

Клиентская аналитика за минуты, а не дни

Николай Паклин, Loginom Company

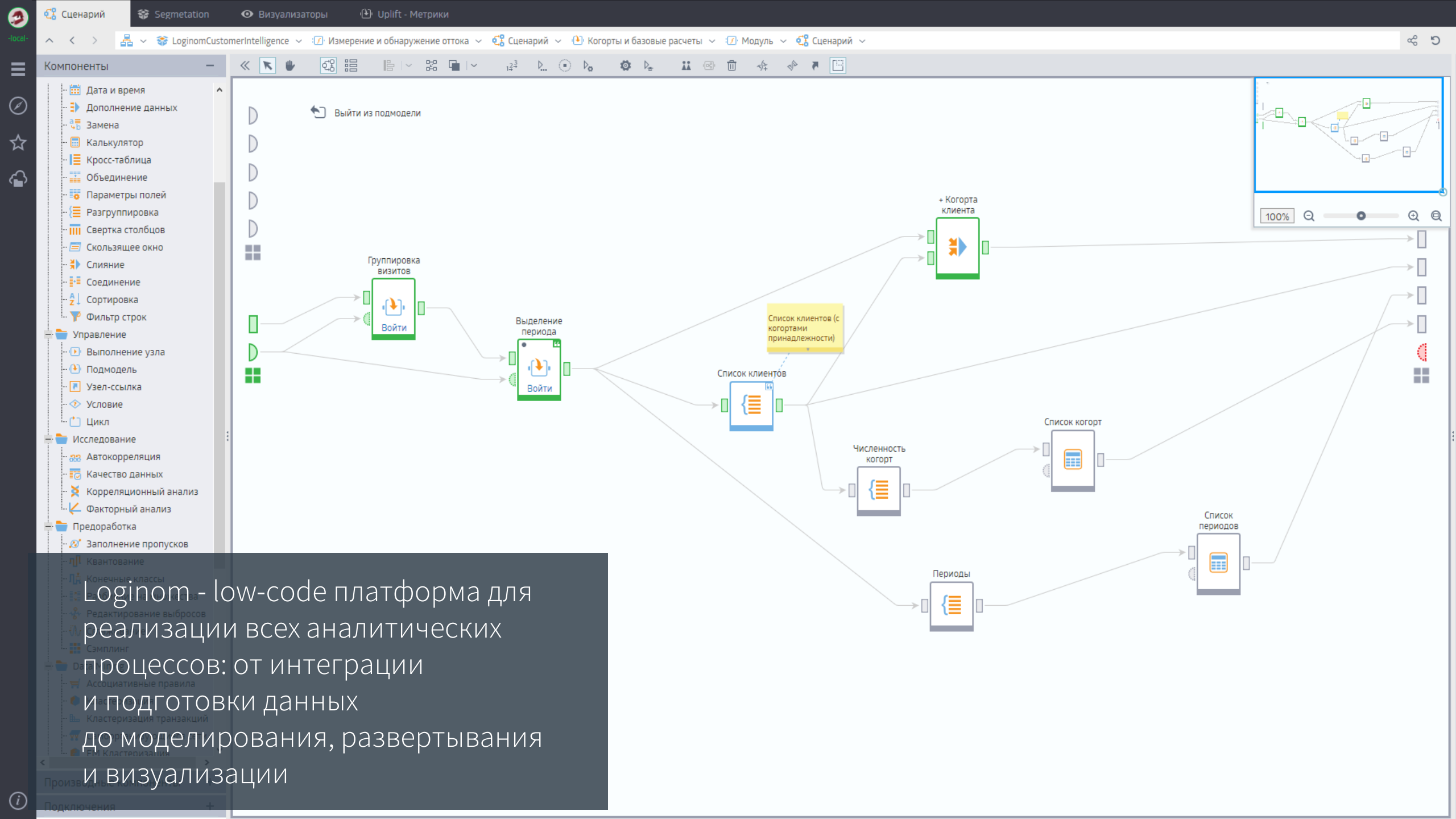
О спикере

- Руководжу направлением [Loginom e-Learning](#)
- Автор [книг](#) и электронных курсов по аналитике данных
- Бизнес-тренер компании Loginom
- Куратор [академической программы](#) Loginom
- Продюсер [Loginom Хакатон](#)
- Имею экспертизу в области клиентской аналитики, кредитного скоринга



План вебинара

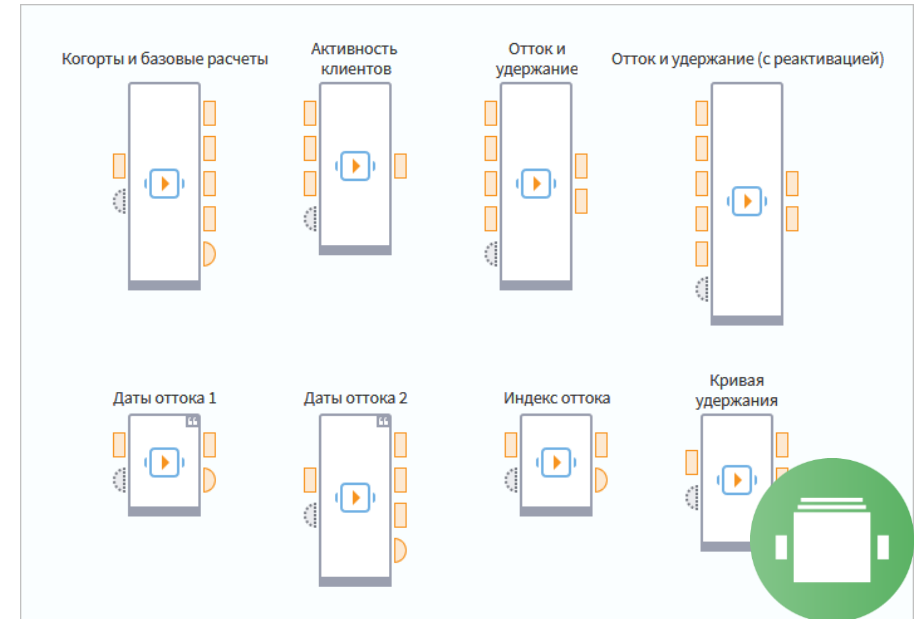
- Что такое библиотеки компонентов?
- Кейсы (ритейл и розничный банк)
 - Обнаружение оттока
 - Сегментация
 - Формирование профиля клиента с кейсами его использования
- Бизнес-курсы



Loginom - low-code платформа для реализации всех аналитических процессов: от интеграции и подготовки данных до моделирования, развертывания и визуализации

Библиотека компонентов

- Компонент библиотеки – подмодель с реализованной логикой обработки данных
- Библиотека компонентов – набор логически объединенных компонентов из одной области
- Пакет зашифрован



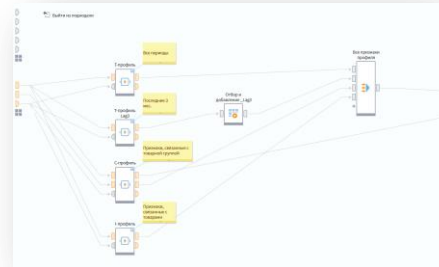
Комплект поставки



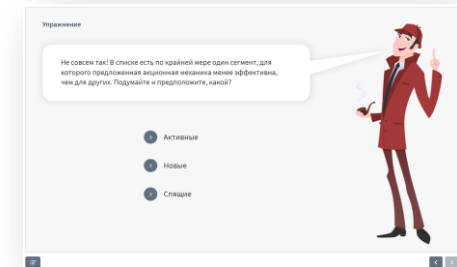
Пакет Loginom с компонентами



Документация



Кейсы использования



Бизнес-курс

Для чего нужны библиотеки компонентов?



- Произвольная логика обработки
- Максимальная гибкость
- Лучшие практики
- Гибкость остается
- Быстрый старт
- Все готово
- Быстрый старт

Self-made в Loginom

Библиотека компонентов

Аналитический сервис,
прикладное решение



- Затраты времени и ресурсов
- Требуются хорошее знание платформы
- Нужно иметь представление о подходах и методах
- Требуется понимать, как «собирать» сценарии из компонентов
- «Все или ничего»
- Гибкость сильно ограничена

Цикл Loginom Customer Intelligence



LCI Churn



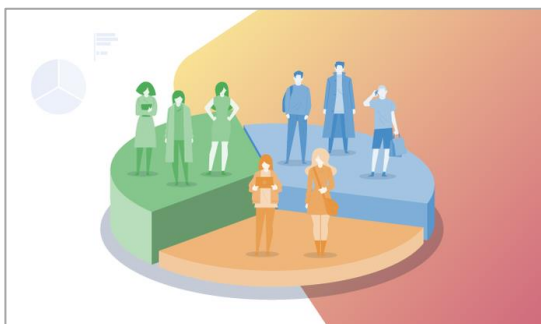
LCI Base Segmentation



LCI Single View



**Обнаружение и измерение
оттока**



**Базовые методы
сегментации клиентов**



**Профиль клиента в ритейле
Профиль клиента розничного банка**

Обнаружение оттока

LCI Churn

1. Вычисление точек ухода клиентов различными методами (для неконтактных систем взаимоотношений)
2. Обоснование корректности выбранного метода обнаружения оттока
3. Исследование активности и оттока клиентов по когортам в динамике



Компоненты библиотеки

Когда уйдет клиент?

Простые методы

Используют усредненные показатели по всем клиентам

Основаны на анализе времени между визитами

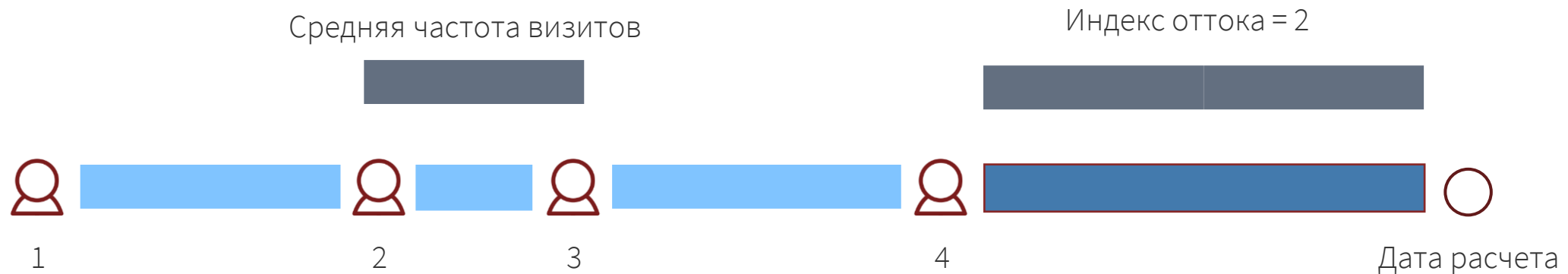
Сложные методы

Учитывают индивидуальное поведение каждого клиента

Основаны на понятиях **собственный интервал неактивности** и **индекс оттока**

Меры индивидуального поведения клиента

- Собственный интервал неактивности:
 $F(\text{Средняя частота визитов}, \text{Максимальная частота визитов})$
- Индекс оттока на дату: отношение числа дней с момента последнего визита к средней частоте визитов



Датасеты

Розничная DIY-сеть

- 1 млн. транзакций
- 41 тыс. клиентов
- Период 01.09.2017-31.10.2019
- 30 тыс. SKU, 17 товарных групп

Транзакции клиентов банка*

- 1,9 млн. транзакций
- 4,1 тыс. клиентов
- Период 01.08.2018 - 01.10.2019

***Sberbank Data Science Contest 2016**

DIY-сеть: результаты

- Структура клиентской базы требует учитывать индивидуальное поведение клиента при определении его точки ухода.
- Пороговый индекс оттока, охватывающий поведение 80% клиентов, равен **3,9**.

То есть клиент будет считаться ушедшим, если прошло в **3,9** раза больше времени средней частоты его визитов.

- На последний момент времени в базе у **70%** клиентов индекс оттока ≤ 1 .

Розничный банк: результаты

- Структура клиентской базы требует учитывать индивидуальное поведение клиента при определении его точки ухода.
- Пороговый индекс оттока, охватывающий поведение 80% клиентов, равен **10,3**.

То есть клиент будет считаться ушедшим, если прошло в **10** раз больше времени средней частоты его визитов.

- На последний момент времени в базе у **67%** клиентов индекс оттока ≤ 1 .

Пример веб-сервиса

- Используются готовые компоненты библиотеки **LCI Churn**.
- В БД хранятся рассчитанные индексы оттока для клиентов за два периода – прошлый и текущий.
- Веб-сервис возвращает список клиентов, которые ушли в отток по сравнению с прошлым периодом расчета.
- Вызов веб-сервиса осуществляется посредством REST-запроса, а текст запроса формируется в JSON.

Бизнес-курс

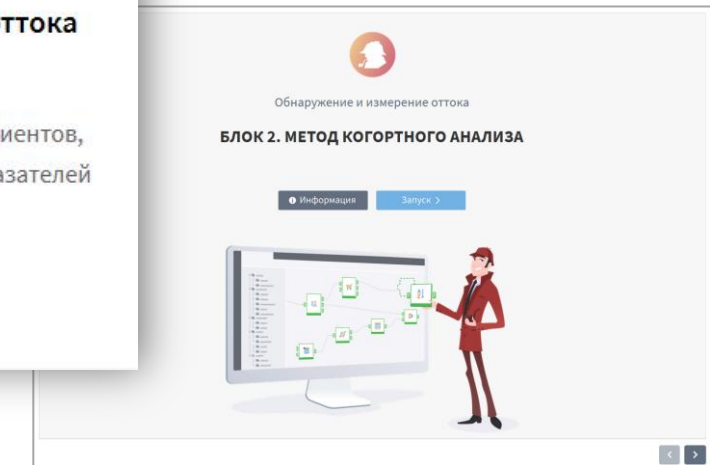
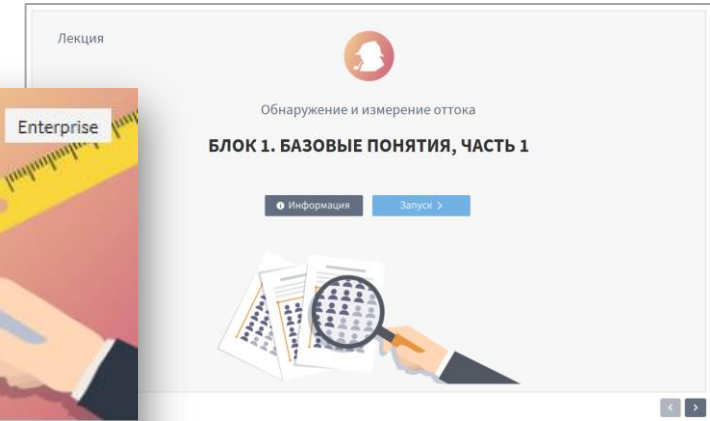
- Лекции
- Пошаговые симуляции с библиотекой компонентов LCI Churn
- Тесты
- Мобильное приложение, персонаж



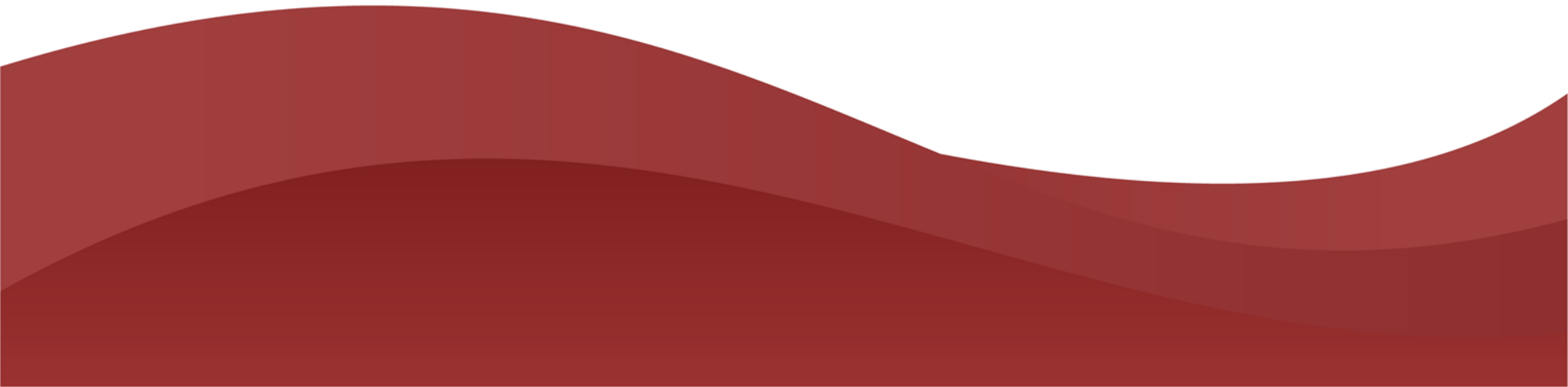
Обнаружение и измерение оттока

Методы обнаружения и измерения показателей оттока и активности клиентов, визуализация и интерпретация показателей удержания клиентов

Электронный курс

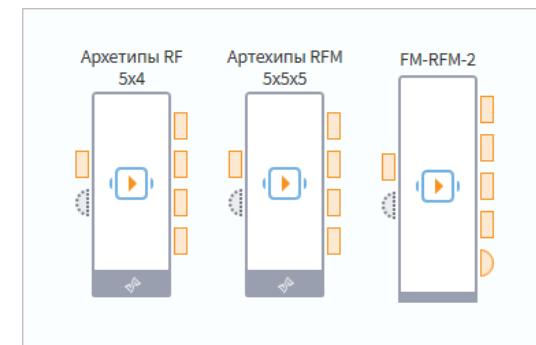
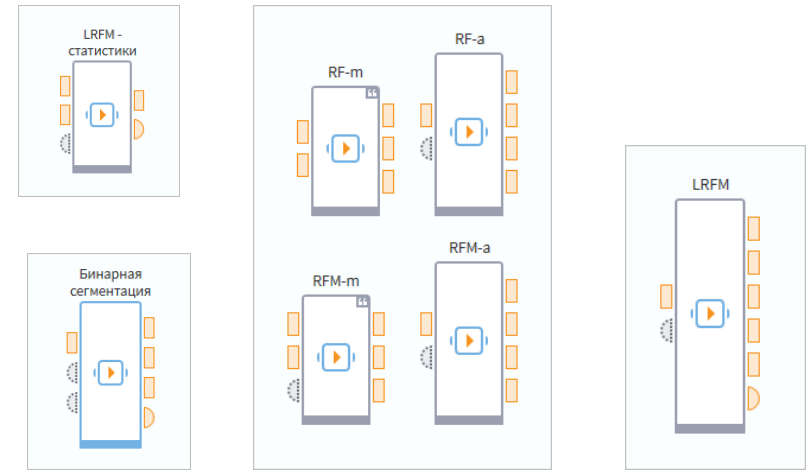


Сегментация



LCI Base Segmentation

1. Построение модели сегментации клиентской базы одним из доступных методов
2. Скоринг клиентской базы через модель сегментации
3. Готовые архетипы для интерпретации сегментов



Классификация методов

Методы базовой сегментации

Одномерные

Двумерные

Трёхмерные

Многомерные

Бинарная
сегментация

RF

FM

RFM

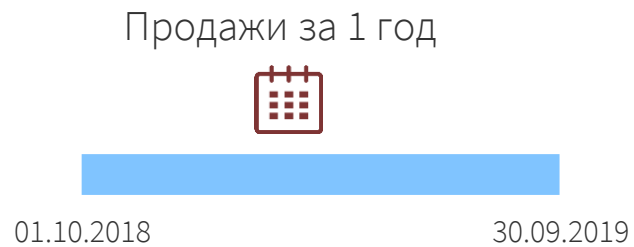
LRFM

Архетипы

- Аналитик оперирует небольшим числом сегментов.
- Близкие коды сегментов можно объединить, мы это назвали «**архетип**».
- Библиотека содержит лучшие экспертные практики архетипов - выберите те, которые вам подходят.



Кейс: RFM-сегментация



Модель RFM 5x5x5



11 архетипов



Сдвиг на 1 месяц



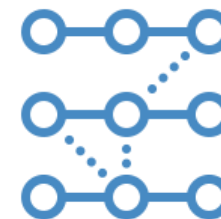
Скоринг через модель RFM



11 архетипов




Переходы



Бизнес-курс

- Лекции
- Пошаговые симуляции с библиотекой компонентов
- **LCI Base Segmentation**
- Тесты
- Мобильное приложение, персонаж




Enterprise

Базовые методы сегментации клиентов

Методы сегментации, основанные на поведении клиента в прошлом: RF, RFM, LRFM, бинарная сегментация

Электронный курс



Лекция



Базовые методы сегментации клиентов

БЛОК 4. ДВУМЕРНАЯ СЕГМЕНТАЦИЯ


Информация | Запуск >



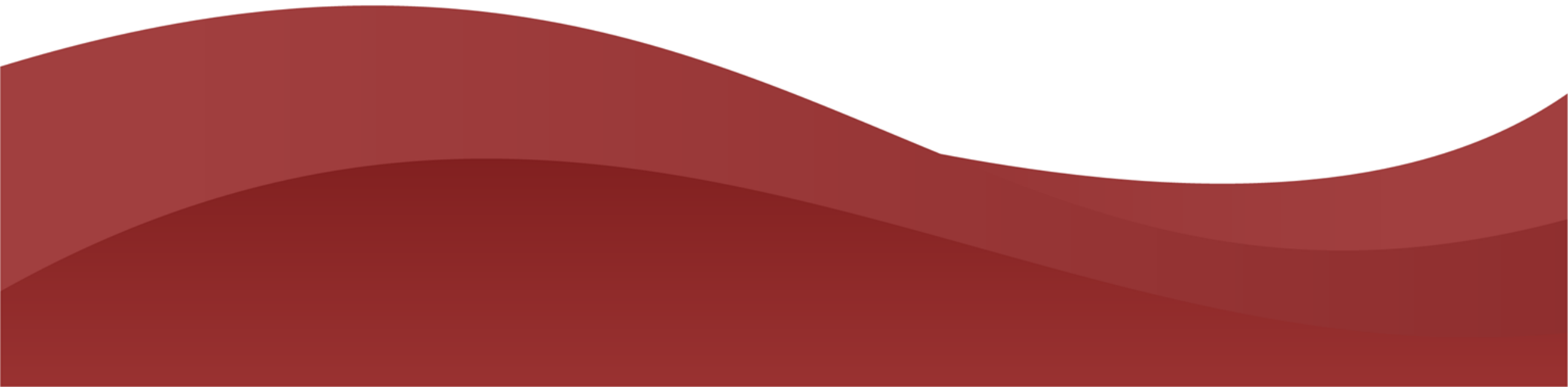
Базовые методы сегментации клиентов

БЛОК 5. RFM-СЕГМЕНТАЦИЯ

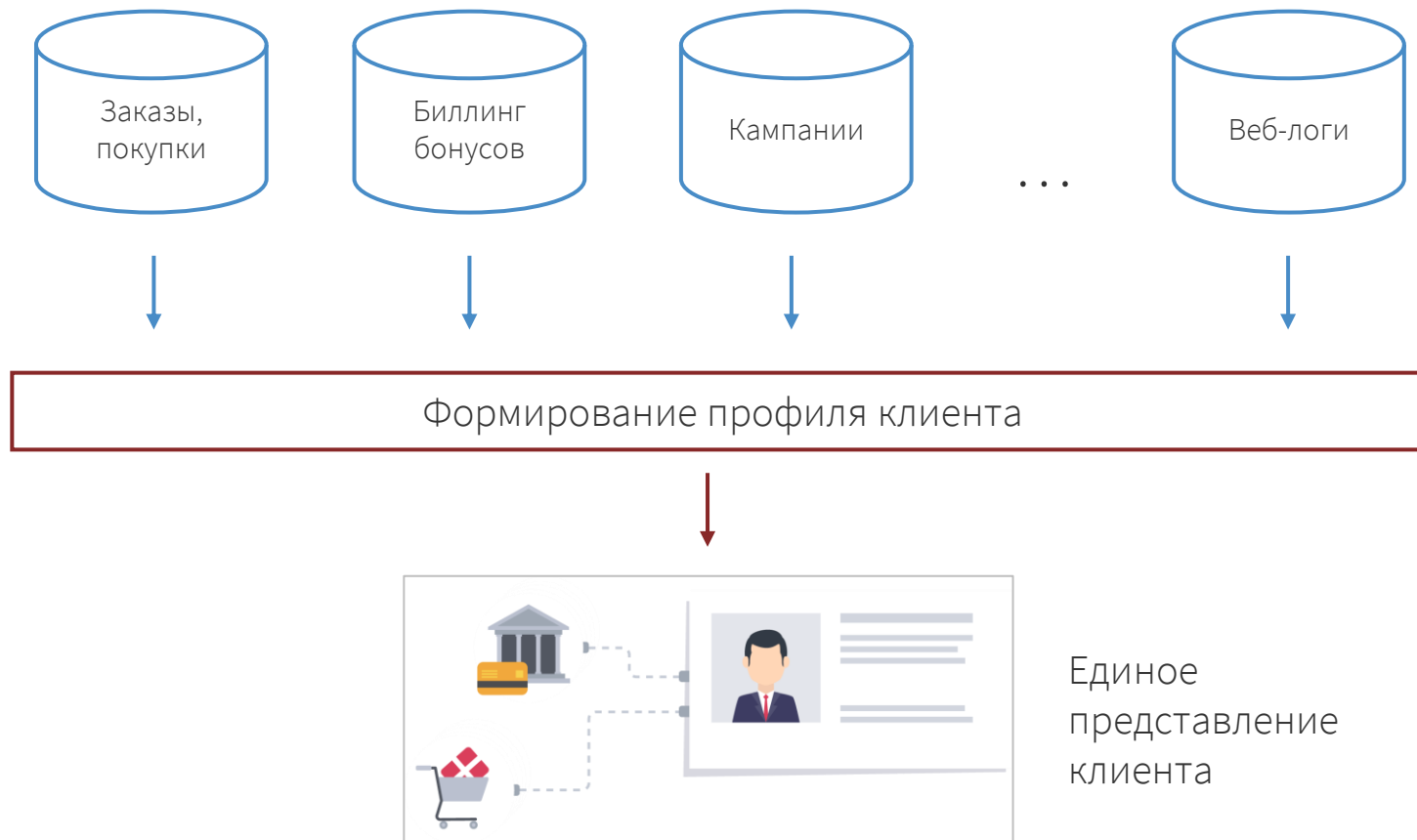
Информация | Запуск >



Профиль клиента

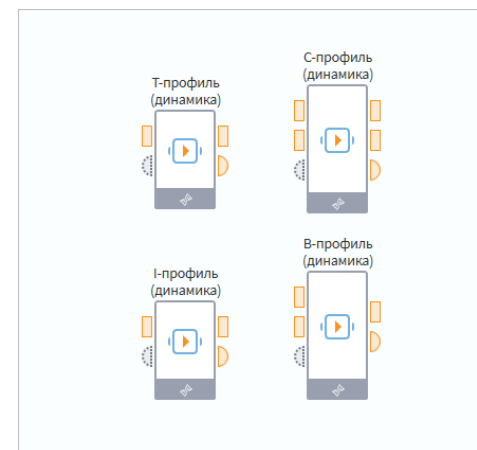
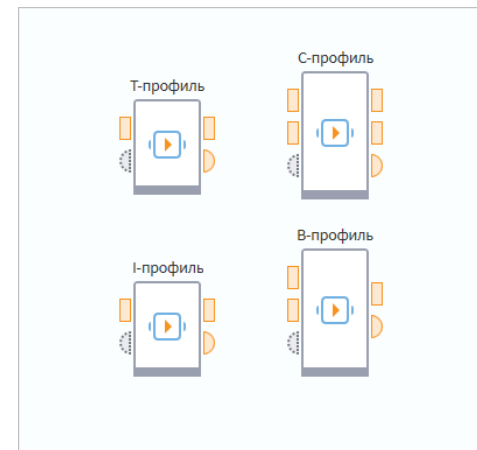


Единое представление клиента



LCI Single View

1. Расчет показателей для клиентских витрин данных из транзакций
2. Подготовка данных для описательного и предсказательного моделирования клиентов
3. Быстрая проверка гипотез в отношении клиентских баз данных



Требования к данным

Минимальные

- Дата транзакции, Клиент, Сумма

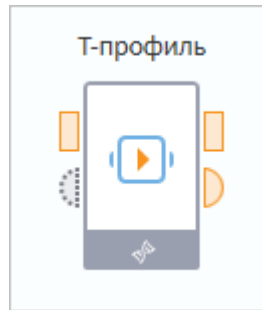
Опциональные (ритейл)

- Время транзакции, Количество, Товар, Товарная группа/подгруппа/бренд, Канал продаж

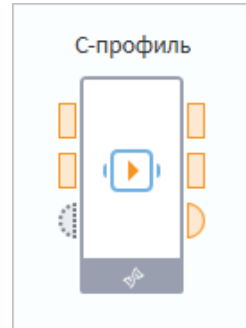
Опциональные (банк)

- Время транзакции, МСС-код, Группа МСС, Тип транзакции, Идентификатор терминала

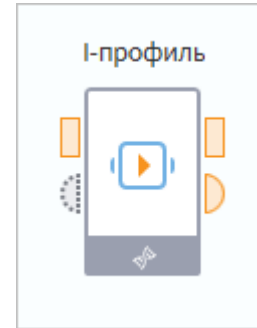
Архитектура библиотеки



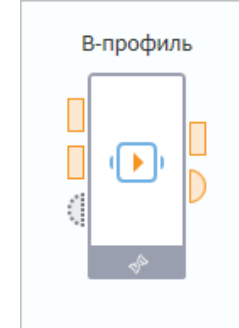
Агрегаты на основе даты, времени, суммы, канала продаж



Агрегаты на основе иерархии категорий

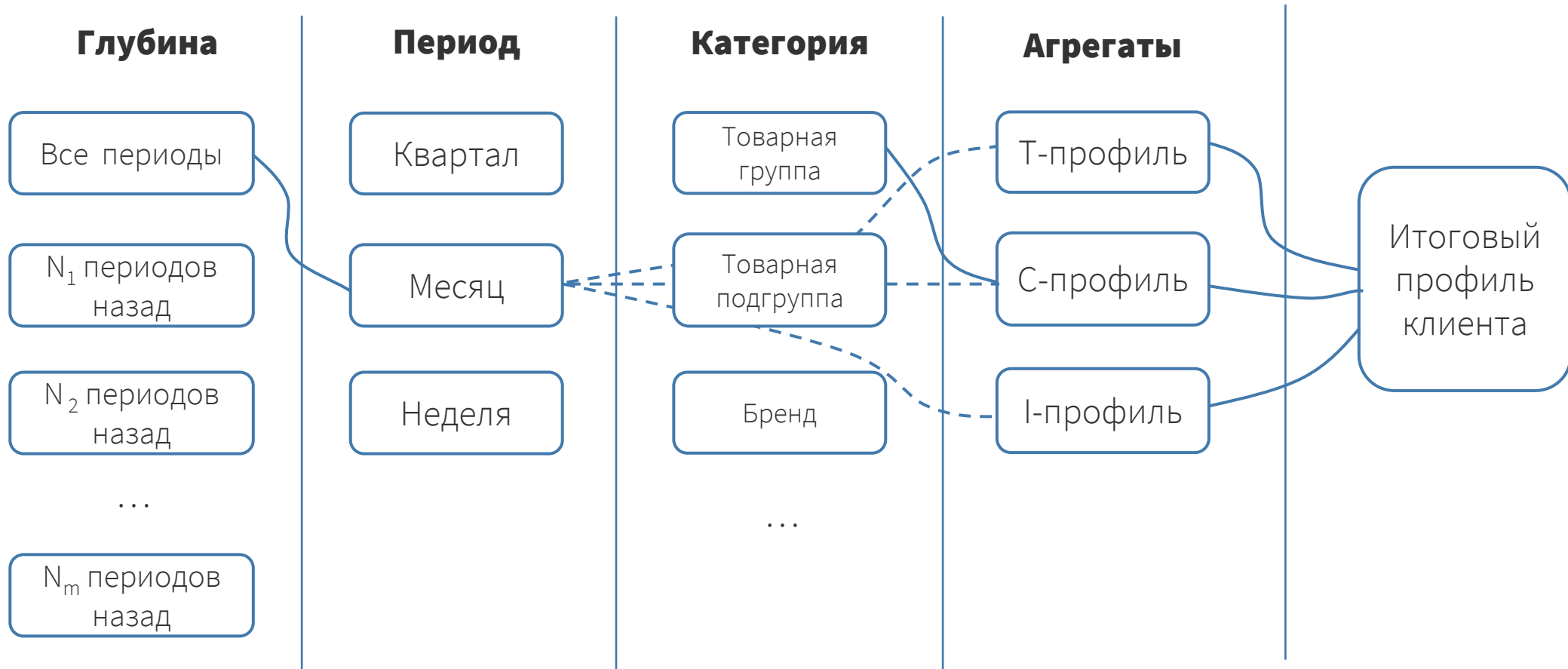


Агрегаты на основе отдельной позиции (товар, МСС-код)

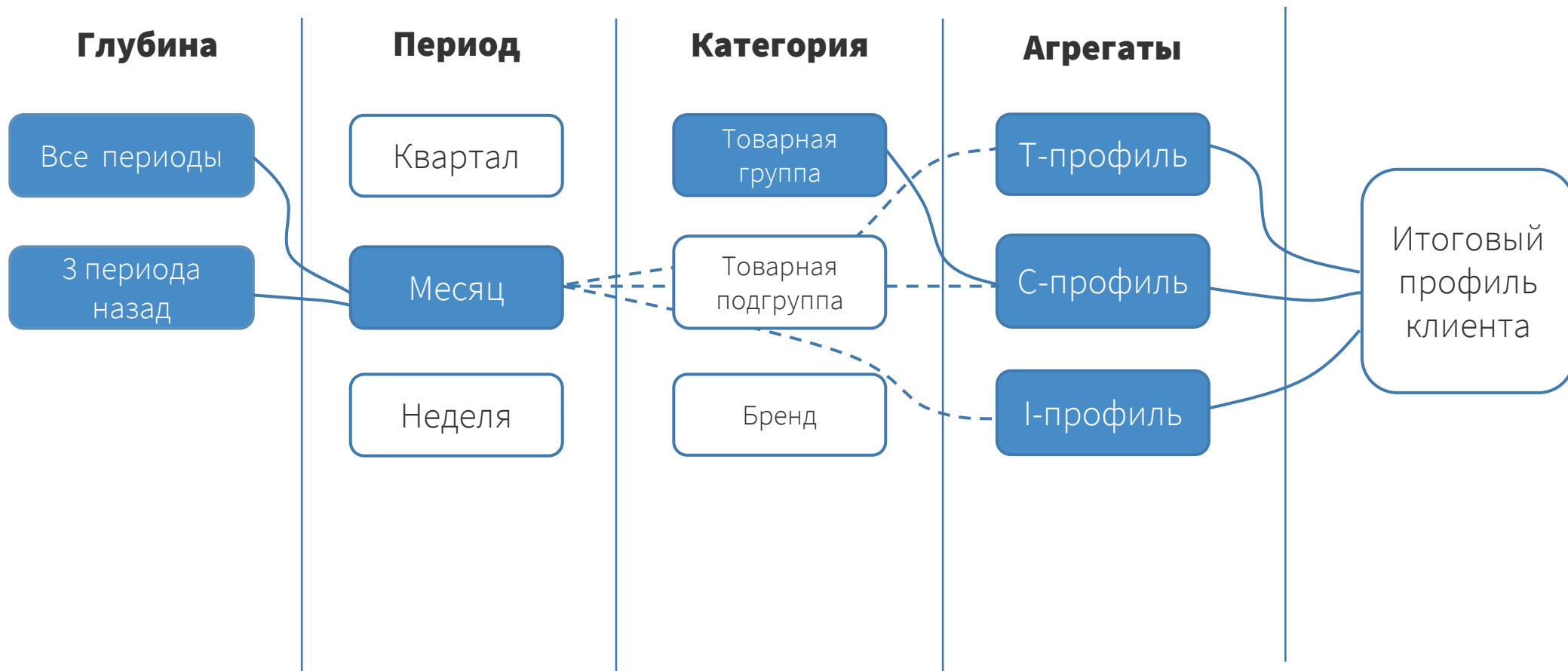


Агрегаты по типам операций (только для банков)

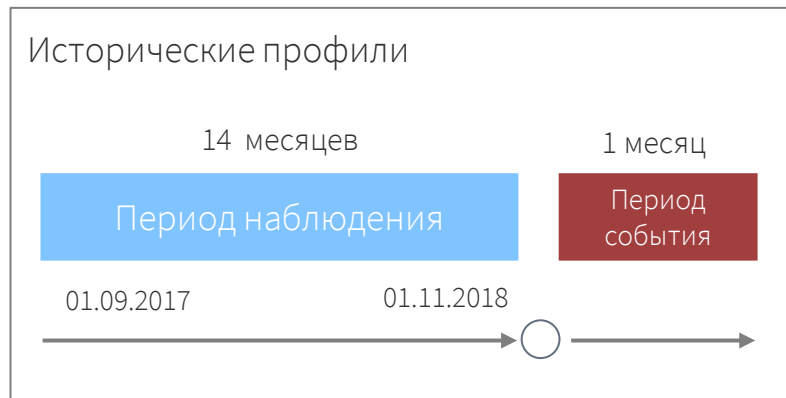
Конструктор профиля клиента



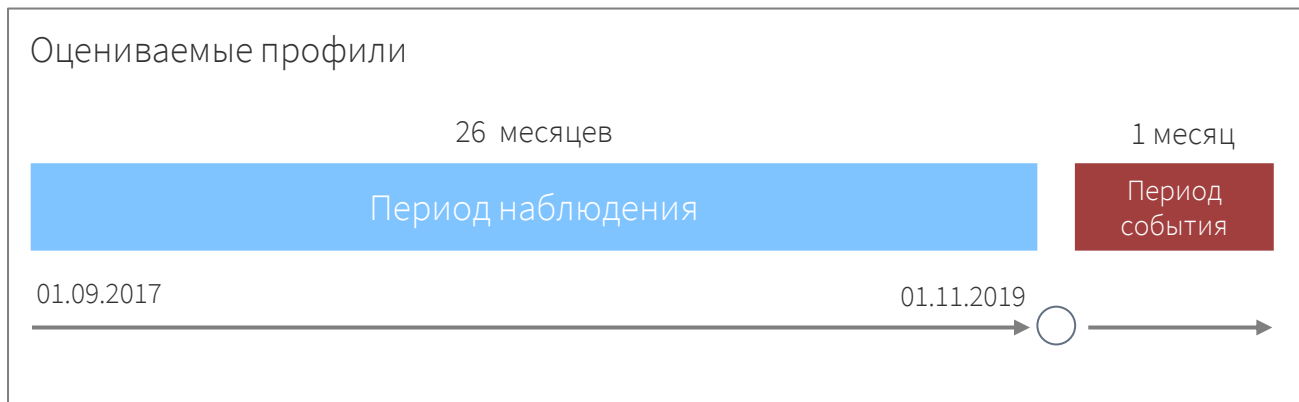
Кейс: профиль клиента DIY-сети



Кейс: вероятность повторной покупки



**Логистическая
ридж-
регрессия**
(23, 7 тыс.
клиентов)



Вероятности
визитов для
41 тыс.
клиентов

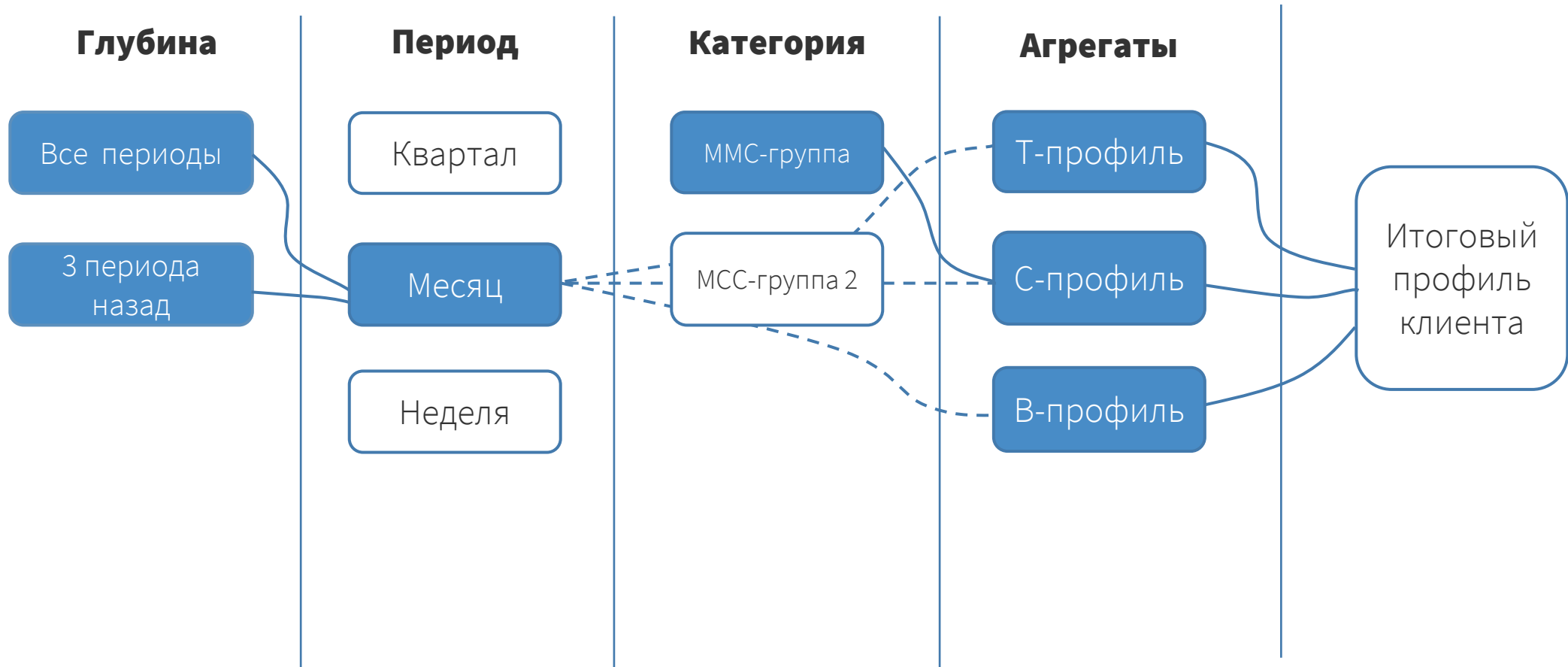
Стандартные группы МСС-кодов

1. Контрактные услуги
2. Оптовые поставщики и производители
3. Авиалинии, авиакомпании
4. Аренда автомобилей
5. Отели и мотели
6. Перевозки
7. Коммунальные и кабельные услуги
8. Поставщики услуг
9. Розничные магазины
10. Автомобили и транспортные средства
11. Магазины одежды
12. Различные магазины
13. Продажи по почте/телефону
14. Личные услуги
15. Бизнес услуги
16. Ремонтные услуги
17. Развлечения
18. Профессиональные услуги
19. Членские организации
20. Государственные услуги
21. Другое

Экспертные группы МСС-кодов

1. Путешествия
2. Красота и здоровье
3. Техника и электроника
4. Телекоммуникации
5. Коммунальные услуги
6. Дом и сад
7. Еда/Супермаркеты
8. Бензин
9. Одежда
10. Аксессуары
11. Развлечения
12. Книги/Музыка
13. Фитнес
14. Культура
15. Образование
16. Другое

Кейс: профиль клиента розничного банка



Кейс: сегментация по предпочтениям МСС-групп

- Используются готовые компоненты библиотеки **LCI Single View**.
- Взяв за основу рассчитанный **С-профиль** с экспертной группировкой МСС-кодов, построим модель сегментации методом машинного обучения.
- Предварительно 16 входных переменных «сожмем» в 6 методом факторного анализа «варимакс».
- Построим модель кластеризации алгоритмом k-means с числом кластеров 7.
- Оценим качество кластеризации через индекс силуэта.

Производительность

Датасет

- Транзакции X5 ([RetailHero Contest 2020](#))
- 45 млн. транзакций, 400 тыс. клиентов
- Период 21.11.2018 - 18.03.2019

Вычислительные ресурсы

- ОЗУ 8 Гб
- CPU 2 ядра
- СУБД SQL Lite

Результаты

- Размер пачки 2 млн. записей
- Клиентов в пачке 17 тыс.
- Общее время **01 ч 58 м**
- Среднее время обработки пачки **5 м 9 с**

Варианты поставок

Максимум

Подписка на
Loginom e-Learning

Оптимум

Подписка только на
библиотеку
компонентов

Эконом

Аренда в облаке,
аналитический сервис

Акция

Пилотный кейс – данные
ваши, библиотеки
компонентов наши

Бесплатно

Демо-версия
библиотеки
компонентов

Вопросы?