

Проект в BQ - sbermegamarket-test

Датасет - test_sql

В базе данных информация о клиентах и оформленных на сайте заказах хранится в нескольких таблицах (см. описание ниже).

1. Написать запрос, который показывает количество выполненных заказов с X SKU в заказе (шт.)

X - число различных (уникальных) SKU в заказе; X принимает значения из множества {1, 2, 3, 4, 5 ...}.

Пример результата запроса:

Колво уникальных SKU	Количество заказов
1	698
2	662
3	726
5	689

2. Написать SQL-запрос, выводящий среднюю стоимость покупки (завершенный заказ) за все время клиентов из центрального региона ("Central"), совершивших и получивших первую покупку в январе 2018 года. Результаты предоставить в разбивке по городам.

Пример результата запроса:

CityID	AvgPurchase
1	6 948
3	7 662
4	7 266
6	6 489

3. По месяцам вывести топ-3 самых покупаемых (по количеству единиц товаров в выкупленных заказах) SKU. Если у нескольких товаров одинаковое количество проданных единиц, то выводить все такие товары.

Пример результата запроса:

MONTH	SKU
-------	-----

1	1101466277
1	1101342018
1	1101133432
...	

4. Проанализировать качество когорт пользователей (когортой считать пользователей, оформивших заказ в одном месяце) по основным маркетинговым метрикам (ARPPU, средний чек, частота, ретеншн в следующий месяц, любые другие метрики на свое усмотрение). Оформить в виде небольшого отчета в excel. Форма представления данных и визуализации свободная. Предполагаем, что задача звучит следующим образом: проанализировать качество и состав когорт. Предложить выводы и гипотезы, что могло на это повлиять.

Описание таблиц

Таблица Orders – данные о заказах:

1. OrderID – идентификатор заказа, целое типа int.
2. CustomerID – идентификатор клиента, целое типа int.
3. OrderDate – дата оформления заказа, тип nchar(8), формат записи - "YYYYMMDD".
4. OrderState – состояние заказа, строка типа nvarchar(9).
Возможные значения: "Fulfilled" (исполнен) и "Cancelled" (отменен).
5. DeliveryDays – количество дней от оформления заказа до получения его клиентом, целое типа tinyint.
Значение 0 означает получение в день оформления, 1 – на следующий день и т.д. Для отмененных заказов – NULL. Датой покупки считается дата получения заказа клиентом.

Первичный ключ таблицы – поле OrderID.

Таблица Order_List** – состав заказов:**

1. OrderID – идентификатор заказа, целое типа int.
2. SKU – идентификатор товара, целое типа int.
3. Quantity – количество заказанного товара, целое типа tinyint.
4. Price – стоимость одной единицы товара, целое типа int.

Первичный ключ таблицы – комбинация полей OrderID и SKU.

Таблица Customers – справочник клиентов.

1. CustomerID – идентификатор клиента, целое типа int.
2. CityID – идентификатор города проживания клиента, целое типа int.
Данные о городе клиента могут отсутствовать; в этом случае считать городом проживания клиента CityID = 1.

Первичный ключ таблицы – поле CustomerID.

Таблица City_Region** – справочник регионов:**

1. CityID – идентификатор города, целое типа int.
2. Region – название региона, строка типа nvarchar(7). Возможные значения: "Central", "North", "South", "East", "West".

Первичный ключ таблицы – поле CityID.