**Домашнее задание. Практическая работа**

**Дисциплина**

ETL-процессы.

**Тема**

Тема 14. Управление данными и их качество

**Форма проверки**

Домашнее задание с самопроверкой.

Совет: выполняйте его сразу, как только изучите тему.

**Имя преподавателя**

Артём Озерков.

**Время выполнения**

2 часа.

**Цель задания**

Научиться использовать метрику PSI для контроля качества данных

**Инструменты для выполнения ДЗ**

Jupyter Notebook

**Правила приёма работы**

Сделайте скриншот результата проверки из Jupyter Notebook. Прикрепите в LMS файл или ссылку на GitHub со скриншотом и ipynb-файлом.

Важно:

* убедитесь, что доступ к материалам открыт;
* название файла должно содержать фамилию и имя студента, номер и название ДЗ.

**Чек-лист самопроверки**

**Задание считается выполненным:**

* если прикреплены файл с выполненным заданием или ссылка на GitHub;
* в файле или на GitHub содержится выполненное задание;
* доступ к материалам открыт.

**Задание не выполнено:**

* если файл или ссылка на GitHub не прикреплены или отсутствует доступ к ним;
* файл или страница на GitHub не содержит выполненное задание.

**Дедлайн**

6 дней после открытия задания на платформе.

**Описание задания**

Повторите работу из демонстрации вебинара «Управление данными и их качество».

Рассчитайте значение метрики PSI для данных, различающихся во времени.

Если значение получается в допустимых пределах, добавьте аномальные значения для ухудшения значения\*.

[По ссылке](https://www.kaggle.com/code/podsyp/population-stability-index) — возможная реализация

\*Особенности тестирования работоспособности алгоритмов качества данных. Оценивают работу алгоритма на реальных данных. Если результат положительный или отрицательный, то улучшают или портят выборки для проверки корректности алгоритма.