**Московский государственный технический   
университет им. Н. Э. Баумана**

Курс «Технологии машинного обучения»

Отчёт по рубежному контролю №1

«Технологии разведочного анализа и обработки данных.»

Вариант № 2

Выполнил: Проверил:  
Абрамов А.В. Гапанюк Ю.Е.  
группа ИУ5-63Б

Дата: 19.04.25 Дата:

Подпись: Подпись:

Москва, 2025 г.

**Задание:**

Номер варианта: **2**

Номер задачи: 1

Номер набора данных, указанного в задаче: **2** ([https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load\_wine.html#sklearn.datasets.load\_wine](https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load_wine.html#sklearn.datasets.load_wines))

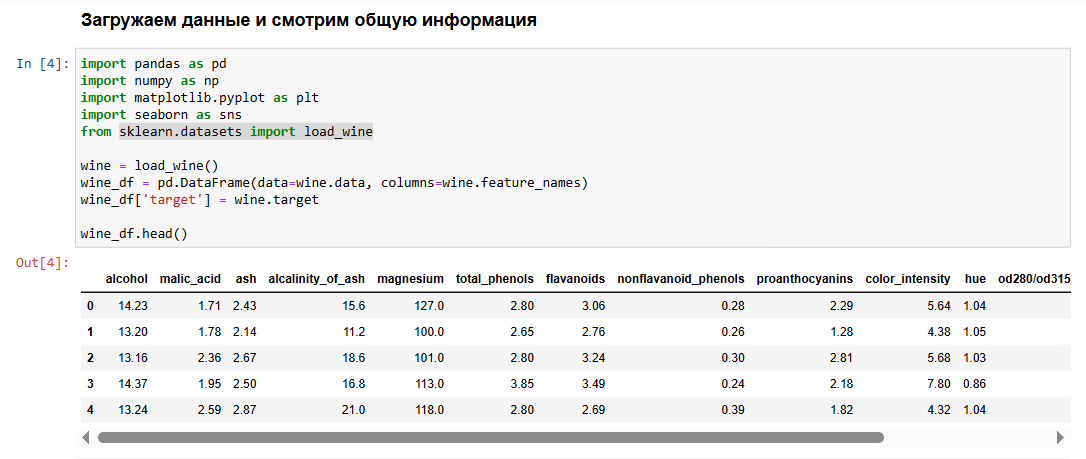
Для студентов групп ИУ5-63Б, ИУ5Ц-83Б - для произвольной колонки данных построить график "Ящик с усами (boxplot)".

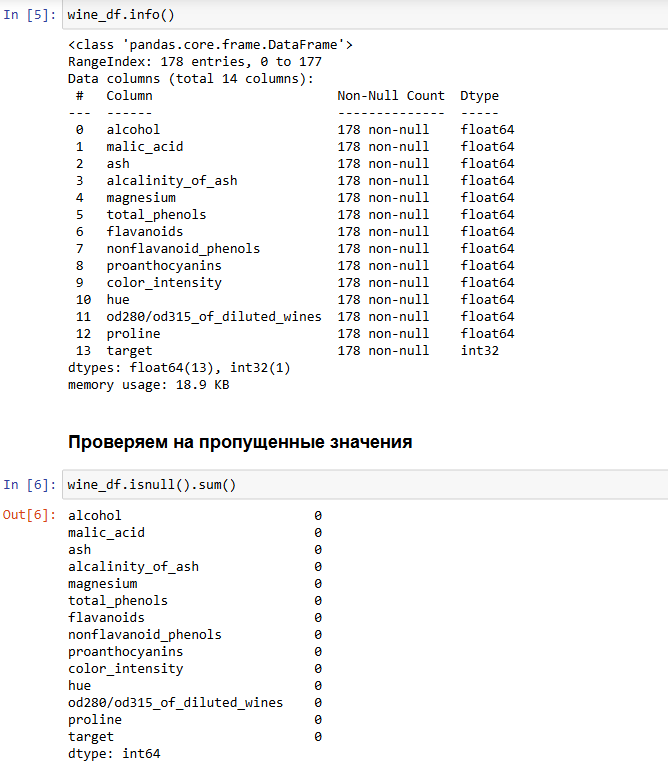
**Задача №1.**

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.

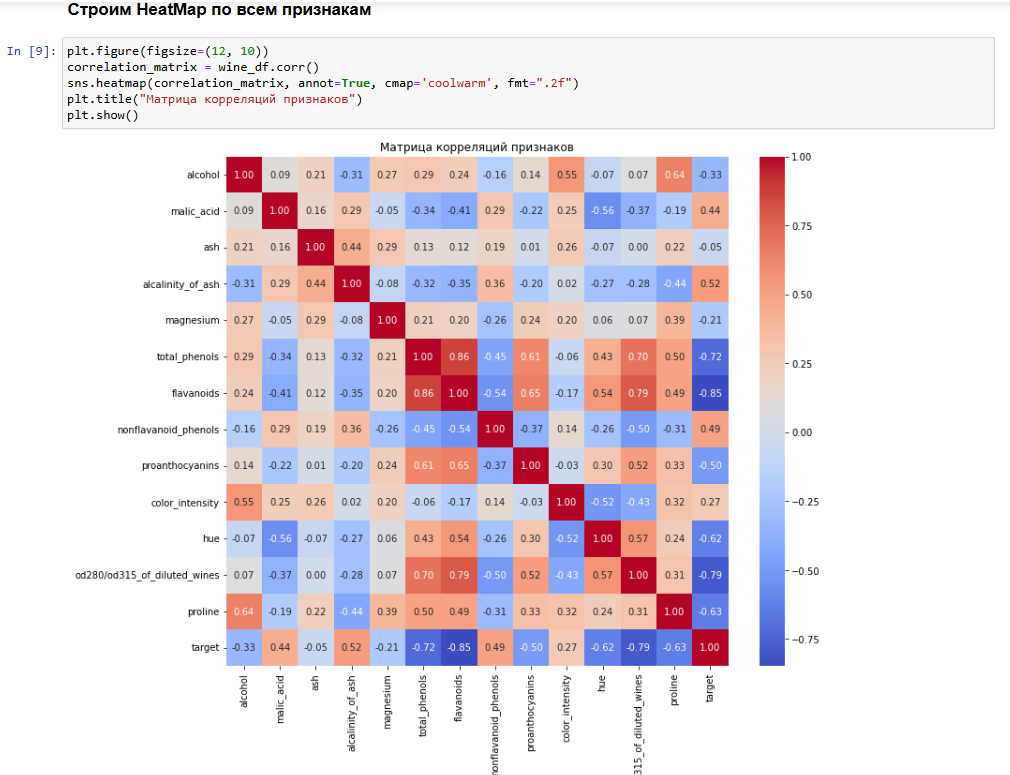
**Ход выполнения:**

1. Загрузил набор данных, просмотрел начало, проверил пропуски и выяснил, что пропуски отсутствуют.

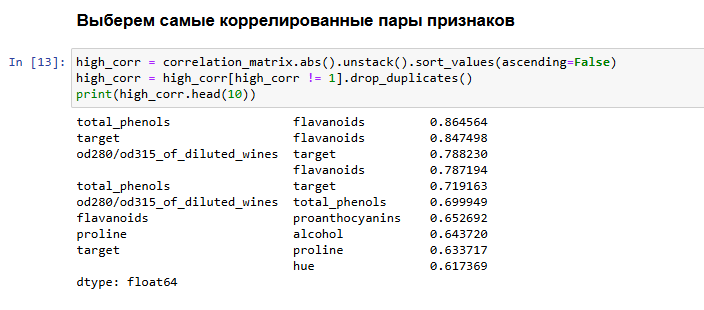




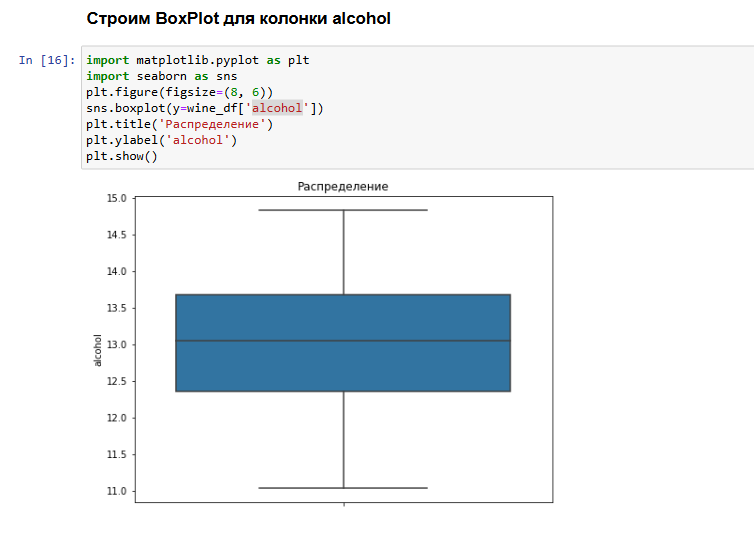
1. Строим HeatMap для проведения корреляционного анализа по всем признакам.



1. Вывели самые коррелирующие пары признаков.



1. Построил Boxplot для alcohol



### Качество данных:

В наборе данных отсутствуют пропущенные значения, что позволяет использовать все имеющиеся наблюдения для анализа и построения моделей.

**Корреляционный анализ:**

Высокая положительная корреляция между общим содержанием фенолов и флавоноидами (0.86).

Значительная положительная корреляция между флавоноидами и оптической плотностью (0.79).

Умеренная отрицательная корреляция между оттенком и интенсивностью цвета (-0.52).

### Для построения моделей следует:

Рассмотреть удаление одного из сильно коррелирующих признаков (например, оставить только флавоноиды вместо флавоноидов и общих фенолов) для снижения мультиколлинеарности.

Учитывать, что набор данных хорошо подходит для классификации благодаря различиям между классами.

Ожидать, что наибольший вклад в модель внесут флавоноиды, общие фенолы, оптическая плотность, пролин и интенсивность цвета из-за их значимых корреляций.