



Командное задание

Loginom Хакатон 2019

Вариант № 3

Входные данные

Прилагаются в архиве **clients_retail.zip** в форматах **CSV, TXT**.

Данные содержат транзакции клиентов розничной сети товаров для детей за определенный период времени, с указанием идентификатора клиента, даты и суммы.

Постановка задачи

Сегментация клиентов – это разделение клиентов на группы, каждая из которых обладает одним или несколькими отличительными признаками. Такой способ группировки позволяет создать стратегию продаж под каждый сегмент клиентов и увеличить эффективность бизнеса в целом.

Существует множество видов сегментации: от экспертной до автоматической. Развитие информационных систем и технологий регистрации первичных данных, а также бонусных программ на первый план выдвинуло поведенческую сегментацию, основанную на анализе транзакций, чеков (покупок).

Маркетологи нередко отдают предпочтение простым, но в то же время мощным инструментам сегментации клиентской базы. Этим объясняется популярность группы методов сегментации, известной как RFM. Эта аббревиатура от английских слов Recency, Frequency, Monetary — давность, частота, деньги. Суть этого метода рассказывается в десятках книг и статей в Интернете.

На самом деле, RFM – лишь частный метод из семейства методов, назовем их методами базовой сегментации, существуют методы RF-сегментации, LRFM-сегментации и другие. Общим у них является то, что в сегментации всегда присутствуют минимум два показателя – R (давность) и F (частота), а остальные добавляются в зависимости от подхода. Заметим, что бинарная сегментация, рассмотренная нами в материалах к Хакатону, еще проще, чем RFM.

Граничные значения показателей R, F и других задаются экспертно, либо методами равномерного квантования. В процедурах расчета всегда выделяется этап непосредственного расчета границ сегментов (при использовании квантования) и этап «прогона» новых клиентов через существующие сегменты с заданными границами.

Результат RFM-сегментации всегда есть некоторое конечное число сегментов, обычно несколько десятков. Для облегчения маркетологи часто объединяют их в 6-8 простых сегментов и называют понятными именами, например, «Транжира», «Лучший клиент» и так далее.

Мы предлагаем вам разработать **библиотеку компонентов методов базовой сегментации**. Библиотека функционально должна покрывать разнообразные вариации семейства RFM-методов, которые вы найдете в прилагаемой литературе либо в Интернете. Кроме того, большим плюсом будет добавить в компоненты готовые интерпретации сегментов, используя лучшие практики из литературы. Также, если вы найдете другие интересные методы сегментации, такие же простые, но в то же время мощные, как RFM, добавьте их в библиотеку.

Результат

- Опубликованные компоненты, решающие поставленные задачи.
- Краткое пояснение по использованию компонентов, в частности, список ограничений и требований к данным, если таковые имеются.
- Пример пакета, использующего компоненты.

Оценка результата

Основные критерии:

1. Корректность реализации, в том числе архитектура компонентов.
2. Применимость к любым данным от розничной сети.
3. Креативность.

Важно

Входные данные, прилагаемые к заданию, могут рассматриваться только как частный случай. Спроектированный компонент должен работать на любом наборе данных, удовлетворяющему требуемой структуре и/или возможным ограничениям по содержащимся в нем данным (должны быть обязательно озвучены вами при их наличии).

Также важно компонент сделать максимально настраиваемым, чтобы при желании пользователь смог при помощи переменных или входных справочников «тюнинговать» логику, заложенную внутри компонента.

Внимательно относитесь к процессу очистке данных. Сильные выбросы могут негативно сказаться на расчетах границ сегментов.

Литература

Мы даем подборку статей, в основном англоязычных, с описанием различных подходов к RFM-сегментации и лучшие практики по интерпретации и готовым сегментам.