

Описание данных

apparel-purchases

Данные о покупках клиентов по дням и по товарам. В каждой записи покупка определенного товара, его цена, количество штук.

В таблице есть списки идентификаторов, к каким категориям относится товар. Часто это вложенные категории (например автотовары-аксессуары-освежители), но также может включать в начале списка маркер распродажи или маркер женщинам/мужчинам.

Нумерация категорий сквозная для всех уровней, то есть 44 на второй позиции списка или на третьей – это одна и та же категория. Иногда дерево категорий обновляется, поэтому могут меняться вложенности, например ['4', '28', '44', '1594'] или ['4', '44', '1594']. Как обработать такие случаи – можете предлагать свои варианты решения.

- client_id идентификатор клиента
- quantity количество единиц товара
- price цена товара
- category_ids идентификаторы категорий
- date дата покупки
- message_id идентификатор сообщения из рассылки

apparel-messages

Рассылки, которые были отправлены клиентам из таблицы покупок.

- bulk_campaign_id идентификатор рассылки
- client_id идентификатор клиента
- message_id идентификатор сообщения
- event действие с сообщением (отправлено, открыто, покупка...)
- channel канал рассылки
- date дата действия
- created_at дата-время полностью

target

- client_id идентификатор клиента
- target клиент совершил покупку в целевом периоде

Общая база рассылок огромна, поэтому собрали для вас агрегированную по дням статистику по рассылкам. Если будете создавать на основе этой статистики дополнительные признаки, обратите внимание, что нельзя суммировать по колонкам `nunique`, потому что это уникальные клиенты в пределах дня, у вас нет данных, повторяются ли они в другие дни)

full_campaign_daily_event

Агрегация общей базы рассылок по дням и типам событий

- `date` дата
- `bulk_campaign_id` идентификатор рассылки
- `count_event*` общее количество каждого события `event`
- `nunique_event*` количество уникальных `client_id` в каждом событии

* в именах колонок найдете все типы событий `event`

full_campaign_daily_event_channel

Агрегация по дням с учетом событий и каналов рассылки

- `date` дата
- `bulk_campaign_id` идентификатор рассылки
- `count_event*_channel*` общее количество каждого события по каналам
- `nunique_event*_channel*` количество уникальных `client_id` по событиям и каналам

* в именах колонок есть все типы событий `event` и каналов рассылки `channel`