

# Анализ клиентов банка и сегментирование по количеству потребляемых продуктов

Айдар Мавлютов  
группа da\_63

## Цель проекта

Провести анализ клиентов банка и сегментировать пользователей по количеству потребляемых продуктов, обращая внимание на отток клиентов, чтобы изменение банковских продуктов приводило к уменьшению оттока.

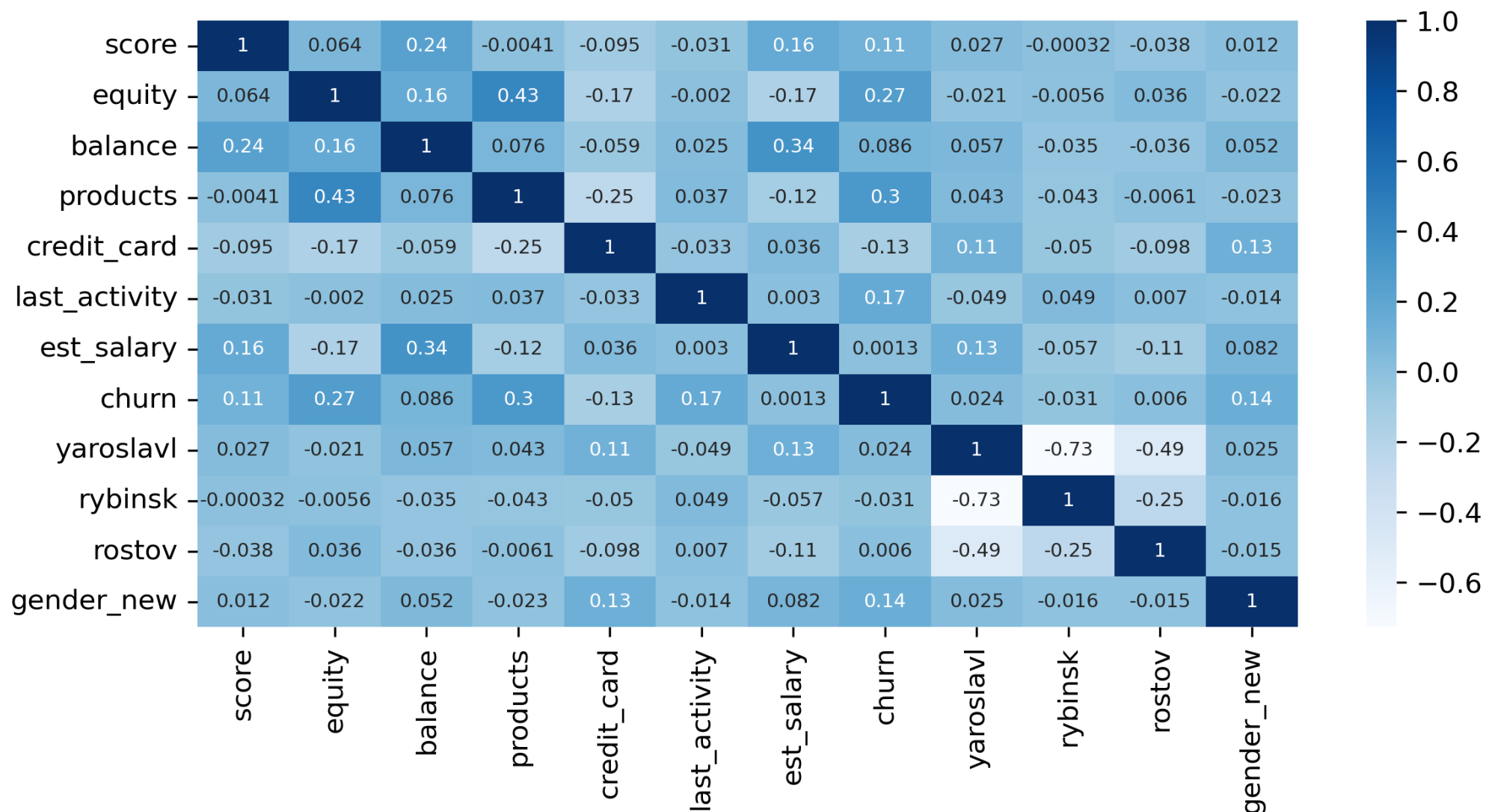
# Задачи проекта

1. Загрузка и предобработка данных
2. Исследование корреляции параметров клиентов
3. Проверка статистических гипотез
4. Исследование оттока
5. Сегментация пользователей по количеству потребляемых продуктов
6. Формулировка выводов и рекомендаций для отдела маркетинга

# Исходные данные

- датасет содержит данные о 10 тыс клиентов банка из 3 филиалов
- выполнен анализ пропусков и поиск дубликатов
- переменные city и gender переведены в бинарный вид
- проведен анализ выбросов и категоризация для переменных score, balance, est\_salary

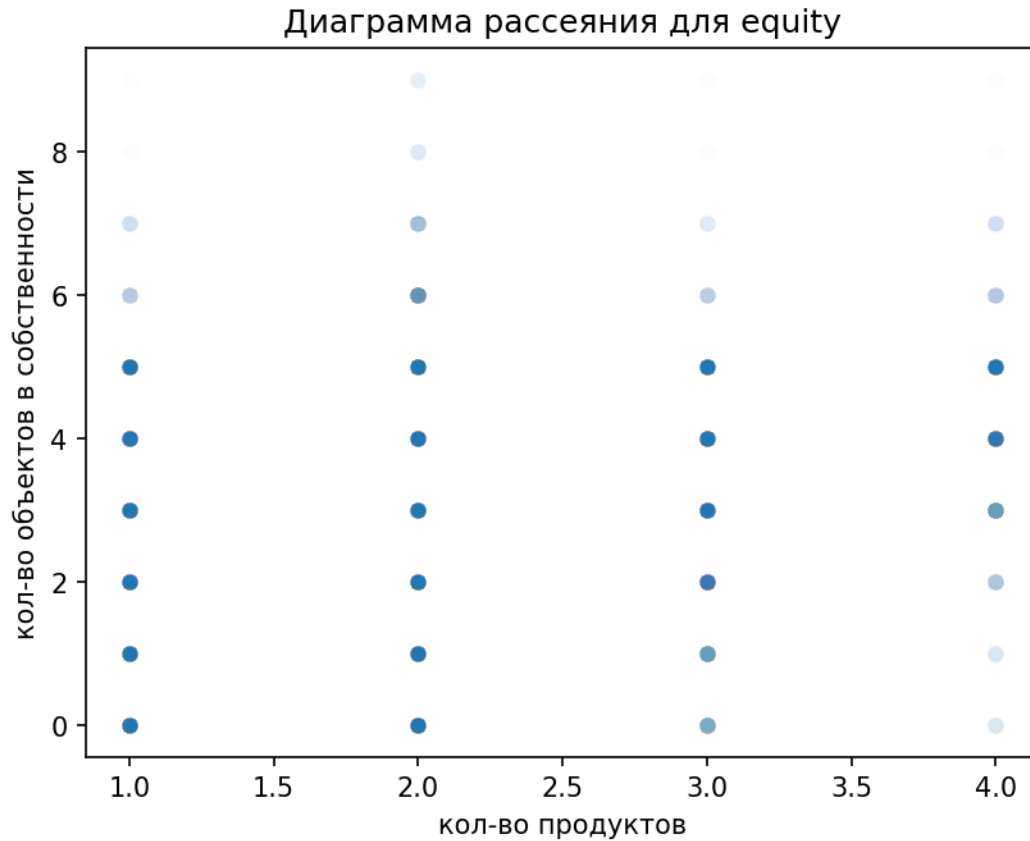
# Корреляция параметров клиентов



умеренная корреляция [0.3-0.5) наблюдается между:

- products и equity
- products и churn
- est\_salary и balance

# Взаимосвязь количества продуктов и количества объектов в собственности (equity)

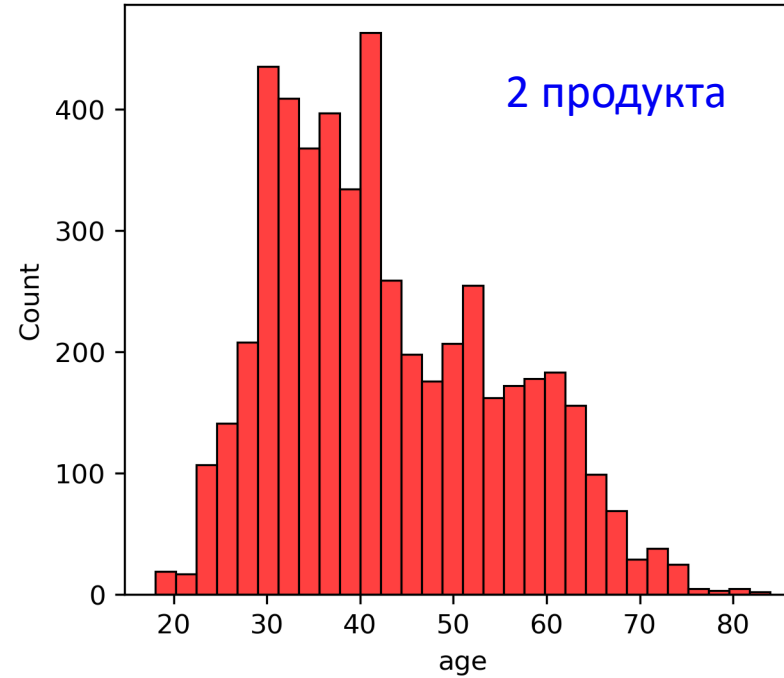
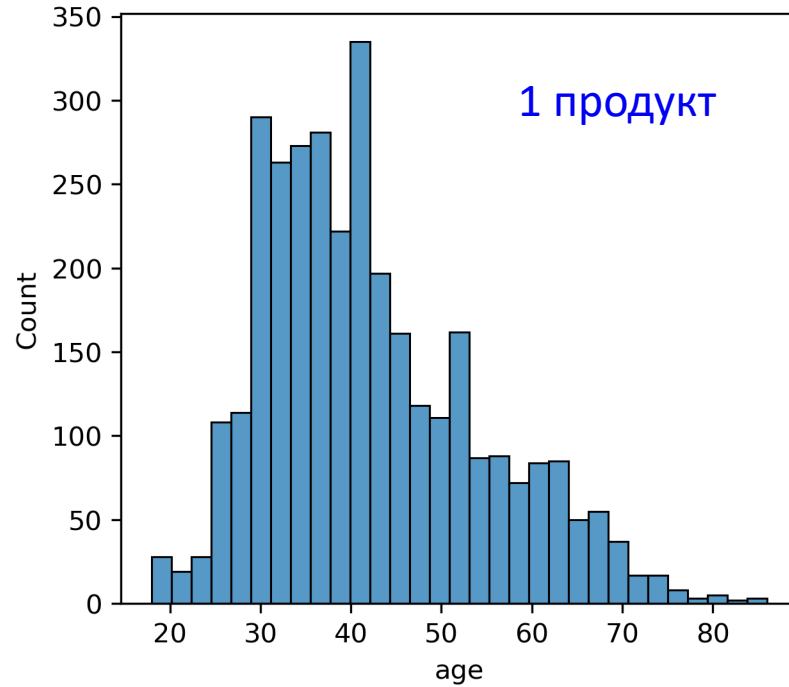


- для клиентов с 1 продуктом количество объектов в собственности варьируется от 0 до 5
- с 2 продуктами также от 0 до 5
- с 3 продуктами - от 2 до 4
- с 4 продуктами от 4 до 5

# Проверка статистических гипотез (I)

H0: Возраст клиентов, которые пользуются одним и двумя продуктами не отличается

H1: Возраст клиентов, которые пользуются одним и двумя продуктами отличается



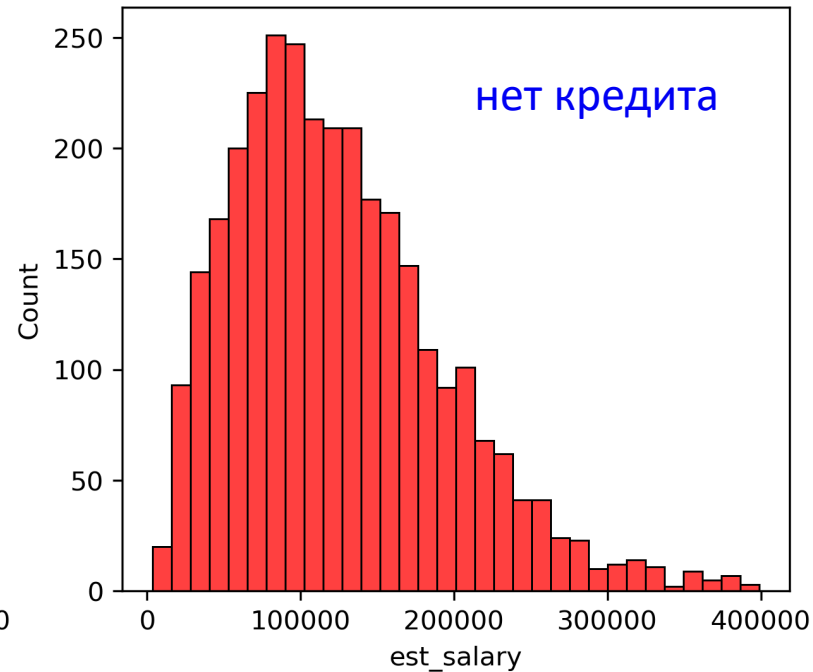
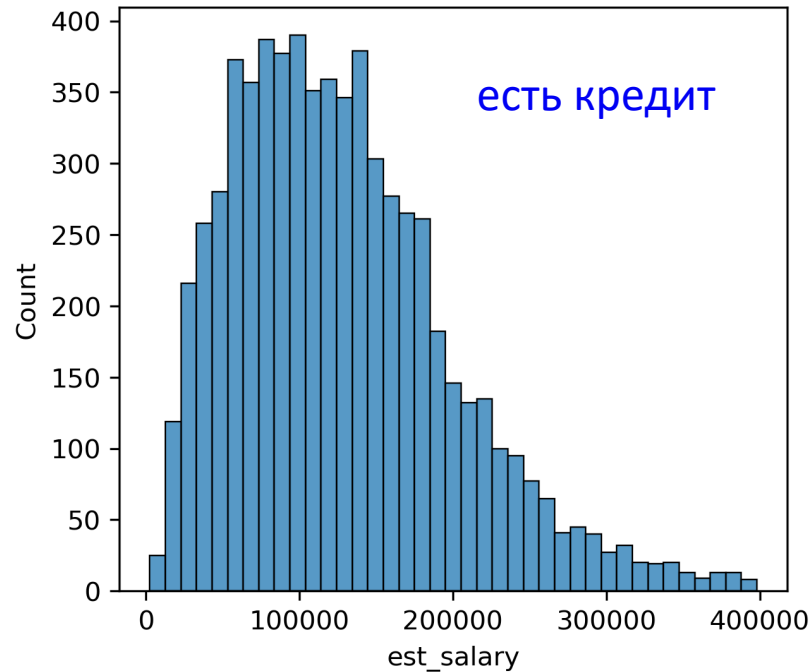
Распределение клиентов по возрасту в выборках с 1 и 2 продуктами

При уровне статистической значимости 5% нулевая гипотеза отвергнута  
(p-value = 3.3%)

# Проверка статистических гипотез (II)

H0: Заработные платы клиентов с кредитом и без равны

H1: Заработные платы клиентов с кредитом и без отличается

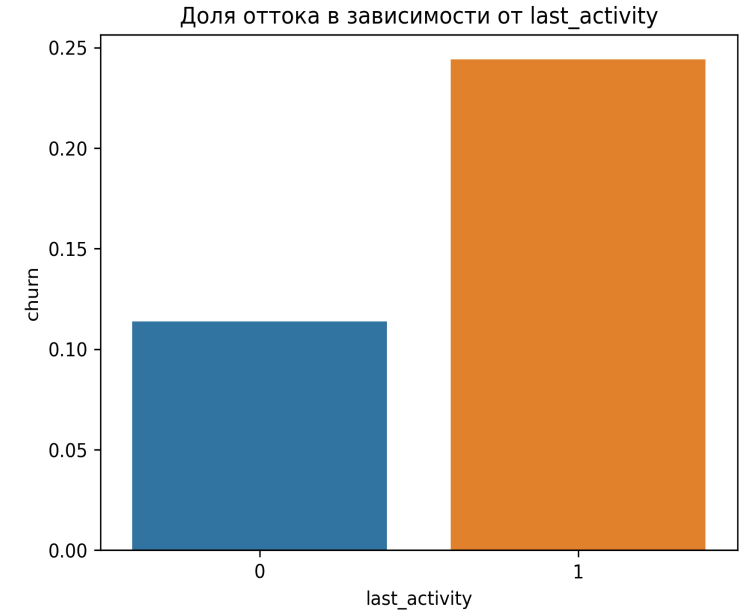
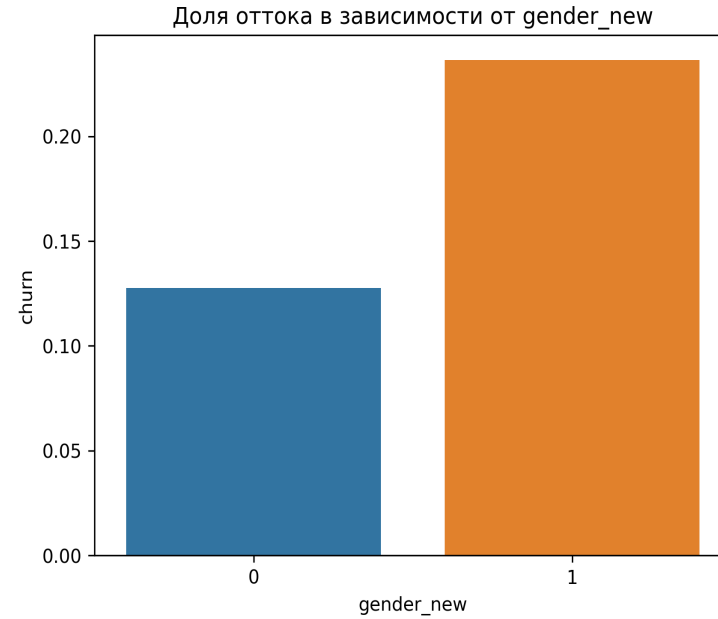
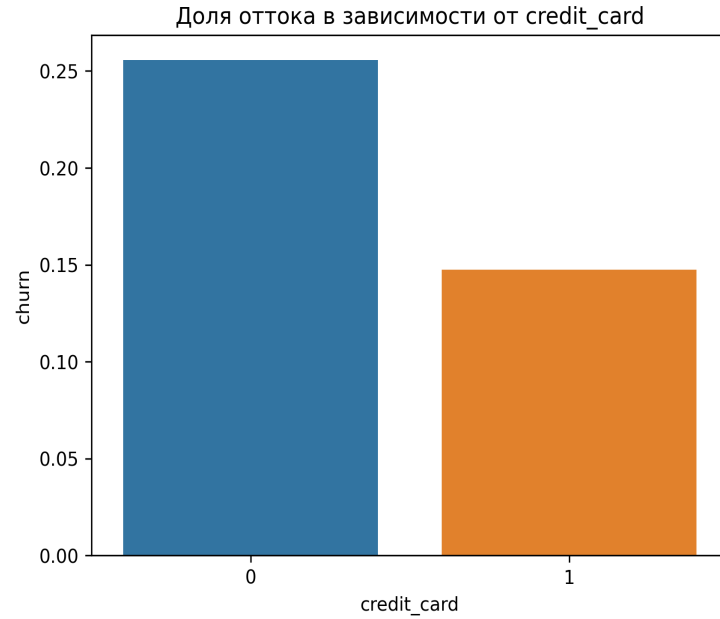


Распределение заработной платы у клиентов с кредитом и без

При уровне статистической значимости 5% нулевая гипотеза отвергнута  
(p-value = 4.5%)

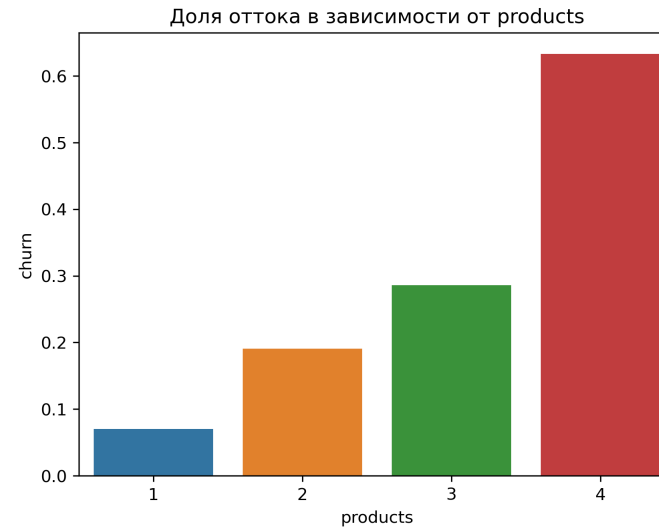
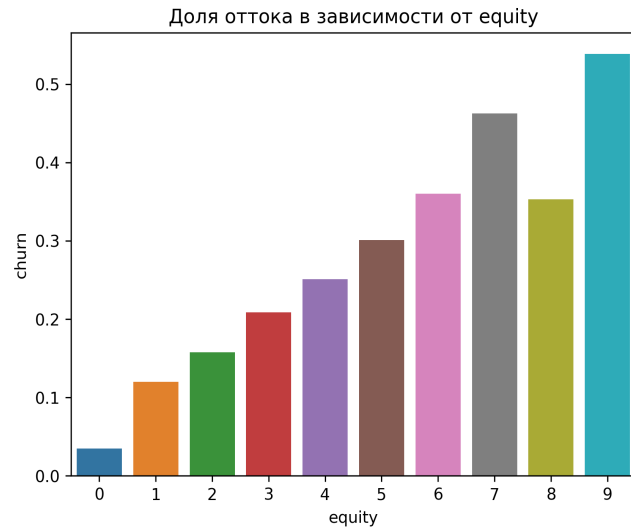
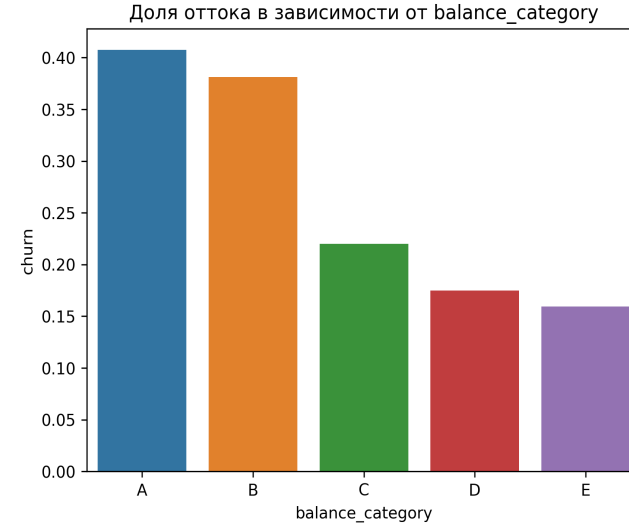
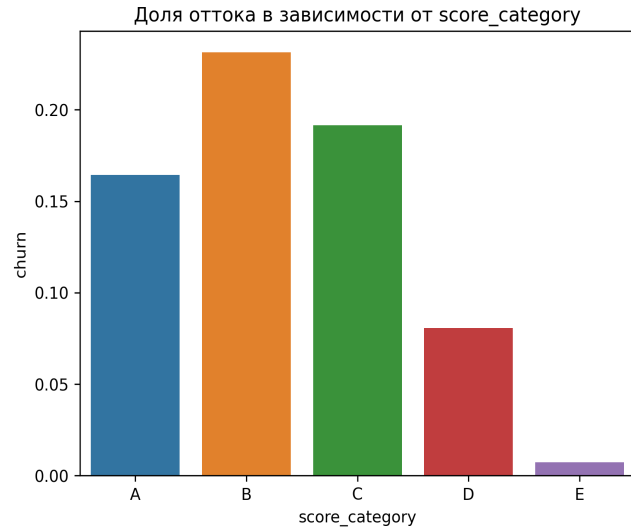


# Исследование оттока (I)



Отток выше у клиентов без кредитной карты, у мужчин и клиентов с недавней активностью

# Исследование оттока (II)



Отток растёт при увеличении

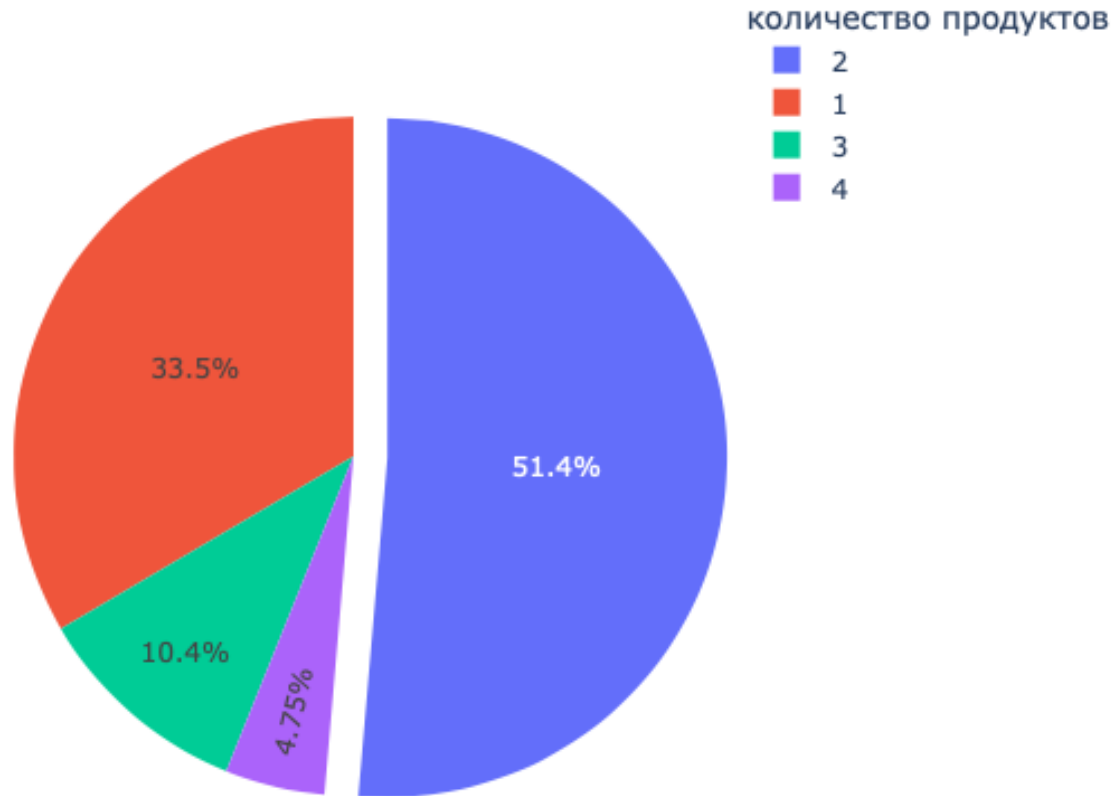
- категории score
- категории баланса
- количества объектов в собственности
- количества продуктов

# Исследование оттока (III)



наибольший отток в диапазонах 25-29 и 52-59 лет

# Сегментация пользователей по количеству продуктов (I)



51% клиентов пользуется 2 продуктами

Общее распределение пользователей  
по количеству продуктов

# Сегментация пользователей по количеству продуктов (II)

Средние значения параметров для каждого сегмента

	score	age	equity	balance	credit_card	last_activity	est_salary	churn	yaroslavl	rostov	rybinsk	gender_new
products												
1	854.298114	42.196208	1.324753	6.108238e+05	0.863813	0.528285	180731.037920	0.070338	0.608201	0.132894	0.258905	0.540856
2	844.563792	42.768900	3.151190	8.131345e+05	0.625439	0.516192	130395.346490	0.190987	0.537846	0.157238	0.304916	0.477175
3	837.307988	45.714836	3.518768	9.940135e+05	0.356112	0.416747	128654.951925	0.285852	0.748797	0.091434	0.159769	0.430221
4	878.267932	39.782700	4.151899	1.267821e+06	0.681435	0.789030	148788.952869	0.632911	0.647679	0.143460	0.208861	0.622363

Общие характеристики всех сегментов:

- большинство клиентов из Ярославля
- est\_salary в диапазоне 130 000-180 000

Для сегментов 1,2,3:

- средний возрастов клиентов составляет 42-46 лет
- значение score в диапазоне 837-854
- активность 42-53%

# Сегментация пользователей по количеству продуктов (III)

Средние значения параметров для каждого сегмента

	score	age	equity	balance	credit_card	last_activity	est_salary	churn	yaroslavl	rostov	rybinsk	gender_new
products												
1	854.298114	42.196208	1.324753	6.108238e+05	0.863813	0.528285	180731.037920	0.070338	0.608201	0.132894	0.258905	0.540856
2	844.563792	42.768900	3.151190	8.131345e+05	0.625439	0.516192	130395.346490	0.190987	0.537846	0.157238	0.304916	0.477175
3	837.307988	45.714836	3.518768	9.940135e+05	0.356112	0.416747	128654.951925	0.285852	0.748797	0.091434	0.159769	0.430221
4	878.267932	39.782700	4.151899	1.267821e+06	0.681435	0.789030	148788.952869	0.632911	0.647679	0.143460	0.208861	0.622363

Для первого сегмента характерно:

- отсутствие объектов в собственности
- наличие кредитной карты
- отток ~7%
- низкий баланс
- большая доля мужчин

Для второго сегмента характерно:

- 3 объекта в собственности
- наличие кредитной карты
- отток ~20%
- более высокий баланс
- большая доля женщин

# Сегментация пользователей по количеству продуктов (IV)

Средние значения параметров для каждого сегмента

	score	age	equity	balance	credit_card	last_activity	est_salary	churn	yaroslavl	rostov	rybinsk	gender_new
products												
1	854.298114	42.196208	1.324753	6.108238e+05	0.863813	0.528285	180731.037920	0.070338	0.608201	0.132894	0.258905	0.540856
2	844.563792	42.768900	3.151190	8.131345e+05	0.625439	0.516192	130395.346490	0.190987	0.537846	0.157238	0.304916	0.477175
3	837.307988	45.714836	3.518768	9.940135e+05	0.356112	0.416747	128654.951925	0.285852	0.748797	0.091434	0.159769	0.430221
4	878.267932	39.782700	4.151899	1.267821e+06	0.681435	0.789030	148788.952869	0.632911	0.647679	0.143460	0.208861	0.622363

Для третьего сегмента характерно:

- 3-4 объекта в собственности
- отсутствие кредитной карты
- отток ~30%
- более высокий баланс
- большая доля женщин

Для четвертого сегмента характерно:

- 4 объекта в собственности
- наличие кредитной карты
- активность
- отток ~63%
- высокий баланс
- большая доля мужчин
- средний возраст 38 лет

# Выводы (I)

В работе исследованы данные клиентов регионального банка с целью выявления факторов оттока.

Установлено:

- при увеличении количества объектов в собственности и продуктов отток увеличивается
- отток уменьшается с повышением параметра score и баланса на счете
- доля отточных клиентов ниже для клиентов с кредитными картами
- доля отточных клиентов выше для клиентов с недавней активностью и для мужчин
- наибольший отток в диапазонах 25-29 и 52-59 лет



## Выводы (II)

На основе сегментации по количеству продуктов установлено:

- доля клиентов с кредитной картой самая высокая для сегмента с 3 продуктами
- доля неактивных клиентов самая высокая для сегмента с 3 продуктами
- доля отточных клиентов растет с увеличением количества продуктов и самая высокая для сегмента с 4 продуктами
- распределение клиентов по возрасту практически не зависит от количества продуктов: пик использования приходится на 30-45 лет

# Рекомендации

Для снижения оттока можно рекомендовать предоставлять кредитные продукты на более выгодных условиях, предложить лучшие условия для мужчин и клиентов в возрасте 25-30 и 45-60 лет, особые условия предложить клиентам с 3 продуктами, чтобы повысить их активность и снизить отток.

Для первого сегмента можно порекомендовать, обратить внимание на привлечение женщин и клиентов в возрасте 30-40 лет, для второго - предлагать более выгодные условия по кредиту, для 3 сегмента - также предложить более выгодные условия по кредиту и обратить внимание на привлечение мужчин, в 4 категории больше привлекать женщин, клиентов с высоким уровнем дохода, предлагать выгодную программу лояльности, тем самым повышать средний возраст клиентов