

Twitter Reactions on Opening Weekend

Mark Coretsopoulos, Andy Han, Jeremy Krovitz
COMP 440 - Collective Intelligence

May 6, 2019

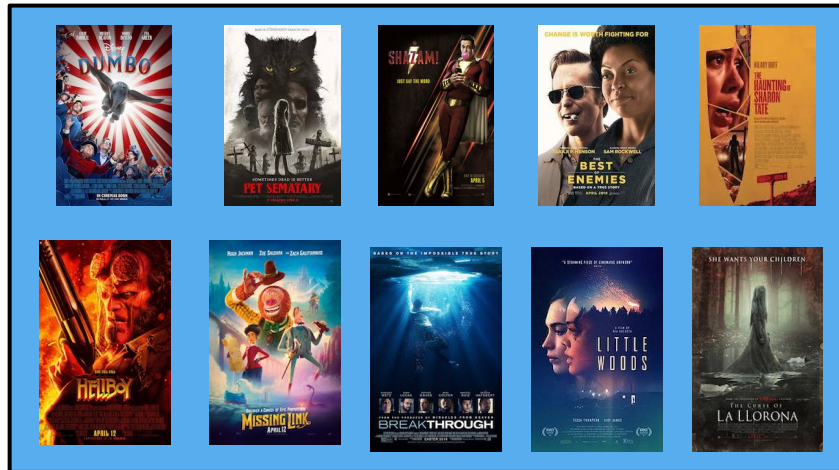
Agenda

1. Introduction
2. Research Question
3. Selection of Movies
4. Data Collection and Cleaning
5. Data Analysis
 - a. Sentiment
 - b. Literacy
6. Takeaways

Introduction



- Interested in analyzing tweets about movies coming out
 - 10 Movies (Over One Month of Releases) | Small, Medium, Large Budgets
- Academia is already working on similar topics



Research Questions

How does the sentiment and readability of tweets correlate with box office returns? How about with critical reception?



$$206.835 - 1.015 \left(\frac{\text{total words}}{\text{total sentences}} \right) - 84.6 \left(\frac{\text{total syllables}}{\text{total words}} \right)$$



Data Collection and Cleaning

- rtweet (R) / Tweepy (Python)
- Filtered by English tweets
- Removed duplicates and tweets containing “http”
- Significantly cut down the number of unwanted tweets



Sentiment Analysis

- Finding out the sentiment of each Tweet using Amazon Comprehend
- We did not expect running the tweets to cost money

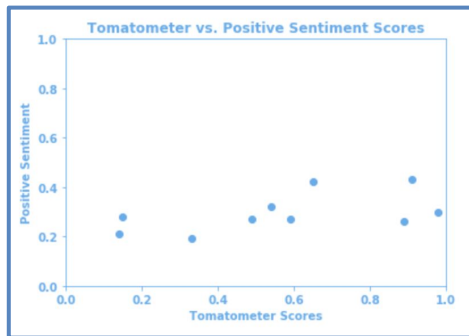
Example Output:

```
For the movie La Llorona, 12.0% of tweets were positive (average confidence score was 0.15814473159145565), 22.0% negative (average confidence score was 0.21091976046685887), 64.0% neutral (average confidence score was 0.576182504100725), and 2.0% mixed (average confidence score was 0.05475300889229402).
```



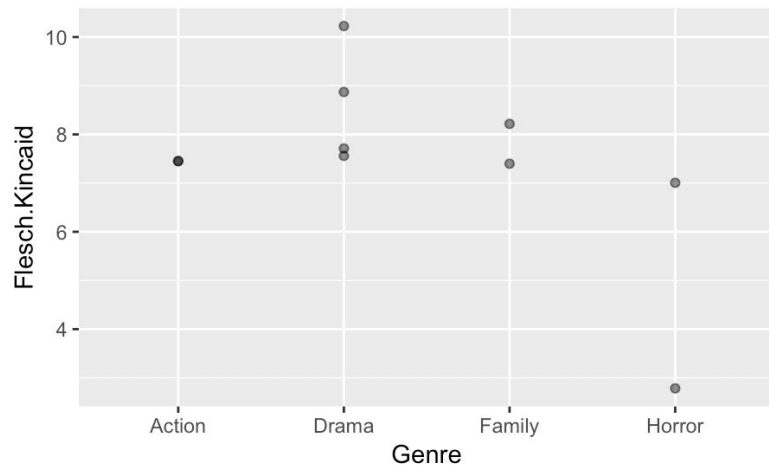
Results of Sentiment Analysis

- Analyzed **40K Tweets** from 10 Movies
- Interesting Observations:
 - Shazam reported the **highest percentage of positive tweets - 43%** and was also critically acclaimed, recouping its budget in its opening weekend
 - La Llorona was not received well, yet managed to do exceptionally well at the box office
 - Mild correlation between positive sentiment and Tomatometer critical and audience scores, almost none with box office



Readability Analysis

- Wanted to look for other potential predictors
- Flesch, Flesch-Kincaid
- Average grade level of 7.5
(22.7 avg words, 1.6 avg sentences)
- Outliers are The Curse of La Llorona (2.8)
and The Best of Enemies (10.2)
- Readability correlates more with genre than
critical reception or box office



Takeaways

- Usage of tweets to predict box office performance is inconclusive, might be helpful in extreme cases
- Different genres prompt different social media responses from audiences
- Sentiment analysis shows some correlation between positive sentiment and critical/audience reception



Questions?



Please support your local small theatres and independent movies!