**CIS4400 Homework 1**

**Chosen Dataset: IMDb Datasets** **Link:**<https://www.imdb.com/interfaces/>

**Business Requirements**

1. Offer valuable insights into film data from around the world, especially trends related to genres, ratings, and production over time.
2. Create a system that captures movie metadata (e.g., title, year, genre, runtime, etc.) that can be used for analysis with ratings and key crew members.
3. Enable the ability for users to interact with the system to receive specific information about movies (e.g., top-rated movies by genre) or a director's activities over time.

**Functional Requirements**

1. The system must allow for querying the details of a film based on its title, release date, or genre.
2. Users must be allowed to view the average rating for one or more genres over time and compare it to another genre's average rating.
3. The back-end of the system must support the linkage of data across titles, ratings, and crew information.
4. The system must also allow for filtering movies by crew member (e.g., all movies by a particular director).
5. Allow users to download/export the information to be retrieved for offline analysis.

**Data Requirements**

* **Source:** IMDb official public dataset (<https://www.imdb.com/interfaces/>)
* **Files Used:** title.basics.tsv, title.ratings.tsv, title.crew.tsv, name.basics.tsv
* **Size:** More than 10 million movie titles
* **Columns:** Greater than 50 (combined) columns across all files
* **Example Columns:** titleType, primaryTitle, startYear, genres, averageRating, numVotes, directors, writers, birthYear, knownForTitles
* **Data Dictionary:**<https://www.imdb.com/interfaces/>

**Information Architecture**

**Description:**

* Users access the system to run queries or view data dashboards.
* The front-end communicates with a backend server that receives the queries and pulls from a normalized relational database.
* Data flows from the original .tsv files to a cleaned and transformed database.
* The user account includes a storage method for remembered users and/or favorites, and may retrieve additional queried data based on user history or other relevant factors.