What Affects Movie Revenue?

Catherine Wu[1], Haocheng Bi[1], Faith Tsang[1], Jeremy Tran[1], Waveley Qiu[1]

[1] Department of Statistics, UCLA

1 Introduction

- A movie's success is widely presumed to be heavily influenced by its genre, which itself appears linked to factors like production country or release year.
- These assumptions lack extensive study, however, and so quantitative analysis on the topic would prove useful to investors, production companies, and theaters.

2 Objectives

- Find the drivers of movie revenue.
- Determine which genres of movies perform best (accounting for movies with multiple genres), and if these differences are statistically significant.
- Examine if the genre-revenue relationship exhibits any interaction with seasonality, and how genre compares to other variables in predictive power.

3 Methods

- Excessively low values for budget or revenue were corrected manually referencing from a Kaggle notebook.
- Standard ANOVA was used for the first objective. Genre counts were treated either as movies that were "mainly" that genre, or as all movies that might fall in that genre.
- Feature importance was calculated via both `randomForest` and `xgboost` algorithms.

4 Results

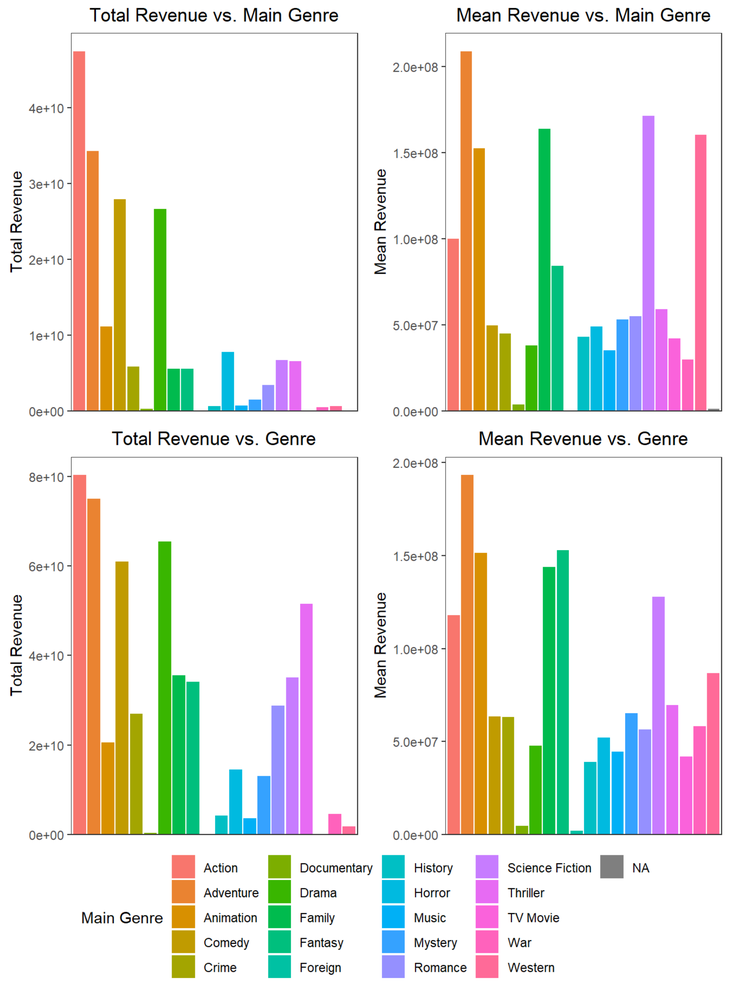
ANOVA: Main Genre Only

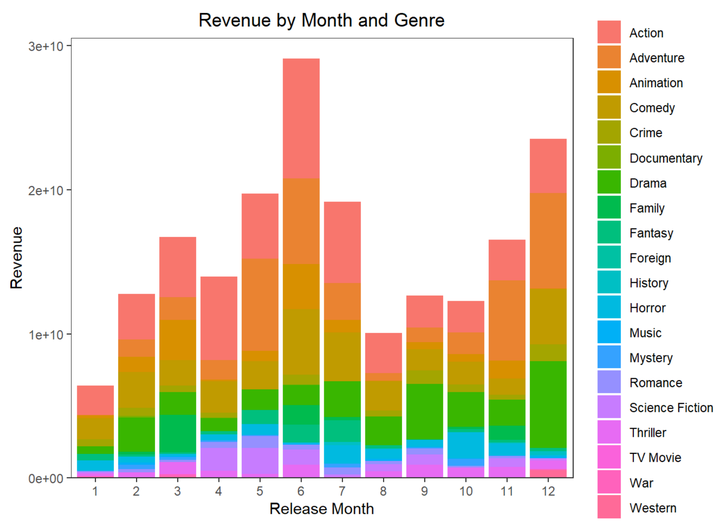
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
main_genre	19	6.352e+18	3.343e+17	18.44	1.028e-58
Residuals	2711	4.913e+19	1.812e+16	NA	NA

ANOVA: With Multiple Genres

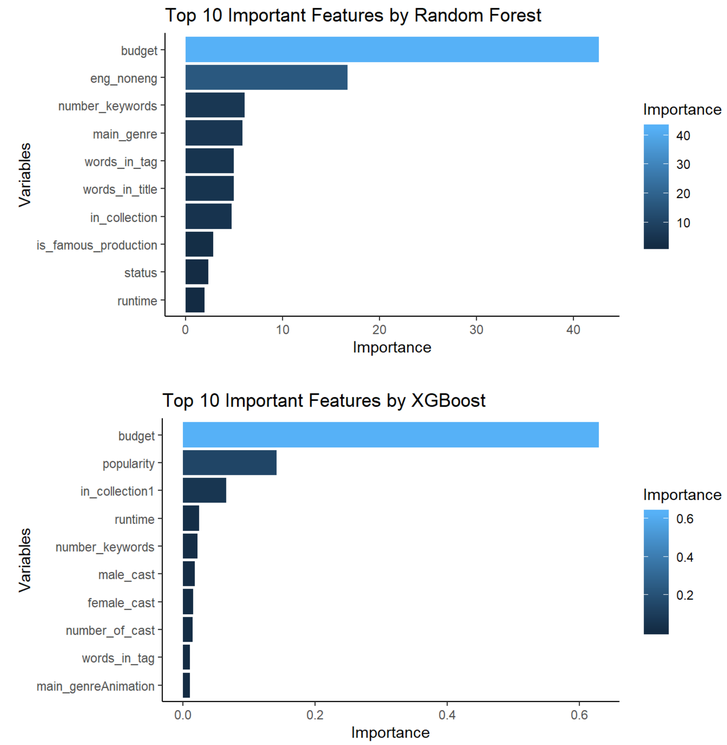
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
genre	19	1.288e+19	6.779e+17	30.51	5.563e-106
Residuals	6820	1.515e+20	2.222e+16	NA	NA

`Pr(>F)` represents the p-value for the ANOVA test.





Above, cumulative monthly box-office revenue from 1980 to 2019 is further segmented by genre. This allows us to see which genres have been most successful overall in any given month.



Feature importance is measured by mean decrease in node impurity.

5 Summary

- Action, adventure, comedy, and drama movies have dominated historically, both in aggregate and on average.
- When allowing movies to have multiple genres, science fiction films' revenue increases significantly, likely because science fiction is often a "secondary" genre to action or adventure.
- June and December are the most profitable months overall, likely due to summer and holiday blockbusters.
- Notably, budget and number of keywords appear to be the only features that are considered most important by both `xgboost` and `randomForest`, while genre does not appear to be very significant by either.

6 Acknowledgements

- The dataset used here was acquired from TMDB via a Kaggle competition. The data itself was acquired using TMDB's
- The inspiration for cleaning many of the text-based variables and for using random forest came from this Kaggle notebook.