



# **ВВЕДЕНИЕ В ОБРАБОТКУ И ВИЗУАЛИЗАЦИЮ ДАННЫХ В GOOGLE COLAB (БЕЗ УСТАНОВКИ PYTHON)**

# ИСТОРИЧЕСКИЙ КОНТЕКСТ

Обработка и визуализация данных — ключевые этапы анализа, берущие начало в:

- **XVIII веке:** первые статистические графики У. Плейфэра (линейные и столбчатые диаграммы).
- **XX веке:** развитие вычислительных методов (язык FORTRAN для научных расчетов).
- **2000-е:** расцвет Python благодаря библиотекам (NumPy, Pandas, Matplotlib), созданным для научных задач.

# ОСНОВНЫЕ ПОНЯТИЯ

**Обработка данных** — преобразование "сырых" данных в структурированный формат:

**Цели:** очистка от шумов, заполнение пропусков, нормализация.

**Инструменты:**

- Табличные данные → Pandas (аналог Excel в коде).
- Числовые массивы → NumPy (математические операции).

**Визуализация данных** — представление информации в графической форме:

**Задачи:** выявление паттернов, аномалий, демонстрация результатов.

**Библиотеки:**

- Matplotlib — базовые графики (как MATLAB).
- Seaborn — статистическая визуализация.
- Plotly — интерактивные диаграммы.

# КЛЮЧЕВЫЕ ЭТАПЫ РАБОТЫ

## 1. Загрузка данных:

- Источники: CSV, Excel, SQL, API.
- Проблемы: кодировки, разделители, пропуски.

## 2. Предобработка:

- Очистка: удаление дубликатов, обработка выбросов.
- Трансформация: нормализация, логарифмирование.

## 3. Анализ:

- Описательная статистика.
- Группировки, агрегация.

## 4. Визуализация:

- Выбор типа графика: гистограммы для распределений, scatter plot для корреляций.
- Кастомизация: подписи осей, легенды, цветовые схемы.

# ПРИМЕРЫ ПРИМЕНЕНИЯ

- **Финансы:** анализ котировок акций (свечные графики).
- **Медицина:** визуализация распространения заболеваний на картах.
- **Соцсети:** построение облака тегов из постов.

## Почему Python?

- **Простота:** синтаксис близок к естественному языку.
- **Экосистема:** тысячи библиотек (SciPy для науки, GeoPandas для карт).
- **Сообщество:** готовые решения на Stack Overflow, Kaggle.

# ЧТО ТАКОЕ GOOGLE COLAB?

**Google Colab (Colaboratory)** — это бесплатная облачная среда для работы с Python, которая позволяет:

- Запускать код прямо в браузере.
- Использовать GPU и TPU бесплатно.
- Работать с данными без установки программ.
- Сохранять проекты в Google Drive.

**Ссылка:** [colab.research.google.com](https://colab.research.google.com)

# ПЛЮСЫ И МИНУСЫ GOOGLE COLAB

## Преимущества:

- Не нужно ничего устанавливать.
- Бесплатный доступ к GPU/TPU.
- Интеграция с Google Drive.

## Недостатки:

- Требуется интернет.
- Ограниченная RAM в бесплатной версии.



# КАК НАЧАТЬ РАБОТУ?

Открываем новый блокнот:

- 1.Перейдите по ссылке выше.
- 2.Нажмите "**Новый блокнот**" (или откройте примеры из галереи).

**Интерфейс Google Colab:**

- **Ячейки кода** — здесь пишем и выполняем Python-код.
- **Текстовые ячейки** — для комментариев и Markdown-разметки.
- **Меню** — запуск кода, настройки, подключение к GPU.