12/6/17

* Questions:
  + All features should be computed for BEFORE the COMPAS screening date correct? B/c we’re trying to beat the algorithm
    - Ask beau to look over code (specifically, the use of filter and mutate)
  + Regarding calculating current violent offense
    - How should I define violent offense? Should I consider misdemeanors against persons as violent offenses as well?
    - Recommendation: look at exactly how the Arnold Foundation does it, or separate this factor into 2 factors
  + Regarding the Broward case data to choose:
    - Recommendation: disposition (convictions) and events/documents(capias issued, or failure to show up)
  + Coding questions
    - Is there a way to perform multiple methods at the same “stage” of piping? How best to avoid redundancy in code?
    - Why am I getting “parser stack overflow” error
    - When I use “distinct” on statutes why don’t I get anything

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

11/20/17

* Questions:
  + How should I process the charge descriptions? There are 2349 unique charge
  + descriptions (but probably just different in how they’re written down)
    - Idea: do it by charge degree b/c there are only 19 charge degrees?
    - Answer: process by keywords; alternately, do by hand?
  + What does it mean to clean the data? Keep seeing this term
  + How to save a sqlite db to somewhere so I can access it in another r file?
  + Should I do all the subfeatures for # felonies for both 5 yrs and total?
  + Broward county data
    - Since I’ll have all the data (I think) should I still be doing computations based on total # charges or total #convictions? Since compas scores are calculated pretrial, maybe total #charges for the current case and total #convictions for the past?
  + What should I use for my search “criterion” on the Broward API? We really want the convictions and failure to appears for the 11500 people in the dataset, but we probably shouldn’t perform 11500 searches

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

11/3/17

to do for next time

* features table
  + recompute some of compas’ data
* data api broward county

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

10/18/17

* Use charges\_df to generate recidivated/not recidivated
  + I think COMPAS only took charges within 30 days of screenings (for their c\_charge) and only counted non-traffic, 1st charges from a different case for their history of offenses and recidivated (will not count secondary offenses as a recidivation, even though that’s questionable)
    - Thought that will not count secondary offenses in a case, but did so for aaron Epstein—not consistent
  + COMPAS throws out code (0) charges, as traffic charges

To do:

* Check where they got the data from
* Visualizing each person as a case study
* Combine all information into same table—done
* Check COMPAS scores were calculated only once per person
* Begin generating a features table for machine learning

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

10/10/17

* Brainstormed features
  + Suggested looking at features computed in Jiaming’s paper
* Code to take underlying data into something we could use
  + Get all people in compas score database, keep all 3 scores
  + Get most recent offence (the offence the compas score was calculated for)
  + Lines 87-96 in reengineering
* Trying to see what compas did

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

10/11/17