**Задание:**

**Преподаватель:**Константин Башевой

**Домашнее задание**

**Задание 1**  
Для датафрейма log из материалов занятия создайте столбец source\_type по правилам:

* если источник traffic\_source равен Yandex или Google, то в source\_type ставится organic;
* для источников paid и email из России ставим ad;
* для источников paid и email не из России ставим other;
* все остальные варианты берём из traffic\_source без изменений.

**Задание 2**  
В файле URLs.txt содержатся URL страниц новостного сайта. Вам нужно отфильтровать его по адресам страниц с текстами новостей. Известно, что шаблон страницы новостей имеет внутри URL конструкцию: /, затем 8 цифр, затем дефис. Выполните действия:

1. Прочитайте содержимое файла с датафрейм.
2. Отфильтруйте страницы с текстом новостей, используя метод str.contains и регулярное выражение в соответствие с заданным шаблоном.

**Задание 3**  
Используйте файл с оценками фильмов ml-latest-small/ratings.csv. Посчитайте среднее время жизни пользователей, которые выставили более 100 оценок. Под временем жизни понимается разница между максимальным и минимальным значениями столбца timestamp для данного значения userId.

**Задание 4**  
Дана статистика услуг перевозок клиентов компании по типам (см. файл “Python\_13\_join.ipynb” в разделе «Материалы для лекции “Продвинутый pandas”» ---- Ноутбуки к лекции «Продвинутый pandas»).  
Нужно сформировать две таблицы:

* таблицу с тремя типами выручки для каждого client\_id без указания адреса клиента;
* аналогичную таблицу по типам выручки с указанием адреса клиента.

*Обратите внимание, что в процессе объединения таблиц данные не должны теряться*

Решение:

<https://colab.research.google.com/drive/1LrmhZwNpcnRD8wN-XcYagKQiaa8wukbO?usp=sharing>

Изображение выглядит как текст

Автоматически созданное описание