Федеральное государственное бюджетное   
образовательное учреждение высшего   
профессионального образования

Московский государственный технический   
университет имени Н.Э. Баумана   
(МГТУ им. Н.Э. Баумана)

**РУБЕЖНЫЙ КОНТРОЛЬ 1  
«Технологии разведочного анализа и обработки данных»**

**ПО ДИСЦИПЛИНЕ «Технологии машинного обучения»**

**Вариант 12**

Выполнил(а): Орлова С. М.   
 (Фамилия И.О. студента)

РТ5-61   
 (Индекс группы)

Проверил(а): Гапанюк Ю. Е.   
 (Фамилия И.О. преподавателя)

«\_\_\_» \_\_\_\_\_\_\_\_\_\_\_\_ 2020 г.

ЗАЧТЕНО / НЕ ЗАЧТЕНО \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
 (нужное выделить) (подпись)

Москва, 2020

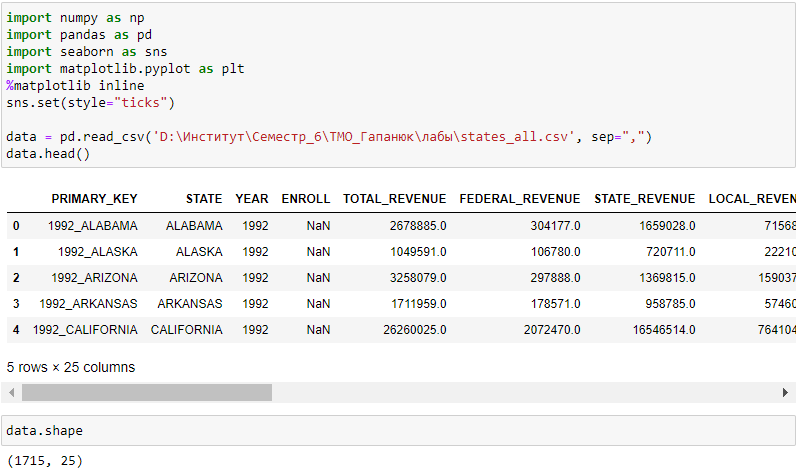
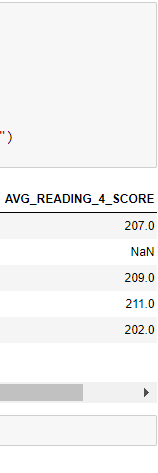
1. Описание задания

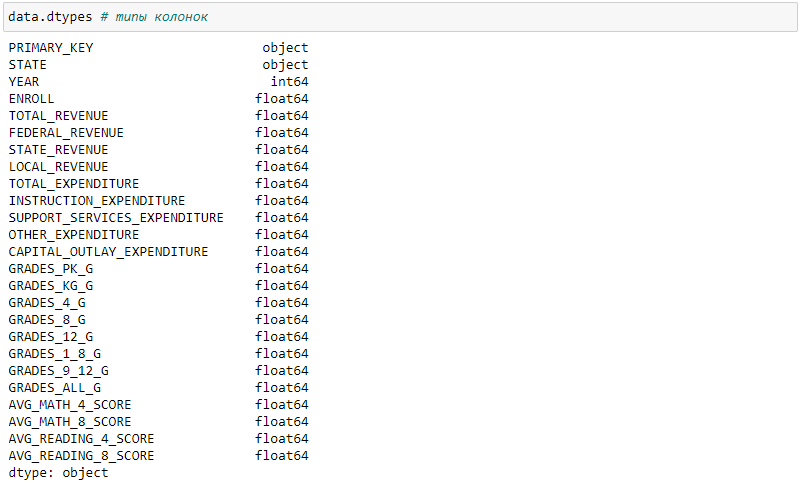
Задача №2:  
Для заданного набора данных (4) проведите обработку пропусков в данных для одного категориального и одного количественного признака. Какие способы обработки пропусков в данных для категориальных и количественных признаков Вы использовали? Какие признаки Вы будете использовать для дальнейшего построения моделей машинного обучения и почему?

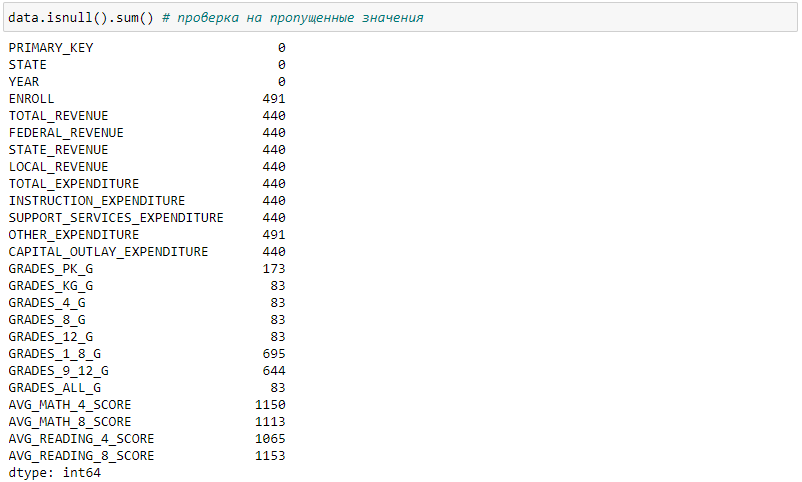
Дополнительные требования:  
Для студентов группы РТ5-61Б - для пары произвольных колонок данных построить график "Jointplot".

1. Текст программы и экранные формы с примерами выполнения программы

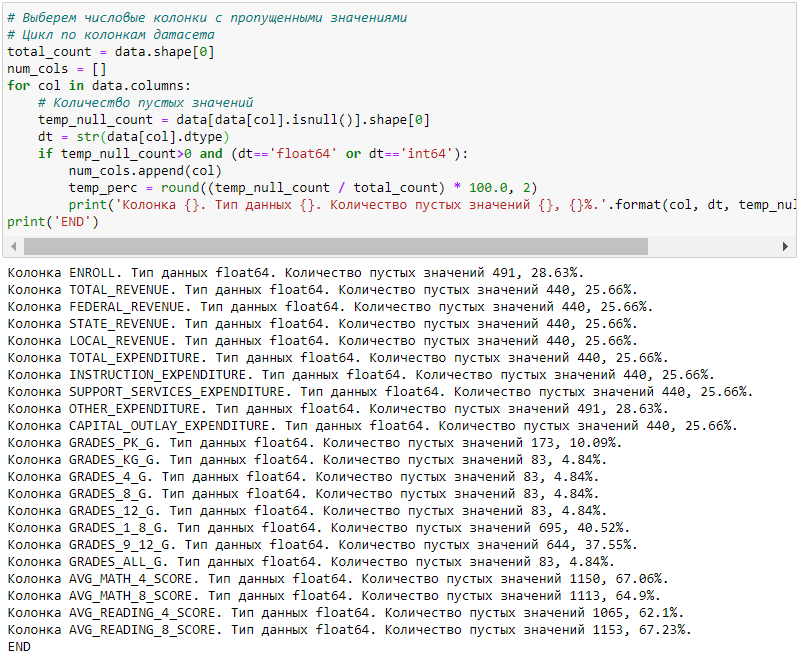
Загрузка и первичный анализ данных:

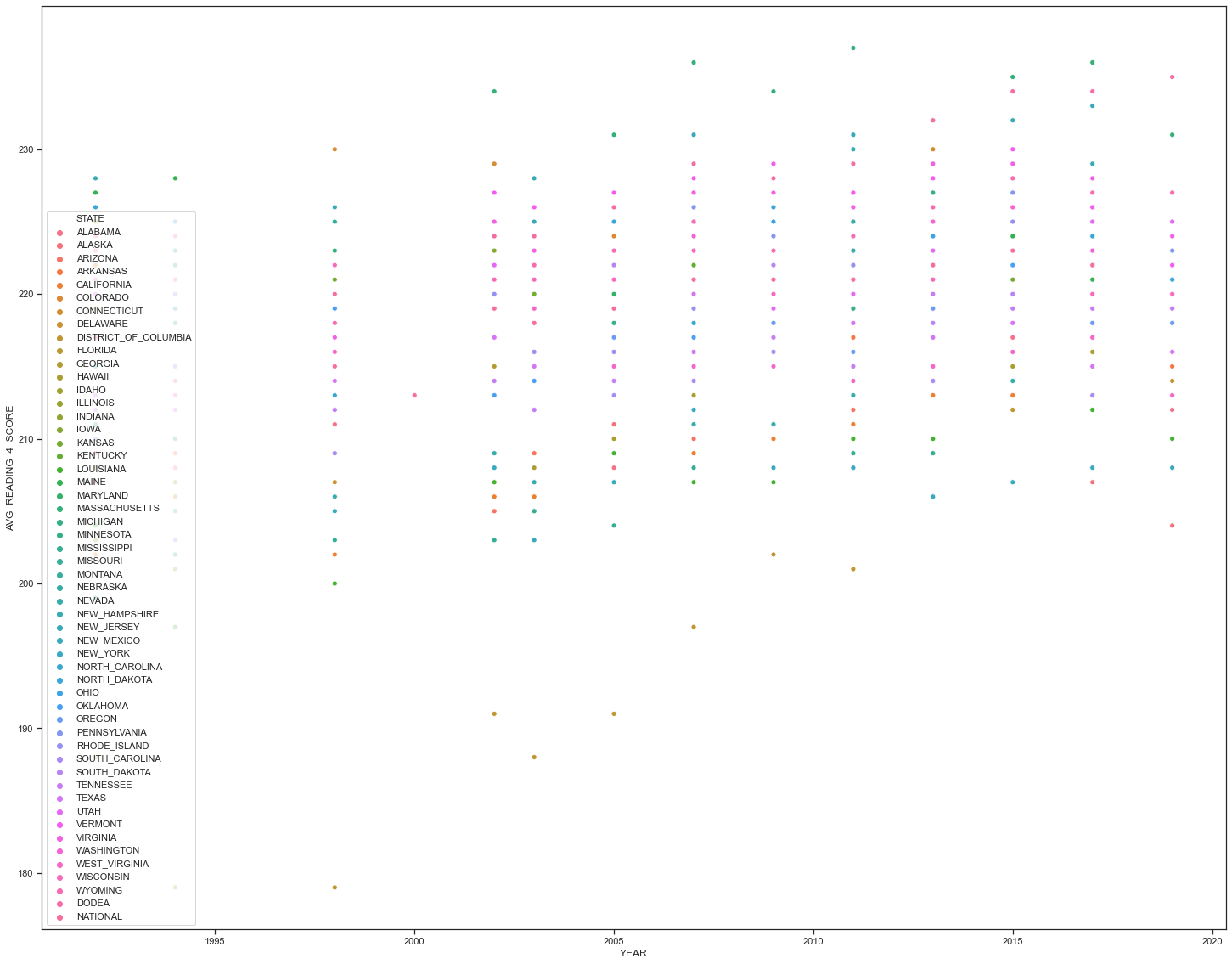


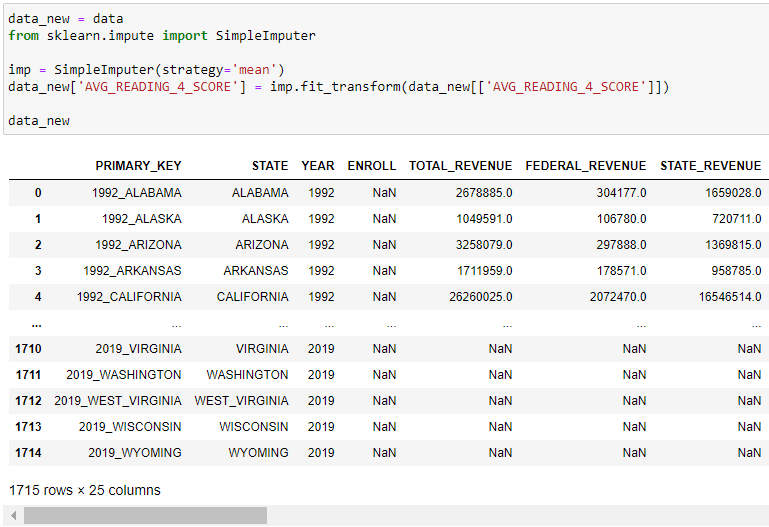
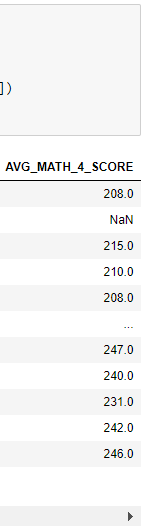


Обработка пропусков в числовых данных:

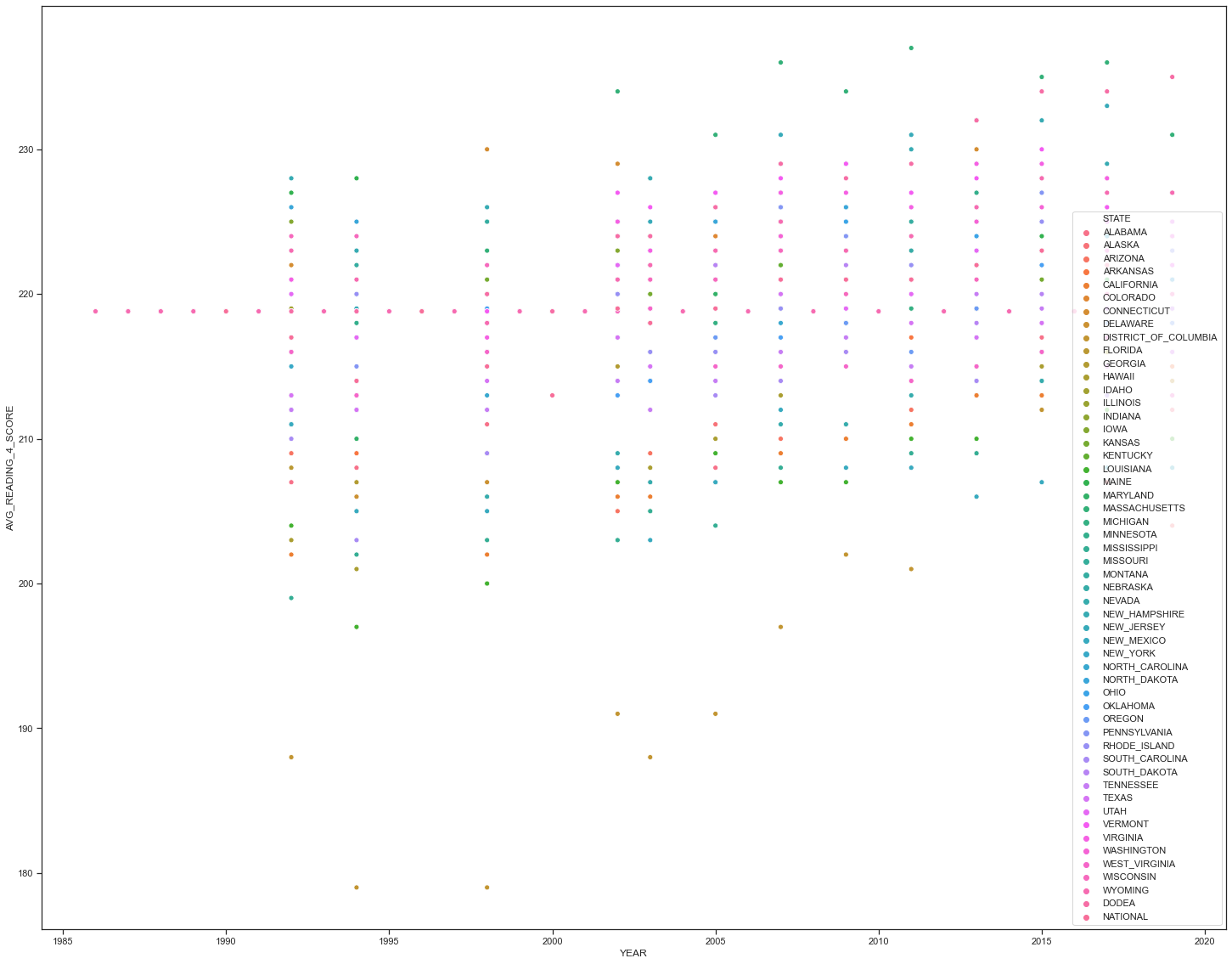










Обработка пропусков в категориальных данных:

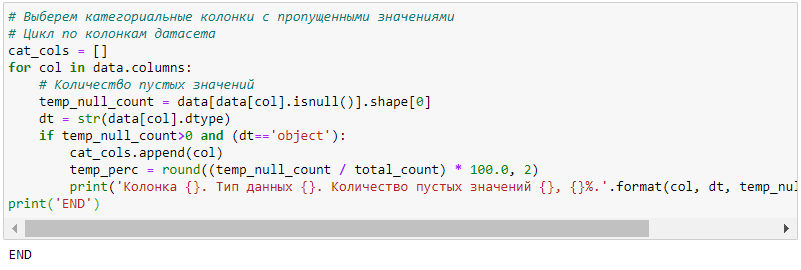
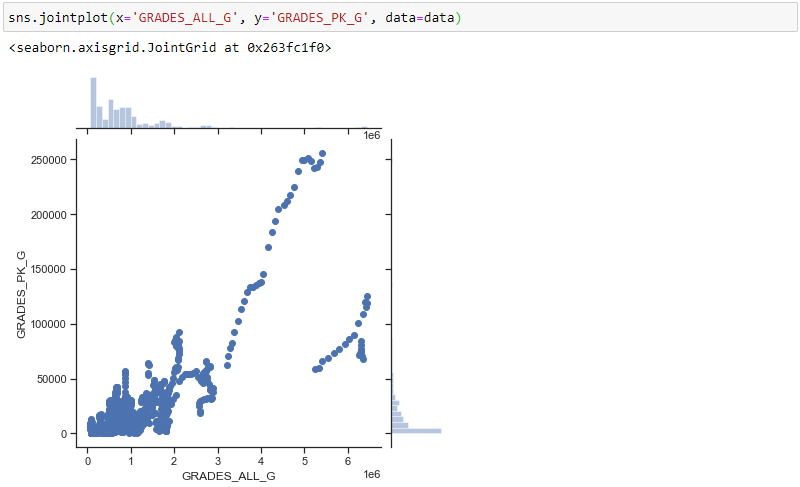


График "Jointplot":



Выводы:

В данной работе пропуски в категориальных данных были обработаны с помощью класса SimpleImputer, а именно были внедрены средние значения. Пропуски в количественных данных отсутствовали, поэтому импьютация не проводилась.