

Проект SQL: Анализ оттока клиентов банка

Code ▾

Рахманова Амина, afrakhmanova

Задача

Проанализировать данные **оттока клиентов банка** и представить результаты в компактной и понятной форме.

- Выяснить, какие категории клиентов банка не заключают новые договоры.
- Подумать над возможными вариантами развития событий для сокращения оттока и проверить их на моделях.

Анализ

Данные и логика анализа

Мы исследуем базу данных клиентов банка и имеем следующие таблицы и значения:

Code

```
[1] "country"      "portfolio" "profile"
```

Code

```
[1] "Country"      "CountryId"
```

Code

```
[1] "CustomerId"      "CreditScore"      "Tenure"            "Balance"
[5] "NumOfProducts"  "HasCrCard"        "IsActiveMember"    "Exited"
[9] "CountryId"
```

Code

```
[1] "CustomerId"      "Gender"            "Age"               "EstimatedSalary"
[5] "CountryId"
```

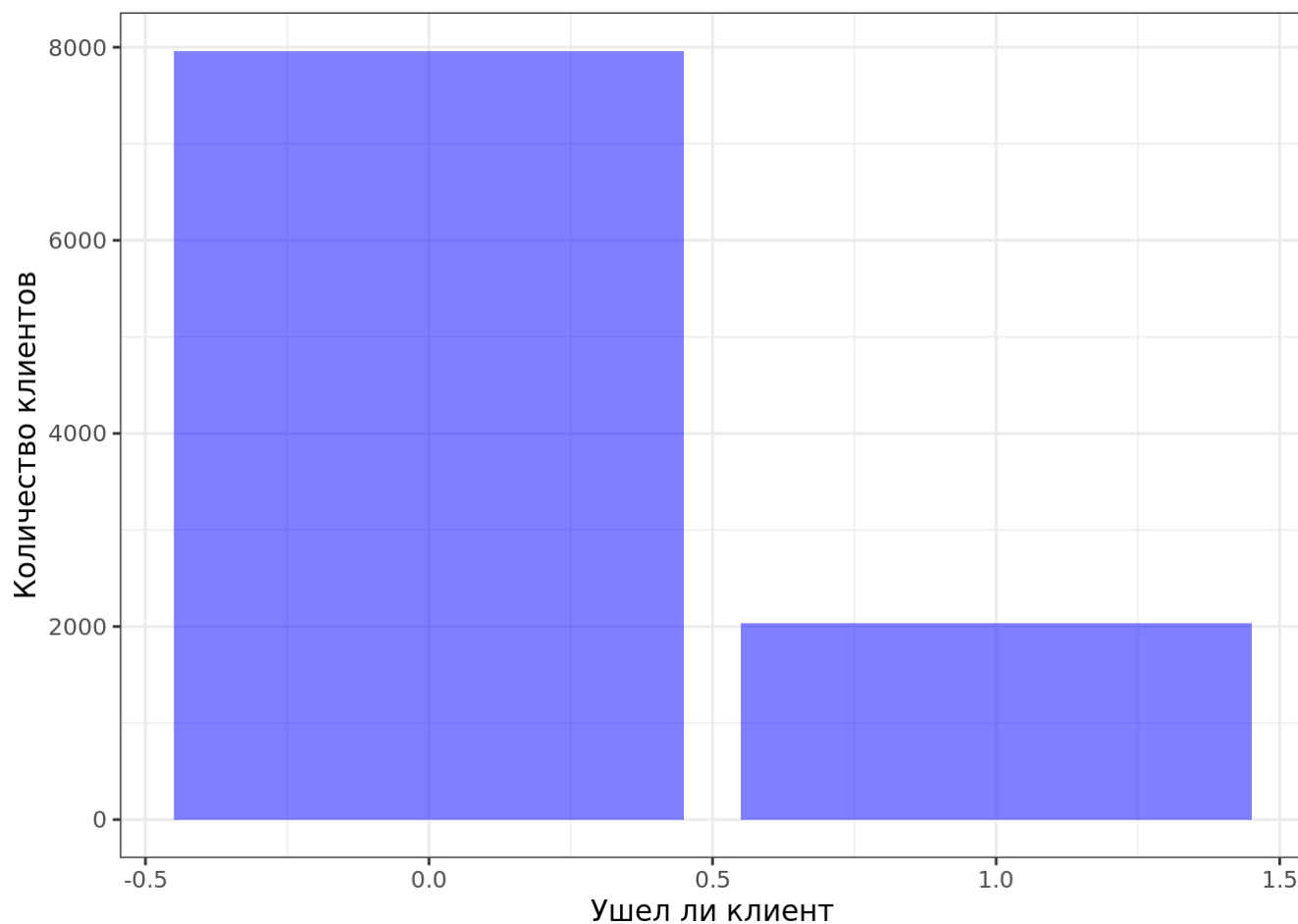
Проверим распределение по оттоку:

Code

Exited	n
<int>	<S3: integer64>
1	2037
0	7963
2 rows	

[Code](#)

```
[1] 20.37
```

[Code](#)

Посмотрим, из каких стран и с каким доходом наши клиенты:

[Code](#)

Country	n
<chr>	<S3: integer64>
France	5014
Germany	2509
Spain	2477
3 rows	

Клиенты банка из трех стран: Франция, Германия и Испания. Будем рассматривать все три, без изменений.

Что касается дохода, посмотрим основную информацию по заработной плате:

[Code](#)

```
EstimatedSalary
Min.      :  11.58
1st Qu.: 51002.11
Median :100193.91
Mean      :100090.24
3rd Qu.:149388.25
Max.      :199992.48
```

При среднем значении заработной платы в 100000 единиц (предположительно евро, учитывая страны) минимальное всего 11. В целом, по трем странам средняя зарплата составляет более 1000 евро, поэтому стоит отсечь данные ниже этого значения для качественного анализа.

Проверим распределение клиентов банка по возрасту:

[Code](#)

Age <int>	n <S3: integer64>
18	22
19	27
20	40
21	53
22	84
23	99
24	132
25	154
26	200
27	209

1-10 of 70 rows

Previous1234567Next

Распределение количества клиентов по возрасту не очень равномерное. С 24 лет количество клиентов превышает 100, что уже является неплохим показателем для анализа. Тем более в среднем после 23 лет студенты (окончившие бакалавриат) устраиваются на постоянную работу (имеется в виду не подработка, а полная занятость). Пенсионный возраст в исследуемых странах превышает 60 лет, отсечем с этого значения, так как дальше слишком мало показателей большего возраста.

В дальнейшем будем использовать отфильтрованные данные по возрасту от 24 до 60 лет, определяющему основной рабочий класс людей, которые как зарабатывают, так и тратят, что является причиной использования продуктов банка.

Также посмотрим сколько месяцев клиенты пользуются услугами банка:

[Code](#)

Tenure <int>	n <S3: integer64>
0	413
1	1035

Tenure	n
<int>	<S3: integer64>
2	1048
3	1009
4	989
5	1012
6	967
7	1028
8	1025
9	984
1-10 of 11 rows	
Previous 1 2 Next	

Логично выбрать клиентов, которые подключились не менее месяца назад. Тогда они уже успели оценить качество и пользу услуг данного банка и приняли решение - остаться или уйти.

Таким образом, для исследования я отобрала клиентов по нескольким признакам:

- возраст от 23 до 60
- заработная плата больше 1000
- клиент сотрудничает с банком не менее месяца

Code

Проверим новое распределение оттока в сравнении с начальными показателями:

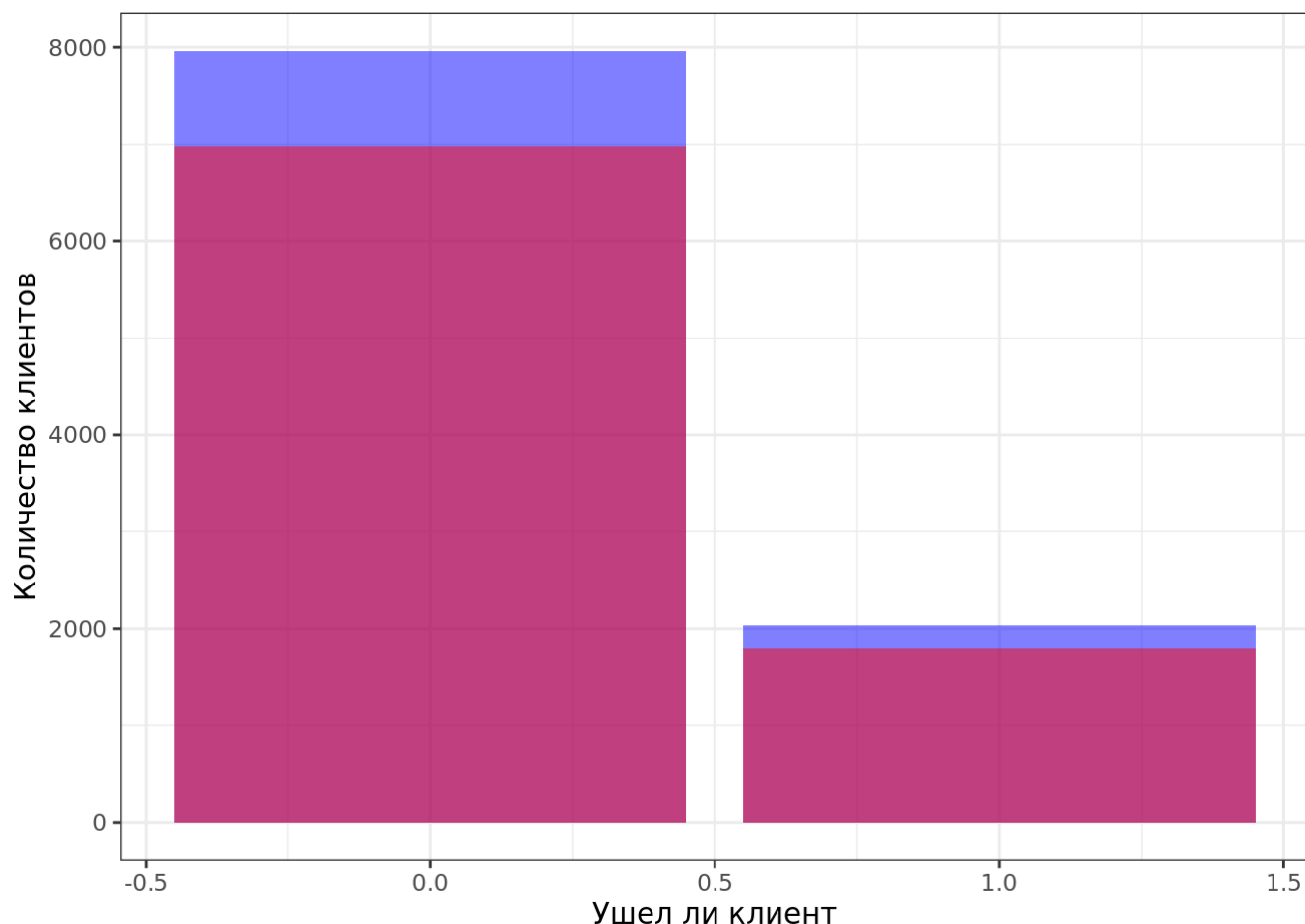
Code

Exited	n
<int>	<int>
0	6986
1	1791
2 rows	

Code

[1] 20.41

Code



Таким образом в отфильтрованных данных осталось 8777 человек, что больше 85% исходных данных.

В целом, показатели распределения целевой переменной уменьшились. При этом доля ушедших клиентов немного увеличилась на общем фоне, что будет на пользу при прогнозировании.

Модель

Для анализа зависимой переменной Exited (1 - клиент ушел, 0 - остался) я выбрала несколько ключевых на мой взгляд показателей:

- Age - возраст клиента
- Balance - текущий баланс
- Country - название страны
- NumOfProducts - количество продуктов банка, которыми пользуется клиент (факторная переменная, имеющая 4 уровня)
- HasCrCard - есть ли кредитная карта (фактор, где 1 - да)
- IsActiveMember - активный ли клиент (фактор, где 1 - да)

Так выглядят данные:

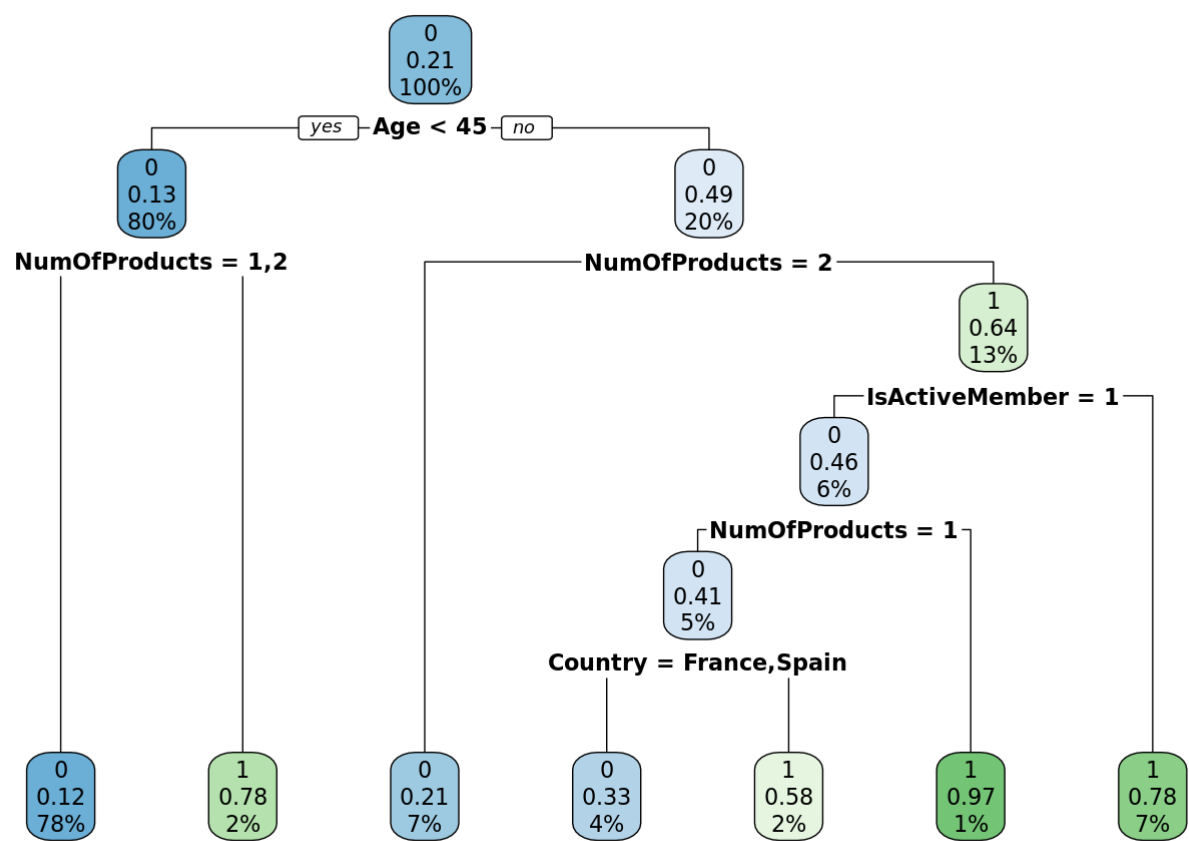
...		Balance	Country	NumOfProducts	HasCrCard	IsActiveMember	Exited
<int>		<dbl>	<fctr>	<fctr>	<fctr>	<fctr>	<fctr>
1	42	0.00	France	1	1	1	1
2	41	83807.86	Spain	1	0	1	0

...	Balance	Country	NumOfProducts	HasCrCard	IsActiveMember	Exited
<int>	<dbl>	<fctr>	<fctr>	<fctr>	<fctr>	<fctr>
3 42	159660.80	France	3	1	0	1
4 39	0.00	France	2	0	0	0
5 43	125510.82	Spain	1	1	1	0

5 rows

Построим модель в виде дерева решений:

Code



Оценим качество модели:

Code

.metric	.estimate
<chr>	<dbl>
accuracy	0.8599089
sens	0.9617021

2 rows

Точность 86%, чувствительность 96% - хорошие показатели модели.

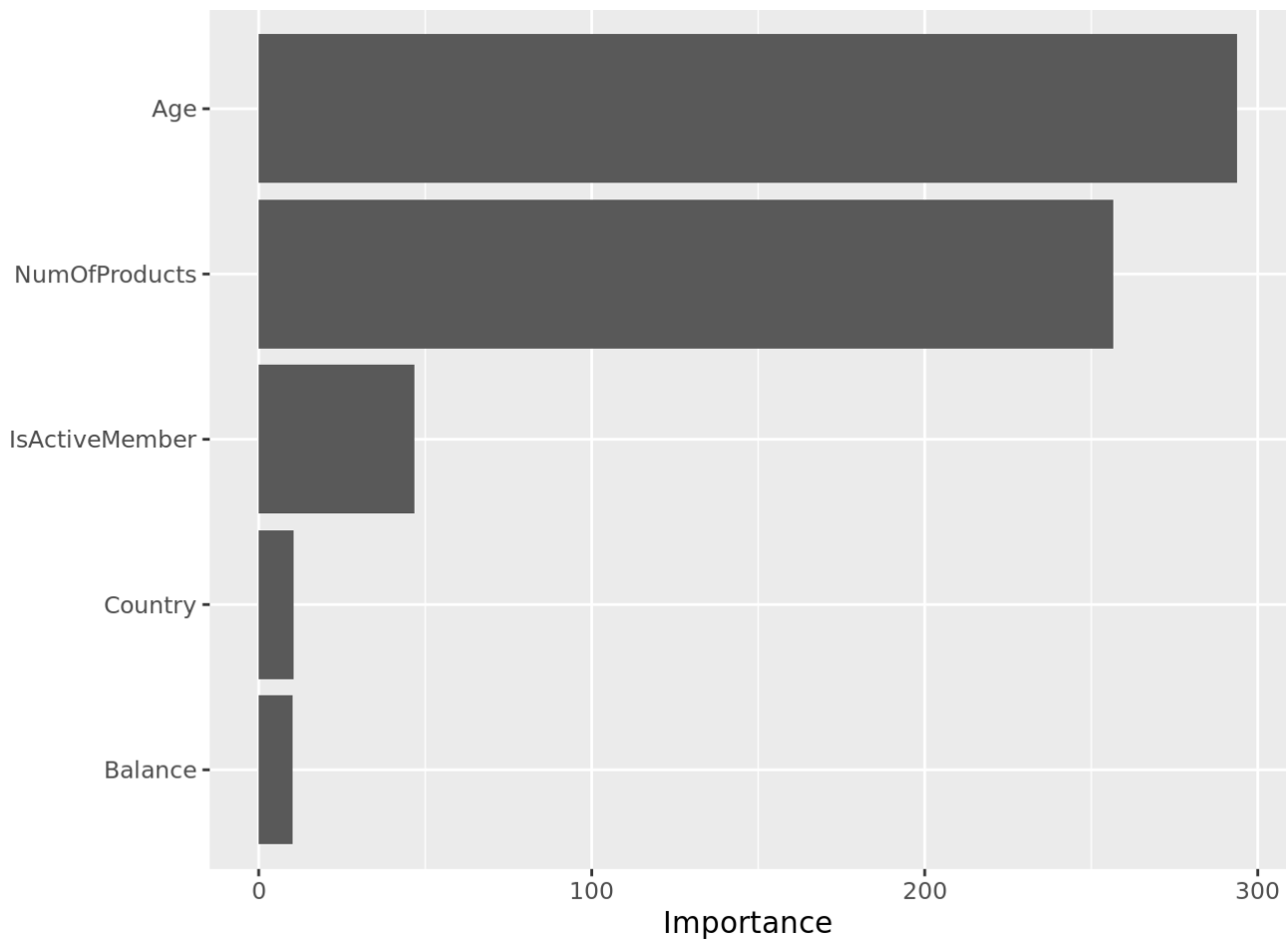
Code

.metric <chr>	.estimate <dbl>
accuracy	0.8541518
sens	0.9653874
2 rows	

Оценки похожи на метрики тестовых данных, что уже хорошо.

Оценим также важность признаков:

Code

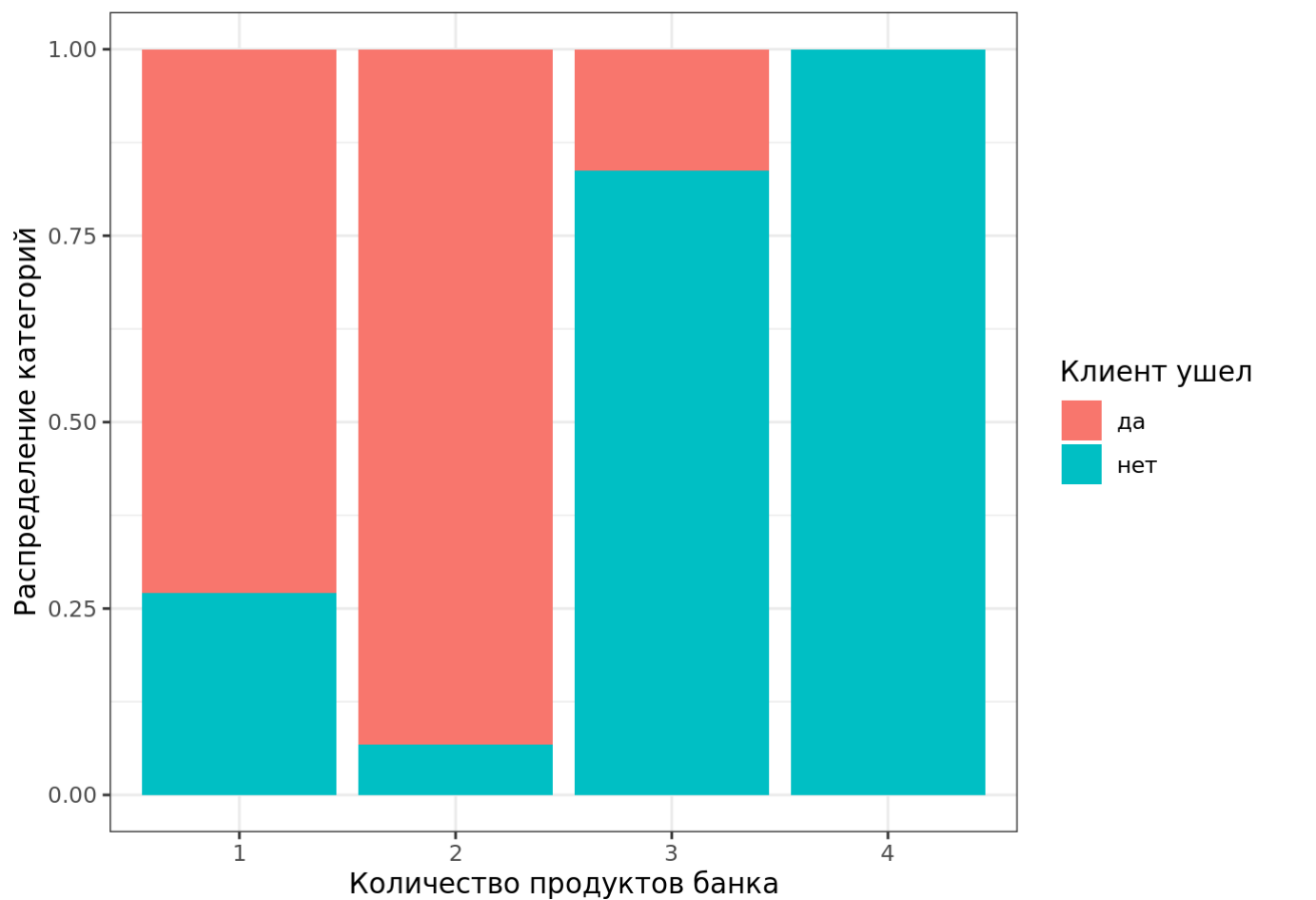


По графику видно, что наибольшее значение имеет возраст клиента и количество продуктов в банке, которыми он пользуется. По дереву также было видно, что изначально клиенты делятся на две категории: моложе 45 лет и старше 45. Причем для первой категории имеет значение только количество имеющихся в банке продуктов, а для второй - несколько разных признаков. Так как изменить возраст клиентов мы не в силах, а можем только привлечь новых, что является уже другой задачей, рассмотрим иные варианты снижения уровня оттока.

Посмотрим, где отток выше в распределении количества продуктов банка:

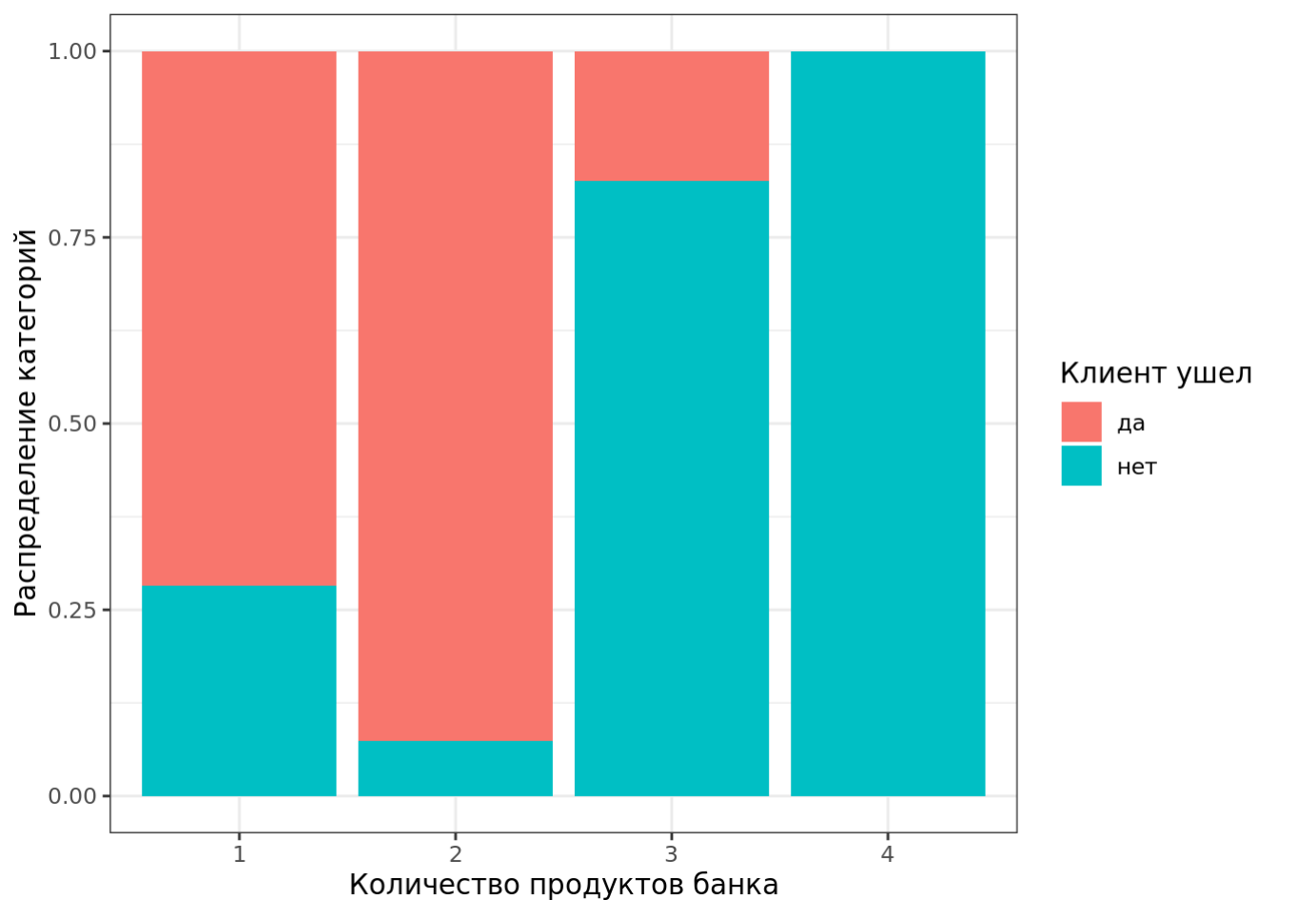
- для тестовой выборки

Code



- для тренировочной выборки

Code



В целом, разницы между тестовой и тренировочной выборкой особо нет. В обеих из них все клиенты, у кого есть 4 продукта, в 100% случаев ушли из банка. Также много ушедших среди тех, у кого уже есть 3 продукта.

Так как мы имеем данные только о кредитной карте как о продукте, проверим, связан ли показатель количества продуктов с наличием этой карты. Также добавим переменную активности, чтобы посмотреть, пользуется ли клиент данной картой (предположительно, если клиент совершает действия = активный, то и с кредитной картой тоже).

Code

NumOfProducts<fctr>	HasCrCard<fctr>	IsActiveMember<fctr>	Exited<fctr>	n<int>
3	0	0	0	4
3	0	0	1	31
3	0	1	0	8
3	0	1	1	23
3	1	0	0	13
3	1	0	1	89
3	1	1	0	16
3	1	1	1	55
8 rows				

Code

NumOfProducts<fctr>	HasCrCard<fctr>	IsActiveMember<fctr>	Exited<fctr>	n<int>
4	0	0	1	9
4	0	1	1	9
4	1	0	1	22
4	1	1	1	18
4 rows				

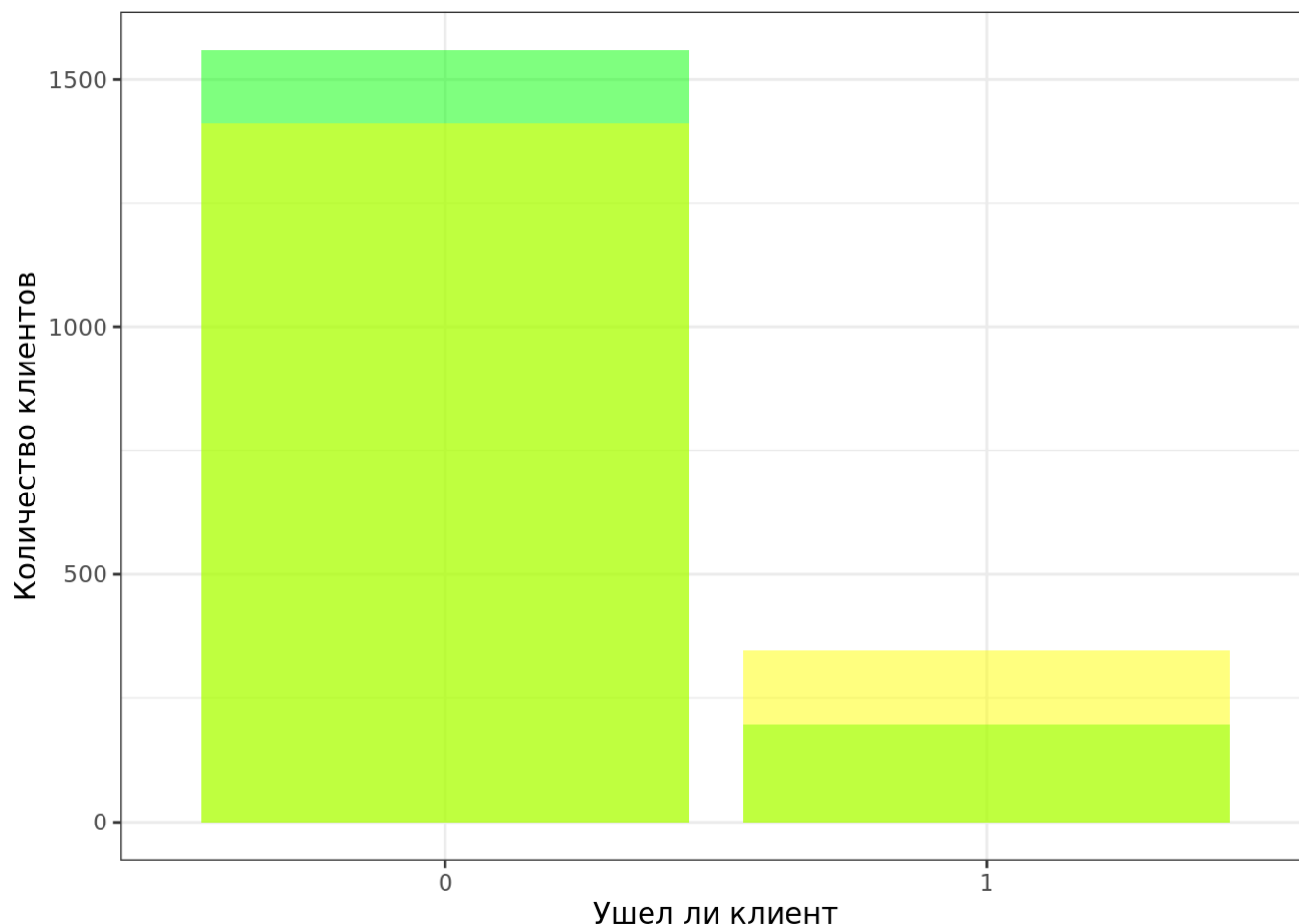
В обеих подборках большинство ушедших клиентов имеют кредитную карту и не пользуются ей. Поработаем с этими показателями.

Симуляция

Одним из решений снижения уровня оттока клиентов может стать мотивация начать пользоваться кредитной картой. Это можно сделать, например, увеличением льготного периода по операциям, бесплатным обслуживанием или увеличением кредитного лимита - на усмотрение банка. Это и привлечет клиентов обратно или же не даст существующим клиентам передумать пользоваться услугами банка (а именно кредитной картой).

Предположим, что 15% клиентов действительно заинтересовались новыми условиями и начали пользоваться кредитной картой, а значит стали активными пользователями.

Посмотрим, что изменится в количественном соотношении:



Желтые столбцы - показатели оттока в начальной тестовой выборке, а зеленые - наши ожидания от симуляции.

Таким образом, это решение может уменьшить отток клиентов, показатели улучшились. Банку рекомендуется подумать над условиями обслуживания кредитных карт, чтобы вернуть ушедших клиентов обратно или предотвратить дальнейший уход оставшихся клиентов.

```
[1] 11.22
```

Дэшборд

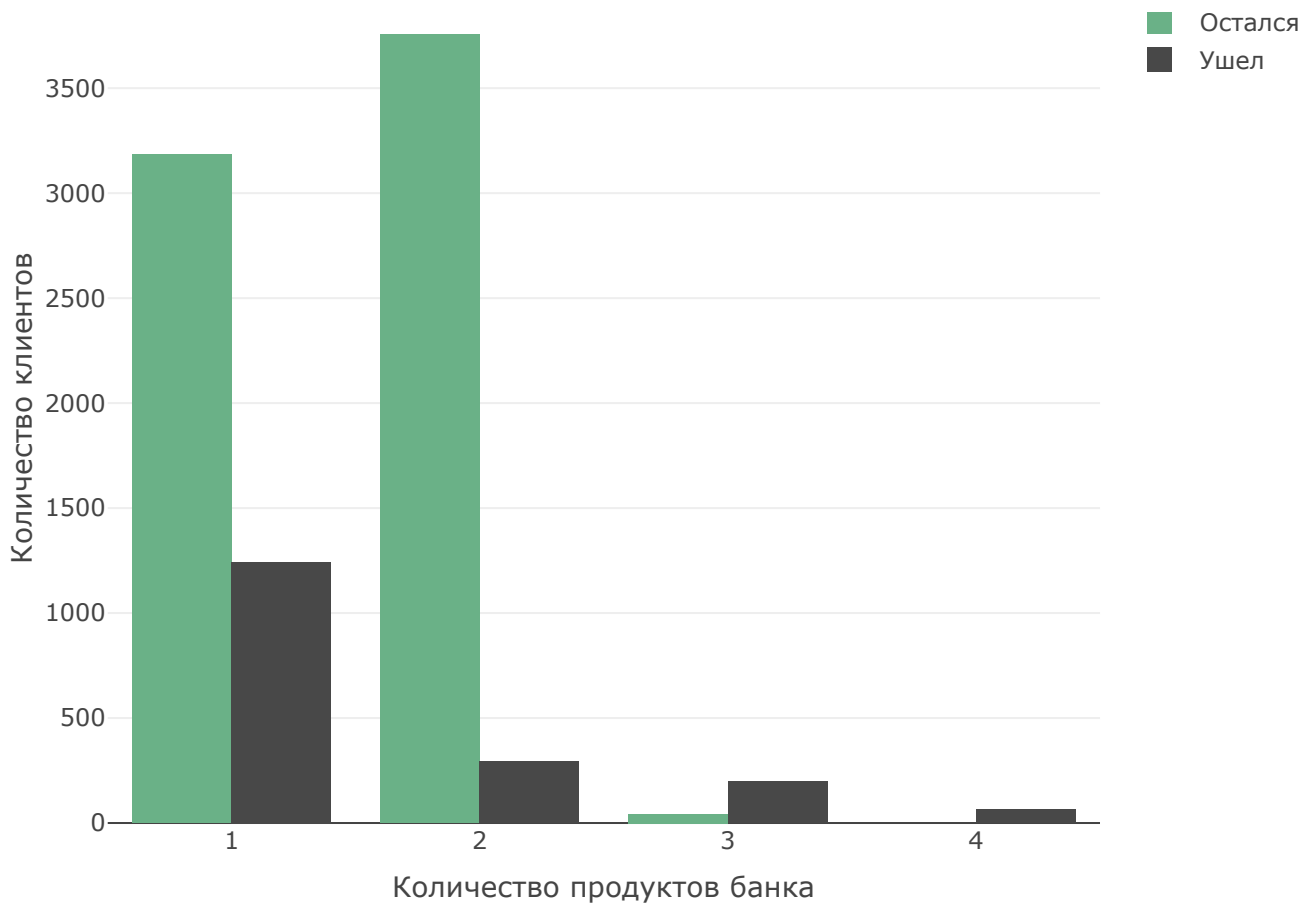
Наш дэшборд для заказчика. Он должен отобразить информацию по клиентам банка и по оттоку.

В “экспресс-данные” внесем:

- страну с наибольшим количеством клиентов (чтобы знать, кто основной потребитель)
- уровень оттока клиентов в выделенной подгруппе
- уровень оттока после предположений о решении проблемы оттока

Также будут отображены интерактивные графики:

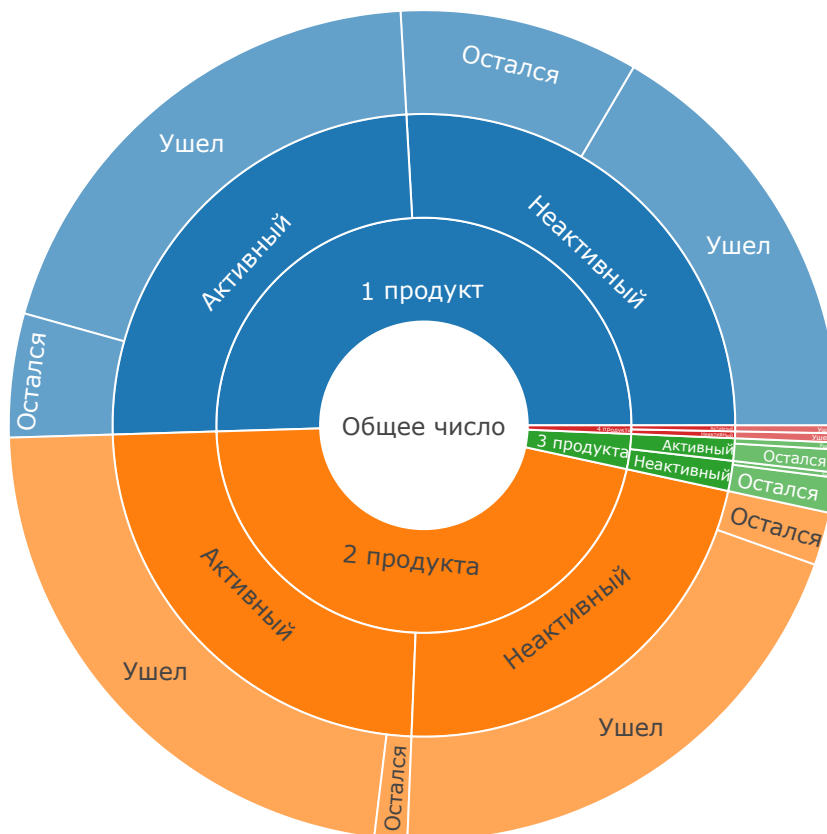
1. Анализ данных выделенной подгруппы клиентов, где видно сколько человек ушло или осталось при условии разделения на количество продуктов и с возможностью фильтра по странам.



2. Распределение оттока в зависимости от количества продуктов и активности клиента в интересующей нас подгруппе.

Рассматриваем данные подгруппы, так как с ними мы и работали при решении задачи.

Code



Общие выводы

В данном отчете мы проанализировали базу данных клиентов банка и постарались выяснить причины ухода клиентов. Мы рассмотрели подкатегорию граждан трех стран с заработной платой выше 1000, возрастом от 24 до 60 лет, которые сотрудничают с банком по крайней мере месяц.

Проведя анализ оттока мы выяснили, что важными показателями является возраст клиента и количество имеющихся у него продуктов в банке. Также проверили дополнительные факторы влияния и увидели, что многие клиенты, имеющие кредитную карту и не пользующиеся ей, ушли от банка. Решением для сокращения оттока таких клиентов было предложение по мотивации клиента использовать кредитную карту, изменив условия обслуживания, которые являются наиболее привлекательными, какими именно - решение за заказчиком. Однако мы проверили теорию и отток клиентов сократился почти в два раза с применением такой стратегии.