**Prompt**

Many jurisdictions around the United States are using risk assessment instruments (RAIs) in helping judges make bail decisions. Pre-trial RAIs are often statistical models that try to predict the likelihood that an individual will commit a crime if released on bail pending their court date. For the past several years, the courts in Broward County, Florida have been using one of the RAIs in the COMPAS suite to inform their decisions. In May of 2016 an investigative journalism team at ProPublica published a [report](https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing) that analysed whether COMPAS might be racially biased. You will use the publicly released data from their analysis in performing the key tasks below.

Key tasks

1. The COMPAS tool was not developed on the Broward County population. Going forward, the County is considering developing their own RAI to replace COMPAS. Using the available data, construct an RAI for predicting two-year recidivism. Evaluate the predictive performance of your model. What are the most important predictors of recidivism?
2. Construct an RAI for predicting *violent* recidivism. Evaluate the predictive performance of your model. What are the most important predictors of violent recidivism? How do they compare to the important predictors of general recidivism?
3. Are your RAIs from (1) and (2) equally predictive across race/ethnicity groups? How about across age and sex groups?
4. Compare your RAIs to the COMPAS RAI. Do your RAIs perform better or worse than COMPAS? Do your RAIs produce similar classifications to COMPAS? Can you identify any systematic differences between your classifications and those of COMPAS?

**Steps for the team:**

1. Randomly split data into training and validation (so we don’t make [this](https://canvas.cmu.edu/courses/14656/modules/items/4438423) mistake)
   1. KR unknowns
      1. What ratio should we use?
      2. Which dataset should this be?
2. Run different classification models that work for binary outcomes
   1. Potential models
      1. LDA
      2. QDA
      3. Naïve Bayes
      4. Random forest
      5. Logistic Regression
   2. Outcome variables
      1. Any recidivism: is\_recid
      2. Violent recidivism: is\_violent\_recid
         1. Why isn’t this used in the ProPublica notebook, given that it’s in their dataset? I’m confused
   3. Apply cross-validation to each method, to pick the best model
   4. Run confusion matrices for each model separately by demographics.
      1. Look at measures other than accuracy
         1. Specificity (1 – FPR)
         2. Sensitivity (TPR)
      2. 3 demographic groups
         1. Age
         2. Race
         3. Gender
   5. COMPAS didn’t use race as in input, but I think we should try it and see what happens. Does this *reduce* racial bias?
3. Compare to Compas
   1. I think we can just do this by looking at ProPublica’s output – not necessarily running it ourselves (unless we have extra time).