

Домашнее задание. Практическая работа

Дисциплина «ETL-процессы».

Тема «Основы трансформации данных».

Форма проверки

Самостоятельное задание с разбором на вебинаре.

Совет: выполняйте домашнее задание сразу, как только изучите тему.

Имя преподавателя

Артём Озерков.

Время выполнения

2 часа.

Цель задания

Отработать навыки преобразования данных.

Инструменты для выполнения ДЗ

Apache Airflow и датасет.

Правила приёма работы

Прикрепите в LMS ссылку на GitHub с выполненным заданием и скриншот из Airflow, чтобы был виден успешно завершившийся процесс.

Важно:

- убедитесь, что по ссылке есть доступ;
- название должно содержать фамилию и имя студента, номер и название ДЗ.

Чек-лист самопроверки

Задание считается выполненным, если:

- прикреплена ссылка с выполненным заданием;
- по ссылке содержится выполненное задание;
- доступ к материалам открыт.

Дедлайн

6 дней после открытия задания на платформе.

Описание задания

Выполните преобразование данных. Для этого используйте [датасет](#).

Что необходимо сделать:

- вычислите 5 самых жарких и самых холодных дней за год;
- отфильтруйте `out/in = in`;
- поле `noted_date` переведите в формат 'yyyy-MM-dd' с типом данных `date`;
- очистите температуру по 5-му и 95-му процентилю.