

**БАНКИ — СЕГМЕНТАЦИЯ
ПОЛЬЗОВАТЕЛЕЙ ПО ПОТРЕБЛЕНИЮ**

ЦЕЛЬ РАБОТЫ

Задача данного исследования: определение новых актуальных сегментов, на которые можно разделить действующих клиентов банка.

Дополнительная задача: проверить отток клиентов из сформированных сегментов для выявления причин, почему это происходит.

Обоснование необходимости исследования: аналогичное исследование проводилось примерно 2 года назад, с тех пор структура клиентов поменялась, а их отток в последнее время усилился.

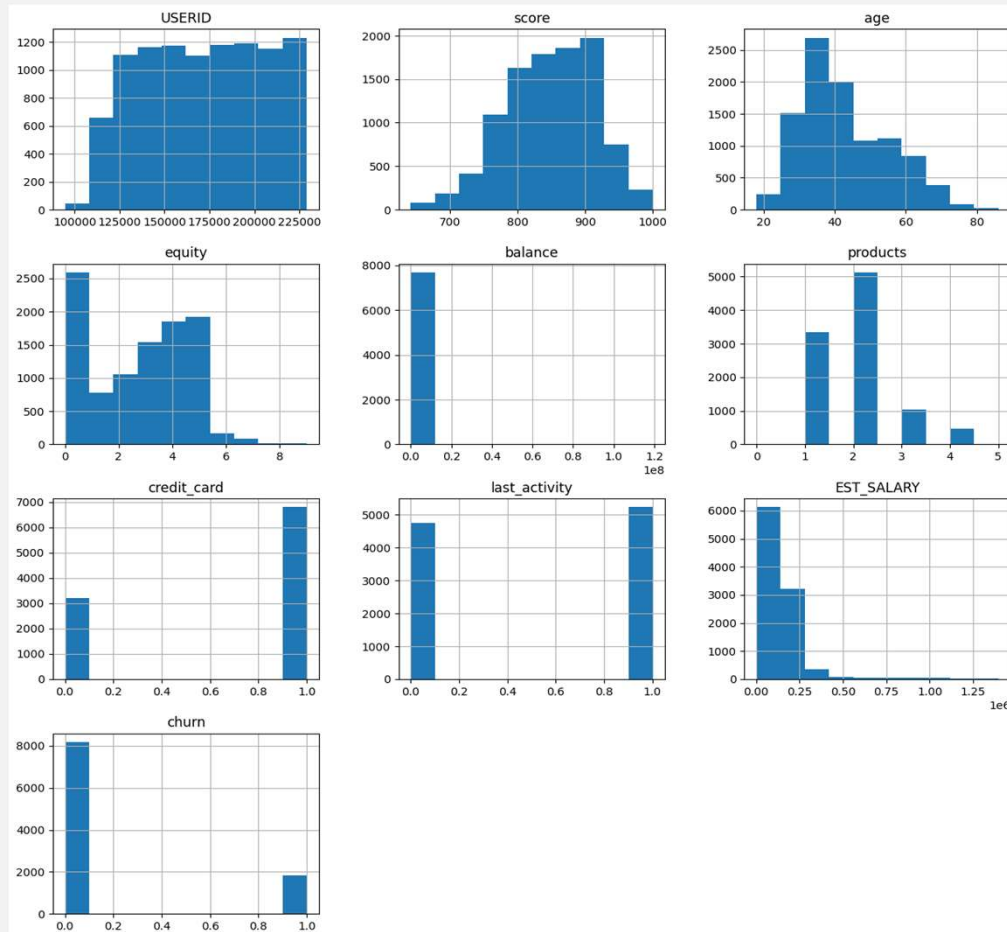
Заказчик исследования: отдел маркетинга банка. Полученные результаты о сегментации клиентов банка предполагается использовать для изменения стратегии предложения клиентам банка его продуктов.

Цель исследования: определение и выделение новых актуальных сегментов клиентов банка таким образом, чтобы изменение банковских продуктов приводило к уменьшению оттока клиентов.

КРАТКОЕ ОПИСАНИЕ ПРОЦЕССА ИССЛЕДОВАНИЯ

- Загрузка и обзор данных
- Предобработка данных;
- Исследовательский анализ данных:
 - Распределение клиентов по городам + доли клиентов из каждого города в общем кол-ве клиентов;
 - Присвоение определенной категории в зависимости от кредитного рейтинга. Построение гисторграммы и анализ получившегося результата;
 - Средняя зарплата и баланс денежных средств на счетах клиентов;
 - Распределение клиентов по количеству продуктов. Категоризация данных;
 - Изучение возрастного диапазона клиентов банка. Разбитие клиентов по возрастным категориям;
 - Сводная таблица: возраст/пол - кредитный рейтинг;
 - Матрица корреляций. Определение признаков, которые сильнее всего влияют на отток клиентов;
 - Портрет клиента из "зоны риска" (с низким скорингом);
- Получившиеся сегменты;
- Портреты действующих / ушедших клиентов;
- Статистические гипотезы: формулирование, доказательство, результаты расчетов;
- Выводы по результатам исследования и рекомендации для заказчика.

ОБЗОР ДАННЫХ

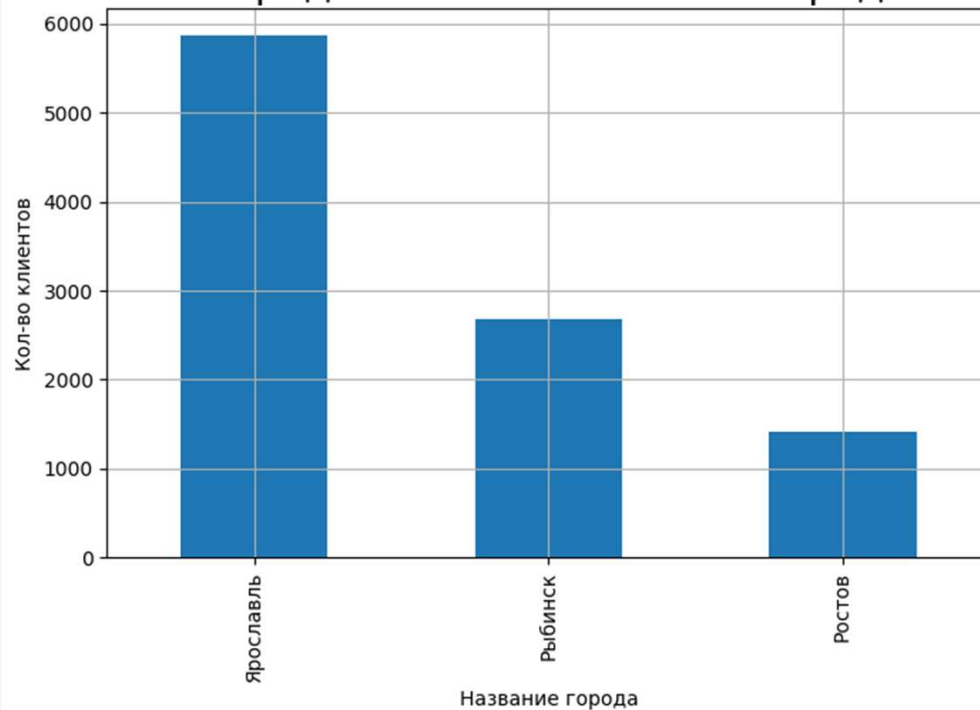


Краткие выводы:

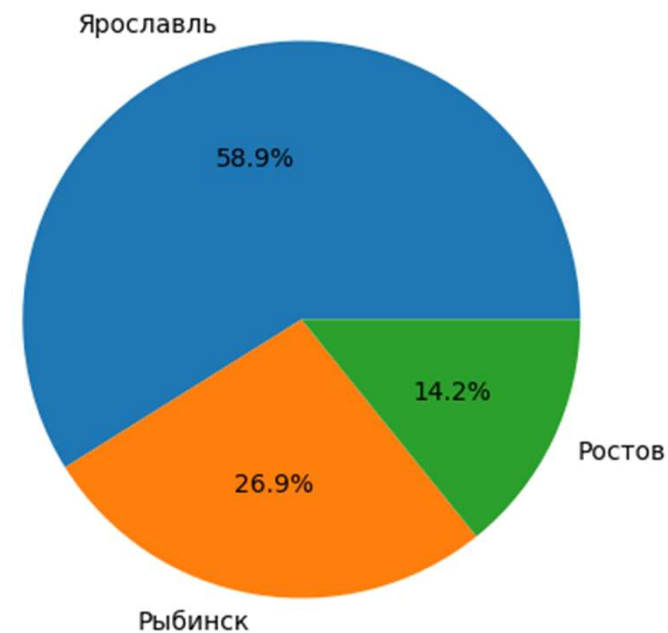
- приблизительное кол-во клиентов - 10 000 человек;
- пропуски: небольшое кол-во в данных с возрастом - такие строки мы просто удалим. Но в поле "Баланс на счете" пропусков очень много - надо будет подумать, что с этим делать;
- основная масса клиентов со скорингом 800 - 900+, т.е. достаточно высоким уровнем надежности;
- возраст большинства клиентов находится в диапазоне 25-45 лет;
- чуть больше 25% клиентов не имеют никаких объектов в собственности;
- в значениях balance (баланс на счете) есть какие-то аномалии, нужно будет проверить это и выбросить лишние данные;
- у 80-85% клиентов один или два продукта банка;
- почти у 70% клиентов есть кредитная карта банка;
- более 45% клиентов почему-то не активные - необходимо будет это выяснить в процессе исследовательского анализа;
- по заработным платам данные похожи на объективные, но их все равно необходимо будет проверить на предмет выбросов;
- доля ушедших клиентов составляет 18,2%.

ИССЛЕДОВАТЕЛЬСКИЙ АНАЛИЗ. РАСПРЕДЕЛЕНИЕ КЛИЕНТОВ ПО ГОРОДАМ

Распределение клиентов по городам



Доли клиентов из каждого города в общем кол-ве клиентов

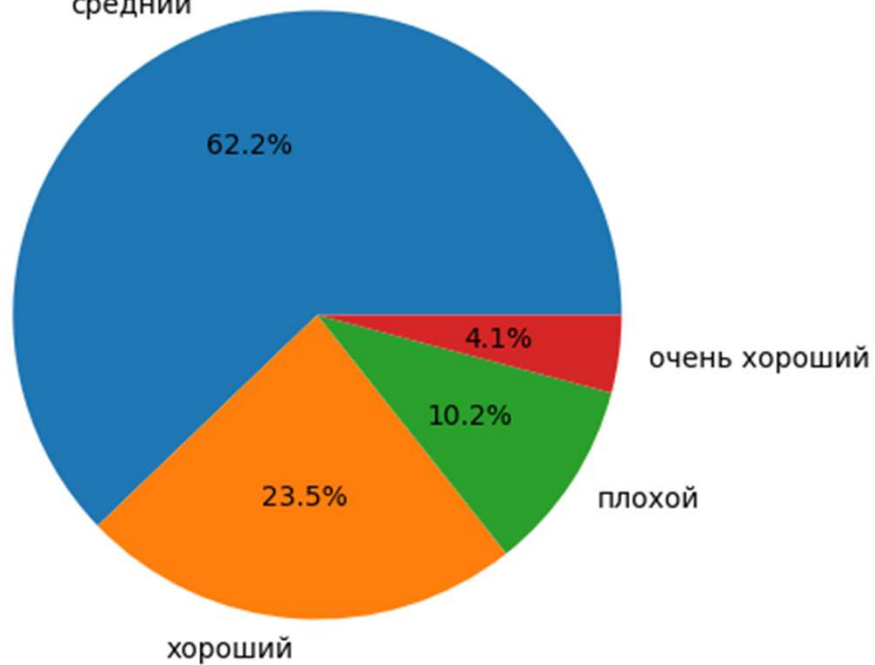


В Ярославле проживают 5875 клиентов, в Рыбинске – 2682 клиента, а в Ростове – 1416 клиентов.

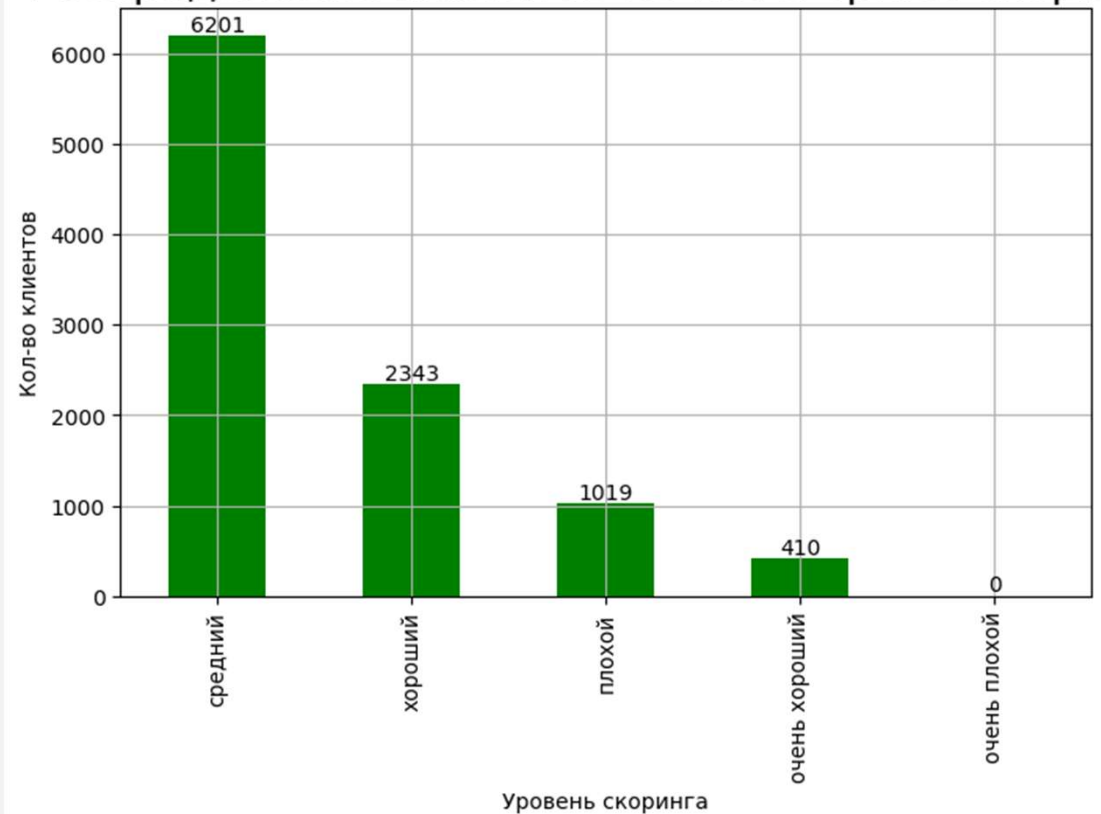
ИССЛЕДОВАТЕЛЬСКИЙ АНАЛИЗ. КРЕДИТНЫЙ РЕЙТИНГ

Доли клиентов в каждом уровне скоринга относительно общего кол-ва клиентов

средний

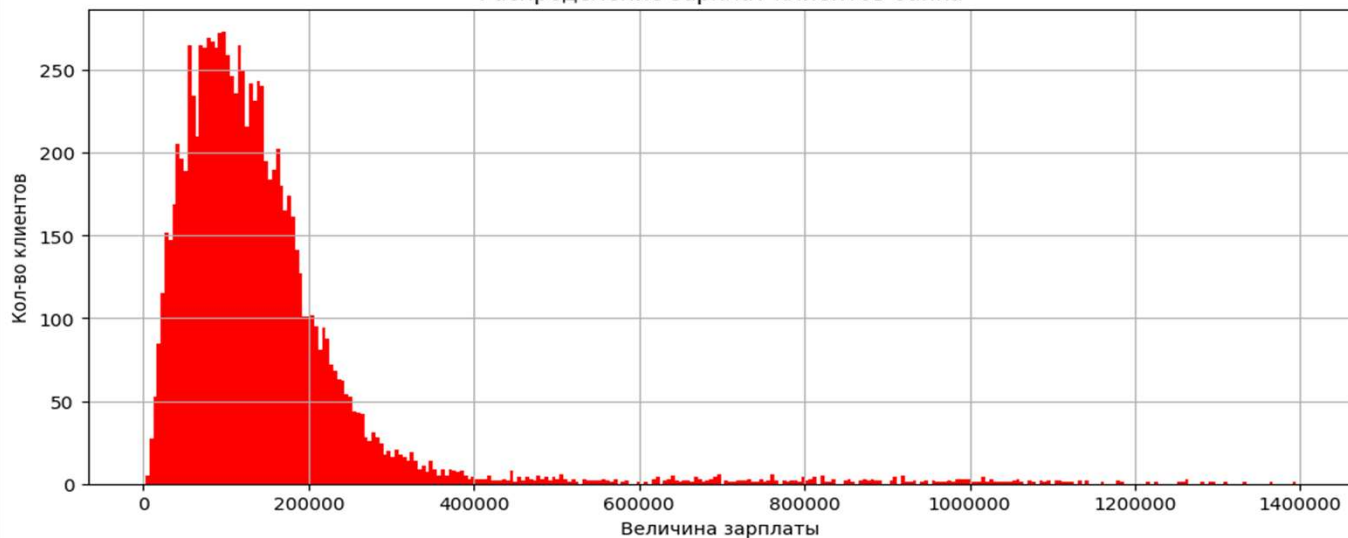


Распределение клиентов по категориям скоринга



Категоризацию подтвердил расчет перцентилей: пороговый уровень, отличающий плохой рейтинг от среднего, составляет 737, а по критериям - 765. Общая доля клиентов со средним и хорошим скорингом составляет 85.7%, что является хорошим показателем качества клиентов.

Распределение зарплат клиентов банка



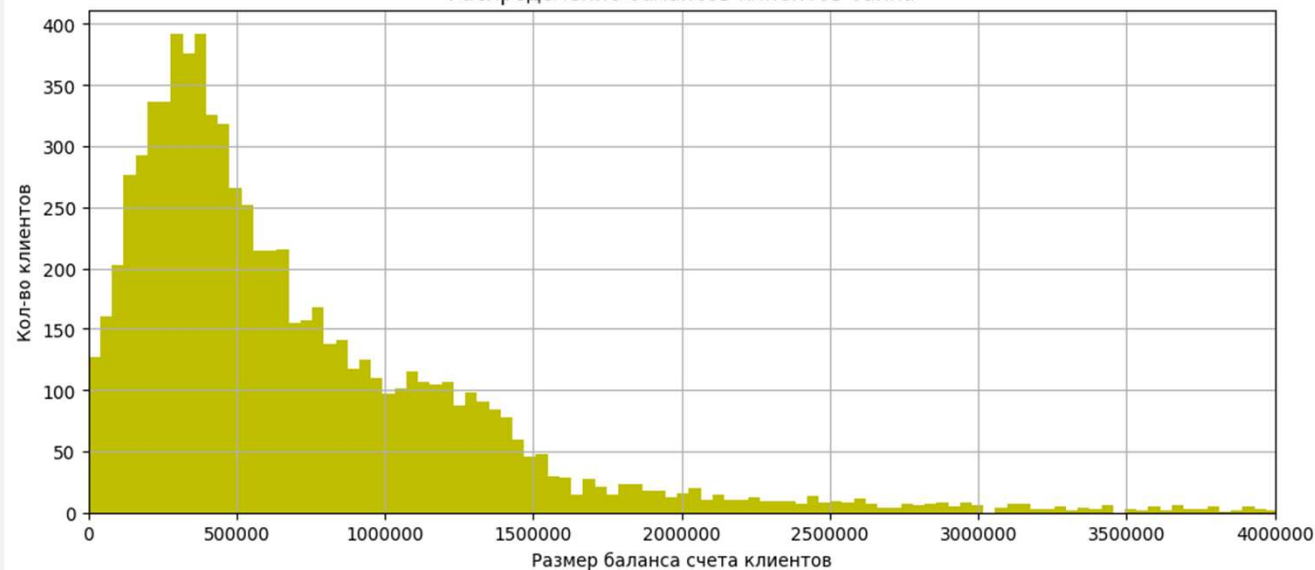
ИССЛЕДОВАТЕЛЬСКИЙ
АНАЛИЗ. СРЕДНЯЯ
ЗАРПЛАТА И БАЛАНС
ДЕНЕЖНЫХ СРЕДСТВ НА
СЧЕТАХ КЛИЕНТОВ

Средняя зарплата клиентов банка составляет **147 799** рублей, а средний баланс – **827 245** рублей.

Порог VIP-клиентов начинается от следующих значений:

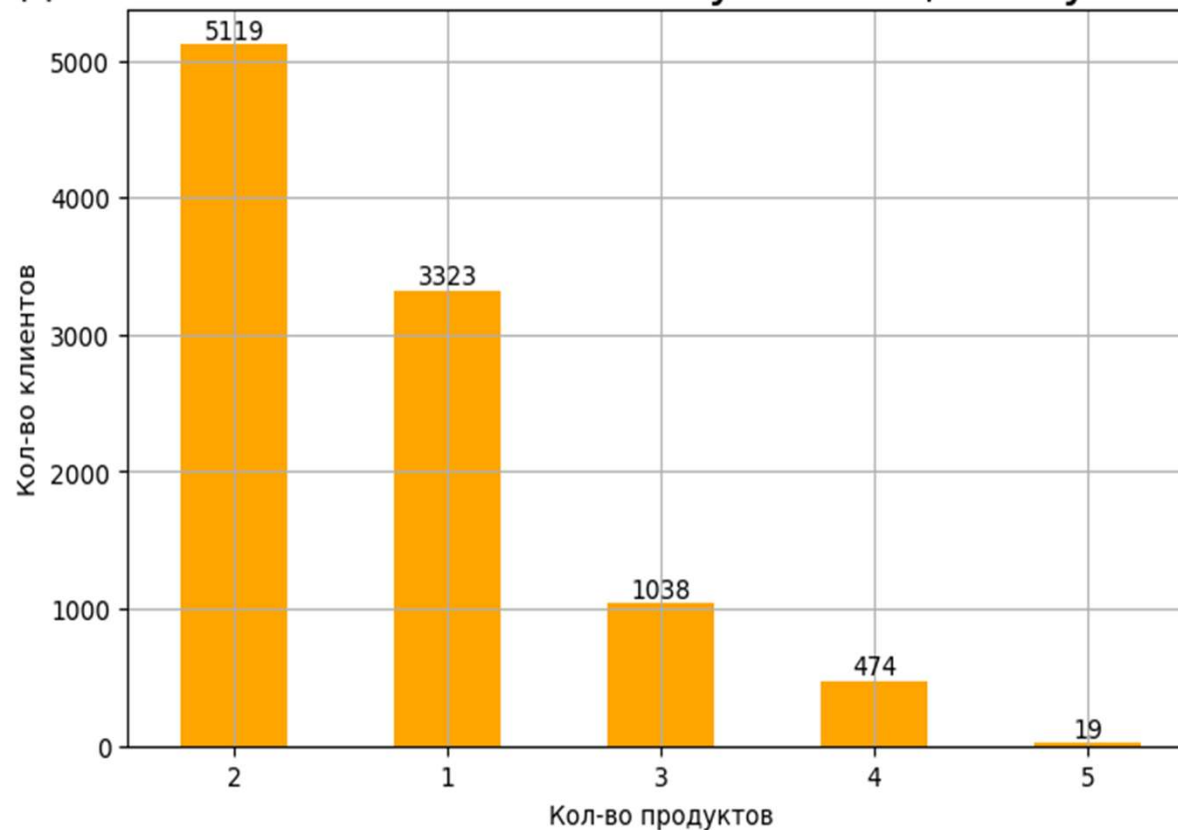
- Величина зарплаты – выше **316 360** рублей;
- Величина баланса – выше **2 114 078** рублей.

Распределение балансов клиентов банка



ИССЛЕДОВАТЕЛЬСКИЙ АНАЛИЗ. РАСПРЕДЕЛЕНИЕ КЛИЕНТОВ ПО КОЛИЧЕСТВУ ПРОДУКТОВ. КАТЕГОРИЗАЦИЯ ДАННЫХ

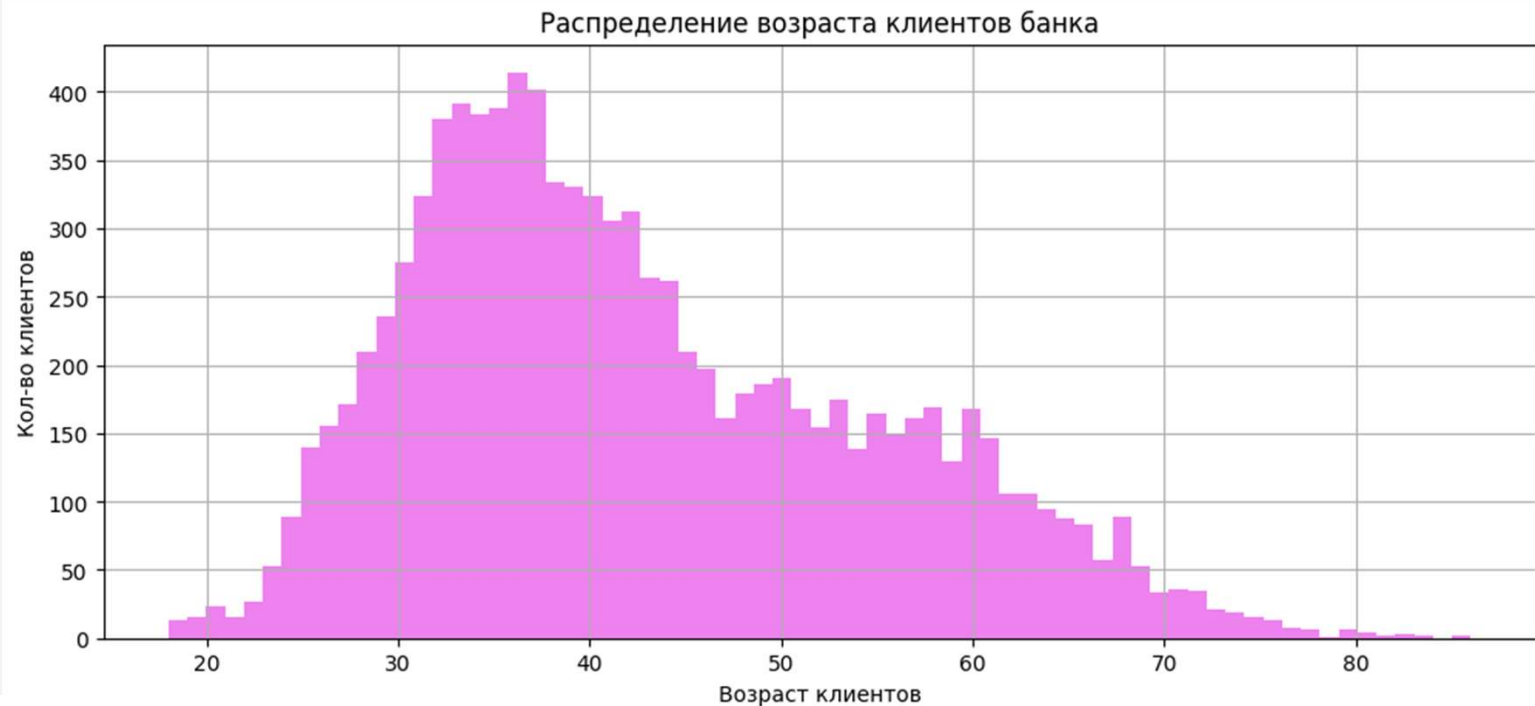
Распределение клиентов по кол-ву имеющихся у них продуктов



Критерии категоризации:

- для клиентов с кол-вом продуктов 1 и 2 пока что нет второго признака, по которым их можно разделить. Укрупнять эти группы не стоит - они итак достаточно большие. Поэтому клиенты с 1 и 2 продуктами были выделены в самостоятельные категории;
- клиенты с 3 и более продуктами были объединены в одну группу для укрупнения.

ИССЛЕДОВАТЕЛЬСКИЙ
АНАЛИЗ. ИЗУЧЕНИЕ
ВОЗРАСТНОГО
ДИАПАЗОНА
КЛИЕНТОВ БАНКА.
РАЗВИТИЕ КЛИЕНТОВ
ПО ВОЗРАСТНЫМ
КАТЕГОРИЯМ



Средний возраст клиента

city_orig	gender_orig	
Ростов	Ж	47.0
	М	41.0
Рыбинск	Ж	46.0
	М	41.0
Ярославль	Ж	45.0
	М	40.0

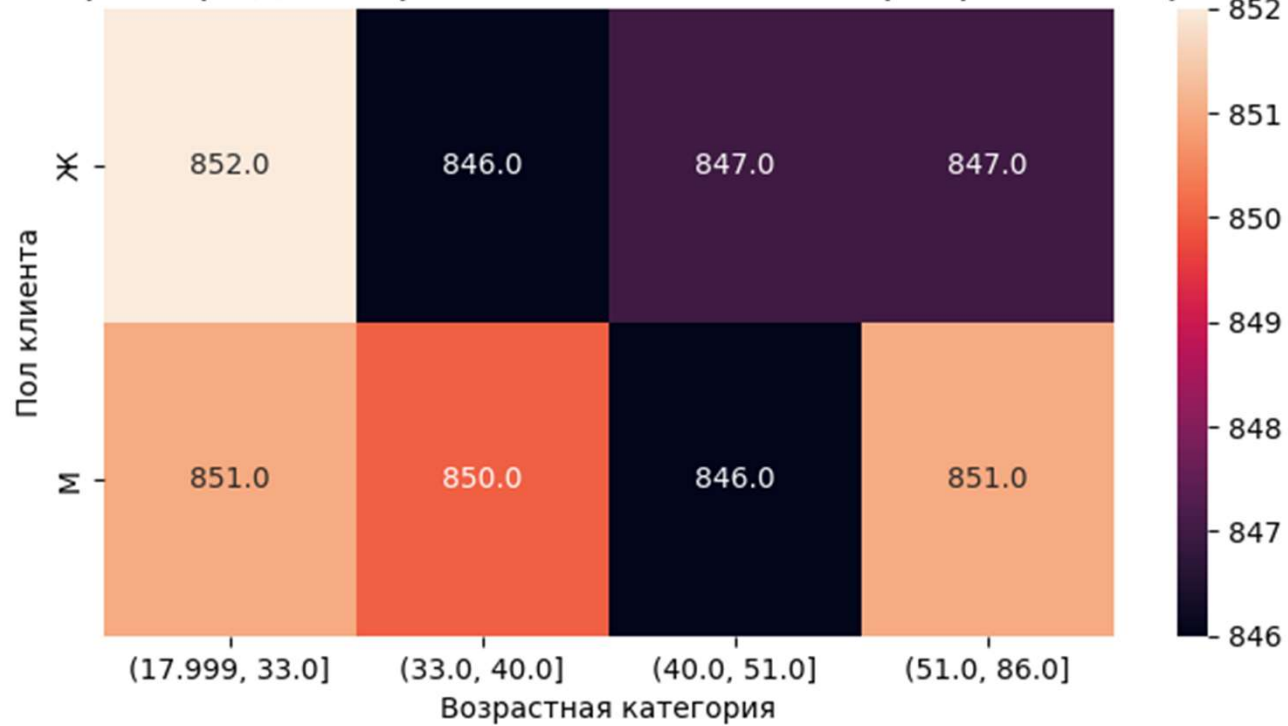
Клиенты были разделены по возрасту на 4 равных категории:

- 18 -33 года;
- 34-40 лет
- 41 - 51 год;
- 52 - 86 лет.

Средний возраст клиентов не опускается ниже 40 лет независимо от пола и города проживания.

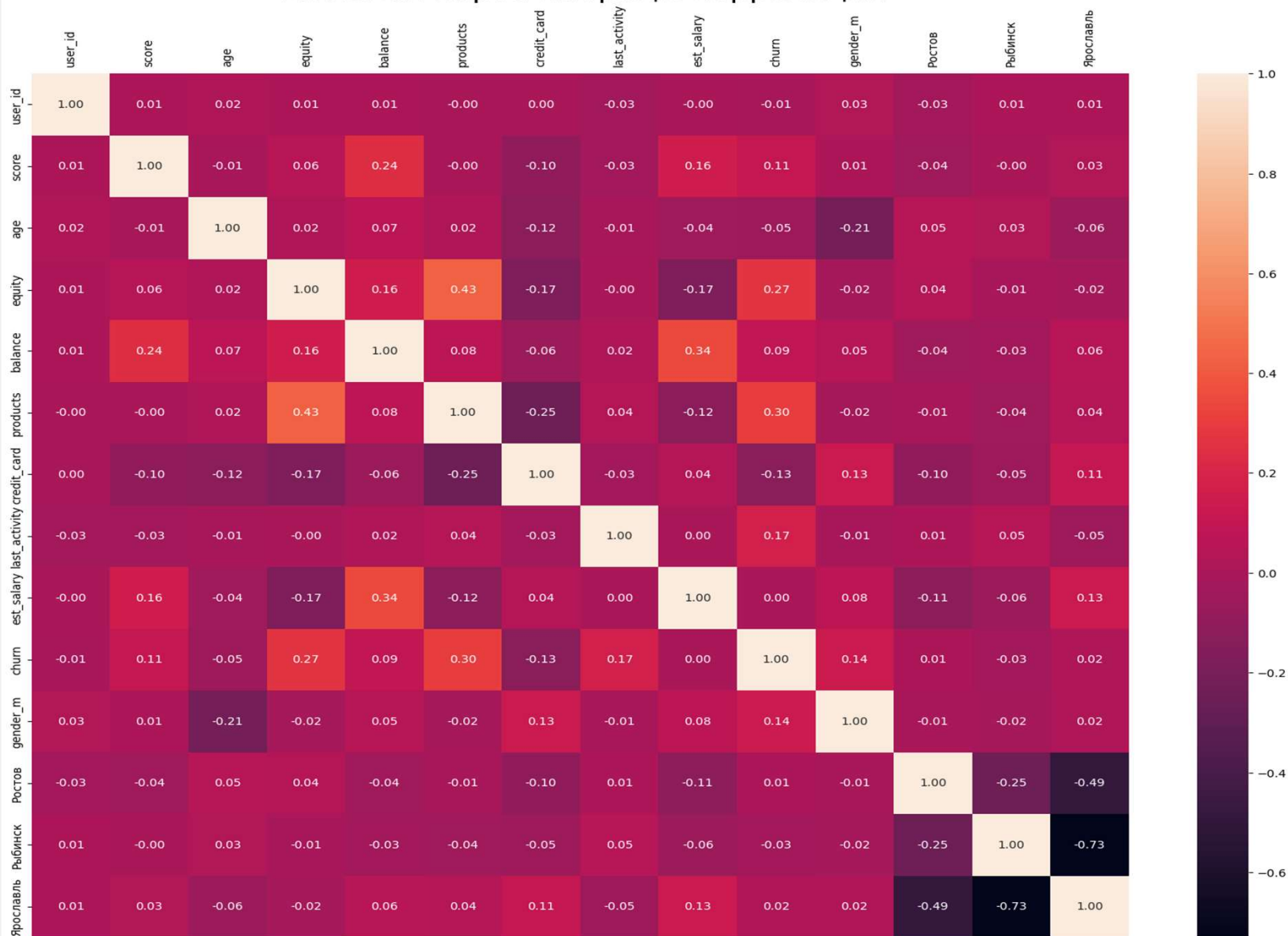
ИССЛЕДОВАТЕЛЬСКИЙ АНАЛИЗ. СВОДНАЯ ТАБЛИЦА: ВОЗРАСТ/ПОЛ - КРЕДИТНЫЙ РЕЙТИНГ

Тепловая карта среднего рейтинга клиентов в разрезе "возраст / пол"



Средние скоринги находятся в среднем диапазоне, что подтверждает ранее выдвинутое предположение о высоком качестве клиентской базы.

Тепловая карта матрицы корреляций



ИССЛЕДОВАТЕЛЬСКИЙ АНАЛИЗ. МАТРИЦА КОРРЕЛЯЦИЙ. ОПРЕДЕЛЕНИЕ ПРИЗНАКОВ, КОТОРЫЕ СИЛЬНЕЕ ВСЕГО ВЛИЯЮТ НА ОТТОК КЛИЕНТОВ

Зависимость **оттока** от:

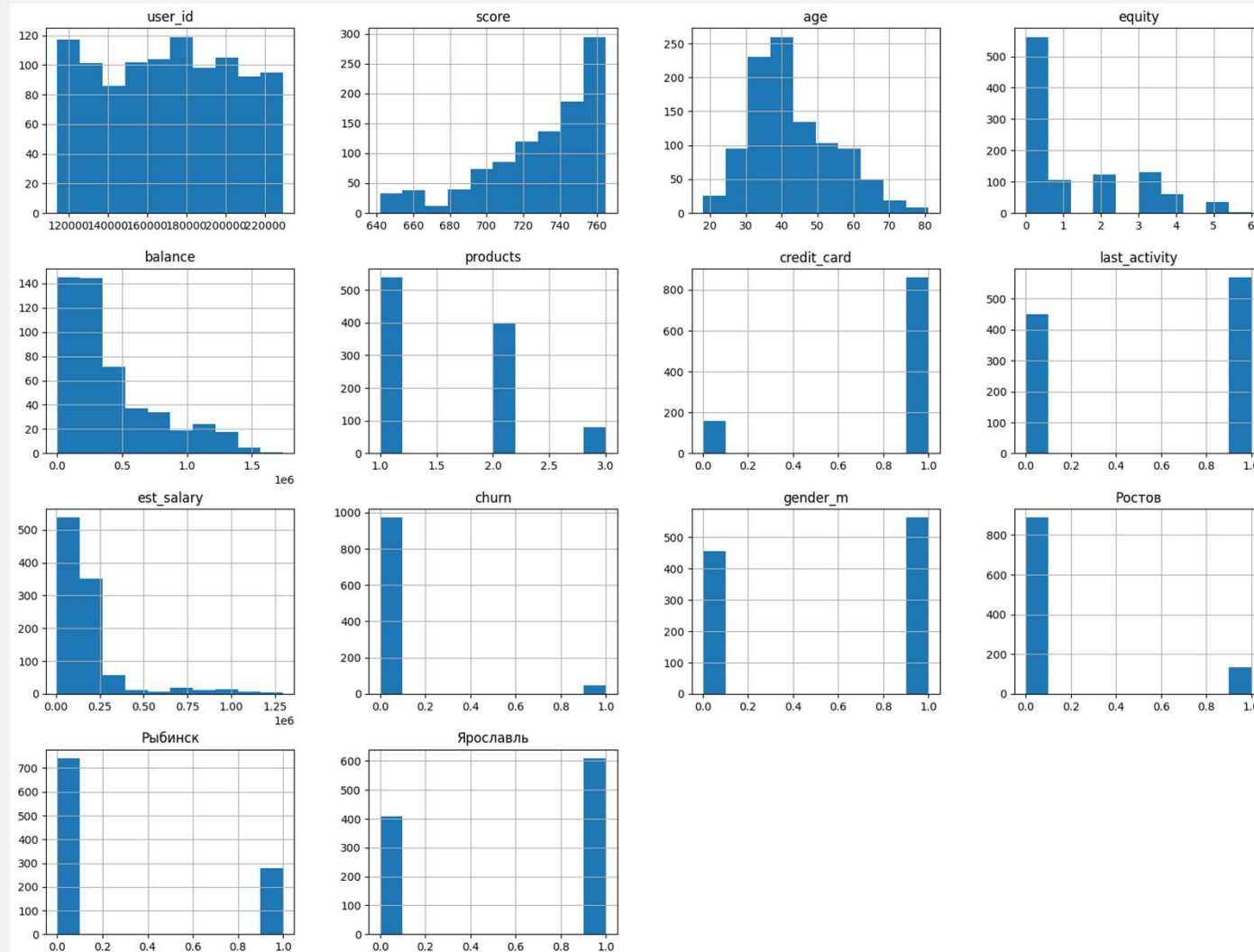
- кредитного рейтинга - коэффициент 0.11: слабая положительная связь. Однако в столбце score отсутствуют выбросы;
- **оценочная стоимость активов клиента (equity) - коэффициент 0.27**: слабая положительная связь;
- **кол-во продуктов - коэффициент 0.3**: слабая положительная связь;
- наличие кредитной карты - коэффициент -0.13: слабая отрицательная связь;
- последняя активность - коэффициент 0.17: слабая положительная связь;
- пол - коэффициент 0.14: слабая положительная связь.

С остальными переменными связь отсутствует. Так как самая сильная связь оттока клиентов была с кол-вом продуктов и оценочной стоимостью активов, то по этим признакам и разделим наших пользователей на сегменты.

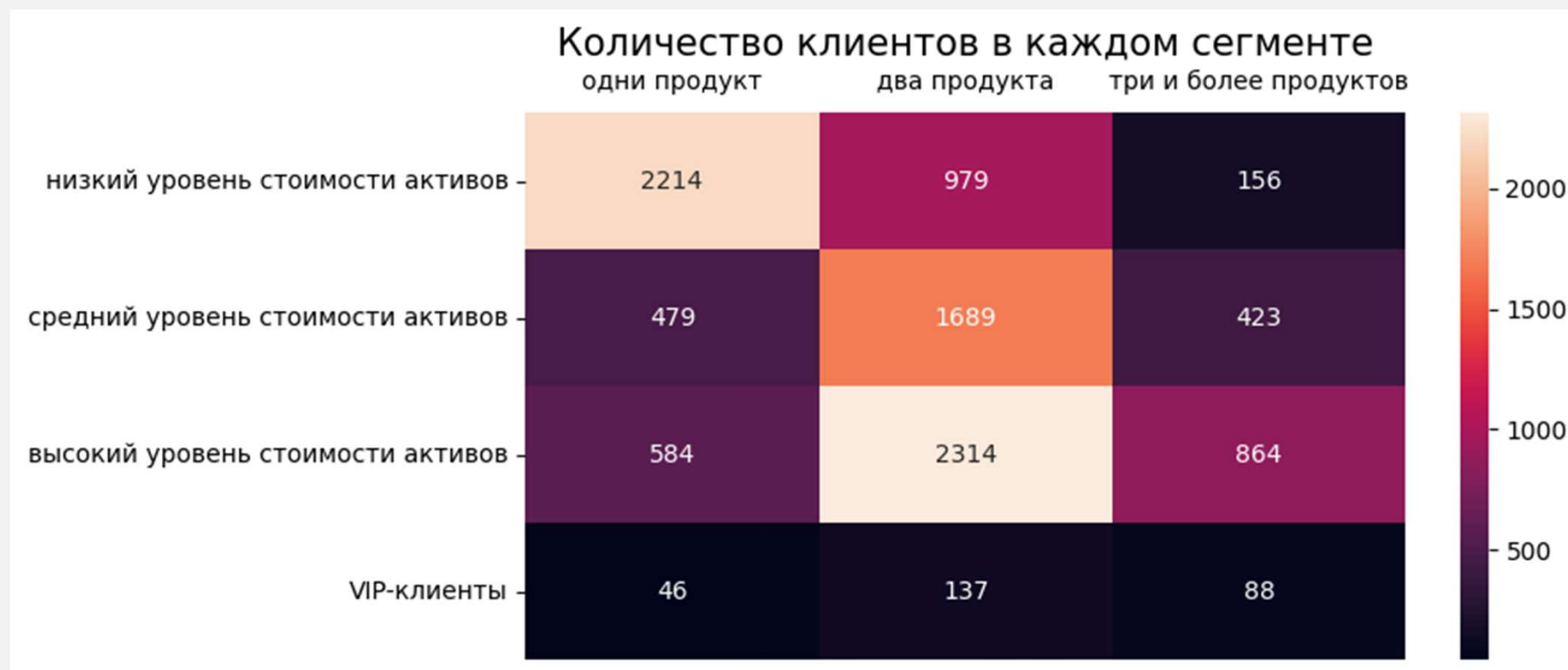
ИССЛЕДОВАТЕЛЬСКИЙ АНАЛИЗ. ПОРТРЕТ КЛИЕНТА ИЗ "ЗОНЫ РИСКА" (С НИЗКИМ СКОРИНГОМ)

Портрет **ненадежного клиента** выглядит следующим образом:

- возраст: 35-40 лет;
- оценка собственного капитала близка к нулю;
- баланс денежных средств также близок к нулю;
- имеет два или чаще один продукт банка;
- 80% клиентов из категории риска имеют кредитную карту;
- больше 50% таких клиентов совершали какие-либо действия более месяца назад, но большинство из таких клиентов остаются активными;
- чаще всего такие клиенты встречаются в Ростове, затем в Рыбинске. В Ярославле кол-во таких клиентов меньше всего.



ПОЛУЧИВШИЕСЯ СЕГМЕНТЫ



При сегментации получилось 12 групп клиентов. В дальнейшем исследовании будет проведен анализ каждого сегмента на отток клиентов – там, где он будет выше среднего по банку, составим портреты действующих / ушедших клиентов.

ПОРТРЕТЫ ДЕЙСТВУЮЩИХ / УШЕДШИХ КЛИЕНТОВ

Портрет клиентов из сегмента, которые имеют три и более продуктов и средний уровень стоимости активов.

	Действующий клиент	Ушедший клиент
Количество пользователей в сегменте	282	141
Категория скоринга	средний	средний
Среднее значение скоринга	820	850
Средний возраст	46	43
Баланс на счете	456174	1014910
Есть ли кредитная карта?	Нет	Да
Средняя зарплата клиента	89849	127475
Пол	Ж	М
Был ли клиент активен в последний месяц?	Нет	Да
В каком городе чаще всего встречается такой клиент	Ярославль	Ярославль

Процент оттока клиентов из данного сегмента составляет 33.3%.
Отток клиентов из этого сегмента выше, чем в среднем по банку.

У нас получились следующие отточные сегменты, которые удовлетворяют условиям исследования:

- клиенты с двумя продуктами и высоким уровнем стоимости активов: численность 2314, из них ушло 24.4%;
- клиенты с тремя и более продуктами и средним уровнем стоимости активов: численность 423, из них ушло 33.3%;
- клиенты с тремя и более продуктами и высоким уровнем стоимости активов: численность 864, из них ушло 43.8%.

Портрет клиентов из сегмента, которые имеют два продукта и высокий уровень стоимости активов.

	Действующий клиент	Ушедший клиент
Количество пользователей в сегменте	1749	565
Категория скоринга	средний	средний
Среднее значение скоринга	852	877
Средний возраст	43	40
Баланс на счете	688630	923205
Есть ли кредитная карта?	Да	Нет
Средняя зарплата клиента	104014	134000
Пол	Ж	М
Был ли клиент активен в последний месяц?	Нет	Да
В каком городе чаще всего встречается такой клиент	Ярославль	Ярославль

Процент оттока клиентов из данного сегмента составляет 24.4%.
Отток клиентов из этого сегмента выше, чем в среднем по банку.

Портрет клиентов из сегмента, которые имеют три и более продуктов и высокий уровень стоимости активов.

	Действующий клиент	Ушедший клиент
Количество пользователей в сегменте	486	378
Категория скоринга	средний	средний
Среднее значение скоринга	852	877
Средний возраст	45	42
Баланс на счете	735968	1263122
Есть ли кредитная карта?	Нет	Да
Средняя зарплата клиента	101174	136389
Пол	Ж	М
Был ли клиент активен в последний месяц?	Нет	Да
В каком городе чаще всего встречается такой клиент	Ярославль	Ярославль

Процент оттока клиентов из данного сегмента составляет 43.8%.
Отток клиентов из этого сегмента выше, чем в среднем по банку.

СТАТИСТИЧЕСКИЕ ГИПОТЕЗЫ: ФОРМУЛИРОВАНИЕ, ДОКАЗАТЕЛЬСТВО, РЕЗУЛЬТАТЫ РАСЧЕТОВ

Гипотеза №1

- Нулевая (H_0): возраст клиентов, имеющих один продукт, равен возрасту клиентов, у которых два продукта банка;
- Альтернативная (H_1): возраст клиентов, имеющих один продукт, не равен возрасту клиентов, у которых два продукта банка.

p-значение: 0.057138410813417605

Не получилось отвергнуть нулевую гипотезу, вывод о различии средних возрастов сделать нельзя.

Гипотеза №2

Исходя из задач исследования нам интересно узнать самую отточную группу клиентов. В соответствии с пожеланиями заказчика исследования и результатам сегментации у нас есть две таких группы:

- клиенты с тремя и более продуктами и средним уровнем стоимости активов : численность 423, из них ушло 33.3%;
- клиенты с тремя и более продуктами и высоким уровнем стоимости активов : численность 864, из них ушло 43.8%.

Попробуем подтвердить, что сегмент клиентов с тремя и более продуктами и высоким уровнем стоимости активов является самым отточным и полученные результаты не являются погрешностью вычислений в процессе исследования - сформулируем наши гипотезы:

- Нулевая (H_0): доля отточных клиентов с тремя и более продуктами и средним уровнем стоимости активов **равна** доле отточных клиентов с тремя и более продуктами и высоким уровнем стоимости активов ;
- Альтернативная (H_1): доля отточных клиентов с тремя и более продуктами и средним уровнем стоимости активов **не равна** доле отточных клиентов с тремя и более продуктами и высоким уровнем стоимости активов .

p-значение: 0.000345792130175937

Отвергаем нулевую гипотезу: не удалось доказать равенство между долями отточных клиентов групп "три и более продуктов и средний уровень стоимости активов" и "три и более продуктов и высокий уровень стоимости активов"

ВЫВОДЫ ПО РЕЗУЛЬТАТАМ ИССЛЕДОВАНИЯ И РЕКОМЕНДАЦИИ ДЛЯ ЗАКАЗЧИКА (СЛАЙД 1)

Ответы на ключевые вопросы, поставленные в качестве задачи исследования:

- клиенты были поделены по следующим категориям (в итоге получилось 12 сегментов):
 - по количеству продуктов: "один продукт", "два продукта", "три и более продуктов" (банка);
 - по уровню оценки стоимости собственности: "низкий уровень стоимости активов", "средний уровень стоимости активов", "высокий уровень стоимости активов" и "VIP-клиенты";
- построены портреты типичного отточного/действующего клиента для каждого сегмента, проведено сравнение со средним значением оттока клиентов в банке;
- самыми отточными и одновременно крупными сегментами являются:
 - клиенты с двумя продуктами и высоким уровнем стоимости активов: численность 2314, из них ушло 24.4%;
 - клиенты с тремя и более продуктами и средним уровнем стоимости активов: численность 423, из них ушло 33.3%;
 - клиенты с тремя и более продуктами и высоким уровнем стоимости активов: численность 864, из них ушло 43.8%;
- подтверждение того, что последний сегмент является одним из самых отточных (не учитывая VIP-клиентов), было проведено с помощью проверки статистической гипотезы на равенство долей.

ВЫВОДЫ ПО РЕЗУЛЬТАТАМ ИССЛЕДОВАНИЯ И РЕКОМЕНДАЦИИ ДЛЯ ЗАКАЗЧИКА (СЛАЙД 2)

Рекомендации к полученным сегментам: Т.к. у нас есть корреляция между кол-вом используемых продуктов и оценкой стоимости активов и оттоком, то можно предложить следующие варианты по его снижению:

- **Кол-во используемых продуктов:**

- произвести их укрупнение - например объединить дебетовую и кредитную карту в одну карту с набором функций (для повышения удобства использования). Или объединить доступ к расчетному счету с депозитом - на остаток на р/с начислять %, чтобы у клиента не было необходимости отдельно открывать вклад;
- у банка просто отсутствует продукт, который можно предложить этому сегменту клиентов (VIP они уже перепробовали все, что было, но не нашли то, что им необходимо). Необходимо активнее внедрять новые продукты, которые в разработке. А в идеальной ситуации узнать у ушедших клиентов, почему они ушли и каких продуктов им не хватало.

- **Оценка стоимости активов:** тут многое зависит от того, чем занимается наш клиент:

- **затруднен доступ к финансовым средствам:** если клиент - предприниматель и у него высокий уровень баланса и размер активов - упростить доступ к его денежным средствам, сделать переводы денег и оплату счетов быстрее и интуитивно понятнее (т.е. провести эксперимент: дать человеку, далекому от банковских операций, оплатить счет, выставленный контрагентом - это должно быть так просто, чтобы даже ребенок мог с этим справиться);
- **низкая доходность:** внедрить прогрессивный % на вклады и на остатки по счету, кешбэки в зависимости от кол-ва денег на балансе;

- **Низкая активность:**

- на отток клиентов также напрямую влияет их активность. Если посмотреть на портреты клиентов из отточных сегментов, то будет очевидно, что большая часть клиентов там - неактивные (активными они становятся после того, как совершают действие по расторжению договора). Поэтому необходимо повысить общую активность клиентов. В частности, поработать над улучшением мобильного приложения и взаимодействием с продуктами (обзвон клиентов воспринимается плохо, а посещения офисов редкие - большинство вопросов опять же можно решить через приложение). Например, на главном экране приложения сделать обзор всех продуктов, которые на данный момент есть у клиента (карты, кредиты, вклады и т.п.). Улучшить взаимодействие с ними: например, досрочная оплата по ипотеке должна быть интуитивно понятной, желательно в 2-3 действия;
- Еще из интересного опыта повышения активности: использовать маркетплейсы банка, а получение товаров в отделениях, у операторов (при получении товара клиент хочет убедиться, что с товаром все будет в порядке, поэтому большим успехом пользуются пункты выдачи, а не постаматы. По этой же логике клиент выбирает получение посылки в отделении банка у оператора, а не в постамате - в данном случае активность клиента уже повышается). При получении товара предлагать какие-то недорогие услуги (например подписка на музыку или доставку еды с кешбэком), которые оформляются в рамках 5-10 минут, после чего клиент сможет начать пользоваться этим продуктом и приносить банку деньги.

СПАСИБО ЗА ВНИМАНИЕ!

А если у вас остались вопросы, то со мной можно связаться по следующим контактам:

Аналитик данных Александр Фошин

Тел.: +7-977-654-32-10

email: i_love_practicum@yandex.ru

P.S.: БОНУС ДЛЯ МЕНЕДЖЕРА ПРОЕКТА, У КОТОРОГО БЫЛО ОЧЕНЬ МНОГО ВОПРОСОВ))

Портрет сегмента всех VIP-клиентов

	Действующий клиент	Ушедший клиент
Количество пользователей в сегменте	163	108
Категория скоринга	средний	средний
Среднее значение скоринга	855	886
Средний возраст	43	41
Баланс на счете	550602	791393
Есть ли кредитная карта?	Да	Нет
Средняя зарплата клиента	91998	120076
Пол	Ж	М
Был ли клиент активен в последний месяц?	Нет	Да
В каком городе чаще всего встречается такой клиент	Ярославль	Ярославль

Процент оттока клиентов из данного сегмента составляет 39.9%
Отток клиентов из этого сегмента выше, чем в среднем по банку.

VIP-сегмент клиентов объединен по признаку оценки стоимости активов. В независимости от кол-ва продуктов этот объединенный сегмент хоть и не многочисленный, но также очень отточный. Причем, что необычно – средняя зарплата у них меньше, чем у клиентов из других сегментов и чем средняя по городам. Однако у них очень большие балансы на счетах и высокая стоимость активов – возможно это обусловлено тем, что они получают основной доход в виде процентов по вкладам, акциям или в виде дивидендов от бизнеса, а зарплаты как таковой у них нет.