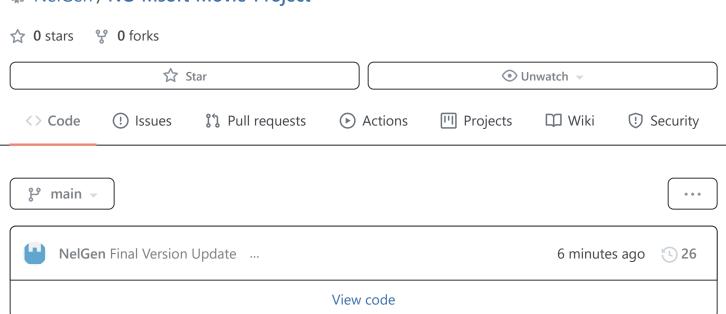
#### □ NelGen / NG-Msoft-Movie-Project





## **Exploratory Data Analysis on Ratings**

## **Business Case:**

Make recommendations to new Microsoft film studio

## Goals:

- 1. Discover top 5 genres in the US along with highly rated actors, actresses and directors suitable for that genre
- 2. Suggest suitable film lengths (run times) for those genres
- 3. Provide top keywords found in positive and negative critic reviews

### Sources:

• Used ratings from the provided IMDB and Rotten Tomatoes data sets

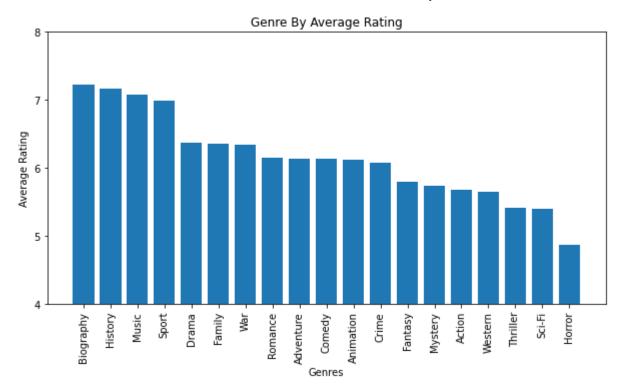
## **Strategies:**

• For Top genres:

- i. Subset the IMDB data for films in the US region
- ii. Removing genres with too few votes that could skew ratings
- iii. Further subset actors, actresses and directors by age less than 75 and alive
- For best run times:
  - i. Using the same subsets, group films by average ratings and run times
  - ii. Remove any run time outliers to exclude films that are significantly short or long
  - iii. Examine the clustering of run times against average rating
- For finding keywords for highest/lowest rated reviews
  - i. Associate the words in each review with the rating of that review
  - ii. Remove any stopwords and punctuations using data from the Natural Language Toolkit
  - iii. Remove any words that are rarely used that could skew ratings
  - iv. Combine all of these words into a master dataframe, sort, and find the highest/lowest rated keywords

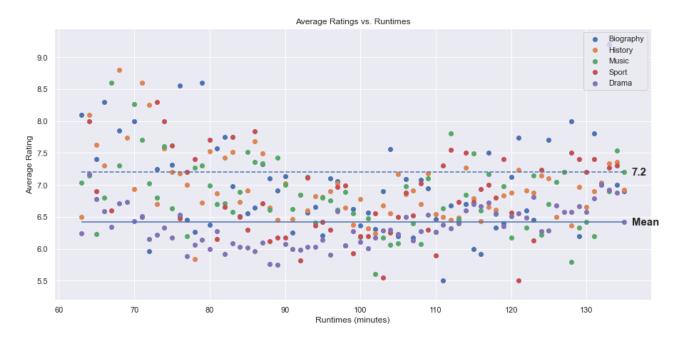
## 1. Top 5 Genres

Using the IMDB database, I found the top 5 genres in the US to be Biography, History, Music, Sports and Drama. The results were surprising as I feel the majority of movies were all action and zombie films. Perhaps most viewers do not share the same tastes as me.



## 2. Effect of Runtime on Ratings

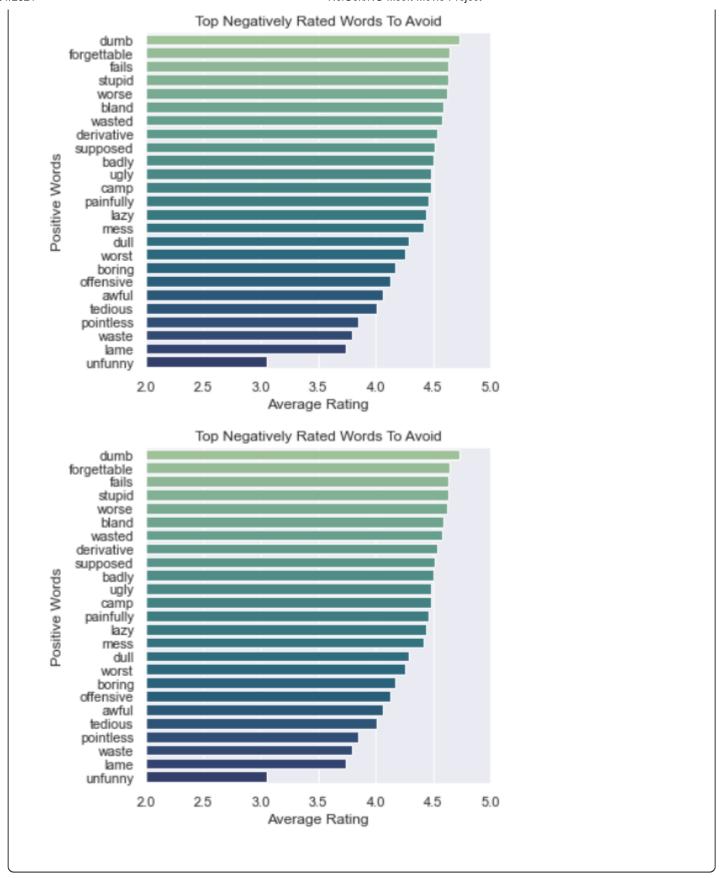
The highest rated films in the top genres are clustered around short (less than 90 minutes) and long (greater than 110 minutes)



I believe the clear winner here are films with run times less than 90.

# 3. Highest and Lowest Rated Keywords Found in Reviews

- Aim to make your audience feel intense emotions such as mesmerized, astonished, haunted. Mix their emotions with heartbreak, and sadness. However, also generate feelings of triumph and achievement.
- Avoid offending or boring your audience, lazy writing, or filling the time with forgettable content.



#### Releases

No releases published Create a new release

## **Packages**

No packages published Publish your first package

#### Languages

Jupyter Notebook 100.0%