



Revenue predictive regression model

By Yordanos W.



Purpose

- Movie directors and investors
- Factors to consider
- Predict profitability

The Data

Data scraping → BeautifulSoup

Websites → Boxoffice mojo, IMDB

Movies → 2000 - 2019 releases

Data points → 2002 movies

The top image is a screenshot of the Box Office Mojo website by IMDbPro. It features a search bar at the top right with the text "Search for Titles". Below the search bar is a navigation menu with links: "Domestic", "International", "Worldwide", "Calendar", "All Time", and "Showdown". The "Domestic" link is circled in red. Below the navigation menu are two tables showing box office data for the week ending Sunday, December 15.

Latest Dailies		Latest Weekend: Dec 13-15	
Jumanji: The Next Level	\$17,175,000	Jumanji: The Next Level	\$60.1M
Frozen II	\$5,946,000	Frozen II	\$19.2M
Knives Out	\$2,544,000	Knives Out	\$9.3M

The bottom image is a collage of various movie posters. Overlaid on the center of the collage is a large yellow rectangle with the "IMDb" logo in black text.

Variables

Features ➡ Budget , Distributor, genre, Running time , Rating

Target ➡ Revenue

Project methodology

- Data scraping
- EDA
- Baseline model
- Feature engineering
- Model development and fitting
- Model evaluation
- Model improvements

Model selection

Lasso mode

$R^2 = 0.47$

$MAE = 0.66$

What does it tell us?

Budget, running time - significant positive influence

Drama, R-rating, Other distributors - negative influence

Remarks

- Time sensitive
- Exceptions



Thank You!