# Лабораторная работа № 05 (предварительный анализ данных)

## Задание 1: Детекция аномалий в данных кредитного скоринга

Используйте набор данных кредитного скоринга с Kaggle ([ссылка на данные](https://www.kaggle.com/mlg-ulb/creditcardfraud)) и определите аномалии в данных, используя Z-score или IQR.

## Задание 2: Поиск аномалий в данных кредитного скоринга

Используйте набор данных кредитного скоринга ([прямая ссылка на данные Kaggle](https://www.kaggle.com/mlg-ulb/creditcardfraud)) и определите аномалии, используя методы Z-score или IQR.

## Задание 3: Обнаружение аномалий в данных о температуре

В наборе данных о температуре в различных городах ([прямая ссылка на данные Kaggle](https://www.kaggle.com/smid80/weatherww2)) обнаружьте аномальные значения температуры.

## Задание 4: Прогнозирование пропущенных значений в данных о болезни диабета.

В данных о диабете Pima Indians (Kaggle: [ссылка на данные](https://www.kaggle.com/uciml/pima-indians-diabetes-database)) есть пропущенные значения. Воспользуйтесь методами регрессии или классификации для предсказания пропущенных значений.

## Задание 5: 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 6: Нормализация и стандартизация признаков датасета "Red Wine Quality"

Для набора данных о качестве красного вина ([ссылка на данные](https://www.kaggle.com/uciml/red-wine-quality-cortez-et-al-2009)) примените различные методы нормализации и стандартизации к признакам, сравните результаты их применения.

## Задание 7: Применение описательной статистики и корреляционного анализа для датасета "House Prices"

На данных о ценах на жилье ([ссылка на данные](https://www.kaggle.com/c/house-prices-advanced-regression-techniques/data)) рассчитайте основные метрики описательной статистики для числовых переменных и рассчитайте корреляции между переменными. Визуализируйте корреляции с помощью тепловых карт и обратите внимание на сильные корреляции.

## Задание 8: Обработка текстовых данных в датасете "Twitter Airline Sentiment"

Набор данных "Twitter Airline Sentiment" на Kaggle ([ссылка на данные](https://www.kaggle.com/crowdflower/twitter-airline-sentiment)) содержит реальные твиты о различных авиакомпаниях. Примените техники предобработки текста, вроде удаления стоп-слов, стемминга, лемматизации и векторизации для подготовки данных к анализу.