- For MD in training data (to create RF):
  1) Initial Guess : replace missing variable w/ (mode) most common value amongst other data samples w/ same ground truth label (OU) (median for numeric)
  2) Run all data through random forest & generate proximity matrix

$$\begin{array}{c|cccc} \text{samples} & 1 & 2 & 3 & \dots \\ \hline 1 & & & & \\ 2 & & & \boxed{1} & \\ 3 & & \boxed{1} & & \\ \vdots & & & & \end{array}$$

'+1' whenever samples end up in same leaf node
* run thru all trees
* divide prox. vals. by numTrees

  3) Weighted frequencies of variable vals using proximity vals as weights (weighted average for numeric var's)

Repeat steps 1-3 6/7ish times until missing data vals converge (no changes each time)

- For testing/ new data for RF categorization:
  1) Make copies of the data sample w/ the different possible ground truths/ OU's
  2) treat like training data to fill in variable
     i.e. go through the previous steps iteratively
  3) run the full data samples (w/ the different OU's) through the F's, the one that gets labeled correctly the most wins → sample classified

Medium article (Airbnb)
- could drop value - not good
- 1 step precomputation method, normalises features to construct distance metric to fill in missing vals w/ the median of the K nearest neighbors