

Кластеризация пользователей

Подготовил Дитятев Иван

Intro

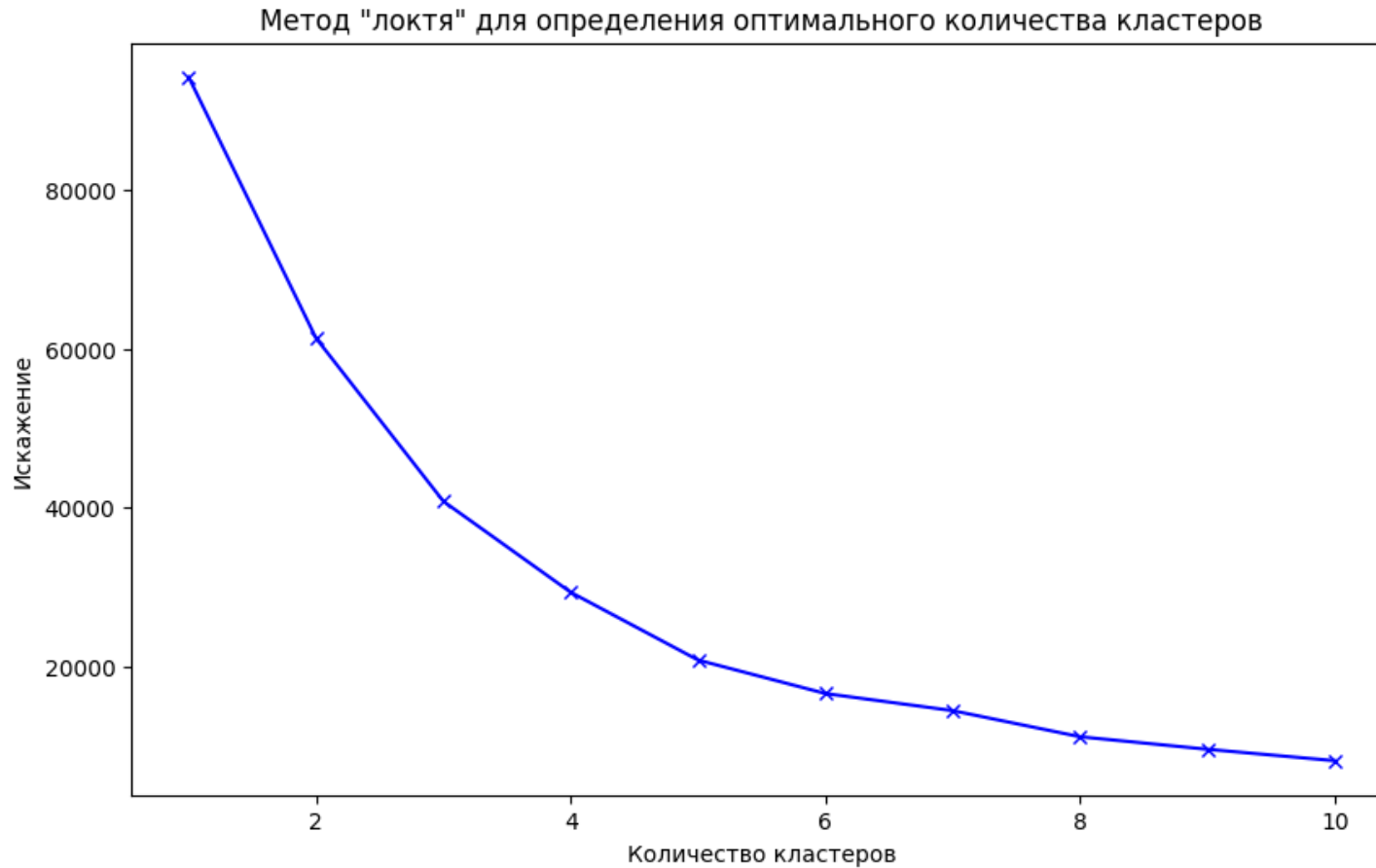
Цель:

- Разделить пользователей на кластеры.
- Предлагаемые кластеры:
 - - активный донор,
 - - спящий донор,
 - - реактивированный донор,
 - - потерянный.

Задачи:

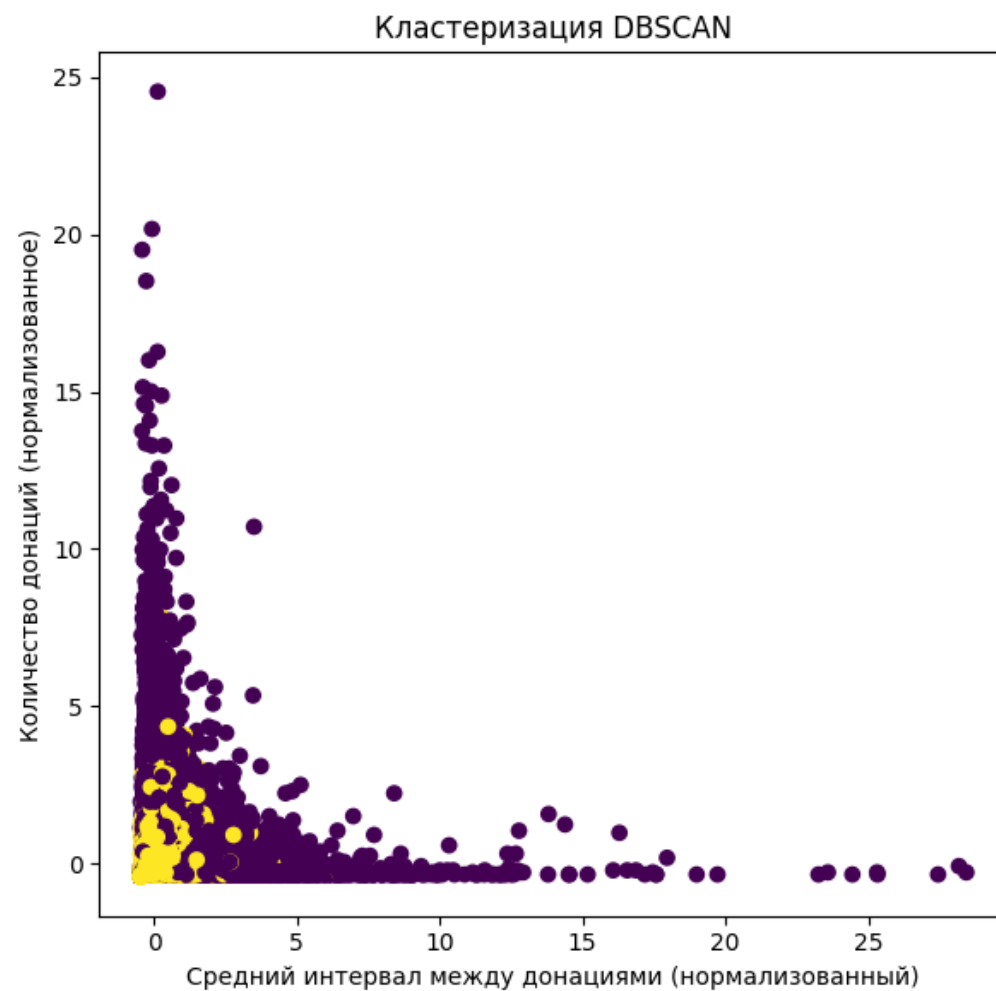
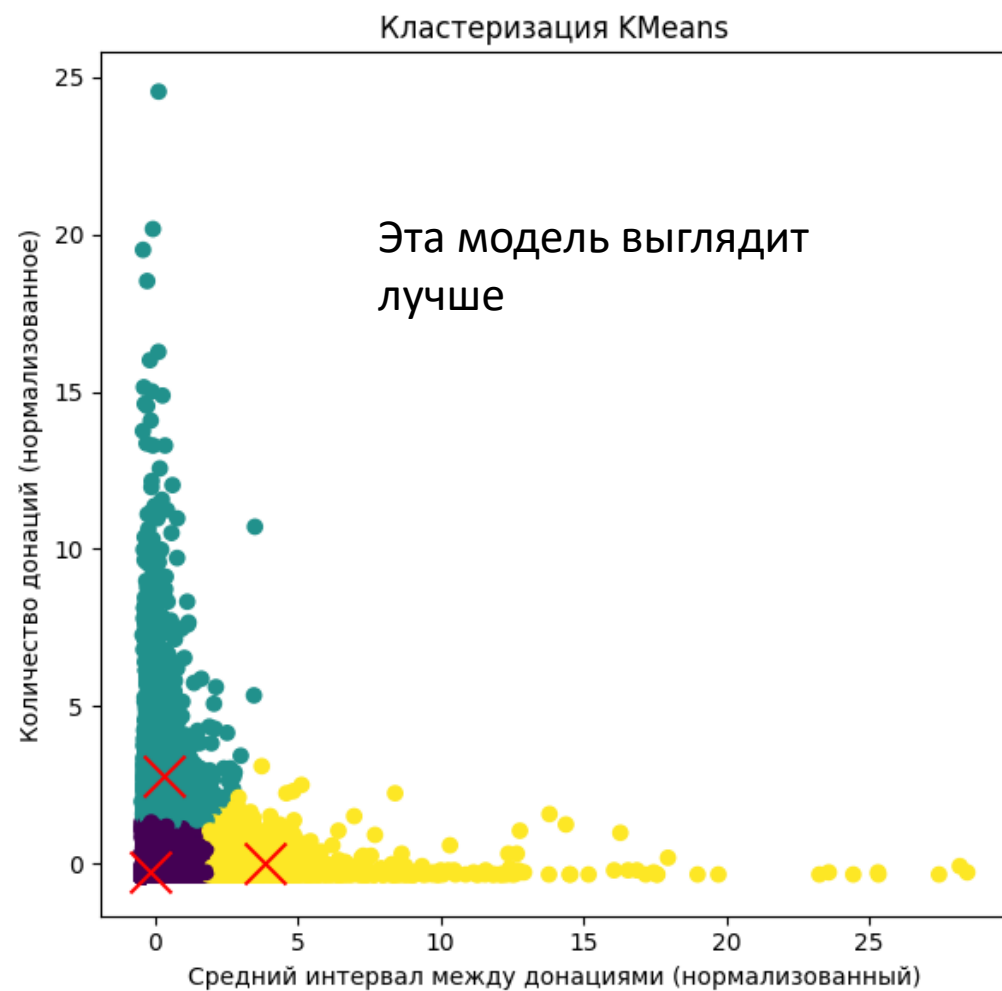
- - Предобработка данных.
- - Определение оптимального числа кластеров.
- - Разделение пользователей на кластеры.
- - Определение четких правил, как отнести пользователя к тому или иному кластеру.
- - Исследовательский анализ данных.
- - Создание презентации.

Оптимальное кол-во кластеров



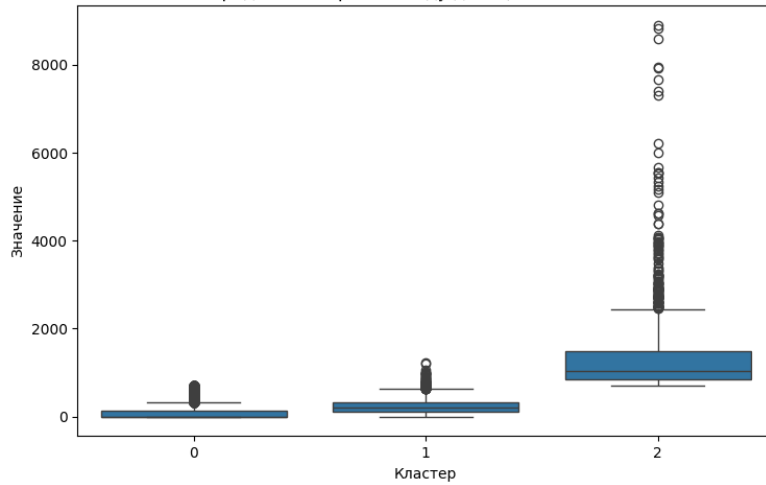
Оптимальное кол-во = 3
кластера

Кластеризация

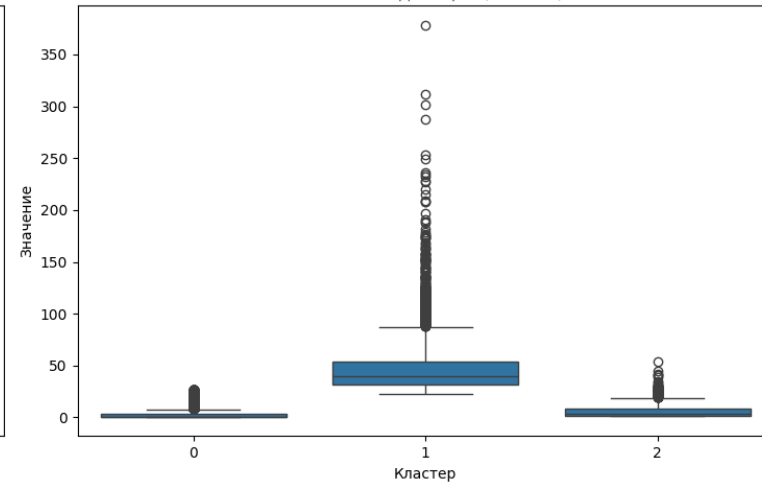


Распределение пользователей по кластерам

Средний интервал между донациями (KMeans)



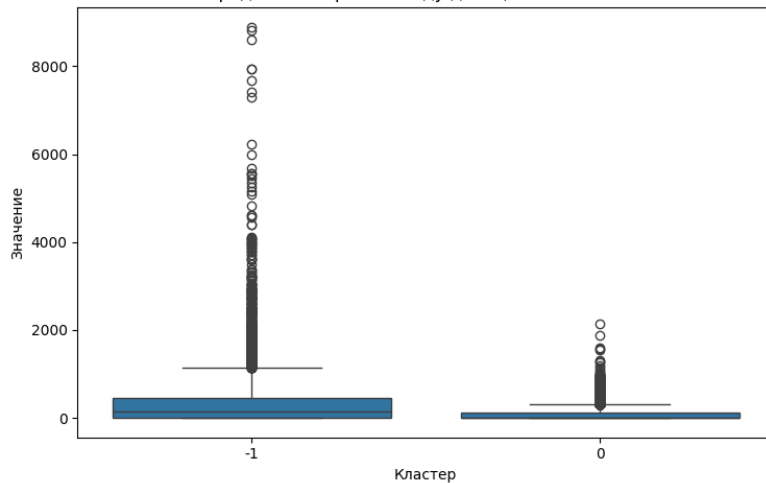
Количество донаций (KMeans)



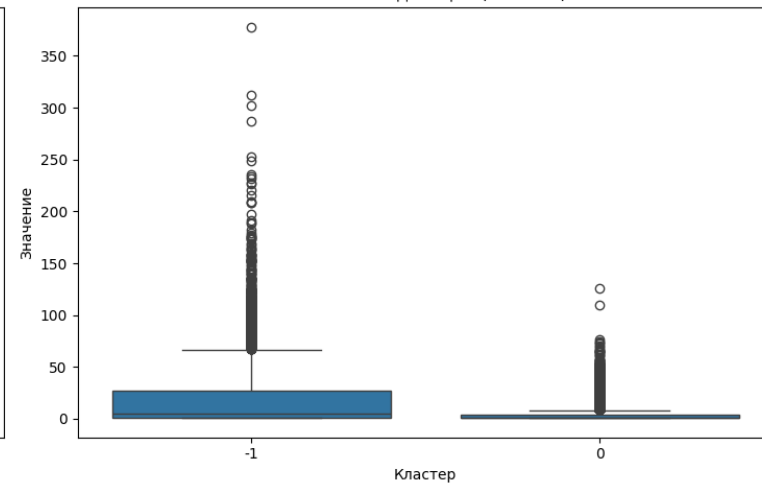
Активный донор: Кластер 1 (Много донаций, короткий период между ними)

Спящий донор: Кластер 2 (Мало донаций, длительный период)

Средний интервал между донациями (DBSCAN)



Количество донаций (DBSCAN)



Реактивированный донор: Кластер 0 (Мало донаций, короткий период между ними)

Потерянный донор: Кластер 3 (Выбросы и те, кто в танке)

Определение правил распределения по кластерам – ВЫВОД

		Значение									
		count	mean	std	min	25%	50%	75%	max	lb	ub
Кластер_kmeans	Показатель										
0	Количество донаций	42185.0	3.704966	5.109585	1.000000	1.000000	1.000000	4.000000	27.000000	0	8.500000
	Средний интервал между донациями	42185.0	88.692985	137.512578	0.000000	0.000000	0.000000	131.125000	712.000000	0	327.812500
1	Количество донаций	3405.0	49.456388	29.689179	23.000000	32.000000	40.000000	54.000000	378.000000	0	87.000000
	Средний интервал между донациями	3405.0	237.872789	169.272337	0.000000	112.515152	193.820000	321.425532	1220.928571	0	634.791103
2	Количество донаций	1544.0	7.102332	7.013081	2.000000	2.000000	4.000000	9.000000	54.000000	0	19.500000
	Средний интервал между донациями	1544.0	1331.854354	896.781507	698.421053	832.420455	1026.333333	1477.099359	8898.000000	0	2444.117716

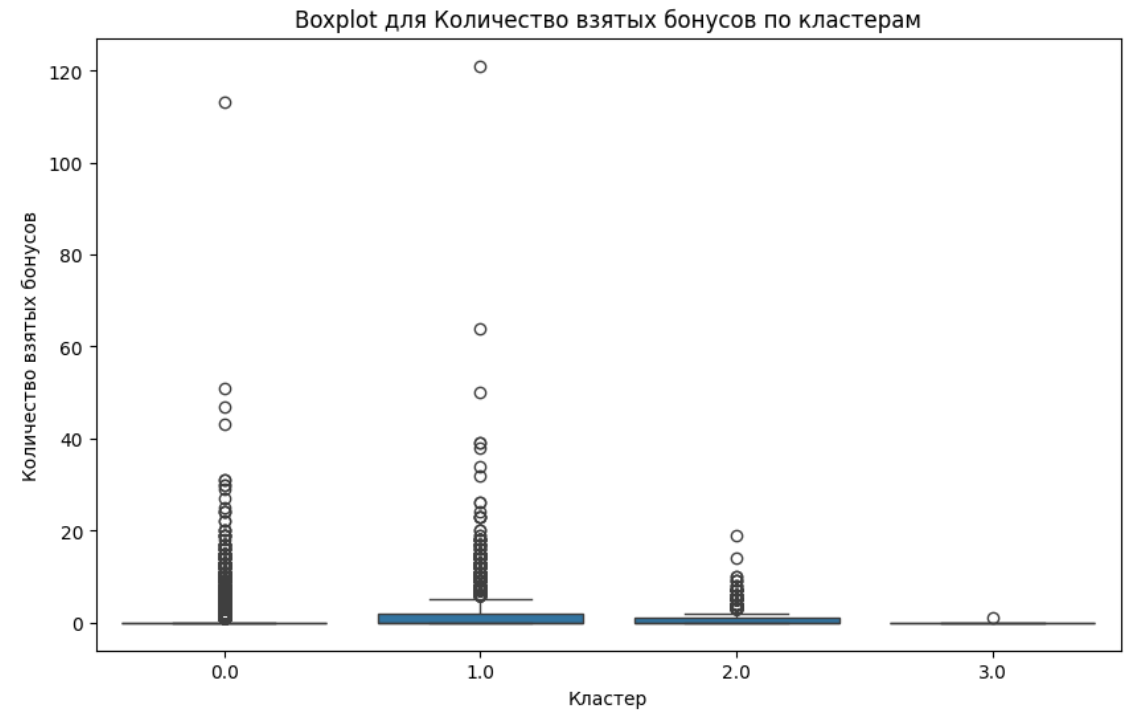
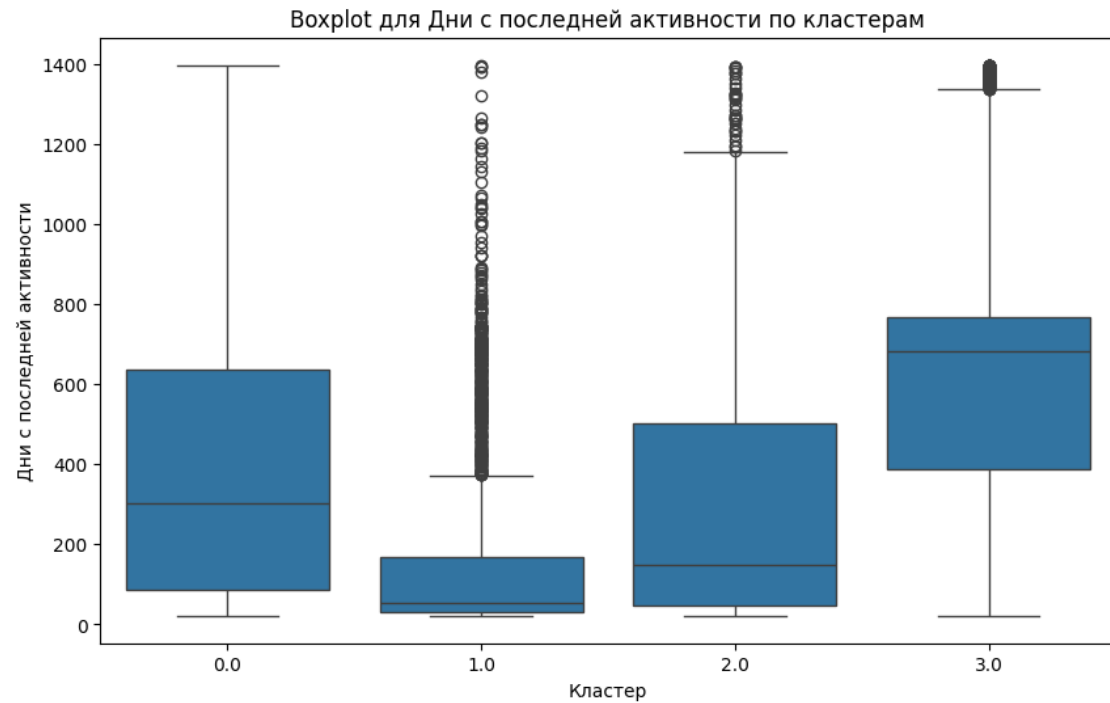
Пример запроса:

```
SELECT *,
CASE
WHEN Количество_донаций BETWEEN 0 AND 8.5 AND Средний_интервал_между_донациями BETWEEN 0 AND 327.8125 THEN 0
WHEN Количество_донаций BETWEEN 0 AND 87 AND Средний_интервал_между_донациями BETWEEN 0 AND 634.791103 THEN 1
WHEN Количество_донаций BETWEEN 0 AND 19.5 AND Средний_интервал_между_донациями BETWEEN 0 AND 2444.117716 THEN 2
ELSE 3
END AS Кластер
FROM some_table;
```

Проверка наличия статически значимых различий между кластерами t-test/ Mann-Whitney U Test (для числовых колонок), χ^2 (для категориальных).

- Проверка показала наличие статистически значимых различий между кластерами.

Некоторые прикольные инсайты с исследовательского анализа



Некоторые прикольные инсайты с исследовательского анализа

