

## Лабораторная 4. Работа с данными

Изучим:

- методы парсинга данных с сайтов (библиотеки Requests, BeautifulSoup, Selenium),
- статистическую обработку полученных данных (библиотеки NumPy и Pandas),
- визуализацию данных (библиотеки Matplotlib и Seaborn).

1. Выберите сайт с которого можно спарсить информацию примерно следующего вида:

- название товара – цена *или*
- название фильма – рейтинг *или*
- дата – обменный курс доллара *или*
- дата – температура днем и т.д.д
- *можно парсить несколько значений, например, **товар-цена-рейтинг** и т.д., в этом случае можно получить более интересные результаты при последующем анализе.*

для парсинга статических страниц можно использовать библиотеки **Requests**, **Beautiful Soup**, а для парсинга динамических страниц – **Selenium**. Например, используя **Selenium**, можно ввести поисковый запрос и спарсить полученные ответы.

2. Сохраните данные в csv-файл.

3. В новом модуле прочитайте данные из csv-файла. Для работы с данными импортируйте библиотеки **NumPy** и **Pandas**.

- 1) Отсортируйте данные.
- 2) Выведите 5 первых значений.
- 3) Примените метод *describe* для расчета основных метрик статистики, объясните полученные результаты.
- 4) Если данные в первом столбике повторяются, то примените группировку (например, выведите среднюю цену повторяющегося товара).

4. Визуализируйте распределение данных при помощи библиотек **Matplotlib** или **Seaborn**. Постройте гистограмму и диаграмму box-plot. Объясните, что изображено на этих диаграммах. Можете построить еще какие-либо диаграммы, доступные в этих библиотеках.

5 \* Реализуйте графический UI для заданий из пунктов 3 и 4.