Understanding the Movie Industry

What makes a High Grossing Movie?

Claire Sarraillé

Claire Sarraillé
Self-Paced
Monday June 14th 1pm PST
Instructor: Jeff Herman

Project Repo URL: https://github.com/clairesarraille/mod1finproj/tree/main/dsc-mod-1-project-v2-1-online-ds-sp-000-master
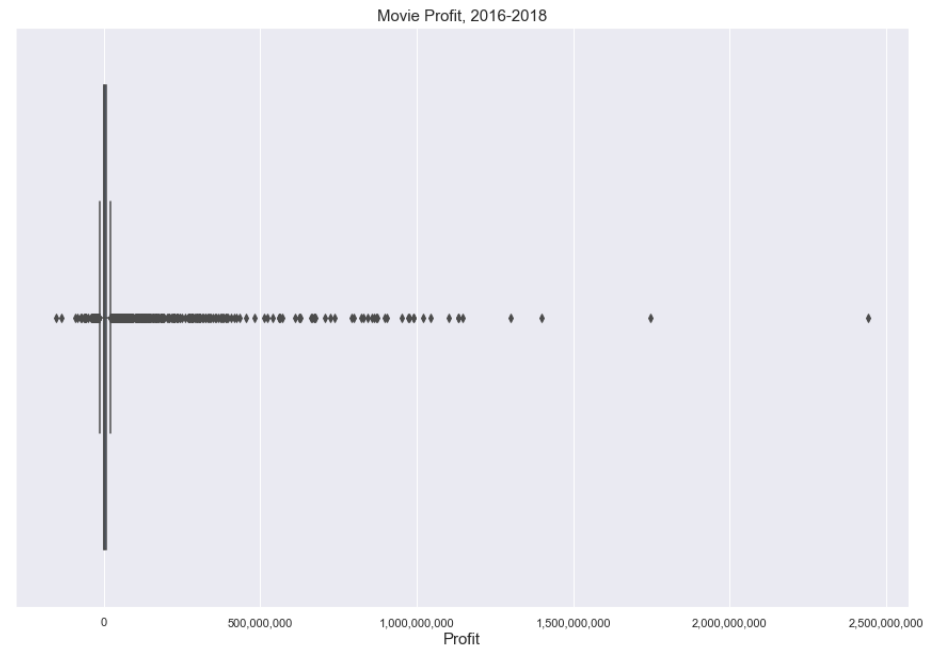
Blog post URL: https://clairesarraille.github.io/2021/06/08/your-new-blog-post.html

Hello my name is Claire Sarraille and I'll be outlining my approach to the data intake, cleaning, munging and visualization of movie data gathered from TMDb (The Movie Database) for my first project as a student with Flatiron.

The Problem:

Most movies don't make money. We at Microsoft want to skip the trial-and-error process as much as possible.

What's the nexus of type, production and indicator that spells a winner at the box office?

Movie Profit, 2016-2018

Microsoft wants to start a studio to create original streaming content. So far, however, they haven't conducted the necessary research to understand the attributes of a successful movie. To rectify this gap in knowledge, myself and my team have been asked to analyze recent movies and present insights into predictors of high profit productions.
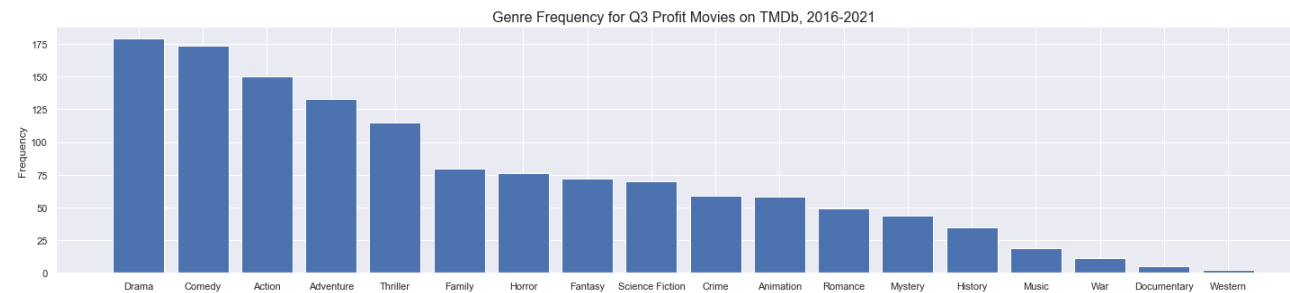
Business Value
Setting the Stage

- Where to start?

  - **Genre** - informs screenplay selection

  - **Production Company** - informs types of social capital to tap into

  - **Crowd Sourced Data** - taps into fickle consumer tastes

The value delivered by this analysis is in presenting the foundation of informed decision making as our movie studio is built from scratch. The genres of films we are known for, as well as the production companies and networks we involve ourselves with are thus valuable to understand so that we are set up for success from the get-go. In turn, understanding the usefulness of crowd-sourced ratings is valuable because this data could help stay on top of the popularity of different kinds of movies in an environment where judgement fall swiftly via social media.
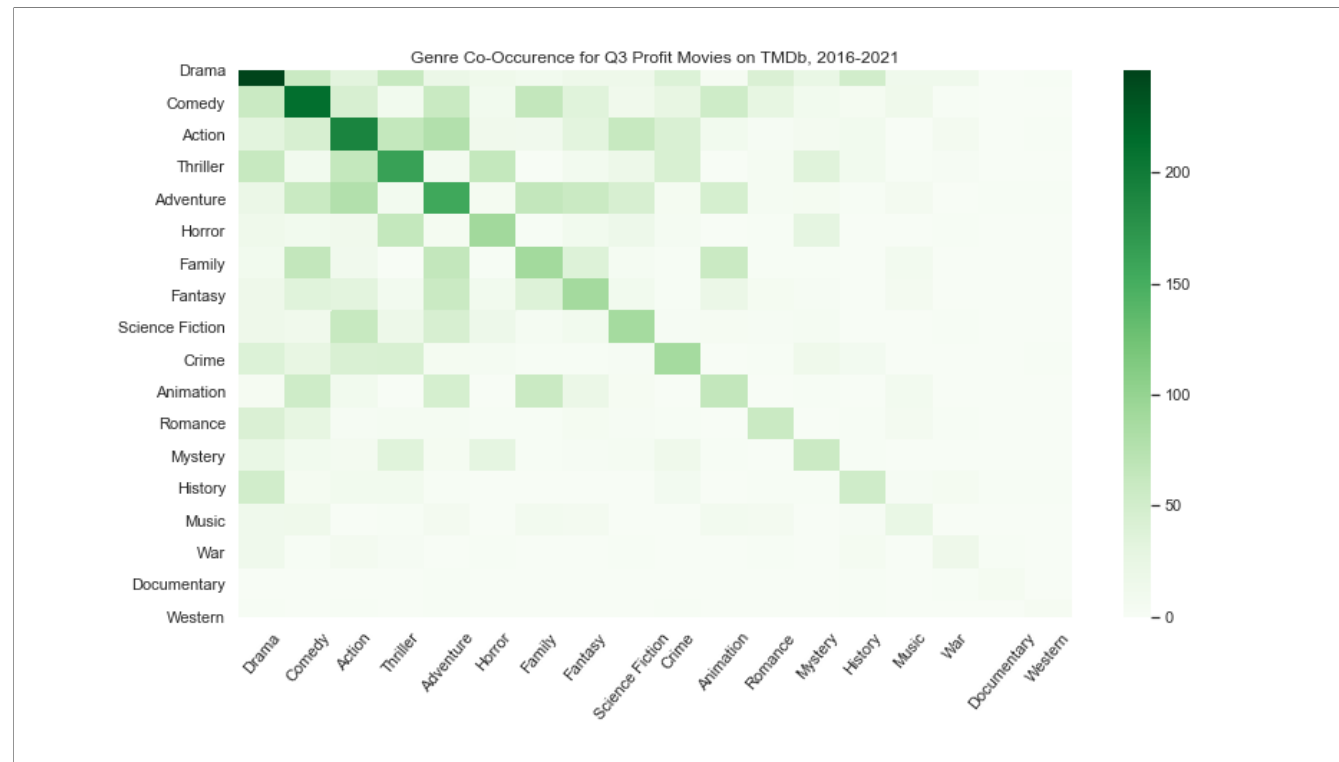
Methodology:

- **Timeframe**: 2016-Present

- **Data**: TMDb

- Variables:
  - Genre
  - Production Company
  - Popularity Score

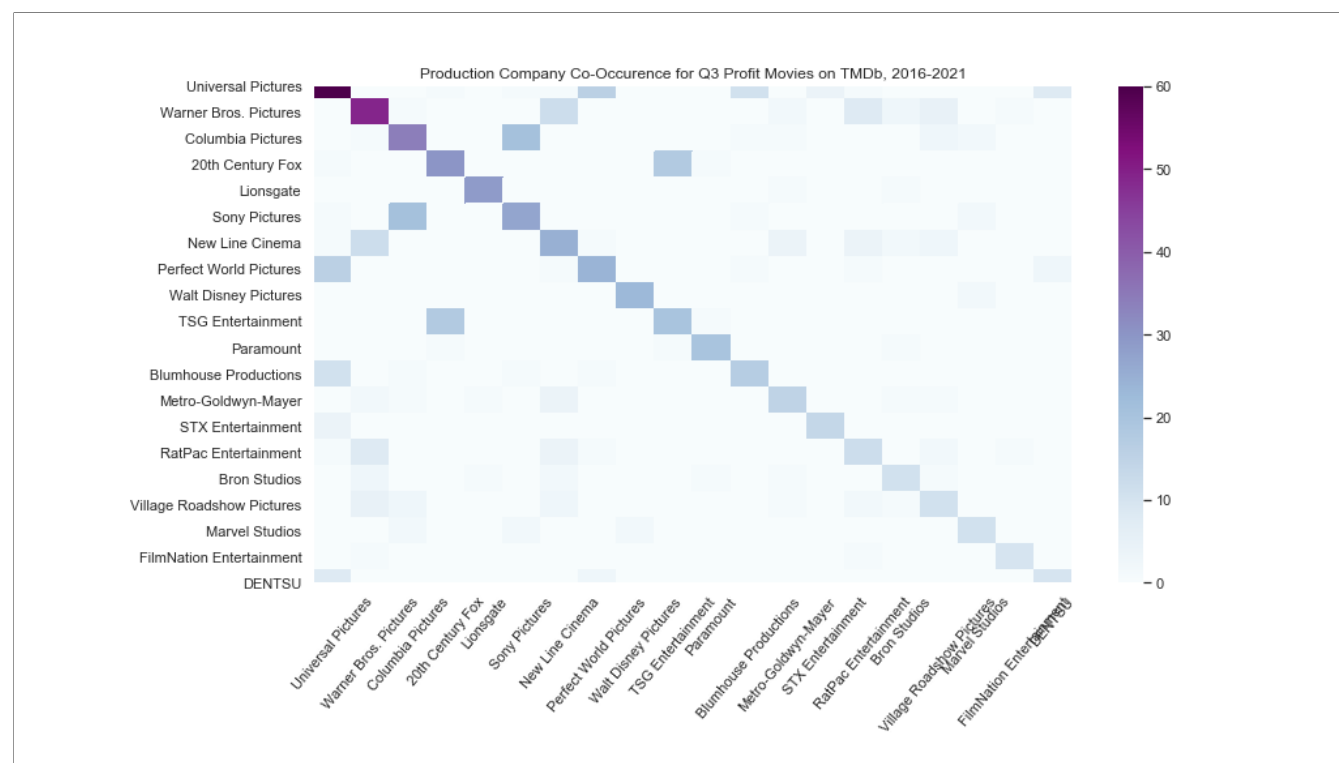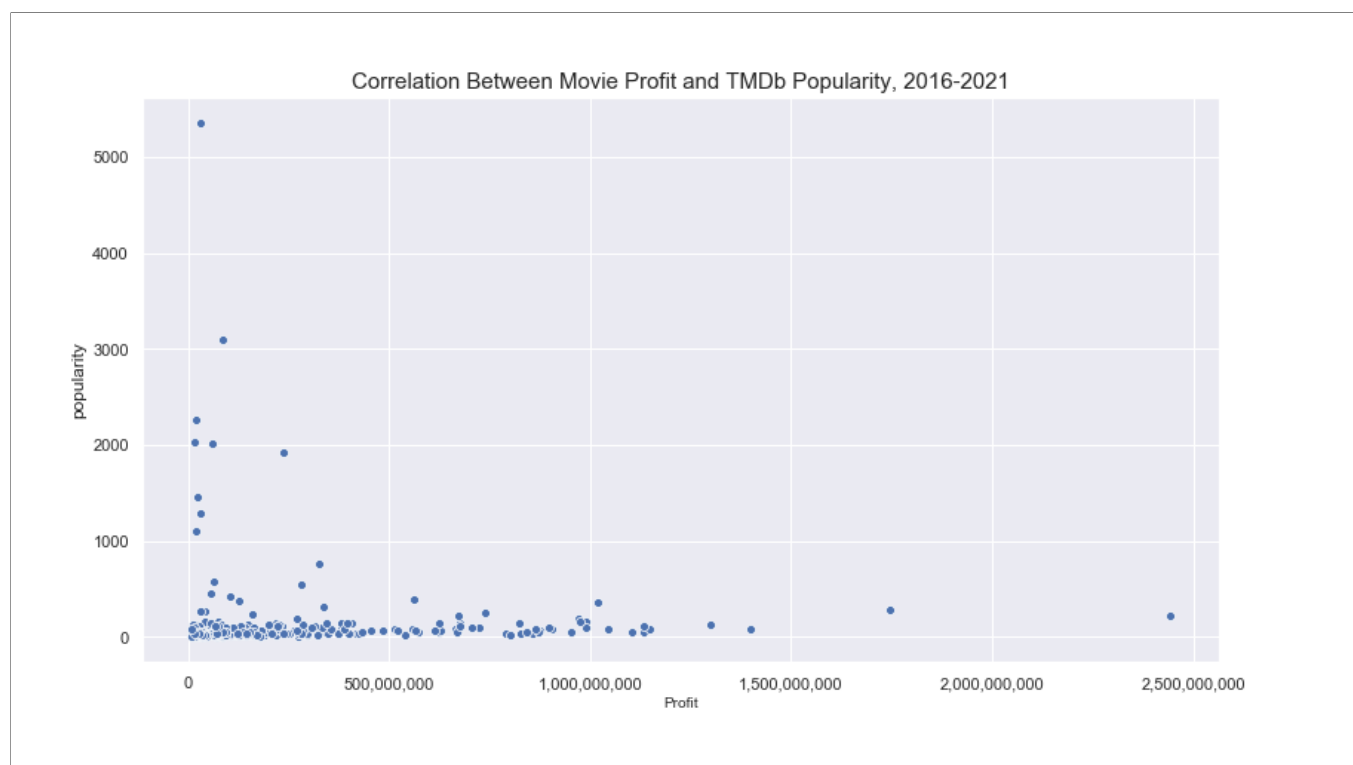Genre Frequency for Q3 Profit Movies on TMDb, 2016-2021

I decided to use TMDb because of its robust API which will allow for repeatability and scaling of analysis. We want to know what is currently profitable and take a few years of the past into account to suss out what outlives trendiness.

We looked beyond the most frequent genre and production company for the highest grossing films by creating heat maps that depict the combos of genre or the combos of production company that occur the most often with high profit films. We isolated high profit films by using the top quartile. We also used TMDbs popularity score, which is a daily updated user popularity index that is influenced by votes, likes, and adds to watchlists. We wanted to know if TMDb popularity is correlated with high profit films as a first step to understanding the relationship between crowd sourced data and box office revenue.

Genre Co-Occurence for Q3 Profit Movies on TMDb, 2016-2021

We can observe from the genre and production company heat maps that some of the highest-profit movies in the past 5 years are Dramas, Comedies and Action films produced by Universal, Warner Bros, and Columbia. Even more interesting are the winning combinations of these attributes. 20th Century Fox paired with Columbia, Sony paired with Warner Bros., and TSG with Lionsgate appear to be synergystic forces. Our recommendation is to conduct further research into how these production companies work together, why they work together, and how we can tap into the networks of these winning teams. Likewise, we can observe patterns of genre that appear to garner the greatest rewards at the box office, such as action-adventure, historical dramas, and family comedies. Our team recommends seeking out screenplays that reflect these genre combinations. We can also see from visualizing and running the Pearson correlation on popularity versus movie profit that there is a moderate correlation. We recommend running more analysis on crowd sourced popularity indices to look for causal factors beyond mere correlation.

Production Company Co-Occurence for Q3 Profit Movies on TMDb, 2016-2021

Correlation Between Movie Profit and TMDb Popularity, 2016-2021

Moderate correlation - 0.06 Pearson coefficient.

Summary

- Heatmap analysis lends a more nuanced picture of categorical attributes of high-profit movies

- Crowd sourced data merits further investigation because it's a nimble indicator of consumer taste.

Future Work

- Comparing the correlation of IMDB user ratings

- Are professional critic reviews a better indictor than user ratings?

- TMDb data has an imdb_id field - opportunity for comparison.

Thank you for your time and attention. Please let us know if you have any feedback.