Big Budgets? Big Returns? - An Analysis of Film Industry

Team 2
Ayoade Israel, Divya Menghani, Hamsini Sankaran, Sivakumar Thiyagarajan

Team's Hypothetical Organization: BigScreen Analytics

Intended Audience: Movie Investors, Production Companies, Directors

19th April, 2023

Research Question

How does the budget of a movie influence its revenue?

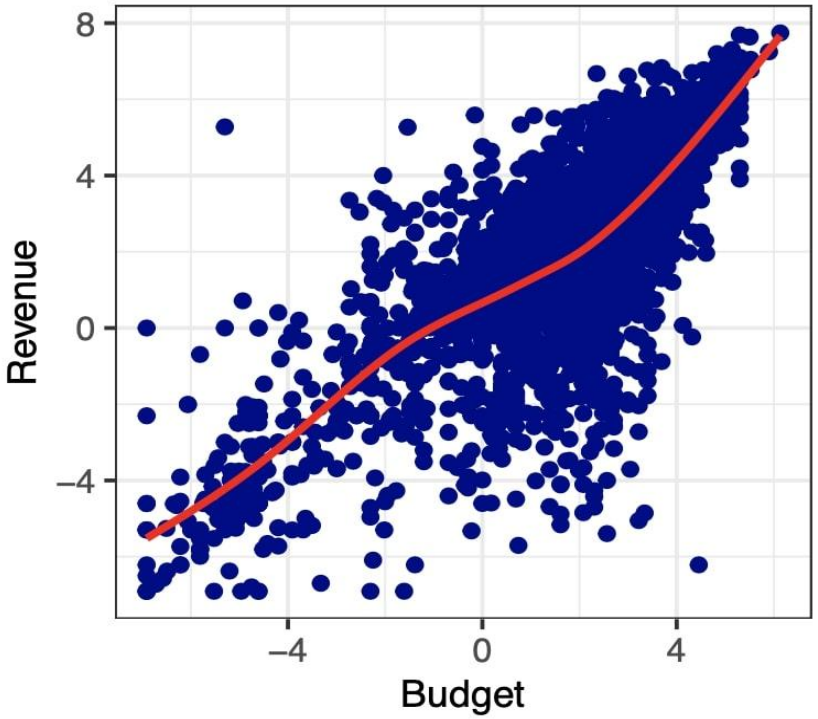
Intro (Question)

- ★ What we want to learn/investigate?
  - Today , we will be learning how the budget and some of the other factors of a movie influences its revenue.
- ★ Why we are investigating this topic?
- ★ Is budget the only factor influencing a movie??
- ★ When does a movie turns into a “HIT” and under what conditions it’s a “MISS”

Data

Data Source: https://www.kaggle.com/datasets/akshaypawar7/millions-of-movies

- **X Concept – Budget**
- **Y Concept – Revenue**



Budget (X Concept)

Covariates

a. *Vote Count*
b. *Popularity*
c. *Movie Run Time*
d. *Movie Title Length*
e. *Movie Release Month*
f. *Movie Language Category*

Influences

Revenue (Y concept)

Models and Results

$$\widehat{revenue} = \beta_0 + \beta_1 \cdot budget + \mathbf{Z}\gamma$$

β_0 = *Constant (y-intercept)*

$\beta 1$ = *coefficient representing increase in revenue for each unit increase in budget*

Z = *row vector of additional covariates*

γ = *column vector of coefficient of additional covariates*

Use case:

Budget = $1,000,000
Run time = 180 minutes
Movie Title Length = 10

→ Revenue = 77.2% ⬆

Movie + English Language → 15.6% ⬆
Movie + Holiday season → 18.1% ⬆

Table 1: Estimated Regressions

	Output Variable: Revenue of the movie		
	(1)	(2)	(3)
Budget	0.909***	0.780***	0.772***
	(0.008)	(0.009)	(0.010)
Vote Count		0.0002***	0.0002***
		(0.00001)	(0.00001)
Run time		0.007***	0.008***
		(0.001)	(0.001)
Popularity			0.001**
			(0.0003)
Movie Title Length			0.107**
			(0.036)
Release season			0.181***
			(0.049)
Release Language			0.156**
			(0.051)
Constant	0.514***	−0.325***	−0.800***
	(0.026)	(0.072)	(0.128)
Observations	6,880	6,880	6,880
R^2	0.625	0.671	0.673
Adjusted R^2	0.625	0.671	0.673
Residual Std. Error	1.734 (df = 6878)	1.624 (df = 6876)	1.620 (df = 6872)
Note:	*p<0.1; **p<0.05; ***p<0.01. Release season is Dec/Jan and rest of the months are considered as non-holiday season. Release Language Category is either English or Non-English movies.		

Limitations and Conclusions

- Limitations
  - Access to a larger dataset
  - Omitted variable bias impact
  - Bias e.g. Sampling bias
  - Questionable I.I.D.
  - Outliers

- Conclusion
  - Future research examinations
    - Reduce limitation impact

Courtesy of pngwing.com