

Dasha Asienga
02/23/2024

SQLite Script to Subset COMPAS Data

This document contains the SQLite script used to subset the COMPAS data of interest from the `compas.db` database. The `compas.db` database is available publicly on the [ProPublica Data Store](#) and can be downloaded from [Github](#).

Instructions:

Download the `compas.db` database file in this folder.

Download *'DB Browser for SQLite'* or any similar software application for managing database software.

Open the `compas.db` database in the user interface.

Run the following code in the *'Execute SQL'* pane.

Results:

The database contains 7 tables, although the *summary* table is empty. The code below returns 12160 rows of participants who were assessed in the COMPAS tool for risk of recidivating.

There are 29 variables of interest returned, though many more could be selected if interested. Further data wrangling and data analysis are conducted in *R*.

Code:

```
select

people.id,
compas.compas_person_id,
people.name,
people.first,
people.last,
people.sex,
people.race,
people.age,
people.age_cat,
compas.marital_status,
compas.custody_status,
people.juv_fel_count,
people.juv_misd_count,
people.juv_other_count,
```

```

people.priors_count,
people.days_b_screening_arrest,
people.c_days_from_compas,
people.c_charge_degree,
people.c_charge_desc,
compas.type_of_assessment,
compas.raw_score,
people.decile_score,
compas.score_text,
people.is_violent_recid,
people.num_vr_cases,
people.is_recid,
people.num_r_cases,
round(jail_agg.days_in_jail) as days_in_jail,
round(prison_agg.days_in_prison) as days_in_prison

from compas

inner join people
on compas.person_id = people.id
and compas.first = people.first
and compas.last = people.last
and compas.decile_score = people.decile_score

left join (

select

jailhistory.person_id,
jailhistory.first,
jailhistory.last,
jailhistory.out_custody,
jailhistory.in_custody,
sum(distinct(julianday(jailhistory.out_custody) -
julianday(jailhistory.in_custody))) as days_in_jail

```

```

from jailhistory

inner join compas
on compas.person_id = jailhistory.person_id
and compas.first = jailhistory.first
and compas.last = jailhistory.last
and jailhistory.in_custody <= compas.screening_date

group by 1,2,3

) as jail_agg
on jail_agg.person_id = people.id
and jail_agg.first = people.first
and jail_agg.last = people.last

left join (

select

prisonhistory.person_id,
prisonhistory.first,
prisonhistory.last,
prisonhistory.out_custody,
prisonhistory.in_custody,
sum(distinct(julianday(prisonhistory.out_custody) -
julianday(prisonhistory.in_custody))) as days_in_prison

from prisonhistory

inner join compas
on compas.person_id = prisonhistory.person_id
and compas.first = prisonhistory.first
and compas.last = prisonhistory.last
and prisonhistory.in_custody <= compas.screening_date

```

```
group by 1,2,3

) as prison_agg
on prison_agg.person_id = people.id
and prison_agg.first = people.first
and prison_agg.last = people.last

where type_of_assessment = 'Risk of Recidivism'
```