Movie Box-office Prediction Model:

The goal of this project is to make a model based on available data to predict the box-office of a movie. The focus of the model will be on three different features: Pre-Release Rating, Social Media Engagement, and ad budget.

Gathering Dataset for training the model:

For training the model based on the 3 chosen features, we need a decent dataset that include those 3 features and the Box-office of the film. I will primarily focus on year 2017 to 2019 to avoid having lots of data for training. I chose the box office mojo website which provide the great data about the box-office of movies. I manually scrap the data from the website and save it to the csv file which is available on project repository. For the chosen years, this dataset has 12,526 values including the worldwide, and domestic box-office for movies.

Dataset for rating:

For this feature, I chose Rotten Tomato, and IMDB website. The data was available on web in csv format already. For those three years, this dataset has 1,761 values which include critics and people rating for the movies. I also used IMDB data set, with 8,573 values about different rating system.