

# Тестовое задание

## Задача 1

Выгрузить из базы данных с использованием SQL-запроса информацию по следующим пользователям:

- Ученики с годовых курсов ЕГЭ и ОГЭ

Необходимые поля в итоговой выгрузке:

- ID курса
- Название курса
- Предмет
- Тип предмета
- Тип курса
- Дата старта курса
- ID ученика
- Фамилия ученика
- Город ученика
- Ученик не отчислен с курса
- Дата открытия курса ученику
- Сколько полных месяцев курса открыто у ученика
- Число сданных ДЗ ученика на курсе

Файлы, необходимые для решения задачи:

1. db\_schema.jpg – схема БД
2. script\_db.py – python-скрипт для имитации запроса к БД

Результат, который нужно прислать по задаче:

- SQL-скрипт в формате файла .sql. В коде скрипта оставляйте комментарии с пояснениями своих действий

## Задача 2

С использованием python-скриптов обработать датасет, подготовленный в Задаче 1:

1. Проверить датасет на: дубликаты, пропуски, типы данных, аномальные значения. Произвести предобработку датасета, если это необходимо
2. Поделить всех учеников на волны

Критерии волн:

- 0 волна – присоединились на курс до даты старта курса включительно
- 1 волна - присоединились на курс в течение 1 недели после даты старта курса включительно
- 2 волна - присоединились на курс в течение от 1 до 2 недель включительно после даты старта курса
- 3 волна - присоединились на курс в течение от 2 до 3 недель включительно после даты старта курса
- 4 волна - присоединились на курс в течение от 3 до 4 недель включительно после даты старта курса
- 5 волна - присоединились на курс от 4 недель после даты старта курса

Результат, который нужно прислать по задаче:

- Jupyter Notebook в формате файла .ipynb. В тетрадке оставляйте комментарии с пояснениями своих действий

## Задача 3

Построить дашборд в Yandex DataLens с использованием датасета, подготовленного в Задаче 2.

Цели дашборда:

1. Отслеживать какие курсы хуже всего /лучше всего продлеваются учениками
2. Отбирать учеников, которые не продлили курс, по дополнительным заданным критериям, чтобы связываться с данными учениками для повышения процента продлений

Для реализации данных целей на дашборде должна быть возможность:

1. Сравнивать курсы по двум метрикам: числу и проценту продлений заданного месяца. Пример: на заданном курсе всего 1000 учеников, из них продлили 2-й месяц курса 800 учеников. Тогда на данном курсе: число продлений 2-го месяца равно 800, процент продлений 2-го месяца равен 80%
2. Получать полную информацию по ученикам (курс ученика, волна ученика на курсе, число сданных ДЗ ученика на курсе, город ученика и т.д.), которые подходят под данное условие:

а. Не отчислены с курса

Также должна быть возможность среди данных учеников отобрать учеников:

б. С заданных курсов

с. С заданных волн

д. С заданных городов

е. Которые не продлили заданный месяц курса

ф. У которых число сданных ДЗ на курсе меньше заданного числа

Результат, который нужно прислать по задаче:

- Ссылка на дашборд в DataLens с публичным доступом
- Код расчетных показателей дашборда с комментариями по логике расчета

## Общая информация

- Решения всех заданий необходимо прислать в едином документе, прикрепив в нем ссылки на результаты
- Оставляйте комментарии и пояснения по задачам - так мы сможем понять, что и как вы делали