*Нефёдова Ирина*

*ИУ5-63б*

*РК №1  
Технологии машинного обучения*

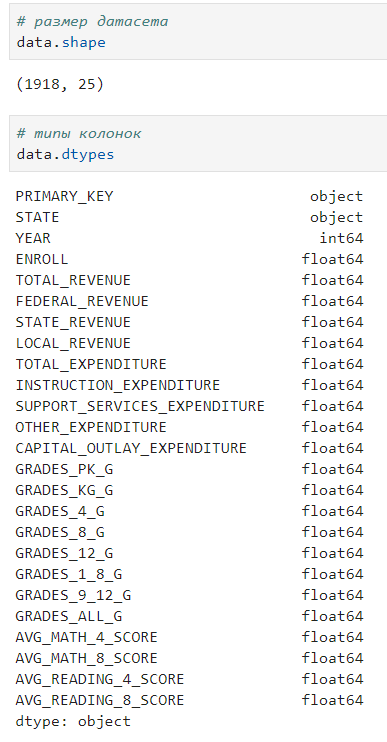
*Вариант №12*

**

**

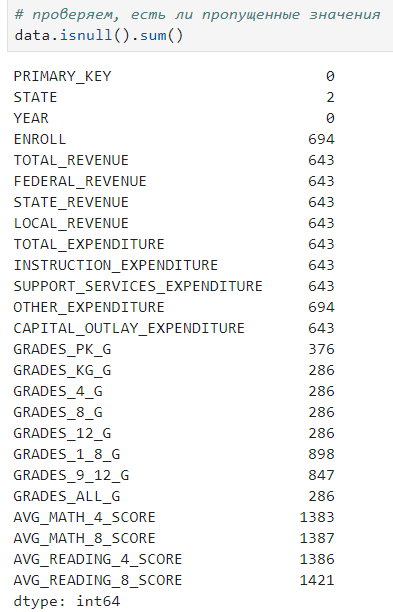
*Задача:* Для заданного набора данных проведите обработку пропусков в данных для одного категориального и одного количественного признака. Какие способы обработки пропусков в данных для категориальных и количественных признаков Вы использовали? Какие признаки Вы будете использовать для дальнейшего построения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

Описание датасета.



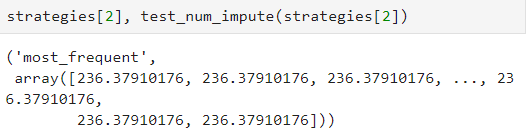
1. Для заданного набора данных провести обработку пропусков в данных для одного категориального и одного количественного признака.

Выполним поиск пропущенных значений в датасете по столбцам. Так как в исходной базе данных не было пропущенных значений типа Object, добавим их самостоятельно. (Поэтому пропущенных значений типа Object всего 2).

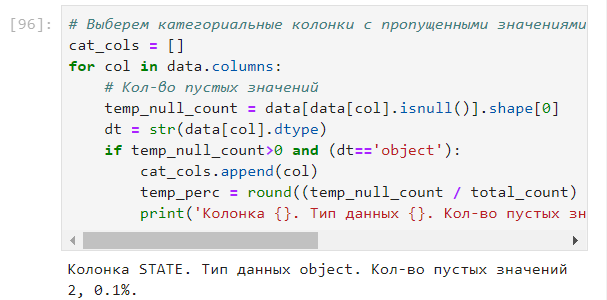


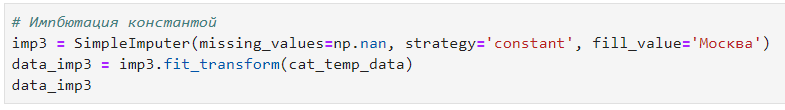
Избавимся от пропущенных значений в поле «AVG\_READING\_8\_SCORE» путем подстановки наиболее часто встречающихся (числовые данные).

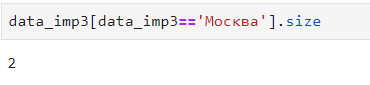




Избавимся от пропущенных значений в поле «STATE» путем подстановки константного значения «Москва» (категориальные данные).







Видим, что оба пропуска, выявленных в 1 пункте, успешно заполнены константой «Москва».

1. Я считаю, что с данными количественного признака работать гораздо удобнее и проще, поэтому думаю, что для дальнейшего построения моделей машинного обучения буду использовать их.