**Наименование работы**:

Очистка и трансформация данных

**Цель работы:**

Научиться подготавливать данные для обучения моделей

**Задачи**:

В соответствие с набором данных из лаб. №2 провести процедуру очистки данных:

1. Удаление выбросов;
2. Заполнение пропущенных значений;
3. Исправление некорректных значений;
4. Кодирование категориальных признаков.

Для реализации рекомендуется использовать язык программирования Python 3.x и библиотеки Pandas, Matplotlib, Numpy, Sklearn

**Оформление результатов**:

Результаты лабораторной работы оформляются в виде отчета в формате PDF.

**Структура отчета:**

* 1. Титульный лист;
  2. Основная часть;
  3. Заключение.

**В основной части приводятся:**

Описание выполнения каждой из поставленных задач в виде:

1. Текста с описанием проделанной работы (что было/что стало);
2. Скриншотов с программным кодом.

**В заключении** приводятсявыводы по проделанной работе

Таблица 1 - наборы данных

| № | Набор данных |
| --- | --- |
| 1 | Данные по штрафам за парковку  <https://www.kaggle.com/new-york-city/nyc-parking-tickets> (любой из 4-х файлов) |
| 2 | Данные о заболеваемости COVID-19  <https://www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset> |
| 3 | Данные о скачиваниях и рейтингах приложений в Google Play  <https://www.kaggle.com/lava18/google-play-store-apps> |
| 4 | Данные о статистике суицидов по странам с 1985 по 2016 годы  <https://www.kaggle.com/russellyates88/suicide-rates-overview-1985-to-2016> |
| 5 | Данные по БУ авто с Craiglist  <https://www.kaggle.com/austinreese/craigslist-carstrucks-data> |