

Блок

Feature Engineering

Занятие № 4

Поиск выбросов и генерация новых признаков

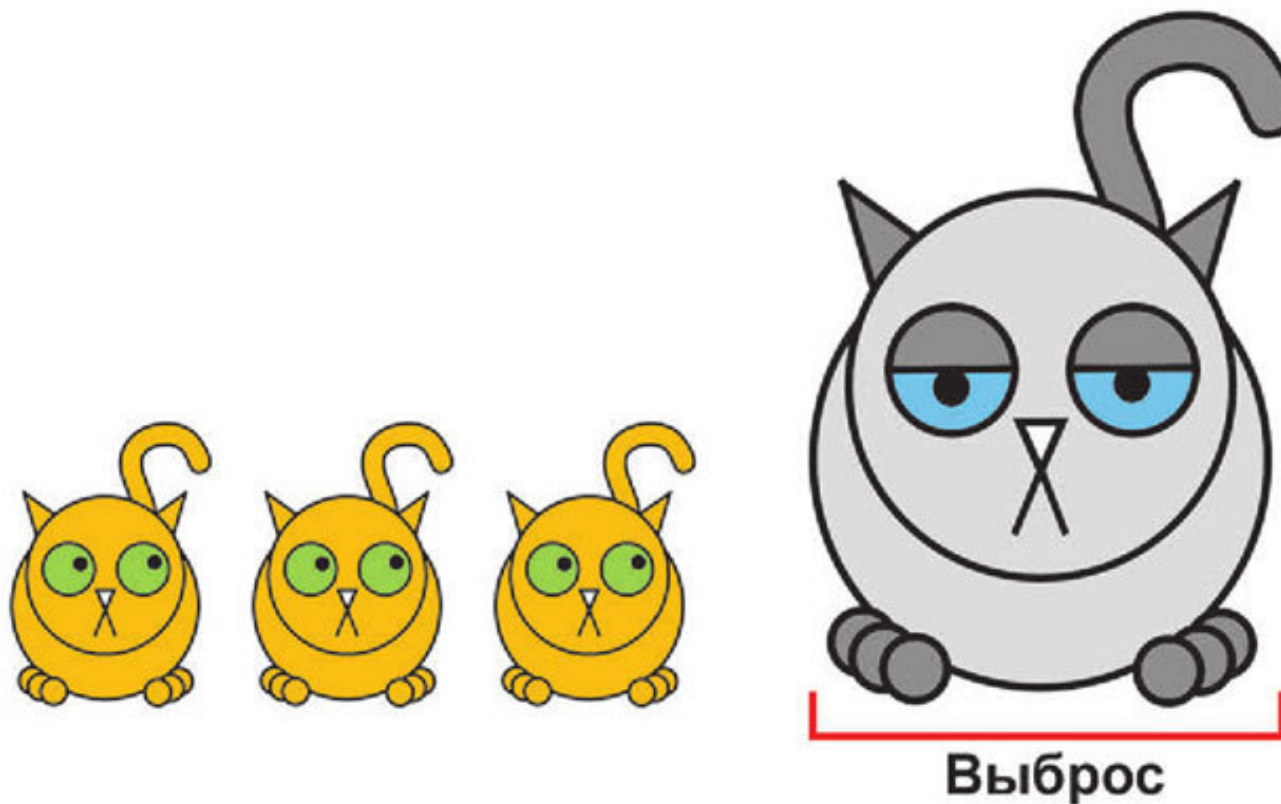
- Научимся искать выбросы
- Продолжим развивать навыки использования средств pandas и numpy для решения проблем в данных
- Вспомним всё, чему научились за последние занятия

Цели занятия

Что такое выброс

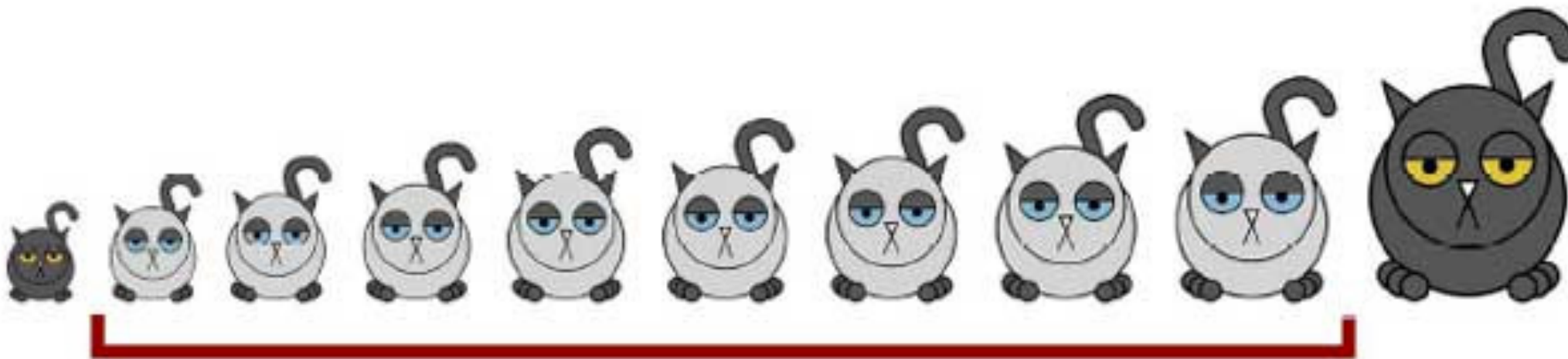
Выброс

Результат измерения, не подпадающий под общее распределение



<https://www.litres.ru/vladimir-savelev-10569666/statistika-i-kotiki-28731109/chitat-onlayn/>

Выброс



Котики для усеченного среднего

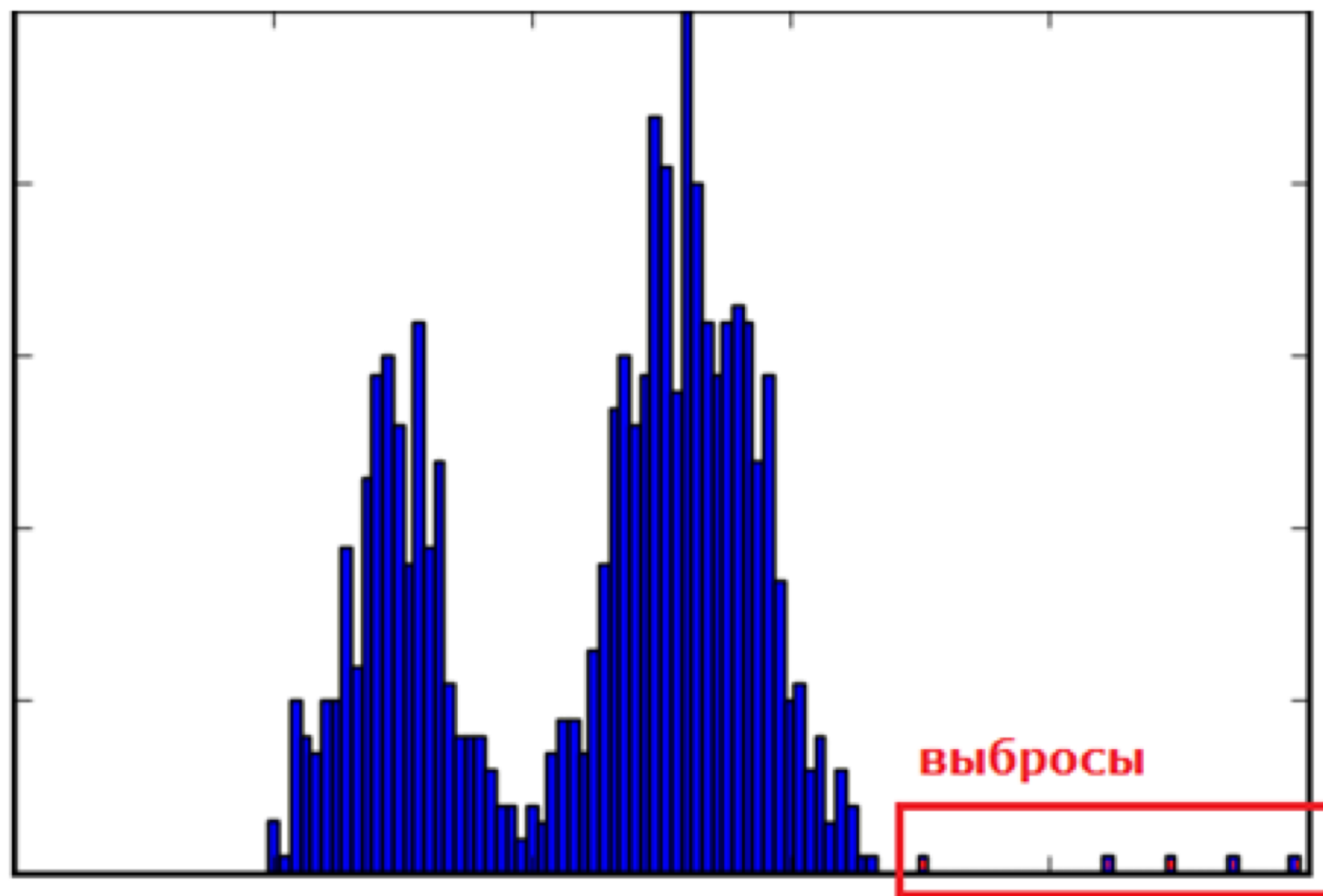
<https://www.litres.ru/vladimir-savelev-10569666/statistika-i-kotiki-28731109/chitat-onlayn/>

Что делаем?

Что делаем?

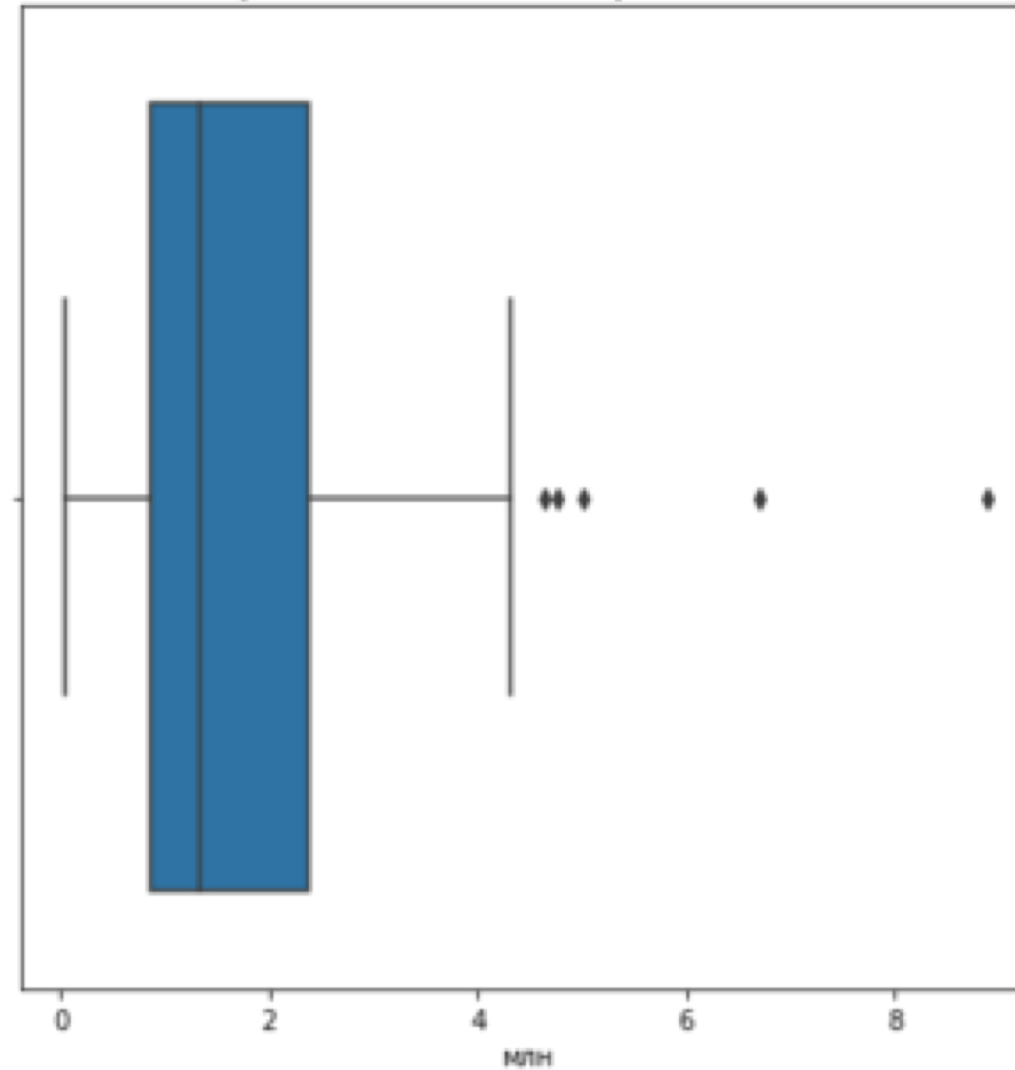
- Смотрим на распределение
- Смотрим на статистические показатели (квантили, медиана, средняя)

Что делаем?



Что делаем?

Количество проживающих людей в регионах России, 1990г.



ПРАКТИЧЕСКАЯ ЧАСТЬ

ВОПРОСЫ