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

Курс «Технологии машинного обучения»

Отчёт по лабораторной работе №1

Выполнил: Проверил:  
Каятский П. Е. Гапанюк Ю.Е.  
группа ИУ5-64Б

Дата: 07.04.25 Дата:

Подпись: Подпись:

Москва, 2025 г.

**Цель лабораторной работы:** изучение различных методов визуализация данных.

**Краткое описание.** Построение основных графиков, входящих в этап разведочного анализа данных.

**Задание:**

* Выбрать набор данных (датасет). Вы можете найти список свободно распространяемых датасетов [здесь.](https://github.com/ugapanyuk/courses_current/wiki/DSLIST)
* Для первой лабораторной работы рекомендуется использовать датасет без пропусков в данных, например из [Scikit-learn.](https://scikit-learn.org/stable/datasets/toy_dataset.html)
* Пример преобразования датасетов Scikit-learn в Pandas Dataframe можно посмотреть [здесь.](https://github.com/ugapanyuk/courses_current/blob/main/notebooks/ds/sklearn_datasets.ipynb)

Для лабораторных работ не рекомендуется выбирать датасеты большого размера.

* Создать ноутбук, который содержит следующие разделы:

1. Текстовое описание выбранного Вами набора данных.
2. Основные характеристики датасета. (Мой датасет: <https://www.kaggle.com/datasets/muhammetvarl/laptop-price>
3. Визуальное исследование датасета.
4. Информация о корреляции признаков.

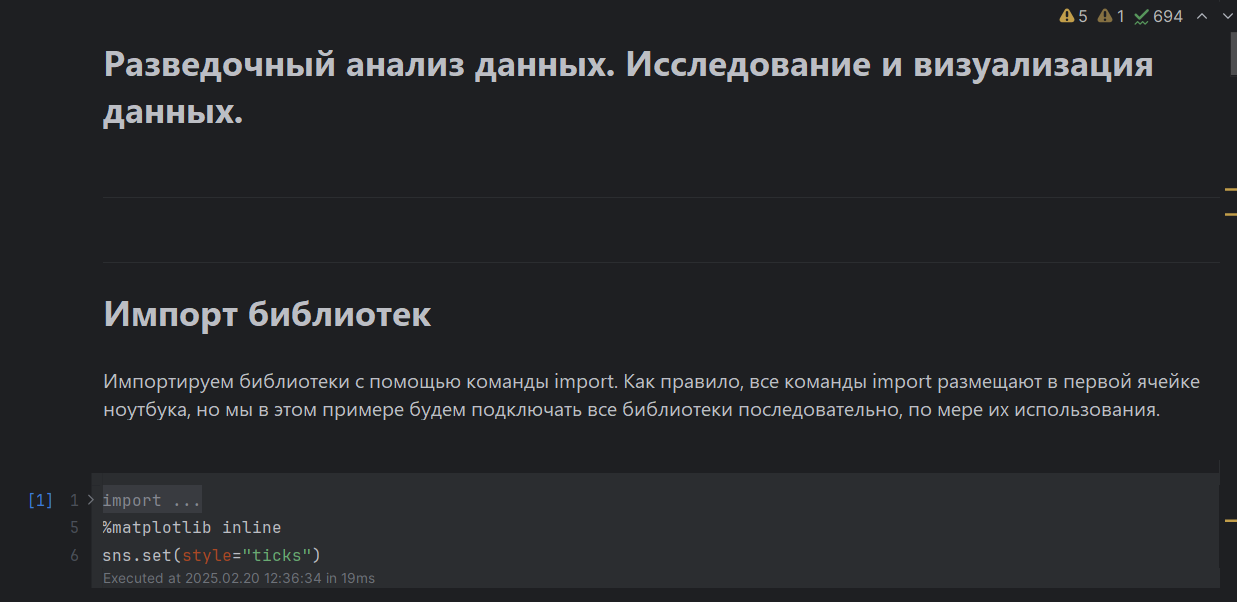
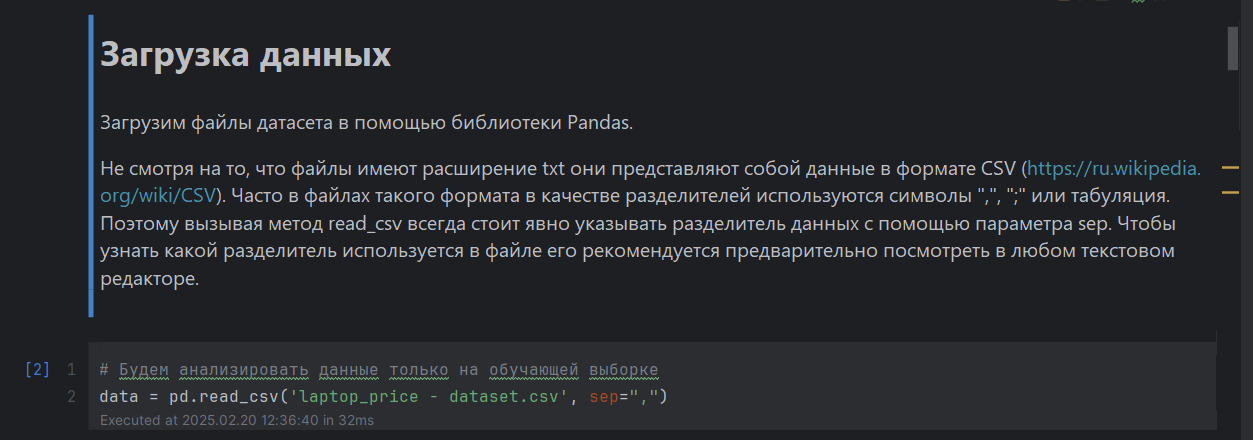
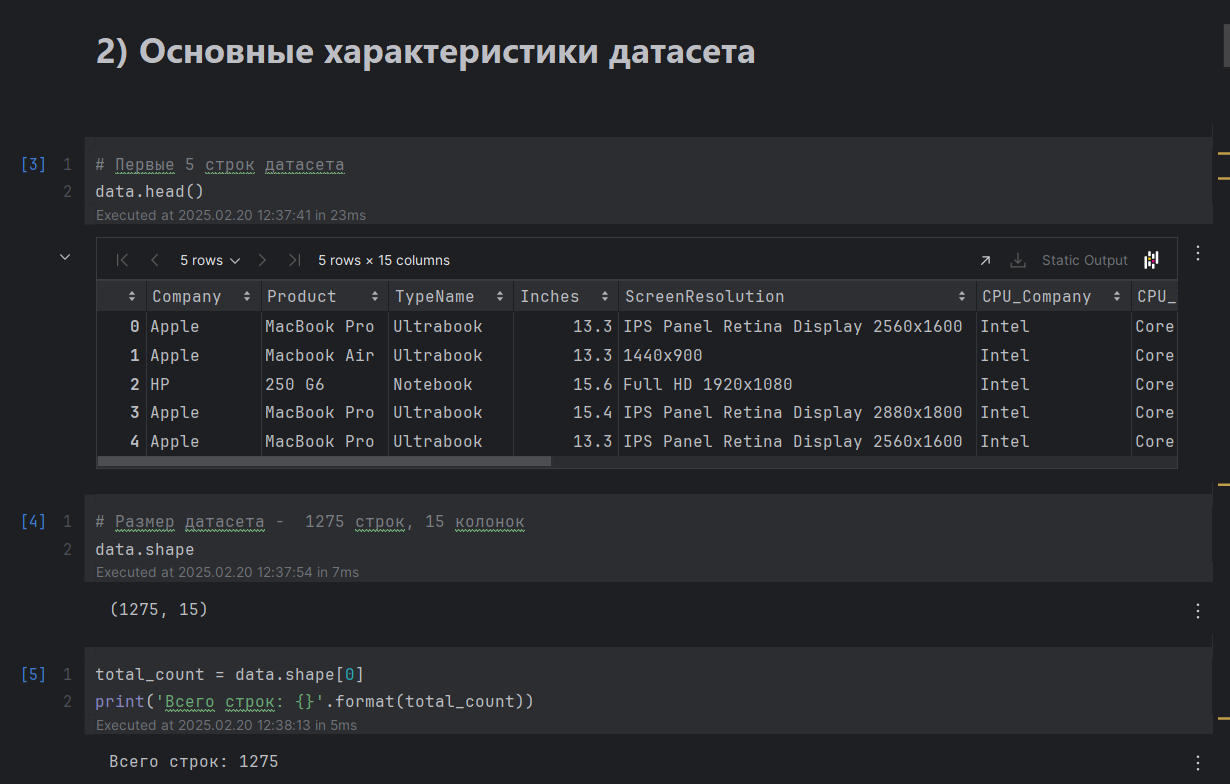
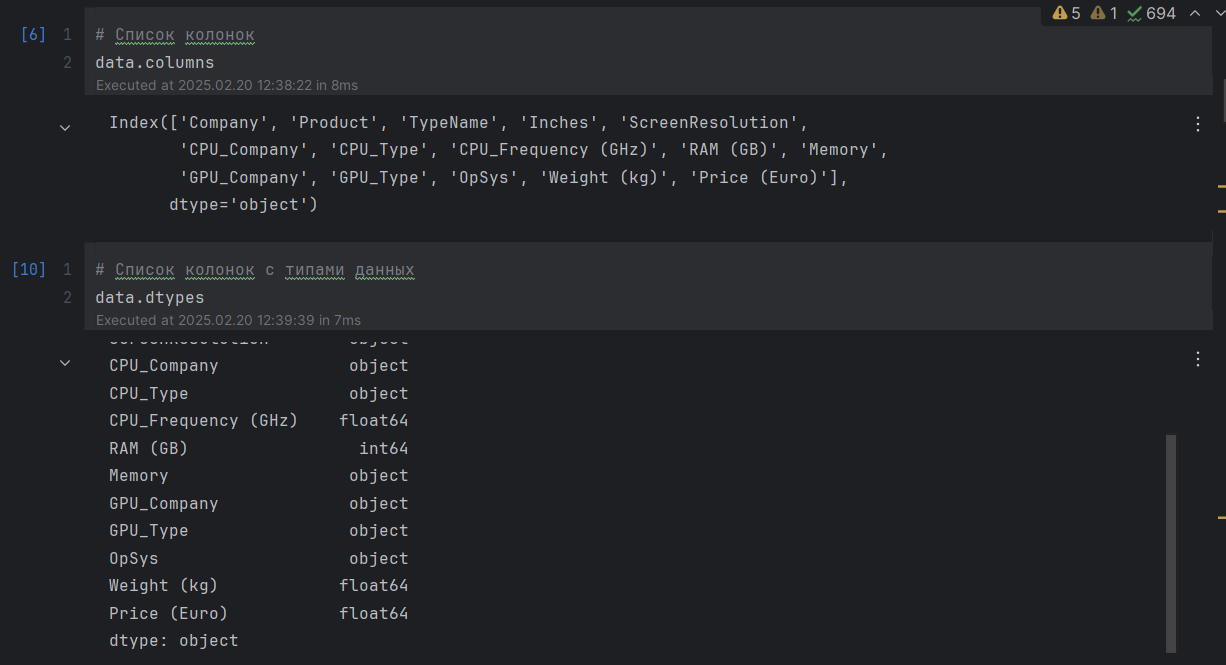
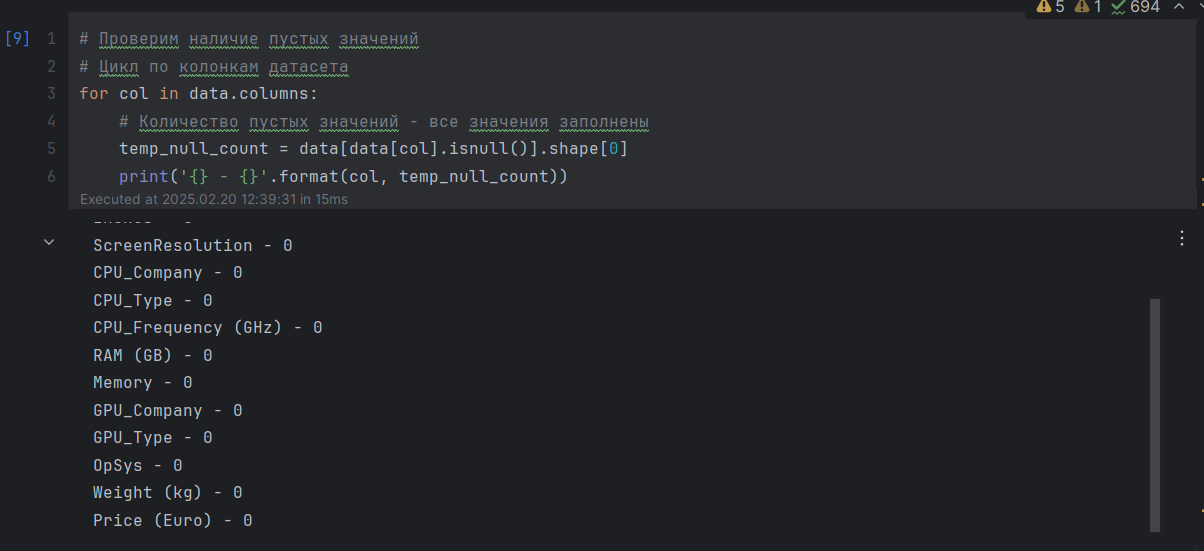
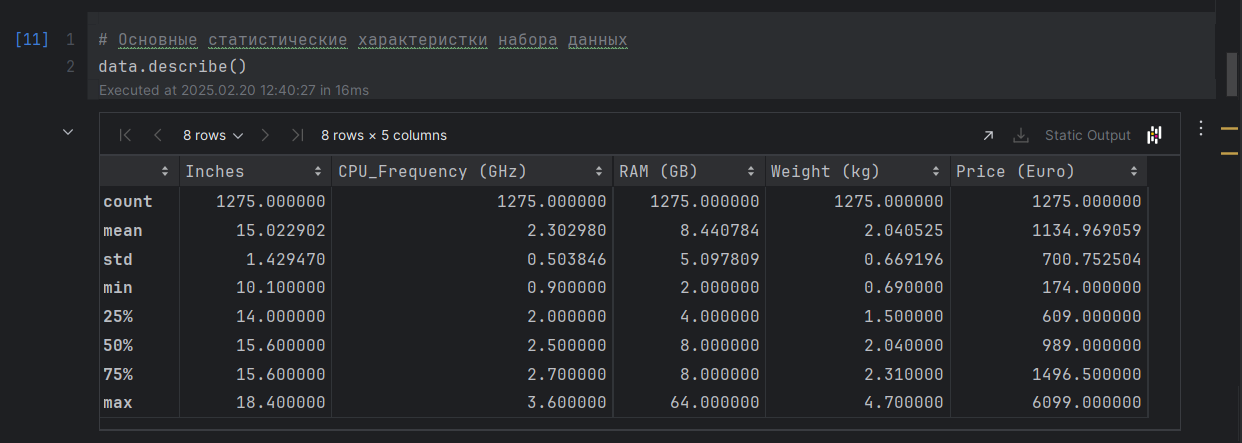
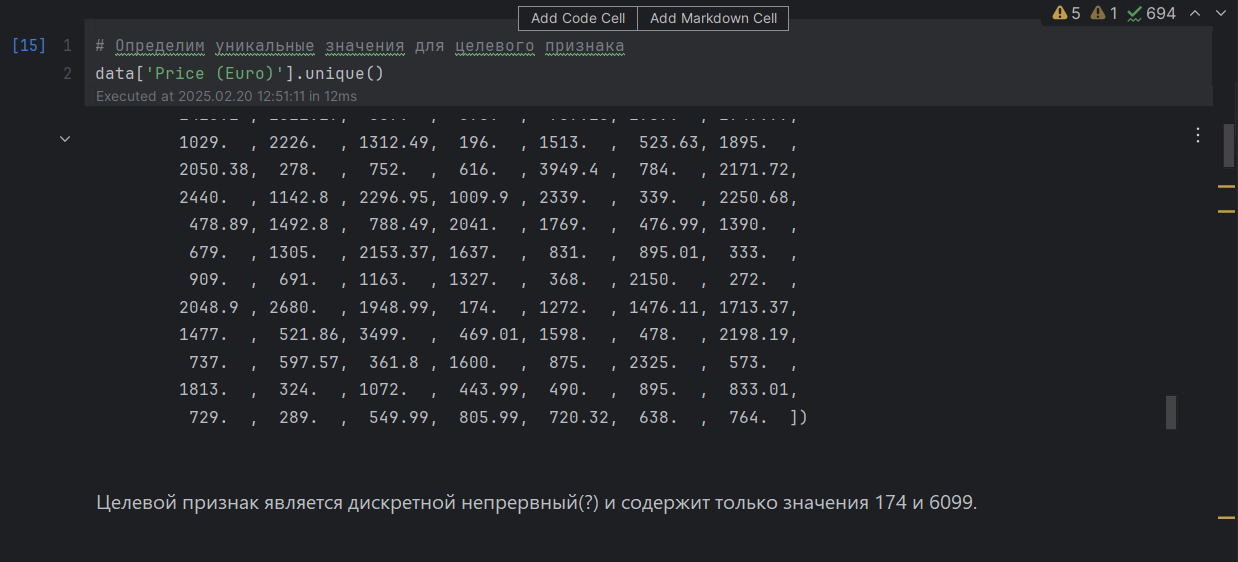
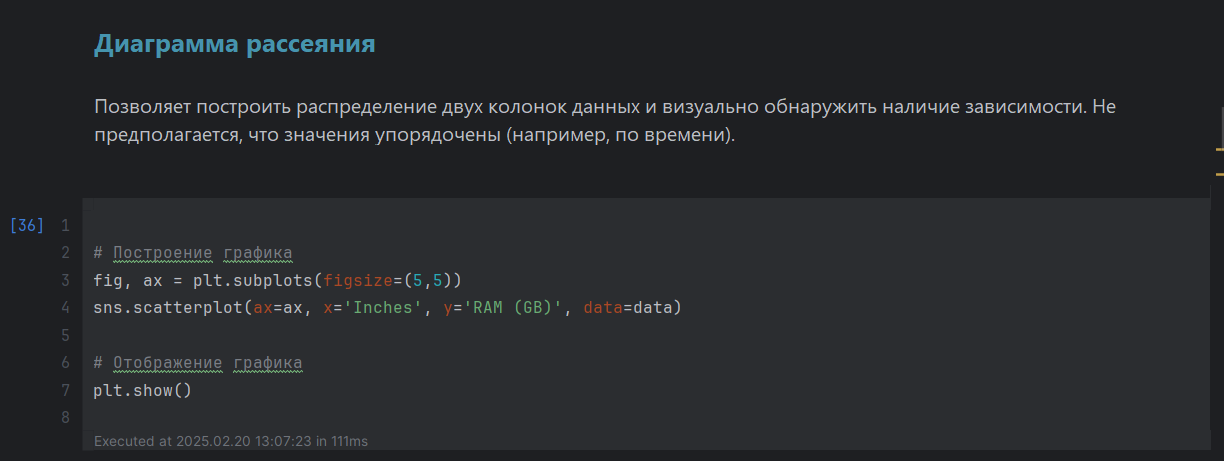
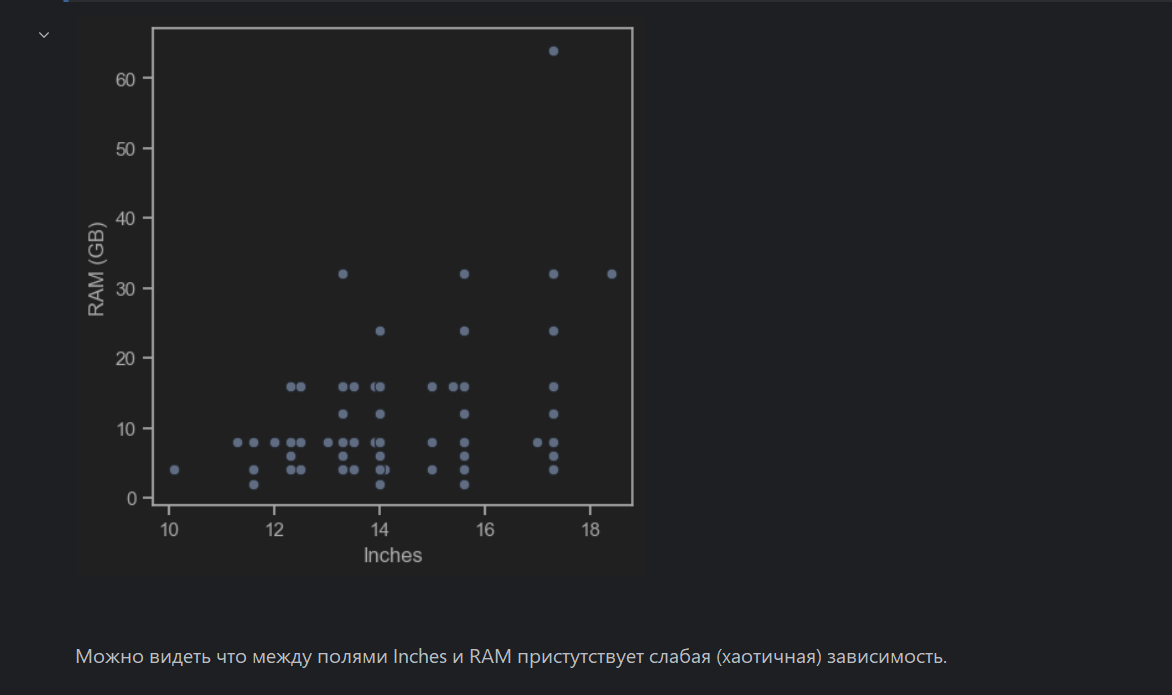
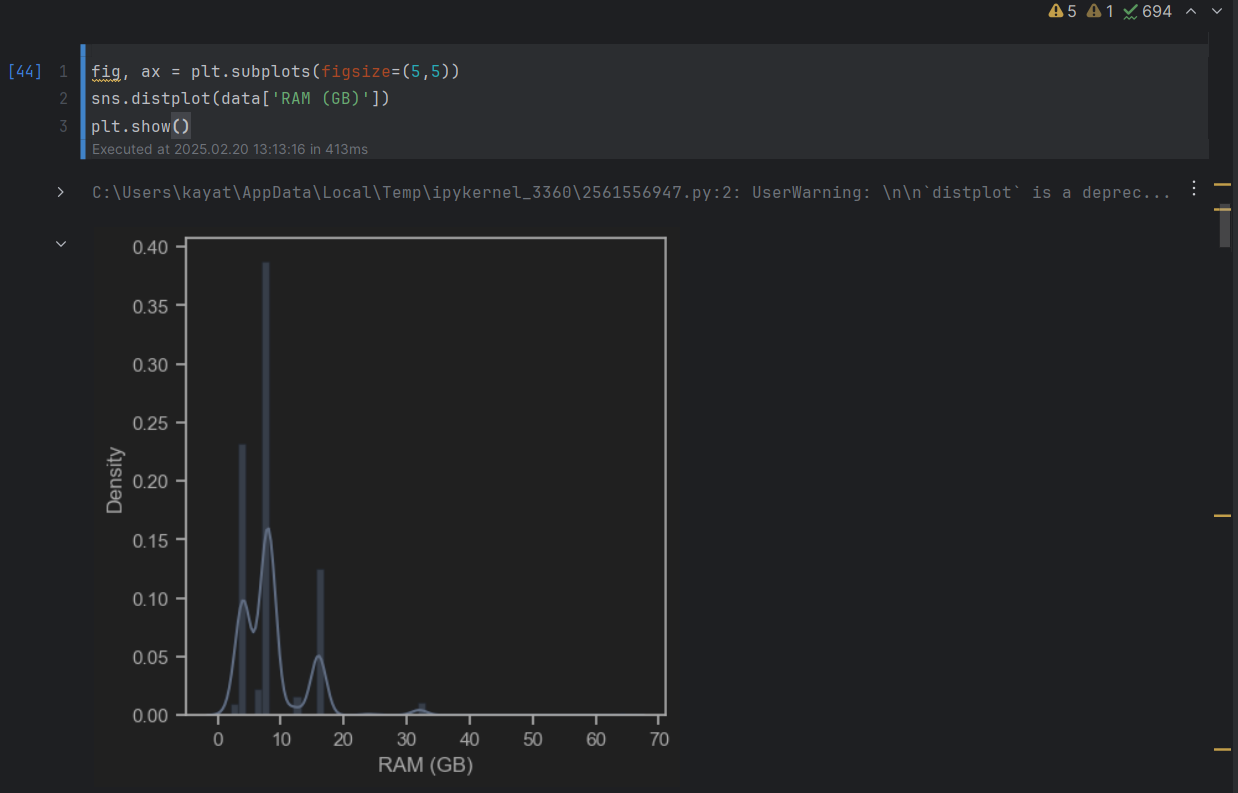
* Сформировать отчет и разместить его в своем репозитории на github.

Средства и способы визуализации данных можно посмотреть [здесь.](https://github.com/ugapanyuk/courses_current/wiki/VISUAL)

В качестве опорного примера для выполнения лабораторной работы можно использовать [пример.](https://github.com/ugapanyuk/courses_current/blob/main/notebooks/eda/eda_visualization.ipynb)

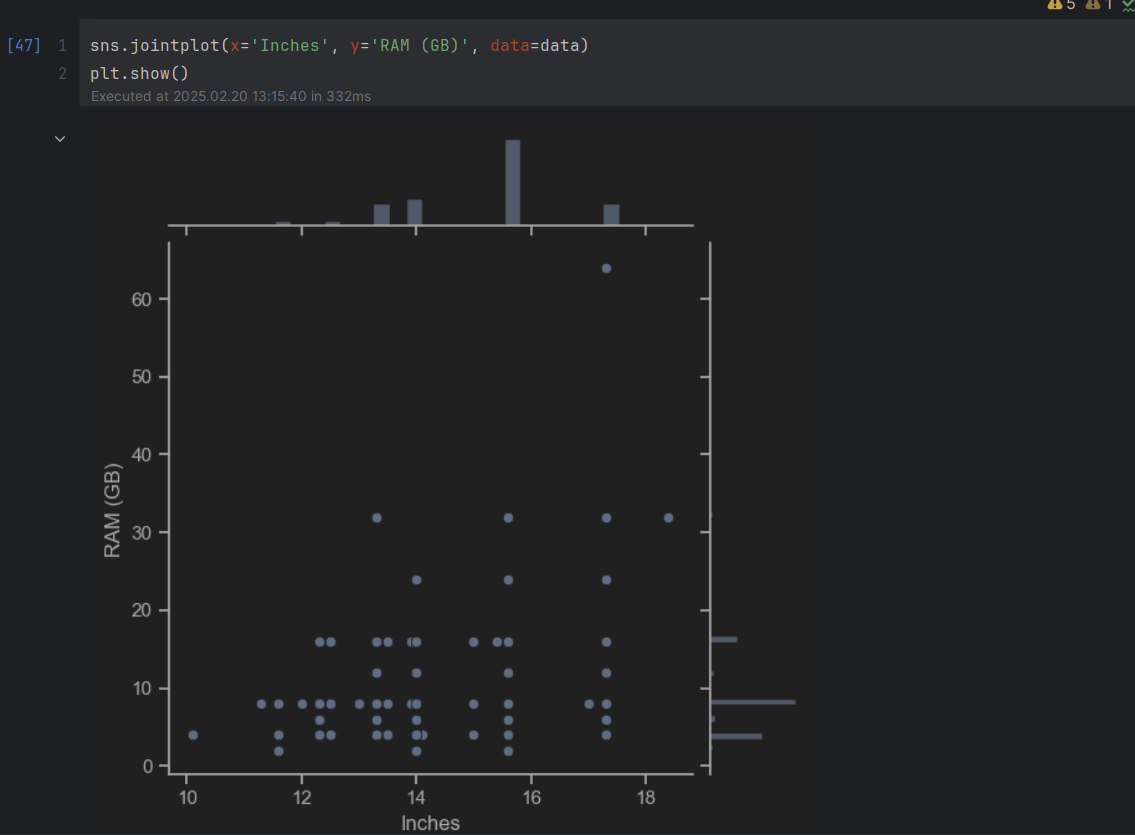
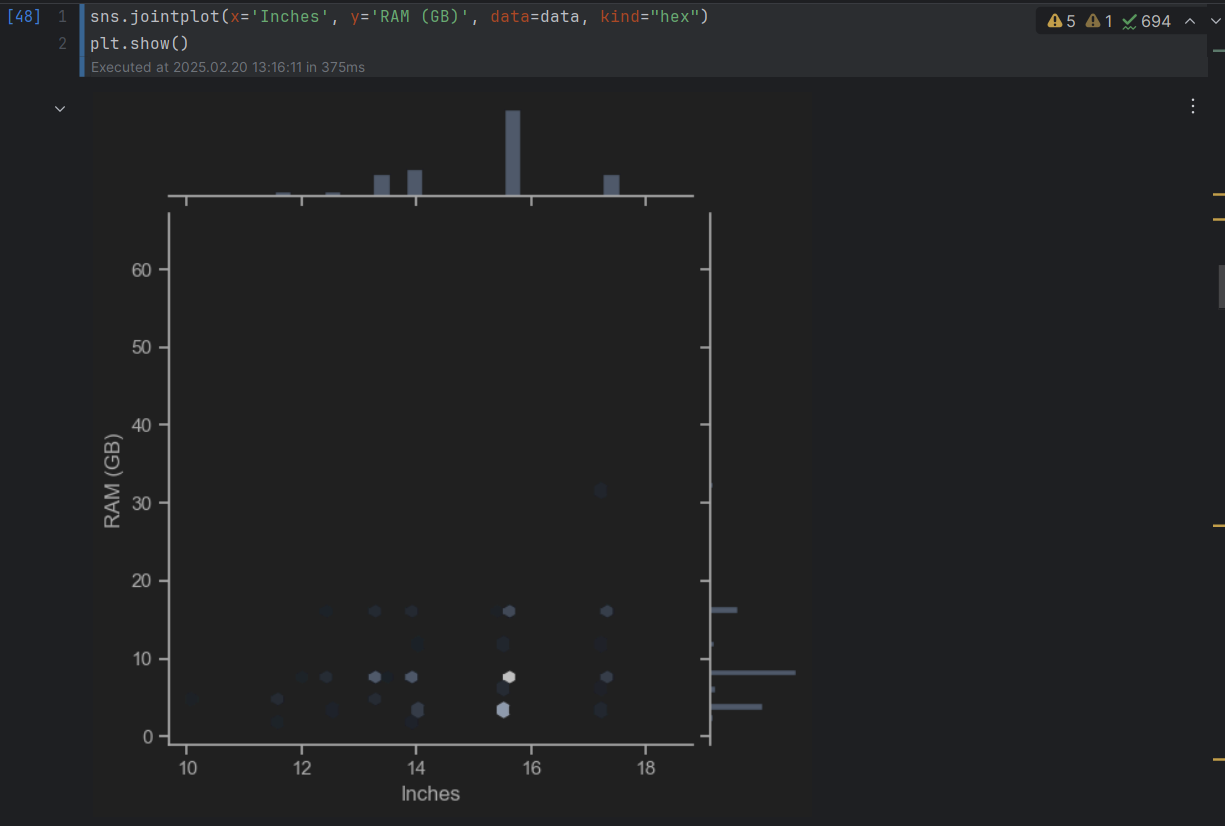
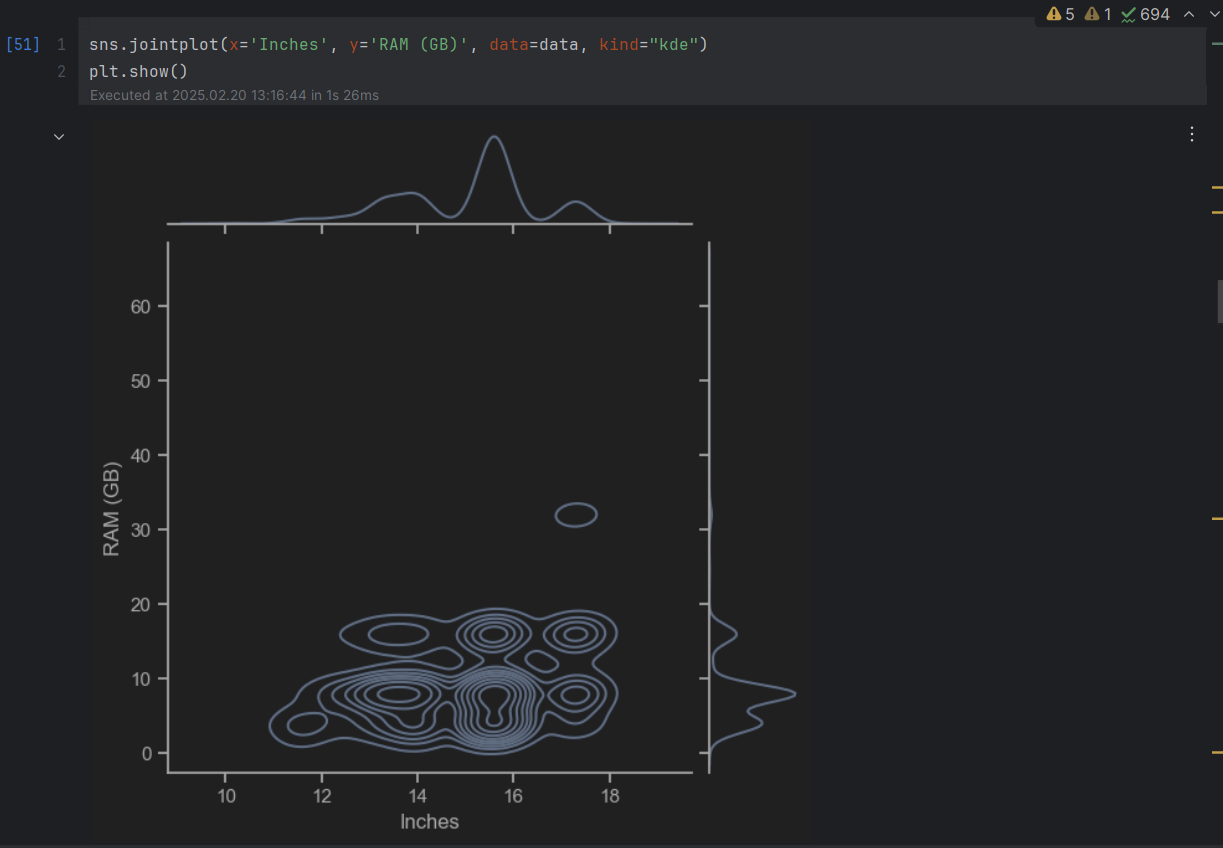
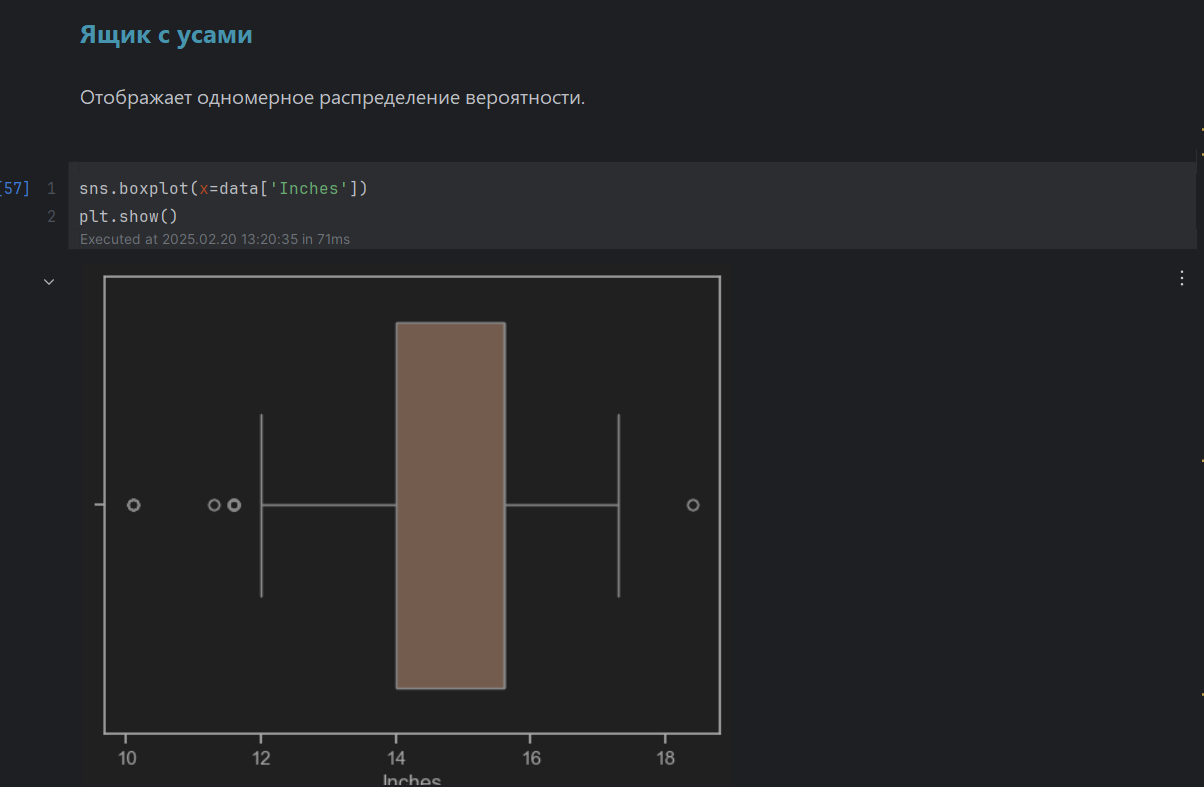
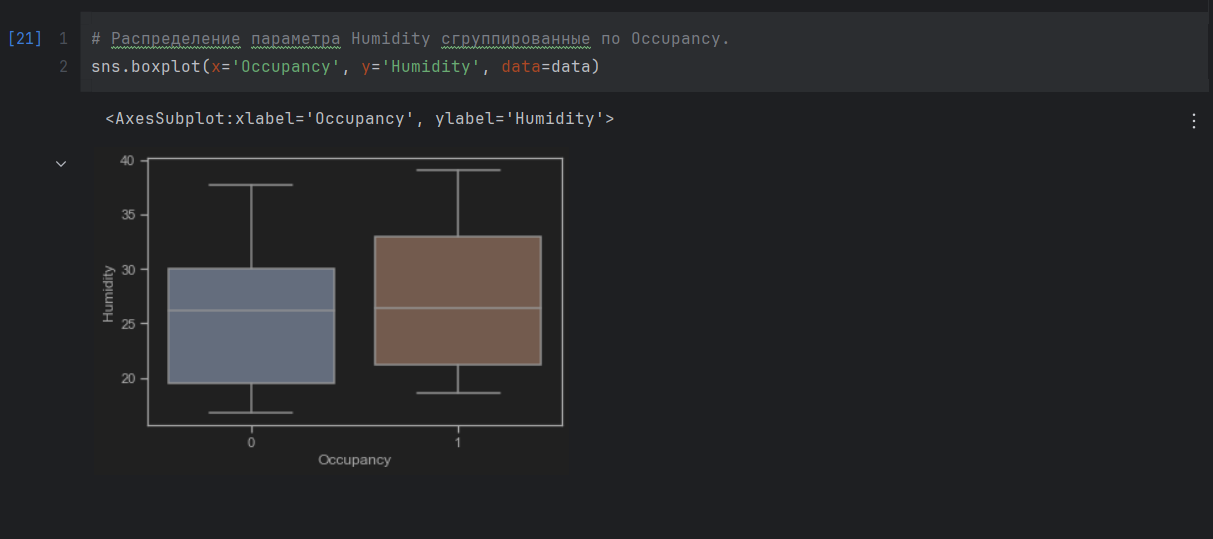
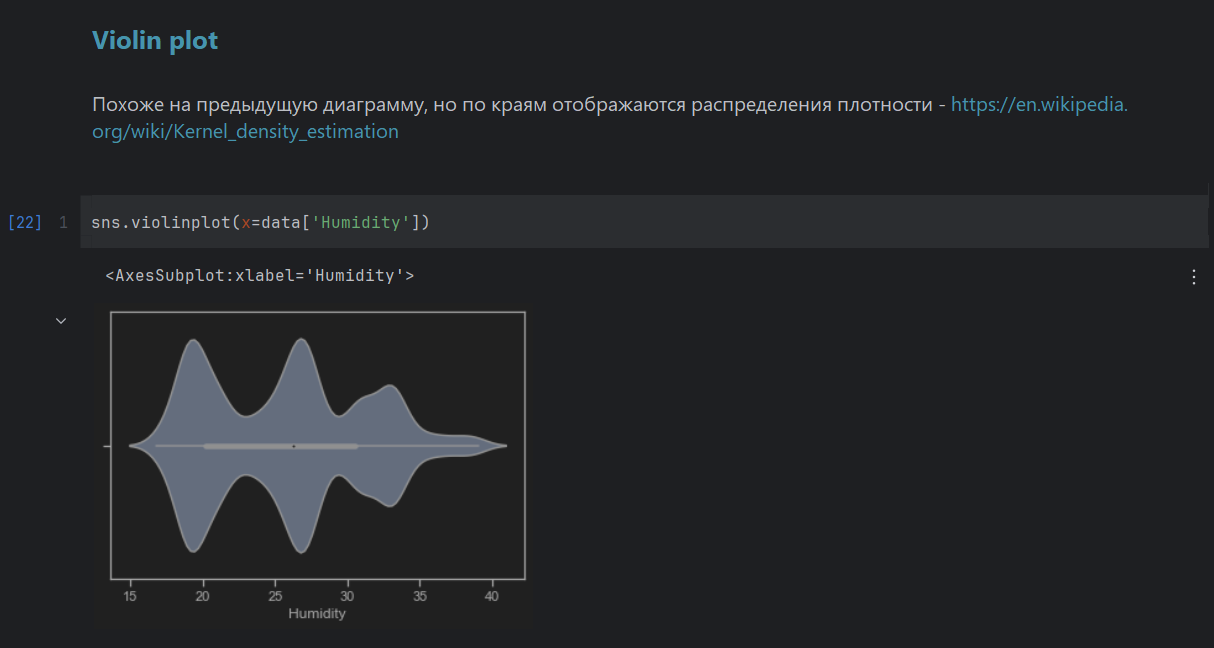
Дополнительно примеры решения задач, содержащие визуализацию, можно посмотреть в репозитории курса mlcourse.ai - <https://github.com/Yorko/mlcourse.ai/wiki/Individual-projects-and-tutorials-(in-Russian)>

**Ход выполнения:**

**** ****         

### Jointplot

Комбинация гистограмм и диаграмм рассеивания.

В случае большого количества признаков анализ числовой корреляционной матрицы становится неудобен.

Для визуализации корреляционной матрицы будем использовать "тепловую карту" heatmap которая показывает степень корреляции различными цветами.

Используем метод heatmap библиотеки seaborn - <https://seaborn.pydata.org/generated/seaborn.heatmap.html>

