I. Source Data

* compas.db (sqlite)
  + casearrest\_df
  + charge\_df
  + compass\_df
  + jailhistory\_df
  + prisonhistory\_df
  + people\_df
* fta\_prob (computed from eventsdescrp\_df)
* convictions (computed from corrected\_dispositions)

II. Features Table

*Static identifiers*

* person\_id
* first
* last
* p\_sex
* p\_race
* p\_dob
* p\_age

*COMPAS information*

* compas\_person\_id
* compas\_case\_id
* compas\_assessment\_id
* screening\_date
* p\_marital\_status
* agency\_text
* custody\_status
* p\_violence\_score
* p\_violence\_raw
* p\_violence\_decile
* p\_vio\_noage
* p\_recid\_score
* p\_recid\_raw
* p\_recid\_decile
* p\_fail\_to\_appear\_score
* p\_fail\_to\_appear\_raw
* p\_fail\_to\_appear\_decile

***Features***

* p\_felony\_count\_person
* p\_yrs5\_felcount\_person
* p\_prison\_visits
* p\_prison\_time
* p\_jail\_visits
* p\_jail\_time
* p\_numcases
* p\_misdem\_count\_person
* p\_misdem\_count\_5yrs
* p\_drug\_count\_person
* p\_drug\_count\_5yrs
* p\_traffic\_count\_person
* p\_yrs5\_trafcount\_person

*History of Violence (COMPAS subscale) features*

* p\_current\_age
* p\_age\_first\_arrest
* p\_age\_first\_offense
* p\_juv\_fel\_count
* p\_felprop\_violarrest
* p\_felassault\_arrest
* p\_misdemassault\_arrest
* p\_murder\_arrest
* p\_famviol\_arrest
* p\_sex\_arrest
* p\_weapons\_arrest

*Arnold PSA Features*

* p\_current\_violent\_offense
* p\_current\_violent\_twenty
* p\_prior\_incarceration
* cq\_prior\_conviction\_F
* cq\_prior\_conviction\_M
* prior\_conviction\_T
* cq\_violent\_convictions
* cq\_total\_convictions
* cq\_fail\_appear\_two\_yr
* cq\_fail\_appear\_two\_plus
* cq\_prob\_violation
* cq\_pending\_charge

***Measures of Recidivism***

* propub\_recid
* propub\_violent\_recid
* traffic\_recid
* arnold\_fta
* arnold\_nca
* arnold\_nvca

III. Features Definitions

*Static identifiers*

* person\_id
* first
* last
* p\_sex
* p\_race
* p\_dob
  + Date of birth
* p\_age
  + Age of individual around the time the data was pulled, or maybe the age of the person at his most recent offense around time data was pulled (uncertain)

*COMPAS information*

* compas\_person\_id
* compas\_case\_id
* compas\_assessment\_id
* screening\_date
* p\_marital\_status
* agency\_text
* custody\_status
* p\_violence\_score
* p\_violence\_raw
* p\_violence\_decile
* p\_vio\_noage
  + Because age is a large component of the COMPAS violence score, we used a polynomial to model the relationship between *current* age (age at screening) and the compas score, and then corrected each raw violence score by the corresponding age.
  + age\_poly\_violence=function(x){0.00000126145944912785\*x^4 - 0.00023448446315780700\*x^3 + 0.01660174356697860000\*x^2 - 0.56651100033213000000\*x + 4.13182477973007000000}
* p\_recid\_score
* p\_recid\_raw
* p\_recid\_decile
* p\_fail\_to\_appear\_score
* p\_fail\_to\_appear\_raw
* p\_fail\_to\_appear\_decile

***Features***

* p\_felony\_count\_person
* p\_yrs5\_felcount\_person
* p\_prison\_visits
* p\_prison\_time
* p\_jail\_visits
* p\_jail\_time
* p\_numcases
  + Total number of cases for each person
* p\_misdem\_count\_person
* p\_misdem\_count\_5yrs
* p\_drug\_count\_person
* p\_drug\_count\_5yrs
* p\_traffic\_count\_person
* p\_yrs5\_trafcount\_person

*History of Violence (COMPAS subscale) features*

* p\_current\_age
  + Age at screening date; p\_current\_age= screening\_date – date of birth
* p\_age\_first\_arrest
* p\_age\_first\_offense
* p\_juv\_fel\_count
* p\_felprop\_violarrest
* p\_felassault\_arrest
* p\_misdemassault\_arrest
* p\_murder\_arrest
* p\_famviol\_arrest
* p\_sex\_arrest
* p\_weapons\_arrest

*Arnold PSA Features*

* p\_current\_violent\_offense
* p\_current\_violent\_twenty
* p\_prior\_incarceration
* cq\_prior\_conviction\_F
* cq\_prior\_conviction\_M
* prior\_conviction\_T
* cq\_violent\_convictions
* cq\_total\_convictions
* cq\_fail\_appear\_two\_yr
* cq\_fail\_appear\_two\_plus
* cq\_prob\_violation
* cq\_pending\_charge

***Measures of Recidivism***

* propub\_recid
* propub\_violent\_recid
* traffic\_recid
* arnold\_fta
* arnold\_nca
* arnold\_nvca