Project Proposal - DS Team Project #2

**Project Team:**

Bill Johnson, Donterrius Daniels, Tamekia Phillips, Ramirra Marshall, Henry White, Adam Armagost, Ryan “Fin” Finley, Chineze Okpala, Evan Mickler, Antione Perrymon, James Niu

**Project Requirements:**

Project Proposal (20pts)-

Data Cleanup and Analysis Requirements

* Citing the data sources from which you will extract.
* Extracting the data from their existing locations.
* Transforming the data (i.e. cleaning, joining, filtering, aggregating, etc).
* Loading the data to a database (relational or non-relational).

Technical Report (20pts)-

* Extract: your original data sources and how the data were formatted (CSV, JSON, pgAdmin 4, etc.).
* Transform: what data cleaning or transformation was required.
  + Use Pandas to clean and format your dataset(s)
* Load: the final database, tables/collections, and why this was chosen.
* Leverage a Team GitHub repo

GitHub Repo (20pts)-

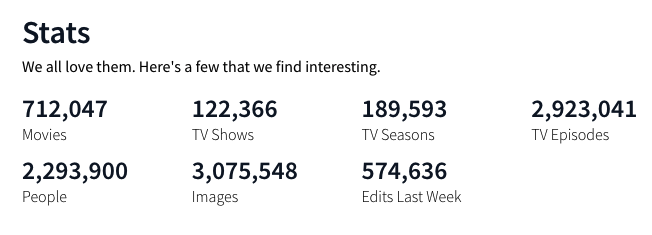
* https://github.com/wpj174/project-two.git

**Project Proposal:**

Our Team's goal is to extract and transform Movie data from various sources and combine them into a single searchable database to build a comprehensive movie profiles.

**Data Sources (citations):**

* Kaggle: Rotten Tomatoes movies and critic reviews dataset (CSV)
  + Source Link: [Rotten Tomatoes Movies](https://www.kaggle.com/stefanoleone992/rotten-tomatoes-movies-and-critic-reviews-dataset?select=rotten_tomatoes_movies.csv)
  + Context: Movies' data is stored on several popular websites, but when it comes to critic reviews there is no better place than Rotten Tomatoes. This website allows users to compare the ratings given by regular users (audience score) and the ratings given/reviews provided by critics (tomatometer) who are certified members of various writing guilds or film critic-associations.
  + Stats:
    - 17,000 Movies
* The Movie DB (API)
  + Source Link: <https://www.themoviedb.org/>
  + API Documentation: <https://developers.themoviedb.org/3/getting-started/introduction>
  + Context: The Movie Database (TMDB) is a community built movie and TV database. Every piece of data has been added by our amazing community dating back to 2008. TMDb's strong international focus and breadth of data is largely unmatched and something we're incredibly proud of. Put simply, we live and breathe community and that's precisely what makes us different.



* Kaggle: Metacritic Movie Reviews (CSV)
  + Source Link: [Metacritic csv](https://www.kaggle.com/miazhx/metacritic-movie-reviews)
  + Context:
  + Stats:
    - 9,100 Movies
* Kaggle: IMDB dataset (CSV)
  + Source Link: [IMDB csv](https://www.kaggle.com/stefanoleone992/imdb-extensive-dataset/settings?select=IMDb+movies.csv)
  + Context: IMDb is the most popular movie website and it combines movie plot description, Metastore ratings, critic and user ratings and reviews, release dates, and many more aspects.

The website is well known for storing almost every movie that has ever been released (the oldest is from 1874 - "Passage de Venus") or just planned to be released (newest movie is from 2027 - "Avatar 5").

IMDb stores information related to more than 6 million titles (of which almost 500,000 are featured films) and it is owned by Amazon since 1998.

* Stats:
  + 90,000 Movies

**Data Base:**

MongoDB -

* Name: GT\_MOVIES\_DB

SQL Postgres -

* Name: GT\_MOVIES\_DB

**Data Transformations & Cleaning:**

* COMMON INDEX: IMDB\_ID
* LANGUAGE: ENGLISH
* TIME RANGE
  + ORIGINAL RELEASE DATE (ALL)
  + STREAMING RELEASE DATE (ROTTENTOMATOES)
* TITLE (ALL)
* DIRECTOR(S) (ALL)
* AGE RATING (ALL)
* GENRE(S) (ALL)
* AUDIENCE RATING/REVIEWS (ALL)
  + RATING BY SOURCE RANGE
* BUDGET (ALL)
* RUNTIME (both)
* MOVIE HOMEPAGE
* PRODUCTION COMPANY (ALL)

**Project Outputs:**

| **Topic** | **Output** | **ETA** |
| --- | --- | --- |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

**Team Roles**

| Adam Armagost | **API** **Pull** |
| --- | --- |
| Bill Johnson | **API Pull** |
| Antoine Perrymon | **Data Cleaning** |
| Ramirra Marshall | **PostgresSQL Pull** |
| Henry White | **Aggregation of Data** |
| Chineze Okpala | **Data Cleaning** |
| Donterrius Daniels | **PostgresSQL Pull** |
| Evan Mickler | **Aggregation of Data** |
| Ryan Finley | **PostgresSQL Pull** |
| Tamekia Phillips | **Data Cleaning** |
| James Niu | **Presentation** |