Domain:

I want to take the First Impressions V2 (CVPR'17) data set and do an investigation on what factors are important on getting past an initial screening on an interview. This data set gives an interview score that corresponds with the 15 second video clip.

Data:

The data set contains 10000 annotated and transcribed 15 second clips of people talking. Each of these clips are rated on personality traits and an interview score ("How likely are you to hire this person").

Known Unknowns:

Working with deep learning present challenges that I have not faced before. Working with image and audio data will be new areas as well. There are well over 10 GB of raw data so working with that in an efficient manner could be tricky as well.

Game Plan:

1) Create baseline guess the average
2) Create model using one image from each interview
3) Create model that uses multiple images from each interview
4) Create model that uses multiple images and audio
5) Create model that uses multiple images audio and text transcriptions
6) Create model that uses multiple images audio text and gender/race/age

Fall Back Plan:

Extract features from the images / videos like beauty, more detailed race, what kind of clothes etc and feed into a simple logistic regression to understand if there is any racial or gender bias in this dataset.