



Project Luther

Too many movies?

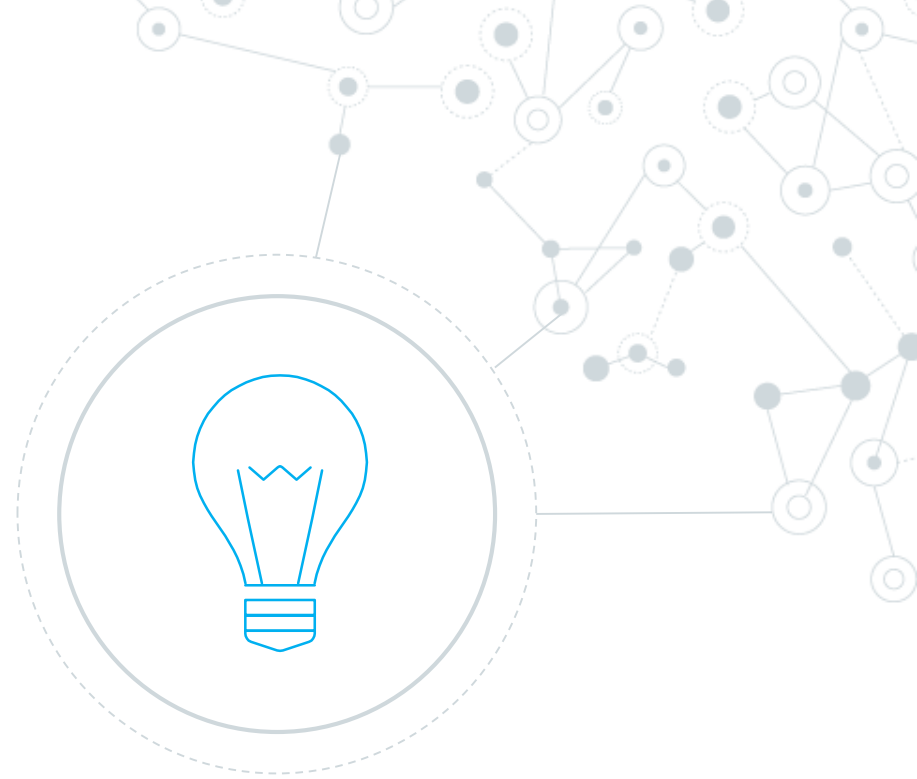
Situation

- ◎ Luther Films plans to release a film in July 2016 but they found out that there will be a number of similar (same-genre) movies that will be released in the same month ("density")
- ◎ Luther funded the production with high-yield debt and is worried that the competition might affect its ability to service the debt. On the other hand, interest will accrue longer if the release is delayed
- ◎ The managers at Luther Films reached out to Metis to help them on this issue and to come back with a recommendation the next day

The idea

Competition decreases
revenue, or does it?
The average American
watches a movie every
2 months

L.E.K Box Office Research



Approach

- ◎ Given the time and budget constraints...
- ◎ Data collection through web scraping
- ◎ Modeling through linear regression

Data Munging

- ◎ Data: Movies from 2009 to 2015
- ◎ Data source: BoxOfficeMojo
- ◎ Webpages scraped: 4,400
- ◎ Missing data treatment: Complete cases only
- ◎ Filter out: blockbusters, box office and budget over US\$1 million, foreign movies
- ◎ Centered “# of theaters” and “# of days in release” variables to solve severe multicollinearity
- ◎ Collapsed genres and distributors
- ◎ Final # of observations: 730

Beware of the Outliers

- © Distributed by Kenn (Viselman)
- © Domestic gross of US\$1 million
- © Budget of US\$20 million
- © While not statistically significant, OLS shows that getting Kenn as the distributor increases domestic box office gross by US\$48 million relative to “Other Distributors”

Source: BoxOfficeMojo



Regression Model

- ⦿ Evaluated OLS and Elastic Net Regression
- ⦿ Used 100% of data since focus in on coefficient
- ⦿ Model equation:

adj. dom. gross = 1 + adj. budget + runtime + # days in release + max # of theaters (C) + (max # of theaters (C))² + density + genre + distributor + month

	R ²	Adj. R ²	Density Coefficient p-value
Ordinary Least Squares	73%	73%	6%
Elastic Net Regression	72%	72%	21%

A background pattern of a network graph with nodes and edges, rendered in light gray. The nodes are represented by small circles, some solid and some hollow, connected by thin lines. The overall pattern is dense and covers the entire slide.

US\$ 2,200,000

less revenue for every same-genre movie that
will be released within 30 days of your movie
release date ($p < 0.06$)



“

*Luther Films should consider releasing the movie in June, a month earlier, for a potential increase in revenue of **US\$15 million*** ($p < 0.08$)

Other Ideas

- ◎ Density measured by # of high budget movies and (just the) # of movies released within +/- 30 days
- ◎ Scraped OMDB data for movie ratings from Rotten Tomatoes, IMDB, Metacritic
- ◎ Scraped RGB pixel data of movie posters



Thanks!

Any questions?

