**Московский государственный технический**

**университет им. Н.Э. Баумана**

Факультет "Информатика и системы управления"

Кафедра ИУ5 "Системы обработки информации и управления"

Технологии машинного обучения

Отчет по лабораторной работе №2

Выполнил:

Студент группы ИУ5-65Б

Козинов Олег

Москва, 2021

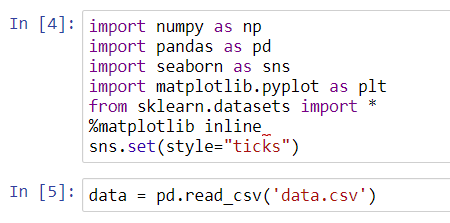
**Задание**

1. Выбрать набор данных (датасет), содержащий категориальные признаки и пропуски в данных. Для выполнения следующих пунктов можно использовать несколько различных наборов данных (один для обработки пропусков, другой для категориальных признаков и т.д.)
2. Для выбранного датасета (датасетов) на основе материалов лекции решить следующие задачи:

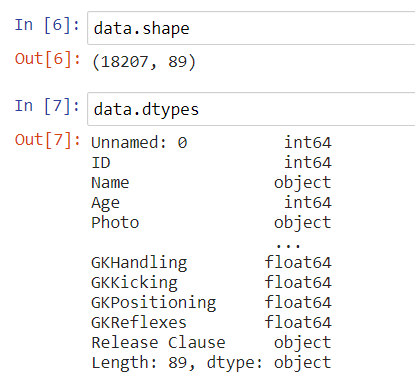
* обработку пропусков в данных;
* кодирование категориальных признаков;
* масштабирование данных.

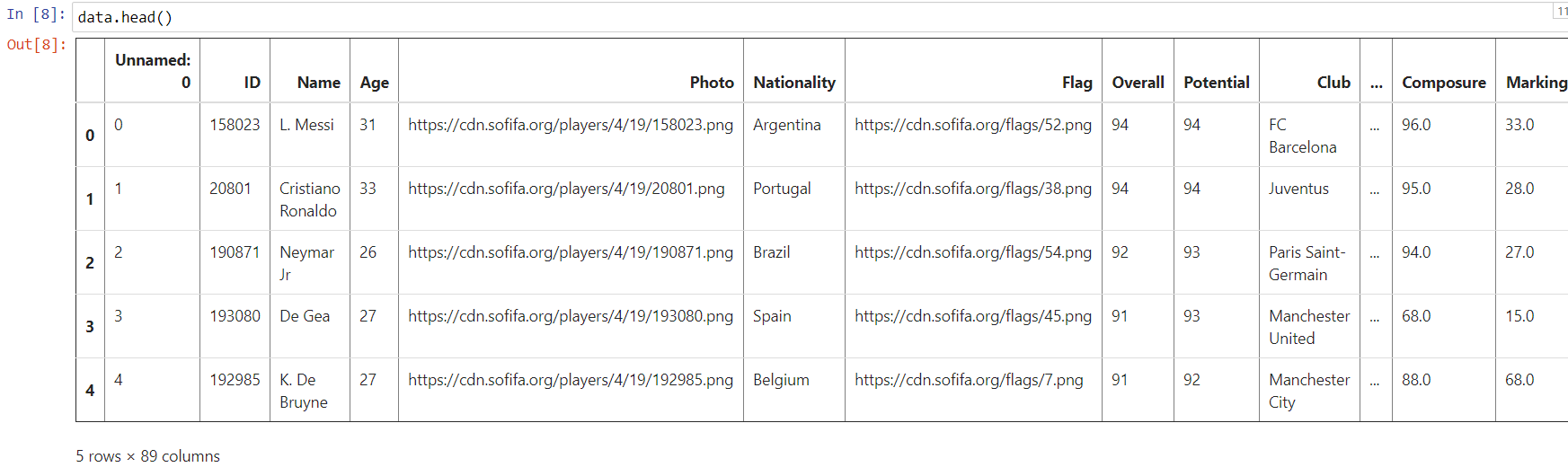
**Выполнение лабораторной работы:**

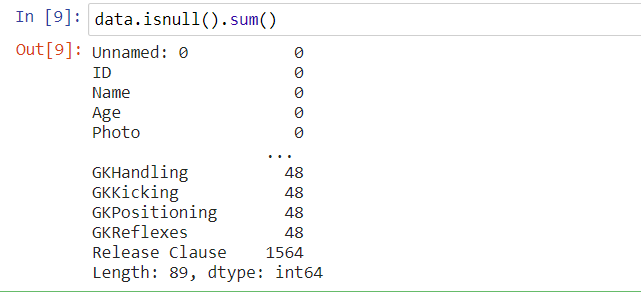
**Загрузка библиотек и датасета**



**Основные характеристики набора данных**

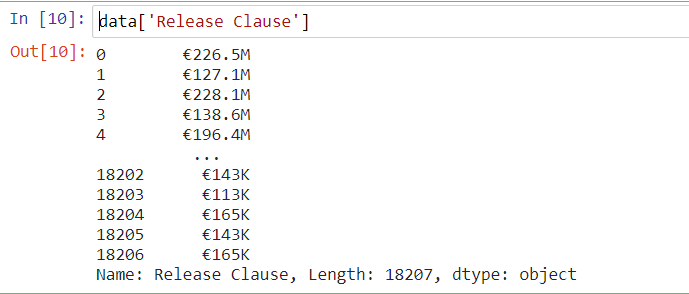




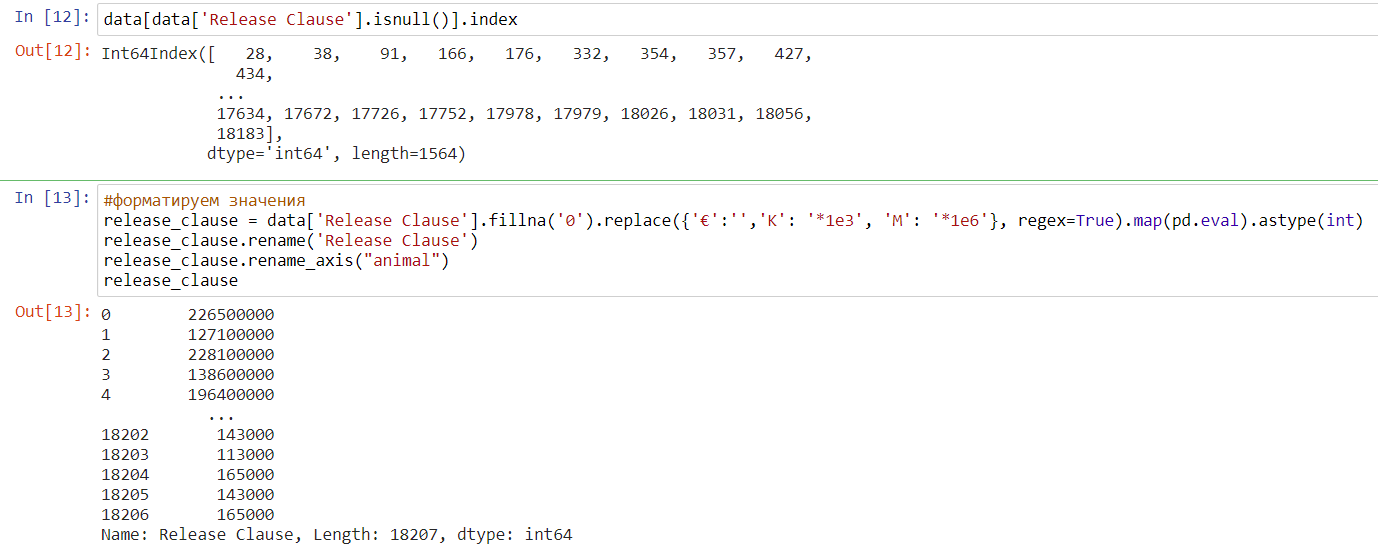


**Обработка пропусков:**

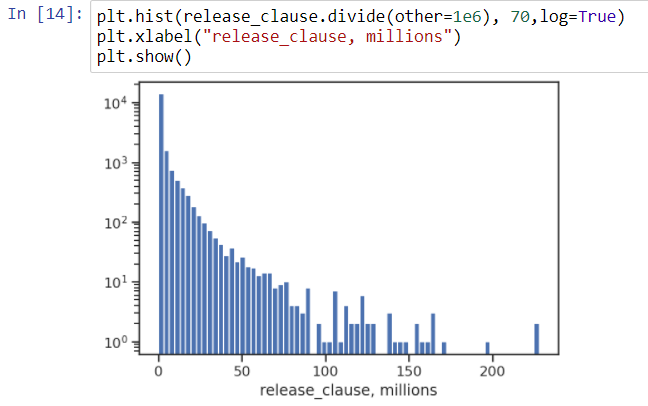
**Подбор и преобразование колонки**



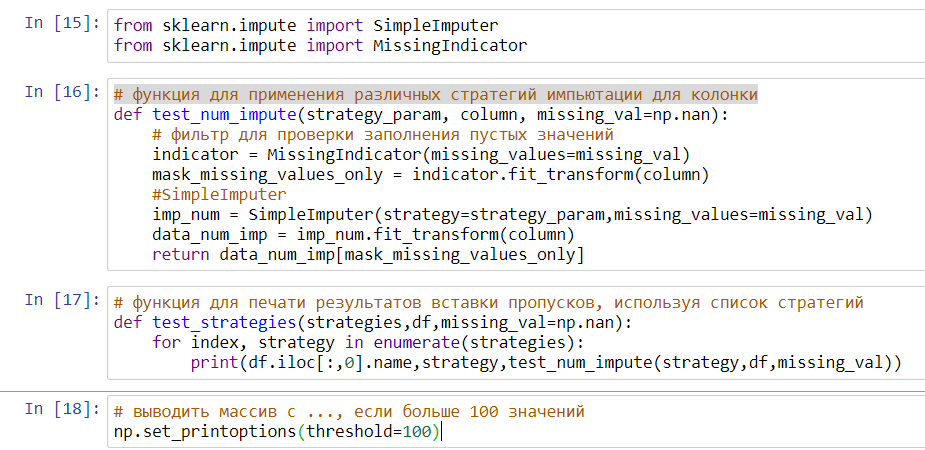


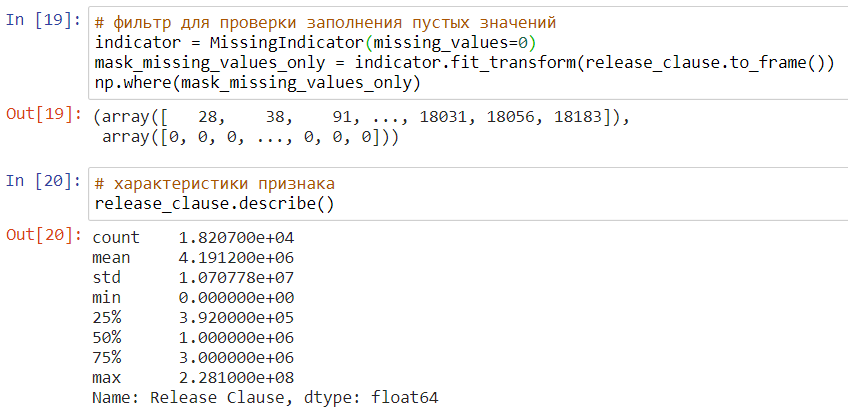


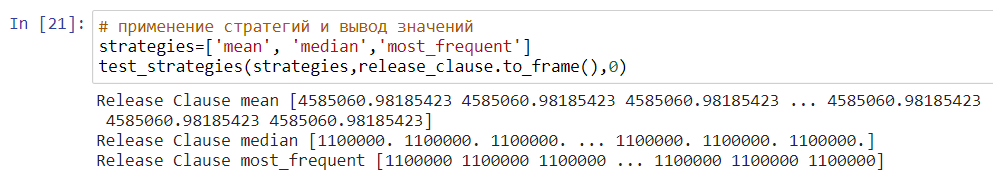
**Гистограмма признака в логарифмическом масштабе**



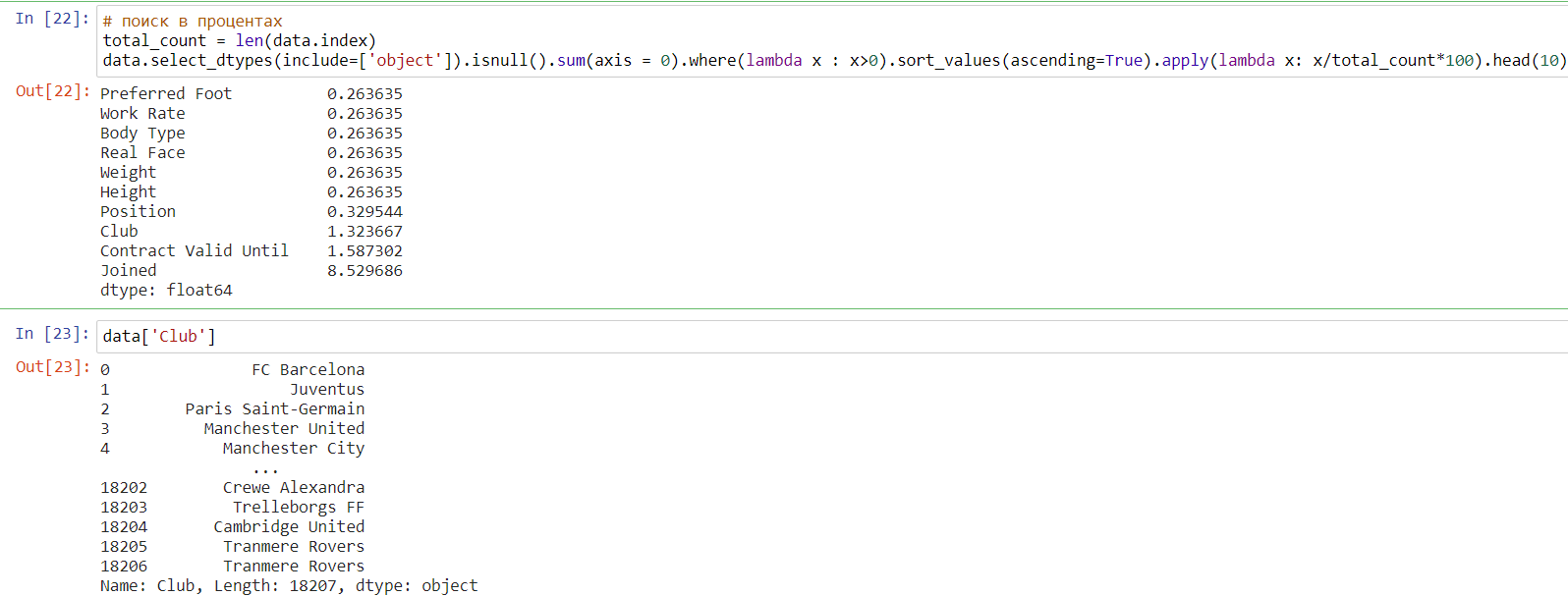
**Используем SimpleImputer для заполнения пропусков**

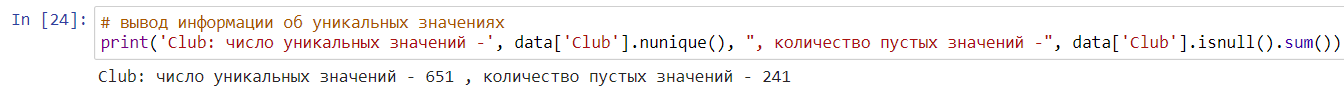




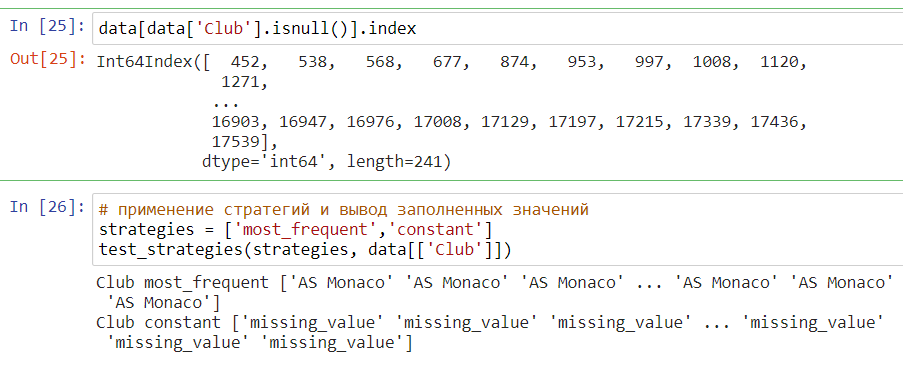


**Обработка пропусков в категориальных данных**

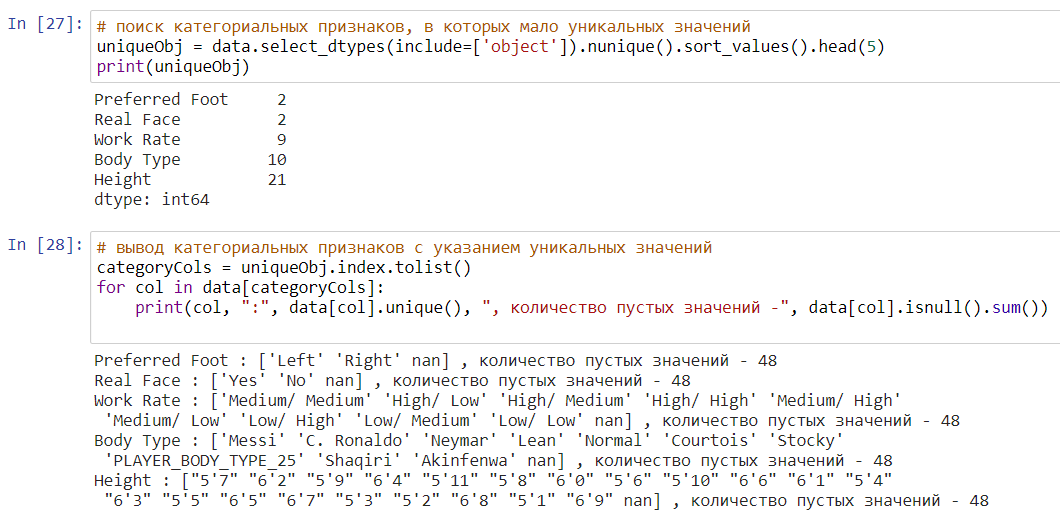


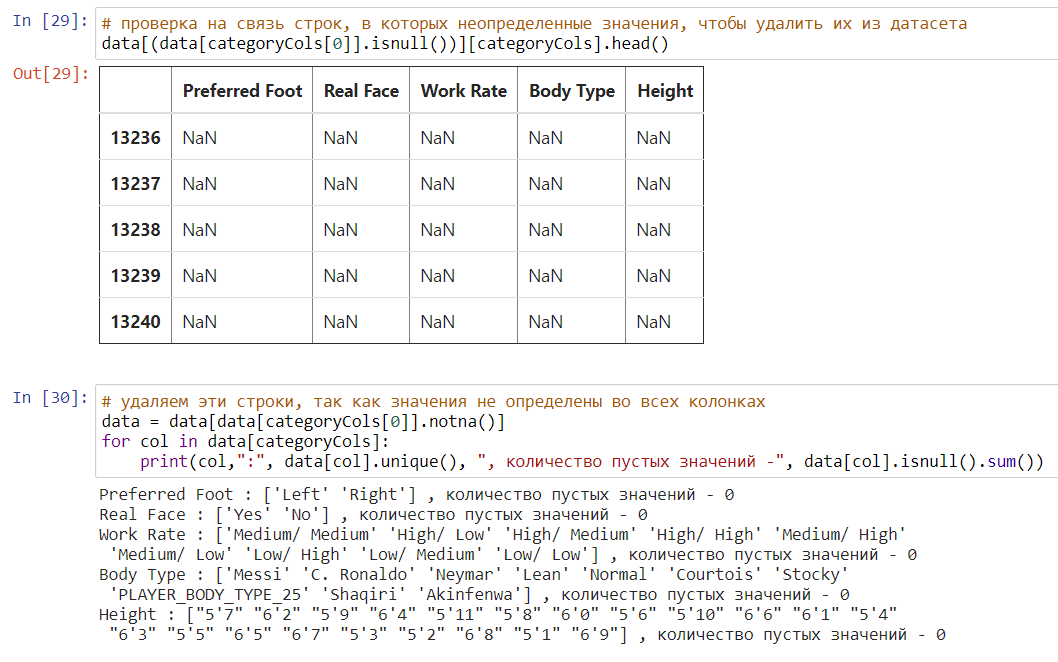


**Заполнение пропусков в данных**

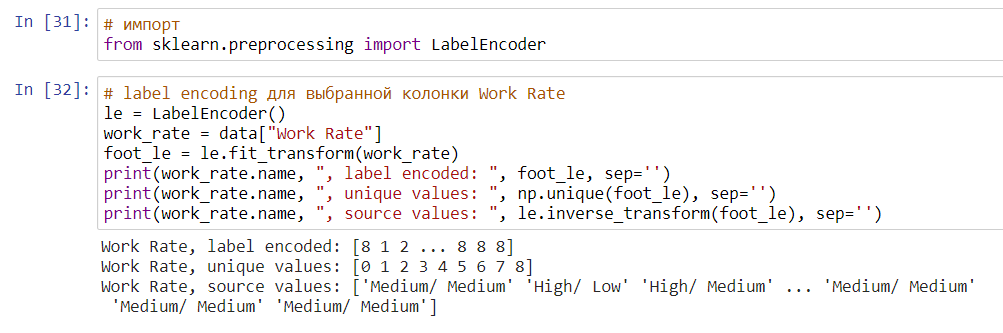


**Преобразование категориальных признаков**

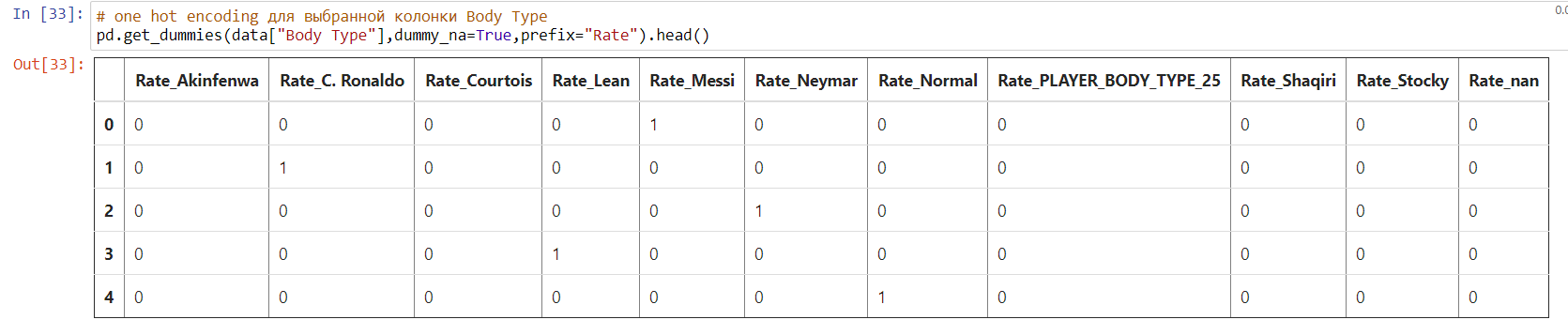




**Label encoding:**

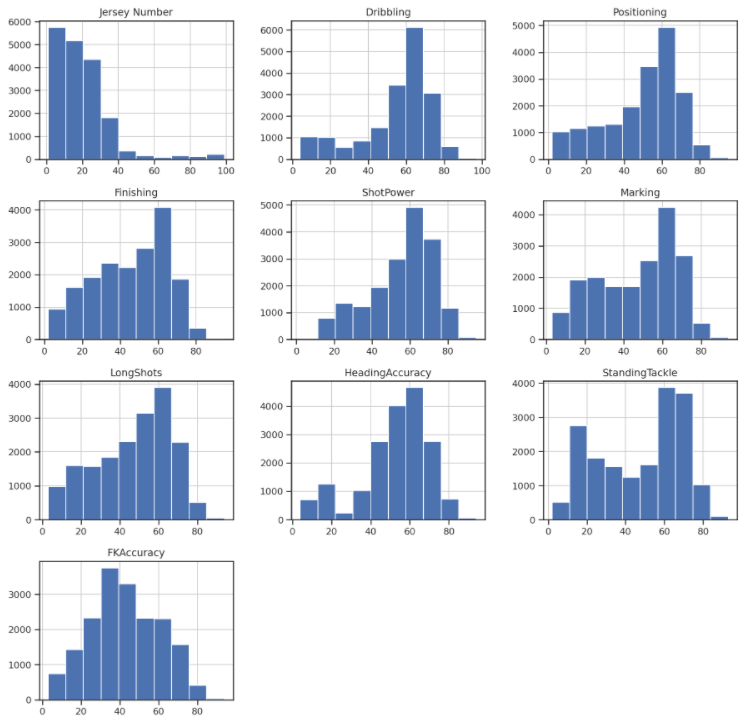


**One-hot encoding:**



**Масштабирование данных**





**Масштабирование различными признаками**

