I'm interested in figuring out whether those stereotypical thinkings are right or not based on a statistician or a data scientist angles. Can famous actors, director, or both bring the highest revenue? How social media and networking service company ex: Facebook influence the movie industry? Is a high IMDB score a good sign for movie companies that they are gonna make money for sure? All in all, I want to know what are the most important features for a successful movie? Can we actually create a model to predict the profit for a movie? Therefore, the movie company can use the model to estimate their strategy in producing a movie.

The dataset is from the Data World website, although it was originally post on the Kaggle competition (The post has been replaced).

As the prediction is movie's revenue (gross - budget), I would like to use and compare both classification (Classify revenues in to different groups) and regression (Use the number from the dataset directly) approaches on this supervised data. In terms of predictors, I want to use director's names, actors names, IMDB score, facebook's likes (actors, directors, and movies), number of critical reviews on imdb, number of users reviews, Number of people who voted for the movie, and content rating of the movie.

For the training data, I am planning to split the whole dataset into two parts: ⅓ of dataset is testing data and ⅔ of dataset is training data.