



Mod 1 Project Proposal to Microsoft

Mindy Zhou, Keita Miyaki
June 19, 2019



Goals

- **What is profitable movie portfolio?**
- What determines successful movies?
 - Genre?
 - Seasonalities?
 - Actors/actresses?
 - Duration?
 - Budget size?

Learning from past samples:

Filtering Criteria

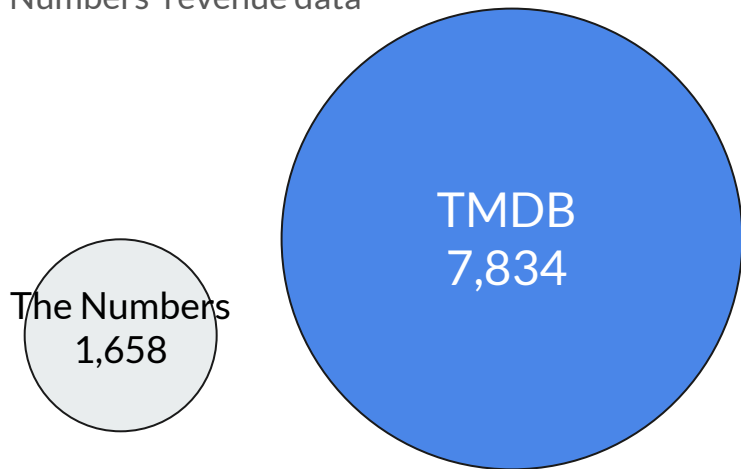
- Movies released in 2009-2018
- Movies in English

Data

- Worldwide gross sales/budget
 - The Numbers
 - Box Office (for data validation)
- Popularity Index
 - TheMovieDB.org
 - It is based on user interaction
- Genre, actors/actresses, length
 - IMDB

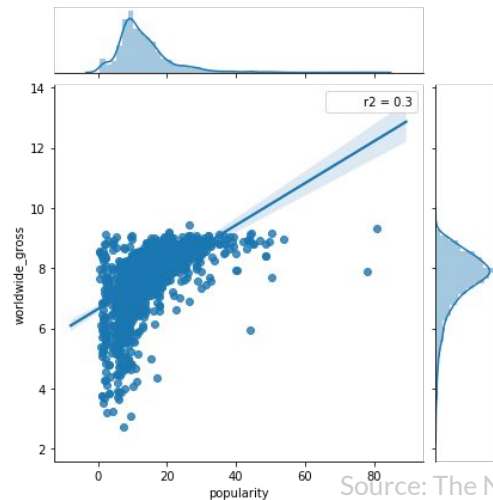
Popularity index proxies gross revenues

TMDB's popularity index covers more than The Numbers' revenue data



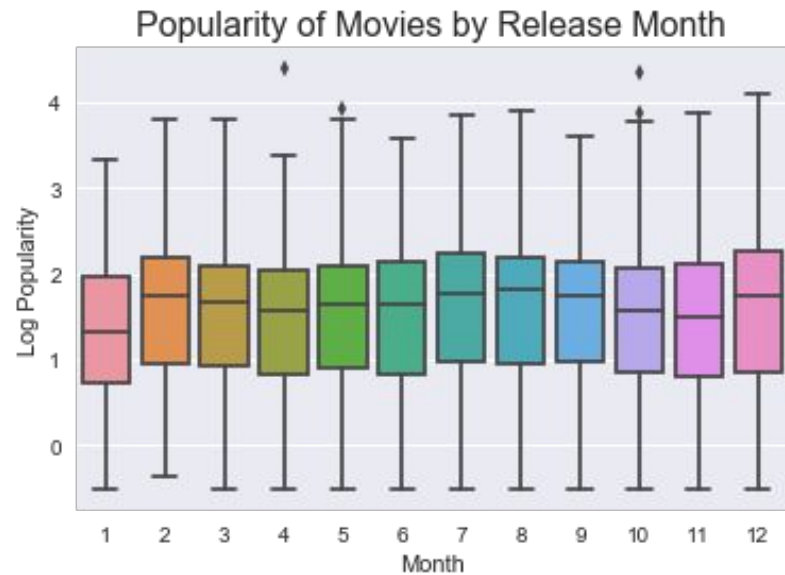
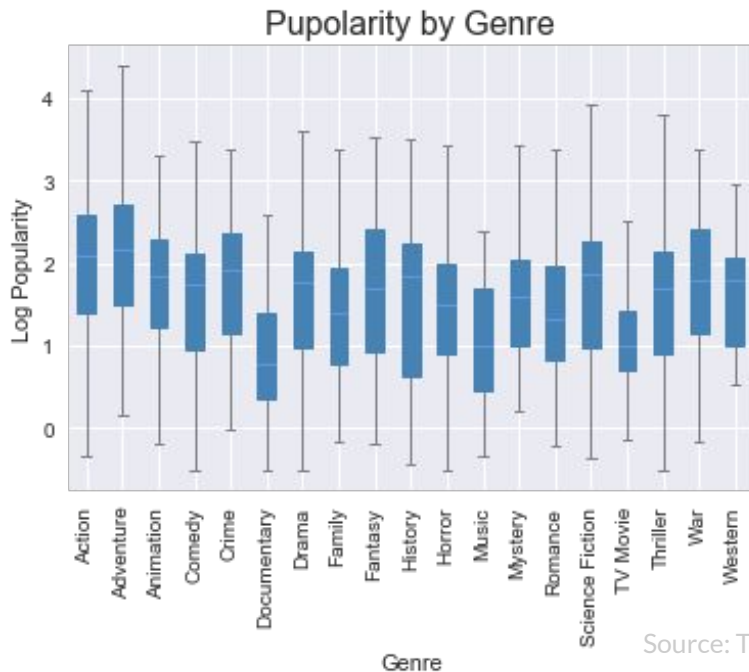
Source: The Numbers, TMDB

Popularity index is correlated with revenues, so we use popularity data as “successful movie” determinant

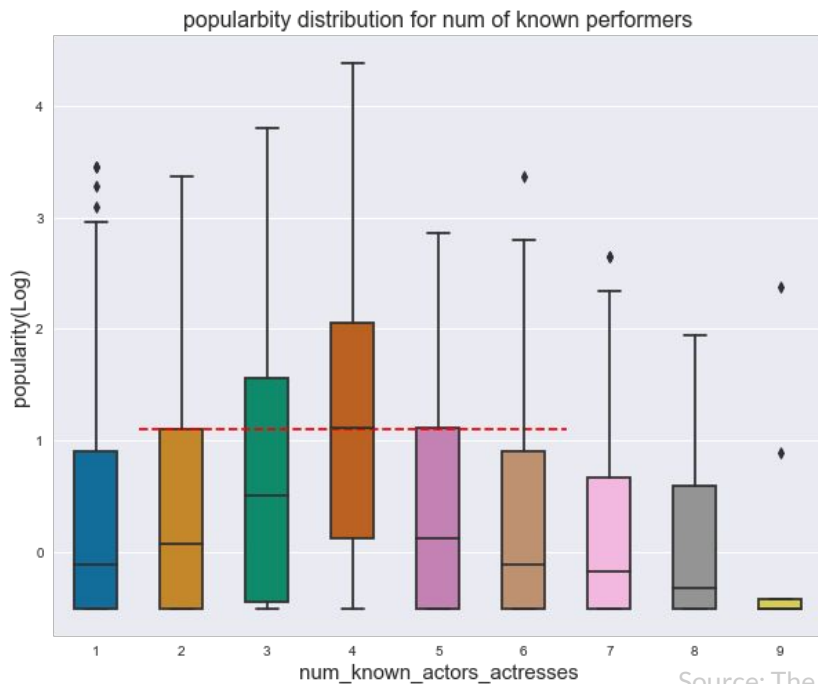


Source: The Numbers, TMDB

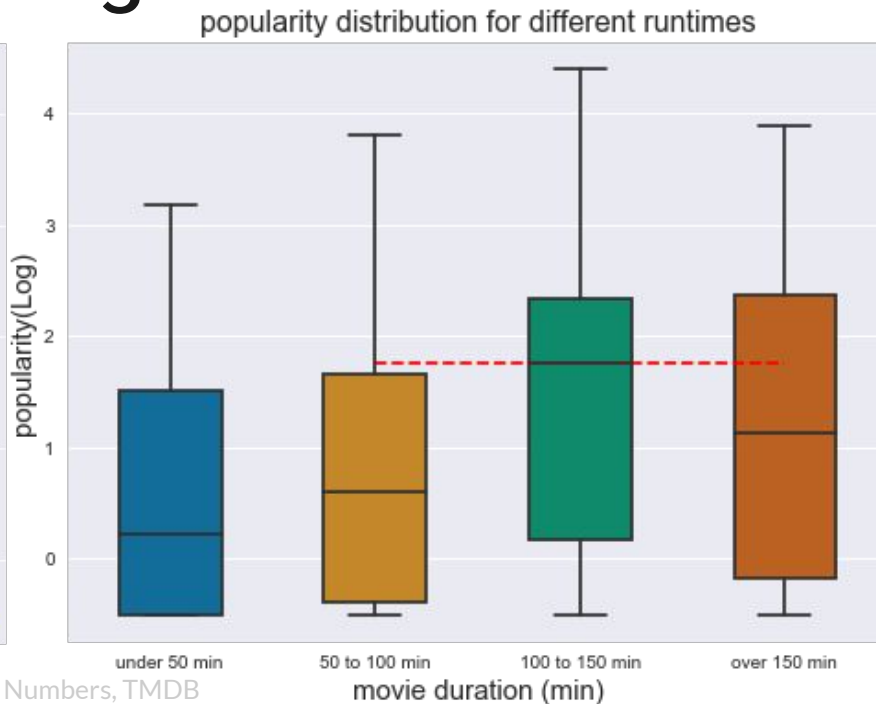
Genres and seasons don't matter



of actors/actresses, length

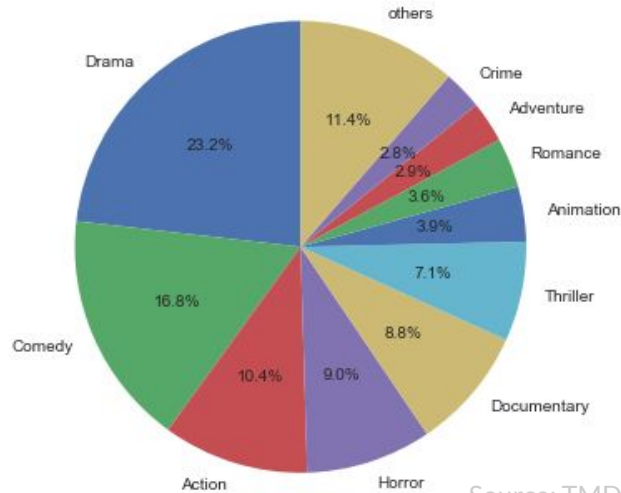


Source: The Numbers, TMDB



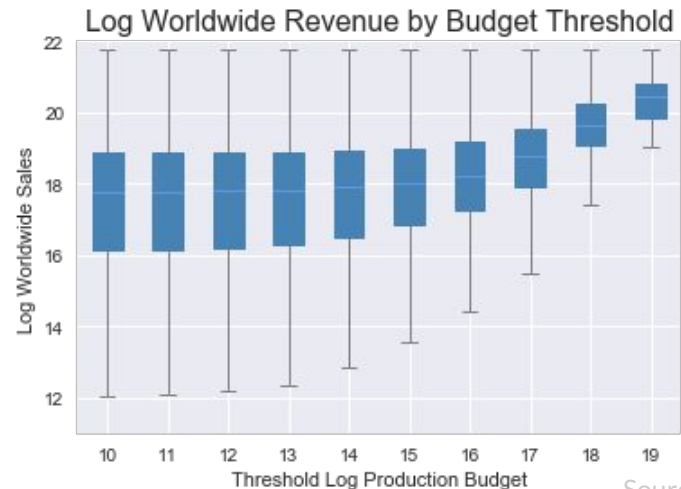
Movie portfolio by genre and Production Budget

Assume supplies match demands, following the genre-mix of current market (below) makes sense



Source: TMDB

Budget above e^{18} (67 million) is associated with well-anticipated high gross revenues



Source: The Numbers



Conclusions

- No significant difference among genres
- No significant difference by release timing
- 4 known Actors/actresses seem to work the best
- Length of 100 to 150 min seem to be most popular
- Production budget >\$67 million is associated with high gross revenue

Next Step:

**Hire more people from Flatiron
School who offer more
sophisticated data science
knowledge**



Sources

- The Numbers
- Box Office
- TheMovieDB.org
- IMDB