Table Construction

1. Went through the following tables and took yellow highlighted variables from each

Green highlighted variables are those I should consider putting in.

Red highlighted variables are those I definitely have to put in.

* 1. compas\_df
     1. id, person\_id, first, last, compas\_person\_id, compas\_case\_id, compass\_assessment\_id, agency\_text, scale\_set, assessment\_reason, legal\_status, custody\_status, marital\_status, rec\_supervision\_level, rec\_supervision\_level\_text, score\_text, scale\_id, screening\_date, scale\_id, type\_of\_assessment, raw\_score, decile\_score;
     2. Left out compas\_person\_id and compas\_case\_id for now, b/c it doesn’t seem useful for features
        1. \*\*\*\* use compas\_person\_id to perform sanity check (check that each person only has 1 compas case??
  2. people\_df:
     1. name, first, last, sex, race, dob, age, age\_cat, juv\_fel\_count, juv\_misd\_count, juv\_other\_count, compas\_screening\_date, decile\_score, score\_text, violent\_recid, priors\_count, days\_b\_screenign\_arrest, c\_jail\_in, c\_jail\_out, c\_case\_number, c\_days\_from\_compas, c\_arrest\_date, c\_offense\_date, c\_charge\_degree, c\_charge\_desc, is\_recid, num\_r\_cases, r\_case\_number, r\_charge\_degree, r\_days\_from\_arrest, r\_offense\_date, r\_charge\_desc, r\_jail\_in, r\_jail,\_out, is\_violent\_recid, num\_vr\_cases< vr\_case\_number, vr\_charge\_degree, vr\_offense\_date, vr\_charge\_desc
     2. left out the rest b/c we should be able to calculate those ourselves
  3. case\_arrest\_df
     1. id, name, case\_number, arrest\_id, arrest\_date, charge\_degree, days\_since\_compas\_arrest, person\_id,
  4. jailhistory\_df
     1. id, first, last, dob, in\_custody, out\_custody, person\_id
  5. charge\_df
     1. id, case\_type, case\_number, filing\_date, offense\_date, charge\_number, charge\_degree, charge, date\_charge\_filed, filing\_type, filing\_agency, name, days\_since\_compas, statute, person\_id
     2. Did not include filing date b/c they weren’t valid values
     3. Did not include case\_number b/c we should already have that from case\_arrest\_df
        1. \*\*\*Perform check on this later
     4. Did not include
  6. prisonhistory\_df
     1. id, first, last, dob, in\_custody, out\_custody, person\_id
  7. compas\_score (?)
     1. Did not include b/c we’re not using compas’ information

Eventually recompute compas’s stuff? Like whether or not they recidivate, number of violent recidivations, etcetera? What would this fall under, algorithm verification?

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

Suggested features

* Number of felonies in the past (total)
* Number juvenile felonies
* Number felonies in past 5 years
  + Number of 1st degree felonies, number of 2nd degree felonies
  + Number of violent felonies, number of drug felonies
* Number charges in current case, number drug charges in current case
* Most number of charges in past cases
* Number visits to prison
* Number visits to jail and prison
* Duration of total jail and prison time
* Number of drug charges in last 5 years
* Drug severity (????? how would I do this??)
* Number of misdemeanors
* Number of traffic offenses
* Age
* Gender
* Race

Arnold PSA Risk Factors

* Age at current arrest—already covered
* Current violent offence
* Current violent offence & 20 yrs old or younger
* Pending charge at time of the offense
* Prior misdemeanor conviction
* Prior felony conviction
* Prior conviction (misdemeanor or felony)
* Prior violent conviction
* Prior failure to appear pretrial in the past two years
* Prior failure to appear pretrial older than two years
* Prior sentence to incarceration

\*look at Jiaming’s paper, Arnold PSA, federal crimes classifier, research criminal psychology