

*Данные для задачи «Предсказание социально-демографических характеристик пользователя» состоят из следующих таблиц:*

**1. all\_events.csv**— Данные с историей просмотров пользователей

Описание признаков:

*viewer\_uid*: id пользователя

*rutube\_video\_id*: id видео

*event\_timestamp*: Дата  
события

*region*: Регион

*ua\_device\_type*: Тип устройства пользователя

*ua\_client\_type*: Браузер/приложение

*ua\_os*: Название операционной системы пользователя

*ua\_client\_name*: Название браузера пользователя

*total\_watchtime*: Время просмотра в секундах

Прим.: В колонке *event\_timestamp* указано московское время (GMT +3). Используйте данные из *region*, чтобы привести время к региональному

**2. train\_events.csv** — Данные с историей просмотров пользователей с известными целевыми переменными

Описание признаков:

*event\_timestamp*: Дата  
события

*viewer\_uid*: id пользователя

*rutube\_video\_id*: id видео

*region*: Регион

*ua\_device\_type*: Тип устройства пользователя

*ua\_client\_type*: Браузер/приложение

*ua\_os*: Название операционной системы пользователя

*ua\_client\_name*: Название браузера пользователя

*total\_watchtime*: Время просмотра в секундах

Прим.: В колонке *event\_timestamp* указано московское время (GMT +3). Используйте данные из *region*, чтобы привести время к региональному

**3. video\_info.csv** — Данные с информацией о видео

Описание признаков:

*rutube\_video\_id*: id видео

*title*: название видео

*duration*: длительность видео в миллисекундах

*author\_id*: id автора видео

*category*: категория видео

**4. train\_targets.csv** — Целевые переменные пользователей из таблицы train\_events

Описание признаков:

*viewer\_uid*: id пользователя

*age*: возраст пользователя

*sex*: Пол пользователя

*age\_class*: возрастная категория пользователя

Прим.: Возрастные категории разделены следующим образом:

0 – (9, 20] (лет)

1 – (20, 30] (лет)

2 – (30, 40] (лет)

3 – (40, 60] (лет)

Вы можете обучать алгоритм в задаче мультиклассовой классификации возрастной группы, но также и в задаче регрессии, самостоятельно определив категорию пользователя по предсказанному количеству лет.

## 5. **sample\_submission.csv** — Пример таблицы с предсказаниями модели участников

Описание признаков:

*viewer\_uid*: id пользователя из тестовой выборки

*sex*: Предсказанный пол пользователя

*age\_class*: Предсказанная возрастная категория пользователя