**Лабораторная работа № 2**

**Тема**: Обработка данных.

**Цель работы** *– Знакомство с методами обработки данных. Применение этих методов при помощи функций модуля pandas.*

Для выполнения данной лабораторной работы можно использовать любой найденный датасет с пропусками (можно из примеров или свой), а если пропусков нет, то необходимо их сгенерировать.

Примеры датасетов с пропусками:

<https://www.kaggle.com/datasets/uciml/german-credit>

<https://www.kaggle.com/datasets/parulpandey/2020-it-salary-survey-for-eu-region>

**Нужно**:

1. Выявить пропуски данных несколькими способами (визуальный, расчетный…)
2. Исключить строки и столбцы с наибольшим количеством пропусков
3. Произвести замену оставшихся пропусков.
4. Построить гистограмму распределения исходного датасета до и после обработки пропусков. Сделать выводы как обработка данных повлияла на их распределение.
5. Проверить датасет на наличие выбросов, удалить найденные аномальные записи.
6. Привести все параметры к числовому виду.
7. Сохранить обработанный датасет.

*При удалении (замене) пропусков необходимо рассуждать: можно ли удалить данный параметр и чем целесообразно заменять пропуски данных в конкретных параметрах, руководствуясь описанием параметров датасета и предметной областью.*