

# Milestone 1: Project Proposal 550 Spring 2023

---

Group Member
--------------

**Songyang Du:** sydu@seas.upenn.edu; [Github](https://github.com/SongyangD): https://github.com/SongyangD

**Shengyin Si:** sisy@seas.upenn.edu; [Github](https://github.com/Shelly0104): https://github.com/Shelly0104

**Yahan Xu:** yahan@seas.upenn.edu; [Github](https://github.com/gooyoowoo): https://github.com/gooyoowoo

**Yu Yang:** yuyang89@seas.upenn.edu; [Github](https://github.com/yuyang2048): https://github.com/yuyang2048

## 1 Application description

A simple movie search website which allows for users to type in search terms and receive corresponding results. The displayed relevant information including poster image, title, ratings, genre, actors, etc. Additional features including daily recommendation, award records (Oscar).

## 2 Dataset

All the tables are not in 1NF as some of the columns contain multi-value.

### 2.1 File

Link: [kaggle](#); Owner: [Neha](#)

- Format: csv
- File size: 9.13MB
- Total Rows: 40109
- Total Tables/Relations/csv files: 1
- Unique movie titles: 39470
- Column: imdbid, imdb link, title, IMDB score, genre, poster URL

### 2.2 Actor

Link: [kaggle](#); Owner: [Trent Park](#)

- Format: csv

- File size: 47.93MB
- Total Rows: at least 83,000 rows each file, name.csv has 297,711 rows
- Total Tables/Relations/csv files: 4 ie: movies.csv, name.csv(used), ratings.csv, title principals.csv
- Column: imdb name id, name, birth name, height, bio, birth details, birth date, place of birth, death details, date of death, place of death, reason of death, spouses string, spouses, divorces, spouses with children, children

## 2.3 Oscar Award

Link: [kaggle](#); Owner: The Devastator

- Format: csv
- File size: 830.82 kB
- Total Rows: 562
- Total Tables/Relations/csv files: 1
- Column: Id, ID FA, title, original title, year, duration, country, director, actors, genre, description, awards, average rating, rating votes, reviews, poster, my vote.

## 2.4 Actor poster (Web scraping)

Using python 3 imdb module to scraping imdb.com to get the image of the actor. Relevant web scraping tutorial has been found [greeksforgreeks](#).

# 3 Queries

A list of at least 5 queries (in natural language) you could write for your datasets. Some of these should require complex SQL (aggregations, subqueries, joins of 3 or more tables, etc.)

## 3.1 Query 1: Movie information (getMovie)

- **Description:** Returns all the information about a movie with given movie\_id.
- **Tables:** movie table, movie\_cast table.
- **Schema:** title, release\_year, director, language, genre, duration, cover\_url, region, description, casts

## 3.2 Query 2: Actor information (getActor)

- **Description:** Returns all the information about an actor with given actor\_id.
- **Tables:** actor table, movie table, movie\_cast table.
- **Schema:** name, DOB, bio, movie\_participated, movie\_cover\_url, actor\_url, Birth\_place

### 3.3 Query 3: Movies list(searchMovies)

- **Description:** Returns an array of movie with all their properties matching the search query which include genre, release\_year, and ratings, ordered by release\_year (descending).
- **Tables:** movie table, rating table
- **Schema:** title, release\_year, director, language, genre, cover\_url

### 3.4 query 4: Actors list(getActors)

- **Description:** Returns an array of all actors with their name, top3 movies participated and url of the actor's portrait.
- **Tables:** movie\_cast, rating table
- **Schema:** name, movie\_participated, actor\_url

### 3.5 Query 5: Top 20 movies by selected Genre(GetTop20Genre)

- **Description:** Description: Returns movie title, poster url and the brief description of the movies in selected genre with the top 20 ratings sorted by rating.
- **Tables:** movie table, rating table
- **Schema:** title, cover\_url, description, movie\_id

### 3.6 Query 6: Random Oscar best-picture nominated movie(randomOscar)

- **Description:** Return a random Osacar best-picture nominated movie.
- **Tables:** oscar\_movie table
- **Schema:** movie\_id, title

### 3.7 Query 7: Top X rated Oscar Nominated movie - best picture (oscarRating)

- **Description:** Return a list of top rated Oscar movies with a given rating. The results should be sorted in descending order by year and limited to the first 10 rows.
- **Tables:** oscar\_movie table
- **Schema:** title, year, votes, rating

### 3.8 Query 8: Genres of Oscar Nominated movie - best picture(getOscarGenres)

- **Description:** Return the genres of movies which have been nominated for Oscar. The results should be sorted in ascending order by genre name.
- **Tables:** oscar\_movie table
- **Schema:** genres

### 3.9 Query 9: All best picture Oscar nominated movies in given year

- **Description:** Return list of movie which have been nominated for Oscar for a given year. The results should be sorted in ascending order by title.
- **Tables:** oscar\_movie table
- **Schema:** title, year, votes, rating

### 3.10 Query 10: All Actresses in the best picture Oscar-nominated movies

- **Description:** Return a list of actresses who have been cast in a given Oscar nominated movies.
- **Tables:** oscar\_movie table, actor table
- **Schema:** name, birth\_year, movie\_year