**Министерство науки и высшего образования Российской Федерации** **Федеральное государственное бюджетное образовательное учреждение высшего образования** **«Московский государственный технический университет** **имени Н.Э. Баумана**

**(национальный исследовательский университет)»**

**(МГТУ им. Н.Э. Баумана)**

**Факультет «Информатика и системы управления»**

**Кафедра ИУ5 «Системы обработки информации и управления»**

Отчёт по РК1

«Технологии машинного обучения»

Вариант 4

Выполнила:

студентка группы ИУ5-63Б

Жамнова М.С.

Преподаватель:

Гапанюк Ю. Е.

2023 г.

Задание:

Задача №1.

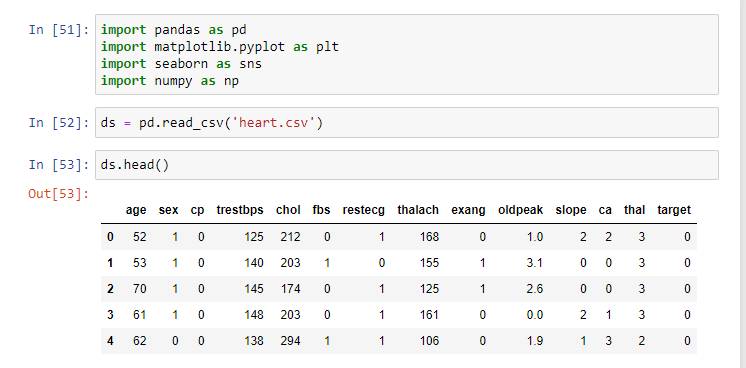
Для заданного набора данных проведите корреляционный анализ. В случае наличия пропусков в данных удалите строки или колонки, содержащие пропуски. Сделайте выводы о возможности построения моделей машинного обучения и о возможном вкладе признаков в модель.

Для студентов групп ИУ5-63Б, ИУ5Ц-83Б - для произвольной колонки данных построить график "Ящик с усами (boxplot)".

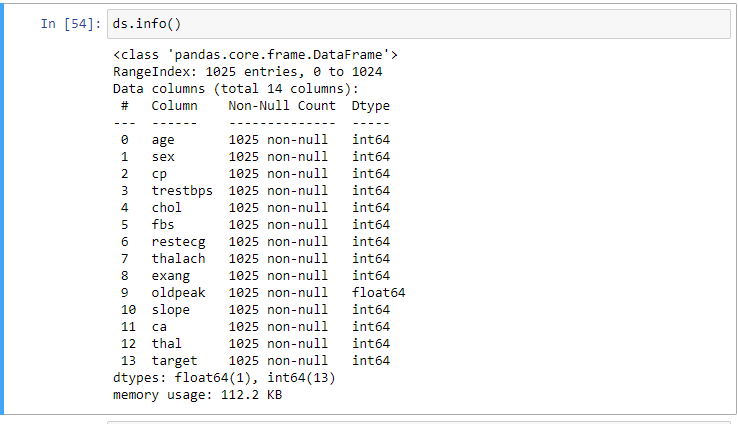
<https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset>

Решение:

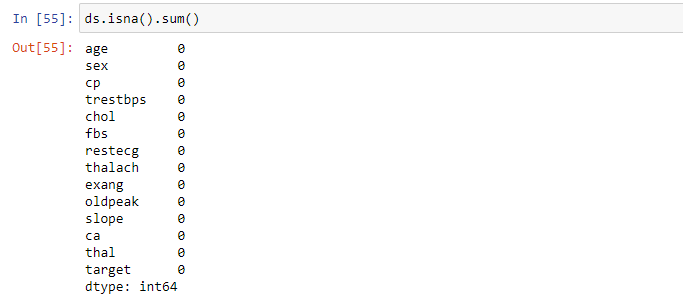
Загружаем датасет и подключаем необходимые библиотеки:

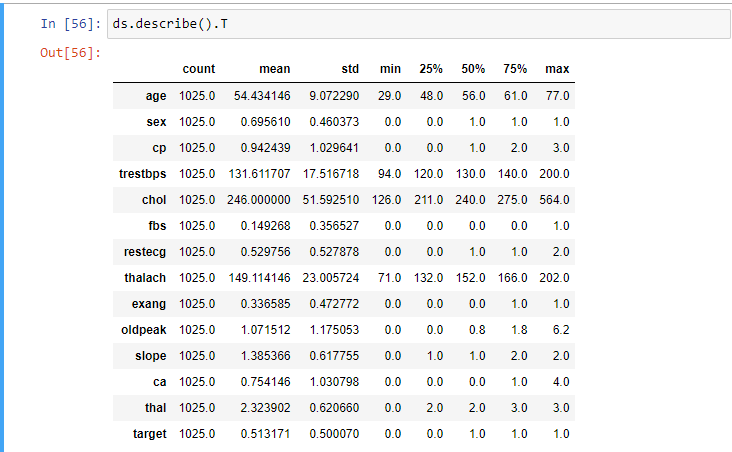


Получим информацию о датасете:

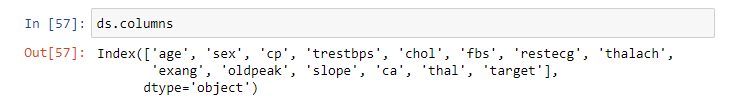


Посчитаем количество пустых значений:

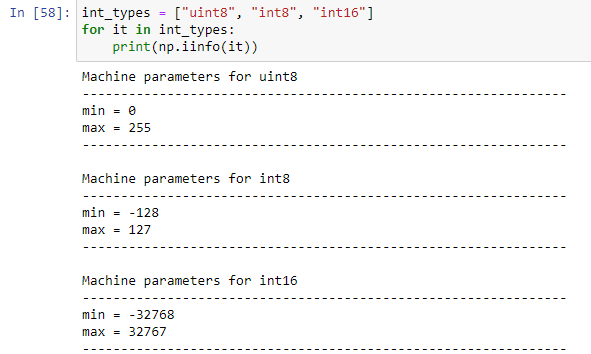


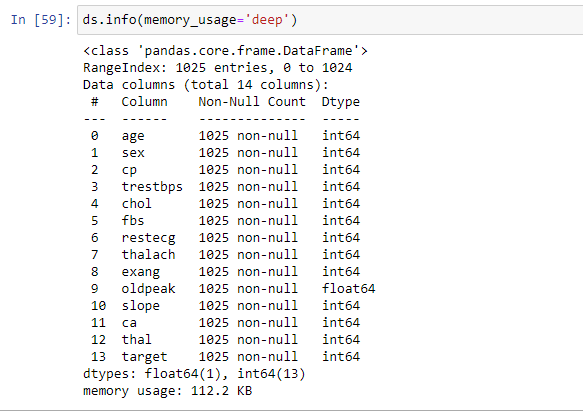


Проверим названия столбцов:

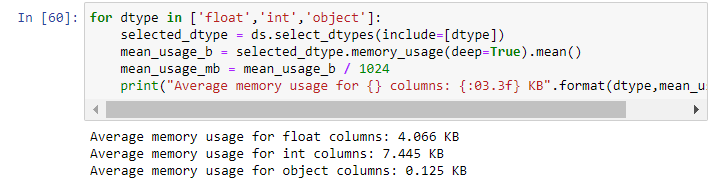


Оптимизируем память и изменим типы данных на более корректные:



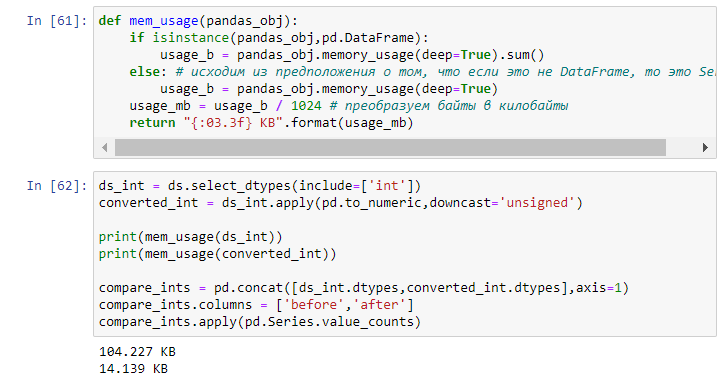


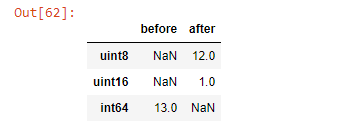
Так как данные разных типов хранятся раздельно, мы исследуем использование памяти разными типами данных. Начнём со среднего показателя использования памяти по разным типам данных:

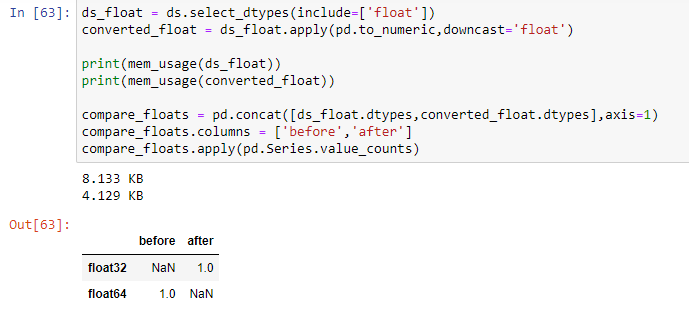


Эти сведения дают нам понять то, что большая часть памяти уходит на столбцы, хранящих объектные значения.

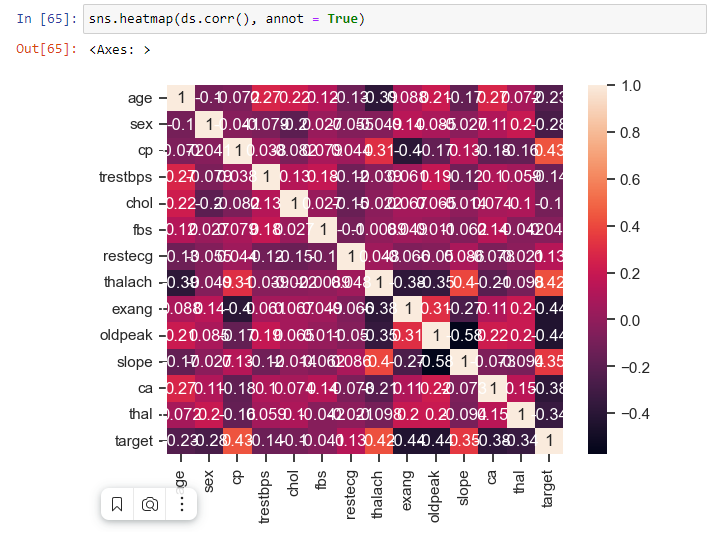
Создадим функцию, которая поможет нам сэкономить немного времени







Корреляционный анализ



Судя по корреляционному, построение такой модели машинного обучения возможно, и будет достаточно успешным. В целом, можно вкладывать практически все имеющиеся признаки, однако можно исключить target (цель), thalach (максимальная достигнутая частота сердечных сокращений), exang (стенокардия, вызванная физической нагрузкой), slope (наклон сегмента ST пиковых упражнений), oldpeck (Депрессия ST).

Построение графика "Ящик с усами (boxplot)"

