# ЛР 04: Предварительный анализ данных

## Задание 1: Обработка пропущенных значений

Используя набор данных о пассажирах Титаника из библиотеки Seaborn (sns.load\_dataset('titanic')), необходимо найти и обработать все пропущенные значения. Можно использовать различные методы (например, удаление, заполнение средним значением или медианой). Сравнить результаты.

## Задание 2: Поиск аномальных значений в данных о пассажирах Титаника

На основе набора данных о пассажирах Титаника ([ссылка на Kaggle](https://www.kaggle.com/c/titanic)), найдите аномалии в столбцах, таких как возраст и стоимость билетов, используя методы расстояния, статистические тесты или машинное обучение.

## Задание 3: Кодирование категориальных значений

В том же наборе данных о пассажирах Титаника, необходимо перекодировать категориальные значения, такие как "sex" и "embarked", используя one-hot encoding.

## Задание 4: Нормализация данных

Используя набор данных о цветках Ириса из библиотеки sklearn (sklearn.datasets.load\_iris()), необходимо нормализовать все признаки, используя Min-Max Scaling или z-score standardization.

## Задание 5: Описательная статистика

На любом из указанных наборов данных студенты рассчитать основные описательные статистики (среднее, медиана, мода, стандартное отклонение, и т.д.) для каждого числового признака. Построить гистограммы, чтобы визуализировать распределение каждого признака.

## Задание 6: Корреляции

На любом из указанных наборов данных рассчитать корреляции между различными признаками и визуализировать их с помощью тепловой карты.,