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 приводятся выводы по проделанной работе

Таблица 1 - наборы данных

№	Набор данных	Бизнес-кейс
1	Данные по штрафам за парковку https://www.kaggle.com/new-york-city/nyc-parking-tickets	1. Общее число штрафов, сгруппированное по штатам 2. Наиболее частый тип кузова автомобилей, получающих штраф 3. Число штрафов, выданных за проезд на красный сигнал светофора в 2015 году в Нью-Йорке
2	Данные о заболеваемости COVID-19 https://www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset	1. Число смертей за март 2020 года, сгруппированное по странам 2. Три наиболее заражаемых штата в США 3. Общее число заражений по дням за последние 30 дней наблюдения
3	Данные о скачиваниях и рейтингах приложений в Google Play https://www.kaggle.com/lava18/google-play-store-apps	1. Десять категорий приложений с наиболее высоким средним рейтингом 2. Максимальное число отзывов о приложении для платных и бесплатных приложений 3. Наиболее популярный жанр приложений дороже 5 долларов
4	Данные о статистике суицидов по странам с 1985 по 2016 годы https://www.kaggle.com/russellyates88/suicide-rates-overview-1985-to-2016	1. Три страны с наиболее частыми случаями суицида из ТОП10 стран с низким ВВП 2. Среднее по всем странам число суицидов, произошедших в год вашего рождения 3. Три самые частотные возрастные категории по суициду
5	Данные по БУ авто с Craiglist https://www.kaggle.com/austinreese/craigslist-carstrucks-data	1. Средняя цена автомобилей, сгруппированная по марке производителя 2. Пять наиболее дешевых марок производителей (считать только по 6-ти цилиндровым автомобилям) 3. Число автомобилей дешевле 5000$, сгруппированное по годам выпуска