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

Курс «Технологии машинного обучения»

Отчёт по рубежному контролю №1

«Технологии разведочного анализа и обработки данных.»

Вариант № 24

Выполнил: Проверил:  
Тихонова Д.Д. Гапанюк Ю.Е.  
группа ИУ5Ц-84Б

Дата: 12.03.25 Дата:

Подпись: Подпись:

Москва, 2025 г.

**Задание:**

Номер варианта: **24**

Номер задачи: **3**

Номер набора данных, указанного в задаче: **8** (https://www.kaggle.com/datasets/altavish/boston-housing-dataset)

Для студентов групп ИУ5-64Б, ИУ5Ц-84Б - для произвольной колонки данных построить график "Скрипичная диаграмма (violin plot).

**Задача №3.**

Для заданного набора данных произведите масштабирование данных (для одного признака) и преобразование категориальных признаков в количественные двумя способами (label encoding, one hot encoding) для одного признака. Какие методы Вы использовали для решения задачи и почему?

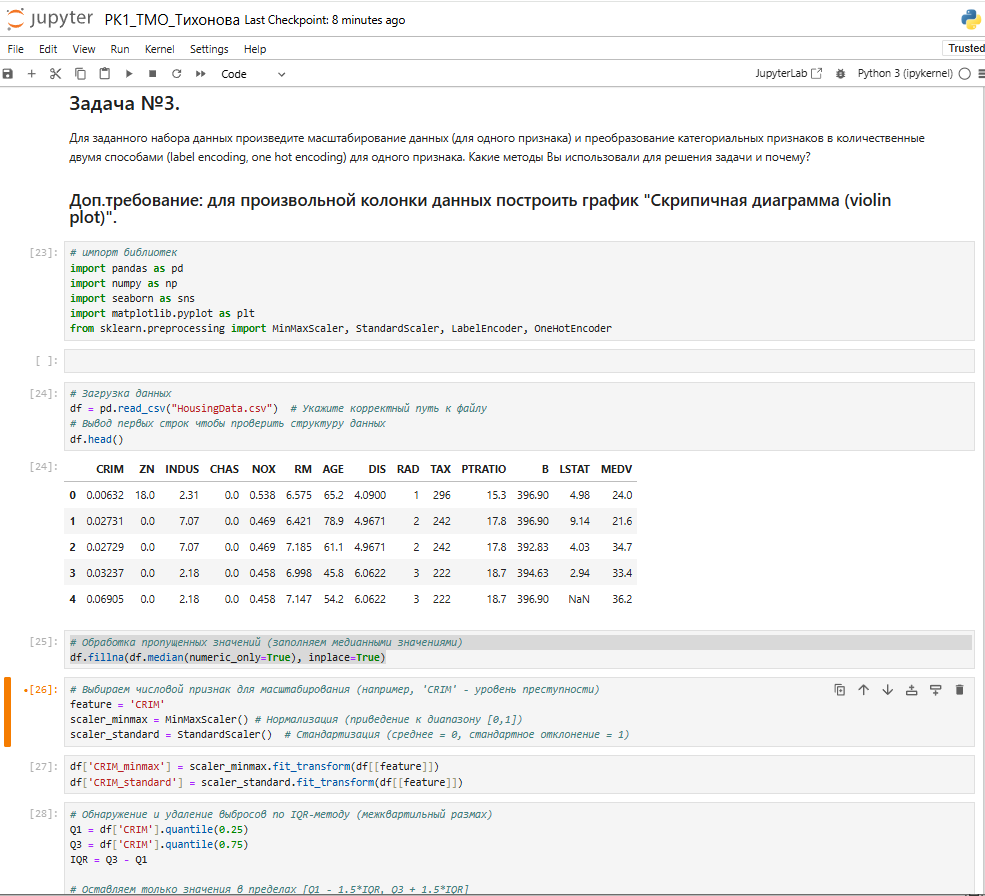
**1. Введение**

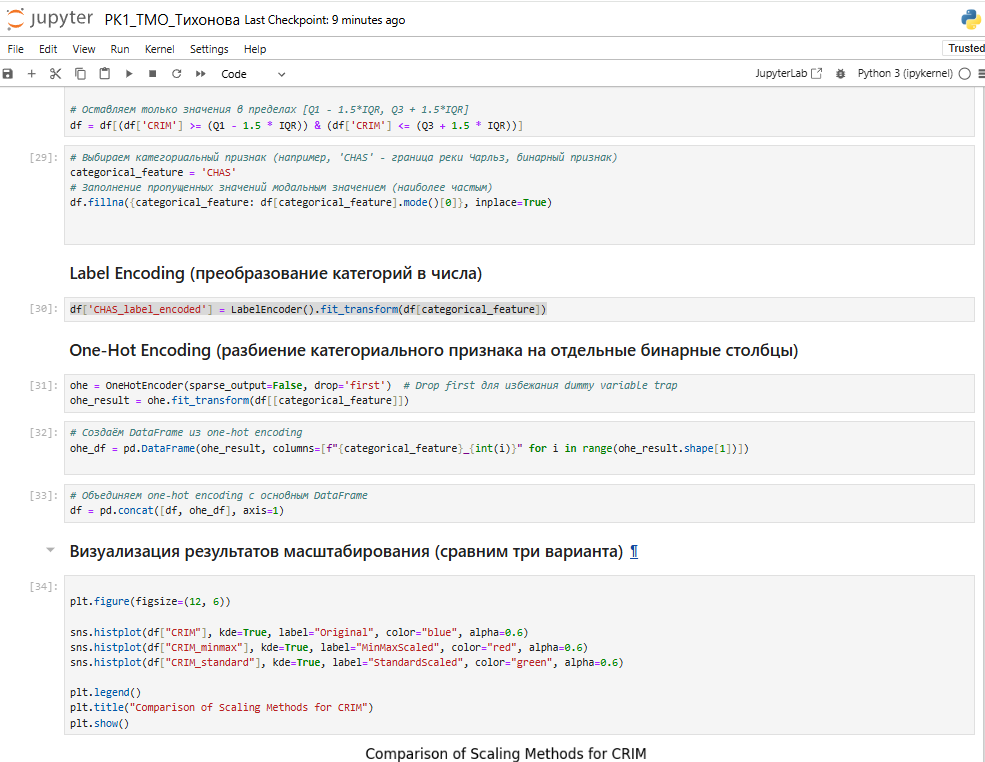
В рамках рубежного контроля была проведена работа с набором данных Boston Housing Dataset. Целью работы являлось масштабирование числового признака, преобразование категориального признака в количественный формат двумя методами (Label Encoding и One-Hot Encoding), а также построение скрипичной диаграммы (violin plot) для визуализации распределения данных.

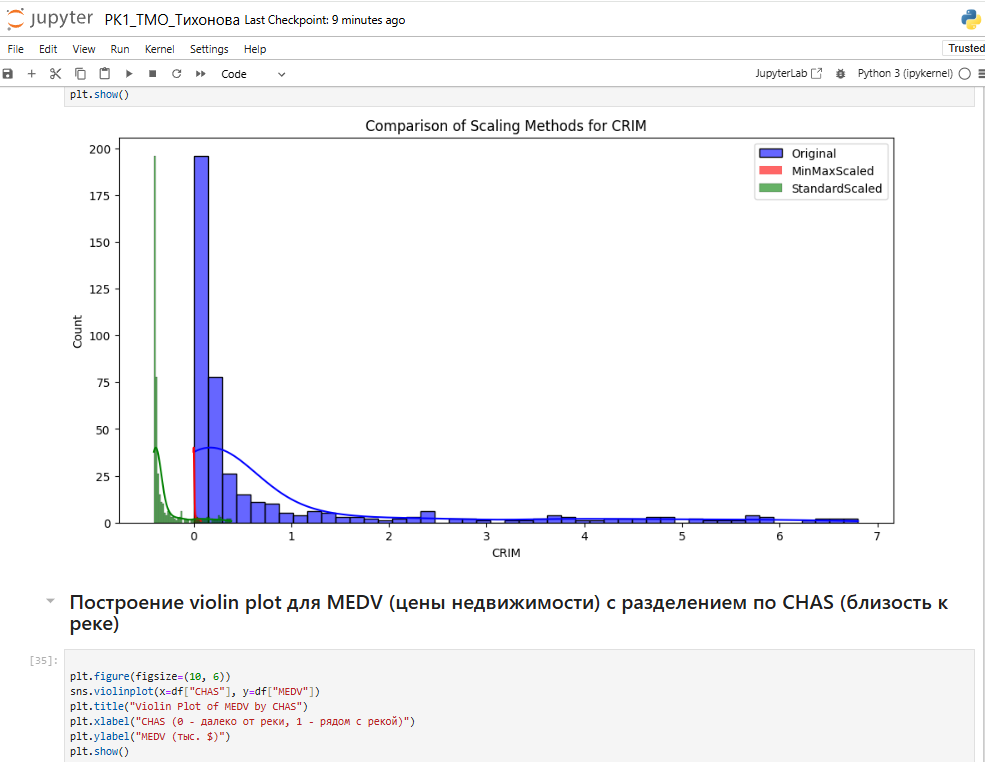
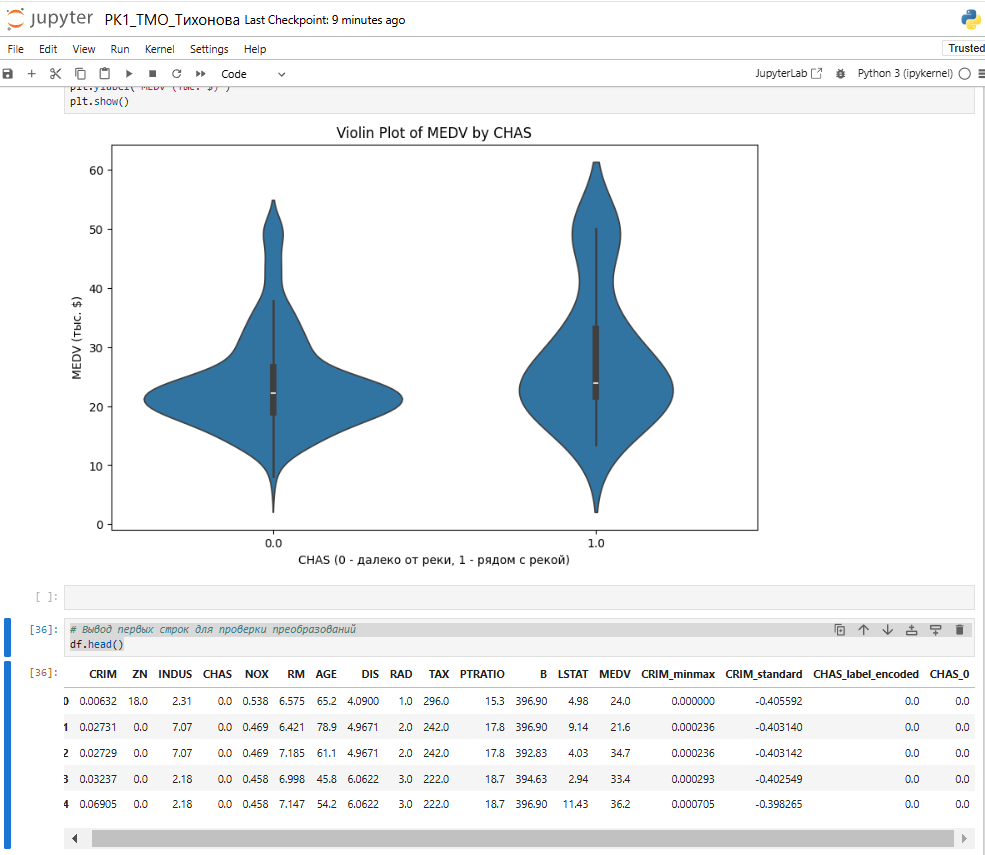
### ****2. Описание исходных данных****

Исходный датасет содержит 506 наблюдений и 14 переменных, включая как числовые, так и категориальные признаки. В данных присутствуют пропущенные значения, которые были обработаны в процессе работы.

**3. Ход выполнения:**





### ****4. Выводы****

В ходе работы были успешно выполнены задачи по масштабированию, обработке категориальных данных и визуализации распределения цен на жилье. Выявлены зависимости между близостью к реке и стоимостью недвижимости. Обработаны выбросы и пропущенные значения, что позволило получить более корректные результаты.