Федеральное государственное бюджетное   
образовательное учреждение высшего   
профессионального образования

Московский государственный технический   
университет имени Н.Э. Баумана   
(МГТУ им. Н.Э. Баумана)

**ЛАБОРАТОРНАЯ РАБОТА 1  
«Разведочный анализ данных. Исследование и визуализация данных»**

**ПО ДИСЦИПЛИНЕ «Технологии машинного обучения»**

Выполнил(а): Орлова С. М.   
 (Фамилия И.О. студента)

РТ5-61   
 (Индекс группы)

Проверил(а): Гапанюк Ю. Е.   
 (Фамилия И.О. преподавателя)

«\_\_\_» \_\_\_\_\_\_\_\_\_\_\_\_ 2020 г.

ЗАЧТЕНО / НЕ ЗАЧТЕНО \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
 (нужное выделить) (подпись)

Москва, 2020

1. Описание задания лабораторной работы

Цель лабораторной работы:  
Изучение различных методов визуализация данных.

Задание:  
Выбрать набор данных (рекомендуется использовать датасет без пропусков в данных). Создать ноутбук, который содержит следующие разделы:

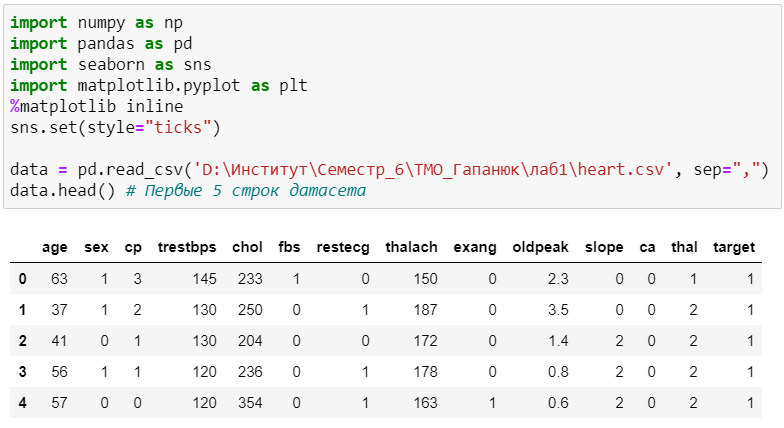
* Текстовое описание выбранного Вами набора данных.
* Основные характеристики датасета.
* Визуальное исследование датасета.
* Информация о корреляции признаков.

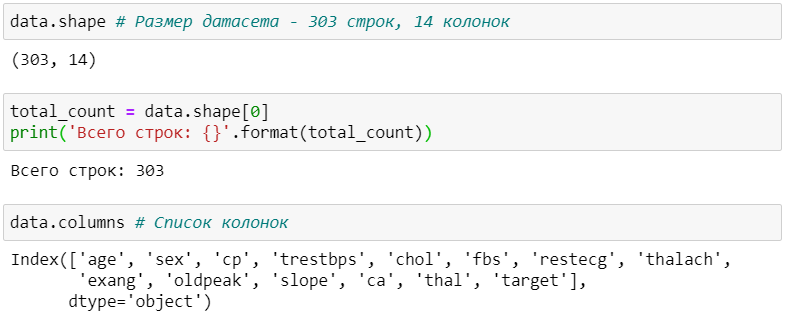
Текстовое описание набора данных:  
В качестве набора данных будут использоваться данные обследований, на основании которых выявляются болезни сердца.  
В нем содержатся следующие колонки:

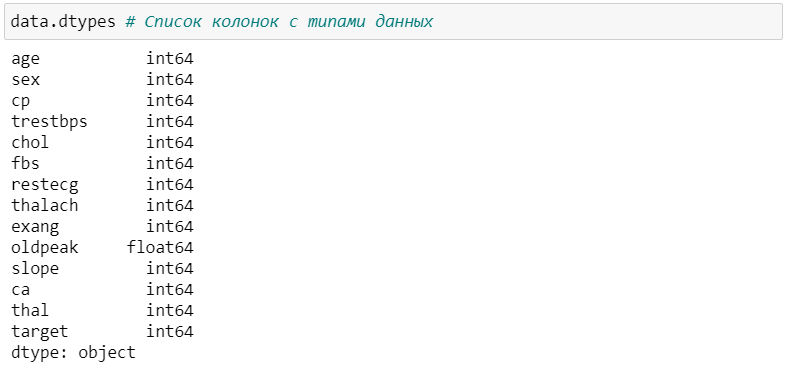
* age – возраст в годах;
* sex – пол (1 – мужчина, 0 – женщина)
* cp – тип боли в груди
* trestbps – артериальное давление в покое (в мм рт. ст. при поступлении в стационар)
* chol – сывороточный холестерин (в мг/дл)
* fbs – уровень сахара в крови натощак > 120 мг/дл (1 – да, 0 – нет)
* restecg – электрокардиографические результаты в покое
* thalach – максимальная частота сердечных сокращений
* exang – вызванная упражнениями стенокардия (1 – да, 0 – нет)
* oldpeak – депрессия ST, вызванная физической нагрузкой относительно покоя
* slope – наклон пика нагрузки ST-сегмента
* ca – количество крупных сосудов окрашенных рентгеноскопией
* thal – талидомид (3 – нормальный, 2 – исправленный дефект, 1 – реверсивный дефект)
* target – молекулярно-таргетная терапия (1 или 0)

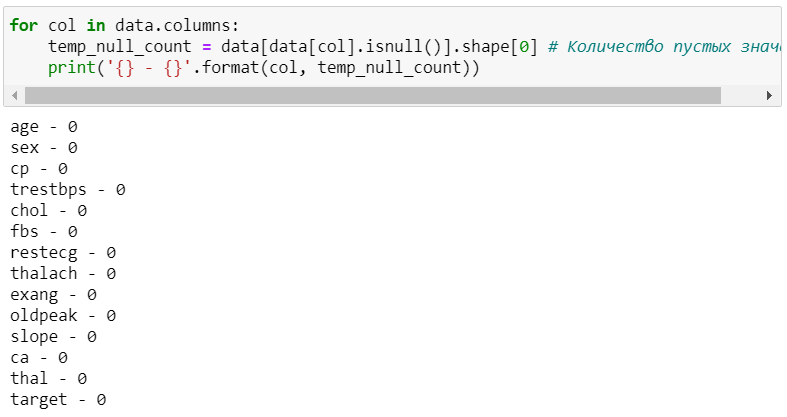
1. Текст программы и экранные формы с примерами выполнения программы

Основные характеристики датасета:

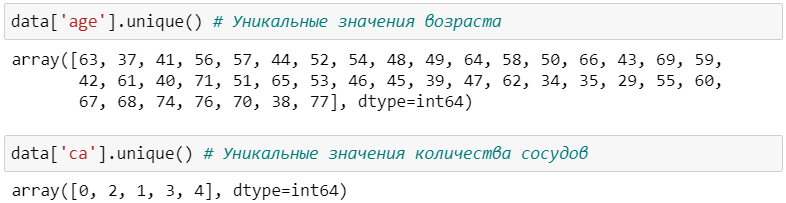




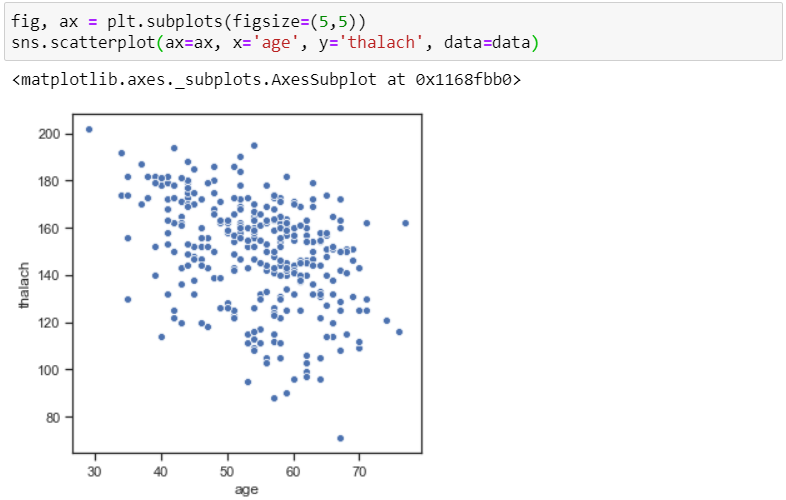


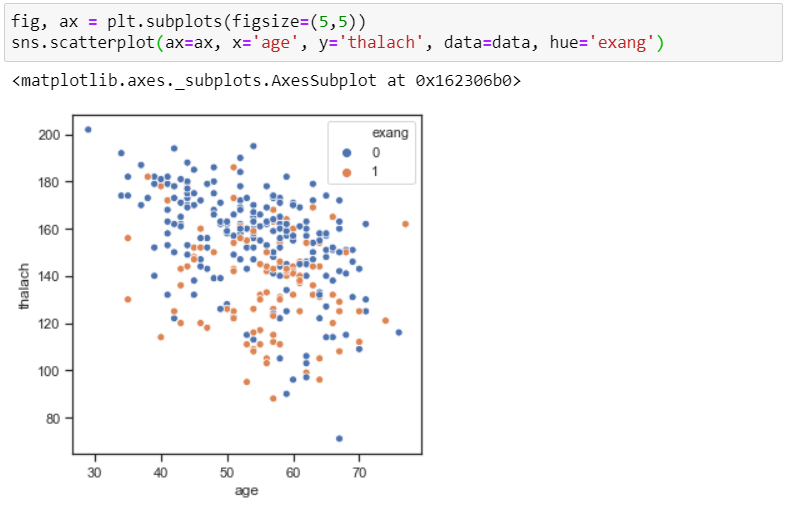


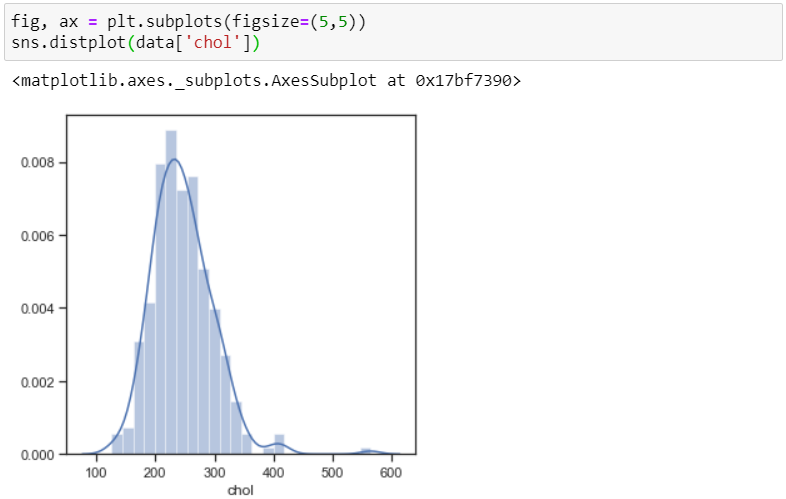


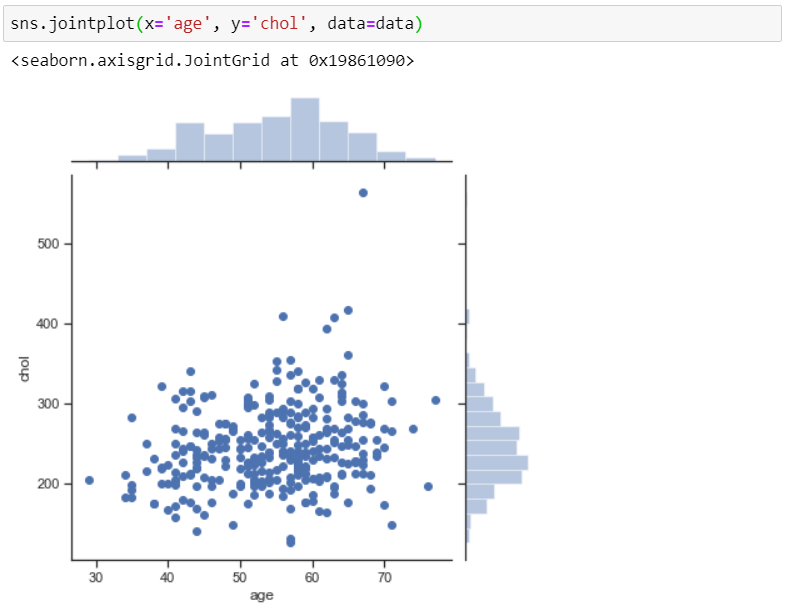


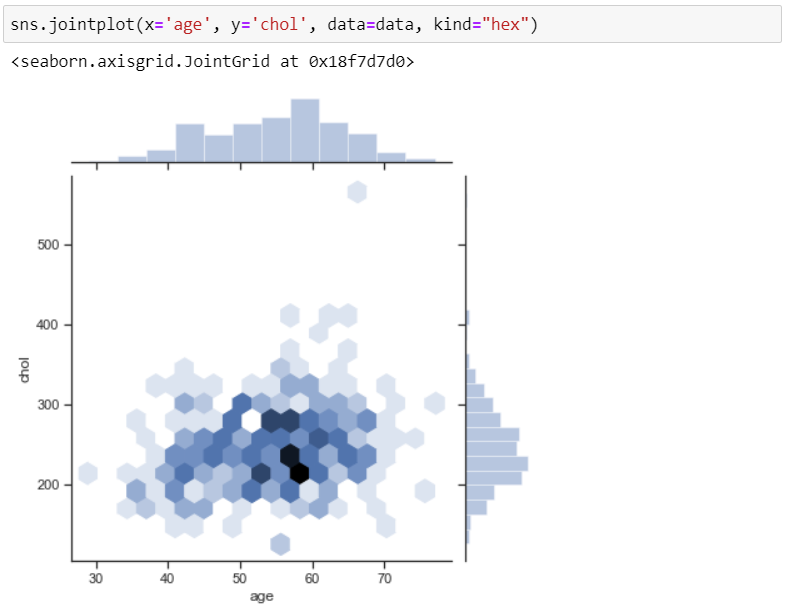
Визуальное исследование датасета:

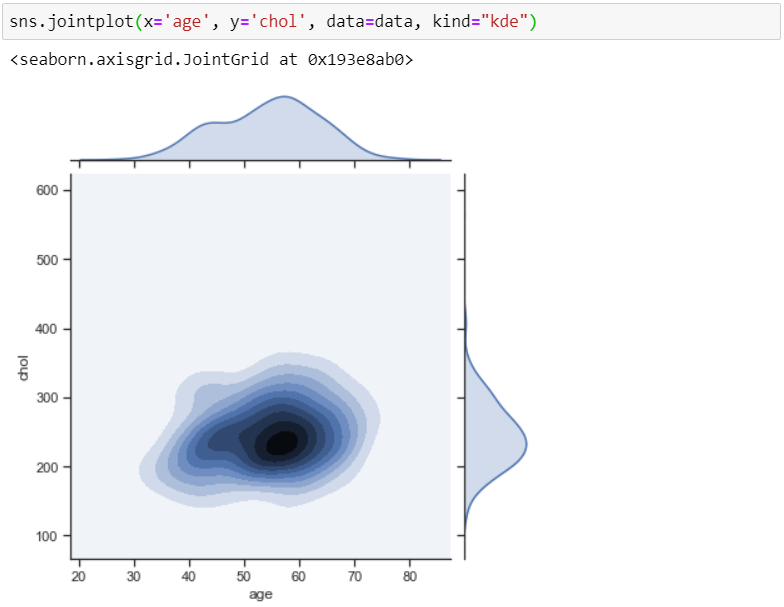


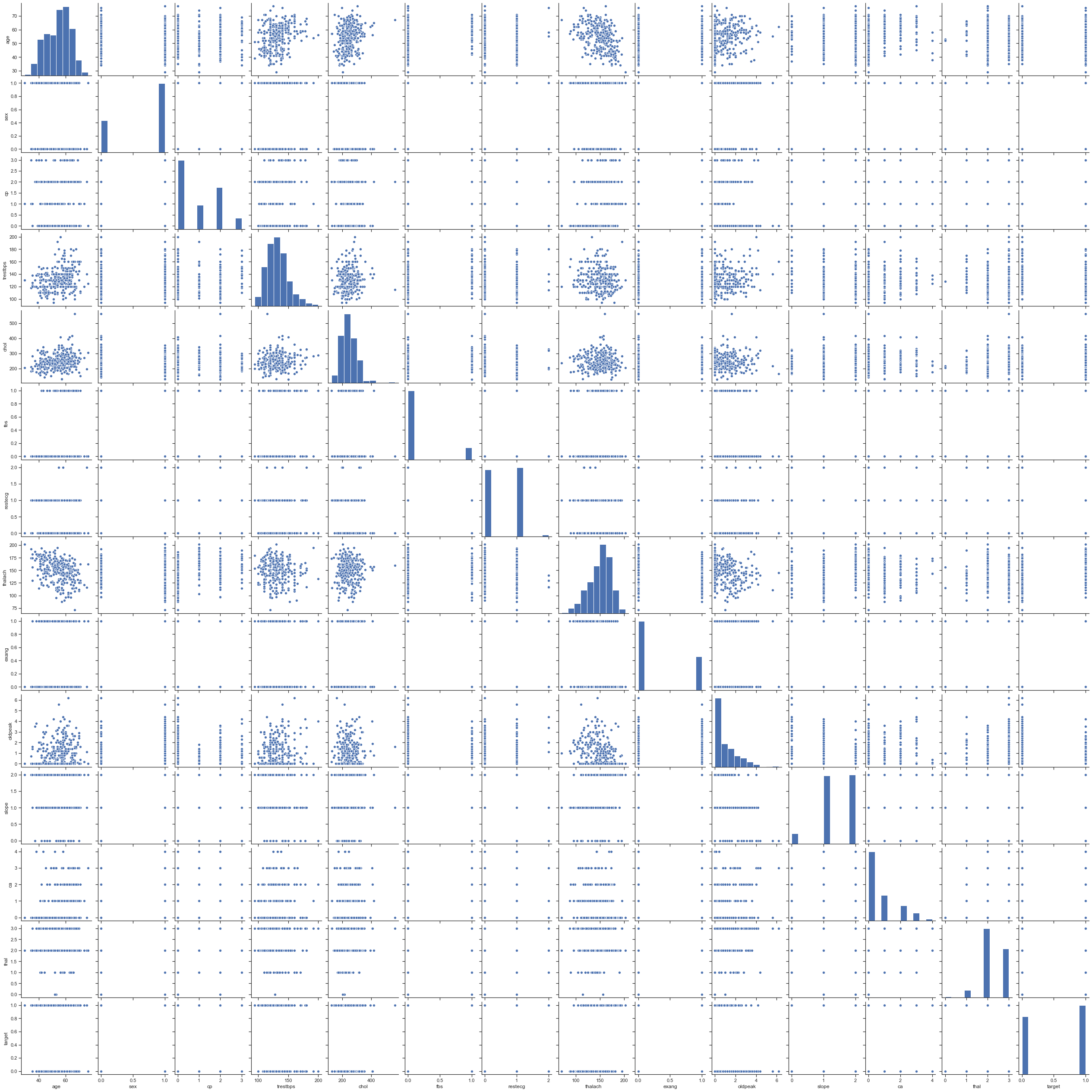


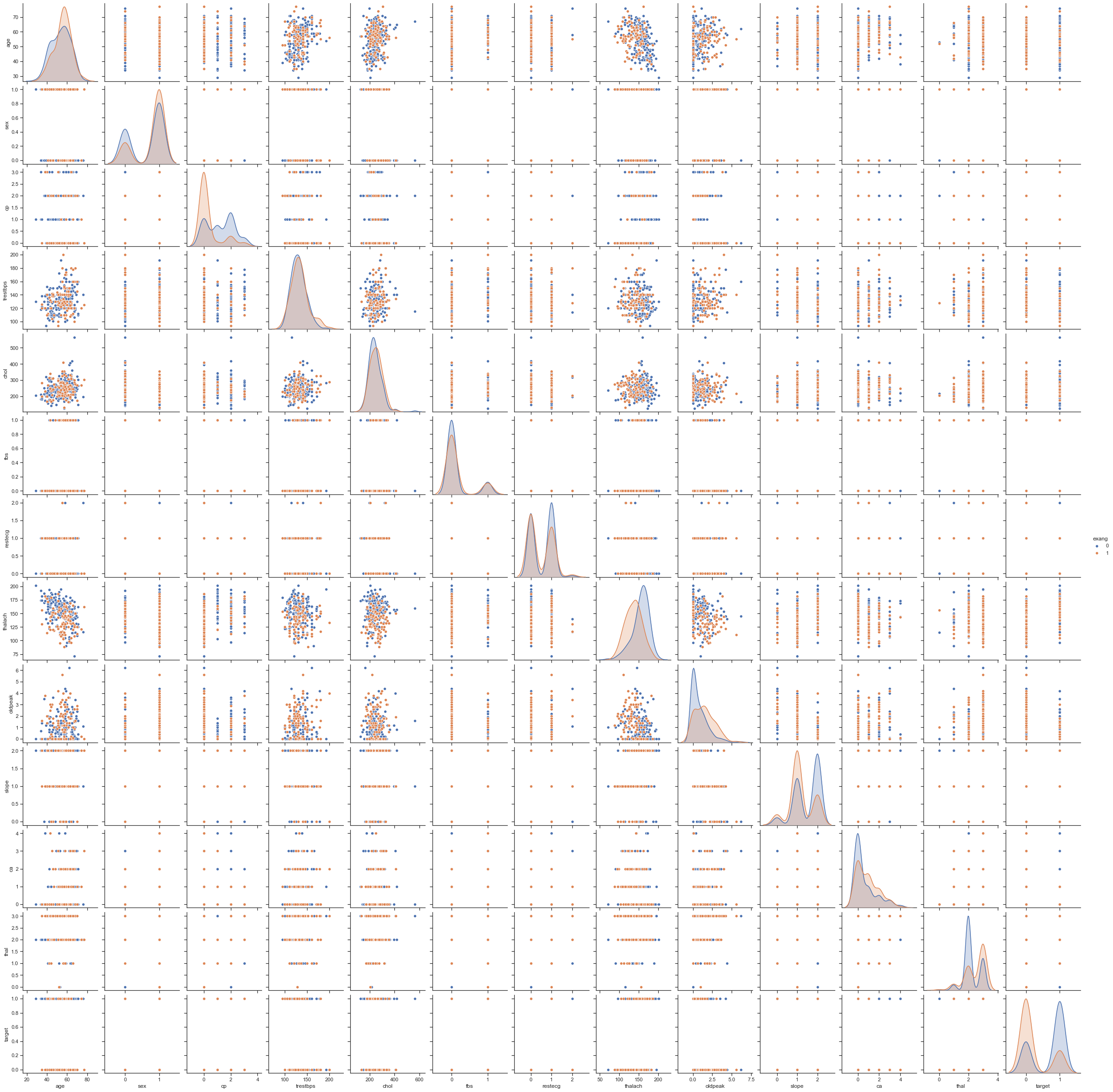


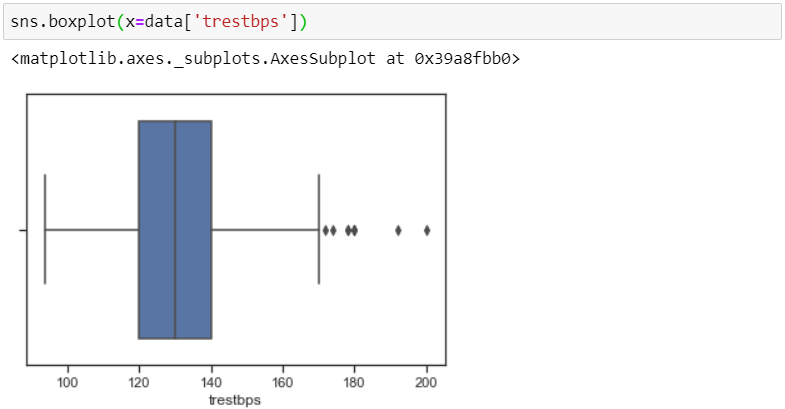


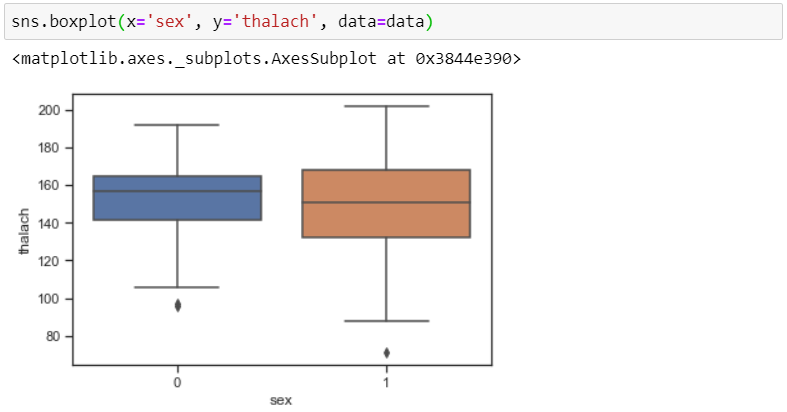


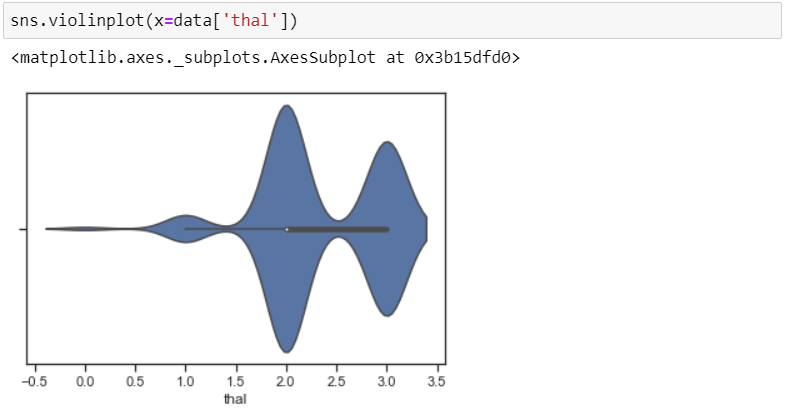


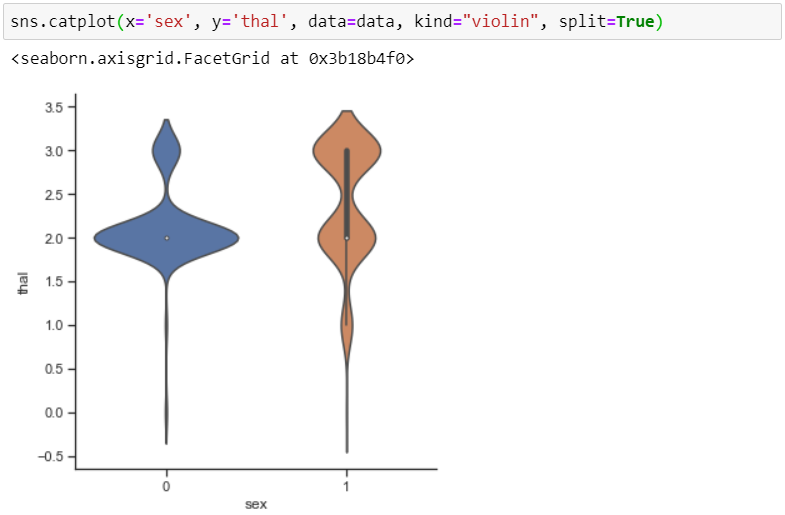
sns.pairplot(data)  


sns.pairplot(data, hue="exang")  










Информация о корреляции признаков:

