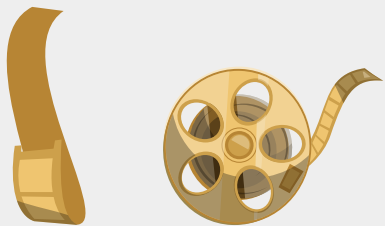
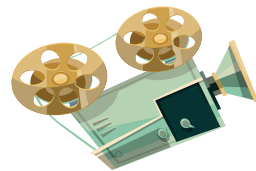


Non-English Film Engagement on IMDb

Team members: Marcus Almanza, Yoshie Bell-Souder, Emily Carpenter, and Janet Matthews-Derrico





Key Questions:

- Do non-English films have fewer votes, regardless of their average rating?
- Which films are “underrated gems,” those that have high ratings, but relatively low view counts?
- Do certain directors have more user-engagement than others, and does the rate of engagement vary based on the language?
- Do certain genres of non-English films have higher engagement? How does this compare to English film genres?



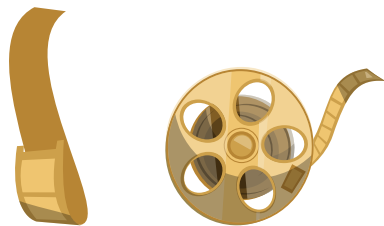
Data Preparation

- Data Integration
 - Identify useful columns among files
 - Merge columns into single file
- Data Reduction
 - Remove all non-movies from data (TV series, shorts, etc)
 - Remove adult films from data
- Data Cleaning
 - Fill in values for language

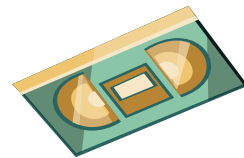


Tools:

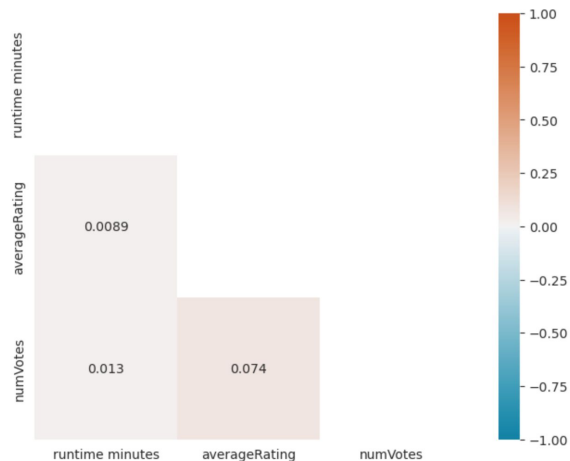
- Discord - daily communication
- Zoom - weekly meetings
- Github - version control
- Google Docs and Slides - project milestones
- Python - data cleaning and calculations
 - Numpy, Pandas, etc.
 - Cinemagoer
 - TMDb (The Movie Database)
- Tableau - data visualizations



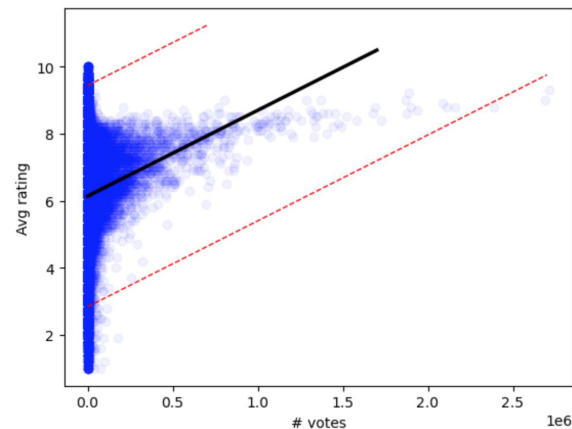
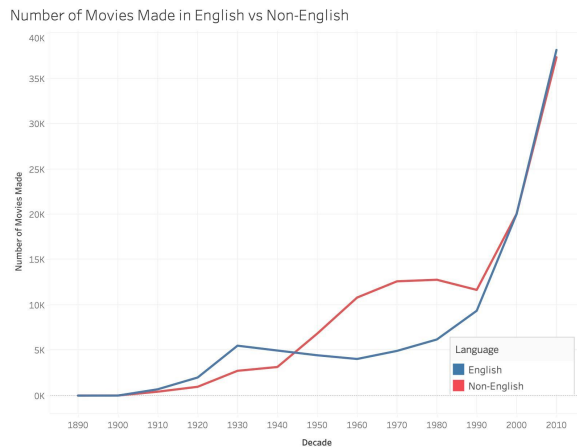
Classification and Clustering



1. Correlation Clustering



2. Linear Clustering

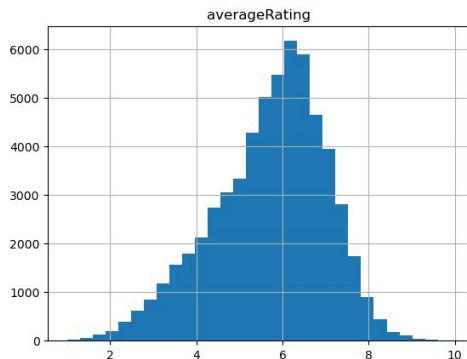




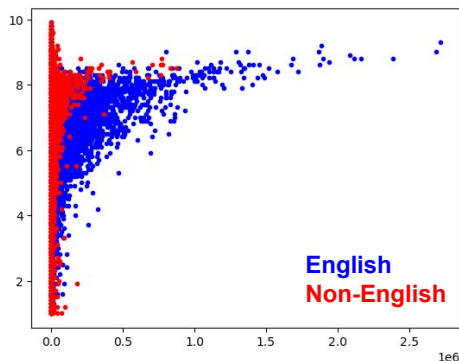
Knowledge Gained: Average Rating

- 59,731 English movies, and 56,508 Non-English movies with number of votes over 100
- Average Rating for Non-English films: 6.2
- Average Rating for English films: 6.08
- No significant difference between the two, even with English films having more votes

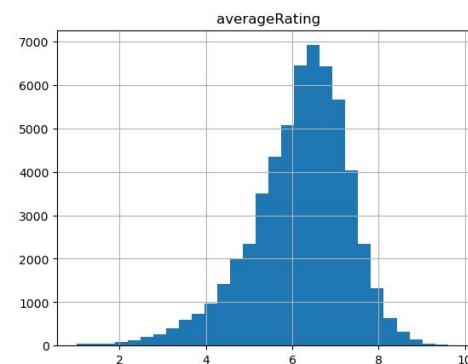
English



Vote Distribution



Non-English

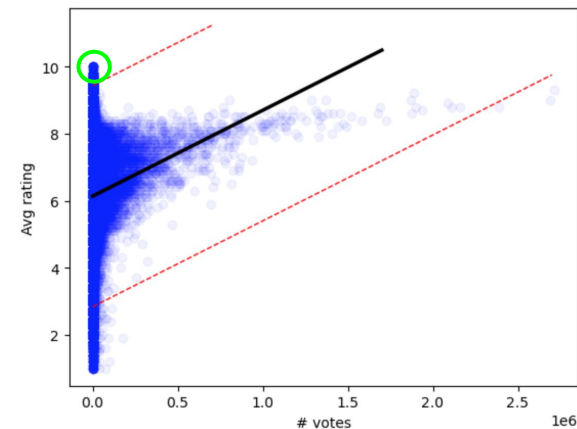




Knowledge Gained: Underrated Gems

Found about 400 films. For example,

- *The Silence of Swastika* (Bhardwaj, 2021), a Hindi movie with 10,353 votes and a 9.5 rating
- *Taxi* (Sajja, 2023), a Telugu movie with 1,504 votes and a 9.8 rating
- *Divorce* (I G, 2023) a Malayalam film with 601 votes and a 9.6 rating

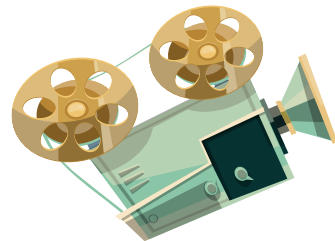
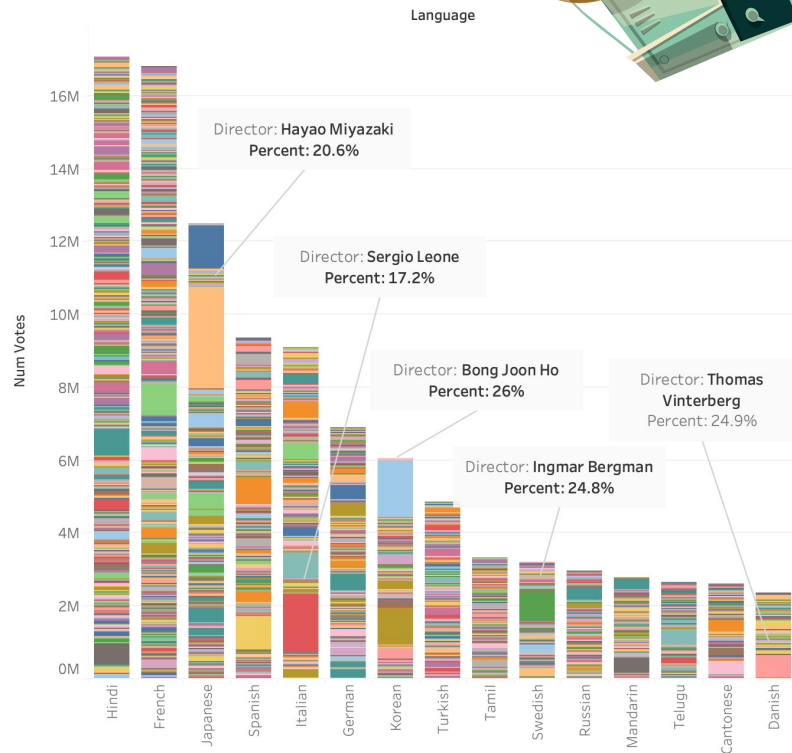


56 films of the top 100 were Indian-language films.

Knowledge Gained: Directors

- *Parasite* accounts for 13.8% of all Korean-language votes
- 25% of all Danish votes are for Thomas Vinterberg's films
- Steven Spielberg accounts for 1.65% of English-language votes
- Vinterberg's portion of Danish ratings is equivalent to the top 44 English-language directors combined

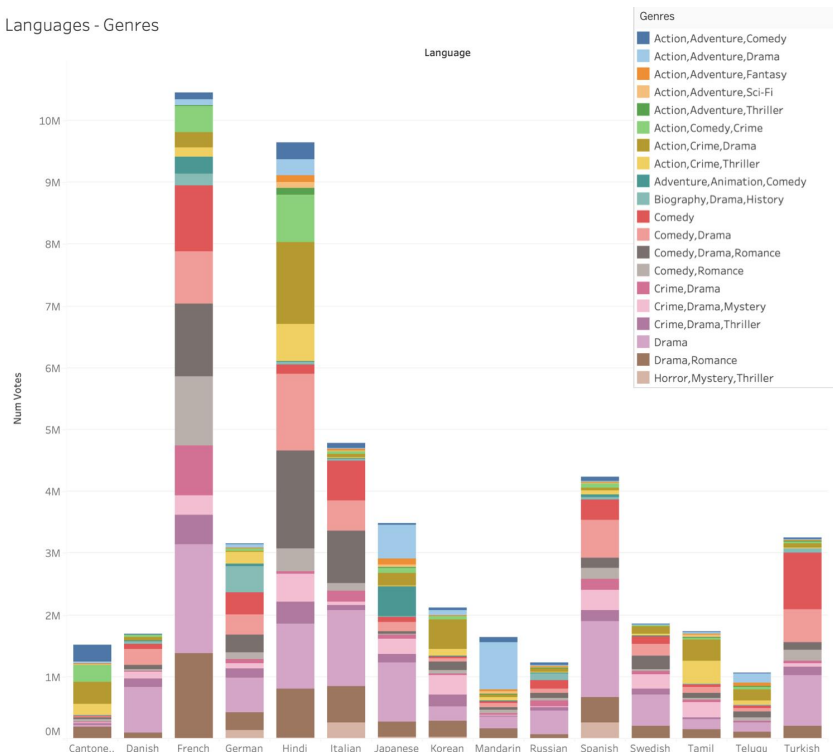
Languages Grouped by Director



Knowledge Gained: Genre



Languages - Genres



Non-English Film Engagement on IMDb

- The most generally popular genre in non-English language films is Drama
- Drama films comprised over 22% of non-English films and accounted for nearly 8.6% of all non-English film user votes
- The genre with the highest average user votes per film was Crime/Drama/Mystery

Knowledge Application:

- Potential investment in non-English film distribution in either the U.S. market or more globally
 - Some languages appear to have an established market and therefore are a relatively safer bet - French, Hindi, Spanish, Italian, and Japanese
 - Films by a director with past success
 - Drama in particular has had obvious and sustained success among non-English films on IMDb
- Findings may help non-English films in choosing conditions that will encourage user engagement
- Our research could be applied with regards to IMDb users looking to discover new films

