

Региональный банк "Метанпромбанк"

**Сегментация пользователей по
потреблению**

Содержание

• Цели и задачи	3
• Исходные данные	4
• Результаты EDA	5
• Основные характеристики выборки	6
• Распределения признаков	7
• Матрица корреляции	8
• Кластеризация	9
• Проверка гипотез	10
• Рекомендации	11
• Дашборд	14
• Контакты	16

Цели и задачи

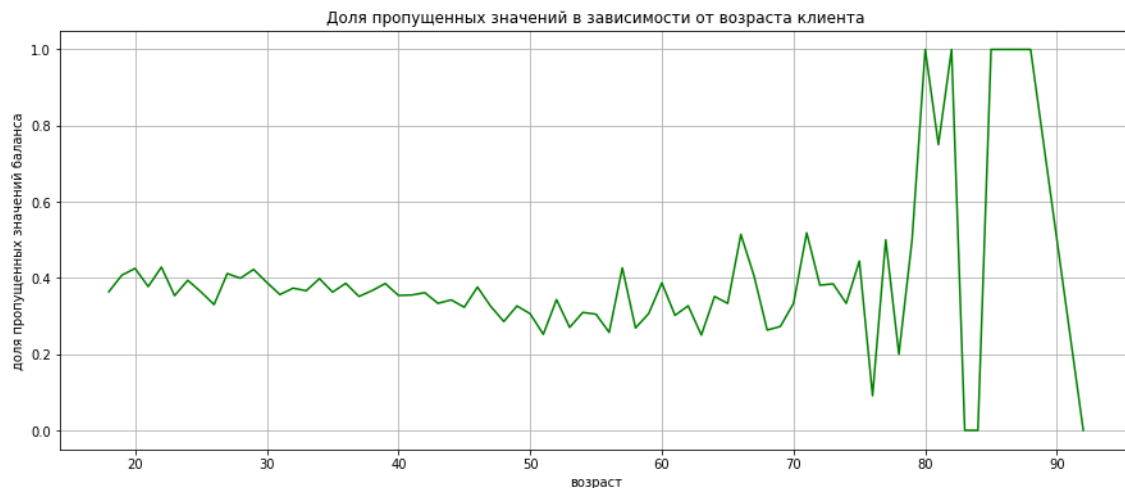
- ❑ **Задача исследования:** анализ клиентов банка и сегментация пользователей по количеству потребляемых продуктов
- ❑ **Цель исследования:** выявление сегментов, склонных к оттоку, для выработки оптимальных стратегий удержания клиентов из каждого сегмента
- ❑ **Результаты исследования:** рекомендации по удержанию клиентов из каждого сегмента

Исходные данные

	userid	score	City	Gender	Age	Objects	Balance	Products	CreditCard	Loyalty	estimated_salary	Churn
0	15677338	619	Ярославль	Ж	42	2	NaN	1	1	1	101348.88	1
1	15690047	608	Рыбинск	Ж	41	1	83807.86	1	0	1	112542.58	0
2	15662040	502	Ярославль	Ж	42	8	159660.80	3	1	0	113931.57	1
3	15744090	699	Ярославль	Ж	39	1	NaN	2	0	0	93826.63	0
4	15780624	850	Рыбинск	Ж	43	2	125510.82	1	1	1	79084.10	0

Обнаруженных пропусков значения в столбце balance для Рыбинска и Ярославля - более 36%. Больше всего пропуски у клиентов пожилого поколения.

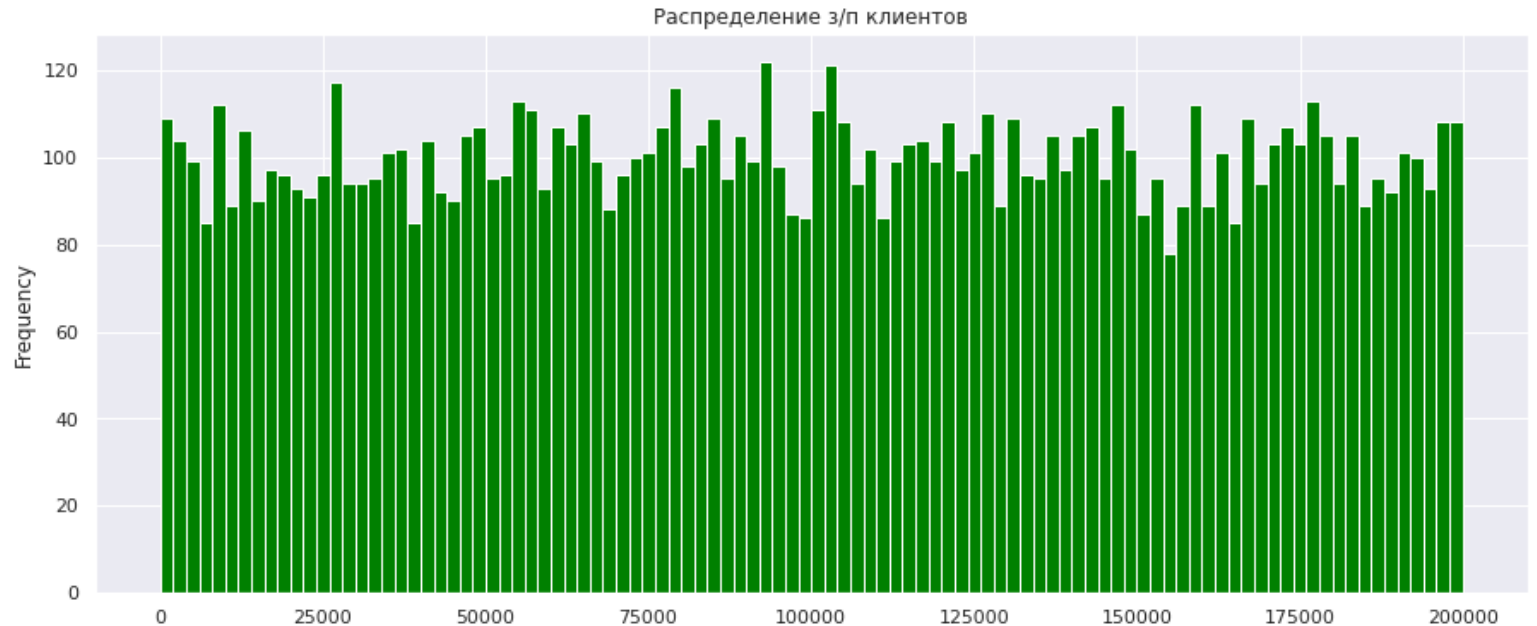
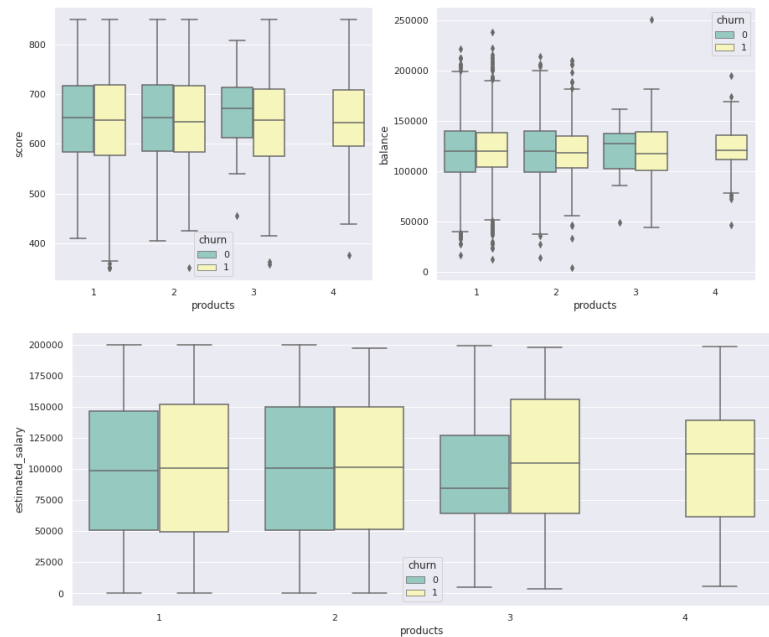
В Ростове Великом пропусков не обнаружено.



Файл с данными о 10000 клиентах банка, содержащий информацию:

- userid — идентификатор пользователя
- score — баллы кредитного скоринга
- City — город
- Gender — по
- Age — возраст
- Objects — количество объектов в собственности
- Balance — баланс на счёте
- Products — количество продуктов
- CreditCard — есть ли кредитная карта
- Loyalty — активный клиент
- estimated_salary — з/п клиента
- Churn — факт оттока

Результаты исследовательского анализа данных



Данные довольно неоднозначны.

Присутствуют явные выбросы и противоречия.

Клиенты банка, имеющие 4 продукта - все в оттоке.

В значениях кредитного скоринга встречаются аномально низкие значения, а значения баланса встречаются как аномально низкие, так и аномально высокие.

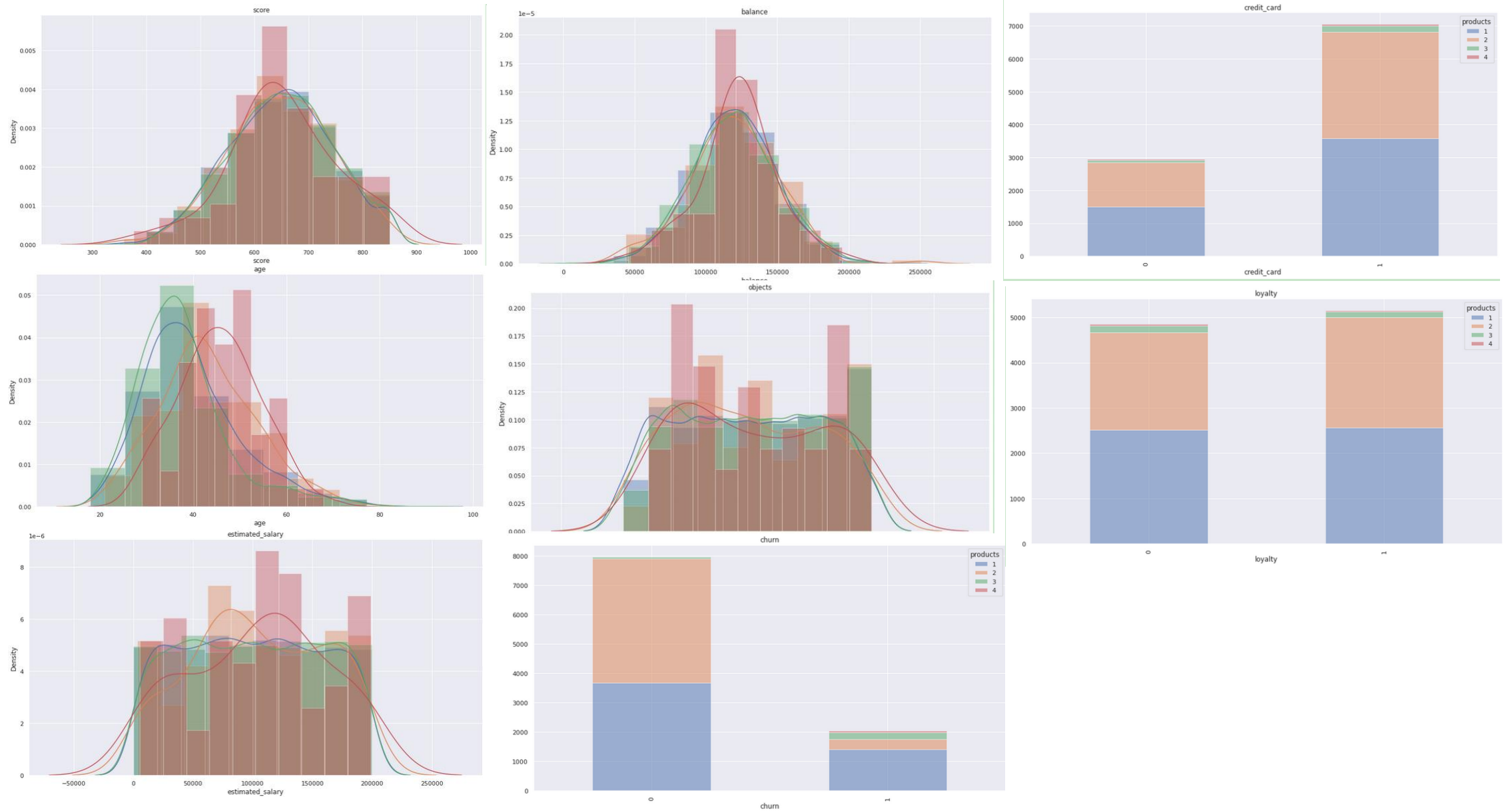
Распределение признака заработная плата клиента имеет вид равномерного, это может указывать на следующее:

- предоставленная выборка является неполной;
- значение признака заполнено автоматически в качестве простого априорного допущения.

Основные характеристики исследуемой выборки

- ☐ в выборке присутствуют пользователи обоих полов (мужчины и женщины) из 3 городов
- ☐ средний возраст клиента - 39 лет, минимальный возраст - 18, максимальный - 92 года
- ☐ баллы кредитного скоринга распределены в диапазоне 350 - 850
- ☐ объектов собственности от 0 до 10
- ☐ в среднем на балансе у клиента - около 120т.р., минимум 3800р., максимум 250т.р.
- ☐ представленная выборка клиентов имеет от 1 до 4 банковских продуктов
- ☐ кредитные карты есть у 70% клиентов
- ☐ в выборке преобладают активные клиенты
- ☐ средняя з/п клиента - 100т.р., минимум 12р., максимум - 200т.р., в выборке присутствуют аномальные значения - выбросы

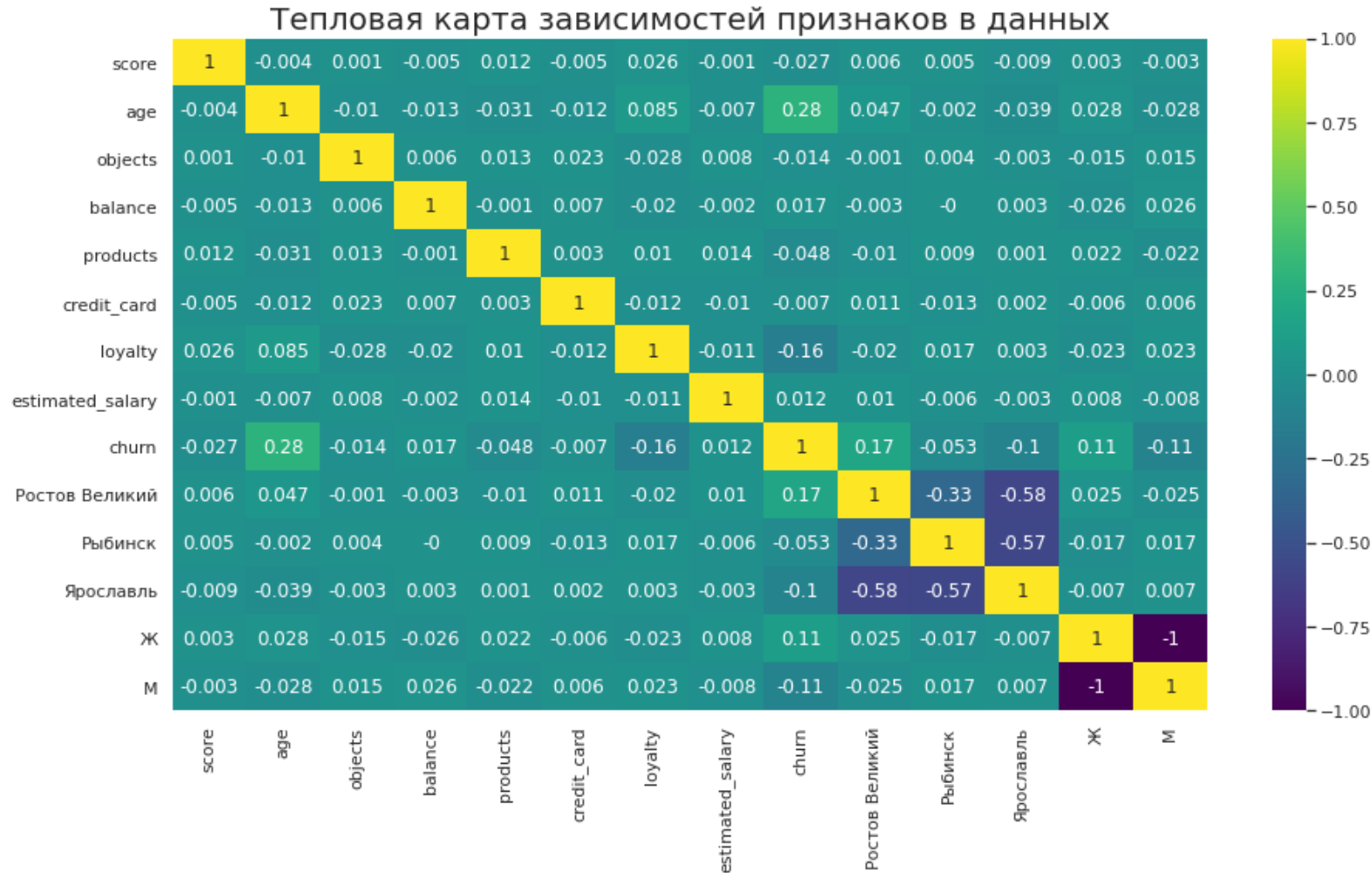
Построение распределений признаков



Выводы по распределениям

- средний возраст клиентов, имеющих 3 и 4 продукта - выше, чем клиентов, имеющих 1 или 2 продукта;
 - средний баланс на счетах вне зависимости от количества продуктов - примерно одинаков (около 120т.р.);
 - клиенты банка, имеющие 4 продукта, со 100% вероятностью попадают в отток, владельцы 3 продуктов - с вероятностью 83%;
 - с увеличением средней з/п клиента, увеличивается количество приобретённых продуктов;
 - кредитные карты есть у 70% клиентов, вне зависимости от того, сколько у него банковских продуктов;
 - активных клиентов - около половины, также не зависит от количества продуктов;
 - в среднем клиент банка владеет 5 объектами в собственности;
 - средняя кредитоспособность (баллы скоринга) не зависят от количества продуктов и в среднем составляют около 650;
 - женщин, приобретающих 3 и 4 продукта заметно больше, чем мужчин, тогда как мужчин с 1 или 2 кредитными продуктами - больше, чем женщин;
 - в Ростове Великом и Ярославле клиентов, имеющих 3 или 4 банковских продукта явно больше, чем в Рыбинске, а клиентов с 1 и 2 продуктами заметно больше в Ярославле, чем в Ростове Великом и Рыбинске.
- присутствует нарушение логики при построении зависимости отток /кол-во продуктов = 3 и 4, т.е. получается, что клиенты банка, имеющие 4 продукта все - в оттоке, хотя интуитивно понятно, что они должны являться активными участниками процесса;
 - средний возраст владельцев 4 продуктов - выше, чем владельцев 1, 2 или 3 продуктов;
 - сильный отток клиентов присутствует у клиентов, владеющих 1 банковским продуктом;
 - минимален отток клиентов с 2 продуктами (в пропорции гораздо ниже, чем у клиентов с 1, 3 и 4).

Поиск признаков, напрямую влияющих на отток



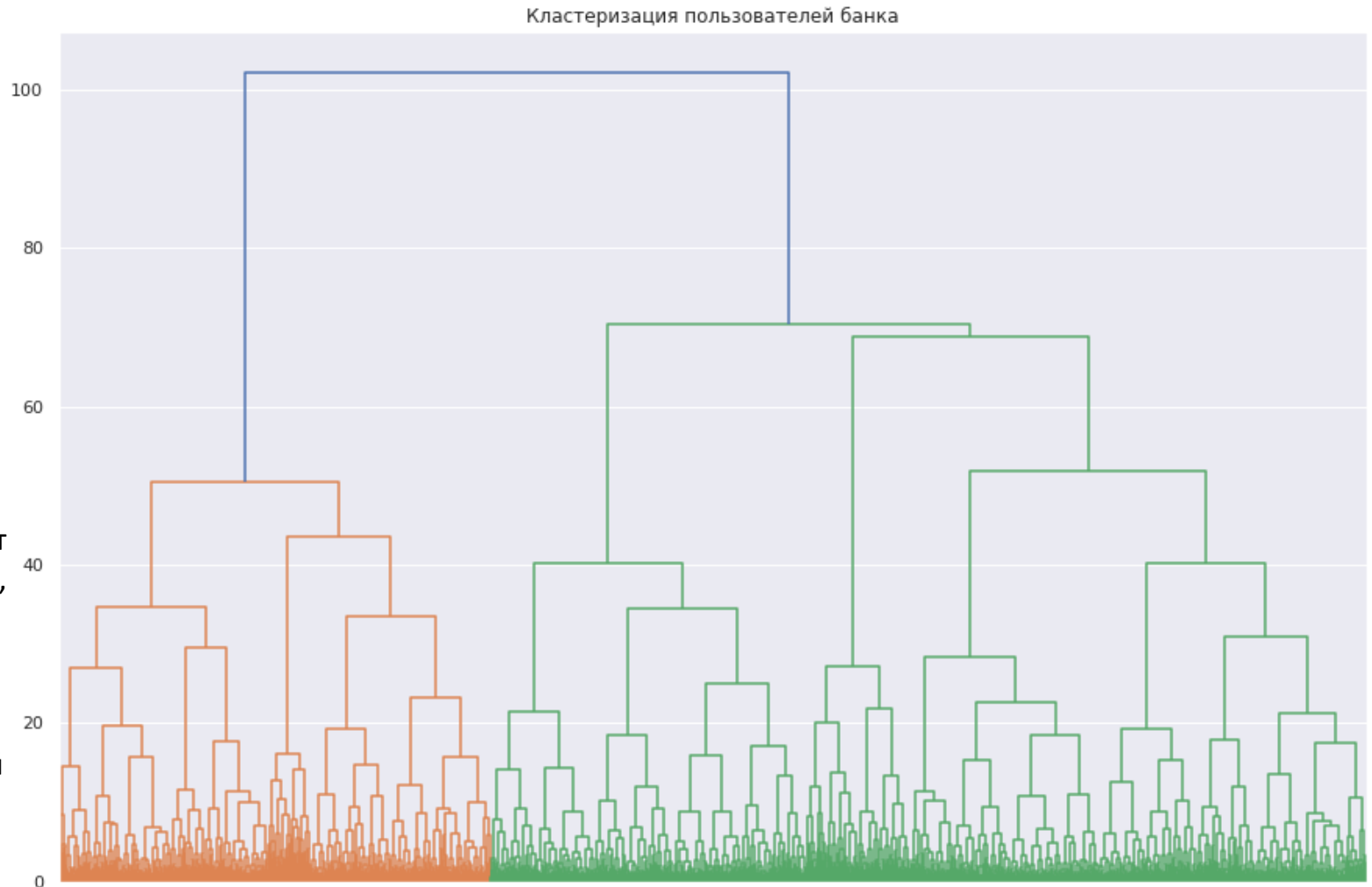
Признаки, сильнее всего влияющие на отток клиентов:

- ✓ возраст клиента
- ✓ активность клиента
- ✓ проживание в Ростове Великом
- ✓ пол клиента

Проведение кластеризации пользователей

Результаты для 5 кластеров:

- ❑ Возраст клиентов с большим доходом (в среднем 155т.р.) - около 36 лет, обычно они имеют 1 продукт, процент оттока таких клиентов составляет около 23%
- ❑ 3/п владельцев 2 продуктов обычно около 130т.р., средний возраст - 37 лет, а процент оттока - также около 24%
- ❑ Чуть ниже процент оттока клиентов со средней з/п 54-55т.р.: он составляет не более 20%, возраст таких клиентов - около 36 лет, причём у тех, у кого з/п 145т.р. имеют баллы кредитного скоринга - около 600, а те, у кого з/п составляет 100т.р.- выше - 700
- ❑ Лидером по оттоку (почти половина) становится пожилой клиент (средний возраст 59 лет) с балансом 120т.р. и средней з/п - 92т.р. Обычно имеют 1, 2 банковских продукта



Проверка гипотез

- ✓ **Гипотеза 1** - доход клиентов, пользующихся 1 и 2 банковскими продуктами, различаются
- ✓ **Гипотеза 2** - у владельцев кредитных карточек и у тех, у кого кредиток нет, баллы скоринга различаются
- ✓ **Гипотеза 3** - доля оттока мужчин и женщин в Ярославле различается
- ✓ **Гипотеза 4** - доля оттока в Ростове Великом и в Ярославле различается
- ✓ **Гипотеза 5** - количество денег на балансе влияет на получение статуса активности
- ✓ **Гипотеза 6** - доля оттока у активных и неактивных пользователей различается
- ✓ **Гипотеза 7** - уровень дохода в Ярославле и в других городах различается

Результаты проверки гипотез

- ✓ Разницы в среднем доходе у пользователей 1 и 2 банковских продуктов нет
- ✓ Разницы в баллах кредитного скоринга у пользователей, имеющих и не имеющих кредитки, нет
- ✓ Женщины в Ярославле чаще перестают пользоваться услугами банка, чем мужчины
- ✓ Доли оттока клиентов в Ярославле и Ростове Великом различаются
- ✓ Разницы в среднем балансе у активных и неактивных клиентов нет
- ✓ Доли оттока активных и неактивных клиентов различаются
- ✓ Разницы в среднем доходе у пользователей в Ярославле, Рыбинске и Ростове Великом нет

Рекомендации по удержанию клиентов

- ❑ делать значительные скидки на второй продукт
- ❑ проводить мероприятия по увеличению активности клиентов (особое внимание - Ростов Великий)
- ❑ применять в Рыбинске и Ростове Великом похожие акции, что и в Ярославле, т.к. уровень дохода в городах примерно равный, а отток в Ярославле меньше всего
- ❑ для удержания лиц женского пола и клиентов в возрасте (>50 лет), предлагать им выгодные условия и дарить подарки

Дашборд

- ✓ С дашбордом в Tableau можно ознакомиться по ссылке: [Kots_NS_dash_final](#)

Dash_KotsNS_bank by [Nataliya](#)

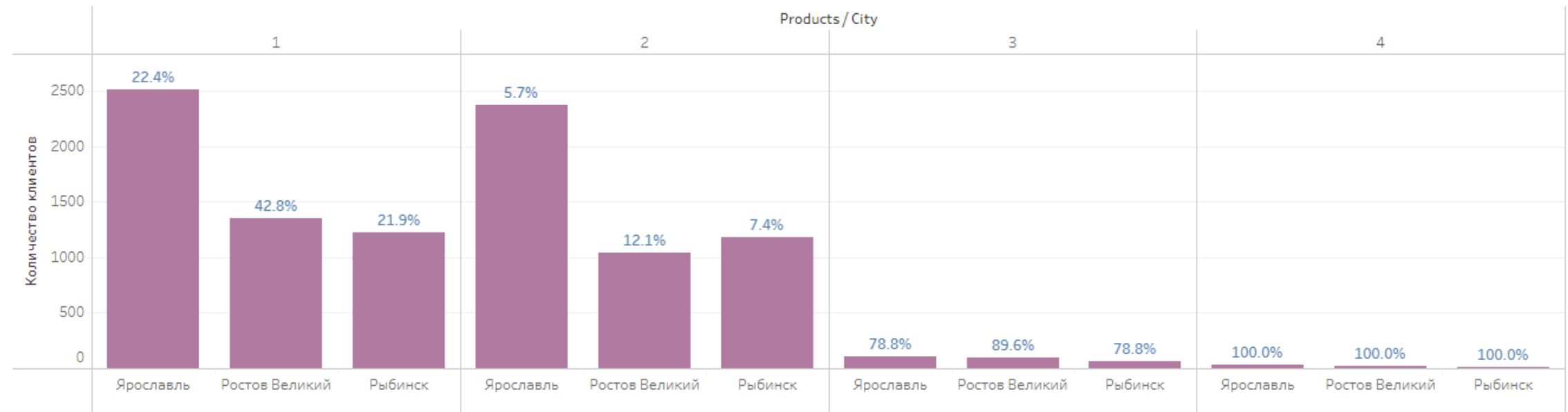


Количество клиентов в зависимости от числа используемых продуктов

- Данная диаграмма отражает количество клиентов банка в зависимости от числа используемых продуктов.
- Индикатор процентов показывает долю ушедших клиентов.
- В данном дашборде существует возможность отфильтровать данные по городу.

Выбор города

- ☒ (All)
- ☒ Ростов Великий
- ☒ Рыбинск
- ☒ Ярославль



Данные и контакты

Дата начала проведения исследования: 27/05/2022

Дата окончания исследования: 10/06/2022

Презентацию подготовила
студентка потока da_37
Котс Наталия

email: kotsnat@gmail.com
+7-915-433-33-39