

Лабораторная работа № 3**Изучение библиотек для обработки и визуализации данных****Цели работы:**

1. Познакомиться с библиотеками Python NumPy, Pandas и Matplotlib.
2. Разобраться с принципами работы в среде Jupyter Notebook на основе сервиса Google Colab.
3. Получить навыки выполнения векторных вычислений с использованием библиотеки NumPy.
4. Получить навыки анализа, обработки и визуализации датасетов с применением библиотек Pandas и Matplotlib.

Основные положения

В работе представлены задания двух типов. Первый тип посвящен изучению библиотеки NumPy и представляет собой задачи на вычисление векторно-матричных выражений. Они приведены в файле `lab3_TP_NumPy_Tasks.pdf`. Второй тип заданий нацелен на формирование навыков работы с библиотеками Pandas и Matplotlib. Он представлен задачами, сформулированными в ноутбуке `lab3_TP_Pandas_Tasks.ipynb`.

Порядок выполнения работы

1. Авторизуйтесь под своей учетной записью на <https://google.com>. Откройте приложение «Google-Диск» (рисунок 1).

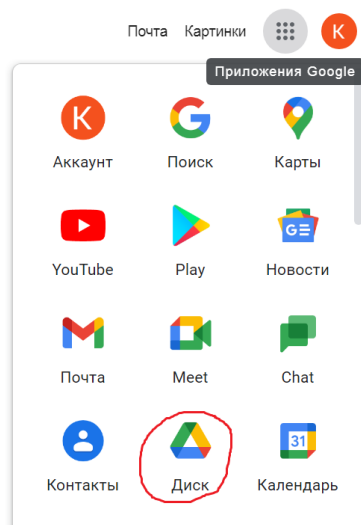


Рисунок 1 – Запуск сервиса Google-Диск

2. На Google-Диске создайте директорию для работы с заданиями лабораторной работы. Для этого нажмите кнопку «Создать» в верхнем левом углу страницы сервиса и выберите пункт «Папку» (рисунок 2).

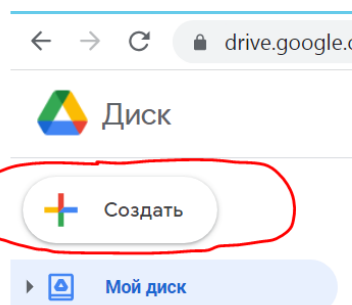


Рисунок 2 – Создание папок и файлов на Google-диске

3. В созданной папке создайте файл «Google Colaboratory» («Создать» → «Еще...» → «Google Colaboratory») (рисунок 3).

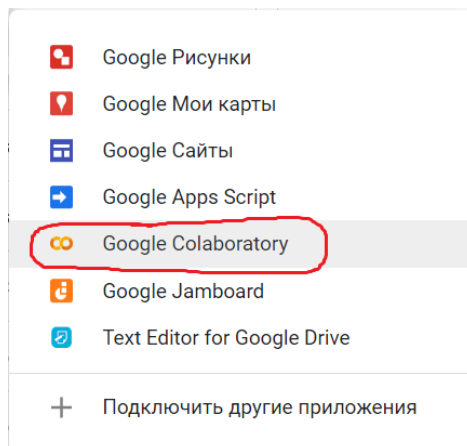


Рисунок 3 – Создание нового ноутбука Google Colab

4. На открывшейся странице можно размещать программный код и дополнительную информацию. Добавление информации и кода осуществляется созданием ячейки соответствующего типа (рисунок 4). Для того, чтобы ячейка с кодом выполнялась, нужно установить курсор в эту ячейку и нажать «Ctrl-Enter». «Текстовая» ячейка может содержать форматированный текст, изображения, формулы и т.д. Формат описывается языком разметки Markdown: <https://guides.hexlet.io/markdown/>.

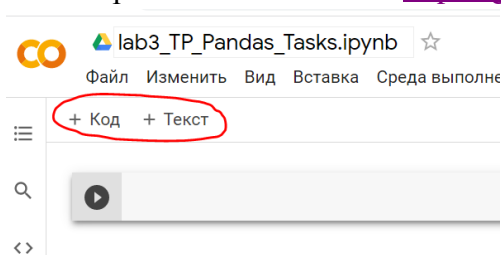


Рисунок 4 – Создание новой ячейки в Google Colab

5. Выполните задания по варианту из файла **lab3_TP_NumPy_Tasks.pdf**. Для этого создайте ячейку с кодом и импортируйте в нее библиотеку numpy (рисунок 5). Составьте последовательность инструкций на языке Python, решающих поставленную задачу. Примеры инструкций можно посмотреть на <https://pythonworld.ru/numpy/100-exercises.html>. Не забывайте вставлять пояснения в ноутбук с помощью ячеек с Markdown-разметкой. После выполнения задания вставьте в конец документа текстовую ячейку выводами.

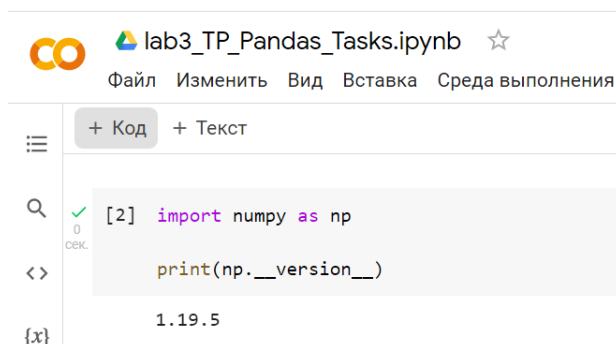


Рисунок 5 – Импортирование библиотеки NumPy и проверка ее версии

6. Загрузите в Вашу папку на Google-Диске файл `lab3_TP_Pandas_Tasks.ipynb`. Для этого можно просто перетащить файл из проводника в браузер на соответствующую страницу (рисунок 6).

7. Двойным щелчком мыши по файлу откройте его в Google Colaboratory.

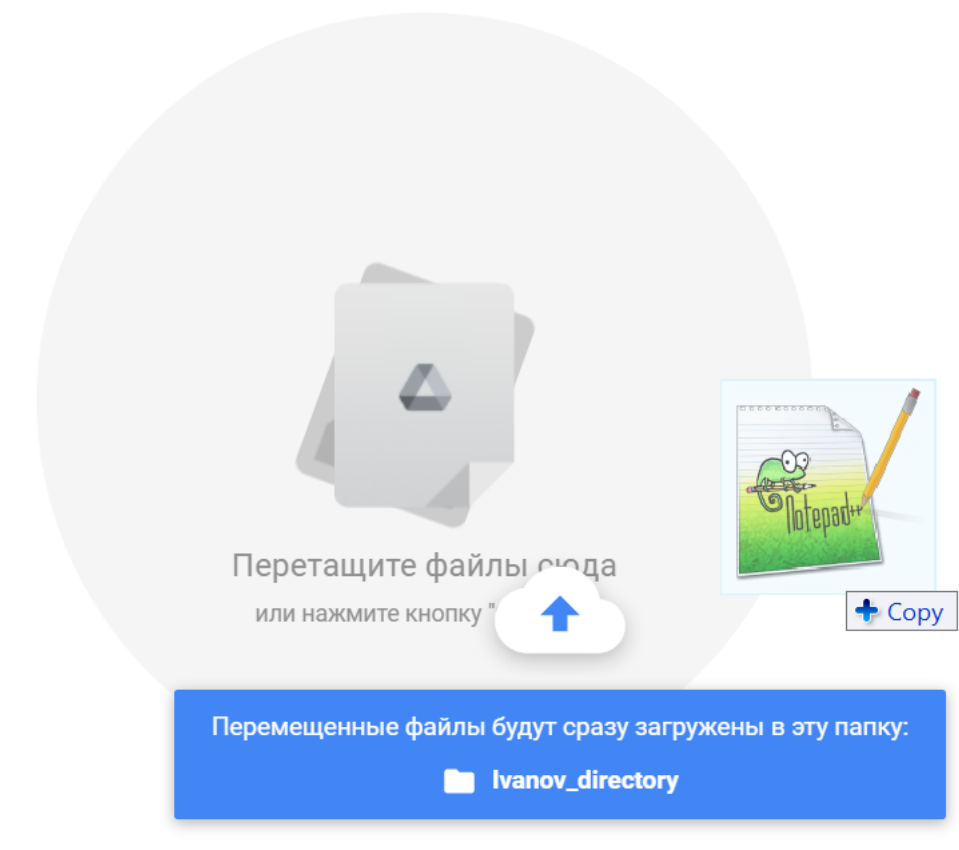


Рисунок 6 – Добавление файла на Google-Диск

8. Выполните все ячейки ноутбука, содержащие код. Убедитесь в корректной загрузке датасета и формировании датафрейма.

9. Заполните ячейки ноутбука в соответствии с заданиями, сформулированными в текстовых ячейках. Сделайте выводы.

Индивидуальное задание

Формулировка индивидуальных заданий по вариантам приводится в файлах **lab3_TP_NumPy_Tasks.pdf** и **lab3_TP_NumPy_Tasks.ipynb**.

Требования к представлению лабораторной работы

Лабораторная работа представляется к отчету в виде файлов-ноутбуков расширения ipynb с выполненными и прокомментированными заданиями. Отдельный протокол для лабораторной работы оформлять **не нужно**. В файле-ноутбуке должна содержаться все необходимая информация о лабораторной (постановка задачи, ход работы, используемые языки / библиотеки / технологии, выводы по работе).

Теоретические вопросы к отчету лабораторной работы

1. Библиотека NumPy: назначение и возможности.
2. Понятие размерности вектора / матрицы (dimension), формы (shape), типа данных элементов.
3. Основные векторные и матричные операции, их математическая нотация и реализация посредством NumPy. Broadcasting, правила его выполнения.
4. Библиотеки Pandas и Matplotlib: назначение и возможности.
5. Понятие Series и DataFrame. Способы их создания и вывода на экран.
6. Основные операции с Series и DataFrame, их реализация в программном коде.
7. Визуализация данных с помощью Matplotlib. Типы графиков и диаграмм в Matplotlib. Библиотеки Seaborn и Plotly.

Список литературы для подготовки

1. Маккинни, У. Python и анализ данных / У. Маккинни ; перевод с английского А. А. Слинкина. — 2-ое изд., испр. и доп. — Москва : ДМК Пресс, 2020. — 540 с.
2. Груздев, А. В. Изучаем Pandas / А. В. Груздев, М. Хейдт ; перевод с английского А. В. Груздева. — 2-ое изд., испр. и доп. — Москва : ДМК Пресс, 2019. — 700 с.
3. Официальная документация NumPy [Электронный ресурс]. — Режим доступа : <https://numpy.org/doc/> (дата обращ. 15.11.2021).
4. Официальная документация Pandas [Электронный ресурс]. — Режим доступа : <https://pandas.pydata.org/docs/> (дата обращ. 15.11.2021).
5. Официальная документация Matplotlib [Электронный ресурс]. — Режим доступа : <https://matplotlib.org/> (дата обращ. 15.11.2021).