**Задание №1 – Предварительный анализ данных и визуализация**

Написать программу на Python, которая загружает набор данных по вашему варианту и производит исследовательский анализ этих данных и визуализирует ряд зависимостей между признаками в нижеперечисленных вариациях с помощью библиотек matplotlib, sns и plotly.

1. Получить описание набора данных и список атрибутов, получить число пропущенных значений в для каждого атрибута.

2. Обработать пропущенные значения (удалить строки/ удалить атрибуты/ выполнить импутацию значений)

3. Построить pairplot для набора данных

4. Выбрать не менее 2-х признаков с неким распределением значений, и отобразить это распределение с помощью гистограмм hist и kdeplot, jointplot

5. Выбрать часть признаков датасета и отобразить корреляцию между ними в виде тепловой карты (heatmap).

6. Выбрать 3 признака (имеющие некое распределение значений), целевую переменную (категориальную) и отобразить для них графики размаха (ящик с усами, box plot).

8. Визуализировать некоторые статистики, для разных атрибутов с использованием следующих инструментов: violinplot, stripplot, swarmplot, pie.

Напишите короткое заключение о наиболее интересных зависимостях, которые Вы обнаружили в данных.

Дополнительный пример визуализации: <https://www.kaggle.com/heyytanay/stroke-prediction-eda>

Для своего варианта необходимо посмотреть последнюю цифру номера своей зачетной книжки (или студенческого билета) и выполнить следующие корректировки:

• если последняя цифра 0 или 5: датасет – Лесные пожары (https://archive.ics.uci.edu/ml/datasets/Forest+Fires), предсказываемое значение – площадь пожара (Area);

• если последняя цифра 1 или 6: датасет – Качество вина (https://archive.ics.uci.edu/ml/datasets/Wine+Quality) предсказываемое значение – качество (Quality), для датасета с красным вином, winequality-red.csv;

• если последняя цифра 2 или 7: датасет – Качество вина (https://archive.ics.uci.edu/ml/datasets/Wine+Quality) предсказываемое значение – качество (Quality), для датасета с белым вином, winequality-white.csv;

• если последняя цифра 3 или 8: датасет – Аренда велосипедов (https://archive.ics.uci.edu/ml/datasets/Bike+Sharing+Dataset), предсказываемое значение – количество аренд велосипедов в сутки (Area), датасет day.csv;

• если последняя цифра 4 или 9: датасет – Аренда велосипедов (https://archive.ics.uci.edu/ml/datasets/Bike+Sharing+Dataset), предсказываемое значение – количество аренд велосипедов в час (Area), датасет hour.csv;