**Obesity, raw and synthetic dataset – data descriptor**

**Source:**

Fabio Mendoza Palechor, Email: fmendoza1 **'@'** cuc.edu.co, Celphone: +573182929611  
Alexis de la Hoz Manotas, Email: akdelahoz **'@'** gmail.com, Celphone: +573017756983

**Data Set Information:**

This dataset include data for the estimation of obesity levels in individuals from the countries of Mexico, Peru and Colombia, based on their eating habits and physical condition. The data contains 17 attributes and 2111 records, the records are labeled with the class variable NObesity (Obesity Level), that allows classification of the data using the values of Insufficient Weight, Normal Weight, Overweight Level I, Overweight Level II, Obesity Type I, Obesity Type II and Obesity Type III. 77% of the data was generated synthetically using the Weka tool and the SMOTE filter, 23% of the data was collected directly from users through a web platform.

**Attribute Information:**

Read the article ([[Web Link]](https://doi.org/10.1016/j.dib.2019.104344)) to see the description of the attributes.

**Relevant Papers:**

[1]Palechor, F. M., & de la Hoz Manotas, A. (2019). Dataset for estimation of obesity levels based on eating habits and physical condition in individuals from Colombia, Peru and Mexico. Data in Brief, 104344.  
[2]De-La-Hoz-Correa, E., Mendoza Palechor, F., De-La-Hoz-Manotas, A., Morales Ortega, R., & SÃ¡nchez HernÃ¡ndez, A. B. (2019). Obesity level estimation software based on decision trees.