Math 327 – Data Analysis Project #2 – Proposal

* Team members:
  + Name 1 Mbonisi
  + Name 2 Tanuj / Rick
* Data set to analyze
  + Description / background: This dataset is set up to contain measurements of certain attributes of a currency bill, and infer if that specific bill is fake or genuine.
  + Response variable – Need 1. Must be binary categorical (2 levels)
    - is\_genuine: the response variable, tells us if the bill is genuine or not.
  + Predictor variables – Need at least four. Can be a mixture of quantitative and categorical
    - diagonal: diagonal length in mm
    - height\_left: the height of the left side in mm
    - height\_right: the height of the right side in mm
    - margin\_low: the lower margin in mm
    - margin\_up: the upper margin in mm  
      **All Predictor variables are numerical.**
  + Number of observations (need at least 50): 1500 (we will take a random sample)
* Questions of interest – at least 2
  + Since we know that float type variables take up more memory, will binning some of the variables change the accuracy of the model, hopefully while being more memory efficient?
  + An attribute we chose not to include was *length*. Given diagonal is derived from length, will the model change significantly if we interchange the variables?

\*\*\*