Group Name: Big Screen

Group Members: Abdallah Aboelela, Ahmed Karkoura, Scarla Pan

Project Description & Goals:

We hope to collect data from Rotten Tomatoes, Oscar wins and nominations, and IMDB to create a movie recommendation site that allows users to input a genre, actor/actress, director, and movie duration and get a list of relevant movies that are both critically acclaimed and highly-rated by moviegoers. If we have enough time, we also hope to use the data acquired to train a model that given a user-selected genre, actor/actress, director, and movie duration combination would be able to predict box office, critic score, and audience score.

Data Sources:

1. Rotten Tomatoes - scraping directly from Rotten Tomatoes website
2. Oscars (wins, nominations) – csv, https://datahub.io/rufuspollock/oscars-nominees-and-winners

Tasks and Timeline:

1. *Data collection (1 – 1 ½ weeks)*

   For rotten tomatoes, this requires scraping the site to create a database that for each movie identifies the genre, main actor(s)/actress(es), director, movie duration, box office, critic score, audience score, etc.

   Rufuspollock has a dataset of all Oscar nominees and winners for all categories downloadable in the above website. We would limit our search to categories related to movies, actors, and directors specifically.

2. Building the recommender (1 – 1 ½ weeks)

   This recommender would filter the database given the user inputs and create a list of relevant movies that have the highest average critic and audience score. We would also include Oscar nominees and winners should they fit the criteria.

3. (POTENTIAL) Training regression model(s) to get predictions (2 weeks)

   Listed in order of priority to the team:

   a. Given information on actors, genres, director, and budget, this model would predict the box office results a movie may receive
   b. Given information on actors, genres, director, and required box office returns, this model would predict the required minimum budget
   c. Given a targeted box office number, and information on two of actors, director, and genre, these models would predict which value of the third variable would maximize box office returns

4. User interface/website (1 – 1 ½ weeks)

   Design and build a website where users could input their preferences and get a list of recommended movies.