

Анализ оттока клиентов банка «Метанпром»

ВЫПОЛНИЛ: КОЛОБАШКИН Н.А.

СТУДЕНТ 89 КОГОРТЫ КУРСА «АНАЛИТИК ДАННЫХ»

Цель исследования

В банке последнее время драматически повысился отток, это начинает становиться проблемой, которая угрожает существованию банка.

Необходимо выделить сегменты отточных клиентов, чтобы они были однородными, высокоотточными и покрывали максимум проблемных зон банка.

Информация необходима менеджеру маркетингового отдела для построения грамотной маркетинговой политики по удержанию отточных клиентов.

Загрузка данных и подготовка к исследованию

Открытие файла с данными и изучение общей информации

- Датасет состоит из 10 000 строк и 12 столбцов
- В столбцах *age* и *balance* имеются пропущенные значения

Предобработка данных

- Названия столбцов заменены на интуитивно понятные и приведены к нижнему регистру
- Обработаны пропущенные значения.
Пропуски в столбце *age* получили классификацию MCAR и соответствующие им строки были удалены.
Пропуски в столбце *balance* получили классификацию MNAR. Соответствующие им строки было решено оставить, т.к. нет основания удалять их.
- Для столбцов *score*, *EST_SALARY* и *age* изменен тип данных.
- Обработаны выбросы
- Произведена проверка на дубликаты
- Выполнено кодирование переменных

Проверка полученного датасета

Исследовательский анализ данных

Сравнение портретов типичных клиентов

Общий портрет типичного пользователя банка:

- Мужчина;
- Проживает в Ярославле;
- Возраст клиента 40 - 43 года;
- Имеет 850 баллов кредитного скоринга;
- Оценка стоимости имущества равна 3;
- На балансе находится от 524 000 до 827 000;
- Пользуется двумя продуктами банка;
- Имеет кредитную карту банка;
- Был активен в последнее время
- Имеет оценочный доход от 120 000 до 147 000;
- Не имеет признаков оттока.

Исследовательский анализ данных

Сравнение портретов типичных клиентов

Отточный	Параметр	Лояльный
Более высокие баллы кредитного скоринга. В среднем 863-865	score	Более низкие баллы кредитного скоринга. В среднем 845-848
66,4 % проживает в Ярославле	city	58,3 % проживает в Ярославле
65 % мужчины	gender	53,3 % женщины
Возраст равен 39-40 годам	age	Возраст равен 40 - 43 годам
31,6 % имеют оценку 5	equity	30,4 % имеют нулевую оценку
Среднее значение равно 1 134 458 Медианное значение равно 783 909	balance	Среднее значение равно 733 032 Медианное значение равно 475 41
53,7 % имеют 2 продукта.	products	50,7 % имеют 2 продукта
55,1 % отточных также имеют кредитку.	credit_card	70,8 % лояльных имеют кредитку.
70,2 % отточных активны.	last_activity	51,7 % лояльных не активны в последнее вре
Среднее значение равно 148 357 Медианное значение равно 125 408	est_salary	Среднее значение равно 147 674 Медианное значение равно 118 227

Исследовательский анализ данных

Распределение параметров

Числовые параметры:

- score
- age
- balance
- est_salary

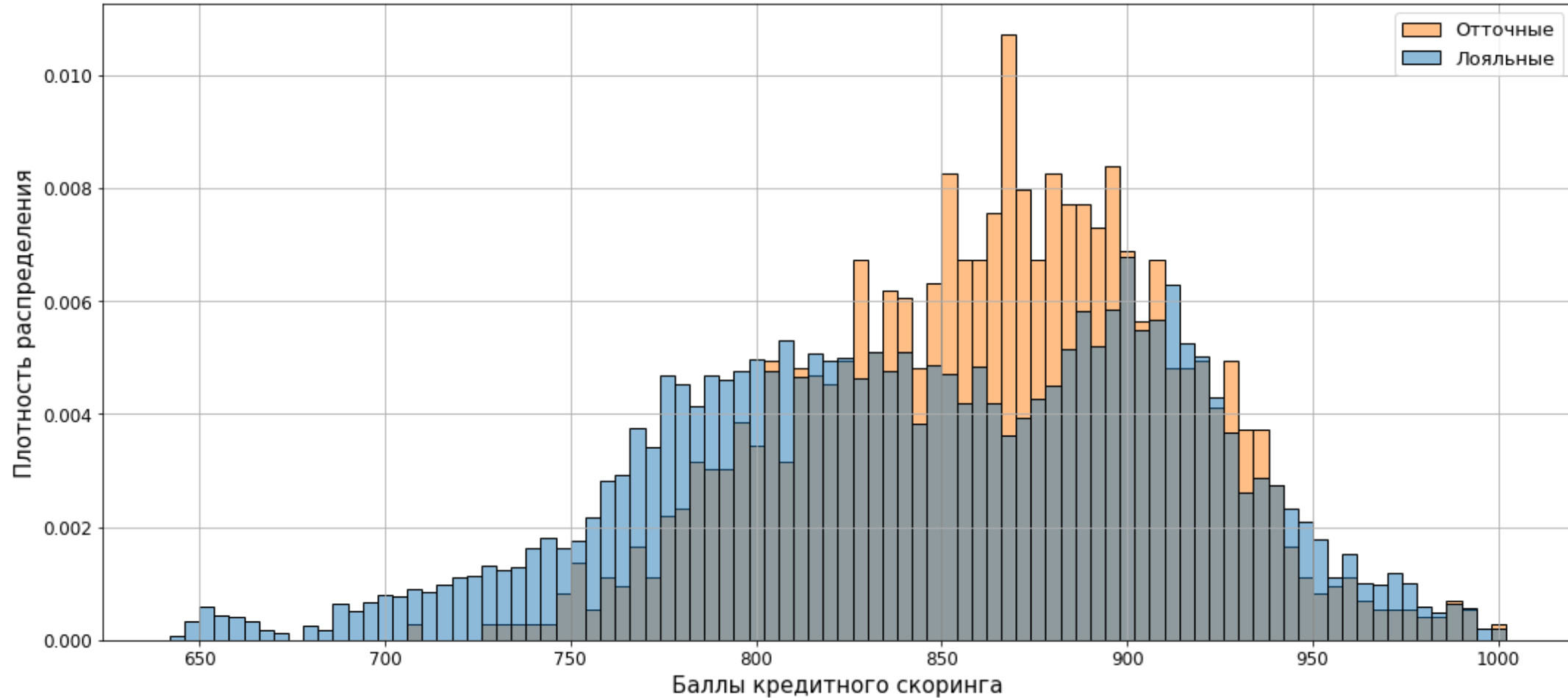
Для числовых параметров построены графики распределения через нормированную гистограмму

Категориальные параметры:

- city
- gender
- Products
- credit_card
- last_activity
- equity

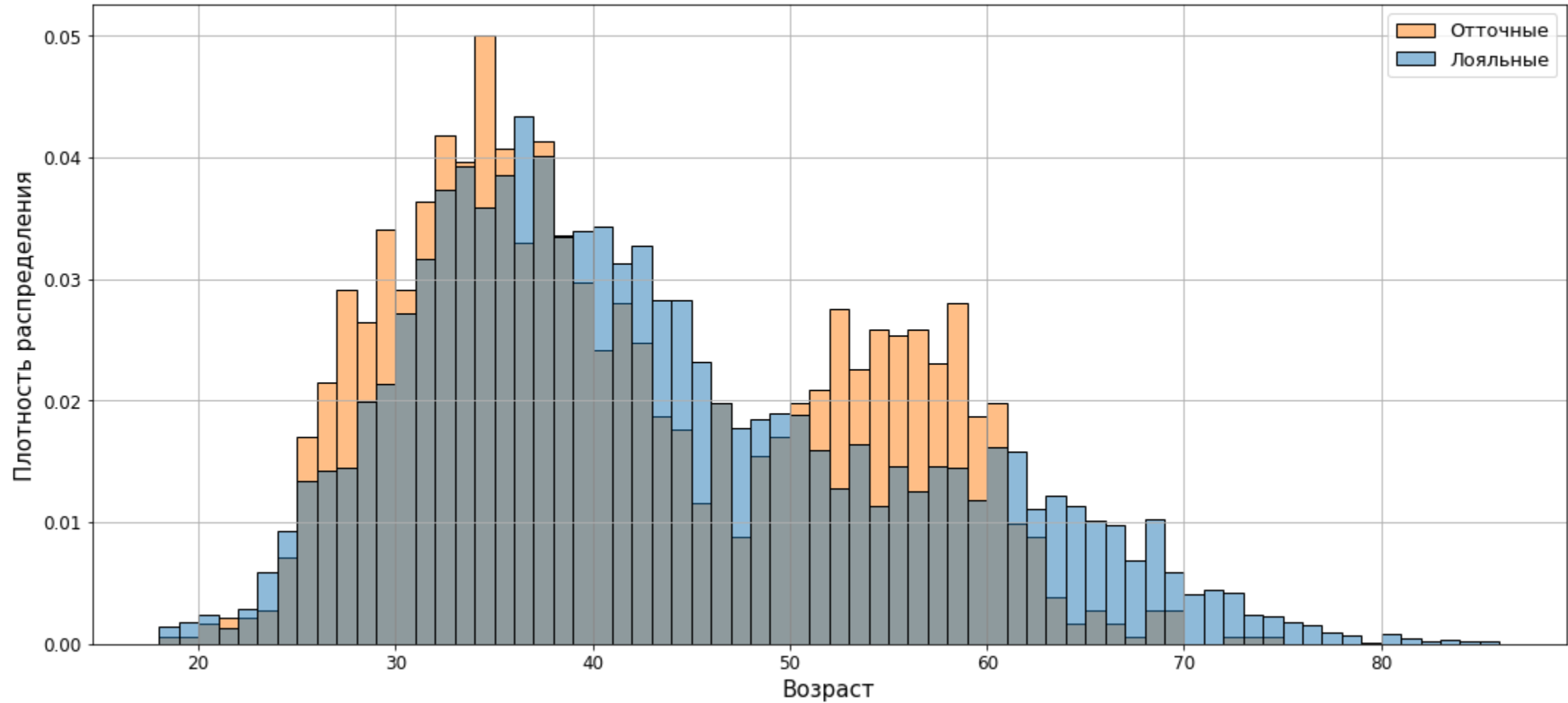
Для категориальных параметров построены графики распределения через нормированную гистограмму

Распределение значений столбца score



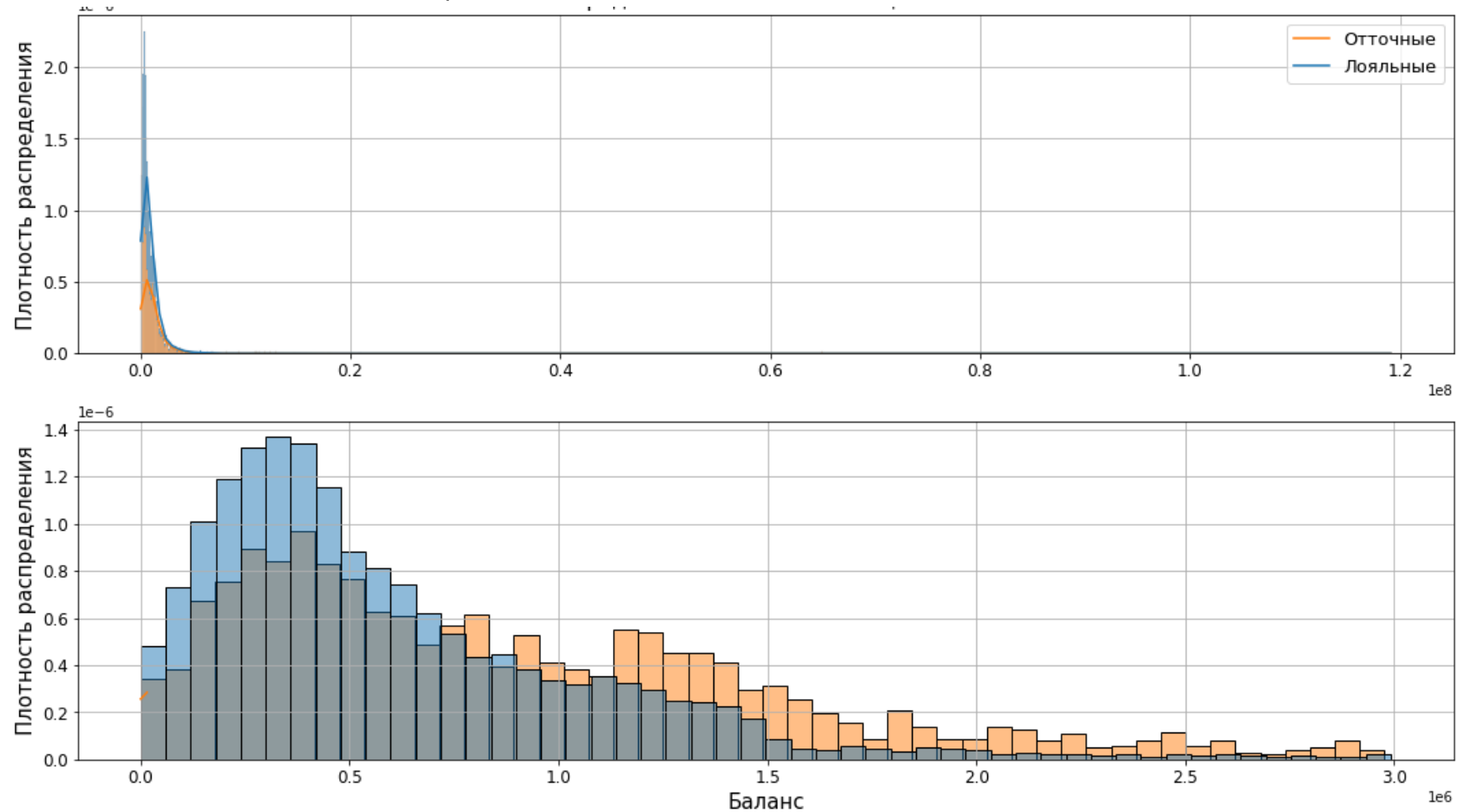
Отток наблюдается для пользователей с двумя промежутками баллов кредитного скоринга - от 830 до 910 и от 930 до 940

Распределение значений столбца *age*



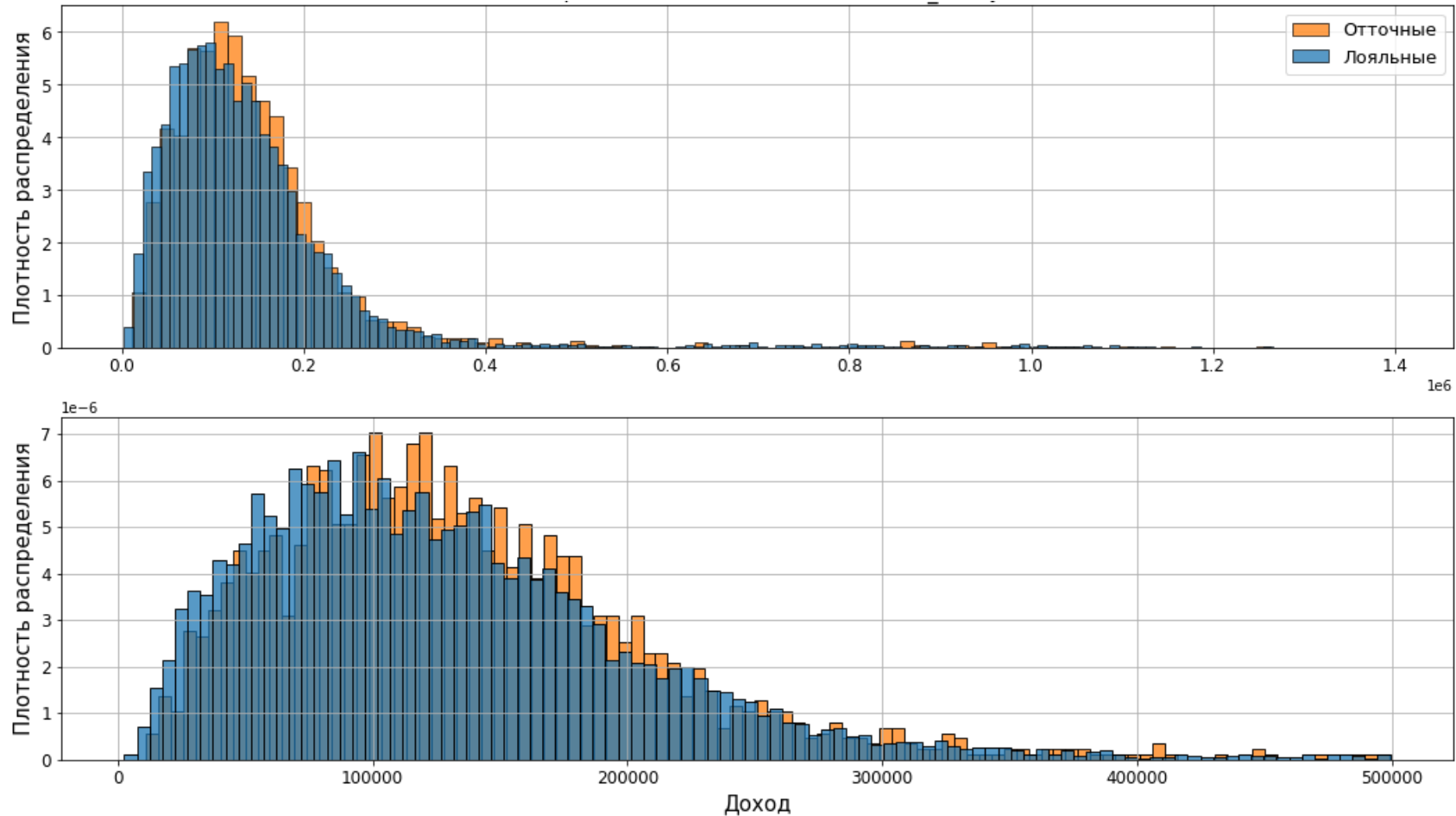
Отток наблюдается для пользователей с двумя возрастными промежутками - с 25 до 38 и с 50 до 61

Распределение значений столбца *balance*



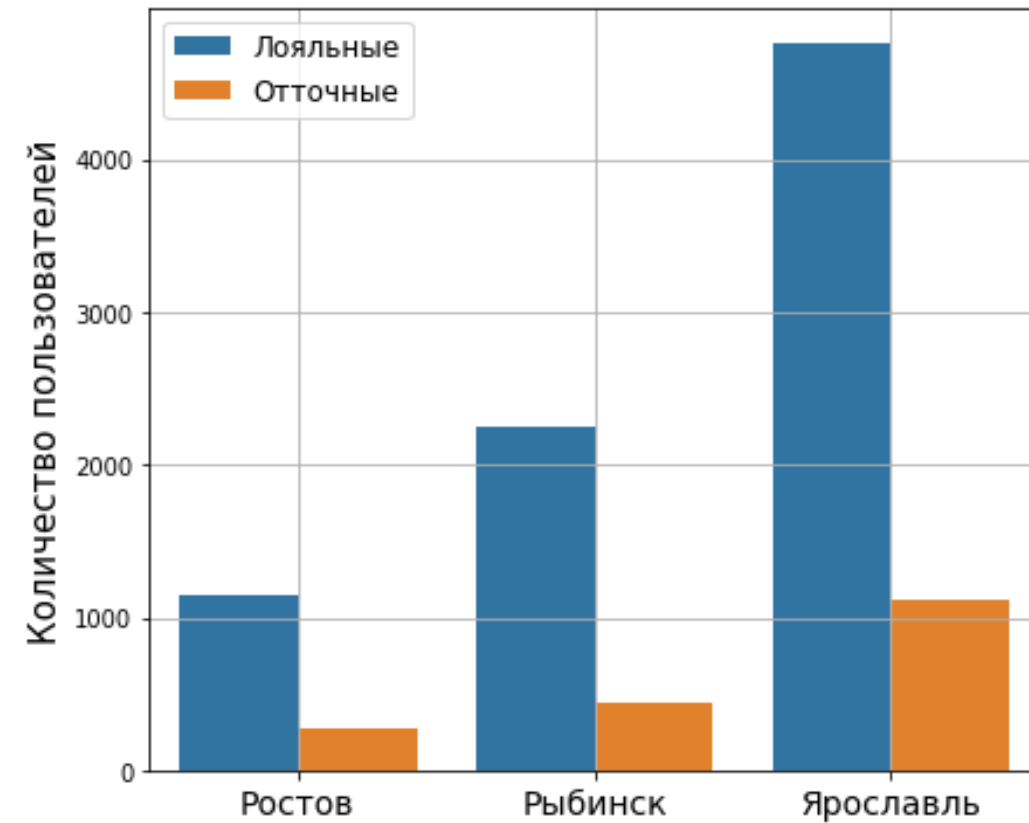
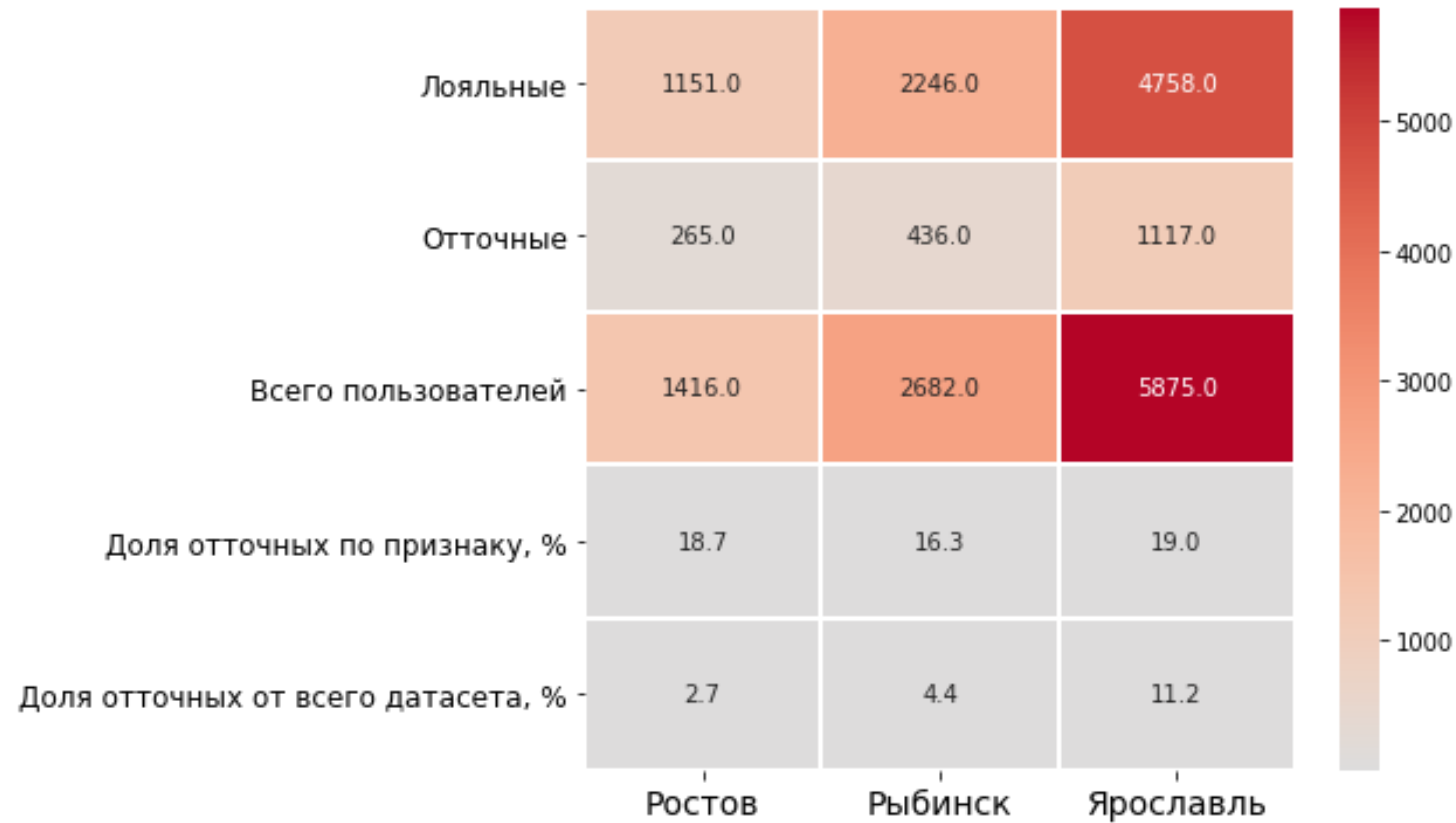
Отток наблюдается для пользователей с балансом на счету
от 750 000 и выше;

Распределение значений столбца *est_salary*



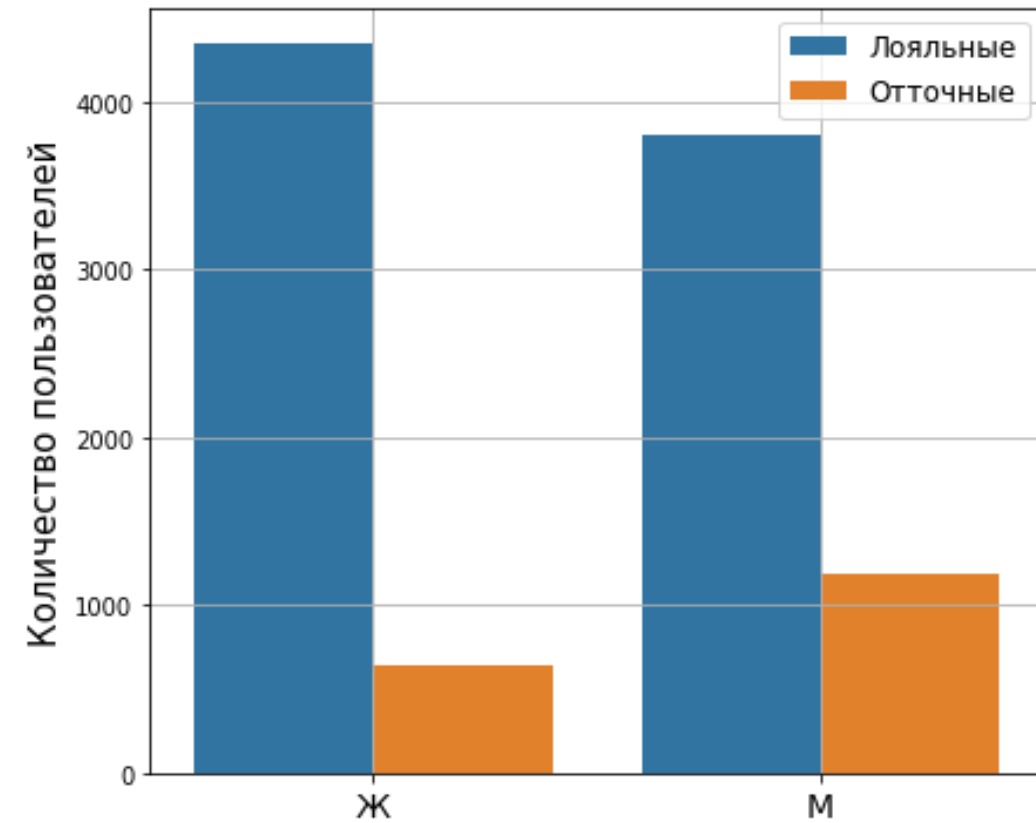
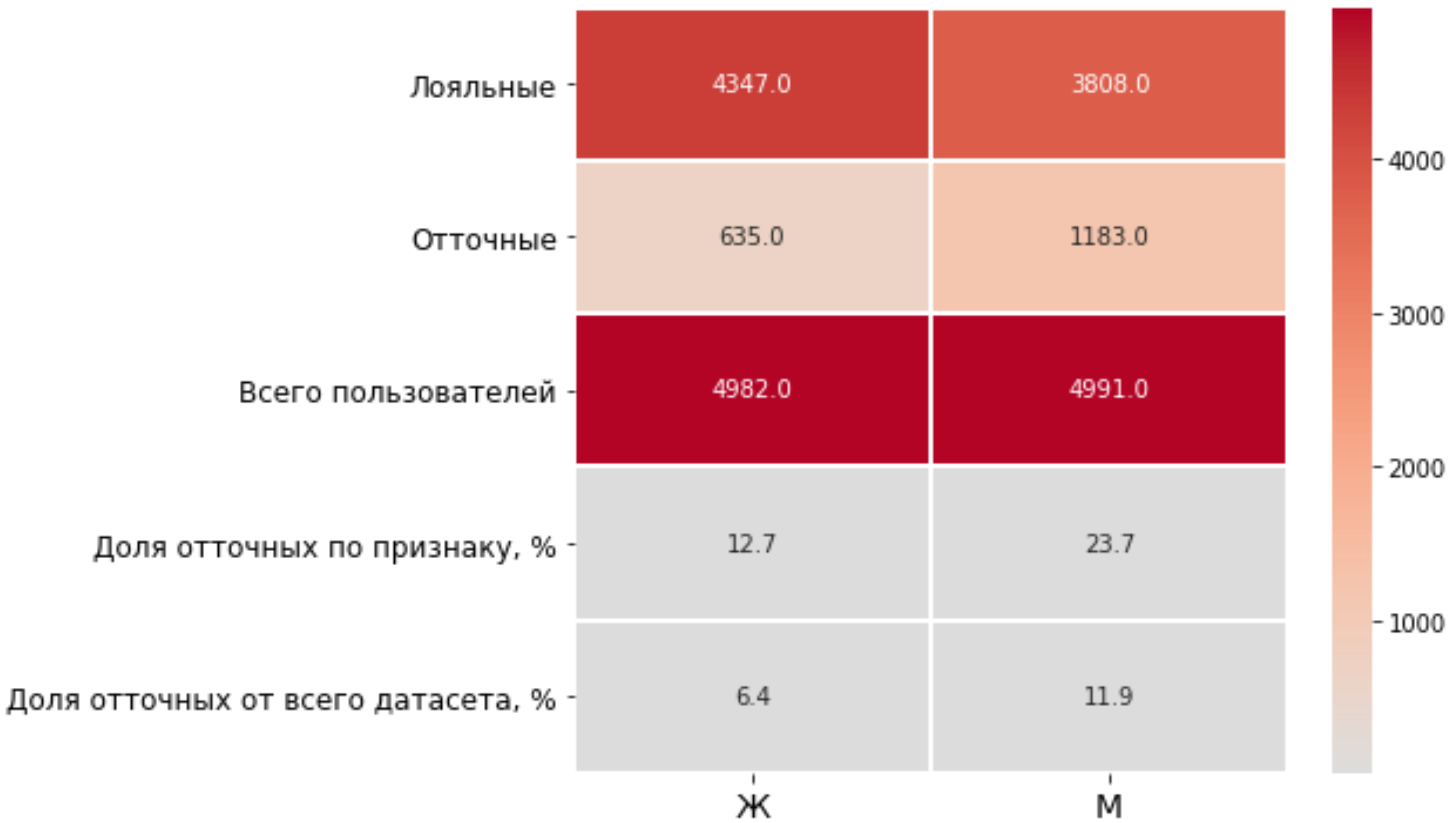
Отток наблюдается для пользователей с оценочным доходом от 100 000 до 220 000 рублей, а также в промежуток от 300 000 до 320 000

Распределение значений по столбцу *sity*



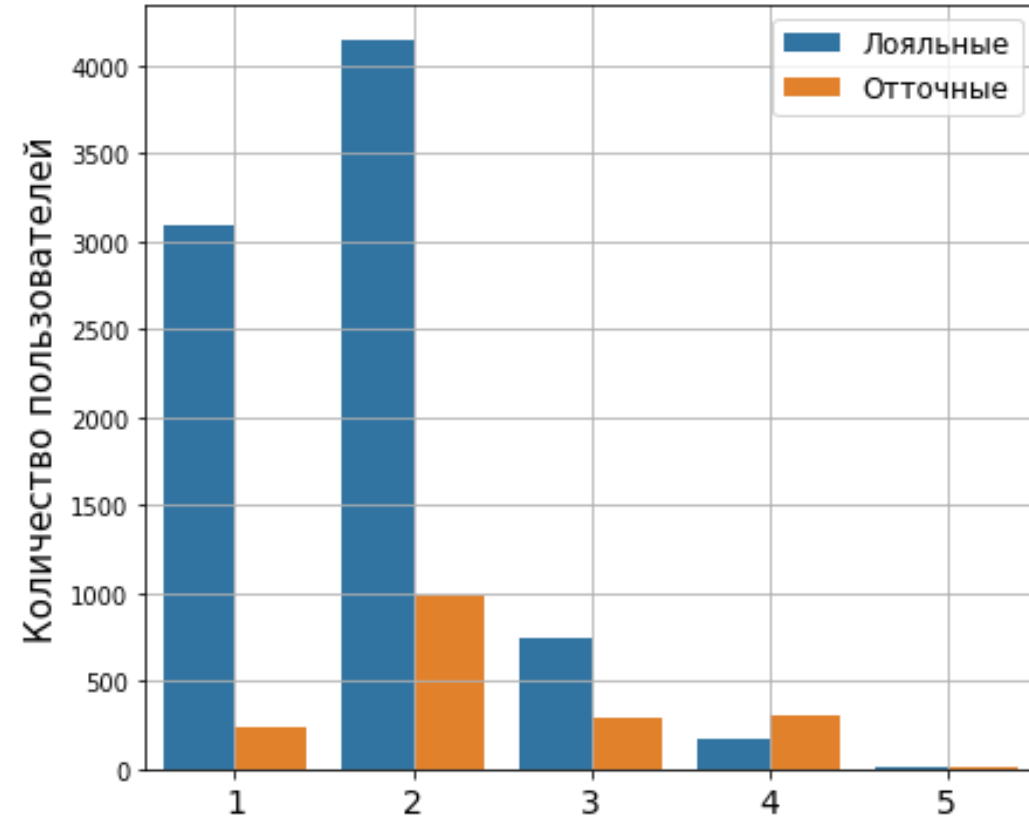
Отток наблюдается преимущественно для пользователей из г. Ярославль

Распределение значений по столбцу *gender*



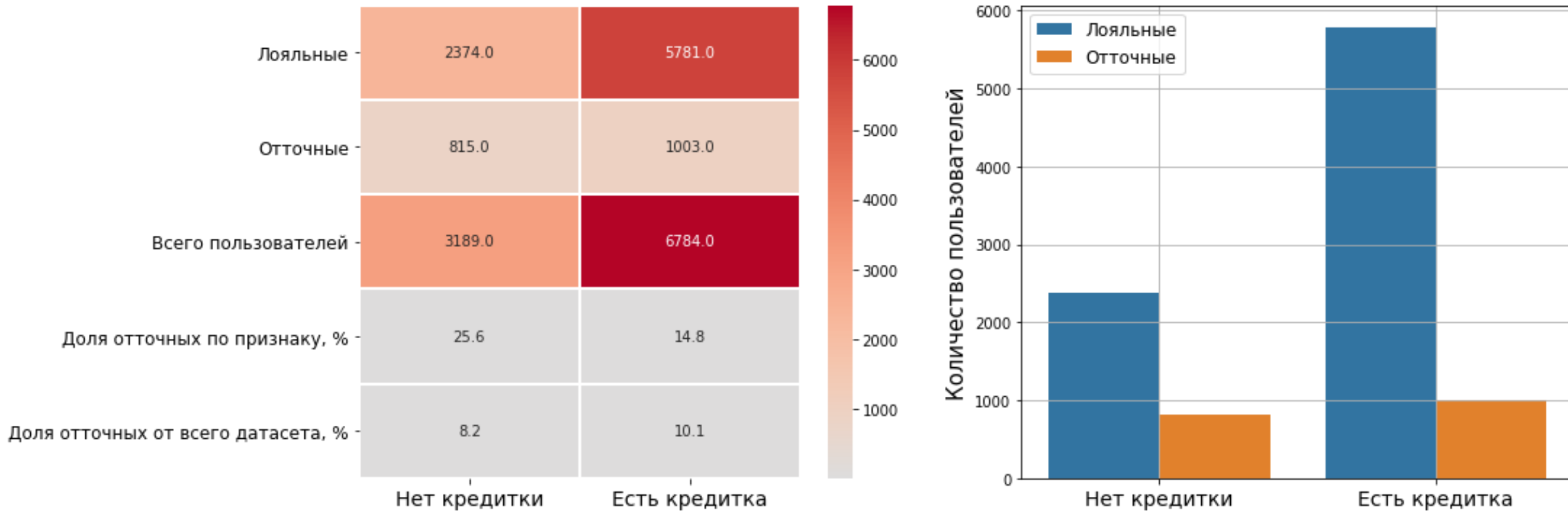
Отток наблюдается преимущественно для мужчин

Распределение значений по столбцу *products*



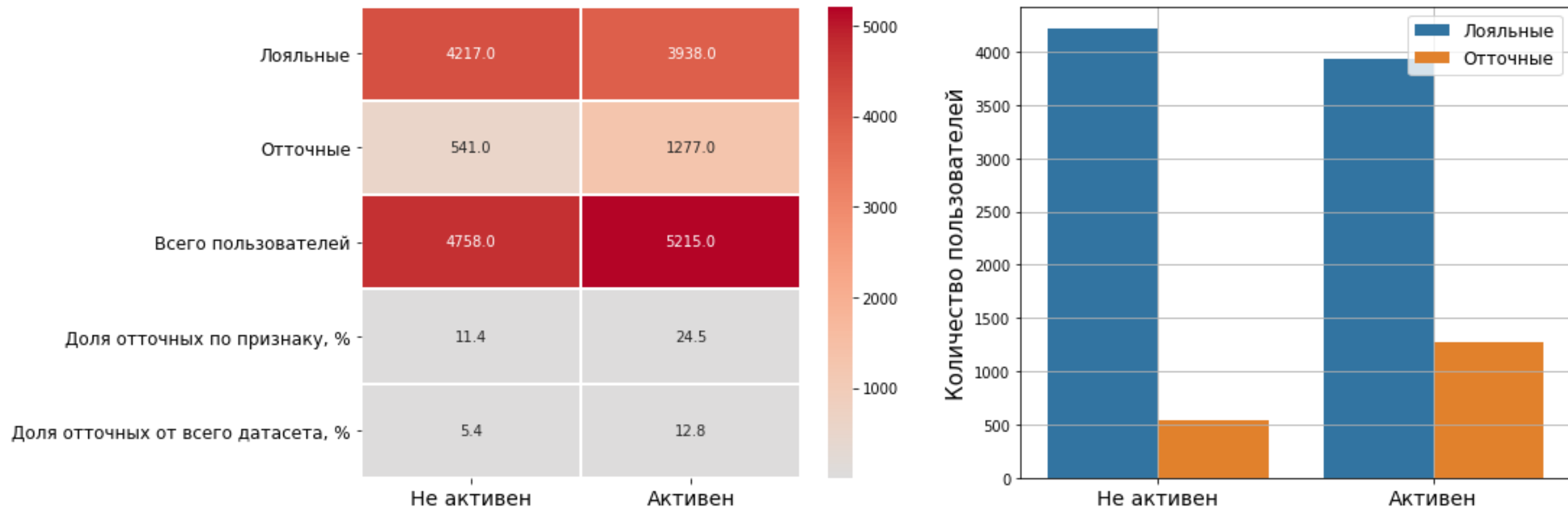
Отток наблюдается для пользователей с 2 двумя продуктами и более

Распределение значений по столбцу *credit_card*



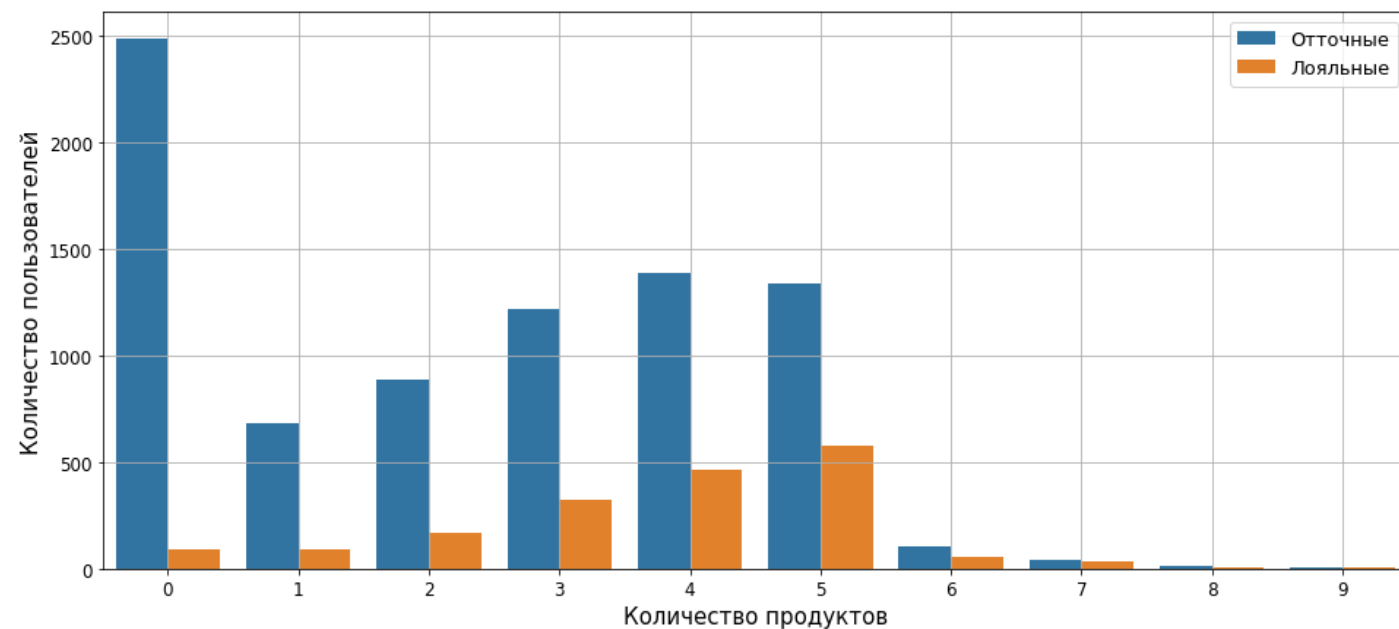
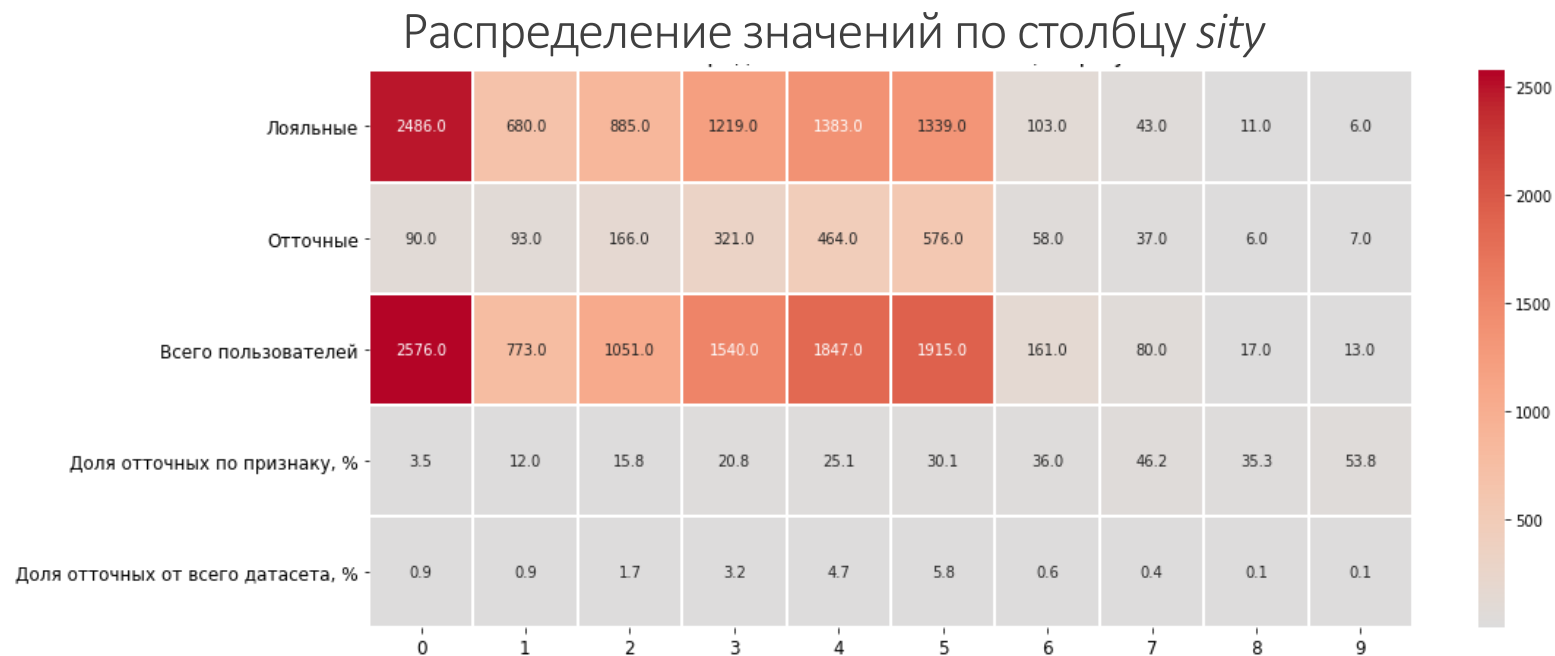
Больше всего (1003) отточных пользователей среди тех, у кого есть кредитка

Распределение значений по столбцу *last_activity*



Бóльший отток наблюдается среди активных пользователей

Количество
отточных
пользователей
растет вместе с
количеством
продуктов вплоть
до 6. Больше всего
их среди тех, кто
пользуется 3, 4 и 5
продуктами банка



Исследовательский анализ данных

Уровни оттока по категориям

Параметры для которых были заданы категории:

- score
- age
- balance
- est_salary

Параметры, значения которых были использованы как категории:

- products
- equity

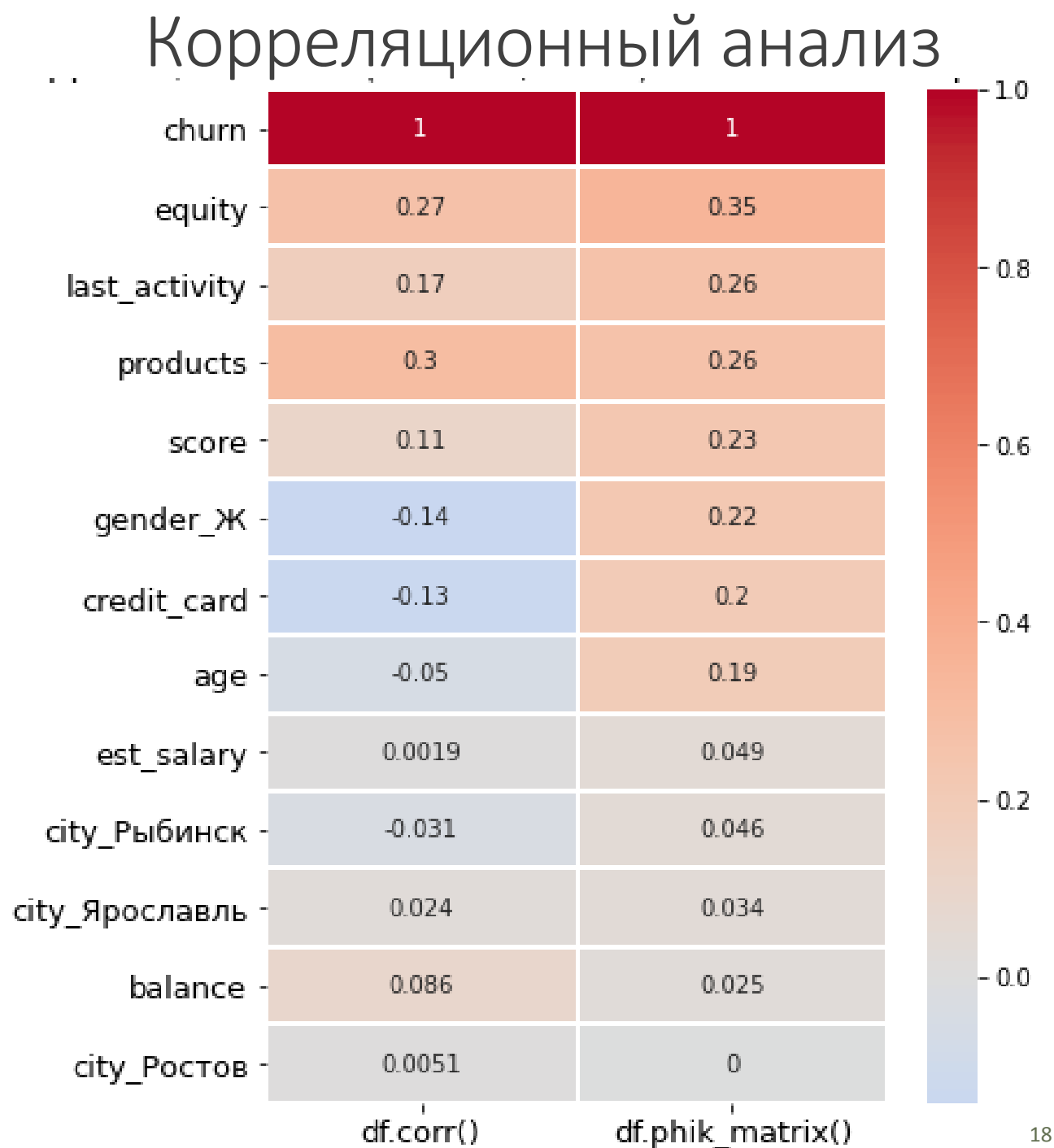
Для приведенных параметров были построены сводные таблицы, отражающие распределение по полученным категориям

1. Наибольшая корреляция к столбцу *equity* и составляет 0,35. Полученное число означает, что это умеренная взаимосвязь. Положительное значение означает, что при росте величины *equity* увеличивается отток клиентов.

2. Вторая по величине корреляция по отношению к столбцам *last_activity* и *products* и составляет 0,26.

Полученное число означает, что это слабая взаимосвязь. При росте величины *last_activity* или *products*, увеличивается отток клиентов.

3. Третья по величине корреляция по отношению к столбцу *score* и составляет 0,23. Полученное число означает, что это слабая взаимосвязь.



Проверка статистических гипотез

Проверка гипотезы различия дохода между теми клиентами, которые ушли и теми, которые остались

- Были сформированы нулевая H_0 и альтернативная H_1 гипотезы:
 - H_0 - доходы отточных и лояльных клиентов равны;
 - H_1 - доходы отточных и лояльных клиентов разные.
- Гипотезы были проверены непараметрическим тестом Уилкоксона-Манна-Уитни, т.к. в данном случае применение t-теста невозможно.
- По результатам проведенного теста была отвергнута нулевая гипотеза H_0 и это значит, что различия дохода между теми клиентами, которые ушли и теми, которые остались - имеются.

Проверка статистических гипотез

Проверка гипотезы различия возраста между теми клиентами, которые ушли и теми, которые остались

- Были сформированы нулевая H_0 и альтернативная H_1 гипотезы:
 - H_0 - возраст отточных и лояльных клиентов равный;
 - H_1 - возраст отточных и лояльных клиентов разный.
- Гипотезы были проверены двумя способами - непараметрическим тестом Уилкоксона-Манна-Уитни и t-тестом.
- По результатам обоих проведенных тестов была отвергнута нулевая гипотеза H_0 и это значит, что различия между теми клиентами, которые ушли и теми, которые остались – имеются.

Сегментация

Сегмент №1

- Количество продуктов в пользовании от 4 включительно;
- Пользователи, имеющий средний баланс на счету (от 500 000 до 3 000 000).

Параметры сегмента:

- Число пользователей в сегменте: равно 317;
- Число отточных пользователей в сегменте - 225;
- Число не отточных пользователей в сегменте - 92;
- Доля отточных пользователей в сегменте - 71 %.

Рекомендации:

- Пересмотр действующих условий на более выгодных условиях;
- Рассказать про новые продукты.

Сегментация

Сегмент №2

- Мужчины;
- Высокий оценочный доход пользователя - от 100 000 до 400 000;
- Количество баллов собственности больше чем 3.

Параметры сегмента:

- Число пользователей в сегменте: равно 1165;
- Число отточных пользователей в сегменте - 493;
- Число не отточных пользователей в сегменте - 672;
- Доля отточных пользователей в сегменте - 42 %.

Рекомендации:

- Повышенный процент сберегательного/накопительного счета;
- Страховка автомобиля или недвижимости на выгодных условиях.

Сегментация

Сегмент №3

- Уровень баллов кредитного скоринга выше низкого (больше 800);
- Количество баллов собственности больше чем 3;
- Возраст в промежутке от 26 до 38.

Параметры сегмента:

- Число пользователей в сегменте: равно 1446;
- Число отточных пользователей в сегменте - 536;
- Число не отточных пользователей в сегменте - 910;
- Доля отточных пользователей в сегменте - 37 %.

Рекомендации:

- Ипотечное кредитование на выгодных условиях;
- Автокредит.

Выводы

- ✓ Необходимо выстроить работу маркетинговой стратегии на основании выделенных сегментов;
- ✓ Обратить внимание на работу отдела в г. Ярославль. Во-первых, 2279 пользователя с пропущенными значениями в столбце *balance* находятся в г. Ярославль. Во-вторых, 1117 из 5975 пользователей этого города являются отточными. Необходимо посетить данный филиал, чтобы разобраться в его работе.
- ✓ Обновить систему сбора данных, чтобы исключить в дальнейшем появление новых дубликатов, а также появление пропусков в столбце *balance*.
- ✓ Через приложение, направить пользователям, имеющим дубликаты идентификатору пользователя (столбец *user_id*) push-уведомление, чтобы пользователь проверил свои данные и, в случае несоответствия, подошел в отделение банка.

stonks