

Лабораторная работа № 1

Для курса понадобится **Python**. Работать в нем можно разными способами:

- ✓ **установить Python** и **скачать необходимые для лабораторной библиотеки**. В этом случае можно работать в IDE **PyCharm** или в **Jupyter Notebook**. Рекомендация: в нашем случае удобнее и нагляднее работать в Jupyter Notebook, но это дело вкуса, если есть опыт работы в PyCharm или других средах разработки, то можете работать где вам удобнее;
- ✓ **зарегистрироваться в Google Colab**, **в этом случае не придется устанавливать библиотеки на компьютер**, все происходит в облаке. Структура такая же как и в **Jupyter Notebook**, что удобно для визуализации и обработки данных.

Внимание: в случае с Jupyter Notebook и Google Colab для работы и сдачи лаб понадобится интернет, поэтому рассчитывайте свои возможности.

Итак, задание ↓

Часть 1. Знакомство с Python

1. Установите пакеты NumPy, Pandas.
2. Используя возможности **NumPy**:
 - создайте двумерный массив из 20 целых случайных чисел
 - разделите полученный массив на 2 массива
 - найдите все заданные значения в первом массиве (например, равные 6)
 - подсчитайте количество найденных элементов
2. **Pandas**. Изучите структуры данных Series и Dataframe:
 - создайте объект **Series** из массива NumPy
 - произведите с ним различные математические операции
 - создайте объект **Dataframe** из массива NumPy
 - напишите строку заголовков в созданном **Dataframe**
 - удалите любую строку
 - удалите любой столбец
 - выведите размер получившегося **Dataframe**
 - найдите все элементы равные какому-либо числу

Часть 2. Статистическая обработка данных и библиотека Matplotlib

1. Скачайте датасет в формате csv на ваш выбор (например, датасеты можно найти на сайте Kaggle: <https://www.kaggle.com/datasets> или в других источниках)
2. Импортируйте данный csv-файл в python используя модуль pandas, **выведите данные для просмотра**.
3. Установите библиотеку визуализации **Matplotlib**. Постройте гистограмму частот по какому-то определенному столбцу (**параметру**).
4. Рассчитайте медиану и среднее значение параметра.
5. Постройте **box plot** для выбранного параметра. Объясните, что на нем изображено.
6. Примените к выбранному параметру метод **.describe()**. Поясните параметры метода и полученные результаты.
7. Сгруппируйте данные по какому-либо признаку и произведите расчет и анализ по каким-либо параметрам. Примеры группировок и рассуждений можно посмотреть в примере [Examples/Titanic.html](#)