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

Курс «Технологии машинного обучения»

Отчёт по рубежному контролю №1

«Технологии разведочного анализа и обработки данных.»

Вариант № 17

Выполнил Проверил:  
Сайфутдинов Р.И. Гапанюк Ю.Е.  
группа ИУ5-63Б

Дата: 15.04.25 Дата:

Подпись: Подпись:

Москва, 2025 г.

**Задание:**

Номер варианта: **17**

Номер задачи: **3**

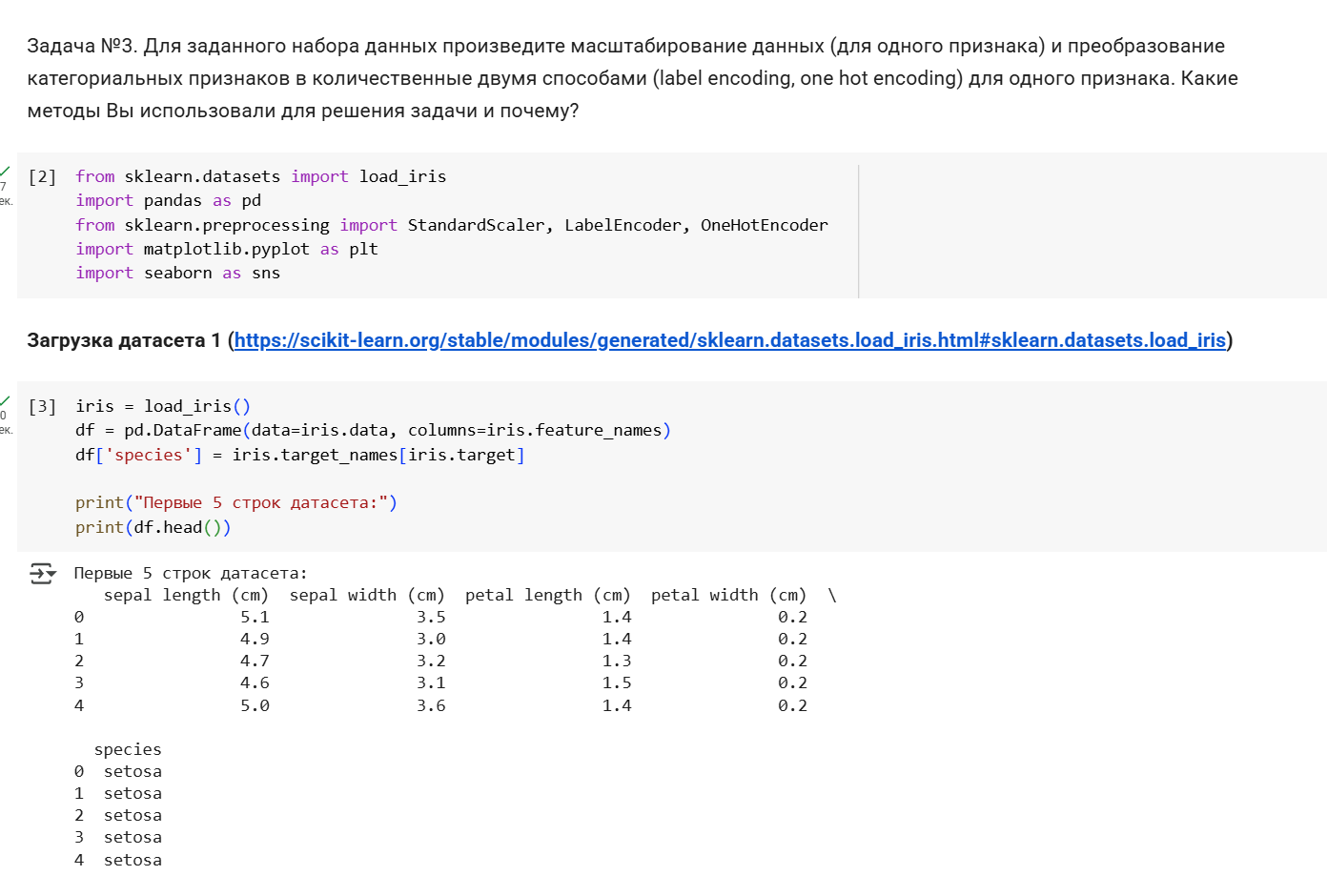
Номер набора данных, указанного в задаче: **1** (https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load\_iris.html#sklearn.datasets.load\_iris)

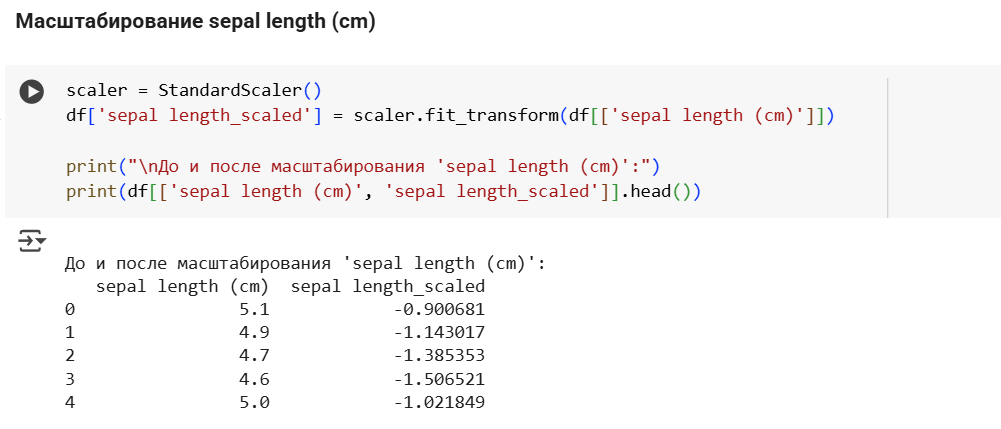
Для студентов групп ИУ5-62Б, ИУ5Ц-82Б - для произвольной колонки данных построить гистограмму.

**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?

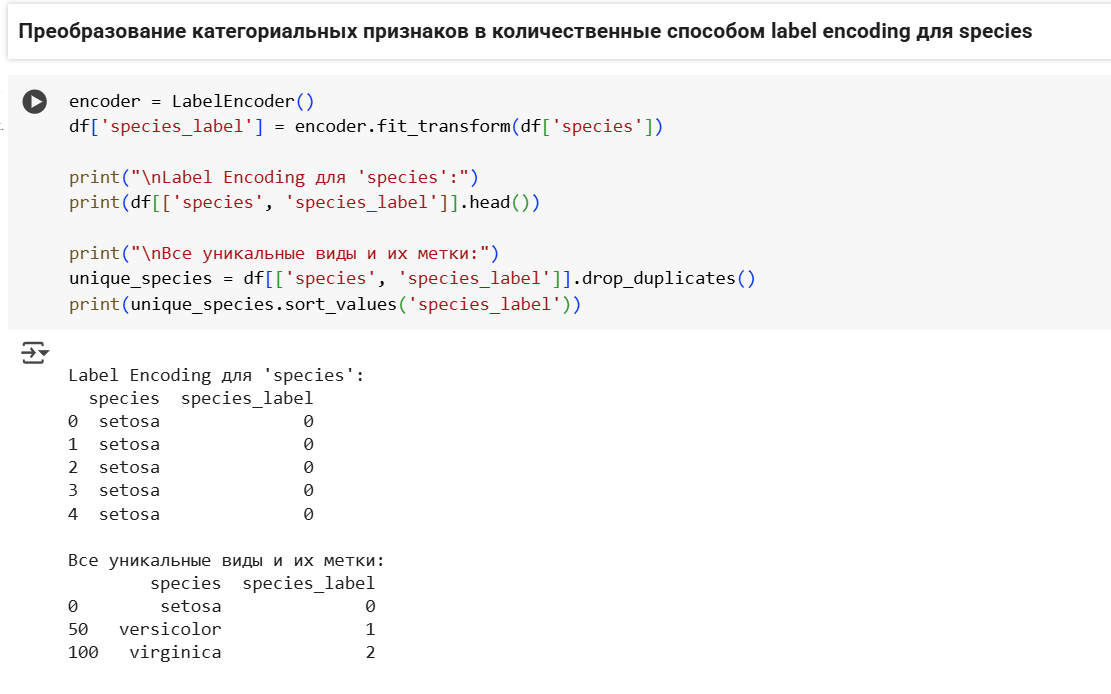
**Ход выполнения:**





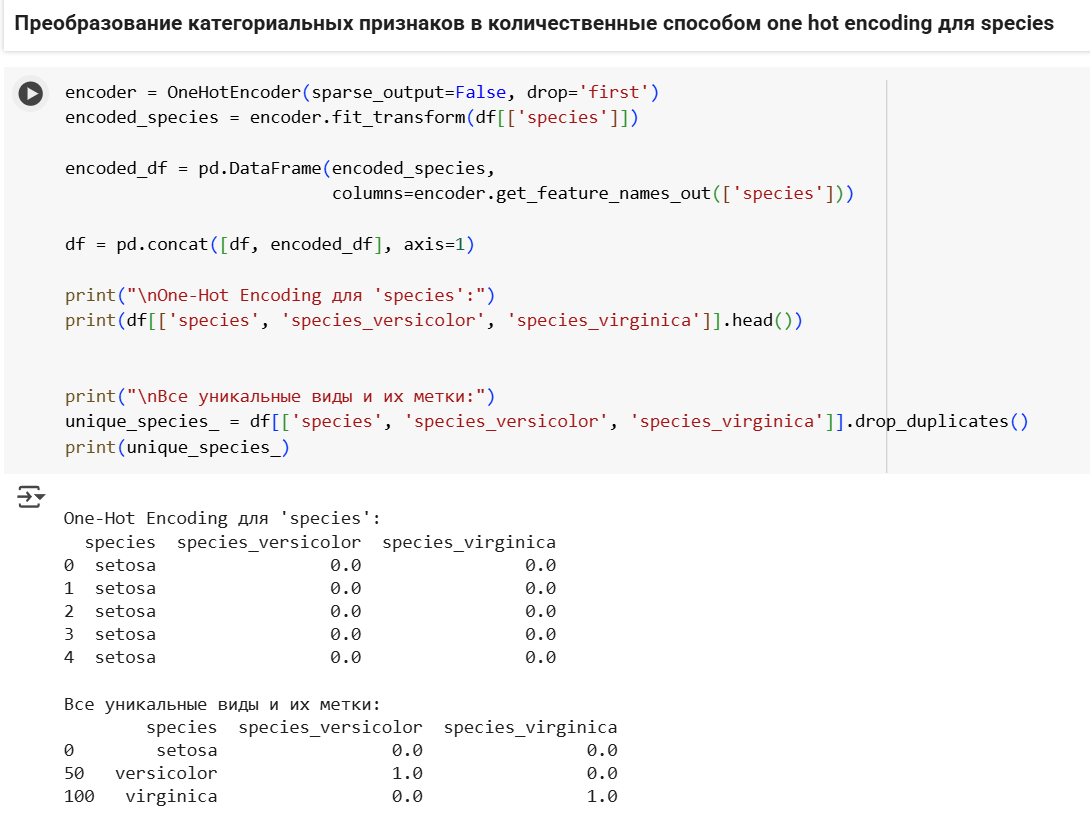
Label Encoding (Кодирование метками)

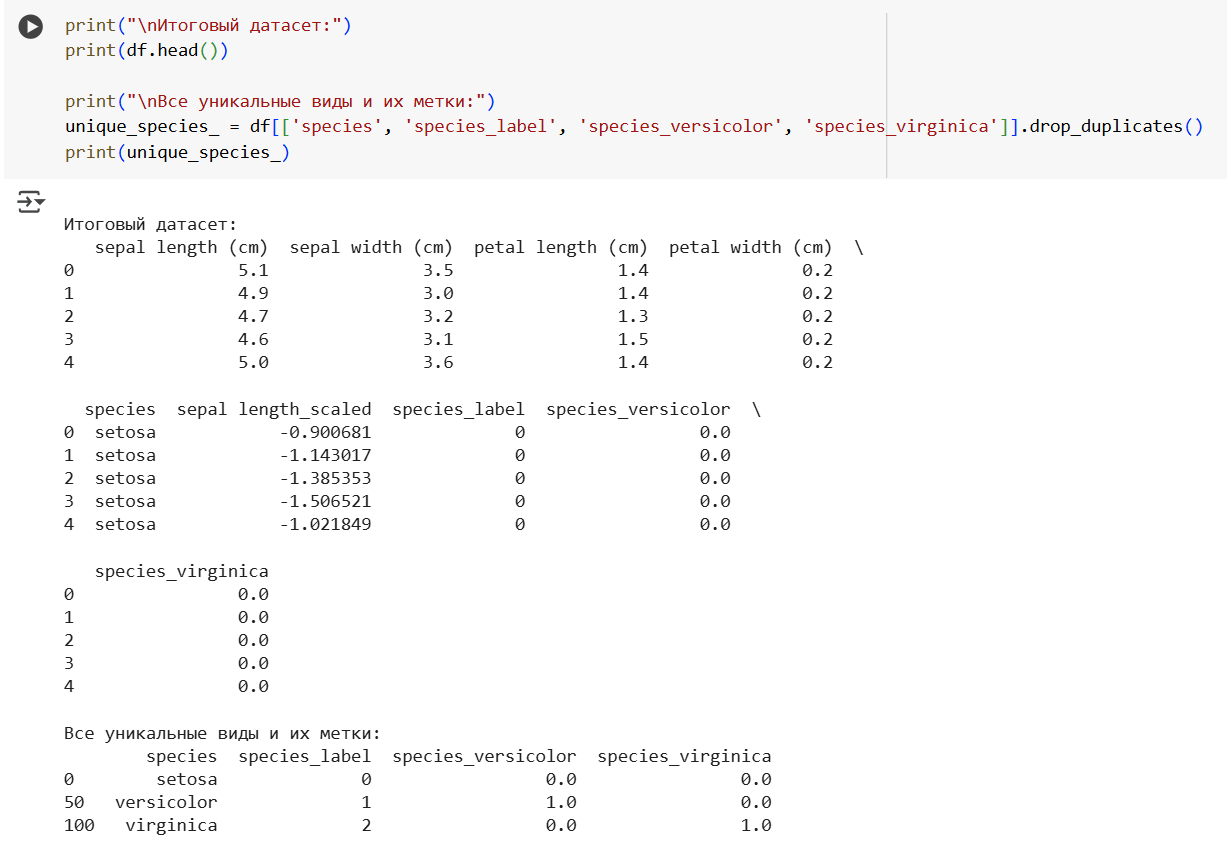
* Каждой категории присваивается уникальное целое число. (Setosa – 0, Versicolor – 1. Virginica – 2).
* Подходит для порядковых данных (где есть иерархия, например, «низкий» < «средний» < «высокий»).

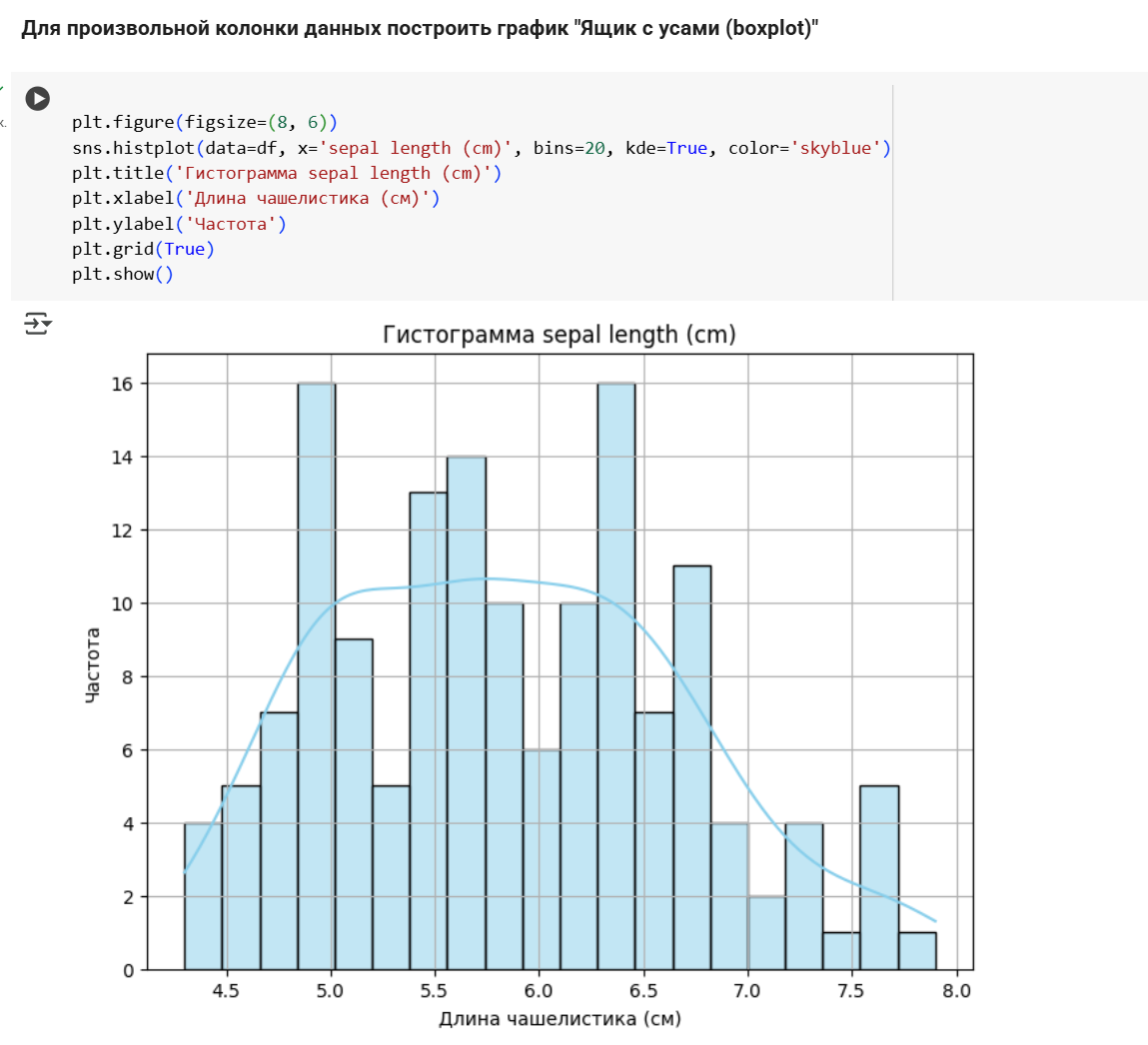
****

One-Hot Encoding (Бинарное кодирование):

* Каждая категория превращается в отдельный бинарный признак (0 или 1).
* Подходит для номинальных данных (где нет порядка).

****

****

****