

Dataset Description

Header	Header Description
unit_id	a unique id for user
golden	whether the user was included in the gold standard for the model; TRUE or FALSE
unit_state	state of the observation; one of finalized (for contributor-judged) or golden (for gold standard observations)
trusted_judgments	number of trusted judgments (int); always 3 for non-golden, and what may be a unique id for gold standard observations
last_judgment_at	date and time of last contributor judgment; blank for gold standard observations
gender	one of male, female, or brand (for non-human profiles)
gender:confidence	a float representing confidence in the provided gender
profile_yn	"no" here seems to mean that the profile was meant to be part of the dataset but was not available when contributors went to judge it
profile_yn:confidence	confidence in the existence/non-existence of the profile
created	date and time when the profile was created
description	the user's profile description
fav_number	number of tweets the user has favorited
gender_gold	if the profile is golden, what is the gender?
link_color	the link color on the profile, as a hex value
name	the user's name
profile_yn_gold	whether the profile y/n value is golden
profileimage	a link to the profile image
retweet_count	number of times the user has retweeted (or possibly, been retweeted)
sidebar_color	color of the profile sidebar, as a hex value
text	text of a random one of the user's tweets
tweet_coord	if the user has location turned on, the coordinates as a string with the format "[latitude, longitude]"
tweet_count	number of tweets that the user has posted
tweet_created	when the random tweet (in the text column) was created
tweet_id	the tweet id of the random tweet
tweet_location	location of the tweet; seems to not be particularly normalized
user_timezone	the timezone of the user