**Московский государственный технический университет**

**им. Н.Э. Баумана**

**(МГТУ им. Н.Э. Баумана)**

**Радиотехнический факультет (РТ)**

Отчёт по рубежному контролю № 1

По дисциплине

«Технологии машинного обучения»

Вариант № 9

|  |  |
| --- | --- |
| Проверил: | Выполнил: |
| Преподаватель кафедры ИУ-5 | студент группы РТ5-61Б |
| Гапанюк Ю.Е. | Лычагин Д.А. |
| Подпись: \_\_\_\_\_\_\_\_\_\_\_\_\_ | Подпись: \_\_\_\_\_\_\_\_\_\_\_\_\_ |
| «\_\_\_» \_\_\_\_\_\_\_\_\_\_\_\_\_ 2020 г. | «\_\_\_» \_\_\_\_\_\_\_\_\_\_\_\_\_ 2020 г. |

Москва, 2020

# Задание

1.1 Условие

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?

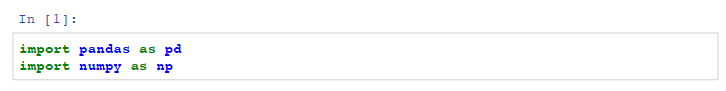
1.2 Набор данных №1

https://www.kaggle.com/karangadiya/fifa19

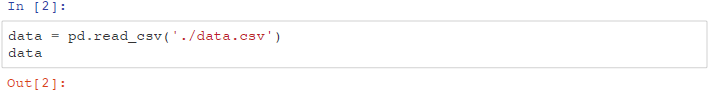
1.3 Дополнительное требование:

Для произвольной колонки данных построить график «Скрипичная диаграмма» (violin plot).

# Выполнение



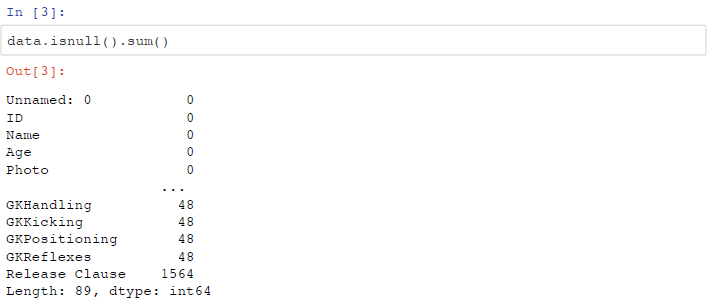
2.1 Извлечение dataset





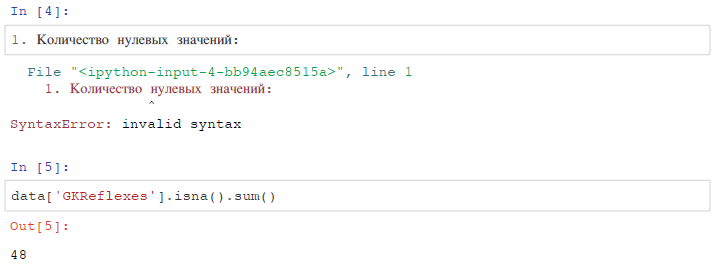
2.2 Обработка пропусков в данных

Проверим, есть ли пропущенные значения

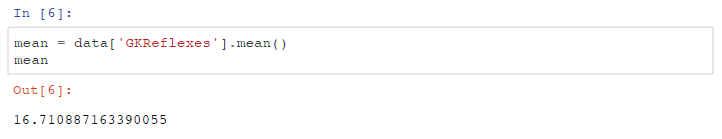


2.3 Замена пустых значений на среднее

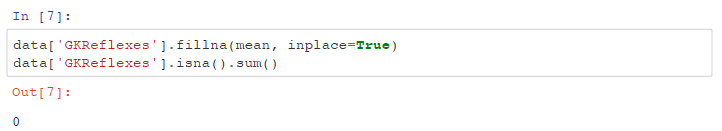
Выполним замену для количественного признака GKReflexes.



Получим среднее:



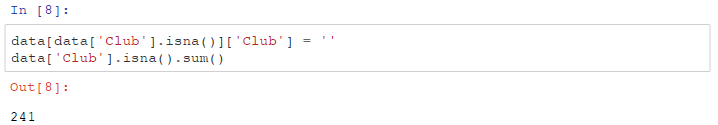
Выполним замену и проверим количество пустых значений:



2.4 Удаление пустых значений

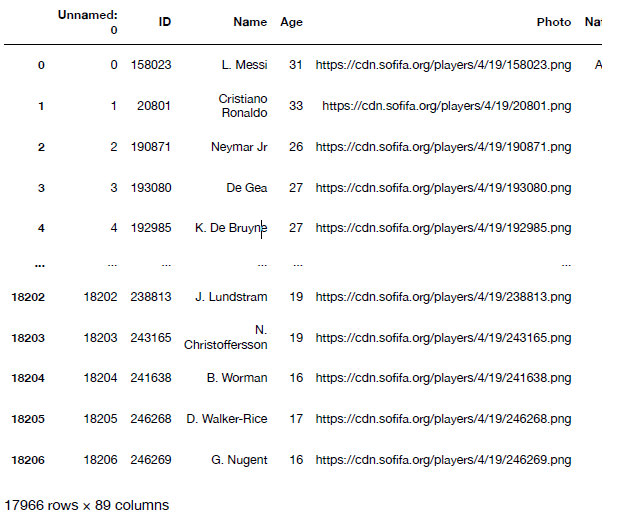
Выполним удаление для категориального признака Club.

Количество нулевых значений:



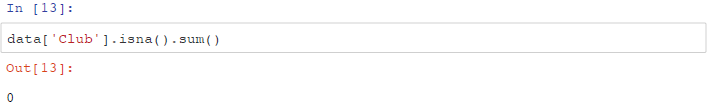
Удалим строки, содержащие нулевое значение колонки Club:





Как можно видеть, количество строк датасета уменьшилось.

Проверим количество пустых значений поля Club:



2.5 Дополнительное задание

Построим график "Скрипичная диаграмма" (Violin plot) для поля Potential

