# Project Proposal

## Data Description

The dataset contains 651 instances of movies released in theaters between the period of 1970 – 2014. The variables of the data set include IMBD ratings, rotten tomatoes critics and audience ratings and scores, actors and directors, movie runtime, genre, releases dates, award nominations and wins, and more. Some variables will be deleted as it has no value to my intended statistical analysis such as the list of actors.

## Target Problem

Is there a correlation between the critic score and audience score? Or genre and the audience score? How about MPAA rating and audience score? My target problem is to find what variables affect the audience score and whether we can predict the audience score using the variables such as genre, critic rating, IMBD rating, release date, and runtime.

## Statistical Problem

My statistical problem will using the given variables such as genre, MPAA rating, release dates, IMBD rating, and critic scores to predict the audience score.

## Proposed Statistical Analysis

My proposed statistical analysis will be using a linear regression model to predict the audience score of the movies. Some initial research shows that there can be minimal correlation between certain variables to the audience score such as movie runtime.