

Задание

Задание состоит из 3-х последовательных этапов по получению и обработке данных:

1. Поиск и изучение данных;
2. Подготовка данных (Python + SQL);
3. Расчет метрик (Python + SQL).

Этап 1. Поиск и изучение данных

Найдите интересный набор данных.

Ожидаемый результат:

1. Ссылка на скачивание набора данных или любые другие реквизиты расположения набора данных;
2. Файл с описанием набора данных (той его части, которая будет использоваться в дальнейшем при выполнении остальной части задания).

Варианты выполнения могут быть следующие:

1. (Простой) набор данных (той его части, которая будет использоваться в дальнейшем при выполнении остальной части задания) суммарно меньше 10 млн строк;
2. (Сложнее) больше 10 млн строк.

Этап 2. Подготовка данных

Подготовьте исходный набор данных (сделайте набор данных удобным).

Ожидаемый результат:

1. Исходный скрипт/код/обработка
2. Ссылка на скачивание итогового набора данных;
3. Описание полей итогового набора.

Варианты выполнения:

1. Python + SQL
2. ...

Этап 3. Расчет метрик

Предложить и рассчитать интересные метрики.

Ожидаемый результат:

1. Исходный скрипт/код/обработка;
2. Ссылка на скачивание наборов данных с посчитанными метриками;
3. Описание метрик

Варианты выполнения:

1. Python + SQL
2. ...