# Body Mass Index and Number of Steps Analysis from Activity Tracker Data

**Renata Rego**
**21 October 2017**

**Abstract.** The general goal of this project is to analyze steps and body weight data from users of activity trackers, and to provide coaching insights for users to reach their fitness goals. The database used in this project was kindly provided by Medisana® and consists on user data stored on their Vitadock Online Platform.

## 1. Introduction

Obesity and sedentary lifestyle are known to belong together. The question we want to investigate in this project is how daily steps, measured with a simple, wide accessible equipment, are related to the increase or decrease of the body mass index (vmi), and wether it is possible to provide users insights of how increasing their daily steps could help with decreasing their bmi (i.e, loosing weight). The users population considered in this project doesn't have any restriction of sex, age, height, etc. Statistics about the available data, such as age range, amount of female and male users, etc. will be presented in the Exploratory Data Analysis.

## 2. Methods

### *2.1. Data Acquisition*

The Vitadock database schema is available at the Vitadock API repository . The data was kindly provided by Medisana® for this project, and due to privacy reasons it cannot be publicly shared. The tables and fields of interest for this project are:

- user settings: user_id, sex, height, birthday
- targetscale: user_id, measurement_date, body_weight, bmi
- targetscale settings: user_id, measure_unit
- tracker stats: user_id, measurement_date, steps, calories, distance

The data of interest were initially read from a MySQL database and saved as json files (one file per table listed above). We observed that some users have several entries in the user_settings table, i.e., there are multiple entries of user_settings with the same

value for the field user_id, which was expected to be unique. The explanation for this is that the database keeps a history of all changes of each particular user settings. In this project, we took the most recently updated setting for each user with multiple entries. This is reasonable because users usually fill in few a required information when registering their account, and as they get interested in the device they update their accounts, adding new data, or updating values inserted automatically by default.

Once we read the data from the relational database into json files, we were ready to start the next steps of the project: data exploration and data preparation.

### *2.2. Data Exploration and Data Preparation*

The main data exploration tasks performed in this project were:

- Getting an overview of the data. For example, number of entries available, number of null values, possible values for the attributes (domain);
- Analyzing the statistical profile of data: ex. max, min, mean values, histograms, boxplots, etc.
- Analysing the correlations between bmi and average number of daily steps.

During the statistical analysis we could also identify outliers and noisy values, and perform the necessary operations to clean the data.

Data Preparation activities includes the data cleaning mentioned above and a couple of other transformations, as for example:

- converting birthday to age;
- filtering weight unit measures: for the sake of simplicity and the considering that the great majority of entries are in kilos, we selected only weights given in kilos;
- recomputing bmi, in order to fill missing values or make sure it is coherent with the user's weight and height.

  The findings of the exploratory data analysis are presented in the links below:

- Users: Visualize statistical profile of user data, filter noisy data.
- Weights: Visualize statistical profile of weight data, filter noisy data, filter by metric (kilos), handle multiple measure at the same timestamp, compute bmi.
- Activities: Visualize statistical profile of activity data, filter noisy steps counts based on threshold, steps *vs* distance ratio, and steps *vs* calories ratio.
- Bmi, Steps correlation: Visualize correlation between bmi and number of steps considering all users, and inidividual users.