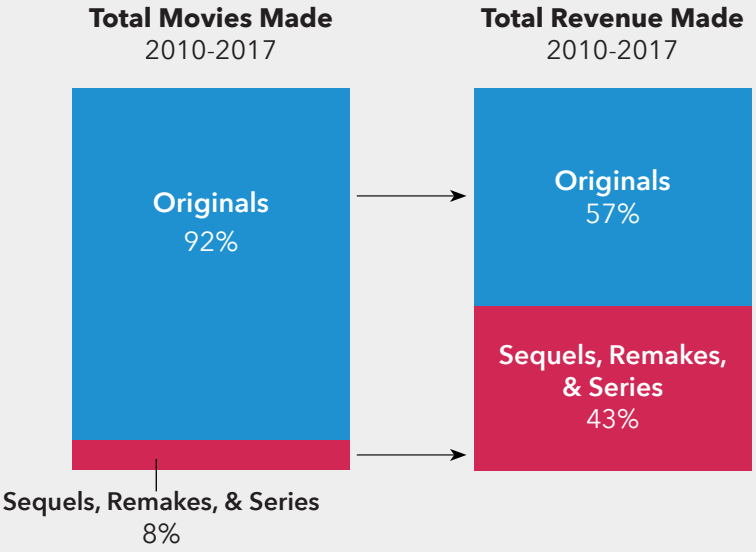


- Sequel, Remake, or Series
- Originals

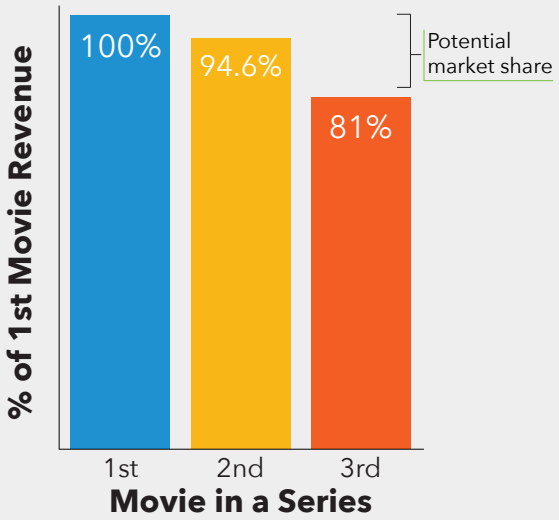
OVERALL
Top three grossing movies by quarter

58% sequels, remakes, or series

AVERAGE PERCENTAGE PER QUARTER



PERCENT DIFFERENCE BETWEEN SEQUEL REVENUES



REGRESSION ANALYSIS

BIG BUDGET FILMS

Over $100 million

Feature	Coefficient	Standard Error	T-Statistic	P-Value
Intercept	179311183	9445376	18.984	2.00e-16
Budget	63116485	8659608	7.289	8.06e-16
Runtime	16289880	7556406	2.156	0.0324

R^2: 0.2828
Adjusted R^2: 0.2753
P-value: 1.63e-14
Residual Standard Error: 114,400,000
Degrees of Freedom: 191
F-Statistic: 37.66

SMALL BUDGET FILMS

Under $40 million

Feature	Coefficient	Standard Error	T-Statistic	P-Value
Intercept	19382328	1036611	18.698	2.00e-16
Budget	13939040	639901	21.783	2.00e-16
Crime	-5261685	1301441	-4.043	5.69e-05
Comedy	4098892	1283416	3.194	0.00145
Thriller	-13895839	4433473	-3.134	0.00177
Runtime	2684520	868000	3.093	0.00204
PG-13	2714359	1276215	2.127	0.03368
Mystery	14685530	7239542	2.029	0.04278

R^2: 0.3945
Adjusted R^2: 0.3902
P-value: 2.20e-16
Residual Standard Error: 18,920,000
Degrees of Freedom: 980
F-Statistic: 91.22

TOTAL SAMPLE

All 1,500 films

Feature	Coefficient	Standard Error	T-Statistic	P-Value
Intercept	61351328	2503743	24.504	2.00e-16
Budget	59920712	1614332	37.118	2.00e-16
Action	-15826485	3239199	-4.886	0.00000114
Drama	-13606376	2990884	-4.549	0.00000581
Runtime	6936115	1617389	4.288	0.00001912
PG-13	6773611	2659456	2.547	0.011

R^2: 0.6181
Adjusted R^2: 0.6168
P-value: 2.20e-15
Residual Standard Error: 49,630,000
Degrees of Freedom: 1,525
F-Statistic: 493.6