Московский государственный технический университет им. Н.Э. Баумана

Факультет «Информатика и системы управления»

Кафедра «Системы обработки информации и управления»

*Дисциплина «Технологии машинного обучения»*

**Отчёт**

по рубежному контролю №1

Тема: «Технологии разведочного анализа и обработки данных.»

*Вариант 12*

Студент:

Крюков Г. М.

Группа ИУ5-61Б

Преподаватель:

Гапанюк Ю. Е.

Москва, 2020 г.

**Задание**

Задача №2.

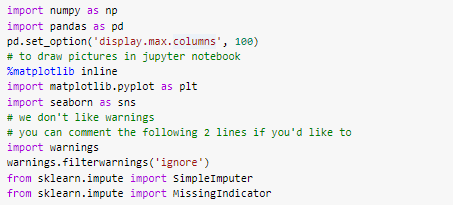
Для заданного набора данных проведите обработку пропусков в данных для одного категориального и одного количественного признака. Какие способы обработки пропусков в данных для категориальных и количественных признаков Вы использовали? Какие признаки Вы будете использовать для дальнейшего построения моделей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

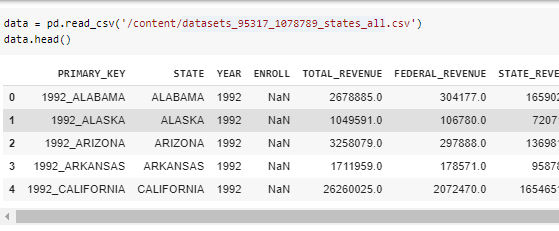
<https://www.kaggle.com/noriuk/us-education-datasets-unification-project>

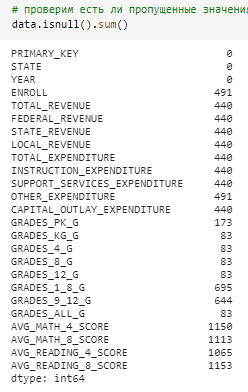
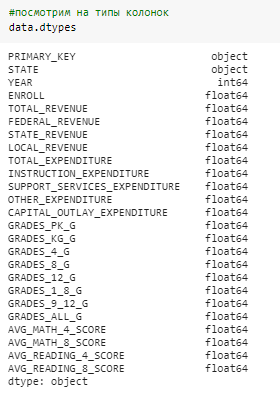
Дополнительные требования по группам:

Для студентов групп ИУ5-61Б, ИУ5Ц-81Б - для пары произвольных колонок данных построить график "Диаграмма рассеяния".

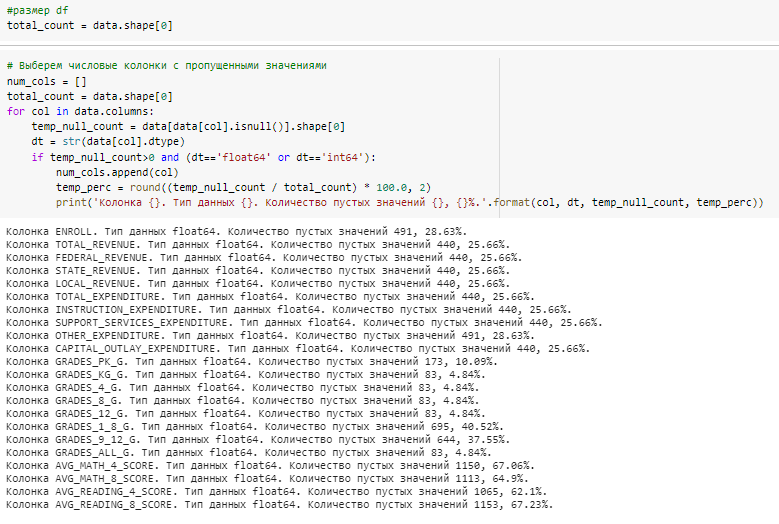
**Выполнение задания**

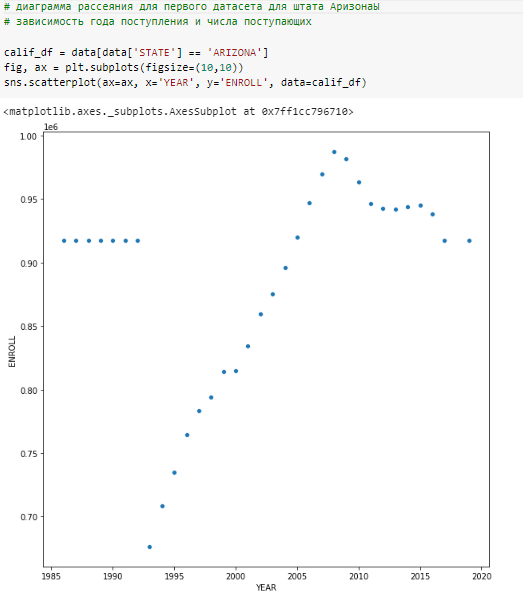
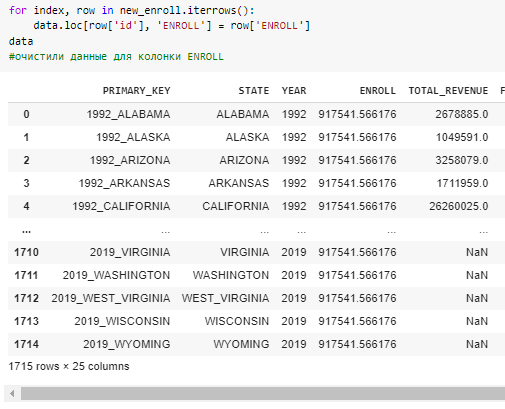
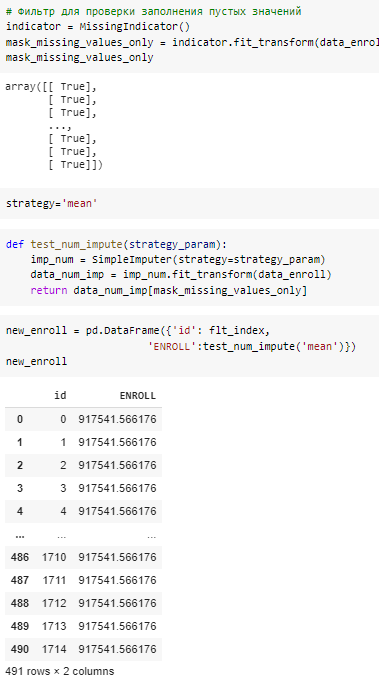
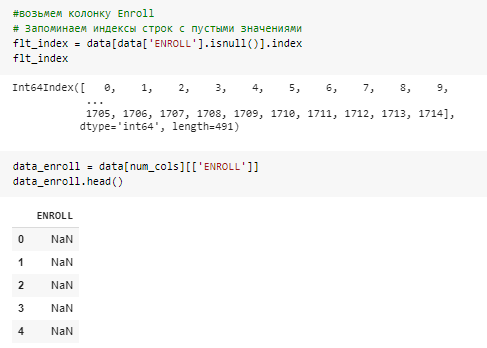






Как мы можем видеть, в датасете нет пропусков категориальных значений.





**Вывод**

Таким образом для обработки пропусков в данных для количественного признака использовался метод импьютации средними значениями. Для дальнейшего построения моделей можно использовать все столбцы, обработав пропуски, в зависимости от нужд исследований.