

Прогнозирование Оттока Клиентов

Chocolife .me



СОДЕРЖАНИЕ

1. В чем Проблема?
2. Формулировка задачи
3. Сбор данных
4. Очистка данных
5. Создание новых признаков
6. Сравнение моделей
7. Наш Результат
8. Предложения по сбору данных

СОДЕРЖАНИЕ



В чем Проблема?

- 2. Формулировка задачи**
- 3. Сбор данных**
- 4. Очистка данных**
- 5. Создание новых признаков**
- 6. Сравнение моделей**
- 7. Наш Результат**
- 8. Предложения по сбору данных**

В Чем Проблема?



А зачем нам Анализ?

А зачем нам уменьшать Отток?

1. Новые Клиенты - это **дорого**.
2. Необходимо защитить свой **доход**.
3. Второй раз клиенты **тратят больше**.
4. Клиенты уйдут, если они думают, что **вам все равно**.
5. Лояльные клиенты могут стать **энтузиастами бренда**.

СОДЕРЖАНИЕ

1. В чем Проблема?



2. Формулировка задачи

3. Сбор данных

4. Очистка данных

5. Создание новых признаков

6. Сравнение моделей

7. Наш Результат

8. Предложения по сбору данных

Формулировка задачи

- Была поставлена задача прогнозирования оттока клиентов из ChosoLife.
- Что такое отток в понимании ChosoFamily?
Пользователь не покупает ни одну продукцию в течение одного года.

Определение Целевой Переменной

- Мы решили прогнозировать отток клиентов которые были активными в промежутке одного года (например 2015).
- Отток определялся на основании его активности в следующем году (2016). Ну а активность это - покупка минимум одной услуги через ChocoLife.

ΟΠΤΟΚ

	2014	2015	2016
ΟΠΤΟΚ	45 %	40,5 %	43 %



СОДЕРЖАНИЕ

1. В чем Проблема?

2. Формулировка задачи



Сбор данных

4. Очистка данных

5. Создание новых признаков

6. Сравнение моделей

7. Наш Результат

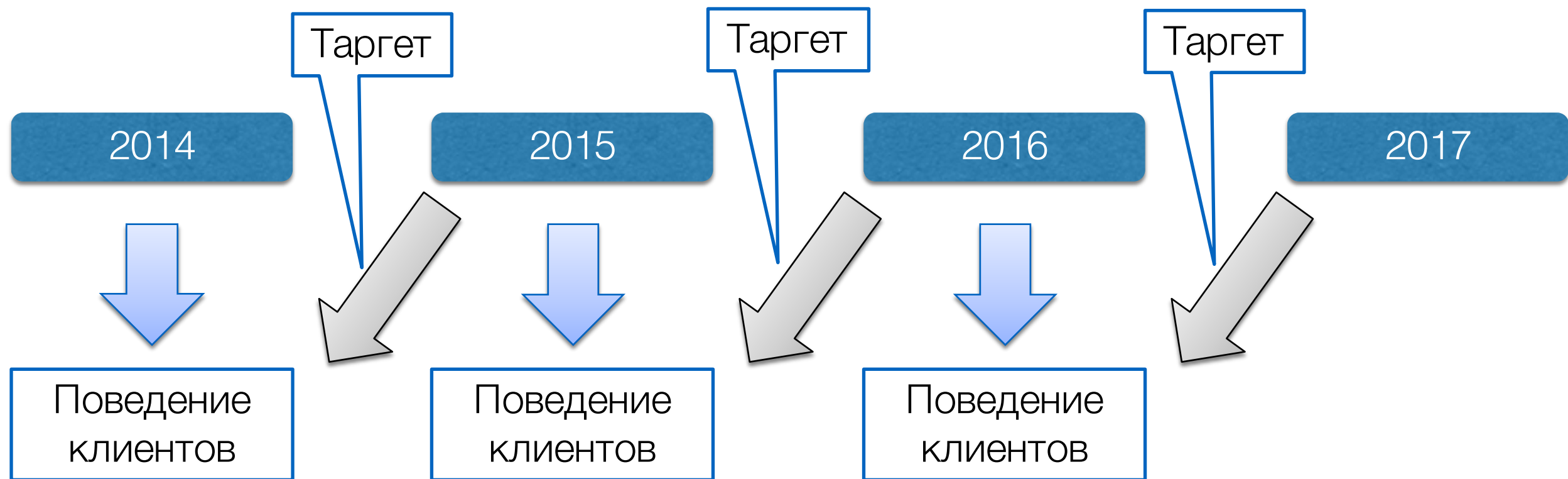
8. Предложения по сбору данных

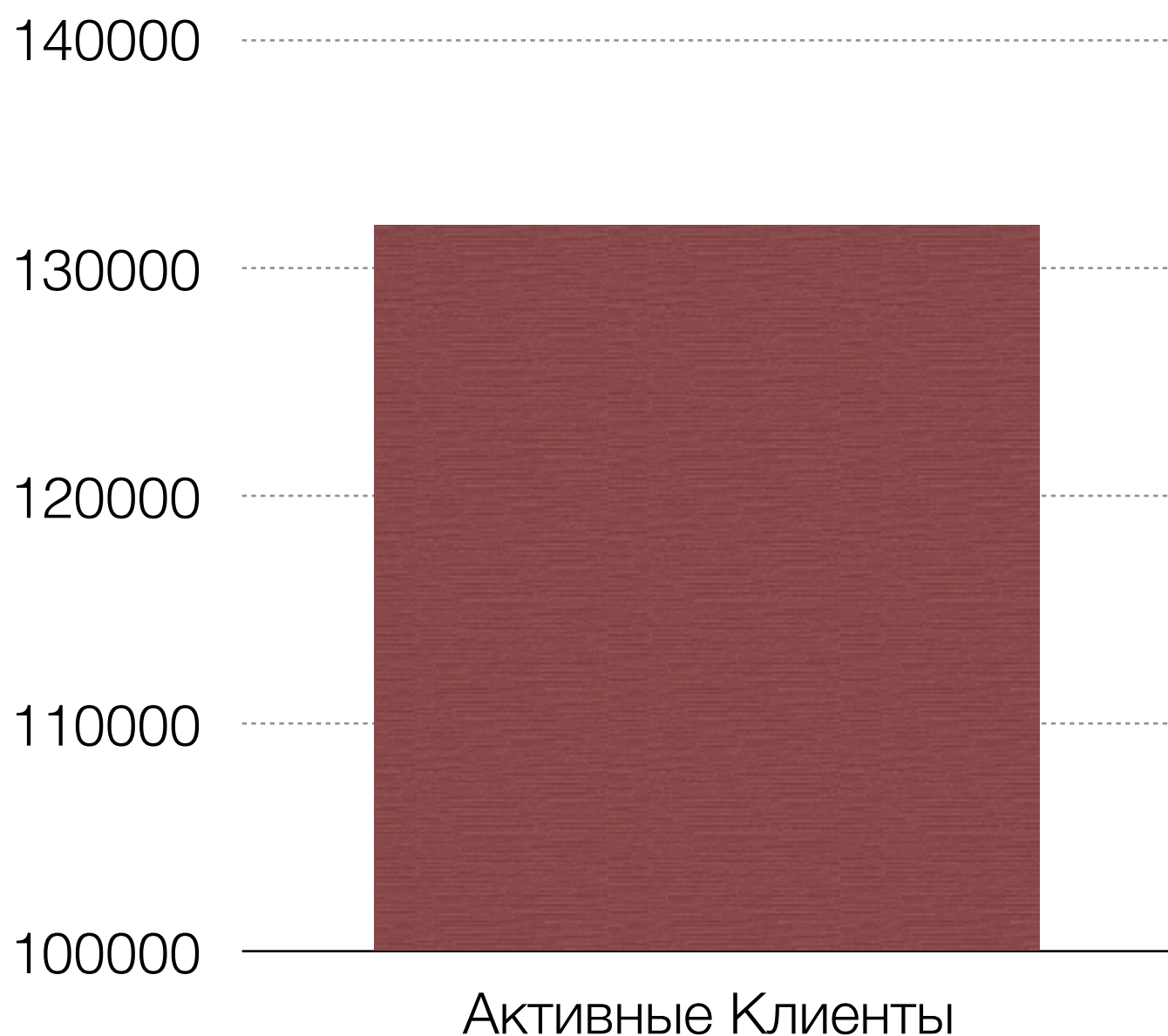
Сбор Данных

Размерность: Данные по продажам и акциям были расположены в 20 файлах.

Время: Данные за 4 года (2014-2017).
Тест: 2016. Обучение: 2014-2015.

Количество клиентов: 1,048,575 клиентов, кто зарегистрировался с 2011 до 2016.

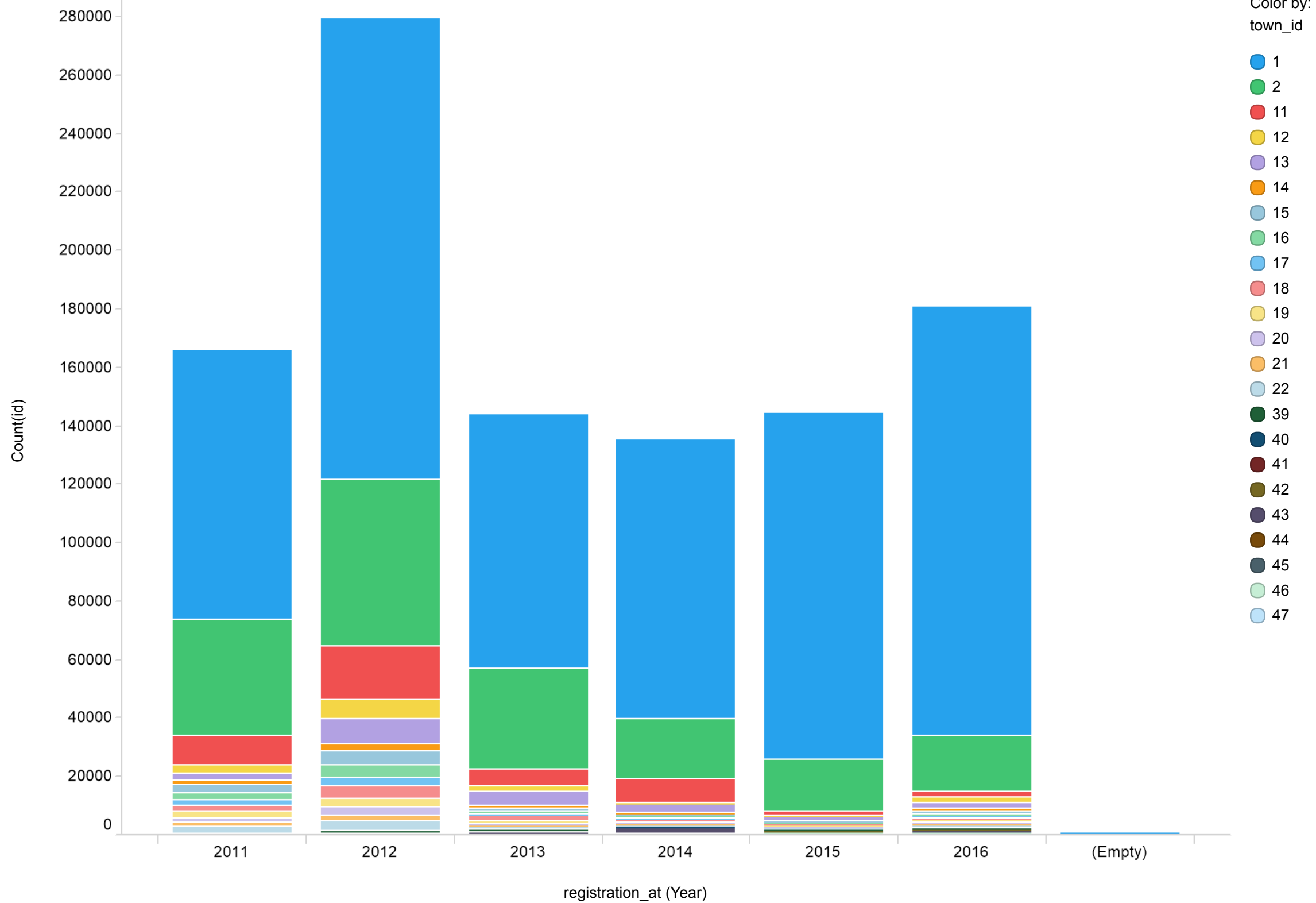




2016



Динамика Количества Регистрации Клиентов



СОДЕРЖАНИЕ

1. В чем Проблема?

2. Формулировка задачи

3. Сбор данных



Очистка данных

5. Создание новых признаков

6. Сравнение моделей

7. Наш Результат


8. Предложения по сбору данных

Очистка данных

В поведенческих данных клиентов за один год были убраны клиенты зарегистрировавшиеся в течение анализируемого года.

По многим признакам клиентов были неполные данные и поэтому мы не смогли использовать их.

СОДЕРЖАНИЕ

1. В чем Проблема?
2. Формулировка задачи
3. Сбор данных
4. Очистка данных
5.  Создание новых признаков
6. Сравнение моделей
7. Наш Результат
8. Предложения по сбору данных

Сбор признаков

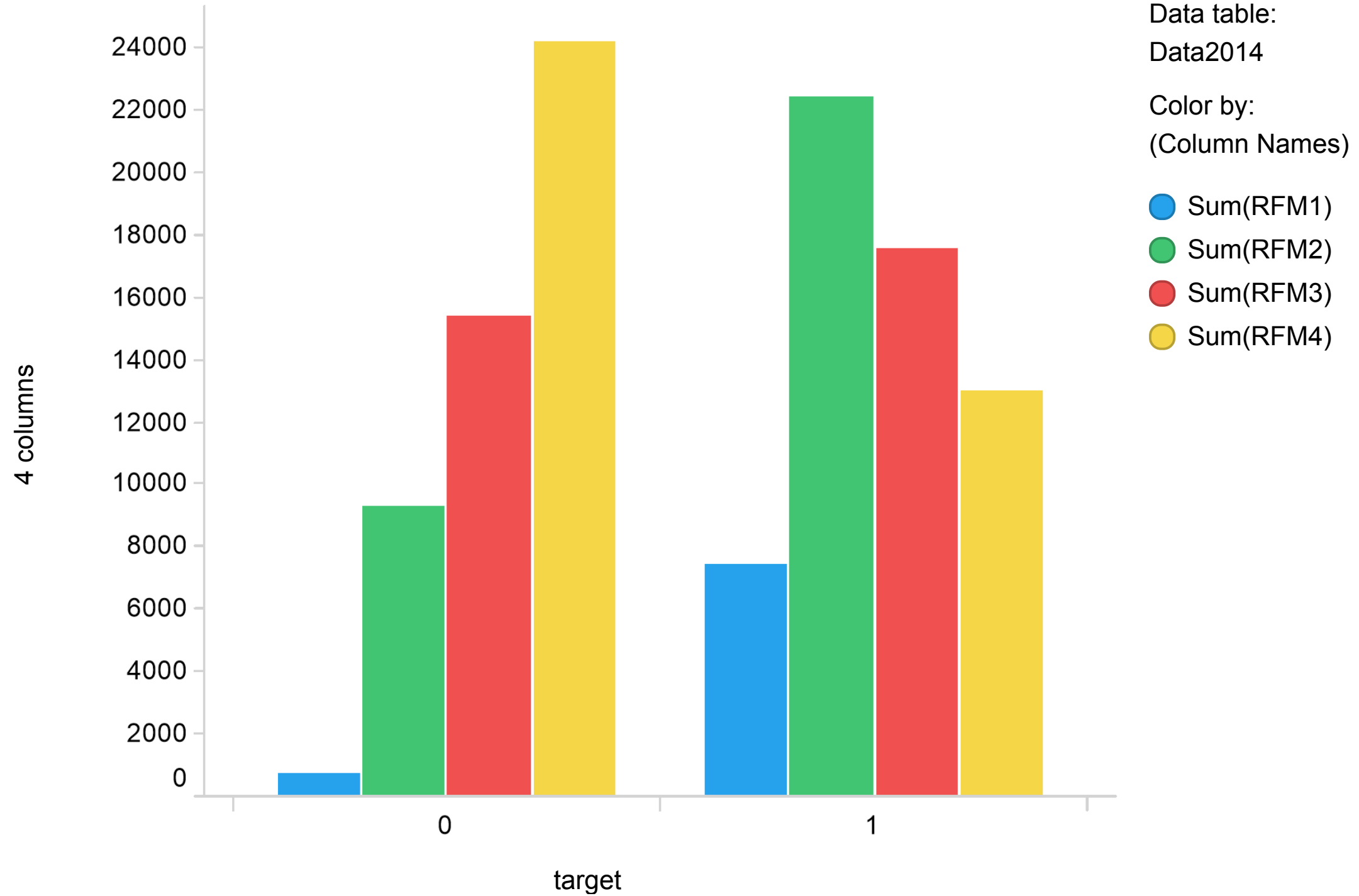
Признаки собирались на каждого клиента за каждый год.

- Количество ваучеров купленных за год.
- Количество уникальных акций за год.
- Количество заказов.
- Средняя цена ваучера.
- Количество уникальных дней покупок.
- Средняя скидка по акциям.
- Количество уникальных категорий акций.

Создание Новых Признаков

- Была проделана RFM (Recency-Frequency-Monetary value) сегментация клиентов, которая разделила их на четыре сегмента по этим трем параметрам.
- Была найдена корреляция между уровнем оттока и сегментом клиентов.

RFM1, RFM2, RFM3, RFM4 per target

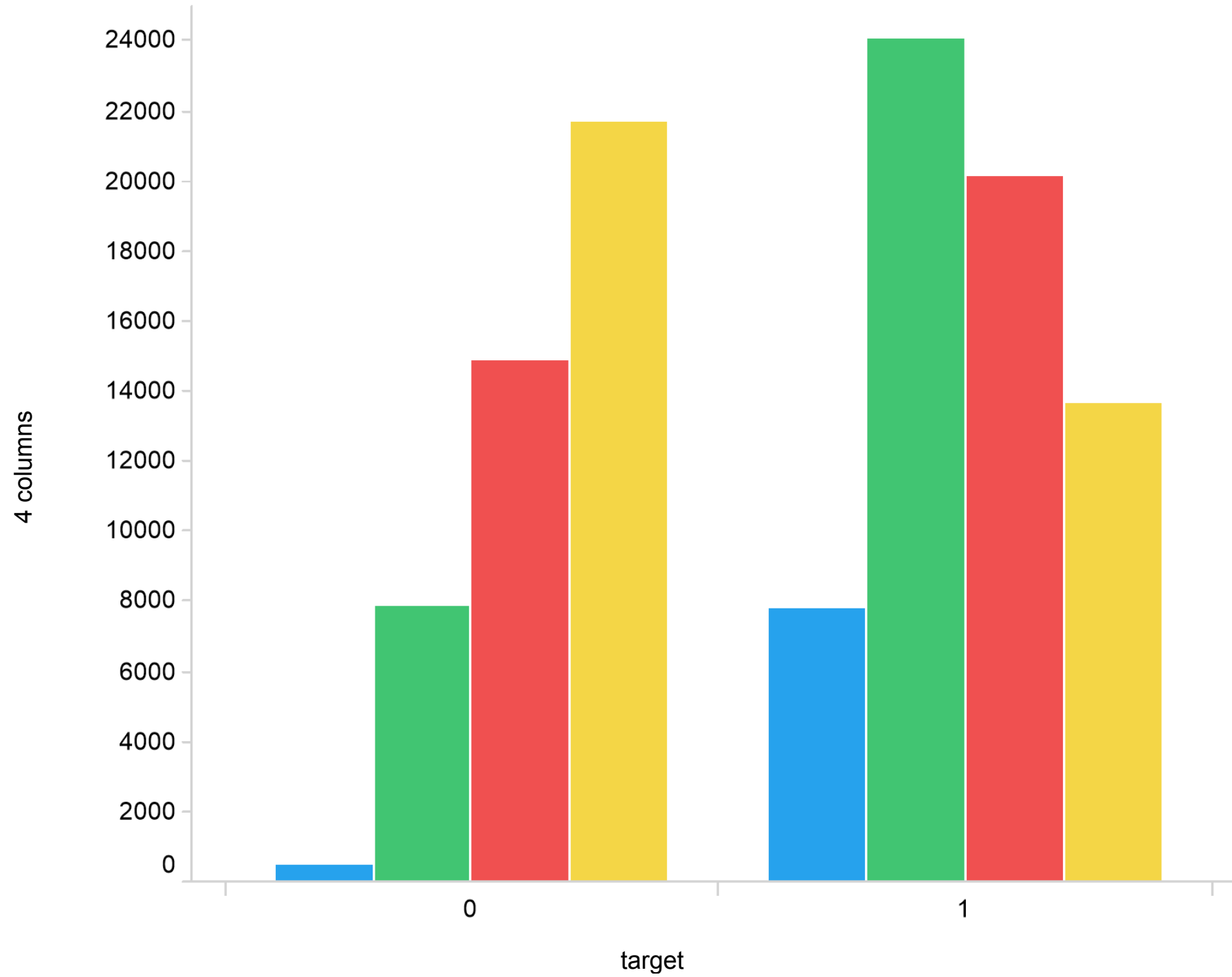


RFM1, RFM2, RFM3, RFM4 per target


Data table:
Data2015

Color by:
(Column Names)

- Sum(RFM1)
- Sum(RFM2)
- Sum(RFM3)
- Sum(RFM4)




СОДЕРЖАНИЕ

1. В чем Проблема?
2. Формулировка задачи
3. Сбор данных
4. Очистка данных
5. Создание новых признаков
-  6. Сравнение моделей
7. Наш Результат
8. Предложения по сбору данных

Сравнение моделей

- 1) Regression analysis: logistic regression.
- 2) Decision tree Random Forest.
- 3) Bayes algorithm: Naive Bayesian.
- 4) Ensemble learning: Ada Boost, Stochastic Gradient Boost and Random Forest.
- 5) Artificial neural network: Multi-layer Perceptron.

СОДЕРЖАНИЕ

1. В чем Проблема?
2. Формулировка задачи
3. Сбор данных
4. Очистка данных
5. Создание новых признаков
6. Сравнение моделей
-  7. Наш Результат
8. Предложения по сбору данных

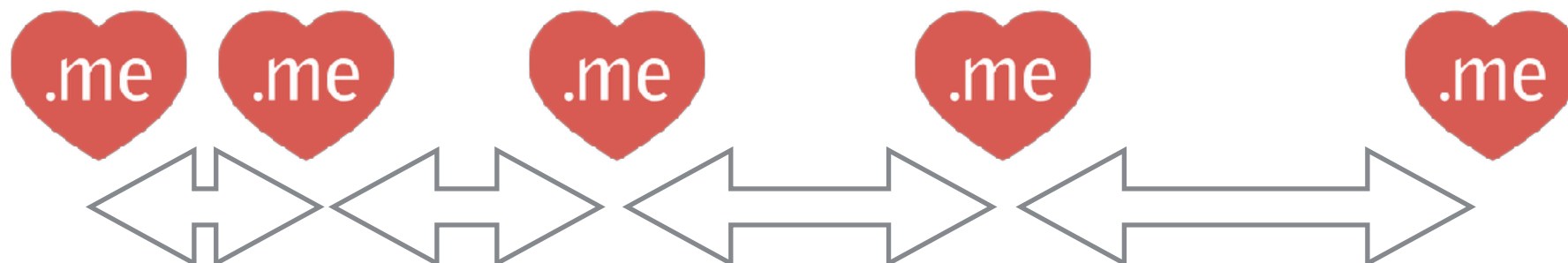
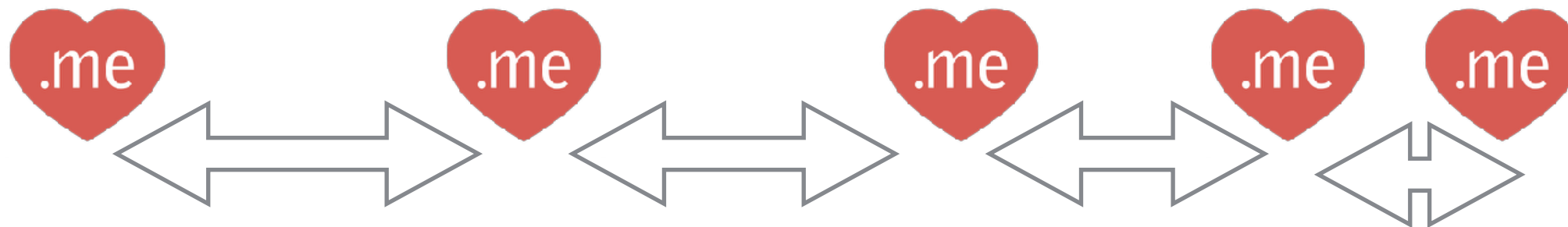
Наш результат

- Сколько бы мы разных моделей ни применяли ни одна из них не могла хорошо спрогнозировать отток. Точность прогнозирования, Precision, Recall, F1 Score варьировали между 66-68%.
- Модель не может показать того чего нет в данных.

	precision	recall	f1-score	support
0	0.61	0.56	0.58	9062
1	0.71	0.75	0.73	13002
avg / total	0.67	0.67	0.67	22064

Нет времени, но все же...

- Отдельные модели для отдельных сегментов.
- Поиск отличительных признаков в оттоке лояльных клиентов.



СОДЕРЖАНИЕ

1. В чем Проблема?
2. Формулировка задачи
3. Сбор данных
4. Очистка данных
5. Создание новых признаков
6. Сравнение моделей
7. Наш Результат



Предложения по сбору данных

Предложения по сбору данных

- Отсутствуют данные “Interaction”, в которые входят: откуда он зашел на акцию, связь с Call-center’ом, комментарии и посты.
- Одно из важных типов данных - “Behavior”. Туда относятся такие как: *Предпочтения, мнение, потребность и социальное поведение(комментарии, баллы, частые покупки и тд.)*.
- *Для определения горизонта прогнозирования, нужны такие данные как: сколько времени провел клиент, что искал, нашел ли, клики и тд.*



Всем Рахмет!