# **Лабораторная работа #1**

**Цель работы:** научиться применять инструменты для подготовки данных.

**Инструменты:** pandas, numpy, scipy, matplotlib/seaborn.

**Содержание:**

Для предложенных датасетов провести первичный статистический анализ. Для этого необходимо:

1. Вычислить меры центральной тенденции (выборочное среднее, медиана, мода и т.д.)
2. Вычислить меры разброса (дисперсия, среднеквадратичное отклонение, min/max-value, IQR)
3. Вычислить параметры формы распределений каждого признака (асимметрия, эксцесс)
4. Построить гистограммы каждого признака, а также ECDF (функция распределения) и KDE (ядерная оценка плотности вероятности). Можно также построить pairplot из seaborn.
5. Вычислить матрицу корреляций между признаками. Построить тепловую карту (heatmap).

После этого исследовать датасеты на возможность:

1. Извлечения новых признаков (feature extraction).
2. Преобразования признаков (feature transformation). *Как минимум z-score scaling.*
3. Отбора признаков и объектов (feature selection). *Как минимум анализ выбросов, аномалий и сильно коррелирующих признаков.*

Отчет должен включать:

1. Цифры, графики и выводы из первичного статистического анализа.
2. Описание новых признаков, методов трансформации и отбора.