### Day 1: Brainstorming

* use omdb api
* topic: trends among highly rated movies
  + how do we define highly rated - audience rating, critic rating, etc.
  + ~~atv shows vs movies?~~ too much
* how do we convert some of these metrics into numerical data

IMDB Ratings -[https://www.imdbcom./chart/top/](https://www.imdb.com/chart/top/)

~~Letterboxd API -~~ [~~https://api-docs.letterboxd.com/~~](https://api-docs.letterboxd.com/)

OMDb API - <https://www.omdbapi.com/>

—------------------------------------------------------------------

* **Rating Factors**
  + Budget
  + Actors/ Directors
  + Is it an original film or is it an ongoing franchise?
  + Was it a book adapted into Film? (example: Shawshank redemption, IT, Lord of The Rings etc)
  + Critic vs Audience scores
* Ratings in API
  + Rotten tomatoes, metacritic, IMDB, Letterboxd

#### **Potential** trends

* Do we go by past decade?
* How ratings have changed over time?
* Professional reviewers vs audience reviews?
* Time of year,
* Do critics from certain reviewers (e.g. imdb vs rotten tomatoes) rate lower/higher on average?
  + Convert to a 10 point scale to compare?
* Genres per time of year?

**Potential Questions:**

* In which situations would make a movie more highly scored
  + Criterion:
    - Genre
      * Time of year / seasonal
      * Decade / Section of years
      * Target Audience / Rating (e.g. G/PG, PG-13, R)
      * Language / Country Made
      * Franchise
      * Awards / Nominations
      * Boxoffice
      * Streaming

#### **Steps** to be completed

* Pulling, cleaning data
* Generating

#### Project requirements

#### **Presentation requirement**

* 10 minutes + questions
* Questions you found interesting and motivation to answer them
* Where and how you found data used to answer questions
  + - Data exploration and cleanup (accompanied by jupyter notebook)
* Analysis process (accompanied by jupyter notebook)
* Conclusions including numerical summary and visualizations of summary
* Implications of findings: What do they mean?

**Development Requirements**

* Use pandas to clean and format datasets
* Create jupyter notebook describing data exploration and cleanup process
* Create jupyter notebook illustrating final data analysis
* Use Matplotlib to create 6-8 visualizations
  + About 2 per question answered
* Save PNG images of visualizations to distribute to class and instructional team
  + Include in presentation
* Create write up summarizing major findings
  + Include heading for each “question” you asked your data as well as short description of findings
* Use at least one API if you can find one with data pertinent to your primary research questions

**Proposal:**  
  
Using the OMDB API, we will be looking to answer:

In what situation is a movie most likely to have a higher rating?

Filtered by genre

We will be looking to find a correlation between a movie’s average rating over 3 different websites (Rotten Tomatoes, IMDB, and MetaCritic) from 2010-2019 through the lenses of:

* Movie Ratings (G, PG, PG13, R)
  + Do movies with specific ratings tend to have higher average movie scores?
* Box Office
  + Does a movie’s box office profit have an effect on average rating? How does this affect different genres
* Release Time of year
  + How does time of year influence movie ratings?
  + Need to define time of year in readme (3 month chunks)?
  + By day for each day
* Run Time
  + How does the length of a movie affect its ratings? Does this differ between genres?

Roles

Project Manager - Nick OD

Cleaning Data - DuVoe

Question Assignment:

* Movie ratings(PG13, R, etc) - James
* Box Office- Makenna
* Release Time of Year - Ethan
* Run time - Zach