

Project PEACH is a large scale open source community driven Data Science project by the Computer Science Department at University College London. The goal of Project PEACH is to provide a data science tool for medical professionals and researchers that can aid them in diagnostic and analytics processes through the use of big data, machine learning and data visualisation. <https://code-4-health.org/peach>

Project Supervisors: Dr. Dean Mohamedally, Dr. Atia Rafiq, Marios Constantinides, George Spithourakis

Team Members: Nikolaos Georgiadis, Narmin Huseynli, Efthymia Kazakou, Johann Kristthorsson, Mahalakshmi Sabapathi, Siri Vinay, Stephen Whelan

Project Contact: Stephen Whelan (Lead Architect, Template 1), stephen.whelan.15@ucl.ac.uk

The following document is a description of the file structure for the User Profile data set.

PEACH_userProfiles_All_Final.csv - (comma delimited)

This file contains the profiles for every user on Project PEACH. These values are static and do not change over the time period of the data sets.

File Structure

Field	Values	Description
UserId	Integer	This field is a user's unique Id. Each user maintains their unique Id throughout all data sets in Project PEACH.
Nationality	String	Name of country the user was born in.
Gender	Integer	0 = male 1 = female
Age	Integer	Age of user (in years).
Latitude	Decimal	The Latitude value of the user's Current_City_of_Domicile
Longitude	Decimal	The Longitude value of the user's Current_City_of_Domicile
Current_City_of_Domicile	String	The city where the user is currently living. If the user does not live in a major city, the <i>City</i> value defaults to "rural".

Current_Smoker	String	yes = current smoker no = not a current smoker
Historical_Smoker	Integer	This value is measured in “pack years.” (i.e. 2 packs per day for 2 years = 4 pack years)
Alcohol_Consumption	Integer	Average total units of alcohol consumed per week
Alcoholic	String	yes = user is alcoholic no = user is not alcoholic This value is determined based on medical standards of units of alcohol consumed per week. The threshold is over 27 units for males and over 20 units for females.
Skin_Color	Decimal	Value between [0, 1]. 0 is lightest, 1 is darkest.
Outgoing_Social_Life	String	yes = user is outgoing no = user lacks a social life
Frequent_Traveler	Integer	Value between [1, 10]. 1 is the user flies very little, 10 is the user is a frequent flyer.
City_of_Domicile_Type	Integer	0 = Urban town 1 = Suburban town 2 = Rural town
Outdoors_Social_Life	String	yes = user spends time outdoors (i.e. hiking, skiing, exercise etc.) no = user spends most of their time indoors
Income_Bracket	Integer	0 = Poor 1 = Middle Class 2 = Wealthy
Education	Integer	0 = None 1 = Primary 2 = Secondary 3 = Bachelor’s Degree 4 = PhD
Hours_of_Sleep_per_Night	Decimal	Average hours of sleep the user gets per night
Type_of_Work	String	Values are {active_indoors, active_outdoors, sedentary}

Pet_Owner	Integer	0 = Do not own a pet 1 = Own a pet
-----------	---------	---------------------------------------