

Need to drop duplicates here first
(193,246 original records and
177,667 records after dropping
duplicates). "user_guid" is
technically foreign key but is
always NULL in this table.

No need to drop
duplicates here
(35,050 records)

Need to drop duplicates here
first (2,460,320 original records
and 2,458,699 records after
dropping duplicates)

complete_tests

<u>created_at</u>
<u>dog_guid</u> (FK)
<u>test_name</u>
updated_at
subcategory_name
user_guid (FK)

dogs

<u>dog_guid</u>
gender
birthday
breed
weight
dog_fixed
dna_tested
created_at
updated_at
dimension
exclude
breed_type
breed_group
user_guid (FK)
total_tests_completed
mean_iti_days
mean_iti_minutes
median_iti_days
median_iti_minutes
time_diff_between_first_and_last_game_days
time_diff_between_first_and_last_game_minutes

exam_answers

<u>script_detail_id</u>
<u>start_time</u>
<u>end_time</u>
<u>loop_number</u>
<u>dog_guid</u> (FK)
subcategory_name
test_name
step_type

Need to drop a few duplicates
here first (30,716 original
records and 30,711 records after
dropping duplicates)

users

<u>user_guid</u>
sign_in_count
created_at
updated_at
max_dogs
membership_id
subscribed
exclude
free_start_user
last_active_at
membership_type
city
state
zip
country
utc_correction

reviews

<u>created_at</u>
<u>user_guid</u> (FK)
<u>test_name</u>
rating
updated_at
subcategory_name
dog_guid (FK)

site_activities

<u>activity_type</u>
<u>description</u>
<u>created_at</u>
<u>user_guid</u> (FK)
membership_id
category_id
script_id
updated_at
script_detail_id
test_name
dog_guid (FK)

After dropping duplicate records,
user_guid can almost uniquely identify
each row except for 13 records that have
the same user_guid but different values of
utc_correction. So, we drop the 13
records where utc_correction="#N/A".
(34,917 original records, 33,206 records
after dropping duplicates, and 33,193

Need to drop duplicates here
first (1,353,130 original records
and 1,337,893 records after
dropping duplicates)