

Project 2 Proposal - Group 3

Alyssa DiFurio

Garrett Kidd

Vanessa Martinez

Japhet Mwamba

We want to investigate movies using IMDB and OMDB API. We want to use the top 250 movies rated on IMDB to conduct this analysis and the OMDB API to extract the metadata regarding these movies. We will use web scraping on the websites that have the lesser available datasets available to us on kaggle. We plan to only scrape the sites once and keep a timestamp on it, in order to keep the data from adjusting too much during our week of work.

Our inspiration was that we all love movies, we're curious to see any patterns in these movies, and it was a popular ETL topic we found online. The data that we are pulling is from a top 250 list, so it will be interesting to see where our thoughts and knowledge on popular movies lie, versus what's critically rated. We also wanted to nerd out and look at movies since we all have some on the top 250 list that we like.

Data Sources:

1. IMDB Top 250 Movies
2. OMDB API

Data Tables:

1. IMDB Top 250 Movies
2. Directors
3. Main Actors
4. Genres
5. Rating

Delegation of Duties:

1. Garrett: Web Scraping, API calls, write-up
2. Alyssa: write-up, Github setup
3. Vanessa: write-up, ERD
4. Japhet: write-up, tbd