Тестовое задание на вакансию   
Data Analyst в Welltory

## Исходные данные

Скачайте данные по ссылке <https://drive.google.com/file/d/14FsqWotcSHj6vP-2eWpZRU5qtO0w9Q5H/view?usp=sharing> Они представляют собой .csv файл с 4 столбцами:

* time\_start\_local — начало временного интервала
* time\_end\_local — конец временного интервала
* time\_offset — сдвиг часового пояса в секундах относительно UTC
* steps — количество шагов, которые прошел пользователь во время этого временного интервала

В файле примерно месячные данные одного пользователя Welltory (мы получили его прямое разрешение на публикацию данных). Мы верим, что их будет достаточно для приемлемого решения задачи.

## Общая задача

Придумать алгоритм, который бы находил в данных такого типа прогулки. Понятие «прогулка» формально не определено, но мы представляем себе, что это примерно постоянная ходьба как минимум в течение 5 минут. Алгоритм нужно реализовать на языке python (версии 3.10 или более ранней).

## Ожидаемый результат

Архив с названием <ваш ник в телеграме или фамилия>.zip со следующими 4 файлами:

* requirements.txt — перечисление зависимостей
* walkings.py — файл с функцией find\_walkings(data: pandas.DataFrame) -> dict, которая возвращает словарь вида   
  {  
   “<день в формате dddd-mm-dd>”: [ # список прогулок в этот день  
   {  
   “start”: “<начало прогулки в формате dddd-mm-dd HH:MM:SS>”,  
   “end”: “<конец прогулки в формате dddd-mm-dd HH:MM:SS>”,  
   “steps”: <количество шагов во время прогулки типа int>  
   },  
   …  
   ]  
  }
* walkings.json — файл с результатами запуска функции find\_walkings на предложенных данных
* walkings.ipynb — jupyter файл с EDA и описанием того, как вы пришли к решению

## Что мы будем оценивать

* Формальную сторону решения: функция find\_walkings должна выдавать адекватный ответ на предложенных данных
* Полноту исследования данных
* Качество оформления ваших рассуждений

**Время выполнения**

Это открытая задачка в том смысле, что она может занять произвольное время, в зависимости от того, как глубоко ее прорабатывать. Но мы считаем, что у человека с богатым опытом (коммерческим или учебным) использования python+pandas достаточно неплохое решение этого задания может занять 2–4 часа. Ну или мы хотим найти такого человека, который может находить приемлемые решения таких задач за подобное время.

**Срок выполнения**

Мы понимаем, что вы можете работать, быть в отъезде, заниматься личными вопросами и не иметь достаточно свободного времени, чтобы взяться за тестовое прямо сейчас. Будет здорово, если вы пришлете решение в течение нескольких дней после получения тестового задания. Пожалуйста, дайте знать, если вам понадобится больше времени.