NAME: San Francisco OKCupid Users

TYPE: Volunteered sample

SIZE: 59946 cases (users active on 2012/06/30 who had active accounts on 2012/06/26), 32 variables

ARTICLE TITLE: OkCupid Profile Data for Introductory Statistics and Data Science Courses

DESCRIPTIVE ABSTRACT: This dataset was created with the use of a python script that pulled the data from public profiles on www.okcupid.com. It has an n=59946, which includes people within a 25 mile radius of San Francisco, who were online in the last year (06/30/2011), with at least one profile picture.

SOURCES: Data scraped from public profiles on www.okcupid.com on 06/30/2012. Permission to use this data was obtained from OkCupid president and co-founder Christian Rudder under the condition that the dataset remains public.

VARIABLE DESCRIPTIONS:

Missing data is blank. Some variables have two factors, which are denoted in this codebook by a semicolon (e.g. "graduated from; two-year college) though they are found without punctuation in the dataset. Details for specific variables are found in parenthesis.

age – in years

body\_type- rather not say, thin, overweight, skinny, average, fit, athletic, jacked, a little extra, curvy, full figured, used up

diet- mostly/strictly; anything, vegetarian, vegan, kosher, halal, other

drinks- very often, often, socially, rarely, desperately, not at all

drugs- never, sometimes, often

education- graduated from, working on, dropped out of; high school, two-year college, university, masters program, law school, med school, Ph.D program, space camp

ethnicity- Asian, middle eastern, black, native American, indian, pacific islander, Hispanic/latin, white, other

height- inches

income- (US $, -1 means rather not say) -1, 20000, 30000, 40000, 50000, 60000 70000, 80000, 100000, 150000, 250000, 500000, 1000000,

job- student, art/music/writing, banking/finance, administration, technology, construction, education, entertainment/media, management, hospitality, law, medicine, military, politics/government, sales/marketing, science/engineering, transportation, unemployed, other, rather not say, retire

offspring- has a kid, has kids, doesnt have a kid, doesn't want kids; ,and/,but might want them, wants them, doesnt want any, doesnt want more

orientation- straight, gay, bisexual

pets- has dogs, likes dogs, dislikes dogs; and has cats, likes cats, dislikes cats

religion- agnosticism, atheism, Christianity, Judaism, Catholicism, Islam, Hinduism, Buddhism, Other; and very serious about it, and somewhat serious about it, but not too serious about it, and laughing about it

sex- m, f

sign- aquarius, pices, aries, Taurus, Gemini, cancer, leo, virgo, libra, scorpio, saggitarius, Capricorn; but it doesn’t matter, and it matters a lot, and it’s fun to think about

smokes- yes, sometimes, when drinking, trying to quit, no

speaks- English (fluently, okay, poorly). Afrikaans, Albanian, Arabic, Armenian, Basque, Belarusan, Bengali, Breton, Bulgarian, Catalan, Cebuano, Chechen, Chinese, C++, Croatian, Czech, Danish, Dutch, Esperanto, Estonian, Farsi, Finnish, French, Frisian, Georgian, German, Greek, Gujarati, Ancient Greek, Hawaiian, Hebrew, Hindi, Hungarian, Icelandic, Ilongo, Indonesian, Irish, Italian, Japanese, Khmer, Korean, Latin, Latvian, LISP, Lithuanian, Malay, Maori, Mongolian, Norwegian, Occitan, Other, Persian, Polish, Portuguese, Romanian, Rotuman, Russian, Sanskrit, Sardinian, Serbian, Sign Language, Slovak, Slovenian, Spanish, Swahili, Swedish, Tagalog, Tamil, Thai, Tibetan, Turkish, Ukranian, Urdu, Vietnamese, Welsh, Yiddish (fluently, okay, poorly)

status- single, seeing someone, married, in an open relationship

essay0- My self summary

essay1- What I’m doing with my life

essay2- I’m really good at

essay3- The first thing people usually notice about me

essay4- Favorite books, movies, show, music, and food

essay5- The six things I could never do without

essay6- I spend a lot of time thinking about

essay7- On a typical Friday night I am

essay8- The most private thing I am willing to admit

essay9- You should message me if...

SPECIAL NOTES: All essay questions are fill-in the blank, answers are not summarized here.

SUBMITTED BY:

Albert Y. Kim and Adriana Escobedo-Land

Reed College

3203 SE Woodstock Blvd

Portland, OR 97202

albert.kim@reed.edu

escobad@reed.edu