Predict Domestic Gross of Movies

Katie Fan

Introduction

- Motivation:

  Predicted gross can help plan movie's production and distribution

- Goal:

  use data insights to predict movie's domestic gross and find the important features

Methodology

- Data:

  Scraped data from https://www.boxofficemojo.com/

  TMDB data from Kaggle

  IMDB data from data.world

- Approach:

  Ridge regression

Methodology

- Features:

  Runtime, Budget,

  Month, MPAA rating,

  Director's average gross,

  Producer's average gross,

  Director's total academy awards,

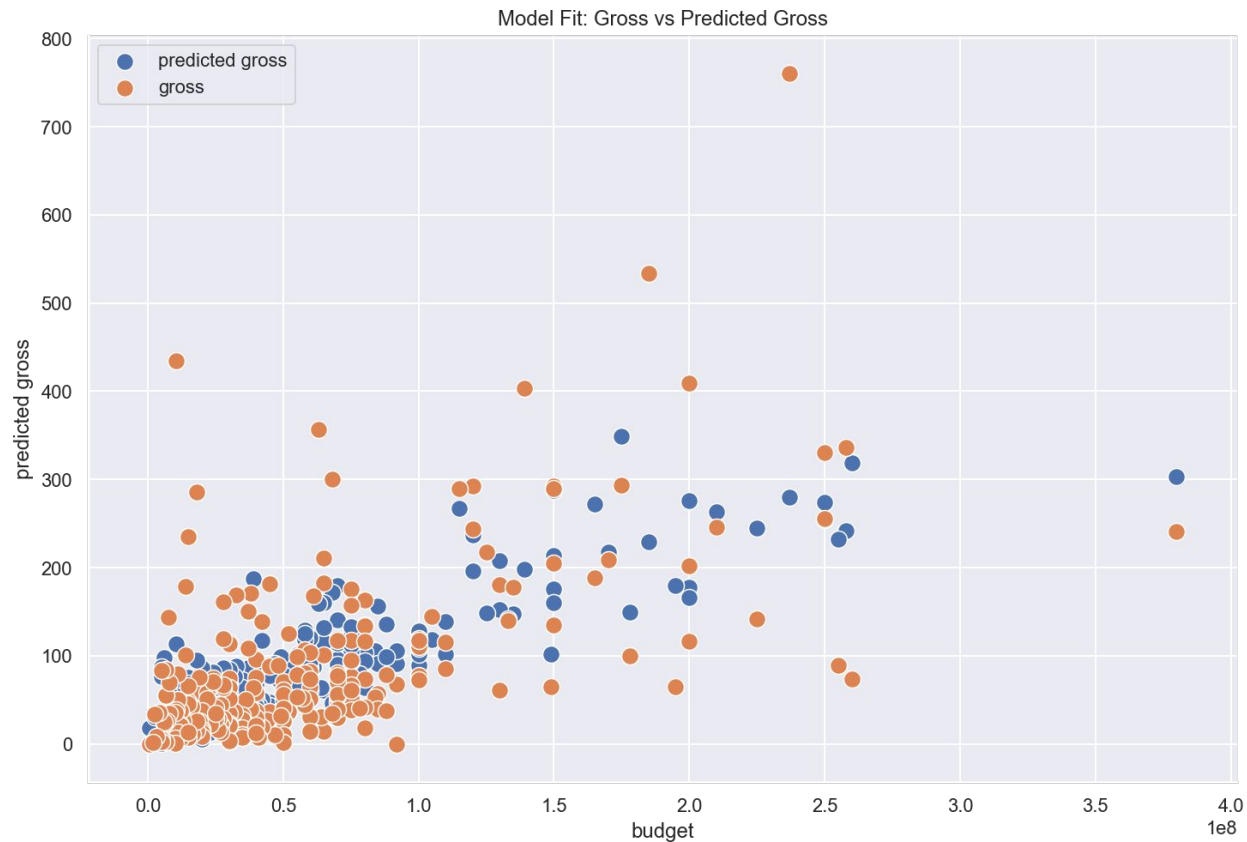 Lead actors' total academy awards

Results: coefficients

Top 5 features and regression coefficients:

- avg_director_gross: 28.54
- budget: 25.30
- avg_producer_gross: 15.33
- runtime: 7.67
- month_Jun: 5.40

Results: R Square

R^2: 0.4937

Results: Model Fit Plot

Model Fit: Gross vs Predicted Gross

orange: true gross
blue: predicted gross

Conclusions

- Avg_director_gross, budget, avg_producer_gross and runtime are the most important features to predict movie's domestic gross

- June is the most important release month for movie's gross

Considerations

- Analyze more features such as popularity of the leading characters, movie writer's gross and awards data, etc.

- Consider inflation

Appendix

Sources:
- scraped data from www.boxofficemojo.com
- Dataset from
  https://www.kaggle.com/tmdb/tmdb-movie-metadata
- Dataset from
  https://data.world/popculture/imdb-5000-movie-dataset

Web Scrape Tool:
- Requests
- BeautifulSoup

Thank you!