Movie Trailers: Is "Popularity" the Key to Success?

Group 64: Granny Smith Geckos
TEAM MEMBERS: Amal Alhaidari (A98069538), Brandon Tong (A13957535), Quyen-Vi Nguyen (A13553619), Weijin Xu (A13913775)

RESEARCH QUESTION:

Are popular US movie trailers on Youtube predictive of the movie's success?

HYPOTHESIS:

If we assume "popularity" can be measured by the number of views, a higher like to dislike ratio, and a "trending" status, and that a movie's success is measured by considering its revenue, budget, and rating, then we hypothesize that the popularity of trailers on Youtube are positively correlated to the success of their movies.

BACKGROUND & PRIOR WORK:

Quantifying Movie Magic with Google Search:

A major finding of this study is that a movie's revenue on opening weekend can be predicted with 94% accuracy up to 4 weeks in advance of the title release-date. The sample they considered is about 60% of moviegoers with the revenue of the movie increasing when trailer search volume, the status of the franchise, and seasonality are considered. They define the status of the franchise as the number of theatres the movie is released in, and whether it is a highly popular and financially successful franchise. They define seasonality as the time of the movie release, being advantageous when released during major holidays and over the summer. An important note to make is that title-related searches on YouTube have the highest predictive power four weeks from the release date. These findings will help guide the direction our project will take, especially with regard to YouTube's title-search predictive power.

Mining Trailer Data from YT for Predicting Gross Income of Movies:

The observation of 7,988 movie trailers that stemmed from the years 2000 to 2017 has seen a positive correlation from movie trailer exposure to gross movie income using a linear regression model. Among other factors, they took into account the movie trailer's like to dislike ratio, comment count, and view count which they observed in August of 2017. Furthermore, the study's data included the movie's title, studio, release-year, release-date as well as the opening income and gross income from August 2017. The methods in which they conducted their observations were done using Box Office Mojo as well as YouTube Data API for movie data and the movie trailer's data respectively. From this, the inclusion of view count, like and dislike ratio, as well as the movie's gross income, has directed us to consider whether a movie trailer's data, such as a "trending" status, can predict other aspects of a movie's success such as its ratings.

ETHICAL CONSIDERATIONS:

For this study, we are considering factors such as like to dislike ratios and the number of views that a trailer garners, but the study would not be able to determine specific factors like demographics and age group which means the privacy of YouTube user's information would not be at risk. However, the

implications made from our findings could potentially allow advertising companies to exploit YouTube's "trending" algorithm. Although YouTube has made statements about their algorithm being automated, much of the information is not made public which causes speculation of whether or not outside influence can determine the popularity or "trending" status on Youtube. As such, whether our findings can be used for exploitation is a possible ethical issue. Potentially, the marketing of a movie's trailer could cheat the system by bypassing the algorithm via payment for the sake of greater revenue by establishing a "trending" status. Furthermore, the movie industry could use this information to generate films purely for financial gain where trailers can be enhanced to upsell the actual content of the movie which can be lower in quality comparatively.

DATA:

Database Name	Link	Description (as stated from source)
The Movies Dataset **(data obtained from the official GroupLens website)**	https://www.kaggle.com/rounakbanik/the-movies-dataset	(csv) Files contain metadata for all 45,000 movies listed in the Full MovieLens Dataset (movies released on or before July 2017). Data includes cast, crew, plot keywords, budget, revenue, posters, release dates, languages, production companies, countries, TMBD vote counts and vote averages (rating on a scale of 1-5).
IMDB Movie database	https://www.statcrunch.com/app/index.php?dataid=1736023	(spreadsheet) Last updated on January 2017. Data on over 50,000 movies before 2005. Includes title, year, length, budget, rating, votes (number of imdb users that rated), r1-10 (percentile of movies who have a rating of 1 through 10), MPAA rating, and genre.
Movie Budgets and Box office Earnings	https://www.statcrunch.com/app/index.php?dataid=2188684	(spreadsheet) Updated Spring 2018. This data comes from this website used to track the financial performance of movies. Columns: Month/Date/Year, budget, domestic gross, and worldwide gross
Trending Youtube Video Statistics	https://www.kaggle.com/datasnaek/youtube-new	(csv) Daily statistics (trending date, title, channel_title, publish time, views, likes, dislikes, comment_count, etc) of trending

		Youtube video statistics for the US, GB, DE, CA, FR, RU, MX, KR, JP, and IN. We are only considering US data.

TEAM'S EXPECTATIONS

Commitment:
1. Attend most Saturday end meetings (meet once a week as a group or before big deadlines) or notify the group a day/night in advance if you can't make it
   a. At least 3/4 people should be meeting at least every other week.
2. Meet up with another person over the week to work on respective sides of the project
3. Update the group on any advancements, progress, or roadblocks while working on the project through the group chat (and preferably update on Trello)
4. Finish all core work before the meeting time (at the latest) so meetings can be used to revise

Work Division:
1. We are all expected to partake/have a say in all sections of the project.
2. Hope and Weijin will take the brunt of coding (70% : 30%, tentative)
3. Brandon and Quyen-Vi will take researching and writing (70% : 30%, tentative)
4. Everyone, near the deadline date, will help with editing/revising

Project Timeline Proposal

*Color-coded by who's completing the tasks

Everyone	Researchers (Brandon and Quyen)	Coders (Hope and Weijin)

	Draft Text?	Write Code?	Proposed due date	Discuss at team meeting	Edit?
Initial team meeting	NA	NA	NA	[Sat|W2] 4/13	NA
Background Research	[W|W3] 4/17 [F|W3] 4/19	NA	[F|W3] 4/19	[Sat|W3] 4/20	[Sat|W3] 4/20
Question & Hypothesis	[Sat|W2] 4/13 [Tu|W3] 4/16	NA	[F|W3] 4/19	[Sat|W3] 4/20	[Sat|W3] 4/20
Ethical Considerations	[W|W3] 4/17	NA	[F|W3] 4/19	[Sat|W3] 4/20	[Sat|W3] 4/20
Dataset	[F|W3] 4/19 [Th|W5] 4/25	[Tu|W4] 4/23	[F|W3] 4/19 [Thu|W5] 4/25	[Sat|W3] 4/20 [Th|W5] 4/25	[Th|W4] 4/25
Data Wrangling	[Th|W4] 4/25 [Th|W5] 5/2	[Tu|W4] 4/23 [Tu|W5] 4/30	[W|W4] 4/24 [F|W5] 5/3	[Sat|W3] 4/20 [Th|W5] 4/25 [Sat|W5] 5/4	[Sat|W6] 5/11
Descriptive	[Th|W5] 5/2 [Th|W6] 5/9	[Tu|W5] 4/30 [Tu|W6] 5/7	[F|W5] 5/3 [F|W6] 5/10	[Th|W4] 4/25 [Sat|W5] 5/4 [Sat|W6] 5/11	[Sat|W7] 5/18
Exploratory	[Th|W5] 5/2 [Th|W6] 5/9	[Tu|W6] 5/7 [Tu|W7] 5/14	[F|W5] 5/3 [F|W6] 5/10	[Th|W4] 4/25 [Sat|W5] 5/4 [Sat|W6] 5/11	[Sat|W7] 5/18
Analysis - Part I	[Th|W6] 5/9 [Th|W7] 5/16	[Tu|W7] 5/14 [Tu|W8] 5/21	[F|W6] 5/10 [F|W7] 5/17	[Th|W4] 4/25 [Sat|W5] 5/4 [Sat|W6] 5/11	[Sat|W8] 5/25
Analysis - Part II	[Th|W7] 5/16 [Sat|W8] 5/25	[Tu|W8] 5/21 [Tu|W9] 5/28	[F|W7] 5/17 [F|W8] 5/24	[Sat|W6] 5/11 [Sat|W7] 5/18 [Sat|W8] 5/25	[Sat|W9] 6/1
Analysis - Part III	[Sat|W8] 5/25 [Th|W9] 5/30	[Tu|W9] 5/28 [Tu|W10] 6/4	[F|W8] 5/24 [F|W9] 5/31	[Sat|W7] 5/18 [Sat|W8] 5/25 [Sat|W9]6/1	[Sat|W10] 6/8
Summarize Results	[Th|W9] 5/30 [Th|W10] 6/6	NA	[F|W9] 5/31 [F|W10] 6/7	[Sat|W8] 5/25 [Sat|W9]6/1 [Sat|W10] 6/8	[Tu|Fi] 6/11
Conclusions	[Th|W10] 6/6	NA	[F|W10] 6/7	[Sat|W8] 5/25	[Tu|Fi] 6/11