

prototype3628800

Школа бэкенд-разработки 2021 (осень)

11 сен 2021, 15:11:52

старт: 11 сен 2021, 10:55:06

финиш: 11 сен 2021, 16:55:06

до финиша: 01:43:10

длительность: 06:00:00

Е. Статистика

Ограничение времени	12 секунд
Ограничение памяти	512.0 Мб
Ввод	стандартный ввод или input.txt
Вывод	стандартный вывод или output.txt

Одной команде Яндекс Маркета предстояло решить ответственную задачу. На основе статистических данных необходимо определить самые продаваемые товары за последние несколько месяцев.

Описание таблиц базы данных

Статистика представляет собой таблицу **statistic** в базе данных, в которой содержатся колонки:

- `date` - день, за который формировалась статистика - тип `date`
- `name` - название товара - тип `text`
- `description` - идентификатор описания товара - тип `text`
- `category_id` - идентификатор категории товара - тип `integer`
- `brand_id` - идентификатор бренда товара - тип `integer`
- `shows` - число показов на витрине Маркета - тип `integer`
- `clicks` - число переходов по ссылке на конкретный товар - тип `integer`
- `orders` - число успешно доставленных товаров - тип `integer`
- `gmv` - сумма денежных средств, вырученная с продажи определенного товара - тип `integer`

Что такое **GMV**?

Пример строки таблицы **statistic**:

date	name	description	category_id	brand_id	shows	clicks	orders	gmv
2021-05-22	Телевизор HARPER 32R490T 32 (2020)	739589	90639	11169208	1083	4	2	12796

Из строки можно понять, что за 22-е мая 2021 года товар с названием "Телевизор HARPER 32R490T 32 (2020)" был показан покупателям 1083 раза, 4 раза на страницу товара заходили по клику, и 2 раза этот товар был успешно продан на общую сумму 12796 рублей.

Также известно, что он принадлежит бренду с идентификатором 11169208 и находится в категории по идентификатору 90639.

Для получения названий брендов и категорий используется две таблицы **brands** и **categories**, каждая из которых содержит колонки

- `id` - идентификатор - тип `integer`
- `name` - название категории/бренда - тип `text`

Таблица **brands**

id	name
90555	Philips

Таблица **categories**

id	name
14694813	Наушники

Гарантируется, что таблицы **brands** и **categories** содержат все необходимые идентификаторы.

Критерии получения данных

Необходимо вывести **первые 100** строк, описывающие динамику продаж товаров, которые удовлетворяют необходимым требованиям.

1. Период исследуемой статистики - **[2021-04-01, 2021-07-31]**, обе границы включительно.
2. Для более объективного анализа необходимо получить общую сумму продаж **каждого товара за каждый месяц** исследуемой статистики - это **Временной ряд**. Месяц определяется первым днём этого месяца. Для даты 2021-04-15 месяцем является дата 2021-04-01.
3. Необходимо исключить некорректные данные. Данные некорректны, если за один день число показов меньше, чем число кликов, или число кликов меньше, чем число успешно выполненных заказов.
4. Часто в статистике фигурируют товары, которые продаются намного дороже других. Поэтому требуется исключить из обрабатываемых данных **категории** "Варочные панели" и "Садовые измельчители", а также **бренды** "Apple" и "Samsung".

сформированном датом агрегации, суммарном количестве показов, кликов, заказов, а также бренду, описанию и названию товара.

5. Результат запроса необходимо **отсортировать** по двум параметрам

1. по дате агрегации(первому числу месяца) по **возрастанию**;
2. по **gmv** за период агрегации по **убыванию**.

6. Для соблюдения консистентности, результат запроса необходимо **сгруппировать** в том числе по описанию, названию товара, названию бренда и названию категории.

Результат запроса должен содержать колонки

- monthDate - месяц, за который была выполнена агрегация
- description - идентификатор описания товара
- item_name - название товара
- category_name - название категории
- brand_name - название бренда
- sum_shows - суммарное количество показов за период агрегации
- sum_clicks - суммарное количество кликов за период агрегации
- sum_orders - суммарное количество успешно выполненных заказов за период агрегации
- sum_gmv - суммарная выручка за период агрегации

Пример

Исходная таблица statistic

Номер строки	date	name	description	category_id	brand_id	shows	clicks	orders	gmv
1	2021-03-23	Ноутбук Lenovo ThinkPad	784011	91013	152981	423	233	33	324434
2	2021-04-14	Ноутбук Lenovo ThinkPad	784011	91013	152981	622	524	44	232413
3	2021-04-14	Ноутбук Lenovo ThinkPad	784011	91013	152981	562	233	45	232413
4	2021-06-14	Ноутбук Lenovo ThinkPad	784011	91013	152981	226	211	23	644464
5	2021-06-14	Ноутбук MGN73RU/A	68686	91013	439977	456	311	193	1025446
6	2021-06-14	Ноутбук Lenovo ThinkPad	784011	91013	152981	175	179	43	334434
7	2021-08-02	Ноутбук Lenovo ThinkPad	784011	91013	152981	135	136	23	345434
8	2021-07-15	Ноутбук Lenovo ThinkPad	784011	91013	152981	165	43	13	343443
9	2021-05-14	Ноутбук APPLE MacBook Pro 13	730972	91013	153043	143	34	23	3434
10	2021-07-17	Ноутбук APPLE MacBook Air 13	793137	91013	153043	3833	453	43	234333
11	2021-04-14	Ноутбук APPLE MacBook Air 13	793137	91013	153043	394	233	32	2342443
12	2021-08-14	Варочная панель Maunfeld	702254	237418	7085802	16	13	12	23432334
13	2021-04-20	Варочная панель ПВИ 4000 GEFEST	708146	237418	152771	11	11	10	2342343

Результат запроса

monthDate	description	item_name	category_name	brand_name	sum_shows	sum_clicks	sum_orders	sum_gmv
2021-04-01	784011	Ноутбук Lenovo ThinkPad	Ноутбуки	Lenovo	1184	757	89	464826

В поле отправки необходимо вставить корректный SQL-запрос для базы данных SQLite, выполняющий поставленную задачу. В качестве компилятора необходимо выбрать "SQLite 3.31.1". Запрос будет выполнен на реальной базе данных, после чего результат будет проверен на корректность и соответствие поставленной задаче.

Вердикт **RE** в данной задаче может означать следующие проблемы:

- невалидный с точки зрения SQLite запрос (любая ошибка, которая не позволяет базе данных выполнить запрос). Сюда относятся ошибки в синтаксических конструкциях, неверные имена таблиц, колонок и т.д.
- слишком большой размер ответа (превышает настройки системы).

Вердикт **WA** означает, что ваше решение является синтаксически верным на языке SQLite (язык способен его запустить и вычислить результат), но результат запроса не совпал с ожидаемым. В том числе слишком большой вывод, но уместающийся в ограничения системы, будет помечен вердиктом WA, а не RE.

Гарантируется, что ответ в данной задаче определен однозначно при выполнении всех требований условия.

[Отправить](#)[Предыдущая](#)[Следующая](#)

Газетерем предложенный пример

1. Строки номер 1 и номер 7 исключаются из результата, так как не попадают в заданный интервал [2021-04-01, 2021-07-31]
2. Строки под номерами 9,10, 11 исключаются, так как принадлежат бренду Apple
3. Строки с номерами 12 и 13 исключаются, так как входят в категорию Варочных панелей
4. Строка 6 исключается, так как содержит некорректные данные (число кликов больше чем число показов)
5. Остаются строки под номерами 2,3,4,5,8
6. Строки 2 и 3 агрегируются, так как относятся к одному товару и попадают в общий период агрегации - месяц апрель, сумма GMV равна $232413 + 232413 = 464826$
7. Строки 4 и 5 попадают в период агрегации - июнь
8. Строка 8 попадает в период агрегации - июль

© 2013–2021 ООО «Яндекс»

Язык Python 3.7 (PyPy 7.3.3)

[Набрать здесь](#)[Отправить файл](#)

1