**Московский государственный технический**

**университет им. Н.Э. Баумана**

Факультет «Информатика с системы управления»

Кафедра ИУ5 «Системы обработки информации и управления»

Курс «Технологии машинного обучения»

Отчёт по рубежному контролю №1

Вариант №13

Выполнил: Проверил:

студент группы РТ5-61Б преподаватель каф. ИУ5

Мицкевич В.Б. Гапанюк Ю.Е.

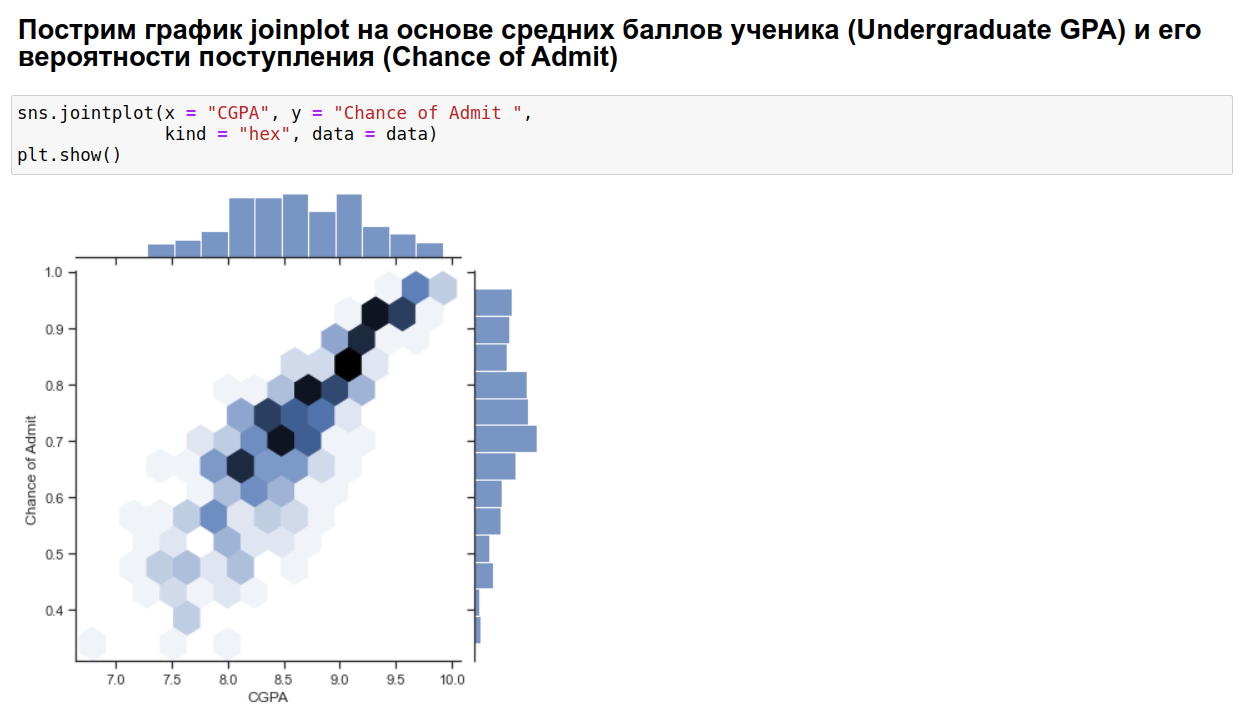
Подпись и дата: Подпись и дата:

Москва, 2023 г.

# **Вариант 13 (Задача №2, Датасет №4)**

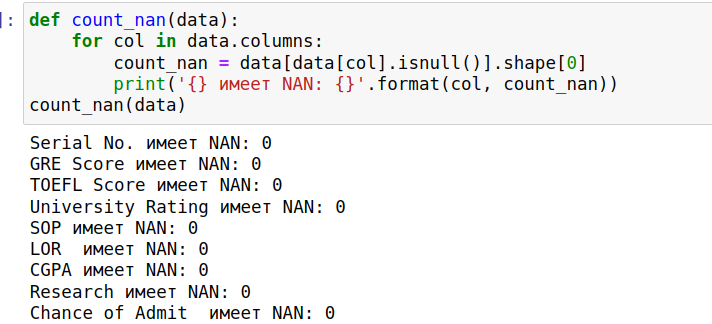
Для заданного набора данных проведите обработку пропусков в данных для одного категориального и одного количественного признака. Какие способы обработки пропусков в данных для категориальных и количественных признаков Вы использовали? Какие признаки Вы будете использовать для дальнейшего построения моделей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

Для студентов группы РТ5-61Б - для пары произвольных колонок данных построить график "Jointplot".

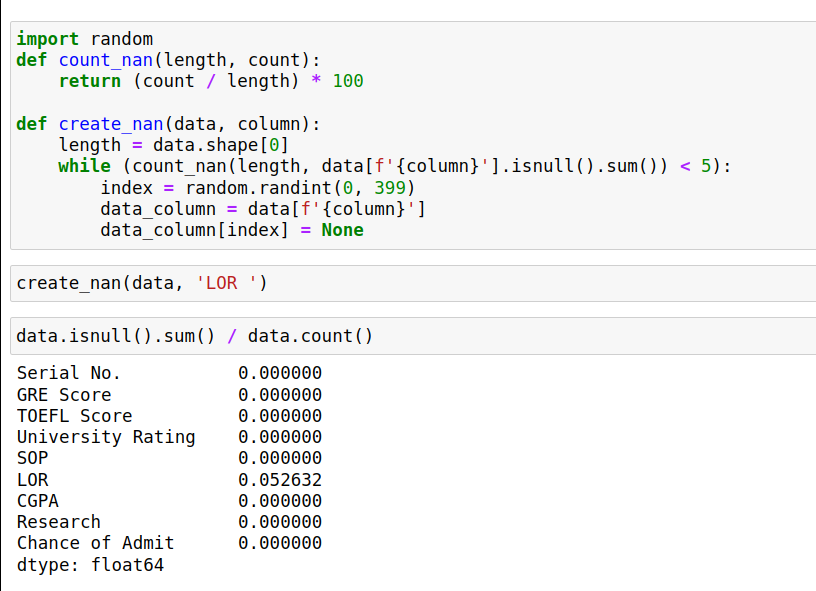


1)**Обработка пропусков в данных для одного количественного признака**

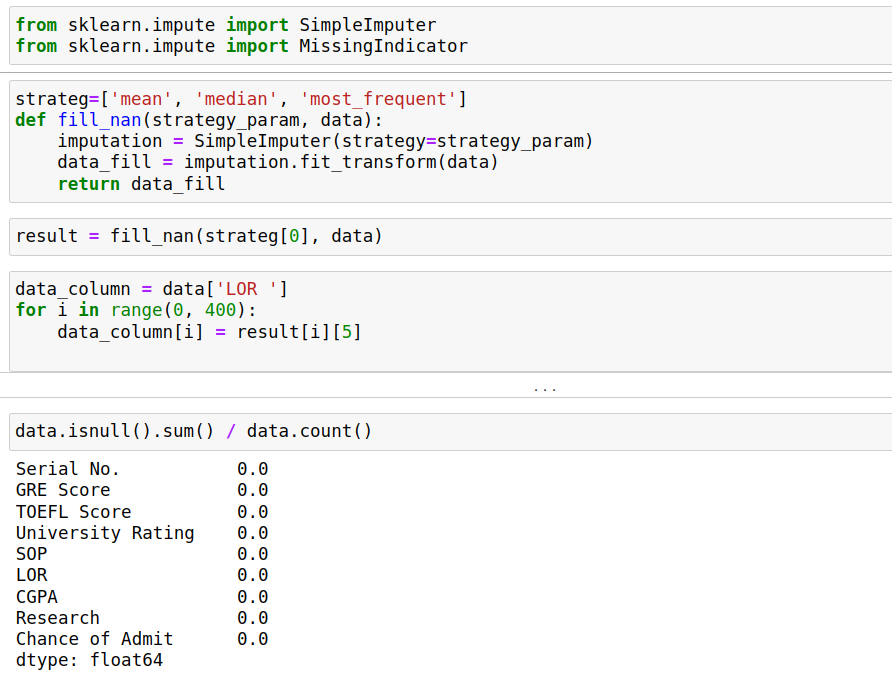
Для начала посмотрим существуют ли пропуска в данном датасете.



В данном датасете отсутствуют признаки. Поэтому искусственно их создадим.В качестве анализа возбмем столбец «LOR».

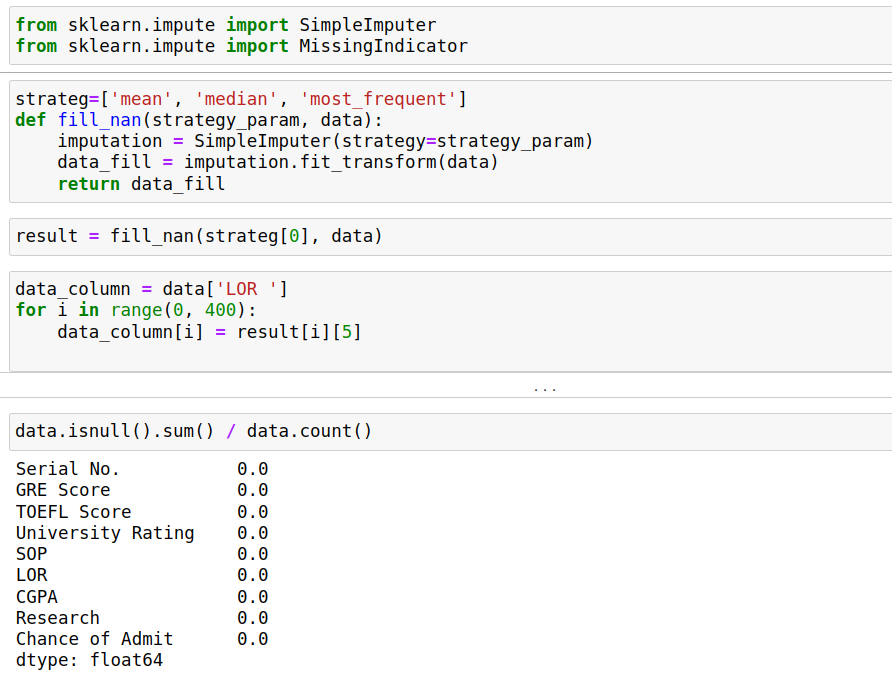


Просмотрев описание данных, мы видим, что значения «LOR» находятся в интервале [1, 5], поэтому заполним пропуска средним значение — 2,5.

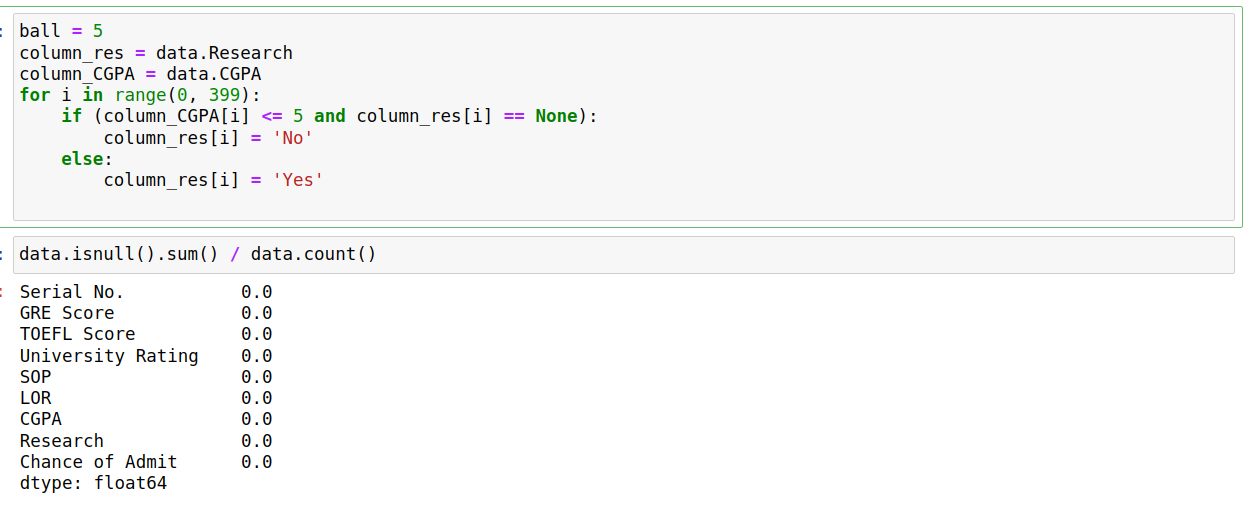


2)**Обработка пропусков в данных для одного категориального признака**

В данном датасете отсутсвуют категориальные признаки, поэтому искуственно создадим один.



Основываясь на баллах бакалвриата можно предположить сделал ученик исследование или нет. На своем опыте могу сказать,что ученики с малым баллом не занимаются исследованиями. Поэтому если у ученика ср балл < 5, то он не занимался исследованием.



**Вывод:**

Для заполнения пропусков можно использовать разные виды заполнения: среднее значение, наиболее встречаемое значение…

А также можно выполнять заполнения пропусков, основываясь на своем опыте.

В моем случае для заполнения пропусков количественного признака я заполнял средним значением, а при заполнение категориального признака решил использовать свой собственный опыт. Основываясь на баллах, можно понять делал ли студент исследования или нет.