осковский государственный технический университет им. Н.Э. Баумана  
Кафедра «Системы обработки информации и управления»

Лабораторная работа №3

по дисциплине «Методы машинного обучения»

на тему«Обработка признаков (часть 2)»

Выполнил: Ван Пэй  
Группа：ИУ5-22М

Москва — 2021 г.

**Цель лабораторной работы:**

изучение продвинутых способов предварительной обработки данных для дальнейшего формирования моделей.

**Задание**:

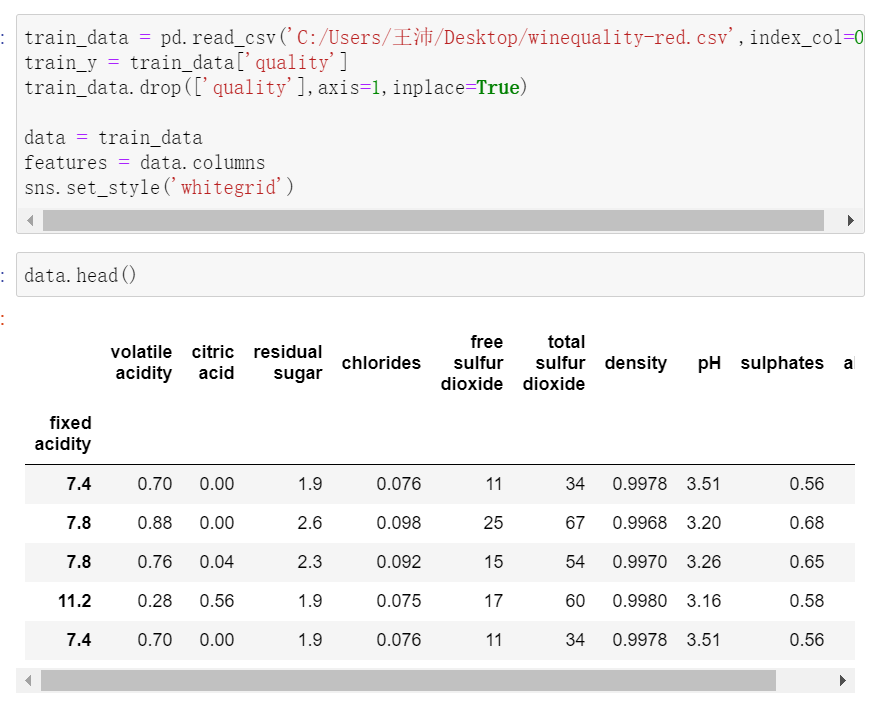
Выбрать набор данных (датасет), содержащий категориальные и числовые признаки и пропуски в данных.

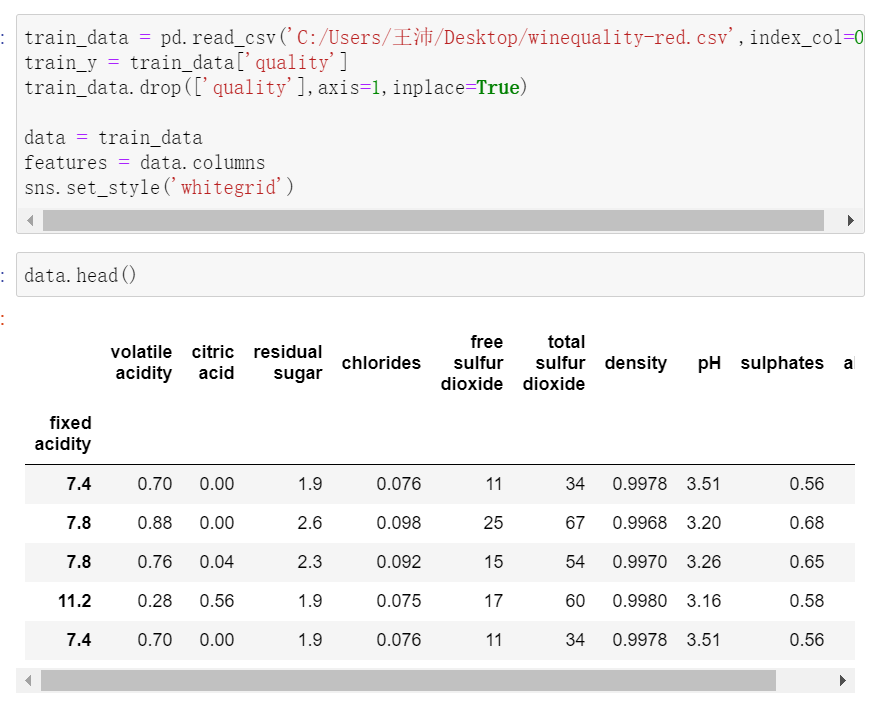
Для выполнения следующих пунктов можно использовать несколько различных наборов данных

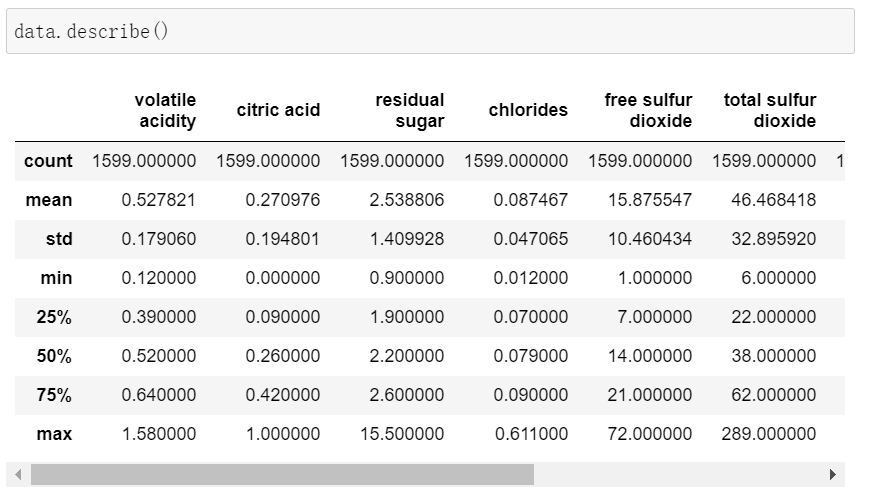
Для выбранного датасета (датасетов) на основе материалов лекций решить следующие задачи:

* масштабирование признаков (не менее чем тремя способами);
* обработку выбросов для числовых признаков (по одному способу для удаления выбросов и для замены выбросов);
* обработку по крайней мере одного нестандартного признака (который не является числовым или категориальным);
* отбор признаков: один метод из группы методов фильтрации (filter methods); один метод из группы методов обертывания (wrapper methods); \*\* один метод из группы методов вложений (embedded methods).

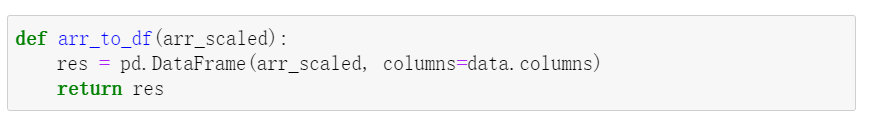






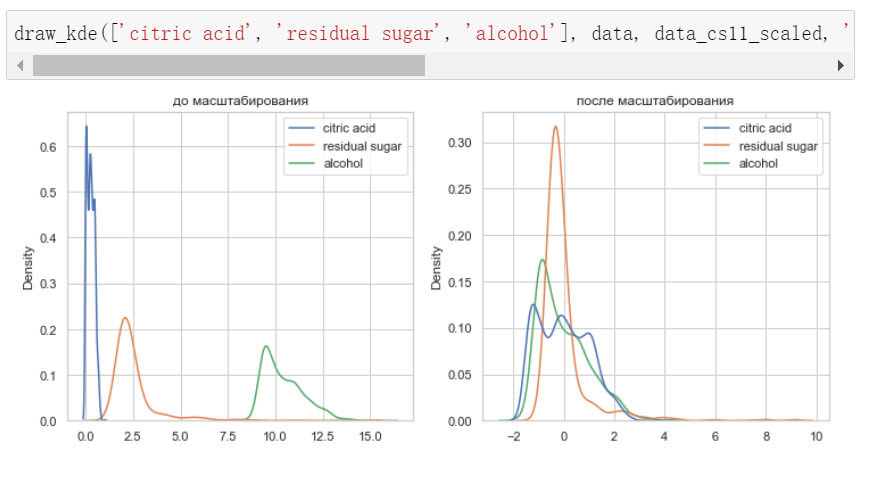


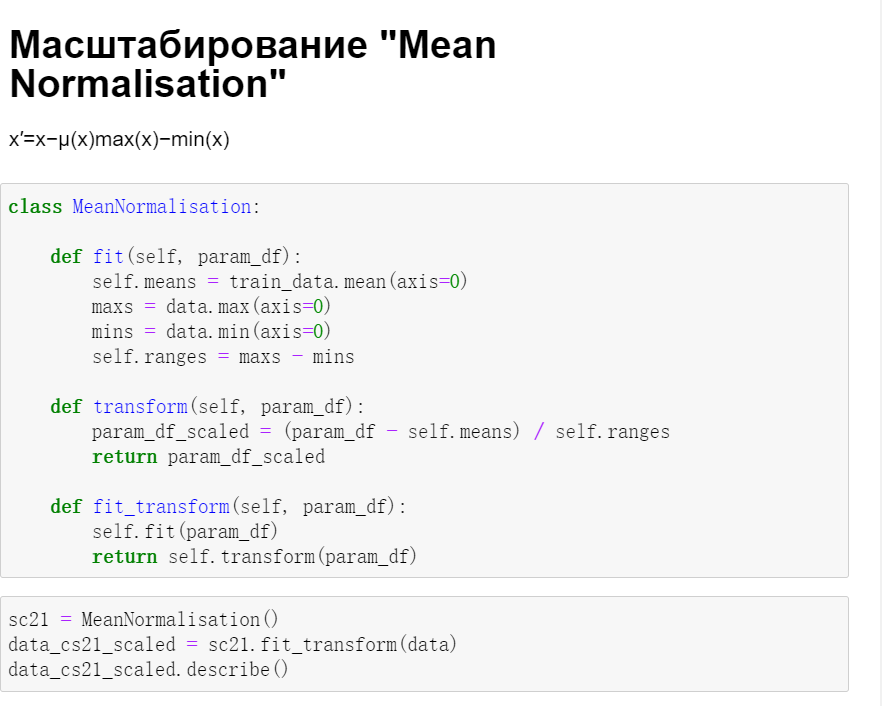
**Масштабирование**

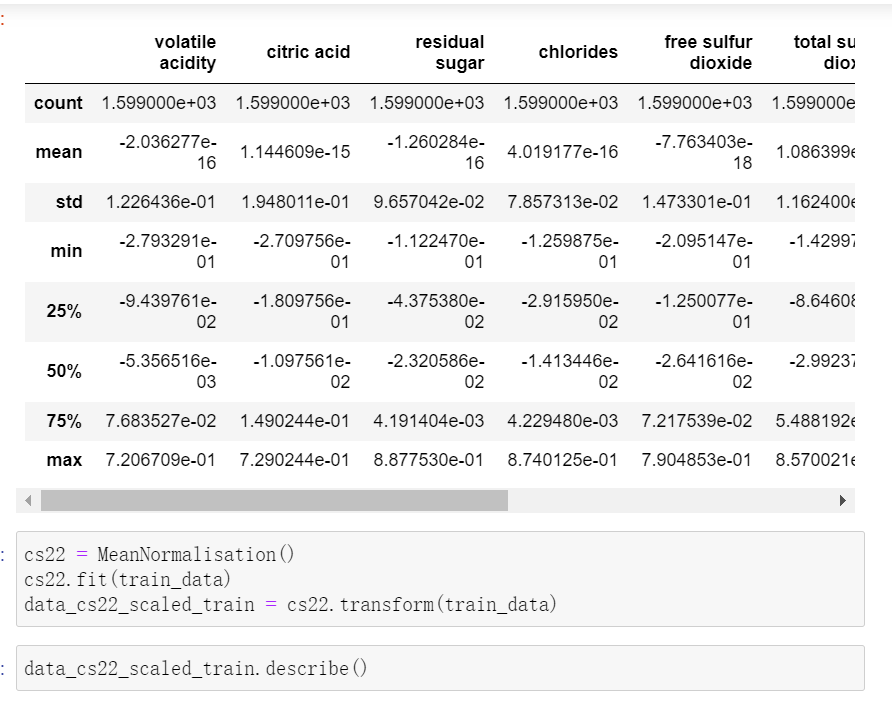


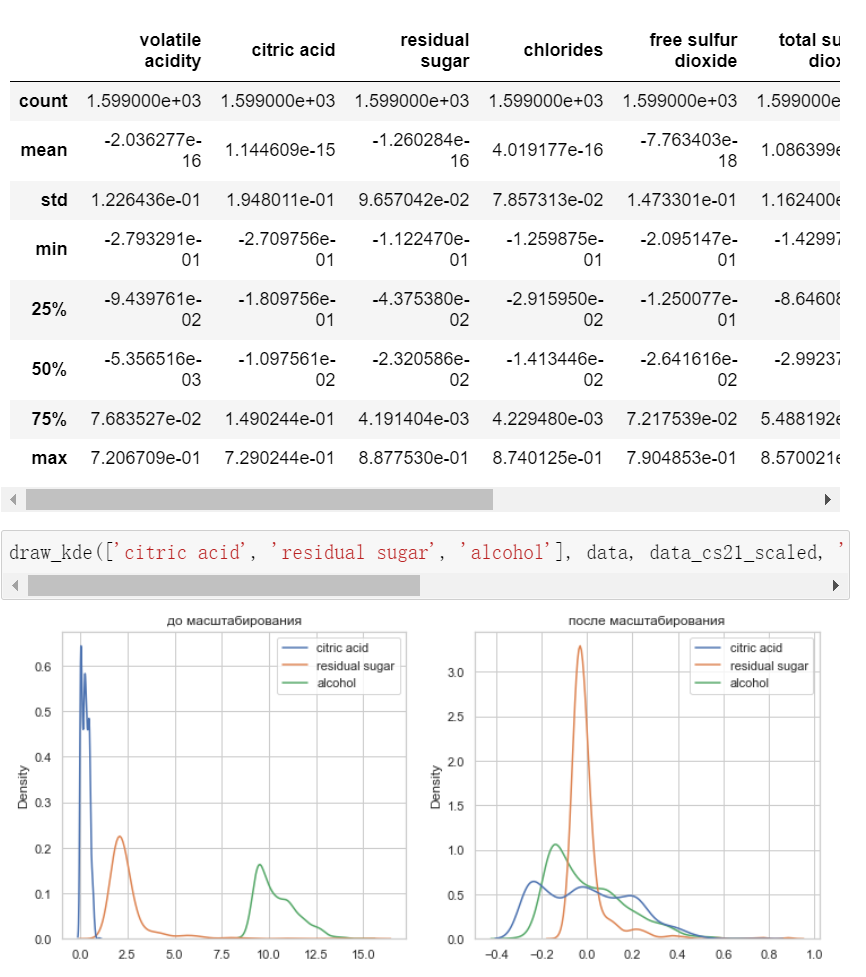




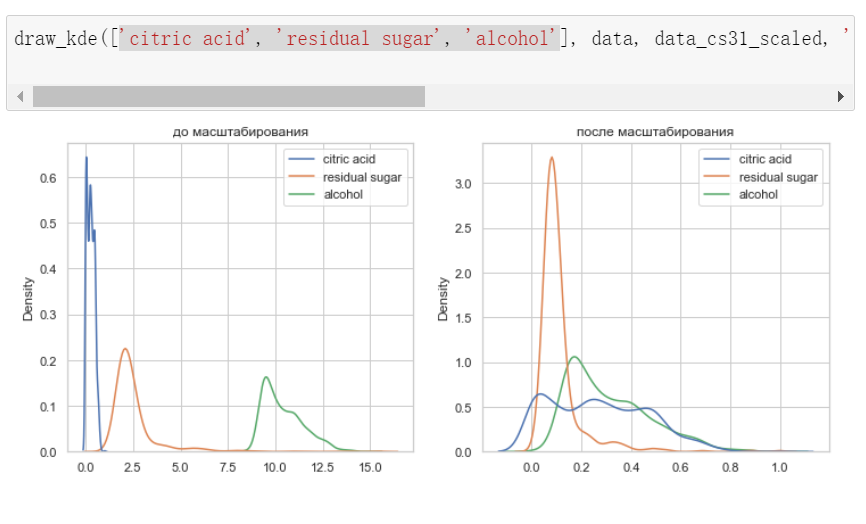






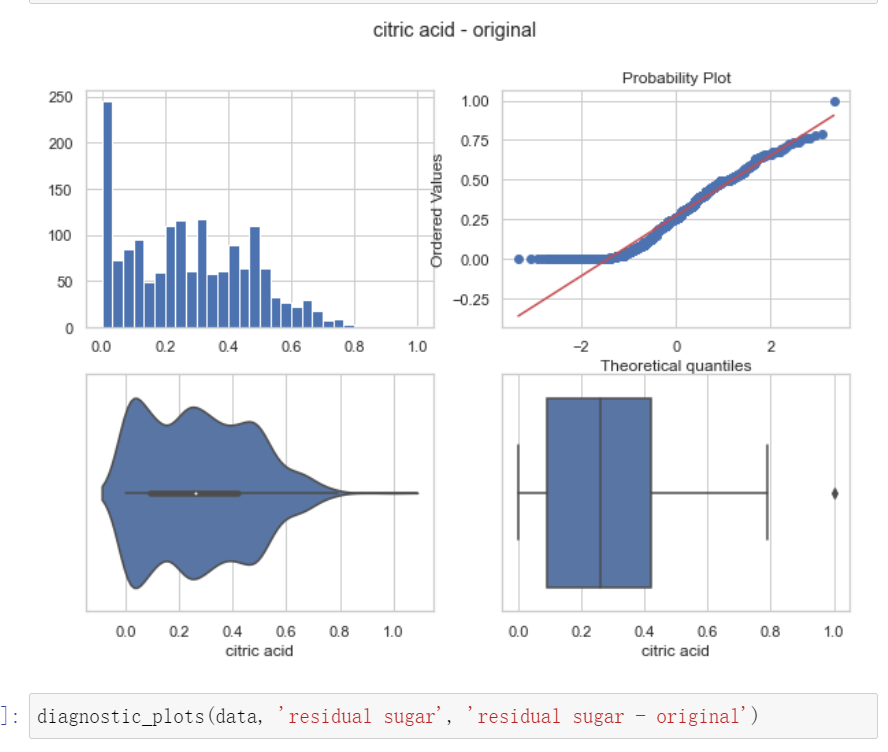


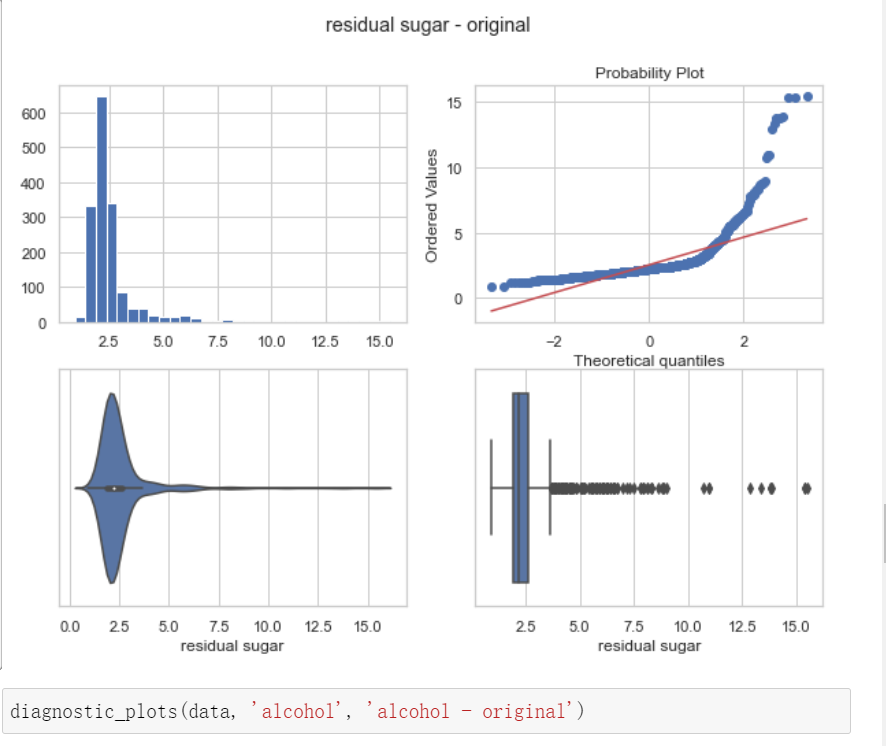


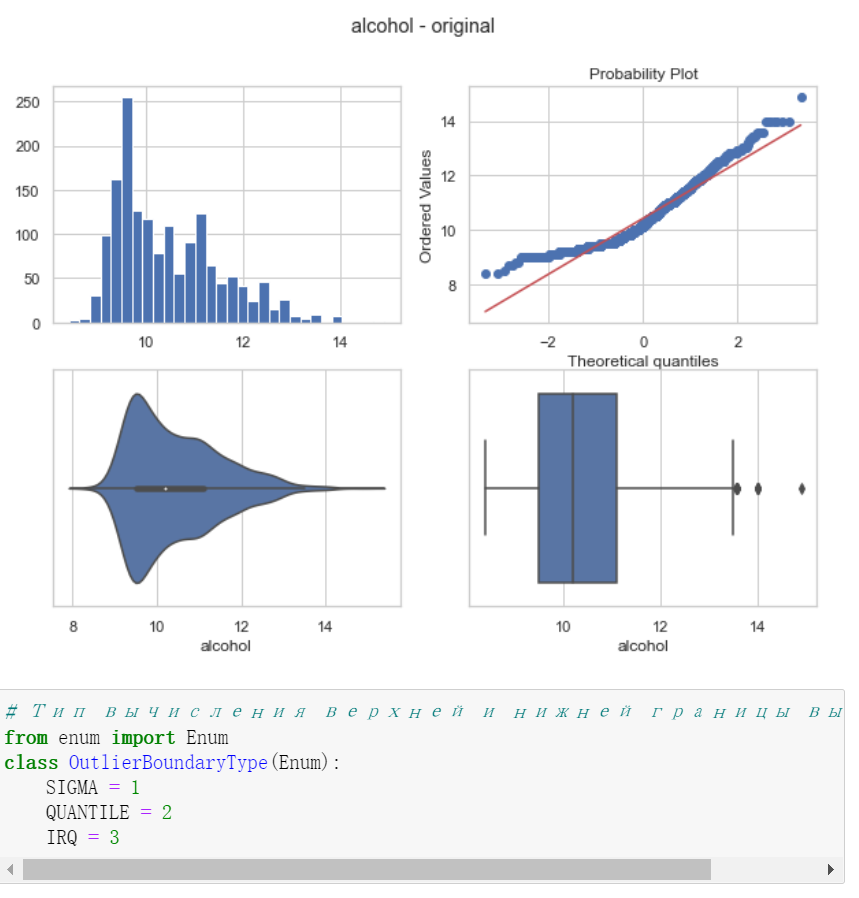


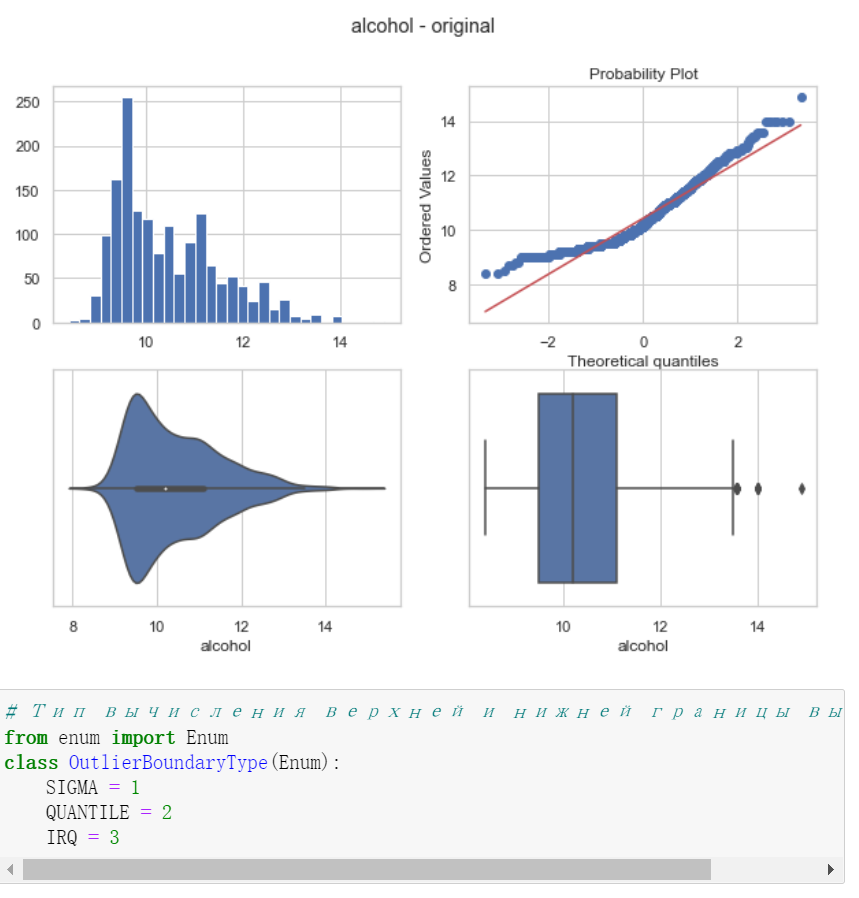
### Обработка выбросов

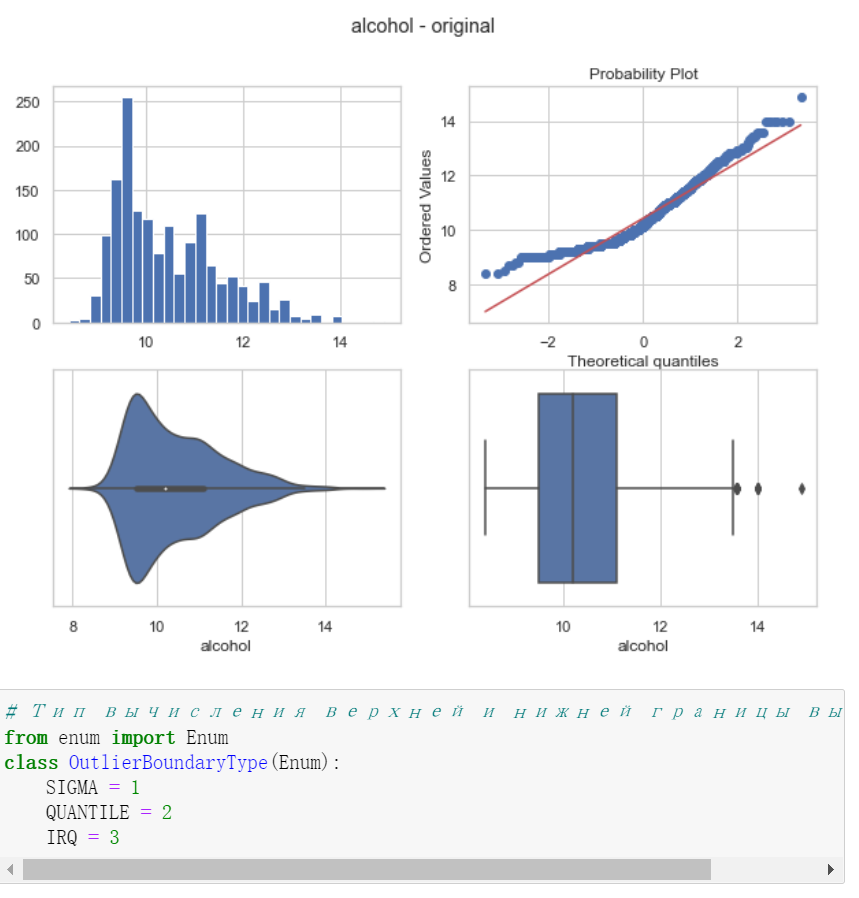


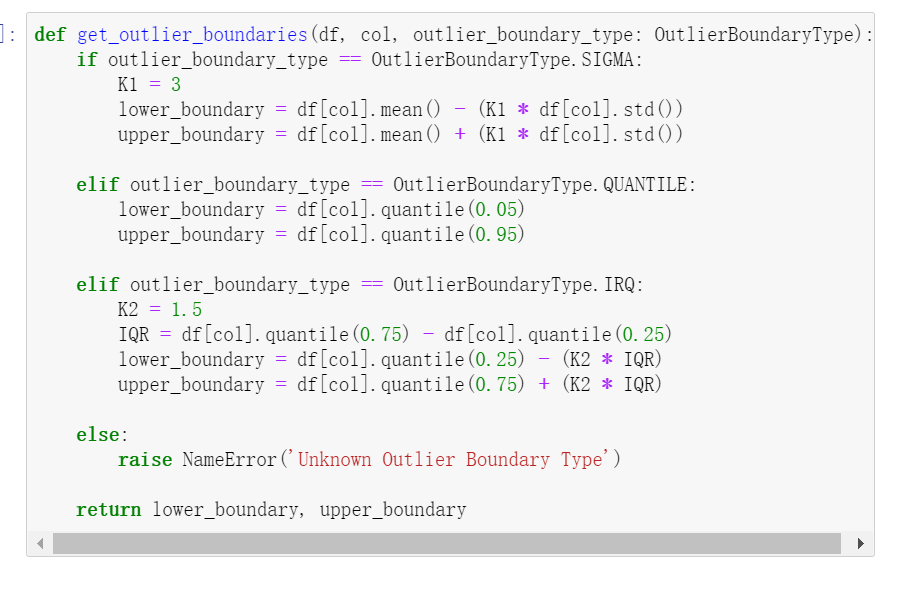


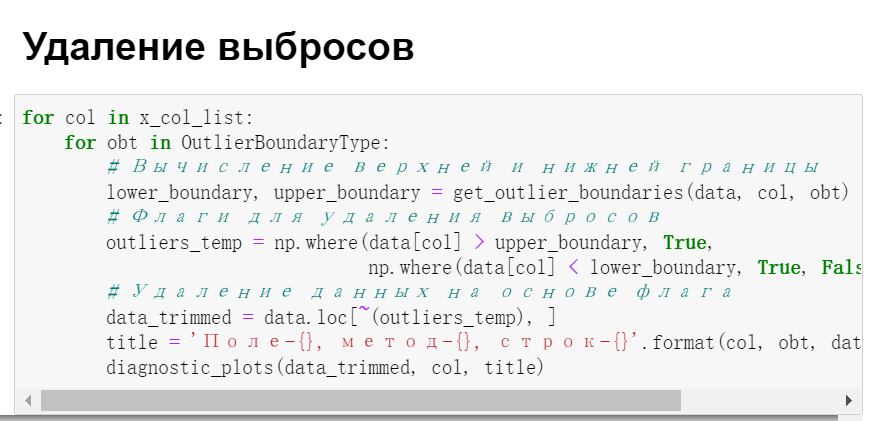


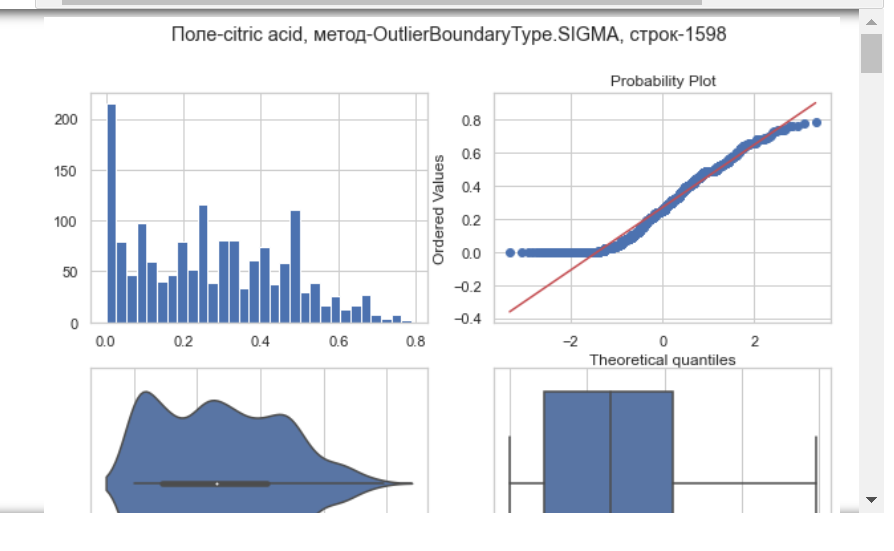














**Feature selection**

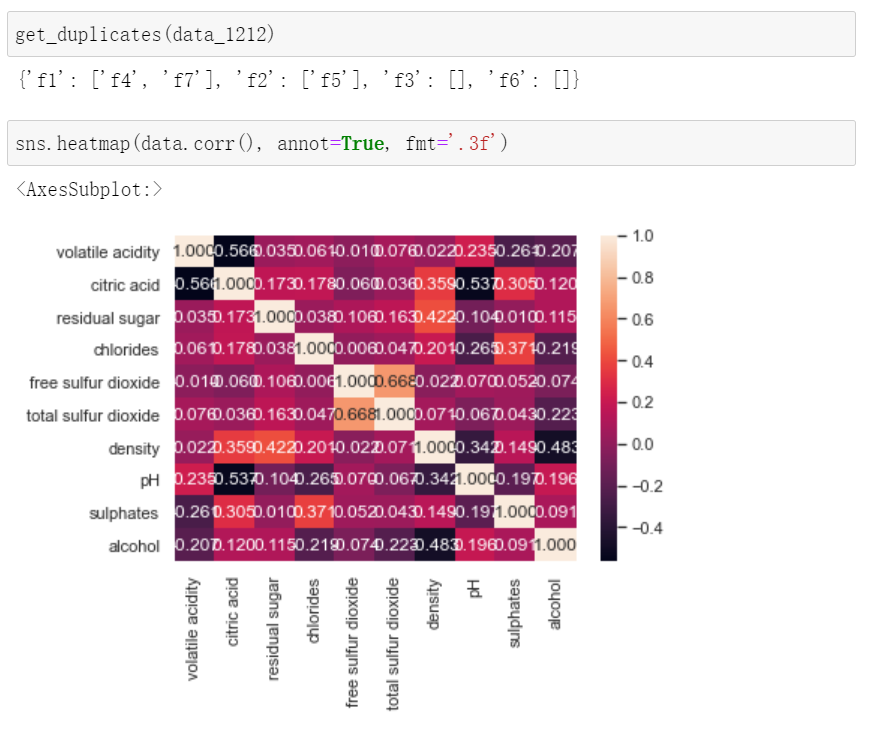






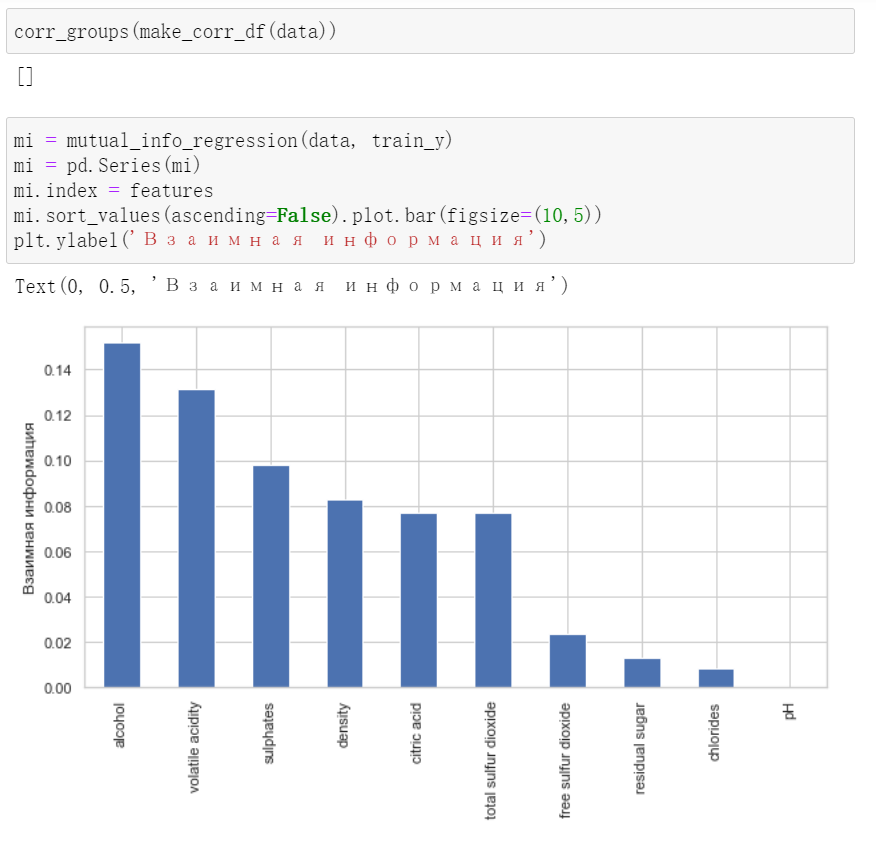














### Список литературы

## [1] Гапанюк Ю. Е. Лабораторная работа «Обработка признаков (часть2)» [Электронный ресурс] https://github.com/ugapanyuk/ml\_course\_2021/wiki/LAB\_MMO\_\_FEATURES