# CS5300 FINAL PROJECT

My current plan is to make an AI that can predict how much money a movie will make based on certain parameters:

* Budget
* Rating
* Genre
* Duration

There are many more possibilities, but at first I want to succeed with a smaller set of attributes.

## Data Clean Up

* Since Gross Revenue is what I intend to classify, removed all empty values
* Restricted to US because other revenue values were obviously not in dollars
* Ratings were a challenge
  + Made all ratings agree with modern ratings. Some older movies had “passed” or “approved” which doesn’t agree with modern ratings. There were a few that were mislabeled.
  + Changed ‘Unrated’ to ‘Not rated’ as there were more of the latter
* Removed duplicates
* Ended with 2980, almost 3k.

## Codings

**Ratings**

1. G
2. PG
3. PG-13
4. R
5. NC-17
6. Not Rated

**Genres**

* Action
* Adventure
* Animation
* Biography
* Comedy
* Crime
* Documentary
* Drama
* Family
* Fantasy
* Film-Noir
* History
* Horror
* Music
* Musical
* Mystery
* Romance
* Sci-Fi
* Sport
* Thriller
* War
* Western