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

Курс «Технологии машинного обучения»

Отчёт по рубежному контролю №1

«Технологии разведочного анализа и обработки данных.»

Вариант № 13

Выполнил: Проверил:  
Сидоров И.Д. Гапанюк Ю.Е.  
группа ИУ5-64Б

Дата: 11.04.25 Дата:

Подпись: Подпись:

Москва, 2025 г.

**Задание:**

Номер варианта: **1**

Номер задачи: **2**

Номер набора данных, указанного в задаче: **5** (https://www.kaggle.com/mohansacharya/graduate-admissions (файл Admission\_Predict.csv))

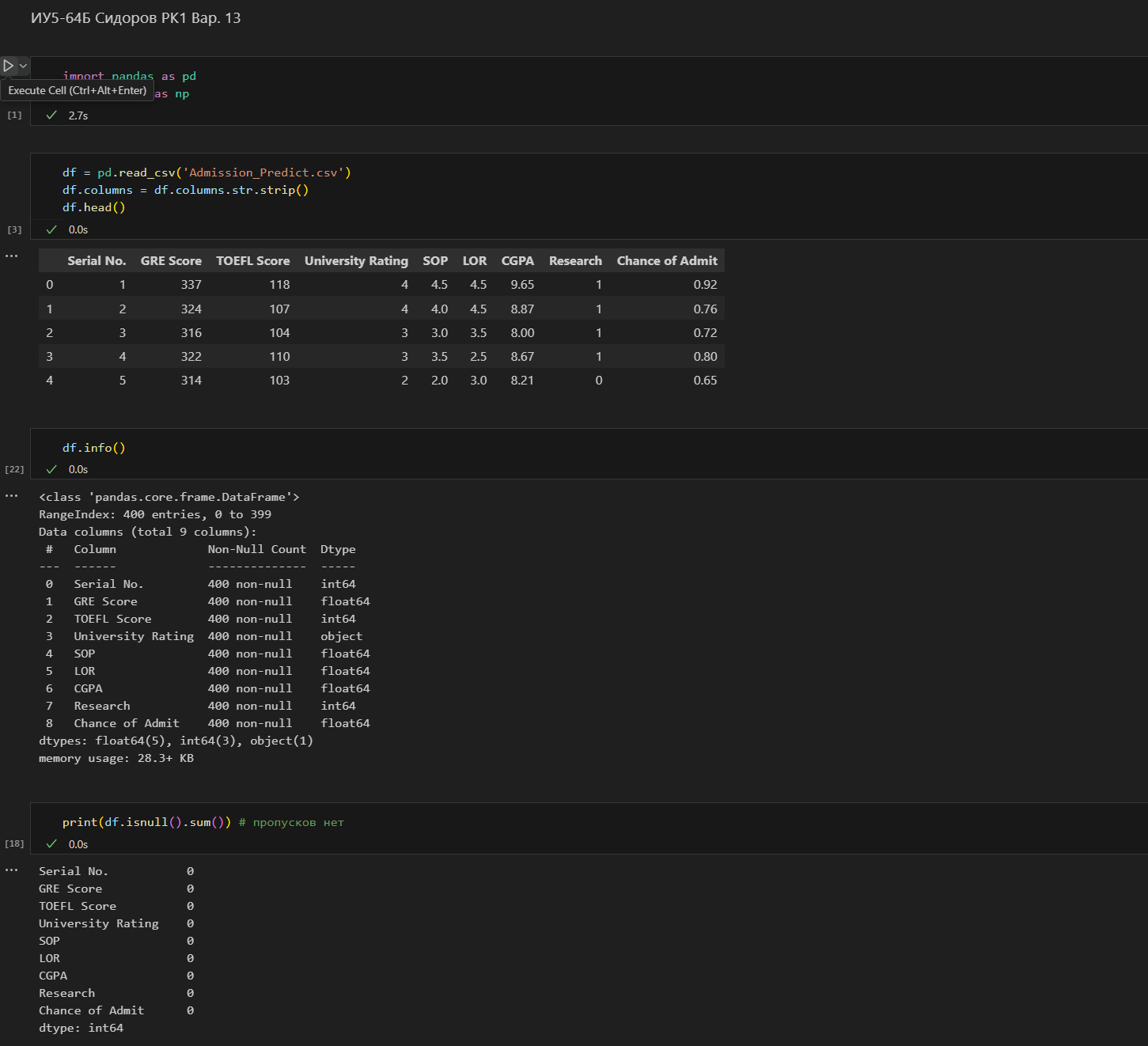
Для студентов группы ИУ5-64Б, ИУ5Ц-84Б - для произвольной колонки данных построить график "Скрипичная диаграмма (violin plot)".

**Задача №2.**

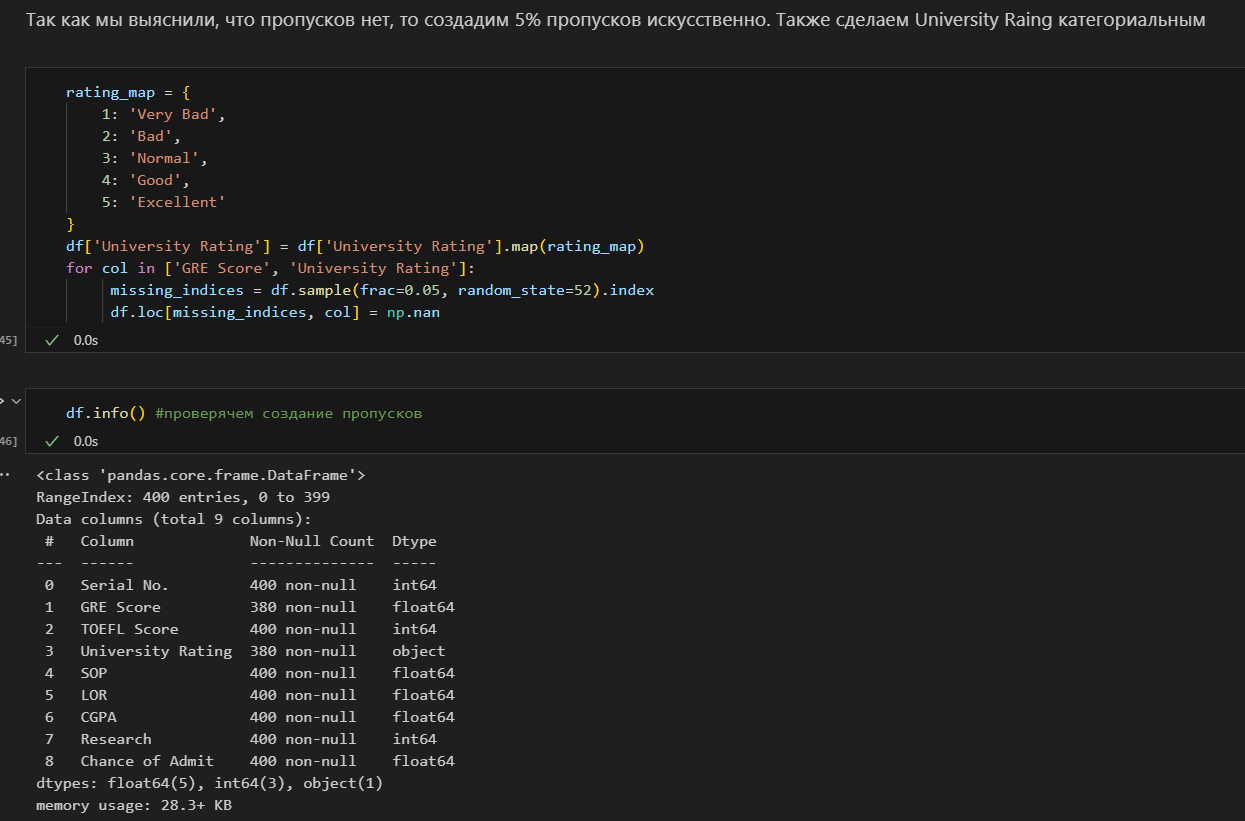
Для заданного набора данных проведите обработку пропусков в данных для одного категориального и одного количественного признака. Какие способы обработки пропусков в данных для категориальных и количественных признаков Вы использовали? Какие признаки Вы будете использовать для дальнейшего построения моделей машинного обучения и почему?

**Ход выполнения:**

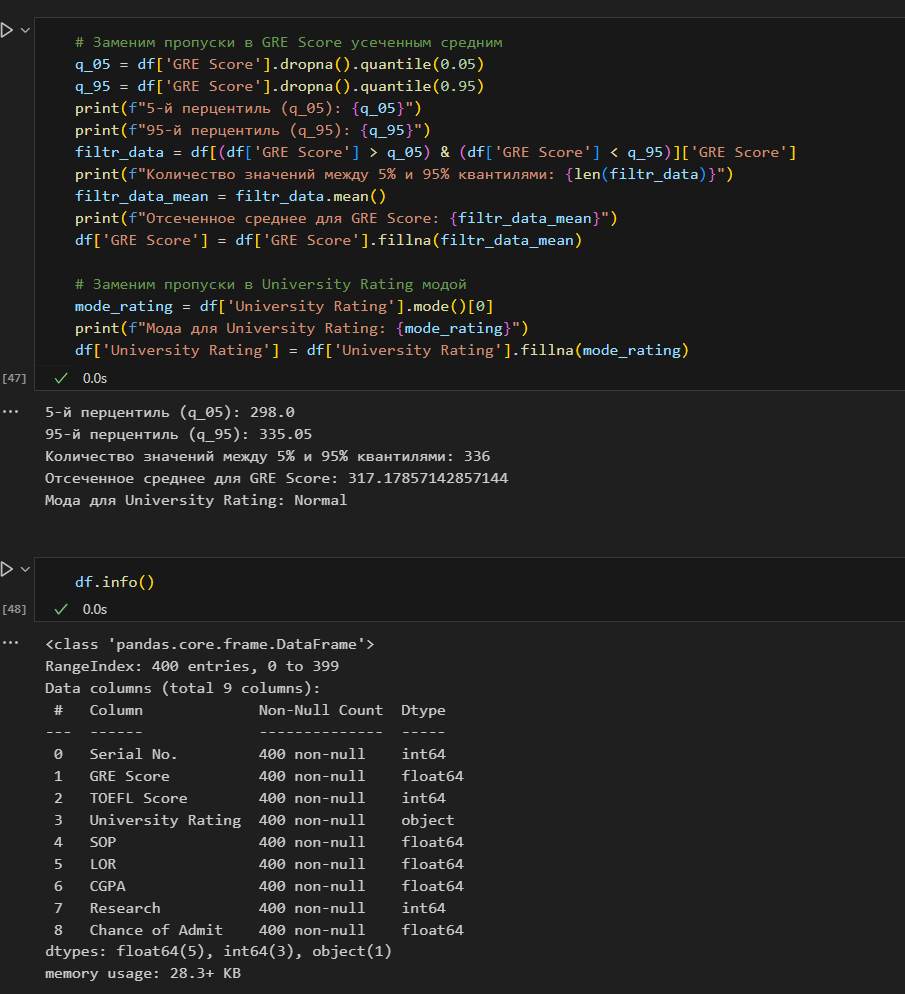
1. Загрузил набор данных, просмотрел начало, проверил пропуски и выяснил, что пропуски отсутствуют.



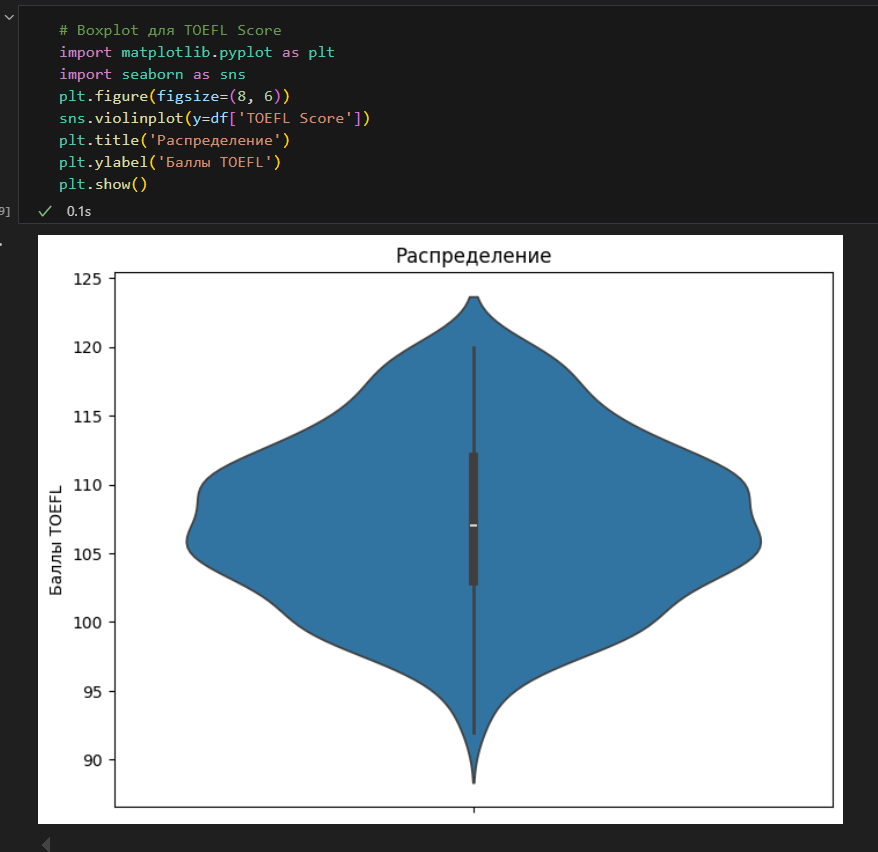
1. Создал 5% пропусков в колонках ‘GRE Score’ и ‘University Rating’ искусственно, предварительно сделав ‘University Rating’ категориальным.



1. Заменил пропуски в ‘GRE Score’ и ‘University Rating’ отсеченным средним и модой соответственно.



1. Построил violin plot для TOEFL Score



Далее для построения моделей машинного обучения я буду использовать все признаки(после преобразования University Rating обратно в числовой). В столбцах признаков нет пропусков, они являются числовыми и согласно моему представлению о поступлении в магистратуру все эти признаки так или иначе имеют влияние на шанс поступления.