

WHAT MAKES A POPULAR MOVIE?

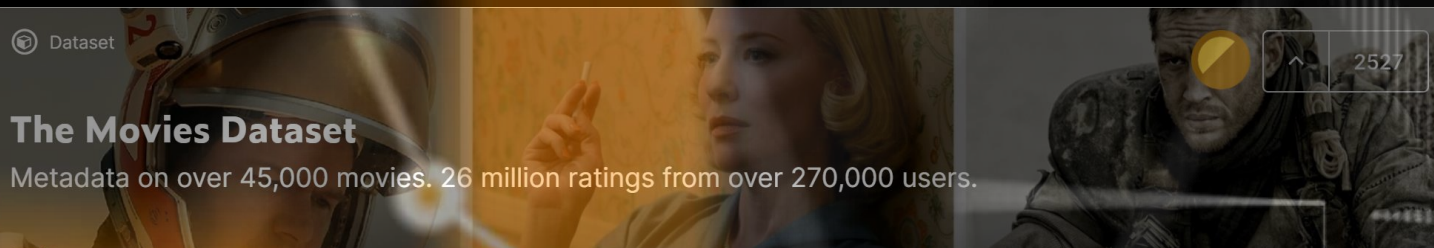


Brain Station Final Project

By: Kash Bains & Chin Chen Lo

OUR DATA

- “*The Movies Dataset*” from *Kaggle.com*



- All types of film
- Data from TMDB and IMDB
- Messy and Incomplete Data
 - Flexible working based on data formats given

The Movie Database (TMDB) ?

- *themoviedb.org* is a crowd-sourced and user-editable movie informational database.
- It is popular among movie lovers with its recommender systems: *Content Filtering* and *Collaborative Filtering*
- The popularity ratings on TMDB serves as an important metric to enhance search results and establish algorithms for discover functions.

THE
MOVIE
DB

DATA CLEANING PROCESS

- ▶ Programs Used
 - ▶ Excel, MySQL, Tableau, R



VARIABLES



Popularity

Popularity - An aggregated value from tmdb based on criteria of the database. The value mainly involves number of views , votes , favourites , watchlists, etc. It serves as an important metric to boost search results and establish an algorithm for discover functions.



Runtime

Runtime - The duration of which the film runs for in minutes.



Spoken Languages

Spoken Languages – The total number of languages spoken in the film.



Primary Genre

Primary Genre - The main genre of the film it falls under as sorted in order by the dataset.

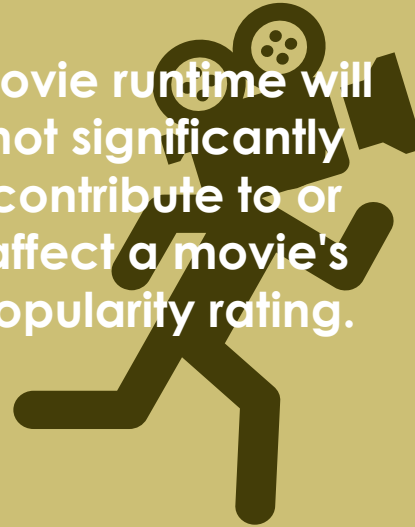
HYPOTHESIS



A film produced in more languages will equate to a higher popularity rating.



Movie runtime will not significantly contribute to or affect a movie's popularity rating.



There are genres of preference by the online community with evident trends in recent years.



> SELECT BEST_GENRE FROM ALL_MOVIE_GENRES;

Overall average popularity ratings
= 3.405



MAX(popularity_rating) = 547.88
MIN(popularity_rating) = 0.000001
MEDIAN(popularity_rating) = 1.3049




Calculations of proportions of
each genre's popularity ratings
below and above the average
popularity ratings




Gives a basis on better-performing
genres from the online TMDb
community

 Adventure

 2 Fantasy

  0 Foreign

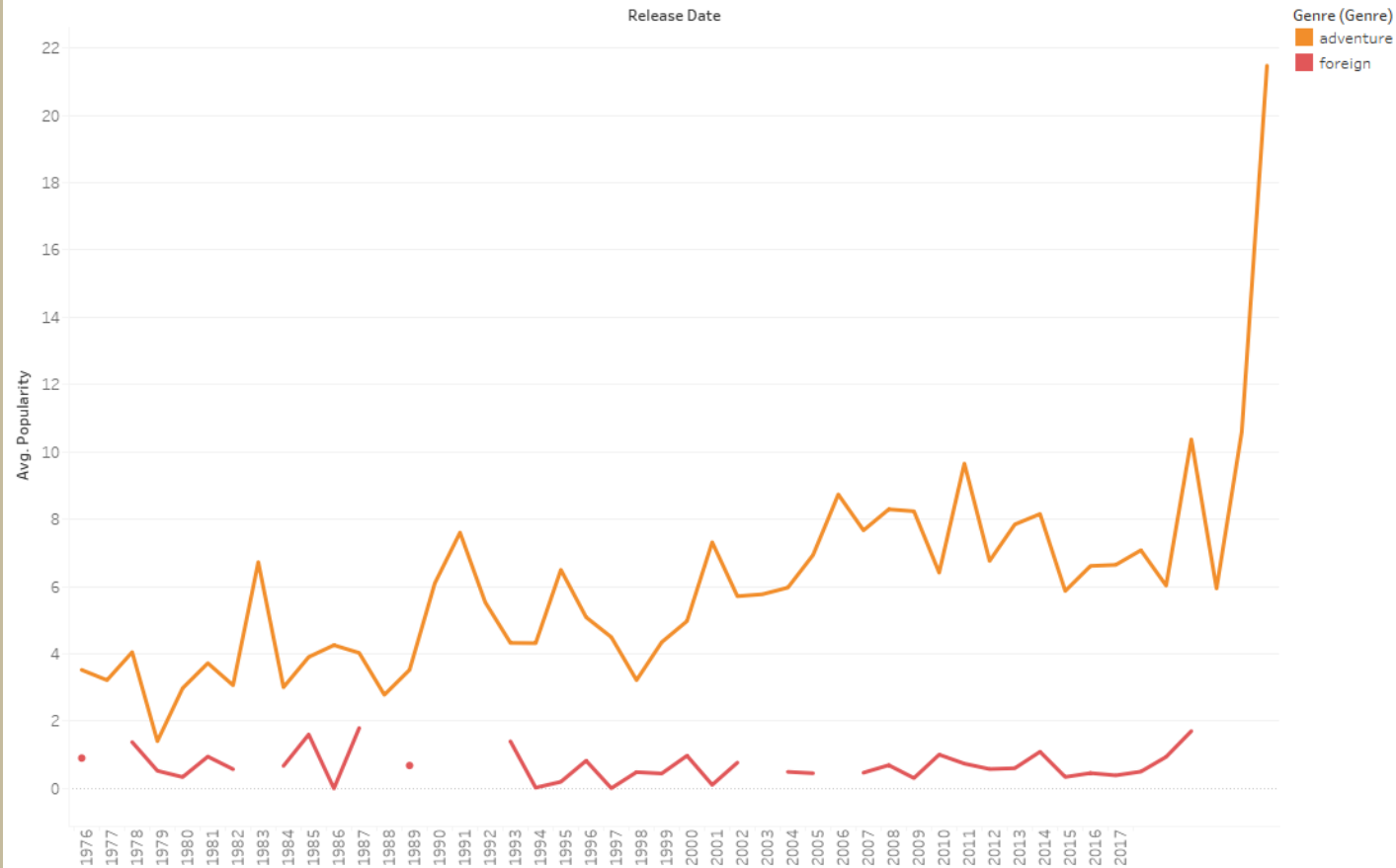
Proportions of Movie
Genre's Above the
Average Popularity
Rating

Genre Id		Avg. Above
action	0.4383	 0.0000 0.5353
adventure	0.5353	
animated	0.4017	
comedy	0.3330	
crime	0.3936	
documentary	0.0607	
drama	0.2763	
family	0.2637	
fantasy	0.4522	
foreign	0.0000	
history	0.2683	
horror	0.3497	
music	0.1761	
mystery	0.3162	
romance	0.2460	
scifi	0.3295	
thriller	0.3403	
tvmovie	0.1958	
war	0.2627	
western	0.2317	

Average of Above broken
down by Genre Id. Color
shows average of Above.
The marks are labeled by
average of Above.

GENRE POPULARITY OVER TIME

AVG Genre Popularity Trend



The trend of average of Popularity for Release Date Year. Color shows details about Genre (Genre). The view is filtered on Genre (Genre), which keeps adventure and foreign.

Comparison between the lowest and highest average popularity rating.

Adventure movies have an early sporadic average, which begins its upswing around the new millennium.

Foreign films have a more constant & low level of popularity over time.

REGRESSIONS & RELATIONSHIPS

Independence?

Chi-Square Test

	Spoken_Languages			
Popularity	1	2-5	5+	
Below 3.405	22240	2800	31	
Above 3.405	7909	2879	51	

No.

Regression

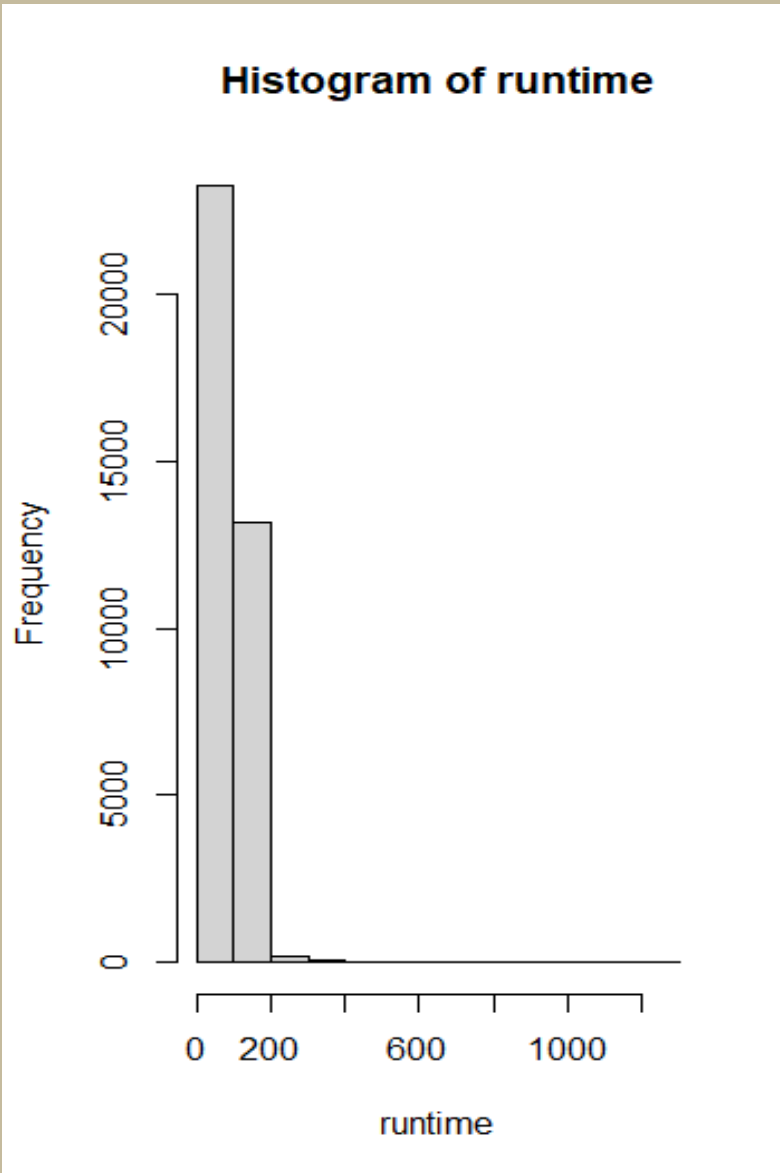
```
lm(popularity ~ spoken_languages * genre_dummy)
```

Non-Interaction Model:
Number of spoken languages has a positive relationship with a movie's popularity rating only up to **6 languages**

A movie's run time contributes positively and significantly to popularity ratings on a small degree.

No simple interaction that can be generalized between genres and number of spoken languages on a movie's popularity rating

Whoaaa there's no rush!



The average film runtime is 99.692 minutes
(roughly standard of 1.65 hours)

- $\text{MAX}(\text{runtime}) = 1256$
- $\text{MIN}(\text{runtime}) = 1$

Films below average of 99.692 minutes **score below average** in popularity ratings (2.71). Film runtimes greater or equal to 99.692 minutes **score above average** (4.54).

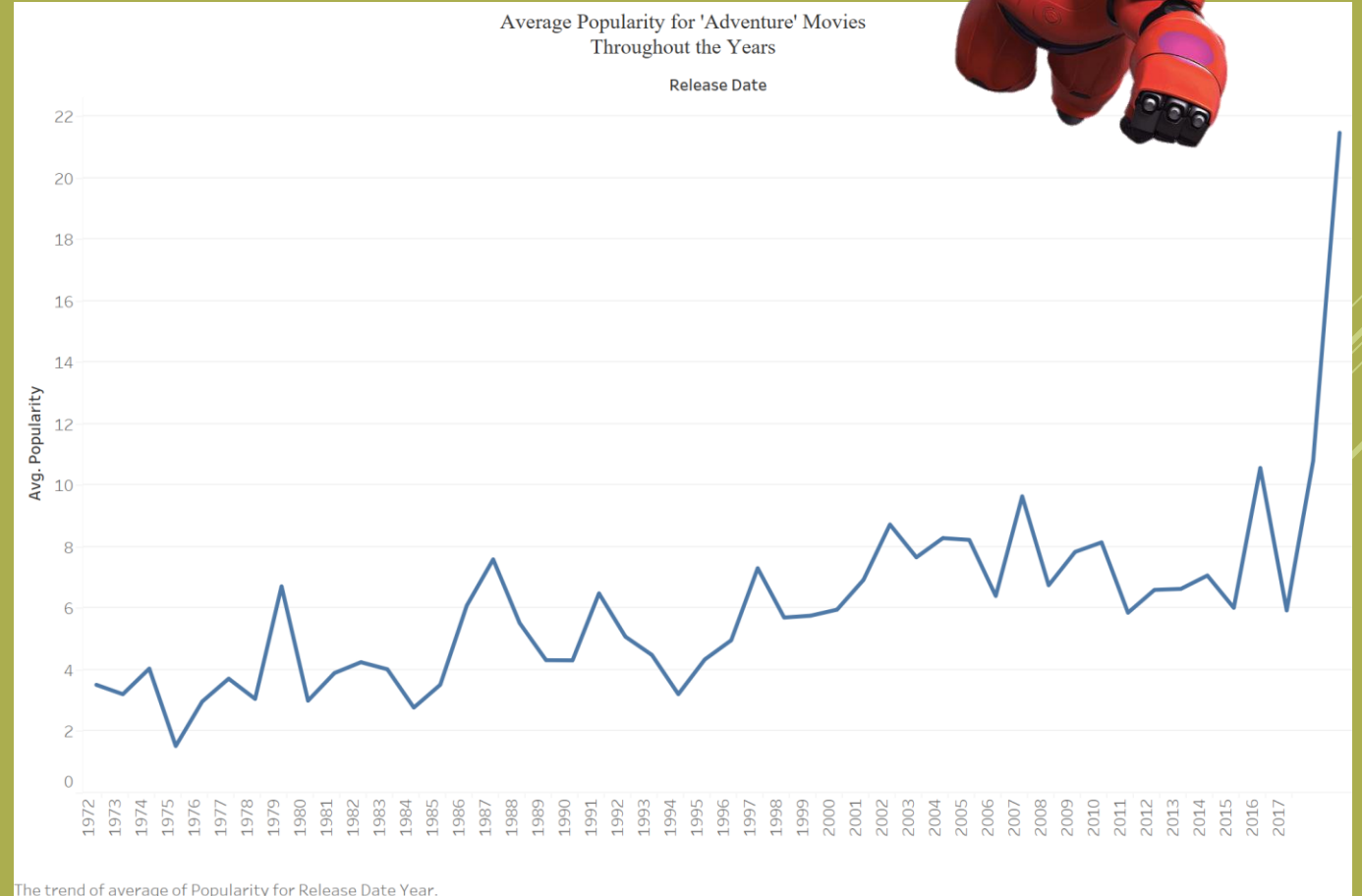
- Audiences (of TMDB) seem to favour standard film times

CONCLUSIONS & OTHER FINDINGS

Given data up to 2020, '**adventure**' genre films have spiked in popularity and received popularity ratings for the vast majority of years.

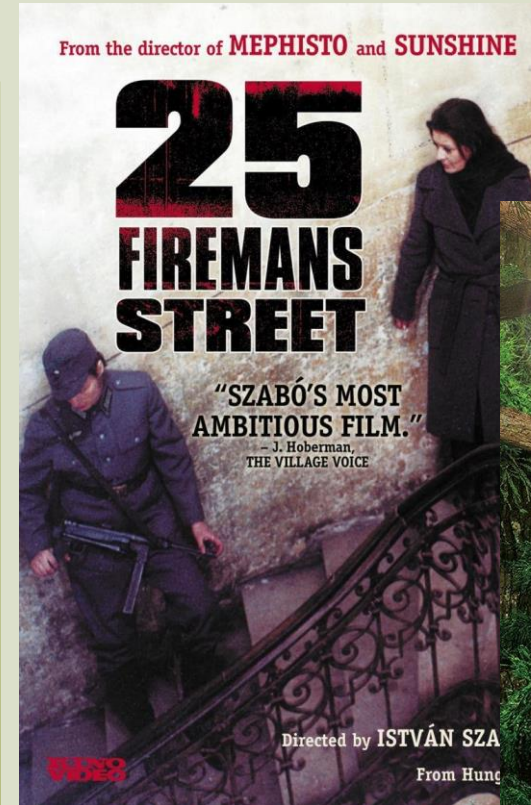


Producers should consider investing in **adventure** films best suited to audience's interests.



“FOREIGN” FILMS UNDERPERFORM

- These type of movies may require extra advertisement or recruitment of famed North-American actors to raise attraction
- “Foreign” movies contain only 1 and at most 2 spoken languages, which may be the prominent factor hindering their success
- The highest popularity rating of *all* ‘foreign’ movies is **only 2.56**
 - (100% below average!)

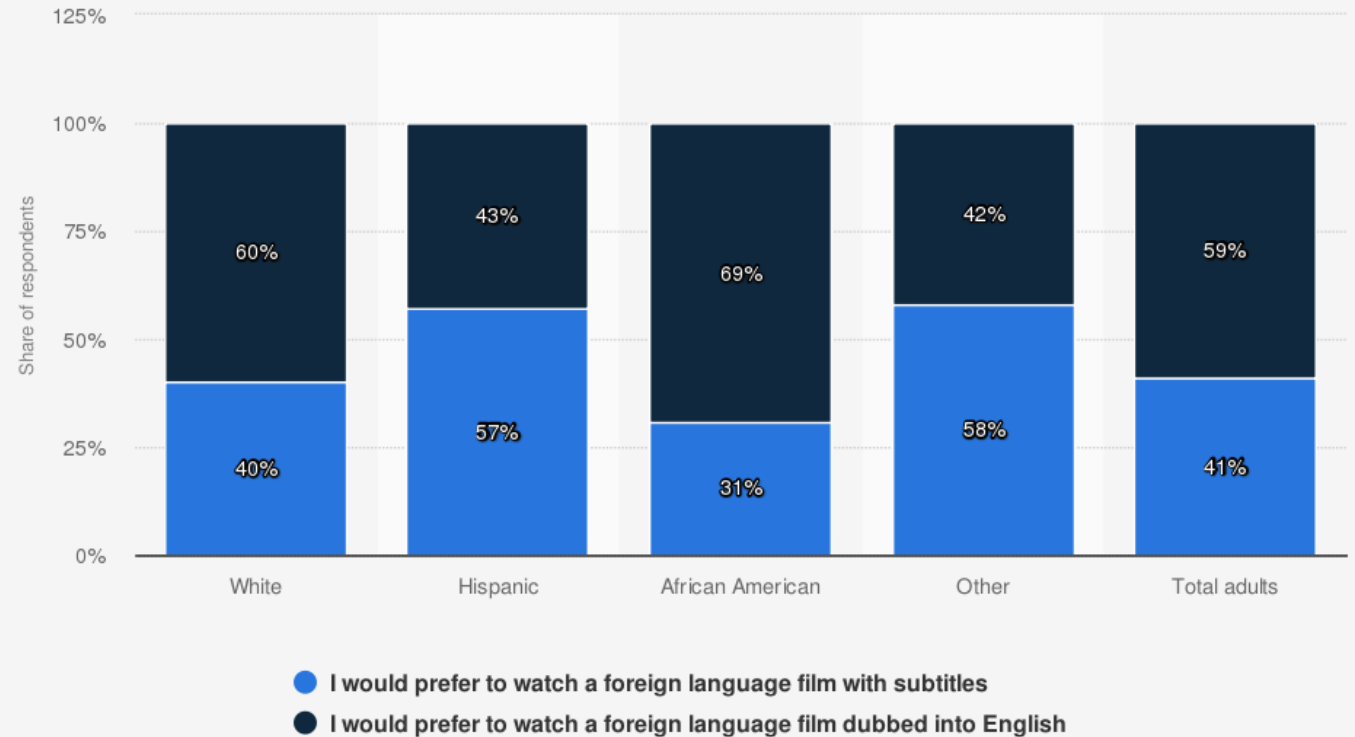


UNFAMILIARITY & LANGUAGE HASSLES

Figure from *Preference between watching foreign language movies dubbed into English or with subtitles in the United States as of January 2020, by ethnicity*



Preference between watching foreign language movies dubbed into English or with subtitles in the United States as of January 2020, by ethnicity



Sources
Morning Consult; The Hollywood Reporter
© Statista 2021

Additional Information:
United States; Morning Consult; The Hollywood Reporter; January 23 to 24, 2020; 2,200 respondents; 18 years and older;

LIMITATIONS

- ▶ Missing or incomplete data
 - ▶ Only movies with reviews on TMDb
- ▶ Duplicate, redundant, nonsensical data
- ▶ Only “primary genre” extracted
- ▶ “Outliers....” but not really



Avalanche	2.185485	/zaSf5OG7	[{'name': 'C	[{'iso_3166	2014/1/1	0	82	[{'iso_639	Released
Rune Bal	1.931659	/zV8bHuSL	[{'name': 'A	[{'iso_3166	#####	0	68	[{'iso_639	Released
- Written	0.065736	/ff9qCepil	[{'name': 'C	[{'iso_3166	#####	0	104	[{'iso_639	Released
- Murren	0.002130	/uadC6bll	[{'name': 'C	[{'iso_3166	#####	0	104	[{'iso_639	Released

Works Cited

Banik, R. (2017, November 10). *The Movies Dataset*. Kaggle. <https://www.kaggle.com/rounakbanik/the-movies-dataset>.

Popularity. The Movie Database. (n.d.). <https://developers.themoviedb.org/3/getting-started/popularity>.

Stoll, J. (2021, January 13). *Watching foreign language films with or without subtitles in the U.S. 2020 by ethnicity*. Statista. <https://www.statista.com/statistics/1095814/foreign-language-film-preferences-us-by-ethnicity/>.