Capstone Check-in #1

## Stephen Strawbridge, Cohort #1019

***Topic 1: Gender and Age Detection***

Problem Statement: Given an image, I will predict the gender and age (or age range) of the individual in the image.

Audience: Online advertising agencies, such as Facebook/LinkedIn (that will use this prediction to give appropriate advertisements to the individual based on their profile picture).

Success Metric: How accurate is my modeling in predicting gender (binary classification) and age range (multi-classification).

Data Source: [Link](https://www.kaggle.com/ttungl/adience-benchmark-gender-and-age-classification)

Potential Challenges: I have to learn new Python libraries. Age detection will likely be very difficult for ages 18 to 35, which is why I might have to do age range instead of a continuous age prediction variable.

***Topic 2: Handwritten Digit Recognition***

Problem Statement: Given an image, I will predict the number written in the image.

Audience: Banks using scanning technology for check deposits.

Success Metric: How accurate is my modeling in predicting a handwritten digit (binary classification).

Data Source: [Link](http://yann.lecun.com/exdb/mnist/)

Potential Challenges: I will have to learn new Python libraries. Sloppy handwriting could make this challenging. I will also need to build a GUI to draw digits to be used for predicting.

***Topic 3: Customer segmentation - Netflix***

Problem Statement: Given a user’s recently watched shows on Netflix, recommend shows they may like.

Audience: Netflix company.

Success Metric: Because this would be unsupervised modeling, there would be no prediction score, but rather suggested groups (using K-means clustering).

Data Source: [???](http://inalitic.com/datasets/nhl%20player%20data.html)

Potential Challenges: Finding data for Netflix will be tough. Success metric will be subjective and harder to measure, as there is no prediction score.