МГТУ им. Н. Э. Баумана, кафедра ИУ5

курс “Технология машинного обучения”

Рубежный контроль №1

# «[Разведочный анализ и обработки данных.](https://github.com/ugapanyuk/ml_course/wiki/LAB_EDA_VISUALIZATION)»

ВЫПОЛНИЛ:

Бакланов Никита

Группа: ИУ5-61Б

ПРОВЕРИЛ:

Гапанюк Ю.Е.

Москва 2020г.

**Вариант:**

| **Номер варианта** | **Номер задачи** | **Номер набора данных, указанного в задаче** |
| --- | --- | --- |
| 2 | 1 | 2 |

**Задание №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?

**Дополнительные требования по группам:**

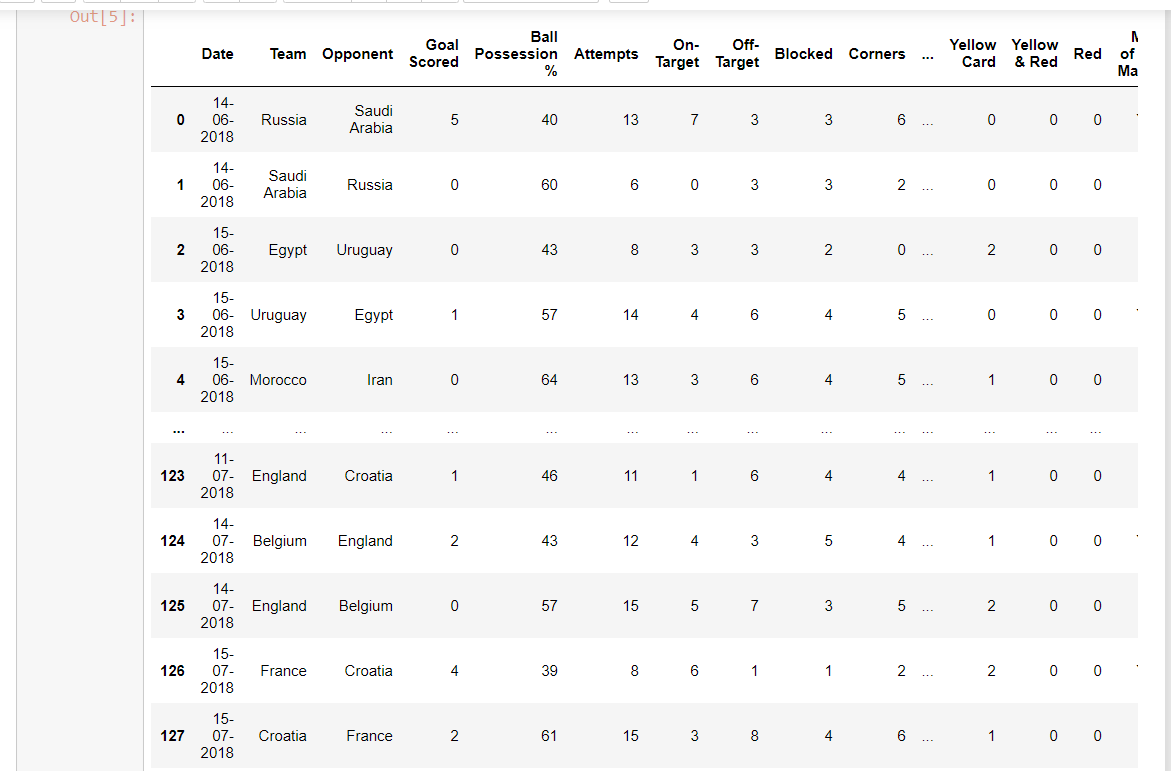
* Для студентов групп ИУ5-61Б, ИУ5Ц-81Б - для пары произвольных колонок данных построить график "Диаграмма рассеяния".

**Набор данных:**

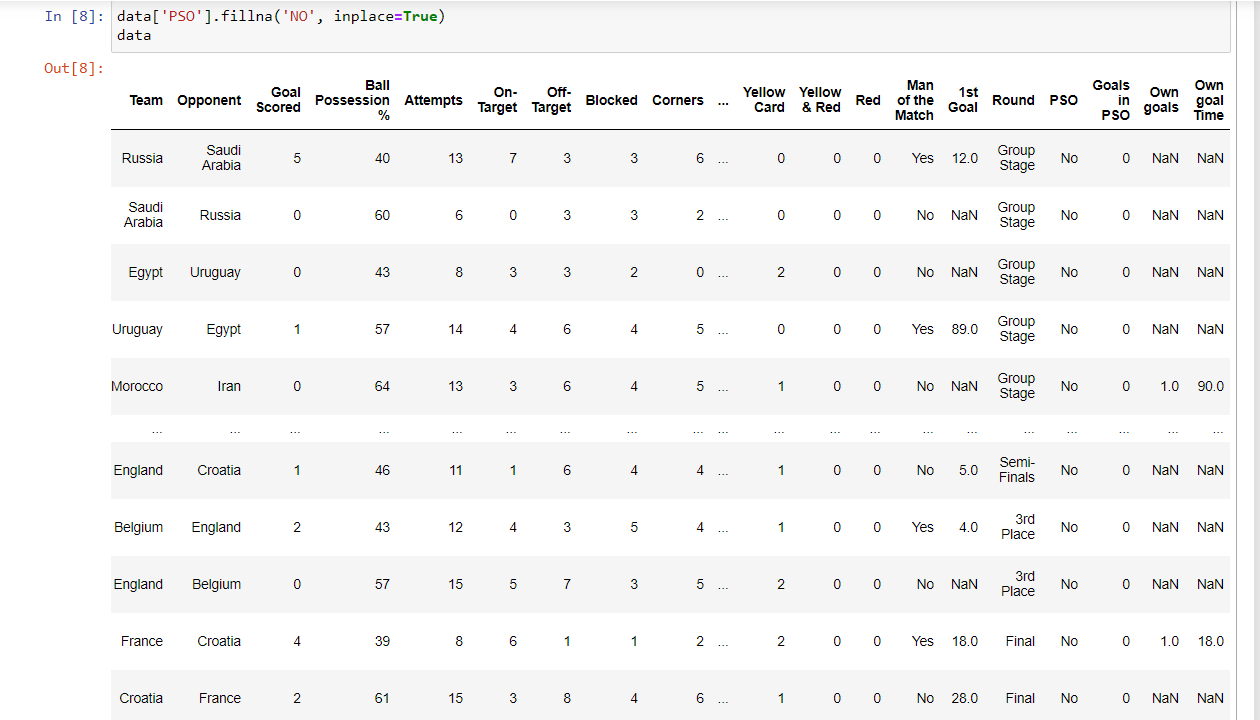
<https://www.kaggle.com/lava18/google-play-store-apps>

**Выполненная работа:**





Обработка пропусков в данных для категориального признака:



Обработка пропусков в данных для количественного признака:

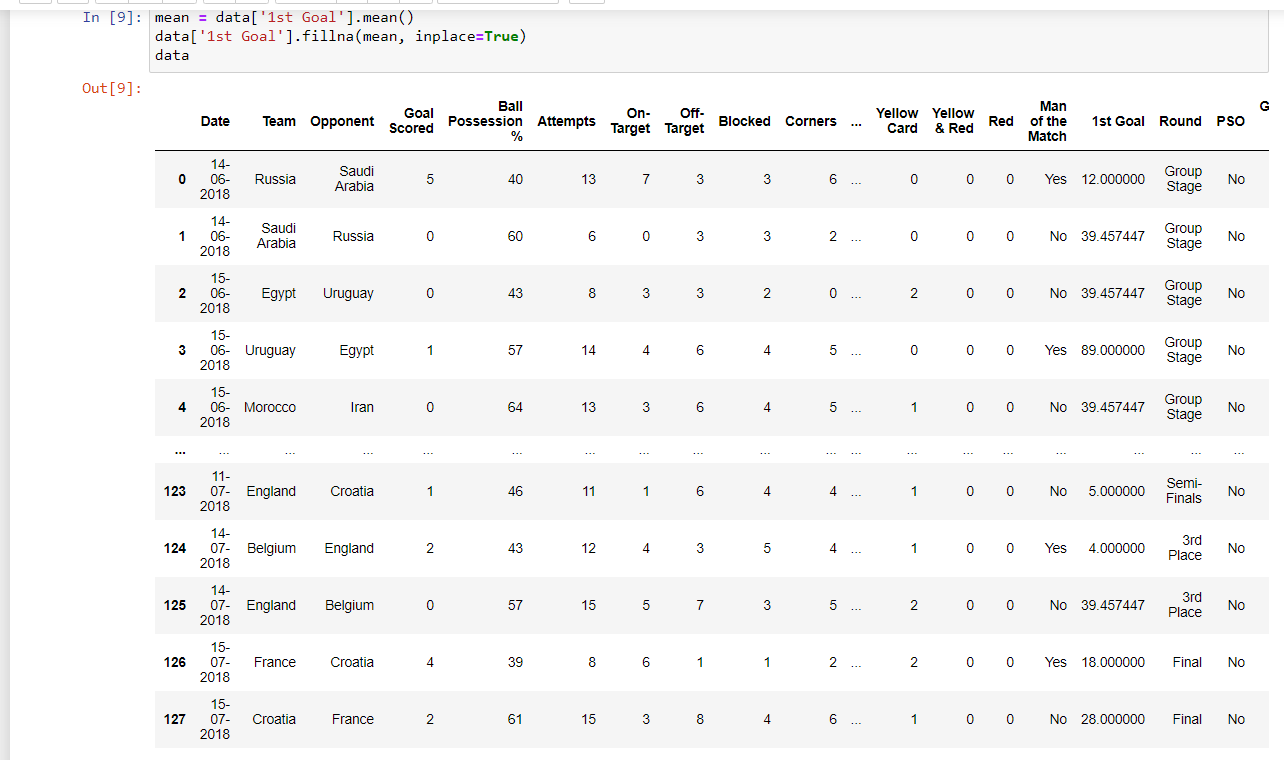
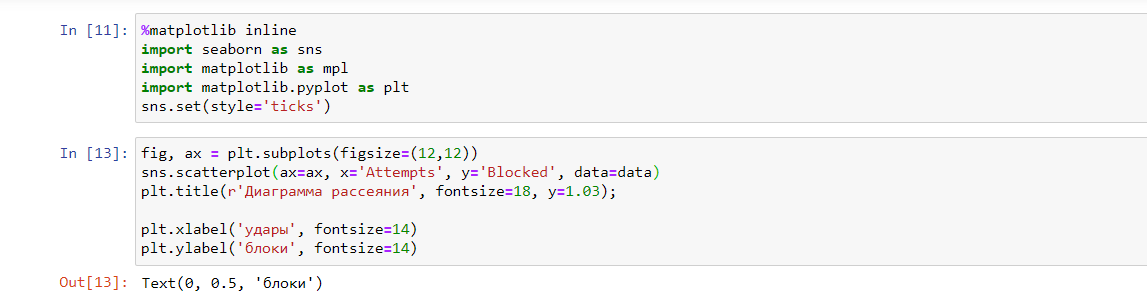
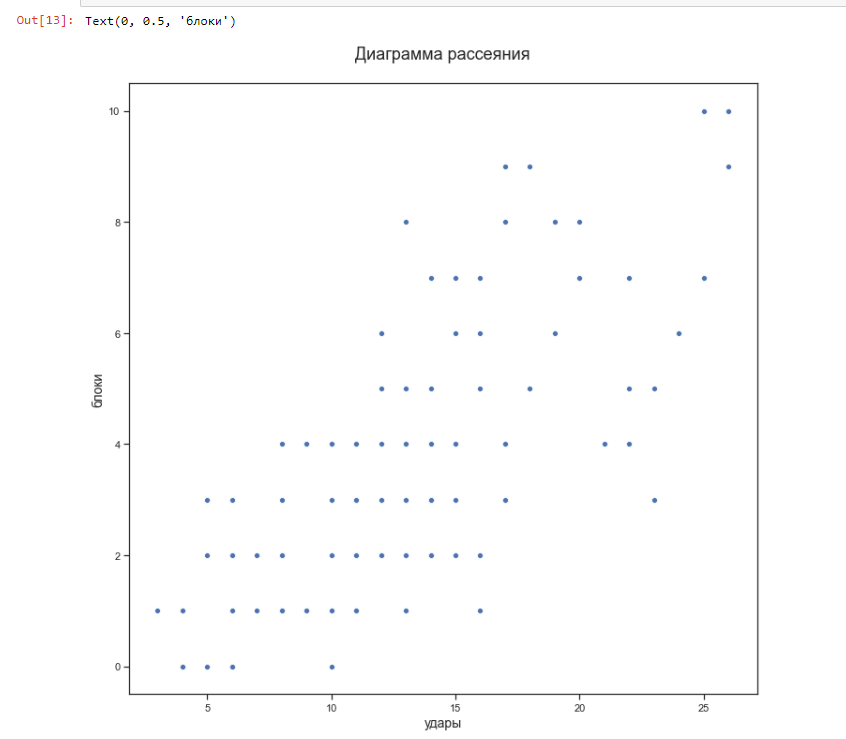


Диаграмма рассеяния:





Ссылка на гитхаб: https://github.com/nikbizkit