Predicting Movies' Gross Earnings using Linear Regression

Project Proposal by: Himani Kaushik

- **Question/need:**
  - What is the framing question of your analysis, or the purpose of the model/system you plan to build?
    - The purpose of the model is to predict the gross earnings of a movie at the box office to decrease the financial risk. The model will also aim to predict the most relevant features that affect the earnings of the movie.
  - Who benefits from exploring this question or building this model/system?
    - Anyone interested in financing or investing in the movie, including individual producers, film studios, private investors, etc.
- **Data Description:**
  - What dataset(s) do you plan to use, and how will you obtain the data?
    - The data will be obtained from IMDB (Internet Movie Database), a website that provides the most extensive information about millions of movies.
    - The data will be scraped from the website using web scraping packages from python.
  - What is an individual sample/unit of analysis in this project? What characteristics/features do you expect to work with?
    - The individual sample includes title, release date, runtime, IMDB rating, gross earnings, genres, certificates, directors, writers, top three stars, opening weekend USA, etc.
    - I expect to work on all columns from the data.
  - If modeling, what will you predict as your target?
    - I will predict the gross earnings of a movie.
- **Tools:**
  - How do you intend to meet the tools requirement of the project?
    - Selenium and BeautifulSoup: Web scraping
    - Pandas and Numpy: Manipulating data
    - Matplotlib and Seaborn: Visualizing and plotting data
    - Statsmodels and Scikit-learn: Linear Regression
  - Are you planning in advance to need or use additional tools beyond those required?
    - Yes, I may use Scrapy for web scraping and other regression models to analyze and visualize data.
- **MVP Goal:**
  - What would a minimum viable product (MVP) look like for this project?
    - MVP for the project would be a model that could predict gross earnings of a movie using regression.