**\*\*Задание 1\*\***

​

1. 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 значением столбца timestamp для данного значения userId.

​

**\*\*Задание 2\*\***

​

2. Дана статистика услуг перевозок клиентов компании по типам (см. файл с кодом занятия). Необходимо 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.

​

​

**\*\*Задание 3\*\***

​

3. В задаче сквозной аналитики вам предоставили данные по местоположению пользователей. Т. е. для каждого user\_id известна последовательность координат (широта/долгота), когда они требовались приложению для полноценной работы. Как бы вы добавили эти сведения в таблицу визитов и покупок? Для составления ответа можно использовать вопросы:

- У каждого пользователя известен набор координат. А для связывания с визитом или фактом покупки скорее всего потребуется одно-два числа. Как их получить?

- Наборы координат одного и того же пользователя могут быть значительно удалены друг от друга. Как это отразится на вопросе расчетах пункта 1?

- Какие дополнительные признаки можно получить из координат? Ведь это просто числа, которые сами по себе мало что дают.

​