



# Сегментация пользователей: микс методологий и эффективность



# Дмитрий Кротов

## Аналитик в Транспорте

- Аналитик направления маркетинга в Авито Авто
- Занимаюсь развитием стрима коммуникаций
- Преподаю в ААА и менторю в Авито Компас





# АВИТО АВТО

Большой продукт с многомиллионной аудиторией

## Модель доски объявлений (classified)

дилеры, перепродавцы и частные продавцы продают авто

## Транзакционные продукты

покупка и продажа авто прямо на Авито через проверенных дилеров и профессиональных продавцов и аренда



# План

- 1 Предпосылки
- 2 Разработка методологии
- 3 Оценка качества модели
- 4 Оценка эффективности модели в CRM-коммуникациях
- 5 Заключение

# Предпосылки

1

# Предпосылки и требования

- 01** **Рост аудитории и продукта** требует инструментов определения состояния юзеров (начал поиск, активный поиск, зашел посмотреть)
- 02** Ключевым заказчиком выступает команда коммуникаций – для них **важно персонализировать рассылки**
- 03** Созданная методология должна иметь **возможность адаптироваться** под разные категории продукта с небольшими затратами
- 04** **Без ML**, так как предыдущие попытки не дали достаточной точности



# Методология

2

# Суть RFM-методологии

**RFM** - метод сегментации, разделяющий клиентов на непересекающиеся между собой группы и позволяющий оценить их активность и ее давность

# R



**Recency**  
давность

# F

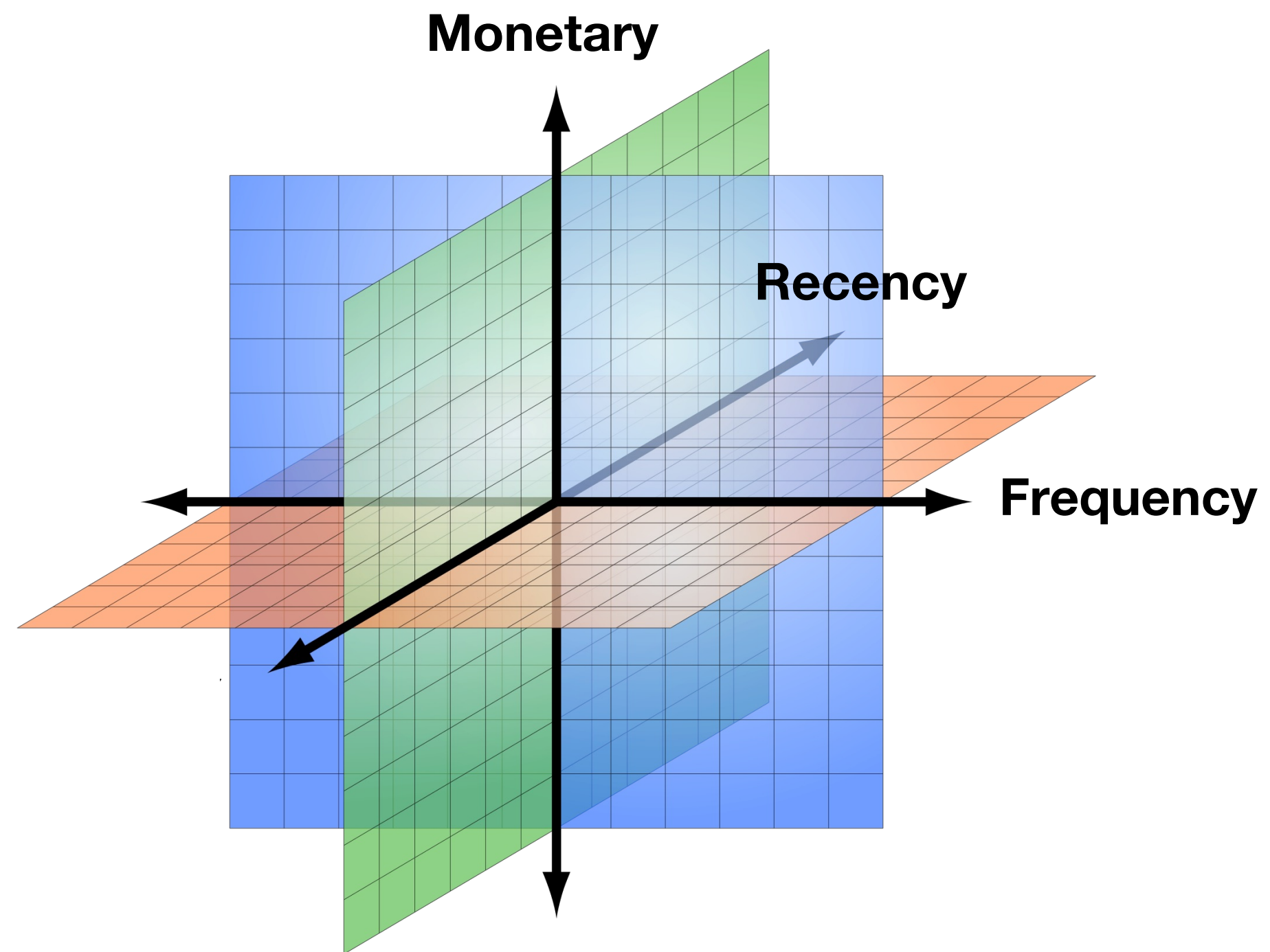


**Frequency**  
частота

# M



**Monetary**  
деньги





# Выбор основы методологии

## Классическая RFM-модель

не подходит из-за того что мы зарабатываем не на покупателях

## CLV-анализ и скоринговая модель

не отвечают на вопрос “горячести” пользователя именно сейчас

## Не классические RFM-модели

- Recency-Frequency-Intensity (RFI)
- Recency-Engagement-Value (REV)
- Recency-Frequency-Monetary-Channel (RFMC)
- RFM без оценки монетаризации
- Recency-Activity-Quantity (RAQ)



# Классическая RFM

- R recency  
давность
- F frequency  
частота
- M monetary  
деньги



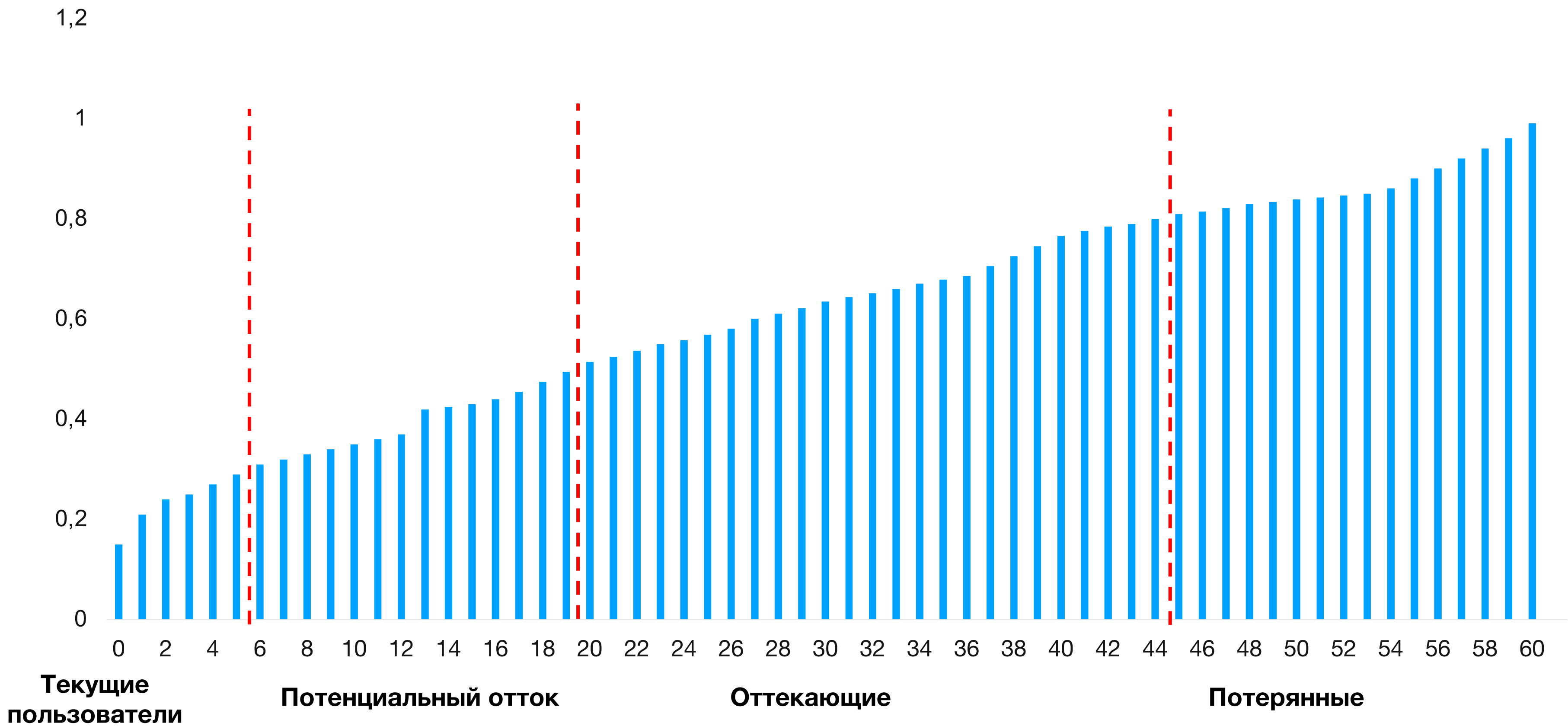
# Адаптированная RFM

- R recency  
давность последнего из  
событий
- F frequency  
частота – интенсивность  
событий в формате score
- F<sub>LP</sub> frequency last period  
доля активности за последнюю  
период от всей активности

# Recency

Оцениваем **давность** посещения нашей категории

Границы каждого сегмента выбираются на основе **размеров выборки** и **продуктового подхода** к понимаю оттока





# Frequency

Расчет параметра:

$$F = w_1 * q_1 + w_2 * q_2 + \dots + w_n * q_n$$

$w_i$  - weight – вес каждого типа события

$q_i$  - quantity – количество каждого типа события у юзера

$n$  – кол-во параметров

Расчет весов:

$$w_n = \frac{q_n}{q_1 + q_2 + \dots + q_n}$$

# Frequency

Пример:

 Просмотры объявлений - 1 200 000 шт

 Поделиться – 350 000 шт

 Добавления в корзину – 50 000 шт

Расчет весов:

$$W_{\text{view}} = \frac{1\,200\,000}{1\,200\,000 + 350\,000 + 50\,000} = 0,75$$

$$W_{\text{share}} = \frac{350\,000}{1\,200\,000 + 350\,000 + 50\,000} = 0,22$$

$$W_{\text{cart}} = \frac{50\,000}{1\,200\,000 + 350\,000 + 50\,000} = 0,03$$

Если взять iv за единицу, то вес параметров получается следующим: поделиться – 3.4 (0.75 / 0.22), корзина – 25 (0.75 / 0.03)

Итоговая формула из нашего примера:

$$F = 1 * q_1 + 3,4 * q_2 + 25 * q_3$$

# Frequency

- Стоит рассматривать как базовое событие в воронке, так и максимально близкое к целевому
- Score должен учитывать наиболее отражающие специфику вашего продукта метрики
- Веса должны быть не «хардами», а автоматически пересчитываемыми
- Границы сегментов должны быть связаны с весами для понятной интерпретации





# Frequency last period

Recency не дает ответа на вопрос о степени активности пользователя в последнее время (возможны кейсы случайного посещения старых гиперактивных пользователей)

## User 1

**Recency** = 3 дня назад

**Frequency** = 500 (горячий)

**Вывод:** свежий и активный

**НО**

**3 дня назад:** 3 контакта, 2 сравнения  
и 5 избранное, 30 просмотров – **100% активности**

## User 2

**Recency** = 3 дня назад

**Frequency** = 500 (горячий)

**Вывод:** свежий и активный

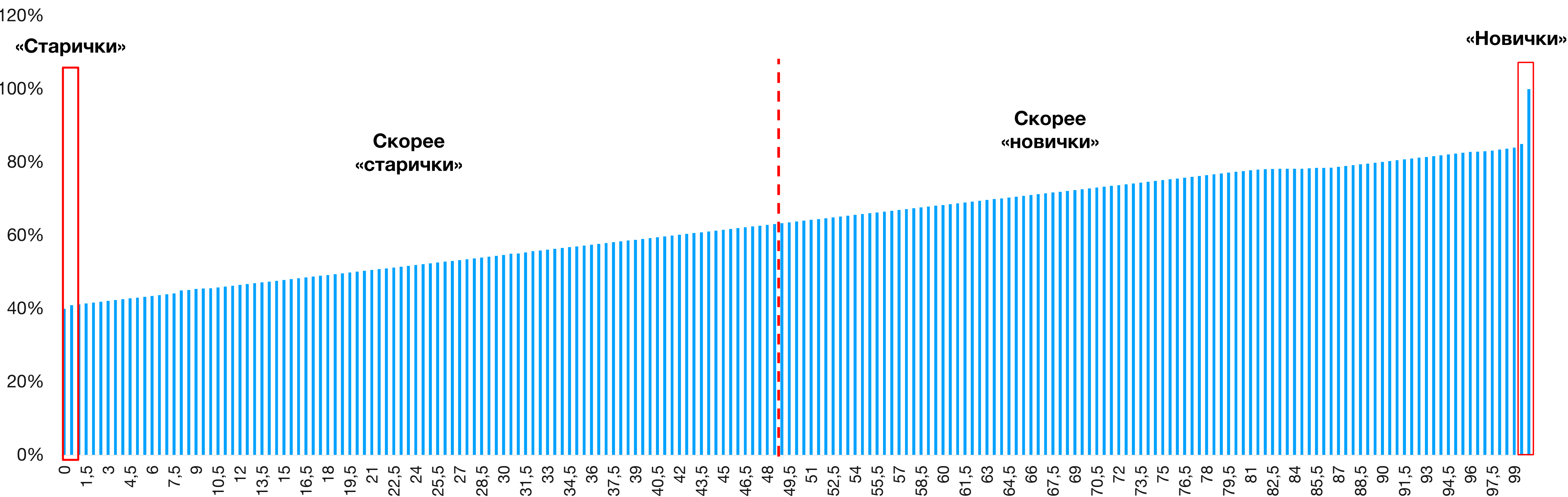
**НО**

**3 дня назад:** 2 просмотра  
– **20% активности**

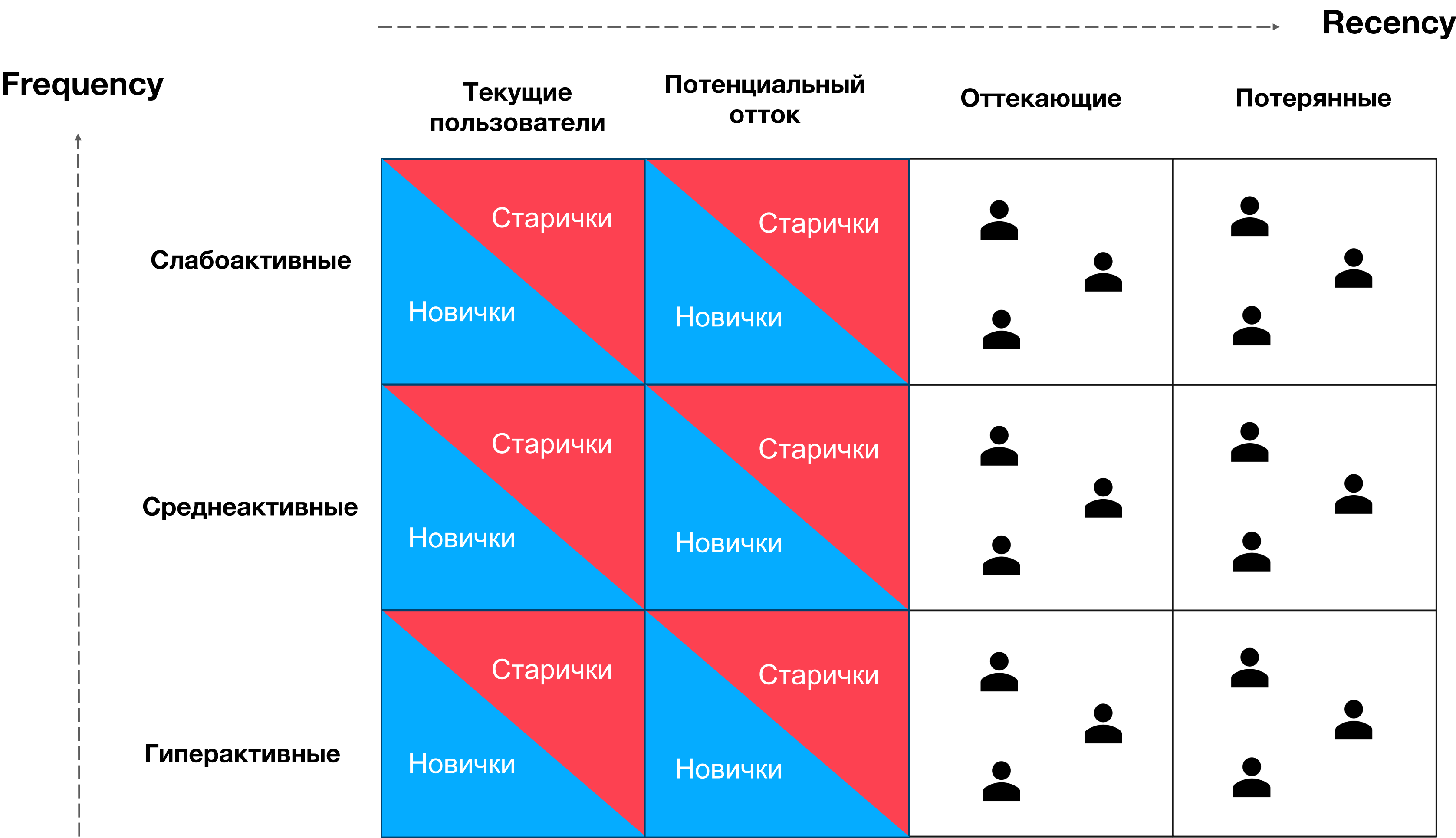
# Frequency last period

$F_{LP} = \text{Frequency за период} / \text{Frequency total}$

Ширина предыдущего периода зависит от бизнеса и продукта



# Как это работает?



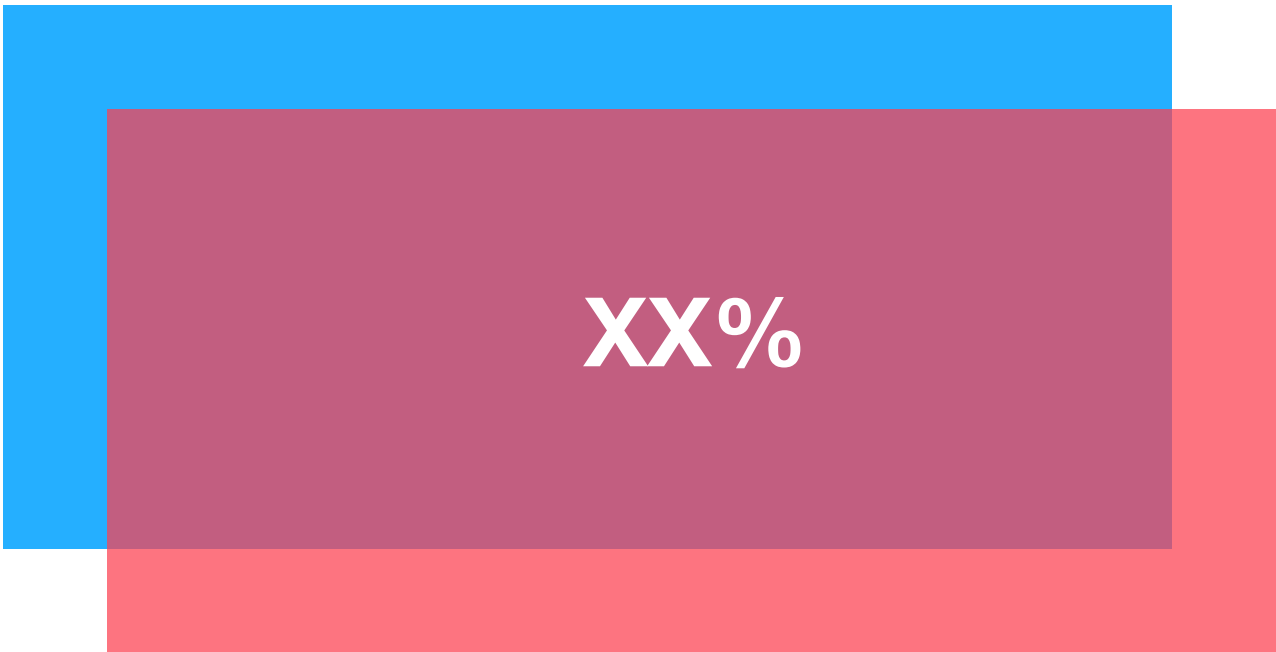


# Оценка качества модели

3

# Как оценить точность?

Наша оценка  
пользователей



Реальная картина

Пример оценки точности на пользователях «горячего» сегмента



# Оценка модели

## **Подход №1: метрики в день скоринга (краткосрочное определение и “горячие юзеры”)**

Покрытие целевой когорты юзеров RF моделью на 9 п.п. выше проб с ML в Авто и почти совпадает с точностью ML моделей в других категориях

## **Подход №2: метрики в течение 30 дней после скоринга - (долгосрочное определение – “холодные юзеры”)**

Покрытие целевой когорты юзеров RF моделью на 11 п.п. выше проб с ML в Авто и точно совпадает с точностью ML моделей в других категориях



# Эффективность модели в CRM-коммуникациях

4

# Использование сегментации в CRM

Сегментация используется для корректировки сегментов коммуникации по параметрам:

- степень активности
- степень новизны активности
- степени оттока юзера

**Оценка эффективности сегментации проводится на основе запускаемых АБ-тестов CRM**

**Общий подход к тестированию:**

отправляем персонализированную коммуникацию под сегмент против стандартной коммуникации на безликую аудиторию



# Результаты

## Персонализированные коммуникации, основанные на сегментации:

- 🔥 Показывают **более высокий рост open rate** в среднем **на 25%**
- 🔥 Приводят в **1,5 раза больше целевых баеров**
- 🔥 Показывают **апфит в просмотрах выше** в среднем **на 27%**
- 🔥 Имеют **апфит в целевых кликах выше** в среднем **на 24%**
- 🔥 **Реже триггерят** пользователей к отпискам от пушей






Кроме того, сегментация позволила детектировать оттоки юзеров в другие категории => возможность кросс-рассылок



# Заключение



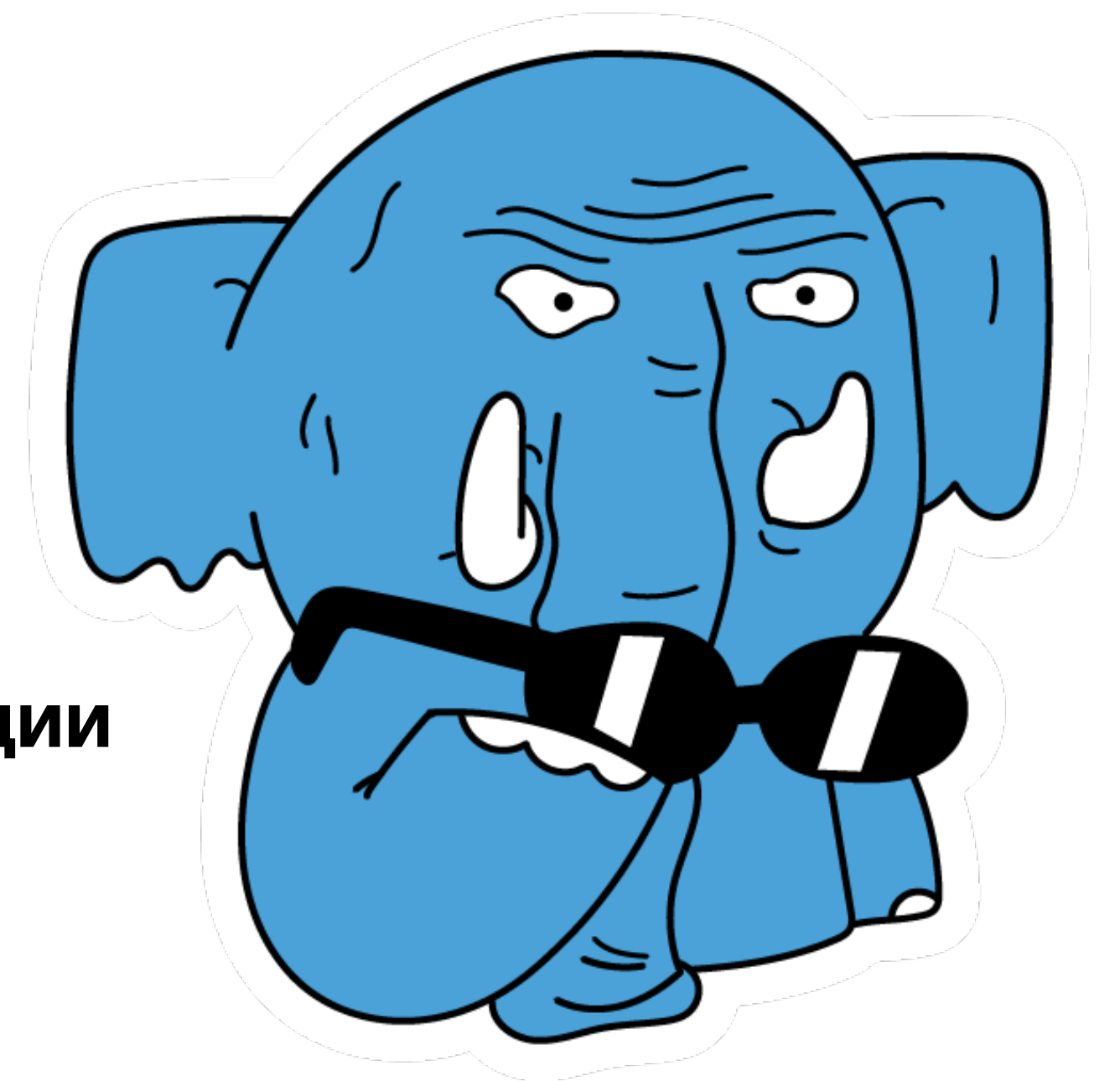
# Итог проделанной работы

-  **Собрали потребности** от бизнеса и продукта
-  **Изучили** имеющиеся методологии
-  **Замиксовали** методологии и **адаптировали** имеющиеся практики – отличный способ сделать эффективный инструмент для конкретной ситуации
-  **Получили эффективный инструмент** без ML
-  **Нарастили метрики** коммуникаций за счет использования сегментации, стали приводить больше целевых баеров и меньше раздражать пользователей нецелевыми пушами



# Оффтоп

- При разработке сегментации учитывайте также **частоту обновления** и **период расчета** (зависят от скорости и частоты взаимодействия с пользователем)
- Не ограничивайте себя имеющимися методологиями, **миксуйте и тестируйте**
- Проводите **онбординг** в сегментацию – в основы деления по сегментам, метрики по ним и не забудьте про **документацию**
- Собирайте **историю** сегментации и создавайте инструменты **визуализации** метрик сегментов



# Дмитрий Кротов

Analyst



@dman\_kr