

Компонент «Рекомендательная система»

Входные данные

- Набор данных, в котором по каждому клиенту с уникальным идентификатором содержатся записи о приобретенных им группах товаров, а также некий показатель, по которому можно определить предпочтаемую клиентом группу (например, количество или стоимость приобретенных товаров группы).
- Каталог товаров, где содержатся соответствия Группа товара–Товар.

Постановка задачи

Необходимо спроектировать компонент, который на входы получает указанные наборы данных и рассчитывает для каждого клиента его «любимую» группу товаров. «Любимой» считается группа, по которой суммарное значение показателя максимально. Посмотрите на таблицу, приведенную ниже. Клиент **001** приобретал в некоем магазине товары из трех групп. В данном примере в качестве показателя используется общая стоимость приобретенных товаров группы. Максимальное значение показателя имеет группа **Рубашки**, значит это «любимая» группа нашего клиента.

ID клиента	Группа товара	Показатель (Сумма, руб.)
001	Верхняя одежда	35 000
001	Рубашки	40 000
001	Кроссовки	15 000

Далее каждому клиенту нужно порекомендовать случайный товар из каталога, принадлежащий его «любимой» группе.

Результат

- Список из **N** клиентов, каждому из которых рекомендован товар из его «любимой» группы.
- Список из **N** клиентов, для каждого из которых указана «любимая» группа товаров.

Дополнительно

Для отладки и проверки работы компонента предоставлено два набора данных. Набор **transactions.lgd**, содержит историческую информацию по продажам и состоит из следующих полей:

- Дата** – дата продажи;
- ID клиента** – уникальный идентификатор клиента;
- Код товара** – код приобретенного товара;
- Количество** – количество приобретенных единиц товара;
- Сумма** – сумма, потраченная клиентом;
- Группа товара** – наименование группы, к которой относится товар.

Набор **catalog.lgd** представляет собой каталог наиболее популярных товаров:

- Группа товара** – наименование группы товара;
- Код товара** – код товара, принадлежащего группе.

Вопросы:

1. Сколько записей в выходном наборе данных?
2. Какой товар (указать код товара) порекомендуют клиенту CN1040246?
3. Какая любимая группа у клиента CN1041538?
4. У скольких клиентов **Любимой группой** является **Какао,кофе**?