

Библиотека «Loginom RFM Kit»

Версия 3.0.0 (для версии Loginom 7.0.2)

Компоненты библиотеки **Loginom RFM Kit** реализуют группу методов сегментации, основанных на RFM-методах и их различных вариациях. Поставляются к бизнес-курсу **Базовые методы сегментации клиентов** в Loginom Skills. Входными данными служат клиентские транзакции.

Реализованы несколько методов: RF, RFM, LRFM, бинарная сегментация. Для ряда методов добавлено объединение сегментов по архетипам. Компоненты отличаются между собой:

- самим методом сегментации;
- фиксированным или переменным (задаваемым пользователем) числом сегментов;
- режимом работы: компонент строит модель сегментации или позволяет "прогнать" данные через готовую модель.

Мнемоника названий компонентов следующая.

- Постфикс **–a** означает, что границы сегментов рассчитываются автоматически (строится модель сегментации).
- Постфикс **–m** означает, что границы сегментов должны быть поданы на вход пользователем (осуществляется "прогон" через готовую модель).
- Числа в названии компонентов означают, что на выходе будет фиксированное число сегментов. Например, **RF 5x4** — RF-сегментация с разбиением **R** на 5 интервалов и **F** на 4 интервала. Фиксированные разбиения необходимы, чтобы распределить сегменты по архетипам.

В данной библиотеке все методы, а также классификации архетипов взяты из лучших практик по клиентской аналитике. По возможности мы приводим ссылки на первоисточник.

Комплект поставки

Файл/Каталог	Назначение
loginom_rfm_kit.lgp	Пакет с компонентами и метакomпонентами библиотеки
Кейсы использования.lgp	Пакет с примером использования компонентов библиотеки
Данные (data)	Демонстрационные наборы данных в формате .lgd

- *Метакomпонент — это производный компонент, который собран из компонентов библиотеки под конкретную задачу.*

Установка и настройка

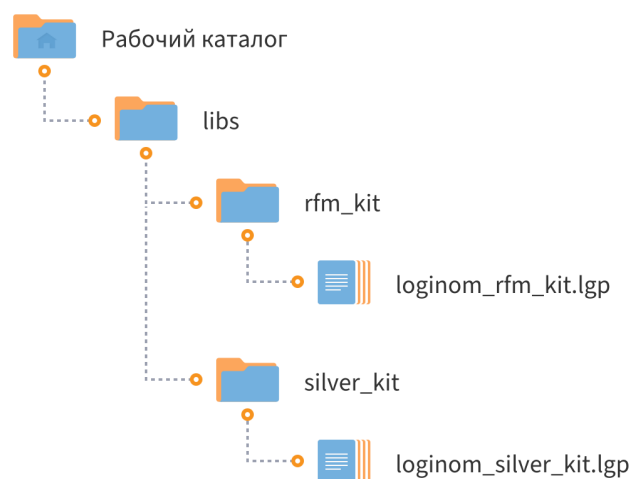
Требования

Для работы библиотеки **Loginom RFM Kit** необходимо:

- Клиент доступа к Loginom Studio или любая локальная редакция Loginom. Версия не ниже 7.0.2
- Библиотека **Loginom Silver Kit** ([скачать на GitHub](#)). Версия не ниже 3.0.0

Порядок установки

1. Перейдите в файловое хранилище клиента Loginom Studio или назначьте каталог на локальном диске в случае Loginom CE.
2. Создайте каталог **libs**.
3. Распакуйте архив **rfm_kit.zip** и поместите все файлы из архива в каталог **libs** так, как они лежали в архиве.
4. Убедитесь, что библиотека **Loginom Silver Kit** находится в каталоге **libs** в папке **silver_kit**.



Требования к данным

- **Транзакции** — подается на компоненты **Бинарная сегментация** и **LRFM-статистики**.
- **Дата первого визита** — подается на компонент **LRFM-статистики**.
- **Интервалы** — подается на компоненты **RF-m**, **RFM-m**.

Транзакции

Набор данных представляет собой список транзакций клиента за определенный период, выбранный для анализа.

Структура набора

Метка	Тип	Описание
Идентификатор	ab Строковый	Уникальный идентификатор клиента
Дата	31 Дата/Время	Дата транзакции
Сумма	9.0 Вещественный	Сумма транзакции

Дата первого визита

Набор данных требуется для расчета L (Length, Продолжительность). Он должен содержать дату первого события, зарегистрированного с клиентом: первого визита, регистрации в программе лояльности и т.п.

Структура набора

Метка	Тип	Описание
Идентификатор	ab Строковый	Уникальный идентификатор клиента
Дата первого визита	31 Дата/Время	Если L считать не требуется, или первая дата должна быть рассчитана из данных, подается пустое поле

Интервалы

Модель сегментации, которая представляет собой границы интервалов по показателям.

Структура набора

Метка	Тип	Описание
Идентификатор интервалов	ab Строковый	Можгут использоваться идентификаторы: Recency, Frequency, Monetary
Верхняя граница	90 Вещественный	Верхние границы интервалов

Компоненты

- Бинарная сегментация
- LRFM-статистики
- RF-a
- RF-m
- Архетипы RF 5x4
- RFM-a
- RFM-m
- Архетипы RFM 5x5x5
- FM-RFM-2
- LRFM
- meta-RFM-a 5x5x5
- meta-RFM-m 5x5x5

Компоненты зашифрованы и доступны к использованию через панель производных компонентов.

Бинарная сегментация

Назначение

Проведение бинарной сегментации клиентов на основе изучения их активности в четырех периодах, включая период актуальности. Дополнительно проводится расчет вероятности визита в пятом периоде. Таким образом на вход необходимо подать транзакции за пять периодов. В качестве периода могут выступать день, неделя, месяц, квартал или год, нужный тип периода определяется пользователем.

Активность клиента в периодах представлена в виде кода вида XXXX, где X может принимать значения 1 (клиент приходил в периоде) или 0 (клиент не приходил).

Добавлена коллекция архетипов для сегментов.



Входные порты

Название	Тип
Транзакции	Таблица
Переменные	Переменные
Архетипы	Переменные

Структура таблицы "Транзакции"

Структура таблицы описана в разделе [Требования к данным](#).

Переменные порта "Переменные"

№	Метка	Тип	Значение
1	Дата актуальности	 Дата/Время	Пропущенное значение
2	Тип периода	 Строковый	m

Дата актуальности — дата, на которую проводится сегментация. На основе нее определяется четвертый период, после чего вычисляются остальные. При установке флага **Пропущенное значение** берется последняя дата из имеющихся данных (по умолчанию). В этом случае вероятность следующего визита определена не будет, так как "следующий" период в данных отсутствует.

Тип периода — переменная может принимать следующие значения:

1. **d** — день;
2. **w** — неделя;
3. **m** — месяц (по умолчанию);
4. **q** — квартал;
5. **y** — год.

Переменные в порте "Архетипы"

№	Метка	Тип	Значение
1	Ушедший	ab Строковый	0000
2	Реактивированный	ab Строковый	0001
3	Спящий	ab Строковый	0010
4	Активный	ab Строковый	0011
5	Спящий	ab Строковый	0100
6	Реактивированный	ab Строковый	0101
7	Спящий	ab Строковый	0110
8	Активный	ab Строковый	0111
9	Ушедший	ab Строковый	1000
10	Реактивированный	ab Строковый	1001
11	Спящий	ab Строковый	1010
12	Активный	ab Строковый	1011
13	Спящий	ab Строковый	1100
14	Реактивированный	ab Строковый	1101
15	Спящий	ab Строковый	1110
16	Активный	ab Строковый	1111

Выходные порты

Название	Тип
Сегменты	Таблица
Статистики по сегментам	Таблица
Справочник архетипов	Таблица
Служебные переменные	Переменные

Структура таблицы "Сегменты"

Метка	Тип	Описание
Идентификатор	ab Строковый	Уникальный идентификатор клиента
Код сегмента	ab Строковый	Кода вида XXXX, где x может принимать значения 1 или 0
Архетип	ab Строковый	Архетип, к которому относится сегмент, 4 варианта: Активные, Спящие, Реактивированные, Ушедшие
Вес клиента	12 Целый	Значимость клиента, см. Алгоритмы

Структура таблицы "Статистики по сегментам"

Метка	Тип	Описание
Код сегмента	ab Строковый	Кода вида XXXX, где X может принимать значения 1 или 0
Архетип	ab Строковый	Архетип, к которому относится сегмент
Численность сегмента	12 Целый	Количество клиентов в сегменте
Вес клиента	12 Целый	Значимость клиента сегмента, см. Алгоритмы
Вероятность	9.0 Вещественный	Вероятность визита клиента сегмента в периоде, следующем за периодом актуальности
Кол-во активных в след. периоде	12 Целый	Количество клиентов сегмента, сделавших покупку в периоде, следующем за периодом актуальности

Структура таблицы "Справочник архетипов"

Метка	Тип	Описание
Код сегмента	ab Строковый	Список кодов сегментов
Архетип	ab Строковый	Архетип, к которому относится конкретный код

Переменные в порте "Служебные переменные"

№	Метка	Тип	Описание
1	Окончание подписки	31 Дата/Время	Дата, до которой будет работать компонент

Алгоритмы

1. Вес (значимость) клиента = $1 \cdot (1\text{-я цифра кода сегмента}) + 2 \cdot (2\text{-я цифра кода сегмента}) + 4 \cdot (3\text{-я цифра кода сегмента}) + 8 \cdot (4\text{-я цифра кода сегмента})$; Одинаков для всех клиентов, которые входят в один сегмент.
2. Вероятность = $\frac{\text{Доля клиентов, сделавших покупку в след. месяце}}{\text{Кол-во активных в след. периоде}} / \text{Численность сегмента}$.

LRFM-статистики

Назначение

Расчет показателей продолжительности отношений клиента и компании (Length), давности последней покупки (Recency), частоты/количества покупок (Frequency), стоимости (Monetary).

Другие варианты перевода Monetary в русскоязычной литературе: деньги, ценность, средний чек, вложения, сумма денег, объем покупок, стоимость, общая сумма, денежная ценность. Если планируется провести RF-анализ, вместо суммы для расчета стоимости можно подать пустое поле.

Предполагается, что данные, которые подаются на вход узла, предварительно обработаны: сгруппированы на уровне SKU, исключены нулевые транзакции, возвраты, взяты данные за исследуемый период.

Входные порты

Название	Тип
Транзакции	Таблица
Дата первого визита	Таблица
Переменные	Переменные




Структура таблицы "Транзакции"

Структура таблицы описана в разделе [Требования к данным](#).

Структура таблицы "Дата первого визита"

Структура таблицы описана в разделе [Требования к данным](#).

Переменные порта "Переменные"

№	Метка	Тип	Значение
1	Дата актуальности	 Дата/Время	Пропущенное значение
2	Стоимость: способ расчета	 Строковый	sum
3	Группировка однодневных визитов	 Логический	true

Дата актуальности — дата, относительно которой рассчитываются **Давность** и **Продолжительность**. Если установлен флаг **Пропущенное значение**, берется последняя дата из имеющихся данных.

Стоимость: способ расчета — способ расчета показателя. Данная переменная может принимать следующие значения:

- **sum** — общая сумма (по умолчанию);
- **avg** — среднее.

Группировка однодневных визитов — требуется для расчета частоты:

- **true** — несколько транзакций в один день считаются одним визитом (по умолчанию);
- **false** — каждая транзакция считается отдельным визитом.

Выходные порты

Название	Тип
Статистики	Таблица
Служебные переменные	Переменные

Структура таблицы "Статистики"

Метка	Тип	Описание
Идентификатор	ab Строковый	Уникальный идентификатор клиента
Продолжительность	12 Целый	Количество дней между первым и последним визитом клиента
Давность	12 Целый	Количество дней, прошедшее между датой актуальности и датой последнего визита
Частота	12 Целый	Количество визитов клиента
Стоимость	90 Вещественный	Стоимость (общая, либо средняя сумма) транзакций клиента

Переменные в порте "Служебные переменные"

№	Метка	Тип	Описание
1	Окончание подписки	31 Дата/Время	Дата, до которой будет работать компонент

Алгоритмы

Length, дни = Дата последней транзакции клиента – Дата первой транзакции клиента

Recency, дни = Дата актуальности - Дата последней транзакции клиента

Frequency, шт. = Количество транзакций клиента

Monetary, (y.e.) = $\Sigma(\text{Сумма})$, либо **Monetary**, (y.e.) = $\Sigma(\text{Сумма}) / \text{Количество транзакций}$

RF-a

Назначение

Построение модели RF-сегментации: расчет границ интервалов R, F и соотнесение клиентов с интервалами. Количество интервалов по каждому показателю задается входными переменными; границы рассчитываются внутри компонента с помощью узла **Квантование** (метод — плитка). По каждому сегменту в компоненте рассчитывается ряд статистик.

Выходные наборы данных формируются в любом случае, однако, если по какому-либо параметру не удалось сформировать заданное количество интервалов, можно увидеть сообщение об этом в соответствующем выходном порте.

Входные порты

Название	Тип
Статистики	Таблица
Переменные	Переменные

Структура таблицы "Статистики"

Подаются данные с одноименного выхода компонента **LRFM-статистики**.

Метка	Тип	Описание
Идентификатор	ab Строковый	Уникальный идентификатор клиента
Давность	12 Целый	Количество дней, прошедшее между датой актуальности и датой последнего визита
Частота	12 Целый	Количество визитов клиента

Переменные порта "Переменные"

№	Метка	Тип	Значение
1	Давность: количество интервалов	12 Целый	5
2	Частота: количество интервалов	12 Целый	4
3	Отделять одновизитников	0/1 Логический	false
4	Частота: граница 1-го интервала	12 Целый	1

1. **Отделять одновизитников** — позволяет отделить клиентов с небольшим количеством посещений перед расчетом интервалов по частоте.
Рекомендуется использовать, если большая часть клиентов совершали 1-2 посещения. Флаг включен — граница первого интервала по частоте будет задаваться пользователем.
2. **Частота: граница 1-го интервала** — верхняя граница первого интервала по частоте (используется, если включен флаг **Отделять одновизитников**).
Граница входит в интервал.

Выходные порты

Название	Тип
Сегменты	Таблица
Статистики по сегментам	Таблица
Интервалы	Таблица
Сообщения	Таблица

Структура таблицы "Сегменты"

Метка	Тип	Описание
Идентификатор	ab Строковый	Уникальный идентификатор клиента
Код сегмента	ab Строковый	Код вида RF, где R – номер интервала по давности, F – номер интервала по частоте
Давность: номер интервала	12 Целый	Номер интервала клиента по давности. Чем больше номер, тем меньше времени прошло с последней покупки
Частота: номер интервала	12 Целый	Номер интервала клиента по частоте. Чем больше номер, тем чаще приходил клиент
Давность	12 Целый	Значение давности для клиента
Частота	12 Целый	Значение частоты для клиента

Структура таблицы "Статистики по сегментам"

Метка	Тип	Описание
Код сегмента	ab Строковый	Код вида RF, где R – номер интервала по давности, F — номер интервала по частоте
Давность: номер интервала	12 Целый	Номер интервала по давности
Частота: номер интервала	12 Целый	Номер интервала по частоте
Количество клиентов	12 Целый	Количество клиентов в сегменте
% клиентов	90 Вещественный	% клиентов сегмента от общего количества клиентов
Количество событий	90 Вещественный	Количество событий (покупок, транзакций, посещений и т.п.), совершенных клиентами сегмента
% кол-ва событий	90 Вещественный	% событий в сегменте от общего количества событий
Среднее кол-во событий	90 Вещественный	Среднее количество событий на клиента сегмента
Средняя давность (дни)	90 Вещественный	Средняя давность события по сегменту

Структура таблицы "Интервалы"

Метка	Тип	Описание
Идентификатор интервалов	ab Строковый	Идентификатор интервала, два варианта: Recency; Frequency
Номер интервала	12 Целый	Порядковый номер интервала (обратный порядок для Recency)
Нижняя граница	9.0 Вещественный	Нижняя граница диапазона открыта, первое значение границы — пустое
Верхняя граница	9.0 Вещественный	Верхняя граница диапазона открыта, последнее значение границы — пустое

Структура таблицы "Сообщения"

Метка	Тип	Описание
Идентификатор интервалов	ab Строковый	Идентификатор интервала
Сообщение	ab Строковый	Выводится предупреждение, если по показателю не удалось сформировать заданное на входе количество интервалов

RF-m

Назначение

Распределение клиентов по интервалам R, F. Количество интервалов и их границы известны заранее и подаются на вход компонента. Нижняя и верхняя границы диапазона открыты. Могут использоваться границы, рассчитанные в компоненте RF-a.

В компоненте присутствует первичная валидация границ. Если валидация не пройдена, выходные наборы данных сформированы не будут. Сообщение о причине будет выведено в соответствующем выходном порте.

Входные порты

Название	Тип
Статистики	Таблица
Интервалы	Таблица

Структура таблицы "Статистики"

Подаются данные с одноименного выхода компонента **LRFM-статистики**.

Метка	Тип	Описание
Идентификатор	ab Строковый	Уникальный идентификатор клиента
Давность	12 Целый	Количество дней, прошедшее между датой актуальности и датой последнего визита
Частота	12 Целый	Количество визитов клиента

Структура таблицы "Интервалы"

Структура таблицы описана в разделе [Требования к данным](#).

Выходные порты

Название	Тип
Сегменты	Таблица
Статистики по сегментам	Таблица
Сообщения	Таблица

Структура таблицы "Сегменты"

Структура набора соответствует одноименному набору на выходе компонента [RF-a](#).

Структура таблицы "Статистики по сегментам"

Структура набора соответствует одноименному набору на выходе компонента [RF-a](#).

Структура таблицы "Сообщения"

Структура набора соответствует одноименному набору на выходе компонента [RF-a](#).

Архетипы RF 5x4

Назначение

Компонент предназначен для распределения полученных RF-сегментов по архетипам ([Jim Porzak, 2008](#)):

- Новый;
- 1 визит;
- Повторный;
- Постоянный;
- Спящий;
- Ушедший.

Компонент можно использовать только для модели сегментации, в которой по давности сформировано пять, а по частоте — четыре интервала, всего 20 сегментов. Внутри компонента проводится валидация на соответствие этим условиям. Если условия не соблюдены, выходные наборы данных не формируются, на выходе **Сообщения** выводится причина.

Входные порты

Название	Тип
Сегменты	Таблица
Архетипы	Переменные

Структура таблицы "Сегменты"

Подаются данные с одноименного выхода компонента **RF-a** или **RF-m**.

Переменные в порте "Архетипы"

Каждая переменная содержит соответствие **Код сегмента–Архетип**. Ниже приведены несколько, полный список можно увидеть на входе узла.

№	Метка	Тип	Значение
1	Постоянный	ab Строковый	54
2	Повторный	ab Строковый	53
3	Повторный	ab Строковый	52
...			
20	Ушедший	ab Строковый	11

Выходные порты

Название	Тип
Сегменты	Таблица
Статистики по сегментам	Таблица
Справочник архетипов	Таблица
Сообщения	Таблица

Структура таблицы "Сегменты"

Метка	Тип	Описание
Идентификатор	ab Строковый	Уникальный идентификатор клиента
Код сегмента	ab Строковый	Код сегмента, к которому относится клиент
Архетип	ab Строковый	Архетип, к которому относится сегмент клиента

Структура таблицы "Статистики по сегментам"

Метка	Тип	Описание
Архетип	ab Строковый	Наименование архетипа
Количество клиентов	12 Целый	Количество клиентов по архетипу
% клиентов	9.0 Вещественный	% клиентов по архетипу от общего количества клиентов
Количество событий	9.0 Вещественный	Количество событий, совершенных клиентами по архетипу
% кол-ва событий	9.0 Вещественный	% событий по архетипу от общего количества событий
Среднее кол-во событий	9.0 Вещественный	Среднее количество событий на клиента архетипа
Средняя давность (дни)	9.0 Вещественный	Средняя давность события по архетипу

Структура таблицы "Справочник архетипов"

Метка	Тип	Описание
Код сегмента	ab Строковый	Список кодов сегментов
Архетип	ab Строковый	Архетип, к которому относится текущий код

Структура таблицы "Сообщения"

Метка	Тип	Описание
Идентификатор интервалов	ab Строковый	Идентификатор интервала
Сообщение	ab Строковый	Выводится предупреждение, если поданный на вход результат сегментации не соответствует условиям

RFM-a

Назначение

Построение модели RFM-сегментации: расчет границ интервалов R, F, M, соотнесение клиентов с этими интервалами и расчет RFM-оценки. Количество интервалов задается входными переменными; границы рассчитываются внутри компонента с помощью узла **Квантование** (метод — плитка).

Выходные наборы данных формируются в любом случае, однако, если по какому-то параметру не удалось сформировать заданное количество интервалов, можно увидеть сообщение об этом в соответствующем выходном порте.

Входные порты

Название	Тип
Статистики	Таблица
Переменные	Переменные

Структура таблицы "Статистики"

Подаются данные с одноименного выхода компонента **LRFM-статистики**.

Метка	Тип	Описание
Идентификатор	ab Строковый	Уникальный идентификатор клиента
Давность	12 Целый	Количество дней, прошедшее между датой актуальности и датой последнего визита
Частота	12 Целый	Количество визитов клиента
Стоимость	9.0 Вещественный	Стоимость (общая, либо средняя сумма) транзакций клиента

Переменные порта "Переменные"

№	Метка	Тип	Значение
1	Давность: кол-во интервалов	12 Целый	5
2	Давность: вес	90 Вещественный	100
3	Частота: количество интервалов	12 Целый	5
4	Частота: вес	90 Вещественный	10
5	Стоимость: количество интервалов	12 Целый	5
6	Стоимость: вес	90 Вещественный	1
7	Отделять одновизитников	0/1 Логический	false
8	Частота: граница 1-го интервала	12 Целый	1

Давность: вес — используется для расчета RFM-оценки.

Частота: вес — используется для расчета RFM-оценки.

Стоимость: вес — используется для расчета RFM-оценки.

Отделять одновизитников — позволяет отделить клиентов с небольшим количеством посещений перед расчетом интервалов по частоте. Рекомендуется использовать, если большая часть клиентов совершали 1-2 посещения. Флаг включен – граница первого интервала по частоте будет задаваться пользователем.

Частота: граница 1-го интервала — верхняя граница первого интервала по частоте (используется, если включен флаг **Отделять одновизитников**). Граница входит в интервал.

Выходные порты

Название	Тип
Сегменты	Таблица
Статистики по сегментам	Таблица
Интервалы	Таблица
Сообщения	Таблица

Структура таблицы "Сегменты"

Метка	Тип	Описание
Идентификатор	ab Строковый	Уникальный идентификатор клиента
Код сегмента	ab Строковый	Код вида RFM, где R — номер интервала по давности, F — номер интервала по частоте, M — номер интервала по стоимости
RFM-оценка	9.0 Вещественный	Рассчитывается по формуле, приведенной в разделе Алгоритмы
Давность: номер интервала	12 Целый	Номер интервала клиента по давности. Чем больше номер, тем меньше давность
Частота: номер интервала	12 Целый	Номер интервала клиента по частоте. Чем больше номер, тем чаще покупал
Стоимость: номер интервала	12 Целый	Номер интервала клиента по стоимости. Чем больше номер, тем больше денег потратил
Давность	12 Целый	Значение давности для клиента
Частота	12 Целый	Значение частоты для клиента
Стоимость	9.0 Вещественный	Значение стоимости для клиента

Структура таблицы "Статистики по сегментам"

Метка	Тип	Описание
Код сегмента	ab Строковый	Код вида RFM
RFM-оценка	9.0 Вещественный	Рассчитывается по формуле, приведенной в разделе Алгоритмы
Давность: номер интервала	12 Целый	Номер интервала клиента по давности
Частота: номер интервала	12 Целый	Номер интервала клиента по частоте
Стоимость: номер интервала	12 Целый	Номер интервала клиента по стоимости
Количество клиентов	12 Целый	Количество клиентов в сегменте
% клиентов	9.0 Вещественный	% клиентов сегмента от общего количества клиентов
Количество событий	9.0 Вещественный	Количество событий (покупок, транзакций, посещений и т.п.), совершенных клиентами сегмента
% кол-ва событий	9.0 Вещественный	% событий в сегменте от общего количества событий
Среднее кол- во событий	9.0 Вещественный	Среднее количество событий на клиента
Средняя давность (дни)	9.0 Вещественный	Средняя давность события в сегменте
Сумма	9.0 Вещественный	Сумма, потраченная клиентами сегмента

Метка	Тип	Описание
% суммы	9.0 Вещественный	% суммы по сегменту от общей суммы
Средняя сумма на клиента	9.0 Вещественный	Средняя сумма, которую тратит клиент сегмента
Средняя стоимость	9.0 Вещественный	Средняя стоимость события по сегменту

Структура таблицы "Интервалы"

Метка	Тип	Описание
Идентификатор интервалов	ab Строковый	Идентификатор интервала, 3 варианта: Recency, Frequency, Monetary
Номер интервала	12 Целый	Порядковый номер интервала (обратный порядок для Recency)
Нижняя граница	9.0 Вещественный	Нижняя граница диапазона открыта, первое значение границы — пустое
Верхняя граница	9.0 Вещественный	Верхняя граница диапазона открыта, последнее значение границы — пустое

Структура таблицы "Сообщения"

Метка	Тип	Описание
Идентификатор интервалов	ab Строковый	Идентификатор интервала
Сообщение	ab Строковый	Выводится предупреждение, если по показателю не удалось сформировать заданное на входе количество интервалов

Алгоритмы

$$\text{RFMScore} = \text{RecencyBin} \cdot \text{RecencyWeight} + \text{FrequencyBin} \cdot \text{FrequencyWeight} + \text{MonetaryBin} \cdot \text{MonetaryWeight}.$$

RFM-m

Назначение

Распределение клиентов по интервалам R, F, M и расчет RFM-оценки.
Количество интервалов и их границы известны заранее и подаются на вход компонента. Могут использоваться границы, рассчитанные в компоненте RFM-a.

В компоненте присутствует предварительная валидация границ. Если валидация не пройдена, выходные наборы данных сформированы не будут. Сообщение о причине будет выведено в соответствующем выходном порте.

Входные порты

Название	Тип
Статистики	Таблица
Интервалы	Таблица
Переменные	Переменные

Структура таблицы "Статистики"

Подаются данные с одноименного выхода компонента **LRFM-статистики**.

Метка	Тип	Описание
Идентификатор	ab Строковый	Уникальный идентификатор клиента
Давность	12 Целый	Количество дней, прошедшее между датой актуальности и датой последнего визита
Частота	12 Целый	Количество визитов клиента
Стоимость	9.0 Вещественный	Стоимость (общая, либо средняя сумма) транзакций клиента

Структура таблицы "Интервалы"

Структура таблицы описана в разделе [Требования к данным](#).

Переменные порта "Переменные"

№	Метка	Тип	Значение
1	Давность: вес	9.0 Вещественный	100
2	Частота: вес	9.0 Вещественный	10
3	Стоимость: вес	9.0 Вещественный	1

Давность: вес — используется для расчета RFM-оценки.

Частота: вес — используется для расчета RFM-оценки.

Стоимость: вес — используется для расчета RFM-оценки.

Выходные порты

Название	Тип
Сегменты	Таблица
Статистики по сегментам	Таблица
Сообщения	Таблица

Структура таблицы "Сегменты"

Структура набора соответствует одноименному набору на выходе компонента [RFM-a](#).

Структура таблицы "Статистики по сегментам"

Структура набора соответствует одноименному набору на выходе компонента [RFM-a](#).

Структура таблицы "Сообщения"

Структура набора соответствует одноименному набору на выходе компонента **RFM-a**.

Алгоритмы

$$\text{RFMScore} = \text{RecencyBin} \cdot \text{RecencyWeight} + \text{FrequencyBin} \cdot \text{FrequencyWeight} + \text{MonetaryBin} \cdot \text{MonetaryWeight}.$$

Архетипы RFM 5x5x5

Назначение

Компонент предназначен для распределения полученных RFM-сегментов по архетипам (разработаны компанией **Putler**):

- Чемпион;
- Лояльный;
- Потенциально лояльный;
- С потенциалом роста;
- Недавний;
- Требуется внимания;
- Засыпающий;
- В зоне риска;
- Нельзя потерять;
- Спящий;
- Потерянный.

Компонент можно использовать только для классической модели сегментации, в которой по каждому показателю сформировано пять интервалов, всего 125 сегментов. Внутри компонента проводится валидация на соответствие этим условиям. Если условия не соблюдены, выходные наборы данных не формируются, на выходе **Сообщения** выводится причина.

Входные порты

Название	Тип
Сегменты	Таблица
Архетипы	Переменные

Структура таблицы "Сегменты"

Подаются данные с одноименного выхода компонента **RFM-a** или **RFM-m**.

Переменные в порте "Архетипы"

Каждая переменная содержит соответствие **Код сегмента–Архетип**. Ниже приведены несколько, полный список можно увидеть на входе узла.

№	Метка	Тип	Значение
1	Чемпион	ab Строковый	555
2	Лояльный	ab Строковый	554
3	Потенциально лояльный	ab Строковый	553
...			
125	Потерянный	ab Строковый	111

Выходные порты

Название	Тип
Сегменты	Таблица
Статистики по сегментам	Таблица
Справочник архетипов	Таблица
Сообщения	Таблица

Структура таблицы "Сегменты"

Структура набора соответствует одноименному набору на выходе компонента **Архетипы RF 5x4**.

Структура таблицы "Статистики по сегментам"

Структура набора соответствует одноименному набору на выходе компонента **Архетипы RF 5x4**.

Структура таблицы "Справочник архетипов"

Структура набора соответствует одноименному набору на выходе компонента
Архетипы RF 5x4.

Структура таблицы "Сообщения"

Структура набора соответствует одноименному набору на выходе компонента
Архетипы RF 5x4.

FM-RFM-2

Назначение

Распределение клиентов по интервалам R, F и M, каждый показатель разбивается на 2 интервала. Возможен автоматический расчет (построение модели, 2 метода: по медиане и по среднему) и расчет по заданным границам ("прогон" через модель).

В компоненте присутствует предварительная валидация границ. Если валидация не пройдена, выходные наборы данных сформированы не будут. Сообщение о причине будет выведено в соответствующем выходном порте.

Кроме того в компоненте добавлены две коллекции архетипов для сегментов. Архетипы по FM-сегментам (Marcus, 1998):

- Транжира;
- Завсегдатай;
- Лучший;
- Неуверенный.

Архетипы по RFM-сегментам:

- Ушедший;
- Новый;
- Перспективный;
- Постоянный.

Проводится расчет статистик по архетипам. Соответствия **Код сегмента–Архетип** в данном случае заданы внутри компонента.

Входные порты

Название	Тип
Статистики	Таблица
Переменные	Переменные

Структура таблицы "Статистики"

Подаются данные с одноименного выхода компонента **LRFM-статистики**.

Метка	Тип	Описание
Идентификатор	ab Строковый	Уникальный идентификатор клиента
Давность	12 Целый	Количество дней, прошедшее между датой актуальности и датой последнего визита
Частота	12 Целый	Количество визитов клиента
Стоимость	90 Вещественный	Стоимость (общая, либо средняя сумма) транзакций клиента

Переменные порта "Переменные"

№	Метка	Тип	Значение
1	Автоматический расчет	0/1 Логический	true
2	Метод расчета	ab Строковый	med
3	Давность: граница	90 Вещественный	Пропущенное значение
4	Частота: граница	90 Вещественный	Пропущенное значение
5	Стоимость: граница	90 Вещественный	Пропущенное значение

Автоматический расчет — переменная может принимать следующие значения:

- **true** — границы определяются автоматически, метод задается в переменной **Метод расчета** (по умолчанию);
- **false** — границы задаются вручную в соответствующих переменных или подаются на вход.

Метод расчета — метод расчета границ при автоматическом расчете. Данная переменная может принимать следующие значения:

- **med** — медиана (по умолчанию);
- **avg** — среднее.

Давность: граница — граница интервала для давности.

Частота: граница — граница интервала для частоты.

Стоимость: граница — граница интервала для стоимости.

Выходные порты

Название	Тип
Сегменты	Таблица
FM 2x2	Таблица
RFM 2x2x2	Таблица
Сообщения	Таблица
Границы	Переменные

Структура таблицы "Сегменты"

Метка	Тип	Описание
Идентификатор	ab Строковый	Уникальный идентификатор клиента
Код FM	ab Строковый	Код сегмента клиента вида FM, см. Алгоритмы
Архетип FM	ab Строковый	Наименование архетипа, к которому относится FM-сегмент
Код RFM	ab Строковый	Код сегмента клиента вида RFM, см. Алгоритмы
Архетип RFM	ab Строковый	Наименование архетипа, к которому относится RFM-сегмент

Структура таблицы "FM 2x2"

Метка	Тип	Описание
Код сегмента	ab Строковый	Код вида FM
Архетип	ab Строковый	Наименование архетипа для FM-сегмента
Количество клиентов	12 Целый	Количество клиентов в сегменте
% клиентов	9.0 Вещественный	% клиентов сегмента от общего количества клиентов
Количество событий	9.0 Вещественный	Количество событий с клиентами сегмента
% кол-ва событий	9.0 Вещественный	% событий в сегменте от общего количества событий
Среднее кол-во событий	9.0 Вещественный	Среднее количество событий на клиента
Сумма	9.0 Вещественный	Сумма, потраченная клиентами сегмента
% суммы	9.0 Вещественный	% суммы по сегменту от общей суммы
Средняя сумма на клиента	9.0 Вещественный	Средняя сумма, которую тратит клиент сегмента
Средняя стоимость	9.0 Вещественный	Средняя стоимость события по сегменту

Структура таблицы "RFM 2x2x2"

Метка	Тип	Описание
Сегмент	ab Строковый	Наименование архетипа по RFM-сегментам
Количество клиентов	12 Целый	Количество клиентов в сегменте
% клиентов	9.0 Вещественный	% клиентов сегмента от общего количества клиентов
Количество событий	9.0 Вещественный	Количество событий с клиентами сегмента
% кол-ва событий	9.0 Вещественный	% событий в сегменте от общего количества событий
Среднее кол-во событий	9.0 Вещественный	Среднее количество событий на клиента
Средняя давность (дни)	9.0 Вещественный	Средняя давность события в сегменте
Сумма	9.0 Вещественный	Сумма, потраченная клиентами сегмента
% суммы	9.0 Вещественный	% суммы по сегменту от общей суммы
Средняя сумма на клиента	9.0 Вещественный	Средняя сумма, которую тратит клиент сегмента
Средняя стоимость	9.0 Вещественный	Средняя стоимость события по сегменту

Структура таблицы "Сообщения"

Метка	Тип	Описание
Идентификатор интервалов	ab Строковый	Идентификатор интервала
Сообщение	ab Строковый	Сообщение о результате обработки

Переменные в порте "Границы"

№	Метка	Тип	Описание
1	Метод расчета	ab Строковый	Если значение пустое, границы были заданы на входе
2	Давность: граница	9.0 Вещественный	Заданная на входе или рассчитанная граница интервала для давности
3	Частота: граница	9.0 Вещественный	Заданная на входе или рассчитанная граница интервала для частоты
4	Стоимость: граница	9.0 Вещественный	Заданная на входе или рассчитанная граница интервала для стоимости

Алгоритмы

Интервалы для показателей и коды сегментов формируются по следующим условиям:

Если **Давность** > **Давность: граница**, -, иначе **R**.

Если **Частота/Стоимость** < **Частота/Стоимость: граница**, -, иначе **F/M**.

LRFM

Назначение

Распределение клиентов по интервалам L, R, F и M, каждый показатель разбивается на 2 интервала. Возможен автоматический расчет (построение модели, 2 метода: по медиане и по среднему) и расчет по заданным границам ("прогон" через модель). Метод предложен [Chang и Tsay, 2004](#).

В компоненте присутствует предварительная валидация границ. Если валидация не пройдена, выходные наборы данных сформированы не будут. Сообщение о причине будет выведено в соответствующем выходном порте.

Кроме того в компоненте добавлены три коллекции архетипов для сегментов. Архетипы по LR-сегментам:

- Перспективный;
- Новый;
- Лояльный;
- Ушедший.

Архетипы по FM-сегментам:

- Транжира;
- Завсегдатай;
- Лучший;
- Неуверенный.

Архетипы по LRFM-сегментам:

- Ушедший;
- Новый;
- Убыточный;
- Перспективный;
- Лояльный.

Проводится расчет статистик по архетипам. Соответствия **Код сегмента–Архетип** в данном случае заданы внутри компонента.






Входные порты

Название	Тип
Статистики	Таблица
Переменные	Переменные

Структура таблицы "Статистики"

Подаются данные с одноименного выхода компонента **LRFM-статистики**.

Переменные порта "Переменные"

№	Метка	Тип	Значение
1	Автоматический расчет	 Логический	true
2	Метод расчета	 Строковый	med
3	Продолжительность: граница	 Вещественный	Пропущенное значение
4	Давность: граница	 Вещественный	Пропущенное значение
5	Частота: граница	 Вещественный	Пропущенное значение
6	Стоимость: граница	 Вещественный	Пропущенное значение

Автоматический расчет — переменная может принимать следующие значения:

- **true** — границы определяются автоматически, метод задается в переменной **Метод расчета** (по умолчанию);
- **false** — границы задаются вручную в соответствующих переменных или подаются на вход.

Метод расчета — метод расчета границ при автоматическом расчете. Данная переменная может принимать следующие значения:

- **med** — медиана (по умолчанию);
- **avg** – среднее.

Продолжительность: граница — граница интервала для продолжительности.

Давность: граница — граница интервала для давности.

Частота: граница — граница интервала для частоты.

Стоимость: граница — граница интервала для стоимости.

Выходные порты

Название	Тип
Сегменты	Таблица
LR 2x2	Таблица
FM 2x2	Таблица
LRFM 2x2x2x2	Таблица
Сообщения	Таблица
Границы	Переменные

Структура таблицы "Сегменты"

Метка	Тип	Описание
Идентификатор	ab Строковый	Уникальный идентификатор клиента
Код LR	ab Строковый	Код вида LR, см. Алгоритмы
Архетип LR	ab Строковый	Наименование архетипа, к которому относится LR-сегмент
Код LRFM	ab Строковый	Код вида LRFM
Архетип FM	ab Строковый	Наименование архетипа, к которому относится FM-сегмент (на основе LRFM-кода)
Архетип LRFM	ab Строковый	Наименование архетипа, к которому относится LRFM-сегмент

Структура таблицы "LR 2x2"

Метка	Тип	Описание
Архетип	ab Строковый	Наименование архетипа LR
Количество клиентов	12 Целый	Количество клиентов в сегменте
% клиентов	9.0 Вещественный	% клиентов сегмента от общего количества клиентов
Количество событий	9.0 Вещественный	Количество событий с клиентами сегмента
% кол-ва событий	9.0 Вещественный	% событий в сегменте от общего количества событий
Среднее кол-во событий	9.0 Вещественный	Среднее количество событий на клиента
Средняя давность (дни)	9.0 Вещественный	Средняя давность события в сегменте
Средняя продолжительность (дни)	9.0 Вещественный	Средняя продолжительность отношений с клиентами сегмента
Сумма (MonetarySum)	9.0 Вещественный	Сумма, потраченная клиентами сегмента
% суммы (MonetaryPercent)	9.0 Вещественный	% суммы по сегменту от общей суммы
Средняя сумма на клиента	9.0 Вещественный	Средняя сумма, которую тратит клиент сегмента
Средняя стоимость	9.0 Вещественный	Средняя стоимость события по сегменту

Структура таблицы "FM 2x2"

Метка	Тип	Описание
Архетип	ab Строковый	Наименование архетипа FM
Количество клиентов	12 Целый	Количество клиентов в сегменте
% клиентов	9.0 Вещественный	% клиентов сегмента от общего количества клиентов
Количество событий	9.0 Вещественный	Количество событий с клиентами сегмента
% кол-ва событий	9.0 Вещественный	% событий в сегменте от общего количества событий
Среднее кол-во событий	9.0 Вещественный	Среднее количество событий на клиента
Средняя давность (дни)	9.0 Вещественный	Средняя давность события в сегменте
Средняя продолжительность (дни)	9.0 Вещественный	Средняя продолжительность отношений с клиентами сегмента
Сумма	9.0 Вещественный	Сумма, потраченная клиентами сегмента
% суммы	9.0 Вещественный	% суммы по сегменту от общей суммы
Средняя сумма на клиента	9.0 Вещественный	Средняя сумма, которую тратит клиент сегмента
Средняя стоимость	9.0 Вещественный	Средняя стоимость события по сегменту

Структура таблицы "LRFM 2x2x2x2"

Метка	Тип	Описание
Архетип	ab Строковый	Наименование архетипа LRFM
Количество клиентов	12 Целый	Количество клиентов в сегменте
% клиентов	9.0 Вещественный	% клиентов сегмента от общего количества клиентов
Количество событий	9.0 Вещественный	Количество событий с клиентами сегмента
% кол-ва событий	9.0 Вещественный	% событий в сегменте от общего количества событий
Среднее кол-во событий	9.0 Вещественный	Среднее количество событий на клиента
Средняя давность (дни)	9.0 Вещественный	Средняя давность события в сегменте
Средняя продолжительность (дни)	9.0 Вещественный	Средняя продолжительность отношений с клиентами сегмента
Сумма	9.0 Вещественный	Сумма, потраченная клиентами сегмента
% суммы	9.0 Вещественный	% суммы по сегменту от общей суммы
Средняя сумма на клиента	9.0 Вещественный	Средняя сумма, которую тратит клиент сегмента
Средняя стоимость	9.0 Вещественный	Средняя стоимость события по сегменту

Структура таблицы "Сообщения"

Метка	Тип	Описание
Идентификатор интервалов	ab Строковый	Идентификатор интервала
Сообщение	ab Строковый	Сообщение о результате обработки

Переменные порта "Границы"

№	Метка	Тип	Описание
1	Метод расчета	ab Строковый	Если значение пустое, границы были заданы на входе
2	Продолжительность: граница	9.0 Вещественный	Заданная на входе или рассчитанная граница интервала для продолжительности
3	Давность: граница	9.0 Вещественный	Заданная на входе или рассчитанная граница интервала для давности
4	Частота: граница	9.0 Вещественный	Заданная на входе или рассчитанная граница интервала для частоты
5	Стоимость: граница	9.0 Вещественный	Заданная на входе или рассчитанная граница интервала для стоимости

Алгоритмы

Интервалы для показателей и коды сегментов формируются по следующим условиям:

Если **Давность** > **Давность: граница**, -, иначе **R**.

Если Продолжительность/Частота/Стоимость <
Продолжительность/Частота/Стоимость: граница, -, иначе L/F/M.

meta-RFM-a 5x5x5

Назначение

Метакомпонент строит RFM-модель с числом интервалов 5x5x5 и переводит их в архетипы. Метакомпонент построен на следующих базовых компонентах:

- LRFM-статистики
- RFM-a
- Архетипы RFM 5x5x5

Все входы и выходы соответствуют таковым в базовых компонентах.

meta-RFM-a 5x5x5

Назначение

Метакомпонент "прогоняет" через готовую RFM-модель с числом интервалов 5x5x5 и переводит их в архетипы. Метакомпонент построен на следующих базовых компонентах:

- LRFM-статистики
- RFM-m
- Архетипы RFM 5x5x5

Все входы и выходы соответствуют таковым в базовых компонентах.

Типовой кейс использования

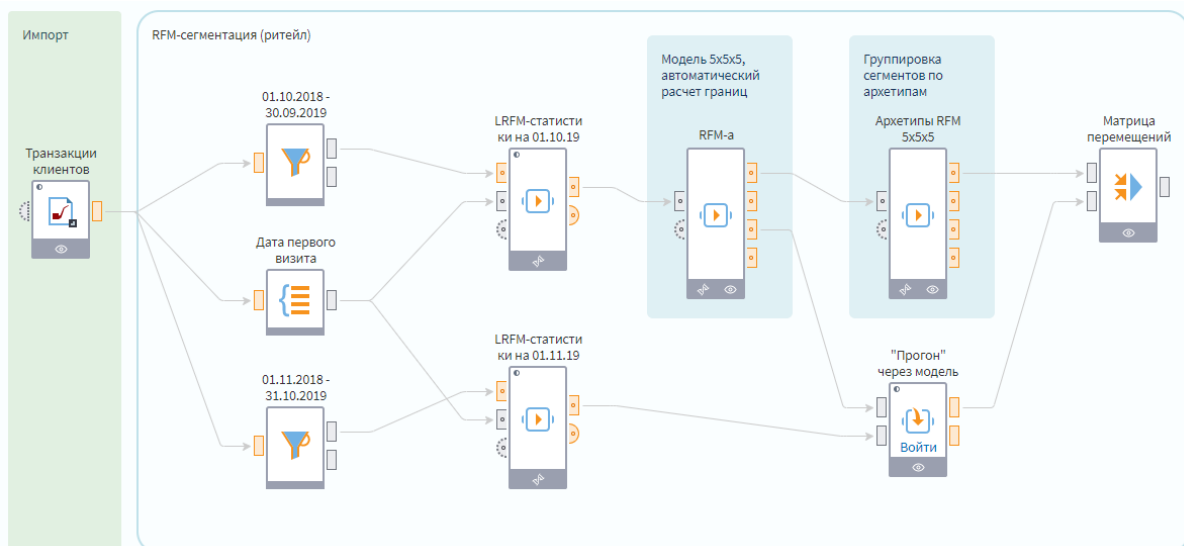
Ритейл. RFM-сегментация

Данный кейс включает в себя два этапа.

1. Построение модели RFM-сегментации клиентской базы 5x5x5 (деление каждого показателя на 5 интервалов) на основе данных по транзакциям клиентов за год: с **01.10.2018** по **30.09.2019** включительно.
2. "Прогон" транзакций со сдвигом на месяц: с **01.11.2018** по **31.10.2019** включительно — через полученную модель.

В пакете Кейсы использования.lgp в модуле **RFM-сегментация (ритейл)** используется четыре компонента из библиотеки Loginom RFM Kit:

- **LRFM-статистики** — компонент с первичными вычислениями;
- **RFM-a** — компонент для построения модели RFM-сегментации;
- **RFM-m** — компонент для "прогона" через модель;
- **Архетипы RFM 5x5x5** — компонент для разделения полученных 125 сегментов на 11 архетипов.



Этап 1. Построение модели

Прежде, чем приступить к расчету показателей R, F и M, отфильтровываются данные за нужный период. Также формируется набор с датами первого визита для каждого клиента (присутствуют в исходных данных).

Таким образом, на вход узла **LRFM-статистики** подается **81 155** транзакций по **16 043** клиентам за период с **01.10.2018** по **30.09.2019**.

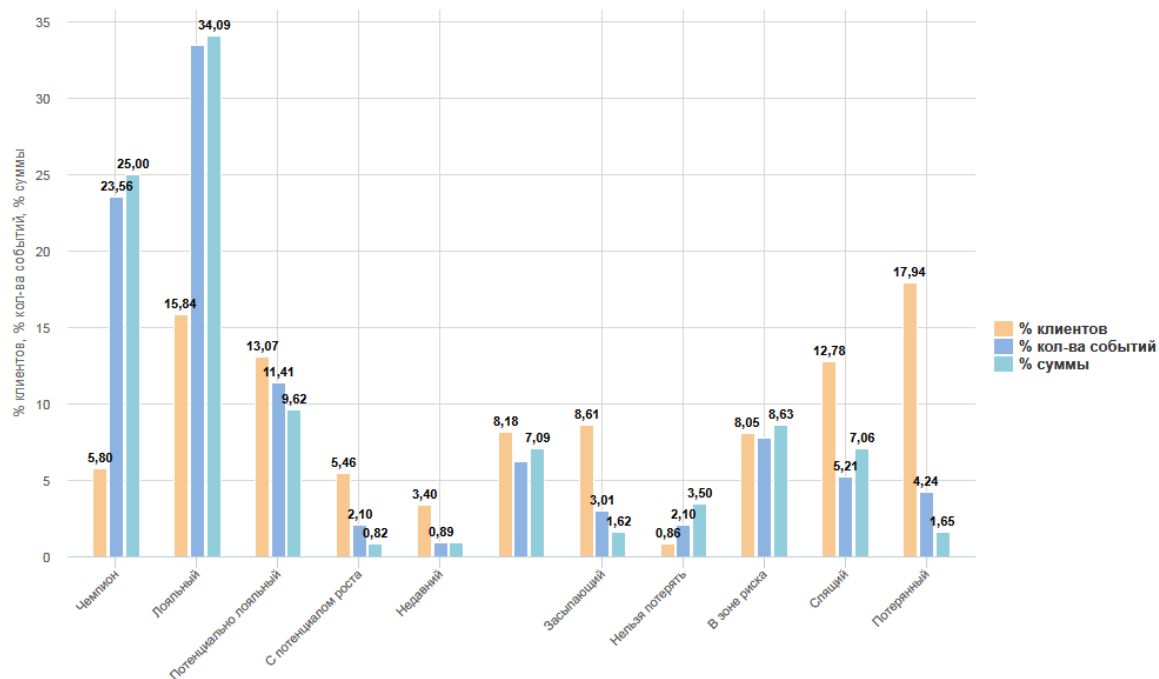
Далее рассчитанные статистики подаются на компонент **RFM-a** для построения модели сегментации. Значения переменных оставляем по умолчанию, в результате получаем следующую модель.

ab Идентификатор интервалов	12 Номер интервала	90 Нижняя граница	90 Верхняя граница
Recency	1	194,00	
Recency	2	111,00	194,00
Recency	3	53,00	111,00
Recency	4	18,00	53,00
Recency	5		18,00
Frequency	1		1,00
Frequency	2	1,00	2,00
Frequency	3	2,00	4,00
Frequency	4	4,00	7,00
Frequency	5	7,00	
Monetary	1		1 268,00
Monetary	2	1 268,00	3 248,00
Monetary	3	3 248,00	7 028,00
Monetary	4	7 028,00	15 600,00
Monetary	5	15 600,00	

На выходе RFM-сегментации получаем 125 сегментов. Для упрощения анализа используем компонент **Архетипы RFM 5x5x5**, где сегменты группируются по 11 архетипам, для которых рассчитывается ряд статистик.

Рассмотрим диаграмму распределения процентов по сегментам. Здесь отображены процент клиентов, который относится к каждому сегменту, процент событий, а также процент денег, которые принесли клиенты сегмента.

Легко заметить, что самый большой доход приносят клиенты сегментов **Чемпион** и **Лояльный**. Совсем новых клиентов (сегмент **Недавний**) довольно мало, возможно, их привлечению не уделяется много внимания. В то же время **Спящих** и **Потерянных** клиентов чуть больше 30%: это приемлемое значение, скорее всего, работа с клиентами в целом ведется достаточно активно.

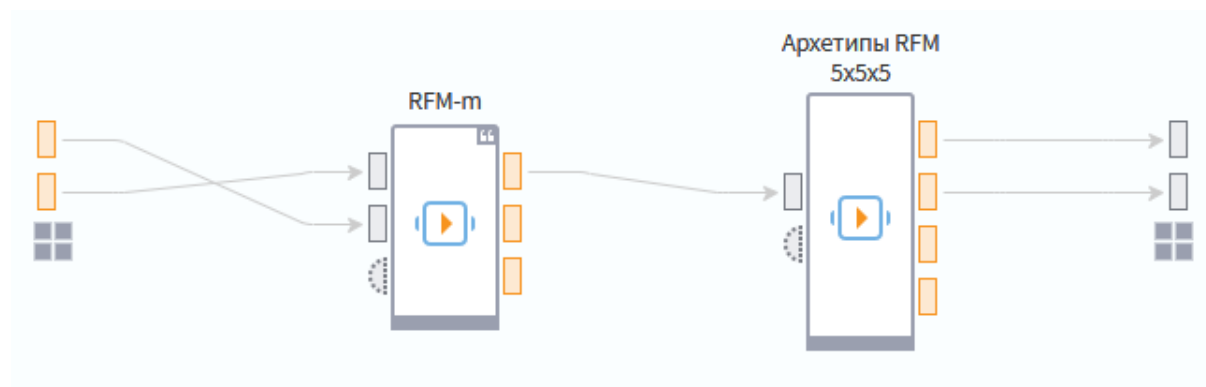


Чтобы убедиться, что выводы по сегментам сделаны верно, необходимо сдвинуть данные на период, "прогнать" клиентов через имеющуюся модель и посмотреть, как изменились статистики по сегментам.

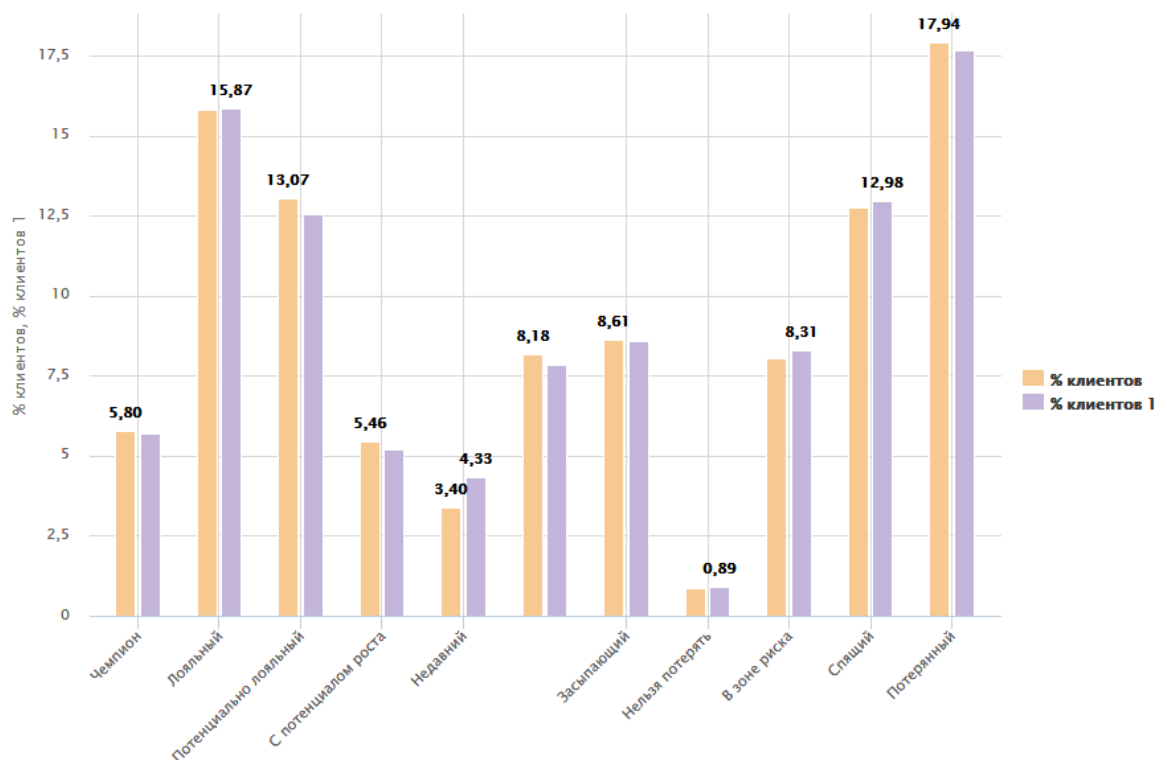
Этап 2. Прогон через модель

Здесь снова используется компонент **LRFM-статистики**. На этот раз на его вход подаются транзакции за период с **01.11.2018** по **31.10.2019**: **84 142** записи по **16 505** клиентам.

Далее данные передаются в подмодель **"Прогон" через модель**, которая содержит узлы выполнения, настроенные на компоненты **RFM-m** и **Архетипы RFM 5x5x5**. Кроме рассчитанных статистик на вход узла RFM-m подается полученная ранее модель сегментации.



После прогона через модель можно снова изучить полученные статистики по архетипам. Для того чтобы сравнить результаты двух сегментаций, используется узел слияния **Матрица перемещений**, где по полю **Архетип** соединяются результаты двух сегментаций. Благодаря этому можно проанализировать, как изменились значения статистик по сравнению с предыдущим периодом. На диаграмме ниже видно, что в целом распределение клиентов по сегментам за месяц практически не изменилось.



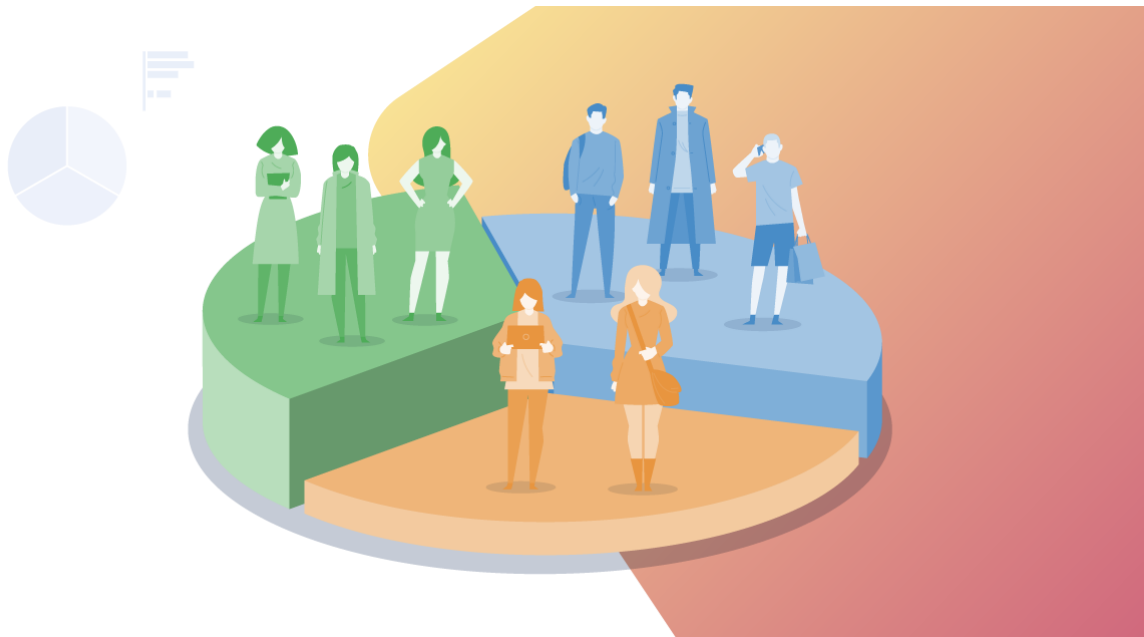
Таким образом, регулярно исследуя сегменты клиентов, можно следить за динамикой изменения клиентской базы и проверять эффективность маркетинговых воздействий.

Loginom Skills

Методология и кейсы использования компонентов рассматриваются в бизнес-курсе **Базовые методы сегментации клиентов**.

В курсе рассматриваются определение и различные виды сегментации, а также методология проведения сегментации базовыми методами: от одномерной Бинарной сегментации до многомерной LRFM-сегментации. Кроме теоретических сведений, все рассмотренные методы применяются на практике с использованием компонентов данной библиотеки.

Для успешного усвоения материалов курса необходимо иметь базовые навыки работы с платформой Loginom, которые рассматриваются в рамках курса **Аналитика данных low-code**.



История изменений документа

№	Дата	Описание
1	01.03.2020	Первая версия документации
2	02.07.2020	Изменение имени библиотеки
3	01.03.2021	Изменение версии библиотеки
4	05.05.2021	Изменение Loginom e-Learning на Loginom Skills
5	17.02.2022	Изменение версии библиотеки, правка документации
6	30.03.2023	Релиз 3.0 для версии Loginom 7.0