

ДОМАШНЕЕ ЗАДАНИЕ

Срок сдачи: 04.09.2026 до 23:59

1 Общие требования

Работа выполняется индивидуально. Сдайте один архив со следующими файлами:

1. Ноутбук `.ipynb`. Все ячейки должны быть выполнены, код должен запускаться сверху вниз без ошибок. Комментарии и выводы оформите в markdown-ячейках рядом с соответствующим кодом.
2. Папка `raw/` с сохраненными исходными данными, полученными при сборе.
3. Итоговый датасет `data.csv` с собранными данными.
4. Краткий отчет `report.pdf` объемом 1-2 страницы.

2 Задание

Выберите источник из таблицы на сайте Летней школы.

1. Попробуйте собрать данные с выбранного сайта с помощью Selenium или Playwright. Сначала проверьте, какие данные можно получить обычным запросом через `requests` и `BeautifulSoup`, и сравните результат с тем, что отображается в браузере.
2. Посмотрите во вкладке Network, какие запросы выполняются при загрузке страницы. Если нужные данные доступны через отдельный запрос, попробуйте получить их напрямую.
3. Соберите не менее 100 записей или все доступные записи, если их меньше. Если на сайте используется пагинация, кнопка «Показать еще» или бесконечная прокрутка, реализуйте переход к следующим данным.
4. Сохраните исходные данные в папку `raw/`, а итоговый датасет - в `data.csv`.
5. Попробуйте отключить загрузку ресурсов, которые не нужны для сбора данных, например изображений, шрифтов и видео. Сравните время работы и количество сетевых запросов до и после оптимизации.
6. В файле `report.pdf` кратко опишите выбранный источник, способ сбора, возникшие сложности и полученный результат.